



# Détection de la convergence de processus de Markov

Béatrice Lachaud

► **To cite this version:**

Béatrice Lachaud. Détection de la convergence de processus de Markov. Mathématiques [math].  
Université René Descartes - Paris V, 2005. Français. tel-00010473

**HAL Id: tel-00010473**

**<https://tel.archives-ouvertes.fr/tel-00010473>**

Submitted on 7 Oct 2005

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

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

**UNIVERSITÉ RENÉ DESCARTES - PARIS 5**  
**Centre Universitaire des Saints-Pères**  
**UFR DE MATHÉMATIQUES ET INFORMATIQUE**

**Thèse**

en vue de l'obtention du grade de

**Docteur de l'Université René Descartes - Paris 5**

**Discipline : Mathématiques**  
**Spécialité : Probabilités**

présentée par

**Béatrice LACHAUD**

---

**Détection de la convergence de processus de Markov**

---

Soutenue publiquement le 14 septembre 2005, devant le jury composé de :

Valentine GENON-CATALOT	Examinatrice
Jean JACOD	Président
Laurent MICLO	Rapporteur
Gilles PAGÈS	Rapporteur
Bernard YCART	Directeur de thèse



# Remerciements

À l'heure où se terminent mes quatre années de thèse, je voudrais profiter des premières pages pour remercier tous ceux avec qui je les ai partagées.

Mes premiers remerciements vont naturellement à Bernard Ycart. Lors de mon DEA, j'ai beaucoup apprécié sa manière d'enseigner qui restera un modèle pour moi. Il m'a ensuite fait confiance pour le stage de DEA, et a su encadrer ma thèse avec patience et rigueur. Sa disponibilité, son enthousiasme et son dynamisme sont autant de qualités qui ont agrémenté cette aventure.

Gilles Pagès et Laurent Miclo ont accepté d'être les rapporteurs de ce travail malgré les délais plutôt courts ; je les en remercie chaleureusement. Je dois aussi remercier Laurent Miclo pour avoir déjà bénéficié de ses conseils lors de ma première année de thèse. Quant à Gilles Pagès, ses cours de DEA sont un très bon souvenir, mais je me rappelle surtout de son article de 1993 dans *Sciences et Vie Junior* intitulé "Tirer des nombres à pile ou face", qui constitue peut-être l'origine de mon penchant pour les probabilités...

Je tiens à remercier tout spécialement Valentine Genon-Catalot, non seulement pour sa participation au jury, mais surtout pour son accompagnement tout au long de ma thèse, et encore plus ces derniers temps.

Je suis aussi très honorée de la présence de Jean Jacod dans le jury. Qu'il en soit ici remercié.

Je ne serais sans doute pas arrivée jusqu'ici si je n'avais pas suivi les cours d'enseignants formidables. Du collègue à la maîtrise en passant par la prépa, je me souviens particulièrement de Mlles Latrémoière et Thauvin, de Mme Goin, et de M. Brancovan. Un grand merci aussi à tous les enseignants du DEA de Paris 7, notamment Laure Élie et Francis Comets grâce à qui j'ai fait mes premiers pas d'enseignante en tant que monitrice.

Pour la vie de thésarde au quotidien, je dois remercier l'ensemble du laboratoire MAP5 et de l'UFR Math-Info. Je n'oublierai pas l'ambiance conviviale et sympathique qui y règne ! Pour les questions plus pratiques ou administratives, j'adresse des remerciements spéciaux à Nellie Bouchard, Christophe Castellani, Marie-Hélène Gbaguidi et Voéuni Kheng, sans oublier les ingénieurs informaticiens Laurent Moineau et Thierry Raedersdorff pour leur disponibilité et leur savoir-faire très précieux. Je dois également des remerciements particuliers à Élodie Brunel-Piccinini pour la

mise en page de ce document.

Je voudrais bien sûr remercier toute la fine équipe des thésards avec qui j'ai partagé le bureau, les énigmes, le déjeuner, les maths, le café, Elixir, les mots fléchés, etc : Christian, Raphaël et Javiera pour le 705F, Olivier et David pour le bureau d'en face, Élise pour ses qualités d'entremetteuse (!), ainsi que "les filles du 4ème" Amandine, Cécile, Claire et Gwendoline. Je remercie aussi Anne, Béné et Estelle pour avoir partagé une tonne de bons moments depuis "les années Orsay" !

Il me reste à adresser un grand merci à mes parents, mes frère et soeur et le reste de la famille pour leurs encouragements et leur soutien constants.

Enfin, merci à David pour m'avoir aidée, rassurée, supportée, soutenue et encouragée depuis le début. Du fond du cœur, merci.

*“Un peu de folie est nécessaire pour faire un pas de plus.”*  
Paulo Coelho



# Table des matières

<b>Introduction</b>	<b>9</b>
<b>1 Motivations</b>	<b>13</b>
1.1 Les méthodes MCMC	13
1.1.1 L'algorithme de Hastings-Metropolis	14
1.1.2 Le recuit simulé	17
1.2 Le phénomène de cutoff	21
<b>2 Le processus d'Ornstein-Uhlenbeck</b>	<b>25</b>
2.1 Cadre de l'étude	25
2.2 Le phénomène de cutoff	27
2.2.1 Cutoff pour le $n$ -échantillon	27
2.2.2 Cutoff pour la moyenne empirique	29
2.3 Temps d'atteinte	31
2.4 Simulations	38
<b>3 Le cutoff</b>	<b>41</b>
3.1 Distances entre mesures de probabilité	41
3.1.1 La distance en variation totale	41
3.1.2 La distance de Hellinger	50
3.1.3 La distance du $\chi^2$	55
3.1.4 La distance de Kullback	58
3.2 Cas général des processus de Markov exponentiellement convergents	66
3.3 Exemples	77
3.3.1 Le processus binaire	78
3.3.2 La file $M/M/\infty$	81
3.3.3 Le processus d'Ornstein-Uhlenbeck	85
<b>4 Temps d'atteinte</b>	<b>89</b>
4.1 Théorème de la limite centrale pour des processus de Markov	89
4.1.1 L'espace $D$ de Skorohod	89
4.1.2 Le théorème de la limite centrale	95
4.1.3 Exemples	97
4.1.4 Le cas particulier des chaînes harmonisées	98



4.2	Temps d'atteinte . . . . .	101
4.2.1	Exemples . . . . .	106
	<b>Conclusion</b>	<b>113</b>
	<b>Bibliographie</b>	<b>118</b>

# Introduction

Les méthodes de Monte Carlo par chaînes de Markov (MCMC) ont pour principe de construire une chaîne de Markov dont la loi stationnaire soit la loi de probabilité à simuler. Cette idée est apparue en 1970 dans un article de W.K. Hastings [35], mais elle est passée inaperçue par la difficulté de sa mise en œuvre pratique à l'époque. Elle est en effet très coûteuse en temps de calcul, et c'est seulement avec le développement de la puissance de calcul des ordinateurs qu'elle s'est imposée dans de nombreux domaines très variés tels que la physique, l'informatique, le traitement d'images ou les statistiques (voir [31, 50, 53]).

La longueur du temps de calcul nécessaire à ces méthodes se comprend aisément en pensant qu'il faut attendre que la chaîne de Markov construite évolue près de son régime stationnaire pour pouvoir en tirer une réalisation raisonnablement approchée de la loi simulée. Le problème est encore plus aigu lorsque l'on cherche à simuler un échantillon de valeurs tirées selon la loi cible. Deux alternatives existent : la première est une méthode de simulation exacte due à J.G. Propp et D.B. Wilson [48] qui consiste à coupler dans le passé plusieurs trajectoires de la chaîne. La seconde consiste à faire évoluer la chaîne dans le futur jusqu'à ce qu'elle ait atteint son régime stationnaire. Deux méthodes se présentent alors : la méthode séquentielle et la méthode parallèle. Dans la méthode séquentielle, une seule trajectoire de la chaîne est simulée, et les valeurs retenues pour former l'échantillon sont des observations de la chaîne à des instants régulièrement espacés. Dans la méthode parallèle, l'idée est de lancer simultanément autant de chaînes que de copies souhaitées dans l'échantillon. Les chaînes sont arrêtées à un certain instant, et l'échantillon des observations à cet instant forme l'échantillon voulu. C'est à cette dernière méthode que nous nous intéressons. La problématique réside alors dans le choix de l'instant d'arrêt.

Dans le cas où l'espace d'états de la chaîne de Markov construite est fini, une méthode pour déterminer empiriquement un tel instant a déjà été proposée par B. Ycart dans [65]. Elle est liée à un phénomène de convergence abrupte des échantillons de chaînes de Markov qui est appelé cutoff. Cependant, les méthodes MCMC sont souvent basées sur une chaîne de Markov à espace d'états continu. Dans ce cas, les résultats théoriques sur la convergence de l'algorithme sont plutôt rares ; l'article [53] en fait une synthèse. Le point de départ de notre travail se situe dans ce cadre. Nous considérons un échantillon de chaînes de Markov (à temps discret ou continu) qui convergent vers leur mesure stationnaire, et nous souhaitons déterminer empiriquement l'instant où l'échantillon atteint son régime stationnaire.

Le chapitre 1 contient les motivations de notre étude. Nous commençons par exposer en détail l'algorithme de Hastings-Metropolis, qui est en lui-même la base des méthodes MCMC. Nous montrons ensuite comment l'utiliser dans des problèmes d'optimisation stochastique, en décrivant notamment l'algorithme du recuit simulé. Ensuite nous introduisons le phénomène de cutoff, qui constitue dans la suite l'outil de détection de la convergence des algorithmes.

Avant de nous intéresser au problème de la détection de convergence dans le cas général, nous consacrons le chapitre 2 à un exemple particulier qui éclaire la démarche entreprise dans la suite : le cas d'un échantillon de processus d'Ornstein-Uhlenbeck. Sur cet exemple, tous les calculs sont explicites, ce qui nous permet de montrer tout d'abord la présence du phénomène de cutoff, non seulement pour l'échantillon, mais aussi pour sa moyenne arithmétique. Nous en déduisons alors un estimateur de l'instant où le régime stationnaire est atteint. Cet estimateur est construit sur le temps d'atteinte par la moyenne de l'échantillon d'un niveau fixé, dont nous pouvons étudier la densité et le comportement asymptotique. Enfin, des simulations numériques permettent d'illustrer les résultats. Ce chapitre fait l'objet de l'article [37] à paraître en décembre 2005 dans *Journal of Applied Probability*.

Le chapitre 3 est consacré au phénomène de cutoff dans le cadre général des échantillons de processus à espace d'états continu tels que le processus échantillonné converge à vitesse exponentielle vers sa mesure stationnaire. Il fait l'objet de l'article [7] en collaboration avec J. Barrera et B. Ycart. Les résultats présentés sont nouveaux dans la mesure où, jusqu'à présent, le phénomène de cutoff n'est démontré que dans le cas de processus à espace d'états fini ou dénombrable (cf. [19, 58, 41, 64, 66]). Le phénomène de cutoff peut se traduire sous la forme suivante : il existe une suite déterministe d'instant  $(t_n)_{n \in \mathbb{N}^*}$  telle que, avant l'instant  $t_n$ , le  $n$ -échantillon est loin de sa mesure stationnaire, et après cet instant, il en est très proche. La notion de proximité entre les lois est mesurée à l'aide de distances entre probabilités. Nous y consacrons la première partie du chapitre. Les distances que nous considérons sont issues de divers domaines : les probabilités (distance en variation totale, distance de Hellinger), les statistiques (distance du  $\chi^2$ ) et la théorie de l'information (distance de Kullback). La seconde partie est consacrée au phénomène de cutoff en lui-même. Les démonstrations reposent d'une part sur les relations qui existent entre les distances entre échantillons et les distances entre marginales, et d'autre part sur les inégalités classiques entre les différentes distances. L'expression des instants de cutoff  $t_n$  est explicite (théorème 3.2.4). Si  $\rho$  désigne le taux de décroissance de la vitesse à laquelle le processus converge vers sa loi stationnaire, alors l'instant  $t_n$  vaut  $\log(n)/(2\rho)$ . Cela permet de donner une forme déterministe de l'instant auquel le  $n$ -échantillon évolue dans son régime stationnaire. De plus, nous obtenons des inégalités très précises permettant d'étudier le comportement de la convergence autour des instants de cutoff (théorème 3.2.5). Nous terminons le chapitre en traitant deux autres exemples de processus exponentiellement convergents, en plus du processus d'Ornstein-Uhlenbeck : le processus binaire et la file d'attente  $M/M/\infty$ .

En pratique, l'idée d'arrêter les algorithmes à l'instant de cutoff pose un problème majeur : le taux de convergence  $\rho$  est inconnu en général. Il est donc naturel d'en chercher un estimateur empirique ; c'est l'objet du chapitre 4. Nous tentons d'étendre au cas général l'étude menée

dans le chapitre 2 sur le processus d'Ornstein-Uhlenbeck. La démarche se compose de trois étapes. La première consiste à établir un théorème de la limite centrale pour des échantillons de processus. Cette étape n'intervient pas dans le cas du processus d'Ornstein-Uhlenbeck, car la moyenne arithmétique d'un échantillon de tels processus est encore un processus d'Ornstein-Uhlenbeck, ce qui simplifie l'étude. Grâce à ce théorème de la limite centrale, nous souhaitons remplacer l'étude du processus moyen de l'échantillon par l'étude du processus gaussien limite. Plus précisément, comme dans le cas du processus d'Ornstein-Uhlenbeck, nous cherchons un estimateur de l'instant de cutoff à partir du temps d'atteinte par une moyenne empirique d'un niveau fixé. La deuxième étape consiste à montrer que le temps d'atteinte par la moyenne d'un niveau fixé a le même comportement asymptotique que le temps d'atteinte par le processus gaussien limite d'une frontière mobile. À ce jour, cette question reste ouverte. La troisième étape est consacrée à l'étude des temps d'atteinte de barrières par des processus gaussien. L'article de J. Durbin [25] de 1985 nous sert de base pour déterminer la densité de tels temps d'atteinte, et expliquer comment retrouver le résultat du chapitre 2 concernant le processus d'Ornstein-Uhlenbeck. Enfin, nous reprenons les deux exemples de processus que nous avons introduits au chapitre 3, et nous obtenons la loi asymptotique des temps d'atteinte de leurs moyennes empiriques.



# Chapitre 1

## Motivations

Dans ce chapitre nous souhaitons retracer le cheminement qui nous a conduits à l'étude des échantillons de processus de Markov et à leur moyenne empirique. Deux principaux sujets nous ont amenés sur cette voie ; il s'agit tout d'abord de l'étude de la convergence de certaines méthodes MCMC, et ensuite du phénomène de cutoff.

### 1.1 Les méthodes MCMC

Le problème auquel on s'intéresse est la simulation d'une variable aléatoire de loi fixée. Il existe de nombreuses méthodes pour résoudre ce problème ; elles sont présentées au chapitre 2 du livre de B. Ycart [67] ou au chapitre 1 du livre de C. Robert [50]. Parmi elles, citons par exemple la méthode d'inversion, qui permet de simuler la loi de fonction de répartition  $F$  comme l'image réciproque par  $F$  d'une variable aléatoire de loi uniforme sur l'intervalle  $[0 ; 1]$ . Ces méthodes nécessitent généralement une bonne connaissance de la loi à simuler, et se généralisent difficilement en dimension supérieure à 1. Les méthodes de Monte-Carlo par chaînes de Markov (MCMC) ont été développées pour remédier à ces problèmes.

Le principe de base d'une méthode MCMC est de construire une chaîne de Markov ergodique qui ait pour loi stationnaire une loi de probabilité de densité  $f$  fixée à l'avance. A priori, une telle méthode paraît beaucoup plus lourde qu'une méthode classique puisqu'il faut attendre que la chaîne évolue dans son régime stationnaire pour obtenir une réalisation de la variable aléatoire. Pourtant, le caractère universel de la simulation par MCMC, comme nous allons le voir dans ce qui suit, a permis à ces méthodes de s'imposer dès que la puissance de calcul des ordinateurs a autorisé des calculs de grande ampleur en un temps raisonnable.

Dans ce qui suit, nous allons tout d'abord détailler un algorithme MCMC très classique : l'algorithme de Hastings-Metropolis. Puis nous montrerons comment cet algorithme permet aussi de mettre au point des algorithmes d'optimisation stochastique, comme le recuit simulé par exemple.

### 1.1.1 L'algorithme de Hastings-Metropolis

L'algorithme de Hastings-Metropolis a été introduit en 1970 par W.K. Hastings [35]. Depuis, il a été largement utilisé dans des domaines très variés. Le livre de C. Robert [50] présente cet algorithme en détail. L'article de P. Diaconis et L. Saloff-Coste [21] fait le point sur les questions théoriques relatives à cet algorithme. Nous supposons dans un premier temps que la loi à simuler est discrète ; la chaîne de Markov construite est alors à espace d'états fini. Ensuite, nous nous placerons dans le cas où la loi à simuler est continue, pour lequel la chaîne construite est à espace d'états infini.

#### Espace d'états fini.

Soit  $E$  un ensemble fini, et soit  $\pi$  une loi de probabilité strictement positive sur  $E$ . On cherche à fabriquer un échantillon de valeurs tirées selon la loi  $\pi$ . L'idée est de construire une chaîne de Markov qui ait  $\pi$  comme loi stationnaire.

Pour cela, on a besoin d'une matrice de transition auxiliaire, notée  $K$ . La chaîne de Markov de noyau de transition  $K$  est modifiée à chaque pas par une étape d'acceptation ou rejet. Plus précisément, si la chaîne est à l'état  $x$  à l'instant  $n$ , alors on commence par choisir  $y$  selon la loi  $K(x, \cdot)$ . Ensuite on définit le taux d'acceptation suivant :

$$A(x, y) = \begin{cases} \frac{\pi(y)K(y, x)}{\pi(x)K(x, y)} & \text{si } K(x, y) \neq 0, \\ 0 & \text{sinon.} \end{cases}$$

Si  $A(x, y) \geq 1$ , alors la chaîne va en  $y$  à l'instant  $n + 1$ . Dans le cas contraire, si  $A(x, y) < 1$ , alors la chaîne ne va en  $y$  qu'avec la probabilité  $A(x, y)$ , sinon elle reste en  $x$ .

La matrice de transition de la nouvelle chaîne  $M$  est donc donnée par la formule suivante :

$$\forall x \neq y, M(x, y) = \begin{cases} K(x, y) \min \{A(x, y), 1\} & \text{si } K(x, y) \neq 0, \\ 0 & \text{sinon.} \end{cases}$$

Les coefficients diagonaux de  $M$  sont tels que la somme des lignes vaille 1.

Le lemme suivant montre que la chaîne de Markov de matrice de transition  $M$  admet  $\pi$  pour mesure réversible.

**Lemme 1.1.1** *La chaîne de Markov de matrice de transition  $M$  admet la loi  $\pi$  comme mesure réversible, c'est-à-dire :*

$$\forall x, y \in E, \quad \pi(x)M(x, y) = \pi(y)M(y, x).$$

#### Démonstration.

L'équation se vérifie par un calcul direct utilisant les deux propriétés suivantes de  $A$ , qui découlent de sa définition :

- $\forall x, y \in E, \quad A(x, y) = 0 \iff A(y, x) = 0.$
- S'ils sont non nuls, les nombres  $A(x, y)$  et  $A(y, x)$  sont inverses l'un de l'autre, et donc :

$$A(x, y) > 1 \iff 0 < A(y, x) < 1 .$$

□

Il faut ensuite imposer certaines conditions sur le noyau  $K$  pour que la chaîne de Markov de matrice de transition  $M$  soit ergodique, c'est-à-dire qu'elle converge vers sa loi stationnaire  $\pi$ . On peut lire ces conditions dans le livre de C. Robert [50] ou dans celui de S.P. Meyn et R. Tweedie [42]. Ces conditions assurent le  $\pi$ -apériodicité et la  $\pi$ -irréductibilité de la chaîne.

Notons que l'on a toute latitude dans cet algorithme pour le choix de  $K$ . Dans la pratique, on prend souvent la matrice  $K$  symétrique, de façon à simplifier le calcul de  $A$ . De plus, quand l'espace  $E$  présente une structure de graphe, il est naturel de prendre pour  $K$  la matrice de transition de la marche aléatoire symétrique sur ce graphe.

Remarquons aussi que pour simuler la loi  $\pi$ , nous n'avons pas besoin de connaître la constante de normalisation de la densité, puisqu'elle se simplifie dans le calcul de  $A$ . C'est utile quand la constante de normalisation n'a pas de forme explicite.

### Espace d'états infini.

Soit  $(E, \mathcal{E})$  un espace mesurable, et soit  $\mu$  une mesure sur  $E$ . Dans cette partie, il est sous-entendu que les densités des lois sont les densités par rapport à la mesure  $\mu$ . On souhaite maintenant généraliser l'algorithme de Hastings-Metropolis pour simuler une valeur tirée selon la loi de densité  $f$  sur  $E$ . On fait l'hypothèse que  $f$  est strictement positive. On définit une chaîne de Markov qui a pour espace d'états  $E$  et pour loi stationnaire la loi de densité  $f$ . Tout ceci figure au chapitre 4 du livre de C. Robert [50].

Considérons un noyau de transition  $K$  tel que, pour tout  $x$  de  $E$ , la loi  $K(x, \cdot)$  ait pour densité  $k_x$ . Définissons alors la probabilité d'acceptation par :

$$\forall x, y \in E, \quad A(x, y) = \begin{cases} \min \left( 1, \frac{f(y)k_y(x)}{f(x)k_x(y)} \right) & \text{si } k_x(y) \neq 0, \\ 0 & \text{sinon.} \end{cases}$$

Construisons ensuite, en modifiant un peu  $K$ , une nouvelle chaîne de Markov sur le modèle précédent. La nouvelle probabilité de transition est donnée par :

$$\forall x \in E, \forall B \in \mathcal{E}, \quad M(x, B) = \int_B K(x, dy) A(x, y) + \mathbb{I}_B(x) \int K(x, dy) (1 - A(x, y)) .$$

Cette expression se justifie par le fait que, lorsque la chaîne est en  $x$ , deux possibilités se présentent pour qu'elle atteigne l'ensemble  $B$ . Soit  $x$  n'est pas un élément de  $B$ ; auquel cas



pour aller dans  $B$  il faut choisir un élément de  $B$  et l'accepter. Soit  $x$  est un élément de  $B$ , auquel cas il y a deux possibilités ; soit choisir un élément n'importe où et le refuser, soit choisir un élément de  $B$  et l'accepter. C'est en combinant ces trois cas que l'on obtient la formule précédente.

Le lemme suivant montre que la chaîne ainsi définie admet pour loi réversible la loi de densité  $f$ .

**Lemme 1.1.2** *La chaîne de Markov de noyau de transition  $M$  admet pour loi réversible la loi de densité  $f$ .*

**Démonstration.**

Nous devons montrer que :

$$\forall B \in \mathcal{E}, \quad \int f(x)M(x, B) d\mu(x) = \int_B f(x) d\mu(x) .$$

Nous avons :

$$\begin{aligned} \int f(x)M(x, B) d\mu(x) &= \int f(x) \left( \int \mathbb{I}_B(y)K(x, dy)A(x, y) \right) d\mu(x) \\ &\quad + \int f(x)\mathbb{I}_B(x) \left( \int K(x, dy) (1 - A(x, y)) \right) d\mu(x) \\ &= \iint f(x)\mathbb{I}_B(y)K(x, dy)A(x, y) d\mu(x) \\ &\quad + \iint f(x)\mathbb{I}_B(x)K(x, dy) d\mu(x) \\ &\quad - \iint f(x)\mathbb{I}_B(x)K(x, dy)A(x, y) d\mu(x) . \end{aligned}$$

Or d'après la définition de  $A$ , nous pouvons écrire l'équation d'équilibre détaillé suivante :

$$\forall x, y \in E, \quad f(x)k_x(y)A(x, y) = f(y)k_y(x)A(y, x) .$$

Donc :

$$\iint f(x)\mathbb{I}_B(y)K(x, dy)A(x, y) d\mu(x) = \iint f(x)\mathbb{I}_B(x)K(x, dy)A(x, y) d\mu(x) ,$$

ce qui montre que :

$$\begin{aligned} \int f(x)M(x, B) d\mu(x) &= \iint f(x)\mathbb{I}_B(x)K(x, dy) d\mu(x) \\ &= \int f(x)\mathbb{I}_B(x) d\mu(x), \end{aligned}$$

puisque  $K(x, \cdot)$  est une loi de probabilité. □

Comme dans le cas où l'espace d'états est fini, il faut imposer certaines conditions au noyau  $K$  pour que la chaîne de Markov converge vers la loi stationnaire de densité  $f$  (cf. [50, 42]).

De plus, il est aussi très pratique dans ce cas de considérer un noyau  $K$  symétrique par rapport à la mesure  $\mu$  de façon à simplifier le calcul de  $A$ . Donnons dans ce cas l'algorithme sous forme explicite. Pour simuler la chaîne de Markov  $(\mathbf{X}_n)_{n \in \mathbb{N}}$  de noyau de transition  $M$ , on commence par initialiser  $\mathbf{X}$ ; ensuite, pour passer de  $\mathbf{X}_n$  à  $\mathbf{X}_{n+1}$ , on procède de la façon suivante :

- on tire au hasard un nombre  $x$  selon la loi  $K(\mathbf{X}_n, \cdot)$ ,
- on calcule  $\rho_n = \frac{f(x)}{f(\mathbf{X}_n)}$ ,
- si  $\rho_n \geq 1$ , alors  $\mathbf{X}_{n+1} = x$ ,
- sinon, on tire un nombre  $u$  selon la loi uniforme sur  $[0 ; 1]$  et :
  - si  $u < \rho_n$ , alors  $\mathbf{X}_{n+1} = x$ ,
  - sinon,  $\mathbf{X}_{n+1} = \mathbf{X}_n$ .

Si la densité  $k_x$  de la probabilité  $K(x, \cdot)$  est une fonction  $k$  paire telle que :

$$k_x(y) = k(x - y) ,$$

alors nous pouvons remarquer que la chaîne de Markov ainsi construite se met sous la forme additive suivante :

$$\forall n \in \mathbb{N}, \quad \mathbf{X}_{n+1} = \mathbf{X}_n + H_n \mathbb{I}_{A_n} ,$$

où :

- $(H_n)_{n \in \mathbb{N}}$  est une suite de variables aléatoires indépendantes et de même loi de densité  $k$  et indépendantes de  $(\mathbf{X}_n)_{n \in \mathbb{N}}$ ,
- $(A_n)_{n \in \mathbb{N}}$  est la suite des événements "acceptation", dans le sens où  $A_n$  est l'événement sur lequel l'algorithme accepte le pas proposé  $\mathbf{X}_n + H_n$  :

$$A_n = \{\rho_n \geq 1\} \cup \{\rho_n < 1 \text{ et } U_n < \rho_n\} ,$$

où  $(U_n)_{n \in \mathbb{N}}$  est une suite de variables aléatoires indépendantes et de même loi uniforme sur  $[0 ; 1]$ , et  $\rho_n$  est défini par :

$$\rho_n = \frac{f(\mathbf{X}_n + H_n)}{f(\mathbf{X}_n)} .$$

L'algorithme de Hastings-Metropolis est universel dans le sens où il permet de simuler n'importe quelle loi de probabilité en utilisant un noyau de transition choisi par l'utilisateur. C'est ce qui en fait un algorithme très utilisé en pratique. De plus, comme nous allons le voir dans la partie suivante, il est à la base d'algorithmes stochastiques d'optimisation très efficaces comme le recuit simulé.

### 1.1.2 Le recuit simulé

Le recuit simulé est un des plus anciens algorithmes stochastiques d'optimisation, et en même temps un des plus connus. Une importante littérature existe à propos de ses performances, et

de leur justification théorique (cf. [12, 13, 11, 6, 8]).

Mathématiquement, il s'agit d'optimiser une fonction de plusieurs variables  $f$  définie sur un espace mesurable  $(E, \mathcal{E})$  et à valeurs réelles. Les méthodes d'analyse classique fonctionnent mal quand le nombre de variables est grand, et quand la fonction  $f$  présente de nombreux extrema locaux, difficiles à différencier des extrema globaux dans les calculs. Une alternative possible est de fabriquer un algorithme stochastique (cf. [17, 24]). L'algorithme de recuit simulé est basé sur l'algorithme de Hastings-Metropolis; l'idée est de simuler la loi uniforme sur l'ensemble des extrema globaux de la fonction  $f$ . Pour cela, on dispose des lois de Boltzmann (ou lois de Gibbs), qui ont la particularité de converger, en un sens que nous allons voir ci-dessous, vers cette loi uniforme.

À l'origine, le recuit simulé définit, à l'instar de l'algorithme de Hastings-Metropolis initial, une chaîne de Markov sur un espace d'états fini, c'est-à-dire que la recherche de l'optimum de la fonction  $f$  se fait parmi un nombre fini d'états. Cette chaîne de Markov n'est pas homogène, c'est-à-dire que ses probabilités de transition dépendent de l'instant considéré. Cependant, elle converge vers une loi stationnaire, qui est la loi uniforme sur les minima de la fonction  $f$ .

Nous souhaitons généraliser cet algorithme à la recherche des minima dans un ensemble  $E$  infini. L'espace  $(E, \mathcal{E})$  est un espace mesurable. Soit  $\mu$  une mesure sur  $(E, \mathcal{E})$ . Nous supposons que la fonction  $f$  est bornée inférieurement par rapport à  $\mu$ , et, sans perte de généralité, que le  $\mu$ -essentiel infimum de la fonction  $f$  est ramené à 0. Rappelons que le  $\mu$ -essentiel infimum de  $f$  est défini de la façon suivante :

$$\mu - \text{ess inf}(f) = \sup \{c \in \mathbb{R} ; \mu(f < c) = 0\} .$$

Le principe du recuit simulé est de simuler une loi de Boltzmann (ou loi de Gibbs). Les lois de Boltzmann sont indicées par un paramètre strictement positif appelé température; pour un réel  $T$  strictement positif, la loi de Boltzmann de paramètre  $T$ , notée  $G_T$ , est la loi de densité  $g_T$  par rapport à  $\mu$ , définie de la façon suivante :

$$\forall x \in E, \quad g_T(x) = Z_T^{-1} e^{-\frac{1}{T}f(x)} ,$$

où  $Z_T$  est la constante de normalisation :

$$Z_T = \int e^{-\frac{1}{T}f(y)} d\mu(y) .$$

Tout l'intérêt de ces distributions réside dans le fait qu'elles se concentrent sur les minima de la fonction  $f$  si la température tend vers 0. Plus précisément, on considère une suite de réels strictement positifs  $(T_n)_{n \in \mathbb{N}}$  qui tend vers 0 quand  $n$  tend vers  $+\infty$ . Tout se passe comme si, à l'étape  $n$ , on essayait de simuler la loi  $G_{T_n}$  par un pas de l'algorithme de Hastings-Metropolis avec un noyau  $K$  symétrique par rapport à  $\mu$ . Ainsi, la probabilité d'acceptation à l'étape  $n$  est donnée par :

$$A_{T_n}(x, y) = \exp \left( \frac{1}{T_n} \min(0, f(x) - f(y)) \right) .$$

De plus, le noyau de transition de la chaîne est défini par :

$$\forall x \in E, \forall B \in \mathcal{E}, P_n(x, B) = \int_B K(x, d\omega) A_{T_n}(x, \omega) + \mathbb{I}_B(x) \int K(x, d\omega) (1 - A_{T_n}(x, \omega)) . \quad (1.1.1)$$

Concrètement, on obtient l'algorithme suivant, décrit dans [17] par exemple :

- $n = 0$ . La chaîne part de  $\mathbf{X}_0$ .
- À l'étape  $n$ , la transition de  $\mathbf{X}_n$  à  $\mathbf{X}_{n+1}$  se décompose en une étape d'exploration et une étape d'acceptation.
  - ★ L'étape d'exploration consiste à proposer un état  $\mathbf{Y}_n$  de loi  $K(\mathbf{X}_n, \cdot)$ .
  - ★ L'étape d'acceptation se décompose en deux cas :
    - si  $f(\mathbf{Y}_n) \leq f(\mathbf{X}_n)$  on accepte l'état  $\mathbf{Y}_n$  et on pose  $\mathbf{X}_{n+1} = \mathbf{Y}_n$ .
    - si  $f(\mathbf{Y}_n) > f(\mathbf{X}_n)$  alors,
      - on accepte  $\mathbf{Y}_n$  et on pose  $\mathbf{X}_{n+1} = \mathbf{Y}_n$  avec la probabilité

$$A_{T_n}(\mathbf{X}_n, \mathbf{Y}_n) = \exp\left(-\frac{1}{T_n} (f(\mathbf{Y}_n) - f(\mathbf{X}_n))\right) ,$$

- on reste en  $\mathbf{X}_n$  et on pose  $\mathbf{X}_{n+1} = \mathbf{X}_n$  avec la probabilité

$$1 - A_{T_n}(\mathbf{X}_n, \mathbf{Y}_n) = 1 - \exp\left(-\frac{1}{T_n} (f(\mathbf{Y}_n) - f(\mathbf{X}_n))\right) .$$

Ainsi, chaque étape de l'algorithme de recuit simulé est un pas de l'algorithme de Hastings-Metropolis, ce qui signifie que, pour tout  $n$ , la probabilité de transition  $P_n$  a pour loi stationnaire la loi de Boltzmann  $G_{T_n}$ .

Voyons maintenant pourquoi la loi  $G_{T_n}$  se concentre sur les minima de la fonction  $f$  quand  $n$  tend vers  $+\infty$ . On considère une suite de températures strictement positives  $(T_n)_{n \in \mathbb{N}}$  qui tend vers 0 quand  $n$  tend vers  $+\infty$ . La proposition suivante donne le résultat.

**Proposition 1.1.3** Soit  $\varepsilon > 0$ . Soit  $M_\varepsilon = \{x \in E ; f(x) \leq \varepsilon\}$ . Alors :

$$\lim_{n \rightarrow +\infty} G_{T_n}(M_\varepsilon) = 1 .$$

**Démonstration.**

Elle s'inspire de [17]. Soit  $\varepsilon > 0$ . Comme 0 est le  $\mu$ -essentiel infimum de  $f$ , on sait que :

$$\mu(f < 0) = 0 ,$$

et que :

$$\mu(M_\varepsilon) > 0 .$$

Regardons le complémentaire de  $M_\varepsilon$ , noté  $M_\varepsilon^c$  :

$$\begin{aligned} G_{T_n}(M_\varepsilon^c) &= \left( \int e^{-\frac{f(y)}{T_n}} d\mu(y) \right)^{-1} \int_{M_\varepsilon^c} e^{-\frac{f(x)}{T_n}} d\mu(x) \\ &= \left( \int e^{-\frac{f(y)}{T_n}} e^{\frac{\varepsilon}{T_n}} d\mu(y) \right)^{-1} \int_{M_\varepsilon^c} e^{-\frac{f(x)}{T_n} + \frac{\varepsilon}{T_n}} d\mu(x) . \end{aligned}$$

Minorons la première parenthèse par  $\mu(M_\varepsilon)$ , en remarquant que sur  $M_\varepsilon$ , on a  $f \leq \varepsilon$  :

$$\begin{aligned} \int e^{-\frac{f(y)}{T_n}} e^{\frac{\varepsilon}{T_n}} d\mu(y) &\geq \int_{M_\varepsilon} e^{-\frac{f(y)}{T_n}} e^{\frac{\varepsilon}{T_n}} d\mu(y) \\ &\geq \int_{M_\varepsilon} d\mu(y) \\ &= \mu(M_\varepsilon). \end{aligned}$$

Enfin, par convergence dominée,

$$\lim_{n \rightarrow +\infty} \int_{M_\varepsilon} e^{-\frac{f(x)}{T_n} + \frac{\varepsilon}{T_n}} d\mu(x) = 0.$$

La proposition est donc démontrée.  $\square$

Il reste alors à montrer que l'algorithme de recuit simulé converge, c'est-à-dire que :

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow +\infty} P_n(M_\varepsilon) = 1.$$

Cette question est résolue dans le cas où l'espace d'états est fini. Le résultat a été démontré par B. Hajek [34] en 1988. Il fait apparaître l'idée intuitive que la convergence dépend de l'allure de la fonction  $f$  de départ ; si les minima locaux sont trop profonds, l'algorithme va souvent stagner, et donc converger moins vite que s'ils sont peu profonds, ou ne pas converger. Le résultat de B. Hajek traduit cette intuition. Des suites de température  $(T_n)_{n \in \mathbb{R}}$  qui assurent la convergence de l'algorithme sont du type :

$$T_n = \frac{h}{\log(n)},$$

où  $h$  est un paramètre positif qui mesure la profondeur des minima locaux de la fonction  $f$ .

Dans le cas où l'espace d'états est infini, la question est aussi résolue moyennant une hypothèse supplémentaire sur le noyau  $K$ , qui est la suivante (cf. [17] p. 100) :

$$(H) \quad \exists p \in \mathbb{N}^*, \exists \varepsilon > 0, \exists \nu \text{ probabilité sur } E, \forall x \in E, \forall B \in \mathcal{E}, \quad K^p(x, B) \geq \varepsilon \nu(B).$$

Un schéma de température qui assure la convergence est aussi de la forme :

$$T_n = \frac{h}{\log(n)},$$

où  $h$  est un paramètre positif qui mesure la profondeur des minima locaux de la fonction  $f$ .

Ces résultats montrent que le recuit simulé est un algorithme d'optimisation d'une fonction générale, qui est simple à implémenter dans la pratique. Cependant, la fonction logarithme,

qui intervient dans l'expression de la température, est chère en temps de calcul et varie lentement. Recalculer  $T_n$  à chaque pas de temps serait inefficace. C'est pourquoi l'évolution de la température est réalisée par paliers de longueur exponentiellement croissante :

$$\forall k \in \mathbb{N}^*, \forall n \in ]e^{(k-1)h}, e^{kh}[, \quad T_n = \frac{1}{k}.$$

Pour un tel schéma de température, la chaîne de Markov construite est homogène sur des intervalles de temps de plus en plus longs. On peut montrer (cf. [67] p. 141) que, pour  $k$  assez grand, chacun de ces intervalles de temps est suffisamment long pour que la chaîne atteigne son équilibre, à savoir la loi de Boltzmann de paramètre  $1/k$ . Comme ces lois convergent vers la restriction de  $\mu$  renormalisée sur les minima de la fonction  $f$  (dans le cas où  $E$  est dénombrable), cela justifie que l'algorithme converge, ce que nous avons affirmé ci-dessus.

Ces résultats de convergence ne donnent cependant aucune indication sur le nombre d'itérations à réaliser ; autrement dit, il n'est pas sûr, au moment où l'on stoppe l'algorithme, que le minimum de la fonction soit atteint. Une des motivations de cette thèse est de donner un critère d'arrêt empirique des algorithmes tels que le recuit simulé.

Cette question de l'arrêt d'une méthode MCMC est déjà apparue dans l'article [65] de B. Ycart. Cependant, elle ne concerne que des algorithmes à espace d'états fini. La solution repose sur le phénomène de convergence abrupte de certaines chaînes de Markov vers leur loi stationnaire, que l'on appelle cutoff. Ce phénomène est expliqué dans la partie suivante.

## 1.2 Le phénomène de cutoff

Lorsque l'on étudie une suite de processus de Markov  $(X_n)_{n \in \mathbb{N}^*}$ , où pour tout  $n$ , le processus  $X_n = \{X_n(t); t \geq 0\}$  converge vers sa mesure stationnaire  $\pi_n$  (en supposant qu'elle existe), il apparaît parfois que cette convergence est brutale ; plus précisément, quand  $t$  est petit, la loi de  $X_n(t)$  est "loin" de la loi  $\pi_n$ , et brutalement, la loi de  $X_n(t)$  devient "très proche" de  $\pi_n$ . De plus, la transition est de plus en plus brutale quand  $n$  tend vers l'infini. Pour quantifier la notion de proximité de deux lois, on choisit une distance  $d$  entre lois de probabilité. On note  $\mathcal{L}(X_n(t))$  la loi de la variable aléatoire  $X_n(t)$ , et on définit la notion de convergence abrupte (ou cutoff) de la façon suivante :

**Définition 1.2.1** Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite de processus telle que chaque  $X_n$  ait une loi stationnaire notée  $\pi_n$ . Soit  $(t_n)_{n \in \mathbb{N}^*}$  une suite de réels positifs tendant vers  $+\infty$ . On dit que la suite  $(t_n)_{n \in \mathbb{N}^*}$  est une suite d'instant de cutoff pour la suite de processus  $(X_n)_{n \in \mathbb{N}^*}$  si :

$$\forall c > 0, \quad \begin{cases} c < 1 & \Rightarrow \liminf_{n \rightarrow +\infty} d(\mathcal{L}(X_n(ct_n)), \pi_n) > 0, \\ c > 1 & \Rightarrow \lim_{n \rightarrow +\infty} d(\mathcal{L}(X_n(ct_n)), \pi_n) = 0. \end{cases}$$

Le choix de la distance  $d$  peut varier selon le contexte étudié. Le plus souvent  $d$  est la distance en variation totale. On en rappelle ci-dessous la définition.

**Définition 1.2.2** Soient  $\mu$  et  $\nu$  deux mesures sur l'espace  $(\Omega, \mathcal{A})$ . La distance en variation totale entre  $\mu$  et  $\nu$  est définie par la formule suivante :

$$d_{VT}(\mu, \nu) = \sup_{A \in \mathcal{A}} |\mu(A) - \nu(A)| .$$

Les propriétés de la distance en variation totale seront détaillées dans la partie 3.1.1 du chapitre 3. Nous verrons aussi dans la partie 3.1 d'autres choix possibles pour la distance qui mesure le phénomène de cutoff, ainsi que leurs propriétés.

Historiquement, la notion de cutoff a été introduite par D. Aldous et P. Diaconis [1] en 1987. Ensuite, elle a été étudiée dans de nombreux autres contextes, comme par exemple la marche aléatoire sur l'hypercube de  $\mathbb{Z}^d$  [20], les modèles de diffusion [22, 59], les files d'attente [29, 41, 51], ou bien les marches aléatoires sur les groupes [23, 55]. Un premier état de l'art a été réalisé par P. Diaconis en 1996 [19], qui concerne le cutoff pour les chaînes de Markov à espace d'états fini. La même année, le cours à Saint-Flour de L. Saloff-Coste [57] donne aussi une vue d'ensemble sur le sujet, ainsi que son chapitre plus récent sur les marches aléatoires sur les groupes finis dans [58] en 2004.

Dès 1999, B. Ycart a montré [64, 66] que ce phénomène survenait aussi dans le cas où le processus  $X_n$  est un  $n$ -échantillon de chaînes de Markov (à temps discret ou continu) à espace d'états fini, convergeant vers une loi stationnaire. Ce dernier cas représente beaucoup plus qu'un cas particulier de cutoff ; en effet l'étude d'un  $n$ -échantillon de chaînes à la place de la chaîne elle-même permet d'établir des théorèmes limites en lien avec le phénomène de cutoff. B. Ycart a notamment étudié dans [65] des fonctionnelles d'échantillons de chaînes de Markov qui permettent de construire des estimateurs des instants de cutoff. Ces estimateurs sont particulièrement utiles puisqu'ils fournissent une règle d'arrêt pour les algorithmes MCMC.

Nous nous intéressons à la généralisation de ces résultats dans le cadre des processus de Markov à espace d'états quelconques. Nous supposons que le processus échantillonné converge à vitesse exponentielle vers sa mesure stationnaire, c'est-à-dire que la distance entre la loi du processus à l'instant  $t$  et sa loi stationnaire est une fonction exponentielle décroissante en  $t$ . Nous montrons alors au chapitre 3 (théorème 3.2.4) que le  $n$ -échantillon de processus présente un phénomène de cutoff. Les instants de cutoff sont de la forme  $\log(n)/(2\rho)$ , où  $\rho$  est le taux de décroissance exponentielle intervenant dans la vitesse de convergence. Cependant, comme le taux  $\rho$  est inconnu en général, il est nécessaire d'en chercher un estimateur. Dans cette optique, nous nous inspirons de la méthode initiée par B. Ycart dans [65]. Un tel estimateur peut être construit comme le temps d'atteinte d'un niveau par la moyenne arithmétique d'une fonction  $\varphi$  du  $n$ -échantillon. Le niveau à atteindre est en fait l'espérance de  $\varphi$  sous la loi  $\nu$ , qui correspond aussi à la limite quand  $t$  tend vers  $+\infty$  de l'espérance du processus échantillonné à l'instant  $t$ . Nous étudions ces temps d'atteinte au chapitre 4.

Ce travail est à mettre en parallèle avec celui de C. Paroissin dans sa thèse [43] en 2002. Il a aussi étudié les temps d'atteinte de niveaux par la moyenne arithmétique d'une fonction

d'un  $n$ -échantillon de processus. Cependant deux différences majeures se dégagent ; la première concerne l'espace d'états des processus considérés, qui est supposé fini dans son travail, et quelconque pour nous. La seconde, et la plus flagrante, se situe dans le niveau à atteindre par la moyenne. Dans les deux cas, les niveaux sont fixes ; comme nous l'avons expliqué dans le paragraphe précédent, le niveau à atteindre, dans notre travail, est la limite de l'espérance du processus sommé. Dans la thèse de C. Paroissin, le niveau étudié est un réel que la fonction espérance atteint en un temps fini.

Dans le chapitre qui vient, nous nous intéressons à un cas particulier de processus : le processus d'Ornstein-Uhlenbeck. Ce processus admet une loi normale pour loi stationnaire, et il converge à vitesse exponentielle vers cette loi, comme nous le verrons dans la proposition 2.1.2. Sur cet exemple, nous déroulons notre plan de travail ; tout d'abord, nous montrons que le  $n$ -échantillon de processus présente un phénomène de cutoff. Ensuite nous remarquons que la moyenne empirique de l'échantillon est elle-même un processus d'Ornstein-Uhlenbeck, ce qui en facilite l'étude. Sur cette moyenne, nous construisons un estimateur de l'instant de cutoff, sous la forme d'un temps d'atteinte. Il est alors possible de calculer la densité de cette variable aléatoire, et d'en déduire le comportement asymptotique de l'estimateur.





# Chapitre 2

## Le processus d'Ornstein-Uhlenbeck

Ce chapitre fait l'objet de l'article [37] à paraître dans *Journal of Applied Probability* en décembre 2005.

Il est consacré à l'étude d'un échantillon de processus d'Ornstein-Uhlenbeck, pour lequel tous les calculs sont explicites. Sur cet exemple, nous sommes en mesure de montrer que le  $n$ -échantillon présente un phénomène de cutoff à l'instant  $t_n$  (nous verrons la forme de  $t_n$  dans la proposition 2.2.1). De plus, nous savons étudier le processus moyen du  $n$ -échantillon (c'est-à-dire construit à chaque instant comme la moyenne arithmétique des  $n$  processus à cet instant), car ce processus est encore un processus d'Ornstein-Uhlenbeck; en particulier, nous savons étudier le temps d'atteinte de 0 par ce processus moyen, où 0 correspond à l'espérance de la loi stationnaire du processus échantillonné. Nous montrons que ce temps d'atteinte se concentre autour de l'instant de cutoff  $t_n$  quand  $n$  tend vers l'infini, et que les queues de ce temps d'atteinte ont la même allure que le comportement de la distance en variation totale au voisinage de l'instant  $t_n$ . Pour motiver l'étude qui figure dans les chapitres suivants, nous allons nous contenter de donner ici les énoncés des résultats clés, en précisant la référence de la partie où figure la démonstration. Pour commencer, nous allons présenter en détail le processus étudié, avec ses propriétés principales.

### 2.1 Cadre de l'étude

Le processus d'Ornstein-Uhlenbeck est solution de l'équation différentielle stochastique dite de Langevin. Il a été introduit dans les années 30 par deux physiciens [61] pour modéliser le mouvement d'une particule dans un fluide visqueux. La définition que nous adoptons pour toute la suite est la suivante.

**Définition 2.1.1** Soient  $\rho$  et  $\sigma$  deux réels strictement positifs, soit  $x_0$  un réel, et soit  $B = (B(t))_{t \geq 0}$  un mouvement brownien standard. On appelle processus d'Ornstein-Uhlenbeck partant

de  $x_0$ , de paramètres  $\rho$  et  $\sigma$  l'unique solution de l'équation différentielle stochastique suivante :

$$\begin{cases} dX(t) = -\rho X(t) dt + \sigma\sqrt{2\rho} dB(t) \\ X(0) = x_0 \end{cases} . \quad (2.1.1)$$

La forme du coefficient de diffusion n'est pas habituelle. Nous l'avons choisie pour obtenir une homogénéité dans la variance de la loi du processus, comme nous allons le voir dans ce qui suit.

Considérons un processus d'Ornstein-Uhlenbeck  $X$  partant de  $x_0$  et de paramètres  $\rho$  et  $\sigma$ . Grâce à la formule d'Itô, il est facile de voir que le processus  $Y$  défini pour tout  $t$  positif par :

$$Y(t) = X(t) e^{\rho t} ,$$

est solution de l'équation différentielle stochastique :

$$\begin{cases} dY(t) = \sigma\sqrt{2\rho} e^{\rho t} dB(t) \\ Y(0) = x_0 \end{cases} .$$

Par conséquent,  $Y$  s'écrit de façon explicite sous la forme :

$$Y(t) = Y(0) + \sigma\sqrt{2\rho} \int_0^t e^{\rho s} dB(s) .$$

Ainsi le processus  $X$  a aussi une formulation explicite :

$$X(t) = x_0 e^{-\rho t} + \sigma\sqrt{2\rho} \int_0^t e^{-\rho(t-s)} dB(s) .$$

Cette écriture du processus  $X$  permet de voir qu'il est gaussien. Sa moyenne est donnée, pour tout  $t$  positif, par :

$$E(X(t)) = x_0 e^{-\rho t} ,$$

et sa fonction de covariance, pour tous  $s$  et  $t$  positifs, par :

$$\text{Cov}(X(s), X(t)) = \sigma^2 [e^{-\rho|t-s|} - e^{-\rho(t+s)}] .$$

Par conséquent, la loi de la variable aléatoire  $X(t)$  est une gaussienne de moyenne  $x_0 e^{-\rho t}$  et de variance  $\sigma^2(1 - e^{-2\rho t})$ .

En faisant tendre  $t$  vers  $+\infty$ , nous voyons que le processus d'Ornstein-Uhlenbeck admet la loi gaussienne centrée de variance  $\sigma^2$  comme loi stationnaire, et qu'il converge en loi vers cette distribution quand  $t$  tend vers  $+\infty$  :

$$X(t) \xrightarrow[t \rightarrow +\infty]{loi} \mathcal{N}(0, \sigma^2) . \quad (2.1.2)$$

Plus précisément, nous pouvons démontrer que cette convergence a lieu à vitesse exponentielle, dans le sens où la distance en variation totale entre la loi du processus à l'instant  $t$  (notée  $\mathcal{L}(X(t))$ ) et sa loi stationnaire est de l'ordre d'une fonction exponentielle décroissante de  $t$ .

**Proposition 2.1.2** *La distance en variation totale entre la loi du processus d'Ornstein-Uhlenbeck à l'instant  $t$  et sa loi stationnaire a la forme suivante lorsque  $t$  tend vers  $+\infty$  :*

$$d_{VT}(\mathcal{L}(X(t)), \mathcal{N}(0, \sigma^2)) = \frac{|x_0|}{\sigma\sqrt{2\pi}} e^{-\rho t} + o(e^{-2\rho t}).$$

Cette propriété fait du processus d'Ornstein-Uhlenbeck un cas particulier de processus exponentiellement convergent ; nous définirons cette notion précisément au chapitre 3 (définition 3.2.1).

**Démonstration.**

Posons  $\varepsilon = e^{-\rho t}$ . On doit donc évaluer, quand  $\varepsilon$  tend vers 0, la distance en variation totale entre la loi  $\mathcal{N}(x_0\varepsilon, \sigma^2(1 - \varepsilon^2))$  et la loi  $\mathcal{N}(0, \sigma^2)$ . Partons de l'égalité (3.1.2) de la proposition 3.1.3 où la distance en variation totale s'écrit comme la moitié de la distance en norme  $L^1$  :

$$d_{VT}(\mathcal{N}(x_0\varepsilon, \sigma^2(1 - \varepsilon^2)), \mathcal{N}(0, \sigma^2)) = \frac{1}{2} \int_{\mathbb{R}} \left| \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} - \frac{1}{\sigma\sqrt{2\pi}\sqrt{1 - \varepsilon^2}} e^{-\frac{(x-x_0\varepsilon)^2}{2\sigma^2(1-\varepsilon^2)}} \right| dx$$

Par le changement de variable ( $y = \frac{x}{\sigma}$ ), on arrive à :

$$d_{VT}(\mathcal{N}(x_0\varepsilon, \sigma^2(1 - \varepsilon^2)), \mathcal{N}(0, \sigma^2)) = d_{VT}\left(\mathcal{N}\left(\frac{x_0\varepsilon}{\sigma}, 1 - \varepsilon^2\right), \mathcal{N}(0, 1)\right).$$

Or cette distance est calculée dans la proposition 3.1.8 du chapitre 3 :

$$\forall \theta \in \mathbb{R}, \quad d_{VT}(\mathcal{N}(\theta\varepsilon, 1 - \varepsilon^2), \mathcal{N}(0, 1)) = \frac{|\theta|\varepsilon}{\sqrt{2\pi}} + o(\varepsilon^2).$$

Nous en déduisons que :

$$d_{VT}(\mathcal{L}(X(t)), \mathcal{N}(0, \sigma^2)) = \frac{|x_0|}{\sigma\sqrt{2\pi}} e^{-\rho t} + o(e^{-2\rho t}).$$

□

Venons-en à l'étude du phénomène de cutoff dans le cas du processus d'Ornstein-Uhlenbeck. Nous nous intéressons tout d'abord à un  $n$ -échantillon de processus, avant d'étudier sa moyenne empirique.

## 2.2 Le phénomène de cutoff

### 2.2.1 Cutoff pour le $n$ -échantillon

Considérons un  $n$ -échantillon de processus d'Ornstein-Uhlenbeck  $(X_1, \dots, X_n)$  partant tous du même point  $x_0$ , de mêmes paramètres  $\rho$  et  $\sigma$  strictement positifs. La loi stationnaire de ce  $n$ -échantillon est la loi normale  $n$ -dimensionnelle  $\mathcal{N}(0, \sigma^2)^{\otimes n}$  obtenue comme produit tensoriel  $n$  fois de la loi normale centrée de variance  $\sigma^2$  par elle-même. Nous la notons  $\nu^{\otimes n}$ . De plus, le

$n$ -échantillon converge, quand  $t$  tend vers l'infini, vers sa loi stationnaire. Définissons alors la fonction  $d_n$ , à chaque instant  $t$ , comme la distance en variation totale entre la loi du  $n$ -échantillon à l'instant  $t$  et la loi stationnaire :

$$d_n(t) = d_{VT}(\mathcal{L}(X_1(t), \dots, X_n(t)), \nu^{\otimes n}) .$$

La proposition suivante montre que le  $n$ -échantillon présente un phénomène de cutoff (voir la définition 1.2.1), et décrit précisément le comportement de la fonction  $d_n$  autour de l'instant de cutoff.

**Proposition 2.2.1** *Le  $n$ -échantillon  $(X_1, \dots, X_n)$  présente un phénomène de cutoff à l'instant  $t_n$  défini par :*

$$t_n = \frac{\log(n)}{2\rho} .$$

De plus, pour tout  $n$  assez grand, et pour tout réel  $u$  tel que  $t_n + u > 0$ , on a :

$$1 - \exp\left(-\frac{x_0^2}{8\sigma^2}e^{-2\rho u}\right) + \varepsilon_n(u) \leq d_n(t_n + u) \quad (2.2.3)$$

et :

$$d_n(t_n + u) \leq \sqrt{1 - \exp\left(-\frac{x_0^2}{4\sigma^2}e^{-2\rho u}\right)} + \varepsilon'_n(u) , \quad (2.2.4)$$

où  $\varepsilon_n$  et  $\varepsilon'_n$  sont des fonctions convergeant simplement vers 0.

Cette proposition précise le comportement de la convergence du  $n$ -échantillon vers sa loi stationnaire autour de l'instant de cutoff  $t_n$  ; en effet elle montre que le cutoff se produit dans un intervalle de temps de longueur  $O(1)$ , alors que la définition 1.2.1 exige seulement qu'il se produise dans un intervalle de longueur  $O(\log(n))$ . Nous verrons la démonstration de ce point dans la partie 3.2 du chapitre 3 (proposition 3.2.6).

La démonstration de la proposition 2.2.1 repose sur la comparaison entre la distance en variation totale et la distance de Hellinger. En effet, la distance en variation totale ne se prête pas aux calculs explicites sur les lois normales multidimensionnelles, contrairement à la distance de Hellinger. Les propriétés de cette distance sont exposées dans la partie 3.1.2 du chapitre 3. Nous ne rappelons ici que la définition : la distance de Hellinger entre deux lois de probabilité  $P$  et  $Q$  de densités respectives  $p$  et  $q$  par rapport à la mesure  $\lambda$  est donnée par :

$$d_H(P, Q) = \left(\frac{1}{2} \int_{\Omega} (\sqrt{p} - \sqrt{q})^2 d\lambda\right)^{\frac{1}{2}} .$$

De plus il est possible de calculer exactement la distance de Hellinger entre deux lois normales  $n$ -dimensionnelles. Soit  $\mathcal{N}(m_1, v_1^2)^{\otimes n}$  et  $\mathcal{N}(m_2, v_2^2)^{\otimes n}$  deux telles lois. Nous avons alors (proposition 3.1.15) :

$$d_H^2(\mathcal{N}(m_1, v_1^2)^{\otimes n}, \mathcal{N}(m_2, v_2^2)^{\otimes n}) = 1 - \left(\frac{2v_1v_2}{v_1^2 + v_2^2}\right)^{\frac{n}{2}} \exp\left(-\frac{n(m_1 - m_2)^2}{4(v_1^2 + v_2^2)}\right) . \quad (2.2.5)$$

Enfin, la distance en variation totale est contrôlée par la distance de Hellinger (proposition 3.1.13). Si  $P$  et  $Q$  sont deux lois de probabilité, alors :

$$d_H^2(P, Q) \leq d_{VT}(P, Q) \leq d_H(P, Q) \sqrt{2 - d_H^2(P, Q)}. \quad (2.2.6)$$

Venons-en à la démonstration de la proposition 2.2.1.

### Démonstration.

Il faut commencer par remarquer qu'il suffit de démontrer les inégalités (2.2.3) et (2.2.4) pour obtenir la présence du cutoff en  $t_n$ . Ceci découle principalement de la décroissance de la fonction  $d_n$  (proposition 3.1.5 du chapitre 3).

À l'instant  $t$ , le  $n$ -échantillon  $(X_1, \dots, X_n)$  suit la loi  $\mathcal{N}(x_0 e^{-\rho t}, \sigma^2(1 - e^{-2\rho t}))^{\otimes n}$ , et la loi stationnaire est la loi  $\mathcal{N}(0, \sigma^2)^{\otimes n}$ . Nous pouvons donc calculer la distance de Hellinger entre ces lois grâce à la formule (2.2.5). Ensuite nous encadrons la distance en variation totale par la distance de Hellinger (inégalité (2.2.6) ci-dessus), et nous étudions le comportement asymptotique de chacune des deux bornes quand  $n$  tend vers l'infini pour obtenir le résultat.  $\square$

Dans la partie suivante, nous allons voir que la moyenne empirique du  $n$ -échantillon présente un phénomène de cutoff au même instant.

### 2.2.2 Cutoff pour la moyenne empirique

Considérons toujours un  $n$ -échantillon de processus d'Ornstein-Uhlenbeck  $(X_1, \dots, X_n)$  partant tous du même point  $x_0$ , de mêmes paramètres  $\rho$  et  $\sigma$  strictement positifs. À partir de ce  $n$ -échantillon, nous construisons le processus moyen :

$$M_n(t) = \frac{1}{n} \sum_{i=1}^n X_i(t).$$

Dans ce cadre particulier des processus d'Ornstein-Uhlenbeck, le processus  $M_n$  est lui-même un processus d'Ornstein-Uhlenbeck.

En effet, la linéarité de l'équation différentielle stochastique vérifiée par le processus d'Ornstein-Uhlenbeck implique que toute combinaison linéaire de processus d'Ornstein-Uhlenbeck indépendants et ayant le même coefficient de dérive  $\rho$  est encore un processus d'Ornstein-Uhlenbeck.

**Proposition 2.2.2** Soient  $\rho, \sigma_1, \dots, \sigma_n$   $n + 1$  réels strictement positifs, et  $x_o^1, \dots, x_o^n$   $n$  réels quelconques. Pour chaque  $i$  de  $\{1, \dots, n\}$ , soit  $X_i$  un processus d'Ornstein-Uhlenbeck partant de  $x_o^i$ , de paramètres  $\rho$  and  $\sigma_i$ . Supposons de plus que  $X_1, \dots, X_n$  sont indépendants, et définissons, pour  $n$  réels  $\lambda_1, \dots, \lambda_n$ , le processus  $Y_n$  comme la combinaison linéaire suivante des  $X_i$  :

$$Y_n = \sum_{i=1}^n \lambda_i X_i.$$

Alors  $Y_n$  est un processus d'Ornstein-Uhlenbeck partant de  $\sum_{i=1}^n \lambda_i x_o^i$ , de paramètres  $\rho$  et  $\left(\sum_{i=1}^n \lambda_i^2 \sigma_i^2\right)^{\frac{1}{2}}$ .

**Démonstration.**

Pour chaque  $i$  de  $\{1, \dots, n\}$  le processus  $X_i$  est solution de l'équation différentielle stochastique suivante :

$$\begin{cases} dX_i(t) &= \sigma_i \sqrt{2\rho} dB_i(t) - \rho X_i(t) dt \\ X(0) &= x_o^i \end{cases},$$

où  $B_1, \dots, B_n$  sont  $n$  mouvements browniens standard indépendants.

On en déduit que  $Y_n$  satisfait l'équation différentielle stochastique suivante :

$$\begin{cases} dY_n(t) &= \sqrt{2\rho} \sum_{i=1}^n \lambda_i \sigma_i dB_i(t) - \rho Y_n(t) dt \\ Y_n(0) &= \sum_{i=1}^n \lambda_i x_o^i \end{cases}.$$

Pour finir, on remarque que  $\sum_{i=1}^n \lambda_i \sigma_i B_i$  peut aussi s'écrire  $\left(\sum_{i=1}^n \lambda_i^2 \sigma_i^2\right)^{\frac{1}{2}} B$ , où  $B$  est un mouvement brownien standard.  $\square$

Comme corollaire immédiat de cette proposition, nous voyons que la moyenne arithmétique de  $n$  processus d'Ornstein-Uhlenbeck indépendants et de même loi est encore un processus d'Ornstein-Uhlenbeck.

**Corollaire 2.2.3** Soit  $\rho$  et  $\sigma$  deux réels strictement positifs, et  $x_0$  un réel quelconque. Soit  $(X_1, \dots, X_n)$  un  $n$ -échantillon de processus d'Ornstein-Uhlenbeck partant de  $x_0$  et de paramètres  $\rho$  et  $\sigma$ . Soit  $M_n$  le processus moyen associé à cet échantillon :

$$\forall t \geq 0, \quad M_n(t) = \frac{1}{n} \sum_{i=1}^n X_i(t).$$

Alors  $M_n$  est un processus d'Ornstein-Uhlenbeck partant de  $x_0$  et de paramètres  $\rho$  et  $\sigma/\sqrt{n}$ .

En utilisant le résultat (2.1.2) de convergence en loi des processus d'Ornstein-Uhlenbeck, nous en déduisons donc que, à  $n$  fixé, le processus  $M_n$  converge en loi quand  $t$  tend vers  $+\infty$  :

$$M_n(t) \xrightarrow[t \rightarrow +\infty]{loi} \mathcal{N}\left(0, \frac{\sigma^2}{n}\right).$$

Notons  $\nu_n$  la loi  $\mathcal{N}\left(0, \frac{\sigma^2}{n}\right)$ . De la même façon que dans la partie précédente, nous pouvons donc définir la fonction  $\bar{d}_n$ , à chaque instant  $t$ , comme la distance en variation totale entre la loi du processus  $M_n$  à l'instant  $t$  et sa loi stationnaire :

$$\bar{d}_n(t) = d_{VT}(\mathcal{L}(M_n(t)), \nu_n).$$

Le phénomène de cutoff pour le processus  $M_n$  est décrit dans la proposition suivante.

**Proposition 2.2.4** *La suite  $(t_n)_{n \in \mathbb{N}^*}$  est une suite d'instants de cutoff pour la convergence du processus  $M_n$  vers sa loi stationnaire. De plus, pour  $n$  assez grand et pour tout réel  $u$  tel que  $t_n + u > 0$ , on a :*

$$1 - \exp\left(-\frac{x_0^2}{8\sigma^2}e^{-2\rho u}\right) + \bar{\varepsilon}_n(u) \leq \bar{d}_n(t_n + u) \quad (2.2.7)$$

et :

$$\bar{d}_n(t_n + u) \leq \sqrt{1 - \exp\left(-\frac{x_0^2}{4\sigma^2}e^{-2\rho u}\right)} + \bar{\varepsilon}'_n(u), \quad (2.2.8)$$

où  $\bar{\varepsilon}_n$  et  $\bar{\varepsilon}'_n$  sont des fonctions convergeant simplement vers 0.

### Démonstration.

Elle est semblable en tout point à la démonstration de la proposition 2.2.1. Nous commençons par démontrer les inégalités (2.2.7) et (2.2.8). Pour cela, nous connaissons la loi de la variable aléatoire  $M_n(t)$  : c'est la loi normale  $\mathcal{N}(x_0 e^{-\rho t}, \sigma^2(1 - e^{-2\rho t})/\sqrt{n})$ . La loi stationnaire est la loi normale  $\mathcal{N}(0, \sigma^2/\sqrt{n})$ . Donc, à l'aide de la formule (2.2.5), nous pouvons calculer la distance de Hellinger entre ces deux lois. Ensuite nous utilisons l'encadrement (2.2.6) de la distance en variation totale par la distance de Hellinger, puis l'étude du comportement asymptotique des deux bornes donne le résultat annoncé.  $\square$

Ces deux phénomènes de cutoff, pour l'échantillon et sa moyenne, sont très similaires. Considérons la distance en variation totale ( $d_n$  ou  $\bar{d}_n$ ) comme une fonction du temps : c'est une fonction qui décroît de 1 à 0 (la décroissance sera démontrée dans la proposition 3.1.5). Elle peut donc être vue comme la fonction de survie d'une certaine variable aléatoire positive, concentrée autour de l'instant de cutoff  $\log(n)/(2\rho)$ . Cette variable aléatoire virtuelle s'interprète comme "l'instant où le processus atteint l'équilibre". Dans la partie suivante, nous allons voir comment trouver une approximation de cet instant virtuel, au moyen d'un temps d'atteinte.

## 2.3 Temps d'atteinte

Dans cette partie nous faisons le lien entre le phénomène de cutoff et certains temps d'atteinte particuliers. L'idée de relier ces deux notions est déjà apparue dans un article de B. Ycart [64] pour l'étude du cutoff d'un échantillon de chaînes de Markov à temps continu et espace d'états fini. Dans ce cas, l'échantillon est noté  $(X_1, \dots, X_n)$ , et converge vers sa loi stationnaire notée  $\nu$ . Dans l'article [64], l'auteur montre qu'un phénomène de cutoff se produit à l'instant  $\tau_n$  défini par :

$$\tau_n = \frac{\log(n)}{2\beta},$$

où  $\beta$  est le trou spectral du générateur infinitésimal du processus échantillonné. Comme  $\beta$  est inconnu en général, un estimateur en est proposé. Pour une certaine fonction  $\phi$  à valeurs réelles, la moyenne de l'image du  $n$ -échantillon par  $\phi$  est définie par :

$$M_n^\phi(t) = \frac{1}{n} \sum_{i=1}^n \phi(X_i(t)),$$



et le temps d'atteinte par ce processus de l'espérance de  $\phi$  sous la loi stationnaire est noté  $T_n$ . B. Ycart montre alors que :

$$\lim_{n \rightarrow +\infty} \frac{E(T_n)}{\tau_n} = 1 ,$$

et que :

$$\lim_{n \rightarrow +\infty} \frac{Var(T_n)}{E(T_n^2)} = 0 ,$$

ce qui montre que la variable aléatoire  $\log(n)/(2T_n)$  est un estimateur consistant de  $\beta$ .

Pour l'échantillon de processus d'Ornstein-Uhlenbeck, nous utilisons la même idée. Nous notons  $(X_1, \dots, X_n)$  le  $n$ -échantillon de processus, partant tous de  $x_0$  supposé non nul, de mêmes paramètres  $\rho$  et  $\sigma$ . La loi stationnaire d'une coordonnée est notée  $\nu$  ( $\nu$  est la loi normale centrée de variance  $\sigma^2$ ), et le processus moyen construit à partir de l'échantillon est noté  $M_n$  (ce qui correspond à prendre pour  $\phi$  l'identité).

La loi  $\nu$  est centrée, donc le temps d'atteinte qui nous intéresse est le premier instant où la moyenne empirique du processus atteint 0 :

$$T_0^{x_0, n} = \inf\{t \geq 0 ; M_n(t) = 0\} .$$

Nous avons vu dans la partie précédente (corollaire 2.2.3) que  $M_n$  est un processus d'Ornstein-Uhlenbeck. Or la densité du temps d'atteinte d'un niveau fixé par un processus d'Ornstein-Uhlenbeck est connue. Elle est utilisée dans de nombreuses applications. Par exemple dans [38], elle sert à déterminer une formule de pricing d'une certaine option. L'article [2] est une référence très complète sur le temps d'atteinte d'un niveau par un processus d'Ornstein-Uhlenbeck. Dans le cas général (quand le niveau à atteindre est différent de 0) la densité du temps d'atteinte est donnée à travers sa transformée de Laplace, et elle est explicite dans le cas où le niveau est 0. On peut la lire dans [46] ou dans [2], et elle est rappelée dans la proposition suivante.

**Proposition 2.3.1** *Soit  $X$  un processus d'Ornstein-Uhlenbeck partant d'un réel  $x_0$  non nul, de paramètres  $\rho$  et  $\sigma$ . Soit  $T_0$  le premier instant où  $X$  atteint 0 :*

$$T_0 = \inf\{t \geq 0 ; X(t) = 0\} .$$

Alors la variable aléatoire  $T_0$  admet la densité  $\psi$  sur  $\mathbb{R}$ , donnée par :

$$\psi(t) = \frac{|x_0|}{2\sigma\sqrt{\pi\rho}} \left( \frac{\rho}{\sinh(\rho t)} \right)^{\frac{3}{2}} \exp\left( -\frac{x_0^2 e^{-\rho t}}{4\sigma^2 \sinh(\rho t)} + \frac{\rho t}{2} \right) . \quad (2.3.9)$$

Le processus  $M_n$  part de  $x_0$ , et a pour paramètres  $\rho$  et  $\sigma/\sqrt{n}$ , donc, par application de la proposition 2.3.1 précédente, nous obtenons la densité  $g_n$  du temps d'atteinte  $T_0^{x_0, n}$  :

$$g_n(t) = \frac{|x_0|}{2\sigma} \sqrt{\frac{n}{\pi\rho}} \left( \frac{\rho}{\sinh(\rho t)} \right)^{\frac{3}{2}} \exp\left( -\frac{nx_0^2 e^{-\rho t}}{4\sigma^2 \sinh(\rho t)} + \frac{\rho t}{2} \right) .$$

De cette formule se déduit la convergence en loi de la variable aléatoire  $T_0^{x_0, n}$  quand  $n$  tend vers l'infini.

**Proposition 2.3.2** Soit  $U_n$  la renormalisation suivante de  $T_0^{x_0, n}$  :

$$U_n = \rho \left( T_0^{x_0, n} - \frac{\log(n)}{2\rho} + \frac{1}{2\rho} \log \left( \frac{2\sigma^2}{x_0^2} \right) \right) .$$

Quand  $n$  tend vers l'infini, la variable aléatoire  $U_n$  converge en loi vers la loi de densité  $g$  sur  $\mathbb{R}$ , avec :

$$g(t) = \frac{2}{\sqrt{\pi}} \exp(-t - e^{-2t}) .$$

La démonstration de cette proposition utilise le lemme de Scheffé ([9] p. 223), que l'on rappelle ci-dessous.

**Lemme 2.3.3** Soit  $(\Omega, \mathcal{A})$  un espace mesurable, et soit  $\mu$  une mesure sur cet espace. Soit  $(P_n)_{n \in \mathbb{N}}$  une suite de lois de probabilité de densités respectives  $(p_n)_{n \in \mathbb{N}}$  par rapport à  $\mu$ , et soit  $P$  une loi de probabilité de densité  $p$  par rapport à  $\mu$ . Si la suite  $(p_n)_{n \in \mathbb{N}}$  tend simplement vers  $p$   $\mu$ -presque partout quand  $n$  tend vers  $+\infty$ , alors :

$$\lim_{n \rightarrow +\infty} d_{VT}(P_n, P) = 0 .$$

Dans la suite, nous utiliserons ce lemme pour établir la convergence en loi d'une suite de variables aléatoires  $(\mathbf{X}_n)_{n \in \mathbb{N}}$  admettant chacune une densité  $p_n$  par rapport à la mesure de Lebesgue sur  $\mathbb{R}$  vers une variable aléatoire  $\mathbf{X}$  admettant aussi une densité  $p$  par rapport à la mesure de Lebesgue sur  $\mathbb{R}$ , en sachant seulement que la suite  $(p_n)_{n \in \mathbb{N}}$  converge simplement vers  $p$  en tout point de  $\mathbb{R}$ .

Notons en effet  $F_n$  et  $F$  les fonctions de répartition respectives de  $\mathbf{X}_n$  et de  $\mathbf{X}$  :

$$\forall x \in \mathbb{R}, \quad F_n(x) = \int_{-\infty}^x p_n(u) du \quad \text{et} \quad F(x) = \int_{-\infty}^x p(u) du .$$

Montrer la convergence en loi de  $(\mathbf{X}_n)_{n \in \mathbb{N}}$  vers  $\mathbf{X}$ , c'est montrer la convergence simple de  $(F_n)_{n \in \mathbb{N}}$  vers  $F$  en tout point de continuité de  $F$ .

Le lemme de Scheffé assure que la distance en variation totale entre la loi de  $\mathbf{X}_n$  et la loi de  $\mathbf{X}$  tend vers 0 quand  $n$  tend vers  $+\infty$ , ce qui, par définition de la distance en variation totale (définition 3.1.1) s'écrit :

$$\lim_{n \rightarrow +\infty} \sup_{A \in \mathcal{A}} \left| \int_A p_n(u) du - \int_A p(u) du \right| = 0 .$$

Soit  $x$  un point de continuité de  $F$ . On a :

$$|F_n(x) - F(x)| \leq \sup_{A \in \mathcal{A}} \left| \int_A p_n(u) du - \int_A p(u) du \right| ,$$

ce qui montre que  $(F_n(x))_{n \in \mathbb{N}}$  tend vers  $F(x)$  quand  $n$  tend vers  $+\infty$ , et assure la convergence en loi de la suite  $(\mathbf{X}_n)_{n \in \mathbb{N}}$  vers  $\mathbf{X}$ .

Nous donnons maintenant la démonstration de la proposition 2.3.2.

### Démonstration.

La densité en  $t$  de la variable aléatoire  $U_n$  est donnée par :

$$\frac{1}{\rho} g_n \left( \frac{t}{\rho} + \frac{\log(n)}{2\rho} - \frac{1}{2\rho} \log \left( \frac{2\sigma^2}{x_0^2} \right) \right) .$$

Un calcul simple montre que, pour tout  $t$  :

$$\lim_{n \rightarrow +\infty} \frac{1}{\rho} g_n \left( \frac{t}{\rho} + \frac{\log(n)}{2\rho} - \frac{1}{2\rho} \log \left( \frac{2\sigma^2}{x_0^2} \right) \right) = g(t) .$$

Ensuite, le lemme de Scheffé (lemme 2.3.3) permet de conclure que la variable aléatoire  $U_n$  converge en loi vers une variable aléatoire de densité  $g$ .  $\square$

La figure 2.1 montre l'allure de la densité limite  $g$ .

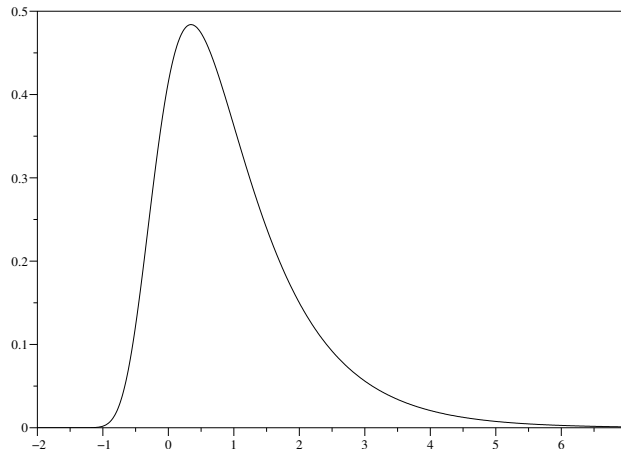


FIG. 2.1 – Graphe de la densité  $g$ .

Illustrons maintenant ce résultat de convergence en loi. Nous avons fixé successivement la valeur de  $n$  à 10, puis 100, puis 1000. Dans chaque cas, nous avons simulé 500 valeurs de la variable aléatoire  $T_0^{x_0, n}$  avec les paramètres suivants :

$$\rho = 1, \quad \sigma = 1 \quad \text{et} \quad x_0 = 5 .$$

Nous avons alors renormalisé ces 500 valeurs pour obtenir 500 valeurs de la variable aléatoire  $U_n$ . Les trois lignes de la figure 2.2 correspondent aux trois valeurs successives de  $n$ . La première colonne montre la superposition de l'histogramme avec le graphe de la densité limite  $g$  (en pointillés). La deuxième colonne montre la superposition de la fonction de répartition empirique de l'échantillon (trait continu) avec la fonction de répartition théorique de la loi limite (en pointillés). Dans chacun des trois cas, nous avons aussi calculé la  $p$ -valeur du test d'adéquation de Kolmogorov-Smirnov.

Rappelons la définition de la  $p$ -valeur. Supposons que le test statistique considéré soit basé sur la statistique de test  $T$ . Si  $t$  est une observation de  $T$ , alors la  $p$ -valeur associée à  $t$  est le plus petit niveau auquel l'hypothèse nulle  $H_0$  est rejetée. La connaissance de la  $p$ -valeur rend donc inutile le calcul préalable de la région de rejet : si  $p(t)$  est la  $p$ -valeur d'une observation  $t$  de la statistique  $T$  sous l'hypothèse nulle  $H_0$ , on obtient un test de niveau  $\alpha$  par la règle de rejet suivante :

$$\text{Rejet de } H_0 \iff p(T) < \alpha .$$

On pourra consulter le livre [10] et le site web de cours interactif de statistique SMEL<sup>1</sup> comme références sur ce sujet.

Ici nous avons noté la  $p$ -valeur  $\mathbf{p}$ , et nous l'avons fait figurer dans les trois cas. Les  $p$ -valeurs obtenues montrent qu'en prenant par exemple un risque de 5%, le test de Kolmogorov-Smirnov accepte l'hypothèse selon laquelle la distribution empirique s'ajuste à la distribution limite.

Cette convergence implique tout d'abord que la variable aléatoire  $T_0^{x_0, n}$  se concentre près de l'instant de cutoff  $t_n$  en termes de convergence en probabilité :

$$\lim_{n \rightarrow +\infty} \frac{T_0^{x_0, n}}{\frac{\log(n)}{2\rho}} = 1 \quad \text{en probabilité,}$$

de sorte que  $\log(n)/(2T_0^{x_0, n})$  est un estimateur consistant de  $\rho$  si  $\rho$  est inconnu.

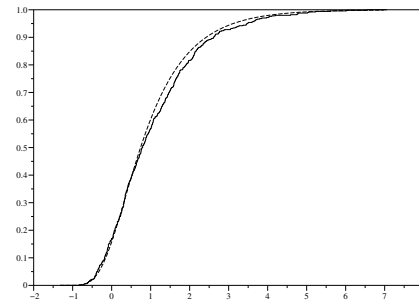
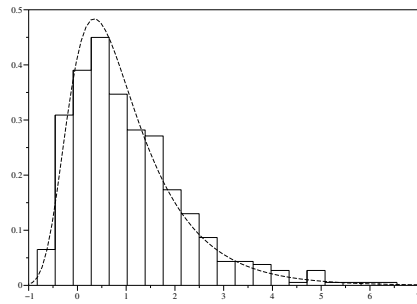
De plus, la loi limite de densité  $g$  est celle d'une variable aléatoire  $Y$  qui s'écrit comme le logarithme d'une gaussienne :

$$Y = -\log\left(\frac{|N|}{\sqrt{2}}\right) ,$$

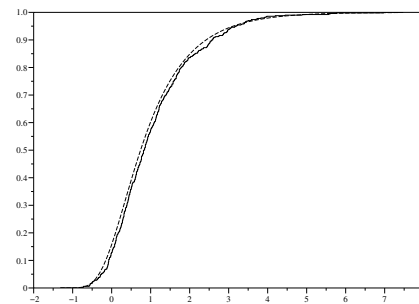
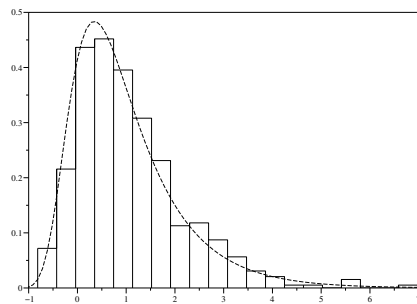
où  $N$  est une variable aléatoire gaussienne standard. La distribution de  $Y$  est donc asymétrique et se comporte très différemment en  $-\infty$  et  $+\infty$ . Plus précisément, regardons la fonction de répartition  $G$  de la variable aléatoire  $Y$ . Pour tout  $y$  de  $\mathbb{R}$ , nous avons :

$$G(y) = 2 \left( 1 - \Phi(\sqrt{2}e^{-y}) \right) ,$$

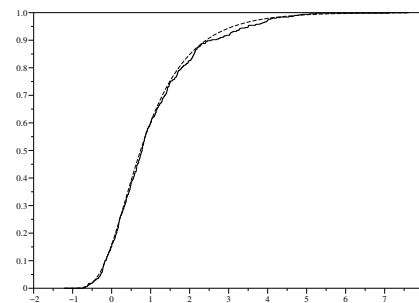
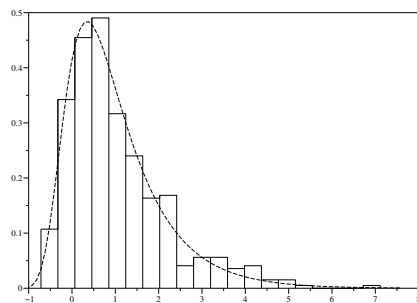
<sup>1</sup><http://www.math-info.univ-paris5.fr/smel>



$n = 10.$     $\mathbf{p} = 0.106.$



$n = 100.$     $\mathbf{p} = 0.204.$



$n = 1000.$     $\mathbf{p} = 0.801.$

FIG. 2.2 – Convergence en loi de  $T_0^{X_0, n}$ .

où  $\Phi$  désigne la fonction de répartition de la loi gaussienne standard. Or on connaît des équivalents de  $1 - \Phi$  en 0 et en  $+\infty$  ([27] Chap.VII p. 175) :

$$1 - \Phi(x) \underset{x \rightarrow 0}{\sim} \frac{1}{2} - \frac{x}{\sqrt{2\pi}},$$

et :

$$1 - \Phi(x) \underset{x \rightarrow +\infty}{\sim} \frac{1}{x\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Donc on obtient des équivalents de  $G$  en  $-\infty$  et en  $+\infty$  :

$$G(y) \underset{y \rightarrow -\infty}{\sim} \frac{1}{\sqrt{\pi}} \exp(y - e^{-2y}),$$

et :

$$G(y) \underset{y \rightarrow +\infty}{\sim} 1 - \frac{2}{\sqrt{\pi}} e^{-y}.$$

En comparant ces équivalents avec le comportement de la distance en variation totale à gauche et à droite de l'instant de cutoff, on constate que les ordres de grandeur sont semblables. En effet, commençons par regarder à droite de l'instant de cutoff. Le majorant de la distance en variation totale est équivalent, quand  $u$  tend vers  $+\infty$  à :

$$\frac{|x_0|}{2\sigma} e^{-\rho u}.$$

De plus la queue de distribution droite de  $T_0^{x_0, n}$  peut être estimée à l'aide de la convergence en loi de  $T_0^{x_0, n} - t_n$  :

$$\lim_{n \rightarrow +\infty} P(T_0^{x_0, n} > t_n + u) = 1 - G\left(\rho u + \log\left(\frac{\sigma\sqrt{2}}{|x_0|}\right)\right).$$

Grâce à l'équivalent de  $G$  au voisinage de  $+\infty$ , on obtient :

$$1 - G\left(\rho u + \log\left(\frac{\sigma\sqrt{2}}{|x_0|}\right)\right) \underset{u \rightarrow +\infty}{\sim} \frac{|x_0|}{\sigma} \sqrt{\frac{2}{\pi}} e^{-\rho u}.$$

Donc la queue de distribution droite de  $T_0^{x_0, n}$  est du même ordre de grandeur que la distance en variation totale à droite de l'instant de cutoff.

À gauche de cet instant, on peut faire le même genre de raisonnement. Le minorant de la distance en variation totale prise à l'instant  $t_n - u$  avec  $u$  positif est :

$$1 - \exp\left(-\frac{x_0^2}{8\sigma^2} e^{2\rho u}\right).$$

Par ailleurs, la queue de distribution gauche de  $T_0^{x_0, n}$  peut être estimée à l'aide de la convergence en loi de  $T_0^{x_0, n} - t_n$  :

$$\lim_{n \rightarrow +\infty} P(T_0^{x_0, n} < t_n - u) = G \left( -\rho u + \log \left( \frac{\sigma \sqrt{2}}{|x_0|} \right) \right) .$$

Grâce à l'équivalent de  $G$  au voisinage de  $-\infty$ , on obtient :

$$G \left( -\rho u + \log \left( \frac{\sigma \sqrt{2}}{|x_0|} \right) \right) \underset{u \rightarrow +\infty}{\sim} \frac{\sigma}{|x_0|} \sqrt{\frac{2}{\pi}} \exp \left( -\rho u - \frac{x_0^2}{2\sigma^2} e^{2\rho u} \right) .$$

Du côté gauche on voit un peu moins bien la ressemblance des deux comportements, mais les allures générales sont les mêmes : à gauche, la fonction est une composée de deux exponentielles, alors qu'à droite c'est une exponentielle simple.

## 2.4 Simulations

Terminons ce chapitre par une illustration numérique du lien entre le phénomène de cutoff du  $n$ -échantillon et le temps d'atteinte de 0 par sa moyenne empirique. Plus précisément, nous voulons illustrer le fait que le  $n$ -échantillon de processus atteint son régime stationnaire à l'instant de cutoff.

Nous commençons par fixer les paramètres du processus d'Ornstein-Uhlenbeck que nous allons échantillonner :

$$\rho = 1, \quad \sigma = 1 \quad \text{et} \quad x_0 = 10 .$$

Nous fixons aussi la taille de l'échantillon :  $n = 100$  puis  $n = 1000$ , et une suite d'instantanés régulièrement espacés  $t_1, \dots, t_k$  (nous choisissons  $t_i = i/100$ ). Une expérience consiste alors à simuler un  $n$ -échantillon de processus d'Ornstein-Uhlenbeck  $(X_1, \dots, X_n)$  avec ces paramètres. Nous savons qu'à chaque instant  $t_i$ , le  $n$ -uplet  $(X_1(t_i), \dots, X_n(t_i))$  est un échantillon gaussien. Nous pouvons donc tester si sa moyenne est nulle ; nous réalisons le test en supposant que sa variance est inconnue, et nous l'estimons par la variance empirique. Nous récupérons donc, à chaque instant  $t_i$ , la  $p$ -valeur du test. En même temps, nous calculons le premier instant où la moyenne du  $n$ -échantillon vaut 0. Nous obtenons donc une observation de la variable aléatoire  $T_0^{x_0, n}$ .

Ensuite nous répétons 1000 fois cette expérience. Pour chaque  $t_i$ , nous obtenons donc un échantillon de taille 1000 de  $p$ -valeurs ; de plus nous avons 1000 observations de la variable aléatoire  $T_0^{x_0, n}$ . Sur les figures 2.3, nous avons associé à chaque abscisse  $t_i$  la moyenne des 1000  $p$ -valeurs obtenues ; la première figure est établie dans le cas où  $n$  vaut 100, et la seconde dans le cas où  $n$  vaut 1000.

Nous pouvons constater que la  $p$ -valeur moyenne du test devient strictement positive à partir de l'instant de cutoff. Cela revient à n'accepter l'hypothèse que le  $n$ -échantillon suit la loi stationnaire qu'à un niveau très faible. Il faut attendre environ deux fois le temps de cutoff pour

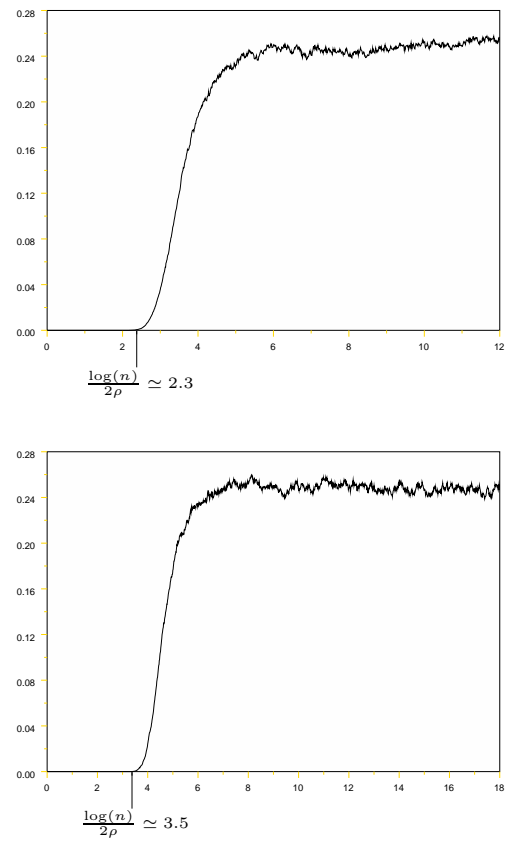


FIG. 2.3 – Moyenne des  $p$ -valeurs du test de nullité de la moyenne d'un échantillon gaussien, obtenues pour  $n = 100$  et  $n = 1000$ .



que le processus soit stabilisé à son régime stationnaire (quand la  $p$ -valeur se stabilise autour de 25%). Ceci illustre le fait que le cutoff permet de détecter le moment où le processus converge vers sa loi stationnaire.

Par ailleurs, avec l'échantillon d'observations de la variable aléatoire  $T_0^{x_0, n}$ , nous pouvons illustrer la convergence en loi de la proposition 2.3.2. En effet, nous calculons la version normalisée de l'échantillon :

$$U_n = \rho \left( T_0^{x_0, n} - \frac{\log(n)}{2\rho} + \frac{1}{2\rho} \log \left( \frac{2\sigma^2}{x_0^2} \right) \right) .$$

Nous choisissons comme test d'adéquation le test de Kolmogorov-Smirnov. Il permet de décider si cet échantillon est distribué selon la loi limite de densité  $g$ . Le tableau 2.1 donne les  $p$ -valeurs obtenues, dans le cas où  $n$  vaut 100, puis dans celui où  $n$  vaut 1000.

$n$	$p$ -valeur
100	0.0037
1000	0.024

TAB. 2.1 –  $p$ -valeur du test de Kolmogorov-Smirnov entre le  $n$ -échantillon de  $U_n$  et la loi limite de densité  $g$ .

En conclusion, nous pouvons dire que l'exemple du processus d'Ornstein-Uhlenbeck illustre le lien étroit qu'il y a entre le phénomène de cutoff dans la convergence d'un  $n$ -échantillon vers sa loi stationnaire et l'étude du temps d'atteinte construit à partir de la moyenne empirique de ce  $n$ -échantillon. Dans les chapitres qui suivent, nous entreprendrons la même démarche pour étudier, en premier lieu, le phénomène de cutoff pour des échantillons de processus exponentiellement convergents, puis pour faire le lien entre ce phénomène et le temps d'atteinte d'un niveau par la moyenne de l'échantillon.

# Chapitre 3

## Le cutoff

Ce chapitre est consacré à l'étude du phénomène de cutoff pour des échantillons de processus tels que le processus échantillonné converge à vitesse exponentielle vers sa mesure stationnaire. Nous commençons par des rappels sur différentes distances entre mesures de probabilité qui sont utiles par la suite (partie 3.1). Dans la partie 3.2 nous traitons le phénomène de cutoff pour les échantillons de processus exponentiellement convergents. Enfin, nous terminons par des exemples (partie 3.3).

### 3.1 Distances entre mesures de probabilité

Pour comprendre le phénomène de cutoff il est nécessaire de connaître différentes distances entre mesures de probabilité. Nous en utiliserons quatre : la distance en variation totale, la distance de Hellinger, la distance du  $\chi^2$  et la distance de Kullback. Bien que les deux dernières ne soient pas des distances au sens topologique du terme (il manque la propriété de symétrie notamment), nous conserverons l'appellation de distance dans leur cas, d'autant plus que ce sont les relations existant entre ces distances qui nous intéressent. Ce sujet est largement abordé dans la littérature ; cependant, les définitions diffèrent souvent d'un ouvrage à l'autre. Dans le cas des distances en variation totale et de Hellinger, où les définitions diffèrent d'une constante multiplicative, nous avons choisi de fixer la constante pour que la distance vaille 1 entre deux mesures étrangères. Les références principales de cette partie sont les livres de D. Pollard [47] et de R.D. Reiss [49]. L'article de A.L. Gibbs et F.E. Su [30] est aussi une référence très complète sur les relations entre les différentes distances.

#### 3.1.1 La distance en variation totale

Il s'agit de la distance la plus utilisée en probabilités. Un chapitre lui est consacré dans le livre de L. Devroye et G. Lugosi [18]. Rappelons sa définition :

**Définition 3.1.1** Soient  $\mu_1$  et  $\mu_2$  deux mesures sur l'espace  $(\Omega, \mathcal{A})$ . La distance en variation totale entre  $\mu_1$  et  $\mu_2$  est définie de la façon suivante :

$$d_{VT}(\mu_1, \mu_2) = \sup_{A \in \mathcal{A}} |\mu_1(A) - \mu_2(A)| \quad (3.1.1)$$

Donnons quelques propriétés simples de cette distance. La propriété suivante montre que la distance en variation totale entre deux mesures de probabilité est toujours comprise entre 0 et 1.

**Proposition 3.1.2** Soient  $P$  et  $Q$  deux mesures de probabilité sur l'espace  $(\Omega, \mathcal{A})$ .

- La distance en variation totale est une distance, et en particulier,  $d_{VT}(P, Q) = 0$  si et seulement si  $P = Q$ .
- Les mesures  $P$  et  $Q$  sont étrangères (c'est-à-dire à supports disjoints) si et seulement si  $d_{VT}(P, Q) = 1$ .

Un des principaux avantages de la distance en variation totale est énoncé dans la proposition suivante; elle montre que la distance peut se calculer directement à partir des densités des mesures de probabilité considérées.

**Proposition 3.1.3** Soient  $P$  et  $Q$  deux mesures de probabilité sur l'espace  $(\Omega, \mathcal{A})$  dominées par une même mesure  $\mu$ . Soient  $p$  et  $q$  leurs densités respectives par rapport à  $\mu$  :

$$p = \frac{dP}{d\mu} \quad \text{et} \quad q = \frac{dQ}{d\mu} .$$

Alors :

$$d_{VT}(P, Q) = \frac{1}{2} \int_{\Omega} |p - q| d\mu \quad (3.1.2)$$

$$= \frac{1}{2} \sup \left\{ \left| \int_{\Omega} f dP - \int_{\Omega} f dQ \right| ; \|f\|_{\infty} \leq 1 \right\} . \quad (3.1.3)$$

La proposition précédente montre que la distance en variation totale est directement reliée à la distance  $L^1$  entre les densités. De plus, au vu de la définition 3.1.1 de la distance en variation totale, on constate que  $d_{VT}(P, Q)$  ne dépend pas de la mesure choisie pour dominer  $P$  et  $Q$ .

### Démonstration.

Nous démontrons d'abord que :

$$d_{VT}(P, Q) = \frac{1}{2} \int_{\Omega} |p - q| d\mu .$$

- Pour cela on montre en premier lieu que :  $d_{VT}(P, Q) \leq \frac{1}{2} \int_{\Omega} |p - q| d\mu$ .

Soit  $A$  un élément de  $\mathcal{A}$ . En notant  $\bar{A}$  le complémentaire de  $A$ , on a :  $|P(A) - Q(A)| = |P(\bar{A}) - Q(\bar{A})|$ . Alors :

$$\begin{aligned} 2 |P(A) - Q(A)| &= |P(A) - Q(A)| + |P(\bar{A}) - Q(\bar{A})| \\ &= \left| \int_A (p - q) d\mu \right| + \left| \int_{\bar{A}} (p - q) d\mu \right| \\ &\leq \int_A |p - q| d\mu + \int_{\bar{A}} |p - q| d\mu \\ &= \int_{\Omega} |p - q| d\mu . \end{aligned}$$

- Ensuite on démontre que :

$$\int_{\{p < q\}} (q - p) d\mu = \int_{\{p > q\}} (p - q) d\mu = \frac{1}{2} \int_{\Omega} |p - q| d\mu .$$

Comme  $P$  et  $Q$  sont des probabilités, on sait que :

$$\int_{\Omega} p d\mu = 1 = \int_{\Omega} q d\mu ,$$

donc on a :

$$0 = \int_{\Omega} (p - q) d\mu = \int_{\{p > q\}} (p - q) d\mu + \int_{\{p < q\}} (p - q) d\mu ,$$

et donc :

$$\int_{\{p > q\}} (p - q) d\mu = \int_{\{p < q\}} (q - p) d\mu .$$

Alors :

$$\begin{aligned} \int_{\Omega} |p - q| d\mu &= \int_{\{p > q\}} (p - q) d\mu + \int_{\{p < q\}} (q - p) d\mu \\ &= 2 \int_{\{p > q\}} (p - q) d\mu \\ &= 2 \int_{\{p < q\}} (q - p) d\mu . \end{aligned}$$

- Finalement on démontre que :  $d_{VT}(P, Q) \geq \frac{1}{2} \int_{\Omega} |p - q| d\mu$ .

Définissons  $A_o = \{p > q\}$ . Alors :

$$\begin{aligned} d_{VT}(P, Q) &\geq |P(A_o) - Q(A_o)| \\ &= \left| \int_{\{p > q\}} (p - q) d\mu \right| \\ &= \int_{\{p > q\}} (p - q) d\mu \\ &= \frac{1}{2} \int_{\Omega} |p - q| d\mu . \end{aligned}$$

Pour finir, démontrons l'égalité suivante :

$$\int_{\Omega} |p - q| d\mu = \sup \left\{ \left| \int_{\Omega} f dP - \int_{\Omega} f dQ \right| ; \|f\|_{\infty} \leq 1 \right\} .$$

Soit  $\varphi$  une fonction mesurable définie sur  $\Omega$  et à valeurs dans  $\mathbb{R}$ . Commençons par remarquer que, pour toute fonction  $f$  mesurable de norme infinie inférieure ou égale à 1, on a :

$$\left| \int_{\Omega} \varphi f d\mu \right| \leq \int_{\Omega} |\varphi| d\mu ,$$

et que l'égalité a lieu pour la fonction  $f = \mathbb{I}_{\varphi \geq 0} - \mathbb{I}_{\varphi < 0}$ . On en déduit que :

$$\int_{\Omega} |\varphi| d\mu = \sup \left\{ \left| \int_{\Omega} \varphi f d\mu \right| ; \|f\|_{\infty} \leq 1 \right\} .$$

L'égalité cherchée est obtenue en appliquant ce résultat à la fonction  $|p - q|$ .  $\square$

La proposition 3.1.3 a un corollaire immédiat, qui permet de calculer la distance en variation totale entre deux lois de probabilité dont l'une est absolument continue par rapport à l'autre.

**Corollaire 3.1.4** *Soient  $P$  et  $Q$  deux lois de probabilité telles que  $P$  est absolument continue par rapport à  $Q$ . Notons  $f$  la densité de  $P$  par rapport à  $Q$ . Alors :*

$$d_{VT}(P, Q) = \frac{1}{2} \int_{\Omega} |f - 1| dQ .$$

Voici maintenant une proposition qui justifie en grande partie l'utilisation de la distance en variation totale dans le phénomène de cutoff.

**Proposition 3.1.5** *Soit  $X = \{X(t); t \geq 0\}$  un processus de Markov de loi stationnaire  $\nu$ . Alors, pour tout réel  $x$ , la fonction qui à  $t$  associe la distance en variation totale entre la loi de  $X(t)$  sachant  $X(0) = x$  et la loi  $\nu$  est décroissante.*

**Démonstration.**

Notons  $P$  le semi-groupe de transition du processus  $X$ . Pour tout  $t$  positif, l'opérateur  $P_t$  est défini de l'ensemble des fonctions mesurables bornées  $\mathcal{C}_b$  dans lui-même par :

$$\forall f \in \mathcal{C}_b, \forall x \in \mathbb{R}, P_t f(x) = E[f(X(t)) | X(0) = x] .$$

L'opérateur  $P_t$  est de norme 1, au sens où :

$$\sup \{ \|P_t f\|_{\infty} ; \|f\|_{\infty} \leq 1 \} = 1 .$$

Notons  $\nu f$  l'intégrale de la fonction  $f$  par rapport à la mesure  $\nu$  :

$$\nu f = \int f d\nu .$$

D'après la formule (3.1.3), la distance en variation totale entre la loi de  $X(t)$  sachant  $X(0) = x$  et la loi  $\nu$  s'écrit :

$$d_{VT}(\mathcal{L}(X(t)), \nu) = \sup \{ |P_t f(x) - \nu f| ; \|f\|_{\infty} \leq 1 \} .$$

Or, pour  $s$  positif, et pour une fonction  $f$  mesurable bornée de norme infinie inférieure à 1, la fonction  $P_s f$  est aussi de norme infinie inférieure à 1, donc :

$$|P_t(P_s f)(x) - \nu(P_s f)| \leq d_{VT}(\mathcal{L}(X(t)), \nu) .$$

Comme  $X$  est un processus de Markov de loi stationnaire  $\nu$ , on sait que :

$$P_t P_s = P_{t+s} \quad \text{et} \quad \nu(P_s f) = \nu f ,$$

ce qui permet de conclure que :

$$d_{VT}(\mathcal{L}(X(t+s)), \nu) \leq d_{VT}(\mathcal{L}(X(t)), \nu) .$$

□

On aura besoin dans la suite de relier la distance en variation totale entre mesures produits avec la distance entre les coordonnées. Commençons par donner une propriété simple qui nous sera utile par la suite, avant d'énoncer la propriété plus générale sur la distance en variation totale entre deux produits tensoriels de mesures.

**Proposition 3.1.6** *Soit  $P$  une loi de probabilité définie sur l'espace  $(\Omega, \mathcal{A})$ , et soient  $Q_1$  et  $Q_2$  deux lois de probabilité définies sur le même espace  $(\Omega', \mathcal{A}')$ . Alors :*

$$d_{VT}(P \otimes Q_1, P \otimes Q_2) = d_{VT}(Q_1, Q_2) .$$

**Démonstration.**

L'espace  $\Omega \times \Omega'$  est muni de la tribu notée  $\mathcal{A} \otimes \mathcal{A}'$  engendrée par les pavés  $A \times A'$ , où  $A \in \mathcal{A}$  et  $A' \in \mathcal{A}'$ . Commençons par montrer que :

$$d_{VT}(P \otimes Q_1, P \otimes Q_2) \leq d_{VT}(Q_1, Q_2) . \quad (3.1.4)$$

Soit  $A \times A'$  un pavé de  $\mathcal{A} \otimes \mathcal{A}'$ . Alors :

$$\begin{aligned} |P \otimes Q_1(A \times A') - P \otimes Q_2(A \times A')| &= |P(A)| |Q_1(A') - Q_2(A')| \\ &\leq |Q_1(A') - Q_2(A')| \\ &\leq d_{VT}(Q_1, Q_2) , \end{aligned}$$

ce qui montre l'inégalité (3.1.4).

Pour l'inégalité dans l'autre sens, fixons  $\varepsilon > 0$ . Soit ensuite  $A'$  un élément de  $\mathcal{A}'$  tel que :

$$|Q_1(A') - Q_2(A')| = d_{VT}(Q_1, Q_2) - \varepsilon .$$

On peut alors écrire :

$$\begin{aligned} d_{VT}(P \otimes Q_1, P \otimes Q_2) &\geq |P \otimes Q_1(\Omega \times A') - P \otimes Q_2(\Omega \times A')| \\ &= |Q_1(A') - Q_2(A')| \\ &= d_{VT}(Q_1, Q_2) - \varepsilon . \end{aligned}$$

Comme cette inégalité est vraie pour tout  $\varepsilon > 0$ , on en déduit que :

$$d_{VT}(P \otimes Q_1, P \otimes Q_2) \geq d_{VT}(Q_1, Q_2) ,$$

ce qui achève la démonstration. □

Pour relier la distance entre les mesures produits aux distances entre les coordonnées, le mieux que l'on sache faire est de donner des inégalités, comme le montre la proposition suivante.

**Proposition 3.1.7** Soit  $n$  un entier strictement positif. Pour tout entier  $i$  compris entre 1 et  $n$ , soient  $P_i$  et  $Q_i$  deux mesures de probabilité définies sur le même espace  $(\Omega_i, \mathcal{A}_i)$ . Soient  $P$  et  $Q$  les probabilités produits suivantes :

$$P = \bigotimes_{i=1}^n P_i \quad \text{et} \quad Q = \bigotimes_{i=1}^n Q_i .$$

Alors la distance en variation totale entre  $P$  et  $Q$  est encadrée par :

$$1 - 2 \exp \left( -\frac{1}{2n} \left( \sum_{i=1}^n d_{VT}(P_i, Q_i) \right)^2 \right) \leq d_{VT}(P, Q) \leq \sum_{i=1}^n d_{VT}(P_i, Q_i) . \quad (3.1.5)$$

### Démonstration.

On démontre successivement les deux inégalités. L'inégalité de droite est une conséquence directe de la proposition 3.1.6. Prenons  $n = 2$  pour simplifier. Une récurrence immédiate permet ensuite d'en déduire l'inégalité pour tout  $n$ . La première étape consiste à utiliser l'inégalité triangulaire, et la seconde à appliquer la proposition 3.1.6 :

$$\begin{aligned} d_{VT}(P_1 \otimes P_2, Q_1 \otimes Q_2) &\leq d_{VT}(P_1 \otimes P_2, Q_1 \otimes P_2) + d_{VT}(Q_1 \otimes P_2, Q_1 \otimes Q_2) \\ &\leq d_{VT}(P_1, Q_1) + d_{VT}(P_2, Q_2) . \end{aligned}$$

L'inégalité de gauche repose sur l'inégalité de Hoeffding (cf. [45] p. 58). On utilise ici la formule (3.1.3) de la distance en variation totale. Pour  $i$  compris entre 1 et  $n$ , soit  $f_i$  une fonction mesurable définie sur  $\Omega_i$  et à valeurs dans  $\mathbb{R}$ , de norme infinie inférieure à 1. Sans perdre de généralité, on suppose de plus que :

$$\sum_{i=1}^n \int_{\Omega_i} f_i dP_i \geq \sum_{i=1}^n \int_{\Omega_i} f_i dQ_i .$$

Soit ensuite  $A$  le sous-ensemble mesurable de  $\Omega$  suivant :

$$A = \left\{ (\omega_1, \dots, \omega_n) \in \Omega ; \sum_{i=1}^n f_i(\omega_i) \geq \frac{1}{2} \sum_{i=1}^n \int_{\Omega_i} f_i (dP_i + dQ_i) \right\} .$$

Soit  $(X_1, \dots, X_n)$  et  $(Y_1, \dots, Y_n)$  deux  $n$ -uplets de variables aléatoires indépendantes telles que, pour tout  $i$  compris entre 1 et  $n$ , la variable aléatoire  $X_i$  suit la loi  $P_i$  et la variable aléatoire  $Y_i$  suit la loi  $Q_i$ . On peut alors évaluer  $P(A)$  et  $Q(A)$  :

$$P(A) = P \left[ \sum_{i=1}^n (f_i(X_i) - E(f_i(X_i))) \geq -\frac{1}{2} \sum_{i=1}^n \int_{\Omega_i} f_i (dP_i - dQ_i) \right] ,$$

et

$$Q(A) = Q \left[ \sum_{i=1}^n (f_i(Y_i) - E(f_i(Y_i))) \geq \frac{1}{2} \sum_{i=1}^n \int_{\Omega_i} f_i (dP_i - dQ_i) \right] .$$

Comme les variables aléatoires  $f_1(X_1), \dots, f_n(X_n)$  sont indépendantes et prennent leurs valeurs entre  $-1$  et  $1$ , on peut appliquer l'inégalité de Hoeffding pour minorer  $P(A)$  :

$$P(A) \geq 1 - \exp \left( -\frac{1}{2n} \left( \frac{1}{2} \sum_{i=1}^n \int_{\Omega_i} f_i (dP_i - dQ_i) \right)^2 \right).$$

De même pour les variables  $f_1(Y_1), \dots, f_n(Y_n)$ , on trouve :

$$Q(A) \leq \exp \left( -\frac{1}{2n} \left( \frac{1}{2} \sum_{i=1}^n \int_{\Omega_i} f_i (dP_i - dQ_i) \right)^2 \right).$$

On en déduit donc que :

$$P(A) - Q(A) \geq 1 - 2 \exp \left( -\frac{1}{2n} \left( \frac{1}{2} \sum_{i=1}^n \int_{\Omega_i} f_i (dP_i - dQ_i) \right)^2 \right),$$

puis que :

$$d_{VT}(P, Q) \geq 1 - 2 \exp \left( -\frac{1}{2n} \left( \frac{1}{2} \sum_{i=1}^n \int_{\Omega_i} f_i (dP_i - dQ_i) \right)^2 \right).$$

Enfin, comme cette inégalité est vraie pour tout  $n$ -uplet  $(f_1, \dots, f_n)$ , on conclut que :

$$d_{VT}(P, Q) \geq 1 - 2 \exp \left( -\frac{1}{2n} \left( \sum_{i=1}^n d_{VT}(P_i, Q_i) \right)^2 \right).$$

□

Pour finir cette partie, nous donnons des calculs de distances en variation totale entre des lois particulières, dont nous aurons besoin dans les exemples (partie 3.3). Les calculs exacts de distances en variation totale sont en général impossibles à mener explicitement. Cependant, les ordres de grandeur figurant dans la proposition 3.1.8 nous suffiront.

**Proposition 3.1.8** • Soient  $\theta$  et  $\theta'$  deux réels compris entre 0 et 1. La distance en variation totale entre la loi de Bernoulli de paramètre  $\theta$ , notée  $b(\theta)$ , et la loi de Bernoulli de paramètre  $\theta'$ , notée  $b(\theta')$ , est :

$$d_{VT}(b(\theta), b(\theta')) = |\theta - \theta'|.$$

- Soit  $\theta$  un réel strictement positif, et  $\varepsilon$  un réel strictement compris entre 0 et 1. Quand  $\varepsilon$  tend vers 0, la distance en variation totale entre la loi de Poisson de paramètre  $\theta$ , notée  $\mathcal{P}(\theta)$ , et la loi de Poisson de paramètre  $\theta(1 - \varepsilon)$ , notée  $\mathcal{P}(\theta(1 - \varepsilon))$ , est :

$$d_{VT}(\mathcal{P}(\theta(1 - \varepsilon)), \mathcal{P}(\theta)) = \frac{\theta^{1+[\theta]} e^{-\theta} \varepsilon}{[\theta]!} + o(\varepsilon),$$

où  $[\theta]$  désigne la partie entière de  $\theta$ .



- Soient  $\theta$  un réel, et  $\varepsilon$  un réel strictement compris entre 0 et 1. Quand  $\varepsilon$  tend vers 0, la distance en variation totale entre la loi normale  $\mathcal{N}(\theta\varepsilon, 1 - \varepsilon^2)$  et la loi normale  $\mathcal{N}(0, 1)$  est :

$$d_{VT}(\mathcal{N}(\theta\varepsilon, 1 - \varepsilon^2), \mathcal{N}(0, 1)) = \frac{|\theta|\varepsilon}{\sqrt{2\pi}} + o(\varepsilon^2).$$

### Démonstration.

Examinons successivement les trois cas.

- Les deux lois de Bernoulli  $b(\theta)$  et  $b(\theta')$  ont une densité par rapport à la mesure de comptage sur  $\mathbb{N}$ , donc par l'égalité (3.1.2) de la définition 3.1.3 de la distance en variation totale entre deux mesures à densité, nous pouvons écrire :

$$\begin{aligned} d_{VT}(b(\theta), b(\theta')) &= \frac{1}{2} (|\theta - \theta'| + |(1 - \theta) - (1 - \theta')|) \\ &= |\theta - \theta'|. \end{aligned}$$

- Les deux lois de Poisson  $\mathcal{P}(\theta)$  et  $\mathcal{P}(\theta(1 - \varepsilon))$  ont une densité par rapport à la mesure de comptage sur  $\mathbb{N}$ , donc par l'égalité (3.1.2) de la définition 3.1.3 de la distance en variation totale entre deux mesures à densité, nous pouvons écrire :

$$\begin{aligned} d_{VT}(\mathcal{P}(\theta(1 - \varepsilon)), \mathcal{P}(\theta)) &= \frac{1}{2} \sum_{k=0}^{+\infty} \left| e^{-\theta} \frac{\theta^k}{k!} - e^{-\theta(1-\varepsilon)} \frac{\theta^k (1-\varepsilon)^k}{k!} \right| \\ &= \frac{1}{2} \sum_{k=0}^{+\infty} e^{-\theta} \frac{\theta^k}{k!} |1 - e^{\theta\varepsilon} (1 - \varepsilon)^k|. \end{aligned}$$

À cette étape, faisons un développement limité à l'ordre 1, pour  $\varepsilon$  proche de 0, de la fonction à l'intérieur des valeurs absolues. Il faut noter que le reste du développement limité est majoré par un polynôme en  $k$ ; donc, grâce au coefficient  $1/k!$  qui figure devant la valeur absolue, ce reste est sommable en  $k$ . Nous obtenons :

$$|1 - e^{\theta\varepsilon} (1 - \varepsilon)^k| = |\theta - k|\varepsilon + o(\varepsilon),$$

puis :

$$\begin{aligned} d_{VT}(\mathcal{P}(\theta(1 - \varepsilon)), \mathcal{P}(\theta)) &= \frac{1}{2} \sum_{k=0}^{+\infty} e^{-\theta} \frac{\theta^k}{k!} |\theta - k|\varepsilon + o(\varepsilon) \\ &= \frac{1}{2} \left( \sum_{k=0}^{\lfloor \theta \rfloor} e^{-\theta} \frac{\theta^k}{k!} (\theta - k)\varepsilon + \sum_{k=\lfloor \theta \rfloor + 1}^{+\infty} e^{-\theta} \frac{\theta^k}{k!} (k - \theta)\varepsilon \right) + o(\varepsilon). \end{aligned}$$

Examinons la première somme :

$$\begin{aligned}
 \sum_{k=0}^{\lfloor \theta \rfloor} e^{-\theta} \frac{\theta^k}{k!} (\theta - k) \varepsilon &= \varepsilon e^{-\theta} \left( \sum_{k=0}^{\lfloor \theta \rfloor} \frac{\theta^{k+1}}{k!} - \sum_{k=1}^{\lfloor \theta \rfloor} \frac{\theta^k}{(k-1)!} \right) \\
 &= \varepsilon e^{-\theta} \left( \theta \left( e^\theta - \sum_{k=\lfloor \theta \rfloor+1}^{+\infty} \frac{\theta^k}{k!} \right) - \theta \left( e^\theta - \sum_{k=\lfloor \theta \rfloor}^{+\infty} \frac{\theta^k}{k!} \right) \right) \\
 &= \varepsilon \theta e^{-\theta} \frac{\theta^{\lfloor \theta \rfloor}}{\lfloor \theta \rfloor!} .
 \end{aligned}$$

Quant à la deuxième somme, elle vaut :

$$\begin{aligned}
 \sum_{k=\lfloor \theta \rfloor+1}^{+\infty} e^{-\theta} \frac{\theta^k}{k!} (k - \theta) \varepsilon &= \varepsilon e^{-\theta} \left( \sum_{k=\lfloor \theta \rfloor+1}^{+\infty} \frac{\theta^k}{(k-1)!} - \sum_{k=\lfloor \theta \rfloor+1}^{+\infty} \frac{\theta^{k+1}}{k!} \right) \\
 &= \varepsilon e^{-\theta} \left( \theta \sum_{k=\lfloor \theta \rfloor}^{+\infty} \frac{\theta^k}{k!} - \theta \sum_{k=\lfloor \theta \rfloor+1}^{+\infty} \frac{\theta^k}{k!} \right) \\
 &= \varepsilon \theta e^{-\theta} \frac{\theta^{\lfloor \theta \rfloor}}{\lfloor \theta \rfloor!} .
 \end{aligned}$$

Donc pour la distance nous obtenons :

$$d_{VT}(\mathcal{P}(\theta(1 - \varepsilon)), \mathcal{P}(\theta)) = \varepsilon e^{-\theta} \frac{\theta^{1+\lfloor \theta \rfloor}}{\lfloor \theta \rfloor!} + o(\varepsilon) .$$

- Partons de l'égalité (3.1.2) de la proposition 3.1.3 où la distance en variation totale s'écrit comme la moitié de la distance en norme  $L^1$  :

$$\begin{aligned}
 d_{VT}(\mathcal{N}(\theta\varepsilon, 1 - \varepsilon^2), \mathcal{N}(0, 1)) &= \frac{1}{2} \int_{\mathbb{R}} \left| \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} - \frac{1}{\sqrt{2\pi}\sqrt{1 - \varepsilon^2}} e^{-\frac{(x - \theta\varepsilon)^2}{2(1 - \varepsilon^2)}} \right| dx \\
 &= \frac{1}{2} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \\
 &\quad \times \left| 1 - \frac{1}{\sqrt{1 - \varepsilon^2}} \exp\left(\frac{x^2}{2} - \frac{(x - \theta\varepsilon)^2}{2(1 - \varepsilon^2)}\right) \right| dx .
 \end{aligned}$$

Après avoir mis le polynôme argument de l'exponentielle sous sa forme canonique, on arrive à :

$$d_{VT}(\mathcal{N}(\theta\varepsilon, 1 - \varepsilon^2), \mathcal{N}(0, 1)) = \frac{1}{2} E \left| 1 - \frac{1}{\sqrt{1 - \varepsilon^2}} \exp\left(\frac{\theta^2}{2}\right) \exp\left(-\frac{(\varepsilon Z - \theta)^2}{2(1 - \varepsilon^2)}\right) \right| ,$$

où  $Z$  est une variable aléatoire de loi gaussienne standard. En regardant la fonction à l'intérieur de l'espérance comme une fonction de  $\varepsilon$ , nous pouvons faire son développement limité au voisinage de 0 :

$$\left| 1 - \frac{1}{\sqrt{1 - \varepsilon^2}} \exp\left(\frac{\theta^2}{2}\right) \exp\left(-\frac{(\varepsilon Z - \theta)^2}{2(1 - \varepsilon^2)}\right) \right| = |\theta\varepsilon Z| + \varepsilon^2 R(\varepsilon, Z) ,$$

où  $R$  est une fonction qui tend vers 0 quand  $\varepsilon$  tend vers 0, et qui est majorée par un polynôme en  $Z$ . En prenant l'espérance, comme tous les moments de la gaussienne sont finis, nous obtenons :

$$d_{VT}(\mathcal{N}(\theta\varepsilon, 1 - \varepsilon^2), \mathcal{N}(0, 1)) = \frac{|\theta|\varepsilon}{\sqrt{2\pi}} + o(\varepsilon^2) .$$

□

Comme nous l'avons constaté dans la proposition 3.1.8, la distance en variation totale est en général difficile à calculer exactement, même pour des lois usuelles. Cependant, elle est comparable à d'autres distances qui sont plus maniables en termes de calculs explicites. C'est ce que nous étudions dans ce qui suit, en commençant par la distance de Hellinger.

### 3.1.2 La distance de Hellinger

La distance de Hellinger entre deux mesures de probabilité est, à un coefficient  $\sqrt{2}$  près, la distance  $L^2$  entre les racines carrées de leurs densités :

**Définition 3.1.9** Soient  $P$  et  $Q$  deux lois de probabilité sur l'espace  $(\Omega, \mathcal{A})$ , de densités respectives  $p$  et  $q$  par rapport à une même mesure  $\lambda$ . La distance de Hellinger entre  $P$  et  $Q$  est définie de la façon suivante :

$$d_H(P, Q) = \left( \frac{1}{2} \int_{\Omega} (\sqrt{p} - \sqrt{q})^2 d\lambda \right)^{\frac{1}{2}} .$$

Remarquons tout de suite qu'en développant le carré dans la formule ci-dessus, nous obtenons :

$$d_H^2(P, Q) = 1 - \int_{\Omega} \sqrt{pq} d\lambda . \quad (3.1.6)$$

De la définition 3.1.9 découle une formule permettant de calculer la distance de Hellinger entre deux lois de probabilité dont l'une est absolument continue par rapport à l'autre.

**Proposition 3.1.10** Soient  $P$  et  $Q$  deux lois de probabilité telles que  $P$  est absolument continue par rapport à  $Q$ . Notons  $f$  la densité de  $P$  par rapport à  $Q$ . Alors :

$$d_H^2(P, Q) = \frac{1}{2} \int_{\Omega} (\sqrt{f} - 1)^2 dQ .$$

Le coefficient  $1/2$  dans la définition permet d'assurer que la distance de Hellinger entre deux lois de probabilité est toujours comprise entre 0 et 1, comme le montre la proposition suivante.

**Proposition 3.1.11** Soient  $P$  et  $Q$  deux mesures de probabilité sur l'espace  $(\Omega, \mathcal{A})$ .

- La distance de Hellinger est une distance, et en particulier,  $d_H(P, Q) = 0$  si et seulement si  $P = Q$ .

- Si  $P$  et  $Q$  sont étrangères (c'est-à-dire à supports disjoints), alors  $d_H(P, Q) = 1$ .

Calculons la distance de Hellinger sur des exemples de lois usuelles, qui nous seront utiles dans la partie 3.3.

**Proposition 3.1.12** • Soient  $\theta$  et  $\theta'$  deux réels compris entre 0 et 1. La distance de Hellinger entre la loi de Bernoulli de paramètre  $\theta$ , notée  $b(\theta)$ , et la loi de Bernoulli de paramètre  $\theta'$ , notée  $b(\theta')$ , est :

$$d_H^2(b(\theta), b(\theta')) = 1 - \sqrt{(1-\theta)(1-\theta')} - \sqrt{\theta\theta'}.$$

- Soit  $\theta$  et  $\theta'$  deux réels strictement positifs. La distance de Hellinger entre la loi de Poisson de paramètre  $\theta$ , notée  $\mathcal{P}(\theta)$ , et la loi de Poisson de paramètre  $\theta'$ , notée  $\mathcal{P}(\theta')$ , vaut :

$$d_H^2(\mathcal{P}(\theta), \mathcal{P}(\theta')) = 1 - e^{-\frac{\theta+\theta'}{2}} e^{\sqrt{\theta\theta'}}.$$

- Soit  $\mathcal{N}(m_1, v_1^2)$  la loi normale de moyenne  $m_1$  et de variance  $v_1^2$ , et  $\mathcal{N}(m_2, v_2^2)$  la loi normale de moyenne  $m_2$  et de variance  $v_2^2$ . Alors la distance de Hellinger entre ces deux distributions est donnée par :

$$d_H^2(\mathcal{N}(m_1, v_1^2), \mathcal{N}(m_2, v_2^2)) = 1 - \sqrt{\frac{2v_1v_2}{v_1^2 + v_2^2}} \exp\left(-\frac{(m_1 - m_2)^2}{4(v_1^2 + v_2^2)}\right).$$

### Démonstration.

Examinons successivement les trois cas. Nous partons toujours de la formule (3.1.6) de la distance de Hellinger.

- Les lois de Bernoulli sont absolument continues par rapport à la mesure de comptage sur  $\mathbb{N}$ , donc la formule (3.1.6) donne directement :

$$d_H^2(b(\theta), b(\theta')) = 1 - \sqrt{(1-\theta)(1-\theta')} - \sqrt{\theta\theta'}.$$

- Les lois de Poisson sont absolument continues par rapport à la mesure de comptage sur  $\mathbb{N}$ , donc la formule (3.1.6) donne :

$$\begin{aligned} d_H^2(\mathcal{P}(\theta), \mathcal{P}(\theta')) &= 1 - \sum_{k=0}^{+\infty} \sqrt{e^{-(\theta+\theta')} \frac{(\theta\theta')^k}{(k!)^2}} \\ &= 1 - e^{-\frac{\theta+\theta'}{2}} e^{\sqrt{\theta\theta'}}. \end{aligned}$$

- Les lois normales sont absolument continues par rapport à la mesure de Lebesgue sur  $\mathbb{R}$ , donc la formule (3.1.6) donne :

$$\begin{aligned} d_H^2(\mathcal{N}(m_1, v_1^2), \mathcal{N}(m_2, v_2^2)) &= 1 - \int_{\mathbb{R}} \sqrt{\frac{1}{2\pi v_1 v_2} \exp\left(-\frac{(x-m_1)^2}{2v_1^2} - \frac{(x-m_2)^2}{2v_2^2}\right)} dx \\ &= 1 - \frac{1}{\sqrt{2\pi v_1 v_2}} \int_{\mathbb{R}} \exp\left(-\frac{(x-m_1)^2}{4v_1^2} - \frac{(x-m_2)^2}{4v_2^2}\right) dx. \end{aligned}$$

Mettons alors l'argument de l'exponentielle sous sa forme canonique. Après calculs, on trouve :

$$-\frac{(x - m_1)^2}{4v_1^2} - \frac{(x - m_2)^2}{4v_2^2} = -\frac{1}{4} \left( \frac{1}{v_1^2} + \frac{1}{v_2^2} \right) \left( x - \frac{m_1v_2^2 + m_2v_1^2}{v_1^2 + v_2^2} \right)^2 - \frac{(m_1 - m_2)^2}{4(v_1^2 + v_2^2)}.$$

Ainsi pour la distance, on obtient :

$$\begin{aligned} d_H^2(\mathcal{N}(m_1, v_1^2), \mathcal{N}(m_2, v_2^2)) &= 1 - \frac{1}{\sqrt{2\pi v_1 v_2}} \exp\left(-\frac{(m_1 - m_2)^2}{4(v_1^2 + v_2^2)}\right) \\ &\quad \times \int_{\mathbb{R}} \exp\left(-\frac{1}{4} \left( \frac{1}{v_1^2} + \frac{1}{v_2^2} \right) \left( x - \frac{m_1v_2^2 + m_2v_1^2}{v_1^2 + v_2^2} \right)^2\right) dx \\ &= 1 - \frac{1}{\sqrt{v_1 v_2}} \exp\left(-\frac{(m_1 - m_2)^2}{4(v_1^2 + v_2^2)}\right) \left( \frac{1}{2v_1^2} + \frac{1}{2v_2^2} \right)^{-\frac{1}{2}} \\ &= 1 - \sqrt{\frac{2v_1 v_2}{v_1^2 + v_2^2}} \exp\left(-\frac{(m_1 - m_2)^2}{4(v_1^2 + v_2^2)}\right). \end{aligned}$$

□

L'inégalité suivante est classique ([47] p. 61) et permet d'encadrer la distance en variation totale par la distance de Hellinger :

**Proposition 3.1.13** Soient  $P$  et  $Q$  deux mesures de probabilité sur l'espace  $(\Omega, \mathcal{A})$ .

$$d_H^2(P, Q) \leq d_{VT}(P, Q) \leq d_H(P, Q) \sqrt{2 - d_H^2(P, Q)}.$$

Une conséquence immédiate de la proposition 3.1.13 qui se trouve fréquemment dans la littérature est l'inégalité :

$$d_H^2(P, Q) \leq d_{VT}(P, Q) \leq \sqrt{2} d_H(P, Q). \quad (3.1.7)$$

### Démonstration.

Montrons successivement les deux inégalités. Soit  $\lambda$  une mesure qui domine les deux mesures  $P$  et  $Q$ . Pour l'inégalité de gauche, il suffit de remarquer que :

$$\forall x, y \in \mathbb{R}^+, (\sqrt{x} - \sqrt{y})^2 \leq |x - y|,$$

pour obtenir :

$$\frac{1}{2} \int_{\Omega} (\sqrt{p} - \sqrt{q})^2 d\lambda \leq \frac{1}{2} \int_{\Omega} |p - q| d\lambda,$$

c'est-à-dire :

$$d_H^2(P, Q) \leq d_{VT}(P, Q).$$

Pour l'inégalité de droite, on utilise l'inégalité de Cauchy-Schwarz :

$$\begin{aligned} d_{VT}^2(P, Q) &= \frac{1}{4} \left( \int_{\Omega} |\sqrt{p} - \sqrt{q}| |\sqrt{p} + \sqrt{q}| d\lambda \right)^2 \\ &\leq \frac{1}{4} \left( \int_{\Omega} (\sqrt{p} - \sqrt{q})^2 d\lambda \right) \left( \int_{\Omega} (\sqrt{p} + \sqrt{q})^2 d\lambda \right). \end{aligned}$$

Or la formule (3.1.6) de la distance de Hellinger donne :

$$d_H^2(P, Q) = 1 - \int_{\Omega} \sqrt{pq} d\lambda,$$

et donc :

$$\int_{\Omega} (\sqrt{p} + \sqrt{q})^2 d\lambda = 4 - 2 d_H^2(P, Q).$$

D'où :

$$d_{VT}^2(P, Q) \leq d_H^2(P, Q) (2 - d_H^2(P, Q)).$$

□

Pour terminer on peut déterminer la distance de Hellinger entre deux mesures produits en fonction de la distance entre les marginales.

**Proposition 3.1.14** *Soit  $n$  un entier strictement positif. Pour tout entier  $i$  compris entre 1 et  $n$ , soient  $P_i$  et  $Q_i$  deux mesures définies sur le même espace  $(\Omega_i, \mathcal{A}_i)$ . Soient  $P$  et  $Q$  les mesures produits suivantes :*

$$P = \bigotimes_{i=1}^n P_i \quad \text{et} \quad Q = \bigotimes_{i=1}^n Q_i.$$

Alors la distance de Hellinger entre  $P$  et  $Q$  est donnée par :

$$d_H^2(P, Q) = 1 - \prod_{i=1}^n (1 - d_H^2(P_i, Q_i)) \leq \sum_{i=1}^n d_H^2(P_i, Q_i).$$

**Démonstration.**

Pour tout  $i$  compris entre 1 et  $n$ , soit  $\lambda_i$  une mesure qui domine  $P_i$  et  $Q_i$ . Soient  $p_i$  et  $q_i$  les densités respectives de  $P_i$  et  $Q_i$  par rapport à  $\lambda_i$ . Soit  $\lambda$  la mesure produit des  $\lambda_i$  :

$$\lambda = \bigotimes_{i=1}^n \lambda_i.$$

La densité  $p$  de  $P$  par rapport à  $\lambda$  est alors donnée, en  $(x_1, \dots, x_n)$ , par :

$$p(x_1, \dots, x_n) = \prod_{i=1}^n p_i(x_i),$$

et de même la densité  $q$  de  $Q$  par rapport à  $\lambda$  est donnée, en  $(x_1, \dots, x_n)$ , par :

$$q(x_1, \dots, x_n) = \prod_{i=1}^n q_i(x_i) .$$

On écrit ensuite la distance de Hellinger entre  $P$  et  $Q$  (formule (3.1.6)) :

$$\begin{aligned} d_H^2(P, Q) &= 1 - \int_{\Omega_1 \times \dots \times \Omega_n} \sqrt{pq} \, d\lambda \\ &= 1 - \int_{\Omega_1 \times \dots \times \Omega_n} \left( \prod_{i=1}^n p_i(x_i) q_i(x_i) \right)^{\frac{1}{2}} d\lambda(x_1, \dots, x_n) . \end{aligned}$$

En appliquant le théorème de Fubini, on obtient :

$$d_H^2(P, Q) = 1 - \prod_{i=1}^n \left( \int_{\Omega_i} \sqrt{p_i q_i} \, d\lambda_i \right) ,$$

ce qui conduit à :

$$d_H^2(P, Q) = 1 - \prod_{i=1}^n (1 - d_H^2(P_i, Q_i)) .$$

Pour montrer la deuxième inégalité, on utilise l'inégalité classique suivante, valable pour des réels  $x_1, \dots, x_n$  positifs et inférieurs à 1 :

$$\prod_{i=1}^n (1 - x_i) \geq 1 - \sum_{i=1}^n x_i .$$

□

Appliquons cette proposition au cas particulier de deux  $n$ -échantillons de lois normales. Un calcul direct basé sur le résultat obtenu dans la proposition 3.1.12 conduit alors à la proposition suivante.

**Proposition 3.1.15** *Soit  $\mathcal{N}(m_1, v_1^2)^{\otimes n}$  la loi d'un  $n$ -échantillon de lois normales de moyenne  $m_1$  et de variance  $v_1^2$ , et soit  $\mathcal{N}(m_2, v_2^2)^{\otimes n}$  la loi d'un  $n$ -échantillon de lois normales de moyenne  $m_2$  et de variance  $v_2^2$ . Alors :*

$$d_H^2(\mathcal{N}(m_1, v_1^2)^{\otimes n}, \mathcal{N}(m_2, v_2^2)^{\otimes n}) = 1 - \left( \frac{2v_1 v_2}{v_1^2 + v_2^2} \right)^{\frac{n}{2}} \exp \left( -\frac{n(m_1 - m_2)^2}{4(v_1^2 + v_2^2)} \right) .$$

Pour mesurer les écarts entre lois de probabilité, nous utiliserons aussi deux autres notions classiques en statistique et théorie de l'information : la distance du  $\chi^2$  et la distance de Kullback. Ce ne sont pas des distances au sens topologique du terme, puisqu'elles ne vérifient pas la propriété de symétrie, et aussi peuvent prendre  $+\infty$  pour valeur, mais elles se comparent cependant aux distances en variation totale et de Hellinger.

### 3.1.3 La distance du $\chi^2$

La distance du  $\chi^2$  trouve son origine en statistiques. Elle est définie de la façon suivante.

**Définition 3.1.16** Soient  $P$  et  $Q$  deux lois de probabilité sur l'espace  $(\Omega, \mathcal{A})$  telles que  $P$  soit absolument continue par rapport à  $Q$ . Soit  $f$  la densité de  $P$  par rapport à  $Q$ . La distance du  $\chi^2$  entre  $P$  et  $Q$  est alors définie par :

$$d_{\chi^2}(P, Q) = \left( \int_{\Omega} (f - 1)^2 dQ \right)^{\frac{1}{2}} .$$

Il faut noter que la racine carrée n'est pas présente chez tous les auteurs. Nous avons choisi de la mettre dans un souci d'homogénéité avec les autres distances considérées.

Remarquons aussi que si  $P$  et  $Q$  ont des densités respectives  $p$  et  $q$  par rapport à une même mesure  $\lambda$ , alors la distance du  $\chi^2$  devient :

$$d_{\chi^2}^2(P, Q) = \int_{\Omega} \left( \frac{p}{q} - 1 \right)^2 q d\lambda ,$$

ce qui, en développant, se simplifie en :

$$d_{\chi^2}^2(P, Q) = \int_{\Omega} \frac{p^2}{q} d\lambda - 1 . \quad (3.1.8)$$

Calculons la distance du  $\chi^2$  entre deux lois de Bernoulli, de Poisson puis normales. Ces résultats nous seront utiles dans la partie 3.3.

**Proposition 3.1.17** • Soient  $\theta$  et  $\theta'$  deux réels strictement compris entre 0 et 1. La distance du  $\chi^2$  entre la loi de Bernoulli de paramètre  $\theta$ , notée  $b(\theta)$ , et la loi de Bernoulli de paramètre  $\theta'$ , notée  $b(\theta')$ , est :

$$d_{\chi^2}^2(b(\theta), b(\theta')) = \frac{(\theta - \theta')^2}{\theta'(1 - \theta')} .$$

- Soit  $\theta$  et  $\theta'$  deux réels strictement positifs. La distance du  $\chi^2$  entre la loi de Poisson de paramètre  $\theta$ , notée  $\mathcal{P}(\theta)$ , et la loi de Poisson de paramètre  $\theta'$ , notée  $\mathcal{P}(\theta')$ , vaut :

$$d_{\chi^2}^2(\mathcal{P}(\theta), \mathcal{P}(\theta')) = \exp \left( \theta' - 2\theta + \frac{\theta^2}{\theta'} \right) - 1 .$$

- Soit  $\mathcal{N}(m_1, v_1^2)$  la loi normale de moyenne  $m_1$  et de variance  $v_1^2$ , et  $\mathcal{N}(m_2, v_2^2)$  la loi normale de moyenne  $m_2$  et de variance  $v_2^2$ . La distance du  $\chi^2$  entre ces deux distributions est donnée par :

$$d_{\chi^2}^2(\mathcal{N}(m_1, v_1^2), \mathcal{N}(m_2, v_2^2)) = \begin{cases} \frac{v_2^2}{v_1 \sqrt{2v_2^2 - v_1^2}} \exp \left( \frac{(m_1 - m_2)^2}{2v_2^2 - v_1^2} \right) - 1 & \text{si } 2v_2^2 - v_1^2 > 0 , \\ +\infty & \text{sinon.} \end{cases}$$



**Démonstration.**

Examinons successivement les trois cas. Nous partons toujours de la formule (3.1.8) de la distance du  $\chi^2$ .

- Les lois de Bernoulli sont absolument continues par rapport à la mesure de comptage sur  $\mathbb{N}$ , donc la formule (3.1.8) donne :

$$\begin{aligned} d_{\chi^2}^2(b(\theta), b(\theta')) &= \frac{\theta^2}{\theta'} + \frac{(1-\theta)^2}{1-\theta'} - 1 \\ &= \frac{(\theta - \theta')^2}{\theta'(1-\theta')}. \end{aligned}$$

- Les lois de Poisson sont absolument continues par rapport à la mesure de comptage sur  $\mathbb{N}$ , donc la formule (3.1.8) donne :

$$\begin{aligned} d_{\chi^2}^2(\mathcal{P}(\theta), \mathcal{P}(\theta')) &= \sum_{k=0}^{+\infty} \left( \frac{e^{-\theta} \theta^k}{k!} \right)^2 \left( \frac{e^{-\theta'} \theta'^k}{k!} \right)^{-1} - 1 \\ &= \exp\left(\theta' - 2\theta + \frac{\theta^2}{\theta'}\right) - 1. \end{aligned}$$

- Les lois normales sont absolument continues par rapport à la mesure de Lebesgue sur  $\mathbb{R}$ , donc la formule (3.1.8) donne :

$$d_{\chi^2}^2(\mathcal{N}(m_1, v_1^2), \mathcal{N}(m_2, v_2^2)) = \int_{\mathbb{R}} \frac{v_2}{v_1^2 \sqrt{2\pi}} \exp\left(-\frac{(x-m_1)^2}{v_1^2} + \frac{(x-m_2)^2}{2v_2^2}\right) dx - 1.$$

Mettons l'argument de l'exponentielle sous sa forme canonique. Après calculs, on obtient :

$$-\frac{(x-m_1)^2}{v_1^2} + \frac{(x-m_2)^2}{2v_2^2} = \left(-\frac{1}{v_1^2} + \frac{1}{2v_2^2}\right) \left(x - \frac{2m_1v_2^2 - m_2v_1^2}{2v_2^2 - v_1^2}\right)^2 + \frac{(m_1 - m_2)^2}{2v_2^2 - v_1^2}.$$

Ici on voit apparaître la condition qui lie  $v_1$  et  $v_2$ ; en effet la fonction suivante :

$$x \mapsto \exp\left(\left(-\frac{1}{v_1^2} + \frac{1}{2v_2^2}\right) \left(x - \frac{2m_1v_2^2 - m_2v_1^2}{2v_2^2 - v_1^2}\right)^2\right)$$

est intégrable sur  $\mathbb{R}$  si et seulement si :

$$-\frac{1}{v_1^2} + \frac{1}{2v_2^2} < 0,$$

ce qui équivaut à :

$$2v_2^2 - v_1^2 > 0.$$

Sous cette condition, on obtient alors :

$$\begin{aligned} d_{\chi^2}^2(\mathcal{N}(m_1, v_1^2), \mathcal{N}(m_2, v_2^2)) &= \frac{v_2}{v_1^2} \exp\left(\frac{(m_1 - m_2)^2}{2v_2^2 - v_1^2}\right) \left(\frac{2}{v_1^2} - \frac{1}{v_2^2}\right)^{-\frac{1}{2}} - 1 \\ &= \frac{v_2^2}{v_1 \sqrt{2v_2^2 - v_1^2}} \exp\left(\frac{(m_1 - m_2)^2}{2v_2^2 - v_1^2}\right) - 1. \end{aligned}$$

□

Intéressons-nous maintenant aux relations existant entre la distance du  $\chi^2$  et les distances précédemment étudiées.

**Proposition 3.1.18** *Soit  $P$  et  $Q$  deux lois de probabilité telles que  $P$  soit absolument continue par rapport à  $Q$ . Alors :*

1.  $d_{VT}(P, Q) \leq \frac{1}{2} d_{\chi^2}(P, Q)$ .
2.  $d_H(P, Q) \leq \frac{1}{\sqrt{2}} d_{\chi^2}(P, Q)$ .

**Démonstration.**

Soit  $f$  la densité de  $P$  par rapport à  $Q$ . La première assertion est une application de l'inégalité de Cauchy-Schwarz. D'après le corollaire 3.1.4, la distance en variation totale est donnée par :

$$d_{VT}(P, Q) = \frac{1}{2} \int_{\Omega} |f - 1| dQ.$$

Comme  $Q$  est une probabilité, l'inégalité de Cauchy-Schwarz donne :

$$d_{VT}(P, Q) \leq \frac{1}{2} \left( \int_{\Omega} (f - 1)^2 dQ \right)^{\frac{1}{2}} = \frac{1}{2} d_{\chi^2}(P, Q).$$

Pour la deuxième assertion, on utilise la proposition 3.1.10 pour écrire :

$$d_H^2(P, Q) = \frac{1}{2} \int_{\Omega} (\sqrt{f} - 1)^2 dQ.$$

Or la fonction  $(1 + \sqrt{f})^2$  est supérieure à 1, ce qui donne :

$$\begin{aligned} d_H^2(P, Q) &\leq \frac{1}{2} \int_{\Omega} \left( [\sqrt{f} - 1] [1 + \sqrt{f}] \right)^2 dQ \\ &= \frac{1}{2} d_{\chi^2}^2(\mu, \nu). \end{aligned}$$

□

Un des avantages de la distance du  $\chi^2$  est qu'elle se comporte bien vis-à-vis des mesures produits, dans le sens où la distance du  $\chi^2$  entre deux mesures produits peut s'exprimer simplement en fonction des distances du  $\chi^2$  entre les marginales.

**Proposition 3.1.19** *Soit  $n$  un entier strictement positif. Pour tout entier  $i$  compris entre 1 et  $n$ , soient  $P_i$  et  $Q_i$  deux lois de probabilité définies sur le même espace  $(\Omega_i, \mathcal{A}_i)$ , telles que  $P_i$  soit absolument continue par rapport à  $Q_i$ . Soient  $P$  et  $Q$  les mesures produits suivantes :*

$$P = \bigotimes_{i=1}^n P_i \quad \text{et} \quad Q = \bigotimes_{i=1}^n Q_i.$$

Alors la distance du  $\chi^2$  entre  $P$  et  $Q$  est donnée par :

$$d_{\chi^2}^2(P, Q) = \prod_{i=1}^n (1 + d_{\chi^2}^2(P_i, Q_i)) - 1 \geq \sum_{i=1}^n d_{\chi^2}^2(P_i, Q_i).$$

**Démonstration.**

Si on note  $f_i$  la densité de  $P_i$  par rapport à  $Q_i$ , la distance du  $\chi^2$  entre  $P_i$  et  $Q_i$  s'écrit, d'après la définition 3.1.16 :

$$d_{\chi^2}^2(P_i, Q_i) = \int_{\Omega_i} f_i^2 dQ_i - 1.$$

De plus la probabilité  $P$  a une densité  $f$  par rapport à  $Q$ , qui s'écrit, en tout point  $(x_1, \dots, x_n)$  de  $\Omega_1 \times \dots \times \Omega_n$  :

$$f(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i).$$

Ainsi on peut écrire la distance du  $\chi^2$  entre  $P$  et  $Q$  sous la forme :

$$\begin{aligned} d_{\chi^2}^2(P, Q) &= \int_{\Omega_1 \times \dots \times \Omega_n} f^2 dQ - 1 \\ &= \prod_{i=1}^n \left( \int_{\Omega_i} f_i^2 dQ_i \right) - 1 \\ &= \prod_{i=1}^n (1 + d_{\chi^2}^2(P_i, Q_i)) - 1. \end{aligned}$$

Pour finir il faut voir que, en développant le produit, on obtient :

$$\prod_{i=1}^n (1 + d_{\chi^2}^2(P_i, Q_i)) = 1 + \sum_{i=1}^n d_{\chi^2}^2(P_i, Q_i) + \text{terme positif},$$

donc on a la minoration annoncée. □

Terminons cette section avec la description d'une dernière distance entre mesures, qui a le comportement le plus simple vis-à-vis des mesures produits : la distance de Kullback.

### 3.1.4 La distance de Kullback

La distance de Kullback est liée à la notion générale d'entropie utilisée en théorie de l'information (cf. [16]). Elle porte aussi le nom d'entropie relative. On pourra lire dans [3] ou [56] la définition et les premières propriétés qui sont rappelées ci-dessous. Commençons par la définition.

**Définition 3.1.20** Soient  $P$  et  $Q$  deux mesures de probabilité sur l'espace  $(\Omega, \mathcal{A})$  telles que  $P$  soit absolument continue par rapport à  $Q$ . Soit  $f$  la densité de  $P$  par rapport à  $Q$ . La distance de Kullback entre  $P$  et  $Q$  est alors définie par :

$$d_K(P, Q) = \left( \int_{\Omega} f \log(f) dQ \right)^{\frac{1}{2}}.$$

L'inégalité de Jensen permet d'assurer que l'intégrale de  $f \log(f)$  est positive, ce qui nous autorise de plus à prendre pour définition la racine carrée de cette intégrale. Elle n'intervient pas dans la définition usuelle, mais figure pour nous par souci d'homogénéité avec les autres distances.

Remarquons que si  $P$  et  $Q$  ont pour densités respectives  $p$  et  $q$  par rapport à une même mesure  $\lambda$ , la distance de Kullback s'écrit alors :

$$\begin{aligned} d_K^2(P, Q) &= \int_{\Omega} \frac{p}{q} \log \left( \frac{p}{q} \right) q \, d\lambda \\ &= \int_{\Omega} p \log \left( \frac{p}{q} \right) \, d\lambda. \end{aligned} \quad (3.1.9)$$

Calculons la distance de Kullback entre deux lois de Bernoulli, de Poisson, puis normales. Ces résultats nous seront utiles dans la partie 3.3.

**Proposition 3.1.21** • Soient  $\theta$  et  $\theta'$  deux réels strictement compris entre 0 et 1. La distance de Kullback entre la loi de Bernoulli de paramètre  $\theta$ , notée  $b(\theta)$ , et la loi de Bernoulli de paramètre  $\theta'$ , notée  $b(\theta')$ , est :

$$d_K^2(b(\theta), b(\theta')) = \theta \log \left( \frac{\theta}{\theta'} \right) + (1 - \theta) \log \left( \frac{1 - \theta}{1 - \theta'} \right).$$

- Soit  $\theta$  et  $\theta'$  deux réels strictement positifs. La distance de Kullback entre la loi de Poisson de paramètre  $\theta$ , notée  $\mathcal{P}(\theta)$ , et la loi de Poisson de paramètre  $\theta'$ , notée  $\mathcal{P}(\theta')$ , vaut :

$$d_K^2(\mathcal{P}(\theta), \mathcal{P}(\theta')) = \theta' - \theta + \theta \log \left( \frac{\theta}{\theta'} \right).$$

- Soit  $\mathcal{N}(m_1, v_1^2)$  la loi normale de moyenne  $m_1$  et de variance  $v_1^2$ , et  $\mathcal{N}(m_2, v_2^2)$  la loi normale de moyenne  $m_2$  et de variance  $v_2^2$ . Alors la distance de Kullback entre ces deux lois est donnée par :

$$d_K^2(\mathcal{N}(m_1, v_1^2), \mathcal{N}(m_2, v_2^2)) = \log \left( \frac{v_2}{v_1} \right) - \frac{1}{2} + \frac{v_1^2}{2v_2^2} + \frac{(m_1 - m_2)^2}{2v_2^2}.$$

### Démonstration.

Examinons successivement les trois cas. Nous partons toujours de la formule (3.1.9) de la distance de Kullback.

- Les lois de Bernoulli sont absolument continues par rapport à la mesure de comptage sur  $\mathbb{N}$ , donc la formule (3.1.9) donne directement le résultat.

- Les lois de Poisson sont absolument continues par rapport à la mesure de comptage sur  $\mathbb{N}$ , donc la formule (3.1.9) donne :

$$\begin{aligned}
 d_K^2(\mathcal{P}(\theta), \mathcal{P}(\theta')) &= \sum_{k=0}^{+\infty} \left( e^{-\theta} \frac{\theta^k}{k!} \right) \log \left( \frac{e^{-\theta} \theta^k}{e^{-\theta'} \theta'^k} \right) \\
 &= \sum_{k=0}^{+\infty} \left( e^{-\theta} \frac{\theta^k}{k!} \right) \left( -\theta + \theta' + k \log \left( \frac{\theta}{\theta'} \right) \right) \\
 &= -\theta + \theta' + \theta \log \left( \frac{\theta}{\theta'} \right) .
 \end{aligned}$$

- Partons de la distance de Kullback sous la forme de l'égalité (3.1.9) ci-dessus :

$$\begin{aligned}
 d_K^2(\mathcal{N}(m_1, v_1^2), \mathcal{N}(m_2, v_2^2)) &= \int_{\mathbb{R}} \frac{1}{v_1 \sqrt{2\pi}} \exp \left( -\frac{(x - m_1)^2}{2v_1^2} \right) \\
 &\quad \times \log \left[ \frac{v_2}{v_1} \exp \left( -\frac{(x - m_1)^2}{2v_1^2} + \frac{(x - m_2)^2}{2v_2^2} \right) \right] dx \\
 &= \int_{\mathbb{R}} \frac{1}{v_1 \sqrt{2\pi}} \exp \left( -\frac{(x - m_1)^2}{2v_1^2} \right) \\
 &\quad \times \left[ \log \left( \frac{v_2}{v_1} \right) - \frac{(x - m_1)^2}{2v_1^2} + \frac{(x - m_2)^2}{2v_2^2} \right] dx \\
 &= \log \left( \frac{v_2}{v_1} \right) - \int_{\mathbb{R}} \frac{1}{2v_1^3 \sqrt{2\pi}} (x - m_1)^2 \exp \left( -\frac{(x - m_1)^2}{2v_1^2} \right) dx \\
 &\quad + \int_{\mathbb{R}} \frac{1}{2v_1 v_2^2 \sqrt{2\pi}} (x - m_2)^2 \exp \left( -\frac{(x - m_1)^2}{2v_1^2} \right) dx .
 \end{aligned}$$

Les deux intégrales se calculent facilement, car elles s'écrivent comme variances de gaussiennes ; si  $Z$  est une variable aléatoire de loi normale  $\mathcal{N}(m_1, v_1^2)$ , on a :

$$\begin{aligned}
 \int_{\mathbb{R}} \frac{1}{2v_1^3 \sqrt{2\pi}} (x - m_1)^2 \exp \left( -\frac{(x - m_1)^2}{2v_1^2} \right) dx &= \frac{\text{Var}(Z)}{2v_1^2} \\
 &= \frac{1}{2} ,
 \end{aligned}$$

et :

$$\begin{aligned}
 \int_{\mathbb{R}} \frac{1}{2v_1 v_2^2 \sqrt{2\pi}} (x - m_2)^2 \exp \left( -\frac{(x - m_1)^2}{2v_1^2} \right) dx &= \frac{E((Z - m_2)^2)}{2v_2^2} \\
 &= \frac{\text{Var}(Z) + (m_1 - m_2)^2}{2v_2^2} \\
 &= \frac{v_1^2}{2v_2^2} + \frac{(m_1 - m_2)^2}{2v_2^2} .
 \end{aligned}$$

On en conclut donc que la distance vaut :

$$d_K^2(\mathcal{N}(m_1, v_1^2), \mathcal{N}(m_2, v_2^2)) = \log\left(\frac{v_2}{v_1}\right) - \frac{1}{2} + \frac{v_1^2}{2v_2^2} + \frac{(m_1 - m_2)^2}{2v_2^2}.$$

□

Examinons les relations entre la distance de Kullback et les distances précédemment étudiées.

**Proposition 3.1.22** *Soit  $P$  et  $Q$  deux lois de probabilité telles que  $P$  soit absolument continue par rapport à  $Q$ . Alors :*

1.  $d_{VT}(P, Q) \leq \frac{1}{\sqrt{2}} d_K(P, Q)$ .
2.  $d_H(P, Q) \leq \sqrt{2} d_K(P, Q)$ .
3.  $d_K^2(P, Q) \leq \log(1 + d_{\chi^2}^2(P, Q)) \leq d_{\chi^2}^2(P, Q)$ .

La première inégalité est due à S. Kullback [36]. La preuve que nous donnons utilise un lemme technique qui est proposé par D. Pollard en exercice dans [47] p. 74.

**Lemme 3.1.23** *Pour tout  $x \geq -1$ , on définit la fonction  $\varphi$  par :*

$$\varphi(x) = \begin{cases} (x+1)\log(x+1) - x & \text{si } x > -1, \\ 1 & \text{si } x = -1. \end{cases}$$

On a alors l'inégalité suivante :

$$\forall x \geq -1, \quad \varphi(x) \geq \frac{x^2}{2\left(1 + \frac{x}{3}\right)}. \quad (3.1.10)$$

**Démonstration.**

Commençons par remarquer que l'inégalité est satisfaite par  $x = -1$ . Ensuite, la fonction  $\varphi$  est de classe  $C^2$  sur l'intervalle  $] -1 ; +\infty[$ , et on peut écrire :

$$\varphi(x) = x^2 \int_{s=0}^1 \int_{t=0}^s \varphi''(xt) dt ds.$$

De plus, on a :

$$\forall x > -1, \quad \varphi''(x) = \frac{1}{1+x},$$

donc on obtient :

$$\forall x > -1, \quad \varphi(x) = x^2 \int_{s=0}^1 \int_{t=0}^s \frac{1}{1+xt} dt ds.$$

Étudions séparément deux cas, selon le signe de  $x$ .

- Pour  $x$  positif, on découpe l'intégrale en deux morceaux :

$$\varphi(x) = x^2 \left( \underbrace{\int_{s=0}^{\frac{1}{3}} \dots}_{A} + \underbrace{\int_{s=\frac{1}{3}}^1 \dots}_{B} \right).$$

Dans le morceau  $A$ , on intervertit les deux intégrales :

$$\begin{aligned} A &= \int_{t=0}^{\frac{1}{3}} \int_{s=t}^1 \frac{1}{1+xt} ds dt \\ &= \int_{t=0}^{\frac{1}{3}} \frac{1-t}{1+xt} dt. \end{aligned}$$

Or  $t$  est positif, et inférieur ou égal à  $1/3$ , donc on peut minorer la fonction à intégrer :

$$\frac{1-t}{1+xt} \geq \frac{1-t}{1+\frac{x}{3}}.$$

On obtient donc :

$$A \geq \frac{1}{1+\frac{x}{3}} \int_0^{\frac{1}{3}} (1-t) dt,$$

c'est-à-dire :

$$A \geq \frac{5}{18} \times \frac{1}{1+\frac{x}{3}}.$$

Dans le morceau  $B$ , on intervertit aussi les deux intégrales, mais le domaine est trapézoïdal, ce qui conduit à une somme de deux termes :

$$B = \int_{t=0}^{\frac{1}{3}} \int_{s=\frac{1}{3}}^1 \frac{1}{1+xt} ds dt + \int_{t=\frac{1}{3}}^1 \int_{s=t}^1 \frac{1}{1+xt} ds dt.$$

La quantité  $B$  peut alors être minorée par le premier terme :

$$\begin{aligned} B &\geq \frac{2}{3} \int_{t=0}^{\frac{1}{3}} \frac{1}{1+xt} dt \\ &\geq \frac{2}{9} \times \frac{1}{1+\frac{x}{3}}, \end{aligned}$$

en utilisant encore le fait que  $t$  est inférieur à  $1/3$ . Ainsi, en sommant  $A$  et  $B$ , on arrive à montrer que :

$$\begin{aligned} \varphi(x) &\geq x^2 \left( \frac{5}{18} + \frac{2}{9} \right) \frac{1}{1+\frac{x}{3}} \\ &= \frac{1}{2} \times \frac{x^2}{1+\frac{x}{3}}. \end{aligned}$$

- Pour  $x$  négatif, on développe la fonction à intégrer en série entière :

$$\frac{1}{1+xt} = \sum_{k=0}^{+\infty} (-x)^k t^k,$$

ce qui est légitime puisque  $-x \in [0; 1[$  et  $t \in [0; 1]$ . Ensuite on intervertit la sommation et la double intégration ; on obtient :

$$\varphi(x) = x^2 \sum_{k=0}^{+\infty} \frac{(-x)^k}{(k+1)(k+2)}.$$

Pour finir, on démontre par récurrence que, pour tout  $k$  positif :

$$\frac{1}{(k+1)(k+2)} \geq \frac{1}{2 \cdot 3^k},$$

et donc pour  $\varphi(x)$ , on obtient :

$$\begin{aligned} \varphi(x) &\geq x^2 \sum_{k=0}^{+\infty} \frac{(-x)^k}{2 \cdot 3^k} \\ &= \frac{1}{2} \times \frac{x^2}{1 + \frac{x}{3}}, \end{aligned}$$

ce qui achève la démonstration. □

Démontrons maintenant la proposition 3.1.22 qui donne les inégalités entre les distances.

### Démonstration.

Nous regardons successivement les quatre inégalités, en notant  $f$  la densité de  $P$  par rapport à  $Q$ .

1. Soit  $g$  la fonction qui à  $x$  associe  $f(x) - 1$ . Comme  $f$  est une densité, on a :

$$\int_{\Omega} g \, dQ = 0.$$

Donc la distance de Kullback au carré entre  $P$  et  $Q$  s'écrit :

$$d_K^2(P, Q) = \int_{\Omega} [(g+1) \log(g+1) - g] \, dQ.$$

En utilisant le lemme précédent (lemme 3.1.23), on minore  $d_K^2(P, Q)$  :

$$d_K^2(P, Q) \geq \frac{1}{2} \int_{\Omega} \frac{g^2}{1 + \frac{g}{3}} \, dQ.$$



Multipliant le terme de droite par 1 qui est aussi  $\int_{\Omega}(1 + g/3) dQ$ , et utilisant l'inégalité de Cauchy-Schwarz, on obtient :

$$\begin{aligned} d_K^2(P, Q) &\geq \frac{1}{2} \left( \int_{\Omega} \frac{|g|}{\sqrt{1 + \frac{g}{3}}} \sqrt{1 + \frac{g}{3}} dQ \right)^2 \\ &= \frac{1}{2} \left( \int_{\Omega} |g| dQ \right)^2 \\ &= 2 d_{VT}^2(P, Q), \end{aligned}$$

ce qui donne la première inégalité.

2. Soit  $h$  la fonction qui à  $x$  associe  $\sqrt{f(x)} - 1$ . Par la formule de la distance de Hellinger établie dans la proposition 3.1.10, on a :

$$d_H^2(P, Q) = \frac{1}{2} \int_{\Omega} h^2 dQ.$$

D'autre part, on sait que l'intégrale  $\int_{\Omega} f dQ$  vaut 1. De plus, on peut écrire :

$$\begin{aligned} \int_{\Omega} f dQ &= \int_{\Omega} (1 + h)^2 dQ \\ &= 1 + 2 \int_{\Omega} h dQ + \int_{\Omega} h^2 dQ, \end{aligned}$$

ce qui montre que :

$$\begin{aligned} \int_{\Omega} h dQ &= -\frac{1}{2} \int_{\Omega} h^2 dQ \\ &= -d_H^2(P, Q). \end{aligned}$$

On écrit alors la distance de Kullback entre  $P$  et  $Q$  :

$$\begin{aligned} d_K^2(P, Q) &= \int_{\Omega} f \log(f) dQ \\ &= 2 \int_{\Omega} (h + 1)^2 \log(h + 1) dQ. \end{aligned}$$

On minore cette quantité grâce à l'inégalité classique suivante : pour tout  $x$  strictement supérieur à  $-1$ ,

$$\log(x + 1) \geq \frac{x}{x + 1}.$$

Ainsi on obtient :

$$\begin{aligned} d_K^2(P, Q) &\geq 2 \int_{\Omega} h(h + 1) dQ \\ &= 2 \int_{\Omega} h^2 dQ + 2 \int_{\Omega} h dQ \\ &= 2 d_H^2(P, Q), \end{aligned}$$

ce qui montre l'inégalité.

3. La fonction  $\log$  est concave, donc l'inégalité de Jensen appliquée avec la mesure de probabilité  $f dQ$  donne :

$$\begin{aligned} d_K^2(P, Q) &= \int_{\Omega} \log(f) f dQ \\ &\leq \log \left( \int_{\Omega} f^2 dQ \right) . \end{aligned}$$

Or en développant la formule de définition de la distance du  $\chi^2$ , on a :

$$d_{\chi^2}^2(P, Q) = \int_{\Omega} f^2 dQ - 2 \int_{\Omega} f dQ + 1 = \int_{\Omega} f^2 dQ - 1 ,$$

d'où :

$$d_K^2(P, Q) \leq \log(1 + d_{\chi^2}^2(P, Q)) .$$

La seconde inégalité s'obtient en utilisant l'inégalité classique suivante : pour tout  $x$  strictement supérieur à  $-1$ ,

$$\log(1 + x) \leq x .$$

□

Finissons cette partie avec la propriété qui nous sera la plus utile par la suite, et qui constitue l'un des principaux intérêts de la distance de Kullback : son comportement vis-à-vis des lois produits.

**Proposition 3.1.24** *Soit  $n$  un entier strictement positif. Pour tout entier  $i$  compris entre 1 et  $n$ , soient  $P_i$  et  $Q_i$  deux probabilités définies sur le même espace  $(\Omega_i, \mathcal{A}_i)$ , telles que  $P_i$  soit absolument continue par rapport à  $Q_i$ . Soient  $P$  et  $Q$  les mesures produits suivantes :*

$$P = \bigotimes_{i=1}^n P_i \quad \text{et} \quad Q = \bigotimes_{i=1}^n Q_i .$$

Alors la distance de Kullback entre  $P$  et  $Q$  est donnée par :

$$d_K^2(P, Q) = \sum_{i=1}^n d_K^2(P_i, Q_i) .$$

#### Démonstration.

Pour tout  $i$  compris entre 1 et  $n$ , on note  $f_i$  la densité de  $P_i$  par rapport à  $Q_i$ . La densité  $f$  de  $P$  par rapport à  $Q$  est alors donnée, en  $(x_1, \dots, x_n)$ , par :

$$f(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i) .$$

La distance de Kullback entre  $P$  et  $Q$  s'écrit alors :

$$\begin{aligned}
 d_K^2(P, Q) &= \int_{\Omega_1 \times \dots \times \Omega_n} \left( \prod_{i=1}^n f_i \right) \log \left( \prod_{i=1}^n f_i \right) dQ_1 \dots dQ_n \\
 &= \sum_{i=1}^n \int_{\Omega_1 \times \dots \times \Omega_n} \left( \prod_{j=1}^n f_j \right) \log(f_i) dQ_1 \dots dQ_n \\
 &= \sum_{i=1}^n \left( \prod_{j \neq i} \underbrace{\int_{\Omega_j} f_j dQ_j}_{=1} \right) \left( \int_{\Omega_i} f_i \log(f_i) dQ_i \right) \\
 &= \sum_{i=1}^n d_K^2(P_i, Q_i) .
 \end{aligned}$$

□

## 3.2 Cas général des processus de Markov exponentiellement convergents

Dans cette partie nous allons montrer qu'un  $n$ -échantillon de processus de Markov présente un phénomène de cutoff dès que le processus qu'on a échantillonné converge à vitesse exponentielle vers sa mesure stationnaire.

Commençons par définir ce que l'on entend par processus exponentiellement convergent. Pour cela on a besoin d'une distance entre lois de probabilité, que l'on note  $d$  et qui peut être n'importe laquelle des distances considérées dans la partie précédente. De plus, pour un processus stochastique  $X$ , on note  $\mathcal{L}(X(t))$  la loi de  $X$  à l'instant  $t$ . Désormais, on suppose que le processus  $X$  admet une loi stationnaire  $\nu$ , et que la loi de  $X$  à l'instant  $t$  est absolument continue par rapport à la loi  $\nu$ . Ceci assure que toutes les distances considérées (variation totale, Hellinger,  $\chi^2$  et Kullback) entre la loi de  $X$  à l'instant  $t$  et  $\nu$  sont bien définies.

**Définition 3.2.1** Soit  $X = \{X(t); t \geq 0\}$  un processus stochastique, et soit  $\nu$  une loi de probabilité. Soit  $\rho$  un réel strictement positif. On dit que le processus  $X$  converge vers  $\nu$  à vitesse exponentielle de taux  $\rho$  au sens de la distance  $d$  s'il existe deux constantes strictement positives  $k$  et  $k'$  telles que, pour tout  $t$  positif :

$$k e^{-\rho t} \leq d(\mathcal{L}(X(t)), \nu) \leq k' e^{-\rho t} . \quad (3.2.11)$$

Les différentes inégalités obtenues dans la partie précédente entre les distances (propositions 3.1.13, 3.1.18 et 3.1.22) permettent d'établir des implications entre les convergences exponentielles d'un processus au sens de différentes distances. En général, ce ne sont pas des équivalences. Cependant, les deux cas de la distance en variation totale et de la distance du

$\chi^2$  ont été largement étudiés dans la littérature, dans le cas des processus de Markov. Les définitions classiques de convergence exponentielle ne correspondent pas à la définition 3.2.1 ci-dessus. Rappelons brièvement les définitions usuelles, avant de les comparer avec la nôtre. Ce sont les définitions d'ergodicité exponentielle et d'ergodicité exponentielle dans  $L^2$  que l'on peut lire dans le livre de M.F. Chen [14], au chapitre 4 p. 144-145 pour l'ergodicité exponentielle, et au chapitre 9 p. 311 pour l'ergodicité exponentielle dans  $L^2$ .

**Définition 3.2.2** Soit  $X = \{X(t); t \geq 0\}$  un processus stochastique, et soit  $\nu$  une loi de probabilité. On dit que le processus  $X$  est exponentiellement ergodique s'il existe deux constantes  $k$  et  $\rho$  strictement positives telles que, pour tout  $t$  positif :

$$d_{VT}(\mathcal{L}(X(t)), \nu) \leq k e^{-\rho t} .$$

Il existe beaucoup de caractérisations équivalentes de cette notion ; on peut citer l'article de G.O. Roberts et J.S. Rosenthal [52] par exemple où ces différentes caractérisations sont résumées.

La définition suivante est relative à  $L^2(\nu)$ , espace des fonctions de carré intégrable par rapport à la mesure  $\nu$ . Nous notons  $\|\cdot\|_{L^2(\nu)}$  sa norme usuelle :

$$\|f\|_{L^2(\nu)} = \left( \int_E f^2 d\nu \right)^{\frac{1}{2}} .$$

Par ailleurs, nous adoptons les notations standard en ce qui concerne le semi-groupe de transition d'un processus de Markov, que nous avons déjà introduites dans la partie 3.1.1 concernant la distance en variation totale, à la proposition 3.1.5. Plus précisément, si  $X$  est un processus de Markov de semi-groupe de transition  $P$ , nous savons que, pour tout  $t$  positif,  $P_t$  est un opérateur défini de l'ensemble des fonctions mesurables bornées  $\mathcal{C}_b$  dans lui-même de la façon suivante :

$$\forall f \in \mathcal{C}_b, \forall x \in \mathbb{R}, \quad P_t f(x) = E[f(X(t)) | X(0) = x] .$$

Ainsi défini, l'opérateur  $P_t$  est de norme 1, au sens où :

$$\sup \{ \|P_t f\|_\infty ; \|f\|_\infty \leq 1 \} = 1 .$$

Grâce à l'invariance de  $\nu$ , on peut définir la norme dans  $L^2(\nu)$  de  $P_t f$  pour des fonctions  $f$  appartenant à  $L^\infty(\nu) \cap L^2(\nu)$ , puis l'étendre aux fonctions de  $L^2(\nu)$ .

Si  $f$  est une fonction de  $L^2(\nu)$  définie sur  $E$ , nous notons aussi  $\nu f$  l'intégrale de la fonction  $f$  par rapport à la mesure  $\nu$  :

$$\nu f = \int_E f d\nu .$$

**Définition 3.2.3** Soit  $X$  un processus de Markov à valeurs dans l'espace  $E$ , dont on note  $P$  le semi-groupe de transition. On dit que le processus  $X$  est exponentiellement ergodique dans  $L^2(\nu)$  s'il existe une constante  $\rho$  strictement positive telle que, pour toute fonction  $f$  de  $L^2(\nu)$  et pour tout  $t$  positif :

$$\|P_t f - \nu f\|_{L^2(\nu)} \leq k(f) e^{-\rho t} ,$$

où  $k(f)$  est une constante strictement positive qui dépend de  $f$ .

Précisons maintenant le lien entre l'ergodicité exponentielle dans  $L^2(\nu)$  définie ci-dessus et la convergence exponentielle au sens de la distance du  $\chi^2$ .

Avec les notations standard toujours, nous notons  $\nu P_t$  la loi de  $X(t)$  si  $X(0)$  suit la loi  $\nu$  (ce que nous rappelons en mettant  $\nu$  en indice de  $P$ ) :

$$\forall t \geq 0, \forall A \in \mathcal{A}, \quad \nu P_t(A) = P_\nu(X(t) \in A) .$$

Supposons alors que la mesure  $\nu$  soit réversible sous l'action du semi-groupe  $P$ , ce qui signifie que :

$$\forall t \geq 0, \quad P_t(x, dy) \nu(dx) = P_t(y, dx) \nu(dy) .$$

Alors si l'on note  $\mu$  la loi de  $X(0)$ , et  $f$  la densité de  $\mu$  par rapport à  $\nu$ , alors on obtient, pour tout  $t$  positif :

$$d_{\chi^2}(\mathcal{L}(X(t)), \nu) = \|P_t f - \nu f\|_{L^2(\nu)} . \quad (3.2.12)$$

La comparaison entre l'ergodicité exponentielle et l'ergodicité exponentielle dans  $L^2(\nu)$  a été étudiée par de nombreux auteurs, dont G.O. Roberts et J.S. Rosenthal [52], et G.O. Roberts et R.L. Tweedie [54]. Dans le théorème 2.1 de [52], et dans le théorème 2 de [54], les auteurs démontrent que les deux notions sont équivalentes pour une chaîne de Markov ergodique, si la mesure  $\nu$  est réversible et si l'espace d'états est muni d'une tribu engendrée par un ensemble dénombrable de parties. Dans l'article [15], M.F. Chen étend ce résultat aux chaînes de Markov à temps continu. Dans le cas où l'espace d'états est dénombrable, il remplace la condition de réversibilité par une condition plus faible. Tous ces auteurs s'accordent sur le fait que la réversibilité est une condition essentiellement technique.

Maintenant nous considérons un  $n$ -échantillon  $X^{(n)}$  de processus, c'est-à-dire un  $n$ -uplet  $(X_1, \dots, X_n)$  de processus qui sont des copies indépendantes du processus  $X$  de départ :

$$X^{(n)} = (X_1, \dots, X_n) .$$

Nous avons supposé que le processus  $X$  avait pour loi stationnaire  $\nu$ , ce qui implique que le  $n$ -échantillon  $X^{(n)}$  a lui aussi une loi stationnaire, qui est la mesure  $\nu$  tensorisée par elle-même  $n$  fois, que l'on note  $\nu^{\otimes n}$ . Le théorème 3.2.4 ci-dessous montre que la convergence du  $n$ -échantillon présente un phénomène de cutoff à l'instant  $\log(n)/(2\rho)$  où  $\rho$  est le taux de la vitesse de convergence exponentielle du processus échantillonné, et ce pour chacune des distances considérées.

**Théorème 3.2.4** • *Soit  $\rho$  un réel strictement positif. Soit  $X$  un processus stochastique qui converge à vitesse exponentielle de taux  $\rho$  vers sa loi asymptotique  $\nu$ , au sens de la distance de Hellinger (resp. du  $\chi^2$ , de Kullback). Soit  $X^{(n)}$  un  $n$ -échantillon du processus  $X$ . Alors le processus  $X^{(n)}$  présente un phénomène de cutoff à l'instant  $\frac{\log(n)}{2\rho}$  au sens de la distance de Hellinger (resp. du  $\chi^2$ , de Kullback).*

- *Soit  $\rho$  un réel strictement positif. Soit  $X$  un processus stochastique qui converge à vitesse exponentielle de taux  $\rho$  vers sa loi asymptotique  $\nu$ , au sens de la distance en variation*

totale et d'une autre distance (Hellinger,  $\chi^2$  ou Kullback). Soit  $X^{(n)}$  un  $n$ -échantillon du processus  $X$ . Alors le processus  $X^{(n)}$  présente un phénomène de cutoff à l'instant  $\frac{\log(n)}{2\rho}$  au sens de la distance en variation totale.

### Démonstration.

Nous traitons successivement les cas des quatre distances. Pour mettre en évidence la présence d'un cutoff, nous cherchons à vérifier la définition de base (définition 1.2.1).

#### 1. Distance de Hellinger.

Par hypothèse, le processus  $X$  vérifie la propriété de convergence exponentielle de taux  $\rho$  donnée par l'inégalité (3.2.11), c'est-à-dire qu'il existe deux réels  $k$  et  $k'$  strictement positifs tels que, pour tout  $t$  positif :

$$ke^{-\rho t} \leq d_H(\mathcal{L}(X(t)), \nu) \leq k'e^{-\rho t} .$$

Et nous avons vu dans la proposition 3.1.14 comment obtenir la distance de Hellinger entre les lois produits en fonction de la distance de Hellinger entre les lois marginales. Ici nous obtenons :

$$d_H^2(\mathcal{L}(X^{(n)}(t)), \nu^{\otimes n}) = 1 - (1 - d_H^2(\mathcal{L}(X(t)), \nu))^n .$$

Nous avons donc un encadrement de la distance de Hellinger entre la loi du  $n$ -échantillon à l'instant  $t$  et sa loi asymptotique :

$$1 - (1 - k^2 e^{-2\rho t})^n \leq d_H^2(\mathcal{L}(X^{(n)}(t)), \nu^{\otimes n}) \leq 1 - (1 - k'^2 e^{-2\rho t})^n .$$

Soit maintenant  $c$  un réel strictement positif. Regardons la distance de Hellinger à l'instant  $c \log(n)/(2\rho)$  ; nous obtenons :

$$1 - (1 - k^2 n^{-c})^n \leq d_H^2 \left( \mathcal{L} \left( X^{(n)} \left( \frac{c \log(n)}{2\rho} \right) \right), \nu^{\otimes n} \right) \leq 1 - (1 - k'^2 n^{-c})^n .$$

Si  $c$  est strictement inférieur à 1, le membre de gauche de l'inégalité tend vers 1 quand  $n$  tend vers l'infini, ce qui implique que la distance de Hellinger entre la loi du  $n$ -échantillon à l'instant  $c \log(n)/(2\rho)$  et sa loi asymptotique tend aussi vers 1. Et si  $c$  est strictement supérieur à 1, le membre de droite de l'inégalité tend vers 0, ce qui implique que la distance de Hellinger tend aussi vers 0. Ainsi le cutoff est démontré au sens de la distance de Hellinger.

#### 2. Distance du $\chi^2$ .

Nous procédons exactement de la même façon que pour la distance de Hellinger. Par l'hypothèse de convergence exponentielle, nous avons :

$$ke^{-\rho t} \leq d_{\chi^2}(\mathcal{L}(X(t)), \nu) \leq k'e^{-\rho t} .$$

Ensuite la proposition 3.1.19 donne la distance du  $\chi^2$  entre les lois produits en fonction de la distance du  $\chi^2$  entre les lois marginales :

$$d_{\chi^2}^2(\mathcal{L}(X^{(n)}(t)), \nu^{\otimes n}) = (1 + d_{\chi^2}^2(\mathcal{L}(X(t)), \nu))^n - 1 .$$

Nous avons donc un encadrement de la distance du  $\chi^2$  entre la loi du  $n$ -échantillon à l'instant  $t$  et sa loi asymptotique :

$$(1 + k^2 e^{-2\rho t})^n - 1 \leq d_{\chi^2}^2(\mathcal{L}(X^{(n)}(t)), \nu^{\otimes n}) \leq (1 + k'^2 e^{-2\rho t})^n - 1 .$$

Pour tout réel strictement positif  $c$ , nous obtenons donc un encadrement de la distance du  $\chi^2$  à l'instant  $c \log(n)/(2\rho)$  :

$$(1 + k^2 n^{-c})^n - 1 \leq d_{\chi^2}^2 \left( \mathcal{L} \left( X^{(n)} \left( \frac{c \log(n)}{2\rho} \right) \right), \nu^{\otimes n} \right) \leq (1 + k'^2 n^{-c})^n - 1 .$$

Si  $c$  est strictement inférieur à 1, le membre de gauche de l'inégalité tend vers  $+\infty$  quand  $n$  tend vers l'infini, ce qui implique que la distance du  $\chi^2$  entre la loi du  $n$ -échantillon à l'instant  $c \log(n)/(2\rho)$  et sa loi asymptotique tend aussi vers  $+\infty$ . Et si  $c$  est strictement supérieur à 1, le membre de droite de l'inégalité tend vers 0, ce qui implique que la distance du  $\chi^2$  tend aussi vers 0. Ainsi le cutoff est démontré au sens de la distance du  $\chi^2$ .

### 3. Distance de Kullback.

Dans ce cas, nous faisons encore le même raisonnement, simplifié par le fait que la distance de Kullback entre lois produits s'écrit comme somme des distances de Kullback entre les marginales (proposition 3.1.24) :

$$d_K^2(\mathcal{L}(X^{(n)}(t)), \nu^{\otimes n}) = n d_K^2(\mathcal{L}(X(t)), \nu) .$$

Par l'hypothèse de convergence exponentielle, nous avons :

$$k e^{-\rho t} \leq d_K(\mathcal{L}(X(t)), \nu) \leq k' e^{-\rho t} .$$

Pour la distance de Kullback entre la loi du  $n$ -échantillon et sa loi asymptotique, nous obtenons donc :

$$n k^2 e^{-2\rho t} \leq d_K^2(\mathcal{L}(X^{(n)}(t)), \nu^{\otimes n}) \leq n k'^2 e^{-2\rho t} .$$

Pour tout réel strictement positif  $c$ , nous en déduisons un encadrement de la distance de Kullback à l'instant  $c \log(n)/(2\rho)$  :

$$k^2 n^{1-c} \leq d_K^2 \left( \mathcal{L} \left( X^{(n)} \left( \frac{c \log(n)}{2\rho} \right) \right), \nu^{\otimes n} \right) \leq k'^2 n^{1-c} .$$

Si  $c$  est strictement inférieur à 1, le membre de gauche de l'inégalité tend vers  $+\infty$  quand  $n$  tend vers l'infini, ce qui implique que la distance de Kullback entre la loi du  $n$ -échantillon à l'instant  $c \log(n)/(2\rho)$  et sa loi asymptotique tend aussi vers  $+\infty$ . Et si  $c$  est strictement supérieur à 1, le membre de droite de l'inégalité tend vers 0, ce qui implique que la distance de Kullback tend aussi vers 0. Ainsi le cutoff est démontré au sens de la distance de Kullback.

### 4. Distance en variation totale.

L'hypothèse de convergence exponentielle donne l'existence de deux réels  $k$  et  $k'$  strictement positifs tels que, pour tout  $t$  positif :

$$k e^{-\rho t} \leq d_{VT}(\mathcal{L}(X(t)), \nu) \leq k' e^{-\rho t} .$$

Là encore nous utilisons l'expression de la distance entre les lois produits en fonction de la distance entre les lois marginales ; mais dans ce cas nous n'avons qu'une inégalité, donnée dans la proposition 3.1.7 :

$$1 - 2 \exp\left(-\frac{n}{2} d_{VT}^2(\mathcal{L}(X(t)), \nu)\right) \leq d_{VT}(\mathcal{L}(X^{(n)}(t)), \nu^{\otimes n}) \leq n d_{VT}(\mathcal{L}(X(t)), \nu).$$

Soit  $c$  un réel strictement positif. La borne de gauche de l'inégalité ci-dessus nous permet de conclure quant au cutoff à gauche (c'est-à-dire pour  $c$  inférieur à 1). En effet, à l'instant  $c \log(n)/(2\rho)$ , nous avons l'inégalité :

$$1 - 2 \exp\left(-\frac{k^2}{2} n^{1-c}\right) \leq d_{VT}(\mathcal{L}(X^{(n)}(t)), \nu^{\otimes n}),$$

qui implique que, quand  $n$  tend vers l'infini, la distance en variation totale entre la loi du  $n$ -échantillon et sa loi asymptotique tend vers 1. Par contre, la borne de droite ne suffit pas pour conclure à la présence du phénomène de cutoff ; nous utilisons alors l'hypothèse de la convergence exponentielle au sens d'une autre distance. En effet, nous avons vu que la distance en variation totale était majorée par chacune des autres distances (d'après l'équation (3.1.7) et les propositions 3.1.18 et 3.1.22), ce qui donne :

$$\begin{aligned} d_{VT}(\mathcal{L}(X^{(n)}(t)), \nu^{\otimes n}) &\leq \sqrt{2} d_H(\mathcal{L}(X^{(n)}(t)), \nu^{\otimes n}), \\ d_{VT}(\mathcal{L}(X^{(n)}(t)), \nu^{\otimes n}) &\leq \frac{1}{2} d_{\chi^2}(\mathcal{L}(X^{(n)}(t)), \nu^{\otimes n}), \\ d_{VT}(\mathcal{L}(X^{(n)}(t)), \nu^{\otimes n}) &\leq \frac{1}{\sqrt{2}} d_K(\mathcal{L}(X^{(n)}(t)), \nu^{\otimes n}). \end{aligned}$$

Ainsi en utilisant le même raisonnement que dans chacun des trois points précédents concernant chacune des distances, nous concluons qu'il y a aussi cutoff à droite (c'est-à-dire quand  $c$  est supérieur à 1).

□

Désormais nous notons  $(t_n)_{n \in \mathbb{N}^*}$  la suite des instants de cutoff :

$$t_n = \frac{\log(n)}{2\rho}.$$

Dans le théorème précédent nous avons supposé que la convergence du processus échantillonné était exponentielle, c'est-à-dire que la distance entre sa loi à l'instant  $t$  et sa loi asymptotique était encadrée par deux exponentielles décroissant à la même vitesse. Nous allons maintenant étudier le cas plus particulier où la distance est de l'ordre d'une exponentielle décroissante. Nous obtenons alors des estimations précises du comportement de la convergence à gauche et à droite de l'instant de cutoff.



Notons  $d(t)$  la distance (en variation totale, de Hellinger, du  $\chi^2$  ou de Kullback) entre la loi du processus  $X$  à l'instant  $t$  et sa loi asymptotique  $\nu$  :

$$d(t) = d(\mathcal{L}(X(t)), \nu) ,$$

et  $d^{(n)}(t)$  la distance (en variation totale, de Hellinger, du  $\chi^2$  ou de Kullback) entre la loi du processus  $X^{(n)}$  à l'instant  $t$  et sa loi asymptotique  $\nu^{\otimes n}$  :

$$d^{(n)}(t) = d(\mathcal{L}(X^{(n)}(t)), \nu^{\otimes n}) .$$

**Théorème 3.2.5** *Supposons que, pour  $t$  assez grand, la fonction  $d$  a la forme suivante :*

$$d(t) = C_d e^{-\rho t} + o(e^{-\rho t}) ,$$

où  $C_d$  est une constante qui dépend de la distance. Soit  $u$  un réel tel que  $u + t_n$  soit strictement positif. Alors :

1. *Distance en variation totale :*

$$1 - 2 \exp\left(-\frac{C_{VT}^2}{2} e^{-2\rho u}\right) \leq \liminf_{n \rightarrow +\infty} d_{VT}^{(n)}(t_n + u) ,$$

et :

$$\limsup_{n \rightarrow +\infty} d_{VT}^{(n)}(t_n + u) \leq \left(1 - \exp(-2C_H^2 e^{-2\rho u})\right)^{\frac{1}{2}} .$$

2. *Distance de Hellinger :*

$$\lim_{n \rightarrow +\infty} d_H^{(n)}(t_n + u) = \left(1 - \exp(-C_H^2 e^{-2\rho u})\right)^{\frac{1}{2}} .$$

3. *Distance du  $\chi^2$  :*

$$\lim_{n \rightarrow +\infty} d_{\chi^2}^{(n)}(t_n + u) = \left(\exp(C_{\chi^2}^2 e^{-2\rho u}) - 1\right)^{\frac{1}{2}} .$$

4. *Distance de Kullback :*

$$\lim_{n \rightarrow +\infty} d_K^{(n)}(t_n + u) = C_K e^{-\rho u} .$$

### Démonstration.

Pour les distances de Hellinger, du  $\chi^2$  et de Kullback, elle repose sur les expressions des distances entre  $n$ -échantillons en fonction des distances entre les marginales, données par les propositions 3.1.14 (Hellinger), 3.1.19 ( $\chi^2$ ) et 3.1.24 (Kullback). Pour ces trois distances, un calcul de limite direct permet d'obtenir le résultat annoncé. Pour la distance en variation totale, la minoration est obtenue grâce à la proposition 3.1.7, qui donne une minoration de la distance en variation totale entre  $n$ -échantillons en fonction de la distance en variation totale entre les marginales. La proposition 3.1.7 donne aussi une majoration, mais celle-ci est insuffisante pour obtenir la majoration de la limite annoncée. Nous utilisons donc la majoration de la distance en variation totale par la distance de Hellinger donnée par la proposition 3.1.13. Nous obtenons

l'existence de deux suites de fonctions  $(\varepsilon_n)_{n \in \mathbb{N}}$  et  $(\varepsilon'_n)_{n \in \mathbb{N}}$  convergeant simplement vers 0 quand  $n$  tend vers  $+\infty$  telles que :

$$1 - 2 \exp\left(-\frac{C_{VT}^2}{2} e^{-2\rho u}\right) + \varepsilon_n(u) \leq d_{VT}^{(n)}(t_n + u) \leq \left(1 - \exp(-2C_H^2 e^{-2\rho u})\right)^{\frac{1}{2}} + \varepsilon'_n(u), \quad (3.2.13)$$

ce qui implique les deux inégalités annoncées.  $\square$

Il faut noter que les résultats du théorème 3.2.5 sont plus forts que le cutoff tel qu'il est décrit dans la définition 1.2.1. En effet, ils affirment que la distance passe de presque 1 à presque 0 dans un intervalle de temps dont la longueur est bornée quand  $n$  tend vers  $+\infty$ , alors que la définition 1.2.1 du cutoff impose seulement que le passage de presque 1 à presque 0 ait lieu dans un intervalle de temps dont la longueur est en  $O(\log(n))$ . Énonçons-le plus précisément, tout d'abord pour la distance en variation totale.

**Proposition 3.2.6** *Supposons que le processus échantillonné soit markovien, et qu'il existe deux suites de fonctions  $(\varepsilon_n)_{n \in \mathbb{N}}$  et  $(\varepsilon'_n)_{n \in \mathbb{N}}$  convergeant simplement vers 0 quand  $n$  tend vers  $+\infty$  telles que :*

$$1 - 2 \exp\left(-\frac{C_{VT}^2}{2} e^{-2\rho u}\right) + \varepsilon_n(u) \leq d_{VT}^{(n)}(t_n + u) \leq \left(1 - \exp(-2C_H^2 e^{-2\rho u})\right)^{\frac{1}{2}} + \varepsilon'_n(u).$$

Alors la suite  $(t_n)_{n \in \mathbb{N}^*}$  est une suite d'instants de cutoff au sens de la définition 1.2.1, c'est-à-dire que, pour tout réel  $c$  strictement positif :

$$\begin{cases} c < 1 & \Rightarrow \liminf_{n \rightarrow +\infty} d_{VT}^{(n)}(ct_n) > 0, \\ c > 1 & \Rightarrow \lim_{n \rightarrow +\infty} d_{VT}^{(n)}(ct_n) = 0. \end{cases}$$

### Démonstration.

Soit  $c$  un réel strictement supérieur à 1. Montrons que :

$$\lim_{n \rightarrow +\infty} d_{VT}^{(n)}(ct_n) = 0.$$

Soit  $\varepsilon > 0$ . La fonction qui à  $u$  associe  $\left(1 - \exp(-2C_H^2 e^{-2\rho u})\right)^{\frac{1}{2}}$  tend vers 0 quand  $u$  tend vers  $+\infty$ , donc :

$$\exists A > 0, \quad \forall u \geq A, \quad \left| \left(1 - \exp(-2C_H^2 e^{-2\rho u})\right)^{\frac{1}{2}} \right| < \varepsilon.$$

De plus, la suite  $((c-1)t_n)_{n \in \mathbb{N}^*}$  tend vers  $+\infty$ , donc :

$$\exists n_0 \in \mathbb{N}, \quad \forall n \geq n_0, \quad (c-1)t_n \geq A.$$

Enfin, la suite  $(\varepsilon'_n)_{n \in \mathbb{N}}$  tend simplement vers 0, ce qui donne :

$$\exists n_1 \in \mathbb{N}, \quad \forall n \geq n_1, \quad |\varepsilon'_n(A)| < \varepsilon.$$

Ainsi, pour  $n \geq \max(n_0, n_1)$ , on écrit :

$$ct_n = t_n + (c - 1)t_n \geq t_n + A .$$

Or le processus échantillonné est markovien, ce qui implique que la fonction  $d_{VT}^{(n)}$  est décroissante (d'après la proposition 3.1.5), donc nous obtenons :

$$\begin{aligned} d_{VT}^{(n)}(ct_n) &\leq d_{VT}^{(n)}(t_n + A) \\ &\leq \left(1 - \exp(-2C_H^2 e^{-2\rho A})\right)^{\frac{1}{2}} + \varepsilon'_n(A) \\ &\leq 2\varepsilon . \end{aligned}$$

Pour  $c$  strictement inférieur à 1, on procède de la même façon, en utilisant la partie gauche de l'inégalité (3.2.13), et le fait que la fonction qui à  $u$  associe  $1 - 2 \exp\left(-\frac{C_{VT}^2}{2} e^{-2\rho u}\right)$  tend vers 1 quand  $u$  tend vers  $-\infty$ .  $\square$

Remarquons que le fait de supposer le processus markovien permet d'obtenir la décroissance de la distance en variation totale, et c'est le point crucial de la démonstration.

Pour les trois autres distances (Hellinger,  $\chi^2$  et Kullback), comme la distance entre  $n$ -échantillons s'exprime en fonction de la distance entre marginales, le fait que les résultats du théorème 3.2.5 impliquent le cutoff peut s'énoncer sous la forme générale suivante.

**Lemme 3.2.7** Soient  $\mu_1$  et  $\mu_2$  deux mesures sur le même espace  $(E, \mathcal{E})$ . Soit  $a$  un élément de  $[0 ; +\infty]$ . Soit  $d$  une distance entre mesures, qui vérifie :

$$\forall n \in \mathbb{N}^*, \quad d(\mu_1^{\otimes n}, \mu_2^{\otimes n}) = h_n(d(\mu_1, \mu_2)) ,$$

où, pour tout  $n$ , la fonction  $h_n$  est une fonction croissante de  $[0 ; a]$  dans  $[0 ; a]$ . Supposons de plus qu'il existe une fonction continue  $h$  de  $\mathbb{R}^+$  dans  $\mathbb{R}^+$  telle que :

$$\forall x \in \mathbb{R}^+, \quad \lim_{n \rightarrow +\infty} h_n\left(\frac{x}{\sqrt{n}}\right) = h(x) ,$$

et :

$$\lim_{x \rightarrow +\infty} h(x) = a .$$

Soit  $X$  un processus stochastique de loi stationnaire  $\nu$ . Avec les mêmes notations que précédemment, supposons que, lorsque  $t$  tend vers  $+\infty$  :

$$d(t) = C_d e^{-\rho t} + o(e^{-\rho t}) .$$

Soit  $(u_n)_{n \in \mathbb{N}^*}$  une suite réelle convergeant dans  $\mathbb{R} \cup \{-\infty; +\infty\}$ , telle que :

$$\lim_{n \rightarrow +\infty} t_n + u_n = +\infty .$$

Alors, en notant  $u$  la limite de la suite  $(u_n)_{n \in \mathbb{N}^*}$ , on a :

$$\lim_{n \rightarrow +\infty} d^{(n)}(t_n + u_n) = h(C_d e^{-\rho u}) .$$

**Démonstration.**

Écrivons tout d'abord  $d^{(n)}(t_n + u_n)$  sous la forme :

$$d^{(n)}(t_n + u_n) = h_n(d(t_n + u_n)) .$$

L'hypothèse selon laquelle :

$$d(t) = C_d e^{-\rho t} + o(e^{-\rho t})$$

signifie qu'il existe une fonction  $\varepsilon$  tendant vers 0 en  $+\infty$  telle que :

$$d(t) = e^{-\rho t} (C_d + \varepsilon(t)) .$$

Nous pouvons donc écrire :

$$\exists n_0 \in \mathbb{N}^*, \forall n \geq n_0, \quad d(t_n + u_n) = \frac{e^{-\rho u_n}}{\sqrt{n}} (C_d + \varepsilon(t_n + u_n)) .$$

Par hypothèse, la suite de fonctions croissantes  $(h_n(\cdot/\sqrt{n}))_{n \in \mathbb{N}^*}$  converge simplement vers la fonction  $h$  continue. Par le théorème de Dini, nous en déduisons que cette convergence est uniforme sur tout compact de  $\mathbb{R}^+$ . Par conséquent, si la limite  $u$  de la suite  $(u_n)_{n \in \mathbb{N}^*}$  est dans  $\mathbb{R} \cup \{+\infty\}$ , nous obtenons directement :

$$\lim_{n \rightarrow +\infty} d^{(n)}(t_n + u_n) = h(C_d e^{-\rho u}) .$$

Si  $u$  vaut  $-\infty$ , la suite  $(e^{-\rho u_n} (C_d + \varepsilon(t_n + u_n)))_{n \in \mathbb{N}^*}$  tend vers  $+\infty$ , ce qui s'écrit :

$$\forall A > 0, \exists n_1 \in \mathbb{N}^*, \forall n \geq n_1, \quad e^{-\rho u_n} (C_d + \varepsilon(t_n + u_n)) \geq A .$$

Fixons  $A$  strictement positif. Par croissance des fonctions  $h_n$ , nous obtenons :

$$h_n \left( \frac{e^{-\rho u_n}}{\sqrt{n}} (C_d + \varepsilon(t_n + u_n)) \right) \geq h_n \left( \frac{A}{\sqrt{n}} \right) ,$$

ce qui montre, en passant à la limite, que :

$$\liminf_{n \rightarrow +\infty} h_n \left( \frac{e^{-\rho u_n}}{\sqrt{n}} (C_d + \varepsilon(t_n + u_n)) \right) \geq h(A) .$$

Ceci étant vrai pour tout  $A$ , la continuité de  $h$  implique que c'est aussi vrai en  $+\infty$  :

$$\liminf_{n \rightarrow +\infty} h_n \left( \frac{e^{-\rho u_n}}{\sqrt{n}} (C_d + \varepsilon(t_n + u_n)) \right) \geq a .$$

Comme la fonction  $h_n$  est à valeurs dans  $[0 ; a]$ , on a le résultat. □

Appliquons le lemme 3.2.7 aux distances qui nous concernent. Dans tous les cas, nous voulons démontrer le phénomène de cutoff au sens de la définition 1.2.1. Pour cela, nous devons calculer la limite suivante, pour  $c$  strictement positif :

$$\lim_{n \rightarrow +\infty} d^{(n)}(ct_n) .$$

Nous allons donc utiliser le lemme 3.2.7 avec  $u_n = (c - 1)t_n$ , qui tend vers  $-\infty$  ou  $+\infty$  selon si  $c$  est inférieur ou supérieur à 1. Examinons les fonctions  $h_n$  dans les trois cas particuliers de distances, afin de conclure sur la présence du cutoff.

**Distance de Hellinger.** La proposition 3.1.14 donne l'expression de la distance entre probabilités produits en fonction de la distance entre les marginales ; si  $P$  et  $Q$  sont deux lois de probabilités sur le même espace, alors :

$$d_H^2(P^{\otimes n}, Q^{\otimes n}) = 1 - (1 - d_H^2(P, Q))^n ,$$

ce qui montre que, dans ce cas, la fonction  $h_n$  intervenant dans la proposition 3.1.14 est définie sur  $[0 ; 1]$  par :

$$h_n(x) = (1 - (1 - x^2)^n)^{\frac{1}{2}} .$$

C'est une fonction croissante à valeurs dans  $[0 ; 1]$ . La fonction  $h$  est alors définie sur  $\mathbb{R}^+$  par :

$$h(x) = \sqrt{1 - e^{-x^2}} .$$

La fonction  $h$  est continue sur  $\mathbb{R}^+$ , et sa limite en  $+\infty$  vaut 1. Le lemme 3.2.7 permet donc de conclure que, si  $c$  est un réel strictement compris entre 0 et 1 :

$$\lim_{n \rightarrow +\infty} d_H^{(n)}(ct_n) = 1 ,$$

et, si  $c$  est strictement supérieur à 1 :

$$\lim_{n \rightarrow +\infty} d_H^{(n)}(ct_n) = 0 .$$

**Distance du  $\chi^2$ .** La proposition 3.1.19 donne l'expression de la distance entre probabilités produits en fonction de la distance entre les marginales :

$$d_{\chi^2}^2(P^{\otimes n}, Q^{\otimes n}) = (1 + d_{\chi^2}^2(P, Q))^n - 1 ,$$

ce qui montre que, dans ce cas, la fonction  $h_n$  intervenant dans la proposition 3.1.14 est définie sur  $\mathbb{R}^+$  par :

$$h_n(x) = ((1 + x^2)^n - 1)^{\frac{1}{2}} .$$

C'est une fonction croissante à valeurs dans  $\mathbb{R}^+$ . La fonction  $h$  est alors définie sur  $\mathbb{R}^+$  par :

$$h(x) = \sqrt{e^{x^2} - 1} .$$

La fonction  $h$  est continue sur  $\mathbb{R}^+$ , et sa limite en  $+\infty$  vaut  $+\infty$ . Le lemme 3.2.7 permet donc de conclure que, si  $c$  est un réel strictement compris entre 0 et 1 :

$$\lim_{n \rightarrow +\infty} d_{\chi^2}^{(n)}(ct_n) = +\infty ,$$

et, si  $c$  est strictement supérieur à 1 :

$$\lim_{n \rightarrow +\infty} d_{\chi^2}^{(n)}(ct_n) = 0 .$$

**Distance de Kullback.** La proposition 3.1.24 donne l'expression de la distance entre probabilités produits en fonction de la distance entre les marginales :

$$d_K^2(P^{\otimes n}, Q^{\otimes n}) = n d_K^2(P, Q) ,$$

ce qui montre que, dans ce cas, la fonction  $h_n$  intervenant dans la proposition 3.1.14 est définie sur  $\mathbb{R}^+$  par :

$$h_n(x) = x\sqrt{n} .$$

C'est une fonction croissante à valeurs dans  $\mathbb{R}^+$ . La fonction  $h$  est alors définie sur  $\mathbb{R}^+$  par :

$$h(x) = x .$$

La fonction  $h$  est continue sur  $\mathbb{R}^+$ , et sa limite en  $+\infty$  vaut  $+\infty$ . Le lemme 3.2.7 permet donc de conclure que, si  $c$  est un réel strictement compris entre 0 et 1 :

$$\lim_{n \rightarrow +\infty} d_K^{(n)}(ct_n) = +\infty ,$$

et, si  $c$  est strictement supérieur à 1 :

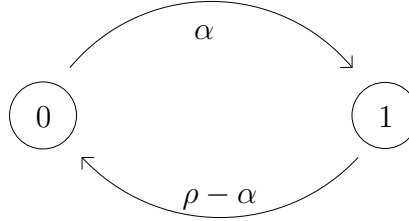
$$\lim_{n \rightarrow +\infty} d_K^{(n)}(ct_n) = 0 .$$

### 3.3 Exemples

Dans cette partie nous considérons deux nouveaux exemples : le processus de Markov à deux états et la file d'attente  $M/M/\infty$ , et nous revenons aussi sur le processus d'Ornstein-Uhlenbeck. Sur ce dernier exemple en effet, nous n'avons considéré au chapitre 2 que la distance en variation totale, il nous reste donc les autres distances à traiter. Ces trois exemples se ressemblent, dans le sens où la convergence vers la loi stationnaire a lieu à vitesse exponentielle. Cependant, ces trois processus nous permettent d'examiner le phénomène de cutoff sur trois familles de lois différentes (Bernoulli, Poisson et gaussienne), et de voir que, indépendamment de la famille de lois, les comportements à gauche et à droite de l'instant de cutoff sont semblables.

### 3.3.1 Le processus binaire

Soit  $\rho$  et  $\alpha$  deux réels strictement positifs, tels que  $\alpha$  soit strictement inférieur à  $\rho$ . Soit  $X$  le processus markovien binaire, dont les deux états sont 0 et 1, qui part de 0 et dont les taux de transition sont les suivants :



Son générateur infinitésimal est donné par la matrice  $\Lambda$  suivante :

$$\Lambda = \begin{pmatrix} -\alpha & \alpha \\ \rho - \alpha & -\rho + \alpha \end{pmatrix}.$$

Pour tout instant  $t$ , l'espérance  $m(t)$  de  $X(t)$  est alors calculable. En effet, le semi-groupe de transition  $P(t)$  à l'instant  $t$  est la matrice  $\exp(t\Lambda)$ . Les valeurs propres de la matrice  $\Lambda$  sont 0 et  $-\rho$ . On en déduit par exemple la diagonalisation suivante de  $\Lambda$  :

$$\Lambda = M \begin{pmatrix} 0 & 0 \\ 0 & -\rho \end{pmatrix} M^{-1},$$

où  $M$  est la matrice inversible :

$$M = \begin{pmatrix} 1 & \alpha \\ 1 & \alpha - \rho \end{pmatrix}.$$

Ensuite on écrit :

$$P(t) = M \begin{pmatrix} 1 & 0 \\ 0 & e^{-\rho t} \end{pmatrix} M^{-1},$$

ce qui donne, après calcul :

$$P(t) = \frac{1}{\rho} \begin{pmatrix} \rho - \alpha & \alpha \\ \rho - \alpha & \alpha \end{pmatrix} + \frac{e^{-\rho t}}{\rho} \begin{pmatrix} \alpha & -\alpha \\ -\rho + \alpha & \rho - \alpha \end{pmatrix}.$$

Comme le processus  $X$  ne prend que les valeurs 0 et 1, la fonction  $m$  est donnée par :

$$m(t) = P(X(t) = 1 | X(0) = 0),$$

ce qui donne, en lisant la formule de  $P$  :

$$m(t) = \frac{\alpha}{\rho} (1 - e^{-\rho t}).$$

De plus, la loi stationnaire du processus est la loi de Bernoulli de paramètre  $\theta$ , avec :

$$\theta = \frac{\alpha}{\rho}.$$

Passons en revue les calculs des quatre distances dans le cas des lois de Bernoulli afin de déterminer les constantes  $C_d$  dans les quatre cas et d'énoncer la proposition de cutoff.

**Distance en variation totale.** Dans la proposition 3.1.8, nous avons déjà calculé la distance en variation totale entre deux lois de Bernoulli. Entre la loi de Bernoulli  $b(\theta)$  de paramètre  $\theta$  et la loi de Bernoulli  $b(\theta(1 - \varepsilon))$  de paramètre  $\theta(1 - \varepsilon)$ , la distance en variation totale vaut :

$$d_{VT}(b(\theta(1 - \varepsilon)), b(\theta)) = \theta\varepsilon.$$

Nous obtenons donc l'expression de la fonction  $d_{VT}$  :

$$d_{VT}(t) = \frac{\alpha}{\rho} e^{-\rho t},$$

ce qui donne la constante  $C_{VT}$  dans ce cas :

$$C_{VT} = \frac{\alpha}{\rho}.$$

**Distance de Hellinger.** De même, la proposition 3.1.12 donne la distance de Hellinger entre lois de Bernoulli. Entre la loi de Bernoulli  $b(\theta(1 - \varepsilon))$  et la loi de Bernoulli  $b(\theta)$ , la distance de Hellinger vaut :

$$d_H^2(b(\theta(1 - \varepsilon)), b(\theta)) = 1 - \sqrt{(1 - \theta)(1 - \theta(1 - \varepsilon))} - \sqrt{\theta^2(1 - \varepsilon)}.$$

Quand  $\varepsilon$  tend vers 0, un développement limité montre que :

$$d_H^2(b(\theta(1 - \varepsilon)), b(\theta)) = \frac{\theta\varepsilon^2}{8(1 - \theta)} + o(\varepsilon^2).$$

Ainsi, pour la distance de Hellinger entre la loi du processus  $X$  à l'instant  $t$  et sa loi stationnaire, nous obtenons, quand  $t$  tend vers  $+\infty$  :

$$d_H(t) = \sqrt{\frac{\alpha}{8(\rho - \alpha)}} e^{-\rho t} + o(e^{-\rho t}),$$

ce qui donne la constante  $C_H$  dans ce cas :

$$C_H = \sqrt{\frac{\alpha}{8(\rho - \alpha)}}.$$

**Distance du  $\chi^2$ .** La proposition 3.1.17 donne la distance du  $\chi^2$  entre deux lois de Bernoulli. Entre la loi de Bernoulli  $b(\theta(1 - \varepsilon))$  et la loi de Bernoulli  $b(\theta)$ , la distance du  $\chi^2$  vaut :

$$d_{\chi^2}^2(b(\theta(1 - \varepsilon)), b(\theta)) = \frac{\theta\varepsilon^2}{1 - \theta}. \quad (3.3.14)$$



Ainsi, pour la distance du  $\chi^2$  entre la loi du processus  $X$  à l'instant  $t$  et sa loi stationnaire, nous obtenons, quand  $t$  tend vers  $+\infty$  :

$$d_{\chi^2}(t) = \sqrt{\frac{\alpha}{\rho - \alpha}} e^{-\rho t},$$

ce qui donne la constante  $C_{\chi^2}$  dans ce cas :

$$C_{\chi^2} = \sqrt{\frac{\alpha}{\rho - \alpha}}.$$

**Distance de Kullback.** La proposition 3.1.21 donne la distance de Kullback entre deux lois de Bernoulli. Entre la loi de Bernoulli  $b(\theta(1 - \varepsilon))$  et la loi de Bernoulli  $b(\theta)$ , la distance de Kullback vaut :

$$d_K^2(b(\theta(1 - \varepsilon)), b(\theta)) = \theta(1 - \varepsilon) \log(1 - \varepsilon) + (1 - \theta(1 - \varepsilon)) \log\left(\frac{1 - \theta(1 - \varepsilon)}{1 - \theta}\right).$$

Quand  $\varepsilon$  tend vers 0, un développement limité montre que :

$$d_K^2(b(\theta(1 - \varepsilon)), b(\theta)) = \frac{\theta \varepsilon^2}{2(1 - \theta)} + o(\varepsilon^2).$$

Ainsi, pour la distance de Kullback entre la loi du processus  $X$  à l'instant  $t$  et sa loi stationnaire, nous obtenons, quand  $t$  tend vers  $+\infty$  :

$$d_K(t) = \sqrt{\frac{\alpha}{2(\rho - \alpha)}} e^{-\rho t} + o(e^{-\rho t}),$$

ce qui donne la constante  $C_K$  dans ce cas :

$$C_K = \sqrt{\frac{\alpha}{2(\rho - \alpha)}}.$$

Nous sommes maintenant en mesure d'énoncer la proposition qui établit le phénomène de cutoff dans le cas du processus binaire. Il s'agit du théorème 3.2.5 appliqué avec les constantes trouvées dans ce cas particulier.

**Proposition 3.3.1** *Si  $X^{(n)}$  est un  $n$ -échantillon de processus de Markov binaires, nous avons les estimations suivantes :*

1. *Distance en variation totale :*

$$1 - 2 \exp\left(-\frac{\alpha^2}{2\rho^2} e^{-2\rho u}\right) \leq \liminf_{n \rightarrow +\infty} d_{VT}^{(n)}(t_n + u),$$

et :

$$\limsup_{n \rightarrow +\infty} d_{VT}^{(n)}(t_n + u) \leq \left(1 - \exp\left(-\frac{\alpha}{4(\rho - \alpha)} e^{-2\rho u}\right)\right)^{\frac{1}{2}}.$$

2. *Distance de Hellinger* :

$$\lim_{n \rightarrow +\infty} d_H^{(n)}(t_n + u) = \left( 1 - \exp \left( -\frac{\alpha}{8(\rho - \alpha)} e^{-2\rho u} \right) \right)^{\frac{1}{2}}.$$

3. *Distance du  $\chi^2$*  :

$$\lim_{n \rightarrow +\infty} d_{\chi^2}^{(n)}(t_n + u) = \left( \exp \left( \frac{\alpha}{\rho - \alpha} e^{-2\rho u} \right) - 1 \right)^{\frac{1}{2}}.$$

4. *Distance de Kullback* :

$$\lim_{n \rightarrow +\infty} d_K^{(n)}(t_n + u) = \sqrt{\frac{\alpha}{2(\rho - \alpha)} e^{-\rho u}}.$$

### 3.3.2 La file $M/M/\infty$

La référence principale de cet exemple est le chapitre 6 du livre de P. Robert [51] sur les réseaux et files d'attente.

Dans un cadre très général, une file d'attente est la donnée d'une ou plusieurs unités de service où arrivent des clients qui demandent une certaine durée d'utilisation de cette unité. Quand les clients ne peuvent pas accéder à l'unité, ils patientent dans une file d'attente en attendant d'être servis. Mathématiquement, une file d'attente est définie par :

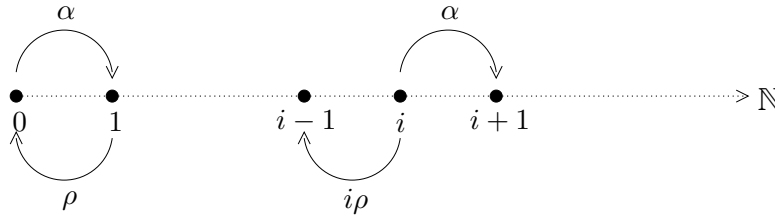
- un processus d'arrivée de clients, c'est-à-dire une suite aléatoire croissante d'instants,
- les durées de service des clients, c'est-à-dire une suite de variables aléatoires positives,
- et une discipline de service, c'est-à-dire une règle qui détermine selon quel procédé sont servis les clients de la file d'attente. Une des disciplines les plus fréquemment utilisées est la discipline FIFO (First In First Out), qui revient à servir les clients selon leur ordre d'arrivée.

Pour la file  $M/M/\infty$ , il y a une infinité de serveurs, le processus d'arrivée des clients est un processus de Poisson d'intensité  $\alpha$  (où  $\alpha$  est un réel strictement positif), et les durées de service des clients sont des variables aléatoires indépendantes et de même loi exponentielle de paramètre  $\rho$  (où  $\rho$  est un réel strictement positif).

Soit  $X$  le processus qui représente le nombre de clients dans la file d'attente. C'est un processus markovien de sauts d'espace d'états  $\mathbb{N}$ , et dont les taux de transitions sont :

$$\begin{aligned} q_{i,i+1} &= \alpha, & \text{pour } i \in \mathbb{N}, \\ q_{i,i-1} &= i\rho, & \text{pour } i \in \mathbb{N}^*, \\ q_{i,j} &= 0, & \text{pour } |i - j| > 1. \end{aligned}$$

Nous pouvons le représenter de la façon suivante :



Fixons de plus le nombre de clients dans la file au départ égal à un entier positif  $x_0$  :

$$X(0) = x_0 .$$

La fonction génératrice  $\Psi_t$  de  $X(t)$  est connue (cf. [51] p. 142) ; pour tout  $u$  de l'intervalle  $[0 ; 1]$ , la fonction  $\Psi_t$  est définie par :

$$\Psi_t(u) = (1 + (u - 1)e^{-\rho t})^{x_0} \exp\left(\frac{\alpha}{\rho}(1 - e^{-\rho t})(u - 1)\right) .$$

En prenant la dérivée de la fonction  $\Psi_t$  à gauche en 1, nous obtenons l'espérance de la variable aléatoire  $X(t)$ , que nous notons  $m(t)$  :

$$m(t) = \frac{\alpha}{\rho} + \left(x_0 - \frac{\alpha}{\rho}\right) e^{-\rho t} .$$

Si le processus  $X$  part de 0 ( $x_0 = 0$ ), alors la loi du processus  $X$  à l'instant  $t$  est la loi de Poisson de paramètre  $\Theta(t)$  donné par :

$$\Theta(t) = \frac{\alpha}{\rho} (1 - e^{-\rho t}) ,$$

et sa loi stationnaire est la loi de Poisson de paramètre  $\theta = \alpha/\rho$ .

Plaçons-nous dans le cas où  $X(0) = 0$  et intéressons-nous successivement aux quatre distances afin de déterminer les constantes  $C_d$  qui interviennent dans le théorème 3.2.5.

**Distance en variation totale.** Nous devons calculer la distance en variation totale entre deux lois de Poisson. Ou plus exactement nous devons évaluer le comportement, quand  $\varepsilon$  tend vers 0 de la distance entre la loi de Poisson de paramètre  $\theta$  et la loi de Poisson de paramètre  $\theta(1 - \varepsilon)$ . Cette estimation est donnée par la proposition 3.1.8 :

$$d_{VT}(\mathcal{P}(\theta(1 - \varepsilon)), \mathcal{P}(\theta)) = \frac{\theta^{1+\lfloor \theta \rfloor} e^{-\theta} \varepsilon}{\lfloor \theta \rfloor!} + o(\varepsilon) ,$$

où  $\lfloor \theta \rfloor$  désigne la partie entière de  $\theta$ .

En utilisant ce résultat avec  $\theta = \alpha/\rho$  et  $\varepsilon = e^{-\rho t}$ , nous obtenons l'expression de la fonction  $d_{VT}$  :

$$d_{VT}(t) = C_{VT} e^{-\rho t} + o(e^{-\rho t}) ,$$

où la constante  $C_{VT}$  est donnée par :

$$C_{VT} = \frac{1}{[\frac{\alpha}{\rho}]!} \left(\frac{\alpha}{\rho}\right)^{[\frac{\alpha}{\rho}]+1} e^{-\frac{\alpha}{\rho}}.$$

**Distance de Hellinger.** La distance de Hellinger entre deux lois de Poisson est donnée dans la proposition 3.1.12 :

$$d_H^2(\mathcal{P}(\theta), \mathcal{P}(\theta')) = 1 - e^{-\frac{\theta+\theta'}{2}} e^{\sqrt{\theta\theta'}}.$$

Pour calculer la distance de Hellinger entre la loi du processus à l'instant  $t$  et sa loi stationnaire, nous devons évaluer la distance, quand  $\varepsilon$  tend vers 0, entre la loi de Poisson de paramètre  $\theta$  et la loi de Poisson de paramètre  $\theta(1 - \varepsilon)$ , ce qui donne, d'après la formule ci-dessus :

$$d_H^2(\mathcal{P}(\theta(1 - \varepsilon)), \mathcal{P}(\theta)) = 1 - e^{-\theta(1-\varepsilon/2)} e^{\theta\sqrt{1-\varepsilon}}.$$

Un développement limité quand  $\varepsilon$  tend vers 0 nous donne :

$$d_H(\mathcal{P}(\theta(1 - \varepsilon)), \mathcal{P}(\theta)) = \varepsilon\sqrt{\frac{\theta}{8}} + o(\varepsilon).$$

Nous obtenons donc l'ordre de la fonction  $d_H$  quand  $t$  tend vers  $+\infty$  :

$$d_H(t) = \sqrt{\frac{\alpha}{8\rho}} e^{-\rho t} + o(e^{-\rho t}),$$

ce qui montre que la constante  $C_H$  vaut dans ce cas :

$$C_H = \sqrt{\frac{\alpha}{8\rho}}.$$

**Distance du  $\chi^2$ .** La distance du  $\chi^2$  entre deux lois de Poisson est donnée dans la proposition 3.1.17 :

$$d_{\chi^2}^2(\mathcal{P}(\theta), \mathcal{P}(\theta')) = \exp\left(\theta' - 2\theta + \frac{\theta^2}{\theta'}\right) - 1.$$

Pour calculer la distance du  $\chi^2$  entre la loi du processus à l'instant  $t$  et sa loi stationnaire, nous devons évaluer la distance, quand  $\varepsilon$  tend vers 0, entre la loi de Poisson de paramètre  $\theta(1 - \varepsilon)$  et la loi de Poisson de paramètre  $\theta$ , ce qui donne, d'après la formule ci-dessus :

$$d_{\chi^2}(\mathcal{P}(\theta(1 - \varepsilon)), \mathcal{P}(\theta)) = \sqrt{e^{\theta\varepsilon^2} - 1}.$$

Un développement limité quand  $\varepsilon$  tend vers 0 nous donne :

$$d_{\chi^2}(\mathcal{P}(\theta(1 - \varepsilon)), \mathcal{P}(\theta)) = \varepsilon\sqrt{\theta} + o(\varepsilon).$$

Nous obtenons donc l'ordre de la fonction  $d_{\chi^2}$  quand  $t$  tend vers  $+\infty$  :

$$d_{\chi^2}(t) = e^{-\rho t} \sqrt{\frac{\alpha}{\rho}} + o(e^{-\rho t}),$$

ce qui montre que la constante  $C_{\chi^2}$  vaut dans ce cas :

$$C_{\chi^2} = \sqrt{\frac{\alpha}{\rho}}.$$

**Distance de Kullback.** La distance de Kullback entre deux lois de Poisson est donnée dans la proposition 3.1.21 :

$$d_K^2(\mathcal{P}(\theta), \mathcal{P}(\theta')) = -\theta + \theta' + \theta \log\left(\frac{\theta}{\theta'}\right).$$

Pour calculer la distance de Kullback entre la loi du processus à l'instant  $t$  et sa loi stationnaire, nous devons évaluer la distance, quand  $\varepsilon$  tend vers 0, entre la loi de Poisson de paramètre  $\theta(1 - \varepsilon)$  et la loi de Poisson de paramètre  $\theta$ , ce qui donne, d'après la formule ci-dessus :

$$d_K^2(\mathcal{P}(\theta(1 - \varepsilon)), \mathcal{P}(\theta)) = \theta\varepsilon + \theta(1 - \varepsilon) \log(1 - \varepsilon).$$

Un développement limité quand  $\varepsilon$  tend vers 0 nous donne :

$$d_K(\mathcal{P}(\theta(1 - \varepsilon)), \mathcal{P}(\theta)) = \varepsilon\sqrt{\frac{\theta}{2}} + o(\varepsilon).$$

Nous obtenons donc l'ordre de la fonction  $d_K$  quand  $t$  tend vers  $+\infty$  :

$$d_K(t) = e^{-\rho t} \sqrt{\frac{\alpha}{2\rho}} + o(e^{-\rho t}),$$

ce qui montre que la constante  $C_K$  vaut dans ce cas :

$$C_K = \sqrt{\frac{\alpha}{2\rho}}.$$

Nous pouvons donc réécrire le théorème 3.2.5 dans le cas de la file  $M/M/\infty$ .

**Proposition 3.3.2** *Si  $X^{(n)}$  est un  $n$ -échantillon de processus de Markov représentant le nombre de clients dans la file d'attente  $M/M/\infty$ , nous avons les estimations suivantes :*

1. *Distance en variation totale :*

$$1 - 2 \exp\left(-\frac{C_{VT}^2}{2} e^{-2\rho u}\right) \leq \liminf_{n \rightarrow +\infty} d_{VT}^{(n)}(t_n + u),$$

et :

$$\limsup_{n \rightarrow +\infty} d_{VT}^{(n)}(t_n + u) \leq \left(1 - \exp\left(-\frac{\alpha}{4\rho} e^{-2\rho u}\right)\right)^{\frac{1}{2}},$$

où la constante  $C_{VT}$  est donnée par :

$$C_{VT} = \frac{1}{[\frac{\alpha}{\rho}]!} \left(\frac{\alpha}{\rho}\right)^{[\frac{\alpha}{\rho}] + 1} e^{-\frac{\alpha}{\rho}}.$$

2. *Distance de Hellinger* :

$$\lim_{n \rightarrow +\infty} d_H^{(n)}(t_n + u) = \left( 1 - \exp\left(-\frac{\alpha}{8\rho} e^{-2\rho u}\right) \right)^{\frac{1}{2}}.$$

3. *Distance du  $\chi^2$*  :

$$\lim_{n \rightarrow +\infty} d_{\chi^2}^{(n)}(t_n + u) = \left( \exp\left(\frac{\alpha}{\rho} e^{-2\rho u}\right) - 1 \right)^{\frac{1}{2}}.$$

4. *Distance de Kullback* :

$$\lim_{n \rightarrow +\infty} d_K^{(n)}(t_n + u) = \sqrt{\frac{\alpha}{2\rho}} e^{-\rho u}.$$

### 3.3.3 Le processus d'Ornstein-Uhlenbeck

Nous revenons dans cette partie au processus d'Ornstein-Uhlenbeck que nous avons considéré dans le chapitre 2. Rappelons les notations utilisées. Nous considérons le processus solution de l'équation différentielle stochastique suivante, où  $\rho$  et  $\sigma$  sont deux réels strictement positifs :

$$\begin{cases} dX(t) = -\rho X(t)dt + \sigma\sqrt{2\rho} dB(t) \\ X(0) = x_0 \end{cases}.$$

Le processus  $X$  a pour loi, à l'instant  $t$ , la loi normale de moyenne  $x_0 e^{-\rho t}$  et de variance  $\sigma^2(1 - e^{-2\rho t})$ . De plus, ce processus converge, quand  $t$  tend vers  $+\infty$ , vers sa loi stationnaire qui est la loi normale centrée et de variance  $\sigma^2$ .

Nous pouvons donc évaluer les différentes distances entre la loi de  $X$  à l'instant  $t$  et sa loi stationnaire, et déterminer les différentes constantes  $C_d$  associées. Le cas de la distance en variation totale figurait déjà dans le chapitre 2. Nous le redonnons ici car le théorème 3.2.5 ne fournit pas tout-à-fait les mêmes bornes que celles obtenues dans la proposition 2.2.1 du chapitre 2.

**Distance en variation totale.** Nous avons déjà vu dans la proposition 2.1.2 du chapitre 2 l'ordre de grandeur de la distance en variation totale entre la loi du processus à l'instant  $t$  et sa loi stationnaire :

$$d_{VT}(t) = \frac{|x_0|}{\sigma\sqrt{2\pi}} e^{-\rho t} + o(e^{-2\rho t}),$$

ce qui montre que la constante  $C_{VT}$  vaut :

$$C_{VT} = \frac{|x_0|}{\sigma\sqrt{2\pi}}.$$

**Distance de Hellinger.** Nous avons déjà vu, dans la proposition 3.1.12, le calcul de la distance de Hellinger entre deux lois gaussiennes quelconques :

$$d_H^2(\mathcal{N}(m_1, v_1^2), \mathcal{N}(m_2, v_2^2)) = 1 - \sqrt{\frac{2v_1 v_2}{v_1^2 + v_2^2}} \exp\left(-\frac{(m_1 - m_2)^2}{4(v_1^2 + v_2^2)}\right).$$

Ici nous devons évaluer, quand  $\varepsilon$  tend vers 0, la distance entre la loi  $\mathcal{N}(x_0\varepsilon, \sigma^2(1 - \varepsilon^2))$  et la loi  $\mathcal{N}(0, \sigma^2)$ . Par la formule ci-dessus, nous obtenons :

$$d_H^2(\mathcal{N}(x_0\varepsilon, \sigma^2(1 - \varepsilon^2)), \mathcal{N}(0, \sigma^2)) = 1 - (1 - \varepsilon^2)^{\frac{1}{4}} \left(1 - \frac{\varepsilon^2}{2}\right)^{-\frac{1}{2}} \exp\left(-\frac{x_0^2\varepsilon^2}{4\sigma^2(2 - \varepsilon^2)}\right).$$

Un développement limité quand  $\varepsilon$  tend vers 0 nous donne :

$$d_H^2(\mathcal{N}(x_0\varepsilon, \sigma^2(1 - \varepsilon^2)), \mathcal{N}(0, \sigma^2)) = \frac{x_0^2\varepsilon^2}{8\sigma^2} + o(\varepsilon^2).$$

Nous obtenons ainsi l'ordre de la fonction  $d_H$  quand  $t$  tend vers  $+\infty$  :

$$d_H(t) = \frac{|x_0|}{2\sigma\sqrt{2}} e^{-\rho t} + o(e^{-\rho t}),$$

ce qui montre que, dans ce cas, la constante  $C_H$  vaut :

$$C_H = \frac{|x_0|}{2\sigma\sqrt{2}}.$$

**Distance du  $\chi^2$ .** De même, nous avons déjà calculé, dans la proposition 3.1.17, la distance du  $\chi^2$  entre deux lois normales quelconques :

$$d_{\chi^2}^2(\mathcal{N}(m_1, v_1^2), \mathcal{N}(m_2, v_2^2)) = \frac{v_2^2}{v_1\sqrt{2v_2^2 - v_1^2}} \exp\left(\frac{(m_1 - m_2)^2}{2v_2^2 - v_1^2}\right) - 1.$$

Ici nous devons évaluer, quand  $\varepsilon$  tend vers 0, la distance entre la loi  $\mathcal{N}(x_0\varepsilon, \sigma^2(1 - \varepsilon^2))$  et la loi  $\mathcal{N}(0, \sigma^2)$ . Par la formule ci-dessus, nous obtenons :

$$d_{\chi^2}^2(\mathcal{N}(x_0\varepsilon, \sigma^2(1 - \varepsilon^2)), \mathcal{N}(0, \sigma^2)) = \frac{\sigma}{\sqrt{1 - \varepsilon^2}} \frac{1}{\sigma\sqrt{1 + \varepsilon^2}} \exp\left(\frac{x_0^2\varepsilon^2}{\sigma^2(1 + \varepsilon^2)}\right) - 1.$$

Un développement limité quand  $\varepsilon$  tend vers 0 nous donne :

$$d_{\chi^2}^2(\mathcal{N}(x_0\varepsilon, \sigma^2(1 - \varepsilon^2)), \mathcal{N}(0, \sigma^2)) = \frac{x_0^2\varepsilon^2}{\sigma^2} + o(\varepsilon^2).$$

Nous obtenons ainsi l'ordre de la fonction  $d_{\chi^2}$  quand  $t$  tend vers  $+\infty$  :

$$d_{\chi^2}(t) = \frac{|x_0|}{\sigma} e^{-\rho t} + o(e^{-\rho t}),$$

ce qui montre que, dans ce cas, la constante  $C_{\chi^2}$  vaut :

$$C_{\chi^2} = \frac{|x_0|}{\sigma}.$$

**Distance de Kullback.** Nous avons déjà calculé, dans la proposition 3.1.21, la distance de Kullback entre deux lois gaussiennes quelconques :

$$d_K^2(\mathcal{N}(m_1, v_1^2), \mathcal{N}(m_2, v_2^2)) = \log\left(\frac{v_2}{v_1}\right) - \frac{1}{2} + \frac{v_1^2}{2v_2^2} + \frac{(m_1 - m_2)^2}{2v_2^2}.$$

Ici nous devons évaluer, quand  $\varepsilon$  tend vers 0, la distance entre la loi  $\mathcal{N}(x_0\varepsilon, \sigma^2(1 - \varepsilon^2))$  et la loi  $\mathcal{N}(0, \sigma^2)$ . Par la formule ci-dessus, nous obtenons :

$$d_K^2(\mathcal{N}(x_0\varepsilon, \sigma^2(1 - \varepsilon^2)), \mathcal{N}(0, \sigma^2)) = \log\left(\frac{1}{\sqrt{1 - \varepsilon^2}}\right) - \frac{1}{2} + \frac{1 - \varepsilon^2}{2} + \frac{x_0^2\varepsilon^2}{2\sigma^2}.$$

Un développement limité quand  $\varepsilon$  tend vers 0 nous donne :

$$d_K^2(\mathcal{N}(x_0\varepsilon, \sigma^2(1 - \varepsilon^2)), \mathcal{N}(0, \sigma^2)) = \frac{x_0^2\varepsilon^2}{2\sigma^2} + o(\varepsilon^2).$$

Nous obtenons ainsi l'ordre de la fonction  $d_K$  quand  $t$  tend vers  $+\infty$  :

$$d_K(t) = \frac{|x_0|}{\sigma\sqrt{2}} e^{-\rho t} + o(e^{-\rho t}),$$

ce qui montre que, dans ce cas, la constante  $C_K$  vaut :

$$C_K = \frac{|x_0|}{\sigma\sqrt{2}}.$$

On peut donc réécrire le théorème 3.2.5 dans le cas du processus d'Ornstein-Uhlenbeck.

**Proposition 3.3.3** Soit  $X^{(n)}$  est un  $n$ -échantillon de processus d'Ornstein-Uhlenbeck solution de l'équation différentielle stochastique :

$$\begin{cases} dX(t) = -\rho X(t)dt + \sigma\sqrt{2\rho} dB(t) \\ X(0) = x_0 \end{cases}.$$

Alors les estimations à gauche et à droite de l'instant de cutoff sont les suivantes :

1. Distance en variation totale :

$$1 - 2 \exp\left(-\frac{x_0^2}{4\pi\sigma^2} e^{-2\rho u}\right) \leq \liminf_{n \rightarrow +\infty} d_{VT}^{(n)}(t_n + u),$$

et :

$$\limsup_{n \rightarrow +\infty} d_{VT}^{(n)}(t_n + u) \leq \left(1 - \exp\left(-\frac{x_0^2}{4\sigma^2} e^{-2\rho u}\right)\right)^{\frac{1}{2}}.$$

2. Distance de Hellinger :

$$\lim_{n \rightarrow +\infty} d_H^{(n)}(t_n + u) = \left(1 - \exp\left(-\frac{x_0^2}{8\sigma^2} e^{-2\rho u}\right)\right)^{\frac{1}{2}}.$$



3. Distance du  $\chi^2$  :

$$\lim_{n \rightarrow +\infty} d_{\chi^2}^{(n)}(t_n + u) = \left( \exp \left( \frac{x_0^2}{\sigma^2} e^{-2\rho u} \right) - 1 \right)^{\frac{1}{2}} .$$

4. Distance de Kullback :

$$\lim_{n \rightarrow +\infty} d_K^{(n)}(t_n + u) = \frac{|x_0|}{\sigma\sqrt{2}} e^{-\rho u} .$$

Remarquons que dans l'encadrement de la distance en variation totale, le majorant est le même que dans la proposition 2.2.1 (nous avons utilisé la même démarche), mais la minoration est légèrement différente. Dans la proposition 2.2.1, la minoration est la suivante :

$$1 - \exp \left( -\frac{x_0^2}{8\sigma^2} e^{-2\rho u} \right) .$$

Elle est en fait toujours meilleure que celle de la proposition 3.3.3, au sens où, pour tous  $x_0$ ,  $\sigma$ ,  $\rho$  et  $u$  :

$$1 - 2 \exp \left( -\frac{x_0^2}{4\pi\sigma^2} e^{-2\rho u} \right) \leq 1 - \exp \left( -\frac{x_0^2}{8\sigma^2} e^{-2\rho u} \right) .$$

Ce n'est pas surprenant puisqu'au chapitre 2 nous avons exploité le calcul exact de la distance de Hellinger entre lois normales.

# Chapitre 4

## Temps d'atteinte

Dans ce chapitre, nous nous intéressons à la détection empirique du cutoff. Dans ce qui précède, nous avons réussi à déterminer l'expression de l'instant de cutoff. Cependant le coefficient  $\rho$  qui intervient dans cette expression est le taux de convergence du processus échantillonné, et il est inconnu en général. Il est donc naturel d'en chercher un estimateur, comme nous l'avons fait au chapitre 2 pour le processus d'Ornstein-Uhlenbeck. Nous suivons ici la même démarche, au sens où nous souhaitons construire l'estimateur comme temps d'atteinte d'un niveau par la moyenne empirique d'une fonction de l'échantillon. Dans un premier temps, nous allons donc exposer les propriétés asymptotiques du processus moyen ; nous montrons que, sous certaines propriétés du processus échantillonné, le processus moyen vérifie un théorème de la limite centrale, où le processus limite est un processus gaussien. Ensuite nous expliquerons comment utiliser ce résultat pour relier le temps d'atteinte d'un niveau par le processus moyen à un temps d'atteinte d'une barrière par le processus limite, avant d'étudier plus généralement les temps d'atteinte de barrières par des diffusions.

### 4.1 Théorème de la limite centrale pour des processus de Markov

Cette partie est consacrée au théorème de la limite centrale pour des processus de Markov à valeurs dans  $\mathbb{R}$ . Nous commençons par décrire l'espace de Skorohod et sa topologie, qui fournit le cadre adapté à la convergence en loi des processus. Il s'agit de rappels de notions classiques, que l'on peut lire dans le livre de P. Billingsley [9] par exemple. Le temps ici décrit  $\mathbb{R}^+$  ; dans ce cadre, on pourra se référer aux articles [39, 60] . Ensuite nous donnons le théorème de la limite centrale ; une référence très complète sur le sujet est le livre d'A. Araujo et E. Giné [5]. Sur l'ensemble de cette partie, le livre de W. Whitt [63] est aussi très détaillé.

#### 4.1.1 L'espace $D$ de Skorohod

Les processus stochastiques auxquels nous nous intéressons ont des trajectoires régulières, au sens où, en tout point, elles sont continues à droite et admettent une limite à gauche (càdlàg).

L'espace des fonctions càdlàg sur un intervalle  $I$  de  $\mathbb{R}$  et à valeurs réelles est noté  $D(I)$  et appelé espace de Skorohod sur  $I$ .

Commençons par donner la proposition suivante, sur le nombre de points de discontinuité d'une fonction càdlàg (cf. [26]).

**Proposition 4.1.1** *Soit  $f$  une fonction de  $D(\mathbb{R})$ . Alors l'ensemble des points de discontinuité de  $f$  est au plus dénombrable.*

**Démonstration.**

Si  $t$  est un réel, nous notons  $f(t_-)$  la limite de  $f$  à gauche de  $t$ . Pour  $n \in \mathbb{N}^*$ , définissons l'ensemble  $A_n$  par :

$$A_n = \left\{ t \in \mathbb{R}; |f(t) - f(t_-)| > \frac{1}{n} \right\}.$$

Remarquons alors que  $A_n$  n'a pas de point d'accumulation, car  $\lim_{s \rightarrow t_+} f(s)$  et  $\lim_{s \rightarrow t_-} f(s)$  existent pour tout  $t \in \mathbb{R}$ . Nous en déduisons que  $A_n$  est au plus dénombrable. Or l'ensemble des points de discontinuité de  $f$  est exactement  $\bigcup_{n=1}^{+\infty} A_n$ . Nous en concluons donc que l'ensemble des points de discontinuité de  $f$  est au plus dénombrable.  $\square$

Cette partie vise à décrire précisément les espaces  $D(\mathbb{R}^+)$  et  $D(\mathbb{R})$ . Avant d'étudier ces deux cas, nous avons besoin de décrire les espaces  $D(I)$ , où  $I$  est un intervalle fermé borné.

**Cas où  $I$  est un intervalle fermé borné.**

Dans le cas où  $I$  est un intervalle fermé borné, la distance qui vient à l'idée en premier pour métriser  $D(I)$  est la distance uniforme :

$$d_U(f, g) = \sup_{u \in I} |f(u) - g(u)|.$$

Malheureusement elle n'est pas bien adaptée à l'espace  $D(I)$ , car il faudrait que deux fonctions aient des discontinuités exactement au même endroit pour être proches. C'est pour pallier ce problème que Skorohod a introduit en 1956 une distance notée ici  $d_I$ . Le principe est de considérer que deux fonctions sont proches si elles le sont uniformément après avoir autorisé un petit changement de temps.

Plus précisément, soit  $\Lambda_I$  l'ensemble des fonctions de  $I$  dans lui-même strictement croissantes et continues. Soit  $id$  la fonction identité de  $I$ . La distance  $d_I$  entre deux fonctions  $f$  et  $g$  de  $D(I)$  est alors définie par :

$$d_I(f, g) = \inf_{\lambda \in \Lambda_I} \max\{d_U(f \circ \lambda, g), d_U(\lambda, id)\},$$

ce qui peut aussi s'écrire :

$$d_I(f, g) = \inf\{\varepsilon > 0; \exists \lambda \in \Lambda_I; \sup_{u \in I} |\lambda(u) - u| \leq \varepsilon \text{ et } \sup_{u \in I} |f(\lambda(u)) - g(u)| \leq \varepsilon\}.$$

**Proposition 4.1.2** *L'application  $d_I$  est une distance sur  $D(I)$ .*

La démonstration de cette proposition est classique. Nous ne donnons ci-dessous que la démonstration du fait que si la distance entre deux fonctions vaut 0, alors les deux fonctions sont égales.

**Démonstration.**

Soient  $f$  et  $g$  deux fonctions de  $D(I)$  telles que  $d_I(f, g) = 0$ . Montrons qu'alors  $f = g$ . Commençons par démontrer que, pour tout  $x$  de  $I$ , deux cas peuvent se produire :

$$g(x) = f(x) \quad \text{ou} \quad g(x) = f(x_-).$$

Pour cela, soit  $(\lambda_n)_{n \in \mathbb{N}^*}$  une suite de fonctions de  $\Lambda_I$  telle que :

$$\forall n \in \mathbb{N}^*, \sup_{u \in I} |\lambda_n(u) - u| \leq \frac{1}{n} \quad \text{et} \quad \sup_{u \in I} |f(\lambda_n(u)) - g(u)| \leq \frac{1}{n}.$$

Soit  $x$  un élément de  $I$ .

- Premier cas : la fonction  $f$  est continue en  $x$ , auquel cas on écrit :

$$|g(x) - f(x)| \leq |g(x) - f(\lambda_n(x))| + |f(\lambda_n(x)) - f(x)|,$$

donc :

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N}^*, \forall n \geq n_0, |g(x) - f(x)| \leq \frac{1}{n} + \varepsilon,$$

ce qui montre que  $f(x) = g(x)$ .

- Deuxième cas : le point  $x$  est un point de discontinuité de la fonction  $f$ . Notons alors  $A$  et  $B$  les ensembles suivants :

$$A = \{n \in \mathbb{N}^*; \lambda_n(x) \geq x\} \quad \text{et} \quad B = \{n \in \mathbb{N}^*; \lambda_n(x) < x\}.$$

Comme  $A$  et  $B$  forment une partition de  $\mathbb{N}^*$ , au moins l'un de ces deux ensembles est infini. Dans le cas où  $A$  est infini, nous pouvons écrire :

$$|g(x) - f(x)| \leq |g(x) - f(\lambda_n(x))| + |f(\lambda_n(x)) - f(x)|,$$

ce qui implique, par continuité à droite de  $f$ , que :

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N}^*, \forall n \geq n_0 \text{ et } n \in A, |g(x) - f(x)| \leq \frac{1}{n} + \varepsilon,$$

donc que  $f(x) = g(x)$ .

Dans le cas où  $B$  est infini, nous écrivons :

$$|g(x) - f(x_-)| \leq |g(x) - f(\lambda_n(x))| + |f(\lambda_n(x)) - f(x_-)|,$$

ce qui implique, par définition de la limite à gauche de  $f$  en  $x$ , que :

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N}^*, \forall n \geq n_0 \text{ et } n \in B, |g(x) - f(x)| \leq \frac{1}{n} + \varepsilon,$$

donc que  $f(x_-) = g(x)$ .

Nous en déduisons donc que  $f$  et  $g$  sont égales partout, sauf aux points  $x$  décrits ci-dessus, où  $f(x_-) = g(x)$ . Mais la continuité de  $g$  à droite implique que  $g(x) = f(x)$ . Ainsi  $f = g$  sur  $I$ .  $\square$

### L'espace $D(\mathbb{R}^+)$ .

Pour définir la convergence dans cet espace, nous utilisons le fait que la restriction à l'intervalle  $[0 ; t]$  d'une fonction càdlàg sur  $\mathbb{R}^+$  est càdlàg sur  $[0 ; t]$ .

**Définition 4.1.3** Soit  $(f_n)_{n \in \mathbb{N}}$  une suite de fonctions de  $D(\mathbb{R}^+)$ , et soit  $f$  une fonction de  $D(\mathbb{R}^+)$ . On dit que la suite  $(f_n)_{n \in \mathbb{N}}$  converge vers la fonction  $f$  dans  $D(\mathbb{R}^+)$  si, pour tout  $t$  positif, point de continuité de  $f$ , la suite des restrictions des  $f_n$  à l'intervalle  $[0 ; t]$  converge vers la restriction de  $f$  à l'intervalle  $[0 ; t]$  dans l'espace  $D([0 ; t])$ .

Remarquons que l'on enlève les points de discontinuité de la fonction limite car on souhaite voir converger des suites de fonctions comme  $\mathbb{I}_{[t_n ; +\infty[}$ , où  $(t_n)_{n \in \mathbb{N}}$  est une suite de réels qui tend en décroissant vers  $t$  strictement positif. Avec la définition ci-dessus, cette suite de fonctions converge vers la fonction  $\mathbb{I}_{[t ; +\infty[}$ . Si l'on tenait compte aussi des points de discontinuité de la fonction limite, cette suite de fonctions ne convergerait pas dans  $D(\mathbb{R}^+)$ .

Ce mode de convergence peut s'écrire en termes de distances, en définissant sur  $D(\mathbb{R}^+)$  la distance suivante.

**Définition 4.1.4** Soient  $f$  et  $g$  deux fonctions de  $D(\mathbb{R}^+)$ . La distance  $d_{\mathbb{R}^+}$  entre  $f$  et  $g$  est définie par :

$$d_{\mathbb{R}^+}(f, g) = \int_0^{+\infty} e^{-t} \min \{d_{[0;t]}(f, g), 1\} dt,$$

où  $d_{[0;t]}(f, g)$  désigne la distance entre les restrictions de  $f$  et  $g$  à l'intervalle  $[0 ; t]$ .

Il faut noter que l'intégrale ci-dessus est bien définie puisque l'intégrande est une fonction mesurable positive majorée par la fonction  $t \mapsto e^{-t}$ , qui est intégrable sur  $[0 ; +\infty[$ . La mesurabilité découle du fait que la fonction  $t \mapsto d_{[0;t]}(f, g)$  est continue en tout point où  $f$  et  $g$  sont continues (cf. [63] p. 415).

Par ailleurs, l'expression de  $d_{\mathbb{R}^+}$  définit bien une distance, puisque, de façon générale, si  $\delta$  est une distance, alors  $\min\{\delta, 1\}$  est une distance.

**Proposition 4.1.5** *Les deux notions de convergence définies ci-dessus sont équivalentes, c'est-à-dire que, si  $(f_n)_{n \in \mathbb{N}}$  est une suite de fonctions de  $D(\mathbb{R}^+)$ , et  $f$  est une fonction de  $D(\mathbb{R}^+)$ , les deux assertions suivantes sont équivalentes :*

- (i) *la suite  $(f_n)_{n \in \mathbb{N}}$  converge vers la fonction  $f$  dans  $D(\mathbb{R}^+)$ ,*
- (ii)  $\lim_{n \rightarrow +\infty} d_{\mathbb{R}^+}(f_n, f) = 0$ .

**Démonstration.**

Elle s'inspire de l'article [62] de W. Whitt. Supposons tout d'abord que la suite de fonctions  $(f_n)_{n \in \mathbb{N}}$  converge vers la fonction  $f$  dans  $D(\mathbb{R}^+)$  au sens de la définition 4.1.3. Cela implique que, pour presque tout  $t$  de  $\mathbb{R}^+$ , la suite des restrictions des fonctions  $f_n$  à l'intervalle  $[0 ; t]$  converge vers la restriction de la fonction  $f$  à l'intervalle  $[0 ; t]$  dans  $D([0 ; t])$ , c'est-à-dire que la suite  $(d_{[0;t]}(f_n, f))_{n \in \mathbb{N}}$  tend vers 0. Par convergence dominée, on en déduit donc que la suite  $(d_{\mathbb{R}^+}(f_n, f))_{n \in \mathbb{N}}$  tend vers 0.

Réciproquement, supposons que la suite  $(d_{\mathbb{R}^+}(f_n, f))_{n \in \mathbb{N}}$  tend vers 0. On montre alors que, en tout point  $t$  où la fonction  $f$  est continue, la suite  $(d_{[0;t]}(f_n, f))_{n \in \mathbb{N}}$  tend vers 0. En effet, soit  $t_0$  un point de continuité de  $f$ . Comme la suite  $(\min\{d_{[0;t_0]}(f_n, f), 1\})_{n \in \mathbb{N}}$  est bornée, on peut en extraire une sous-suite convergente, indicée disons par  $\phi(n)$ . Supposons que cette sous-suite converge vers un réel  $\epsilon$  strictement positif. Alors, par continuité à droite de la fonction  $t \mapsto d_{[0;t]}(f_n, f)$  en  $t_0$  (cf. [63] p. 415), on a :

$$\exists \eta > 0, \quad \forall t \in [t_0 ; t_0 + \eta[, \quad \liminf_{\phi(n) \rightarrow +\infty} d_{[0;t]}(f_{\phi(n)}, f) > \eta .$$

Ceci implique que :

$$\liminf_{\phi(n) \rightarrow +\infty} d_{\mathbb{R}^+}(f_{\phi(n)}, f) > 0 ,$$

ce qui est en contradiction avec l'hypothèse selon laquelle la suite  $(d_{\mathbb{R}^+}(f_n, f))_{n \in \mathbb{N}}$  tend vers 0. On en déduit donc que la sous-suite  $(\min\{d_{[0;t_0]}(f_{\phi(n)}, f), 1\})_{n \in \mathbb{N}}$  tend vers 0, ce qui implique que la sous-suite  $(d_{[0;t_0]}(f_{\phi(n)}, f))_{n \in \mathbb{N}}$  tend vers 0. Comme toute sous-suite tend vers 0, la suite  $(d_{[0;t_0]}(f_n, f))_{n \in \mathbb{N}}$  tend vers 0. □

**L'espace  $D(\mathbb{R})$ .**

La procédure pour définir la convergence sur cet espace est exactement la même que pour l'espace  $D(\mathbb{R}^+)$ , sauf que les fonctions sont restreintes à des intervalles de la forme  $[-t ; t]$ . La définition précise de la convergence est la suivante.

**Définition 4.1.6** *Soit  $(f_n)_{n \in \mathbb{N}}$  une suite de fonctions de  $D(\mathbb{R})$ , et soit  $f$  une fonction de  $D(\mathbb{R})$ . On dit que la suite  $(f_n)_{n \in \mathbb{N}}$  converge vers la fonction  $f$  dans  $D(\mathbb{R})$  si, pour tout  $t$  positif tel que  $t$  et  $-t$  sont des points de continuité de  $f$ , la suite des restrictions des  $f_n$  à l'intervalle  $[-t ; t]$  converge vers la restriction de  $f$  à l'intervalle  $[-t ; t]$  dans l'espace  $D([-t ; t])$ .*

De même que dans  $D(\mathbb{R}^+)$ , on peut définir une distance sur  $D(\mathbb{R})$  qui métrise cette convergence.

**Définition 4.1.7** Soient  $f$  et  $g$  deux fonctions de  $D(\mathbb{R})$ . La distance  $d_{\mathbb{R}}$  entre  $f$  et  $g$  est définie par :

$$d_{\mathbb{R}}(f, g) = \int_0^{+\infty} e^{-t} \min \{d_{[-t;t]}(f, g), 1\} dt ,$$

où  $d_{[-t;t]}(f, g)$  désigne la distance entre les restrictions de  $f$  et  $g$  à l'intervalle  $[-t ; t]$ .

L'intégrale qui intervient dans cette définition est bien définie pour les mêmes raisons que dans le paragraphe précédent. De même, les deux notions de convergence sont équivalentes.

**Proposition 4.1.8** Les deux notions de convergence définies ci-dessus sont équivalentes, c'est-à-dire que, si  $(f_n)_{n \in \mathbb{N}}$  est une suite de fonctions de  $D(\mathbb{R})$ , et  $f$  est une fonction de  $D(\mathbb{R})$ , les deux assertions suivantes sont équivalentes :

- (i) la suite  $(f_n)_{n \in \mathbb{N}}$  converge vers la fonction  $f$  dans  $D(\mathbb{R})$ ,
- (ii)  $\lim_{n \rightarrow +\infty} d_{\mathbb{R}}(f_n, f) = 0$ .

**Démonstration.**

Elle est semblable à celle de la proposition 4.1.5. Supposons tout d'abord que la suite de fonctions  $(f_n)_{n \in \mathbb{N}}$  converge vers la fonction  $f$  dans  $D(\mathbb{R})$  au sens de la définition 4.1.7. Cela implique que, pour presque tout  $t$  de  $\mathbb{R}^+$ , la suite des restrictions des fonctions  $f_n$  à l'intervalle  $[-t ; t]$  converge vers la restriction de la fonction  $f$  à l'intervalle  $[-t ; t]$  dans  $D([-t ; t])$ , c'est-à-dire que la suite  $(d_{[-t;t]}(f_n, f))_{n \in \mathbb{N}}$  tend vers 0. Par convergence dominée, on en déduit donc que la suite  $(d_{\mathbb{R}}(f_n, f))_{n \in \mathbb{N}}$  tend vers 0.

Réciproquement, supposons que la suite  $(d_{\mathbb{R}}(f_n, f))_{n \in \mathbb{N}}$  tend vers 0. On montre alors que, en tout point  $t$  positif tel que  $t$  et  $-t$  sont des points de continuité de la fonction  $f$ , la suite  $(d_{[-t;t]}(f_n, f))_{n \in \mathbb{N}}$  tend vers 0. En effet, soit  $t_0$  positif tel que  $t_0$  et  $-t_0$  sont des points de continuité de  $f$ . Comme la suite  $(\min\{d_{[-t_0;t_0]}(f_n, f), 1\})_{n \in \mathbb{N}}$  est bornée, on peut en extraire une sous-suite convergente, indicée disons par  $\phi(n)$ . Supposons que cette sous-suite converge vers un réel  $\epsilon$  strictement positif. Alors, on peut en déduire que :

$$\exists \eta > 0, \quad \forall t \in [t_0 ; t_0 + \eta[, \quad \liminf_{\phi(n) \rightarrow +\infty} d_{[-t;t]}(f_{\phi(n)}, f) > \eta .$$

Ceci implique que :

$$\liminf_{\phi(n) \rightarrow +\infty} d_{\mathbb{R}}(f_{\phi(n)}, f) > 0 ,$$

ce qui est en contradiction avec l'hypothèse selon laquelle la suite  $(d_{\mathbb{R}}(f_n, f))_{n \in \mathbb{N}}$  tend vers 0. On en déduit donc que la sous-suite  $(\min\{d_{[-t_0;t_0]}(f_{\phi(n)}, f), 1\})_{n \in \mathbb{N}}$  tend vers 0, ce qui implique que la sous-suite  $(d_{[-t_0;t_0]}(f_{\phi(n)}, f))_{n \in \mathbb{N}}$  tend vers 0. Comme toute sous-suite tend vers 0, la suite  $(d_{[-t_0;t_0]}(f_n, f))_{n \in \mathbb{N}}$  tend vers 0.  $\square$

Nous avons maintenant les espaces adéquats pour étudier le théorème de la limite centrale pour des processus, ainsi que la topologie adaptée. Nous pouvons donc exposer ce théorème.

### 4.1.2 Le théorème de la limite centrale

Dans cette partie, nous nous intéressons au comportement de la moyenne arithmétique d'un  $n$ -échantillon de processus stochastiques, quand  $n$  tend vers l'infini. Nous sommes motivés par le principe de parallélisation des algorithmes, tel que nous l'avons expliqué au chapitre 1.

L'idée d'étudier la moyenne des processus plutôt que le processus lui-même paraît naturelle, surtout lorsque l'on s'intéresse au comportement asymptotique de celui-ci; cependant on ne la trouve pas fréquemment dans la littérature. À notre connaissance, un des premiers articles qui traitent de ce sujet est celui de M.G. Hahn [33] en 1978, qui considère des processus de Markov dont le temps varie continûment de 0 à 1. À la même époque, des travaux ont été effectués dans un cadre plus général que celui des processus de Markov, notamment par A. Araujo dans sa thèse [4], et par E. Giné [32]. Le livre d'A. Araujo et E. Giné de 1980 [5] est une référence très complète sur le sujet. Par la suite, W. Whitt a généralisé les résultats de M.G. Hahn à des processus dont le temps varie de 0 à  $+\infty$ . Nous commençons par exposer ces résultats, avant de les appliquer au problème qui nous intéresse.

La définition suivante est un point de vocabulaire, qui permet d'énoncer de façon concise le théorème de la limite centrale pour un processus stochastique.

**Définition 4.1.9** Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé, soit  $I$  un intervalle de  $\mathbb{R}$ , et soit  $X$  un processus càdlàg sur  $(\Omega, \mathcal{A}, P)$  tel que, pour tout  $t$  appartenant à  $I$ , la variable aléatoire réelle  $X(t)$  soit mesurable. On dit que  $X$  satisfait le théorème de la limite centrale s'il existe une variable aléatoire gaussienne  $Z$  à valeurs dans  $D(I)$  qui est la limite en loi fonctionnelle, dans  $D(I)$ , de la suite de variables aléatoires  $(Z_n)_{n \in \mathbb{N}^*}$ , où  $Z_n$  est définie par :

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - EX_i),$$

où  $X_1, X_2, \dots$  sont des copies indépendantes de  $X$  sur le même espace probabilisé. On appelle  $Z$  le processus gaussien limite de  $X$ .

Notons qu'une variable aléatoire gaussienne à valeurs dans  $D(I)$  est en fait un processus gaussien càdlàg au sens habituel.

Considérons maintenant un processus stochastique  $Y = \{Y(t); t \geq 0\}$  à valeurs réelles, puis  $(Y_i)_{i \in \mathbb{N}^*}$  une suite de copies de  $Y$  indépendantes. Remarquons que ce cadre général permet de traiter le cas où le processus  $Y$  est l'image d'un processus de Markov. Si  $X$  est un processus de Markov à valeurs dans un espace  $E$ , et si  $\varphi$  est une fonction définie sur  $E$  et à valeurs réelles, alors le processus  $\varphi(X)$  est un processus stochastique à valeurs réelles.

Pour énoncer le résultat de W. Whitt, nous supposons que le processus  $Y$  est à valeurs dans  $D(\mathbb{R}^+)$ , c'est-à-dire que ses trajectoires sont des applications càdlàg de  $\mathbb{R}^+$  dans  $\mathbb{R}$ .

Le résultat de [63] (généralisation du résultat de M.G. Hahn dans [33]) s'énonce alors ainsi :



**Théorème 4.1.10** Soit  $Y$  un processus stochastique de  $D(\mathbb{R}^+)$ . Si, pour tout  $T$  de  $]0 ; +\infty[$ , il existe deux fonctions  $f_1$  et  $f_2$  définies sur  $[0 ; T]$  à valeurs réelles, continues croissantes, un réel  $\alpha$  strictement supérieur à  $1/2$ , et un réel  $\beta$  strictement supérieur à  $1$  tels que :

$$E[(Y(u) - Y(s))^2] \leq (f_1(u) - f_1(s))^\alpha ,$$

et :

$$E[(Y(u) - Y(t))^2(Y(t) - Y(s))^2] \leq (f_2(u) - f_2(t))^\beta ,$$

pour  $0 \leq s \leq t \leq u \leq T$ , avec  $u - s \leq 1$ , alors  $Y$  satisfait le théorème de la limite centrale. Le processus limite  $Z$  est un processus gaussien centré, dont la fonction de covariance est celle de  $Y$ .

Si le processus échantillonné est un processus de Markov, le théorème 4.1.10 prend une forme plus simple, que nous allons donner ci-dessous.

Rappelons la définition du supremum essentiel d'une variable aléatoire réelle  $U$  :

$$\text{ess sup}(U) = \inf \{c \in \mathbb{R} ; P(U > c) = 0\} .$$

Le théorème 4.1.10 s'énonce alors de la façon suivante.

**Théorème 4.1.11** Soit  $X$  un processus de Markov de  $D(\mathbb{R}^+)$ . Si, pour tout  $T$  de  $]0 ; +\infty[$ , il existe une fonction  $f_1$  définie sur  $[0 ; T]$  à valeurs réelles, continue croissante, et un réel  $\alpha$  strictement supérieur à  $1/2$ , tels que :

$$\text{ess sup } E[(X(t) - X(s))^2 | X(s)] \leq (f_1(t) - f_1(s))^\alpha ,$$

pour  $0 \leq s \leq t \leq T$ , avec  $t - s \leq 1$ , alors  $Y$  satisfait le théorème de la limite centrale. Le processus limite  $Z$  est un processus gaussien centré, dont la fonction de covariance est celle de  $X$ .

Comme premier corollaire du théorème 4.1.10, M.G. Hahn a remarqué que les processus de Markov à espace d'états fini vérifient automatiquement les hypothèses.

**Corollaire 4.1.12** Si  $X$  est un processus de Markov admettant un nombre fini d'états et à taux de sauts bornés, alors les hypothèses du théorème 4.1.11 sont vérifiées avec pour  $f_1$  une fonction constante et  $\alpha$  valant  $1$ , de sorte que le processus  $X$  satisfait le théorème de la limite centrale.

Ce corollaire est utile dans la pratique, comme nous allons le voir dans les exemples de la partie suivante.

### 4.1.3 Exemples

#### Le processus binaire.

Le processus échantillonné est un processus de Markov à deux états : 0 et 1. Le taux de transition de 0 à 1 est  $\alpha$  et le taux de transition de 1 à 0 est  $\rho - \alpha$ . De la partie 3.3.1 du chapitre 3, nous pouvons déduire que la loi de  $X$  à l'instant  $t$  est une loi de Bernoulli de paramètre  $m(t)$ , où  $m$  est la fonction définie par :

$$m(t) = \frac{\alpha}{\rho} (1 - e^{-\rho t}) .$$

Et la loi stationnaire du processus est la loi de Bernoulli de paramètre  $\theta$ , avec :

$$\theta = \frac{\alpha}{\rho} .$$

Considérons maintenant une suite  $(X_i)_{i \in \mathbb{N}^*}$  de processus de Markov de même loi que  $X$  et indépendants. Pour  $n$  entier, soit  $M_n$  le processus moyen, défini de la façon suivante :

$$\forall t \geq 0, \quad M_n(t) = \frac{1}{n} \sum_{i=1}^n X_i(t) .$$

Le corollaire 4.1.12 affirme qu'il existe un processus gaussien  $Z$  centré et de même fonction de covariance que  $X$  tel que :

$$\sqrt{n}(M_n - m) \xrightarrow[n \rightarrow +\infty]{} Z \quad \text{dans} \quad D(\mathbb{R}^+) .$$

Ce résultat a déjà été utilisé, notamment par C. Paroissin dans sa thèse [43] en 2002. Il s'intéressait à démontrer un théorème de la limite centrale pour le temps d'atteinte d'un niveau par le processus  $M_n$ , et le théorème de la limite centrale ci-dessus constituait la première étape de son étude.

#### La file d'attente $M/M/\infty$ .

Nous considérons ici un processus de Markov dont l'espace d'états est  $\mathbb{N}$ , et dont le taux de transition de  $i$  à  $i + 1$  est  $\alpha$  pour tout  $i$ , et le taux de transition de  $i$  à  $i - 1$  est  $i\rho$  (cf. partie 3.3.2 du chapitre 3).

Nous avons vu que la fonction génératrice  $\Psi_t$  de  $X(t)$  est connue (cf. [51] p. 142) ; pour tout  $u$  de l'intervalle  $[0 ; 1]$ , la fonction  $\Psi_t$  est définie par :

$$\Psi_t(u) = (1 + (u - 1)e^{-\rho t})^{x_0} \exp\left(\frac{\alpha}{\rho} (1 - e^{-\rho t}) (u - 1)\right) ,$$

ce qui permet de déterminer l'espérance de la variable aléatoire  $X(t)$ , que l'on note  $m(t)$  :

$$m(t) = \frac{\alpha}{\rho} + \left(x_0 - \frac{\alpha}{\rho}\right) e^{-\rho t} .$$

Soient maintenant  $X_1, \dots, X_n$   $n$  copies indépendantes du processus  $X$ , qui partent toutes de l'entier  $x_0$ . Soit  $M_n$  leur moyenne empirique :

$$M_n(t) = \frac{1}{n} \sum_{i=1}^n X_i(t).$$

**Proposition 4.1.13** *Le processus  $X$  satisfait le théorème de la limite centrale, c'est-à-dire qu'il existe un processus gaussien  $Z$  tel que la suite de processus*

$$\left(\sqrt{n}(M_n - m)\right)_{n \in \mathbb{N}^*}$$

*converge, dans  $D(\mathbb{R}^+)$ , vers le processus  $Z$  quand  $n$  tend vers l'infini. De plus, si l'on note  $B$  le mouvement brownien standard et  $h$  la fonction définie sur  $\mathbb{R}^+$  par :*

$$h(t) = 2\alpha + \rho \left(x_0 - \frac{\alpha}{\rho}\right) e^{-\rho t},$$

*alors  $Z$  est défini pour tout  $t$  positif par :*

$$Z(t) = \int_0^t e^{-\rho(t-s)} \sqrt{h(s)} dB(s).$$

*Si  $x_0 = \alpha/\rho$ , alors la fonction  $m$  est constante égale à  $\alpha/\rho$ , et le processus limite  $Z$  est un processus d'Ornstein-Uhlenbeck partant de 0, de paramètres  $\rho$  et  $\sqrt{\frac{\alpha}{\rho}}$ .*

#### Démonstration.

Il suffit de remarquer que le processus  $nM_n$  est lui-même un processus qui représente le nombre de clients dans une file d'attente  $M/M/\infty$  dans laquelle le processus d'arrivée des clients est un processus de Poisson d'intensité  $n\alpha$ , et les durées de service sont inchangées. Le théorème de la limite centrale pour de telles files d'attente ([51] p. 155) permet de conclure.  $\square$

Notons que sur cet exemple, le théorème de la limite centrale (proposition 4.1.13) est vrai, alors que le théorème 4.1.11 dû à W. Whitt ne s'applique pas. En effet, pour  $t$  supérieur ou égal à  $s$ , l'espérance de  $(X(t) - X(s))^2$  sachant  $X(s)$  est calculable, et son supremum essentiel est infini. Cet exemple illustre donc la limite d'utilisation du théorème 4.1.11. Dans la partie suivante, nous allons voir qu'il s'applique néanmoins à une famille très générale de processus : les chaînes de Markov harmonisées.

#### 4.1.4 Le cas particulier des chaînes harmonisées

L'harmonisation est une technique qui permet de passer d'une chaîne de Markov à temps discret à un processus de Markov (à temps continu). Cette technique est décrite dans le livre [67] de B. Ycart (chapitre 5, p. 179 sq.). L'idée est de construire un processus de saut dont les sauts sont ceux de la chaîne de départ, mais espacés selon un processus de Poisson d'intensité constante, indépendant de la chaîne.

**Définition 4.1.14** Soit  $(\mathbf{X}_n)_{n \in \mathbb{N}}$  une chaîne de Markov d'espace d'états  $E$ . Soit  $\lambda$  un réel strictement positif, et soit  $(N_t)_{t \in \mathbb{R}^+}$  un processus de Poisson d'intensité  $\lambda$ , indépendant de  $(\mathbf{X}_n)_{n \in \mathbb{N}}$ . On appelle chaîne de Markov harmonisée le processus  $(X(t))_{t \in \mathbb{R}^+}$  défini par :

$$\forall t \in \mathbb{R}^+, \quad X(t) = \mathbf{X}_{N_t}.$$

Sous une condition concernant les sauts de la chaîne  $(\mathbf{X}_n)_{n \in \mathbb{N}}$ , le processus  $X$  vérifie le théorème de la limite centrale.

**Proposition 4.1.15** Soit  $(\mathbf{X}_n)_{n \in \mathbb{N}}$  une chaîne de Markov d'espace d'états  $E$  inclus dans  $\mathbb{R}$ . S'il existe un réel  $\sigma$  strictement positif tel que la chaîne  $(\mathbf{X}_n)_{n \in \mathbb{N}}$  vérifie :

$$\forall n \in \mathbb{N}, \quad \text{ess sup } E[(\mathbf{X}_{n+1} - \mathbf{X}_n)^2 | \mathbf{X}_n] \leq \sigma^2,$$

alors la chaîne de Markov harmonisée  $(X(t))_{t \in \mathbb{R}^+}$  satisfait le théorème de la limite centrale.

### Démonstration.

Nous allons utiliser le théorème 4.1.11. Montrons que la chaîne de Markov harmonisée en vérifie les hypothèses. Soient  $t \geq 0$  et  $s \geq 0$  deux réels tels que  $0 \leq t - s \leq 1$ .

$$\begin{aligned} E[(X(t) - X(s))^2 | X(s)] &= E[(\mathbf{X}_{N_t} - \mathbf{X}_{N_s})^2 | \mathbf{X}_{N_s}] \\ &= \sum_{n=0}^{+\infty} E[(\mathbf{X}_{N_t} - \mathbf{X}_{N_s})^2 \mathbb{I}_{\{N_t - N_s = n\}} | \mathbf{X}_{N_s}] \\ &= \sum_{n=0}^{+\infty} P(N_t - N_s = n) E[(\mathbf{X}_{N_s+n} - \mathbf{X}_{N_s})^2 | \mathbf{X}_{N_s}] \\ &\quad \text{par indépendance de } N_t - N_s \text{ avec } (N_s, \mathbf{X}_{N_s}). \end{aligned}$$

Écrivons  $\mathbf{X}_{N_s+n} - \mathbf{X}_{N_s}$  sous la forme :

$$\mathbf{X}_{N_s+n} - \mathbf{X}_{N_s} = \sum_{i=0}^{n-1} (\mathbf{X}_{N_s+i+1} - \mathbf{X}_{N_s+i}).$$

L'inégalité de Cauchy-Schwarz donne alors :

$$\begin{aligned} E[(\mathbf{X}_{N_s+n} - \mathbf{X}_{N_s})^2 | \mathbf{X}_{N_s}] &\leq n \sum_{i=0}^{n-1} E[(\mathbf{X}_{N_s+i+1} - \mathbf{X}_{N_s+i})^2 | \mathbf{X}_{N_s}] \\ &= n \sum_{i=0}^{n-1} E[E[(\mathbf{X}_{N_s+i+1} - \mathbf{X}_{N_s+i})^2 | \mathbf{X}_{N_s}, \mathbf{X}_{N_s+i}] | \mathbf{X}_{N_s}]. \end{aligned}$$

car la tribu engendrée par  $\mathbf{X}_{N_s}$  est incluse dans la tribu engendrée par  $\mathbf{X}_{N_s}$  et  $\mathbf{X}_{N_s+i}$ . Par la propriété de Markov, nous obtenons ensuite :

$$E[(\mathbf{X}_{N_s+n} - \mathbf{X}_{N_s})^2 | \mathbf{X}_{N_s}] \leq n \sum_{i=0}^{n-1} E[E[(\mathbf{X}_{N_s+i+1} - \mathbf{X}_{N_s+i})^2 | \mathbf{X}_{N_s+i}] | \mathbf{X}_{N_s}].$$

Or par hypothèse nous avons :

$$\forall i \in \{0, \dots, n-1\}, \quad E[(\mathbf{X}_{N_s+i+1} - \mathbf{X}_{N_s+i})^2 | \mathbf{X}_{N_s+i}] \leq \sigma^2 .$$

Nous obtenons donc :

$$\text{ess sup } E[(\mathbf{X}_{N_s+n} - \mathbf{X}_{N_s})^2 | \mathbf{X}_{N_s}] \leq n^2 \sigma^2 .$$

D'où :

$$\begin{aligned} \text{ess sup } E[(X(t) - X(s))^2 | X(s)] &\leq \sigma^2 \sum_{n=0}^{+\infty} n^2 P(N_t - N_s = n) \\ &= \sigma^2 E[(N_t - N_s)^2] \\ &= \sigma^2 (\lambda(t-s) + \lambda^2(t-s)^2) \\ &\quad \text{car } N_t - N_s \text{ suit la loi de Poisson de paramètre } \lambda(t-s) , \\ &\leq \sigma^2 (\lambda + \lambda^2) (t-s) \quad \text{puisque } t-s \leq 1 . \end{aligned}$$

Le théorème 4.1.11 est donc vérifié, avec  $\alpha = 1$  et la fonction  $f_1$  définie par :

$$\forall t \in \mathbb{R}, \quad f_1(t) = \sigma^2(\lambda + \lambda^2)t .$$

□

### Application aux algorithmes stochastiques.

Le résultat de la proposition 4.1.15 permet de traiter le cas d'un certain nombre des algorithmes MCMC que nous avons décrits au chapitre 1. Prenons l'exemple de l'algorithme de Hastings-Metropolis. Il a pour but de simuler une valeur tirée suivant la loi de densité  $f$  sur un espace  $E$ . Prenons ici  $E = \mathbb{R}$  pour rester dans le cadre précédent. Nous avons vu au chapitre 1 que la chaîne de Markov sous-jacente à cet algorithme avait la forme suivante :

$$\forall n \in \mathbb{N}, \quad \mathbf{X}_{n+1} = \mathbf{X}_n + H_n \mathbb{I}_{A_n} ,$$

où :

- $(H_n)_{n \in \mathbb{N}}$  est une suite de variables aléatoires indépendantes et de même loi, de densité  $p$  et indépendantes de  $(\mathbf{X}_n)_{n \in \mathbb{N}}$ ,
- $(A_n)_{n \in \mathbb{N}}$  est la suite des événements "acceptation", dans le sens où  $A_n$  est l'événement sur lequel l'algorithme accepte le pas proposé  $\mathbf{X}_n + H_n$  :

$$A_n = \{\rho_n \geq 1\} \cup \{\rho_n < 1 \text{ et } \mathbf{U}_n < \rho_n\} ,$$

où  $(\mathbf{U}_n)_{n \in \mathbb{N}}$  est une suite de variables aléatoires indépendantes et de même loi uniforme sur  $[0 ; 1]$ , et  $\rho_n$  est défini par :

$$\rho_n = \frac{f(\mathbf{X}_n + H_n)}{f(\mathbf{X}_n)} .$$

La chaîne de Markov a donc une forme additive, et le terme additionné est le produit d'une variable aléatoire indépendante de la chaîne par une indicatrice. Dans ce cas, la proposition suivante donne le théorème de la limite centrale pour la chaîne harmonisée.

**Proposition 4.1.16** Soit  $(\mathbf{X}_n)_{n \in \mathbb{N}}$  une chaîne de Markov définie par :

$$\forall n \in \mathbb{N}, \quad \mathbf{X}_{n+1} = \mathbf{X}_n + H_n I_n,$$

où  $(H_n)_{n \in \mathbb{N}}$  est une suite de variables aléatoires indépendantes et de même loi possédant un moment d'ordre 2 et indépendantes de  $(\mathbf{X}_n)_{n \in \mathbb{N}}$ , et  $(I_n)_{n \in \mathbb{N}}$  est une suite de variables aléatoires bornées presque sûrement. Alors la chaîne de Markov harmonisée de  $(\mathbf{X}_n)_{n \in \mathbb{N}}$  satisfait le théorème de la limite centrale.

**Démonstration.**

Soit  $M$  un réel positif tel que :  $|I_n| \leq M$  p.s. On a alors :

$$|\mathbf{X}_{n+1} - \mathbf{X}_n|^2 \leq M^2 H_n^2 \quad \text{p.s.},$$

et par indépendance de  $H_n$  avec  $\mathbf{X}_n$ , nous obtenons :

$$E[|\mathbf{X}_{n+1} - \mathbf{X}_n|^2 | \mathbf{X}_n] \leq M^2 E[H_n^2] \quad \text{p.s.}$$

Les variables aléatoires  $H_n$  sont toutes de même loi, donc  $E[H_n^2]$  ne dépend pas de  $n$ , et ainsi la chaîne  $(\mathbf{X}_n)_{n \in \mathbb{N}}$  vérifie les hypothèses de la proposition 4.1.15.  $\square$

Dans le cas de l'algorithme de Hastings-Metropolis, il suffit donc que les variables aléatoires  $H_n$  aient un moment d'ordre 2 pour que la chaîne harmonisée satisfasse le théorème de la limite centrale.

Bien que cette forme additive de chaîne de Markov paraisse très particulière, elle n'en reste pas moins un cadre rencontré très souvent en pratique.

À ce stade, nous disposons donc d'un théorème de la limite centrale pour les processus de Markov, qui permet de rapporter l'étude asymptotique du processus moyen d'un échantillon à celle d'un processus gaussien. Comme dans le chapitre 2, nous souhaitons maintenant étudier le temps d'atteinte d'un niveau fixé par le processus moyen de l'échantillon. Dans la partie suivante, nous allons voir comment relier ce temps d'atteinte au temps d'atteinte d'une barrière par le processus gaussien limite, puis nous étudierons plus généralement les temps d'atteinte de barrières par des processus gaussiens.

## 4.2 Temps d'atteinte

Dans le chapitre 2 sur le processus d'Ornstein-Uhlenbeck, nous avons vu que le temps d'atteinte d'un niveau par la moyenne empirique de l'échantillon fournissait une bonne approximation

de l'instant de cutoff, ainsi qu'un estimateur du paramètre inconnu  $\rho$  correspondant au taux de la convergence du processus échantillonné. Dans le cas général des processus exponentiellement convergents, nous souhaitons entreprendre la même démarche. Le théorème de la limite centrale tel que nous l'avons explicité dans la partie précédente permet d'étudier le comportement de la moyenne empirique de l'échantillon.

Soit  $X$  un processus de Markov à valeurs dans un espace  $E$  qui converge vers sa mesure stationnaire  $\nu$ . Soit  $\varphi$  une fonction définie sur  $E$  et à valeurs réelles. Notons  $m(t)$  l'espérance de  $\varphi(X(t))$ . Soit  $(X_1, \dots, X_n)$  un  $n$ -échantillon de ce processus. La moyenne empirique de l'image de l'échantillon par  $\varphi$  est définie par :

$$M_n(t) = \frac{1}{n} \sum_{i=1}^n \varphi(X_i(t)) .$$

Nous souhaitons étudier le temps d'atteinte par ce processus d'un certain niveau fixé. Soit  $a$  ce niveau ( $a \in \mathbb{R}$ ). Nous notons  $T_a^n$  le temps d'atteinte de  $a$  par le processus  $M_n$  :

$$T_a^n = \inf\{t \geq 0 ; M_n(t) = a\} .$$

Nous avons vu au chapitre 2 que le niveau  $a$  à atteindre était l'espérance de la loi stationnaire :

$$a = \int_E \varphi d\nu = \nu\varphi .$$

Dans la partie 4.1 nous avons vu dans le théorème 4.1.10 que, sous certaines conditions, il existait un processus gaussien centré  $Z$  de même fonction de covariance que les  $\varphi(X_i)$  tel que la convergence suivante ait lieu dans  $D(\mathbb{R}^+)$  :

$$\sqrt{n}(M_n(t) - m(t)) \xrightarrow[t \rightarrow +\infty]{} Z(t) .$$

Comme au chapitre 2, nous aimerions démontrer que  $T_a^n$  (éventuellement renormalisé) converge en loi quand  $n$  tend vers l'infini, et, si possible, identifier sa limite. Pour cela, nous sommes tentés d'écrire :

$$\begin{aligned} T_a^n &= \inf\{t \geq 0 ; M_n(t) = a\} \\ &= \inf\{t \geq 0 ; \sqrt{n}(M_n(t) - m(t)) = \sqrt{n}(a - m(t))\} . \end{aligned}$$

Si l'on définit le temps d'atteinte  $\tau_a^n$  par :

$$\tau_a^n = \inf\{t \geq 0 ; Z(t) = \sqrt{n}(a - m(t))\} ,$$

on a envie de dire que  $T_a^n$  et  $\tau_a^n$  ont le même comportement quand  $n$  tend vers l'infini. Ensuite, comme le processus  $Z$  est gaussien, il est possible d'utiliser des résultats déjà existants pour avoir, si ce n'est le comportement exact de  $\tau_a^n$  quand  $n$  tend vers l'infini, au moins une approximation. Nous avons donc deux étapes à franchir :

- montrer que  $T_a^n$  et  $\tau_a^n$  ont le même comportement asymptotique ;
- déterminer le comportement asymptotique de  $\tau_a^n$ .

À ce jour, nous n'avons pas réussi à démontrer la première étape. Pourtant, cette question semble se rapprocher d'un problème semblable, résolu par C. Paroissin et B. Ycart dans l'article [44]. En effet, ils s'intéressent aussi dans cet article au temps d'atteinte d'un niveau fixé par le processus moyen ; toute la différence réside dans le niveau choisi. Nous souhaitons prendre comme niveau l'espérance de la loi stationnaire, qui est aussi la limite de la fonction  $m$  en  $+\infty$ , alors que dans l'article [44] le niveau est un réel atteint par  $m$  en un temps fini.

Quant au second point, nous avons une réponse partielle, qui nous semble être un démarrage intéressant au vu des résultats satisfaisants qu'il apporte dans plusieurs cas particuliers.

Il s'agit d'utiliser un théorème démontré par J. Durbin en 1985 ([25]), qui fournit une approximation de la densité du temps d'atteinte par un processus gaussien d'un niveau fixé. Nous énonçons ce théorème avant de voir comment l'utiliser.

**Théorème 4.2.1** *Soit  $(Y_t)_{t \geq 0}$  un processus gaussien continu centré de fonction de covariance  $K$ . Soit  $p$  la densité du temps de premier passage de la frontière  $\phi$  ( $\phi$  est une fonction de  $t$ ) "venant d'en bas". Les hypothèses sont les suivantes :*

- (H1) pour tout  $t \geq 0$ , la fonction frontière  $\phi$  est continue en  $t$  et dérivable à gauche en  $t$  ;
- (H2) la fonction de covariance  $K$  est définie positive et ses dérivées partielles premières sont continues sur l'ensemble  $\{(s, t) \in \mathbb{R}^+ ; 0 \leq s \leq t\}$ , avec les dérivées à gauche ou à droite adéquates en  $s = 0$  et  $s = t$  ;
- (H3) la variance de l'incrément  $Y_t - Y_s$  satisfait la condition :

$$\lim_{s \nearrow t} \frac{\text{Var}(Y_t - Y_s)}{t - s} = \lambda_t ,$$

avec  $0 < \lambda_t < \infty$ , ce qui est équivalent à :

$$\lim_{s \nearrow t} \left[ \frac{\partial K(s, t)}{\partial s} - \frac{\partial K(s, t)}{\partial t} \right] = \lambda_t ,$$

avec  $0 < \lambda_t < \infty$ .

Alors la fonction  $p$  est donnée par :

$$\forall t \geq 0, \quad p(t) = b(t)f(t) ,$$

où :

$$f(t) = \frac{1}{\sqrt{2\pi K(t, t)}} \exp\left(-\frac{\phi^2(t)}{2K(t, t)}\right) ,$$

et :

$$b(t) = \lim_{s \nearrow t} \frac{1}{t - s} E [I(s, Y)(\phi(s) - Y(s)) | Y(t) = \phi(t)] ,$$



avec :

$$I(s, Y) = \begin{cases} 1 & \text{si la trajectoire ne traverse pas la frontière avant } s, \\ 0 & \text{sinon.} \end{cases}$$

Cette formule pour  $p$  pose cependant un problème : celui du calcul de la fonction  $b$ , qui est difficile en général. Pour y remédier, J. Durbin ([25]) donne une approximation de  $p$ , obtenue en supprimant l'indicatrice dans la fonction  $b$ . C'est le théorème suivant.

**Théorème 4.2.2** *Les notations sont celles du théorème 4.2.1. Supposons que les hypothèses (H1), (H2) et (H3) sont vérifiées. Soit  $b_1$  la fonction définie pour tout  $t \geq 0$  par :*

$$\begin{aligned} b_1(t) &= \lim_{s \nearrow t} \frac{1}{t-s} E[\phi(s) - Y(s) | Y(t) = \phi(t)] \\ &= \frac{\phi(t)}{K(t, t)} \left. \frac{\partial K(s, t)}{\partial s} \right|_{s=t} - \phi'(t). \end{aligned}$$

Soit  $p_1$  la fonction définie pour tout  $t \geq 0$  par :

$$p_1(t) = b_1(t)f(t).$$

Alors  $p_1$  est une approximation de  $p$  qui est exacte à la limite quand la frontière s'éloigne à l'infini.

Dans son article, J. Durbin montre que cette approximation est en fait exacte dans l'exemple du premier temps de passage d'un mouvement brownien à une frontière affine. Soit en effet  $B$  un mouvement brownien standard. Soit  $T$  son premier temps de passage de la droite d'équation  $y = At + C$ , où  $C$  est positif (cette condition permet d'assurer que le mouvement brownien atteint la barrière en venant d'en bas) :

$$T = \inf\{t \geq 0 ; B(t) = At + C\}.$$

La fonction de covariance du mouvement brownien est donnée par :

$$\forall(s, t), \quad K(s, t) = \min(s, t).$$

Le calcul de  $b_1$  donne alors :

$$b_1(t) = \frac{C}{t},$$

et celui de  $f$  :

$$f(t) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(At + C)^2}{2t}\right).$$

Nous en déduisons donc que :

$$p_1(t) = \frac{C}{t\sqrt{2\pi t}} \exp\left(-\frac{(At + C)^2}{2t}\right),$$

ce qui correspond à la formule que l'on peut trouver dans la littérature ([28] Chap. VI p. 175). Remarquons que si  $C$  est négatif, il suffit d'écrire  $T$  sous la forme :

$$T = \inf\{t \geq 0 ; -B(t) = -At - C\} .$$

Comme  $-B$  est aussi un mouvement brownien standard, nous pouvons refaire le raisonnement qui précède. Pour  $C$  quelconque, nous obtenons donc :

$$p_1(t) = \frac{|C|}{t\sqrt{2\pi t}} \exp\left(-\frac{(At + C)^2}{2t}\right) . \quad (4.2.1)$$

Notons que cette formule permet aussi de retrouver la densité exacte du temps d'atteinte de zéro par le processus d'Ornstein-Uhlenbeck, que nous avons rappelée dans la proposition 2.3.1 du chapitre 2. En effet, le processus d'Ornstein-Uhlenbeck peut s'écrire comme un mouvement brownien changé de temps, grâce à la transformation de Doob (voir [2] par exemple). Soit  $X$  un processus d'Ornstein-Uhlenbeck partant de  $x_0$ , de paramètres  $\rho$  et  $\sigma$  (définition 2.1.1). Il existe alors un mouvement brownien standard  $B$  tel que :

$$\forall t \geq 0, \quad X(t) = e^{-\rho t} \left( x_0 + \sigma \sqrt{2\rho} B(\tau(t)) \right) ,$$

où  $\tau$  est la fonction définie par :

$$\tau(t) = \frac{1}{2\rho} (e^{2\rho t} - 1) .$$

Soit alors  $T_0$  le temps d'atteinte de zéro par le processus d'Ornstein-Uhlenbeck :

$$T_0 = \inf\{t \geq 0 ; X(t) = 0\} .$$

Nous pouvons alors en déduire que :

$$T_0 = \inf \left\{ t \geq 0 ; B(\tau(t)) = -\frac{x_0}{\sigma\sqrt{2\rho}} \right\} ,$$

ce qui signifie que :

$$\tau(T_0) = \inf \left\{ t \geq 0 ; B(t) = -\frac{x_0}{\sigma\sqrt{2\rho}} \right\} .$$

La formule (4.2.1) donne ainsi la densité  $p_{\tau(T_0)}$  de la variable aléatoire  $\tau(T_0)$  :

$$p_{\tau(T_0)}(t) = \frac{|x_0|}{2\sigma t \sqrt{\pi\rho t}} \exp\left(-\frac{x_0^2}{4\rho\sigma^2 t}\right) .$$

Ensuite, comme la fonction  $\tau$  est bijective, nous en déduisons la densité  $p_{T_0}$  de la variable aléatoire  $T_0$  :

$$p_{T_0}(t) = \frac{|x_0|}{2\sigma\sqrt{\pi\rho}} \left( \frac{\rho}{\sinh(\rho t)} \right)^{\frac{3}{2}} \exp\left(-\frac{x_0^2 e^{-\rho t}}{4\sigma^2 \sinh(\rho t)} + \frac{\rho t}{2}\right) ,$$

qui correspond à la formule de la proposition 2.3.1.

En conclusion, l'expression approchée de J. Durbin est plutôt séduisante. Cependant, dans le cas général, nous ne disposons pas d'estimation de l'erreur entre l'approximation fournie et la densité exacte. Ainsi la qualité de l'approximation n'est pas quantifiée.

Dans la section suivante, nous allons mettre cette approximation en oeuvre sur les exemples que nous avons déjà abordés dans la partie 4.1.3 : le processus binaire et la file  $M/M/\infty$ .

## 4.2.1 Exemples

### Le processus binaire.

Nous reprenons ici les notations utilisées dans la partie 4.1.3. Le processus binaire est le processus à deux états 0 et 1, avec le taux de transition de 0 à 1 égal à  $\alpha$ , et le taux de transition de 1 à 0 égal à  $\rho - \alpha$ . Nous avons montré que si  $(X_1, \dots, X_n)$  est un  $n$ -échantillon de processus binaires, alors la moyenne empirique  $M_n$  définie par :

$$M_n(t) = \frac{1}{n} \sum_{i=1}^n X_i(t),$$

vérifie le théorème de la limite centrale :

$$\sqrt{n}(M_n - m) \xrightarrow[n \rightarrow +\infty]{} Z \quad \text{dans } D(\mathbb{R}^+),$$

avec :

$$m(t) = \frac{\alpha}{\rho} (1 - e^{-\rho t}).$$

De plus, le processus  $Z$  est gaussien centré. Sa fonction de covariance est notée  $K$ , elle est égale à celle d'un des processus  $X_i$ , que l'on peut calculer.

**Proposition 4.2.3** *La fonction de covariance  $K$  du processus binaire est donnée par :*

$$\forall s \leq t, \quad K(s, t) = \frac{\alpha}{\rho^2} (\alpha + (\rho - \alpha)e^{\rho s}) (1 - e^{-\rho s}) e^{-\rho t}.$$

### Démonstration.

Rappelons que l'on a calculé dans la partie 3.3.1 le semi-groupe de transition  $P$  du processus :

$$\forall t \geq 0, \quad P(t) = \frac{1}{\rho} \begin{pmatrix} \rho - \alpha & \alpha \\ \rho - \alpha & \alpha \end{pmatrix} + \frac{e^{-\rho t}}{\rho} \begin{pmatrix} \alpha & -\alpha \\ -\rho + \alpha & \rho - \alpha \end{pmatrix}.$$

Nous pouvons donc écrire l'espérance du produit de  $X(s)$  par  $X(t)$  pour  $s \leq t$  (l'indice 0 rappelle que  $X(0) = 0$ ) :

$$\begin{aligned} E_o(X(s)X(t)) &= P_o(X(t) = 1, X(s) = 1) \\ &= P_o(X(t) = 1 | X(s) = 1) P_o(X(s) = 1) \\ &= P(X(t-s) = 1 | X(0) = 1) P_o(X(s) = 1) \\ &= \frac{\alpha}{\rho^2} (\alpha + (\rho - \alpha)e^{-\rho(t-s)}) (1 - e^{-\rho s}). \end{aligned}$$

Ensuite il suffit d'écrire :

$$K(s, t) = E_o(X(s)X(t)) - E_o(X(s))E_o(X(t)) ,$$

pour en déduire la formule annoncée.  $\square$

Rappelons que l'on s'intéresse à l'étude du temps d'atteinte  $\tau_a^n$  défini par :

$$\tau_a^n = \inf\{t \geq 0 ; Z(t) = \sqrt{n}(a - m(t))\} ,$$

où l'on voudrait que  $a$  soit l'espérance de la loi stationnaire du processus. Ici la loi stationnaire est la loi de Bernoulli de paramètre  $\frac{\alpha}{\rho}$  ; donc son espérance vaut  $\frac{\alpha}{\rho}$ . Désormais nous allons donc considérer :

$$a = \frac{\alpha}{\rho} .$$

Il reste finalement à vérifier les hypothèses du théorème 4.2.2 pour pouvoir l'appliquer. Les hypothèses (H1) et (H2) sont vérifiées. Pour (H3), un calcul montre que :

$$\lim_{s \nearrow t} \left[ \frac{\partial K(s, t)}{\partial s} - \frac{\partial K(s, t)}{\partial t} \right] = \frac{\alpha}{\rho} (2(\rho - \alpha) + (2\alpha - \rho)e^{-\rho t}) ,$$

et cette fonction de  $t$  est strictement positive quels que soient  $\alpha$  et  $\rho$ .

Nous pouvons maintenant calculer la densité approchée  $p_{\tau_{\alpha/\rho}^n, 1}(t)$  de la variable aléatoire  $\tau_{\alpha/\rho}^n$  par la formule du théorème 4.2.2 (le 1 en indice rappelle que  $p_{\tau_{\alpha/\rho}^n, 1}$  n'est pas la densité exacte de  $\tau_{\alpha/\rho}^n$ ). Nous obtenons :

$$p_{\tau_{\alpha/\rho}^n, 1}(t) = \rho \sqrt{\frac{n\alpha}{2\pi}} \left( \frac{2(\rho - \alpha) + (2\alpha - \rho)e^{-\rho t}}{(1 - e^{-\rho t})^{\frac{3}{2}} (\rho - \alpha + \alpha e^{-\rho t})^{\frac{3}{2}}} \right) \exp \left( -\rho t - \frac{n\alpha e^{-2\rho t}}{2(1 - e^{-\rho t})(\rho - \alpha + \alpha e^{-\rho t})} \right) .$$

Soit  $\tilde{\tau}_{\alpha/\rho}^n$  une variable aléatoire de loi de densité  $p_{\tau_{\alpha/\rho}^n, 1}$ . Nous en déduisons alors la convergence en loi de la variable aléatoire  $\tilde{\tau}_{\alpha/\rho}^n$  renormalisée :

**Proposition 4.2.4** Soit  $V_n$  la renormalisation suivante de la variable aléatoire  $\tau_{\alpha/\rho}^n$  :

$$V_n = \rho \left( \tilde{\tau}_{\alpha/\rho}^n - \frac{\log(n)}{2\rho} - \frac{1}{2\rho} \log \left( \frac{\alpha}{2(\rho - \alpha)} \right) \right) .$$

Quand  $n$  tend vers l'infini, la variable aléatoire  $V_n$  converge en loi vers la loi de densité  $g$  sur  $\mathbb{R}$ , avec :

$$g(t) = \frac{2}{\sqrt{\pi}} \exp(-t - e^{-2t}) .$$

**Démonstration.**

La densité de  $V_n$  au point  $t$  est donnée par :

$$v_n(t) = \frac{1}{\rho} p_{\tau_{\alpha/\rho}^n, 1} \left( \frac{t}{\rho} + \frac{\log(n)}{2\rho} + \frac{1}{2\rho} \log \left( \frac{\alpha}{2(\rho - \alpha)} \right) \right) .$$

Un calcul montre que, pour tout  $t$  :

$$\lim_{n \rightarrow +\infty} \frac{1}{\rho} p_{\tau_{\alpha/\rho}^n, 1} \left( \frac{t}{\rho} + \frac{\log(n)}{2\rho} + \frac{1}{2\rho} \log \left( \frac{\alpha}{2(\rho - \alpha)} \right) \right) = g(t) .$$

Ensuite, le lemme de Scheffé (lemme 2.3.3) permet de conclure que la variable aléatoire  $V_n$  converge en loi vers une variable aléatoire de densité  $g$ .  $\square$

Comme conséquence de la proposition 4.2.4, nous pouvons déduire la convergence en probabilité suivante pour  $\tilde{\tau}_{\alpha/\rho}^n$  :

$$\frac{\tilde{\tau}_{\alpha/\rho}^n}{\frac{\log(n)}{2\rho}} \xrightarrow[n \rightarrow \infty]{} 1 \quad \text{en probabilité .}$$

**La file d'attente  $M/M/\infty$ .**

Nous reprenons ici les notations utilisées dans la partie 4.1.3. Le processus qui correspond à la file  $M/M/\infty$  est le processus d'espace d'états  $\mathbb{N}$ , qui part d'un entier  $x_0$ , avec le taux de transition de  $i$  à  $i + 1$  égal à  $\alpha$ , et le taux de transition de  $i + 1$  à  $i$  égal à  $\rho(i + 1)$ . Nous avons montré dans la proposition 4.1.13 que si  $(X_1, \dots, X_n)$  est un  $n$ -échantillon de processus  $M/M/\infty$ , alors la moyenne empirique  $M_n$  définie par :

$$M_n(t) = \frac{1}{n} \sum_{i=1}^n X_i(t) ,$$

vérifie le théorème de la limite centrale :

$$\sqrt{n}(M_n - m) \xrightarrow[n \rightarrow +\infty]{} Z \quad \text{dans } D(\mathbb{R}^+) ,$$

avec :

$$m(t) = \frac{\alpha}{\rho} + \left( x_0 - \frac{\alpha}{\rho} \right) e^{-\rho t} .$$

De plus, le processus  $Z$  est gaussien centré. Sa fonction de covariance est notée  $K$ , elle est égale à celle d'un des processus  $X_i$ , que l'on peut calculer.

**Proposition 4.2.5** *La fonction de covariance  $K$  du processus associé à la file d'attente  $M/M/\infty$  est donnée par :*

$$\forall s \leq t, \quad K(s, t) = \frac{\alpha}{\rho} e^{-\rho(t-s)} + \left( x_0 - \frac{\alpha}{\rho} \right) e^{-\rho t} - x_0^2 e^{-\rho(t+s)} .$$

**Démonstration.**

Pour  $s \leq t$ , nous cherchons à évaluer l'espérance du produit  $X(s)X(t)$  (l'indice  $x_0$  rappelle que  $X(0) = x_0$ ) :

$$\begin{aligned} E_{x_0}(X(s)X(t)) &= E_{x_0}[E_{x_0}(X(s)X(t)|X(s))] \\ &= E_{x_0}[X(s) E_{x_0}(X(t)|X(s))] . \end{aligned}$$

Soit  $x \in \mathbb{N}$ . Par homogénéité, nous avons :

$$E_{x_0}(X(t)|X(s) = x) = E(X(t-s)|X(0) = x) .$$

Or d'après [51] p. 142, la fonction génératrice de  $X(t)$  sachant  $X(0) = x$  est donnée par :

$$\Psi_t(u) = (1 + (u-1)e^{-\mu t})^x \exp\left(\frac{\alpha}{\rho}(1 - e^{-\rho t})(u-1)\right) .$$

Comme  $\Psi'_t(1) = E(X(t)|X(0) = x)$ , nous en déduisons que :

$$E(X(t)|X(0) = x) = \rho(1 - e^{-\rho t}) + xe^{-\rho t} .$$

Nous obtenons alors :

$$E_{x_0}(X(t)|X(s)) = \rho(1 - e^{-\rho(t-s)}) + X(s)e^{-\rho(t-s)} .$$

Ensuite l'espérance du produit  $X(s)X(t)$  vaut :

$$\begin{aligned} E_{x_0}(X(s)X(t)) &= E_{x_0}\left[X(s)\frac{\alpha}{\rho}(1 - e^{-\rho(t-s)}) + X(s)^2e^{-\rho(t-s)}\right] \\ &= \frac{\alpha m(s)}{\rho}(1 - e^{-\rho(t-s)}) + e^{-\rho(t-s)}E_{x_0}(X(s)^2) . \end{aligned}$$

Le calcul de  $E_{x_0}(X(s)^2)$  se fait en utilisant la dérivée seconde de  $\Psi_s$  en 1. En effet nous avons :

$$\Psi''_s(1) = E_{x_0}(X(s)(X(s) - 1)) ,$$

d'où :

$$E_{x_0}(X(s)^2) = \Psi''_s(1) + m(s) .$$

Après calcul, nous obtenons :

$$E_{x_0}(X(s)^2) = \frac{\alpha}{\rho} \left( \frac{\alpha}{\rho} - 2x_0 \right) e^{-2\rho s} + \left( \frac{2\alpha}{\rho} + 1 \right) \left( x_0 - \frac{\alpha}{\rho} \right) e^{-\rho s} + \rho \left( 1 + \frac{\alpha}{\rho} \right) .$$

Nous obtenons ensuite l'expression de  $E_{x_0}(X(s)X(t))$ , puis celle de  $K$ , pour  $s \leq t$  :

$$\begin{aligned} K(s, t) &= E_{x_0}(X(s)X(t)) - E_{x_0}(X(s)) E_{x_0}(X(t)) \\ &= \frac{\alpha}{\rho} e^{-\rho(t-s)} + \left( x_0 - \frac{\alpha}{\rho} \right) e^{-\rho t} - x_0^2 e^{-\rho(t+s)} . \end{aligned}$$

□

Rappelons que l'on s'intéresse à l'étude du temps d'atteinte  $\tau_a^n$  défini par :

$$\tau_a^n = \inf\{t \geq 0 ; Z(t) = \sqrt{n}(a - m(t))\},$$

où l'on voudrait que  $a$  soit l'espérance de la loi stationnaire du processus. Ici la loi stationnaire est la loi de Poisson de paramètre  $\frac{\alpha}{\rho}$  ; donc son espérance vaut  $\frac{\alpha}{\rho}$ . Désormais nous allons donc considérer :

$$a = \frac{\alpha}{\rho}.$$

Il reste finalement à vérifier les hypothèses du théorème 4.2.2 pour pouvoir l'appliquer. Les hypothèses (H1) et (H2) sont vérifiées. Pour (H3), un calcul montre que :

$$\lim_{s \nearrow t} \left[ \frac{\partial K(s, t)}{\partial s} - \frac{\partial K(s, t)}{\partial t} \right] = 2\alpha + \rho \left( x_o - \frac{\alpha}{\rho} \right) e^{-\rho t},$$

et cette fonction de  $t$  est strictement positive quels que soient  $\alpha$  et  $\rho$ .

Nous pouvons maintenant calculer la densité approchée  $p_{\tau_{\alpha/\rho}^n, 1}(t)$  de la variable aléatoire  $\tau_{\alpha/\rho}^n$  par la formule du théorème 4.2.2. Nous obtenons :

$$\begin{aligned} p_{\tau_{\alpha/\rho}^n, 1}(t) &= -\sqrt{n} \left( x_o - \frac{\alpha}{\rho} \right) e^{-\rho t} \left( \frac{2\alpha + \rho \left( x_o - \frac{\alpha}{\rho} \right) e^{-\rho t}}{\frac{\alpha}{\rho} + \left( x_o - \frac{\alpha}{\rho} \right) e^{-\rho t}} \right) \\ &\quad \times \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\frac{\alpha}{\rho} + \left( x_o - \frac{\alpha}{\rho} \right) e^{-\rho t}}} \exp \left( -\frac{n \left( x_o - \frac{\alpha}{\rho} \right)^2 e^{-2\rho t}}{2 \left( \frac{\alpha}{\rho} + \left( x_o - \frac{\alpha}{\rho} \right) e^{-\rho t} \right)} \right). \end{aligned}$$

Soit  $\tilde{\tau}_{\alpha/\rho}^n$  une variable aléatoire de loi de densité  $p_{\tau_{\alpha/\rho}^n, 1}$ . Nous en déduisons alors la convergence en loi de la variable aléatoire  $\tilde{\tau}_{\alpha/\rho}^n$  renormalisée :

**Proposition 4.2.6** Soit  $W_n$  la renormalisation suivante de la variable aléatoire  $\tilde{\tau}_{\alpha/\rho}^n$  :

$$W_n = \rho \left( \tilde{\tau}_{\alpha/\rho}^n - \frac{\log(n)}{2\rho} + \frac{1}{2\rho} \log \left( \frac{\frac{2\alpha}{\rho}}{\left( x_o - \frac{\alpha}{\rho} \right)^2} \right) \right).$$

Quand  $n$  tend vers l'infini, la variable aléatoire  $W_n$  converge en loi vers la loi de densité  $g$  sur  $\mathbb{R}$ , avec :

$$g(t) = \frac{2}{\sqrt{\pi}} \exp(-t - e^{-2t}).$$

### Démonstration.

La densité de  $W_n$  au point  $t$  est donnée par :

$$w_n(t) = \frac{1}{\rho} p_{\tau_{\alpha/\rho}^n, 1} \left( \frac{t}{\rho} + \frac{\log(n)}{2\rho} - \frac{1}{2\rho} \log \left( \frac{\frac{2\alpha}{\rho}}{\left( x_o - \frac{\alpha}{\rho} \right)^2} \right) \right).$$

Un calcul simple montre que, pour tout  $t$  :

$$\lim_{n \rightarrow +\infty} \frac{1}{\rho} p_{\tau_{\alpha/\rho}, 1} \left( \frac{t}{\rho} + \frac{\log(n)}{2\rho} - \frac{1}{2\rho} \log \left( \frac{\frac{2\alpha}{\rho}}{(x_o - \frac{\alpha}{\rho})^2} \right) \right) = g(t).$$

Ensuite, le lemme de Scheffé (lemme 2.3.3) permet de conclure que la variable aléatoire  $W_n$  converge en loi vers une variable aléatoire de densité  $g$ .  $\square$

Notons que la loi limite du temps d'atteinte renormalisé est la même dans les trois exemples que nous avons traités. Ce n'est pas surprenant dans la mesure où les trois processus se ressemblent : ce sont des processus exponentiellement convergents (définition 3.2.1). Dans ces trois exemples, l'espérance du processus à l'instant  $t$  est une exponentielle décroissante, et les fonctions de covariance se ressemblent aussi. Il serait pourtant intuitif de penser que la covariance influe sur le temps d'atteinte. En effet, le processus gaussien  $Z$  devrait atteindre plus vite la barrière si sa fonction de covariance a une grande amplitude que si elle varie peu. C'est un de nos objectifs à venir d'étudier des processus d'autres types pour illustrer cette intuition.





# Conclusion et perspectives

Dans cette thèse, nous sommes partis d'un problème de détection de la convergence de méthodes MCMC : comment déterminer empiriquement l'instant où l'algorithme a atteint son régime stationnaire. Pour cela nous nous sommes intéressés à un échantillon d'algorithmes plutôt qu'à un seul. Ensuite nous avons fait le lien avec le phénomène de cutoff, qui se montre très utile dans la détermination d'instant d'arrêt d'algorithmes. Cependant, l'expression de l'instant de cutoff fait intervenir un paramètre inconnu qui mesure la vitesse de convergence de l'algorithme vers sa mesure stationnaire. Il est alors nécessaire de l'estimer. Dans cette perspective, nous avons introduit le temps d'atteinte d'un niveau par la moyenne arithmétique d'une fonction de l'échantillon. Nous avons alors étudié les propriétés asymptotiques de cette variable aléatoire pour en déduire le comportement de l'estimateur. Au fur et à mesure du cheminement de notre raisonnement, nous avons cependant laissé des questions en suspens, qui méritent d'être reprises.

Au chapitre 3, plusieurs généralisations sont possibles, et elles font l'objet de l'article [7] avec J. Barrera et B. Ycart, qui est terminé et qui sera soumis d'ici quelques semaines. Nous pouvons tout d'abord étudier le phénomène de cutoff pour des échantillons de processus non markoviens, pourvu qu'ils aient toujours la propriété de convergence exponentielle vers leur loi stationnaire. Nous pouvons aussi supposer que les taux des convergences de chacun des processus marginaux sont différents. Enfin, il est aussi possible de montrer que le phénomène de cutoff survient également dans le cas où la vitesse de convergence du processus échantillonné est une fonction produit d'un polynôme par une exponentielle décroissante.

Dans le chapitre 4, plusieurs voies sont ouvertes ; en plus de comprendre le passage du temps d'atteinte par la moyenne au temps d'atteinte par le processus limite, nous pourrions étudier plus précisément l'approximation de J. Durbin. En effet, nous avons vu, dans le cas particulier du processus d'Ornstein-Uhlenbeck, que le temps d'atteinte d'un niveau fixe par ce processus se ramène, par un changement de temps, au temps d'atteinte d'une barrière affine par le mouvement brownien standard. Cette idée devrait permettre d'étudier une large famille de temps d'atteinte, concernant les processus obtenus comme mouvements browniens changés de temps.

Enfin, il paraît incontournable de comparer le phénomène de cutoff avec d'autres moyens de détection de la convergence de processus. Le premier qui vient à l'esprit est le couplage du processus partant d'un réel fixé avec un processus de même noyau de transition, mais partant de la loi stationnaire. Il est alors classique (cf. [40]) de voir que la distance en variation totale entre les lois des deux processus est majorée par la fonction de survie du temps de couplage. Sur les

$n$ -échantillons de processus exponentiellement convergents de taux  $\rho$ , on peut donc montrer que le temps de couplage est en  $\log(n)/\rho$ , c'est-à-dire le double de l'instant de cutoff. Comprendre cette comparaison représente un objectif de recherches futures.

# Bibliographie

- [1] D.J. ALDOUS AND P. DIACONIS. Strong uniform times and finite random walks. *Adv. Appl. Math.*, 8: 69–97, 1987.
- [2] L. ALILI, P. PATIE, AND J.L. PEDERSEN. Representations of the first hitting time density of an Ornstein-Uhlenbeck process. *Stochastic Models*, 21(4), 2005. To appear.
- [3] C. ANÉ, S. BLACHÈRE, D. CHAFAÏ, P. FOUGÈRES, I. GENTIL, F. MALRIEU, C. ROBERTO, AND G. SCHEFFER. *Sur les inégalités de Sobolev logarithmiques*, volume 10 of *Panoramas et Synthèses*. Société Mathématique de France, 2000.
- [4] A. ARAUJO. *On the central limit theorem in Banach spaces*. PhD thesis, University of California, Berkeley, 1974. En partie dans *J. Multivariate Anal.* 8, 598–613.
- [5] A. ARAUJO AND E. GINÉ. *The central limit theorem for real and Banach valued random variables*. Wiley, 1980.
- [6] R. AZENCOTT. Simulated annealing. *Séminaire Bourbaki*, 697: 161–175, 1988.
- [7] J. BARRERA, B. LACHAUD, AND B. YCART. Cutoff for  $n$ -tuples of exponentially converging processes. *Submitted*, 2005.
- [8] D. BERTSIMAS AND J. TSITSIKLIS. Simulated annealing. *Statistical Science*, 8: 10–15, 1993.
- [9] P. BILLINGSLEY. *Convergence of probability measures*. Wiley, 1968.
- [10] PH. CAPÉRAÀ AND B. VAN CUTSEM. *Méthodes et modèles en statistique non paramétrique : exposé fondamental*. Presses de l'Université Laval, Dunod, 1988.
- [11] O. CATONI. Rates of convergence for sequential annealing: a large deviation approach. In R. Azencott, editor, *Simulated annealing : parallelization techniques*, pages 25–35. Wiley, New York, 1992.
- [12] O. CATONI. Rough large deviation estimates for simulated annealing: application to exponential schedules. *Ann. Probab.*, 20(3): 1109–1146, 1992.
- [13] O. CATONI. Simulated annealing algorithms and Markov chains with rare transitions. In *Séminaire de Probabilités, XXXIII*, volume 1709 of *Lecture Notes in Math.*, pages 69–119. Springer, 1999.
- [14] M.F. CHEN. *From Markov chains to non-equilibrium particle systems*. World Scientific, Singapore, 1992.
- [15] M.F. CHEN. Equivalence of exponential ergodicity and  $l^2$ -exponential convergence for Markov chains. *Stochastic Process. Appl.*, 87(2): 287–297, 2000.

- [16] T.M. COVER AND J.A. THOMAS. *Elements of information theory*. Wiley, 1991.
- [17] P. DEL MORAL AND N. BARTOLI. *Simulation et algorithmes stochastiques*. Cépaduès Éditions, 2001.
- [18] L. DEVROYE AND G. LUGOSI. *Combinatorial methods in density estimation*. Springer-Verlag, 2001.
- [19] P. DIACONIS. The cutoff phenomenon in finite Markov chains. *Proc. Natl. Acad. Sci. USA*, 93: 1659–1664, 1996.
- [20] P. DIACONIS, R. GRAHAM, AND J. MORRISON. Asymptotic analysis of a random walk on a hypercube with many dimensions. *Rand. Struct. Algorithms*, 1: 51–72, 1990.
- [21] P. DIACONIS AND L. SALOFF-COSTE. What do we know about the Metropolis algorithm? *J. Comp. Syst. Sci.*, 55(1): 20–36, 1998.
- [22] P. DIACONIS AND M. SHAHSHAHANI. Time to reach stationarity in the Bernoulli-Laplace diffusion model. *SIAM J. Math. Anal.*, 18: 208–218, 1987.
- [23] C. DOU AND M. HILDEBRAND. Enumeration and random walks on finite groups. *Ann. Probab.*, 24: 987–1000, 1996.
- [24] M. DUFLO. *Algorithmes stochastiques*. Number 23 in Mathématiques et Applications. Springer-Verlag, 1996.
- [25] J. DURBIN. The first-passage density of a continuous Gaussian process to a general boundary. *J. Appl. Probab.*, 22: 99–122, 1985.
- [26] S.N. ETHIER AND T.G. KURTZ. *Markov processes, characterization and convergence*. Probability and mathematical statistics. Wiley, 1986.
- [27] W. FELLER. *An introduction to probability theory and its applications*, volume I. Wiley, 1968.
- [28] W. FELLER. *An introduction to probability theory and its applications*, volume II. Wiley, 1971.
- [29] C. FRICKER, P. ROBERT, AND D. TIBI. On the rates of convergence of Erlang’s model. *J. Appl. Probab.*, 36(4): 1167–1184, 1999.
- [30] A.L. GIBBS AND F.E. SU. On choosing and bounding probability metrics. *Int. Statist. Review*, 70(3): 419–435, 2002.
- [31] W.R. GILKS, S. RICHARDSON, AND D.J. SPIEGELHALTER, editors. *Markov Chains Monte Carlo in practice*. Chapman & Hall, London, 1996.
- [32] E. GINÉ. On the central limit theorem for sample continuous processes. *Ann. Probab.*, 2: 619–641, 1974.
- [33] M.G. HAHN. Central limit theorems in  $D[0, 1]$ . *Z. Wahrsch. Verw. Gebiete*, 44: 89–101, 1978.
- [34] B. HAJEK. Cooling schedules for optimal annealing. *Math. Oper. Res.*, 13(2): 311–329, 1988.

- [35] W.K. HASTINGS. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57: 97–109, 1970.
- [36] S. KULLBACK. A lower bound for discrimination information in terms of variation. *IEEE Trans. Inform. Theory*, 13: 126–127, 1967.
- [37] B. LACHAUD. Cutoff and hitting times for a sample of Ornstein-Uhlenbeck processes and its average. *J. Appl. Probab.*, 42(5), 2005. To appear.
- [38] B. LEBLANC AND O. SCAILLET. Past dependent options on yields in the affine term structure. *Finance and Stochastics*, 2: 349–367, 1998.
- [39] T. LINDVALL. Weak convergence in the function space  $D[0, \infty)$ . *J. Appl. Probab.*, 10 :109–121, 1973.
- [40] T. LINDVALL. *Lectures on the coupling method*. Probability and mathematical statistics. Wiley, 1992.
- [41] S. MARTÍNEZ AND B. YCART. Decay rates and cutoff for convergence and hitting times of Markov chains with countably infinite state space. *Adv. Appl. Probab.*, 33(1): 188–205, 2001.
- [42] S.P. MEYN AND R. TWEEDIE. *Markov chains and stochastic stability*. Springer-Verlag, 1996.
- [43] C. PAROISSIN. *Résultats asymptotiques pour des grands systèmes réparables monotones*. PhD thesis, Univ. Paris 7, 2002.
- [44] C. PAROISSIN AND B. YCART. Central limit theorem for hitting times of functionals of Markov jump processes. *ESAIM Probab. Stat.*, 8: 66–75, 2004.
- [45] V.V. PETROV. *Sums of independent random variables*. Springer-Verlag, Berlin, 1975.
- [46] J. PITMAN AND M. YOR. Bessel processes and infinitely divisible laws. In *Stochastic integrals. Proc. Sympos., Univ. Durham, Durham, 1980*, volume 851 of *Lecture Notes in Math.*, pages 285–370, Springer, Berlin, 1981.
- [47] D. POLLARD. *A user's guide to measure theoretic probability*. Cambridge University Press, 2001.
- [48] J.G. PROPP AND D.B. WILSON. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures Algorithms*, 9(1–2): 223–252, 1996.
- [49] R.D. REISS. *Approximate distributions of order statistics*. Springer-Verlag, New York, 1989.
- [50] C. ROBERT. *Méthodes de Monte Carlo par chaînes de Markov*. Editions Economica, Paris, 1996.
- [51] PH. ROBERT. *Réseaux et files d'attente : méthodes probabilistes*. Number 35 in *Mathématiques et Applications*. Springer-Verlag, 2000.
- [52] G.O. ROBERTS AND J.S. ROSENTHAL. Geometric ergodicity and hybrid Markov chains. *Electron. Comm. Probab.*, 2(2): 13–25, 1997.

- [53] G.O. ROBERTS AND J.S. ROSENTHAL. General state space Markov chains and MCMC algorithms. *Probab. Surv.*, 1: 20–71, 2004.
- [54] G.O. ROBERTS AND R.L. TWEEDIE. Geometric  $L^2$  and  $L^1$  convergence are equivalent for reversible Markov chains. *J. Appl. Probab.*, 38A: 37–41, 2001.
- [55] S. ROUSSEL. Phénomène de cutoff pour certaines marches aléatoires sur le groupe symétrique. *Colloquium Math.*, 86: 111–135, 2000.
- [56] G. ROYER. *Une initiation aux inégalités de Sobolev logarithmiques*, volume 5 of *Cours spécialisés*. Société Mathématique de France, 1999.
- [57] L. SALOFF-COSTE. Lectures on finite Markov chains. In *Lectures on probability theory and statistics (Saint-Flour, 1996)*, volume 1665 of *Lecture Notes in Math.*, pages 301–413, Berlin, 1997. Springer.
- [58] L. SALOFF-COSTE. Random walks on finite groups. In H. Kesten, editor, *Probability on discrete structures*, volume 110, pages 263–346. Springer, Berlin, 2004.
- [59] F. SCARABOTTI. Time to reach stationarity in the Bernoulli-Laplace diffusion model with many urns. *Adv. Appl. Math.*, 18: 351–371, 1997.
- [60] C. STONE. Weak convergence of stochastic processes defined on semi-infinite time intervals. *Proc. Amer. Math. Soc.*, 14: 694–696, 1963.
- [61] G.E. UHLENBECK AND L.S. ORNSTEIN. On the theory of Brownian motion. *Phys. Rev.*, 36: 823–841, 1930.
- [62] W. WHITT. Some useful functions for functional limit theorems. *Math. Oper. Res.*, 5(1): 67–85, 1980.
- [63] W. WHITT. *An introduction to stochastic-process limits and their application to queues*. Springer, Berlin, 2002.
- [64] B. YCART. Cutoff for samples of Markov chains. *ESAIM Probab. Stat.*, 3: 89–107, 1999.
- [65] B. YCART. Stopping tests for Monte-Carlo Markov chain methods. *Meth. and Comp. in Appl. Probab.*, 2(1): 23–36, 2000.
- [66] B. YCART. Cutoff for Markov chains: some examples and applications. In E. Goles and S. Martínez, editors, *Complex Systems*, pages 261–300. Kluwer, Dordrecht, 2001.
- [67] B. YCART. *Modèles et algorithmes markoviens*. Number 39 in *Mathématiques et Applications*. Springer-Verlag, 2002.

## Détection de la convergence de processus de Markov

**Résumé :** Notre travail porte sur le phénomène de cutoff pour des  $n$ -échantillons de processus de Markov, dans le but de l'appliquer à la détection de la convergence d'algorithmes parallélisés. Dans un premier temps, le processus échantillonné est un processus d'Ornstein-Uhlenbeck. Nous mettons en évidence le phénomène de cutoff pour le  $n$ -échantillon, puis nous faisons le lien avec la convergence en loi du temps d'atteinte par le processus moyen d'un niveau fixé. Dans un second temps, nous traitons le cas général où le processus échantillonné converge à vitesse exponentielle vers sa loi stationnaire. Nous donnons des estimations précises des distances entre la loi du  $n$ -échantillon et sa loi stationnaire. Enfin, nous expliquons comment aborder les problèmes de temps d'atteinte liés au phénomène du cutoff.

**Mots-clefs :** cutoff, distances entre lois de probabilité, processus d'Ornstein-Uhlenbeck, convergence exponentielle, théorème de la limite centrale, temps d'atteinte.

**Classification AMS (2000) :** 60F05, 60J25, 60J60.

---

## Detection of convergence for Markov processes

**Abstract :** Our work deals with the cutoff phenomenon for  $n$ -samples of Markov processes, in order to apply it to the detection of convergence of parallelized algorithms. In a first part, the sampled process is an Ornstein-Uhlenbeck process. We point out the cutoff phenomenon for the  $n$ -sample, and we link it with the convergence in distribution of the hitting time of a fixed level by the average process. In a second part, the general case is studied, in which the sampled process is supposed to converge at exponential rate to its stationary distribution. Precise estimates are given for the distances between the distribution of the  $n$ -sample and its stationary distribution. Finally, we explain the way of dealing with some hitting time problems linked to the cutoff phenomenon.

**Key Words :** cutoff, distances between probability distributions, Ornstein-Uhlenbeck process, exponential convergence, central limit theorem, hitting time.

**AMS Classification (2000) :** 60F05, 60J25, 60J60.