



HAL
open science

Etude de l'estimation du Maximum de Vraisemblance dans des modèles Markoviens, Semi-Markoviens et Semi-Markoviens Cachés avec Applications

Samis Trevezas

► **To cite this version:**

Samis Trevezas. Etude de l'estimation du Maximum de Vraisemblance dans des modèles Markoviens, Semi-Markoviens et Semi-Markoviens Cachés avec Applications. Mathématiques [math]. Université de Technologie de Compiègne, 2008. Français. NNT : . tel-00472644

HAL Id: tel-00472644

<https://theses.hal.science/tel-00472644>

Submitted on 12 Apr 2010

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 DE TECHNOLOGIE DE COMPIÈGNE
LABORATOIRE DE MATHÉMATIQUES APPLIQUÉES DE
COMPIÈGNE

THÈSE

présentée pour l'obtention du grade de Docteur, spécialité
« Technologie de l'Information et des Systèmes »

par

Samis TREVEZAS

ÉTUDE DE L'ESTIMATION DU MAXIMUM DE VRAISEMBLANCE DANS DES MODÈLES MARKOVIENS, SEMI-MARKOVIENS ET SEMI-MARKOVIENS CACHÉS AVEC APPLICATIONS

Thèse soutenue le 5 Décembre 2008 devant le jury composé de :

M ^{me}	HUBER CATHERINE	Université Paris Descartes	(Présidente du jury et rapporteur)
M.	LIMNIOS NIKOLAOS	UTC	(Directeur)
M.	LEE MEI-LING TING	Maryland University	(Examinatrice)
M	MOYAL PASCAL	UTC	(Examineur)
M.	NIKOULINE MIKHAIL	Université Victor Segalen	(Rapporteur)
M.	NUEL GRÉGORY	Université Paris Descartes	(Rapporteur)

*À ma famille
Makis, Evgenia, Bianca
et
à ma femme Maha*

REMERCIEMENTS

Cette thèse aujourd'hui terminée aurait été moins riche sans quelques personnes que je souhaite remercier ici.

En premier lieu, je tiens à remercier mon directeur de thèse Nikolaos LIMNIOS pour son encadrement, ses précieux conseils, sa disponibilité et sa confiance en mon travail. Son expertise dans les processus semi-markoviens était une condition suffisante pour que je me sente rassuré aux moments difficiles. Il m'a permis assez vite une indépendance conditionnelle autour des thèmes de recherche assez modernes.

Je suis très sensible à l'honneur que m'a fait Mme Catherine HUBER en acceptant de rapporter et de présider mon jury de thèse. J'adresse toute ma gratitude aux MM. Mikhail NIKOULINE et Grégory NUEL, pour avoir accepté d'être rapporteurs de ce travail. Je remercie particulièrement Mme Mei-Ling Ting Lee qui est venue des États-Unis pour participer à mon jury de thèse. Je les remercie tous pour leurs remarques et discussions constructives.

Ma gratitude à Pascal MOYAL a deux faces. Il m'a fait l'honneur de participer à mon jury de thèse et il m'a beaucoup facilité la tâche d'enseignement pour que je puisse avoir la concentration nécessaire pour la rédaction de la thèse et la préparation de ma soutenance. Pour cette raison, je remercie également Vincent ROBIN.

Un grand merci à Ourania CHRYSSAFINOÛ pour m'avoir signalé cette proposition de thèse et pour avoir encadré mon stage de Master 2, que j'ai effectué en parallèle avec cette thèse. Elle m'a beaucoup encouragé durant les années de thèse ainsi qu'à Athènes durant mes études au département de Maths à l'Université d'Athènes. En plus d'Ourania, les personnes qui m'ont convaincu de faire le pas vers l'étranger et qui m'ont aidé à aimer davantage l'aléatoire sont Antonis ECONOMOÛ, Nikolaos PAPADATOS et Apostolis BURNETAS. Je les remercie tous pour tout ce qu'ils m'ont appris et pour m'avoir donné confiance en moi dans ce domaine de recherche. Je n'oublie pas Dimitrios DERIZIOTIS qui a été le premier à me convaincre de faire un doctorat en mathématiques, même si la théorie des nombres n'est pas pour l'instant un sujet des mes recherches.

Je remercie spécialement Sonia MALEFAKI avec qui nous avons développé un travail algorithmique sur les chaînes semi-markoviennes cachées qui constitue un chapitre de ma thèse. Ta détermination pour ce travail et ton amitié ont été importants durant la dernière année de ma thèse. Merci

aussi à Vassili GEORGIOU pour ton amitié et ta présence.

Un grand merci Vlad BARBU pour ton support durant la première année de ma thèse, pour votre amitié toi et Andra BARBU, pour les soirées et les moments que nous avons passé ensemble. Merci pour l'aide bibliographique et ta disponibilité aux moments où je me sentais seul dans la France. Costi MORARESCU, un bon ami avec qui nous avons partagé de très bon moments ensemble, tu es et tu restes toujours dans mes pensées une personne de très bonne qualité scientifique et humaine. Julien CHIQUET je te remercie également, tu as été toujours présent aux moments importants pour moi et disponible aux moments difficiles.

Merci Iliia MANDRAKO, Pano KARABELA et XATZIPANAGI. Votre amour pour les maths m'a aidé à construire de bonnes bases mathématiques. Vous êtes toujours dans mon coeur.

Je remercie nos anciennes secrétaires Béatrice RAMONELL et Stéphanie ROUAULT ainsi que la nouvelle Béatrice BARBIER pour leur disponibilité et leur aide. Céline LEDENT et Magali COLLIGNON ont été toujours efficaces quand j'ai eu besoin. A l'aide de Sabine VIDALE plusieurs articles importants pour mon travail ont été à ma disponibilité très rapidement. Je tiens à remercier également tous les membres du LMAC qui ont participé plus ou moins directement au bon déroulement de cette thèse, tout particulièrement Djalil KATEB et Sergio ALVAREZ.

Je garde une pensée pour les amis de route, notamment Amir, Lekbir, Adel, Nabil, Karim, Alfred, Vassili, Hassane, Alfredo, Gaël, Céline. Nous avons passé de bons moments ensemble, nous avons fait plusieurs discussions intéressantes et j'ai beaucoup enrichi mes bagages culturelles à côté de vous. Dans ce contexte, merci aussi à Dima, Sandy, Fadi, Claudia, Joe, Roger, Maged, Sergio.

Mes meilleurs amis de la Grèce, Louis, Nickos et Alexis merci d'avoir été toujours là pour moi et d'être intéressés à mes préoccupations même si parfois elles étaient trop différentes des vôtres. Merci de m'avoir supporté dans les moments très difficiles et de m'avoir écouté pendant de longs moments et de m'avoir donné de précieux conseils... Un grand merci aussi à Maria et Christina.

Je remercie évidemment ma famille pour son irremplaçable soutien. Evgenia et Makis, et ma soeur Bianca, cette thèse est la votre aussi et je tiens à vous exprimer par ces quelques lignes mon amour et ma gratitude. Je vous remercie pour votre amour inconditionnel, pour votre confiance en moi et votre soutien, et pour tout ce que vous m'avez offert.

Mon séjour en France a été spécialement marqué par la rencontre de la femme avec qui je vais partager ma vie. Au sein de notre laboratoire,

Maha a changé ma vie et m'a donné le courage et la force de finir ma thèse assez vite, elle m'a toujours poussé vers l'avant, sa motivation et sa présence m'ont été très indispensables. Mon niveau de français a été lancé grâce à elle, mon envie de travailler a été augmentée grâce à elle, la qualité de ce mémoire a été plus illustrée grâce à elle. Les mots ne sont pas suffisants pour que je la remercie assez.

Il existe plusieurs possibilités d'obtenir le diplôme du Docteur en effectuant une certaine thèse parmi plusieurs thèses possibles. Mais, la chance de rencontrer dans ma vie une personne comme Maha était seulement une, celle de venir en France, à Compiègne précisément, et à effectuer une thèse dans le laboratoire de mathématiques appliquées à l'UTC. Merci à tous les gens qui ont participé à réaliser cette possibilité.

Un chaleureux merci aussi à sa famille qui pensait toujours à moi, tante Angèle, amo Milad, Ghassan et tayta Malké.

Et pour en finir, merci aux gens oubliés ici et qui ont permis, d'une manière quelconque, l'accomplissement de ce travail et merci à tous ceux qui s'intéressent à lire ce manuscrit.

Compiègne, le 26 février 2009.

TABLE DES MATIÈRES

TABLE DES MATIÈRES	ix
NOTATIONS	1
1 INTRODUCTION	5
1.1 CONTEXTE GÉNÉRAL ET MOTIVATIONS	5
1.2 OBJECTIFS ET ORGANISATION DES CHAPITRES	11
2 EMV POUR CERTAINES CARACTÉRISTIQUES DES CM ERGODIQUES	15
2.1 RAPPELS SUR LES CHAÎNES DE MARKOV	17
2.2 ESTIMATION DU MAXIMUM DE VRAISEMBLANCE POUR CM	18
2.3 ESTIMATION DE LA LOI STATIONNAIRE	24
2.4 REPRÉSENTATIONS DE LA VARIANCE DANS LE TLC	29
2.5 ESTIMATION DE LA VARIANCE	35
2.6 EXEMPLES ET ILLUSTRATIONS NUMÉRIQUES	39
CONCLUSION	42
3 EMV POUR DES CHAÎNES SEMI-MARKOVIENNES	43
3.1 RAPPELS SUR LES CHAÎNES SEMI-MARKOVIENNES	46
3.2 LE PROCESSUS COUPLE DE CSM ET TRA	49
3.3 EMV POUR LES CHAÎNES SEMI-MARKOVIENNES FINIES	53
3.4 COMPORTEMENT ASYMPTOTIQUE DE L'EMV	58
3.5 L-TRAJECTOIRES CENSURÉES À UN INSTANT FIXE	68
CONCLUSION	75
4 EMV POUR DES MODÈLES SEMI-MARKOVIENS CACHÉS	77
4.1 DÉFINITION DU MODÈLE CACHÉ SM-MI-B GÉNÉRAL	80
4.2 REPARAMÉTRISATION DE MSMC PAR UNE SOUS-CLASSE DE MMC	84
4.3 CONVERGENCE DE L'EMV	86
4.4 NORMALITÉ ASYMPTOTIQUE DE L'EMV	89
CONCLUSION	98
5 ALGORITHMES EM ET SEM POUR DES MSMC	99
5.1 PRÉLIMINAIRES ET HYPOTHÈSES	101
5.2 UNE VERSION AMÉLIORÉE DE L'ALGORITHME EM POUR MSMC	102
5.3 DÉMONSTRATIONS DES FORMULES POUR EM	108
5.4 ALGORITHME SEM POUR MSMC NON PARAMÉTRIQUE	111
5.5 APPLICATIONS	114
CONCLUSION	124
CONCLUSION ET PERSPECTIVES	129

A	ANNEXES	135
A.1	THÉORÈME LIMITE CENTRALE POUR LES MARTINGALES	136
A.2	PREUVE DU LEMME 2.3	136
A.3	COMPORTEMENT INITIAL	137
A.4	CONDITIONS LEROUX : CONVERGENCE	139
A.5	CONDITIONS BICKEL-RITOV-RYDÉN : NORMALITÉ ASYMPTOTIQUE .	140
	BIBLIOGRAPHIE	141

NOTATIONS

ABBREVIATIONS

CM	Chaîne de Markov
CRM	Chaîne de Renouvellement Markovienne
CSM	Chaîne Semi-Markovienne
CSMC	Chaîne Semi-Markovienne Cachée
EE	Estimateur Empirique
EM	Estimation-Maximisation
EMV	Estimateur du Maximum de Vraisemblance
<i>e.p.</i>	espace paramétrique
<i>e.p.n.</i>	espace paramétrique naturel
IC	Intervalle de Confiance
i.i.d	indépendantes et identiquement distribuées
MCEM	Monte Carlo EM
MM	Modèle de Markov
MMC	Modèle de Markov Caché
MSM	Modèle Semi-Markovien
MSMC	Modèle Semi-Markovien Caché
PM	Processus de Markov
PMC	Processus de Markov Caché
<i>p.s.</i>	presque sûrement
PSM	Processus Semi-Markovien
SAEM	version Stochastique Approximation de l'EM
SEM	version Stochastique de l'EM
SM	Semi-Markovien
ssi	si et seulement si
TLC	Théorème de la Limite Centrale
TLCF	Théorème de la Limite Centrale Fonctionnel
TRA	Temps de Récurrence en Arrière
\mathbf{X}_0^n	(X_0, X_1, \dots, X_n)
$\mathbf{X}_0^n = i_0^n$	$(X_0, X_1, \dots, X_n) = (i_0, i_1, \dots, i_n)$
$i_0^n \in E$	$i_j \in E$, pour tout j tel que $0 \leq j \leq n$

ENSEMBLES

B	espace d'état de la chaîne de Markov couple (\mathbf{Z}, \mathbf{U})
$E = \{1, 2, \dots, s\}$	espace d'état de la chaîne de Markov ou de la chaîne semi-markovienne
E^2	espace produit $E \times E$

E_+^2	$\{(i, j) \in E^2 : p_{ij} > 0\}$
$E_+^{(i)}$	$\{j \in E : (i, j) \in E_+^2\}$
E_*^2	$\{(i, j) \in E^2 : i \neq j\}$
E_{-i}	$E \setminus \{i\}$
\mathbb{N}, \mathbb{N}^*	ensembles des entiers naturels et des entiers strictement positifs
\mathbb{R}, \mathbb{R}^d	ensembles des réels et des vecteurs réels à d dimensions
T_m	$\{0, 1, \dots, m-1\}$
\mathbb{T}_m	$\{1, 2, \dots, m\}$

DÉTERMINISTES

$\mathbf{1}$	vecteur où chaque élément vaut 1
\mathbb{I}	matrice où chaque élément vaut 1
A_{dg}	matrice d'éléments nuls et dont la diagonale correspond à celle de la matrice A
$(e_i)_{i \in E}$	les vecteurs de la base canonique de \mathbb{R}^E

PROBABILISTES

\mathbb{P}, \mathbb{E}	probabilité et espérance associée
$\mathbb{P}_i(\cdot), \mathbb{E}_i(\cdot)$	probabilité conditionnelle $\mathbb{P}(\cdot \mid X_0 = i)$ et espérance associée
$(\Omega, \mathcal{F}, \mathbb{P})$	espace probabilisé
X, X^*	chaînes $(X_n)_{n \in \mathbb{N}}$ et $(X_n)_{n \in \mathbb{N}^*}$ respectivement
X (chap.2)	chaîne de Markov finie
X (chap.3)	temps de séjour conditionnels
Z	chaîne semi-markovienne finie
U	temps de récurrence en arrière
S	instants de saut pour la CSM
J	états visités aux instants de saut
(J, S)	chaîne de renouvellement markovienne
(Z, U)	chaîne de Markov couple
(Z, Y)	chaîne semi-markovienne cachée
$((Z, U), Y)$	chaîne markovienne cachée
P	matrice de transition de la chaîne de Markov X
q	noyau semi-markovien de la chaîne semi-markovienne Z
convergence faible	convergence vague
$\bar{H}_i(\cdot)$	fonction de survie du temps de séjour dans l'état i
$\bar{F}_{ij}(\cdot)$	fonction de survie du temps de séjour conditionnel dans l'état i , conditionné au prochaine état visité j

STATISTIQUES

\mathcal{P}	<i>e.p.n.</i> pour la chaîne de Markov X
\mathcal{Q}	<i>e.p.n.</i> pour la chaîne semi-markovienne Z
\mathcal{H}	<i>e.p.n.</i> pour la chaîne de Markov couple (Z, U)
Θ_g	<i>e.p.</i> des densités conditionnelles des observations de Y sachant le processus caché Z

Θ	espace paramétrique pour le modèle semi-markovien
p	paramètre vectoriel contenant les probabilités de transition de P
q	paramètre vectoriel contenant les éléments du noyau semi-markovien q
η	paramètre vectoriel contenant les probabilités de transition non identiquement nulles de la chaîne de Markov couple (Z, U)
θ	paramètre vectoriel pour le modèle semi-markovien caché (Z, Y)
θ_g	paramètre vectoriel pour les lois conditionnelles de Y sachant Z
$\tilde{\theta}$	paramètre vectoriel pour le modèle markovien caché $((Z, U), Y)$
λ	paramètre vectoriel dans la paramétrisation naturelle pour $\lambda = p, q, \eta, \theta, \theta_g, \tilde{\theta}$
λ^*	paramètre vectoriel dans la paramétrisation minimale pour $\lambda = p, q, \eta, \theta, \tilde{\theta}$
f	vecteur des temps de séjour conditionnels
p^e	vecteur des probabilités de transition pour la CM immergée J
$\hat{\lambda}(m), \hat{\lambda}^*(m)$	EMV du paramètre vectoriel λ et λ^* respectivement pour $\lambda = p, q, \eta, \theta, \tilde{\theta}$
$\hat{\pi}(m)$	EMV de la loi stationnaire
$\tilde{\pi}(m)$	estimateur empirique de la loi stationnaire
$\hat{\sigma}^2(m)$	EMV de la variance
$\xrightarrow{\mathbb{P}_\lambda - p.s.}$	convergence presque sure sous \mathbb{P}_λ pour $\lambda = p, q, \eta, \theta, \tilde{\theta}$
$\xrightarrow{\lambda}$	convergence en loi (vague) sous \mathbb{P}_λ pour $\lambda = p, q, \eta, \theta, \tilde{\theta}$

INTRODUCTION



1.1 CONTEXTE GÉNÉRAL ET MOTIVATIONS

Le travail présenté dans ce mémoire entre dans le cadre de l'estimation des paramètres inconnus dans des modèles statistiques, ce qui constitue un thème important pour l'inférence statistique. Nous présentons particulièrement une étude basée sur la méthode du maximum de vraisemblance pour des phénomènes qui peuvent être modélisés par des processus stochastiques qui possèdent la propriété markovienne d'une manière directe ou indirecte. Il s'agit des modèles markoviens, semi-markoviens et semi-markoviens cachés.

La méthode du maximum de vraisemblance est une méthode qui donne assez souvent de bonnes estimations. Elle a été introduite dans le cadre des variables aléatoires indépendantes et identiquement distribuées (i.i.d.) par Sir Ronald Fisher dans une série de travaux (voir Fisher (1912; 1922; 1925; 1935)). Nous citons Aldrich (1997) pour une présentation bibliographique sur l'introduction et le développement de cette méthode, ainsi que sur les notions que Fisher a développées pour justifier son utilisation et ses limitations, parmi lesquelles, la vraisemblance, la mesure d'information, la statistique exhaustive et la statistique efficace. Fisher renforce les arguments pour l'utilisation de cette méthode en mettant l'accent sur le fait qu'elle est privée des défauts que les deux méthodes concurrentes principales de cette époque possèdent. En effet, Fisher critique la méthode des moments pour le fait qu'elle se base sur un choix arbitraire des équations des moments. Il critique également la méthode des moindres carrés pour le fait qu'un estimateur déduit de cette méthode ne reste pas invariant sous une transformation d'échelle. Depuis, la prépondérance de la méthode du maximum de vraisemblance est justifiée par la richesse des propriétés théoriques et la multitude des applications. Dans le cadre des variables aléatoires i.i.d., l'estimateur du maximum de vraisemblance a été beaucoup étudié, ainsi que ses propriétés asymptotiques pour plusieurs modèles paramétriques (voir par exemple, Cramér (1949), Scholz (2006), Van Der Vaart (1998)). Cette dernière référence présente également les estimateurs du maximum de vraisemblance (EMV) comme un cas particulier des M-estimateurs.

Parmi les propriétés asymptotiques souhaitées pour un estimateur se trouvent les propriétés de convergence et de normalité asymptotique. Sous des conditions « classiques » de régularité introduites par Wald (1949), dans le cadre des variables aléatoires i.i.d. l'EMV, s'il est unique,

alors il est convergent. Également, sous quelques conditions supplémentaires, souvent appelées conditions de type Cramér, l'EMV est asymptotiquement normal. Depuis, plusieurs raffinements de ces conditions ont été développés par différents auteurs (voir par exemple Scholz (2006)). Une fois que l'EMV du paramètre du modèle possède ces deux propriétés asymptotiques, nous pouvons déduire les mêmes propriétés pour des caractéristiques du modèle exprimées en fonction de ce paramètre, à condition qu'il y a un certain degré de régularité. En effet, pour des fonctions continues on peut appliquer le théorème de l'application continue pour conclure à la convergence et pour des fonctions différentiables la méthode delta pour conclure à la normalité asymptotique de l'EMV des caractéristiques.

Il est très avantageux de fournir, avec la justification théorique des bonnes propriétés asymptotiques d'un estimateur, des formules explicites pour l'estimation de l'erreur associée. Cela se fait en construisant des intervalles de confiance asymptotiques pour les quantités à estimer. Notre travail entre dans ce cadre là, en utilisant la méthode du maximum de vraisemblance pour des variables aléatoires qui ne sont pas i.i.d., mais qui ont une dépendance markovienne, semi-markovienne ou semi-markovienne cachée.

Les modèles markoviens constituent un sujet central dans la théorie de probabilités appliquées et de la statistique. En effet, ils peuvent donner parfois une bonne description pour des phénomènes dont l'évolution dans le temps peut être décrite par des variables aléatoires faiblement dépendantes.

Les premiers travaux sur l'estimation statistique pour des modèles markoviens à un espace d'état fini, particulièrement par la méthode du maximum de vraisemblance, ont été faits par Anderson et Goodman (1957) et Billingsley (1961a;b). Ce dernier a étudié l'estimateur du maximum de vraisemblance et ses propriétés asymptotiques dans le cadre d'une observation de longue durée pour un modèle markovien ergodique. Par contre, Anderson et Goodman (1957) ont traité principalement l'estimation dans le cas d'une suite de copies indépendantes d'une chaîne markovienne censurée à un instant fixe. Un autre travail plus récent concernant l'inférence statistique des chaînes de Markov est celui de Basawa et Prakasa Rao (1980). Cependant, l'estimateur du maximum de vraisemblance pour des caractéristiques qui peuvent être exprimées en fonction des paramètres du modèle n'a été étudié que plus récemment. Nous citons dans ce contexte le travail de Sadek et Limnios (2002) pour l'estimation de la fiabilité, la disponibilité et le taux de défaillance dans des systèmes qui peuvent être modélisés par des chaînes de Markov finies et ergodiques.

Parmi les caractéristiques intéressantes à estimer dans un modèle markovien ergodique, il y a sa loi stationnaire et la variance asymptotique du théorème de la limite centrale pour des fonctionnels additifs. En ce qui concerne la variance asymptotique, Chauveau et Diebolt (2003) étudient un processus de variance empirique en temps continu basé sur L i.i.d. chaînes parallèles. Stefanov (1995) considère également une estimation du maximum de vraisemblance pour les processus markoviens et les chaînes

markoviennes via des familles exponentielles et obtient des solutions explicites pour les paramètres de la variance pour le théorème de la limite centrale fonctionnel. Il existe des travaux récents qui se concentrent sur l'estimation de fonctions des chaînes de Markov en utilisant des méthodes de type Monte-Carlo par chaînes de Markov. Un aperçu sur ces méthodes se trouve dans Roberts et Rosenthal (2004).

Le théorème de la limite centrale est un aspect de la théorie des chaînes de Markov qui mérite d'être étudié davantage même dans un espace d'état fini. De même, l'estimation de la loi stationnaire dans le cas des chaînes de Markov ergodiques. Les expressions analytiques des estimateurs et la précision de l'erreur asymptotique de ces estimations peuvent être obtenues par la méthode du maximum de vraisemblance.

Nous nous sommes ainsi intéressés à travailler sur l'estimation des caractéristiques qui découlent d'une chaîne de Markov finie, à fournir différentes représentations pour ces caractéristiques et à calculer l'erreur associée. En fait, l'estimation des caractéristiques pour les chaînes markoviennes finies est importante pour les applications et elle n'a pas été beaucoup explorée. De plus, les différentes représentations pour les caractéristiques permettent d'obtenir des expressions facilement programmables et calculables. Enfin, l'étude pour des modèles markoviens nous a aussi servi comme outil afin d'étudier les modèles semi-markoviens.

Les modèles semi-markoviens constituent une généralisation des modèles markoviens. En effet, à l'opposé de ces derniers, le temps de séjour dans chaque état n'est pas nécessairement géométrique mais peut suivre une loi quelconque. Ce qui explique la grande variété des phénomènes qu'ils peuvent modéliser. Ainsi, ces modèles ont été beaucoup étudiés et appliqués, principalement à temps continu. Cependant, les modèles à temps discret ont été moins étudiés.

Les processus semi-markoviens ont été introduits indépendamment par Lévy (1954), Smith (1955) et Takacs (1954). Depuis, l'estimation de ces processus à temps continu a fait l'objet d'une littérature relativement abondante. Nous citons, à titre d'exemple, Moore et Pyke (1968) qui ont présenté des estimateurs empiriques et du maximum de vraisemblance pour le noyau semi-markovien d'un processus semi-markovien fini. Lagakos et al. (1978) ont obtenu l'estimateur du maximum de vraisemblance pour le noyau semi-markovien d'un processus semi-markovien fini, avec des états absorbants. Gill (1980) a construit un estimateur pour le noyau semi-markovien fini en utilisant des processus de comptage. Ouhbi et Limnios (1999; 2003) ont étudié des estimateurs du noyau semi-markovien, de la fonction de renouvellement markovienne et de la matrice de transition semi-markovienne. Limnios et al. (2005) ont obtenu un estimateur de la loi stationnaire d'un processus semi-markovien.

Dans le cadre du temps discret, Barbu et Limnios (2006a) ont considéré des estimateurs empiriques adaptés aux systèmes semi-markoviens non paramétriques, ayant des bonnes propriétés asymptotiques. Ces estimateurs peuvent être considérés comme estimateurs du maximum de vraisemblance approché si on ne prend pas en compte le temps de séjour

dans le dernier état du système. Ils ont également montré que ces estimateurs sont fortement convergents et que chaque composante de l'EMV approché est asymptotiquement normale.

À notre connaissance, l'étude de l'estimateur du maximum de vraisemblance exact qui correspond au modèle semi-markovien non paramétrique n'as pas encore été considérée. En plus, les résultats asymptotiques donnés dans Barbu et Limnios (2006a) pour une seule composante de l'EMV approché nécessitent d'être examinés pour plusieurs composantes. En effet, les paramètres sont dépendants et donc le comportement de l'EMV doit être examiné globalement. En outre, il est nécessaire de considérer des versions vectorielles afin d'examiner le comportement asymptotique d'autres caractéristiques qui dépendent du noyau semi-markovien. Ce qui nous a poussé donc à étudier l'estimateur du maximum de vraisemblance exact globalement en considérant des composantes vectorielles de ce dernier.

Cette tâche a été accomplie en utilisant la chaîne markovienne couple obtenue en associant à la chaîne semi-markovienne, la suite de temps de récurrence en arrière. Nous citons Chryssaphinou et al. (2008) pour les propriétés asymptotiques de la chaîne couple. L'utilisation de ce processus pour des études probabilistes dans le cadre des chaînes semi-markoviennes n'est pas nouveau. Voir par exemple Anselone (1960). Stenflo (1996) utilise ce couple pour déduire des propriétés asymptotiques pour des systèmes itératifs des fonctions aléatoires, contrôlés par une chaîne semi-markovienne, par celles des systèmes correspondants contrôlés par une chaîne de Markov. Notre contribution porte sur l'utilisation de l'ensemble des probabilités de transition non identiquement nulles de la chaîne markovienne couple pour faire de l'inférence statistique pour les chaînes semi-markoviennes, via la reparamétrisation associée du noyau semi-markovien. La motivation a été d'utiliser cette reparamétrisation pour obtenir l'EMV exact dans un modèle semi-markovien non paramétrique et démontrer les propriétés asymptotiques de convergence et de normalité asymptotique pour chaque sous-vecteur fini de cet EMV. Cela nécessite d'étendre les résultats existants pour les chaînes de Markov finies à la classe des chaînes de Markov infinies induite par le processus couple.

Cette reparamétrisation nous a également servi pour l'étude de l'estimation du maximum de vraisemblance dans le cas d'une suite des copies indépendantes d'une chaîne semi-markovienne censurée à un instant fixe. Ce qui, à notre connaissance, n'a pas encore été considéré jusqu'à présent.

La méthodologie et les résultats obtenus pour l'inférence statistique sur les chaînes semi-markoviennes peuvent être utilisés pour les chaînes semi-markoviennes cachées. Ces chaînes constituent une classe de modèles relativement nouvelle et ont le potentiel de modéliser des phénomènes plus complexes que ceux qui peuvent être modélisés par les modèles markoviens cachés, qui sont beaucoup utilisés.

L'avantage des modèles markoviens et semi-markoviens se trouve dans le fait qu'à l'aide des processus de comptage appropriés, on peut déduire facilement des statistiques suffisantes qui sont observables pour

les paramètres du modèle. Cependant, les modèles markoviens cachés et semi-markoviens cachés sont caractérisés par le fait que les statistiques suffisantes ne peuvent pas être directement observées, ce qui rend leur étude à la fois plus difficile et plus intéressante.

La classe des modèles markoviens cachés (MMC) a attiré l'attention à la fois théoriquement et algorithmiquement. Leur utilisation est devenue l'une des techniques les plus puissantes et les plus populaires dans plusieurs domaines scientifiques. Notamment, la reconnaissance de la parole (Rabiner (1989), Rabiner et Juang (1993)), la biologie (Krogh et al. (1994a), Krogh et al. (1994b)), le traitement d'images (Li et Gray (2000)) et plusieurs autres domaines (Bhar et Hamori (2000), Sansom (1998)). Pour une revue générale sur l'utilisation des processus de Markov cachés (PMC) dans la statistique et la théorie de l'information, nous renvoyons le lecteur à Ephraïm et Merhav (2002). Un livre récent qui couvre une grande partie de travaux sur les modèles markoviens cachés est celui de Cappé et al. (2005).

Le premier travail théorique sur l'inférence statistique pour les MMC est celui de Baum et Petrie (1966), où les auteurs ont étudié un processus stochastique bidimensionnel (Z, Y) dont la première composante Z est une chaîne de Markov finie non observée directement, et la deuxième composante Y est un processus observé, où Y conditionné par Z , forme une suite des variables aléatoires conditionnellement indépendantes avec un espace d'état fini. Ils ont également démontré la convergence et la normalité asymptotique de l'estimateur du maximum de vraisemblance pour ce modèle. Ces MMC sont souvent désignés comme étant des fonctions probabilistes des chaînes de Markov. Les conditions de convergence ont été affaiblies dans Petrie (1969) qui a étudié également le problème de l'identifiabilité de ces modèles. Dans Leroux (1992), Bickel et al. (1998), les auteurs ont démontré respectivement la convergence et la normalité asymptotique de l'EMV pour un processus observable à un espace d'état général. Lorsque l'espace d'état du processus caché est pseudo-compact Jensen et Petersen, Douc et Matias (2001), Douc et al. (2004) ont étudié des propriétés asymptotiques pour l'EMV. Pour le cas où l'espace d'état du processus caché n'est pas nécessairement compact, nous citons Fuh (2006). Pour plus de bibliographie sur les travaux théoriques voir Cappé et al. (2005).

Bien que les MMC puissent être utiles dans plusieurs cas, ils souffrent d'une limitation importante. En effet, ils ne permettent que la loi géométrique pour les temps de séjour dans les états du processus caché. Un tel modèle peut être inexact pour décrire plusieurs problèmes réels (voir Levinson (1986)).

Les modèles semi-markoviens cachés (MSMC) constituent une généralisation assez naturelle des MMC. Dans le cas des MSMC, le processus caché Z , décrit ci-dessus, n'est plus une chaîne de Markov comme c'est le cas pour les MMC, mais une chaîne semi-markovienne. À l'opposé des MMC, les MSMC permettent n'importe quelle loi pour les temps de séjour dans les différents états du processus caché. Cela est important

afin de réaliser une meilleure description de l'ensemble des données d'un problème réel quand la propriété markovienne n'est pas suffisante pour décrire la dépendance entre les variables aléatoires du processus caché.

Les MSMC ont été présentés pour la première fois dans Ferguson (1980). Depuis, plusieurs algorithmes appropriés ont été construits dans le but d'estimer les paramètres inconnus du modèle. En ce qui concerne la démonstration des propriétés théoriques de l'EMV pour des MSMC avec un support fini pour les lois de temps de séjour conditionnelles, Barbu et Limnios (2006) ont utilisé le processus couple avec un espace paramétrique formé par toutes les probabilités de transition de la matrice de transition de ce couple qui sont identiquement nulles ou non. Cependant, Barbu et Limnios (2008) ont fourni une justification théorique plus solide pour la convergence et la normalité asymptotique de l'EMV, en considérant une reparamétrisation du modèle semi-markovien à l'aide de l'espace paramétrique naturel du processus couple (sans les probabilités de transition nulles). Cette reparamétrisation est une idée qui a été développée au cours de cette thèse dans un cadre plus général pour l'étude des propriétés asymptotiques de l'EMV pour les MSMC généraux avec dépendance des temps de récurrence en arrière à support fini pour les lois de temps de séjour conditionnelles. Néanmoins, l'utilisation d'un support fini reste une limitation importante.

Les modèles semi-markoviens cachés généraux avec dépendance des temps de récurrence en arrière peuvent avoir une flexibilité plus grande que celle des modèles semi-markoviens cachés généraux. Cela peut être expliqué par le fait que sachant l'état caché du système, la loi d'observation peut dépendre du temps qu'il a passé avant le dernier saut du processus semi-markovien.

Dans le but d'obtenir l'EMV des paramètres des MSMC, un algorithme EM (Estimation-Maximisation) peut être appliqué. Pour une comparaison avec des algorithmes de type Newton voir Bulla (2006). L'algorithme EM est spécialement destiné à traiter des problèmes avec des données manquantes. Ferguson (1980) a été le premier à proposer un algorithme EM pour les MSMC. Depuis, plusieurs travaux de recherche ont été effectués dans cette direction. Nous donnons, dans le chapitre 5 de ce mémoire, quelques références pour différentes applications comme la reconnaissance de la parole, l'analyse des suites biologiques d'ADN, les points de rupture pour des données de précipitation, la climatologie et l'analyse de la structure des plantes. Dans la majorité des applications, l'intérêt est focalisé sur des modèles paramétriques.

Barbu et Limnios (2006; 2008) ont présenté un algorithme EM pour des modèles semi-markoviens cachés non paramétriques, dans le but d'estimer le noyau semi-markovien du processus semi-markovien caché et les probabilités d'émission.

Leur travail nous a donné la motivation d'étudier davantage leur algorithme initial afin d'améliorer sa performance, notamment, en terme de temps d'exécution. Dans cette perspective, nous avons développé également un algorithme EM stochastique (SEM) pour estimer les caractéristiques du MSMC non paramétrique.

Dans la littérature on peut trouver une grande variété des versions stochastiques pour l'algorithme EM. La version la plus simple est l'algorithme SEM proposé par Celeux et Diebolt (1985). La principale idée de cet algorithme consiste à remplacer le calcul de l'étape-E de l'algorithme EM, l'étape de l'estimation, par une étape de simulation. Ainsi, à chaque itération d'un SEM, on doit générer une trajectoire du processus caché en utilisant une loi conditionnelle, sachant les données observées et la valeur du paramètre actuel. Une généralisation de l'algorithme SEM est l'algorithme Monte Carlo EM (MCEM). Il a été introduit par Wei et Tanner (1990). Dans chaque étape de la simulation de l'algorithme MCEM, on génère plusieurs trajectoires du processus caché échantillonnées indépendamment par la même loi. Une généralisation de l'algorithme MCEM a été développée par Delyon et al. (1999). Il s'agit de l'algorithme EM Approximation Stochastique (SAEM). Cet algorithme peut prendre en compte, à chaque étape de la simulation, les données simulées dans l'étape précédente. Cela est fait en construisant une fonction appropriée pour l'étape de maximisation par une combinaison convexe d'une approximation de Monte Carlo dans l'étape actuelle avec des fonctions correspondantes calculées dans l'étape précédente.

C'est pour cette version stochastique que nous avons opté. En effet, l'implémentation de SAEM semble être directe. En plus, cette version donne la possibilité d'utiliser des trajectoires plus longues, afin d'avoir plus de précision sur l'estimation des paramètres.

Enfin, ces algorithmes, proposés dans ce mémoire, peuvent être utiles dans plusieurs applications, comme par exemple, l'analyse des suites d'ADN dans le cas où une chaîne semi-markovienne cachée peut bien modéliser la génération des bases d'ADN. Un cas particulier assez intéressant à considérer peut être la détection des îlots CpG. La particularité des suites d'ADN se trouve dans le fait que des trajectoires assez grandes sont disponibles.

1.2 OBJECTIFS ET ORGANISATION DES CHAPITRES

Objectifs

Les objectifs de ce travail se résument principalement en les points suivants.

i) Nous étudions les propriétés asymptotiques de l'EMV associé à quelques principales caractéristiques des modèles markoviens, afin d'estimer les valeurs et mesurer les erreurs de cette estimation en construisant des intervalles de confiance. Notre intérêt se focalise sur l'estimation de la loi stationnaire et sur l'estimation de la variance asymptotique dans le théorème de la limite centrale pour des fonctionnelles additives pour les CM finies ergodiques (irréductibles). Nous nous sommes particulièrement intéressés à fournir des formules explicites et facilement programmables pour des applications potentielles. Pour cela, nous proposons des formules calculables pour la variance du TLC et également pour la variance asymptotique de son EMV ;

ii) Nous étudions les propriétés asymptotiques de l'EMV (convergence et normalité asymptotique) pour des modèles semi-markoviens non paramétriques et nous construisons des intervalles de confiance en se basant sur ces propriétés asymptotiques pour une chaîne semi-markovienne finie ergodique. Nous considérons également l'estimation du maximum de vraisemblance dans le cas de L trajectoires des copies indépendantes d'une chaîne semi-markovienne (non nécessairement ergodique) censurée à un instant fixe. Nous étudions pour ce cas les propriétés théoriques pour l'EMV de la partie du noyau semi-markovien qui peut être estimé jusqu'à l'instant d'observation. Nous cherchons également à mesurer la concentration de la loi du temps de séjour dans chaque état de la CSM sur l'ensemble déterminé par l'intervalle d'observation ;

iii) Nous incorporons l'approche semi-markovienne dans le cadre des modèles cachés afin de prendre en compte des phénomènes plus complexes dont le comportement peut être approché par des modèles de type semi-markovien caché (MSMC). Ces phénomènes aléatoires sont décrits par un processus bidimensionnel (Z, Y) , où Z est une chaîne semi-markovienne (non observée), et Y un processus (observé) dont la loi conditionnelle de chaque v.a. Y_n , conditionné par Z , dépend uniquement de l'état Z_n . Dans notre étude, nous nous sommes intéressés à considérer également le cas plus général où la loi conditionnelle de Y_n peut dépendre de la valeur de Z_n et du temps pendant lequel le système a séjourné dans cet état. Il s'agit ainsi d'estimer le noyau semi-markovien de Z et les lois conditionnelles qui génèrent le processus Y . Notre travail sur les MSMC comporte deux axes : un axe de statistique mathématique pour la démonstration des résultats théoriques, en particulier la convergence des estimateurs, et un axe numérique pour le développement et l'implémentation des algorithmes appropriés.

Organisation des chapitres

Ce mémoire est structuré comme suit.

Dans le deuxième chapitre, nous considérons l'EMV pour quelques caractéristiques d'une chaîne de Markov finie ergodique. Nous construisons l'EMV de la loi stationnaire et nous étudions ses propriétés asymptotiques en la comparant avec son estimateur empirique. Nous étudions la variance asymptotique du théorème de la limite centrale (TLC) pour des fonctionnelles additives pour les CM. Également, nous considérons différentes formes de l'EMV pour cette variance et nous montrons la convergence et la normalité asymptotique de cet estimateur. De plus, nous donnons deux formes explicites pour la variance asymptotique de cet EMV.

Dans le troisième chapitre, nous considérons un modèle semi-markovien non paramétrique. Nous présentons l'EMV exact du noyau qui gouverne l'évolution de la chaîne semi-markovienne (CSM) et démontrons la convergence forte, ainsi que la normalité asymptotique de chaque sous-vecteur fini de cet estimateur en obtenant des formes explicites pour

les matrices de covariance asymptotiques. De plus, nous montrons la relation de l'EMV exact avec son estimateur empirique. Cela permet aussi de démontrer la normalité asymptotique de chaque sous-vecteur fini de l'estimateur empirique. Notre méthode est basée sur un changement d'espace paramétrique qui correspond à l'espace paramétrique naturel de la CM formée par la CSM et par la suite des temps de récurrence en arrière (TRA). Ceci a été appliqué pour une observation de longue durée d'une seule trajectoire d'une CSM, ainsi que pour une suite des i.i.d trajectoires d'une CSM censurées à un instant fixe.

Dans le quatrième chapitre, nous introduisons un modèle semi-markovien caché général avec dépendance des temps de récurrence en arrière. Nous montrons sa convergence et sa normalité asymptotique. Nous déduisons également des expressions explicites pour les matrices de covariance asymptotiques (en fonction du noyau semi-markovien), qui apparaissent dans le théorème de la limite centrale pour l'EMV des principales caractéristiques des chaînes semi-markoviennes.

Dans le cinquième chapitre, nous proposons une version améliorée de l'algorithme EM (Estimation-Maximisation) et une version stochastique de cet algorithme (SEM), afin de trouver l'EMV pour les modèles semi-markoviens cachés non paramétriques. En particulier, nous optons pour l'algorithme de l'Approximation Stochastique EM (SAEM). Le support pour les lois des temps de séjour conditionnelles peut être fini ou infini. Notre méthode s'applique directement aux cas où le support des lois conditionnelles des temps de séjour dans les états de la CSM est fini ou quand l'erreur de l'approximation par un support fini approprié est suffisamment petite. Cela couvre également le cas où le support est concentré dans l'intervalle d'observation. Plusieurs tests numériques sont présentés pour illustrer la performance des algorithmes.

Notons finalement que cette thèse constitue une suite des travaux déjà effectués au sein de notre laboratoire, notamment la thèse de Barbu V. (2005) et de Sadek A. (2003).

Notre travail peut avoir plusieurs applications dans plusieurs domaines, tels que la biostatistique, la fiabilité, la théorie des files d'attente, recherche opérationnelle, la maintenance, la communication, etc.

Citons pour finir une application potentielle de notre travail : la détection automatique des îlots CpG dans une séquence d'ADN, ce qui sert le diagnostic automatique en médecine.

EMV POUR CERTAINES CARACTÉRISTIQUES DES CM ERGODIQUES

SOMMAIRE

2.1	RAPPELS SUR LES CHAÎNES DE MARKOV	17
2.2	ESTIMATION DU MAXIMUM DE VRAISEMBLANCE POUR CM	18
2.3	ESTIMATION DE LA LOI STATIONNAIRE	24
2.4	REPRÉSENTATIONS DE LA VARIANCE DANS LE TLC	29
2.5	ESTIMATION DE LA VARIANCE	35
2.6	EXEMPLES ET ILLUSTRATIONS NUMÉRIQUES	39
	CONCLUSION	42

Les processus de Markov (PM) constituent un sujet central dans la théorie des probabilités appliquées et de la statistique. En effet, plusieurs problèmes réels peuvent être modélisés par ce type de processus stochastiques à temps discret ou continu. L'estimation statistique pour les chaînes de Markov (CM) finies et ergodiques a été traitée principalement par Billingsley (1961a;b). L'estimation est basée sur la méthode du maximum de vraisemblance. En considérant une trajectoire observée sur l'intervalle temporel $\{0, 1, \dots, m\}$, l'étape initiale consiste à trouver l'estimateur du maximum de vraisemblance (EMV) pour la matrice de transition. Plusieurs quantités peuvent également être estimées, comme la loi stationnaire, les temps moyens d'entrée dans des différentes parties de l'espace d'état et autres caractéristiques probabilistes qui sont impliquées dans l'analyse d'une CM. À ce but, un problème important est d'estimer la variance σ^2 du théorème de la limite centrale (TLC) pour les CM. En particulier, lorsqu'on considère des fonctionnels additifs d'une CM, la variance est la même que pour le cas des théorèmes limites centrales fonctionnels (TLCF) pour les PM. Des travaux récents se concentrent sur cet aspect dans le contexte des méthodes de Monte-Carlo par chaînes de Markov. Un aperçu récent est dû à Roberts et Rosenthal (2004). Chauveau et Diebolt (2003) étudient un processus de variance empirique en temps continu basé sur i.i.d. chaînes parallèles. Stefanov (1995) considère également une

estimation du maximum de vraisemblance pour PM et CM via une approche par des familles exponentielles et obtient des solutions explicites pour les paramètres de la variance pour TLCF. Pour une approche unifiée concernant les familles exponentielles des processus stochastiques voir Küchler et Sørensen (1997).

Le TLC est un aspect de la théorie des CM qui mérite d'être étudié davantage même dans un espace d'état discret ou fini. Notre travail se focalise notamment sur les aspects calculatoires pour l'EMV de la variance associée. Pour cela, nous proposons des formules calculables pour la variance de la TLC et également pour la variance asymptotique de son EMV.

Ce chapitre est organisé comme suit. Dans les deux premières sections nous passons en revue quelques résultats de base sur les chaînes de Markov discrètes à espace d'état dénombrable et sur l'estimation du maximum de vraisemblance pour les CM à espace d'état fini. Dans la section 2.3, nous considérons l'estimation de la loi stationnaire et prouvons l'équivalence asymptotique entre l'EMV et l'estimateur empirique de la loi stationnaire, ainsi que leur relation exacte. Dans la section 2.4, nous étudions la variance associée au TLC sous des différentes formes matricielles. Ces formes résultent des différentes démonstrations du TLC existantes dans la littérature. Pour ces démonstrations, nous renvoyons à Doob (1953), Billingsley (1961b), Dacunha-Castelle et Duflo (1983), Meyn et Tweedy (1993) et Port (1994).

Le point de départ pour notre analyse est l'expression de la variance donnée dans Port (1994). Notre tâche principale est accomplie dans la section 2.5, où nous construisons l'EMV de la variance dans l'une de ces formes matricielles et montrons qu'il est fortement consistant et asymptotiquement normal. La variance asymptotique qui correspond à cette loi asymptotique est calculée sous deux différentes formes matricielles. Dans la dernière section, nous présentons quelques exemples qui illustrent les résultats théoriques.

Nous mentionnons que les résultats présentés dans ce chapitre font l'objet de l'article Trevezas et Limnios (à paraître 2008b).

2.1 RAPPELS SUR LES CHAÎNES DE MARKOV

Dans cette section nous donnons quelques résultats de base sur les chaînes markoviennes. D'abord, nous introduisons quelques abréviations pour les notations utilisées tout au long de ce mémoire.

Nous désignons par

- $\mathbf{X}_0^n := (X_0, X_1, \dots, X_n)$,
- $(\mathbf{X}_0^n = i_0^n) := ((X_0, X_1, \dots, X_n) = (i_0, i_1, \dots, i_n))$,
- $i_0^n \in E$ indique que $i_j \in E$, pour tout j tel que $0 \leq j \leq n$,
- $\mathbb{P}_i(\cdot)$ représente la mesure de probabilité $\mathbb{P}(\cdot \mid X_0 = i)$,
- $\mathbb{E}_i(\cdot)$ représente l'espérance par rapport à $\mathbb{P}_i(\cdot)$.

Notons que s'il s'agit d'un espace d'état dénombrable E , nous le considérons toujours muni de la σ -algèbre formée par les parties de E .

Nous définissons une CM de la manière suivante.

Définition 2.1 Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $\mathbf{X} := (X_n)_{n \in \mathbb{N}}$ un processus stochastique à valeurs dans un espace d'état dénombrable E . Le processus \mathbf{X} est une chaîne de Markov si pour tout $n \in \mathbb{N}^*$ et $i_0^n \in E$, il satisfait

$$\mathbb{P}(X_n = i_n \mid \mathbf{X}_0^{n-1} = i_0^{n-1}) = \mathbb{P}(X_n = i_n \mid X_{n-1} = i_{n-1}). \quad (2.1)$$

De plus, \mathbf{X} est homogène par rapport au temps si la probabilité ci-dessus ne dépend pas de n . Pour une CM homogène, nous notons $P = (p_{ij})_{i,j \in E}$ sa matrice de transition, où $p_{ij} := \mathbb{P}(X_n = j \mid X_{n-1} = i)$ et $\alpha(\cdot)$ sa loi initiale, complètement déterminée par le vecteur α de composantes $\alpha_i := \mathbb{P}(X_0 = i)$.

Une classe importante des temps d'arrêt (voir par exemple Port (1994)) pour la CM \mathbf{X} est constituée par les temps de retour dans une partie $A \subseteq E$.

Définition 2.2 Soit $A \subseteq E$. La variable aléatoire

$$T_A = \inf\{n > 0 : X_n \in A\} \quad (2.2)$$

est appelée temps de retour dans l'ensemble d'état A . Lorsque $A = \{i\}$, nous posons $T_i := T_{\{i\}}$.

Définition 2.3 La chaîne de Markov \mathbf{X} est dite irréductible si pour tout $i, j \in E$

$$\mathbb{P}_i(T_j < \infty) > 0. \quad (2.3)$$

Si, de plus¹ il existe un état $a \in E$ tel que

$$\mathbb{P}_a(T_a < \infty) = 1, \quad (2.4)$$

¹Pour une CM irréductible, l'existence d'un état a qui vérifie (2.4) constitue plutôt un critère qui transfère la propriété de récurrence de cet état dans tous les états, et non pas une définition.

alors la chaîne X est dite irréductible et récurrente. En examinant l'espérance de cette v.a. nous obtenons deux types de récurrence, positive ou nulle si $\mathbb{E}_a(T_a)$ est finie ou infinie respectivement. Une CM irréductible et récurrente positive est appelé ergodique².

La classe de CM la plus utile pour les applications est celle des chaînes ergodiques. Cette classe est caractérisée par l'existence d'un vecteur de probabilité à composantes strictement positives qui vérifie de manière unique la relation (2.5) donnée ci-dessous.

Définition 2.4 Une CM X admet une loi stationnaire $\pi(\cdot)$, complètement déterminée par le vecteur π appelé vecteur stationnaire, si l'équation d'équilibre suivante est vérifiée

$$\pi = \pi P. \quad (2.5)$$

Le théorème ci-dessous caractérise la classe des CM ergodiques (voir par exemple Port (1994)).

Théorème 2.1 Une chaîne de Markov X est ergodique ssi il existe une loi stationnaire unique à support l'ensemble E .

2.2 ESTIMATION DU MAXIMUM DE VRAISEMBLANCE POUR CM

Afin de présenter des aspects intéressants de l'estimation statistique des CM finies, nous exploitons la liaison entre la loi multinomiale et l'EMV du modèle markovien. Une liaison plus forte a été démontrée à l'aide de la loi asymptotique de l'EMV via la formule de Whittle (voir par exemple Billingsley (1961b)). En effet, cette formule donne le nombre des trajectoires d'une CM censurées à un instant donné et qui ont une matrice des fréquences fixe³.

Dans notre travail, nous utilisons également la loi asymptotique de l'EMV. Pour une démonstration directe, nous pourrions nous servir du théorème de la limite centrale pour les martingales (voir l'annexe A.1). Cela est fait dans le chapitre 3 afin d'examiner le comportement asymptotique de l'EMV dans un modèle semi-markovien, en utilisant une CM à espace d'état infini. La méthodologie est similaire, même plus facile, pour une CM finie.

D'abord, avant de passer au modèle markovien nous introduisons une famille de lois multinomiales $\mathcal{M}(N, p_1, p_2, \dots, p_s)$ où le paramètre $N = n$ est connu et dont l'ordre est borné par un nombre s . Nous donnons dans la proposition 2.1, l'EMV pour une observation d'un vecteur aléatoire qui appartient à cette famille. Son espace paramétrique naturel s'écrit

$$\mathcal{P}_{\mathcal{M}} := \{p = (p_1, p_2, \dots, p_s) \in \mathbb{R}^s : p_i \geq 0, 1 \leq i \leq s, \sum_{i=1}^s p_i = 1\}. \quad (2.6)$$

² Notons que dans la littérature ce terme est souvent accompagné par la propriété d'apériodicité.

³ Pour une trajectoire X_0^m d'une CM, la matrice des fréquences est décrite par le nombre des transitions pour tous les couples d'états possibles quand le nombre total des transitions est m .

Considérons un paramètre $\mathbf{p} \in \mathcal{P}_{\mathcal{M}}$ et notons d le nombre des composantes strictement positives et i_1, i_2, \dots, i_d les indices correspondants. Alors, \mathbf{p} peut être considéré comme un membre d'une loi multinomiale d'ordre d ($d \leq s$) dans cette famille de lois multinomiales d'ordre borné par s . Cette description de l'espace paramétrique est avantageuse, notamment :

i) dans un modèle plus complexe où apparaît un nombre fini de lois multinomiales dont l'ordre n'est pas nécessairement identique à l'ordre maximal s ,

ii) pour encadrer de manière naturelle une suite d'observations d'un vecteur aléatoire multinomial d'ordre s , où l'ordre n'est atteint qu'après un nombre aléatoire d'observations,

iii) dans le cas où l'ordre n'est pas connu a priori et doit donc être estimé.

Pour le cas ii), cela est justifié par le fait qu'une observation d'une loi multinomiale peut ne pas être suffisante pour atteindre l'ordre et par conséquent c'est plus pratique de considérer l'ordre comme un paramètre à estimer. Cette convention nous permet de trouver des solutions de type EMV d'ordre réduit dans le cas où il y a des composantes X_i nulles d'un vecteur aléatoire X .

Puisque la somme des composantes est égale à l'unité (une relation de dépendance linéaire) nous pouvons enlever un paramètre et utiliser une représentation minimale $\mathcal{P}_{\mathcal{M}}^* \subset \mathbb{R}^{d-1}$ de type

$$\mathcal{P}_{\mathcal{M}}^* := \{\mathbf{p}^* = (p_{i_1}, p_{i_2}, \dots, p_{i_{d-1}}) \in \mathbb{R}^{d-1} : p_{i_k} \geq 0, 1 \leq k \leq d-1, \sum_{k=1}^{d-1} p_{i_k} \leq 1\}. \quad (2.7)$$

Dans ce qui suit, nous adoptons la convention que $0^0 = 1$ et $0 \log 0 = 0$ pour des termes sous-jacents de ce type dans la fonction de vraisemblance et log-vraisemblance d'une observation de la loi multinomiale respectivement. Pour la démonstration de la proposition suivante nous n'utilisons pas la méthode de multiplicateurs de Lagrange mais nous optons pour une utilisation systématique de la paramétrisation minimale qui joue un rôle important dans notre travail.

Proposition 2.1 *Soit \mathbf{X}_1^s un vecteur aléatoire qui suit une loi dans la famille de lois multinomiales du paramètre $N = n$ connu, d'ordre $d \leq s$, i.e.,*

$$\mathbb{P}_{\mathbf{p}}(\mathbf{X}_1^s = \mathbf{x}_1^s) = \binom{n}{x_1, x_2, \dots, x_s} p_1^{x_1} p_2^{x_2} \dots p_s^{x_s}, \quad (2.8)$$

où $\mathbf{p} \in \mathcal{P}_{\mathcal{M}}$ (voir (2.6)) et $\sum_{i=1}^s x_i = n$. L'EMV du paramètre \mathbf{p} , soit $\mathbf{p}^* \in \mathcal{P}_{\mathcal{M}}^*$ (voir (2.7)) associé à une observation est donné par :

$$\hat{\mathbf{p}}^* = \left(\frac{X_{i_1}}{n}, \frac{X_{i_2}}{n}, \dots, \frac{X_{i_{d-1}}}{n} \right). \quad (2.9)$$

Par conséquent, l'EMV du paramètre \mathbf{p} s'écrit

$$\hat{\mathbf{p}} = \left(\frac{X_1}{n}, \frac{X_2}{n}, \dots, \frac{X_{s-1}}{n}, \frac{X_s}{n} \right). \quad (2.10)$$

Démonstration. La fonction de log-vraisemblance s'écrit

$$\ell(\mathbf{p}) = c + \sum_{i=1}^s X_i \log p_i, \quad (2.11)$$

où c est le logarithme du coefficient binomial et ne dépend pas du paramètre \mathbf{p} . Soit

$$d = \sum_{i=1}^s \mathbb{1}_{\{X_i > 0\}}. \quad (2.12)$$

Nous allons montrer que d est l'EMV de l'ordre de la loi multinomiale.

Soit $Y_1, Y_2, \dots, Y_{d-1}, n - \sum_{i=1}^{d-1} Y_i$ une nouvelle énumération des v.a. non nulles dans l'expression (2.11) et $\theta_1, \theta_2, \dots, \theta_{d-1}$ les paramètres p_i qui correspondent aux v.a. Y_1, Y_2, \dots, Y_{d-1} . Nous posons aussi θ_d la somme des paramètres qui correspondent aux v.a. $X_i = 0$. La log-vraisemblance peut être reformulée dans la forme suivante :

$$\ell(\mathbf{p}) = c + \sum_{i=1}^{d-1} Y_i \log \theta_i + (n - \sum_{i=1}^{d-1} Y_i) \log(1 - \sum_{i=1}^{d-1} \theta_i - \theta_d). \quad (2.13)$$

C'est évident qu'en fonction de θ_d , la log-vraisemblance $\ell(\mathbf{p})$ est décroissante. Puisque $0 \leq \theta_d \leq 1 - \sum_{i=1}^{d-1} \theta_i$, donc afin de maximiser $\ell(\mathbf{p})$ il suffit de le faire quand $\theta_d = 0$. Cela implique que $\hat{p}_i = 0$, pour les $s - d$ paramètres où $X_i = 0$. Par conséquent, le problème de maximisation se réduit au cas classique de l'EMV d'une observation de la famille multinomiale d'ordre d , paramétrisée dans un ensemble $\Theta = \mathcal{P}_{\mathcal{M}}^* \subset \mathbb{R}^{d-1}$ (voir 2.7). La fonction de log-vraisemblance $\ell(\theta)$ est bien définie et différentiable dans Θ^0 et par conséquent le point d'extrémum se trouve comme solution unique des équations de vraisemblance $\nabla_{\theta} \ell(\theta) = 0$. Par une extension continue de $\ell(\theta)$ en prenant $\ell(\theta) = -\infty$ à la frontière nous déduisons directement que l'extrémum est en fait un maximum. En combinant les arguments précédents nous arrivons facilement à (2.9) et à (2.10). \square

Nous examinons à présent l'existence de l'EMV dans un espace des CM finies définies sur $(\Omega, \mathcal{F}, \mathbb{P}_p)$, à espace d'état $E = \{1, 2, \dots, s\}$, une loi initiale $\alpha(\cdot)$ et une matrice de transition P . Nous décomposons la fonction de vraisemblance correspondante à une observation X_0^m d'une CM en s parties indépendantes qui sont équivalentes au sens de maximisation à s lois multinomiales indépendantes, dont l'ordre dépend du nombre des transitions possibles⁴ pour chaque état.

Afin d'étendre et de justifier dans le cadre de l'EMV la procédure d'estimation jusqu'à l'instant où tous les états et toutes les transitions possibles sont observées, nous utilisons les conventions suivantes. Le nombre d'états pour la CM et le nombre des transitions possibles pour chaque état vont être considérés comme des paramètres bornés par s . Dans ce cas, la fonction de vraisemblance est bien définie partout à l'exception des instants où un nouvel état est enregistré par le système.

Notons que pour le moment, nous ne faisons aucune hypothèse concernant la structure de P .

⁴Les transitions possibles sont celles avec une probabilité strictement positive.

Nous associons au modèle markovien l'espace paramétrique naturel suivant

$$\mathcal{A} \times \mathcal{P} := \mathcal{A} \times \mathcal{P}_1 \times \cdots \times \mathcal{P}_s = \mathcal{P}_{\mathcal{M}}^{s+1} \subset \mathbb{R}^{s^2+s}, \quad (2.14)$$

où \mathcal{A} et \mathcal{P}_i sont des copies de $\mathcal{P}_{\mathcal{M}}$ défini par (2.6) pour le cas multinomial, et correspondent respectivement au domaine de variation du vecteur initial $\boldsymbol{\alpha}$ et du vecteur formé par la i -ème ligne de la matrice P , pour $1 \leq i \leq s$. De plus, nous identifions P avec le paramètre vectoriel $\boldsymbol{p} \in \mathcal{P}$. Nous notons que les composantes p_{ij} de \boldsymbol{p} sont ordonnées selon l'ordre lexicographique de E^2 .

Dans le cadre d'une observation \mathbf{X}_0^m d'une seule trajectoire d'une CM, censurée à l'instant m , où la loi initiale ne dépend pas du paramètre \boldsymbol{p} , l'information pour le vecteur initial est concentrée dans la première observation X_0 , et par conséquent, pour tout $m \in \mathbb{N}$,

$$\widehat{\boldsymbol{\alpha}}(m) = (\delta_{X_0 i})_{i \in E}, \quad (2.15)$$

où δ_{ij} est le delta de Kronecker.

Remarque 2.1 *Pour une observation \mathbf{X}_0^m où le vecteur initial $\boldsymbol{\alpha}$ d'une CM ne dépend pas du paramètre \boldsymbol{p} , les fonctions de vraisemblance $L_m(\boldsymbol{\alpha}, \boldsymbol{p})$ et log-vraisemblance $\ell_m(\boldsymbol{\alpha}, \boldsymbol{p})$ associées vont être exprimées en fonction de \boldsymbol{p} uniquement, puisque l'estimateur du vecteur initial est trivialement donnée par (2.15).*

Afin d'exprimer l'EMV $\widehat{\boldsymbol{p}}(m)$ du paramètre \boldsymbol{p} , nous définissons pour tout $i \in E$ le processus de comptage $\mathbf{N}_i := (N_i(m))_{m \in \mathbb{N}^*}$, où $N_i(m)$ indique le nombre des visites dans l'état $i \in E$, jusqu'à l'instant $m \in \mathbb{N}$ en négligeant le dernier état observé X_m :

$$N_i(m) := \sum_{n=0}^{m-1} \mathbb{1}_{\{X_n=i\}}, \quad (2.16)$$

ainsi que le processus de comptage $\mathbf{N}_{ij} := (N_{ij}(m))_{m \in \mathbb{N}^*}$, où $N_{ij}(m)$ indique le nombre des transitions de type $(i, j) \in E^2$:

$$N_{ij}(m) := \sum_{k=1}^m \mathbb{1}_{\{X_{k-1}=i, X_k=j\}}. \quad (2.17)$$

Par (2.16) et (2.17), nous avons pour tout $m \in \mathbb{N}^*$,

$$N_i(m) = \sum_{j \in E} N_{ij}(m). \quad (2.18)$$

La v.a. $N_i(m)$ par (2.18) exprime aussi le nombre des transitions de \mathbf{X} qui ont commencé par un état $i \in E$ jusqu'à l'instant $m \in \mathbb{N}^*$. Également, lorsque les instants d'observation successifs sont temporellement équidistants, il peut être interprété comme étant le temps de séjour dans l'état $i \in E$ jusqu'à l'instant $m \in \mathbb{N}^*$.

Soit

$$M_{(s)} := \inf\{m \in \mathbb{N}^* : N_i(m) > 0, \forall i \in E\}. \quad (2.19)$$

Après ce temps aléatoire, tous les états du système et au moins une transition de chaque état vers un autre état ont été enregistrés.

Proposition 2.2 Soit $\mathbf{X} := (X_n)_{n \in \mathbb{N}}$ une CM homogène définie sur $(\Omega, \mathcal{F}, \mathbb{P}_{\mathbf{p}})$, où cette famille des espaces probabilisés est paramétrisée par $\mathbf{p} \in \mathcal{P}$ donné par (2.14) à valeurs dans un espace d'état $E = \{1, 2, \dots, s\}$, de σ -algèbre l'ensemble des parties de E . L'EMV $\hat{\mathbf{p}}(m)$ qui correspond à une observation \mathbf{X}_0^m de \mathbf{X} censurée à l'instant m , est donnée par ses composantes pour $(i, j) \in E^2$, par

$$\hat{p}_{ij}(m) = \begin{cases} \frac{N_{ij}(m)}{N_i(m)} & \text{si } N_i(m) > 0, \\ 0 & \text{si } N_i(m) = 0. \end{cases} \quad (2.20)$$

Démonstration. D'abord, nous supposons que $m \geq M_{(s)}$.

En utilisant la propriété markovienne (2.1), l'expression (2.17) et la remarque 2.1, nous pouvons écrire la fonction de vraisemblance correspondante à l'observation \mathbf{X}_0^m comme suit

$$\mathcal{L}_m(\mathbf{p}) = \prod_{i,j \in E} p_{ij}^{N_{ij}(m)}, \quad (2.21)$$

Cela implique que la log-vraisemblance associée est donnée par

$$\ell_m(\mathbf{p}) = \sum_{i \in E} \ell_{i,m}(\mathbf{p}_i) - C, \quad (2.22)$$

où $\ell_{i,m}(\mathbf{p}_i)$ représente la log-vraisemblance d'une observation $\mathcal{M}(N_i(m), \mathbf{p}_i)$ avec $\mathbf{p}_i \in \mathcal{P}_i$ pour $i \in E$, et la constante C ne dépend pas du paramètre \mathbf{p} (voir (2.11)). Puisque les vecteurs \mathbf{p}_i sont indépendants, la relation (2.20) résulte donc de la proposition 2.1.

Dans le cas où $m < M_{(s)}$, à l'exception des instants où un nouvel état est enregistré par le système, il existe un nombre $s - r > 0$ d'états pour lesquels $N_i(m) = 0$ et également $N_{ij}(m) = 0$. Par notre convention d'identifier \mathbf{p}_i au vecteur nul pour $N_i(m) = 0$, nous obtenons une solution de type EMV pour un espace d'état de \mathbf{X} réduit à un nombre d'états r . \square

Remarque 2.2 *i) L'EMV $\hat{\mathbf{p}}(m)$ donné par (2.20) peut être un estimateur « fiable » pour les probabilités de transition entre les différents états de E pour une famille de CM ergodiques, puisque le nombre de visites pour chaque état $i \in E$ est théoriquement infini (avec probabilité 1) pour une réalisation quelconque de \mathbf{X} . L'information accumulée n'est pas suffisante dans le cas d'états transitoires.*

ii) Pour une discussion plus détaillée sur la justification du terme EMV dans le cas où les états du système n'ont pas tous encore été visités voir l'annexe A.3.

Nous considérons maintenant la classe des CM ergodiques, pour laquelle nous étudions, dans les propositions ci-dessous, le comportement asymptotique de l'EMV. Pour cela, nous aurons besoin de l'hypothèse suivante :

Hypothèse 2.1 *La CM \mathbf{X} finie sous $\mathbb{P}_{\mathbf{p}}$ est ergodique (irréductible)⁵.*

⁵ Si \mathbf{X} est une CM finie, alors les expressions \mathbf{X} irréductible et \mathbf{X} ergodique sont identiques.

Soit ⁶

$$E_+^2 := \{(i, j) \in E^2 : p_{ij} > 0\} \text{ et } d := |E_+^2|, \quad (2.23)$$

$$E_+^{(i)} := \{j \in E : (i, j) \in E_+^2\} \text{ et } d_i := |E_+^{(i)}|, \quad i \in E. \quad (2.24)$$

Il est clair de vérifier le lemme suivant :

Lemme 2.1 *Sous l'hypothèse 2.1, le processus bidimensionnel $\bar{X} := (X_n, X_{n+1})_{n \in \mathbb{N}}$ est une chaîne de Markov ergodique à un espace d'état E_+^2 , une matrice de transition*

$$\bar{P} := (p_{(i,j)(k,l)})_{(i,j),(k,l) \in E_+^2} \quad \text{où } p_{(i,j)(k,l)} = \delta_{jk} p_{kl}, \quad (2.25)$$

et un vecteur stationnaire $\bar{\pi}$, où ses composantes sont données par

$$\pi_{i,j} = \pi_i p_{ij}. \quad (2.26)$$

Lemme 2.2 *Sous l'hypothèse 2.1, quand $m \rightarrow \infty$, nous avons :*

$$i) \quad \frac{N_{ij}(m)}{m} \xrightarrow{\mathbb{P}_p - p.s.} \pi_i p_{ij}, \quad (2.27)$$

$$ii) \quad \frac{N_i(m)}{m} \xrightarrow{\mathbb{P}_p - p.s.} \pi_i. \quad (2.28)$$

Démonstration. En appliquant le théorème ergodique pour la chaîne de Markov \bar{X} et la fonction $f = \mathbb{1}_{i,j} (= \mathbb{1}_{\{(i,j)\}})$, nous avons par (2.17),

$$\frac{N_{ij}(m)}{m} \xrightarrow{\mathbb{P}_p - p.s.} \mathbb{E}_{\bar{\pi}} \mathbb{1}_{\{\bar{X}=(i,j)\}}, \quad (2.29)$$

où $\bar{\pi}(\cdot)$ est la loi stationnaire de \bar{X} . Par la relation (2.26) du lemme 2.1, (2.29) implique (2.27).

La partie ii) est une implication directe de la partie i) et de (2.18). \square

Proposition 2.3 *Sous l'hypothèse 2.1, l'EMV $\hat{p}(m)$ du paramètre p est un estimateur fortement convergent, i.e., quand $m \rightarrow \infty$*

$$\hat{p}(m) \xrightarrow{\mathbb{P}_p - p.s.} p. \quad (2.30)$$

Une démonstration de cette proposition peut être faite au moyen du théorème ergodique.

Nous notons $\xi(m)$ le vecteur dont les composantes sont définies pour $(i, j) \in E_+^2$ par

$$\xi_{ij}(m) = \sqrt{m}(\hat{p}_{ij}(m) - p_{ij}). \quad (2.31)$$

Proposition 2.4 *(Billingsley (1961a)) Sous l'hypothèse 2.1, l'EMV $\hat{p}(m)$ du paramètre p est asymptotiquement normal, i.e., quand $m \rightarrow \infty$*

$$\xi(m) \xrightarrow{p} \xi \sim \mathcal{N}(\mathbf{0}, \Gamma), \quad (2.32)$$

⁶Tous les ensembles considérés ici dépendent de la vraie valeur du paramètre p .

où Γ est une matrice de covariance de dimension $d \times d$, donnée par

$$\Gamma = \begin{pmatrix} \frac{1}{\pi_1} \Lambda_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \frac{1}{\pi_2} \Lambda_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \frac{1}{\pi_s} \Lambda_s \end{pmatrix}, \quad (2.33)$$

et les matrices Λ_i , pour $i \in E$, de dimension $d_i \times d_i$ sont données par

$$\Lambda_i = (p_{ij}[\delta_{jl} - p_{il}])_{j,l \in E_+^{(i)}}. \quad (2.34)$$

- Remarque 2.3**
- i) Une démonstration de cette proposition peut être accomplie en utilisant le TLC pour les CM ou le TLC pour les Martingales,
 - ii) Sadek et Limnios (2002) ont utilisé la forme (2.33) de la matrice de covariance asymptotique de l'EMV pour démontrer des propriétés asymptotiques pour l'EMV des taux de défaillance et de fiabilité pour les CM qui sont ergodiques et apériodiques,
 - iii) dans Trevezas et Limnios (à paraître 2008b) nous explicitons que les propriétés asymptotiques pour l'EMV sont valides également pour les CM périodiques.

2.3 ESTIMATION DE LA LOI STATIONNAIRE

Afin de trouver l'EMV de la loi stationnaire qui vérifie de manière unique (2.5) dans le cas où la CM X est ergodique, nous allons exprimer ce vecteur comme une fonction de la matrice P .

L'EMV qui correspond à la matrice P s'écrit pour tout $m \geq M_{(s)}$, de la façon suivante⁷ :

$$\hat{P}(m) = (\hat{p}_{ij}(m))_{i,j \in E}. \quad (2.35)$$

Cela va nous permettre d'avoir une forme matricielle de l'EMV qui correspond à la loi stationnaire d'une CM ergodique.

Dans l'analyse qui suit, nous considérons que la vraie valeur du paramètre \boldsymbol{p} est telle que P a tous ses éléments strictement positifs de manière à assurer que \boldsymbol{p} est un point intérieur d'un espace paramétrique approprié, qui est partie de \mathbb{R}^{s^2-s} et qui sera défini dans la suite.

Les résultats qui suivent sont valables même dans le cas où une représentation minimale des paramètres implique un nombre d strictement inférieur à $s^2 - s$. Dans ce cas, si d est estimé alors nous modifions la dimension de l'espace paramétrique en ajustant les dimensions des vecteurs et des matrices qui apparaissent.

Nous posons $\boldsymbol{\Pi} = \mathbf{1}^\top \mathbf{1}$ la matrice carrée dont chaque élément vaut un. Nous avons donc

$$\boldsymbol{\pi} \boldsymbol{\Pi} = \mathbf{1}. \quad (2.36)$$

Nous remarquons par (2.5) et (2.36) que

$$\boldsymbol{\pi}(I - P + \boldsymbol{\Pi}) = \boldsymbol{\pi}(I - P) + \boldsymbol{\pi} \boldsymbol{\Pi} = \mathbf{0} + \mathbf{1} = \mathbf{1}. \quad (2.37)$$

⁷Pour une définition de $M_{(s)}$ voir (2.19).

Afin d'obtenir π , la matrice $I - P + \mathbb{I}$ doit être non singulière. Ce qui est toujours vérifié s'il s'agit d'une CM ergodique. Pour une démonstration voir par exemple Resnick (1992) [proposition 2.14.1.]. Cette condition est suffisante mais pas nécessaire. Dans le lemme suivant nous donnons une caractérisation des CM ayant une classe récurrente avec la possibilité d'avoir quelques états transitoires. Pour une démonstration voir l'annexe A.2.

Lemme 2.3 *Soit P la matrice de transition d'une CM X finie. X admet une seule classe récurrente si et seulement si $I - P + \mathbb{I}$ est non-singulière.*

En utilisant la relation (2.37) et le fait que chaque CM finie ergodique a une seule classe récurrente nous obtenons

$$\pi = \mathbf{1}(I - P + \mathbb{I})^{-1}. \quad (2.38)$$

Dans ce qui suit, nous allons définir et comparer deux estimateurs de la loi stationnaire, l'EMV et l'estimateur empirique (EE).

$$\text{(EMV):} \quad \hat{\pi}(m) := \mathbf{1}(I - \hat{P}(m) + \mathbb{I})^{-1}, \quad m \geq M_{(s)}, \quad (2.39)$$

$$\text{(EE):} \quad \tilde{\pi}(m) := \left(\frac{N_i(m)}{m} \right)_{i \in E}, \quad m \geq 1. \quad (2.40)$$

Afin que l'expression (2.39) soit bien définie, il faut que la matrice $I - \hat{P}(m) + \mathbb{I}$ soit non-singulière. Pour tout $m \geq M_{(s)}$, la matrice $\hat{P}(m)$ est stochastique et correspond à une CM à un espace d'état E , ayant une classe récurrente et la possibilité d'avoir des états transitoires. D'après le lemme 2.3, ceci implique que la matrice $I - \hat{P}(m) + \mathbb{I}$ est non-singulière. Pour une discussion concernant la justification de l'utilisation de la matrice $I - \hat{P}(m) + \mathbb{I}$ pour des valeurs de $m < M_{(s)}$ et son comportement, voir l'annexe A.3.

En ce qui concerne l'EE, l'expression (2.40) est valable pour tout $m \geq 1$. Pour chaque instant m , l'EE $\tilde{\pi}(m)$ exprime le pourcentage de temps durant lequel le système se trouve dans les différents états de E .

Nous donnons dans la proposition suivante la relation exacte qui lie les deux estimateurs pour chaque $m \geq M_{(s)}$. Pour cela, nous définissons d'abord le processus de comptage du nombre des visites aux différents états $j \in E$, dans le cas où nous ne prenons pas en compte l'état initial

$$N_j^*(m) = \sum_{n=1}^m \mathbf{1}_{\{X_n=j\}} = N_j(m) + \delta_{jX_m} - \delta_{jX_0}. \quad (2.41)$$

Ce processus peut aussi exprimer le nombre des transitions vers un état $j \in E$ jusqu'à l'instant $m \in \mathbb{N}$, puisque

$$N_j^*(m) = \sum_{i=1}^s N_{ij}(m). \quad (2.42)$$

Notons $(e_i)_{i \in E}$ les vecteurs de la base canonique de \mathbb{R}^s .

Proposition 2.5 *L'estimateur empirique $\tilde{\pi}(m)$ et l'estimateur du maximum de vraisemblance $\hat{\pi}(m)$ de la loi stationnaire π sont liés par la relation suivante :*

$$\tilde{\pi}(m) = \hat{\pi}(m) - \Delta(m)(I - \hat{P}(m) + \mathbb{I})^{-1}, \quad (2.43)$$

où $\Delta(m) = (\Delta_i(m))_{i \in E}$, et $\Delta_i(m) = (\delta_{iX_m} - \delta_{iX_0})/m$, $i \in E$.

Démonstration. Nous remarquons que

$$\begin{aligned} \tilde{\pi}(m)\hat{P}(m) &= \left(\frac{N_1(m)}{m}, \frac{N_2(m)}{m}, \dots, \frac{N_s(m)}{m} \right) \begin{pmatrix} \frac{N_{11}(m)}{N_1(m)} & \dots & \frac{N_{1s}(m)}{N_1(m)} \\ \frac{N_{21}(m)}{N_2(m)} & \dots & \frac{N_{2s}(m)}{N_2(m)} \\ \vdots & \ddots & \vdots \\ \frac{N_{s1}(m)}{N_s(m)} & \dots & \frac{N_{ss}(m)}{N_s(m)} \end{pmatrix}, \\ &= \left(\frac{\sum_{i=1}^s N_{i1}(m)}{m}, \frac{\sum_{i=1}^s N_{i2}(m)}{m}, \dots, \frac{\sum_{i=1}^s N_{is}(m)}{m} \right). \end{aligned} \quad (2.44)$$

Par (2.41), (2.42), et (2.44) nous concluons que

$$\tilde{\pi}(m)\hat{P}(m) = \tilde{\pi}(m) + \Delta(m). \quad (2.45)$$

Par (2.45) nous obtenons

$$\tilde{\pi}(m)(I - \hat{P}(m) + \mathbb{I}) = \mathbf{1} - \Delta(m). \quad (2.46)$$

Puisque pour $m \geq M_{(s)}$, donné par (2.19), la matrice $I - \hat{P}(m) + \mathbb{I}$ est non-singulière, (2.46) implique alors (2.43). \square

Remarque 2.4 *Pour chaque $m \geq M_{(s)}$, le vecteur $\hat{\pi}(m)$ (EMV) est l'unique vecteur stationnaire pour la matrice $\hat{P}(m)$. Par (2.43) le vecteur (EE) $\tilde{\pi}(m)$ joue ce rôle ssi $\Delta(m) = \mathbf{0}$. Puisque cela est vrai ssi $X_m = X_0$, la sous-suite des temps de passage successifs dans l'état initial X_0 , détermine alors les instants où les deux estimateurs sont identiques.*

La proposition suivante établit l'équivalence asymptotique par rapport à la loi des grands nombres de l'EE et de l'EMV de la loi stationnaire.

Proposition 2.6 *Sous l'hypothèse 2.1, l'estimateur empirique $\tilde{\pi}(m)$ et l'estimateur du maximum de vraisemblance $\hat{\pi}(m)$ du vecteur stationnaire π sont des estimateurs fortement convergents, i.e., quand $m \rightarrow \infty$*

$$i) \quad \tilde{\pi}(m) \xrightarrow{\mathbb{P}_p\text{-p.s.}} \pi, \quad (2.47)$$

$$ii) \quad \hat{\pi}(m) \xrightarrow{\mathbb{P}_p\text{-p.s.}} \pi. \quad (2.48)$$

et plus particulièrement, nous avons

$$\hat{\pi}(m) = \tilde{\pi}(m) + O_{\mathbb{P}}(m^{-1})^8. \quad (2.49)$$

⁸ $X_m = O_{\mathbb{P}}(R_m)$ ssi $X_m = R_m Y_m$ avec Y_m bornée en probabilité (\mathbb{P} -bornée).

Démonstration. i) Cela résulte du fait que les composantes de $\tilde{\pi}(m)$, donné par (2.40) vérifient (2.28).

ii) Par (2.35) et (2.20) nous avons dans une forme matricielle que

$$\widehat{P}(m) \xrightarrow{\mathbb{P}_{p-p.s.}} P. \quad (2.50)$$

En considérant la suite des matrices $\{(I - \widehat{P}(m) + \mathbb{I})^{-1}\}_{m \geq M(s)}$ et puisque l'application $A \mapsto A^{-1}$ est une application continue par rapport à une norme des matrices appropriée, nous concluons par (2.50) et par le théorème de l'application continue⁹ que

$$(I - \widehat{P}(m) + \mathbb{I})^{-1} \xrightarrow{\mathbb{P}_{p-p.s.}} (I - P + \mathbb{I})^{-1}. \quad (2.51)$$

En multipliant à gauche la partie gauche de (2.51) par le vecteur $\mathbf{1}$, vu (2.51), (2.39) et (2.38) nous concluons la partie ii).

Également, par (2.51) et la définition de Δ_m dans la proposition 2.5 nous avons

$$\Delta_m(I - \widehat{P}(m) + \mathbb{I})^{-1} = O_{\mathbb{P}}(m^{-1}), \quad (2.52)$$

ce qui conduit à l'aide de (2.43) à (2.49). \square

L'étude du comportement asymptotique de l'EMV $\widehat{p}(m)$ du paramètre p , ainsi que d'autres caractéristiques de la CM (loi stationnaire, variance dans un TLC) qui peuvent être exprimées comme une fonction de la matrice de transition P , nous oblige à considérer une représentation minimale spécifique pour le paramètre vectoriel p . Notons

$$\mathcal{P}^* := \prod_{i \in E} \mathcal{P}_{\mathcal{M}_i}^* \subset \mathbb{R}^{d-s}, \quad (2.53)$$

où $\mathcal{P}_{\mathcal{M}_i}^*$, pour tout $i \in E$, sont données par (2.7), et

$$E_*^2 := \{(i, j) \in E^2 : i \neq j\}. \quad (2.54)$$

Dans la notation ci-dessous nous précisons le type de représentation minimale que nous considérons.

Notation 2.1 *Le paramètre vectoriel $p \in \mathcal{P}$ a une représentation minimale p^* indexée par E_*^2 , s'il est considéré comme élément de \mathcal{P}^* avec $d = s^2$ en lui enlevant les composantes qui correspondent aux éléments diagonaux de P . Les composantes p_{ij} de $p^* \in \mathcal{P}^*$ sont ordonnées selon l'ordre lexicographique de E_*^2 .*

Dans l'hypothèse suivante, nous allons considérer que sous \mathbb{P}_p toutes les transitions sont possibles, c'est à dire $d = s^2$, où d est défini par (2.23).

Hypothèse 2.2 *Le paramètre p^* , donné par la notation 2.1, sous \mathbb{P}_p , vérifie $d = s^2$.*

L'hypothèse 2.2 est plus restrictive que l'hypothèse 2.1, mais elle nous sera utile pour les résultats qui suivront, en particulier pour les représentations matricielles. Les résultats théoriques (convergence et normalité asymptotique) sont valables également sous l'hypothèse 2.1.

⁹En anglais : continuous mapping theorem.

Notons que les éléments diagonaux de P seront considérés comme des fonctions des autres paramètres correspondants à la même ligne de la matrice P . Ainsi, les EMV qui correspondent aux s éléments diagonaux (s est le nombre d'états) donnent

$$\widehat{p}_{ii}(m) = 1 - \sum_{j \in E_i} \widehat{p}_{ij}(m), \quad (2.55)$$

où $E_i := E \setminus \{i\}$.

Nous considérons utile de faire ce choix de représentation minimale, indiqué ci-dessus, même dans le cas où $d < s^2$, à condition que p_{ii} soit > 0 pour tout $i \in E$. Si au moins une auto-transition (transition vers le même état) n'est pas possible, alors le choix doit être différent pour pouvoir exclure s paramètres dépendants. Les relations de dépendance sont imposées toujours par la nature stochastique de la matrice P . Par exemple, voir chapitre 4, où les paramètres diagonaux sont identiquement nuls par la nature de ce modèle. Un autre exemple peut être la structure de la matrice de transition d'une CM immergée d'une chaîne semi-markovienne (CSM).

Pour être compatible avec la notation 2.1, nous définissons ζ^* une représentation minimale de ζ indexée par E_*^2 , où ζ est le vecteur de la loi normale définie par (2.32). D'une façon similaire, $\widehat{p}^*(m)$ et $\zeta^*(m)$ correspondent aux représentations minimales de $\widehat{p}(m)$ et $\zeta(m)$ respectivement. Nous notons Γ^* la matrice de covariance de dimension $(s^2 - s) \times (s^2 - s)$ qui correspond à ζ^* , où dans l'expression (2.33), pour tout $i \in E$, les matrices Λ_i données par (2.34) sont substitués par Λ_i^* , en leur ôtant la i -ème ligne et la i -ème colonne.

Nous définissons pour $(i, j) \in E_*^2$, la matrice $\partial P / \partial p_{ij} := (\partial p_{kl} / \partial p_{ij})_{k,l \in E}$. Par (2.1) nous avons

$$\frac{\partial p_{kl}}{\partial p_{ij}} = \begin{cases} 1, & \text{si } (k, l) = (i, j), \\ -1, & \text{si } (k, l) = (i, i), \\ 0, & \text{sinon.} \end{cases} \quad (2.56)$$

Proposition 2.7 *Sous l'hypothèse 2.2, l'EMV $\widehat{\pi}(m)$ du vecteur stationnaire π est asymptotiquement normal, i.e., quand $m \rightarrow \infty$*

$$\sqrt{m}(\widehat{\pi}(m) - \pi) \overset{P}{\rightsquigarrow} \mathcal{N}(\mathbf{0}, G \Gamma^* G^\top), \quad (2.57)$$

où la matrice $G := (G_{u;i,j})_{u \in E, (i,j) \in E_*^2}$ de dimension $s \times (s^2 - s)$ est définie par

$$G_{u;i,j} = \pi_j (I - P + \Pi)_{ju}^{-1} - \pi_i (I - P + \Pi)_{iu}^{-1}, u \in E, (i, j) \in E_*^2, \quad (2.58)$$

et Γ^* est une représentation minimale (indexée par E_*^2) de la matrice Γ définie par (2.33).

Démonstration. Soit $g : (\mathcal{P}^*)^\circ \rightarrow \mathbb{R}^s$, où

$$g(\mathbf{p}^*) := \mathbf{1}(I - P + \Pi)^{-1}. \quad (2.59)$$

Si $A_{.u}$ correspond à la u -ème colonne de la matrice A , et $g(\mathbf{p}^*) = (g_u(\mathbf{p}^*))_{u \in E}$, alors ses composantes $g_u(\mathbf{p}^*)$ pour $u \in E$ sont données par

$$g_u(\mathbf{p}^*) = \mathbf{1}(I - P + \Pi)_{.u}^{-1}. \quad (2.60)$$

Par les relations (2.39) et (2.59) nous avons

$$\begin{aligned}\sqrt{m}(\widehat{\pi}(m) - \pi) &= \sqrt{m}[\mathbf{1}(I - \widehat{P}(m) + \mathbb{I})^{-1} - \mathbf{1}(I - P + \mathbb{I})^{-1}] \\ &= \sqrt{m}[g(\widehat{\mathbf{p}}^*(m)) - g(\mathbf{p}^*)].\end{aligned}$$

Puisque par (2.32), $\sqrt{m}(\widehat{\mathbf{p}}(m) - \mathbf{p}) \overset{P}{\rightsquigarrow} \mathcal{N}(\mathbf{0}, \Gamma)$, cela implique que

$$\sqrt{m}(\widehat{\mathbf{p}}^*(m) - \mathbf{p}^*) \overset{P}{\rightsquigarrow} \mathcal{N}(\mathbf{0}, \Gamma^*), \quad (2.61)$$

comme loi marginale d'une loi asymptotiquement normale. Pour que nous puissions appliquer la méthode delta pour la fonction g , nous devons garantir sa différentiabilité sur $(\mathcal{P}^*)^\circ$. Par (2.59), il suffit que l'application $\mathbf{p}^* \mapsto (I - P + \mathbb{I})^{-1}$ soit différentiable sur $(\mathcal{P}^*)^\circ$. Ce qui est vérifié puisque chaque élément (i, j) de la matrice $(I - P + \mathbb{I})^{-1}$ est un rapport des polynômes de \mathbf{p}^* , où le numérateur est le (i, j) cofacteur et le dénominateur est le déterminant de la matrice $I - P + \mathbb{I}$.

Les valeurs exactes des dérivées partielles de g_u par rapport à p_{ij} pour $u \in E$ et $(i, j) \in E_*^2$, peuvent être exprimées comme suit

$$\begin{aligned}\frac{\partial g_u}{\partial p_{ij}} &= \frac{\partial \mathbf{1}(I - P + \mathbb{I})^{-1}.u}{\partial p_{ij}} = \frac{\partial \mathbf{1}(I - P + \mathbb{I})^{-1} \mathbf{e}_u^\top}{\partial p_{ij}} \\ &= \mathbf{1}(I - P + \mathbb{I})^{-1} \frac{\partial P}{\partial p_{ij}} (I - P + \mathbb{I})^{-1} \mathbf{e}_u^\top \\ &= \pi \frac{\partial P}{\partial p_{ij}} (I - P + \mathbb{I})^{-1} \mathbf{e}_u^\top = \pi \frac{\partial P}{\partial p_{ij}} (I - P + \mathbb{I})^{-1}. \quad (2.62)\end{aligned}$$

Nous désignons par G la matrice de dimension $s \times (s^2 - s)$ dont les éléments sont donnés par (2.62). Par (2.61) et en appliquant la méthode delta pour la fonction différentiable g nous arrivons à l'expression (2.57) tout en tenant compte de (2.62) et de la forme de $\partial P / \partial p_{ij}$ donnée par (2.56). \square

Corollaire 2.1 *Sous l'hypothèse 2.2, l'estimateur empirique $\widetilde{\pi}(m)$ du vecteur stationnaire π est asymptotiquement normal, i.e.,*

$$\sqrt{m}(\widetilde{\pi}(m) - \pi) \overset{P}{\rightsquigarrow} \mathcal{N}(\mathbf{0}, G \Gamma^* G^\top). \quad (2.63)$$

Démonstration. Par les relations (2.49) et (2.43) nous obtenons

$$\sqrt{m}(\widetilde{\pi}(m) - \pi) = \sqrt{m}(\widehat{\pi}(m) - \pi) - \sqrt{m}O_{\mathbb{P}}(m^{-1}) \quad (2.64)$$

$$= \sqrt{m}(\widehat{\pi}(m) - \pi) + O_{\mathbb{P}}(m^{-1/2}), \quad (2.65)$$

ce qui implique que l'estimateur empirique du vecteur stationnaire converge en loi, quand $m \rightarrow \infty$, vers la même limite que l'EMV, donnée par la proposition 2.7. \square

2.4 REPRÉSENTATIONS DE LA VARIANCE DANS LE TLC

Dans cette section, nous présentons des formes de la variance dans le TLC pour les CM finies. Ces formes vont servir à construire des formules

calculables pour la variance.

D'abord, nous présentons le TLC pour les CM à un espace d'état dénombrable. Pour cela, nous aurons besoin de définir les quantités suivantes.

La probabilité ci-dessous est appelée probabilité tabou,

$${}_a p_{ij}^{(n)} := \mathbb{P}_i(T_a \geq n, X_n = j). \quad (2.66)$$

Cette quantité exprime la probabilité que la chaîne en commençant par l'état initial $i \in E$ va être dans l'état $j \in E$ à l'instant n en évitant l'état a .

Les quantités ci-dessous qui dépendent des probabilités tabou (2.66) vont jouer un rôle important dans notre analyse :

$$g_a(i, j) := \sum_{n=1}^{+\infty} {}_a p_{ij}^{(n)}. \quad (2.67)$$

Pour donner une interprétation probabiliste à (2.67) nous définissons N_j^a la v.a. qui représente le nombre de passages de la chaîne dans l'état j en ne tenant pas compte de l'état initial, avant qu'elle visite pour la première fois l'état a , i.e.,

$$N_j^a := \sum_{n=1}^{+\infty} \mathbb{1}_{\{T_a \geq n, X_n = j\}}. \quad (2.68)$$

Par les relations (2.66) et (2.67) nous avons

$$\begin{aligned} g_a(i, j) &= \sum_{n=1}^{+\infty} \mathbb{E}_i(\mathbb{1}_{\{T_a \geq n, X_n = j\}}) \\ &= \mathbb{E}_i\left(\sum_{n=1}^{+\infty} \mathbb{1}_{\{T_a \geq n, X_n = j\}}\right) \\ &= \mathbb{E}_i(N_j^a). \end{aligned} \quad (2.69)$$

Cela exprime que $g_a(i, j)$ correspond à l'espérance du nombre de passages par l'état j , en démarrant par l'état i , avant que la chaîne entre pour la première fois dans l'état a .

Le temps de retour T_a et les v.a. N_j^a sont liés par

$$T_a = \sum_{j \in E} N_j^a. \quad (2.70)$$

Soit f une fonction réelle définie sur E . Nous supposons que

$$\mathbb{E}_\pi |f(X)| = \sum_{i \in E} \pi_i |f(i)| < \infty, \quad (2.71)$$

et nous désignons par

$$\mu := \mathbb{E}_\pi f(X) = \sum_{i \in E} \pi_i f(i), \quad (2.72)$$

et par \bar{f} la fonction f centrée par rapport à la loi stationnaire, i.e., $\bar{f} = f - \mu$.

Théorème 2.2 (Port (1994)) Soit X une chaîne de Markov ergodique. On suppose qu'il existe un état $a \in E$, tel que

$$\sum_{i,j \in E} \pi_i g_a(i,j) |\bar{f}(i)| |\bar{f}(j)| < \infty. \quad (2.73)$$

Si $(S_n)_{n \in \mathbb{N}}$ correspond à la suite des sommes partielles de $(f(X_n))_{n \in \mathbb{N}}$, alors pour une loi initiale quelconque, quand $m \rightarrow \infty$

$$\frac{1}{\sqrt{m}}(S_m - m\mu) \rightsquigarrow \mathcal{N}(0, \sigma^2), \quad (2.74)$$

où

$$\sigma^2 = \sum_{i \in E} \pi(i) \bar{f}(i)^2 + 2 \sum_{i,j \in E-a} \pi_i g_a(i,j) \bar{f}(i) \bar{f}(j). \quad (2.75)$$

Remarque 2.5 i) Nous notons que si la condition (2.73) est vérifiée par un état $a \in E$, alors elle est valide pour chaque état de E . Également, puisque l'expression (2.75) correspond à la variance du théorème limite (2.74), donc elle ne dépend pas du choix de $a \in E$.

ii) Dans le cas d'une CM finie, (2.73) est trivialement vérifiée.

iii) Dans la littérature, la forme la plus souvent utilisée de la variance (voir Doob (1953), Lemme 7.3 ou Billingsley (1961b), remarques après le Théorème 3.3) est donnée comme suit :

$$\sigma^2 = \mathbb{E}_\pi \left(\bar{f}(X_0)^2 \right) + 2 \sum_{k=1}^{+\infty} \mathbb{E}_\pi \left(\bar{f}(X_0) \bar{f}(X_k) \right). \quad (2.76)$$

Maintenant, nous considérons le cas où la CM X est finie et nous analysons la première forme de la variance qui résulte de l'expression (2.75). Pour cette raison nous allons exprimer les quantités $g_a(i,j)$, données par (2.67), en fonction de la matrice de transition P .

Nous désignons par P_a la matrice d'ordre $s-1$ qui résulte de la matrice P en ôtant la a -ème ligne et la a -ème colonne de la matrice P . Également, P'_a désigne la matrice de dimension $s \times (s-1)$ qui résulte de la matrice P en ôtant seulement la a -ème colonne. Nous désignons par B_a la matrice de dimension $s \times (s-1)$ qui résulte de la matrice identité en lui introduisant une ligne nulle entre la ligne $(a-1)$ et la ligne a . Alors,

$$P_a = B_a^\top P B_a \quad \text{et} \quad P'_a = P B_a. \quad (2.77)$$

Proposition 2.8 Soit $g_a(i,j)$ ($= \mathbb{E}_i(N_j^a)$) défini par (2.67). Alors,

$$g_a(i,j) = \begin{cases} e_i P'_a (I - P_a)^{-1} e_j^\top & \text{si } j \neq a, \\ 1 & \text{si } j = a, \end{cases} \quad (2.78)$$

où $j' = j$ si $j < a$, et $j' = j-1$ si $j > a$, $i, j \in E$.

Démonstration. Si $j \neq a$,

$$\begin{aligned} \mathbb{P}_i(T_a \geq n, X_n = j) &= \sum_{i_1^{n-1} \in E-a} \mathbb{P}_i(X_1^{n-1} = i_1^{n-1}, X_n = j, T_a \geq n) \\ &= \sum_{i_1^{n-1} \in E-a} \mathbb{P}_i(X_1^{n-1} = i_1^{n-1}, X_n = j) \\ &= \sum_{i_1^{n-1} \in E-a} p_{ii_1} p_{i_1 i_2} \cdots p_{i_{n-1} j} \\ &= e_i P'_a P_a^{n-1} e_j^\top. \end{aligned} \quad (2.79)$$

Par conséquent, par la définition de $g_a(i, j)$ et (2.79)

$$\begin{aligned} g_a(i, j) &= \sum_{n=1}^{+\infty} \mathbb{P}_i(T_a \geq n, X_n = j) = \sum_{n=1}^{+\infty} \mathbf{e}_i P_a' P_a^{n-1} \mathbf{e}_j^\top \\ &= \mathbf{e}_i P_a' \left(\sum_{n=1}^{+\infty} P_a^{n-1} \right) \mathbf{e}_j^\top = \mathbf{e}_i P_a' (I - P_a)^{-1} \mathbf{e}_j^\top. \end{aligned} \quad (2.80)$$

Si $j = a$,

$$\mathbb{P}_i(T_a \geq n, X_n = a) = \mathbb{P}_i(T_a = n).$$

Cela implique que

$$g_a(i, a) = \sum_{n=1}^{\infty} \mathbb{P}_i(T_a = n) = \mathbb{P}_i(T_a < \infty) = 1.$$

□

Remarque 2.6 Nous avons utilisé dans la relation (2.80) le fait que la série $\sum_{n=1}^{+\infty} P_a^{n-1}$ est convergente. Ceci résulte de l'irréductibilité de la matrice P . Ce qui implique que le rayon spectral de la matrice P_a est strictement inférieur à un. Pour une justification de ces arguments et une bonne présentation de l'analyse spectrale des CM voir Seneta (2006).

À présent, nous introduisons les notations suivantes :

$$\boldsymbol{\pi}^\# = (\pi_1 \bar{f}(1), \pi_2 \bar{f}(2), \dots, \pi_s \bar{f}(s)), \quad (2.81)$$

$$D = (d_{ij})_{i,j \in E}, \text{ où } d_{ij} = g_a(i, j), \text{ si } i \neq a \text{ et } d_{a,j} = 0, \quad (2.82)$$

$$\Pi = \mathbf{1}^\top \boldsymbol{\pi}. \quad (2.83)$$

Par (2.83) et (2.38) nous avons

$$\Pi = \Pi (I - P + \Pi)^{-1}. \quad (2.84)$$

La matrice Π correspond à la limite de Cesàro de la suite matricielle $\{P^n\}_{n \geq 0}$, et de plus si nous renforçons l'irréductibilité par la condition supplémentaire d'apériodicité, alors Π est sa matrice limite. Nous désignons par Π_{dg} la matrice qui résulte de Π en prenant tous les éléments non diagonaux de Π nuls.

Par (2.75) et les expressions (2.81), (2.82) et (2.83) nous obtenons une première forme de σ^2 :

$$\sigma^2 = \boldsymbol{\pi}^\# \bar{\mathbf{f}}^\top + 2\boldsymbol{\pi}^\# D \bar{\mathbf{f}}^\top = \boldsymbol{\pi}^\# (I + 2D) \bar{\mathbf{f}}^\top. \quad (2.85)$$

Notons que par (2.81), $\boldsymbol{\pi}^\#$ s'écrit

$$\boldsymbol{\pi}^\# = \bar{\mathbf{f}} \Pi_{dg}. \quad (2.86)$$

La relation (2.86) mène à l'expression suivante de la variance :

$$\sigma^2 = \bar{\mathbf{f}} \Pi_{dg} (I + 2D) \bar{\mathbf{f}}^\top. \quad (2.87)$$

La forme de la variance ci-dessus est une première forme matricielle. Ensuite, nous présenterons une deuxième forme qui résulte de l'expression (2.76). Pour cela nous aurons besoin d'introduire la matrice fondamentale de la matrice P .

Définition 2.5 Soit P une matrice ergodique. Nous définissons sa matrice fondamentale Z par

$$Z = (I - P + \Pi)^{-1}. \quad (2.88)$$

Remarque 2.7 La matrice fondamentale de P est bien définie puisque la matrice $I - P + \Pi$ est inversible. Si P est apériodique cela peut être démontré par le fait que P^n converge vers la matrice Π (voir, par exemple Doob (1953), page 225). Pour une ébauche de la démonstration dans le cas où P est périodique et pour des propriétés non-élémentaires de Z nous renvoyons à Kemeny et Snell (1976).

Nous donnons ci-dessous les propriétés élémentaires de la matrice Z .

Lemme 2.4 La matrice fondamentale Z de P satisfait les propriétés suivantes :

$$i) Z \mathbf{1}^\top = \mathbf{1}^\top,$$

et

$$ii) Z\Pi = \Pi Z = \Pi.$$

Démonstration. i) Nous remarquons que $(I - P + \Pi) \mathbf{1}^\top = \mathbf{1}^\top$ et en multipliant à gauche par $(I - P + \Pi)^{-1}$ nous obtenons la première propriété. ii) De même comme i) en remarquant que la matrice $I - P + \Pi$ satisfait les propriétés souhaitées de la matrice Z . \square

Le lemme suivant peut être démontré par induction.

Lemme 2.5 Pour tout $n \geq 1$,

$$(P - \Pi)^n = P^n - \Pi. \quad (2.89)$$

En analysant la relation (2.76) nous pouvons obtenir une deuxième forme pour σ^2 .

Proposition 2.9 La variance dans le TLC pour une CM finie, donnée par (2.76) peut être représentée sous la forme

$$\sigma^2 = \bar{f} \Pi_{dg} (2Z - I) \bar{f}^\top. \quad (2.90)$$

Démonstration. Nous avons

$$\begin{aligned} \mathbb{E}_\pi \left(\bar{f}(X_0) \bar{f}(X_k) \right) &= \sum_{i,j} \pi_i p_{ij}^{(k)} \bar{f}(i) \bar{f}(j) \\ &= \sum_{i,j} \pi_i (p_{ij}^{(k)} - \pi_j) \bar{f}(i) \bar{f}(j) + \sum_{i,j} \pi_i \pi_j \bar{f}(i) \bar{f}(j) \\ &= \sum_{i,j} \pi_i (p_{ij}^{(k)} - \pi_j) \bar{f}(i) \bar{f}(j) + \left(\sum_i \pi_i \bar{f}(i) \right)^2 \\ &= \sum_{i,j} \pi_i (p_{ij}^{(k)} - \pi_j) \bar{f}(i) \bar{f}(j) \\ &= \sum_{i,j} \pi_i (P^k - \Pi)_{ij} \bar{f}(i) \bar{f}(j). \end{aligned} \quad (2.91)$$

Par le lemme 2.5, la relation (2.91) donne

$$\mathbb{E}_\pi \left(\bar{f}(X_0) \bar{f}(X_k) \right) = \sum_{i,j \in E} \pi_i (P - \Pi)_{ij}^k \bar{f}(i) \bar{f}(j). \quad (2.92)$$

Sous une forme matricielle (2.92) donne

$$\mathbb{E}_\pi \left(\bar{f}(X_0) \bar{f}(X_k) \right) = \bar{f} \Pi_{dg} (P - \Pi)^k \bar{f}^\top. \quad (2.93)$$

De même, nous avons

$$\mathbb{E}_\pi \left(\bar{f}(X_0)^2 \right) = \bar{f} \Pi_{dg} \bar{f}^\top. \quad (2.94)$$

Finalement, par (2.93) et (2.94), la relation (2.76) donne

$$\begin{aligned} \sigma^2 &= \bar{f} \Pi_{dg} \bar{f}^\top + 2 \bar{f} \Pi_{dg} \sum_{k=1}^{+\infty} (P - \Pi)^k \bar{f}^\top \\ &= \bar{f} \Pi_{dg} \bar{f}^\top + 2 \bar{f} \Pi_{dg} \left(\sum_{k=0}^{+\infty} (P - \Pi)^k - I \right) \bar{f}^\top \\ &= \bar{f} \Pi_{dg} \bar{f}^\top + 2 \bar{f} \Pi_{dg} \left((I - P + \Pi)^{-1} - I \right) \bar{f}^\top \\ &= \bar{f} \Pi_{dg} (2Z - I) \bar{f}^\top. \end{aligned}$$

□

La proposition suivante décrit des expressions équivalentes pour la variance dans le TLC des CM.

Proposition 2.10 *Soit X une chaîne de Markov ergodique. Alors, nous avons les expressions équivalentes suivantes pour la variance dans le TLC pour des CM.*

$$\begin{aligned} 1) \quad \sigma^2 &= \bar{f} \Pi_{dg} (2Z - I) \bar{f}^\top = f \Pi_{dg} (2Z - I - \Pi) f^\top. \\ 2) \quad \sigma^2 &= \bar{f} \Pi_{dg} (I + 2D) \bar{f}^\top = f \Pi_{dg} (I - \Pi) (I + 2D) (I - \Pi) f^\top \\ &= \pi^\# (I + 2D) \bar{f}^\top. \end{aligned}$$

Démonstration. Les premiers membres (à gauche) des expressions 1) et 2) sont égaux puisqu'ils représentent les formes matricielles avec lesquelles nous avons exprimé la variance σ^2 dans le TLC aux relations (2.90) et (2.87) respectivement. Le dernier membre de l'expression 2) correspond à la relation (2.85). La connection de \bar{f} avec f est effectuée via l'expression

$$\bar{f} = f(I - \Pi^\top). \quad (2.95)$$

Nous remarquons que $\Pi_{dg} \Pi = (\pi(i) \pi(j))_{i,j \in E}$. Cela implique que $\Pi_{dg} \Pi$ est symétrique et donc

$$(I - \Pi^\top) \Pi_{dg} = \Pi_{dg} (I - \Pi). \quad (2.96)$$

Par (2.95) et (2.96) nous arrivons à l'égalité intermédiaire de l'expression 2). Pour le membre à droite de 1) nous avons

$$\bar{f} \Pi_{dg} (2Z - I) \bar{f}^\top = f \Pi_{dg} (I - \Pi) (2Z - I) (I - \Pi) f^\top. \quad (2.97)$$

En utilisant la propriété ii) de Z dans le lemme 2.4 nous obtenons

$$(I - \Pi) (2Z - I) (I - \Pi) = 2Z - I - \Pi. \quad (2.98)$$

En combinant (2.97) avec (2.98) nous arrivons à l'expression souhaitée. □

Remarque 2.8 Pour lier les expressions données par 1) et 2) dans la proposition 2.10, nous avons utilisé le fait qu'elles proviennent des deux différentes démonstrations du TLC pour les CM. Pour une démonstration directe, nous pouvons utiliser la relation suivante entre la matrice fondamentale Z et la matrice P :

$$Z = \Pi + (I - \Pi)D^*(I - \Pi), \text{ où } D^* = B_a(I_a - P_a)B_a^\top. \quad (2.99)$$

Pour une démonstration de (2.99) voir par exemple Kemeny et Snell (1976). Il est clair qu'à l'aide de (2.99) et du fait que $\Pi \bar{f}^\top = \mathbf{0}^\top$, on peut obtenir une expression pour la variance similaire à ii)

$$\sigma^2 = \bar{f} \Pi_{dg}(2D^* - I) \bar{f}^\top = f \Pi_{dg}(I - \Pi)(2D^* - I)(I - \Pi) f^\top. \quad (2.100)$$

2.5 ESTIMATION DE LA VARIANCE

Le but principal de cette section est l'estimation de la variance qui apparaît dans le TLC pour les CM, lorsque la matrice de transition P est inconnue. Nous obtenons l'EMV de la variance en la représentant par différentes formes matricielles qui découlent des différentes démonstrations précédentes.

Les représentations de la variance qui sont données par la proposition 2.10 atteignent cet objectif, en substituant la matrice P avec son EMV \hat{P} , donné par (2.35). Nous rappelons que ces deux expressions sont bien définies si la vraie valeur du paramètre p^* correspond à une matrice P irréductible (hypothèse 2.1). De même que pour le cas de l'estimation du vecteur stationnaire, nous considérons que la vraie valeur du paramètre p^* satisfait l'hypothèse 2.2. Cela simplifie les représentations matricielles sans restreindre les résultats théoriques (convergence et normalité asymptotique). Nous trouverons également la variance asymptotique qui correspond à la loi normale limite de l'estimateur de la variance.

Pour $m \geq 1$, nous définissons

$$\hat{\sigma}^2(m) = \hat{\pi}^\#(m)(I + 2\hat{D}(m))\hat{f}^\top(m), \quad (2.101)$$

où $\hat{\pi}^\#(m)$ et $\hat{D}(m)$ sont l'EMV de $\pi^\#$ et D qui sont définis par les équations (2.81) et (2.82), et $\hat{f}(m)$ est l'EMV de \bar{f} qui correspond au vecteur des valeurs de f centré par rapport à la loi stationnaire. Pour une fonction f fixée, ces estimateurs dépendent seulement de la matrice de transition P . Afin de montrer cette dépendance et justifier le terme EMV pour l'estimateur de la variance donné par (2.101), nous devons l'exprimer comme une fonction de P .

Soit $\bar{F} = \text{diag}\{\bar{f}(i), i \in E\}$, et $F = \text{diag}\{f(i), i \in E\}$. Nous remarquons que le vecteur $\pi^\#$ s'écrit

$$\pi^\# = \pi \bar{F} = \pi [F - (\pi f^\top)I]. \quad (2.102)$$

Alors, à l'aide de l'expression (2.38), nous obtenons la forme suivante :

$$\pi^\# = \mathbf{1}(I - P + \Pi)^{-1}[F - (\mathbf{1}(I - P + \Pi)^{-1} f^\top)I]. \quad (2.103)$$

Nous introduisons la matrice C_a qui résulte de la matrice nulle en remplaçant la a -ème colonne par le vecteur $(1, 1, \dots, 1, 0, 1, \dots, 1)^\top$. Les expressions matricielles de D et \bar{f}^\top en fonction de P sont alors données par :

$$D = B_a[P_a(I_a - P_a)^{-1}]B_a^\top + C_a = B_a[B_a^\top P B_a(I_a - B_a^\top P B_a)^{-1}]B_a^\top + C_a, \quad (2.104)$$

$$\bar{f}^\top = f^\top - (\pi f^\top)\mathbf{1}^\top = f^\top - [\mathbf{1}(I - P + \mathbb{I})^{-1}f^\top]\mathbf{1}^\top. \quad (2.105)$$

En prenant l'EMV (2.35) comme estimateur de P et en substituant P dans les expressions (2.103), (2.104) et (2.105) par $\hat{P}(m)$, nous obtenons les EMV pour $\pi^\#, D$ et \bar{f}^\top ci-dessous :

$$\hat{\pi}^\#(m) = \mathbf{1}(I - \hat{P}(m) + \mathbb{I})^{-1}[F - (\mathbf{1}(I - \hat{P}(m) + \mathbb{I})^{-1}f^\top)I], \quad (2.106)$$

$$\hat{D}(m) = B_a[B_a^\top \hat{P}(m)B_a(I_a - B_a^\top \hat{P}(m)B_a)^{-1}]B_a^\top + C_a, \quad (2.107)$$

$$\hat{f}^\top(m) = f^\top - [\mathbf{1}(I - \hat{P} + \mathbb{I})^{-1}f^\top]\mathbf{1}^\top. \quad (2.108)$$

En substituant (2.106), (2.107) et (2.108) dans l'expression (2.101) nous obtenons une expression pour l'EMV de la variance σ^2 .

Nous désignons la dépendance entre σ^2 et le paramètre p^* par la relation fonctionnelle $\sigma^2 = \Phi(p^*)$.

Théorème 2.3 Soit X une chaîne de Markov qui vérifie l'hypothèse 2.2. Nous supposons que σ^2 est la variance qui correspond au TLC pour les CM (théorème 2.2) et qui est donnée par (2.75). Soit $\hat{\sigma}^2(m)$ l'EMV de la variance σ^2 , représentée sous la forme (2.101), alors, quand $m \rightarrow \infty$, nous avons :

$$i) \hat{\sigma}^2(m) \xrightarrow{\mathbb{P}_{p-p.s.}} \sigma^2,$$

$$ii) \sqrt{m}(\hat{\sigma}^2(m) - \sigma^2) \xrightarrow{p} \mathcal{N}(0, \Phi' \Gamma^* (\Phi')^\top), \text{ à condition que } \sigma^2 > 0,$$

où le vecteur $\Phi' = (\partial\Phi/\partial p_{ij})_{(i,j) \in E_*^2}$ de dimension $s^2 - s$ et Γ^* une représentation minimale (indexée par E_*^2) de la matrice Γ , définie par (2.33).

Les éléments du vecteur Φ' sont donnés par

$$\begin{aligned} \frac{\partial\Phi}{\partial p_{ij}} &= \pi[H^{ij}\bar{F} - (\pi H^{ij}f^\top)I](I + 2D)\bar{f}^\top + \\ &+ 2\pi^\#B_a[I_a + P_a(I_a - P_a)^{-1}][B_a^\top \frac{\partial P}{\partial p_{ij}}B_a(I_a - P_a)^{-1}]B_a^\top \bar{f}^\top - \\ &- \pi^\#(I + 2D)(\pi H^{ij}f^\top)\mathbf{1}^\top, \end{aligned} \quad (2.109)$$

où nous posons

$$H^{ij} = \frac{\partial P}{\partial p_{ij}}(I - P + \mathbb{I})^{-1}. \quad (2.110)$$

Démonstration i) Nous rappelons d'abord que la multiplication, l'addition et l'application inverse dans l'espace des matrices non-singulières sont des applications continues (par rapport à une norme appropriée).

Par conséquent, par la convergence forte de $\widehat{P}(m)$ (2.50) et le théorème de l'application continue nous inférons par les équations (2.106), (2.107) et (2.108) que $\mathbb{P}_p - p.s., \widehat{\pi}^\#(m) \rightarrow \pi^\#, \widehat{D}(m) \rightarrow D, \text{ et } \widehat{\mathbf{f}}^\top(m) \rightarrow \widetilde{\mathbf{f}}^\top$. En combinant ces résultats nous concluons par (2.101) et le théorème de Slutsky, la convergence forte de $\widehat{\sigma}^2(m)$ vers σ^2 .

ii) Nous devons assurer que la fonction Φ , qui exprime la variance comme une fonction du paramètre \mathbf{p}^* , est différentiable sur $(\mathcal{P}^*)^\circ$. Cela est vrai si chaque terme dans la factorisation de σ^2 dans la relation (2.85), considéré comme une fonction de \mathbf{p}^* , est différentiable sur $(\mathcal{P}^*)^\circ$.

En ce qui concerne le premier facteur $\pi^\#$, donné par (2.81), nous avons déjà remarqué dans la section précédente que chaque composante de π est une fonction rationnelle de \mathbf{p}^* . De même, $\widetilde{f}(i) = f(i) - \mu = f(i) - \sum_{i \in E} f(i)\pi(i)$ est une fonction rationnelle de \mathbf{p}^* , où le dénominateur est égale à $\det(I - P + \mathbb{I})$. Par conséquent chaque composante de $\pi^\#$ est un rapport de polynômes de \mathbf{p}^* , où le dénominateur est égale à $\det^2(I - P + \mathbb{I})$. Cela implique la différentiabilité sur $(\mathcal{P}^*)^\circ$. Avec des arguments similaires nous pouvons tirer la conclusion que le deuxième facteur $I + 2D$, considéré comme un vecteur, a des composantes différentiables sur $(\mathcal{P}^*)^\circ$. Dans ce cas, le dénominateur est égale à $\det(I_a - P_a)$. Pareil pour $\widetilde{\mathbf{f}}^\top$, où le dénominateur est égale à $\det(I - P + \mathbb{I})$.

La discussion ci-dessus nous permet d'appliquer la méthode delta (voir, par exemple chapitre 3 de Van Der Vaart (1998)). Nous avons

$$\sqrt{m}(\widehat{\sigma}^2(m) - \sigma^2) = \sqrt{m} [\Phi(\widehat{\mathbf{p}}^*(m)) - \Phi(\mathbf{p}^*)],$$

et puisque

$$\sqrt{m}(\widehat{\mathbf{p}}^*(m) - \mathbf{p}^*) \overset{p}{\rightsquigarrow} \mathcal{N}(\mathbf{0}, \Gamma^*),$$

nous pouvons conclure que,

$$\sqrt{m} [\Phi(\widehat{\mathbf{p}}^*(m)) - \Phi(\mathbf{p}^*)] \overset{p}{\rightsquigarrow} \mathcal{N}(\mathbf{0}, \Phi' \Gamma^* (\Phi')^\top).$$

Nous calculons $\partial\Phi/\partial p_{ij}$ en utilisant la relation (2.85), qui s'exprime de cette façon $\sigma^2 = \Phi(\mathbf{p}^*) = \pi^\#(I + 2D)\widetilde{\mathbf{f}}^\top$, nous obtenons donc

$$\frac{\partial\Phi}{\partial p_{ij}} = \frac{\partial\pi^\#}{\partial p_{ij}}(I + 2D)\widetilde{\mathbf{f}}^\top + 2\pi^\# \frac{\partial D}{\partial p_{ij}}\widetilde{\mathbf{f}}^\top + \pi^\#(I + 2D) \frac{\partial\widetilde{\mathbf{f}}^\top}{\partial p_{ij}}. \quad (2.111)$$

Nous analysons maintenant les dérivées qui apparaissent dans l'équation ci-dessus,

$$\frac{\partial\pi^\#}{\partial p_{ij}} = \frac{\partial\pi\bar{F}}{\partial p_{ij}} = \frac{\partial\pi}{\partial p_{ij}}\bar{F} + \pi \frac{\partial\bar{F}}{\partial p_{ij}}. \quad (2.112)$$

En utilisant le fait que

$$\frac{\partial\pi}{\partial p_{ij}} = \frac{\partial\mathbf{1}(I - P + \mathbb{I})^{-1}}{\partial p_{ij}} = \mathbf{1}(I - P + \mathbb{I})^{-1} \frac{\partial P}{\partial p_{ij}} (I - P + \mathbb{I})^{-1},$$

nous concluons que

$$\frac{\partial \pi}{\partial p_{ij}} = \pi H^{ij}, \quad (2.113)$$

où H^{ij} est donnée par (2.110).

Également,

$$\frac{\partial \bar{F}}{\partial p_{ij}} = \frac{\partial (F - (\pi \mathbf{f}^\top) I)}{\partial p_{ij}} = -(\pi H^{ij} \mathbf{f}^\top) I, \quad (2.114)$$

$$\frac{\partial D}{\partial p_{ij}} = \frac{\partial (B_a [P_a (I_a - P_a)^{-1}] B_a^\top + C_a)}{\partial p_{ij}} = B_a \frac{\partial [P_a (I_a - P_a)^{-1}] B_a^\top}{\partial p_{ij}}. \quad (2.115)$$

Le terme intermédiaire de (2.115) s'écrit

$$\begin{aligned} \frac{\partial [P_a (I_a - P_a)^{-1}] B_a^\top}{\partial p_{ij}} &= \frac{\partial P_a}{\partial p_{ij}} (I_a - P_a)^{-1} + P_a \frac{\partial (I_a - P_a)^{-1}}{\partial p_{ij}} = \\ &= \frac{\partial P_a}{\partial p_{ij}} (I_a - P_a)^{-1} + P_a (I_a - P_a)^{-1} \frac{\partial P_a}{\partial p_{ij}} (I_a - P_a)^{-1}. \end{aligned}$$

En substituant la dérivée $\partial P_a / \partial p_{ij}$ nous obtenons

$$\frac{\partial [P_a (I_a - P_a)^{-1}] B_a^\top}{\partial p_{ij}} = [I_a + P_a (I_a - P_a)^{-1}] [B_a^\top \frac{\partial P}{\partial p_{ij}} B_a (I_a - P_a)^{-1}]. \quad (2.116)$$

Par (2.113) et (2.114), la relation (2.112) donne

$$\frac{\partial \pi^\#}{\partial p_{ij}} = \pi [H^{ij} \bar{F} - (\pi H^{ij} \mathbf{f}^\top) I]. \quad (2.117)$$

En utilisant (2.116), la relation (2.115) devient

$$\frac{\partial D}{\partial p_{ij}} = B [I_a + P_a (I_a - P_a)^{-1}] [B^\top \frac{\partial P}{\partial p_{ij}} B (I_a - P_a)^{-1}] B^\top. \quad (2.118)$$

Également,

$$\frac{\partial \bar{\mathbf{f}}^\top}{\partial p_{ij}} = -(\pi H^{ij} \mathbf{f}^\top) \mathbf{1}^\top. \quad (2.119)$$

Maintenant, l'expression finale pour $\partial \Phi / \partial p_{ij}$ est obtenue par l'insertion des expressions (2.117), (2.118) et (2.119) dans la relation (2.111). \square

Remarque 2.9 *i) Les résultats asymptotiques i) et ii) du théorème 2.3 pour l'EMV $\hat{\sigma}^2(m)$ de la variance σ^2 sont aussi valables sous l'hypothèse 2.1.
ii) La représentation de Φ' donnée par (2.109) est valable même dans le cas où $d < s^2$ à condition que dans la représentation minimale nous enlevons les paramètres diagonaux. En effet, on peut suivre les mêmes étapes que pour $d = s^2$, en réduisant la dimension du paramètre \mathbf{p}^* et en ajustant les dimensions des matrices et des vecteurs associés qui apparaissent dans les dérivées partielles.*

Les différentes représentations de la variance selon la proposition 2.10, amènent à des différentes représentations des dérivées partielles $\partial \Phi / \partial p_{ij}$. Nous donnons, dans ce qui suit, une forme différente qui résulte de l'expression i) dans la proposition 2.10.

Proposition 2.11 Les éléments de Φ' décrits dans le théorème 2.3 s'écrivent sous la forme

$$\frac{\partial \Phi}{\partial p_{ij}} = \mathbf{f} \left((\Pi H^{ij})_{dg} (2Z - I - \Pi) + \Pi_{dg} \left(2Z \left(\frac{\partial P}{\partial p_{ij}} - \Pi H^{ij} \right) Z - \Pi H^{ij} \right) \right) \mathbf{f}^\top. \quad (2.120)$$

Démonstration Nous représentons l'EMV de la variance sous la forme

$$\widehat{\sigma}^2(m) = \mathbf{f} \widehat{\Pi}_{dg}(m) (2\widehat{Z}(m) - I - \widehat{\Pi}(m)) \mathbf{f}^\top, \quad (2.121)$$

où par (2.84) l'EMV de Π s'écrit

$$\widehat{\Pi}(m) = \mathbb{I}(I - \widehat{P}(m) + \mathbb{I})^{-1}, \quad (2.122)$$

et par (2.88) l'EMV de Z

$$\widehat{Z}(m) = (I - \widehat{P}(m) + \widehat{\Pi}(m))^{-1}. \quad (2.123)$$

Pareil comme dans (2.113) nous avons

$$\frac{\partial \Pi}{\partial p_{ij}} = \Pi \frac{\partial P}{\partial p_{ij}} (I - P + \mathbb{I})^{-1} = \Pi H^{ij}. \quad (2.124)$$

En considérant Z comme une fonction des éléments non diagonaux de P , nous concluons que pour $(i, j) \in E_*^2$

$$\begin{aligned} \frac{\partial Z}{\partial p_{ij}} &= -(I - P + \Pi)^{-1} \left(-\frac{\partial P}{\partial p_{ij}} + \frac{\partial \Pi}{\partial p_{ij}} \right) (I - P + \Pi)^{-1} \\ &= Z \left(\frac{\partial P}{\partial p_{ij}} - \Pi H^{ij} \right) Z. \end{aligned} \quad (2.125)$$

Par l'expression i) de la proposition 2.10 nous obtenons

$$\frac{\partial \Phi}{\partial p_{ij}} = \mathbf{f} \left(\frac{\partial \Pi_{dg}}{\partial p_{ij}} (2Z - I - \Pi) + \Pi_{dg} \left(2 \frac{\partial Z}{\partial p_{ij}} - \frac{\partial \Pi}{\partial p_{ij}} \right) \right) \mathbf{f}^\top. \quad (2.126)$$

En introduisant dans la relation (2.126) les dérivées données par (2.124) et (2.125), nous obtenons le résultat souhaité. \square

Remarque 2.10 Si pour $(i, j) \in E_*^2$ fixé nous désignons $G_{dg}^{ij} = \text{diag}\{G_{u,i,j}, 1 \leq u \leq s\}$, où $G_{u,i,j}$ sont données par (2.58), alors nous obtenons

$$(\Pi H^{ij})_{dg} = G_{dg}^{ij}.$$

2.6 EXEMPLES ET ILLUSTRATIONS NUMÉRIQUES

Cette section illustre les résultats théoriques pour l'estimation de la variance présentée dans la section précédente. D'abord, nous considérons un exemple simple analytique et ensuite nous illustrons la manière avec laquelle l'EMV de la variance se comporte par deux exemples numériques.

Exemple 2.1 Nous considérons l'espace d'état $E = \{1, 2\}$. La représentation minimale de l'espace paramétrique (2.53) de la chaîne de Markov X la plus simple à deux états est donnée par $\mathcal{P}^* = [0, 1] \times [0, 1]$.

Soit $p_{12} := p$ et $p_{21} := q$. La CM X est ergodique ssi $(p, q) \in]0, 1[\times]0, 1[$. Dans ce cas, c'est facile de calculer la variance qui correspond au TLC donnée par la proposition 2.10 pour conclure que

$$\sigma^2 = \frac{pq(2-p-q)}{(p+q)^3} (f(1) - f(2))^2. \quad (2.127)$$

Si $(p, q) \in]0, 1[\times]0, 1[= (\mathcal{P}^*)^\circ$ et $f(1) \neq f(2)$, alors $\sigma^2 > 0$ et par le théorème 2.3, nous obtenons la variance de la loi asymptotique normale de $\sqrt{m}(\hat{\sigma}^2(m) - \sigma^2)$ sous la forme

$$\Phi' \Gamma^* (\Phi')^\top = \frac{pq}{(p+q)^7} [u(p, q) + u(q, p)] (f(1) - f(2))^4, \quad (2.128)$$

où, $u(p, q) = (1-p)[(q-p)(2-p-q) - 2p]^2$.

Exemple 2.2 Soit $E = \{1, 2, 3, 4, 5\}$ (espace d'état) et la fonction $f(x) = x$. Nous considérons la matrice de transition P , donnée par

$$P = \begin{pmatrix} 0.6 & 0.3 & 0 & 0.1 & 0 \\ 0.3 & 0.5 & 0.2 & 0 & 0 \\ 0.4 & 0 & 0.4 & 0 & 0.2 \\ 0 & 0.3 & 0 & 0.5 & 0.2 \\ 0 & 0 & 0.2 & 0.2 & 0.6 \end{pmatrix}.$$

La vraie valeur de la variance est : $\sigma^2 = 11.6562$.

En nous basant sur un chemin simulé de longueur 10^5 de la chaîne de Markov ergodique ci-dessus et après enregistrement des valeurs successives de cet estimateur de la variance nous donnons dans la figure 2.1 une représentation graphique de l'indication de la convergence. Par la mise en oeuvre du théorème 2.3, nous obtenons $\sqrt{m}(\hat{\sigma}^2(m) - \sigma^2) \xrightarrow{P} N(0, 3249)$, où nous avons utilisé la partie entière de la variance de la loi limite.

Cela nous aide à obtenir des intervalles de confiance (IC) asymptotiques pour σ^2 . Par exemple, des IC à 95% sont donnés par

$$\mathbb{P}(\sigma^2 \in [\hat{\sigma}^2(m) - 1.96\sqrt{3249/m}, \hat{\sigma}^2(m) + 1.96\sqrt{3249/m}]) = 0.95.$$

Dans la figure 2.1, les intervalles de confiance sont représentés comme des intervalles entre la courbe la plus basse et la plus haute.

Exemple 2.3 Nous considérons la marche aléatoire unidimensionnelle limitée à un espace d'état fini $S = \{0, 1, \dots, s-1\}$. Ce type de marche aléatoire finie (ou processus de

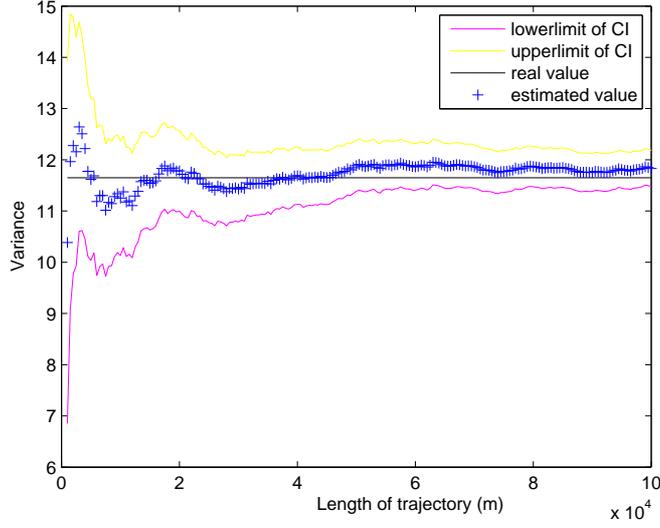


FIG. 2.1 – Estimation de la variance (95% IC) pour l'exemple 2.2.

naissance et de mort fini dans le temps discret) est une chaîne de Markov, où la matrice de transition P est donnée sous la forme suivante

$$P = \begin{pmatrix} r_0 & p_0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ q_1 & r_1 & p_1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & q_2 & r_2 & p_2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & q_{s-2} & r_{s-2} & p_{s-2} \\ 0 & 0 & 0 & 0 & \cdots & 0 & q_{s-1} & r_{s-1} \end{pmatrix}. \quad (2.129)$$

La matrice P est ergodique ssi $p_i, q_i > 0$, pour $1 \leq i \leq s - 2$, et $p_0, q_{s-1} > 0$. Cette classe des modèles représente la position exacte d'une particule mobile dans une dimension, avec la restriction de se déplacer à chaque pas du temps une position vers la gauche ou vers la droite ou rester dans la même position. La limitation à un espace d'état fini implique l'existence d'une borne inférieure et d'une borne supérieure. Si de plus, les auto-transitions ne sont pas admissibles, c'est à dire, $r_i = 0$ pour $0 \leq i \leq s - 1$, alors la chaîne correspondante est périodique de période 2.

Le modèle de Ehrenfest (voir Karlin et Taylor (1975)) peut aussi être considéré comme un cas spécial de ce modèle. Selon le genre de modèle que nous étudions nous pouvons ajuster un espace paramétrique approprié. Nous estimons la variance dans le TLC pour la classe des chaînes de Markov décrites par P en considérant comme paramètres inconnus tous les éléments non diagonaux et non identiquement nuls. Comme un cas spécial nous prenons $s = 30$, la fonction de récompense $f(x) = x$, et comme valeur exacte de la matrice de transition P nous avons fait un choix aléatoire des paramètres p_i et q_i , avec la contrainte $0.05 \leq p_i, q_i \leq 0.95$. La vraie valeur de la variance est donnée par : $\sigma^2 = 2420$.

Par la mise en oeuvre du théorème 2.3, nous obtenons

$$\sqrt{m}(\widehat{\sigma^2}(m) - \sigma^2) \xrightarrow{p} \mathcal{N}(0, 32027 \times 10^5).$$

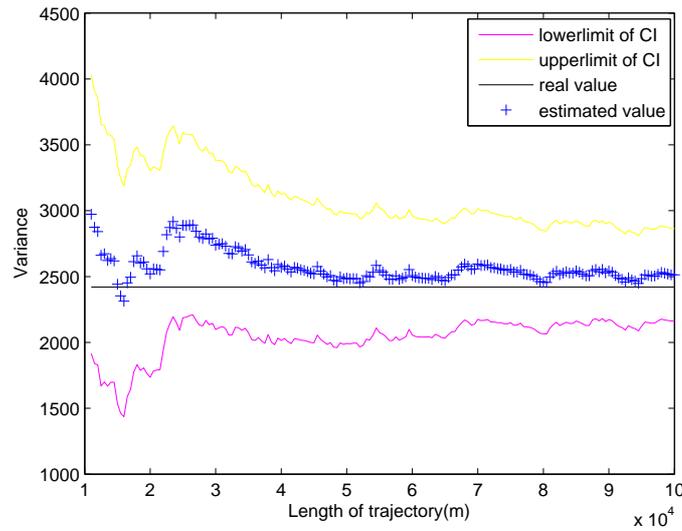


FIG. 2.2 – Estimation de la variance (95% IC) pour l'exemple 2.3.

Dans la figure 2.2, nous représentons un procédé d'estimation basé sur un chemin simulé de longueur 10^5 , donnant également les intervalles de confiance de 95 %.

CONCLUSION DU CHAPITRE

Dans ce chapitre nous avons considéré l'estimation du maximum de vraisemblance pour quelques principales caractéristiques d'une CM finie ergodique. En premier, nous avons construit l'EMV du vecteur stationnaire et nous avons étudié ses propriétés asymptotiques en le comparant avec son estimateur empirique. Puis, notre intérêt a été focalisé sur la variance asymptotique dans le TLC pour les CM. Nous avons considéré différentes formes matricielles de l'EMV pour la variance et nous avons montré sa convergence et sa normalité asymptotique. Également, nous avons donné des formes explicites pour la variance asymptotique de l'EMV de la variance.

En outre, nous avons donné quelques exemples qui ont bien montré l'efficacité de ces estimateurs.

EMV POUR DES CHAÎNES SEMI-MARKOVIENNES

SOMMAIRE

3.1	RAPPELS SUR LES CHAÎNES SEMI-MARKOVIENNES	46
3.2	LE PROCESSUS COUPLE DE CSM ET TRA	49
3.3	EMV POUR LES CHAÎNES SEMI-MARKOVIENNES FINIES	53
3.4	COMPORTEMENT ASYMPTOTIQUE DE L'EMV	58
3.5	L-TRAJECTOIRES CENSURÉES À UN INSTANT FIXE	68
	CONCLUSION	75

Les processus semi-markoviens (PSM) généralisent les processus de Markov et de renouvellement. De nos jours, ils deviennent de plus en plus importants dans les applications. Pour un échantillon d'une littérature assez abondante sur les PSM à temps continu voir Pyke (1961a;b), Çinlar (1969; 1975), Limnios et Oprüşan (2001). En ce qui concerne l'estimation statistique des systèmes semi-markoviens à temps continu, nous mentionnons, par exemple, Moore et Pyke (1968), Lagakos et al. (1978), Gill (1980), Ouhbi et Limnios (1999; 2003) et Limnios et al. (2005).

À l'opposé des PSM à temps continu, la modélisation probabiliste en temps discret n'a pas eu autant d'attention. Nous mentionnons principalement les travaux de Anselone (1960), Howard (1971), Gerontidis (1994), Mode et Pickens (1998), Mode et Sleeman (2000), Barbu et al. (2004) et Barbu (2005). Une référence très récente sur ce sujet est le livre de Barbu et Limnios (2008).

Dans le cadre du temps discret, Barbu et Limnios (2006a) ont considéré des estimateurs empiriques adaptés aux systèmes semi-markoviens non paramétriques, ayant des bonnes propriétés asymptotiques. Ces estimateurs peuvent être considérés comme estimateurs du maximum de vraisemblance approché si on ne prend pas en compte le temps de séjour dans le dernier état du système. Ils ont également montré que ces estimateurs sont fortement convergents et chaque composante de l'EMV approché est asymptotiquement normale.

À notre connaissance, l'étude de l'EMV exact qui correspond au modèle

semi-markovien non paramétrique n'a pas encore été considérée.

Dans ce chapitre, nous présentons l'EMV exact du noyau semi-markovien qui gouverne l'évolution de la chaîne semi-markovienne (CSM) et démontrons la convergence forte et la normalité asymptotique de chaque sous-vecteur fini de cet estimateur. De plus, nous montrons la relation de l'EMV exact avec l'estimateur empirique proposé par Barbu et Limnios (2006a). Cela permet aussi de démontrer la normalité asymptotique de chaque sous-vecteur fini de l'estimateur empirique.

La méthodologie suivie dans ce chapitre est basée sur un changement d'espace paramétrique qui correspond à l'espace paramétrique naturel de la CM formée par la CSM et par la suite de temps de récurrence en arrière (TRA). Ce qui nous permettra d'utiliser une approche pour l'estimation analogue à celle utilisée pour les CM dans le chapitre précédent. Ce passage d'une chaîne markovienne finie à une chaîne markovienne infinie n'est pas automatique. La clé des résultats asymptotiques de l'estimateur du maximum de vraisemblance se trouve dans une utilisation du théorème de la limite centrale pour les martingales. Afin de démontrer les résultats asymptotiques de chaque sous-vecteur fini de l'estimateur du maximum de vraisemblance de la chaîne markovienne couple, nous construisons une différence de martingales appropriée et appliquons le principe de Wald-Cramér. Nous précisons ici que le processus couple a un espace d'état infini à condition qu'il existe une loi conditionnelle de séjour dans un état de la chaîne semi-markovienne qui a un support infini. Par conséquent, la reparamétrisation du noyau semi-markovien par l'espace paramétrique naturel du processus couple implique un nombre infini de paramètres. Cela justifie le cadre non paramétrique.

Le processus couple joue un rôle important dans la compréhension de la structure semi-markovienne. En effet, il peut être utile, d'une part, pour étudier le comportement probabiliste et les théorèmes limites pour des CSM (voir par exemple, Stenflo (1996) et Chryssaphinou et al. (2008)), et d'autre part, il peut servir à l'inférence statistique. Un des enjeux de ce chapitre et du chapitre suivant est de bien illustrer ce rôle.

En outre, nous considérons dans le présent chapitre, le problème d'estimation du maximum de vraisemblance pour L trajectoires des copies indépendantes d'une CSM Z censurée à l'instant m . À notre connaissance, cette étude n'a pas également été considérée jusqu'à maintenant.

Ce chapitre est organisé comme suit. Dans la première section nous passons en revue quelques faits de base sur les chaînes semi-markoviennes finies. Dans la section 3.2, nous mentionnons des résultats existants dont nous aurons besoin pour le processus couple qui est formé par la CSM et par les TRA. Dans la section 3.3, nous donnons l'EMV exact qui correspond au modèle semi-markovien non paramétrique pour une trajectoire. Et, nous étudions son comportement asymptotique dans la section 3.4, où nous démontrons la convergence forte et la normalité asymptotique de chaque sous-vecteur fini de cet estimateur en obtenant des formes explicites pour la matrice de covariance asymptotique. Dans la dernière section de ce chapitre, nous considérons le même problème d'estimation

appliqué au cas d'une suite des copies indépendantes d'une chaîne semi-markovienne censurée à un instant fixe.

Nous mentionnons que les résultats présentés dans ce chapitre font l'objet de l'article Trevezas et Linnios (soumis, 2008).

3.1 RAPPELS SUR LES CHAÎNES SEMI-MARKOVIENNES

Nous introduisons dans cette section les notions élémentaires et les caractéristiques de base de la théorie semi-markovienne dont nous aurons besoin. Les définitions et les notations de base, dans le cadre probabiliste, sont prises et légèrement modifiées de Barbu et Limnios (2008) et Limnios et Oprışan (2001).

Soit $E = \{1, \dots, s\}$ un espace d'état fini. Notons \mathcal{M}_E l'ensemble des matrices positives sur $E \times E$ et $\mathcal{M}_E(\mathbb{N})$ l'ensemble des fonctions matricielles définies sur \mathbb{N} à valeurs dans \mathcal{M}_E .

Définition 3.1 Une fonction matricielle $q := (q(k), k \in \mathbb{N}) \in \mathcal{M}_E(\mathbb{N})$ est appelée noyau semi-markovien à temps discret si elle satisfait les deux propriétés suivantes :

1. $q_{ij}(k) \geq 0, i, j \in E, k \in \mathbb{N}$,
2. $\sum_{k=0}^{\infty} \sum_{j \in E} q_{ij}(k) = 1, i \in E$.

Nous définissons la fonction $\bar{H}_i(\cdot)$ suivante

$$\bar{H}_i(u) = \sum_{j \in E} \sum_{k=u+1}^{+\infty} q_{ij}(k), \quad u \in \mathbb{N}. \quad (3.1)$$

Définition 3.2 Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $(\mathbf{J}, \mathbf{S}) := (J_n, S_n)_{n \in \mathbb{N}}$ un processus bidimensionnel à valeurs dans $E \times \mathbb{N}$. Le processus (\mathbf{J}, \mathbf{S}) est appelé chaîne de renouvellement markovienne (CRM) associée au noyau semi-markovien q si pour tout $n \in \mathbb{N}^*, k \in \mathbb{N}, i, j \in E$, il satisfait \mathbb{P} -p.s.

$$\begin{aligned} & \mathbb{P}(J_{n+1} = j, S_{n+1} - S_n = k \mid J_n = i, \mathbf{J}_0^{n-1}, \mathbf{S}_0^n) \\ &= \mathbb{P}(J_{n+1} = j, S_{n+1} - S_n = k \mid J_n = i) = q_{ij}(k), \end{aligned} \quad (3.2)$$

alors que pour $n = 0$

$$\mathbb{P}(J_1 = j, S_1 - S_0 = k \mid J_0 = i, S_0 = u) = \frac{q_{ij}(k+u)}{\bar{H}_i(u)}, \quad (3.3)$$

où $\bar{H}_i(u)$ est donnée par (3.1) en fonction de q .

Nous notons $\beta(\cdot)$ la loi initiale de (\mathbf{J}, \mathbf{S}) , complètement déterminée par le vecteur $\boldsymbol{\beta}$ à composantes $\beta_{i,u} := \mathbb{P}(J_0 = i, S_0 = u)$.

De plus, nous avons que $(\mathbf{J}^*, \mathbf{S}^*)$ (sans le vecteur (J_0, S_0)) est homogène par rapport au temps.

Remarque 3.1 i) Si nous considérons en particulier des lois initiales $\beta(\cdot)$ pour (J_0, S_0) tel que $\beta(E \times \{0\}) = 1$, alors $S_0 = 0$ \mathbb{P} -p.s. et (3.3) peut être incorporée dans (3.2) pour $n = 0$. Dans ce cas (\mathbf{J}, \mathbf{S}) est homogène aussi. Nous notons la loi marginale de J_0 par $\alpha(\cdot)$.

ii) Dans le cas général nous notons

$$\beta_{u|i} := \mathbb{P}(S_0 = u \mid J_0 = i) \quad \text{et} \quad \beta_{i|u} := \mathbb{P}(J_0 = i \mid S_0 = u). \quad (3.4)$$

iii) $E \times \mathbb{N}$, donné dans la définition 3.2 comme ensemble des valeurs pour le processus (\mathbf{J}, \mathbf{S}) , représente le plus grand ensemble où ce processus peut prendre

des valeurs par rapport à toutes les sélections possibles du noyau semi-markovien donné par la définition 3.1. Pour une sélection spécifique, l'espace d'état se détermine par l'ensemble

$$S = \{(i, k) \in E \times \mathbb{N} : \exists n \in \mathbb{N}, \mathbb{P}((J_n, S_n) = (i, k)) > 0\}. \quad (3.5)$$

Nous introduisons deux processus basés sur S et qui jouent un rôle important dans l'interprétation probabiliste d'un système semi-markovien :

$$\mathbf{X} := (X_n)_{n \in \mathbb{N}}, \text{ où } X_0 = S_0 \text{ et } X_n = S_n - S_{n-1} \text{ pour } n \in \mathbb{N}^*, \quad (3.6)$$

$$\mathbf{N} := (N(n))_{n \in \mathbb{N}}, \text{ où } N(n) = \sup\{k \in \mathbb{N} : S_k - S_0 \leq n\}. \quad (3.7)$$

Définition 3.3 *i) La fonction $\bar{H}_i(\cdot)$ définie par (3.1) pour tout $i \in E$, est appelée la fonction de survie du temps de séjour dans l'état i , puisque par la définition de \mathbf{X} et les relations (3.1) et (3.2) nous avons pour $n \in \mathbb{N}^*$*

$$\mathbb{P}(X_{n+1} > u \mid J_n = i) = \bar{H}_i(u). \quad (3.8)$$

Le temps de séjour moyen dans l'état i est donné par

$$m_i := \mathbb{E}(X_{n+1} \mid J_n = i) = \sum_{u=0}^{+\infty} \bar{H}_i(u). \quad (3.9)$$

ii) La fonction $\bar{F}_{ij}(\cdot)$ définie pour tout $(i, j) \in E_+^2$ (voir (2.23)) par

$$\bar{F}_{ij}(u) := \frac{\sum_{k=u+1}^{+\infty} q_{ij}(k)}{\sum_{k=0}^{+\infty} q_{ij}(k)} = \mathbb{P}(X_{n+1} > u \mid J_n = i, J_{n+1} = j), \quad (3.10)$$

est appelée fonction de survie du temps de séjour conditionnel dans l'état i , conditionné au prochaine état visité j .

Le temps de séjour moyen conditionnel dans l'état i , conditionné au prochaine état visité j est donné par

$$m_{ij} := \mathbb{E}(X_{n+1} \mid J_n = i, J_{n+1} = j) = \sum_{u=0}^{+\infty} \bar{F}_{ij}(u). \quad (3.11)$$

À l'aide de (3.2), il est facile de déduire que \mathbf{J} et (\mathbf{J}, \mathbf{X}) vérifient la propriété de Markov.

Proposition 3.1 *i) Le processus (\mathbf{J}, \mathbf{X}) est une chaîne de Markov à espace d'état¹ $E \times \mathbb{N}$; il est appelé (\mathbf{J}, \mathbf{X}) -processus à espace d'état E . Les probabilités de transition sont données, pour $n \in \mathbb{N}^*$, par*

$$\mathbb{P}(J_{n+1} = j, X_{n+1} = k \mid J_n = i, X_n = u) = q_{ij}(k),$$

alors que pour $n = 0$, cette probabilité est identique à (3.3).

ii) Le processus \mathbf{J} est une chaîne de Markov à espace d'état E ; il est appelé la chaîne de Markov immergée de la CRM (\mathbf{J}, \mathbf{S}) . Nous notons $P^e \in \mathcal{M}_E$ la

¹ idem comme remarque 3.1-(iii), en substituant S_n par X_n .

matrice de transition de la CM \mathbf{J}^* , déterminée par le vecteur de probabilités \mathbf{p}^e , à composantes p_{ij} qui vérifient pour $n \in \mathbb{N}^*$

$$p_{ij} := \mathbb{P}(J_{n+1} = j \mid J_n = i) = \sum_{k=0}^{+\infty} q_{ij}(k), \quad (3.12)$$

alors que pour $n = 0$

$$\mathbb{P}(J_1 = j \mid J_0 = i) = p_{ij} \sum_{u=0}^{+\infty} \beta_{u|i} \frac{\bar{F}_{ij}(u)}{\bar{H}_i(u)}. \quad (3.13)$$

iii) Le processus \mathbf{X}^* conditionnellement à \mathbf{J}^* forme une suite de variables aléatoires conditionnellement indépendantes et chaque X_{n+1} pour $n \in \mathbb{N}^*$ dépend exclusivement de J_n et de J_{n+1} . Nous notons $f_{ij}(\cdot)$ sa loi conditionnelle, pour tout $(i, j) \in E_+^2$ (voir (2.23) appliqué ici pour la matrice P^e). Cette loi est complètement déterminée par le vecteur \mathbf{f}_{ij} à composantes $f_{ij}(k)$ qui vérifient pour $k \in \mathbb{N}$, $n \in \mathbb{N}^*$,

$$f_{ij}(k) := \mathbb{P}(X_{n+1} = k \mid J_n = i, J_{n+1} = j) = \frac{q_{ij}(k)}{p_{ij}}, \quad (3.14)$$

alors que pour $n = 0$

$$\mathbb{P}(X_1 = k \mid J_0 = i, J_1 = j) = \frac{\sum_{u=0}^{+\infty} \beta_{u|i} \frac{f_{ij}(k+u)}{\bar{H}_i(u)}}{\sum_{u=0}^{+\infty} \beta_{u|i} \frac{\bar{F}_{ij}(u)}{\bar{H}_i(u)}}. \quad (3.15)$$

La définition suivante introduit les chaînes semi-markoviennes.

Définition 3.4 Soit (\mathbf{J}, \mathbf{S}) une chaîne de renouvellement markovienne donnée par la définition 3.2. Le processus $\mathbf{Z} := (Z_n)_{n \in \mathbb{N}}$ est appelé chaîne semi-markovienne associée à (\mathbf{J}, \mathbf{S}) , si

$$Z_n = J_{N(n)}, \quad n \in \mathbb{N}, \quad (3.16)$$

où $N(n)$ est donnée par (3.7).

Dans notre travail nous adoptons deux conditions pour le noyau semi-markovien qui sont bien adaptées pour les applications et donnent une interprétation assez naturelle pour la CSM \mathbf{Z} et les autres processus \mathbf{J} , \mathbf{S} , \mathbf{X} et \mathbf{N} que nous avons déjà définis. Ces conditions sont :

(C1) $q(0) = O_s$, (forme équivalente : pour tout $i \in E$, $\bar{H}_i(0) = 1$.)

(C2) $(q(k))_{dg} = O_s$, pour $k \in \mathbb{N}$,

où O_s est la matrice nulle de dimension $s \times s$.

La condition (C1) exprime l'absence des transitions instantanées pour la CSM \mathbf{Z} . Ainsi, Z_n est l'état du système à l'instant observé n . Cela implique

que la suite S est strictement croissante \mathbb{P} -p.s. et la suite X^* des accroissements de S n'a pas des termes nuls.

La condition (C2) nous permet d'identifier un saut avec un changement d'état puisqu'elle exclue les auto-transitions pour J . Cela implique que S^* peut être interprété comme les instants successifs de changement d'état pour la CSM Z qui commence à être observée après un temps aléatoire S_0 dans l'état J_0 . Une condition nécessaire et suffisante afin de pouvoir imposer (C2) est l'absence d'état absorbant pour la CM J .

Le processus J décrit la suite des différents états observés qui correspondent aux instants indiqués par S . En tenant compte du temps initial S_0 d'observation de Z , nous avons pour tout $n \in \mathbb{N}$ la relation suivante

$$J_n = Z_{S_n - S_0}. \quad (3.17)$$

La suite X^* correspond aux temps de séjour dans les états successifs observés. De plus, N peut être interprété sous la condition (C2) comme étant le processus de comptage du nombre de changements d'état de Z et se simplifie par (C1) (voir (3.7)) comme

$$N(n) = \max\{k \leq n : S_k - S_0 \leq n\}. \quad (3.18)$$

Nous définissons également la suite de temps de récurrence en arrière (TRA) $\mathbf{U} := (U_n)_{n \in \mathbb{N}}$ de Z , où

$$U_n = \begin{cases} n + S_0 & \text{si } n < S_1 - S_0, \\ n + S_0 - S_{N(n)} & \text{si } n \geq S_1 - S_0, \end{cases} \quad (3.19)$$

où $N(n)$ est donné par (3.18).

3.2 LE PROCESSUS COUPLE DE CSM ET TRA

Nous considérons à présent le couple (Z, \mathbf{U}) formé par le processus Z et les temps de récurrence en arrière \mathbf{U} . Ce processus bidimensionnel constitue une chaîne de Markov (voir par exemple Limnios et Oprisan (2001), Theorem 3.12, ou Chryssaphinou et al. (2008)). Dans leur travail, Chryssaphinou et al. (2008) étudient quelques propriétés de (Z, \mathbf{U}) . L'espace d'état $B \subseteq E \times \mathbb{N}$ de (Z, \mathbf{U}) est formé par tous les états (i, u) tels que la fonction de survie pour l'état $i \in E$ dans la position $u \in \mathbb{N}$, donnée par (3.1), est strictement positive.

Nous notons pour tout $i \in E$

$$B^{(i)} := \{(i, u) \in \{i\} \times \mathbb{N} : \bar{H}_i(u) > 0\}. \quad (3.20)$$

Ainsi, l'espace d'état de (Z, \mathbf{U}) est

$$B = \bigcup_{i \in E} B^{(i)}. \quad (3.21)$$

Le processus (Z, \mathbf{U}) joue un rôle important dans la structure semi-markovienne. D'une part, il peut être utilisé pour l'étude du comportement probabiliste et des théorèmes limites, et d'autre part, il permet de faire l'inférence statistique pour les chaînes semi-markoviennes. Dans ce

chapitre et le chapitre suivant nous allons illustrer ce rôle, puisque les probabilités de transition du premier ordre identiquement non nulles qui correspondent à (\mathbf{Z}, \mathbf{U}) vont constituer une reparamétrisation appropriée pour le noyau semi-markovien q . De cette manière, l'EMV qui correspond au modèle semi-markovien et au modèle semi-markovien caché, ainsi que ses propriétés asymptotiques vont être déduits via cette reparamétrisation.

Dans ce chapitre nous adoptons la condition suivante :

Hypothèse 3.1 *Le temps de séjour dans chaque état $i \in E$ de la CSM \mathbf{Z} sous \mathbb{P}_q n'est pas borné, i.e.,*

$$B^{(i)} = \{i\} \times \mathbb{N}, \quad (3.22)$$

ce qui implique que $B = E \times \mathbb{N}$.

Remarque 3.2 *Dans le cas où le temps de séjour dans un ou plusieurs états du système sous \mathbb{P}_q sont bornés, les résultats de ce chapitre sont aussi valables en prenant en compte le temps maximal de séjour pour ces états. Dans le chapitre suivant nous examinons une classe de chaînes semi-markoviennes cachées où le temps de séjour dans tous les états de la chaîne cachée est borné.*

L'avantage de l'utilisation de (\mathbf{Z}, \mathbf{U}) pour l'étude de \mathbf{Z} réside dans le fait que la CM (\mathbf{Z}, \mathbf{U}) n'a qu'un nombre fini de transitions possibles dans chaque pas, même si elle admet un espace d'état infini. En effet, soit $(Z_n, U_n) = (i, u)$, deux cas se présentent alors. Ou bien, le système décrit par \mathbf{Z} reste dans l'état i pour une unité de temps de plus, ainsi $U_{n+1} = u + 1$. Ou bien, il fait un saut dans un état j , avec un nombre fini de possibilités, et par conséquent $U_{n+1} = 0$. Puisque l'état de U_{n+1} est complètement déterminé par la connaissance de Z_{n+1} , conditionnellement à (Z_n, U_n) , alors les probabilités de transition $p_{(i,u)(j,k)}$ non identiquement nulles de (\mathbf{Z}, \mathbf{U}) , dont la matrice de transition est $P := (p_{(i,u)(j,k)})_{(i,u),(j,k) \in B'}$ seront désignées par

$$p_{i,u;j} := \begin{cases} \mathbb{P}(Z_{n+1} = j, U_{n+1} = 0 \mid Z_n = i, U_n = u) & \text{si } j \neq i, \\ \mathbb{P}(Z_{n+1} = i, U_{n+1} = u + 1 \mid Z_n = i, U_n = u) & \text{si } j = i. \end{cases} \quad (3.23)$$

Cela va nous permettre d'avoir une nouvelle représentation du modèle SM par l'espace paramétrique naturel correspondant à la CM (\mathbf{Z}, \mathbf{U}) à l'aide des paramètres donnés ci-dessus.

La proposition suivante décrit la relation entre les éléments du noyau semi-markovien q et les probabilités de transition $p_{i,u;j}$ données par (3.23). Pour une démonstration de (3.24) voir Chryssaphinou et al. (2008) (théorème 1).

Proposition 3.2 *Les probabilités de transition du premier ordre non identiquement nulles de la CM (\mathbf{Z}, \mathbf{U}) , à espace d'état B (hypothèse 3.1), qui sont données par (3.23) s'écrivent*

de la façon suivante,

$$p_{i,u;j} = \begin{cases} \frac{q_{ij}(u+1)}{\bar{H}_i(u)} & \text{si } j \in E_{-i}, \\ \frac{\bar{H}_i(u+1)}{\bar{H}_i(u)} & \text{si } j = i, \end{cases} \quad (3.24)$$

où $\bar{H}_i(\cdot)$ est donné par (3.1).

En considérant le noyau q en fonction des paramètres $p_{i,u;j}$ pour $(i, j) \in E_*^2$, $k \in \mathbb{N}^*$, nous avons

$$q_{ij}(k) = \bar{H}_i(k-1)p_{i,k-1;j}, \quad (3.25)$$

où $\bar{H}_i(0) = 1$ et pour $k \in \mathbb{N}^*$, la fonction de survie dans l'état $i \in E$ s'écrit

$$\bar{H}_i(k) = \prod_{u=0}^{k-1} p_{i,u;i}. \quad (3.26)$$

Dans ce chapitre nous considérons le cas particulier, mais le plus souvent rencontré dans la littérature, où $S_0 = 0$. Cela implique que $X_0 = U_0 = 0$. D'après (3.18) et (3.19) nous obtenons le processus de comptage du nombre de sauts jusqu'à l'instant n sous la forme

$$N(n) = \max\{k \leq n : S_k \leq n\}, \quad (3.27)$$

et la suite de temps de récurrence en arrière sous la forme

$$U_n = n - S_{N(n)}. \quad (3.28)$$

Nous explicitons maintenant l'espace d'état de (J, X) (voir proposition 3.1) sous la condition : $S_0 = X_0 = 0$.

Soit pour tout $i \in E$,

$$\mathcal{S}_i^* = \{(j, k) \in E \times \mathbb{N}^* : q_{ij}(k) > 0\}^2. \quad (3.29)$$

Pour une loi initiale $\alpha(\cdot)$ sur E , il est clair que par la définition (3.2) et par (3.29), l'espace d'état de (J, X) est

$$\mathcal{S} := \left(\bigcup_{i \in E} \mathcal{S}_i^* \right) \cup (E \times \{0\}), \quad (3.30)$$

alors que l'espace d'état de (J^*, X^*) est

$$\mathcal{S}^* := \bigcup_{i \in E} \mathcal{S}_i^*. \quad (3.31)$$

Également, nous aurons besoin, dans la section suivante, de l'espace d'état de la CM $(J_n, X_{n+1}, J_{n+1})_{n \in \mathbb{N}}$, qui est donné par

$$\mathcal{D}^* := \{(i, k, j) \in E \times \mathbb{N}^* \times E : q_{ij}(k) > 0\}. \quad (3.32)$$

Dans le cas où la CSM Z est stationnaire, l'hypothèse $S_0 = 0$ ne rend pas la CM (Z, U) stationnaire ($S_0 = U_0$). Il est donc mieux adapté de supposer

²C'est évident que cet ensemble ainsi que tous les ensembles considérés dans la suite dépendent de la vraie valeur du paramètre q .

sous la condition de stationnarité de \mathbf{Z} que $(J_0, S_0) = (Z_0, U_0)$ suit la loi stationnaire de (\mathbf{Z}, \mathbf{U}) . Pour cette raison et par analogie avec la définition 2.2, nous définissons T_i le temps de retour dans un état $i \in E$, comme suit :

$$T_i = \inf\{n \geq S_1 : Z_n = i\}. \quad (3.33)$$

Nous notons également le temps moyen de récurrence entre deux visites successives de \mathbf{Z} dans l'état i aux instants de saut

$$\mu_{ii} := \mathbb{E}(T_i \mid Z_0 = i, U_0 = 0). \quad (3.34)$$

Nous remarquons que cela correspond également au temps moyen de retour de la CM (\mathbf{Z}, \mathbf{U}) dans l'état $(i, 0)$.

Dans le cas où la CM (\mathbf{Z}, \mathbf{U}) est ergodique, il est clair que les probabilités stationnaires sont données par

$$\pi_{i,u} = \frac{\bar{H}_i(u)}{\mu_{ii}}, \quad (i, u) \in B. \quad (3.35)$$

Nous aurons besoin de ces expressions pour les probabilités stationnaires qui apparaissent dans les résultats asymptotiques de ce chapitre. Par ailleurs, nous utiliserons le contexte de stationnarité (loi initiale \equiv loi stationnaire) dans le chapitre suivant où nous supposons que la CSM cachée est stationnaire.

Nous donnons maintenant une classification pour les états de la CSM.

Dans (3.33) nous avons défini le temps de retour T_i dans un état $i \in E$. Nous définissons également le temps d'entrée T_i^0 dans un état $i \in E$, par

$$T_i^0 = \begin{cases} T_i & \text{si } Z_0 \neq i, \\ 0 & \text{si } Z_0 = i. \end{cases}$$

Soit

$$v_{ij}^0 = \mathbb{P}(T_j^0 < \infty \mid Z_0 = i). \quad (3.36)$$

et

$$v_{ij} = \mathbb{P}(T_j < \infty \mid Z_0 = i). \quad (3.37)$$

Nous remarquons donc que

$$v_{ij}^0 = \begin{cases} v_{ij} & \text{si } i \neq j, \\ 1 & \text{si } i = j. \end{cases} \quad (3.38)$$

Nous munissons l'espace d'état E de la CSM \mathbf{Z} d'une relation d'ordre partielle « \rightarrow » à l'aide de laquelle nous obtenons une relation d'équivalence « \leftrightarrow » sur l'ensemble E .

- Définition 3.5**
- i) L'état j est dit accessible par i , et on écrit $i \rightarrow j$, si $v_{ij}^0 > 0$.
 - ii) On dit que les états i et j communiquent, et on écrit $i \leftrightarrow j$, si $i \rightarrow j$ et $j \rightarrow i$, i.e., $v_{ij}^0 v_{ji}^0 > 0$. Les classes d'équivalence sont appelées classes de communication.
 - iii) L'état i est dit transitoire si $v_{ii} < 1$ et récurrent si $v_{ii} = 1$. Dans le dernier

cas, en examinant l'espérance du temps de retour dans i en commençant par i (temps de récurrence moyen) nous obtenons deux types de récurrence, positive et nulle si $\mu_{ii} := \mathbb{E}(T_i | Z_0 = i)$ est finie ou infinie respectivement.

iv) L'état i est dit d -périodique, pour un certain $d \in \mathbb{N}^*$, si la loi du temps de retour T_i dans l'état i , en commençant par i , est d -périodique. Si l'état i est 1-périodique, il est dit apériodique.

v) Une chaîne semi-markovienne Z est dite irréductible si elle a une seule classe de communication. Elle est dite récurrente (positive/nulle) si tous les états sont récurrents (positifs/nuls).

vi) Une chaîne semi-markovienne Z est dite ergodique si elle est irréductible et récurrente positive.

vii) Une chaîne semi-markovienne Z est dite d -périodique, pour un certain $d \in \mathbb{N}^*$, si tous les états sont d -périodiques. Si Z est 1-périodique, elle est dite apériodique.

Remarque 3.3 On peut démontrer que les propriétés de récurrence (positive/nulle) et d'être transitoire caractérisent chaque classe de communication. Par conséquent, il suffit de les tester pour un état de la classe.

3.3 EMV POUR LES CHAÎNES SEMI-MARKOVIENNES FINIES

Nous examinons l'existence de l'EMV dans un espace des CSM finies définies sur $(\Omega, \mathcal{A}, \mathbb{P}_{\alpha, q})$, à espace d'état $E = \{1, 2, \dots, s\}$, de noyau semi-markovien q et de loi initiale $\alpha(\cdot)$. Nous supposons que la classe des CSM considérée vérifie l'hypothèse 3.1. Sinon et afin que les résultats de cette section restent valables, nous devons prendre en compte le temps de séjour maximal pour les états de la CSM Z où la fonction de survie s'annule.

Nous désignons pour tout $i \in E$,

$$\mathcal{Q}_i := \left\{ \mathbf{q}_i = (q_{ij}(k)) \in \mathbb{R}^{(\mathbb{N}^*)^{s-1}} : q_{ij}(k) \geq 0, \sum_{j \in E, j \neq i} \sum_{k \in \mathbb{N}^*} q_{ij}(k) = 1 \right\} \quad (3.39)$$

l'espace paramétrique de toutes les lois discrètes possibles définies sur l'ensemble $(\mathbb{N}^*)^{s-1}$, où les composantes du vecteur \mathbf{q}_i sont ordonnées selon l'ordre lexicographique de $\mathbb{N}^* \times E$. De plus, en prenant s copies indépendantes de \mathcal{Q}_i , nous associons au modèle semi-markovien l'espace paramétrique naturel

$$\mathcal{A} \times \mathcal{Q} := \mathcal{A} \times \prod_{i=1}^s \mathcal{Q}_i, \quad (3.40)$$

où \mathcal{A} représente l'espace des valeurs du vecteur initial α qui détermine la loi initiale $\alpha(\cdot)$.

Nous identifions q avec le paramètre vectoriel $\mathbf{q} \in \mathcal{Q}$ de dimension infinie. Aucune hypothèse sur la structure de \mathbf{q} n'est considérée pour le moment.

Soit $T_m := \{0, 1, \dots, m-1\}$ et \mathbb{T}_m l'ensemble T_m translaté d'une unité à droite, i.e., $\mathbb{T}_m := \{1, 2, \dots, m\}$. Nous désignons pour tout $i \in E$, $m \in \mathbb{N}^*$,

$$\mathcal{Q}_{i|m} := \left\{ \mathbf{q}_{i|m} = (q_{ij}(k)) \in \mathbb{R}^{(\mathbb{T}_m)^{s-1}} : q_{ij}(k) \geq 0, \sum_{j \in E-i} \sum_{k \in \mathbb{T}_m} q_{ij}(k) \leq 1 \right\}, \quad (3.41)$$

où $\mathbf{q}_{i|m}$ représente le vecteur fini des paramètres de $q_i \in \mathcal{Q}_i$ qui peuvent être estimés jusqu'à l'instant $m \in \mathbb{N}$. De plus, en prenant le produit de s copies indépendantes de $\mathcal{Q}_{i|m}$, nous obtenons l'espace paramétrique correspondant au vecteur \mathbf{q}_m , donné comme suit :

$$\mathcal{Q}_m := \prod_{i=1}^s \mathcal{Q}_{i|m}. \quad (3.42)$$

Dans le cadre d'une observation \mathbf{Z}_0^m d'une seule trajectoire d'une CSM censurée à l'instant m , où la loi initiale ne dépend pas du paramètre \mathbf{q} , l'information pour le vecteur initial est concentrée dans la première observation Z_0 , et par conséquent, pour tout $m \in \mathbb{N}$,

$$\hat{\boldsymbol{\alpha}}(m) = (\delta_{Z_0 i})_{i \in E'}, \quad (3.43)$$

où δ_{ij} est le delta de Kronecker.

Remarque 3.4 *Pour une observation \mathbf{Z}_0^m où le vecteur initial $\boldsymbol{\alpha}$ d'une CSM ne dépend pas du paramètre \mathbf{q} , les fonctions de vraisemblance $\mathcal{L}_m(\boldsymbol{\alpha}, \mathbf{q})$ et de log-vraisemblance $\ell_m(\boldsymbol{\alpha}, \mathbf{q})$ associées vont être exprimées en fonction de \mathbf{q} uniquement, puisque l'estimateur du vecteur initial est trivialement donné par (3.43). De plus, nous identifions $\mathbb{P}_{\boldsymbol{\alpha}, \mathbf{q}}$ à $\mathbb{P}_{\mathbf{q}}$.*

Étant donné qu'aucune information n'est accessible pour les paramètres $q_{ij}(k)$ avec $k > m$, nous nous restreignons à l'estimateur empirique $\tilde{\mathbf{q}}_m(m)$ du paramètre \mathbf{q}_m .

Nous définissons, pour J , le processus de comptage du nombre de passages aux différents états $i \in E$ qui correspond à l'observation $J_0^{N(m)}$ en négligeant le dernier état observé $J_{N(m)}$. Ce processus est donné pour tout $m \geq 1$ par les v.a.

$$N_i(m) := \sum_{n=0}^{N(m)-1} \mathbb{1}_{\{J_n=i\}}. \quad (3.44)$$

Également, nous définissons le processus de comptage du nombre de transitions du type $(i, j) \in E_*^2$ qui ont eu lieu dans $k \in \mathbb{T}_m$ unités de temps, donné pour tout $m \geq 1$ par les v.a.

$$N_{ij}(k, m) := \sum_{n=1}^{N(m)} \mathbb{1}_{\{J_{n-1}=i, J_n=j, X_n=k\}}. \quad (3.45)$$

Dans les deux définitions, nous considérons les sommes vides ($N(m) = 0$) comme étant nulles.

Soit

$$M_{(s)} := \inf\{m \in \mathbb{N}^* : N_i(m) > 0, \forall i \in E\}. \quad (3.46)$$

Pour chaque $m \geq M_{(s)}$, tous les états du système et au moins une transition de la CM J de chaque état vers un autre état ont été enregistrés.

Nous définissons comme estimateur empirique du paramètre \mathbf{q}_m , le vecteur $\tilde{\mathbf{q}}_m(m)$ à composantes

$$\tilde{q}_{ij}(k, m) = \frac{N_{ij}(k, m)}{N_i(m)}. \quad (3.47)$$

Si $N_i(m) = 0$, alors par convention $0/0 = 0$. Cette convention nous permet d'étendre l'estimation dans le cas où les états n'ont pas encore été tous observés ($m < M_{(s)}$).

Notre principal but est de trouver l'EMV qui correspond au paramètre \mathbf{q} en considérant l'observation \mathbf{Z}_0^m . La fonction de vraisemblance $\mathcal{L}_m(\mathbf{q})$ (voir remarque 3.4) qui correspond à l'observation \mathbf{Z}_0^m s'écrit pour $m \geq S_1$ ($N(m) \geq 1$) :

$$\mathcal{L}_m(\mathbf{q}) = \left(\prod_{k=1}^{N(m)} q_{J_{k-1}J_k}(X_k) \right) \bar{H}_{J_{N(m)}}(U_m), \quad (3.48)$$

où $\bar{H}_i(\cdot)$ est défini par (3.1).

Différentes reparamétrisations du modèle semi-markovien (MSM) résultent des différentes représentations du noyau semi-markovien q . Une première possibilité est via l'équation (3.14) qui nous donne

$$q_{ij}(k) = p_{ij}f_{ij}(k). \quad (3.49)$$

Cela nous permet de caractériser complètement \mathbf{q} par le couple $(\mathbf{p}^e, \mathbf{f})$ où \mathbf{p}^e contient les probabilités de transition de la CM immergée J et \mathbf{f} contient les probabilités des temps de séjour conditionnelles (voir proposition 3.1). Barbu et Limnios (2006a) ont construit des estimateurs empiriques pour les caractéristiques d'un système semi-markovien discret avec des applications à la fiabilité des systèmes. Leur approche est basée sur la paramétrisation $(\mathbf{p}^e, \mathbf{f})$ du modèle semi-markovien.

Puisque le facteur $\bar{H}_{J_{N(m)}}(U_m)$ qui apparaît dans l'équation (3.48) correspond au temps de séjour dans l'état final du système jusqu'à l'instant m , l'information que nous obtenons pour le noyau semi-markovien de ce terme peut être alors négligée pour m assez grand dans le cas des systèmes ergodiques. En considérant la fonction de log-vraisemblance de (3.48), le problème de sa maximisation se réduit à la maximisation de la fonction de log-vraisemblance approchée suivante

$$\tilde{\ell}_m(\mathbf{q}) = \sum_{i,j \in E_*^2} \sum_{k \in \mathbb{T}_m} N_{ij}(k, m) \log q_{ij}(k), \quad (3.50)$$

où nous faisons la convention que $0 \log 0 = 0$.

Il est clair que les estimateurs empiriques $\tilde{q}_{ij}(k, m)$ donnés par (3.47) constituent la solution unique du problème de maximisation ci-dessus et pour $k > m$ les paramètres sont nuls. La démonstration se base sur

les mêmes arguments que la proposition 2.2. En fait, l'EMV existe par la convention ci-dessus en considérant le nombre de transitions possibles comme un paramètre. De plus, il est unique par la convention $0/0 = 0$ qui rend \mathbf{Z} une CSM à r états si $N_i(m) > 0$ pour r indices, et rend la solution de type EMV, pour un espace d'état réduit à r états. Nous notons que l'EMV est nul pour un nombre infini de paramètres tels que $N_{ij}(k, m) = 0$. Le raisonnement ci-dessus nous permet de caractériser l'estimateur empirique $\tilde{q}_m(m)$ comme EMV approché (M-estimateur).

Une deuxième reparamétrisation possible du MSM se fait à l'aide de l'équation (3.25). Une façon de faire consiste à utiliser les termes $\bar{H}_i(k)$ qui résultent de la fonction de survie $\bar{H}_i(\cdot)$ et les probabilités $p_{i,u;j}$ de saut dans l'état $j \in E_i$, sachant que le système a séjourné dans l'état i pour u unités de temps. Cependant, nous préférons ici de construire l'espace paramétrique naturel qui correspond à la CM (\mathbf{Z}, \mathbf{U}) , puisque par (3.26) la fonction de survie dépend exclusivement des paramètres $p_{i,u;j}$.

Cet espace paramétrique s'écrit

$$\mathcal{H} := \left\{ \boldsymbol{\eta} = (p_{i,u;j}) \in \mathbb{R}^{B \times E} : p_{i,u;j} \geq 0, \sum_{j \in E} p_{i,u;j} = 1 \right\}. \quad (3.51)$$

Posons $B_m := E \times T_m$ et définissons les projections de \mathcal{H} de dimension finie du type

$$\mathcal{H}_m := \left\{ \boldsymbol{\eta}_m = (p_{i,u;j}) \in \mathbb{R}^{B_m \times E} : p_{i,u;j} \geq 0, \sum_{j \in E} p_{i,u;j} = 1 \right\}, \quad (3.52)$$

où $\boldsymbol{\eta}_m$ paramétrise les sm premières lois à support fini (voir relation (3.23)), pour lesquelles nous pouvons avoir de l'information jusqu'à l'instant m .

Dans cette section nous allons déduire l'EMV exact pour le noyau semi-markovien et sa relation avec le M-estimateur (estimateur empirique) donné par (3.50). Afin d'exprimer l'EMV $\hat{q}(m)$ du paramètre \boldsymbol{q} , nous aurons besoin des processus de comptages bien adaptés pour la CM (\mathbf{Z}, \mathbf{U}) à espace d'état infini dénombrable ($B = E \times \mathbb{N}$) par analogie avec les processus (2.16) et (2.17) pour une CM \mathbf{X} à espace d'état fini E .

Nous définissons pour tout $(i, u) \in E \times \mathbb{N}$ le processus de comptage $\mathbf{N}_{i,u}^B := (N_{i,u}^B(m))_{m \in \mathbb{N}^*}$, où $N_{i,u}^B(m)$ indique le nombre de visites à l'état $(i, u) \in E \times \mathbb{N}$, jusqu'à l'instant $m \in \mathbb{N}$ en négligeant le dernier état observé (Z_m, U_m) :

$$N_{i,u}^B(m) := \sum_{n=0}^{m-1} \mathbb{1}_{\{Z_n=i, U_n=u\}}, \quad (i, u) \in E \times \mathbb{N}, \quad (3.53)$$

ainsi que le processus de comptage $\mathbf{N}_{i,u;j}^B := (N_{i,u;j}^B(j, m))_{m \in \mathbb{N}^*}$, où $N_{i,u;j}^B(j, m)$ indique le nombre de transitions du type $(i, u), (j, k)$ qui sont réalisables, i.e., de probabilité non nulle :

$$N_{i,u;j}^B(j, m) := \sum_{n=0}^{m-1} \mathbb{1}_{\{Z_{n-1}=i, U_{n-1}=u, Z_n=j\}}, \quad (i, j) \in E^2, u \in \mathbb{N}. \quad (3.54)$$

La reparamétrisation du MSM sera effectuée via la relation qui exprime le paramètre \mathbf{q} correspondant au noyau semi-markovien comme une fonction de paramètre $\boldsymbol{\eta}$ qui correspond à la paramétrisation naturelle de la CM (\mathbf{Z}, \mathbf{U}) (voir (3.51)).

Soit $\mathbf{q} := \Phi(\boldsymbol{\eta})$, où les fonctions composantes de Φ (voir (3.25) et (3.26)) pour $i \neq j$ sont données par :

$$\Phi_{ijk}(\boldsymbol{\eta}) = \begin{cases} p_{i,0;j} & \text{si } k = 1, \\ p_{i,k-1;j} \prod_{r=0}^{k-2} p_{i,r;i} & \text{si } k \geq 2. \end{cases} \quad (3.55)$$

Nous notons que chaque fonction composante Φ_{ijk} dépend d'un nombre fini de k variables.

Dans la proposition suivante nous obtenons l'EMV pour le paramètre $\boldsymbol{\eta}_m \in \mathcal{H}_m$ (voir (3.52)) qui contient les probabilités des transitions pour lesquelles nous pourrions avoir d'information jusqu'à l'instant m . Puis, nous transférons ces résultats à l'EMV du paramètre $\mathbf{q}_m \in \mathcal{Q}_m$ (voir (3.42)) qui peut être estimé jusqu'à l'instant m . En outre, nous établissons la liaison entre l'estimateur empirique $\hat{\mathbf{q}}_m(m)$ à composantes donnés par (3.47) et l'EMV $\hat{\mathbf{q}}(m)$.

Proposition 3.3 *Soit \mathbf{Z} une CSM définie sur $(\Omega, \mathcal{F}, \mathbb{P}_q)$, (modèle non-paramétrique) où cette famille des espaces probabilisés est paramétrisée par $\mathbf{q} \in \mathcal{Q}$ donné par (3.40), à valeurs dans un espace d'état $E = \{1, 2, \dots, s\}$.*

i) L'EMV $\hat{\mathbf{q}}(m)$ qui correspond à une observation \mathbf{Z}_0^m de \mathbf{Z} censurée à l'instant m , est donnée par ses composantes pour $(i, j) \in E_^2$, $k \in \mathbb{T}_m$, définies par*

$$\hat{q}_{ij}(k, m) = \Phi_{ijk}(\hat{\boldsymbol{\eta}}_m(m)) = \begin{cases} \hat{p}_{i,0;j}(m) & \text{si } k = 1, \\ \hat{p}_{i,k-1;j}(m) \prod_{r=0}^{k-2} \hat{p}_{i,r;i}(m) & \text{si } 2 \leq k \leq m, \end{cases} \quad (3.56)$$

où l'EMV $\hat{\boldsymbol{\eta}}_m(m)$ du paramètre $\boldsymbol{\eta}_m$ est donné par

$$\hat{\boldsymbol{\eta}}_m(m) = (\hat{p}_{i,u;j}(m)) = \left(\frac{N_{i,u}^B(j, m)}{N_{i,u}^B(m)} \right). \quad (3.57)$$

Si $N_{i,u}^B(m) = 0$ alors par convention $0/0 = 0$. En utilisant (3.26), nous avons alors

$$\hat{q}_{ij}(k, m) = \hat{H}_i(k-1, m) \hat{p}_{i,k-1;j}(m), \quad (3.58)$$

où $\hat{H}_i(k-1, m)$ est l'EMV de $\bar{H}_i(k-1)$.

ii) La relation entre l'EMV exact $\hat{q}_{ij}(k, m)$ donné par (3.56) et l'estimateur empirique $\tilde{q}_{ij}(k, m)$ donné par (3.47), s'écrit pour $i \neq j$ comme suit :

$$\hat{q}_{ij}(k, m) = \tilde{q}_{ij}(k, m) \left(1 - \frac{\mathbb{1}_{\{Z_m=i, U_m>0\}}}{N_i(m) + 1} \right) \left(1 + \frac{\mathbb{1}_{\{Z_m=i, k>U_m\}}}{N_{i,U_m}^B(m)} \right). \quad (3.59)$$

Démonstration. i) En utilisant les processus de comptage (3.54), la fonction de vraisemblance $\mathcal{L}_m(\mathbf{q})$ donnée par (3.48) s'écrit en fonction de $\boldsymbol{\eta}_m$ sous la forme suivante

$$\mathcal{L}_m(\boldsymbol{\eta}_m) = \alpha(\mathbf{Z}_0) \prod_{i,j \in E} \prod_{u \in \mathbb{T}_m} p_{i,u;j}^{N_{i,u}^B(j, m)}. \quad (3.60)$$

Puisque la loi initiale ne varie pas avec $\boldsymbol{\eta}_m$ (voir remarque 3.4) la fonction de log-vraisemblance qu'on cherche à maximiser est donnée par

$$\ell_m(\boldsymbol{\eta}_m) = \sum_{i,j \in E} \sum_{u \in T_m} N_{i,u}^B(j, m) \log p_{i,u;j}. \quad (3.61)$$

Les relations de dépendance que nous considérons dans un cadre non-paramétrique sont imposées par la nature stochastique de la matrice de transition P de la CM (\mathbf{Z}, \mathbf{U}) . Ainsi, pour tout $(i, u) \in B$,

$$\sum_{j \in E} p_{i,u;j} = 1. \quad (3.62)$$

Par les relations (3.53), (3.54) et (3.62) et en maximisant la fonction de log-vraisemblance $\ell_m(\boldsymbol{\eta}_m)$ donnée par (3.61) par rapport à $\boldsymbol{\eta}_m$, nous obtenons (3.57), où nous supposons que le nombre d'états pour la CM (\mathbf{Z}, \mathbf{U}) est un paramètre estimé par le nombre des (i, u) pour lesquels $N_{i,u}(m) > 0$. Cela nous permet de justifier la solution du type EMV pour un tel espace d'état réduit. Par le principe de l'invariance de l'EMV et la relation (3.55) nous concluons (3.56).

ii) Par (3.54) et (3.53), la relation (3.56) peut être écrite comme suit

$$\hat{q}_{ij}(k, m) = \frac{N_{i,k-1}^B(j, m)}{N_{i,k-1}^B(m)} \frac{N_{i,k-2}^B(i, m)}{N_{i,k-2}^B(m)} \dots \frac{N_{i,1}^B(i, m)}{N_{i,1}^B(m)} \frac{N_{i,0}^B(i, m)}{N_{i,0}^B(m)}, \quad k \geq 2. \quad (3.63)$$

Nous remarquons que pour $l \geq 1$,

$$N_{i,l-1}^B(j, m) = N_{ij}(l, m), \quad (3.64)$$

$$N_{i,l-1}^B(i, m) = N_{i,l}^B(m) + \mathbb{1}_{\{Z_m=i, U_m=l\}}, \quad (3.65)$$

$$N_{i,0}^B(m) = N_i(m) + \mathbb{1}_{\{Z_m=i, U_m>0\}}. \quad (3.66)$$

En introduisant (3.64), (3.65) et (3.66) dans (3.63) nous obtenons

$$\hat{q}_{ij}(k, m) = \frac{N_{ij}(k, m) (N_{i,U_m}^B(m) + \mathbb{1}_{\{Z_m=i, k>U_m\}})}{(N_i(m) + \mathbb{1}_{\{Z_m=i, U_m>0\}}) N_{i,U_m}^B(m)}. \quad (3.67)$$

Par l'équation (3.67) nous concluons (3.59). \square

Remarque 3.5 *L'EMV exact $\hat{\boldsymbol{q}}(m)$, donné par la proposition 3.3, peut être un estimateur « fiable » pour le noyau SM plutôt dans une famille des CSM ergodiques à espace d'état E , puisque le nombre de visites pour chaque état $i \in E$ est théoriquement infini (avec probabilité 1) pour une réalisation quelconque de \mathbf{Z} . L'information accumulée n'est pas suffisante dans le cas d'états transitoires.*

3.4 COMPORTEMENT ASYMPTOTIQUE DE L'EMV

Pour les résultats asymptotiques présentés dans cette section, nous avons besoin de supposer que la classe des CSM en question vérifie la condition suivante.

Hypothèse 3.2 La CSM \mathbf{Z} finie à espace d'état E est ergodique (irréductible et récurrente positive) sous \mathbb{P}_q (voir définition 3.5).

L'équivalence ci-dessous peut être vérifiée facilement :

Lemme 3.1 La CSM \mathbf{Z} finie à espace d'état E est ergodique sous \mathbb{P}_q ssi la CM (\mathbf{Z}, \mathbf{U}) est ergodique sous \mathbb{P}_η .

Nous trouverons, dans ce qui suit, via le TLC pour les martingales les lois asymptotiques de l'EMV du paramètre vectoriel $\eta_{\tilde{n}}$ pour tout $\tilde{n} \in \mathbb{N}$ fixé.

Soit

$$E_{i,u}^+ := \{j \in E : p_{i,u;j} > 0\} \text{ et } d_{i,u} := |E_{i,u}^+|, (i, u) \in B. \quad (3.68)$$

Notons

$$D^+ := \bigcup_{(i,u) \in B} \{i\} \times \{u\} \times E_{i,u}^+, \quad (3.69)$$

l'ensemble des transitions de la CM (\mathbf{Z}, \mathbf{U}) avec une probabilité strictement positive et également

$$D_m^+ := D^+ \cap (B_m \times E) = \bigcup_{(i,u) \in B_m} \{i\} \times \{u\} \times E_{i,u}^+, \quad (3.70)$$

la partie de D^+ pour laquelle $U_n \in T_m$, où la dernière expression résulte de (3.69).

Si nous désignons par $d_m^+ := |D_m^+|$ (cardinal de D_m^+), alors par (3.68) nous avons pour tout $m \in \mathbb{N}$,

$$d_m^+ = \sum_{(i,u) \in B_m} d_{i,u}. \quad (3.71)$$

Notation 3.1 Nous désignons par \mathcal{H}_m^+ le domaine de variation du vecteur η_m^+ qui résulte de $\eta_m \in \mathcal{H}_m$ en gardant seulement les composantes qui sont strictement positives sous \mathbb{P}_η . Ces composantes sont indiquées par l'ensemble D_m^+ donné par (3.70).

L'étude du comportement asymptotique de l'EMV $\hat{\eta}_{\tilde{n}}(m)$ du paramètre $\eta_{\tilde{n}}$, pour tout \tilde{n} fixé, nous oblige à considérer une représentation minimale spécifique pour le paramètre vectoriel $\eta_{\tilde{n}}$.

La nature stochastique de la matrice P de la CM (\mathbf{Z}, \mathbf{U}) impose des relations de dépendance pour les paramètres de η donnés par (3.62). Par l'hypothèse 3.1, la fonction de survie $\bar{H}_i(\cdot)$ est une fonction strictement positive, pour tout $i \in E$, et par conséquent par son expression (3.26) nous pouvons déduire que pour tout $(i, u) \in B$ nous avons $p_{i,u;i} > 0$. Puisque tous ces paramètres sont strictement positifs nous pouvons les retirer de \mathcal{H}^+ et \mathcal{H}_m^+ pour avoir des représentations minimales, notées \mathcal{H}^* et \mathcal{H}_m^* , associées aux espaces paramétriques \mathcal{H} et \mathcal{H}_m respectivement.

Nous notons

$$E_{i,u}^* := \{j \in E_{-i} : p_{i,u;j} > 0\} \text{ et } d_{i,u}^* := |E_{i,u}^*|, (i, u) \in B, \quad (3.72)$$

et les paramètres exclus de la paramétrisation minimale sont donnés par

$$p_{i,u;i} = 1 - \sum_{j \in E_{i,u}^*} p_{i,u;j}. \quad (3.73)$$

Puisque $p_{i,u;i} \in E_{i,u}^+$ pour tout $(i, u) \in B$, nous avons donc

$$E_{i,u}^* = E_{i,u}^+ \setminus \{i\} \quad \text{et} \quad d_{i,u}^* = d_{i,u}^+ - 1. \quad (3.74)$$

Dans le cas où ils existent (i, u) tels que $p_{i,u;i} = 1$, alors $E_{i,u}^* = \emptyset$ et par conséquent il y a une transition déterministe de (i, u) vers $(i, u + 1)$.

Nous supposons également que la classe des CSM concernée vérifie la condition suivante :

Hypothèse 3.3 *Le support de la loi du temps de séjour dans chaque état $i \in E$ de la CSM \mathbf{Z} sous \mathbb{P}_q est \mathbb{N}^* .*

Remarque 3.6 *Nous pouvons reformuler l'hypothèse 3.3 avec deux manières différentes :*

- i) *Pour tout état $i \in E$ de la CSM \mathbf{Z} sous \mathbb{P}_q , sa fonction de survie $\bar{H}_i(\cdot)$ est strictement décroissante.*
- ii) *L'hypothèse 3.1 est vérifiée, et de plus, pour tout état $(i, u) \in B$ de la CM (\mathbf{Z}, \mathbf{U}) il n'y a pas des transitions déterministes, i.e., $p_{i,u;i} < 1$, ce qui est équivalent aussi à $E_{i,u}^* \neq \emptyset$.*

Remarque 3.7 *Dans le cas où le support de la loi du temps de séjour dans un ou plusieurs états n'est pas \mathbb{N}^* (voir hypothèse 3.3 et remarque 3.6) les résultats asymptotiques de cette section sont aussi valables sous l'hypothèse 3.2. Il suffit de faire les représentations pour les lois asymptotiques normales en prenant en compte seulement les lois de transition pour lesquelles $E_{i,u}^* \neq \emptyset$ ($d_{i,u}^* \neq 0$).*

Notation 3.2 *i) Le paramètre vectoriel $\boldsymbol{\eta} \in \mathcal{H}$ admet une représentation minimale $\boldsymbol{\eta}^*$ indexée par l'ensemble*

$$D^* := \bigcup_{(i,u) \in B} \{i\} \times \{u\} \times E_{i,u}^*, \quad (3.75)$$

si $\boldsymbol{\eta}$ est considéré comme un élément de \mathcal{H}^ qui résulte de \mathcal{H} en excluant les paramètres $p_{i,u;j}$ pour lesquels les (i, u, j) n'appartiennent pas à D^* . Les composantes $p_{i,u;j}$ de $\boldsymbol{\eta}^* \in \mathcal{H}^*$ sont ordonnées selon l'ordre lexicographique de D^* .*

ii) Le paramètre vectoriel $\boldsymbol{\eta}_m \in \mathcal{H}_m$ admet une représentation minimale $\boldsymbol{\eta}_m^$ indexée par l'ensemble*

$$D_m^* := \bigcup_{(i,u) \in B_m} \{i\} \times \{u\} \times E_{i,u}^*, \quad (3.76)$$

si $\boldsymbol{\eta}_m$ est considéré comme un élément de \mathcal{H}_m^ qui résulte de \mathcal{H}_m en excluant les paramètres $p_{i,u;j}$ pour lesquels les (i, u, j) n'appartiennent pas à D_m^* . Les composantes $p_{i,u;j}$ de $\boldsymbol{\eta}_m^* \in \mathcal{H}_m^*$ sont ordonnées selon l'ordre lexicographique de D_m^* .*

Si nous notons par d_m^* le nombre des paramètres dans la paramétrisation minimale, alors par (3.74) et (3.76) nous avons

$$d_m^* := |D_m^*| = d_m - sm. \quad (3.77)$$

Le théorème suivant nous assure que l'EMV d'une sélection quelconque parmi un nombre fini de lois de transition qui correspondent aux états de la CM (\mathbf{Z}, \mathbf{U}) est asymptotiquement normal. Nous obtenons également la forme explicite de la matrice de covariance associée.

Théorème 3.1 *Soit (\mathbf{Z}, \mathbf{U}) la CM qui associe à la CSM \mathbf{Z} la suite de TRA \mathbf{U} . Sous les hypothèses 3.2 et 3.3, la suite de vecteurs aléatoires $\boldsymbol{\zeta}_{\bar{n}}(m) := \sqrt{m}(\hat{\boldsymbol{\eta}}_{\bar{n}}^+(m) - \boldsymbol{\eta}_{\bar{n}}^+(m))$, converge \mathbb{P}_{η} -faiblement, quand $m \rightarrow \infty$, vers un vecteur aléatoire $\boldsymbol{\zeta}_{\bar{n}}$, où $\boldsymbol{\zeta}_{\bar{n}} \sim \mathcal{N}(\mathbf{0}, \Gamma_{\bar{n}})$ et $\Gamma_{\bar{n}}$ est une matrice de covariance de dimension $d_{\bar{n}} \times d_{\bar{n}}$, définie dans une forme diagonale par blocs*

$$\Gamma_{\bar{n}} = \text{diag}\left\{\frac{1}{\pi_{i,u}}\Lambda_{iu} : (i, u) \in B_{\bar{n}}\right\}, \quad (3.78)$$

où pour tout $(i, u) \in B_{\bar{n}}$, la matrice Λ_{iu} de dimension $d_{i,u} \times d_{i,u}$ est donnée par

$$\Lambda_{iu} = (p_{i,u;j}[\delta_{jl} - p_{i,u;l}])_{j,l \in E_{i,u}^+}. \quad (3.79)$$

Démonstration. Nous définissons le vecteur aléatoire

$$\boldsymbol{\zeta}(m) = (\zeta_{iuj}(m) : (i, u, j) \in D_{\bar{n}}^+), \quad (3.80)$$

où

$$\zeta_{iuj}(m) = N_{i,u}^B(j, m) - N_{i,u}^B(m)p_{i,u;j}. \quad (3.81)$$

Si nous désignons par

$$R(m) := \text{diag}\{r_{iuj}(m) : (i, u, j) \in D_{\bar{n}}^+\}, \quad \text{où } r_{iuj}(m) = \frac{m}{N_{i,u}^B(m)}, \quad (3.82)$$

alors nous avons clairement que

$$\boldsymbol{\zeta}_{\bar{n}}(m) = m^{-1/2}R(m)\boldsymbol{\zeta}(m). \quad (3.83)$$

Soit également

$$R := \text{diag}\{r_{iuj} : (i, u, j) \in D_{\bar{n}}^+\}, \quad \text{où } r_{iuj} = \frac{1}{\pi_{i,u}}. \quad (3.84)$$

Nous rappelons que le processus $N_{i,u}^B(m)$ est un processus de comptage du nombre de visites dans l'état (i, u) jusqu'à l'instant m de la CM ergodique (\mathbf{Z}, \mathbf{U}) . En appliquant le théorème ergodique des CM nous avons

$$\frac{N_{i,u}^B(m)}{m} \xrightarrow{\mathbb{P}_{\eta}\text{-p.s.}} \pi_{i,u} > 0. \quad (3.85)$$

En conséquence, par (3.85) et les expressions (3.82) et (3.84) nous obtenons

$$R(m) \xrightarrow{\mathbb{P}_{\eta}\text{-p.s.}} R. \quad (3.86)$$

Notons que si nous établissons que $m^{-1/2}\boldsymbol{\zeta}(m)$ converge en loi vers le vecteur aléatoire $\boldsymbol{\zeta}$, alors par la représentation de $\boldsymbol{\zeta}_{\bar{n}}(m)$ donnée par (3.83) et le résultat limite (3.86), $\boldsymbol{\zeta}_{\bar{n}}(m)$ va converger en loi vers le vecteur aléatoire $R\boldsymbol{\zeta}$.

Nous remarquons par la relation (3.81) que

$$\begin{aligned} \sum_{j \in E_{i,u}^+} \zeta_{iuj}(m) &= \sum_{j \in E_{i,u}^+} N_{i,u}^B(j, m) - N_{i,u}^B(m) \sum_{j \in E_{i,u}^+} p_{i,u;j} \\ &= N_{i,u}^B(m) - N_{i,u}^B(m) = 0. \end{aligned} \quad (3.87)$$

Puisque pour tout $(i, u) \in B_{\bar{n}}$ il y a une relation de dépendance pour les $d_{i,u}$ v.a. $\zeta_{iuj}(m)$ (pour les différents $j \in E_{i,u}^+$), nous excluons les v.a. $\zeta_{iui}(m)$ afin de former un nouveau vecteur $\zeta^*(m)$ qui représente $\zeta(m)$ dans une représentation minimale indexée par l'ensemble $D_{\bar{n}}^*$ (voir (3.76)).

Ceci nous permet d'appliquer la technique de Cramér-Wold, où nous devons démontrer que pour un $t \in \mathbb{R}^{d_{\bar{n}}^*}$ arbitraire, la v.a. $m^{-1/2}t\zeta^*(m)$ converge en loi vers la v.a. $t\zeta^*$. À ce but, nous allons représenter le processus $\{t\zeta^*(m)\}_{m \in \mathbb{N}^*}$ comme une suite de sommes partielles d'une autre suite Y^* qui a la propriété d'être une différence de martingales adaptée à la filtration naturelle de la CM (Z, U) .

Soit $\mathcal{F}_n := \sigma\{(Z_k, U_k)_{0 \leq k \leq n}\}$, i.e., la σ -algèbre engendrée par le vecteur aléatoire $(Z_k, U_k)_{0 \leq k \leq n}$.

Pour tout $n \in \mathbb{N}^*$, nous définissons la v.a.

$$Y_n := \mathbb{1}_{\{U_{n-1} \leq \bar{n}-1, U_n=0\}} [t_{Z_{n-1}U_{n-1}Z_n} - \mathbb{E}_\eta(\mathbb{1}_{\{U_n=0\}} t_{Z_{n-1}U_{n-1}Z_n} \mid Z_{n-1}, U_{n-1})]. \quad (3.88)$$

C'est évident que la suite des v.a. Y^* est adaptée à la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}^*}$ de l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P}_\eta)$. Et par construction, nous avons $\mathbb{E}_\eta |Y_n| < \infty$ et $\mathbb{E}_\eta(Y_n \mid \mathcal{F}_{n-1}) = 0$. Par conséquent, Y^* est une (\mathcal{F}_n) -différence de martingales.

Puisque

$$\mathbb{E}_\eta(\mathbb{1}_{\{U_n=0\}} t_{Z_{n-1}U_{n-1}Z_n} \mid Z_{n-1}, U_{n-1}) = \sum_{j \in E_{i,u}^+} p_{Z_{n-1}, U_{n-1}; j} t_{Z_{n-1}U_{n-1}j}, \quad (3.89)$$

en combinant (3.88) et (3.89) nous avons

$$\begin{aligned} \sum_{n=1}^m Y_n &= \sum_{(i,u,j) \in D_{\bar{n}}^+} \sum_{n=1}^m [t_{iuj} \mathbb{1}_{\{Z_{n-1}=i, U_{n-1}=u\}} (\mathbb{1}_{\{Z_n=j\}} - p_{i,u;j})] \\ &= \sum_{(i,u,j) \in D_{\bar{n}}^+} t_{iuj} (N_{i,u}^B(j, m) - N_{i,u}^B(m) p_{i,u;j}) = t\zeta^*(m). \end{aligned} \quad (3.90)$$

Notons également que

$$\mathbb{E}_\eta(Y_n^2 \mid \mathcal{F}_{n-1}) = \mathbb{1}_{\{U_{n-1} \leq \bar{n}-1\}} \text{Var}_\eta(\mathbb{1}_{\{U_n=0\}} t_{Z_{n-1}U_{n-1}Z_n} \mid Z_{n-1}, U_{n-1}), \quad (3.91)$$

et par conséquent,

$$\begin{aligned} \sum_{n=1}^m \mathbb{E}_\eta(Y_n^2 \mid \mathcal{F}_{n-1}) &= \sum_{n=1}^m \mathbb{1}_{\{U_{n-1} \leq \bar{n}-1\}} \text{Var}_\eta(\mathbb{1}_{\{U_n=0\}} t_{Z_{n-1}U_{n-1}Z_n} \mid Z_{n-1}, U_{n-1}) \\ &= \sum_{(i,u) \in B_{\bar{n}}} N_{i,u}(m) \text{Var}_\eta(\mathbb{1}_{\{U_n=0\}} t_{iuZ_n} \mid Z_{n-1} = i, U_{n-1} = u). \end{aligned} \quad (3.92)$$

Puisque

$$\mathbb{E}_\eta(\mathbb{1}_{\{U_n=0\}} t_{iuZ_n}^2 \mid Z_{n-1} = i, U_{n-1} = u) = \sum_{j \in E_{i,u}^+} t_{iuj}^2 p_{i,u;j}, \quad (3.93)$$

et de plus

$$\begin{aligned} \mathbb{E}_\eta^2(\mathbb{1}_{\{U_n=0\}} t_{iuZ_n} \mid Z_{n-1} = i, U_{n-1} = u) &= \left(\sum_{j \in E_{i,u}^*} t_{iuj} p_{i,u;j} \right)^2 \\ &= \sum_{j \in E_{i,u}^*} t_{iuj} p_{i,u;j} \left(\sum_{l \in E_{i,u}^*} t_{iul} p_{i,u;l} \right), \end{aligned} \quad (3.94)$$

en prenant la différence de (3.93) et (3.94) pour former la variance qui apparaît dans la relation (3.92), nous obtenons

$$\begin{aligned} \sum_{n=1}^m \mathbb{E}_\eta(Y_n^2 \mid \mathcal{F}_{n-1}) &= \sum_{(i,u) \in B_{\bar{n}}} N_{i,u}^B(m) \sum_{j \in E_{i,u}^*} t_{iuj} p_{i,u;j} \left(t_{iuj} - \sum_{l \in E_{i,u}^*} t_{iul} p_{i,u;l} \right) \\ &= \sum_{(i,u) \in B_{\bar{n}}} N_{i,u}^B(m) \sum_{j,l \in E_{i,u}^*} t_{iuj} p_{i,u;j} (\delta_{jl} t_{iuj} - t_{iul} p_{i,u;l}). \end{aligned} \quad (3.95)$$

Pour le reste, notons

$$\lambda_{i,u}(\mathbf{t}, \boldsymbol{\eta}) := \sum_{j,l \in E_{i,u}^*} t_{iuj} p_{i,u;j} (\delta_{jl} t_{iuj} - t_{iul} p_{i,u;l}). \quad (3.96)$$

Soit $\mathbf{t}_{iu} := (t_{iuj} : j \in E_{i,u}^*)$ et $\Lambda_{iu}^* \in \mathbb{R}^{d_{i,u}^* \times d_{i,u}^*}$ la matrice qui résulte de Λ_{iu} donnée par (3.79) en extrayant la ligne et la colonne qui correspondent au paramètre $p_{i,u;i}$.

Selon les notations ci-dessus, les coefficients $\lambda_{i,u}(\mathbf{t}, \boldsymbol{\eta})$ donnés par (3.96) peuvent s'exprimer comme suit :

$$\lambda_{i,u}(\mathbf{t}, \boldsymbol{\eta}) = \mathbf{t}_{iu} \Lambda_{iu}^* \mathbf{t}_{iu}^\top. \quad (3.97)$$

Il est clair que

$$\det(\Lambda_{iu}^*) = \prod_{j \in E_{i,u}^*} p_{i,u;j} \left(1 - \sum_{j \in E_{i,u}^*} p_{i,u;j} \right) = \prod_{j \in E_{i,u}^+} p_{i,u;j}. \quad (3.98)$$

Ainsi, puisque Λ_{iu}^* est une matrice symétrique et $p_{i,u;j} > 0$ pour tout $j \in E_{i,u}^+$, nous concluons que pour tout $(i,u) \in B_{\bar{n}}$,

$$\lambda_{i,u}(\mathbf{t}, \boldsymbol{\eta}) > 0, \quad \text{ssi } \mathbf{t}_{iu} \neq \mathbf{0}. \quad (3.99)$$

Par (3.95) et (3.96) nous obtenons

$$\sum_{n=1}^m \mathbb{E}_\eta(Y_n^2 \mid \mathcal{F}_{n-1}) = \sum_{(i,u) \in B_{\bar{n}}} \lambda_{i,u}(\mathbf{t}, \boldsymbol{\eta}) N_{i,u}^B(m). \quad (3.100)$$

Par (3.85) et par le fait que les coefficients $\lambda_{i,u}(\mathbf{t}, \boldsymbol{\eta})$ sont constants pour \mathbf{t} fixé, nous obtenons \mathbb{P}_η -p.s.,

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{n=1}^m \mathbb{E}_\eta(Y_n^2 \mid \mathcal{F}_{n-1}) = \sum_{(i,u) \in B_{\bar{n}}} \pi_{i,u} \lambda_{i,u}(\mathbf{t}, \boldsymbol{\eta}) := \sigma^2(\mathbf{t}, \boldsymbol{\eta}). \quad (3.101)$$

Comme pour tout $(i, u) \in B_{\bar{n}}$ nous avons $\pi_{i,u} > 0$, nous constatons alors par (3.99) que

$$\sigma^2(\mathbf{t}, \boldsymbol{\eta}) > 0, \text{ ssi } \mathbf{t} \neq \mathbf{0}. \quad (3.102)$$

De plus, pour $\epsilon > 0$ arbitraire

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{n=1}^m \mathbb{E}_{\boldsymbol{\eta}}(Y_n^2 \mathbb{1}_{\{|Y_n| > \epsilon \sqrt{m}\}}) = 0, \quad (3.103)$$

étant donné que Y_m est borné pour tout $m \in \mathbb{N}^*$.

Vu (3.102), la limite exprimée par l'équation (3.101) est strictement positive et donc la première condition (M1) (voir l'annexe A.1) du TLC pour martingales est vérifiée. Également par (3.103) la deuxième condition (M2) (voir l'annexe A.1) est satisfaite. En conséquence, nous pouvons appliquer le TLC pour martingales pour conclure par (3.90) que $\mathbb{P}_{\boldsymbol{\eta}}$ -faiblement,

$$m^{-1/2} \mathbf{t} \boldsymbol{\zeta}^*(m) \rightarrow \mathcal{N}(0, \sigma^2(\mathbf{t}, \boldsymbol{\eta})). \quad (3.104)$$

Soient $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, R^{-1}\Lambda)$ et la représentation minimale $\boldsymbol{\zeta}^* \sim \mathcal{N}(\mathbf{0}, (R^*)^{-1}\Lambda^*)$. Alors, la v.a. $\mathbf{t} \boldsymbol{\zeta}^* \sim \mathcal{N}(0, \mathbf{t}(R^*)^{-1}\Lambda^*\mathbf{t}^\top)$. Clairement, par (3.101) nous avons

$$\mathbf{t}(R^*)^{-1}\Lambda^*\mathbf{t}^\top = \sigma^2(\mathbf{t}, \boldsymbol{\eta}). \quad (3.105)$$

Par (3.104), (3.105) et la définition de $\boldsymbol{\zeta}^*$ nous concluons que $\mathbb{P}_{\boldsymbol{\eta}}$ -faiblement,

$$m^{-1/2} \mathbf{t} \boldsymbol{\zeta}^*(m) \rightarrow \mathbf{t} \boldsymbol{\zeta}^*. \quad (3.106)$$

La relation (3.106) est valide pour tout $\mathbf{t} \in \mathbb{R}^{d_{\bar{n}}^*}$ et par conséquent par (3.83), (3.86) et le théorème de Slutsky nous obtenons $\mathbb{P}_{\boldsymbol{\eta}}$ -faiblement,

$$\boldsymbol{\zeta}_{\bar{n}}^*(m) \rightarrow \mathcal{N}(\mathbf{0}, \Lambda^*(R^*)^\top). \quad (3.107)$$

La démonstration est achevée pour le vecteur minimale $\boldsymbol{\zeta}_{\bar{n}}^*(m)$ puisque cette matrice de covariance donnée ci-dessus est identique à $\Gamma_{\bar{n}}^*$ qui résulte de $\Gamma_{\bar{n}}$ donnée par (3.78). Pour former $\boldsymbol{\zeta}_{\bar{n}}(m)$ à partir de $\boldsymbol{\zeta}_{\bar{n}}^*(m)$, il suffit d'ajouter la composante qui correspond au paramètre $p_{i,u;j}$, pour tout (i, u) . Puisqu'il s'agit d'une combinaison linéaire des autres composantes pour les différents j , identique à celle d'une loi multinomiale, donc l'extension est directe pour obtenir la forme de la matrice de covariance $\Gamma_{\bar{n}}$ donnée par (3.78). \square

En vue de distinguer les composantes du noyau semi-markovien qui sont strictement positives, nous définissons l'ensemble suivant

$$\mathbb{D}_m^* := \{(i, k, j) \in E \times \mathbb{T}_m \times E : (i, k-1, j) \in D_m^*\}, \quad (3.108)$$

qui est identique à D_m^* translaté d'une unité à droite pour la deuxième composante. Par la signification de D_m^* et (3.25), l'ensemble \mathbb{D}_m^* désigne les indices pour lesquels les composantes $q_{ij}(k)$ de \mathbf{q}_m sont strictement positives sous $\mathbb{P}_{\boldsymbol{\eta}}$. Par (3.32) nous concluons qu'il est identique au sous-ensemble de l'espace d'état de la CM $(J_n, X_{n+1}, J_{n+1})_{n \in \mathbb{N}}$, où $X_{n+1} \leq m$. Il peut être écrit également sous la forme

$$\mathbb{D}_m^* = \bigcup_{i \in E} \{i\} \times \mathbb{D}_{i|m}^* := \bigcup_{i \in E} \{i\} \times \left(\bigcup_{u \in \mathbb{T}_m} \{u\} \times E_{i,u-1}^* \right). \quad (3.109)$$

Notation 3.3 i) Nous désignons par \mathcal{Q}_m^+ le domaine de variation du vecteur \mathbf{q}_m^+ formé par les composantes strictement positives de \mathbf{q}_m sous \mathbb{P}_η . Ces composantes sont indexées par l'ensemble \mathbb{D}_m^* donné par (3.108) et de dimension d_m^* .

ii) Nous désignons par $\mathcal{Q}_{i|m}^+$ le domaine de variation du vecteur $\mathbf{q}_{i|m}^+$ formé par les composantes strictement positives de $\mathbf{q}_{i|m}$ sous \mathbb{P}_η . Ces composantes sont indexées par l'ensemble $\mathbb{D}_{i|m}^*$ donné par (3.109) et de dimension $d_{i|m}^* = \sum_{u \in T_m} d_{i,u}^*$.

Nous désignons par

$$\begin{aligned} \Phi_m : \mathcal{H}_m^* &\rightarrow \mathcal{Q}_m^+ \\ \boldsymbol{\eta}_m^* &\mapsto \Phi_m(\boldsymbol{\eta}_m^*) = \mathbf{q}_m^+ \end{aligned} \quad (3.110)$$

la restriction de Φ sur \mathcal{H}_m^* dont les composantes Φ_{ijk} ont la même définition que (3.55) mais appliquées à $\boldsymbol{\eta}_m^*$.

Théorème 3.2 Sous les hypothèses 3.2 et 3.3, la suite de vecteurs aléatoires $\boldsymbol{\xi}_{\bar{n}}(m) := \sqrt{m}(\widehat{\mathbf{q}}_{\bar{n}}^+(m) - \mathbf{q}_{\bar{n}}^+(m))$, converge \mathbb{P}_η -faiblement, quand $m \rightarrow \infty$, vers un vecteur aléatoire $\boldsymbol{\xi}_{\bar{n}}$ tel que $\boldsymbol{\xi}_{\bar{n}} \sim \mathcal{N}(\mathbf{0}, \Delta_{\bar{n}})$, où $\Delta_{\bar{n}}$ est une matrice de covariance de dimension $d_{\bar{n}} \times d_{\bar{n}}$ définie dans une forme diagonale par blocs

$$\Delta_{\bar{n}} = \text{diag}\{\mu_{ii} K_i : i \in E\}, \quad (3.111)$$

où pour tout $i \in E$ la matrice K_i de dimension $d_{i|\bar{n}}^* \times d_{i|\bar{n}}^*$ est donnée par

$$K_i = (q_{ij}(k)[\delta_{(j,k)(r,l)} - q_{ir}(l)])_{(k,j),(l,r) \in \mathbb{D}_{i|\bar{n}}^* \times \mathbb{D}_{i|\bar{n}}^*}. \quad (3.112)$$

Démonstration. Nous remarquons que

$$\sqrt{m}(\widehat{\mathbf{q}}_{\bar{n}}^+(m) - \mathbf{q}_{\bar{n}}^+(m)) = \sqrt{m} [\Phi_{\bar{n}}(\widehat{\boldsymbol{\eta}}_{\bar{n}}^*(m)) - \Phi_{\bar{n}}(\boldsymbol{\eta}_{\bar{n}}^*(m))], \quad (3.113)$$

où $\Phi_{\bar{n}}$ est donnée par (3.110). Également, par le théorème 3.1 quand $m \rightarrow \infty$,

$$\sqrt{m}(\widehat{\boldsymbol{\eta}}_{\bar{n}}^*(m) - \boldsymbol{\eta}_{\bar{n}}^*(m)) \overset{\eta}{\rightsquigarrow} \mathcal{N}(\mathbf{0}, \Gamma_{\bar{n}}^*). \quad (3.114)$$

Par (3.113) et la méthode delta, puisque $\Phi_{\bar{n}}$ est une fonction analytique sur \mathcal{H}_m^* et par conséquent la dérivée totale $\Phi'_{\bar{n}}$ existe, nous concluons que quand $m \rightarrow \infty$,

$$\sqrt{m}(\widehat{\mathbf{q}}_{\bar{n}}^+(m) - \mathbf{q}_{\bar{n}}^+(m)) \overset{\eta}{\rightsquigarrow} \mathcal{N}(\mathbf{0}, \Phi'_{\bar{n}} \Gamma_{\bar{n}}^* (\Phi'_{\bar{n}})^\top). \quad (3.115)$$

La prochaine étape consiste à montrer que la matrice de covariance $\Phi'_{\bar{n}} \Gamma_{\bar{n}}^* (\Phi'_{\bar{n}})^\top$ est identique à $\Delta_{\bar{n}}$ donnée par (3.111).

D'abord, il est clair que

$$\Phi'_{\bar{n}} = \text{diag}\{\Phi'_{1|\bar{n}}, \Phi'_{2|\bar{n}}, \dots, \Phi'_{s|\bar{n}}\}, \quad (3.116)$$

où pour tout $i \in E$,

$$\Phi'_{i|\bar{n}} = \left(\frac{\partial \Phi_{ijk}}{\partial p_{i,l;r}} \right)_{(k,j) \in \mathbb{D}_{i|m}^*, (l,r) \in \mathbb{D}_{i|m}^*}. \quad (3.117)$$

Notons ici que nous considérons que les fonctions Φ_{ijk} sont ordonnées selon l'ordre lexicographique de \mathbb{D}_m^* .

Si nous posons pour tout $i \in E$

$$\Lambda_i := \text{diag}\left\{\frac{1}{\bar{H}_i(u)}\Lambda_{i,u} : u \in T_{\bar{n}}\right\}, \quad (3.118)$$

alors par (3.78), (3.111) et (3.35) nous avons

$$\Phi'_{\bar{n}} \Gamma_{\bar{n}}^* (\Phi'_{\bar{n}})^\top = \Delta_{\bar{n}} \text{ ssi } \Phi'_{i|\bar{n}} \Lambda_i (\Phi'_{i|\bar{n}})^\top = K_i, \text{ pour tout } i \in E. \quad (3.119)$$

Par la relation (3.55) nous obtenons que les éléments de la matrice $\Phi'_{i|\bar{n}}$ sont donnés par

$$\frac{\partial \Phi_{ijk}}{\partial p_{i,l,r}} = \begin{cases} -\frac{\bar{H}_i(k-1)}{p_{i,l;i}} p_{i,k-1;j} & \text{si } 0 \leq l < k-1, \\ \bar{H}_i(k-1) & \text{si } l = k-1, r = j, \\ 0 & \text{si } k-1 < l < \bar{n}-1. \end{cases} \quad (3.120)$$

Maintenant, nous définissons pour tout $k \in \mathbb{T}_{\bar{n}}$, $l \in T_{\bar{n}}$, la matrice suivante

$$P_{kl}^i := \left(\frac{p_{i,k-1;j}}{p_{i,l;i}} \right)_{j \in E_{i,k-1}^*, r \in E_{i,l}^*}, \quad (3.121)$$

dont les colonnes sont identiques puisque ses éléments ne dépendent pas de l'indice r .

En combinant (3.117), (3.120) et (3.121) nous concluons que

$$(\Phi'_{i|\bar{n}})_{kl} = \begin{cases} -\bar{H}_i(k-1)P_{kl}^i & \text{si } 0 \leq l < k-1, \\ \bar{H}_i(k-1)I_{s-1} & \text{si } l = k-1, \\ O_{s-1} & \text{si } k-1 < l < \bar{n}-1. \end{cases} \quad (3.122)$$

Notons que

$$\Phi'_{i|\bar{n}} \Lambda_i = \begin{pmatrix} \Lambda_{i0} & O_{s-1} & O_{s-1} & \cdots & O_{s-1} \\ -\bar{H}_i(1)P_{20}^i \Lambda_{i0} & \Lambda_{i1} & O_{s-1} & \cdots & O_{s-1} \\ -\bar{H}_i(2)P_{30}^i \Lambda_{i0} & -\frac{\bar{H}_i(2)}{\bar{H}_i(1)}P_{31}^i \Lambda_{i1} & \Lambda_{i2} & \cdots & O_{s-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\bar{H}_i(s-1)P_{s0}^i \Lambda_{i0} & -\frac{\bar{H}_i(s-1)}{\bar{H}_i(1)}P_{s1}^i \Lambda_{i1} & -\frac{\bar{H}_i(s-1)}{\bar{H}_i(2)}P_{s2}^i \Lambda_{i2} & \cdots & \Lambda_{i,s-1} \end{pmatrix} \quad (3.123)$$

Nous avons également

$$P_{ku}^i \Lambda_{iu} = (p_{i,k-1;j} p_{i,u;r})_{j \in E_{i,k-1}^*, r \in E_{i,l}^*} \quad (3.124)$$

et par conséquent

$$P_{ku}^i \Lambda_{iu} (P_{lu}^i)^\top = \frac{1-p_{i,u;i}}{p_{i,u;i}} (p_{i,k-1;j} p_{i,l;r})_{j \in E_{i,k-1}^*, r \in E_{i,l}^*}. \quad (3.125)$$

Par l'équation (3.122) et (3.123) nous prenons pour $k < l + 1$,

$$\begin{aligned}
(\Phi'_{i|\tilde{n}} \Lambda_i (\Phi'_{i|\tilde{n}})^\top)_{kl} &= \bar{H}_i(k-1) \bar{H}_i(l) \left(\sum_{u=0}^{k-2} \frac{1-p_{i,u;i}}{p_{i,u;i} \bar{H}_i(u)} \right) (p_{i,k-1;j} p_{i,l;r})_{j,r} \\
&\quad - \bar{H}_i(l) (p_{i,k-1;j} p_{i,l;r})_{j,r} \\
&= \left\{ \sum_{u=0}^{k-2} \left(\frac{1-p_{i,u;i}}{\bar{H}_i(u+1)} \right) - \frac{1}{\bar{H}_i(k-1)} \right\} (q_{ij}(k) q_{ir}(l+1))_{j,r} \\
&= \left\{ \sum_{u=0}^{k-2} \left(\frac{1}{\bar{H}_i(u+1)} - \frac{1}{\bar{H}_i(u)} \right) - \frac{1}{\bar{H}_i(k-1)} \right\} (-K_i)_{kl} \\
&= (K_i)_{kl}. \tag{3.126}
\end{aligned}$$

Comme les matrices $\Phi'_{i|\tilde{n}} \Lambda_i (\Phi'_{i|\tilde{n}})^\top$ et K_i sont symétriques, alors leurs blocs pour $k > l + 1$ sont aussi identiques. Leurs blocs diagonaux (pour $k = l + 1$) le sont également en vertu des mêmes arguments utilisés pour obtenir (3.126). \square

Remarque 3.8 Une démonstration directe du théorème 3.2 est aussi possible à l'aide de la technique que nous avons utilisée pour démontrer le théorème 3.1. Dans ce cas, le TLC pour martingales peut être appliqué en considérant une différence de martingales appropriée par rapport à la filtration naturelle engendrée par la CM (J, \mathbf{X}) . Une sélection possible, semblable à (3.88), est la suivante

$$Y_n := \mathbb{1}_{\{X_n \leq \tilde{n}\}} [t_{J_{n-1} J_n X_n} - \mathbb{E}_q(\mathbb{1}_{\{X_n \leq \tilde{n}\}} t_{J_{n-1} J_n X_n} \mid J_{n-1})]. \tag{3.127}$$

Dans la dernière partie de cette section nous interprétons la convergence et la normalité asymptotique de l'EMV non-paramétrique du noyau semi-markovien dans le cadre de convergence des processus stochastiques, puisque nous avons un espace paramétrique de dimension infinie.

Soit

$$\hat{q}(m) := (\hat{q}_m(m), 0, 0, \dots), \quad m \geq 1, \tag{3.128}$$

la suite des estimateurs qui peuvent être décomposés en deux parties, une partie qui correspond à l'EMV exact du vecteur q_m , et une autre partie qui complète avec un nombre infini de zéros les paramètres pour lesquels nous n'avons aucune information jusqu'à l'instant m .

Théorème 3.3 *i) L'EMV $\hat{q}(m)$ est un estimateur fortement convergent de la vraie valeur du noyau semi-markovien q , i.e., quand $m \rightarrow \infty$,*

$$\hat{q}(m) \xrightarrow{\mathbb{P}_q\text{-p.s.}} q. \tag{3.129}$$

ii) La suite de processus stochastiques $\xi(m) := (\xi_m(m), 0, 0, \dots)$, $m \geq 1$, où $\xi_m(m)$ est donné dans l'énoncé du théorème 3.2, converge \mathbb{P}_q -faiblement, quand $m \rightarrow \infty$, vers un processus stochastique ξ , où pour tout $\tilde{n} \in \mathbb{N}^$, le vecteur aléatoire correspondant $\xi_{\tilde{n}} \sim \mathcal{N}(\mathbf{0}, \Delta_{\tilde{n}})$ et $\Delta_{\tilde{n}}$ est donné par (3.111).*

Démonstration. i) Par la relation (3.59) nous pouvons remarquer que l'EMV $\hat{q}_{ij}(k, m)$ et l'estimateur empirique $\tilde{q}_{ij}(k, m)$ du paramètre $q_{ij}(k, m)$ convergent \mathbb{P}_η -p.s. vers la même limite, à condition que cette limite existe. Barbu et Limnios (2008) ont démontré que $\tilde{q}_{ij}(k, m)$ est fortement convergent (voir théorème 4.1 de la même référence). Par conséquent, la convergence forte de $\hat{q}(m)$ résulte de la convergence par composantes.

ii) La suite de processus stochastiques $\xi(m)$ converge \mathbb{P}_η -faiblement vers le processus ξ ssi pour tout $\tilde{n} \in \mathbb{N}^*$ la suite de vecteurs aléatoires $\xi_{\tilde{n}}(m)$ converge \mathbb{P}_η -faiblement, quand m tend vers l'infini, vers le vecteur $\xi_{\tilde{n}}$ de dimension finie. Ce résultat est une conséquence du théorème 3.2. \square

3.5 L-TRAJECTOIRES CENSURÉES À UN INSTANT FIXE

Dans cette section nous considérons L trajectoires des copies indépendantes d'une chaîne semi-markovienne \mathbf{Z} censurée à l'instant m . L'étude basée sur plusieurs trajectoires est intéressante puisqu'il y a des problèmes où l'estimation des paramètres est basée uniquement sur plusieurs trajectoires. Par exemple, dans l'analyse de durée de vie où plusieurs patients atteints d'une certaine maladie sont observés ou dans la fiabilité où nous observons plusieurs systèmes identiques mais indépendants.

Nous construisons dans la suite, l'EMV qui correspond à ce modèle et cherchons ses propriétés asymptotiques.

Nous désignons par $(\mathbf{Z}_0^m)_l := (Z_0^{(l)}, Z_1^{(l)}, \dots, Z_m^{(l)})$ la l -ème réalisation indépendante de la CSM \mathbf{Z} censurée à l'instant m . L'estimation des paramètres faite pour une seule trajectoire d'une CSM, non ergodique, n'est pas « satisfaisante » vu son caractère intrinsèque. Au contraire, dans le cas où plusieurs trajectoires correspondantes aux copies indépendantes d'une CSM peuvent être faites, c'est possible d'explorer mieux la structure du modèle. Pour cette raison, nous ne faisons aucune hypothèse concernant le type de communication entre les différents états de la CSM \mathbf{Z} .

Nous introduisons maintenant l'hypothèse dont nous aurons besoin dans cette section.

Hypothèse 3.4 *La loi initiale $\alpha(\cdot)$ sous $\mathbb{P}_{\alpha, q}$*

- i) *a un support qui donne une probabilité strictement positive sur chaque classe de communication de \mathbf{Z} ,*
- ii) *ne dépend pas du paramètre q .*

Interprétation. Puisque pour chaque trajectoire, c'est la première observation qui détermine l'ensemble des états observés, il suffit donc d'assurer qu'au minimum un état pour chaque classe de communication de \mathbf{Z} est accessible dans cette première observation. Cela est fait dans la partie i) de l'hypothèse 3.4 qui garantit que tous les états de toutes les classes seront observés (avec probabilité un) après un L assez grand.

Par ailleurs, avec la partie ii) de l'hypothèse 3.4 nous pouvons décomposer le problème de maximisation en deux parties indépendantes. Une partie pour la loi initiale et une autre pour le noyau semi-markovien.

Malgré l'avantage de la structure arbitraire du système, rien n'assure que sa structure complète va être révélée même après un nombre de trajectoires assez grand. L'instant m de censure joue un rôle décisif sur les transitions qui peuvent être observées. Naturellement, nous ne pouvons pas observer les transitions d'un état i vers un état j lorsque le temps minimal des transitions dépasse m , malgré que j soit accessible par i sous la vraie valeur du paramètre. Par conséquent, notre but est d'estimer la partie du noyau semi-markovien qui est accessible jusqu'à l'instant m . En supposant que $S_0 = 0$, cette partie correspond au paramètre q_m . Nous montrons, dans la suite, que l'EMV du paramètre q_m est à la fois fortement convergent et asymptotiquement normal.

Également, ce type d'estimation permet de mesurer la concentration de la loi du temps de séjour dans tout $i \in E$ sur l'ensemble fini $\{1, 2, \dots, m\}$. Cela est fait par l'EMV de la fonction de répartition du temps de séjour à l'instant m . L'erreur de cette estimation peut être également mesurée en construisant des intervalles de confiance asymptotiques par la normalité asymptotique de cet estimateur.

Un autre avantage de ce type d'estimation est la possibilité d'estimer également la loi initiale $\alpha(\cdot)$ de la CSM \mathbf{Z} . À ce but, nous définissons pour tout $i \in E$ le processus de comptage $N_i^\alpha := \{N_i^\alpha(L)\}_{L \in \mathbb{N}^*}$ du nombre d'apparitions de i comme état initial dans les L trajectoires, avec

$$N_i^\alpha(L) = \sum_{l=1}^L \mathbb{1}_{\{Z_0^{(l)}=i\}}. \quad (3.130)$$

En suivant la ligne de la section précédente nous montrons ces résultats asymptotiques via la reparamétrisation du modèle semi-markovien par l'espace paramétrique naturel de la CSM (\mathbf{Z}, \mathbf{U}) censurée à l'instant m .

Soit $N_{i,u}^{B,l}$ et $N_{i,u;j}^{B,l}$ les l -èmes réalisations des processus de comptages $N_{i,u}^B$ et $N_{i,u;j}^B$ respectivement, définis dans la section précédente par (3.53) et (3.54). Nous définissons également les processus de comptages

$$N_{i,u}^B(L) := \sum_{l=1}^L N_{i,u}^{B,l}, \quad \text{d'où } N_{i,u}^B(m; L) = \sum_{l=1}^L N_{i,u}^{B,l}(m), \quad (3.131)$$

et

$$N_{i,u;j}^B(L) := \sum_{l=1}^L N_{i,u;j}^{B,l}, \quad \text{d'où } N_{i,u;j}^B(j, m; L) = \sum_{l=1}^L N_{i,u;j}^{B,l}(j, m). \quad (3.132)$$

La v.a. $N_{i,u}^B(m; L)$ indique le nombre de visites dans l'état $(i, u) \in E \times \mathbb{N}$, jusqu'à l'instant $m \in \mathbb{N}^*$ pour les L trajectoires en négligeant le dernier état observé (Z_m, U_m) . Également la v.a. $N_{i,u;j}^B(j, m; L)$ indique le nombre

de transitions du type (i, u) , (j, k) qui sont réalisables³, jusqu'à l'instant m pour les L trajectoires.

La proposition suivante étend les résultats de la proposition 3.3 pour l'estimation du maximum de vraisemblance de $\mathbf{1}$ à L trajectoires d'une CSM \mathbf{Z} censurée à l'instant m . De plus, il n'est pas nécessaire de supposer que \mathbf{Z} est ergodique puisque par l'hypothèse de l'indépendance des copies de \mathbf{Z} les états initiaux peuvent appartenir à des différentes classes de communication.

Proposition 3.4 Soit \mathbf{Z} une CSM définie sur $(\Omega, \mathcal{F}, \mathbb{P}_{\alpha, \mathbf{q}})$, (modèle non paramétrique) où cette famille des espaces probabilisés est paramétrisée par $(\alpha, \mathbf{q}) \in \mathcal{A} \times \mathcal{Q}$ donné par (3.40), à valeurs dans un espace d'état $E = \{1, 2, \dots, s\}$. Sous l'hypothèse 3.4(ii), l'EMV $(\hat{\alpha}(L), \hat{\boldsymbol{\eta}}_m(m; L))$ qui correspond à L trajectoires $(\mathbf{Z}_0^m)_i$ pour $i \in \mathbb{T}_L$ correspondantes à L copies indépendantes d'une CSM \mathbf{Z} censurée à l'instant fixe m , est donné par ses composantes, pour $(i, j) \in E_{\neq}^2$, $k \in \mathbb{T}_m$, décrites ci-dessous

$$\hat{q}_{ij}(k, m; L) = \Phi_{ijk}(\hat{\boldsymbol{\eta}}_m(m; L)) = \begin{cases} \hat{p}_{i,0;j}(m; L), & \text{si } k = 1, \\ \hat{p}_{i,k-1;j}(m; L) \prod_{r=0}^{k-2} \hat{p}_{i,r;i}(m; L), & \text{si } 2 \leq k \leq m, \end{cases} \quad (3.133)$$

$$\hat{\alpha}_i(L) = \frac{N_i^\alpha(L)}{L}, \quad (3.134)$$

où l'EMV $\hat{\boldsymbol{\eta}}_m(m; L)$ du paramètre $\boldsymbol{\eta}_m$ est donné par

$$\hat{\boldsymbol{\eta}}_m(m; L) = (\hat{p}_{i,u;j}(m; L)) = \left(\frac{N_{i,u}^B(j, m; L)}{N_{i,u}^B(m; L)} \right). \quad (3.135)$$

Si $N_{i,u}^B(m; L) = 0$, alors par convention $0/0 = 0$.

Démonstration. La fonction de vraisemblance $\mathcal{L}_L(\alpha, \mathbf{q}_m)$ pour L observations indépendantes $(\mathbf{Z}_0^m)_i$, $i \in \mathbb{T}_L$, s'écrit

$$\mathcal{L}_L(\alpha, \mathbf{q}_m) = \prod_{l=1}^L \alpha_{J_0^{(l)}} \left(\prod_{k=1}^{N^{(l)}(m)} q_{J_{k-1}^{(l)} J_k^{(l)}}(X_k^{(l)}) \right) \bar{H}_{J_{N^{(l)}(m)}}(U_m^{(l)}). \quad (3.136)$$

En utilisant les processus de comptages (3.130) et (3.132), la fonction de vraisemblance $\mathcal{L}_L(\alpha, \mathbf{q}_m)$ peut être écrite en fonction de $(\alpha, \boldsymbol{\eta}_m)$ sous la forme

$$\mathcal{L}_L(\alpha, \boldsymbol{\eta}_m) = \left(\prod_{i \in E} \alpha_i^{N_i^\alpha(L)} \right) \prod_{i,j \in E} \prod_{l \in \mathbb{T}_L} \prod_{u \in \mathbb{T}_m} p_{i,u;j}^{N_{i,u}^{B,l}(j,m)}. \quad (3.137)$$

Par conséquent, la fonction de log-vraisemblance qui doit être maximisée est donnée par

$$\ell_L(\alpha, \boldsymbol{\eta}_m) = \sum_{i \in E} N_i^\alpha(L) \log \alpha_i + \sum_{i,j \in E} \sum_{l \in \mathbb{T}_L} \sum_{u \in \mathbb{T}_m} N_{i,u}^{B,l}(j,m) \log p_{i,u;j}. \quad (3.138)$$

Les relations de dépendance que nous devons prendre en compte dans un cadre non-paramétrique sont données par (3.62) et par le fait que

³soit $(j, k) = (j, 0)$, $j \neq i$, soit $(j, k) = (i, u + 1)$.

le vecteur initial α est un vecteur de probabilité. Par ces relations de dépendance, l'hypothèse 3.4-(ii), (3.131) et en maximisant la fonction de log-vraisemblance (3.138) par rapport à α et η_m , nous obtenons (3.134) et (3.135).

Par le principe de l'invariance de l'EMV et la relation (3.55) nous concluons (3.133). \square

Dans la proposition suivante nous montrons la convergence forte de l'EMV du paramètre (α, η_m) .

Proposition 3.5 *Sous l'hypothèse 3.4, l'EMV $(\hat{\alpha}(L), \hat{\eta}_m(m; L))$ du paramètre (α, η_m) donné par (3.133) et (3.134) est fortement convergent.*

Démonstration. Il suffit de montrer que l'EMV $(\hat{\alpha}(L), \hat{\eta}_m(m; L))$ du paramètre (α, η_m) donné par (3.134) et (3.135) est fortement consistant. Si cela est vérifié alors en utilisant le théorème de l'application continue pour la fonction Φ_m donnée par (3.110) nous obtenons le résultat énoncé.

En fait, nous remarquons par la forme de $\hat{\eta}_m(m; L)$ que si nous divisons chaque terme de la fraction par L , nous pouvons appliquer la loi forte de grands nombres pour conclure que quand $L \rightarrow \infty$,

$$\hat{p}_{i,u;j}(m; L) \xrightarrow{\mathbb{P}_{\alpha,\eta}\text{-P.S.}} \frac{\mathbb{E}_{\alpha,\eta} N_{i,u}^{B,l}(j, m)}{\mathbb{E}_{\alpha,\eta} N_{i,u}^{B,l}(m)}. \quad (3.139)$$

Cela est possible puisque les suites $\{N_{i,u}^{B,l}(j, m)\}_{l \in \mathbb{N}^*}$ et $\{N_{i,u}^{B,l}(j, m)\}_{l \in \mathbb{N}^*}$ sont des suites de v.a. indépendantes d'espérance bornée par m .

Par (3.54) nous avons

$$\mathbb{E}_{\alpha,\eta} N_{i,u}^{B,l}(j, m) = p_{i,u;j} \sum_{n=0}^{m-1} \mathbb{P}_{\alpha,\eta}((Z_n^{(l)}, U_n^{(l)}) = (i, u)) = p_{i,u;j} \mathbb{E}_{\alpha,\eta} N_{i,u}^{B,l}(m). \quad (3.140)$$

En introduisant (3.140) dans (3.139) nous obtenons, quand $L \rightarrow \infty$,

$$\hat{\eta}_m(m; L) \xrightarrow{\mathbb{P}_{\alpha,\eta}\text{-P.S.}} \eta_m. \quad (3.141)$$

Avec le même raisonnement concernant la première observation de chaque trajectoire nous avons également, quand $L \rightarrow \infty$,

$$\hat{\alpha}(L) \xrightarrow{\mathbb{P}_{\alpha,\eta}\text{-P.S.}} \alpha. \quad (3.142)$$

Par (3.141) et (3.142) nous concluons la convergence forte de l'EMV $(\hat{\alpha}(L), \hat{\eta}_m(m; L))$. \square

Anderson et Goodman (1957) ont démontré la normalité asymptotique de l'EMV d'une chaîne de Markov finie même dans le cas où elle n'est pas homogène. Ils ont considéré, en premier lieu, le nombre d'observations connu pour chaque état initial pour les L trajectoires, ensuite ils ont étendu les résultats asymptotiques pour le cas où le vecteur des fréquences observées pour les états initiaux est distribué selon la loi multinomiale. Les étapes de leur démonstration sont applicables au cas de L trajectoires correspondantes à L copies indépendantes de la CM (Z, U) observées jusqu'à

l'instant m puisque elles sont basées sur le TLC multidimensionnel pour i.i.d. vecteurs aléatoires et non pas sur la structure de (\mathbf{Z}, \mathbf{U}) . Cependant, pour les appliquer dans notre cas, il suffit seulement de remarquer que le support de la loi initiale est réduit aux états $\{(i, 0)\}_{i \in E}$.

Remarque 3.9 *Sous l'hypothèse 3.4, l'EMV $\widehat{\boldsymbol{\alpha}}(L)$ est asymptotiquement normal, i.e., le vecteur $\boldsymbol{\zeta}(L) := \sqrt{L}(\widehat{\boldsymbol{\alpha}}(L) - \boldsymbol{\alpha})$ converge $\mathbb{P}_{\boldsymbol{\alpha}, \boldsymbol{\eta}}$ -faiblement, quand $L \rightarrow \infty$, vers une loi normale $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \Gamma^{(L)})$, où $\Gamma_{ij}^{(L)} := \text{Cov}_{\boldsymbol{\alpha}, \boldsymbol{\eta}}(\zeta_i, \zeta_j) = \alpha_i(\delta_{ij} - \alpha_j)$. Cela est une conséquence directe du TLC appliqué à la suite de i.i.d. vecteurs aléatoires $\{(\mathbb{1}_{\{Z_0^{(l)}=i\}} : i \in E)\}_{l \in \mathbb{N}^*}$. Par l'indépendance asymptotique de l'EMV $\widehat{\boldsymbol{\alpha}}(L)$ et l'EMV $\boldsymbol{\eta}_m^+(m; L)$, nous excluons, dans ce qui suit, le paramètre $\boldsymbol{\alpha}$ de l'estimation.*

Le théorème de Anderson et Goodman (1957) adapté à notre cas donne le lemme suivant :

Lemme 3.2 *Sous les hypothèses 3.3 et 3.4, la suite de vecteurs aléatoires $\boldsymbol{\zeta}_m(L) := \sqrt{L}(\widehat{\boldsymbol{\eta}}_m^+(L) - \boldsymbol{\eta}_m^+(m; L))$, converge $\mathbb{P}_{\boldsymbol{\alpha}, \boldsymbol{\eta}}$ -faiblement, quand $L \rightarrow \infty$, vers un vecteur aléatoire $\boldsymbol{\zeta}_m$, où $\boldsymbol{\zeta}_m \sim \mathcal{N}(\mathbf{0}, \Gamma_m^{(L)})$ et $\Gamma_m^{(L)}$ est une matrice de covariance de dimension $d_m \times d_m$, définie dans une forme diagonale par blocs*

$$\Gamma_m^{(L)} = \text{diag}\left\{\frac{1}{\pi_{i,u|m}} \Lambda_{iu} : (i, u) \in B_m\right\}, \quad (3.143)$$

où pour tout $(i, u) \in B_m$ la matrice Λ_{iu} de dimension $d_{i,u} \times d_{i,u}$ est donnée par (3.79) et $\pi_{i,u|m}$ est donné par

$$\pi_{i,u|m} := \mathbb{E}_{\boldsymbol{\alpha}, \boldsymbol{\eta}} N_{i,u}^B(m) = \sum_{r \in E} \sum_{n \in T_m} \alpha_r p_{r,0;i,u}^{(n)}, \quad (3.144)$$

où

$$p_{r,0;i,u}^{(n)} = \mathbb{P}_{\boldsymbol{\alpha}, \boldsymbol{\eta}}((Z_n, U_n) = (i, u) | (Z_0, U_0) = (r, 0)). \quad (3.145)$$

La définition suivante introduit la fonction de renouvellement markovienne qui joue un rôle très important dans la théorie de CRM. Cette définition a été donnée dans Barbu et Limnios (2008) dans un autre contexte, pour l'étude d'une seule trajectoire.

Définition 3.6 *La fonction matricielle $\Psi = (\Psi(m), m \in \mathbb{N}) \in \mathcal{M}_E(\mathbb{N})$, où*

$$\Psi_{ij}(m) := \mathbb{E}(N_{j,0}^B(m+1) | Z_0 = i) \quad (3.146)$$

est appelée fonction de renouvellement markovienne à temps discret. Nous notons également

$$\Psi_i(m) := \mathbb{E} N_{i,0}^B(m+1). \quad (3.147)$$

Avec le théorème suivant nous adaptons les résultats asymptotiques de l'EMV pour les CM au cas de CSM en explicitant également la forme de la matrice de covariance de la loi asymptotique normale de l'EMV. L'astuce consiste à montrer que ce théorème découle du lemme 3.2 d'une manière similaire que le théorème 3.2 découle du théorème 3.1.

Théorème 3.4 Sous les hypothèses 3.3 et 3.4, la suite de vecteurs aléatoires $\xi_m(L) := \sqrt{L}(\hat{q}_m^+(m; L) - q_m^+(L))$, converge $\mathbb{P}_{\alpha, \eta}$ -faiblement, quand $L \rightarrow \infty$, vers un vecteur aléatoire ξ_m , où $\xi_m \sim \mathcal{N}(\mathbf{0}, \Delta_m^{(L)})$ où $\Delta_m^{(L)}$ est une matrice de covariance de dimension $d_m \times d_m$, définie dans une forme diagonale par blocs

$$\Delta_m^{(L)} = \text{diag} \left\{ \frac{1}{\Psi_i(m-1)} K_i : i \in E \right\}, \quad (3.148)$$

où pour tout $i \in E$ la matrice K_i de dimension $d_{i|m}^* \times d_{i|m}^*$ est donnée par (3.112) et $\Psi_i(m-1)$ est donnée par (3.147) et supposée strictement positive.

Démonstration. Nous remarquons par (3.144) et (3.147) que

$$\pi_{i,0|m} = \Psi_i(m-1). \quad (3.149)$$

Notons également que

$$\begin{aligned} \mathbb{E}_{\alpha, \eta} N_{i,u}^B(m) &= \sum_{n=0}^{m-1} \mathbb{P}_{\alpha, \eta}(Z_n = i, U_n = u) \\ &= \sum_{n=u}^{m-1} \mathbb{P}_{\alpha, \eta}(Z_{n-u} = i, U_{n-u} = 0) \times \\ &\quad \mathbb{P}_{\alpha, \eta}(X_{N(n-u)+1} > u | Z_{n-u} = i, U_{n-u} = 0) \\ &= \bar{H}_i(u) \sum_{n=0}^{m-1} \mathbb{P}_{\alpha, \eta}(Z_n = i, U_n = 0) \\ &= \bar{H}_i(u) \mathbb{E}_{\alpha, \eta} N_{i,0}^B(m). \end{aligned} \quad (3.150)$$

Cela signifie à l'aide de (3.144) et (3.149) que

$$\pi_{i,u|m} = \bar{H}_i(u) \Psi_i(m-1), \quad (3.151)$$

ce qui montre que les rapports de $\bar{H}_i(u)$ avec $\pi_{i,u|m}$ ne dépendent pas de u . En mettant $(m, \pi_{i,u|m}, (\Psi_i(m-1))^{-1}, L, \Gamma_m^{(L)})$ à la place de $(\tilde{n}, \pi_{i,u}, \mu_{ii}, m, \Gamma_{\tilde{n}})$ dans l'énoncé du théorème 3.1, les étapes de la démonstration du présent théorème se déduisent à partir du lemme 3.2, de la même manière que pour démontrer le théorème 3.2 par le théorème 3.1. \square

Remarque 3.10 *i) Le terme $\Psi_i(m-1)$ qui apparaît dans l'expression (3.148) et qui est défini par (3.147), peut être exprimé en fonction du noyau semi-markovien q . Si $q^{(n)}$ représente la puissance de convolution d'ordre n de q , (voir Barbu et Limnios (2008), pages 48-49), alors nous avons la relation suivante*

$$\Psi_i(m-1) = \sum_{r \in E} \sum_{n, l \in T_m} \alpha_r q_{ri}^{(n)}(l). \quad (3.152)$$

Par conséquent, par (3.152) et (3.151) nous obtenons une forme alternative de (3.144),

$$\pi_{i,u|m} = \bar{H}_i(u) \sum_{r \in E} \sum_{n, l \in T_m} \alpha_r q_{ri}^{(n)}(l). \quad (3.153)$$

ii) Le terme $\Psi_i(m-1)$ par (3.147) est strictement positif ssi il existe un état j où $\alpha_j > 0$ et l'état $i \in E$ est accessible par j jusqu'à l'instant m . Si le support de la loi initiale est E , cela est toujours vrai. Sinon, dans le théorème 3.4 nous devons considérer seulement les états qui vérifient la propriété ci-dessus.

Pour mesurer la concentration de la loi du temps de séjour dans chaque état $i \in E$ dans l'ensemble fini $\{1, 2, \dots, m\}$, nous considérons l'EMV de sa fonction de répartition $H_i(m)$ à l'instant m . Puisque

$$H_i(m) = \sum_{(j,k) \in \mathbb{D}_{i|m}^*} q_{ij}(k), \quad (3.154)$$

l'EMV s'écrit

$$\widehat{H}_i(m; L) = \sum_{(j,k) \in \mathbb{D}_{i|m}^*} \widehat{q}_{ij}(k, m; L). \quad (3.155)$$

Cette forme additive est convenable pour déduire facilement les résultats asymptotiques de convergence et de normalité asymptotique.

Corollaire 3.1 *Sous les hypothèses 3.3 et 3.4, sous $\mathbb{P}_{\alpha, \eta}$,*

i) l'EMV $\widehat{\mathbf{H}}(m; L) := (\widehat{H}_i(m; L))_{i \in E}$ du vecteur $\mathbf{H}(m) := (H_i(m) : i \in E)$ est fortement convergent, et,

ii) la suite de vecteurs aléatoires $\zeta_m(L) := \sqrt{L}(\widehat{\mathbf{H}}(m; L) - \mathbf{H}(m))$ converge faiblement, quand $L \rightarrow \infty$, vers un vecteur aléatoire ζ_m , où $\zeta_m \sim \mathcal{N}(\mathbf{0}, Z_m^{(L)})$ et $Z_m^{(L)}$ est une matrice de covariance diagonale de dimension $s \times s$ donnée par

$$Z_m^{(L)} = \text{diag} \left\{ \frac{H_i(m)\widehat{H}_i(m)}{\Psi_i(m-1)} : i \in E \right\}. \quad (3.156)$$

Démonstration. Soit

$$\begin{aligned} \Phi_m : \mathcal{Q}_m^+ &\rightarrow \mathbb{R}^s \\ \mathbf{q}_m^+ &\mapsto \Phi_m(\mathbf{q}_m^+) = \mathbf{H}(m), \end{aligned} \quad (3.157)$$

où chaque composante $H_i(m) = \Phi_{i|m}$ est donnée par (3.154). Nous concluons directement par la forme additive de chaque composante de Φ_m qu'elle est continue et par suite la dérivée totale existe. Par la proposition 3.5, l'EMV $\widehat{\mathbf{q}}_m(m; L)$ est convergent ainsi par (3.157) et par le théorème de l'application continue, $\widehat{\mathbf{H}}(m; L)$ est convergent. De même, en utilisant le théorème 3.4 et la méthode delta, l'EMV $\widehat{\mathbf{H}}(m; L)$ est asymptotiquement normal de moyenne nulle et d'une matrice de covariance diagonale de dimension $s \times s$ donnée par

$$Z_m^{(L)} = \text{diag} \left\{ \frac{1}{\Psi_i(m-1)} \Phi'_{i|m} K_i (\Phi'_{i|m})^\top : i \in E \right\}. \quad (3.158)$$

Par (3.154) nous notons que

$$\Phi'_{i|m} = \mathbf{1}_{d_{i|m}^*}, \quad (3.159)$$

et par conséquent en tenant compte de (3.159) et de la forme de K_i donnée par (3.112) nous avons

$$\begin{aligned}
\Phi'_{i|m} K_i (\Phi'_{i|m})^\top &= \mathbf{1}_{d_{i|m}^*} K_i \mathbf{1}_{d_{i|m}^*}^\top \\
&= \sum_{(j,k),(l,u) \in \mathcal{D}_{i|m}^*} K_i((j,k), (l,u)) \\
&= \sum_{(j,k) \in \mathcal{D}_{i|m}^*} q_{ij}(k)(1 - q_{ij}(k)) - \sum_{\substack{(j,k),(l,u) \in \mathcal{D}_{i|m}^* \\ (j,k) \neq (l,u)}} q_{ij}(k)q_{il}(u) \\
&= \sum_{(j,k) \in \mathcal{D}_{i|m}^*} q_{ij}(k) - \sum_{(j,k),(l,u) \in \mathcal{D}_{i|m}^*} q_{ij}(k)q_{il}(u) \\
&= H_i(m) - H_i^2(m) \\
&= H_i(m)\bar{H}_i(m). \tag{3.160}
\end{aligned}$$

Par (3.158) et (3.160) nous obtenons la forme explicite de $Z_m^{(L)}$ donnée dans (3.156). \square

CONCLUSION DU CHAPITRE

Dans ce chapitre nous avons considéré l'estimation du maximum de vraisemblance pour le modèle semi-markovien non paramétrique. Nous avons donné l'EMV exact pour une observation \mathbf{Z}_0^m d'une trajectoire d'une CSM, ainsi que pour L trajectoires d'une CSM censurée à l'instant m . À l'aide de l'espace paramétrique naturel du processus couple (\mathbf{Z}, \mathbf{U}) nous avons pu montrer des bonnes propriétés asymptotiques, telles que la convergence et la normalité asymptotique.

En utilisant la même technique de reparamétrisation, nous ferons dans le chapitre suivant, l'étude théorique de l'EMV pour un certain type de modèles semi-markoviens cachés.

EMV POUR DES MODÈLES SEMI-MARKOVIENS CACHÉS

SOMMAIRE

4.1	DÉFINITION DU MODÈLE CACHÉ SM-MI-B GÉNÉRAL	80
4.2	REPARAMÉTRISATION DE MSMC PAR UNE SOUS-CLASSE DE MMC	84
4.3	CONVERGENCE DE L'EMV	86
4.4	NORMALITÉ ASYMPTOTIQUE DE L'EMV	89
	CONCLUSION	98

Dans les deux chapitres précédents nous avons considéré l'estimation du maximum de vraisemblance pour les chaînes markoviennes et semi-markoviennes finies. Une bonne compréhension de la structure markovienne et semi-markovienne permet de développer des modèles plus complexes dans l'intention d'être mieux adaptés aux applications réelles. Dans cette perspective, nous présentons dans ce chapitre une classe des modèles dont la particularité est d'avoir une structure cachée. Il s'agit des modèles markoviens et semi-markoviens cachés.

Au cours des dernières décennies, l'utilisation des modèles markoviens cachés (MMC) est devenue l'une des techniques les plus puissantes et les plus populaires dans plusieurs domaines scientifiques. Notamment, la reconnaissance de la parole (Rabiner (1989), Rabiner et Juang (1993)), la biologie (Krogh et al. (1994a), Krogh et al. (1994b)), le traitement d'images (Li et Gray (2000)) et plusieurs autres domaines (Bhar et Hamori (2000), Sansom (1998)). Pour une revue générale sur l'utilisation des processus de Markov cachés (PMC) dans la statistique et la théorie de l'information, nous renvoyons le lecteur à Ephraïm et Merhav (2002).

Le premier travail sur l'inférence statistique pour les MMC est celui de Baum et Petrie (1966), où les auteurs ont étudié un processus stochastique bidimensionnel (Z, Y) dont la première composante Z est une chaîne de Markov finie non observée directement, et la deuxième composante Y est un processus observé, où conditionnée par rapport à la première, forme une suite de variables aléatoires conditionnellement indépendantes avec un espace d'état fini (alphabet fini). Ils ont également démontré la

convergence et la normalité asymptotique de l'EMV pour ce modèle. Ces MMC sont souvent désignés comme étant des fonctions probabilistes des chaînes de Markov (CM). Les conditions de convergence ont été affaiblies dans Petrie (1969). Dans Leroux (1992), Bickel et al. (1998), les auteurs ont démontré la convergence et la normalité asymptotique de l'EMV pour un processus observable avec un espace d'état général.

En observant une trajectoire d'une certaine longueur, le principal objectif est d'estimer les probabilités de transition de la CM, ainsi que les distributions conditionnelles qui caractérisent la relation entre le processus caché et le processus observé. Bien que les MMC puissent être utiles dans plusieurs cas, ils souffrent d'une limitation importante. En effet, ils ne permettent que la distribution géométrique pour les temps de séjour dans les états du processus caché. Un tel modèle peut être inexact pour décrire plusieurs problèmes réels (voir Levinson (1986)).

Une généralisation assez naturelle des MMC est les modèles semi-markoviens cachés (MSMC). Dans ce cas, le processus caché Z n'est plus une chaîne de Markov mais une chaîne semi-markovienne. À l'opposé des MMC, les MSMC permettent n'importe quelle distribution pour les temps de séjour dans les différents états du processus caché, afin de réaliser une meilleure description de l'ensemble des données d'un problème réel.

Les MSMC ont été présentés pour la première fois dans Ferguson (1980). Depuis, plusieurs variantes ont été proposées. Tous ces travaux se basent sur la construction des algorithmes appropriés dans le but d'estimer les paramètres inconnus du modèle. Nous donnerons dans le chapitre suivant une revue bibliographique concernant les études algorithmiques.

En ce qui concerne les propriétés théoriques, Barbu et Limnios (2008) ont démontré la convergence et la normalité asymptotique de l'EMV pour MSMC non paramétrique dans le cas où le support des temps de séjour conditionnels est fini. Par ailleurs, une approche plus générale que celle dans Barbu et Limnios (2008) a été présentée dans Trevezas et Limnios (à paraître 2008a), où les auteurs ont démontré la convergence et la normalité asymptotique de l'EMV pour MSMC général avec dépendance des temps de récurrence en arrière. C'est ce travail qui sera présenté dans ce chapitre et qui consiste en les points suivants :

- i) Nous généralisons les résultats pour MSMC finis donnés dans Barbu et Limnios (2008) aux MSMC généraux, où l'espace d'état du processus observable est considéré d'être un sous-ensemble d'un espace euclidien. À cette fin, nous suivons la ligne de Leroux (1992) et Bickel et al. (1998) ;
- ii) nous permettons aux valeurs du processus observable Y_n , conditionné par rapport à la CSM, de dépendre d'une manière probabiliste, non seulement de l'état Z_n mais également de la durée de temps pendant laquelle le système a séjourné dans cet état (dépendance des temps de récurrence en arrière) ;
- iii) nous utilisons des représentations minimales pour les espaces paramétriques, qui sont impliqués dans notre analyse, prenant en compte les relations de dépendance parmi les paramètres. Également, nous

utilisons pour tous les couples (i, j) des constantes générales \tilde{n}_{ij} afin de spécifier le support des temps de séjour conditionnels, plutôt que d'étendre l'espace paramétrique avec les paramètres identiquement nuls ;

- iv) nous faisons une décomposition des éléments du noyau semi-markovien, différente de celle présentée dans Barbu et Limnios (2008).

Par iii) et iv) nous déduisons également des expressions explicites pour les matrices de covariance asymptotiques (en fonction du noyau semi-markovien), qui apparaissent dans le théorème de la limite centrale pour l'EMV des principales caractéristiques des chaînes semi-markoviennes (CSM).

Ce chapitre sera organisé de la façon suivante. Dans la section 1, nous introduisons les modèles semi-markoviens cachés généraux avec dépendance des temps de récurrence en arrière, ainsi que les notations et les conditions dont nous aurons besoin. Dans la section 2, nous reparamétrisons les MSMC par une sous-classe des MMC. Ensuite, nous démontrons, dans la section 3, la convergence forte de l'EMV de MSMC et des principales caractéristiques de CSM, à savoir, les temps de séjour conditionnels et les chaînes de Markov immergées. Leurs normalités asymptotiques seront également démontrées dans la dernière section de ce chapitre.

Les résultats de ce chapitre font partie de l'article Trevezas et Limnios (à paraître 2008a).

4.1 DÉFINITION DU MODÈLE CACHÉ SM-MI-B GÉNÉRAL

Dans cette section nous définissons les chaînes semi-markoviennes cachées générales avec dépendance des temps de récurrence en arrière.

Définition 4.1 Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $(\mathbf{Z}, \mathbf{Y}) := (Z_n, Y_n)_{n \in \mathbb{N}}$, un processus bidimensionnel qui vérifie les hypothèses suivantes :

1. \mathbf{Z} est une chaîne semi-markovienne non-observable avec espace d'état $E = \{1, 2, \dots, s\}$, donnée par la définition 3.4,
2. \mathbf{Y} est un processus à valeurs dans l'espace mesurable $(\mathcal{Y}, \mathcal{E})$, où il existe une mesure σ -finie ν qui domine les images de \mathbb{P} par \mathbf{Y} , et \mathbf{Y} conditionné par \mathbf{Z} constitue une suite d'éléments aléatoires conditionnellement markoviens d'ordre l ,
3. Tout Y_n conditionné par le vecteur \mathbf{Y}_0^{n-1} et par le processus \mathbf{Z} , dépend de \mathbf{Z} uniquement par Z_n et U_n , où U_n est le TRA à l'instant n , donné par (3.19).

Sous ces hypothèses, la chaîne (\mathbf{Z}, \mathbf{Y}) est appelée chaîne cachée générale de type¹ SM-MI-B. Si de plus, la mesure de probabilité \mathbb{P} est indexée par un paramètre $\theta \in \Theta$, où Θ est un ensemble euclidien, alors la famille des chaînes cachées (\mathbf{Z}, \mathbf{Y}) définies sur $(\Omega, \mathcal{F}, \mathbb{P}_\theta)$ est appelée modèle caché général de type SM-MI-B.

Nous désignons par q^θ le noyau semi-markovien du processus \mathbf{Z} , indexé par θ , qui vérifie les conditions (C1) et (C2) données après la définition 3.4 dans le chapitre 3. Également, nous définissons pour tout $i \in E$, $k \in \mathbb{N}$, $\mathbf{y}_0^{l-1} \in \mathcal{Y}^l$,

$$g_\theta(y_l | i, k, \mathbf{y}_0^{l-1}) := \frac{\mathbb{P}_\theta(Y_l \in dy_l | Z_n = i, U_n = k, \mathbf{Y}_0^{l-1} = \mathbf{y}_0^{l-1})}{\nu(dy_l)}, \quad (4.1)$$

la fonction de densité conditionnelle de Y_l conditionné par le vecteur $(Z_l, U_l, \mathbf{Y}_0^{l-1})$.

Le modèle caché général de type SM-MI-B défini ci-dessus est une extension du modèle général de type SM-MI, en ajoutant la dépendance des temps de récurrence en arrière. Pour illustrer la différence entre ces deux modèles, nous considérons la loi conditionnelle du vecteur \mathbf{Y}_0^m , conditionné par le processus \mathbf{Z} . Si le modèle caché est général de type SM-MI, alors la fonction de densité conditionnelle de \mathbf{Y}_0^m , conditionné par \mathbf{Z} , s'écrit

$$p_m^\theta(\mathbf{y}_0^m | \mathbf{Z}) = \prod_{i=0}^m g_\theta(y_i | Z_i, \mathbf{y}_{i-1}^{i-1}), \quad (4.2)$$

alors que pour le modèle caché général de type SM-MI-B, la densité s'écrit, par la propriété 3 de la définition 4.1, comme suit

$$p_m^\theta(\mathbf{y}_0^m | \mathbf{Z}) = \prod_{i=0}^m g_\theta(y_i | Z_i, U_i, \mathbf{y}_{i-1}^{i-1}). \quad (4.3)$$

Notons que dans la majorité des applications, la sélection de l'espace $(\mathcal{Y}, \mathcal{E}, \nu)$ est telle que :

¹le symbole B a été choisi du terme anglais « backward » pour indiquer la dépendance de \mathbf{Y} des temps de récurrence en arrière de \mathbf{Z} .

- $\mathcal{Y} \subset \mathbb{R}^s$, $\mathcal{E} = \mathcal{B}(\mathcal{Y})$ est la tribu borélienne sur \mathcal{Y} , et $\nu = \lambda$ est la mesure de Lebesgue.
- $\mathcal{Y} \subset \mathbb{N}$, fini ou non, \mathcal{E} est l'ensemble de toutes les parties de \mathcal{Y} , et ν est la mesure de comptage.

Dans ce chapitre, nous considérons l'étude théorique du comportement asymptotique de l'EMV qui correspond au modèle caché général de type SM-M0-B.

Nous désignons par θ_0 la vraie valeur du paramètre θ , et par $\mathbb{P}_0, \mathbb{E}_0$ lorsqu'il s'agit respectivement de $\mathbb{P}_{\theta_0}, \mathbb{E}_{\theta_0}$.

Dans la suite, nous supposons que la classe des CSM, en question, vérifie les deux conditions suivantes :

- Hypothèse 4.1**
- i) Il existe une constante $\tilde{n} \in \mathbb{N}$ telle que $q_{ij}^\theta(k) = 0$, pour tout $k > \tilde{n}$, $i, j \in E$, et $\theta \in \Theta$.
 - ii) La CSM \mathbf{Z} sous \mathbb{P}_0 est irréductible et stationnaire (voir définition 3.5).

Remarque 4.1 Nous donnons ici quelques conséquences de l'hypothèse 4.1.

- i) Le temps de séjour dans chaque état $i \in E$ de la CSM \mathbf{Z} sous \mathbb{P}_0 a un support fini, i.e., il existe des constantes $\tilde{n}_i \in \mathbb{T}_{\tilde{n}}$, pour tout $i \in E$ tel que

$$B^{(i)} = \{i\} \times T_{\tilde{n}_i},$$

où $B^{(i)}$ est donné par (3.20). Cela implique que l'espace d'état de (\mathbf{Z}, \mathbf{U}) est donné par

$$B = \bigcup_{i \in E} B^{(i)}, \quad (4.4)$$

- ii) La loi du temps de séjour conditionnel dans l'état i , lorsque le prochain état visité est j , sous \mathbb{P}_0 , a un support \mathbb{D}_{ij}^+ fini, i.e., il existe des constantes minimales $\tilde{n}_{ij} \in \mathbb{T}_{\tilde{n}_i}$, pour tout $(i, j) \in E_+^2$ tel que $\mathbb{D}_{ij}^+ \subseteq \mathbb{T}_{\tilde{n}_{ij}}$.
- iii) La chaîne de Markov (\mathbf{Z}, \mathbf{U}) est finie et donc par l'hypothèse 4.1-(ii) elle est également ergodique.

De cette remarque, nous déduisons que

$$\tilde{n}_i = \max_{j \in E} \tilde{n}_{ij}.$$

La constante \tilde{n}_{ij} exprime le temps maximal pendant lequel la CSM peut séjourner dans l'état i avant une transition directe dans l'état j . Dans la pratique, il y a deux possibilités pour ces constantes :

- i) elles sont censées être connues par les caractéristiques du système auquel ce modèle peut être appliqué ;
- ii) elles peuvent être imposées par l'expérimentateur comme approximation d'un système plus compliqué.

En fait, l'existence de ces bornes \tilde{n}_{ij} suffit pour les résultats théoriques qui suivront. Nous notons également que pour tout $(i, j) \in E^2 \setminus E_+^2$, nous avons $\tilde{n}_{ij} = 0$ et par conséquent aucune transition directe n'est possible de i à j .

Dans la suite, nous désignons par $\bar{H}_i^0(\cdot)$, μ_{ii}^0 , p_{ij}^0 , $f_{ij}^0(k)$ les vraies valeurs des caractéristiques du système semi-markovien définies respectivement par (3.1), (3.34), (3.12) et (3.14) dans le chapitre 3.

À l'opposé du chapitre précédent où nous avons considéré que $S_0 = 0$, ici, sous l'hypothèse 4.1, la CSM \mathbf{Z} est stationnaire. Le cas de stationnarité peut correspondre à un système semi-markovien qui a fonctionné pendant une période de temps assez grande (théoriquement infini). Nous pouvons construire une CSM stationnaire en posant la loi initiale du couple (Z_0, S_0) égale à la loi stationnaire du vecteur (Z_0, U_0) . De plus, les relations de dépendance concernant le noyau semi-markovien sous le contexte de stationnarité sont différentes. Nous désignons par $C_{q,sta}$ l'ensemble des relations de dépendance pour le vecteur q dues à la stationnarité seulement et $C_{q,sto}$ celles données par (3.39) dans le chapitre 3, ici pour tout $i \in E$.

L'ensemble $C_{q,sta}$ contient les relations pour tous les $j \in E$

$$\sum_{i \in E, j} \sum_{u \in T_{\bar{n}_i}} \frac{q_{ij}(u+1)}{\mu_{ii}} = \frac{1}{\mu_{jj}}, \quad (4.5)$$

qu'elles peuvent être déduites par les équations d'équilibre pour la matrice de transition P de la chaîne de Markov (\mathbf{Z}, \mathbf{U}) . En effet, ces équations sont données pour tout $j \in E$ par²

$$\sum_{i \in E, j} \sum_{u \in T_{\bar{n}_i}} \pi_{i,u} p_{i,u;j} = \pi_{j,0}, \quad (4.6)$$

et en utilisant (3.35), nous arrivons aux équations

$$\sum_{i \in E, j} \sum_{u \in T_{\bar{n}_i}} \frac{\bar{H}_i(u)}{\mu_{ii}} p_{i,u;j} = \frac{1}{\mu_{jj}}. \quad (4.7)$$

Ainsi, (4.5) peut être déduite par (3.25) et (4.7).

Dans la suite, nous notons par $C_{\eta,sta}$ les relations de dépendance dues à la stationnarité et qui sont exprimées en terme de η par les équations (4.7).

Nous faisons maintenant une hypothèse qui sera utile pour simplifier les représentations matricielles de ce chapitre. Notons cependant que pour les résultats théoriques de ce chapitre, c'est-à-dire, convergence et normalité asymptotique des estimateurs, cette hypothèse n'est pas nécessaire.

Hypothèse 4.2 *Le temps de séjour conditionnel dans l'état i , conditionné au prochain état visité j sous \mathbb{P}_θ a comme support $\mathbb{D}_{ij}^+ = \mathbb{T}_{\bar{n}_{ij}}$.*

Sous cette hypothèse, l'ensemble des indices (j, k) pour lesquels $q_{ij}(k) > 0$, s'écrit

$$\mathbb{D}_i^+ = \bigcup_{j: (i,j) \in E_+^2} \{j\} \times \mathbb{T}_{\bar{n}_{ij}}.$$

²nous excluons les équations qui correspondent aux états (j, k) pour $k \neq 0$, qui ne donnent pas des relations de dépendance supplémentaires.

Également, nous avons

$$\mathbb{D}^+ := \bigcup_{i \in E} (\{i\} \times \mathbb{D}_i^+).$$

Nous notons par d^+ et d_i^+ les cardinaux de \mathbb{D}^+ et \mathbb{D}_i^+ respectivement.

Soit T un ensemble fini des indices de cardinal d_g . Nous posons $d_{q,g}^+ = d^+ + d_g$.

Nous précisons maintenant l'espace paramétrique du modèle que nous considérons, en tenant compte des relations de dépendance. Cet espace s'écrit

$$\Theta := \mathcal{Q} \times \Theta_g \subset \mathbb{R}^{d_{q,g}^+},$$

où

$$\mathcal{Q} := \{\mathbf{q} = (q_{ij}(k)) \in \mathbb{R}^{d^+} : C_{q,\text{sto}} \wedge C_{q,\text{sta}}\}, \quad (4.8)$$

$$\Theta_g := \{\boldsymbol{\theta}_g = (\theta_{g;t}) \in \mathbb{R}^{d_g} : t \in T\}. \quad (4.9)$$

L'espace \mathcal{Q} paramétrise les éléments du noyau semi-markovien. Si nous notons pr_{ijk} la fonction projection, nous avons alors $q_{ij}^\theta(k) = pr_{ijk}(\boldsymbol{\theta}) = q_{ij}(k)$. L'espace Θ_g est formé par les paramètres qui caractérisent les densités conditionnelles $g_\theta(\mathbf{y} \mid i, u)$. Cela peut considérer des densités d'une famille paramétrique spécifique, des différentes familles paramétriques ou aussi des probabilités de transition notamment quand \mathcal{Y} est un espace d'état fini. Dans le cas le plus simple d'une famille paramétrique spécifique, nous avons $g_\theta(\mathbf{y} \mid i, u) := g(\mathbf{y} \mid \theta_{i,u})$, où $\theta_{i,u}$ appartient à un ensemble euclidien. Dans ce cas, l'ensemble des indices T qui apparaît dans Θ_g se compose de tous les couples possibles de $(i, u) \in B$.

Dans ce qui suit, nous supposons pour simplifier que le cardinal d_B de T est $d_B = \sum_i \tilde{n}_i$, c'est-à-dire un paramètre de dimension un pour chaque densité conditionnelle. Nous notons par A_g l'ensemble commun des valeurs des paramètres $\theta_{i,u}$.

Pour donner des représentations minimales nous devons prendre en compte $C_{q,\text{sto}}$. Dans $C_{q,\text{sto}}$ il y a s relations de dépendance linéaires, une pour chaque $i \in E$. Par conséquent, nous devons exprimer s composantes du noyau semi-markovien en fonction des autres. À ce but, nous désignons par

$$J_i = \{j \in E : \tilde{n}_{ij} = \tilde{n}_i\}. \quad (4.10)$$

Nous pouvons choisir un $j_i \in J_i$, pour tout $i \in E$, et exprimer les s éléments comme suit :

$$q_{ij_i}(\tilde{n}_i) = 1 - \sum_{(j,k) \in \mathbb{D}_i^*} q_{ij}(k), \quad (4.11)$$

où pour tout $i \in E$

$$\mathbb{D}_i^* = \mathbb{D}_i^+ \setminus \{(j_i, \tilde{n}_i)\}.$$

Maintenant, nous désignons par

$$\mathbb{D}^* = \bigcup_{i \in E} (\{i\} \times \mathbb{D}_i^*),$$

ainsi que

$$d^* = |\mathbb{D}^*|.$$

Ainsi, une représentation minimale de \mathcal{Q} est donnée par

$$\mathcal{Q}^* := \{\mathbf{q}^* = (q_{ij}(k)) \in \mathbb{R}^{d^*} : C_{\mathbf{q}, \text{sta}}\}, \quad (4.12)$$

où les éléments de \mathbf{q}^* sont ordonnés selon l'ordre lexicographique de \mathbb{D}^* . Également une représentation minimale de Θ est donnée par

$$\Theta^* = \mathcal{Q}^* \times \Theta_g. \quad (4.13)$$

4.2 REPARAMÉTRISATION DE MSMC PAR UNE SOUS-CLASSE DE MMC

Dans cette section, nous décrivons la reparamétrisation du modèle semi-markovien caché général avec dépendance des temps de récurrence en arrière qui vérifie les hypothèses 4.1 et 4.2, par l'espace paramétrique d'une sous-classe de modèles markoviens cachés. À cette fin, il est suffisant de faire la reparamétrisation pour la CSM \mathbf{Z} par l'espace paramétrique d'une sous-classe de CM. Dans la section 3.3 du chapitre 3 nous avons décrit la reparamétrisation dans le cas où le support des lois des temps de séjour pour les états de \mathbf{Z} est infini. Dans ce chapitre, nous considérons le cas fini exprimé par l'hypothèse 4.1.

Sous cette hypothèse et la remarque 4.1, la matrice de transition de la CM finie (\mathbf{Z}, \mathbf{U}) à espace d'état B , donnée par (4.4), est de dimension $d_B \times d_B$. Le lemme suivant indique les positions des éléments non identiquement nuls que nous utilisons pour la reparamétrisation du vecteur \mathbf{q} .

Lemme 4.1 *Sous l'hypothèse 4.1, la matrice de transition P de la CM (\mathbf{Z}, \mathbf{U}) , à espace d'état B donné par (4.4), dont les probabilités de transition non identiquement nulles sont données par (3.23), peut être représentée par blocs $P = (P_{ij})_{i,j \in E}$, où P_{ij} est une matrice de dimension $\tilde{n}_i \times \tilde{n}_j$, comme suit :*

pour $i = j$,

$$P_{ii} = \begin{pmatrix} 0 & p_{i,0;i} & 0 & \dots & 0 \\ 0 & 0 & p_{i,1;i} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & p_{i,\tilde{n}_i-2;i} \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \quad (4.14)$$

et pour $i \neq j$,

$$P_{ij} = \begin{pmatrix} p_{i,0;j} & 0 & \dots & 0 \\ p_{i,1;j} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ p_{i,\tilde{n}_{ij}-1;j} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}. \quad (4.15)$$

Remarque 4.2 *i) Les probabilités de transition $p_{i,u;j}$ données ci-dessus sont exprimées en fonction de \mathbf{q} dans la proposition 3.2 pour le cas infini. Alors que dans ce chapitre, nous avons $U_n \in T_{\tilde{n}_i}$.*

ii) Par l'expression (3.26) de la fonction de survie $\tilde{H}_i(\cdot)$ et (4.4), nous concluons que pour tout $i \in E$, $u \in T_{\tilde{n}_i-1}$, nous avons $p_{i,u;i} > 0$, alors que $p_{i,\tilde{n}_i-1;i} = 0$.

iii) Pour tout $(i, j) \in E^2 \setminus E_+^2$, nous avons $\tilde{n}_{ij} = 0$, et par conséquent, la matrice P_{ij} est nulle, alors que si $\tilde{n}_{ij} = \tilde{n}_i$, donc la première colonne de P_{ij} n'a pas des éléments identiquement nuls.

iv) Sous l'hypothèse 4.2, nous concluons que tous les éléments $p_{i,u;j}$ de la première colonne de la matrice P_{ij} sont strictement positifs sous \mathbb{P}_0 .

Nous construisons ici l'espace paramétrique qui correspond à la CM (\mathbf{Z}, \mathbf{U}) en tenant en compte également la stationnarité. Nous désignons par $C_{\eta, \text{sta}}$ l'ensemble des relations de dépendance pour le vecteur $\boldsymbol{\eta}$ dues uniquement à la stationnarité et qui sont données par (4.7) et $C_{\eta, \text{sto}}$ celles données par (3.51) pour tout $(i, u) \in B$. Nous remarquons également que le nombre d de paramètres non identiquement nuls de la matrice P donnée par (4.14) et (4.15) est

$$d = \sum_{(i,j) \in E_+^2} \tilde{n}_{ij} + \sum_{i \in E} \tilde{n}_i - s = d^+ + d_B - s.$$

L'espace paramétrique qui correspond au paramètre $\boldsymbol{\eta}$ s'écrit

$$\mathcal{H} := \left\{ \boldsymbol{\eta} = (p_{i,u;j}) \in \mathbb{R}^d : C_{\eta, \text{sto}} \wedge C_{\eta, \text{sta}} \right\},$$

et alors la reparamétrisation du modèle est effectuée par l'espace

$$\tilde{\Theta} := \mathcal{H} \times \Theta_g \subset \mathbb{R}^{\tilde{d}},$$

où $\tilde{d} = d + d_B$.

Dans ce chapitre nous donnons des représentations minimales différentes de celles du chapitre précédent. Puisque P est une matrice stochastique, il y a exactement d_B relations de dépendance linéaire parmi les éléments de P , une pour chaque ligne. Pour le moment, nous précisons la notation et dans la proposition suivante nous explicitons le choix de la représentation minimale.

Notons

$$\mathcal{H}^* := \left\{ \boldsymbol{\eta}^* = (p_{i,u;j}) \in \mathbb{R}^{d^*} : C_{\eta, \text{sta}} \right\}, \quad (4.16)$$

où $d^* = d - d_B$, et donc la paramétrisation minimale de $\tilde{\Theta}$ est donnée par

$$\tilde{\Theta}^* := \mathcal{H}^* \times \Theta_g \subset \mathbb{R}^{\tilde{d}}.$$

Dans le chapitre 3 nous avons exprimé le vecteur \mathbf{q} en fonction de $\boldsymbol{\eta}$ par (3.55). Dans le cadre du présent chapitre, nous précisons la fonction à l'aide de laquelle nous identifions la classe de CSM à la classe de CM qui est décrite par le lemme 4.1.

Proposition 4.1 *Il existe une fonction continue Ψ_1 de \mathcal{H}^* dans \mathcal{Q}^* , analytique dans son intérieur $(\mathcal{H}^*)^0$, qui reparamétrise les chaînes semi-markoviennes \mathbf{Z} à espace d'état E et de noyau semi-markovien qui vérifie l'hypothèse 4.1, par l'espace paramétrique \mathcal{H}^* des chaînes markoviennes à espace d'état B et de matrice de transition donnée par P (voir lemme 4.1).*

Démonstration. D'abord, nous précisons le choix de l'espace paramétrique minimal \mathcal{H}^* , défini par (4.16). En fait, pour former \mathcal{H}^* , nous excluons de \mathcal{H} pour tout $(i, u) \in B$ les paramètres $p_{i,u;j_i}$, où j_i est un choix quelconque de J_i , donné par (4.10). Donc, $p_{i,u;j_i}$ est donné en fonction de η^* comme suit :

$$p_{i,u;j_i} = \begin{cases} 1 - \sum_{\substack{j:\tilde{n}_{ij} \geq u+1 \\ j \neq j_i}} p_{i,u;j} - p_{i,u;i} & \text{si } u \in T_{\tilde{n}_i-1}, \\ 1 - \sum_{j \in G_i} p_{i,u;j} & \text{si } u = \tilde{n}_i - 1, \end{cases}$$

où $G_i = \{j : j \neq j_i, \tilde{n}_{ij} = \tilde{n}_i\}$.

Nous définissons maintenant la fonction souhaitée :

$$\begin{aligned} \Psi_1 : \mathcal{H}^* &\rightarrow \mathcal{Q}^* \\ \eta^* &\mapsto \Psi_1(\eta^*) = q^*, \end{aligned} \quad (4.17)$$

où les composantes de Ψ_1 , pour tout $(i, j) \in E_+^2$, sont données par

$$q_{ij}(1) = \begin{cases} p_{i,0;j} & \text{si } j \neq j_i, \\ 1 - \sum_{j \in G_i} p_{i,0;j} - p_{i,0;i} & \text{si } j = j_i, \end{cases} \quad (4.18)$$

$$q_{ij}(k) = \begin{cases} p_{i,k-1;j} \prod_{r=0}^{k-2} p_{i,r;i} & \text{si } j \neq j_i, 2 \leq k \leq \tilde{n}_{ij}, \\ (1 - \sum_{j \in G_i} p_{i,k-1;j} - p_{i,k-1;i}) \prod_{r=0}^{k-2} p_{i,r;i} & \text{si } j = j_i, 2 \leq k < \tilde{n}_i. \end{cases} \quad (4.19)$$

Par (4.18) et (4.19), nous concluons que Ψ_1 est une fonction continue sur \mathcal{H}^* et analytique dans son intérieur. \square

Remarque 4.3 *Les s paramètres de \mathcal{Q} qui ont été exclus pour obtenir \mathcal{Q}^* peuvent être exprimés en fonction de η^* comme suit :*

$$q_{ij}(\tilde{n}_i) = (1 - \sum_{j \in G_i} p_{i,\tilde{n}_i-1;j}) \prod_{r=0}^{\tilde{n}_i-1} p_{i,r;i}. \quad (4.20)$$

4.3 CONVERGENCE DE L'EMV

Dans cette section nous démontrons la convergence de l'EMV $\hat{\theta}(m)$ qui correspond au modèle semi-markovien général de type SM-M0-B. Par

suite, nous déduisons la convergence de l'EMV $\hat{p}^e(m)$ du vecteur p^e qui contient les probabilités de transition de la CM immergée J , et de l'EMV $\hat{f}(m)$ du vecteur f qui contient les probabilités des temps de séjour conditionnelles (voir proposition 3.1).

Par la proposition 4.1 de la section précédente, le modèle caché général de type SM-M0-B peut être décrit par un modèle caché général de type M1-M0 décrit dans la section 4.2. En outre, la stationnarité de (Z, Y) implique la stationnarité de $((Z, U), Y)$, ce qui transfère ainsi la stationnarité du premier modèle à l'autre. En ce qui concerne l'espace paramétrique Θ^* décrit par (4.13) nous supposons qu'il est compact. En fait, par (4.12) et (3.39) il est clair que \mathcal{Q}^* est compact, ce qui implique par (4.13) que Θ^* est compact si et seulement si Θ_g est compact. Si ce n'est pas le cas, nous pouvons toujours utiliser une technique standard de compactification (voir Leroux (1992) et Kiefer et Wolfowitz (1956)).

Dans cette section nous gardons la formulation du problème de maximisation en utilisant θ et $\hat{\theta}$ et dans la section suivante, puisque nous calculons des dérivées, nous utiliserons les représentations minimales θ^* et $\tilde{\theta}^*$.

Nous donnons maintenant la fonction de vraisemblance correspondante à l'observation $\{Y_0^m = y_0^m\}$,

$$\mathcal{L}_m(\theta) = \sum_{(i,u)_0^m} \pi_{i_0,u_0}^\theta \prod_{j=0}^{m-1} p_{i_j,u_j;i_{j+1}}^\theta \prod_{j=0}^m g(y_j | \theta_{i_j,u_j}), \quad (4.21)$$

où $\pi_{i,u}^\theta$ sont les probabilités stationnaires de (Z, U) données par (3.35), ici exprimées en fonction de θ .

L'EMV $\hat{\theta}(m)$ maximise $\mathcal{L}_m(\theta)$ sur Θ . Dans le meilleur cas, c'est une classe formée par les paramètres θ induits par des permutations d'une valeur spécifique, qui maximise la probabilité indiquée. Pour cette raison, nous définissons une relation d'équivalence « \sim » dans Θ où $\theta_1 \sim \theta_2$ si et seulement si $\mathbb{P}_{\theta_1} = \mathbb{P}_{\theta_2}$. Par conséquent, les résultats pour les estimateurs doivent être compris dans le cadre de Θ^* / \sim , c'est-à-dire, dans la topologie quotient induite par cette équivalence \sim dans Θ^* .

Maintenant, nous donnons les conditions sous lesquelles nous pouvons déduire que l'EMV $\hat{\theta}(m)$ est convergent. Ces conditions sont données dans Leroux (1992) (voir l'annexe A.4), et elles sont adaptées ici à notre modèle.

(B1) (*Condition d'identifiabilité*) La famille formée par les mélanges d'au plus d_B éléments de $\{g(y | \theta), \theta \in A_g\}$ est identifiable.

(B2) La fonction de densité $g(y | \theta)$ est continue dans A_g , pour tout $y \in \mathcal{Y}$.

(B3) $\mathbb{E}_0 \left| \log g(Y_0 | \theta_{i,u}^0) \right| < \infty$, pour tout $(i, u) \in B$.

(B4) Pour tout $\theta \in A_g$, $\mathbb{E}_0 \left[\sup_{|\theta' - \theta| < \delta} \left(\log g(Y_0 | \theta') \right)^+ \right] < \infty$, pour un certain $\delta > 0$, où $x^+ = \max(x, 0)$.

Dans notre modèle, l'identifiabilité est assurée si les hypothèses 4.1 et 4.2, et la condition (B1) sont vérifiées, ainsi que les $\theta_{i,u}^0$ sont distincts. Pour les détails nous renvoyons le lecteur à Leroux (1992).

Dans le théorème suivant nous déduisons la convergence de l'EMV.

Théorème 4.1 *Sous les hypothèses 4.1 et 4.2, et les conditions (B1)-(B4), l'EMV $\widehat{\theta}(m)$ est un estimateur fortement convergent (au sens de la topologie quotient) de θ_0 , et par conséquent, $\widehat{q}(m)$ est un estimateur fortement convergent de q^0 .*

Démonstration. Nous constatons que la fonction de vraisemblance correspondante à l'observation $\{Y_0^m = y_0^m\}$, donnée par (4.21), reparamétrisée par $\tilde{\theta}$, prend la forme

$$\mathcal{L}_m(\tilde{\theta}) = \sum_{(i,u)_0^m} \pi_{i_0,u_0}^{\tilde{\theta}} \prod_{j=0}^{m-1} p_{i_j,u_j;i_{j+1}} \prod_{j=0}^m g(y_j | \theta_{i_j,u_j}), \quad (4.22)$$

qui correspond à la fonction de vraisemblance de la même observation générée par un modèle caché général de type M1-M0 qui est défini par le processus $((Z, U), Y)$. En conséquence, si les conditions 1-6 de l'article de Leroux (1992) (voir également l'annexe A.4) sont vérifiées, alors l'EMV $\widehat{\theta}(m)$ de $\tilde{\theta}$ est convergent par le théorème 3, section 6, du même article. En effet, il est clair que la condition 1 de Leroux peut être déduite des hypothèses 4.1 et 4.2. Les conditions 2 et 3 sont identiques aux (B1) et (B2). La condition 4 peut être déduite du fait que les probabilités de transition données dans la proposition 3.2 sont des fonctions continues du vecteur q . Enfin, les conditions 5 et 6 sont identiques respectivement aux (B3) et (B4). Par la convergence de $\widehat{\theta}(m)$ et la continuité de Ψ_1 définie par (4.17), nous concluons par le théorème de l'application continue que l'EMV $\widehat{\theta}(m)$ de θ est fortement convergent. \square

Maintenant, nous transférons le résultat de ce théorème à l'estimation des probabilités de transition p_{ij} de la CM immergée J , définies par (3.12) et les probabilités des temps de séjour conditionnelles $f_{ij}(k)$, définies par (3.14).

Selon la remarque 4.1(ii), nous avons

$$p_{ij} = \begin{cases} \sum_{k \in \mathbb{T}_{\tilde{n}_{ij}}} q_{ij}(k) & \text{si } \tilde{n}_{ij} > 0, \\ 0 & \text{si } \tilde{n}_{ij} = 0, \end{cases} \quad (4.23)$$

et

$$f_{ij}(k) = \begin{cases} \frac{q_{ij}(k)}{p_{ij}} & \text{si } \tilde{n}_{ij} > 0, k \in \mathbb{T}_{\tilde{n}_{ij}}, \\ 0 & \text{si } \tilde{n}_{ij} = 0. \end{cases} \quad (4.24)$$

Pour le corollaire suivant nous aurons besoin de quelques notations supplémentaires. Nous notons par p^e et f les formes vectorielles des éléments p_{ij} et $f_{ij}(k)$ non nuls, qui sont donnés respectivement par (4.23) et (4.24). Notons également par $\widehat{p}^e(m)$ et $\widehat{f}(m)$ leurs EMV, et p^{e0} et f^0 les vraies

valeurs de \mathbf{p}^e et \mathbf{f} .

Nous désignons pour tout $i \in E$,

$$c_i = \text{card}\{j : \tilde{n}_{ij} > 0\}, \text{ et } \tilde{c} = \sum_i c_i. \quad (4.25)$$

Alors, les résultats asymptotiques suivants sont vérifiés :

Corollaire 4.1 *Sous les hypothèses 4.1 et 4.2, et les conditions (B1)-(B4),*
i) *l'EMV $\widehat{\mathbf{p}}^e(m)$ des probabilités de transition de la chaîne markovienne \mathbf{J} est un estimateur fortement convergent de \mathbf{p}^{e0} .*
ii) *l'EMV $\widehat{\mathbf{f}}(m)$ des probabilités des temps de séjour conditionnelles est un estimateur fortement convergent de \mathbf{f}^0 .*

Démonstration. i) Nous définissons la fonction $\Phi : \mathcal{Q}^* \rightarrow \mathbb{R}^{\tilde{c}}$, où ses composantes sont définies par

$$\begin{aligned} \Phi_{ij} : \mathcal{Q}^* &\rightarrow \mathbb{R} \\ \mathbf{q}^* &\mapsto p_{ij} = \sum_{k \in \mathbb{T}_{\tilde{n}_{ij}}} q_{ij}(k). \end{aligned} \quad (4.26)$$

Nous déduisons par la propriété de l'invariance de l'EMV que

$$\widehat{\mathbf{p}}^e(m) = \widehat{\Phi(\mathbf{q}^*)}(m) = \Phi(\widehat{\boldsymbol{\theta}}^*(m)).$$

Par le théorème 4.1 et le théorème de l'application continue pour la fonction Φ (fonction additive) nous concluons que, quand $m \rightarrow \infty$,

$$\widehat{\mathbf{p}}^e(m) \xrightarrow{\mathbb{P}_0\text{-p.s.}} \mathbf{p}^{e0}.$$

ii) Nous définissons la fonction $T : \mathcal{Q}^* \rightarrow \mathbb{R}^{d^+}$, où ses composantes sont définies par

$$\begin{aligned} T_{ijk} : \mathcal{Q}^* &\rightarrow \mathbb{R} \\ \mathbf{q}^* &\mapsto f_{ij}(k) = \frac{q_{ij}(k)}{p_{ij}}. \end{aligned} \quad (4.27)$$

Nous remarquons que

$$T_{ijk}(\mathbf{q}^*) = \frac{pr_{ijk}(\mathbf{q}^*)}{\Phi_{ij}(\mathbf{q}^*)},$$

ce qui montre que T est continue. Par le théorème 4.1 et le théorème de l'application continue pour la fonction T nous concluons que, quand $m \rightarrow \infty$,

$$\widehat{\mathbf{f}}(m) \xrightarrow{\mathbb{P}_0\text{-p.s.}} \mathbf{f}^0.$$

□

4.4 NORMALITÉ ASYMPTOTIQUE DE L'EMV

Dans cette section nous démontrons la normalité asymptotique de l'EMV $\widehat{\boldsymbol{\theta}}(m)$ qui correspond au modèle semi-markovien général de type SM-M0-B. Pour cela nous aurons besoin de deux notions étroitement liées à l'estimation du maximum de vraisemblance. L'une concerne le taux d'entropie

pour un processus stochastique et l'autre concerne la divergence généralisée de Kullback-Leibler. Pour simplifier leurs expressions, nous utilisons le fait que la stationnarité de (\mathbf{Z}, \mathbf{Y}) et $((\mathbf{Z}, \mathbf{U}), \mathbf{Y})$ permet d'indexer $(Z_n, Y_n)_n$ et $((Z_n, U_n), Y_n)_n$ par \mathbf{Z} . Dans ce cas, le taux d'entropie du processus stochastique (\mathbf{Z}, \mathbf{Y}) est défini par

$$-\mathbb{H}(\theta_0^*) := -\mathbb{E}_0[\log \mathbb{P}_0(Y_0 | \mathbf{Y}_-)], \quad (4.28)$$

et la divergence généralisée de Kullback-Leibler par

$$\mathbb{H}_0(\theta^*) := \mathbb{E}_0[\log \mathbb{P}_\theta(Y_0 | \mathbf{Y}_-)], \quad \theta^* \in \Theta^*, \quad (4.29)$$

où nous avons utilisé la notation $\mathbf{Y}_- := (Y_n)_{n \in \mathbb{Z}_-}$ et le fait que $\mathbb{P}_\theta = \mathbb{P}_{\theta^*}$. Pour l'utilité de ces notions concernant la démonstration de la convergence de l'EMV du paramètre θ dans un modèle caché général de type M1-M0, voir Leroux (1992), ainsi que Baum et Petrie (1966) dans le cas fini.

Nous désignons par $\sigma(\theta_0^*)$ l'opposé de la matrice hessienne de $\mathbb{H}_0(\theta^*)$, calculée en θ_0^* , i.e.,

$$\sigma(\theta_0^*) = \left(\sigma_{u,v}(\theta_0^*) \right)_{u,v} := - \left(\frac{\partial^2 \mathbb{H}_0(\theta^*)}{\partial \theta_u \partial \theta_v} \Big|_{\theta^* = \theta_0^*} \right)_{u,v}. \quad (4.30)$$

À l'ensemble des conditions données dans la section précédente, nous ajoutons ici d'autres conditions en nous basant sur l'article de Bickel et al. (1998) (voir l'annexe A.5), afin d'assurer la normalité asymptotique de l'EMV du paramètre θ . Ces conditions, adaptées à notre modèle, peuvent être énoncées comme suit :

(C1) La CM \mathbf{Z} est apériodique (voir définition 3.5).

(C2) Les densités conditionnelles $g(y | \theta_{i,u})$ sont deux fois continûment dérivables par rapport à $\theta_{i,u} \in A_g$, dans un voisinage de $\theta_{i,u}^0$, pour toutes les valeurs possibles de $(i, u) \in B$, $y \in \mathcal{Y}$.

(C3) Il existe un $\delta > 0$, pour tout $(i, u) \in B$ tel que

$$\begin{aligned} \text{i)} \quad & \mathbb{E}_0 \left[\sup_{|\theta_{i,u} - \theta_{i,u}^0| < \delta} \left(\frac{d}{d\theta_{i,u}} \log g(Y_0 | \theta_{i,u}) \right)^2 \right] < \infty, \\ \text{ii)} \quad & \mathbb{E}_0 \left[\sup_{|\theta_{i,u} - \theta_{i,u}^0| < \delta} \left| \frac{d^2}{d\theta_{i,u}^2} \log g(Y_0 | \theta_{i,u}) \right| \right] < \infty, \\ \text{iii)} \quad & \int \sup_{|\theta_{i,u} - \theta_{i,u}^0| < \delta} \left| \frac{d^j}{d\theta_{i,u}^j} g(y | \theta_{i,u}) \right| \nu(dy) < \infty, \quad \text{pour } 1 \leq j \leq 2. \end{aligned}$$

(C4) Pour $\theta_0^* \in \Theta^*$, il existe un $\delta > 0$ tel que, si

$$r_0(y) := \sup_{\|\theta^* - \theta_0^*\| < \delta} \max_{(i_1, u_1), (i_2, u_2)} \frac{g(y | \theta_{i_1, u_1})}{g(y | \theta_{i_2, u_2})}, \quad (4.31)$$

alors, $\mathbb{P}_0(r_0(Y_0) = \infty \mid (Z_0, U_0) = (i, u)) < 1$, pour tout $(i, u) \in B$.

(C5) La vraie valeur θ_0^* est un point intérieur de Θ^* .

(C6) La matrice $\sigma(\theta_0^*)$ est non singulière.

Remarque 4.4 *Les conditions (C1)-(C4) peuvent être substituées par des conditions similaires pour les densités conditionnelles $g_\theta(y \mid i, u)$, comme elles apparaissent dans Bickel et al. (1998).*

Dans ce qui suit, nous aurons besoin de considérer la fonction Ψ_1 qui est donnée par (4.17) comme une fonction sur $\tilde{\Theta}^*$. Pour cela, nous définissons

$$\Psi : \tilde{\Theta}^* \rightarrow \Theta^*, \quad \text{où } \Psi(\eta^*, \theta_g) = (\Psi_1(\eta^*), \theta_g). \quad (4.32)$$

Nous remarquons à l'aide de la proposition 4.1 que la fonction Ψ est continue et analytique dans l'intérieur de $\tilde{\Theta}^*$. Également, nous déduisons par la proposition 3.2 que la fonction inverse existe et qu'elle est continue sur Θ^* et analytique dans l'intérieur de Θ^* .

La fonction Ψ^{-1} est donnée par

$$\Psi^{-1} : \Theta^* \rightarrow \tilde{\Theta}^*, \quad \text{où } \Psi^{-1}(q^*, \theta_g) = (\Psi_1^{-1}(q^*), \theta_g). \quad (4.33)$$

Nous désignons par Ψ' la matrice jacobienne de la fonction Ψ calculée en $\tilde{\theta}_0^*$, et $\sigma(\tilde{\theta}_0^*)$ l'opposé de la matrice hessienne de la divergence généralisée de Kullback-Leibler, définie par (4.29)³, calculée en $\tilde{\theta}_0^*$.

D'abord, nous démontrons la normalité asymptotique de l'EMV $\hat{\theta}^*(m)$ du paramètre $\tilde{\theta}_0^*$.

Théorème 4.2 *Sous les hypothèses 4.1 et 4.2 et les conditions (B1)-(B4) et (C1)-(C6), l'EMV $\hat{\theta}^*(m)$ de $\tilde{\theta}_0^*$ est asymptotiquement normal, i.e., quand $m \rightarrow \infty$,*

$$\sqrt{m}(\hat{\theta}^*(m) - \tilde{\theta}_0^*) \overset{\tilde{\theta}_0^*}{\rightsquigarrow} \mathcal{N}(0, \sigma(\tilde{\theta}_0^*)^{-1}). \quad (4.34)$$

Démonstration. En utilisant la proposition 3.2, l'EMV $\hat{\theta}^*(m)$ est asymptotiquement normal par le théorème 1 de la section 3 de Bickel et al. (1998), si les conditions de régularité (A1)-(A6) posées dans leur article (voir l'annexe A.5) sont vérifiées pour la chaîne $((\mathbf{Z}, \mathbf{U}), Y)$ paramétrisée par $\tilde{\Theta}^*$, et également si la matrice $\sigma(\tilde{\theta}_0^*)$ est non singulière. Les hypothèses 4.1 et 4.2 et la condition (C1) assurent que le processus (\mathbf{Z}, \mathbf{U}) est une chaîne de Markov finie irréductible et apériodique, et par conséquent, la condition (A1) de Bickel et al. (1998) est vérifiée. Par la continuité de l'application Ψ , définie par (4.32), nous déduisons que si θ_0^* est un point intérieur de Θ^* , alors $\tilde{\theta}_0^*$ est un point intérieur de $\tilde{\Theta}^*$. Par conséquent, la condition (C5) implique (A5). En outre, par (C6) et (4.33) nous déduisons que $\sigma(\tilde{\theta}_0^*)$ est non singulière.

Nos conditions (B1)-(B4) avec les hypothèses 4.1 et 4.2 impliquent que (A6) de Bickel et al. (1998) est vérifiée (voir également la démonstration

³considérée ici en fonction de $\tilde{\theta}^*$.

du théorème 4.1). Les autres conditions (A2)-(A4) sont vérifiées par (C2)-(C4) de notre modèle. \square

Dans ce qui suit, nous aurons besoin de la décomposition suivante de la matrice $\sigma(\tilde{\theta}_0^*)^{-1}$:

$$\sigma(\tilde{\theta}_0^*)^{-1} = \begin{bmatrix} \sigma(\tilde{\theta}_0^*)_{11}^{-1} & \sigma(\tilde{\theta}_0^*)_{12}^{-1} \\ \sigma(\tilde{\theta}_0^*)_{21}^{-1} & \sigma(\tilde{\theta}_0^*)_{22}^{-1} \end{bmatrix}, \quad (4.35)$$

où la partition de l'espace paramétrique correspond au vecteur $(\boldsymbol{\eta}^*, \boldsymbol{\theta}_g)$.

Maintenant, nous utilisons le résultat du théorème 4.2 ci-dessus pour démontrer la normalité asymptotique de l'EMV $\hat{\theta}^*(m)$ du paramètre θ_0^* .

Théorème 4.3 *Sous les hypothèses 4.1 et 4.2 et les conditions (B1)-(B4) et (C1)-(C6), l'EMV $\hat{\theta}^*(m)$ de θ_0^* est asymptotiquement normal, i.e., quand $m \rightarrow \infty$,*

$$\sqrt{m}(\hat{\theta}^*(m) - \theta_0^*) \overset{\tilde{\theta}_0}{\rightsquigarrow} \mathcal{N}(0, \Psi' \sigma(\tilde{\theta}_0)^{-1} (\Psi')^\top), \quad (4.36)$$

Par conséquent,

$$\sqrt{m}(\hat{\boldsymbol{q}}^*(m) - \boldsymbol{q}_0^*) \overset{\tilde{\theta}_0}{\rightsquigarrow} \mathcal{N}(0, \Psi_1' \sigma(\tilde{\theta}_0^*)_{11}^{-1} (\Psi_1')^\top), \quad (4.37)$$

où la matrice Ψ' est donnée par les relations (4.48)-(4.52) ci-dessous et Ψ_1' est la sous-matrice de Ψ' formée par ses d^* premières lignes et colonnes (matrice Jacobienne de l'application $\Psi_1 : \mathcal{Q}^* \rightarrow \mathcal{H}^*$).

Démonstration. Soit pour tout $i \in E$, $\tilde{n}_{i\tau_i(1)}, \tilde{n}_{i\tau_i(2)}, \dots, \tilde{n}_{i\tau_i(c_i)}$, la suite de \tilde{n}_{ij} , ordonnée par rapport aux différents j tels que $\tilde{n}_{ij} > 0$, où c_i est défini par (4.25).

Dans le cas où quelques éléments de la suite sont égaux, l'ordre est considéré selon l'ordre des indices j . Notons que puisque $\tilde{n}_{i\tau_i(c_i)} = \tilde{n}_i$, donc $\tau_i(c_i) \in J_i$ (voir (4.10)), et par conséquent nous pouvons choisir $j_i = \tau_i(c_i)$. Pour tout $i \in E$, nous considérons la décomposition suivante

$$\boldsymbol{q}_{i\tau_i(j)}^* = \begin{cases} (q_{i\tau_i(j)}(1), q_{i\tau_i(j)}(2), \dots, q_{i\tau_i(j)}(\tilde{n}_{i\tau_i(j)})) & \text{si } j \in \mathbb{T}_{c_i-1}, \\ (q_{ij_i}(1), q_{ij_i}(2), \dots, q_{ij_i}(\tilde{n}_{ij_i} - 1)) & \text{si } j = c_i. \end{cases} \quad (4.38)$$

et également

$$\boldsymbol{q}_i^* = (\boldsymbol{q}_{i\tau_i(1)}, \boldsymbol{q}_{i\tau_i(2)}, \dots, \boldsymbol{q}_{ij_i}). \quad (4.39)$$

Enfin, les paramètres de l'espace paramétrique θ^* peuvent être ordonnés comme suit :

$$\boldsymbol{\theta}^* = (\boldsymbol{q}_1^*, \boldsymbol{q}_2^*, \dots, \boldsymbol{q}_s^*, \boldsymbol{\theta}_g). \quad (4.40)$$

Pour les représentations matricielles données dans la suite, nous aurons besoin d'une mise en ordre spécifique des éléments de $\tilde{\Theta}^*$. À ce but, nous désignons pour tout $i \in E$, $j \in \mathbb{T}_{c_i-1}$,

$$\boldsymbol{\eta}_{ii}^* = (p_{i,0;i}, p_{i,1;i}, \dots, p_{i,\tilde{n}_i-2;i}), \quad (4.41)$$

$$\boldsymbol{\eta}_{i\tau_i(j)}^* = (p_{i,0;\tau_i(j)}, p_{i,1;\tau_i(j)}, \dots, p_{i,\tilde{n}_{i\tau_i(j)}-1;\tau_i(j)}). \quad (4.42)$$

En notant par

$$\boldsymbol{\eta}_i^* = (\boldsymbol{\eta}_{i\tau_i(1)}^*, \boldsymbol{\eta}_{i\tau_i(2)}^*, \dots, \boldsymbol{\eta}_{i\tau_i(c_i-1)}^*, \boldsymbol{\eta}_{ii}^*),$$

une mise en ordre des paramètres de $\tilde{\Theta}^*$ est ainsi donnée par

$$\tilde{\boldsymbol{\theta}}^* = (\boldsymbol{\eta}_1^*, \boldsymbol{\eta}_2^*, \dots, \boldsymbol{\eta}_s^*, \boldsymbol{\theta}_g). \quad (4.43)$$

En utilisant les relations (4.18), (4.19), (4.40) et (4.43) une décomposition dans une forme diagonale par blocs pour Ψ' est donnée comme suit :

$$\Psi' = \text{diag}\{M^{(1)}, M^{(2)}, \dots, M^{(s)}, I_{d_B}\}, \quad (4.44)$$

où pour tout $i \in E$, $M^{(i)} = \begin{pmatrix} \frac{\partial \boldsymbol{q}_i^*}{\partial \boldsymbol{\eta}_i^*} \end{pmatrix}$.

Par les relations (4.18), (4.19), (4.39) et (4.4) nous pouvons décomposer $M^{(i)}$ en blocs de la façon suivante

$$M^{(i)} = \begin{pmatrix} M_{11}^{(i)} & \mathbf{0} & \dots & \mathbf{0} & M_{1c_i}^{(i)} \\ \mathbf{0} & M_{22}^{(i)} & \dots & \mathbf{0} & M_{2c_i}^{(i)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & M_{c_i-1, c_i-1}^{(i)} & M_{c_i-1, c_i}^{(i)} \\ M_{c_i1}^{(i)} & M_{c_i2}^{(i)} & \dots & M_{c_i, c_i-1}^{(i)} & M_{c_i c_i}^{(i)} \end{pmatrix}, \quad (4.45)$$

où pour tout $j \in \mathbb{T}_{c_i-1}$,

$$M_{jj}^{(i)} = \begin{pmatrix} \frac{\partial \boldsymbol{q}_{i\tau_i(j)}^*}{\partial \boldsymbol{\eta}_{i\tau_i(j)}^*} \end{pmatrix}, \quad M_{jc_i}^{(i)} = \begin{pmatrix} \frac{\partial \boldsymbol{q}_{i\tau_i(j)}^*}{\partial \boldsymbol{\eta}_{ii}^*} \end{pmatrix}, \quad M_{c_ij}^{(i)} = \begin{pmatrix} \frac{\partial \boldsymbol{q}_{ij}^*}{\partial \boldsymbol{\eta}_{i\tau_i(j)}^*} \end{pmatrix},$$

et

$$M_{c_i c_i}^{(i)} = \begin{pmatrix} \frac{\partial \boldsymbol{q}_{ii}^*}{\partial \boldsymbol{\eta}_{ii}^*} \end{pmatrix}.$$

Ces quatre différents types des matrices gardent toute l'information nécessaire afin d'avoir une forme explicite de la matrice Ψ' . Pour cela, nous étudions en détail chacun séparément.

Pour tout $i \in E$, $k \in \mathbb{T}_{\tilde{n}_i-1}$, $l \in \mathbb{T}_k$ nous définissons

$$a_i(k; l) = \frac{\bar{H}_i(k)}{p_{i, l-1; i}}, \quad (4.46)$$

$$b_{iu}^{(j)}(k; l) = p_{i, u; \tau_i(j)} a_i(k; l), \quad u \in \mathbb{T}_{\tilde{n}_{i\tau_i(j)}-1}. \quad (4.47)$$

Nous utilisons l'abréviation $c_{ij} = \tilde{n}_{i\tau_i(j)} - 2$, $c_{ij}^+ = c_{ij} + 1$. Alors nous avons,

$$M_{jj}^{(i)} = \text{diag}\{1, \bar{H}_i(1), \bar{H}_i(2), \dots, \bar{H}_i(c_{ij}^+)\}, \quad (4.48)$$

$$M_{c_ij}^{(i)} = \begin{pmatrix} \Delta_{c_ij}^{(i)} & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{0}^\top \end{pmatrix}, \quad (4.49)$$

où

$$\Delta_{c_ij}^{(i)} = -\text{diag}\{1, \bar{H}_i(1), \bar{H}_i(2), \dots, \bar{H}_i(c_{ij})\}, \quad (4.50)$$

$$M_{j c_i}^{(i)} = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ b_{i1}^{(j)}(1;1) & 0 & \dots & 0 & 0 & \dots & 0 \\ b_{i2}^{(j)}(2;1) & b_{i2}^{(j)}(2;2) & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ b_{i,c_{ij}^+}^{(j)}(c_{ij}^+;1) & b_{i,c_{ij}^+}^{(j)}(c_{ij}^+;2) & \dots & b_{i,c_{ij}^+}^{(j)}(c_{ij}^+;c_{ij}^+) & 0 & \dots & 0 \end{pmatrix}, \quad (4.51)$$

et

$$M_{c_i c_i}^{(i)} = \begin{pmatrix} -1 & 0 & \dots & 0 & 0 \\ b_{i1}^{(j)}(1;1) & -\tilde{H}_i(1) & \dots & 0 & 0 \\ b_{i2}^{(j)}(2;1) & b_{i2}^{(j)}(2;2) & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ b_{i,c_{ij}^{(j)}}^{(j)}(c_{ij}^{(j)};1) & b_{i,c_{ij}^{(j)}}^{(j)}(c_{ij}^{(j)};2) & \dots & b_{i,c_{ij}^{(j)}}^{(j)}(c_{ij}^{(j)};c_{ij}^{(j)}) & -\tilde{H}_i(c_{ij}^{(j)}). \end{pmatrix} \quad (4.52)$$

Puisque

$$\sqrt{m}(\hat{\theta}^*(m) - \theta_0^*) = \sqrt{m}(\Psi(\hat{\theta}^*(m)) - \Psi(\tilde{\theta}_0^*)), \quad (4.53)$$

où Ψ est différentiable en $\tilde{\theta}_0^*$, alors le présent théorème résulte du théorème 4.2 par une application de la méthode delta. \square

Remarque 4.5 Afin de trouver la matrice de covariance asymptotique de $\sqrt{m}(\hat{\theta}(m) - \theta_0)$, nous ajoutons les paramètres $q_{ij_i}(\tilde{n}_i)$, donnés par (4.20), et en utilisant la relation (4.11) nous concluons que, quand $m \rightarrow \infty$,

$$\sqrt{m}(\hat{\theta}(m) - \theta_0) \overset{\tilde{\theta}_0}{\rightsquigarrow} \mathcal{N}(0, C\Psi_1'\sigma(\tilde{\theta}_0)_{11}^{-1}(\Psi_1')^\top C^\top), \quad (4.54)$$

où

$$C = \text{diag}\{C_i, i \in E\}, \quad C_i = \begin{pmatrix} I_{r_i} \\ -\mathbf{1} \end{pmatrix}, \quad r_i = \sum_{j=1}^{c_i} \tilde{n}_{i\tau_i(j)} - 1. \quad (4.55)$$

Dans les deux propositions suivantes, nous démontrons la normalité asymptotique de l'EMV correspondant aux caractéristiques du modèle semi-markovien, définies par (4.23) et (4.24).

Proposition 4.2 Sous les hypothèses 4.1 et 4.2 et les conditions (B1)-(B4) et (C1)-(C6), l'EMV $\hat{p}^e(m)$ de p^{e_0} , des probabilités de transition de la chaîne markovienne J , est asymptotiquement normal, i.e., quand $m \rightarrow \infty$

$$\sqrt{m}(\hat{p}^e(m) - p^{e_0}) \overset{\tilde{\theta}_0}{\rightsquigarrow} \mathcal{N}(0, \Phi_1'\Psi_1'\sigma(\tilde{\theta}_0)^{-1}(\Phi_1'\Psi_1')^\top), \quad (4.56)$$

où $\Phi_1'\Psi_1'$ est donnée par les relations (4.63) et (4.64).

Démonstration. Nous définissons pour tout $i \in E$,

$$p_i^e = (p_{i\tau_i(1)}, p_{i\tau_i(2)}, \dots, p_{i\tau_i(c_i)}). \quad (4.57)$$

Alors, une mise en ordre du paramètre p^e de la CM J est la suivante :

$$p^e = (p_1^e, p_2^e, \dots, p_s^e). \quad (4.58)$$

Nous désignons par

$$V^{(i)} := \begin{pmatrix} \frac{\partial \mathbf{p}_i^e}{\partial \mathbf{q}_i^*} \end{pmatrix}. \quad (4.59)$$

Par la définition de Φ donnée dans (4.26) et par (4.59) nous avons que

$$\Phi' = \text{diag}\{V^{(i)}, i \in E\}, \quad (4.60)$$

où $V^{(i)}$ est donnée par

$$V^{(i)} = \begin{pmatrix} \mathbf{1}_{11}^{(i)} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{22}^{(i)} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{1}_{c_i-1, c_i-1}^{(i)} & \mathbf{0} \\ -\mathbf{1}_{c_i 1}^{(i)} & -\mathbf{1}_{c_i 2}^{(i)} & \dots & -\mathbf{1}_{c_i, c_i-1}^{(i)} & \mathbf{0} \end{pmatrix}, \quad (4.61)$$

et $\mathbf{1}_{jj}^{(i)}$, $\mathbf{1}_{c_{ij}}^{(i)}$ sont des vecteurs lignes avec chaque élément égal à un, de dimension $\tilde{n}_{i\tau_i(j)}$, pour tout j tel que $1 \leq j \leq c_i - 1$.

Puisque

$$\sqrt{m}(\hat{\mathbf{p}}^e(m) - \mathbf{p}^{e0}) = \sqrt{m}(\Phi(\hat{\mathbf{q}}^*(m)) - \Phi(\mathbf{q}_0^*)), \quad (4.62)$$

donc par le théorème 4.3 et la différentiabilité de Φ dans \mathcal{Q}^* , nous déduisons, par une application de la méthode delta, la validité de (4.56), où

$$\Phi'_1 \Psi'_1 = \text{diag}\{V^{(i)} M^{(i)}, i \in E\}, \quad (4.63)$$

et $V^{(i)}$, $M^{(i)}$ sont données par (4.61) et (4.45) respectivement.

La forme explicite de ce produit pour tout $i \in E$, est donnée comme suit :

$$V^{(i)} M^{(i)} = \begin{pmatrix} \mathbf{d}_1^{(i)} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{d}_2^{(i)} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{d}_{c_i-1}^{(i)} & \mathbf{0} \\ -\mathbf{d}_1^{(i)} & -\mathbf{d}_2^{(i)} & \dots & -\mathbf{d}_{c_i-1}^{(i)} & \mathbf{0} \end{pmatrix}, \quad (4.64)$$

où

$$\mathbf{d}_j^{(i)} = (1, \bar{H}_i(1), \bar{H}_i(2), \dots, \bar{H}_i(c_{ij}^+)). \quad (4.65)$$

□

Proposition 4.3 *Sous les hypothèses 4.1 et 4.2 et les conditions (B1)-(B4) et (C1)-(C6), l'EMV $\hat{\mathbf{f}}(m)$ de \mathbf{f}^0 , des probabilités des temps de séjour conditionnelles, est asymptotiquement normal, i.e., quand $m \rightarrow \infty$,*

$$\sqrt{m}(\hat{\mathbf{f}}(m) - \mathbf{f}^0) \overset{\hat{\theta}_0}{\rightsquigarrow} \mathcal{N}(0, T' \Psi'_1 \sigma(\theta_0^*)^{-1} (T' \Psi'_1)^\top). \quad (4.66)$$

Démonstration. Nous définissons pour tout $i \in E$, $j \in \mathbb{T}_{c_i}$,

$$f_{i\tau_i(j)} = (f_{i\tau_i(j)}(1), f_{i\tau_i(j)}(2), \dots, f_{i\tau_i(j)}(\tilde{n}_{i\tau_i(j)})), \quad (4.67)$$

et pour tout $i \in E$,

$$f_i = (f_{i\tau_i(1)}, f_{i\tau_i(2)}, \dots, f_{ij_i}). \quad (4.68)$$

Alors, une mise en ordre du paramètre f est comme suit :

$$f = (f_1, f_2, \dots, f_s). \quad (4.69)$$

Pour tout $i \in E$, nous désignons par

$$F^{(i)} := \left(\frac{\partial f_i}{\partial q_i^*} \right) \quad (4.70)$$

la matrice jacobienne correspondante.

Par la définition de T donnée dans (4.27), nous avons

$$T' = \text{diag}\{F^{(i)}, i \in E\}, \quad (4.71)$$

où pour tout $i \in E$, la matrice $F^{(i)}$ est donnée par

$$F^{(i)} = \begin{pmatrix} F_{11}^{(i)} & 0 & \dots & 0 & 0 \\ 0 & F_{22}^{(i)} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & F_{c_i-1, c_i-1}^{(i)} & 0 \\ F_{c_i 1}^{(i)} & F_{c_i 2}^{(i)} & \dots & F_{c_i, c_i-1}^{(i)} & F_{c_i c_i}^{(i)} \end{pmatrix}, \quad (4.72)$$

et les matrices

$$F_{j_1 j_2}^{(i)} := \left(\frac{\partial f_{i\tau_i(j_1)}}{\partial q_{i\tau_i(j_2)}^*} \right), \quad (4.73)$$

pour les différentes valeurs de j_1 et j_2 qui correspondent aux matrices non nulles dans (4.72), sont données par

$$F_{jj}^{(i)} = -\frac{1}{p_{i\tau_i(j)}^2} \begin{pmatrix} -\sum_{k \neq 1} q_{i\tau_i(j)}(k) & q_{i\tau_i(j)}(1) & \dots & q_{i\tau_i(j)}(1) \\ q_{i\tau_i(j)}(2) & -\sum_{k \neq 2} q_{i\tau_i(j)}(k) & \dots & q_{i\tau_i(j)}(2) \\ \vdots & \vdots & \ddots & \vdots \\ q_{i\tau_i(j)}(\tilde{n}_{i\tau_i(j)}) & q_{i\tau_i(j)}(\tilde{n}_{i\tau_i(j)}) & \dots & -\sum_{k \neq \tilde{n}_{i\tau_i(j)}} q_{i\tau_i(j)}(k) \end{pmatrix}, \quad (4.74)$$

$$F_{c_i j}^{(i)} = \frac{1}{p_{ij_i}^2} \begin{pmatrix} q_{ij_i}(1) & q_{ij_i}(1) & \dots & q_{ij_i}(1) \\ \vdots & \vdots & \ddots & \vdots \\ q_{ij_i}(\tilde{n}_i - 1) & q_{ij_i}(\tilde{n}_i - 1) & \dots & q_{ij_i}(\tilde{n}_i - 1) \\ -\sum_{k \neq \tilde{n}_i} q_{ij_i}(k) & -\sum_{k \neq \tilde{n}_i} q_{ij_i}(k) & \dots & -\sum_{k \neq \tilde{n}_i} q_{ij_i}(k) \end{pmatrix}, \quad (4.75)$$

$$F_{c_i c_i}^{(i)} = \frac{1}{p_{ij_i}} \begin{pmatrix} I_{s_i} \\ -\mathbf{1} \end{pmatrix}, \text{ où } s_i = c_{ij_i}^+. \quad (4.76)$$

Puisque

$$\sqrt{m}(\widehat{f}(m) - f^0) = \sqrt{m}(T(\widehat{q}^*(m)) - T(q_0^*)),$$

donc par le théorème 4.3 et la différentiabilité de T dans l'intérieur de \mathcal{Q}^* , nous déduisons, par une application de la méthode delta, la validité de (4.66), où

$$T' \Psi_1' = \text{diag}\{F^{(i)} M^{(i)}, i \in E\}, \quad (4.77)$$

où $F^{(i)}$ et $M^{(i)}$ sont données par (4.72) et (4.45) respectivement. La forme explicite de ce produit pour tout $i \in E$ est donnée comme suit :

$$F^{(i)} M^{(i)} = \begin{pmatrix} D_{11}^{(i)} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & D_{22}^{(i)} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & D_{c_i-1, c_i-1}^{(i)} & \mathbf{0} \\ D_{c_i 1}^{(i)} & D_{c_i 2}^{(i)} & \dots & D_{c_i, c_i-1}^{(i)} & D_{c_i, c_i}^{(i)} \end{pmatrix}, \quad (4.78)$$

où pour tout $j \in \mathbb{T}_{c_i-1}$,

$$D_{jj}^{(i)} = -\frac{1}{p_{i\tau_i(j)}^2} \begin{pmatrix} -\sum_{k \neq 1} q_{i\tau_i(j)}(k) & \bar{H}_i(1) q_{i\tau_i(j)}(1) & \dots & \bar{H}_i(c_{ij}^+) q_{i\tau_i(j)}(1) \\ q_{i\tau_i(j)}(2) & -\bar{H}_i(1) \sum_{k \neq 2} q_{i\tau_i(j)}(k) & \dots & \bar{H}_i(c_{ij}^+) q_{i\tau_i(j)}(2) \\ \vdots & \vdots & \ddots & \vdots \\ q_{i\tau_i(j)}(\tilde{n}_{i\tau_i(j)}) & \bar{H}_i(1) q_{i\tau_i(j)}(\tilde{n}_{i\tau_i(j)}) & \dots & -\bar{H}_i(c_{ij}^+) \sum_{k \neq \tilde{n}_{i\tau_i(j)}} q_{i\tau_i(j)}(k) \end{pmatrix}, \quad (4.79)$$

$$D_{c_{ij}}^{(i)} = \frac{1}{p_{ij_i}^2} \begin{pmatrix} -\sum_{k \neq 1} q_{ij_i}(k) & \bar{H}_i(1) q_{ij_i}(1) & \dots & \bar{H}_i(c_{ij}^+) q_{ij_i}(1) \\ q_{ij_i}(2) & -\bar{H}_i(1) \sum_{k \neq 2} q_{ij_i}(k) & \dots & \bar{H}_i(c_{ij}^+) q_{ij_i}(2) \\ \vdots & \vdots & \ddots & \vdots \\ q_{ij_i}(\tilde{n}_i) & \bar{H}_i(1) q_{ij_i}(\tilde{n}_i) & \dots & -\bar{H}_i(c_{ij}^+) \sum_{k \neq \tilde{n}_i} q_{ij_i}(k) \end{pmatrix}, \quad (4.80)$$

et

$$D_{c_i c_i}^{(i)} = \sum_{j=1}^{c_i} F_{c_{ij}}^{(i)} M_{j c_i}^{(i)}, \quad (4.81)$$

où $F_{c_{ij}}^{(i)}$ et $M_{j c_i}^{(i)}$ sont données par (4.51)-(4.52) et (4.75)-(4.76). \square

CONCLUSION DU CHAPITRE

Dans ce chapitre nous avons introduit le modèle semi-markovien caché général (MSMC général) avec dépendance des temps de récurrence en arrière. Nous avons démontré la convergence et la normalité asymptotique de l'EMV qui correspond à la vraie valeur du paramètre de ce modèle dans le cas où le support des temps de séjour conditionnels pour les états cachés est fini. Pour cela, nous avons reparamétrisé ce modèle par une sous-classe de modèles markoviens cachés (MMC) et également nous avons donné des formes exactes qui lient la matrice de covariance asymptotique du MMC avec celles des différentes caractéristiques du MSMC.

ALGORITHMES EM ET SEM POUR DES MSMC

SOMMAIRE

5.1	PRÉLIMINAIRES ET HYPOTHÈSES	101
5.2	UNE VERSION AMÉLIORÉE DE L'ALGORITHME EM POUR MSMC . . .	102
5.3	DÉMONSTRATIONS DES FORMULES POUR EM	108
5.4	ALGORITHME SEM POUR MSMC NON PARAMÉTRIQUE	111
5.5	APPLICATIONS	114
	CONCLUSION	124

Dans le chapitre précédent nous avons étudié les propriétés asymptotiques de l'EMV du modèle semi-markovien caché général (MSMC général) avec dépendance des temps de récurrence en arrière dans le cas où le support des lois des temps de séjour conditionnelles pour les états cachés est fini.

Dans ce chapitre notre objectif est d'estimer le noyau semi-markovien qui gouverne l'évolution de la chaîne semi-markovienne cachée (CSMC) ainsi que les probabilités d'émission des états cachés vers les états observés, quand l'espace d'état \mathcal{Y} pour le processus observable Y est fini et en considérant qu'il n'y a pas de dépendance des temps de récurrence en arrière. Tout particulièrement, nous concentrons notre intérêt sur le cas non paramétrique.

Dans le but d'obtenir l'EMV des paramètres des MSMC, un algorithme EM (Estimation-Maximisation) peut être appliqué. Ce type d'algorithme est spécialement destiné à traiter des problèmes avec des données manquantes. Ferguson (1980) a été le premier à proposer un algorithme EM pour les MSMC. Depuis, plusieurs travaux de recherche ont été effectués dans cette direction. La majorité de ces travaux modélise les temps de séjour par une famille paramétrique des lois. Levinson (1986) a remplacé les lois de probabilité pour les temps de séjour par des fonctions de densité de probabilité en introduisant les MMC à variable de durée continue et plus précisément pour des applications à la reconnaissance de la parole il a utilisé une loi Gamma. Par ailleurs, Durbin et al. (1998)

utilisent une variable de durée qui suit une loi négative binomiale pour l'analyse des suites biologiques d'ADN. Mitchell et Jamieson (1993) supposent que la variable de durée appartient à la famille exponentielle discrète. Une application intéressante présentée par Sansom et Thomson (2001; 2007) et Sansom (1998), a consisté à modéliser avec des MSMC les points de rupture pour des données de précipitation et de climatologie. Guédon et Coccozza-Thivent (1990) ont présenté des applications à la reconnaissance de la parole (schéma de Derin). Guédon (1999; 2003; 2005; 2007) a présenté des algorithmes EM pour CSMC avec des applications à la reconnaissance de la parole et à l'analyse de la structure des plantes. Également, des applications à la finance ont été effectuées par Bulla et Bulla (2006).

Pour des discussions sur le cas non paramétrique voir Sansom et Thomson (2001), Bulla et Bulla (2006), Guédon (2007). Pour des algorithmes spécialement destinés aux cas non paramétriques voir Barbu et Limnios (2006; 2008), Malefaki et al. (soumis, 2008).

Dans le but d'estimer le noyau semi-markovien non paramétrique et les probabilités d'émission, Barbu et Limnios (2008) ont proposé un algorithme EM. Dans ce chapitre, nous proposons une version améliorée et plus rapide, en terme de temps d'exécution, de leur algorithme initial. En outre, nous présentons une version stochastique de cet algorithme (SEM). Cette version nous donne la possibilité d'utiliser des trajectoires encore plus longues, afin d'avoir plus de précision dans une durée de temps raisonnable. Cela peut être très utile dans plusieurs applications, comme par exemple, l'analyse des suites d'ADN, où des trajectoires assez grandes sont disponibles. Des exemples numériques qui illustrent l'efficacité des algorithmes EM et SEM proposés, seront également fournis dans ce chapitre.

Notons que dans ce chapitre nous ne faisons pas l'hypothèse que la chaîne semi-markovienne Z est stationnaire. Également, le support des lois des temps de séjour conditionnelles peut être fini ou infini.

Ce chapitre sera organisé de la façon suivante. Dans la section 5.1, nous introduisons le modèle caché dont nous ferons l'étude algorithmique ainsi que les conditions dont nous aurons besoin. Dans la section 5.2, nous présentons une version améliorée de l'algorithme EM pour des MSMC introduit dans Barbu et Limnios (2008). Ensuite, nous démontrons dans la section 5.3, les formules utilisées pour la construction de cet algorithme. Dans la section 5.4, nous présentons un algorithme EM stochastique (SEM) pour calculer l'EMV du MSMC non paramétrique. Enfin, dans la dernière section, nous étudions la performance des algorithmes EM et SEM en présentant quelques exemples numériques.

Les résultats de ce chapitre font partie de l'article Malefaki et al. (soumis, 2008).

5.1 PRÉLIMINAIRES ET HYPOTHÈSES

Dans cette section nous précisons les caractéristiques du modèle caché dont nous présentons une étude algorithmique dans la suite.

Nous considérons (Z, Y) un modèle caché de type SM-M0 non paramétrique, où

- Z a pour espace d'état fini $E = \{1, 2, \dots, s\}$ et vérifie $S_0 = 0$ (voir (3.1)),
- Y a pour espace d'état $\mathcal{Y} = \{1, \dots, r\}$.

Nous énonçons, dans ce qui suit, les conditions dont nous aurons besoin pour la sous-classe de CSM à considérer :

- Hypothèse 5.1**
- i) La CSM Z sous \mathbb{P}_0 est ergodique.
 - ii) Le temps de séjour conditionnel dans l'état i , conditionné au prochaine état visité j , sous $\mathbb{P}_{\theta^{(0)}}$, où $\theta^{(0)}$ est la valeur initiale des algorithmes EM et SEM que nous présentons dans ce chapitre, doit avoir un support $\mathbb{D}_{ij}^+ \subseteq \mathbb{T}_M$, pour tout $(i, j) \in E_+^2$.

- Remarque 5.1**
- i) L'hypothèse 5.1-(ii) implique qu'il y a des constantes $\tilde{n}_{ij} := |\mathbb{D}_{ij}^+|$ qui déterminent le support des lois conditionnelles décrites ci-dessus, tout au long de la procédure de l'estimation des paramètres pour les différentes itérations des algorithmes EM et SEM décrits dans la suite. Si le vrai support pour ces lois n'est pas borné, ce qui est le cas dans les exemples que nous mentionnons dans la section 5.5, nous pouvons choisir des constantes \tilde{n}_{ij} appropriées de façon à ce que l'erreur d'approximation par un support borné soit négligeable.

ii) Notons que pour une observation Y_0^M dans un cadre non paramétrique, sous l'hypothèse que $S_0 = 0$, le support pour les lois ci-dessus est toujours supposé d'être concentré dans l'ensemble $\{1, 2, \dots, M\}$. En effet, nous ne pouvons avoir aucune information pour des transitions dont la durée dépasse l'intervalle d'observation. Cela revient à dire que nous pouvons toujours choisir $\tilde{n}_{ij} = M$ pour tout $(i, j) \in E_+^2$. Mais, en choisissant les \tilde{n}_{ij} inférieurs à M , nous pouvons accélérer l'estimation pour un nombre suffisant des paramètres, à condition que l'erreur soit suffisamment petite.

Dans le modèle caché (Z, Y) de type SM-M0 non paramétrique, étant donné le processus Z , la fonction de probabilité conditionnelle de Y_n dépend uniquement de Z_n , et pour tout $n \in \mathbb{N}$, $a \in \mathcal{Y}$, $i \in E$, nous avons

$$\mathbb{P}_\theta(Y_n = a \mid Z_n = i, Z_0^{n-1}, Y_0^{n-1}) = \mathbb{P}_\theta(Y_n = a \mid Z_n = i) := R_{i,a}. \quad (5.1)$$

- Remarque 5.2** Dorénavant, afin de simplifier la présentation des algorithmes EM et SEM pour le MSMC nous supposons que toutes les transitions pour les différents états de la CSM Z sont possibles, i.e., $E_+^2 = E_*^2$ (en éliminant les auto-transitions). Également, nous prenons comme support pour la loi d'émission $R_i(\cdot)$, pour tout $i \in E$, l'ensemble \mathcal{Y} , i.e., les probabilités d'émission $R_{i,a}$ sont strictement positives. Dans le cas contraire, nous pouvons toujours considérer le support exact.

Dans ce chapitre nous considérons un cadre non paramétrique et donc nous choisissons comme espace paramétrique approprié pour ce modèle,

tout en tenant compte de la remarque 5.2 :

$$\Theta := (\mathcal{Q}, \mathcal{R}),$$

avec

$$\mathcal{Q} := \{q = (q_{ij}(k)) \in \mathbb{R}^{d^+} : C_{q,sto}\}, \quad (5.2)$$

$$\mathcal{R} := \{R = (R_{i,a}) \in \mathbb{R}^{sr} : C_{R,sto}\}, \quad (5.3)$$

où $C_{R,sto}$ représente l'ensemble (pour les différents $i \in E$) des relations de dépendance linéaires

$$\sum_{a \in \mathcal{Y}} R_{i,a} = 1, \quad (5.4)$$

et $C_{q,sto}$ celui des relations données par (3.39), ici pour les différents $i \in E$.

Quand il s'agit d'un vecteur $Z_{k_1}^{k_2}$ dont chaque composante vaut $i \in E$, nous le noterons, dans la suite, par $Z_{k_1}^{k_2} = i$, où la dimension de i est impliquée par celle du vecteur $Z_{k_1}^{k_2}$.

5.2 UNE VERSION AMÉLIORÉE DE L'ALGORITHME EM POUR MSMC

Dans cette section, en nous basant sur l'algorithme EM de Barbu et Limnios (2008), nous présentons une version améliorée de cet algorithme pour le MSMC non paramétrique, qui est défini dans la section 5.1 et vérifie l'hypothèse 5.1.

Cette amélioration consiste en les points suivants :

i) L'algorithme initial de Barbu et Limnios (2008) ne prend pas en compte le support des lois qui apparaissent dans la valeur initiale $\theta^{(0)}$. Toutes les lois de transition d'un état i vers un état j de la CSM Z qui ont été supposées d'avoir un support \mathbb{D}_{ij}^+ vont garder ce support tout au long de la procédure d'estimation. Cependant, dans notre version, nous avons pris en compte les constantes $\tilde{n}_{ij} := |\mathbb{D}_{ij}^+|$, ce qui a permis de réduire considérablement le temps d'exécution de l'algorithme. Et, dans le cas où le support théorique des lois de transition d'un état i vers un état j est infini, nous pouvons obtenir une approximation assez satisfaisante, en choisissant \tilde{n}_{ij} tel que $F_{ij}(\tilde{n}_{ij}) > 1 - \epsilon$, pour un ϵ donné assez petit, où $F_{ij}(\cdot)$ est la fonction de répartition de $f_{ij}(\cdot)$.

ii) La construction de l'algorithme initial a été basée sur la paramétrisation (p^e, f) du noyau semi-markovien q . Par contre, notre algorithme consiste en une paramétrisation plus directe puisqu'il utilise uniquement le noyau q . Et, pour obtenir p^e ou f , il suffit ainsi d'utiliser l'équation (5.43) ou (4.24) respectivement.

iii) Les formules utilisées dans l'algorithme initial ont été simplifiées, ainsi notre algorithme nécessite moins de calcul que l'algorithme initial.

L'algorithme EM a été présenté en détails par Dempster et al. (1977). Il se base sur une méthode itérative et déterministe conçue pour trouver l'EMV dans des problèmes à données manquantes. Depuis, il a été utilisé par plusieurs auteurs et le travail de Dempster et al. (1977) a été cité plus de 10000 fois.

Chaque itération de l'algorithme EM se compose de deux étapes. La première étape est l'étape d'estimation (étape-E) où l'espérance conditionnelle de la log-vraisemblance des données complètes conditionnées par les données observées et par le paramètre actuel, appelée fonction-Q, est calculée. La deuxième étape est l'étape de maximisation (étape-M), où les paramètres sont mis à jour en maximisant la fonction-Q de l'étape-E.

Plus précisément, dans le cas de MSMC, basé sur une suite échantillonnée $\{\mathbf{Y}_0^M = \mathbf{y}_0^M\}$, notre principal objectif est de trouver l'EMV du paramètre $\boldsymbol{\theta} = (\mathbf{q}, \mathbf{R})$ du modèle. La fonction de vraisemblance $\mathcal{L}_M(\boldsymbol{\theta}) := p_M(\mathbf{Z}_0^M, \mathbf{Y}_0^M | \boldsymbol{\theta})$ des données complètes est donnée par

$$\mathcal{L}_M(\boldsymbol{\theta}) = \alpha(Z_0) \left(\prod_{k=1}^{N(M)} q_{J_{k-1}J_k}(S_k - S_{k-1}) \right) \left(\prod_{n=0}^M R_{Z_n; Y_n} \right) \bar{H}_{J_{N(M)}}(M - S_{N(M)}), \quad (5.5)$$

où $\alpha(\cdot)$ est la loi initiale et $\bar{H}_i(\cdot)$ est la fonction de survie dans l'état i pour la CSM \mathbf{Z} .

Nous définissons le processus de comptage du nombre de transitions de type $(i, j) \in E_*^2$ qui ont eu lieu dans $k \in \mathbb{T}_M$ unités de temps par :

$$N_{ij}(k, M) := \sum_{n=0}^{M-k} \mathbb{1}_{\{Z_{n+k}=j, Z_n^{n+k-1}=i, Z_{n-1} \neq i\}}, \quad (i, j) \in E_*^2, k \in \mathbb{T}_{\tilde{n}_{ij}}, \quad (5.6)$$

où par convention $\{Z_0 = i, Z_{-1} \neq i\} = \{Z_0 = i\}$, ce qui résulte de l'hypothèse $S_0 = 0$. Cette définition coïncide avec (3.45), mais ici elle est exprimée en terme de \mathbf{Z} et pas de (\mathbf{J}, \mathbf{X}) .

Notons également que $S_{N(M)}$, qui correspond à l'instant du dernier saut de \mathbf{Z} avant M , peut être exprimé comme la somme de tous les temps de séjour dans les différents états du système avant M , i.e.,

$$S_{N(M)} = \sum_{\substack{i,j,k \\ i \neq j}} k N_{ij}(k, M). \quad (5.7)$$

En outre, nous définissons le processus de comptage du nombre de passages du processus (\mathbf{Z}, \mathbf{Y}) à l'état $(i, a) \in E \times A$ par :

$$N_{i;a}(M) := \sum_{n=0}^M \mathbb{1}_{\{Z_n=i, Y_n=a\}}, \quad i \in E, a \in A. \quad (5.8)$$

D'après les définitions introduites ci-dessus nous pouvons reformuler la

fonction de vraisemblance comme suit :

$$\mathcal{L}_M(\boldsymbol{\theta}) = \alpha(Z_0) \left(\prod_{\substack{i,j,k \\ i \neq j}} q_{ij}(k)^{N_{ij}(k,M)} \right) \left(\prod_{i,\alpha} R_{i;\alpha}^{N_{i;\alpha}(M)} \right) \bar{H}_{Z_M}(M - S_{N(M)}), \quad (5.9)$$

où $S_{N(M)}$ est donné par (5.7).

Par cette représentation de la fonction de vraisemblance des données complètes, nous pouvons déduire que les processus de comptage donnés ci-dessus constituent des statistiques suffisantes pour les paramètres du modèle qui correspondent aux données complètes.

La fonction de log-vraisemblance $\ell_M(\boldsymbol{\theta}) := \log p_M(\mathbf{Z}_0^M, \mathbf{Y}_0^M | \boldsymbol{\theta})$ des données complètes est alors donnée par

$$\begin{aligned} \ell_M(\boldsymbol{\theta}) = & \log(\alpha(Z_0)) + \sum_{\substack{i,j,k \\ i \neq j}} N_{ij}(k, M) \log q_{ij}(k) + \sum_i \sum_b N_{i;a}(M) \log R_{i;a} \\ & + \log(\bar{H}_{Z_M}(M - S_{N(M)})). \end{aligned} \quad (5.10)$$

Au lieu de maximiser la fonction de log-vraisemblance des données complètes, qui ne sont pas connues, l'algorithme EM maximise l'espérance conditionnelle de la log-vraisemblance des données complètes, conditionnée à l'événement $\{\mathbf{Y}_0^M = \mathbf{y}_0^M\}$ et la valeur actuelle $\boldsymbol{\theta}^{(m)}$ du paramètre $\boldsymbol{\theta}$. Pour simplicité et quand il n'y a aucune ambiguïté, nous notons par \mathbf{y} cet événement. Le premier terme de (5.10) correspond à la loi initiale et puisque dans notre contexte il ne dépend pas de $\boldsymbol{\theta}$, il peut donc être éliminé de la fonction de maximisation. Par conséquent, la fonction qui doit être maximisée est :

$$\begin{aligned} Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(m)}) & := \mathbb{E}_{\boldsymbol{\theta}^{(m)}} \left[\log(p_M(\mathbf{Z}_0^M, \mathbf{Y}_0^M | \boldsymbol{\theta})) | \mathbf{y} \right] \\ & = \sum_{\substack{i,j,k \\ i \neq j}} \log q_{ij}(k) \sum_{n=0}^{M-k} \mathbb{E}_{\boldsymbol{\theta}^{(m)}} \left(\mathbb{1}_{\{Z_{n+k}=j, Z_n^{n+k-1}=i, Z_{n-1} \neq i\}} | \mathbf{y} \right) \\ & \quad + \sum_i \sum_a \log R_{i;a} \sum_{n=0}^M \mathbb{E}_{\boldsymbol{\theta}^{(m)}} \left(\mathbb{1}_{\{Z_n=i, Y_n=a\}} | \mathbf{y} \right) \\ & \quad + \mathbb{E}_{\boldsymbol{\theta}^{(m)}} \left(\log \bar{H}_{Z_M}(M - S_{N(M)}) | \mathbf{y} \right). \end{aligned} \quad (5.11)$$

Remarque 5.3 *Le dernier terme dépend du dernier état visité et du temps de séjour dans cet état jusqu'à l'instant M . Son influence diminue quand M devient grand, à condition que \mathbf{Z} soit ergodique (hypothèse 5.1-(i)). Donc, ce terme peut être négligé pour le calcul de l'EMV (EMV approché).*

Par la relation (5.11) et la remarque 5.3, le calcul de l'EMV se réduit à la

maximisation de la fonction

$$\begin{aligned} \tilde{Q}(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(m)}) &:= \sum_{\substack{i,j,k \\ i \neq j}} \log q_{ij}(k) \sum_{n=0}^{M-k} \mathbb{P}_{\boldsymbol{\theta}^{(m)}}(Z_{n+k} = j, \mathbf{Z}_n^{n+k-1} = \mathbf{i}, Z_{n-1} \neq i \mid \mathbf{y}) \\ &+ \sum_i \sum_a \log R_{i,a} \sum_{n=0}^M \mathbb{1}_{\{Y_n=a\}} \mathbb{P}_{\boldsymbol{\theta}^{(m)}}(Z_n = i \mid \mathbf{y}). \end{aligned} \quad (5.12)$$

Plus précisément, nous commençons par une valeur initiale arbitraire $\boldsymbol{\theta}^{(0)}$, et dans chaque étape-EM, en utilisant un algorithme Avant-Arrière qui va être présenté dans le reste de cette section, nous calculons les probabilités

$$\mathbb{P}_{\boldsymbol{\theta}^{(m)}}(Z_{n+k} = j, \mathbf{Z}_n^{n+k-1} = \mathbf{i}, Z_{n-1} \neq i \mid \mathbf{y}),$$

et

$$\mathbb{P}_{\boldsymbol{\theta}^{(m)}}(Z_n = i \mid \mathbf{y}),$$

et puis en maximisant la fonction $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(m)})$ par rapport à $\boldsymbol{\theta}$, nous obtenons une nouvelle valeur $\boldsymbol{\theta}^{(m+1)}$ du paramètre du modèle. Cette procédure continue jusqu'à satisfaire un certain critère d'arrêt.

Il y a une littérature abondante au sujet du critère d'arrêt. La majorité des critères sont basés sur la proximité de deux valeurs successives de la fonction-Q ou des estimations du paramètre (voir Fletcher (1980), McLachlan et Peel (2000)). Néanmoins, l'étude des critères est au-delà de notre intention pour ce travail, et par conséquent nous n'entrons pas dans les détails de ce sujet.

Maintenant, nous définissons les quantités sur lesquelles nous basons les relations récursives pour les algorithmes EM et SEM.

Soit pour tout $n \in \mathbb{T}_M$, $j \in E$:

$$P_0 := \mathbb{P}_{\boldsymbol{\theta}}(Y_0), \quad (5.13)$$

$$P_n := \mathbb{P}_{\boldsymbol{\theta}}(Y_n \mid \mathbf{y}_0^{n-1}), \quad n \neq 0, \quad (5.14)$$

$$F_n(j) := \mathbb{P}_{\boldsymbol{\theta}}(Z_n = j, Z_{n-1} \neq j \mid \mathbf{y}_0^n), \quad (5.15)$$

$$L_n(j) := \mathbb{P}_{\boldsymbol{\theta}}(Z_n = j \mid \mathbf{y}_0^M), \quad (5.16)$$

$$L_{1,n}(j) := \mathbb{P}_{\boldsymbol{\theta}}(Z_n = j, Z_{n-1} \neq j \mid \mathbf{y}_0^M), \quad (5.17)$$

$$L_{2,n}(j) := \mathbb{P}_{\boldsymbol{\theta}}(Z_n = j, Z_{n-1} = j \mid \mathbf{y}_0^M), \quad n \neq 0, \quad (5.18)$$

où pour tout $i \in E$ et $a \in A$, la loi initiale du couple (\mathbf{Z}, \mathbf{Y}) est déterminée par

$$\mu(i, a) := \mathbb{P}_{\boldsymbol{\theta}}(Z_0 = i, Y_0 = a). \quad (5.19)$$

De plus, en utilisant également que $S_0 = 0$, nous avons pour tout $j \in E$

$$F_0(j) := \mathbb{P}_{\boldsymbol{\theta}}(Z_0 = j \mid y_0) = \mu(j, y_0) / \sum_{i \in E} \mu(i, y_0), \quad (5.20)$$

$$L_0(j) := \mathbb{P}_{\boldsymbol{\theta}}(Z_0 = j \mid \mathbf{y}_0^M). \quad (5.21)$$

Dans le reste de cette section nous présentons une description détaillée de l'algorithme EM pour le MSMC. Également, pour simplifier les notations dans la suite, nous utilisons pour les produits $\prod_{p=n-k+1}^n R_{i;y_p}$ le terme

$$T_{n,k}(i) := \prod_{p=n-k+1}^n R_{i;y_p}, \quad (5.22)$$

et aussi la notation standard $x \wedge y = \min\{x, y\}$. Pour des sommes et des produits vides nous faisons la convention qu'ils sont nuls et unités respectivement.

Étape Estimation

Étape Avant

– Pour $n = 0$, nous avons

$$P_0 = \sum_i \mu(i, y_0), \quad F_0(j) = \frac{\mu(j, y_0)}{P_0}.$$

– Pour $n = 1, \dots, M$, nous avons

$$\begin{aligned} P_n &= \sum_i \sum_{j \neq i} \sum_{t=1}^{(\tilde{n}_{ij}-1) \wedge n} \sum_{u=1}^{(\tilde{n}_{ij}-t) \wedge (M-n)} \frac{q_{ij}(t+u) F_{n-t}(i) T_{n,t}(i)}{\prod_{p=n-t+1}^{n-1} P_p} \\ &+ \sum_i \sum_{j \neq i} \sum_{t=1}^{\tilde{n}_{ij} \wedge n} \frac{q_{ij}(t) F_{n-t}(i) R_{j;y_n} T_{n-1,t-1}(i)}{\prod_{p=n-t+1}^{n-1} P_p} \\ &+ \sum_i \sum_{t=1}^{(\tilde{n}_{ij}-1) \wedge n} \frac{\tilde{H}_i(M-n+t) F_{n-t}(i) T_{n,t}(i)}{\prod_{p=n-t+1}^{n-1} P_p}, \end{aligned} \quad (5.23)$$

$$F_n(j) = \sum_{i \neq j} \sum_{t=1}^{\tilde{n}_{ij} \wedge n} \frac{q_{ij}(t) F_{n-t}(i) R_{j;y_n} T_{n-1,t-1}(i)}{\prod_{p=n-t+1}^n P_p}. \quad (5.24)$$

Étape Arrière

– Pour $n = M$, nous avons

$$L_{1;M}(i) = F_M(i), \quad (5.25)$$

$$L_{2;M}(i) = \sum_{u=2}^{\tilde{n}_{ij} \wedge (M+1)} \frac{\tilde{H}_i(u-1) F_{M-u+1}(i) T_{M,u-1}(i)}{\prod_{l=M-u+2}^M P_l}. \quad (5.26)$$

– Pour $n = M-1, M-2, \dots, 0$, nous avons

$$\begin{aligned} L_{1;n}(i) &= \sum_{j \neq i} \sum_{t=1}^{\tilde{n}_{ij} \wedge (M-n)} \frac{q_{ij}(t) F_n(i) L_{1,n+t}(j) R_{j;y_{n+t}} T_{n+t-1,t-1}(i)}{F_{n+t}(j) \prod_{p=n+1}^{n+t} P_p} \\ &+ \frac{\tilde{H}_i(\tilde{n}_{ij} \wedge (M-n)) F_n(i) T_{M,M-n}(i)}{\prod_{p=n+1}^M P_p}. \end{aligned} \quad (5.27)$$

$$L_{2;n}(i) = \sum_{j \neq i} \sum_{t=1}^{(M-n) \wedge (\tilde{n}_{ij}-1)} \sum_{u=2}^{(n+1) \wedge (\tilde{n}_{ij}-t)} \left[\frac{q_{ij}(t+u-1) F_{n-u+1}(i) L_{1;n+t}(j)}{F_{n+t}(j)} \right. \\ \left. \times \frac{R_{j;y_{n+t}} T_{n+t-1;t+u-2}(i)}{\prod_{p=n-u+2}^{n+t} P_p} \right] \\ + \frac{H_i(\tilde{n}_{ij} \wedge (M-n+u-1)) F_{n-u+1}(i) T_{M;M-n+u-1}}{\prod_{l=n-u+2}^M P_l}, \quad (5.28)$$

$$L_{1;0}(i) = L_0(i). \quad (5.29)$$

Étape Maximisation

La procédure itérative de l'estimation de la vraie valeur du paramètre est donnée comme suit :

$$q_{ij}^{(m+1)}(k) = q_{ij}^{(m)}(k) \sum_{n=0}^{M-k} \left(\frac{F_n^{(m)}(i) L_{1;n+k}^{(m)}(j) R_{j;y_{n+k}}^{(m)} T_{n+k-1,k-1}(i)}{F_{n+k}^{(m)}(j) \prod_{p=n+1}^{n+k} P_p^{(m)}} \right) \\ / \left(\sum_{n=0}^{M-1} L_{1;n}^{(m)}(i) - L_{2;M}^{(m)}(i) \right), \quad (5.30)$$

$$R_{i;a}^{(m+1)} = \frac{\sum_{n=0}^M \delta_{y_{na}} L_n^{(m)}(i)}{\sum_{n=0}^M L_n^{(m)}(i)}, \quad (5.31)$$

où pour $n = 1, 2, \dots, M$

$$L_n(i) = L_{1;n}(i) + L_{2;n}(i), \quad (5.32)$$

et

$$L_0(i) = L_{1;0}(i). \quad (5.33)$$

Remarque 5.4 (cas SM-M1) Dans certaines applications il est naturel de supposer que la suite observée \mathbf{Y} , conditionnée par \mathbf{Z} , forme une suite de v.a. conditionnellement markovienne. Il est clair que la fonction-Q approchée correspondante, de même que (5.12), vaut

$$\tilde{Q}(\boldsymbol{\theta} | \boldsymbol{\theta}^{(m)}) = \sum_{\substack{i,j,k \\ i \neq j}} \log q_{ij}(k) \sum_{n=0}^{M-k} \mathbb{P}_{\boldsymbol{\theta}^{(m)}}(Z_{n+k} = j, \mathbf{Z}_n^{n+k-1} = \mathbf{i}, Z_{n-1} \neq i | \mathbf{y}) \\ + \sum_i \sum_{a,b} \log R_{i;a,b} \sum_{n=1}^M \mathbb{1}_{\{Y_{n-1}=a, Y_n=b\}} \mathbb{P}_{\boldsymbol{\theta}^{(m)}}(Z_n = i | \mathbf{y}). \quad (5.34)$$

La procédure itérative de l'estimation de la vraie valeur du paramètre, en maximisant la fonction \tilde{Q} , donnée par (5.34), est identique pour le noyau semi-markovien pour les deux cas, SM-M0 et SM-M1, (voir la relation (5.55) ci-dessous), alors que pour les probabilités d'émission elle donne :

$$R_{i;a,b}^{(m+1)} = \frac{\sum_{n=1}^M \mathbb{1}_{\{Y_{n-1}=a, Y_n=b\}} \mathbb{P}_{\boldsymbol{\theta}^{(m)}}(Z_n = i | \mathbf{y})}{\sum_{n=1}^M \mathbb{P}_{\boldsymbol{\theta}^{(m)}}(Z_n = i | \mathbf{y})}. \quad (5.35)$$

Cela nous permet de construire l'algorithme EM quand il s'agit de dépendance conditionnelle markovienne comme une extension simple du cas où il y a indépendance conditionnelle entre les v.a. de la suite \mathbf{Y} conditionnée par \mathbf{Z} (comparer avec (5.56)). Par conséquent, cette construction ne sera pas détaillée ici.

5.3 DÉMONSTRATIONS DES FORMULES POUR EM

Dans cette section nous donnons les démonstrations des expressions (5.23) et (5.24) qui concernent les termes P_n et $F_n(j)$ définis par (5.14) et (5.15) respectivement. Nous définissons

$$F_n(j, i; u, t) := \mathbb{P}_\theta(Z_{n+u} = j, \mathbf{Z}_{n-t}^{n+u-1} = \mathbf{i}, Z_{n-t-1} \neq i \mid \mathbf{y}_0^{n-t}). \quad (5.36)$$

Notons que

$$\begin{aligned} F_n(j, i; u, t) &= \mathbb{P}_\theta(Z_{n-t} = i, Z_{n-t-1} \neq i \mid \mathbf{y}_0^{n-t}) \\ &\quad \times \mathbb{P}_\theta(Z_{n+u} = j, \mathbf{Z}_{n-t+1}^{n+u-1} = \mathbf{i} \mid Z_{n-t} = i, Z_{n-t-1} \neq i, \mathbf{y}_0^{n-t}) \\ &= F_{n-t}(i)q_{ij}(u+t), \end{aligned} \quad (5.37)$$

puisque la probabilité $\mathbb{P}_\theta(Z_{n+u} = j, \mathbf{Z}_{n-t+1}^{n+u-1} = \mathbf{i} \mid Z_{n-t} = i, Z_{n-t-1} \neq i, \mathbf{y}_0^{n-t})$ ne dépend pas du vecteur \mathbf{y}_0^{n-t} .

Maintenant, nous analysons le terme P_n . Pour $n = 1, \dots, M$, nous avons

$$\begin{aligned} P_n &= \mathbb{P}_\theta(Y_n \mid \mathbf{y}_0^{n-1}) \\ &= \sum_{\substack{i,j \\ i \neq j}} \sum_{\substack{1 \leq t \leq \tilde{n}_{ij} \wedge n \\ 0 \leq u \leq (\tilde{n}_{ij}-t) \wedge (M-n)}} \mathbb{P}_\theta(Z_{n+u} = j, \mathbf{Z}_{n-t}^{n+u-1} = \mathbf{i}, Z_{n-t-1} \neq i, Y_n \mid \mathbf{y}_0^{n-1}) \\ &\quad + \sum_i \sum_{t=1}^{(\tilde{n}_{ij}-1) \wedge n} \mathbb{P}_\theta(\mathbf{Z}_{n-t}^M = \mathbf{i}, Z_{n-t-1} \neq i, Y_n \mid \mathbf{y}_0^{n-1}) \\ &= \sum_{\substack{i,j \\ i \neq j}} \sum_{\substack{1 \leq t \leq \tilde{n}_{ij} \wedge n \\ 0 \leq u \leq (\tilde{n}_{ij}-t) \wedge (M-n)}} \frac{\mathbb{P}_\theta(Z_{n+u} = j, \mathbf{Z}_{n-t}^{n+u-1} = \mathbf{i}, Z_{n-t-1} \neq i, \mathbf{Y}_{n-t}^n \mid \mathbf{y}_0^{n-t})}{\mathbb{P}_\theta(\mathbf{Y}_{n-t}^{n-1} \mid \mathbf{y}_0^{n-t})} \\ &\quad + \sum_i \sum_{t=1}^{(\tilde{n}_{ij}-1) \wedge n} \frac{\mathbb{P}_\theta(\mathbf{Z}_{n-t}^M = \mathbf{i}, Z_{n-t-1} \neq i, \mathbf{Y}_{n-t}^n \mid \mathbf{y}_0^{n-t})}{\mathbb{P}_\theta(\mathbf{Y}_{n-t}^{n-1} \mid \mathbf{y}_0^{n-t})} \end{aligned} \quad (5.38)$$

Nous notons par $A_1(u)$ le terme additif qui apparaît dans le premier terme de l'équation (5.38) (nous le considérons ici comme une fonction de u), i.e.,

$$A_1(u) = \frac{\mathbb{P}_\theta(Z_{n+u} = j, \mathbf{Z}_{n-t}^{n+u-1} = \mathbf{i}, Z_{n-t-1} \neq i, \mathbf{Y}_{n-t}^n \mid \mathbf{y}_0^{n-t})}{\mathbb{P}_\theta(\mathbf{Y}_{n-t}^{n-1} \mid \mathbf{y}_0^{n-t})} \quad (5.39)$$

et par $N_1(u)$ le numérateur de $A_1(u)$. Alors, nous avons

$$\begin{aligned}
N_1(u) &= \mathbb{P}_\theta(Z_{n+u} = j, \mathbf{Z}_{n-t}^{n+u-1} = \mathbf{i}, Z_{n-t-1} \neq i \mid \mathbf{y}_0^{n-t}) \\
&\quad \times \mathbb{P}_\theta(\mathbf{Y}_{n-t}^n \mid Z_{n+u} = j, \mathbf{Z}_{n-t}^{n+u-1} = \mathbf{i}, Z_{n-t-1} \neq i, \mathbf{y}_0^{n-t}) \\
&= F_n(j, \mathbf{i}; u, t) \mathbb{P}_\theta(\mathbf{Y}_{n-t}^n \mid Z_{n+u} = j, \mathbf{Z}_{n-t}^{n+u-1} = \mathbf{i}, \mathbf{y}_{n-t}) \\
&= F_n(j, \mathbf{i}; u, t) \mathbb{1}_{\{Y_{n-t} = y_{n-t}\}} \mathbb{P}_\theta(\mathbf{Y}_{n-t+1}^n \mid Z_{n+u} = j, \mathbf{Z}_{n-t}^{n+u-1} = \mathbf{i}).
\end{aligned} \tag{5.40}$$

Pour $u > 0$, nous avons

$$\mathbb{P}_\theta(\mathbf{Y}_{n-t+1}^n \mid Z_{n+u} = j, \mathbf{Z}_{n-t}^{n+u-1} = \mathbf{i}) = \prod_{p=n-t+1}^n \mathbb{P}_\theta(Y_p \mid Z_p = i), \tag{5.41}$$

et pour $u = 0$,

$$\mathbb{P}_\theta(\mathbf{Y}_{n-t+1}^n \mid Z_n = j, \mathbf{Z}_{n-t}^{n-1} = \mathbf{i}) = \prod_{p=n-t+1}^{n-1} \mathbb{P}_\theta(Y_p \mid Z_p = i) \mathbb{P}_\theta(Y_n \mid Z_n = j). \tag{5.42}$$

Puisque,

$$\mathbb{P}_\theta(\mathbf{Y}_{n-t}^{n-1} \mid \mathbf{y}_0^{n-t}) = \mathbb{1}_{\{Y_{n-t} = y_{n-t}\}} \left(\delta_{t1} + (1 - \delta_{t1}) \prod_{p=n-t+1}^{n-1} \mathbb{P}_\theta(Y_p \mid \mathbf{y}_0^{p-1}) \right), \tag{5.43}$$

en combinant les relations (5.37) et (5.39)-(5.43) nous concluons que

$$A_1(u) = \frac{q_{ij}(t+u) F_{n-t}(i) T_{n,t}(i)}{\prod_{p=n-t+1}^{n-1} P_p}, \quad u > 0, \tag{5.44}$$

$$A_1(0) = \frac{q_{ij}(t) F_{n-t}(i) R_{j; y_n} T_{n-1,t-1}(i)}{\prod_{p=n-t+1}^{n-1} P_p}, \tag{5.45}$$

où nous rappelons que le produit vide est égal à l'unité.

Si nous notons par A_2 le terme additif du deuxième terme de l'équation (5.38), i.e.,

$$A_2 := \frac{\mathbb{P}_\theta(\mathbf{Z}_{n-t}^M = \mathbf{i}, Z_{n-t-1} \neq i, \mathbf{Y}_{n-t}^n \mid \mathbf{y}_0^{n-t})}{\mathbb{P}_\theta(\mathbf{Y}_{n-t}^{n-1} \mid \mathbf{y}_0^{n-t})}, \tag{5.46}$$

alors, avec le même type d'arguments que pour $A_1(u)$, et en tenant en compte que

$$\mathbb{P}_\theta(\mathbf{Z}_{n-t+1}^M = \mathbf{i} \mid Z_{n-t} = i, Z_{n-t-1} \neq i, \mathbf{y}_0^{n-t}) = \bar{H}_i(M - n + t), \tag{5.47}$$

nous pouvons déduire que

$$A_2 = \frac{\bar{H}_i(M - n + t) F_{n-t}(i) T_{n,t}(i)}{\prod_{p=n-t+1}^{n-1} P_p}. \tag{5.48}$$

Par les relations (5.38), (5.44) et (5.45) nous inférons la validité de (5.23).

Maintenant, nous allons d eduire l'expression (5.24) qui concerne le terme $F_n(j)$ d efini par (5.15).

Pour $n = 1, 2, \dots, M$, et $j \in E$, nous avons

$$\begin{aligned}
F_n(j) &= \mathbb{P}_\theta(Z_n = j, Z_{n-1} \neq j \mid \mathbf{y}_0^n) \\
&= \sum_{i \neq j} \sum_{t=1}^{\tilde{n}_{ij} \wedge n} \mathbb{P}_\theta(Z_n = j, \mathbf{Z}_{n-t}^{n-1} = \mathbf{i}, Z_{n-t-1} \neq i \mid \mathbf{y}_0^n) \\
&\quad \sum_{i \neq j} \sum_{t=1}^{\tilde{n}_{ij} \wedge n} \frac{\mathbb{P}_\theta(Z_n = j, \mathbf{Z}_{n-t}^{n-1} = \mathbf{i}, Z_{n-t-1} \neq i, \mathbf{Y}_{n-t+1}^n \mid \mathbf{y}_0^{n-t})}{\mathbb{P}_\theta(\mathbf{Y}_{n-t+1}^n \mid \mathbf{y}_0^{n-t})} \\
&= \sum_{i \neq j} \sum_{t=1}^{\tilde{n}_{ij} \wedge n} \frac{A_1(0)}{P_n}, \tag{5.49}
\end{aligned}$$

o u $A_1(0)$ est donn e par (5.45).

Nous proc edons,  a pr esent,  a la d emonstration de l'expression donn ee pour le terme $L_{1,n}(i)$.

Notons que

$$\begin{aligned}
L_{1,n}(i) &= \sum_{j \neq i} \sum_{t=1}^{\tilde{n}_{ij} \wedge (M-n)} \frac{\mathbb{P}_\theta(Z_{n+t} = j, Z_n^{n+t-1} = \mathbf{i}, Z_{n-1} \neq i, \mathbf{Y}_0^M)}{\mathbb{P}_\theta(\mathbf{Y}_0^M)} \\
&\quad + \frac{\mathbb{P}_\theta(Z_n^M = \mathbf{i}, Z_{n-1} \neq i, \mathbf{Y}_0^M)}{\mathbb{P}_\theta(\mathbf{Y}_0^M)}. \tag{5.50}
\end{aligned}$$

Si nous d esignons par N_1 et N_2 le num erateur et le d enominateur respectivement du premier terme de la relation (5.50), alors

$$\begin{aligned}
N_1 &= \mathbb{P}_\theta(\mathbf{Y}_0^n) \mathbb{P}_\theta(Z_n = i, Z_{n-1} \neq i \mid \mathbf{y}_0^n) \\
&\quad \times \mathbb{P}_\theta(Z_{n+t} = j, \mathbf{Z}_{n+1}^{n+t-1} = \mathbf{i} \mid Z_n = i, Z_{n-1} \neq i) \\
&\quad \times \mathbb{P}_\theta(\mathbf{Y}_{n+1}^{n+t} \mid Z_{n+t} = j, \mathbf{Z}_{n+1}^{n+t-1} = \mathbf{i}) \\
&\quad \times \mathbb{P}_\theta(\mathbf{Y}_{n+t+1}^M \mid Z_{n+t} = j, Z_{n+t-1} \neq j), \tag{5.51}
\end{aligned}$$

et

$$N_2 = \frac{\mathbb{P}_\theta(\mathbf{Y}_0^{n+t}) \mathbb{P}_\theta(Z_{n+t} = j, Z_{n+t-1} \neq j \mid \mathbf{y}_0^{n+t}) \mathbb{P}_\theta(\mathbf{Y}_{n+t+1}^M \mid Z_{n+t} = j, Z_{n+t-1} \neq j)}{\mathbb{P}_\theta(Z_{n+t} = j, Z_{n+t-1} \neq j \mid \mathbf{y}_0^M)}. \tag{5.52}$$

En combinant (5.51) et (5.52) nous obtenons

$$\frac{N_1}{N_2} = \frac{q_{ij}(t) F_n(i) L_{1,n+t}(j) R_{j; \mathbf{y}_{n+t}} T_{n+t-1; t-1}(i)}{F_{n+t}(j) \prod_{p=n+1}^{n+t} P_p}. \tag{5.53}$$

Si nous d esignons par N'_1 et $N'_2 (= N_2)$ le num erateur et le d enominateur respectivement du deuxi eme terme de la relation (5.50), alors

$$\begin{aligned}
\frac{N'_1}{N'_2} &= \frac{\mathbb{P}_\theta(Z_n = i, Z_{n-1} \neq i | \mathbf{y}_0^n) \mathbb{P}_\theta(\mathbf{Z}_{n+1}^M = \mathbf{i} | Z_n = i, Z_{n-1} \neq i) \mathbb{P}_\theta(\mathbf{Y}_{n+1}^M | \mathbf{Z}_{n+1}^M = \mathbf{i})}{\mathbb{P}_\theta(\mathbf{Y}_{n+1}^M | \mathbf{y}_0^n)} \\
&= \frac{\tilde{H}_i(\tilde{n}_i \wedge (M-n)) F_n(i) T_{M, M-n}}{\prod_{p=n+1}^M P_p}.
\end{aligned} \tag{5.54}$$

Par les relations (5.50) et (5.53)-(5.54), nous inférons la relation (5.27).

Afin d'appliquer l'étape-M, nous devons maximiser la fonction $\tilde{Q}(\boldsymbol{\theta} | \boldsymbol{\theta}^{(m)})$, donnée par (5.12), par rapport à $\boldsymbol{\theta}$ et sous les contraintes $C_{q, \text{sto}}$ et $C_{R, \text{sto}}$. En utilisant soit une maximisation directe, soit la méthode des multiplicateurs de Lagrange, nous déduisons que l'EMV $\boldsymbol{\theta}^{(m+1)} = (q_{ij}^{(m+1)}(k), R_{i;a}^{(m+1)})$ est donné pour $i \neq j$ par

$$q_{ij}^{(m+1)}(k) = \frac{\sum_{n=0}^{M-k} \mathbb{P}_{\boldsymbol{\theta}^{(m)}}(Z_{n+k} = j, \mathbf{Z}_n^{n+k-1} = \mathbf{i}, Z_{n-1} \neq i | \mathbf{y})}{\sum_{n=0}^{M-1} \mathbb{P}_{\boldsymbol{\theta}^{(m)}}(Z_n = i, Z_{n-1} \neq i | \mathbf{y}) - \mathbb{P}_{\boldsymbol{\theta}^{(m)}}(Z_M = i, Z_{M-1} = i | \mathbf{y})}, \tag{5.55}$$

$$R_{i;a}^{(m+1)} = \frac{\sum_{n=0}^M \mathbb{1}_{\{Y_n=a\}} \mathbb{P}_{\boldsymbol{\theta}^{(m)}}(Z_n = i | \mathbf{y})}{\sum_{n=0}^M \mathbb{P}_{\boldsymbol{\theta}^{(m)}}(Z_n = i | \mathbf{y})}. \tag{5.56}$$

Étant donné que la probabilité $\mathbb{P}_{\boldsymbol{\theta}^{(m)}}(Z_{n+k} = j, \mathbf{Z}_n^{n+k-1} = \mathbf{i}, Z_{n-1} \neq i | \mathbf{y})$ vaut N_1/N_2 , qui est donné par (5.53), et en utilisant les relations (5.16), (5.17) et (5.18), nous obtenons les relations (5.30) et (5.31).

5.4 ALGORITHME SEM POUR MSMC NON PARAMÉTRIQUE

Dans cette section nous présentons un algorithme EM stochastique (SEM) pour obtenir l'EMV du MSMC non paramétrique. Dans la littérature on peut trouver une grande variété de versions stochastiques pour l'algorithme EM. La principale idée consiste à remplacer le calcul de l'étape-E par une étape de simulation. Dans chaque itération d'un SEM nous devons générer une trajectoire échantillonnée \mathbf{Z}_0^M par la loi conditionnelle $\mathbb{P}_\theta(\cdot | \mathbf{y}_0^M)$.

La version stochastique la plus simple de l'algorithme EM est l'algorithme **SEM** proposé par Celeux et Diebolt (1985). L'algorithme **SEM** remplace la fonction $Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(m)})$ par la fonction

$$\hat{Q}(\boldsymbol{\theta} | \boldsymbol{\theta}^{(m)}) = \log p_M(\mathbf{Z}_0^M, \mathbf{Y}_0^M | \boldsymbol{\theta}), \tag{5.57}$$

où \mathbf{Z}_0^M est une observation générée par $\mathbb{P}_{\boldsymbol{\theta}^{(m)}}(\cdot | \mathbf{y}_0^M)$.

Monte Carlo EM (MCEM) (Wei et Tanner (1990)) est une généralisation naturelle de **SEM**. Dans l'étape- m de la simulation de l'algorithme MCEM,

on génère n_m trajectoires $(\mathbf{Z}_0^M)_i$, $1 \leq i \leq n_m$, échantillonnées indépendamment par $\mathbb{P}_{\boldsymbol{\theta}^{(m)}}(\cdot | \mathbf{y}_0^M)$ au lieu de générer seulement une. Puis, on maximise la fonction

$$\check{Q}(\boldsymbol{\theta} | \boldsymbol{\theta}^{(m)}) = \frac{1}{n_m} \sum_{i=1}^{n_m} \log p_M((\mathbf{Z}_0^M)_i, \mathbf{Y}_0^M | \boldsymbol{\theta}). \quad (5.58)$$

Dans le cas particulier où $n_m = 1$, MCEM se réduit à **SEM**. En outre, il est important de signaler que MCEM converge seulement si la taille de l'échantillon de Monte Carlo augmente dans chaque itération.

Une autre version stochastique de l'EM est l'algorithme EM Approximation Stochastique (SAEM) (Delyon et al. (1999)). Cet algorithme peut être considéré comme une généralisation de MCEM. La principale différence entre SAEM et MCEM se trouve dans le fait que la fonction $Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(m)})$ est estimée de manière récursive en moyennant une approximation de Monte Carlo avec les fonctions correspondantes calculées dans l'étape précédente. Cela signifie que l'algorithme SAEM utilise toutes les données simulées, à l'opposé de MCEM qui oublie les valeurs simulées durant les itérations précédentes et il est basé uniquement sur les données actuelles. Plus précisément, la fonction $Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(m)})$ est estimée par

$$\check{Q}(\boldsymbol{\theta} | \boldsymbol{\theta}^{(m)}) = (1 - \gamma_m) \check{Q}(\boldsymbol{\theta} | \boldsymbol{\theta}^{(m-1)}) + \frac{\gamma_m}{n_m} \sum_{i=1}^{n_m} \log p_M((\mathbf{Z}_0^M)_i, \mathbf{Y}_0^M | \boldsymbol{\theta}), \quad (5.59)$$

où $(\mathbf{Z}_0^M)_i$ sont générées indépendamment par $\mathbb{P}_{\boldsymbol{\theta}^{(m)}}(\cdot | \mathbf{y}_0^M)$, pour $1 \leq i \leq n_m$. Les poids γ_m ont une importance cruciale pour la convergence de l'algorithme. Ils sont choisis typiquement de sorte qu'ils forment une suite positive et décroissante pour toutes les itérations. Polyak et Juditsky (1992) ont démontré que pour une suite de poids $\gamma_m \propto (1/m)^\alpha$ avec $1/2 < \alpha \leq 1$, SAEM a un taux de convergence optimal. Une manière différente pour la sélection de la suite γ_m est proposée par Jank (2006). L'avantage principal de cet algorithme est le fait qu'il converge avec une taille d'échantillon constante, et assez souvent petite, à l'opposé de MCEM qui nécessite une nouvelle valeur de n_m dans chaque itération. Par conséquent, la seule décision que nous devons prendre afin d'appliquer un algorithme SAEM est le choix de la suite γ_m .

Comme l'implémentation de SAEM semble être directe, nous utiliserons cette variante dans les exemples que nous présentons dans la section 5.5, en prenant $\alpha = 1$.

En utilisant un algorithme SEM quelconque, dans chaque itération m , d'abord, nous devons simuler \mathbf{Z}_0^M par la loi conditionnelle $\mathbb{P}_{\boldsymbol{\theta}^{(m)}}(\cdot | \mathbf{y}_0^M)$ et ensuite maximiser la log-vraisemblance des données complètes par rapport à $\boldsymbol{\theta}$.

La simulation démarre en échantillonnant le dernier état visité Z_M et le dernier instant de saut $S_{N(M)}$. Alors, pour tout $j \in E$ et $0 \leq u \leq \tilde{n}_j - 1$, nous avons

$$\begin{aligned}
& \mathbb{P}_\theta(Z_M = j, S_{N(M)} = M - u \mid \mathbf{y}_0^M) \\
&= \mathbb{P}_\theta(\mathbf{Z}_{M-u}^M = \mathbf{j}, Z_{M-u-1} \neq j \mid \mathbf{y}_0^M) \\
&= \mathbb{P}_\theta(\mathbf{Z}_{M-u}^M = \mathbf{j}, Z_{M-u-1} \neq j, \mathbf{Y}_0^M) / \mathbb{P}_\theta(\mathbf{Y}_0^M) \\
&= \mathbb{P}_\theta(Z_{M-u} = j, Z_{M-u-1} \neq j \mid \mathbf{y}_0^{M-u}) \mathbb{P}_\theta(\mathbf{Z}_{M-u}^M = \mathbf{j} \mid Z_{M-u} = j, Z_{M-u-1} \neq j) \\
&\quad \times \mathbb{P}_\theta(\mathbf{Y}_{M-u+1}^M \mid \mathbf{Z}_{M-u+1}^M = \mathbf{j}) / \mathbb{P}_\theta(\mathbf{Y}_{M-u+1}^M \mid \mathbf{y}_0^{M-u}).
\end{aligned}$$

En utilisant les quantités déjà définies dans les sections 5.2 et 5.3, pour tout $j \in E$ et $0 \leq u \leq \tilde{n}_i - 1$, nous avons

$$\mathbb{P}_\theta(Z_M = j, S_{N(M)} = M - u \mid \mathbf{y}_0^M) = \frac{\tilde{H}_j(u) F_{M-u}(j) T_{M,u}(j)}{\prod_{p=M-u+1}^M P_p}. \quad (5.60)$$

Pour les étapes intermédiaires, la relation récursive qui nous donne l'état visité précédemment et l'instant de saut correspondant, c'est-à-dire, le couple (J_{k-1}, S_{k-1}) conditionné par la suite $((J_l, S_l)_{l=k}^{N(M)}, \mathbf{Y}_0^M)$, peut être obtenu en notant que

$$\begin{aligned}
& \mathbb{P}_\theta(J_{k-1} = i, S_k - S_{k-1} = u \mid (J_l, S_l)_{l=k+1}^{N(M)}, S_k = t, J_k = j, \mathbf{y}_0^M) \\
&= \mathbb{P}_\theta(\mathbf{Z}_{t-u}^{t-1} = \mathbf{i}, Z_{t-u-1} \neq i \mid S_k = t, Z_t = j, \mathbf{Z}_{t+1}^M, \mathbf{y}_0^M) \\
&= \frac{\mathbb{P}_\theta(\mathbf{Z}_{t-u}^{t-1} = \mathbf{i}, Z_{t-u-1} \neq i, S_k = t, Z_t = j, \mathbf{Z}_{t+1}^M, \mathbf{Y}_0^M)}{\mathbb{P}_\theta(S_k = t, Z_t = j, \mathbf{Z}_{t+1}^M, \mathbf{Y}_0^M)}. \quad (5.61)
\end{aligned}$$

Le numérateur N_1 de l'équation (5.61) s'écrit :

$$\begin{aligned}
N_1 &= \mathbb{P}_\theta(\mathbf{Y}_0^{t-u}) \mathbb{P}_\theta(Z_{t-u} = i, Z_{t-u-1} \neq i \mid \mathbf{y}_0^{t-u}) \\
&\quad \times \mathbb{P}_\theta(Z_t = j, \mathbf{Z}_{t-u+1}^{t-1} = \mathbf{i} \mid Z_{t-u} = i, Z_{t-u-1} \neq i) \\
&\quad \times \mathbb{P}_\theta(\mathbf{Y}_{t-u+1}^t \mid Z_t = j, \mathbf{Z}_{t-u+1}^{t-1} = \mathbf{i}) \\
&\quad \times \mathbb{P}_\theta(\mathbf{Z}_{t+1}^M, \mathbf{Y}_{t+1}^M \mid Z_t = j, Z_{t-1} \neq j). \quad (5.62)
\end{aligned}$$

Le dénominateur N_2 de l'équation (5.61) s'écrit :

$$\begin{aligned}
N_2 &= \mathbb{P}_\theta(\mathbf{Y}_0^t) \mathbb{P}_\theta(Z_t = j, Z_{t-1} \neq j \mid \mathbf{y}_0^t) \\
&\quad \times \mathbb{P}_\theta(\mathbf{Z}_{t+1}^M, \mathbf{Y}_{t+1}^M \mid Z_t = j, Z_{t-1} \neq j). \quad (5.63)
\end{aligned}$$

Finalement, par les équations (5.61), (5.62), (5.63) et les quantités définies dans la section 5.2, nous obtenons

$$\begin{aligned}
\mathbb{P}_\theta(J_{k-1} = i, S_k - S_{k-1} = u \mid (J_l, S_l)_{l=k+1}^{N(M)}, S_k = t, J_k = j, \mathbf{y}_0^M) \\
= \frac{q_{ij}(u) F_{t-u}(i) R_{j,y_t} T_{t-1,u-1}(i)}{F_t(j) \prod_{p=t-u+1}^t P_p}. \quad (5.64)
\end{aligned}$$

Remarque 5.5 *i) Dans le cas particulier où l'espace d'état de \mathbf{Z} a seulement 2 états, après que le dernier état Z_M soit échantillonné, il y aura un changement déterministe des états de la CM immergée \mathbf{J} . Donc, il suffit d'échantillonner uniquement les lois des temps de séjour dans chaque état.*

ii) Une approche similaire pour échantillonner $\mathbf{Z}_0^M \sim \mathbb{P}_{\theta^{(m)}}(\cdot | \mathbf{y}_0^M)$ est proposée par Guédon (2007). Dans son algorithme, l'auteur propose de tirer successivement l'état actuel Z_t aux instants de saut, en démarrant par l'état Z_M , et l'occupation d'état conditionnée par l'état actuel et par les observations \mathbf{y} . En suivant cette approche, un algorithme avant-arrière différent est nécessaire.

iii) Le principal avantage de la version stochastique de l'algorithme EM réside dans le fait que seulement un procédé vers l'avant est nécessaire. Ce qui diminue significativement le temps CPU par itération. Ainsi, l'algorithme SEM permet d'analyser des trajectoires assez grandes, ce qui n'est pas faisable avec l'algorithme EM original dans un temps raisonnable. Ceci ouvre la voie à de nombreuses applications, particulièrement à l'analyse d'ADN.

5.5 APPLICATIONS

Afin d'illustrer la performance de nos algorithmes EM et SEM nous présentons dans cette section deux exemples numériques. Le premier concerne un MSMC à deux états observables et cachés et l'autre traite le cas d'un MSMC à quatre états observables et deux états cachés. Dans chaque exemple, nous considérons deux cas qui correspondent à deux trajectoires \mathbf{Y}_0^M de différentes longueurs, l'une avec $M = 5 \times 10^4$ et l'autre avec $M = 8 \times 10^4$.

Notons que pour l'implémentation numérique des algorithmes, nous avons utilisé le langage Fortran sur une machine avec un processeur Intel Core 2 6400 de fréquence 2,13GHz et mémoire vive (RAM) 1GB.

Nous notons aussi que ce travail a été effectué en collaboration avec S. Malefaki.

Exemple 5.1 *Nous considérons le MSMC (\mathbf{Z}, \mathbf{Y}) où l'espace d'état de \mathbf{Z} et de \mathbf{Y} est l'ensemble $E = \{0, 1\}$.*

Nous appliquons les algorithmes EM et SEM afin de trouver l'EMV de ce modèle. Nous utilisons pour cet exemple deux trajectoires différentes.

Cas 1 : *Nous supposons que nous observons une trajectoire \mathbf{Y}_0^M avec $M = 5 \times 10^4$.*

Cette trajectoire a été générée en utilisant des variables aléatoires de loi Weibull discrète¹ pour le noyau ($q_{01}(\cdot) := \mathcal{W}(0.7, 0.9)$ et $q_{10}(\cdot) := \mathcal{W}(0.5, 0.7)$), et en prenant des lois conditionnelles de Y_n sachant Z_n avec des vecteurs de probabilité $\mathbf{R}_0 := (0.8, 0.2)$ et $\mathbf{R}_1 := (0.2, 0.8)$.

Cas 2 : *Nous utilisons ici une autre trajectoire \mathbf{Y}_0^M avec $M = 8 \times 10^4$.*

Cette trajectoire a été échantillonnée du même noyau que le cas précédent, mais des lois conditionnelles de Y_n sachant Z_n avec des vecteurs de probabilité $\mathbf{R}_0 := (0.8, 0.2)$ et $\mathbf{R}_1 := (0.5, 0.5)$.

¹ $X \sim \mathcal{W}(q, b)$ si $\mathbb{P}(X = n) = q^{(n-1)b} - q^{nb}$, $n \in \mathbb{N}^*$.

Pour chaque cas, trois différentes valeurs initiales de $\theta^{(0)} = (q_{01}^{(0)}, q_{10}^{(0)}, R_0^{(0)}, R_1^{(0)})$ ont été utilisées pour les algorithmes EM et SEM. Elles sont définies comme suit :

Cas 1 :

$$\begin{aligned}
 (1a) \quad & \left\{ \begin{array}{l} q_{01}^{(0)}(1) = 0.3, q_{01}^{(0)}(2) = 0.2, q_{01}^{(0)}(3) = 0.1, q_{01}^{(0)}(i) = \frac{0.4}{12}, \text{ pour } i=4, \dots, 15, \\ q_{10}^{(0)}(1) = 0.5, q_{10}^{(0)}(2) = 0.2, q_{10}^{(0)}(3) = 0.1, q_{10}^{(0)}(i) = \frac{0.2}{7}, \text{ pour } i=4, \dots, 10, \\ R_0^{(0)} = (0.8, 0.2), R_1^{(0)} = (0.2, 0.8). \end{array} \right. \\
 (1b) \quad & \left\{ \begin{array}{l} q_{01}^{(0)}(i) = \frac{1}{15}, \text{ pour } i = 1, \dots, 15, \\ q_{10}^{(0)}(i) = \frac{1}{10}, \text{ pour } i = 1, \dots, 10, \\ R_0^{(0)} = (0.8, 0.2), R_1^{(0)} = (0.2, 0.8). \end{array} \right. \quad (5.65) \\
 (1c) \quad & \left\{ \begin{array}{l} q_{01}^{(0)}(i) = \frac{1}{15}, \text{ pour } i = 1, \dots, 15, \\ q_{10}^{(0)}(i) = \frac{1}{10}, \text{ pour } i = 1, \dots, 10, \\ R_0^{(0)} = (0.6, 0.4), R_1^{(0)} = (0.4, 0.6). \end{array} \right.
 \end{aligned}$$

Cas 2 :

$$\begin{aligned}
 (1a) \quad & \left\{ \begin{array}{l} q_{01}^{(0)}(1) = 0.3, q_{01}^{(0)}(2) = 0.2, q_{01}^{(0)}(3) = 0.1, q_{01}^{(0)}(i) = \frac{0.4}{12}, \text{ pour } i=4, \dots, 15, \\ q_{10}^{(0)}(1) = 0.5, q_{10}^{(0)}(2) = 0.2, q_{10}^{(0)}(3) = 0.1, q_{10}^{(0)}(i) = \frac{0.2}{7}, \text{ pour } i=4, \dots, 10, \\ R_0^{(0)} = (0.8, 0.2), R_1^{(0)} = (0.5, 0.5). \end{array} \right. \\
 (1\beta) \quad & \left\{ \begin{array}{l} q_{01}^{(0)}(i) = \frac{1}{15}, \text{ pour } i = 1, \dots, 15, \\ q_{10}^{(0)}(i) = \frac{1}{10}, \text{ pour } i = 1, \dots, 10, \\ R_0^{(0)} = (0.8, 0.2), R_1^{(0)} = (0.5, 0.5). \end{array} \right. \quad (5.66) \\
 (1\gamma) \quad & \left\{ \begin{array}{l} q_{01}^{(0)}(i) = \frac{1}{15}, \text{ pour } i = 1, \dots, 15, \\ q_{10}^{(0)}(i) = \frac{1}{10}, \text{ pour } i = 1, \dots, 10, \\ R_0^{(0)} = (0.75, 0.25), R_1^{(0)} = (0.45, 0.55). \end{array} \right.
 \end{aligned}$$

Dans les tableaux 5.1, 5.2 et 5.3 et les figures 5.1 et 5.2 nous donnons les valeurs estimées des paramètres pour les différentes valeurs initiales données ci-dessus pour le cas 1. Dans les tableaux 5.5, 5.6 et 5.7, et dans les figures 5.3, 5.4 et 5.5 nous présentons les résultats correspondants au cas 2. Et dans les tableaux 5.4 et 5.8, nous donnons le temps moyen d'exécution (en secondes) par itération pour EM et SEM pour les deux cas.

Nous remarquons d'après les tableaux et les figures correspondants au cas 1 que l'estimation pour les douze premiers paramètres de q_{01} est bonne, alors que pour les trois derniers paramètres dont les vraies valeurs sont proches de zéro, elle n'est pas satisfaisante. Également, l'estimation pour les neuf premiers paramètres de q_{10} est bonne, ce qui n'est pas le cas pour le dernier paramètre. Par ailleurs, l'estimation pour les vecteurs R_0 et R_1 est assez bonne.

Ces résultats montrent également que l'algorithme SEM donne presque les mêmes estimations que l'algorithme EM, mais en utilisant SEM avec un échantillon de taille 100 à chaque itération, nous avons besoin de presque dix fois moins de temps CPU que l'EM correspondant (voir le tableau 5.4). Ceci est dû à l'étape arrière qui exige beaucoup de calcul et qui est nécessaire dans l'EM et pas dans SEM. Notons également que les estimations sont proches des vraies valeurs avec les deux algorithmes, même si nous commençons par des valeurs initiales qui ne sont pas proches des vraies valeurs, tels les cas (1_b) et (1_c).

Dans le cas 2, les probabilités d'émission sont bien estimées en utilisant les deux algorithmes, et ceci indépendamment du choix des valeurs initiales. Mais, ceci n'est pas le cas pour le noyau semi-markovien. Pour celui-ci, les valeurs estimées sont proches des vraies valeurs uniquement quand les valeurs initiales sont proches des vraies valeurs, comme par exemple, le cas (1_a). Et, pour les autres cas, même après 200 itérations des algorithmes EM et SEM, les résultats ne sont pas précis. Dans ces cas, nous pouvons avoir une idée sur la forme des lois mais nous ne pouvons pas corriger les estimations ponctuelles des probabilités. En fait, plus que 200 itérations sont nécessaires afin d'achever la convergence de l'algorithme. La figure 5.5 décrit les résultats de l'algorithme SEM après 300 et 600 itérations, en partant de la valeur initiale décrite par (1_b).

Dans ce cas ($M = 80000$), le temps moyen d'exécution par itération est approximativement égal à 825 secondes. Ce qui signifie qu'afin d'accomplir 200 itérations pour l'algorithme EM amélioré, nous avons besoin de presque 2 jours. Alors que, l'algorithme SEM, avec un échantillon de taille 500 à chaque itération, nécessite presque 25 secondes par itération. Ainsi, afin d'implémenter 1000 itérations de l'EM nous avons besoin de 10 jours et uniquement 7 heures pour l'algorithme SEM.

En analysant les résultats numériques donnés dans le premier exemple pour les deux cas et plusieurs autres cas que nous ne présentons pas ici, nous déduisons que l'algorithme SEM est préférable, particulièrement

		EM		SEM	
$(\mathbf{1}_a)$	\mathbf{R}_0	0.774	0.226	0.772	0.228
	\mathbf{R}_1	0.162	0.838	0.164	0.836
$(\mathbf{1}_b)$	\mathbf{R}_0	0.757	0.243	0.757	0.243
	\mathbf{R}_1	0.192	0.808	0.195	0.805
$(\mathbf{1}_c)$	\mathbf{R}_0	0.753	0.247	0.756	0.244
	\mathbf{R}_1	0.201	0.799	0.200	0.800

TAB. 5.1 – *Exemple1-cas1*. Estimation des vecteurs de probabilité $\mathbf{R}_0 = (0.8, 0.2)$ et $\mathbf{R}_1 = (0.2, 0.8)$, en utilisant les algorithmes EM et SEM avec 100 itérations et en partant des valeurs initiales données dans $(\mathbf{1}_a)$, $(\mathbf{1}_b)$ et $(\mathbf{1}_c)$.

dans des cas où la convergence est lente.

		EM			SEM		
Vraie valeur		$(\mathbf{1}_a)$	$(\mathbf{1}_b)$	$(\mathbf{1}_c)$	$(\mathbf{1}_a)$	$(\mathbf{1}_b)$	$(\mathbf{1}_c)$
1	0.3000	0.3201	0.3104	0.3079	0.3174	0.3125	0.3182
2	0.1860	0.1629	0.1403	0.1341	0.1593	0.1320	0.1251
3	0.1306	0.1258	0.1185	0.1151	0.1286	0.1212	0.1256
4	0.0946	0.0580	0.0649	0.0667	0.0579	0.0675	0.0580
5	0.0697	0.0756	0.0723	0.0721	0.0720	0.0695	0.0753
6	0.0520	0.0275	0.0320	0.0342	0.0293	0.0317	0.0411
7	0.0391	0.0619	0.0579	0.0555	0.0598	0.0563	0.0485
8	0.0296	0.0131	0.0175	0.0190	0.0157	0.0211	0.0188
9	0.0225	0.0372	0.0390	0.0385	0.0407	0.0367	0.0419
10	0.0172	0.0139	0.0170	0.0179	0.0143	0.0202	0.0158
11	0.0132	0.0232	0.0232	0.0231	0.0232	0.0229	0.0221
12	0.0101	0.0092	0.0103	0.0116	0.0082	0.0108	0.0080
13	0.0078	0.0141	0.0228	0.0247	0.0155	0.0211	0.0246
14	0.0061	0.0322	0.0370	0.0384	0.0304	0.0366	0.0371
15	0.0047	0.0253	0.0369	0.0414	0.0276	0.0402	0.0397

TAB. 5.2 – *Exemple1-cas1*. Estimation des $q_{01}(k)$, $k = 1, \dots, 15$, en utilisant les algorithmes EM et SEM avec 100 itérations et en partant des valeurs initiales données dans $(\mathbf{1}_a)$, $(\mathbf{1}_b)$ et $(\mathbf{1}_c)$.

Exemple 5.2 Nous considérons dans cet exemple le MSMC (\mathbf{Z}, \mathbf{Y}) tel que $Z_n \in \{0, 1\}$ et $Y_n \in \{0, 1, 2, 3\}$.

Nous cherchons à calculer l'EMV de ce modèle en utilisant les algorithmes EM et SEM. Nous utilisons deux trajectoires différentes, décrites dans ce qui suit.

Cas 1 : Nous supposons que nous observons une trajectoire \mathbf{Y}_0^M avec $M = 5 \times 10^4$.

Cette trajectoire a été simulée avec les caractéristiques suivantes. Les lois des temps de séjour pour les états 0 et 1 sont $q_{01}(\cdot) := \mathcal{W}(0.7, 0.9)$ et $q_{10}(\cdot) := \mathcal{W}(0.5, 0.7)$ respectivement et les lois conditionnelles de Y_n sachant Z_n sont données par les vecteurs de probabilité $\mathbf{R}_0 := (0.4, 0.3, 0.2, 0.1)$ et $\mathbf{R}_1 := (0.1, 0.2, 0.3, 0.4)$.

		EM			SEM		
Vraie valeur		(1 _a)	(1 _b)	(1 _c)	(1 _a)	(1 _b)	(1 _c)
1	0.5000	0.5295	0.4970	0.4851	0.5294	0.4847	0.4879
2	0.1757	0.1588	0.1545	0.1542	0.1563	0.1692	0.1598
3	0.1002	0.1170	0.1199	0.1205	0.1171	0.1084	0.1141
4	0.0636	0.0705	0.0770	0.0777	0.0703	0.0805	0.0731
5	0.0427	0.0245	0.0267	0.0293	0.0239	0.0301	0.0360
6	0.0298	0.0251	0.0263	0.0267	0.0262	0.0281	0.0265
7	0.0213	0.0256	0.0290	0.0291	0.0270	0.0304	0.0282
8	0.0156	0.0051	0.0085	0.0102	0.0067	0.0074	0.0089
9	0.0115	0.0087	0.0126	0.0148	0.0072	0.0129	0.0140
10	0.0087	0.0352	0.0485	0.0524	0.0360	0.0483	0.0514

TAB. 5.3 – *Exemple1-cas1*. Estimation des $q_{10}(k)$, $k = 1, \dots, 10$, en utilisant les algorithmes EM et SEM avec 100 itérations et en partant des valeurs initiales données dans (1_a), (1_b) et (1_c).

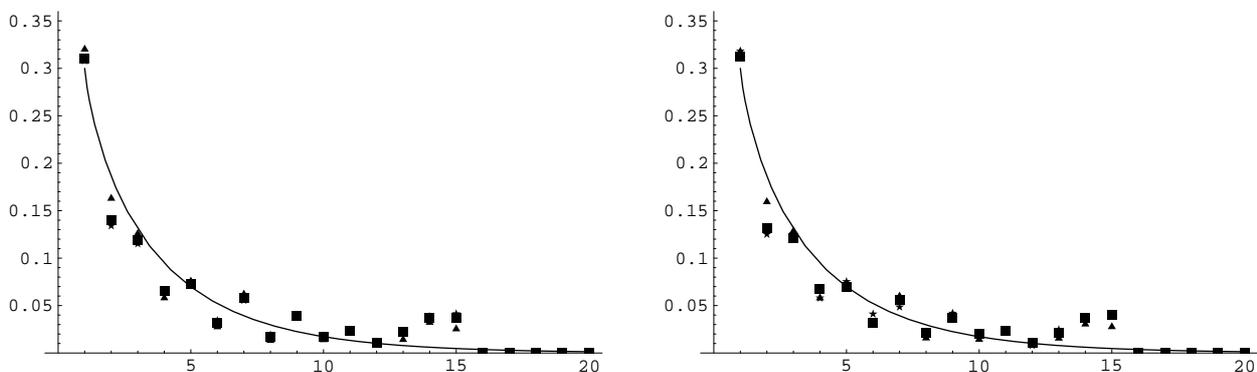


FIG. 5.1 – *Exemple1-cas1*. À gauche : Les vraies valeurs (ligne continue) et les valeurs estimées de $q_{01}(k)$, $k = 1, \dots, 15$, en utilisant l'algorithme EM. À droite : Les vraies valeurs (ligne continue) et les valeurs estimées en utilisant l'algorithme SEM. Les valeurs estimées obtenues en utilisant les valeurs initiales (1_a) sont représentées par des triangles, et celles obtenues en utilisant les valeurs initiales (1_b) par des carrés et celles avec les valeurs initiales (1_c) par des étoiles.

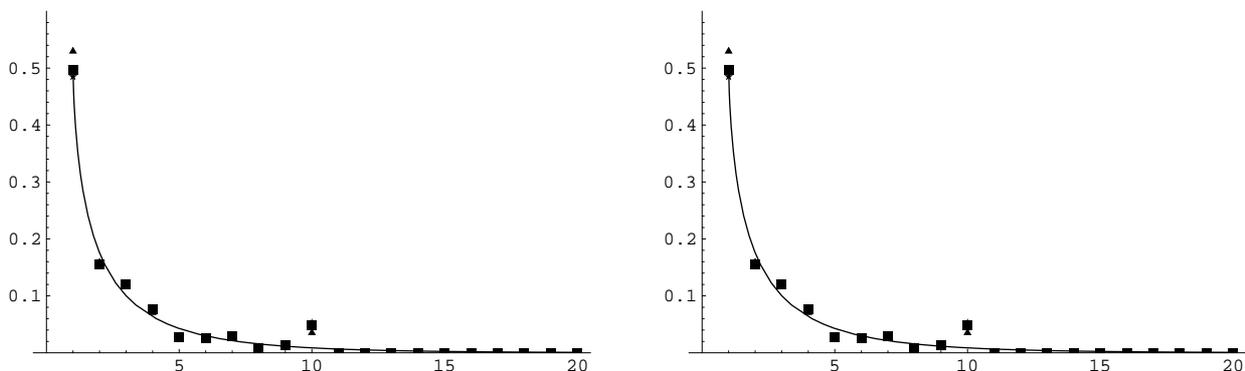


FIG. 5.2 – *Exemple1-cas1*. À gauche : Les vraies valeurs (ligne continue) et les valeurs estimées de $q_{10}(k)$, $k = 1, \dots, 10$, en utilisant l'algorithme EM. À droite : Les vraies valeurs (ligne continue) et les valeurs estimées en utilisant l'algorithme SEM. Les valeurs estimées obtenues en utilisant les valeurs initiales (1_a) sont représentées par des triangles, et celles obtenues en utilisant les valeurs initiales (1_b) par des carrés et celles avec les valeurs initiales (1_c) par des étoiles.

	EM	SEM
$(\mathbf{1}_a)$	29.69 (29.25)	3.37
$(\mathbf{1}_b)$	32.85 (32.15)	3.14
$(\mathbf{1}_c)$	32.56 (31.89)	3.11

TAB. 5.4 – *Exemple1-cas1*. Le temps moyen d'exécution (en secondes) par itération pour EM et SEM calculé en 100 itérations. Nous donnons aussi en parenthèse le temps moyen CPU correspondant pour l'étape arrière dans l'algorithme EM. La taille d'échantillon dans l'algorithme SEM est 100.

		EM		SEM	
$(\mathbf{1}_\alpha)$	R_0	0.7898	0.2102	0.7899	0.2101
	R_1	0.4938	0.5062	0.4938	0.5062
$(\mathbf{1}_\beta)$	R_0	0.7762	0.2238	0.7771	0.2229
	R_1	0.5442	0.4558	0.5445	0.4555
$(\mathbf{1}_\gamma)$	R_0	0.7712	0.2288	0.7717	0.2283
	R_1	0.5365	0.4635	0.5381	0.4619

TAB. 5.5 – *Exemple1-cas2*. Estimation des vecteurs de probabilité R_0 et R_1 , en utilisant les algorithmes EM et SEM avec 200 itérations et en partant des valeurs initiales données dans $(\mathbf{1}_\alpha)$, $(\mathbf{1}_\beta)$ et $(\mathbf{1}_\gamma)$.

		EM			SEM		
Vraie valeur		$(\mathbf{1}_\alpha)$	$(\mathbf{1}_\beta)$	$(\mathbf{1}_\gamma)$	$(\mathbf{1}_\alpha)$	$(\mathbf{1}_\beta)$	$(\mathbf{1}_\gamma)$
1	0.3000	0.3568	0.2212	0.1951	0.3368	0.2228	0.1808
2	0.1860	0.1972	0.1410	0.1310	0.2025	0.1432	0.1345
3	0.1306	0.0977	0.1020	0.0996	0.1041	0.1158	0.1251
4	0.0946	0.0456	0.0865	0.0889	0.0493	0.0818	0.0889
5	0.0697	0.0380	0.0626	0.0664	0.0364	0.0553	0.0720
6	0.0520	0.0351	0.0512	0.0557	0.0365	0.0521	0.0467
7	0.0391	0.0498	0.0442	0.0488	0.0535	0.0383	0.0432
8	0.0296	0.0267	0.0241	0.0265	0.0292	0.0219	0.0254
9	0.0225	0.0132	0.0194	0.0209	0.0144	0.0142	0.0242
10	0.0172	0.0219	0.0288	0.0308	0.0205	0.0337	0.0242
11	0.0132	0.0266	0.0337	0.0363	0.0293	0.0275	0.0358
12	0.0101	0.0202	0.0360	0.0391	0.0178	0.0393	0.0386
13	0.0078	0.0299	0.0424	0.0461	0.0343	0.0328	0.0499
14	0.0061	0.0216	0.0438	0.0474	0.0193	0.0493	0.0513
15	0.0047	0.0196	0.0632	0.0675	0.0162	0.0724	0.0711

TAB. 5.6 – *Exemple1-cas2*. Estimations des $q_{01}(k)$, $k = 1, \dots, 15$, en utilisant les algorithmes EM et SEM avec 200 itérations et en partant des valeurs initiales données dans $(\mathbf{1}_\alpha)$, $(\mathbf{1}_\beta)$ et $(\mathbf{1}_\gamma)$.

Vraie valeur		EM			SEM		
		$(\mathbf{1}_\alpha)$	$(\mathbf{1}_\beta)$	$(\mathbf{1}_\gamma)$	$(\mathbf{1}_\alpha)$	$(\mathbf{1}_\beta)$	$(\mathbf{1}_\gamma)$
1	0.5000	0.4819	0.2194	0.2450	0.4966	0.1950	0.2478
2	0.1757	0.2126	0.1635	0.1732	0.1799	0.1739	0.1584
3	0.1002	0.0817	0.1020	0.1020	0.0906	0.0956	0.0924
4	0.0636	0.0452	0.1085	0.1038	0.0466	0.1082	0.0918
5	0.0427	0.0530	0.0912	0.0834	0.0572	0.1034	0.0981
6	0.0298	0.0311	0.0605	0.0533	0.0342	0.0600	0.0539
7	0.0213	0.0321	0.0538	0.0470	0.0312	0.0654	0.0558
8	0.0156	0.0177	0.0485	0.0380	0.0138	0.0367	0.0389
9	0.0115	0.0195	0.0588	0.0539	0.0184	0.0574	0.0576
10	0.0087	0.0252	0.1054	0.1003	0.0314	0.1043	0.0953

TAB. 5.7 – *Exemple1-cas2*. Estimations des $q_{10}(k)$, $k = 1, \dots, 10$, en utilisant les algorithmes EM et SEM avec 200 itérations et en partant des valeurs initiales données dans $(\mathbf{1}_\alpha)$, $(\mathbf{1}_\beta)$ et $(\mathbf{1}_\gamma)$.

	EM	SEM
$(\mathbf{1}_\alpha)$	891.85	25.21
$(\mathbf{1}_\beta)$	837.59	26.75
$(\mathbf{1}_\gamma)$	828.73	27.32

TAB. 5.8 – *Exemple1-cas2*. Le temps moyen d'exécution (en secondes) par itération pour les algorithmes EM et SEM calculé en 200 itérations. La taille d'échantillon dans l'algorithme SEM est 500.

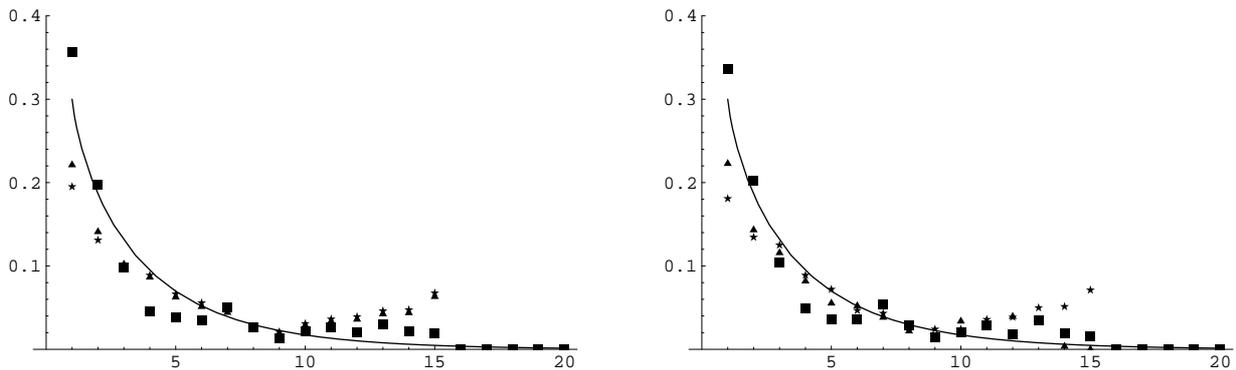


FIG. 5.3 – *Exemple1-cas2*. À gauche : Les vraies valeurs (ligne continue) et les valeurs estimées de $q_{01}(k)$, $k = 1, \dots, 15$, en utilisant l'algorithme EM. À droite : Les vraies valeurs (ligne continue) et les valeurs estimées en utilisant l'algorithme SEM. Les valeurs estimées obtenues en utilisant les valeurs initiales $(\mathbf{1}_\alpha)$ sont représentées par des carrés, et celles obtenues en utilisant les valeurs initiales $(\mathbf{1}_\beta)$ par des triangles et celles avec les valeurs initiales $(\mathbf{1}_\gamma)$ par des étoiles.

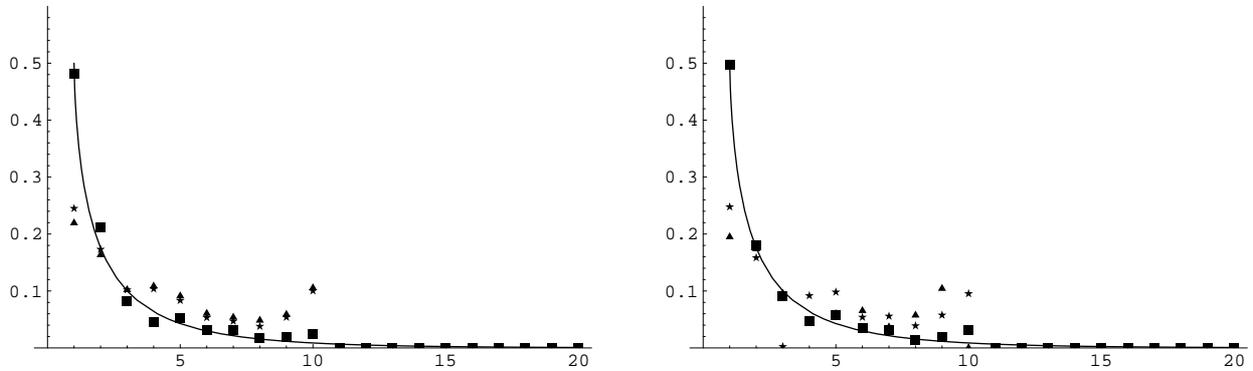


FIG. 5.4 – *Exemple1-cas2*. À gauche : Les vraies valeurs (ligne continue) et les valeurs estimées de $q_{10}(k)$, $k = 1, \dots, 10$, en utilisant l'algorithme EM. À droite : Les vraies valeurs (ligne continue) et les valeurs estimées en utilisant l'algorithme SEM. Les valeurs estimées obtenues en utilisant les valeurs initiales $(\mathbf{1}_\alpha)$ sont représentées par des carrés, et celles obtenues en utilisant les valeurs initiales $(\mathbf{1}_\beta)$ par des triangles et celles avec les valeurs initiales $(\mathbf{1}_\gamma)$ par des étoiles.

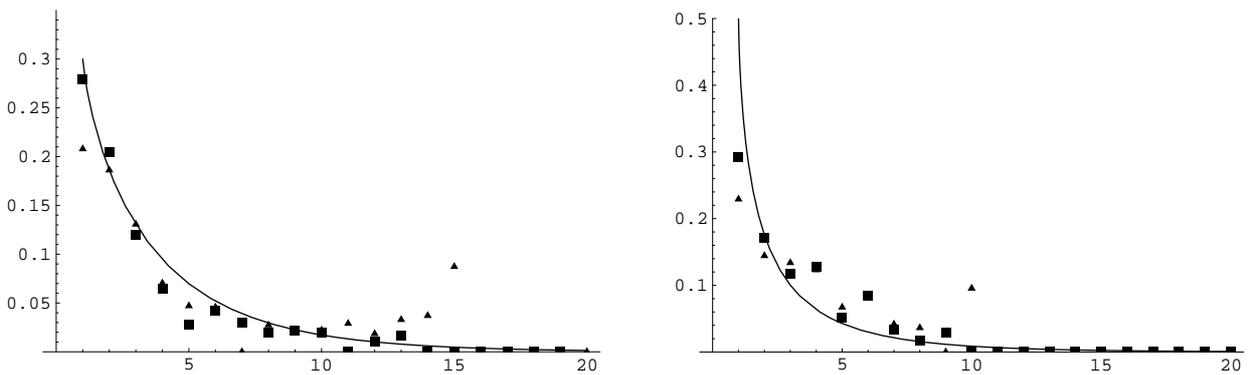


FIG. 5.5 – *Exemple1-cas2*. À gauche : Nous présentons les vraies valeurs (ligne continue) et les valeurs estimées de $q_{01}(k)$, $k = 1, \dots, 15$, avec 300 (triangles) et 600 (carrés) itérations en utilisant l'algorithme SEM. À droite : Nous présentons les vraies valeurs (ligne continue) et les valeurs estimées de $q_{10}(k)$, $k = 1, \dots, 15$, avec 300 (triangles) et 600 (carrés) itérations en utilisant l'algorithme SEM.

Cas 2 : Nous utilisons ici une autre trajectoire \mathbf{Y}_0^M avec $M = 8 \times 10^4$.

Nous gardons les mêmes lois pour les temps de séjour que le cas précédent, et prenons des vecteurs de probabilité pour les lois conditionnelles de Y_n sachant Z_n égales à $\mathbf{R}_0 := (0.25, 0.25, 0.25, 0.25)$ et $\mathbf{R}_1 := (0.1, 0.4, 0.4, 0.1)$.

Pour le premier cas, nous avons utilisé deux valeurs initiales de $\theta^{(0)} = (q_{01}^{(0)}, q_{10}^{(0)}, \mathbf{R}_0^{(0)}, \mathbf{R}_1^{(0)})$ en commun pour les algorithmes EM et SEM. Ces valeurs sont décrites par (2a) et (2b), données ci-dessous. Et, nous avons considéré de plus, deux valeurs initiales uniquement pour l'algorithme SEM. Elles sont définies par (2c) et (2d).

Pour le deuxième cas, deux valeurs initiales communes aux deux algorithmes ont été utilisées. Elles sont décrites par (2α) et (2β).

Cas 1 :

$$(2a) \left\{ \begin{array}{l} q_{01}^{(0)}(1) = 0.3, q_{01}^{(0)}(2) = 0.2, q_{01}^{(0)}(3) = 0.1, q_{01}^{(0)}(i) = \frac{0.4}{12}, \text{ pour } i=4, \dots, 15, \\ q_{10}^{(0)}(1) = 0.5, q_{10}^{(0)}(2) = 0.2, q_{10}^{(0)}(3) = 0.1, q_{10}^{(0)}(i) = \frac{0.2}{7}, \text{ pour } i=4, \dots, 10, \\ \mathbf{R}_0^{(0)} = (0.4, 0.3, 0.2, 0.1), \mathbf{R}_1^{(0)} = (0.1, 0.2, 0.3, 0.4). \end{array} \right.$$

$$(2b) \left\{ \begin{array}{l} q_{01}^{(0)}(i) = \frac{1}{15}, \text{ pour } i = 1, \dots, 15, \\ q_{10}^{(0)}(i) = \frac{1}{10}, \text{ pour } i = 1, \dots, 10, \\ \mathbf{R}_0^{(0)} = (0.4, 0.3, 0.2, 0.1), \mathbf{R}_1^{(0)} = (0.1, 0.2, 0.3, 0.4). \end{array} \right.$$

(5.67)

$$(2c) \left\{ \begin{array}{l} q_{01}^{(0)}(i) = \frac{1}{15}, \text{ pour } i = 1, \dots, 15, \\ q_{10}^{(0)}(i) = \frac{1}{10}, \text{ pour } i = 1, \dots, 10, \\ \mathbf{R}_0^{(0)} = (0.5, 0.25, 0.15, 0.1), \mathbf{R}_1^{(0)} = (0.1, 0.15, 0.25, 0.5). \end{array} \right.$$

$$(2d) \left\{ \begin{array}{l} q_{01}^{(0)}(i) = \frac{1}{15}, \text{ pour } i = 1, \dots, 15, \\ q_{10}^{(0)}(i) = \frac{1}{10}, \text{ pour } i = 1, \dots, 10, \\ \mathbf{R}_0^{(0)} = (0.25, 0.25, 0.25, 0.25), \mathbf{R}_1^{(0)} = (0.1, 0.15, 0.25, 0.5). \end{array} \right.$$

Cas 2 :

$$\begin{aligned}
 (2_\alpha) \left\{ \begin{array}{l} q_{01}^{(0)}(1) = 0.3, q_{01}^{(0)}(2) = 0.2, q_{01}^{(0)}(3) = 0.1, q_{01}^{(0)}(i) = \frac{0.4}{12}, \text{ pour } i=4, \dots, 15, \\ q_{10}^{(0)}(1) = 0.5, q_{10}^{(0)}(2) = 0.2, q_{10}^{(0)}(3) = 0.1, q_{10}^{(0)}(i) = \frac{0.2}{7}, \text{ pour } i=4, \dots, 10, \\ \mathbf{R}_0^{(0)} = (0.25, 0.25, 0.25, 0.25), \mathbf{R}_1^{(0)} = (0.1, 0.4, 0.4, 0.1). \end{array} \right. \\
 (2_\beta) \left\{ \begin{array}{l} q_{01}^{(0)}(i) = \frac{1}{15}, \text{ pour } i = 1, \dots, 15, \\ q_{10}^{(0)}(i) = \frac{1}{10}, \text{ pour } i = 1, \dots, 10, \\ \mathbf{R}_0^{(0)} = (0.25, 0.25, 0.25, 0.25), \mathbf{R}_1^{(0)} = (0.1, 0.4, 0.4, 0.1). \end{array} \right.
 \end{aligned} \tag{5.68}$$

Dans les tableaux 5.9, 5.10, 5.11 et 5.12 et les figures 5.6 et 5.7 nous donnons les valeurs estimées des paramètres pour les différentes valeurs initiales données ci-dessus pour le cas 1. Dans les tableaux 5.14, 5.15 et 5.16 nous présentons les résultats correspondants au cas 2. Également, dans les tableaux 5.13 et 5.17 nous donnons le temps moyen d'exécution (en secondes) par itération pour EM et SEM pour les deux cas.

Nous remarquons d'après les tableaux et les figures pour le cas 1 que l'estimation pour les quatorze premiers paramètres de q_{01} est bonne, alors que pour le dernier paramètre elle n'est pas satisfaisante. Également, l'estimation pour les neuf premiers paramètres de q_{10} est bonne, alors que pour le dernier paramètre elle n'est pas précise. Par ailleurs, les estimations des probabilités d'émission sont assez précises, et ceci indépendamment des valeurs initiales.

Ces résultats montrent également que l'algorithme SEM donne presque les mêmes estimations que l'algorithme EM, mais en utilisant SEM avec un échantillon de taille 500 à chaque itération, nous avons besoin de presque deux fois et demi moins de temps CPU que l'EM correspondant (voir le tableau 5.13).

Afin de voir l'effet de la taille de l'échantillon choisi pour SEM à chaque itération sur le temps CPU moyen, nous pouvons comparer entre le cas 1 de l'exemple 1 et le cas 1 de l'exemple 2 (ici présenté). Ces deux cas ont la même longueur de trajectoire mais pas la même taille de l'échantillon. Ainsi, d'après les tableaux 5.4 et 5.13, nous remarquons que le rapport entre le temps CPU de EM sur celui de SEM a été divisé par 4 quand la taille de l'échantillon a été multipliée par 5.

En ce qui concerne le deuxième cas de l'exemple 2, comme nous pouvons constater dans les tableaux 5.15 et 5.16, nous avons des bonnes estimations pour les dix premiers paramètres de q_{01} et pour tous les paramètres de q_{01} dans le cas de la valeur initiale 2_α . Par contre, après 200 itérations, l'estimation n'est pas bonne dans le cas 2_β , mais elle devient acceptable après 500 itérations. Puisque la convergence est lente, il est nécessaire

d'utiliser un algorithme SEM pour avoir un temps de calcul abordable. Par ailleurs, par le tableau 5.14 nous concluons que les estimations des probabilités d'émission sont assez précises, indépendamment des valeurs initiales et ceci même après 100 itérations.

		EM				SEM			
(2 _a)	R_0	0.3908	0.3080	0.2057	0.0955	0.3899	0.3086	0.2053	0.0962
	R_1	0.0737	0.1780	0.3072	0.4411	0.0762	0.1773	0.3076	0.4389
(2 _b)	R_0	0.3786	0.3012	0.2108	0.1093	0.3793	0.3010	0.2106	0.1091
	R_1	0.1104	0.1958	0.2936	0.4002	0.1101	0.1964	0.2936	0.3999

TAB. 5.9 – *Exemple2-cas1*. Estimations des vecteurs de probabilité R_0 et R_1 , en utilisant les algorithmes EM et SEM avec 200 itérations et en partant des valeurs initiales données dans (2_a) et (2_b).

		SEM			
(2 _c)	R_0	0.4030	0.3022	0.2105	0.1083
	R_1	0.1083	0.1938	0.2945	0.4034
(2 _d)	R_0	0.3763	0.3005	0.2107	0.1125
	R_1	0.1027	0.1925	0.2974	0.4074

TAB. 5.10 – *Exemple2-cas1*. Estimations des vecteurs de probabilité R_0 et R_1 , en utilisant l'algorithme SEM avec 200 itérations et en partant des valeurs initiales données dans (2_c) et (2_d).

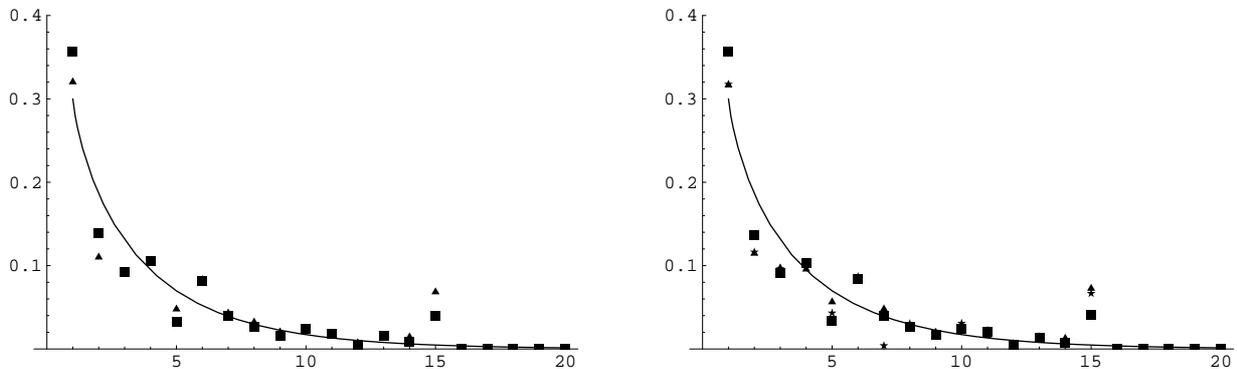


FIG. 5.6 – *Exemple2-cas1*. À gauche : Les vraies valeurs (ligne continue) et les valeurs estimées de $q_{01}(k)$, $k = 1, \dots, 15$, en utilisant l'algorithme EM. À droite : Les vraies valeurs (ligne continue) et les valeurs estimées en utilisant l'algorithme SEM. Pour les deux figures, les valeurs estimées obtenues en utilisant les valeurs initiales (2_a) sont représentées par des carrés, et celles obtenues en utilisant les valeurs initiales (2_b) par des triangles. Et, pour la figure de droite, les valeurs estimées obtenues avec les valeurs initiales (2_c) sont représentées par des étoiles.

CONCLUSION DU CHAPITRE

Dans ce chapitre, nous avons proposé une version améliorée de l'algorithme EM, ainsi qu'une version stochastique de cet algorithme, afin d'étu-

Vraie valeur		EM		SEM			
		(2 _a)	(2 _b)	(2 _a)	(2 _b)	(2 _c)	(2 _d)
1	0.3000	0.3565	0.3203	0.3570	0.3166	0.3175	0.3038
2	0.1860	0.1390	0.1101	0.1360	0.1147	0.1162	0.1089
3	0.1306	0.0919	0.0922	0.0916	0.0970	0.0962	0.0972
4	0.0946	0.1060	0.1032	0.1030	0.0957	0.1017	0.1034
5	0.0697	0.0324	0.0478	0.0534	0.0564	0.0431	0.0494
6	0.0520	0.0820	0.0830	0.0845	0.0816	0.0869	0.0866
7	0.0391	0.0398	0.0430	0.0398	0.0479	0.0400	0.0415
8	0.0296	0.0264	0.0322	0.0269	0.0247	0.0303	0.0313
9	0.0225	0.0163	0.0204	0.0165	0.0200	0.0165	0.0209
10	0.0172	0.0238	0.0235	0.0243	0.0262	0.0308	0.0284
11	0.0132	0.0185	0.0177	0.0200	0.0162	0.0224	0.0171
12	0.0101	0.0046	0.0074	0.0051	0.0065	0.0056	0.0098
13	0.0078	0.0155	0.0160	0.0138	0.0110	0.0147	0.0122
14	0.0061	0.0081	0.0144	0.0071	0.0127	0.0111	0.0153
15	0.0047	0.0391	0.0684	0.0403	0.0726	0.0667	0.0740

TAB. 5.11 – *Exemplez-cas1*. Estimations des $q_{01}(k)$, $k = 1, \dots, 15$, en utilisant les algorithmes EM et SEM avec 200 itérations et en partant des valeurs initiales (2_a) et (2_b) pour EM et SEM. De plus nous considérons pour l'algorithme SEM les valeurs initiales données par (2_c) et (2_d).

Vraie valeur		EM		SEM			
		(2 _a)	(2 _b)	(2 _a)	(2 _b)	(2 _c)	(2 _d)
1	0.5000	0.5103	0.4245	0.5073	0.4243	0.4318	0.4577
2	0.1757	0.1976	0.1813	0.1952	0.1785	0.1716	0.1664
3	0.1002	0.1130	0.0957	0.1115	0.0992	0.1080	0.0892
4	0.0636	0.0450	0.0654	0.0518	0.0666	0.0675	0.0704
5	0.0427	0.0466	0.0640	0.0441	0.0650	0.0635	0.0648
6	0.0298	0.0346	0.0597	0.0350	0.0630	0.0567	0.0559
7	0.0213	0.0100	0.0225	0.0116	0.0226	0.0195	0.0182
8	0.0156	0.0113	0.0249	0.0126	0.0225	0.0239	0.0206
9	0.0115	0.0123	0.0225	0.0129	0.0197	0.0223	0.0203
10	0.0087	0.0192	0.0038	0.0185	0.0386	0.0349	0.0365

TAB. 5.12 – *Exemplez-cas1*. Estimations des $q_{10}(k)$, $k = 1, \dots, 10$, en utilisant les algorithmes EM et SEM avec 200 itérations et en partant des valeurs initiales (2_a) et (2_b) pour EM et SEM, et de plus nous considérons pour l'algorithme SEM les valeurs initiales données par (2_c) et (2_d).

	EM	SEM
(2 _a)	35.87	15.27
(2 _b)	38.71	14.52

TAB. 5.13 – *Exemplez-cas1*. Le temps moyen d'exécution (en secondes) par itération pour EM et SEM calculé en 200 itérations respectivement. La taille d'échantillon dans l'algorithme SEM est 500.

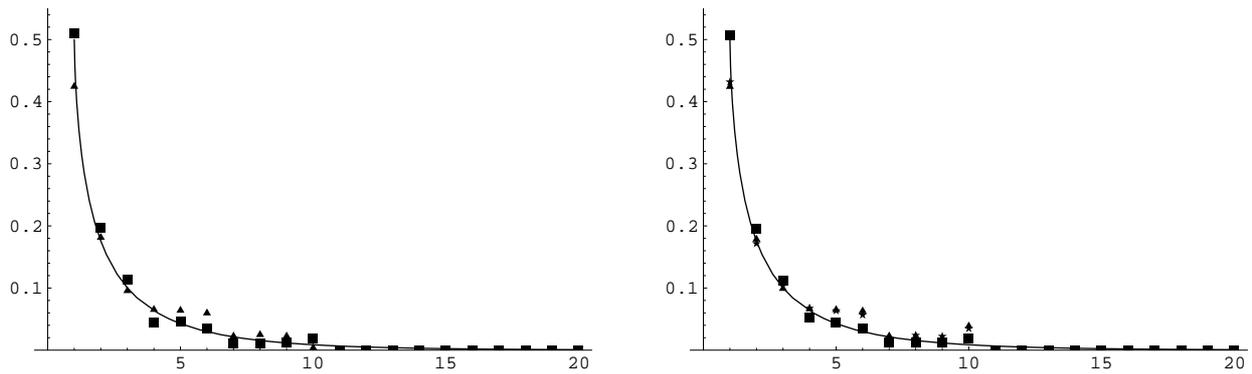


FIG. 5.7 – *Exemplez-cas1*. À gauche : Les vraies valeurs (ligne continue) et les valeurs estimées de $q_{10}(k)$, $k = 1, \dots, 10$ en utilisant l'algorithme EM. À droite : Les vraies valeurs (ligne continue) et les valeurs estimées en utilisant l'algorithme SEM. Pour les deux figures, les valeurs estimées obtenues en utilisant les valeurs initiales (2_a) sont représentées par des carrés, et celles obtenues en utilisant les valeurs initiales (2_b) par des triangles. Et, pour la figure de droite, les valeurs estimées obtenues avec les valeurs initiales (2_c) sont représentées par des étoiles.

		EM				SEM			
(2_a)	R_0	0.2374	0.2557	0.2565	0.2504	0.2381	0.2560	0.2561	0.2498
	R_1	0.0964	0.4126	0.4142	0.0768	0.1000	0.4066	0.4094	0.0840
(2_b)	R_0	0.2313	0.2620	0.2641	0.2436	0.2317	0.2606	0.2635	0.2442
	R_1	0.1210	0.3859	0.3855	0.1076	0.1223	0.3858	0.3843	0.1076

TAB. 5.14 – *Exemplez-cas2*. Estimations des vecteurs de probabilité R_0 et R_1 , en utilisant les algorithmes EM et SEM avec 100 itérations et en partant des valeurs initiales données dans (2_a) et (2_b) .

		EM		SEM		
	Vraie valeur	(2_a)	(2_b)	(2_a)	(2_b)	(2_b) -500
1	0.3000	0.2714	0.1513	0.2748	0.1567	0.2047
2	0.1860	0.1968	0.1353	0.2059	0.1430	0.1571
3	0.1306	0.1237	0.1041	0.1140	0.1014	0.1236
4	0.0946	0.0395	0.0769	0.0477	0.0770	0.0592
5	0.0697	0.0436	0.0688	0.0388	0.0628	0.0597
6	0.0520	0.0448	0.0643	0.0436	0.0726	0.0775
7	0.0391	0.0684	0.0552	0.0689	0.0523	0.0725
8	0.0296	0.0271	0.0294	0.0262	0.0248	0.0167
9	0.0225	0.0274	0.0353	0.0286	0.0348	0.0221
10	0.0172	0.0175	0.0300	0.0183	0.0275	0.0177
11	0.0132	0.0337	0.0418	0.0379	0.0449	0.0298
12	0.0101	0.0255	0.0386	0.0249	0.0311	0.0270
13	0.0078	0.0235	0.0403	0.0223	0.0422	0.0288
14	0.0061	0.0216	0.0501	0.0268	0.0544	0.0384
15	0.0047	0.0353	0.0821	0.0283	0.0744	0.0818

TAB. 5.15 – *Exemplez-cas2*. Estimations des $q_{01}(k)$, $k = 1, \dots, 15$, en utilisant les algorithmes EM et SEM avec 200 itérations et en partant des valeurs initiales données par (2_a) et (2_b) . La dernière colonne donne les valeurs estimées avec l'algorithme SEM en partant de la valeur initiale (2_b) pour 500 itérations.

	Vraie valeur	EM		SEM		
		(2_α)	(2_β)	(2_α)	(2_β)	$(2_\beta) - 500$
1	0.5000	0.4700	0.2062	0.4517	0.1888	0.2923
2	0.1757	0.1965	0.1677	0.2010	0.1701	0.1893
3	0.1002	0.1223	0.1296	0.1165	0.1599	0.1517
4	0.0636	0.0405	0.0977	0.0397	0.0802	0.0666
5	0.0427	0.0382	0.0837	0.0389	0.0832	0.0810
6	0.0298	0.0443	0.0776	0.0434	0.0695	0.0595
7	0.0213	0.0395	0.0631	0.0457	0.0639	0.0603
8	0.0156	0.0215	0.0447	0.0394	0.0511	0.0266
9	0.0115	0.0115	0.0443	0.0141	0.0481	0.0185
10	0.0087	0.0156	0.0853	0.0180	0.0852	0.0543

TAB. 5.16 – *Exemplez-cas2*. Estimations des $q_{10}(k)$, $k = 1, \dots, 10$, en utilisant les algorithmes EM et SEM avec 200 itérations et en partant des valeurs initiales données par (2_α) et (2_β) . La dernière colonne donne les valeurs estimées avec l’algorithme SEM en partant de la valeur initiale (2_β) pour 500 itérations.

	EM	SEM
(2_α)	423.08	24.19
(2_β)	451.45	26.11

TAB. 5.17 – *Exemplez-cas2*. Le temps moyen d’exécution (en secondes) par itération pour EM et SEM calculé en 200 itérations. La taille d’échantillon dans l’algorithme SEM est 500.

dier les modèles semi-markoviens cachés non paramétriques par la méthode du maximum de vraisemblance.

Nos algorithmes s’appliquent directement aux cas où le support des lois du temps de séjour conditionnelles pour les états de la CSM Z est fini ou quand l’erreur de l’approximation par un support fini approprié est suffisamment petite. Le support des lois dans les valeurs initiales doit être augmenté dans le cas où une estimation plus précise est exigée.

Plusieurs tests numériques ont été présentés dans le cas où il y a deux états cachés. Ils montrent des bonnes estimations avec un temps de calcul abordable. En particulier, les résultats ont pointé la supériorité de l’algorithme SEM notamment en terme de temps d’exécution.

Enfin, l’importance et l’utilité de l’approche présentée ici se trouvent notamment dans la compréhension de la forme des lois sous-jacentes du modèle, et moins dans la précision ponctuelle des probabilités à estimer. Ainsi, elle pourra être utilisée comme une approche initiale surtout quand il n’y a aucune information sur la structure des vraies lois du modèle.

CONCLUSION GÉNÉRALE ET PERSPECTIVES

CONCLUSION GÉNÉRALE

Au cours de ce mémoire, nous avons présenté une étude basée sur l'estimation du maximum de vraisemblance pour des modèles markoviens, semi-markoviens et semi-markoviens cachés.

Les résultats de ce travail se résument principalement en les points suivants :

i) Dans le deuxième chapitre nous avons donné l'estimateur du maximum de vraisemblance (EMV) de la loi stationnaire et de la variance asymptotique dans le théorème de la limite centrale pour des fonctionnelles additives d'une chaîne de Markov finie ergodique (irréductible). Nous avons démontré leurs propriétés asymptotiques de convergence et de normalité asymptotique. Les expressions obtenues dans les résultats asymptotiques pour ces deux caractéristiques ont été ensuite représentées par des formules explicites et facilement programmables pour des applications potentielles.

Notre point de départ se trouve dans les expressions de la loi stationnaire et de la variance dans le théorème de la limite centrale déjà existantes dans la littérature. Nous avons exprimé toutes ces expressions en fonction de la matrice de transition de la chaîne de Markov. Puis, en utilisant l'EMV pour la matrice de transition et ses propriétés asymptotiques, nous avons obtenu l'EMV pour les deux caractéristiques et démontré ses propriétés asymptotiques. Cela a été réalisé à l'aide du théorème de l'application continue et de la méthode delta en utilisant également des propriétés matricielles afin d'arriver aux formules explicites.

ii) Dans le troisième chapitre nous avons considéré l'estimation du maximum de vraisemblance pour le modèle semi-markovien fini non paramétrique. Nous avons construit l'EMV exact et démontré les propriétés asymptotiques de convergence et de normalité asymptotique pour chaque sous-vecteur fini de l'EMV. Cela a été fait pour le cas d'une trajectoire d'une chaîne semi-markovienne ergodique et également pour L trajectoires qui proviennent d'une suite des copies indépendantes d'une chaîne semi-markovienne censurée à un instant fixe, non nécessairement ergodique.

Dans le cas d'une trajectoire, notre point de départ a été le travail de Barbu et Limnios (2006a) qui ont considéré l'étude du modèle semi-markovien non paramétrique avec des estimateurs empiriques. Afin de trouver l'EMV exact, nous avons utilisé une reparamétrisation appropriée du modèle. En effet, nous avons utilisé l'espace paramétrique naturel qui correspond à une chaîne de Markov couple. Ce processus couple est formé par la chaîne semi-markovienne initiale et par la suite des temps de récurrence en arrière. Afin de démontrer les résultats asymptotiques de chaque sous-vecteur fini de l'estimateur du maximum de vraisemblance de la chaîne markovienne couple, nous avons construit une différence de martingales appropriée et appliqué le principe de Wald-Cramér. Enfin, nous avons transféré ces résultats asymptotiques au modèle semi-markovien fini non paramétrique.

Dans le cas de L trajectoires qui sont générées par des copies indépendantes d'une chaîne semi-markovienne, notre point de départ a été le travail de Anderson et Goodman (1957) qui ont considéré le même problème pour des chaînes de Markov finies. Nous avons utilisé la même reparamétrisation du noyau semi-markovien que dans le cas d'une trajectoire pour déduire les propriétés asymptotiques de la partie du noyau qui peut être estimée jusqu'à l'instant de censure.

iii) Dans le quatrième chapitre nous avons considéré une étude théorique pour le modèle semi-markovien caché. Nous avons introduit le modèle semi-markovien caché général avec dépendance des temps de récurrence en arrière. Nous avons démontré la convergence et la normalité asymptotique de l'EMV dans le cas où le support des lois des temps de séjour conditionnelles pour les états cachés est fini. Nous avons donné également des formules qui lient les matrices de covariance asymptotiques des différentes caractéristiques du modèle semi-markovien caché avec la matrice de covariance asymptotique du modèle markovien caché qui a servi pour reparamétriser le modèle initial.

Notre point de départ a été les résultats théoriques de convergence et de normalité asymptotique de l'EMV pour des modèles markoviens cachés généraux donnés par Leroux (1992) et Bickel et al. (1998) respectivement, et le premier travail de Barbu et Limnios (2006) sur les propriétés théoriques de l'EMV pour des modèles semi-markoviens cachés. Afin de démontrer les résultats asymptotiques, nous avons utilisé l'espace paramétrique naturel du processus couple décrit ci-dessus.

iv) Dans le cinquième chapitre nous avons considéré une étude algorithmique pour le modèle semi-markovien caché non paramétrique. Nous avons présenté une version améliorée de l'algorithme EM donné par Barbu et Limnios (2008), ainsi qu'une version stochastique de cet algorithme, afin d'étudier ces modèles par la méthode du maximum de vraisemblance. Des tests numériques ont été présentés notamment dans le cas où il y a deux états cachés.

Notre point de départ a été l'algorithme EM qui a été développé par Barbu et Limnios (2008). La comparaison entre la version améliorée

de l'algorithme EM et la version approximation stochastique (SAEM) montre la supériorité de la version stochastique notamment en terme de temps d'exécution. Enfin, l'importance et l'utilité de l'approche non-paramétrique, présentée ici, se trouvent notamment dans la compréhension de la forme des lois sous-jacentes du modèle, et moins dans la précision des probabilités à estimer. Ainsi, elle pourra être utilisée comme une approche initiale surtout quand il n'y a aucune information sur la structure des vraies lois du modèle.

PERSPECTIVES

Dans la continuité directe de notre travail, nous pouvons étendre les résultats obtenus dans plusieurs directions intéressantes.

i) Nous avons effectué une étude théorique pour la loi stationnaire et la variance asymptotique dans le théorème de la limite centrale pour chaînes de Markov. Nous avons démontré la convergence forte et la normalité asymptotique de l'EMV pour les deux caractéristiques. Notre but à court terme est de démontrer également son efficacité. Par suite, il serait intéressant de considérer le problème d'estimation de la variance asymptotique dans le théorème de la limite centrale pour des chaînes de renouvellement markoviennes, un problème dont le degré de difficulté est plus grand que celui pour les chaînes de Markov (voir Moore et Pyke (1968)). En effet, si nous considérons le cas non paramétrique du noyau semi-markovien, l'expression de la variance asymptotique implique un nombre infini des paramètres inconnus. Nous pourrions peut être exploiter la méthodologie du troisième chapitre pour s'attaquer à ce problème. Dans le cas où le problème est paramétrique, Stefanov (1995) a considéré une approche unifiée basée sur les familles exponentielles pour obtenir des solutions explicites à l'aide de l'estimateur du maximum de vraisemblance pour les processus de Markov. Il serait donc intéressant de considérer également des problèmes paramétriques en utilisant les résultats de Stefanov (1995) afin de donner des formules explicites pour l'EMV de la variance asymptotique pour les chaînes de Markov et de renouvellement, et pour la variance asymptotique de cet EMV.

ii) Nous avons considéré l'estimation de la partie du noyau semi-markovien qui peut être estimée jusqu'à l'instant de censure et ses propriétés asymptotiques pour le problème de L trajectoires d'une chaîne semi-markovienne censurée à un instant fixe m . Cela ne constitue qu'une première étape dans ce type d'estimation. Une possibilité est de considérer le cas où le temps de censure n'est pas fixe mais une variable aléatoire. Elle peut être une variable aléatoire indépendante de la chaîne semi-markovienne ou dépendante. Le cas indépendant peut correspondre, par exemple, à une intervention extérieure dans un système modélisé par une chaîne semi-markovienne. Le cas dépendant peut être, par exemple, le temps d'entrée d'une chaîne semi-markovienne dans un état absorbant

(ou marqué). Dans la fiabilité des systèmes, un état absorbant peut être un état de panne dans un système, réparable ou non.

Notons qu'ici nous avons fait l'hypothèse que le temps d'observation initial correspond à un instant de saut ($S_0 = 0$). Il y a des applications où une hypothèse de tel type n'est pas réaliste. En fiabilité, la prise en compte de la variable aléatoire S_0 , où plus particulièrement $S_0 \neq 0$, peut correspondre aux situations où le temps initial d'observation n'est pas identique au début du fonctionnement du système. Dans le contexte médical, S_0 peut représenter le délai de la détection d'un certain type de maladie pour un certain patient. Une possibilité est de considérer que l'évolution entre les différentes phases de cette maladie peut être modélisée par une chaîne semi-markovienne. L'observation de l'évolution de la maladie auprès d'un patient correspond à une trajectoire d'une chaîne semi-markovienne censurée à gauche. Par conséquent, le problème d'estimation consiste à faire une prévision pour l'évolution de la maladie en observant plusieurs patients, ce qui entre dans le cadre d'estimation avec des trajectoires censurées à gauche qui proviennent d'une suite des copies indépendantes d'une chaîne semi-markovienne.

iii) La complexité des modèles semi-markoviens cachés et leurs applications potentielles nous donne la motivation d'investir davantage dans les deux aspects, théorique et algorithmique, de l'étude de ces modèles. Un travail qui constitue une continuation naturelle des résultats présentés dans le quatrième et le cinquième chapitre, est en cours. En fait, nous avons développé un nouvel algorithme EM et une version stochastique SAEM spécialement destinés au modèle semi-markovien caché avec dépendance des temps de récurrence en arrière. Dans la formulation de ces algorithmes nous avons pris en compte l'EMV exact en utilisant la reparamétrisation du noyau semi-markovien par les probabilités de transition de la chaîne de Markov couple que nous avons décrit dans ce mémoire. Nous sommes en train de comparer les performances des algorithmes en utilisant ces différentes paramétrisations du modèle semi-markovien caché avec ou sans dépendance des temps de récurrence en arrière. Ce travail peut avoir des applications intéressantes. La prise en compte de la dépendance des temps de récurrence en arrière peut être utile dans des situations où la loi d'émission des états cachés change avec le temps à cause des restrictions naturelles imposées par le type de l'application considérée, ou du vieillissement des composants du système modélisé par la chaîne semi-markovienne cachée, ou à cause des interventions imposées par l'expérimentateur.

Dans le même contexte, l'estimation des constantes appropriées de façon à ce que l'erreur d'approximation du vrai support des lois des temps de séjour conditionnelles par un support borné soit négligeable est un sujet important. Nous avons vu que dans le cas de plusieurs copies indépendantes d'une chaîne semi-markovienne censurée à un instant fixe m , nous pouvons mesurer la concentration de la loi du temps de séjour pour tous les états de la chaîne semi-markovienne dans l'ensemble fini $\{1, 2, \dots, m\}$. Cela est fait par l'EMV de la fonction de répartition du temps de séjour dans chaque état à l'instant m . Nous avons l'intention

d'étendre les résultats obtenus pour les chaînes semi-markoviennes aux chaînes semi-markoviennes cachées. Cela peut être réalisé en considérant plusieurs copies indépendantes d'une chaîne semi-markovienne cachée jusqu'à un instant fixe. En considérant différentes valeurs de m , nous pourrions estimer des constantes appropriées et les utiliser par la suite pour étudier des trajectoires assez longues afin d'avoir des bonnes estimations avec un temps de calcul abordable.

Il est dans notre intérêt également d'investir davantage dans le sujet des critères d'arrêt des algorithmes, ainsi que de choisir la suite des poids γ_m , utilisée pour l'algorithme SAEM, d'une manière optimale pour accélérer la convergence.

En outre, l'approche non paramétrique, étant une approche initiale, doit être combinée avec des modèles paramétriques pour des applications qui nécessitent une structure cachée spécifique afin d'avoir plus de précision dans l'estimation. Une application particulière qui peut entrer dans ce cadre est la détection automatique des îlots dans une séquence d'ADN, ce qui sert le diagnostic automatique en médecine.

Finalement, au niveau théorique, nous avons supposé que le support des lois des temps de séjour conditionnelles est fini. Cela pose une limitation importante. Jusqu'à présent, à notre connaissance, aucun résultat asymptotique pour l'EMV du modèle semi-markovien caché avec un support infini n'a été démontré, ni dans le cas paramétrique ni dans le cas non paramétrique. Cela constitue un de nos futurs objectifs.

ANNEXES



SOMMAIRE

A.1	THÉORÈME LIMITE CENTRALE POUR LES MARTINGALES	136
A.2	PREUVE DU LEMME 2.3	136
A.3	COMPORTEMENT INITIAL	137
A.4	CONDITIONS LEROUX : CONVERGENCE	139
A.5	CONDITIONS BICKEL-RITOV-RYDÉN : NORMALITÉ ASYMPTOTIQUE . .	140

A.1 THÉORÈME LIMITE CENTRALE POUR LES MARTINGALES

Théorème A.1 *Soit X une martingale définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ adaptée à la filtration naturelle \mathcal{F} et $Y := \Delta X$ la différence de martingales associée. Si les deux conditions suivantes sont satisfaites,*

$$(M1) \quad m^{-1} \sum_{n=1}^m \mathbb{E} (Y_n^2 \mid \mathcal{F}_{n-1}) \xrightarrow[m \rightarrow \infty]{\mathbb{P}} \sigma^2 > 0,$$

$$(M2) \quad m^{-1} \sum_{n=1}^m \mathbb{E} \left(Y_n^2 \mathbf{1}_{\{|Y_n| > \epsilon \sqrt{m}\}} \right) \xrightarrow[m \rightarrow \infty]{} 0, \quad \text{pour tout } \epsilon > 0,$$

alors,

$$m^{-1} X_m \xrightarrow{\mathbb{P}\text{-p.s.}} 0,$$

et

$$m^{-1/2} X_m = \sum_{n=1}^m Y_n \rightsquigarrow \mathcal{N}(0, \sigma^2).$$

A.2 PREUVE DU LEMME 2.3

Nous remarquons par (2.37) qu'un vecteur stationnaire vérifie

$$\pi P = \pi \quad \text{ssi} \quad \pi(I - P + \mathbb{I}) = \mathbf{1}. \quad (\text{A.1})$$

En effet, si la matrice $I - P + \mathbb{I}$ est non-singulière, alors les équations d'équilibre admettent une solution unique et par conséquent la CM X a une seule classe récurrente positive.

Maintenant, supposons que X a une seule classe récurrente positive et montrons que l'équation

$$(I - P + \mathbb{I})\mathbf{x}^\top = \mathbf{0}, \quad (\text{A.2})$$

admet $\mathbf{x} = \mathbf{0}$ comme solution unique.

Puisque X a une seule classe récurrente positive, alors il existe un r , $1 \leq r \leq s$, tel que $E = R \cup T$, avec $R = \{i_1, \dots, i_r\}$ la classe récurrente et $T = \{i_{r+1}, \dots, i_s\}$ l'ensemble des états transitoires. Si $r = s$, alors $T = \emptyset$ et la chaîne X est irréductible. Nous effectuons une permutation des lignes de la matrice P de sorte que les s premières lignes correspondent à la classe récurrente R .

Soit π un vecteur stationnaire. Par (A.1) et (A.2) nous avons

$$\pi(I - P + \mathbb{I})\mathbf{x}^\top = \mathbf{1}\mathbf{x}^\top = \mathbf{0}. \quad (\text{A.3})$$

Cela implique

$$\mathbb{I}\mathbf{x}^\top = (\mathbf{1}\mathbf{x}^\top, \mathbf{1}\mathbf{x}^\top, \dots, \mathbf{1}\mathbf{x}^\top) = \mathbf{0}. \quad (\text{A.4})$$

Par (A.2) et (A.4) nous obtenons

$$P\mathbf{x}^\top = \mathbf{x}^\top, \quad (\text{A.5})$$

et par conséquent pour tout $m \in \mathbb{N}$

$$m^{-1} \sum_{n=1}^m P^n \mathbf{x}^\top = \mathbf{x}^\top. \quad (\text{A.6})$$

Pour la chaîne de Markov \mathbf{X} avec $E = R \cup T$, décrits ci-dessus, la limite de Cesàro de la suite $\{P^n\}_{n \geq 1}$ converge vers la matrice

$$\Pi = \begin{pmatrix} \Pi_r & \mathbf{O}_{r \times (s-r)} \\ \Pi_{(s-r) \times r} & \mathbf{O}_{s-r} \end{pmatrix}, \quad (\text{A.7})$$

et en prenant la limite pour $m \rightarrow \infty$, (A.6) donne le système d'équations suivant :

$$\Pi \mathbf{x}^\top = \mathbf{x}^\top. \quad (\text{A.8})$$

Cela revient au système d'équations pour tout $i \in E$:

$$\sum_{j \in R} \pi_j x_j = x_i. \quad (\text{A.9})$$

Puisque, par (A.3), $\sum_{i \in E} x_i = 0$ et par (A.9), les x_i sont égaux, nous concluons alors que $\mathbf{x} = \mathbf{0}$.

A.3 COMPORTEMENT INITIAL

Nous définissons pour tout $i \in E$,

$$M_i := \inf\{m \in \mathbb{N} : N_i(m) > 0\}. \quad (\text{A.10})$$

Cette v.a. correspond au premier instant où le système reconnaît l'apparition de l'état $i \in E$ via le processus de comptage N_i . Ainsi, par (2.16), nous avons

$$M_i - 1 = \inf\{m \in \mathbb{N} : X_m = i, X_k \neq i, \text{ pour } 0 \leq k \leq m - 1\}, \quad (\text{A.11})$$

Nous notons également $\{M_{(i)}\}_{i \in E}$ la suite ordonnée des $\{M_i\}_{i \in E}$ et considérons la décomposition suivante

$$(M_{(1)}, M_{(2)}, \dots, M_{(s)}) = (M_{i_1}, M_{i_2}, \dots, M_{i_s}). \quad (\text{A.12})$$

Nous considérons pour tout $1 \leq r \leq s$, l'ensemble

$$E_{(r)} := \{i_1, i_2, \dots, i_r\}, \quad (\text{A.13})$$

qui représente les états visités successivement. Nous soulignons les cas particuliers suivants

$$\begin{aligned} M_{(1)} &= 1, \\ M_{(s)} &= \inf\{m \in \mathbb{N} : N_i(m) > 0, \forall i \in E\}. \end{aligned} \quad (\text{A.14})$$

Nous considérons également la suite $(E_m)_{m \in \mathbb{N}}$, où $E_m \subseteq E$ est donné par

$$E_m := \{i \in E : N_i(m) > 0\}. \quad (\text{A.15})$$

Si nous posons

$$r_m := \text{card}\{i \in E : M_i \leq m\} = \text{card}\{i \in E : N_i(m) > 0\}, \quad (\text{A.16})$$

alors par (A.15), (A.16) et (A.13) nous concluons que

$$E_m = E_{(r_m)} = \{i_1, i_2, \dots, i_{r_m}\}. \quad (\text{A.17})$$

Soit $1 \leq r_m \leq s$ et $m \neq M_{(r_m+1)} - 1$ ($E_m = E_{m+1}$). Nous effectuons une permutation des lignes de la matrice $\hat{P}(m)$ en déplaçant les r_m lignes correspondantes aux indices $(i_1, i_2, \dots, i_{r_m})$ vers les premières lignes. La matrice $\hat{P}(m)$ peut être ainsi décomposée dans la forme suivante :

$$\hat{P}(m) = \begin{pmatrix} \hat{P}_{(r_m)}(m) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad (\text{A.18})$$

où $\hat{P}_{(r_m)}(m)$ représente une sous-matrice stochastique qui correspond à la matrice de transition pour une chaîne de Markov à espace d'état $E_{(r_m)}$, c'est à dire, une CM réduite aux états enregistrés par les processus de comptages N_i jusqu'à l'instant m .

Pour tous les instants $m \in \mathbb{N}$ tels que E_m est strictement inclus dans E_{m+1} , i.e., $X_m \in E_m^c$ où E_m^c désigne le complémentaire de E_m , la matrice $\hat{P}(m)$ ne peut pas être décomposée dans la forme (A.18). En effet, même si le système a reconnu une transition d'un état $i \in E_m$ vers un état $j \in E_m^c$, cet état n'a pas encore donné une transition. Cela est reflété également dans le calcul du vecteur stationnaire $\hat{\pi}(m)$. En effet, les m tels que $X_m \in E_m^c$ correspondent aux instants pour lesquels le calcul n'a pas de sens.

Soit $(p_1, p_2, \dots, p_{r_m})$ le vecteur des paramètres qui correspondent aux états enregistrés par les processus de comptages N_i jusqu'à l'instant m . La fonction de vraisemblance va être considérée comme une fonction de ce vecteur et l'EMV qui est associé à l'observation \mathbf{Z}_0^m va maximiser ainsi la fonction de vraisemblance par rapport à ce vecteur. Les autres paramètres sont indépendants de ce vecteur et donc leur contribution ne joue aucun rôle pour la fonction de vraisemblance.

Nous posons $\hat{p}_{ij}(m) = 0$ si $i \in \{i_1, i_2, \dots, i_{r_m}\}^c$ afin d'approcher la CM à un espace d'état E par des CM à un espace d'état graduellement augmenté selon la suite $(E_m)_{m \in \mathbb{N}}$.

Pour tout $m \in \mathbb{N}$, l'existence et l'unicité de l'EMV sont dues au fait que nous avons gardé uniquement les paramètres $(p_1, p_2, \dots, p_{r_m})$. Cela nous permet d'étendre la méthode du maximum de vraisemblance dans le cas où la totalité des observations n'a pas encore été observée. Pour les instants « problématiques », où $m = M_{(i)} - 1$, $2 \leq i \leq s$, nous pouvons faire la convention de garder l'estimation de l'instant précédent. La variable aléatoire $M_{(s)}$ exprime le premier instant où la matrice $\hat{P}(m)$ va reconnaître \mathbf{X} comme une CM à espace d'état E .

Concernant le calcul de l'EMV $\hat{\pi}(m)$ du vecteur stationnaire π , donné dans (2.39), nous vérifions d'abord que la matrice $I - \hat{P}(m) + \mathbb{I}$ est non-singulière (sauf pour les instants d'augmentation d'espace d'état). Ensuite,

nous vérifions que ce calcul permet bien d'associer le vecteur stationnaire $\hat{\pi}_{r_m}(m)$ à la CM à espace d'état $E_{m,,}$, ainsi que le vecteur zéro aux états non visités.

Nous remarquons que la matrice $I - \hat{P}(m) + \mathbb{I}$ s'écrit

$$I - \hat{P}(m) + \mathbb{I} = \begin{pmatrix} I_{r_m} - \hat{P}_{(r_m)}(m) + \mathbb{I}_{r_m} & \mathbb{I}_{r_m \times (s-r_m)} \\ \mathbb{I}_{(s-r_m) \times r_m} & I_{s-r_m} + \mathbb{I}_{s-r_m} \end{pmatrix}, \quad (\text{A.19})$$

Nous notons

$$\hat{\pi}(m) = (\hat{\pi}_{r_m}(m), \mathbf{0}_{s-r_m}). \quad (\text{A.20})$$

Il est clair que par (A.19), nous avons

$$\hat{\pi}(m)(I - \hat{P}(m) + \mathbb{I}) = \mathbf{1}, \quad (\text{A.21})$$

et par conséquent si $I - \hat{P}(m) + \mathbb{I}$ est non-singulière alors la représentation (2.39) est identique à (A.20). Les étapes de la preuve sont similaires à celles du lemme 2.3 en prenant en compte la décomposition (A.20) et (A.18). Une deuxième possibilité est de démontrer par des méthodes élémentaires de l'algèbre linéaire que l'inversibilité de $I - \hat{P}(m) + \mathbb{I}$ peut être déduite de sa forme (A.19) et du fait que les matrices $I_{r_m} - \hat{P}_{(r_m)}(m) + \mathbb{I}_{r_m}$ et $I_{s-r_m} + \mathbb{I}_{s-r_m}$ sont inversibles.

A.4 CONDITIONS LEROUX : CONVERGENCE

Nous citons les conditions données dans Leroux (1992) pour assurer la convergence de l'EMV (théorème 3 dans Leroux (1992)). Ici, $P = (p_{ij})$ est la matrice de transition pour la chaîne de Markov cachée. L'espace paramétrique des caractéristiques du modèle est noté par Φ et celui de la famille des densités conditionnelles $\{g(\cdot | \theta)\}$ par Θ . La matrice de transition sous la vraie valeur est notée par P_{ϕ_0} .

- Cond. 1.** La matrice stochastique P_{ϕ_0} est irréductible.
- Cond. 2.** La famille formée par les mélanges d'au plus s éléments de $\{g(\cdot | \theta) : \theta \in \Theta\}$ est identifiable.
- Cond. 3.** Pour tout y , $g(y | \cdot)$ est continue et admet une limite nulle à l'infini.
- Cond. 4.** Pour tout j, k , $p_{jk}(\cdot)$ et $\theta_j(\cdot)$ sont continues.
- Cond. 5.** $\mathbb{E}_{\phi_0} [|\log g(Y_1 | \theta_j(\phi_0))|] < \infty$, pour $j = 1, \dots, s$.
- Cond. 6.** Pour tout $\theta \in \Theta$, $\mathbb{E}_0 \left[\sup_{\|\theta' - \theta\| < \delta} \left(\log g(Y_1 | \theta') \right)^+ \right] < \infty$, pour un certain $\delta > 0$, (où $\|\cdot\|$ est la norme euclidienne et $x^+ = \max(x, 0)$).

A.5 CONDITIONS BICKEL-RITOV-RYDÉN : NORMALITÉ ASYMPTOTIQUE

Nous citons les conditions données dans Bickel et al. (1998) pour assurer la normalité asymptotique de l'EMV. Ici, l'espace paramétrique des caractéristiques est noté par Θ et $p_0(i, j)$ représentent les probabilités de transition sous la vraie valeur θ_0 .

(A1) La matrice de transition $P_0 := (p_0(i, j))$ est irréductible et apériodique.

(A2) Pour tout i et j , les applications $\theta \mapsto p_\theta(i, j)$ et $\theta \mapsto \pi_\theta(i)$ sont deux fois continûment dérivables dans un certain voisinage de θ_0 . Pour tout i et $y \in \mathcal{Y}$, l'application $\theta \mapsto g_\theta(y | i)$ est deux fois continûment différentiable dans le même voisinage.

(A3) Écrivons $\theta = (\theta_1, \dots, \theta_d)$. Il existe un $\delta > 0$ tel que

i) pour tout $1 \leq k \leq d$ et pour tout i ,

$$\mathbb{E}_0 \left[\sup_{\|\theta - \theta_0\| < \delta} \left(\frac{\partial}{\partial \theta_k} \log g_\theta(Y_1 | i) \right)^2 \right] < \infty;$$

ii) pour tout $1 \leq k, l \leq d$ et pour tout i ,

$$\mathbb{E}_0 \left[\sup_{\|\theta - \theta_0\| < \delta} \left| \frac{\partial^2}{\partial \theta_k \partial \theta_l} \log g_\theta(Y_1 | i) \right| \right] < \infty;$$

iii) pour $l = 1, 2$, tout $1 \leq k_n \leq d$, $n = 1, \dots, l$ et pour tout i ,

$$\int \sup_{\|\theta - \theta_0\| < \delta} \left| \frac{\partial^l}{\partial \theta_{k_1} \dots \partial \theta_{k_l}} g_\theta(y | i) \right| \nu(dy) < \infty.$$

(A4) Il existe un $\delta > 0$ tel que, si

$$\rho_0(y) := \sup_{\|\theta - \theta_0\| < \delta} \max_{1 \leq i, j \leq K} \frac{g_\theta(y | i)}{g_\theta(y | j)}, \quad (\text{A.22})$$

alors, $\mathbb{P}_0(\rho_0(Y_1) = \infty | X_1 = i) < 1$, pour tout i .

(A5) θ_0 est un point intérieur de Θ .

(A6) L'estimateur du maximum de vraisemblance est fortement convergent.

BIBLIOGRAPHIE

- J. Aldrich. R. A. Fisher and the making of Maximum Likelihood 1912-1922. *Statistical Science*, 12(3) :162–176, 1997. (Cité page 5.)
- T.W. Anderson et L.A. Goodman. Statistical inference about Markov Chains. *Ann. Math. Stat.*, 28 :89–110, 1957. (Cité pages 6, 71, 72 et 130.)
- P.M. Anselone. Ergodic theory for discrete semi-Markov chains. *Duke Math. J.*, 27(1) :33–40, 1960. (Cité pages 8 et 43.)
- V. Barbu. *Estimation de Chaînes Semi-Markoviennes et des Chaînes Semi-Markoviennes Cachées en Vue d'Applications en Fiabilité et en Biologie*. Thèse, Laboratoire de Mathématiques Appliquées, UTC, Compiègne, 2005. (Cité page 43.)
- V. Barbu, M. Boussemart, et N. Limnios. Discrete time semi-Markov model for reliability and survival analysis. *Communications in Statistics-Theory and Methods*, 33(11) :2833–2868, 2004. (Cité page 43.)
- V. Barbu et N. Limnios. Maximum likelihood estimation for hidden semi-Markov models. *C.R. Acad. Sci. Paris*, 342 :201–205, 2006. (Cité pages 10, 100 et 130.)
- V. Barbu et N. Limnios. Empirical estimation for discrete time semi-markov processes with applications in reliability. *Journal of Nonparametric Statistics*, 18(4) :483–498, 2006a. (Cité pages 7, 8, 43, 44, 55 et 130.)
- V. Barbu et N. Limnios. *Semi-Markov chains and hidden semi-Markov models toward applications : Their use in reliability and DNA analysis (Lecture notes in statistics)*, volume 191. Springer, 2008. (Cité pages 10, 43, 46, 68, 72, 73, 78, 79, 100, 102 et 130.)
- I.V. Basawa et B.L.S. Prakasa Rao. *Statistical inference for stochastic processes*. Academic Press, London, 1980. (Cité page 6.)
- L.E. Baum et T. Petrie. Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Stat.*, 37 :1554–1563, 1966. (Cité pages 9, 77 et 90.)
- R. Bhar et S. Hamori. *Hidden Markov Models : Applications to Financial Economics*. Springer, New York, 2000. (Cité pages 9 et 77.)
- P.J. Bickel, Y. Ritov, et T. Ryden. Asymptotic normality of the maximum-likelihood estimator for general hidden-Markov models. *Ann. Stat.*, 26 (4) :1614–1635, 1998. (Cité pages 9, 78, 90, 91, 130 et 140.)

- P. Billingsley. *Statistical Inference for Markov Processes*. The University of Chicago Press, 1961a. (Cité pages 6, 15 et 23.)
- P. Billingsley. Statistical methods in Markov chains. *Ann. Math. Stat.*, 32 : 12–40, 1961b. (Cité pages 6, 15, 16, 18 et 31.)
- J. Bulla. *Application of Hidden Markov Models and Hidden Semi-Markov Models to Financial Time Series*. Thesis, Georg-August-Universität Göttingen, 2006. (Cité page 10.)
- J. Bulla et I. Bulla. Stylized facts of financial time series and hidden semi-Markov models. *Computational Statistics & Data Analysis*, 51(4) :2192–2209, 2006. (Cité page 100.)
- O. Cappé, E. Moulines, et T. Rydén. *Inference in Hidden Markov Models*. Springer, New York, 2005. (Cité page 9.)
- G. Celeux et J. Diebolt. The SEM algorithm : a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Comp. Statis. Quarterly*, 2 :73–82, 1985. (Cité pages 11 et 111.)
- D. Chauveau et J. Diebolt. Estimation of the asymptotic variance in the CLT for Markov Chains. *Stochastic Models*, 19(4) :449–465, 2003. (Cité pages 6 et 15.)
- O. Chryssaphinou, M. Karaliopoulou, et N. Limnios. On discrete time semi-Markov chains and applications in words occurrences. *Communications in Statistics-Theory and Methods*, 37 :1306–1322, 2008. (Cité pages 8, 44, 49 et 50.)
- E. Çinlar. Markov renewal theory. *Adv. in Appl. Probab.*, 1 :123–187, 1969. (Cité page 43.)
- E. Çinlar. *Introduction to Stochastic Processes*. Prentice Hall, New York, 1975. (Cité page 43.)
- H. Cramér. *Mathematical Methods of Statistics*. Princeton University Press, Princeton, N.J., 1949. (Cité page 5.)
- D. Dacunha-Castelle et M. Duflo. *Probabilités et statistiques 2*. Masson, 1983. (Cité page 16.)
- B. Delyon, V. Lavielle, et E. Moulines. Convergence of a stochastic approximation version of the EM algorithm. *Ann. Stat.*, 27 :94–128, 1999. (Cité pages 11 et 112.)
- A.P. Dempster, N.M. Laird, et D.B. Rubin. Maximum Likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 39 :1–38, 1977. (Cité page 103.)
- J.I. Doob. *Stochastic Processes*. Wiley, New York, 1953. (Cité pages 16, 31 et 33.)
- R. Douc et C. Matias. Asymptotics of the maximum likelihood estimator for general hidden Markov models. *Bernoulli*, 7 :381–420, 2001. (Cité page 9.)

- R. Douc, É. Moulines, et T. Rydén. Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *Ann. Stat.*, 32 :2254–2304, 2004. (Cité page 9.)
- R. Durbin, S. Eddy, A. Krogh, et G. Mitchison. *Biological Sequence Analysis : Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University press, Cambridge, 1998. (Cité page 99.)
- Y. Ephraim et N. Merhav. Hidden Markov Processes. *IEEE Trans. Inf. Theory*, 48(6) :1518–1569, 2002. (Cité pages 9 et 77.)
- J.D. Ferguson. Variable duration models for speech. Dans *Proc. of the symposium on the Application of Hidden Markov Models to Text and Speech*, pages 143–179, 1980. (Cité pages 10, 78 et 99.)
- R. A. Fisher. On an absolute criterion for fitting frequency curves. *Messenger of Mathematics*, 41 :155–160, 1912. (Cité page 5.)
- R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London*, 222A :309–368, 1922. (Cité page 5.)
- R. A. Fisher. Theory of statistical estimation. *Proc. Camb. Phil. Soc.*, 22 : 700–725, 1925. (Cité page 5.)
- R. A. Fisher. The logic of inductive inference. *Journal of the Royal Statistical Society*, 98 :39–82, 1935. (Cité page 5.)
- R. Fletcher. *Practical Methods of Optimization*. Wiley, New York, 1980. (Cité page 105.)
- C. D. Fuh. Efficient likelihood estimation in state space models. *Ann. Stat.*, 34(4) :2026–2068, 2006. (Cité page 9.)
- I.I. Gerontidis. Semi-Markov replacement chains. *Adv. in Appl. Probab.*, 26 : 728–755, 1994. (Cité page 43.)
- R.D. Gill. Nonparametric estimation based on censored observations of Markov renewal processes. *Z. Wahrsch. verw. Gebiete.*, 53 :97–116, 1980. (Cité pages 7 et 43.)
- Y. Guédon. Computational methods for discrete hidden semi-Markov chains. *Applied Stochastic Models in Business and Industry*, 15 :195–224, 1999. (Cité page 100.)
- Y. Guédon. Estimating hidden semi-Markov chains from discrete sequences. *Journal of Computational and Graphical Statistics*, 12(3) :604–639, 2003. (Cité page 100.)
- Y. Guédon. Hidden hybrid Markov/semi-Markov chains. *Computational Statistics & Data Analysis*, 49(3) :663–688, 2005. (Cité page 100.)
- Y. Guédon. Exploring the state sequence space for hidden Markov and semi-Markov chains. *Computational Statistics & Data Analysis*, 51(5) : 2379–2409, 2007. (Cité pages 100 et 114.)

- Y. Guédon et C. Coccozza-Thivent. Explicit state occupancy modelling by hidden semi-Markov models : application of Derin's scheme. *Computer Speech and Language*, 4 :167–192, 1990. (Cité page 100.)
- R. Howard. *Dynamic Probabilistic Systems*, volume 2. Wiley, New York, 1971. (Cité page 43.)
- W. Jank. Implementing and Diagnosing the Stochastic Approximation EM algorithm. *Journal of Computational and Graphical Statistics*, 15(4) :803–829, 2006. (Cité page 112.)
- J.L. Jensen et N.V. Petersen. Asymptotic normality of the maximum likelihood estimator in state space models. *Ann. Stat.*, 27(2). (Cité page 9.)
- S. Karlin et H.M. Taylor. *A First Course in Stochastic Processes*. Academic Press, Inc., 1975. (Cité page 41.)
- U. Küchler et M. Sørensen. *Exponential Families of Stochastic Processes*. Springer-Verlag, New York, 1997. (Cité page 16.)
- J.G. Kemeny et J.L. Snell. *Finite Markov Chains*. Springer-Verlag, 1976. (Cité pages 33 et 35.)
- J. Kiefer et J. Wolfowitz. Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters. *Ann. Math. Stat.*, 27 :887–906, 1956. (Cité page 87.)
- A. Krogh, M. Brown, I. S. Mian, K. Sjölander, et D. Haussler. Hidden Markov models in computational biology : Applications to protein modeling. *J. Molecular Biology*, 235 :1501–1531, 1994a. (Cité pages 9 et 77.)
- A. Krogh, I.S. Mian, et D. Haussler. A Hidden Markov model that finds genes in E. coli DNA. *Nucleic Acids Research*, 22 :4768–4778, 1994b. (Cité pages 9 et 77.)
- S.W. Lagakos, C.J. Sommer, et M. Zelen. Semi-Markov models for partially censored data. *Biometrika*, 65(2) :311–317, 1978. (Cité pages 7 et 43.)
- B.G. Leroux. Maximum-likelihood estimation for hidden Markov models. *Stochastic Proc. Appl.*, 40 :127–143, 1992. (Cité pages 9, 78, 87, 88, 90, 130 et 139.)
- S.E. Levinson. Continuously variable duration hidden Markov models for automatic speech recognition. *Computer Speech and Language*, 1 :29–45, 1986. (Cité pages 9, 78 et 99.)
- P. Lévy. Processus semi-markoviens. Dans *Proc. of International Congress of Mathematics, Amsterdam*, 1954. (Cité page 7.)
- J. Li et R.M. Gray. *Image Segmentation and Compression using Hidden Markov models*. Springer, New York, 2000. (Cité pages 9 et 77.)
- N. Limnios et G. Oprüşan. *Semi-Markov Processes and Reliability*. Birkhäuser, Boston, 2001. (Cité pages 43, 46 et 49.)

- N. Limnios, B. Ouhbi, et A. Sadek. Empirical estimator of stationary distribution for semi-Markov processes. *Communications on Statistics - Theory and Methods*, 34(4), 2005. (Cité pages 7 et 43.)
- S. Malefaki, S. Trevezas, et N. Limnios. An improved EM and a stochastic EM algorithm for nonparametric hidden semi-Markov models. *soumis*, 2008. (Cité page 100.)
- G.J. McLachlan et D. Peel. *Finite Mixture Models*. Wiley, New York, 2000. (Cité page 105.)
- S.P. Meyn et R.L. Tweedy. *Markov Chains and Stochastic Stability*. Springer-Verlag, London, 1993. (Cité page 16.)
- C.D. Mitchell et L.H. Jamieson. Modeling duration in a hidden Markov model with the exponential family. Dans *Proc. of the 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages II.331–II.334. Minneapolis, 1993. (Cité page 100.)
- C.J. Mode et G.T. Pickens. Computational methods for renewal theory and semi-Markov processes with illustrative examples. *The American Statistician*, 42(2) :143–152, 1998. (Cité page 43.)
- C.J. Mode et C.K. Sleeman. *Stochastic processes in epidemiology*. World Scientific, New Jersey, 2000. (Cité page 43.)
- E.H. Moore et R. Pyke. Estimation of the transition distributions of a Markov renewal process. *Ann. Inst. Statist. Math.*, 20 :411–424, 1968. (Cité pages 7, 43 et 131.)
- B. Ouhbi et N. Limnios. Nonparametric estimation for semi-Markov processes based on its hazard rate functions. *Statistical Inference for Stochastic Processes*, 2(2) :151–173, 1999. (Cité pages 7 et 43.)
- B. Ouhbi et N. Limnios. Nonparametric reliability estimation for semi-Markov processes. *Journal of Statistical Planning and Inference*, 109(1-2) : 155–165, 2003. (Cité pages 7 et 43.)
- T. Petrie. Probabilistic functions of finite state Markov chains. *Ann. Math. Stat.*, 40 :97–115, 1969. (Cité pages 9 et 78.)
- B.T. Polyak et A.B. Juditsky. Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization*, 30 :838, 1992. (Cité page 112.)
- S.C. Port. *Theoretical probability for applications*. Wiley, New York, 1994. (Cité pages 16, 17, 18 et 31.)
- R. Pyke. Markov renewal processes : definitions and preliminary properties. *Ann. Math. Stat.*, 32 :1231–1241, 1961a. (Cité page 43.)
- R. Pyke. Markov renewal processes with finitely many states. *Ann. Math. Stat.*, 32 :1243–1259, 1961b. (Cité page 43.)

- L.R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, 77 :257–284, 1989. (Cité pages 9 et 77.)
- L.R. Rabiner et B.H. Juang. *Fundamentals of Speech Recognition*. Prentice Hall, Englewood Cliffs, NJ, 1993. (Cité pages 9 et 77.)
- S.I. Resnick. *Adventures in Stochastic Processes*. Birkhäuser, Boston, 1992. (Cité page 25.)
- G.O. Roberts et J.S. Rosenthal. General state space Markov Chains and MCMC algorithms. *Probability Surveys*, 1 :20–71, 2004. (Cité pages 7 et 15.)
- A. Sadek et N. Limnios. Asymptotic properties for maximum likelihood estimators for reliability and failure rates of Markov Chains. *Communication in Statistics-Theory and Methods*, 31(10) :1837–1861, 2002. (Cité pages 6 et 24.)
- J. Sansom. A Hidden Markov model for Rainfall using breakpoint data. *Journal of Climate*, 11(1) :42–53, 1998. (Cité pages 9, 77 et 100.)
- J. Sansom et P.J. Thomson. Fitting hidden semi-Markov models to breakpoint rainfall data. *Journal of Applied Probability*, 38A :142–157, 2001. (Cité page 100.)
- J. Sansom et P.J. Thomson. On rainfall seasonality using a hidden semi-Markov model. *Journal of Geophysical Research*, D15 :112, 2007. (Cité page 100.)
- F. W. Scholz. *Maximum Likelihood estimation-Statistical Encyclopedia of Statistics*. Wiley, 2006. (Cité pages 5 et 6.)
- E. Seneta. *Non-Negative Matrices and Markov Chains*. Springer, 2006. (Cité page 32.)
- W.L. Smith. Regenerative stochastic processes. *Proc. Roy. Soc. London (A)*, 232 :6–31, 1955. (Cité page 7.)
- V.T. Stefanov. Explicit limit results for minimal sufficient statistics and maximum likelihood estimators in some Markov processes : Exponential families approach. *Ann. Stat.*, 23(4) :1073–1101, 1995. (Cité pages 6, 15 et 131.)
- Ö. Stenflo. Iterated function systems controlled by a semi-Markov chain. *Theory Stoch. Process.*, 18(1-2) :305–313, 1996. (Cité pages 8 et 44.)
- L. Takacs. Some investigations concerning recurrent stochastic processes of a certain type. *Magyar Tud. Akad. Mat. Kutato Int. Kzl.*, 3 :115–128, 1954. (Cité page 7.)
- S. Trevezas et N. Limnios. Maximum likelihood estimation for general hidden semi-Markov processes with backward recurrence time dependence. *Journal of Mathematical Sciences*, à paraître 2008a. (Cité pages 78 et 79.)

- S. Trevezas et N. Limnios. Variance estimation in the central limit theorem for Markov chains. *Journal of Statistical Planning and Inference*, à paraître 2008b. (Cité pages 16 et 24.)
- S. Trevezas et N. Limnios. Exact MLE and asymptotic properties for non-parametric semi-Markov models. soumis, 2008. (Cité page 45.)
- A.W. Van Der Vaart. *Asymptotic Statistics*. Cambridge University Press, New York, 1998. (Cité pages 5 et 37.)
- A. Wald. Note on the consistency of the maximum likelihood estimate. *Ann. Math. Stat.*, 20 :595–601, 1949. (Cité page 5.)
- G.C.G. Wei et M.A. Tanner. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, 85 :699–704, 1990. (Cité pages 11 et 111.)

Résumé Nous construisons l'estimateur du maximum de vraisemblance (EMV) de la loi stationnaire et de la variance asymptotique du théorème de la limite centrale (TLC) pour des fonctionnelles additives des chaînes de Markov ergodiques et nous démontrons sa convergence forte et sa normalité asymptotique. Ensuite, nous considérons un modèle semi-markovien non paramétrique. Nous présentons l'EMV exact du noyau semi-markovien qui gouverne l'évolution de la chaîne semi-markovienne (CSM) et démontrons la convergence forte, ainsi que la normalité asymptotique de chaque sous-vecteur fini de cet estimateur en obtenant des formes explicites pour les matrices de covariance asymptotiques. Ceci a été appliqué pour une observation de longue durée d'une seule trajectoire d'une CSM, ainsi que pour une suite des trajectoires i.i.d. d'une CSM censurée à un instant fixe. Nous introduisons un modèle semi-markovien caché (MSMC) général avec dépendance des temps de récurrence en arrière. Nous donnons des propriétés asymptotiques de l'EMV qui correspond à ce modèle. Nous déduisons également des expressions explicites pour les matrices de covariance asymptotiques qui apparaissent dans le TLC pour l'EMV des principales caractéristiques des CSM. Enfin, nous proposons une version améliorée de l'algorithme EM (Estimation-Maximisation) et une version stochastique de cet algorithme (SAEM) afin de trouver l'EMV pour les MSMC non paramétriques. Des exemples numériques sont présentés pour ces deux algorithmes.

Abstract We construct the maximum likelihood estimator (MLE) of the stationary distribution and of the asymptotic variance of the central limit theorem for additive functionals of ergodic Markov chains and we prove its strong consistency and its asymptotic normality. In the sequel, we consider a non-parametric semi-Markov model. We present the exact MLE of the semi-Markov kernel that governs the evolution of the semi-Markov chain (SMC) and we prove the strong consistency as well as the asymptotic normality of every finite subvector of this estimator by obtaining explicit forms for the asymptotic covariance matrices. The asymptotics were considered for one trajectory of a SMC as well as for a sequence of i.i.d. observations of a SMC censored at a fixed time. We introduce a general hidden semi-Markov model (HSMM) with backward recurrence time dependence. We prove asymptotic properties of the MLE that corresponds to this model. We also deduce explicit expressions for the asymptotic covariance matrices that appear in the CLT for the MLE of some basic characteristics of the SMC. Finally, we propose an improved version of the EM algorithm for HSMM and a stochastic version of this algorithm (SAEM), in order to find the MLE for non-parametric HSMMs. Numerical examples are presented for both algorithms.