

Processus aléatoires et applications

Nils Berglund

▶ **To cite this version:**

Nils Berglund. Processus aléatoires et applications. DEA. Université d'Orléans, 2010, pp.99. cel-00439562v2

HAL Id: cel-00439562

https://cel.hal.science/cel-00439562v2

Submitted on 17 Nov 2010 (v2), last revised 31 Dec 2013 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Processus aléatoires et applications

Master 2 Pro de Mathématiques

Université d'Orléans

Nils Berglund

Version d'Octobre 2010

Table des matières

I Chaînes de Markov **1**

1 Chaînes de Markov sur un ensemble fini **3**
1.1 Exemples de chaînes de Markov 3
1.2 Définitions 9
1.3 Chaînes de Markov absorbantes 12
1.4 Chaînes de Markov irréductibles 15
1.5 Chaînes de Markov réversibles 22
1.6 Exercices 24

2 Chaînes de Markov sur un ensemble dénombrable **29**
2.1 Marches aléatoires 29
2.2 Généralités sur les processus stochastiques 33
2.3 Récurrence, transience et période 37
2.4 Distributions stationnaires 42
2.5 Convergence vers la distribution stationnaire 45
2.6 Exercices 47

3 Application aux algorithmes MCMC **51**
3.1 Méthodes Monte Carlo 51
3.2 Algorithmes MCMC 54
3.3 L'algorithme de Metropolis 56
3.4 Le recuit simulé 58

II Processus de sauts et files d'attente **61**

4 Rappels de probabilités **63**
4.1 Loi binomiale et loi de Poisson 63
4.2 Loi normale et loi exponentielle 66
4.3 Exercices 69

5 Le processus ponctuel de Poisson **71**
5.1 Construction par la fonction de comptage 72
5.2 Construction par les temps d'attente 74
5.3 Généralisations 76
5.4 Exercices 77

6 Processus markoviens de sauts **79**
6.1 Taux de transition 80
6.2 Générateur et équations de Kolmogorov 82
6.3 Distributions stationnaires 84
6.4 Exercices 87

7 Files d'attente **89**
7.1 Classification et notation de Kendall 89
7.2 Cas markoviens : Files d'attente M/M/s 90
7.3 Cas général : Files d'attente G/G/1 94
7.4 Exercices 97

Partie I

Chaînes de Markov

Chapitre 1

Chaînes de Markov sur un ensemble fini

1.1 Exemples de chaînes de Markov

Les chaînes de Markov sont intuitivement très simples à définir. Un système peut admettre un certain nombre d'états différents. L'état change au cours du temps discret. A chaque changement, le nouvel état est choisi avec une distribution de probabilité fixée au préalable, et ne dépendant que de l'état présent.

Exemple 1.1.1 (La souris dans le labyrinthe)**.** Une souris se déplace dans le labyrinthe de la figure 1.1. Initialement, elle se trouve dans la case 1. A chaque minute, elle change de case en choisissant, de manière équiprobable, l'une des cases adjacentes. Dès qu'elle atteint soit la nourriture (case 4), soit sa tanière (case 5), elle y reste.

On se pose alors les questions suivantes :

1. Avec quelle probabilité la souris atteint-elle la nourriture plutôt que sa tanière?
2. Au bout de combien de temps atteint-elle sa tanière ou la nourriture?

On peut essayer de répondre à ces questions en construisant un arbre décrivant les chemins possibles. Par exemple, il est clair que la souris se retrouve dans sa tanière au bout d'une minute avec probabilité 1/3. Sinon, elle passe soit dans la case 2, soit dans la case 3, et depuis chacune de ces cases elle a une chance sur deux de trouver la nourriture. Il y a donc une probabilité de 1/6 que la souris trouve la nourriture au bout de deux minutes. Dans les autres cas, elle se retrouve dans la case de départ, ce qui permet d'établir une formule de récurrence pour les probabilités cherchées.

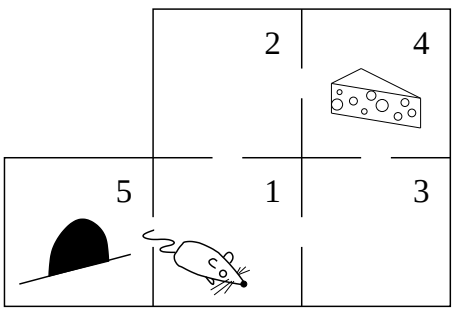


FIGURE 1.1. Le labyrinthe dans lequel vit la souris.

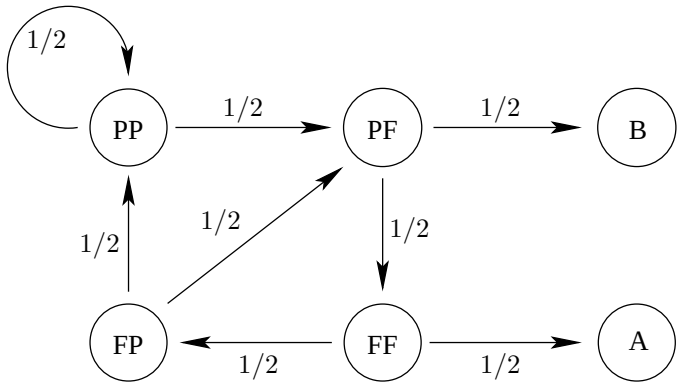


FIGURE 1.2. Graphe associé au jeu de Pile ou Face. Chaque symbole de deux lettres représente le résultat des deux derniers jets de pièce. Anatole gagne si la pièce tombe trois fois de suite sur Face, Barnabé gagne si la pièce tombe sur Pile-Face-Pile.

Cette manière de faire est toutefois assez compliquée, et devient rapidement impossible à mettre en oeuvre quand la taille du labyrinthe augmente. Dans la suite, nous allons développer une méthode plus efficace pour résoudre le problème, basée sur une représentation matricielle.

Exemple 1.1.2 (Jeu de Pile ou Face)**.** Anatole et Barnabé jouent à la variante suivante de Pile ou Face. Ils jettent une pièce de monnaie (parfaitement équilibrée) de manière répétée. Anatole gagne dès que la pièce tombe trois fois de suite sur Face, alors que Barnabé gagne dès que la suite Pile-Face-Pile apparaît.

On se pose les questions suivantes :

1. Avec quelle probabilité est-ce Anatole qui gagne le jeu?
2. Au bout de combien de jets de la pièce l'un des deux joueurs gagne-t-il?

La situation est en fait assez semblable à celle de l'exemple précédent. Un peu de réflexion montre que si personne n'a gagné au bout de n jets de la pièce, la probabilité que l'un des deux joueurs gagne au coup suivant ne dépend que des deux derniers résultats. On peut alors décrire le jeu par une chaîne de Markov sur l'ensemble

$$\mathcal{X} = \{\mathrm{PP}, \mathrm{PF}, \mathrm{FP}, \mathrm{FF}, \mathrm{A\ gagne}, \mathrm{B\ gagne}\}\ , \tag{1.1.1}$$

où par exemple PP signifie que la pièce est tombée sur Pile lors des deux derniers jets. On détermine alors les probabilités de transition entre les cinq états, et on retrouve un problème semblable à celui de la souris.

Exemple 1.1.3 (Modèle d'Ehrenfest)**.** C'est un système motivé par la physique, qui a été introduit pour modéliser de manière simple la répartition d'un gaz entre deux récipients. N boules, numérotées de 1 à N, sont réparties sur deux urnes. De manière répétée, on tire au hasard, de façon équiprobable, un numéro entre 1 et N, et on change d'urne la boule correspondante.

On voudrait savoir comment ce système se comporte asymptotiquement en temps :

1. Est-ce que la loi du nombre de boules dans chaque urne approche une loi limite?
2. Quelle est cette loi?
3. Avec quelle fréquence toutes les boules se trouvent-elles toutes dans la même urne?

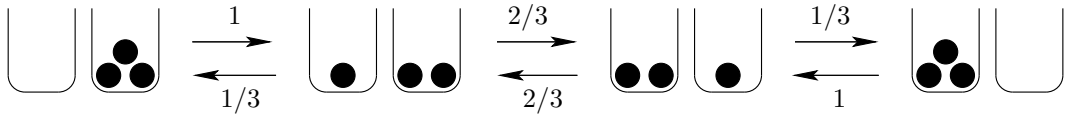


FIGURE 1.3. Le modèle d'urnes d'Ehrenfest, dans le cas de 3 boules.

On peut à nouveau décrire le système par une chaîne de Markov, cette fois sur l'espace des états $\mathcal{X} = \{0, 1, \ldots, N\}$, où le numéro de l'état correspond au nombre de boules dans l'urne de gauche, par exemple.

Exemple 1.1.4 (Texte aléatoires)**.** Voici trois "textes" générés de manière aléatoire :

A. YxUV,luUqHCLvE?,MRiKaoiWjyhg nEYKrMFD!rUFUy.qvW;e:FfIN.udbBdo!, ZpGwTEOFcA;;RrSMvPjA'Xtn.vP?JNZA;xWP, Cm?;i'MzLqVsAnlqHyk,ghDT :PwSwrnJojRhVjSe?dFkoVRN!MTfiFeemBXITdj m.h d'ea;Jkjx,XvHIBPfFT s I'SLcSX;'X!S, ODjX.eMoLnQttneLnNE!qGRgCJ:BuYAauJXoOCCsQkLcyPO MuIKLRtSm;PNpFfp'PfgvIJNrUr t l aXtIA?;TPhPxU:,ZmVGr,,'DIjqZDBY DrkPRiKDYRknDhivt;, LYXDuxNKpjegMvrtfz:JpNTDj'LFmHzXxotRM u.iya UUrgZRcA QmCZffwsNWhddBUPAhJIFJvs.CkKFLJoXef;kCnXrv'uWNcpULYsnl Kg OURmysAnxFjHawwsSpM H;PWPsMaFYLMFyvRWOjbdPlLQIaaspNZkuO'Ns.I jEXO,lxQ'GS;n;H:DH:VWJN :t'JMTUVpKCkVZ'NyKJMGilbQFXEgDEcWxMBiyo ybRIWIAC deMJnnL;SBAZ?:.UuGnC:B.!IBUT,pT?tyHHLICvN, mKZgwlMJOJd HHobua;KU.;kADVM?jr'v.SCq:hZLR;lqkmLkhn:ajhBM,gKexDAro,HlczWTv cFmNPt.MudUWPO, sTrWIJdgjoiJd.:d;CpJkJCW;FIRnpMGa;umFysOMAqQtmT pPaYZKtOFYppeE.KFX?SuvcbaDrQ XECelD;cfoQKf?'jCTUaISS;fV:gqoWfSq k:Tf!YuPBANtKhewiNg'ImOFs:UhcExmBjsAaMhBf UVP, 'dcFk;gxJMQGyXI; nVwwfWxS:YXQMELEIObTJiilUYSlOsg.gCqlrN:nEU:irHM'nOLXWUbJLTU re' kk vAwMgt'KgWSxwxqJe,z'OBCrnoIshSCDIZirla,rWNPkc?UgZm GOBX.QylY jOtuF

B. nsunragetnetelpnlac. pieln tJmends d e.imnqu caa aneezsconns re.tc oml d e c, paeisfuaul irt ssna l df.ieulat a ese t hre edn ro m eeel slsplotasstp etuoMeiiseeaenemzeaeuqpeer enuoco sfehnnir p ts 'mpisu qrd iraLp nFetesa,opQeey rieeaduset MuuisecG il e m ru daeiafasousfnircot i eeedracev ever.nsn iaeulu!,mtel lpa rdbjdide tolr'murunlr bteaaua ieasilureseuavrmoce ntvqm qnurnaunsa.mraayVarinanr eumsu cnponf ciuo .pssre elreeY snrrq aani psu oqoddaiaaomrssloe'avia,loei va eroltrsurdeduuoe ffusir 'th'nilt has,sllu oooe tee ?eoxaea slsii i u edtvsear e,Mesatnd o o rvdocaeagiua apugiqn rclt smtee.te, gceade etsn e v in eag ent so ra te, oi seGndd i eeet!dii e ese nanu d sp ul afeen aqelonens ssisaaoe cs eectadegotuudlru i 'c, uuuuts 'tt , dir atermdmuciqedn esovsioieieerxdroie mqso,es rrvteen,r dtei xcalrionuaae e vtmplsz miuqa u aboir br gmcdexptedn pEua't vm vnic eeren ereaa,eegeta u rss nlmxomas ea nsbnt s,eEpeteae teiasbo cd ee tu em ue quee en, sd eeneepeot

C. cesalu'act, bouleuivoie melarous die ndant leuvoiblue poit pesois deuntaciroverchu llie e lle s r lerchar, laisueuayaissabes vet s cuetr i as, rdetite se d'iretie, de.. nendoules, le pablur e d ! copomouns ppait limmix a r aux urars laie Le r lercret ce c. n'are four nsirepapole pa vr s, nte le efit. itesit, le faun e ju estatusuet usoin prcilaisanonnout ssss l tosesace cole sientt, dent pontrtires. e, l mentoufssss chat Laneus c Chontrouc Ce e. Et deses j'ecci uleus mmon s mauit paga lanse l cont ciquner e c Cha s l'a Jes des

s'erattrlunt es de sacouen erends. ve e quns som'a aisajouraite eux lala pour ! a levionible plaint n ss, danetrc ponce con du lez, l danoit, dirvecs'u ce ga vesai : chleme eesanl Pa chiontotes anent fomberie vaud'untitez e esonsan t a ! bondesal'is Ilaies, vapa e ! Lers jestsiee celesu unallas, t. ces. ta ce aielironi mmmileue cecoupe et dennt vanen A la ajole quieet, scemmu tomtemotit me aisontouimmet Le s Prage ges peavoneuse ! blec douffomurrd ntis.. rur, ns ablain i pouilait lertoipr ape. leus icoitth me e e, poiroia s. ! atuepout somise e la as

Il est clair qu'aucun de ces textes n'a de signification. Toutefois, le texte B. semble moins arbitraire que le texte A., et C. paraît moins éloigné d'un texte français que B. Il suffit pour cela d'essayer de lire les textes à haute voix.

Voici comment ces textes ont été générés. Dans les trois cas, on utilise le même alphabet de 60 lettres (les 26 minuscules et majuscules, quelques signes de ponctuation et l'espace).

1. Pour le premier texte, on a simplement tiré au hasard, de manière indépendante et avec la loi uniforme, des lettres de l'alphabet.
2. Pour le second texte, on a tiré les lettres de manière indépendante, mais pas avec la loi uniforme. Les probabilités des différentes lettres correspondent aux fréquences de ces lettres dans un texte de référence français (en l'occurence, un extrait du *Colonel Chabert* de Balzac). Les fréquences des différentes lettres du texte aléatoire sont donc plus naturelles, par exemple la lettre e apparaît plus fréquemment (dans 13% des cas) que la lettre z (0.2%).
3. Pour le dernier texte, enfin, les lettres n'ont pas été tirées de manière indépendante, mais dépendant de la lettre précédente. Dans le même texte de référence que précédemment, on a déterminé avec quelle fréquence la lettre a est suivie de a (jamais), b (dans 3% des cas), et ainsi de suite, et de même pour toutes les autres lettres. Ces fréquences ont ensuite été choisies comme probabilités de transition lors de la génération du texte.

Ce procédé peut facilement être amélioré, par exemple en faisant dépendre chaque nouvelle lettre de plusieurs lettres précédentes. Mais même avec une seule lettre précédente, il est remarquable que les textes engendrés permettent assez facilement de reconnaître la langue du texte de référence, comme en témoignent ces deux exemples:

D. deser Eld s at heve tee opears s cof shan; os wikey coure tstheevons irads; Uneer I tomul moove t nendoot Heilotetateloreagis his ud ang l ars thine br, we tinond end cksile: hersest tear, Sove Whey tht in t ce tloour ld t as my aruswend Ne t nere es alte s ubrk, t r s; penchike sowo Spotoucthistey psushen, ron icoowe l Whese's oft Aneds t aneiksanging t ungl o whommade bome, ghe; s, ne. torththilinen's, peny. d llloine's anets but whsto a It hoo tspinds l nafr Aneve powit tof f I afatichif m as tres, ime h but a wrove Les des wined orr; t he ff teas be hende pith hty ll ven bube. g Bube d hitorend tr, Mand nd nklichis okers r whindandy, Sovede brk f Wheye o edsucoure, thatovigh ld Annaix; an eer, andst Sowery looublyereis isthalle Base whon ey h herotan wict of les, h tou dends m'dys h Wh on'swerossictendoro whaloclocotolfrrovatel aled ouph rtrsspok, ear'sustithimiovelime From alshis ffad, Spake's wen ee: hoves aloorth erthis n t Spagovekl stat hetubr tes, Thuthiss oud s hind t s potrearall's ts dofe [1]

E. dendewoch wich iere Daf' lacht zuerckrech, st, Gebr d, Bes. jenditerullacht, keie Un! etot' in To sendenus scht, ubteinraben Qun Jue die m arun dilesch d e Denuherelererufein ien.

[1] Texte de référence: Quelques sonnets de Shakespeare.

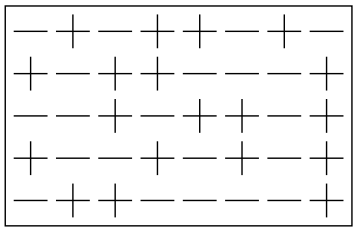


FIGURE 1.4. Une configuration du modèle d'Ising en dimension $d = 2$.

seurdan s ire Zein. es min? dest, in. maur as s san Gedein it Ziend en desckruschn kt vontimelan. in, No Wimmmschrstich vom delst, esichm ispr jencht sch Nende Buchichtannnlin Sphrr s Klldiche dichwieichst. ser Bollesilenztoprs uferm e mierchlls aner, d Spph! wuck e ing Erenich n sach Men. Sin s Gllaser zege schteun d, Gehrstren ite Spe Kun h Umischr Ihngertt, ms ie. es, bs de! ieichtt f; Ginns Ihe d aftalt veine im t'seir; He Zicknerssolanust, fllll. mmichnennd wigeirdie h Zierewithennd, wast naun Wag, autonbe Wehn eietichank We dessonindeuchein ltichlich bsch n, Ichritienstam Lich uchodigem Din eieiers die it f tlo nensseicichenko Mechtarzaunuchrtzubuch aldert; l von. fteschan nn ih geier Schich Geitelten Deichst Fager Zule fer in vischtrn; Schtih Un Hit ach, dit? at ichuch Eihra! Hich g ure vollle Est unvochtelirn An [2]

Cela donne, inversément, une méthode assez économique permettant à une machine de déterminer automatiquement dans quelle langue un texte est écrit.

Exemple 1.1.5 (Le modèle d'Ising)**.** Comme le modèle d'Ehrenfest, ce modèle vient de la physique, plus particulièrement de la physique statistique. Il est sensé décrire un ferroaimant, qui a la propriété de s'aimanter spontanément à température suffisamment basse. On considère une partie (connexe) Λ du réseau $\mathbb{Z}^d$ (d étant la dimension du système, par exemple 3), contenant N sites. A chaque site, on attache un "spin" (une sorte d'aimant élémentaire), prenant valeurs $+1$ ou -1. Un choix d'orientations de tous les spins s'appelle une configuration, c'est donc un élément de l'espace de configuration $\mathcal{X} = \{-1, 1\}^\Lambda$. A une configuration σ, on associe l'énergie

$$H(\sigma) = -\sum_{<i,j>\in\Lambda} \sigma_i\sigma_j - h\sum_{i\in\Lambda} \sigma_i \ . \tag{1.1.2}$$

Ici, la notation $< i, j >$ indique que l'on ne somme que sur les paires de spins plus proches voisins du réseau, c'est–à–dire à une distance 1. Le premier terme est donc d'autant plus grand qu'il y a de spins voisins différents. Le second terme décrit l'interaction avec un champ magnétique extérieur h. Il est d'autant plus grand qu'il y a de spins opposés au champ magnétique.

Un principe de base de la physique statistique est que si un système est en équilibre thermique à température T, alors il se trouve dans la configuration σ avec probabilité proportionnelle à $\mathrm{e}^{-\beta H(\sigma)}$ (mesure de Gibbs), où $\beta = 1/T$. A température faible, le système privilégie les configurations de basse énergie, alors que lorsque la température tend vers l'infini, toutes les configurations deviennent équiprobables.

L'aimantation totale de l'échantillon est donnée par la variable aléatoire

$$m(\sigma) = \sum_{i\in\Lambda} \sigma_i \ , \tag{1.1.3}$$

[2]Texte de référence: Un extrait du *Faust* de Goethe.

et son espérance vaut

$$\mathbb{E}(m) = \frac{\sum_{\sigma\in\mathcal{X}} m(\sigma)\,\mathrm{e}^{-\beta H(\sigma)}}{\sum_{\sigma\in\mathcal{X}} \mathrm{e}^{-\beta H(\sigma)}}\,. \tag{1.1.4}$$

L'intérêt du modèle d'Ising est qu'on peut montrer l'existence d'une transition de phase, en dimension d supérieure ou égale à 2. Dans ce cas il existe une température critique en-dessous de laquelle l'aimantation varie de manière discontinue en fonction de h dans la limite $N \to \infty$. Pour des températures supérieures à la valeur critique, l'aimantation est continue en h.

Si l'on veut déterminer numériquement l'aimantation, il suffit en principe de calculer la somme (1.1.4). Toutefois, cette somme comprend 2^N termes, ce qui croît très rapidement avec la taille du système. Par exemple pour un cube de $10 \times 10 \times 10$ spins, le nombre de termes vaut 2^{1000}, ce qui est de l'ordre de 10^{300}. Un ordinateur calculant 10^{10} termes par seconde mettrait beaucoup plus que l'âge de l'univers à calculer la somme.

Une alternative est d'utiliser un algorithme dit de Metropolis. Au lieu de parcourir toutes les configurations possibles de $\mathcal{X}$, on n'en parcourt qu'un nombre limité, de manière bien choisie, à l'aide d'une chaîne de Markov. Pour cela, on part dans une configuration initiale σ, puis on transforme cette configuration en retournant un spin choisi au hasard. Plus précisément, on n'opère cette transition qu'avec une certaine probabilité, qui dépend de la différence d'énergie entre les configurations de départ et d'arrivée. L'idée est que si les probabilités de transition sont bien choisies, alors la chaîne de Markov va échantilloner l'espace de configuration de telle manière qu'il suffira de lui faire parcourir une petite fraction de toutes les configurations possibles pour obtenir une bonne approximation de l'aimantation $\mathbb{E}(m)$. Les questions sont alors

1. De quelle manière choisir ces probabilités de transition?
2. Combien de pas faut-il effectuer pour approcher $\mathbb{E}(m)$ avec une précision donnée?

Exemple 1.1.6 (Le problème du voyageur de commerce)**.** C'est un exemple classique de problème d'optimisation. Un voyageur de commerce doit visiter N villes, en revenant à son point de départ après être passé exactement une fois par chaque ville. Comment choisir l'ordre des villes de manière à minimiser la longueur du circuit?

La difficulté est que le nombre de circuits possibles croît extrêmement vite avec le nombre N de villes, beaucoup plus vite qu'exponentiellement. En effet, il y a $N!$ permutations possibles de l'ordre des villes. Si l'on ne tient compte ni de la ville de départ, ni du sens de parcours, il reste $(N-1)!/2$ circuits possibles. Calculer les longueurs de tous ces circuits devient irréalisable dès que N dépasse 20 environ.

On peut tenter de trouver une solution approchée par approximations successives. Partant d'un circuit initial, on le modifie légèrement, par exemple en échangeant deux villes. Si cette modification raccourcit la longueur du circuit, on continue avec le circuit modifié. Si elle le rallonge, par contre, on rejette la modification et on en essaie une autre.

Le problème avec cette méthode est que le système peut se retrouver piégé dans un minimum local, qui est très différent du minimum global recherché de la longueur. On peut en effet se retrouver "bloqué" dans un circuit plus court que tous ses voisins (obtenus en permutant deux villes), mais une permutation de plus de deux villes pourrait raccourcir le circuit.

Une variante plus efficace de cette méthode est celle du *recuit simulé*. Dans ce cas, on ne rejette pas toutes les modifications qui allongent le circuit, mais on les accepte avec une certaine probabilité, qui décroît avec l'allongement. De cette manière, le processus peut s'échapper du minimum local et a une chance de trouver un minimum plus profond. La

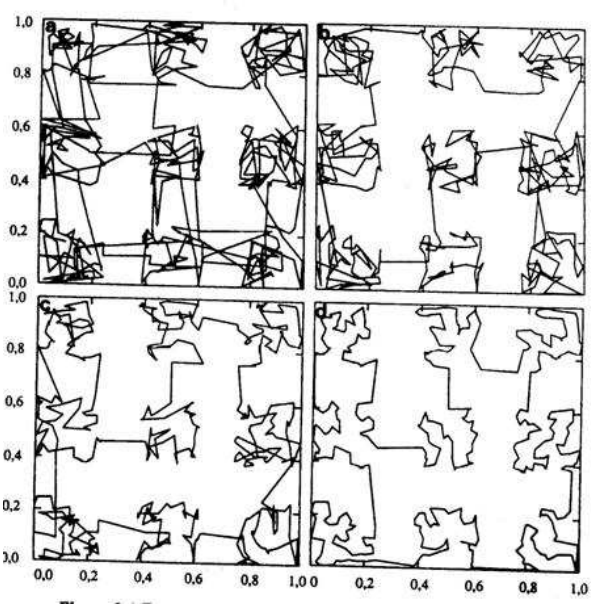


FIGURE 1.5. Approximations successives de la solution du problème du voyageur de commerce par la méthode du recuit simulé (tiré de l'article original : S. Kirkpatrick, C. Gelatt et M. Vecchi, *Optimization by Simulated Annealing*, Science, 220 (1983), pp. 671–680, copyright 1983 by A.A.A.S.)).

terminologie vient de la métallurgie : Dans un alliage, les atomes des différents métaux sont disposés de manière plus ou moins régulière, mais avec des imperfections. Moins il y a d'imperfections, plus l'alliage est solide. En réchauffant et refroidissant plusieurs fois l'alliage, on donne aux atomes la possibilité de se réarranger de manière plus régulière, c'est-à-dire en diminuant l'énergie potentielle.

A nouveau, on se pose les questions suivantes :

1. Comment choisir les probabilités d'acceptation des modifications?
2. Comment la probabilité de s'approcher à une certaine distance du minimum cherché dépend-elle de la longueur de la simulation?

1.2 Définitions

Définition 1.2.1. *Soit N un entier strictement positif. Une matrice P de taille $N \times N$ est une* matrice stochastique *si ses éléments de matrice $p_{ij} = (P)_{ij}$ satisfont*

$$0 \leqslant p_{ij} \leqslant 1 \qquad \forall i, j \tag{1.2.1}$$

et

$$\sum_{j=1}^{N} p_{ij} = 1 \qquad \forall i \,. \tag{1.2.2}$$

On vérifie facilement que si P est une matrice stochastique, alors toutes ses puissances P^2, P^3, ... sont encore des matrices stochastiques.

Les éléments p_{ij} vont définir les probabilités de transition de la chaîne de Markov de l'état i vers l'état j.

Définition 1.2.2. *Soit $\mathcal{X} = \{1, \dots, N\}$ un ensemble fini et P une matrice stochastique de taille N. Une* chaîne de Markov *sur $\mathcal{X}$ de matrice de transition P est une suite $(X_0, X_1, X_2, \dots)$ de variables aléatoires à valeurs dans $\mathcal{X}$, satisfaisant la* propriété de Markov

$$\begin{aligned}\mathbb{P}\big\{X_n = j \,\big|\, X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \dots, X_0 = i_0\big\} &= \mathbb{P}\big\{X_n = j \,\big|\, X_{n-1} = i_{n-1}\big\} \\ &= p_{i_{n-1}j} \end{aligned} \tag{1.2.3}$$

pour tout temps $n \geqslant 1$ et tout choix (i_0, i_1, i_{n-1}, j) d'éléments de $\mathcal{X}$. La loi de X_0, que nous noterons ν, est appelée la distribution initiale *de la chaîne.*

Pour s'assurer que cette définition fait bien sens, il faut vérifier que les X_n construits comme ci-dessus sont bien des variables aléatoires, c'est-à-dire que la somme sur tous les $j \in \mathcal{X}$ des probabilités $\mathbb{P}\{X_n\} = j$ vaut 1. Ceci est immédiat par récurrence sur n. Si les n variables X_0, …, X_{n-1} sont des variables aléatoires, alors on a :

$$\begin{aligned}\sum_{j\in\mathcal{X}} \mathbb{P}\{X_n = j\} &= \sum_{j\in\mathcal{X}} \mathbb{P}\{X_n = j, X_{n-1} \in \mathcal{X}, \dots, X_0 \in \mathcal{X}\} \\ &= \sum_{j\in\mathcal{X}} \sum_{i_{n-1}\in\mathcal{X}} \cdots \sum_{i_0\in\mathcal{X}} \mathbb{P}\{X_n = j, X_{n-1} = i_{n-1}, \dots, X_0 = i_0\} \\ &= \sum_{i_{n-1}\in\mathcal{X}} \underbrace{\sum_{j\in\mathcal{X}} p_{i_{n-1}j}}_{=1} \underbrace{\sum_{i_{n-2}\in\mathcal{X}} \cdots \sum_{i_0\in\mathcal{X}} \mathbb{P}\{X_{n-1} = i_{n-1}, \dots, X_0 = i_0\}}_{=\mathbb{P}\{X_{n-1}=i_{n-1}, X_{n-2}\in\mathcal{X}, \dots, X_0\in\mathcal{X}\} = \mathbb{P}\{X_{n-1}=i_{n-1}\}} \\ &= \mathbb{P}\{X_{n-1} \in \mathcal{X}\} = 1\,. \end{aligned} \tag{1.2.4}$$

Notation 1.2.3. *Si la loi initiale ν est fixée, nous noterons souvent $\mathbb{P}_\nu$ la loi de la chaîne de Markov associée. Si ν est concentrée en un seul site i ($\nu = \delta_i$), on notera la loi de la chaîne $\mathbb{P}_i$ au lieu de $\mathbb{P}_{\delta_i}$. Enfin nous écrirons parfois $X_{[n,m]}$ au lieu de $(X_n, X_{n+1}, \dots, X_m)$, et $\mathbb{P}\{X_{[n,m]} = i_{[n,m]}\}$ au lieu de $\mathbb{P}\{X_n = i_n, X_{n+1} = i_{n+1}, \dots, X_m = i_m\}$. $X_{[n,m]}$ est appelé la* trajectoire *de la chaîne entre les temps n et m.*

Exemple 1.2.4. Dans le cas de l'exemple 1.1.1 de la souris dans le labyrinthe, la matrice de transition est donnée par

$$P = \begin{pmatrix} 0 & 1/3 & 1/3 & 0 & 1/3 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{1.2.5}$$

Dans le cas de l'exemple 1.1.2 du jeu de Pile ou Face, la matrice de transition vaut

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{1.2.6}$$

Voici tout d'abord une caractérisation d'une chaîne de Markov en termes de ses trajectoires.

Théorème 1.2.5. *Soit $\{X_n\}_{n\in\mathbb{N}}$ une suite de variables aléatoires à valeurs dans $\mathcal{X}$, ν une distribution de probabilité sur $\mathcal{X}$, et P une matrice stochastique. Alors $\{X_n\}_{n\in\mathbb{N}}$ est une chaîne de Markov de matrice de transition P et de distribution initiale ν si et seulement si pour tout $n \geqslant 0$, et pour tout choix de $i_0, i_1, \ldots, i_n$ d'éléments de $\mathcal{X}$, on a*

$$\mathbb{P}\big\{X_{[0,n]} = i_{[0,n]}\big\} = \nu_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n} \,. \tag{1.2.7}$$

Démonstration.

$\Rightarrow$: Par récurrence sur n. C'est clair pour $n = 0$. Si c'est vrai pour n, alors

$$\begin{aligned}\mathbb{P}\big\{X_{[0,n+1]} = i_{[0,n+1]}\big\} &= \mathbb{P}\big\{X_{n+1} = i_{n+1} \,\big|\, X_{[0,n]} = i_{[0,n]}\big\}\mathbb{P}\big\{X_{[0,n]} = i_{[0,n]}\big\} \\ &= p_{i_n i_{n+1}} \nu_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n} \,. \end{aligned} \tag{1.2.8}$$

$\Leftarrow$: Par définition de la probabilité conditionnelle, on a

$$\mathbb{P}\big\{X_n = i_n \,\big|\, X_{[0,n-1]} = i_{[0,n-1]}\big\} = \frac{\mathbb{P}\{X_{[0,n]} = i_{[0,n]}\}}{\mathbb{P}\{X_{[0,n-1]} = i_{[0,n-1]}\}} = p_{i_{n-1} i_n} \,, \tag{1.2.9}$$

la dernière égalité suivant de (1.2.7). □

L'équation (1.2.7) donne la probabilité de la trajectoire $X_{[0,n]}$. Le résultat suivant montre que la propriété de Markov reste vraie pour des trajectoires : l'évolution sur un intervalle de temps $[n+1, m]$ ne dépend que de l'état au temps n, et pas de la trajectoire passée de la chaîne.

Proposition 1.2.6. *Si $\{X_n\}_{n\in\mathbb{N}}$ est une chaîne de Markov sur $\mathcal{X}$, alors pour tous temps $n < m \in \mathbb{N}$, tous $i_n \in \mathcal{X}$, $A \subset \mathcal{X}^n$ et $B \subset \mathcal{X}^{m-n}$ tels que $\mathbb{P}\{X_n = i_n, X_{[0,n-1]} \in A\} > 0$ on a*

$$\mathbb{P}\big\{X_{[n+1,m]} \in B \,\big|\, X_n = i_n, X_{[0,n-1]} \in A\big\} = \mathbb{P}\big\{X_{[n+1,m]} \in B \,\big|\, X_n = i_n\big\} \,. \tag{1.2.10}$$

Démonstration. On a

$$\begin{aligned}\mathbb{P}\big\{X_{[n+1,m]} \in B \,\big|\, X_n = i_n, X_{[0,n-1]} \in A\big\} &= \frac{\mathbb{P}\big\{X_{[n+1,m]} \in B, X_n = i_n, X_{[0,n-1]} \in A\big\}}{\mathbb{P}\big\{X_n = i_n, X_{[0,n-1]} \in A\big\}} \\ &= \frac{\displaystyle\sum_{i_{[n+1,m]}\in B} \sum_{i_{[0,n-1]}\in A} \nu_{i_0} p_{i_0 i_1} \cdots p_{i_{m-1} i_m}}{\displaystyle\sum_{i_{[0,n-1]}\in A} \nu_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}} \\ &= \sum_{i_{[n+1,m]}\in B} p_{i_n i_{n+1}} \cdots p_{i_{m-1} i_m} \,, \end{aligned} \tag{1.2.11}$$

qui ne dépend pas de l'ensemble A. En particulier, choisissant $A = \mathcal{X}^n$ dans l'égalité ci-dessus, on trouve

$$\mathbb{P}\big\{X_{[n+1,m]} \in B \,\big|\, X_n = i_n\big\} = \sum_{i_{[n+1,m]}\in B} p_{i_n i_{n+1}} \cdots p_{i_{m-1} i_m} \,, \tag{1.2.12}$$

d'où le résultat, en comparant (1.2.11) et (1.2.12). □

Un cas particulier important est celui où $B = \mathcal{X}^{m-n-1} \times \{j\}$, c'est-à-dire qu'on s'intéresse à toutes les trajectoires se terminant en $i_m = j$ au temps m. Dans ce cas, la relation (1.2.12) donne

$$\mathbb{P}\big\{X_m = j \;\big|\; X_n = i\big\} = \sum_{i_{n+1}\in\mathcal{X}} \cdots \sum_{i_{m-1}\in\mathcal{X}} p_{ii_{n+1}} p_{i_{n+1}i_{n+2}} \dots p_{i_{m-1}j} \,. \tag{1.2.13}$$

Par définition du produit matriciel, la somme ci-dessus n'est autre que l'élément de matrice (i, j) de la matrice P^{m-n}, que nous noterons $p_{ij}^{(m-n)}$. On remarquera que le membre de droite de (1.2.13) ne dépend que de la différence $m - n$. On a donc pour tout $0 < n < m$

$$\mathbb{P}\big\{X_m = j \;\big|\; X_n = i\big\} = p_{ij}^{(m-n)} = \mathbb{P}\big\{X_{m-n} = j \;\big|\; X_0 = i\big\} \tag{1.2.14}$$

(*propriété des incréments stationnaires*). Enfin, pour tout $n > 0$,

$$\mathbb{P}_\nu\big\{X_m = j\big\} = \sum_{i\in\mathcal{X}} \mathbb{P}_\nu\big\{X_m = j, X_0 = i\big\} = \sum_{i\in\mathcal{X}} \nu_i p_{ij}^{(m)} \,. \tag{1.2.15}$$

La matrice P^m donne donc les probabilités de transition en m pas.

Exemple 1.2.7. Pour la matrice de transition (1.2.5) de la souris dans le labyrinthe, on trouve

$$P^2 = \begin{pmatrix} 1/3 & 0 & 0 & 1/3 & 1/3 \\ 0 & 1/6 & 1/6 & 1/2 & 1/6 \\ 0 & 1/6 & 1/6 & 1/2 & 1/6 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} . \tag{1.2.16}$$

La $i^{\text{ème}}$ ligne de la matrice correspond aux probabilités d'être dans les différents états si la souris est partie de la case i. Ainsi, si elle est partie de la case 1, on voit qu'elle se retrouve au temps 2 soit en 1, soit dans sa tanière, soit auprès de la nourriture, à chaque fois avec même probabilité 1/3. Si elle est partie de l'une des cases 2 ou 3, elle a une chance sur 2 d'avoir trouvé la nourriture au temps 2, et une chance sur 6 de se retrouver dans l'une des cases 2, 3 ou 5.

1.3 Chaînes de Markov absorbantes

Définition 1.3.1. *On dit qu'un état $j \in \mathcal{X}$ est* atteignable *depuis un autre état $i \in \mathcal{X}$, et on note $i \rightsquigarrow j$, s'il existe un temps $n \in \mathbb{N}$ tel que $p_{ij}^{(n)} > 0$, c'est-à-dire que partant de i, on atteint j avec probabilité positive en un nombre fini de pas. On notera $i \sim j$ si on a à la fois $i \rightsquigarrow j$ et $j \rightsquigarrow i$.*

On vérifie facilement que la relation $\rightsquigarrow$ est réflexive et transitive, et que $\sim$ est une relation d'équivalence.

Définition 1.3.2. *Un état $i \in \mathcal{X}$ est dit* absorbant *si $p_{ii} = 1$ (et donc nécessairement $p_{ij} = 0$ pour tout $j \neq i$). Une chaîne de Markov est dite* absorbante *s'il existe, pour tout état de $\mathcal{X}$, un état absorbant atteignable depuis cet état.*

Dans le reste de cette section, nous allons considérer des chaînes absorbantes avec $r \geqslant 1$ états absorbants. Les exemples 1.1.1 de la souris et 1.1.2 du jeu de Pile ou Face sont des exemples de chaînes absorbantes avec deux états absorbants.

Nous conviendrons de numéroter les états de manière à placer d'abord les $q = N - r$ états non absorbants, et ensuite les r états absorbants. La matrice de transition prend alors la *forme canonique*

$$P = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix} , \tag{1.3.1}$$

où Q est une matrice de taille $q \times q$, R est une matrice de taille $q \times r$, 0 désigne la matrice nulle de taille $r \times q$, et I la matrice identité de taille r. Il est facile de montrer par récurrence que

$$P^n = \begin{pmatrix} Q^n & [I + Q + \cdots + Q^{n-1}]R \\ 0 & I \end{pmatrix} . \tag{1.3.2}$$

Proposition 1.3.3. *Soit P la matrice de transition d'une chaîne de Markov absorbante, écrite sous forme canonique. Alors*

1. *On a*
$$\lim_{n\to\infty} Q^n = 0 . \tag{1.3.3}$$
2. *La matrice $I - Q$ est inversible, et son inverse vaut*
$$\big[I - Q\big]^{-1} = \sum_{k=0}^{\infty} Q^k . \tag{1.3.4}$$

Démonstration.

1. Soit $i \leqslant q$ un état non absorbant. L'élément de matrice $(Q^n)_{ij}$ de Q^n est la probabilité de se trouver dans l'état non absorbant j, après n pas, partant de i. Par conséquent, $(Q^n)_{ij}$ est inférieur ou égal à la probabilité de ne pas avoir atteint d'état absorbant en n pas. Soit
$$m_i = \min\{n \geqslant 1 \colon \exists k > q,\, (P^n)_{ik} > 0\} \tag{1.3.5}$$
le nombre minimal de pas nécessaire à atteindre un état absorbant k depuis i. Soit
$$p_i = \mathbb{P}_i\{X_{m_i} \leqslant q\} < 1 \tag{1.3.6}$$
la probabilité de ne pas atteindre d'état absorbant en m_i pas, partant de i. Soit enfin
$$M = \max_{i=1,\dots,q} m_i , \qquad p = \max_{i=1,\dots,q} p_i . \tag{1.3.7}$$
Alors la probabilité de ne pas atteindre d'état absorbant en M pas, partant de n'importe quel état non absorbant, est bornée par p. Il suit que la probabilité de ne pas atteindre d'état absorbant en Mn pas est bornée par p^n. Cette probabilité tend vers 0 lorsque n tend vers l'infini. La probabilité de ne pas être absorbé après un nombre arbitraire m de pas étant une fonction décroissante de m, elle tend nécessairement vers 0. Par conséquent, $(Q^n)_{ij}$ tend vers zéro lorsque n tend vers l'infini, pour tout $j \in \{1, \dots, q\}$.
2. Supposons qu'il existe un vecteur x tel que $Qx = x$. Dans ce cas on a
$$x = Qx = Q^2x = \cdots = Q^nx = \cdots = \lim_{n\to\infty} Q^nx = 0 , \tag{1.3.8}$$
ce qui montre que Q n'admet pas la valeur propre 1. Par conséquent, $I - Q$ est inversible. Enfin, comme
$$\big[I - Q\big] \sum_{k=0}^{n} Q^k = I - Q^{n+1} \to I \qquad \text{lorsque } n \to \infty , \tag{1.3.9}$$
on obtient la relation (1.3.4) en multipliant à gauche par $[I - Q]^{-1}$. □

Nous noterons F la matrice $[I-Q]^{-1}$, et nous l'appellerons la *matrice fondamentale* de la chaîne. La relation (1.3.2) montre que

$$\lim_{n\to\infty} P^n = \begin{pmatrix} 0 & FR \\ 0 & I \end{pmatrix} . \tag{1.3.10}$$

Le fait que Q^n tend vers zéro traduit donc le fait que la probabilité d'absorption tend vers 1 lorsque le temps tend vers l'infini. La matrice $B = FR$ devrait représenter les probabilités de transition, dans la limite des temps infinis, entre états non absorbants et absorbants. Ceci est confirmé par le résultat suivant.

Théorème 1.3.4. *Soit F la matrice fondamentale d'une chaîne de Markov absorbante.*

1. *L'élément de matrice f_{ij} de F est l'espérance du nombre de passages en j partant de i :*
$$f_{ij} = \mathbb{E}_i\Big(\sum_{n\geqslant 0} 1_{\{X_n=j\}}\Big) . \tag{1.3.11}$$
2. *Soit $\tau = \inf\{n \geqslant 1 \colon X_n > q\}$ la variable aléatoire donnant le temps jusqu'à absorption. Alors*
$$\mathbb{E}_i(\tau) = \sum_{j=1}^{q} f_{ij} . \tag{1.3.12}$$
3. *Les éléments de matrice b_{ik} de $B = FR$ donnent les probabilités d'être absorbés dans les différents états :*
$$b_{ik} = \mathbb{P}_i\big\{X_\tau = k\big\} . \tag{1.3.13}$$

Démonstration.

1. Soit la variable de Bernoulli $Y_{n,j} = 1_{\{X_n=j\}}$. On a $\mathbb{E}_i(Y_{n,j}) = \mathbb{P}_i\{Y_{n,j}=1\} = (Q^n)_{ij}$, et donc
$$\mathbb{E}_i\Big(\sum_{n\geqslant 0} Y_{n,j}\Big) = \sum_{n\geqslant 0}(Q^n)_{ij} = (F)_{ij} = f_{ij} . \tag{1.3.14}$$
2. Sommant la relation ci-dessus sur tous les états non absorbants, on a
$$\begin{aligned}\sum_{j=1}^{q} f_{ij} &= \mathbb{E}_i\Big(\sum_{n\geqslant 0}\sum_{j=1}^{q} Y_{n,j}\Big) = \mathbb{E}_i\Big(\sum_{n\geqslant 0} 1_{\{X_n\leqslant q\}}\Big) \\ &= \sum_{n\geqslant 0}\mathbb{P}_i\{\tau > n\} = \sum_{n\geqslant 0} n\mathbb{P}_i\{\tau = n\} = \mathbb{E}_i(\tau) .\end{aligned} \tag{1.3.15}$$
3. En décomposant sur les valeurs possibles n de $\tau - 1$, puis sur les valeurs possibles j de X_n,
$$\begin{aligned}\mathbb{P}_i\{X_\tau = k\} &= \sum_{n\geqslant 0}\mathbb{P}_i\{X_n \leqslant q, X_{n+1} = k\} \\ &= \sum_{n\geqslant 0}\sum_{j=1}^{q}\mathbb{P}_i\{X_n = j\}\mathbb{P}\{X_{n+1} = k | X_n = j\} \\ &= \sum_{n\geqslant 0}\sum_{j=1}^{q}(Q^n)_{ij}(R)_{jk} = \sum_{n\geqslant 0}(Q^n R)_{ik} = (FR)_{ik} .\end{aligned} \tag{1.3.16}$$

□

Exemple 1.3.5. Pour l'exemple 1.1.2 du jeu de Pile ou Face, les matrices Q et R sont données par

$$Q = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \end{pmatrix} , \qquad R = \begin{pmatrix} 0 & 0 \\ 0 & 1/2 \\ 0 & 0 \\ 1/2 & 0 \end{pmatrix} . \tag{1.3.17}$$

On calcule alors la matrice fondamentale

$$F = [I - Q]^{-1} = \begin{pmatrix} 7/3 & 4/3 & 1/3 & 2/3 \\ 1/3 & 4/3 & 1/3 & 2/3 \\ 4/3 & 4/3 & 4/3 & 2/3 \\ 2/3 & 2/3 & 2/3 & 4/3 \end{pmatrix} , \tag{1.3.18}$$

et la matrice donnant les probabilités d'absorption

$$B = FR = \begin{pmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \\ 1/3 & 2/3 \\ 2/3 & 1/3 \end{pmatrix} . \tag{1.3.19}$$

Ainsi, partant de l'un des états PP, PF ou FP, Anatole gagne avec probabilité 1/3, et Barnabé gagne avec probabilité 2/3. Partant de l'état FF, c'est Barnabé qui gagne avec probabilité 1/3, et Anatole qui gagne avec probabilité 2/3. Comme personne ne gagne lors des deux premiers jets, et que les quatre états PP, PF, FP et FF sont atteints avec la même probabilité, il faut choisir la distribution initiale $\nu = (1/4, 1/4, 1/4, 1/4, 0, 0)$. Par conséquent, Anatole gagne le jeu avec probabilité

$$\mathbb{P}_\nu\{X_\tau = \text{“A gagne”}\} = \sum_{i=1}^{4} \nu_i b_{i1} = \frac{5}{12} . \tag{1.3.20}$$

Quelle est la durée moyenne du jeu? La relation (1.3.12) montre que la somme des éléments de la ligne i de F donne l'espérance du temps d'absorption partant de i, donc par exemple $\mathbb{E}_1(\tau) = 14/3$. En moyennant sur la distribution initiale, on trouve $\mathbb{E}_\nu(\tau) = 23/6$. La durée moyenne du jeu est donc de $2 + 23/6 = 35/6$, soit un peu moins de 6 jets de pièce.

1.4 Chaînes de Markov irréductibles

Définition 1.4.1. *Une chaîne de Markov est dite* irréductible *ou* ergodique *si* $i \sim j$ $\forall i, j \in \mathcal{X}$. *La chaîne est dite* régulière *s'il existe une puissance* P^n *de* P *dont tous les éléments sont strictement positifs.*

Une chaîne de Markov régulière est nécessairement irréductible, car tout état est atteignable depuis tout autre en n pas au plus. La réciproque n'est pas vraie, car dans la définition de l'irréductibilité on n'a pas spécifié le nombre de pas.

Exemple 1.4.2. La chaîne décrivant le modèle d'Ehrenfest est irréductible. En effet, quel que soit le nombre de boules dans l'urne de gauche, on peut atteindre tout autre état en déplaçant au plus N boules d'une urne à l'autre. Cependant, la chaîne n'est pas régulière. En effet, comme à chaque pas de temps on déplace exactement une boule, le nombre de boules dans l'urne de gauche sera alternativement pair et impair. Par conséquent, chaque élément de matrice des puissance P^n sera nul pour un n sur deux.

Définition 1.4.3. *Pour un sous-ensemble $A \subset \mathcal{X}$, on appelle* temps de premier passage *de la chaîne dans A la variable aléatoire*

$$\tau_A = \inf\{n > 0 \colon X_n \in A\} \, . \tag{1.4.1}$$

Dans le cas où $A = \{i\}$ consiste en un seul point, nous écrirons aussi τ_i au lieu de $\tau_{\{i\}}$.

Une différence importante entre chaînes absorbantes et irréductibles est que ces dernières finissent toujours par revenir dans chacun de leurs états.

Proposition 1.4.4. *Pour une chaîne de Markov irréductible sur un ensemble fini $\mathcal{X}$, le temps de premier passage en tout sous-ensemble $A \subset \mathcal{X}$ est fini presque sûrement :*

$$\mathbb{P}\big\{\tau_A < \infty\big\} := \lim_{n\to\infty} \mathbb{P}\big\{\tau_A \leqslant n\big\} = 1 \, . \tag{1.4.2}$$

Démonstration. Considérons une autre chaîne de Markov de matrice de transition $\widehat{P}$, obtenue à partir de la chaîne de départ en rendant absorbants les états de A :

$$\hat{p}_{ij} = \begin{cases} \delta_{ij} & \text{si } i \in A \, , \\ p_{ij} & \text{sinon} \, . \end{cases} \tag{1.4.3}$$

Les trajectoires de la chaîne initiale et de la chaîne modifiée coïncident jusqu'au temps τ_A. Il suffit donc de montrer (1.4.2) pour la chaîne absorbante. Or dans ce cas, le résultat est une conséquence directe de la Proposition 1.3.3. En effet, la probabilité $\mathbb{P}_i\{\tau_A > n\}$ de ne pas avoir été absorbé jusqu'au temps n, partant de i, est donnée par la somme des $(Q^n)_{ij}$ sur les $j \in A$, qui tend vers 0. □

Il est important de remarquer que le résultat ci-dessus n'est plus forcément vrai lorsque $\mathcal{X}$ n'est pas fini! Nous reviendrons sur ce point dans le chapitre suivant.

Nous étudions maintenant de plus près les chaînes de Markov régulières. Leur propriété principale est le fait remarquable suivant.

Théorème 1.4.5. *Soit P la matrice de transition d'une chaîne de Markov régulière. Alors il existe des nombres $\pi_1, \ldots, \pi_N > 0$, dont la somme vaut 1, tels que*

$$\lim_{n\to\infty} P^n = \Pi := \begin{pmatrix} \pi_1 & \pi_2 & \ldots & \pi_N \\ \pi_1 & \pi_2 & \ldots & \pi_N \\ \vdots & \vdots & & \vdots \\ \pi_1 & \pi_2 & \ldots & \pi_N \end{pmatrix} \, . \tag{1.4.4}$$

De plus, le vecteur ligne $\pi = (\pi_1, \pi_2, \ldots, \pi_N)$ satisfait

$$\pi P = \pi \, . \tag{1.4.5}$$

Démonstration. Si la chaîne n'a qu'un état, le résultat est immédiat, donc nous pouvons admettre que $N \geqslant 2$. Supposons pour commencer que tous les éléments de P^k sont strictement positifs pour $k = 1$, et soit $d > 0$ le plus petit élément de P. Alors $d \leqslant 1/2$, puisque $Nd \leqslant 1$. Soit y un vecteur colonne tel que

$$0 \leqslant m_0 \leqslant y_i \leqslant M_0 \qquad \forall i \in \{1, \ldots, N\} \, . \tag{1.4.6}$$

Soit $z = Py$. La plus grande valeur possible d'une composante z_j de z est obtenue si $y^T = (m_0, M_0, \dots, M_0)$ et $p_{k1} = d$. Dans ce cas, la somme des $N-1$ derniers éléments de la ligne j de P vaut $1-d$, et par conséquent $z_j = dm_0 + (1-d)M_0$. On a donc nécessairement

$$z_j \leqslant dm_0 + (1-d)M_0 =: M_1 \qquad \forall j \in \{1, \dots, N\} \; . \tag{1.4.7}$$

Un raisonnement similaire montre que

$$z_j \geqslant dM_0 + (1-d)m_0 =: m_1 \qquad \forall j \in \{1, \dots, N\} \; . \tag{1.4.8}$$

Par conséquent, nous avons $m_1 \leqslant z_j \leqslant M_1$, avec

$$M_1 - m_1 = (1-2d)(M_0 - m_0) \; . \tag{1.4.9}$$

De plus, on voit facilement que $m_1 \geqslant m_0$ et $M_1 \leqslant M_0$. Après n itérations, les composantes de $P^n y$ seront comprises entre des nombres m_n et M_n, satisfaisant

$$M_n - m_n = (1-2d)^n (M_0 - m_0) \tag{1.4.10}$$

et

$$m_0 \leqslant m_1 \leqslant \cdots \leqslant m_n \leqslant M_n \leqslant \cdots \leqslant M_1 \leqslant M_0 \; . \tag{1.4.11}$$

Les suites $\{m_n\}_{n \geqslant 1}$ et $\{M_n\}_{n \geqslant 1}$ sont donc adjacentes, et convergent vers une même limite u. On a donc

$$\lim_{n \to \infty} P^n y = \begin{pmatrix} u \\ \vdots \\ u \end{pmatrix} \; , \tag{1.4.12}$$

où u dépend de y. Appliquons cette relation sur les vecteurs de base $e_1, \dots, e_N$. Il existe des nombres π_i tels que

$$\lim_{n \to \infty} P^n e_i = \begin{pmatrix} \pi_i \\ \vdots \\ \pi_i \end{pmatrix} \tag{1.4.13}$$

pour chaque i. Or $P^n e_i$ est la i-ème colonne de P^n, nous avons donc prouvé la relation (1.4.4). Par ailleurs, comme dans le cas $y = e_i$ on a $m_0 = 0$ et $M_0 = 1$, la relation (1.4.8) donne $m_1 \geqslant d$, donc $\pi_i \geqslant d$. Par conséquent, tous les π_i sont strictement positifs. La somme des π vaut 1 car toute puissance de P est une matrice stochastique.

Enfin, pour montrer (1.4.5), il suffit d'observer que $\Pi P = \lim_{n \to \infty} P^{n+1} = \Pi$. Chaque ligne de cette équation matricielle est équivalente à (1.4.5).

Considérons finalement le cas où tous les éléments de P^k sont positifs pour un $k > 1$. Le raisonnement ci-dessus peut être répété pour montrer que les composantes de $P^{kn} z$ sont comprises entre deux bornes m_n et M_n satisfaisant $M_{kn} - m_{kn} = (1-2d)^n (M_0 - m_0)$. Pour les étapes intermédiaires, on peut appliquer (1.4.7) et (1.4.8) avec $d = 0$ pour conclure que $M_{n+1} - m_{n+1} \leqslant M_n - m_n$ pour tout n. Cela montre à nouveau que $M_n - m_n$ tend vers zéro. □

Remarque 1.4.6. Le résultat précédent montre que toute matrice stochastique régulière P admet 1 comme valeur propre. En fait, nous aurions déjà pu le remarquer avant, car la définition (1.2.2) d'une matrice stochastique (quelconque) implique que $P\mathbf{1} = \mathbf{1}$, où $\mathbf{1} = (1, 1 \dots, 1)^T$. La relation (1.4.4) montre en plus que pour une matrice stochastique régulière, la valeur propre 1 est simple, et toutes les autres valeurs propres sont strictement inférieures à 1 en module.

Remarque 1.4.7. On déduit facilement de l'expression (1.4.4) que

$$\Pi^2 = \Pi \, . \tag{1.4.14}$$

Une matrice satisfaisant cette relation est appelée un *projecteur*. En l'occurence, Π projette tout vecteur ligne ν sur un multiple de π (c'est-à-dire $\nu\Pi \parallel \pi$), et tout vecteur colonne x sur un multiple du vecteur $\mathbf{1}$ (c'est-à-dire $\Pi x \parallel \mathbf{1}$). En particulier, si ν est une distribution de probabilité (donc $\sum_i \nu_i = 1$), alors on vérifie que $\nu\Pi = \pi$.

Le vecteur ligne π a plusieurs propriétés importantes :

1. Par (1.4.4) on a, $\forall i, j \in \mathcal{X}$,

$$\lim_{n\to\infty} \mathbb{P}_i\big\{X_n = j\big\} = \lim_{n\to\infty} (P^n)_{ij} = \pi_j \, . \tag{1.4.15}$$

   π décrit donc la distribution de probabilité asymptotique de la chaîne, qui est indépendante de l'état initial.
2. L'équation (1.4.5) implique que pour tout temps n,

$$\mathbb{P}_\pi\big\{X_n = j\big\} = \sum_{i\in\mathcal{X}} \pi_i (P^n)_{ij} = (\pi P^n)_j = \pi_j \, , \tag{1.4.16}$$

   ce qui motive la définition suivante.

Définition 1.4.8. *La distribution de probabilité $\pi = (\pi_1, \pi_2, \ldots, \pi_N)$ satisfaisant la relation* (1.4.4) *est appelée* distribution stationnaire *(ou* invariante*) de la chaîne de Markov.*

Enfin, on a le résultat général suivant :

Théorème 1.4.9. *Pour une chaîne régulière et toute distribution initiale ν, on a*

$$\lim_{n\to\infty} \mathbb{P}_\nu\{X_n = j\} = \pi_j \quad \forall j \in \mathcal{X} \, . \tag{1.4.17}$$

Démonstration. Une première preuve très simple consiste à observer que la loi asymptotique de X_n est donnée par le vecteur ligne

$$\lim_{n\to\infty} \nu P^n = \nu\Pi = \pi \, , \tag{1.4.18}$$

en vertu de la Remarque 1.4.7. Il est toutefois plus intéressant de présenter une autre preuve, très élégante, due à Doeblin. Considérons une autre chaîne de Markov, définie sur l'espace $\mathcal{X} \times \mathcal{X}$. Ses probabilités de transition $P^\star$ sont données par

$$p^\star_{(i,j),(k,l)} = p_{ik} p_{jl} \, . \tag{1.4.19}$$

Nous supposons que la distribution initiale de cette chaîne est une mesure produit $\rho = \nu \otimes \pi$, c'est-à-dire que

$$\rho((i,j)) = \nu_i \pi_j \quad \forall (i,j) \in \mathcal{X} \times \mathcal{X} \, . \tag{1.4.20}$$

Nous dénotons cette chaîne par $\{(X_n, Y_n)\}_{n\geqslant 0}$. Par construction, les variables aléatoires X_0 et Y_0 sont indépendantes. Il suit alors de la définition (1.4.19) des probabilités de transition que X_n et Y_n sont indépendantes pour tout n, et que les suites $\{X_n\}_{n\geqslant 0}$ et $\{Y_n\}_{n\geqslant 0}$ sont en fait deux chaînes de Markov sur $\mathcal{X}$ de matrice de transition P, et de distributions initiales respectivement données par ν et π.

La matrice de transition $P^\star$ est également régulière : il suffit de se convaincre que les éléments de matrice des puissances $(P^\star)^n$ sont donnés par des produits $p_{ik}^{(n)} p_{jl}^{(n)}$. Considérons alors l'ensemble

$$A = \{(i,i)\colon i \in \mathcal{X}\} \subset \mathcal{X} \times \mathcal{X} \,. \tag{1.4.21}$$

Le temps de premier passage τ_A peut aussi s'écrire

$$\tau_A = \inf\{n > 0\colon X_n = Y_n\} \,. \tag{1.4.22}$$

Nous prétendons que les deux chaînes ont la même loi pour $n \geqslant \tau_A$. Plus précisément,

$$\mathbb{P}_\rho\{X_n = j, \tau_A \leqslant n\} = \mathbb{P}_\rho\{Y_n = j, \tau_A \leqslant n\} \qquad \forall j \in \mathcal{X} \,, \forall n \geqslant 0 \,. \tag{1.4.23}$$

Pour montrer cela, nous introduisons un nouveau processus $\{Z_n\}_{n\in\mathbb{N}}$ défini par

$$Z_n = \begin{cases} X_n & \text{pour } n \leqslant \tau_A \,, \\ Y_n & \text{pour } n > \tau_A \,. \end{cases} \tag{1.4.24}$$

On vérifie par un calcul direct, en décomposant sur les valeurs possibles de τ_A, que

$$\mathbb{P}_\rho\{[Z_{[0,n]} = i_{[0,n]}\} = \nu_{i_0} \prod_{m=1}^{n} p_{i_{m-1} i_m} \tag{1.4.25}$$

pour tout $n \geqslant 0$ et tout choix de $i_{[0,n]} \in \mathcal{X}^{n+1}$. Par le Théorème 1.2.5, il suit que $\{Z_n\}_{n\in\mathbb{N}}$ est une chaîne de Markov de distribution initiale ν et matrice de transition P, et est donc égale en loi à $\{X_n\}_{n\in\mathbb{N}}$. Ceci prouve (1.4.23). Finalement, on a

$$\begin{aligned} \mathbb{P}_\nu\{X_n = j\} &= \mathbb{P}_\rho\{X_n = j, \tau_A \leqslant n\} + \mathbb{P}_\rho\{X_n = j, \tau_A > n\} \,, \\ \pi_j = \mathbb{P}_\pi\{Y_n = j\} &= \mathbb{P}_\rho\{Y_n = j, \tau_A \leqslant n\} + \mathbb{P}_\rho\{Y_n = j, \tau_A > n\} \,. \end{aligned} \tag{1.4.26}$$

En prenant la différence et en utilisant (1.4.23), il vient

$$\begin{aligned} \big|\mathbb{P}_\nu\{X_n = j\} - \pi_j\big| &\leqslant \big|\mathbb{P}_\rho\{X_n = j, \tau_A > n\} - \mathbb{P}_\rho\{Y_n = j, \tau_A > n\}\big| \\ &\leqslant 2\mathbb{P}_\rho\{\tau_A > n\} \,. \end{aligned} \tag{1.4.27}$$

Or cette dernière quantité tend vers zéro lorsque $n \to \infty$, puisque τ_A est fini presque sûrement en vertu de la Proposition 1.4.4. □

Remarque 1.4.10. La preuve de Doeblin permet aussi d'exprimer la vitesse de convergence vers la distribution stationnaire à l'aide du temps τ_A introduit dans la preuve. En effet, en sommant la première ligne de (1.4.27) sur tous les $j \in \mathcal{X}$, on obtient

$$\sum_{j\in\mathcal{X}} \big|\mathbb{P}_\nu\{X_n = j\} - \pi_j\big| \leqslant 2\mathbb{P}_{\nu\otimes\pi}\{\tau_A > n\} \,. \tag{1.4.28}$$

Le membre de gauche peut être considéré comme la distance ℓ_1 entre la distribution de X_n et la distribution stationnaire. Ce genre d'argument est appelé un argument de couplage.

Nous revenons maintenant au cas général de chaînes de Markov irréductibles. Dans ce cas, la loi de X_n ne converge pas nécessairement vers une loi π donnée. Toutefois, une partie des résultats précédents reste vraie :

Proposition 1.4.11. *Soit P la matrice de transition d'une chaîne de Markov irréductible. Alors P admet 1 comme valeur propre simple. L'unique vecteur propre à gauche π de P pour la valeur propre 1 tel que $\sum_i \pi_i = 1$ sera à nouveau appelé la* distribution stationnaire *de la chaîne.*

Démonstration. Considérons la matrice stochastique $Q = \frac{1}{2}[P + I]$. Soit

$$m = \max_{i,j \in \mathcal{X}} \bigl\{ \min\{n \geqslant 1 \colon p_{ij}^{(n)} > 0\} \bigr\} \; . \tag{1.4.29}$$

Considérons la matrice

$$Q^m = \frac{1}{2^m} \biggl[I + \binom{m}{1} P + \binom{m}{2} P^2 + \cdots + \binom{m}{m-1} P^{m-1} + P^m \biggr] \; . \tag{1.4.30}$$

Pour tout couple (i, j), il existe un terme de cette somme dont l'élément de matrice (i, j) soit strictement positif. Comme tous les autres éléments de matrice sont non-négatifs, on conclut que $(Q^m)_{i,j} > 0$. Par conséquent, Q est la matrice de transition d'une chaîne régulière. Par le théorème 1.4.5, il existe une unique distribution de probabilité π telle que $\pi Q = \pi$, ce qui implique $\frac{1}{2}[\pi + \pi P] = \pi$, donc $\pi P = \pi$. □

Exemple 1.4.12. On vérifie facilement par calcul direct que la distribution stationnaire du modèle d'Ehrenfest est binomiale de paramètre 1/2 : $\nu_i = 2^{-N} \binom{N}{i}$. Nous verrons plus loin une interprétation plus intuitive de ce résultat.

Quelle est l'interprétation de la distribution stationnaire? D'une part, nous savons déjà que si X_n suit la loi π à un temps n, alors X_m suivra la même loi π à tous les temps ultérieurs $m > n$. Par contre, les Théorèmes 2.1.3 et 2.1.5 ne sont plus nécessairement vrais dans ce cas : Il suffit de considérer l'exemple du modèle d'Ehrenfest. Toutefois, on a encore convergence vers la distribution stationnaire dans le sens de la moyenne ergodique (ou moyenne de Cesaro) :

Théorème 1.4.13. *Pour une chaîne de Markov irréductible, et pour toute distribution initiale ν, la fréquence moyenne de passage en tout état j converge vers π_j:*

$$\lim_{n \to \infty} \frac{1}{n} \mathbb{E}_\nu \biggl(\sum_{m=0}^{n-1} 1_{\{X_m = j\}} \biggr) = \pi_j \qquad \forall j \in \mathcal{X} \; . \tag{1.4.31}$$

Démonstration. Soit Π la matrice dont toutes les lignes sont égales à π, cf. (1.4.4). Alors on a $\Pi P = \Pi$, et le fait que $P\mathbf{1} = \mathbf{1}$ implique qu'on a également $P\Pi = \Pi$. Il suit que

$$(I + P + \cdots + P^{n-1})(I - P + \Pi) = I - P^n + n\Pi \; . \tag{1.4.32}$$

Montrons que la matrice $I - P + \Pi$ est inversible. Soit x un vecteur colonne tel que $(I - P + \Pi)x = 0$. Alors on a

$$0 = \pi(I - P + \Pi)x = \underbrace{\pi(I - P)}_{=0} x + \pi\Pi x = \pi x \; , \tag{1.4.33}$$

puisque $\pi\Pi = \pi$ en raison du fait que $\sum_i \pi_i = 1$. Il suit que $\Pi x = 0$, et donc $(I - P)x = 0$. Comme P admet 1 comme valeur propre simple, avec vecteur propre à droite $\mathbf{1}$, ceci implique que $x \parallel \mathbf{1}$, ce qui n'est possible que si $x = 0$ puisque $\pi x = 0$ et tous les π_i sont positifs. La matrice $I - P + \Pi$ est donc bien inversible.

Soit $Z = (I - P + \Pi)^{-1}$. Comme $\pi(I - P + \Pi) = \pi$, on a aussi $\pi = \pi Z$ et $\Pi = \Pi Z$. En multipliant (1.4.32) à droite par Z, il vient

$$I + P + \cdots + P^{n-1} = (I - P^n)Z + n\Pi Z = (I - P^n)Z + n\Pi \,. \tag{1.4.34}$$

Or nous avons, pour tout état initial i,

$$\frac{1}{n}\mathbb{E}_i\Big(\sum_{m=0}^{n-1} 1_{\{X_m=j\}}\Big) = \frac{1}{n}\sum_{m=0}^{n-1}\big(P^m\big)_{ij} = \bigg[\frac{1}{n}(I - P^n)Z + \Pi\bigg]_{ij} \,. \tag{1.4.35}$$

Comme les éléments de matrice de P^n sont uniformément bornés par 1, cette quantité converge vers $(\Pi)_{ij} = \pi_j$ lorque $n \to \infty$. Pour une distribution initiale quelconque ν, on obtient de la même manière la convergence vers $(\nu\Pi)_j = \pi_j$. □

Remarque 1.4.14. Cette preuve peut ne pas sembler très transparente. On peut en fait l'éclairer avec quelques notions de calcul matriciel. Soit la matrice $Q = P - \Pi$, décrivant l'écart entre la matrice stochastique et le projecteur sur la distribution stationnaire. Il suit des égalités $\Pi P = P\Pi = \Pi^2 = \Pi$ que $\Pi Q = Q\Pi = 0$, et on en déduit

$$P^n = \Pi + Q^n \tag{1.4.36}$$

(voir aussi l'exercice 1.8). Dans le cas où P est régulière, Q^n tend vers 0 lorsque $n \to \infty$. Si P est irréductible mais pas régulière, la fonction $n \mapsto Q^n$ est oscillante. La preuve montre cependant que Q n'admet pas la valeur propre 1, et que la moyenne des Q^n tend vers zéro, donc la moyenne des P^n tend vers Π.

La distribution stationnaire π a également un lien intéressant avec l'espérance du temps de premier retour en un site i, appelé *temps de récurrence moyen* en i :

Théorème 1.4.15. *Pour une chaîne de Markov irréductible de distribution stationnaire π, les temps de récurrence moyens sont donnés par*

$$\mathbb{E}_i(\tau_i) = \frac{1}{\pi_i} \,. \tag{1.4.37}$$

Démonstration. Nous commençons par établir une équation reliant divers temps de premier passage moyens. Pour $i, j \in \mathcal{X}$, on a

$$\begin{aligned}
\mathbb{E}_i(\tau_j) &= \mathbb{P}_i\{\tau_j = 1\} + \sum_{n\geqslant 2} n\mathbb{P}_i\{\tau_j = n\} \\
&= p_{ij} + \sum_{n\geqslant 2} n \sum_{k\neq j} \mathbb{P}_i\{\tau_j = n, X_1 = k\} \\
&= p_{ij} + \sum_{n\geqslant 2} n \sum_{k\neq j} p_{ik}\mathbb{P}_k\{\tau_j = n-1\} \\
&= p_{ij} + \sum_{k\neq j} p_{ik} \sum_{m\geqslant 1} (m+1)\mathbb{P}_k\{\tau_j = m\} \\
&= p_{ij} + \sum_{k\neq j} p_{ik}\bigg[\mathbb{E}_k(\tau_j) + \underbrace{\sum_{m\geqslant 1}\mathbb{P}_k\{\tau_j = m\}}_{=1}\bigg] \\
&= 1 + \sum_{k\neq j} p_{ik}\mathbb{E}_k(\tau_j) \,.
\end{aligned} \tag{1.4.38}$$

Cette relation peut être récrite sous la forme

$$1 - \mathbb{E}_i(\tau_j) = -\sum_{k\in\mathcal{X}} (1-\delta_{kj}) p_{ik} \mathbb{E}_k(\tau_j) \; . \tag{1.4.39}$$

Il suit que

$$\begin{aligned} 1 - \pi_j \mathbb{E}_j(\tau_j) &= \sum_{i\in\mathcal{X}} \pi_i \big[1 - \delta_{ij}\mathbb{E}_i(\tau_j)\big] \\ &= \sum_{i\in\mathcal{X}} \pi_i \big[1 - \mathbb{E}_i(\tau_j) + (1-\delta_{ij})\mathbb{E}_i(\tau_j)\big] \\ &= \sum_{k\in\mathcal{X}} \sum_{i\in\mathcal{X}} \pi_i (\delta_{ik} - p_{ik})(1-\delta_{kj})\mathbb{E}_k(\tau_j) \; . \end{aligned}$$

La somme sur i s'annule, puisque $\pi_k = \sum_i \pi_i p_{ik}$. □

Exemple 1.4.16. Dans le cas du modèle d'Ehrenfest avec N boules, le temps de récurrence moyen vers l'état à i boules est donné par

$$\mathbb{E}_i(\tau_i) = \frac{1}{\nu_i} = 2^N \frac{i!(N-i)!}{N!} \; . \tag{1.4.40}$$

En particulier, le temps moyen entre configurations où toutes les boules sont dans l'urne de gauche est de 2^N. Ce temps devient gigantesque pour des nombres de boules de l'ordre du nombre d'Avogadro, c'est-à-dire du nombre de molécules dans un échantillon d'une mole de gaz. Ce modèle simple peut donc justifier pourquoi, lorsque deux récipients contenant des gaz sont mis en contact, on n'observe jamais toutes les molécules dans le même récipient.

1.5 Chaînes de Markov réversibles

Définition 1.5.1. *Soit P une matrice stochastique. Un vecteur $\alpha = \{\alpha_i\}_{i\in\mathcal{X}} \in [0,\infty)^{\mathcal{X}}$, $\alpha \neq 0$, est dit* réversible *par rapport à P si*

$$\alpha_i p_{ij} = \alpha_j p_{ji} \qquad \forall i,j \in \mathcal{X} \; . \tag{1.5.1}$$

Une chaîne de Markov est dite réversible *si sa matrice admet un vecteur réversible.*

La condition (1.5.1) est appelée *condition d'équilibre détaillé* en physique. Elle signifie que si les états i et j sont occupés avec probabilités proportionnelles à α_i et α_j respectivement, alors les taux de transition de i à j et de j à i sont égaux.

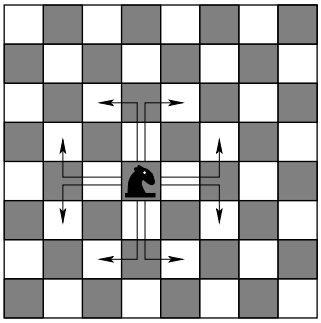
Théorème 1.5.2. *Soit P une matrice stochastique et $\alpha \in [0,\infty)^{\mathcal{X}}$ un vecteur non nul.*

1. *Si α est réversible par rapport à P, alors α est une mesure invariante.*
2. *Si α est réversible par rapport à P, et $\sum_{j\in\mathcal{X}} \alpha_j < \infty$, alors la mesure π définie par $\pi_i = \alpha_i / \sum_{j\in\mathcal{X}} \alpha_j$ est une distribution stationnaire.*
3. *Si π est une distribution stationnaire, alors*

$$\mathbb{P}_\pi\{X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n\} = \mathbb{P}_\pi\{X_0 = i_n, X_1 = i_{n-1}, \ldots, X_n = i_0\} \tag{1.5.2}$$

pour tout $n \in \mathbb{N}$ et tout choix de $i_0, \ldots, i_n \in \mathcal{X}$.

Démonstration.



2	3	4	4	4	4	3	2
3	4	6	6	6	6	4	3
4	6	8	8	8	8	6	4
4	6	8	8	8	8	6	4
4	6	8	8	8	8	6	4
4	6	8	8	8	8	6	4
3	4	6	6	6	6	4	3
2	3	4	4	4	4	3	2

FIGURE 1.6. Mouvements permis du cavalier sur l'échiquier. Nombre de mouvements possibles à partir de chaque case.

1. On a
$$\sum_{i\in\mathcal{X}} \alpha_i p_{ij} = \alpha_j \sum_{i\in\mathcal{X}} p_{ji} = \alpha_j \ . \tag{1.5.3}$$
2. Suit immédiatement de 1.
3. Par le Théorème 1.2.5,
$$\begin{aligned}\mathbb{P}_\pi\{X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\} &= \pi_{i_0} p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n} \\ &= p_{i_1 i_0} \pi_{i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n} \\ &= \dots = p_{i_1 i_0} p_{i_2 i_1} \dots p_{i_n i_{n-1}} \pi_{i_n} \ , \end{aligned} \tag{1.5.4}$$
qui est égal à $\mathbb{P}_\pi\{X_0 = i_n, X_1 = i_{n-1}, \dots, X_n = i_0\}$. □

La relation (1.5.2) signifie qu'une trajectoire a la même probabilité que la trajectoire renversée dans le temps. C'est ce qui justifie le terme de réversibilité.

Exemple 1.5.3 (Marche aléatoire du cavalier)**.** On suppose qu'un cavalier se déplace sur un échiquier, en choisissant à chaque unité de temps, de manière équiprobable, l'un des mouvements autorisés par les règles des Echecs. Combien de temps se passe-t-il en moyenne entre deux passages du cavalier au coin inférieur gauche de l'échiquier?

Soit n_i le nombre de mouvements possibles à partir de la case i (Figure 1.6). La trajectoire du cavalier est donc décrite par une chaîne de Markov de probabilités de transition
$$p_{ij} = \begin{cases} \dfrac{1}{n_i} & \text{si le mouvement } i \mapsto j \text{ est permis ,} \\ 0 & \text{sinon .} \end{cases} \tag{1.5.5}$$
On vérifie alors facilement que $n = \{n_i\}_{i\in\mathcal{X}}$ est un vecteur réversible de la chaîne. Par conséquent,
$$\pi_i = \frac{n_i}{\sum_{j\in\mathcal{X}} n_j} = \frac{n_i}{336} \tag{1.5.6}$$
est la distribution stationnaire de la chaîne. Il suit alors du Théorème 2.4.3 que le temps de récurrence moyen vers le coin inférieur gauche est donné par $1/\pi_{(1,1)} = 336/2 = 168$.

Exemple 1.5.4 (Modèle d'Ehrenfest)**.** Nous sommes maintenant en mesure d'expliquer pourquoi la distribution invariante du modèle d'Ehrenfest est binomiale (cf. Exemples 1.1.3 et 1.4.12). Pour cela, au lieu de considérer le modèle comme une chaîne de Markov sur

l'ensemble $\{0, \ldots, N\}$, nous le considérons comme une chaîne sur $\mathcal{X} = \{0,1\}^N$ (qu'on peut considérer comme un hypercube de dimension N). La composante x_i de l'état $x = (x_1, \ldots, x_N) \in \mathcal{X}$ vaut 0 si la $i^{\text{ème}}$ boule est dans l'urne de gauche, et 1 si elle est dans l'urne de droite.

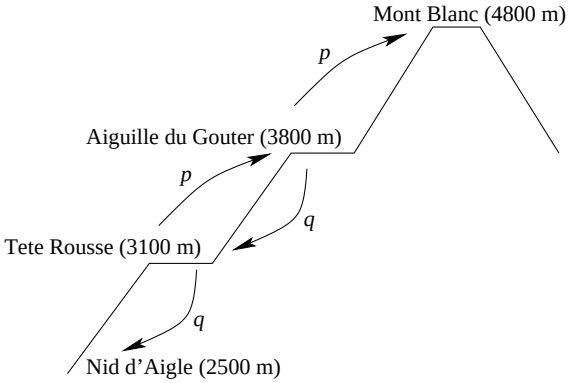
Depuis chaque état $x \in \mathcal{X}$, on peut atteindre exactement N autres états, obtenus en changeant exactement une composante de x, chaque fois avec probabilité $1/N$. Par conséquent, tout vecteur constant est réversible, et la distribution stationnaire est uniforme: $\pi_x = 2^{-N} \forall x \in \mathcal{X}$. Toutefois, il peut y avoir beaucoup d'états de $\mathcal{X}$ correspondant à un nombre donné de boules dans une urne. En fait,

$$\mathbb{P}_\pi\{m \text{ boules dans l'urne de droite}\} = \sum_{x:\ \sum x_i = m} \pi_x = \binom{N}{m} \frac{1}{2^N} \,. \tag{1.5.7}$$

On retrouve le fait que la distribution stationnaire du modèle d'Ehrenfest est binomiale.

1.6 Exercices

Exercice 1.1. Un alpiniste veut faire l'ascension du Mont Blanc. Lors de son ascension, il décide de passer la nuit au refuge de Tête Rousse, et également au refuge de l'Aiguille du Goûter. Chaque matin, il observe la météo. Si celle-ci lui paraît bonne, alors il poursuit son ascension jusqu'à l'étape suivante. Par contre, si la météo semble mauvaise, il redescend d'une étape. On suppose que l'alpiniste est initialement au refuge de Tête Rousse, et que s'il est obligé de redescendre au Nid d'Aigle, alors il abandonne son projet d'ascension. Enfin, on suppose que la météo est bonne avec probabilité p, et mauvaise avec probabilité $q = 1 - p$, et indépendante de la météo des jours précédents.



1. Montrer que le problème peut être décrit par une chaîne de Markov absorbante, et calculer sa matrice fondamentale.
2. Calculer, en fonction de p, la probabilité que l'alpiniste atteigne le sommet.
3. Déterminer la valeur $p^\star$ de p pour qu'il ait une chance sur deux d'atteindre le sommet.
4. Calculer le nombre moyen de jours de l'ascension pour $p = p^\star$.

Exercice 1.2. Résoudre le problème de la souris dans le labyrinthe (Exemple 3.1.1.) :

1. Déterminer la matrice fondamentale de la chaîne.
2. Calculer la probabilité que la souris atteigne la nourriture.
3. Calculer le temps moyen du parcours de la souris.

Exercice 1.3. Deux joueurs A et B s'affrontent dans une partie de tennis. Chaque point joué est gagné par le joueur A avec une probabilité de 3/5, sinon il est gagné par B. On suppose les points indépendants.

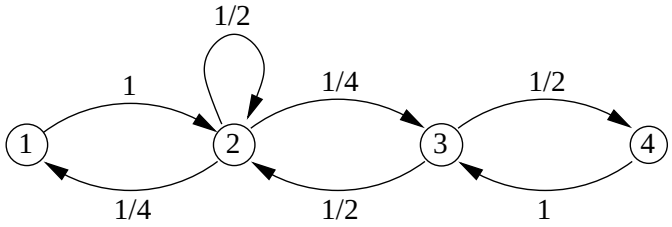
Initialement, les deux joueurs sont à égalité. Pour gagner la partie, un joueur doit obtenir une avance de deux points sur son opposant.

1. Modéliser le jeu par une chaîne de Markov absorbante à 5 états: Egalité (deuce), Avantage A, Avantage B, A gagne, et B gagne. Donner la matrice de transition de cette chaîne.
2. Montrer que la matrice fondamentale de la chaîne est donnée par
$$N = \frac{1}{13}\begin{pmatrix} 25 & 15 & 10 \\ 10 & 19 & 4 \\ 15 & 9 & 19 \end{pmatrix} .$$
3. Calculer la probabilité que A gagne, si les joueurs sont initialement à égalité.
4. Calculer la durée moyenne du jeu si les joueurs sont initialement à égalité.

Exercice 1.4. On considère une chaîne de Markov sur $\mathcal{X} = \{1, 2, 3, 4\}$, de matrice de transition
$$P = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/16 & 7/16 & 0 & 1/2 \\ 1/16 & 0 & 7/16 & 1/2 \\ 0 & 1/4 & 1/4 & 1/2 \end{pmatrix}$$

1. Montrer que cette chaîne est irréductible.
2. La chaîne est-elle régulière?
3. Calculer la distribution stationnaire de la chaîne.

Exercice 1.5. On considère la chaîne de Markov suivante :



1. Donner la matrice de transition P de la chaîne.
2. La chaîne est-elle irréductible?
3. La chaîne est-elle régulière?
4. Déterminer la distribution stationnaire de la chaîne.
5. La chaîne est-elle réversible?

Exercice 1.6. Angèle possède 3 parapluies. Chaque jour, elle va au bureau le matin, et revient à son domicile le soir. Pour chaque trajet, elle emporte avec elle un parapluie s'il pleut, et s'il y en a au moins un sur place. Elle n'emporte pas de parapluie s'il ne pleut pas. On suppose que la probabilité qu'il pleuve au début de chaque trajet est de 1/3, et qu'elle est indépendante de la météo lors de tous les autres trajets. Soit X_n le nombre de parapluies qu'Angèle possède sur place avant de débuter le $n$ième trajet.

1. Montrer que $\{X_n\}_n$ est une chaîne de Markov, et donner sa matrice de transition.
2. De quel type de chaîne s'agit-il?
3. Quelle est la probabilité, asymptotiquement au bout d'un grand nombre de voyages, qu'Angèle ne dispose pas de parapluie sur place au moment de partir?
4. Quelle est la probabilité asymptotique qu'elle se fasse mouiller bêtement, c'est-à-dire qu'elle n'ait pas de parapluie à sa disposition alors qu'il pleut dès son départ?

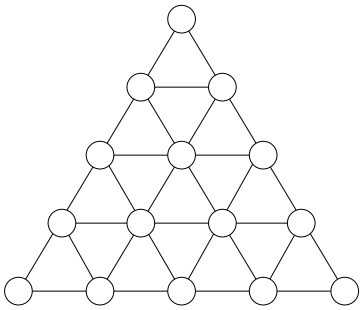
Exercice 1.7. Montrer que la distribution binomiale est stationnaire pour le modèle d'Ehrenfest.

Exercice 1.8. Toute matrice stochastique de taille 2×2 peut s'écrire

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \qquad \text{avec } p, q \in [0,1].$$

1. Discuter, en fonction de p et q, quand P est absorbante, irréductible, régulière.
2. Montrer que la matrice
$$\Pi = \frac{1}{p+q} \begin{pmatrix} q & p \\ q & p \end{pmatrix}$$
est un projecteur ($\Pi^2 = \Pi$) qui commute avec P. Calculer son noyau et son image.
3. Soit $Q = P - \Pi$. Montrer que $Q\Pi = \Pi Q = 0$. Calculer Q^2, puis Q^n pour tout n.
4. Déduire des résultats précédents P^n pour tout n. Discuter la limite de P^n lorsque $n \to \infty$ en fonction de p et q.

Exercice 1.9. Une puce se déplace sur le réseau ci-dessous, en choisissant à chaque saut l'une des cases adjacentes, au hasard de manière uniforme.



Déterminer le temps de récurrence moyen vers le coin inférieur gauche.

Exercice 1.10. On considère

1. un roi;
2. une dame;
3. un fou

se déplaçant sur un échiquier, en choisissant à chaque fois, de manière équiprobable, l'un des mouvements permis par les règles des échecs.

Déterminer dans chaque cas le temps de récurrence moyen vers le coin inférieur gauche.

Exercice 1.11. Soit X_n la chaîne de Markov sur $\mathcal{X} = \{1, 2, \ldots, N\}$ de probabilités de transition

$$p_{ij} = \begin{cases} p & \text{si } j = i+1 \text{ ou } i = N \text{ et } j = 1 \\ 1-p & \text{si } j = i-1 \text{ ou } i = 1 \text{ et } j = N \, . \end{cases}$$

1. Déterminer la distribution stationnaire de la chaîne.
2. Pour quelles valeurs de p la chaîne est-elle réversible?

Exercice 1.12. Soit $\mathcal{G} = (V, E)$ un graphe non orienté connexe fini. Soit X_n la chaîne de Markov sur V construite en choisissant pour X_{n+1}, de manière équiprobable, l'un des sommets adjacents à X_n.

1. Montrer que le nombre de voisins de chaque site forme un vecteur réversible.
2. En déduire une expression pour la distribution stationnaire de la chaîne.

Chapitre 2

Chaînes de Markov sur un ensemble dénombrable

2.1 Marches aléatoires

Les marches aléatoires constituent un exemple relativement simple, et néanmoins très important de chaînes de Markov sur un ensemble dénombrable infini. Dans ce cas, en effet, $\mathcal{X} = \mathbb{Z}^d$ est un réseau infini, de dimension $d \in \mathbb{N}^*$. D'habitude, on considère que la chaîne démarre en $X_0 = 0$. Ensuite, elle choisit à chaque instant l'un des $2d$ sites voisins, selon une loi fixée d'avance.

Définition 2.1.1. *Une* marche aléatoire *sur $\mathbb{Z}^d$ est une chaîne de Markov à valeurs dans $\mathbb{Z}^d$, de distribution initiale $\nu = \delta_0$, et de probabilités de transition satisfaisant*

$$p_{ij} = 0 \qquad si\ i = j\ ou\ \|i - j\| > 1\ . \tag{2.1.1}$$

La marche est dite symétrique *si*

$$p_{ij} = \frac{1}{2d} \qquad pour\ \|i - j\| = 1\ . \tag{2.1.2}$$

Les trajectoires de la marche aléatoire sont des suites de points de $\mathbb{Z}^d$ à distance 1, qu'on a coutume à identifier à la ligne brisée reliant ces points (Figure 2.1).

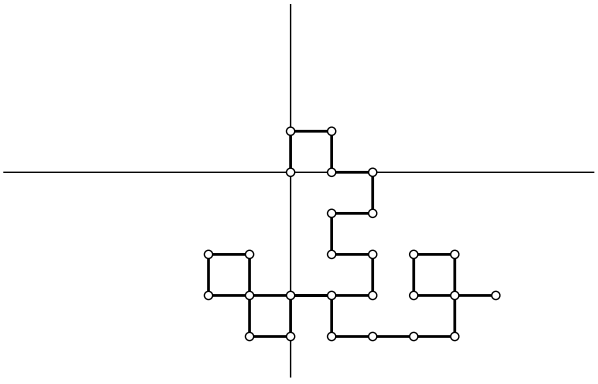


FIGURE 2.1. Une trajectoire d'une marche aléatoire en dimension $d = 2$.

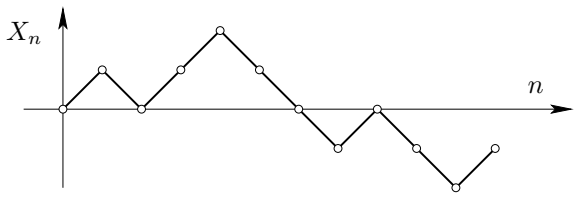


FIGURE 2.2. Une réalisation d'une marche aléatoire unidimensionnelle.

Dans le cas symétrique, il suit directement du Théorème 1.2.5 que chaque segment de trajectoire $X_{[0,n]}$ a probabilité $(2d)^{-n}$. On peut facilement déterminer quelques propriétés de la loi de X_n.

Proposition 2.1.2. *Pour la marche aléatoire symétrique sur $\mathbb{Z}^d$, les variables aléatoires X_n satisfont*

$$\mathbb{E}(X_n) = 0 \qquad et \qquad \mathrm{cov}(X_n) = \frac{n}{d} I \tag{2.1.3}$$

pour tout temps n. De plus, lorsque $n \to \infty$ on a

$$\frac{X_n}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}\Big(0, \frac{1}{d} I\Big) \,, \tag{2.1.4}$$

où $\mathcal{L}$ désigne la convergence en loi, et $\mathcal{N}(0, \Sigma)$ dénote la loi normale centrée de matrice de covariance Σ.

DÉMONSTRATION. On vérifie facilement que les variables aléatoires $Y_n = X_n - X_{n-1}$ sont i.i.d., d'espérance nulle et de matrice de covariance $\frac{1}{d} I$, ce qui implique (2.1.3). La relation (2.1.4) suit alors directement du théorème de la limite centrale. □

En conséquence, la position de la marche aléatoire au temps n se trouvera avec grande probabilité dans une boule de rayon d'ordre $\sqrt{n}$ autour de l'origine. On dit que la marche aléatoire a un comportement *diffusif* (par opposition à *ballistique*, où la distance à l'origine croîtrait proportionnellement à n).

Nous considérons maintenant plus particulièrement le cas de la marche aléatoire unidimensionnelle ($d = 1$) symétrique. Dans ce cas, on voit facilement que la loi de X_n est binomiale centrée :

$$\mathbb{P}\{X_n = k\} = \frac{1}{2^n} \binom{n}{\frac{n+k}{2}} \qquad \forall k \in \{-n, -n+2, \ldots, n-2, n\} \,. \tag{2.1.5}$$

En particulier, la probabilité que le processus se trouve en 0 au $n^{\text{ème}}$ pas est donnée par

$$\mathbb{P}\{X_n = 0\} = \begin{cases} 0 & \text{si } n \text{ est impair} \,, \\ \dfrac{(2m)!}{2^{2m}(m!)^2} & \text{si } n = 2m \text{ est pair} \,. \end{cases} \tag{2.1.6}$$

Remarquons que la formule de Stirling implique que pour m grand,

$$\mathbb{P}\{X_{2m} = 0\} \sim \frac{1}{2^{2m}} \frac{\sqrt{4\pi m}\, \mathrm{e}^{-2m} (2m)^{2m}}{2\pi m\, \mathrm{e}^{-2m}\, m^{2m}} = \frac{1}{\sqrt{\pi m}} \,. \tag{2.1.7}$$

En tout temps pair, l'origine est l'endroit le plus probable où trouver la marche, mais cette probabilité décroît avec le temps.

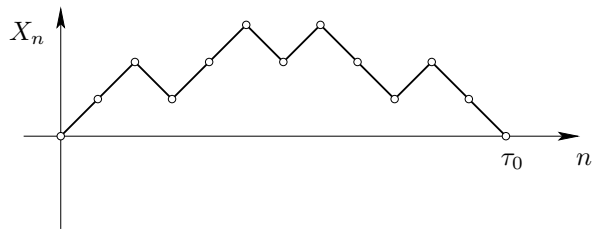


FIGURE 2.3. Une réalisation d'une marche aléatoire unidimensionnelle pour laquelle $\tau_0 = 12$.

Cependant, la loi de chaque X_n ne détermine pas le processus, les X_n n'étant pas indépendants, et on peut étudier beaucoup d'autres propriétés de la marche aléatoire. Une première quantité intéressante est le temps τ_0 du premier retour du processus en 0 (Figure 2.3) :

$$\tau_0 = \inf\{n \geqslant 1 \colon X_n = 0\} \; . \tag{2.1.8}$$

Il est clair que τ_0 ne peut prendre que des valeurs paires. De plus, si $\tau_0 = n$ alors $X_n = 0$, donc $\mathbb{P}\{\tau_0 = n\} \leqslant \mathbb{P}\{X_n = 0\}$. En fait, il nous faut déterminer

$$\mathbb{P}\{\tau_0 = n\} = \mathbb{P}\{X_1 \neq 0, X_2 \neq 0, \dots, X_{n-1} \neq 0, X_n = 0\} \; . \tag{2.1.9}$$

Théorème 2.1.3. *La loi de τ_0 est donnée par*

$$\mathbb{P}\{\tau_0 = n\} = \begin{cases} 0 & \textit{pour } n \textit{ impair} \; , \\ \dfrac{1}{n}\mathbb{P}\{X_{n-2} = 0\} & \textit{pour } n \textit{ pair} \; . \end{cases} \tag{2.1.10}$$

DÉMONSTRATION. Supposons que $\tau_0 = n$. Comme le processus ne peut pas changer de signe sans passer par 0, on a

$$\begin{aligned} \mathbb{P}\{\tau_0 = n\} &= \mathbb{P}\{X_1 > 0, X_2 > 0, \dots, X_{n-1} > 0, X_n = 0\} \\ &\quad + \mathbb{P}\{X_1 < 0, X_2 < 0, \dots, X_{k-1} < 0, X_n = 0\} \\ &= 2\mathbb{P}\{X_1 > 0, X_2 > 0, \dots, X_{n-1} > 0, X_n = 0\} \\ &= 2\mathbb{P}\{X_1 = 1, X_2 > 0, \dots, X_{n-2} > 0, X_{n-1} = 1, X_n = 0\} \\ &= 2\mathbb{P}\{X_n = 0 | X_{n-1} = 1\}\mathbb{P}\{X_1 = 1, X_2 > 0, \dots, X_{n-2} > 0, X_{n-1} = 1\} \; , \end{aligned} \tag{2.1.11}$$

où nous avons utilisé la propriété de Markov dans la dernière ligne. La propriété des incréments stationnaires (cf. (1.2.14)) implique

$$\mathbb{P}\{X_n = 0 | X_{n-1} = 1\} = \mathbb{P}\{X_1 = -1\} = \frac{1}{2} \; . \tag{2.1.12}$$

Il suit que

$$\begin{aligned} \mathbb{P}\{\tau_0 = n\} &= \mathbb{P}\{X_1 = 1, X_2 > 0, \dots, X_{n-2} > 0, X_{n-1} = 1\} \\ &= \mathbb{P}\{X_1 = X_{n-1} = 1\} - \mathbb{P}\{X_1 = X_{n-1} = 1, \exists m \in \{2, \dots, n-2\} \colon X_m = 0\} \; . \end{aligned} \tag{2.1.13}$$

Nous utilisons maintenant un argument important, appelé le *principe de réflexion* : A tout chemin allant de $(1, 1)$ à $(n-1, 1)$ passant par 0, on peut faire correspondre un unique

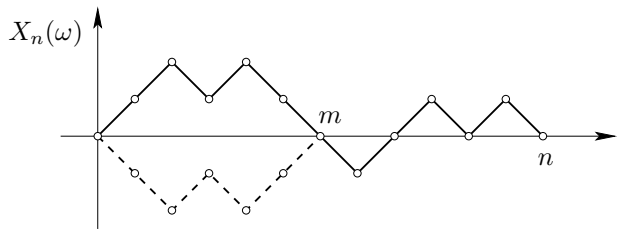


FIGURE 2.4. Pour chaque réalisation d'une marche aléatoire avec $\tau_0 = m < n$ telle que $X_1 = 1$, il existe une autre réalisation telle que $\tau_0 = m$ et $X_1 = -1$, obtenue par réflexion par rapport à l'axe des abscisses.

chemin de $(-1, 1)$ à $(n-1, 1)$, obtenu en réfléchissant par rapport à l'axe des abscisses la partie du chemin antérieure au premier passage en 0. On a donc

$$\mathbb{P}\{X_1 = X_{n-1} = 1,\, \exists m \in \{2, \dots, n-2\} :\, X_m = 0\} = \mathbb{P}\{X_1 = -1, X_{n-1} = 1\}\,. \quad (2.1.14)$$

Finalement, en appliquant de nouveau la propriété des incréments stationnaires, on a

$$\begin{aligned}\mathbb{P}\{X_1 = 1, X_{n-1} = 1\} &= \mathbb{P}\{X_{n-1} = 1 | X_1 = 1\}\mathbb{P}\{X_1 = 1\} = \mathbb{P}\{X_{n-2} = 0\} \cdot \frac{1}{2}\,,\\ \mathbb{P}\{X_1 = -1, X_{n-1} = 1\} &= \mathbb{P}\{X_{n-1} = 1 | X_1 = -1\}\mathbb{P}\{X_1 = -1\} = \mathbb{P}\{X_{n-2} = 2\} \cdot \frac{1}{2}\,.\end{aligned} \quad (2.1.15)$$

En remplaçant dans (2.1.13), il vient

$$\mathbb{P}\{\tau_0 = n\} = \frac{1}{2}\big[\mathbb{P}\{X_{n-2} = 0\} - \mathbb{P}\{X_{n-2} = 2\}\big]\,. \quad (2.1.16)$$

Le reste de la preuve est un calcul direct. Comme

$$\frac{\mathbb{P}\{X_{n-2} = 2\}}{\mathbb{P}\{X_{n-2} = 0\}} = \frac{\binom{n-2}{n/2}}{\binom{n-2}{n/2-1}} = \frac{\left(\frac{n}{2}-1\right)!\left(\frac{n}{2}-1\right)!}{\left(\frac{n}{2}\right)!\left(\frac{n}{2}-2\right)!} = \frac{\frac{n}{2}-1}{\frac{n}{2}} = 1 - \frac{2}{n}\,, \quad (2.1.17)$$

on obtient

$$\mathbb{P}\{\tau_0 = n\} = \frac{1}{2}\mathbb{P}\{X_{n-2} = 0\}\left[1 - 1 + \frac{2}{n}\right] = \frac{1}{n}\mathbb{P}\{X_{n-2} = 0\}\,, \quad (2.1.18)$$

ce qui conclut la démonstration. □

Le tableau suivant donne les premières valeurs de la loi et de la fonction de répartition de τ_0 :

n	2	4	6	8	10	12	14
$\mathbb{P}\{\tau_0 = n\}$	$\frac{1}{2}$ $= 0.5$	$\frac{1}{8}$ $= 0.125$	$\frac{1}{16}$ $\cong 0.063$	$\frac{5}{128}$ $\cong 0.039$	$\frac{7}{256}$ $\cong 0.027$	$\frac{21}{1024}$ $\cong 0.021$	$\frac{33}{2048}$ $\cong 0.016$
$\mathbb{P}\{\tau_0 \leqslant n\}$	$= 0.5$	$= 0.625$	$\cong 0.688$	$\cong 0.727$	$\cong 0.754$	$\cong 0.774$	$\cong 0.791$

Il est donc assez probable de revenir rapidement en 0, puis la loi prend des valeurs plutôt faibles, tout en décroissant lentement. Il suit de (2.1.7) que pour des grands n, $\mathbb{P}\{\tau_0 = n\}$ décroît comme $1/n^{3/2}$. Ce fait a une conséquence surprenante :

Corollaire 2.1.4. $\mathbb{E}(\tau_0) = +\infty$.

Démonstration. On a

$$\mathbb{E}(\tau_0) = \sum_{n\geqslant 1} n\mathbb{P}\{\tau_0 = n\} = \sum_{m\geqslant 1} 2m\frac{1}{2m}\mathbb{P}\{X_{2m-2} = 0\} \sim \sum_{m\geqslant 1}\frac{1}{\sqrt{\pi m}} = +\infty\,. \tag{2.1.19}$$

□

En d'autres termes, la marche aléatoire finit toujours par revenir en 0, mais la loi de τ_0 décroît trop lentement pour que son espérance soit finie. Cela est lié au fait que si la marche aléatoire s'éloigne beaucoup de 0, il lui faut longtemps pour y revenir.

Par un raisonnement analogue, on peut déterminer la loi du temps de passage

$$\tau_i = \inf\{n \geqslant 0\colon X_n = i\} \tag{2.1.20}$$

Nous donnons simplement le résultat, la démonstration est laissée en exercice.

Théorème 2.1.5. *La loi de τ_i est donnée par*

$$\mathbb{P}\{\tau_i = n\} = \begin{cases} \dfrac{|i|}{n}\mathbb{P}\{X_n = i\} & \text{pour } n \in \{|i|, |i|+2, \dots\}\,, \\ 0 & \text{sinon}\,. \end{cases} \tag{2.1.21}$$

Pour des raisons similaires à celles du cas du retour en 0, la loi de τ_L décroît en $1/n^{3/2}$, et son espérance est donc infinie.

2.2 Généralités sur les processus stochastiques

Jusqu'à présent, nous avons parlé de chaînes de Markov, qui sont des processus stochastiques particuliers, sans préciser l'espace probabilisé sous-jacent. Cela ne pose pas de problème tant qu'on parle de segments de trajectoire finis : Il suffit de considérer des espaces produits comprenant un nombre fini de termes. Le cas de trajectoires infinies nécessite quelques précautions supplémentaires. Nous donnons ici un survol de la construction générale des processus stochastiques en temps discret.

Soit $(E, \mathcal{E})$ un espace mesurable. L'ensemble $E^n = E \times E \times \cdots \times E$ peut être muni d'une tribu $\mathcal{E}^{\otimes n}$, définie comme la tribu engendrée par tous les événements du type

$$A^{(i)} = \{\omega \in E^n\colon \omega_i \in A\}\,, \qquad A \in \mathcal{E}\,, \tag{2.2.1}$$

appelés *cylindres*. On dénote par $E^{\mathbb{N}}$ l'ensemble des applications $x : \mathbb{N} \to E$, c'est-à-dire l'ensemble des suites $(x_0, x_1, x_2, \dots)$ à valeurs dans E. Cet ensemble peut à nouveau être muni d'une tribu construite à partir de tous les cylindres, notée $\mathcal{E}^{\otimes\mathbb{N}}$.

Définition 2.2.1.

- *Un* processus stochastique *à valeurs dans $(E, \mathcal{E})$ est une suite $\{X_n\}_{n\in\mathbb{N}}$ de variables aléatoires à valeurs dans $(E, \mathcal{E})$, définies sur un même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ (autrement dit, chaque X_n est une application $\mathcal{F}$-$\mathcal{E}$-mesurable de Ω dans E). C'est donc également une variable aléatoire à valeurs dans $(E^{\mathbb{N}}, \mathcal{E}^{\otimes\mathbb{N}})$.*

- *Soit $\mathbb{Q}$ une mesure de probabilité sur $(E^{\mathbb{N}}, \mathcal{E}^{\otimes\mathbb{N}})$. Les* distributions de dimension finie *de $\mathbb{Q}$ sont les mesures sur $(E^{n+1}, \mathcal{E}^{\otimes n+1})$ définies par*

$$\mathbb{Q}^{(n)} = \mathbb{Q} \circ (\pi^{(n)})^{-1} , \tag{2.2.2}$$

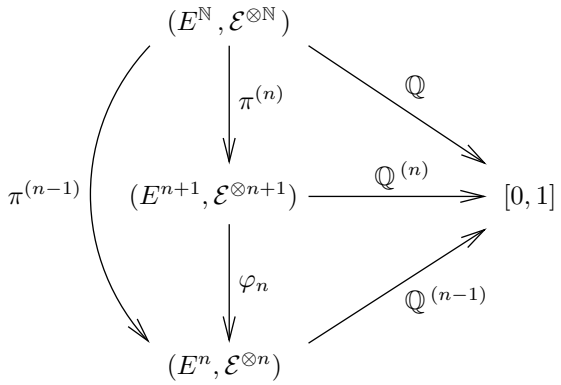
où $\pi^{(n)}$ est la projection $\pi^{(n)} : E^{\mathbb{N}} \to E^{n+1}$, $(x_0, x_1, x_2, \dots) \mapsto (x_0, \dots, x_n)$.

On se convainc facilement que la suite des $\{\mathbb{Q}^{(n)}\}_{n\in\mathbb{N}}$ détermine $\mathbb{Q}$ univoquement. Inversément, pour qu'une suite donnée $\{\mathbb{Q}^{(n)}\}_{n\in\mathbb{N}}$ corresponde effectivement à une mesure $\mathbb{Q}$, les $\mathbb{Q}^{(n)}$ doivent satisfaire une *condition de compatibilité* :

Soit φ_n la projection $\varphi_n : E^{n+1} \to E^n$, $(x_0, \dots, x_n) \mapsto (x_0, \dots, x_{n-1})$. Alors on a $\pi^{(n-1)} = \varphi_n \circ \pi^{(n)}$, donc pour tout $A \in \mathcal{E}^{\otimes n}$, $(\pi^{(n-1)})^{-1}(A) = (\pi^{(n)})^{-1}(\varphi_n^{-1}(A))$. La condition de compatibilité s'écrit donc

$$\mathbb{Q}^{(n-1)} = \mathbb{Q}^{(n)} \circ \varphi_n^{-1} . \tag{2.2.3}$$

Le diagramme suivant illustre la situation (toutes les projections étant mesurables, on peut les considérer à la fois comme applications entre ensembles et entre tribus) :



Nous allons voir comment construire une suite de $\mathbb{Q}^{(n)}$ satisfaisant la condition (2.2.3).

Définition 2.2.2. *Soient $(E_1, \mathcal{E}_1)$ et $(E_2, \mathcal{E}_2)$ deux espaces mesurables. Un* noyau markovien de $(E_1, \mathcal{E}_1)$ vers $(E_2, \mathcal{E}_2)$ *est une application $K : E_1 \times \mathcal{E}_2 \to [0,1]$ satisfaisant les deux conditions*

1. *Pour tout $x \in E_1$, $K(x, \cdot)$ est une mesure de probabilité sur $(E_2, \mathcal{E}_2)$.*
2. *Pour tout $A \in \mathcal{E}_2$, $K(\cdot, A)$ est une application $\mathcal{E}_1$-mesurable.*

Exemple 2.2.3.

1. Soit μ une mesure de probabilité sur $(E_2, \mathcal{E}_2)$. Alors K défini par $K(x, A) = \mu(A)$ pour tout $x \in E_1$ est un noyau markovien.
2. Soit $f : E_1 \to E_2$ une application mesurable. Alors K défini par $K(x, A) = 1_A(f(x))$ est un noyau markovien.
3. Soit $\mathcal{X} = \{1, \dots, N\}$ un ensemble fini, et posons $E_1 = E_2 = \mathcal{X}$ et $\mathcal{E}_1 = \mathcal{E}_2 = \mathcal{P}(\mathcal{X})$. Alors K défini par

$$K(i, A) = \sum_{j\in A} p_{ij} , \tag{2.2.4}$$

où $P = (p_{ij})_{i,j\in\mathcal{X}}$ est une matrice stochastique, est un noyau markovien.

Si μ est une mesure de probabilité sur $(E_1, \mathcal{E}_1)$ et K est un noyau markovien de $(E_1, \mathcal{E}_1)$ vers $(E_2, \mathcal{E}_2)$, on définit une mesure de probabilité $\mu \otimes K$ sur $\mathcal{E}_1 \otimes \mathcal{E}_2$ par

$$(\mu \otimes K)(A) := \int_{E_1} K(x_1, A_{x_1}) \mu(\mathrm{d}x_1) \,, \tag{2.2.5}$$

où $A_{x_1} = \{x_2 \in E_2 \colon (x_1, x_2) \in A\} \in \mathcal{E}_2$ est la *section* de A en x_1. On vérifie que c'est bien une mesure de probabilité. Afin de comprendre sa signification, calculons ses marginales. Soient π_1 et π_2 les projections définies par $\pi_i(x_1, x_2) = x_i$, $i = 1, 2$.

1. Pour tout ensemble mesurable $A_1 \in \mathcal{E}_1$, on a

$$\begin{aligned} \big((\mu \otimes K) \circ \pi_1^{-1}\big)(A_1) &= (\mu \otimes K)(A_1 \otimes E_2) \\ &= \int_{E_1} 1_{A_1}(x_1) K(x_1, E_2) \mu(\mathrm{d}x_1) \\ &= \int_{A_1} K(x_1, E_2) \mu(\mathrm{d}x_1) = \mu(A_1) \,, \end{aligned} \tag{2.2.6}$$

   où on a utilisé le fait que la section $(A_1 \otimes \mathcal{E}_1)_{x_1}$ est donnée par E_2 si $x_1 \in A_1$, et $\emptyset$ sinon. Ceci implique

$$(\mu \otimes K) \circ \pi_1^{-1} = \mu \,. \tag{2.2.7}$$

   La première marginale de $\mu \otimes K$ est donc simplement μ.

2. Pour tout ensemble mesurable $A_2 \in \mathcal{E}_2$, on a

$$\begin{aligned} \big((\mu \otimes K) \circ \pi_2^{-1}\big)(A_2) &= (\mu \otimes K)(E_1 \otimes A_2) \\ &= \int_{E_1} K(x_1, A_2) \mu(\mathrm{d}x_1) \,. \end{aligned} \tag{2.2.8}$$

   La seconde marginale de $\mu \otimes K$ s'interprête comme suit: c'est la mesure sur E_2 obtenue en partant avec la mesure μ sur E_1, et en "allant de tout $x \in E_1$ vers $A_2 \in \mathcal{E}_2$ avec probabilité $K(x_1, A_2)$".

Enfin, par une variante du théorème de Fubini–Tonelli, on vérifie que pour toute fonction $(\mu \otimes K)$-intégrable $f : E_1 \times E_2 \to \mathbb{R}$, on a

$$\int_{E_1 \times E_2} f \,\mathrm{d}(\mu \otimes K) = \int_{E_1} \left(\int_{E_2} f(x_1, x_2) K(x_1, \mathrm{d}x_2) \right) \mu(\mathrm{d}x_1) \,. \tag{2.2.9}$$

Nous pouvons maintenant procéder à la construction de la suite $\{\mathbb{Q}^{(n)}\}_{n \in \mathbb{N}}$ de distributions de dimension finie, satisfaisant la condition de compatibilité (2.2.3). Sur l'espace mesurable $(E, \mathcal{E})$, on se donne une mesure de probabilité ν, appelée *mesure initiale*. On se donne pour tout $n \in \mathbb{N}$ un noyau markovien K_n de $(E^{n+1}, \mathcal{E}^{\otimes n+1})$ vers $(E, \mathcal{E})$. On définit alors la suite $\{\mathbb{Q}^{(n)}\}_{n \in \mathbb{N}}$ de mesures de probabilité sur $(E^{n+1}, \mathcal{E}^{n+1})$ récursivement par

$$\begin{aligned} &\mathbb{Q}^{(0)} = \nu \,, \\ &\mathbb{Q}^{(n)} = \mathbb{Q}^{(n-1)} \otimes K_{n-1} \,, \qquad\qquad n \geqslant 2 \,. \end{aligned} \tag{2.2.10}$$

Par (2.2.7), on a $\mathbb{Q}^{(n)} \circ \varphi_n^{-1} = (\mathbb{Q}^{(n-1)} \otimes K_{n-1}) \circ \varphi_n^{-1} = \mathbb{Q}^{(n-1)}$, donc la condition de compatibilité est bien satisfaite.

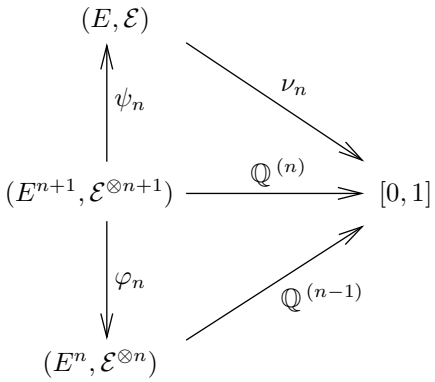
L'interprétation de (2.2.10) est simplement que chaque noyau markovien K_n décrit les probabilités de transition entre les temps n et $n+1$, et permet ainsi de définir une

mesure sur les segments de trajectoire plus longs d'une unité. Remarquons enfin qu'on peut également construire pour tout m, n un noyau $K_{n,m}$ de $(E^n, \mathcal{E}^{\otimes n})$ vers $(E^m, \mathcal{E}^{\otimes m})$ tel que $\mathbb{Q}^{(n+m)} = \mathbb{Q}^{(n)} \otimes K_{n,m}$.

On peut noter par ailleurs que si $\psi_n : E^{n+1} \to E$ désigne la projection sur la dernière composante $(x_0, \ldots, x_n) \mapsto x_n$, alors la formule (2.2.8) montre que la loi de X_n, qui est donnée par la marginale $\nu_n = \mathbb{Q}^{(n)} \circ \psi_n^{-1}$, s'exprime comme

$$\mathbb{P}\{X_n \in A\} = \nu_n(A) = \int_{E_n} K_{n-1}(x, A)\mathbb{Q}^{(n-1)}(\mathrm{d}x) \; . \tag{2.2.11}$$

La situation est illustrée par le diagramme suivant :



Nous donnons maintenant, sans démonstration, le résultat général assurant la légitimité de toute la procédure.

Théorème 2.2.4 (Ionescu–Tulcea)**.** *Pour la suite de mesures* $\{\mathbb{Q}^{(n)}\}_{n\in\mathbb{N}}$ *construites selon* (2.2.10)*, il existe une unique mesure de probabilité* $\mathbb{Q}$ *sur* $(E^{\mathbb{N}}, \mathcal{E}^{\otimes\mathbb{N}})$ *telle que* $\mathbb{Q}^{(n)} = \mathbb{Q} \circ (\pi^{(n)})^{-1}$ *pour tout* n*, c'est-à-dire que les* $\mathbb{Q}^{(n)}$ *sont les distributions de dimension finie de* $\mathbb{Q}$*.*

Exemple 2.2.5.

1. **Mesures produit:** On se donne une suite $\{\mu_n\}_{n\in\mathbb{N}}$ de mesures de probabilité sur l'espace mesurable $(E, \mathcal{E})$. Soit, pour tout n, $\mathbb{Q}^{(n)} = \mu_0 \otimes \mu_1 \otimes \cdots \otimes \mu_n$ la mesure produit. C'est une mesure de la forme ci-dessus, avec noyau markovien

$$K_n(x, A) = \mu_{n+1}(A) \qquad \forall x \in E^n, \forall A \in \mathcal{E} \; . \tag{2.2.12}$$

   La relation (2.2.11) montre que la loi ν_n de X_n est donnée par μ_n. On dit que les variables aléatoires X_n sont *indépendantes*. Si tous les μ_n sont les mêmes, on dit qu'elles sont *indépendantes et identiquement distribuées (i.i.d.)*.
2. **Système dynamique:** On se donne une application mesurable $f : E \to E$, une mesure de probabilité initiale ν sur E. Soit pour tout n le noyau markovien

$$K_n(x, A) = 1_A f(x_n) \; , \tag{2.2.13}$$

   et construisons les $\mathbb{Q}^{(n)}$ comme ci-dessus. Alors la formule (2.2.11) montre qu'on a

pour tout $A \in E$

$$\begin{aligned}\nu_{n+1}(A) &= \int_{E^{n+1}} 1_A f(x_n) \mathbb{Q}^{(n)}(\mathrm{d}x) \\ &= \int_E 1_A f(x_n) \nu_n(\mathrm{d}x_n) \\ &= \int_{f^{-1}(A)} \nu_n(\mathrm{d}x_n) = \nu_n(f^{-1}(A)) \ . \end{aligned} \tag{2.2.14}$$

Il suit que

$$\nu_n = \nu \circ f^{-n} \qquad \forall n \in \mathbb{N} \ . \tag{2.2.15}$$

Cette situation correspond à un système dynamique déterministe. Par exemple, si f est bijective et $\nu = \delta_{x_0}$ est concentrée en un point, on a $\nu_n = \delta_{f^n(x_0)}$.

3. **Chaînes de Markov:** Soit $\mathcal{X}$ un ensemble fini ou dénombrable, muni de la tribu $\mathcal{P}(\mathcal{X})$, et $P = (p_{ij})_{i,j\in\mathcal{X}}$ une matrice stochastique sur $\mathcal{X}$, c'est-à-dire que $0 \leqslant p_{ij} \leqslant 1$ $\forall i, j \in \mathcal{X}$ et $\sum_{j\in\mathcal{X}} p_{ij} = 1$ $\forall i \in \mathcal{X}$. On se donne une mesure de probabilité ν sur $\mathcal{X}$, et une suite $\mathbb{Q}^{(n)}$ construite à partir de (2.2.10) avec pour noyaux markoviens

$$K_n(i_{[0,n-1]}, i_n) = p_{i_{n-1} i_n} \ . \tag{2.2.16}$$

Le processus stochastique de mesure $\mathbb{Q}$, dont les distributions de dimension finie sont les $\mathbb{Q}^{(n)}$, est la chaîne de Markov sur $\mathcal{X}$ de distribution initiale ν et de matrice de transition P. Dans ce cas la relation (2.2.11) se traduit en

$$\mathbb{P}\{X_n \in A\} = \nu_n(A) = \sum_{i\in\mathcal{X}} \sum_{j\in A} p_{ij} \nu_{n-1}(\{i\}) = \sum_{j\in A} \sum_{i\in\mathcal{X}} \mathbb{P}\{X_{n-1} = i\} p_{ij} \ . \tag{2.2.17}$$

2.3 Récurrence, transience et période

Nous revenons maintenant à l'étude des propriétés de chaînes de Markov sur un ensemble dénombrable $\mathcal{X}$. Rappelons que le *temps de premier passage* de la chaîne en un site $i \in \mathcal{X}$ est la variable aléatoire

$$\tau_i = \inf\{n \geqslant 1 \colon X_n = i\} \ , \tag{2.3.1}$$

avec la convention que $\tau_i = \infty$ si $X_n \neq i$ $\forall n \geqslant 1$. Dans le cas où la chaîne démarre dans l'état i au temps 0, τ_i s'appelle également le *temps de premier retour en i*.

Dans le cas où $\mathcal{X}$ est fini, nous avons vu que pour une chaîne irréductible, τ_i était fini presque sûrement (cf. Proposition 1.4.4). Dans le cas où $\mathcal{X}$ est infini, ce n'est plus forcément le cas. En effet, la preuve que nous avons donnée de la Proposition 1.4.4 utilise la Proposition 1.3.3 sur les chaînes absorbantes, dont la preuve ne marche plus dans le cas infini (la définition (1.3.7) de p n'interdit pas que $p = 1$). On est donc amené à introduire la distinction suivante.

Définition 2.3.1. *Un état $i \in \mathcal{X}$ est dit* récurrent *si*

$$\mathbb{P}_i\{\tau_i < \infty\} := \sum_{n=1}^{\infty} \mathbb{P}_i\{\tau_i = n\} = 1 \ . \tag{2.3.2}$$

Dans le cas contraire, il est dit transient. *La chaîne de Markov est appelée* récurrente, *respectivement* transiente, *si tous ses états sont récurrents, respectivement transients.*

Le fait qu'un état soit récurrent signifie que la chaîne revient vers cet état presque sûrement, et donc qu'elle y revient infiniment souvent. Le fait qu'un état soit transient signifie que la chaîne a une probabilité positive de ne jamais retourner dans cet état.

La condition (2.3.2) n'est en général pas aisée à vérifier, car elle présuppose la connaissance exacte de la loi de τ_i. Toutefois, on peut obtenir une condition équivalente beaucoup plus facile à vérifier. Pour cela, nous commençons par démontrer une équation dite de renouvellement.

Proposition 2.3.2. *Pour tout $i, j \in \mathcal{X}$ et tout temps $n \in \mathbb{N}$ on a la relation*

$$\mathbb{P}_i\{X_n = j\} = \sum_{m=1}^{n} \mathbb{P}_i\{\tau_j = m\}\mathbb{P}_j\{X_{n-m} = j\} \; . \tag{2.3.3}$$

Démonstration. En décomposant sur les temps de premier passage en j, il vient

$$\begin{aligned}\mathbb{P}_i\{X_n = j\} &= \sum_{m=1}^{n} \mathbb{P}_i\{j \notin X_{[1,m-1]}, X_m = j, X_n = j\} \\ &= \sum_{m=1}^{n} \underbrace{\mathbb{P}_i\{X_n = j | j \notin X_{[1,m-1]}, X_m = j\}}_{=\mathbb{P}_i\{X_n=j|X_m=j\}=\mathbb{P}_j\{X_{n-m}=j\}} \underbrace{\mathbb{P}_i\{j \notin X_{[1,m-1]}, X_m = j\}}_{=\mathbb{P}_i\{\tau_j=m\}} \; , \end{aligned} \tag{2.3.4}$$

où nous avons utilisé la propriété des incréments indépendants. □

Nous pouvons maintenant prouver un critère de récurrence plus simple à vérifier que la définition (2.3.2).

Théorème 2.3.3. *Les deux conditions suivantes sont équivalentes:*

1. *L'état i est récurrent.*
2. *On a*

$$\sum_{n=0}^{\infty} \mathbb{P}_i\{X_n = i\} = +\infty \; . \tag{2.3.5}$$

Démonstration.

$\Rightarrow$: L'équation de renouvellement (2.3.3) permet d'écrire

$$\begin{aligned} S \coloneqq \sum_{n=0}^{\infty} \mathbb{P}_i\{X_n = i\} &= 1 + \sum_{n=1}^{\infty} \mathbb{P}_i\{X_n = i\} \\ &= 1 + \sum_{n=1}^{\infty} \sum_{m=1}^{n} \mathbb{P}_i\{\tau_i = m\}\mathbb{P}_i\{X_{n-m} = i\} \\ &= 1 + \sum_{m=1}^{\infty} \mathbb{P}_i\{\tau_i = m\} \sum_{n=m}^{\infty} \mathbb{P}_i\{X_{n-m} = i\} \\ &= 1 + \underbrace{\sum_{m=1}^{\infty} \mathbb{P}_i\{\tau_i = m\}}_{=1} \sum_{n=0}^{\infty} \mathbb{P}_i\{X_n = i\} = 1 + S \; . \end{aligned} \tag{2.3.6}$$

Comme $S \in [0, \infty]$, l'égalité $S = 1 + S$ implique nécessairement $S = +\infty$.

$\Leftarrow$: On ne peut pas directement inverser les implications ci-dessus. Cependant, on peut montrer la contraposée en définissant pour tout $0 < s < 1$ les séries entières

$$\psi(s) = \sum_{n=0}^{\infty} \mathbb{P}_i\{X_n = i\} s^n \ ,$$
$$\phi(s) = \sum_{n=1}^{\infty} \mathbb{P}_i\{\tau_i = n\} s^n \ . \tag{2.3.7}$$

Ces séries ont un rayon de convergence supérieur ou égal à 1 car leurs coefficients sont inférieurs ou égaux à 1. Un calcul analogue au calcul (2.3.6) ci-dessus donne alors

$$\begin{aligned}\psi(s) &= 1 + \sum_{m=1}^{\infty} \mathbb{P}_i\{\tau_i = m\} \sum_{n=m}^{\infty} \mathbb{P}_i\{X_{n-m} = i\} s^n \\ &= 1 + \sum_{m=1}^{\infty} \mathbb{P}_i\{\tau_i = m\} s^m \sum_{n=0}^{\infty} \mathbb{P}_i\{X_n = i\} s^n \\ &= 1 + \psi(s)\phi(s) \ , \end{aligned} \tag{2.3.8}$$

d'où

$$\psi(s) = \frac{1}{1 - \phi(s)} \ . \tag{2.3.9}$$

Par conséquent, si $\mathbb{P}_i\{\tau_i < \infty\} = \phi(1) < 1$, alors on obtient, en prenant la limite $s \nearrow 1$,

$$\sum_{n=0}^{\infty} \mathbb{P}_i\{X_n = i\} = \lim_{s \nearrow 1} \psi(s) = \frac{1}{1 - \phi(1)} < \infty \ . \tag{2.3.10}$$

□

Une application de ce critère de récurrence est le résultat important suivant.

Corollaire 2.3.4. *La marche aléatoire symétrique sur $\mathbb{Z}^d$ est récurrente pour $d = 1$ et $d = 2$ et transiente pour $d \geqslant 3$.*

Démonstration. Comme la marche aléatoire est invariante par translation, il suffit de vérifier que l'origine est récurrente, respectivement transiente.

1. En dimension $d = 1$, nous avons vu dans la Section 2.1 que $\mathbb{P}_0\{X_{2m} = 0\}$ se comportait en $1/\sqrt{\pi m}$ pour m grand. Il suit que
$$\sum_{n=0}^{\infty} \mathbb{P}_0\{X_n = 0\} = +\infty \ , \tag{2.3.11}$$
donc par le théorème précédent, l'origine est récurrente.
2. En dimension $d = 2$, on peut encore calculer explicitement $\mathbb{P}_0\{X_{2m} = 0\}$. Parmi les 4^{2m} trajectoires de longueur $2m$, il faut déterminer combien sont revenues à l'origine. Munissons le réseau $\mathbb{Z}^2$ des quatres directions Nord, Sud, Est et Ouest. Une trajectoire de longueur $2m$ partant et revenant de l'origine doit avoir fait autant de pas (disons k) vers le Nord que vers le Sud, et autant de pas (c'est-à-dire $m - k$) vers l'Est que vers l'Ouest. Le nombre de telles trajectoires est donc donné par
$$\sum_{k=0}^{m} \binom{2m}{2k} \binom{2k}{k} \binom{2(m-k)}{m-k} = \sum_{k=0}^{m} \frac{(2m)!}{[k!(m-k)!]^2} = \binom{2m}{m} \sum_{k=0}^{m} \binom{m}{k}^2 \ . \tag{2.3.12}$$

Cette dernière somme peut se simplifier grace aux identités suivantes:

$$\binom{2m}{m} = \binom{2m-1}{m} + \binom{2m-1}{m-1} = \binom{2m-2}{m} + 2\binom{2m-2}{m-1} + \binom{2m-2}{m-2}$$
$$= \cdots = \sum_{k=0}^{m} \binom{m}{k}\binom{m}{m-k} = \sum_{k=0}^{m} \binom{m}{k}^2 . \tag{2.3.13}$$

Il suit que

$$\mathbb{P}_0\{X_{2m} = 0\} = \frac{1}{4^{2m}}\binom{2m}{m}^2 \sim \frac{1}{\pi m} , \tag{2.3.14}$$

où la dernière équivalence découle de la formule de Stirling. Ceci implique la divergence de la somme des $\mathbb{P}_0\{X_n = 0\}$, donc la récurrence de l'origine.

3. En dimension $d = 3$, on obtient de manière analogue que le nombre de chemins revenus à l'origine après $6m$ pas est donné par

$$\binom{6m}{3m} \sum_{k_1+k_2+k_3=3m} \left(\frac{(3m)!}{k_1!k_2!k_3!}\right)^2 . \tag{2.3.15}$$

Cette fois-ci, nous ne disposons pas d'une expression exacte pour la somme. Nous pouvons toutefois l'estimer comme suit:

$$\sum_{k_1+k_2+k_3=3m} \left(\frac{(3m)!}{k_1!k_2!k_3!}\right)^2 \leqslant \left(\max_{k_1+k_2+k_3=3m} \frac{(3m)!}{k_1!k_2!k_3!}\right) \sum_{k_1+k_2+k_3=3m} \frac{(3m)!}{k_1!k_2!k_3!}$$
$$= \frac{(3m)!}{(m!)^3} \cdot 3^{3m} . \tag{2.3.16}$$

En effet, le maximum est atteint lorsque tous les k_i sont égaux, alors que la somme est égale au nombre de mots de longueur $3m$ qu'on peut écrire avec trois lettres différentes, avec répétition (les k_i correspondant au nombre de lettres de chaque type choisies). En utilisant la formule de Stirling, on trouve alors

$$\mathbb{P}_0\{X_{6m} = 0\} \leqslant \frac{3^{3m}}{6^{6m}}\binom{6m}{3m}\frac{(3m)!}{(m!)^3} \sim \frac{1}{2(\pi m)^{3/2}} . \tag{2.3.17}$$

Comme par ailleurs

$$\mathbb{P}_0\{X_{6m} = 0\} \geqslant \underbrace{\mathbb{P}_0\{X_{6m} = 0 | X_{6m-2} = 0\}}_{=\mathbb{P}_0\{X_2=0\}=1/6} \mathbb{P}_0\{X_{6m-2} = 0\} \tag{2.3.18}$$
$$\geqslant \left(\frac{1}{6}\right)^2 \mathbb{P}_0\{X_{6m-4} = 0\} , \tag{2.3.19}$$

les termes de la série des $\mathbb{P}_0\{X_{2n} = 0\}$ décroissent en $n^{-3/2}$, ce qui implique que la série est sommable, et donc que l'origine est un point transient.

4. En dimension $d \geqslant 4$, on montre de manière analogue que $\mathbb{P}_0\{X_{2n} = 0\}$ décroît comme $n^{-d/2}$, ce qui implique à nouveau la transience de l'origine. □

Nous avons utilisé à plusieurs reprises le fait qu'une marche aléatoire ne peut revenir au même endroit qu'aux temps pairs. On dit qu'elle a la période 2. Plus généralement, on introduit la définition suivante.

Définition 2.3.5. *La* période *d'un état $i \in \mathcal{X}$ est le nombre*

$$d_i = \operatorname{pgcd}\bigl\{n \geqslant 1 \colon \mathbb{P}_i\{X_n = i\} > 0\bigr\} \,. \tag{2.3.20}$$

Si $d_i = 1$, on dit que l'état i est apériodique. *Si tout $i \in \mathcal{X}$ est apériodique, on dit que la chaîne est apériodique.*

Remarque 2.3.6. Une chaîne régulière est apériodique. En effet, pour tout état i, il existe un état j tel que $p_{ij} > 0$. Par définition, il existe un temps n tel que $\mathbb{P}_k\{X_n = \ell\} > 0$ pour tout $k, \ell \in \mathcal{X}$. Par conséquent, on a $\mathbb{P}_i\{X_n = i\} > 0$ et aussi

$$\mathbb{P}_i\{X_{n+1} = i\} \geqslant \mathbb{P}_i\{X_1 = j, X_{n+1} = i\} = p_{ij}\mathbb{P}_j\{X_n = i\} > 0 \,. \tag{2.3.21}$$

Ceci implique que $d_i = \operatorname{pgcd}\{n, n+1\} = 1$.

Nous avons introduit la relation d'équivalence $i \sim j$ signifiant que i est atteignable depuis j est inversément. On montre assez facilement que les propriétés de récurrence/transience et la période sont constantes sur les classes d'équivalence. On peut alors parler de classes récurrentes ou transientes, et de la période d'une classe. Si la chaîne est irréductible, alors elle est respectivement récurrente, transiente ou apériodique si et seulement si elle admet un état récurrent, transient ou apériodique.

Proposition 2.3.7. *Si i et j sont dans la même classe récurrente, alors*

$$\mathbb{P}_i\{\tau_j < \infty\} = \mathbb{P}_j\{\tau_i < \infty\} = 1 \,. \tag{2.3.22}$$

Démonstration. Soit $A_M = \cup_{m=1}^{M}\{X_m = j\}$ l'événement "la chaîne visite le site j lors des M premiers pas". Alors

$$\lim_{M\to\infty} \mathbb{P}_i(A_M) = \sum_{m=1}^{\infty} \mathbb{P}_j\{\tau_j = m\} = 1 \,. \tag{2.3.23}$$

Soit n_0 le plus petit entier tel que $\mathbb{P}_j\{X_{n_0} = i\} > 0$. Alors pour tout $M > n_0$,

$$\begin{aligned}
\mathbb{P}_j\bigl(A_M \cap \{X_{n_0} = i\}\bigr) &= \sum_{n=1}^{M-n_0} \mathbb{P}_j\{X_{n_0} = i, \tau_j = n_0 + n\} \\
&= \sum_{n=1}^{M-n_0} \mathbb{P}_j\{X_{n_0} = i, j \notin X_{[1,n_0]}\}\mathbb{P}_i\{\tau_j = n\} \\
&\leqslant \mathbb{P}_j\{X_{n_0} = i\} \sum_{n=1}^{M-n_0} \mathbb{P}_i\{\tau_j = n\} \,.
\end{aligned} \tag{2.3.24}$$

La première égalité suit du fait que la chaîne ne peut pas retourner en j avant n_0 et visiter i au temps n_0, par définition de n_0. Nous faisons maintenant tendre M vers l'infini des deux côtés de l'inégalité. Le membre de gauche tend vers $\mathbb{P}_j\{X_{n_0} = i\}$ en vertu de (2.3.23). Il vient donc

$$\mathbb{P}_j\{X_{n_0} = i\} \leqslant \mathbb{P}_j\{X_{n_0} = i\}\mathbb{P}_i\{\tau_j < \infty\} \,. \tag{2.3.25}$$

Comme $\mathbb{P}_j\{X_{n_0} = i\} \neq 0$ et $\mathbb{P}_i\{\tau_j < \infty\} \leqslant 1$, on a nécessairement $\mathbb{P}_i\{\tau_j < \infty\} = 1$. □

2.4 Distributions stationnaires

Nous considérons une chaîne de Markov irréductible sur un ensemble dénombrable $\mathcal{X}$, de matrice de transition $P = (p_{ij})_{i,j\in\mathcal{X}}$.

Définition 2.4.1. *Une distribution de probabilité π sur $\mathcal{X}$ est dite* stationnaire *si elle satisfait*

$$\pi_j = \sum_{i\in\mathcal{X}} \pi_i p_{ij} \qquad \forall j \in \mathcal{X} \; . \tag{2.4.1}$$

Plus généralement, une mesure μ sur $\mathcal{X}$ (qui n'est pas nécessairement une mesure de probabilité) satisfaisant $\mu_j = \sum_{i\in\mathcal{X}} \mu_i p_{ij}$ pour tout $j \in \mathcal{X}$ est appelée une mesure invariante *de la chaîne.*

Dans le cas où $\mathcal{X}$ est fini, nous avons vu qu'une chaîne irréductible admettait toujours une distribution stationnaire. Dans le cas infini, ce n'est plus nécessairement le cas. Nous verrons par exemple que les marches aléatoires sur $\mathbb{Z}^d$ n'admettent pas de distribution stationnaire (par contre, elles admettent beaucoup de mesures invariantes).

Nous allons maintenant dériver une condition nécessaire et suffisante pour qu'une chaîne de Markov irréductible admette une distribution stationnaire, qui sera toujours unique dans ce cas. Un rôle important est joué par la quantité

$$\gamma_i^{(k)} = \mathbb{E}_k\Big(\sum_{n=1}^{\tau_k} 1_{\{X_n=i\}}\Big) \; , \tag{2.4.2}$$

c'est-à-dire le nombre moyen de passages en i entre deux passages en k. Intuitivement, si k est récurrent alors la chaîne revient infiniment souvent en k, et donc $\gamma_i^{(k)}$ devrait mesurer le temps moyen passé en i. On peut s'attendre à ce que ce temps corresponde à une mesure invariante.

Proposition 2.4.2. *Supposons la chaîne irréductible et récurrente. Alors on a $\forall k \in \mathcal{X}$:*

1. $\gamma_k^{(k)} = 1$;
2. $\gamma^{(k)}$ *est une mesure invariante;*
3. *Pour tout $i \in \mathcal{X}$, on a $0 < \gamma_i^{(k)} < \infty$;*
4. $\gamma^{(k)}$ *est l'unique mesure invariante telle que $\gamma_k^{(k)} = 1$.*

Démonstration.

1. Evident, puisque τ_k est fini presque sûrement, $X_{\tau_k} = k$ et $X_n \neq k$ pour $1 \leqslant n < \tau_k$.
2. Nous avons

$$\begin{aligned}
\gamma_i^{(k)} &= \mathbb{E}_k\Big(\sum_{n=1}^{\infty} 1_{\{X_n=i, n\leqslant\tau_k\}}\Big) = \sum_{n=1}^{\infty} \mathbb{P}_k\{X_n = i, n \leqslant \tau_k\} \\
&= \sum_{j\in\mathcal{X}} \sum_{n=1}^{\infty} \mathbb{P}_k\{X_{n-1} = j, n \leqslant \tau_k\} p_{ji} \\
&= \sum_{j\in\mathcal{X}} p_{ji} \sum_{m=0}^{\infty} \mathbb{P}_k\{X_m = j, m \leqslant \tau_k - 1\} \; .
\end{aligned} \tag{2.4.3}$$

Or la seconde somme dans cette expression peut s'écrire

$$\mathbb{E}_k\Big(\sum_{m=0}^{\tau_k-1} 1_{\{X_m=j\}}\Big) = \mathbb{E}_k\Big(\sum_{m=1}^{\tau_k} 1_{\{X_m=j\}}\Big) = \gamma_j^{(k)} \; , \tag{2.4.4}$$

vu que $\mathbb{P}_k\{X_0 = j\} = \delta_{kj} = \mathbb{P}_k\{X_{\tau_k} = j\}$. Ceci prouve l'invariance de la mesure $\gamma^{(k)}$.

3. L'invariance de la mesure implique que pour tout $n \geqslant 0$,

$$\gamma_i^{(k)} = \sum_{j \in \mathcal{X}} \gamma_j^{(k)} \mathbb{P}_j\{X_n = i\} \; . \tag{2.4.5}$$

En particulier, $1 = \gamma_k^{(k)} \geqslant \gamma_j^{(k)} \mathbb{P}_j\{X_n = k\}$ pour tout j. Comme par irréductibilité, il existe un n tel que $\mathbb{P}_j\{X_n = k\} > 0$, on en déduit que $\gamma_j^{(k)} < \infty$ pour tout j. D'autre part, on a aussi $\gamma_i^{(k)} \geqslant \mathbb{P}_k\{X_n = i\}$, qui est strictement positif pour au moins un n.

4. Soit λ une mesure invariante telle que $\lambda_k = 1$. Alors pour tout j on a

$$\lambda_j = \sum_{i \neq k} \lambda_i p_{ij} + p_{kj} \geqslant p_{kj} \; . \tag{2.4.6}$$

Il vient alors, en minorant λ_i par p_{ki} dans l'expression ci-dessus,

$$\begin{aligned} \lambda_j &\geqslant \sum_{i \neq k} p_{ki} p_{ij} + p_{kj} \\ &= \mathbb{P}_k\{X_2 = j, \tau_k \geqslant 2\} + \mathbb{P}_k\{X_1 = j, \tau_k \geqslant 1\} \end{aligned} \tag{2.4.7}$$

Par récurrence, on trouve donc pour tout $n \geqslant 1$ ($a \wedge b$ désigne le minimum de a et b)

$$\lambda_j \geqslant \sum_{m=1}^{n+1} \mathbb{P}_k\{X_m = j, \tau_k \geqslant m\} = \mathbb{E}_k\Bigg(\sum_{m=1}^{(n+1)\wedge\tau_k} 1_{\{X_m = j\}} \Bigg) \; . \tag{2.4.8}$$

Lorsque n tend vers l'infini, le membre de droite tend vers $\gamma_j^{(k)}$. On a donc $\lambda_j \geqslant \gamma_j^{(k)}$ pour tout j. Par conséquent, $\mu = \lambda - \gamma^{(k)}$ est une mesure invariante, satisfaisant $\mu_k = 0$. Comme $\mu_k = \sum_j \mu_j \mathbb{P}_j\{X_n = k\}$ pour tout n, l'irréductibilité implique $\mu_j = 0$ $\forall j$, donc nécessairement $\lambda = \gamma^{(k)}$. □

Nous pouvons maintenant énoncer et démontrer le théorème principal sur les distributions stationnaires.

Théorème 2.4.3. *Pour une chaîne de Markov irréductible, les propriétés suivantes sont équivalentes :*

1. *Il existe une distribution stationnaire.*
2. *Il existe un état $k \in \mathcal{X}$ tel que*

$$\mu_k := \mathbb{E}_k(\tau_k) < \infty \; . \tag{2.4.9}$$

3. *La relation (2.4.9) est vérifiée pour tout $k \in \mathcal{X}$.*

De plus, si ces propriétés sont vérifiées, alors la distribution stationnaire est unique, et donnée par

$$\pi_i = \frac{1}{\mu_i} \qquad \forall i \in \mathcal{X} \; . \tag{2.4.10}$$

Démonstration.

$2 \Rightarrow 1$: Si $\mu_k < \infty$ alors k est récurrent, donc la chaîne, étant irréductible, est récurrente. Par la proposition précédente, $\gamma^{(k)}$ est l'unique mesure invariante prenant valeur 1 en k. Or nous avons

$$\sum_{j\in\mathcal{X}} \gamma_j^{(k)} = \mathbb{E}_k\Bigg(\sum_{n=1}^{\tau_k} \underbrace{\sum_{j\in\mathcal{X}} 1_{\{X_n=j\}}}_{=1}\Bigg) = \mathbb{E}_k(\tau_k) = \mu_k < \infty \,. \tag{2.4.11}$$

Par conséquent, la mesure π définie par $\pi_j = \gamma_j^{(k)}/\mu_k$ est une mesure de probabilité invariante, c'est-à-dire une distribution stationnaire.

$1 \Rightarrow 3$: Soit π une distribution stationnaire, et $k \in \mathcal{X}$. Alors $\hat{\gamma}$ défini par $\hat{\gamma}_j = \pi_j/\pi_k$ est une mesure invariante telle que $\hat{\gamma}_k = 1$. Par la proposition précédente, on a nécessairement $\hat{\gamma} = \gamma^{(k)}$. Il suit par le même calcul que ci-dessus

$$\mathbb{E}_k(\tau_k) = \sum_{j\in\mathcal{X}} \hat{\gamma}_j = \frac{\sum_j \pi_j}{\pi_k} = \frac{1}{\pi_k} < \infty \,. \tag{2.4.12}$$

$3 \Rightarrow 2$: Evident.

Dans ce cas, l'unicité de la mesure suit de celle de $\gamma^{(k)}$, et la relation (2.4.10) suit de (2.4.12). □

Ce résultat motive la définition suivante.

Définition 2.4.4. *Un état $i \in \mathcal{X}$ tel que*

$$\mathbb{E}_i(\tau_i) < \infty \tag{2.4.13}$$

est appelé récurrent positif. *Un état récurrent i qui n'est pas récurrent positif est appelé* récurrent nul. *La chaîne est dite* récurrente positive *si tous ses états le sont. C'est par exemple le cas s'il existe un tel état, et que la chaîne est irréductible.*

Une chaîne irréductible admet donc une distribution stationnaire si et seulement si elle est récurrente positive.

Exemple 2.4.5. Les marches aléatoires sur $\mathbb{Z}^d$ n'admettent pas de distribution stationnaire. En effet, si π était une distribution stationnaire, l'invariance par translation impliquerait que tout translaté de π serait encore une distribution stationnaire. Mais nous savons que si une telle distribution existait, alors elle serait unique. π devrait donc être uniforme, mais il n'existe pas de mesure de probabilité uniforme sur $\mathbb{Z}^d$.

Par contre, toutes les mesures uniformes sont invariantes. Les fonctions pas nécessairement positives satisfaisant (2.4.1) dans le cas d'une marche aléatoire symétrique sont appelées *harmoniques*. Toutes les fonctions affines sont harmoniques, mais en dimension supérieure ou égale à 2, il y en a beaucoup d'autres.

Nous avons donc obtenu le résultat suivant :

Théorème 2.4.6. *La marche aléatoire symétrique est récurrente nulle en dimensions $d = 1$ et $d = 2$.*

2.5 Convergence vers la distribution stationnaire

Dans le cas fini, nous avons montré que si la chaîne était régulière, alors la loi de X_n convergeait vers la distribution stationnaire. Dans le cas d'un espace infini, une chaîne de Markov ne peut jamais être régulière : les probabilités de transition étant sommables, elles ne peuvent être minorées par une quantité strictement positive. Il s'avère toutefois que la récurrence positive et l'apériodicité suffisent à garantir la convergence vers la distribution stationnaire.

Théorème 2.5.1. *Soit $\{X_n\}_{n\geqslant 0}$ une chaîne de Markov irréductible, apériodique et récurrente positive, et soit π son unique distribution stationnaire. Alors pour toute distribution initiale ν, on a*

$$\lim_{n\to\infty} \mathbb{P}_\nu\big\{X_n = j\big\} = \pi_j \qquad \forall j \in \mathcal{X} \,. \tag{2.5.1}$$

Démonstration. Nous allons généraliser la preuve de Doeblin, déjà vue dans le cas fini (voir le Théorème 1.4.9).

- Nous introduisons une chaîne de Markov $(X_n, Y_n)_{n\geqslant 0}$ sur $\mathcal{X}\times\mathcal{X}$, de probabilités de transition
$$p^\star_{(i,j),(k,l)} = p_{ik}p_{jl} \,, \tag{2.5.2}$$
et de distribution initiale $\rho = \nu\otimes\pi$. Dans ce cas, X_n et Y_n sont deux chaînes indépendantes de matrice de transition P, et de distributions initiales ν et π.
- Le seul point non trivial de la généralisation est de montrer que $P^\star$ est irréductible et apériodique. Pour cela, fixons un état $k\in\mathcal{X}$. Considérons d'abord l'ensemble
$$\Gamma_k = \big\{n\in\mathbb{N} : \mathbb{P}_k\{X_n = k\} > 0\big\} \,. \tag{2.5.3}$$
La propriété de Markov implique que si $n, m \in \Gamma_k$, alors $n+m\in\Gamma_k$. D'autre part, par définition de l'apériodicité, $\operatorname{pgcd}\Gamma_k = 1$. Nous prétendons qu'il existe un n_0 tel que tout $t\geqslant n_0$ appartienne à Γ_k.
Pour cela, supposons d'abord qu'il existe $n, m\in\Gamma_k$ premiers entre eux. Par le théorème de Bezout, il existe des entiers $p, q\geqslant 1$ tels que $pn - qm = \pm 1$. Quitte à intervertir n et m, on peut supposer que $pn - qm = 1$. Soit $n_0 = qnm$. Alors pour $1\leqslant r\leqslant n$, on a $n_0 + r = qnm + r(pn-qm) = qm(n-r) + rpn \in \Gamma_k$. Il suffit alors d'écrire tout $t > n_0$ comme $t = n_0 + r + ns$ avec $1\leqslant r\leqslant n$ pour conclure que $t\in\Gamma_k$.
Il se pourrait que $\operatorname{pgcd}\Gamma_k = 1$ sans que cet ensemble ne contienne deux entiers premiers entre eux. Mais par le théorème de Bezout, il existe forcément un ensemble d'élements de Γ_k dont une combinaison linéaire vaut 1, et le raisonnement ci-dessus s'adapte facilement à ce cas.
- Fixons des états $i, j, k, \ell\in\mathcal{X}$. P étant supposée irréductible, il existe $r\in\mathbb{N}$ tel que $\mathbb{P}_i\{X_r = k\} > 0$. Comme pour tout $n\geqslant n_0$,
$$\mathbb{P}_i\{X_{r+n} = k\} \geqslant \mathbb{P}_i\{X_r = k\}\mathbb{P}_k\{X_n = k\} > 0 \,, \tag{2.5.4}$$
il suit que $\mathbb{P}_i\{X_n = k\} > 0$ pour tous les $n\geqslant n_0 + r$. Pour des raisons similaires, il existe $m_0, s\in\mathbb{N}$ tels que $\mathbb{P}_j\{X_m = \ell\} > 0$ pour tous les $m\geqslant m_0 + s$. Par conséquent, il existe un temps M tel que $\mathbb{P}^\star_{(i,j)}\{(X_t, Y_t) = (k,\ell)\} > 0$ pour tous les $t\geqslant M$. Ceci implique que la chaîne composée est irréductible et apériodique.
- Comme la chaîne composée admet manifestement la distribution invariante $\pi\otimes\pi$, le Théorème 2.4.3 implique qu'elle est récurrente positive.

- Le reste de la preuve est identique au cas fini. On introduit le temps τ_A de premier passage sur la diagonale $A = \{(i,i)\colon i \in \mathcal{X}\}$, et on montre comme dans le cas fini que

$$|\mathbb{P}_\nu\{X_n = j\} - \pi_j| \leqslant 2\mathbb{P}_\rho\{\tau_A > n\} \; . \tag{2.5.5}$$

  La Proposition 2.3.7 implique que τ_A est fini presque sûrement, et donc que la différence ci-dessus tend vers zéro pour $n \to \infty$. □

Un problème important, mais en général difficile, est d'estimer la vitesse de convergence vers la distribution stationnaire. La preuve de Doeblin fournit (voir le cas fini) l'estimation

$$\sum_{j\in\mathcal{X}} \big|\mathbb{P}_\nu\{X_n = j\} - \pi_j\big| \leqslant 2\mathbb{P}_{\nu\otimes\pi}\{\tau_A > n\} \; , \tag{2.5.6}$$

qui est utile dans les cas où on arrive à contrôler le temps de couplage τ_A.

Une autre approche est basée sur la théorie spectrale. Nous avons vu que la matrice de transition P admet la valeur propre 1, avec comme vecteurs propres à gauche et à droite, respectivement, la distribution stationnaire π et le vecteur $\mathbf{1}$ dont toutes les composantes sont égales à 1:

$$\pi P = \pi \qquad \text{et} \qquad P\mathbf{1} = \mathbf{1} \; . \tag{2.5.7}$$

Soit μ un vecteur ligne tel que

$$\mu\mathbf{1} = \sum_{i\in\mathcal{X}} \mu_i = 0 \; . \tag{2.5.8}$$

Alors $\mu P\mathbf{1} = \mu\mathbf{1} = 0$, ce qui montre que le sous-espace $\mathbf{1}_\perp = \{\mu \in \mathbb{R}^{\mathcal{X}}\colon \sum_i \mu_i = 0\}$ est invariant (à droite) par P : $\mathbf{1}_\perp P \subset \mathbf{1}_\perp$. On remarquera que les éléments de $\mathbf{1}_\perp$ ne sont pas des mesures, mais des mesures signées. Toutefois, pour certains $\mu \in \mathbf{1}_\perp$, la somme $\pi + \mu$ est une mesure de probabilité : il suffit pour celà que $\mu_i \geqslant -\pi_i$ pour tout $i \in \mathcal{X}$.

Si $\mu \in \mathbf{1}_\perp$ est un vecteur propre à gauche de P, de valeur propre λ, et si P est irréductible, apériodique et récurrente positive, alors on aura nécessairement $|\lambda| < 1$. En effet, si ce n'était pas le cas, la loi de la chaîne de condition initiale $\pi + \varepsilon\mu$ (ε un réel suffisamment petit) ne convergerait pas vers π.

Cette observation donne une caractérisation de la vitesse de convergence en termes de *trou spectral* : Soit λ_0 la plus grande valeur propre de P de module strictement inférieur à 1. On sait que le vecteur propre correspondant se trouve dans $\mathbf{1}_\perp$ (éventuellement avec des composantes complexes). Alors la loi de X_n converge exponentiellement vite vers la distribution stationnaire π, à vitesse $|\lambda_0|^n$. Toutefois, la détermination du trou spectral est en général un problème difficile si l'ensemble $\mathcal{X}$ est grand. Le problème est un peu plus facile dans le cas des chaînes de Markov dites *réversible*.

Les chaînes de Markov réversibles se prêtent mieux à une étude spectrale que les chaînes non réversibles. Pour le voir, supposons la chaîne irréductible et récurrente positive, de distribution stationnaire π, et introduisons le produit scalaire

$$\langle f|g\rangle_\pi = \sum_{i\in\mathcal{X}} \pi_i \overline{f}_i g_i \; . \tag{2.5.9}$$

Alors on a

$$\langle f|Pg\rangle_\pi = \sum_{i\in\mathcal{X}} \pi_i \overline{f}_i \sum_{j\in\mathcal{X}} p_{ij} g_j = \sum_{j\in\mathcal{X}} \pi_j \sum_{i\in\mathcal{X}} p_{ji} \overline{f}_i g_j = \langle Pf|g\rangle_\pi \; . \tag{2.5.10}$$

Autrement dit, l'opérateur linéaire P est autoadjoint dans $\ell^2(\mathbb{C}\,,\pi)$.

Un résultat classique de la théorie des espaces de Hilbert dit que toutes les valeurs propres de P sont réelles, et que les espaces propres associés sont orthogonaux. En effet, soient x_1 et x_2 deux vecteurs propres à droite de P, de valeurs propres respectives λ_1 et λ_2. Alors

$$(\overline{\lambda}_1 - \lambda_2)\langle x_1|x_2\rangle_\pi = \langle \lambda_1 x_1|x_2\rangle_\pi - \langle x_1|\lambda_2 x_2\rangle_\pi = \langle Px_1|x_2\rangle_\pi - \langle x_1|Px_2\rangle_\pi = 0\,. \quad (2.5.11)$$

D'une part, prenant $x_1 = x_2$, on obtient que λ_1 est réelle. D'autre part, si $\lambda_1 \neq \lambda_2$, on obtient l'orthogonalité de x_1 et x_2. On sait de plus que P est diagonalisable.

On a alors une représentation variationnelle du trou spectral :

$$|\lambda_0| = \sup_{x\colon \langle x|\mathbf{1}\rangle_\pi = 0} \frac{\langle x|Px\rangle_\pi}{\langle x|x\rangle_\pi}\,. \quad (2.5.12)$$

2.6 Exercices

Exercice 2.1. Montrer que la loi du temps de premier passage en i ($i \neq 0$) de la marche aléatoire symétrique unidimensionnelle est donnée par

$$\mathbb{P}\big\{\tau_i = n\big\} = \begin{cases} \dfrac{|i|}{n}\mathbb{P}\{X_n = i\} & \text{pour } n \in \{|i|, |i|+2, \dots\}\,, \\ 0 & \text{sinon}\,. \end{cases}$$

(Théorème 2.1.5). En déduire que $\mathbb{E}(\tau_i) = +\infty$.
Indications:

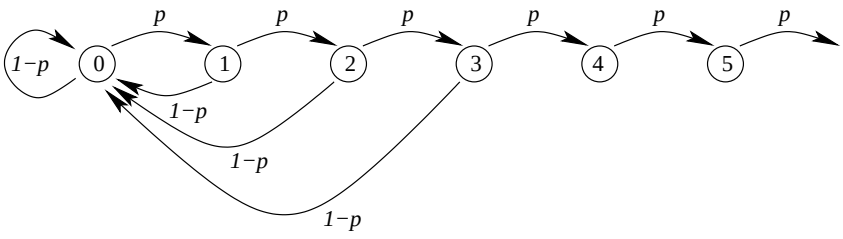
1. Par symétrie, on peut supposer $i > 0$.
2. Ecrire l'événement $\{\tau_i = n\}$ à l'aide des événements $\{X_{n-1} = i-1\}$ et $\{\tau_i \leqslant n-2\}$.
3. A l'aide du principe de réflexion, montrer que

$$\mathbb{P}\big\{\tau_i = n\big\} = \frac{1}{2}\big[\mathbb{P}\big\{X_{n-1} = i-1\big\} - \mathbb{P}\big\{X_{n-1} = i+1\big\}\big]$$

et conclure par un calcul direct.

Exercice 2.2. On considère une chaîne de Markov sur $\mathbb{N}$ de probabilités de transition

$$p_{ij} = \begin{cases} p & \text{si } j = i+1\,, \\ 1-p & \text{si } j = 0\,, \\ 0 & \text{sinon}\,. \end{cases}$$

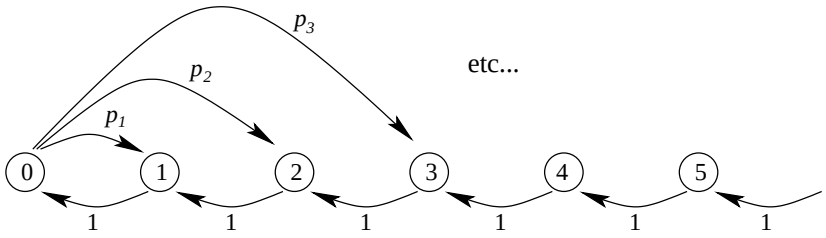


1. Pour quelles valeurs de p la chaîne est-elle irréductible?
   Pour le reste du problème, on suppose que p est tel que la chaîne soit irréductible.
2. On suppose $X_0 = 0$. Soit
$$\tau_0 = \inf\{n > 0 \colon X_n = 0\}$$
   le temps de premier retour en 0. Montrer que
$$\tau_0 = n \quad \Rightarrow \quad X_m = m \ , \qquad m = 0, 1, \ldots, n-1 \ .$$
   En déduire la loi de τ_0.
3. Montrer que l'état 0 est récurrent.
4. Montrer que l'état 0 est récurrent positif.
5. Montrer que l'état 0 est apériodique.
6. Soit π l'unique distribution stationnaire de la chaîne. Calculer π_0 à l'aide de $\mathbb{E}_0(\tau_0)$.
7. Exprimer π_i en fonction de π_{i-1} et en déduire π_i pour tout i.

Exercice 2.3. Un client arrive dans une banque. Devant l'unique guichet, il y a une queue de longueur aléatoire L. La loi de L est donnée par $\mathbb{P}\{L = k\} = p_k$, $k = 1, 2, \ldots$ (on suppose $\sum_{k \geqslant 1} p_k = 1$).

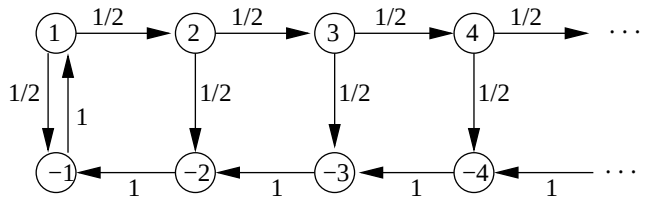
On admet que chaque client est servi pendant un intervalle de temps de longueur 1. Une fois le premier client servi, notre client avance d'une unité dans la file. Une fois servi, on suppose que le client se place à nouveau au bout de la queue.

On modélise la situation par la chaîne de Markov suivante :



1. Sous quelle condition sur les p_k la chaîne est-elle irréductible?
   Pour le reste du problème, on suppose les p_k tels que la chaîne soit irréductible.
2. On suppose $X_0 = 0$. Soit
$$\tau_0 = \inf\{n > 0 \colon X_n = 0\}$$
   le temps de premier retour en 0. Déterminer sa loi.
3. Montrer que l'état 0 est récurrent.
4. Donner une condition nécessaire et suffisante sur les p_k pour que l'état 0 soit récurrent positif.
5. Donner une condition nécessaire et suffisante sur les p_k pour que l'état 0 soit apériodique.
6. Soit π l'unique distribution stationnaire de la chaîne. Calculer π_0 à l'aide de $\mathbb{E}_0(\tau_0)$.
7. Par récurrence, déterminer π_i pour tout i.

Exercice 2.4. On considère la chaîne de Markov suivante sur $\mathcal{X} = \mathbb{Z}^*$:



1. La chaîne est-elle irréductible?
2. L'état 1 est-il récurrent?
3. L'état 1 est-il récurrent positif?
4. L'état 1 est-il apériodique?
5. Soit π la distribution stationnaire de la chaîne. Calculer π_1.
6. Calculer π_i pour tout $i \in \mathcal{X}$.

Exercice 2.5. On considère une marche aléatoire unidimensionnelle symétrique sur l'ensemble $\{0, 1, \ldots, N\}$ avec conditions aux bords absorbantes, c'est-à-dire que l'on suppose les états 0 et N absorbants. Soit

$$\tau = \tau_0 \wedge \tau_N = \inf\big\{n \geqslant 0 \colon X_n \in \{0, N\}\big\}$$

le temps d'absorption, et soit

$$p(i) = \mathbb{P}_i\{X_\tau = N\} \ .$$

1. Déterminer $p(0)$ et $p(N)$.
2. Montrer que pour tout $i \in \{1, \ldots, N-1\}$, on a

$$p(i) = \frac{1}{2}\big[p(i-1) + p(i+1)\big] \ .$$

   Une fonction $f : \mathbb{Z} \supset A \to \mathbb{R}$ telle que $f(i) = \frac{1}{2}[f(i-1) + f(i+1)]$ pour tout $i \in A$ est appelée *harmonique* (discrète).
3. Montrer (par l'absurde) le *principe du maximum*: Une fonction harmonique sur A ne peut atteindre son minimum et son maximum qu'au bord de A (on pourra supposer A de la forme $A = \{a, a+1, \ldots, b-1, b\}$, dans ce cas son bord est $\partial A = \{a, b\}$).
4. Montrer que si f et g sont deux fonctions harmoniques sur A, alors toute combinaison linéaire de f et g est encore harmonique.
5. Montrer que si f et g sont deux fonctions hamoniques sur A, qui coïncident sur le bord de A, alors elles sont égales partout dans A (considérer $f - g$).
6. Montrer que toute fonction linéaire $f(i) = ci + h$ est harmonique.
7. En utilisant les points 1., 2., 5. et 6., déterminer la fonction p.

Exercice 2.6. On considère une marche aléatoire symétrique sur $\mathcal{X} = \{0, 1, \ldots, N\}$, avec conditions au bord absorbantes, c'est-à-dire que dès que la marche atteint l'un des états 0 ou N, elle y reste indéfiniment. Soit

$$\tau = \inf\{n \geqslant 0 \colon X_n \in \{0, N\}\}$$

le temps d'absorption. Par convention, $\tau = 0$ si $X_0 \in \{0, N\}$.

1. Montrer que pour tout $i \in \{1, \dots, N-1\}$,

$$\mathbb{P}_i\{\tau = n\} = \frac{1}{2}\big[\mathbb{P}_{i-1}\{\tau = n-1\} + \mathbb{P}_{i+1}\{\tau = n-1\}\big] \ .$$

2. Soit $f(i) = \mathbb{E}_i(\tau)$. Que valent $f(0)$ et $f(N)$?
3. Déduire de 1. que

$$f(i) = \frac{1}{2}\big[f(i-1) + f(i+1)\big] + 1 \ . \tag{2.6.1}$$

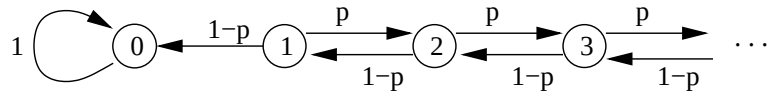
4. Montrer que $f(i) = -i^2$ satisfait l'équation (2.6.1).
5. Une fonction $g : \mathcal{X} \to \mathbb{R}$ est dite harmonique si

$$g(i) = \frac{1}{2}\big[g(i-1) + g(i+1)\big] \qquad \forall i \in \mathcal{X} \ .$$

Montrer que si f satisfait (2.6.1) et g est harmonique, alors $f + g$ satisfait (2.6.1).

6. Montrer que si f_1 et f_2 satisfont (2.6.1), alors $f_1 - f_2$ est harmonique. Déduire du principe du maximum que si f_1 et f_2 coïncident en 0 et en N, alors elles sont égales partout.
7. Sachant que les fonctions linéaires sont harmoniques, déterminer $\mathbb{E}_i(\tau)$.

Exercice 2.7. On considère une marche aléatoire sur $\mathbb{N}$, absorbée en 0:



On note $h(i) = \mathbb{P}_i\{\tau_0 < \infty\}$.

1. Déterminer $h(0)$.
2. Montrer que

$$h(i) = (1-p)h(i-1) + ph(i+1) \tag{2.6.2}$$

pour tout $i \geqslant 1$.

3. On considère le cas $p = 1/2$. En se basant sur les propriétés des fonctions harmoniques dérivées dans les exercices précédents, déterminer $h(i)$ pour tout i.
Indication: $h(i)$ est une probabilité.
4. Nous considérons à partir de maintenant le cas général $0 < p < 1$. Montrer que tout h solution de (2.6.2) satisfait encore le principe du maximum: pour tout ensemble $A = \{a, a+1, \dots, b-1, b\} \subset \mathbb{N}$, h atteint son maximum au bord $\partial A = \{a, b\}$ de A.
5. Montrer que deux solutions de l'équation (2.6.2), coïncidant sur le bord de A, sont égales partout dans A.
6. Montrer que $h(i) = \alpha^i$ satisait l'équation (2.6.2) pour certaines valeurs de α qu'on déterminera.
7. En déduire l'unique solution pour h lorsque $p < 1/2$. Trouver deux solutions possibles lorsque $p > 1/2$.

Chapitre 3

Application aux algorithmes MCMC

3.1 Méthodes Monte Carlo

On appelle *méthodes Monte Carlo* un ensemble d'algorithmes stochastiques permettant d'estimer numériquement des grandeurs pouvant être considérées comme des espérances. En voici pour commencer un exemple très simple.

Exemple 3.1.1 (Calcul d'un volume)**.** On aimerait calculer numériquement le volume $|V|$ d'un sous-ensemble compact V de $\mathbb{R}^N$. On suppose que V est donné par un certain nombre M d'inégalités:

$$V = \left\{ x \in \mathbb{R}^N \colon f_1(x) \geqslant 0, \ldots, f_M(x) \geqslant 0 \right\} . \tag{3.1.1}$$

Par exemple, si $M = 1$ et $f_1(x) = 1 - \|x\|^2$, alors V est une boule. Dans ce cas, bien sûr, le volume de V est connu explicitement. On s'intéresse à des cas où V a une forme plus compliquée, par exemple une intersection d'un grand nombre de boules et de demi-plans. Dans la suite nous supposerons, sans limiter la généralité, que V est contenu dans le cube unité $[0,1]^N$.

Une première méthode de calcul numérique du volume consiste à discrétiser l'espace. Divisons le cube $[0,1]^N$ en cubes de coté ε (avec ε de la forme $1/K$, $K \in \mathbb{N}$). Le nombre total de ces cubes est égal à $1/\varepsilon^N = K^N$. On compte alors le nombre n de cubes dont le centre est contenu dans V, et le volume de V est approximativement égal à $n\varepsilon^N$. Plus précisément, on peut encadrer $|V|$ par $n_-\varepsilon^N$ et $n_+\varepsilon^N$, où n_- est le nombre de cubes entièrement contenus dans V, et n_+ est le nombre de cubes dont l'intersection avec V est vide, mais effectuer ces tests n'est en général pas aisé numériquement.

Quelle est la précision de l'algorithme? Si le bord ∂V est raisonnablement lisse, l'erreur faite pour ε petit est de l'ordre de la surface du bord fois ε. Pour calculer $|V|$ avec une précision donnée δ, il faut donc choisir ε d'ordre $\delta/|\partial V|$. Cela revient à un nombre de cubes d'ordre

$$\left(\frac{|\partial V|}{\delta} \right)^N , \tag{3.1.2}$$

ou encore, comme on effectue M tests pour chaque cube, à un nombre de calculs d'ordre $(M|\partial V|/\delta)^N$. Ce nombre ne pose pas de problème pour les petites dimensions ($N = 1, 2$ ou 3 par exemple), mais croît vite avec la dimension N.

Une alternative intéressante pour les N grands est fournie par l'algorithme de Monte Carlo. Dans ce cas, on génère une suite $X_1, X_2, \ldots, X_n, \ldots$ de variables aléatoires indépendantes, identiquement distribuées (i.i.d.), de loi uniforme sur $[0,1]^N$. Ceci est facile à réaliser numériquement, car on dispose de générateurs de nombres (pseudo-)aléatoires distribués uniformément sur $[0,1]$ (ou plutôt sur $\{0,1,\ldots,n_{\max}\}$ où $n_{\max}$ est un entier du genre $2^{31}-1$, mais en divisant ces nombres par $n_{\max}$, on obtient de bonnes approximations de variables uniformes sur $[0,1]$). Il suffit alors de considérer des N-uplets de tels nombres.

Considérons alors les variables aléatoires i.i.d.

$$Y_i = 1_{\{X_i \in V\}} , \qquad i = 1, 2, \ldots . \tag{3.1.3}$$

On aura

$$\mathbb{E}(Y_i) = \mathbb{P}\{X_i \in V\} = |V| . \tag{3.1.4}$$

Les moyennes

$$S_n = \frac{1}{n} \sum_{i=1}^{n} Y_i \tag{3.1.5}$$

ont espérance $\mathbb{E}(S_n) = |V|$ et variance $\operatorname{Var} S_n = \operatorname{Var} Y_1/n$. La loi faible des grands nombres implique que

$$\lim_{n\to\infty} \mathbb{P}\big\{\big|S_n - \mathbb{E}(S_n)\big| > \delta\big\} = 0 \tag{3.1.6}$$

pour tout $\delta > 0$. Par conséquent, S_n devrait donner une bonne approximation du volume $|V|$ lorsque n est suffisamment grand. (La loi forte des grands nombres affirme par ailleurs que S_n tend vers $|V|$ presque sûrement, c'est-à-dire que S_n n'est plus vraiment aléatoire à la limite.)

Pour savoir comment choisir n en fonction de la précision désirée, il nous faut estimer la probabilité que $|S_n - |V|| > \delta$, pour n grand mais fini. Une première estimation est fournie par l'inégalité de Bienaymé–Chebychev, qui affirme que

$$\mathbb{P}\big\{\big|S_n - \mathbb{E}(S_n)\big| > \delta\big\} \leqslant \frac{\operatorname{Var}(S_n)}{\delta^2} = \frac{\operatorname{Var}(Y_1)}{\delta^2 n} < \frac{1}{\delta^2 n} , \tag{3.1.7}$$

où nous avons utilisé le fait que $\operatorname{Var}(Y_1) \leqslant \mathbb{E}(Y_1^2) \leqslant 1$. On peut donc affirmer que pour que la probabilité de faire une erreur supérieure à δ soit inférieure à ε, il suffit de choisir

$$n > \frac{1}{\delta^2 \varepsilon} . \tag{3.1.8}$$

Comme pour chaque i, il faut générer N variables aléatoires, et effectuer M tests, le nombre de calculs nécessaires est d'ordre $MN/(\delta^2\varepsilon)$. L'avantage de cette méthode est que ce nombre ne croît que linéairement avec N, par opposition à la croissance exponentielle dans le cas de la discrétisation. On notera toutefois que contrairement à la discrétisation, qui donne un résultat certain aux erreurs près, la méthode de Monte Carlo ne fournit que des résultats vrais avec très grande probabilité (d'où son nom).

Remarquons que l'estimation (3.1.7) est assez pessimiste, et peut être considérablement améliorée. Par exemple, le théorème de la limite centrale montre que

$$\lim_{n\to\infty} \mathbb{P}\left\{ \frac{(S_n - \mathbb{E}(S_n))^2}{\operatorname{Var}(S_n)} > \eta^2 \right\} = \int_{|x|>\eta} \frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}} \, \mathrm{d}x , \tag{3.1.9}$$

dont le membre de droite décroît comme $\mathrm{e}^{-\eta^2/2}$ pour η grand. Cela indique que pour n grand,

$$\mathbb{P}\big\{|S_n - |V|| > \delta\big\} \simeq \mathrm{e}^{-n\delta^2/2\,\mathrm{Var}(Y_1)} \ . \tag{3.1.10}$$

Ceci permet d'améliorer le critère (3.1.8) en

$$n > const\ \frac{\log(1/\varepsilon)}{\delta^2} \ , \tag{3.1.11}$$

d'où un nombre de calculs d'ordre $NM\log(1/\varepsilon)/\delta^2$. Cette estimation n'est pas une borne rigoureuse, contrairement à (3.1.8), parce qu'on n'a pas tenu compte de la vitesse de convergence dans (3.1.9), qui par ailleurs ne s'applique que pour η indépendant de ε. On devrait plutôt utiliser des estimations provenant de la théorie des grandes déviations, que nous n'aborderons pas ici. Les résultats sont toutefois qualitativement corrects.

A titre d'illustration, supposons qu'on veuille déterminer le volume d'un domaine de dimension $N = 1000$, défini par $M = 10$ inégalités, avec une précision de $\delta = 10^{-4}$. La méthode de discrétisation nécessite un nombre de calculs d'ordre 10^{5000}, ce qui est irréalisable avec les ordinateurs actuels. La méthode de Monte Carlo, par contre, fournit un résultat de la même précision, sûr avec probabilité $1 - 10^{-6}$, avec un nombre de calculs d'ordre $\log(10^6) \cdot 10^{12} \simeq 10^{13}$, ce qui ne prend que quelques minutes sur un PC.

La méthode de Monte Carlo se généralise facilement à d'autres problèmes que des calculs de volume. Supposons par exemple donné un espace probabilisé $(\Omega, \mathcal{F}, \mu)$, et une variable aléatoire $Y : \Omega \to \mathbb{R}$. On voudrait estimer l'espérance de Y. Pour cela, l'algorithme de Monte Carlo consiste à générer des variables aléatoires indépendantes $Y_1, Y_2, \ldots, Y_n, \ldots$, toutes de loi μY^{-1}, et de calculer leur moyenne. Cette moyenne doit converger vers l'espérance cherchée (pourvu que Y soit intégrable).

Cet algorithme n'est toutefois efficace que si l'on arrive à générer les variables aléatoires Y_i de manière efficace. Une fois de plus, ceci est relativement facile en dimension faible, mais devient rapidement difficile lorsque la dimension croît.

Exemple 3.1.2 (Cas unidimensionnel)**.** Une variable aléatoire unidimensionnelle Y s'obtient facilement à partir d'une variable de loi uniforme. Soit en effet U une variable uniforme sur $[0, 1]$. Sa fonction de répartition est donnée par

$$F_U(u) = \mathbb{P}\{U \leqslant u\} = u \qquad \text{pour } 0 \leqslant u \leqslant 1. \tag{3.1.12}$$

On cherche une fonction φ telle que la variable $Y = \varphi(u)$ admette la fonction de répartition prescrite $F_Y(y)$. Or on a

$$F_Y(y) = \mathbb{P}\{Y \leqslant y\} = \mathbb{P}\{\varphi(U) \leqslant y\} = \mathbb{P}\{U \leqslant \varphi^{-1}(y)\} = \varphi^{-1}(y) \ . \tag{3.1.13}$$

Il suffit donc de prendre $Y = F_Y^{-1}(U)$.

Par exemple, pour générer une variable de loi exponentielle, dont la fonction de répartition est donnée par $F_Y(y) = 1 - \mathrm{e}^{-\lambda y}$, il suffit de prendre

$$Y = -\frac{1}{\lambda}\log(1 - U) \ . \tag{3.1.14}$$

Pour la loi normale, cette méthode nécessiterait le calcul approché de sa fonction de répartition, ce qui est numériquement peu efficace. Il existe toutefois une alternative permettant d'éviter ce calcul. Soient en effet U et V deux variables aléatoires indépendantes,

de loi uniforme sur $[0,1]$. On introduit successivement les variables

$$\begin{aligned} R &= \sqrt{-2\log(1-U)}\ , & Y_1 &= R\cos\Phi\ , \\ \Phi &= 2\pi V\ , & Y_2 &= R\sin\Phi\ . \end{aligned} \tag{3.1.15}$$

Alors Y_1 et Y_2 sont des variables aléatoires indépendantes, de loi normale centrée réduite. Pour le voir, on vérifie d'abord que R a la fonction de répartition $1-\mathrm{e}^{-r^2/2}$, donc la densité $r\,\mathrm{e}^{-r^2/2}$. Le couple (R,Φ) a donc la densité jointe $r\,\mathrm{e}^{-r^2/2}/(2\pi)$, et la formule de changement de variable montre que le couple (Y_1,Y_2) a la densité jointe $\mathrm{e}^{-(y_1^2+y_2^2)/2}/(2\pi)$, qui est bien celle d'un couple de variables normales centrées réduites indépendantes.

Bien entendu, les espérances de variables aléatoires de loi unidimensionnelle sont soit connues explicitement, soit calculables numériquement par la simple estimation d'une intégrale. Nous nous intéressons ici à des situations où la loi de Y n'est pas aussi simple à représenter.

3.2 Algorithmes MCMC

Considérons le cas d'un espace probabilisé discret $(\mathcal{X},\mathcal{P}(\mathcal{X}),\mu)$, où $\mathcal{X}$ est un ensemble dénombrable, mais très grand. Par exemple, dans le cas du modèle d'Ising (cf. Exemple 1.1.5), $\mathcal{X}=\{-1,1\}^N$ est de cardinal 2^N, et on s'intéresse à des N d'ordre 1000 au moins. La mesure de probabilité μ est dans ce cas une application de $\mathcal{X}$ vers $[0,1]$ telle que la somme des $\mu(i)$ vaut 1.

On voudrait estimer l'espérance d'une variable aléatoire $Y:\mathcal{X}\to\mathbb{R}$, comme par exemple l'aimantation dans le cas du modèle d'Ising :

$$\mathbb{E}(Y)=\sum_{i\in\mathcal{X}} Y(i)\mu(i)\ . \tag{3.2.1}$$

La méthode de Monte Carlo usuelle consiste alors à générer une suite de variables aléatoires $X_0,X_1,\dots$ sur $\mathcal{X}$, indépendantes et de distribution μ, puis de calculer la moyenne des $Y(X_j)$.

Pour générer ces X_j, on peut envisager de procéder comme suit : on définit un ordre total sur $\mathcal{X}$, et on détermine la fonction de répartition

$$F_\mu(j)=\sum_{k=1}^{j}\mu(k)\ . \tag{3.2.2}$$

Si U est une variable de loi uniforme sur $[0,1]$, alors $F_\mu^{-1}(U)$ suit la loi μ. Toutefois, en procédant de cette manière, on ne gagne rien, car le calcul des sommes (3.2.2) est aussi long que le calcul de la somme (3.2.1), que l'on voulait précisément éviter!

Les méthodes MCMC (pour *Monte Carlo Markov Chain*) évitent cet inconvénient. L'idée est de simuler *en même temps* la loi μ et la variable aléatoire Y, à l'aide d'une chaîne de Markov sur $\mathcal{X}$, de distribution invariante μ.

Soit donc $\{X_n\}_{n\in\mathbb{N}}$ une telle chaîne, supposée irréductible, et de distribution initiale ν arbitraire. On lui associe une suite $Y_n=Y(X_n)$ de variables aléatoires. Celles-ci peuvent se décomposer comme suit :

$$Y_n=\sum_{i\in\mathcal{X}} Y(i)1_{\{X_n=i\}}\ . \tag{3.2.3}$$

Considérons les moyennes

$$S_n = \frac{1}{n}\sum_{m=0}^{n-1} Y_m \ . \tag{3.2.4}$$

Le Théorème 1.4.13 permet d'écrire

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}_\nu\big(S_n\big) &= \lim_{n\to\infty} \frac{1}{n}\mathbb{E}_\nu\Big(\sum_{m=0}^{n-1} Y_m\Big) \\ &= \sum_{i\in\mathcal{X}} Y(i) \lim_{n\to\infty} \frac{1}{n}\mathbb{E}_\nu\Big(\sum_{m=0}^{n-1} 1_{\{X_m=i\}}\Big) \\ &= \sum_{i\in\mathcal{X}} Y(i)\mu(i) \\ &= \mathbb{E}(Y) \ . \end{aligned} \tag{3.2.5}$$

L'espérance de S_n converge bien vers l'espérance cherchée. Pour pouvoir appliquer l'idée de la méthode Monte Carlo, il nous faut plus, à savoir la convergence (au moins) en probabilité de S_n vers $\mathbb{E}(Y)$. On ne peut pas invoquer directement la loi des grands nombres, ni le théorème de la limite centrale, car les Y_n ne sont plus indépendants. Mais il s'avère que des résultats analogues restent vrais dans le cas de chaînes de Markov.

Théorème 3.2.1. *Supposons la chaîne réversible, et de distribution initiale égale à sa distribution stationnaire. Soit λ_0 la plus grande (en module) valeur propre différente de* 1 *de la matrice de transition de la chaîne. Alors*

$$\operatorname{Var} S_n \leqslant \frac{1}{n}\left(\frac{1+|\lambda_0|}{1-|\lambda_0|}\right)\operatorname{Var} Y \ . \tag{3.2.6}$$

Démonstration. Comme la chaîne démarre dans la distribution stationnaire μ, tous les Y_i ont même loi μY^{-1}, même s'ils ne sont pas indépendants. Il suit que

$$\begin{aligned}\operatorname{Var} S_n &= \frac{1}{n^2}\Big[\sum_{m=0}^{n-1} \operatorname{Var} Y_m + 2\sum_{0\leqslant p<q<n} \operatorname{cov}(Y_p, Y_q)\Big] \\ &= \frac{1}{n}\operatorname{Var} Y + \frac{2}{n^2}\sum_{m=1}^{n-1}(n-m)\operatorname{cov}(Y_0, Y_m) \ , \end{aligned} \tag{3.2.7}$$

en vertu de la propriété des incréments indépendants. Or si $\mathbf{1} = (1, 1, \ldots, 1)^T$ on a

$$\begin{aligned}\operatorname{cov}(Y_0, Y_m) &= \mathbb{E}_\mu\Big((Y_0 - \mathbb{E}_\mu(Y_0))(Y_m - \mathbb{E}_\mu(Y_m))\Big) \\ &= \sum_{i\in\mathcal{X}}\sum_{j\in\mathcal{X}} (Y(i)-\mathbb{E}(Y))(Y(j)-\mathbb{E}(Y))\underbrace{\mathbb{P}_\mu\{X_0=i, X_m=j\}}_{=\mu(i)(P^m)_{ij}} \\ &= \sum_{i\in\mathcal{X}} \mu_i(Y(i)-\mathbb{E}(Y))[P^m(Y-\mathbb{E}(Y)\mathbf{1})]_i \\ &= \langle Y-\mathbb{E}(Y)\mathbf{1} | P^m(Y-\mathbb{E}(Y)\mathbf{1})\rangle_\mu \\ &\leqslant |\lambda_0|^m \langle Y-\mathbb{E}(Y)\mathbf{1} | Y-\mathbb{E}(Y)\mathbf{1}\rangle_\mu = |\lambda_0|^m \operatorname{Var} Y \ . \end{aligned} \tag{3.2.8}$$

Dans l'inégalité, nous avons utilisé le fait que $Y - \mathbb{E}(Y)\mathbf{1} \in \mathbf{1}_\perp$ puisque la somme des $\mu_i(Y(i) - \mathbb{E}(Y))$ est nulle, et que par conséquent ce vecteur se trouve dans le sous-espace complémentaire au vecteur propre $\mathbf{1}$. Le résultat suit alors en remplaçant dans (3.2.7), en majorant $n - m$ par n et en sommant une série géométrique. □

Il suit de cette estimation et de l'inégalité de Bienaymé–Chebychev que pour calculer $\mathbb{E}(Y)$ avec une précision δ et avec probabilité $1-\varepsilon$, il faut choisir

$$n \geqslant \frac{\operatorname{Var} Y}{\delta^2 \varepsilon} \left(\frac{1+|\lambda_0|}{1-|\lambda_0|} \right) . \tag{3.2.9}$$

En pratique, on ne peut pas faire démarrer la chaîne exactement avec la distribution invariante. Ceci conduit à une convergence un peu plus lente, mais du même ordre de grandeur puisque la loi des Y_n converge exponentiellement vite vers μY^{-1}. Les résultats sont bien sûr meilleurs si on choisit bien la condition initiale, c'est-à-dire de manière à ce que la loi des Y_n converge rapidement.

3.3 L'algorithme de Metropolis

Nous avons vu comment estimer l'espérance d'une variable aléatoire Y à l'aide d'une chaîne de Markov de distribution invariante donnée par la loi de Y. Pour que cet algorithme soit efficace, il faut encore que l'on puisse trouver facilement, en fonction de cette loi, une matrice de transition donnant la distribution invariante souhaitée.

Une manière simple de résoudre ce problème est de chercher une chaîne de Markov réversible. Nous allons illustrer cette méthode dans le cas du modèle d'Ising, mais on voit facilement comment la généraliser à d'autres systèmes.

Dans le cas du modèle d'Ising (Exemple 1.1.5), l'univers est donné par $\mathcal{X} = \{-1, 1\}^\Lambda$, où Λ est un sous-ensemble (supposé ici fini) de $\mathbb{Z}^d$. La mesure de probabilité sur $\mathcal{X}$ est définie par

$$\mu(\sigma) = \frac{\mathrm{e}^{-\beta H(\sigma)}}{Z_\beta} , \qquad \text{où } Z_\beta = \sum_{\sigma \in \mathcal{X}} \mathrm{e}^{-\beta H(\sigma)} . \tag{3.3.1}$$

Le paramètre β désigne la température inverse, et la fonction $H : \mathcal{X} \to \mathbb{R}$, associant à toute configuration son énergie, est donnée par

$$H(\sigma) = - \sum_{i,j \in \Lambda \colon \|i-j\|=1} \sigma_i \sigma_j - h \sum_{i \in \Lambda} \sigma_i , \tag{3.3.2}$$

où h est le champ magnétique. L'objectif est de calculer l'espérance de la variable aimantation, donnée par

$$m(\sigma) = \frac{1}{|\Lambda|} \sum_{i \in \Lambda} \sigma_i \tag{3.3.3}$$

(nous avons introduit un facteur $1/|\Lambda|$ pour assurer que m prenne des valeurs entre -1 et 1). Afin de satisfaire la condition de réversibilité, on cherche une matrice de transition P sur $\mathcal{X}$ dont les éléments satisfont

$$\mu(\sigma) p_{\sigma\sigma'} = \mu(\sigma') p_{\sigma'\sigma} \tag{3.3.4}$$

pour toute paire $(\sigma, \sigma') \in \mathcal{X} \times \mathcal{X}$. Cela revient à imposer que

$$\frac{p_{\sigma\sigma'}}{p_{\sigma'\sigma}} = \mathrm{e}^{-\beta \Delta H(\sigma,\sigma')} , \tag{3.3.5}$$

où $\Delta H(\sigma, \sigma') = H(\sigma') - H(\sigma)$. On notera que cette condition ne fait pas intervenir la constante de normalisation Z_β, ce qui est souhaitable, car le calcul de cette constante est aussi coûteux que celui de $\mathbb{E}(m)$.

L'*algorithme de Métropolis* consiste dans un premier temps à définir un ensemble de transitions permises, c'est-à-dire une relation symétrique $\sim$ sur $\mathcal{X}$ (on supposera toujours que $\sigma \not\sim \sigma$). Les plus courantes sont

- la *dynamique de Glauber*, qui consiste à choisir $\sigma \sim \sigma'$ si et seulement si les deux configurations σ et σ' diffèrent en exactement une composante; on parle de dynamique de renversement de spin;
- la *dynamique de Kawasaki*, qui consiste à choisir $\sigma \sim \sigma'$ si et seulement si σ' est obtenue en intervertissant deux composantes de σ; on parle de dynamique d'échange de spin. Dans ce cas, la chaîne n'est pas irréductible sur $\mathcal{X}$, car elle conserve le nombre total de spins $+1$ et -1 : elle est en fait irréductible sur chaque sous-ensemble de configurations à nombre fixé de spins de chaque signe.

Une fois la relation $\sim$ fixée, on choisit des probabilités de transition telles que

$$p_{\sigma\sigma'} = \begin{cases} p_{\sigma'\sigma}\, \mathrm{e}^{-\beta \Delta H(\sigma,\sigma')} & \text{si } \sigma \sim \sigma' \ , \\ 1 - \displaystyle\sum_{\sigma'' \sim \sigma} p_{\sigma\sigma''} & \text{si } \sigma = \sigma' \ , \\ 0 & \text{autrement .} \end{cases} \tag{3.3.6}$$

Pour satisfaire la condition lorsque $\sigma \sim \sigma'$, une possibilité est de prendre

$$p_{\sigma\sigma'} = \begin{cases} q_{\sigma\sigma'} & \text{si } H(\sigma') \leqslant H(\sigma) \ , \\ q_{\sigma\sigma'}\, \mathrm{e}^{-\beta \Delta H(\sigma,\sigma')} & \text{si } H(\sigma') > H(\sigma) \ , \end{cases} \tag{3.3.7}$$

où $q_{\sigma\sigma'} = q_{\sigma'\sigma}$ et $\sum_{\sigma' \sim \sigma} q_{\sigma\sigma'} = 1$. On peut par exemple choisir les $q_{\sigma\sigma'}$ constants, égaux à l'inverse du nombre de transitions permises. Cela revient à toujours effectuer la transition si elle décroît l'énergie, et de ne l'effectuer qu'avec probabilité $\mathrm{e}^{-\beta \Delta H(\sigma,\sigma')}$ si elle fait croître l'énergie. Une autre possibilité est de choisir

$$p_{\sigma\sigma'} = \frac{q_{\sigma\sigma'}}{1 + \mathrm{e}^{\beta \Delta H(\sigma,\sigma')}} \ . \tag{3.3.8}$$

Remarquons que le calcul de la différence d'énergie ΔH est particulièrement simple dans le cas de la dynamique de Glauber, car seuls le spin que l'on renverse et ses voisins entrent en compte. Ainsi, si $\sigma^{(k)}$ dénote la configuration obtenue en renversant le spin numéro k de σ, on aura

$$\Delta H(\sigma, \sigma^{(k)}) = 2\sigma_k \left[\sum_{j \colon \|j-k\|=1} \sigma_j + h \right] , \tag{3.3.9}$$

qui est une somme de $2d+1$ termes pour un réseau $\Lambda \subset \mathbb{Z}^d$.

Concrètement, l'algorithme de Metropolis avec dynamique de Glauber s'implémente comme suit (avec $N = |\Lambda|$) :

1. **Etape d'initialisation :**
   - choisir une configuration initiale $\sigma(0)$ (si possible telle que $\delta_{\sigma(0)}$ soit proche de μ);
   - calculer $m_0 = m(\sigma(0))$ (nécessite N calculs);
   - calculer $H(\sigma(0))$ (nécessite de l'ordre de dN calculs);
   - poser $S = m_0$.
2. **Etape d'itération :** Pour $n = 1, 2, \ldots, n_{\max}$,
   - choisir un spin k au hasard uniformément dans Λ;

- calculer $\Delta H(\sigma(n-1),\sigma')$, où $\sigma' = \sigma(n-1)^{(k)}$;
- si $\Delta H(\sigma(n-1),\sigma') \leqslant 0$, poser $\sigma(n) = \sigma'$;
- si $\Delta H(\sigma(n-1),\sigma') > 0$, poser $\sigma(n) = \sigma'$ avec probabilité $q_{\sigma\sigma'}\,\mathrm{e}^{-\beta\Delta H(\sigma(n-1),\sigma')}$, sinon prendre $\sigma(n+1) = \sigma(n)$;
- si on a renversé le spin k, $m_n = m_{n-1} + 2\sigma_k(n-1)/N$, sinon $m_n = m_{n-1}$;
- ajouter m_n à S.

Le quotient $S/(n+1)$ converge alors vers $\mathbb{E}(m)$, avec une vitesse déterminée par (3.2.6). La seule quantité difficile à estimer est le trou spectral $1-|\lambda_0|$. Il s'avère que ce trou est grand sauf pour les températures très basses ou proches de la température critique de la transition de phase. Dans ce dernier cas, il existe des algorithmes alternatifs, tels que l'algorithme dit de Swendsen–Wang, qui convergent beaucoup mieux.

3.4 Le recuit simulé

Pour terminer, nous allons illustrer l'idée de l'algorithme du recuit simulé dans le cas du problème du voyageur de commerce (Exemple 1.1.6). Rappelons que pour N villes, on se donne une fonction $(i,j) \mapsto d(i,j)$ donnant la distance (ou le temps de voyage) entre les villes i et j. Le but est de trouver une permutation σ de $\{1,2,\dots,N\}$ qui minimise la longueur totale d'un circuit fermé,

$$H(\sigma) = \sum_{i=1}^{N-1} d(\sigma(i),\sigma(i+1)) + d(\sigma(N),\sigma(1)) \;. \tag{3.4.1}$$

La difficulté est que l'ensemble $\mathcal{S}_N$ des permutations possibles est de cardinal $N!$ (en fait $(N-1)!/2$ si l'on fixe la ville de départ et le sens de parcours), et croît donc encore plus vite qu'exponentiellement en N. Le tableau suivant donne quelques valeurs de temps de calcul, en supposant que la longueur d'un chemin se calcule en une microseconde.

N	Nombre de chemins	Temps de calcul
5	12	12 μs
10	181440	0.18 s
15	$43.6 \cdot 10^6$	12 heures
20	$60 \cdot 10^{15}$	1928 ans
25	$310 \cdot 10^{21}$	9.8 milliards d'années

D'où l'idée d'explorer $\mathcal{S}_N$ à l'aide d'une chaîne de Markov. On commence par se donner à nouveau une relation symétrique $\sim$ sur $\mathcal{S}_N$, par exemple $\sigma \sim \sigma'$ si les deux permutations diffèrent d'une transposition. Un simple algorithme de descente du gradient, consistant à choisir à chaque étape l'une des transitions permises au hasard, et à les accepter si elles décroissent H, ne donne pas de bons résultats, car la chaîne peut rester bloquée dans des minima locaux.

Si l'on choisit les probabilités de transition comme dans (3.3.7), on évite cet inconvénient. Toutefois, la chaîne ne converge alors pas vers le minimum de H, mais vers la mesure de Gibbs $\mathrm{e}^{-\beta H(\sigma)}/Z_\beta$. Lorsque β tend vers $+\infty$, cette mesure converge effectivement vers une mesure concentrée sur le ou les minima globaux de H. Mais si l'on choisit d'entrée un β trop élevé (c'est-à-dire une température trop faible), la chaîne de Markov risque d'être piégée très longtemps au voisinage de minima locaux.

L'algorithme du recuit simulé consiste à choisir une chaîne de Markov inhomogène dans le temps. Cela veut dire qu'au lieu de choisir un β fixe dans (3.3.7), on prend un β_n dépendant du temps, typiquement $\beta_n = \beta_0 K^n$ avec $K > 1$. Si K est suffisamment proche de 1, c'est-à-dire si l'on refroidit le système suffisamment lentement, la chaîne converge vers une bonne approximation du minimum global de H — c'est du moins ce que l'on observe numériquement dans beaucoup de cas.

Partie II

Processus de sauts et files d'attente

Chapitre 4

Rappels de probabilités

4.1 Loi binomiale et loi de Poisson

Nous commençons par rappeler quelques lois de probabilité usuelles qui joueront un rôle important dans la suite. Une *expérience de Bernoulli* de longueur n et probabilité de succès $q \in [0,1]$ consiste à répéter n fois, de manière indépendante, une expérience élémentaire qui n'admet que deux issues possibles : Le succès, qui se produit avec probabilité q, et l'échec, qui se produit avec probabilité $1-q$. Si par exemple l'expérience consiste à jeter un dé équilibré, et que l'on considère comme succès uniquement l'obtention de 6 points, on aura $q = \frac{1}{6}$.

Soit X la variable aléatoire donnant le nombre de succès de l'expérience de longueur n. Elle pourra prendre les valeurs $0, 1, 2, \ldots, n$, la valeur k étant obtenue avec probabilité

$$\mathbb{P}\{X = k\} = \binom{n}{k} q^k (1-q)^{n-k} =: b_{n,q}(k) \ , \tag{4.1.1}$$

où nous avons noté les coefficients binomiaux

$$\binom{n}{k} \equiv C_n^k = \frac{n!}{k!(n-k)!} \ . \tag{4.1.2}$$

En effet, il y a $\binom{n}{k}$ manières d'arranger les k succès et $n-k$ échecs parmi les n expériences, et chacun de ces arrangements se produit avec probabilité $q^k(1-q)^{n-k}$.

Définition 4.1.1 (Loi binomiale)**.** *Soit X une variable aléatoire prenant ses valeurs dans $\{0, 1, \ldots, n\}$ satisfaisant* (4.1.1)*. Nous dirons que X suit une* loi binomiale de paramètres (n, q) *et noterons $X \sim b_{n,q}$.*

On peut représenter X comme somme de n variables aléatoires Y_i indépendantes et identiquement distribuées (i.i.d.), de loi de Bernoulli de paramètre q, c'est-à-dire telles que $\mathbb{P}\{Y_i = 1\} = q = 1 - \mathbb{P}\{Y_i = 0\}$. Alors comme l'espérance de chaque Y_i vaut $\mathbb{E}(Y_i) = 0 \cdot \mathbb{P}\{Y_i = 0\} + 1 \cdot \mathbb{P}\{Y_i = 1\} = q$, on obtient pour l'espérance de X

$$\mathbb{E}(X) := \sum_{k=0}^{n} k \, \mathbb{P}\{X = k\} = \sum_{i=1}^{n} \mathbb{E}(Y_i) = nq \ . \tag{4.1.3}$$

De plus, comme la variance de chaque Y_i vaut $\mathrm{Var}(Y_i) := \mathbb{E}(Y_i^2) - \mathbb{E}(Y_i)^2 = q(1-q)$, on voit que la variance de X est donnée par

$$\mathrm{Var}(X) := \mathbb{E}(X^2) - \mathbb{E}(X)^2 = nq(1-q) \ . \tag{4.1.4}$$

Une deuxième loi importante dans ce cours est la loi de Poisson.

Définition 4.1.2 (Loi de Poisson)**.** *Nous dirons que la variable aléatoire X suit une* loi de Poisson de paramètre $\lambda > 0$*, et noterons $X \sim \pi_\lambda$, si elle prend des valeurs entières non-négatives, avec probabilité*

$$\mathbb{P}\{X = k\} = \mathrm{e}^{-\lambda}\,\frac{\lambda^k}{k!} =: \pi_\lambda(k)\;. \tag{4.1.5}$$

Avant de discuter sa signification, mentionnons quelques propriétés de base de cette loi.

Proposition 4.1.3.

1. *Si X suit une loi de Poisson de paramètre λ, alors*

$$\mathbb{E}(X) = \mathrm{Var}(X) = \lambda\;. \tag{4.1.6}$$

2. *Si X et Y sont indépendantes, et suivent des lois de Poisson de paramètres λ et μ respectivement, alors $X + Y$ suit une loi de Poisson de paramètre $\lambda + \mu$.*

Démonstration. Voir exercices 4.2 et 4.3. □

L'importance de la loi de Poisson π_λ vient du fait qu'elle donne une bonne approximation de la loi binomiale $b_{n,q}$ lorsque la longueur n de l'expérience est grande et que la probabilité de succès q est faible, avec $nq = \lambda$. En effet nous avons le résultat de convergence suivant :

Proposition 4.1.4. *Soit $\{q_n\}_{n\geqslant 0}$ une suite telle que $\lim_{n\to\infty} nq_n = \lambda > 0$. Alors, pour tout $k \in \mathbb{N}$,*

$$\lim_{n\to\infty} b_{n,q_n}(k) = \pi_\lambda(k)\;. \tag{4.1.7}$$

Démonstration. Soit $\lambda_n = nq_n$. Alors

$$\begin{aligned} b_{n,q_n}(k) &= b_{n,\lambda_n/n}(k) \\ &= \frac{n(n-1)\dots(n-k+1)}{k!}\,\frac{\lambda_n^k}{n^k}\left(1-\frac{\lambda_n}{n}\right)^{n-k} \\ &= \frac{\lambda_n^k}{k!}\,\frac{1}{(1-\lambda_n/n)^k}\left(1-\frac{1}{n}\right)\dots\left(1-\frac{k-1}{n}\right)\left(1-\frac{\lambda_n}{n}\right)^{n}\;. \end{aligned} \tag{4.1.8}$$

Lorsque $n \to \infty$, on a $\lambda_n \to \lambda$, $\lambda_n/n \to 0$ et $j/n \to 0$ pour $j = 0, \dots, k-1$, donc

$$\lim_{n\to\infty} b(k; n, q_n) = \frac{\lambda^k}{k!}\lim_{n\to\infty}\left(1-\frac{\lambda_n}{n}\right)^n = \frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}\;, \tag{4.1.9}$$

la dernière égalité pouvant se montrer par un développement limité de $\log[(1-\lambda_n/n)^n]$. □

Exemple 4.1.5. La probabilité de gagner au tiercé est de 1/1000. Quelle est la probabilité de gagner k fois en jouant 2000 fois?

Nous modélisons la situation par une expérience de Bernoulli de longueur $n = 2000$ et paramètre $q = 0.001$. La probabilité de gagner k fois sera donnée par

$$b_{2000,0.001}(k) \simeq \pi_2(k) = \mathrm{e}^{-2}\,\frac{2^k}{k!}\;. \tag{4.1.10}$$

Le tableau suivant compare quelques valeurs des deux lois.

k	0	1	2	3	4	5
$b_{2000,0.001}(k)$	0.13520	0.27067	0.27081	0.18053	0.09022	0.03605
$\pi_2(k)$	0.13534	0.27067	0.27067	0.18045	0.09022	0.03609

L'avantage de la loi de Poisson est qu'elle est bien plus rapide à calculer, puisqu'elle ne fait pas intervenir de coefficients binomiaux.

Nous donnons maintenant un résultat bien plus fort, à savoir la convergence dans L^1 de la loi de Bernoulli vers la loi de Poisson. Ce résultat est entre autres intéressant pour sa preuve, qui peut être faite exclusivement avec des arguments de probabilités, et un minimum d'analyse.

Théorème 4.1.6. *On a*

$$\sum_{k=0}^{\infty} \bigl| b_{n,q}(k) - \pi_{nq}(k) \bigr| \leqslant 2nq^2 \; . \tag{4.1.11}$$

Démonstration. Nous commençons par introduire des espaces probabilisés (Ω_i, p_i), pour $i = 1, \dots, n$, donnés par $\Omega_i = \{-1, 0, 1, 2, \dots\}$ et

$$p_i(k) = \begin{cases} \mathrm{e}^{-q} - (1-q) & \text{si } k = -1 \; , \\ 1 - q & \text{si } k = 0 \; , \\ \mathrm{e}^{-q} \, \dfrac{q^k}{k!} & \text{si } k \geqslant 1 \; . \end{cases} \tag{4.1.12}$$

On vérifiera que les p_i définissent bien une distribution de probabilité. Sur chaque Ω_i, nous introduisons les deux variables aléatoires

$$X_i(\omega_i) = \begin{cases} 0 & \text{si } \omega_i = 0 \; , \\ 1 & \text{sinon} \; , \end{cases} \qquad\qquad Y_i(\omega_i) = \begin{cases} \omega_i & \text{si } \omega_i \geqslant 1 \; , \\ 0 & \text{sinon} \; . \end{cases} \tag{4.1.13}$$

De cette manière, on a $\mathbb{P}\{X_i = 0\} = 1 - q$, $\mathbb{P}\{X_i = 1\} = q$, et $\mathbb{P}\{Y_i = k\} = \pi_q(k)$ pour tout $k \geqslant 0$. De plus,

$$\begin{aligned} \mathbb{P}\{X_i = Y_i\} &= \mathbb{P}\{X_i = 0, Y_i = 0\} + \mathbb{P}\{X_i = 1, Y_i = 1\} \\ &= p_i(0) + p_i(1) = 1 - q + q\,\mathrm{e}^{-q} \; , \end{aligned} \tag{4.1.14}$$

donc

$$\mathbb{P}\{X_i \neq Y_i\} = q(1 - \mathrm{e}^{-q}) \leqslant q^2 \; . \tag{4.1.15}$$

Soit (Ω, p) l'espace produit des (Ω_i, p_i). Alors

- $X = X_1 + \cdots + X_n$ suit la loi binomiale $\mathbb{P}\{X = k\} = b_{n,q}(k)$;
- $Y = Y_1 + \cdots + Y_n$ suit la loi de Poisson $\mathbb{P}\{Y = k\} = \pi_{nq}(k)$, en vertu de la Proposition 4.1.3.

Comme $X \neq Y$ implique que $X_i \neq Y_i$ pour un i au moins, il suit de (4.1.15) que

$$\mathbb{P}\{X \neq Y\} \leqslant \sum_{i=1}^{n} \mathbb{P}\{X_i \neq Y_i\} \leqslant nq^2 \; . \tag{4.1.16}$$

Il reste donc à montrer que la le membre de gauche de (4.1.11) est majoré par $2\mathbb{P}\{X \neq Y\}$. Un tel procédé s'appelle un argument de couplage. Nous posons, pour abréger l'écriture, $f(k) = \mathbb{P}\{X = k\}$, $g(k) = \mathbb{P}\{Y = k\}$ et $A = \{k\colon f(k) > g(k)\}$. Alors

$$\begin{aligned}\sum_{k=0}^{\infty}\bigl|b_{n,q}(k) - \pi_{nq}(k)\bigr| &= \sum_{k=0}^{\infty}\bigl|f(k) - g(k)\bigr| \\ &= \sum_{k\in A}\bigl(f(k) - g(k)\bigr) - \sum_{k\notin A}(f(k) - g(k)) \\ &= 2\sum_{k\in A}\bigl(f(k) - g(k)\bigr) - \underbrace{\sum_{k\in\mathbb{N}}\bigl(f(k) - g(k)\bigr)}_{=1-1=0} . \end{aligned} \tag{4.1.17}$$

Or nous pouvons écrire

$$\begin{aligned}\sum_{k\in A}\bigl(f(k) - g(k)\bigr) &= \mathbb{P}\{X \in A\} - \mathbb{P}\{Y \in A\} \\ &= \mathbb{P}\{X \in A, Y \in A\} + \mathbb{P}\{X \in A, Y \neq X\} - \mathbb{P}\{Y \in A\} \\ &\leqslant \mathbb{P}\{X \in A, Y \in A\} + \mathbb{P}\{X \neq Y\} - \mathbb{P}\{Y \in A\} \\ &\leqslant \mathbb{P}\{X \neq Y\} , \end{aligned} \tag{4.1.18}$$

ce qui conclut la démonstration. □

Si nous prenons par exemple $q = \lambda/n$, la borne (4.1.11) nous fournit

$$\sum_{k=0}^{\infty}\bigl|b_{n,\lambda/n}(k) - \pi_{\lambda}(k)\bigr| \leqslant 2\frac{\lambda^2}{n} . \tag{4.1.19}$$

Dans l'exemple 4.1.5, λ^2/n vaut $4/2000 = 0.002$. Ainsi la somme de toutes les valeurs absolues de différences $b_{2000,0.001}(k) - \pi_2(k)$ est bornée par 0.004, et comme tous ces termes sont positifs, la plupart seront bien plus petits encore.

4.2 Loi normale et loi exponentielle

Nous aurons également affaire à des variables aléatoires réelles continues. Pour les définir, le plus simple est de passer par la notion de *fonction de répartition*.

Définition 4.2.1 (Fonction de répartition)**.** *Une fonction $F : \mathbb{R} \to [0, 1]$ est une* fonction de répartition *si*

- *F est croissante: $x \leqslant y \Rightarrow F(x) \leqslant F(y)$.*
- *F est continue à droite: $\lim_{y\to x+} F(y) = F(x)\ \forall x$.*
- *$\lim_{x\to-\infty} F(x) = 0$ et $\lim_{x\to+\infty} F(x) = 1$.*

Une fonction de répartition F est dite absolument continue de densité *f si*

$$F(x) = \int_{-\infty}^{x} f(y)\ \mathrm{d}y . \tag{4.2.1}$$

Le lien entre la notion de fonction de répartition et les variables aléatoires vient du fait que pour toute variable aléatoire réelle, $\mathbb{P}\{X \leqslant t\}$ est une fonction de répartition.

En effet,

- si $s \leqslant t$, alors $\{X \leqslant s\} \subset \{X \leqslant t\}$, et donc $\mathbb{P}\{X \leqslant s\} \leqslant \mathbb{P}\{X \leqslant t\}$;
- $\lim_{s \to t+} \mathbb{P}\{X \leqslant s\} - \mathbb{P}\{X \leqslant t\} = \lim_{s \to t+} \mathbb{P}\{t < X \leqslant s\} = 0$;
- $\lim_{t \to -\infty} \mathbb{P}\{X \leqslant t\} = 0$ et $\lim_{t \to +\infty} \mathbb{P}\{X \leqslant t\} = 1$.

Ceci motive la définition suivante.

Définition 4.2.2 (Variable aléatoire à densité)**.** *Si X est une variable aléatoire,*

$$F_X(t) = \mathbb{P}\{X \leqslant t\} \tag{4.2.2}$$

est appelée fonction de répartition de X. *Si F_X est absolument continue de densité f, on dit que X* admet la densité *f et on a les relations*

$$\mathbb{P}\{X \leqslant t\} = \int_{-\infty}^{t} f(s)\, \mathrm{d}s\ , \tag{4.2.3}$$

$$\mathbb{P}\{a < X \leqslant b\} = \mathbb{P}\{X \leqslant b\} - \mathbb{P}\{X \leqslant a\} = \int_{a}^{b} f(s)\, \mathrm{d}s\ . \tag{4.2.4}$$

Dans ce cas, on peut remplacer $<$ par $\leqslant$ et inversement.

L'espérance d'une variable aléatoire réelle X de densité f est définie par l'intégrale

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x)\, \mathrm{d}x\ , \tag{4.2.5}$$

(pourvu que la fonction $xf(x)$ soit absolument intégrable, sinon on dit que X n'admet pas d'espérance).

Toutes les variables aléatoires réelles ne sont par des variables à densité : Par exemple, les fonctions de répartition de variables à valeurs discrètes, comme celles vues dans la section précédente, sont constantes par morceaux et admettent des discontinuités, et ne peuvent donc par être écrites comme l'intégrale d'une fonction continue.

Un premier exemple important de variable aléatoire réelle à densité est celui des variables gaussiennes ou normales.

Définition 4.2.3 (Loi normale)**.** *On dit que la variable aléatoire X suit une* loi normale de moyenne μ et écart-type σ, *et on notera $X \sim \mathcal{N}(\mu, \sigma^2)$ si elle admet la densité*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, \mathrm{e}^{-(x-\mu)^2/2\sigma^2}\ . \tag{4.2.6}$$

Si $X \sim \mathcal{N}(0,1)$, on dit qu'elle suit une loi normale centrée réduite, ou standard.

On vérifie que si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors son espérance vaut $\mathbb{E}(X) = \mu$ et sa variance vaut $\mathrm{Var}(X) = \sigma^2$. L'importance de la loi normale vient avant tout du théorème de la limite centrale :

Théorème 4.2.4 (Théorème de la limite centrale)**.** *Soit $X_1, X_2, \ldots$ une suite de variables aléatoires indépendantes, identiquement distribuées (abrégé* i.i.d.*), d'espérance finie μ et de variance finie σ^2. Alors la variable aléatoire $S_n = \sum_{i=1}^{n} X_i$ satisfait*

$$\lim_{n \to \infty} \mathbb{P}\left\{ a \leqslant \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leqslant b \right\} = \int_a^b \frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}}\, \mathrm{d}x\ , \tag{4.2.7}$$

c'est-à-dire que $(S_n - n\mu)/\sqrt{n\sigma^2}$ converge en loi vers une variable normale standard.

Un second exemple de loi à densité, particulièrement important dans ce cours, est celui des variables exponentielles.

Définition 4.2.5 (Variable exponentielle)**.** *On dit que la variable aléatoire X suit une* loi exponentielle de paramètre $\lambda > 0$*, et on note $X \sim \mathcal{E}xp(\lambda)$, si elle satisfait*

$$\mathbb{P}\{X > t\} = \mathrm{e}^{-\lambda t} \tag{4.2.8}$$

pour tout $t \geqslant 0$. Sa fonction de répartition est donc $F_X(t) = 1 - \mathrm{e}^{-\lambda t}$ pour $t > 0$, et sa densité est $\lambda\,\mathrm{e}^{-\lambda t}$, toujours pour $t > 0$.

On vérifie qu'une variable de loi exponentielle a espérance $1/\lambda$ et variance $1/\lambda^2$. Une propriété remarquable de la loi exponentielle est la *propriété de Markov* : Pour $t > s \geqslant 0$,

$$\mathbb{P}\{X > t | X > s\} = \mathrm{e}^{-\lambda(t-s)} = \mathbb{P}\{X > t - s\}\ . \tag{4.2.9}$$

Nous aurons parfois affaire à des couples, ou des n-uplets de variables aléatoires à densité, aussi appelés *vecteurs aléatoires*. Leur *densité conjointe* est définie comme la fonction de n variables f telle que

$$\mathbb{P}\{X_1 \leqslant t_1, X_2 \leqslant t_2, \dots, X_n \leqslant t_n\} = \int_{-\infty}^{t_1}\int_{-\infty}^{t_2}\cdots\int_{-\infty}^{t_n} f(x_1, x_2, \dots, x_n)\ \mathrm{d}x_n \dots \mathrm{d}x_2\,\mathrm{d}x_1 \tag{4.2.10}$$

pour tout choix de $(t_1, t_2, \dots, t_n)$. Autrement dit, on a

$$f(t_1, t_2, \dots, t_n) = \frac{\partial^n}{\partial t_1 \partial t_2 \dots \partial t_n}\mathbb{P}\{X_1 \leqslant t_1, X_2 \leqslant t_2, \dots, X_n \leqslant t_n\}\ . \tag{4.2.11}$$

Les variables aléatoires $X_1, X_2, \dots, X_n$ sont dites *indépendantes* si on a

$$\mathbb{P}\{X_1 \leqslant t_1, X_2 \leqslant t_2, \dots, X_n \leqslant t_n\} = \mathbb{P}\{X_1 \leqslant t_1\}\mathbb{P}\{X_2 \leqslant t_2\}\dots\mathbb{P}\{X_n \leqslant t_n\} \tag{4.2.12}$$

pour tout choix de $t_1, t_2, \dots, t_n$. On montre que c'est équivalent à ce que la densité conjointe s'écrive sous la forme

$$f(x_1, x_2, \dots, x_n) = f_1(x_1) f_2(x_2) \dots f_n(x_n) \tag{4.2.13}$$

pour des densités $f_1, f_2, \dots, f_n$ (appelées *densités marginales*).

Un autre résultat important est le suivant :

Proposition 4.2.6 (Convolution)**.** *Si X_1 et X_2 sont des variables aléatoires indépendantes, de densités respectives f_1 et f_2, alors $X_1 + X_2$ admet une densité donnée par la convolution*

$$(f_1 * f_2)(z) = \int_{-\infty}^{\infty} f_1(z - x_2) f_2(x_2)\ \mathrm{d}x_2\ . \tag{4.2.14}$$

Exemple 4.2.7 (Loi Gamma)**.** Soient $X_1, \dots, X_n$ des variables i.i.d. de loi $\mathcal{E}xp(\lambda)$. En calculant la convolution de leurs densités, on montre par récurrence sur n que la somme $S_n = X_1 + \cdots + X_n$ admet la densité

$$\gamma_{\lambda,n}(x) = \frac{\lambda^n}{(n-1)!} x^{n-1}\,\mathrm{e}^{-\lambda x} \tag{4.2.15}$$

pour $x \geqslant 0$. On dit que S_n suit une *loi Gamma* de paramètres (λ, n).

4.3 Exercices

Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. On appelle *fonction génératrice* de X la fonction $G_X : \mathbb{C} \to \mathbb{C}$ définie par

$$G_X(z) = \mathbb{E}\big(z^X\big) = \sum_{k=0}^{\infty} z^k \, \mathbb{P}\{X = k\} \; .$$

Exercice 4.1. Calculer les fonctions génératrices des distributions suivantes:

1. Loi de Bernoulli: $\mathbb{P}\{X = 0\} = 1 - q$, $\mathbb{P}\{X = 1\} = q$, où $q \in [0, 1]$.
2. Loi binomiale: $\mathbb{P}\{X = k\} = b_{n,q}(k) = \binom{n}{k} q^k (1-q)^{n-k}$, pour $k = 0, 1, \ldots, n$.
3. Loi de Poisson: $\mathbb{P}\{X = k\} = \pi_\lambda(k) = \mathrm{e}^{-\lambda} \lambda^k / k!$, où $\lambda > 0$ et $k \in \mathbb{N}$.
4. Loi géométrique: $\mathbb{P}\{X = k\} = q(1-q)^{k-1}$, où $q \in [0, 1]$ et $k \in \mathbb{N}^*$.

Exercice 4.2. On suppose que G_X existe pour tous les z dans un disque de rayon strictement supérieur à 1. Montrer que

$$G_X(1) = 1 \; , \qquad G'_X(1) = \mathbb{E}(X) \; , \qquad G''_X(1) = \mathbb{E}(X^2) - \mathbb{E}(X) \; ,$$

et en déduire une expression de la variance de X en termes de sa fonction génératrice. En déduire les espérances et variances de variables aléatoires de Bernoulli, binomiale, de Poisson et géométrique.

Exercice 4.3. Soient X et Y deux variables aléatoires indépendantes à valeurs dans $\mathbb{N}$, et de fonctions génératrices G_X et G_Y respectivement. Montrer que $G_{X+Y} = G_X G_Y$.

Application: Vérifier les assertions suivantes.

1. La somme de n variables aléatoires de Bernoulli indépendantes suit une loi binomiale;
2. La somme de deux variables aléatoires binomiales indépendantes suit une loi binomiale;
3. La somme de deux variables aléatoires indépendantes de loi de Poisson suit une loi de Poisson.

Exercice 4.4. Soit N une variable aléatoire à valeurs dans $\mathbb{N}$, et soit $G_N(z) = \mathbb{E}(z^N)$ sa fonction génératrice. Soient $X_1, X_2, \ldots$ des variables aléatoires à valeurs dans $\mathbb{N}$, indépendantes et identiquement distribuées, et indépendantes de N. Soit $G_X(z) = \mathbb{E}(z^X)$ leur fonction génératrice.

1. Soit $n \in \mathbb{N}^*$ et
$$S_n = X_1 + \cdots + X_n \; .$$
Ecrire la fonction génératrice $\mathbb{E}(z^{S_n})$ de S_n en fonction de $G_X(z)$.
2. Soit
$$S_N = X_1 + \cdots + X_N \; .$$
Montrer que sa fonction génératrice $G_S(z) = \mathbb{E}(z^{S_N})$ est donnée par
$$G_S(z) = G_N(G_X(z)) \; .$$
Indication: Ecrire $\mathbb{P}\{S_N = k\}$ en fonction des $\mathbb{P}\{S_N = k | N = n\}$, $n \in \mathbb{N}^*$.
3. On suppose que N suit une loi de Poisson de paramètre $\lambda > 0$, et que les X_i suivent des lois de Bernoulli de paramètre $q \in [0, 1]$. Déterminer la loi de S_N.

Exercice 4.5. Soit U une variable aléatoire uniforme sur $[0,1]$, c'est-à-dire de densité donnée par la fonction indicatrice $1_{[0,1]}$.

1. Donner la fonction de répartition de U.
2. Soit $\varphi : [0,1] \to \mathbb{R}$ une fonction strictement croissante et admettant une réciproque φ^{-1}. Déterminer la fonction de répartition de la variable aléatoire $Y = \varphi(U)$.
3. Déterminer φ de telle manière que Y soit exponentielle de paramètre λ.

Chapitre 5

Le processus ponctuel de Poisson

Le processus ponctuel de Poisson est un processus stochastique qui associe une distribution de probabilité aux configurations de points sur $\mathbb{R}_+$. Ces points modélisent, par exemple, les temps de passage d'un bus à un arrêt, les instants d'arrivée d'appels téléphoniques dans une centrale, et ainsi de suite.

Dans certains cas, par exemple lorsque les bus suivent un horaire régulier et qu'il n'y a pas de perturbation du trafic, les instants d'arrivée sont assez régulièrement espacés. Dans d'autres situations, par exemple lorsque des travaux perturbent le trafic, les instants d'arrivée deviennent beaucoup plus irréguliers, et on observe à la fois des longs temps d'attente et des bus qui se suivent presque immédiatement. Le processus ponctuel de Poisson modélise la situation la plus aléatoire possible, dans un sens qui reste à définir.

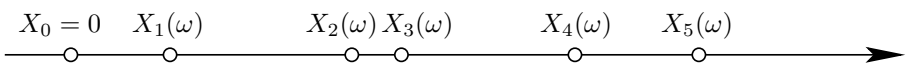


FIGURE 5.1. Une réalisation d'un processus de Poisson.

Le processus peut être caractérisé de plusieurs manières différentes. Une réalisation peut être spécifiée par une suite croissante de nombres réels positifs

$$X_0 = 0 < X_1(\omega) < X_2(\omega) < X_3(\omega) < \dots \,, \tag{5.0.1}$$

désignant les positions des points dans $\mathbb{R}_+$. Alternativement, on peut décrire une réalisation en donnant le nombre de points $N_I(\omega)$ contenus dans chaque intervalle I de la forme $I =]t, t+s]$. Si nous abrégeons $N_{]0,t]}$ par N_t (communément appelé *fonction de comptage*), nous aurons $N_{]t,t+s]} = N_{t+s} - N_t$, et les N_t sont donnés en fonction des X_n par

$$N_t(\omega) = \sup\{n \geqslant 0 \colon X_n(\omega) \leqslant t\} \,. \tag{5.0.2}$$

Inversement, les X_n se déduisent des N_t par la relation

$$X_n(\omega) = \inf\{t \geqslant 0 \colon N_t(\omega) \geqslant n\} \,. \tag{5.0.3}$$

Nous allons voir deux constructions équivalentes du processus de Poisson. La première construction part de la distribution des N_t.

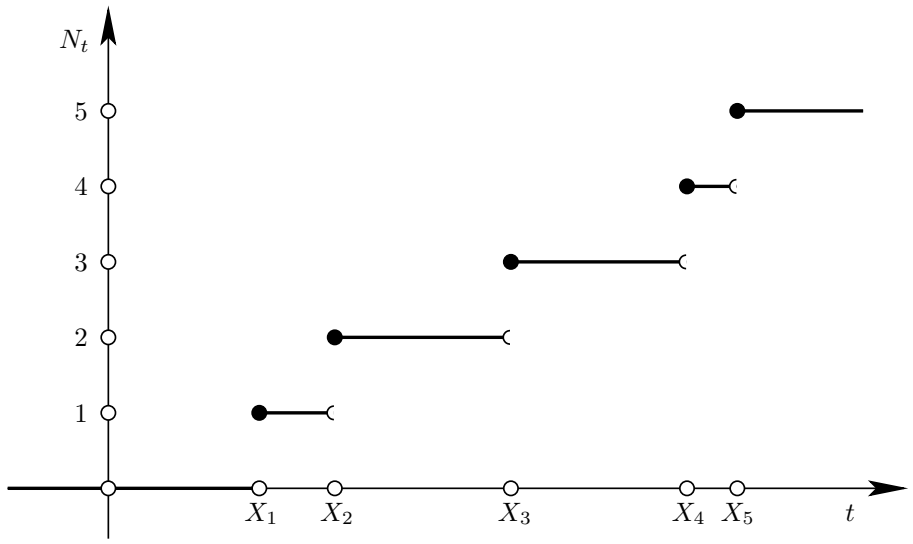


FIGURE 5.2. Fonction de comptage d'une réalisation d'un processus de Poisson.

5.1 Construction par la fonction de comptage

Définition 5.1.1 (Processus de Poisson)**.** *Le processus ponctuel de Poisson satisfait les conditions suivantes:*

1. *N_I ne dépend que de la longueur de I, i.e. $N_{]t,t+s]}$ a la même loi que N_s.*
2. *Si $I_1, \ldots, I_k$ sont deux à deux disjoints, $N_{I_1}, \ldots, N_{I_k}$ sont indépendants.*
3. *$\mathbb{E}(N_I)$ existe pour tout I (de longueur finie).*
4. *Il existe un intervalle I tel que $\mathbb{P}\{N_I > 0\} > 0$.*
5. *Absence de points doubles:* $\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \mathbb{P}\{N_\varepsilon \geqslant 2\} = 0$.

Supposant qu'un tel processus existe bel et bien, nous pouvons dériver quelques-unes de ses propriétés.

Proposition 5.1.2.

1. *Soit $\alpha(t) = \mathbb{E}(N_t)$. Alors il existe $\lambda > 0$ tel que $\alpha(t) = \lambda t$.*
2. *Pour tout intervalle borné $I \subset \mathbb{R}_+$, on a $\mathbb{P}\{N_I \geqslant 1\} \leqslant \mathbb{E}(N_I)$.*

DÉMONSTRATION.

1. Comme $N_0 = 0$, on a $\alpha(0) = 0$. De plus, comme $N_{t+s} = N_t + N_{]t,t+s]}$ on a

$$\alpha(t+s) = \alpha(t) + \mathbb{E}(N_{]t,t+s]}) = \alpha(t) + \alpha(s) \,, \tag{5.1.1}$$

   en vertu de la condition 1. Par un résultat d'analyse, ceci implique nécessairement que $\alpha(t) = \lambda t$ pour un $\lambda \geqslant 0$, et la propriété 4. implique $\lambda > 0$.
2. Comme $\mathbb{E}(N_I)$ est supposé fini, on a

$$\mathbb{E}(N_I) = \sum_{k=0}^{\infty} k\mathbb{P}\{N_I = k\} \geqslant \sum_{k=1}^{\infty} \mathbb{P}\{N_I = k\} = \mathbb{P}\{N_I \geqslant 1\} \,, \tag{5.1.2}$$

   ce qui montre à la fois que $\mathbb{P}\{N_I \geqslant 1\}$ est finie et qu'elle est borné par $\mathbb{E}(N_I)$. □

La propriété remarquable du processus de Poisson est alors que les variables aléatoires $N_{]t,t+s]}$ suivent nécessairement une loi de Poisson.

Théorème 5.1.3. *Si le processus satisfait les 5 conditions de la Définition 5.1.1, alors les variables aléatoires $N_{]t,t+s]}$ suivent des lois de Poisson de paramètre λs:*

$$\mathbb{P}\{N_{]t,t+s]} = k\} = \pi_{\lambda s}(k) = \mathrm{e}^{-\lambda s}\,\frac{(\lambda s)^k}{k!}\ . \tag{5.1.3}$$

Démonstration. Par la propriété 1., il suffit de montrer le résultat pour $t = 0$, c'est-à-dire pour N_s. Partageons $]0, s]$ en k intervalles de longueur égale, de la forme

$$]s_{j-1}, s_j] \qquad \text{où } s_j = \frac{js}{k} \text{ pour } 0 \leqslant j \leqslant k\ . \tag{5.1.4}$$

L'idée de la preuve est que pour k suffisamment grand, il est peu probable d'avoir plus d'un point par intervalle, donc la loi de $Y_j^{(k)} = N_{]s_{j-1},s_j]}$ est à peu près une loi de Bernoulli. La loi de N_s est donc proche d'une loi binomiale, que l'on peut approximer par la loi de Poisson pour k grand.

Il suit des conditions 1. et 2. que les $Y_j^{(k)}$ sont i.i.d., de même loi que $N_{s_1} = N_{s/k}$, et on a

$$N_s = \sum_{j=1}^{k} Y_j^{(k)}\ . \tag{5.1.5}$$

Introduisons alors des variables aléatoires

$$\overline{Y}_j^{(k)} = \begin{cases} 0 & \text{si } Y_j^{(k)} = 0\ , \\ 1 & \text{si } Y_j^{(k)} \geqslant 1\ . \end{cases} \tag{5.1.6}$$

Les $\overline{Y}_j^{(k)}$ sont également i.i.d., et suivent une loi de Bernoulli. La variable aléatoire

$$\overline{N}_s^{(k)} = \sum_{j=1}^{k} \overline{Y}_j^{(k)}\ , \tag{5.1.7}$$

satisfaisant $\overline{N}_s^{(k)} \leqslant N_s$ pour tout k, on a

$$\mathbb{P}\{\overline{N}_s^{(k)} \geqslant m\} \leqslant \mathbb{P}\{N_s \geqslant m\} \tag{5.1.8}$$

pour tout k et tout m. De plus, $\overline{N}_s^{(k)}$ suit une loi binomiale de paramètre

$$p_k = \mathbb{P}\{\overline{Y}_j^{(k)} = 1\} = \mathbb{P}\{Y_j^{(k)} \geqslant 1\} = \mathbb{P}\{N_{s/k} \geqslant 1\}\ . \tag{5.1.9}$$

Estimons maintenant la différence entre les lois de $\overline{N}_s^{(k)}$ et N_s. Nous avons

$$\begin{aligned} \mathbb{P}\{\overline{N}_s^{(k)} \neq N_s\} &= \mathbb{P}\{\exists j \in \{1, \dots, k\} : Y_j^{(k)} \geqslant 2\} \\ &\leqslant \sum_{j=1}^{k} \mathbb{P}\{Y_j^{(k)} \geqslant 2\} \\ &= k\mathbb{P}\{Y_1^{(k)} \geqslant 2\} = k\mathbb{P}\{N_{s/k} \geqslant 2\}\ . \end{aligned} \tag{5.1.10}$$

La condition 5. avec $\varepsilon = s/k$ implique alors

$$\lim_{k\to\infty} \mathbb{P}\{\overline{N}_s^{(k)} \neq N_s\} = 0 \; . \tag{5.1.11}$$

Comme on a d'une part la minoration

$$\mathbb{P}\{N_s = m\} \geqslant \mathbb{P}\{N_s = \overline{N}_s^{(k)} = m\} \geqslant \mathbb{P}\{\overline{N}_s^{(k)} = m\} - \mathbb{P}\{\overline{N}_s^{(k)} \neq N_s\} \; , \tag{5.1.12}$$

et d'autre part la majoration

$$\begin{aligned} \mathbb{P}\{N_s = m\} &= \mathbb{P}\{\overline{N}_s^{(k)} = N_s = m\} + \mathbb{P}\{\overline{N}_s^{(k)} \neq N_s = m\} \\ &\leqslant \mathbb{P}\{\overline{N}_s^{(k)} = m\} + \mathbb{P}\{N_s \neq \overline{N}_s^{(k)}\} \; , \end{aligned} \tag{5.1.13}$$

il suit que

$$\lim_{k\to\infty} \mathbb{P}\{\overline{N}_s^{(k)} = m\} = \mathbb{P}\{N_s = m\} \; . \tag{5.1.14}$$

Il reste à montrer que kp_k tend vers λs pour $k \to \infty$. Si c'est le cas, alors la Proposition 4.1.4 montre que N_s suit une loi de Poisson de paramètre λs. Or nous avons

$$\begin{aligned} kp_k = \mathbb{E}(\overline{N}_s^{(k)}) &= \sum_{j=1}^{\infty} j\mathbb{P}\{\overline{N}_s^{(k)} = j\} = \sum_{j=1}^{\infty}\sum_{\ell=1}^{j} \mathbb{P}\{\overline{N}_s^{(k)} = j\} \\ &= \sum_{\ell=1}^{\infty}\sum_{j=\ell}^{\infty} \mathbb{P}\{\overline{N}_s^{(k)} = j\} = \sum_{\ell=1}^{\infty} \mathbb{P}\{\overline{N}_s^{(k)} \geqslant \ell\} \; . \end{aligned} \tag{5.1.15}$$

Un calcul analogue montre que

$$\lambda s = \mathbb{E}(N_s) = \sum_{\ell=1}^{\infty} \mathbb{P}\{N_s \geqslant \ell\} \; . \tag{5.1.16}$$

Il suit de (5.1.8) que $kp_k \leqslant \lambda s$ pour tout k. Par un théorème d'analyse, on peut alors intervertir limite et somme, et écrire

$$\lim_{k\to\infty} \sum_{\ell=1}^{\infty} \mathbb{P}\{\overline{N}_s^{(k)} \geqslant \ell\} = \sum_{\ell=1}^{\infty} \lim_{k\to\infty} \mathbb{P}\{\overline{N}_s^{(k)} \geqslant \ell\} = \sum_{\ell=1}^{\infty} \mathbb{P}\{N_s \geqslant \ell\} = \lambda s \; , \tag{5.1.17}$$

en vertu de (5.1.14). Ceci montre que kp_k converge bien vers λs, et par conséquent que la loi de N_s, étant la limite d'une loi binomiale de paramètre kp_k, est une loi de Poisson de paramètre λs. □

5.2 Construction par les temps d'attente

La seconde construction du processus ponctuel de Poisson se base sur la distribution des différences de position $Z_n = X_n - X_{n-1}$. Celles-ci caractérisent également de manière univoque le processus, via la relation

$$X_n(\omega) = \sum_{j=1}^{n} Z_j(\omega) \; . \tag{5.2.1}$$

Le résultat remarquable est alors que les Z_j sont i.i.d. et suivent une loi bien particulière, à savoir une loi exponentielle de paramètre λ.

Théorème 5.2.1. *Pour tout n, les variables aléatoires $Z_1, \dots, Z_n$ sont indépendantes, et suivent la même loi exponentielle $\mathcal{E}xp(\lambda)$.*

Démonstration. Fixons des instants

$$t_0 = 0 < s_1 < t_1 < s_2 < t_2 < \cdots < s_n < t_n \ . \tag{5.2.2}$$

Nous pouvons alors calculer

$$\begin{aligned}
&\mathbb{P}\big\{X_1 \in]s_1, t_1], X_2 \in]s_2, t_2], \dots, X_n \in]s_n, t_n]\big\} \\
&\quad = \mathbb{P}\big\{N_{]0,s_1]} = 0, N_{]s_1,t_1]} = 1, N_{]t_1,s_2]} = 0, \dots, N_{]t_{n-1},s_n]} = 0, N_{]s_n,t_n]} \geqslant 1\big\} \\
&\quad = \prod_{k=1}^{n} \mathbb{P}\big\{N_{]t_{k-1},s_k]} = 0\big\} \prod_{k=1}^{n-1} \mathbb{P}\big\{N_{]s_k,t_k]} = 1\big\}\, \mathbb{P}\big\{N_{]s_n,t_n]} \geqslant 1\big\} \\
&\quad = \prod_{k=1}^{n} \mathrm{e}^{-\lambda(s_k - t_{k-1})} \prod_{k=1}^{n-1} \lambda(t_k - s_k)\, \mathrm{e}^{-\lambda(t_k - s_k)} \left[1 - \mathrm{e}^{-\lambda(t_n - s_n)}\right] \\
&\quad = \lambda^{n-1} \prod_{k=1}^{n-1} (t_k - s_k) \big[\mathrm{e}^{-\lambda s_n} - \mathrm{e}^{-\lambda t_n}\big] \\
&\quad = \int_{s_1}^{t_1} \int_{s_2}^{t_2} \dots \int_{s_n}^{t_n} \lambda^n\, \mathrm{e}^{-\lambda x_n}\, \mathrm{d}x_n\, \mathrm{d}x_{n-1} \dots \mathrm{d}x_2\, \mathrm{d}x_1 \ .
\end{aligned} \tag{5.2.3}$$

La loi conjointe de $(X_1, \dots, X_n)$ admet donc la densité

$$f(x_1, \dots, x_n) = \begin{cases} \lambda^n\, \mathrm{e}^{-\lambda x_n} & \text{si } 0 < x_1 < \cdots < x_n \ , \\ 0 & \text{sinon .} \end{cases} \tag{5.2.4}$$

Nous pouvons alors calculer la fonction de répartion des Z_k:

$$\begin{aligned}
\mathbb{P}\big\{Z_1 \leqslant z_1, \dots, Z_n \leqslant z_n\big\} &= \mathbb{P}\big\{X_1 \leqslant z_1, X_2 - X_1 \leqslant z_2, \dots, X_n - X_{n-1} \leqslant z_n\big\} \\
&= \mathbb{P}\big\{X_1 \leqslant z_1, X_2 \leqslant z_2 + X_1, \dots, X_n \leqslant z_n + X_{n-1}\big\} \\
&= \int_0^{z_1} \int_{x_1}^{z_2 + x_1} \dots \int_{x_{n-1}}^{z_n + x_{n-1}} f(x_1, \dots, x_n)\, \mathrm{d}x_n \dots \mathrm{d}x_1 \ .
\end{aligned} \tag{5.2.5}$$

La densité conjointe de $(Z_1, \dots, Z_n)$ s'obtient alors en calculant la dérivée

$$\begin{aligned}
\frac{\partial^n}{\partial z_1 \dots \partial z_n} \mathbb{P}\big\{Z_1 \leqslant z_1, \dots, Z_n \leqslant z_n\big\} &= f(z_1, z_1 + z_2, \dots, z_1 + \cdots + z_n) \\
&= \lambda^n\, \mathrm{e}^{-\lambda(z_1 + \cdots + z_n)} \ ,
\end{aligned} \tag{5.2.6}$$

Or cette densité est bien la densité conjointe de n variables exponentielles indépendantes de paramètre λ. □

Ce résultat fournit une méthode permettant de construire le processus de Poisson: chaque X_n est obtenu à partir de X_{n-1} en lui ajoutant une variable aléatoire de loi exponentielle, indépendante des variables précédentes. On notera que X_n suit une loi Gamma de paramètres (λ, n).

Notons que la propriété de Markov (4.2.9) peut s'écrire

$$\mathbb{P}\big\{Z_n > t + \varepsilon \;\big|\; Z_n > t\big\} = \mathrm{e}^{-\lambda\varepsilon} \qquad \forall \varepsilon > 0 \ , \tag{5.2.7}$$

et donc

$$\lim_{\varepsilon\to 0} \frac{\mathbb{P}\big\{ Z_n \leqslant t+\varepsilon \;\big|\; Z_n > t \big\}}{\varepsilon} = \lim_{\varepsilon\to 0} \frac{1-\mathrm{e}^{-\lambda\varepsilon}}{\varepsilon} = \lambda\,. \tag{5.2.8}$$

Le paramètre λ représente donc le taux d'apparition d'un nouveau point, qui est indépendant du passé. Ce résultat peut bien sûr également être obtenu directement à l'aide de la fonction de comptage.

5.3 Généralisations

Il existe plusieurs généralisations du processus ponctuel de Poisson discuté ici:

1. Le *processus de Poisson inhomogène* : Dans ce cas le nombre de points $N_{]t,t+s]}$ suit une loi de Poisson de paramètre

$$\int_t^{t+s} \lambda(u)\ \mathrm{d}u\,, \tag{5.3.1}$$

où $\lambda(u)$ est une fonction positive, donnant le taux au temps u. Ce processus permet de décrire des situations où les points apparaissent avec une intensité variable, par exemple si l'on veut tenir compte des variations journalières du trafic influençant les horaires de passage de bus. On retrouve le processus de Poisson homogène si $\lambda(u)$ est constant.

2. Le *processus de Poisson de dimension* $n \geqslant 2$: Ce processus peut être défini par sa fonction de comptage, en remplaçant les intervalles I par des sous-ensembles (mesurables) de $\mathbb{R}^n$. Les nombres de points dans deux ensembles disjoints sont à nouveau indépendants, et le nombre de points dans un ensemble est proportionnel à son volume. Ce processus peut par exemple modéliser la distribution des étoiles dans une région de l'espace ou du ciel.
3. Le *processus de naissance et de mort* : Le processus ponctuel de Poisson peut être considéré comme un processus de naissance pur : Si N_t est interprété comme le nombre d'individus dans une population au temps t, ce nombre augmente avec un taux constant λ. Plus généralement, dans un processus de naissance et de mort, de nouveaux individus naissent avec un taux λ et meurent avec un taux μ; éventuellement, $\lambda = \lambda(N_t)$ et $\mu = \mu(N_t)$ peuvent dépendre de la taille actuelle de la population.
4. Le *processus de Poisson composé* : Soit N_t la fonction de comptage d'un processus de Poisson simple, et soient $Y_1, Y_2, \ldots$, des variables aléatoires indépendantes et identiquement distribuées (i.i.d.) et indépendantes de N_t. Alors le processus

$$S_t = \sum_{i=1}^{N_t} Y_i \tag{5.3.2}$$

est appelé un processus de Poisson composé. A chaque instant X_n où le processus de Poisson sous-jacent augmente d'une unité, le processus composé S_t augmente d'une quantité aléatoire Y_n. Si N_t décrit les instants d'arrivée de clients dans une station service, et Y_n décrit la quantité d'essence achetée par le client n, alors S_t est la quantité d'essence totale vendue au temps t. Si tous les Y_i valent 1 presque sûrement, on retrouve le processus de Poisson de départ.
5. Le *processus de renouvellement* : Un tel processus est défini comme le processus de Poisson à partir de ses temps d'attente Z_n, sauf que les Z_n ne suivent pas forcément

une loi exponentielle. Il modélise par exemple les instants de remplacement de machines de durée de vie aléatoire (d'où le terme de "renouvellement" : chaque nouvelle machine étant supposée indépendante des précédentes, le processus oublie son passé à chaque instant de renouvellement). Ce processus n'a en général plus les propriétés d'indépendance de la fonction de comptage, mais sous certaines hypothèses sur la loi des Z_n (existence des deux premiers moments), il existe des versions asymtotiques de ces propriétés.

5.4 Exercices

Exercice 5.1. Des clients arrivent dans une banque suivant un processus de Poisson d'intensité λ. Sachant que deux clients sont arrivés dans la première heure, quelle est la probabilité que

1. les deux soient arrivés dans les 20 premières minutes?
2. l'un au moins soit arrivé dans les 20 premières minutes?

Exercice 5.2. Dans un serveur informatique, les requêtes arrivent selon un processus ponctuel de Poisson, avec un taux de 60 requêtes par heure. Déterminer les probabilités suivantes :

1. L'intervalle entre les deux premières requêtes est compris entre 2 et 4 minutes.
2. Aucune requête n'arrive entre 14h et 14h05.
3. Sachant que deux requêtes sont arrivées entre 14h et 14h10, les deux sont arrivées dans les 5 premières minutes.
4. Sachant que deux requêtes sont arrivées entre 14h et 14h10, au moins une est arrivée dans les 5 premières minutes.

Exercice 5.3. Un serveur informatique envoie des messages selon un processus ponctuel de Poisson. En moyenne, il envoie un message toutes les 30 secondes.

1. Quelle est la probabilité que le serveur n'envoie aucun message au cours des 2 premières minutes de sa mise en service?
2. A quel moment espérez-vous le second message (quel est le temps moyen de l'envoi du second message)?
3. Quelle est la probabilité que le serveur n'ait pas envoyé de message durant la première minute, sachant qu'il a envoyé 3 messages au cours des 3 premières minutes?
4. Quelle est la probabilité qu'il y ait moins de 3 messages au cours des 2 premières minutes, sachant qu'il y en a eu au moins 1 au cours de la première minute?

Exercice 5.4. L'écoulement des voitures le long d'une route est modélisé par un processus de Poisson d'intensité $\lambda = 2$ voitures par minute. A cause d'un chantier, le traffic est arrêté alternativement dans chaque direction. On admet qu'à l'arrêt, chaque véhicule occupe une longueur de 8 mètres en moyenne.

1. Quelle est la loi du temps d'arrivée X_n de la $n$ième voiture?
2. A l'aide du théorème de la limite centrale, donner une approximation gaussienne de la loi de X_n.
3. Pendant combien de temps peut-on arrêter le traffic si l'on désire que la queue ainsi formée ne dépasse la longueur de 250m qu'avec une probabilité de 0.2? (La valeur de x pour laquelle $\mathbb{P}\{\mathcal{N}(0,1) < x\} = 0.2$ est $x \simeq -0.85$).

Exercice 5.5 (Le paradoxe de l'autobus)**.** Les temps d'arrivée d'autobus à un arrêt sont décrits par un processus de Poisson $(X_n)_n$ d'intensité λ. Un client arrive à l'instant t après le début du service.

1. Calculer la probabilité que le client rate le $n$ième bus mais attrape le $(n+1)$ième bus.
2. Calculer la probabilité qu'il rate le $n$ième bus et doive attendre le $(n+1)$ième bus pendant un temps s au moins.
3. Calculer la probabilité qu'il doive attendre le bus suivant pendant un temps s au moins.
4. En déduire le temps d'attente moyen, et comparer ce temps avec le temps de passage moyen entre bus. Qu'en pensez-vous?

Exercice 5.6. Soient $(X_n)_n$ et $(Y_n)_n$ deux processus de Poisson indépendants, d'intensités respectives λ et μ. Soit $(Z_n)_n$ le processus obtenu en superposant $(X_n)_n$ et $(Y_n)_n$. Montrer qu'il s'agit encore d'un processus de Poisson et donner son intensité.

Exercice 5.7. Soit X_n un processus ponctuel de Poisson d'intensité λ.
Soit Y_n le processus obtenu en effaçant chaque X_n $(n \geqslant 1)$ avec probabilité $1/2$, indépendamment de tous les autres, puis en renumérotant les points restants par les entiers positifs. On note N_t, respectivement M_t, le nombre de X_n, respectivement Y_n, dans l'intervalle $]0, t]$.

1. Donner la loi de N_t.
2. Montrer que pour tout k, $l \mapsto \mathbb{P}\{M_t = l | N_t = k\}$ suit une loi binomiale, et déterminer ses paramètres.
3. En déduire la loi de M_t.
4. Montrer que Y_n est un processus ponctuel de Poisson, et déterminer son intensité.
5. Que se passe-t-il si on efface chaque X_n avec probabilité $1-q$, $0 < q < 1$?

Chapitre 6

Processus markoviens de sauts

Les processus markoviens de sauts sont un analogue en temps continu des chaînes de Markov. Rappelons que la propriété essentielle permettant de définir ces chaînes est la propriété de Markov :

$$\begin{aligned}\mathbb{P}\big\{X_n = j \;\big|\; X_{n-1} = i, X_{n-2} = i_{n-2}, \dots, X_0 = i_0\big\} &= \mathbb{P}\big\{X_n = j \;\big|\; X_{n-1} = i\big\} \\ &= p_{ij} \end{aligned} \tag{6.0.1}$$

pour tout temps n. Un processus markovien de sauts est défini par la condition

$$\begin{aligned}\mathbb{P}\big\{X_{s+t} = j \;\big|\; X_s = i, X_{s_{n-2}} = i_{n-2}, \dots, X_{s_0} = i_0\big\} &= \mathbb{P}\big\{X_{s+t} = j \;\big|\; X_s = i\big\} \\ &= P_t(i,j) \end{aligned} \tag{6.0.2}$$

pour tout choix de temps $0 \leqslant s_0 < s_1 < \cdots < s_{n-2} < s < s+t \in \mathbb{R}_+$. Pour chaque $t \geqslant 0$, l'ensemble des $P_t(i,j)$ doit former une matrice stochastique, appelée *noyau de transition*. Ses éléments sont appelés *probabilités de transition*.

Exemple 6.0.1 (Chaîne de Markov en temps continu)**.** Soit Y_n une chaîne de Markov de matrice de transition P, et soit N_t la fonction de comptage d'un processus ponctuel de Poisson de taux λ. Alors

$$X_t = Y_{N_t} \tag{6.0.3}$$

est un processus markovien de sauts, qu'on appelle une *chaîne de Markov en temps continu*. Ce processus effectue des transitions à des temps aléatoires, donnés par le processus de Poisson, au lieu de temps régulièrement espacés comme dans le cas de la chaîne de Markov de départ. La propriété (6.0.2) est une conséquence de deux propriétés de Markov : celle de la chaîne Y_n, et celle de la loi exponentielle, c.f. (4.2.9).

Comme le nombre de sauts N_t jusqu'au temps t suit une loi de Poisson, les probabilités de transition sont données par

$$P_t(i,j) = \sum_{n=0}^{\infty} \mathbb{P}\big\{N_t = n\big\}\mathbb{P}_i\big\{X_n = j\big\} = \sum_{n=0}^{\infty} \mathrm{e}^{-\lambda t}\,\frac{(\lambda t)^n}{n!} p_{ij}^{(n)}\,, \tag{6.0.4}$$

où les $p_{ij}^{(n)}$ sont les éléments de la matrice P^n, décrivant les probabilités de transition en n pas de la chaîne de Markov Y_n.

6.1 Taux de transition

Les noyaux de transition P_t pour différents t ne sont pas arbitraires, mais obéissent à une équation dite de Chapman–Kolmogorov (qui est l'analogue de la relation (1.2.13) pour les chaînes de Markov).

Proposition 6.1.1 (Equation de Chapman–Kolmogorov)**.** *La famille $\{P_t\}_{t\geqslant 0}$ des noyaux de transition d'un processus markovien de sauts sur un ensemble dénombrable $\mathcal{X}$ satisfait*

$$P_{s+t}(i,j) = \sum_{k\in\mathcal{X}} P_s(i,k)P_t(k,j) \tag{6.1.1}$$

pour tout choix de $s,t \geqslant 0$ et de $i,j \in \mathcal{X}$.

Démonstration. On a

$$\begin{aligned} P_{s+t}(i,j) &= \mathbb{P}\big\{X_{s+t} = j \;\big|\; X_0 = i\big\} \\ &= \sum_{k\in\mathcal{X}} \frac{\mathbb{P}\{X_{s+t}=j, X_s = k, X_0 = i\}}{\mathbb{P}\{X_0 = i\}} \\ &= \sum_{k\in\mathcal{X}} \frac{\mathbb{P}\{X_{s+t}=j, X_s = k, X_0 = i\}}{\mathbb{P}\{X_s = k, X_0 = i\}} \frac{\mathbb{P}\{X_s = k, X_0 = i\}}{\mathbb{P}\{X_0 = i\}} \\ &= \sum_{k\in\mathcal{X}} \underbrace{\mathbb{P}\big\{X_{s+t} = j \;\big|\; X_s = k, X_0 = i\big\}}_{=P_t(k,j)} \underbrace{\mathbb{P}\big\{X_s = k \;\big|\; X_0 = i\big\}}_{=P_s(i,k)}, \end{aligned} \tag{6.1.2}$$

en vertu de la propriété de Markov (6.0.2). □

L'équation de Chapman–Kolmogorov montre que si l'on connaît les noyaux P_t pour tous les temps t dans un intervalle $]0,t_0]$, aussi petit soit-il, alors on connaît les noyaux pour tous les temps $t \in \mathbb{R}_+$. Ceci suggère de considérer la limite

$$q(i,j) = \lim_{h\to 0} \frac{P_h(i,j)}{h}\,, \qquad i \neq j \in \mathcal{X}\,. \tag{6.1.3}$$

Si la limite existe, alors c'est par définition la dérivée à droite de $P_t(i,j)$ en $t = 0$. Dans ce cas, $q(i,j)$ est appelé le *taux de transition de l'état i vers l'état j*.

Exemple 6.1.2 (Chaîne de Markov en temps continu)**.** En effectuant un développement limité de l'expression (6.0.4) en $h = 0$, on obtient (comme $p_{ij}^{(0)} = 0$ pour $i \neq j$)

$$P_h(i,j) = (1 - \lambda h + \dots)(\lambda h p_{ij} + \dots) = \lambda h p_{ij} + \mathcal{O}(h)\,, \tag{6.1.4}$$

ce qui donne le taux de transition

$$q(i,j) = \lambda p_{ij}\,. \tag{6.1.5}$$

Exemple 6.1.3 (Processus de Poisson)**.** Soit N_t la fonction de comptage d'un processus ponctuel de Poisson de paramètre λ. On peut la considérer comme un processus markovien de sauts, dans lequel seules les transitions entre n et $n+1$ sont permises. Les probabilités de transition sont alors données par

$$P_s(n,n+1) = \mathbb{P}\big\{N_{t+s} = n+1 \;\big|\; N_t = n\big\} = \mathbb{P}\big\{N_{]t,t+s]} = 1\big\} = \lambda s\, \mathrm{e}^{-\lambda s}\,. \tag{6.1.6}$$

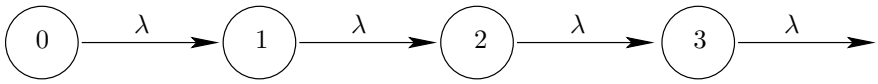


FIGURE 6.1. Graphe représentant le processus de Poisson, ou processus de naissance pur.

Les taux de transition correspondants sont donc

$$q(n, n+1) = \lim_{h \to 0} \frac{\lambda h \, \mathrm{e}^{-\lambda h}}{h} = \lambda \, . \tag{6.1.7}$$

La chaîne peut être représentée graphiquement comme dans la Figure 6.1.

En règle générale, on construit le processus de sauts directement à partir des taux de transition $q(i, j)$, au lieu de partir des noyaux de transition $P_t(i, j)$. Pour cela, on commence par définir

$$\lambda(i) = \sum_{j \neq i} q(i, j) \, , \tag{6.1.8}$$

qui donne le taux avec lequel le processus quitte l'état i. Supposons que $0 < \lambda(i) < \infty$. Dans ce cas

$$r(i, j) = \frac{q(i, j)}{\lambda(i)} \tag{6.1.9}$$

donne la probabilité que le processus choisisse l'état j, sachant qu'il quitte l'état i.

Soit $\{Y_n\}_{n \geqslant 0}$ une chaîne de Markov de matrice de transition $R = \{r(i, j)\}$. A partir de cette chaîne, on peut construire (et simuler) le processus de saut de taux $\{q(i, j)\}$ de la manière suivante :

- On se donne des variables i.i.d. $\tau_0, \tau_1, \ldots$ de loi exponentielle $\mathcal{E}xp(1)$, indépendantes des probabilités régissant la chaîne Y_n.
- Si le processus part de l'état $X_0 = Y_0$, il doit quitter cet état avec un taux $\lambda(Y_0)$, c'est-à-dire après un temps aléatoire t_1 de loi $\mathcal{E}xp(\lambda(Y_0))$. Il suffit pour cela de choisir $t_1 = \tau_0 / \lambda(Y_0)$.
- Au temps t_1, le processus saute dans l'état Y_1, dans lequel il doit rester pendant un temps aléatoire t_2 de loi $\mathcal{E}xp(\lambda(Y_1))$.

En continuant de cette manière, on est donc amenés à définir des temps de séjour

$$t_n = \frac{\tau_{n-1}}{\lambda(Y_{n-1})} \, , \tag{6.1.10}$$

et de poser

$$X_t = Y_n \qquad \text{pour } T_n \leqslant t < T_{n+1} \, , \tag{6.1.11}$$

où les instants des sauts sont donnés par

$$T_n = \sum_{i=1}^{n} t_i \, . \tag{6.1.12}$$

La seule différence entre le cas général et le cas particulier de la chaîne de Markov en temps continu (Exemple 6.0.1) est que les différents temps de séjour t_n n'ont pas nécessairement le même paramètre.

6.2 Générateur et équations de Kolmogorov

Nous montrons maintenant comment déterminer le noyau de transition P_t à partir des taux de transition $q(i,j)$. Un rôle important est joué par le générateur du processus.

Définition 6.2.1 (Générateur)**.** *Soit X_t un processus markovien de sauts de taux de transition $\{q(i,j)\}$. On appelle* générateur (infinitésimal) *du processus la matrice L d'éléments*

$$L(i,j) = \begin{cases} q(i,j) & \text{si } i \neq j \ , \\ -\lambda(i) & \text{si } i = j \ . \end{cases} \tag{6.2.1}$$

On remarquera que la définition (6.1.8) des $\lambda(i)$ implique

$$\sum_{j\in\mathcal{X}} L(i,j) = 0 \qquad \forall i \in \mathcal{X} \ . \tag{6.2.2}$$

Théorème 6.2.2 (Equations de Kolmogorov)**.** *Le noyau de transition satisfait l'*équation de Kolmogorov rétrograde

$$\frac{\mathrm{d}}{\mathrm{d}t} P_t = L P_t \ , \tag{6.2.3}$$

*ainsi que l'*équation de Kolmogorov progressive

$$\frac{\mathrm{d}}{\mathrm{d}t} P_t = P_t L \ . \tag{6.2.4}$$

Démonstration. Pour l'équation rétrograde, commençons par écrire, en utilisant l'équation de Chapman–Kolmogorov,

$$\begin{aligned} P_{t+h}(i,j) - P_t(i,j) &= \sum_{k\in\mathcal{X}} P_h(i,k) P_t(k,j) - P_t(i,j) \\ &= \sum_{k\neq i} P_h(i,k) P_t(k,j) + \big[P_h(i,i) - 1\big] P_t(i,j) \ . \end{aligned} \tag{6.2.5}$$

La dérivée s'obtient en divisant par h des deux côtés et en prenant la limite $h \to 0$. Pour le premier terme de droite, on a

$$\lim_{h\to 0} \frac{1}{h} \sum_{k\neq i} P_h(i,k) P_t(k,j) = \sum_{k\neq i} q(i,k) P_t(k,j) \ . \tag{6.2.6}$$

Pour traiter le second terme, on observe que la condition $\sum_k P_h(i,k) = 1$ implique

$$\lim_{h\to 0} \frac{1}{h} \Big(P_h(i,i) - 1 \Big) = \lim_{h\to 0} \frac{1}{h} \Big(- \sum_{k\neq i} P_h(i,k) \Big) = - \sum_{k\neq i} q(i,k) = -\lambda(i) \ . \tag{6.2.7}$$

En insérant (6.2.6) et (6.2.7) dans (6.2.5), on obtient

$$\frac{\mathrm{d}}{\mathrm{d}t} P_t(i,j) = \sum_{k\neq i} q(i,k) P_t(k,j) - \lambda(i) P_t(i,j) = \sum_{k\in\mathcal{X}} L(i,k) P_t(k,j) \ , \tag{6.2.8}$$

ce qui n'est autre que l'équation (6.2.3) écrite en composantes. La preuve de l'équation progessive est analogue, en utilisant la décomposition

$$\begin{aligned} P_{t+h}(i,j) - P_t(i,j) &= \sum_{k\in\mathcal{X}} P_t(i,k) P_h(k,j) - P_t(i,j) \\ &= \sum_{k\neq j} P_t(i,k) P_h(k,j) + \big[P_h(j,j) - 1\big] P_t(i,j) \end{aligned} \tag{6.2.9}$$

au lieu de (6.2.5). □

Les équations de Kolmogorov sont des systèmes d'équations linéaires couplées pour les éléments de matrice $P_t(i,j)$. Comme $P_0 = I$ est la matrice identité, la solution peut s'écrire sous la forme de la matrice exponentielle

$$P_t = \mathrm{e}^{Lt} = \sum_{n=0}^{\infty} \frac{(Lt)^n}{n!} \; . \tag{6.2.10}$$

Pour le vérifier, il suffit de dériver la série terme à terme (ce qui est justifié si la série converge absolument). Il est difficile de trouver une forme plus explicite de e^{Lt} en toute généralité, mais le calcul peut être fait dans certains cas particuliers.

Exemple 6.2.3 (Chaîne de Markov en temps continu)**.** En appliquant la définition du générateur à (6.1.5), et en utilisant le fait que $\sum_{j\neq i} q(i,j) = \lambda(1-p_{ii})$ puisque P est une matrice stochastique, on obtient

$$L = \lambda(P - I) \; . \tag{6.2.11}$$

En général, e^{AB} n'est pas égal à $\mathrm{e}^{A}\,\mathrm{e}^{B}$. Mais si $AB = BA$, on a effectivement $\mathrm{e}^{AB} = \mathrm{e}^{A}\,\mathrm{e}^{B}$. Comme $PI = IP$ et $\mathrm{e}^{-\lambda It} = \mathrm{e}^{-\lambda t}\, I$, on obtient

$$P_t = \mathrm{e}^{-\lambda t}\,\mathrm{e}^{\lambda tP} \; , \tag{6.2.12}$$

ce qui est équivalent à (6.0.4).

Exemple 6.2.4 (Processus de Poisson)**.** Pour le processus de Poisson, le générateur est donné par

$$L(i,j) = \begin{cases} -\lambda & \text{si } j = i \; , \\ \lambda & \text{si } j = i+1 \; , \\ 0 & \text{sinon} \; . \end{cases} \tag{6.2.13}$$

L'équation de Kolmogorov rétrograde s'écrit donc

$$\frac{\mathrm{d}}{\mathrm{d}t} P_t(i,j) = -\lambda P_t(i,j) + \lambda P_t(i+1,j) \; . \tag{6.2.14}$$

Ecrivons le générateur sous la forme $L = \lambda(R-I)$ où I est la matrice identité, et $(R)_{ij} = 1$ si $j = i+1$, 0 sinon. Comme $IR = RI$, on a

$$\mathrm{e}^{Lt} = \mathrm{e}^{-\lambda tI}\,\mathrm{e}^{\lambda tR} = \mathrm{e}^{-\lambda t} \sum_{n\geqslant 0} \frac{(\lambda t)^n}{n!} R^n \; , \tag{6.2.15}$$

puisque $\mathrm{e}^{-\lambda tI} = \mathrm{e}^{-\lambda t}\, I$. Les éléments de la matrice R^n valent 1 si $j - i = n$ et 0 sinon. Par conséquent, en prenant l'élément de matrice (i,j) de (6.2.15), on trouve

$$P_t(i,j) = \begin{cases} \mathrm{e}^{-\lambda t}\, \dfrac{(\lambda t)^{j-i}}{(j-i)!} & \text{si } j \geqslant i \; , \\ 0 & \text{sinon} \; . \end{cases} \tag{6.2.16}$$

On vérifie que c'est bien une solution du système (6.2.14). On aurait également pu obtenir ce résultat directement, en observant que $P_t(i,j)$ est la probabilité que le processus de Poisson admette $j-i$ points dans l'intervalle $[0,t[$, qui est donnée par la loi de Poisson.

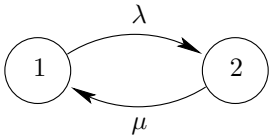


FIGURE 6.2. Graphe du processus de sauts à deux états.

Exemple 6.2.5 (Processus de sauts à deux états)**.** Soit $\mathcal{X} = \{1, 2\}$, avec des taux de transition $q(1, 2) = \lambda$ et $q(2, 1) = \mu$ (Figure 6.2). Le générateur a alors la forme

$$L = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix} . \tag{6.2.17}$$

Un calcul montre que $L^2 = -(\lambda + \mu)L$, donc $L^n = (-(\lambda + \mu))^{n-1}L$ pour tout $n \geqslant 1$. Il suit que

$$\begin{aligned} P_t = \mathrm{e}^{Lt} &= I + \sum_{n \geqslant 1} \frac{t^n}{n!}(-(\lambda + \mu))^{n-1}L \\ &= I - \frac{1}{\lambda + \mu}\big(\mathrm{e}^{-(\lambda+\mu)t} - 1\big)L \\ &= \frac{1}{\lambda + \mu}\begin{pmatrix} \mu + \lambda\,\mathrm{e}^{-(\lambda+\mu)t} & \lambda - \lambda\,\mathrm{e}^{-(\lambda+\mu)t} \\ \mu - \mu\,\mathrm{e}^{-(\lambda+\mu)t} & \lambda + \mu\,\mathrm{e}^{-(\lambda+\mu)t} \end{pmatrix} . \end{aligned} \tag{6.2.18}$$

On remarquera en particulier que si $\lambda + \mu > 0$, alors

$$\lim_{t\to\infty} P_t = \begin{pmatrix} \dfrac{\mu}{\lambda + \mu} & \dfrac{\lambda}{\lambda + \mu} \\ \dfrac{\mu}{\lambda + \mu} & \dfrac{\lambda}{\lambda + \mu} \end{pmatrix} . \tag{6.2.19}$$

Cela signifie qu'asymptotiquement, le système sera dans l'état 1 avec probabilité $\mu/(\lambda+\mu)$, et dans l'état 2 avec probabilité $\lambda/(\lambda + \mu)$, quel que soit l'état initial.

6.3 Distributions stationnaires

Définition 6.3.1 (Distribution stationnaire)**.** *Une distribution de probabilité π sur $\mathcal{X}$ est dite* stationnaire *pour le processus de noyau de transition P_t si*

$$\pi P_t = \pi \qquad \forall t > 0 , \tag{6.3.1}$$

c'est-à-dire, en composantes,

$$\sum_{i\in\mathcal{X}} \pi(i)P_t(i, j) = \pi(j) \qquad \forall j \in \mathcal{X},\ \forall t > 0 . \tag{6.3.2}$$

Comme en général, on n'a pas accès aux noyaux de transition, il est plus utile de disposer d'un critère faisant intervenir le générateur.

Théorème 6.3.2. *π est une distribution stationnaire si et seulement si $\pi L = 0$, c'est-à-dire*

$$\sum_{i\in\mathcal{X}} \pi(i)L(i, j) = 0 \qquad \forall j \in \mathcal{X} . \tag{6.3.3}$$

Démonstration.

$\Rightarrow$: Soit π une distribution stationnaire. L'équation de Kolmogorov progressive s'écrit en composantes

$$\frac{\mathrm{d}}{\mathrm{d}t}P_t(i,j) = \sum_{k\in\mathcal{X}} P_t(i,k)L(k,j) \; . \tag{6.3.4}$$

En multipliant par $\pi(i)$ et en sommant sur i, on obtient

$$\sum_{i\in\mathcal{X}} \pi(i)\frac{\mathrm{d}}{\mathrm{d}t}P_t(i,j) = \sum_{i,k\in\mathcal{X}} \pi(i)P_t(i,k)L(k,j) \; . \tag{6.3.5}$$

Le membre de gauche peut s'écrire

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum_{i\in\mathcal{X}} \pi(i)P_t(i,j) = \frac{\mathrm{d}}{\mathrm{d}t}\pi(j) = 0 \; . \tag{6.3.6}$$

En effectuant la somme sur i dans le membre de droite, on obtient

$$\sum_{i,k\in\mathcal{X}} \pi(i)P_t(i,k)L(k,j) = \sum_{k\in\mathcal{X}} \pi(k)L(k,j) \; . \tag{6.3.7}$$

L'égalité de (6.3.6) et (6.3.7) est bien équivalente à (6.3.2).

$\Leftarrow$: Supposons que $\pi L = 0$. Alors on a par l'équation de Kolmogorov rétrograde

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\sum_{i\in\mathcal{X}} \pi(i)P_t(i,j) &= \sum_{i\in\mathcal{X}} \pi(i) \sum_{k\in\mathcal{X}} L(i,k)P_t(k,j) \\ &= \sum_{k\in\mathcal{X}} \underbrace{\Bigl(\sum_{i\in\mathcal{X}} \pi(i)L(i,k)\Bigr)}_{=0} P_t(k,j) = 0 \; .\end{aligned} \tag{6.3.8}$$

Par conséquent, $\sum_i \pi(i)P_t(i,j)$ est constante, et donc égale à sa valeur en $t = 0$. Or cette valeur est $\pi(j)$ puisque $P_0(i,j) = \delta_{ij}$. Cela prouve que π est stationnaire. □

Exemple 6.3.3 (Chaîne de Markov en temps continu)**.** Comme $L = \lambda(P - I)$, l'équation $\pi L = 0$ est équivalente à $\pi P = \pi$. La distribution π est donc stationnaire pour la chaîne en temps continu $X_t = Y_{N_t}$ si et seulement si elle est stationnaire pour la chaîne en temps discret Y_n.

Exemple 6.3.4 (Processus de Poisson)**.** Le processus de Poisson n'admet pas de distribution stationnaire. En effet, nous avons vu que le générateur était donné par $L = \lambda(R - I)$, où $(R)_{ij} = 1$ si $j = i + 1$ et 0 sinon. La condition $\pi L = 0$ est équivalente à $R\pi = \pi$, donc à $\pi_{i+1} = \pi_i$ pour tout i. Or il n'existe pas de distribution de probabilité sur $\mathbb{N}$ dont tous les éléments seraient égaux. L'expression (6.2.14) des probabilités de transition $P_t(i,j)$ montre d'ailleurs que celles-ci tendent vers 0 lorsque $t \to \infty$. Cela est dû au fait que la distribution "s'en va vers l'infini".

Exemple 6.3.5 (Processus à deux états)**.** Dans le cas du processus à deux états, l'équation $\pi L = 0$ s'écrit

$$\pi(1)(-\lambda) + \pi(2)\mu = 0 \; . \tag{6.3.9}$$

Comme $\pi(1) + \pi(2) = 1$, on a nécessairement

$$\pi = \left(\frac{\mu}{\lambda + \mu}, \frac{\lambda}{\lambda + \mu} \right) . \tag{6.3.10}$$

Cela correspond bien à la valeur limite de P_t trouvée dans (6.2.19).

Dans le cas des chaînes de Markov en temps discret, nous avons vu que toute distribution initiale converge vers une unique distribution stationnaire, à condition que la chaîne soit irréductible, apériodique et récurrente positive. Dans le cas en temps continu, la condition d'apériodicité n'est plus nécessaire. Cela est dû au caractère aléatoire des intervalles de temps entre transitions.

On dira qu'un processus markovien de sauts est *irréductible* s'il est possible de trouver, pour toute paire d'états (i, j), un chemin $(i_0 = i, i_1, \dots, i_n = j)$ tel que $q(i_k, i_{k+1}) > 0$ pour tout $k = 0, \dots, n-1$.

Théorème 6.3.6. *Si le processus de sauts est irréductible et admet une distribution stationnaire π, alors*

$$\lim_{t \to \infty} P_t(i, j) = \pi(j) \qquad \forall i, j \in \mathcal{X} . \tag{6.3.11}$$

De plus, si $r : \mathcal{X} \to \mathbb{R}$ est une fonction telle que $\sum_i \pi(i)|r(i)| < \infty$, alors

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t r(X_s) \, \mathrm{d}s = \sum_i \pi(i) r(i) =: \mathbb{E}_\pi(r) . \tag{6.3.12}$$

Nous admettrons ce résultat. La relation (6.3.12) est un analogue du Théorème 1.4.13. En particulier, si $r(i) = \delta_{ij}$, elle montre que le temps moyen passé dans l'état j est égal à $\pi(j)$.

Nous avons vu que la distribution stationnaire d'une chaîne de Markov réversible était plus facile à déterminer, car il suffit de résoudre les conditions d'équilibre détaillé. De manière tout à fait analogue, on dira que le processus de sauts de taux de transition $\{q(i, j)\}$ est *réversible* s'il existe une distribution π telle que

$$\pi(i) q(i, j) = \pi(j) q(j, i) \qquad \forall i, j \in \mathcal{X} . \tag{6.3.13}$$

Théorème 6.3.7. *Si la condition d'équilibre détaillé* (6.3.13) *est satisfaite, alors π est une distribution stationnaire.*

Démonstration. Sommant la condition sur tous les $i \neq j$, on obtient

$$\sum_{i \neq j} \pi(i) q(i, j) = \pi(j) \sum_{i \neq j} q(j, i) = \pi(j) \lambda(j) = -\pi(j) L(j, j) . \tag{6.3.14}$$

Mais ceci est équivalent à $\sum_i \pi(i) L(i, j) = 0$. □

Exemple 6.3.8 (Processus de naissance et de mort)**.** Soit $\mathcal{X} = \{0, \dots, N\}$, et supposons que les seuls taux de transition non nuls sont

$$\begin{aligned} q(n, n+1) &= \lambda_n && \text{pour } 0 \leqslant n < N , \\ q(n, n-1) &= \mu_n && \text{pour } 0 < n \leqslant N . \end{aligned} \tag{6.3.15}$$

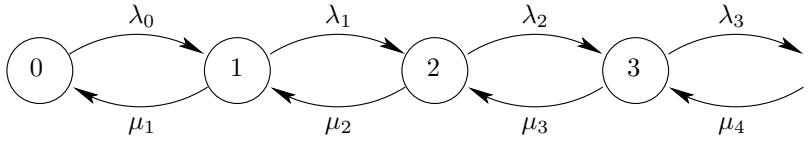


FIGURE 6.3. Graphe représentant le processus de naissance et de mort.

Ici X_t représente par exemple le nombre d'individus d'une population, qui naissent avec taux λ_n et meurent avec taux μ_n, pouvant dépendre du nombre n d'individus. Le graphe associé est représenté dans la Figure 6.3.

Supposons tous les taux strictement positifs. Dans ce cas, on peut satisfaire la condition d'équilibre détaillé en posant

$$\pi(n) = \frac{\lambda_{n-1}}{\mu_n}\pi(n-1) \ . \tag{6.3.16}$$

Par récurrence, on obtient

$$\pi(n) = \frac{\lambda_{n-1}\lambda_{n-2}\dots\lambda_0}{\mu_n\mu_{n-1}\dots\mu_1}\pi(0) \ , \tag{6.3.17}$$

et $\pi(0)$ est déterminé par la condition de normalisation $\sum_i \pi(i) = 1$.

6.4 Exercices

Exercice 6.1. Un homme d'affaires voyage entre Paris, Bordeaux et Marseille. Il passe dans chaque ville un temps de loi exponentielle, de moyenne 1/4 de mois pour Paris et Bordeaux, et de 1/5 de mois pour Marseille. S'il est à Paris, il va à Bordeaux ou Marseille avec probabilité 1/2. S'il est à Bordeaux, il va à Paris avec probabilité 3/4 et à Marseille avec probabilité 1/4. Après avoir visité Marseille, il retourne toujours à Paris.

1. Donner le générateur du processus markovien de sauts décrivant l'itinéraire de l'homme d'affaires.
2. Déterminer la fraction de temps qu'il passe dans chaque ville.
3. Combien de voyages fait-il en moyenne de Bordeaux à Paris par année?

Exercice 6.2. Un petit magasin d'informatique peut avoir au plus trois ordinateurs en stock. Des clients arrivent avec un taux de 2 clients par semaine. Si au moins un ordinateur est en stock, le client l'achète. S'il reste au plus un ordinateur, le tenancier du magasin commande deux nouveaux ordinateurs, qui sont livrés après un temps de loi exponentielle de moyenne 1 semaine.

1. Donner le générateur du processus décrivant le nombre d'ordinateurs en stock.
2. Déterminer la distribution stationnaire.
3. Quel est le taux de vente d'ordinateurs?

Exercice 6.3. Un cycliste peut être victime de deux types de pannes, qui se produisent selon un processus de Poisson :

- Il déraille en moyenne une fois toutes les 10 heures. Dans ce cas, il lui faut un temps exponentiel de moyenne 0.1 heure pour remettre la chaîne en place.
- Il crève une chambre à air en moyenne une fois toutes les 25 heures. Il lui faut alors un temps exponentiel de moyenne 1 heure pour réparer la panne.

1. Déterminer la distribution invariante.
2. Pendant quelle proportion du temps le cycliste peut-il rouler?
3. Généraliser la solution au cas de N différentes pannes, chaque panne se produisant avec un taux μ_i et nécessitant un temps de réparation exponentiel de moyenne $1/\lambda_i$.

Exercice 6.4. Une molécule d'hémoglobine peut fixer soit une molécule d'oxygène, soit une molécule de monoxyde de carbone. On suppose que ces molécules arrivent selon des processus de Poisson de taux λ_{O_2} et λ_{CO} et restent attachées pendant des temps exponentiels de taux μ_{O_2} et μ_{CO} respectivement. Déterminer la fraction de temps passée dans chacun des trois états : hémoglobine seule, hémoglobine et oxygène, hémoglobine et monoxyde de carbone.

Exercice 6.5. Des voitures arrivent dans une station service avec un taux de 20 voitures par heure. La station comporte une seule pompe. Si un conducteur trouve la pompe libre, il s'arrête pour faire le plein. L'opération lui prend un temps exponentiel de moyenne 6 minutes. Si la pompe est occupée mais qu'aucune voiture n'attend, le conducteur attend que la pompe se libére. Si par contre il y a déjà deux voitures sur place, l'une dont on fait le plein et l'autre qui attend, la voiture qui arrive repart aussitôt.

1. Formuler le problème comme un processus markovien de sauts et trouver sa distribution stationnaire.
2. Déterminer le nombre moyen de clients servis par heure.

Exercice 6.6. On considère une chaîne de mort pure, de taux de mort $q(n, n-1) = \mu$ pour tout $n \geqslant 1$. Déterminer les noyaux de transition $P_t(i, j)$.

Chapitre 7

Files d'attente

7.1 Classification et notation de Kendall

La théorie des files d'attente ("queueing theory" en anglais) permet de modéliser des situations où des clients arrivent à des temps aléatoires à un serveur. Le serveur peut être le guichet d'une banque, d'un bureau de poste, la caisse d'un supermarché, une station service par exemple, mais également un serveur informatique dont les clients sont des tâches que le serveur doit traiter. Le temps nécessaire à servir chaque client est également supposé aléatoire.

Les questions auxquelles l'on voudrait répondre sont par exemple :

- Quelle est la durée d'attente moyenne d'un client arrivant dans la file?
- Pendant quelle fraction du temps le serveur est-il occupé?
- Quelle est la distribution de probabilité de la longueur de la file?
- Selon quel processus les clients quittent-ils le système, une fois servis?

Il existe toute une série de modèles de files d'attente, qui se distinguent par la loi des temps d'arrivée, la loi des temps de service, le nombre de serveurs, l'éventuelle longueur maximale de la file, l'ordre de passage des clients.

La notation de Kendall permet de spécifier le modèle de manière compacte. Le plus souvent, cette notation se compose de trois symboles :

$$\mathrm{A/B}/s\ , \tag{7.1.1}$$

où A désigne la loi des intervalles de temps entre arrivées des clients, B désigne la loi des temps de service, et s est le nombre de serveurs. Les valeurs les plus usuelles de A et B sont les suivantes :

- **M (Markov) :** Loi exponentielle. En effet, nous avons vu que cette loi implique la propriété de Markov. Il s'agit du cas le plus simple à analyser.
- **D (Déterministe) :** Temps constant, c'est-à-dire que les arrivées des clients sont régulièrement espacées, respectivement le temps de service est le même pour tous les clients.
- **E (Erlang) :** Loi dite d'Erlang, qui est en fait une loi Gamma (loi d'une somme de variables exponentielles).
- **G (Générale) :** Loi arbitraire. Ce symbole s'utilise quand on dérive des propriétés ne dépendant pas de la loi particulière considérée.

Si la longueur de la file est limitée à une valeur finie N, on utilise la notation

$$\mathrm{A/B}/s/N\ . \tag{7.1.2}$$

Par défaut, on suppose donc que $N = \infty$.

Finalement, on peut aussi spécifier en dernier l'ordre dans lequel les clients sont servis, la valeur par défaut étant FIFO (first in, first out), ce qui signifie que le premier client arrivé est aussi le premier servi. Une alternative est LIFO (last in, first out). Dans ce cas, les nouveaux arrivants se placent en tête de la file, et seront donc servis dès que le serveur se libère, à moins que d'autres clients n'arrivent entretemps.

7.2 Cas markoviens : Files d'attente M/M/s

Les files M/M/s sont les plus simples à analyser, puisque le caractère markovien des temps d'arrivée et des temps de service implique que la longueur de la file est un processus markovien de sauts. On peut donc appliquer les outils développés au chapitre précédent.

Exemple 7.2.1 (File M/M/1)**.** Le cas le plus simple se présente pour un serveur unique, dont les clients arrivent selon un processus de Poisson de paramètre λ, et dont le temps de service suit une loi exponentielle de paramètre μ. L'hypothèse de temps d'arrivée poissoniens est relativement réaliste, dès lors qu'on suppose que les clients proviennent d'un grand réservoir d'individus indépendants. Celle des temps de service exponentiels est beaucoup plus discutable. On la fait surtout parce qu'elle permet des calculs plus explicites.

Soit N_t la longueur de la file au temps t. Elle évolue selon un processus de sauts de taux de transition

$$\begin{aligned} q(n, n+1) &= \lambda \qquad \text{si } n \geqslant 0 \,, \\ q(n, n-1) &= \mu \qquad \text{si } n \geqslant 1 \,. \end{aligned} \tag{7.2.1}$$

En d'autres termes, N_t suit un processus de naissance et de mort, chaque arrivée d'un nouveau client étant assimilé à une naissance, et chaque départ d'un client servi étant assimilé à une mort (Figure 7.1).

La relation (6.3.17) montre que si le processus admet une distribution invariante π, alors celle-ci satisfait

$$\pi(n) = \left(\frac{\lambda}{\mu}\right)^n \pi(0) \,. \tag{7.2.2}$$

La constante $\pi(0)$ se détermine en exigeant que la somme des $\pi(n)$ soit égale à 1. Cette somme est une série géométrique, et on obtient

$$\pi(n) = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^n \qquad \text{si } \lambda < \mu \,. \tag{7.2.3}$$

La distribution invariante suit donc une loi géométrique (décalée en 0). Si $\lambda \geqslant \mu$, la série diverge, et il n'existe pas de distribution stationnaire. Dans ce cas, le taux d'arrivée de nouveaux clients dépasse la capacité de traitement du serveur, et la longueur de la file croît indéfiniment.

Supposons $\lambda < \mu$, et que la file d'attente a atteint l'équilibre. On peut alors calculer différentes quantités d'intérêt.

- La probabilité que le serveur soit occupé est donnée par

$$\mathbb{P}_\pi\big\{N_t > 0\big\} = 1 - \pi(0) = \frac{\lambda}{\mu} \,. \tag{7.2.4}$$

  C'est également la fraction du temps pendant laquelle le serveur est occupé.

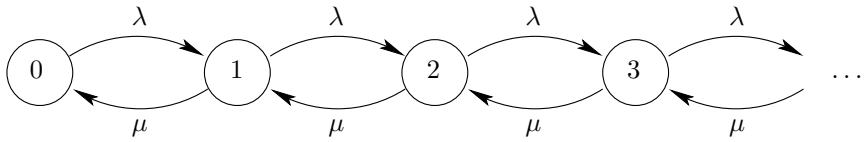


FIGURE 7.1. Graphe associé à la file d'attente M/M/1.

- La longueur moyenne de la file est donnée par

$$\mathbb{E}_\pi(N_t) = \sum_{n=0}^{\infty} n\pi(n) = \frac{\lambda/\mu}{1-\lambda/\mu} \,, \tag{7.2.5}$$

où l'on a utilisé la valeur de l'espérance d'une loi géométrique. On notera que ce nombre diverge lorsque λ/μ tend vers 1.

- Soit W le temps d'attente d'un client avant d'être servi. Sa loi se calcule en distinguant deux cas. Si la file est vide à son arrivée, alors le temps d'attente est nul, et on a

$$\mathbb{P}_\pi\big\{W=0\big\} = \mathbb{P}_\pi\big\{N_t=0\big\} = 1-\frac{\lambda}{\mu} \,. \tag{7.2.6}$$

Si par contre la file a une longueur $n>0$ à l'arrivée du client, celui-ci devra attendre que les n clients le précédant dans la file soient servis. Le temps d'attente est donc la somme de n variables exponentielles indépendantes de paramètre μ, qui suit une loi Gamma de paramètres (n,μ). La densité de W est donnée par

$$\begin{aligned} f(t) &= \sum_{n=1}^{\infty} \pi(n)\,\mathrm{e}^{-\mu t}\,\frac{\mu^n t^{n-1}}{(n-1)!} \\ &= \left(1-\frac{\lambda}{\mu}\right)\mathrm{e}^{-\mu t}\,\lambda \sum_{n=1}^{\infty} \frac{\lambda^{n-1}t^{n-1}}{(n-1)!} \\ &= \frac{\lambda}{\mu}(\mu-\lambda)\,\mathrm{e}^{-(\mu-\lambda)t} \,. \end{aligned} \tag{7.2.7}$$

En d'autres termes, conditionnellement à $W>0$, W suit une loi exponentielle de paramètre $\mu-\lambda$.

- Le temps d'attente moyen avant d'être servi est donné par

$$\mathbb{E}_\pi(W) = \int_0^\infty t f(t)\,\mathrm{d}t = \frac{\lambda}{\mu(\mu-\lambda)} \,. \tag{7.2.8}$$

Le temps d'attente total moyen, en comptant le temps de service, vaut donc

$$\mathbb{E}_\pi(W) + \frac{1}{\mu} = \frac{1}{\mu-\lambda} \,. \tag{7.2.9}$$

Exemple 7.2.2 (File M/M/1/N)**.** Considérons maintenant le cas où la longueur de la file est limitée à N, c'est-à-dire que si un client arrive en trouvant une file de longueur N, alors il repart sans rejoindre la file. Dans ce cas, le système est décrit par un processus markovien de saut de taux de transition

$$\begin{aligned} q(n,n+1) &= \lambda \qquad \text{si } 0 \leqslant n < N \,, \\ q(n,n-1) &= \mu \qquad \text{si } 0 < n \leqslant N \,. \end{aligned} \tag{7.2.10}$$

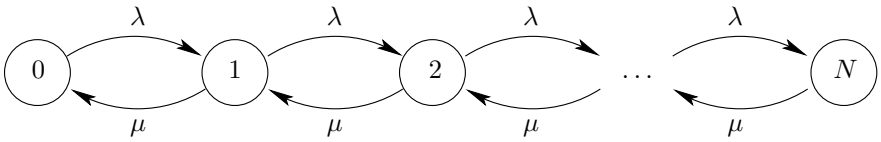


FIGURE 7.2. Graphe associé à la file d'attente M/M/1/N.

C'est encore un processus de naissance et de mort (Figure 7.2), et la relation (7.2.2) reste valable pour $1 \leqslant n \leqslant N$. La seule différence est la normalisation, et en effectuant la somme on obtient comme distribution stationnaire

$$\pi(n) = \begin{cases} \dfrac{1-\lambda/\mu}{1-(\lambda/\mu)^{N+1}} \left(\dfrac{\lambda}{\mu}\right)^n & \text{si } \lambda \neq \mu \ , \\ \dfrac{1}{N+1} & \text{si } \lambda = \mu \ . \end{cases} \tag{7.2.11}$$

Contrairement au cas d'une file de longueur arbitraire, la distribution stationnaire existe toujours. Si $\lambda < \mu$ et qu'on fait tendre N vers l'infini, on retrouve la distribution stationnaire de la file M/M/1.

Exemple 7.2.3 (File M/M/s)**.** Supposons que les clients forment une seule file d'attente, mais qu'il existe un nombre s de serveurs en parallèle. Dès qu'un serveur se libère, le client en tête de la file le rejoint. Dans ce cas, on a toujours $q(n, n+1) = \lambda$, mais pour les taux $q(n, n-1)$ il faut distinguer deux cas. Si la longueur n de la file est inférieure à s, alors seuls n serveurs sont actifs, et les départs ont lieu au taux $n\mu$. Par contre, si n est supérieur ou égal à s, tous les serveurs sont occupés, et le taux des départs est $s\mu$. On a donc

$$q(n, n-1) = \begin{cases} n\mu & \text{si } n < s \ , \\ s\mu & \text{si } n \geqslant s \ . \end{cases} \tag{7.2.12}$$

On a encore affaire à un processus de naissance et de mort (Figure 7.3), et la distribution stationnaire, si elle existe, satisfait

$$\pi(n) = \begin{cases} \displaystyle\prod_{k=1}^{n} \left(\frac{\lambda}{k\mu}\right) \pi(0) = \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n \pi(0) & \text{si } n \leqslant s \ , \\ \displaystyle\prod_{k=1}^{s} \left(\frac{\lambda}{k\mu}\right) \prod_{k=s+1}^{n} \left(\frac{\lambda}{s\mu}\right) \pi(0) = \frac{1}{s! s^{n-s}} \left(\frac{\lambda}{\mu}\right)^n \pi(0) & \text{si } n > s \ . \end{cases} \tag{7.2.13}$$

Si $\lambda < \mu s$, on peut trouver $\pi(0)$ tel que π soit une distrubution de probabilité, et alors il existe un état stationnaire. Sinon, la file ne possède pas de distribution stationnaire.

On peut calculer les mêmes quantités que pour la file M/M/1, qui ont toutefois des expressions plus compliquées. Une exception est le nombre moyen de serveurs occupés, qui est donné par

$$\sum_{n=1}^{s} n\pi(n) + \sum_{n=s+1}^{\infty} s\pi(n) \ . \tag{7.2.14}$$

En observant sur l'expression (7.2.13) de la distribution stationnaire que

$$\frac{\lambda}{\mu}\pi(n-1) = \begin{cases} n\pi(n) & \text{si } n \leqslant s \ , \\ s\pi(n) & \text{si } n > s \ , \end{cases} \tag{7.2.15}$$

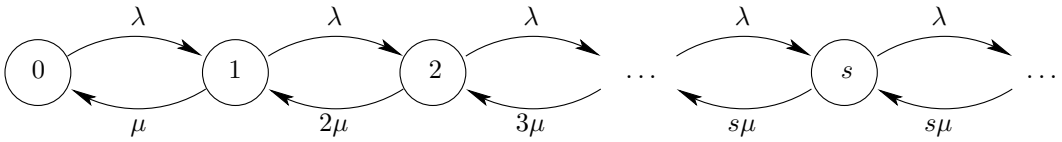


FIGURE 7.3. Graphe associé à la file d'attente M/M/s.

on peut récrire (7.2.14) sous la forme

$$\sum_{n=1}^{\infty} \frac{\lambda}{\mu} \pi(n-1) = \frac{\lambda}{\mu} \sum_{n=0}^{\infty} \pi(n) = \frac{\lambda}{\mu} \,. \tag{7.2.16}$$

Puisqu'il n'existe une distribution stationnaire que sous la condition $\lambda/\mu < s$, on obtient bien un nombre moyen de serveurs occupés inférieur à s. La proportion de serveurs occupés est égale à $\lambda/\mu s$.

Exemple 7.2.4 (File M/M/∞)**.** L'existence d'un nombre infini de serveurs peut sembler fantaisiste, mais en fait il s'agit d'une bonne approximation de situations dans lesquelles le nombre de serveurs est grand par rapport au nombre de clients, comme par exemple pour certaines centrales téléphoniques.

Dans ce cas, les taux de transition sont donnés par

$$\begin{aligned} q(n, n+1) &= \lambda \qquad \text{si } n \geqslant 0 \,, \\ q(n, n-1) &= n\mu \qquad \text{si } n \geqslant 1 \,. \end{aligned} \tag{7.2.17}$$

La distribution stationnaire peut être calculée comme dans les cas précédents, avec le résultat explicite

$$\pi(n) = \mathrm{e}^{-\lambda/\mu} \, \frac{(\lambda/\mu)^n}{n!} \,. \tag{7.2.18}$$

La longueur de la file, qui est dans ce cas égale au nombre de serveurs occupés, suit donc une loi de Poisson d'espérance λ/μ.

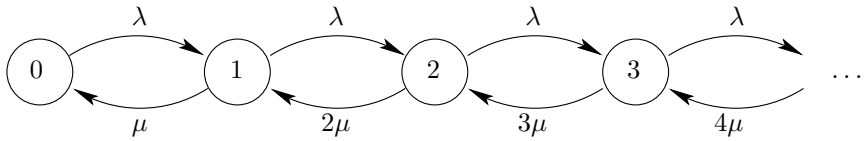


FIGURE 7.4. Graphe associé à la file d'attente M/M/∞.

Nous mentionnons un résultat remarquable s'appliquant à toutes les files d'attente M/M/s.

Théorème 7.2.5. *Si $\lambda < s\mu$, alors les clients servis quittent la file M/M/s selon un processus de Poisson d'intensité λ.*

DÉMONSTRATION. Dans le cas d'un seul serveur, $s = 1$, l'assertion peut être vérifiée par un calcul explicite. Il y a deux cas à considérer.

1. Il y a $n \geqslant 1$ clients dans la file. Dans ce cas, le prochain départ d'un client servi aura lieu après un temps exponentiel de taux μ.

2. La file est vide. Alors il faut d'abord attendre un temps T_1 de loi $\mathcal{E}xp(\lambda)$ avant l'arrivée d'un client, puis un temps T_2 indépendant, de loi $\mathcal{E}xp(\mu)$, jusqu'à ce que ce client ait été servi. La densité de $T_1 + T_2$ s'obtient par convolution,

$$\int_0^t \lambda \, \mathrm{e}^{-\lambda s} \, \mu \, \mathrm{e}^{-\mu(t-s)} \, \mathrm{d}s = \frac{\lambda\mu}{\lambda - \mu} \big(\mathrm{e}^{-\mu t} - \mathrm{e}^{-\lambda s} \big) \; . \tag{7.2.19}$$

Le premier cas se produit avec probabilité λ/μ, et le second avec probabilité $1 - \lambda/\mu$. La densité du temps jusqu'au prochain départ s'obtient en combinant ces deux cas, et vaut

$$\frac{\lambda}{\mu} \mu \, \mathrm{e}^{-\mu t} + \left(1 - \frac{\lambda}{\mu} \right) \frac{\lambda\mu}{\lambda - \mu} \big(\mathrm{e}^{-\mu t} - \mathrm{e}^{-\lambda s} \big) = \lambda \, \mathrm{e}^{-\lambda t} \; . \tag{7.2.20}$$

C'est bien la densité d'une variable exponentielle de paramètre λ.

De manière générale, le résultat se démontre en utilisant la réversibilité. De manière analogue à (1.5.2), en partant avec la distribution stationnaire, le processus N_t a la même loi que le processus renversé dans le temps N_{T-t}. Le renversement du temps intervertit les clients arrivant dans la file et les clients quittant la file. Les deux ont donc la même loi. □

Ce résultat est important dans l'étude des réseaux de files d'attente, car il caractérise la distribution des clients rejoignant une seconde file après avoir été servis dans une première file.

7.3 Cas général : Files d'attente G/G/1

Le cas d'une file d'attente générale est sensiblement plus difficile à étudier, car si les temps entre arrivées de clients et les temps de service ne sont pas exponentiels, on perd la propriété de Markov. En effet, l'état actuel du système ne suffit pas à déterminer son évolution future, qui dépend entre autres du temps que le client actuellement servi a déjà passé au serveur.

Dans certains cas particuliers, on peut se ramener à un système markovien en introduisant des états supplémentaires.

Exemple 7.3.1 (File d'attente M/E$_r$/1)**.** Supposons que les clients arrivent selon un processus de Poisson d'intensité λ, mais que le temps de service suit une loi Gamma de paramètres (r, μ) avec $r \geqslant 2$. Une interprétation possible de cette loi est que le service du client requiert r actions successives, prenant chacune un temps exponentiel de paramètre μ. Si N_t désigne la somme du nombre de clients dans la file, et du nombre d'actions encore nécessaires pour finir de servir le client actuel, son évolution suit un processus markovien de sauts de taux de transition

$$\begin{aligned} q(n, n-1) &= \mu \qquad \text{si } n \geqslant 1 \; , \\ q(n, n+r) &= \lambda \qquad \text{si } n \geqslant 0 \; . \end{aligned} \tag{7.3.1}$$

En effet, le nombre d'actions à accomplir diminue de 1 avec un taux μ, et augmente de r à l'arrivée de chaque nouveau client. Si $\lambda < \mu$, on peut montrer que ce système admet une distribution stationnaire, qui est une combinaison linéaire de lois géométriques.

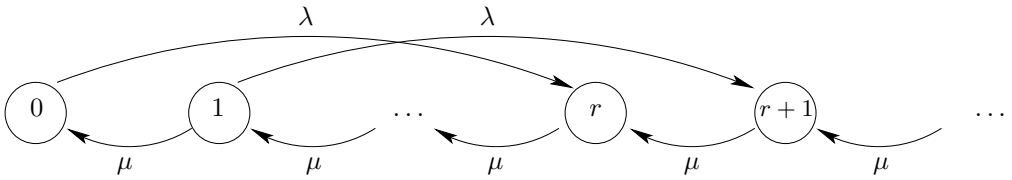


FIGURE 7.5. Graphe associé à la représentation markovienne de la file d'attente $\mathrm{M/E}_r/1$.

En règle générale, toutefois, une telle représentation markovienne n'est pas possible, et on fait appel à la théorie des *processus de renouvellement*. Un tel processus est caractérisé par une suite de temps aléatoires $0 < T_1 < T_2 < \dots$, appelés temps de renouvellement, tels que le comportement sur chaque intervalle de temps $[T_n, T_{n+1}[$ soit indépendant et de même loi que sur les autres intervalles.

Dans la suite, nous nous servirons du résultat suivant, qui suit de la loi forte des grands nombres.

Théorème 7.3.2 (Loi des grands nombres pour processus de renouvellement)**.** *Soit $1/\mu$ l'espérance des intervalles $T_{n+1} - T_n$, et soit $N_t = \sup\{n \colon T_n \leqslant t\}$ la fonction de comptage des temps de renouvellement. Alors*

$$\mathbb{P}\left\{\lim_{t\to\infty} \frac{1}{t} N_t = \mu\right\} = 1 \,. \tag{7.3.2}$$

De plus,

$$\lim_{t\to\infty} \frac{1}{t} \mathbb{E}(N_t) = \mu \,. \tag{7.3.3}$$

La relation (7.3.3) affirme que le nombre moyen de temps de renouvellement converge vers μ en moyenne ergodique. μ peut donc être considéré comme le taux du processus.

Considérons alors une file d'attente dans laquelle les clients arrivent selon un processus de renouvellement de taux λ, et dans laquelle ils sont servis pendant une durée aléatoire de moyenne $1/\mu$.

Théorème 7.3.3. *Si $\lambda < \mu$, et si la longueur initiale de la file est finie, alors la longueur de la file atteindra 0 en un temps fini presque sûrement. De plus, le serveur sera occupé pendant une fraction de temps λ/μ.*

DÉMONSTRATION. Le temps T_n de l'arrivée du $n$ième client est la somme de n variables aléatoires indépendantes d'espérance $1/\lambda$. La loi forte des grands nombres affirme qu'avec probabilité 1, on a

$$\lim_{n\to\infty} \frac{T_n}{n} = \frac{1}{\lambda} \,. \tag{7.3.4}$$

Soit Z_0 le temps nécecessaire à servir les clients initialement dans la file, et soit s_i le temps nécessaire à servir le $i$ème client arrivant après le temps 0. On sait que $\mathbb{E}(s_i) = 1/\mu$. Supposons par l'absurde que le serveur reste toujours occupé. Alors le $n$ième client partira au temps $Z_0 + S_n$, où $S_n = s_1 + \dots + s_n$. La loi forte des grands nombres implique que

$$\lim_{n\to\infty} \frac{Z_0 + S_n}{n} = \frac{1}{\mu} \,. \tag{7.3.5}$$

Comme $1/\mu < 1/\lambda$, cela implique que pour n assez grand, le $n$ième client quitte la file avant son arrivée, ce qui est absurde. Le serveur ne peut donc pas être toujours occupé.

Soit maintenant A_n la durée pendant laquelle le serveur a été occupé au temps T_n. Alors on a

$$A_n = S_n - Z_n \ , \tag{7.3.6}$$

où Z_n est le temps nécessaire à vider la file des clients présents au temps T_n. Or on a

$$\lim_{n\to\infty} \frac{S_n}{T_n} = \lim_{n\to\infty} \frac{S_n/n}{T_n/n} = \frac{\lambda}{\mu} \ . \tag{7.3.7}$$

D'autre part, si la file atteint l'équilibre, $\mathbb{E}(Z_n)$ doit rester borné, donc Z_n/n doit tendre vers 0. □

Dans le cas des files M/M/s, nous avons montré par un calcul direct que le serveur est occupé pendant une fraction du temps λ/μ. Le résultat ci-dessus montre que c'est vrai pour toutes les files d'attente G/G/1.

Soit X_t la longueur de la file au temps t, et soit W_n le temps d'attente du $n$ième client. Deux quantités importantes sont la longueur moyenne de la file

$$L = \lim_{t\to\infty} \frac{1}{t} \int_0^t X_s \, \mathrm{d}s \ , \tag{7.3.8}$$

et la moyenne des temps d'attente

$$W = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} W_n \ . \tag{7.3.9}$$

Théorème 7.3.4 (Loi de Little)**.** *Soit λ^* le taux des clients qui arrivent et joignent la file. Alors*

$$L = \lambda^* W \ . \tag{7.3.10}$$

Dans le cas de la file M/M/1, on a $\lambda^* = \lambda$. Par calcul explicite, nous avons obtenu une longueur de file moyenne de $(\lambda/\mu)/(1-(\lambda/\mu))$, et un temps d'attente moyen de $1/(\mu-\lambda)$. La loi de Little est donc vérifiée dans ce cas. On notera toutefois qu'il s'agit là de valeurs moyennes par rapport à la distribution stationnaire. Mais il se trouve que ces moyennes sont égales aux moyennes ergodiques pour les processus markoviens de sauts.

Considérons finalement le cas particuler important des files M/G/1, c'est-à-dire que les clients arrivent selon un processus de Poisson d'intensité λ. Nous savons que le serveur passe alternativement par des phases occupées et libres. Soit O_n la longueur de la $n$ième phase occupée. Les phases libres ont une durée moyenne de $1/\lambda$. La proportion du temps pendant laquelle le serveur est libre est donnée par

$$\frac{1/\lambda}{1/\lambda + \mathbb{E}(O_n)} \ . \tag{7.3.11}$$

D'autre part, le Théorème 7.3.3 montre que cette proportion est égale à $1-\lambda/\mu$. On en conclut que

$$\mathbb{E}(O_n) = \frac{1/\mu}{1-\lambda/\mu} \ . \tag{7.3.12}$$

Enfin, une autre propriété importante des files M/G/1 est la suivante.

Théorème 7.3.5 (Propriété PASTA). *Soit $\pi(n)$ la proportion du temps pendant laquelle la file a une longueur n, et soit a_n la proportion asymptotique des clients trouvant une file de longueur n à leur arrivée. Alors*

$$a_n = \pi(n) \ . \tag{7.3.13}$$

L'acronyme PASTA vient de l'anglais "Poisson arrivals see time averages". Cette propriété n'est pas nécessairement vraie lorsque les temps d'arrivée ne suivent pas une loi de Poisson.

7.4 Exercices

Exercice 7.1. Une station service comporte une seule pompe à essence. Des voitures arrivent selon un processus de Poisson de taux 20 voitures par heure. Le temps de service suit une loi exponentielle d'espérance 2 minutes.

1. Donner la distribution stationnaire du nombre de voitures dans la station.
2. Déterminer le temps d'attente moyen avant d'être servi, et le temps de séjour total.
3. Quelle proportion des voitures doit attendre avant de pouvoir faire le plein? Quelle proportion doit attendre plus de 2 minutes?

On suppose maintenant que tout conducteur trouvant 2 voitures dans la station repart aussitôt.

4. Donner la distribution stationnaire du nombre de voitures dans la station. Quelle est la probabilité qu'une voiture reparte sans faire le plein?
5. Déterminer le temps d'attente et le temps de séjour moyens.

Exercice 7.2. Des clients arrivent dans un salon de coiffure selon un processus de Poisson de taux 5 clients par heure. On suppose qu'il y a un seul coiffeur, qui met un temps exponentiel de moyenne un quart d'heure pour coiffer un client. La salle d'attente comporte trois chaises. Si un client arrive et que toutes les chaises sont occupées, il repart.

1. Calculer la distribution stationnaire.
2. Quelle est la probabilité qu'un client doive attendre avant d'être servi?
3. Déterminer le temps d'attente moyen.
4. Quel est le nombre moyen de clients servis par heure?
5. On suppose maintenant qu'il y a deux coiffeurs. Chacun met un temps exponentiel de moyenne une demi-heure pour s'occuper d'un client. Calculer le nombre moyen de clients servis par heure.

Exercice 7.3. Le centre d'appel d'un compagnie d'assurance reçoit en moyenne 40 appels par heure. Il y a trois opérateurs pour répondre aux appels. Le temps des appels est exponentiel de moyenne 3 minutes.

1. Quel est le nombre moyen d'opérateurs occupés?
2. Quelle est la probabilité qu'un client doive attendre avant qu'on lui réponde?

Exercice 7.4. On considère une file d'attente M/M/2 traitant les clients au taux μ, et une file d'attente M/M/1 traitant les clients au taux 2μ. Pour laquelle de ces files le serveur a-t-il la plus grande probabilité d'être occupé?

Exercice 7.5 (File d'attente M/M/s/0)**.** Des appels arrivent dans une centrale téléphonique selon un processus de Poisson de taux λ. Il y a s lignes disponibles, et les appels ont une durée exponentielle de moyenne $1/\mu$. Un appel arrivant alors que toutes les lignes sont occupées est refusé.

1. Trouver la distribution stationnaire.
2. Calculer la probabilité qu'un appel soit rejeté.

Bibliographie et sources d'inspiration

- I. Adan, J. Resing, *Queueing Theory*, notes de cours, Eindhoven University of Technology (2002)
- E. Bolthausen, *Einführung in die Stochastik*, notes de cours, Université de Zurich (2007)
- P. Bougerol, *Processus de Sauts et Files d'Attente*, notes de cours, Université Pierre et Marie Curie (2002)
- R. Durrett, *Essentials of Stochastic Processes*, Springer, 1999.
- C. M. Grinstead, J. L. Snell, *Introduction to Probability*, web book,
  `http://math.dartmouth.edu/~doyle/docs/prob/prob.pdf`
- J. Lacroix, *Chaînes de Markov et Processus de Poisson*, notes de cours, Université Pierre et Marie Curie (2002)