Probabilités et statistique pour l'agrégation externe

Jean-Baptiste BARDET, Florent MALRIEU

Table des matières

Chapitre 1. Introduction 5
1. Programme 5
2. Modélisation 6
3. Simulation 6
4. Deux mots sur l'épreuve 7

Chapitre 2. Simulation de variables aléatoires 9
1. Inversion de la fonction de répartition 9
2. Simulation de lois discrètes 10
3. Et avec Scilab (premiers pas ? 10
4. Quelques théorèmes limites 11
5. Méthode du rejet 12
6. Variables aléatoires gaussiennes 13
7. Vecteurs aléatoires gaussiens 14
8. Méthode de Monte-Carlo 14
9. Mélange de lois 15
10. Mesures de Gibbs 16
11. Le théorème de Wigner 17
12. Corrigés 18

Chapitre 3. Martingales 21
1. Probabilités, espérances et lois conditionnelles 22
2. Martingales : définition, théorème d'arrêt 26
3. Théorèmes de convergence des martingales 29
4. Exercices 32
5. Corrigés 35
A. Deux lemmes analytiques 41

Chapitre 4. Chaînes de Markov 43
1. Matrices de transition et chaînes de Markov 43
2. Équation de Chapman-Kolmogorov 46
3. Propriété de Markov 47
4. Récurrence/transience. Irréductibilité 48
5. Mesures invariantes, mesures réversibles 49
6. Exemples 52
7. Exercices 53
8. Corrigés 56

Chapitre 5. Statistiques d'ordre, processus de Poisson 61
1. Lois exponentielles 61
2. Statistiques d'ordre 63
3. Processus de Poisson 64
4. Corrigés 68

Chapitre 6. Chaînes de Markov à temps continu 71
1. Définitions et généralités 71
2. Processus de naissance 73
3. Classes 74
4. Temps d'atteinte et d'absorption 75
5. Récurrence et transience 75
6. Mesures invariantes 76
7. Convergence à l'équilibre et théorème ergodique 77
8. Processus de vie et mort 78
9. Exercices 79
10. Corrigés 80

Chapitre 7. Quelques notions d'estimation paramétrique 81
1. Introduction et notations 81
2. Méthode des moments 82
3. Estimateur du maximum de vraisemblance 82
4. Exercices 84
5. Corrigés 85

Chapitre 8. Quantiles empiriques 89
1. Théorème de Glivenko-Cantelli 89
2. Convergence des quantiles empiriques 90
3. Théorème de Kolmogorov-Smirnov 91
4. Tests associés 92
5. Corrigés 94

Chapitre 9. Intervalles de confiance 97
1. Définitions et premier exemple 97
2. Les lois gaussiennes 98
3. Les grands échantillons 100
4. Intervalles de confiance pour une proportion 102
5. Utilisation des quantiles empiriques 102
6. Inégalité de Chernov 102
7. Corrigés 104

Chapitre 10. Tests du chi-deux 105
1. Un peu de probabilités 106
2. Test d'adéquation à une loi donnée 108
3. Test d'adéquation à une famille de lois 110
4. Test d'indépendance 111
5. Test d'homogénéité 112

Chapitre 11. Modèle linéaire gaussien 113
1. Travailler et capitaliser pour produire ! 113
2. Les données 115
3. Estimation des coefficients de régression 115
4. Décomposition de la variance et le coefficient de corrélation multiple R 116
5. Loi des estimateurs sous le modèle gaussien 117
6. Intervalles de confiance 117
7. Intervalle de prédiction 118

Bibliographie 119

Chapitre 1

Introduction

1. Programme

1.1. À l'écrit.

(1) Définition d'un espace probabilisé : événements, tribus, mesure de probabilité. Indépendance d'événements et de tribus. Loi du 0-1, lemmes de Borel-Cantelli.

(2) Probabilités conditionnelles : définition, formule des probabilités totales et théorème de Bayes.

(3) Variables aléatoires, loi d'une variable aléatoire : loi discrète et loi absolument continue. Fonction de répartition et densité.

(4) Exemples de variables aléatoires : variable de Bernoulli, binomiale, de Poisson, uniforme, exponentielle, de Gauss.

(5) Espérance et variance d'une variable aléatoire à valeurs réelles, théorème de transfert.

(6) Indépendance de variables aléatoires. Loi conditionnelle d'une variable par rapport à une autre.

(7) Transformations exponentielles de lois : fonction caractéristique, transformée de Laplace, fonction génératrice. Liens avec l'indépendance et la convolution, application aux sommes de variables aléatoires indépendantes.

(8) Convergences de suites de variables aléatoires : en probabilité, dans L^p, presque sûrement, en loi.

(9) Inégalité de Markov, inégalité de Bienaymé-Tchebychev. Loi faible des grands nombres, applications en statistiques.

(10) Théorème de Lévy, théorème central limite, applications en statistiques.

1.2. À l'oral.

(1) Utilisation de lois usuelles pour modéliser certains phénomènes aléatoires. Exemples : temps d'attente ou durée de vie, erreurs de mesure, sondages... Méthodes de simulation de variables aléatoires.

(2) Chaînes de Markov à espace d'états finis. Classification des états. Convergence vers une loi stationnaire (théorème ergodique et théorème central limite admis). Chaînes de Markov homogènes à espace d'états dénombrable, transience, récurrence positive ou nulle, exemple de la marche aléatoire simple.

(3) Lois de Poisson, exponentielle et Gamma, construction et propriétés du processus de Poisson sur $\mathbb{R}_+$.

(4) Espérance conditionnelle, définition des martingales, temps d'arrêt. Exemples d'utilisation des théorèmes de convergence presque sûre et L^2 des martingales à temps discret.

(5) Échantillons, moments empiriques, loi et fonction de répartition empiriques.

(6) Applications des théorèmes de convergences à l'estimation (lois des grands nombres, théorème central limite, utilisation du lemme de Slutsky). Définition et construction d'intervalles de confiance.

(7) Estimation paramétrique. Estimation par maximum de vraisemblance : définition et exemples.

(8) Vecteurs gaussiens : définition, simulation en dimension 2, théorème de Cochran. Théorème central limite dans $\mathbb{R}^n$.

(9) Modèle linéaire gaussien : calculs par moindres carrés, régression linéaire simple ou multiple, exemples d'utilisation.

(10) Tests paramétriques (test du rapport de vraisemblance). Tests d'ajustement (tests du χ^2, tests de Kolmogorov-Smirnov). Exemples d'utilisation.

2. Modélisation

La modélisation mathématique est l'articulation entre un phénomène physique, biologique ou autre et des objets mathématiques. On cherche ainsi à capturer les comportements observés ou supposés par des mécanismes mathématiques.
Un bon modèle mathématique a donc deux qualités antinomiques : il est suffisamment complexe pour bien décrire les phénomènes observés et suffisamment simple pour se prêter à une étude mathématique précise. Le mathématicien aura ainsi plutôt tendance à proposer un mécanisme très simple avant de le confronter aux données. S'il y a une adéquation acceptable, c'est gagné ! Dans le cas contraire, il faudra prendre en compte certains de phénomènes précédemment négligés.
Un bon exemple de cette démarche est la description des interactions entre solides. Pour expliquer la marche des planètes (et même découvrir Pluton) et calculer les coefficients de marée, la mécanique newtonienne suffit. Pas besoin donc d'aller chercher plus loin. Elle doit pourtant être abandonnée au profit de la relativité pour décrire l'évolution des infiniment grand et petit. Cependant, il est totalement superfétatoire d'utiliser la relativité pour calculer les coefficients de marée.
Un autre exemple classique est la loi de Hardy-Weinberg en génétique des populations (voir le chapitre 10). Toutes les hypothèses permettant d'établir cette loi d'évolution sont clairement irréalistes dans une population réelle. Cependant, elle est très souvent vérifiée (par exemple entre gauchers et droitiers ou pour les groupes sanguins). Ainsi, si un test statistique (voir le chapitre 10) permet de remettre en cause la loi, c'est qu'un mécanisme biologique très fort est à l'œuvre (mutation, sélection, immigration...).

3. Simulation

La simulation est un passage obligé de l'épreuve. Elle peut se faire avec n'importe lequel des logiciels disponibles le jour de l'épreuve. Notons que depuis la session 2015, seuls des logiciels libres sont utilisables. Ce cours est illustré par des programmes en Scilab.
Les notions de programmation nécessaires à l'épreuve de modélisation sont assez modestes. Il faut maîtriser les boucles (for et while), la création de fonctions, la génération de nombres aléatoires et l'affichage de courbes.
Dans la mesure du possible, il est bon de proposer des simulations dans lesquelles on peut faire varier les paramètres pour illustrer les différents comportements possibles du modèle.
Dans le chapitre 2, quelques commandes de base sont passées en revue. Tout au long du fascicule, des simulations illustrent les résultats théoriques. Les fonctions utilisées pour réaliser ces figures sont à votre disposition et peuvent servir d'exemples. Au fur et à

mesure de la lecture, il est conseillé de réfléchir à une façon d'utiliser la simulation pour accompagner les théorèmes étudiés, et de prendre le temps de les implémenter pendant la préparation : l'utilisation d'un logiciel nécessite un peu d'habitude pour pouvoir s'en servir pendant la préparation à l'oral sans perdre trop de temps...

Une autre remarque importante sur l'utilisation de ce fascicule est qu'il n'est pas nécessaire de le lire linéairement : la partie « Statistique » (Chapitres 7 à 11) est en grande partie indépendante de la partie « Probabilités » (Chapitres 3 à 6). Inversement, la notion d'estimateur d'un paramètre, notion clef de statistique inférentielle, apparaît assez tôt dans le document, bien qu'elle ne soit définie qu'au chapitre 7. Le lecteur ne doit donc pas hésiter à s'investir dans ces diverses parties, et dans l'apprentissage de Scilab, en parallèle.

4. Deux mots sur l'épreuve

Commençons par rappeler que les rapports de jury des années précédentes (en ligne sur le site de l'épreuve www.agreg.org) apportent de nombreuses informations sur l'organisation de l'épreuve, les attentes du jury et les erreurs à éviter.
L'épreuve de modélisation dure une heure après quatre heures de préparation. C'est à la fois long (il faut arriver à garder une concentration et une efficacité importantes pendant cinq heures) et court (découvrir un texte, une situation nouvelle et construire un exposé cohérent dans la foulée n'est pas une mince affaire). Comme pour les autres épreuves, l'entrainement est capital. Il est également très important pour le candidat d'avoir le soutien de quelques livres couvrant le programme (il est parfois nécessaire de réviser les hypothèses de certains théorèmes ou de se remémorer une définition classique) : nous fournissons pour cela une bibliographie étoffée à la fin de ce cours.
En début d'épreuve, le candidat tire deux textes. Contrairement aux deux autres épreuves orales (où l'on sait vite quelle est la leçon la plus favorable), il est face à l'inconnu. Il est important de prendre dix à quinze minutes pour explorer un peu en détail les DEUX textes avant de faire son choix. Encore une fois, découvrir un texte et « entrer dedans » s'apprend. Il est important de répéter cet exercice régulièrement (de nombreux textes sont en ligne sur différentes pages de préparation à l'agrégation et sur le site de l'épreuve).
La plupart des textes sont fortement « teintés » probabilités ou statistique. Tous les couplages sont possibles. Il ne faut donc pas faire d'impasse sur l'un ou l'autre des pans du programme. D'autant plus que le jury, dans la partie questions, interrogera le candidat sur des notions qui n'ont pas nécessairement été abordées dans l'exposé.

Chapitre 2

Simulation de variables aléatoires

Un des enjeux importants des probabilités appliquées est de pouvoir simuler grâce à un ordinateur l'aléa présent dans les modèles et notamment (nous verrons pourquoi dans la suite) une suite de variables aléatoires indépendantes et de même loi uniforme sur $[0,1]$. Ceci est impossible pour plusieurs raisons : l'ordinateur ne manipule qu'un nombre fini d'entiers et ne peut exécuter que des algorithmes déterministes. Il est toutefois possible de construire des suites récurrentes sur un ensemble fini qui ressemblent de manière satisfaisante à la suite cherchée. L'un des algorithmes les plus efficaces est le Mersenne Twister (sa période est notamment de $2^{19937}-1$).
Dans la suite, nous supposerons toujours disposer d'une suite $(U_n)_{n\geq 1}$ de variables aléatoires indépendantes de même loi uniforme sur $[0,1]$.

1. Inversion de la fonction de répartition

Soit X une variable aléatoire réelle de fonction de répartition F définie par $F(x) = \mathbb{P}(X \leq x)$ pour $x \in \mathbb{R}$.
La fonction F est croissante, continue à droite, admet en tout point une limite à gauche et le nombre de points de discontinuité de F est fini ou dénombrable.
La fonction F n'est en général pas une bijection mais, en tant que fonction croissante, elle admet une *fonction inverse généralisée* F^- définie par

$$\forall y \in [0,1], \quad F^-(y) = \inf\{x \in \mathbb{R}\,;\, F(x) \geq y\}.$$

Cette fonction coïncide avec F^{-1} lorsque F est bijective. La méthode de simulation, dite d'inversion (sous-entendu de la fonction de répartition) s'appuie sur le résultat suivant.

PROPOSITION 2.1. *Si U est une variable aléatoire de loi uniforme sur $[0,1]$, alors $F^-(U)$ admet F pour fonction de répartition.*

DÉMONSTRATION. Comme F est croissante et continue à droite, on a, pour tout $u \in]0,1[$

$$\{x \in \mathbb{R}\,;\, F(x) \geq u\} = [F^-(u); +\infty[\,.$$

On en déduit que, pour $u \in]0,1[$ et $x \in \mathbb{R}$, $F^-(u) \leq x$ si et seulement si $u \leq F(x)$. Par conséquent

$$\mathbb{P}(F^-(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x)\,.$$

□

EXEMPLE 2.2. On peut ainsi simuler des v.a. de loi exponentielle (de densité $\lambda e^{-\lambda x}\mathbf{1}_{\{x>0\}}$), de loi de Cauchy (de densité $1/(\pi(1+x^2))$), de loi de Rayleigh (de densité $xe^{-x^2/2}\mathbf{1}_{\{x>0\}}$), de loi de densité $2x^{-3}\mathbf{1}_{\{x>1\}}\ldots$

Remarquons que la proposition 2.1 admet une réciproque, mais sous des hypothèses plus fortes, qui jouera un rôle essentiel dans l'étude de la fonction de répartition empirique et des théorèmes de Glivenko-Cantelli et Kolmogorov-Smirnov (voir le chapitre 8).

PROPOSITION 2.3. *Si la fonction de répartition F d'une v.a.r. X est continue, alors $F(X)$ suit la loi uniforme sur $[0,1]$.*

DÉMONSTRATION. Comme F est supposée continue (donc continue à gauche), on peut adapter l'argument précédent, en définissant $F^+(y) = \sup\{x \in \mathbb{R}\,;\ F(x) \le y\}$, qui vérifie cette fois-ci que, pour $u \in]0,1[$ et $x \in \mathbb{R}$, $x \le F^+(u)$ si et seulement si $F(x) \le u$. Ainsi

$$\mathbb{P}(F(X) \le u) = \mathbb{P}(X \le F^+(u)) = F(F^+(u))$$

et on conclut en remarquant qu'à nouveau par continuité de F, $F(F^+(u)) = u$ (on a plus précisément $F^{-1}(u) = [F^-(u); F^+(u)]$, faire un dessin peut aider). □

2. Simulation de lois discrètes

La méthode d'inversion permet aussi de simuler les lois discrètes. Commençons par des mesures sur un ensemble fini $\{x_1, \dots, x_k\}$. En général (et c'est ce que fait Scilab), on se place sur $\{1, \dots, k\}$.

EXERCICE 2.1. Soit $\mu = p_1\delta_1 + \cdots + p_k\delta_k$ une loi de probabilité sur $\{1, \dots, k\}$. Montrer que la variable aléatoire suivante a pour loi μ :

$$X = \mathbf{1}_{\{U<p_1\}} + 2\mathbf{1}_{\{p_1 \le U < p_1+p_2\}} + \cdots + k\mathbf{1}_{\{p_1+\cdots+p_{k-1} \le U < 1\}}.$$

Pour simuler une variable aléatoire chargeant un ensemble dénombrable, on peut utiliser l'exercice 2.1 à l'aide d'une boucle while. Du point de vue de l'efficacité algorithmique, il est plus astucieux de renommer les points afin que les poids respectifs soient décroissants (la boucle sera exécutée moins souvent en moyenne).
Cependant, pour certaines lois classiques, comme la loi géométrique ou la loi de Poisson, il existe des astuces plus efficaces.

EXERCICE 2.2 (Loi géométrique). Si X suit une loi exponentielle de paramètre λ, quelle est la loi de $[X]$ (où $[x]$ désigne la partie entière de x) ? En déduire un algorithme de simulation de la loi géométrique de paramètre $p \in]0,1[$ qui affecte à $k \in \mathbb{N}^*$ le poids $p(1-p)^{k-1}$ grâce à une variable aléatoire uniforme sur $[0,1]$.

EXERCICE 2.3 (Loi de Poisson). Soit $(T_k)_{k\ge1}$ une suite de variables aléatoires indépendantes de même loi $\mathscr{E}(\lambda)$. On note $S_0 = 0$ et, pour $n \ge 1$, $S_n = T_1 + \cdots + T_n$. De plus, on définit $N = \sum_{k=1}^{\infty} \mathbf{1}_{\{S_k \le 1\}}$. Montrer que S_n suit la loi Gamma de densité $\lambda e^{-\lambda x}\frac{(\lambda x)^{n-1}}{(n-1)!}\mathbf{1}_{\{x>0\}}$ pour $n \ge 1$. Montrer que $\{S_n \le 1\} = \{N \ge n\}$. Calculer $\mathbb{P}(N = n)$. En déduire que N suit une loi de Poisson de paramètre λ. On pourra se reporter au chapitre 5 pour la démonstration de ces propriétés.

3. Et avec Scilab (premiers pas ?

Tous les logiciels de simulation possèdent un générateur de nombres aléatoires pour les lois classiques. Pour Scilab, la fonction s'appelle grand (taper help grand pour une description des fonctionnalités). Attention, dans Scilab, les lois gaussiennes sont paramétrées par leur écart-type et les lois exponentielles par leur espérance. La commande

```
histplot(30,grand(1,1000,'exp',2))
```

trace un histogramme de 30 classes pour un échantillon de taille 1000 de v.a. i.i.d. de loi $\mathscr{E}(1/2)$.
Voici un petit guide pour se familiariser avec les commandes de base de Scilab sur les matrices, l'affichage et les variables booléennes. Cette dernière notion permet d'éviter de nombreuses boucles. Exécuter les commandes en suivant.
Fabrication manuelle d'une matrice (nuance avec ou sans le " ;" final) :

```
M=[1 3 5 ;-2 6 1]
```

Extraction de la première ligne ou de la troisième colonne :

l=M(1, :) ; c=M(:,3)
Transposée de M :
N=M'
Fabrication automatique d'un vecteur :
[0 :0.1 :2]
Fabrication automatique d'une matrice de zéros :
P=zeros(4,5)
►► Que fait la commande A=ones(M) ?
Suppression de la deuxième colonne :
P(:,2)=[ ]
Fabrication d'une matrice par blocs :
x=[1 :2] ;P=[x 2*x ;3*x 4*x]
►► Fabriquer une matrice 3×4 qui contient M en haut à gauche, x' en haut à droite, x en bas à gauche et [0 0] en bas à droite, en utilisant la construction par blocs.
►► Que calcule la commande s=ones(x)'*x ?
►► Si x et y sont deux vecteurs (colonne), comment calculer leur produit scalaire ?
►►Affichage propre d'un résultat :
p=3.14 ; disp('une valeur approchée de pi = '+string(p)+' au centième près.')
►►Vecteurs booléens (pratiques pour trier) :
l=(M>2)
►► Expliquer la commande suivante :
x=[-2 :.5 :2] ;l=(x>-1 & x<1) ;x(l)=3.14 ; x
►► D'autres commandes à tester :
a=sum(x), a=mean(x), x.^2, sort(x), -sort(-x), (x>0), M.^2, M^2
Rappel : pour créer et utiliser une fonction, taper le texte de la fonction dans un fichier grâce à l'éditeur scipad de Scilab. Une fonction commence par function nom_fonction et se termine par endfunction. Il est possible (et souhaitable) de mettre plusieurs fonctions dans un même fichier. La sauvegarde se fait avec un Contrôle-s et la compilation avec un Contrôle-l. Dans la fenêtre Scilab, appeler la fonction. Attention : si la fonction n'utilise pas d'arguments, tapez nom_fonction().

4. Quelques théorèmes limites

THÉORÈME 2.4 (Loi forte des grands nombres). *Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. i.i.d. On note $\bar{X}_n = (X_1 + \cdot + X_n)/n$ la moyenne des n premiers termes de la suite.*

(1) Si X_1 est intégrable alors $(\bar{X}_n)_{n\geq 1}$ converge presque sûrement vers $\mathbb{E}(X_1)$.

(2) Si $(\bar{X}_n)_{n\geq 1}$ converge presque sûrement vers $a \in \mathbb{R}$ alors X_1 est intégrable et $\mathbb{E}(X_1) = a$.

EXERCICE 2.4. Que font les commandes Scilab suivantes :

(1) plot2d(cumsum(rand(1,1000))./[1 :1000])

(2) plot2d([1 :1000],cumsum(rand(1,1000))./[1 :1000],4)

(3) clf() ; plot2d(cumsum(-log(rand(1,1000)))./[1 :1000])

(4) plot2d([1 :1000]',cumsum(grand(1000,10,'nor',0,1),'r')./([1 : 1000]'*ones(1,10)))

EXERCICE 2.5. Illustrer le point 2 de la loi des grands nombres en simulant la suite des moyennes empiriques pour des v.a. de loi de Cauchy (grâce à la méthode de l'inversion de la fonction de répartition).

THÉORÈME 2.5 (Théorème limite central). *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. i.i.d. de carré intégrable. On note $\sigma^2 = \mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2$ la variance de X_1. Alors*

$$\sqrt{n}(\bar{X}_n - \mathbb{E}(X_1)) \xrightarrow[n\to\infty]{\mathscr{L}} \mathscr{N}(0,\sigma^2).$$

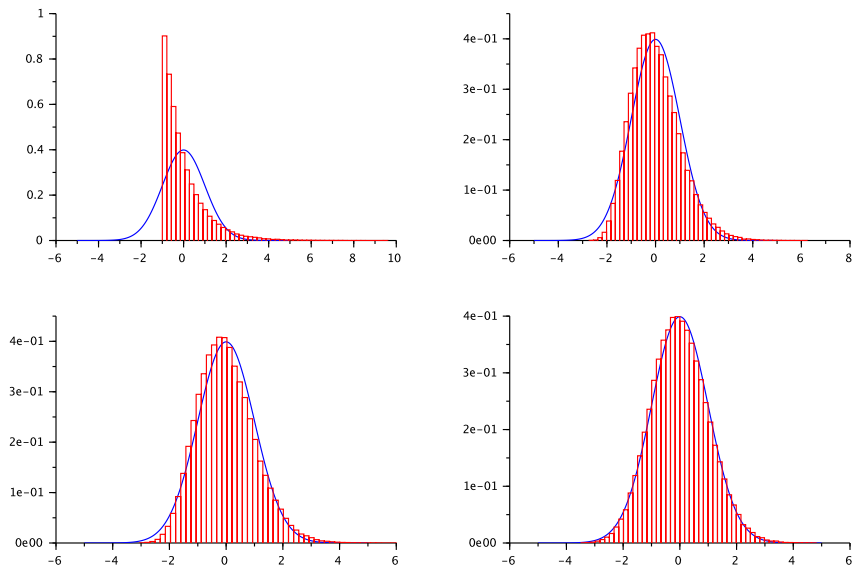

FIGURE 1. Histogrammes de 10000 réalisations de $\sqrt{n}(\bar{X}_n - 1)$ pour un échantillon de loi $\mathscr{E}(1)$ pour $n = 1$, $n = 10$, $n = 20$ et $n = 50$ comparés à la densité de la loi $\mathscr{N}(0,1)$.

Pour la présentation de ces deux théorèmes fondamentaux, leurs démonstrations et quelques applications, on pourra consulter le fascicule de probabilités.

5. Méthode du rejet

Très souvent se pose le problème de simuler des variables de loi uniforme sur un sous-ensemble de $\mathbb{R}^d$. Lorsqu'il est borné, il existe une façon très simple de procéder.

PROPOSITION 2.6 (Simulation de lois uniformes). *Soit λ_d la mesure de Lebesgue sur $\mathbb{R}^d$, D et D' deux boréliens de R^d tels que*

$$D \subset D' \quad \text{et} \quad 0 < \lambda_d(D) \le \lambda_d(D') < +\infty.$$

Soit X un point aléatoire de loi uniforme sur D'. Alors la loi conditionnelle de X sachant que $X \in D$ est la loi uniforme sur D.

EXERCICE 2.6 (Loi uniforme sur le disque). Décrire un algorithme qui fournit une variable aléatoire de loi uniforme sur le disque unité dans $\mathbb{R}^d$ à partir des variables aléatoires de loi uniforme sur le pavé $[-1,1]^d$ (voir la figure 2). Combien d'itérations en moyenne sont nécessaires ? Quelle est la loi du nombre d'itérations ?

EXERCICE 2.7 (Loi uniforme sur la sphère). Voici deux façons d'obtenir un point uniforme sur la sphère unité.
1. Comment utiliser l'exercice 2.6 pour simuler une v.a. uniforme sur la sphère unité ?
2. Soit $Y_1,\dots,Y_d$ des v.a. i.i.d. de loi $\mathscr{N}(0,1)$. Montrer que la loi de $Y = (Y_1,\dots,Y_d)$ est invariante par rotation (pour toute matrice orthogonale O, OY et Y ont même loi). En déduire la loi de $Y/\|Y\|_2$.

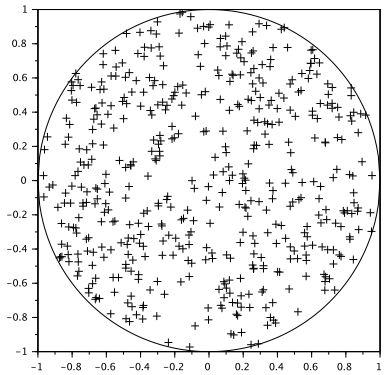

FIGURE 2. 500 réalisations de la loi uniforme sur la boule unité euclidienne de $\mathbb{R}^2$ par la méthode de rejet suggérée dans l'exercice 2.6.

THÉORÈME 2.7. *Soit μ une loi de densité f sur $\mathbb{R}$.*

(1) Supposons que f soit continue, à support compact $[a,b]$ et bornée par M (le graphe de f est donc contenu dans le pavé $[a,b]\times[0,M]$). Soit $((X_n,Y_n))_{n\geq 1}$ une suite de vecteurs aléatoires indépendants de même loi uniforme sur $[a,b]\times[0,M]$. On définit

$$T=\inf\{n\geq 1 : f(X_n)\geq Y_n\}.$$

Alors T est fini presque sûrement, X_T et T sont indépendantes et suivent respectivement la loi μ et la loi géométrique de paramètre $1/(M(b-a))$.

(2) Supposons qu'il existe une densité g facile à simuler telle que $f\leq cg$ pour une constante $c>0$. Soit alors $(W_n)_{n\geq 1}$ et $(U_n)_{n\geq 1}$ deux suites de variables aléatoires indépendantes de lois respectives de densité g et uniforme sur $[0,1]$.

$$T=\inf\{n\geq 1 : f(W_n)\geq cU_n g(W_n)\}$$

Alors W_T et T sont indépendants de lois respectives de densité f et la loi géométrique de paramètre $1/c$.

Un algorithme de rejet est d'autant meilleur que le rapport entre l'aire sous la densité f et l'aire du rectangle (ou l'aire sous cg) est petit puisque c'est l'inverse du nombre moyen de couples qu'il faudra générer (l'espérance de la loi géométrique de paramètre p vaut $1/p$). Il existe de nombreuses astuces pour construire des domaines efficaces. On pourra consulter [**Yca02**] pour la preuve de ces résultats et leur mise en pratique sur des exemples.

6. Variables aléatoires gaussiennes

On dit que X suit la loi gaussienne $\mathcal{N}(m,\sigma^2)$ si sa loi admet pour densité la fonction

$$\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

Dans ce cas, m et σ^2 sont respectivement la moyenne et la variance de X et $(X-m)/\sigma$ suit la loi $\mathcal{N}(0,1)$. Cette stabilité par translation et homothétie permet de ramener le problème à la simulation de variables aléatoires de loi $\mathcal{N}(0,1)$.

Pour générer des variables aléatoires de loi gaussienne, on ne peut pas utiliser la méthode d'inversion puisque l'on ne dispose pas d'une primitive de la densité gaussienne. Il faut donc utiliser des moyens détournés. L'un des plus célèbres s'appelle l'algorithme de Box-Muller.

EXERCICE 2.8 (Algorithme de Box-Muller). Montrer que si U et V sont deux v.a.r. indépendantes de lois respectives exponentielle de paramètre $1/2$ et uniforme sur $[0,1]$ alors X et Y définies par

$$\begin{cases} X = \sqrt{U}\cos(2\pi V) \\ Y = \sqrt{U}\sin(2\pi V) \end{cases}$$

sont indépendantes et de même loi $\mathcal{N}(0,1)$.

Citons encore un autre algorithme souvent présenté dans les ouvrages traitant de simulation. Il mêle rejet et changement de variables.

EXERCICE 2.9 (Algorithme polaire-rejet). Soit $((U_n, V_n))_{n\in\mathbb{N}^*}$ des v.a. i.i.d. de loi uniforme sur $[-1,1]$. On note $T = \inf\{n \geq 1,\ U_n^2 + V_n^2 \leq 1\}$ et $R^2 = U_T^2 + V_T^2$. Montrer que les variables aléatoires

$$\begin{cases} X = \sqrt{-2\ln(R^2)/R^2}\,U_T \\ Y = \sqrt{-2\ln(R^2)/R^2}\,V_T \end{cases}$$

sont indépendantes et de même loi $\mathcal{N}(0,1)$.

Un algorithme est d'autant plus efficace qu'il ne fait appel qu'à un petit nombre d'opérations compliquées au rang desquelles se trouvent les générations de v.a. et les évaluations de fonctions comme ln, cos sin... Pour une comparaison des temps de calculs des algorithmes, on pourra consulter [**Yca02**].

7. Vecteurs aléatoires gaussiens

Dans toute la suite on notera A^T la matrice transposée de A. Un vecteur (colonne) aléatoire $X = (X_1, \dots, X_d)^T$ est dit gaussien si toute combinaison linéaire de ses coordonnées est une variable aléatoire (réelle) gaussienne. Il ne suffit pas que $X_1, \dots, X_d$ le soient (contre-exemple ?). La loi d'un vecteur gaussien est caractérisé par son (vecteur-)espérance $m = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_d))^T$ et sa matrice de covariance Γ dont les coefficients sont définis par

$$\Gamma_{ij} = \mathrm{Cov}(X_i, X_j) = \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j) = (\mathbb{E}(XX^T) - m\,{}^t m)_{ij}.$$

On note $\mathcal{N}_d(m, \Gamma)$ la loi de X.

EXERCICE 2.10 (Décomposition de Cholesky et vecteurs gaussiens). Montrer que la matrice Γ est symétrique positive. Quelle contrainte sur les coefficients de Γ l'inégalité de Cauchy-Schwarz impose-t-elle ? Montrer que toute matrice symétrique positive Γ peut s'écrire sous la forme $\Gamma = AA^T$ avec A triangulaire inférieure et que, de plus, A est inversible ssi Γ est définie positive. Soit $Y = (Y_1, \dots, Y_d)^T$ un vecteur aléatoire gaussien centrée de matrice de covariance I_d. Quelle est la loi de $AY + m$?

8. Méthode de Monte-Carlo

La méthode de Monte-Carlo fournit des valeurs approchées d'intégrales en utilisant des réalisations i.i.d. d'une loi que l'on sait simuler. Pour cela, on interprète tout simplement une intégrale comme une espérance, que l'on va approcher par la moyenne empirique de v.a. bien choisies. Par exemple, si $(X_n)_{n\geq 1}$ est une suite de v.a. i.i.d. de loi uniforme sur $[0,1]^m$ et si $f : [0,1]^m \to \mathbb{R}$ est une fonction intégrable (par rapport à la mesure de Lebesgue sur $[0,1]^m$), alors la loi des grands nombres entraîne la convergence presque-sure suivante :

$$\frac{1}{n}\big(f(X_1) + \cdots + f(X_n)\big) \underset{n\to\infty}{\longrightarrow} \mathbb{E}\big(f(X_1)\big) = \int_{[0,1]^m} f(x)\,dx \quad \text{p.s.}$$

Si l'on sait majorer la variance de $f(X_1)$, on est en mesure de contrôler l'erreur en donnant un intervalle de confiance asymptotique. La vitesse de convergence de cette méthode est

$\sqrt{n}$ (ce qui peut paraître lent par rapport à des méthodes déterministes). Cependant cette vitesse ne dépend pas de la régularité de f et dépend plus faiblement de la dimension m que les méthodes déterministes.

EXERCICE 2.11. Utiliser une méthode de Monte-Carlo pour évaluer la valeur de π à partir des trois intégrales suivantes (on précisera la mesure de probabilité par rapport à laquelle on intègre) :

(1) $\displaystyle\int_0^1 4\sqrt{1-x^2}\,dx$ (aire du disque unité)

(2) $\displaystyle\int_{[-1,+1]^2} \mathbf{1}_{\{x^2+y^2\leq 1\}}\,dx\,dy$ (aire du disque unité)

(3) $\displaystyle\int_{[-1,+1]^3} \mathbf{1}_{\{x^2+y^2+z^2\leq 1\}}\,dx\,dy$ (volume de la boule unité).

EXERCICE 2.12. Soit f la fonction définie sur $\mathbb{R}^3$ par

$$f(x,y,z) = \sin(x)e^{x^2+3y} + \cos(z^4).$$

On note I l'intégrale de f par rapport à la mesure uniforme sur la sphère unité S^2 de $\mathbb{R}^3$. Donner une valeur approchée et un intervalle dans lequel se trouve I avec une probabilité de l'ordre de 0,95.

9. Mélange de lois

La notion de mélange est très importante en probabilités, statistique et modélisation. Imaginons que l'on veuille modéliser la taille des Français (hommes et femmes) adultes. Notons m_f et m_h les tailles moyennes respectives et σ_f^2 et σ_h^2 leurs variances. En vertu du théorème limite central, il semble naturel de modéliser la taille d'une Française prise au hasard par une loi $\mathcal{N}(m_f,\sigma_f^2)$. Ainsi, si p désigne la proportion de femmes dans la population totale, la loi de la taille d'un individu pris au hasard dans la population sera donnée par le mélange de paramètre p des lois $\mathcal{N}(m_f,\sigma_f^2)$ et $\mathcal{N}(m_h,\sigma_h^2)$. On la note $p\mathcal{N}(m_f,\sigma_f^2)+(1-p)\mathcal{N}(m_h,\sigma_h^2)$. Sa densité est donnée par

$$\frac{p}{\sqrt{2\pi\sigma_f^2}}\exp\left(-\frac{(x-m_f)^2}{2\sigma_f^2}\right)+\frac{1-p}{\sqrt{2\pi\sigma_h^2}}\exp\left(-\frac{(x-m_h)^2}{2\sigma_h^2}\right).$$

EXERCICE 2.13. Interpréter en terme de mélange de lois la densité suivante :

$$f(x)=\begin{cases}\frac{1}{3}e^{-x}+\frac{4}{3}e^{-2x} & \text{si } x>0,\\ 0 & \text{sinon.}\end{cases}$$

Simuler des variables aléatoires indépendantes admettant cette densité.

EXERCICE 2.14 (Estimation par la méthode des moments). On suppose que l'on observe des v.a. i.i.d. $X_1,\dots,X_n$ de loi $p\mathcal{N}(a,1)+(1-p)\mathcal{N}(-a,1)$ avec $a>0$ et $p\in]0,1[$ inconnus (voir la figure 3 pour l'allure de la densité de la loi de X_1). Calculer $\mathbb{E}(X_1)$ et $\mathbb{E}(X_1^2)$ en fonction de a et p. Comment estimer $\mathbb{E}(X_1)$ et $\mathbb{E}(X_1^2)$ en fonction de $X_1,\cdots,X_n$? En déduire une procédure d'estimation des paramètres a et p (voir le chapitre 7 pour l'aspect statistique).

On a ici mélangé deux lois de probabilité mais on peut aussi en mélanger plus, jusqu'à une famille non dénombrable, par le procédé suivant.

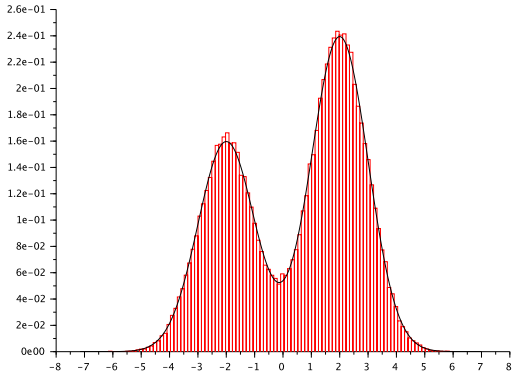

FIGURE 3. Histogramme de taille $n = 10^5$ et densité pour la loi mélange de l'exercice 2.14 pour $a = 2$ et $p = 0.6$.

DÉFINITION 2.8. Soit $(\mu_\theta)_{\theta\in\Theta}$ une famille de probabilités sur $(\Omega, \mathscr{A})$ et ν une probabilité sur Θ. On appelle mesure-mélange de $(\mu_\theta)_\theta$ de poids ν la mesure définie par

$$\forall A \in \mathscr{A}, \quad \mu(A) = \int_\Theta \mu_\theta(A)\nu(d\theta).$$

L'algorithme permettant de simuler une v.a. de loi μ est le suivant : on génère une v.a. X de loi ν puis une v.a. Y de loi μ_X indépendante de X : Y suit alors la loi ν.

EXERCICE 2.15. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. i.i.d. de fonction caractéristique φ et N une v.a. indépendante de la suite précédente à valeurs dans $\mathbb{N}$ de fonction génératrice G définies sur un espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$. On définit la v.a. Z par

$$\forall \omega \in \Omega, \quad Z(\omega) = \begin{cases} 0 & \text{si } N(\omega) = 0, \\ \displaystyle\sum_{n=1}^{N(\omega)} X_n(\omega) & \text{si } N(\omega) \geq 1. \end{cases}$$

Exprimer la fonction caractéristique de Z grâce à φ et G. En déduire son espérance lorsqu'elle existe. Quelle est la loi de Z lorsque X_1 suit la loi de Bernoulli $\mathscr{B}(p)$ et N la loi de Poisson $\mathscr{P}(\lambda)$? Quel est le rapport avec le mélange de lois ?

10. Mesures de Gibbs

Soit E un ensemble fini, H une fonction de E dans $\mathbb{R}$ et β un paramètre positif. On souhaite simuler une variable aléatoire de loi ν_β dont la densité par rapport à la mesure de comptage sur E est proportionnelle à $e^{-\beta H}$, c'est-à-dire que pour tout $\eta \in E$,

$$\nu_\beta(\eta) = \frac{1}{Z_\beta} e^{-\beta H(\eta)} \quad \text{avec} \quad Z_\beta = \sum_{\sigma\in E} e^{-\beta H(\sigma)}.$$

REMARQUE 2.9. Lorsque $\beta \to 0$, ν_β tend vers la mesure uniforme sur E. Lorsque $\beta \to +\infty$, ν_β tend vers la mesure uniforme sur les points de E où H atteint son minimum.

EXEMPLE 2.10 (Métal ferromagnétique). Soit $\Lambda = \{1, 2, \ldots, n\}^d$. Les points de Λ sont appelés des sites (ils modélisent les positions des atomes). À chaque site x du réseau Λ on associe un spin $\eta(x) \in \{-1, +1\}$ et on note $\eta = (\eta(x))_{x\in\Lambda}$ une configuration. L'ensemble

$E = \{-1,+1\}^\Lambda$ est l'ensemble de toutes les configurations. Les spins interagissent avec leurs voisins sur le réseau. Pour $\eta \in E$, on note

$$H(\eta) = -\sum_{x\sim y} \eta(x)\eta(y) \quad \text{où} \quad x \sim y \text{ ssi } \left|x-y\right| = 1.$$

La quantité $H(\eta)$ est d'autant plus petite que les sites voisins ont des valeurs de spin qui coïncident.

Lorsque E est grand, il est impossible de calculer la constante de normalisation Z_β et donc de simuler E à partir de la méthode naïve évoquée dans la section 2. La solution viendra des chaînes de Markov (voir le chapitre 4).

11. Le théorème de Wigner

Ce résultat est culturel et sert à manipuler des matrices (il est hors du programme de l'agrégation). Considérons un tableau symétrique de taille infinie dont les coefficients sont des v.a.r. :

$$\begin{pmatrix} w_{11} & w_{12} & \cdots \\ w_{12} & w_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

On note W_n la matrice aléatoire carrée **symétrique** réelle de taille $n \times n$ obtenue en tronquant le tableau à partir de son coin supérieur gauche. Les valeurs propres de W_n sont réelles.

THÉORÈME 2.11 (Wigner). *Supposons que les v.a. $(w_{ij}, 1 \le i < j)$ sont i.i.d.* ***centrées*** *de variance finie σ^2 et que les v.a. $(w_{ij}, 1 \le i)$ sont i.i.d.* ***centrées****. Si $(\lambda_{n,1}(\omega), \dots, \lambda_{n,n}(\omega))$ désigne le spectre de $n^{-1/2}W_n(\omega)$, alors, pour presque tout ω, la mesure empirique*

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_{i,n}(\omega)}$$

converge faiblement vers la loi du demi-cercle de paramètre σ, de densité par rapport à la mesure de Lebesgue dx sur $\mathbb{R}$:

$$\frac{1}{2\pi\sigma^2}\sqrt{4\sigma^2 - x^2}\,\mathbf{1}_{\{-2\sigma \le x \le 2\sigma\}}(x).$$

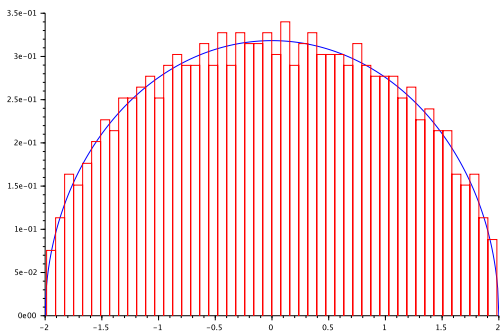

FIGURE 4. Illustration du théorème de Wigner avec une matrice carrée de 1000 lignes à coefficients gaussiens standards.

REMARQUE 2.12. On pourra utiliser les commandes Scilab triu(M,1) et triu(M) pour créer une matrice symétrique et spec pour en déduire le spectre.

12. Corrigés

CORRIGÉ DE L'EXERCICE 2.2. La variable aléatoire $[X]$ est à valeurs dans $\mathbb{N}$ et, pour tout $k \in \mathbb{N}$,

$$\mathbb{P}([X] = k) = \int_k^{k+1} \lambda e^{-\lambda t} dt = e^{-\lambda k} - e^{-\lambda(k+1)} = e^{-\lambda k}(1 - e^{-\lambda}).$$

Ainsi, la variable aléatoire $[X] + 1$ suit la loi géométrique de paramètre $p = 1 - e^{-\lambda}$. Soit U de loi uniforme sur $[0, 1]$. Alors

$$N = 1 + \left[\frac{\ln(U)}{\ln(1-p)}\right]$$

puisque $-\ln(U)$ suit la loi exponentielle de paramètre 1.

CORRIGÉ DE L'EXERCICE 2.6. Si $(U_n)_{n\geq 1}$ est une suite de v.a. i.i.d. de loi uniforme sur le pavé (les coordonnées sont indépendantes de même loi uniforme sur $[-1, 1]$), l'algorithme consiste à retenir le premier élément de la suite dont la norme euclidienne est inférieure à 1 (à l'aide d'une boucle while). Si l'on note V_d le volume de la boule unité de $\mathbb{R}^d$ alors la probabilité que U_n soit dans cette boule est $V_d 2^{-d}$: le nombre d'itérations suit la loi géométrique avec ce paramètre (qui est l'inverse de l'espérance).

CORRIGÉ DE L'EXERCICE 2.7.
1. Si X suit la loi uniforme sur le disque unité alors $X/\|X\|$ suit la loi uniforme sur la sphère.
2. Le vecteur OY est un vecteur gaussien centré de matrice de covariance I_d (voir les propriétés des vecteurs gaussiens dans le fascicule de probabilités). On conclut comme dans la question précédente. Notons que cette seconde méthode est bien plus rapide puisque l'on obtient un point sur la sphère unité de $\mathbb{R}^d$ grâce à d variables aléatoires gaussiennes sur $\mathbb{R}$.

CORRIGÉ DE L'EXERCICE 2.8. Soit f une fonction borélienne bornée sur $\mathbb{R}^2$. Alors, grâce à un changement de variables

$$\begin{cases} x = \sqrt{u}\cos(2\pi v) \\ y = \sqrt{u}\sin(2\pi v) \end{cases}$$

dont le jacobien vaut π, on obtient

$$\begin{aligned}\mathbb{E}(f(X,Y) &= \int_{\mathbb{R}_+^* \times]0,1[} f\left(\sqrt{-2\ln(u)}\cos(2\pi v), \sqrt{-2\ln(u)}\sin(2\pi v)\right)\frac{1}{2}e^{-u/2} du dv \\ &= \int_{\mathbb{R}^2} f(x,y)\frac{e^{-(x^2+y^2)/2}}{2\pi} dx dy.\end{aligned}$$

Ainsi, X et Y sont indépendantes et de même loi $\mathcal{N}(0,1)$. Notons que l'on peut remplacer U par $-2\ln(W)$ où W est de loi uniforme sur $]0,1[$ et indépendante de V.

CORRIGÉ DE L'EXERCICE 2.10. On peut sans perte de généralité supposer que X est centré. La matrice Γ est symétrique par définition. Soit u un vecteur ligne de $\mathbb{R}^d$. Alors

$$u\Gamma u^T = \sum_{i,j=1}^d u_i u_j \mathrm{Cov}(X_i X_j) = \mathbb{V}(uX) \geq 0.$$

La matrice est donc positive. On a notamment que, pour $i \neq j$,

$$\Gamma_{ij}^2 \leq \Gamma_{ii}\Gamma_{jj}.$$

La décomposition de Cholesky fournit la décomposition demandée. Enfin, $AY + m$ suit la loi $\mathcal{N}_d(m, \Gamma)$.

CORRIGÉ DE L'EXERCICE 2.14. On a

$$\mathbb{E}(X_1) = (2p-1)a \quad \text{et} \quad \mathbb{E}(X_1^2) = a^2+1.$$

Notons

$$\hat{m}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{et} \quad \hat{m}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

les estimateurs des deux premiers moments de X_1. La méthode des moments suggère de proposer les estimateurs suivants

$$\hat{a} = \sqrt{\hat{m}_2 - 1} \quad \text{et} \quad \hat{p} = \frac{1}{2}\left(\frac{\hat{m}_1}{\hat{a}} + 1\right)$$

des paramètres a et p. Voir la figure 5 pour une illustration de la convergence presque sûre.

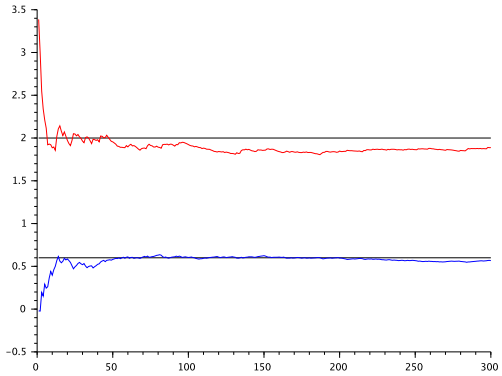

FIGURE 5. Estimation de $a = 2$ (courbe rouge) et $p = 0.6$ (courbe bleue) en fonction de la taille de l'échantillon $n = 1, \ldots, 300$.

CORRIGÉ DE L'EXERCICE 2.15. Soit $t \in \mathbb{R}$, on a, par convergence dominée,

$$\begin{aligned}\mathbb{E}\left(e^{itZ}\right) &= \mathbb{E}\left(e^{it(X_1+\cdots+X_k}\mathbf{1}_{\{N=k\}}\right)\\ &= \sum_{k\geq 0}\varphi(t)^k\mathbb{P}(N=k) = G(\varphi(t)).\end{aligned}$$

En calculant la dérivée en 0 de cette relation on obtient, sous réserve d'existence du premier moment de X_1 et N, la relation de Wald :

$$\mathbb{E}(Z) = \mathbb{E}(N)\mathbb{E}(X_1).$$

Dans le cas particulier, l'identification des fonctions caractéristiques assure que Z suit la loi de Poisson de paramètre $p\lambda$. Ce fait est relié à la propriété d'amincissement des processus de Poisson (voir le chapitre 5). Enfin, la loi de Z peut être vue comme le mélange (voir la définition 2.8) avec $\Theta = \mathbb{N}$, ν la loi de Poisson de paramètre λ et μ_k la loi binomiale $\mathscr{B}(k,p)$.

CORRIGÉ DE L'EXERCICE 2.12. On sait simuler une suite $(X_n)_{n\in\mathbb{N}^*}$ de v.a. i.i.d. de loi uniforme ν sur S^2 d'après l'exercice 2.7. On pose ensuite

$$\hat{I}_n = \frac{1}{n}\sum_{i=1}^{n} f(X_i) \quad \text{et} \quad \hat{\sigma}_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n} f(X_i)^2 - \hat{I}_n^2}.$$

Alors

$$\hat{I}_n \xrightarrow[n\to\infty]{p.s.} \int_{S^2} f\,d\nu \quad \text{et} \quad \hat{\sigma}_n^2 \xrightarrow[n\to\infty]{p.s.} \int_{S^2} f^2\,d\nu - \left(\int_{S^2} f\,d\nu\right)^2.$$

Donc

$$\frac{\sqrt{n}}{\hat{\sigma}_n}(\hat{I}_n - I) \xrightarrow[n\to\infty]{\mathscr{L}} \mathscr{N}(0,1).$$

et

$$\mathbb{P}\left(I \in \left[\hat{I}_n - \frac{1.96\hat{\sigma}_n}{\sqrt{n}}, \hat{I}_n + \frac{1.96\hat{\sigma}_n}{\sqrt{n}}\right]\right) \xrightarrow[n\to\infty]{} 0.95.$$

On pourra se reporter au chapitre 9 pour comprendre comment déterminer les intervalles de confiance (et utiliser le lemme de Slutsky).

Chapitre 3

Martingales

Dans tout le chapitre, $(\Omega, \mathscr{F}, \mathbb{P})$ est un espace de probabilité.

Commençons par présenter quelques modèles que nous pourrons étudier avec les outils présentés dans ce chapitre. Les principaux résultats sur ces modèles seront donnés au cours du chapitre, et les démonstrations présentées sous forme d'exercices en fin de chapitre.

Ruine du joueur. Un joueur arrive au casino avec une fortune initiale de $k > 0$ Euros. Il mise un Euro à chaque partie, gagne un Euro (en plus de sa mise) s'il gagne la partie et perd sa mise sinon. On suppose qu'il gagne avec probabilité $p \in]0;1[$ et perd avec probabilité $q = 1 - p$. On pose aussi $\rho = q/p = (1-p)/p$.
Enfin, il applique une stratégie très simple : il se fixe un objectif de $M > k$ Euros et il joue jusqu'à atteindre cet objectif ou être ruiné. Les questions clefs liées à ce modèle sont :

— quelles la probabilité que le joueur reparte ruiné ?

— combien de parties va-t-il jouer en moyenne ?

La modélisation mathématique de ce problème est assez simple : on prend $(Y_i)_{i\in\mathbb{N}^*}$ une suite de v.a. i.i.d. de loi $p\delta_1 + q\delta_{-1}$, où Y_k représente le gain relatif du joueur lors de la k-ème partie ; on note $X_0 = k$ la fortune initiale du joueur, puis $X_n = X_{n-1} + Y_n$ la fortune du joueur après la n-ème partie. On a donc aussi

$$X_n = k + \sum_{i=1}^{n} Y_i \,. \tag{1}$$

Enfin, le nombre de parties jouées est la variable aléatoire

$$T = \inf\big\{n \geq 0 : X_n \in \{0, M\}\big\} \,. \tag{2}$$

La figure 1 présente des simulations de ce modèle. Le cadre mathématique étant donné, les objets d'intérêt correspondant aux deux questions précédentes sont maintenant : $\mathbb{P}(X_T = 0)$, la probabilité de ruine du joueur ; et $\mathbb{E}(T)$, le nombre moyen de parties.

Processus de Galton-Watson. Il s'agit d'un modèle simple de généalogie. On évite toutes les difficultés liées à la reproduction sexuée en considérant qu'un individu engendre seul (Sir Galton, à l'origine de ce modèle au XIXème siècle, s'intéressait à la descendance des lords anglais, et ne considérait donc que les hommes et leurs descendants mâles...) un certain nombre de descendants. On va aussi considérer qu'à toutes les générations, le nombre de descendants est aléatoire, mais toujours de même loi, et indépendant d'une génération à l'autre et d'un individu à l'autre dans la même génération.

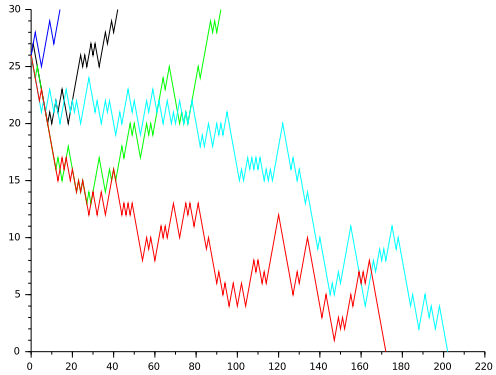

FIGURE 1. 5 trajectoires de la ruine du joueur jusqu'au temps T pour $k = 26$, $M = 30$ et $p = 0,45$.

Mathématiquement, on se donne donc une loi de reproduction $\mathbf{p} = (p_k)_{k\in\mathbb{N}}$, où p_k est la probabilité pour chaque individu d'avoir k enfants. On suppose, que $\mathbf{p}$ est de moyenne le $m = \sum_{k\in\mathbb{N}} kp_k$ et de variance σ^2 finies, et aussi que $p_0 > 0$ et $p_0 + p_1 < 1$. On pose $X_0 = 1$ (il y a un seul individu à la génération 0), puis pour tout $n \in \mathbb{N}^*$

$$X_n = \sum_{i=1}^{X_{n-1}} Y_i^{(n)}, \tag{3}$$

où les $Y_i^{(n)}$ sont des v.a. i.i.d. (et indépendantes de tout ce qui s'est passé avant) de loi $\mathbf{p}$: $Y_i^{(n)}$ est le nombre de descendants à la génération n du i-ème individu de la génération $n-1$; ainsi X_n est le nombre total d'individus de la n-ème génération. Dans ce modèle, tout ce qu'on pourra dire sur le comportement de la suite de variables aléatoires $(X_n)_{n\geq 0}$ est intéressant. En particulier, la probabilité d'extinction de la lignée, $\mathbb{P}(\text{extinction}) = \mathbb{P}(\exists n \in \mathbb{N} : X_n = 0)$ est le premier objet d'intérêt. La théorie des martingales nous permettra d'aller plus loin et par exemple de construire dans certains cas un estimateur du paramètre m.

1. Probabilités, espérances et lois conditionnelles

La notion de conditionnement est sans aucun doute l'une des notions fondamentales de la théorie des probabilités. Cette notion apparait de manière progressive et graduelle, avec des objets de plus en plus compliqués mathématiquement, que nous présentons dans cette section, ainsi que les liens existant entre eux.

1.1. Probabilités conditionnelles.

DÉFINITION 3.1. Soit $A, B \in \mathscr{F}$, avec $\mathbb{P}(A) > 0$, on appelle alors probabilité conditionnelle de B sachant A, notée $\mathbb{P}(B|A)$, la quantité

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(A)}. \tag{4}$$

La probabilité conditionnelle $\mathbb{P}(B|A)$ représente la probabilité d'observer l'événement B lorsque l'on sait déjà que l'événement A est réalisé. C'est donc immédiatement une notion pas complètement intuitive, et qui introduit une notion d'information. Pour rendre

cette idée plus précise, on pourrait donner un énoncé (à préciser) du genre : si on observe un grand nombre de fois dans des répétitions d'une même expérience aléatoire les occurrences de deux événements A et B, alors, parmi les observations où A est réalisée, la proportion d'observations où B l'est aussi sera proche de $\mathbb{P}(B|A)$.
Pour l'instant, nous nous contenterons de rappeler quelques propriétés élémentaires (mais fondamentales) de cet objet :

PROPOSITION 3.2.

1. *Pour $A \in \mathscr{F}$ tel que $\mathbb{P}(A) > 0$, l'application $\mathbb{P}(\cdot|A) : \mathscr{F} \longrightarrow [0,1]$ est une mesure de probabilité de $(\Omega, \mathscr{F})$, concentrée sur A (i.e. vérifiant $\mathbb{P}(A^c|A) = 0$).*
2. *Soit $A, B \in \mathscr{F}$ tels que $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$. Alors A et B sont deux événements indépendants si et seulement si $\mathbb{P}(B|A) = \mathbb{P}(B)$ si et seulement si $\mathbb{P}(A|B) = \mathbb{P}(A)$.*
3. ***Formule de Bayes.*** *Soit $A, B \in \mathscr{F}$ tels que $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$. Alors :*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}. \tag{5}$$

4. ***Formule des probabilités totales.*** *Soit $(A_i)_{i\in I}$, I fini ou dénombrable, une partition mesurable de Ω. Alors :*

$$\mathbb{P}(B) = \sum_{i\in I} \mathbb{P}(B|A_i)\mathbb{P}(A_i). \tag{6}$$

Les formules (5) et (6) semblent assez anodines, elles sont en particulier très faciles à démontrer, mais elles introduisent en fait une idée fondamentale, qui est qu'on peut renverser les liens de causalité, et que ce renversement n'est pas trivial : si on connait la probabilité conditionnelle de B sachant A, on pourra calculer la probabilité conditionnelle de A sachant B, mais à condition de connaître aussi les probabilités de A et de B, et le résultat dépend de celles-ci. C'est la source de paradoxes classiques (voir par exemple [**FF98**, Section 6.2]).
Souvent on utilisera (6) pour calculer le dénominateur dans (5). Ceci fait que la nomenclature n'est pas complètement standardisée, et qu'on appelle parfois formule de Bayes ou formule des probabilités totales la formule obtenue en combinant les deux.

1.2. Espérance conditionnelle. Pour passer à l'étape suivante, nous énonçons le théorème d'existence de l'espérance conditionnelle d'une variable aléatoire par rapport à une sous-tribu de $\mathscr{F}$. C'est un théorème difficile de théorie de la mesure (avec un doigt d'analyse fonctionnelle). Il contient donc en particulier la définition de l'espérance conditionnelle.

THÉORÈME 3.3. *Soit $X \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ et $\mathscr{G}$ une sous-tribu de $\mathscr{F}$. Il existe une variable aléatoire $Y \in L^1(\Omega, \mathscr{G}, \mathbb{P})$ telle que,*

$$\forall Z \in L^\infty(\Omega, \mathscr{G}, \mathbb{P}), \quad \mathbb{E}(XZ) = \mathbb{E}(YZ). \tag{7}$$

De plus, Y est unique à égalité presque sûre près, on l'appelle l'espérance conditionnelle de X sachant $\mathscr{G}$ et on la note $\mathbb{E}(X|\mathscr{G})$.

Ajoutons quelques propriétés fondamentales.

PROPOSITION 3.4.

(1) *L'espérance conditionnelle a tout d'une espérance : elle est linéaire, positive et monotone, satisfait le théorème de convergence monotone, le lemme de Fatou et le théorème de convergence dominée, ainsi que l'inégalité de Jensen.*

(2) *L'espérance est une espérance conditionnelle (!) : $\mathbb{E}(X|\{\emptyset, \Omega\}) = \mathbb{E}(X)$ p.s.*

(3) Si $\mathscr{H} \subset \mathscr{G}$ sont deux sous-tribus de $\mathscr{F}$, alors

$$\mathbb{E}\left(\mathbb{E}(X|\mathscr{G})|\mathscr{H}\right) = \mathbb{E}(X|\mathscr{H}) \text{ p.s.}$$

En particulier, $\mathbb{E}(\mathbb{E}(X|\mathscr{G})) = \mathbb{E}(X)$ p.s.

(4) Si Y est $\mathscr{G}$-mesurable et XY est intégrable, alors $\mathbb{E}(XY|\mathscr{G}) = Y\mathbb{E}(X|\mathscr{G})$ p.s.

(5) Si $\sigma(X)$ et $\mathscr{G}$ sont deux tribus indépendantes, alors $\mathbb{E}(X|\mathscr{G}) = \mathbb{E}(X)$ p.s.

Une autre remarque importante donne un sens « géométrique » à l'espérance conditionnelle des variables aléatoires de carré intégrable. C'est d'ailleurs l'une des méthodes pour démontrer l'existence de l'espérance conditionnelle (on passe ensuite à L^1 par troncation).

PROPOSITION 3.5. *Si $\mathscr{G}$ est une sous-tribu de $\mathscr{F}$ et $X \in L^2(\Omega, \mathscr{F}, \mathbb{P})$, la projection orthogonale de X sur le sous-espace vectoriel $L^2(\Omega, \mathscr{G}, \mathbb{P})$ dans l'espace de Hilbert $L^2(\Omega, \mathscr{F}, \mathbb{P})$ donne une version de l'espérance conditionnelle $\mathbb{E}(X|\mathscr{G})$.*
On en déduit la formule suivante, un théorème de Pythagore pour l'espérance conditionnelle :

$$\begin{aligned}\mathbb{V}(X) = \mathbb{E}\big((X - \mathbb{E}(X))^2\big) &= \mathbb{E}\big((X - \mathbb{E}(X|\mathscr{G}))^2\big) + \mathbb{E}\big((\mathbb{E}(X|\mathscr{G}) - \mathbb{E}(X))^2\big) \\ &= \mathbb{E}(\mathbb{V}(X|\mathscr{G})) + \mathbb{V}(\mathbb{E}(X|\mathscr{G}))\,,\end{aligned}$$

où $\mathbb{V}(X|\mathscr{G}) = \mathbb{E}\big((X - \mathbb{E}(X|\mathscr{G}))^2 \big| \mathscr{G}\big)$ désigne la « variance conditionnelle » de X conditionnellement à $\mathscr{G}$. Ainsi, la variance peut se décomposer comme somme de l'espérance de la variance conditionnelle et de la variance de l'espérance conditionnelle...

Cas particulier. Si X et Y sont deux variables aléatoires, et Y est intégrable, on utilisera la notation simplifiée

$$\mathbb{E}(Y|X) = \mathbb{E}(Y|\sigma(X))\,.$$

Comme on sait par ailleurs que toute variable aléatoire $\sigma(X)$-mesurable peut s'écrire comme une fonction de X [**BL98**, Lemme VI.3.1], le résultat d'existence de l'espérance conditionnelle assure dans ce contexte qu'il existe une fonction mesurable φ de $\mathbb{R}$ dans $\mathbb{R}$ telle que $\mathbb{E}(Y|X) = \varphi(X)$ p.s., c'est-à-dire aussi telle que $\mathbb{E}(\varphi(X)\psi(X)) = \mathbb{E}(Y\psi(X))$ pour toute fonction ψ bornée.
Plus généralement, si $(X_i)_{i\in I}$ est une famille de variables aléatoires, on notera encore $\mathbb{E}(Y|X_i, i \in I)$ à la place de $\mathbb{E}(Y|\sigma(X_i, i \in I))$.

Processus de Galton-Watson. Par des calculs simples utilisant l'espérance conditionnelle, les fonctions génératrices et l'étude d'une suite définie par récurrence, on peut montrer que

$$\mathbb{P}(\text{extinction}) \begin{cases} = 1 & \text{si } m \leq 1\,; \\ < 1 & \text{si } m > 1. \end{cases}$$

Ce résultat est assez naturel, en tout cas, pour $m \neq 1$: si chaque individu a moins d'un enfant en moyenne, la population va avoir tendance à décroître, donc à s'éteindre inexorablement, alors que s'il a plus d'un enfant, la population a des chances de croître significativement, et aura beaucoup de mal à s'éteindre après avoir atteint une grande taille. Ce résultat correspond aux questions 1 à 4 de l'exercice 3.7.

1.3. Loi conditionnelle. La notion de loi conditionnelle repose sur le théorème suivant, appelé théorème de décomposition des mesures, et qui peut-être vu comme une généralisation du théorème de Fubini aux mesures qui ne sont pas des mesures produits :

THÉORÈME 3.6. *Pour tout couple (X,Y) de variables aléatoires réelles, de loi jointe $\mu_{X,Y}$, il existe une famille $(\mu^x)_{x\in\mathbb{R}}$ de mesures de probabilités sur $\mathbb{R}$, avec $x \mapsto \mu^x$ mesurable, telle que pour tout fonction $\Phi : \mathbb{R}^2 \mapsto \mathbb{R}$ mesurable positive, ou $\mu_{X,Y}$-intégrable, on ait*

$$\begin{aligned}\mathbb{E}(\Phi(X,Y)) &= \iint_{\mathbb{R}^2} \Phi(x,y)\,d\mu_{X,Y}(x,y)\\ &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \Phi(x,y)\,d\mu^x(y)\right] d\mu_X(x)\end{aligned}$$

DÉFINITION 3.7. On appelle μ^x, qu'on note aussi $\mu_{Y|X=x}$ ou $\mathscr{L}(Y|X=x)$, la loi conditionnelle de Y sachant que $X=x$; on appelle aussi μ^X, qu'on note $\mu_{Y|X}$ ou $\mathscr{L}(Y|X)$, la loi conditionnelle de Y sachant X (ce deuxième objet est une fonction de X, c'est-à-dire une variable aléatoire $\sigma(X)$-mesurable, à valeurs dans les mesures de probabilités, c'est donc un objet compliqué...).

La loi conditionnelle est naturellement reliée à l'espérance conditionnelle, par la formule suivante, que l'on obtient facilement à partir de la définition de l'espérance conditionnelle et de la loi conditionnelle :

$$\mathbb{E}(Y|X) = \int_{\mathbb{R}} y\,d\mu^X(y) \quad p.s.$$

Ainsi, l'espérance conditionnelle est l'espérance de la loi conditionnelle. Mais la loi conditionnelle contient beaucoup plus d'informations, puisque plus généralement, pour tout fonction réelle g mesurable telle que $g(Y)$ soit intégrable, on a

$$\mathbb{E}(g(Y)|X) = \int_{\mathbb{R}} g(y)\,d\mu^X(y) \quad p.s.$$

REMARQUE 3.8. La notion de loi conditionnelle peut être généralisée de manière naturelle au cas où l'on conditionne par un nombre fini de variables aléatoires $(X_1, X_2, \dots, X_n)$, ce qui est la même chose que dire que X est un vecteur aléatoire. En effet, le théorème 3.6 est encore valide sur $\mathbb{R}\times\mathbb{R}^n$. On utilisera donc aussi dans la suite la notion de loi conditionnelle de Y sachant $(X_1, X_2, \dots, X_n)$, notée $\mathscr{L}(Y|X_1, X_2, \dots, X_n)$ ou $\mu_{Y|X_1,X_2,\dots,X_n}$.

Le cas des variables aléatoires discrètes. Dans le cas des variables aléatoires discrètes, ces trois notions sont reliées de manière assez naturelle : soit X et Y deux variables aléatoires discrètes, à valeurs respectivement dans deux sous-ensembles dénombrables $\mathscr{X}$ et $\mathscr{Y}$ de $\mathbb{R}$. On définit alors, pour tout $x \in \mathscr{X}$, la loi conditionnelle $\mu_{Y|X=x}$ par

$$\mu_{Y|X=x}(y) = \mathbb{P}(Y=y|X=x) \quad \forall y \in \mathscr{Y}$$

Quant à l'espérance conditionnelle, c'est la moyenne de la loi conditionnelle, donc

$$\begin{aligned}\mathbb{E}(Y|X) &= \sum_{y\in\mathscr{Y}} y\,\mathbb{P}(Y=y|X=x) \quad \text{quand } X=x\\ &= \sum_{x\in\mathscr{X}} \left[\sum_{y\in\mathscr{Y}} y\,\mathbb{P}(Y=y|X=x)\right] \mathbf{1}_{\{x\}}(X)\end{aligned}$$

Le cas des variables aléatoires à densité. Si le couple (X,Y) admet une densité $f_{(X,Y)}$ sur $\mathbb{R}^2$, alors X est à densité, de densité $f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x,y)\,dy$, et la loi conditionnelle $\mu_{Y|X=x}$ est aussi à densité, de densité

$$f_{Y|X=x}(y) = \frac{f_{(X,Y)}(x,y)}{f_X(x)}.$$

D'autres cas particuliers. Si X et Y sont indépendantes :

— $\mu_{Y|X=x} = \mu_Y$ pour tout $x \in \mathbb{R}$;

— $\mu_{X+Y|X=x} = \mu_{x+Y}$, la loi de la variable aléatoire $x+Y$. C'est un cas très utile dans la pratique.

2. Martingales : définition, théorème d'arrêt

2.1. Filtrations.

DÉFINITION 3.9.

(1) Une filtration de l'espace mesurable $(\Omega, \mathscr{F})$ est une suite croissante $(\mathscr{F}_n)_{n\in\mathbb{N}}$ de sous-tribus de $\mathscr{F}$: pour tout $n \in \mathbb{N}$, $\mathscr{F}_n \subset \mathscr{F}_{n+1} \subset \mathscr{F}$.

(2) Un processus stochastique $(X_n)_{n\in\mathbb{N}}$ (c'est-à-dire une suite de variables aléatoires) est dit adapté à la filtration $(\mathscr{F}_n)_{n\in\mathbb{N}}$ si, pour tout $n \in \mathbb{N}$, X_n est $\mathscr{F}_n$-mesurable.

(3) Étant donné un processus stochastique $(X_n)_{n\in\mathbb{N}}$, la fitration naturelle associée à ce processus est définie par, pour tout $n \in \mathbb{N}$, $\mathscr{F}_n = \sigma(X_0, X_1, \dots, X_n)$.

REMARQUE 3.10. Ce n'est pas (seulement...) pour le plaisir d'introduire de nouveaux termes que nous nous mettons à parler de *processus stochastique* plutôt que de *suite de variables aléatoires*. En faisant cela, nous introduisons un point de vue un peu différent sur ce type d'objet, et nous commençons en particulier à considérer l'indice $n \in \mathbb{N}$ comme un indice de temps : ainsi, un processus stochastique sera souvent vu comme la modélisation d'un phénomène évoluant (de manière aléatoire, stochastique) au cours du temps.
Avec la notion de filtration, nous ajoutons à cette flèche du temps une notion d'information : la tribu $\mathscr{F}_n$ représente toute l'information que nous avons (sur l'univers Ω) au temps n. Au temps n, on a au moins observé les réalisations de $X_0, X_1, \dots, X_n$, c'est le sens de la filtration naturelle, mais on peut avoir plus d'information, c'est la raison pour laquelle on donne une définition plus générale.

Terminons cette introduction aux filtrations avec une proposition qui est parfois utile, mais dont la démonstration est surtout un exercice à savoir faire (tout comme celle de la proposition 3.5, qui est dans le même esprit).

PROPOSITION 3.11. *Soit $(\mathscr{F}_n)_{n\in\mathbb{N}}$ une filtration de $(\Omega, \mathscr{F})$ et $(X_n)_{n\in\mathbb{N}}$ un processus adapté à cette filtration et tel que $X_n \in L^2(\mathbb{P})$ pour tout $\in \mathbb{N}$. On suppose que, pour tout $n \in \mathbb{N}$, $\mathbb{E}(X_{n+1}|\mathscr{F}_n) = \mathbb{E}(X_{n+1})$.*
Alors, pour tout $n \in \mathbb{N}$,

$$\mathbb{V}\left(\sum_{i=0}^{n} X_i\right) = \sum_{i=0}^{n} \mathbb{V}(X_i).$$

2.2. Martingales, sous-martingales, sur-martingales. Comme dit précédemment, on souhaite désormais étudier des processus stochastiques, c'est-à-dire des modélisations de phénomènes évoluant au cours du temps. Pour cela, le seul modèle étudié dans les cours de probabilités (jusqu'en M1) est assez pauvre, car peu de phénomènes naturels peuvent raisonnablement être modélisés par une suite de variables aléatoires indépendantes et identiquement distribuées. Le but de ce chapitre (comme de celui sur les chaînes de Markov) est donc d'introduire une classe plus générale de processus stochastiques, qui sont un peu plus crédibles pour la modélisation, tout en pouvant encore être étudiés mathématiquement.

DÉFINITION 3.12. Soit $(\mathscr{F}_n)_{n\in\mathbb{N}}$ une filtration de $(\Omega, \mathscr{F})$. Un processus stochastique $(X_n)_{n\in\mathbb{N}}$ adapté à $(\mathscr{F}_n)_{n\in\mathbb{N}}$ et tel que $X_n \in L^1(\mathbb{P})$ pour tout $n \in \mathbb{N}$ est appelé :

— une $(\mathscr{F}_n)_{n\in\mathbb{N}}$-martingale (ou simplement une martingale s'il n'y a pas d'ambiguité sur la filtration) si, pour tout $n \in \mathbb{N}$,

$$\mathbb{E}(X_{n+1}|\mathscr{F}_n) = X_n \quad p.s.; \tag{8}$$

— une $(\mathscr{F}_n)_{n\in\mathbb{N}}$-sous-martingale si, pour tout $n \in \mathbb{N}$,

$$\mathbb{E}(X_{n+1}|\mathscr{F}_n) \geq X_n \quad p.s.; \tag{9}$$

— une $(\mathscr{F}_n)_{n\in\mathbb{N}}$-sur-martingale si, pour tout $n \in \mathbb{N}$,

$$\mathbb{E}(X_{n+1}|\mathscr{F}_n) \leq X_n \quad p.s.; \tag{10}$$

Premiers exemples. Prenons $(X_i)_{i\in\mathbb{N}^*}$ une suite de v.a. i.i.d. intégrables et centrées, c'est-à-dire vérifiant $\mathbb{E}(|X_i|) < +\infty$ et $\mathbb{E}(X_i) = 0$, et posons $S_0 = x$, puis, pour tout $n \in \mathbb{N}^*$, $S_n = S_{n-1} + X_n$. Cette définition est équivalente à la formule

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

et définit la forme générale d'une marche au hasard sur $\mathbb{R}$ (ou un sous-groupe de $\mathbb{R}$ si x et les X_i y appartiennent p.s.).
On peut alors vérifier facilement (exercice) que $\mathscr{F}_n = \sigma(S_0, S_1, \ldots, S_n) = \sigma(X_0, X_1, \ldots, X_n)$ et que $(S_n)_{n\in\mathbb{N}}$ est une martingale pour cette filtration naturelle.
De la même manière, en prenant $(Y_i)_{i\in\mathbb{N}^*}$ une suite de v.a. i.i.d. intégrables et de moyenne 1, et en posant $M_0 = 1$ puis, pour tout $n \in \mathbb{N}^*$, $M_n = M_{n-1} \cdot Y_n$, ce qui équivaut à

$$M_n = \prod_{i=1}^{n} Y_i\,,$$

on peut vérifier (exercice) que la suite $(M_n)_{n\in\mathbb{N}}$ est une martingale pour sa filtration naturelle.

Ruine du joueur. Les trois processus

$$A_n = X_n - (p-q)n \quad ; \quad B_n = A_n^2 - 4pqn \quad ; \quad C_n = \rho^{X_n}$$

sont des martingales par rapport à la filtration $\mathscr{F}_n = \sigma(X_0, X_1, \ldots, X_n)$.

2.3. Temps d'arrêt, théorème d'arrêt.

DÉFINITION 3.13. Soit $(\mathscr{F}_n)_{n\in\mathbb{N}}$ une filtration de $(\Omega, \mathscr{F})$. On dit qu'une variable aléatoire T à valeurs dans $\mathbb{N}\cup\{+\infty\}$ est un temps d'arrêt pour $(\mathscr{F}_n)_{n\in\mathbb{N}}$ si, pour tout $n \in \mathbb{N}$, l'événement $\{T \leq n\}$ est $\mathscr{F}_n$-mesurable (ou de manière équivalente, pour tout $n \in \mathbb{N}$, $\{T = n\}$ est $\mathscr{F}_n$-mesurable).

PROPOSITION 3.14. *Si $(X_n)_{n\in\mathbb{N}}$ est un processus stochastique $(\mathscr{F}_n)_{n\in\mathbb{N}}$-adapté et $A \subset \mathbb{R}$, on définit le temps d'atteinte de A par*

$$T_A = \inf\{n \in \mathbb{N} : X_n \in A\}$$

(avec la convention que $\inf \emptyset = +\infty$*). Alors T_A est un temps d'arrêt pour $(\mathscr{F}_n)_{n\in\mathbb{N}}$.*

Le cas du temps d'atteinte d'un sous-ensemble traité dans l'exemple précédent est en fait le cas typique de temps d'arrêt.

Ruine du joueur. Le temps aléatoire T défini dans ce modèle est un temps d'arrêt par rapport à la filtration $(\mathscr{F}_n)_{n\in\mathbb{N}}$.

Un temps d'arrêt est donc une variable aléatoire à valeurs entières (éventuellement infinies). Si on a aussi un processus stochastique $(X_n)_{n\in\mathbb{N}}$, on peut considérer le processus stochastique arrêté au temps (aléatoire) T, $(X_{T\wedge n})_{n\in\mathbb{N}}$, bien défini par $(X_{T\wedge n})(\omega) = X_{T(\omega)\wedge n}(\omega)$.
Il est tentant de définir aussi X_T par $X_{T(\omega)}(\omega)$, mais c'est tout de suite un peu plus délicat, par exemple si T prend la valeur $+\infty$ avec probabilité positive.
Commençons donc par énoncer le résultat principal pour le processus arrêté.

PROPOSITION 3.15. *Si $(\mathscr{F}_n)_{n\in\mathbb{N}}$ une filtration de $(\Omega,\mathscr{F})$, $(X_n)_{n\in\mathbb{N}}$ est une $(\mathscr{F}_n)_{n\in\mathbb{N}}$-martingale, et T est un temps d'arrêt pour $(\mathscr{F}_n)_{n\in\mathbb{N}}$, alors le processus $(X_{T\wedge n})_{n\in\mathbb{N}}$ est encore une $(\mathscr{F}_n)_{n\in\mathbb{N}}$-martingale.*
En particulier, pour tout $n\in\mathbb{N}$,

$$\mathbb{E}(X_{T\wedge n}) = \mathbb{E}(X_0)\,. \tag{11}$$

On a alors bien envie de faire tendre n vers l'infini dans la formule (11), pour obtenir que $\mathbb{E}(X_T) = \mathbb{E}(X_0)$, ce qui donnerait un outil de calcul riche. Il est clair qu'on ne peut pas écrire cette formule en toute généralité, à cause de la remarque faite avant la proposition, et comme l'illustre aussi l'exercice 3.5.
Pour faire les choses proprement, il faut utiliser des outils de théorie de la mesure qui permettent de faire tendre n vers l'infini dans la formule (11). On peut le refaire sur chaque exemple, ou vérifier l'une des hypothèses du théorème suivant.

THÉORÈME 3.16 (Théorème d'arrêt ou théorème de Doob). *Soit $(\mathscr{F}_n)_{n\in\mathbb{N}}$ une filtration de $(\Omega,\mathscr{F})$, $(X_n)_{n\in\mathbb{N}}$ une $(\mathscr{F}_n)_{n\in\mathbb{N}}$-martingale, et T un temps d'arrêt pour $(\mathscr{F}_n)_{n\in\mathbb{N}}$. On a*

$$\mathbb{E}(X_T) = \mathbb{E}(X_0) \tag{12}$$

dès que l'une des hypothèses suivantes est vérifiée :

(1) T est borné.

(2) La martingale est bornée ($\sup_{n\in\mathbb{N}} \|X_n\|_\infty < +\infty$) et T est p.s. fini.

(3) $\mathbb{E}(T) < +\infty$ et la martingale est à accroissements bornés ($\sup_{n\in\mathbb{N}} \|X_{n+1} - X_n\|_\infty < +\infty$).

Ainsi, puisque la conclusion du théorème de Doob est en défaut dans le cas de l'exercice 3.5, alors que la martingale est bien à accroissements bornés, on peut en déduire que $\mathbb{E}(T) = +\infty$ (le temps d'atteinte de 1 est p.s. fini mais n'est pas intégrable, et c'est pour cette raison que le théorème de Doob est en défaut...).

Ruine du joueur. On a les formules explicites suivantes pour la probabilité de ruine et le nombre moyen de parties jouées :

— si $p = q = 1/2$:

$$\mathbb{P}(X_T = 0) = \frac{M-k}{M} \quad \text{et} \quad \mathbb{E}(T) = k(M-k)\,;$$

— si $p \neq q$:

$$\mathbb{P}(X_T = 0) = \frac{\rho^k - \rho^M}{1-\rho^M} \quad \text{et} \quad \mathbb{E}(T) = \frac{1}{p-q}\,\frac{M\left(1-\rho^k\right) - k\left(1-\rho^M\right)}{1-\rho^M}\,,$$

avec $\rho = q/p$.

Ces résultats sont illustrés par les figures 2 et 3.

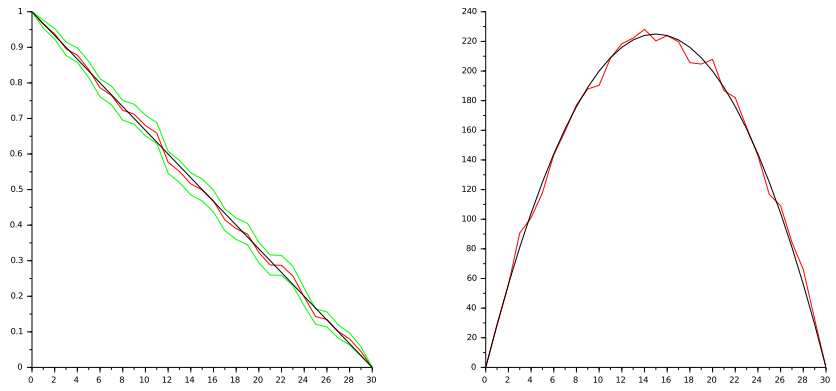

FIGURE 2. Probabilité de ruine (à gauche) et temps moyen de jeu (à droite) pour $M = 30$ et $p = 0,5$, k étant en abscisses. On compare les valeurs données par les formules exactes (en noir) aux valeurs obtenues par une méthode de Monte Carlo avec 1000 simulations (en rouge), plus les intervalles de confiance (en vert) pour la probabilité de ruine.

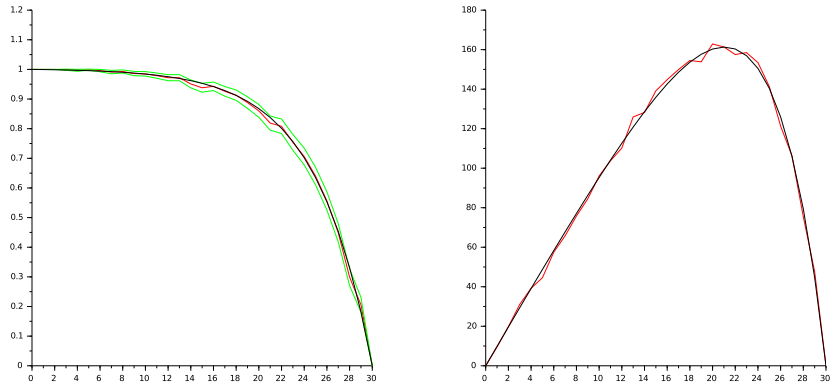

FIGURE 3. Même chose que la figure 2 pour $p = 0,45$.

REMARQUE 3.17. Les résultats obtenus dans cette section se généralisent aux sur-martingales. La proposition 3.15 devient : si $(X_n)_{n\in\mathbb{N}}$ est une sur-martingale, et T un temps d'arrêt, alors $(X_{T\wedge n})_{n\in\mathbb{N}}$ est encore une sur-martingale, et la formule (11) devient $\mathbb{E}(X_{T\wedge n}) \leq \mathbb{E}(X_0)$. Une version possible (et utile) du Théorème d'arrêt serait : si $(X_n)_{n\in\mathbb{N}}$ est une sur-martingale positive, et T un temps d'arrêt p.s. fini, alors $\mathbb{E}(X_T) \leq \mathbb{E}(X_0)$ (il suffit d'appliquer le lemme de Fatou à la formule précédente).

3. Théorèmes de convergence des martingales

On l'a déjà vu, les martingales sont une classe de processus stochastiques généralisant les sommes de v.a. i.i.d., et apparaissant assez naturellement, tant du point de vue de la définition mathématique que des applications.
Cependant, pour qu'elles soient vraiment utiles, il faut aussi qu'on ait des résultats décrivant leur comportement en temps long, généralisant la loi des grands nombres et

le théorème limite central classiques pour les sommes de v.a. i.i.d. C'est ce que nous présentons ici.

DÉFINITION 3.18. Soit $(\mathscr{F}_n)_{n\in\mathbb{N}}$ une filtration de $(\Omega,\mathscr{F})$, $(X_n)_{n\in\mathbb{N}}$ une $(\mathscr{F}_n)_{n\in\mathbb{N}}$-martingale. On dit que la martingale $(X_n)_{n\in\mathbb{N}}$ est bornée dans L^p si $\sup_{n\in\mathbb{N}}\|X_n\|_p < +\infty$.

3.1. Théorèmes de convergences des martingales. Les théorèmes fondamentaux à connaître pour l'agrégation sont les suivants :

THÉORÈME 3.19. *Soit $(\mathscr{F}_n)_{n\in\mathbb{N}}$ une filtration de $(\Omega,\mathscr{F})$, $(X_n)_{n\in\mathbb{N}}$ une $(\mathscr{F}_n)_{n\in\mathbb{N}}$-martingale. Si $(X_n)_{n\in\mathbb{N}}$ est bornée dans L^1, alors elle converge presque sûrement : $X_\infty = \lim_{n\to+\infty} X_n$ existe p.s.*

THÉORÈME 3.20. *Soit $(\mathscr{F}_n)_{n\in\mathbb{N}}$ une filtration de $(\Omega,\mathscr{F})$, $(X_n)_{n\in\mathbb{N}}$ une $(\mathscr{F}_n)_{n\in\mathbb{N}}$-martingale, avec $X_n \in L^2(\mathbb{P})$ pour tout $n\in\mathbb{N}$.*

(1) La martingale $(X_n)_{n\in\mathbb{N}}$ est bornée dans L^2 si et seulement si la série
$$\sum_{n\in\mathbb{N}} \mathbb{E}\left[(X_{n+1}-X_n)^2\right]$$
est convergente.

(2) Si le point (1) est vérifié, alors la suite $(X_n)_{n\in\mathbb{N}}$ converge p.s. et dans L^2 : $X_\infty = \lim_{n\to+\infty} X_n$ existe p.s., $X_\infty \in L^2(\mathbb{P})$, et $\lim_{n\to+\infty}\|X_n - X_\infty\|_2 = 0$.

REMARQUE 3.21. Attention, en revanche, il n'est pas vrai en général qu'une martingale bornée dans L^1 converge vers sa limite dans L^1, comme l'illustre le cas des processus de Galton-Watson dans le cas où $m \le 1$ (voir ci-dessous). On a besoin d'un peu plus : il suffit qu'il existe $p > 1$ tel que la martingale soit bornée dans L^p ; on peut aussi exhiber une condition nécessaire et suffisante, l'uniforme intégrabilité (voir [**Wil91**, Chap. 13] pour en savoir plus sur cette notion).

Processus de Galton-Watson. Le processus $Z_n = \frac{X_n}{m^n}$ est une martingale pour la filtration $\mathscr{F}_n = \sigma(X_0, X_1, \dots, X_n)$.
Cette martingale est bornée dans L^1, donc il existe Z_∞ telle que Z_n converge vers Z_∞ p.s.
Si $m \le 1$, la convergence n'a pas lieu dans L^1.
Si $m > 1$, la martingale est bornée dans L^2, donc la convergence a aussi lieu dans L^2 et dans L^1.
Ces résultats correspondent aux questions 5 à 9 de l'exercice 3.7 et sont illustrés par la figure 4.

Un corollaire utile du théorème 3.20 est le résultat suivant :

THÉORÈME 3.22. *Soit $(X_k)_{k\in\mathbb{N}}$ une suite de v.a. indépendantes centrées et telles que, pour tout $k\in\mathbb{N}$, $\sigma_k^2 = \mathbb{E}(X_k^2) < +\infty$. Alors*
$$\sum_{k\in\mathbb{N}} \sigma_k^2 < +\infty \quad\Longrightarrow\quad \sum_{k\in\mathbb{N}} X_k \text{ converge p.s.}$$

Ce résultat sera particulièrement utile associé au lemme 3.30 (de Kronecker). On peut ainsi obtenir le théorème 3.23 ci-dessous : c'est une généralisation de la loi forte des grands nombres pour des variables aléatoires centrées et indépendantes avec un moment d'ordre 2. On peut d'ailleurs retrouver la loi forte des grands nombres classique par un argument de troncation.

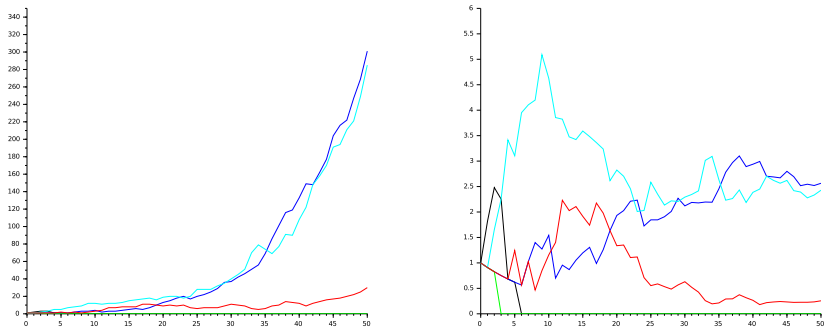

FIGURE 4. 5 trajectoires du processus de Galton-Watson et de la martingale associée jusqu'au temps 30 pour $p_0 = 0,2$; $p_1 = 0,5$; $p_2 = 0,3$; donc $m = 1,1$.

THÉORÈME 3.23. *Soit $(X_k)_{k\in\mathbb{N}^*}$ une suite de v.a. indépendantes centrées et telles que, pour tout $k \in \mathbb{N}^*$, $\sigma_k^2 = \mathbb{E}(X_k^2) < +\infty$. Alors*

$$\sum_{k\in\mathbb{N}^*} \frac{\sigma_k^2}{k^2} < +\infty \quad \Longrightarrow \quad \lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} X_k = 0 \text{ p.s.}$$

3.2. Décomposition de Doob, LGN, TLC. Le cadre L^2 est encore bien adapté pour obtenir des théorèmes limite de type Loi des grands nombres et théorème limite central pour des martingales. Pour cela, on introduit la décomposition suivante.

PROPOSITION 3.24 (Décomposition de Doob). *Soit $(\mathscr{F}_n)_{n\in\mathbb{N}}$ une filtration de $(\Omega, \mathscr{F})$ et $(X_n)_{n\in\mathbb{N}}$ une $(\mathscr{F}_n)_{n\in\mathbb{N}}$-martingale, avec $X_n \in L^2(\mathbb{P})$ pour tout $n \in \mathbb{N}$. Alors :*

(1) $(X_n^2)_{n\in\mathbb{N}}$ est une sous-martingale.

(2) Soit $(\langle X\rangle_n)_{n\in\mathbb{N}}$ le processus stochastique défini par $\langle X\rangle_0 = 0$, puis, pour tout $n \in \mathbb{N}$,

$$\langle X\rangle_{n+1} - \langle X\rangle_n = \mathbb{E}\left[(X_{n+1} - X_n)^2 \middle| \mathscr{F}_n\right].$$

Alors :

(a) le processus $(\langle X\rangle_n)_{n\in\mathbb{N}}$ est positif, croissant et prévisible, c'est-à-dire vérifie que, pour tout $n \in \mathbb{N}^*$, $\langle X\rangle_n$ est $\mathscr{F}_{n-1}$-mesurable ;

(b) le processus $\left(Z_n = X_n^2 - \langle X\rangle_n\right)_{n\in\mathbb{N}}$ est une $(\mathscr{F}_n)_{n\in\mathbb{N}}$-martingale.

On appelle $(\langle X\rangle_n)_{n\in\mathbb{N}}$ le compensateur de la sous-martingale $(X_n^2)_{n\in\mathbb{N}}$, ou le processus croissant associé à $(X_n)_{n\in\mathbb{N}}$.
La décomposition de Doob est l'écriture

$$X_n^2 = Z_n + \langle X\rangle_n \quad \forall n \in \mathbb{N}^*, \tag{13}$$

soit la décomposition d'un carré de martingale en somme d'une martingale et d'un processus positif, croissant et prévisible.
On pose aussi

$$\langle X\rangle_\infty = \lim_{n\to+\infty} \langle X\rangle_n.$$

REMARQUE 3.25. On a la formule explicite

$$\langle X\rangle_\infty = \sum_{n\in\mathbb{N}} \mathbb{E}\left[(X_{n+1} - X_n)^2 \middle| \mathscr{F}_n\right]. \tag{14}$$

Associée au point (1) du théorème 3.20, elle implique que la martingale $(X_n)_{n\in\mathbb{N}}$ est bornée dans L^2 si et seulement si $\langle X\rangle_\infty \in L^1(\mathbb{P})$.

REMARQUE 3.26. L'existence de la décomposition de Doob est en fait un résultat général pour toute sous-martingale. L'avantage pour un carré de martingale est qu'on obtient cette décomposition explicitement.

Cette décomposition donne le bon cadre pour obtenir des théorèmes limite.

THÉORÈME 3.27 (Loi des grands nombres pour les martingales L^2). *Soit $(\mathscr{F}_n)_{n\in\mathbb{N}}$ une filtration de $(\Omega,\mathscr{F})$ et $(X_n)_{n\in\mathbb{N}}$ une $(\mathscr{F}_n)_{n\in\mathbb{N}}$-martingale, avec $X_n \in L^2(\mathbb{P})$ pour tout $n \in \mathbb{N}$.*

(1) Sur $\{\langle X\rangle_\infty < +\infty\}$, la martingale $(X_n)_{n\in\mathbb{N}}$ converge p.s. vers $X_\infty \in L^2(\mathbb{P})$.

(2) Sur $\{\langle X\rangle_\infty = +\infty\}$, la suite $\left(\frac{X_n}{\langle X\rangle_n}\right)_{n\in\mathbb{N}^*}$ converge p.s. vers 0.

THÉORÈME 3.28 (Théorème limite central pour les martingales L^2). *Soit $(\mathscr{F}_n)_{n\in\mathbb{N}}$ une filtration de $(\Omega,\mathscr{F})$ et $(X_n)_{n\in\mathbb{N}}$ une $(\mathscr{F}_n)_{n\in\mathbb{N}}$-martingale, avec $X_n \in L^2(\mathbb{P})$ pour tout $n \in \mathbb{N}$.. On suppose qu'il existe une suite $(a_n)_{n\in\mathbb{N}}$ de réels strictement positifs, croissante et tendant vers $+\infty$, et telle que les deux conditions suivantes soient vérifiées :*

(1) Il existe $l \in \mathbb{R}_+$ tel que la suite $\left(\frac{\langle X\rangle_n}{a_n}\right)_{n\in\mathbb{N}^*}$ converge vers l en probabilité.

(2) (Condition de Lindeberg) Pour tout $\varepsilon > 0$, la suite

$$\frac{1}{a_n}\sum_{k=1}^{n}\mathbb{E}\left[(X_k - X_{k-1})^2\mathbf{1}_{\{|X_k-X_{k-1}|\geq\varepsilon\sqrt{a_n}\}}\middle|\mathscr{F}_{k-1}\right]$$

converge vers 0 en probabilité.

Alors,

$$\frac{1}{\sqrt{a_n}}X_n \xrightarrow[n\to+\infty]{\mathscr{L}} \mathscr{N}(0,l),$$

et, si $l > 0$,

$$\sqrt{a_n}\frac{X_n}{\langle X\rangle_n} \xrightarrow[n\to+\infty]{\mathscr{L}} \mathscr{N}(0,l^{-1}).$$

Processus de Galton-Watson. Le processus $M_n = X_n - mX_{n-1}$ est une martingale pour la filtration $\mathscr{F}_n$.
Si $m > 1$, on a

$$\lim_{n\to+\infty}\frac{\langle M\rangle_n}{m^n} = \frac{\sigma^2}{m-1}Z_\infty \quad \text{p.s.},$$

ce qui permet de démontrer que

$$\hat{m}_n = \frac{\sum_{k=1}^n X_k}{\sum_{k=1}^n X_{k-1}}.$$

converge p.s. vers la moyenne m en restriction à $\{Z_\infty > 0\}$.
Ces résultats correspondent aux questions 5, puis 10 à 13 de l'exercice 3.7 et la convergence de $\hat{m}_n$ est illustrée par la figure 5.

4. Exercices

EXERCICE 3.1. Soit X et Y deux variables aléatoires indépendantes de loi de Poisson de paramètre respectif λ et μ. Montrer que la loi conditionnelle de X sachant $X+Y$ est une loi binomiale dont vous préciserez les paramètres.

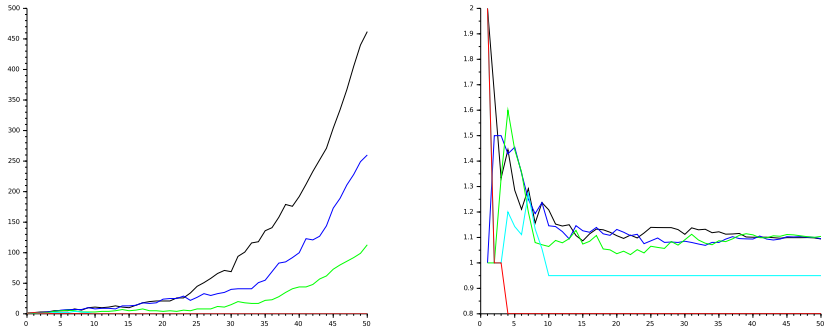

FIGURE 5. 5 trajectoires du processus de Galton-Watson et de l'estimateur $\hat{m}_n$ jusqu'au temps 50 pour $p_0 = 0,2$; $p_1 = 0,5$; $p_2 = 0,3$; donc $m = 1,1$. Seules 3 des 5 trajectoires correspondent à un cas de non-extinction et convergent effectivement vers m. Pour les deux autres, il y a extinction, elles sont donc constantes à partir d'un certain rang.

EXERCICE 3.2. On prend dans cet exercice $\Omega = [0,1[$, $\mathscr{F} = \mathscr{B}([0,1[)$ sa tribu borélienne, $\mathbb{P}$ la mesure de Lebesgue sur $[0,1[$, et la variable aléatoire X définie par $X(\omega) = \omega$.
Pour tout $n \in \mathbb{N}$, on définit

$$\mathscr{I}_n = \left\{ \left[\frac{k}{2^n}; \frac{k+1}{2^n} \right[, 0 \le k \le 2^n - 1 \right\},$$

puis $\mathscr{F}_n = \sigma(\mathscr{I}_n)$.
1. Montrer que $(\mathscr{F}_n)_{n\in\mathbb{N}}$ est une filtration de $(\Omega, \mathscr{F})$.
2. Identifier, pour tout $n \in \mathbb{N}$, $X_n = \mathbb{E}(X|\mathscr{F}_n)$, et préciser sa loi.
3. Vérifier que la suite $(X_n)_{n\in\mathbb{N}}$ converge presque sûrement vers X.

EXERCICE 3.3. Démontrer les propositions 3.5 et 3.11.

EXERCICE 3.4. Vérifier que si $(X_n)_{n\in\mathbb{N}}$ est une $(\mathscr{F}_n)_{n\in\mathbb{N}}$-martingale, alors :
1. pour tous $0 \le m \le n$, $\mathbb{E}(X_n|\mathscr{F}_m) = X_m$;
2. pour tout $n \in \mathbb{N}$, $\mathbb{E}(X_n) = \mathbb{E}(X_0)$.

EXERCICE 3.5. Soit $(X_i)_{i\in\mathbb{N}}$ une suite de v.a. i.i.d. telles que $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = 1/2$, et $S_n = \sum_{i=1}^n X_i$ la marche au hasard simple symétrique sur $\mathbb{Z}$. On pose

$$T = \inf\{n \in \mathbb{N} : S_n = +1\}$$

et on admet que T est p.s. fini (c'est par exemple une conséquence de la récurrence de marche simple symétrique en dimension 1).
Montrer que $\mathbb{E}(X_T) = 1 \neq 0 = \mathbb{E}(X_0)$.

EXERCICE 3.6 (Ruine du joueur). Le début de l'énoncé est dans le premier cadre au début du chapitre. On pose pour tout $n \ge 0$ $\mathscr{F}_n = \sigma(Y_1, Y_2, \dots, Y_n)$ (par convention, $\mathscr{F}_0 = \{\emptyset, \Omega\}$).
1. Vérifier que $(\mathscr{F}_n)_{n\in\mathbb{N}}$ est une filtration et que pour tout $n \in \mathbb{N}$, $\mathscr{F}_n = \sigma(X_0, X_1, \dots, X_n)$.
2. Montrer que les trois processus suivants sont des $(\mathscr{F}_n)_{n\in\mathbb{N}}$-martingales :

$$A_n = X_n - (p-q)n \quad ; \quad B_n = A_n^2 - 4pqn \quad ; \quad C_n = \rho^{X_n}.$$

3. Pourquoi T est-il un temps d'arrêt associé à la filtration $(\mathscr{F}_n)_{n\in\mathbb{N}}$?
4. Pour $p = q = 1/2$ (donc $\rho = 1$) utiliser la martingale $(B_n = X_n^2 - n)_{n\in\mathbb{N}}$ pour obtenir que, pour tout $n \in \mathbb{N}$,

$$\mathbb{E}(X_{T\wedge n}^2) = k^2 + \mathbb{E}(T \wedge n).$$

En déduire que $\mathbb{E}(T) < +\infty$, donc que T est p.s. fini.
5. Toujours pour $p = q = 1/2$, utiliser la martingale $(A_n = X_n)_{n\in\mathbb{N}}$ pour démontrer que

$$\mathbb{P}(X_T = 0) = \frac{M-k}{M} \quad \text{et} \quad \mathbb{E}(X_T) = k$$

puis en réutilisant $(B_n)_{n\in\mathbb{N}}$ que

$$\mathbb{E}(T) = k(M-k)\,.$$

6. Pour $p \neq q$, utiliser la martingale $(A_n)_{n\in\mathbb{N}}$ pour obtenir que, pour tout $n \in \mathbb{N}$,

$$\mathbb{E}(X_{T\wedge n}) = k + (p-q)\mathbb{E}(T \wedge n)\,.$$

En déduire que $\mathbb{E}(T) < +\infty$, donc que T est p.s. fini.
7. Toujours pour $p \neq q$, utiliser la martingale $(C_n)_{n\in\mathbb{N}}$ pour démontrer que

$$\mathbb{P}(X_T = 0) = \frac{\rho^k - \rho^M}{1-\rho^M} \quad \text{et} \quad \mathbb{E}(X_T) = M\frac{1-\rho^k}{1-\rho^M}$$

puis en réutilisant $(A_n)_{n\in\mathbb{N}}$ que

$$\mathbb{E}(T) = \frac{1}{p-q}\frac{M\left(1-\rho^k\right) - k\left(1-\rho^M\right)}{1-\rho^M}$$

8. Pourquoi les arguments des questions 6. et 7. ne fonctionnent-ils pas pour $p = q = 1/2$?

EXERCICE 3.7 (Processus de Galton-Watson). On définit les fonctions génératrices de la loi de reproduction **p** et des tailles de chaque population par, pour tout $s \in [0;1]$, pour tout $n \in \mathbb{N}$,

$$\varphi(s) = \sum_{k\in\mathbb{N}} p_k s^k \quad \text{et} \quad \varphi_n(s) = \mathbb{E}\left(s^{X_n}\right).$$

1. Montrer que, pour tout $n \in \mathbb{N}$,

$$\mathbb{E}\left(s^{X_{n+1}} \,\middle|\, X_n\right) = \varphi(s)^{X_n}\,.$$

2. En déduire que la suite de fonctions $(\varphi_n)_{n\in\mathbb{N}}$ vérifie $\varphi_0(s) = s$ et $\varphi_{n+1}(s) = \varphi_n\big(\varphi(s)\big)$, donc aussi $\varphi_n(s) = \varphi^n(s)$ (l'itérée n-ème de φ), pour tout $s \in [0;1]$ et tout $n \in \mathbb{N}$.
3. Montrer que la fonction φ est $\mathscr{C}^\infty$ sur $[0;1[$, dérivable à gauche en 1 avec $\varphi'(1) = m$, croissante et strictement convexe sur $[0;1[$.
En déduire que :

— si $m \leq 1$, 1 est l'unique point fixe de φ ;

— si $m > 1$, il existe un unique autre point fixe $q \in]0;1[$ de φ.

4. Remarquer que, pour tout $n \in \mathbb{N}$, $\mathbb{P}(X_n = 0) = \varphi_n(0) = \varphi^n(0)$, puis que

$$\mathbb{P}(\text{extinction}) = \mathbb{P}(\exists n \in \mathbb{N} : X_n = 0) = \lim_{n\to+\infty} \varphi^n(0) = \begin{cases} 1 & \text{si } m \leq 1\,; \\ q & \text{si } m > 1. \end{cases}$$

5. Montrer que $Z_n = \frac{X_n}{m^n}$ et $M_n = \sum_{i=1}^n (X_n - mX_{n-1})$ sont des martingales pour la filtration naturelle $\mathscr{F}_n = \sigma(X_0, X_1, \ldots, X_n)$.
6. En déduire que la suite $Z_n = \frac{X_n}{m^n}$ converge p.s. vers une variable aléatoire Z_∞ intégrable.
7. Par des considérations sur les espérances et en utilisant la question 4., montrer que si $m \leq 1$, Z_n ne converge pas vers Z_∞ dans L^1.
8. Si $m > 1$, calculer le processus croissant $\langle Z\rangle_n$ associé à Z_n, puis $\langle Z\rangle_\infty$ et $\mathbb{E}(\langle Z\rangle_\infty)$.
9. En déduire que pour $m > 1$, Z_n converge vers Z_∞ dans L^2 et dans L^1, et que $\mathbb{E}(Z_\infty) = 1$.
10. Vérifier que, pour tout $n \in \mathbb{N}$, $\langle M\rangle_n = \sigma^2 \sum_{k=0}^{n-1} X_k$.

11. Si $m > 1$, montrer, à l'aide de la question 6 et du lemme 3.29 (de Toeplitz) que

$$\lim_{n\to+\infty} \frac{\langle M\rangle_n}{m^n} = \frac{\sigma^2}{m-1} Z_\infty \text{ p.s.} \tag{15}$$

En déduire que $\{Z_\infty > 0\} \subset \{\langle M\rangle_\infty = +\infty\}$
12. On pose

$$\hat{m}_n = \frac{\sum_{k=1}^n X_k}{\sum_{k=1}^n X_{k-1}}.$$

Montrer que $\hat{m}_n - m = \sigma^2 \frac{M_n}{\langle M\rangle_n}$.
13. Appliquer la loi des grands nombres pour les martingales pour en déduire que pour $m > 1$ et en restriction à $\{Z_\infty > 0\}$, $\hat{m}_n$ converge p.s. vers m : $\hat{m}_n$ est un estimateur consistant de m, en restriction à un événement de probabilité strictement positive.

On pourrait aussi réutiliser la formule (15) pour montrer que $\hat{m}_n$ est asymptotiquement normal à l'aide du TLC pour les martingales...

EXERCICE 3.8 (Casino). On considère un jeu de hasard entre un joueur et le croupier d'un casino. Le capital total en jeu est 1. Au temps n, le capital du joueur est $X_n \in [0,1]$ et celui du croupier est $1 - X_n$. Au début du jeu, le capital du joueur est une constante $X_0 = p \in]0,1[$.
La règle du jeu est la suivante. Après les n premières parties, la probabilité de gagner la partie suivante est égale à X_n et la probabilité de perdre est $1 - X_n$. Si le joueur gagne, il obtient la moitié du capital du croupier. Dans le cas contraire, il cède la moitié de son capital au croupier. Ainsi, pour toute fonction borélienne bornée $f : [0,1] \to \mathbb{R}$, on a :

$$\mathbb{E}(f(X_{n+1})|\mathscr{F}_n) = X_n f\left(X_n + \frac{1-X_n}{2}\right) + (1-X_n) f\left(\frac{X_n}{2}\right).$$

où $(\mathscr{F}_n)_{n\geq 0}$ est la filtration naturelle de la suite $(X_n)_{n\geq 0}$.
1. Montrer que $(X_n)_{n\geq 0}$ est une martingale.
2. Montrer que $(X_n)_{n\geq 0}$ converge p.s. et dans L^2 vers une variable aléatoire Z.
3. Montrer que $\mathbb{E}(X_{n+1}^2) = \mathbb{E}(3X_n^2 + X_n)/4$. En déduire que $\mathbb{E}(Z^2) = \mathbb{E}(Z) = p$.
4. Établir que toute variable aléatoire W telle que $0 \leq W \leq 1$ et $\mathbb{E}(W(1-W)) = 0$ est une variable de loi de Bernoulli. En déduire la loi de Z.
5. Pour tout $n \geq 0$, on pose $Y_n = 2X_{n+1} - X_n$. Déterminer les probabilités $\mathbb{P}(Y_n = 0|\mathscr{F}_n)$ et $\mathbb{P}(Y_n = 1|\mathscr{F}_n)$. En déduire la loi de Y_n.
6. On considère les événements $G_n = \{Y_n = 1\}$ et $H_n = \{Y_n = 0\}$. Montrer que, lorsque n tend vers l'infini, Y_n converge presque sûrement vers Z. En déduire que

$$\mathbb{P}\left(\liminf_{n\to\infty} G_n\right) = p \quad \text{et} \quad \mathbb{P}\left(\liminf_{n\to\infty} H_n\right) = 1-p.$$

Les variables $(Y_n)_{n\geq 0}$ sont-elles indépendantes ?
7. Quelle est l'interprétation des résultats des points 4, 5, et 6 en termes de victoire/défaite du joueur ?

5. Corrigés

CORRIGÉ DE L'EXERCICE 3.1. Pour des lois discrètes, connaître la loi conditionnelle revient à connaître les probabilités conditionnelles. Il suffit donc de calculer, pour tous entiers $0 \leq k \leq n$,

$$\mathbb{P}(X = k|X+Y = n) = \frac{\mathbb{P}(X = k \text{ et } X+Y = n)}{\mathbb{P}(X+Y = n)} = \frac{\mathbb{P}(X = k \text{ et } Y = n-k)}{\mathbb{P}(X+Y = n)}. \tag{16}$$

Par indépendance de X et Y,

$$\mathbb{P}(X = k \text{ et } Y = n-k) = \mathbb{P}(X = k)\mathbb{P}(Y = n-k) = e^{-(\lambda+\mu)}\frac{\lambda^k}{k!}\frac{\mu^{n-k}}{(n-k)!}.$$

De la même manière, et en utilisant la formule du binôme de Newton,

$$\mathbb{P}(X+Y = n) = \sum_{k=0}^{n}\mathbb{P}(X = k \text{ et } Y = n-k) = \sum_{k=0}^{n} e^{-(\lambda+\mu)}\frac{\lambda^k}{k!}\frac{\mu^{n-k}}{(n-k)!} = e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^n}{n!};$$

on a ainsi (re)démontré que la somme de deux v.a. de Poisson indépendantes est encore une v.a. de Poisson, de paramètre la somme des deux paramètres. Enfin, en insérant les calculs dans la dernière expression de (16), on obtient que

$$\mathbb{P}(X = k|X+Y = n) = \binom{n}{k}\left(\frac{\lambda}{\lambda+\mu}\right)^k\left(\frac{\mu}{\lambda+\mu}\right)^{n-k}.$$

Ainsi, la loi conditionnelle de X sachant $X+Y$ est la loi binomiale de paramètres $X+Y$ et $\lambda/(\lambda+\mu)$.

CORRIGÉ DE L'EXERCICE 3.2.
1. Par définition, les $\mathscr{F}_n$ sont des sous-tribus de $\mathscr{F}$, puisque ce sont les tribus engendrées par les $\mathscr{I}_n$. De plus, pour tout $n \in \mathbb{N}$, $\mathscr{I}_n \subset \mathscr{I}_{n+1}$, ce qui implique aussi que $\mathscr{F}_n \subset \mathscr{F}_{n+1}$. $(\mathscr{F}_n)_{n\in\mathbb{N}}$ est donc bien une filtration de $(\Omega, \mathscr{F})$.
2. Comme $\mathscr{F}_n$ est la tribu engendrée par la partition finie $\mathscr{I}_n$ de Ω, on peut vérifier qu'une fonction $f : \Omega \to \mathbb{R}$ est $\mathscr{F}_n$-mesurable ssi elle est constante sur tous les éléments de $\mathscr{I}_n$. Ainsi $X_n = \mathbb{E}(X|\mathscr{F}_n)$ doit être constante sur les éléments de $\mathscr{I}_n$ et vérifier, pour tout $0 \le k \le 2^n - 1$, $\mathbb{E}(X_n \mathbf{1}_{[k/2^n;(k+1)/2^n[}) = \mathbb{E}(X\mathbf{1}_{[k/2^n;(k+1)/2^n[})$, ce qui donne $X_n(\omega) = (k+1/2)/2^n$ pour tout $\omega \in [k/2^n; (k+1)/2^n[$. Ainsi X_n est une variable aléatoire discrète, de loi uniforme sur l'ensemble $\{(k+1/2)/2^n, 0 \le k \le 2^n - 1\}$.
3. Pour tout $\omega \in \Omega$, l'élément de $\mathscr{I}_n$ qui contient ω est l'intervalle $[\lfloor 2^n\omega\rfloor/2^n; (\lfloor 2^n\omega\rfloor+1)/2^n]$, ce dont on déduit que $|X(\omega) - X_n(\omega)| = |\omega - (\lfloor 2^n\omega\rfloor + 1/2)/2^n| \le 2^{-n-1}$, donc que $X_n(\omega)$ converge vers $X(\omega)$ quand n tend vers l'infini. Si c'est vrai pour tout ω, c'est vrai a fortiori presque sûrement.

CORRIGÉ DE L'EXERCICE 3.3.
Prop. 3.5. Notons $F = L^2(\Omega, \mathscr{F}, \mathbb{P})$, qui est un espace de Hilbert pour le produit scalaire $\langle X, Y\rangle_2 = \mathbb{E}(XY)$ (la norme associée est bien la norme $\|\cdot\|_2$), et $G = L^2(\Omega, \mathscr{G}, \mathbb{P})$, qui est un sous-espace vectoriel de F. Pour $X \in F$, son projeté orthogonal sur G, noté $p_G X$ est caractérisé par la relation d'orthogonalité $\langle X - p_G X, Y\rangle = \mathbb{E}((X - p_G X)Y) = 0$, soit aussi $\mathbb{E}(XY) = \mathbb{E}((p_G X)Y)$, pour tout $Y \in G$. C'est la définition de l'espérance conditionnelle, donc $p_G X = \mathbb{E}(X|\mathscr{G})$.
On obtient ensuite la formule de Pythagore assez directement, en remarquant que $\mathbb{E}(X)$, donc $p_G X - \mathbb{E}(X)$, appartiennent à G. Ainsi, toujours par orthogonalité, $\|X - \mathbb{E}(X)\|_2^2 = \|X - p_G X\|_2^2 + \|p_G X - \mathbb{E}(X)\|_2^2$. C'est exactement l'expression de la proposition 3.5, formulée différemment.
Prop. 3.11. On peut supposer sans perte de généralité que les v.a. X_n sont centrées et vérifient $\mathbb{E}(X_{n+1}|\mathscr{F}_n) = 0$ (en effet, il suffit de prendre $\tilde{X}_n = X_n - \mathbb{E}(X_n)$ pour réduire l'hypothèse initiale à celle-ci). Dans le cas centré, il suffit de montrer que $\mathbb{E}\left(\left(\sum_{i=0}^n X_i\right)^2\right) = \sum_{i=0}^n \mathbb{E}\left(X_i^2\right)$. Le plus élégant est de le démontrer par récurrence sur $n \in \mathbb{N}$. Pour $n = 0$, c'est évident. Supposons que la formule est vraie au rang n. Alors :

$$\mathbb{E}\left(\left(\sum_{i=0}^{n+1} X_i\right)^2\right) = \mathbb{E}\left(\left(\sum_{i=0}^{n} X_i\right)^2\right) + 2\mathbb{E}\left(\left(\sum_{i=0}^{n} X_i\right)X_{n+1}\right) + \mathbb{E}\left(X_{n+1}^2\right).$$

On peut appliquer l'hypothèse de récurrence au premier terme. Quant au second, les propriétés de l'espérance conditionnelle, le fait que le processus soit adapté et l'hypothèse donnent que

$$\mathbb{E}\left(\left(\sum_{i=0}^{n} X_i\right) X_{n+1}\right) = \mathbb{E}\left(\mathbb{E}\left(\left(\sum_{i=0}^{n} X_i\right) X_{n+1} \middle| \mathscr{F}_n\right)\right) = \mathbb{E}\left(\left(\sum_{i=0}^{n} X_i\right) \mathbb{E}(X_{n+1}|\mathscr{F}_n)\right) = 0\,,$$

ce qui permet de conclure.

CORRIGÉ DE L'EXERCICE 3.4.

1. Démontrons-le par récurrence sur $n \geq m$, à $m \in \mathbb{N}$ fixé : pour $n = m$, $\mathbb{E}(X_m|\mathscr{F}_m) = X_m$ car le processus est adapté à la fitration ; pour $n = m+1$ (étape inutile mais qui aide à comprendre), $\mathbb{E}(X_{m+1}|\mathscr{F}_m) = X_m$ par définition d'une martingale ; supposons maintenant que c'est vrai au rang $n \geq m$: $\mathbb{E}(X_n|\mathscr{F}_m) = X_m$, et montrons que c'est encore vrai au rang $n+1$: par le point (3) de la proposition 3.4, la définition d'une martingale et l'hypothèse de récurrence, on a $\mathbb{E}(X_{n+1}|\mathscr{F}_m) = \mathbb{E}(\mathbb{E}(X_{n+1}|\mathscr{F}_n)|\mathscr{F}_m) = \mathbb{E}(X_n|\mathscr{F}_m) = X_m$.

2. Par le point précédent et de nouveau le point (3) de la proposition 3.4, on a $\mathbb{E}(X_n) = \mathbb{E}(\mathbb{E}(X_n|\mathscr{F}_0)) = \mathbb{E}(X_0)$.

CORRIGÉ DE L'EXERCICE 3.6.

1. Le fait que $(\mathscr{F}_n)_{n\in\mathbb{N}}$ est une filtration est immédiat. De la formule (1), on déduit immédiatement que, pour $0 \leq j \leq n$, $X_j \in \mathscr{F}_n$, donc aussi $\sigma(X_0, X_1, \ldots, X_n) \subset \mathscr{F}_n$. Inversement, pour tout $1 \leq j \leq n$, $Y_j = X_j - X_{j-1} \in \sigma(X_0, X_1, \ldots, X_n)$, donc $\mathscr{F}_n \subset \sigma(X_0, X_1, \ldots, X_n)$, on a donc l'égalité souhaitée.

2. Comme X_n est $\mathscr{F}_n$-mesurable et bornée ($X_n \in [k-n; k+n]$ avec probabilité 1), il est clair que les processus A_n, B_n et C_n sont adaptés et à termes intégrables. Reste à vérifier la propriété de martingale à proprement parler. Pour tout $n \geq 1$, en utilisant que X_n est $\mathscr{F}_n$-mesurable, Y_{n+1} indépendant de $\mathscr{F}_n$ et les propriétés correspondantes de l'espérance conditionnelle, on obtient

$$\mathbb{E}(A_{n+1}|\mathscr{F}_n) = \mathbb{E}(X_n + Y_{n+1} - (p-q)(n+1)|\mathscr{F}_n) = X_n + \mathbb{E}(Y_{n+1}) - (p-q)(n+1) = A_n$$

car $\mathbb{E}(Y_{n+1}) = p - q$. De même

$$\begin{aligned}\mathbb{E}(B_{n+1}|\mathscr{F}_n) &= \mathbb{E}\Big(\big(A_n + (Y_{n+1} - (p-q))\big)^2 - 4pq(n+1)\Big|\mathscr{F}_n\Big)\\ &= A_n^2 + \mathbb{V}(Y_{n+1}) - 4pq(n+1) = B_n\end{aligned}$$

car $\mathbb{V}(Y_{n+1}) = 4pq$. Enfin

$$\mathbb{E}(C_{n+1}|\mathscr{F}_n) = \mathbb{E}\big(\rho^{X_n+Y_{n+1}}\big|\mathscr{F}_n\big) = \rho^{X_n}\mathbb{E}\big(\rho^{Y_{n+1}}\big) = C_n$$

car $\mathbb{E}\big(\rho^{Y_{n+1}}\big) = p\rho + q\rho^{-1} = q + p = 1$. Les trois processus sont donc des martingales (en fait, B_n est un exemple de la martingale Z_n obtenue dans la décomposition de Doob énoncée formule (13)).

3. T est le temps d'atteinte de l'ensemble $A = \{0; M\}$, c'est donc un temps d'arrêt par la proposition 3.14. Profitons néanmoins de cet exercice pour le redémontrer (ce qui revient à démontrer la proposition...). Il suffit en effet de remarquer que, pour tout $n \in \mathbb{N}$,

$$\{T > n\} = \{X_0 \notin A, X_1 \notin A, \ldots, X_n \notin A\} \in \mathscr{F}_n\,,$$

donc par passage au complémentaire $\{T \leq n\} \in \mathscr{F}_n$.

4. Pour $p = q = 1/2$, $B_n = X_n^2 - n$ est une martingale. Puisque T est un temps d'arrêt, on obtient en appliquant la formule (11) que

$$\mathbb{E}(X_{T\wedge n}^2) - \mathbb{E}(T \wedge n) = \mathbb{E}(B_{T\wedge n}) = \mathbb{E}(B_0) = k^2\,, \tag{17}$$

donc la formule de l'énoncé.

On en déduit que $\mathbb{E}(T \wedge n) \leq k^2 + M^2$, car pour tout $n \in \mathbb{N}$, $X_{T\wedge n} \in \{0, 1, \ldots, M\}$ p.s. Par le théorème de convergence monotone, on a encore $\mathbb{E}(T) \leq k^2 + M^2$, donc T est d'espérance finie, soit aussi p.s. finie.

5. Toujours pour $p = q = 1/2$, $A_n = X_n$ est une martingale, à accroissements bornés (par 1), et T est un temps d'arrêt p.s. fini. On peut donc appliquer l'hypothèse (3) du théorème 3.16 d'arrêt de Doob pour obtenir $\mathbb{E}(X_T) = \mathbb{E}(X_0) = k$. Par définition de T, la v.a. X_T ne prend que les deux valeurs 0 et M, donc

$$k = \mathbb{E}(X_T) = 0 \cdot \mathbb{P}(X_T = 0) + M \cdot \mathbb{P}(X_T = M) = M(1 - \mathbb{P}(X_T = 0)),$$

d'où l'on déduit la formule pour la probabilité de ruine.

En utilisant le théorème de convergence dominée ($X_{T\wedge n}^2 \leq M^2$ p.s. pour tout $n \in \mathbb{N}$) dans le premier terme, on peut faire tendre n vers l'infini dans (17) pour obtenir que $\mathbb{E}(T) = \mathbb{E}(X_T^2) - k^2 = M^2(k/M) - k^2 = k(M - k)$.

6. Pour $p \neq q$, une application de la formule (11) à la martingale A_n donne immédiatement

$$\mathbb{E}(X_{T\wedge n}) = k + (p - q)\mathbb{E}(T \wedge n).$$

On en déduit que

$$\mathbb{E}(T \wedge n) = \frac{1}{p-q}(\mathbb{E}(X_{T\wedge n}) - k) \leq \frac{M-k}{p-q}, \tag{18}$$

puis la même borne pour $\mathbb{E}(T)$, par convergence monotone. Ainsi, dans ce cas aussi, T est d'espérance finie donc p.s. finie.

7. En appliquant la formule (11) à C_n, on obtient $\mathbb{E}(\rho^{X_{T\wedge n}}) = \rho^k$. Par convergence dominée ($\rho^{X_{T\wedge n}} \leq \max(1, \rho^M)$ p.s. pour tout $n \in \mathbb{N}$), on peut faire tendre n vers l'infini, et en déduire $\mathbb{E}(\rho^{X_T}) = \rho^k$. Comme dans le cas précédent, cette formule se réécrit

$$\rho^k = 1 \cdot \mathbb{P}(X_T = 0) + \rho^M \cdot \mathbb{P}(X_T = M) = \mathbb{P}(X_T = 0) + \rho^M(1 - \mathbb{P}(X_T = 0)),$$

d'où l'on déduit que

$$\mathbb{P}(X_T = 0) = \frac{\rho^k - \rho^M}{1 - \rho^M},$$

puis

$$\mathbb{E}(X_T) = M \cdot \mathbb{P}(X_T = M) = M\frac{1 - \rho^k}{1 - \rho^M}$$

et enfin, en faisant tendre n vers l'infini dans (18),

$$\mathbb{E}(T) = \frac{1}{p-q}(\mathbb{E}(X_T) - k) = \frac{1}{p-q}\frac{M\left(1 - \rho^k\right) - k\left(1 - \rho^M\right)}{1 - \rho^M}.$$

8. Dans le cas $p = q = 1/2$, les martingales A_n et C_n sont trop simples et ne contiennent pas assez d'information sur le processus de départ : C_n est constante égale à 1, quant à A_n, son second terme s'annule, on ne peut donc pas faire apparaître T en appliquant la formule (11). C'est pour cela qu'on est obligé dans ce cas d'utiliser une martingale plus compliquée, B_n.

CORRIGÉ DE L'EXERCICE 3.7.

1. La tribu $\sigma(X_n)$ est aussi engendrée par l'ensemble des événements de la forme $\{X_n = k\}$ (car X_n est une v.a. discrète, à valeurs entières). Il suffit donc de vérifier que, pour tous $n, k \in \mathbb{N}$,

$$\mathbb{E}\left(s^{X_{n+1}}\mathbf{1}_{X_n=k}\right) = \mathbb{E}\left(\varphi(s)^{X_n}\mathbf{1}_{X_n=k}\right).$$

Et en effet, par indépendance de X_n et des $Y_i^{(n+1)}$,

$$\mathbb{E}\Big(s^{X_{n+1}}\mathbf{1}_{X_n=k}\Big) = \mathbb{E}\Big(s^{\sum_{i=1}^{X_n} Y_i^{(n+1)}}\mathbf{1}_{X_n=k}\Big) = \mathbb{E}\Big(s^{\sum_{i=1}^{k} Y_i^{(n+1)}}\mathbf{1}_{X_n=k}\Big)$$
$$= \varphi(s)^k\mathbb{P}(X_n = k) = \mathbb{E}\big(\varphi(s)^{X_n}\mathbf{1}_{X_n=k}\big).$$

2. $\varphi_0(s) = \mathbb{E}\big(s^{X_0}\big) = s^1 = s$, et

$$\varphi_{n+1}(s) = \mathbb{E}\big(s^{X_{n+1}}\big) = \mathbb{E}\big(\mathbb{E}\big(\, s^{X_{n+1}} \,\big|\, X_n\big)\big) = \mathbb{E}\big(\varphi(s)^{X_n}\big) = \varphi_n\big(\varphi(s)\big).$$

On en déduit par une récurrence immédiate que $\varphi_n(s) = \varphi^n(s)$ pour tout $n \in \mathbb{N}$ et tout $s \in [0;1]$.

3. Pour tout $n \in \mathbb{N}$ et pour tout $s_0 \in [0;1[$, la série $\sum_{k\geq n} k(k-1)\cdots(k-n+1)s^{k-n}$ est normalement convergente sur $[0;s_0]$, donc la série $\sum_{k\geq n} p_k k(k-1)\cdots(k-n+1)s^{k-n}$ aussi. Par applications successives du théorème de dérivabilité des séries de fonctions, on en déduit que φ est $\mathscr{C}^\infty$ sur $[0;s_0]$ pour tout $s_0 < 1$, donc sur $[0;1[$, avec

$$\varphi'(s) = \sum_{k\in\mathbb{N}^*} kp_k s^{k-1}$$
$$\varphi''(s) = \sum_{k\geq 2} k(k-1)p_k s^{k-2}.$$

On déduit ensuite des hypothèses faites sur $\mathbf{p}$ que $\varphi'(s) > 0$ et $\varphi''(s) > 0$ pour tout $s \in]0;1[$, donc que φ est strictement croissante et strictement convexe. D'autre part, $\lim_{s\to 1}\varphi'(s) = m$, donc, par un théorème classique d'analyse, φ est dérivable (à gauche) en 1, de dérivée $\varphi'(1) = m$.

Comme par ailleurs, $\varphi(0) > 0$ et $\varphi(1) = 1$, on a par continuité, croissance et stricte convexité de φ que 1 est l'unique point fixe de cette fonction si $m \leq 1$ (car une fonction strictement convexe est strictement au dessus de sa pente en 0) et qu'il existe un unique autre point fixe $q \in]0;1[$ si $m > 1$ (par le théorème des valeurs intermédiaires et croissance stricte).

4. En prenant $s = 0$ dans la définition, on remarque que $\mathbb{P}(X_n = 0) = \varphi_n(0) = \varphi^n(0)$. La suite $\mathbb{P}(X_n = 0)$ est croissante (une fois que la lignée s'éteint, c'est pour toujours), de limite $\mathbb{P}(\text{extinction})$. Ainsi

$$\mathbb{P}(\text{extinction}) = \lim_{n\to+\infty}\varphi^n(0).$$

Il est particulièrement simple d'étudier une suite définie par récurrence $u_{n+1} = \varphi(u_n)$ lorsque φ est continue et croissante : la suite converge en croissant vers le plus petit point fixe au dessus de la valeur initiale u_0. Ce résultat permet de conclure.

5. Il est clair que les processus Z_n et M_n sont adaptés. D'autre part, on peut montrer, comme dans la question 1., que $\mathbb{E}(X_{n+1}|\mathscr{F}_n) = mX_n$. On en déduit facilement que Z_n et M_n sont des martingales.

6. La suite Z_n est bornée dans L^1 (puisqu'à valeurs positives et d'espérance constante égale à 1), il suffit donc d'appliquer le théorème 3.19 de convergence p.s. des martingales pour conclure.

7. Pour $m \leq 1$, on a montré qu'il y a extinction avec probabilité 1, ce qui implique que, presque sûrement, $Z_n = 0$ à partir d'un certain rang, donc $Z_\infty = 0$. Donc $\mathbb{E}(Z_\infty) = 0 \neq 1 = \mathbb{E}(Z_n)$, ce qui rend impossible la convergence de Z_n vers Z_∞ dans L^1.

8. On calcule $\mathbb{E}\big[\,(Z_{n+1} - Z_n)^2\big|\,\mathscr{F}_n\big] = \sigma^2\frac{Z_n}{m^{n+1}}$, donc $\langle Z\rangle_n = \sigma^2\sum_{k=0}^{n-1}\frac{Z_k}{m^{k+1}}$, puis en passant à la limite $\langle Z\rangle_\infty = \sigma^2\sum_{k\in\mathbb{N}}\frac{Z_k}{m^{k+1}}$. Comme les Z_k sont d'intégrale 1, on en déduit que $\mathbb{E}(\langle Z\rangle_\infty) = \sigma^2\sum_{k\in\mathbb{N}} m^{-k-1} < +\infty$,

9. Par la remarque 3.25, le résultat précédent implique que la martingale Z_n est bornée dans L^2. Le théorème 3.20 implique alors que la convergence de Z_n vers Z_∞ a aussi lieu

dans L^2 (et dans L^1 car dans un espace de probabilité, toute suite convergente dans L^2 l'est encore dans L^1). On en déduit que $\mathbb{E}(Z_\infty) = 1$.
10. Comme dans la question 8., il suffit de calculer que

$$\mathbb{E}\left[(M_{n+1} - M_n)^2 \middle| \mathscr{F}_n\right] = \mathbb{E}\left[(X_{n+1} - mX_n)^2 \middle| \mathscr{F}_n\right] = \sigma^2 X_n.$$

11. Z_n converge p.s. vers Z_∞ (question 6.) et $\langle M\rangle_n = \sigma^2 \sum_{k=0}^{n-1} X_k = \sigma^2 \sum_{k=0}^{n-1} m^k Z_k$. On peut donc appliquer le lemme de Toeplitz pour obtenir que

$$\lim_{n\to+\infty} \frac{\langle M\rangle_n}{\sum_{k=0}^{n-1} m^k} = \sigma^2 Z_\infty,$$

ce qui équivaut à la formule (15).
Ainsi, pour tout $\omega \in \Omega$ tel que $Z_\infty(\omega) > 0$, la suite $\langle M\rangle_n(\omega)$ diverge comme m^n, donc tend vers l'infini, ce qui implique que $\langle M\rangle_\infty(\omega) = +\infty$.
12. Calcul direct.
13. Application de la loi des grands nombres (théorème 3.27).

CORRIGÉ DE L'EXERCICE 3.8.
1. La suite $(X_n)_{n\geq 0}$ est adaptée à la filtration $(\mathscr{F}_n)_{n\geq 0}$ par définition. Pour tout $n \in \mathbb{N}$, X_n est à valeurs dans $[0,1]$ donc intégrable. Enfin,

$$\mathbb{E}(X_{n+1}|\mathscr{F}_n) = X_n\left(X_n + \frac{1-X_n}{2}\right) + (1-X_n)\left(\frac{X_n}{2}\right) = X_n.$$

2. La suite $(X_n)_{n\geq 0}$ est une martingale bornée. Elle converge donc p.s. et dans L^2 vers une variable aléatoire Z à valeurs dans $[0,1]$.
3. On a

$$\mathbb{E}(X_{n+1}^2|\mathscr{F}_n) = X_n\left(X_n + \frac{1-X_n}{2}\right)^2 + (1-X_n)\left(\frac{X_n}{2}\right)^2 = \frac{X_n}{4}(3X_n+1).$$

En prenant l'espérance, on obtient $\mathbb{E}(X_{n+1}^2) = \mathbb{E}(3X_n^2 + X_n)/4$. Puisque la suite $(X_n)_{n\geq 1}$ est bornée, le théorème de convergence dominée permet de passer à la limite dans l'espérance pour obtenir $\mathbb{E}(Z^2) = \mathbb{E}(3Z^2 + Z)/4$. On en déduit donc que $\mathbb{E}(Z^2) = \mathbb{E}(Z)$. Enfin, puisque $(X_n)_{n\geq 0}$ est une martingale, $\mathbb{E}(Z) = \mathbb{E}(X_0) = p$.
4. Puisque $W \in [0,1]$, on a $W(1-W) \geq 0$. Si cette variable positive est d'espérance nulle c'est que $\mathbb{P}(W \in \{0,1\}) = 1$. En d'autres termes, W suit une loi de Bernoulli. Ceci assure donc que $Z \sim \mathscr{B}(p)$ puisque $\mathbb{E}(Z) = p$.
5. Par définition, $Y_n = 1$ si le joueur gagne la partie $n+1$ et $Y_n = 0$ s'il la perd. On a donc

$$\mathbb{P}(Y_n = 0|\mathscr{F}_n) = X_n \quad \text{et} \quad \mathbb{P}(Y_n = 0|\mathscr{F}_n) = 1 - X_n.$$

En prenant l'espérance, on obtient que $Y_n \sim \mathscr{B}(p)$.
6. Puisque $(X_n)_{n\geq 0}$ converge vers Z, la suite $(Y_n)_{n\geq 0}$ converge vers $2Z - Z = Z$. Comme Y_n est à valeurs dans $\{0,1\}$, c'est que la suite $(Y_n)_{n\geq 0}$ est stationnaire : à partir d'un certain rang, $(Y_n(\omega))_{n\geq 0}$ est constante. Cette constante dépend bien entendu de ω et vaut Z donc

$$\mathbb{P}\left(\liminf_{n\to\infty} G_n\right) = p \quad \text{et} \quad \mathbb{P}\left(\liminf_{n\to\infty} H_n\right) = 1 - p.$$

Les variables $(Y_n)_{n\geq 0}$ ne sont pas indépendantes puisqu'elles sont égales à partir d'un certain rang et non constantes.
7. À partir d'un certain nombre (aléatoire) de parties, c'est toujours le même joueur qui gagnera à chaque fois même si le nombre de parties est infini puisque les capitaux des deux joueurs ne sont jamais nuls.

A. Deux lemmes analytiques

Nous collectons ici deux lemmes purement analytiques, qui sont souvent utiles dans le contexte des martingales et des théorèmes limite pour des suites de variables aléatoires non i.i.d.

LEMME 3.29 (Lemme de Toeplitz). *Soit $(a_k)_{k\in\mathbb{N}}$ une suite de réels strictement positifs telle que $\sum_{k\in\mathbb{N}} a_k = +\infty$, et soit $(x_k)_{k\in\mathbb{N}}$ une suite de réels convergente, de limite $l \in \mathbb{R}$. Alors*

$$\frac{\sum_{k=1}^{n} a_k x_k}{\sum_{k=1}^{n} a_k} \xrightarrow[n\to+\infty]{} l\,.$$

LEMME 3.30 (Lemme de Kronecker). *Soit $(b_n)_{n\in\mathbb{N}}$ une suite croissante de réels strictement positifs tendant vers $+\infty$, soit $(x_k)_{k\in\mathbb{N}}$ une suite de réels, et posons $s_n = \sum_{k=0}^{n-1} x_k$. Alors*

$$\sum_{k\in\mathbb{N}} \frac{x_k}{b_k} \text{ converge} \quad \Longrightarrow \quad \lim_{n\to+\infty} \frac{s_n}{b_n} = 0\,.$$

Le lemme de Toeplitz est une généralisation du lemme classique de Cesàro (qui correspond au cas où $a_k = 1$) et se démontre de la même manière. Le lemme de Kronecker se déduit aisément du précédent avec une transformation d'Abel.

Chapitre 4

Chaînes de Markov

Dans tout ce qui suit E est un espace de cardinal **fini ou dénombrable**. Lorsqu'il est fini, de cardinal d, on l'identifiera très souvent à $\{1,\dots,d\}$, lorsqu'il est (infini) dénombrable, on prendra souvent $\mathbb{N}$.

1. Matrices de transition et chaînes de Markov

1.1. Définition classique.

DÉFINITION 4.1. Une matrice de transition (ou noyau de transition ou matrice markovienne) sur l'ensemble E est une application $P : E \times E \to [0,1]$ telle que

$$\forall x \in E, \quad \sum_{y\in E} P(x,y) = 1.$$

Une chaîne de Markov de matrices de transition $(P_n)_{n\geq 0}$ et de loi initiale μ est une suite de variables aléatoires $X = (X_n)_{n\in\mathbb{N}}$ définie sur un espace probabilisé $(\Omega, \mathscr{F}, \mathbb{P})$ et à valeurs dans E telle que $\mathscr{L}(X_0) = \mu$ et

$$\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = \mu(\{x_0\}) \prod_{i=0}^{n-1} P_i(x_i, x_{i+1}). \tag{19}$$

pour tout $n \in \mathbb{N}$ et tout $(n+1)$-uplet $(x_0,\dots,x_n)$ d'éléments de E.

DÉFINITION 4.2. La chaîne est dite homogène si la suite de matrices markoviennes $(P_n)_{n\geq 0}$ est constante.

1.2. Premiers exemples.

EXEMPLE 4.3 (Marche aléatoire sur les sommets du cube). L'espace d'états est $E = \{0,1\}^N$ les sommets du cube unité dans $\mathbb{R}^N$ et la matrice de transition est définie ainsi : pour $x, y \in E$, en notant $|x-y| = \sum_{i=1}^N |x_i - y_i|$,

$$P(x,y) = \begin{cases} 1/(2N) & \text{si } |x-y| = 1, \\ 0 & \text{sinon.} \end{cases}$$

EXEMPLE 4.4 (Urne d'Ehrenfest). L'espace d'états est $\{0,1,\dots,N\}$ et la matrice de transition est définie par

$$P(x,y) = \begin{cases} x/N & \text{si } y = x-1, \\ (N-x)/N & \text{si } y = x+1, \\ 0 & \text{sinon.} \end{cases}$$

Une chaîne de Markov X de matrice de transition P modélise la situation physique extrêmement simplifiée suivante : N particules évoluent dans un environnement constitué de deux compartiments, notés A et B, qui communiquent, et à chaque pas de temps, une particule choisie au hasard (uniformément) parmi les N particules

passe d'un compartiment à l'autre. X_n représente alors le nombre de particules dans le compartiment A (par exemple, il y a une symétrie évidente dans le modèle) au temps n, car on vérifie en effet facilement que le noyau P construit ci-dessus traduit mathématiquement la description de l'évolution. Ce modèle a été introduit par les époux Ehrenfest en 1907 pour illustrer l'un des points clefs de la thermodynamique, à savoir l'apparent paradoxe entre une description microscopique par la mécanique qui est réversible en temps et une description macroscopique par la thermodynamique qui est irréversible. C'est donc une version extrêmement simplifiée des intuitions de Boltzmann, très mal accueillies par la communauté scientifique en son temps.
La Figure 1 représente des trajectoires de la chaîne de Markov modélisant l'urne d'Ehrenfest.

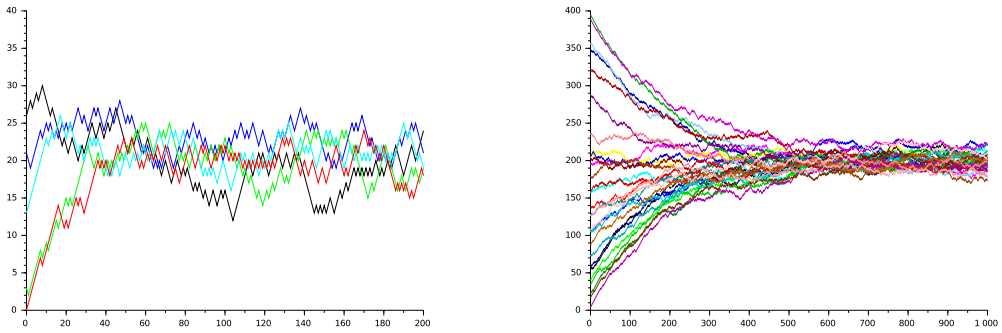

FIGURE 1. Urne d'Ehrenfest : à gauche, 5 trajectoires jusqu'au temps 200, pour $N = 40$ – à droite, 30 trajectoires jusqu'au temps 1000, pour $N = 400$ (valeurs initiales tirées uniformément).

EXEMPLE 4.5 (Modèle de Wright-Fisher). L'espace d'états est $\{0, 1, \ldots, N\}$ et

$$P(x, y) = \binom{N}{y}\left(\frac{x}{N}\right)^y\left(1 - \frac{x}{N}\right)^{N-y}.$$

En d'autres termes, $\mathscr{L}(X_{n+1}|X_n = x)$ est la loi binomiale $\mathscr{B}(N, x/N)$.

Ce modèle est utilisé en génétique pour modéliser l'évolution de la fréquence d'un allèle dans une population de petite taille. Nous faisons les hypothèses suivantes :

— la population est haploïde : chaque individu possède un seul exemplaire du gène d'intérêt ; et ce gène se présente sous deux allèles distincts A et B ;

— la taille de la population reste constante au cours du temps, égale à N ;

— les générations ne se chevauchent pas : à chaque instant k, la k-ième génération meurt et donne naissance aux N individus de la $(k+1)$-ième génération ;

— chacun des enfants choisit son parent uniformément parmi tous les individus de la génération précédente et indépendamment des autres ;

— et la reproduction à l'instant k ne dépend pas des reproductions précédentes.

Sous toutes ces hypothèses, si on note X_n le nombre d'allèles A dans la population au temps n (c'est-à-dire à la génération n), on peut vérifier que $X = (X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov de matrice de transition P.
La figure 2 représente des trajectoires de la chaîne de Markov modélisant le modèle de Wright-Fisher. On voit assez bien intuitivement que le comportement est radicalement différence de celui de l'urne d'Ehrenfest.

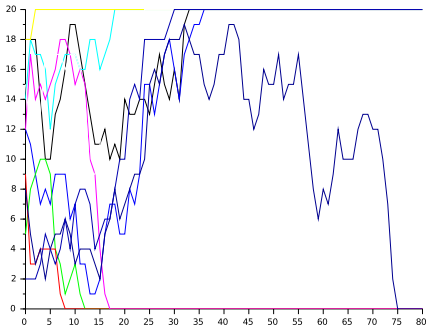

FIGURE 2. Modèle de Wright-Fisher : 10 trajectoires jusqu'au temps 80, pour $N = 20$ (valeurs initiales tirées uniformément).

EXEMPLE 4.6 (File d'attente M/M/1 (à temps discret)). Nous souhaitons modéliser le nombre de clients attendant à un guichet, sous les hypothèses simplifiées suivantes : à chaque pas de temps, avec probabilité $p \in]0;1[$, un nouveau client arrive dans la file, et avec probabilité $q = 1-p$, un client (s'il y en a un) est servi et quitte donc la file d'attente, tout ceci indépendamment de l'état de la file d'attente et de ce qui s'est passé précédemment.
Si on note X_n le nombre de clients dans la file d'attente au temps n, on peut vérifier que $X = (X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov sur l'espace d'états $\mathbb{N}$ de matrice de transition définie par

$$\begin{cases} P(x,x+1) = p & \text{pour tout } x \in \mathbb{N}\,; \\ P(0,0) = q\,; & \\ P(x,x-1) = q & \text{pour tout } x \in \mathbb{N}^*\,; \\ P(x,y) = 0 & \text{sinon.} \end{cases}$$

Dans la suite, nous considérerons, sauf mention contraire, des chaînes de Markov homogènes.

1.3. Suites récurrentes aléatoires.

DÉFINITION 4.7. Une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires à valeurs dans E est une suite récurrente aléatoire s'il existe :

— une suite $(U_n)_{n\in\mathbb{N}^*}$ de variables aléatoires indépendantes de loi uniforme sur l'intervalle $[0,1]$, et indépendantes de X_0 ;

— une fonction $f : E \times [0,1] \longrightarrow E$;

telles que, pour tout $n \in \mathbb{N}^*$,

$$X_n = f(X_{n-1}, U_n).$$

REMARQUE 4.8. Puisque toute variable aléatoire réelle Y peut être réalisée comme fonction d'une variable aléatoire U de loi uniforme sur $[0,1]$ (par la formule $Y = F_Y^{-1}(U)$, où F_Y^{-1} désigne l'inverse généralisée de la fonction de répartition de Y), cette définition équivaut à dire qu'il existe

— une suite $(Y_n)_{n\in\mathbb{N}^*}$ de variables aléatoires réelles indépendantes et identiquement distribuées, et indépendantes de X_0 ;

— une fonction $f : E \times \mathbb{R} \longrightarrow E$;

telles que, pour tout $n \in \mathbb{N}^*$,

$$X_n = f(X_{n-1}, Y_n).$$

REMARQUE 4.9. Cette définition est une généralisation naturelle des suites récurrentes réelles $x_{n+1} = f(x_n)$. Ici, on fait la même chose, en ajoutant un ingrédient aléatoire à chaque étape.

Les deux définitions que nous avons données sont en fait équivalentes.

THÉORÈME 4.10. *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires à valeurs dans E. Alors $(X_n)_{n\in\mathbb{N}}$ est une suite récurrente aléatoire si et seulement si $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov homogène.*

REMARQUE 4.11. Toute chaîne de Markov (même non homogène) pourrait être vue comme une suite récurrente aléatoire, à condition de remplacer la fonction f par une suite de fonctions $(f_n)_{n\in\mathbb{N}}$.

EXEMPLE 4.12. La marche au hasard issue de 0 sur $\mathbb{Z}$, $X_n = \sum_{k=1}^n Y_k = X_{n-1} + Y_n$, avec $(Y_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires i.i.d. à valeurs dans $\mathbb{Z}$, est une suite récurrente aléatoire, donc une chaîne de Markov.

REMARQUE 4.13. Dans un grand nombre de situations de modélisation, on montre que l'objet mathématique étudié est une chaîne de Markov homogène en montrant que c'est une suite récurrente aléatoire.

1.4. Graphes probabilistes. Une troisième définition possible, équivalente aux deux premières (ce n'est finalement qu'une traduction graphique de la définition 4.1) est de dire qu'une chaîne de Markov est définie par le graphe probabiliste associé :

DÉFINITION 4.14. Un graphe probabiliste sur E est un graphe d'ensemble des sommets E, qui est orienté et étiqueté (à chaque arête (i,j), on associe une étiquette $P(i,j)$) avec des étiquettes positives et dont la somme sur les arêtes sortantes de chaque sommet vaut 1. Une chaîne de Markov associée à ce graphe probabiliste est une suite de variables aléatoires obéissant à la règle suivante : si je suis au sommet i à un certain instant, je saute à l'instant suivant le long de l'une des arêtes sortantes du sommet i, chaque arête étant choisie avec une probabilité égale à son étiquette.

2. Équation de Chapman-Kolmogorov

Il faut voir une matrice markovienne P comme un opérateur agissant sur les fonctions et les mesures de la manière suivante. L'opérateur P associe à une fonction h (bornée) de E dans $\mathbb{R}$ la fonction Ph (bornée) de E dans $\mathbb{R}$ définie par

$$\forall x \in E, \quad Ph(x) = \sum_{y\in E} P(x,y)h(y)$$

et à une mesure positive (respectivement une probabilité) μ sur E la mesure positive (respectivement la probabilité) μP sur E définie par

$$\forall y \in E, \quad \mu P(y) = \sum_{x \in E} \mu(x) P(x, y).$$

On peut aussi composer ces opérateurs : on vérifie facilement que si P et Q sont deux matrices markoviennes, alors PQ défini par

$$\forall x, y \in E, \quad PQ(x, y) = \sum_{z \in E} P(x, z) Q(z, y)$$

est encore une matrice markovienne et vérifie $(PQ)h = P(Qh)$ pour toute fonction h et $\mu(PQ) = (\mu P)Q$ pour toute mesure μ. On note ensuite P^n pour l'itérée n-ème de P.

REMARQUE 4.15. Le nom de matrice markovienne vient bien sûr du fait que, dans le cas où E est fini et identifié avec $\{1, \dots, d\}$, une matrice stochastique est une « vraie » matrice, et toutes les opérations définies ci-dessus sont des produits matriciels, en identifiant les fonctions à des vecteurs colonnes et les mesures à des vecteurs lignes.
Dans le cas infini dénombrable, tout le formalisme du produit matriciel subsiste, mais avec des matrices infinies, et donc aussi des séries à la place des sommes dans les formules précédentes.

THÉORÈME 4.16. *Soit $X = (X_n)_{n \in \mathbb{N}}$ une chaîne de Markov homogène sur E de matrice de transition P et de loi initiale μ_0. Notons μ_n la loi de X_n. Alors*

— *la suite $(\mu_n)_{n \in \mathbb{N}}$ vérifie la relation de récurrence, appelée équation de Chapman-Kolmogorov, suivante :*

$$\mu_{n+1} = \mu_n P = \mu_0 P^{n+1}, \tag{20}$$

— *pour tous $x, y \in E$, $\mathbb{P}(X_n = y | X_0 = x) = P^n(x, y)$,*

— *pour toute fonction h de E dans $\mathbb{R}$,*

$$\mathbb{E}(h(X_n) | X_0 = x) = (P^n h)(x).$$

— *pour toute fonction h de E dans $\mathbb{R}$ et toute probabilité μ_0 sur E, si $\mathscr{L}(X_0) = \mu_0$ alors*

$$\mathbb{E}(h(X_n)) = \mu_0(P^n h) = (\mu_0 P^n) h = \mu_0 P^n h.$$

3. Propriété de Markov

La propriété de Markov exprime le fait que « conditionnellement au présent, passé et futur sont indépendants ». Introduisons un peu de formalisme. Soit X une chaîne de Markov sur E. On peut voir X comme une variable aléatoire à valeurs dans l'espace des trajectoires $\mathbb{T} = E^{\mathbb{N}}$. Pour cela, on munit $\mathbb{T}$ de la tribu engendrée par les tribus cylindriques $(\mathscr{B}_n(\mathbb{T}))_{n \in \mathbb{N}}$ définies par $\mathscr{B}_n(\mathbb{T}) = \sigma(X_0, \dots, X_n)$.
Si X_0 est de loi μ, on note $\mathbb{P}_\mu$ la loi de X et si $\mu = \delta_x$, on note $\mathbb{P}_{\delta_x} = \mathbb{P}_x$.

THÉORÈME 4.17. *La loi d'une chaîne de Markov homogène sur E de matrice de transition P et de loi initiale μ est l'unique mesure de probabilité $\mathbb{P}_\mu$ sur $(\mathbb{T}, \mathscr{B}(\mathbb{T}))$ caractérisée par* (19).

On introduit à présent le processus décalé de n : $X_{n+} = (X_{n+,k})_{k \in \mathbb{N}}$ défini par

$$X_{n+,k} = X_{n+k}, \quad \text{pour } k \in \mathbb{N}.$$

THÉORÈME 4.18 (Propriété de Markov). *Soit X une chaîne de Markov sur E de loi initiale μ. Pour tout $A \in \mathscr{B}(\mathbb{T})$ et $k \in \mathbb{N}$,*

$$\mathbb{P}(X_{k+} \in A | X_k = x_k, \dots, X_0 = x_0) = \mathbb{P}_{x_k}(A).$$

REMARQUE 4.19. On peut montrer que, pour une chaîne de Markov à espace d'états fini ou dénombrable, cette propriété est encore vraie si l'on remplace k par un temps d'arrêt. On parle alors de propriété de Markov forte [**BL98**, Thm. VIII.3.2].

4. Récurrence/transience. Irréductibilité

La première idée naturelle pour étudier une chaîne de Markov $(X_n)_{n\in\mathbb{N}}$ à valeurs dans E est d'étudier par quels points de E elle passe, et combien de temps elle met à revenir à son point de départ. On définit pour cela les temps de premier retour et les nombres de passages en i : pour tout $i \in E$,

$$\tau_i = \inf\{n \in \mathbb{N}^* : X_n = i\},$$
$$N_i = \sum_{n\in\mathbb{N}} \mathbf{1}_{X_n = i}.$$

On dit alors que $i \in E$ est un :

— état récurrent si $\mathbb{P}_i(N_i = +\infty) = 1$ (partant de i, on repasse presque sûrement une infinité de fois par i)

— état transient (ou transitoire) si $\mathbb{P}_i(N_i = +\infty) < 1$.

Tous ces objets sont liés par une dichotomie :

THÉORÈME 4.20. *Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov homogène à valeurs dans E. Alors, pour tout $i \in E$,*

1. *Si $\mathbb{P}_i(\tau_i < +\infty) = 1$, alors i est un état récurrent et $\mathbb{E}_i(N_i) = \sum_{n=0}^{+\infty} P^n(i,i) = +\infty$.*
2. *Si $\mathbb{P}_i(\tau_i < +\infty) < 1$, alors i est un état transient et $\mathbb{E}_i(N_i) = \sum_{n=0}^{+\infty} P^n(i,i) < +\infty$.*

DÉFINITION 4.21. Soit P une matrice stochastique sur E et $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov de matrice de transition P.

1. Pour $i, j \in E$, on dit que i conduit à j, noté $i \to j$ s'il existe $n \in \mathbb{N}$ tel que $P^n(i,j) > 0$; on dit que i et j communiquent, noté $i \leftrightarrow j$ si $i \to j$ et $j \to i$. $\leftrightarrow$ est une relation d'équivalence, qui partitionne donc E en classes (d'équivalence).
2. On dit que la matrice P (ou une chaîne de Markov associée) est irréductible si E est l'unique classe (tous les états communiquent entre eux).
3. Une classe C est dite close si pour tout $i \in C$, $\mathbb{P}_i(\forall n \in \mathbb{N}, X_n \in C) = 1$. Si $C = \{i\}$ est une classe close, on dit que i est un état absorbant (ce qui équivaut à $P(i,i) = 1$).

PROPOSITION 4.22. *Soit P une matrice stochastique sur E et $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov de matrice de transition P.*

1. *S'il existe $j \in E$ tel que $i \to j$ et $j \not\to i$, alors i est transient.*
2. *Deux états qui communiquent sont de même nature (tous les deux récurrents ou transients). On peut donc parler de classe récurrente ou de classe transiente.*
3. *Si la chaîne est irréductible, tous les états sont de même nature. On dit alors que la chaîne de Markov elle-même est récurrente ou transiente.*
4. *Une classe récurrente est close ; une classe close et finie est récurrente.*
5. *Toute chaîne de Markov irréductible sur un espace d'états **fini** est récurrente.*

Modèle de Wright-Fisher. Les deux états 0 et N sont absorbants. $\{1, 2, \ldots, N-1\}$ est une classe transiente.
On en déduit que $T = \inf\{n \in \mathbb{N} : X_n \in \{0, N\}\}$ est p.s. fini (on appelle ce phénomène la dérive génétique : dans une petite population, l'un des allèles disparaît au bout d'un certain temps), et un argument de martingale donnerait que $\mathbb{P}_i(X_T = N) = 1 - \mathbb{P}_i(X_T = 0) = i/N$.

File d'attente M/M/1 (à temps discret). La chaîne de Markov modélisant la file d'attente M/M/1 est irréductible. Elle est :

— transiente pour $p > 1/2$;

— récurrente pour $p \leq 1/2$.

5. Mesures invariantes, mesures réversibles

5.1. Définitions.

DÉFINITION 4.23. Soit P une matrice de transition sur E. Une mesure π sur E est dite invariante par P (ou P-invariante) si elle vérifie

$$\pi P = \pi \,. \tag{21}$$

REMARQUE 4.24. Une mesure invariante est un vecteur propre à gauche associé à la valeur propre 1 pour P. Dans la description « suites récurrentes aléatoires », une mesure de probabilité invariante π peut aussi être vue comme une sorte de point fixe, au sens où une variable aléatoire X de loi π vérifie $X \overset{\mathscr{L}}{=} f(X,U)$, avec U uniforme sur $[0,1]$ indépendante de X.

REMARQUE 4.25. En associant la définition 4.23 et la formule de Chapman-Kolmogorov (20), on observe le fait suivant : si π est une mesure de probabilité invariante pour P et $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov de matrice de transition P et de mesure initiale π, alors X_n est de loi π pour tout $n \in \mathbb{N}$. Cette remarque justifie l'utilisation du terme *invariante* : si la condition initiale est tirée selon une mesure invariante, la chaîne est distribuée selon cette mesure invariante à tout temps.

Définissons aussi ce qu'est une mesure réversible :

DÉFINITION 4.26. Soit $P = (P(i,j))_{i,j\in E}$ une matrice de transition sur E. Une mesure de probabilité $\pi = (\pi(i))_{i\in E}$ sur E est dite réversible pour P (ou P-réversible) si elle vérifie

$$\pi(i)P(i,j) = \pi(j)P(j,i) \quad \text{pour tous } i,j \in E\,. \tag{22}$$

C'est une notion plus forte, comme le montrent la Proposition ci-dessous et la remarque qui la suit, plus facile à caractériser (les calculs sont plus simples), mais aussi plus rare.

PROPOSITION 4.27. *Soit P une matrice de transition et π une mesure sur E. Si π est P-réversible, alors π est aussi P-invariante.*

On peut vérifier que si π est une mesure réversible pour P et $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov de matrice de transition P et de mesure initiale π, alors, pour tout $n \in \mathbb{N}$, pour tout chemin $(i_0, i_1, i_2, \ldots, i_{n-1}, i_n)$ dans E, on a :

$$\begin{aligned}\mathbb{P}(X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots, X_{n-1} = i_{n-1}, X_n = i_n)\\ = \mathbb{P}(X_0 = i_n, X_1 = i_{n-1}, X_2 = i_{n-2}, \ldots, X_{n-1} = i_1, X_n = i_0)\,.\end{aligned}$$

Ainsi les probabilités d'observer un chemin donné dans un ordre ou dans l'ordre inverse coïncident : la réversibilité est une invariance par retournement du temps.

EXEMPLE 4.28. La marche aléatoire sur le cube admet une unique mesure invariante (la loi uniforme sur E) et la chaîne est réversible par rapport à cette probabilité.

Modèle de Wright-Fisher Toutes les mesures $p\delta_0 + (1-p)\delta_N$, pour $p \in [0,1]$ sont des mesures invariantes de la chaîne de Wright-Fisher.

5.2. Existence. Unicité. Le problème de l'existence de mesures invariantes pour une chaîne de Markov est en fait en partie lié à celui de la récurrence/transience étudié précédemment.
Nous avons par exemple un premier résultat :

PROPOSITION 4.29. *Une chaîne de Markov irréductible et admettant une mesure de probabilité invariante est récurrente.*

REMARQUE 4.30. Dans le cas où E est **fini**, un argument de compacité (dans ce cas, l'ensemble $\mathscr{P}(E)$ des mesures de probabilité sur E est compact) assure l'existence d'au moins une mesure de probabilité invariante.
Si E est **infini dénombrable**, en revanche, rien n'assure l'existence d'une mesure de probabilité invariante (c'est même impossible dans le cas transient, par la proposition ci-dessus), et il faut donc considérer aussi le cas des mesures σ-finies.

REMARQUE 4.31. Inversement, l'existence d'une mesure σ-finie invariante n'assure pas la récurrence : par exemple, la marche simple sur $\mathbb{Z}^3$ est transiente bien que la mesure uniforme soit invariante.

THÉORÈME 4.32. *Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov de matrice de transition P sur E, que l'on suppose irréductible et récurrente. Alors il existe une mesure π invariante par P, et elle est unique à une constante multiplicative près. Elle vérifie $0 < \pi(i) < +\infty$ pour tout $i \in E$. De plus, pour tout $i \in E$, la mesure π_i telle que $\pi_i(i) = 1$ est donnée par*

$$\pi_i(j) = \mathbb{E}_i\Big(\sum_{k=0}^{\tau_i - 1} \mathbf{1}_{\{j\}}(X_k)\Big). \tag{23}$$

REMARQUE 4.33. Attention, dans ce théorème, la mesure invariante π n'est pas forcément une mesure finie, ça peut être une mesure σ-finie. On règle ce problème avec le résultat suivant, qui repose sur le fait que la mesure π_i définie dans le théorème précédent vérifie $\pi_i(E) = \sum_{j\in E} \pi_i(j) = \mathbb{E}_i(\tau_i)$.

REMARQUE 4.34. Dans le cas non irréductible, la formule (23) permet encore de construire des mesures invariantes : si C est une classe récurrente et $i \in C$, alors π_i est une mesure invariante concentrée sur C (unique à une constante multiplicative près).

THÉORÈME 4.35. *Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov de matrice de transition P sur E, que l'on suppose irréductible et récurrente. Deux cas se présentent :*

1. *Soit $\mathbb{E}_i(\tau_i) = +\infty$ pour tout $i \in E$. On dit alors que la chaîne de Markov est nulle récurrente, et les mesures invariantes sont σ-finies.*
2. *Soit $\mathbb{E}_i(\tau_i) < +\infty$ pour tout $i \in E$. On dit que la chaîne de Markov est positive récurrente, et l'unique mesure de probabilité invariante π vérifie, pour tout $i \in E$,*

$$\pi(i) = \frac{1}{\mathbb{E}_i(\tau_i)} \tag{24}$$

*Notons que lorsque E est **fini**, on est forcément dans le second cas : toute chaîne de Markov irréductible sur un espace d'états **fini** est positive récurrente, donc admet une unique mesure de probabilité invariante π définie par* (24).

REMARQUE 4.36. La formule (24) se traduit par « En régime stationnaire, le temps moyen passé en un site est l'inverse du temps moyen mis pour y revenir ». C'est en fait très moral !

Urne d'Ehrenfest. La chaîne de Markov associée à l'urne d'Ehrenfest est irréductible (sur un espace d'états fini) donc récurrente positive. Elle est aussi réversible par rapport à sa mesure invariante, qui est la loi binomiale $\mathscr{B}(N, 1/2)$.

File d'attente M/M/1 (à temps discret). La chaîne de Markov modélisant la file d'attente M/M/1 est :

— récurrente nulle si $p = 1/2$. Ses mesures invariantes σ-finies sont proportionnelles à la mesure de comptage sur $\mathbb{N}$, et elles sont réversibles ;

— récurrente positive si $p < 1/2$. Son unique mesure de probabilité invariante est la loi π définie par $\pi(i) = (p/q)^i(1 - p/q)$, et elle est réversible.

5.3. Convergence à l'équilibre. Le premier intérêt de comprendre les mesures invariantes d'une chaîne de Markov est que cela permet de décrire un régime stationnaire, à l'équilibre, de cette chaîne. Mais il se passe quelque chose d'encore plus fort : en fait, la mesure de probabilité invariante (unique si la chaîne est irréductible et récurrente positive) décrit aussi le comportement en temps long de la chaîne de Markov, même lorsqu'elle part d'une autre mesure initiale. Nous terminons avec une série de résultats allant dans ce sens.

THÉORÈME 4.37 (Théorème ergodique). *Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov sur E homogène, irréductible et récurrente positive, d'unique mesure de probabilité invariante π (et de mesure initiale quelconque).*

1. *Alors, pour tout fonction $h : E \to \mathbb{R}$ π-intégrable, on a*
$$\frac{1}{n}\sum_{k=0}^{n-1} h(X_k) \xrightarrow[n\to+\infty]{p.s.} \int_E h d\pi = \sum_{x\in E} \pi(x)h(x).$$
2. *En particulier, pour tous $i, j \in E$,*
$$\frac{1}{n}\sum_{k=0}^{n-1} P^k(i,j) \xrightarrow[n\to+\infty]{} \pi(j),$$

*Ces résultats sont en particulier valables pour toute chaîne de Markov irréductible sur un espace d'états **fini**.*

Pour obtenir un résultat plus fort (sans la moyenne de Cesàro), il faut faire une hypothèse supplémentaire.

DÉFINITION 4.38. Soit P une matrice de transition sur E. Pour tout $i \in E$, on définit $R(i) = \{n \in \mathbb{N}^* : P^n(i,i) > 0\}$, et on appelle période de a le pgcd de $R(i)$. On dit que i est un état apériodique s'il est de période 1.

PROPOSITION 4.39. *Si la matrice de transition P est irréductible, alors tous les états sont de même période. On dit que P est apériodique dès que l'un (c.-à-d. tous) est (sont) apériodique(s). C'est en particulier le cas dès qu'il existe $i \in E$ avec $P(i,i) > 0$.*

PROPOSITION 4.40. *Si P est une matrice de transition sur un espace d'états E **fini**, alors P est irréductible apériodique si et seulement s'il existe $r \in \mathbb{N}^*$ tel que $P^r(i,j) > 0$ pour tous $i, j \in E$.*

THÉORÈME 4.41 (Théorème de convergence ; cas fini). *Soit E un ensemble **fini**, P une matrice de transition irréductible apériodique sur E et π l'unique mesure de probabilité invariante pour P. Alors, pour tous $i, j \in E$, $P^n(i,j)$ converge vers $\pi(j)$ exponentiellement vite quand n tend vers l'infini.*
Ainsi, si $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov de matrice de transition P et de mesure initiale μ quelconque, alors la loi de X_n converge exponentiellement vite vers la mesure invariante π lorsque n tend vers l'infini.

THÉORÈME 4.42 (Théorème de convergence ; cas général).

1. *Soit E un ensemble **dénombrable**, P une matrice de transition irréductible, apériodique et récurrente positive sur E et π l'unique mesure de probabilité invariante pour P. Alors, pour tous $i, j \in E$, $P^n(i,j)$ converge vers $\pi(j)$ quand n tend vers l'infini.*
2. *Si P est irréductible et récurrente nulle sur E, alors $P^n(i,j)$ tend vers 0 quand n tend vers l'infini pour tous $i, j \in E$.*

6. Exemples

6.1. Marche au hasard simple sur un graphe. On se donne un graphe fini, non orienté, connexe et simple (il y a au plus une arête entre deux sommets). On note toujours E l'ensemble de ses sommets, et le degré du sommet i, noté $\deg(i)$ est le nombre d'arêtes incidentes à i. On définit alors la marche au hasard simple sur ce graphe comme la chaîne de Markov de matrice de transition $P = (P(i,j))_{i,j\in E}$ donnée par

$$p(i,j) = \begin{cases} \frac{1}{\deg(i)} & \text{s'il y a une arête entre } i \text{ et } j\,; \\ 0 & \text{sinon} \end{cases}$$

(à chaque étape, la marche saute sur un sommet voisin de sa position actuelle, choisi uniformément).
En notant $C = \sum_{i\in E} \deg(i)$ (on peut vérifier que C est le double du nombre d'arêtes, pourquoi ?), on peut définir une mesure de probabilité π sur E par $\pi(i) = \frac{\deg(i)}{C}$. Vérifions que cette mesure est réversible pour P :

— si i et j ne sont pas voisins, alors $\pi(i)p_{i,j} = \pi(j)p_{j,i} = 0.$;

— si i et j sont voisins, alors

$$\pi(i)p_{i,j} = \frac{\deg(i)}{C}\frac{1}{\deg(i)} = \frac{1}{C} = \frac{\deg(j)}{C}\frac{1}{\deg(j)} = \pi(j)p_{j,i}\,.$$

π est bien une mesure réversible, donc invariante pour la marche au hasard simple sur le graphe (le cas symétrique, $p = \frac{1}{2}$, de la marche au hasard sur le cercle était un cas particulier).
On peut déduire des résultats généraux obtenus précédemment que le temps moyen passé par la marche au hasard en un sommet i est proportionnel à son degré, ou bien aussi que le temps moyen mis par la marche pour revenir en un sommet i est inversement proportionnel à son degré.

6.2. Algorithme de Metropolis-Hastings. Soit E un ensemble fini, H une fonction de E dans $\mathbb{R}$. On souhaite simuler une variable aléatoire de loi μ dont la densité par rapport à la mesure de comptage sur E est proportionnelle à e^{-H}.
L'algorithme de Metropolis-Hastings permet de construire et de simuler une chaîne de Markov récurrente apériodique $(X_n)_{n\in\mathbb{N}}$ sur E de loi invariante μ à partir des seuls rapports $\mu(x)/\mu(y)$.

Pour construire le noyau de transition P on se donne un noyau de transition irréductible Q sur E tel que $Q(x,y) > 0$ ssi $Q(y,x) > 0$. On pose

$$P(x,y) = \begin{cases} Q(x,y)\rho(x,y) & \text{si } x \neq y, \\ 1 - \sum_{z \neq x} P(x,z) & \text{si } x = y, \quad \text{où} \quad \rho(x,y) = \alpha\left(\frac{\mu(y)Q(y,x)}{\mu(x)Q(x,y)}\right)\mathbf{1}_{Q(x,y)>0}, \end{cases}$$

et où $\alpha : \mathbb{R} \to [0,1]$ est une fonction fixée qui vérifie $\alpha(u) = u\alpha(1/u)$ pour tout réel non nul. Par exemple, $\alpha = \min(1,u)$ ou encore $\alpha(u) = u/(1+u)$.

THÉORÈME 4.43. *Le noyau P est irréductible et récurrent positif, de loi invariante réversible μ, apériodique si $\alpha < 1$.*

7. Exercices

EXERCICE 4.1. Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov de loi initiale λ et de matrice de transition P. Soit $k \in \mathbb{N}^*$. Montrer que $(X_{kn})_{n\in\mathbb{N}}$ est une chaîne de Markov dont on précisera les paramètres.

EXERCICE 4.2. Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov sur $\{1,2,3\}$ de matrice de transition

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 2/3 & 1/3 \\ p & 1-p & 0 \end{pmatrix}$$

Calculer $\mathbb{P}(X_n = 1 | X_0 = 1)$ dans les cas suivants : $p = 1/16$, $p = 1/6$ et $p = 1/12$.

EXERCICE 4.3. Soit Rec (resp. Tra) l'ensemble des états récurrents (resp. transients) d'une chaîne de Markov finie et homogène. Soit T le temps d'atteinte de l'ensemble Rec. Montrer que pour tout état $i \in$ Tra, on a

$$\mathbb{E}_i(T) = 1 + \sum_{j \in \text{Tra}} P(i,j)\mathbb{E}_j(T).$$

EXERCICE 4.4. Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov sur $\{1,2,3,4\}$ de matrice de transition

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

1. Classer les états de la chaîne.
2. Quelle est la probabilité que la chaîne issue de 2 soit absorbée en 4 ?
3. Quel est le temps moyen d'absorption de la chaîne issue de 2 ?
4. Quelles sont les mesures invariantes de la chaîne ?

EXERCICE 4.5. On considère la matrice de transition suivante :

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/3 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$

Représenter le graphe de la chaîne. Quelles sont les classes de la chaîne, les mesures invariantes ? Peut-on parler de chaîne périodique, apériodique ?

EXERCICE 4.6. Quelle est la mesure invariante de la chaîne de transition

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}?$$

EXERCICE 4.7. Soit $p \in]0,1[$, $q = 1-p$ et $(X_n)_{n\in\mathbb{N}}$ la chaîne de Markov sur $\{0,1,2,\dots,M\}$ de matrice de transition définie par

$$p_{00} = q, \quad p_{01} = p, \quad \forall i = 1,\dots,M-1, \quad p_{i,i-1} = q, \quad p_{i,i+1} = p, \quad p_{M,M-1} = q, \quad p_{M,M} = p.$$

Montrer que la chaîne admet une mesure réversible que l'on explicitera.

EXERCICE 4.8. Sur l'ensemble fini $E = \mathbb{Z}/m\mathbb{Z}$, on considère la chaîne de Markov $(X_n)_{n\in\mathbb{N}}$ de transition P définie par

$$P_{i,i+k} = P_{i,i-k} = 1/2, \quad P_{i,j} = 0 \text{ sinon pour } k \in \{1,\dots,m-1\}.$$

Donner, dans tous les cas, ses classes de récurrence et la mesure invariante de ces classes. Lorsque la chaîne est récurrente irréductible, déterminer quand elle est apériodique.

EXERCICE 4.9. On considère la chaîne d'Ehrenfest décrite dans l'exemple 4.4.
1. Grâce à la section 2, calculer l'espérance de X_n sachant que $X_0 = x$.
2. Montrer que la loi binomiale $\mathscr{B}(d, 1/2)$ (que l'on notera π) est invariante (penser à la réversibilité...). Est-elle unique ?
3. On suppose que la chaîne est à l'état stationnaire. Comparer $\pi(d/2)$ et $\pi(0)$. Commenter...
4. Lorsque d est grand, estimer que la probabilité d'observer la chaîne dans un état n'appartenant pas à l'intervalle $[(d-5\sqrt{d})/2, (d+5\sqrt{d})/2]$.
5. Déterminer la période de la chaîne X. Soit $F = \{0,1,\dots,d\} \cap (2\mathbb{N})$ et $z \in F$. Montrer que la suite Y^z définie par $Y_n^z = X_{2n}^z$ est une chaîne de Markov sur F dont on exprimera la mesure invariante en fonction de celle de X. Si X_0 est à valeurs dans E quel est le comportement en temps en long de la loi de X_n ? Les résultats de cette question admettent-ils une généralisation ?
6. Montrer que la chaîne d'Ehrenfest est l'image de la marche aléatoire sur le cube par une fonction que l'on précisera. Soit E et F deux ensembles finis, $f : E \to F$ et X une chaîne de Markov sur E. La suite Y définie par $Y_n = f(X_n)$ est-elle, en général, une chaîne de Markov ?

EXERCICE 4.10 (Simulation du modèle d'Ehrenfest). Sauf cas contraire, on prendra $N = 10$.
1. Écrire une fonction qui trace une trajectoire de cette chaîne. On pourra utiliser la fonction grand avec l'option "markov" après avoir créer la matrice de transition P avec la commande :

$$\mathrm{P} = \mathrm{diag}([1 : -1/\mathrm{N} : 1/\mathrm{N}], 1) + \mathrm{diag}([1/\mathrm{N} : 1/\mathrm{N} : 1], -1)$$

2. Illustrer le fait que, pour $l \in \{0,\dots,m\}$,

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i = l\}} \xrightarrow[n\to\infty]{p.s.} \mu(\{l\}).$$

Quel théorème est-il mis en lumière ici ?
3. Si T_l désigne le temps d'atteinte du point l, comparer, par simulation $\mathbb{E}_l(T_l)$ et $1/\mu(l)$.
4. Comparer les grandes puissances paires et impaires de P. La loi de X_n converge-t-elle vers μ ? Pourquoi ?
5. Comment simuler les trajectoires d'une chaîne d'Ehrenfest lorsque le nombre de particules atteint 1000000 ?

EXERCICE 4.11 (File d'attente M/M/1 (à temps discret)). On note P la matrice de transition et $X = (X_n)_{n\in\mathbb{N}}$ la chaîne de Markov correspondant à la file d'attente M/M/1 décrite dans l'exemple 4.6.
1. Montrer que P est irréductible et apériodique.
2. Montrer que X est

— récurrente positive si $p < 1/2$;

— récurrente nulle si $p = 1/2$;

— transiente si $p > 1/2$.

On pourra remarquer que si $\tau_0 = \inf\{n \in \mathbb{N}^* : X_n = 0\}$ et $T_M = \inf\{n \in \mathbb{N}^* : X_n \in \{0, M\}\}$, alors τ_0 est la limite croissante des T_M lorsque M tend vers l'infini, et utiliser les résultats sur la ruine du joueur (exercice 3.6).
3. Montrer que, pour $p < 1/2$, la loi π définie par $\pi(i) = (p/q)^i(1 - p/q)$ est réversible, donc invariante pour X.
4. Montrer que, pour $p = 1/2$, la mesure de comptage sur $\mathbb{N}$ est réversible, donc invariante pour X.

EXERCICE 4.12. On considère un cavalier fou sur un échiquier : à chaque coup, il se déplace de manière équiprobable sur toutes les cases de l'échiquier qu'il a le droit d'atteindre. On note X_n sa position après n déplacements. Montrer que $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov irréductible récurrente. Quelle est sa mesure invariante ?

EXERCICE 4.13. Soit $(V, \mathscr{E})$ un graphe connexe non orienté d'ensemble de sommets fini V et d'ensemble d'arêtes $\mathscr{E} \in V \times V$. On associe à chaque arête (i, j) un poids $w_{ij} = w_{ji} > 0$ et l'on pose $w_i = \sum_j w_{ij}$. Déterminer la mesure invariante de la chaîne de Markov sur V de matrice de transition P définie par $P(i, j) = w_{ij}/w_i$.

EXERCICE 4.14. On considère la chaîne de Markov $(X_n)_{n\geq 0}$ sur l'espace d'états $\mathbb{N}$ définie par la transition suivante :

$$\mathscr{L}(X_{n+1}|X_n = k) = \mathscr{P}(\alpha) * \mathscr{B}(k, p).$$

1. Déterminer la transformée de Laplace de $\mathscr{L}(X_{n+1}|X_n = k)$.
2. Montrer que si X_0 suit une loi de Poisson, il en est de même pour X_n pour tout $n \geq 1$.
3. Déterminer la mesure de probabilité invariante.

EXERCICE 4.15 (Pagerank idéalisé). Voici comment fonctionne une version simplifiée de l'algorithme de classement utilisé par Google. Soit G le graphe (fini) dont les sommets $s_1, \ldots, s_N$ sont les pages du web et dont les arêtes (orientées) sont les hyperliens pointant d'une page vers une autre. On ajoute un sommet s_0 au graphe G ainsi qu'une arête de chaque sommet de G vers s_0. On note $\hat{G}$ le graphe obtenu et A sa matrice d'adjacence définie par $A_{ij} = 1$ s'il existe un lien de la page i vers la page j et $A_{ij} = 0$ sinon. Soit $N(i)$ le nombre d'arêtes sortant de s_i. Étant donné un paramètre $p \in]0, 1[$, on définit une matrice Q sur $\hat{G}$ de la façon suivante :

— pour $0 \leq i \leq N$, $Q(s_0, s_i) = 1/(N + 1)$,

— pour $1 \leq i \neq j \leq N$, $Q(s_i, s_i) = 0$ et

$$Q(s_i, s_0) = \begin{cases} 1 & \text{si } N(i) = 1, \\ 1 - p & \text{si } N(i) \neq 1, \end{cases} \quad \text{et} \quad Q(s_i, s_j) = \begin{cases} 0 & \text{si } A_{ij} = 0, \\ p/(N(i) - 1) & \text{si } A_{ij} = 1. \end{cases}$$

1. Écrire les matrices A et Q correspondant au mini-web suivant

$$1 \to 2 \quad 1 \to 3 \quad 3 \to 2 \quad 2 \to 3 \quad 4 \to 5 \quad 5 \to 4$$

avec $p = 1/2$.
2. On se place à présent dans le cas général. Montrer que la matrice Q est une matrice stochastique apériodique irréductible.
3. Interpréter l'évolution de la chaîne de Markov associée à Q.
4. Montrer que cette chaîne admet une unique probabilité invariante π. L'algorithme Pagerank interprète $0 \leq \pi(i) \leq 1$ comme la popularité de la page i et ordonne les pages web à partir de cette donnée.
5. Vous semble-t-il raisonnable de calculer explicitement la probabilité invariante ? Comment faire sinon ?

EXERCICE 4.16 (Simulation d'une chaîne simple). Les consommateurs de 3 produits sont initialement répartis respectivement à 50% pour P1, 30% pour P2 et 20% pour P3. Après chaque mois, 60% restent fidèles à P1 contre 70% pour P2 et 90% pour P3. Les autres se réorientent entre les deux autres produits (de manière équiprobable).
1. Déterminer la distribution initiale ν et la matrice de transition P associées à ce modèle markovien (exemple : $P_{12} = 0.2$).
2. Tracer l'évolution déterministe des répartitions des consommateurs pendant les 12 premiers mois.
3. Tracer l'évolution de l'opinion d'un individu initialement adepte de P1 pendant les douze premiers mois(consultez l'aide pour grand option 'markov').
4. Même question pour un individu pris au hasard dans la population totale selon la loi ν.
5. Simuler la répartition sur douze mois des opinions de 1000 personnes (qui ne se concertent pas) choisies indépendamment dans la population.
6. Déterminer la distribution stationnaire de la matrice de transition. Calculer P^{100}. Commenter. Refaire la question 2 en prenant la distribution stationnaire comme distribution initiale des consommateurs. Quelles sont les valeurs propres de P ? On pourra utiliser les fonctions spec, bdiag, mais aussi de grandes puissances de P.

EXERCICE 4.17 (Chaîne de Wright-Fisher avec mutation). On adapte ici le modèle de Wright-Fisher décrit dans l'exemple 4.5 pour prendre en compte des mutations. À chaque génération, un allèle A (resp. B) mute en B (resp. A) avec une probabilité u (resp. B).
1. Montrer que la matrice de transition P associée à ce modèle est telle que

$$P(i,j) = \binom{N}{j} q_i^j (1-q_i)^{N-j} \quad \text{avec} \quad q_i = \frac{i}{N}(1-u) + \left(1 - \frac{i}{N}\right) v$$

pour $i, j \in \{0, 1, \ldots, N\}$.
2. Que dire de cette chaîne (statut des points, mesure(s) invariante(s), périodicité, etc) ?
3. Grâce à Scilab, calculer la mesure invariante π pour $N = 10$ et 20. On pourra créer la matrice P grâce à la fonction binomial. À partir de quelle valeur de N, le logiciel ne permet plus le calcul de π ?
4. Proposer une méthode s'appuyant sur le théorème ergodique pour donner une estimation des coefficients de π. Confronter les résultats de la simulation à ceux de la question précédente pour $N = 10$ et 20.

8. Corrigés

CORRIGÉ DE L'EXERCICE 4.1. La mesure initiale est la même et la matrice de transition est donnée par P^k. On parle de chaîne échantillonée.

CORRIGÉ DE L'EXERCICE 4.3. C'est une application de la propriété de Markov sur un pas. Partant de $i \in \mathrm{Tra}$, on atteint un point récurrent (et T vaut 1) ou un point transient j (avec

probabilité $P(i,j)$) et T vaut alors 1 plus le temps d'atteindre les points récurrents partant de j (on oublie d'où l'on vient).

CORRIGÉ DE L'EXERCICE 4.5. Les points $\{1,2,3,4\}$ sont transients et la chaîne possède une classe de récurrence $\{5,6\}$. La chaîne restreinte à cet ensemble est apériodique puisque $P(5,5) > 0$ et la seule mesure de probabilité invariante est la mesure uniforme sur $\{5,6\}$.

CORRIGÉ DE L'EXERCICE 4.7. On écrit la relation d'équilibre détaillé (22) pour $i \le M-1$ et $j = i+1$:

$$\pi(i)p = \pi(i+1)q.$$

On en déduit que $\pi(i) = (p/q)^i \pi(0)$. En imposant que π soit une mesure de probabilité, on obtient, pour $i = 0,1,\dots,M$,

$$\pi(i) = \frac{1-p/q}{1-(p/q)^{M+1}}\left(\frac{p}{q}\right)^i.$$

Remarquons que si $p < q$ et que M tend vers l'infini, la mesure π converge vers la loi géométrique sur $\mathbb{N}$ et on retrouve la chaîne étudiée dans l'exemple 4.6.

CORRIGÉ DE L'EXERCICE 4.9.
1. Notons f la fonction sur E donnée par $f(x) = x - N/2$ assimilée à un vecteur colonne. La fonction f est un vecteur propre (à droite) associé à la valeur propre $1-2/N$. On a donc

$$P^n f(i) = \left(1-\frac{2}{N}\right)^n f(i) \quad \text{ou} \quad \mathbb{E}(X_n) = \left(1-\frac{2}{N}\right)^n \left(i - \frac{N}{2}\right) + \frac{N}{2}.$$

On peut aussi remarquer que

$$\begin{aligned}\mathbb{E}(X_{n+1}|X_n) &= \left(1-\frac{X_n}{N}\right)(X_n+1) + \left(1+\frac{X_n}{N}\right)(X_n-1)\\ &= \left(1-\frac{2}{N}\right)X_n + 1,\end{aligned}$$

puis on raisonne par récurrence.
2. On vérifie que la loi binomiale $\mathscr{B}(d,1/2)$ est P-symétrique. Elle est unique puisque la chaîne est irréductible.
3. La formule de Stirling assure que $\pi(d/2) \sim 1/\sqrt{\pi N/2}$ et $\pi = 2^{-N}$. Les états autour de $N/2$ sont beaucoup plus probables que ceux situés sur les bords.
4. On utilise l'approximation de la loi binomiale par la loi gaussienne.
5. La chaîne est périodique de période 2 (la parité de X change à chaque pas et $P^2(x,x) > 0$). La suite Y^z est une chaîne de Markov sur F de matrice de transition P^2 (restreinte à F) et sa mesure de probabilité invariante est la loi binomiale restreinte à F et normalisée pour être de masse 1. Cette chaîne est apériodique puisque $P^2(x,x) > 0$. Ainsi, la loi de Y_n converge la mesure invariante. Ce résultat se généralise à une période k quelconque (il y a alors k sous-chaînes).
6. La chaîne d'Ehrenfest est l'image de la marche aléatoire sur le cube $0,1^N$ par la fonction $x \in 0,1^N \mapsto x(1) + \cdots + x(N)$. Ce fait est remarquable car en général, l'image d'une chaîne de Markov n'en est pas une. Par exemple, si $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov sur $\{1,2,3\}$ de matrice de transition

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \end{pmatrix}$$

et $f = \mathbf{1}_{\{2,3\}}$ alors $(Y_n)_{n\in\mathbb{N}}$ n'est pas markovienne.

CORRIGÉ DE L'EXERCICE 4.11.

1. Pour tous $i, j \in \mathbb{N}$, $P^{|i-j|}(i,j) > 0$ (on peut suivre avec probabilité strictement positive un chemin se déplaçant au plus proche voisin de i à j), donc P est irréductible. Elle est apériodique car $P(0,0) > 0$.
2. Avec les notations de l'énoncé, on a $T_M = \inf(\tau_0, \tau_M)$, et comme la chaîne ne fait que des pas de 1, la suite $\tau_M = \inf\{n \geq 0 : X_n = M\} \geq (M - X_0)_+$ tend p.s. vers l'infini en croissant quand M tend vers l'infini. On en déduit que

$$\mathbb{P}_1(\tau_0 < +\infty) = \lim_{M \to +\infty} \mathbb{P}_1(\tau_0 < \tau_M) = \lim_{M \to +\infty} \mathbb{P}_1(X_{T_M} = 0).$$

Le comportement asymptotique des probabilités de ruine se déduit facilement des formules explicites de l'exercice 3.6 pour $k = 1$, d'où

$$\mathbb{P}_1(\tau_0 < +\infty) \begin{cases} = 1 & \text{si } p \leq 1/2\,; \\ = q/p < 1 & \text{si } p > 1/2. \end{cases}$$

On obtient en appliquant la propriété de Markov au temps 1 que $\mathbb{P}_0(\tau_0 < +\infty) = q + p \cdot \mathbb{P}_1(\tau_0 < +\infty)$, donc que

$$\mathbb{P}_1(\tau_0 < +\infty) \begin{cases} = 1 & \text{si } p \leq 1/2\,; \\ = 2q < 1 & \text{si } p > 1/2. \end{cases}$$

Ceci prouve que la chaîne de Markov X est récurrente pour $p \leq 1/2$ et transiente sinon.
De la même manière, la suite T_M converge en croissant vers τ_0 quand M tend vers l'infini, donc, toujours avec les résultats explicites de l'exercice 3.6 pour $k = 1$,

$$\mathbb{E}_1(\tau_0) = \lim_{M \to +\infty} \mathbb{E}_1(T_M) = \begin{cases} +\infty & \text{si } p \geq 1/2\,; \\ 1/(q-p) & \text{si } p < 1/2. \end{cases}$$

Comme $\mathbb{E}_0(\tau_0) = q + p(1 + \mathbb{E}_1(\tau_0))$, on en déduit

$$\mathbb{E}_0(\tau_0) = \begin{cases} +\infty & \text{si } p \geq 1/2\,; \\ q/(q-p) & \text{si } p < 1/2. \end{cases}$$

On en déduit que X est récurrente nulle pour $p = 1/2$ et récurrente positive pour $p < 1/2$.
3. Il suffit de vérifier que $\pi(i)P(i, i+1) = \pi(i+1)P(i+1, i)$ pour tout $i \in \mathbb{N}$, ce qui est immédiat.
Notons que ce résultat est cohérent avec la dernière formule de la question précédente puisqu'on obtient bien que $\pi(0) = 1 - p/q = 1/\mathbb{E}_0(\tau_0)$.
4. C'est une conséquence immédiate du fait que $P(i, i+1) = P(i+1, i) = 1/2$ pour tout $i \in \mathbb{N}$.

CORRIGÉ DE L'EXERCICE 4.12. Un cavalier peut atteindre toutes les cases de l'échiquier donc la chaîne est irréductible. Elle est récurrente puisque l'espace d'états est fini. Elle est périodique de période 2 puisque la couleur de la case sur laquelle est le cavalier change à chaque déplacement et il peut revenir sur place en deux coups. Pour la mesure invariant, c'est un cas particulier de l'exercice 4.13.

CORRIGÉ DE L'EXERCICE 4.13. La chaîne est irréductible sur un espace d'états fini. Elle admet donc une unique mesure de probabilité invariante. On remarque facilement que la mesure la mesure $\tilde{\pi}$ donnée par $\tilde{\pi}(i) = w_i$ est P-symétrique (mais pas nécessairement de masse 1). L'unique mesure de probabilité invariante est donc

$$\pi(i) = \frac{w_i}{\sum_{j \in E} w_j}.$$

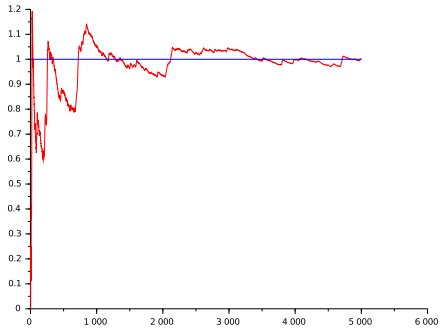

FIGURE 3. Illustration du théorème 4.37 ergodique pour la file d'attente M/M/1, avec $p = 1/3$: en rouge, la fonction qui à n associe la moyenne ergodique des n premiers termes le long d'une trajectoire ; en bleu la valeur moyenne de π, $p/(q-p)$.

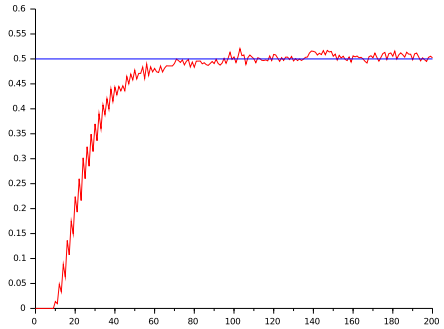

FIGURE 4. Illustration du théorème 4.42 de convergence pour la file d'attente M/M/1, avec $p = 1/3$: en rouge, la fonction qui à n associe la valeur estimée (par méthode de Monte-Carlo avec 5000 tirages) de la probabilité $P^n(10,0)$; en bleu la valeur $\pi(0) = 1 - p/q = 1/2$.

CORRIGÉ DE L'EXERCICE 4.14. On considère la chaîne de Markov $(X_n)_{n\geq 0}$ sur l'espace d'états $\mathbb{N}$ définie par la transition suivante :

$$\mathscr{L}(X_{n+1}|X_n = k) = \mathscr{P}(\alpha) * \mathscr{B}(k,p).$$

1. Notons L_k la transformée de Laplace de $\mathscr{L}(X_{n+1}|X_n = k)$. Par définition du produit de convolution et l'expression des transformées de Laplace classique,

$$L_k(t) = e^{\lambda(e^t-1)}(pe^t + 1 - p)^k.$$

2. Supposons que la loi de X_0 soit la loi de Poisson de paramètre α. Alors,

$$\begin{aligned}
\mathbb{E}\big(e^{tX_1}\big) &= \mathbb{E}\left(\sum_{k\geq 0} e^{tX_1}\mathbf{1}_{\{X_0=k\}}\right) = \sum_{k\geq 0}\mathbb{E}\big(e^{tX_1}\mathbf{1}_{\{X_0=k\}}\big)\\
&= \sum_{k\geq 0} L_k(t)\mathbb{P}(X_0 = k) = e^{\lambda(e^t-1)}e^{-\alpha}\sum_{k\geq 0}(pe^t+1-p)^k\frac{\alpha^k}{k!}\\
&= \exp\big((\lambda+\alpha p)(e^t-1)\big).
\end{aligned}$$

Ainsi, X_1 suit la loi $\mathscr{P}(\lambda + \alpha p)$. On en déduit par récurrence qu'il en est de même pour X_n pour tout $n \geq 1$.
3. Si $\alpha = \lambda/(1-p)$ alors X_0 et X_1 ont même loi (de Poisson) et c'est encore le cas pour tout $n \in \mathbb{N}$. L'unique (la chaîne est irréductible) mesure de probabilité invariante est la loi de Poisson de paramètre $\lambda/(1-p)$.

CORRIGÉ DE L'EXERCICE 4.15.
1. Les matrices A et Q sont des éléments de $\mathscr{M}_6(\mathbb{R})$ et

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad \text{et} \quad Q = \begin{pmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/2 & 0 & 1/4 & 1/4 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 0 & 1/2 & 0 \end{pmatrix}$$

2. Tous les coefficients de Q sont clairement positifs. De plus,

$$\sum_{j=0}^{N} Q(s_0, s_j) = \sum_{i=0}^{N} \frac{1}{N+1} = 1,$$

et pour $1 \leq i \leq N$,

$$\sum_{j=0}^{N} Q(s_i, s_j) = 1 - p + \frac{p}{N(i)-1} \sum_{j=1}^{N} A_{ij} = 1,$$

car $1 + \sum_{j=1}^{N} A_{ij}$ arêtes partent du sommet s_i. La matrice Q est donc stochastique. Elle est irréductible car tous les sommets mènent à s_0 et réciproquement. Enfin, elle est apériodique puisque $Q(s_0, s_0) > 0$.
3. Partant de s_0, la chaîne choisit un sommet au hasard (y compris s_0) selon la loi uniforme. Partant de s_i pour $1 \leq i \leq N$, si la page s_i ne contient aucun lien, la chaîne saute en s_0. Sinon, elle saute en s_0 avec probabilité $1-p$ et sinon elle choisit un des sommets reliés par une arête issue de s_i selon la loi uniforme.
4. La chaîne est irréductible à espace d'états fini. Elle admet donc une unique mesure de probabilité invariante.
5. En pratique, la matrice associée au vrai web est beaucoup trop grosse (et sans cesse modifiée). Google estime π par le théorème ergodique en lançant des robots qui évoluent grâce à la matrice Q.

Chapitre 5

Statistiques d'ordre, processus de Poisson

1. Lois exponentielles

Cette section regroupe quelques propriétés importantes des lois exponentielles.

PROPOSITION 5.1 (Absence de mémoire). *Une variable aléatoire réelle T à densité suit une loi exponentielle si et seulement si elle vérifie la propriété d'absence de mémoire :*

$$\forall s, t \in \mathbb{R}_+, \quad \mathbb{P}(T > t + s | T > s) = \mathbb{P}(T > t).$$

DÉMONSTRATION. En posant $\varphi(t) = \mathbb{P}(T > t)$, on obtient que, pour tous $s, t \in \mathbb{R}_+$, $\varphi(t+s) = \varphi(t)\varphi(s)$. On utilise ensuite que $\varphi(0) = 1$ et que φ est décroissante et continue pour conclure. □

On peut renforcer cette propriété d'absence de mémoire comme suit.

PROPOSITION 5.2. *Soit T et S deux variables aléatoires indépendantes de lois exponentielles. Alors, pour tout $t \in \mathbb{R}_+$,*

$$\mathbb{P}(T > t + S | T > S) = \mathbb{P}(T > t).$$

DÉMONSTRATION. Notons λ et μ les paramètres respectifs des lois de S et T. On écrit, grâce au théorème de Fubini,

$$\begin{aligned}
\mathbb{P}(T > t + S) &= \mathbb{E}\big(\mathbf{1}_{\{T > t+S\}}\big) = \mathbb{E}\big(e^{-\mu(t+S)}\big) \\
&= \int_{\mathbb{R}_+} e^{-\mu(t+s)} \lambda e^{-\lambda s}\, ds \\
&= \frac{\mu}{\lambda + \mu} e^{-\mu t} = \mathbb{P}(T > S)\mathbb{P}(T > t),
\end{aligned}$$

où l'expression de $\mathbb{P}(T > S)$ est obtenue en prenant $t = 0$. □

De nombreux phénomènes présentent cette propriété d'absence de mémoire :

- les puces électroniques ne s'usent (quasiment) pas (elles tombent en panne suite à un problème extérieur),
- un atome radioactif se décompose de manière spontanée sans lien avec son âge,
- certaines tortues, quelques poissons comme l'esturgeon ou certains mollusques bivalves ont une mortalité constante au cours du temps (hors prédation, contagion, accident etc).

PROPOSITION 5.3. *Soit $(T_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes de lois exponentielles de paramètres respectifs $(\lambda_n)_{n\in\mathbb{N}^*}$ strictement positifs.*

$$si \quad \sum_{n\in\mathbb{N}^*} \frac{1}{\lambda_n} < \infty \quad alors \quad \mathbb{P}\left(\sum_{n\in\mathbb{N}^*} T_n < \infty\right) = 1$$

et

$$\textit{si} \quad \sum_{n\geq 1} \frac{1}{\lambda_n} = \infty \quad \textit{alors} \quad \mathbb{P}\Big(\sum_{n\geq 1} T_n = \infty\Big) = 1.$$

DÉMONSTRATION. Si $\sum_{n\in\mathbb{N}^*} 1/\lambda_n$ est fini alors, par convergence monotone, la variable aléatoire $\sum_{n\in\mathbb{N}^*} T_n$ est intégrable donc elle est finie p.s. Notons que l'indépendance n'est pas utile pour cette partie de la proposition.
Pour tout $N \in \mathbb{N}^*$, par indépendance,

$$\mathbb{E}\left(\exp\left\{-\sum_{n=1}^{N} T_n\right\}\right) = \prod_{n=1}^{N} \mathbb{E}\big(\exp\{-T_n\}\big) = \prod_{n=1}^{N} (1+1/\lambda_n)^{-1}.$$

Si $\sum_{n\in\mathbb{N}^*} 1/\lambda_n$ est infini alors $\prod_{n\in\mathbb{N}^*}(1+1/\lambda_n)$ est également infini. Par convergence monotone on conclut que la série de terme général T_n diverge avec probabilité 1. □

PROPOSITION 5.4. *Soit I un ensemble (au plus) dénombrable et $(T_k)_{k\in I}$ des variables aléatoires exponentielles de paramètres respectifs $(\lambda_k)_{k\in I}$ vérifiant $\lambda = \sum_k \lambda_k < \infty$. Soit $T = \inf_k T_k$. Alors, avec probabilité 1, l'infimum est atteint en un unique K (aléatoire) élément de I. De plus, T et K sont indépendantes, T suit la loi $\mathscr{E}(\lambda)$ et pour tout $k \in I$, $\mathbb{P}(K = k) = \lambda_k/\lambda$.*

DÉMONSTRATION. Soit $k \in I$ et $t \in \mathbb{R}_+$.

$$\begin{aligned}
\mathbb{P}(K = k, T \geq t) &= \mathbb{P}(T_k \geq t, T_j > T_k, j \neq k) \\
&= \int_t^\infty \lambda_k e^{-\lambda_k s} \mathbb{P}(T_j > s, j \neq k)\, ds \\
&= \int_t^\infty \lambda_k e^{-\lambda_k s} \prod_{j\neq k} e^{-\lambda_j s}\, ds \\
&= \int_t^\infty \lambda_k e^{-\lambda s}\, ds = \frac{\lambda_k}{\lambda} e^{-\lambda t}.
\end{aligned}$$

On conclut en remarquant que les ensembles de la forme $\{K = k\} \times \{T \geq t\}$ avec $k \in I$ et $t \in \mathbb{R}_+$ engendrent la tribu produit. □

EXERCICE 5.1 (Temps de demi-vie). Considérons un fragment de métal contenant un grand nombre N (que l'on supposera pair) d'atomes radioactifs et supposons que chaque atome se désintègre après un temps exponentiel de paramètre α petit indépendamment des autres. Notons $(T_n^N)_{1\leq n\leq N}$ les instants où un atome se désintègre et $T_0^N = 0$. On pose aussi $\tau_n^N = T_n^N - T_{n-1}^N$ pour $n \in \mathbb{N}^*$. On s'intéresse au temps $T_{N/2}^N$, appelé temps de demi-vie. C'est l'instant où la moitié des atomes initialement présents se sont désintégrés.
1. Montrer que $T_1^N = \tau_1^N$ suit la loi exponentielle de paramètre αN.
2. Montrer plus généralement que les variables $(\tau_n^N)_{1\leq n\leq N}$ sont indépendantes et que τ_n^N suit la loi exponentielle de paramètre $\alpha(N-n+1)$.
3. Montrer que

$$\lim_{N\to\infty} \mathbb{E}(T_{N/2}^N) = \frac{\ln(2)}{\alpha}.$$

4. Si X suit la loi exponentielle de paramètre 1, calculer

$$\mathbb{E}\big[(X - \mathbb{E}(X))^2\big] \quad \text{et} \quad \mathbb{E}\big[(X - \mathbb{E}(X))^4\big].$$

5. En déduire que $\mathbb{E}\Big[\big(T_{N/2}^N - \mathbb{E}(T_{N/2}^N)\big)^4\Big] = \mathscr{O}(N^{-2})$.
6. Déterminer la limite de $T_{N/2}^N$ quand $N \to \infty$ en précisant la nature de la convergence.

2. Statistiques d'ordre

On considère dans cette section un échantillon $X_1, X_2, \dots, X_n$ de variables aléatoires réelles indépendantes et de même loi μ de fonction de répartition F continue et de densité f. Remarquons tout d'abord que

$$\mathbb{P}(\exists i \neq j : X_i = X_j) \leq \frac{n(n-1)}{2}\mathbb{P}(X_1 = X_2) = \frac{n(n-1)}{2}\int_{\mathbb{R}^2} \mathbf{1}_{\{x=y\}} f(x)f(y)\,dxdy = 0.$$

Ainsi, pour presque tout ω, il existe une unique permutation (aléatoire mais on omet la dépendance en ω) σ_n telle que

$$X_{\sigma_n(1)} < X_{\sigma_n(2)} < \cdots < X_{\sigma_n(n)}.$$

Dans la suite, on notera l'échantillon ordonné $X_{(1,n)} < X_{(2,n)} < \cdots < X_{(n,n)}$, on parle également de statistiques d'ordre. On a en particulier

$$X_{(1,n)} = \min_{1\leq i\leq n} X_i \quad \text{et} \quad X_{(n,n)} = \max_{1\leq i\leq n} X_i.$$

PROPOSITION 5.5. *La loi de σ_n est la loi uniforme sur $\mathscr{S}_n$ l'ensemble des permutations de $\{1,2,\dots,n\}$. La statistique d'ordre $(X_{(1,n)},\dots,X_{(n,n)})$ admet la densité*

$$(x_1,\dots,x_n) \longmapsto n!f(x_1)\cdots f(x_n)\mathbf{1}_{\{x_1<\cdots<x_n\}}.$$

De plus, la permutation σ_n est indépendante de la statistique d'ordre.

DÉMONSTRATION. Soit $\sigma \in \mathscr{S}_n$. On a

$$\begin{aligned}\mathbb{P}(\sigma_n = \sigma) &= \mathbb{P}(X_{\sigma(1)} < \cdots < X_{\sigma(n)})\\ &= \int_{\mathbb{R}^n} \mathbf{1}_{\{x_{\sigma(1)}<\cdots<x_{\sigma(n)}\}} f(x_1)\cdots f(x_n)\,dx_1\dots dx_n\\ &= \int_{\mathbb{R}^n} \mathbf{1}_{\{x_1<\cdots<x_n\}} f(x_1)\cdots f(x_n)\,dx_1\dots dx_n,\end{aligned}$$

par symétrie. Ainsi, la loi de σ_n est uniforme sur $\mathscr{S}_n$: $\mathbb{P}(\sigma_n = \sigma) = 1/n!$ pour tout $\sigma \in \mathscr{S}_n$. Soit g une fonction borélienne bornée sur $\mathbb{R}^n$:

$$\begin{aligned}&\mathbb{E}\big(\mathbf{1}_{\{\sigma_n=\sigma\}} g(X_{(1,n)},\dots,X_{(n,n)})\big)\\ &\quad= \int_{\mathbb{R}^n} \mathbf{1}_{\{x_{\sigma(1)}<\cdots<x_{\sigma(n)}\}} g(\sigma(x_1),\dots,\sigma(x_n)) f(x_1)\dots f(x_n)\,dx_1\dots dx_n,\\ &\quad= \int_{\mathbb{R}^n} \mathbf{1}_{\{x_1<\cdots<x_n\}} g(x_1,\dots,x_n) f(x_1)\dots f(x_n)\,dx_1\dots dx_n.\end{aligned}$$

En particulier,

$$\mathbb{E}\big(\mathbf{1}_{\{\sigma_n=\sigma\}} g(X_{(1,n)},\dots,X_{(n,n)})\big) = \frac{1}{n!}\int_{\mathbb{R}^n} n!\mathbf{1}_{\{x_1<\cdots<x_n\}} g(x_1,\dots,x_n) f(x_1)\dots f(x_n)\,dx_1\dots dx_n.$$

On a donc

$$\mathbb{E}\big(\mathbf{1}_{\{\sigma_n=\sigma\}} g(X_{(1,n)},\dots,X_{(n,n)})\big) = \mathbb{P}(\sigma_n = \sigma)\mathbb{E}\big(g(X_{(1,n)},\dots,X_{(n,n)})\big),$$

ce qui caractérise l'indépendance de σ_n et de la statistique d'ordre. □

PROPOSITION 5.6. *Pour tout $x \in \mathbb{R}$ et $1 \leq k \leq n$,*

$$\begin{aligned}\mathbb{P}(X_{(k,n)} \leq x) &= \sum_{r=k}^{n} \binom{n}{r} F(x)^r (1-F(x))^{n-r}\\ &= \frac{n!}{(k-1)!(n-k)!}\int_0^{F(x)} t^{k-1}(1-t)^{n-k}\,dt.\end{aligned}$$

DÉMONSTRATION. Soit $x \in \mathbb{R}$. Considérons la variable aléatoire $S_n(x) = \sum_{i=1}^n \mathbf{1}_{\{X_i \le x\}}$ de loi $\mathscr{B}(n, F(x))$. Le point-clé est le suivant $\{S_n(x) \ge k\} = \left\{X_{(k,n)} \le x\right\}$. On a donc

$$\mathbb{P}(X_{(k,n)} \le x) = \mathbb{P}(S_n(x) \ge k) = \sum_{r=k}^{n} \binom{n}{r} F(x)^r (1 - F(x))^{n-r}.$$

La seconde expression découle de l'expression de la loi de $(X_{(1,n)}, \dots, X_{(n,n)})$ donnée dans la proposition 5.5. □

3. Processus de Poisson

3.1. Définition et premières propriétés.

DÉFINITION 5.7. Soit $(\tau_i)_{i\ge1}$ une suite de variables aléatoires réelles positives. On définit la suite croissante $(T_n)_{n\ge1}$ par $T_n = \tau_1 + \cdots + \tau_n$, $n \ge 1$. Le processus à temps continu $(N_t)_{t\in\mathbb{R}_+}$, défini par

$$N_t = \sum_{n\ge1} \mathbf{1}_{\{T_n \le t\}},$$

est appelé processus de comptage.
Si de plus les variables aléatoires $(\tau_i)_{i\in\mathbb{N}^*}$ sont indépendantes et de loi exponentielle de paramètre $\lambda \in \mathbb{R}_+^*$, on dit que le processus $(N_t)_{t\in\mathbb{R}_+}$ est un processus de Poisson d'intensité λ.

PROPOSITION 5.8. *Soit $n \ge 1$ et $\tau_1, \dots, \tau_n$ des variables aléatoires indépendantes de loi exponentielle de paramètre $\lambda \in \mathbb{R}_+^*$ et pour tout $1 \le i \le n$, $T_i = \tau_1 + \cdots + \tau_i$. Alors,*

(1) T_n suit la loi gamma de paramètres (λ, n) dont la densité est donnée par

$$f_{T_n}(x) = \frac{1}{(n-1)!} \lambda^n x^{n-1} e^{-\lambda x} \mathbf{1}_{[0,+\infty[}(x),$$

(2) la loi jointe de $(T_1, \dots, T_n)$ admet pour densité la fonction

$$f_{(T_1,\dots,T_n)}(t_1, \dots, t_n) = \lambda^n e^{-\lambda t_n} \mathbf{1}_{\{0<t_1<\cdots<t_n\}}. \tag{25}$$

DÉMONSTRATION.
(1) Par récurrence sur n. Pour $n = 1$, il suffit de remarquer que f_{T_1} est la densité d'une loi $\mathscr{E}(\lambda)$. Supposons le résultat vrai pour un $n \ge 1$. En écrivant $T_{n+1} = T_n + \tau_{n+1}$, somme de deux variables aléatoires indépendantes, on a

$$\begin{aligned} f_{T_{n+1}}(x) &= \int_0^x f_{T_n}(u) f_{\tau_1}(x-u)\,du \\ &= \frac{\lambda^{n+1}}{(n-1)!} e^{-\lambda} \int_0^x u^{n-1}\,du \\ &= \frac{\lambda^{n+1}}{n!} e^{-\lambda} x^n. \end{aligned}$$

(2) Les variables $(\tau_i)_{1\le i\le n}$ étant indépendantes, pour toute fonction g mesurable bornée sur $\mathbb{R}^n$, on a par changement de variables

$$\begin{aligned} \mathbb{E}(g(T_1, \dots, T_n)) &= \int_{\mathbb{R}^n} g(s_1, \dots, s_1 + \cdots + s_n) f_{(\tau_1,\dots,\tau_n)}(s_1, \cdots, s_n)\,ds_1 \dots ds_n \\ &= \int_{\mathbb{R}_+^n} g(s_1, \dots, s_1 + \cdots + s_n) \lambda^n e^{-\lambda(s_1+\cdots+s_n)}\,ds_1 \cdots ds_n \\ &= \int_{\mathbb{R}^n} g(t_1, \dots, t_n) \lambda^n e^{-\lambda t_n} \mathbf{1}_{\{0\le t_1<\cdots<t_n\}}\,dt_1 \cdots dt_n. \end{aligned}$$

□

REMARQUE 5.9 (Paramétrage). Attention, les lois Gamma sont parfois paramétrées par $\theta = 1/\lambda$ plutôt que λ. C'est par exemple le cas dans Scilab où les lois exponentielles sont paramétrées par leur espérance et non par son inverse.

PROPOSITION 5.10. *Soit $(N_t)_{t\in\mathbb{R}_+}$ un processus de Poisson d'intensité λ. Alors*

(1) La variable aléatoire N_t suit la loi de Poisson de paramètre λt,

(2) La loi de $(T_1,\dots,T_n)$ sachant que $N_t = n$ est la loi d'un vecteur de variables aléatoires indépendantes réordonnées de loi uniforme sur $[0,t]$.

DÉMONSTRATION.
(1) La clé est la relation suivante : pour tous $t \in \mathbb{R}_+^*$ et $n \in \mathbb{N}$,

$$\{T_n \le t\} = \{N_t \ge n\}.$$

Ainsi,

$$\mathbb{P}(N_t = n) = \mathbb{P}(N_t \ge n) - \mathbb{P}(N_t \ge n+1) = \mathbb{P}(T_n \le t) - \mathbb{P}(T_{n+1} \le t).$$

On a donc, d'après la proposition précédente,

$$\mathbb{P}(N_t = n) = \int_0^t \lambda e^{-\lambda x}\frac{(\lambda x)^{n-1}}{(n-1)!}\,dx - \int_0^t \lambda e^{-\lambda x}\frac{(\lambda x)^n}{n!}\,dx = \left[e^{-\lambda x}\frac{(\lambda x)^n}{n!}\right]_0^t = e^{-\lambda t}\frac{(\lambda t)^n}{n!}.$$

La loi de N_t est donc la loi de Poisson $\mathscr{P}(\lambda t)$.
(2) Soit g une fonction borélienne bornée sur $\mathbb{R}^n$. Alors, pour tous $t \in \mathbb{R}_+$ et $n \in \mathbb{N}$,

$$\begin{aligned}\mathbb{E}\big[g(T_1,\dots,T_n)|N_t = n\big] &= \frac{\mathbb{E}\big[g(T_1,\dots,T_n)\mathbf{1}_{\{T_n\le t<T_{n+1}\}}\big]}{\mathbb{P}(N_t = n)}\\ &= \frac{e^{\lambda t}n!}{(\lambda t)^n}\mathbb{E}\big[g(T_1,\dots,T_n)\mathbf{1}_{\{T_n\le t<T_{n+1}\}}\big].\end{aligned}$$

En utilisant le point (1) de la proposition précédente, on obtient que l'espérance ci-dessus est égale à

$$\int_{\tilde\Delta_t} g(t_1,\dots,t_n)\lambda^{n+1}e^{-\lambda t_{n+1}}\,dt_1\cdots dt_{n+1} = e^{-\lambda t}\lambda^n\int_{t_1<\dots<t_n\le t} g(t_1,\dots,t_n)\,dt_1\cdots dt_n$$

où $\tilde\Delta_t = \{(t_1,\dots,t_{n+1}) : 0 < t_1 < \dots < t_n < t < t_{n+1}\}$. On obtient donc la relation

$$\mathbb{E}\big[g(T_1,\dots,T_n)|N_t = n\big] = \int_{0<t_1<\dots<t_n\le t} g(t_1,\dots,t_n)\frac{n!}{t^n}\,dt_1\dots dt_n,$$

qui, avec la proposition 5.5, fournit le résultat attendu. □

DÉFINITION 5.11. Un processus de comptage $(X_t)_{t\in\mathbb{R}_+}$ est à accroissements indépendants si pour tout $n \in \mathbb{N}$ et pour tout $t_1 \le \dots \le t_n$ les variables aléatoires $X_{t_2} - X_{t_1},\dots,X_{t_n} - X_{t_{n-1}}$ sont indépendantes. Il est à accroissements stationnaires si pour tous $0 \le s \le t$ et $h \in \mathbb{R}_+$, les variables $X_t - X_s$ et $X_{t+h} - X_{s+h}$ ont la même loi.

PROPOSITION 5.12. *Un processus de Poisson $(N_t)_{t\in\mathbb{R}_+}$ est un processus à accroissements indépendants et stationnaires.*

DÉMONSTRATION. On rappelle que la loi jointe de $(T_1,\dots,T_n)$ admet pour densité

$$f_{(T_1,\dots,T_n)}(t_1,\dots,t_n) = \lambda^n e^{-\lambda t_n}\mathbf{1}_{\{0<t_1<\dots<t_n\}}.$$

Soit $s \le t$, on a

$$\begin{aligned}
\mathbb{P}(N_s = l, N_t = k+l) &= \mathbb{P}(T_l \le s < T_{l+1}, T_{l+k} \le t < T_{l+k+1}) \\
&= \int \lambda^{l+k+1} e^{-\lambda t_{l+k+1}} \mathbf{1}_{\{0<t_1<\cdots<t_l<s<t_{l+1}<\cdots<t_{l+k}<t<t_{l+k+1}\}} dt_1 \cdots dt_{l+k+1} \\
&= e^{-\lambda t} \frac{(\lambda s)^l}{l!} \frac{(\lambda(t-s))^k}{k!} \\
&= \mathbb{P}(N_s = l)\mathbb{P}(N_{t-s} = k).
\end{aligned}$$

On en déduit d'une part que N_s et $N_t - N_s$ sont indépendants et d'autre part que $N_t - N_s$ a même loi que N_{t-s}. Ceci montre que le processus est à accroissements stationnaires. De plus, les mêmes calculs donnent, pour tout $0 \le t_1 \le \cdots \le t_n$ et $k_1, \dots, k_n \in \mathbb{N}$,

$$\mathbb{P}(N_{t_1} = k_1, \dots, N_{t_n} = k_1 + \cdots + k_n) = \mathbb{P}(N_{t_1} = k_1)\mathbb{P}(N_{t_2-t_1} = k_2) \cdots \mathbb{P}(N_{t_n-t_{n-1}} = k_n),$$

ce qui prouve l'indépendance des accroissements. □

Réciproquement, le théorème suivant donne une définition alternative des processus de Poisson.

THÉORÈME 5.13. *Soit $(N_t)_{t\in\mathbb{R}_+}$ un processus de comptage à accroissements indépendants et stationnaires tel que $N_0 = 0$ et $N_t \sim \mathscr{P}(\lambda t)$, $\lambda \in \mathbb{R}_+^*$. Alors $(N_t)_{t\in\mathbb{R}_+}$ est un processus de Poisson d'intensité λ.*

DÉMONSTRATION. Commençons par voir que la densité du vecteur $(T_1, \dots, T_n)$ est nécessairement celle donnée par (25). On le fait pour $n = 2$ (le cas général étant similaire). Soit $t_1 < t_2$ et $h_1, h_2 \in \mathbb{R}_+^*$. On a

$$\begin{aligned}
&\mathbb{P}\big(T_1 \in [t_1, t_1 + h_1[, T_2 \in [t_2, t_2 + h_2[\big) \\
&\quad = \mathbb{P}(N_{t_1} = 0, N_{t_1+h_1} = 1, N_{t_2} = 1, N_{t_2+h_2} \ge 2) \\
&\quad = \mathbb{P}(N_{t_1} = 0, N_{t_1+h_1} - N_{t_1} = 1, N_{t_2} - N_{t_1+h_1} = 0, N_{t_2+h_2} - N_{t_2} \in \mathbb{N}^*) \\
&\quad = e^{-\lambda t_1} e^{-\lambda h_1} \lambda h_1 e^{-\lambda(t_2-t_1-h_1)}(1 - e^{-\lambda h_2}) \\
&\quad = e^{-\lambda(t_2+h_2)} \lambda^2 h_1 h_2 + o(h_2),
\end{aligned}$$

où l'on a utilisé l'indépendance et la stationnarité des accroissements. On en déduit

$$f_{(T_1,T_2)}(t_1, t_2) = \lim_{h_1,h_2\to 0} \frac{\mathbb{P}\big(T_1 \in [t_1, t_1 + h_1[, T_2 \in [t_2, t_2 + h_2[\big)}{h_1 h_2} = e^{-\lambda t_2} \lambda^2.$$

Par un changement de variables, on en déduit que la loi du couple $(\tau_1, \tau_2) = (T_1, T_2 - T_1)$ est le produit de deux lois exponentielles de paramètre λ. La même méthode montre que pour tout $n \ge 1$, les variables $\tau_1, \dots, \tau_n$ sont indépendantes et de même loi $\mathscr{E}(\lambda)$. □

3.2. Superposition et amincissement.

PROPOSITION 5.14 (Superposition). *Soit $(M_t)_{t\in\mathbb{R}_+}$ et $(N_t)_{t\in\mathbb{R}_+}$ deux processus de Poisson indépendants d'intensité respective λ et μ. Alors le processus $(M_t + N_t)_{t\in\mathbb{R}_+}$ est un processus de Poisson d'intensité $\lambda + \mu$.*

DÉMONSTRATION. Notons $S_t = M_t + N_t$. On utilise la caractérisation donnée par le théorème 5.13. S_t est de loi de Poisson de paramètre $\lambda + \mu$ comme somme de deux v.a. de Poisson indépendantes (c'est un résultat classique). Comme $S_t - S_s = (M_t - M_s) + (N_t - N_s)$, le fait que $(S_t)_{t\in\mathbb{R}_+}$ est à accroissements stationnaires et indépendants se déduit directement des mêmes propriétés pour $(M_t)_{t\in\mathbb{R}_+}$ et $(N_t)_{t\in\mathbb{R}_+}$, et de leur indépendance. □

PROPOSITION 5.15 (Amincissement). *Soit $(N_t)_{t\in\mathbb{R}_+}$ un processus de Poisson d'intensité λ et une suite $(B_n)_{n\in\mathbb{N}^*}$ de variables aléatoires indépendantes de loi $\mathscr{B}(p)$ et indépendante de $(N_t)_{t\in\mathbb{R}_+}$. On définit deux processus de la façon suivante : pour $t\in\mathbb{R}_+$,*

$$N_t^{(1)} = \sum_{n\in\mathbb{N}^*} B_n \mathbf{1}_{\{T_n\le t\}} \quad \textit{et} \quad N_t^{(2)} = N_t - N_t^{(1)} = \sum_{n\ge 1}(1-B_n)\mathbf{1}_{\{T_n\le t\}}.$$

Alors $(N_t^{(1)})_{t\in\mathbb{R}_+}$ et $(N_t^{(2)})_{t\in\mathbb{R}_+}$ sont deux processus de Poisson indépendants d'intensité respective λp et $\lambda(1-p)$.

DÉMONSTRATION. Soit $0\le s<t$ et $k\in\mathbb{N}$. On a

$$\begin{aligned}
\mathbb{P}(N_t^{(1)}-N_s^{(1)}=k) &= \sum_{n\ge k}\binom{n}{k}p^k(1-p)^{n-k}\mathbb{P}(N_t-N_s=n)\\
&= \sum_{n\ge k}\binom{n}{k}p^k(1-p)^{n-k}e^{-\lambda(t-s)}\frac{(\lambda(t-s))^n}{n!}\\
&= \frac{(p\lambda(t-s))^k}{k!}e^{-\lambda(t-s)}\sum_{n\ge k}\frac{(\lambda(1-p))^{n-k}}{(n-k)!}\\
&= \frac{(p\lambda(t-s))^k}{k!}e^{-p\lambda(t-s)}.
\end{aligned}$$

Ceci montre que $N_t^{(1)}-N_s^{(1)}$ suit une loi $\mathscr{P}(p\lambda)$ et on montre de même que $N_t^{(2)}-N_s^{(2)}$ suit une loi $\mathscr{P}((1-p)\lambda)$.
Soit $0\le s_1<t_1\le s_2<t_2\le\cdots\le s_n<t_n$. En notant

$$\Delta_i = N_{t_i}-N_{s_i},\quad \Delta_i^{(1)} = N_{t_i}^{(1)}-N_{s_i}^{(1)} \quad \text{et} \quad \Delta_i^{(2)} = N_{t_i}^{(2)}-N_{s_i}^{(2)},$$

on a, pour $k_1,\dots,k_n$ et $l_1,\dots,l_n$ entiers,

$$\begin{aligned}
&\mathbb{P}(\forall i=1,\dots,n,\ \Delta_i^{(1)}=k_i,\Delta_i^{(2)}=l_i)\\
&\quad= \mathbb{P}(\forall i=1,\dots,n,\ \Delta_i^{(1)}=k_i \mid \forall i=1,\dots,n,\ \Delta_i=k_i+l_i)\mathbb{P}(\forall i=1,\dots,n,\ \Delta_i=k_i+l_i)\\
&\quad= \prod_{i=1}^n\binom{k_i+l_i}{k_i}p^{k_i}(1-p)^{l_i}\prod_{i=1}^n\mathbb{P}(\Delta_i=k_i+l_i)\\
&\quad= \prod_{i=1}^n\binom{k_i+l_i}{k_i}p^{k_i}(1-p)^{l_i}\prod_{i=1}^n e^{-\lambda(t_i-s_i)}\frac{(\lambda(t_i-s_i))^{k_i+l_i}}{(k_i+l_i)!}\\
&\quad= \prod_{i=1}^n e^{-p\lambda(t_i-s_i)}\frac{(p\lambda(t_i-s_i))^{k_i}}{k_i!}e^{-(1-p)\lambda(t_i-s_i)}\frac{((1-p)\lambda(t_i-s_i))^{l_i}}{l_i!}\\
&\quad= \prod_{i=1}^n\mathbb{P}(\Delta_i^{(1)}=k_i)\mathbb{P}(\Delta_i^{(2)}=l_i).
\end{aligned}$$

Ceci montre que les processus $(N_t^{(1)})_{t\in\mathbb{R}_+}$ et $(N_t^{(2)})_{t\in\mathbb{R}_+}$ sont à accroissements indépendants et qu'ils sont indépendants l'un de l'autre. □

EXERCICE 5.2. Soit $(T_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes de même loi $\mathscr{E}(\lambda)$ et N une variable aléatoire indépendante de cette suite et de loi géométrique de paramètre p. Montrer que $T=\sum_{n=1}^N T_n$ suit une loi exponentielle de paramètre λp. Peut-on faire un lien avec la propriété d'amincissement des processus de Poisson (voir la proposition 5.15) ?

Paradoxe de l'inspection (ou de l'autobus) On dispose d'un stock infini d'ampoules dont les durées de vie sont modélisées par des variables aléatoires indépendantes de loi $\mathscr{E}(\lambda)$, $\lambda \in \mathbb{R}_+^*$. À l'instant 0, on allume la première ampoule. Chaque ampoule est ensuite remplacée instantanément par une neuve dès qu'elle se casse. Si l'on désigne par N_t le nombre d'ampoules utilisées à l'instant t, alors le processus aléatoire $(N_t)_{t\in\mathbb{R}_+}$ est un processus de Poisson d'intensité λ.
Soit $t \in \mathbb{R}_+^*$. On s'intéresse à l'espérance de la durée de vie de l'ampoule allumée à l'instant t
On note, pour $i \in \mathbb{N}^*$, τ_i la durée de vie de la i-ème ampoule (c'est-à-dire le temps inter-saut $T_i - T_{i-1}$. L'ampoule en fonctionnement à l'instant t a une durée de vie τ_{N_t+1} (le temps écoulé entre la dernière défaillance avant t et la suivante). Écrivons $\tau_{N_t+1} = A + B$, où $A = t - T_{N_t}$ et $B = T_{N_t+1} - t$ avec comme précédemment $T_n = \tau_1 + \cdots + \tau_n$.
Commençons par déterminer les lois des variables A et B.
Clairement A est à valeurs dans $[0, t]$ et pour $s \in [0, t]$, on a

$$\mathbb{P}(A > s) = \mathbb{P}(T_{N_t} < t - s) = \mathbb{P}(N_t - N_{t-s} = 0) = \mathbb{P}(N_s = 0) = e^{-\lambda s}.$$

La loi de A est donc la loi de $X \wedge t$ où X suit une loi exponentielle de paramètre λ. Son espérance est

$$\mathbb{E}(A) = \int_{[0,t]} \mathbb{P}(A > s)ds = \int_{[0,t]} e^{-\lambda s} ds = \frac{1 - e^{-\lambda t}}{\lambda}.$$

La variable aléatoire B vérifie, pour tout $s \in \mathbb{R}_+^*$,

$$\mathbb{P}(B > s) = \mathbb{P}(T_{N_t+1} > s + t) = \mathbb{P}(N_{s+t} - N_t = 0) = \mathbb{P}(N_s = 0) = e^{-\lambda s}.$$

Donc B suit une loi exponentielle de paramètre λ et son espérance vaut $\mathbb{E}(B) = \frac{1}{\lambda}$.
En conclusion, on a

$$\mathbb{E}(\tau_{N_t+1}) = \mathbb{E}(A) + \mathbb{E}(B) = \frac{2}{\lambda} - \frac{e^{-\lambda t}}{\lambda} > \frac{1}{\lambda} = \mathbb{E}(\tau_1).$$

Moralité : l'espérance de la durée de vie de l'ampoule inspectée au temps t est strictement plus grande que la durée de vie moyenne des ampoules. Lorsque t est grand, elle est même presque deux fois plus grande.
Ce paradoxe est parfois présenté en remplaçant les changements d'ampoules par des passages d'autobus à un arrêt. La conclusion est que lorsque l'on se rend à un arrêt de bus, en moyenne, le temps entre le passage du dernier bus et celui du prochain bus est strictement supérieur au temps moyen entre les passages successifs des bus. Ce paradoxe repose très fortement sur la propriété d'absence de mémoire des temps inter-sauts qui n'est clairement pas satisfaite pour un réseau de bus raisonnable.

4. Corrigés

CORRIGÉ DE L'EXERCICE 5.1.
1. La variable T_1^N s'écrit comme le minimum de N variables indépendantes de loi exponentielle de paramètre α. Elle suit donc la loi exponentielle de paramètre $N\alpha$.
2. Si S et T sont deux variables aléatoires indépendantes de lois exponentielles alors, pour tout $t \in \mathbb{R}_+^*$,

$$\mathbb{P}(T > S + t | T > S) = \mathbb{P}(T > t).$$

Ceci assure que les variables $(\tau_n^N)_{1\le n\le N}$ sont indépendantes et que τ_n^N suit la loi exponentielle de paramètre $N - n + 1$.

3. Pour tout $N \in \mathbb{N}^*$,

$$\begin{aligned}\mathbb{E}(T_{N/2}^N) &= \sum_{n=1}^{N/2} \mathbb{E}(\tau_n^N) = \sum_{n=1}^{N/2} \frac{1}{\alpha(N-n+1)} = \sum_{n=N/2+1}^{N} \frac{1}{\alpha n} = \sum_{n=1}^{N} \frac{1}{\alpha n} - \sum_{n=1}^{N/2} \frac{1}{\alpha n} \\ &= \frac{1}{\alpha}\big((\ln(N)+\gamma+o(1)) - (\ln(N/2)+\gamma+o(1))\big) = \frac{\ln(2)}{\alpha} + o(1).\end{aligned}$$

4. Si X suit la loi exponentielle de paramètre 1, $\mathbb{E}\big[(X-\mathbb{E}(X))^2\big] = \mathbb{V}(X) = 1$. De plus, une intégration par parties assure que $\mathbb{E}(X^n) = n\mathbb{E}(X^{n-1})$. Ainsi, $\mathbb{E}(X^n) = n!$. On a donc

$$\mathbb{E}\big[(X-\mathbb{E}(X))^4\big] = 4! - 4\times 3! + 6\times 2! - 4\times 1! + 1 = 9.$$

5. Notons $Z_n = \tau_n^N - \mathbb{E}(\tau_n^N)$. On a

$$\mathbb{E}(Z_n^2) = \frac{1}{\alpha^2} \quad \text{et} \quad \mathbb{E}(Z_n^4) = \frac{9}{\alpha^4}$$

car si E suit la loi $\mathscr{E}(\alpha)$ alors αE suit la loi $\mathscr{E}(1)$. Le développement de $(Z_1+\cdots+Z_{N/2})^4$ fait intervenir des termes de la forme

$$Z_n^4, \quad Z_n^3 Z_m, \quad Z_n^2 Z_m^2, \quad Z_n^2 Z_m Z_l, \quad Z_n Z_m Z_l Z_j$$

où j, l, m et n sont distincts. Seuls les premiers et troisièmes types ont une espérance non nulle puisque les variables (Z_n) sont indépendantes et centrées. On a donc pour tout $N \in \mathbb{N}^*$

$$\begin{aligned}\mathbb{E}\Big[\big(T_{N/2}^N - \mathbb{E}(T_{N/2}^N)\big)^4\Big] &= \mathbb{E}\left[\left(\sum_{n=1}^{N/2} Z_n\right)^4\right] \\ &= \sum_{n=1}^{N/2} \mathbb{E}(Z_n^4) + 6 \sum_{1\le n\ne m\le N/2}^{N/2} \mathbb{E}(Z_n^2)\mathbb{E}(Z_m^2) \\ &= \frac{9}{\alpha^4} \sum_{n=N/2+1}^{N} \frac{1}{n^4} + \frac{6}{\alpha^4}\left(\sum_{n=N/2+1}^{N} \frac{1}{n^2}\right)^2.\end{aligned}$$

Par comparaison série-intégrale,

$$\sum_{n=N/2+1}^{N} \frac{1}{n^2} \le \sum_{n\ge N/2+1} \frac{1}{n^2} \le \int_{N/2}^{\infty} \frac{1}{x^2}\,dx = \mathscr{O}(N^{-1}),$$

et

$$\sum_{n=N/2+1}^{N} \frac{1}{n^4} \le \sum_{n\ge N/2+1} \frac{1}{n^4} \le \int_{N/2}^{\infty} \frac{1}{x^4}\,dx = \mathscr{O}(N^{-3}).$$

On a donc $\mathbb{E}\Big[\big(T_{N/2}^N - \mathbb{E}(T_{N/2}^N)\big)^4\Big] = \mathscr{O}(N^{-2})$.

6. On a, pour tout $\varepsilon \in \mathbb{R}_+^*$,

$$\mathbb{P}\big(\big|T_{N/2}^N - \mathbb{E}(T_{N/2}^N)\big| > \varepsilon\big) \le \frac{\mathbb{E}\Big[\big(T_{N/2}^N - \mathbb{E}(T_{N/2}^N)\big)^4\Big]}{\varepsilon^4} \le \frac{C}{\varepsilon^4 N^2}.$$

Le lemme de Borel-Cantelli assure donc que $T_{N/2}^N - \mathbb{E}(T_{N/2}^N)$ converge presque sûrement vers 0. On a donc

$$T_{N/2}^N \xrightarrow[N\to\infty]{p.s.} \frac{\ln(2)}{\alpha}.$$

CORRIGÉ DE L'EXERCICE 5.2. On procède de la même manière que dans l'exercice 2.15 en utilisant que

$$\varphi(t) = \frac{\lambda}{\lambda - it} \quad \text{et} \quad G(s) = \frac{ps}{1-(1-p)s}.$$

Cette propriété est une conséquence de la propriété d'amincissement. Les variables $(T_n)_{n\in\mathbb{N}^*}$ jouent le rôle des temps de saut du processus de Poisson principal $(N_t)_{t\in\mathbb{R}_+}$ et le premier temps de saut du sous-processus $(N_t^{(1)})_{t\in\mathbb{R}_+}$ se produit après un nombre aléatoire de sauts de N de loi géométrique de paramètre p.

Chapitre 6

Chaînes de Markov à temps continu

1. Définitions et généralités

On s'intéresse à des processus $(X_t)_{t\in\mathbb{R}_+}$ indexés par le temps continu $t \in \mathbb{R}_+$ à valeurs dans un espace fini ou dénombrable E. Une trajectoire du processus est donc une fonction $t \mapsto X_t(\omega)$ de $\mathbb{R}_+$ dans I qui dépend d'un aléa ω, tout comme une trajectoire d'une chaîne de Markov $(Y_n)_{n\in\mathbb{N}}$ est une suite $(Y_n(\omega))_{n\in\mathbb{N}}$ à valeurs dans E qui dépend de l'aléa ω. On supposera dans toute la suite que les trajectoires de X sont continues à droite : pour (presque tout) ω et tout $t \in \mathbb{R}_+$, il existe $\varepsilon > 0$ tel que

$$\forall s \in [t, t+\varepsilon], \quad X_s(\omega) = X_t(\omega).$$

Pour une trajectoire donnée, trois comportements différents sont alors possibles :

(1) la trajectoire devient constante à partir d'un certain instant,

(2) la trajectoire possède un nombre infini de sauts sur $\mathbb{R}_+$ mais seulement un nombre fini sur tout intervalle borné,

(3) la trajectoire possède un nombre infini de sauts dans un intervalle de temps borné ; il existe alors un temps d'explosion ζ fini. Après ce premier instant d'explosion, la trajectoire peut repartir, exploser à nouveau ou non...

Dans toute la suite, nous considérons les processus jusqu'à leur premier temps d'explosion. On parle alors de processus minimal.
On appelle $J_0, J_1, \ldots$ les temps de saut de $(X_t)_{t\in\mathbb{R}_+}$ et $S_1, S_2, \ldots$ les temps d'attente. Ils se déduisent de $(X_t)_{t\in\mathbb{R}_+}$ par les relations

$$J_0 = 0, \quad J_{n+1} = \inf\{t \geq J_n \,;\, X_t \neq X_{J_n}\},$$

pour $n \in \mathbb{N}$ avec la convention $\inf \emptyset = +\infty$ et pour $n \in \mathbb{N}^*$,

$$S_n = \begin{cases} J_n - J_{n-1} & \text{si } J_{n-1} < \infty, \\ \infty & \text{sinon.} \end{cases}$$

REMARQUE 6.1. La propriété de continuité à droite assure que $S_n > 0$ pour tout n. Si $J_{n+1} = \infty$ alors on peut définir $X_\infty = X_{J_n}$, sinon, X_∞ n'est pas défini. Le (premier) temps d'explosion ζ est défini par

$$\zeta = \sup_n J_n = \sum_{n=1}^{\infty} S_n.$$

Le processus à temps discret $(Y_n)_{n\in\mathbb{N}}$ défini par $Y_n = X_{J_n}$ pour tout $n \in \mathbb{N}$ est appelée chaîne incluse. Il représente la suite des valeurs prises par $(X_t)_{t\in\mathbb{R}_+}$.

DÉFINITION 6.2. Soit E un espace fini dénombrable. Un générateur sur E est une matrice $Q = (q_{ij})_{i,j\in E}$ (éventuellement de taille infinie) vérifiant

(1) pour tout $i \in E$, $0 \leq -q_{ii} < \infty$,

(2) pour tous $i \neq j$, $q_{ij} \geq 0$,

(3) pour tout $i \in E$, $\sum_{j\in E} q_{ij} = 0$.

On notera $q_i = q(i) = -q_{ii}$ pour tout $i \in I$.
À un générateur Q, on associe la matrice de saut $\Pi = (\pi_{ij})_{i,j\in E}$ donnée par

$$\pi_{ij} = \begin{cases} q_{ij}/q_i & \text{si } j \neq i \text{ et } q_i \neq 0, \\ 0 & \text{si } j \neq i \text{ et } q_i = 0, \end{cases} \quad \text{et} \quad \pi_{ii} = \begin{cases} 0 & \text{si } q_i \neq 0, \\ 1 & \text{si } q_i = 0. \end{cases}$$

REMARQUE 6.3. La matrice Π est stochastique : ses coefficients sont positifs et leur somme sur chaque ligne vaut 1.

La loi d'une chaîne de Markov à temps continu $(X_t)_{t\in\mathbb{R}_+}$ est déterminée par une mesure de probabilité λ sur E (identifiée à un vecteur ligne) et un générateur Q. La mesure λ détermine la distribution initiale (c'est la loi de X_0) et la matrice Q détermine la dynamique de la chaîne. On peut décrire de deux manières équivalentes la dynamique de $(X_t)_{t\in\mathbb{R}_+}$.

La première description se fait par l'intermédiaire de la chaîne incluse et des temps d'attente.

DÉFINITION 6.4. Le processus $(X_t)_{t\in\mathbb{R}_+}$ est une chaîne de Markov homogène à temps continu de générateur Q si

(1) $(Y_n)_{n\in\mathbb{N}}$ est une chaîne de Markov à temps discret de loi initiale λ et de matrice de transition Π,
(2) conditionnellement à $Y_0 = i_0, \dots, Y_{n-1} = i_{n-1}$, les temps d'attente $S_1, \dots, S_n$ sont des variables aléatoires exponentielles indépendantes de paramètres respectifs $q_{i_0}, \dots, q_{i_{n-1}}$.

Plus simplement, sachant que la chaîne X est issue de i, elle y reste un temps exponentiel de paramètre q_i puis saute à un nouvel état, en choisissant l'état j avec probabilité π_{ij}. Elle oublie ensuite ce qui s'est passé, attend un nouveau temps exponentiel de paramètre q_j (indépendant du précédent), etc.

La second description est l'analogue de la définition à temps discret. Notons

$$p_{ij}(t) = \mathbb{P}(X_t = j | X_0 = i).$$

DÉFINITION 6.5. Le processus $(X_t)_{t\in\mathbb{R}_+}$ est une chaîne de Markov homogène à temps continu si, pour tout $n \in \mathbb{N}$, tous $0 \le t_0 \le t_1 \le \cdots \le t_{n+1}$ et tous états $i_0, \dots, i_{n+1}$,

$$\mathbb{P}(X_{t_{n+1}} = i_{n+1} | X_{t_0} = i_0, \dots, X_{t_n} = i_n) = p_{i_n i_{n+1}}(t_{n+1} - t_n).$$

THÉORÈME 6.6 (Propriété de Markov). *Soit $(X_t)_{t\in\mathbb{R}_+}$ une chaîne de Markov à temps continu de générateur Q. Alors, conditionnellement à $X_s = i$, $(X_{s+t})_{t\in\mathbb{R}_+}$ est une chaîne de Markov à temps continu de générateur Q issue de i et indépendante de $(X_r)_{0\le r\le s}$.*

REMARQUE 6.7. Pour tout $t \in \mathbb{R}_+$, $P(t) = (p_{ij}(t))_{ij}$ est une matrice stochastique et $t \mapsto P(t)$ vérifie la propriété dite de semi-groupe suivante :

$$P(t+s) = P(t)P(s) \quad \text{et} \quad P(0) = Id.$$

En effet, pour tous i, j dans E,

$$\begin{aligned} p_{ij}(t+s) &= \sum_{k\in E} \mathbb{P}(X_{t+s} = j, X_t = k | X_0 = i) \\ &= \sum_{k\in E} \mathbb{P}(X_{t+s} = j | X_0 = i, X_t = k)\mathbb{P}(X_t = k | X_0 = i) \\ &= \sum_{k\in E} p_{kj}(s) p_{ik}(t) = (P(t)P(s))_{ij}. \end{aligned}$$

REMARQUE 6.8. On peut montrer qu'il y a une bijection entre générateurs et semi-groupes en toute généralité. Si E est de cardinal fini, le lien est explicite : si Q est un générateur alors $(P(t))_{t\geq 0}$ défini par, pour tout $t \in \mathbb{R}_+$,

$$P(t) = e^{tQ} = \sum_{n=0}^{\infty} \frac{t^n}{n!} Q^n,$$

est un semi-groupe. Si $(P(t))_{t\in\mathbb{R}_+}$ est un semi-groupe, il existe un générateur Q tel que, pour tout $t \in \mathbb{R}_+$, $P(t) = \exp(tQ)$.

2. Processus de naissance

Un processus de naissance est une généralisation du processus de Poisson dans laquelle le paramètre λ peut dépendre de l'état courant du processus. Les paramètres d'un processus de naissance sont les taux de naissances $(q_i)_{i\in\mathbb{N}}$ supposés à valeurs dans $\mathbb{R}_+$.

DÉFINITION 6.9. Un processus de Markov $(X_t)_{t\in\mathbb{R}_+}$ à valeurs dans $\mathbb{N}\cup\{\infty\}$ est un processus de naissance de taux $(q_i)_{i\in\mathbb{N}}$ si, conditionnellement à $X_0 = i$, ses temps d'attente $S_1, S_2, \ldots$ sont des variables aléatoires indépendantes de lois exponentielles de paramètres respectifs $q_i, q_{i+1}, \ldots$, et sa chaîne incluse $(Y_n)_{n\in\mathbb{N}}$ est donnée par $Y_n = i + n$.

REMARQUE 6.10. Le processus de Poisson correspond à la suite $(q_i)_{i\in\mathbb{N}}$ constante égale à λ.

Croissance d'une colonie de bactéries. Chaque bactérie dans une colonie se divise en deux bactéries identiques après un temps exponentiel de paramètre λ et ce indépendamment des autres. Si i individus sont présents alors la première division apparaît après un temps exponentiel de paramètre λi. Il y a alors $i+1$ bactéries et, par la propriété d'absence de mémoire, les horloges des bactéries sont remises à zéro. Soit X_t la taille de la colonie au temps t et supposons que $X_0 = 1$. Soit T le temps de la première naissance. Alors

$$\mathbb{E}(X_t) = \mathbb{E}\big(X_t \mathbf{1}_{\{T\leq t\}}\big) + \mathbb{E}\big(X_t \mathbf{1}_{\{T>t\}}\big) = \int_0^t \lambda e^{-\lambda s} \mathbb{E}(X_t | T = s)\, ds + e^{-\lambda t}.$$

Soit $\mu(t) = \mathbb{E}(X_t)$ alors $\mathbb{E}(X_t|T = s) = 2\mu(t-s)$ pour $0 \leq s \leq t$ donc μ vérifie l'équation fonctionnelle

$$\mu(t) = e^{-\lambda t} + \int_0^t \lambda e^{-\lambda s} \mu(t-s)\, ds,$$

qui peut encore s'écrire grâce au changement de variables $r = t - s$

$$e^{\lambda t}\mu(t) = 2\lambda \int_0^t e^{\lambda r} \mu(r)\, dr.$$

La fonction μ est donc solution de l'équation différentielle $\mu'(t) = \lambda\mu(t)$, avec la condition initiale $\mu(0) = 1$. Ceci assure donc que $\mathbb{E}(X_t) = e^{\lambda t}$: la taille de la population croît exponentiellement vite (mais n'explose pas en temps fini).

La principale différence entre le processus de Poisson et un processus de naissance général est la possibilité qu'a le second d'exploser. Il est très facile de caractériser les processus qui explosent grâce à la proposition 5.3. Remarquons qu'un processus de naissance explose si et seulement si il tend vers $+\infty$ en temps fini.

THÉORÈME 6.11. *Soit $(X_t)_{t\in\mathbb{R}_+}$ un processus de naissance de taux $(q_i)_{i\in\mathbb{N}}$ issu de 0.*

$$\textit{Si} \quad \sum_{i\in\mathbb{N}} \frac{1}{q_i} < \infty \quad \textit{alors} \quad \mathbb{P}(\zeta < \infty) = 1 \quad \textit{et si} \quad \sum_{i\in\mathbb{N}} \frac{1}{q_i} = \infty \quad \textit{alors} \quad \mathbb{P}(\zeta = \infty) = 1.$$

THÉORÈME 6.12. *Soit $(X_t)_{t\in\mathbb{R}_+}$ un processus croissant, continu à droite à valeurs dans $\mathbb{N}\cup\{\infty\}$. Soit $(q_i)_{i\in\mathbb{N}}$ une suite de réels positifs. Les deux propriétés suivantes sont équivalentes :*

(1) Conditionnellement à $X_0 = i$, les temps d'attente $S_1, S_2, \dots$ sont des variables aléatores indépendantes de lois exponentielles de paramètres respectifs $q_i, q_{i+1}, \dots$ et la chaîne incluse est donnée par $Y_n = i + n$ pour tout $n \in \mathbb{N}$;

(2) Pour tous $t, h \in \mathbb{R}_+$, conditionnellement à $X_t = i$, X_{t+h} est indépendant de $(X_r)_{0\le r\le t}$ et, quand h décroît vers 0, uniformément en t,

$$\mathbb{P}(X_{t+h} = i|X_t = i) = 1 - q_i h + o(h)$$
$$\mathbb{P}(X_{t+h} = i+1|X_t = i) = q_i h + o(h).$$

REMARQUE 6.13. Notons $p_{ij}(t) = \mathbb{P}(X_t = j|X_0 = i)$. On peut montrer grâce à la formulation (2) que $P(t) = (p_{i,j}(t))_{ij}$ est solution de

$$p'_{ij}(t) = -q_j p_{ij}(t) + q_{j-1} p_{i,j-1}(t)$$

que l'on peut écrire sous forme compacte $P'(t) = P(t)Q$ où Q est le générateur

$$Q = \begin{pmatrix} -q_0 & q_0 & 0 & 0 & \dots \\ 0 & -q_1 & q_1 & 0 & \dots \\ 0 & 0 & -q_2 & q_2 & \dots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix}$$

3. Classes

Les notions de classes sont les mêmes à temps continu et à temps discret. Plus précisément, $(X_t)_{t\in\mathbb{R}_+}$ a la même structure de classe que sa chaîne incluse $(Y_n)_{n\in\mathbb{N}}$.
On dit que i mène à j, et on note $i \to j$ si

$$\mathbb{P}_i(X_t = j, \text{ pour un certain temps } t \in \mathbb{R}_+) > 0.$$

On dit que i et j communiquent, et on note $i \leftrightarrow j$ si $i \to j$ et $j \to i$.

THÉORÈME 6.14. *Pour deux états distincts i et j, les propriétés suivantes sont équivalentes :*

(1) $i \to j$,

(2) $i \to j$ pour la chaîne incluse,

(3) il existe $n \in \mathbb{N}$ et $i_0, \dots, i_n$ dans E tels que $q_{i_0 i_1} \dots q_{i_{n-1} i_n} > 0$,

(4) pour tout $t \in \mathbb{R}_+^*$, $p_{ij}(t) > 0$,

(5) il existe $t \in \mathbb{R}_+^*$, $p_{ij}(t) > 0$.

DÉMONSTRATION. Les implications (4) ⇒ (5) ⇒ (1) ⇒ (2) découlent des définitions. Si (2) est vraie, il existe $n \in \mathbb{N}$ et $i_0, \dots, i_n$ dans E tels que $\pi_{i_0 i_1} \dots \pi_{i_{n-1} i_n} > 0$, ce qui implique (3). Si $q_{ij} > 0$ alors

$$p_{ij}(t) \ge \mathbb{P}_i(J_1 \le t, Y_1 = j, J_2 > t) = (1 - e^{-q_i t})\pi_{ij} e^{-q_j t} > 0,$$

pour tout $t \in \mathbb{R}_+^*$. Donc si (3) est vraie, alors

$$p_{ij}(t) \ge p_{i_0 i_1}(t/n) \cdots p_{i_{n-1} i_n}(t/n) > 0$$

pour tout $t \in \mathbb{R}_+^*$ et (4) est établie. □

REMARQUE 6.15. Le point (4) montre que la situation est plus simple en temps continu qu'en temps discret. En temps continu, si i mène à j alors pour tout $t > 0$, avec une probabilité strictement positive, X_t peut valoir j (sachant que $X_0 = i$). En temps discret, si i mène à j, il faut parfois un certain temps avant que la chaîne issue de i atteigne j et des phénomènes de périodicité apparaissent. Par exemple, pour la marche aléatoire simple issue de 0, le temps d'atteinte de 4 sera supérieur ou égal à 4 et pair.

4. Temps d'atteinte et d'absorption

Soit $(X_t)_{t\in\mathbb{R}_+}$ une chaîne de Markov de générateur Q. Le temps d'atteinte d'une partie A de E est la variable aléatoire D^A définie par

$$D^A = \inf\{t \in \mathbb{R}_+ \ ; \ X_t \in A\}, \quad \text{avec la convention } \inf\emptyset = +\infty.$$

Si H^A est le temps d'atteinte de A pour la chaîne incluse, alors $\{D^A < \infty\} = \{H^A < \infty\}$, et sur cet ensemble, on a $D^A = J_{H^A}$.

EXERCICE 6.1. On considère la chaîne de générateur

$$Q = \begin{pmatrix} -2 & 1 & 1 & 0 \\ 2 & -6 & 2 & 2 \\ 3 & 3 & -9 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Quel est le temps moyen mis par X pour atteindre 4 depuis 1 ?

5. Récurrence et transience

DÉFINITION 6.16. Soit $(X_t)_{t\in\mathbb{R}_+}$ une chaîne de Markov de générateur Q.

(1) L'état i est dit récurrent si $\mathbb{P}_i(\{t \in \mathbb{R}_+,\ X_t = i\}$ est non borné$) = 1$.

(2) L'état i est dit transient (ou transitoire) si $\mathbb{P}_i(\{t \in \mathbb{R}_+,\ X_t = i\}$ est non borné$) = 0$.

REMARQUE 6.17. Si $(X_t)_{t\in\mathbb{R}_+}$ peut exploser en étant issu de i, alors i ne peut être récurrent.

THÉORÈME 6.18. *On a*

(1) si i est récurrent pour la chaîne incluse, alors i est récurrent pour X ;

(2) si i est transient pour la chaîne incluse alors i est transient pour X ;

(3) tout état est soit récurrent, soit transient ;

(4) récurrence et transience sont des propriétés de classes.

DÉMONSTRATION. Supposons que i soit récurrent pour Y. Si $X_0 = i$, X n'explose pas, donc J_n tend vers $+\infty$. De plus, $X_{J_n} = Y_n = i$ infiniment souvent et $\{t \in \mathbb{R}_+,\ X_t = i\}$ est non borné avec probabilité 1.
Supposons que i soit transient pour Y. Si $X_0 = i$, alors

$$N = \sup\{n \in \mathbb{N},\ Y_n = i\} < +\infty \quad p.s.$$

donc $\{t \in \mathbb{R}_+,\ X_t = i\}$ est borné par J_{N+1}, qui est fini avec probabilité 1 puisque Y n'a pas atteint de point absorbant avant le temps N.
Les points (3) et (4) se déduisent des résultats analogues sur Y. □

Comme en temps discret, on peut reformuler les notions de récurrence et transience en terme de temps de premier retour. On définit le temps T_i de premier retour de $(X_t)_{t\in\mathbb{R}_+}$ dans l'état i par

$$T_i(\omega) = \inf\{t \geq J_1(\omega) \ ; \ X_t(\omega) = i\}.$$

REMARQUE 6.19. Attention à ne pas confondre les temps de premier passage et de premier retour. Dans le second cas, on s'assure que l'on a quitté le point de départ.

THÉORÈME 6.20. *On a la dichotomie suivante :*

(1) si $q_i = 0$ ou $\mathbb{P}_i(T_i < \infty) = 1$, alors i est récurrent et $\int_0^\infty p_{ii}(t)\,dt = \infty$,

(2) si $q_i > 0$ et $\mathbb{P}_i(T_i < \infty) < 1$, alors i est transient et $\int_0^\infty p_{ii}(t)\,dt < \infty$.

DÉMONSTRATION. Si $q_i = 0$, alors $(X_t)_{t\in\mathbb{R}_+}$ ne peut quitter i qui est donc récurrent. De plus, $p_{ii}(t) = 1$ pour tout $t \in \mathbb{R}_+$ et $\int_0^\infty p_{ii}(t)\,dt = \infty$. Supposons que $q_i > 0$. Soit N_i le temps de premier retour en i pour la chaîne incluse $(Y_n)_{n\in\mathbb{N}}$. Alors $\mathbb{P}_i(N_i < \infty) = \mathbb{P}_i(T_i < \infty)$. Comme i est récurrent pour Y ssi $\mathbb{P}_i(N_i < \infty) = 1$, i est récurrent pour X si et seulement si $\mathbb{P}_i(T_i < \infty) = 1$. De plus, en notant $\pi_{ij}^{(n)} = (\Pi^n)_{ij}$, on va montrer que

$$\int_0^\infty p_{ii}(t)\,dt = \frac{1}{q_i}\sum_{n=0}^\infty \pi_{ii}^{(n)},$$

ce qui assure que i est récurrent ssi $\int_0^\infty p_{ii}(t)\,dt$ est infini d'après le résultat correspondant pour la chaîne incluse. Pour établir la relation ci-dessus, on écrit, en discutant sur le nombre de sauts qui ont eu lieu avant le temps t,

$$\begin{aligned}\int_0^\infty p_{ii}(t)\,dt &= \int_0^\infty \mathbb{E}_i\big(\mathbf{1}_{\{X_t=i\}}\big)\,dt = \mathbb{E}_i\left(\int_0^\infty \mathbf{1}_{\{X_t=i\}}\,dt\right)\\ &= \mathbb{E}_i\left(\int_0^\infty \mathbf{1}_{\{X_t=i\}}\sum_{n=0}^\infty \mathbf{1}_{\{J_n\le t<J_{n+1}\}}\,dt\right) = \sum_{n=0}^\infty \mathbb{E}_i\left(\mathbf{1}_{\{Y_n=i\}}\int_0^\infty \mathbf{1}_{\{J_n\le t<J_{n+1}\}}\,dt\right)\\ &= \sum_{n=0}^\infty \mathbb{E}_i\big(S_{n+1}\mathbf{1}_{\{Y_n=i\}}\big) = \sum_{n=0}^\infty \mathbb{E}_i(S_{n+1}|Y_n=i)\mathbb{P}_i(Y_n=i) = \frac{1}{q_i}\sum_{n=0}^\infty \pi_{ii}^{(n)},\end{aligned}$$

ce qui est le résultat annoncé. □

6. Mesures invariantes

On dit que ν est une mesure invariante pour X si

$$\nu Q = 0.$$

REMARQUE 6.21. Une mesure ν est invariante ssi, si X_0 suit la loi ν, alors X_t suit la loi ν. En effet, si X_0 suit la loi ν alors X_t suit la loi $\nu P(t)$. Cette mesure ne dépend pas du temps si $(\nu P(t))' = 0$, c'est-à-dire si $\nu Q P(t) = 0$. Or $P(t)$ est une matrice inversible (penser à $P(t) = e^{tQ}$) donc la loi de X_t est indépendante du temps ssi $\nu Q = 0$.

On peut relier les mesures invariantes de $(X_t)_{t\in\mathbb{R}_+}$ aux mesures invariantes de sa chaîne incluse.

THÉORÈME 6.22. *Soit Q un générateur, Π sa matrice de saut et ν une mesure. Les deux assertions suivantes sont équivalentes :*

(1) ν est invariante ;

(2) $\mu\Pi = \mu$ avec $\mu_i = \nu_i q_i$, c'est-à-dire que μ est invariante pour la chaîne incluse.

DÉMONSTRATION. On a, par définition, $q_i(\pi_{ij} - \delta_{ij}) = q_{ij}$ pour tous i, j donc

$$(\mu(\Pi - Id))_j = \sum_{i\in E}\mu_i(\pi_{ij} - \delta_{ij}) = \sum_{i\in E}\nu_i q_{ij} = (\nu Q)_j.$$

□

Ce lien permet de réutiliser les théorèmes d'existence et d'unicité connus pour les chaînes à temps discret.

THÉORÈME 6.23. *Supposons que Q soit le générateur d'une chaîne irréductible et récurrente. Alors Q admet une mesure invariante, unique à une constante multiplicative près.*

DÉMONSTRATION. Une fois exclu le cas où $E = \{i\}$, l'irréductibilité assure que pour tout $i \in E$, $q_i > 0$. Ceci entraîne que Π est irréductible et récurrente donc Π admet une mesure invariante unique à une constante multiplicative près. □

REMARQUE 6.24. Attention, si l'espace d'état E est infini, rien ne dit que les mesures invariantes sont de masse finie : l'existence d'une mesure de probabilité invariante n'est pas assurée. Elle l'est par contre dans le cas où E est fini.

Pour répondre complètement à cette question, il faut introduire la notion de récurrence positive. On dit qu'un point i est récurrent positif si $q_i = 0$ ou si le temps moyen de premier retour en i $m_i = \mathbb{E}_i(T_i)$ est fini. Un point récurrent non récurrent positif est dit récurrent nul. On obtient alors le lien entre mesure de probabilité invariante et espérance des temps de retour avec une petite subtilité par rapport au temps discret.

THÉORÈME 6.25. *Soit Q un générateur irréductible. Les propriétés suivantes sont équivalentes :*

(1) tous les états sont récurrents positifs ;

(2) il existe un état récurrent positif ;

(3) Q est non explosif et admet une probabilité invariante ν qui charge tous les points.

De plus, quand (3) a lieu, $m_i = 1/(\nu_i q_i)$.

REMARQUE 6.26. La mesure invariante affecte au point i le poids $1/(q_i m_i)$ qui s'interprète comme le temps moyen que l'on passe en i multiplié par l'inverse du temps de retour.

Comme pour les chaînes de Markov à temps discret, il existe une notion plus forte que l'invariance.

DÉFINITION 6.27. On dit que la mesure ν est réversible (ou symétrique) pour le générateur Q si, pour tous $i, j \in E$,

$$\nu_i Q_{ij} = \nu_j Q_{ji}.$$

7. Convergence à l'équilibre et théorème ergodique

THÉORÈME 6.28 (Convergence à l'équilibre). *Soit Q un générateur irréductible non-explosif de semi-groupe $(P(t))_{t\in\mathbb{R}_+}$ et de probabilité invariante λ. Alors, pour tous états i, j,*

$$p_{ij}(t) \xrightarrow[t\to\infty]{} \lambda_j = \frac{1}{q_j m_j}.$$

De plus, pour toute mesure initiale ν, $\mathbb{P}(X_t = j) \xrightarrow[t\to\infty]{} \dfrac{1}{q_j m_j}$.

THÉORÈME 6.29 (Théorème ergodique). *Soit Q un générateur irréductible et ν la loi de probabilité de X_0. Alors*

$$\frac{1}{t}\int_0^t \mathbf{1}_{\{X_s=i\}}\, ds \xrightarrow[t\to\infty]{} \frac{1}{q_i m_i} \quad p.s. \tag{26}$$

De plus, dans le cas récurrent positif, pour toute fonction bornée $f : E \to \mathbb{R}$, on a

$$\frac{1}{t}\int_0^t f(X_s)\, ds \xrightarrow[t\to\infty]{} \overline{f} = \sum_{i\in E} \frac{1}{q_i m_i} f(i) = \sum_{i\in E} \lambda_i f(i) \quad p.s.$$

REMARQUE 6.30. La relation (26) assure que le temps moyen passé en i par la chaîne jusqu'au temps t converge vers le poids que la mesure invariante affecte à i. Ce poids peut être nul si la chaîne n'est pas récurrente positive.

8. Processus de vie et mort

Les processus de vie et mort sur $\mathbb{N}$ forment un exemple important de chaînes de Markov à temps continu. Leur dynamique est résumée par les taux de naissance $(\lambda_i)_{i\in\mathbb{N}}$ et les taux de mort $(\mu_i)_{i\in\mathbb{N}^*}$ (car $\mu_0 = 0$). Partant de l'état $n \in \mathbb{N}$,

— le processus reste en n pendant un temps exponentiel de paramètre $\lambda_n + \mu_n$,

— il saute ensuite vers $n+1$ ou $n-1$ avec probabilités respectives $\lambda_n/(\lambda_n+\mu_n)$ et $\mu_n/(\lambda_n+\mu_n)$.

Pour simuler un tel processus jusqu'à un instant t donné, on utilise exactement l'algorithme ci-dessus.
Si tous les taux de naissance (resp. mort) sont nuls, on parle de processus de mort (resp. naissance) pur.

Bactéries et processus de Yule. Reprenons la modélisation de la colonie de bactéries. Chacune d'elles se divise au taux λ et meurt au taux μ. Ceci revient à considérer le processus de Yule pour lequel les taux de naissance et de mort en n sont respectivement donnés par λn et μn.

On peut également voir ces processus comme la modélisation d'une file d'attente devant un ou plusieurs guichets : une mort correspond au service d'un client tandis qu'une naissance traduit l'arrivée d'un nouveau client dans la file. Citons deux exemples classiques :

EXEMPLE 6.31 (Pile $M/M/1$). On pose $\lambda_n = \lambda$ et $\mu_n = \mu$: les clients arrivent aux temps de saut d'un processus de Poisson d'intensité λ et une seule caisse sert un client après un temps exponentiel de paramètre μ ;

EXEMPLE 6.32 (File $M/M/\infty$). On pose $\lambda_n = \lambda$ et $\mu_n = \mu n$: les clients arrivent aux temps de saut d'un processus de Poisson d'intensité λ et une infinité de caisses servent chacune indépendamment un client après un temps exponentiel de paramètre μ. La figure 1 fournit une trajectoire de ce processus.

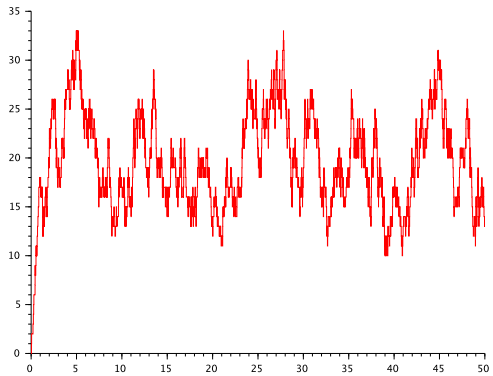

FIGURE 1. Trajectoire de la file $M/M/\infty$ (voir l'exemple 6.32) avec $\lambda = 20$, $\mu = 1$, $X_0 = 0$ jusqu'au temps $t = 50$.

REMARQUE 6.33 (Nomenclature). Les files d'attente dépendent essentiellement de quatre paramètres : la loi d'arrivée des clients, la loi de service de chaque caisse, le nombre de

caisses et (éventuellement) la taille maximale de la file (si elle est pleine, un client arrivant repart sans attendre son tour). Dans les exemples ci-dessus, le M est l'abréviation de memoryless (les lois d'arrivée et de service sont exponentielles).

THÉORÈME 6.34. *Un processus de vie et de mort dont tous les taux de saut sont strictement positifs est irréductible. Si*

$$s = \sum_{i\geq 1}\left(\prod_{l=1}^{i}\frac{\lambda_{l-1}}{\mu_l}\right) < \infty,$$

alors le processus de vie et de mort associé admet une unique mesure de probabilité invariante. Elle est de plus symétrique et est donnée par

$$\nu_0 = \frac{1}{1+s} \quad \textit{et} \quad \nu_i = \frac{1}{1+s}\prod_{l=1}^{i}\frac{\lambda_{l-1}}{\mu_l}.$$

DÉMONSTRATION. Tous les états communiquent donc le processus est nécessairement irréductible. On écrit la relation caractérisant les mesures réversibles pour $j = i+1$:

$$\nu_i Q_{i,i+1} = \nu_{i+1} Q_{i+1,i} \quad \text{ou} \quad \nu_{i+1} = \frac{\lambda_i}{\mu_{i+1}}\nu_i.$$

On a donc

$$\nu_i = \nu_0 \prod_{l=1}^{i}\frac{\lambda_{l-1}}{\mu_l}.$$

On conclut en utilisant le fait que ν est une probabilité. □

9. Exercices

EXERCICE 6.2. Considérons une flotte de N bus. Chaque véhicule tombe en panne indépendamment des autres avec un taux λ et est envoyé au dépôt. L'atelier ne peut en réparer qu'un à la fois et le travail est distribué selon une loi exponentielle de paramètre μ. Quelle est la mesure d'équilibre du nombre de bus en service ?

EXERCICE 6.3 (File $M/M/\infty$). On utilise les notations de l'exemple 6.32. Notons $\nu_{k,t} = \mathscr{L}(X_t|X_0 = k)$, c'est-à-dire que, pour $k \in \mathbb{N}$ et $t \in \mathbb{R}_+$,

$$\forall i \in \mathbb{N}, \quad \nu_{k,t}(i) = \mathbb{P}(X_t = i|X_0 = k).$$

On rappelle deux définitions de la distance en variation totale entre deux mesures de probabilité sur $\mathbb{N}$:

$$d_{\mathrm{VT}}(\mu,\nu) = \frac{1}{2}\sum_{k\in\mathbb{N}}\left|\mu(\{k\}) - \nu(\{k\})\right| = \inf\left\{\mathbb{P}(X \neq Y) \,:\, X \sim \mu,\ Y \sim \nu\right\}.$$

1. Montrer que la file est toujours stable et que sa mesure invariante est une loi de Poisson dont on précisera le paramètre.
2. Soit $k > l$. Montrer que $\mu_{k,t} = \mu_{l,t} * \mathscr{B}(k-l, e^{-\mu t})$. Comment interpréter cette relation en terme de file d'attente ? En déduire que

$$d_{\mathrm{VT}}(\mu_{k,t}, \mu_{l,t}) \leq (k-l)e^{-\mu t}.$$

EXERCICE 6.4 (File $M/M/1$). Notons L la transformée de Laplace du temps d'atteinte de l'état k partant de l'état $k+1$:

$$\forall s \in \mathbb{R}, \quad L(s) = \mathbb{E}\left(e^{sT_k}|X_0 = k+1\right) \quad \text{où} \quad T_k = \inf\{t \in \mathbb{R}_+ \,:\, X_t = k\}.$$

1. Donner une CNS de stabilité et l'expression de la mesure invariante.

2. Pourquoi L ne dépend-elle pas de k ? Montrer que

$$L(s) = \frac{\lambda+\mu}{\lambda+\mu-s}\left(\frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu}L(s)^2\right).$$

En déduire l'expression de L.

10. Corrigés

CORRIGÉ DE L'EXERCICE 6.1. Notons $k_i = \mathbb{E}_i(H^{\{4\}})$. Partant de 1, le processus reste en 1 pour un temps moyen $1/q_1 = 1/2$ puis saute avec des probabilités égales en 2 ou 3. Donc

$$k_1 = \frac{1}{2} + \frac{1}{2}k_2 + \frac{1}{2}k_3.$$

De même

$$k_2 = \frac{1}{6} + \frac{1}{3}k_1 + \frac{1}{3}k_3, \quad k_3 = \frac{1}{9} + \frac{1}{3}k_1 + \frac{1}{3}k_2.$$

La résolution de ce système fournit la valeur $k_1 = 17/12$.

CORRIGÉ DE L'EXERCICE 6.2. Pour $n = 0,\dots,N$, si n bus sont en panne (dont un est en cours de réparation si $n \in \mathbb{N}^*$), $N-n$ sont susceptibles de tomber en panne. Le taux de naissance (arrivée d'un bus dans la file) en n est donc égal à $\lambda_n = (N-n)\lambda$. D'autre part, le taux de mort (réparation d'un bus) est donnée par $\mu_n = \mu\mathbf{1}_{n>0}$. Reste à appliquer le théorème 6.34.

CORRIGÉ DE L'EXERCICE 6.3.
1. Le théorème 6.34 assure que la mesure invariante est la loi de Poisson de paramètre λ/μ.
2. La relation souhaitée découle de $\mathscr{B}(n,p) * \mathscr{B}(m,p) = \mathscr{B}(n+m,p)$. Parmi les k clients initialement présents, choisissons-en l que l'on sépare des autres. Le nombre de ces clients encore présent au temps t est de loi $\mathscr{B}(k-l, e^{-\mu t})$. Le reste du système évolue comme si l personnes étaient présentes initialement. Soit Y et Z deux v.a. indépendantes de lois respectives $\mu_{l,t}$ et $\mathscr{B}(k-l, e^{-\mu t})$. Alors $Y+Z$ suit la loi $\mu_{k,t}$ et

$$\begin{aligned} d_{\mathrm{VT}}(\mu_{k,t}, \mu_{l,t}) &\le \mathbb{P}(Z \neq 0) \\ &= 1 - (1-e^{-\mu t})^{k-l} \le (k-l)e^{-\mu t}. \end{aligned}$$

CORRIGÉ DE L'EXERCICE 6.4.
1. Le théorème 6.34 assure que le processus admet une mesure de probabilité invariante ssi $(\lambda/\mu)^n$ est le terme général d'une série convergence. La file est stable ssi $\lambda < \mu$. La mesure de probabilité invariante est la loi géométrique sur $\mathbb{N}$ de paramètre λ/μ.
2. La transformée de Laplace L ne dépend pas de k puisque les taux de naissance et de mort sont constants. Notons T le premier temps de saut du processus. Si $X_T = k$ alors $T_k = T$ sinon, $X_T = k+2$ et T_k s'écrit comme $T + T' + T''$ où T et T'' sont deux copies indépendantes de T_k. On a donc

$$L(s) = \frac{\mu}{\lambda+\mu}\mathbb{E}\big(e^{sT}\big) + \frac{\lambda}{\lambda+\mu}\mathbb{E}\big(e^{sT}\big)L(s)^2,$$

ce qui donne la formule souhaitée. On en déduit $L(s)$ en résolvant l'équation du second degré.

Chapitre 7

Quelques notions d'estimation paramétrique

L'objet de ce chapitre n'est pas de donner un panorama complet de la théorie de l'estimation paramétrique (méthode des moments, maximum de vraisemblance,...) mais d'en comprendre les fondements et de les mettre en application dans des exemples concrets.

1. Introduction et notations

On appelle modèle paramétrique la donnée $(E^n, (P_\theta)_{\theta\in\Theta})$ où

— E^n est l'ensemble des observations (E sera souvent inclus dans $\mathbb{R}^k$ pour $k \geq 1$),

— Θ est inclus dans $\mathbb{R}^q$, pour $q \geq 1$,

— P_θ est une mesure de probabilité sur E^n pour tout $\theta \in \Theta$.

On notera $\mathbb{E}_\theta$ l'espérance par rapport à la mesure P_θ.
Très souvent dans la suite, P_θ sera de la forme $\mu_\theta^{\otimes n}$. Cette situation correspond à des observations indépendantes et de même loi μ_θ sur E. Cette généralité permet toutefois de couvrir par exemple des problèmes d'estimation à partir de l'observation d'une chaîne de Markov (voir l'exercice 7.5).

EXEMPLE 7.1 (Quelques modèles classiques). Citons par exemple les modèles

— de Bernoulli : $\Theta = [0,1]$ et $P_\theta = \mathscr{B}(\theta)^{\otimes n}$,

— de Poisson : $\Theta = \mathbb{R}_+^*$ et $P_\theta = \mathscr{P}(\theta)^{\otimes n}$,

— exponentiel : $\Theta = \mathbb{R}_+^*$ et $P_\theta = \mathscr{E}(\theta)^{\otimes n}$,

— uniforme : $\Theta = \mathbb{R}_+^*$ et $P_\theta = \mathscr{U}_{[0,\theta]}^{\otimes n}$,

— gaussien : $\Theta = \mathbb{R} \times \mathbb{R}_+^*$ et $P_\theta = \mathscr{N}(m,\sigma^2)^{\otimes n}$ où $\theta = (m,\sigma^2)$.

La problématique est d'estimer une quantité $g(\theta)$ (la fonction g de Θ dans $\mathbb{R}^{k'}$ étant connue) en construisant une fonction de l'échantillon $(X_1,\dots,X_n)$ de loi P_θ.
Une statistique est une fonction mesurable de l'échantillon et un estimateur est une statistique à valeurs dans un ensemble contenant $g(\Theta)$. On notera souvent $\hat{g}_n$ un estimateur de $g(\theta)$.
Il existe plusieurs façons de mesurer les qualités d'un estimateur : certaines sont asymptotiques (quand le nombre n d'observations tend vers l'infini) et s'appuient en général sur des théorèmes limite, d'autres sont non asymptotiques (à n fixé).

DÉFINITION 7.2 (Biais). Le biais de l'estimateur $\hat{g}_n$ est la fonction $\theta \mapsto \mathbb{E}_\theta(\hat{g}_n) - g(\theta)$. L'estimateur est dit sans biais si cette fonction est nulle. Il est dit asymptotiquement sans biais si pour tout $\theta \in \Theta$, $\mathbb{E}_\theta(\hat{g}_n)$ tend vers $g(\theta)$.

DÉFINITION 7.3 (Erreur quadratique). L'erreur quadratique de l'estimateur $\hat{g}_n$ est définie par

$$\delta(\hat{g}_n) = \mathbb{E}_\theta\Big[\big(\hat{g}_n - g(\theta)\big)^2\Big].$$

Cette erreur s'exprime en fonction de la décomposition classique dite de biais-variance :

$$\delta(\hat{g}_n) = \mathbb{V}_\theta(\hat{g}_n) + \big(\mathbb{E}_\theta(\hat{g}_n) - g(\theta)\big)^2.$$

On voit donc bien ici que deux sources d'erreurs se superposent et les estimateurs peuvent être mauvais de plusieurs façons. On peut par exemple penser à deux joueurs de fléchettes : l'un groupant toutes ses fléchettes mais à côté de la cible (petite variance et gros biais), l'autre les éparpillant mais de manière centrée (petit biais, grosse variance).

DÉFINITION 7.4 (Consistance). On dit que $\hat{g}_n$ est consistant (resp. fortement consistant) si, pour tout $\theta \in \Theta$, $\hat{g}_n$ converge vers $g(\theta)$ en probabilité (resp. presque sûrement).

2. Méthode des moments

L'idée est très simple. Plaçons-nous dans le cas d'un échantillon de variables aléatoires réelles indépendantes de même loi μ_θ. On sait, en vertu de la loi des grands nombres, que

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow[n\to\infty]{\text{p.s.}} \mathbb{E}_\theta(X_1^k)$$

sous réserve que μ_θ admette un moment d'ordre k. Ainsi, si la quantité $g(\theta)$ à estimer s'écrit comme une fonction des k premiers moments de μ_θ, on peut obtenir en substituant les moments par les moments empiriques un estimateur de $g(\theta)$.

EXEMPLE 7.5 (Lois exponentielles). Si $\mu_\theta = \mathscr{E}(\theta)$ alors le premier moment de μ_θ vaut $1/\theta$. La méthode des moments suggère d'estimer θ à partir de $X_1, \dots, X_n$ par

$$\hat{\theta}_n = \frac{n}{\sum_{i=1}^n X_i}.$$

Remarquons que le second moment de μ_θ vaut $2/\theta^2$. On pourrait donc également proposer l'estimateur suivant :

$$\hat{\theta}'_n = \sqrt{\frac{2n}{\sum_{i=1}^n X_i^2}}.$$

EXERCICE 7.1. On se place dans le cadre de l'exemple 7.5. Montrer que $\hat{\theta}_n$ est nécessairement biaisé. En utilisant que $X_1 + \cdots + X_n$ suit une loi Gamma, calculer le biais, la variance et l'erreur quadratique de $\hat{\theta}_n$. Montrer que $\hat{\theta}_n$ est fortement consistant et donner un résultat de convergence en loi pour la suite $(\hat{\theta}_n - \theta)_{n\geq 0}$ convenablement renormalisée.

3. Estimateur du maximum de vraisemblance

Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes et de même loi μ_θ où $\theta \in \Theta \subset \mathbb{R}^d$ est un paramètre inconnu. On suppose que toutes les mesures $(\mu_\theta)_{\theta\in\Theta}$ sont absolument continues par rapport à une mesure de référence μ, et de densités $(f_\theta)_{\theta\in\Theta}$. La vraisemblance de l'échantillon $X_1, \dots, X_n$ est la fonction

$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n f_\theta(x_i).$$

La probabilité que $(X_1, \dots, X_n)$ soit "proche" de $(x_1, \dots, x_n)$ est de l'ordre de

$$L(\theta; x_1, \dots, x_n) \prod_{i=1}^n \mu(dx_i).$$

L'idée clé est donc de chercher un paramètre θ où la fonction $\theta \mapsto L(\theta; x_1, \dots, x_n)$ atteint son maximum. Ces points dépendent bien entendu de $(x_1, \dots, x_n)$. On appelle estimateur du maximum de vraisemblance de θ

$$\hat{\theta}_{\mathrm{MV}} = \mathrm{argmax}_{\theta \in \Theta} L(\theta; X_1, \dots, X_n).$$

REMARQUE 7.6. La notation $\mathrm{argmax}_{x \in X} f(x)$ désigne la valeur de x en laquelle f atteint son maximum (si elle existe et est unique).

EXEMPLE 7.7. $\Theta =]0, +\infty[$ et $\mu_\theta = \mathscr{E}(\theta)$ pour $\theta \in \Theta$. Si μ est la mesure de Lebesgue sur $[0, +\infty[$,

$$L(\theta; X_1, \dots, X_n) = \prod_{i=1}^{n} \theta e^{-\theta X_i} \quad \text{et} \quad \hat{\theta}_{\mathrm{MV}} = \frac{n}{\sum_{i=1}^{n} X_i}.$$

Remarquons que l'on retrouve l'estimateur obtenu par la méthode des moments (voir l'exemple 7.5).

EXEMPLE 7.8. $\Theta =]0, +\infty[$ et $\mu_\theta = \mathscr{U}_{[0,\theta]}$ pour $\theta \in \Theta$. Si μ est la mesure de Lebesgue sur $[0, +\infty[$,

$$L(\theta; X_1, \dots, X_n) = \prod_{i=1}^{n} \frac{1}{\theta} \mathbf{1}_{[0,\theta]}(X_i) = \frac{1}{\theta^n} \mathbf{1}_{[\max_{1 \le i \le n} X_i, +\infty[}(\theta), \quad \text{et} \quad \hat{\theta}_{\mathrm{MV}} = \max_{1 \le i \le n} X_i.$$

Cette fois-ci, l'estimateur obtenu n'est pas le même que celui donné par la méthode des moments.

EXEMPLE 7.9. $\Theta =]0, +\infty[$ et $\mu_\theta = \mathscr{P}(\theta)$ pour $\theta \in \Theta$. Si μ est la mesure de comptage sur $\mathbb{N}$,

$$L(\theta; X_1, \dots, X_n) = \prod_{i=1}^{n} e^{-\theta} \frac{\theta^{X_i}}{X_i!} \quad \text{et} \quad \hat{\theta}_{\mathrm{MV}} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

EXEMPLE 7.10. $\Theta = \mathbb{R} \times]0, +\infty[$ et $\mu_\theta = \mathscr{N}(m, v)$ pour $\theta = (m, v) \in \Theta$.

$$L(\theta; X_1, \dots, X_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi v}} e^{-(X_i - m)^2/(2v)}$$

et

$$\hat{m}_{\mathrm{MV}} = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad \hat{v}_{\mathrm{MV}} = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \hat{m}_{\mathrm{MV}}^2.$$

REMARQUE 7.11 (Log-vraisemblance). Dans la plupart des exemples ci-dessus, il est plus aisé de chercher le(s) point(s) où la log-vraisemblance $\theta \mapsto \ln(L(\theta; X_1, \dots, X_n))$ atteint un maximum.

REMARQUE 7.12 (Recherche de maxima). Soit Θ un sous-ensemble de $\mathbb{R}^d$ et $f : \Theta \to \mathbb{R}$ une fonction de classe $\mathscr{C}^2$. La fonction f peut atteindre son maximum au bord de Θ ou en des points de son intérieur. Dans le second cas, ces points sont critiques : ∇f y est nul. Pour assurer que ces points sont des maxima, on peut utiliser une condition de second ordre (en montrant que la matrice hessienne est définie négative).

REMARQUE 7.13 (Changement bijectif de paramétrage). Un des avantages de l'estimateur du maximum de vraisemblance est sa robustesse par rapport à un changement de paramètre. Si φ est un $\mathscr{C}^1$ difféomorphisme de Θ dans Θ' et que l'estimateur du maximum de vraisemblance de θ est $\hat{\theta}_{\mathrm{MV}}$ alors celui de θ' est $\varphi(\hat{\theta}_{\mathrm{MV}})$.

REMARQUE 7.14 (Comportement asymptotique). Il existe des théorèmes généraux décrivant le comportement asymptotique de l'estimateur du maximum de vraisemblance. Il est notamment possible d'établir par exemple qu'il est consistant, asymptotiquement sans

biais, asymptotiquement de variance minimum et parfois que ses fluctuations sont gaussiennes. Toutefois, ces résultats dépassent le programme de l'agrégation et les propriétés de l'estimateur du maximum de vraisemblance peuvent en général être obtenues par des théorèmes généraux (loi des grands nombres, théorème limite central, etc.).

4. Exercices

EXERCICE 7.2 (Lois uniformes). On considère le cadre de l'exemple 7.8.
1. Déterminer l'estimateur obtenu par la méthode des moments (ou du premier moment). Décrire les propriétés de cet estimateur (biais, variance, comportement asymptotique).
2. Montrer que l'estimateur du maximum de vraisemblance est donné par

$$\hat{\theta}_{\mathrm{MV}} = \max_{1\le i\le n} X_i.$$

Déterminer sa loi, son biais et son erreur quadratique. En déduire un estimateur sans biais. Quelle est son erreur quadratique ? Quelle conclusion en tirer ? Décrire les propriétés asymptotiques de l'estimateur du maximum de vraisemblance.
3. Quel estimateur semble le plus performant ?

EXERCICE 7.3 (Lois puissances). Soit $(X_n)_{n\ge1}$ une suite de v.a. i.i.d. de densité

$$f_\theta(x) = (1+\theta)x^\theta \mathbf{1}_{[0,1]},$$

avec θ inconnu.
1. Quelles sont les valeurs admissibles pour θ ?
2. Déterminer l'estimateur $\hat{\theta}_n$ du maximum de vraisemblance pour θ construit à partir des v.a. $(X_i)_{1\le i\le n}$.

EXERCICE 7.4 (Estimation d'un coefficient de mélange). On considère n observations indépendantes $(X_1,\dots,X_n)$ de loi $p\mathcal{U}_{[0,a]} + (1-p)\mathcal{U}_{[0,b]}$, c'est-à-dire que X_i suit la loi uniforme sur $[0,a]$ avec probabilité p et la loi uniforme sur $[0,b]$ avec probabilité $1-p$. On suppose que $a<b$, avec a et b connus.
1. Déterminez la fonction de répartition et la densité de X_1.
2. Soit N_a la variable aléatoire égale au nombre d'individus X_i compris entre 0 et a. Montrez que la vraisemblance peut s'écrire :

$$L(p;x_1,\dots,x_n) = \left(\frac{p}{a}+\frac{1-p}{b}\right)^{n_a}\left(\frac{1-p}{b}\right)^{n-n_a}$$

où $n_a = n_a(x_1,\dots,x_n)$ est le nombre d'observations de l'échantillon comprises entre 0 et a.
3. En déduire l'équation de vraisemblance relative au paramètre p et déterminez l'estimateur du maximum de vraisemblance $\hat{p}$ du paramètre p en fonction de la variable aléatoire N_a.
4. Calculez le biais de l'estimateur $\hat{p}$.
5. Calculez $\mathbb{E}_p(X_1)$. Déterminez l'estimateur p^* du paramètre p obtenu par la méthode des moments basée sur le moment d'ordre 1. Calculez le biais de cet estimateur.
6. Des deux estimateurs précédents, lequel est le meilleur en terme d'erreur quadratique moyenne ?

EXERCICE 7.5 (Chaîne de Markov). On considère une chaîne de Markov X à valeurs dans $E=\{1,2,\dots,k\}$ de matrice de transition P inconnue.
1. Sous quelles hypothèses peut-on espérer pouvoir estimer la matrice de transition P à partir de l'observation d'une trajectoire de X ?
2. Déterminer l'estimateur du maximum de vraisemblance de P.
3. Quelles sont ses propriétés asymptotiques ?

5. Corrigés

CORRIGÉ DE L'EXERCICE 7.1. La méthodes des moments et le maximum de vraisemblance fournissent tous les deux l'estimateur $\hat{\theta}_n = 1/\bar{X}_n$. Remarquons que $\mathbb{E}_\theta(\bar{X}_n) = 1/\theta$. Or, si X est une v.a. (strictement) positive non constante,

$$\mathbb{E}\left(\frac{1}{X}\right) > \frac{1}{\mathbb{E}(X)}$$

en vertu de l'inégalité de Jensen pour la fonction strictement convexe $x \mapsto 1/x$ sur $\mathbb{R}_+^*$. Ainsi, $\mathbb{E}_\theta(\hat{\theta}_n) > \theta$: l'estimateur $\hat{\theta}_n$ a un biais strictement positif.
La somme $X_1 + \cdots + X_n$ suit une loi Gamma de paramètre (n, θ) donc

$$\mathbb{E}_\theta(\hat{\theta}_n) = \int_{\mathbb{R}_+^*} \frac{n}{s} \frac{\theta^n s^{n-1}}{\Gamma(n)} e^{-\theta s}\, ds = \frac{n\theta}{n-1} \int_{\mathbb{R}_+^*} \frac{\theta^{n-1} s^{n-2}}{\Gamma(n-1)} e^{-\theta s}\, ds = \frac{n\theta}{n-1}.$$

Le biais de $\hat{\theta}_n$ vaut donc $\theta/(n-1)$. L'estimateur $(n-1)/(X_1 + \cdots + X_n)$ est donc sans biais. De même,

$$\mathbb{E}_\theta(\hat{\theta}_n^2) = \int_{\mathbb{R}_+^*} \frac{n^2}{s^2} \frac{\theta^n s^{n-1}}{\Gamma(n)} e^{-\theta s}\, ds = \frac{n^2\theta^2}{(n-1)(n-2)} \int_{\mathbb{R}_+^*} \frac{\theta^{n-2} s^{n-3}}{\Gamma(n-2)} e^{-\theta s}\, ds = \frac{n^2\theta^2}{(n-1)(n-2)}.$$

On a donc

$$\mathbb{V}_\theta(\hat{\theta}_n) = \frac{n^2\theta^2}{(n-1)^2(n-2)}.$$

D'après la loi forte des grands nombres, $\bar{X}_n$ converge vers $\mathbb{E}_\theta(X_1)$ presque sûrement. L'estimateur $\hat{\theta}_n$ est donc fortement consistant. D'après le théorème 9.9, $\sqrt{n}(\hat{\theta}_n - \theta)$ converge en loi vers la distribution gaussienne $\mathcal{N}(0, 1/\theta^2)$ (voir l'exemple 9.10).

CORRIGÉ DE L'EXERCICE 7.2.
1. Puisque $\mathbb{E}_\theta(X_1) = \theta/2$, la méthode des moments suggère d'estimer θ par $\hat{\theta}_n = 2\bar{X}_n$. Cet estimateur est sans biais. Sa variance est donnée par

$$\mathbb{V}_\theta(\hat{\theta}_n) = 4\mathbb{V}_\theta(\bar{X}_n) = \frac{4}{n}\mathbb{V}_\theta(X_1) = \frac{\theta^2}{3n}.$$

Il est fortement consistant et asymptotiquement normal :

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n\to\infty]{\mathscr{L}} \mathcal{N}(0, \theta^2/3).$$

2. Voir l'exemple 7.8 pour le calcul de $\hat{\theta}_{\mathrm{MV}}$. Pour calculer sa loi, on détermine sa fonction de répartition : pour tout $t \in [0, \theta]$,

$$\begin{aligned}\mathbb{P}_\theta(\hat{\theta}_{\mathrm{MV}} \le t) &= \mathbb{P}_\theta(X_1 \le t, \ldots, X_n \le t)\\ &= \mathbb{P}_\theta(X_1 \le t)^n = \frac{t^n}{\theta^n}.\end{aligned}$$

On en déduit que la loi de $\hat{\theta}_{\mathrm{MV}}$ admet pour densité la fonction

$$f_\theta : [0, +\infty[\mapsto \frac{n}{\theta^n} t^{n-1} \mathbf{1}_{[0,\theta]}(t).$$

L'estimateur est nécessairement biaisé puisqu'il est presque sûrement inférieur à θ. Plus précisément,

$$\mathbb{E}_\theta(\hat{\theta}_{\mathrm{MV}}) = \int_0^\theta x f_\theta(x)\, dx = \frac{n}{n+1}\theta.$$

Le biais vaut donc $\theta/(n+1)$. De plus,

$$\mathbb{E}_\theta(\hat{\theta}_{\mathrm{MV}}^2) = \frac{n}{n+2}\theta^2 \quad \text{et} \quad \mathbb{V}_\theta(\hat{\theta}_{\mathrm{MV}}) = \frac{n}{(n+2)(n+1)^2}$$

L'erreur quadratique de $\hat{\theta}_{\mathrm{MV}}$ est donc de l'ordre de $1/n^2$, ce qui est bien mieux que celle de $\hat{\theta}_n$.
Pour tout $\varepsilon \in]0,\theta[$,

$$\mathbb{P}_\theta\left(\left|\hat{\theta}_{\mathrm{MV}}-\theta\right| \le \varepsilon\right) = \mathbb{P}_\theta\left(\hat{\theta}_{\mathrm{MV}} \le \theta-\varepsilon\right) = \left(1-\frac{\varepsilon}{\theta}\right)^n.$$

On en déduit donc que $\hat{\theta}_{\mathrm{MV}}$ converge presque sûrement vers θ : l'estimateur est fortement consistant.
De plus, pour tout $t > 0$,

$$\mathbb{P}_\theta((n/\theta)(\theta-\hat{\theta}_{\mathrm{MV}}) > t) = \mathbb{P}_\theta(\hat{\theta}_{\mathrm{MV}} \le \theta - \theta t/n) = \left(1-\frac{t}{n}\right)^n \xrightarrow[n\to\infty]{} e^{-t}.$$

En d'autres termes,

$$\frac{n}{\theta}(\theta-\hat{\theta}_{\mathrm{MV}}) \xrightarrow[n\to\infty]{\mathscr{L}} \mathscr{E}(1).$$

Dans cet exemple, l'estimateur du maximum de vraisemblance converge en loi à vitesse n (et non $\sqrt{n}$ comme $\hat{\theta}_n$) et la loi limite n'est plus une loi gaussienne. On obtient alors un intervalle de confiance dont l'amplitude est de l'ordre de $1/n$ (et non plus $1/\sqrt{n}$).
3. L'estimateur du maximum de vraisemblance est plus performant que celui issu de la méthode des moments (voir la figure 1).

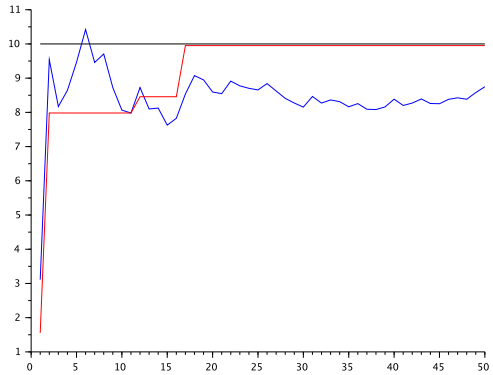

FIGURE 1. Les estimateurs $\hat{\theta}_n$ (en bleu) et $\hat{\theta}_{\mathrm{MV}}$ (en rouge) en fonction de n pour l'exercice 7.2.

CORRIGÉ DE L'EXERCICE 7.3.
1. La fonction f_θ est une densité de probabilité ssi $\theta > -1$.
2. La log-vraisemblance de l'échantillon est donnée par

$$\log L(\theta; X_1,\dots,X_n) = n\ln(1+\theta) + \theta\sum_{i=1}^n \ln(X_i).$$

On a donc

$$\hat{\theta}_{\mathrm{MV}} = \frac{n}{\sum_{i=1}^n \ln(1/X_i)} - 1.$$

CORRIGÉ DE L'EXERCICE 7.4. On considère n observations indépendantes $(X_1,\dots,X_n)$ de loi $p\mathscr{U}_{[0,a]} + (1-p)\mathscr{U}_{[0,b]}$, c'est-à-dire que X_i suit la loi uniforme sur $[0,a]$ avec probabilité p et la loi uniforme sur $[0,b]$ avec probabilité $1-p$. On suppose que $a < b$, avec a et b connus.

1. La fonction de répartition F_p et la densité f_p de X_1 sont données par

$$F_p(x) = \begin{cases} 0 & \text{si } x < 0, \\ px/a + (1-p)x/b & \text{si } x \in [0, a[, \\ p + (1-p)x/b & \text{si } x \in [a, b[, \\ 1 & \text{si } x \geq b, \end{cases}$$

et

$$f_p(x) = \begin{cases} 0 & \text{si } x < 0, \\ p/a + (1-p)/b & \text{si } x \in [0, a[, \\ (1-p)/b & \text{si } x \in [a, b[, \\ 0 & \text{si } x \geq b. \end{cases}$$

2. D'après l'expression de f_p on obtient bien l'expression de la vraisemblance souhaitée en regroupant les termes égaux.
3. Après calcul, on obtient

$$\hat{p} = \frac{b}{b-a}\frac{N_a}{n} - \frac{a}{b-a}.$$

4. On a

$$\mathbb{P}_p(X_1 \leq a) = p + (1-p)\frac{a}{b} \quad \text{et ainsi} \quad \mathbb{E}_p(\hat{p}) = p.$$

L'estimateur $\hat{p}$ est sans biais.
5. Par construction, $\mathbb{E}_p(X) = pa/2 + (1-p)b/2$. On pose donc

$$p^* = \frac{1}{b-a}\big(b - 2\bar{X}_n\big).$$

Cet estimateur est également sans biais.
6. Reste à comparer les variances (puisque les biais sont nuls).

CORRIGÉ DE L'EXERCICE 7.5.
1. Il faut supposer que la chaîne explore tout l'espace, c'est-à-dire que P est irréductible.
2. Pour $x_1, \ldots, x_n \in E$,

$$\mathbb{P}(X_1 = x_1, \ldots X_n = x_n) = \prod_{i=1}^{n} P(x_i, x_{i+1}) = \prod_{x,y \in E} P(x,y)^{n_{xy}},$$

où $n_{xy} = \sum_{i=1}^{n} \mathbf{1}_{\{x_i = x, x_{i+1} = y\}}$. Les paramètres sont les coefficients de P contraints à former une matrice markovienne. Après optimisation, on trouve

$$\hat{P}(x,y) = \frac{N_{xy}}{N_x} \quad \text{où} \quad N_x = \sum_{i=1}^{n} \mathbf{1}_{\{X_i = x\}}, \quad N_{xy} = \sum_{i=1}^{n} \mathbf{1}_{\{X_i = x, X_{i+1} = y\}}.$$

Ce résultat est logique puisque l'on estime $P(x,y)$ par la proportion de saut de x vers y parmi tous les passages de la chaîne en x.
3. Le théorème ergodique (voir le théorème 4.37) et le théorème limite central associé permettent de montrer que cet estimateur est fortement consistant et asymptotiquement gaussien.

Chapitre 8

Quantiles empiriques

Dans ce chapitre, μ est une mesure de probabilité sur $\mathbb{R}$ de fonction de répartition F et $(X_n)_{n\geq 0}$ une suite de v.a. i.i.d. de loi μ. Rappelons deux théorèmes très importants pour la statistique puisqu'ils assurent essentiellement qu'un échantillon infini dénombrable caractérise la mesure dont il est issu. Ils sont établis dans le chapitre La loi des grands nombres du fascicule de probabilités.

1. Théorème de Glivenko-Cantelli

DÉFINITION 8.1. La mesure empirique d'un échantillon $X_1, \dots, X_n$ est définie par

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}.$$

La fonction de répartition empirique d'un échantillon $X_1, \dots, X_n$ est définie par

$$\forall t \in \mathbb{R}, \quad F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq t\}}.$$

REMARQUE 8.2 (Calcul de la fonction de répartition empirique). La fonction de répartition empirique d'un échantillon $X_1, \dots, X_n$ est donnée par

$$F_n(x) = \begin{cases} 0 & \text{si } x < X_{(1)}, \\ \dfrac{i}{n} & \text{si } x \in [i/n;\ (i+1)/n[\text{ avec } i = 1, \dots, n-1, \\ 1 & \text{si } x \geq X_{(n)}. \end{cases}$$

La commande Scilab pour tracer la fonction de répartition empirique d'un vecteur X de taille n est donc

```
plot2d2([1/n : 1/n : 1], −gsort(−X))
```

EXERCICE 8.1. Confronter la fonction de répartition empirique d'un échantillon à la fonction de répartition de sa loi. On pourra consulter l'aide pour les fonctions Scilab cdfnor, cdfchi, etc.

THÉORÈME 8.3 (Glivenko-Cantelli). *Presque sûrement, la suite des fonctions de répartition empiriques converge uniformément vers la fonction de répartition de μ :*

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = \|F_n - F\|_\infty \xrightarrow[n\to\infty]{p.s.} 0.$$

Dans le même ordre d'idée, on peut obtenir le résultat suivant, qui reste valable sur $\mathbb{R}^d$.

THÉORÈME 8.4 (Th. fondamental de la statistique). *Avec probabilité 1, la suite $(\mu_n)_{n\geq 1}$ converge étroitement vers μ. Plus précisément, presque sûrement, pour toute fonction continue bornée,*

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) = \int_{\mathbb{R}} f\, d\mu_n \xrightarrow[n\to\infty]{} \int_{\mathbb{R}} f\, d\mu.$$

2. Convergence des quantiles empiriques

On associe à F son inverse généralisée F^- définie sur $[0,1]$ par

$$\forall p \in [0,1], \quad F^-(p) = \inf\{x \in \mathbb{R}\,;\, F(x) \geq p\}.$$

PROPOSITION 8.5. *Si U est une variable aléatoire de loi uniforme sur $[0,1]$ alors $F^-(U)$ est une variable aléatoire de fonction de répartition F.*
Réciproquement, si F est continue, $F(X)$ suit la loi uniforme sur $[0,1]$

REMARQUE 8.6. La première partie de la proposition est parfois appelé théorème fondamental de la simulation car il établit que toute mesure sur $\mathbb{R}$ peut être simulée (en théorie) à partir d'une variable aléatoire de loi uniforme. La réciproque est fausse si F n'est pas continue : si une mesure μ possède un atome alors c'est encore le cas pour toute mesure image de μ.

DÉFINITION 8.7. Un quantile d'ordre p est un réel x tel que $\mathbb{P}(X \leq x) \geq p$ et $\mathbb{P}(X \geq x) \geq 1-p$. Le réel $x_p := F^-(p)$ est un quantile d'ordre p. Pour $p = 1/2$, on parle de médiane. Pour $p \in \{1/4,\ 1/2,\ 3/4\}$ on parle de quartiles.

EXEMPLE 8.8. Pour la loi de Bernoulli de paramètre $1/2$ (resp $3/4$), l'ensemble des médianes est l'intervalle $[0,1]$ (resp. le singleton $\{1\}$). La loi exponentielle de paramètre λ admet une unique médiane. Elle vaut $\ln(2)/\lambda$.

EXERCICE 8.2. À quelle condition le quantile d'ordre p est-il unique ? Déterminer la médiane de la loi gaussienne $\mathcal{N}(m,\sigma^2)$. Connaît-on explicitement d'autres quantiles de cette loi ?

On suppose dans toute la suite que F est continue. Soit $X_1,\dots,X_n$ un échantillon de loi μ. Presque sûrement, il existe une permutation (aléatoire) σ telle que

$$X_{\sigma(1)} < X_{\sigma(2)} < \cdots < X_{\sigma(n)}.$$

On note $(X_{(1)}, X_{(2)},\dots,X_{(n)})$ l'échantillon réordonné (dans l'ordre croissant). En particulier, $X_{(1)}$ est la plus petite valeur de l'échantillon. On pourra se reporter au chapitre 5 pour les propriétés importantes des statistiques d'ordre.
On note $Q_{p,n} = X_{([np]+1)}$ le quantile empirique d'ordre p.

THÉORÈME 8.9 (Convergence p.s. des quantiles empiriques). *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a.r. i.i.d. de fonction de répartition continue. Alors*

$$X_{([nq]/n)} \xrightarrow[n\to\infty]{} x_q \quad p.s.$$

DÉMONSTRATION. On utilise le théorème de Glivenko-Cantelli :

$$\begin{aligned}\left|F(Q_{p,n}) - F(x_p)\right| &\leq \left|F(Q_{p,n}) - F_n(Q_{p,n})\right| + \left|F_n(Q_{p,n}) - F(x_p)\right| \\ &\leq \|F_n - F\|_\infty + \left|([np]+1)/n - p\right| \xrightarrow[n\to\infty]{p.s.} 0.\end{aligned}$$

On utilise ensuite la continuité de F^- en x_p pour obtenir le résultat. □

THÉORÈME 8.10 (Convergence en loi des quantiles empiriques). *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a.r. i.i.d. de fonction de répartition continue et de densité f. Alors, pour $q \in]0,1[$,*

$$\sqrt{n}\big(X_{([nq]/n)} - x_q\big) \xrightarrow[n\to\infty]{\mathscr{L}} \mathcal{N}\left(0, \frac{q(1-q)}{f(x_q)^2}\right).$$

DÉMONSTRATION. On utilise l'expression de la densité de la loi du quantile empirique 5.6 et la formule de Stirling. □

REMARQUE 8.11. Soit $(X_n)_{n\geq 1}$ une suite de v.a. i.i.d. de loi $\mathscr{E}(\lambda)$. Alors la loi de $X_{(1,n)}$ est la loi $\mathscr{E}(n\lambda)$ et $nX_{(1,n)}$ suit la loi $\mathscr{E}(\lambda)$. Le théorème 8.10 est donc faux pour les quantiles extrêmes (*i.e.* pour $q = 0$ ou $q = 1$).

CONSEILS BIBLIOS 8.12. Pour la convergence en loi, on pourra se référer à [**CGCDM99**] pour un exemple et [**Tas85**] pour le résultat.

EXERCICE 8.3 (Convergence en loi de la médiane). Soit μ une probabilité admettant une densité f strictement positive. On note m la médiane de μ. On suppose pour simplifier que la taille n de l'échantillon observé est impaire et on l'écrit $n = 2p-1$.
1. Déterminer la densité de $\sqrt{2p-1}(X_{(p)} - m)$.
2. En déduire que $\sqrt{2p-1}(X_{(p)} - m)$ converge en loi vers une loi normale dont on précisera les paramètres.

EXERCICE 8.4 (Moyenne contre médiane). On cherche à estimer θ dans les trois modèles suivants :

$$f_\theta(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-\theta)^2/2} \quad g_\theta(x) = \frac{1}{2} e^{-|x-\theta|} \quad \text{et} \quad h_\theta(x) = \frac{1}{\pi} \frac{1}{1+(x-\theta)^2}.$$

Comparer les méthodes basées sur la méthode des moments et celle basée sur l'estimation de la médiane empirique.

3. Théorème de Kolmogorov-Smirnov

Le théorème de Glivenko-Cantelli fournit la convergence presque sûre uniforme de la suite des fonctions de répartition empiriques. Le résultat suivant fournit une vitesse sous une hypothèse supplémentaire..

THÉORÈME 8.13 (Kolmogorov-Smirnov). *Si F est continue alors*

$$D_n = \sqrt{n}\|F_n - F\|_\infty \xrightarrow[n\to\infty]{\mathscr{L}} \mu_{KS},$$

où μ_{KS} est la loi de Kolmogorov-Smirnov caractérisée par exemple par sa fonction de répartition :

$$\forall t > 0, \quad F_{KS}(t) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2t^2} = 1 + 2\sum_{k=1}^{+\infty} (-1)^k e^{-2k^2t^2}.$$

REMARQUE 8.14 (fondamentale). Puisque F est continue et croissante et que F_n est constante par morceaux, le supremum ne peut être atteint qu'au voisinage points de discontinuité de F_n, c'est-à-dire aux points $(X_i)_{1\leq i\leq n}$. Plus précisément,

$$\|F_n - F\|_\infty = \max_{1\leq i\leq n}\left(\left|\frac{i-1}{n} - F(X_{(i)})\right|, \left|\frac{i}{n} - F(X_{(i)})\right|\right).$$

De plus, puisque F est continue, $F(X_i)$ suit la loi uniforme sur $[0,1]$ et, comme F est croissante, $(F(X_{(i)}))_{1\leq i\leq n}$ a même loi qu'un échantillon réordonné de v.a. i.i.d. de loi uniforme sur $[0,1]$. Ainsi, D_n a même loi que

$$\sqrt{n}\max_{1\leq i\leq n}\left(\left|\frac{i-1}{n} - U_{(i)}\right|, \left|\frac{i}{n} - U_{(i)}\right|\right).$$

La loi de D_n ne dépend donc pas de μ, on dit que c'est une statistique libre.

Il existe de nombreux théorèmes du type de celui de Kolmogorov-Smirnov. Citons celui de Cràmer-Von Mises.

THÉORÈME 8.15 (Crámer-Von Mises). *Si F est continue alors*

$$nI_n^2 = n\int_{\mathbb{R}} (F_n(x) - F(x))^2 \, dF(x) \xrightarrow[n\to\infty]{\mathscr{L}} \mu_{CVM},$$

où μ_{CVM} est une loi tabulée.

EXERCICE 8.5. Montrer, en découpant l'intégrale sur les intervalles de la forme $[X_{(i)}, X_{(i+1)}]$, que

$$nI_n^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left(\frac{2i-1}{2n} - F(X_{(i)}) \right)^2.$$

En déduire que c'est encore une statistique libre.

REMARQUE 8.16. Quelques quantiles des lois de Kolmogorov-Smirnov et Cràmer-Von Mises :

Loi \ Ordre	0.9	0.95	0.99
μ_{KS}	1.212	1.346	1.615
μ_{CVM}	0.347	0.46	0.734

REMARQUE 8.17 (Une idée de preuve pour ces théorèmes). La preuve esquissée ci-dessous n'est pas la preuve originale de Kolmogorov mais celle, beaucoup plus élégante, de Doob. On pourra consulter [**Fis63**] pour de plus amples informations (mais c'est hors programme...).
Pour tout $x \in [0,1]$, on pose $H_n(x) = \sqrt{n}(F_n(x) - x)$. La fonction (aléatoire) H_n est nulle en 0 et 1 et a n sauts de hauteur $1/n$. En vertu du TCL multidimensionnel, pour $0 \le x_1 < x_2 < \cdots < x_k < 1$, la suite de vecteurs aléatoires $((H_n(x_1), \ldots, H_n(x_k)))_n$ converge en loi vers un vecteur gaussien centré de matrice de covariance (symétrique) K où $K_{ij} = x_i(1-x_j)$ pour $i \le j$. On peut en déduire la convergence de la suite de processus aléatoire $(H_n)_n$ vers un processus H appelé pont brownien. Il ne reste plus qu'à montrer que le supremum et l'intégrale sont conservés par passage à la limite...

4. Tests associés

Ces théorèmes permettent de mettre en place des tests d'adéquation à une loi de fonction de répartition continue (comme les test du χ^2 le font pour des lois sur un ensemble fini). On formalise le problème de la manière suivante. Soit F^0 une fonction de répartition continue. On veut tester

$$H_0 \,:\, F = F^0 \quad \text{contre} \quad H_1 \,:\, F \neq F^0.$$

Soit $A_n = \sqrt{n}\left\|F_n - F^0\right\|_\infty$. Sous H_0, $F = F^0$ donc la loi de A_n converge vers la loi de Kolmogorov-Smirnov. Sous H_1, $\left\|F_n - F^0\right\|_\infty$ converge presque sûrement vers $\left\|F - F^0\right\|_\infty 0$ qui est strictement positif et A_n converge vers $+\infty$ p.s.
Comme pour le test du χ^2, on choisit donc une région de rejet de la forme $[k_\alpha, +\infty[$ où α est le niveau du test et k_α est le quantile d'ordre $1-\alpha$ de la loi de Kolmogorov-Smirnov.
On peut faire exactement le même raisonnement pour un construire un test d'adéquation à partir du théorème de Crámer-Von Mises.

EXERCICE 8.6. **1.** On veut illustrer le fonctionnement du test de Kolmogorov-Smirnov pour tester l'hypothèse qu'un échantillon de taille n suit la loi exponentielle de paramètre 1. Écrire successivement trois fonctions Scilab executant respectivement les trois tâches suivantes :

(1) prend en entrée la taille de l'échantillon et un paramètre de perturbation e, simule un échantillon de loi Gamma de paramètre $(1+e, 1)$ et renvoie 1 si l'on ne rejette pas H_0 et 0 sinon au niveau 0.95 ;

(2) effectue N fois la procédure ci-dessus et retourne le pourcentage de non-rejet ;

(3) effectue la procédure ci-dessus pour e variant autour de 0 et tracer la puissance en fonction de e (voir la figure 1).

2. Faire la même chose avec le test de Crámer-Von Mises.

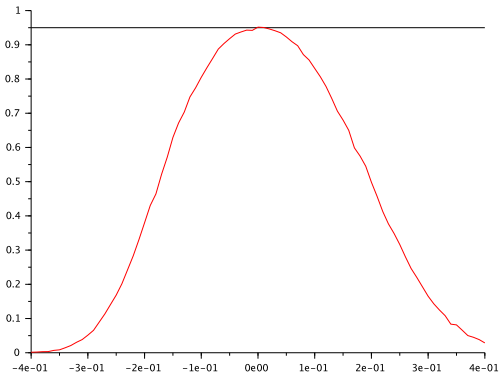

FIGURE 1. Probabilité de ne pas rejeter H_0 pour l'exercice 8.6 avec $N = 10000$, $n = 100$ et $\varepsilon \in [-0.4; 0.4]$ au niveau 0.05.

EXERCICE 8.7 (Tests d'exponentialité). Les lois exponentielles apparaissent naturellement en modélisation. Il est important de pouvoir tester l'adéquation à une loi (ou une famille de lois) exponentielle(s).
1. Comment tester sur un échantillon $X_1, \dots, X_n$ qu'il suit la loi $\mathscr{E}(1)$?
2. Supposons que l'échantillon soit de loi exponentielle de paramètre θ inconnu. Déterminer la loi jointe de $S = (S_1, \dots, S_n)$ où $S_k = X_1 + \cdots + X_k$.
3. En déduire que le vecteur

$$T_n = \left(\frac{S_1}{S_n}, \dots, \frac{S_{n-1}}{S_n} \right)$$

a même loi que la statistique d'ordre $(U_{(1,n-1)}, \dots, U_{(n-1,n-1)})$ d'un échantillon $U_1, \dots, U_{n-1}$ de loi uniforme sur $[0, 1]$.
4. Proposer un test de niveau asymptotique α fondé sur T_n pour tester l'adéquation d'une loi à la famille des lois exponentielles.

On peut également tester l'homogénéité de deux distributions : on dispose de deux échantillons $X_1, \dots, X_n$ et $Y_1, \dots, Y_m$ de fonctions de répartition empiriques respectives F_n et G_n issus de lois de fonctions de répartition respectives F et G supposées continues. On peut tester

$$H_0 \ : \ F = G \quad \text{contre} \quad H_1 \ : \ F \neq G$$

grâce au résultat suivant :

$$\sqrt{\frac{nm}{n+m}} \| F_n - G_m \| \begin{cases} \xrightarrow[n,m\to\infty]{\mathscr{L}} \mu_{KS} & \text{sous } H_0, \\ \xrightarrow[n,m\to\infty]{p.s.} +\infty & \text{sous } H_1. \end{cases}$$

5. Corrigés

CORRIGÉ DE L'EXERCICE 8.2. Le quantile d'ordre p de la loi μ est unique dès que sa fonction de répartition F est continue en x_p (ou $\mu(\{x_p\} = 0)$. Par parité d'une densité de la loi $\mathcal{N}(0,\sigma^2)$, la médiane de la loi $\mathcal{N}(m,\sigma^2)$ vaut m. Puisqu'une densité de cette loi est strictement positive et continue sur $\mathbb{R}$, sa fonction de répartition est un $\mathscr{C}^1$ difféomorphisme de $\mathbb{R}$ sur $]0,1[$. En particulier tous les quantiles d'ordre p sont uniques. Cependant, il n'existe pas d'expression explicite d'autres quantiles que la médiane. La fonction Scilab cdfnor fournit selon les paramètres en entrée (voir l'aide) la fonction de répartition ou la fonction quantile.

CORRIGÉ DE L'EXERCICE 8.3. Soit μ une probabilité admettant une densité f strictement positive. On note m la médiane de μ. On suppose pour simplifier que la taille n de l'échantillon observé est impaire et on l'écrit $n = 2p-1$.
1. D'après la proposition 5.6, la fonction de répartition de $X_{(p)}$ est donnée par

$$F_p(x) = \frac{n!}{(p-1)!(n-p)!}\int_0^{F(x)} t^{p-1}(1-t)^{n-p}\,dt.$$

Sa densité est donc

$$f_p(x) = \frac{n!}{(p-1)!(n-p)!} f(x)F(x)^{p-1}(1-F(x))^{n-p}.$$

Enfin, la densité g_p de $\sqrt{2p-1}(X_{(p)}-m)$ est donnée par

$$\begin{aligned} g_p(u) &= \frac{f_p(m+u/\sqrt{n})}{\sqrt{n}} \\ &= \frac{(2p-1)!}{(p-1)!^2\sqrt{2p-1}} f\left(m+\frac{u}{\sqrt{2p-1}}\right) F\left(m+\frac{u}{\sqrt{2p-1}}\right)^{p-1}\left(1-F\left(m+\frac{u}{\sqrt{2p-1}}\right)\right)^{p-1}. \end{aligned}$$

2. Pour $u \in \mathbb{R}$ fixé,

$$f\left(m+\frac{u}{\sqrt{2p-1}}\right) \xrightarrow[p\to\infty]{} f(m)$$

et

$$F\left(m+\frac{u}{\sqrt{2p-1}}\right)^{p-1} = \left(\frac{1}{2}+\frac{f(m)u}{\sqrt{2p-1}}+o(1/\sqrt{p})\right)^{p-1} = \frac{1}{2^{p-1}}\left(1+\frac{2f(m)u}{\sqrt{2p-1}}+o(1/\sqrt{p})\right)^{p-1}.$$

On a donc,

$$g_p(u) = \frac{(2p-1)!}{(2^{p-1}(p-1)!)^2}\frac{1}{\sqrt{2p-1}}(f(m)+o(1/p))\left(1-\frac{4f(m)^2u^2}{2p-1}+o(1/p)\right)^{p-1}$$

La formule de Stirling assure que

$$g_p(u) \xrightarrow[p\to\infty]{} \frac{2f(m)}{\sqrt{2\pi}} e^{-2f(m)^2u^2}.$$

Ainsi, $\sqrt{2p-1}(X_{(p,2p-1)}-m)$ converge en loi vers la loi normale centrée de variance $1/(4f(m)^2)$.

CORRIGÉ DE L'EXERCICE 8.4. La loi de Cauchy n'ayant pas d'espérance, la méthode des moments est inopérante. Pour les deux premiers exemples, les lois ont un moment d'ordre 2 (indépendant de θ) et θ est l'espérance de la loi de l'échantillon. La moyenne empirique est donc un estimateur fortement consistant. Le théorème limite central assure qu'elle

converge vers θ à vitesse $\sqrt{n}$ avec une variance asymptotique $\mathbb{V}_\theta(X_1)$, c'est-à-dire 1 pour la loi gaussienne et 2 pour la loi de Laplace.
D'autre part, la médiane empirique est un estimateur fortement consistant dans les trois cas en vertu du théorème 8.9. De plus, le théorème 8.10 assure qu'elle converge vers θ à vitesse $\sqrt{n}$ avec une variance asymptotique respectivement donnée par $\pi/2$, 1 et $\pi^2/4$.
Pour le modèle gaussien, la moyenne empirique est donc meilleure que la médiane et c'est le contraire pour la loi de Laplace. L'intuition que l'on peut avoir est que les quantiles ne sont que peu sensibles aux queues de la distribution (au point que l'on puisse estimer la médiane de la loi de Cauchy) ce qui n'est bien entendu pas le cas de la moyenne.
Les figures 2 et 3 illustrent les comportements des deux estimateurs pour la loi de Laplace.

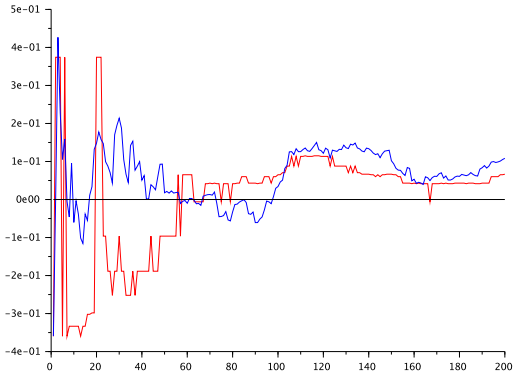

FIGURE 2. Estimation de θ pour la loi de Laplace par la moyenne empirique (courbe bleue) la médiane empirique (courbe rouge) en fonction de la taille de l'échantillon $n = 1, \ldots, 200$.

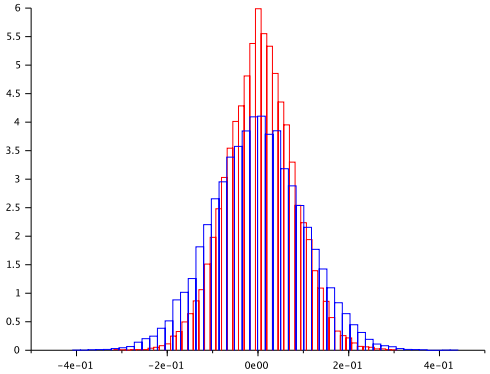

FIGURE 3. Distribution, pour la loi de Laplace, de la moyenne empirique (courbe bleue) et de la médiane empirique (courbe rouge) pour un échantillon de taille 200.

Pour la loi gaussienne, l'estimateur du maximum de vraisemblance est la moyenne empirique tandis que c'est la médiane empirique pour la loi de Laplace. Ceci provient de la remarque suivante. Soit μ une mesure de probabilité sur $\mathbb{R}$ possédant un moment d'ordre 2. Alors $x \mapsto \mathbb{E}((X-x)^2)$ et $x \mapsto \mathbb{E}(|X-x|)$ atteignent respectivement leur minimum en la

moyenne de μ et toutes les médianes de μ. Remarquons qu'il est impossible de déterminer l'estimateur du maximum de vraisemblance pour la loi de Cauchy...

CORRIGÉ DE L'EXERCICE 8.6. Voir la fonction Scilab kolmogorov.sci

Chapitre 9

Intervalles de confiance

Les probabilités s'attachent à décrire le comportement (souvent asymptotique) de fonctionnelles de variables aléatoires dont on connaît la loi. Une des deux grandes questions auxquelles s'intéresse la statistique est de décrire une loi de probabilité à partir d'observations supposées être des réalisations i.i.d. de cette loi inconnue. Le statisticien est une sorte de détective qui, face à de multiples individus, doit sélectionner un ou des suspects au vu d'indices dont aucun n'est une preuve.

1. Définitions et premier exemple

On se placera souvent dans un cadre paramétrique : soit $(\Omega, \mathscr{A})$ un espace mesurable et $(\mathbb{P}_\theta)_{\theta\in\Theta}$ une famille de probabilités sur $(\Omega, \mathscr{A})$ indexée par $\theta \in \Theta \subset \mathbb{R}^d$. Donnons tout de suite des exemples classiques de telles familles :

— $(\mathscr{B}(\theta))_{\theta\in[0,1]}$,
— $\{p = (p_1, \dots, p_k),\ p_i \geq 0,\ p_1 + \cdots + p_k = 1\}$,
— $(\mathscr{E}(\theta))_{\theta\in\mathbb{R}_+}$,
— $(\mathscr{N}(m, \sigma^2))_{(m,\sigma)\in\mathbb{R}\times\mathbb{R}_+}$.

Ces familles sont respectivement associées à un sondage sur l'abstention, un premier tour d'élections présidentielles, des temps de connexion à un serveur informatique et une mesure entachée d'erreurs.
Soit $\alpha \in]0,1[$ et un échantillon $X_1, \dots, X_n$ de loi $\mathbb{P}_\theta$, un intervalle (ou une région) de confiance pour le paramètre θ (inconnu-) de probabilité de confiance $1-\alpha$ est un intervalle (ou une région) qui dépend de l'échantillon (il est aléatoire) tel que la probabilité que cet intervalle contienne θ soit égale à $1-\alpha$.

EXEMPLE 9.1 (Échantillon gaussien de variance connue). L'exemple le plus simple est le suivant : soit $X_1, \dots, X_n$ i.i.d. de loi $\mathscr{N}(\theta, 1)$ avec θ inconnu. Alors

$$\forall \theta \in \mathbb{R}, \quad \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n X_i - \theta\right) \sim \mathscr{N}(0,1).$$

Or, on sait que si $Y \sim \mathscr{N}(0,1)$, alors $\mathbb{P}(|Y| \leq 1{,}96) = 0{,}95$. Ainsi,

$$\forall \theta \in \mathbb{R}, \quad \mathbb{P}_\theta\left(\theta \in \left[\overline{X}_n - \frac{1{,}96}{\sqrt{n}}, \overline{X}_n + \frac{1{,}96}{\sqrt{n}}\right]\right) = \mathbb{P}_\theta\left(\sqrt{n}\left|\frac{1}{n}\sum_{i=1}^n X_i - \theta\right| \leq 1{,}96\right) = 0{,}95.$$

L'intervalle $\left[\overline{X}_n - \frac{1{,}96}{\sqrt{n}}, \overline{X}_n + \frac{1{,}96}{\sqrt{n}}\right]$ est donc un intervalle de confiance pour θ de niveau de confiance 0,95.

Voici à présent la définition mathématique d'un intervalle de confiance telle qu'on peut la trouver dans [**Tas85**] par exemple.

DÉFINITION 9.2. Soit $\alpha \in]0,1[$ donné ; on appelle région de confiance pour le paramètre θ, de niveau de confiance $1-\alpha$, la famille non vide de parties de Θ $C_{x_1,\dots,x_n}$ telle que

$$\forall \theta \in \Theta, \quad \mathbb{P}_\theta(\theta \in C_{X_1,\dots,X_n}) = 1-\alpha.$$

EXERCICE 9.1. Montrer que, dans l'exemple 9.1, l'intervalle obtenu est l'intervalle de confiance pour θ (de probabilité de confiance 0,95) le moins long.

REMARQUE 9.3. Très souvent, lorsque le paramètre θ est réel, la région construite se trouvera être un intervalle. On parlera alors d'intervalle de confiance.

Dans l'exemple 9.1, on a utilisé, pour construire l'intervalle de confiance, une variable aléatoire qui dépend de l'échantillon et du paramètre inconnu mais dont la loi ne dépend pas du paramètre. C'est ce que l'on appelle une fonction pivotale. Cette recherche de fonction pivotale sera l'une des clés pour déterminer des intervalles de confiance.
Souvent la situation ne sera pas aussi simple que dans l'exemple 9.1 et il faudra se contenter par exemple de fonctions asymptotiquement pivotales (c'est-à-dire dont la loi converge, quand la taille de l'échantillon tend vers l'infini, vers une loi qui ne dépend pas de θ).
Les autres outils dont nous aurons besoin sont très variés : théorèmes limites (LFGN, TLC, convergence des quantiles empiriques), propriétés de lois classiques, inégalités de déviation à la Chernov...
Pour une définition complète des intervalles de confiance, des méthodes d'estimation etc on pourra consulter [**Tas85**] et [**Sap90**].

2. Les lois gaussiennes

L'exemple le plus commun en pratique est le cas d'un échantillon gaussien. Avec un peu de connaissance de lois dites classiques, on peut donner des intervalles de confiance pour estimer les paramètres de façon exacte (c'est-à-dire non asymptotique).
Notons $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ et $\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$. Alors $\overline{X}_n$ est l'estimateur sans biais de variance minimale de m. Il converge presque sûrement vers m. D'autre part, σ_n^2 converge aussi presque sûrement vers σ^2 mais σ_n^2 est biaisé : $\mathbb{E}(\sigma_n^2) = \frac{n-1}{n}\sigma^2$. On lui préfère parfois l'estimateur sans biais $\frac{n}{n-1}\sigma_n^2$.

2.1. Estimation de la moyenne...

2.1.1. *... si l'écart-type est connu.* On utilise la statistique pivotale $\frac{\sqrt{n}}{\sigma}(\overline{X}_n - m)$ qui suit la loi $\mathcal{N}(0,1)$. L'intervalle

$$\left[\overline{X}_n - \frac{\sigma x_{1-\alpha/2}}{\sqrt{n}}\ ;\ \overline{X}_n + \frac{\sigma x_{1-\alpha/2}}{\sqrt{n}}\right]$$

est un intervalle de confiance pour m une probabilité de confiance $1-\alpha$ où x_q est défini par la relation $\int_{-\infty}^{x_q} e^{-x^2/2}\frac{dx}{\sqrt{2\pi}} = q$ ou encore $x_q = \Phi^{-1}(q)$ en notant Φ la fonction de répartition de la loi gaussienne centrée réduite.

EXEMPLE 9.4. Pamela est un mannequin célèbre dont le poids est strictement surveillé par Cruella. Cette charmante dame a investi un jour dans l'achat d'une balance *Harmonia* afin de connaître précisément le poids de sa protégée. Horreur : elle a constaté sur l'emballage de la balance que les fabricants (d'honnêtes artisans suisses) admettaient que leur outil de mesure (nul n'est parfait) pouvait commettre des erreurs de mesure dont l'écart-type valait 0,1 kg. En effet les pièces détachées ne sont pas toutes exactement identiques, leur montage n'est jamais parfait et le transport à travers les Alpes endommage parfois les balances. Ne faisant ni une ni deux, Cruella a, dès le lendemain, dévalisé le magasin en investissant dans l'achat de 99 nouvelles balances *Harmonia* et a forcé Pamela à sauter

sur les 100 balances pendant qu'elle relevait scrupuleusement les 100 mesures. Résultat moyen des pesées : 55,4 kg. Donner à Cruella un intervalle de confiance pour le poids de Pamela de probabilité de confiance 0,95.

2.1.2. *... si l'écart-type est inconnu.* On utilise le fait que $T = \dfrac{\overline{X}_n - m}{\sigma_n}\sqrt{n-1}$ suit une loi de Student à $n-1$ degrés de liberté. Pour mémoire, la densité de la loi de Student à n degrés de liberté possède la densité :

$$f_{St(n)}(t) = \frac{1}{\sqrt{n}\mathrm{B}(1/2, n/2)}\left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} \quad \text{où} \quad B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}.$$

Cette loi est tabulée : on peut trouver $t_{\alpha/2}$ tel que $\mathbb{P}(-t_{\alpha/2} \leq T \leq t_{\alpha/2}) = 1-\alpha$ et l'intervalle de confiance pour m s'écrit alors

$$\left[\overline{X}_n - t_{\alpha/2}\frac{\sigma_n}{\sqrt{n-1}}\ ;\ \overline{X}_n + t_{\alpha/2}\frac{\sigma_n}{\sqrt{n-1}}\right].$$

EXEMPLE 9.5. On a mesuré le poids de raisin par souche sur 10 souches prises au hasard dans une vigne. On a obtenu les résultats suivants (en kg) :

$$2,4\ ;\ 3,2\ ;\ 3,6\ ;\ 4,1\ ;\ 4,3\ ;\ 4,7\ ;\ 5,4\ ;\ 5,9\ ;\ 6,5\ ;\ 6,9.$$

Le vigneron se demande quel est le poids moyen de raisin par cep. Voici une question concrète de sa part. Aux mathématicien(nes) d'apporter une réponse claire et argumentée. Pour cela, on commence par « modéliser »le problème. Si les 10 ceps appartiennent à une même vigne, ils auront été traités aux pesticides, exposés au soleil et à la pluie et entretenus avec amour par le vigneron de façon équivalente. Cependant, des petits écarts sont inévitables : un peu d'ombre d'un côté, un pied de vigne dans un trou d'eau, un pied à l'hérédité plus solide que ses voisins... Ceci nous conduit à modéliser les mesures effectuées sur les 10 ceps par la réalisation de 10 v.a. identiquement distribuées (les conditions sont globalement les mêmes) et indépendantes (les ceps ne sont pas sous le même arbre, dans le même trou d'eau...) de loi $\mathcal{N}(m, \sigma^2)$, où m représente le poids moyen de raisin par cep et σ son écart-type. Ces deux quantités sont a priori inconnues et l'on veut estimer m. Donner une estimation et un intervalle de confiance pour m.

2.2. Estimation de l'écart-type...

2.2.1. *... si la moyenne est connue.* La statistique $T = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - m)^2$ est l'estimateur sans biais de variance minimale de σ^2 et $\dfrac{nT}{\sigma^2}$ suit une loi du χ^2 à n degrés de liberté. Pour mémoire, une v.a. de loi du χ^2 à n degrés de liberté notée $\chi^2(n)$ admet pour densité

$$f_{\chi^2(n)}(x) = \frac{1}{2^{n/2}\Gamma(n/2)}e^{-x/2}x^{n/2-1}\mathbf{1}_{\{x>0\}}.$$

Les lois du χ^2 ne sont pas symétriques : les bornes de l'intervalle de confiance ne se déterminent pas tout à fait comme dans les cas précédents. On procède en général de la façon suivante : on détermine k_1 et k_2 tels que

$$\mathbb{P}(X \leq k_1) = \frac{\alpha}{2} \quad \text{et} \quad \mathbb{P}(X \geq k_2) = \frac{\alpha}{2} \quad \text{où} \quad X \sim \chi^2(n).$$

On a donc

$$\mathbb{P}\left(k_1 < \frac{nT}{\sigma^2} < k_2\right) = 1-\alpha.$$

Un intervalle de confiance pour σ de probabilité de confiance $1-\alpha$ peut ainsi être choisi de la forme

$$\left[\sqrt{\frac{nT}{k_2}}\,;\,\sqrt{\frac{nT}{k_1}}\right].$$

EXERCICE 9.2 (Pamela suite). Expliquer comment les constructeurs de balance ont pu fournir une valeur de σ. Fournir aussi un intervalle de confiance.

2.2.2. *... si la moyenne est inconnue.* On utilise le fait que $\dfrac{n\sigma_n^2}{\sigma^2}$ suit une loi $\chi^2(n-1)$ (attention au nombre de degrés de liberté) et on procède comme dans le cas précédent.

EXERCICE 9.3 (Raisins suite). Donner un intervalle de probabilité de confiance 0,95 pour le poids de raisin par pied de vigne dans l'exemple 9.5.

3. Les grands échantillons

Lorsque la loi des variables aléatoires n'est plus gaussienne, les calculs explicites de loi ne fonctionnent plus. On a alors recours aux deux théorèmes limites les plus célèbres : la loi forte des grands nombres et le théorème limite central. Si l'on dispose d'un échantillon $X_1,\dots,X_n$ de loi inconnue $\mathbb{P}_\theta$ indexée par sa moyenne θ et de variance connue, on utilise la loi forte des grands nombres et le théorème limite central pour dire que l'intervalle

$$\left[\overline{X}_n-\frac{1,96\sigma}{\sqrt{n}}\,;\,\overline{X}_n+\frac{1,96\sigma}{\sqrt{n}}\right]$$

est un intervalle de confiance de probabilité de confiance asymptotique 0,95, c'est-à-dire que

$$\forall\theta\in\Theta,\quad \mathbb{P}_\theta\left(\theta\in\left[\overline{X}_n-\frac{1,96\sigma}{\sqrt{n}}\,;\,\overline{X}_n+\frac{1,96\sigma}{\sqrt{n}}\right]\right)\xrightarrow[n\to\infty]{}0,95.$$

REMARQUE 9.6. Le point important est que l'on ne sait pas à quelle vitesse a lieu cette convergence. Il existe des résultats théoriques pour contrôler cette erreur mais ils sont en général beaucoup trop pessimistes (car très généraux) et font intervenir des paramètres de la loi de l'échantillon, comme par exemple le moment d'ordre trois, dont on ne dispose pas.

Le lemme de Slutsky (démontré dans le fascicule de probabilités) que nous rappelons ci-dessous est très utile pour les intervalles de confiance.

THÉORÈME 9.7 (Lemme de Slutsky). *Soit $X_n\xrightarrow{\mathbb{P},\mathscr{L}}X$ et $Y_n\xrightarrow{\mathscr{L}}c$ où c est une constante. Alors*

$$(X_n,Y_n)\xrightarrow{\mathbb{P},\mathscr{L}}(X,c)$$

où la convergence est en loi ou en probabilité selon le mode de convergence de X_n. De plus

(1) $X_n+Y_n\xrightarrow{\mathbb{P},\mathscr{L}}X+c$,

(2) $X_nY_n\xrightarrow{\mathbb{P},\mathscr{L}}cX$,

(3) si $c\neq 0$, $X_n/Y_n\xrightarrow{\mathbb{P},\mathscr{L}}X/c$.

EXEMPLE 9.8. Soit $(X_n)_{n\geq 1}$ des variables aléatoires indépendantes de même loi, d'espérance m et de variance σ^2 inconnues. On souhaite estimer m. Posons $\bar{X}_n$ la moyenne de $X_1,\dots,X_n$. D'après le théorème limite central, on sait que

$$\lim_{n\to\infty}\mathbb{P}\left(\bar{X}_n-\frac{1,96\sigma}{\sqrt{n}}\leq\bar{X}_n+\frac{1,96\sigma}{\sqrt{n}}\right)=0,95.$$

Mais σ est inconnu. On dit que σ est un paramètre de nuisance. Cependant,la loi des grands nombres assure que

$$\hat{\sigma}_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\bar{X}_n)^2} \xrightarrow[n\to\infty]{\mathbb{P},p.s.} \sigma.$$

On a donc encore, d'après le lemme de Slutsky

$$\lim_{n\to\infty} \mathbb{P}\left(\bar{X}_n - \frac{1,96\hat{\sigma}}{\sqrt{n}} \leq \bar{X}_n + \frac{1,96\hat{\sigma}}{\sqrt{n}}\right) = 0,95.$$

On introduit ainsi une deuxième approximation (en plus de celle du TLC) : la probabilité de confiance est doublement asymptotique...
Voici également un résultat très utile en pratique pour transférer des vitesses de convergence en loi.

THÉORÈME 9.9 (Méthode delta). *Soient $(v_n)_{n\geq 1}$ une suite de réels qui tend vers $+\infty$, $a \in \mathbb{R}$ et $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles telle que $v_n(X_n - a)$ converge en loi vers une loi μ sur $\mathbb{R}$. Si g est une fonction de $\mathbb{R}$ dans $\mathbb{R}$ de classe $\mathscr{C}^1$ telle que $g'(a) \neq 0$ alors*

$$\frac{v_n}{g'(a)}(g(X_n) - g(a)) \xrightarrow[n\to\infty]{\mathscr{L}} \mu.$$

DÉMONSTRATION. La suite $1/v_n$ converge en loi vers la constante 0 et $v_n(X_n - a)$ converge en loi vers μ. D'après le lemme de Slutsky, $(1/v_n, v_n(X_n - a))$ converge en loi vers $(0, Z)$ où Z suit la loi μ. On en déduite que la suite $(1/v_n)v_n(X_n - a)$ converge en loi vers la constante 0. Cette converge a donc lieu également en probabilité.
Soit ψ la fonction définie pour tout $y \in \mathbb{R}$ par

$$\psi(y) = \int_0^1 g'(a + u(y - a))\, du.$$

On remarque que pour tous $x \neq y$, $g(y) - g(x) = \psi(y)(y - x)$. En choisissant $x = a$ et $y = X_n$, on obtient

$$v_n(g(X_n) - g(a)) = \psi(X_n)v_n(X_n - a).$$

La fonction ψ est continue donc $\psi(X_n)$ converge en probabilité vers $\psi(a) = g'(a)$. Donc $v_n(g(X_n) - g(a))$ converge en loi vers $g'(a)Z$. □

EXEMPLE 9.10. Soit $(X_n)_{n\geq 1}$ une suite de v.a. i.i.d. de loi exponentielle de paramètre λ. Notons $\bar{X}_n$ la moyenne de l'échantillon $X_1, \ldots, X_n$. Alors, le théorème limite central assure que

$$\sqrt{n}\left(\bar{X}_n - \frac{1}{\lambda}\right) \xrightarrow[n\to\infty]{\mathscr{L}} \mathscr{N}(0, \lambda^{-2}).$$

On a donc, avec $g(x) = 1/x$ et $a = 1/\lambda$,

$$\sqrt{n}\left(\frac{1}{\bar{X}_n} - \lambda\right) \xrightarrow[n\to\infty]{\mathscr{L}} \mathscr{N}(0, \lambda^2).$$

Enfin, d'après le lemme de Slutsky, on a

$$\bar{X}_n\sqrt{n}\left(\frac{1}{\bar{X}_n} - \lambda\right) \xrightarrow[n\to\infty]{\mathscr{L}} \mathscr{N}(0, 1).$$

On déduit facilement de cette convergence un intervalle de confiance asymptotique pour λ.

4. Intervalles de confiance pour une proportion

Les méthodes employées sont les mêmes que dans le cas de grands échantillons de loi quelconque mais comme il s'agit d'un cas particulier très fréquent, il est bon de le détailler à part. On observe $X_1, \dots, X_n$ indépendantes de loi $\mathscr{B}(\theta)$ avec $\theta \in [0,1]$ inconnu. On note toujours $\overline{X}_n$ la moyenne empirique de l'échantillon. Nous présentons les constructions de deux intervalles de confiance pour θ, toutes deux basées principalement sur l'utilisation du théorème limite central :

$$\sqrt{n}\frac{\overline{X}_n - \theta}{\sqrt{\theta(1-\theta)}} \xrightarrow[n\to\infty]{\mathscr{L}} \mathscr{N}(0,1).$$

EXERCICE 9.4 (Méthode 1). On considère la statistique asymptotiquement pivotale

$$\sqrt{n}\frac{\overline{X}_n - \theta}{\sqrt{\theta(1-\theta)}}.$$

Montrer que

$$\left[\frac{\overline{X}_n + \frac{r^2}{2n} - \frac{r}{\sqrt{n}}\sqrt{\frac{r^2}{4n} + \overline{X}_n(1-\overline{X}_n)}}{1+\frac{r^2}{n}} \; ; \; \frac{\overline{X}_n + \frac{r^2}{2n} + \frac{r}{\sqrt{n}}\sqrt{\frac{r^2}{4n} + \overline{X}_n(1-\overline{X}_n)}}{1+\frac{r^2}{n}}\right]$$

est un intervalle de confiance pour θ de probabilité de confiance asymptotique $1-\alpha$ si $r = \Phi^{-1}(1-\alpha/2)$ (où Φ est la fonction de répartition de la loi gaussienne centrée réduite).

EXERCICE 9.5 (Méthode 2). Montrer que la statistique

$$\sqrt{n}\frac{\overline{X}_n - \theta}{\sqrt{\overline{X}_n(1-\overline{X}_n)}}.$$

est asymptotiquement pivotale. En déduire que

$$\left[\overline{X}_n - r\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}} \; ; \; \overline{X}_n + r\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}\right]$$

est un intervalle de confiance pour θ de probabilité de confiance asymptotique $1-\alpha$ si $r = \Phi^{-1}(1-\alpha/2)$ (où Φ est la fonction de répartition de la loi gaussienne centrée réduite).

5. Utilisation des quantiles empiriques

Les paramètres à estimer peuvent parfois dépendre de manière simple des quantiles de la loi (un peu comme dans la méthode des moments). On peut alors utiliser les théorèmes 8.9 et 8.10 de convergence des quantiles empiriques pour obtenir des estimateurs consistants et des intervalles de confiance. On pourra par exemple consulter l'exercice 8.4.

6. Inégalité de Chernov

Ce chapitre utilise des notions de probabilités qui font partie du tronc commun et les techniques employées sont très classiques. Elles seront réemployées à de nombreuses occasions.

On souhaite à présent être en mesure de fournir des intervalles de confiance non asymptotiques, c'est-à-dire qui n'utilisent pas un théorème limite pour lequel on ne contrôle pas la vitesse de convergence. Pour cela, on va minorer la probabilité de confiance par une quantité exacte dépendant de n. On assure ainsi ses arrières en prenant une marge de sécurité (espérons que c'est ce que font les constructeurs de centrales nucléaires et les

fabriquants de médicaments). La méthode qui repose sur les propriétés d'intégrabilité de la loi de l'échantillon est une amélioration de l'inégalité de Tchebychev qui fournit la majoration suivante :

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X_1)\right| \geq r\right) \leq \frac{\mathbb{V}(X_1)}{nr^2}.$$

Ceci peut encore s'interpréter dans notre contexte en terme d'intervalle de confiance pour l'espérance :

$$\mathbb{P}\Big(\mathbb{E}(X_1) \in \Big[\overline{X}_n - r\ ;\ \overline{X}_n + r\Big]\Big) \geq \frac{\mathbb{V}(X_1)}{nr^2}.$$

On a ainsi construit un intervalle de confiance pour $\mathbb{E}(X_1)$ de probabilité de confiance minorée par $\mathbb{V}(X_1)/(nr^2)$. Cette inégalité, valable dès que X_1 est de carré intégrable peut être fortement améliorée sous des conditions d'intégrabilité plus fortes.

EXERCICE 9.6. Soit $\mu = (1/2)\delta_{-1} + (1/2)\delta_1$ et $X_1, \dots, X_n$ un échantillon de loi μ.
1. Calculer la transformée de Laplace de μ définie par $L_\mu(t) = \mathbb{E}(e^{tX})$ et montrer que $\ln L_\mu(t) \leq t^2/2$.
2. Montrer que, pour tout $\lambda > 0$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X_1)\right| \geq r\right) = 2\mathbb{P}\Big(e^{\lambda \sum_i X_i} \geq e^{\lambda n r}\Big) \leq 2e^{-n(\lambda r - \ln L_\mu(\lambda))} \leq 2e^{-n(\lambda r - \lambda^2/2)}.$$

3. Optimiser en $\lambda > 0$.
4. En déduire un intervalle de confiance non asymptotique pour θ dans le modèle suivant : pour tout $\theta \in \mathbb{R}$, $\mathbb{P}_\theta = (1/2)\delta_{-1+\theta} + (1/2)\delta_{1+\theta}$.

Le principe général, qui suit le raisonnement de l'exercice 9.6, fonctionne dès que la mesure μ considérée admet une transformée de Laplace $L_\mu(t) = \mathbb{E}(e^{tX})$ définie sur un voisinage de l'origine. Ceci assure que μ a tous ses moments finis et donc de bonnes qualités de décroissance que l'on va chercher à utiliser au mieux.

EXERCICE 9.7 (Un peu de transformée de Laplace). On rassemble ici les propriétés essentielles de la transformée de Laplace. Soit μ unes mesure de probabilité sur $\mathbb{R}$.
1. Montrer que L_μ est convexe. Que dire son domaine de définition $\mathscr{D}$? Peut-il être réduit à $\{0\}$?
2. Montrer que L_μ est une fonction log-convexe (son logarithme est une fonction convexe). Soit $X_1, \dots, X_n$ un échantillon de loi μ. Montrer que, pour tous $r \geq 0$ et $\lambda > 0$,

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X_1) \geq r\right) \leq \exp\left(-n \sup_{\lambda > 0}\big(\lambda r + \lambda \mathbb{E}(X_1) - \ln L_\mu(\lambda)\big)\right),$$

et que le supremum est fini et atteint dès que $L_\mu(\lambda_0)$ est fini pour un certain $\lambda_0 > 0$.
3. Calculer les transformées de Laplace des lois $\mathscr{N}(m, \sigma^2)$, $\mathscr{B}(p)$, $\mathscr{P}(\lambda)$ et $\mathscr{E}(\lambda)$ en précisant leurs domaines de définition.
4. Déterminer une inégalité de déviation pour la moyenne empirique pour chacun de ces cas particuliers.

REMARQUE 9.11. La majoration de la probabilité de déviation via la transformée de Laplace est le premier pas de ce que les probabilistes appellent la théorie des grandes déviations. Nous venons de montrer en particulier que

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X_1) \geq r\right) \leq -\sup_{\lambda > 0}\big(\lambda r + \lambda \mathbb{E}(X_1) - \ln L_\mu(\lambda)\big).$$

On peut en fait établir que cette majoration est en fait une égalité, c'est-à-dire que l'inégalité de Chernov donne le bon ordre de grandeur (dans l'exponentielle).

Pour retrouver tous les résultats des exercices proposés ici, on pourra consulter le premier chapitre [**DZ98**] qui, bien qu'écrit en anglais, est très facile à lire.

7. Corrigés

CORRIGÉ DE L'EXERCICE 9.7.
1. La fonction exponentielle est positive et convexe donc, pour $s \in [0,1]$ et $x, y \in \mathbb{R}$, on a la relation (dans $[0,+\infty]$)

$$L_\mu(sx + (1-s)y) \leq sL_\mu(x) + (1-s)L_\mu(y).$$

La fonction L_μ est donc convexe et son domaine de définition $\mathscr{D}$ l'est aussi. C'est donc un intervalle. Il contient 0 puisque $L_\mu(0) = 1$. Il est par exemple réduit à $\{0\}$ pour la loi de Cauchy.
2. Si I est un intervalle ouvert inclus dans l'intérieur $\mathscr{D}$, il existe y et z dans $\mathscr{D}$ tels que $y < x < z$ pour tout $x \in I$, tout $t \in \mathbb{R}$ et tout $n \in \mathbb{N}$, il existe $C_n > 0$ tel que

$$\sup_{x \in I} \left| x^n e^{tx} \right| \leq Ce^{ty} + e^{tz}.$$

La fonction L_μ est donc de classe $\mathscr{C}^\infty$ à l'intérieur de $\mathscr{D}$ (théorème de dérivation sous l'espérance). De plus,

$$L''_\mu(t) = \frac{\mathbb{E}X^2 e^{tX}}{\mathbb{E}e^{tX}} - \left(\frac{\mathbb{E}Xe^{tX}}{\mathbb{E}e^{tX}}\right)^2$$

Cette quantité est positive (c'est une variance) donc L_μ est log-convexe.
3. Soit $r > 0$ et $\lambda > 0$ alors, d'après l'inégalité de Markov,

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}(X_1) \geq r\right) &= \mathbb{P}\left(\sum_{i=1}^n X_i \geq n(r + \mathbb{E}(X_1))\right) \\
&= \mathbb{P}\left(e^{\lambda \sum_{i=1}^n X_i} \geq e^{\lambda n(r+\mathbb{E}(X_1))}\right) \\
&\leq \exp\left(-n\left(\lambda r + \lambda \mathbb{E}(X_1) - \ln L_\mu(\lambda)\right)\right).
\end{aligned}$$

On conclut en optimisant en λ.
4. On a classiquement

$$\begin{aligned}
L_{\mathscr{N}(m,\sigma^2)}(t) &= e^{\sigma^2 t^2/2 + mt} \\
L_{\mathscr{B}(p)}(t) &= pe^t + (1-p) \\
L_{\mathscr{P}(\lambda)}(t) &= e^{\lambda(e^t - 1)} \\
L_{\mathscr{E}(\lambda)}(t) &= \frac{\lambda}{\lambda - t} \quad \text{pour } t < \lambda.
\end{aligned}$$

Chapitre 10

Tests du chi-deux

L'une des fonctions des statistiques est de proposer, à partir d'observations d'un phénomène aléatoire (ou modélisé comme tel) une estimation de la loi de ce phénomène. C'est que nous avons fait en construisant des intervalles de confiance. Les statistiques servent aussi à prendre des décisions. Peut-on considérer qu'un médicament est plus efficace qu'un placebo ? Le nombre de consultations de Google par seconde suit-il une loi de Poisson ? Les gènes pilotant la couleur des yeux et celle des cheveux sont-ils sur les mêmes chromosomes ? Il y a deux points communs (au moins) à toutes ces questions : leurs réponses sont des oui-non et le phénomène sous-jacent est aléatoire. Les tests statistiques vont permettre d'apporter une réponse à des questions manichéennes en contrôlant l'aléa inhérent à la situation.
En statistique les deux éventualités sont appelées des hypothèses et sont notées H_0 (*hypothèse nulle*) et H_1 (*hypothèse alternative*). Souvent H_1 sera le contraire de H_0 ; dans tous les cas, le postulat est qu'une et une seule des deux hypothèses est vraie. Un test statistique est un algorithme qui conduit à accepter H_0 ou à rejeter H_0 à partir d'observations d'un phénomène aléatoire. Voici un exemple un peu naïf mais très instructif.

EXEMPLE 10.1. On vient d'acheter un dé à six faces tout neuf sensé fournir des résultats distribués selon la loi uniforme sur $\{1,\dots,6\}$. On souhaite valider (ou invalider) cette affirmation du fabriquant à partir de l'observation de n lancers de dé. On est donc amené à poser les hypothèses :

$$H_0 \ : \text{ le dé est équilibré} \quad \text{et} \quad H_1 \ : \text{ le dé est pipé.}$$

Il y a deux éventualités pour la réalité et deux décisions possibles. Sur les quatre configurations, deux sont satisfaisantes (la prise de décision correspond à la réalité). On peut résumer la situation avec le tableau suivant :

Décision \ Réalité	H_0 est vraie	H_1 est vraie
on accepte H_0	bonne décision	erreur de seconde espèce
on rejette H_0	erreur de première espèce	bonne décision

On distingue les deux erreurs car H_0 et H_1 ne jouent pas un rôle symétrique. L'hypothèse nulle représentera ce que l'on pense être la réalité. On souhaitera en premier lieu imposer la probabilité de commettre l'erreur de première espèce, c'est-à-dire de rejeter H_0 alors qu'elle est vraie. Cette probabilité est appelée niveau du test.

REMARQUE 10.2. On peut faire une analogie avec la justice qui pose comme principe la présomption d'innocence. On souhaite contrôler en priorité la probabilité d'envoyer un innocent en prison (erreur de première espèce) en négligeant pour l'instant celle de relâcher un coupable (erreur de seconde espèce). Dans l'exemple 10.1, on laisse le bénéfice du doute au constructeur de dés.

L'idée est de trouver une statistique (une fonction des observations) dont on connaît la loi si H_0 est vraie et qui ne se comporte pas de la même manière selon que H_0 ou H_1 est vraie. Les tests du χ^2 sont un exemple relativement simple de tests statistiques qui vont permettre de tester

(1) l'adéquation à une loi de probabilité sur un ensemble fini : est-il raisonnable de penser que les résultats que j'observe sont des réalisations i.i.d. d'une loi $(p_1,\dots,p_k)$ sur $\{1,\dots,k\}$?

(2) l'indépendance de deux caractères mesurés sur un même individu.

(3) l'homogénéité de plusieurs échantillons : deux médicaments ont-ils le même effet (guérison, amélioration, état stationnaire) sur la population atteinte ?

Loi de Hardy-Weinberg. Considérons une population d'individus diploïde (chaque individu porte deux versions d'un même gène) idéale (taille infinie, pas de sélection, pas de mutation, pas d'immigration, deux allèles A et a du gène seulement, générations séparées). Alors, la proportion p d'allèles A est constante au cours du temps. De plus, si au temps initial, les proportions des génotypes AA, Aa et aa sont respectivement égales à q, r et s (avec $q+r+s=1$), alors $p=r+s/2$ et, à la génération suivante, ces proportions sont données par

$$(r+s/2)^2, \quad 2(r+s/2)(t+s/2) \quad \text{et} \quad (t+s/2)^2.$$

On dit alors que la population est en équilibre de Hardy-Weinberg de paramètre p.
Ce modèle peut paraître très simpliste puisqu'aucune vraie population n'est idéale au sens ci-dessus. Toutefois, on souhaite confronter ce modèle à des données. Sur une population de 100 individus, on observe les répartitions respectives suivantes de génotypes : 13, 49 et 38. Ces données nous poussent-elle à rejeter l'hypothèse que la population est en équilibre de Hardy-Weinberg ? Nous verrons au cours du chapitre plusieurs façons de répondre à cette question grâce aux tests statistiques.

1. Un peu de probabilités

Cette section retrace les grandes lignes du raisonnement qui permet de construire la statistique de test du χ^2.
Soit $p=(p_1,\dots,p_k)$ une loi de probabilité sur $\{1,\dots,k\}$ et $X_1,\dots,X_n$ un échantillon de loi p. Pour $1\le i\le n$, on pose $N_i(n)=\mathrm{Card}\{j=1,\dots,n,\ X_j=i\}$. On dit que le vecteur $N(n)=(N_1(n),\dots,N_k(n))$ suit la loi multinomiale de paramètre (n,p).

REMARQUE 10.3. Attention n est un entier et p est un vecteur de probabilité. Pour Scilab, les paramètres de grand option mul sont n et un vecteur-colonne formé des $k-1$ premières coordonnées de p.

EXERCICE 10.1 (Loi multinomiale). Soit $N(n)$ de loi multinomiale de paramètre (n,p).

1. Montrer que $\mathbb{P}(N_1(n)=n_1,\dots,N_k(n)=n_k)=\begin{cases}\dfrac{n!}{n_1!\cdots n_k!} & \text{si } n_1+\cdots+n_k=n,\\ 0 & \text{sinon.}\end{cases}$

2. Quelle est la loi de $N_i(n)$ pour $i=1,\dots,k$?

3. Montrer que

$$\mathbb{E}(N_i(n))=np_i, \quad \mathbb{V}(N_i(n))=np_i(1-p_i) \quad \text{et} \quad \mathrm{cov}(N_i(n),N_j(n))=-np_ip_j$$

pour $i\neq j$.

4. Les v.a. $N_1(n)$ et $N_2(n)$ sont-elles indépendantes ? Interpréter le signe de la covariance de $N_1(n)$ et $N_2(n)$.

5. Montrer que $(N(n)/n)_n$ converge presque sûrement vers le vecteur p.

PROPOSITION 10.4. *Soit $(X_n)_{n\in\mathbb{N}}$ des v.a. i.i.d. de loi p. Avec les notations ci-dessus (en identifiant p à un vecteur colonne et en notant ${}^t p$ son transposé),*

$$\sqrt{n}\left(\frac{N(n)}{n} - p\right) \xrightarrow[n\to\infty]{\mathscr{L}} \mathcal{N}(0, K_p),$$

où $K_p = \Delta_p - p\,{}^t p$ en notant Δ_p est la matrice diagonale de diagonale p.

DÉMONSTRATION. Appliquons le théorème limite central multidimensionnel à la suite de v.a. i.i.d. $(Y(n))_{n\geq 1}$ à valeurs dans $\mathbb{R}^k$ définies par $Y_i(n) = \mathbf{1}_{\{X_n = i\}}$. Ces v.a. suivent la loi multinomiale de paramètre $(1, p)$. La v.a. $Y(1)$ admet p pour (vecteur-)espérance et K_p pour matrice de covariance. Remarquons de plus que $N(n) = Y(1) + \cdots + Y(n)$. □

Le résultat suivant, qui est à la base de l'idée des tests du χ^2, consiste à produire, à partir d'une v.a. normale dans $\mathbb{R}^k$ de loi $\mathcal{N}(0, K_p)$ une v.a. dont la loi ne dépend pas de p.

PROPOSITION 10.5. *Si Z suit la loi $\mathcal{N}(0, K_p)$ alors la v.a. $Z_1^2/p_1 + \cdots + Z_k^2/p_k$ suit une loi du χ^2 à $k-1$ degrés de liberté.*

DÉMONSTRATION. Si l'on note $U_i = Z_i/\sqrt{p_i}$ pour $i = 1, \ldots, k$ alors $U = (U_1, \ldots, U_k)$ suit la loi $\mathcal{N}(0, I - \sqrt{p}\sqrt{p}^T)$ (où $\sqrt{p}$ désigne le vecteur(-colonne) des racines carrées des coefficients de p). Remarquons à présent que la matrice de covariance de U est la matrice de projection orthogonale sur l'hyperplan orthogonal au vecteur normé $\sqrt{p}$. Il existe donc une matrice orthogonale O telle que $O(I - \sqrt{p}\sqrt{p}^T)O^T$ soit égale à la matrice $\Delta_{(1,\ldots,1,0)}$ (avec la notation introduite dans la proposition 10.4). Le vecteur aléatoire OU suit alors la loi normale centrée de matrice de covariance

$$\mathbb{E}((OU)(OU)^T) = \mathbb{E}(OUU^TO^T) = O\mathbb{E}(UU^T)O^T = O\mathrm{cov}(U)O^T.$$

En d'autres termes, OU suit la loi $\mathcal{N}(0, \Delta_{(1,\ldots,1,0)})$, c'est-à-dire que les $k-1$ premières coordonnées de OU sont des v.a. gaussiennes centrées réduites indépendantes tandis que la dernière est nulle. Ainsi la loi de la v.a.

$$\frac{Z_1^2}{p_1} + \cdots + \frac{Z_k^2}{p_k} = \|U\|^2 = \|OU\|^2$$

est-elle la même que celle de la somme des carrées de $k-1$ v.a. gaussiennes indépendantes centrées réduites, c'est-à-dire une loi du χ^2 à $k-1$ degrés de liberté. □

LEMME 10.6. *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. définies sur $(\Omega, \mathscr{A}, \mathbb{P})$ à valeurs dans $\mathbb{R}^k$ qui converge en loi vers μ et f de $\mathbb{R}^k$ dans $\mathbb{R}^l$ continue. Alors $(f(X_n))_{n\in\mathbb{N}}$ converge en loi vers la mesure image de μ par f.*

DÉMONSTRATION. Il suffit d'utiliser la caractérisation de la convergence en loi par la convergence des espérances des fonctions continues. □

Il ne reste plus qu'à rassembler les propositions 10.4 et 10.5 et le lemme 10.6 pour obtenir le résultat principal de ce chapitre.

THÉORÈME 10.7. *Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de v.a. i.i.d. de loi p alors*

$$D_n = n\sum_{i=1}^{k} \frac{(N_i(n)/n - p_i)^2}{p_i} = \sum_{i=1}^{n} \frac{(N_i(n) - np_i)^2}{np_i} \xrightarrow[n\to\infty]{\mathscr{L}} \chi^2(k-1).$$

On a donc construit à partir de l'échantillon $X_1, \ldots, X_n$ une statistique D_n dont la loi ne dépend asymptotiquement plus de p. On est donc capable de choisir une région de $\mathbb{R}$ telle que D_n appartienne à cette région avec une probabilité (asymptotique) donnée.

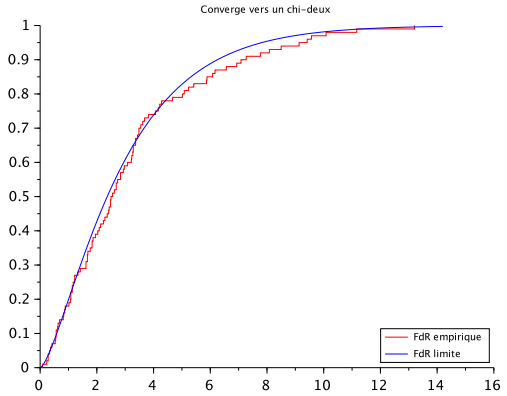

FIGURE 1. Fonction de répartition empirique de $N = 100$ réalisations de D_n défini dans le théorème 10.7 avec $n = 100$ et $p = (0.2\ 0.3\ 0.2\ 0.3)$ et fonction de répartition de la loi $\chi^2(3)$.

Ajoutons à ce théorème de convergence, la proposition suivante. Elle n'est qu'une conséquence directe de la loi des grands nombres mais elle joue un rôle essentiel dans l'élaboration des tests du χ^2.

PROPOSITION 10.8. *Soit q une loi de probabilité sur $\{1,\dots,k\}$ différente de p. Alors*

$$E_n = \sum_{i=1}^{k} \frac{(N_i(n)/n - q_i)^2}{q_i} \xrightarrow[n\to\infty]{p.s.} \sum_{i=1}^{k} \frac{(p_i - q_i)^2}{q_i}.$$

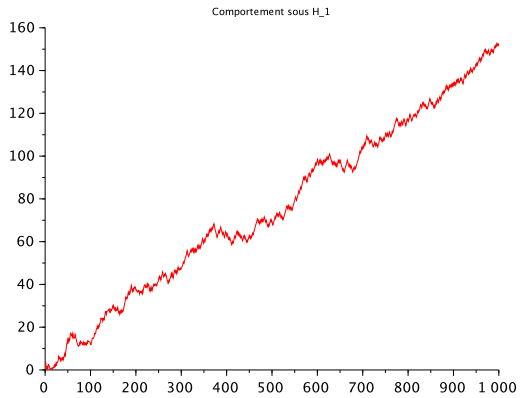

FIGURE 2. Comportement de $n \mapsto E_n$ défini dans la proposition 10.8 pour un échantillon de loi $p = (0.2\ 0.3\ 0.2\ 0.3)$ et $q = (0.3\ 0.2\ 0.3\ 0.2)$.

CONSEILS BIBLIOS 10.9. Pour retrouver tous ces résultats et bien d'autres, on pourra consulter [**Tas85**] ou [**Sap90**] (et [**Mon82**] pour les plus motivés).

2. Test d'adéquation à une loi donnée

On dispose d'observations que l'on considère comme des réalisations i.i.d. de loi p inconnue. On souhaite ici construire un test qui permette de répondre à la question suivante : la loi des observations est-elle p^0 ? En termes statistiques, on souhaite tester

$$H_0 \ :\ p = p^0 \quad \text{contre} \quad H_1 \ :\ p \neq p^0.$$

C'est par exemple le cas dans l'exemple 10.1 du dé à six faces avec p^0 la loi uniforme sur $\{1,\dots,6\}$.

2.1. Mise en place du test. On note α le niveau du test (en général $\alpha = 0.1, 0.05, 0.01$). En utilisant le théorème 10.7 et la proposition 10.8, on obtient le comportement asymptotique de D_n :

$$D_n = n\sum_{i=1}^{k} \frac{\left(\frac{N_i(n)}{n} - p_i^0\right)^2}{p_i^0} \begin{cases} \xrightarrow[n\to\infty]{\mathscr{L}} \chi^2(k-1) & \text{si } H_0 \text{ est vraie,} \\ \xrightarrow[n\to\infty]{p.s.} +\infty & \text{sinon.} \end{cases}$$

Le niveau étant fixé, on choisit une région de rejet égale à $\{D_n \geq x_{1-\alpha}\}$ où $x_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi $\chi^2(k-1)$. La règle de décision est la suivante. On calcule D_n grâce aux observations. Si $D_n \geq x_{1-\alpha}$ alors on rejette H_0, sinon on ne rejette pas H_0.

EXERCICE 10.2 (Loi de Hardy-Weinberg (suite)). Mettre en place le test du χ^2 pour tester l'hypothèse que la population observée est en équilibre de Hardy-Weinberg de paramètre $1/2$.

EXERCICE 10.3 (Un peu de génétique). Deux cobayes (génération 0) de lignées pures (c'est-à-dire qu'ils ont les mêmes caractéristiques que leurs ancêtres depuis des générations) dont les pelages sont gris et lisse pour le premier et blanc et rêche pour le second ont donné une progéniture homogène au pelage gris et lisse. En croisant ces cobayes de la génération 1 entre eux, on a obtenu 64 descendants dont les pelages se répartissent de la manière indiquée dans le tableau suivant :

Pelage	gris et lisse	blanc et lisse	gris et rêche	blanc et rêche
Effectifs	33	13	15	3

Faisons les hypothèses de modélisation suivantes (on parle de modèle mendélien) :

— les cobayes sont des animaux diploïdes (ils possèdent deux versions d'un même chromosome) ;

— le gène responsable de la couleur du pelage est présent sous la forme de deux allèles, l'un dominant (A) associé au gris, l'autre récessif (a) associé au blanc ;

— le gène responsable de la texture du pelage est présent sous la forme de deux allèles, l'un dominant (B) associé au lisse, l'autre récessif (b) associé au rude ;

— les gènes responsables de la couleur et la texture du pelage sont sur des chromosomes différents ;

— chaque parent donne, au hasard, à son descendant une copie d'un des deux chromosomes de chaque paire, et ce indépendamment de l'autre parent.

1. Quel est le patrimoine génétique des cobayes de la génération 0 ? de la génération 1 ?
2. Montrer que la distribution théorique des cobayes de la génération 2 sous les hypothèses du modèle mendélien est $(9/16, 3/16, 3/16, 1/16)$ où l'on a rangé les individus selon leur phénotype : 1 pour gris et lisse, 2 pour gris et rêche, 3 pour blanc et lisse, 4 pour blanc et rêche.
3. Ces résultats expérimentaux sont-ils conforment au modèle mendélien au niveau 0.05 ?

2.2. Quelques adaptations. Le test du χ^2 s'appuie sur deux résultats asymptotiques (une convergence en loi si H_0 est vraie et une convergence presque sûre si H_1 est vraie). Or on ne dispose jamais que d'un nombre fini d'observations. Toute la question est de savoir si l'on a le droit de faire comme si la limite en loi était une égalité. En pratique, les livres recommandent la recette suivante : pour que le test soit valide, il faut que, pour tout

$i = 1, \dots, k$, np_i soit supérieur ou égal à 5. Si ce n'est pas le cas, il faut regrouper des classes à trop faible effectifs pour atteindre le seuil exigé.
Le test du χ^2 peut aussi être utilisé pour tester l'adéquation à une loi sur $\mathbb{N}$, sur $\mathbb{R}$ ou même sur $\mathbb{R}^d$. Pour cela, il suffit de découper l'espace en un nombre fini de classes et faire fonctionner la moulinette χ^2. Pour une loi sur $\mathbb{N}$, on utilise le découpage suivant :

$$\mathbb{N} = \{0\} \cup \cdots \cup \{k\} \cup \{l \geq k+1\}.$$

2.3. Remarque très importante. Les hypothèses d'un test peuvent être vues comme des parties de l'ensemble des mesures de probabilité sur un certain espace. Dans notre cas, H_0 représente un singleton et H_1 son complémentaire. Les tests du χ^2 (comme tous les tests non paramétriques) ne donnent vraiment d'informations que si l'hypothèse H_0 est rejetée. En effet, si H_0 n'est pas rejetée, il se peut très bien que ce se soit parce que la loi p de l'échantillon est dans H_1 mais tout près de p^0. Ceci est encore renforcé lorsque l'on est obligé de regrouper des classes faute d'un échantillon trop petit ou de créer des classes pour des lois continues : des tas de lois fourniront les mêmes vecteurs de probabilité sur l'ensemble fini. **On se sert d'un test non paramétrique pour invalider un modèle**. Si H_0 est rejetée, alors il faut changer de modèle. Si H_0 n'est pas rejetée c'est que le modèle (bien que simpliste, approximatif... et vraisemblablement faux) est satisfaisant. Pensez au modèle des lois de la gravitation, parfaitement capable de décrire la trajectoire des astres mais en défaut sur la description des particules élémentaires.

CONSEILS BIBLIOS 10.10. Pour le principe de la méthode et de nombreux exemples, on pourra consulter [**Tas85**], [**DRV01**] et [**CDF99**].

3. Test d'adéquation à une famille de lois

On souhaite ici mettre en place un test permettant de décider si la loi de l'échantillon appartient ou non à une famille de lois $(p(\theta))_{\theta\in\Theta}$ indexée par un paramètre θ à valeurs dans $\Theta \subset \mathbb{R}^d$. On suppose que l'on dispose de $\hat{\theta}_n$ l'estimateur du maximum de vraisemblance de θ. Il nous faut nous munir d'un théorème un peu plus puissant (et difficile) que le théorème 10.7. Sa démonstration est délicate car elle fait intervenir les propriétés des estimateurs du maximum de vraisemblance.

THÉORÈME 10.11. *Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de v.a. i.i.d. de loi $p(\theta)$ (avec $\theta \in \Theta$ inconnu) alors*

$$D'_n = \sum_{i=1}^{n} \frac{(N_i(n) - np_i(\hat{\theta}_n))^2}{np_i(\hat{\theta}_n)} \xrightarrow[n\to\infty]{\mathscr{L}} \chi^2(k-d-1).$$

REMARQUE 10.12. Le fait que la loi limite soit toujours une loi du χ^2 pourrait paraître miraculeux. Cela tient au fait que l'estimateur du maximum de vraisemblance possède dans une très grande majorité des cas le comportement suivant : $\hat{\theta}_n$ converge p.s. vers θ (on dit qu'il est fortement consistant) et $\sqrt{n}(\hat{\theta}_n - \theta)$ converge en loi vers une mesure gaussienne.

REMARQUE 10.13. On peut toutefois faire un commentaire qualitatif sur le nombre de degrés de liberté de la loi limite. En effet, il est naturel que celui-ci soit plus petit que $k-1$ puisque l'on compare les fréquences empiriques, non plus à une loi fixée, mais à la loi la plus vraisemblable dans une famille paramétrée au vu des observations. Il paraîtrait logique que D'_n soit d'une certaine façon plus petit que D_n. C'est bien ce qui se passe puisque la fonction de répartition de la loi $\chi^2(k-1)$ est inférieure à celle de la loi $\chi^2(k-d-1)$ (penser à l'interprétation en terme de somme de carrés de v.a. gaussiennes indépendantes). On dit que D'_n est stochastiquement inférieur à D_n.

COROLLAIRE 10.14. *Pour tester $H_0 : p \in \{p(\theta), \theta \in \Theta\}$ contre $H_1 : p \notin \{p(\theta), \theta \in \Theta\}$ on utilise la statistique D'_n qui a les comportements suivants selon que H_0 ou H_1 est vraie :*

$$D'_n = \sum_{i=1}^{n} \frac{(N_i(n) - np_i(\hat{\theta}_n))^2}{np_i(\hat{\theta}_n)} \begin{cases} \xrightarrow[n\to\infty]{\mathscr{L}} \chi^2(k-d-1) & \text{si } H_0 \text{ est vraie,} \\ \xrightarrow[n\to\infty]{p.s.} +\infty & \text{sinon.} \end{cases}$$

EXERCICE 10.4 (Loi de Hardy-Weinberg (suite et fin)). Tester le fait que la population observée est en équilibre de Hardy-Weinberg (on estimera tout d'abord le paramètre p de la loi).

EXERCICE 10.5. Pour 10000 fratries de quatre enfants (exactement), on a relevé le nombre de garçons :

nombre de garçons	0	1	2	3	4
effectifs	572	2329	3758	2632	709

On modélise les naissances successives de la façon suivante.

— les naissances sont indépendantes ;

— à chaque naissance, la livraison est un garçon ou une fille avec probabilités respectives θ et $1-\theta$.

1. Dans ce modèle, quelle est la loi p du nombre de garçons dans une fratrie de quatre enfants ?
2. Tester l'hypothèse $H_0 : p = \mathscr{B}(4, 1/2)$ contre $H_1 : p \neq \mathscr{B}(4, 1/2)$ au niveau 0,05.
3. Tester l'hypothèse $H_0 : p \in \{\mathscr{B}(4,\theta), \theta \in]0,1[\}$ contre $H_1 : p \notin \{\mathscr{B}(4,\theta), \theta \in]0,1[\}$.
4. Conclusion ?

EXERCICE 10.6. On étudie le nombre de connexion à Google pendant la durée de temps unitaire d'une seconde. On fait 200 mesures.

nombre de connexion par seconde	0	1	2	3	4	5	6	7	8	9	10	11
effectif empirique	6	15	40	42	37	30	10	9	5	3	2	1

Soit X la v.a. à valeurs dans $\mathbb{N}$ comptant le nombre de connexions par seconde. Peut-elle être considérée comme une loi de Poisson au niveau 5% ?

4. Test d'indépendance

Soit $(Y_1, Z_1), \dots, (Y_n, Z_n)$ des v.a. i.i.d. avec $(Y_l)_{1\le l\le n}$ à valeurs dans $\{1, \dots, r\}$ et $(Z_l)_{1\le l\le n}$ à valeurs dans $\{1, \dots, s\}$. La loi de (Y_1, Z_1) est donnée par une matrice $P = (p_{ij})_{1\le i\le r, 1\le j\le s}$ à coefficients positifs dont la somme vaut 1 : $p_{ij} = \mathbb{P}(Y_1 = i,\ Z_1 = j)$. Notons, pour $i = 1, \dots, r$ et $j = 1, \dots, s$,

$$p_{i\cdot} = \mathbb{P}(Y_1 = i) = p_{i1} + \cdots + p_{is} \quad \text{et} \quad p_{\cdot j} = \mathbb{P}(Z_1 = j) = p_{1j} + \cdots + p_{rj}.$$

Les v.a. Y_1 et Z_1 sont indépendantes si et seulement si, pour tous i et j, $p_{ij} = p_{i\cdot}p_{\cdot j}$. À partir de l'échantillon, définissons les v.a. suivantes :

$$N_{ij} = \text{Card}\{l = 1, \dots, n\ ;\ (X_l, Y_l) = (i, j)\}, \quad N_{i\cdot} = N_{i1} + \cdots + N_{is} \quad \text{et} \quad N_{\cdot j} = N_{1j} + \cdots + N_{rj}.$$

PROPOSITION 10.15. *Avec les notations ci-dessus,*

$$D_n = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{\left(N_{ij} - \frac{N_{i\cdot}N_{\cdot j}}{n}\right)^2}{\frac{N_{i\cdot}N_{\cdot j}}{n}} \begin{cases} \xrightarrow[n\to\infty]{\mathscr{L}} \chi^2((r-1)(s-1)) & \text{si } Y_1 \text{ et } Z_1 \text{ sont indépendantes,} \\ \xrightarrow[n\to\infty]{p.s.} +\infty & \text{sinon.} \end{cases}$$

REMARQUE 10.16. Cette proposition peut être vue comme un corollaire du théorème 10.11. En effet on teste l'adéquation de la loi du couple à la famille paramétrique des lois produit sur $\{1,\dots,r\}\times\{1,\dots,s\}$ en estimant les paramètres par la méthode du maximum de vraisemblance. Le nombre de paramètres estimés est $(r-1)+(s-1)$ puisque la donnée des $r-1$ premiers coefficients de la loi de Y donne le dernier (et idem pour Z) et que la donnée des lois marginales d'une loi produit détermine la loi du couple. Sous l'hypothèse d'indépendance de Y et Z, la statistique D_n converge donc vers une loi du χ^2 à $rs-(r+s+2)-1=(r-1)(s-1)$ degrés de liberté.

EXERCICE 10.7 (Yeux et cheveux). Depuis une terrasse de café ensoleillée, un statisticien en plein travail a noté les couleurs des yeux et des cheveux de 124 passants.

Yeux \ Cheveux	blonds	bruns	roux	noirs
bleus	25	9	7	3
gris	13	17	7	10
marrons	7	13	5	8

Les deux critères sont-ils indépendants au niveau 0.05 ?

CONSEILS BIBLIOS 10.17. Encore d'autres exemples dans [**Tas85**], [**DRV01**] et [**CDF99**].

5. Test d'homogénéité

Les tests du χ^2 permettent aussi de tester l'homogénéité de plusieurs échantillons. On étudie un caractère pouvant prendre k valeurs $A_1,\dots,A_k$ (ou k modalités ou à valeurs dans k classes). On dispose de l échantillons $E_1,\dots,E_l$ différents. Pour tout $i\in\{1,\dots,k\}$, pour tout $j\in\{1,\dots,l\}$, on connaît l'effectif observé O_{ij} de la valeur A_i dans l'échantillon E_j. On souhaite tester

$$H_0 \ : \text{ les échantillons sont issus de la même loi}$$

contre

$$H_1 \ : \text{ les échantillons n'ont pas même loi.}$$

La mise en place pratique du test est la même que pour le test d'indépendance. On définit

$$O_{i\cdot}=O_{i1}+\cdots+O_{il},\quad O_{\cdot j}=O_{1j}+\cdots+O_{kj}\quad\text{et}\quad n=\sum_{i=1}^{k}\sum_{j=1}^{l}O_{ij}\left(=\sum_{i=1}^{k}O_{i\cdot}=\sum_{j=1}^{l}O_{\cdot j}\right).$$

PROPOSITION 10.18. *Avec les notations ci-dessus,*

$$D_n=\sum_{i=1}^{k}\sum_{j=1}^{l}\frac{\left(O_{ij}-\frac{O_{i\cdot}O_{\cdot j}}{n}\right)^2}{\frac{O_{i\cdot}O_{\cdot j}}{n}}\begin{cases}\xrightarrow[n\to\infty]{\mathscr{L}}\chi^2((k-1)(l-1)) & \textit{si } H_0 \textit{ est vraie,}\\[2ex] \xrightarrow[n\to\infty]{p.s.}+\infty & \textit{sinon.}\end{cases}$$

EXERCICE 10.8. Les résultats des élèves de trois collèges notés A, B et C sont classés en trois catégories : insuffisant (I), moyen (M) et bon (B).

Collège \ Résultats	I	M	B
A	30	65	205
B	23	56	121
C	75	125	300

Peut-on admettre, au niveau 5%, que les résultats des trois collèges sont identiques ?

Chapitre 11

Modèle linéaire gaussien

Ce chapitre est construit comme une étude de cas : un modèle est présenté dans la section 1 puis sert de fil rouge au long de l'étude du modèle linéaire. On peut sans souci commencer la lecture par la section 2.

1. Travailler et capitaliser pour produire !

La production d'un pays dépend d'un grand nombre de variables appelées facteurs de production (le nombre d'employés, le capital disponible, les infrastructures, la météorologie, etc). On souhaite ici proposer un modèle simple expliquant la production par les deux seuls facteurs de production que sont le capital et le travail. On appelle produit marginal d'un facteur de production Φ la quantité

$$PM(\Phi) := \frac{\partial P}{\partial \Phi}.$$

HYPOTHÈSE 11.1. La production est une fonction du capital et du travail : il existe F telle que $P = F(K,T)$. Cette fonction F est appelée fonction de production.

HYPOTHÈSE 11.2. La production a des rendements d'échelle constants, c'est-à-dire que si on multiplie les quantités de capital et de travail par une constante positive, la production est multipliée par le même facteur : la fonction F est 1-homogène.

On a donc

$$F(K,T) = K\frac{\partial F}{\partial K}(K,T) + T\frac{\partial F}{\partial T}(K,T) = PM(K) \times K + PM(T) \times T.$$

HYPOTHÈSE 11.3. Supposons que les contributions du travail et du capital soient fixes : il existe $\alpha \in]0,1[$ tel que $PM(K) \times K = \alpha P$ et $PM(T) \times T = (1-\alpha)P$.

En posant $f(x) = F(x,1)$ (que l'on appelle fonction de production intensive), la relation $P = F(K,T)$ s'écrit $Q = f(R)$ où les quantités $Q = P/T$ et $R = K/T$ désignent respectivement la production et le capital par unité de travail.
On fait habituellement les hypothèses suivantes sur la forme intensive de la fonction de production.

HYPOTHÈSE 11.4. $\forall x > 0,\ f'(x) > 0$ et $f''(x) < 0$.

HYPOTHÈSE 11.5. $f(0) = 0$, $\lim_{x\to 0} f'(x) = +\infty$ et $\lim_{x\to+\infty} f'(x) = 0$.

Ces conditions sont naturelles : la production est une fonction croissante du capital mais la concavité signifie que la production augmente d'autant moins vite que le capital est élevé. De plus, le produit marginal du capital est très grand pour un stock de capital très bas et tend à s'annuler quand le stock de capital augmente.

THÉORÈME 11.6. *La fonction F vérifie les hypothèses précédentes si et seulement si*

$$F(K,T) = cK^{\alpha}T^{1-\alpha}.$$

DÉMONSTRATION. On montre que $PM(K) = f'(K/T)$ puis que f est solution de l'équation différentielle $xf'(x) = \alpha f(x)$, ce qui permet de conclure que F est de la forme annoncée. La réciproque est une simple vérification. □ Ce modèle est appelé modèle de Cobb et Douglas [1]. On se propose de confronter ce modèle à des données réelles avec l'idée de valider (ou invalider) la forme de la fonction F. On peut réécrire F sous la forme

$$P = c\left(\frac{K}{T}\right)^{\beta_2} T^{\beta_3}.$$

Les constantes c, β_2 et β_3 sont les paramètres du modèle. Si le modèle de Cobb-Douglas est valide, $\beta_2 \in [0,1]$ et $\beta_3 = 1$. L'étude de ces coefficients, en particulier leur variabilité au cours du temps ou d'un pays à l'autre est d'un grand intérêt pour les économistes. Il est également intéressant de savoir dans quelle mesure on peut affirmer que $\beta_3 = 1$.

On va prendre le logarithme dans cette équation. Si l'on pose

$$\begin{aligned}
&\beta_1 = \log(c)\\
&y = \log(P)\\
&x = (1, \log(K) - \log(T), \log(T)),
\end{aligned}$$

il vient

$$y = \sum_{j=1}^{3} \beta_j x_j \tag{27}$$

qui est un modèle linéaire. Voici les données utilisées dans l'étude de Cobb et Douglas (1928) :

Année	P	K	T	Année	P	K	T	Année	P	K	T
1899	100	100	100	1907	151	176	138	1915	189	266	154
1900	101	107	105	1908	126	185	121	1916	225	298	182
1901	112	114	110	1909	155	198	140	1917	227	335	196
1902	122	122	118	1910	159	208	144	1918	223	366	200
1903	124	131	123	1911	153	216	145	1919	218	387	193
1904	122	138	116	1912	177	226	152	1920	231	407	193
1905	143	149	125	1913	184	236	154	1921	179	417	147
1906	152	163	133	1914	169	244	149	1922	240	431	161

Bien entendu, le modèle (27) est bien trop rigide pour être confronté à de véritables données. On le remplace par le suivant :

$$y_i = \sum_{j=1}^{3} x_{ij}\beta_j + u_i, \quad i = 1, \ldots 24 \tag{28}$$

(x_{ij} est ici la j-ième variable explicative pour l'année i) où u_i représente l'écart au modèle idéal (27) pour l'individu i. Le modèle le plus simple est le **modèle linéaire gaussien** qui postule que les v.a. $(u_i)_{1\le i\le 24}$ sont des variables aléatoires gaussiennes centrées de même variance σ^2 ; plus σ est faible, plus on se rapproche de (27). Dans cette situation il y a quatre paramètres inconnus : β_1, β_2, β_3 et σ^2.

1. "A theory of production", *American Economic Review*, 18, 139-165, 1928.

2. Les données

Les données consistent en des variables observées y_i à valeurs dans $\mathbb{R}$ et des variables explicatives (ou régresseurs) x_i à valeurs dans $\mathbb{R}^p$ (vus comme des vecteurs-lignes), pour $i = 1,...n$. Chaque paire (y_i, x_i) représente une expérience (production, capital, travail pour une année donnée). On les écrit sous forme matricielle ainsi :

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 1 & x_{12} & \dots & x_{1p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n2} & \dots & x_{np} \end{pmatrix}.$$

On notera $X_{.j}$ la j-ième colonne de X. On convient généralement que le premier régresseur est la constante, mais ce n'est pas obligatoire. Les régresseurs sont pondérés par un vecteur (colonne) $(\beta_1^*, \dots, \beta_p^*)$. Le modèle s'exprime sous la forme

$$y_i = x_i\beta^* + u_i = \beta_1^* + \sum_{i=2}^{p} x_{ij}\beta_j^* + u_i, \quad i = 1,...n,$$

qui se réécrit de manière plus compacte (en notant u le vecteur colonne $(u_1, \dots, u_n)$)

$$y = X\beta^* + u. \tag{29}$$

Le modèle linéaire gaussien se réécrit simplement :

$$y \sim \mathcal{N}(X\beta^*, \sigma^2 Id).$$

La forme (29) est le modèle général. On suppose toujours le vecteur u centré de variance finie.

3. Estimation des coefficients de régression

On cherche tout d'abord à estimer le vecteur β^* grâce aux observations $(y_i, x_i)_{1\le i\le n}$.

DÉFINITION 11.7. Pour tout β, on définit l'erreur quadratique moyenne d'ajustement (ou erreur résiduelle) $S(\beta)$ comme

$$S(\beta)^2 = \frac{1}{n}\|y - X\beta\|^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i\beta)^2.$$

Le vecteur des coefficients de regression calculés aux moindres carrés ordinaires est

$$\hat{\beta} = \arg\min_{\beta} S(\beta).$$

PROPOSITION 11.8. *On suppose la matrice X^TX inversible. Alors $\hat{\beta}$ est donné par*

$$\hat{\beta} = (X^TX)^{-1}X^Ty.$$

DÉMONSTRATION. La fonction S est telle que

$$\frac{\partial S(\beta)^2}{\partial \beta_j} = \frac{1}{n}\sum_i 2x_{ij}(y_i - x_i\beta) = \frac{2}{n}(X^T(y - X\beta))_j$$

Son gradient s'annule si $X^T(y - X\beta) = 0$. La fonction S^2 admet un unique point critique. Elle est de plus convexe donc ce point est le minimum global de S^2. □

En remplaçant y par son expression (29), on obtient sans difficulté que, si les u_i sont centrés, décorrélés et de variance σ^2, alors $\hat{\beta}$ est un estimateur non biaisé de β^* et que si les u_i sont décorrélés et de variance σ^2, sa covariance est $(X^TX)^{-1}$.

EXERCICE 11.1. Dans le modèle de Cobb-Douglas, estimer le vecteur β .

DÉFINITION 11.9. Posons quelques termes de vocabulaire.

- Vecteur des valeurs ajustées (ou prédites) : $\hat{y} = X\hat{\beta}$, soit $\hat{y}_i = x_i\hat{\beta}$.
- Vecteur des résidus : $\hat{u} = y - \hat{y}$, soit $u_i = y_i - x_i\hat{\beta}$.

Dans toute la suite on supposera la matrice $X^T X$ inversible.

4. Décomposition de la variance et le coefficient de corrélation multiple R

On va voir que le vecteur $\hat{u}$ est orthogonal aux colonnes de X, ce qui a des conséquences importantes. On suppose ici que la première colonne de X est constituée de 1. À un vecteur de données $v = (v_1, \dots v_n)$ on associe sa moyenne et sa variance :

$$\bar{v} = \frac{1}{n}\sum_{i=1}^{n} v_i,$$
$$s_v^2 = \frac{1}{n}\sum_{i=1}^{n} v_i^2 - \bar{v}^2 = \frac{1}{n}\sum_{i=1}^{n}(v_i - \bar{v})^2.$$

PROPOSITION 11.10. *On a*

$$\hat{u} \perp x_{.j},\ j = 1, \dots, p, \quad \textit{et} \quad s_y^2 = s_{\hat{y}}^2 + s_{\hat{u}}^2.$$

La variabilité des données se décompose donc en la somme de deux termes, le premier ne dépendant que des variables explicatives et un deuxième terme étant la variabilité de ce que l'on ne peut expliquer. En d'autres termes, la connaissance du modèle et son estimation fait que l'incertitude sur y, qui est représentée par s_y, ne se réduit plus, pour quelqu'un qui a accès aux variables explicatives, qu'à $s_{\hat{u}}$.

DÉMONSTRATION. Pour l'orthogonalité, il suffit de vérifier que $X^T\hat{u} = 0$, mais :

$$X^T\hat{u} = X^T y - X^T X\hat{\beta} = X^T y - X^T X(X^T X)^{-1} X^T y = 0.$$

Pour la deuxième affirmation, rappelons que

$$y = \hat{y} + \hat{u}.$$

Mais l'orthogonalité de $\hat{u}$ à $\mathbf{1}$ (i.e. $x_{.1}$) implique que $\bar{\hat{u}} = 0$, et que donc $\bar{y} = \bar{\hat{y}} + \bar{\hat{u}} = \bar{\hat{y}}$. D'où

$$y - \bar{y} = (\hat{y} - \bar{\hat{y}}) + \hat{u}.$$

Le fait que $\hat{y} - \bar{\hat{y}}$ soit combinaison linéaire des colonnes de X et que $\hat{u}$ soit orthogonal à ces dernières implique que l'on peut appliquer le théorème de Pythagore, ce qui conduit à

$$\|y - \bar{y}\|^2 = \|\hat{y} - \bar{\hat{y}}\|^2 + \|\hat{u}\|^2$$

qui est le résultat cherché car $\bar{\hat{u}} = 0$. □

DÉFINITION 11.11. Le coefficient de détermination R^2 est le rapport de la « variance expliquée » par la « variance totale » :

$$R = \frac{s_{\hat{y}}}{s_y}, \quad R^2 = 1 - \frac{s_{\hat{u}}^2}{s_y^2}.$$

Plus R est proche de 1, plus le modèle est adéquat sur les données. Le coefficient R^2 est donc une mesure de comparaison entre la prédiction par le modèle et la prédiction par une constante, via les erreurs d'ajustement.

EXERCICE 11.2. Dans le modèle de Cobb-Douglas, que vaut le coefficient R ?

5. Loi des estimateurs sous le modèle gaussien

Rappelons que la covariance de deux vecteurs aléatoire U et V est la matrice formée des covariances $\mathrm{Cov}(U_i, V_j)$, soit $E[UV^T] - E[U]E[V]^T$. Rappelons également quelques propriétés de bases des vecteurs gaussiens.

PROPOSITION 11.12. *Si $X_1, \ldots X_n$ est une suite indépendante de variables aléatoires gaussiennes, alors $X = (X_1, \ldots X_n)$ un vecteur gaussien ; pour toute matrice A et tout vecteur m de taille adéquate, la variable $AX + m$ est encore un vecteur gaussien.*
Si X est un vecteur gaussien et A et B deux matrices de tailles adéquates et si AX et BX sont décorrélés, alors AX et BX sont indépendants.

PROPOSITION 11.13. *Sous l'hypothèse que $y \sim \mathcal{N}(X\beta^*, \sigma^2 Id)$, l'estimateur $\hat\beta$ suit la loi $\mathcal{N}(\beta^*, (X^TX)^{-1}\sigma^2)$. De plus*

$$\hat\sigma^2 = \frac{1}{n-p}\sum_i \hat u_i^2$$

est un estimateur non biaisé de σ^2 ; $(n-p)\hat\sigma^2/\sigma^2$ suit un χ^2_{n-p} et est indépendant de $\hat\beta$.

DÉMONSTRATION. En effet

$$\hat\beta = (X^TX)^{-1}X^Ty = (X^TX)^{-1}X^T(X\beta^* + u) = \beta^* + (X^TX)^{-1}X^Tu.$$

D'où sa loi. De plus,

$$\hat y = X\hat\beta = X\beta^* + X(X^TX)^{-1}X^Tu = X\beta^* + Pu.$$

Comme P est une matrice de projection orthogonale ($P = P^T = P^2$), les composantes $\hat y$ et $\hat u = y - \hat y = (I-P)u$ du vecteur gaussien $(\hat y, \hat u)$ sont décorrélées et donc indépendantes. Comme $\hat\beta = (X^TX)^{-1}X^T\hat y$, il s'ensuit que $\hat\beta$ et $\hat\sigma$ sont indépendants. L'espérance de $\hat\sigma^2$ se calcule comme suit, en notant Tr la trace et $Q = I - P$:

$$E[\|\hat u\|^2] = E[u^TQu] = E[\mathrm{Tr}(Quu^T)]) = \mathrm{Tr}(QE[uu^T]) = \sigma^2\mathrm{Tr}(Q) = (n-p)\sigma^2.$$

Ce qui conclut la démonstration. □

REMARQUE 11.14. Dans le modèle gaussien, l'estimateur $\hat\beta$ obtenu par la méthode des moindres carrés est également l'estimateur du maximum de vraisemblance. En effet, minimiser x^2 et maximiser e^{-x^2} fournit le même résultat…

6. Intervalles de confiance

Rappelons que la loi de Student de paramètre q est celle du rapport d'une gaussienne standard par une racine de χ^2_q indépendante. On désignera par t_q sa fonction quantile.
La loi donnée pour $(\hat\beta, \hat\sigma^2)$ implique qu'en posant $\hat\sigma_j^2 = \hat\sigma^2((X^TX)^{-1})_{jj}$:

$$\frac{\hat\beta_j - \beta_j^*}{\hat\sigma_j} = \frac{\hat\beta_j - \beta_j^*}{\sigma((X^TX)^{-1})_{jj}^{1/2}}\frac{\sigma}{\hat\sigma} \sim \mathrm{Student}(n-p).$$

Cette v.a. appartient donc avec probabilité $1-\alpha$ à l'intervalle $[-t_{n-p}(1-\alpha/2), t_{n-p}(1-\alpha/2)]$ et par conséquent on a l'intervalle de confiance

$$\beta_j^* \in [\hat\beta_j - \delta, \hat\beta_j + \delta], \quad \delta = \hat\sigma_j t_{n-p}(1-\alpha/2) \quad \text{avec probabilité de confiance } 1-\alpha.$$

Si 0 est en dehors de l'intervalle de confiance on dit que $\hat\beta_j$ est **significativement non-nul** au niveau de α.

EXERCICE 11.3. Dans le modèle de Cobb-Douglas, tester l'hypothèse $\beta_3 = 1$.

7. Intervalle de prédiction

Soit un nouvel individu de régresseur x'. On veut prédire y'. On sait

$$y' \sim \mathcal{N}(x'\beta^*, \sigma).$$

D'où l'approximation

$$y' \sim \mathcal{N}(x'\hat{\beta}, \hat{\sigma}).$$

qui conduit à des intervalles à 95% de la forme

$$y' \in [x'\hat{\beta} - 1.96\hat{\sigma}\ ,\ x'\hat{\beta} - 1.96\hat{\sigma}].$$

L'erreur sur σ et β^* étant d'ordre $1/\sqrt{n}$, l'approximation faite est généralement raisonnable.

Bibliographie

[BC07] B. BERCU et D. CHAFAÏ : *Modélisation stochastique et simulation. Cours et applications.* Dunod, 2007.

[BEK06] M. BENAÏM et N. EL KAROUI : *Promenade aléatoire : Chaînes de Markov et simulations ; martingales et stratégies.* Les Éditions de l'École Polytechnique, 2006.

[BL98] Ph. BARBE et M. LEDOUX : *Probabilités.* De la licence à l'agrégation. Belin, 1998. 24, 48

[BMP08] P. BALDI, L. MAZLIAK et P. PRIOURET : *Martingales et chaînes de Markov. Théorie élémentaire et exercices corrigés.* Hermann, 2008.

[CDF99] F. COUTY, J. DEBORD et D. FREDON : *Probabilités et statistiques.* Dunod, 1999. 110, 112

[CGCDM99] M. COTRELL, V. GENON-CATALOT, Ch. DUHAMEL et Th. MEYRE : *Exercices de probabilités.* Cassini, 1999. 91

[DRV01] J.-J. DAUDIN, S. ROBIN et C. VUILLET : *Statistique inférentielle.* Presses Universitaires de Rennes, 2001. 110, 112

[DZ98] A. DEMBO et O. ZEITOUNI : *Large deviations techniques and applications*, volume 38 de *Applications of Mathematics (New York).* Springer-Verlag, second édition, 1998. 104

[FF98] D. FOATA et A FUCHS : *Calcul des probabilités.* Dunod, 1998. 23

[FF02] D. FOATA et A FUCHS : *Processus stochastiques.* Dunod, 2002.

[Fis63] M. FISZ : *Probability theory and mathematical statistic.* Wiley, 1963. 92

[Mon82] A. MONFORT : *Cours de statistique mathématique.* Économia, 1982. 108

[Nor98] J. R. NORRIS : *Markov Chains.* Cambridge University Press, 1998.

[Rev97] D. REVUZ : *Probabilités.* Hermann, 1997.

[RS12] V. RIVOIRARD et G. STOLZ : *Statistique mathématique en action.* Vuibert, 2012.

[Sap90] G. SAPORTA : *Probabilités, analyse de données et statistique.* Éditions Technip, 1990. 98, 108

[Tas85] Ph. TASSI : *Méthodes statistiques.* Économia, 1985. 91, 97, 98, 108, 110, 112

[Wil91] D. WILLIAMS : *Probability with martingales.* Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991. 30

[Yca02] B. YCART : *Modèles et Algorithmes Markoviens*, volume 39 de *Mathématiques et Applications.* Springer, 2002. 13, 14