

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

**Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA**

UMI[®]
800-521-0600

SYLVAIN THIVIERGE

**SIMULATION DE MONTE-CARLO PAR
LES CHAÎNES DE MARKOV**

Mémoire

présenté

à la Faculté des études supérieures

de l'Université Laval

pour l'obtention

du grade de maître ès sciences (M.Sc.)

Département de mathématiques et de statistique

FACULTÉ DES SCIENCES ET DE GÉNIE

UNIVERSITÉ LAVAL

QUÉBEC

AVRIL 1999

© Sylvain Thivierge, 1999



National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-42024-8

Canada

Sommaire

Dans ce mémoire, nous nous intéressons à la simulation de Monte-Carlo par les chaînes de Markov. Étant donnée une mesure de probabilité π définie sur \mathbb{R}^d , ce type de simulation consiste principalement à simuler une chaîne de Markov $(X_n; n \geq 0)$ dont le noyau markovien P possède, comme loi stationnaire, la mesure de probabilité π . Lorsque le noyau markovien P est apériodique et récurrent au sens de Harris, la loi de X_n converge en variation totale vers π . Sous ces conditions, la simulation markovienne permet de simuler une variable aléatoire de loi π , et elle constitue, par conséquent, une solution de rechange aux méthodes de simulation de Monte-Carlo plus conventionnelles. Trois algorithmes de simulation markovienne sont décrits et étudiés dans ce mémoire; il s'agit de l'échantillonneur de Gibbs, l'algorithme de Metropolis-Hastings, et l'algorithme hit-and-run. Pour chacun d'eux, des conditions assurant la convergence en variation totale de la loi de X_n vers π sont énoncées. D'autres aspects importants tels que l'évaluation de la vitesse de cette convergence et les conditions permettant l'utilisation d'un théorème limite central dans le but d'estimer certaines caractéristiques de la loi π à partir de $(X_n; n \geq 0)$ sont aussi abordés.

Claude Bélisle
Directeur de recherche

Sylvain Thivierge
Étudiant

Remerciements

Je tiens d'abord à adresser mes remerciements à mon directeur de recherche, le professeur Claude Bélisle, pour son soutien, ses suggestions, ainsi que tout le temps qu'il a consacré à répondre à mes nombreuses questions, tant avant que pendant la rédaction de ce mémoire.

Je remercie aussi tous les professeurs du Département de mathématiques et de statistique de l'Université Laval pour m'avoir fait découvrir et apprécié le monde des probabilités et de la statistique, ainsi que tous les autres membres du personnel du département pour leur disponibilité à mon égard.

Durant le trimestre d'hiver 1997, j'ai passé trois mois à l'Université de la Colombie-Britannique, à Vancouver, alors que Monsieur Bélisle y était en visite sabbatique. Je remercie le Département de statistique de cette université pour son hospitalité ainsi que Monsieur Bélisle pour son appui.

Je désire aussi exprimer ma gratitude aux personnes et organismes qui m'ont permis d'obtenir un appui financier du fond FCAR par l'intermédiaire de l'équipe de probabilités et statistique du Département de mathématiques et de statistique ainsi que d'un appui financier du fond CRSNG par l'intermédiaire de mon directeur de recherche.

Finalement, mes derniers remerciements vont aux personnes de mon entourage pour leurs encouragements. Un merci particulier à Christine pour sa compréhension et sa patience lors des nombreuses soirées et fins de semaine consacrées à la rédaction de ce mémoire.

Table des matières

| | |
|--|------------|
| Sommaire | i |
| Remerciements | ii |
| Table des matières | iii |
| Introduction | 1 |
| 1 Chaînes de Markov | 6 |
| 1.1 Rappel | 6 |
| 1.2 Noyaux markoviens et chaînes de Markov | 8 |
| 1.2.1 Définition d'une chaîne de Markov | 8 |
| 1.2.2 Existence | 12 |
| 1.2.3 Probabilités de transition d'ordre supérieur | 15 |
| 1.3 Classification des chaînes de Markov | 17 |
| 1.3.1 φ -irréductibilité | 18 |
| 1.3.2 Périodicité | 22 |

| | | |
|----------|---|-----------|
| 1.3.3 | Récurrance | 25 |
| 1.4 | Convergence | 27 |
| 1.4.1 | Mesures stationnaires | 27 |
| 1.4.2 | Convergence en variation totale | 30 |
| 1.4.3 | Le théorème d'Orey | 33 |
| 2 | Algorithmes de simulation markovienne | 36 |
| 2.1 | L'échantillonneur de Gibbs | 37 |
| 2.1.1 | Origines de l'algorithme | 37 |
| 2.1.2 | Description de l'algorithme | 40 |
| 2.1.3 | Le choix de la densité f | 42 |
| 2.1.4 | Noyau markovien de l'échantillonneur de Gibbs | 46 |
| 2.1.5 | Convergence | 50 |
| 2.1.6 | L'échantillonneur de Gibbs avec mises à jour aléatoires | 59 |
| 2.1.7 | Regroupement des composantes | 60 |
| 2.2 | L'algorithme de Metropolis-Hastings | 61 |
| 2.2.1 | Origines de l'algorithme | 61 |
| 2.2.2 | Description de l'algorithme | 64 |
| 2.2.3 | Réversibilité | 66 |
| 2.2.4 | Convergence | 70 |
| 2.2.5 | Cas particuliers de l'algorithme | 75 |
| 2.2.6 | Mises à jour composante par composante | 77 |

| | | |
|----------|--|------------|
| 2.3 | L'algorithme hit-and-run | 79 |
| 2.3.1 | Origines de l'algorithme | 79 |
| 2.3.2 | Description de l'algorithme | 81 |
| 2.3.3 | Stationnarité de π | 83 |
| 2.3.4 | Convergence | 86 |
| 3 | Vitesse de convergence et ergodicité géométrique | 90 |
| 3.1 | Ergodicité géométrique | 91 |
| 3.2 | Comportement asymptotique des moyennes | 99 |
| 3.2.1 | La loi forte des grands nombres | 99 |
| 3.2.2 | Le théorème limite central | 100 |
| 3.3 | Vitesse de convergence de l'algorithme hit-and-run | 105 |
| 3.3.1 | Convergence lente de l'algorithme hit-and-run | 105 |
| 3.3.2 | Ergodicité géométrique de l'algorithme hit-and-run | 107 |
| 3.3.2.1 | Le cas où ν est la loi uniforme sur (S_d, \mathcal{B}_{S_d}) | 108 |
| 3.3.2.2 | Le cas où ν est la loi uniforme sur $\{e_1, \dots, e_d\}$ | 110 |
| | Conclusion | 113 |
| | Bibliographie | 116 |

Introduction

Beaucoup de problèmes en statistique nécessitent le calcul d'intégrales de type

$$\pi(g) = \int_E g d\pi, \quad (1)$$

où π est une mesure de probabilité définie sur un espace mesurable (E, \mathcal{E}) , et où $g : E \rightarrow \mathbb{R}$ est une fonction réelle mesurable. Par exemple, si π est une mesure de probabilité sur \mathbb{R}^d et que $X = (X^1, \dots, X^d)$ est une variable aléatoire de loi π , nous pourrions être intéressés à calculer certaines caractéristiques de π et de X , comme l'espérance d'une des composantes de X ou la probabilité d'un événement $\{X \in A\}$, où $A \in \mathcal{E}$. Ces caractéristiques se calculent généralement par l'entremise de $\pi(g)$ avec une fonction g adéquatement choisie. Par exemple, si $g(x^1, \dots, x^d) = x^i$, $i \in \{1, \dots, d\}$, alors $\pi(g)$ est l'espérance de la composante X^i de X . Si $g(x^1, \dots, x^d) = \mathbf{1}_A(x^1, \dots, x^d)$, alors $\pi(g)$ est la probabilité de l'événement $\{X \in A\}$.

Dans plusieurs situations, comme par exemple lorsque π est une loi à plusieurs dimensions ou lorsque le support de π est un ensemble compliqué, il peut être difficile, voire impossible, de calculer une telle intégrale de façon analytique ou même de l'évaluer à l'aide de méthodes numériques. Une autre solution qui s'offre à nous consiste à estimer $\pi(g)$ à l'aide d'une *méthode de Monte-Carlo*, c'est-à-dire une méthode qui fait intervenir la simulation de variables aléatoires (que nous appelons *simulation de Monte-Carlo*) distribuées selon certaines lois données.

Le champ d'application des méthodes de Monte-Carlo est évidemment beaucoup plus vaste que l'estimation d'intégrales. Il inclut aussi l'évaluation de la performance d'estimateurs ou de tests d'hypothèses, la maximisation de fonctions et la résolution

de systèmes d'équations. Il n'est pas non plus restreint aux problèmes mathématiques ou statistiques; la simulation et les méthodes de Monte-Carlo sont aussi utilisées en physique, en chimie et en biologie, par exemple, afin d'étudier le comportement de certains phénomènes aléatoires. Pour une introduction complète sur la simulation et les méthodes de Monte-Carlo, incluant leurs origines et leurs applications, on peut consulter entre autres Hammersley et Handscomb (1964), et Rubinstein (1981).

Revenons maintenant au contexte qui nous a servi d'introduction, c'est-à-dire l'estimation de (1). La méthode de Monte-Carlo la plus simple qui permet d'estimer $\pi(g)$ est la *méthode de la moyenne* (voir Rubinstein, 1981, page 118). Elle consiste à simuler d'abord un échantillon de variables aléatoires indépendantes X_1, \dots, X_n , toutes de loi π , pour ensuite utiliser l'estimateur naturel

$$\bar{g}_n = \frac{1}{n} \sum_{i=1}^n g(X_i) \quad (2)$$

afin d'approximer $\pi(g)$. Si $\pi(|g|) < \infty$, cet estimateur est sans biais pour $\pi(g)$ et il converge avec probabilité 1 vers $\pi(g)$ lorsque $n \rightarrow \infty$, en vertu de la loi forte des grands nombres. Si de plus $\pi(g^2) < \infty$, le théorème limite central assure que la moyenne \bar{g}_n est asymptotiquement normalement distribuée, ce qui permet entre autres de construire des intervalles de confiance pour $\pi(g)$. Cette approche peut bien sûr être généralisée au cas où la fonction g de (1) est une fonction à valeurs dans \mathbb{R}^d , $d \geq 2$, à l'aide, par exemple, du théorème limite central multidimensionnel.

Lorsque la loi π est complexe, il peut être impossible de construire un algorithme de simulation de Monte-Carlo permettant de simuler directement des variables aléatoires de loi π ou de mettre en pratique un algorithme qui permettrait de le faire (par *directement*, nous entendons que la loi des variables aléatoires simulées est exactement π). Une autre approche possible pour estimer $\pi(g)$ consiste à simuler n variables aléatoires indépendantes Y_1, \dots, Y_n distribuées selon une autre loi, disons π' , et à utiliser ensuite la moyenne des variables aléatoires $(g(Y_i); i \in \{1, \dots, n\})$, cette fois par contre, en pondérant adéquatement chacune d'elles. Pour illustrer le fonctionnement de cette méthode, supposons que π et π' soient deux mesures de probabilité sur \mathbb{R}^d , absolument continues par rapport à la mesure de Lebesgue avec densités f et f' respectivement.

Alors $\pi(g)$ peut aussi s'écrire

$$\pi(g) = \int_{\mathbb{R}^d} \frac{g(x)f(x)}{f'(x)} f'(x) dx .$$

Si Y_1, \dots, Y_n est un échantillon simulé de variables aléatoires indépendantes de loi π' , il suffit alors d'estimer $\pi(g)$ à l'aide de l'estimateur $\bar{h}_n = n^{-1} \sum_{i=1}^n h(Y_i)$, où $h(x) = g(x)f(x)/f'(x)$. Cette méthode de Monte-Carlo s'appelle la *méthode de l'échantillonnage ciblé* (voir Rubinstein, 1981, page 122). Même lorsque des algorithmes de simulation de Monte-Carlo permettant de simuler des variables aléatoires indépendantes X_1, \dots, X_n de loi π sont disponibles, la méthode de l'échantillonnage ciblé peut être plus efficace que la méthode de la moyenne puisqu'un choix approprié de loi π' peut permettre d'obtenir un estimateur dont la variance est beaucoup plus petite que celle de (2). Cependant, il arrive parfois que la mesure π soit connue seulement à une constante multiplicative près; c'est souvent le cas, par exemple, lorsque π est la loi a posteriori d'un paramètre multidimensionnel dans un modèle statistique bayésien. Lorsqu'il est impossible de calculer la constante de normalisation, la méthode de l'échantillonnage ciblé ne peut être utilisée.

La méthode de la moyenne et la méthode de l'échantillonnage ciblé nécessitent donc la simulation d'un échantillon de variables aléatoires indépendantes de loi π (ou π'). Pour parvenir à simuler une variable aléatoire X de loi π lorsqu'il est impossible de le faire directement, une autre approche consiste à simuler une chaîne de Markov $(X_n; n \geq 0)$ dont le noyau markovien P possède la loi π comme loi stationnaire. Sous certaines conditions, le fait que π soit stationnaire pour P implique que lorsque $n \rightarrow \infty$, la loi de X_n converge (en un sens que nous décrirons dans ce mémoire) vers π . Pour simuler une variable X de loi π , il suffit de se fixer un entier $k \geq 1$ et de simuler les $k + 1$ premières variables de la chaîne de Markov, c'est-à-dire $(X_n; 0 \leq n \leq k)$, et de conserver uniquement X_k ; pour k suffisamment grand, la loi de X_k sera si près de π que X_k pourra, à toutes fins pratiques, être vue comme une variable aléatoire de loi π . Pour obtenir un échantillon de m variables aléatoires indépendantes de loi π , il suffit alors de simuler indépendamment les $(k + 1)$ premières variables de m chaînes, disons $(X_{1,n}; 0 \leq n \leq k)$, \dots , $(X_{m,n}; 0 \leq n \leq k)$, et de conserver uniquement $X_{1,k}, \dots, X_{m,k}$. La méthode de la moyenne (ou de l'échantillonnage ciblé) peut ensuite être appliquée à l'échantillon $X_{1,k}, \dots, X_{m,k}$ afin d'estimer $\pi(g)$.

La simulation d'une chaîne de Markov dont le noyau markovien possède π comme loi stationnaire permet aussi l'utilisation d'une troisième méthode de Monte-Carlo afin d'estimer $\pi(g)$. Celle-ci consiste à simuler d'abord les $m(k+1)$ premières variables d'une seule chaîne de Markov $(X_n; n \geq 0)$ pour estimer ensuite $\pi(g)$ à l'aide de la moyenne

$$\frac{1}{m(k+1)} \sum_{i=0}^{m(k+1)} g(X_i). \quad (3)$$

Cette méthode est parfois appelée la *méthode de la moyenne ergodique*. Sous certaines conditions, l'estimateur (3) converge presque sûrement vers $\pi(g)$ lorsque $m(k+1) \rightarrow \infty$, et sa loi asymptotique est une loi normale (lorsqu'il est centré et normalisé convenablement). Il est aussi possible de définir une version ergodique de l'échantillonnage ciblé, qui consiste à simuler les $m(k+1)$ premières variables d'une chaîne de Markov $(X_n; n \geq 0)$ dont le noyau markovien possède π' comme loi stationnaire, et à estimer ensuite $\pi(g)$ à l'aide d'une moyenne pondérée des variables $(g(X_n), 0 \leq n \leq m(k+1))$.

Aussi surprenant que cela puisse paraître, il est parfois plus facile de simuler une chaîne de Markov $(X_n; n \geq 0)$ dont le noyau markovien possède π comme loi stationnaire que de simuler directement une variable aléatoire de loi π , surtout lorsque π est une loi compliquée ou à plusieurs dimensions. Jusqu'à présent, plusieurs algorithmes permettant de le faire ont vu le jour; les plus connus sont l'*algorithme de Metropolis* (Metropolis et coll., 1953), l'*algorithme de Hastings* (Hastings, 1970), qui est en fait une généralisation de l'algorithme de Metropolis, l'*échantillonneur de Gibbs* (Geman et Geman, 1984; Gelfand et Smith, 1990) et l'*algorithme hit-and-run* (Smith, 1984; Bélisle, Romeijn et Smith, 1993). Puisque la simulation d'une variable aléatoire de loi π à l'aide de l'un de ces algorithmes se fait par l'intermédiaire de la simulation d'une chaîne de Markov, ils sont connus, en anglais, sous le nom de *Markov Chain Monte Carlo algorithms*, ou plus simplement *MCMC algorithms*. Dans ce mémoire, ces algorithmes seront appelés *algorithmes de simulation de Monte-Carlo par les chaînes de Markov*, et souvent, dans le but d'alléger le texte, nous utiliserons simplement les termes *algorithmes de simulation markovienne*.

Ces algorithmes de simulation markovienne requièrent généralement que la loi π soit une mesure de probabilité définie sur \mathbb{R}^d , absolument continue par rapport à la mesure

de Lebesgue; certains d'entre eux peuvent cependant être généralisés à quelques cas où π est absolument continue par rapport à un autre type de mesure. Bien qu'ils soient relativement faciles à mettre en application, les propriétés des chaînes de Markov qu'ils simulent sont souvent complexes et difficiles à étudier, et étant donnée une chaîne de Markov $(X_n; n \geq 0)$ dont le noyau markovien possède π comme loi stationnaire, il n'est généralement pas simple de déterminer si:

- 1) la loi de X_n converge (en un sens approprié) vers π ;
- 2) la moyenne $(n + 1)^{-1} \sum_{i=0}^n g(X_i)$ converge presque sûrement vers $\pi(g)$ dans le cas où la méthode de la moyenne ergodique est utilisée pour estimer $\pi(g)$;
- 3) la vitesse de ces convergences (lorsqu'elles ont lieu) est rapide ou non.

Le succès et la performance d'un de ces algorithmes ou d'une méthode d'estimation de $\pi(g)$ dépendent évidemment de ces trois aspects.

Ce mémoire est une introduction à la simulation markovienne. Puisque celle-ci consiste avant tout à simuler des chaînes de Markov, le chapitre 1 présentera certaines notions importantes sur ces dernières; ces notions sont indispensables si l'on veut bien comprendre le fonctionnement des algorithmes de simulation markovienne et évaluer leur efficacité pour une loi π donnée. Au chapitre 2, nous nous attarderons ensuite aux trois principaux algorithmes (et certaines de leurs généralisations), à savoir l'échantillonneur de Gibbs, l'algorithme de Metropolis-Hastings et l'algorithme hit-and-run; en plus d'un bref résumé de leurs origines et de la description de la façon dont ils procèdent pour simuler une chaîne de Markov dont le noyau possède π comme loi stationnaire, nous présenterons aussi des conditions sous lesquelles la loi de X_n d'une chaîne simulée par ces algorithmes converge effectivement vers π . Finalement, au chapitre 3, nous nous intéresserons à d'autres aspects importants, comme par exemple la vitesse à laquelle la convergence a lieu (lorsqu'elle a lieu) et les conditions sous lesquelles le théorème limite central assurant la normalité asymptotique de $(n + 1)^{-1} \sum_{i=0}^n g(X_i)$ est valide, lorsque l'approche choisie pour estimer $\pi(g)$ est la méthode de la moyenne ergodique.

Chapitre 1

Chaînes de Markov

Ce chapitre se veut un résumé de certaines notions portant sur les chaînes de Markov à valeurs dans des espaces mesurables généraux. Nous y présentons aussi quelques résultats de convergence importants qui nous seront utiles par la suite lors de l'étude des algorithmes de simulation markovienne. Ce chapitre s'inspire en grande partie des ouvrages suivants: Doob (1953), Orey (1971), Revuz (1984), Nummelin (1984), Meyn et Tweedie (1993), Tierney (1994) et Athreya, Doss et Sethuraman (1996).

Nous débuterons par un rappel du théorème sur la convergence des chaînes de Markov à espaces d'états dénombrables. Notre principal objectif sera de développer les outils qui nous permettront de généraliser ce résultat aux chaînes de Markov à espaces d'états plus généraux.

1.1 Rappel

Soit E un ensemble dénombrable. Considérons $(X_n; n \geq 0)$ une chaîne de Markov à valeurs dans E dont les probabilités de transition sont homogènes dans le temps et

sont données par

$$p_{ij} = \mathbb{P}[X_{n+1} = j | X_n = i], \quad i, j \in E, n \geq 0.$$

Supposons maintenant que la chaîne soit *irréductible*, c'est-à-dire que pour tous $i, j \in E$, il existe un entier $m \geq 1$ tel que

$$p_{ij}^m = \mathbb{P}[X_m = j | X_0 = i] > 0. \quad (1.1)$$

Supposons de plus que la chaîne soit *apériodique*, c'est-à-dire que pour tout $i \in E$,

$$\text{p. g. c. d.}\{n \geq 1 : p_{ii}^n > 0\} = 1,$$

où "p. g. c. d." signifie "plus grand commun diviseur". Alors:

- a) si E est fini, la chaîne possède une loi stationnaire, c'est-à-dire qu'il existe des nombres réels $(\pi_i; i \in E)$ tels que

$$\pi_i \geq 0, \quad \sum_{i \in E} \pi_i = 1, \quad \text{et} \quad \sum_{i \in E} \pi_i p_{ij} = \pi_j \quad \text{pour tout } j \in E,$$

et de plus,

$$\lim_{n \rightarrow \infty} p_{ij}^n = \pi_j \quad \text{pour tous } i, j \in E; \quad (1.2)$$

- b) si E est infini et si la chaîne possède une loi stationnaire $(\pi_i; i \in E)$, alors (1.2) est toujours valide.

Ce résultat classique, prouvé dans le cas fini par A. A. Markov en 1906 et dans le cas infini par A. N. Kolmogorov en 1931, est présenté dans de nombreux livres de probabilités. Parmi les plus importants, on retrouve notamment Chung (1967), Feller (1968), Freedman (1983), Karlin and Taylor (1975), Çinlar (1975) et Billingsley (1995).

Il devient maintenant naturel de se demander si un résultat de convergence similaire existe pour une chaîne de Markov $(X_n; n \geq 0)$ à valeurs dans un espace d'états plus général, comme par exemple dans \mathbb{R}^d ou dans un borélien de \mathbb{R}^d . Il s'avère qu'en généralisant de façon adéquate les concepts d'irréductibilité, d'apériodicité et de loi stationnaire, il est possible d'énoncer des conditions, en termes de ces concepts, sous lesquelles la loi de X_n converge, en un sens approprié, vers la loi stationnaire éventuelle de la chaîne.

1.2 Noyaux markoviens et chaînes de Markov

1.2.1 Définition d'une chaîne de Markov

Une chaîne de Markov est une suite de variables aléatoires $(X_n; n \geq 0)$ définies sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, à valeurs dans un espace mesurable (E, \mathcal{E}) et satisfaisant la *propriété de Markov*. Cette propriété peut être énoncée ainsi: la suite $(X_n; n \geq 0)$ possède la *propriété de Markov* si pour tout $n \geq 1$, X_{n+1} et (X_{n-1}, \dots, X_0) sont conditionnellement indépendantes étant donnée X_n .

L'ensemble E est appelé *l'espace des états*. L'indice n représente généralement le temps; si $X_n = x$, $x \in E$, nous disons qu'au temps n , la chaîne est à l'état x .

Avant de donner la définition formelle d'une chaîne de Markov, nous devons d'abord introduire le concept de *noyau*.

Définition 1.1. Un *noyau* défini sur un espace mesurable (E, \mathcal{E}) est une fonction $K : E \times \mathcal{E} \rightarrow [0, \infty]$ telle que:

- i) pour tout $A \in \mathcal{E}$, la fonction $K(\cdot, A)$ est mesurable;
- ii) pour tout $x \in E$, la fonction $K(x, \cdot)$ est une mesure sur (E, \mathcal{E}) .

Un noyau K défini sur un espace mesurable (E, \mathcal{E}) est dit *sous-markovien* s'il satisfait $K(x, E) \leq 1$ pour tout $x \in E$. S'il satisfait $K(x, E) = 1$ pour tout $x \in E$, c'est-à-dire si la mesure $K(x, \cdot)$ est une mesure de probabilité sur (E, \mathcal{E}) pour tout $x \in E$, il est dit *markovien*.

Dans ce mémoire, nous nous intéresserons surtout aux noyaux sous-markoviens, et plus particulièrement, aux noyaux markoviens. Nous verrons plus loin que lorsqu'il est associé à une chaîne de Markov à valeurs dans un espace général, un noyau markovien joue un rôle similaire à celui de la matrice des probabilités de transition dans le cas

dénombrable. Pour identifier un noyau markovien, nous utiliserons la plupart du temps le symbole P .

Définition 1.2. Soit un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, un espace mesurable (E, \mathcal{E}) et un noyau markovien P défini sur (E, \mathcal{E}) . Une suite de variables aléatoires $(X_n; n \geq 0)$ définies sur (Ω, \mathcal{F}) et à valeurs dans (E, \mathcal{E}) est appelée *une chaîne de Markov avec noyau markovien P* si, pour tout $A \in \mathcal{E}$ et pour tout $n \geq 0$,

$$\mathbb{P}[X_{n+1} \in A | X_n, \dots, X_0] = \mathbb{P}[X_{n+1} \in A | X_n] = P(X_n, A) \quad \text{p.s.} \quad (1.3)$$

Outre la propriété de Markov, la suite de variables aléatoires que nous venons de définir possède d'autres propriétés importantes. Cette définition mérite que l'on s'y attarde un peu.

Premièrement, la partie

$$\mathbb{P}[X_{n+1} \in A | X_n, \dots, X_0] = \mathbb{P}[X_{n+1} \in A | X_n] \quad \text{p.s.} \quad (1.4)$$

est précisément la propriété de Markov. Cette propriété est suffisante pour que $(X_n; n \geq 0)$ puisse s'appeler une *chaîne de Markov*.

Ensuite, (1.3) suppose implicitement que pour toute paire (X_{n+1}, X_n) , $n \geq 0$, il existe une *loi conditionnelle régulière de X_{n+1} étant donnée X_n* , c'est-à-dire une fonction $H_n : \Omega \times \mathcal{E} \rightarrow [0, 1]$ telle que:

- i) pour tout $A \in \mathcal{E}$, la fonction $H_n(\cdot, A)$ est une version de la probabilité conditionnelle de $\{X_{n+1} \in A\}$ étant donnée X_n ;
- ii) pour tout $\omega \in \Omega$, la fonction $H_n(\omega, \cdot)$ est une mesure de probabilité sur (E, \mathcal{E}) .

En général, étant données deux variables aléatoires Y et X (faisant partie d'une chaîne de Markov ou non) définies sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et à valeurs dans un espace mesurable (E, \mathcal{E}) , il n'est pas garanti qu'une loi conditionnelle régulière de Y étant donnée X existe. Néanmoins, en imposant certaines restrictions à l'espace (E, \mathcal{E}) , il est possible d'en assurer l'existence. Par exemple, si E est un espace métrique complet

et séparable et que \mathcal{E} est la σ -algèbre engendrée par les sous-ensembles ouverts de E , une fonction H satisfaisant les conditions (i) et (ii) pour Y et X existe toujours. Ceci est vrai en particulier dans le cas où E est dénombrable et que \mathcal{E} est la famille de tous les sous-ensembles de E . À l'aide de la proposition 4.9 de Breiman (1968), il est possible de montrer que si H est une loi conditionnelle de Y étant donnée X , il existe alors un noyau markovien P tel que pour tout $A \in \mathcal{E}$,

$$P(X(\omega), A) = H(\omega, A) \quad \text{p.s.,}$$

ce qui implique que la fonction $P(X(\cdot), \cdot)$ est elle-même une loi conditionnelle de Y étant donnée X . Pour plus de détails sur les propriétés et l'existence des mesures de probabilité conditionnelles régulières, on peut consulter entre autres Breiman (1968) et Durrett (1991).

Enfin, (1.3) suppose qu'il est possible de choisir les noyaux $(P_n; n \geq 0)$ associés aux lois conditionnelles régulières $(H_n; n \geq 0)$ de manière à ce qu'ils soient tous identiques et égaux à P . Ceci implique que pour tout $A \in \mathcal{E}$ et pour tout $n \geq 0$, la fonction $P(x, A)$ est une version de la probabilité conditionnelle $\mathbb{P}[X_{n+1} \in A | X_n = x]$ (voir par exemple Breiman, 1968, page 69, pour la définition de $\mathbb{P}[X_n \in A | X_{n-1} = x]$). Une chaîne de Markov pour laquelle il existe des versions de $\mathbb{P}[X_{n+1} \in A | X_n = x]$, $n \geq 0$, $A \in \mathcal{E}$, satisfaisant

$$\mathbb{P}[X_{n+1} \in A | X_n = x] = \mathbb{P}[X_1 \in A | X_0 = x] \quad \text{pour tous } n \geq 0, A \in \mathcal{E} \text{ et } x \in E$$

est dite *homogène dans le temps*. En demeurant prudent dans notre interprétation de $\mathbb{P}[X_{n+1} \in A | X_n = x]$, l'homogénéité dans le temps signifie en gros que pour tout $x \in E$, la loi de X_{n+1} étant donné que $X_n = x$ est toujours la même, peu importe $n \geq 0$.

La définition 1.2 est donc en fait la définition d'une *chaîne de Markov homogène dans le temps, possédant un noyau markovien P* . Dans ce chapitre, nous restreindrons notre étude aux chaînes homogènes dans le temps. Par la suite, chaque fois qu'il sera question d'une chaîne de Markov, il sera sous-entendu qu'elle est homogène dans le temps, à moins qu'il soit spécifié qu'il en est autrement. Pour éviter des complications inutiles, nous ne considérerons aussi que les chaînes de Markov qui possèdent un noyau markovien.

Un noyau markovien est donc en quelque sorte la généralisation naturelle de la matrice des probabilités de transition. En effet, si $(X_n; n \geq 0)$ est une chaîne de Markov à valeurs dans un espace général (E, \mathcal{E}) et avec noyau markovien P , alors pour tout $A \in \mathcal{E}$ et pour tout $n \geq 0$, $P(X_n, A)$ nous fournit une version de la probabilité conditionnelle $\mathbb{P}[X_{n+1} \in A | X_n]$. En particulier, si E est dénombrable, nous avons, pour tout $n \geq 0$,

$$\mathbb{P}[X_{n+1} = j | X_n = i] = P(i, \{j\}), \quad i, j \in E,$$

sauf possiblement pour certains $i \in E$ tels que $\mathbb{P}[X_n = i] = 0$. Les $P(i, \{j\})$, $i, j \in E$, sont donc simplement les probabilités que l'on retrouve dans la matrice de transition habituellement associée à une telle chaîne.

Revenons maintenant à la définition 1.2. Supposons que $(X_n; n \geq 0)$ soit une chaîne de Markov à valeurs dans un espace mesurable (E, \mathcal{E}) avec noyau markovien P . La mesure de probabilité λ définie sur (E, \mathcal{E}) par

$$\lambda(A) = \mathbb{P}(X_0 \in A), \quad A \in \mathcal{E} \quad (1.5)$$

est appelée *la mesure de probabilité initiale* de la chaîne. La mesure λ est donc la loi de X_0 . Considérons maintenant, pour chaque $n \geq 1$, l'espace mesurable produit $(E^{n+1}, \mathcal{E}^{n+1})$, où E^{n+1} est le produit cartésien $E \times E \times \dots \times E$ ($n+1$ fois) et où \mathcal{E}^{n+1} est la plus petite σ -algèbre contenant les sous-ensembles de E^{n+1} pouvant s'écrire sous la forme $A_0 \times A_1 \times \dots \times A_n$, où $A_0, \dots, A_n \in \mathcal{E}$. Définissons maintenant une mesure de probabilité μ_n sur $(E^{n+1}, \mathcal{E}^{n+1})$ par

$$\mu_n(G) = \mathbb{P}[(X_0, \dots, X_n) \in G], \quad G \in \mathcal{E}^{n+1}.$$

Puisque X_0, \dots, X_n sont toutes des variables aléatoires à valeurs dans (E, \mathcal{E}) , définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, le vecteur (X_0, \dots, X_n) est une variable aléatoire à valeurs dans $(E^{n+1}, \mathcal{E}^{n+1})$, et donc la mesure μ_n est bien définie; il s'agit de la loi conjointe du vecteur (X_0, \dots, X_n) .

Maintenant, puisque $(X_n; n \geq 0)$ est une chaîne de Markov avec noyau markovien P , (1.3) implique que pour tout sous-ensemble de E^{n+1} pouvant s'écrire comme $A_0 \times A_1 \times \dots \times A_n$ avec $A_0, \dots, A_n \in \mathcal{E}$, nous avons

$$\mu_n(A_0 \times \dots \times A_n) = \int_{A_0} \dots \int_{A_{n-1}} \lambda(dx_0) P(x_0, dx_1) \dots P(x_{n-1}, A_n). \quad (1.6)$$

Remarquons qu'avec $n = 0$, (1.6) devient (1.5).

Inversement, si une suite de variables aléatoires à valeurs dans (E, \mathcal{E}) satisfait (1.6) pour un certain noyau markovien P et pour une certaine mesure de probabilité λ , et ceci pour tout $n \geq 1$, alors cette suite est une chaîne de Markov avec noyau markovien P et mesure de probabilité initiale λ . Plus précisément, nous avons:

Théorème 1.1. *Soit $(X_n; n \geq 0)$ une suite de variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et à valeurs dans un espace mesurable (E, \mathcal{E}) . Soit P un noyau markovien et λ une mesure de probabilité définies sur (E, \mathcal{E}) . Alors $(X_n; n \geq 0)$ est une chaîne de Markov avec noyau markovien P et mesure de probabilité initiale λ si et seulement si, pour tout $n \geq 0$ et pour tous $A_0, A_1, \dots, A_n \in \mathcal{E}$,*

$$\mathbb{P}[X_0 \in A_0, X_1 \in A_1, \dots, X_n \in A_n] = \int_{A_0} \cdots \int_{A_{n-1}} \lambda(dx_0) P(x_0, dx_1) \cdots P(x_{n-1}, A_n). \quad (1.7)$$

1.2.2 Existence

À chaque chaîne de Markov (au sens de la définition 1.2) correspond un noyau markovien P et une mesure de probabilité initiale λ . La présente section s'attarde à la question suivante: étant donné une mesure de probabilité λ et un noyau markovien P arbitraires définis sur un espace mesurable (E, \mathcal{E}) , existe-t-il un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ sur lequel il est possible de définir une suite de variables aléatoires à valeurs dans (E, \mathcal{E}) , et tel que cette suite sera une chaîne de Markov avec loi initiale λ et noyau markovien P ? Nous verrons qu'en général, la réponse est oui; il nous suffira simplement de supposer que E est un espace métrique complet et séparable et que \mathcal{E} est la famille de ses boréliens. Ce résultat nous permettra de "construire" des chaînes de Markov à valeurs dans (E, \mathcal{E}) à partir d'un noyau markovien P et d'une mesure de probabilité initiale λ totalement arbitraires, sans nous préoccuper de l'espace $(\Omega, \mathcal{F}, \mathbb{P})$ et sans nous demander si une telle chaîne existe vraiment.

Nous présentons ici les grandes lignes de la construction d'une chaîne de Markov à valeurs dans un espace (E, \mathcal{E}) , avec mesure de probabilité initiale λ et noyau markovien

P arbitraires. Les détails techniques de la construction sont disponibles par exemple dans Revuz (1984) et Doob (1953).

Fixons $n \geq 0$. Considérons l'espace produit E^{n+1} , et l'algèbre \mathcal{A}^{n+1} contenant tous les sous-ensembles de E^{n+1} pouvant s'écrire comme une union finie d'ensembles de la forme $A_0 \times A_1 \times \cdots \times A_n$, où $A_0, \dots, A_n \in \mathcal{E}$. Définissons maintenant la mesure de probabilité μ_n^* sur l'algèbre \mathcal{A}^{n+1} par

$$\mu_n^*(A_0 \times A_1 \times A_2 \times \cdots \times A_n) = \int_{A_1} \cdots \int_{A_{n-1}} \lambda(dx_1)P(x_1, dx_2)P(x_2, dx_3) \cdots P(x_{n-1}, A_n). \quad (1.8)$$

Cette intégrale est bien définie par la mesurabilité de $P(\cdot, A)$ pour chaque $A \in \mathcal{E}$ et par le fait que $P(x, \cdot)$ est une mesure de probabilité pour chaque $x \in E$. Pour la définition d'une algèbre et d'une mesure de probabilité sur une algèbre, voir par exemple Bartle (1995, page 97).

Par le théorème d'extension de Carathéodory (voir par exemple Bartle, 1995, page 101), la mesure μ_n^* possède une extension unique μ_n sur l'espace mesurable $(E^{n+1}, \mathcal{E}^{n+1})$ (que nous avons défini à la page 11). En répétant cette procédure pour tout $n \geq 0$, ces constructions nous fournissent une suite d'espaces probabilisés $((E^{n+1}, \mathcal{E}^{n+1}, \mu_n); n \geq 0)$ dont les mesures de probabilité μ_n sont compatibles, c'est-à-dire que ces mesures satisfont, pour tout $n \geq 0$,

$$\mu_{n+1}(A_0 \times \cdots \times A_n \times E) = \mu_n(A_0 \times \cdots \times A_n), \quad A_0, \dots, A_n \in \mathcal{E}. \quad (1.9)$$

Dénotons par E^∞ l'espace produit $E \times E \times \cdots$ et par \mathcal{E}^∞ la σ -algèbre engendrée par les ensembles cylindriques, c'est-à-dire les ensembles de la forme $A_0 \times \cdots \times A_n \times E \times E \times \cdots$, où $A_0, \dots, A_n \in \mathcal{E}$. En vertu du théorème d'extension de Kolmogorov (voir par exemple Durrett, 1991, pages 26-27), l'hypothèse faite sur la nature de l'espace (E, \mathcal{E}) et la compatibilité des mesures $(\mu_n; n \geq 0)$ impliquent qu'il existe une mesure de probabilité \mathbb{P} unique, définie sur (Ω, \mathcal{F}) , où $\Omega = E^\infty$ et $\mathcal{F} = \mathcal{E}^\infty$, telle que pour tout $n \geq 0$ et pour tous $A_0, \dots, A_n \in \mathcal{E}$,

$$\mathbb{P}[A_0 \times \cdots \times A_n \times E \times E \times \cdots] = \mu_n(A_0 \times \cdots \times A_n).$$

Pour chaque $\omega = (\omega_0, \omega_1, \dots) \in \Omega$, posons maintenant

$$X_n(\omega) = \omega_n \quad \text{pour tout } n \geq 0.$$

La suite $(X_n; n \geq 0)$ ainsi définie est une chaîne de Markov à valeurs dans (E, \mathcal{E}) , avec noyau markovien P et mesure de probabilité initiale λ , puisqu'elle satisfait l'équation (1.7) du théorème 1.1.

La chaîne que nous venons de construire s'appelle la *chaîne de Markov canonique associée à P et λ* . L'espace $(E^\infty, \mathcal{E}^\infty)$ est appelé l'*espace canonique*. Il existe plusieurs espaces probabilisés $(\Omega, \mathcal{F}, \mathbb{P})$ sur lesquels nous pouvons définir une chaîne de Markov à valeurs dans (E, \mathcal{E}) avec noyau markovien P et mesure de probabilité initiale λ . Nous en avons identifié un.

Remarque 1.1. Par abus de langage, nous dirons parfois d'un noyau markovien P qu'il est *le* noyau markovien d'une chaîne de Markov $(X_n; n \geq 0)$. Cependant, une chaîne de Markov ne possède pas nécessairement un seul noyau markovien; il serait donc plus exact de dire que P est *un* noyau markovien pour $(X_n; n \geq 0)$. En effet, il est facile de construire une chaîne de Markov $(X_n; n \geq 0)$ à laquelle nous pouvons associer deux noyaux markoviens complètement différents. Voici un exemple très simple: une chaîne de Markov à valeurs dans $\{1, 2, 3\}$, avec vecteur de probabilités initiales $(1/2, 1/2, 0)$ et matrice de transition

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

est aussi une chaîne de Markov avec vecteur de probabilités initiales $(1/2, 1/2, 0)$ et matrice de transition

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

Puisque l'état initial de la chaîne de Markov sera 1 ou 2 avec probabilité 1, et qu'elle ne visitera jamais l'état 3 par la suite, la dernière ligne de la matrice de transition peut être définie arbitrairement. En général, si $(X_n; n \geq 0)$ est une chaîne de Markov à valeurs dans un espace mesurable (E, \mathcal{E}) avec noyau markovien P_1 , et qu'elle satisfait $\mathbb{P}[X_n \in B] = 0$ pour tout $n \geq 0$ et pour un certain ensemble $B \in \mathcal{E}$, alors tout autre noyau P_2 tel que

$$P_2(x, A) = P_1(x, A) \quad \text{pour tous } A \in \mathcal{E} \text{ et } x \in B^c$$

peut aussi servir de noyau markovien pour $(X_n; n \geq 0)$.

Inversement, un noyau markovien n'identifie pas une seule chaîne de Markov. Il identifie plutôt une *famille* de chaînes de Markov, c'est-à-dire la famille de toutes les chaînes de Markov avec noyau markovien P et mesure de probabilité initiale arbitraire.

1.2.3 Probabilités de transition d'ordre supérieur

Considérons une mesure de probabilité λ et un noyau markovien P définis sur un espace mesurable (E, \mathcal{E}) . À partir de λ et P , nous pouvons définir une nouvelle mesure de probabilité λP sur (E, \mathcal{E}) par

$$\lambda P(A) = \int_E P(x, A) \lambda(dx).$$

Supposons maintenant que P_1 et P_2 soient deux noyaux markoviens définis sur un espace mesurable (E, \mathcal{E}) . À l'aide de P_1 et P_2 , définissons une nouvelle fonction sur $E \times \mathcal{E}$ par

$$P_1 P_2(x, A) = \int_E P_1(x, dy) P_2(y, A), \quad A \in \mathcal{E}.$$

Cette fonction $P_1 P_2$ est aussi un noyau markovien et est appelée le *noyau produit* de P_1 et P_2 .

Le produit de noyaux markoviens est une opération associative, c'est-à-dire que si P_1, P_2, P_3 sont des noyaux markoviens définis sur (E, \mathcal{E}) , alors $(P_1 P_2) P_3 = P_1 (P_2 P_3)$. Cette propriété nous permet de définir, sans risque d'ambiguïté, la $n^{\text{ième}}$ puissance P^n

d'un noyau markovien P . Celle-ci est définie en posant d'abord

$$P^0(x, A) = \delta_x(A) = \begin{cases} 1 & \text{si } x \in A, \\ 0 & \text{sinon,} \end{cases}$$

et ensuite, pour $n \geq 1$, de façon itérative par

$$P^n = PP^{n-1}.$$

La propriété d'associativité de l'opération produit de noyaux nous permet aussi d'écrire

$$P^n = P^m P^{n-m}, \quad 0 \leq m \leq n,$$

pour tout entier positif¹ n . Cette dernière équation est souvent appelée l'équation de Chapman-Kolmogorov (voir par exemple Meyn et Tweedie, 1993, page 67).

Si $(X_n; n \geq 0)$ est une chaîne de Markov avec noyau markovien P et mesure de probabilité initiale λ , la mesure de probabilité λP^n est simplement la loi (inconditionnelle) de X_n . En effet, en utilisant (1.7), nous avons, pour tout $A \in \mathcal{E}$ et pour tout $n \geq 0$,

$$\begin{aligned} \mathbb{P}[X_n \in A] &= \mathbb{P}[X_0 \in E, \dots, X_{n-1} \in E, X_n \in A] \\ &= \int_E \cdots \int_E \lambda(dx_0) P(x_0, dx_1) \cdots P(x_{n-1}, A) \\ &= \int_E \lambda(dx_0) P^n(x_0, A) \\ &= \lambda P^n(A). \end{aligned} \tag{1.10}$$

Le noyau markovien P^n fournit les probabilités de transition en n étapes de la chaîne: pour tout $A \in \mathcal{E}$ et pour tous $m \geq 0$ et $n \geq 1$, $P^n(X_m, A)$ est une version de la probabilité conditionnelle $\mathbb{P}[X_{n+m} \in A | X_m]$. En effet, en utilisant (1.10) et encore une fois (1.7), nous avons, pour tous $A, B \in \mathcal{E}$ et pour tous $m \geq 0$ et $n \geq 1$,

$$\begin{aligned} \mathbb{P}[X_m \in A, X_{m+n} \in B] &= \mathbb{P}[X_m \in A, X_{m+1} \in E, \dots, X_{m+n-1} \in E, X_{m+n} \in B] \\ &= \int_A \int_E \cdots \int_E \lambda P^m(dx_m) P(x_m, dx_{m+1}) \cdots P(x_{m+n-1}, B) \\ &= \int_A \lambda P^m(dx_m) P^n(x_m, B). \end{aligned}$$

¹Dans ce mémoire, un nombre réel x est dit positif si $x > 0$ et non négatif si $x \geq 0$.

Mentionnons qu'étudier le comportement asymptotique d'une chaîne de Markov avec noyau markovien P et mesure de probabilité initiale λ est en fait équivalent à étudier le comportement asymptotique de la suite de mesures de probabilité $(\lambda P^n; n \geq 0)$.

1.3 Classification des chaînes de Markov

Les concepts d'irréductibilité, de récurrence et de périodicité, bien connus dans le cas dénombrable, seront ici définis dans le cas général.

Un peu comme pour la matrice des probabilités de transition dans le cas discret, la classification des chaînes de Markov générales selon ces critères dépend de la structure de communication de leurs noyaux markoviens respectifs et est indépendante de leurs mesures de probabilité initiales. Ceci implique que si une chaîne de Markov avec noyau markovien P possède une de ces propriétés, alors toutes les chaînes de Markov avec noyau markovien P la possèdent aussi, peu importe leurs mesures de probabilité initiales. Malgré que ces propriétés s'appliquent en fait aux noyaux markoviens, il nous arrivera parfois de les attribuer aux chaînes de Markov correspondantes et de parler d'une chaîne de Markov irréductible, récurrente, etc.

Pour un noyau markovien donné P , ces propriétés seront parfois énoncées de façon plus probabiliste en considérant la probabilité de certains événements relativement à une chaîne de Markov avec noyau markovien P (comme par exemple une chaîne de Markov canonique). Dans ce qui suit, nous utiliserons le symbole \mathbb{P}_x , $x \in E$ pour identifier une mesure de probabilité sur un espace mesurable quelconque (Ω, \mathcal{F}) , sur lequel a été définie une chaîne de Markov avec noyau markovien P et état initial x (c'est-à-dire avec δ_x , la mesure de Dirac en x , comme loi initiale). Par exemple, $\mathbb{P}_x[X_n \in A]$ représente la probabilité qu'une chaîne de Markov avec noyau markovien P et état initial x se retrouve dans A au temps n .

Remarque 1.2. À partir de maintenant, nous supposerons toujours que tout espace mesurable (E, \mathcal{E}) sur lequel est défini un noyau markovien P est arbitraire, en autant

que sa σ -algèbre \mathcal{E} puisse être générée à l'aide d'une famille dénombrable de sous-ensembles de E . Cette hypothèse permet d'éviter des problèmes techniques dans les démonstrations de certains théorèmes importants. Pour plus de détails à ce sujet, voir par exemple, Orey (1971) et Revuz (1984). Notons que cette condition n'est pas très restrictive, puisqu'elle est satisfaite par la plupart des espaces mesurables que l'on retrouve dans les applications. Voici quelques exemples:

- 1) un ensemble fini E muni d'une σ -algèbre arbitraire, comme par exemple $E = \{1, 2, \dots, k\}$ et $\mathcal{E} = \mathcal{P}(E)$, la famille de tous les sous-ensembles de E ;
- 2) un ensemble dénombrable E muni d'une σ -algèbre arbitraire, comme par exemple $E = \{1, 2, \dots\}$ et $\mathcal{E} = \mathcal{P}(E)$;
- 3) un borélien E de \mathbb{R}^d et $\mathcal{E} = \mathcal{B}_E$, la famille des boréliens de \mathbb{R}^d inclus dans E ;
- 4) un espace métrique complet et séparable E et \mathcal{E} , la famille des boréliens de E .

Rappelons que l'hypothèse spécifiant que E est un espace métrique complet et séparable et que \mathcal{E} est la famille des boréliens de E (ce qui inclut notamment tous les exemples que nous venons de donner) nous a déjà bien servis. Premièrement, elle a assuré l'existence d'un noyau markovien lorsque $(X_n; n \geq 0)$ était une chaîne de Markov définie sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans (E, \mathcal{E}) (voir section 1.2.1). Deuxièmement, elle nous a permis d'utiliser le théorème d'extension de Kolmogorov pour construire un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et une chaîne de Markov à partir d'un noyau markovien P défini sur (E, \mathcal{E}) (voir section 1.2.2).

1.3.1 φ -irréductibilité

Le concept d'irréductibilité tel que défini pour une chaîne de Markov à valeurs dans un espace dénombrable s'applique difficilement au cas général. En effet, il serait impensable d'imposer que pour tous $x, y \in E$, il existe un entier positif n tel que

$$P^n(x, \{y\}) > 0,$$

puisqu'alors toutes les chaînes de Markov irréductibles seraient nécessairement à valeurs dans un espace dénombrable.

Pour définir un concept d'irréductibilité plus général, nous nous contentons d'imposer que quel que soit l'état initial de la chaîne, tous les "gros" sous-ensembles de E contenus dans \mathcal{E} aient une probabilité positive d'être visités.

Définition 1.3. Soit φ une mesure σ -finie définie sur un espace mesurable (E, \mathcal{E}) , et posons $\mathcal{E}_\varphi^+ = \{A \in \mathcal{E} : \varphi(A) > 0\}$. Un noyau markovien défini sur (E, \mathcal{E}) est dit φ -irréductible si, quels que soient $x \in E$ et $A \in \mathcal{E}_\varphi^+$, il existe un entier $m \geq 1$ tel que $P^m(x, A) > 0$.

En termes d'une chaîne de Markov $(X_n; n \geq 0)$ avec noyau markovien P et état initial x , l'existence d'un entier $m \geq 1$ tel que $P^m(x, A) > 0$ pour un certain $A \in \mathcal{E}$ signifie que la probabilité que $(X_n; n \geq 0)$ visite l'ensemble A est positive. La φ -irréductibilité exige que ceci soit respecté pour tout $x \in E$ et pour tout $A \in \mathcal{E}_\varphi^+$.

Même lorsque P est un noyau markovien défini sur un espace dénombrable (E, \mathcal{E}) , la notion de φ -irréductibilité est beaucoup plus générale que la notion usuelle d'irréductibilité, c'est-à-dire l'irréductibilité au sens de (1.1). Voici un exemple pour l'illustrer.

Exemple 1.1. Considérons une chaîne de Markov à valeurs dans $\{1, 2, 3\}$, avec matrice des probabilités de transition

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

Si l'état initial de la chaîne est 1 ou 2, la probabilité qu'elle visite à un certain moment l'état 3 est nulle. La matrice n'est donc pas irréductible au sens usuel. Par contre, en partant des états 1, 2 ou 3, la probabilité qu'elle visite par la suite les états 1 et 2 est positive. Si nous dénotons maintenant par φ n'importe quelle mesure sur $\{1, 2, 3\}$ accordant un poids nul à l'état 3, nous pouvons affirmer que la matrice est φ -irréductible. \square

En général, si P est un noyau markovien défini sur un espace mesurable (E, \mathcal{E}) , où E est dénombrable et où \mathcal{E} est $\mathcal{P}(E)$, la famille de tous les sous-ensembles de E , l'irréductibilité (au sens usuel) correspond donc à la φ -irréductibilité où φ est n'importe quelle mesure accordant un poids positif à chaque élément de E (par exemple, la mesure de dénombrement sur (E, \mathcal{E})).

Considérons maintenant, pour tout $x \in E$, une chaîne de Markov avec noyau markovien P et état initial x . Posons, pour tout $A \in \mathcal{E}$,

$$T_A = \inf\{n \geq 1 : X_n \in A\}, \quad A \in \mathcal{E}.$$

T_A est alors une variable aléatoire puisque l'ensemble $\{T_A = n\}$ est mesurable pour tout n . En effet, $\{T_A = n\} = \{X_1 \in A^c, \dots, X_{n-1} \in A^c, X_n \in A\} \in \sigma(X_1, \dots, X_n) \in \mathcal{F}$, où $\sigma(X_1, \dots, X_n)$ est la σ -algèbre engendrée par X_1, \dots, X_n . La variable aléatoire T_A est appelée *le temps de première visite de la chaîne dans A*.

La définition de φ -irréductibilité peut maintenant être énoncée en termes des variables aléatoires $(T_A; A \in \mathcal{E}_\varphi^+)$: un noyau markovien P est φ -irréductible si, pour tout $x \in E$ et pour tout $A \in \mathcal{E}_\varphi^+$, nous avons

$$\mathbb{P}_x[T_A < \infty] > 0.$$

Les deux définitions sont équivalentes, puisque $\mathbb{P}_x[T_A < \infty] > 0$ si et seulement si il existe un entier positif m tel que $P^m(x, A) > 0$. Cette deuxième interprétation de la φ -irréductibilité nous sera utile pour faire le lien avec le concept de *réurrence* un peu plus loin.

Lorsqu'un noyau P est φ -irréductible pour une certaine mesure φ , cette dernière est appelée une *mesure d'irréductibilité* pour P . Il est clair qu'une mesure d'irréductibilité n'est pas unique; en effet, si φ est une mesure d'irréductibilité et que ψ est une mesure σ -finie sur (E, \mathcal{E}) , absolument continue par rapport à φ (i.e. $A \in \mathcal{E}$ et $\varphi(A) = 0 \Rightarrow \psi(A) = 0$), alors ψ est aussi une mesure d'irréductibilité.

Inversement, pour un noyau markovien φ -irréductible P , il existe toujours une mesure d'irréductibilité qui accordera une mesure nulle à un ensemble $A \in \mathcal{E}$ si et

seulement si la condition

$$\mathbb{P}_x[T_A < \infty] > 0 \quad \text{pour tout } x \in E \quad (1.11)$$

n'est pas respectée. Une telle mesure est appelée une *mesure d'irréductibilité maximale* pour P . La démonstration de la proposition suivante est présentée dans Meyn et Tweedie (1993, page 88).

Proposition 1.1. *Soit P un noyau markovien φ -irréductible sur un espace mesurable (E, \mathcal{E}) . Alors il existe une mesure ψ telle que:*

- i) P est ψ -irréductible;*
- ii) si φ' est une mesure σ -finie sur (E, \mathcal{E}) , alors P est φ' -irréductible si et seulement si φ' est absolument continue par rapport à ψ ;*
- iii) si $\psi(A) = 0$, alors $\psi(\{x \in E : \mathbb{P}_x[T_A < \infty] > 0\}) = 0$.*

En un sens, une mesure d'irréductibilité maximale ψ est une mesure d'irréductibilité naturelle. Elle décrit mieux qu'une mesure d'irréductibilité arbitraire les sous-ensembles mesurables de E qui risquent d'être visités par un chaîne avec noyau P , et ceci, en fonction de son état initial x . Premièrement, elle accorde une mesure positive à tous les sous-ensembles dans \mathcal{E} qui ont une probabilité positive d'être visités quel que soit l'état initial $x \in E$. Deuxièmement, elle accorde non seulement une mesure nulle à tout sous-ensemble $A \in \mathcal{E}$ qui est inaccessible à partir d'au moins un $x \in E$, mais elle permet aussi d'affirmer que si A est un sous-ensemble dans \mathcal{E} tel que $\psi(A) = 0$, alors $\mathbb{P}_x[X_n \in A \text{ pour un certain } n \geq 1] = 0$ pour ψ -presque tout $x \in E$.

Toutefois, une mesure d'irréductibilité maximale n'est pas unique non plus. Toute mesure équivalente (i.e. qui assigne une mesure nulle exactement aux mêmes sous-ensembles mesurables de E) est aussi une mesure d'irréductibilité maximale. Cette fois, par contre, l'ensemble $\mathcal{E}_\psi^+ = \{A \in \mathcal{E} : \psi(A) > 0\}$ est indépendant de la mesure d'irréductibilité maximale ψ . Pour cette raison, nous utiliserons simplement le symbole \mathcal{E}^+ pour faire référence à l'ensemble $\{A \in \mathcal{E} : \psi(A) > 0\}$ où ψ est une mesure d'irréductibilité maximale.

1.3.2 Périodicité

La φ -irréductibilité d'un noyau P assure que tout sous-ensemble $A \in \mathcal{E}$ tel que $\varphi(A) > 0$ a une probabilité positive d'être visité par une chaîne de Markov avec noyau P et état initial $x \in E$ arbitraire. Il se peut alors qu'une visite dans ces sous-ensembles ne soit possible seulement qu'à certains temps n spécifiques et que la chaîne ait alors un comportement "cyclique". Voici deux exemples très simples illustrant des comportements cycliques.

Exemple 1.2. Considérons une chaîne de Markov à valeurs dans $\{1, 2, 3\}$ avec matrice de transition

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

Si φ est la mesure de dénombrement sur $\{1, 2, 3\}$, alors cette matrice est φ -irréductible. Toutefois, la chaîne de Markov ne peut visiter les états 1, 2 et 3 qu'à certains temps n particuliers. Supposons que l'état initial de la chaîne soit 1. Nous avons alors $\mathbb{P}[X_n \in \{2, 3\}] = 1$ pour tout n impair et $\mathbb{P}[X_n \in \{2, 3\}] = 0$ pour tout n pair. La chaîne visite alors successivement les ensembles $\{1\}$ et $\{2, 3\}$. Avec probabilité 1, la chaîne visitera tous les états dans $\{1, 2, 3\}$, mais ces visites se feront selon un patron cyclique. \square

Exemple 1.3. Soit $E = \{x \in \mathbb{R} : 0 < x \leq 3\}$ et \mathcal{E} , la σ -algèbre des boréliens de E . Considérons le noyau markovien P , défini sur (E, \mathcal{E}) pour tous $x \in E$ et $A \in \mathcal{E}$ par

$$P(x, A) = \begin{cases} \int_A \mathbf{1}_{(1, 2]}(y) dy & \text{si } 0 < x \leq 1, \\ \int_A \mathbf{1}_{(2, 3]}(y) dy & \text{si } 1 < x \leq 2, \\ \int_A \mathbf{1}_{(0, 1]}(y) dy & \text{si } 2 < x \leq 3. \end{cases}$$

Remarquons que $P(x, \cdot)$ est la loi uniforme sur $(1, 2]$ lorsque $x \in (0, 1]$, la loi uniforme sur $(2, 3]$, lorsque $x \in (1, 2]$, et la loi uniforme sur $(0, 1]$ lorsque $x \in (2, 3)$. Si φ représente n'importe quelle mesure absolument continue par rapport à la mesure de

Lebesgue sur (E, \mathcal{E}) , alors P est φ -irréductible. Par contre, pour tout $x \in (0, 1]$, le noyau P satisfait

$$P^n(x, (1, 2]) > 0 \quad \text{si et seulement si } n \in \{1, 4, 7, 10, \dots\},$$

$$P^n(x, (2, 3]) > 0 \quad \text{si et seulement si } n \in \{2, 5, 8, 11, \dots\}$$

et

$$P^n(x, (0, 1]) > 0 \quad \text{si et seulement si } n \in \{0, 3, 6, 9, \dots\}.$$

Une situation similaire se produit lorsque $x \in (1, 2]$ et lorsque $x \in (2, 3]$. Si $(X_n; n \geq 0)$ est une chaîne de Markov avec noyau markovien P et état initial $x_0 \in (0, 1]$, elle satisfait alors avec probabilité 1,

$$X_1 \in (1, 2], X_2 \in (2, 3], X_3 \in (0, 1], X_4 \in (1, 2], \dots$$

Si A est n'importe quel sous-ensemble dans \mathcal{E}^+ , la chaîne de Markov le visitera avec probabilité 1. Néanmoins, si A est presque entièrement contenu dans $(0, 1]$ (c'est-à-dire si A est tel que la mesure de Lebesgue de l'ensemble $\{x \in A : x \notin (0, 1]\}$ est égale à 0), par exemple, cette visite aura lieu nécessairement à un temps $n \in \{3, 6, 9, 12, \dots\}$. Plus généralement, toute chaîne de Markov avec noyau markovien P visitera avec probabilité 1 les sous-ensembles $(0, 1]$, $(1, 2]$ et $(2, 3]$ les uns après les autres, dans l'ordre (celui visité par X_1 pouvant changer selon la loi initiale). \square

Définition 1.4. Soit P un noyau markovien φ -irréductible pour une certaine mesure φ . Une famille $\{C_1, C_2, \dots, C_d\}$ de sous-ensembles disjoints dans \mathcal{E} est appelée un *cycle de longueur d* pour P , si

$$\varphi \left[\left(\bigcup_{j=1}^d C_j \right)^c \right] = 0,$$

$$P(x, C_{j+1}) = 1, \quad x \in C_j, \quad 1 \leq j \leq d-1$$

et

$$P(x, C_1) = 1, \quad x \in C_d.$$

Si $\{C_1, C_2, \dots, C_d\}$ est un cycle de longueur d pour un noyau markovien φ -irréductible P , alors toute chaîne avec noyau markovien P visite de façon cyclique les sous-ensembles C_1, \dots, C_d avec probabilité 1. Par exemple, si $d = 4$, nous avons pour tout $x \in C_3$, $\mathbb{P}_x[X_1 \in C_4, X_2 \in C_1, X_3 \in C_2, \dots] = 1$. Notons que si $\{C_1, \dots, C_d\}$ est un cycle

de longueur d par rapport à une mesure d'irréductibilité φ , alors $\{C_1, \dots, C_d\}$ est aussi un cycle de longueur d par rapport à n'importe quelle autre mesure d'irréductibilité pour P . La notion de cycle est donc indépendante de la mesure d'irréductibilité utilisée.

Pour un noyau markovien φ -irréductible P , considérons l'ensemble \mathcal{D} de tous les entiers $d \geq 1$ pour lesquels il existe un cycle de longueur d . Cet ensemble est non vide puisque $P(x, E) = 1$ pour tout $x \in E$ et il contient donc au moins l'élément 1. Il est aussi possible de montrer que la φ -irréductibilité de P implique que l'ensemble \mathcal{D} est nécessairement fini (voir par exemple Orey, 1971, pages 10-14). Nous avons alors:

Théorème 1.2. *Soit P un noyau markovien φ -irréductible pour une certaine mesure φ . Soit \mathcal{D} l'ensemble de tous les entiers $d \geq 1$ pour lesquels il existe un cycle de longueur d pour P . Dénotons par p le plus grand élément de \mathcal{D} . Alors:*

- i) p est indépendant de la mesure d'irréductibilité φ ;*
- ii) si d est un autre élément de \mathcal{D} , alors p est un multiple de d ;*
- iii) si C_1, C_2, \dots, C_p est un cycle de longueur p , et si D_1, D_2, \dots, D_p est un autre cycle de longueur p , alors, en numérotant les sous-ensembles D_1, \dots, D_p dans un autre ordre si nécessaire, nous avons*

$$\varphi[\{x \in C_i : x \notin D_i\}] = \varphi[\{x \in D_i : x \notin C_i\}] = 0 \quad \text{pour tout } i \in \{1, \dots, p\}.$$

La démonstration de ce théorème est présentée notamment dans Orey (1971, page 13). L'entier p est appelé la *période* de P . En vertu du théorème précédent, p est indépendant de la mesure d'irréductibilité φ et le cycle C_1, \dots, C_p correspondant à p est unique modulo les sous-ensembles φ -nuls. Remarquons que la période du noyau markovien de l'exemple 1.2 est 2, et qu'un cycle correspondant possible est $\{\{1\}, \{2, 3\}\}$. Pour le noyau de l'exemple 1.3, la période est 3 et un cycle correspondant possible est $\{(0, 1], (1, 2], (2, 3]\}$.

Mentionnons finalement que si P est un noyau φ -irréductible de période $p \geq 2$, il est dit *périodique*. Lorsque $p = 1$, il est dit *apériodique*.

1.3.3 Récurrence

Tout comme l'irréductibilité, le concept de récurrence tel que défini dans le cas dénombrable a besoin d'une modification pour s'adapter au cas général. Nous pourrions toutefois constater que la signification probabiliste de la récurrence reste sensiblement la même.

Lorsque P est un noyau markovien défini sur un espace mesurable (E, \mathcal{E}) où E est dénombrable, il est irréductible si

$$\mathbb{P}_i[T_{\{j\}} < \infty] > 0 \quad \text{pour tous } i, j \in E, \quad (1.12)$$

et récurrent si

$$\mathbb{P}_i[T_{\{j\}} < \infty] = 1 \quad \text{pour tous } i, j \in E \quad (1.13)$$

(voir par exemple Hoel, Port et Stone, 1972, page 21). Pour définir un concept d'irréductibilité s'adaptant au cas où (E, \mathcal{E}) est un espace mesurable général, nous avons imposé que (1.12) soit satisfaite lorsqu'on remplace $\{j\}$ par n'importe quel sous-ensemble $A \in \mathcal{E}$ tel que $\varphi(A) > 0$ pour une certaine mesure φ . C'est ce que nous ferons aussi pour définir un concept de récurrence plus général: nous remplacerons $\{j\}$ dans (1.13) par les sous-ensembles mesurables de E auxquels la mesure φ accorde une probabilité positive; toutefois, nous changerons le "pour tout $i \in E$ " par "pour φ -presque tout $x \in E$ ".

Définition 1.5. Un noyau φ -irréductible P défini sur un espace mesurable (E, \mathcal{E}) est dit *récurrent* si, pour tout $A \in \mathcal{E}_\varphi^+$,

$$\mathbb{P}_x[T_A < \infty] = 1 \quad \text{pour } \varphi\text{-presque tout } x \in E. \quad (1.14)$$

Supposons que P soit un noyau markovien sur un espace mesurable (E, \mathcal{E}) et que P soit aussi φ -irréductible pour une certaine mesure φ . Cela signifie que si $(X_n; n \geq 0)$ est une chaîne de Markov avec noyau markovien P , nous avons pour tout $x \in E$,

$$\mathbb{P}_x[T_A < \infty] > 0 \quad \text{pour tout } A \in \mathcal{E}_\varphi^+.$$

La *récurrence* est une condition beaucoup plus forte; elle exige non seulement que cette probabilité soit positive pour tout $x \in E$, mais elle exige aussi que cette probabilité soit égale à 1 pour φ -presque tout $x \in E$.

Pour qu'un noyau P soit qualifié de récurrent, plusieurs auteurs exigent que P soit φ -irréductible pour une certaine mesure d'irréductibilité φ , et que

$$\mathbb{P}_x[T_A < \infty] = 1 \quad \text{pour } \psi\text{-presque tout } x \in E,$$

quel que soit $A \in \mathcal{E}^+$, pour une certaine mesure d'irréductibilité maximale ψ . Par contre, Nummelin (1984, pages 42-43) montre que ce type de récurrence est équivalent à exiger que (1.14) soit satisfaite pour tout $A \in \mathcal{E}_\varphi^+$, où φ est n'importe quelle mesure d'irréductibilité, maximale ou non. La récurrence est donc un concept indépendant de la mesure d'irréductibilité utilisée pour vérifier (1.14); si φ et φ' sont deux mesures d'irréductibilité arbitraires pour P , alors nous avons

$$\mathbb{P}_x[T_A < \infty] = 1 \quad \text{pour tout } A \in \mathcal{E}_\varphi^+ \text{ et pour } \varphi\text{-presque tout } x \in E$$

si et seulement si

$$\mathbb{P}_x[T_A < \infty] = 1 \quad \text{pour tout } A \in \mathcal{E}_{\varphi'}^+ \text{ et pour } \varphi'\text{-presque tout } x \in E.$$

Nummelin (1984, page 42) montre aussi que la récurrence peut être vérifiée à l'aide de la condition suivante:

Théorème 1.3. *Un noyau φ -irréductible P défini sur un espace mesurable (E, \mathcal{E}) est récurrent si et seulement si, pour tout $A \in \mathcal{E}_\varphi^+$,*

$$\mathbb{P}_x[X_n \in A \text{ i.s.}] = 1 \quad \text{pour } \varphi\text{-presque tout } x \in E, \quad (1.15)$$

où "i.s." signifie "infiniment souvent".

Encore une fois, la condition (1.15) peut être satisfaite à l'aide de n'importe quelle mesure d'irréductibilité φ pour que le noyau P soit récurrent. Remarquons que le fait que les conditions de la définition 1.5 et du théorème 1.3 soient équivalentes ne signifie pas que si nous fixons un $A \in \mathcal{E}_\varphi^+$, les ensembles $\{x \in E : \mathbb{P}_x[T_A < \infty] = 1\}$ et $\{x \in E : \mathbb{P}_x[X_n \in A \text{ i.s.}] = 1\}$ sont identiques. Cela signifie, par contre, que la mesure φ accorde une probabilité égale à 1 aux deux ensembles, quelle que soit la mesure d'irréductibilité φ .

Nous définissons maintenant un type de récurrence plus fort: lorsque la condition (1.14) est respectée pour tout $x \in E$ plutôt que pour φ -presque tout $x \in E$, le noyau P est alors dit *récurrent au sens de Harris*.

Définition 1.6. Un noyau φ -irréductible P est dit *récurrent au sens de Harris* si, pour tout $A \in \mathcal{E}_\varphi^+$,

$$\mathbb{P}_x[T_A < \infty] = 1 \quad \text{pour tout } x \in E, \quad (1.16)$$

ou, de façon équivalente, si

$$\mathbb{P}_x[X_n \in A \text{ i.s.}] = 1 \quad \text{pour tout } x \in E.$$

Comme dans le cas de la récurrence, le concept de récurrence au sens de Harris est indépendant de la mesure d'irréductibilité choisie pour identifier les éléments de \mathcal{E} qui doivent satisfaire (1.16); si (1.16) est satisfaite pour tout $A \in \mathcal{E}_\varphi^+$, elle l'est aussi si on remplace φ par n'importe quelle autre mesure d'irréductibilité (voir par exemple Nummelin, 1984, page 43). Pour conclure cette section, notons que si P est un noyau markovien défini sur un espace dénombrable (E, \mathcal{E}) et que P est φ -irréductible, où φ est la mesure de dénombrement sur (E, \mathcal{E}) (i.e. irréductible au sens de (1.12)), alors la récurrence de P et la récurrence au sens de Harris de P sont deux concepts équivalents, puisque le seul ensemble de mesure φ nulle est alors l'ensemble vide.

1.4 Convergence

1.4.1 Mesures stationnaires

Nous généralisons ici le concept de loi stationnaire. Par la suite, les noyaux markoviens récurrents seront divisés en deux groupes: les noyaux markoviens récurrents qui possèdent une loi stationnaire, et les noyaux markoviens qui n'en possèdent pas. Nous verrons qu'une chaîne de Markov avec noyau markovien P possédant une loi stationnaire a un comportement aléatoire plus "stable" qu'une chaîne de Markov avec noyau markovien n'en possédant pas. Voici donc les définitions d'une *loi stationnaire* pour un noyau markovien P et du concept plus général de *mesure stationnaire*:

Définition 1.7. Soit P un noyau markovien défini sur un espace mesurable (E, \mathcal{E}) . Une mesure σ -finie π sur (E, \mathcal{E}) est appelée une *mesure stationnaire* pour P si, pour

tout $A \in \mathcal{E}$,

$$\pi(A) = \int_E P(x, A) \pi(dx).$$

Si de plus $\pi(E) = 1$, alors π est appelée une *mesure de probabilité stationnaire* (ou *loi stationnaire*) pour P .

Avant de poursuivre, voici quelques commentaires sur la terminologie *loi stationnaire*. Rappelons qu'une suite de variables aléatoires $(X_n; n \geq 0)$ à valeurs dans un espace mesurable (E, \mathcal{E}) est dite *stationnaire* si, pour tout choix d'entiers $k \geq 1$ et $0 \leq n_1 < \dots < n_k$, la loi du vecteur $(X_{n_1+m}, X_{n_2+m}, \dots, X_{n_k+m})$ est indépendante de la valeur de l'entier $m \geq 0$. En particulier, cela implique que les variables aléatoires X_0, X_1, \dots sont identiquement distribuées. En général, une suite de variables aléatoires identiquement distribuées ne forme pas nécessairement une suite stationnaire. Par contre, une chaîne de Markov $(X_n; n \geq 0)$ pour laquelle les variables aléatoires X_0, X_1, \dots sont identiquement distribuées est aussi une suite stationnaire de variables aléatoires. En effet, supposons que P et λ soient respectivement le noyau markovien et la mesure de probabilité initiale d'une chaîne de Markov $(X_n; n \geq 0)$. Choisissons des entiers $k \geq 1$ et $0 \leq n_1 < \dots < n_k$. Alors, pour tous A_1, \dots, A_k dans \mathcal{E} et pour tout $m \geq 0$, nous avons

$$\mathbb{P}[X_{n_1+m} \in A_1, \dots, X_{n_k+m} \in A_k] = \int_{A_1} \dots \int_{A_{k-1}} \lambda P^m(dx_1) P^{n_1}(x_1, dx_2) \dots P^{n_k}(x_k, A_k).$$

Cette probabilité sera indépendante de m pour tous $A_1, \dots, A_k \in \mathcal{E}$, peu importe le choix de k, n_1, \dots, n_k , si et seulement si les mesures $\lambda, \lambda P, \lambda P^2, \dots$ sont identiques, c'est-à-dire si et seulement si λ est une mesure de probabilité stationnaire pour P . Si λ est stationnaire pour P , les variables aléatoires X_0, X_1, \dots sont donc identiquement distribuées; mais parce que la suite $(X_n; n \geq 0)$ possède la propriété de Markov, elle est aussi stationnaire. Remarquons que si π est une mesure stationnaire finie pour P avec $\pi(E) \neq 1$, cette mesure peut toujours être normalisée pour devenir une loi stationnaire. Évidemment, si $\pi(E) = \infty$, alors π n'a pas d'interprétation probabiliste similaire à celle d'une loi stationnaire.

Le théorème suivant nous fournit une condition suffisante pour assurer l'existence d'une mesure stationnaire π pour un noyau markovien φ -irréductible P . Sous cette condition, la mesure π est l'unique mesure stationnaire de P ; de plus elle est une

mesure d'irréductibilité maximale pour P . Une démonstration est présentée dans Meyn et Tweedie (1993, page 242).

Théorème 1.4. *Soit P un noyau markovien défini sur un espace mesurable (E, \mathcal{E}) . Supposons que P soit φ -irréductible pour une certaine mesure φ . Si P est récurrent, il possède alors une mesure stationnaire, disons π , et cette mesure stationnaire est unique à une constante multiplicative près. De plus, π est une mesure d'irréductibilité maximale pour P .*

Il est clair que si π est une mesure de probabilité stationnaire pour P , alors toute mesure π' de la forme

$$\pi'(A) = \alpha \pi(A), \quad A \in \mathcal{E}$$

pour un certain réel positif α est aussi stationnaire pour P . C'est pourquoi l'unicité d'une mesure stationnaire π ne peut être établie qu'à une constante multiplicative près. Si π est une mesure finie pour P et que P est récurrent, alors la mesure π' obtenue en normalisant π de façon à obtenir $\pi'(E) = 1$ est la seule loi stationnaire de P . Si P est récurrent et que π est une mesure stationnaire infinie pour P , alors toutes les mesures stationnaires pour P sont aussi infinies, et donc P ne possède pas de loi stationnaire.

En général, l'existence d'une mesure stationnaire π pour un noyau markovien P n'implique pas que celui-ci est récurrent. Néanmoins, elle l'implique si π est une mesure finie, et si P est φ -irréductible pour une certaine mesure φ , comme le montrent Meyn et Tweedie (1993, page 231).

Théorème 1.5. *Soit P un noyau markovien défini sur un espace mesurable (E, \mathcal{E}) . Supposons que P soit φ -irréductible pour une certaine mesure φ , et que π soit une mesure stationnaire finie pour P . Alors P est récurrent.*

Nous découvrirons bientôt que si $(X_n; n \geq 0)$ est une chaîne de Markov avec noyau markovien P et que la loi de X_n converge (en un sens à être défini plus loin) vers une certaine mesure de probabilité π , alors π est nécessairement une loi stationnaire pour P . De plus, lorsqu'une loi stationnaire π pour un noyau markovien P existe et est connue, nous disposons alors automatiquement d'une mesure d'irréductibilité potentielle pour P . En effet, les théorèmes 1.4 et 1.5 impliquent que si P est φ -irréductible pour une

certaines mesure φ , il est nécessairement π -irréductible puisqu'alors, π est une mesure d'irréductibilité maximale pour P . L'intérêt de l'existence d'une mesure de probabilité stationnaire π pour un noyau markovien P est donc beaucoup plus que le simple fait de pouvoir construire une chaîne de Markov stationnaire à partir de π et P .

Bien que nous sachions maintenant qu'un noyau récurrent possède une mesure stationnaire unique (à une constante multiplicative près), nous n'avons pas encore d'outils permettant d'identifier cette mesure et de déterminer si celle-ci est finie. L'identification d'une loi stationnaire pour un noyau markovien récurrent donné est une tâche ardue (voir par exemple Meyn et Tweedie 1993, chapitre 10). Cependant, dans le contexte de la simulation markovienne, il n'est pas nécessaire de considérer cet aspect. En effet, ce type de simulation fait intervenir des noyaux markoviens qui sont construits de façon à ce qu'ils possèdent comme loi stationnaire la mesure de probabilité désirée. Tout résultat ayant pour but d'établir l'existence d'une loi stationnaire et d'identifier celle-ci n'est donc d'aucune utilité pour le présent mémoire.

Certains auteurs ajoutent un qualificatif au terme *réurrence* pour faire la différence entre un noyau récurrent possédant une mesure de probabilité stationnaire et un noyau récurrent n'en possédant pas.

Définition 1.8. Soit P un noyau markovien récurrent, défini sur un espace mesurable (E, \mathcal{E}) . Soit π son unique mesure stationnaire (à une constante multiplicative près). Si $\pi(E) < \infty$, alors P est dit *réurrent positif*. Sinon, il est dit *réurrent nul*.

1.4.2 Convergence en variation totale

Pour définir un concept de convergence d'une suite de mesures de probabilité $(\mu_n; n \geq 1)$ vers une autre mesure de probabilité μ , il nous faut bien sûr une notion de distance entre deux mesures de probabilité. Celle qui est habituellement utilisée pour énoncer des résultats de convergence dans le contexte des chaînes de Markov est appelée la *variation totale*.

Définition 1.9. Soit λ une mesure signée et finie, définie sur un espace mesurable

(E, \mathcal{E}) . La *variation totale* de λ , que l'on note par $\|\lambda\|$, est définie par

$$\|\lambda\| = \sup_{f:|f|\leq 1} \left| \int_E f d\lambda \right| = \sup_{A \in \mathcal{E}} \lambda(A) - \inf_{A \in \mathcal{E}} \lambda(A). \quad (1.17)$$

Si μ et ν sont deux mesures finies sur un espace mesurable (E, \mathcal{E}) , alors leur différence $(\mu - \nu)$, définie par

$$(\mu - \nu)(A) = \mu(A) - \nu(A), \quad A \in \mathcal{E},$$

est une mesure signée finie. En particulier, si μ et ν sont deux mesures de probabilité sur (E, \mathcal{E}) , alors $(\mu - \nu)(E) = 0$ et donc

$$\sup_{A \in \mathcal{E}} (\mu - \nu)(A) = - \inf_{A \in \mathcal{E}} (\mu - \nu)(A).$$

La variation totale de $\mu - \nu$ devient alors

$$\|\mu - \nu\| = 2 \sup_{A \in \mathcal{E}} |\mu(A) - \nu(A)|.$$

La variation totale de la différence entre deux mesures de probabilité définies sur un espace mesurable (E, \mathcal{E}) peut servir de notion de distance sur l'espace $\mathcal{M}_{(E, \mathcal{E})}$ de toutes les mesures de probabilité sur (E, \mathcal{E}) . En effet, si μ, ν et λ sont trois mesures de probabilité sur (E, \mathcal{E}) , la variation totale satisfait les quatre propriétés habituelles d'une notion de distance (voir par exemple Lindvall, 1992, page 232):

- 1) $\|\mu - \nu\| \geq 0$;
- 2) $\|\mu - \nu\| = 0$ si et seulement si $\mu = \nu$;
- 3) $\|\mu - \nu\| = \|\nu - \mu\|$;
- 4) $\|\mu - \nu\| \leq \|\mu - \lambda\| + \|\lambda - \nu\|$.

La variation totale nous permet donc de définir une notion de convergence d'une suite de mesures de probabilité $(\mu_n; n \geq 1)$ sur (E, \mathcal{E}) vers une autre mesure de probabilité μ .

Définition 1.10. Soit (E, \mathcal{E}) un espace mesurable. Soit μ et $(\mu_n; n \geq 1)$ des mesures de probabilité définies sur (E, \mathcal{E}) . Alors nous disons que la suite $(\mu_n; n \geq 1)$ converge en variation totale vers μ si

$$\lim_{n \rightarrow \infty} \|\mu_n - \mu\| = 0.$$

Si μ, μ_1, μ_2, \dots sont des mesures de probabilité sur un espace mesurable (E, \mathcal{E}) , telles que $\lim_{n \rightarrow \infty} \|\mu_n - \mu\| = 0$, alors en vertu de la définition 1.9, cela signifie donc que

$$\lim_{n \rightarrow \infty} \mu_n(A) = \mu(A) \quad \text{uniformément sur } \mathcal{E}$$

et que, pour tout réel positif c ,

$$\lim_{n \rightarrow \infty} \int_E f d\mu_n = \int_E f d\mu \quad \text{uniformément sur } \{f \text{ mesurable} : |f| \leq c\}. \quad (1.18)$$

Si E est un espace métrique et que \mathcal{E} est la famille des boréliens de E , (1.18) implique notamment que

$$\lim_{n \rightarrow \infty} \int_E f d\mu_n = \int_E f d\mu \quad \text{pour toute fonction } f \text{ continue et bornée,} \quad (1.19)$$

c'est-à-dire que $(\mu_n; n \geq 1)$ converge faiblement vers μ (voir par exemple Billingsley, 1968, page 7). La convergence en variation totale est donc un mode de convergence plus fort que la convergence faible.

Pour étudier le comportement limite d'une chaîne de Markov à valeurs dans un espace mesurable (E, \mathcal{E}) , avec noyau markovien P et mesure de probabilité initiale λ , nous considérerons la variation totale entre les mesures de probabilité $(\lambda P^n; n \geq 0)$ et une mesure de probabilité limite potentielle. Supposons qu'il existe une mesure de probabilité ν_λ sur (E, \mathcal{E}) , telle que $\|\lambda P^n - \nu_\lambda\| \rightarrow 0$ lorsque $n \rightarrow \infty$. Pour tout $A \in \mathcal{E}$, nous avons alors

$$\begin{aligned} \nu_\lambda(A) &= \lim_{n \rightarrow \infty} \lambda P^n(A) \\ &= \lim_{n \rightarrow \infty} \int_E P^n(x, A) \lambda(dx) \\ &= \lim_{n \rightarrow \infty} \int_E P^{n+1}(x, A) \lambda(dx) \\ &= \lim_{n \rightarrow \infty} \int_E P(x, A) \lambda P^n(dx) \\ &= \int_E P(x, A) \nu_\lambda(dx), \end{aligned}$$

puisque $P(\cdot, A)$ est une fonction mesurable et bornée. Il s'ensuit que si la suite $(\lambda P^n; n \geq 0)$ converge en variation totale vers ν_λ , cette dernière est nécessairement une loi stationnaire pour P . De plus, si P possède une loi stationnaire unique, pour toute mesure de probabilité λ telle que la suite $(\lambda P^n; n \geq 0)$ converge en variation totale, cette convergence aura nécessairement lieu vers la loi stationnaire unique. À l'opposé, si P ne possède aucune loi stationnaire, la convergence en variation totale de la suite $(\lambda P^n; n \geq 0)$ vers une mesure de probabilité ν_λ est impossible, quelle que soit la mesure de probabilité initiale λ .

1.4.3 Le théorème d'Orey

Supposons que P soit un noyau φ -irréductible défini sur un espace mesurable (E, \mathcal{E}) et que π est une mesure de probabilité stationnaire pour P . Alors, par les théorèmes 1.4 et 1.5, P est aussi récurrent et π -irréductible, et π est l'unique mesure stationnaire de P , à une constante multiplicative près. Si P est également apériodique et récurrent au sens de Harris, alors pour toute mesure de probabilité λ définie sur (E, \mathcal{E}) , la suite de mesures $(\lambda P^n; n \geq 0)$ converge en variation totale, lorsque $n \rightarrow \infty$. Puisque cette convergence ne peut avoir lieu que vers une loi stationnaire de P , et que P ne possède en fait qu'une loi stationnaire, il s'ensuit alors que cette convergence doit avoir lieu vers π . Le théorème suivant, dû à Orey (1959), résume ces faits.

Théorème 1.6. *Soit P un noyau markovien défini sur un espace mesurable (E, \mathcal{E}) . Supposons que P soit apériodique, récurrent au sens de Harris, et que π soit une mesure de probabilité stationnaire pour P . Alors, pour toute mesure de probabilité λ définie sur (E, \mathcal{E}) ,*

$$\lim_{n \rightarrow \infty} \sup_{A \in \mathcal{E}} |\lambda P^n(A) - \pi(A)| = 0. \quad (1.20)$$

Si, pour un certain $x \in E$, nous remplaçons λ par δ_x , la mesure de Dirac en x , (1.20) devient

$$\lim_{n \rightarrow \infty} \sup_{A \in \mathcal{E}} |P^n(x, A) - \pi(A)| = 0.$$

Nous pouvons donc conclure que si P et π satisfont les conditions du théorème 1.6, alors pour tout $x \in E$, nous avons $\|P^n(x, \cdot) - \pi(\cdot)\| \rightarrow 0$ lorsque $n \rightarrow \infty$.

Supposons maintenant que P ne soit pas récurrent au sens de Harris mais seulement récurrent. Dénotons par N l'ensemble des $x \in E$ qui empêchent P d'être récurrent au sens de Harris, c'est-à-dire l'ensemble des $x \in E$ pour lesquels il existe un $A \in \mathcal{E}$ avec $\pi(A) > 0$ tel que $P_x[T_A < \infty] < 1$. Par la définition de la récurrence, cet ensemble satisfait $\pi(N) = 0$. Selon Nummelin (1984, théorème 3.7), l'ensemble $H = N^c$ est absorbant, c'est-à-dire qu'il satisfait

$$P(x, H) = 1 \quad \text{pour tout } x \in H.$$

De plus, le noyau markovien $P_{[H]}$, défini sur $(H, \mathcal{E} \cap H)$ par

$$P_{[H]}(x, A) = P(x, A), \quad x \in H, A \in \mathcal{E} \cap H,$$

est récurrent au sens de Harris. Notons que le noyau $P_{[H]}$ s'appelle *la restriction de P à l'espace mesurable $(H, \mathcal{E} \cap H)$* . Toujours selon Nummelin (1984, page 8), le noyau $P_{[H]}$ satisfait, pour tous $n \geq 0$, $x \in H$ et $A \in \mathcal{E} \cap H$,

$$(P_{[H]})^n(x, A) = (P^n)_{[H]}(x, A) = P^n(x, A).$$

Par le théorème 1.6, il s'ensuit que pour toute mesure de probabilité λ définie sur $(H, \mathcal{E} \cap H)$,

$$\lim_{n \rightarrow \infty} \|\lambda(P_{[H]}^n(\cdot) - \pi(\cdot))\| = 0,$$

et donc que

$$\lim_{n \rightarrow \infty} \|\lambda P^n(x, \cdot) - \pi(\cdot)\| = 0$$

pour toute mesure de probabilité λ sur (E, \mathcal{E}) telle que $\lambda(H) = 1$. Nous avons donc:

Théorème 1.7. *Soit P un noyau markovien défini sur un espace mesurable (E, \mathcal{E}) . Supposons que P soit apériodique, récurrent, et que π soit une mesure de probabilité stationnaire pour P . Posons $H = \{x \in E : \mathbb{P}_x[T_A < \infty] = 1 \forall A \in \mathcal{E}^+\}$. Alors, pour toute mesure de probabilité λ définie sur (E, \mathcal{E}) telle que $\lambda(H) = 1$,*

$$\lim_{n \rightarrow \infty} \sup_{A \in \mathcal{E}} |\lambda P^n(A) - \pi(A)| = 0.$$

En particulier, ce résultat implique aussi que pour π -presque tout $x \in E$ (i.e. pour tout $x \in H$),

$$\lim_{n \rightarrow \infty} \sup_{A \in \mathcal{E}} |P^n(x, A) - \pi(A)| = 0.$$

Dans le cas où E est dénombrable et que $\mathcal{E} = \mathcal{P}(E)$, les théorèmes 1.6 et 1.7 renforcent le théorème classique de convergence de la section 1.1 de deux façons. Premièrement, ils établissent la convergence de $P^n(x, \{y\})$ vers $\pi(\{y\})$ sous des conditions d'irréductibilité et de récurrence plus faibles: en effet, si P est un noyau ψ -irréductible et apériodique sur (E, \mathcal{E}) où ψ est une mesure d'irréductibilité maximale pour P , nous avons dans le cas récurrent,

$$\lim_{n \rightarrow \infty} P^n(x, \{y\}) = \pi(\{y\}) \quad \text{pour tous } y \in E \text{ et } x \in E \text{ tel que } \psi(\{x\}) > 0, \quad (1.21)$$

et dans le cas récurrent au sens de Harris,

$$\lim_{n \rightarrow \infty} P^n(x, \{y\}) = \pi(\{y\}) \quad \text{pour tous } x, y \in E. \quad (1.22)$$

Le théorème de la section 1.1 exigeait que la mesure ψ accorde un poids positif à tous les sous-ensembles non vides dans \mathcal{E} , alors qu'ici, la mesure ψ peut accorder un poids nul à certains de ces sous-ensembles. Deuxièmement, ils établissent que la convergence dans (1.21) et (1.22) est uniforme sur E :

$$\lim_{n \rightarrow \infty} \sup_{y \in E} |P^n(x, \{y\}) - \pi(\{y\})| = 0,$$

pour tout $x \in E$ dans le cas récurrent au sens de Harris, et pour tout $x \in E$ tel que $\psi(\{x\}) > 0$ dans le cas récurrent.

Chapitre 2

Algorithmes de simulation markovienne

Trois algorithmes de simulation markovienne seront présentés ici: l'*échantillonneur de Gibbs*, l'*algorithme de Metropolis-Hastings* et l'*algorithme hit-and-run*. À partir d'une mesure de probabilité π définie sur \mathbb{R}^d , ces trois méthodes permettent de simuler une chaîne de Markov $(X_n; n \geq 0)$ dont le noyau markovien P possède, comme loi stationnaire, la mesure de probabilité π . Pour chacune de ces méthodes, l'algorithme à suivre pour générer la chaîne de Markov en question sera d'abord décrit. Ensuite, nous énoncerons des conditions sous lesquelles le noyau P est apériodique, π -irréductible et récurrent au sens de Harris. Le théorème d'Orey nous permettra alors de conclure que si la chaîne est simulée suffisamment longtemps, sa dernière valeur sera approximativement distribuée selon la loi π .

Tout au long de ce chapitre, π désignera une mesure de probabilité définie sur $(\mathbb{R}^d, \mathcal{B}^d)$, $d \geq 2$, où \mathcal{B}^d représente la σ -algèbre des boréliens de \mathbb{R}^d . Nous supposons que π est absolument continue par rapport à la mesure de Lebesgue sur $(\mathbb{R}^d, \mathcal{B}^d)$ et que f est une densité pour π . Bien que certains des algorithmes que nous verrons peuvent parfois se généraliser aux cas où π est une mesure de probabilité absolument continue par rapport à une autre mesure, comme par exemple la mesure de dénombrement, nous

nous concentrerons sur le cas où π est absolument continue par rapport à la mesure de Lebesgue. La mesure de probabilité π sera parfois appelée la *loi cible*, puisque notre but sera de simuler une variable aléatoire dont la loi est π .

Pour tout borélien A de \mathbb{R}^d , l'ensemble des boréliens de \mathbb{R}^d inclus dans A sera dénoté \mathcal{B}_A et l'ensemble $\{B \in \mathcal{B}_A : \pi(B) > 0\}$ sera alors identifié à l'aide de \mathcal{B}_A^+ . Si ν est une mesure de probabilité sur $(\mathbb{R}^d, \mathcal{B}^d)$, nous représenterons son support par $\text{Supp}(\nu)$; en d'autres termes,

$$\text{Supp}(\nu) = \{x \in \mathbb{R}^d : \nu(B(x, \varepsilon)) > 0 \forall \varepsilon > 0\},$$

où $B(x, \varepsilon) = \{y \in \mathbb{R}^d : \|y - x\| < \varepsilon\}$ est la boule ouverte de rayon ε centrée à x . Finalement, lorsqu'aucune mesure ne sera spécifiée, les expressions *pour presque tout* et *presque partout* feront référence à la mesure de Lebesgue sur $(\mathbb{R}^d, \mathcal{B}^d)$, laquelle sera notée μ .

2.1 L'échantillonneur de Gibbs

2.1.1 Origines de l'algorithme

L'échantillonneur de Gibbs a été initialement conçu par Geman et Geman (1984) dans un contexte de restauration bayésienne d'images dégradées. Dans cet article, l'ensemble $Z = \{(i, j) \in \mathbb{N}^2 : 1 \leq i, j \leq m\}$ représente les m^2 pixels d'une image en noir et blanc. À chaque pixel $(i, j) \in Z$ est associée une valeur $f^{(i,j)} \in \{1, \dots, k\}$ représentant sa couleur, c'est-à-dire son intensité de gris. Étant donnée une mesure de probabilité π sur $\{1, \dots, k\}^{m^2}$ (l'ensemble de toutes les images possibles), l'objectif de Geman et Geman (1984) est de parvenir à simuler une variable aléatoire (une "image" aléatoire) de loi π . Cependant, l'entier m étant habituellement très grand, la simulation directe d'une telle variable aléatoire est extrêmement difficile.

La stratégie que Geman et Geman (1984) proposent pour remédier à cette difficulté consiste à simuler une chaîne de Markov $(F_n; n \geq 0) = ((F_n^{(1,1)}, \dots, F_n^{(m,m)}); n \geq 0)$ (où

$F_n^{(i,j)}$, $(i, j) \in Z$, représente l'intensité de gris du pixel (i, j) au temps n) dont le noyau markovien possède π comme loi stationnaire. Pour ce faire, leur algorithme ne requiert que la connaissance, pour chaque $(i, j) \in Z$, d'une version de la loi conditionnelle de $F^{(i,j)}$ étant donné $(F^{(k,l)}; (k, l) \in N_{(i,j)})$, disons $\pi_{(i,j)|N_{(i,j)}}$. Ici, $F = (F^{(i,j)}; (i, j) \in Z)$ est une variable aléatoire hypothétique de loi π et pour tout $(i, j) \in Z$, $N_{(i,j)}$ est un sous-ensemble de Z tel que $(i, j) \notin N_{(i,j)}$.

À l'aide du théorème sur la convergence des chaînes de Markov à espace d'états fini, ils montrent que sous certaines hypothèses portant sur π et sur la famille de sous-ensembles $\mathcal{N} = (N_{(i,j)}; (i, j) \in Z)$, la loi de la $n^{\text{ième}}$ image $(F_n^{(1,1)}, \dots, F_n^{(m,m)})$ converge vers π lorsque $n \rightarrow \infty$, en ce sens que

$$\lim_{n \rightarrow \infty} \mathbb{P}[(F_n^{(1,1)}, \dots, F_n^{(m,m)}) = f] = \pi(\{f\}), \quad f \in \{1, \dots, k\}^{m^2}.$$

Lorsque les hypothèses en question sont satisfaites, leur algorithme permet donc de simuler une variable aléatoire F de loi π .

Les hypothèses faites par les auteurs ont notamment pour but d'assurer que les versions des lois conditionnelles mentionnées plus haut déterminent de façon unique la loi π , en ce sens que la fonction $\pi_{(i,j)|N_{(i,j)}}$ est une version de la loi conditionnelle de $F^{(i,j)}$ étant donné $(F^{(k,l)}; (k, l) \in N_{(i,j)})$ pour tout $(i, j) \in Z$ si et seulement si la loi de F est π . Ces hypothèses peuvent être résumées ainsi: la première suppose que la famille de sous-ensembles \mathcal{N} est une structure de voisinage pour Z , c'est-à-dire qu'elle satisfait les trois propriétés suivantes:

- a) $N_{(i,j)} \subset Z$ pour tout $(i, j) \in Z$;
- b) $(i, j) \notin N_{(i,j)}$ pour tout $(i, j) \in Z$;
- c) $(i, j) \in N_{(k,l)}$ si et seulement si $(k, l) \in N_{(i,j)}$, pour tous $(i, j), (k, l) \in N_{(i,j)}$.

La deuxième hypothèse suppose que la mesure de probabilité π est de la forme

$$\pi(f) = \frac{1}{Z} e^{-U(f)/T},$$

où Z et T sont des constantes positives et où U est de la forme

$$U(f) = \sum_{C \in \mathcal{C}} V_C(f).$$

Ici, \mathcal{C} représente la famille des *cliques* associée à \mathcal{N} , c'est-à-dire la famille des sous-ensembles C de Z qui sont tels que si $(i, j) \in C$ et $(k, l) \in C$, alors $(i, j) \in N_{(k,l)}$ et $(k, l) \in N_{(i,j)}$. Pour tout $C \in \mathcal{C}$, V_C est une fonction non négative définie sur Z telle que $V_C(f)$ ne dépend que des valeurs de f aux sites appartenant à la clique C . La terminologie *échantillonneur de Gibbs* provient du fait qu'une telle loi π s'appelle une mesure de Gibbs, en l'honneur du mathématicien J. W. Gibbs (1839-1903). Geman et Geman (1984) montrent que π est une mesure de Gibbs par rapport au système de voisinage \mathcal{N} si et seulement si π satisfait la propriété suivante: si $F = (F^{(i,j)}; (i, j) \in Z)$ est distribuée selon π , alors pour tout $(i, j) \in Z$, la loi conditionnelle de $F^{(i,j)}$ étant donnée $(F^{(k,l)}; (k, l) \in N_{(i,j)})$ est égale à la loi conditionnelle de $F^{(i,j)}$ étant donnée $(F^{(k,l)}; (k, l) \in Z - \{(i, j)\})$.

L'algorithme fut adapté plus tard par Gelfand et Smith (1990) pour devenir un algorithme de simulation pouvant simuler une grande variété de lois de probabilité et pouvant ainsi s'appliquer à plusieurs problèmes conventionnels en statistique ou en mathématique. Ces auteurs supposent que π est une loi sur $(\mathbb{R}^d, \mathcal{B}^d)$ absolument continue par rapport à la mesure de Lebesgue (ou une loi discrète sur \mathcal{Y}^d , où \mathcal{Y} est un sous-ensemble fini ou dénombrable de \mathbb{R}). Leur algorithme permet donc de générer une chaîne de Markov à valeurs dans $(\mathbb{R}^d, \mathcal{B}^d)$, avec loi stationnaire π . Quoique l'idée de base de l'algorithme remonte à Metropolis et coll. (1953) et à Hastings (1970), l'article de Gelfand et Smith est vraiment celui qui a suscité de l'intérêt pour le sujet dans le domaine de la statistique. Depuis, l'échantillonneur de Gibbs a connu une énorme popularité, et les articles à son sujet se sont multipliés dans la littérature. Par exemple, plusieurs auteurs ont fourni des conditions sous lesquelles la loi de X_n converge en variation totale vers la loi π ; citons, entre autres, Chan (1993), Smith et Roberts (1993), Roberts et Polson (1994), Roberts et Smith (1994) et Tierney (1994). D'autres ont étudié la vitesse de cette convergence, comme par exemple Schervish et Carlin (1992), Athreya, Doss et Sethuraman (1996), Roberts et Rosenthal (1998b) et Bélisle (1998a). Enfin, beaucoup d'autres ont suggéré de nouvelles applications à l'échantillonneur de Gibbs, plusieurs d'entre elles étant reliées à l'estimation de paramètres dans des modèles statistiques bayésiens, comme Gelfand et Smith (1990), Tanner (1991), Casella et George (1992), Smith et Roberts (1993), Tierney (1994) et Gilks, Richardson et Spiegelhalter (1996).

Notons finalement que l'algorithme de Gelfand et Smith (1990) continue d'être appelé l'*échantillonneur de Gibbs*, malgré le fait qu'une loi cible π à laquelle il peut être appliqué n'est pas nécessairement une loi de Gibbs. Pour cette raison, certains reprochent au nom *échantillonneur de Gibbs* de ne pas être tout à fait approprié.

2.1.2 Description de l'algorithme

Considérons maintenant la densité f de la loi cible π qui nous a été donnée à la page 36. Pour tout $i \in \{1, 2, \dots, d\}$ et pour tout $t = (t^1, \dots, t^{i-1}, t^{i+1}, \dots, t^d) \in \mathbb{R}^{d-1}$, dénotons par $f_i(\cdot | t)$ la densité unidimensionnelle définie par

$$f_i(v|t) = \frac{f(t^1, \dots, t^{i-1}, v, t^{i+1}, \dots, t^d)}{\int_{\mathbb{R}} f(t^1, \dots, t^{i-1}, u, t^{i+1}, \dots, t^d) du}, \quad v \in \mathbb{R}. \quad (2.1)$$

Ainsi, si $X = (X^1, \dots, X^d)$ est une variable aléatoire distribuée selon la loi π , alors pour tout $i \in \{1, \dots, d\}$, la fonction

$$f_i(v|x^1, \dots, x^{i-1}, x^{i+1}, \dots, x^d)$$

est une version de la densité conditionnelle de X^i étant donné que

$$(X^1, \dots, X^{i-1}, X^{i+1}, \dots, X^d) = (x^1, \dots, x^{i-1}, x^{i+1}, \dots, x^d).$$

Par abus de langage, les fonctions $f_i(\cdot | \cdot)$, $i \in \{1, \dots, d\}$, seront souvent appelées des densités conditionnelles ou même des densités conditionnelles de π . Ceci évitera de toujours avoir à supposer qu'il y a une variable aléatoire X de loi π avant d'utiliser le terme *densité conditionnelle*.

Pour simuler une chaîne de Markov avec loi stationnaire π , l'échantillonneur de Gibbs procède de la façon suivante. Tout d'abord, l'état initial $x_0 = (x_0^1, \dots, x_0^d)$ de la chaîne est choisi arbitrairement dans \mathbb{R}^d . Ensuite, pour tout $n \geq 1$, un point $x_n = (x_n^1, \dots, x_n^d)$ de \mathbb{R}^d est généré composante par composante, en utilisant les densités conditionnelles $f_i(\cdot | \cdot)$, $i \in \{1, \dots, d\}$, et le point précédent $x_{n-1} = (x_{n-1}^1, \dots, x_{n-1}^d)$, selon l'algorithme suivant:

- Étape 1:** générer x_n^1 à partir de la densité $f_1(\cdot | x_{n-1}^2, \dots, x_{n-1}^d)$;
Étape 2: générer x_n^2 à partir de la densité $f_2(\cdot | x_n^1, x_{n-1}^3, \dots, x_{n-1}^d)$;
Étape 3: générer x_n^3 à partir de la densité $f_3(\cdot | x_n^1, x_n^2, x_{n-1}^4, \dots, x_{n-1}^d)$;
 \vdots
Étape d: générer x_n^d à partir de la densité $f_d(\cdot | x_n^1, x_n^2, \dots, x_n^{d-1})$.

Chaque étape de l'algorithme effectue en quelque sorte une mise à jour du vecteur obtenu à l'étape précédente. Par exemple, si l'état initial choisi est $x_0 = (x_0^1, \dots, x_0^d)$, la première étape de l'algorithme remplace la première composante de x_0 par un point distribué selon la densité $f_1(\cdot | x_0^2, \dots, x_0^d)$ pour ainsi produire le vecteur $(x_1^1, x_0^2, \dots, x_0^d)$. La deuxième composante de ce dernier est ensuite remplacée dans la deuxième étape par un point distribué selon la densité $f_2(\cdot | x_1^1, x_0^3, \dots, x_0^d)$ pour produire alors le vecteur $(x_1^1, x_1^2, x_0^3, \dots, x_0^d)$, et ainsi de suite. La $d^{\text{ème}}$ étape de l'algorithme complète la transition de x_0 à x_1 . L'algorithme est ensuite répété pour produire x_2, x_3, \dots . Quel que soit $n \geq 1$, la transition de x_{n-1} à x_n sera appelée une *itération* de l'algorithme, alors que chaque transition intermédiaire (par exemple, la transition de (x_0^1, \dots, x_0^d) à $(x_1^1, x_0^2, \dots, x_0^d)$) sera appelée une *mise à jour*.

La suite de points $(x_n; n \geq 0)$ produite par l'algorithme est manifestement la réalisation d'une chaîne de Markov (homogène dans le temps) $(X_n; n \geq 0)$ à valeurs dans $(\mathbb{R}^d, \mathcal{B}^d)$. Dans les sections suivantes, nous verrons que:

- 1) la mesure de probabilité π est une loi stationnaire pour le noyau markovien de $(X_n; n \geq 0)$;
- 2) sous certaines conditions de régularité, la π -irréductibilité du noyau markovien de $(X_n; n \geq 0)$ est suffisante pour conclure que pour tout point de départ x_0 , la loi de X_n converge en variation totale vers la mesure π , lorsque $n \rightarrow \infty$, et donc que l'échantillonneur de Gibbs permet dans ce cas de simuler une variable aléatoire de loi π .

Remarque 2.1. Le point de départ x_0 peut évidemment être choisi aléatoirement à l'aide d'une loi λ définie sur $(\mathbb{R}^d, \mathcal{B}^d)$.

Remarque 2.2. Pour tout élément $x = (x^1, \dots, x^d)$ de \mathbb{R}^d , nous désignerons parfois par $x^{(-i)}$, $i \in \{1, \dots, d\}$, le vecteur de \mathbb{R}^{d-1} égal à x sans sa $i^{\text{ième}}$ composante; en d'autres termes, pour tout $x \in \mathbb{R}^d$ et pour tout $i \in \{1, \dots, d\}$,

$$x^{(-i)} = (x^1, \dots, x^{i-1}, x^{i+1}, \dots, x^d).$$

Cette notation inhabituelle permettra d'alléger le texte.

2.1.3 Le choix de la densité f

À la section 2.1.2, nous avons supposé implicitement que pour tout $x \in \mathbb{R}^d$ et pour tout $i \in \{1, \dots, d\}$, la condition

$$0 < \int_{\mathbb{R}} f(x^1, \dots, x^{i-1}, u, x^{i+1}, \dots, x^d) du < \infty \quad (2.2)$$

était satisfaite. En effet, cette condition est nécessaire pour que l'équation (2.1) ait un sens, c'est-à-dire pour que les fonctions $f_i(\cdot | x^{(-i)})$, $i \in \{1, \dots, d\}$, que l'algorithme utilise soient des densités bien définies sur \mathbb{R} pour tout $x \in \mathbb{R}^d$. En général, cette condition n'est évidemment pas satisfaite. Cependant, cette situation n'a rien d'alarmant. En effet, en vertu du théorème suivant, pour π -presque tout point de départ $x_0 \in \mathbb{R}^d$, l'algorithme génère, avec probabilité 1, une suite de points pour chacun desquels la condition (2.2) est satisfaite pour tout $i \in \{1, \dots, d\}$.

Théorème 2.1. *Soit π une mesure de probabilité sur $(\mathbb{R}^d, \mathcal{B}^d)$, absolument continue par rapport à la mesure de Lebesgue. Soit f une densité pour π . Posons, pour tout $i \in \{1, \dots, d\}$,*

$$A_f^i = \{x \in \mathbb{R}^d : 0 < \int_{\mathbb{R}} f(x^1, \dots, x^{i-1}, u, x^{i+1}, \dots, x^d) du < \infty\}$$

et

$$A_f = \bigcap_{i=1}^d A_f^i.$$

Alors

$$\pi(A_f) = 1.$$

Démonstration. Pour montrer que $\pi(A_f) = 1$, il suffit de montrer que $\pi(A_f^i) = 1$ pour tout $i \in \{1, \dots, d\}$. Fixons donc $i \in \{1, \dots, d\}$. Dénotons par $\pi^{(-i)}$ la projection de la mesure de probabilité π sur l'ensemble $\{(x^1, \dots, x^d) \in \mathbb{R}^d : x^i = 0\}$, c'est-à-dire la mesure de probabilité définie sur $(\mathbb{R}^{d-1}, \mathcal{B}^{d-1})$ par

$$\pi^{(-i)}(A_1 \times \dots \times A_{i-1} \times A_{i+1} \times \dots \times A_d) = \pi(A_1 \times \dots \times A_{i-1} \times \mathbb{R} \times A_{i+1} \times \dots \times A_d),$$

où $A_1, \dots, A_{i-1}, A_{i+1}, \dots, A_d \in \mathcal{B}$. Notons que si π est la loi d'un vecteur aléatoire (X^1, \dots, X^d) à valeurs dans \mathbb{R}^d , alors $\pi^{(-i)}$ est simplement la loi marginale du vecteur $(X^1, \dots, X^{i-1}, X^{i+1}, \dots, X^d)$. Posons

$$\Lambda_0 = \{(t^1, \dots, t^{i-1}, t^{i+1}, \dots, t^d) \in \mathbb{R}^{d-1} : \int_{\mathbb{R}} f(t^1, \dots, t^{i-1}, u, t^{i+1}, \dots, t^d) du = 0\}$$

et

$$\Lambda_\infty = \{(t^1, \dots, t^{i-1}, t^{i+1}, \dots, t^d) \in \mathbb{R}^{d-1} : \int_{\mathbb{R}} f(t^1, \dots, t^{i-1}, u, t^{i+1}, \dots, t^d) du = \infty\}, \quad (2.3)$$

et remarquons que si pour un certain $x \in \mathbb{R}^d$,

$$\int_{\mathbb{R}} f(x^1, \dots, x^{i-1}, u, x^{i+1}, \dots, x^d) du = \alpha, \quad \alpha \geq 0,$$

alors

$$\int_{\mathbb{R}} f(y^1, \dots, y^{i-1}, u, y^{i+1}, \dots, y^d) du = \alpha$$

pour tout $y \in \mathbb{R}^d$ tel que $y^i = x^i$. Par conséquent, pour montrer que $\pi(A_f^i) = 1$, il suffit de montrer que $\pi^{(-i)}(\Lambda_0) = \pi^{(-i)}(\Lambda_\infty) = 0$.

Par le théorème de Fubini (voir par exemple Bartle, 1995, page 119), la fonction

$$\int_{\mathbb{R}} f(t^1, \dots, t^{i-1}, u, t^{i+1}, \dots, t^d) du$$

est mesurable en $t = (t^1, \dots, t^{i-1}, t^{i+1}, \dots, t^d)$ et donc Λ_0 et $\Lambda_\infty \in \mathcal{B}^{d-1}$. Puisque pour tout $t \in \Lambda_0$,

$$\int_{\mathbb{R}} f(t^1, \dots, t^{i-1}, u, t^{i+1}, \dots, t^d) du = 0,$$

nous avons

$$\begin{aligned} \pi^{(-i)}(\Lambda_0) &= \int_{\Lambda_0} \left[\int_{\mathbb{R}} f(t^1, \dots, t^{i-1}, u, t^{i+1}, \dots, t^d) du \right] dt \\ &= 0. \end{aligned}$$

Maintenant, pour l'ensemble Λ_∞ , nous avons aussi

$$\pi^{(-i)}(\Lambda_\infty) = \int_{\Lambda_\infty} \left[\int_{\mathbb{R}} f(t^1, \dots, t^{i-1}, u, t^{i+1}, \dots, t^d) du \right] dt.$$

Puisque $\int_{\mathbb{R}} f(t^1, \dots, t^{i-1}, u, t^{i+1}, \dots, t^d) du = \infty$ pour tout $t \in \Lambda_\infty$ et que l'ensemble Λ_∞ doit satisfaire $\pi^{(-i)}(\Lambda_\infty) \leq 1$, il faut nécessairement que $\mu^{d-1}(\Lambda_\infty) = 0$, où μ^{d-1} représente la mesure de Lebesgue sur $(\mathbb{R}^{d-1}, \mathcal{B}^{d-1})$. Puisque $\pi^{(-i)}$ est absolument continue par rapport à μ^{d-1} , il s'ensuit que $\pi^{(-i)}(\Lambda_\infty) = 0$. ■

Pour chaque $i \in \{1, \dots, d\}$, l'ensemble A_f^i est l'ensemble des $x \in \mathbb{R}^d$ pour lesquels la fonction $f_i(\cdot | x^{(-i)})$ est une densité sur \mathbb{R} . Considérons aussi, pour tout $i \in \{1, \dots, d\}$, l'ensemble

$$B_f^i = \{x \in A_f^i : \int_{\Gamma_x^i} f(x^1, \dots, x^{i-1}, u, x^{i+1}, \dots, x^d) du = 0\},$$

où

$$\Gamma_x^i = \{u \in \mathbb{R} : (x^1, \dots, x^{i-1}, u, x^{i+1}, \dots, x^d) \notin \text{Supp}(\pi)\}.$$

L'ensemble B_f^i est l'ensemble des $x \in A_f^i$ pour lesquels le point

$$(x^1, \dots, x^{i-1}, V, x^{i+1}, \dots, x^d),$$

où V est une variable aléatoire à valeurs dans $(\mathbb{R}, \mathcal{B})$ avec densité $f_i(\cdot | x^{(-i)})$, est inclus, avec probabilité 1, dans le support de π . Puisque $\pi(A_f^i) = 1$ et que $f(x) = 0$ pour presque tout x à l'extérieur de $\text{Supp}(\pi)$, nous avons $\pi(B_f^i) = 1$ pour tout $i \in \{1, \dots, d\}$ et donc $\pi(B_f) = 1$, où $B_f = \bigcap_{i=1}^d B_f^i$.

Posons maintenant $C_f = A_f \cap B_f \cap \text{Supp}(\pi)$. En vertu du théorème précédent et du fait que $\pi(B_f) = 1$, nous avons $\pi(C_f) = 1$. Nous pouvons ainsi conclure que si le point de départ x_0 de l'échantillonneur de Gibbs est choisi dans C_f , alors chaque mise à jour de l'algorithme nous amène, avec probabilité 1, à un autre point de C_f . En d'autres termes, avec probabilité 1, nous évitons tout problème avec les points $x \in \mathbb{R}^d$ pour lesquels l'une ou l'autre des fonctions $f_i(\cdot | x^{(-i)})$, $i \in \{1, \dots, d\}$, ne sont pas des densités bien définies, et ceci, sans que la chaîne ne sorte du support de π .

En pratique, le fait de choisir le point de départ dans l'ensemble C_f ne règle toutefois pas tous les problèmes. En voici un exemple:

Exemple 2.1. Considérons A , le carré unité fermé de \mathbb{R}^2 , c'est-à-dire $A = \{x \in \mathbb{R}^2 : 0 \leq x^1, x^2 \leq 1\}$, et π , la loi uniforme sur A . Comme densité pour π , choisissons la densité f définie par

$$f(x^1, x^2) = \begin{cases} 1 & \text{si } (x^1, x^2) \in A, x^1 \notin \mathbb{Q}, x^2 \notin \mathbb{Q}, \\ 0 & \text{sinon.} \end{cases}$$

Ainsi, la densité f est positive partout sur l'ensemble A , sauf aux points dont l'une ou l'autre des coordonnées est rationnelle. Pour éviter tout problème avec les points $(x^1, x^2) \in A$ où l'une ou l'autre des fonctions $f_1(\cdot|x^2)$ et $f_2(\cdot|x^1)$ ne sont pas des densités bien définies sur \mathbb{R} , il suffit donc de choisir un point de départ x_0 à coordonnées irrationnelles à l'intérieur du carré A . Ainsi, avec probabilité égale à 1, la trajectoire de la chaîne de Markov simulée par l'échantillonneur de Gibbs ne sortira pas de A , et elle ne passera jamais par un point $(x^1, x^2) \in \mathbb{R}^2$ où l'une ou l'autre des fonctions $f_1(\cdot|x^2)$ et $f_2(\cdot|x^1)$ n'est pas une densité bien définie sur \mathbb{R} . En théorie, une densité inusuelle comme celle du présent exemple ne causera aucun problème. En pratique, par contre, la situation est très différente. En effet, si nous utilisons un ordinateur pour simuler la chaîne de Markov, il est alors impossible de choisir un point de départ à coordonnées irrationnelles, puisqu'un ordinateur ne connaît que les nombres réels avec un nombre fini de décimales! Dans cet exemple, il est donc préférable de choisir une version plus appropriée de la densité de π , comme par exemple celle qui est positive partout sur l'ensemble A . \square

Pour éviter des complications inutiles, nous supposons dorénavant que la mesure de probabilité π satisfait la condition suivante:

Condition G: *Il existe une version f de la densité de π qui satisfait:*

- 1) l'ensemble $\mathcal{X} = \{x \in \mathbb{R}^d : f(x) > 0\}$ est ouvert;
- 2) pour tout $i \in \{1, \dots, d\}$ et pour tout $x \in \mathcal{X}$,

$$0 < \int_{\mathbb{R}} f(x^1, \dots, x^{i-1}, u, x^{i+1}, \dots, x^d) du < \infty.$$

La plupart des mesures de probabilité sur $(\mathbb{R}^d, \mathcal{B}^d)$ absolument continues par rapport à la mesure de Lebesgue que l'on retrouve en pratique satisfont la condition G. Celle-ci n'est donc pas trop restrictive.

À partir de maintenant, nous supposerons toujours que la loi cible π satisfait la condition G (et nous omettrons parfois de le mentionner); ceci sous-entend aussi que l'algorithme sera toujours défini à l'aide d'une version de la densité de π satisfaisant les points 1 et 2 de la condition G, et que le point de départ x_0 sera toujours choisi dans \mathcal{X} . La chaîne de Markov $(X_n; n \geq 0)$ sera vue comme une chaîne de Markov à valeurs dans l'espace mesurable $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$.

Remarque 2.3. Même avec la condition G satisfaite, il peut y avoir d'autres types de problèmes causés par le fait que les ordinateurs ne peuvent générer que des nombres rationnels (que ce soit dans le contexte de l'échantillonneur de Gibbs ou de tout autre algorithme de simulation markovienne); pour une discussion à ce sujet, voir Roberts et Rosenthal (1998a), et Roberts, Rosenthal et Schwartz (1998).

Remarque 2.4. Si la condition G n'est pas satisfaite par la loi cible π et que le point de départ x_0 n'est pas choisi dans l'ensemble C_f , une solution possible pour contourner le fait qu'une ou plusieurs des fonctions $f_i(\cdot | x_0^{(-i)})$, $i \in \{1, \dots, d\}$, ne sont pas des densités sur \mathbb{R}^d est de redéfinir ces fonctions par des densités arbitraires sur \mathbb{R} , en s'assurant qu'elles font en sorte qu'après une itération de l'algorithme, la chaîne revienne dans l'ensemble C_f avec probabilité 1. Toutefois, nous ne considérerons pas cette alternative ici.

2.1.4 Noyau markovien de l'échantillonneur de Gibbs

Considérons la suite des points obtenus après chaque mise à jour, c'est-à-dire la suite des vecteurs

$$(X_0^1, X_0^2, \dots, X_0^d), (X_1^1, X_0^2, \dots, X_0^d), (X_1^1, X_1^2, X_0^3, \dots, X_0^d), \dots \quad (2.4)$$

Cette suite définit aussi une chaîne de Markov sur $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$. Cependant, contrairement

à $(X_n; n \geq 0)$, cette chaîne n'est pas homogène dans le temps puisque la loi d'un vecteur particulier dépend aussi de l'indice de la composante qui y est mise à jour.

Supposons qu'après un certain nombre de mises à jour, cette chaîne soit à l'état $x = (x^1, \dots, x^d)$, et que ce soit maintenant au tour de la $i^{\text{ème}}$ composante d'être remplacée. Nous sommes donc au point

$$(X_{n+1}^1, \dots, X_{n+1}^{i-1}, X_n^i, \dots, X_n^d) = (x^1, \dots, x^d)$$

pour un certain $n \geq 1$. Il faut maintenant choisir un point $v \in \mathbb{R}$ selon la densité $f_i(\cdot | x^{(-i)})$ et passer au point

$$(X_{n+1}^1, \dots, X_{n+1}^i, X_n^{i+1}, \dots, X_n^d) = (x^1, \dots, x^{i-1}, v, x^{i+1}, \dots, x^d).$$

Pour tout borélien A de $\mathcal{B}_{\mathcal{X}}$, nous avons ainsi

$$\begin{aligned} \mathbb{P}[(X_{n+1}^1, \dots, X_{n+1}^i, X_n^{i+1}, \dots, X_n^d) \in A | (X_{n+1}^1, \dots, X_{n+1}^{i-1}, X_n^i, \dots, X_n^d) = x] \\ = \int_{\mathbb{R}} \mathbf{1}_A(x^1, \dots, x^{i-1}, v, x^{i+1}, \dots, x^d) f_i(v | x^{(-i)}) dv. \end{aligned} \quad (2.5)$$

Notons que pour tout $A \in \mathcal{B}_{\mathcal{X}}$, la fonction $\mathbf{1}_A(x)$ est mesurable en chacune des composantes de x . Pour tout $x \in \mathcal{X}$, elle est aussi une mesure de probabilité sur $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$: la mesure de Dirac en x . Puisque pour tout $i \in \{1, \dots, d\}$, les fonctions $f_i(\cdot | x^{(-i)})$, $x \in \mathcal{X}$, et $f_i(v | \cdot)$, $v \in \mathbb{R}$, sont mesurables, et que $f_i(\cdot | x^{(-i)})$ est une densité sur \mathbb{R} pour tout $x \in \mathcal{X}$, la fonction dans le membre de droite de (2.5) est un noyau markovien bien défini sur $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$.

Pour tout $i \in \{1, \dots, d\}$, posons maintenant

$$P_{\langle \pi, i \rangle}(x, A) = \int_{\mathbb{R}} \mathbf{1}_A(x^1, \dots, x^{i-1}, v, x^{i+1}, \dots, x^d) f_i(v | x^{(-i)}) dv, \quad x \in \mathcal{X}, A \in \mathcal{B}_{\mathcal{X}}. \quad (2.6)$$

Alors, l'algorithme de l'échantillonneur de Gibbs peut aussi être exprimé de la façon suivante: à l'aide d'un point de départ y_0 choisi arbitrairement dans \mathcal{X} , générer un point y_1 à l'aide de la mesure de probabilité $P_{\langle \pi, 1 \rangle}(y_0, \cdot)$, générer ensuite un point y_2 à l'aide de $P_{\langle \pi, 2 \rangle}(y_1, \cdot)$, et ainsi de suite, jusqu'à y_d . Ensuite, réutiliser le cycle $\{P_{\langle \pi, 1 \rangle}, \dots, P_{\langle \pi, d \rangle}\}$ pour générer, à partir du point y_d , les points y_{d+1}, \dots, y_{2d} , etc.

Revenons maintenant à la chaîne originale $(X_n; n \geq 0)$. Son noyau markovien P_π est facile à identifier: il est simplement le produit des noyaux $P_{\langle \pi, 1 \rangle}, \dots, P_{\langle \pi, d \rangle}$. En d'autres termes, pour tout $x \in \mathcal{X}$ et pour tout $A \in \mathcal{B}_\mathcal{X}$,

$$P_\pi(x, A) = P_{\langle \pi, 1 \rangle} \cdots P_{\langle \pi, d \rangle}(x, A). \quad (2.7)$$

À l'aide des équations (2.6) et (2.7), nous pouvons aussi écrire le noyau markovien P_π sous la forme suivante:

$$P_\pi(x, A) = \int_A p_\pi(x, y) dy, \quad x \in \mathcal{X}, A \in \mathcal{B}_\mathcal{X}, \quad (2.8)$$

où

$$p_\pi(x, y) = \prod_{i=1}^d f_i(y^i | y^1, \dots, y^{i-1}, x^{i+1}, \dots, x^d), \quad x, y \in \mathcal{X}. \quad (2.9)$$

Notons que pour tout $x \in \mathcal{X}$, la fonction $p_\pi(x, \cdot)$ est une densité sur \mathcal{X} . Ceci implique que les mesures de probabilité $(P_\pi(x, \cdot); x \in \mathcal{X})$ sont toutes absolument continues par rapport à la mesure de Lebesgue. En fait, il est facile de montrer que ce résultat implique que les mesures de probabilité $(P_\pi^n(x, \cdot); x \in \mathcal{X}, n \geq 1)$ sont toutes absolument continues par rapport à la mesure de Lebesgue.

La stationnarité de π pour P_π est la première condition que doit satisfaire le noyau P_π pour que la loi de X_n puisse converger en variation totale vers π . En effet, nous avons vu à la section 1.4.2 que si la loi de X_n converge en variation totale, ce ne peut être que vers une mesure de probabilité stationnaire pour P_π . Le prochain théorème nous montre que la mesure de probabilité π est stationnaire pour chacun des noyaux $P_{\langle \pi, 1 \rangle}, \dots, P_{\langle \pi, d \rangle}$. Nous pourrons ensuite conclure que π est aussi stationnaire pour P_π .

Théorème 2.2. *Pour tout $i \in \{1, \dots, d\}$, la mesure de probabilité π est stationnaire pour le noyau markovien $P_{\langle \pi, i \rangle}$.*

Démonstration. Fixons $i \in \{1, \dots, d\}$ et supposons que $f^{(-i)}$ soit une densité pour la mesure de probabilité $\pi^{(-i)}$ (que nous avons définie dans la démonstration du théorème 2.1). Pour tout $A \in \mathcal{B}_\mathcal{X}$, nous avons alors

$$\begin{aligned}
& \int_{\mathcal{X}} P_{\langle \pi, i \rangle}(x, A) \pi(dx) \\
&= \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}} \mathbf{1}_A(x^1, \dots, x^{i-1}, u, x^{i+1}, \dots, x^d) f_i(u|x^{(-i)}) du \right] \pi(dx) \\
&= \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}} \mathbf{1}_A(x^1, \dots, x^{i-1}, u, x^{i+1}, \dots, x^d) f_i(u|x^{(-i)}) du \right] f(x) dx \\
&= \int_{\mathbb{R}^{d-1}} \left[\int_{\mathbb{R}} \left[\int_{\mathbb{R}} \mathbf{1}_A(x^1, \dots, x^{i-1}, u, x^{i+1}, \dots, x^d) f_i(u|x^{(-i)}) f(x) dx^i \right] du \right] dx^{(-i)} \\
&\quad (\text{par le théorème de Fubini}) \\
&= \int_{\mathbb{R}^{d-1}} \left[\int_{\mathbb{R}} \mathbf{1}_A(x^1, \dots, x^{i-1}, u, x^{i+1}, \dots, x^d) f_i(u|x^{(-i)}) f^{(-i)}(x^{(-i)}) du \right] dx^{(-i)} \\
&= \int_{\mathbb{R}^d} \mathbf{1}_A(x) f(x) dx \quad (\text{une fois de plus, par le théorème de Fubini}) \\
&= \int_A f(x) dx \\
&= \pi(A).
\end{aligned}$$

La mesure de probabilité π est bien stationnaire pour le noyau markovien $P_{\langle \pi, i \rangle}$. ■

En général, si P_1 et P_2 sont deux noyaux markoviens définis sur un espace mesurable (E, \mathcal{E}) , et que π est une mesure de probabilité stationnaire pour chacun des noyaux P_1 et P_2 , alors π est aussi stationnaire pour le noyau produit $P_1 P_2$. En effet, si $A \in \mathcal{E}$, alors

$$\begin{aligned}
\pi P_1 P_2(A) &= \int_E P_1 P_2(x, A) \pi(dx) \\
&= \int_E \left[\int_E P_2(y, A) P_1(x, dy) \right] \pi(dx) \\
&= \int_E P_2(y, A) \left[\int_E P_1(x, dy) \pi(dx) \right] \\
&= \int_E P_2(y, A) \pi(dy) \\
&= \pi(A).
\end{aligned}$$

Dans le contexte de l'échantillonneur de Gibbs, ce résultat nous permet de conclure que la mesure de probabilité π est stationnaire pour le noyau markovien P_π .

2.1.5 Convergence

Le simple fait que la mesure de probabilité π soit stationnaire pour le noyau P_π n'implique pas que la suite de mesures $(P_\pi^n(x, \cdot); n \geq 1)$ convergera en variation totale vers π pour tout $x \in \mathcal{X}$, ou même pour un seul $x \in \mathcal{X}$. Par exemple, la loi π est stationnaire pour chacun des noyaux $P_{\langle \pi, 1 \rangle}, \dots, P_{\langle \pi, d \rangle}$. Pourtant, quels que soient $n \geq 1$ et $x \in \mathcal{X}$, la mesure $P_{\langle \pi, i \rangle}^n(x, \cdot)$, $i \in \{1, \dots, d\}$, accorde à l'ensemble $\{y \in \mathcal{X} : y^j = x^j, j \neq i\}$ une probabilité égale à 1. Elle est donc singulière par rapport à la mesure de Lebesgue, et pour tout $x \in \mathcal{X}$, la suite de mesures de probabilité $(P_{\langle \pi, i \rangle}^n(x, \cdot); n \geq 1)$ ne convergera certainement pas en variation totale vers π .

Dans cette section, nous présentons des conditions nécessaires et suffisantes qui assurent à la fois la π -irréductibilité, l'apériodicité et la récurrence au sens de Harris du noyau P_π . À l'aide du théorème d'Orey, nous pourrions ensuite conclure que sous ces conditions, l'échantillonneur de Gibbs génère une chaîne de Markov $(X_n; n \geq 0)$ dont la loi de X_n converge en variation totale vers π , peu importe l'état initial $x_0 \in \mathcal{X}$ choisi. Rappelons qu'il n'y a aucune perte de généralité à choisir π comme mesure d'irréductibilité potentielle pour P_π , puisque si P_π est φ -irréductible pour une certaine mesure φ , alors P_π est automatiquement π -irréductible (voir théorème 1.4). Ceci n'élimine pas du tout la possibilité d'utiliser, en pratique, une autre mesure σ -finie pour vérifier l'irréductibilité de P_π : n'importe quelle mesure absolument continue par rapport à la mesure de Lebesgue sur $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ peut faire l'affaire. Tout au long de cette section, nous supposerons que la mesure de probabilité π satisfait la condition G. Nous débutons par deux résultats simples qui nous seront très utiles par la suite.

Proposition 2.1. *Soit R un rectangle ouvert inclus dans \mathcal{X} , c'est-à-dire un sous-ensemble de \mathcal{X} de la forme $R = l_1 \times \dots \times l_d$, où l_1, \dots, l_d sont tous des intervalles ouverts de \mathbb{R} . Alors, quel que soit $x \in R$, la densité $p_\pi(x, \cdot)$, dont la définition est donnée par l'équation (2.9), satisfait $p_\pi(x, y) > 0$ pour tout $y \in R$.*

Démonstration. Fixons $x \in R$. Pour tout $i \in \{1, \dots, d\}$, la densité $f_i(\cdot | x^{(-i)})$ satisfait alors

$$f_i(v | x^{(-i)}) > 0, \quad v \in l_i,$$

puisque nous avons supposé que la condition G était respectée. Ceci implique, en vertu de l'équation (2.9), que $p_\pi(x, y) > 0$ quel que soit $y \in R$. ■

Corollaire 2.1. *Sous les hypothèses de la proposition 2.1, nous avons $P_\pi(x, A) > 0$ pour tout $x \in R$ et pour tout $A \in \mathcal{B}_R^+$.*

Une conséquence importante du corollaire 2.1 est que le noyau P_π est π -irréductible lorsque l'ensemble \mathcal{X} est un rectangle ouvert de \mathbb{R}^d . Nous poursuivons maintenant avec une définition tirée de Rudin (1987, page 196).

Définition 2.1. Soit A un sous-ensemble de \mathbb{R}^d , et posons

$$\mathcal{J}_A = \{B \subseteq A : \text{il existe un sous-ensemble ouvert } C \text{ de } \mathbb{R}^d \text{ tel que } C \cap A = B\}.$$

Le sous-ensemble A est dit *connexe* s'il n'existe pas deux sous-ensembles disjoints de A , disons A_1 et A_2 , tous deux dans \mathcal{J}_A , tels que $A_1 \cup A_2 = A$.

La famille de sous-ensembles \mathcal{J}_A est simplement la topologie induite sur A par la topologie euclidienne de \mathbb{R}^d . Intuitivement, un sous-ensemble connexe est un sous-ensemble "en un seul morceau". Par exemple, un rectangle et une boule de \mathbb{R}^d (ouverts ou non) sont des sous-ensembles connexes de \mathbb{R}^d . D'un autre côté, l'union de deux rectangles disjoints de \mathbb{R}^d n'est pas connexe. Selon Rudin (1987, page 197), un sous-ensemble ouvert A de \mathbb{R}^d peut toujours être exprimé comme une union finie ou dénombrable d'ensembles ouverts de \mathbb{R}^d , connexes et disjoints, que nous appelons les *composantes connexes* de A . En particulier, c'est le cas de l'ensemble \mathcal{X} .

Proposition 2.2. *Soit G , une union finie de rectangles ouverts R_1, \dots, R_m , tous inclus dans \mathcal{X} . Supposons que G soit connexe. Alors, pour tout $x \in G$, il existe une version de la densité de la mesure de probabilité $P_\pi^m(x, \cdot)$, disons $p_\pi^m(x, \cdot)$, telle que $p_\pi^m(x, y) > 0$ quel que soit $y \in G$.*

Démonstration. Nous ne considérerons que le cas $m = 2$, le résultat se démontrant facilement pour $m > 2$ à l'aide d'un raisonnement par récurrence. Posons, pour tous $x, y \in G$,

$$p_\pi^2(x, y) = \int_{\mathcal{X}} p_\pi(x, z)p_\pi(z, y) dz.$$

La fonction $p_\pi^2(x, \cdot)$ ainsi définie est bel et bien une version de la densité de la mesure de probabilité $P_\pi^2(x, \cdot)$, et ceci, pour tout $x \in G$. Puisque $G = R_1 \cup R_2$ est connexe, l'ensemble $R_1 \cap R_2$ est non vide. De plus, R_1 et R_2 étant ouverts, la mesure de Lebesgue de l'ensemble $R_1 \cap R_2$ est positive. En vertu de la proposition 2.1, si x et z appartiennent au même rectangle, alors $p_\pi(x, z) > 0$. Il en est de même pour $p_\pi(z, y)$ si z et y appartiennent au même rectangle. Nous avons donc, pour tous $x, y \in G$,

$$p_\pi^2(x, y) \geq \int_{R_1 \cap R_2} p_\pi(x, z) p_\pi(z, y) dz > 0,$$

puisque $\mu(R_1 \cap R_2) > 0$. ■

Corollaire 2.2. *Sous les hypothèses de la proposition 2.2, le noyau markovien P_π satisfait $P_\pi^m(x, A) > 0$ pour tous $x \in G$ et $A \in \mathcal{B}_G^+$.*

Les deux résultats précédents peuvent en fait être généralisés au cas où G est un ensemble connexe quelconque, comme le montrent le prochain théorème et son corollaire.

Théorème 2.3. *Soit U une composante connexe de \mathcal{X} . Alors pour tous $x, z \in U$, il existe un rectangle ouvert $R \subset U$ contenant z , un entier positif m et une version de la densité de la mesure de probabilité $P_\pi^m(x, \cdot)$, disons $p_\pi^m(x, \cdot)$, tels que $p_\pi^m(x, y) > 0$ pour tout $y \in R$.*

Démonstration. Fixons x et z dans U . Puisque U est connexe, il existe une fonction continue $g : [0, 1] \rightarrow U$ telle que $g(0) = x$ et $g(1) = z$ (voir par exemple Royden, 1988, page 183). Pour tout $t \in [0, 1]$, nous pouvons, puisque U est ouvert, choisir un rectangle ouvert R_t tel que $g(t) \in R_t \subset U$. La famille de rectangles ouverts $\mathcal{R} = \{R_t; t \in [0, 1]\}$ ainsi choisie est alors un recouvrement de $\{g(t); t \in [0, 1]\}$, c'est-à-dire qu'elle satisfait $\{g(t); t \in [0, 1]\} \subset \cup_{0 \leq t \leq 1} R_t$. Mais puisque g est continue et que l'intervalle $[0, 1]$ est un sous-ensemble compact de \mathbb{R} , l'ensemble $\{g(t); t \in [0, 1]\}$ est aussi compact (voir Rudin, 1987, page 38). Puisque \mathcal{R} est un recouvrement pour $\{g(t); t \in [0, 1]\}$, la compacité de $\{g(t); t \in [0, 1]\}$ implique, par définition, qu'il est possible d'extraire de \mathcal{R} un sous-recouvrement fini pour $\{g(t); t \in [0, 1]\}$, c'est-à-dire qu'il existe un entier positif m et des réels $0 \leq t_1 \leq \dots \leq t_m \leq 1$ tels que $\{g(t); t \in [0, 1]\} \subset \cup_{i=1}^m R_{t_i}$ (voir

Rudin, 1987, page 35). Puisque $x, z \in G = \cup_{i=1}^m R_{t_i}$ et que G est une union finie de m rectangles ouverts, la proposition 2.2 implique qu'il existe une version de la densité de la mesure de probabilité $P_\pi^m(x, \cdot)$, disons $p_\pi^m(x, \cdot)$, telle que $p_\pi^m(x, y) > 0$ pour tout $y \in R_{t_m}$.

Corollaire 2.3. *Soit U , une composante connexe de \mathcal{X} . Soit $x \in U$ et $A \in \mathcal{B}_U^+$. Alors il existe un entier positif m tel que $P_\pi^m(x, A) > 0$.*

Démonstration. Choisissons un point z contenu dans le support de la restriction de π à l'ensemble A . Nous avons alors, par la définition même de support, $\pi(A \cap V) > 0$ pour tout ensemble ouvert V contenant z , et en particulier, pour tout rectangle ouvert V satisfaisant $z \in V \subset U$. En vertu du théorème 2.3, il existe un rectangle ouvert R satisfaisant $z \in R \subset U$, un entier positif m et une version de la densité de la mesure de probabilité $P_\pi^m(x, \cdot)$, disons $p_\pi^m(x, \cdot)$, tels que $p_\pi^m(x, y) > 0$ pour tout $y \in R$. Ceci implique que

$$\int_A p_\pi^m(x, y) dy \geq \int_{A \cap R} p_\pi^m(x, y) dy > 0,$$

et donc que $P_\pi^m(x, A) > 0$. ■

La connexité de l'ensemble \mathcal{X} est par conséquent une condition suffisante pour que le noyau P_π soit π -irréductible. Pour que P_π soit π -irréductible lorsque \mathcal{X} n'est pas connexe, il est essentiel qu'il soit possible pour $(X_n; n \geq 0)$ de passer d'une composante connexe de \mathcal{X} à une autre afin de pouvoir visiter, avec probabilité positive, n'importe quel sous-ensemble de \mathcal{X} de mesure π positive. Ceci nous amène à la définition suivante.

Définition 2.2. Soit U et V , deux composantes connexes de \mathcal{X} . Nous dirons que V est *accessible à partir de U* , et nous écrirons alors $U \rightarrow V$, s'il existe un entier $m \geq 1$ et une suite de composantes connexes de \mathcal{X} , disons $U = W_0, W_1, \dots, W_{m-1}, W_m = V$, tels que pour tout $j \in \{1, \dots, m\}$, il existe un $x \in W_{j-1}$ et un $i \in \{1, \dots, d\}$ tels que

$$\{y \in \mathbb{R}^d : y^k = x^k, k \neq i\} \cap W_j \neq \emptyset. \quad (2.10)$$

Lorsque $U \rightarrow V$ sera satisfaite par toute paire de composantes connexes (U, V) de \mathcal{X} , nous dirons que les composantes connexes de \mathcal{X} *communiquent*.

Supposons que W_1 et W_2 soient deux composantes connexes de \mathcal{X} satisfaisant (2.10) pour un certain $i \in \{1, \dots, d\}$ et un certain $x \in W_1$. Géométriquement, ceci revient à dire qu'il existe une droite parallèle à l'axe des x^i , passant par x , et traversant la composante connexe W_2 . Puisque cette dernière est un sous-ensemble ouvert de \mathbb{R}^d et que la densité f est positive partout sur W_2 , il s'ensuit que $P_{\langle \pi, i \rangle}(x, W_2) > 0$.

Le point x ne peut cependant pas être le seul point de W_1 à partir duquel il est possible de passer directement à la composante connexe W_2 . En effet, puisque W_1 est aussi un sous-ensemble ouvert, il existe nécessairement un rectangle ouvert $R \subseteq W_1$ contenant x tel que pour tout $z \in R$, $P_{\langle \pi, i \rangle}(z, W_2) > 0$. Considérant que $P_{\langle \pi, i \rangle}(z, R) > 0$ pour tous $z \in R$ et $i \in \{1, \dots, d\}$ (voir la démonstration de la proposition 2.1), il est facile de conclure que nous avons aussi $P_\pi(z, W_2) > 0$ pour tout $z \in R$; en partant de $z \in R$, la chaîne $(X_n; n \geq 0)$ peut se retrouver après une itération, avec probabilité positive, dans W_2 .

La communication des composantes connexes de \mathcal{X} est une condition nécessaire et suffisante pour que le noyau P_π soit π -irréductible. Pour arriver à le montrer, nous aurons besoin du résultat suivant.

Lemme 2.1. *Soit ν une mesure de probabilité sur $(\mathcal{X}, \mathcal{B}_\mathcal{X})$. Soit A et B , deux sous-ensembles mesurables de \mathcal{X} . Supposons que $\nu(A) > 0$ et que pour tout $x \in A$, il existe un entier $m \geq 1$ tel que $P_\pi^m(x, B) > 0$. Alors il existe un entier $n \geq 1$ tel que*

$$\int_A \nu(dx) P_\pi^n(x, B) > 0.$$

Démonstration. Posons, pour tout $n \geq 1$, $A_n = \{x \in A : P_\pi^n(x, B) > 0\}$. Comme $P_\pi^n(\cdot, B)$ est une fonction mesurable et comme $A \in \mathcal{B}_\mathcal{X}$, les ensembles A_1, A_2, \dots sont tous mesurables. Maintenant, puisque $\cup_{n=1}^\infty A_n = A$ et que $\nu(A) > 0$, il existe au moins un entier $k \geq 1$ tel que $\nu(A_k) > 0$. Nous avons alors

$$\int_A \nu(dx) P_\pi^k(x, B) \geq \int_{A_k} \nu(dx) P_\pi^k(x, B) > 0$$

et donc k est un choix possible pour l'entier désiré. ■

Théorème 2.4. *Le noyau P_π est π -irréductible si et seulement si les composantes connexes de \mathcal{X} communiquent.*

Démonstration. Par la définition de la π -irréductibilité, la communication des composantes connexes de \mathcal{X} est clairement une condition nécessaire pour que P_π soit π -irréductible. Sinon, il serait impossible pour la chaîne de Markov $(X_n; n \geq 0)$ de passer d'une composante connexe à une autre. Maintenant, supposons que \mathcal{X} possède seulement deux composantes connexes, disons U et V , telles que U et V communiquent. Fixons $x \in U$ et $A \in \mathcal{B}_\mathcal{X}^+$. Nous allons montrer qu'il existe un entier $n \geq 1$ pour lequel $P_\pi^n(x, A) > 0$. Si $\pi(A \cap U) > 0$, le résultat est immédiat grâce au théorème 2.3. Si $\pi(A \cap U) = 0$, posons $B = A \cap V$. Soit $R \subseteq U$ un rectangle ouvert tel que $P_\pi(y, V) > 0$ pour tout $y \in R$. Puisque $U \rightarrow V$, un tel rectangle existe (voir les deux paragraphes suivant la définition 2.2). Puisque U est connexe, le théorème 2.3 assure qu'il existe un entier positif n_1 tel que $P_\pi^{n_1}(x, R) > 0$. Par le lemme 2.1, nous avons alors

$$\begin{aligned} P_\pi^{n_1+1}(x, V) &= \int_{\mathcal{X}} P_\pi^{n_1}(x, dy) P_\pi(y, V) \\ &\geq \int_R P_\pi^{n_1}(x, dy) P_\pi(y, V) > 0. \end{aligned}$$

Puisque V est aussi connexe, pour tout $y \in V$, il existe un entier $m \geq 1$ tel que $P_\pi^m(y, B) > 0$. Grâce encore une fois au lemme 2.1, nous pouvons en déduire qu'il existe un entier $n_2 \geq 1$ tel que $\int_V P_\pi^{n_1+1}(x, dy) P_\pi^{n_2}(y, B) > 0$. Nous obtenons ainsi

$$\begin{aligned} P_\pi^{n_1+n_2+1}(x, B) &= \int_{\mathcal{X}} P_\pi^{n_1+1}(x, dy) P_\pi^{n_2}(y, B) \\ &\geq \int_V P_\pi^{n_1+1}(x, dy) P_\pi^{n_2}(y, B) > 0. \end{aligned}$$

En posant $n = n_1 + n_2 + 1$, nous avons alors $P_\pi^n(x, B) > 0$, et donc $P_\pi^n(x, A) > 0$. Puisque la communication entre deux composantes connexes est une relation symétrique, la démonstration est aussi valable lorsque $x \in V$. Le résultat se généralise facilement au cas où \mathcal{X} possède plus de deux composantes connexes. ■

La communication des composantes connexes de \mathcal{X} est également une condition nécessaire et suffisante pour assurer que π soit l'unique loi stationnaire pour le noyau markovien P_π . En effet, puisqu'elle implique la π -irréductibilité de P_π , elle implique aussi que π est l'unique loi stationnaire (voir les théorèmes 1.4 et 1.5). Pour prouver sa nécessité, supposons que U et V soient deux composantes connexes de \mathcal{X} telles que V n'est pas accessible à partir de U (ce qui implique aussi que U n'est pas accessible à

partir de V). L'ensemble \mathcal{X} peut alors s'écrire comme l'union de deux sous-ensembles ouverts et disjoints, disons \mathcal{X}_1 et \mathcal{X}_2 , tels que $P_\pi(x, \mathcal{X}_1) = 1$ pour tout $x \in \mathcal{X}_1$, et $P_\pi(x, \mathcal{X}_2) = 1$ pour tout $x \in \mathcal{X}_2$. Pour tout sous-ensemble mesurable A de \mathcal{X} , nous avons alors

$$\begin{aligned} \pi(A \cap \mathcal{X}_1) &= \int_{\mathcal{X}} \pi(dx) P_\pi(x, A \cap \mathcal{X}_1) \quad (\text{puisque } \pi \text{ est stationnaire}) \\ &= \int_{\mathcal{X}_1} \pi(dx) P_\pi(x, A \cap \mathcal{X}_1) + \int_{\mathcal{X}_2} \pi(dx) P_\pi(x, A \cap \mathcal{X}_1) \\ &= \int_{\mathcal{X}_1} \pi(dx) P_\pi(x, A \cap \mathcal{X}_1) \quad (\text{puisque } P_\pi(x, A \cap \mathcal{X}_1) = 0 \text{ si } x \in \mathcal{X}_2), \end{aligned}$$

et donc la mesure de probabilité $\pi_{\mathcal{X}_1}$ définie sur $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ par

$$\pi_{\mathcal{X}_1}(A) = \frac{\pi(A \cap \mathcal{X}_1)}{\pi(\mathcal{X}_1)}, \quad A \in \mathcal{B}_{\mathcal{X}},$$

est aussi stationnaire pour le noyau markovien P_π . Il en est de même pour la mesure $\pi_{\mathcal{X}_2}$, définie de façon similaire, et pour toute autre mesure de probabilité de la forme

$$\alpha \pi_{\mathcal{X}_1}(A) + (1 - \alpha) \pi_{\mathcal{X}_2}(A), \quad A \in \mathcal{B}_{\mathcal{X}}, 0 \leq \alpha \leq 1.$$

S'il existe une autre loi stationnaire pour le noyau P_π , disons ν , cela signifie aussi que pour tout $i \in \{1, \dots, d\}$, la fonction $f_i(\cdot | \cdot)$ est également une version de la densité conditionnelle de X^i étant donné $X^{(-i)}$, lorsque X est une variable aléatoire de loi ν . Un des attraits importants de l'échantillonneur de Gibbs est qu'il ne requiert que la connaissance des densités conditionnelles $f_1(\cdot | \cdot), \dots, f_d(\cdot | \cdot)$, pour en arriver à simuler une variable aléatoire dont la loi est π . De plus, il ramène la simulation d'une variable aléatoire multidimensionnelle à la simulation de plusieurs variables aléatoires unidimensionnelles, une tâche souvent plus facile. Toutefois, l'un des prix à payer est qu'il n'est pas certain que connaître les densités conditionnelles $f_i(\cdot | \cdot)$, $i \in \{1, \dots, d\}$, est équivalent à connaître π . Dans ce cas, il est possible de montrer que la loi de X_n convergera tout de même en variation totale lorsque $n \rightarrow \infty$; néanmoins, la loi limite dépendra alors du point de départ choisi et ne sera assurément pas la loi π .

Puisque la mesure de probabilité π est stationnaire pour le noyau P_π , la communication des composantes connexes de \mathcal{X} est aussi une condition nécessaire et suffisante pour assurer la récurrence de P_π (par le théorème 1.5). Nous sommes maintenant en

mesure d'affirmer que si cette condition est satisfaite et que le noyau P_π est de plus apériodique, nous aurons alors $\|P_\pi^n(x, \cdot) - \pi(\cdot)\| \rightarrow 0$ pour tout $x \in \mathcal{X}$, sauf possible-ment pour quelques $x \in \mathcal{X}$ formant un ensemble N de mesure π nulle (voir le théorème 1.7). Il s'avère que dans le cas où la condition G est satisfaite, cet ensemble est vide; en d'autres termes, le noyau P_π est récurrent au sens de Harris. Pour le montrer, nous aurons besoin du théorème suivant, dont la démonstration est présentée dans Bélisle (1998a).

Théorème 2.5. *Soit (E, \mathcal{E}) un espace mesurable. Soit P un noyau markovien et π une mesure de probabilité, tous deux définis sur (E, \mathcal{E}) . Supposons que P soit récurrent et que π soit stationnaire pour P . Pour tout $n \geq 0$ et pour tout $x \in E$, considérons $P_{\text{sing}}^n(x, \cdot)$, la partie singulière de la décomposition de Lebesgue de la mesure de probabilité $P^n(x, \cdot)$ par rapport à π . Si pour tout $x \in E$,*

$$\lim_{n \rightarrow \infty} P_{\text{sing}}^n(x, E) = 0,$$

alors P est récurrent au sens de Harris.

Le fait que le noyau markovien P_π soit récurrent au sens de Harris lorsqu'il est π -irréductible est une conséquence immédiate du théorème 2.5. En effet, puisque π et la mesure de Lebesgue μ sur $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ sont équivalentes, et que la mesure de probabilité $P_\pi(x, \cdot)$ est absolument continue par rapport à μ pour tout $x \in \mathcal{X}$, elle est aussi absolument continue par rapport à π pour tout $x \in \mathcal{X}$. Il en est alors de même pour les mesures de probabilité $(P_\pi^n(x, \cdot); x \in \mathcal{X}, n \geq 2)$. Nous avons donc:

Théorème 2.6. *Supposons que le noyau markovien P_π soit π -irréductible. Alors P_π est récurrent au sens de Harris.*

Avant d'utiliser le théorème d'Orey pour conclure que lorsque le noyau P_π est π -irréductible, la suite des mesures de probabilité $(P_\pi^n(x, \cdot); n \geq 1)$ converge en variation totale vers π quel que soit $x \in \mathcal{X}$, il reste à établir l'apériodicité du noyau P_π . Pour le faire, nous n'avons pas besoin de condition autre que la π -irréductibilité (rappelons que par définition, un noyau markovien apériodique doit être φ -irréductible pour une certaine mesure φ).

Théorème 2.7. *Si le noyau P_π est π -irréductible, alors il est apériodique.*

Démonstration. Fixons $x \in \mathcal{X}$, $n \geq 1$ et $y \in \text{Supp}(P_\pi^n(x, \cdot))$. Choisissons un $\varepsilon > 0$ suffisamment petit pour que la boule ouverte $B(y, \varepsilon)$ soit entièrement incluse dans \mathcal{X} . Un tel ε existe puisque \mathcal{X} est ouvert. Puisque pour tout $z \in B(y, \varepsilon)$, il existe un rectangle R tel que $z \in R$ et $R \subset B(y, \varepsilon)$, la proposition 2.1 implique que $P_\pi(z, R) > 0$ et donc que $P_\pi(z, B(y, \varepsilon)) > 0$ pour tout $z \in B(y, \varepsilon)$. Il s'ensuit que

$$\begin{aligned} P_\pi^{n+1}(x, B(y, \varepsilon)) &= \int_{\mathcal{X}} P_\pi^n(x, dz) P_\pi(z, B(y, \varepsilon)) \\ &\geq \int_{B(y, \varepsilon)} P_\pi^n(x, dz) P_\pi(z, B(y, \varepsilon)) \\ &> 0. \end{aligned}$$

Ce résultat étant toujours valide si l'on remplace ε par n'importe quel réel $\delta < \varepsilon$, il s'ensuit maintenant que $y \in \text{Supp}(P_\pi^{n+1}(x, \cdot))$. Ceci étant vrai pour tout $y \in \text{Supp}(P_\pi^n(x, \cdot))$, nous pouvons finalement conclure que pour tout $x \in \mathcal{X}$,

$$\text{Supp}(P_\pi^n(x, \cdot)) \subseteq \text{Supp}(P_\pi^{n+1}(x, \cdot)).$$

Le noyau markovien P_π ne peut donc pas posséder un cycle de longueur $p > 1$. Par conséquent, le noyau P_π est apériodique. ■

Nous avons maintenant tous les outils nécessaires pour énoncer sous quelles conditions l'échantillonneur de Gibbs produit une chaîne de Markov telle que la loi de X_n converge en variation totale vers la mesure de probabilité π , quel que soit le point de départ $x_0 \in \mathcal{X}$.

Théorème 2.8. *Soit π une mesure de probabilité sur $(\mathbb{R}^d, \mathcal{B}^d)$, absolument continue par rapport à la mesure de Lebesgue, satisfaisant la condition G. Dénotons par P_π le noyau markovien de l'échantillonneur de Gibbs correspondant à la mesure de probabilité π . Supposons finalement que les composantes connexes de l'ensemble \mathcal{X} (de la condition G) communiquent. Alors, pour tout $x \in \mathcal{X}$,*

$$\lim_{n \rightarrow \infty} \sup_{A \in \mathcal{B}_{\mathcal{X}}} |P_\pi^n(x, A) - \pi(A)| = 0.$$

2.1.6 L'échantillonneur de Gibbs avec mises à jour aléatoires

L'algorithme décrit à la section 2.1.2 est parfois appelé *l'échantillonneur de Gibbs avec mises à jour séquentielles*, puisque les composantes du vecteur initial x_0 sont mises à jour dans l'ordre, selon leurs indices.

Une autre version de l'échantillonneur de Gibbs consiste à choisir aléatoirement selon la loi uniforme sur $\{1, \dots, d\}$, avant chaque mise à jour, l'indice de la composante du vecteur le plus récent qui sera modifiée. Par exemple, supposons que le point de départ choisi soit x_0 . Nous choisissons alors un point I_1 selon la loi uniforme sur $\{1, \dots, d\}$, et nous générons ensuite un point $v \in \mathbb{R}$ selon la densité $f_{I_1}(v|x_0^{(-I_1)})$ pour ensuite poser $X_1 = x_1 = (x_0^1, \dots, x_0^{I_1-1}, v, x_0^{I_1+1}, \dots, x_0^d)$. Ensuite, nous répétons cette procédure pour produire x_2 , à partir de x_1 et de la densité $f_{I_2}(\cdot|x_1^{(-I_2)})$ où I_2 est choisi aléatoirement selon la loi uniforme sur $\{1, \dots, d\}$, et ainsi de suite.

Cet autre algorithme est appelé *l'échantillonneur de Gibbs avec mises à jour aléatoires*. Contrairement à la version séquentielle de l'algorithme, la suite de points $(X_n; n \geq 0)$ obtenue après chacune des mises à jour constitue une chaîne de Markov homogène dans le temps. Le noyau markovien P_π^* régissant la transition de x_n , $n \geq 0$, au point suivant est simplement la moyenne des noyaux $P_{\langle \pi, 1 \rangle}, P_{\langle \pi, 2 \rangle}, \dots, P_{\langle \pi, d \rangle}$, c'est-à-dire que

$$P_\pi^*(x, A) = \frac{1}{d} [P_{\langle \pi, 1 \rangle}(x, A) + P_{\langle \pi, 2 \rangle}(x, A) + \dots + P_{\langle \pi, d \rangle}(x, A)], \quad x \in \mathcal{X}, A \in \mathcal{B}_\mathcal{X}.$$

Puisque la mesure de probabilité π est stationnaire pour chacun des noyaux $(P_{\langle \pi, i \rangle}; i \in \{1, \dots, d\})$, elle l'est aussi pour le noyau P_π^* . En modifiant légèrement les démonstrations de la présente section, nous pourrions montrer que tous les résultats portant sur la convergence de l'échantillonneur de Gibbs avec mises à jour séquentielles sont aussi valides (sous les mêmes conditions) pour cette version de l'algorithme.

Il est aussi possible de construire d'autres versions de l'algorithme, en utilisant par exemple une autre loi que la loi uniforme sur $\{1, \dots, d\}$ pour choisir l'indice des composantes à mettre à jour. Pour que ces algorithmes fonctionnent bien, il est important de s'assurer que chaque élément de $\{1, \dots, d\}$ ait une probabilité non nulle d'être choisi

à chaque mise à jour, afin que la chaîne simulée puisse se “promener” dans \mathcal{X} dans toutes les directions. Cette condition est nécessaire pour que le noyau markovien de la chaîne en question soit π -irréductible. L'échantillonneur de Gibbs avec mises à jour aléatoires est un cas particulier de l'*algorithme hit-and-run* dont nous parlerons à la section 2.3.

Notons que dans leur algorithme original, Geman et Geman (1984) suggèrent aussi la possibilité d'utiliser des mises à jour séquentielles des d composantes du vecteur de départ dans un ordre différent que l'ordre naturel (i.e. $1, 2, \dots, d$). Le fait que π soit stationnaire pour chacun des noyaux markoviens $P_{\langle \pi, 1 \rangle}, \dots, P_{\langle \pi, d \rangle}$ permet en effet de construire beaucoup d'autres noyaux markoviens pour lesquels π sera stationnaire. Par exemple, la transition de x_0 à X_1 pourrait se faire en utilisant les noyaux $P_{\langle \pi, 1 \rangle}, \dots, P_{\langle \pi, d \rangle}$ dans l'ordre inverse, pour ensuite les utiliser dans un ordre totalement différent pour effectuer la transition de $X_1 = x_1$ à X_2 , et ainsi de suite. Si l'utilisation des noyaux $P_{\langle \pi, 1 \rangle}, \dots, P_{\langle \pi, d \rangle}$ dans l'ordre naturel permet de simuler une chaîne de Markov telle que la loi de X_n converge en variation totale vers π , alors il en sera de même si ces noyaux sont utilisés suivant n'importe quel schéma de mises à jour, à condition que chacun des noyaux $P_{\langle \pi, 1 \rangle}, \dots, P_{\langle \pi, d \rangle}$ soit utilisé un nombre infini de fois dans la situation hypothétique où la chaîne de Markov serait simulée sans arrêt.

2.1.7 Regroupement des composantes

Au lieu de mettre à jour une seule composante à chaque étape de l'algorithme, Smith et Roberts (1993) suggèrent de former des “blocs” avec les d composantes du vecteur x_0 , et de mettre à jour en même temps les composantes faisant partie d'un même bloc. Pour illustrer leur idée, supposons que $d = 3$, et qu'en plus de la densité conditionnelle $f_3(\cdot | \cdot)$, la fonction $f_{12}(\cdot | w)$ définie pour tout $w \in \mathbb{R}$ par

$$f_{12}(u, v | w) = \frac{f(u, v, w)}{\int_{\mathbb{R}} \int_{\mathbb{R}} f(s, t, w) ds dt}$$

soit aussi disponible. Notons que si $X = (X^1, X^2, X^3)$ est une variable aléatoire de loi π , alors $f_{12}(\cdot | \cdot)$ est une version de la densité conditionnelle de (X^1, X^2) étant donnée

X^3 . Dans l'algorithme original avec mises à jour séquentielles, il est maintenant possible de combiner les étapes 1 et 2 en une seule pour ainsi donner ce nouvel algorithme décrivant les étapes à suivre pour passer de $X_n = x_n$ à $X_{n+1} = x_{n+1}$:

Étape 1: générer (x_{n+1}^1, x_{n+1}^2) à partir de la densité $f_{12}(\cdot | x_n^3)$;

Étape 2: générer x_{n+1}^3 à partir de la densité $f_3(\cdot | x_{n+1}^1, x_{n+1}^2)$.

Le regroupement des composantes est recommandé par Smith et Roberts (1993) lorsque la loi cible π est telle que si $X = (X^1, \dots, X^d)$ est une variable aléatoire de loi π , alors certaines des composantes de X sont fortement corrélées entre elles. Dans ce cas, il est possible que la chaîne se déplace peu lors d'une itération, pouvant ainsi faire en sorte que la convergence vers π soit très lente. Ce genre de blocage peut contribuer à l'élimination de ce problème, mais le prix à payer est qu'il faut alors simuler des variables aléatoires multidimensionnelles.

Un autre algorithme qui utilise le regroupement des composantes est l'*algorithme d'augmentation de données* de Tanner et Wong (1987), rebaptisé plus tard *algorithme de substitution* par Gelfand et Smith (1990). Ceux-ci étudient ses similarités et ses différences avec l'échantillonneur de Gibbs avec mises à jour séquentielles. Mentionnons aussi que certains auteurs, comme par exemple Athreya, Doss et Sethuraman (1996), utilisent aussi le nom *algorithme de substitution* pour désigner l'échantillonneur de Gibbs; puisque celui-ci permet de simuler une famille de lois beaucoup plus vaste que la famille des lois de Gibbs, le terme *Gibbs* n'a plus vraiment sa place, selon eux, dans le nom de l'algorithme.

2.2 L'algorithme de Metropolis-Hastings

2.2.1 Origines de l'algorithme

Metropolis et coll. (1953) furent vraisemblablement les premiers à proposer l'utili-

sation des chaînes de Markov pour simuler une variable aléatoire distribuée selon une loi multidimensionnelle. Leur algorithme (l'*algorithme de Metropolis*, parfois appelé l'*échantillonneur de Metropolis*) fut développé dans le but de calculer certaines caractéristiques de systèmes physiques constitués de particules en interaction.

Étant donné un tel système (par exemple, l'ensemble des molécules d'une certaine quantité d'un gaz ou d'un liquide), ils supposent que celui-ci peut être représenté comme N particules en interaction dans un cube $C \in \mathbb{R}^3$. Ils supposent ensuite que si le système est observé à un temps aléatoire, alors S , la configuration des particules du système à ce moment, c'est-à-dire l'ensemble des N vecteurs de C qui identifient l'emplacement des N particules, est une variable aléatoire dont la loi π possède une densité proportionnelle à $e^{-\beta E(s)}$, où $E(s)$ représente l'énergie potentielle du système lorsque la configuration de ses particules est s , et où β est une constante positive.

Pour calculer la valeur moyenne d'une fonction réelle $g : C^N \rightarrow \mathbb{R}$, où $g(s)$, $s \in C^N$, représente la valeur d'une caractéristique du système lorsque la configuration des particules est s , il faut donc évaluer

$$\pi(g) = \frac{\int_{C^N} g(s) e^{-\beta E(s)} ds}{\int_{C^N} e^{-\beta E(s)} ds}. \quad (2.11)$$

Puisque N est généralement très grand, il n'est pas toujours possible de calculer analytiquement ou numériquement ce rapport d'intégrales. Une première méthode de Monte-Carlo possible pour estimer $\pi(g)$ consiste à simuler n configurations aléatoires indépendantes S_1, \dots, S_n distribuées uniformément dans C^N et à utiliser les moyennes

$$\frac{1}{n} \sum_{i=1}^n g(S_i) e^{-\beta E(S_i)} \quad \text{et} \quad \frac{1}{n} \sum_{i=1}^n e^{-\beta E(S_i)}$$

pour estimer le numérateur et le dénominateur de (2.11); cependant, cette méthode s'avère inefficace (voir Hammersley et Handscomb, 1964, page 118). Une deuxième méthode possible consiste à simuler un échantillon de variables aléatoires indépendantes S_1, \dots, S_n de loi π et d'utiliser la moyenne $\bar{g}_n = \sum_{i=1}^n n^{-1} g(S_i)$ comme estimateur de $\pi(g)$. Cependant, cette approche n'est pas envisageable si la valeur du dénominateur de (2.11) est inconnue, c'est-à-dire si la mesure π est connue seulement à une constante multiplicative près. L'algorithme de Metropolis permet de contourner ces difficultés.

À partir d'une configuration de départ s_0 arbitrairement choisie dans C^N et un $\alpha > 0$, l'algorithme de Metropolis génère itérativement une suite de configurations s_1, s_2, \dots . Pour ce faire, une configuration $t \in C^N$ est tout d'abord "proposée" comme valeur pour s_1 . Cette configuration t est obtenue à partir de s_0 en faisant subir un déplacement linéaire aléatoire à chacune des N particules de la façon suivante: si (x^1, x^2, x^3) est la position d'une particule selon la configuration s_0 , alors la position de cette même particule dans la configuration t est donnée par

$$x^1 + \alpha U_1, \quad x^2 + \alpha U_2 \quad \text{et} \quad x^3 + \alpha U_3,$$

où U_1, U_2, U_3 sont trois variables aléatoires indépendantes distribuées uniformément sur l'intervalle $(-1, 1)$. Si l'un des déplacements U_1, U_2, U_3 fait en sorte que la particule sort du cube, alors celle-ci retourne dans le cube par le côté opposé et continue son déplacement dans la même direction; par exemple, si le cube C est le cube unité de \mathbb{R}^3 et que $x_1 = 1/2$, $U_1 = 3/4$ et $\alpha = 1$, alors la nouvelle valeur de la composante x^1 devient $1/4$. Cette nouvelle configuration t est ensuite acceptée comme valeur pour s_1 avec une certaine probabilité; cette probabilité dépend de la variation d'énergie potentielle occasionné par ce changement de configuration, c'est-à-dire $\Delta E = E(t) - E(s_0)$. Si $\Delta E < 0$, alors la configuration t est automatiquement acceptée; si $\Delta E \geq 0$, la nouvelle configuration t est acceptée avec probabilité $e^{-\beta \Delta E}$. Si la configuration t n'est pas acceptée, alors la configuration s_1 demeure la même que s_0 . Cette procédure est ensuite répétée pour générer s_2 à l'aide de s_1 , s_3 à l'aide de s_2 , et ainsi de suite. La suite de configurations aléatoires $(S_n; n \geq 0)$ ainsi définie est une chaîne de Markov à valeurs dans C^N . Metropolis et coll. (1953) montrent que S_n est asymptotiquement distribuée selon π , et suggèrent d'utiliser la moyenne ergodique

$$\bar{g}_n = \frac{1}{n} \sum_{i=1}^n g(S_i)$$

pour estimer $\pi(g)$. Remarquons la principale caractéristique de l'algorithme de Metropolis: il n'est aucunement nécessaire de connaître la valeur du dénominateur de (2.11).

Plus tard, Hastings (1970) adapta et généralisa cet algorithme afin de simuler une variable aléatoire distribuée selon une loi essentiellement arbitraire. Malgré l'article de Hastings et malgré le fait que l'algorithme était bien connu et utilisé par les physiciens,

cette approche pour simuler des variables aléatoires distribuées selon une loi compliquée est restée méconnue des statisticiens pendant longtemps. Il fallut attendre l'apparition de l'échantillonneur de Gibbs de Gelfand et Smith (1990) pour que la situation change; toute l'attention portée à l'échantillonneur de Gibbs a permis de faire connaître aussi l'algorithme de Metropolis-Hastings, et plusieurs auteurs s'y sont intéressés par la suite. Citons en autres, Smith et Roberts (1993), Roberts et Smith (1994), Tierney(1994), Chib et Greenberg (1995) et Mengersen et Tweedie (1996).

Puisque nous décrirons ici la généralisation de l'algorithme développé par Hastings (1970), nous utiliserons la terminologie *algorithme de Hastings*. L'appellation *algorithme de Metropolis* sera utilisée seulement lorsque nous discuterons du cas particulier de Metropolis et coll. (1953).

2.2.2 Description de l'algorithme

Soit Q un noyau markovien sur $(\mathbb{R}^d, \mathcal{B}^d)$, pouvant s'écrire sous la forme

$$Q(x, A) = \int_A q(x, y) dy, \quad x \in \mathbb{R}^d, A \in \mathcal{B}^d,$$

pour une certaine fonction non négative q , définie sur $\mathbb{R}^d \times \mathbb{R}^d$. En d'autres termes, Q est un noyau markovien tel que, pour tout $x \in \mathbb{R}^d$, la mesure de probabilité $Q(x, \cdot)$ est absolument continue par rapport à la mesure de Lebesgue sur $(\mathbb{R}^d, \mathcal{B}^d)$, avec densité $q(x, \cdot)$. Supposons maintenant que $\alpha(x, y)$ soit une fonction réelle mesurable, arbitraire (pour le moment), définie sur $\mathbb{R}^d \times \mathbb{R}^d$ et telle que $0 \leq \alpha(x, y) \leq 1$ pour tout $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$.

L'algorithme de Hastings permet de simuler une chaîne de Markov $(X_n; n \geq 0)$ en procédant de la façon suivante. Il requiert tout d'abord un état initial $X_0 = x_0$, pouvant être arbitrairement choisi dans \mathbb{R}^d . Ensuite, la procédure suivante est répétée pour chaque $n \geq 1$: étant donné que $X_n = x$, un point *candidat* y pour X_{n+1} , distribué selon la loi $Q(x, \cdot)$, est simulé. Ce point candidat y est ensuite "accepté" comme valeur pour X_{n+1} avec probabilité $\alpha(x, y)$. Si le point y n'est pas accepté, alors x devient aussi la valeur prise par X_{n+1} .

Les transitions de $X_n = x$ à X_{n+1} de la chaîne de Markov ainsi définie se font donc, pour tous $n \geq 0$ et $x \in \mathbb{R}^d$, selon la loi $P_{(Q, \pi)}(x, \cdot)$ de densité

$$p_{(Q, \pi)}(x, y) = q(x, y)\alpha(x, y), \quad x, y \in \mathbb{R}^d,$$

avec probabilité d'avoir $X_{n+1} = x$ donnée par¹

$$P_{(Q, \pi)}(x, \{x\}) = 1 - \int_{\mathbb{R}^d} q(x, y)\alpha(x, y) dy.$$

Notons que tout dépendant de la façon dont a été définie la fonction α , nous avons possiblement $\int_{\mathbb{R}^d} q(x, y)\alpha(x, y) dy < 1$ pour certains $x \in \mathbb{R}^d$ et donc le noyau défini par $\int_A p_{(Q, \pi)}(x, y) dy$, $x \in \mathbb{R}^d$, $A \in \mathcal{B}^d$, est dans ce cas sous-markovien. Il ne décrit que les transitions "acceptées" de la chaîne.

En posant $r(x) = P_{(Q, \pi)}(x, \{x\})$, le noyau markovien de la chaîne définie par l'algorithme de Hastings peut maintenant s'écrire

$$P_{(Q, \pi)}(x, A) = \int_A p_{(Q, \pi)}(x, y) dy + \int_A r(x)\delta_x(dy), \quad A \in \mathcal{B}^d, \quad (2.12)$$

où δ_x représente la mesure de Dirac en x . Remarquons que pour tout $x \in \mathbb{R}^d$ tel que $r(x) > 0$, la mesure de probabilité $P_{(Q, \pi)}(x, \cdot)$ n'est pas absolument continue par rapport à la mesure de Lebesgue, et que (2.12) est sa décomposition de Lebesgue par rapport à la mesure de Lebesgue. Cela dit, les mesures définies par $\int_A p_{(Q, \pi)}(x, y) dy$, $A \in \mathcal{B}^d$, et $\int_A r(x)\delta_x(dy)$, $A \in \mathcal{B}^d$, sont respectivement la partie absolument continue et la partie singulière (par rapport à la mesure de Lebesgue) de la mesure de probabilité $P_{(Q, \pi)}(x, \cdot)$.

Afin de rendre la mesure de probabilité π stationnaire pour le noyau $P_{(Q, \pi)}$, il est essentiel de définir convenablement les probabilités d'acceptation $\alpha(x, y)$. C'est ce que nous ferons à la prochaine section.

Remarque 2.5. Pour l'algorithme de Hastings, nous permettons à la mesure de probabilité π d'être une loi définie sur $(\mathbb{R}, \mathcal{B})$. Les autres algorithmes étudiés dans ce mémoire

¹La fonction α étant pour le moment arbitraire, nous utilisons ici le symbole π comme indice dans $p_{(Q, \pi)}$ et $P_{(Q, \pi)}$ uniquement pour mettre l'emphase sur le fait que notre objectif est de définir un noyau markovien possédant la loi π comme loi stationnaire. Un peu plus loin, α sera définie en fonction de π , et la présence de cet indice prendra alors tout son sens.

sont triviaux et sans intérêt lorsque $d = 1$. C'est pourquoi nous avons supposé, à la page 36, que π était définie sur $(\mathbb{R}^d, \mathcal{B}^d)$, avec $d \geq 2$.

2.2.3 Réversibilité

Un concept important que nous avons omis jusqu'à présent est la *réversibilité* d'un noyau markovien par rapport à une mesure de probabilité. Ce concept simplifiera grandement notre recherche de la façon dont la fonction α doit être définie pour que π soit stationnaire pour le noyau markovien $P_{(Q, \pi)}$. Bien que nous nous intéressons surtout aux noyaux markoviens (et maintenant, sous-markoviens) et aux mesures de probabilité, la définition de la réversibilité est présentée ici dans un cadre plus général.

Définition 2.3. Soit ν et K , une mesure et un noyau, tous deux définis sur un espace mesurable (E, \mathcal{E}) . Le noyau K est alors dit *réversible par rapport à ν* si, pour tous $A, B \in \mathcal{E}$,

$$\int_A \nu(dx) K(x, B) = \int_B \nu(dy) K(y, A). \quad (2.13)$$

La réversibilité d'un noyau K par rapport à une mesure ν signifie donc que les mesures définies sur (E^2, \mathcal{E}^2) par $\nu(dx)K(x, dy)$ et $\nu(dy)K(y, dx)$ sont identiques. En termes d'une chaîne de Markov avec noyau markovien P , la réversibilité de ce dernier par rapport à π implique ceci: lorsque la loi de X_n est π pour un certain $n \geq 1$, les événements $\{X_n \in A, X_{n+1} \in B\}$ et $\{X_n \in B, X_{n+1} \in A\}$ ont la même probabilité, quels que soient les sous-ensembles $A, B \in \mathcal{E}$.

La réversibilité d'un noyau markovien P par rapport à une mesure de probabilité π s'avère être une condition suffisante pour que π soit stationnaire pour P . En effet, en appliquant (2.13) à P et π et en remplaçant le sous-ensemble A par l'ensemble E , nous obtenons

$$\int_E \pi(dx) P(x, B) = \int_B \pi(dy) P(y, E), \quad B \in \mathcal{E},$$

et, puisque $P(y, E) = 1$ pour tout $y \in E$,

$$\int_E \pi(dx) P(x, B) = \pi(B), \quad B \in \mathcal{E}.$$

Ce résultat implique que si nous parvenons à définir la fonction α de façon à rendre $P_{(Q,\pi)}$ réversible par rapport à π , nous aurons construit, par la même occasion, un noyau markovien avec loi stationnaire π .

Proposition 2.3. *Le noyau markovien $P_{(Q,\pi)}$ est réversible par rapport à π si et seulement si la fonction α satisfait*

$$\pi(dx)Q(x, dy)\alpha(x, y) = \pi(dy)Q(y, dx)\alpha(y, x) \quad \text{pour tous } x, y \in \mathbb{R}^d, \quad (2.14)$$

c'est-à-dire si et seulement si le noyau sous-markovien défini par $Q(x, dy)\alpha(x, y)$ est lui-même réversible par rapport à π .

Démonstration. Choisissons deux sous-ensembles dans \mathcal{B}^d , disons A et B . Nous avons alors

$$\begin{aligned} \int_A \pi(dx)P_{(Q,\pi)}(x, B) &= \int_A \pi(dx) \left[\int_B p_{(Q,\pi)}(x, y) dy + r(x)\mathbf{1}_B(x) \right] \\ &= \int_A \pi(dx) \left[\int_B p_{(Q,\pi)}(x, y) dy \right] + \int_A \pi(dx)r(x)\mathbf{1}_B(x) \\ &= \int_A \pi(dx) \left[\int_B Q(x, dy)\alpha(x, y) \right] + \int_{A \cap B} \pi(dx)r(x). \end{aligned} \quad (2.15)$$

En substituant A à B et B à A dans la dernière égalité, nous obtenons

$$\int_B \pi(dx)P_{(Q,\pi)}(x, A) = \int_B \pi(dx) \left[\int_A Q(x, dy)\alpha(x, y) \right] + \int_{B \cap A} \pi(dx)r(x). \quad (2.16)$$

Le résultat est ensuite immédiat: nous aurons égalité entre (2.15) et (2.16) pour tous $A, B \in \mathcal{B}^d$, si et seulement si

$$\int_A \pi(dx) \left[\int_B Q(x, dy)\alpha(x, y) \right] = \int_B \pi(dx) \left[\int_A Q(x, dy)\alpha(x, y) \right], \quad A, B \in \mathcal{B}^d,$$

c'est-à-dire si et seulement si le noyau sous-markovien défini par $Q(x, dy)\alpha(x, y)$ est réversible par rapport à π . ■

Les mesures π et $(Q(x, dy)\alpha(x, y); x \in \mathbb{R}^d)$ étant toutes absolument continues par rapport à la mesure de Lebesgue, la réversibilité du noyau $Q(x, dy)\alpha(x, y)$ est donc équivalente à exiger que sa densité $q(x, y)\alpha(x, y)$ satisfasse

$$f(x)q(x, y)\alpha(x, y) = f(y)q(y, x)\alpha(y, x), \quad (2.17)$$

pour presque tout $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ (par rapport à la mesure $\pi(dx)Q(x, dy)\alpha(x, y)$ sur $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}^d \times \mathcal{B}^d)$). Rappelons que f est une version de la densité de la loi cible π . Nous allons maintenant déterminer un choix possible pour la fonction α .

Fixons $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$, et supposons qu'à ce point,

$$f(x)q(x, y) > f(y)q(y, x). \quad (2.18)$$

Cela signifie, de façon intuitive, que lorsque la chaîne définie par l'algorithme de Hastings est à l'état y à un certain temps n , la probabilité que le noyau Q propose x comme état suivant est trop petite, par rapport à la probabilité que Q propose la valeur y pour l'état suivant lorsque l'état présent est x . Une possibilité est alors de définir $\alpha(y, x)$ la plus grande possible, c'est-à-dire égale à 1. Dans ce cas, pour que l'équation (2.17) soit satisfaite à ce point (x, y) particulier, il faut poser

$$\alpha(x, y) = \frac{f(y)q(y, x)}{f(x)q(x, y)}. \quad (2.19)$$

Pour les points $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ pour lesquels l'inégalité (2.18) est inversée, suivons la même démarche: posons $\alpha(x, y) = 1$ et calculons $\alpha(y, x)$ à l'aide de (2.19). Aux points $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ où possiblement $f(x)q(x, y) = 0$, la fonction $\alpha(x, y)$ peut être définie arbitrairement, puisque la mesure de l'ensemble $\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d : f(x)q(x, y) = 0\}$, par rapport à la mesure définie par $\pi(dx)Q(x, dy)$, est nulle. À tous les points (x, y) inclus dans cet ensemble, nous définirons ici la fonction $\alpha(x, y)$ comme étant égale à 1.

La fonction α que nous venons de définir, c'est-à-dire

$$\alpha(x, y) = \begin{cases} \min \left\{ \frac{f(y)q(y, x)}{f(x)q(x, y)}, 1 \right\} & \text{si } f(x)q(x, y) > 0, \\ 1 & \text{sinon,} \end{cases}$$

satisfait (2.17). Cette définition de la fonction α est précisément celle suggérée par Hastings (1970). Notons qu'elle n'est pas la seule à satisfaire (2.17); par exemple, si $\alpha_0(x, y) = \alpha(x, y)/2$ pour tout $x, y \in \mathbb{R}^d$, alors α_0 satisfait aussi (2.17). Par contre, Peskun (1973) montre que ce choix de fonction α est optimal, en ce sens qu'il rejette les points candidats "convenables" moins souvent que tout autre choix. Remarquons finalement la principale caractéristique de l'algorithme de Hastings: la densité f intervient à la fois au dénominateur et au numérateur de la fonction α . Il est donc suffisant

de connaître f à une constante multiplicative près pour simuler la chaîne définie par cet algorithme.

Les prochaines sections s'attardent aux conditions qu'il faut imposer pour que la suite de mesures de probabilité $(P_{(Q,\pi)}^n(x, \cdot); n \geq 1)$ converge en variation totale, pour π -presque tout $x \in \mathbb{R}^d$, vers la loi cible π .

Remarque 2.6. La réversibilité d'un noyau markovien P par rapport à une mesure de probabilité π est une condition suffisante pour que π soit stationnaire pour P . Elle n'est toutefois pas nécessaire. Par exemple, considérons le noyau markovien P défini sur $E = \{1, 2, 3\}$ par $P(x, \{x\}) = 1/2$ pour tout $x \in E$ et $P(1, \{2\}) = P(2, \{3\}) = P(3, \{1\}) = 1/2$, et π , la loi uniforme sur E . Alors π est stationnaire pour P , mais P n'est pas réversible par rapport à π . Il suffit de constater qu'entre autres, $1/6 = \pi(\{1\})P(1, \{2\}) \neq \pi(\{2\})P(2, \{1\}) = 0$.

Remarque 2.7. Si P_1 et P_2 sont deux noyaux markoviens réversibles par rapport à une mesure de probabilité, disons π , alors π est aussi stationnaire pour les noyaux markoviens produits P_1P_2 et P_2P_1 et le noyau markovien moyen $\alpha P_1 + (1 - \alpha)P_2$, où $0 \leq \alpha \leq 1$. Cependant, alors que le noyau $\alpha P_1 + (1 - \alpha)P_2$ est automatiquement réversible par rapport à π , ce n'est pas toujours le cas des noyaux produits P_1P_2 et P_2P_1 . Par exemple, nous aurions pu montrer à la section 2.1.4 que les noyaux markoviens $P_{\langle \pi, 1 \rangle}, \dots, P_{\langle \pi, d \rangle}$ de l'échantillonneur de Gibbs appliqué à π étaient tous réversibles par rapport à π ; cela nous aurait aussi permis de conclure que π était stationnaire pour chacun des noyaux $P_{\langle \pi, 1 \rangle}, \dots, P_{\langle \pi, d \rangle}$ et par conséquent, pour le noyau markovien produit P_π et pour le noyau markovien moyen P_π^* . Il n'est pas difficile de montrer que le noyau moyen P_π^* est alors réversible; en revanche, le noyau markovien produit P_π n'est généralement pas réversible par rapport à π . En effet, considérons par exemple les trois carrés ouverts de \mathbb{R}^2 définis par $A = \{(x^1, x^2) \in \mathbb{R}^2 : 0 < x_1, x_2 < 1\}$, $B = \{(x^1, x^2) \in \mathbb{R}^2 : 0 < x_1 < 1, 1 < x_2 < 2\}$ et $C = \{(x^1, x^2) \in \mathbb{R}^2 : 1 < x_1 < 2, 0 < x_2 < 1\}$, et π , la loi uniforme sur $\mathcal{X} = A \cup B \cup C$ (qui est un ensemble en forme de "L"). Alors le noyau $P_\pi = P_{\langle \pi, 1 \rangle} P_{\langle \pi, 2 \rangle}$ satisfait $\int_C \pi(dx) P_\pi(x, B) > 0$ et $\int_B \pi(dx) P_\pi(x, C) = 0$ et il n'est donc pas réversible par rapport à π .

2.2.4 Convergence

Comme dans le cas de l'échantillonneur de Gibbs, une condition de régularité doit d'abord être imposée pour éliminer la possibilité de rencontrer certaines situations problématiques et sans intérêt.

Condition H: *Le point de départ x_0 de l'algorithme de Hastings est toujours choisi dans l'ensemble $\mathcal{X} = \{x \in \mathbb{R}^d : f(x) > 0\}$.*

Puisque $x \in \mathcal{X}$ et $y \notin \mathcal{X}$ implique que $\alpha(x, y) = 0$, la condition H assure notamment que la chaîne de Markov définie par l'algorithme de Hastings ne sortira jamais, avec probabilité 1, du support de π . Nous supposons toujours que la condition H est satisfaite par Q . L'ensemble \mathcal{X} étant absorbant pour le noyau $P_{(Q, \pi)}$, nous étudierons la restriction de ce dernier à l'espace mesurable $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ (le noyau Q , quant à lui, demeure défini sur $(\mathbb{R}^d, \mathcal{B}^d)$).

Le fait que la chaîne de Markov soit générée à partir du noyau Q suggère qu'il est nécessaire que pour tout $x \in \mathcal{X}$, n'importe quel sous-ensemble $A \in \mathcal{B}_{\mathcal{X}}^+$ soit accessible (par rapport à Q) à partir de x^2 . En effet, supposons que pour un certain $x \in \mathcal{X}$ et un certain $A \in \mathcal{B}_{\mathcal{X}}^+$, nous ayons $Q^n(x, A) = 0$ pour tout $n \geq 1$. Cela signifie que le sous-ensemble A n'est pas accessible (par rapport à Q) à partir de x et que le noyau Q ne proposerait jamais, avec probabilité 1, un point candidat de A pour un X_n quelconque. Ainsi, la chaîne ne visiterait jamais le sous-ensemble A , et nous aurions donc $P_{(Q, \pi)}^n(x, A) = 0$ pour tout $n \geq 1$.

Toutefois, le fait que tous les $A \in \mathcal{B}_{\mathcal{X}}^+$ soient accessibles (par rapport à Q) à partir de n'importe quel $x \in \mathcal{X}$ n'est pas suffisant, comme l'illustre le prochain exemple.

²Un sous-ensemble $A \in \mathcal{B}_{\mathcal{X}}$ est dit accessible (par rapport à Q) à partir de x s'il existe un entier positif n tel que $Q^n(x, A) > 0$. Puisque nous avons imposé la condition H, il n'est pas nécessaire que cette propriété soit respectée pour tout $A \in \mathcal{B}_{\mathcal{X}}^+$ lorsque $x \in \mathcal{X}^c$. La condition stipulant que tous les $A \in \mathcal{B}_{\mathcal{X}}^+$ sont accessibles (par rapport à Q) à partir de n'importe quel $x \in \mathcal{X}$ est donc plus faible que la π -irréductibilité de Q , qui revient à dire que tous les $A \in \mathcal{B}_{\mathbb{R}^d}^+$ sont accessibles à partir de n'importe quel $x \in \mathbb{R}^d$, puisque Q est défini sur $(\mathbb{R}^d, \mathcal{B}^d)$.

Exemple 2.2. Soit $A = \{(x^1, x^2) \in \mathbb{R}^2 : 0 < x^1, x^2 < 1\}$, $B = \{(x^1, x^2) \in \mathbb{R}^2 : 2 < x^1, x^2 < 3\}$, et π , la mesure de probabilité uniforme sur $\mathcal{X} = A \cup B$ avec densité

$$f(x^1, x^2) = \begin{cases} \frac{1}{2} & \text{si } (x^1, x^2) \in \mathcal{X}, \\ 0 & \text{sinon.} \end{cases}$$

Considérons maintenant le noyau markovien Q défini sur $(\mathbb{R}^2, \mathcal{B}^2)$ tel que pour tout $x = (x^1, x^2) \in \mathbb{R}^2$, $Q(x, \cdot)$ est la loi uniforme sur $C_x = \{(y^1, y^2) \in \mathbb{R}^2 : x^1 - \frac{1}{2} < y^1 < x^1 + \frac{1}{2}, x^2 - \frac{1}{2} < y^2 < x^2 + \frac{1}{2}\}$, c'est-à-dire sur le carré unité dont le centre est (x^1, x^2) . Pour tout $x \in \mathbb{R}^2$, choisissons comme densité pour la mesure de probabilité $Q(x, \cdot)$ la fonction

$$q(x, y) = \begin{cases} 1 & \text{si } (y^1, y^2) \in C_x, \\ 0 & \text{sinon.} \end{cases}$$

Le noyau markovien Q ainsi défini est π -irréductible sur $(\mathbb{R}^2, \mathcal{B}^2)$. Supposons que le point de départ x_0 pour l'algorithme de Hastings soit choisi dans le carré A . Nous avons alors, quel que soit ce choix de $x_0 \in A$, $Q(x_0, B) = 0$, et la valeur proposée pour X_1 ne sera donc pas contenue dans B . Si la chaîne parvient un jour à visiter le carré B , ce sera donc en au moins deux transitions, ceci étant vrai pour tout $x_0 \in A$. Pour se rendre dans B , la chaîne de Markov devra donc passer par au moins un point de $(A \cup B)^c$. Mais, la densité f étant nulle partout sur $(A \cup B)^c$, chaque fois qu'une valeur de cet ensemble sera proposée par le noyau Q , celle-ci sera nécessairement rejetée, et la chaîne demeurera alors à son état précédent et ne quittera jamais le carré A . Il s'ensuit que le noyau markovien $P_{(Q, \pi)}$ ne peut être π -irréductible. \square

L'irréductibilité est un aspect beaucoup plus difficile à étudier dans le cas de l'algorithme de Hastings que dans le cas de l'échantillonneur de Gibbs. Non seulement il fait intervenir le support de la mesure de probabilité π , mais il dépend aussi du choix du noyau Q . Il faut, bien sûr, que tous les $A \in \mathcal{B}_X^+$ soient accessibles (par rapport à Q) à partir de n'importe quel $x \in \mathcal{X}$, mais il doit être possible pour un tel x et un tel A que la trajectoire entre x et A puisse se faire (relativement au noyau Q) sans passer absolument par \mathcal{X}^c . Si les seules trajectoires possibles entre x et A passent par

\mathcal{X}^c , la chaîne de Markov ne pourra alors jamais atteindre le sous-ensemble A , malgré le fait que $Q^n(x, A) > 0$ pour un certain entier $n \geq 1$, puisque toute valeur proposée à l'extérieur de \mathcal{X} sera toujours rejetée. Mais ce n'est toujours pas suffisant; il y a un autre aspect à considérer.

Exemple 2.3. Soit A, B et C , trois sous-ensembles disjoints de \mathbb{R}^d . Supposons que π soit une mesure de probabilité quelconque sur $(\mathbb{R}^d, \mathcal{B}^d)$, avec densité f telle que $f(x) > 0$ si et seulement si $x \in \mathcal{X} = A \cup B \cup C$. Soit Q un noyau markovien sur $(\mathbb{R}^d, \mathcal{B}^d)$ défini par une famille de densités $\{q(x, \cdot), x \in \mathbb{R}^d\}$ sur \mathbb{R}^d , telle que pour tous $x, y \in \mathcal{X}$, $q(x, y) > 0$ si et seulement si

$$x \in A, y \in B \quad \text{ou} \quad x \in B, y \in B \cup C \quad \text{ou} \quad x \in C, y \in A \cup C.$$

Avec un tel noyau Q , tous les $D \in \mathcal{B}_{\mathcal{X}}^+$ sont accessibles à partir de n'importe quel $x \in \mathcal{X}$. De plus, la transition entre $x \in \mathcal{X}$ et $D \in \mathcal{B}_{\mathcal{X}}^+$ d'une chaîne de Markov avec noyau markovien Q et état initial $x \in \mathcal{X}$ peut se faire (avec probabilité positive) sans qu'une partie de sa trajectoire passe par l'extérieur de \mathcal{X} . Malgré cela, le noyau markovien $P_{(Q, \pi)}$ n'est pas π -irréductible.

En effet, choisissons un $x \in A$. Pour tout $y \in B$, nous avons alors $q(y, x) = 0$ et donc $\alpha(x, y) = 0$. Ensuite, pour tout $y \in C$, nous avons $q(x, y) = 0$. Il s'ensuit que

$$\begin{aligned} P_{(Q, \pi)}(x, B \cup C) &= \int_B q(x, y) \alpha(x, y) dy + \int_C q(x, y) \alpha(x, y) dy \\ &= 0 \end{aligned}$$

puisque pour tout $y \in B \cup C$, l'un des deux termes de $\alpha(x, y)q(x, y)$ est nul. Cette observation, combinée avec le fait que $Q(x, A \cup B \cup C) = 1$ pour tout $x \in A$, implique que lorsque le point de départ est choisi dans A , la chaîne de Markov définie par l'algorithme de Hastings appliqué à π et Q ne sort jamais de A avec probabilité 1. Le noyau $P_{(Q, \pi)}$ n'est donc pas π -irréductible. \square

Il faut donc imposer des conditions beaucoup plus fortes que la simple accessibilité de tous les $A \in \mathcal{B}_{\mathcal{X}}^+$ à partir de n'importe quel $x \in \mathcal{X}$ pour assurer la π -irréductibilité de $P_{(Q, \pi)}$. Puisque celle-ci fait intervenir plusieurs aspects outre cette propriété d'accessibilité, il est difficile d'énoncer des conditions nécessaires et suffisantes

simples qui l'assurent. Néanmoins, l'un des avantages de l'algorithme est que le choix de Q est relativement arbitraire, et il est par conséquent possible de choisir un Q qui satisfait les conditions simples qui sont disponibles. Le prochain théorème fournit un exemple de condition toute simple qui assure la π -irréductibilité de $P_{(Q,\pi)}$.

Théorème 2.9. *Supposons que $q(x, y) > 0$ pour tous $x, y \in \mathcal{X}$. Alors le noyau markovien $P_{(Q,\pi)}$ est π -irréductible.*

Démonstration. Fixons $x \in \mathcal{X}$ et $A \in \mathcal{B}_{\mathcal{X}}^+$. Nous avons alors

$$\begin{aligned} P_{(Q,\pi)}(x, A) &= \int_A q(x, y)\alpha(x, y) dy + r(x)\mathbf{1}_A(x) \\ &\geq \int_A q(x, y)\alpha(x, y) dy \\ &> 0 \quad \text{puisque } \alpha(x, y) > 0 \text{ pour tout } (x, y) \in \mathcal{X}^2. \end{aligned}$$

Nous avons donc $P_{(Q,\pi)}(x, A) > 0$ quels que soient $x \in \mathcal{X}$ et $A \in \mathcal{B}_{\mathcal{X}}^+$. ■

Des conditions plus faibles, mais aussi plus compliquées à vérifier, sont disponibles par exemple dans Roberts et Smith (1994).

Contrairement à la π -irréductibilité, l'apériodicité de $P_{(Q,\pi)}$ est un aspect beaucoup plus facile à étudier. Le théorème suivant énonce une condition nécessaire et suffisante pour que le noyau $P_{(Q,\pi)}$ soit apériodique.

Théorème 2.10. *Supposons que le noyau markovien $P_{(Q,\pi)}$ soit π -irréductible. Il est alors apériodique, si et seulement si au moins une des deux conditions suivantes est respectée:*

- i) $\pi(\{x \in \mathcal{X} : r(x) > 0\}) > 0$;
- ii) le noyau Q est apériodique.

Démonstration. Supposons, dans le but d'obtenir une contradiction, que la condition (i) soit satisfaite, mais que le noyau $P_{(Q,\pi)}$ soit périodique de période $p \geq 2$. Soit $\{C_1, C_2, \dots, C_p\}$ un cycle de longueur p pour $P_{(Q,\pi)}$. Puisque $\pi(\cup_{i=1}^p C_i) = 1$, au moins un de ces sous-ensembles, disons C_k , satisfait $\pi(\{x \in C_k : r(x) > 0\}) > 0$, et

donc $\pi(\{x \in C_k : P_{(Q,\pi)}(x, C_k) > 0\}) > 0$, ce qui contredit la périodicité de $P_{(Q,\pi)}$. La condition (i) implique donc que le noyau $P_{(Q,\pi)}$ est apériodique, peu importe que Q soit apériodique ou non.

Maintenant, si la condition (ii) est satisfaite, mais que (i) ne l'est pas, nous avons alors $r(x) = 0$ pour presque tout $x \in \mathcal{X}$, ce qui implique que pour presque tout $x \in \mathcal{X}$, les mesures $P_{(Q,\pi)}(x, \cdot)$ et $Q(x, \cdot)$ sont identiques. Le noyau Q étant apériodique, le noyau $P_{(Q,\pi)}$ l'est alors aussi. Ceci nous permet de conclure que si les deux conditions sont violées, alors $P_{(Q,\pi)}$ est périodique. ■

Si le noyau $P_{(Q,\pi)}$ est π -irréductible, les résultats du chapitre 1 permettent de conclure que le noyau $P_{(Q,\pi)}$ est aussi récurrent, et que la mesure de probabilité π est l'unique loi stationnaire de $P_{(Q,\pi)}$. Comme dans le cas de l'échantillonneur de Gibbs, nous pouvons conclure encore plus.

Théorème 2.11. *Si le noyau markovien $P_{(Q,\pi)}$ est π -irréductible, alors il est récurrent au sens de Harris.*

Démonstration. Fixons $x \in \mathcal{X}$. Nous savons, grâce à (2.12), que la partie singulière de la mesure de probabilité $P_{(Q,\pi)}(x, \cdot)$ par rapport à la mesure de Lebesgue μ sur $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ accorde toute sa masse au singleton $\{x\}$. Considérons maintenant le noyau markovien produit $P_{(Q,\pi)}^2$, et choisissons un borélien $A \in \mathcal{B}_{\mathcal{X}}$ tel que $\mu(A) = 0$. Puisque $P_{(Q,\pi)}(y, A) = 0$ lorsque $y \notin A$, le noyau $P_{(Q,\pi)}^2$ satisfait $P_{(Q,\pi)}^2(x, A) = 0$ lorsque $x \notin A$. Il s'ensuit que la partie singulière de la mesure de probabilité $P_{(Q,\pi)}^2(x, \cdot)$ accorde aussi toute sa masse au singleton $\{x\}$. Le même raisonnement permet de conclure que ce résultat est aussi vrai pour la mesure de probabilité $P_{(Q,\pi)}^n(x, \cdot)$, quel que soit $n > 2$.

Or, la probabilité $P_{(Q,\pi)}^n(x, \{x\})$ est celle de l'événement

$$\{X_0 = x, X_1 \in \mathcal{X}, \dots, X_{n-1} \in \mathcal{X}, X_n = x\};$$

elle peut s'écrire

$$P_{(Q,\pi)}^n(x, \{x\}) = \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} P_{(Q,\pi)}(x, dy_1) \cdots P_{(Q,\pi)}(y_{n-1}, \{x\}).$$

Puisque pour tout $y \in \mathcal{X}$, nous avons $P_{(Q,\pi)}(y, A) = 0$ si $\mu(A) = 0$ et $y \notin A$, elle peut aussi s'écrire

$$\begin{aligned} P_{(Q,\pi)}^n(x, \{x\}) &= \int_{\{x\}} \cdots \int_{\{x\}} P_{(Q,\pi)}(x, dy_1) \cdots P_{(Q,\pi)}(y_{n-1}, \{x\}) \\ &= [P_{(Q,\pi)}(x, \{x\})]^n \\ &= r(x)^n. \end{aligned}$$

Puisque par hypothèse, le noyau $P_{(Q,\pi)}$ est π -irréductible, nous avons $r(x) < 1$ pour tout $x \in \mathcal{X}$. Nous avons donc $\lim_{n \rightarrow \infty} r(x)^n = 0$; en d'autres termes, la masse totale de la partie singulière de la mesure de probabilité $P_{(Q,\pi)}^n(x, \cdot)$ par rapport à μ tend vers 0 lorsque $n \rightarrow \infty$. Puisque μ et π sont équivalentes, ce résultat est aussi valide si nous considérons plutôt la partie singulière des mesures de probabilité $(P_{(Q,\pi)}^n(x, \cdot); n \geq 1)$ par rapport à π . Ceci étant vrai pour tout $x \in \mathcal{X}$, le théorème 2.5 implique que le noyau markovien $P_{(Q,\pi)}$ est récurrent au sens de Harris. ■

Lorsque le noyau $P_{(Q,\pi)}$ est π -irréductible et apériodique, le théorème d'Orey et le théorème 2.11 permettent de conclure que pour tout $x \in \mathcal{X}$,

$$\lim_{n \rightarrow \infty} \sup_{A \in \mathcal{B}_{\mathcal{X}}} |P_{(Q,\pi)}^n(x, A) - \pi(A)| = 0.$$

2.2.5 Cas particuliers de l'algorithme

Jusqu'à présent, le choix du noyau markovien Q est resté totalement arbitraire, si nous excluons le fait qu'il doit faire en sorte que le noyau $P_{(Q,\pi)}$ qui lui est associé soit π -irréductible et apériodique. Il existe naturellement une multitude de choix possibles. Chacun des différents noyaux Q définit un nouvel algorithme et un nouveau noyau markovien $P_{(Q,\pi)}$. En restreignant le choix de Q à certaines familles de noyaux markoviens, il est possible d'obtenir des cas particuliers intéressants. Nous décrivons ici certaines classes d'algorithme, parmi celles que l'on retrouve le plus souvent dans la littérature.

L'algorithme de Metropolis. Supposons que le noyau Q soit choisi de telle sorte que des versions $(q(x, \cdot); x \in \mathbb{R}^d)$ des densités des mesures $(Q(x, \cdot); x \in \mathbb{R}^d)$ satisfassent

$$q(x, y) = q(y, x) \quad \text{pour tous } x, y \in \mathbb{R}^d.$$

Les probabilités d'acceptation $\alpha(x, y)$ se simplifient alors pour devenir

$$\alpha(x, y) = \min \left\{ \frac{f(y)}{f(x)}, 1 \right\}.$$

La fonction q n'intervient donc plus dans le calcul des probabilités d'acceptation et le genre de situation décrit à l'exemple 2.3 ne peut plus survenir avec un tel choix de noyau Q . Pour que $P_{(Q, \pi)}$ soit π -irréductible, il est nécessaire et suffisant que Q soit π -irréductible et qu'il soit possible que la chaîne de Markov puisse se rendre dans $A \in \mathcal{B}_X^+$ à partir de $x \in X$ sans avoir à passer par X^c . Avec un tel choix de Q , l'algorithme est appelé l'*algorithme de Metropolis*, puisqu'il est simplement une adaptation de l'algorithme original de Metropolis et coll. (1953) à un contexte autre que celui décrit à la section 2.2.1.

L'algorithme de la chaîne d'indépendance. Un autre choix possible de noyau Q pour générer les points candidats, est un noyau satisfaisant, pour une certaine mesure de probabilité ν sur $(\mathbb{R}^d, \mathcal{B}^d)$,

$$Q(x, \cdot) = \nu(\cdot) \quad \text{pour tout } x \in \mathbb{R}^d.$$

Cette possibilité fut suggérée par Hastings (1970). Avec un tel noyau Q , la loi utilisée pour générer un point candidat pour X_{n+1} , $n \geq 0$, est donc toujours la même quelle que soit la valeur prise précédemment par X_n . C'est pourquoi Tierney (1994) appelle cette version l'*algorithme de la chaîne d'indépendance*. Toutefois, cette terminologie ne veut pas dire que les variables simulées $(X_n; n \geq 0)$ sont indépendantes, puisque la probabilité d'acceptation de la valeur proposée pour X_{n+1} , $n \geq 0$, dépend de l'état précédent X_n ; Pour cette version de l'algorithme, le noyau $P_{(Q, \pi)}$ est π -irréductible et apériodique si et seulement si $\nu(A) > 0$ pour tout $A \in \mathcal{B}_X^+$ (voir Tierney, 1994).

L'algorithme de la marche aléatoire. Cet algorithme est aussi un cas particulier de l'algorithme de Metropolis. Il utilise un noyau Q dont les densités $(q(x, \cdot); x \in \mathbb{R}^d)$ satisfont, pour une certaine densité h sur \mathbb{R}^d ,

$$q(x, y) = h(y - x), \quad x, y \in \mathbb{R}^d.$$

Avec cet algorithme, une valeur proposée Y pour X_{n+1} , lorsque $X_n = x$, est simplement obtenue en simulant d'abord une variable aléatoire Z de densité h et en posant ensuite $Y = x + Z$. Notons que l'exemple 2.2 est une illustration de cet algorithme, avec la densité uniforme sur le carré unité centré à 0 comme densité h .

La terminologie *marche aléatoire* vient du fait que le noyau markovien Q associé à q est celui d'une marche aléatoire sur \mathbb{R}^d , dont les "pas" sont gouvernés par la loi associée à la densité h . Une condition suffisante pour que le noyau $P_{(Q, \pi)}$ soit π -irréductible est que h soit positive partout dans un voisinage de l'origine, et que \mathcal{X} soit ouvert et connexe (voir Tierney, 1994).

On discute aussi de d'autres cas particuliers intéressants de l'algorithme de Hastings dans notamment Tierney (1994), Chib et Greenberg (1995), Brooks (1998) et Roberts et Rosenthal (1998a).

2.2.6 Mises à jour composante par composante

L'algorithme de Hastings, tel que nous l'avons vu jusqu'à présent, requiert la simulation de variables aléatoires distribuées selon les lois multidimensionnelles ($Q(x, \cdot)$, $x \in \mathbb{R}^d$) (lorsque $d \geq 2$). Puisqu'en général, la simulation de lois multidimensionnelles est une tâche relativement ardue, le choix d'un noyau Q adéquat pour appliquer l'algorithme de Hastings à π peut être assez restreint. Pour pallier cette difficulté, Hastings (1970) propose une autre version de son algorithme, qui consiste à décomposer le passage de $X_n = x$ à X_{n+1} en d mises à jour du vecteur x et à effectuer ces mises à jour à l'aide de lois unidimensionnelles. Nous verrons que cette version s'apparente très fortement à l'échantillonneur de Gibbs, et pour cette raison, nous réutiliserons à certains endroits la notation de la section 2.1.3.

Supposons que pour tout $i \in \{1, \dots, d\}$, Q_i soit un noyau markovien sur $(\mathbb{R}^d, \mathcal{B}^d)$, tel que pour tout $x \in \mathbb{R}^d$, la mesure de probabilité $Q_i(x, \cdot)$ accorde toute sa masse à l'ensemble $D_x^i = \{y \in \mathbb{R}^d : y^j = x^j, j \neq i\}$, c'est-à-dire l'ensemble des vecteurs de \mathbb{R}^d dont toutes les composantes sont égales aux composantes de x , exceptée la $i^{\text{ième}}$.

Supposons finalement que pour tout $x \in \mathbb{R}^d$ et pour tout $i \in \{1, \dots, d\}$, la mesure de probabilité $Q_i(x, \cdot)$ possède une densité par rapport à la mesure de Lebesgue (unidimensionnelle) sur D_x^i , disons $q_i(x, y)$. Notons que pour tout $x \in \mathbb{R}^d$, générer un point y à partir de la mesure de probabilité $Q_i(x, \cdot)$ revient en fait à remplacer la composante i du vecteur x , par un point de \mathbb{R} .

Pour tout $i \in \{1, \dots, d\}$ et pour tous $x, y \in \mathcal{X}$, posons

$$\alpha_i(x, y) = \min \left\{ \frac{f(y)q_i(y, x)}{f(x)q_i(x, y)}, 1 \right\}, \quad (2.20)$$

et dénotons par $P_{(Q_i, \pi)}$, le noyau markovien de l'algorithme de Hastings correspondant à Q_i et α_i , c'est-à-dire $P_{(Q_i, \pi)}(x, dy) = Q_i(x, dy)\alpha_i(x, y) + r_i(x)\delta_x(dy)$. Ayant suivi la même démarche qu'à la section 2.2.3 pour construire les noyaux $P_{(Q_i, \pi)}$, $i \in \{1, \dots, d\}$, nous pouvons conclure que la mesure de probabilité π est aussi stationnaire pour chacun d'eux, et aussi pour le noyau produit $P_{(Q_1, \pi)} \cdots P_{(Q_d, \pi)}$. Il est maintenant facile de construire un nouvel algorithme pour générer une chaîne de Markov avec loi stationnaire π , en utilisant successivement les noyaux $P_{(Q_1, \pi)}$, ..., $P_{(Q_d, \pi)}$ pour effectuer la transition de $X_n = x$ à X_{n+1} . Comme dans le cas de l'échantillonneur de Gibbs, cette transition se fait en d mises à jour, en suivant le schéma suivant:

$$\begin{aligned} (x^1, x^2, \dots, x^d) &\rightarrow (X_{n+1}^1, x^2, \dots, x^d) \rightarrow \dots \rightarrow (X_{n+1}^1, \dots, X_{n+1}^{d-1}, x^d) \\ &\rightarrow (X_{n+1}^1, \dots, X_{n+1}^{d-1}, X_{n+1}^d). \end{aligned}$$

La principale différence entre les deux algorithmes est qu'une fois la transition terminée, il est possible que certaines des composantes du vecteur X_{n+1} soient identiques à celles de x , puisque certains points candidats peuvent être rejetés. Remarquons pour tout $i \in \{1, \dots, d\}$ et pour tous $x, y \in \mathcal{X}$, le terme $f(y)/f(x)$ que l'on retrouve dans (2.20) peut s'écrire

$$\frac{f(y)}{f(x)} = \frac{f_i(y^i | x^{(-i)})}{f_i(x^i | x^{(-i)})},$$

et donc tout comme l'échantillonneur de Gibbs, cet algorithme utilise, mais indirectement cette fois, les densités conditionnelles $f_1(\cdot | \cdot)$, ..., $f_d(\cdot | \cdot)$ (telles que définies à la section 2.1.3). Puisque pour tous $x \in \mathbb{R}^d$ et $i \in \{1, \dots, d\}$, il nous est permis de connaître la densité $f_i(\cdot | x^{(-i)})$ à une constante multiplicative près, cette version de l'algorithme de Hastings peut être utilisée comme solution alternative à l'échantillonneur de Gibbs

lorsque les constantes de normalisation des densités conditionnelles sont difficiles à calculer.

Certains auteurs ont aussi suggéré des algorithmes hybrides entre l'échantillonneur de Gibbs et l'algorithme de Hastings, lorsque les constantes de normalisation sont inconnues pour seulement une ou quelques densités conditionnelles (voir par exemple Chib et Greenberg, 1995). Par exemple, supposons que $d = 3$ et que les constantes de normalisation des densités conditionnelles soient toutes connues, sauf pour $f_2(\cdot | x^{(-i)})$, $x \in \mathbb{R}^d$. Il suffit d'utiliser les noyaux $P_{\langle \pi, 1 \rangle}$ et $P_{\langle \pi, 3 \rangle}$ lorsque vient le temps de mettre à jour les composantes 1 et 3, et de remplacer le noyau $P_{\langle \pi, 2 \rangle}$ par un noyau $P_{(Q_2, \pi)}$ pour ainsi permettre de mettre à jour la composante 2. Remarquons aussi que si les noyaux $P_{(Q_1, \pi)}$, ..., $P_{(Q_d, \pi)}$ utilisés dans l'algorithme de Hastings avec mises à jour composante par composante sont en fait les noyaux markoviens $P_{\langle \pi, 1 \rangle}$, ..., $P_{\langle \pi, d \rangle}$ de l'échantillonneur de Gibbs, nous avons $\alpha_i(x, y) = 1$ pour tous $x, y \in \mathcal{X}$ et pour tout $i \in \{1, \dots, d\}$, et les deux algorithmes sont alors identiques. En ce sens, l'échantillonneur de Gibbs peut être vu comme un cas particulier de cette version de l'algorithme de Hastings, dans lequel aucun point proposé n'est rejeté, puisqu'alors les mises à jour se font en générant des points "directement" des densités conditionnelles.

Il est évidemment possible de construire beaucoup d'autres algorithmes à partir de l'algorithme de Hastings avec mises à jour composante par composante, comme par exemple en choisissant aléatoirement l'indice du noyau à être utilisé avant chaque mise à jour, ou en utilisant des noyaux qui mettent à jour plus d'une composante à la fois.

2.3 L'algorithme hit-and-run

2.3.1 Origines de l'algorithme

L'algorithme *hit-and-run* (ou l'échantillonneur *hit-and-run*) fut d'abord proposé par Boneh et Golan (1979) dans un contexte d'identification de contraintes linéairement

indépendantes, et par Smith (1984) comme un outil de simulation de lois uniformes sur des sous-ensembles ouverts et bornés de \mathbb{R}^d .

Supposons que A soit un sous-ensemble ouvert et borné de \mathbb{R}^d . La méthode la plus simple pour simuler une variable aléatoire X uniformément distribuée sur A est la *méthode du rejet*. Elle consiste d'abord à déterminer un sous-ensemble C de \mathbb{R}^d tel que $A \subseteq C$. Autant que possible, le sous-ensemble C doit être un sous-ensemble simple, comme par exemple un cube, un rectangle, une boule, etc. Il suffit ensuite de simuler une suite de variables aléatoires indépendantes Y_1, Y_2, \dots uniformément distribuées sur C et de poser $X = Y_t$, où $t = \min\{i \geq 1 : Y_i \in A\}$. La variable aléatoire X ainsi définie est uniformément distribuée sur A .

Cette méthode n'est pas toujours efficace; il se peut qu'il faille simuler un très grand nombre de variables aléatoires Y_1, Y_2, \dots uniformément distribuées sur C avant d'avoir $Y_i \in A$ pour un certain $i \geq 1$. C'est le cas entre autres lorsque la forme de A nous oblige à choisir un sous-ensemble C beaucoup plus grand que A , ou lorsque d est très grand. Par exemple, supposons que A soit la boule ouverte de \mathbb{R}^d de rayon 1 et centrée à l'origine, et que C soit le plus petit cube de \mathbb{R}^d contenant A , c'est-à-dire $C = \{(x^1, \dots, x^d) \in \mathbb{R}^d : -1 \leq x^1, \dots, x^d \leq 1\}$. Si Y est une variable aléatoire uniformément distribuée sur C , alors la probabilité de l'événement $\{Y \in A\}$ est donnée par le rapport des volumes de A et C , c'est-à-dire

$$\mathbb{P}[Y \in A] = \frac{1}{d2^{d-1}} \frac{\pi^{d/2}}{\Gamma(d/2)}. \quad (2.21)$$

À mesure que d augmente, $\mathbb{P}[Y \in A]$ diminue, et de plus, $\lim_{d \rightarrow \infty} \mathbb{P}[Y \in A] = 0$. En particulier, ceci implique que plus d est grand, plus le nombre moyen de variables aléatoires uniformément distribuées sur C qu'il faut simuler pour obtenir une seule valeur dans A est grand. Par exemple, si $d = 20$, la simulation de plus de 40 millions de variables aléatoires (en moyenne) sera nécessaire pour simuler une seule variable X uniformément distribuée sur A ! Dans ce cas, la méthode du rejet n'est pas du tout envisageable.

Pour simuler une variable aléatoire uniformément distribuée sur un sous-ensemble ouvert et borné A de \mathbb{R}^d , l'approche proposée par Smith (1984) consiste à générer une suite de points $(x_n; n \geq 0)$, en choisissant x_0 arbitrairement dans A et en appliquant

ensuite la procédure suivante pour tout $n \geq 1$: une *direction* θ_{n+1} est d'abord choisie aléatoirement selon la loi uniforme sur la sphère de rayon 1 centrée à l'origine; le point x_{n+1} est ensuite choisi aléatoirement, selon la loi uniforme sur $\{x \in A : x = x_n + \lambda\theta_{n+1} \text{ pour un certain } \lambda \in \mathbb{R}\}$, c'est-à-dire sur la partie de la droite passant par x_n et $x_n + \theta_{n+1}$ qui est située dans A . Smith (1984) montre que si $(X_n; n \geq 0)$ représente la suite de points aléatoires que l'algorithme génère, alors la loi asymptotique du point X_n est la loi uniforme sur A .

Cette première version de l'algorithme fut plus tard généralisé par Bélisle, Romeijn et Smith (1993) pour permettre la simulation d'une loi absolument continue par rapport à la mesure de Lebesgue sur un sous-ensemble ouvert et borné de \mathbb{R}^d . Ces auteurs donnent aussi la possibilité d'utiliser une mesure de probabilité arbitraire ν sur la sphère unité pour choisir les directions. La convergence de l'algorithme est alors établie sous certaines restrictions sur ν et sous la condition que la densité f de la loi cible soit bornée, positive partout sur A , et continue presque partout sur A . Plusieurs de ces contraintes furent plus tard relâchées dans Bélisle (1998a) et Diaconis et Freedman (1998), pour ainsi permettre l'application de l'algorithme à un plus vaste éventail de lois.

2.3.2 Description de l'algorithme

Soit S_d la sphère unité centrée à l'origine, c'est-à-dire $S_d = \{\theta \in \mathbb{R}^d : \|\theta\| = 1\}$, où $\|\cdot\|$ représente la distance euclidienne, et soit ν une mesure de probabilité arbitraire sur (S_d, \mathcal{B}_{S_d}) , que nous appellerons la *loi des directions*. Dénotons aussi par e_1, e_2, \dots, e_d les vecteurs de la base canonique de \mathbb{R}^d , c'est-à-dire les vecteurs

$$(1, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots, (0, \dots, 0, 1)$$

respectivement.

Étant donné un point de départ $x_0 \in \mathbb{R}^d$ arbitrairement choisi, l'algorithme hit-and-run génère itérativement une suite de points x_1, x_2, \dots dans \mathbb{R}^d , en suivant les étapes suivantes pour passer de x_n à x_{n+1} , $n \geq 0$:

Étape 1 : choisir une *direction* θ_{n+1} sur S_d selon la mesure de probabilité ν ;

Étape 2 : choisir aléatoirement un $\lambda_{n+1} \in \mathbb{R}$ distribué selon la densité

$$f_{(x_n, \theta_{n+1})}(\lambda) = \frac{f(x_n + \lambda\theta_{n+1})}{\int_{\mathbb{R}} f(x_n + r\theta_{n+1}) dr}; \quad (2.22)$$

Étape 3 : poser $x_{n+1} = x_n + \lambda_{n+1}\theta_{n+1}$.

Remarquons que si X est une variable aléatoire de loi π , alors la fonction $f_{(x, \theta)}(\cdot)$ n'est qu'une reparamétrisation de la densité conditionnelle de X étant donné que X est située sur la droite passant par x et $x+\theta$, c'est-à-dire étant donné que $X \in L_{x, x+\theta} = \{x + \lambda\theta; \lambda \in \mathbb{R}\}$.

Tout comme les autres algorithmes que nous avons présentés jusqu'à maintenant, l'algorithme hit-and-run définit une chaîne de Markov $(X_n; n \geq 0)$ dont le noyau markovien possède comme loi stationnaire la mesure de probabilité π . Lorsque ν est la loi uniforme sur $\{e_1, \dots, e_d\}$, l'algorithme hit-and-run n'est autre que l'échantillonneur de Gibbs avec mises à jour aléatoires appliqué à π . Lorsque ν est la loi uniforme sur (S_d, \mathcal{B}_{S_d}) et que π est la loi uniforme sur un sous-ensemble borné et ouvert A de \mathbb{R}^d , nous retrouvons l'algorithme proposé par Boneh et Golan (1979) et Smith (1984). Ce cas particulier est maintenant parfois appelé *hypersphere directions hit-and-run algorithm* ou plus simplement, *HD algorithm*. Lorsque π est la loi uniforme sur A et que ν est la loi uniforme sur $\{e_1, \dots, e_d\}$, l'algorithme est parfois appelé *coordinate directions hit-and-run algorithm*, ou plus simplement, *CD algorithm*; pour plus de détails sur ces deux derniers cas particuliers, voir Berbee et coll. (1987) et Bélisle et coll. (1998).

Pour certains $x \in \mathbb{R}^d$ et $\theta \in S_d$, il se peut que la fonction $f_{(x, \theta)}$ ne soit pas une densité bien définie sur \mathbb{R} , c'est-à-dire que $\int_{\mathbb{R}} f(x+r\theta) dr \in \{0, \infty\}$. On peut par contre montrer, en suivant un développement similaire à la démonstration du théorème 2.1, que pour tout $\theta \in S_d$,

$$\pi \left(\left\{ x \in \mathbb{R}^d : 0 < \int_{\mathbb{R}} f(x+r\theta) dr < \infty \right\} \right) = 1.$$

Pour éviter toute complication avec ces densités mal définies, nous imposerons une condition à la mesure de probabilité π ; cette condition est une simple généralisation

de la condition G de la section 2.1.3, et elle a pour but d'éliminer le même genre de situations problématiques que nous y avons décrit.

Condition HR: *Il existe une version f de la densité de π telle que:*

- 1) *l'ensemble $\mathcal{X} = \{x \in \mathbb{R}^d : f(x) > 0\}$ est ouvert;*
- 2) *pour tout $\theta \in S_d$ et pour tout $x \in \mathcal{X}$, $0 < \int_{\mathbb{R}} f(x + r\theta) dr < \infty$.*

Encore une fois, la plupart des mesures de probabilité absolument continues par rapport à la mesure de Lebesgue que l'on retrouve en pratique satisfont la condition HR. Nous supposons donc à partir de maintenant que la condition HR est toujours satisfaite par π , que l'algorithme est défini à l'aide d'une densité f satisfaisant les points 1 et 2 de la condition HR, et que le point de départ de l'algorithme hit-and-run est toujours choisi dans \mathcal{X} .

2.3.3 Stationnarité de π

Pour tout $\theta \in S_d$, posons

$$P_{(\theta, \pi)}(x, A) = \int_{\mathbb{R}} \mathbf{1}_A(x + \lambda\theta) f_{(x, \theta)}(\lambda) d\lambda, \quad x \in \mathcal{X}, A \in \mathcal{B}_{\mathcal{X}}. \quad (2.23)$$

Grâce à la condition HR, la fonction $P_{(\theta, \pi)}(x, \cdot)$ est une mesure de probabilité sur $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ pour tout $(x, \theta) \in \mathcal{X} \times S_d$, et la fonction $P_{(\cdot, \pi)}(\cdot, A)$ est mesurable par rapport à la σ -algèbre produit $\mathcal{B}_{S_d} \times \mathcal{B}_{\mathcal{X}}$ pour tout $A \in \mathcal{B}_{\mathcal{X}}$. Pour tout $\theta \in S_d$, $P_{(\theta, \pi)}(\cdot, \cdot)$ est donc un noyau markovien bien défini sur $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$.

Chaque itération de l'algorithme hit-and-run consiste donc à choisir d'abord aléatoirement un noyau markovien $P_{(\theta, \pi)}$ à l'aide de la mesure de probabilité ν sur (S_d, \mathcal{B}_{S_d}) et à l'utiliser ensuite pour effectuer la transition de $X_n = x_n$ à X_{n+1} , $n \geq 0$. Le noyau markovien $P_{(\nu, \pi)}$ régissant ces transitions est donc la moyenne des noyaux markoviens $(P_{(\theta, \pi)}; \theta \in S_d)$, pondérée selon la mesure de probabilité ν :

$$P_{(\nu, \pi)}(x, A) = \int_{S_d} \nu(d\theta) P_{(\theta, \pi)}(x, A), \quad x \in \mathcal{X}, A \in \mathcal{B}_{\mathcal{X}}.$$

Le fait que la fonction $P_{(\cdot, \pi)}(\cdot, A)$ soit mesurable pour tout $A \in \mathcal{B}_X$ par rapport à la σ -algèbre produit $\mathcal{B}_{S_d} \times \mathcal{B}^d$ permet la définition d'un tel noyau.

Le prochain théorème montre que pour tout $\theta \in S_d$, le noyau $P_{(\theta, \pi)}$ est réversible par rapport à la mesure de probabilité π , ce qui implique aussi que π est stationnaire pour $P_{(\theta, \pi)}$ quel que soit $\theta \in S_d$ (voir section 2.2.3).

Théorème 2.12. *Pout tout $\theta \in S_d$, le noyau markovien $P_{(\theta, \pi)}$ est réversible par rapport à π .*

Démonstration. Fixons $\theta \in S_d$. Pour tous $x, y \in \mathcal{X}$, posons $L_{x, x+\theta} = \{x + \lambda\theta; \lambda \in \mathbb{R}\}$ et

$$M_{x, y} = \int_{\mathbb{R}} f\left(x + r \frac{y - x}{\|y - x\|}\right) dr.$$

Remarquons que si x et y sont tels que $y \in L_{x, x+\theta}$, alors $(y - x)/\|y - x\| = \theta$. Pour tout $x \in \mathcal{X}$, dénotons maintenant par $\mu_\theta(x, \cdot)$ la mesure de Lebesgue unidimensionnelle sur $L_{x, x+\theta}$ vue comme mesure sur $(\mathbb{R}^d, \mathcal{B}^d)$. La fonction $\mu_\theta(\cdot, \cdot)$ est ainsi un noyau sur $(\mathbb{R}^d, \mathcal{B}^d)$. Définissons maintenant la mesure ϕ sur $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}^d \times \mathcal{B}^d)$ par

$$\phi(A \times B) = \int_B \mu_\theta(x, A) dx, \quad A, B \in \mathcal{B}^d.$$

Pour tous $A, B \in \mathcal{B}_X$, le noyau $P_{(\theta, \pi)}$ satisfait alors

$$\begin{aligned} \int_A \pi(dx) P_{(\theta, \pi)}(x, B) &= \int_A \pi(dx) \left[\int_{\mathbb{R}} \mathbf{1}_B(x + \lambda\theta) \frac{f(x + \lambda\theta)}{\int_{\mathbb{R}} f(x + r\theta) dr} d\lambda \right] \\ &= \int_A \mu(dx) f(x) \left[\int_B \mu_\theta(x, dy) \frac{f(y)}{M_{x, y}} \right] \\ &= \int_{A \times B} \phi(dx \times dy) \frac{f(x)f(y)}{M_{x, y}}. \end{aligned} \quad (2.24)$$

La dernière égalité se justifie à l'aide d'une version généralisée du théorème de Fubini, que l'on retrouve notamment dans Billingsley (1995, page 241). Puisque $y \in L_{x, x+\theta}$ si et seulement si $x \in L_{y, y+\theta}$, la fonction $M_{x, y}$ satisfait

$$M_{x, y} = M_{y, x} \quad \text{pour tous } x, y \in \mathbb{R}^d.$$

La fonction $f(x)f(y)/M_{x, y}$ est ainsi symétrique en x et y . Pour prouver la réversibilité de $P_{(\theta, \pi)}$ par rapport à π , c'est-à-dire pour montrer qu'on peut remplacer l'ensemble

$A \times B$ par $B \times A$ sans changer la valeur de l'intégrale (2.24), il suffit maintenant de montrer que $\phi(A \times B) = \phi(B \times A)$ pour tous $A, B \in \mathcal{B}^d$, c'est-à-dire que le noyau μ_θ est réversible par rapport à la mesure de Lebesgue μ .

Posons maintenant $\Lambda = \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d : y \in L_{x, x+\theta}\}$, et pour tous $A, B \in \mathcal{B}^d$, $\Lambda_{A, B} = \{(x, y) \in A \times B : y \in L_{x, x+\theta}\}$ (on peut facilement montrer que ces ensembles sont tous dans $\mathcal{B}^d \times \mathcal{B}^d$). Puisque la mesure ϕ accorde une mesure nulle à l'ensemble $\Lambda = \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d : y \notin L_{x, x+\theta}\}$ et que $\Lambda_{A, B} = \Lambda \cap (A \times B)$, nous avons $\phi(A \times B) = \phi(\Lambda_{A, B})$. Maintenant, puisque $y \in L_{x, x+\theta}$ implique que $x \in L_{y, y+\theta}$, les ensembles $\Lambda_{A, B}$ et $\Lambda_{B, A} = \{(x, y) \in B \times A : y \in L_{x, x+\theta}\}$ sont exactement identiques, et donc $\phi(\Lambda_{A, B}) = \phi(\Lambda_{B, A})$. Finalement, puisque $\Lambda_{B, A}$ contient les mêmes éléments que $B \times A$ sauf pour un ensemble de points de $\mathbb{R}^d \times \mathbb{R}^d$ de mesure ϕ nulle, nous avons $\phi(\Lambda_{B, A}) = \phi(B \times A)$, et donc $\phi(A \times B) = \phi(B \times A)$. ■

Le noyau markovien $P_{(\nu, \pi)}$ étant une moyenne pondérée de noyaux markoviens réversibles par rapport à π , il est lui-même réversible par rapport à π . En effet, pour tous $A, B \in \mathcal{B}^d$, il satisfait

$$\begin{aligned} \int_A \pi(dx) P_{(\nu, \pi)}(x, B) &= \int_A \left[\int_{S_d} \nu(d\theta) P_{(\theta, \pi)}(x, B) \right] \\ &= \int_{S_d} \nu(d\theta) \left[\int_A \pi(dx) P_{(\theta, \pi)}(x, B) \right] \end{aligned} \quad (2.25)$$

$$\begin{aligned} &= \int_{S_d} \nu(d\theta) \left[\int_B \pi(dx) P_{(\theta, \pi)}(x, A) \right] \\ &= \int_B \pi(dx) \left[\int_{S_d} \nu(d\theta) P_{(\theta, \pi)}(x, A) \right] \quad (2.26) \\ &= \int_B \pi(dx) P_{(\nu, \pi)}(x, A), \end{aligned}$$

où le changement de l'ordre d'intégration de (2.25) et (2.26) est justifié par le théorème de Fubini.

Puisque la réversibilité de $P_{(\nu, \pi)}$ par rapport à π implique que π est stationnaire pour $P_{(\nu, \pi)}$, il suffira maintenant que $P_{(\nu, \pi)}$ soit π -irréductible et apériodique pour qu'il satisfasse $\lim_{n \rightarrow \infty} \|P_{(\nu, \pi)}^n(x, \cdot) - \pi(\cdot)\|$ pour π -presque tout $x \in \mathcal{X}$. Nous verrons à la prochaine section que sous ces conditions, nous avons en fait $\lim_{n \rightarrow \infty} \|P_{(\nu, \pi)}^n(x, \cdot) - \pi(\cdot)\|$ pour tout $x \in \mathcal{X}$.

2.3.4 Convergence

Dans cette section sont présentées des conditions nécessaires et suffisantes pour assurer que le noyau markovien $P_{(\nu, \pi)}$ soit apériodique et récurrent au sens de Harris, lorsque la loi π satisfait la condition HR. Tous les résultats présentés ici sont démontrés dans Bélisle, Romeijn et Smith (1993) et Bélisle (1998a). Mentionnons que les résultats de la section 2.1 sur la convergence de l'échantillonneur de Gibbs en étaient en fait des cas particuliers. Nous débutons par une généralisation du concept de communication entre les composantes connexes de l'ensemble ouvert \mathcal{X} .

Définition 2.4. Soit U et V , deux composantes connexes de \mathcal{X} . Nous dirons que V est *accessible à partir de U* (par rapport à la mesure de probabilité ν), et nous écrirons simplement $U \rightarrow V$, s'il existe un entier $m \geq 1$ et une suite de composantes connexes de \mathcal{X} , disons $U = W_0, W_1, \dots, W_{m-1}, W_m = V$, tels que pour tout $j \in \{1, \dots, m\}$, il existe un $x \in W_{j-1}$ et un $\theta \in \text{Supp}(\nu)$ satisfaisant

$$\{x + \lambda\theta; \lambda \in \mathbb{R}\} \cap W_j \neq \emptyset. \quad (2.27)$$

Lorsque $U \rightarrow V$ sera satisfaite par toute paire de composantes connexes (U, V) de \mathcal{X} , nous dirons que les composantes connexes de \mathcal{X} *communiquent par rapport à la mesure de probabilité ν* .

Géométriquement, le fait que (2.27) soit satisfaite par deux composantes connexes W_1 et W_2 de \mathcal{X} pour un certain $x \in W_1$ et un certain $\theta \in \text{Supp}(\nu)$ signifie que la droite passant par x et $x + \theta$ traverse la composante connexe W_2 . Si tel est le cas, le fait que W_2 soit ouvert et que θ soit contenu dans le support de ν implique qu'il existe un voisinage C de θ , avec $\nu(C) > 0$, tel que $P_{(\vartheta, \pi)}(x, W_2) > 0$ pour tout $\vartheta \in C$, et donc que $P_{(\nu, \pi)}(x, W_2) > 0$. Finalement, le fait que W_1 soit aussi ouvert implique qu'il existe nécessairement un $B \in \mathcal{B}_{\mathcal{X}}^+$ avec $B \subseteq W_1$ tel que $P_{(\nu, \pi)}(y, W_2) > 0$ pour tout $y \in B$.

La communication (par rapport à ν) des composantes connexes de \mathcal{X} est clairement nécessaire pour que la chaîne de Markov $(X_n; n \geq 0)$ puisse passer d'une composante connexe à une autre, et ainsi visiter avec probabilité positive n'importe quel sous-ensemble $A \in \mathcal{B}_{\mathcal{X}}^+$. Dans le cas où ν est la loi uniforme sur $\{e_1, \dots, e_d\}$, c'est-à-dire dans

le cas de l'échantillonneur de Gibbs, la communication des composantes connexes était aussi suffisante pour que le noyau P_π^* soit π -irréductible. Lorsque ν est arbitraire, il y a un deuxième aspect important à considérer en vue d'obtenir la π -irréductibilité. Le prochain exemple montre qu'à l'intérieur d'une composante connexe U de \mathcal{X} , la chaîne de Markov $(X_n; n \geq 0)$ de l'algorithme hit-and-run ne peut pas nécessairement se déplacer librement; plus précisément, il montre que si $x \in U$ et $A \in \mathcal{B}_U^+$, alors il n'existe pas nécessairement un entier $n \geq 0$ tel que $P_{(\nu, \pi)}^n(x, A) > 0$.

Exemple 2.4. Supposons que π soit la loi uniforme sur \mathcal{X} , le carré unité ouvert de \mathbb{R}^2 . Supposons aussi que ν soit la mesure de probabilité sur (S_d, \mathcal{B}_{S_d}) définie par $\nu((0, 1)) = \nu((0, -1)) = 1/2$. La mesure de probabilité ν ne permettant que des déplacements parallèles à la droite passant par $(0, 1)$ et $(0, -1)$, la chaîne $(X_n; n \geq 0)$ définie par l'algorithme hit-and-run ne quittera jamais, avec probabilité 1, la droite $L_{x_0, x_0+(0, 1)}$, où x_0 est le point de départ choisi. Pour tout ensemble $A \in \mathcal{B}_\mathcal{X}^+$ tel que $A \cap \{x_0 + \lambda(0, 1); \lambda \in \mathbb{R}\} = \emptyset$, le noyau $P_{(\nu, \pi)}$ satisfera alors $P_{(\nu, \pi)}^n(x_0, A) = 0$ pour tout $n \geq 1$. Il s'ensuit que le noyau $P_{(\nu, \pi)}$ n'est pas π -irréductible. \square

La communication entre les composantes connexes de \mathcal{X} n'est donc pas suffisante pour assurer la π -irréductibilité de $P_{(\nu, \pi)}$. Il est aussi nécessaire qu'à l'intérieur d'une composante connexe donnée U , tous les sous-ensembles mesurables contenus dans U soient accessibles à partir de n'importe quel $x \in U$. Dans le cas de l'échantillonneur de Gibbs, cette condition était automatiquement respectée (voir le théorème 2.3), puisque les déplacements de la chaîne de Markov qu'il définit peuvent se faire dans les d directions. Si nous imposons une restriction à la mesure de probabilité ν de telle sorte que les déplacements de $\{X_n, n \geq 0\}$ puissent se faire dans les d directions, alors cette condition sera satisfaite. Ceci nous mène à la définition suivante.

Définition 2.5. La mesure de probabilité ν est dite à *pleine dimension* si son support n'est pas contenu dans un sous-espace linéaire de \mathbb{R}^d de dimension inférieure à d .

En d'autres termes, le fait que la mesure de probabilité ν soit à pleine dimension signifie qu'il existe un ensemble de d vecteurs $\{\theta_1, \dots, \theta_d\}$, tous contenus dans $\text{Supp}(\nu)$, tel que $\{\theta_1, \dots, \theta_d\}$ est une base vectorielle pour \mathbb{R}^d (i.e. pour tout $x \in \mathbb{R}^d$, il existe des nombres réels non tous nuls $\alpha_1, \dots, \alpha_d$ tels que $\alpha_1\theta_1 + \dots + \alpha_d\theta_d = x$). Dans le cas de

l'échantillonneur de Gibbs avec mises à jour aléatoires, la condition de la définition 2.5 était toujours satisfaite, puisque le support de ν était la base canonique de \mathbb{R}^d .

Lorsque \mathcal{X} est un ensemble connexe, le fait que la mesure de probabilité ν soit à pleine dimension est une condition nécessaire et suffisante pour que le noyau markovien $P_{(\nu, \pi)}$ soit π -irréductible. Dans le cas où \mathcal{X} possède plus d'une composante connexe et que ν est à pleine dimension, le noyau $P_{(\nu, \pi)}$ est π -irréductible si et seulement si nous ajoutons l'hypothèse que les composantes connexes de \mathcal{X} communiquent par rapport à la loi ν . Ces conditions sont aussi nécessaires et suffisantes pour que π soit l'unique loi stationnaire de $P_{(\nu, \pi)}$ et pour que ce dernier soit apériodique. Le théorème suivant, démontré dans Bélisle, Romeijn et Smith (1993) et Bélisle (1998a) résume ces résultats.

Théorème 2.13. *Le noyau markovien $P_{(\nu, \pi)}$ est π -irréductible et apériodique et π est l'unique loi stationnaire de $P_{(\nu, \pi)}$ si et seulement si les deux conditions suivantes sont satisfaites:*

- 1) *la mesure de probabilité ν est à pleine dimension;*
- 2) *les composantes connexes de \mathcal{X} communiquent par rapport à ν .*

Lorsque le noyau $P_{(\nu, \pi)}$ est π -irréductible, il est aussi récurrent puisque π est une loi stationnaire pour $P_{(\nu, \pi)}$ (voir les théorèmes 1.4 et 1.5). En vertu du théorème 1.7, nous pouvons conclure que pour π -presque tout $x \in \mathcal{X}$,

$$\lim_{n \rightarrow \infty} \|P_{(\nu, \pi)}^n(x, \cdot) - \pi(\cdot)\| = 0. \quad (2.28)$$

Dans le cas de l'échantillonneur de Gibbs avec mises à jour séquentielles, le fait que la mesure de probabilité $P_\pi(x, \cdot)$ soit absolument continue par rapport à π pour tout $x \in \mathcal{X}$ permettait d'affirmer que le noyau P_π était également récurrent au sens de Harris et donc qu'en fait, la convergence dans (2.28) avait lieu pour tout $x \in \mathcal{X}$. Dans le cas de l'algorithme hit-and-run, lorsque la mesure de probabilité ν est absolument continue par rapport à la mesure de Lebesgue sur (S_d, \mathcal{B}_{S_d}) , la mesure $P_{(\nu, \pi)}(x, \cdot)$ est absolument continue par rapport à la mesure de Lebesgue sur $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ pour tout $x \in \mathcal{X}$ (voir Bélisle, Romeijn et Smith, 1993, section 3). Lorsque ν n'est pas absolument continue par rapport à la mesure de Lebesgue sur (S_d, \mathcal{B}_{S_d}) , la mesure de probabilité

$P_{(\nu, \pi)}(x, \cdot)$ possède une partie singulière par rapport à la mesure de Lebesgue sur $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$; en particulier, dans le cas de l'échantillonneur de Gibbs avec mises à jour aléatoires, la mesure de probabilité $P_{\pi}^*(x, \cdot)$ est singulière par rapport à la mesure de Lebesgue, pour tout $x \in \mathcal{X}$. En effet, pour tout $x \in \mathcal{X}$, la mesure $P_{\pi}^*(x, \cdot)$ accorde toute sa masse à l'ensemble $\{y \in \mathcal{X} : y^j = x^j, j \neq 1\} \cup \dots \cup \{y \in \mathcal{X} : y^j = x^j, j \neq d\}$, c'est-à-dire l'ensemble constitué des d droites parallèles aux axes et passant par x . Cependant, Bélisle, Romeijn et Smith (1993, proposition 1) montrent que si ν est n'importe quelle mesure de probabilité à pleine dimension sur (S_d, \mathcal{B}_{S_d}) , alors pour tout $x \in \mathcal{X}$, les décompositions de Lebesgue des mesures de probabilité $(P_{(\nu, \pi)}^n(x, \cdot); n \geq 0)$ par rapport à la mesure de Lebesgue, c'est-à-dire

$$P_{(\nu, \pi)}^n(x, A) = P_{\text{abs}}^n(x, A) + P_{\text{sing}}^n(x, A), \quad A \in \mathcal{B}_{\mathcal{X}}, n \geq 0,$$

sont telles que:

- i) $P_{\text{abs}}^n(x, \mathcal{X}) > 0$ pour tout $n \geq d$;
- ii) $\lim_{n \rightarrow \infty} P_{\text{abs}}^n(x, \mathcal{X}) = 1$.

Puisque π est une mesure équivalente à la mesure de Lebesgue sur $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$, ce résultat est aussi valide si l'on considère plutôt les décompositions de Lebesgue des mesures $(P_{(\nu, \pi)}^n(x, \cdot), x \in \mathcal{X}, n \geq 1)$ par rapport à π . En vertu du théorème 2.5, lorsque le noyau $P_{(\nu, \pi)}$ est π -irréductible, il est aussi récurrent au sens de Harris. Ceci nous permet de conclure:

Théorème 2.14. *Soit π une mesure de probabilité sur $(\mathbb{R}^d, \mathcal{B}^d)$, absolument continue par rapport à la mesure de Lebesgue, satisfaisant la condition HR. Soit ν une mesure de probabilité sur (S_d, \mathcal{B}_{S_d}) . Dénotons par $P_{(\nu, \pi)}$ le noyau markovien de l'algorithme hit-and-run appliqué à π et utilisant ν comme loi des directions. Supposons finalement que ν soit à pleine dimension, et que les composantes connexes de l'ensemble \mathcal{X} (de la condition HR) communiquent par rapport à ν . Alors pour tout $x \in \mathcal{X}$,*

$$\lim_{n \rightarrow \infty} \|P_{(\nu, \pi)}^n(x, \cdot) - \pi(\cdot)\| = 0.$$

Chapitre 3

Vitesse de convergence et ergodicité géométrique

Au chapitre 2, nous nous sommes attardés aux conditions que doivent satisfaire la loi cible π et son support (et le noyau Q dans le cas de l'algorithme de Hastings) pour que les algorithmes de simulation markovienne qui y sont décrits définissent une chaîne de Markov $(X_n; n \geq 0)$ dont la loi de X_n converge en variation totale vers π . Cependant, nous n'avons encore vu aucun résultat permettant d'évaluer la vitesse à laquelle elle a lieu.

Supposons que P soit le noyau markovien d'une chaîne de Markov $(X_n; n \geq 0)$ définie par un algorithme de simulation markovienne, utilisé dans le but de simuler une variable aléatoire de loi π . Évidemment, ce qui sera effectivement simulé n'est pas la chaîne $(X_n; n \geq 0)$ mais plutôt une portion de celle-ci, disons les $k + 1$ premières variables, c'est-à-dire $(X_n; 0 \leq n \leq k)$. Puisque l'objectif est de pouvoir considérer X_k comme une variable aléatoire de loi π , il est naturel de choisir l'entier k en fonction de la distance (en variation totale) entre la vraie loi de X_k et la loi cible π . Par exemple, si une distance en variation totale d'au plus $\varepsilon > 0$ entre la loi de X_k et π est jugée acceptable (en ce sens que cette distance est si petite que l'on peut, à toutes fins pratiques, considérer que X_k est effectivement distribuée selon π), il suffit alors de

choisir k comme étant un entier satisfaisant

$$\|P^k(x_0, \cdot) - \pi(\cdot)\| \leq \varepsilon,$$

où x_0 est l'état initial choisi pour l'algorithme, et de simuler ensuite la chaîne $(X_n; n \geq 0)$ jusqu'à X_k . Évidemment, cette stratégie requiert l'évaluation de $\|P^n(x_0, \cdot) - \pi(\cdot)\|$, ce qui est généralement impossible en pratique si P ne satisfait pas certaines conditions.

La vitesse de convergence de $P^n(x, \cdot)$, $x \in E$, vers π est un aspect pratique important de la simulation markovienne. Comme nous le verrons plus loin, il est possible de construire des exemples de lois π pour lesquelles certains algorithmes appliqués à π définissent un noyau markovien P dont la convergence en variation totale de $P^n(x, \cdot)$ vers π est extrêmement lente quel que soit $x \in E$. Dans ces cas, la loi de la dernière variable de $(X_n; 0 \leq n \leq k)$ peut être très différente de π , même lorsque k est un très grand entier.

Dans le présent chapitre, nous nous intéressons entre autres à cet aspect. Nous présenterons d'abord quelques résultats importants portant sur la vitesse de convergence et les méthodes permettant d'évaluer celle-ci dans un contexte général, c'est-à-dire lorsque P est un noyau qui n'est pas nécessairement défini par un algorithme de simulation. Nous discuterons ensuite de cet aspect dans le contexte précis de la simulation markovienne. Nous aborderons aussi un autre sujet très lié à la vitesse de convergence: le comportement asymptotique de la moyenne $\bar{g}_n = (n+1)^{-1} \sum_{i=0}^n g(X_i)$, où $(X_n; n \geq 0)$ est une chaîne de Markov et où g est une fonction réelle mesurable. Les deux premières sections peuvent être vues comme un complément au chapitre 1.

3.1 Ergodicité géométrique

Étant donné un noyau markovien P défini sur un espace mesurable (E, \mathcal{E}) avec loi stationnaire π tel que $\|P^n(x, \cdot) - \pi(\cdot)\| \rightarrow 0$ pour tout $x \in E$, il est généralement très difficile, voire impossible, d'évaluer

$$\|P^n(x, \cdot) - \pi(\cdot)\|, \quad x \in E, n \geq 1. \quad (3.1)$$

Dans cette section, nous étudions une certaine classe de noyaux markoviens que l'on appelle les noyaux markoviens *géométriquement ergodiques*; pour ces derniers, il est possible de calculer des bornes supérieures pour (3.1) quels que soient $x \in E$ et $n \geq 1$. Nous verrons aussi que ces noyaux possèdent plusieurs autres propriétés intéressantes. Nous débutons par quelques définitions.

Définition 3.1. Un noyau markovien P défini sur un espace mesurable (E, \mathcal{E}) est dit *ergodique* s'il existe une mesure de probabilité π telle que

$$\lim_{n \rightarrow \infty} \|P^n(x, \cdot) - \pi(\cdot)\| = 0 \quad \text{pour tout } x \in E. \quad (3.2)$$

En vertu du théorème 1.6, un noyau markovien P apériodique et récurrent au sens de Harris avec loi stationnaire π est ergodique. Selon la proposition 6.3 de Nummelin (1984), l'inverse est aussi vrai: un noyau markovien P est apériodique, récurrent au sens de Harris et possède π comme loi stationnaire si et seulement si $\lim_{n \rightarrow \infty} \|P^n(x, \cdot) - \pi(\cdot)\| = 0$ pour tout $x \in E$. Tous les noyaux markoviens ergodiques, et seulement ceux-ci, sont donc des noyaux markoviens apériodiques, récurrents au sens de Harris et possédant une loi stationnaire. Notons que remplacer (3.2) par

$$\lim_{n \rightarrow \infty} \|\lambda P^n(\cdot) - \pi(\cdot)\| = 0 \quad \text{pour toute mesure de probabilité } \lambda \text{ sur } (E, \mathcal{E})$$

fournit une définition équivalente de l'ergodicité. Nous poursuivons avec la définition d'un type d'ergodicité particulier.

Définition 3.2. Un noyau markovien ergodique P avec loi stationnaire π est dit *géométriquement ergodique* s'il existe un réel $0 < \rho < 1$ et une fonction réelle étendue M définie sur E satisfaisant $\int |M| d\pi < \infty$ tels que

$$\|P^n(x, \cdot) - \pi(\cdot)\| \leq M(x)\rho^n \quad (3.3)$$

pour tous $x \in E$ et $n \geq 1$.

Lorsque P est un noyau géométriquement ergodique avec loi stationnaire π , la convergence en variation totale de $P^n(x, \cdot)$ vers π se fait donc, pour π -presque tout $x \in E$, à une vitesse géométrique. Remarquons aussi que pour un tel noyau, la convergence de

$\|\lambda P^n(\cdot) - \pi(\cdot)\|$ est aussi géométrique pour toute mesure de probabilité λ sur (E, \mathcal{E}) satisfaisant $\lambda(|M|) = \int |M| d\lambda < \infty$. En effet, (3.3) implique que

$$-M(x)\rho^n \leq P^n(x, A) - \pi(A) \leq M(x)\rho^n$$

pour tous $n \geq 1$, $x \in E$ et $A \in \mathcal{E}$. En intégrant ensuite les trois termes par rapport à la mesure λ , nous obtenons

$$-\lambda(M)\rho^n \leq \lambda P(A) - \pi(A) \leq \lambda(M)\rho^n$$

pour tous $n \geq 1$ et $A \in \mathcal{E}$, et en prenant finalement le supremum sur $A \in \mathcal{E}$, nous obtenons

$$\|\lambda P^n(\cdot) - \pi(\cdot)\| \leq \lambda(M)\rho^n$$

pour tout $n \geq 1$. Le prochain théorème présente une condition nécessaire et suffisante pour qu'un noyau markovien ergodique soit en fait géométriquement ergodique. Sa démonstration est présentée entre autres dans Meyn et Tweedie (1993, chapitre 15). Nous avons cependant besoin d'une définition supplémentaire avant de l'énoncer.

Définition 3.3. Soit P un noyau markovien défini sur un espace mesurable (E, \mathcal{E}) . Supposons que P soit φ -irréductible pour une certaine mesure σ -finie φ . Un ensemble $C \in \mathcal{E}$ tel que $\varphi(C) > 0$ est appelé un *ensemble petit pour P* s'il existe un entier $m \geq 1$, un réel $\beta > 0$ et une mesure de probabilité ν sur (E, \mathcal{E}) tels que

$$P^m(x, A) \geq \beta\nu(A) \quad \text{pour tous } x \in C \text{ et } A \in \mathcal{E}. \quad (3.4)$$

Tout ensemble mesurable $A \in \mathcal{E}$ tel que $\varphi(A) > 0$ possède au moins un sous-ensemble petit pour un noyau markovien φ -irréductible P (voir Nummelin, 1984, page 19). Notons que la mesure ν de (3.4) est nécessairement une mesure d'irréductibilité pour P . En effet, puisque P est φ -irréductible par hypothèse, il possède une mesure d'irréductibilité maximale, disons ψ . Puisque $\nu(A) > 0$ implique que $P^m(x, A) > 0$ pour tout $x \in C$ et que $\varphi(C) > 0$ implique que $\psi(C) > 0$, le sous-ensemble A satisfait nécessairement $\psi(A) > 0$ (revoir les propriétés d'une mesure d'irréductibilité maximale, page 21). Finalement, $\psi(A) > 0$ implique que pour tout $x \in E$, il existe un entier n_0 tel que $P^{n_0}(x, A) > 0$.

Pour que P soit géométriquement ergodique, il faut et il suffit que l'un des ensembles petits pour P satisfasse la condition suivante:

Théorème 3.1. Soit P un noyau markovien défini sur un espace mesurable (E, \mathcal{E}) . Supposons que P soit φ -irréductible pour une certaine mesure φ . Alors P est géométriquement ergodique si et seulement s'il existe un ensemble petit $C \in \mathcal{E}$, une fonction réelle $V : E \rightarrow [1, \infty)$ et deux constantes $a < 1$ et $b < \infty$ tels que

$$\int_E V(y)P(x, dy) \leq aV(x) + b\mathbf{1}_C(x) \quad \text{pour tout } x \in E. \quad (3.5)$$

Une fois qu'un noyau markovien P a été reconnu comme étant géométriquement ergodique, il est possible d'identifier une fonction M et un réel ρ tels que (3.3) est satisfaite (voir par exemple Meyn et Tweedie, 1994); en fait, ceux-ci peuvent être calculés à partir des réels a et b et de la fonction V de (3.5), ainsi que de l'entier m et du réel β qui font que (3.4) est satisfaite par l'ensemble petit C de (3.5). Toutefois, ces bornes géométriques sont assez complexes à déterminer; pour cette raison, nous n'en discuterons pas davantage dans ce mémoire. Une autre condition suffisante assurant l'ergodicité géométrique d'un noyau markovien P est énoncée dans Tierney (1994, proposition 1) et Chan (1993, théorème 2.1). Nous passons immédiatement au cas particulier où la fonction M de (3.3) peut être choisie bornée; le noyau markovien P est alors dit *uniformément ergodique* et le calcul de bornes supérieures pour $\|P^n(x, \cdot) - \pi(\cdot)\|$, $n \geq 1$, $x \in E$ se simplifie beaucoup.

Définition 3.4. Un noyau markovien ergodique P avec loi stationnaire π est dit *uniformément ergodique* s'il existe un $0 < M < \infty$ et un $0 < \rho < 1$ tels que

$$\|P^n(x, \cdot) - \pi(\cdot)\| \leq M\rho^n \quad (3.6)$$

pour tous $x \in E$ et $n \geq 1$.

Certains auteurs définissent un noyau markovien uniformément ergodique comme étant un noyau ergodique satisfaisant

$$\limsup_{n \rightarrow \infty} \sup_{x \in E} \|P^n(x, \cdot) - \pi(\cdot)\| = 0, \quad (3.7)$$

c'est-à-dire lorsque la convergence en variation totale de $P^n(x, \cdot)$ vers π est uniforme en x (voir par exemple Meyn et Tweedie, 1993, page 383 et Nummelin, 1984, page 122). Par contre, Meyn et Tweedie (1993, page 384) montrent que dans ce cas, la convergence se fait nécessairement à une vitesse géométrique. Les définitions d'ergodicité

uniforme au sens de (3.6) et (3.7) sont ainsi équivalentes. Le prochain théorème présente une condition nécessaire et suffisante pour qu'un noyau markovien soit uniformément ergodique. Il fournit aussi des valeurs possibles pour M et ρ dans (3.6).

Théorème 3.2. *Soit P un noyau markovien défini sur un espace mesurable (E, \mathcal{E}) . Alors P est uniformément ergodique si et seulement s'il existe une mesure de probabilité ν sur (E, \mathcal{E}) , un réel $\beta > 0$ et un entier $m \geq 1$ tels que*

$$P^m(x, A) \geq \beta\nu(A) \quad \text{pour tous } x \in E \text{ et } A \in \mathcal{E}. \quad (3.8)$$

Dans ce cas,

$$\|P^n(x, \cdot) - \pi(\cdot)\| \leq (1 - \beta)^{(n/m)} \quad \text{pour tout } n \geq 1,$$

où π est la loi stationnaire de P .

Une démonstration de ce théorème est disponible notamment dans Meyn et Tweedie (1993, page 384). La condition (3.8) est souvent appelée la *condition de Doeblin* (voir par exemple Doob, 1953, page 192 et Meyn et Tweedie, 1993, page 391). Elle est aussi équivalente à la condition suivante: l'espace E est petit pour P . Remarquons que contrairement à la condition (3.4), la mesure ν est automatiquement une mesure d'irréductibilité pour P , sans avoir à faire l'hypothèse que P est a priori φ -irréductible. De plus, si (3.8) est satisfaite avec $m \geq 1$, elle l'est aussi pour tout $n > m$ (avec le même β et la même mesure de probabilité ν) puisqu'alors, pour tout $k > 1$ et pour tout $A \in \mathcal{E}$,

$$P^{m+k}(x, A) = \int_E P^k(x, dy)P^m(y, A) \geq \beta \int_E P^k(x, dy)\nu(A) = \beta\nu(A).$$

Le théorème 3.2 fournit donc une façon de déterminer si un noyau markovien est uniformément ergodique, et si tel est le cas, de calculer des bornes supérieures pour $\|P^n(x, \cdot) - \pi(\cdot)\|$.

Pour conclure cette section, nous présentons ici trois exemples afin d'illustrer la théorie qui vient d'être exposée.

Exemple 3.1. Supposons que P soit un noyau markovien sur $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$, où \mathcal{X} est un borélien de \mathbb{R}^d , et que ν soit une mesure de probabilité sur $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$. Supposons aussi

que les mesures ν et $(P^n(x, \cdot); x \in \mathcal{X}, n \geq 1)$ soient toutes absolument continues par rapport à la mesure de Lebesgue μ sur $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$, avec densités h et $(p^n(x, \cdot); x \in \mathcal{X}, n \geq 1)$ respectivement. Par exemple, P pourrait être le noyau markovien de l'échantillonneur de Gibbs avec mises à jour séquentielles appliqué à une certaine mesure de probabilité π , ou le noyau de l'algorithme hit-and-run appliqué à π , lorsque la loi des directions est absolument continue par rapport à la mesure de Lebesgue sur (S_d, \mathcal{B}_{S_d}) . S'il existe un entier $m \geq 1$, un réel $\beta > 0$, et une densité h sur \mathcal{X} tels que

$$p^m(x, y) \geq \beta h(y) \quad \text{pour tous } x, y \in \mathcal{X}, \quad (3.9)$$

alors P est uniformément ergodique, puisque P satisfait la condition de Doeblin (3.8). Cette condition est aussi nécessaire: si P est uniformément ergodique, il existe un entier $m \geq 1$, un réel $\beta > 0$, une densité h sur \mathcal{X} et des versions des densités des mesures de probabilité $(P^m(x, \cdot); x \in \mathcal{X})$, disons $(p^m(x, \cdot); x \in \mathcal{X})$, tels que (3.9) est satisfaite.

Par exemple, considérons le noyau markovien P de l'algorithme hit-and-run appliqué à la loi uniforme π sur un sous-ensemble ouvert et borné \mathcal{X} de \mathbb{R}^d , avec ν , la loi uniforme sur (S_d, \mathcal{B}_{S_d}) , comme loi des directions. On peut alors montrer que pour tout $x \in \mathcal{X}$, la mesure de probabilité $P_{(\nu, \pi)}(x, \cdot)$ possède comme densité

$$p_{(\nu, \pi)}(x, y) = \frac{2}{C_d \|x - y\|^{d-1} M_{x,y}},$$

où $C_d = 2\pi^{d/2}/\Gamma(d/2)$ est l'aire de la surface de S_d et où $M_{x,y}$ est la mesure de Lebesgue unidimensionnelle de $\mathcal{X} \cap L_{x,y}$. Dénotons maintenant par $D_{\mathcal{X}}$ le diamètre de \mathcal{X} , c'est-à-dire $\sup_{u,v \in \mathcal{X}} \|u - v\|$. Puisque \mathcal{X} est borné, $D_{\mathcal{X}}$ est fini. Pour tous $x, y \in \mathcal{X}$, nous avons donc

$$p_{(\nu, \pi)}(x, y) \geq \frac{2}{C_d D_{\mathcal{X}}^d},$$

ou encore,

$$P_{(\nu, \pi)}(x, A) \geq \frac{2\mu(A)}{C_d D_{\mathcal{X}}^d} = \frac{2\pi(A)\mu(\mathcal{X})}{C_d D_{\mathcal{X}}^d} \quad \text{pour tout } A \in \mathcal{B}_{\mathcal{X}}.$$

Le noyau $P_{(\nu, \pi)}$ est ainsi uniformément ergodique, et il satisfait

$$\|P_{(\nu, \pi)}^n(x, \cdot) - \pi(\cdot)\| \leq \left(1 - \frac{2\mu(\mathcal{X})}{C_d D_{\mathcal{X}}^d}\right)^n$$

pour tous $n \geq 1$ et $x \in \mathcal{X}$. \square

Lorsque P est le noyau markovien d'un algorithme de simulation markovienne, les mesures de probabilité $(P^n(x, \cdot); x \in \mathcal{X}, n \geq 1)$ ne sont généralement pas toutes absolument continues par rapport à μ . C'est le cas entre autres de l'algorithme hit-and-run (lorsque la loi des directions n'est pas absolument continue par rapport à la mesure de Lebesgue sur (S_d, \mathcal{B}_{S_d})) et de l'algorithme de Hastings. Dans ce cas, considérons plutôt, pour tous $x \in \mathcal{X}$ et $n \geq 1$, la densité $p^n(x, \cdot)$ correspondant à la partie absolument continue de $P^n(x, \cdot)$ par rapport à μ . Si, pour un certain $m \geq 1$, un certain $\beta > 0$ et une certaine densité h sur \mathcal{X} , celle-ci satisfait (3.9), alors P est uniformément géométrique, puisque pour tout $A \in \mathcal{B}_{\mathcal{X}}$ et pour tout $x \in \mathcal{X}$, nous avons alors

$$P^m(x, A) \geq \int_A p^m(x, y) dy \geq \beta \int_A h(y) dy = \beta \nu(A).$$

Nous poursuivons avec un exemple de noyau markovien qui n'est pas uniformément ergodique, mais qui est tout de même géométriquement ergodique.

Exemple 3.2. Soit $E = \{0, 1, 2, \dots\}$ et \mathcal{E} , la famille de tous les sous-ensembles de E . Considérons le noyau markovien défini sur (E, \mathcal{E}) par

$$P(i, \{j\}) = \begin{cases} \frac{1}{3} & \text{si } j = i + 1, \\ \frac{2}{3} & \text{si } i \neq 0 \text{ et } j = i - 1, \\ \frac{2}{3} & \text{si } i = 0 \text{ et } j = 0. \end{cases}$$

Le noyau P est alors celui d'une marche aléatoire $(X_n; n \geq 0)$ sur E qui se déplace, à chaque transition, d'une unité vers la gauche avec probabilité $2/3$ et d'une unité vers la droite avec probabilité $1/3$; lorsque $(X_n; n \geq 0)$ est à l'état 0, elle y reste au temps suivant avec probabilité $2/3$ et se déplace à l'état 1 avec probabilité $1/3$.

Pour tous $i, j \in E$ distincts, la chaîne nécessite au minimum $|i - j|$ transitions afin de passer à l'état j en partant de i . Pour tout $0 \leq n < |i - j|$ le noyau P satisfait alors $P^n(i, \{j\}) = 0$. Il s'ensuit que P ne satisfait pas la condition de Doeblin; il n'est donc pas uniformément ergodique. En utilisant le résultat du théorème 3.1, nous allons cependant montrer que P est géométriquement ergodique.

Dénotons par ψ la mesure de dénombrement sur E ; celle-ci est une mesure d'irréductibilité maximale pour P . Puisque tout $A \in \mathcal{E}^+$ contient un ensemble petit pour P

et que $\psi(\{i\}) > 0$ pour tout $i \in E$, chaque état individuel $\{i\}$, $i \in E$, est un ensemble petit pour P . Considérons alors l'ensemble petit $\{0\}$ et tentons de trouver une fonction $V : E \rightarrow [1, \infty)$ et des constantes $a < 1$ et $b < \infty$ telles que (3.5) est satisfaite.

Soit $1 < \delta < 2$, et posons $V(i) = \delta^i$ pour tout $i \in E$. Nous avons alors, pour tout $i > 0$,

$$\int_E V(y)P(x, dy) = \sum_{j=0}^{\infty} \delta^j P(i, \{j\}) = \frac{2}{3}\delta^{i-1} + \frac{1}{3}\delta^{i+1}.$$

En posant $a = (2/3)\delta^{-1} + (1/3)\delta$, nous avons

$$\sum_{j=0}^{\infty} \delta^j P(i, \{j\}) = a\delta^i \quad \text{pour tout } i > 0.$$

En posant $b = (2/3)(1 - \delta^{-1})$, nous avons maintenant

$$\sum_{j=0}^{\infty} \delta^j P(i, \{j\}) = a\delta^i + b\mathbf{1}_{\{0\}}(i) \quad \text{pour tout } i \in E.$$

Puisque $a < 1$ et $b < \infty$, le noyau P satisfait la condition du théorème 3.1 et est ainsi géométriquement ergodique. On peut aussi montrer que sa loi stationnaire est la loi π définie sur (E, \mathcal{E}) par

$$\pi(\{i\}) = \left(\frac{1}{2}\right)^{i+1}, \quad i \in E,$$

c'est-à-dire la loi géométrique de paramètre $p = 1/2$. \square

Notre dernier exemple montre qu'un noyau markovien ergodique P défini sur un espace fini est toujours uniformément ergodique.

Exemple 3.3. Soit $E = \{1, \dots, K\}$, $K \in \mathbb{N}$ et \mathcal{E} la famille de tous les sous-ensembles de E . Dénotons par ν la loi uniforme sur (E, \mathcal{E}) . Supposons que P soit un noyau markovien ν -irréductible et apériodique, défini sur (E, \mathcal{E}) . Ces deux hypothèses et le fait que E soit fini implique qu'il existe un entier $k \geq 1$ tel que $P^k(x, A) > 0$ pour tout $x \in E$ et pour tout sous-ensemble non vide $A \in \mathcal{E}$. Posons

$$\delta = \min_{x, y \in E} P^k(x, \{y\}).$$

Nous avons alors, pour tous $x, y \in E$,

$$P^k(x, \{y\}) \geq K\delta\nu(\{y\}),$$

et donc, pour tout $A \in \mathcal{E}$,

$$P^k(x, A) \geq K\delta\nu(A).$$

Le noyau P satisfait ainsi la condition de Doeblin, et en vertu du théorème 3.2, nous avons

$$\|P^n(x, A) - \pi(A)\| \leq (1 - K\delta)^{n/k}$$

pour tous $n \geq 1$, $x \in E$ et $A \in \mathcal{E}$. \square

3.2 Comportement asymptotique des moyennes

3.2.1 La loi forte des grands nombres

Lorsque l'objectif est d'estimer certaines caractéristiques de la loi cible π , comme par exemple $\pi(g) = \int g d\pi$ où g est une fonction réelle mesurable, il est possible dans certains cas de simuler une seule chaîne de Markov $(X_n; n \geq 0)$ pour y arriver. En effet, sous des conditions très faibles, une loi forte des grands nombres assure que la moyenne $\bar{g}_n = (n + 1)^{-1} \sum_{i=0}^n g(X_i)$ converge presque sûrement vers $\pi(g)$. Ensuite, sous des conditions un peu plus fortes portant notamment sur la nature de la convergence en variation totale de $P^n(x, \cdot)$ vers la loi cible π , un théorème limite central soutient que la loi de \bar{g}_n (centrée et normalisée convenablement) est asymptotiquement normale. Lorsque ce résultat s'applique, il est alors possible de calculer des intervalles de confiance pour $\pi(g)$. Nous débutons par la présentation de la loi forte des grands nombres.

Théorème 3.3. *Soit P un noyau markovien récurrent au sens de Harris défini sur un espace mesurable (E, \mathcal{E}) , avec loi stationnaire π . Supposons que $(X_n; n \geq 0)$ soit une chaîne de Markov avec noyau markovien P et loi initiale λ . Alors pour toute fonction g satisfaisant $\pi(|g|) = \int |g| d\pi < \infty$,*

$$\lim_{n \rightarrow \infty} \bar{g}_n = \pi(g) \quad p.s. .$$

Dans le cas où P est seulement récurrent, le résultat est aussi valide, pourvu que $\lambda(\{x \in E : \mathbb{P}_x[T_A < \infty] = 1 \forall A \in \mathcal{E}^+\}) = 1$, où $T_A = \inf\{n \geq 1 : X_n \in A\}$.

Une démonstration de la loi forte des grands nombres est présentée notamment dans Athreya, Doss et Sethuraman (1996). Notons qu'il n'est pas nécessaire d'exiger l'apériodicité du noyau P pour qu'elle soit valide.

3.2.2 Le théorème limite central

Considérons un noyau markovien P , défini sur un espace mesurable (E, \mathcal{E}) avec loi stationnaire π , et une fonction mesurable $g : E \rightarrow \mathbb{R}$. Soit $(X_n; n \geq 0)$, une chaîne de Markov avec noyau markovien P et loi initiale π . La suite $(X_n; n \geq 0)$ est alors stationnaire, et en particulier, les variables X_0, X_1, \dots sont identiquement distribuées et ont pour loi π . Évidemment, sauf dans certains cas triviaux, les variables X_0, X_1, \dots ne sont pas indépendantes, et il est donc impossible d'utiliser le théorème limite central classique (i.e, pour les suites i.i.d.) pour conclure que la moyenne \bar{g}_n , lorsque centrée et normalisée convenablement, est asymptotiquement normale.

Lorsque P est ergodique, cependant, la dépendance entre X_0 et X_k , $k \geq 1$, est de moins en moins forte, à mesure que k augmente, puisque la loi asymptotique conditionnelle de X_k étant donné que $X_0 = x_0$ est toujours π peu importe $x_0 \in E$. Il en est de même pour la dépendance entre X_k et X_{n+k} , quel que soit $n \geq 0$, puisque la chaîne $(X_n; n \geq 0)$ est stationnaire. En imposant des conditions à P de telle sorte que cette dépendance diminue suffisamment rapidement, les variables aléatoires

$$\sum_{i=0}^k g(X_i), \quad \sum_{i=k+1}^{2k} g(X_i), \quad \sum_{i=2k+1}^{3k} g(X_i), \quad \dots$$

seront approximativement identiquement distribuées et indépendantes, si k est grand. En ce sens, il est raisonnable de croire que si la dépendance diminue assez rapidement, la loi asymptotique de \bar{g}_n sera normale. Si tel est le cas, il sera aussi raisonnable de croire que la normalité asymptotique de \bar{g}_n sera aussi valable lorsque la loi initiale de $(X_n; n \geq 0)$ est arbitraire, puisque P est récurrent au sens de Harris.

Supposons tout d'abord que P soit un noyau géométriquement ergodique, défini sur un espace mesurable (E, \mathcal{E}) , avec loi stationnaire π . Considérons une chaîne de Markov

$(X_n; n \geq 0)$ avec noyau markovien P et loi initiale π , et une fonction mesurable $g : E \rightarrow \mathbb{R}$. L'ergodicité géométrique de P implique qu'il existe une fonction M intégrable par rapport à π et un réel $0 < \rho < 1$ tels que pour tous A et $B \in \mathcal{E}$,

$$\alpha(n) = \sup_{A, B \in \mathcal{E}} \left| \int_B [P^n(x, A) - \pi(A)] \pi(dx) \right| \leq \pi(M) \rho^n. \quad (3.10)$$

Le réel $\alpha(n)$ s'appelle le *coefficient de mélange* α de X_0 et X_n ; il est une mesure de dépendance entre ces deux variables aléatoires (voir Robert, 1995, page 234). En général, une suite stationnaire de variables aléatoires $(Y_n; n \geq 0)$ à valeurs dans un espace mesurable (E, \mathcal{E}) satisfaisant $\lim_{n \rightarrow \infty} \alpha(n) = 0$, où

$$\alpha(n) = \sup_{A, B \in \mathcal{E}} |\mathbb{P}[Y_n \in A, Y_0 \in B] - \mathbb{P}[Y_n \in A] \mathbb{P}[Y_0 \in B]|, \quad (3.11)$$

est dite α -*mélangeante*. Dans le cas de notre chaîne $(X_n; n \geq 0)$, la quantité $\mathbb{P}[X_n \in A, X_0 \in B] - \mathbb{P}[X_n \in A] \mathbb{P}[X_0 \in B]$ peut aussi s'écrire $\int_B [P^n(x, A) - \pi(A)] \pi(dx)$ et l'ergodicité géométrique de P implique donc que $\alpha(n)$ converge vers 0 à une vitesse géométrique. Le résultat suivant nous permettra d'établir un théorème limite central pour \bar{g}_n .

Théorème 3.4. *Soit $(Y_n; n \geq 0)$ une suite stationnaire de variables aléatoires à valeurs dans un espace mesurable (E, \mathcal{E}) , chacune de ces variables ayant pour loi π . Supposons que $(Y_n; n \geq 0)$ soit α -mélangeante, c'est-à-dire que*

$$\alpha(n) = \sup_{A, B \in \mathcal{E}} |\mathbb{P}[Y_n \in A, Y_0 \in B] - \pi(A) \pi(B)|$$

converge vers 0 lorsque $n \rightarrow \infty$. Si $g : E \rightarrow \mathbb{R}$ est une fonction mesurable telle que

$$\pi(|g|^{2+\varepsilon}) < \infty \quad \text{pour un certain } \varepsilon > 0, \quad (3.12)$$

et que, pour ce même ε ,

$$\sum_{n=0}^{\infty} [\alpha(n)]^{\varepsilon/(2+\varepsilon)} < \infty, \quad (3.13)$$

alors

$$\sigma_g^2 = \text{Var}[g(Y_0)] + 2 \sum_{k=1}^{\infty} \text{Cov}[g(Y_0), g(Y_k)]$$

est absolument convergente et non négative. De plus,

$$\sqrt{n} [\bar{g}_n - \pi(g)] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_g^2),$$

où $\mathcal{N}(0, \sigma_g^2)$ représente la loi normale de moyenne 0 et de variance σ_g^2 .

Signalons que dans le cas où $\sigma_g^2 = 0$, le résultat est toujours valide pourvu que $\mathcal{N}(0, 0)$ soit interprétée comme étant la mesure de Dirac en 0. C'est ce que nous ferons dans tout le restant de cette section.

Le fait que le noyau P de notre chaîne stationnaire $(X_n; n \geq 0)$ soit géométriquement ergodique assure que la condition (3.13) est toujours respectée, quel que soit le $\varepsilon > 0$ choisi dans la condition (3.12). Remarquons que lorsque σ_g^2 est bien définie et finie, c'est-à-dire lorsque la série $\sum_{k=1}^{\infty} \text{Cov}[g(X_0), g(X_k)]$ converge absolument, alors la variance asymptotique de \bar{g}_n , c'est-à-dire

$$\begin{aligned} & \lim_{n \rightarrow \infty} n \text{Var}[\bar{g}_n] \\ &= \lim_{n \rightarrow \infty} \frac{n}{(n+1)^2} \sum_{k=0}^n \sum_{j=0}^n \text{Cov}[g(X_k), g(X_j)] \\ &= \lim_{n \rightarrow \infty} \frac{n}{(n+1)^2} \left[(n+1) \text{Var}[g(X_0)] + 2 \sum_{k=1}^n (n+1-k) \text{Cov}[g(X_0), g(X_k)] \right] \\ &= \lim_{n \rightarrow \infty} \frac{n}{(n+1)} \text{Var}[g(X_0)] + \lim_{n \rightarrow \infty} \frac{2n}{(n+1)} \sum_{k=1}^n \frac{n+1-k}{n+1} \text{Cov}[g(X_0), g(X_k)] \\ &= \text{Var}[g(X_0)] + 2 \sum_{k=0}^{\infty} \text{Cov}[g(X_0), g(X_k)] - 2 \lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{k}{(n+1)} \text{Cov}[g(X_0), g(X_k)] \end{aligned}$$

est aussi bien définie et finie, et sa valeur coïncide avec celle de σ_g^2 . Il est en effet possible de montrer, à l'aide de techniques d'analyse mathématique simples, que

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n |\text{Cov}[g(X_0), g(X_k)]| < \infty$$

implique que

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{k}{n+1} \text{Cov}[g(X_0), g(X_k)] = 0.$$

Nous allons maintenant voir que dans le cas d'un noyau uniformément ergodique, le ε de (3.12) peut être choisi nul, et il n'est pas nécessaire de vérifier une condition similaire à (3.13). Considérons donc le cas où $(X_n; n \geq 0)$ est une chaîne de Markov avec noyau markovien P uniformément ergodique, avec loi initiale π , cette dernière étant stationnaire pour P . Dans ce cas, il est aisé de constater que P et π satisfont

$$\varphi(n) = \sup_{A \in \mathcal{E}, B \in \mathcal{E}^+} \left| \pi(B)^{-1} \int_B \pi(dx) P^n(x, A) - \pi(A) \right| \leq M \rho^n$$

pour certains réels $M > 0$ et $0 < \rho < 1$. Notons que $\pi(B)^{-1} \int_B \pi(dx) P^n(x, A)$ n'est que la probabilité conditionnelle $\mathbb{P}[X_n \in A | X_0 \in B]$ lorsque $\pi(B) > 0$. Le réel $\varphi(n)$ s'appelle le *coefficient de mélange* φ et est une mesure de dépendance entre X_0 et X_n plus forte que $\alpha(n)$, en ce sens que si $\varphi(n) \rightarrow 0$, alors $\alpha(n) \rightarrow 0$ aussi. En général, une suite stationnaire de variables aléatoires $(Y_n; n \geq 0)$ à valeurs dans un espace mesurable (E, \mathcal{E}) satisfaisant $\lim_{n \rightarrow \infty} \varphi(n) = 0$, où

$$\varphi(n) = \sup_{A \in \mathcal{E}, B \in \mathcal{E}^+} |\mathbb{P}[Y_n \in A | Y_0 \in B] - \mathbb{P}[Y_n \in A]|$$

et où $\mathcal{E}^+ = \{A \in \mathcal{E} : \mathbb{P}[Y_0 \in A] > 0\}$, est dite φ -mélangeante (voir Robert, 1995, page 236). En vertu du prochain théorème, le fait que $(X_n; n \geq 0)$ soit φ -mélangeante est suffisant pour déduire la normalité asymptotique de $\sqrt{n}[\bar{g}_n - \pi(g)]$, lorsque g est une fonction réelle mesurable telle que $\int g^2 < \infty$.

Théorème 3.5. *Soit $(Y_n; n \geq 0)$ une suite stationnaire de variables aléatoires à valeurs dans un espace mesurable (E, \mathcal{E}) , chacune de ces variables ayant pour loi π . Supposons que $(Y_n; n \geq 0)$ soit φ -mélangeante, c'est-à-dire que*

$$\varphi(n) = \sup_{A \in \mathcal{E}, B \in \mathcal{E}^+} |\mathbb{P}[Y_n \in A | Y_0 \in B] - \pi(B)|$$

converge vers 0 lorsque $n \rightarrow \infty$. Si

$$\sum_{n=0}^{\infty} [\varphi(n)]^{1/2} < \infty, \quad (3.14)$$

alors pour toute fonction mesurable $g : E \rightarrow \mathbb{R}$ telle que $\pi(g^2) < \infty$, σ_g^2 est bien définie, non négative et finie. De plus,

$$\sqrt{n} [\bar{g}_n - \pi(g)] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_g^2). \quad (3.15)$$

Ce théorème est démontré notamment dans Billingsley (1968, page 184). Selon Tierney (1995, page 67), une chaîne de Markov stationnaire $(X_n; n \geq 0)$ avec noyau markovien P et loi initiale π est φ -mélangeante à une vitesse géométrique si et seulement si P est uniformément ergodique.

Attardons-nous maintenant au cas où $(X_n; n \geq 0)$ est une chaîne de Markov avec noyau markovien P , mais avec loi initiale différente de π . Selon la proposition 17.1.6 de

Meyn et Tweedie (1993, page 415), si une chaîne de Markov avec loi initiale λ et noyau markovien P récurrent au sens de Harris et avec loi stationnaire π , satisfait, pour une certaine fonction $g : E \rightarrow \mathbb{R}$,

$$\sqrt{n} [\bar{g}_n - \pi(g)] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_g^2), \quad (3.16)$$

alors toute chaîne $(X_n; n \geq 0)$ avec noyau markovien P et loi initiale arbitraire satisfait aussi (3.16), avec la même fonction g et le même σ_g^2 . Nous pouvons maintenant conclure:

Théorème 3.6. *Soit P un noyau markovien géométriquement ergodique, défini sur un espace mesurable (E, \mathcal{E}) . Dénotons par π son unique loi stationnaire. Soit $(X_n; n \geq 0)$ une chaîne de Markov avec noyau markovien P et loi initiale λ . Posons $\sigma_g^2 = \text{Var}_\pi[g(X_0)] + 2 \sum_{k=1}^{\infty} \text{Cov}_\pi[g(X_0), g(X_k)]$, où l'indice π de Var_π et de Cov_π indique que la variance et les covariances sont calculées sous l'hypothèse que $\lambda = \pi$. Alors, pour toute fonction réelle mesurable g telle que $\int |g|^{2+\varepsilon} d\pi < \infty$ pour un certain $\varepsilon > 0$, σ_g^2 est bien définie et $0 \leq \sigma_g^2 < \infty$. De plus,*

$$\sqrt{n} [\bar{g}_n - \pi(g)] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_g^2) \quad (3.17)$$

lorsque $n \rightarrow \infty$. Si P est uniformément ergodique, le résultat est aussi valide pour toute fonction g satisfaisant $\pi(g^2) < \infty$.

Mentionnons que le théorème limite central peut être valide sous d'autres conditions que celles du théorème 3.6. Par exemple, Roberts et Rosenthal (1998a) montrent que dans le cas où P est géométriquement ergodique, (3.17) est valide pour toute fonction g telle que $\pi(g^2) < \infty$, pourvu qu'on ajoute l'hypothèse que P est réversible par rapport à π . Cette hypothèse est satisfaite par le noyau markovien de plusieurs algorithmes de simulation markovienne, comme par exemple l'échantillonneur de Gibbs avec mises à jour aléatoires, l'algorithme de Hastings et l'algorithme hit-and-run. Elle n'est cependant pas toujours satisfaite pour l'échantillonneur de Gibbs avec mises à jour séquentielles. Aussi, Kipnis et Varadhan (1986) montrent que sous la seule hypothèse que P est φ -irréductible pour une certaine mesure φ , apériodique et réversible par rapport à une mesure de probabilité, disons π , alors (3.17) est valide pourvu que $\pi(g^2) < \infty$ et que σ_g^2 soit bien définie et finie. Par contre, cette dernière condition sur

σ_g^2 est généralement difficile à vérifier; lorsque P est géométriquement ergodique, elle est automatiquement satisfaite. Pour d'autres conditions sous lesquelles le théorème limite central est valide, on peut consulter entre autres Nummelin (1984) et Tierney (1994, 1995).

3.3 Vitesse de convergence de l'algorithme hit-and-run

Nous discutons ici de la vitesse de convergence et de l'ergodicité géométrique de l'algorithme hit-and-run. Nous reprenons donc la notation de la section 2.3: la loi cible π est une mesure de probabilité sur $(\mathbb{R}^d, \mathcal{B}^d)$, absolument continue par rapport à la mesure de Lebesgue μ . La mesure de probabilité sur (S_d, \mathcal{B}_{S_d}) servant à choisir les directions avant chaque itération est notée ν . Finalement, nous supposons toujours (sans le mentionner) que la version de la densité de π utilisée pour définir le noyau markovien $P_{(\nu, \pi)}$ de l'algorithme hit-and-run associé à π et ν satisfait la condition HR.

3.3.1 Convergence lente de l'algorithme hit-and-run

Lorsqu'aucune condition n'est imposée à la loi cible π ou à son support, la vitesse de convergence de $P_{(\nu, \pi)}^n(x, \cdot)$ vers π peut être extrêmement lente. Nous verrons ici que pour toute mesure de probabilité ν à pleine dimension, il est toujours possible de construire une mesure de probabilité π sur $(\mathbb{R}^d, \mathcal{B}^d)$ telle que le noyau markovien $P_{(\nu, \pi)}$ associé à π et ν est ergodique, mais aussi telle que la convergence de $\|P_{(\nu, \pi)}^n(x, \cdot) - \pi(\cdot)\|$ vers 0 est aussi lente que désirée, et ceci, quel que soit $x \in \mathcal{X}$. Le théorème suivant est démontré dans Bélisle (1998a, 1998b).

Théorème 3.7. *Soit ν une mesure de probabilité à pleine dimension sur (S_d, \mathcal{B}_{S_d}) . Soit $(b_n; n \geq 1)$ une suite de nombres réels positifs telle que $b_n \rightarrow 0$ lorsque $n \rightarrow \infty$. Alors il existe une mesure de probabilité π sur $(\mathbb{R}^d, \mathcal{B}^d)$ telle que le noyau de l'algorithme*

hit-and-run associé à ν et π sur $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ satisfait

$$\lim_{n \rightarrow \infty} \sup_{A \in \mathcal{B}_{\mathcal{X}}} |P_{(\nu, \pi)}^n(x, A) - \pi(A)| = 0 \quad \text{pour tout } x \in \mathcal{X}, \quad (3.18)$$

mais pour lequel il existe un entier $n_0 > 1$ tel que

$$\sup_{A \in \mathcal{B}_{\mathcal{X}}} |P_{(\nu, \pi)}^n(x, A) - \pi(A)| > b_n \quad \text{pour tous } x \in \mathcal{X} \text{ et } n > n_0. \quad (3.19)$$

Étant donnée une suite de nombres réels $(b_n; n \geq 1)$ telle que $b_n \rightarrow 0$, une loi “pathologique” π satisfaisant (3.18) et (3.19) n’est pas nécessairement complexe; cela dépend de la nature de la mesure de probabilité ν . Par exemple, considérons le cas où la mesure de probabilité ν est à pleine dimension et possède la propriété suivante: il existe un sous-ensemble ouvert Γ de S_d satisfaisant $\nu(\Gamma \cup (-\Gamma)) = 0$, où $-\Gamma = \{\theta \in S_d : -\theta \in \Gamma\}$. On dit d’une telle loi ν que son support possède un *trou symétrique*. Bélisle (1998a) montre que si ν est une loi à pleine dimension dont le support possède un trou symétrique, alors la mesure de probabilité π du théorème 3.7 peut toujours être choisie comme étant la loi uniforme sur un sous-ensemble ouvert, borné et connexe de \mathbb{R}^d . La loi uniforme sur les vecteurs de la base canonique $\{e_1, \dots, e_d\}$ étant un exemple de loi à pleine dimension sur (S_d, \mathcal{B}_{S_d}) dont le support possède un trou symétrique, ce résultat s’applique notamment à l’échantillonneur de Gibbs avec mises à jour aléatoires. Selon Bélisle (1998a), il s’applique également à l’échantillonneur de Gibbs avec mises à jour séquentielles.

Lorsque ν est une mesure de probabilité à pleine dimension arbitraire, la mesure de probabilité pathologique π ne peut pas toujours être choisie comme la loi uniforme sur un sous-ensemble ouvert et borné. Par exemple, considérons le cas où la partie absolument continue ν_{abs} de ν par rapport à la mesure de Lebesgue sur (S_d, \mathcal{B}_{S_d}) est telle qu’il existe une version de la densité de ν_{abs} , disons h , satisfaisant, pour un certain $\varepsilon > 0$,

$$h(\theta) \geq \varepsilon \quad \text{pour tout } \theta \in S_d.$$

Nous verrons à la section 3.3.2 qu’avec une telle loi ν , pour tout ensemble ouvert et borné $\mathcal{X} \subset \mathbb{R}^d$, le noyau $P_{(\nu, \pi)}$ de l’algorithme hit-and-run associé à ν et la loi uniforme sur \mathcal{X} est uniformément ergodique. En d’autres termes, nous verrons que pour toute suite de nombres réels $(b_n; n \geq 1)$ telle que $b_n \rightarrow 0$ à une vitesse non géométrique,

comme par exemple $b_n = \alpha/n$ où $\alpha \in \mathbb{R}$, il n'existe pas de loi uniforme π sur un sous-ensemble ouvert, borné et connexe telle que (3.18) et (3.19) sont satisfaites. Néanmoins, Bélisle (1998b) montre que la mesure de probabilité pathologique π peut du moins être construite de façon à ce qu'elle possède une densité f telle que l'ensemble \mathcal{X} de la condition HR est ouvert, borné et convexe.

Le théorème 3.7 permet donc de conclure que sans hypothèse supplémentaire sur la loi cible π ou sur son support, la convergence de l'algorithme hit-and-run vers π peut être arbitrairement lente. En particulier, il permet de conclure que lorsqu'elle a lieu, la convergence ne se fait pas toujours à une vitesse géométrique.

3.3.2 Ergodicité géométrique de l'algorithme hit-and-run

Nous sommes maintenant en mesure d'énumérer plusieurs raisons qui font que l'ergodicité géométrique est une propriété intéressante et souhaitable pour le noyau markovien P d'un algorithme de simulation markovienne. D'abord, elle assure que la vitesse de convergence de $\|P^n(x, \cdot) - \pi(\cdot)\|$ vers 0 n'est pas arbitrairement lente, au sens de la section précédente. Ensuite, elle permet d'évaluer cette vitesse de convergence par l'entremise du calcul de bornes supérieures pour $\|P^n(x, \cdot) - \pi(\cdot)\|$. Finalement, elle est accompagnée d'un théorème limite central, ce qui permet entre autres de calculer des intervalles de confiance pour certaines caractéristiques d'intérêt de π .

Les conditions générales assurant l'ergodicité géométrique d'un noyau markovien, comme par exemple la condition de Doeblin et la condition du théorème 3.1, sont habituellement difficiles à vérifier en pratique. C'est pourquoi plusieurs auteurs ont tenté d'énoncer des conditions suffisantes, souvent en termes de π et de son support, pour que les différents algorithmes de simulation markovienne soient géométriquement ergodiques, ou plus particulièrement, uniformément ergodiques. Dans les cas de l'échantillonneur de Gibbs et de l'algorithme hit-and-run, on retrouve notamment Schervish et Carlin (1992), Chan (1993), Chen et Schmeiser (1993), Roberts et Polson (1994), Liu, Wong et Kong (1995), Roberts et Rosenthal (1998b) et Diaconis et

Freedman (1998). Dans le cas de l'algorithme de Hastings, citons entre autres Tierney (1994), Roberts et Tweedie (1996) et Mergensen et Tweedie (1996). Bien que ces conditions soient parfois assez simples à vérifier, elles sont généralement très restrictives, et plusieurs lois π que l'on retrouve dans les applications ne les satisfont pas. Nous présentons ici une condition suffisante pour l'ergodicité uniforme de l'algorithme hit-and-run dans les deux cas suivants:

- 1) ν est la loi uniforme sur (S_d, \mathcal{B}_{S_d}) ;
- 2) ν est la loi uniforme sur $\{e_1, \dots, e_d\}$ (le noyau $P_{(\nu, \pi)}$ est alors celui de l'échantillonneur de Gibbs avec mises à jour aléatoires).

3.3.2.1 Le cas où ν est la loi uniforme sur (S_d, \mathcal{B}_{S_d})

Le théorème suivant fut d'abord démontré par Smith (1984) dans le cas particulier où π est la loi uniforme sur un sous-ensemble ouvert et borné de \mathbb{R}^d . La version présentée ici requiert que le support de π soit borné et que la densité f de π satisfasse $\sup_{\theta \in S_d, x \in \mathcal{X}} \int_{\mathbb{R}} f(x + r\theta) dr < \infty$; cette dernière condition est satisfaite si, entre autres, f est bornée.

Théorème 3.8. *Soit π , une mesure de probabilité sur $(\mathbb{R}^d, \mathcal{B}^d)$, absolument continue par rapport à la mesure de Lebesgue et avec densité f . Soit ν , la loi uniforme sur (S_d, \mathcal{B}_{S_d}) . Supposons que l'ensemble \mathcal{X} (de la condition HR) soit borné, et que*

$$\gamma = \sup_{\theta \in S_d, x \in \mathcal{X}} \int_{\mathbb{R}} f(x + r\theta) dr < \infty.$$

Alors le noyau markovien $P_{(\nu, \pi)}$ de l'algorithme hit-and-run associé à π et ν satisfait la condition de Doeblin. Plus précisément,

$$P_{(\nu, \pi)}(x, A) \geq \frac{2}{C_d D_{\mathcal{X}}^{d-1} \gamma} \pi(A) \quad \text{pour tout } x \in \mathcal{X}, \quad (3.20)$$

où $C_d = 2\pi^{d/2}/\Gamma(d/2)$ est l'aire de la surface de la sphère S_d et où $D_{\mathcal{X}} = \sup_{u, v \in \mathcal{X}} \|u - v\|$ est le diamètre de \mathcal{X} .

Démonstration. Nous allons montrer que les densités des mesures de probabilité $(P_{(\nu, \pi)}(x, \cdot); x \in \mathcal{X})$ satisfont la condition présentée à l'exemple 3.1. Rappelons que le noyau markovien $P_{(\nu, \pi)}$ peut s'écrire (voir section 2.3.2)

$$P_{(\nu, \pi)}(x, A) = \int_{S_d} \nu(d\theta) \int_{\mathbb{R}} \mathbf{1}_A(x + \lambda\theta) \frac{f(x + \lambda\theta)}{\int_{\mathbb{R}} f(x + r\theta) dr} d\lambda, \quad x \in \mathcal{X}, A \in \mathcal{B}_{\mathcal{X}}.$$

Puisque dans notre cas, la mesure ν est la loi uniforme sur (S_d, \mathcal{B}_{S_d}) , la fonction $h(\theta) = 1/C_d$ est une version de la densité de ν . Pour tous $x \in \mathcal{X}$ et $A \in \mathcal{B}_{\mathcal{X}}$, la probabilité $P_{(\nu, \pi)}(x, A)$ peut alors s'écrire

$$P_{(\nu, \pi)}(x, A) = \frac{1}{C_d} \int_{S_d} \nu_0(d\theta) \int_{\mathbb{R}} \mathbf{1}_A(x + \lambda\theta) \frac{f(x + \lambda\theta)}{\int_{\mathbb{R}} f(x + r\theta) dr} d\lambda,$$

ou encore,

$$P_{(\nu, \pi)}(x, A) = \frac{2}{C_d} \int_{S_d} \nu_0(d\theta) \int_0^\infty \mathbf{1}_A(x + \lambda\theta) \frac{f(x + \lambda\theta)}{\int_{\mathbb{R}} f(x + r\theta) dr} d\lambda,$$

où ν_0 représente la mesure de Lebesgue $(d-1)$ -dimensionnelle sur S_d . Finalement, en effectuant le changement de variable $dy = \lambda^{d-1} d\lambda \nu_0(d\theta)$ (pour passer des coordonnées polaires aux coordonnées rectangulaires), nous obtenons

$$P_{(\nu, \pi)}(x, A) = \frac{2}{C_d} \int_A \frac{f(y)}{\|y - x\|^{d-1} M_{x,y}} dy,$$

où $M_{x,y}$ est l'intégrale de f le long de $L_{x,y}$, la droite passant par x et y . Le résultat (3.20) est ensuite immédiat puisque $D_{\mathcal{X}} \geq \|x - y\|$ et $\gamma \geq M_{x,y}$ pour tous $x, y \in \mathcal{X}$ implique que

$$\frac{2f(y)}{C_d \|y - x\|^{d-1} M_{x,y}} \geq \frac{2f(y)}{C_d D_{\mathcal{X}}^{d-1} \gamma}$$

pour tous $x, y \in \mathcal{X}$. ■

Le théorème 3.8 est aussi valide dans le cas plus général où la loi ν possède une décomposition de Lebesgue

$$\nu(\Gamma) = \nu_{\text{sing}}(\Gamma) + \int_{\Gamma} \nu_0(d\theta) h(\theta) \quad \Gamma \in \mathcal{B}_{S_d},$$

avec une densité h satisfaisant, pour un certain $\varepsilon > 0$, $h(\theta) \geq \varepsilon$ pour tout $\theta \in S_d$. En effet, le noyau $P_{(\nu, \pi)}$ peut alors s'écrire

$$P_{(\nu, \pi)}(x, A) = \int_{S_d} P_{(\theta, \pi)}(x, A) \nu_{\text{sing}}(d\theta) + \int_{S_d} P_{(\theta, \pi)}(x, A) h(\theta) d\theta, \quad (3.21)$$

et en reprenant les étapes de la démonstration du théorème 3.8 avec la partie de (3.21) faisant intervenir la densité h , nous obtenons

$$P_{(\nu, \pi)}(x, A) \geq 2\varepsilon \int_A \frac{f(y)}{\|y - x\|^{d-1} M_{x,y}} dy,$$

et donc

$$P_{(\nu, \pi)}(x, A) \geq \frac{2\varepsilon}{D_{\mathcal{X}}^{d-1} \gamma} \pi(A).$$

3.3.2.2 Le cas où ν est la loi uniforme sur $\{e_1, \dots, e_d\}$

Les conditions du théorème 3.8 ne sont pas suffisantes pour assurer l'ergodicité uniforme du noyau markovien de l'échantillonneur de Gibbs avec mises à jour aléatoires. En effet, nous avons vu à la section 3.3.1 que lorsque ν est la loi uniforme sur (S_d, \mathcal{B}_{S_d}) , il existe toujours un sous-ensemble ouvert, borné et connexe \mathcal{X} tel que l'algorithme hit-and-run associé à ν et la loi uniforme sur \mathcal{X} n'est pas géométriquement ergodique. En plus des conditions du théorème 3.8, il faut ajouter une hypothèse sur la forme géométrique de \mathcal{X} en vue d'obtenir l'ergodicité géométrique.

Dans Roberts et Rosenthal (1998b), plusieurs conditions suffisantes assurant l'ergodicité géométrique ou uniforme de l'échantillonneur de Gibbs (avec mises à jour aléatoires) sont présentées, dans le cas où la loi cible π est la loi uniforme sur un sous-ensemble ouvert, connexe et borné \mathcal{X} de \mathbb{R}^d . L'une d'elles, qui assure en fait l'ergodicité uniforme, est ici adaptée au cas où la loi cible π n'est pas nécessairement une loi uniforme.

Dans ce qui suit, pour tout sous-ensemble ouvert A de \mathbb{R}^d , nous dénoterons par \overline{A} la fermeture de A , c'est-à-dire le plus petit sous-ensemble fermé de \mathbb{R}^d contenant A . Comme auparavant, $B(x, r)$ dénote la boule ouverte de \mathbb{R}^d de rayon r centrée à x .

Théorème 3.9. *Soit π une mesure de probabilité sur $(\mathbb{R}^d, \mathcal{B}^d)$, absolument continue par rapport à la mesure de Lebesgue. Soit f , une densité pour π , et soit ν , la loi uniforme sur $\{e_1, \dots, e_d\}$. Supposons que l'ensemble ouvert \mathcal{X} (de la condition HR) soit connexe et borné, et que la densité f satisfasse, pour tout $x \in \mathcal{X}$,*

$$f(x) \geq \varepsilon \quad \text{pour un certain } \varepsilon > 0 \tag{3.22}$$

et

$$\gamma = \sup_{i \in \{1, \dots, d\}, x \in \mathcal{X}} \int_{\mathbb{R}} f(x + re_i) dr < \infty. \quad (3.23)$$

Posons maintenant, pour tout $x \in \overline{\mathcal{X}}$,

$$\eta(x) = \sup\{r > 0 : x \in \overline{B(y, r)} \text{ pour un certain } y \in \mathcal{X} \text{ tel que } B(y, r) \subseteq \mathcal{X}\}.$$

Supposons finalement que l'ensemble \mathcal{X} satisfasse la condition suivante:

$$\text{il existe un } a > 0 \text{ tel que } \eta(x) > a \text{ pour tout } x \in \overline{\mathcal{X}}. \quad (3.24)$$

Alors le noyau markovien $P_{(\nu, \pi)}$ de l'algorithme hit-and-run associé à π et ν est uniformément ergodique.

Démonstration. Puisque $\gamma \geq \int_{\mathbb{R}} f(x + re_i) dr$ pour tous $x \in \mathcal{X}$ et $i \in \{1, \dots, d\}$, et puisque $f(x) > \varepsilon > 0$ pour tout $x \in \mathcal{X}$, le noyau markovien $P_{(\nu, \pi)}$ satisfait

$$\begin{aligned} P_{(\nu, \pi)}(x, A) &= \frac{1}{d} \sum_{i=1}^d \int_{\mathbb{R}} \mathbf{1}_A(x + \lambda e_i) \frac{f(x + \lambda e_i)}{\int_{\mathbb{R}} f(x + re_i) dr} d\lambda \\ &\geq \frac{\varepsilon}{\gamma d} \sum_{i=1}^d \int_{\mathbb{R}} \mathbf{1}_A(x + \lambda e_i) d\lambda \end{aligned}$$

pour tous $x \in \mathcal{X}$ et $A \in \mathcal{B}_{\mathcal{X}}$. Remarquons que le terme de droite de la dernière inégalité est, à une constante multiplicative près, le noyau de l'échantillonneur de Gibbs (avec mises à jour aléatoires) appliqué à la loi uniforme sur $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$. Pour éviter tout risque d'ambiguïté, dénotons celui-ci par K . Selon Roberts et Rosenthal (1998b, théorème 7), la condition (3.24) implique que K est uniformément ergodique; il existe donc un réel $\beta > 0$, un entier $m > 1$ et une mesure de probabilité ψ sur $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ tels que

$$K^m(x, A) > \beta \psi(A) \quad \text{pour tous } x \in \mathcal{X} \text{ et } A \in \mathcal{B}_{\mathcal{X}}. \quad (3.25)$$

Maintenant, (3.25) et le fait que $P_{(\nu, \pi)}(x, A) \geq \alpha K(x, A)$ pour un certain $0 < \alpha < 1$ impliquent que

$$P_{(\nu, \pi)}^m(x, A) \geq \alpha^m K^m(x, A) \geq \alpha^m \beta \psi(A),$$

et par conséquent, le noyau $P_{(\nu, \pi)}$ satisfait la condition de Doeblin. En vertu du théorème 3.2, le noyau $P_{(\nu, \pi)}$ est uniformément ergodique. ■

Intuitivement, la condition (3.24) signifie qu'il existe un $a > 0$ tel que si nous faisons "rouler" une boule de rayon a partout dans \mathcal{X} , alors celle-ci toucherait à tous les points de $\overline{\mathcal{X}}$. Cette condition, qui assure que la frontière de \mathcal{X} est "lisse", n'est pas du tout nécessaire pour que $P_{(\nu, \pi)}$ soit uniformément ergodique. Par exemple, un rectangle ouvert et borné de \mathbb{R}^d ne satisfait pas la condition (3.24); pourtant, il n'est pas difficile de montrer que si π est une mesure de probabilité sur ce rectangle dont une version de la densité satisfait (3.22) et (3.23), alors le noyau $P_{(\nu, \pi)}$ associé à π est uniformément ergodique. À l'aide du lemme 1 et de la proposition 2 de Roberts et Rosenthal (1998b), nous pouvons aussi déduire que si \mathcal{X} est une union finie de rectangles ouverts de \mathbb{R}^d et que les composantes connexes de \mathcal{X} communiquent, alors le noyau markovien $P_{(\nu, \pi)}$ est uniformément ergodique. Roberts et Rosenthal (1998b) présentent aussi des exemples de sous-ensembles ouverts, bornés et connexes de \mathbb{R}^d ne satisfaisant pas la condition (3.24) tels que les noyaux de l'échantillonneur de Gibbs avec mises à jour aléatoires appliqués à la loi uniforme sur ces sous-ensembles (3.24) sont géométriquement ergodiques sans être uniformément ergodiques.

Remarque 3.1. Le théorème 3.9 s'applique aussi dans le cas de l'échantillonneur de Gibbs avec mises à jour séquentielles, à condition d'ajouter l'hypothèse que \mathcal{X} est *convexe*. Voir Roberts et Rosenthal (1998b) pour plus de détails.

Remarque 3.2. Dans ce mémoire, nous nous sommes surtout intéressés aux résultats théoriques associés à la convergence des algorithmes de simulation markovienne. Compte tenu de l'algorithme et de la complexité de la loi cible, les conditions assurant la convergence et la convergence géométrique, par exemple, peuvent être très difficile à vérifier en pratique. C'est pourquoi plusieurs auteurs ont développé des méthodes statistiques permettant, dans certains cas, de détecter la convergence d'un algorithme lors de son utilisation. Pour une discussion sur ce sujet, et pour un résumé des principales considérations pratiques de la simulation markovienne, on peut consulter entre autres Tierney (1994), Roberts et Rosenthal (1998a) et Brooks (1998).

Conclusion

Les algorithmes de simulation markovienne, tels l'échantillonneur de Gibbs, l'algorithme de Metropolis-Hastings et l'algorithme hit-and-run sont devenus des outils de simulation de variables aléatoires multidimensionnelles très populaires depuis l'article de Gelfand et Smith (1990). Étant donnée une loi cible π (généralement définie sur \mathbb{R}^d), ils permettent de simuler une chaîne de Markov $(X_n; n \geq 0)$ dont le noyau markovien P possède, par construction, la mesure de probabilité π comme loi stationnaire. Sous certaines conditions, souvent exprimables en termes du support de π , la loi de X_n converge en variation totale vers π , rendant alors possible la simulation d'une variable aléatoire de loi π . Cette dernière peut être alors obtenue simplement en simulant la chaîne $(X_n; n \geq 0)$ suffisamment longtemps.

Une des raisons pour lesquelles ces algorithmes suscitent beaucoup d'intérêt est qu'il est souvent plus facile de les mettre en application que de construire un algorithme, ou d'appliquer un algorithme déjà existant, permettant de simuler directement une variable aléatoire de loi π . C'est le cas entre autres lorsque d est grand, lorsque le support de π est un ensemble compliqué, ou lorsque π est connue seulement à une constante multiplicative près et que la constante de normalisation est difficile ou impossible à calculer. Par exemple, l'échantillonneur de Gibbs fait intervenir la simulation d'une série de variables aléatoires unidimensionnelles plutôt que multidimensionnelles; les difficultés qu'apporte la simulation de variables aléatoires multidimensionnelles sont donc éliminées. De plus, il ne requiert que la connaissance de versions des lois conditionnelles de X^i étant donné $(X^1, \dots, X^{i-1}, X^{i+1}, \dots, X^d)$, où (X^1, \dots, X^d) représente une variable aléatoire hypothétique de loi π . Dans le cas où π possède une densité par

rapport à la mesure de Lebesgue, disons f , il suffit donc de pouvoir calculer, pour tout $x \in \mathbb{R}^d$ et pour tout $i \in \{1, \dots, d\}$,

$$\int_{\mathbb{R}} f(x^1, \dots, x^{i-1}, u, x^{i+1}, \dots, x^d) du, \quad (4.26)$$

afin d'obtenir ces densités conditionnelles. Un autre avantage de l'algorithme est que le calcul de (4.26) fournit aussi les densités conditionnelles de π si f est en fait une version de la densité de π à une constante multiplicative près; la constante de normalisation de f n'est donc nullement nécessaire. Quant à lui, l'algorithme de Hastings permet de simuler une chaîne de Markov dont le noyau possède π comme loi stationnaire à partir d'un noyau markovien Q essentiellement arbitraire; il permet de connaître π à une constante multiplicative près, et il nous offre aussi la possibilité de simuler des variables aléatoires unidimensionnelles plutôt que multidimensionnelles, par l'entremise de sa version avec mises à jour composante par composante. La simulation markovienne est donc une alternative à la simulation de Monte-Carlo classique, lorsque celle-ci présente trop de difficultés.

Pour chacun de ces algorithmes, nous avons énoncé dans ce mémoire des conditions sous lesquelles il permet effectivement de simuler une variable aléatoire de loi π . Ces conditions consistaient principalement à assurer que le noyau markovien P de la chaîne de Markov $(X_n; n \geq 0)$ qu'il définit soit apériodique et récurrent au sens de Harris (c'est-à-dire ergodique). En vertu du théorème d'Orey, ces deux propriétés sont nécessaires et suffisantes pour obtenir la convergence en variation totale de la loi de X_n vers π , quelle que soit la loi initiale de la chaîne. En vertu de la loi forte des grands nombres, l'ergodicité implique aussi la convergence presque sûre de la moyenne ergodique \bar{g}_n vers $\pi(g) = \int g d\pi$, où g est une fonction mesurable et intégrable, rendant ainsi possible l'estimation de $\pi(g)$ à l'aide de la simulation d'une seule longue chaîne.

Un autre aspect important dont il a été question est la vitesse de la convergence de $\|P^n(x, \cdot) - \pi(\cdot)\|$, $x \in \mathbb{R}^d$, vers 0, lorsque celle-ci a lieu. Il est en effet primordial de pouvoir l'évaluer, si l'on veut déterminer le nombre de variables de $(X_n; n \geq 0)$ qu'il faut simuler pour se rapprocher suffisamment de la loi π . En général, il n'existe pas de résultat permettant d'évaluer la vitesse de convergence. Par contre, lorsque le noyau P est géométriquement ergodique en plus d'être ergodique, il est possible de calculer, pour tout $n \geq 1$, une borne supérieure pour $\|P^n(x, \cdot) - \pi(\cdot)\|$, $x \in \mathbb{R}^d$. L'ergodicité

géométrique est aussi une propriété intéressante pour plusieurs autres raisons; elle assure que la convergence se fait à une vitesse géométrique (donc qu'elle ne peut être arbitrairement lente), et elle implique l'existence d'un théorème limite central pour la moyenne ergodique \bar{g}_n , rendant alors possible le calcul d'intervalles de confiance pour $\pi(g)$.

Bibliographie

- Athreya, K.B., Doss, H. et Sethuraman, J. (1996). *On the convergence of the Markov chain simulation method. Ann. Statist.*, 24, 69-100.
- Bartle, R.G. (1995). *The Elements of Integration and Lebesgue Measure*. John Wiley & Sons, New York.
- Baxter, J.R. et Rosenthal, J.S. (1995). *Rates of convergence for everywhere-positive Markov chains. Statist. Probab. Lett.*, 22, 333-338.
- Bélisle, C. (1998a). *Slow convergence of the Gibbs sampler. Rev. Canad. Statist.*, 26, 629-641.
- Bélisle, C. (1998b). *Slow Hit-and-Run sampling*. Rapport technique, Département de mathématiques et de statistique, Université Laval. À paraître dans *Statistics and Probability Letters*.
- Bélisle, C., Boneh, A. et Caron, R.J. (1998). *Convergence properties of Hit-and-Run samplers. Stochastic Models*, 14, 767-800.
- Bélisle, C., Romeijn, H.E. et Smith, R.L. (1993). *Hit-and-Run algorithms for generating multivariate distributions. Math. Oper. Res.*, 18, 255-266.
- Berbee, H.C.P., Boender, C.G.E., Rinnooy Kan, A.H.G., Scheffer, C.L., Smith, R.L. et Telgen, J. (1987). *Hit-and-Run algorithms for the identification of nonredundant linear inequalities. Math. Programming*, 37, 184-207.

- Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley & Sons, New York.
- Billingsley, P. (1995). *Probability and Measure*. Third Edition. John Wiley & Sons, New York.
- Boneh, A. et Golan, A. (1979). *Constraints' redundancy and feasible region boundness by random feasible point generator (RFPG)*. Third European Congress on Operation Research, EURO III, Amsterdam.
- Breiman, L. (1968). *Probability*. Addison-Wesley, Reading.
- Brooks, S.P. (1998). *Markov chain Monte Carlo method and its application*. *The Statistician*, 47, 69-100.
- Casella, G. et George, E. (1992). *Explaining the Gibbs sampler*. *Amer. Statist.*, 46, 167-174.
- Chan, K.S. (1993). *Asymptotic behavior of the Gibbs sampler*. *J. Amer. Statist. Assoc.*, 88, 320-326.
- Chan, K.S. et Geyer, C.J. (1994). *Discussion on Markov chains for exploring posterior distributions by L. Tierney*. *Ann. Statist.*, 22, 1747-1758.
- Chen, M.-H. et Schmeiser, B.W. (1993). *Performance of the Gibbs, hit-and-run, and Metropolis samplers*. *J. Comput. Graph. Statist.*, 2, 251-272.
- Chib, S. et Greenberg, E. (1995). *Understanding the Metropolis-Hastings algorithm*. *Amer. Statist.*, 49, 327-335.
- Chung, K.L. (1967). *Markov Chains with Stationary Transition Probabilities*. Second Edition. Springer-Verlag, Berlin.
- Chung, K.L. (1974). *A Course in Probability Theory*. Second Edition. Academic Press, New York.

- Çinlar, E. (1975). *Introduction to Stochastic Processes*. Prentice-Hall, Englewood Cliffs.
- Diaconis, P. et Freedman, D. (1997). *On Markov chains with continuous state space*. Technical Report #497, Department of statistics, University of California at Berkeley.
- Diaconis, P. et Freedman, D. (1998). *On the hit and run process*. Technical Report #501, Department of statistics, University of California at Berkeley.
- Doob, J.L. (1953). *Stochastic Processes*. John Wiley & Sons, New York.
- Durrett, R. (1991). *Probability: Theory and Examples*. Wadsworth & Brooks/Cole, Belmont.
- Feller, W. (1968). *An Introduction to Probability Theory and its Applications, Vol. 1*. Third Edition. John Wiley & Sons, New York.
- Feller, W. (1971). *An Introduction to Probability Theory and its Applications, Vol. 2*. Second Edition. John Wiley & Sons, New York.
- Freedman, D. (1983). *Markov Chains*. Second Edition. Springer-Verlag, New York.
- Gelfand, A.E. et Smith, A.F.M. (1990). *Sampling-based approaches to calculating marginal densities*. *J. Amer. Statist. Assoc.*, 85, 398-409.
- Geman, S. et Geman, D. (1984). *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*. *IEEE Trans. Pattn. Anal. Mach. Intell.*, 6, 721-741.
- Geyer, C.J. (1992). *Practical Markov chain Monte Carlo (with discussion)*. *Statist. Sci.*, 7, 473-511.
- Gilks, W.R., Richardson, S. et Spiegelhalter, D.J. (éditeurs) (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

- Hammersley, J.M. et Handscomb, D.C. (1964). *Monte Carlo Methods*. Methuen & Co Ltd, London.
- Hastings, W.K. (1970). *Monte Carlo sampling methods using Markov chains and their applications*. *Biometrika*, 57, 97-109.
- Hoel, P.G., Port, S.C. et Stone, C.J. (1972). *Introduction to Stochastic Processes*. Houghton Mifflin, Boston.
- Karlin, S. et Taylor, H.M. (1975). *A First Course in Stochastic Processes*. Second Edition. Academic Press, New York.
- Karlin, S. et Taylor, H.M. (1981). *A Second Course in Stochastic Processes*. Academic Press, New York.
- Kipnis, C. et Varadhan, J.L. (1986). *Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions*. *Comm. Math. Phys.*, 104, 1-19.
- Lindvall, T. (1992). *Lectures on the Coupling Method*. John Wiley & Sons, New York.
- Liu, J.S., Wong, W.H. et Kong, A. (1995). *Covariance structure and convergence rate of the Gibbs Sampler with various scans*. *J. Roy. Statist. Soc. Ser. B*, 57, 157-169.
- Mengersen, K.L. et Tweedie, R.L. (1996). *Rates of convergence of the Hastings and Metropolis algorithms*. *Ann. Statist.*, 24, 101-121.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. et Teller, E. (1953). *Equations of state calculations by fast computing machines*. *J. Chem. Phys.*, 21, 1087-1091.
- Meyn, S.P. et Tweedie, R.L. (1993). *Markov Chains and Stochastic Stability*. Springer-Verlag, London.
- Meyn, S.P. et Tweedie, R.L. (1994). *Computable bounds for geometric convergence rates of Markov chains*. *Ann. Appl. Probab.*, 4, 981-1011.

- Nummelin, E. (1984). *General Irreducible Markov Chains and Non-negative Operators*. Cambridge University Press, Cambridge.
- Orey, S (1959). *Recurrent Markov chains*. *Pacific J. Math.*, 9, 805-827.
- Orey, S. (1971). *Limit Theorems for Markov Chain Transition Probabilities*. Van Nostrand Reinhold, New York.
- Peskun, P.H. (1973). *Optimum Monte-Carlo sampling using Markov chains*. *Biometrika*, 60, 607-612.
- Revuz, D. (1984). *Markov Chains*. Revised Edition. North-Holland Publishing Company, Amsterdam.
- Robert, C.P. (1994). *Discussion on Markov chains for exploring posterior distributions by L. Tierney*. *Ann. Statist.*, 22, 1742-1747.
- Robert, C.P. (1995). *Convergence control methods for Markov chain Monte Carlo algorithms*. *Statist. Sci.*, 10, 231-253.
- Roberts, G.O. et Polson, N.G. (1994). *On the geometric convergence of the Gibbs sampler*. *J. Roy. Statist. Soc. Ser. B*, 56, 377-384.
- Roberts, G.O. et Rosenthal, J.S. (1998a). *Markov-chain Monte Carlo: some practical implications of theoretical results (with discussion)*. *Rev. Canad. Statist.*, 26, 5-31.
- Roberts, G.O. et Rosenthal, J.S. (1998b). *On convergence rates of Gibbs samplers for uniform distributions*. À paraître dans *The Annals of Applied Probability*.
- Roberts, G.O. et Rosenthal, J.S. (1998c). *Two convergence properties of hybrid samplers*. *Ann. Appl. Probab.*, 8, 397-407.
- Roberts, G.O., Rosenthal, J.S. et Schwartz, P.O. (1998). *Convergence properties of perturbed Markov chains*. *J. Appl. Probab.*, 35, 1-11.

- Roberts, G.O. et Tweedie, R.L (1996). *Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. Biometrika*, 83, 96-110.
- Roberts, G.O. et Sahu, S.K. (1997). *Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. J. Roy. Statist. Soc. Ser. B*, 59, 291-317.
- Roberts, G.O. et Smith, A.F.M. (1994). *Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. Stoch. Proc. Appl.*, 49, 207-216.
- Rosenthal, J.S. (1995). *Minorization conditions and convergence rates for Markov chain Monte Carlo. J. Amer. Statist. Assoc.*, 90, 558-566.
- Royden, H.L. (1988). *Real Analysis*. Third Edition. Macmillan, New York.
- Rubinstein, R.Y. (1981). *Simulation and the Monte Carlo Method*. John Wiley & Sons, New York.
- Rudin, W. (1987). *Real and Complex Analysis*. Third Edition. McGraw-Hill, New York.
- Schervish, M.J. et Carlin, B.P. (1992). *On the convergence of successive substitution sampling. J. Comput. Graph. Statist.*, 1, 111-127.
- Smith, A.F.M. et Roberts, G.O. (1993). *Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). J. Roy. Statist. Soc. Ser. B*, 55, 3-24.
- Smith, R.L. (1984). *Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. Oper. Res.*, 32, 1296-1308.

- Tanner, M.A. (1991). *Tools for Statistical Inference: Observed Data and Data Augmentation Methods. Lecture Notes in Statist.*, 67. Springer-Verlag, New York.
- Tanner, M.A. et Wong, W.H. (1987). *The calculation of posterior distributions by data augmentation (with discussion). J. Amer. Statist. Assoc.*, 82, 528-550.
- Tierney, L. (1994). *Markov chains for exploring posterior distributions (with discussion). Ann. Statist.*, 22, 1701-1762.
- Tierney, L. (1995). *Introduction to general state-space Markov chain theory. Dans Markov Chain Monte Carlo in Practice* (édité par W.R. Gilks, S. Richardson et D.J. Spiegelhalter), pp. 59-74. Chapman & Hall, London.
- Tierney, L. (1998). *A note on Metropolis-Hastings kernels for general state spaces. Ann. Appl. Probab.*, 8, 1-9.