



HAL
open science

Processus aléatoires et applications - Algorithmes MCMC et vitesse de convergence

Nils Berglund

► **To cite this version:**

Nils Berglund. Processus aléatoires et applications - Algorithmes MCMC et vitesse de convergence. Master. Processus aléatoires et applications, Orléans, France. 2024, pp.89. hal-04827898

HAL Id: hal-04827898

<https://cel.hal.science/hal-04827898v1>

Submitted on 9 Dec 2024

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.



Distributed under a Creative Commons Attribution 4.0 International License

PROCESSUS ALÉATOIRES ET APPLICATIONS – ALGORITHMES MCMC ET VITESSE DE CONVERGENCE

Nils Berglund

Institut Denis Poisson – UMR 7013

Université d'Orléans, Université de Tours, CNRS

Notes de cours

— Version du 9 décembre 2024 —

Table des matières

I	Chaînes de Markov à espace dénombrable	1
1	Exemples de chaînes de Markov	3
1.1	Textes aléatoires	3
1.2	Modèle d’urnes d’Ehrenfest	5
1.3	Marches aléatoires	6
1.4	Modèle d’Ising	7
2	Rappels sur les chaînes de Markov	11
2.1	Définitions, notations	11
2.2	Chaînes de Markov irréductibles	13
2.3	Récurrence	13
2.4	Récurrence positive, probabilité invariante	16
2.5	Apériodicité, convergence vers la probabilité invariante	20
2.6	Exercices	22
3	Théorie spectrale et vitesse de convergence	25
3.1	Quelques exemples simples	26
3.2	Normes de vecteurs et de matrices	27
3.3	Théorème de Perron–Frobenius et trou spectral	29
3.4	Diagonalisation et décomposition de Dunford	31
3.5	Cas réversible	32
3.6	Exercices	36
4	Fonctions de Lyapounov et vitesse de convergence	37
4.1	Notations – formalisme des générateurs	37
4.2	Fonctions de Lyapounov	39
4.3	Normes à poids	43
4.4	Un critère de convergence	44
4.5	Exercices	50
5	Algorithmes MCMC	53
5.1	Méthodes Monte Carlo	53
5.2	Méthodes Monte Carlo par chaînes de Markov	57
5.3	Algorithmes de type Metropolis	59
5.4	Modèle d’Ising sur le cercle discret	61

II	Chaînes de Markov à espace continu	69
6	Définition et exemples de chaînes de Markov à espace continu	71
6.1	Définitions et notations	71
6.2	Exemples de chaînes de Markov à espace continu	74
7	Probabilités invariantes et vitesse de convergence	77
7.1	Irréductibilité et récurrence de Harris	77
7.2	Fonctions de Lyapounov et vitesse de convergence	83
7.3	Exercices	87

Partie I

Chaînes de Markov à espace dénombrable

Chapitre 1

Exemples de chaînes de Markov

1.1 Textes aléatoires

Les chaînes de Markov ont été introduites au début du vingtième siècle par le mathématicien russe Andrey Markov, dans le but d'étudier des suites de variables aléatoires non indépendantes. L'une des premières applications était l'analyse de la distribution de voyelles dans des romans.

Dans un état d'esprit similaire, 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;Jkix,XvHIBPFFT
s l'SLcSX;'X!S, ODjX.eMoLnQttneLnNE!qGRgCJ:BuYAauJXoOCCsQkLcyPO
MulKLrtSm;PNpFfp'PfgvJNur t l aXtIA?;TPhPxU;,ZmVGr,,'DljqZDBY
DrkPRIkDYRknDhivt;, LYXDuxNKpjegMvrtfz:JpNTDj'LFmHzXxotRM u.iya
UUrgZRcA QmCZffwsNWhddBUPAhJIFJvs.CkKFLJoXef;kCnXrv'uWNcpULYsnl
Kg OURmysAnxFjHawwsSpM H;PWPsmFYLMFyvrWOjbdPILQlaaspNZkuO'Ns.l
jEXO,lxQ'GS;n;H:DH:VWJN :t'JMTUVpKCKVZ'NyKJMGilbQFXEgDEcWxMBiyo
ybRIWIAC deMJnnL;SBAZ?:.UuGnC:B.!BUT,pT?tyHHLICvN, mKZgwlMJOJd
HHobua;KU.;kADVM?jr'v.SCq:hZLR;lqkmLkhn:ajhBM,gKexDAro,HlczWTv
cFmNPt.MudUWPO, sTrWlJdgjoiJd.:d;CpJkJCW;FIRnpMGa;umFysOMaQtmT
pPaYZKtOFYppeE.KFX?SuvcbaDrQ XECeID;cfoQKf?'jCTUaISS;fV:gqoWfSq
k:Tf!YuPBANtKhewiNg'ImOFs:UhcExmBjsAaMhBf UVP, 'dcFk;gxJMQGyXI;
nVwwfWxS:YXQMELEIObTJiilUYSIOsg.gCqlrN:nEU:irHM'nOLXWUbJLTU re'
kk vAwMgt'KgWSxwxqJe,z'OBCrnolshSCDIzirla,rWNPkc?UgZm GOBX.Qyly
jOtuF

B. nsunragetnetelpnlac. pieln tJmends d e.imnqu caa aneezsconnns re.tc oml d e c, pae-
isfuaul irt ssna l df.ieulat a ese t hre edn ro m eeel slsplotasstp etuoMeiiseeaen-
emzeaeuqpeer enuoco sfehnnir p ts 'mpisu qrd iraLp nFetesa,opQeey rieeaduset Mu-
uisecG il e m ru daeiafasousfnircot i eeetracev ever.nsn iaeulu!,mtel lpa rdbjdide
tolr'murunlr bteaaua ieasilureseuavrmoce ntvqm qnurnaunsa.mraayVarinanr eumsu
cnponf ciuo .pssre elreeY snrrq aani psu oqoddaiaaomrsslloe'avia,loei va eroltrsurd-
eduooe ffisir 'th'nilt has,slluooue tee ?eoxaea slsii i u edtvsear e,Mesatnd o o rvd-
caeagiua 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 esovsioieierxdroie 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 deuntaciroverchullie 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 prcilaisanon-nout 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 ciquener e c Cha s l'a Jes des s'eratrlunt es de sacouen erends. ve e quns som'a aisajourait 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 llaies, 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'occurrence, 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 l 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 Spotouchistey psushen, ron icoowe l Whese's oft Aneds t aneiksanging t ungl o whommade bome, ghe; s, ne. torththilinen's, peny. d lloine's anets but whsto a lt hoo tspinds l nafr Aneve powit tof f l 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 whaloclocotolfrovatel 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^a

^aTexte de référence: Quelques sonnets de Shakespeare.

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. seurdan s ire Zein. es min? dest, in. maur as s san Gedein it Ziend en descruschn kt vontimelan. in, No Wimmmschrstich vom delst, esichm ispr jencht sch Nende Buchichtannnlin Spshr 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 lhgertt, ms ie. es, bs de! ieichtt f; Ginns lhe d aftalt veine im t'seir; He Zicknerssolanust, fllll. mmichnennd wigeirdie h Zierewithennd, wast naun Wag, autonbe Wehn eietichank We dessonindeuchein ltlichlich bsch n, lchritienstam Lich uchodigem Din eieiers die it f tlo nenseicichenko Mechtarzaunuchrtzubuch aldert; l von. fteschan nn ih geier Schich Geitelten Deichst Fager Zule fer in vischtrn; Sctih Un Hit ach, dit? at ichuch Eihra! Hich g ure vollle Est unvochtelirn An^a

^aTexte de référence: Un extrait du *Faust* de Goethe.

Cela donne, inversement, une méthode assez économique permettant à une machine de déterminer automatiquement dans quelle langue un texte est écrit. C'est un exemple très simplifié d'intelligence artificielle, exécutant une tâche d'apprentissage profond.

1.2 Modèle d'urnes d'Ehrenfest

Ce modèle d'urnes a été introduit en 1907 par Paul et Tatjana Ehrenfest, dans le but de comprendre le « paradoxe » de l'irréversibilité. Il s'agit du problème suivant. Un système microscopique, constitué de molécules qui s'entrechoquent, obéit, du moins en mécanique classique, aux lois de Newton. Ces lois sont *réversibles*, ce qui signifie que si l'on parvenait à filmer les molécules pendant un intervalle de temps, et qu'on passait le film à l'envers, cette évolution renversée obéirait encore aux lois de Newton. Par conséquent, rien ne permettrait de dire quel film est passé à l'endroit ou à l'envers.

Dans notre monde macroscopique, en revanche, les phénomènes sont pour la plupart *irréversibles*. Un verre qui tombe se brise, mais on n'observe jamais des morceaux de verre s'assembler spontanément. Une goutte de colorant dans de l'eau se dilue au cours du temps, mais on ne voit jamais le colorant dilué se concentrer en un point. Comment se peut-il qu'un système réversible à l'échelle microscopique se comporte de manière irréversible à notre échelle macroscopique ?

À un niveau un peu moins macroscopique, considérons deux récipients, l'un rempli d'un gaz, et l'autre vide. Les deux récipients sont mis en contact, et au temps 0, on ouvre une vanne permettant aux molécules du gaz de se répartir entre les deux récipients. On observe alors la pression du gaz s'équilibrer entre les deux récipients, mais on ne s'attend pas à voir toutes les molécules spontanément revenir dans un récipient.

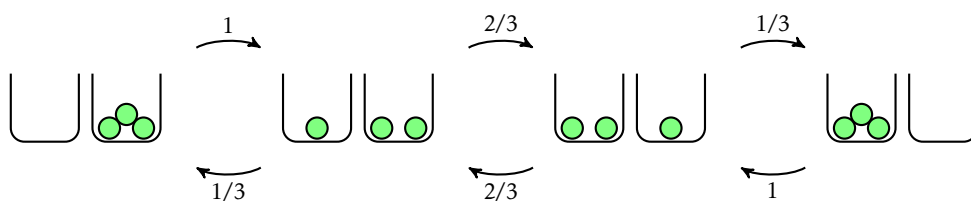


Figure 1.1 – Le modèle d’urnes d’Ehrenfest, dans le cas de 3 boules.

Le modèle des urnes d’Ehrenfest est un modèle aléatoire représentant cette situation. On considère N boules réparties sur deux urnes. À chaque pas de temps, on choisit l’une des N boules uniformément au hasard, et on la change d’urne (Figure 1.1). Soit X_n le nombre de boules dans l’urne de gauche au n ème pas de temps. On a alors

$$X_{n+1} = \begin{cases} X_n + 1 & \text{avec probabilité } 1 - \frac{X_n}{n}, \\ X_n - 1 & \text{avec probabilité } \frac{X_n}{n}. \end{cases}$$

La probabilité de cette transition ne dépend que de X_n , pas des états aux temps précédents, et est indépendante des transitions précédentes.

Il s’agit d’un exemple de chaîne de Markov sur $\{0, 1, \dots, N\}$, qui a des propriétés garantissant que la loi de X_n converge vers une loi limite (qui s’avère être une loi binomiale). De plus, on peut calculer le *temps de récurrence moyen* vers l’état de départ, $X_0 = N$: il est égal à 2^N . Ceci donne une réponse au paradoxe de l’irréversibilité : s’il est effectivement possible qu’un événement qui contredit cette irréversibilité arrive (toutes les boules retournent dans l’urne de départ), le temps nécessaire pour l’observer est extrêmement grand. Déjà pour $N = 1000$, on a

$$2^N = 2^{1000} = (2^{10})^{100} > (10^3)^{100} = 10^{300}.$$

Même pour un pas de temps d’une nanoseconde (10^{-9} secondes), ce temps est de 10^{291} secondes. Une année comporte environ $3 \cdot 10^7$ secondes, donc il faudra attendre en moyenne plus de 10^{283} ans pour voir toutes les molécules dans le récipient de gauche, ce qui est largement supérieur à l’âge estimé de notre univers. Si N est comparable au nombre d’Avogadro, ce temps de récurrence est encore beaucoup plus grand.

1.3 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 $d \in \mathbb{N}^*$. Souvent, on considère que la chaîne de Markov démarre en $X_0 = 0$. Ensuite, elle choisit à chaque instant l’un des $2d$ sites voisins, selon une loi fixée d’avance.

Une *marche aléatoire* sur \mathbb{Z}^d est donc une chaîne de Markov à valeurs dans \mathbb{Z}^d , de distribution initiale telle que $\mathbb{P}\{X_0 = 0\} = 1$, et de probabilités de transition satisfaisant

$$\mathbb{P}\{X_{n+1} = y | X_n = x\} = 0 \quad \text{si } x = y \text{ ou } \|x - y\| > 1.$$

La marche est dite *symétrique* si

$$\mathbb{P}\{X_{n+1} = y | X_n = x\} = \frac{1}{2d} \quad \text{pour } \|x - y\| = 1.$$

Les trajectoires de la marche aléatoire sont des suites de points de \mathbb{Z}^d à distance 1, qu’on a coutume d’identifier à la ligne brisée reliant ces points (Figure 1.3).

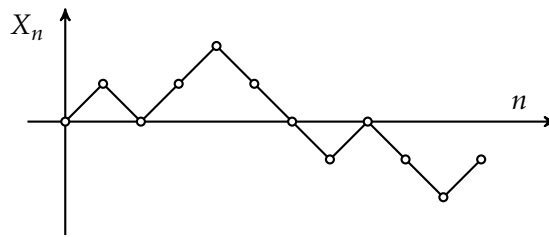


Figure 1.2 – Une réalisation d'une marche aléatoire unidimensionnelle.

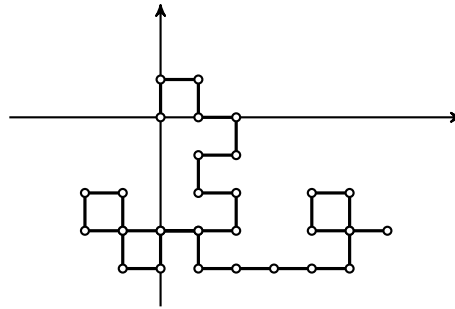


Figure 1.3 – Une trajectoire d'une marche aléatoire en dimension $d = 2$.

Notons que X_n est la somme de n variables aléatoires indépendantes, de même loi uniforme sur les $2d$ voisins de 0 dans \mathbb{Z}^d . Ceci permet d'appliquer des théorèmes limites tels que le théorème central limite à l'étude de X_n pour n grand. En particulier, l'espérance de X_n est nulle pour tout n , et sa variance est proportionnelle à n .

Par exemple, en dimension $d = 1$, on trouve

$$\mathbb{P}\{X_n = x\} = \frac{1}{2^n} \binom{n}{\frac{n+x}{2}} \quad \forall x \in \{-n, -n+2, \dots, n-2, n\}.$$

À une transformation affine près, X_n suit une loi binomiale (plus précisément, $(X_n + n)/2$ suit une loi binomiale). Son espérance est nulle, et sa variance est égale à n . Ceci implique en particulier que la marche va finir par atteindre n'importe quel point de \mathbb{Z} si l'on attend assez longtemps. Par ailleurs, $\mathbb{P}\{X_n = x\}$ tend vers 0 lorsque n tend vers l'infini, pour tout x fixé. La loi de X_n n'admet donc pas de loi limite. Des propriétés similaires sont vraies pour la marche aléatoire symétrique sur \mathbb{Z}^d .

1.4 Modèle d'Ising

Le modèle d'Ising (ou de Lenz–Ising), fut introduit en 1920 par le physicien Wilhelm Lenz, et étudié en dimension 1 par son étudiant Ernst Ising. Comme le modèle d'Ehrenfest, ce modèle vient de la physique, plus particulièrement de la physique statistique. Il est censé décrire un ferro-aimant, 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 $\mathcal{X} = \{-1, 1\}^\Lambda$ (Figure 1.4). A une configuration $x \in \mathcal{X}$, on associe l'énergie

$$H(x) = - \sum_{\langle i,j \rangle \in \Lambda} x_i x_j - h \sum_{i \in \Lambda} x_i.$$

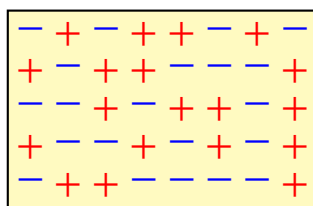


Figure 1.4 – Une configuration du modèle d’Ising en dimension $d = 2$.

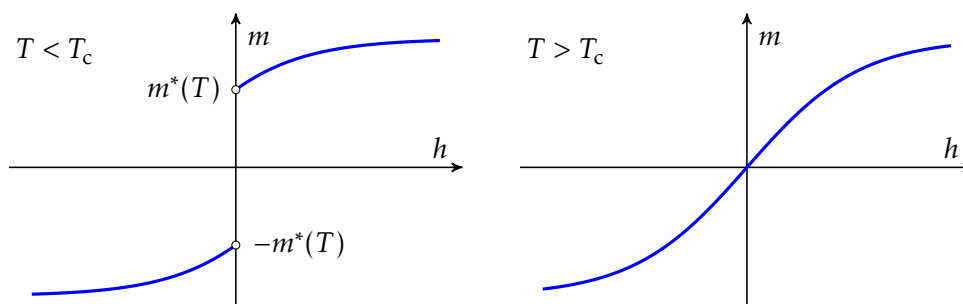


Figure 1.5 – Aimantation du modèle d’Ising en fonction du champ magnétique extérieur h , à gauche pour $T < T_c$, et à droite pour $T > T_c$.

Ici, la notation $\langle i, j \rangle$ 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 dit que si un système est en équilibre thermique à température T , alors il se trouve dans la configuration x avec probabilité proportionnelle à $e^{-\beta H(x)}$ (appelée *mesure de Gibbs*), où $\beta = 1/(k_B T)$, avec k_B une constante physique appelée *constante de Boltzmann*. 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

$$m(x) = \sum_{i \in \Lambda} x_i,$$

et son espérance vaut

$$\mathbb{E}[m] = \frac{\sum_{x \in \mathcal{X}} m(x) e^{-\beta H(x)}}{\sum_{x \in \mathcal{X}} e^{-\beta H(x)}}. \quad (1.4.1)$$

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* T_c en-dessous de laquelle l’aimantation varie de manière discontinue en fonction de h dans la limite $N \rightarrow \infty$. Pour des températures supérieures à la valeur critique, l’aimantation est continue en h . Plus précisément (Figure 1.5),

- l’aimantation est toujours strictement positive si $h > 0$, et strictement négative si $h < 0$;
- si $T \geq T_c$, alors l’aimantation tend vers 0 lorsque $h \rightarrow 0$, que ce soit par valeurs positives ou négatives;
- en revanche, si $T < T_c$, l’aimantation tend vers une valeur strictement positive $m^*(T)$ lorsque h tend vers 0 par valeurs positives, et vers $-m^*(T)$ lorsque h tend vers 0 par valeurs négatives.

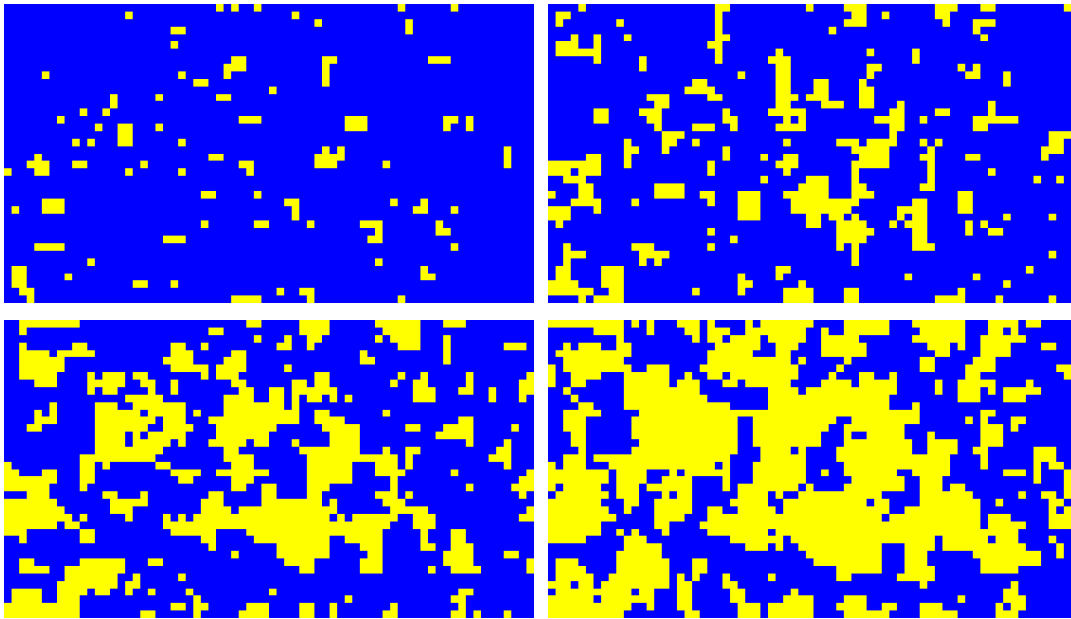


Figure 1.6 – Exemple de simulation d’une dynamique de Glauber. Evolution au cours du temps pour $h = 1$ et $\beta = 0.6$, avec tous les spins initialement égaux à -1 (bleu). Le champ h positif favorise les spins égaux à $+1$ (jaunes).

La quantité $m^*(T)$ s’appelle l’aimantation spontanée du système. Elle tend continûment vers 0 lorsque T tend vers T_c par la gauche.

L’existence de l’aimantation spontanée est importante pour de nombreux dispositifs de stockage de données (disques durs, mémoires flash). Lorsque des données sont sauvegardées sur un tel dispositif, un champ magnétique est appliqué localement afin de créer une aimantation, qui persiste lorsque le champ retombe à zéro. Des données sous forme binaire sont ainsi représentées par des domaines d’aimantation différentes, et cette information peut être récupérée par la suite, tant que l’aimant n’est pas porté à une température dépassant T_c .

Si l’on veut déterminer numériquement l’aimantation, il suffit en principe de calculer la somme (1.4.1). 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 d’une configuration initiale x , 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 échantillonner 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 ?

Répondre à ces deux questions est l’un des objectifs principaux de ce cours.

Rappels sur les chaînes de Markov

Nous rappelons dans ce chapitre quelques notions de base de la théorie des chaînes de Markov, souvent sans démonstration. La plupart des démonstrations peuvent se trouver dans n'importe quel bon cours sur les chaînes de Markov, comme par exemple [1].

2.1 Définitions, notations

Soit \mathcal{X} un ensemble dénombrable, fini ou infini.

Définition 2.1.1: Mesure de probabilité, matrice stochastique

- Une mesure de probabilité ν sur \mathcal{X} est un ensemble $(\nu(x))_{x \in \mathcal{X}}$ de nombres réels positifs ou nuls satisfaisant

$$\sum_{x \in \mathcal{X}} \nu(x) = 1. \quad (2.1.1)$$

- Une *matrice stochastique* sur \mathcal{X} est un ensemble $P = (p_{xy})_{x,y \in \mathcal{X}}$ de nombres réels positifs ou nuls satisfaisant

$$\sum_{y \in \mathcal{X}} p_{xy} = 1 \quad \forall x \in \mathcal{X}. \quad (2.1.2)$$

Remarquons que puisque les $\nu(x)$ sont positifs ou nuls, la condition (2.1.1) implique qu'ils sont nécessairement tous dans l'intervalle $[0, 1]$. Il en va de même pour les p_{xy} .

Définition 2.1.2: Chaîne de Markov

On se donne une matrice stochastique P sur \mathcal{X} , et une mesure de probabilité ν sur \mathcal{X} . Une *chaîne de Markov* (homogène en temps) sur \mathcal{X} , de loi initiale ν et de matrice de transition P , est une suite $(X_n)_{n \geq 0}$ de variables aléatoires à valeurs dans \mathcal{X} , telles que $\mathbb{P}\{X_0 = x\} = \nu(x)$ pour tout $x \in \mathcal{X}$, et satisfaisant la *propriété de Markov*

$$\begin{aligned} \mathbb{P}\{X_n = y | X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}\} &= \mathbb{P}\{X_n = y | X_{n-1} = x_{n-1}\} \\ &= p_{x_{n-1}y} \end{aligned}$$

pour tout $n \geq 1$ et tout choix de $x_0, \dots, x_{n-1}, y \in \mathcal{X}$.

Une conséquence immédiate de cette définition est la suivante.

Proposition 2.1.3: Probabilité de trajectoires et loi de X_n

Soit $(X_n)_{n \geq 0}$ une chaîne de Markov de loi initiale ν et de matrice de transition P . Alors, pour tout $n \geq 0$ et tout choix de $x_0, \dots, x_n \in \mathcal{X}$,

$$\mathbb{P}\{X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\} = \nu(x_0)p_{x_0x_1} \cdots p_{x_{n-1}x_n}. \quad (2.1.3)$$

De plus, pour tout $n \geq 1$ et tout $y \in \mathcal{X}$, on a

$$\mathbb{P}\{X_n = y\} = \sum_{x_0 \in \mathcal{X}} \cdots \sum_{x_{n-1} \in \mathcal{X}} \nu(x_0)p_{x_0x_1} \cdots p_{x_{n-2}x_{n-1}}p_{x_{n-1}y}. \quad (2.1.4)$$

Dans la suite, les notations suivantes vont s'avérer pratiques.

- On écrira $\mathbb{P}^\nu\{X_n = y\}$ au lieu de $\mathbb{P}\{X_n = y\}$ pour insister sur le fait que la loi initiale est ν .
- De manière similaire, on écrira $\mathbb{E}^\nu[X_n]$ pour l'espérance de X_n , partant de la loi ν .
- Soit δ_x la mesure de probabilité sur \mathcal{X} donnée par

$$\delta_x(y) = \begin{cases} 1 & \text{si } y = x, \\ 0 & \text{sinon.} \end{cases}$$

Alors, on écrira souvent $\mathbb{P}^x\{\cdot\}$ et $\mathbb{E}^x[\cdot]$ au lieu de $\mathbb{P}^{\delta_x}\{\cdot\}$ et $\mathbb{E}^{\delta_x}[\cdot]$.

- Il sera pratique de voir les mesures de probabilité sur \mathcal{X} comme des vecteurs ligne. De cette façon, (2.1.4) peut s'écrire

$$\mathbb{P}^\nu\{X_n = y\} = (\nu P^n)_y.$$

Définition 2.1.4: Chaîne de Markov réversible

La chaîne de Markov est dite *réversible* s'il existe une application $\alpha : \mathcal{X} \rightarrow [0, \infty)$, non identiquement nulle, telle que

$$\alpha(x)p_{xy} = \alpha(y)p_{yx} \quad \forall x, y \in \mathcal{X}.$$

Dans ce cas, $\alpha = (\alpha_x)_{x \in \mathcal{X}}$ est appelé un *vecteur réversible*.

Le nom réversible vient de la propriété suivante.

Proposition 2.1.5: Renversement du temps

Supposons la chaîne de Markov réversible, pour un vecteur réversible α qui est une mesure de probabilité. Alors

$$\mathbb{P}^\alpha\{X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\} = \mathbb{P}^\alpha\{X_0 = x_n, X_1 = x_{n-1}, \dots, X_n = x_0\}$$

pour tout $n \in \mathbb{N}$, et tout choix de $x_0, x_1, \dots, x_n \in \mathcal{X}$.

Démonstration. Il suit de (2.1.3) que

$$\begin{aligned} \mathbb{P}^\alpha\{X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\} &= \alpha(x_0)p_{x_0x_1}p_{x_1x_2} \cdots p_{x_{n-1}x_n} \\ &= p_{x_1x_0}\alpha(x_1)p_{x_1x_2} \cdots p_{x_{n-1}x_n} \\ &= \dots \\ &= p_{x_1x_0}p_{x_2x_1} \cdots p_{x_nx_{n-1}}\alpha(x_n) \\ &= \alpha(x_n)p_{x_nx_{n-1}} \cdots p_{x_2x_1}p_{x_1x_0}. \end{aligned}$$

ce qui est bien égal à $\mathbb{P}^\alpha\{X_0 = x_n, X_1 = x_{n-1}, \dots, X_n = x_0\}$. □

2.2 Chaînes de Markov irréductibles

Définition 2.2.1: État accessible, chaîne de Markov irréductible

- On dit qu'un état $y \in \mathcal{X}$ est *accessible* depuis $x \in \mathcal{X}$ s'il existe $n \geq 0$ tel que

$$\mathbb{P}^x\{X_n = y\} > 0.$$

Dans ce cas, on écrira $x \rightsquigarrow y$.

- On dit que les états x et y *communiquent* et on écrit $x \sim y$, si on a à la fois $x \rightsquigarrow y$ et $y \rightsquigarrow x$.
- La chaîne de Markov est *irréductible* si $x \sim y$ pour tout $x, y \in \mathcal{X}$.

On vérifie facilement que la relation \rightsquigarrow est *réflexive* et *transitive* : on a toujours $x \rightsquigarrow x$, et si $x \rightsquigarrow y$ et $y \rightsquigarrow z$, alors on a $x \rightsquigarrow z$.

La relation \sim est réflexive, transitive et *symétrique* : si $x \sim y$, alors $y \sim x$. C'est donc une *relation d'équivalence*. On a donc une partition de \mathcal{X} en *classes d'équivalence* :

$$\mathcal{X} = \bigsqcup_{k \geq 0} \mathcal{X}_k,$$

où \sqcup signifie la réunion disjointe, et $x \sim y$ si et seulement si x et y appartiennent à la même classe. En particulier, la chaîne de Markov est irréductible si et seulement si elle admet une unique classe d'équivalence.

On peut associer à une chaîne de Markov un graphe orienté, dont les sommets sont les éléments de \mathcal{X} , et dont les arêtes sont les couples (x, y) tels que $p_{xy} > 0$ (avec $y \neq x$). Si \mathcal{X} est fini, une manière de montrer que la chaîne de Markov est irréductible est d'exhiber un chemin fermé dans ce graphe, c'est-à-dire une suite $(x_1, \dots, x_m, x_{m+1} = x_1)$, contenant tous les éléments de \mathcal{X} au moins une fois, et telle que $p_{x_i x_{i+1}} > 0$ pour tout $i \in \{1, \dots, m\}$.

Exemple 2.2.2: Marche aléatoire symétrique sur \mathbb{Z}^d

La marche aléatoire symétrique sur \mathbb{Z}^d est irréductible. En effet, pour tout $x, y \in \mathbb{Z}^d$, il existe un chemin reliant x à y . Ce chemin peut être construit en changeant chaque composante de x , par étapes successives, d'une unité à la fois, jusqu'à atteindre y .

Remarque 2.2.3: Classes ouvertes et fermées

Si la chaîne de Markov n'est pas irréductible, alors une classe \mathcal{X}_k est *fermée* si pour tout $x \in \mathcal{X}_k$ et tout $y \notin \mathcal{X}_k$, y n'est pas accessible depuis x . Dans ce cas, la restriction de la chaîne de Markov à \mathcal{X}_k est irréductible. Une classe qui n'est pas fermée est dite *ouverte*.

2.3 Récurrence

Définition 2.3.1: Temps de passage

Soit $x \in \mathcal{X}$. Le *temps de passage* (ou *temps de premier passage*) de la chaîne de Markov en x est la variable aléatoire

$$\tau_x = \inf\{n \geq 1 : X_n = x\},$$

avec la convention $\tau_x = \infty$ si $X_n \neq x$ pour tout $n \geq 1$.

Dans le cas particulier où la mesure initiale est δ_x , τ_x s'appelle également *temps de retour* en x .

Dans la suite, on écrira

$$\mathbb{P}^y\{\tau_x < \infty\} = \lim_{n \rightarrow \infty} \mathbb{P}^x\{\tau_x < n\} = 1 - \mathbb{P}^y\{\tau_x = \infty\}.$$

Attention, par convention la limite lorsque $n \rightarrow \infty$ ne comprend *jamais* le terme $n = \infty$.

Définition 2.3.2: Récurrence et transience

- Un état $x \in \mathcal{X}$ est dit *récurrent* si $\mathbb{P}^x\{\tau_x < \infty\} = 1$.
- Un état non récurrent est dit *transient*.
- La chaîne de Markov est dite *récurrente* si tous ses états sont récurrents, et *transiente* si tous ses états sont transients.

Le critère suivant permet de ramener la question de la récurrence d'une chaîne de Markov à celle d'un petit nombre d'états.

Proposition 2.3.3: Récurrence et communication

Si les états x et y communiquent, alors y est récurrent si et seulement si x est récurrent. Par conséquent,

- si un état d'une classe \mathcal{X}_k est récurrent (respectivement transient), alors tous les états de la classe sont récurrents (respectivement transients); on dit alors que la classe est récurrente (respectivement transiente);
- si la chaîne de Markov est irréductible, et possède un état récurrent (respectivement transient), alors la chaîne de Markov est récurrente (respectivement transiente).

Démonstration partielle. Nous allons montrer que si x et y sont dans la même classe récurrente, alors

$$\mathbb{P}^x\{\tau_y < \infty\} = \mathbb{P}^y\{\tau_x < \infty\} = 1.$$

Soit $A_M = \bigcup_{m=1}^M \{X_m = y\}$ l'événement « la chaîne de Markov visite le site y lors des M premiers pas ». Alors

$$\lim_{M \rightarrow \infty} \mathbb{P}^x(A_M) = \sum_{m=1}^{\infty} \mathbb{P}^y\{\tau_y = m\} = 1. \quad (2.3.1)$$

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

$$\begin{aligned} \mathbb{P}^y(A_M \cap \{X_{n_0} = x\}) &= \sum_{n=1}^{M-n_0} \mathbb{P}^y\{X_{n_0} = x, \tau_y = n_0 + n\} \\ &= \sum_{n=1}^{M-n_0} \mathbb{P}^y\{X_{n_0} = x, X_1 \neq y, \dots, X_{n_0} \neq y\} \mathbb{P}^x\{\tau_y = n\} \\ &\leq \mathbb{P}^y\{X_{n_0} = x\} \sum_{n=1}^{M-n_0} \mathbb{P}^x\{\tau_y = n\}. \end{aligned}$$

La première égalité suit du fait que la chaîne de Markov ne peut pas retourner en y avant n_0 et visiter x 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 $\mathbb{P}^y\{X_{n_0} = x\}$ en vertu de (2.3.1). Il vient donc

$$\mathbb{P}^y\{X_{n_0} = x\} \leq \mathbb{P}^y\{X_{n_0} = x\} \mathbb{P}^x\{\tau_y < \infty\}.$$

Comme $\mathbb{P}^y\{X_{n_0} = x\} \neq 0$ et $\mathbb{P}^x\{\tau_y < \infty\} \leq 1$, on a nécessairement $\mathbb{P}^x\{\tau_y < \infty\} = 1$. \square

Pour montrer qu'un état est récurrent, le critère suivant est souvent utile en pratique.

Théorème 2.3.4: Critère de récurrence

Un état $x \in \mathcal{X}$ est récurrent si et seulement si

$$\sum_{n=0}^{\infty} \mathbb{P}^x\{X_n = x\} = \infty .$$

La démonstration de ce résultat est basée sur la relation suivante.

Proposition 2.3.5: Équation de renouvellement

Pour tout $x, y \in \mathcal{X}$ et tout temps $n \in \mathbb{N}$ on a la relation

$$\mathbb{P}^x\{X_n = y\} = \sum_{m=1}^n \mathbb{P}^x\{\tau_y = m\} \mathbb{P}^y\{X_{n-m} = y\} . \quad (2.3.2)$$

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

$$\begin{aligned} \mathbb{P}^x\{X_n = y\} &= \sum_{m=1}^n \mathbb{P}^x\{X_1 \neq y, \dots, X_{m-1} \neq y, X_m = y, X_n = y\} \\ &= \sum_{m=1}^n \underbrace{\mathbb{P}^x\{X_n = y | X_1 \neq y, \dots, X_{m-1} \neq y, X_m = y\}}_{=\mathbb{P}^x\{X_n=y|X_m=y\}=\mathbb{P}^y\{X_{n-m}=y\}} \underbrace{\mathbb{P}^x\{X_1 \neq y, \dots, X_{m-1} \neq y, X_m = y\}}_{=\mathbb{P}^x\{\tau_y=m\}} , \end{aligned}$$

où nous avons utilisé la propriété des incréments indépendants. □

Démonstration du Théorème 2.3.4.

⇒: L'équation de renouvellement (2.3.2) permet d'écrire

$$\begin{aligned} S := \sum_{n=0}^{\infty} \mathbb{P}^x\{X_n = x\} &= 1 + \sum_{n=1}^{\infty} \mathbb{P}^x\{X_n = x\} \\ &= 1 + \sum_{n=1}^{\infty} \sum_{m=1}^n \mathbb{P}^x\{\tau_x = m\} \mathbb{P}^x\{X_{n-m} = x\} \\ &= 1 + \sum_{m=1}^{\infty} \mathbb{P}^x\{\tau_x = m\} \sum_{n=m}^{\infty} \mathbb{P}^x\{X_{n-m} = x\} \\ &= 1 + \underbrace{\sum_{m=1}^{\infty} \mathbb{P}^x\{\tau_x = m\}}_{=1} \sum_{n=0}^{\infty} \mathbb{P}^x\{X_n = x\} = 1 + S . \end{aligned} \quad (2.3.3)$$

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

⇐: 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

$$\begin{aligned} \psi(s) &= \sum_{n=0}^{\infty} \mathbb{P}^x\{X_n = x\} s^n , \\ \phi(s) &= \sum_{n=1}^{\infty} \mathbb{P}^x\{\tau_x = n\} s^n = \mathbb{E}^x[s^{\tau_x}] . \end{aligned}$$

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 (2.3.3) ci-dessus donne alors

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

d'où

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

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

$$\sum_{n=0}^{\infty} \mathbb{P}^x\{X_n = x\} = \lim_{s \nearrow 1} \psi(s) = \frac{1}{1 - \phi(1)} < \infty,$$

ce qui conclut la démonstration. □

2.4 Récurrence positive, probabilité invariante

Définition 2.4.1: Récurrence positive

Un état récurrent $x \in \mathcal{X}$ est dit *récurrent positif* si

$$\mathbb{E}^x[\tau_x] < \infty.$$

Sinon, l'état est appelé *récurrent nul*. Une chaîne de Markov récurrente est dite *récurrente positive* si tous ses états sont récurrents positifs, et *récurrente nulle* sinon.

La récurrence positive est à nouveau une propriété de classe.

Proposition 2.4.2: Récurrence positive et communication

Si les états x et y communiquent, alors y est récurrent positif si et seulement si x est récurrent positif. En particulier, si la chaîne de Markov est irréductible et admet un état récurrent positif, alors la chaîne de Markov est récurrente positive.

Remarque 2.4.3: Cas d'un \mathcal{X} fini

Si \mathcal{X} est fini et la chaîne de Markov est irréductible, alors elle est nécessairement récurrente positive. En effet, l'irréductibilité montre que pour tout $x \in \mathcal{X}$, on peut trouver un entier fini m tel que

$$p = \max_{y \in \mathcal{X}} \mathbb{P}^y\{\tau_x > m\} < 1.$$

La propriété de Markov implique alors que pour tout $k \geq 1$, on a

$$\mathbb{P}^x\{\tau_x > km\} \leq p^k.$$

La décroissance exponentielle des queues de la loi de τ_x implique que $\mathbb{E}^x[\tau_x] < \infty$.

Voici un résultat de récurrence/transience très classique, qui se démontre à l'aide du Théorème 4.2.5.

Théorème 2.4.4: Récurrence/transience de marches aléatoires symétriques

La marche aléatoire symétrique sur \mathbb{Z}^d est récurrente nulle si $d \in \{1, 2\}$ et transiente si $d \geq 3$.

L'intérêt principal de la définition de récurrence positive est lié à l'existence de probabilités invariantes.

Définition 2.4.5: Mesures et probabilités invariantes

Une mesure sur \mathcal{X} (c'est-à-dire une application $\mu : \mathcal{X} \rightarrow \mathbb{R}_+ = [0, \infty)$) est dite *invariante* si

$$\sum_{x \in \mathcal{X}} \mu(x) p_{xy} = \mu(y) \quad \forall y \in \mathcal{X}. \quad (2.4.1)$$

Si μ est une mesure de probabilité, on dit que c'est une *probabilité invariante*. On la notera alors souvent π .

La relation (2.4.1) s'écrit matriciellement

$$\mu P = \mu,$$

c'est-à-dire que le vecteur ligne μ est vecteur propre à gauche de P , pour la valeur propre 1. Si π est une probabilité invariante, alors

$$\mathbb{P}^\pi\{X_n = x\} = \pi(x) \quad \forall x \in \mathcal{X}, \forall n \geq 0.$$

Exemple 2.4.6

Soit μ une mesure uniforme sur \mathbb{Z}^d , c'est-à-dire qu'il existe une constante $c \in \mathbb{R}$ telle que $\mu(x) = c$ pour tout $x \in \mathbb{Z}^d$. Alors μ est une mesure invariante pour la marche aléatoire symétrique sur \mathbb{Z}^d . Toutefois, μ n'est pas une mesure de probabilité, car on ne peut pas la normaliser (la somme des $\mu(x)$ vaut soit 0, si $c = 0$, soit est infinie, si $c \neq 0$).

Exemple 2.4.7

On vérifie que la loi binomiale de paramètres n et $\frac{1}{2}$ est une probabilité invariante du modèle d'Ehrenfest à n boules (voir Exercice 2.6.1).

Le lien entre récurrence positive et probabilité invariante est mis en évidence par le résultat suivant.

Théorème 2.4.8: Récurrence positive et probabilité invariante

Soit $(X_n)_{n \geq 0}$ une chaîne de Markov irréductible sur \mathcal{X} . Alors les conditions suivantes sont équivalentes :

1. La chaîne de Markov admet une probabilité invariante.
2. La chaîne de Markov admet un état récurrent positif.
3. Tous les états $x \in \mathcal{X}$ sont récurrents positifs.

De plus, si ces propriétés sont vérifiées, alors la probabilité invariante est unique, et satisfait

$$\pi(x) = \frac{1}{\mathbb{E}^x[\tau_x]} \quad \forall x \in \mathcal{X}. \quad (2.4.2)$$

Une manière de démontrer ce résultat est de fixer un état $z \in \mathcal{X}$, et de considérer la mesure $\gamma^{(z)}$, définie par

$$\gamma^{(z)}(x) = \mathbb{E}^z \left[\sum_{n=1}^{\tau_z} \mathbb{1}_{X_n=x} \right], \quad (2.4.3)$$

qui mesure le nombre moyen de passages en x entre deux passages en z . On a alors les propriétés suivantes.

Proposition 2.4.9

Supposons la chaîne de Markov irréductible et récurrente. Alors on a pour tout $z \in \mathcal{X}$:

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

Démonstration.

1. Évident, puisque τ_z est fini presque sûrement, $X_{\tau_z} = z$ et $X_n \neq z$ pour $1 \leq n < \tau_z$.
2. Nous avons

$$\begin{aligned} \gamma^{(z)}(x) &= \mathbb{E}^z \left[\sum_{n=1}^{\infty} \mathbb{1}_{\{X_n=x, n \leq \tau_z\}} \right] = \sum_{n=1}^{\infty} \mathbb{P}^z \{X_n = x, n \leq \tau_z\} \\ &= \sum_{y \in \mathcal{X}} \sum_{n=1}^{\infty} \mathbb{P}^z \{X_{n-1} = y, n \leq \tau_z\} p_{yx} \\ &= \sum_{y \in \mathcal{X}} p_{yx} \sum_{m=0}^{\infty} \mathbb{P}^z \{X_m = y, m \leq \tau_z - 1\}. \end{aligned}$$

Or la seconde somme dans cette expression peut s'écrire

$$\mathbb{E}^z \left[\sum_{m=0}^{\tau_z-1} \mathbb{1}_{\{X_m=y\}} \right] = \mathbb{E}^z \left[\sum_{m=1}^{\tau_z} \mathbb{1}_{\{X_m=y\}} \right] = \gamma^{(z)}(y),$$

vu que $\mathbb{P}^z \{X_0 = y\} = \delta_{zy} = \mathbb{P}^z \{X_{\tau_z} = y\}$. Ceci prouve l'invariance de la mesure $\gamma^{(z)}$.

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

$$\gamma^{(z)}(x) = \sum_{y \in \mathcal{X}} \gamma^{(z)}(y) \mathbb{P}^y \{X_n = x\}.$$

En particulier, $1 = \gamma^{(z)}(z) \geq \gamma^{(z)}(y) \mathbb{P}^y \{X_n = z\}$ pour tout y . Comme par irréductibilité, il existe un n tel que $\mathbb{P}^y \{X_n = z\} > 0$, on en déduit que $\gamma^{(z)}(y) < \infty$ pour tout y . D'autre part, on a aussi $\gamma^{(z)}(x) \geq \mathbb{P}^z \{X_n = x\}$, qui est strictement positif pour au moins un n .

4. Soit λ une mesure invariante telle que $\lambda(z) = 1$. Alors pour tout y on a

$$\lambda(y) = \sum_{x \neq z} \lambda(x) p_{xy} + p_{zy} \geq p_{zy}.$$

Il vient alors, en minorant $\lambda(x)$ par p_{zx} dans l'expression ci-dessus,

$$\begin{aligned} \lambda(y) &\geq \sum_{x \neq z} p_{zx} p_{xy} + p_{zy} \\ &= \mathbb{P}^z \{X_2 = y, \tau_z \geq 2\} + \mathbb{P}^z \{X_1 = y, \tau_z \geq 1\} \end{aligned}$$

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

$$\lambda(y) \geq \sum_{m=1}^{n+1} \mathbb{P}^z\{X_m = y, \tau_z \geq m\} = \mathbb{E}^z \left[\sum_{m=1}^{(n+1) \wedge \tau_z} \mathbb{1}_{\{X_m=y\}} \right].$$

Lorsque n tend vers l'infini, le membre de droite tend vers $\gamma^{(z)}(y)$. On a donc $\lambda(y) \geq \gamma^{(z)}(y)$ pour tout y . Par conséquent, $\mu = \lambda - \gamma^{(z)}$ est une mesure invariante, satisfaisant $\mu(z) = 0$. Comme $\mu(z) = \sum_y \mu(y) \mathbb{P}^y\{X_n = z\}$ pour tout n , l'irréductibilité implique $\mu(y) = 0 \forall y$, donc nécessairement $\lambda = \gamma^{(z)}$. \square

Démonstration du Théorème 2.4.8.

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

$$\sum_{y \in \mathcal{X}} \gamma^{(z)}(y) = \mathbb{E}^z \left[\underbrace{\sum_{n=1}^{\tau_z} \sum_{y \in \mathcal{X}} \mathbb{1}_{\{X_n=y\}}}_{=1} \right] = \mathbb{E}^z[\tau_z] = \mu(z) < \infty.$$

Par conséquent, la mesure π définie par $\pi(y) = \gamma^{(z)}(y)/\mu(z)$ est une probabilité invariante.

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

$$\mathbb{E}^z[\tau_z] = \sum_{y \in \mathcal{X}} \hat{\gamma}(y) = \frac{1}{\pi(z)} \sum_{y \in \mathcal{X}} \pi(y) = \frac{1}{\pi(z)} < \infty. \quad (2.4.4)$$

$3 \Rightarrow 2$: Évident.

Dans ce cas, l'unicité de la mesure suit de celle de $\gamma^{(z)}$, et la relation (2.4.2) suit de (2.4.4). \square

Dans le cas particulier d'une chaîne de Markov réversible, la probabilité invariante peut être déduite immédiatement d'un vecteur réversible.

Proposition 2.4.10: Probabilités invariante d'une chaîne de Markov réversible

Soit $(X_n)_{n \geq 0}$ une chaîne de Markov réversible, de vecteur réversible α . Alors, si

$$\mathcal{N} = \sum_{x \in \mathcal{X}} \alpha(x) < \infty,$$

la chaîne de Markov admet une probabilité invariante, donnée par

$$\pi(x) = \frac{1}{\mathcal{N}} \alpha(x) \quad \forall x \in \mathcal{X}.$$

Démonstration. Pour tout $x \in \mathcal{X}$, on a

$$\sum_{y \in \mathcal{X}} \pi(y) p_{yx} = \frac{1}{\mathcal{N}} \sum_{y \in \mathcal{X}} \alpha(y) p_{yx} = \frac{1}{\mathcal{N}} \sum_{y \in \mathcal{X}} p_{xy} \alpha(x) = \frac{1}{\mathcal{N}} \alpha(x) = \pi(x).$$

De plus, π est bien une mesure de probabilité, puisque la somme des $\pi(x)$ vaut 1. \square

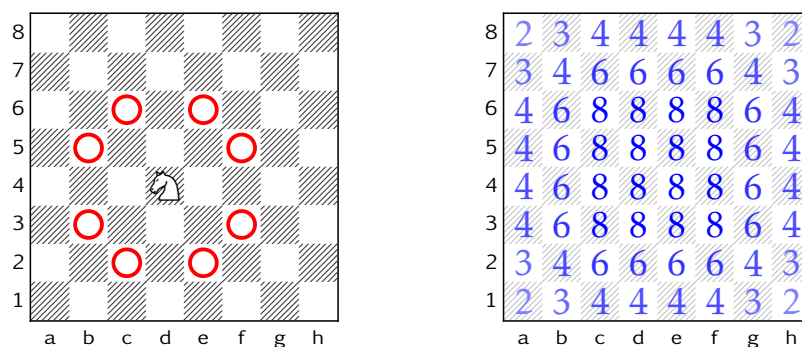


Figure 2.1 – Mouvements permis du cavalier sur l'échiquier. Nombre de mouvements possibles à partir de chaque case.

Exemple 2.4.11: Le cavalier fou

Un cavalier se déplace sur un échiquier standard (de 64 cases), en choisissant à chaque pas l'un des mouvements permis par les règles du jeu des échecs, uniformément au hasard (Figure 2.1). La position du cavalier est décrite par une chaîne de Markov sur l'ensemble \mathcal{X} des 64 cases de l'échiquier. Si $\alpha(x)$ désigne le nombre de mouvements permis en partant de la case x , alors les probabilités de transition sont données par

$$p_{xy} = \begin{cases} \frac{1}{\alpha(x)} & \text{si le mouvement de } x \text{ vers } y \text{ est permis,} \\ 0 & \text{sinon.} \end{cases}$$

On vérifie que α est un vecteur réversible, et que $\mathcal{N} = \sum_{x \in \mathcal{X}} \alpha(x) = 336$ (voir Figure 2.1). La chaîne de Markov est donc réversible, et admet la probabilité invariante π donnée par

$$\pi(x) = \frac{\alpha(x)}{336}.$$

Le Théorème 2.4.8 permet alors de calculer le temps de récurrence moyen vers n'importe quel état. Celui-ci vaut

$$\mathbb{E}^x[\tau_x] = \frac{1}{\pi(x)} = \frac{336}{\alpha(x)}.$$

2.5 Apériodicité, convergence vers la probabilité invariante

Définition 2.5.1: Période

La *période* d'un état $x \in \mathcal{X}$ est le nombre

$$d_x = \text{pgcd}\{n \geq 1 : \mathbb{P}^x\{X_n = x\} > 0\}.$$

Si $d_x = 1$, alors on dit que x est *apériodique*. Si tout $x \in \mathcal{X}$ est apériodique, on dit que la chaîne de Markov est apériodique.

La période est à nouveau un propriété de classe.

Proposition 2.5.2: Période et communication

Si $x \sim y$, alors $d_x = d_y$. Par conséquent, si la chaîne de Markov est irréductible et admet un état apériodique, alors la chaîne de Markov est apériodique.

Exemple 2.5.3: Marche aléatoire symétrique sur \mathbb{Z}^d

Pour la marche aléatoire symétrique sur \mathbb{Z}^d , la période de l'état 0 vaut $d_0 = 2$. En effet, partant de 0, la marche ne peut retourner en 0 qu'au temps pairs. Par conséquent, la marche n'est pas apériodique (tous les états sont de période 2).

L'importance de la notion d'apériodicité vient du résultat crucial suivant.

Théorème 2.5.4: Convergence vers la probabilité invariante

Soit $(X_n)_{n \geq 0}$ une chaîne de Markov irréductible, apériodique et récurrente positive, et soit π son unique probabilité invariante. Alors pour toute loi initiale ν et tout $x \in \mathcal{X}$, on a

$$\lim_{n \rightarrow \infty} \mathbb{P}^\nu \{X_n = x\} = \pi(x).$$

Nous allons esquisser l'idée principale d'une démonstration de ce théorème, due à Wolfgang Doeblin. Considérons deux chaînes de Markov indépendantes, $(X_n)_{n \geq 0}$ et $(Y_n)_{n \geq 0}$, ayant les deux la même matrice de transition P , mais la première partant de ν , alors que la seconde part de π . Le couple (X_n, Y_n) est une chaîne de Markov sur $\mathcal{X} \times \mathcal{X}$, de probabilités de transition

$$p_{(x,y),(u,v)}^\star = p_{xu} p_{yv},$$

et de loi initiale $\rho = \nu \otimes \pi$, définie par

$$\rho(x, y) = \nu(x)\pi(y).$$

On montre alors (à l'aide du théorème de Bézout) que cette chaîne de Markov est encore irréductible et apériodique. Comme elle admet la probabilité invariante $\pi \otimes \pi$, elle est aussi récurrente positive. Soit alors

$$\tau_\Delta = \inf\{n \geq 0 : X_n = Y_n\} \quad (2.5.1)$$

le temps de passage sur la *diagonale* $\Delta = \{(x, x) : x \in \mathcal{X}\}$. On déduit de la récurrence positive que τ_Δ est presque sûrement fini. Introduisons alors le processus $(Z_n)_{n \geq 0}$, défini par

$$Z_n = \begin{cases} X_n & \text{si } n < \tau_\Delta, \\ Y_n & \text{si } n \geq \tau_\Delta. \end{cases}$$

Il suit de l'expression (2.1.3) de la probabilité d'une trajectoire que $(Z_n)_{n \geq 0}$ est une chaîne de Markov de loi initiale ν et de matrice de transition P . Par conséquent, Z_n est égal en loi à X_n pour tout $n \geq 0$. Ceci implique que pour tout $n \in \mathbb{N}$ et tout $x \in \mathcal{X}$, on a

$$\mathbb{P}^\rho \{X_n = x, \tau_\Delta \leq n\} = \mathbb{P}^\rho \{Z_n = x, \tau_\Delta \leq n\} = \mathbb{P}^\rho \{Y_n = x, \tau_\Delta \leq n\}. \quad (2.5.2)$$

La première égalité suit de l'égalité en loi de X_n et Y_n , alors que la seconde vient du fait que $Z_n = Y_n$ pour $\tau_\Delta \leq n$.

On observe maintenant que pour tout $n \in \mathbb{N}$ et tout $x \in \mathcal{X}$, on a

$$\begin{aligned} \mathbb{P}^\nu \{X_n = x\} &= \mathbb{P}^\rho \{X_n = x, \tau_\Delta \leq n\} + \mathbb{P}^\rho \{X_n = x, \tau_\Delta > n\}, \\ \pi(x) &= \mathbb{P}^\pi \{Y_n = x\} = \mathbb{P}^\rho \{Y_n = x, \tau_\Delta \leq n\} + \mathbb{P}^\rho \{Y_n = x, \tau_\Delta > n\}. \end{aligned}$$

En prenant la différence et en utilisant (2.5.2), on obtient

$$\left| \mathbb{P}^\nu \{X_n = x\} - \pi(x) \right| \leq \left| \mathbb{P}^\rho \{X_n = x, \tau_\Delta > n\} - \mathbb{P}^\rho \{Y_n = x, \tau_\Delta > n\} \right| \leq 2\mathbb{P}^\rho \{\tau_\Delta > n\}.$$

La chaîne de Markov $(X_n, Y_n)_{n \geq 0}$ étant récurrente positive, cette quantité tend vers 0 lorsque n tend vers l'infini, ce qui prouve le théorème. En fait, on a même obtenu un peu mieux : pour tout $n \geq 0$, on a

$$\sum_{x \in \mathcal{X}} |\mathbb{P}^v\{X_n = x\} - \pi(x)| \leq 2\mathbb{P}^\rho\{\tau_\Delta > n\}. \quad (2.5.3)$$

Si on arrive à majorer la probabilité $\mathbb{P}^\rho\{\tau_\Delta > n\}$, on obtient donc une majoration d'une distance entre la loi de X_n et π (il s'agit d'une distance du type ℓ^1). C'est un exemple de ce qu'on appelle un *argument de couplage*.

2.6 Exercices

Exercice 2.6.1

On considère le modèle des urnes d'Ehrenfest à N boules, c'est-à-dire la chaîne de Markov sur l'ensemble $\mathcal{X} = \{0, 1, \dots, N\}$ de probabilités de transition

$$p_{xy} = \begin{cases} \frac{x}{N} & \text{si } y = x - 1, \\ 1 - \frac{x}{N} & \text{si } y = x + 1, \\ 0 & \text{sinon.} \end{cases}$$

1. Montrer que cette chaîne de Markov est irréductible. Est-elle apériodique ?
2. Montrer que la distribution de probabilité invariante de cette chaîne de Markov suit une loi binômiale, dont on précisera les paramètres.

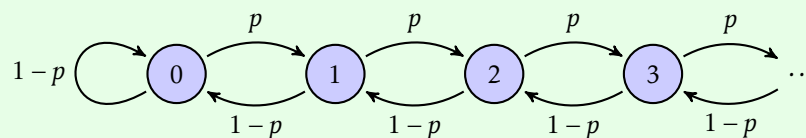
Exercice 2.6.2

Soit $\mathcal{G} = (V, E)$ un graphe non orienté connexe fini. Soit $(X_n)_{n \geq 0}$ 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 probabilité invariante de la chaîne de Markov.

Exercice 2.6.3

Soit $p \in [0, 1]$. On considère la chaîne de Markov suivante sur $\mathcal{X} = \mathbb{N}$:



1. Pour quelles valeurs de p la chaîne de Markov est-elle irréductible?
On suppose dans la suite que p est tel que la chaîne de Markov soit irréductible.
2. La chaîne de Markov est-elle apériodique ?
3. On suppose que la chaîne de Markov est réversible, et soit α un vecteur réversible. Ecrire une relation de récurrence pour les composantes de α , et en déduire α_n en fonction de α_0 .
4. Pour quelles valeurs de p la chaîne de Markov admet-elle une probabilité invariante π ?

Déterminer π pour ces valeurs de p .

5. Pour quelles valeurs de p la chaîne de Markov est-elle récurrente? Récurrente positive?
6. Déterminer le temps de récurrence moyen $\mathbb{E}^0[\tau_0]$.
7. Calculer la position moyenne $\mathbb{E}^\pi[X_n]$ pour les valeurs de p telles que π existe.

Exercice 2.6.4

On considère une marche aléatoire unidimensionnelle symétrique sur l'ensemble $\mathcal{X} = \{0, 1, \dots, N\}$ avec conditions aux bords absorbantes, c'est-à-dire que l'on suppose que $p_{00} = p_{NN} = 1$. Soit

$$\tau = \tau_0 \wedge \tau_N = \inf\{n \geq 0 : X_n \in \{0, N\}\}$$

le temps d'absorption, et soit

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

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

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

Une fonction $f : \mathbb{Z} \supset A \rightarrow \mathbb{R}$ telle que $f(x) = \frac{1}{2}[f(x-1) + f(x+1)]$ pour tout $x \in A$ 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 $A = \{a, a+1, \dots, b-1, b\}$, 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 harmoniques 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 $f(x) = cx + h$ est harmonique.
7. En utilisant les points 1., 2., 5. et 6., déterminer la fonction p .

Exercice 2.6.5

On considère une marche aléatoire symétrique sur $\mathcal{X} = \{0, 1, \dots, N\}$, 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

$$\tau = \inf\{n \geq 0 : X_n \in \{0, N\}\}$$

le temps d'absorption. Par convention, $\tau = 0$ si $X_0 \in \{0, N\}$. Pour $\lambda \in \mathbb{R}$ et $i \in \mathcal{X}$ on pose

$$f(x, \lambda) = \mathbb{E}^x[e^{-\lambda\tau} \mathbb{1}_{\{X_\tau=N\}}] = \begin{cases} \mathbb{E}^x[e^{-\lambda\tau}] & \text{si } X_\tau = N, \\ 0 & \text{sinon.} \end{cases}$$

1. Que valent $f(0, \lambda)$ et $f(N, \lambda)$?
2. Montrer que pour tout $x \in \{1, \dots, N-1\}$,

$$\mathbb{P}^x\{\tau = n\} = \frac{1}{2}[\mathbb{P}^{x-1}\{\tau = n-1\} + \mathbb{P}^{x+1}\{\tau = n-1\}].$$

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

$$f(x, \lambda) = \frac{1}{2} e^{-\lambda} [f(x-1, \lambda) + f(x+1, \lambda)].$$

4. Trouver une relation entre c et λ telle que l'équation ci-dessus pour f admette des solutions de la forme $f(x, \lambda) = e^{cx}$. Montrer à l'aide d'un développement limité que

$$c^2 = 2\lambda + \mathcal{O}(\lambda^2).$$

5. Déterminer des constantes a et b telles que

$$\mathbb{E}^x[e^{-\lambda\tau} \mathbb{1}_{\{X_\tau=N\}}] = a e^{cx} + b e^{-cx}.$$

6. Effectuer un développement limité au premier ordre en λ de l'égalité ci-dessus. En déduire

$$\mathbb{P}^x\{X_\tau = N\}.$$

7. Calculer

$$\mathbb{E}^x[\tau \mathbb{1}_{\{X_\tau=N\}}].$$

8. Sans faire les calculs, indiquer comment procéder pour déterminer la variance de la variable aléatoire $\tau \mathbb{1}_{\{X_\tau=N\}}$ et l'espérance et la variance de τ .

On rappelle les développements limités suivants:

$$\cosh(x) = \frac{e^x + e^{-x}}{2} = 1 + \frac{1}{2!}x^2 + \mathcal{O}(x^4),$$

$$\sinh(x) = \frac{e^x - e^{-x}}{2} = x + \frac{1}{3!}x^3 + \mathcal{O}(x^5).$$

Théorie spectrale et vitesse de convergence

Dans ce chapitre et le suivant, nous allons considérer des chaînes de Markov $(X_n)_{n \geq 0}$ irréductibles, récurrentes positives et apériodiques sur un ensemble dénombrable \mathcal{X} . Soit $f : \mathcal{X} \rightarrow \mathbb{R}$ une fonction bornée, et soit π la probabilité invariante de la chaîne de Markov. Le but est d'estimer la quantité

$$\mathbb{E}^\pi[f] = \sum_{x \in \mathcal{X}} \pi(x) f(x).$$

Nous savons par le Théorème 2.5.4 que l'on a

$$\mathbb{E}^\pi[f] = \lim_{n \rightarrow \infty} \sum_{x \in \mathcal{X}} \mathbb{P}^\nu\{X_n = x\} f(x) = \lim_{n \rightarrow \infty} \mathbb{E}^\nu[f(X_n)],$$

pour toute loi initiale ν . Notre but est maintenant de majorer l'erreur

$$|\mathbb{E}^\nu[f(X_n)] - \mathbb{E}^\pi[f]|. \tag{3.0.1}$$

Une première manière de le faire est la suivante.

Lemme 3.0.1: Couplage et vitesse de convergence

Si la chaîne de Markov est apériodique, alors

$$|\mathbb{E}^\nu[f(X_n)] - \mathbb{E}^\pi[f]| \leq 2 \mathbb{P}^{\nu \otimes \pi}\{\tau_\Delta > n\} \sup_{x \in \mathcal{X}} |f(x)|,$$

où τ_Δ est défini dans (2.5.1).

Démonstration. On a

$$\mathbb{E}^\nu[f(X_n)] - \mathbb{E}^\pi[f] = \sum_{x \in \mathcal{X}} [\mathbb{P}^\nu\{X_n = x\} - \pi(x)] f(x).$$

Le résultat suit donc de (2.5.3). □

Si l'on arrive à contrôler $\mathbb{P}^\rho\{\tau_\Delta > n\}$, on obtient donc la majoration souhaitée. Toutefois, cela n'est pas toujours possible, et on doit alors avoir recours à d'autres approches. Dans ce chapitre, nous allons discuter comment l'erreur (3.0.1) dépend de quantités liées aux valeurs propres et vecteurs propres de la matrice de transition P . Une autre approche, plus robuste, basée sur les fonctions de Lyapounov, sera discutée dans le chapitre suivant.

3.1 Quelques exemples simples

Exemple 3.1.1

Considérons la matrice stochastique

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

La chaîne de Markov sur $\mathcal{X} = \{1, 2\}$ associée est irréductible, récurrente positive, mais pas apériodique : sa période est égale à 2. En fait, on a

$$P^n = \begin{cases} P & \text{si } n \text{ est impair,} \\ \mathbb{1} & \text{si } n \text{ est pair,} \end{cases}$$

où $\mathbb{1}$ dénote la matrice identité. Par conséquent,

$$\mathbb{E}^\nu[f(X_n)] = \nu P^n f = \begin{cases} \nu(1)f(2) + \nu(2)f(1) & \text{si } n \text{ est impair,} \\ \nu(1)f(1) + \nu(2)f(2) & \text{si } n \text{ est pair,} \end{cases}$$

D'un autre côté, la chaîne de Markov étant récurrente positive, elle admet une unique probabilité invariante π , satisfaisant $\pi P = \pi$. On trouve facilement que $\pi = (\frac{1}{2}, \frac{1}{2})$, ce qui implique

$$\mathbb{E}^\pi[f(X_n)] = \pi f = \frac{1}{2}(f(1) + f(2)).$$

On s'aperçoit que si $\nu \neq \pi$, alors $\mathbb{E}^\nu[f]$ ne converge pas vers $\mathbb{E}^\pi[f]$, sauf dans le cas particulier $f(1) = f(2)$.

Les valeurs propres de P sont 1 et -1 . Des vecteurs propres à gauche associés sont π et $(1, -1)$. La valeur propre -1 est associée au fait que la chaîne de Markov est 2-périodique.

Exemple 3.1.2

On peut facilement généraliser cet exemple à des périodes supérieures. Par exemple, la matrice stochastique

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

satisfait $P^3 = \mathbb{1}$. Ses valeurs propres sont les trois racines cubiques de 1, à savoir 1 et $e^{\pm i2\pi/3}$. La chaîne de Markov associée est irréductible, récurrente positive, et de période 3. Elle admet l'unique probabilité invariante $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. À nouveau, si $\nu \neq \pi$, alors $\mathbb{E}^\nu[f]$ ne converge pas vers $\mathbb{E}^\pi[f]$, sauf dans le cas particulier où f est constante.

Exemple 3.1.3

Par contraste, considérons la matrice stochastique

$$P = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix}.$$

La chaîne de Markov associée est irréductible, récurrente positive, et apériodique (car, par

exemple, on a $\mathbb{P}^1\{X_1 = 1\} = \frac{1}{3} > 0$ et $\mathbb{P}^1\{X_2 = 1\} = \frac{5}{9} > 0$). Les valeurs propres de P sont $\lambda_0 = 1$ et $\lambda_1 = -\frac{1}{3}$. Une manière de calculer P^n est d'utiliser la *décomposition de Dunford* (que nous rappellerons à la section 3.4)

$$P = \lambda_0 \Pi_0 + \lambda_1 \Pi_1, \quad \Pi_0 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad \Pi_1 = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Les matrices Π_0 et Π_1 sont des *projecteurs* : elles satisfont $\Pi_0^2 = \Pi_0$, et $\Pi_1^2 = \Pi_1$. Elles sont obtenues chacune en multipliant un vecteur propre à droite et un vecteur propre à gauche de P , proprement normalisés. De plus, on vérifie que $\Pi_0 \Pi_1 = \Pi_1 \Pi_0 = 0$. Ceci implique, par la formule du binôme de Newton, que

$$P^n = \lambda_0^n \Pi_0 + \lambda_1^n \Pi_1 = \Pi_0 + \left(-\frac{1}{3}\right)^n \Pi_1.$$

Par conséquent, nous avons

$$\nu P^n f = \frac{1}{2}(f(1) + f(2)) + \frac{1}{2} \left(-\frac{1}{3}\right)^n (\nu(1) - \nu(2))(f(1) - f(2)).$$

Comme par ailleurs, $\pi = (\frac{1}{2}, \frac{1}{2})$, on a

$$\mathbb{E}^\pi[f] = \pi f = \frac{1}{2}(f(1) + f(2)).$$

Par conséquent, $\mathbb{E}^\nu[f(X_n)]$ converge exponentiellement vite vers $\mathbb{E}^\pi[f]$, avec une différence d'ordre 3^{-n} .

Ces exemples suggèrent que

- si la chaîne de Markov est périodique, alors P admet plusieurs valeurs propres différentes de module 1, P^n ne converge pas lorsque $n \rightarrow \infty$, et $\mathbb{E}^\nu[f(X_n)]$ ne converge pas vers $\mathbb{E}^\pi[f]$ si $\nu \neq \pi$, sauf pour des f très particuliers;
- si la chaîne de Markov est apériodique, alors P admet 1 comme valeur propre simple, toutes les autres valeurs propres de P sont strictement inférieures à 1 en module, et $\mathbb{E}^\nu[f(X_n)]$ converge vers $\mathbb{E}^\pi[f]$ si $\nu \neq \pi$.

Nous allons voir dans les sections suivantes que ceci est effectivement le cas.

3.2 Normes de vecteurs et de matrices

Soit P la matrice de transition d'une chaîne de Markov irréductible et récurrente positive. Nous savons que P admet la valeur propre $\lambda_0 = 1$. Un vecteur propre à gauche associé est π , alors qu'un vecteur propre à droite est le vecteur

$$\mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

En effet, la propriété (2.1.2) d'une matrice stochastique équivaut à $P\mathbf{1} = \mathbf{1}$.

Dans la suite, il sera naturel de travailler avec les normes suivantes.

Définition 3.2.1: Normes de vecteurs

La norme ℓ^1 d'un vecteur ligne μ est définie par

$$\|\mu\|_1 = \sum_{x \in \mathcal{X}} |\mu(x)|.$$

La norme ℓ^∞ (ou *norme sup*) d'un vecteur colonne est définie par

$$\|v\|_\infty = \sup_{x \in \mathcal{X}} |v(x)|.$$

Dans la suite, nous utiliserons souvent la majoration élémentaire

$$|\mu v| = \left| \sum_{x \in \mathcal{X}} \mu(x)v(x) \right| \leq \sum_{x \in \mathcal{X}} |\mu(x)v(x)| \leq \|\mu\|_1 \|v\|_\infty. \quad (3.2.1)$$

Lemme 3.2.2: Normes et matrice stochastique

Pour une matrice stochastique P , et tout vecteur ligne μ et vecteur colonne v de dimension adéquate, on a

$$\|Pv\|_\infty \leq \|v\|_\infty \quad \text{et} \quad \|\mu P\|_1 \leq \|\mu\|_1.$$

De plus, il existe des vecteurs μ et v non nuls tels que $\|Pv\|_\infty = \|v\|_\infty$ et $\|\mu P\|_1 = \|\mu\|_1$.

Démonstration. On a

$$\|Pv\|_\infty = \sup_{x \in \mathcal{X}} \left| \sum_{y \in \mathcal{X}} p_{xy} v(y) \right| \leq \sup_{x \in \mathcal{X}} \left[\|v\|_\infty \sum_{y \in \mathcal{X}} p_{xy} \right] = \|v\|_\infty,$$

et

$$\|\mu P\|_1 = \sum_{x \in \mathcal{X}} \left| \sum_{y \in \mathcal{X}} \mu(y) p_{yx} \right| \leq \sum_{y \in \mathcal{X}} |\mu(y)| \sum_{x \in \mathcal{X}} p_{yx} = \|\mu\|_1.$$

Pour avoir égalité, il suffit de prendre $v = \mathbf{1}$ et $\mu = \pi$. □

Remarque 3.2.3: Norme subordonnée

On peut associer à P une *norme subordonnée* $\|P\|$, correspondant à la norme $\|\cdot\|_1$ pour la multiplication à gauche et à la norme $\|\cdot\|_\infty$ pour la multiplication à droite, satisfaisant

$$\|P\| := \sup_{v \neq 0} \frac{\|Pv\|_\infty}{\|v\|_\infty} = \sup_{\mu \neq 0} \frac{\|\mu P\|_1}{\|\mu\|_1} = 1.$$

Corollaire 3.2.4: Module des valeurs propres

Toute valeur propre λ d'une matrice stochastique P satisfait $|\lambda| \leq 1$.

Démonstration. Soit λ une valeur propre de P , et v un vecteur propre à droite associé. Alors

$$|\lambda| \|v\|_\infty = \|\lambda v\|_\infty = \|Pv\|_\infty \leq \|v\|_\infty,$$

d'où le résultat, car on peut diviser des deux côtés par $\|v\|_\infty > 0$. □

3.3 Théorème de Perron–Frobenius et trou spectral

Le résultat suivant est un cas particulier du théorème de Perron–Frobenius (ce théorème est plus général, car il admet des versions s’appliquant à des matrices non stochastiques, à condition que tous leurs éléments soient réels non négatifs).

Théorème 3.3.1: Perron–Frobenius

Soit P une matrice stochastique irréductible. Alors

- P admet $\lambda_0 = 1$ comme valeur propre *simple* (de multiplicité algébrique 1);
- si P est apériodique, alors toutes ses valeurs propres autres que λ_0 sont de module strictement inférieur à 1;
- si P est périodique, de période p , alors elle admet exactement p valeurs propres de module 1, qui sont des racines p èmes de 1.

Nous admettrons ce résultat. Voici toutefois quelques indications sur sa démonstration.

- Si la valeur propre $\lambda_0 = 1$ n’était pas de multiplicité 1, on pourrait trouver au moins deux vecteurs lignes π et μ , linéairement indépendants, tels que $\pi P = \pi$ et $\mu P = \mu$ (dans le cas diagonalisable, sinon l’argument est un peu plus compliqué). Le vecteur μ n’est pas nécessairement une mesure de probabilité. Mais on peut trouver $\theta \in [0, 1]$ tel que la combinaison convexe

$$\nu = \theta\mu + (1 - \theta)\pi$$

soit une mesure de probabilité. Dans le cas diagonalisable, on trouve

$$\nu P^n = \nu \quad \forall n \geq 0.$$

Mais ceci contredit l’unicité de la probabilité invariante.

- Si P est apériodique, supposons par l’absurde que P admet une valeur propre λ de module 1, différente de 1. Si λ est réelle, pour un vecteur propre à gauche μ , on peut procéder comme au point précédent, pour construire une mesure de probabilité ν satisfaisant

$$\nu P^n = \theta\lambda^n\mu + (1 - \theta)\pi.$$

Mais alors νP^n ne converge pas vers π lorsque n tend vers l’infini, ce qui contredit le Théorème 2.5.4. Si λ est complexe, alors $\bar{\lambda}$ est également valeur propre, de vecteur propre $\bar{\mu}$, et on peut appliquer un argument analogue avec le vecteur réel $\mu + \bar{\mu}$.

- Si P est périodique de période P , l’idée de base est que P^P admet p sous-espaces invariants supplémentaires. La restriction de P à chacun de ces sous-espaces doit admettre la valeur propre 1, ce qui correspond à une valeur propre racine p ème de l’unité de P .

Concentrons-nous maintenant sur le cas où P est apériodique.

Lemme 3.3.2: Limite de P^n

Si P est apériodique, alors

$$\lim_{n \rightarrow \infty} P^n = \Pi_0 = \mathbf{1}\pi. \quad (3.3.1)$$

La matrice Π_0 est un *projecteur*, c’est-à-dire qu’elle satisfait $\Pi_0^2 = \Pi_0$.

Démonstration. Le théorème 2.5.4 implique que νP^n converge vers π pour toute loi initiale ν . La relation (3.3.1) s’obtient en appliquant ceci à δ_x pour tout $x \in \mathcal{X}$. La relation $\Pi_0^2 = \Pi_0$ suit du fait que $\pi\mathbf{1} = 1$, en vertu de (2.1.1). \square

Remarque 3.3.3

La matrice Π_0 est une matrice dont toutes les lignes sont égales. En particulier, si \mathcal{X} est fini, de cardinal N , alors

$$\Pi_0 = \begin{pmatrix} \pi(1) & \dots & \pi(N) \\ \vdots & & \vdots \\ \pi(1) & \dots & \pi(N) \end{pmatrix}.$$

Définition 3.3.4: Rayon spectral et trou spectral

Soit P une matrice stochastique irréductible et apériodique, et soit $P_\perp = P - \Pi_0$. Alors le *rayon spectral* de P_\perp est

$$\begin{aligned} \rho &= \sup\{|\lambda_j| : \lambda_j \text{ est valeur propre de } P_\perp\} \\ &= \sup\{|\lambda_j| : \lambda_j \text{ est valeur propre de } P, \lambda \neq 1\}. \end{aligned}$$

Le *trou spectral* de P est par définition $1 - \rho$.

Le théorème de Perron–Frobenius implique que $0 \leq \rho < 1$, donc que $1 - \rho > 0$. L'intérêt de cette définition est lié à l'observation suivante.

Proposition 3.3.5: Vitesse de convergence et trou spectral

On a

$$\mathbb{E}^\nu[f(X_n)] - \mathbb{E}^\pi[f] = (\nu - \pi)P_\perp^n f.$$

Démonstration. On a une décomposition de l'espace des mesures en deux sous-espace supplémentaires, invariants par P , l'un associé à Π_0 , et l'autre associé à P_\perp . Le premier est simplement le sous-espace vectoriel de dimension 1 engendré par π , alors que le second est

$$\mathbf{1}_\perp = \left\{ \mu : \mathcal{X} \rightarrow \mathbb{R} : \mu \mathbf{1} = 0 \right\} = \left\{ \mu : \mathcal{X} \rightarrow \mathbb{R} : \sum_{x \in \mathcal{X}} \mu(x) = 0 \right\}.$$

En effet, si $\mu \in \mathbf{1}_\perp$, alors

$$\mu P \mathbf{1} = \mu \mathbf{1} = 0,$$

ce qui implique que $\mu P \in \mathbf{1}_\perp$, ou encore $\mathbf{1}_\perp P \subset \mathbf{1}_\perp$. De plus, on a

$$\begin{aligned} \mu P_\perp &= \mu P - \mu \Pi_0 = \mu P \\ \pi P_\perp &= \pi P - \pi \Pi_0 = \pi - \pi \mathbf{1} \pi = 0 \end{aligned} \tag{3.3.2}$$

puisque $\mu \Pi_0 = \mu \mathbf{1} \pi = 0$ et $\pi \mathbf{1} = 1$.

Décomposons alors ν en $\nu = \pi + \mu$. On a $\mu \in \mathbf{1}_\perp$, puisque $\mu \mathbf{1} = \nu \mathbf{1} - \pi \mathbf{1} = 1 - 1 = 0$. Il suit de (3.3.2) que pour tout $n \geq 0$,

$$\nu P^n = (\pi + \mu)P^n = \pi + \mu P_\perp^n.$$

Par conséquent,

$$\mathbb{E}^\nu[f(X_n)] = \nu P^n f = \pi f + \mu P_\perp^n f,$$

d'où le résultat. □

Par la majoration (3.2.1), on a

$$|\mathbb{E}^\nu[f(X_n)] - \mathbb{E}^\pi[f]| \leq \|\nu - \pi\|_1 \|P_\perp^n f\|_\infty. \quad (3.3.3)$$

On s'attend à avoir

$$\|P_\perp^n f\|_\infty \leq C \rho^n \|f\|_\infty \quad (3.3.4)$$

pour une constante C à déterminer. Si c'est bien le cas, alors on aura montré que $\mathbb{E}^\nu[f(X_n)]$ converge exponentiellement vite vers $\mathbb{E}^\pi[f]$, avec une erreur qui décroît comme ρ^n .

3.4 Diagonalisation et décomposition de Dunford

Notre objectif est maintenant de vérifier (3.3.4). Nous supposons pour l'instant que \mathcal{X} est fini, de cardinal N . Considérons d'abord le cas où P_\perp est diagonalisable. Alors il existe une matrice non singulière S telle que

$$S^{-1}P_\perp S = \Lambda_\perp = \begin{pmatrix} 0 & 0 & \dots & \dots & 0 \\ 0 & \lambda_1 & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \lambda_{N-2} & 0 \\ 0 & \dots & \dots & 0 & \lambda_{N-1} \end{pmatrix}.$$

En effet, la première valeur propre de P_\perp est nulle, puisque $\pi P_\perp = 0$, cf. (3.3.2). On a alors $P_\perp = S \Lambda_\perp S^{-1}$, et

$$P_\perp^n = S \Lambda_\perp^n S^{-1} \quad \forall n \geq 0.$$

On remarque que $\|\Lambda_\perp^n g\|_\infty \leq \rho^n \|g\|_\infty$ par définition du rayon spectral, et que par conséquent

$$\|P_\perp^n\|_\infty \leq \|S\| \|\Lambda_\perp^n S^{-1}\|_\infty \leq \rho^n \|S\| \|S^{-1}\| \|f\|_\infty,$$

où les normes de S et S^{-1} sont des normes subordonnées, comme définies dans la remarque 3.2.3. On conclut donc que (3.3.4) est vérifié, avec $C = \|S\| \|S^{-1}\|$.

Si P_\perp n'est pas diagonalisable, on a

$$S^{-1}P_\perp S = T_\perp,$$

où T_\perp est une matrice triangulaire, diagonale par blocs, où les blocs sont des *blocs de Jordan* de la forme $B(\lambda_j, b_j)$, avec

$$B(\lambda, b) = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & & \vdots \\ \vdots & & \ddots & \ddots & \\ \vdots & & & \lambda & 1 \\ 0 & \dots & \dots & 0 & \lambda \end{pmatrix} \in \mathbb{C}^{b \times b}.$$

La dimension b_j de $B(\lambda_j, b_j)$ dépend de la différence entre la *multiplicité algébrique* de λ_j (sa multiplicité en tant que racine du polynôme caractéristique), et sa *multiplicité géométrique* (la dimension du noyau de $P - \lambda_j \mathbb{1}$). Dans ce cas, on a

$$P_\perp^n = S T_\perp^n S^{-1} \quad \forall n \geq 0.$$

On pourrait alors essayer de majorer $\|T_\perp^n g\|_\infty$ par une constante fois $\rho^n \|g\|_\infty$. Il est toutefois plus commode de passer par la *décomposition de Dunford*, que nous rappelons ici.

Proposition 3.4.1: Décomposition de Dunford

Soit P une matrice, admettant les valeurs propres différentes $\lambda_0, \dots, \lambda_k$. On note m_i la multiplicité algébrique de λ_i , et g_i sa multiplicité géométrique (on rappelle que $1 \leq g_i \leq m_i$). Alors on a la décomposition

$$P = \sum_{i=0}^k (\lambda_i \Pi_i + N_i),$$

où

- les Π_i sont des projecteurs, satisfaisant $\Pi_i \Pi_j = \delta_{ij} \Pi_i$;
- les N_i sont nilpotentes : elles satisfont $N_i^{m_i - g_i} = 0$;
- on a $N_i N_j = 0$ si $i \neq j$ et $P_i N_j = N_j P_i = \delta_{ij} N_i$.

Il suit de la dernière propriété que

$$P_{\perp}^n = \sum_{i=1}^k (\lambda_i \Pi_i + N_i)^n,$$

et la formule du binôme de Newton implique

$$(\lambda_i \Pi_i + N_i)^n = \Pi_i \sum_{p=0}^{m_i - g_i - 1} \lambda_i^{n-p} \binom{n}{p} N_i^p.$$

En effet, le fait que $N_i^{m_i - g_i} = 0$ implique que tous les termes avec $p \geq m_i - g_i$ sont nuls.

Le point important ici est que puisque $m_i - g_i$ est borné, $\|P_{\perp}^n f\|_{\infty}$ décroît toujours comme ρ^n , même si ce terme est multiplié par une constante qui dépend de manière plus compliquée de P_{\perp} (mais pas de n). Ainsi, (3.3.4) reste vrai, avec un C dépendant des termes de la décomposition de Dunford.

Nous avons supposé jusqu'ici que \mathcal{X} était fini. Si \mathcal{X} est infini, la matrice stochastique définit un opérateur linéaire dit *compact*, ce qui signifie essentiellement qu'il applique des ensembles compacts sur des ensembles bornés (dont la fermeture est compacte). Pour ces opérateurs, la notion de valeur propre est encore bien définie. En particulier, on sait que toute valeur propre non nulle de P est de multiplicité finie. Par conséquent, on a encore une décomposition de Dunford. Toutefois, il est moins clair que la constante C dans (3.3.4) est toujours finie.

3.5 Cas réversible

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

$$\langle f, g \rangle_{\pi} = \sum_{x \in \mathcal{X}} \pi(x) \overline{f(x)} g(x),$$

où $f, g \in \mathbb{C}^{\mathcal{X}}$ sont des vecteurs colonne. On dénote par $\ell^2(\mathbb{C}, \pi)$ l'ensemble des vecteurs f tels que $\langle f, f \rangle_{\pi} < \infty$. C'est un espace de Hilbert.

Lemme 3.5.1: Caractère autoadjoint de P

L'opérateur linéaire P est autoadjoint dans l'espace de Hilbert $\mathcal{H} = \ell^2(\mathbb{C}, \pi)$, c'est-à-dire

$$\langle f, P g \rangle_{\pi} = \langle P f, g \rangle_{\pi} \quad \forall f, g \in \mathcal{H}.$$

Démonstration. On a

$$\langle f, Pg \rangle_\pi = \sum_{x \in \mathcal{X}} \pi(x) \overline{f(x)} \sum_{y \in \mathcal{X}} p_{xy} g(y) = \sum_{y \in \mathcal{X}} \pi(y) \sum_{x \in \mathcal{X}} p_{yx} \overline{f(x)} g(y) = \langle Pf, g \rangle_\pi,$$

où on a utilisé la réversibilité dans la deuxième égalité. \square

Rappelons un résultat classique de la théorie des espaces de Hilbert.

Proposition 3.5.2: Théorème spectral

Soit P un opérateur autoadjoint compact dans un espace de Hilbert \mathcal{H} . Alors toutes les valeurs propres de P sont réelles, et les espaces propres associés sont orthogonaux. De plus, \mathcal{H} admet une base orthonormée de vecteurs propres, dans laquelle P est diagonale.

Démonstration. Soient v_1 et v_2 deux vecteurs propres à droite de P , de valeurs propres respectives λ_1 et λ_2 . Alors

$$(\bar{\lambda}_1 - \lambda_2) \langle v_1, v_2 \rangle_\pi = \langle \lambda_1 v_1, v_2 \rangle_\pi - \langle v_1, \lambda_2 v_2 \rangle_\pi = \langle P v_1, v_2 \rangle_\pi - \langle v_1, P v_2 \rangle_\pi = 0.$$

D'une part, prenant $v_1 = v_2$, on obtient que λ_1 est réelle. D'autre part, si $\lambda_1 \neq \lambda_2$, on obtient l'orthogonalité de v_1 et v_2 .

Le fait que P est diagonalisable se montre par récurrence. On sait que P admet au moins une valeur propre complexe, avec vecteur propre associé v . On montre alors que le complément orthogonal $v_\perp = \{w \in \mathcal{H} : \langle w, v \rangle_\pi = 0\}$ est invariant par P . La restriction P_\perp de P à v_\perp admet à nouveau une valeur propre, ce qui permet d'établir l'hérédité (si P est de dimension finie, la récurrence s'arrête lorsque le complément orthogonal est $\{0\}$). \square

On a également un lien explicite entre vecteurs propres à gauche et à droite.

Lemme 3.5.3: Vecteurs propres à droite et à gauche

Si v est un vecteur propre à droite de l'opérateur autoadjoint P , alors μ défini par

$$\mu(x) = \pi(x)v(x) \quad \forall x \in \mathcal{X}$$

est un vecteur propre à gauche, pour la même valeur propre.

Démonstration. Soit v un vecteur colonne tel que $Pv = \lambda v$. Pour tout $x \in \mathcal{X}$, on a

$$(\mu P)_x = \sum_{y \in \mathcal{X}} \mu(y) p_{yx} = \sum_{y \in \mathcal{X}} v(y) \pi(y) p_{yx} = \pi(x) \sum_{y \in \mathcal{X}} p_{xy} v(y) = \pi(x) (Pv)_x = \lambda \pi(x) v(x) = \lambda \mu(x).$$

Par conséquent, $\mu P = \lambda \mu$. \square

Une première conséquence du caractère autoadjoint de P est une représentation variationnelle du trou spectral.

Proposition 3.5.4: Principe min-max

Le trou spectral de P satisfait

$$\rho = \sup_{v: \langle v, \mathbf{1} \rangle_\pi = 0} \frac{|\langle v, Pv \rangle_\pi|}{\langle v, v \rangle_\pi}.$$

Démonstration. Soit $(v_k)_{k \geq 0}$ une base orthonormée de vecteurs propres à droite de P . Alors tout $v \in \mathcal{H}$ s'écrit

$$v = \sum_{k \geq 0} c_k v_k, \quad \text{où } c_k = \langle v_k, v \rangle_\pi.$$

On obtient alors

$$\begin{aligned} \langle v, v \rangle_\pi &= \sum_{k, \ell \geq 0} \bar{c}_k c_\ell \langle v_k, v_\ell \rangle_\pi = \sum_{k \geq 0} |c_k|^2, \\ \langle v, Pv \rangle_\pi &= \sum_{k, \ell \geq 0} \bar{c}_k c_\ell \langle v_k, Pv_\ell \rangle_\pi = \sum_{k \geq 0} \lambda_k |c_k|^2. \end{aligned}$$

La première relation n'est autre que la relation de Parseval. Par conséquent,

$$\frac{|\langle v, Pv \rangle_\pi|}{\langle v, v \rangle_\pi} \leq \frac{\sum_{k \geq 0} |\lambda_k| |c_k|^2}{\sum_{k \geq 0} |c_k|^2}.$$

Si $\langle v, \mathbf{1} \rangle_\pi = 0$, alors $c_0 = 0$, de sorte que cette quantité est bornée par ρ . L'égalité a lieu dans le cas $v = v_1$, si on a numéroté les valeurs propres de manière que $|\lambda_1| = \rho$. \square

Il est également possible d'obtenir une majoration analogue à (3.3.3). Même si elle ne peut pas sembler optimale, elle a le mérite d'être explicite.

Proposition 3.5.5: Vitesse de convergence dans le cas réversible

Si la chaîne de Markov est réversible, on a la majoration

$$\left| \mathbb{E}^\nu [f(X_n)] - \mathbb{E}^\pi [f] \right| \leq \rho^n \|f\|_\infty \| \nu - \pi \|_1^{1/2} \sup_{x \in \mathcal{X}} \left| \frac{\nu(x)}{\pi(x)} - 1 \right|^{1/2}.$$

Démonstration. Il s'agit de majorer $|(\nu - \pi)P_\perp^n f|$. La décomposition de Dunford s'écrit

$$P_\perp^n = \sum_{k \geq 1} \lambda_k \Pi_k,$$

où le projecteur Π_k peut s'écrire $\Pi_k = v_k \mu_k$. En effet, Π_k projette bien sur v_k par action à droite, et sur μ_k par action à gauche. De plus, $\Pi_k^2 = v_k (\mu_k v_k) \mu_k = \Pi_k$, puisque

$$\mu_k v_k = \sum_{x \in \mathcal{X}} \mu_k(x) v_k(x) = \sum_{x \in \mathcal{X}} \pi(x) v_k(x) v_k(x) = \langle v_k, v_k \rangle_\pi = 1.$$

Nous avons donc

$$(\nu - \pi)P_\perp^n f = \sum_{k \geq 1} \lambda_k (\nu - \pi) v_k \mu_k f = \sum_{k \geq 1} \lambda_k a_k b_k, \quad (3.5.1)$$

où nous avons posé

$$a_k = \mu_k f = \sum_{x \in \mathcal{X}} \mu_k(x) f(x) = \sum_{x \in \mathcal{X}} \pi(x) v_k(x) f(x) = \langle v_k, f \rangle_\pi,$$

et

$$b_k = (\nu - \pi) v_k = \sum_{x \in \mathcal{X}} (\nu(x) - \pi(x)) v_k(x) = \langle \nu, v_k \rangle_\pi,$$

où g est le vecteur colonne de composantes $g(x) = (\nu(x) - \pi(x))/\pi(x)$. Il suit alors de (3.5.1) et de l'inégalité de Cauchy-Schwarz que

$$|(\nu - \pi)P_{\perp}^n f| \leq \rho \sum_{k \geq 1} |a_k b_k| \leq \rho \left(\sum_{k \geq 1} a_k^2 \right)^{1/2} \left(\sum_{k \geq 1} b_k^2 \right)^{1/2}.$$

Or, par la relation de Parseval,

$$\sum_{k \geq 1} a_k^2 \leq \langle f, f \rangle_{\pi} = \sum_{x \in \mathcal{X}} \pi(x) f(x)^2 \leq \|f\|_{\infty}^2.$$

D'autre part,

$$\sum_{k \geq 1} b_k^2 \leq \langle g, g \rangle_{\pi} = \sum_{x \in \mathcal{X}} \pi(x) g(x)^2 \leq \sup_{x \in \mathcal{X}} |g(x)| \|\pi g\|_1.$$

Comme $\|\pi g\|_1 = \|\nu - \pi\|_1$, le résultat est prouvé. \square

Le facteur $\|\nu - \pi\|_1$ ne pose pas de problème, car on peut toujours le majorer par $\|\nu\|_1 + \|\pi\|_1 = 2$. Pour que le supremum sur x soit petit, il faut que $\nu(x)$ ne soit pas trop différent de $\pi(x)$, du moins si $\pi(x)$ est petit. Une possibilité est de choisir pour ν la probabilité uniforme sur un ensemble probable sous π , et sur lequel π ne varie pas trop.

Proposition 3.5.6: Cas d'un ν uniforme

Soit $\mathcal{X}_0 \subset \mathcal{X}$ un ensemble fini, tel que

$$\pi(\mathcal{X}_0^c) := \sum_{x \notin \mathcal{X}_0} \pi(x) = \delta \quad \text{et} \quad \max_{x \in \mathcal{X}_0} \pi(x) \leq (1+c) \min_{x \in \mathcal{X}_0} \pi(x).$$

Soit ν la loi uniforme sur \mathcal{X}_0 . Alors

$$\|\nu - \pi\|_1 \leq 2\delta + c \quad \text{et} \quad \sup_{x \in \mathcal{X}} \left| \frac{\nu(x)}{\pi(x)} - 1 \right| \leq \max \left\{ 1, \frac{c(1+\delta)}{(1+c)(1-\delta)} \right\}.$$

Démonstration. Soit

$$m = \min_{x \in \mathcal{X}_0} \pi(x), \quad M = \max_{x \in \mathcal{X}_0} \pi(x).$$

Alors on a $M \leq (1+c)m$ et

$$m|\mathcal{X}_0| \leq \pi(\mathcal{X}_0) = 1 - \delta \leq M|\mathcal{X}_0|.$$

En combinant ces inégalités, on obtient

$$M \leq \frac{(1+c)(1-\delta)}{|\mathcal{X}_0|} \quad \text{et} \quad m \geq \frac{1-\delta}{(1+c)|\mathcal{X}_0|}.$$

On a

$$\|\nu - \pi\|_1 = \sum_{x \in \mathcal{X}_0} \left| \frac{1}{|\mathcal{X}_0|} - \pi(x) \right| + \sum_{x \in \mathcal{X}_0^c} \pi(x).$$

La seconde somme vaut δ , alors qu'en utilisant le fait que $m \leq \pi(x) \leq M$ dans la première somme, on obtient, en simplifiant l'expression obtenue, que celle-ci est toujours inférieure à $\delta + c$. Ceci prouve la majoration de $\|\nu - \pi\|_1$. Pour la seconde majoration, on utilise le fait que

$$\sup_{x \in \mathcal{X}} \left| \frac{\nu(x)}{\pi(x)} - 1 \right| = \max \left\{ \sup_{x \in \mathcal{X}_0} \left| \frac{\nu(x)}{\pi(x)} - 1 \right|, 1 \right\},$$

et on borne la première somme à nouveau à l'aide de l'encadrement $m \leq \pi(x) \leq M$. \square

Le message essentiel à retenir de ce chapitre est que la théorie spectrale permet de montrer que $\mathbb{E}^{\pi_0}[f(X_n)]$ converge exponentiellement vite vers $\mathbb{E}^{\pi}[f]$, avec un exposant déterminé par le trou spectral, et une constante proportionnelle à $\|f\|_{\infty}$. Toutefois, si \mathcal{X} est grand ou infini, il n'est pas facile de déterminer explicitement le trou spectral, ainsi que la constante. C'est pour cette raison que nous allons introduire une autre approche, basée sur des fonctions de Lyapounov, qui est plus flexible et a l'avantage de fournir des valeurs explicites de l'exposant et de la constante.

3.6 Exercices

Exercice 3.6.1

On considère la marche aléatoire symétrique sur le cercle discret à N sites :

$$p_{xy} = \begin{cases} \frac{1}{2} & \text{si } y = x + 1, \\ \frac{1}{2} & \text{si } y = x - 1, \\ 0 & \text{sinon,} \end{cases}$$

avec l'identification modulo N : $N + 1 = 1$, $0 = N$.

1. Quelle est la matrice de transition de cette chaîne de Markov ?
2. Par un argument de symétrie, trouver la probabilité invariante de la chaîne.
3. Soit $\omega = e^{2\pi i/N}$. Montrer que pour tout $k \in \{0, \dots, N-1\}$, le vecteur v_k de composantes

$$v_{k,x} = \omega^{k(x-1)}, \quad x \in \{1, \dots, N\}$$

est un vecteur propre de P . En déduire les valeurs propres de P .

4. Déterminer le rayon spectral ρ de P (sa valeur propre différente de 1 de plus grand module). Distinguer les cas N pair et N impair.
5. Par un développement limité, déterminer le trou spectral $1 - \rho$ à l'ordre dominant en N .

Exercice 3.6.2

Soit $p \in]0, 1[$ et $q = 1 - p$. On considère la marche aléatoire asymétrique sur le cercle discret à N sites :

$$p_{xy} = \begin{cases} p & \text{si } y = x + 1, \\ q & \text{si } y = x - 1, \\ 0 & \text{sinon.} \end{cases}$$

Par la même méthode qu'à l'exercice précédent, déterminer, en fonction de p , le rayon spectral ρ de P , ainsi que le trou spectral $1 - \rho$ à l'ordre dominant en N .

Chapitre 4

Fonctions de Lyapounov et vitesse de convergence

Dans ce chapitre, nous considérons à nouveau des chaînes de Markov $(X_n)_{n \geq 0}$ irréductibles, récurrentes positives et apériodiques sur un ensemble dénombrable \mathcal{X} . Soit $f : \mathcal{X} \rightarrow \mathbb{R}$ une fonction bornée, et soit π la probabilité invariante de la chaîne de Markov. Le but est à nouveau de majorer l'erreur

$$|\mathbb{E}^\nu[f(X_n)] - \mathbb{E}^\pi[f]|.$$

Au lieu d'utiliser des informations sur les valeurs propres de la matrice de transition P , nous allons ici baser l'analyse sur des propriétés de fonctions dites de Lyapounov. Si les estimations fournies par ces fonctions ne sont pas toujours aussi précises que celles provenant de l'analyse spectrale, la méthode est plus robuste, et donne souvent des bornes explicites.

4.1 Notations – formalisme des générateurs

Commençons par préciser quelques définitions liées aux mesures et aux fonctions tests.

Définition 4.1.1: Mesures signées

Une *mesure signée finie* sur \mathcal{X} est une application $\mu : \mathcal{X} \rightarrow \mathbb{R}$ telle que

$$\|\mu\|_1 := \sum_{x \in \mathcal{X}} |\mu(x)| < \infty.$$

On notera \mathcal{E}_1 l'espace de Banach des mesures signées finies.

Si $\mu : \mathcal{X} \rightarrow [0, 1]$, et $\|\mu\|_1 = 1$, alors μ est une *mesure de probabilité*.

Notons que la somme de deux mesures de probabilité n'est pas une mesure de probabilité. Le sous-ensemble des mesures de probabilité n'est donc pas un sous-espace de \mathcal{E}_1 . Cependant, la combinaison convexe de deux mesures de probabilité est une mesure de probabilité.

Définition 4.1.2: Fonctions test

Une *fonction test* (ou *observable*) sur \mathcal{X} est une application $f : \mathcal{X} \rightarrow \mathbb{R}$ telle que

$$\|f\|_\infty := \sup_{x \in \mathcal{X}} |f(x)| < \infty.$$

On notera \mathcal{E}_∞ l'espace de Banach des fonctions test.

Les notations suivantes, en parties déjà introduites, vont s'avérer utiles.

- Pour une mesure signée finie μ et une fonction test f , nous écrivons

$$\mu(f) = \sum_{x \in \mathcal{X}} \mu(x) f(x).$$

Cette quantité est bien définie, car

$$|\mu(f)| \leq \sum_{x \in \mathcal{X}} |\mu(x)| |f(x)| \leq \sup_{x \in \mathcal{X}} |f(x)| \sum_{x \in \mathcal{X}} |\mu(x)| = \|f\|_\infty \|\mu\|_1 < \infty.$$

- Si μ est une mesure de probabilité, nous écrivons aussi $\mu(f) = \mathbb{E}^\mu[f]$.
- Si δ_x dénote la mesure de Dirac en x (c'est-à-dire que $\delta_x(x) = 1$ et $\delta_x(y) = 0$ si $y \neq x$), on abrège $\mathbb{E}^{\delta_x}[f]$ par $\mathbb{E}^x[f]$.
- Pour $A \subset \mathcal{X}$, on écrit

$$\mu(A) = \mu(\mathbb{1}_A) = \sum_{x \in A} \mu(x).$$

- Si μ est une mesure de probabilité, alors $\mu(A)$ est aussi la probabilité de A .
- Pour une mesure de probabilité μ et une fonction test f , on écrira

$$\mathbb{E}^\mu[f(X_n)] = \mu P^n f = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{X}} \mu(x) (P^n)_{xy} f(y),$$

où $(P^n)_{xy}$ est l'élément de matrice (x, y) de P^n .

Définition 4.1.3: Distance en variation totale

La distance en variation totale entre deux mesures $\mu, \nu \in \mathcal{E}_1$ est

$$\|\mu - \nu\|_{\text{TV}} = 2 \sup\{|\mu(A) - \nu(A)| : A \subset X\}.$$

Intuitivement, deux mesures sont d'autant plus proches en variation totale qu'elles donnent des probabilités proches aux événements.

Pour des mesures de probabilité, le résultat suivant montre que la distance en variation totale est en fait équivalente à la norme ℓ^1 .

Lemme 4.1.4: Équivalence des distances

Si μ et ν sont deux mesures de probabilité, alors

$$\|\mu - \nu\|_{\text{TV}} = \sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)| = \|\mu - \nu\|_1.$$

Démonstration. Soit $B = \{x \in \mathcal{X} : \mu(x) > \nu(x)\}$. Alors on a

$$0 \leq \mu(B) - \nu(B) = 1 - \mu(B^c) + (1 - \nu(B^c)) = \nu(B^c) - \mu(B^c), \quad (4.1.1)$$

ce qui implique

$$\begin{aligned} \sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)| &= \sum_{x \in B} (\mu(x) - \nu(x)) + \sum_{x \in B^c} (\nu(x) - \mu(x)) \\ &= \mu(B) - \nu(B) + \nu(B^c) - \mu(B^c) \\ &= 2[\mu(B) - \nu(B)] \end{aligned} \quad (4.1.2)$$

par (4.1.1). De plus, pour tout $A \subset \mathcal{X}$,

$$\mu(A) - \nu(A) \leq \sum_{x \in A \cap B} (\mu(x) - \nu(x)) \leq \sum_{x \in B} (\mu(x) - \nu(x)) = \mu(B) - \nu(B),$$

où nous avons utilisé à deux reprises le fait que $\mu(x) \leq \nu(x)$ sur $A \cap B^c$. De même,

$$\nu(A) - \mu(A) \leq \sum_{x \in A \cap B^c} (\nu(x) - \mu(x)) \leq \nu(B^c) - \mu(B^c) = \mu(B) - \nu(B).$$

Il suit de (4.1.2) que

$$|\mu(A) - \nu(A)| \leq \mu(B) - \nu(B) = \frac{1}{2} \|\mu - \nu\|_1.$$

De plus, si $A = B$, on a égalité. □

Définition 4.1.5: Générateur

Soit P la matrice de transition d'une chaîne de Markov sur un ensemble dénombrable \mathcal{X} . Le *générateur* de la chaîne de Markov est l'application $\mathcal{L} : \mathcal{E}_\infty \rightarrow \mathcal{E}_\infty$ donnée par

$$(\mathcal{L}f)(x) = \sum_{y \in \mathcal{X}} p_{xy} [f(y) - f(x)]. \quad (4.1.3)$$

Remarquons que comme $\sum_{y \in \mathcal{X}} p_{xy} = 1$, on a l'expression équivalente

$$(\mathcal{L}f)(x) = \left[\sum_{y \in \mathcal{X}} p_{xy} f(y) \right] - f(x) = \mathbb{E}^x[f(X_1)] - f(x).$$

On peut donc écrire $\mathcal{L} = P - \mathbb{1}$, où $\mathbb{1}$ dénote la matrice identité.

4.2 Fonctions de Lyapounov

Dans la suite, nous supposons que P est la matrice de transition d'une chaîne de Markov *irréductible* sur \mathcal{X} . De plus, nous supposons que \mathcal{X} est équipé d'une norme $\|\cdot\|$. Par exemple, si $\mathcal{X} \subset \mathbb{Z}$, on peut prendre $\|x\| = |x|$. Si $\mathcal{X} \subset \mathbb{Z}^d$, on peut prendre la norme Euclidienne (ou toute autre norme équivalente).

Définition 4.2.1: Fonction de Lyapounov

Une *fonction de Lyapounov* est une fonction $V : \mathcal{X} \rightarrow \mathbb{R}_+ = [0, \infty[$ satisfaisant

$$V(x) \rightarrow +\infty \quad \text{pour } \|x\| \rightarrow \infty.$$

Proposition 4.2.2: Formule de Dynkin

Pour toute fonction de Lyapounov V , on a

$$\mathbb{E}^x[V(X_n)] = V(x) + \mathbb{E}^x \left[\sum_{m=0}^{n-1} (\mathcal{L}V)(X_m) \right]. \quad (4.2.1)$$

De plus, si τ est un temps d'arrêt tel que $\mathbb{E}^x[\tau] < \infty$, alors

$$\mathbb{E}^x[V(X_\tau)] = V(x) + \mathbb{E}^x \left[\sum_{m=0}^{\tau-1} (\mathcal{L}V)(X_m) \right].$$

Démonstration. Montrons (4.2.1). On procède par récurrence sur n . L'initialisation se fait pour $n = 1$, où la définition (4.1.3) du générateur implique

$$\mathbb{E}^x[V(X_1)] = V(x) + (\mathcal{L}V)(x).$$

Pour vérifier l'hérédité, une première façon de procéder est d'écrire

$$\begin{aligned} \mathbb{E}^x[V(X_{n+1})] &= \sum_{y \in \mathcal{X}} V(y) \mathbb{P}^x\{X_{n+1} = y\} \\ &= \sum_{y \in \mathcal{X}} V(y) \sum_{z \in \mathcal{X}} \underbrace{\mathbb{P}^x\{X_{n+1} = y | X_n = z\} \mathbb{P}^x\{X_n = z\}}_{=p_{zy}} \\ &= \sum_{z \in \mathcal{X}} \mathbb{P}^x\{X_n = z\} \underbrace{\sum_{y \in \mathcal{X}} V(y) p_{zy}}_{=(\mathcal{L}V)(z) + V(z)} \\ &= \mathbb{E}^x \left[\sum_{z \in \mathcal{X}} \mathbb{1}_{X_n=z} (\mathcal{L}V)(z) \right] + \sum_{z \in \mathcal{X}} \mathbb{P}^x\{X_n = z\} V(z) \\ &= \mathbb{E}^x[(\mathcal{L}V)(X_n)] + \mathbb{E}^x[V(X_n)]. \end{aligned}$$

Une autre manière de procéder est d'utiliser le formalisme des espérances conditionnelles, en écrivant

$$\mathbb{E}^x[V(X_{n+1})] = \mathbb{E}^x[V(X_n)] + \mathbb{E}^x[V(X_{n+1}) - V(X_n)].$$

Or, si \mathcal{F}_n dénote la tribu engendrée par (X_0, X_1, \dots, X_n) , on a

$$\begin{aligned} \mathbb{E}^x[V(X_{n+1}) - V(X_n)] &= \mathbb{E}^x[\mathbb{E}^x[V(X_{n+1}) - V(X_n) | \mathcal{F}_n]] \\ &= \mathbb{E}^x[\mathbb{E}^{X_n}[V(X_{n+1}) - V(X_n)]] = \mathbb{E}^x[(\mathcal{L}V)(X_n)]. \end{aligned}$$

Avec l'hypothèse de récurrence, ceci conclut la démonstration. \square

Théorème 4.2.3: Croissance sous-exponentielle

Supposons qu'il existe une fonction de Lyapounov V et $c > 0$, $d \geq 0$ tels que

$$(\mathcal{L}V)(x) \leq cV(x) + d \quad \forall x \in \mathcal{X}.$$

Alors on a

$$\mathbb{E}^x[V(X_n)] \leq (1+c)^n V(x) + \frac{(1+c)^n - 1}{c} d$$

pour tout $n \in \mathbb{N}$ et tout $x \in \mathcal{X}$.

Démonstration. Commençons par considérer le cas $d = 0$. Notons $f_n(x) = \mathbb{E}^x[V(X_n)]$. Alors la formule de Dynkin implique

$$\begin{aligned} f_n(x) &= V(x) + \mathbb{E}^x \left[\sum_{m=0}^{n-1} (\mathcal{L}V)(X_m) \right] \\ &\leq V(x) + c \sum_{m=0}^{n-1} f_m(x). \end{aligned}$$

En utilisant $f_0(x) = V(x)$ comme initialisation, on obtient facilement par récurrence sur n que $f_n(x) \leq (1+c)^n V(x)$. Dans le cas $d > 0$, la relation de récurrence devient

$$f_n(x) \leq V(x) + c \sum_{m=0}^{n-1} f_m(x) + nd.$$

En posant

$$f_n(x) \leq (1+c)^n V(x) + k_n d,$$

on obtient pour k_n la relation de récurrence

$$k_n = c \sum_{m=0}^{n-1} k_m + n, \quad k_0 = 0.$$

On vérifie par récurrence que ceci équivaut à

$$k_n = \frac{1}{c} \left((1+c)^n - 1 \right),$$

d'où le résultat. □

Théorème 4.2.4: Non-explosion

Supposons qu'il existe $d \geq 0$ et un ensemble borné $K \subset \mathcal{X}$ tel que pour tout $x \in \mathcal{X}$, on ait

$$(\mathcal{L}V)(x) \leq d \mathbb{1}_K(x) = \begin{cases} d & \text{si } x \in K, \\ 0 & \text{sinon.} \end{cases}$$

Alors

$$\mathbb{P}^x \left\{ \lim_{n \rightarrow \infty} \|X_n\| = \infty \right\} = 0 \quad \forall x \in \mathcal{X}.$$

Démonstration (idée). Soit Ω_1 l'événement

$$\Omega_1 = \left\{ \omega : \lim_{n \rightarrow \infty} \|X_n(\omega)\| = \infty \right\}.$$

Considérons le cas $d = 0$. Alors la formule de Dynkin implique

$$\mathbb{E}^x [V(X_n)] \leq V(x).$$

Or, on a aussi

$$\mathbb{E}^x [V(X_n)] = \underbrace{\mathbb{E}^x [V(X_n) \mathbb{1}_{\Omega_1^c}]}_{\geq 0} + \mathbb{E}^x [V(X_n) \mathbb{1}_{\Omega_1}]$$

Par conséquent, $\mathbb{E}^x [V(X_n) \mathbb{1}_{\Omega_1}] \leq \mathbb{E}^x [V(X_n)] \leq V(x)$. Comme $V(X_n)$ tend vers l'infini sur Ω_1 , ceci n'est possible que si $\mathbb{E}^x [\mathbb{1}_{\Omega_1}] = \mathbb{P}^x \{\Omega_1\} = 0$. □

Théorème 4.2.5: Réurrence positive

Soit $f : \mathcal{X} \rightarrow [1, \infty[$ et V une fonction de Lyapounov telle que

$$(\mathcal{L}V)(x) \leq -cf(x) + d\mathbb{1}_K(x) \quad \forall x \in \mathcal{X},$$

pour un ensemble borné $K \subset \mathcal{X}$ et des constantes $c > 0$ et $d \geq 0$. Supposons de plus qu'il existe $\delta > 0$ tel que K satisfait

$$p_{xy} \geq \delta \quad \forall x, y \in K. \quad (4.2.2)$$

Alors la chaîne de Markov est récurrente positive, et admet donc une mesure de probabilité invariante π . De plus,

$$\pi(f) < \infty.$$

Démonstration. Commençons par considérer le cas où $K = \{x_0\}$ contient un seul point. Nous voulons montrer que x_0 est récurrent positif, car dans ce cas, la chaîne de Markov étant irréductible, elle sera récurrente positive. Soit donc

$$\tau_K = \tau_{x_0} = \inf\{n \geq 1 : X_n \in K\}.$$

Pour tout $T \in \mathbb{N}$, posons $\tau_K \wedge T = \min\{\tau_K, T\}$. Par la formule de Dynkin, qui s'applique puisque $\tau_K \wedge T$ est un temps d'arrêt fini presque sûrement, donc d'espérance finie, nous avons

$$\begin{aligned} 0 \leq \mathbb{E}^{x_0}[V(X_{\tau_K \wedge T})] &= V(x_0) + \mathbb{E}^{x_0} \left[\sum_{m=0}^{\tau_K \wedge T - 1} (\mathcal{L}V)(X_m) \right] \\ &\leq V(x_0) - c \mathbb{E}^{x_0} \left[\sum_{m=0}^{\tau_K \wedge T - 1} f(X_m) \right] + d \\ &\leq V(x_0) - c \mathbb{E}^{x_0}[\tau_K \wedge T] + d. \end{aligned} \quad (4.2.3)$$

Il suit que

$$\mathbb{E}^{x_0}[\tau_{x_0} \wedge T] = \mathbb{E}^{x_0}[\tau_K \wedge T] \leq \frac{V(x_0) + d}{c},$$

ce qui est fini. Comme la quantité de droite est indépendante de T , on peut prendre la limite lorsque T tend vers l'infini, pour lesquelles $\tau_K \wedge T$ tend vers τ_K . Ceci implique que la chaîne de Markov est récurrente positive.

Dans le cas où K contient au moins deux points, nous savons déjà par le raisonnement ci-dessus que $\mathbb{E}^{x_0}[\tau_K]$ est fini pour tout $x_0 \in K$. Il nous faut montrer que $\mathbb{E}^{x_0}[\tau_{x_0}]$ est également fini pour tout $x_0 \in K$.

Soit $\tau_{K,n}$ le temps du n ième passage de la chaîne en K , et soit

$$Y_n = X_{\tau_{K,n}}.$$

(La chaîne de Markov $(Y_n)_{n \geq 0}$ est appelée la *trace* de $(X_n)_{n \geq 0}$ sur K .) On peut déduire de l'hypothèse (4.2.2) que Y_n atteint x_0 en un temps d'espérance finie. Ceci suit du fait qu'une chaîne récurrente sur un ensemble fini est récurrente positive. Comme le temps de retour vers K est borné, il suit qu'on a bien $\mathbb{E}^{x_0}[\tau_{x_0}] < \infty$.

La chaîne de Markov étant récurrente positive, elle admet une unique probabilité invariante π . Il reste à montrer que $\pi(f)$ est fini. En fait (voir (2.4.3)), la mesure μ donnée par

$$\mu(y) = \mathbb{E}^{x_0} \left[\sum_{n=1}^{\tau_{x_0}} \mathbb{1}_{X_n=y} \right]$$

est invariante pour tout $x_0 \in \mathcal{X}$. La probabilité invariante π est obtenue en normalisant μ . Comme

$$\sum_{y \in \mathcal{X}} \mu(y) = \mathbb{E}^{x_0} \left[\underbrace{\sum_{n=1}^{\tau_{x_0}} \sum_{y \in \mathcal{X}} \mathbb{1}_{X_n=y}}_{=1} \right] = \mathbb{E}^{x_0} [\tau_{x_0}],$$

on conclut que

$$\pi(y) = \frac{1}{\mathbb{E}^{x_0} [\tau_{x_0}]} \mathbb{E}^{x_0} \left[\sum_{n=1}^{\tau_{x_0}} \mathbb{1}_{X_n=y} \right].$$

Il suit que

$$\pi(f) = \frac{1}{\mathbb{E}^{x_0} [\tau_{x_0}]} \mathbb{E}^{x_0} \left[\underbrace{\sum_{n=1}^{\tau_{x_0}} \sum_{y \in \mathcal{X}} f(y) \mathbb{1}_{X_n=y}}_{=f(X_n)} \right] = \frac{1}{\mathbb{E}^{x_0} [\tau_{x_0}]} \mathbb{E}^{x_0} \left[\sum_{n=1}^{\tau_{x_0}} f(X_n) \right].$$

Si $K = \{x_0\}$ (4.2.3) implique que pour tout $T \in \mathbb{N}$,

$$\mathbb{E}^{x_0} \left[\sum_{n=1}^{\tau_{x_0} \wedge T} f(X_n) \right] = \mathbb{E}^{x_0} \left[\sum_{n=0}^{\tau_{x_0} \wedge T - 1} f(X_n) \right] \leq \frac{V(x_0) + d}{c},$$

ce qui permet de majorer $\pi(f)$, en prenant le limite $T \rightarrow \infty$. Si K contient deux points ou plus, on a une majoration analogue, mais avec d multiplié par l'espérance du nombre de points de K visités avant d'atteindre x_0 , qui est finie. \square

Remarque 4.2.6

On peut affaiblir l'hypothèse (4.2.2) sur K de plusieurs manières.

- Il suffit de supposer qu'il existe un $k \geq 1$ tel que $(P^k)_{xy} \geq \delta$ pour tout $x, y \in K$ et un $\delta > 0$.
- Une condition suffisante encore plus faible est qu'il existe des réels $a_1, a_2, \dots \geq 0$, de somme égale à 1, tels que

$$\sum_{k=1}^{\infty} a_k (P^k)_{xy} \geq \delta$$

pour tout $x, y \in K$ et un $\delta > 0$. Des ensembles K satisfaisant ce critère sont appelés *petits* (« petite set » en anglais).

4.3 Normes à poids

Pour obtenir des résultats de convergence de la loi de X_n vers π , il est plus utile de travailler avec des normes à poids.

Définition 4.3.1: Norme à poids sur les fonctions test

Soit $W : \mathcal{X} \rightarrow [1, \infty[$. La norme à poids W d'une fonction test est définie comme

$$\|f\|_W = \sup_{x \in \mathcal{X}} \frac{|f(x)|}{W(x)}.$$

On notera \mathcal{E}_∞^W l'espace de Banach des fonctions test f telles que $\|f\|_W < \infty$.

Notons les propriétés suivantes :

- Pour tout $x \in \mathcal{X}$, on a $|f(x)| \leq \|f\|_W W(x)$.
- On a $\|f\|_W \leq \|f\|_\infty$. Par conséquent $\mathcal{E}_\infty \subset \mathcal{E}_\infty^W$.
- Plus généralement, si W_1 et W_2 sont deux poids tels que $W_1(x) \leq W_2(x)$ pour tout $x \in \mathcal{X}$, alors $\|f\|_{W_2} \leq \|f\|_{W_1}$. Par conséquent $\mathcal{E}_\infty^{W_1} \subset \mathcal{E}_\infty^{W_2}$.

On peut également définir une métrique duale à $\|\cdot\|_W$ entre mesures signées finies de la manière suivante.

Définition 4.3.2: Distance à poids entre mesures

Pour une fonction poids $W : \mathcal{X} \rightarrow [1, \infty[$ et deux mesures signées finies μ, ν , on pose

$$\begin{aligned} \rho_W(\mu, \nu) &= \sup_{f: \|f\|_W \leq 1} \sum_{x \in \mathcal{X}} f(x) |\mu(x) - \nu(x)| \\ &= \sup_{f: \|f\|_W \neq 0} \frac{1}{\|f\|_W} \sum_{x \in \mathcal{X}} f(x) |\mu(x) - \nu(x)|. \end{aligned} \quad (4.3.1)$$

On a alors les propriétés suivantes :

- On peut remplacer $f(x)$ par $|f(x)|$ dans la définition (4.3.1). En effet, cette transformation ne change pas $\|f\|_W$.
- On peut également remplacer $|\mu(x) - \nu(x)|$ par $\mu(x) - \nu(x)$. En effet, il suffit de changer le signe de $f(x)$ selon le signe de $\mu(x) - \nu(x)$ pour trouver le même résultat.
- Si $W(x) = 1$ pour tout $x \in \mathcal{X}$, alors on a

$$\rho_W(\mu, \nu) = \|\mu - \nu\|_{TV}.$$

En effet, le supremum dans (4.3.1) est alors atteint pour la fonction test f valant 1 partout.

- Pour un poids W général, on a

$$\rho_W(\mu, \nu) = \sum_{x \in \mathcal{X}} W(x) |\mu(x) - \nu(x)|. \quad (4.3.2)$$

En effet, le supremum dans (4.3.1) est atteint pour $f(x) = W(x)$ pour tout $x \in \mathcal{X}$.

- On a la majoration

$$|(\mu - \nu)(f)| \leq \|f\|_W \rho_W(\mu, \nu). \quad (4.3.3)$$

En effet,

$$|(\mu - \nu)(f)| \leq \sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)| |f(x)| \leq \|f\|_W \sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)| W(x),$$

d'où le résultat, par (4.3.2).

4.4 Un critère de convergence

Les théorèmes de la section 4.2 sont dûs à Meyn et Tweedie [3]. Leurs travaux fournissent également un résultat de convergence de $\mathbb{E}^x[f(X_n)]$ vers $\pi(f)$, mais les hypothèses sont assez difficiles à vérifier (notamment, tous les ensembles bornés doivent être petits), et les bornes obtenues ne sont pas explicites.

Le résultat suivant est dû à Hairer et Mattingly [2]. Les hypothèses sont plus faciles à vérifier en pratique, et la majoration obtenue pour $\mathbb{E}^x[f(X_n)] - \pi(f)$ a l'avantage de faire intervenir des constantes explicites.

Théorème 4.4.1: Critère de convergence pour espérances

Supposons que les deux conditions suivantes soient satisfaites.

1. **Condition de dérive géométrique :** Il existe $d \geq 0$, $c > 0$ et une fonction de Lyapounov V tels que

$$(\mathcal{L}V)(x) \leq -cV(x) + d \quad \forall x \in \mathcal{X}. \quad (4.4.1)$$

2. **Condition de minoration :** Pour un $R > 2d/c$, soit $K = \{x \in \mathcal{X} : V(x) < R\}$. Alors il existe $\alpha \in]0, 1[$ et une mesure de probabilité ν telle que

$$\inf_{x \in K} p_{xy} = \inf_{x \in K} \mathbb{P}^x\{X_1 = y\} \geq \alpha \nu(y) \quad \forall y \in \mathcal{X}. \quad (4.4.2)$$

Alors il existe des constantes $M > 0$ et $\bar{\gamma} < 1$ telles que

$$\|\mathbb{E}^x[f(X_n)] - \pi(f)\|_{1+V} \leq M \bar{\gamma}^n \|f - \pi(f)\|_{1+V} \quad (4.4.3)$$

pour toute fonction test $f \in \mathcal{E}_\infty^{1+V}$.

Précisons qu'on a

$$\|\mathbb{E}^x[f(X_n)] - \pi(f)\|_{1+V} = \sup_{x \in \mathcal{X}} \frac{|\mathbb{E}^x[f(X_n)] - \pi(f)|}{1 + V(x)}.$$

La majoration (4.4.3) peut donc s'écrire

$$|\mathbb{E}^x[f(X_n)] - \pi(f)| \leq (1 + V(x)) M \bar{\gamma}^n \|f - \pi(f)\|_{1+V} \quad \forall x \in \mathcal{X}.$$

Comme $\bar{\gamma} < 1$, on a donc convergence exponentielle de $\mathbb{E}^x[f(X_n)]$ vers $\pi(f)$. Comme en pratique, on peut souvent choisir x tel que $V(x)$ ne soit pas trop grand, la dépendance en $V(x)$ ne pose pas de problème.

La présence de $\|f - \pi(f)\|_{1+V}$ n'est pas vraiment restrictive non plus. Si par exemple $f \in \mathcal{E}_\infty$, ou si f est à support compact (c'est-à-dire nulle en-dehors d'un ensemble compact), cette quantité est finie.

La condition de minoration (4.4.2) est un peu plus faible que la condition (4.2.2) du Théorème 4.2.5. Le point crucial est que l'on ait une borne inférieure sur les probabilités de transition qui soit indépendante du point de départ dans K .

Dans la suite, nous utiliserons les notations

$$\begin{aligned} (\mu^{\mathcal{P}})(y) &= \mathbb{P}^\mu\{X_1 = y\} = \sum_{x \in \mathcal{X}} \mu(x) p_{xy}, \\ (\mathcal{P}f)(x) &= \mathbb{E}^x[f(X_1)] = \sum_{y \in \mathcal{X}} p_{xy} f(y). \end{aligned}$$

La condition de dérive géométrique (4.4.1) est équivalente à

$$(\mathcal{P}V)(x) \leq \gamma V(x) + d \quad \forall x \in \mathcal{X}.$$

avec $\gamma = 1 - c$. L'ingrédient essentiel de la démonstration du Théorème 4.4.1 est la borne suivante.

Proposition 4.4.2: L'application \mathcal{P} est contractante pour la distance $\rho_{1+\beta V}$

Il existe $\bar{\gamma} \in]0, 1[$ et $\beta > 0$ tels que

$$\rho_{1+\beta V}(\mu^{\mathcal{P}}, \nu^{\mathcal{P}}) \leq \bar{\gamma} \rho_{1+\beta V}(\mu, \nu).$$

En fait, la proposition donne des expressions explicites pour les constantes β et $\bar{\gamma}$: pour tout choix de α_0 et γ_0 tels que

$$0 < \alpha_0 < \alpha \quad \text{et} \quad \gamma + \frac{2d}{R} < \gamma_0 < 1, \quad (4.4.4)$$

on peut prendre

$$\begin{aligned} \beta &= \frac{\alpha_0}{d}, \\ \bar{\gamma} &= \max\left\{1 - (\alpha - \alpha_0), \frac{2 + R\beta\gamma_0}{2 + R\beta}\right\} = 1 - \min\left\{\alpha - \alpha_0, \frac{R\beta(1 - \gamma_0)}{2 + R\beta}\right\}. \end{aligned} \quad (4.4.5)$$

De plus, nous verrons ci-dessous que l'on a

$$M = \frac{1}{1 - \bar{\gamma}} \sup_{x \in \mathcal{X}} \frac{2 + \beta[(1 + \gamma)V(x) + d]}{1 + \beta V(x)} \leq \frac{\max\{1 + \gamma, 2 + \beta d\}}{1 - \bar{\gamma}}. \quad (4.4.6)$$

Ces expressions ne sont pas particulièrement élégantes, mais elles ont le mérite d'être explicites, ce qui peut servir dans les applications. Notons que si d tend vers 0, alors β tend vers l'infini, et on peut prendre $\bar{\gamma}$ arbitrairement proche de γ .

Nous allons d'abord montrer que la Proposition 4.4.2 implique bien le Théorème 4.4.1.

Démonstration du Théorème 4.4.1. Nous donnons tout d'abord une démonstration de l'existence de π , même si celle-ci suit en fait du Théorème 4.2.5. C'est une application du théorème du point fixe de Banach.

Fixons $x_0 \in \mathcal{X}$, et soit $\mu_0^{x_0} = \delta_{x_0}$ la mesure de Dirac en x_0 . Soit $\mu_n^{x_0} = \mu_0^{x_0} \mathcal{P}^n$. La proposition implique

$$\rho_{1+\beta V}(\mu_{n+1}^{x_0}, \mu_n^{x_0}) \leq \bar{\gamma} \rho_{1+\beta V}(\mu_n^{x_0}, \mu_{n-1}^{x_0}) \leq \dots \leq \bar{\gamma}^n \rho_{1+\beta V}(\mu_1^{x_0}, \mu_0^{x_0}).$$

On a donc une suite de Cauchy, et comme on sait que la distance en variation totale est complète, donc a fortiori la distance $\rho_{1+\beta V}$, on en conclut que la suite des $\mu_n^{x_0}$ converge vers une mesure π en variation totale. De plus, $\pi(1 + V)$ est finie, car $\rho_{1+\beta V}(\pi, \mu_0^{x_0})$ l'est.

Pour montrer que π est invariante, il suffit d'observer que

$$\pi \mathcal{P} = \lim_{n \rightarrow \infty} \mu_0^{x_0} \mathcal{P}^{n+1} = \lim_{n \rightarrow \infty} \mu_0^{x_0} \mathcal{P}^n = \pi.$$

Afin de démontrer (4.4.3), il est utile de centrer f . Posons donc

$$\hat{f}(x) = f(x) - \pi(f) \quad \forall x \in \mathcal{X}.$$

Alors $\pi(\hat{f}) = \pi(f) - \pi(\pi(f)) = 0$, puisque $\pi(\pi(f)) = \pi(f)$. Ainsi

$$\mathcal{P}^n f - \pi(f) = \mathcal{P}^n \hat{f} + \mathcal{P}^n \pi(f) - \pi(\hat{f}) - \pi(\pi(f)) = \mathcal{P}^n \hat{f}.$$

Ceci permet d'écrire

$$\|\mathbb{E}[f(X_n)] - \pi(f)\|_{1+V} = \|\mathcal{P}^n f - \pi(f)\|_{1+V} = \|\mathcal{P}^n \hat{f}\|_{1+V}.$$

Il s'agit donc de montrer qu'il existe une constante M telle que pour toute fonction test \hat{f} satisfaisant $\pi(\hat{f}) = 0$, on ait

$$\|\mathcal{P}^n \hat{f}\|_{1+V} \leq M \bar{\gamma}^n \|\hat{f}\|_{1+V}.$$

Or, comme $(\mathcal{P}^n \hat{f})(x) = \delta_x(\mathcal{P}^n \hat{f}) = \mu_n^x(\hat{f})$ et $\pi(\hat{f}) = 0$, on a

$$\|\mathcal{P}^n \hat{f}\|_{1+\beta V} = \sup_{x \in \mathcal{X}} \frac{|(\mu_n^x - \pi)(\hat{f})|}{1 + \beta V(x)} \leq \|\hat{f}\|_{1+\beta V} \sup_{x \in \mathcal{X}} \frac{\rho_{1+\beta V}(\mu_n^x, \pi)}{1 + \beta V(x)}$$

en vertu de (4.3.3). Observons que

$$\begin{aligned}\rho_{1+\beta V}(\mu_n^x, \pi) &= \rho_{1+\beta V}(\mu_n^x, \mu_{n+1}^x) + \rho_{1+\beta V}(\mu_{n+1}^x, \mu_{n+2}^x) + \dots \\ &\leq [\bar{\gamma}^n + \bar{\gamma}^{n+1} + \dots] \rho_{1+\beta V}(\mu_1^x, \mu_0^x) \\ &= \frac{\bar{\gamma}^n}{1 - \bar{\gamma}} \rho_{1+\beta V}(\mu_1^x, \mu_0^x).\end{aligned}$$

De plus,

$$\begin{aligned}\rho_{1+\beta V}(\mu_1^x, \mu_0^x) &= \sum_{y \in \mathcal{X}} (1 + \beta V(y)) |\mu_1^x(y) - \mu_0^x(y)| \\ &\leq \sum_{y \in \mathcal{X}} (1 + \beta V(y)) \mu_1^x(y) + 1 + \beta V(x) \\ &= (1 + \beta(\mathcal{P}V))(x) + 1 + \beta V(x).\end{aligned}$$

Comme $(\mathcal{P}V)(x) = (\mathcal{L}V)(x) + V(x) \leq \gamma V(x) + d$ par l'hypothèse de dérive géométrique (4.4.1), on obtient finalement

$$\|\mathcal{P}^n \hat{f}\|_{1+\beta V} \leq \|\hat{f}\|_{1+\beta V} \frac{\bar{\gamma}^n}{1 - \bar{\gamma}} \sup_{x \in \mathcal{X}} \frac{2 + \beta[V(x) + \gamma V(x) + d]}{1 + \beta V(x)} =: M \bar{\gamma}^n \|\hat{f}\|_{1+\beta V}.$$

En particulier, la borne est vraie pour $\beta = 1$, ce qui conclut la démonstration. \square

Il nous reste à démontrer la Proposition 4.4.2. L'idée est de travailler avec une définition alternative de la distance ρ_β . On introduit sur \mathcal{X} la distance

$$d_\beta(x, y) = \begin{cases} 0 & \text{si } x = y, \\ 2 + \beta V(x) + \beta V(y) & \text{si } x \neq y, \end{cases}$$

(on vérifie facilement que d_β satisfait bien la définition d'une distance), et la semi-norme de Lipschitz

$$\|f\|_\beta = \sup_{x \neq y} \frac{|f(x) - f(y)|}{d_\beta(x, y)},$$

C'est une semi-norme, et non une norme, car $\|f\|_\beta = 0$ n'implique pas $f = 0$ (mais seulement que f est constante). On a alors le résultat suivant.

Lemme 4.4.3: Équivalence des distances

On a

$$\rho_{1+\beta V}(\mu, \nu) = \rho_{1+\beta V}^*(\mu, \nu) := \sup_{f: \|f\|_\beta \leq 1} \sum_{x \in \mathcal{X}} f(x)(\mu(x) - \nu(x)). \quad (4.4.7)$$

Notons que la seule différence entre la définition (4.3.1) de $\rho_{1+\beta V}(\mu, \nu)$ et cette égalité est l'ensemble des f sur lequel on prend le supremum.

Montrons tout d'abord que ce lemme implique la proposition.

Démonstration de la Proposition 4.4.2. L'idée est de montrer que \mathcal{P} est contractante dans la seminorme $\|\cdot\|_\beta$, à savoir

$$\|\mathcal{P}f\|_\beta \leq \bar{\gamma} \|f\|_\beta. \quad (4.4.8)$$

En effet, ceci implique

$$\begin{aligned}
\rho_{1+\beta V}(\mu^{\mathcal{P}}, \nu^{\mathcal{P}}) &= \sup_{f: \|f\|_{\beta} \neq 0} \frac{1}{\|f\|_{\beta}} (\mu^{\mathcal{P}} - \nu^{\mathcal{P}})(f) \\
&= \sup_{f: \|f\|_{\beta} \neq 0} \frac{1}{\|f\|_{\beta}} (\mu - \nu)(\mathcal{P}f) \\
&\leq \sup_{f_1: \|f_1\|_{\beta} \neq 0} \frac{\bar{\gamma}}{\|f_1\|_{\beta}} (\mu - \nu)(f_1) \\
&= \bar{\gamma} \rho_{1+\beta V}(\mu, \nu).
\end{aligned}$$

Afin de démontrer (4.4.8), nous fixons un f tel que $\|f\|_{1+\beta V} \leq 1$ et $\|f\|_{\beta} \leq 1$. Il s'agit de montrer que

$$|(\mathcal{P}f)(x) - (\mathcal{P}f)(y)| \leq \bar{\gamma} d_{\beta}(x, y)$$

pour tout $x, y \in \mathcal{X}$. La relation est vraie pour $x = y$, donc nous supposons $x \neq y$. Nous considérons deux cas séparément.

- **Cas 1:** $V(x) + V(y) \geq R$. Dans ce cas, on a

$$|(\mathcal{P}f)(x)| = \left| \sum_{y \in \mathcal{X}} p_{xy} f(y) \right| \leq \underbrace{\|f\|_{1+\beta V}}_{\leq 1} \sum_{y \in \mathcal{X}} (1 + \beta V(y)) p_{xy} \leq 1 + \beta (\mathcal{P}V)(x), \quad (4.4.9)$$

d'où, par la condition (4.4.1) de dérive géométrique,

$$\begin{aligned}
|(\mathcal{P}f)(x) - (\mathcal{P}f)(y)| &\leq 2 + \beta (\mathcal{P}V)(x) + \beta (\mathcal{P}V)(y) \\
&\leq 2 + \beta \gamma V(x) + \beta \gamma V(y) + 2\beta d \\
&\leq 2 + \beta \gamma_0 [V(x) + V(y)]
\end{aligned}$$

pour tout γ_0 satisfaisant (4.4.4). Un calcul élémentaire montre alors qu'en posant

$$\gamma_1 = \frac{2 + \beta R \gamma_0}{2 + \beta R},$$

on obtient bien la majoration requise

$$|(\mathcal{P}f)(x) - (\mathcal{P}f)(y)| \leq \gamma_1 (2 + V(x) + V(y)) = \gamma_1 d_{\beta}(x, y).$$

- **Cas 2:** $V(x) + V(y) < R$. Dans ce cas, on a $x, y \in K$. La matrice $\widetilde{\mathcal{P}}$ d'éléments

$$\tilde{p}_{xy} = \frac{1}{1-\alpha} p_{xy} - \frac{\alpha}{1-\alpha} \nu(y)$$

est une matrice stochastique. En effet, la condition de minoration (4.4.2) montre que ces éléments sont tous positifs ou nuls, et on vérifie immédiatement que la somme sur y des \tilde{p}_{xy} vaut 1. De plus, on a

$$(\mathcal{P}f)(x) = (1-\alpha)(\widetilde{\mathcal{P}}f)(x) + \alpha \nu(f),$$

d'où, par un calcul analogue à (4.4.9),

$$\begin{aligned}
|(\mathcal{P}f)(x) - (\mathcal{P}f)(y)| &= (1-\alpha) |(\widetilde{\mathcal{P}}f)(x) - (\widetilde{\mathcal{P}}f)(y)| \\
&\leq (1-\alpha) [2 + \beta (\widetilde{\mathcal{P}}V)(x) + \beta (\widetilde{\mathcal{P}}V)(y)] \\
&\leq (1-\alpha) + \beta (\mathcal{P}V)(x) + \beta (\mathcal{P}V)(y) \\
&\leq 2(1-\alpha) + \beta \gamma [V(x) + V(y)] + 2\beta d \\
&\leq \gamma_2 d_{\beta}(x, y),
\end{aligned}$$

si $\beta = \frac{\alpha_0}{d}$ est donné par (4.4.5) et $\gamma_2 = \max\{\gamma, 1 - (\alpha - \alpha_0)\}$. La troisième ligne suit du fait que

$$(\widetilde{\mathcal{P}}V)(x) \leq \frac{1}{1-\alpha}(\mathcal{P}V)(x).$$

Le résultat suit, avec $\bar{\gamma} = \max\{\gamma_1, \gamma_2\}$. □

Il nous reste à démontrer le Lemme 4.4.3.

Démonstration du Lemme 4.4.3. Définissons les boules

$$\mathcal{B} = \{f \in \mathcal{E}_\infty : \|f\|_{1+\beta V} \leq 1\}, \quad \mathcal{B}^* = \{f \in \mathcal{E}_\infty : \|f\|_\beta \leq 1\}.$$

Alors on peut écrire

$$\rho_{1+\beta V}(\mu, \nu) = \sup_{f \in \mathcal{B}} f(\mu - \nu), \quad \rho_{1+\beta V}^*(\mu, \nu) = \sup_{f \in \mathcal{B}^*} f(\mu - \nu).$$

L'observation cruciale est la suivante. Pour $c \in \mathbb{R}$, soit $f + c$ la fonction translatée de c , définie par $(f + c)(x) = f(x) + c$ pour tout $x \in \mathcal{X}$. Alors on a

$$(f + c)(\mu - \nu) = \sum_{x \in \mathcal{X}} (f(x) + c)(\mu(x) - \nu(x)) = f(\mu - \nu) + c \sum_{x \in \mathcal{X}} \mu(x) - c \sum_{x \in \mathcal{X}} \nu(x) = f(\mu - \nu),$$

puisque μ et ν sont des mesures de probabilité. Par conséquent, on a aussi

$$\rho_{1+\beta V}(\mu, \nu) = \sup_{f \in \widehat{\mathcal{B}}} f(\mu - \nu), \quad \rho_{1+\beta V}^*(\mu, \nu) = \sup_{f \in \widehat{\mathcal{B}}^*} f(\mu - \nu),$$

où on a défini les « cylindres »

$$\widehat{\mathcal{B}} = \{f + c : f \in \mathcal{B}, c \in \mathbb{R}\}, \quad \widehat{\mathcal{B}}^* = \{f + c : f \in \mathcal{B}^*, c \in \mathbb{R}\}.$$

Ces cylindres sont obtenus en étendant les boules \mathcal{B} et \mathcal{B}^* dans la direction des fonctions constantes.

Nous allons montrer ci-dessous que

$$\|f\|_\beta = \inf_{c \in \mathbb{R}} \|f + c\|_{1+\beta V}. \quad (4.4.10)$$

Cela impliquera que $\widehat{\mathcal{B}} = \widehat{\mathcal{B}}^*$, et par conséquent que $\rho_{1+\beta V}(\mu, \nu) = \rho_{1+\beta V}^*(\mu, \nu)$, qui est l'égalité voulue.

Comme remarqué juste après la Définition 4.3.1, on a, pour tout $x \in \mathcal{X}$,

$$|f(x)| \leq \|f\|_{1+\beta V}(1 + \beta V(x)).$$

Par conséquent, on a pour tout $x \neq y \in \mathcal{X}$

$$\frac{|f(x) - f(y)|}{d_\beta(x, y)} \leq \frac{|f(x)| + |f(y)|}{2 + \beta V(x) + \beta V(y)} \leq \|f\|_{1+\beta V}.$$

Ceci implique

$$\|f\|_\beta \leq \|f\|_{1+\beta V}.$$

Comme de plus la définition de $\|f\|_\beta$ ne dépend que de différences de f en des points différents, on a $\|f + c\|_\beta = \|f\|_\beta$ pour tout $c \in \mathbb{R}$. On a donc obtenu

$$\|f\|_\beta \leq \inf_{c \in \mathbb{R}} \|f + c\|_{1+\beta V}.$$

Pour montrer l'inégalité inverse, posons

$$c^* = \inf_{x \in \mathcal{X}} (1 + \beta V(x) - f(x)).$$

Commençons par montrer que $|c^*| < \infty$. Comme $c^* \leq (1 + \beta V(x_0) - f(x_0))$ pour tout $x_0 \in \mathcal{X}$, il suffit pour cela de montrer que c^* est bornée inférieurement. Cela suit du fait que pour tout $x, y \in \mathcal{X}$, on a

$$f(x) \leq |f(y)| + |f(x) - f(y)| \leq |f(y)| + 2 + \beta V(x) + \beta V(y),$$

qui implique

$$1 + \beta V(x) - f(x) \geq -1 - \beta V(y) - |f(y)|.$$

Comme $V(y)$ est finie pour au moins un $y \in \mathcal{X}$, on a bien la minoration voulue.

On observe maintenant que, d'une part,

$$f(x) + c^* \leq f(x) + 1 + \beta V(x) - f(x) = 1 + \beta V(x), \quad (4.4.11)$$

alors que d'autre part,

$$\begin{aligned} f(x) + c^* &= \inf_{y \in \mathcal{X}} (f(x) + 1 + \beta V(y) - f(y)) & (4.4.12) \\ &\geq \inf_{y \in \mathcal{X}} (1 + \beta V(y) - \underbrace{\|f\|_\beta}_{=1} d_\beta(x, y)) \\ &\geq \inf_{y \in \mathcal{X}} (1 + \beta V(y) - 2 - \beta V(x) - \beta V(y)) \\ &\geq -1 - \beta V(x). \end{aligned}$$

Il suit de (4.4.11) et (4.4.12) que

$$|f(x) + c^*| \leq 1 + \beta V(x),$$

d'où l'inégalité inverse

$$\inf_{c \in \mathbb{R}} \|f + c\|_{1+\beta V} \leq \|f + c^*\|_{1+\beta V} \leq 1 = \|f\|_\beta.$$

Nous avons donc démontré (4.4.10), et par suite aussi (4.4.7). \square

4.5 Exercices

Exercice 4.5.1

On considère la marche aléatoire symétrique sur \mathbb{Z} .

1. Calculer $(\mathcal{L}V)(x)$ pour la fonction de Lyapounov $V(x) = x^2$. Expliciter la formule de Dynkin pour un temps n déterministe. Appliquer, si possible, les théorèmes de croissance sous-exponentielle, non-explosion, récurrence positive et de convergence.
2. Calculer $(\mathcal{L}V)(x)$ pour la fonction de Lyapounov $V(x) = |x|$. Comme au point précédent, expliciter la formule de Dynkin, et appliquer, si possible les différents théorèmes.
3. Que se passe-t-il pour la marche aléatoire asymétrique ?

Exercice 4.5.2

Soit $p \in]0, 1[$ et $q = 1 - p$. On considère la marche aléatoire sur \mathbb{N} de probabilités de transition

$$p_{xy} = \begin{cases} p & \text{si } y = x + 1, \\ q & \text{si } y = 0, \\ 0 & \text{sinon.} \end{cases}$$

Calculer $(\mathcal{L}V)(x)$ pour la fonction de Lyapounov $V(x) = |x|$. Que peut-on en déduire ?

Exercice 4.5.3

Pour $x \in \mathbb{N}$, on dénote par

$$\lfloor x \rfloor = \max\{y \in \mathbb{N} : y \leq x\}$$

la partie entière de x .

Soit $p \in [0, 1]$. On considère la chaîne de Markov sur $\mathbb{N} = \{0, 1, 2, \dots\}$ de probabilités de transition

$$p_{xy} = \begin{cases} p & \text{si } y = x + 1, \\ 1 - p & \text{si } y = \lfloor x/2 \rfloor, \\ 0 & \text{sinon.} \end{cases}$$

Pour $\alpha \in \mathbb{R}$, on pose $V(x) = e^{\alpha x}$.

1. Pour quelles valeurs de α la fonction V est-elle une fonction de Lyapounov ? On dira que ces valeurs de α sont *admissibles*.
2. Calculer $(\mathcal{L}V)(x)$, où \mathcal{L} est le générateur de la chaîne de Markov. On distinguera les cas x pair et x impair.
3. Pour quels p la chaîne est-elle à croissance sous-exponentielle ?
4. Déterminer une fonction $f_p(\alpha)$ telle que

$$(\mathcal{L}V)(x) \leq -f_p(\alpha)V(x)$$

pour tout $x \in \mathbb{N}^* = \{1, 2, \dots\}$.

5. Étudier la fonction $\alpha \mapsto f_p(\alpha)$ pour les valeurs de α admissibles : comportement aux bords du domaine, croissance/décroissance, convexité.
6. Pour quelles valeurs de p existe-t-il un α admissible tel que $f_p(\alpha) > 0$?
7. Pour quelles valeurs de p peut-on affirmer l'existence d'une unique probabilité invariante π telle que la loi de X_n converge exponentiellement vite vers π ?

Algorithmes MCMC

5.1 Méthodes Monte Carlo

On appelle *méthodes Monte Carlo* un ensemble d’algorithmes stochastiques, introduits dans les années 1940 par le mathématicien polonais Stanisław Ulam, permettant d’estimer numériquement des grandeurs pouvant être considérées comme des espérances. En voici pour commencer un exemple très simple.

Exemple 5.1.1: 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:

$$V = \{x \in \mathbb{R}^N : f_1(x) \geq 0, \dots, f_M(x) \geq 0\}.$$

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-espaces. Dans la suite nous supposons, 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 côté 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 non vide (Figure 5.1). Toutefois, 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 mesure $|\partial V|$ 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

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

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 . C’est ce qu’on appelle le *fléau de la dimension*.

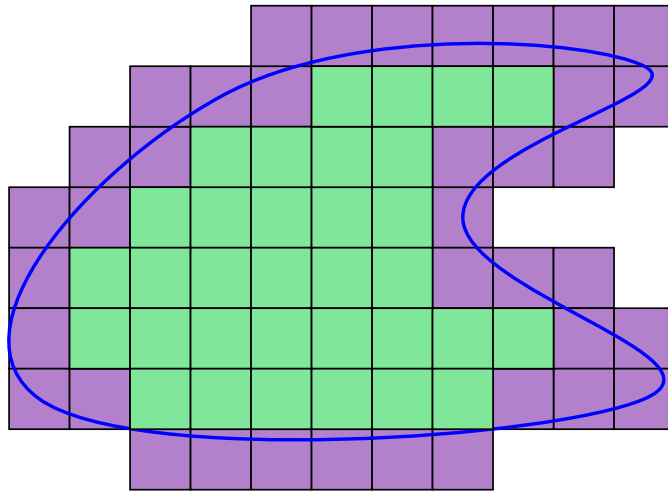


Figure 5.1 – Calcul d'une aire par discrétisation. L'aire est encadrée par $n_-\varepsilon^2$ et $n_+\varepsilon^2$, où n_- est le nombre de carrés de côté ε entièrement contenus dans le domaine, et n_+ est le nombre de carrés intersectant le domaine.

Une alternative intéressante pour les N grands est fournie par l'*algorithme de Monte Carlo*. Dans ce cas, on génère une suite $X_1, X_2, \dots, X_n, \dots$ 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 $\{0, 1, \dots, n_{\max}\}$ 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.

$$Y_n = \mathbb{1}_{X_n \in V}, \quad n = 1, 2, \dots$$

On aura

$$\mathbb{E}[Y_n] = \mathbb{P}\{X_n \in V\} = |V|.$$

Les moyennes empiriques

$$S_n = \frac{1}{n} \sum_{m=1}^n Y_m$$

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

$$\lim_{n \rightarrow \infty} \mathbb{P}\{|S_n - \mathbb{E}[S_n]| > \delta\} = 0$$

pour tout $\delta > 0$. Par conséquent, S_n devrait donner une bonne approximation du volume $|V|$ lorsque n est suffisamment grand.¹

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

$$\mathbb{P}\{|S_n - \mathbb{E}[S_n]| > \delta\} \leq \frac{\text{Var}(S_n)}{\delta^2} = \frac{\text{Var}(Y_1)}{\delta^2 n} < \frac{1}{\delta^2 n}, \quad (5.1.1)$$

¹La 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.

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

$$n > \frac{1}{\delta^2 \varepsilon}. \quad (5.1.2)$$

Comme pour chaque m , 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).

Remarque 5.1.2: Estimation d'erreur améliorée

L'estimation (5.1.1) est assez pessimiste, et peut être considérablement améliorée. Par exemple, le théorème central limite montre que

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \frac{(S_n - \mathbb{E}[S_n])^2}{\text{Var}(S_n)} > \eta^2 \right\} = \int_{|x| > \eta} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx, \quad (5.1.3)$$

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

$$\mathbb{P}\{|S_n - \mathbb{E}[S_n]| > \delta\} \simeq e^{-n\delta^2/2\text{Var}(Y_1)}.$$

Ceci permet d'améliorer le critère (5.1.2) en

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

d'où un nombre de calculs d'ordre $NM \log(1/\varepsilon)/\delta^2$. Cette estimation n'est pas une borne rigoureuse, contrairement à (5.1.2), parce qu'on n'a pas tenu compte de la vitesse de convergence dans (5.1.3), 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.

Exemple 5.1.3: Estimation d'un volume

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, en revanche, 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 Table 5.1 compare des coûts pour différentes valeurs de N .

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

N	Discrétisation	Monte Carlo BC	Monte Carlo TCL
1	10^5	10^{15}	$1,4 \cdot 10^{10}$
2	10^9	$2 \cdot 10^{15}$	$2,8 \cdot 10^{10}$
3	10^{13}	$3 \cdot 10^{15}$	$4,2 \cdot 10^{10}$
10	10^{41}	10^{16}	$1,4 \cdot 10^{11}$
100	10^{401}	10^{17}	$1,4 \cdot 10^{12}$
1000	10^{4001}	10^{18}	$1,4 \cdot 10^{13}$

Table 5.1 – Comparaison, pour différentes dimensions N et un nombre $M = 10$ d'inégalités, des coûts de calcul d'un volume avec précision $\delta = 10^{-4}$, par discrétisation (10^{4N+1}), par la méthode de Monte Carlo avec $\varepsilon = 10^{-6}$ avec l'estimation de Bienaymé–Chebychev ($10^{15}N$), et par la même méthode avec l'estimation basée sur le théorème central limite ($\log(10^6) \cdot 10^9 N$).

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.

Remarque 5.1.4: 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

$$F_U(u) = \mathbb{P}\{U \leq u\} = u \quad \text{pour } 0 \leq u \leq 1.$$

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

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

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

Exemple 5.1.5: Loi exponentielle

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

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

Exemple 5.1.6: Loi normale – Algorithme de Box–Muller

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

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

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 $1 - e^{-r^2/2}$, donc la densité

$re^{-r^2/2}$. Le couple (R, Φ) a donc la densité jointe $re^{-r^2/2}/(2\pi)$, et la formule de changement de variable montre que le couple (Y_1, Y_2) a la densité jointe $e^{-(y_1^2+y_2^2)/2}/(2\pi)$, 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.

5.2 Méthodes Monte Carlo par chaînes de Markov

Considérons le cas d'un espace probabilisé discret $(\mathcal{X}, \mathcal{P}(\mathcal{X}), \pi)$, où \mathcal{X} est un ensemble dénombrable, mais très grand. Par exemple, dans le cas du modèle d'Ising (voir la section 1.4), $\mathcal{X} = \{-1, 1\}^N$ est de cardinal 2^N , et on s'intéresse à des N grands, par exemple d'ordre 1000. La mesure de probabilité π est dans ce cas une application de \mathcal{X} vers $[0, 1]$ telle que la somme des $\pi(x)$ vaut 1.

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

$$\mathbb{E}^\pi[Y] = \sum_{x \in \mathcal{X}} Y(x)\pi(x). \quad (5.2.1)$$

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 loi π , puis de calculer la moyenne des $Y(X_m)$.

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

$$x \mapsto F_\pi(x) = \sum_{y \leq x} \pi(y). \quad (5.2.2)$$

Si U est une variable de loi uniforme sur $[0, 1]$, alors $F_\pi^{-1}(U)$ suit la loi π . Toutefois, en procédant de cette manière, on ne gagne rien, car le calcul des sommes (5.2.2) est aussi long que le calcul de la somme (5.2.1), 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 π et la variable aléatoire Y , à l'aide d'une chaîne de Markov sur \mathcal{X} , de probabilité invariante π .

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

$$Y_n = \sum_{x \in \mathcal{X}} Y(x) \mathbb{1}_{X_n=x}.$$

Considérons les moyennes empiriques

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

Le Théorème 2.5.4 permet d'écrire

$$\begin{aligned}
\lim_{n \rightarrow \infty} \mathbb{E}^\nu[S_n] &= \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}^\nu \left[\sum_{m=0}^{n-1} Y_m \right] \\
&= \sum_{x \in \mathcal{X}} Y(x) \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}^\nu \left[\sum_{m=0}^{n-1} \mathbb{1}_{X_m=x} \right] \\
&= \sum_{x \in \mathcal{X}} Y(x) \pi(x) \\
&= \mathbb{E}^\pi[Y].
\end{aligned}$$

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 central limite, 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 5.2.1: Réduction de variance partant de la probabilité invariante

Supposons la chaîne réversible, et de loi initiale égale à sa probabilité invariante. Soit ρ le rayon spectral associé à la chaîne. Alors

$$\text{Var}(S_n) \leq \frac{1}{n} \left(\frac{1+\rho}{1-\rho} \right) \text{Var}^\pi(Y). \quad (5.2.3)$$

Démonstration. Comme la chaîne démarre dans la probabilité invariante π , tous les Y_i ont même loi πY^{-1} , même s'ils ne sont pas indépendants. Il suit que

$$\begin{aligned}
\text{Var}(S_n) &= \frac{1}{n^2} \left[\sum_{m=0}^{n-1} \text{Var}(Y_m) + 2 \sum_{0 \leq p < q < n} \text{Cov}(Y_p, Y_q) \right] \\
&= \frac{1}{n} \text{Var}^\pi(Y) + \frac{2}{n^2} \sum_{m=1}^{n-1} (n-m) \text{Cov}(Y_0, Y_m), \quad (5.2.4)
\end{aligned}$$

en vertu du fait que (Y_p, Y_q) a la même loi que (Y_0, Y_{q-p}) . Or si $\mathbf{1} = (1, 1, \dots, 1)^\top$ on a

$$\begin{aligned}
\text{Cov}(Y_0, Y_m) &= \mathbb{E}^\pi \left[(Y_0 - \mathbb{E}^\pi[Y_0])(Y_m - \mathbb{E}^\pi[Y_m]) \right] \\
&= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{X}} (Y(x) - \mathbb{E}^\pi[Y])(Y(y) - \mathbb{E}^\pi[Y]) \underbrace{\mathbb{P}^\pi\{X_0 = x, X_m = y\}}_{=\pi(x)(P^m)_{xy}} \\
&= \sum_{x \in \mathcal{X}} \pi(x) (Y(x) - \mathbb{E}^\pi[Y]) [P^m(Y - \mathbb{E}^\pi[Y]\mathbf{1})]_x \\
&= \langle Y - \mathbb{E}^\pi[Y]\mathbf{1}, P^m(Y - \mathbb{E}^\pi[Y]\mathbf{1}) \rangle_\pi \\
&\leq \rho^m \langle Y - \mathbb{E}^\pi[Y]\mathbf{1}, Y - \mathbb{E}^\pi[Y]\mathbf{1} \rangle_\pi = \rho^m \text{Var}^\pi(Y).
\end{aligned}$$

Dans l'inégalité à la dernière ligne, nous avons utilisé le fait que $Y - \mathbb{E}^\pi[Y]\mathbf{1} \in \mathbf{1}_\perp$ puisque la somme des $\pi(x)(Y(x) - \mathbb{E}^\pi[Y])$ 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 (5.2.4), en majorant $n - m$ par n et en sommant une série géométrique. \square

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

$$n \geq \frac{\text{Var}^\pi(Y)}{\delta^2 \varepsilon} \left(\frac{1 + \rho}{1 - \rho} \right).$$

En pratique, on ne peut pas faire démarrer la chaîne exactement avec la probabilité 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 πY^{-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.

5.3 Algorithmes de type Metropolis

Nous avons vu comment estimer l'espérance d'une variable aléatoire Y à l'aide d'une chaîne de Markov de probabilité 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 probabilité invariante souhaitée.

Une méthode pour le faire a été développée à Los Alamos au début des années 1950, par Nicholas Metropolis, Arianna Rosenbluth, Marshall Rosenbluth, Augusta Teller et Edward Teller (plus connu pour le développement de la bombe à hydrogène)². L'algorithme devrait donc être appelé Metropolis–Rosenbluth–Rosenbluth–Teller–Teller, mais il est plus connu sous le nom d'algorithme de Metropolis, ou de Metropolis–Hastings, pour une forme plus générale développée par la suite par Wilfred Keith Hastings.

Le but de cet algorithme est d'échantillonner une *mesure de Gibbs*, de la forme

$$\pi(x) = \frac{e^{-\beta H(x)}}{Z_\beta}, \quad \text{où } Z_\beta = \sum_{x \in \mathcal{X}} e^{-\beta H(x)}.$$

Le paramètre β désigne la température inverse du système, et la fonction $H : \mathcal{X} \rightarrow \mathbb{R}$ associe à toute configuration $x \in \mathcal{X}$ son énergie. Nous en avons vu un exemple dans la section 1.4 avec le modèle d'Ising. Il s'agit donc de construire une chaîne de Markov sur \mathcal{X} admettant π comme probabilité invariante. Une manière simple d'approcher ce problème est de chercher une chaîne de Markov réversible. On cherche donc une matrice de transition P sur \mathcal{X} dont les éléments satisfont

$$\pi(x)p_{xy} = \pi(y)p_{yx} \tag{5.3.1}$$

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

$$\frac{p_{xy}}{p_{yx}} = e^{-\beta \Delta H(x, y)},$$

où

$$\Delta H(x, y) = H(y) - H(x)$$

est la différence d'énergie entre les états y et x . 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 $\mathbb{E}^\pi[Y]$.

²Il semble par ailleurs que la contribution principale de Nicholas Metropolis ait été de mettre à disposition du temps de calcul sur l'ordinateur MANIAC qu'il gérait.

L'algorithme de Metropolis 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 $x \sim x$). Une fois la relation \sim fixée, on choisit des probabilités de transition telles que

$$p_{xy} = \begin{cases} p_{yx} e^{-\beta\Delta H(x,y)} & \text{si } x \sim y, \\ 1 - \sum_{z \sim x} p_{xz} & \text{si } x = y, \\ 0 & \text{autrement.} \end{cases} \quad (5.3.2)$$

On remarque que la chaîne de Markov est irréductible à condition que la relation \sim le soit (deux états quelconques de \mathcal{X} peuvent être reliés par un chemin d'états équivalents par \sim). De plus, la chaîne de Markov est apériodique si $p_{xx} > 0$ pour tout $x \in \mathcal{X}$. Si \mathcal{X} est fini, la chaîne est automatiquement récurrente positive. Pour satisfaire la condition de réversibilité (5.3.1) lorsque $x \sim y$, une possibilité est de prendre

$$p_{xy} = \begin{cases} q & \text{si } H(y) \leq H(x), \\ q e^{-\beta\Delta H(x,y)} & \text{si } H(y) > H(x), \end{cases} \quad (5.3.3)$$

où $q \in]0, 1]$ est une constante qui contrôle la vitesse de l'algorithme. Elle doit être choisie assez petite pour que p_{xx} soit positif. Ce choix revient à effectuer la transition avec probabilité q si elle décroît l'énergie, et de ne l'effectuer qu'avec probabilité $q e^{-\beta\Delta H(x,y)}$ si elle fait croître l'énergie. Une autre possibilité est de choisir

$$p_{xy} = \frac{q}{1 + e^{\beta\Delta H(x,y)}}.$$

Remarque 5.3.1: q non constant

Au lieu de choisir un q constant, on peut également choisir des coefficients q_{xy} dépendant de x et y , satisfaisant $q_{xy} = q_{yx}$, et avec $\sum_{y \sim x} q_{xy}$ assez petit pour avoir $p_{xx} > 0$. Cela peut permettre, dans certains cas, d'accélérer la convergence de l'algorithme.

Nous allons illustrer cette méthode dans le cas du modèle d'Ising, mais on voit facilement comment la généraliser à d'autres systèmes. Rappelons que dans le cas du modèle d'Ising (voir la section 1.4), l'univers est donné par $\mathcal{X} = \{-1, 1\}^\Lambda$, où Λ est un sous-ensemble (supposé ici de cardinal fini N) de \mathbb{Z}^d . L'énergie est donnée par

$$H(x) = - \sum_{i,j \in \Lambda: \|i-j\|=1} x_i x_j - h \sum_{i \in \Lambda} x_i,$$

où $h \in \mathbb{R}$ est le champ magnétique. L'objectif est de calculer l'espérance de la variable aimantation, donnée par

$$m(x) = \sum_{i \in \Lambda} x_i.$$

Le choix de la relation symétrique \sim sur \mathcal{X} dépend de la physique que l'on souhaite modéliser. Les deux choix les plus courants sont

- la *dynamique de Glauber*, qui consiste à choisir $x \sim y$ si et seulement si les deux configurations x et y diffèrent en exactement un point de Λ ; on parle de dynamique de renversement de spin;

- la *dynamique de Kawasaki*, qui consiste à choisir $x \sim y$ si et seulement si y est obtenue en intervertissant deux composantes de x ; 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.

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 $R_k(x)$ dénote la configuration obtenue en renversant le spin numéro k de x , on aura

$$\Delta H(x, R_k(x)) = 2x_k \left[\sum_{j: \|j-k\|=1} x_j + h \right], \quad (5.3.4)$$

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 de la manière suivante (avec $q = 1/N$) :

1. Étape d'initialisation :

- choisir une configuration initiale X_0 (de préférence telle que $\pi(X_0)$ ne soit pas trop petit);
- calculer $m_0 = m(X_0)$ (nécessite N calculs);
- calculer $H(X_0)$ (nécessite de l'ordre de dN calculs);
- poser $S = m_0$.

2. Etape d'itération : Pour $n = 0, 1, \dots, n_{\max} - 1$,

- choisir un spin k au hasard uniformément dans Λ ;
- calculer $\Delta H(X_n, y)$, où $y = R_k(X_n)$ est obtenu en renversant le spin choisi;
- si $\Delta H(X_n, y) \leq 0$, poser $X_{n+1} = y$;
- si $\Delta H(X_n, y) > 0$, soit B_n une variable de Bernoulli de paramètre $e^{-\beta \Delta H(X_n, y)}$, qui est indépendante des autres aléas; poser $X_{n+1} = y$ si $B_n = 1$, et $X_{n+1} = X_n$ si $B_n = 0$;
- si $B_n = 1$ (on a renversé le spin k), alors $m_{n+1} = m_n + 2(X_n)_k$, sinon $m_{n+1} = m_n$;
- ajouter m_{n+1} à S .

Le quotient $S/(n+1)$ converge alors vers $\mathbb{E}^\pi[m]$, avec une vitesse déterminée par (5.2.3). La seule quantité difficile à estimer est le trou spectral $1 - \rho$. Nous allons donner un exemple de son estimation à l'aide de la méthode des fonctions de Lyapounov dans la section suivante.

5.4 Modèle d'Ising sur le cercle discret

Nous donnons dans cette section quelques exemples d'estimations de vitesse de convergence pour le modèle d'Ising sur $\Lambda = \mathbb{Z}/N\mathbb{Z}$, pour la dynamique de Glauber et une matrice de transition définie par (5.3.3). Cela signifie que l'on considère N spins alignés, avec conditions aux bords périodiques : on identifie $i = 0$ avec $i = N$ et $i = -1$ avec $i = N - 1$, de manière que chaque spin $i \in \Lambda$ ait exactement deux voisins $i - 1$ et $i + 1$. Soit $N_\pm(x)$ le nombre de spins de la configuration x valant ± 1 . Alors on a

$$N_+(x) + N_-(x) = N, \quad N_+(x) - N_-(x) = m(x), \quad (5.4.1)$$

ce qui implique $m(x) = N - 2N_-(x)$. Soit par ailleurs

$$I(x) = \left| \{i \in \Lambda : x_{i+1} \neq x_i\} \right|$$

le nombre d'« interfaces » de x , c'est-à-dire le nombre de fois que la fonction $i \mapsto x_i$ change de signe en faisant le tour du cercle discret. Alors on a

$$\sum_{i,j \in \Lambda : \|i-j\|=1} x_i x_j = \sum_{i \in \Lambda} x_i x_{i+1} = N - 2I(x).$$

Par conséquent, l'énergie du modèle d'Ising peut également s'écrire

$$\begin{aligned} H(x) &= 2I(x) - hm(x) - N \\ &= 2I(x) + 2hN_-(x) - N(1+h). \end{aligned}$$

Les constantes $-N$ et $-N(1+h)$ n'ont pas d'incidence sur la dynamique de Glauber, qui ne fait intervenir que des différences d'énergie entre configurations. Dénotons par \boxplus la configuration dont tous les spins valent $+1$, et par \boxminus celle dont tous les spins valent -1 . Alors on a

$$H(\boxplus) = -N(1+h), \quad H(\boxminus) = -N(1-h).$$

Dans la suite, on supposera que N est pair, et que $0 < h \leq 1$. Dans ce cas, la configuration d'énergie minimale est \boxplus , et on vérifie qu'on a deux configurations d'énergie maximale égale à N , données par

$$(1, -1, 1, -1, \dots) \quad \text{et} \quad (-1, 1, -1, 1, \dots).$$

Afin de pouvoir appliquer l'approche par fonctions de Lyapounov, nous commençons par déterminer le générateur.

Proposition 5.4.1: Générateur pour la dynamique de Glauber

Pour tout $x \in \mathcal{X}$, notons

$$\begin{aligned} A_+(x) &= \{y \in \mathcal{X} : y \sim x, H(y) > H(x)\}, \\ A_0(x) &= \{y \in \mathcal{X} : y \sim x, H(y) = H(x)\}, \\ A_-(x) &= \{y \in \mathcal{X} : y \sim x, H(y) < H(x)\}. \end{aligned}$$

Alors pour toute fonction $V : \mathcal{X} \rightarrow \mathbb{R}$, on a

$$(\mathcal{L}V)(x) = q \left[\sum_{y \in A_-(x) \cup A_0(x)} (V(y) - V(x)) + \sum_{y \in A_+(x)} (V(y) - V(x)) e^{-\beta \Delta H(x,y)} \right]. \quad (5.4.2)$$

Démonstration. Il suit de (5.3.2) et (5.3.3) que

$$\begin{aligned} (\mathcal{L}V)(x) &= \sum_{y \in A_- \cup A_0} p_{xy} V(y) + \sum_{y \in A_+} p_{xy} V(y) + p_{xx} V(x) - V(x) \\ &= q \sum_{y \in A_- \cup A_0} V(y) + q \sum_{y \in A_+} e^{-\beta \Delta H(x,y)} V(y) - q \left[\sum_{y \in A_- \cup A_0} 1 + \sum_{y \in A_+} e^{-\beta \Delta H(x,y)} \right] V(x), \end{aligned}$$

d'où le résultat, en regroupant les termes. □

Commençons par étudier le cas particulier $\beta = 0$, qui correspond à une température infinie. Dans ce cas, toutes les transitions permises ont la même probabilité $q = 1/N$, et le système effectue une marche aléatoire symétrique sur l'hypercube de dimension N . La probabilité invariante est simplement la mesure uniforme.

Remarquons que $m(X_n)$ satisfait

$$m(X_{n+1}) = \begin{cases} m(X_n) + 2 & \text{avec probabilité } \frac{N_-(X_n)}{N} = \frac{1}{2} - \frac{m(X_n)}{2N}, \\ m(X_n) - 2 & \text{avec probabilité } \frac{N_+(X_n)}{N} = \frac{1}{2} + \frac{m(X_n)}{2N}. \end{cases} \quad (5.4.3)$$

Comme ces probabilités de transition ne dépendent que de $m(X_n)$, la suite des $M_n = m(X_n)$ est une chaîne de Markov sur $\mathcal{M} = \{-N, -N+2, \dots, N-2, N\}$. En fait, $(M_n + N)/2$ n'est autre que le modèle d'Ehrenfest à N boules, qui admet la loi binomiale de paramètres $(N, \frac{1}{2})$ comme probabilité invariante. Ceci donne l'idée d'utiliser $V(x) = m(x)^2$ comme fonction de Lyapounov. En effet, les valeurs de m proches de 0 sont plus probables que les valeurs proches de $\pm N$. On obtient facilement la condition de dérive géométrique suivante.

Proposition 5.4.2: Condition de dérive géométrique pour $V = m^2$

Si $\beta = 0$, la fonction de Lyapounov $V(x) = m(x)^2$ satisfait

$$(\mathcal{L}V)(x) = -cV(x) + d, \quad \text{avec } c = \frac{4}{N} \text{ et } d = 4. \quad (5.4.4)$$

Démonstration. Par la Proposition 5.4.1 avec $\beta = 0$, on a

$$\begin{aligned} (\mathcal{L}V)(x) &= q \sum_{y \sim x} [m(y)^2 - m(x)^2] \\ &= q [N_-(x)((m(x)+2)^2 - m(x)^2) + N_+(x)((m(x)-2)^2 - m(x)^2)] \\ &= 4q [N - V(x)] \end{aligned}$$

en vertu de (5.4.1). Le résultat suit du fait que $q = \frac{1}{N}$. \square

Pour appliquer la condition de minoration (4.4.2) du Théorème 4.4.1, il nous faut choisir $R > 2d/c = 2N$, donc par exemple $R = 4N$. Ainsi, on aura

$$K = \{x \in \mathcal{X} : m(x)^2 < 4N\} = \{x \in \mathcal{X} : |m(x)| < 2\sqrt{N}\}.$$

On constate que la condition de minoration ne peut pas être satisfaite, car p_{xy} est non nulle seulement si $x \sim y$, ce qui n'est pas le cas pour tous les $x, y \in K$. Une manière de résoudre ce problème est de considérer une puissance de la matrice de transition.

Lemme 5.4.3: Condition de dérive pour processus accéléré

Soit $(X_n)_{n \geq 0}$ une chaîne de Markov dont le générateur satisfait

$$(\mathcal{L}V)(x) \leq -cV(x) + d \quad \forall x \in \mathcal{X}$$

pour une fonction de Lyapounov V et des constantes $c > 0$ et $d \geq 0$. Alors pour tout $T \in \mathbb{N}^*$, le générateur \mathcal{L}^T du processus accéléré $(X_{nT})_{n \geq 0}$ satisfait

$$(\mathcal{L}^T V)(x) \leq -[1 - (1-c)^T]V(x) + [1 - (1-c)^T] \frac{d}{c} \quad \forall x \in \mathcal{X}.$$

Démonstration. Pour $x \in \mathcal{X}$ et $n \in \mathbb{N}$, soit

$$g_n(x) = (\mathcal{P}^n V)(x) = \mathbb{E}^x[V(X_n)].$$

Alors on obtient, comme dans la démonstration de la formule de Dynkin,

$$g_{n+1}(x) \leq -(1-c)g_n(x) + d.$$

En utilisant $g_0(x) = V(x)$ comme condition initiale, on obtient facilement par récurrence sur n

$$g_n(x) \leq -(1-c)^n V(x) + [1 - (1-c)^n] \frac{d}{c} \quad \forall n \in \mathbb{N}^*.$$

Le résultat suit alors du fait que $(\mathcal{L}^T V)(x) = g_T(x) - V(x)$. \square

On peut simplifier l'étude en ne considérant que la chaîne de Markov $(M_n)_{n \geq 0}$. Notons $\hat{\mathcal{L}}$ son générateur, et $\hat{\pi}$ sa probabilité invariante, qui se déduit de la loi binomiale. Alors (5.4.4) implique que si $V(m) = m^2$, on a

$$(\hat{\mathcal{L}}V)(m) = -cV(m) + d, \quad \text{avec } c = \frac{4}{N} \text{ et } d = 4.$$

Cette relation peut aussi être déduite directement de (5.4.3). On obtient alors le résultat de convergence suivant, en appliquant le Théorème 4.4.1 à une puissance convenablement choisie de la matrice de transition de $(M_n)_{n \geq 0}$.

Proposition 5.4.4: Convergence pour fonctions de l'aimantation lorsque $\beta = 0$

Il existe des constantes $C > 0$ et $\bar{\gamma} < 1$, indépendantes de N , telles que pour toute fonction test $f : \mathcal{M} \rightarrow \mathbb{R}$, on ait

$$|\mathbb{E}^m[f(M_n)] - \hat{\pi}(f)| \leq C(1+m^2)\bar{\gamma}^{n/N} \|f - \hat{\pi}(f)\|_{1+m^2}. \quad (5.4.5)$$

Démonstration. Considérons le processus $(M_n^T)_{n \geq 0}$ accéléré d'un facteur $T \in \mathbb{N}^*$, défini comme $M_n^T = M_{Tn}$. En procédant comme dans la démonstration du Lemme 5.4.3, on obtient

$$(\hat{\mathcal{L}}^T V)(m) = -c_T V(m) + d_T, \quad \text{avec } c_T = 1 - (1-c)^T \text{ et } d_T = c_T \frac{d}{c}.$$

On remarque que $2d_T/c_T = 2d/c = 2N$. On peut donc utiliser la même valeur $4N$ pour R , et

$$K = \{m \in \mathcal{M} : |m| < 2\sqrt{N}\}.$$

En revanche, on a gagné en ce qui concerne la condition de minoration. En effet, pour m d'ordre \sqrt{N} , l'évolution de M_n est essentiellement une marche aléatoire symétrique (plus précisément, les probabilités de transitions sont minorées par celles d'une marche aléatoire de paramètre $\frac{1}{2} - \mathcal{O}(N^{-1/2})$). La variance d'une telle marche aléatoire au temps T est d'ordre T . Il suit du théorème central limite (ou du théorème de de Moivre-Laplace) que

$$\mathbb{P}^m \{M_T = p\} \geq \frac{a}{\sqrt{T}} \quad \forall m, p \in K$$

pour une constante a indépendante de T . Choisissons alors pour ν la probabilité uniforme sur K . Prenons $T = bN$ pour un b positif. Comme le cardinal de K est d'ordre \sqrt{N} , on peut choisir b de telle manière que

$$\inf_{m \in K} \mathbb{P}^m \{M_T = p\} \geq \frac{1}{2} \nu(p) \quad \forall p \in K.$$

Comme $\nu(m) = 0$ pour tout $m \notin K$, la condition de minoration (4.4.2) est satisfaite avec $\alpha = \frac{1}{2}$. On vérifie alors que les choix

$$\alpha_0 = \frac{1}{4}, \quad \gamma_0 = 1 - \frac{c_T}{4}$$

satisfont (4.4.4), avec $\gamma = 1 - c_T$. On trouve ensuite à l'aide de (4.4.5)

$$\beta = \frac{1}{4Nc_T}, \quad R\beta = \frac{1}{c_T}, \quad \bar{\gamma} = \max\left\{\frac{1}{4}, 1 - \frac{c_T}{4(1+2c_T)}\right\}.$$

Comme de plus

$$\log(1 - c_T) = T \log\left(1 - \frac{4}{N}\right) = -\frac{4T}{N} + \mathcal{O}\left(\frac{T}{N^2}\right) = -4b + \mathcal{O}\left(\frac{1}{N}\right),$$

c_T converge vers une limite indépendante de N lorsque $N \rightarrow \infty$ (qui vaut $1 - e^{-4b}$).

Par le Théorème 4.4.1, l'espérance de $f(M_n^T)$ converge vers sa limite $\hat{\pi}(f)$ exponentiellement vite, avec un taux $\bar{\gamma}^n$. Le résultat suit en revenant au temps non accéléré, quitte à remplacer $\bar{\gamma}$ par $\bar{\gamma}^{1/b}$. \square

La borne (5.4.5) montre que pour approcher $\hat{\pi}(f)$ à une distance d'ordre δ , il faut choisir un n d'ordre $N \log(1/\delta)$. Cette convergence est relativement rapide. Bien entendu, comme la probabilité invariante $\hat{\pi}$ est connue explicitement, on n'a pas besoin d'estimer $\hat{\pi}(f)$, on peut la calculer directement avec un coût N . On notera aussi que ce résultat ne marche que pour des fonctions de l'aimantation. Il n'affirme rien, par exemple, sur des fonctions qui dépendraient de $I(x)$.

Le second cas particulier que nous allons considérer est celui où $0 < h \leq 1$ et β est assez grand (dans un sens à préciser plus bas), donc où la température est assez basse. Pour $h > 0$, l'état d'énergie minimale est \boxplus . Un candidat pour une fonction de Lyapounov est la différence d'énergie

$$V(x) = H(x) - H(\boxplus) = 2I(x) + 2hN_-(x), \quad (5.4.6)$$

qui est bien positive ou nulle. Nous commençons par donner une majoration générale de $\mathcal{L}H$ (comme $\mathcal{L}c = 0$ pour toute constante c , cela fournit également une majoration de $\mathcal{L}V$).

Lemme 5.4.5: Majoration de $\mathcal{L}H$

Supposons $\beta \geq \frac{1}{2h}$. Alors pour tout $x \in \mathcal{X}$, on a

$$(\mathcal{L}H)(x) \leq -2qh|A_-(x)| + 2qhe^{-2\beta h}|A_+(x)|. \quad (5.4.7)$$

Démonstration. Il suit de l'expression (5.3.4) de ΔH que

$$\Delta H(x, y) \neq 0 \quad \Rightarrow \quad |\Delta H(x, y)| \geq 2h.$$

En effet, la somme des x_j est nécessairement un entier pair, et le minimum de $|\Delta H(x, y)|$ est atteint lorsque cette somme vaut 0. Il suit que la première somme dans (5.4.2) est majorée par $-2h|A_-(x)|$, puisque les termes avec $y \in A_0(x)$ sont nuls. La seconde somme s'écrit

$$\sum_{y \in A_+(x)} \Delta H(x, y) e^{-\beta \Delta H(x, y)}.$$

Or la fonction $u \mapsto f(u) = u e^{-\beta u}$ est croissante sur $[0, 1/\beta]$ et décroissante sur $[1/\beta, \infty[$. Comme $\Delta H(x, y)$ est minoré par $2h$ dans cette somme, et que $2h \geq 1/\beta$, celle-ci est inférieure ou égale à $2he^{-2\beta h}|A_+|$. \square

La majoration (5.4.7) exprime le fait que si β est assez grand, alors l'énergie a tendance à diminuer, sauf dans le cas particulier où $|A_-| = 0$, c'est-à-dire si x n'a pas de configuration voisine d'énergie inférieure. Cela n'arrive que si $x \in \{\boxplus, \boxminus\}$.

Proposition 5.4.6: Condition de dérive géométrique pour $V = H - H(\boxplus)$

Si $\beta \geq \frac{1}{2h}$, la fonction de Lyapounov (5.4.6) satisfait la condition de dérive géométrique

$$(\mathcal{L}V)(x) \leq -cV(x) + d \quad \text{avec } c = \frac{h}{2N} \text{ et } d = 2he^{-2\beta h} + h^2. \quad (5.4.8)$$

Démonstration. Nous allons considérer séparément les cas $x = \boxplus$, $x = \boxminus$, et $x \in \mathcal{X} \setminus \{\boxplus, \boxminus\}$.

- Si $x = \boxplus$, alors $|A_-(x)| = 0$, puisque retourner un spin augmente toujours l'énergie. Par conséquent, (5.4.7) implique

$$(\mathcal{L}V)(\boxplus) \leq 2he^{-2\beta h},$$

où nous avons utilisé $|A_+(\boxplus)| = N$ et $q = 1/N$. Comme $V(\boxplus) = 0$, la borne (5.4.8) est bien vérifiée.

- Si $x = \boxminus$, alors on a

$$(\mathcal{L}V)(\boxminus) \leq 2he^{-2\beta h}, \quad V(\boxminus) = 2hN.$$

Par conséquent, (5.4.8) est vérifié puisque $d = 2he^{-2\beta h} + 2hNc$.

- Pour tous les autres x , on a toujours $|A_-(x)| \geq \frac{1}{2}I(x)$. En effet, pour chaque interface, changer le spin -1 qui se trouve d'un côté de l'interface en $+1$ diminue l'énergie de la configuration. Le facteur $\frac{1}{2}$ vient du fait que ce spin -1 peut être compris entre deux interfaces. Ainsi,

$$\begin{aligned} (\mathcal{L}V)(x) &\leq -qhI(x) + 2he^{-2\beta h} \\ &= -\frac{1}{2}qhV(x) + qh^2N_-(x) + 2he^{-2\beta h} \\ &\leq -\frac{1}{2}qhV(x) + qh^2N + 2he^{-2\beta h}. \end{aligned}$$

Ceci montre que (5.4.8) est bien vérifié, puisque $q = 1/N$. □

Si nous supposons $\beta \geq \frac{1}{2h} \log(\frac{2}{h})$, alors on peut prendre $d = 2h^2$. Dans la condition de minoration, il faut donc prendre $R > 8hN$, de sorte que $K = \{x \in \mathcal{X} : V(x) < R\}$ contient beaucoup d'états (selon la valeur de h , il peut même arriver que $K = \mathcal{X}$). Il nous faut donc à nouveau accélérer le temps afin de pouvoir appliquer le Théorème 4.4.1.

Nous n'allons pas donner une analyse détaillée, mais un argument heuristique. Si l'on prend $\nu = \delta_{\boxplus}$, on aura, pour la chaîne de Markov accélérée d'un facteur T ,

$$\alpha = \inf_{x \in K} \mathbb{P}^x \{X_T = \boxplus\}.$$

La question est de savoir comment choisir T pour que α soit d'ordre 1, disons $\alpha = \frac{1}{2}$.

On s'attend à ce que la transition la plus difficile soit celle de \boxminus vers \boxplus . La manière la plus économique de faire cette transition est de renverser d'abord un spin quelconque, puis de renverser des spins adjacents, un par un, jusqu'à atteindre \boxplus (Figure 5.2). On vérifie que seule la première transition fait augmenter l'énergie. Quitte à augmenter encore β , on peut modéliser la transition en négligeant tout renversement de spin non optimal, faisant augmenter l'énergie plus que nécessaire. On aboutit alors à la chaîne de Markov de la Figure 5.3. En effet, la probabilité de la première transition est de $Nqe^{-\beta\Delta H(\boxplus, R_k(\boxplus))}$, où $\Delta H(\boxplus, R_k(\boxplus)) = 4 - 2h$ ne dépend pas de k , puisqu'on peut retourner n'importe lequel des N spins. Toutes les transitions suivantes ont la même probabilité $2q$, car on peut choisir de quel côté la goutte croît.

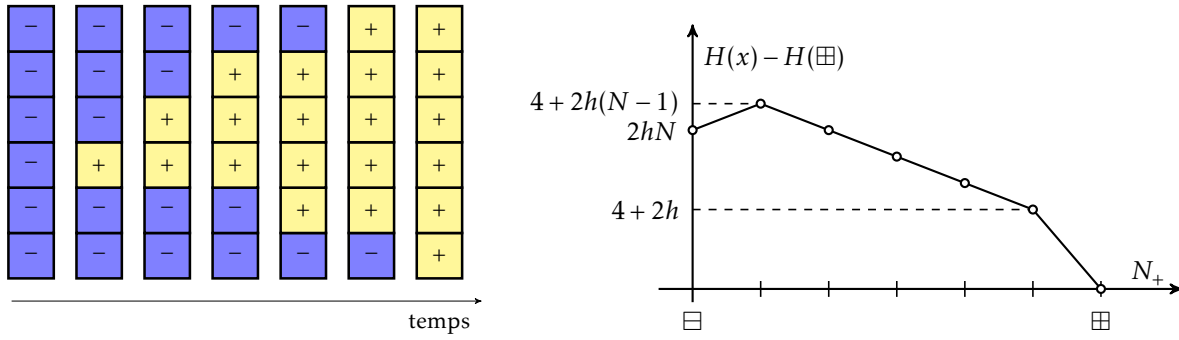


Figure 5.2 – À gauche, exemple d'une transition optimale de l'état \square vers l'état \oplus par croissance d'une goutte, pour $N = 6$. À droite, valeur de la différence d'énergie $H(x) - H(\oplus)$ en fonction de N_+ .

Soit alors $f(y) = \mathbb{E}^y[\tau_N]$. Pour $y \in \{2, \dots, N-1\}$, cette fonction satisfait

$$f(y) = 2qf(y+1) + (1-2q)f(y) + 1.$$

Avec la condition initiale $f(y) = 0$, on trouve

$$\mathbb{E}^2[\tau_N] = \frac{N-2}{2q} = \frac{N(N-2)}{2}. \quad (5.4.9)$$

Par ailleurs, pour $y \in \{0, 1\}$ on obtient les équations

$$\begin{aligned} f(0) &= e^{-\beta\Delta H} f(1) + (1 - e^{-\beta\Delta H}) f(0) + 1 \\ f(1) &= qf(0) + (1 - 3q)f(1) + 2qf(2) + 1. \end{aligned}$$

En résolvant ce système pour $f(0)$ et $f(1)$ (ce qui revient à calculer la matrice fondamentale de la chaîne de Markov absorbée en 2), on obtient

$$f(0) = \frac{3}{2} e^{\beta\Delta H} + \frac{1}{2q} + f(2).$$

En combinant ceci avec (5.4.9), on aboutit finalement, dans cette approximation, à

$$\mathbb{E}^{\square}[\tau_{\oplus}] \simeq \frac{3}{2} e^{2\beta(2-h)} + \frac{N(N-1)}{2}.$$

L'inégalité de Markov implique alors

$$\mathbb{P}^{\square}\{\tau_{\oplus} \geq k\} \leq \frac{\mathbb{E}^{\square}[\tau_{\oplus}]}{k},$$

Par conséquent, en choisissant $T = 2\mathbb{E}^{\square}[\tau_{\oplus}]$, on aura $\alpha = \frac{1}{2}$. Soient alors c_T et d_T les constantes données par le Lemme 5.4.3. Comme

$$\log(1 - c_T) = T \log\left(1 - \frac{h}{2N}\right) = -\frac{hT}{2N} \left(1 + \mathcal{O}\left(\frac{h^2}{N}\right)\right),$$

on a $c_T = 1 - \mathcal{O}(e^{-hT/(2N)})$. Un choix possible de paramètres est

$$\alpha_0 = \frac{1}{4}, \quad R = 16hN, \quad \gamma_0 = 1 - \frac{1}{4}c_T,$$

ce qui conduit à $R\beta = 1/c_T$ et

$$\bar{\gamma} = \frac{11}{12} + \mathcal{O}(e^{-hT/(2N)}).$$

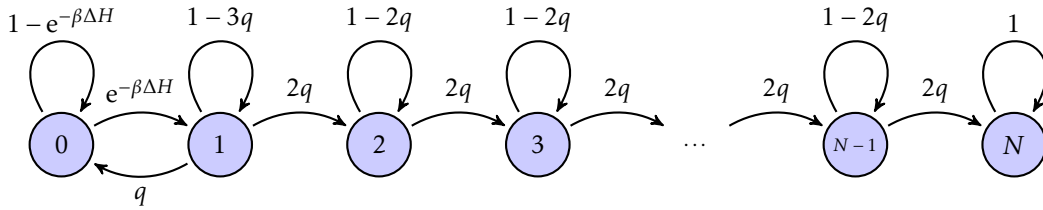


Figure 5.3 – Chaîne de Markov modélisant une transition optimale de l'état \boxplus vers l'état \boxminus , par croissance d'une goutte de spins +1. La valeur de ΔH pour la transition entre les états 0 et 1 est $\Delta H(\boxplus, R_k(\boxplus)) = 4 - 2h$. Les états sont numérotés selon le nombre N_+ de spins valant +1.

Le point important est que $1 - \bar{\gamma}$ est minoré par une quantité indépendante de N . On s'attend donc à une convergence de la forme

$$|\mathbb{E}^x[f(X_n)] - \pi(f)| \leq C(1 + V(x))\bar{\gamma}^{n/T} \|f - \pi(f)\|_{1+V}$$

avec $T = N(N-1) + 2e^{\beta(2-h)}$. Pour atteindre une précision δ , il faut choisir n d'ordre $T \log(1/\delta)$. Si β n'est pas trop grand, ce temps varie comme N^2 . Toutefois, si $e^{\beta(2-h)}$ dépasse N^2 , c'est ce terme qui détermine le temps de convergence. L'algorithme converge donc moins rapidement à très faible température, en raison du temps nécessaire à renverser le premier spin de la configuration \boxminus .

Dans le cas du modèle d'Ising sur $\Lambda \subset \mathbb{Z}^2$, la situation est moins favorable. En effet, en partant de la configuration \boxminus , il faut d'abord créer une goutte de spins +1 d'une certaine taille avant que l'énergie se mette à diminuer en approchant \boxplus . Dans ce cas, il existe des algorithmes alternatifs, tels que l'algorithme dit de Swendsen–Wang, qui convergent beaucoup mieux. Au lieu de retourner un seul spin à la fois, cet algorithme retourne des groupes de spins bien choisis.

Partie II

**Chaînes de Markov
à espace continu**

Définition et exemples de chaînes de Markov à espace continu

Dans ce chapitre, nous examinons comment on peut étendre la théorie des chaînes de Markov sur un ensemble \mathcal{X} dénombrable à des ensembles infinis non dénombrables, plus précisément des sous-ensembles ouverts de \mathbb{R}^d . Une grande partie des concepts du cas discret (évolution de la loi de X_n , probabilité invariante) peuvent être transposés à cette situation de manière assez directe, essentiellement en « remplaçant les sommes par des intégrales ». Il faut être un peu prudent, toutefois, en généralisant les notions de récurrence et de récurrence positive. Nous aborderons cette question dans le chapitre suivant.

6.1 Définitions et notations

Soit $\mathcal{X} \subset \mathbb{R}^d$ un ouvert. Cet ensemble est muni de la *tribu des boréliens*, qui contient en particulier tous les ouverts de \mathcal{X} . Voici d’abord la généralisation de concept de matrice stochastique à cette situation.

Définition 6.1.1: Densité de probabilité, noyau markovien à densité

- Une *densité de probabilité* ν sur \mathcal{X} est une application $\nu : \mathcal{X} \rightarrow \mathbb{R}_+ = [0, \infty[$, continue par morceaux, et satisfaisant

$$\int_{\mathcal{X}} \nu(x) dx = 1 .$$

- Un *noyau markovien à densité* sur \mathcal{X} est une application $p : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_+$, continue par morceaux, satisfaisant

$$\int_{\mathcal{X}} p(x, y) dy = 1 \quad \forall x \in \mathcal{X} .$$

Dans la suite, nous utiliserons la même notation pour la mesure de probabilité associée à la densité ν . Cela revient à poser

$$\nu(A) = \int_A \nu(x) dx$$

pour tout borélien $A \subset \mathcal{X}$. La généralisation naturelle de la notion de chaîne de Markov est alors la suivante.

Définition 6.1.2: Chaîne de Markov sur un ouvert \mathcal{X} de \mathbb{R}^d

Soit ν une densité de probabilité sur \mathcal{X} , et p un noyau markovien à densité. Une *chaîne de Markov* (homogène en temps) sur \mathcal{X} , de loi initiale ν et de noyau de transition p , est une suite $(X_n)_{n \geq 0}$ de variables aléatoires à valeurs dans \mathcal{X} , telles que $\mathbb{P}\{X_0 \in A\} = \nu(A)$ pour tout borélien $A \subset \mathcal{X}$, et satisfaisant la *propriété de Markov*

$$\begin{aligned} \mathbb{P}\{X_n \in A | X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}\} &= \mathbb{P}\{X_n \in A | X_{n-1} = x_{n-1}\} \\ &= \int_A p(x_{n-1}, x_n) dx_n \end{aligned} \quad (6.1.1)$$

pour tout $n \geq 1$, tout choix de $x_0, \dots, x_{n-1} \in \mathcal{X}$, et tout borélien $A \subset \mathcal{X}$.

Comme la probabilité qu'une variable aléatoire à densité prenne une valeur particulière vaut 0, il n'est pas immédiatement évident que les probabilités conditionnelles dans (6.1.1) sont bien définies. Il faut en fait les interpréter à l'aide de densités conditionnelles. Pour ce faire, soit

$$\mathcal{B}_\varepsilon(x_0) = \{x \in \mathcal{X} : \|x - x_0\| < \varepsilon\}$$

la boule ouverte de centre x_0 et de rayon ε (où $\|\cdot\|$ est la norme Euclidienne). On définit alors

$$\begin{aligned} \mathbb{P}\{X_1 \in A | X_0 = x_0\} &= \lim_{\varepsilon \rightarrow 0} \mathbb{P}\{X_1 \in A | X_0 \in \mathcal{B}_\varepsilon(x_0)\} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{\mathbb{P}\{X_1 \in A, X_0 \in \mathcal{B}_\varepsilon(x_0)\}}{\mathbb{P}\{X_0 \in \mathcal{B}_\varepsilon(x_0)\}}. \end{aligned}$$

Si $f(x_0, x_1)$ désigne la densité jointe de X_0 et X_1 , alors on a

$$\begin{aligned} \mathbb{P}\{X_0 \in \mathcal{B}_\varepsilon(x_0)\} &= \int_{\mathcal{B}_\varepsilon(x_0)} \nu(x) dx, \\ \mathbb{P}\{X_1 \in A, X_0 \in \mathcal{B}_\varepsilon(x_0)\} &= \int_A \int_{\mathcal{B}_\varepsilon(x_0)} f(x, x_1) dx dx_1, \end{aligned}$$

de sorte que

$$\mathbb{P}\{X_1 \in A | X_0 = x_0\} = \int_A \lim_{\varepsilon \rightarrow 0} \frac{\int_{\mathcal{B}_\varepsilon(x_0)} f(x, x_1) dx}{\int_{\mathcal{B}_\varepsilon(x_0)} \nu(x) dx} dx_1 = \int_A \frac{f(x_0, x_1)}{\nu(x_0)} dx_1.$$

La dernière égalité suit du théorème de la valeur moyenne, qui montre que

$$\lim_{\varepsilon \rightarrow 0} \frac{1}{|\mathcal{B}_\varepsilon(x_0)|} \int_{\mathcal{B}_\varepsilon(x_0)} \nu(x) dx = \nu(x_0), \quad \lim_{\varepsilon \rightarrow 0} \frac{1}{|\mathcal{B}_\varepsilon(x_0)|} \int_{\mathcal{B}_\varepsilon(x_0)} f(x, x_1) dx = f(x_0, x_1).$$

En comparant avec (6.1.1) pour $n = 1$, il vient

$$\frac{f(x_0, x_1)}{\nu(x_0)} = p(x_0, x_1). \quad (6.1.2)$$

Le noyau markovien $p(x_0, x_1)$ s'interprète donc comme la *densité conditionnelle de X_1 sachant que $X_0 = x_0$* . Par un raisonnement analogue, pour tout $n \in \mathbb{N}^*$, la densité jointe de (X_0, \dots, X_n) vaut

$$f(x_0, \dots, x_n) = \nu(x_0)p(x_0, x_1) \dots p(x_{n-1}, x_n). \quad (6.1.3)$$

C'est l'analogie continue de la relation (2.1.3) pour la probabilité des trajectoires dans le cas discret.

La loi de chaque X_n est obtenue en calculant la marginale adéquate de la loi jointe. Ainsi, (6.1.2) implique

$$\mathbb{P}\{X_1 \in A\} = \int_A \int_{\mathcal{X}} f(x_0, x_1) dx_0 dx_1 = \int_A \int_{\mathcal{X}} v(x_0) p(x_0, x_1) dx_0 dx_1 .$$

De manière analogue, (6.1.3) montre que

$$\begin{aligned} \mathbb{P}\{X_2 \in A\} &= \int_A \int_{\mathcal{X}} \int_{\mathcal{X}} v(x_0) p(x_0, x_1) p(x_1, x_2) dx_0 dx_1 dx_2 \\ &= \int_A \int_{\mathcal{X}} v(x_0) p^2(x_0, x_2) dx_0 dx_2 , \end{aligned}$$

où p^2 est un noyau markovien défini par

$$p^2(x_0, x_2) = \int_{\mathcal{X}} p(x_0, x_1) p(x_1, x_2) dx_1 .$$

Plus généralement, pour tout $n \geq 2$ on a

$$\mathbb{P}\{X_n \in A\} = \int_A \int_{\mathcal{X}} v(x_0) p^n(x_0, x_n) dx_0 dx_n ,$$

où p^n est un noyau markovien défini par récurrence par

$$p^n(x_0, x_n) = \int_{\mathcal{X}} p^{n-1}(x_0, x_{n-1}) p(x_{n-1}, x_n) dx_{n-1} ,$$

avec $p^1 = p$. Cette relation est appelée *relation de Chapman–Kolmogorov*.

Il sera commode d'utiliser les notations suivantes, où $A \subset \mathcal{X}$ est un borélien, et $f : \mathcal{X} \rightarrow \mathbb{R}$:

$$\begin{aligned} \mathbb{P}^v\{X_1 \in A\} &= (v\mathcal{P})(A) := \int_A \int_{\mathcal{X}} v(x_0) p(x_0, x_1) dx_0 dx_1 , \\ \mathbb{E}^{x_0}[f(X_1)] &= (\mathcal{P}f)(x_0) := \int_{\mathcal{X}} p(x_0, x_1) f(x_1) dx_1 , \\ \mathbb{E}^v[f(X_1)] &= (v\mathcal{P})(f) := v(\mathcal{P}f) = \int_{\mathcal{X}} \int_{\mathcal{X}} v(x_0) p(x_0, x_1) f(x_1) dx_0 dx_1 . \end{aligned} \quad (6.1.4)$$

La relation de Chapman–Kolmogorov permet de définir \mathcal{P}^n pour tout $n \geq 1$ en remplaçant p par p^n dans (6.1.4). On a également des concepts de mesure signée et fonction test tout à fait analogues à ceux du cas discret.

Définition 6.1.3: Mesures signées finies à densité

Soit $\mu : \mathcal{X} \rightarrow \mathbb{R}$ une application continue par morceaux telle que

$$\|\mu\|_1 := \int_{\mathcal{X}} |\mu(x)| dx < \infty .$$

Elle définit une *mesure signée finie à densité*, qui associe à tout borélien A le nombre

$$\mu(A) := \int_A \mu(x) dx .$$

On notera \mathcal{E}_1 l'espace de Banach des mesures signées finies à densité. Si $\mu : \mathcal{X} \rightarrow \mathbb{R}_+$ et $\|\mu\|_1 = 1$, alors μ est une mesure de probabilité.

Définition 6.1.4: Fonctions test

Une *fonction test* sur \mathcal{X} est une application $f : \mathcal{X} \rightarrow \mathbb{R}$, continue par morceaux, telle que

$$\|f\|_\infty := \sup_{x \in \mathcal{X}} |f(x)| < \infty .$$

On notera \mathcal{E}_∞ l'espace de Banach des fonctions test.

De manière analogue au cas discret, nous utiliserons la notation

$$\mu(f) = \int_{\mathcal{X}} \mu(x) f(x) dx .$$

Cette intégrale est bien définie pour tout $\mu \in \mathcal{E}_1$ et tout $f \in \mathcal{E}_\infty$, et on a

$$|\mu(f)| \leq \int_{\mathcal{X}} |\mu(x)| |f(x)| dx \leq \|\mu\|_1 \|f\|_\infty .$$

Remarque 6.1.5: Continuité par morceaux

L'hypothèse de continuité par morceaux n'est pas vraiment nécessaire. Toutes les intégrales ci-dessus peuvent être interprétées comme des intégrales de Lebesgue, et alors on peut remplacer « continue par morceaux » par « mesurable ». Toutefois, quand nous étudierons les question de récurrence, l'hypothèse de continuité par morceaux simplifiera nettement la théorie. Cette hypothèse est amplement suffisante pour les applications.

6.2 Exemples de chaînes de Markov à espace continu

Voici quelques exemples simples de chaînes de Markov à espace d'états continu.

Exemple 6.2.1: Variables i.i.d.

Soit μ une densité de probabilité sur \mathcal{X} , et soit

$$p(x, y) = \mu(y) \quad \forall x, y \in \mathcal{X} .$$

Il est immédiat de vérifier que p est un noyau markovien à densité. Pour tout $x \in \mathcal{X}$ et tout borélien $A \subset \mathcal{X}$, on a

$$\mathbb{P}^x \{X_1 \in A\} = \int_A \mu(x_1) dx_1 = \mu(A) ,$$

ce qui montre que X_1 a la densité μ . De plus, on trouve

$$p^2(x, x_2) = \int_{\mathcal{X}} p(x, x_1) p(x_1, x_2) dx_1 = \int_{\mathcal{X}} \mu(x_1) \mu(x_2) dx_1 = \mu(x_2) .$$

Plus généralement, on vérifie par récurrence que pour tout $n \geq 1$, on a

$$p^n(x, x_n) = \mu(x_n) \quad \forall x, x_n \in \mathcal{X} .$$

Par conséquent, les variables aléatoires X_1, X_2, \dots ont toutes la même loi, de densité μ . De plus, on vérifie facilement que

$$\mathbb{P}^x \{X_1 \in A_1, \dots, X_n \in A_n\} = \mu(A_1) \dots \mu(A_n)$$

pour tout $n \geq 1$, et tout choix de boréliens A_1, \dots, A_n . Les variables X_1, X_2, \dots sont donc indépendantes.

Exemple 6.2.2: Marche aléatoire à pas Gaussiens

Prenons $\mathcal{X} = \mathbb{R}$. Supposons que $X_0 = 0$ et que

$$X_{n+1} = X_n + Y_{n+1} \quad \forall n \in \mathbb{N},$$

où les Y_n sont i.i.d., de loi normale centrée et de variance σ^2 . En d'autres termes, on a

$$X_n = \sum_{m=1}^n Y_m \quad \forall n \in \mathbb{N}.$$

On dit que $(X_n)_{n \geq 0}$ est une *marche aléatoire à pas Gaussiens sur \mathbb{R}* . Montrons que c'est une chaîne de Markov sur \mathbb{R} . La propriété de Markov suit de l'indépendance des Y_n . Les probabilités de transition sont données par

$$\begin{aligned} \mathbb{P}\{X_{n+1} \in A \mid X_n = x\} &= \mathbb{P}\{X_n + Y_{n+1} \in A \mid X_n = x\} \\ &= \mathbb{P}\{x + Y_{n+1} \in A \mid X_n = x\} \\ &= \mathbb{P}\{Y_{n+1} \in A - x\} \\ &= \int_{A-x} \frac{e^{-y^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}} dy \\ &= \int_A \frac{e^{-(z-x)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}} dz, \end{aligned}$$

où $A - x = \{y - x : y \in A\}$. Il suit que le noyau de transition de la chaîne de Markov est donné par

$$p(x, y) = \frac{e^{-(y-x)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}}.$$

Plus généralement, si les Y_n ont une densité μ , alors $p(x, y) = \mu(y - x)$.

Exemple 6.2.3: Modèle auto-régressif AR(1)

Les *modèles auto-régressifs* sont couramment utilisés en statistiques, en économétrie et en traitement du signal. Le modèle AR(1) en est un cas particulier, où le paramètre 1, appelé *ordre*, désigne le temps de mémoire. Il est défini par la relation de récurrence

$$X_{n+1} = aX_n + Y_{n+1},$$

où $a \in \mathbb{R}$, et les Y_n sont i.i.d., de loi normale centrée et de variance σ^2 . Il suit alors d'un calcul analogue à celui de l'exemple précédent que

$$\begin{aligned} \mathbb{P}\{X_{n+1} \in A \mid X_n = x\} &= \mathbb{P}\{Y_{n+1} \in A - ax\} \\ &= \int_A \frac{e^{-(z-ax)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}} dz. \end{aligned}$$

Le noyau de transition du modèle AR(1) est donc donné par

$$p(x, y) = \frac{e^{-(y-ax)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}}.$$

Plus généralement, le modèle autorégressif d'ordre p , AR(p), est défini par

$$X_{n+1} = \sum_{i=1}^p a_i X_{n-i} + Y_{n+1}.$$

Si $p \geq 2$, la suite des X_n n'est pas une chaîne de Markov, puisque la valeur de X_{n+1} dépend des valeurs à p temps précédents. Toutefois, les vecteurs $Z_n = (X_n, X_{n+1}, \dots, X_{n+p-1})$ définissent une chaîne de Markov sur \mathbb{R}^p .

Exemple 6.2.4: Applications itérées bruitées

Une autre généralisation du modèle AR(1) est donnée par la relation de récurrence

$$X_{n+1} = F(X_n) + Y_{n+1},$$

où $F : \mathbb{R} \rightarrow \mathbb{R}$, et les Y_n sont à nouveau i.i.d., de loi normale centrée et de variance σ^2 . Il s'agit d'une chaîne de Markov de noyau de transition

$$p(x, y) = \frac{e^{-(y-F(x))^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}}.$$

On peut évidemment considérer d'autres lois pour les Y_n que la loi normale. Ce genre de modèle apparaît par exemple en dynamique des populations, ou en épidémiologie. Sa dynamique dépend fortement des propriétés de F (points fixes, stabilité).

Probabilités invariantes et vitesse de convergence

La principale difficulté des chaînes de Markov à espace continu, par rapport aux chaînes de Markov à espace dénombrable, est que l'on a $\mathbb{P}^x\{X_n = y\} = 0$ pour tout choix de $x, y \in \mathcal{X}$ et de $n \in \mathbb{N}^*$. Par conséquent, l'espérance du temps de passage en un point différent du point de départ est en général infinie. La solution consiste à ne pas considérer les temps de premier passage en des points, mais en des ensembles ouverts. C'est ce que nous étudierons plus en détail dans la Section 7.1. Avec cette modification, la théorie des fonctions de Lyapounov s'applique sans grandes modifications, comme nous allons le voir dans la Section 7.2.

7.1 Irréductibilité et récurrence de Harris

Nous considérons dans cette section une chaîne de Markov $(X_n)_{n \geq 0}$ sur un ouvert \mathcal{X} de \mathbb{R}^d , de noyau de transition à densité p . La définition du temps de premier passage est la même que dans le cas dénombrable, mais nous la rappelons néanmoins ici.

Définition 7.1.1: Temps de premier passage

Soit $A \subset \mathcal{X}$ un borélien. Alors le *temps de premier passage en A* de la chaîne de Markov $(X_n)_{n \geq 0}$ est la variable aléatoire

$$\tau_A = \inf\{n \geq 1 : X_n \in A\} \in \mathbb{N}^* \cup \{\infty\}.$$

La définition de l'irréductibilité est en revanche légèrement différente de celle du cas discret, en raison du fait que les probabilités de transition vers des points sont nulles.

Définition 7.1.2: Irréductibilité d'une chaîne de Markov à espace continu

La chaîne de Markov $(X_n)_{n \geq 0}$ est dite *irréductible* si pour tout $x \in \mathcal{X}$ et tout ouvert $A \subset \mathcal{X}$, il existe un $n \in \mathbb{N}^*$ tel que $\mathbb{P}^x\{X_n \in A\} > 0$. De manière équivalente, pour tout $x \in \mathcal{X}$ et $A \subset \mathcal{X}$ ouvert, il existe un $n \in \mathbb{N}^*$ tel que $\mathbb{P}^x\{\tau_A \leq n\} > 0$.

Remarquons que si $p(x, y) > 0$ pour tout $x, y \in \mathcal{X}$, alors la chaîne de Markov est irréductible. C'est le cas pour tous les exemples du chapitre précédent faisant intervenir des variables Gaussiennes. Nous pouvons maintenant donner les analogues continus des définitions de récurrence et de récurrence positive.

Définition 7.1.3: Récurrence (positive) au sens de Harris

- La chaîne de Markov $(X_n)_{n \geq 0}$ est *Harris-récurrente* si

$$\mathbb{P}^x\{\tau_A < \infty\} = 1$$

pour tout $x \in \mathcal{X}$ et tout ouvert $A \subset \mathcal{X}$.

- La chaîne de Markov $(X_n)_{n \geq 0}$ est *Harris-récurrente positive* si de plus

$$\mathbb{E}^x[\tau_A] < \infty$$

pour tout $x \in \mathcal{X}$ et tout ouvert $A \subset \mathcal{X}$.

Remarquons que contrairement au cas discret, la définition fait intervenir le temps de passage en tout ensemble ouvert A . Par conséquent, une chaîne de Markov Harris-récurrente est automatiquement irréductible, puisque $\mathbb{P}^x\{\tau_A < \infty\} = 1$ implique $\mathbb{P}^x\{\tau_A \leq n\} > 0$ pour un n fini. L'intérêt principal de cette définition est lié aux mesures et probabilités invariantes, définies comme suit.

Définition 7.1.4: Mesure et probabilité invariantes

Une mesure μ sur \mathcal{X} est *invariante* si $\mu \mathcal{P} = \mu$, c'est-à-dire si

$$\int_{\mathcal{X}} \mu(x)p(x, y) dx = \mu(y) \quad \forall y \in \mathcal{X}. \quad (7.1.1)$$

Si μ est une mesure de probabilité, alors on dit que c'est une *probabilité invariante*.

Théorème 7.1.5: Récurrence, mesures invariantes et probabilités invariantes

Si la chaîne de Markov $(X_n)_{n \geq 0}$ est Harris-récurrente, alors elle admet une mesure invariante μ . Si elle est de plus Harris-récurrente positive, alors elle admet une probabilité invariante π . De plus, π est essentiellement unique, c'est-à-dire que si π' est une autre probabilité invariante, alors $\pi'(A) = \pi(A)$ pour tout ouvert $A \subset \mathcal{X}$.

Afin de préparer la démonstration de ce résultat, nous introduisons la notion de processus tué en touchant un sous-ensemble de \mathcal{X} .

Définition 7.1.6: Noyau du processus tué en touchant $B \subset \mathcal{X}$

Soit B un borélien de \mathcal{X} , $B^c = \mathcal{X} \setminus B$, et soit $p^\dagger = p_B^\dagger$ la fonction définie par

$$p^\dagger(x, y) = p_B^\dagger(x, y) = \begin{cases} p(x, y) & \text{si } y \in B^c, \\ 0 & \text{si } y \in B. \end{cases} \quad (7.1.2)$$

On définit par récurrence des noyaux p_n^\dagger par $p_1^\dagger = p^\dagger$ et

$$p_{n+1}^\dagger(x, y) = \int_{B^c} p_n^\dagger(x, z)p^\dagger(z, y) dz \quad \forall n \in \mathbb{N}^*. \quad (7.1.3)$$

Notons que les noyaux p_n^\dagger ne sont pas en général markoviens, car leur intégrale par rapport à y est en général strictement inférieure à 1. On dit que ce sont des noyaux *sous-markoviens*. Leur intérêt pour nous est le lemme suivant.

Lemme 7.1.7: Processus tué et loi de τ_B

Pour tout $n \geq 1$, tout $x \in \mathcal{X}$, et tout borélien $A \subset \mathcal{X}$ tel que $A \cap B = \emptyset$, on a

$$\mathbb{P}^x\{X_n \in A, \tau_B > n\} = \int_A p_n^\dagger(x, y) dy.$$

Démonstration. Cela suit du fait que

$$\begin{aligned} \mathbb{P}^x\{X_n \in A, \tau_B > n\} &= \mathbb{P}^x\{X_1 \notin B, X_2 \notin B, \dots, X_{n-1} \notin B, X_n \in A \setminus B\} \\ &= \int_{B^c} \int_{B^c} \dots \int_{B^c} \int_{A \setminus B} p(x, x_1) p(x_1, x_2) \dots p(x_{n-1}, x_n) dx_n dx_{n-1} \dots dx_2 dx_1 \\ &= \int_{\mathcal{X}} \int_{\mathcal{X}} \dots \int_{\mathcal{X}} \int_A p^\dagger(x, x_1) p^\dagger(x_1, x_2) \dots p^\dagger(x_{n-1}, x_n) dx_n dx_{n-1} \dots dx_2 dx_1 \end{aligned}$$

en vertu de (7.1.2), et puisque $A \setminus B = A$. Par (7.1.3) et une récurrence sur n , ceci est bien égal à l'intégrale sur A de p_n^\dagger par rapport à sa seconde variable. \square

Remarque 7.1.8: Processus de Markov tué

On peut associer à $(X_n)_{n \geq 0}$ un processus tué en touchant B , noté $(X_n^\dagger)_{n \geq 0}$, de la manière suivante. On ajoute à \mathcal{X} un état cimetière \dagger , qui est absorbant, et on pose

$$X_n^\dagger = \begin{cases} X_n & \text{si } n < \tau_B, \\ \dagger & \text{si } n \geq \tau_B. \end{cases}$$

Son noyau restreint à B^c est alors p^\dagger , et on a $\mathbb{P}^x\{X_n \in A, \tau_B > n\} = \mathbb{P}^x\{X_n^\dagger \in A\}$ si $A \cap B = \emptyset$.

Un objet important lié au processus tué est le noyau de potentiel, qui joue un rôle similaire à celui de la matrice fondamentale d'une chaîne de Markov absorbante.

Définition 7.1.9: Noyau de potentiel

Soit $B \subset \mathcal{X}$ un ouvert. Le noyau de potentiel de la chaîne de Markov $(X_n)_{n \geq 0}$ relatif à B est l'application qui associe à chaque $x \in \mathcal{X}$ et chaque borélien $A \subset \mathcal{X}$ le nombre

$$G_B(x, A) = \mathbb{E}^x \left[\sum_{n=0}^{\tau_B-1} \mathbb{1}_{X_n \in A} \right] \in [0, \infty]. \quad (7.1.4)$$

Le lien entre noyau de potentiel et processus tué est le suivant.

Proposition 7.1.10: Densité du noyau de potentiel

Pour $x \in B$, le noyau de potentiel $G_B(x, \cdot)$ est une mesure (pas nécessairement finie), qui admet sur B^c la densité

$$g_B(x, y) = \sum_{n=1}^{\infty} p_n^\dagger(x, y), \quad (7.1.5)$$

pour tous les y tels que cette série converge.

Démonstration. Notons tout d'abord que $\mathbb{1}_{X_n \in A_1 \cup A_2} \leq \mathbb{1}_{X_n \in A_1} + \mathbb{1}_{X_n \in A_2}$, avec égalité si $A_1 \cap A_2 = \emptyset$. Ceci montre que $G_B(x, \cdot)$ est une mesure, puisque $G_B(x, A_1 \cup A_2) \leq G_B(x, A_1) + G_B(x, A_2)$, avec égalité si $A_1 \cap A_2 = \emptyset$.

Soit maintenant $A \subset \mathcal{X}$ un borélien tel que $A \cap B = \emptyset$. Alors on a

$$\begin{aligned} \int_A g_B(x, y) dy &= \sum_{n=1}^{\infty} \int_A p_n^\dagger(x, y) dy \\ &= \sum_{n=1}^{\infty} \mathbb{P}^x \{X_n \in A, \tau_B > n\} \\ &= \mathbb{E}^x \left[\sum_{n=1}^{\infty} \mathbb{1}_{X_n \in A, \tau_B > n} \right] \\ &= \mathbb{E}^x \left[\sum_{n=1}^{\tau_B - 1} \mathbb{1}_{X_n \in A} \right] = G_B(x, A). \end{aligned}$$

Pour obtenir la dernière ligne, nous avons utilisé le fait que $A \cap B = \emptyset$ (et donc $x \notin A$). Ceci montre que $g_B(x, y)$ est bien la densité de $G_B(x, \cdot)$, du moins sur B^c . \square

Nous sommes maintenant en mesure de donner une démonstration (au moins partielle) du Théorème 7.1.5 (nous admettrons l'unicité essentielle).

Démonstration du Théorème 7.1.5. La démonstration est inspirée de la construction du cas discret, reposant sur les mesures données par (2.4.3). Fixons un $x \in \mathcal{X}$, et soit $B_\varepsilon = \mathcal{B}_\varepsilon(x)$ la boule de centre x et de rayon $\varepsilon > 0$. Pour tout borélien $A \subset \mathcal{X}$, nous posons

$$\mu_\varepsilon(A) = G_{B_\varepsilon}(x, A).$$

Notre but est de montrer que μ_ε converge vers une mesure invariante lorsque $\varepsilon \rightarrow 0$. Commençons par remarquer que

$$\mu_\varepsilon(B_\varepsilon) = 1.$$

En effet, l'hypothèse de récurrence de Harris implique que $\mathbb{P}^x \{\tau_{B_\varepsilon} < \infty\} = 1$, et par conséquent la somme (7.1.4) a presque sûrement un nombre fini de termes, dont seul le dernier contribue à $\mu_\varepsilon(B_\varepsilon)$. D'autre part, μ_ε admet sur B^c la densité donnée par (7.1.5). Nous observons maintenant que pour tout $y \in B_\varepsilon^c$, on a

$$\begin{aligned} \mu_\varepsilon(y) &= p^\dagger(x, y) + \sum_{n=2}^{\infty} p_n^\dagger(x, y) \\ &= p(x, y) + \sum_{n=2}^{\infty} \int_{B_\varepsilon^c} p_{n-1}^\dagger(x, z) p^\dagger(z, y) dz \\ &= p(x, y) + \int_{B_\varepsilon^c} \sum_{m=1}^{\infty} p_m^\dagger(x, z) p(z, y) dz \\ &= p(x, y) + \int_{B_\varepsilon^c} \mu_\varepsilon(z) p(z, y) dz. \end{aligned}$$

Nous avons utilisé à deux reprises le fait que $p^\dagger(z, y) = p(z, y)$ pour tout $z \in \mathcal{X}$, puisque $y \in B_\varepsilon^c$. Il suit que

$$\mu_\varepsilon(y) - \int_{\mathcal{X}} \mu_\varepsilon(z) p(z, y) dz = p(x, y) - \int_{B_\varepsilon} \mu_\varepsilon(z) p(z, y) dz.$$

Or, comme $\mu_\varepsilon(B_\varepsilon) = 1$, μ_ε est une mesure de probabilité sur B . Par conséquent,

$$\inf_{z \in B_\varepsilon} p(z, y) \leq \int_{B_\varepsilon} \mu_\varepsilon(z) p(z, y) dz \leq \sup_{z \in B_\varepsilon} p(z, y).$$

Il suit que pour tout x en lequel $x \mapsto p(x, y)$ est continue, on a

$$\lim_{\varepsilon \rightarrow 0} \int_{B_\varepsilon} \mu_\varepsilon(z) p(z, y) dz = p(x, y).$$

Par conséquent, on a pour ces x

$$\lim_{\varepsilon \rightarrow 0} \left[\mu_\varepsilon(y) - \int_{\mathcal{X}} \mu_\varepsilon(z) p(z, y) dz \right] = 0.$$

Ceci montre que la limite de μ_ε lorsque $\varepsilon \rightarrow 0$ est invariante pour presque tout x (en tout point de continuité de p , mais la valeur de μ_ε en des points isolés n'influe pas sur les probabilités).

Considérons finalement le cas Harris-récurrent positif. Alors on a

$$\mu_\varepsilon(\mathcal{X}) = \mathbb{E}^x \left[\sum_{n=0}^{\tau_{B_\varepsilon} - 1} \mathbb{1}_{X_n \in \mathcal{X}} \right] = \mathbb{E}^x [\tau_{B_\varepsilon}].$$

On peut alors prendre

$$\pi_\varepsilon(y) = \frac{1}{\mathbb{E}^x [\tau_{B_\varepsilon}]} \mu_\varepsilon(y).$$

C'est la densité d'une mesure de probabilité, qui converge vers une probabilité invariante lorsque $\varepsilon \rightarrow 0$. \square

Remarque 7.1.11: Hypothèses de récurrence

Dans la démonstration, nous n'avons pas utilisé les hypothèses de récurrence (positive) de Harris dans toute leur généralité. En fait, nous avons seulement supposé qu'il existe un point particulier $x \in \mathcal{X}$ et un $\varepsilon_0 > 0$ tels que pour tout $\varepsilon \in]0, \varepsilon_0[$, le temps de passage dans la boule de rayon ε centrée en x , partant de x , est presque sûrement fini, respectivement d'espérance finie.

Le résultat suivant permet d'exprimer des espérances sous la probabilité invariante en termes d'excursions vers un ensemble fixé B .

Proposition 7.1.12: Espérance de fonctions test

Soit $(X_n)_{n \geq 0}$ une chaîne de Markov Harris-récurrente positive, π son unique probabilité invariante, et $f : \mathcal{X} \rightarrow \mathbb{R}_+$. Alors pour tout ouvert $B \subset \mathcal{X}$, on a

$$\pi(f) = \int_B \pi(x) \mathbb{E}^x \left[\sum_{n=0}^{\tau_B - 1} f(X_n) \right] dx = \int_B \pi(x) \mathbb{E}^x \left[\sum_{n=1}^{\tau_B} f(X_n) \right] dx. \quad (7.1.6)$$

Démonstration. Montrons par récurrence que pour tout $N \in \mathbb{N}$, on a

$$\pi(x) = \pi(x) \mathbb{1}_{x \in B} + \sum_{n=1}^N \int_B \pi(y) p_n^\dagger(y, x) dy + \int_{\mathcal{X}} \pi(y) p_{N+1}^\dagger(y, x) dy. \quad (7.1.7)$$

L'initialisation suit de la décomposition $\pi(x) = \pi(x) \mathbb{1}_{x \in B} + \pi(x) \mathbb{1}_{x \in B^c}$ et du fait que

$$\pi(x) \mathbb{1}_{x \in B^c} = (\pi \mathcal{P})(x) \mathbb{1}_{x \in B^c} = \int_{\mathcal{X}} \pi(y) p(y, x) \mathbb{1}_{x \in B^c} dy = \int_{\mathcal{X}} \pi(y) p^\dagger(y, x) dy.$$

L'hérédité vient de

$$\begin{aligned} \int_{\mathcal{X}} \pi(y) p_{N+1}^\dagger(y, x) dy &= \int_{\mathcal{X}} \left[\pi(y) \mathbb{1}_{y \in B} + \int_{\mathcal{X}} \pi(z) p^\dagger(z, y) dz \right] p_{N+1}^\dagger(y, x) dy \\ &= \int_B \pi(y) p_{N+1}^\dagger(y, x) dy + \int_{\mathcal{X}} \pi(y) p_{N+2}^\dagger(y, x) dy. \end{aligned}$$

Faisons alors tendre N vers l'infini dans (7.1.7). Le Lemme 7.1.7 montre que

$$\lim_{N \rightarrow \infty} \int_A p_{N+1}^\dagger(y, x) dx = \lim_{N \rightarrow \infty} \mathbb{P}^y \{X_{N+1} \in A, \tau_B > N + 1\} = 0$$

pour tout borélien A , par récurrence du processus. Il suit que $p_{N+1}^\dagger(y, x)$ tend vers 0, d'où

$$\pi(x) = \pi(x) \mathbb{1}_{x \in B} + \int_B \pi(y) g_B(y, x) dy,$$

où g_B est la densité (7.1.5) du noyau de potentiel G_B . En intégrant cette relation contre f , il vient, en permutant les variables x et y ,

$$\begin{aligned} \pi(f) &= \int_B \pi(x) f(x) dx + \int_{B^c} \int_B \pi(y) g_B(y, x) dy f(x) dx \\ &= \int_B \pi(x) \left[f(x) + \int_{B^c} g_B(x, y) f(y) dy \right] dx. \end{aligned}$$

Or on a

$$f(x) + \int_{B^c} g_B(x, y) f(y) dy = \mathbb{E}^x[f(X_0)] + \mathbb{E}^x \left[\sum_{n=1}^{\tau_B-1} f(X_n) \right].$$

Ceci montre la première égalité dans (7.1.6). La seconde égalité vient du fait que le terme $n = 0$ de la première somme est égal au terme $n = \tau_B$ de la seconde. \square

Remarque 7.1.13: Lien entre probabilité invariante et temps de récurrence moyen

En prenant $f = 1$ dans (7.1.6) avec $B = B_\varepsilon = \mathcal{B}_\varepsilon(x)$, il vient

$$\int_B \pi(x) \mathbb{E}^x[\tau_B] dx = 1.$$

Il suit de (7.1.1) que π est continue presque partout (en tout point de continuité de $y \mapsto p(x, y)$). En ces points, le théorème de la valeur moyenne implique

$$\pi(x) = \lim_{\varepsilon \rightarrow 0} \frac{1}{|\mathcal{B}_\varepsilon(x)| \mathbb{E}^x[\tau_{B_\varepsilon}]}$$

C'est l'analogie de la relation (2.4.2) du cas discret.

Exemple 7.1.14: Processus auto-régressif AR(1)

Nous avons déjà observé que le processus AR(1) était Harris-récurrent, puisque sa densité de transition est minorée par une constante strictement positive sur tout compact. Comme les Y_n sont Gaussiennes, et que toute somme de Gaussiennes est encore Gaussienne, on s'attend à avoir une probabilité invariante Gaussienne. En fait, on a la relation de récurrence

$$\text{Var}(X_{n+1}) = a^2 \text{Var}(X_n) + \sigma^2.$$

On vérifie par récurrence que

$$\text{Var}(X_n) = a^{2n} \text{Var}(X_0) + \frac{\sigma^2}{1 - a^2}.$$

Ainsi, si $|a| < 1$, la loi de X_n converge vers une loi normale centrée de variance $\sigma^2/(1 - a^2)$. C'est aussi la probabilité invariante. On notera que dans le cas $a = 1$, on obtient la marche aléatoire à pas Gaussiens. Dans ce cas, la loi de X_n ne converge pas, et la chaîne de Markov n'est pas Harris-récurrente positive.

7.2 Fonctions de Lyapounov et vitesse de convergence

Dans cette section, nous considérons une chaîne de Markov sur $\mathcal{X} \subset \mathbb{R}^d$, admettant une densité de transition p continue par morceaux. Par simplicité, nous la supposons également irréductible, même si certains résultats peuvent être étendus à des situations plus générales.

Il s'avère que l'approche par fonctions de Lyapounov à l'étude de propriétés de récurrence et de convergence des lois se transpose assez facilement au cas d'un espace continu. La définition de fonction de Lyapounov est la même que dans le cas discret.

Définition 7.2.1: Fonction de Lyapounov

Une *fonction de Lyapounov* est une fonction $V : \mathcal{X} \rightarrow \mathbb{R}_+ = [0, \infty[$ satisfaisant

$$V(x) \rightarrow +\infty \quad \text{pour } \|x\| \rightarrow \infty.$$

La définition de générateur s'adapte aussi très facilement au cas continu.

Définition 7.2.2: Générateur

Le *générateur* \mathcal{L} d'une chaîne de Markov sur un ensemble ouvert $\mathcal{X} \subset \mathbb{R}^d$ est défini, pour toute fonction $f : \mathcal{X} \rightarrow \mathbb{R}$, par

$$(\mathcal{L}f)(x) = (\mathcal{P}f)(x) - f(x) = \int_{\mathcal{X}} p(x, y) f(y) dy - f(x).$$

Les trois résultats suivants, concernant la formule de Dynkin, la croissance sous-exponentielle et la non-explosion, restent inchangés par rapport au cas discret, avec essentiellement les mêmes démonstrations. Nous répétons donc simplement ici leurs énoncés.

Proposition 7.2.3: Formule de Dynkin

Pour toute fonction de Lyapounov V , on a

$$\mathbb{E}^x[V(X_n)] = V(x) + \mathbb{E}^x \left[\sum_{m=0}^{n-1} (\mathcal{L}V)(X_m) \right],$$

De plus, si τ est un temps d'arrêt tel que $\mathbb{E}^x[\tau] < \infty$, alors

$$\mathbb{E}^x[V(X_\tau)] = V(x) + \mathbb{E}^x \left[\sum_{m=0}^{\tau-1} (\mathcal{L}V)(X_m) \right].$$

Théorème 7.2.4: Croissance sous-exponentielle

Supposons qu'il existe une fonction de Lyapounov V et $c > 0$, $d \geq 0$ tels que

$$(\mathcal{L}V)(x) \leq cV(x) + d \quad \forall x \in \mathcal{X}.$$

Alors on a

$$\mathbb{E}^x[V(X_n)] \leq (1+c)^n V(x) + \frac{(1+c)^n - 1}{c} d$$

pour tout $n \in \mathbb{N}$ et tout $x \in \mathcal{X}$.

Théorème 7.2.5: Non-explosion

Supposons qu'il existe $d \geq 0$ et un ensemble borné $K \subset \mathcal{X}$ tel que pour tout $x \in \mathcal{X}$, on ait

$$(\mathcal{L}V)(x) \leq d \mathbb{1}_K(x) = \begin{cases} d & \text{si } x \in K, \\ 0 & \text{sinon.} \end{cases}$$

Alors

$$\mathbb{P}^x \left\{ \lim_{n \rightarrow \infty} \|X_n\| = \infty \right\} = 0 \quad \forall x \in \mathcal{X}.$$

Le résultat de récurrence positive et sa démonstration doivent être très légèrement adaptés. La principale différence est que l'ensemble K doit être un ouvert borné. On pourrait également prendre un compact d'intérieur non vide, le point important étant que K doit être borné et contenir un ensemble ouvert, afin de pouvoir appliquer la Harris-récurrence.

Théorème 7.2.6: Récurrence positive

Soit $f : \mathcal{X} \rightarrow [1, \infty[$ et V une fonction de Lyapounov telle que

$$(\mathcal{L}V)(x) \leq -cf(x) + d \mathbb{1}_K(x) \quad \forall x \in \mathcal{X},$$

pour un ouvert borné $K \subset \mathcal{X}$ et des constantes $c > 0$ et $d \geq 0$. Supposons de plus qu'il existe $\delta > 0$ tels que

$$p(x, y) \geq \delta \quad \forall x, y \in K. \quad (7.2.1)$$

Alors la chaîne de Markov est Harris-récurrente positive, et admet donc une mesure de probabilité invariante π . De plus,

$$\pi(f) < \infty.$$

Démonstration. Nous allons considérer d'abord le passage en K , puis celui en $A \subset K$, puis celui en un A général.

1. Fixons $x \in \mathcal{X}$, et soit $T \in \mathbb{N}^*$. Nous noterons $\tau_K \wedge T = \min\{\tau_K, T\}$. Alors la formule de Dynkin implique

$$\begin{aligned} 0 \leq \mathbb{E}^x[V(X_{\tau_K \wedge T})] &= V(x) + \mathbb{E}^x \left[\sum_{m=0}^{\tau_K \wedge T - 1} (\mathcal{L}V)(X_m) \right] \\ &\leq V(x) - c \mathbb{E}^x \left[\sum_{m=0}^{\tau_K \wedge T - 1} f(X_m) \right] + d \\ &\leq V(x) - c \mathbb{E}^x[\tau_K \wedge T] + d. \end{aligned} \quad (7.2.2)$$

Par conséquent, on a

$$\mathbb{E}^x[\tau_K \wedge T] \leq \frac{V(x) + d}{c}$$

pour tout $T \in \mathbb{N}^*$. Comme le membre de droite ne dépend pas de T , on obtient, en faisant tendre T vers l'infini,

$$\mathbb{E}^x[\tau_K] \leq \frac{V(x) + d}{c} \quad \forall x \in \mathcal{X}. \quad (7.2.3)$$

2. Soit $\tau_{K,n}$ le temps du n ième passage de la chaîne de Markov en K , qui est d'espérance finie en vertu de (7.2.3). Soit $(Y_n)_{n \geq 0}$ la chaîne de Markov trace sur K , définie par $Y_n = X_{\tau_{K,n}}$. Alors, pour tout $x_0 \in K$ et tout ouvert $A \subset K$, on a

$$\mathbb{P}^{x_0}\{Y_1 \notin A\} = 1 - \mathbb{P}^{x_0}\{Y_1 \in A\} \leq 1 - \mathbb{P}^{x_0}\{X_1 \in A\} \leq 1 - \delta|A|$$

par l'hypothèse (7.2.1). Notons que cette borne est indépendante de $x_0 \in K$ choisi. Si $\hat{\tau}_A = \inf\{n \geq 1 : Y_n \in A\}$ est le temps du premier passage de Y_n en A , alors on a pour tout $n \in \mathbb{N}$

$$\mathbb{P}^{x_0}\{\hat{\tau}_A \geq n+1\} = \mathbb{P}^{x_0}\{\hat{\tau}_A \geq n, Y_n \notin A\} \leq (1 - \delta|A|)\mathbb{P}^{x_0}\{\hat{\tau}_A \geq n\}.$$

Par récurrence sur n , on obtient alors

$$\mathbb{P}^{x_0}\{\hat{\tau}_A \geq n\} \leq (1 - \delta|A|)^n \quad \forall n \in \mathbb{N}.$$

Il suit que

$$\mathbb{E}^{x_0}[\hat{\tau}_A] = \sum_{n \geq 0} n \mathbb{P}^{x_0}\{\hat{\tau}_A \geq n\} \leq \sum_{n \geq 0} n(1 - \delta|A|)^n < \infty.$$

Par conséquent, $\mathbb{E}^{x_0}[\tau_A < \infty]$, en vertu de (7.2.3) et du fait que $\mathbb{E}^{x_0}[V(X_1)]$ est borné par la formule de Dynkin.

Ici, nous pouvons invoquer la Remarque 7.1.11 pour conclure que la chaîne de Markov admet une probabilité invariante π . En effet, il suffit d'appliquer le résultat que nous venons d'obtenir aux A donnés par des boules de centre x_0 et de rayon ε assez petit.

3. Afin de montrer que la chaîne de Markov est Harris-récurrente positive, il faut encore vérifier que $\mathbb{E}^x[\tau_A]$ est fini pour tout $x \in \mathcal{X}$ et tout ouvert $A \subset \mathcal{X}$. La majoration (7.2.3) nous permet de nous limiter aux $x \in K$. On peut alors adapter l'argument du point précédent. Soit $(Z_n)_{n \geq 0}$ le processus trace sur $K \cup A$. Il suit de l'hypothèse d'irréductibilité que Z_n va visiter A avec probabilité strictement positive au bout d'un temps assez long. Comme K est borné, on peut trouver un entier n_1 tel que $\mathbb{P}^x\{\hat{\tau}_A > n_1\}$ soit majoré par un $p < 1$ pour tout $x \in K$. On peut alors procéder comme au point précédent pour montrer que $\mathbb{E}^x[\tau_A]$ est fini.
4. Afin de majorer $\pi(f)$, nous utilisons le Lemme 5.4.5 avec $B = K$ et (7.2.2) pour obtenir

$$\pi(f) = \int_K \pi(x) \mathbb{E}^x \left[\sum_{n=0}^{\tau_K-1} f(X_n) dx \right] \leq \pi(K) \sup_{x \in K} \frac{V(x) + d}{c},$$

ce qui est fini puisque K est borné. □

Afin de formuler un résultat de convergence, nous travaillerons à nouveau avec des normes et des distances à poids. Les définitions suivantes sont des adaptations naturelles de celles du cas discret.

Définition 7.2.7: Normes et distances à poids

- Un *poids* sur \mathcal{X} est une application $W : \mathcal{X} \rightarrow [1, \infty[$.
- La norme à poids d'une fonction test f est définie par

$$\|f\|_W = \sup_{x \in \mathcal{X}} \frac{|f(x)|}{W(x)}.$$

On note \mathcal{E}_∞^W l'espace de Banach des fonctions test de norme $\|f\|_W$ finie.

- Pour deux mesures signées finies à densité, on définit

$$\begin{aligned} \rho_W(\mu, \nu) &= \sup_{f: \|f\|_W \leq 1} \int_{\mathcal{X}} f(x) |\mu(x) - \nu(x)| dx \\ &= \sup_{f: \|f\|_W \neq 0} \frac{1}{\|f\|_W} \int_{\mathcal{X}} f(x) |\mu(x) - \nu(x)| dx \\ &= \int_{\mathcal{X}} W(x) |\mu(x) - \nu(x)| dx. \end{aligned}$$

L'analogue continu du Théorème 4.4.1 prend alors la forme suivante. Sa démonstration est la même que celle du cas discret.

Théorème 7.2.8: Ergodicité géométrique

Supposons que les deux conditions suivantes soient satisfaites.

1. **Condition de dérive géométrique :** Il existe $d \geq 0$, $c > 0$ et une fonction de Lyapounov V tels que

$$(\mathcal{L}V)(x) \leq -cV(x) + d \quad \forall x \in \mathcal{X}. \quad (7.2.4)$$

2. **Condition de minoration :** Pour un $R > 2d/c$, soit $K = \{x \in \mathcal{X} : V(x) < R\}$. Alors il existe $\alpha \in]0, 1[$ et une mesure de probabilité ν telle que

$$\inf_{x \in K} p(x, y) = \inf_{x \in K} \mathbb{P}^x \{X_1 = y\} \geq \alpha \nu(y) \quad \forall y \in \mathcal{X}. \quad (7.2.5)$$

Alors il existe des constantes $M > 0$ et $\bar{\gamma} < 1$ telles que

$$\|\mathbb{E}[f(X_n)] - \pi(f)\|_{1+V} \leq M \bar{\gamma}^n \|f - \pi(f)\|_{1+V}$$

pour toute fonction test $f \in \mathcal{E}_\infty^{1+V}$.

Les constantes $\bar{\gamma}$ et M sont à nouveau déterminées par les relations (4.4.4), (4.4.5) et (4.4.6).

Exemple 7.2.9: Processus auto-régressif AR(1)

Nous avons vu dans l'Exemple 6.2.3 que le noyau de transition du modèle AR(1) était donné par

$$p(x, y) = \frac{e^{-(y-ax)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}}.$$

Prenons comme fonction de Lyapounov $V(x) = x^2$. Alors on a

$$\begin{aligned} (\mathcal{L}V)(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} y^2 e^{-(y-ax)^2/(2\sigma^2)} dy - V(x) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (ax + \sigma z)^2 e^{-z^2/2} dz - x^2 \\ &= -(1-a^2)x^2 + \sigma^2. \end{aligned}$$

Pour obtenir la deuxième ligne, nous avons utilisé le changement de variables $y = ax + \sigma z$. La dernière ligne suit des propriétés de la densité d'une loi gaussienne standard. La condition de dérive géométrique (7.2.4) est donc vérifiée avec $c = 1 - a^2$ et $d = \sigma^2$.

Pour vérifier la condition de minoration, nous devons choisir un $R > 2d/c = 2\sigma^2/(1-a^2)$. Alors nous avons $K = [-\sqrt{R}, \sqrt{R}]$ et

$$\inf_{x \in K} p(x, y) = \inf_{|x| < \sqrt{R}} \frac{e^{-(y-ax)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}} = \frac{e^{-(|y|+a\sqrt{R})^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}},$$

l'infimum étant atteint pour $x = \pm\sqrt{R}$, selon le signe de y . Ceci suggère de prendre pour ν la mesure de densité

$$\nu(y) = \frac{e^{-(|y|+a\sqrt{R})^2/(2\sigma^2)}}{\mathcal{N}\sqrt{2\pi\sigma^2}} \mathbb{1}_{|y| \leq \sqrt{R}},$$

où

$$\mathcal{N} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\sqrt{R}}^{\sqrt{R}} e^{-(|y|+a\sqrt{R})^2/(2\sigma^2)} dy = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{R}/\sigma}^{\sqrt{R}/\sigma} e^{-(|z|+a\sqrt{R}/\sigma)^2/2} dz$$

est la constante de normalisation assurant que ν soit une mesure de probabilité (nous avons posé $y = \sigma z$ pour obtenir la seconde inégalité). En effet, la condition de minoration (7.2.5) est alors satisfaite en prenant

$$\alpha = \mathcal{N}.$$

En choisissant $R = 4\sigma^2/(1-a^2)$, on obtient un α indépendant de σ . On vérifie alors que cela donne un taux de convergence $1 - \bar{\gamma}$ strictement positif si $a^2 < 1$, mais qui tend vers 0 lorsque a^2 tend vers 1.

7.3 Exercices

Exercice 7.3.1

On considère la chaîne de Markov sur \mathbb{R} donnée par

$$X_{n+1} = aX_n + Y_{n+1}$$

avec $a \in [0, \infty[$, les Y_n étant indépendantes, identiquement distribuées, de loi uniforme sur $[-\delta, \delta]$ pour un $\delta > 0$.

1. Donner les probabilités de transition $p(x, y)$ de la chaîne.
2. Soit la fonction de Lyapounov $V(x) = x^2$. Calculer $(\mathcal{L}V)(x)$.
3. Pour quelles valeurs de a la chaîne est-elle à croissance sous-exponentielle ?
4. À l'aide de la formule de Dynkin, calculer la variance de X_n lorsque $a = 1$.

5. Pour quelles valeurs de a la chaîne satisfait-elle une condition de dérive géométrique ? Quels en sont les paramètres ?
6. Lorsque la condition de dérive géométrique est satisfaite, trouver $\alpha \in]0, 1[$, une mesure de probabilité ν , et une condition sur p tels que la condition de minoration soit satisfaite. Que peut-on en déduire ?

Exercice 7.3.2

On considère la chaîne de Markov sur \mathbb{R} donnée par

$$X_{n+1} = aX_n + Y_{n+1}$$

avec $a \in [0, \infty[$, les Y_n étant indépendantes, identiquement distribuées, de loi de Cauchy de paramètre $c > 0$.

Soit la fonction de Lyapounov $V(x) = |x|^\beta$. Pour quelles valeurs de $\beta > 0$ la quantité $(\mathcal{L}V)(x)$ est-elle finie ?

Exercice 7.3.3

On considère la chaîne de Markov sur $\mathbb{R}_+ = [0, \infty[$ donnée par

$$X_{n+1} = aX_n + Y_{n+1}$$

avec $a \in [0, \infty[$, les Y_n étant indépendantes, identiquement distribuées, de loi exponentielle de paramètre $\lambda > 0$.

1. Donner les probabilités de transition $p(x, y)$ de la chaîne.
2. Calculer

$$\int_0^\infty x^k e^{-\lambda x} dx$$

pour $k \in \{0, 1\}$.

3. Soit la fonction de Lyapounov $V(x) = x$. Calculer $(\mathcal{L}V)(x)$.
4. Pour quelles valeurs de a la chaîne est-elle à croissance sous-exponentielle ?
5. À l'aide de la formule de Dynkin, calculer l'espérance de X_n lorsque $a = 1$. Que se passe-t-il lorsque $n \rightarrow \infty$?
6. Pour quelles valeurs de a la chaîne satisfait-elle une condition de dérive géométrique ? Quels en sont les paramètres ?
7. Lorsque la condition de dérive géométrique est satisfaite, trouver $\alpha \in]0, 1[$ et une mesure ν tels que la condition de minoration soit satisfaite. Que peut-on en déduire ?

Bibliographie

- [1] R. Durrett. *The Essentials of Probability*. Duxbury, 1994.
- [2] Martin Hairer and Jonathan C. Mattingly. Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI*, volume 63 of *Progr. Probab.*, pages 109–117. Birkhäuser/Springer Basel AG, Basel, 2011.
- [3] Sean P. Meyn and R. L. Tweedie. Stability of Markovian processes. I. Criteria for discrete-time chains. *Adv. in Appl. Probab.*, 24(3):542–574, 1992.
- [4] Esa Nummelin. *General irreducible Markov chains and nonnegative operators*, volume 83 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1984.

Nils Berglund
Institut Denis Poisson (IDP)
Universite d'Orleans, Universite de Tours, CNRS – UMR 7013
Bâtiment de Mathematiques, B.P. 6759
45067 Orleans Cedex 2, France
E-mail address: nils.berglund@univ-orleans.fr
<https://www.idpoisson.fr/berglund>