Méthodes de lissage et d'estimation dans des modèles à variables latentes par des méthodes de Monte-Carlo séquentielles

Cyrille Dubarry

Université de Paris VI - Pierre et Marie Curie
Télécom SudParis

Ecole doctorale de sciences mathématiques de Paris centre

Thèse de doctorat n°2012TELE0040

Discipline : Mathématiques

Spécialité : Probabilités et statistiques

Présentée par :

Cyrille DUBARRY

Méthodes de lissage et d'estimation dans des modèles à variables latentes par des méthodes de Monte-Carlo séquentielles

Dirigée par :

Randal DOUC

Soutenue le 9 octobre 2012 devant le jury composé de :

Pierre DEL MORAL	Directeur de recherche - INRIA, Université Bordeaux I	(Examinateur)
Randal DOUC	Professeur - Telecom SudParis	(Directeur de thèse)
Emmanuel GOBET	Professeur - Ecole Polytechnique	(Rapporteur)
François LE GLAND	Directeur de recherche - INRIA Rennes	(Examinateur)
Eric MOULINES	Professeur - Télécom ParisTech	(Examinateur)
Wojciech PIECZYNSKI	Professeur - Telecom SudParis	(Examinateur)
Christian ROBERT	Professeur - Université Paris Dauphine	(Rapporteur - absent)
Mathieu ROSENBAUM	Professeur - UPMC	(Examinateur)

Le hasard

Il faut remonter aux croisades du *XI^e^* siècle pour que le mot *hasard* apparaisse dans la langue française. D'origine arabe, *az-zahr* désigne *le dé* en lui-même ou *le jeu de dés*. En outre, *zahra* signifie la *fleur*, faisant référence au symbole gravé sur la face gagnante des dés de cette époque. Le passage de l'arabe au français viendrait des Croisés, qui auraient joué à ce jeu de az-zahr pendant le siège d'un palais syrien, comme l'évoque le chroniqueur Guillaume de Tyr. Des cubes gravés, retrouvés dans des sarcophages égyptiens prouvent néanmoins qu'ils étaient utilisés bien avant le Moyen Age. Ces jeux, appréciés à toutes les époques, n'ont pas toujours été bien tolérés par les élites politiques. A l'époque romaine ils étaient même punis d'une lourde amende.

Direct Matin n°1015, 23/01/2012.

Résumé

Les modèles de chaînes de Markov cachées ou plus généralement ceux de Feynman-Kac sont aujourd'hui très largement utilisés. Ils permettent de modéliser une grande diversité de séries temporelles (en finance, biologie, traitement du signal, ...) La complexité croissante de ces modèles a conduit au développement d'approximations via différentes méthodes de Monte-Carlo, dont le Markov Chain Monte-Carlo (MCMC) et le Sequential Monte-Carlo (SMC). Les méthodes de SMC appliquées au filtrage et au lissage particulaires font l'objet de cette thèse. Elles consistent à approcher la loi d'intérêt à l'aide d'une population de particules définies séquentiellement. Différents algorithmes ont déjà été développés et étudiés dans la littérature. Nous raffinons certains de ces résultats dans le cas du Forward Filtering Backward Smoothing et du Forward Filtering Backward Simulation grâce à des inégalités de déviation exponentielle et à des contrôles non asymptotiques de l'erreur moyenne. Nous proposons également un nouvel algorithme de lissage consistant à améliorer une population de particules par des itérations MCMC, et permettant d'estimer la variance de l'estimateur sans aucune autre simulation. Une partie du travail présenté dans cette thèse concerne également les possibilités de mise en parallèle du calcul des estimateurs particulaires. Nous proposons ainsi différentes interactions entre plusieurs populations de particules. Enfin nous illustrons l'utilisation des chaînes de Markov cachées dans la modélisation de données financières en développant un algorithme utilisant l'Expectation-Maximization pour calibrer les paramètres du modèle exponentiel d'Ornstein-Uhlenbeck multi-échelles.

Mots-clés

Monte-Carlo séquentiel ; Filtrage particulaire ; Lissage particulaire ; Feynman-Kac ; Chaîne de Markov cachée ; Estimation

Smoothing and estimation methods in hidden variable models through sequential Monte-Carlo methods

Abstract

Hidden Markov chain models or more generally Feynman-Kac models are now widely used. They allow the modelling of a variety of time series (in finance, biology, signal processing, ...) Their increasing complexity gave birth to approximations using Monte-Carlo methods, among which Markov Chain Monte-Carlo (MCMC) and Sequential Monte-Carlo (SMC). SMC methods applied to particle filtering and smoothing are dealt with in this thesis. These methods consist in approximating the law of interest through a particle population sequentially defined. Different algorithms have already been developed and studied in the literature. We make some of these results more precise in the particular of the Forward Filtering Backward Smoothing and Forward Filtering Backward Simulation by showing exponential deviation inequalities and by giving non-asymptotic upper bounds to the mean error. We also introduce a new smoothing algorithm improving a particle population through MCMC iterations and allowing to estimate the estimator variance without further simulation. Part of the work presented in this thesis is devoted to the parallel computing of particle estimators. We study different interaction schemes between several particle populations. Finally, we also illustrate the use of hidden Markov chains in the modelling of financial data through an algorithm using Expectation-Maximization to calibrate the exponential Ornstein-Uhlenbeck multiscale stochastic volatility model.

Keywords

Sequential Monte-Carlo; Particle filtering; Particle smoothing; Feynman-Kac; Hidden Markov chain; Estimation

Département CITI
TELECOM SudParis
9, rue Charles Fourier
91011 EVRY Cedex
France
http ://citi.telecom-sudparis.eu

Table des matières

Chapitre 1

Introduction

Sommaire

1.1 Présentation générale

La nécessité de modéliser une grande diversité de séries temporelles étant de plus en plus forte, ces dernières décennies ont connu l'émergence de divers modèles à variables latentes dont le succès n'est plus à faire. Les domaines d'applications sont nombreux (finance, biologie, traitement du signal, ... voir Del Moral [2004] et Del Moral and Doucet [2010-2011]) pour par exemple l'inférence de paramètres ou le calcul d'espérances. Les propriétés statistiques des données réelles à modéliser ont mené au développement de modèles toujours plus généraux et donc plus complexes, pour lesquels les calculs exacts ne sont plus possibles. Différentes méthodes de Monte-Carlo ont alors fait leur apparition pour fournir des approximations de ces modèles.

Engle [1982] a introduit le modèle *ARCH* (*AutoRegressive Conditional Heteroscedasticity*), très utilisé par exemple en finance dans la modélisation de la volatilité. Ce modèle consiste à faire dépendre linéairement la variable d'intérêt d'un bruit non observé, cette dépendance remontant dans le temps jusqu'à un ordre fixé. Bollerslev [1986] a généralisé le modèle ARCH au modèle *GARCH* (*Generalized AutoRegressive Conditional Heteroscedasticity*) en autorisant la variable d'intérêt à dépendre également de ses valeurs passées. La relative simplicité des équations impliquées permet d'estimer ces modèles de manière exacte grâce à la méthode des moindres carrés. Tout un éventail de modèles autorégressifs ont ensuite été développés et étudiés.

Des modèles offrant plus de souplesse comme les *chaînes de Markov cachées* (*HMM* pour *Hidden Markov Model*) ou plus généralement les *modèles de Feynman-Kac* sont aujourd'hui très largement utilisés et font l'objet de cette thèse (voir aussi Cappé et al. [2005] et Del Moral [2004]). Dans le cas des HMM, il s'agit d'un couple de processus stochastiques dont une seule composante est observée. Elle dépend de manière aléatoire de l'autre composante qui est une chaîne de Markov non observée. Il existe deux cas simples pour lesquels il est possible de retrouver de manière exacte la loi du processus caché conditionnellement aux observations. Le premier est celui où l'espace d'état est fini et le second est le *modèle linéaire gaussien*, résolu par la méthode du *filtre de Kalman* (Kalman and Bucy [1961]). Dans le cas où les dépendances ne sont pas tout à fait linéaire, le *filtre de Kalman étendu* propose d'utiliser des développements en série de Taylor pour linéariser le système. Il ne s'agit bien sûr que d'une approximation, et pour des modèles plus

complexes il est préférable d'utiliser des approximations par méthode de *Monte-Carlo* (voir Handschin and Mayne [1969] et Rubin [1987]).

Les méthodes de *Markov Chain Monte-Carlo* ou *MCMC* (voir Andrieu et al. [2003] pour une présentation plus complète) consistent à approcher une loi cible en construisant une chaîne de Markov ergodique admettant pour loi stationnaire la loi cible. Plus particulièrement, dans le cas où la densité de la loi cible est connue à une constante près, l'algorithme de *Metropolis-Hastings* est utilisé. Il s'agit exactement du cas des HMM où la densité de la loi de la chaîne de Markov cachée conditionnellement aux observations est connue à une constante près. L'idée est de proposer un nouveau candidat et de l'accepter avec une probabilité dépendant d'un ratio de vraisemblance. Dans le cas où la loi cible correspond à celle d'une chaîne de Markov (les HMM sont donc concernées), les algorithmes de *Gibbs* et plus généralement de *Metropolis-within-Gibbs* suggèrent d'actualiser les composantes de la chaîne une à une afin d'améliorer la probabilité d'acceptation.

Une alternative aux méthodes MCMC est l'utilisation des méthodes de *Monte-Carlo séquentielles* (*SMC*), appliquées par Gordon et al. [1993] au filtrage et au lissage particulaires. Ces méthodes, que nous étudions de manière intensive dans cette thèse, consistent à approcher la loi d'intérêt à l'aide d'une population de particules définies de manière séquentielle (voir Doucet and Johansen [2009], Cappé et al. [2007], Fearnhead [2008], Del Moral [2004], Cappé et al. [2005]). Le *Forward-Filter*, aussi connu dans la littérature sous le nom de *auxiliary filter* alterne des étapes de sélection et de mutation des particules de manière à approcher séquentiellement la loi de filtrage. Il permet également d'approcher la loi de lissage en conservant la généalogie des particules donnée par les étapes de sélection (Kitagawa [1996]). Cet algorithme appelé *Filter-Smoother* présente le double avantage d'être très simple et de complexité linéaire par rapport au nombre de particules mais souffre d'un problème de dégénérescence bien connu. Pour cette raison, Doucet et al. [2000] introduisent le *FFBS* (*Forward Filtering Backward Smoothing*) qui ajoute une passe backward au Forward-Filter afin d'approcher la loi de lissage mais la complexité devient quadratique pour les lois marginales et exponentielle en temps pour la loi jointe de lissage. Cet algorithme a ensuite été étendu au *FFBSi* (*Forward Filtering Backward Simulation*) par Godsill et al. [2004] permettant d'approcher la loi jointe de lissage avec une complexité quadratique, puis une implémentation par une méthode de rejet du FFBSi proposée par Douc et al. [2010] a permis d'atteindre une complexité linéaire par rapport au nombre de particules. Dans ce même article, des inégalités de déviation exponentielle et des théorèmes *central limite* sont établis en ce qui concerne l'erreur d'approximation générée par les deux algorithmes FFBS et FFBSi. La dépendance en fonction de l'horizon de temps n'est donnée que pour la loi marginale de lissage. Dans le cas où l'approximation de la loi jointe de lissage est appliquée à une fonctionnelle additive, la norme L_q de cette même erreur est bornée de manière non asymptotique dans Del Moral et al. [2010a,b]. En ce qui concerne d'abord la variance, Del Moral et al. [2010b] établissent qu'elle est bornée par des termes ne faisant intervenir que le ratio horizon de temps sur nombre de particules. En revanche, dès lors que $q > 2$, la norme L_q est bornée dans Del Moral et al. [2010a] par des termes plus volumineux où l'horizon de temps est au carré. C'est pourquoi nous consacrons l'article théorique correspondant au Chapitre 4 à affiner ce résultat. Nous montrons ainsi que sous une hypothèse de noyau fortement mélangeant pour la chaîne de Markov cachée, la norme L_q pour $q \geq 2$ peut en réalité être bornée non asymptotiquement par des termes ne faisant intervenir que le ratio horizon de temps sur nombre de particules. Ce résultat est très intéressant en pratique lors de l'estimation de paramètres. En effet, le calcul du gradient de la fonction de log-vraisemblance (score de Fischer) et l'étape *Expectation* de l'algorithme *EM* (*Expectation-Maximization*) font naturellement apparaître des fonctionnelles additives telles que celles étudiées dans le Chapitre 4 (voir Cappé et al. [2005] et Doucet et al. [2010]). Il existe enfin un dernier algorithme d'approximation particulaire de la loi de lissage, mais il ne donne accès qu'aux lois marginales. Il s'agit du *Two-Filter smoother* introduit par [Briers et al., 2010] qui consiste à utiliser un filtre forward et un autre backward afin de les coupler. Cette méthode souffre originellement d'une complexité quadratique mais elle a été récemment modifiée par [Fearnhead et al., 2010] pour devenir linéaire par rapport au nombre de particules.

La littérature propose aussi d'allier les méthodes MCMC et SMC. Par exemple, Andrieu et al. [2010] proposent une nouvelle procédure appelée *PMCMC* (*Particle Markov Chain Monte-Carlo*) qui est un algorithme de Metropolis-Hastings dans lequel le candidat proposé est obtenu par une méthode SMC. Les vraisemblances intervenant dans le ratio d'acceptation sont estimées de manière non biaisée par cette même méthode SMC. Moins récemment, Gilks and Berzuini [2001] avaient proposé le mariage inverse. L'algo-

rithme de base était le Forward-Filter et à chaque étape de temps la population de particules était améliorée par une passe de MCMC appliquée à chaque trajectoire. Pour un horizon de temps fixé, cette méthode n'est pas rentable car la diversification apportée par la passe MCMC est à nouveau détériorée par l'étape de sélection de l'algorithme SMC. Partant de cette constatation, nous proposons dans l'article correspondant au Chapitre 6 une meilleure exploitation de cette idée. L'horizon de temps étant fixé, nous améliorons les trajectoires de particules produites par le Forward-Smoother classique. Chaque trajectoire est utilisée comme point de départ d'un algorithme de Metropolis-within-Gibbs de loi cible la loi jointe de lissage. Ces algorithmes sont exécutés de manière indépendante et au lieu d'utiliser la propriété d'ergodicité des chaînes générées, nous choisissons d'oublier les états intermédiaires et d'approcher la loi de lissage uniquement à l'aide des trajectoires finales. Les expérimentations numériques du Chapitre 6 montrent que le nombre d'étapes des algorithmes MCMC nécessaire à la convergence est très faible, rendant l'algorithme rapide à exécuter et plus efficace que les autres algorithmes de lissage particulaire de complexité linéaires. De plus, une analyse plus théorique de la variance de l'estimateur ainsi obtenu permet de montrer que les trajectoires se comportent asymptotiquement comme des simulations indépendantes de la loi jointe de lissage, permettant ainsi d'obtenir simplement des intervalles de confiance pour notre estimateur à l'aide d'un seul jeu de particules. Cette possibilité n'est offerte par aucune méthode MCMC ou SMC.

Comme mentionné précédemment, les HMM font aussi l'objet d'inférence de paramètres. A partir d'une loi *a priori* sur la valeur des paramètres, les méthodes MCMC permettent d'approcher leur loi *a posteriori* conditionnellement aux observations. Une autre technique utilisée pour l'estimation de paramètres est la maximisation de la log-vraisemblance. Que l'on utilise une méthode du gradient ou d'Expectation-Maximization, il est primordial de pouvoir approcher la distribution jointe de lissage. En effet, même si dans les modèles simples, la loi marginale (des couples ou des triplets consécutifs) suffit, certains modèles nécessitent par exemple la loi de tous les couples comme c'est le cas dans le modèle de volatilité stochastique multi-échelle (voir Masoliver and Perello [2006], Buchbinder and Chistilin [2007], Eisler et al. [2007]) étudié dans l'article faisant l'objet du Chapitre 5. En effet, ce modèle présente une dégénérescence du noyau de transition de la chaîne de Markov cachée et ne permet pas, tel quel, d'écrire la log-vraisemblance utilisée dans l'algorithme EM. Nous avons donc changé de variable cachée et les espérances conditionnelles qui en ont suivi exigeaient plus que les simples lois marginales de lissage. La calibration d'un modèle similaire mais non dégénéré a été étudiée par Fouque et al. [2008] grâce à des techniques MCMC, avec moins de succès numériquement.

Structure du document

Le travail de recherche présenté dans cette thèse se décompose en trois articles de revue déjà soumis à publication, retranscrits dans leur intégralité dans les Chapitres 4, 5 et 6, ainsi que d'un travail de recherche additionnel présenté dans le Chapitre 3 et réalisé en collaboration avec Pierre Del Moral et Eric Moulines.

L'article du Chapitre 4 concerne l'analyse théorique de deux algorithmes de lissage particulaire existants, le FFBS et le FFBSi. Nous y exhibons des bornes non asymptotiques concernant l'erreur d'approximation générée par ces algorithmes. Le Chapitre 5 est au contraire beaucoup plus pratique : nous y développons une méthode d'estimation de paramètres dans le cas d'un modèle complexe, dont l'étude n'avait pas jamais été menée à son terme auparavant. Enfin les Chapitres 3 et 6 sont partagés entre pratique et théorie. Dans le Chapitre 3, nous proposons deux algorithmes de parallélisation du Forward-Filter et Forward-Smoother ainsi qu'une analyse asymptotique des variances et biais associés afin de déterminer, selon les cas, quel algorithme est le plus favorable. Ce chapitre donnera prochainement lieu à un article de revue. Le Chapitre 6 propose lui aussi un nouvel algorithme de lissage, rapide à exécuter et amorce son analyse théorique.

Dans la suite du présent chapitre, nous introduisons les notations utilisées tout en détaillant mathématiquement les motivations décrites dans la présentation générale. Les autres chapitres s'organisent de la manière suivante :

Chapitre 2. (Préambule) Nous y présentons de manière détaillée les algorithmes de lissage particulaire présents dans la littérature afin de pouvoir les comparer. Ce chapitre est aussi l'occasion d'introduire de manière plus détaillée certains articles formant cette thèse.

Chapitre 3. (Préambule) *Convergence des modèles d'îlot de Feynman-Kac* (Dubarry, Del Moral, Moulines). Nous présentons ici certains résultats qui donneront prochainement lieu à un article de revue. Ce travail sera présenté au congrès 8th World Congress in Probability and Statistics (Istanbul, Turquie, Juillet 2012) et au workshop Sequential Monte Carlo methods and Efficient simulation in Finance (Ecole Polytechnique, Paris, Octobre 2012).

Chapitre 4. (Article) *Non-asymptotic deviation inequalities for smoothed additive functionals in nonlinear state-space models* (Dubarry, Le Corff). Accepté à Bernoulli Journal. Certains résultats ont été présentés aux conférences European Meeting of Statisticians (Athènes, Grèce, Août 2010) et Statistical Signal Processing (Nice, France, Juin 2011).

Chapitre 5. (Article) *Calibrating the exponential Ornstein-Uhlenbeck multiscale stochastic volatility model* (Dubarry, Douc). Soumis à Quantitative Finance. Nous donnons ici la version actuelle de l'article (2^{nd} round). Les principaux résultats ont été présentés à la conférence Stochastic Processes and their Applications (Oaxaca, Mexique, 2011)

Chapitre 6. (Article) *Particle approximation improvement of the joint smoothing distribution with on-the-fly variance estimation* (Dubarry, Douc). Soumis à Statistics and computing. Nous donnons ici la version actuelle de l'article (1^{st} round). Certains résultats préliminaires ont été présentés à la conférence Statistical Signal Processing (Nice, France, Juin 2011).

1.2 Cadre et notations

1.2.1 HMM

Nous nous intéressons dans cette thèse au cas particulier des chaînes de Markov cachées. Une HMM se définit comme un double processus stochastique où une chaîne de Markov $\{X_t\}_{t=0}^{\infty}$ est partiellement observée à travers une séquence d'observations $\{Y_t\}_{t=0}^{\infty}$. Plus précisément, soient $\mathbb{X}$ et $\mathbb{Y}$ deux espaces d'états (discrets ou continus) munis respectivement des tribus $\mathcal{X}$ et $\mathcal{Y}$. Nous désignons alors par M un noyau de transition markovien sur $(\mathbb{X},\mathcal{X})$ et par G un noyau de transition de $(\mathbb{X},\mathcal{X})$ vers $(\mathbb{Y},\mathcal{Y})$ tels que la dynamique du processus $\{(X_t,Y_t)\}_{t=0}^{\infty}$ soit donnée par le noyau de transition markovien suivant :

$$P[(x,y),\mathsf{A}] \stackrel{\text{def}}{=} M\otimes G[(x,y),\mathsf{A}] = \iint M(x,\mathrm{d}x')\,G(x',\mathrm{d}y')\mathbf{1}_{\mathsf{A}}(x',y')\,, \tag{1.1}$$

où $(x,y)\in\mathbb{X}\times\mathbb{Y}$ et $\mathsf{A}\in\mathcal{X}\otimes\mathcal{Y}$. Nous supposons qu'il existe deux mesures finies positives λ sur $(\mathbb{X},\mathcal{X})$ et μ sur $(\mathbb{Y},\mathcal{Y})$ dominant respectivement $M(x,\cdot)$ et $G(x,\cdot)$ pour tout $x\in\mathbb{X}$. Il existe donc deux densités m et g telles que

$$\forall(x,x',y)\in\mathbb{X}\times\mathbb{X}\times\mathbb{Y},\quad m(x,x') \stackrel{\text{def}}{=} \frac{\mathrm{d}M(x,\cdot)}{\mathrm{d}\lambda}(x') \text{ et } g(x,y) \stackrel{\text{def}}{=} \frac{\mathrm{d}G(x,\cdot)}{\mathrm{d}\mu}(y)\,,$$

où la fonction qui à $x\in\mathbb{X}$ associe $g(x,y)$ est connue sous le nom de fonction de vraisemblance d'un état étant donnée une observation $y\in\mathbb{Y}$. Par souci de simplicité, nous noterons $\lambda(\mathrm{d}x)$ par $\mathrm{d}x$. Nous supposons également que X_0 suit une loi χ admettant une densité par rapport à λ et par abus de notation, nous écrirons $\chi(\mathrm{d}x)=\chi(x)\lambda(\mathrm{d}x)=\chi(x)\mathrm{d}x$.

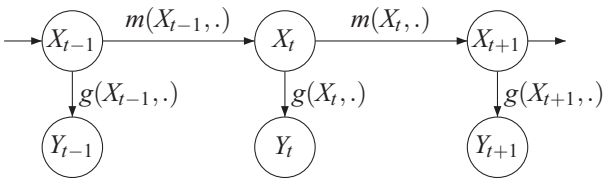


FIGURE 1.1 – HMM

Les notations introduites sont résumées dans la Figure 1.1.

1.2.2 Filtrage et lissage

Lorsque la loi initiale χ, et les noyaux de transition m et g sont connus, un premier problème est de retrouver la loi des variables cachées $X_{0:T}$ conditionnellement aux observations $Y_{0:T}$ où la notation $a_{u:v}$ est une abréviation pour $\{a_s\}_{s=u}^{v}$. Nous considérons par la suite que les observations $Y_{0:T} = y_{0:T}$ sont fixées, elles seront donc omises dans les notations. On définit pour tout $0 \leq s \leq u \leq T$ et toute fonction mesurable h sur $\mathbb{X}^{u-s+1}$:

$$\phi_{s:u|T}[h] \stackrel{\text{def}}{=} \frac{\int \chi(\mathrm{d}x_0) g(x_0,y_0) \prod_{t=1}^{T} M(x_{t-1},\mathrm{d}x_t) g(x_t,y_t) h(x_{s:u})}{\int \chi(\mathrm{d}x_0) g(x_0,y_0) \prod_{t=1}^{T} M(x_{t-1},\mathrm{d}x_t) g(x_t,y_t)} = \mathbb{E}\left[h(X_{s:u})|Y_{0:T}\right] . \tag{1.2}$$

$\phi_{T|T}$ est appelée *loi de filtrage* et $\phi_{0:T|T}$ *loi jointe de lissage*. On pourra aussi s'intéresser aux marginales de la loi de lissage $\phi_{t|T}$ et $\phi_{t-1,t|T}$.

Exemple 1.1 (Modèle linéaire gaussien)**.** Le HMM le mieux connu est le *modèle linéaire gaussien* (*LGM, Linear Gaussian Model*) défini sur $\mathbb{X} = \mathbb{R}^n$ et $\mathbb{Y} = \mathbb{R}^m$ par :

$$\begin{aligned} X_{t+1} &= F_{t+1} X_t + W_{t+1} , \\ Y_t &= H_t X_t + V_t , \end{aligned}$$

où X_0 suit une loi gaussienne de moyenne $M_0 \in \mathbb{R}^n$ et de matrice de covariance $Q_0 \in \mathcal{M}_{n,n}(\mathbb{R})$ (que nous noterons $X_0 \sim \mathcal{N}(M_0, Q_0)$), et pour tout $t \geq 0$, $F_{t+1} \in \mathcal{M}_{n,n}(\mathbb{R})$, $W_{t+1} \sim \mathcal{N}(0, Q_{t+1})$, $Q_{t+1} \in \mathcal{M}_{n,n}(\mathbb{R})$, $H_t \in \mathcal{M}_{m,n}(\mathbb{R})$, $V_t \sim \mathcal{N}(0, R_t)$, et $R_t \in \mathcal{M}_{m,m}(\mathbb{R})$. Le filtre de Kalman donne alors la distribution de filtrage exacte de X_T conditionnellement à $Y_{0:T}$ comme une gaussienne de moyenne $\hat{X}_T$ et de matrice de covariance P_T calculées grâce à l'algorithme 1.

Algorithm 1 Filtre de Kalman

Entrées : M_0, $Q_{0:T}$, $F_{1:T}$, $H_{0:T}$, $R_{0:T}$, et $Y_{0:T}$.
Initialisation :
$P_0 = (Q_0^{-1} + H_0^{\mathrm{T}} R_0^1 H_0)^{-1}$,
$\hat{X}_0 = P_0(Q_0^{-1} M_0 + H_0^{\mathrm{T}} R_0^{-1} Y_0)$.
Mise à jour récursive :
for t from 1 to T **do**
  $\bar{X}_t = F_t \hat{X}_{t-1}$,
  $\bar{P}_t = F_t P_{t-1} F_t^{\mathrm{T}} + Q_t$,
  $\tilde{Y}_t = Y_t - H_t \bar{X}_t$,
  $S_t = H_t \bar{P}_t H_t^{\mathrm{T}} + R_t$,
  $K_t = \bar{P}_t H_t^{\mathrm{T}} S_t^{-1}$,
  $\hat{X}_t = \bar{X}_t + K_t \tilde{Y}_t$,
  $P_t = (\mathrm{I}_n - K_t H_t) \bar{P}_t$.
end for
Sorties : $\hat{X}_T$ et P_T.

1.2.3 Inférence et EM

Dans le cas où la distribution de la HMM, c'est à dire χ_θ, m_θ et g_θ, dépend d'un paramètre $\theta \in \Theta$ inconnu, un autre problème très courant est l'inférence de ce paramètre. L'estimateur du maximum de vraisemblance donné par

$$\theta^* \stackrel{\text{def}}{=} \mathrm{argmax}_{\theta\in\Theta} \log p_\theta(y_{0:T}) = \mathrm{argmax}_{\theta\in\Theta} \log \int \chi_\theta(\mathrm{d}x_0) g_\theta(x_0,y_0) \prod_{t=1}^{T} M_\theta(x_{t-1},\mathrm{d}x_t) g_\theta(x_t,y_t) ,$$

n'est presque jamais calculable directement, et l'algorithme *Expectation-Maximization* (EM) est couramment utilisé pour l'approcher. Il définit pour cela une suite récursive $(\hat{\theta}_n)$ dans l'espace des paramètres Θ par

$$\hat{\theta}_{n+1} = \mathrm{argmax}_{\theta\in\Theta} \mathbb{E}_{\hat{\theta}_n}\left[\log p_\theta(Y_{0:T}, X_{0:T})|Y_{0:T}\right] , \tag{1.3}$$

où

$$p_\theta(y_{0:T},x_{0:T}) = \chi_\theta(x_0)g_\theta(x_0,y_0)\prod_{t=1}^{T} m_\theta(x_{t-1},x_t)g_\theta(x_t,y_t)\ .$$

Il est alors facile de vérifier que la log-vraisemblance $\log p_{\hat\theta_n}(y_{0:T})$ augmente à chaque étape de la récurrence :

$$\begin{aligned}
\log p_{\hat\theta_{n+1}}(y_{0:T}) &= \log\int p_{\hat\theta_{n+1}}(x_{0:T},y_{0:T})\mathrm{d}x_{0:T} = \log\int \frac{p_{\hat\theta_{n+1}}(x_{0:T},y_{0:T})}{p_{\hat\theta_n}(x_{0:T}|y_{0:T})}p_{\hat\theta_n}(x_{0:T}|y_{0:T})\mathrm{d}x_{0:T}\\
&\geq \int \log\left(\frac{p_{\hat\theta_{n+1}}(x_{0:T},y_{0:T})}{p_{\hat\theta_n}(x_{0:T}|y_{0:T})}\right)p_{\hat\theta_n}(x_{0:T}|y_{0:T})\mathrm{d}x_{0:T}\\
&= \mathbb{E}_{\hat\theta_n}\left[\log p_{\hat\theta_{n+1}}(Y_{0:T},X_{0:T})\middle|Y_{0:T}\right] - \mathbb{E}_{\hat\theta_n}\left[\log p_{\hat\theta_n}(Y_{0:T},X_{0:T})\middle|Y_{0:T}\right] + \log p_{\hat\theta_n}(y_{0:T}) \geq \log p_{\hat\theta_n}(y_{0:T})\ ,
\end{aligned}$$

où la dernière inégalité vient de la définition de $\hat\theta_{n+1}$. L'étape E de l'algorithme se ramène alors à calculer des espérances de la forme

$$\mathbb{E}_{\hat\theta_n}\left[\log\left(m_\theta(X_{t-1},X_t)g_\theta(X_t,Y_t)\right)|Y_{0:T}\right]\ .$$

Dans les configurations les plus simples (modèles exponentiels par exemple), l'étape M de maximisation peut être facilement réalisée à partir de ces espérances et approcher les lois de lissage marginales de tous les couples (X_{t-1},X_t), $0<t\leq T$ suffit à implémenter l'algorithme (voir Exemple 1.2). Cependant, certains modèles plus complexes nécessitent d'approcher la loi de lissage jointe de $X_{0:T}$ afin de pouvoir séparer les étapes E et M (voir l'exemple de la Sous-section 1.2.4 et pour plus de détails le Chapitre 5).

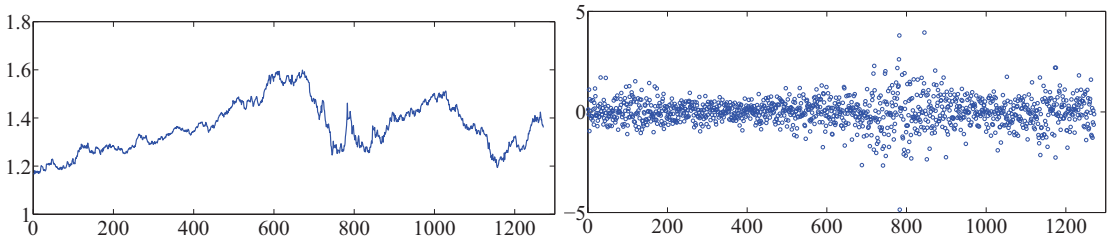
Exemple 1.2 (Modèle de volatilité stochastique)**.** Les *modèles de volatilité stochastique* (*SVM, Stochastic Volatility Models*) ont été introduits afin de mieux modéliser les séries temporelles de données financières que les modèles ARCH/GARCH ([Hull and White, 1987]). Nous considérons le SVM élémentaire introduit par [Hull and White, 1987] :

$$\begin{cases} X_{t+1} = \alpha X_t + \sigma U_{t+1}\ ,\\ Y_t = \beta \mathrm{e}^{\frac{X_t}{2}}V_t\ , \end{cases}$$

où $X_0 \sim \mathcal{N}\left(0,\frac{\sigma^2}{1-\alpha^2}\right)$, U_t et V_t sont des variables aléatoires indépendantes distribuées selon la loi gaussienne standard. Le paramètre $\theta \overset{\text{def}}{=} (\alpha,\sigma,\beta)$ est supposé inconnu et peut être récursivement estimé grâce à l'algorithme EM précédemment décrit. La suite $(\hat\theta_n)$ définie par 1.3 prend ses valeurs dans l'espace des paramètres $\Theta \overset{\text{def}}{=} (-1,1)\times(0,\infty)\times(0,\infty)$ et la maximisation du logarithme de la fonction de vraisemblance permet d'obtenir la récurrence suivante :

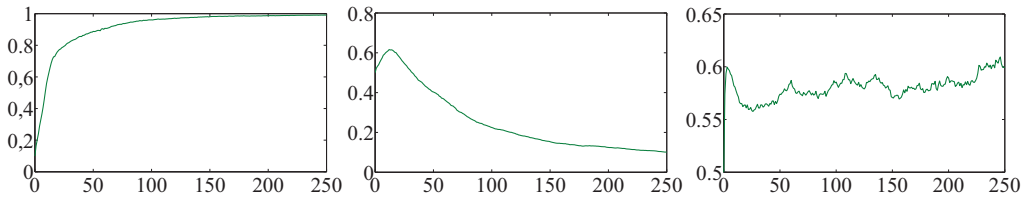
$$\begin{cases} \hat\phi_{n+1} = \dfrac{\sum_{t=1}^{T}\mathbb{E}_{\hat\theta_n}\left[X_{t-1}X_t|Y_{0:T}\right]}{\sum_{t=0}^{T-1}\mathbb{E}_{\hat\theta_n}\left[X_t^2\middle|Y_{0:T}\right]}\ ,\\ \hat\sigma^2_{n+1} = \dfrac{1}{T}\left(\sum_{t=1}^{T}\mathbb{E}_{\hat\theta_n}\left[X_t^2\middle|Y_{0:T}\right] + \hat\phi^2_{n+1}\sum_{t=1}^{T}\mathbb{E}_{\hat\theta_n}\left[X_{t-1}^2\middle|Y_{0:T}\right] - 2\hat\phi_{n+1}\sum_{t=1}^{T}\mathbb{E}_{\hat\theta_n}\left[X_{t-1}X_t|Y_{0:T}\right]\right)\ ,\\ \hat\beta^2_{n+1} = \dfrac{1}{T+1}\sum_{t=0}^{T}Y_t^2\mathbb{E}_{\hat\theta_n}\left[\exp(-X_t)|Y_{0:T}\right]\ . \end{cases} \tag{1.4}$$

La mise à jour du paramètre permet bien de séparer les étapes E et M de l'algorithme EM et nécessite de connaître les lois de lissage marginales de tous les couples (X_{t-1},X_t), $0<t\leq T$, ce qui n'est pas réalisable. Pour contourner ce problème, [Wei and Tanner, 1991] ont proposé de remplacer les espérances conditionnelles par des approximations de Monte-Carlo. Dans cet exemple, nous avons utilisé l'approximation donnée par le Filter-Smoother (décrit dans la Section 2.1 du Chapitre 2) avec $N = 20000$ particules et l'algorithme a été appliqué à des données réelles du taux de change EUR/USD sur cinq ans. Le taux de change quotidien ainsi que son rendement (dont nous avons supprimé la tendance selon [Kim et al., 1998]) sont visibles sur la Figure 1.2. La convergence de l'algorithme EM est montrée dans la Figure 1.3 pour des itérations allant de 0 à 250.



(a) Taux de change quotidien (b) Rendement sans tendance

FIGURE 1.2 – Données EUR/USD pour la période 16/11/2005-16/11/2010



(a) Estimation de α (b) Estimation de σ (c) Estimation de β

FIGURE 1.3 – Estimation de θ dans le SVM

1.2.4 Exemple du modèle de volatilité stochastique multi-échelles

Le modèle précédent ne tient compte que d'un seul facteur explicatif de la volatilité (la dimension de la variable cachée X est 1) mais peut être généralisé à une dimension supérieure, rendant alors l'estimation de ses paramètres plus problématique (voir Chapitre 5).

Le modèle de volatilité stochastique d'Ornstein-Uhlenbeck multi-échelles permet de prendre en compte certaines propriétés des données financières (voir Masoliver and Perello [2006], Buchbinder and Chistilin [2007]). Ce modèle est présenté ci-dessous sous la forme d'un système d'équations différentielles stochastiques :

$$\begin{cases} \mathrm{d}S_t = S_t[\kappa_t \mathrm{d}t + \sigma_t^S \mathrm{d}B_t^S] \,, \\ \sigma_t^S = \exp\left(\frac{1}{2}\langle 1_p, \xi_t \rangle\right) \,, \\ \mathrm{d}\xi_t = \mathrm{diag}\,(a)\,(\mu - \xi_t)\mathrm{d}t + b\mathrm{d}B_t^\xi \,, \end{cases} \tag{1.5}$$

où S est le prix d'un actif financier, κ est un drift, σ^S est la volatilité de S expliquée par un vecteur ξ de dimension p composé de processus d'Ornstein-Uhlenbeck de paramètres $a = (a_1, \cdots, a_p)^T, \mu = (\mu_1, \cdots, \mu_p)^T$ et $b = (b_1, \cdots, b_p)^T$, et $\langle \cdot, \cdot \rangle$ est le produit scalaire usuel de deux vecteurs de $\mathbb{R}^p$. Le drift κ sera supprimé des données réelles grâce à la méthode proposée par Kim et al. [1998]. De plus, sans perte de généralité, nous supposons que les composantes de a sont deux à deux distinctes. Le processus (B^S, B^ξ) est un mouvement brownien de dimension 2 de corrélation ρ, c'est-à-dire $\mathrm{d} < B^S, B^\xi >_t = \rho \mathrm{d}t$, et marginalement B^S et B^ξ sont des mouvements browniens standards. 1_p est le vecteur de dimension p constitué que de 1. Généralement, les composantes ξ^i de ξ sont associées aux échelles de temps $1/a_i$. Le même mouvement brownien B^ξ est utilisé pour toutes les composantes de ξ. Ce choix est justifié dans l'introduction du Chapitre 5, il est nécessaire dans la pratique mais pose des problèmes évidents de dégénérescence. De plus la corrélation entre l'actif et sa volatilité complexifie un peu plus le modèle mais est vitale pour décrire l'effet de levier très observé sur les marchés d'actions (Black [1976], Christie [1982]).

Pour pouvoir utiliser l'algorithme EM dans l'estimation des paramètres a, μ, b et ρ, nous avons besoin d'une discrétisation de (1.5). En appliquant un schéma d'Euler de pas fixe, on obtient :

Définition 1.1. LE MODÈLE DISCRÉTISÉ.

$$\begin{cases} X_{k+1} = \mathrm{diag}\,(\alpha) X_k + \sigma W_{k+1} \,, \\ Y_{k+1} = \beta e^{\langle 1_p, X_k \rangle / 2} V_{k+1} \,, \end{cases} \tag{1.6}$$

où $\{(W_k, V_k)\}_{k\geq 1}$ est une suite de vecteurs gaussiens iid tels que

$$\begin{bmatrix} W_1 \\ V_1 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) \,,$$

et α, σ, β sont définis en fonction de a, b, μ et du pas du schéma d'Euler.

Le modèle précédent est donc paramétré par le vecteur θ de dimension $2p+2$ que l'on suppose appartenir à l'espace :

$$\Theta_p \stackrel{\text{def}}{=} \left\{ (\alpha, \sigma, \beta, \rho) \in [0,1)^p \times (\mathbb{R}_+^\star)^p \times \mathbb{R}_+^\star \times (-1,1); \alpha_1 < \cdots < \alpha_p \right\} \,,$$

et l'estimation de ce paramètre sera basée uniquement sur les observations $(Y_k)_{k\geq 1}$. Nous supposons aussi que la distribution de X_0 est la distribution stationnaire de la chaîne de Markov X :

$$X_0 \sim \mathcal{N}\left(0, \Upsilon_{\alpha,\sigma}\right) \,, \tag{1.7}$$

où la matrice de covariance satisfait $\Upsilon_{\alpha,\sigma} = \mathrm{diag}\,(\sigma) A_\alpha \,\mathrm{diag}\,(\sigma)$, avec $A_\alpha = \left((1-\alpha_i\alpha_j)^{-1}\right)_{1\leq i,j\leq p}$. Cette distribution n'est pas dégénérée.

Le Théorème suivant montre alors que le modèle discret est identifiable, et que, plus précisément, le vecteur $(Y_1, \ldots, Y_{2p})$ suffit à caractériser le paramètre θ (la preuve se trouve dans le Chapitre 5).

Théorème 1.1. *Soient $\theta^{(i)} = (\alpha^{(i)}, \sigma^{(i)}, \beta^{(i)}, \rho^{(i)})$, $i \in \{1,2\}$ deux jeux de paramètres dans Θ_p, et deux paires de processus $(X^{(i)}, Y^{(i)})$, $i \in \{1,2\}$ tels que pour tout $k \geq 0$,*

$$\begin{cases} X_{k+1}^{(i)} = \mathrm{diag}\left(\alpha^{(i)}\right) X_k^{(i)} + \sigma^{(i)} W_{k+1}^{(i)} \,, \\ Y_{k+1}^{(i)} = \beta^{(i)} \mathrm{e}^{\left\langle 1_p, X_k^{(i)} \right\rangle / 2} V_{k+1}^{(i)} \,, \end{cases} \tag{1.8}$$

où $\{(W_k^{(1)}, V_k^{(1)})\}_{k\geq 1}$ et $\{(W_k^{(2)}, V_k^{(2)})\}_{k\geq 1}$ sont deux suites indépendantes de vecteurs gaussiens iid tels que

$$\begin{bmatrix} W_1^{(i)} \\ V_1^{(i)} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) \,,$$

et $X_0^{(i)} \sim \mathcal{N}\left(0, \Upsilon_{\alpha^{(i)}, \sigma^{(i)}}\right)$. Alors les trois assertions suivantes sont équivalentes :

$$i)\ Y_{1:2p}^{(1)} \stackrel{\mathcal{L}}{=} Y_{1:2p}^{(2)} \,, \qquad ii)\ \theta^{(1)} = \theta^{(2)} \,, \qquad iii)\ \forall k \geq 1, \quad Y_{1:k+1}^{(1)} \stackrel{\mathcal{L}}{=} Y_{1:k+1}^{(2)} \,.$$

L'estimation de θ grâce à l'algorithme EM propose de définir une suite récursive $(\hat{\theta}_n)$ dans l'espace des paramètres Θ_p satisfaisant

$$\hat{\theta}_{n+1} = \mathrm{argmax}_{\theta\in\Theta_p} \, \mathbb{E}_{\hat{\theta}_n} \left[\log p_\theta(Y_{1:T}, U) | \, Y_{1:T} \right] \,, \tag{1.9}$$

où $\log p_\theta(Y_{1:T}, U)$ est la log-vraisemblance de la distribution jointe de $(Y_{1:T}, U)$ pour toute variable cachée U. L'idée la plus naturelle serait de choisir simplement $U = X_{0:T}$, mais cette variable cachée est dégénérée, ne permettant pas de calculer la log-vraisemblance. En revanche le choix de $U = (X_0, \tilde{X}_{1:T})$ où pour tout $t \leq T$, $\tilde{X}_t \stackrel{\text{def}}{=} \langle \mathbb{1}_p, X_t \rangle$ permet à la fois de calculer la log-vraisemblance et de séparer les étapes E et M. En effet, la suite $(\hat{\theta}_n)$ vérifie alors

$$\hat{\theta}_{n+1} = \mathrm{argmax}_{\theta\in\Theta_p} \, \ell_{n,\theta}^{1,T}(Y_{1:T}) \,, \tag{1.10}$$

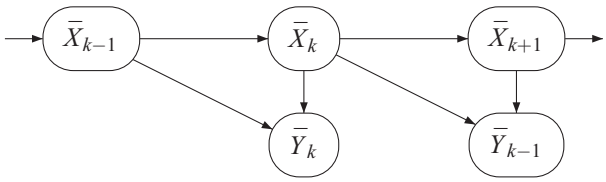


FIGURE 1.4

où pour $1 \leq r < s \leq T$, $\ell_{n,\theta}^{r,s}(Y_{r:s})$ peut être décomposé en :

$$\ell_{n,\theta}^{r,s}(Y_{r:s}) = \sum_{\nu} \sum_{k,\ell=1}^{T} \sum_{i,j=1}^{p} f_{\nu,k,\ell,i,j}(\theta) \mathbb{E}_{\hat{\theta}_n} \left[g_{\nu,k,\ell,i,j} \left(X_0^i, X_0^j, \tilde{X}_k, \tilde{X}_\ell \right) \middle| Y_{1:T} \right] , \tag{1.11}$$

où les fonctions déterministes $(f_{\nu,k,\ell,i,j})$ et $(g_{\nu,k,\ell,i,j})$ sont définies explicitement dans le Chapitre 5. L'étape de maximisation peut être réalisée par exemple à l'aide de l'algorithme du gradient conjugué.

Il ne reste alors plus qu'à estimer les espérances conditionnelles grâce à l'approximation de la loi de lissage de $\left(X_0^i, X_0^j, \tilde{X}_k, \tilde{X}_\ell\right), 1 \leq i,j \leq p, 1 \leq k,\ell \leq T$, ce qui devient difficile lorsque T est grand (voir Chapitre 2). Ce problème peut être résolu en utilisant un algorithme EM de type bloc. Le *Maximum Split Data Likelihood Estimate* (MSDLE) introduit par Rydén [1994] consiste intuitivement à découper les observations de telle sorte à former des blocs de taille fixe considérés comme indépendants, puis à maximiser la vraisemblance résultante. En d'autres termes, la log-vraisemblance $\log p_\theta(Y_{1:T})$ est remplacée (mais non approchée, il ne s'agit en effet que d'un changement de fonction de contraste) par la fonction de contraste $\sum_{u=0}^{B-1} \log p_\theta(Y_{u\eta+1:(u+1)\eta})$ où B est le nombre de blocs et $\eta = T/B$ est la taille de ces blocs. Bien sûr cet estimateur n'est pas accessible en pratique pour le modèle considéré, et nous l'approchons donc par un algorithme EM où la quantité définie en (1.11) est remplacée par

$$L_{n,\theta}^{B,T}(Y_{1:T}) = \sum_{u=0}^{B-1} \ell_{n,\theta}^{u\eta+1,(u+1)\eta} \left(Y_{u\eta+1:(u+1)\eta} \right) . \tag{1.12}$$

L'étape E nécessite d'approcher la loi de lissage de $\left(X_0, \tilde{X}_{1:T}\right)$ sachant $Y_{1:T}$. Les algorithmes de lissage particulaire connus (voir Chapitre 2) ne pourraient être utilisés que si $\tilde{X}$ était une chaîne de Markov (on se retrouverait alors bien dans un modèle de Markov caché). Ceci n'étant pas le cas, il faut la remplacer par une nouvelle variable cachée plus adéquate $\bar{X}$ provenant de la décimation dans le temps de X définie pour tout $k \geq 0$ par $\bar{X}_k = X_{pk}$ et on définit aussi les observations vectorielles associées pour tout $k \geq 1$ par $\underline{Y}_k = [Y_{p(k-1)+1}, \cdots, Y_{pk}]^T$. $\bar{X}$ est une chaîne de Markov cachée non dégénérée, et au temps k, l'observation $\underline{Y}_k$ dépend de $(\bar{X}_{k-1}, \bar{X}_k)$ (voir Figure 1.4).

De plus, $\tilde{X}$ peut être exprimée en fonction de $\bar{X}$ et l'étape E peut être réalisée avec un algorithme de lissage appliqué au modèle $(\bar{X}, \underline{Y})$.

En pratique, l'algorithme ainsi construit est capable d'estimer jusqu'à trois échelles de temps. De nombreux résultats numériques sont présentés dans le Chapitre 5. Nous avons simulé par exemple des observations $Y_{1:T}$ pour $p = 3$ et appliqué notre algorithme pour retrouver les paramètres utilisés. La convergence de l'algorithme est montrée dans la Figure 1.5. La même efficacité se retrouve pour $p \in \{1,2\}$.

Cette méthode de calibrage a aussi été appliquée à des données de marché d'actions : les indices CAC 40 et Dow Jones. Dans le cas particulier du CAC 40, un critère de sélection du nombre d'échelles p détaillé dans le Chapitre 5 a permis de conclure que deux facteurs sont nécessaires et suffisants pour modéliser ces données. Les paramètres estimés sont donnés dans la Table 1.1.

α_1	α_2	σ_1	σ_2	β	ρ
0.00	0.98	0.30	0.19	0.53	-0.58

TABLE 1.1 – Paramètres estimés pour le CAC 40

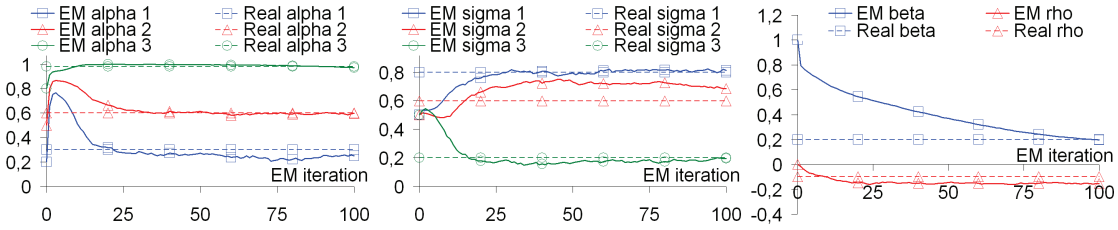


FIGURE 1.5 – Paramètres estimés à partir de données simulées avec $p = 3$

Conclusion L'étude de ce modèle en particulier nous conforte dans l'idée que l'inférence de paramètres à l'aide de l'algorithme EM peut nécessiter d'approcher la loi jointe de lissage, pas seulement la marginale. Ceci est lié à la chaîne de Markov cachée du modèle initial. En effet, X étant dégénérée, nous avons dû la remplacer par une autre variable cachée $\tilde{X}$ non dégénérée mais non markovienne. Elle nous a permis à la fois d'écrire la densité jointe et de séparer les étapes E et M mais son caractère non markovien a fait apparaître la loi jointe de tous les couples $\{(\tilde{X}_k, \tilde{X}_\ell)\}_{k\leq\ell}$, ce qui justifie l'intérêt porté dans cette thèse aux méthodes de Monte-Carlo séquentielles pour l'approximation des lois de lissage présentées dans le Chapitre 2.

Un autre enseignement que nous pouvons tirer de cet exemple est une méthode de construction d'un estimateur de paramètres via l'EM. Il nous a fallu ici réfléchir à quelle variable cachée était la mieux adaptée. A priori aucun des processus cachés auquel on pense naturellement ne présente toutes les qualités requises : la chaîne de Markov cachée initiale est dégénérée et donc ne donne pas accès au calcul des densités, la chaîne de Markov après décimation $\bar{X}$ ne permet pas de séparer les étapes E et M, et le processus $\tilde{X}$ n'est pas une chaîne de Markov et ne permet pas d'utiliser les algorithmes de lissage particulaires. Cependant, nous avons remarqué $\log p_\theta(Y_{1:T}, X_0, \tilde{X}_{1:T})$ pouvait se mettre sous la forme

$$\log p_\theta(Y_{1:T}, X_0, \tilde{X}_{1:T}) = \sum_i f_i(\theta) g_i(X_0, \tilde{X}_{1:T}) \,,$$

avec $X_0, \tilde{X}_{1:T}$ pouvant s'écrire comme $(X_0, \tilde{X}_{1:T}) = h(\theta, \bar{X}_{0:T})$. La quantité à maximiser est donc

$$\hat{\theta}_{n+1} = \operatorname{argmax}_{\theta\in\Theta_p} \sum_i f_i(\theta) \mathbb{E}_{\hat{\theta}_n}\left[g_i(X_0, \tilde{X}_{1:T}) \middle| Y_{1:T} \right] = \operatorname{argmax}_{\theta\in\Theta_p} \sum_i f_i(\theta) \mathbb{E}_{\hat{\theta}_n}\left[g_i(h(\hat{\theta}_n, \bar{X}_{0:T})) \middle| Y_{1:T} \right] .$$

La variable $\tilde{X}$ n'a donc joué qu'un rôle intermédiaire dans la création de l'algorithme d'inférence. Elle permet uniquement de séparer les étapes E et M puis nous la remplaçons par une variable plus adéquate pour l'estimation des espérances.

Chapitre 2

Lissage particulaire

Sommaire

Comme vu précédemment, il est indispensable de savoir approcher la loi de lissage. Les algorithmes de lissage particulaire permettent d'effectuer cette tâche via une population de N particules pondérées $\{(\xi_t^i, \omega_t^i)\}_{i=1}^N$, $0 \leq t \leq T$. La loi marginale est alors approchée par

$$\phi_{t|T}^{N,\text{algo}} \stackrel{\text{def}}{=} \left(\sum_{i=1}^N \omega_t^i \right)^{-1} \sum_{i=1}^N \omega_t^i \delta_{\xi_t^i} ,$$

où algo désignera l'algorithme utilisé. Dans le cas du Filter-Smoother, FFBS et FFBSi, nous obtiendrons également une estimation de la loi jointe.

2.1 Filter-Smoother

2.1.1 Algorithme

Le *Filter-Smoother*, connu aussi sous le nom de *path-space method* ou *genealogical tree*, est l'algorithme le plus simple. Il consiste à utiliser la généalogie des particules générées dans le cadre du filtre particulaire auxiliaire. La première étape sera donc d'approcher séquentiellement les distributions de filtrage $\phi_{t|t}$ grâce à N particules pondérées $\{(\tilde{\xi}_t^i, \tilde{\omega}_t^i)\}_{i=1}^N$ par $\phi_{t|t}^{N,\text{F-S}} \stackrel{\text{def}}{=} \left(\sum_{i=1}^N \tilde{\omega}_t^i\right)^{-1} \sum_{i=1}^N \tilde{\omega}_t^i \delta_{\tilde{\xi}_t^i}$. Les particules initiales $\{\tilde{\xi}_0^i\}_{i=1}^N$ sont des variables aléatoires iid telles que $\tilde{\xi}_0^1 \sim \rho_0$ où ρ_0 est une distribution auxiliaire. On pose alors $\tilde{\omega}_0^i \stackrel{\text{def}}{=} d\chi/d\rho_0(\tilde{\xi}_0^i) g(\tilde{\xi}_0^i, y_0)$. Le filtre $\phi_{0|0}$ est alors ciblé par $\phi_{0|0}^{N,\text{F-S}}$. Le filtre auxiliaire procède ensuite récursivement. Supposons que nous ayons un échantillon pondéré $\{(\tilde{\xi}_{t-1}^i, \tilde{\omega}_{t-1}^i)\}_{i=1}^N$ ciblant

$\phi_{t-1|t-1}$. On veut alors simuler de nouvelles particules selon la cible $\phi_{t|t}^{N,\mathrm{F-S,c}}$ définie pour toute fonction h mesurable sur $\mathbb{X}$ par

$$\phi_{t|t}^{N,\mathrm{F-S,c}}[h] \stackrel{\mathrm{def}}{=} \frac{\phi_{t-1|t-1}^{N,\mathrm{F-S}}[Mg(\cdot,y_t)h]}{\phi_{t-1|t-1}^{N,\mathrm{F-S}}[Mg(\cdot,y_t)]} .$$

D'après [Pitt and Shephard, 1999], ceci peut être fait en considérant la distribution cible auxiliaire suivante

$$\phi_{t|t}^{N,\mathrm{F-S,a}}(i,h) \stackrel{\mathrm{def}}{=} \frac{\tilde{\omega}_{t-1}^i M(\tilde{\xi}_{t-1}^i, g(\cdot,y_t)h)}{\sum_{\ell=1}^N \tilde{\omega}_{t-1}^\ell M(\tilde{\xi}_{t-1}^\ell, g(\cdot,y_t)h)} ,$$

sur l'espace produit $\{1,\dots,N\}\times\mathbb{X}$ muni de la tribu $\mathcal{P}(\{1,\dots,N\})\otimes\mathcal{X}$. Par construction, $\phi_{t|t}^{N,\mathrm{F-S,c}}$ est la distribution marginale de $\phi_{t|t}^{N,\mathrm{F-S,a}}(i,h)$ prise par rapport à l'indice de la particule. Il est donc possible d'approcher la distribution cible $\phi_{t|t}^{N,\mathrm{F-S,c}}$ en simulant selon la distribution auxiliaire et en oubliant les indices. Plus précisément, on simule d'abord des paires $\{(I_{t-1}^i, \tilde{\xi}_t^i)\}_{i=1}^N$ d'indices et de particules selon la distribution instrumentale suivante :

$$\pi_{t|t}(i,h) \propto \tilde{\omega}_{t-1}^i \vartheta_t(\tilde{\xi}_{t-1}^i) P_t(\tilde{\xi}_{t-1}^i, h) ,$$

sur l'espace produit $\{1,\dots,N\}\times\mathbb{X}$ où $\{\vartheta_t\}$ sont appelés *poids multiplicateurs d'ajustement* et P_t est un noyau de transition markovien proposé. Par souci de simplicité, nous supposons que $P_t(x,\cdot)$, pour tout $x\in\mathbb{X}$, admet une densité $p_t(x,\cdot)$ par rapport à la mesure de référence λ qui domine M. On associe à chaque tirage le poids d'importance

$$\tilde{\omega}_t^i \stackrel{\mathrm{def}}{=} \frac{m(\tilde{\xi}_{t-1}^{I_{t-1}^i}, \tilde{\xi}_t^i) g(\tilde{\xi}_t^i, y_t)}{\vartheta_t(\tilde{\xi}_{t-1}^{I_{t-1}^i}) p_t(\tilde{\xi}_{t-1}^{I_{t-1}^i}, \tilde{\xi}_t^i)} ,$$

de telle sorte que $\tilde{\omega}_t^i \propto \mathrm{d}\phi_{t|t}^{N,\mathrm{F-S,a}} / \mathrm{d}\pi_{t|t}(I_{t-1}^i, \tilde{\xi}_t^i)$. Finalement, le système de particules $\{(\tilde{\xi}_t^i, \tilde{\omega}_t^i)\}_{i=1}^N$ cible $\phi_{t|t}$. Cette procédure s'implémente par l'algorithme 2.

Algorithm 2 Filtre auxiliaire

1: Simuler $(\tilde{\xi}_0^i)_{i=1}^N$ iid selon la loi instrumentale ρ_0.
2: $(\tilde{\omega}_0^i)_{i=1}^N \leftarrow \left(\chi(\tilde{\xi}_0^i) g(\tilde{\xi}_0^i, y_0) / \rho_0(\tilde{\xi}_0^i)\right)_{i=1}^N$.
3: **for** t from 1 to T **do**
4: Simuler $I_{t-1}^{1:N}$ multinomialement avec des probabilités proportionnelles à $(\tilde{\omega}_{t-1}^i \vartheta_t(\tilde{\xi}_{t-1}^i))_{i=1}^N$.
5: Simuler $\tilde{\xi}_t^i$ indépendamment pour tout i selon $P_t(\tilde{\xi}_{t-1}^{I_{t-1}^i}, \cdot)$.
6: $\tilde{\omega}_t^{1:N} \leftarrow \left(\frac{m(\tilde{\xi}_{t-1}^{I_t^i}, \tilde{\xi}_t^i) g(\tilde{\xi}_t^i, y_t)}{\vartheta_t(\tilde{\xi}_{t-1}^{I_t^i}) p_t(\tilde{\xi}_{t-1}^{I_t^i}, \tilde{\xi}_t^i)} \right)_{i=1}^N$.
7: **end for**

Les indices $\{I_t^i\}_{i=1:N}^{t=0:T-1}$ sont alors utilisés pour reconstruire la généalogie des particules (Algorithme 3) et nous obtenons ainsi N chemins $\{\xi_t^i\}_{t=0}^T$, pour $1\le i\le N$, qui, associés aux poids $\omega_T^i \stackrel{\mathrm{def}}{=} \tilde{\omega}_T^i$, approchent $\phi_{0:T|T}(h)$ pour toute fonction mesurable h sur $\mathbb{X}^{T+1}$ par

$$\phi_{0:T|T}^{N,\mathrm{F-S}}(h) \stackrel{\mathrm{def}}{=} \left(\sum_{i=1}^N \omega_T^i\right)^{-1} \sum_{i=1}^N \omega_T^i h(\xi_{0:T}^i) .$$

Algorithm 3 Filter-Smoother

1: Simuler $(\tilde{\xi}_{0:T}^{1:N}, I_{0:T-1}^{1:N}, \tilde{\omega}_{0:T}^{1:N})$ grâce au filtre auxiliaire (Algorithme 2).
2: $(\omega_T^{1:N}) \leftarrow (\tilde{\omega}_T^{1:N})$.
3: Pour tout $1 \leq i \leq N$, $J_T^i \leftarrow i$ et $(\xi_T^i) \leftarrow (\tilde{\xi}_T^{J_T^i})$.
4: **for** t from $T-1$ down to 0 **do**
5: Pour tout $1 \leq i \leq N$, $J_t^i \leftarrow I_t^{J_{t+1}^i}$ et $(\xi_t^i) \leftarrow (\tilde{\xi}_t^{J_t^i})$.
6: **end for**

2.1.2 Propriétés

Le choix le plus simple est celui de l'algorithme appelé *bootstrap filter* proposé par [Gordon et al., 1993]. Il consiste à poser pour tout $x \in \mathbb{X}$, $\vartheta_t(x) = 1$ et $p_t(x,\cdot) = m(x,\cdot)$. Un choix plus attrayant mais souvent très coûteux en calculs, est le *fully adapted filter*. Il s'agit de poser $\vartheta_t^*(x) \stackrel{\text{def}}{=} \int m(x,x')g(x',y_t)\mathrm{d}x'$, pour $x \in \mathbb{X}$ et $p_t^*(x,x') \stackrel{\text{def}}{=} m(x,x')g(x',y_t)/\vartheta_t^*(x)$, pour $(x,x') \in \mathbb{X} \times \mathbb{X}$. D'autres choix sont discutés par exemple dans [Douc et al., 2009] et [Cornebise et al., 2008].

Cet algorithme est donc facile à mettre en place et il fournit une approximation de la loi jointe de lissage avec une complexité linéaire en le nombre de particules N. En revanche, les étapes de sélection (simulation des indices $\{I_t^i\}_{i=1:N}^{t=0:T-1}$) sont connues pour diminuer progressivement la diversité des particules aux temps petits devant l'horizon T. Ce phénomène, discuté par exemple dans Doucet et al. [2010], connu sous le nom de *dégénérescence* est clairement visible sur la Figure 2.1 qui représente 50 chemins de particules pour un horizon de temps de 100 dans le modèle de volatilité stochastique présenté dans l'Exemple 1.2. La dégénérescence est telle que les premières particules sont toutes les mêmes.

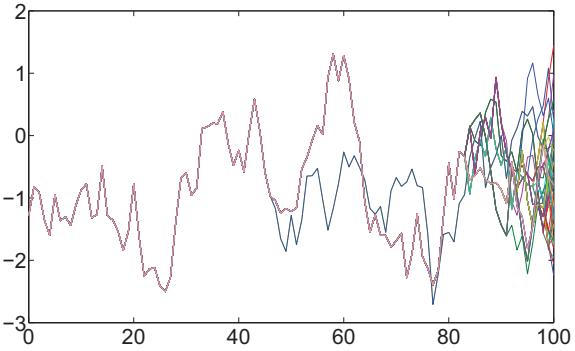


FIGURE 2.1 – Dégénérescence du Filter-Smoother

Cet inconvénient est aussi visible dans la variance asymptotique de l'estimateur (voir Briers et al. [2010]). En particulier, dans le cas indépendant où pour tout $(x,x') \in \mathbb{X} \times \mathbb{X}$, $m(x,x') = m(x')$ et $p_t(x,x') = m(x')$, la variance asymptotique du Filter-Smoother pour une fonction linéaire h sur $\mathbb{X}^{T+1}$ de la forme

$$\forall x \in \mathbb{X}^{T+1}, \quad h(x) = \sum_{t=0}^{T} h_t(x_t),$$

où les $h_{0:T}$ sont des fonctions bornées et mesurables sur $\mathbb{X}$, est donnée dans Del Moral [2004] par :

$$\Gamma_{0:T|T}[S_{T,0}] \stackrel{\text{def}}{=} \sum_{t=0}^{T-1} \frac{m(g_T^2)}{m(g_T)^2} \frac{m(g_t[h_t - \phi_t(h_t)]^2)}{m(g_t)} + \frac{m(g_T^2[h_T - \phi_T(h_T)]^2)}{m(g_T)^2} + \sum_{t=1}^{T-1} \left\{ \sum_{s=0}^{t-1} \frac{m(g_t^2)}{m(g_t)^2} \frac{m(g_s[h_s - \phi_s(h_s)]^2)}{m(g_s)} + \frac{m(g_t^2[h_t - \phi_t(h_t)]^2)}{m(g_t)^2} \right\} + \frac{\chi(g_0^2[h_0 - \phi_0(h_0)]^2)}{\chi(g_0)^2}.$$

Conclusion On constate donc que la variance du Filter-Smoother est d'ordre T^2/N. Il est conjecturé (et prouvé dans certains cas précis) que cet ordre reste valide plus généralement pour des noyaux markoviens non dégénérés. C'est pourquoi d'autres algorithmes ont été construits afin de régler ce problème de dégénérescence (voir les Sections suivantes).

2.2 Two-Filter

2.2.1 Version quadratique

La loi marginale de lissage peut également être obtenue en combinant les résultats du filtre particulaire décrit dans la Section 2.1 évoluant dans le sens du temps et d'un autre filtre connu sous le nom de *Backward Information Filter* (introduit par Mayne [1966]) évoluant dans le sens inverse du temps et proportionnel à la distribution de la chaîne cachée conditionnellement aux observations futures.

Soit $(\gamma_t)_{t\geq 0}$ une famille de fonctions mesurables positives telles que, pour tout $t \in \{0,\dots,T\}$,

$$\int \gamma_t(x_t)\mathrm{d}x_t \left[\prod_{u=t+1}^{T} g(x_{u-1},y_{u-1})M(x_{u-1},\mathrm{d}x_u)\right] g(x_T,y_T) < \infty\,. \tag{2.1}$$

Alors, d'après Briers et al. [2010], le Backward Information Filter peut être défini pour toute fonction mesurable h sur $\mathbb{X}$ par

$$\psi_{\gamma,t|T}(h) \stackrel{\text{def}}{=} \frac{\int\cdots\int \gamma_t(x_t)\mathrm{d}x_t \left[\prod_{u=t+1}^{T} g(x_{u-1},y_{u-1})M(x_{u-1},\mathrm{d}x_u)\right] g(x_T,y_T)h(x_t)}{\int\cdots\int \gamma_t(x_t)\mathrm{d}x_t \left[\prod_{u=t+1}^{T} g(x_{u-1},y_{u-1})M(x_{u-1},\mathrm{d}x_u)\right] g(x_T,y_T)}\,.$$

Dans Isard and Blake [1998] et Kitagawa [1996], une méthode de Monte-Carlo séquentielle est développée en supposant implicitement que (2.1) est vérifiée pour $\gamma_t \equiv 1$. Dans le cas où γ_t est une distribution, comme proposé par Briers et al. [2010] et Fearnhead et al. [2010], alors $\psi_{\gamma,t|T}$ est la distribution de X_t conditionnellement à $Y_{t:T} = y_{t:T}$ si la distribution a priori de X_t est de densité γ_t. Cependant l'algorithme du Two-Filter ne nécessite pas que γ_t soit la densité d'une distribution (donc $\int \gamma_t(x_t)\mathrm{d}x_t$ peut ne pas être finie) et il n'est pas non plus nécessaire de supposer que $\gamma_t(x_t) = \int m(x_{t-1},x_t)\gamma_{t-1}(x_{t-1})\mathrm{d}x_{t-1}$. Les seules conditions sur γ_t pour que $\psi_{\gamma,t|T}$ puisse être approché par une population de particles sont la relation (2.1) et le fait que γ_t soit calculable explicitement. La distribution marginale de lissage au temps t peut alors être obtenue en combinant la loi de filtrage au temps t et le Backward Information Filter au temps $t+1$ en remarquant que pour toute fonction mesurable h sur $\mathbb{X}$:

$$\phi_{t|0:T}(h) \propto \int\int \phi_{t|t}(\mathrm{d}x_t)m(x_t,x_{t+1})\frac{\psi_{\gamma,t+1|T}(\mathrm{d}x_{t+1})}{\gamma_{t+1}(x_{t+1})}h(x_t)\,.$$

On suppose que $\phi_{t|t}$ est approchée grâce à N particules pondérées $\{(\tilde{\xi}_t^i,\tilde{\omega}_t^i)\}_{i=1}^N$ par

$$\phi_{t|t}^{N,\mathrm{F-S}} = \left(\sum_{i=1}^{N}\tilde{\omega}_t^i\right)^{-1}\sum_{i=1}^{N}\tilde{\omega}_t^i\delta_{\tilde{\xi}_t^i}\,,$$

comme décrit dans la Section 2.1. Nous allons aussi décrire ci-dessous une méthode similaire permettant d'approcher $\psi_{\gamma,t|T}$ grâce à N particules pondérées $\{(\check{\xi}_t^i,\check{\omega}_t^i)\}_{i=1}^N$ par $\left(\sum_{i=1}^N\check{\omega}_t^i\right)^{-1}\sum_{i=1}^N\check{\omega}_t^i\delta_{\check{\xi}_t^i}$. Alors $\phi_{t|0:T}(h)$ peut être approchée par

$$\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\tilde{\omega}_t^i\check{\omega}_{t+1}^j\frac{m(\tilde{\xi}_t^i,\check{\xi}_{t+1}^j)}{\gamma_{t+1}(\check{\xi}_{t+1}^j)}\right)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\tilde{\omega}_t^i\check{\omega}_{t+1}^j\frac{m(\tilde{\xi}_t^i,\check{\xi}_{t+1}^j)}{\gamma_{t+1}(\check{\xi}_{t+1}^j)}h(\tilde{\xi}_t^i)\,.$$

Il ne reste alors plus qu'à décrire la procédure d'approximation du Backward Information Filter en partant de T jusqu'à 0. Pour cela, remarquons que pour $t = T-1,\dots,0$, on a pour toute fonction mesurable h sur $\mathbb{X}$

$$\psi_{\gamma,t|T}(h) \propto \int\int \psi_{\gamma,t+1|T}(\mathrm{d}x_{t+1})\left[\gamma_t(x_t)g(x_t,y_t)\frac{m(x_t,x_{t+1})}{\gamma_{t+1}(x_{t+1})}\right]h(x_t)\mathrm{d}x_t\,. \tag{2.2}$$

L'équation (2.2) est analogue à celle du filtre, et une approximation particulaire du Backward Information Filter peut être obtenue de manière similaire (voir Algorithme 4) : supposons que $\{(\check{\xi}_{t+1}^i, \check{\omega}_{t+1}^i)\}_{i=1}^N$ approche le Backward Information Filter par $\psi_{\gamma,t+1|T}^N(\mathrm{d}x) \propto \sum_{i=1}^N \check{\omega}_{t+1}^i \delta_{\check{\xi}_{t+1}^i}(\mathrm{d}x)$, alors (2.2) mène à la distribution cible

$$\psi_{\gamma,t|T}^{N,\mathrm{c}}(\mathrm{d}x_t) \propto \sum_{i=1}^N \check{\omega}_{t+1}^i \left[\gamma_t(x_t) g(x_t, y_t) \frac{m(x_t, \check{\xi}_{t+1}^i)}{\gamma_{t+1}(\check{\xi}_{t+1}^i)} \right] \mathrm{d}x_t \ ,$$

qui est la marginale par rapport à x_t de la densité auxiliaire

$$\psi_{\gamma,t|T}^{N,\mathrm{a}}(i, x_t) \propto \frac{\check{\omega}_{t+1}^i}{\gamma_{t+1}(\check{\xi}_{t+1}^i)} \gamma_t(x_t) g(x_t, y_t) m(x_t, \check{\xi}_{t+1}^i)$$

Une approximation particulaire du Backward Information Filter au temps t peut alors en découler en choisissant un poids d'ajustement $\vartheta_{t|T}$ and un noyau instrumental $p_{t|T}$, puis en simulant $\{(I^i, \check{\xi}_t^i)\}_{i=1}^N$ à partir de la distribution instrumentale

$$\pi_{t|T}(i, x_t) \propto \frac{\check{\omega}_{t+1}^i \vartheta_{t|T}(\check{\xi}_{t+1}^i)}{\gamma_{t+1}(\check{\xi}_{t+1}^i)} p_{t|T}(\check{\xi}_{t+1}^i, x_t) \ .$$

Les particules sont alors associées aux poids d'importance

$$\check{\omega}_t^i = \frac{\gamma_t(\check{\xi}_t^i) g(\check{\xi}_t^i, y_t) m(\check{\xi}_t^i, \check{\xi}_{t+1}^{I^i})}{\vartheta_{t|T}(\check{\xi}_{t+1}^{I^i}) p_{t|T}(\check{\xi}_{t+1}^{I^i}, \check{\xi}_t^i)} \ .$$

Algorithm 4 Backward Information Filter

1: Simuler $(\check{\xi}_T^i)_{i=1}^N$ iid selon la loi instrumentale ρ_T.
2: $(\check{\omega}_T^i)_{i=1}^N \leftarrow \left(\gamma_T(\check{\xi}_T^i) g(\check{\xi}_T^i, y_T) / \rho_T(\check{\xi}_T^i) \right)_{i=1}^N$.
3: **for** t from $T-1$ down to 0 **do**
4: Simuler $I^{1:N}$ multinomialement avec des probabilités proportionnelles à $\left(\frac{\check{\omega}_{t+1}^i \vartheta_{t|T}(\check{\xi}_{t+1}^i)}{\gamma_{t+1}(\check{\xi}_{t+1}^i)} \right)_{i=1}^N$.
5: Simuler $\check{\xi}_t^i$ indépendamment pour tout i selon $P_{t|T}(\check{\xi}_{t+1}^{I^i}, \cdot)$.
6: $\check{\omega}_t^{1:N} \leftarrow \left(\frac{\gamma_t(\check{\xi}_t^i) g(\check{\xi}_t^i, y_t) m(\check{\xi}_t^i, \check{\xi}_{t+1}^{I^i})}{\vartheta_{t|T}(\check{\xi}_{t+1}^{I^i}) p_{t|T}(\check{\xi}_{t+1}^{I^i}, \check{\xi}_t^i)} \right)_{i=1}^N$.
7: **end for**

2.2.2 Version linéaire

Le défaut de l'estimateur précédent est sa complexité d'ordre $O(N^2)$. Fearnhead et al. [2010] ont contourné ce problème en combinant la loi de filtrage et le Backward Information Filter de manière différente. Ainsi, pour toute fonction mesurable h sur $\mathbb{X}$, on a

$$\phi_{t|0:T}(h) \propto \int\int\int \phi_{t-1|t-1}(\mathrm{d}x_{t-1}) \left[m(x_{t-1}, x_t) g(x_t, y_t) \frac{m(x_t, x_{t+1})}{\gamma_{t+1}(x_{t+1})} \right] \psi_{\gamma,t+1|T}(\mathrm{d}x_{t+1}) h(x_t) \mathrm{d}x_t \ .$$

Cette expression montre que la loi de lissage peut être approchée par la distribution cible suivante :

$$\phi_{t|T}^{N,\mathrm{T-F,c}}(\mathrm{d}x_t) \propto \sum_{i=1}^N \sum_{j=1}^N \frac{\tilde{\omega}_{t-1}^i \check{\omega}_{t+1}^j}{\gamma_{t+1}(\check{\xi}_{t+1}^j)} m(\check{\xi}_{t-1}^i, x_t) g(x_t, y_t) m(x_t, \check{\xi}_{t+1}^j) \mathrm{d}x_t \ .$$

En suivant la construction du filtre particulaire, on introduit une variable auxiliaire (I,J), correspondant à une paire d'indices sélectionnés, et on définit la distribution auxiliaire

$$\phi_{t|T}^{N,\mathrm{T-F,a}}(i,j,x_t) \propto \frac{\tilde{\omega}_{t-1}^i \check{\omega}_{t+1}^j}{\gamma_{t+1}(\check{\xi}_{t+1}^j)} m(\tilde{\xi}_{t-1}^i, x_t) g(x_t, y_t) m(x_t, \check{\xi}_{t+1}^j) \mathrm{d}x_t \ ,$$

sur l'espace produit $\{1,\ldots,N\}^2 \times \mathbb{X}$. $\phi_{t|T}^{N,\mathrm{T-F,c}}$ étant la distribution marginale de $\phi_{t|T}^{N,\mathrm{T-F,a}}$ par rapport à x_t, il est possible de simuler $\phi_{t|T}^{N,\mathrm{T-F,c}}$ en tirant un ensemble $\{(I^\ell, J^\ell, \xi_t^\ell)\}_{\ell=1}^N$ d'indices et de particules selon la distribution instrumentale

$$\pi_{t|T}(i,j,h) \propto \frac{\tilde{\omega}_{t-1}^i \vartheta_{t|T}(\tilde{\xi}_{t-1}^i, \check{\xi}_{t+1}^j) \check{\omega}_{t+1}^j}{\gamma_{t+1}(\check{\xi}_{t+1}^j)} P_{t|T}(\tilde{\xi}_{t-1}^i, \check{\xi}_{t+1}^j, h) \ ,$$

où $\vartheta_{t|T}$ est un poids d'ajustement et $P_{t|T}$ est un noyau de transition instrumental. On associe alors à chaque tirage $(I^\ell, J^\ell, \xi_t^\ell)$ le poids d'importance

$$\omega_t^\ell = \frac{m(\tilde{\xi}_{t-1}^{I^\ell}, \xi_t^\ell) g(\xi_t^\ell, y_t) m(\xi_t^\ell, \check{\xi}_{t+1}^{J^\ell})}{\vartheta_{t|T}(\tilde{\xi}_{t-1}^{I^\ell}, \check{\xi}_{t+1}^{J^\ell}) p_{t|T}(\tilde{\xi}_{t-1}^{I^\ell}, \check{\xi}_{t+1}^{J^\ell}, \xi_t^\ell)} \ .$$

On supprime ensuite les indices auxiliaires $\{(I^\ell, J^\ell)\}_{\ell=1}^N$ et les particules pondérées $\{(\xi_t^\ell, \omega_t^\ell)\}_{\ell=1}^N$ approchent la distribution marginale de lissage pour toute fonction mesurable h sur $\mathbb{X}$ par

$$\phi_{t|T}^{N,\mathrm{T-F}}(h) \stackrel{\mathrm{def}}{=} \left(\sum_{\ell=1}^N \omega_t^\ell \right)^{-1} \sum_{\ell=1}^N \omega_t^\ell h(\xi_t^\ell) \ .$$

Dans le cas général, cet algorithme reste de complexité $O(N^2)$ de par la simulation des indices auxiliaires $\{(I^\ell, J^\ell)\}_{\ell=1}^N$ avec une probabilité proportionnelle à $\frac{\tilde{\omega}_{t-1}^i \vartheta_{t|T}(\tilde{\xi}_{t-1}^i, \check{\xi}_{t+1}^j) \check{\omega}_{t+1}^j}{\gamma_{t+1}(\check{\xi}_{t+1}^j)}$. En choisissant le poids d'ajustement de telle sorte qu'il puisse être décomposé en $\vartheta_{t|T}(\tilde{x}, \check{x}) = \tilde{\vartheta}_t(\tilde{x}) \check{\vartheta}_{t|T}(\check{x})$, les indices auxiliaires peuvent être simulés indépendamment avec des probabilités respectivement proportionnelles à $\tilde{\omega}_{t-1}^i \tilde{\vartheta}_t(\tilde{\xi}_{t-1}^i)$ et $\frac{\check{\omega}_{t+1}^j \check{\vartheta}_{t|T}(\check{\xi}_{t+1}^j)}{\gamma_{t+1}(\check{\xi}_{t+1}^j)}$ ce qui conduit bien à un algorithme de complexité linéaire (voir Algorithme 5).

Algorithm 5 Linear Two-Filter

1: Simuler $(\tilde{\xi}_{0:T}^{1:N}, \tilde{\omega}_{0:T}^{1:N})$ avec le filtre auxiliaire (Algorithme 2).
2: Simuler $(\check{\xi}_{0:T}^{1:N}, \check{\omega}_{0:T}^{1:N})$ avec le Backward Information Filter (Algorithme 4).
3: Pour $1 \leq i \leq N$, $\xi_0^i \leftarrow \check{\xi}_0^i$ et $\omega_0^i \leftarrow \check{\omega}_0^i \chi(\check{\xi}_0^i) / \gamma_0(\check{\xi}_0^i)$.
4: **for** t from 1 to $T-1$ **do**
5: Simuler $I^{1:N}$ multinomialement avec des probabilités proportionnelles à $\left(\tilde{\omega}_{t-1}^i \tilde{\vartheta}_t(\tilde{\xi}_{t-1}^i) \right)_{i=1}^N$.
6: Simuler $J^{1:N}$ multinomialement avec des probabilités proportionnelles à $\left(\frac{\check{\omega}_{t+1}^j \check{\vartheta}_{t|T}(\check{\xi}_{t+1}^j)}{\gamma_{t+1}(\check{\xi}_{t+1}^j)} \right)_{j=1}^N$.
7: Simuler ξ_t^i indépendamment pour tout i selon $P_{t|T}(\tilde{\xi}_{t-1}^{I^i}, \check{\xi}_{t+1}^{J^i}, \cdot)$.
8: $\omega_t^{1:N} \leftarrow \left(\frac{m(\tilde{\xi}_{t-1}^{I^i}, \xi_t^i) g(\xi_t^i, y_t) m(\xi_t^i, \check{\xi}_{t+1}^{J^i})}{\tilde{\vartheta}_t(\tilde{\xi}_{t-1}^{I^i}) \check{\vartheta}_{t|T}(\check{\xi}_{t+1}^{J^i}) p_{t|T}(\tilde{\xi}_{t-1}^{I^i}, \check{\xi}_{t+1}^{J^i}, \xi_t^i)} \right)_{i=1}^N$.
9: **end for**
10: $\xi_T^{1:N} \leftarrow \tilde{\xi}_T^{1:N}$ et $\omega_T^{1:N} \leftarrow \tilde{\omega}_T^{1:N}$.

Conclusion C'est donc cet algorithme linéaire qui est comparé aux autres algorithmes de lissage dans le Chapitre 6. Il est en pratique très rapide à exécuter et élimine le problème de dégénérescence du Filter-Smoother comme le montre la Figure 2.3.b de la Section 2.4. On peut même étendre cet algorithme pour obtenir la loi de lissage de tous les triplets consécutifs (en conservant les particules sélectionnées en $t-1$ et $t+1$ pour générer celle en t) mais dans certains cas comme celui présenté dans l'exemple de la Sous-section 1.2.4, ceci ne suffit pas et la loi jointe peut être approchée avec les algorithmes FFBS et FFBSi présentés dans la Section 2.3 suivante.

2.3 FFBS/FFBSi

2.3.1 Algorithme

Le *Forward Filtering Backward Smoothing* et le *Forward Filtering Backward Simulation* s'appuient sur une première passe forward qui consiste à approcher la distribution de filtrage à l'aide de particules pondérées $\{(\tilde{\xi}_t^i, \tilde{\omega}_t^i)\}_{i=1:N}^{t=0:T}$ comme décrit dans la Section 2.1 puis à lui ajouter une passe backward pour obtenir une approximation de la loi de lissage. Cette idée a été développée par Hürzeler and Künsch [1998], Doucet et al. [2000]. Pour cela, introduisons B_η, appelé *Backward kernel*, défini pour toute mesure η sur $\mathcal{X}$, toute fonction mesurable h sur $\mathbb{X}$ et tout $x \in \mathbb{X}$ par

$$B_\eta(x,h) \stackrel{\text{def}}{=} \frac{\int \eta(\mathrm{d}x')m(x',x)h(x')}{\int \eta(\mathrm{d}x')m(x',x)} . \tag{2.3}$$

La distribution de lissage $\phi_{t:T|T}$ peut alors s'écrire pour tout $0 \leq t \leq T-1$ et toute fonction mesurable h sur $\mathbb{X}^{T-t+1}$ de la façon suivante :

$$\phi_{t:T|T}(h) = \int \cdots \int \phi_{T|T}(\mathrm{d}x_T) B_{\phi_{T-1|T-1}}(x_T, \mathrm{d}x_{T-1}) \ldots B_{\phi_{t|t}}(x_{t+1}, x_t) h(x_{t:T}) . \tag{2.4}$$

En conséquence, la loi jointe de lissage peut être calculée récursivement dans le sens inverse du temps par la formule

$$\phi_{t:T|T}(h) = \int \cdots \int B_{\phi_{t|t}}(x_{t+1}, x_t) \phi_{t+1|T}(\mathrm{d}x_{t+1:T}) h(x_{t:T}) . \tag{2.5}$$

Forward Filtering Backward Smoothing

La décomposition (2.4) suggère d'approcher la loi de lissage en remplaçant $\phi_{s|s}$ par son approximation particulaire $\phi_{s|s}^{N,\mathrm{F-S}}$:

$$\phi_{t:T|T}^{N,\mathrm{FFBS}}(h) \stackrel{\text{def}}{=} \int \cdots \int \phi_{T|T}^{N,\mathrm{F-S}}(\mathrm{d}x_T) B_{\phi_{T-1|T-1}^{N,\mathrm{F-S}}}(x_T, \mathrm{d}x_{T-1}) \ldots B_{\phi_{t|t}^{N,\mathrm{F-S}}}(x_{t+1}, x_t) h(x_{t:T}) .$$

De plus, par définition, pour toute fonction mesurable h sur $\mathbb{X}$

$$B_{\phi_{s|s}^{N,\mathrm{F-S}}}(x,h) = \sum_{i=1}^{N} \frac{\tilde{\omega}_s^i m(\tilde{\xi}_s^i, x)}{\sum_{\ell=1}^N \tilde{\omega}_s^\ell m(\tilde{\xi}_s^\ell, x)} h(\tilde{\xi}_s^i) ,$$

d'où l'on déduit

$$\phi_{t:T|T}^{N,\mathrm{FFBS}}(h) = \sum_{i_t=1}^{N} \cdots \sum_{i_T=1}^{N} \left(\prod_{s=t+1}^{T} \frac{\tilde{\omega}_{s-1}^{i_{s-1}} m(\tilde{\xi}_{s-1}^{i_{s-1}} \tilde{\xi}_s^{i_s})}{\sum_{\ell=1}^N \tilde{\omega}_{s-1}^\ell m(\tilde{\xi}_{s-1}^\ell \tilde{\xi}_s^{i_s})} \right) \frac{\tilde{\omega}_T^{i_T}}{\sum_{\ell=1}^N \tilde{\omega}_T^\ell} h(\tilde{\xi}_t^{i_t}, \ldots, \tilde{\xi}_T^{i_T}) , \tag{2.6}$$

pour toute fonction mesurable h sur $\mathbb{X}^{T-t+1}$. Cet estimateur souffre d'une complexité exponentielle en $O(N^{T-t+1})$ et il n'est donc pas implémentable en pratique. Cependant, il permet d'approcher la loi marginale de lissage avec une complexité quadratique (Algorithme 6). En effet, l'équation (2.5) donne la relation récursive suivante pour toute fonction mesurable h sur $\mathbb{X}$:

$$\phi_{t|T}^{N,\mathrm{FFBS}}(h) = \int B_{\phi_{t|t}}^{N,\mathrm{F-S}}(x_{t+1}, \mathrm{d}x_t) \phi_{t+1|T}^{N,\mathrm{FFBS}}(\mathrm{d}x_{t+1}) h(x_t) .$$

En supposant alors que l'on a déjà obtenu une approximation de $\phi_{t+1|T}$ par $\phi_{t+1|T}^{N,\mathrm{FFBS}} = \left(\sum_{j=1}^N \omega_{t+1}^j\right)^{-1} \sum_{j=1}^N \omega_{t+1}^j \delta_{\xi_{t+1}^j}$, on obtient

$$\phi_{t|T}^{N,\mathrm{FFBS}}(h) \propto \sum_{i=1}^N \tilde{\omega}_t^i \sum_{j=1}^N \frac{\omega_{t+1}^j m(\tilde{\xi}_t^i, \xi_{t+1}^j)}{\sum_{\ell=1}^N \tilde{\omega}_t^\ell m(\tilde{\xi}_t^\ell, \xi_{t+1}^j)} h(\tilde{\xi}_t^i) \; .$$

On en déduit alors qu'en posant

$$\begin{aligned}
&\xi_t^i = \tilde{\xi}_t^i \; , \quad \forall 1 \leq i \leq N \; , \quad \forall 0 \leq t \leq T \; , \\
&\omega_T^i = \tilde{\omega}_T^i \; , \quad \forall 1 \leq i \leq N \; , \\
&\omega_t^i = \tilde{\omega}_t^i \sum_{j=1}^N \frac{\omega_{t+1}^j m(\tilde{\xi}_t^i, \xi_{t+1}^j)}{\sum_{\ell=1}^N \tilde{\omega}_t^\ell m(\tilde{\xi}_t^\ell, \xi_{t+1}^j)} \; , \quad \forall 1 \leq i \leq N \; , \quad 0 \leq t \leq T-1 \; ,
\end{aligned}$$

on obtient

$$\phi_{t|T}^{N,\mathrm{FFBS}}(h) = \left(\sum_{i=1}^N \omega_t^i\right)^{-1} \sum_{i=1}^N \omega_t^i h(\xi_t^i) \; .$$

Algorithm 6 FFBS marginal

1: Simuler $(\tilde{\xi}_{0:T}^{1:N}, \tilde{\omega}_{0:T}^{1:N})$ avec le filtre auxiliaire (Algorithme 2).
2: $\xi_T^{1:N} \leftarrow \tilde{\xi}_T^{1:N}$ et $\omega_T^{1:N} \leftarrow \tilde{\omega}_T^{1:N}$.
3: **for** t from $T-1$ down to 0 **do**
4: $\quad \xi_t^{1:N} \leftarrow \tilde{\xi}_t^{1:N}$.
5: $\quad$ Pour $1 \leq i \leq N$ $\omega_t^i \leftarrow \tilde{\omega}_t^i \sum_{j=1}^N \frac{\omega_{t+1}^j m(\tilde{\xi}_t^i, \xi_{t+1}^j)}{\sum_{\ell=1}^N \tilde{\omega}_t^\ell m(\tilde{\xi}_t^\ell, \xi_{t+1}^j)}$.
6: **end for**

De plus, le lissage d'une fonctionnelle additive de la forme $h(x_{0:T}) = \sum_{t=0}^T h_t(x_t)$ peut être approché dans le cadre du FFBS de manière forward (et donc sans calculer les poids en backward) grâce une méthode introduite par Del Moral et al. [2010a].

Forward Filtering Backward Simulation

L'estimateur de la loi jointe de lissage donné en (2.6) n'est pas implémentable mais peut être interprété différemment en remarquant que les poids normalisés impliqués dans le produit définissent une distribution de probabilité sur l'espace $\{1,\ldots,N\}^{T-t+1}$ des trajectoires associées à une chaîne de Markov inhomogène évoluant dans le sens inverse du temps. Considérons pour $0 \leq t \leq T-1$, la matrice de transition markovienne $\{\Lambda_t^N(i,j)\}_{i,j=1}^N$ définie pour tout $(i,j) \in \{1,\ldots,N\}^2$ par

$$\Lambda_t^N(i,j) \stackrel{\mathrm{def}}{=} \frac{\tilde{\omega}_t^j m(\tilde{\xi}_t^j, \tilde{\xi}_{t+1}^i)}{\sum_{\ell=1}^N \tilde{\omega}_t^\ell m(\tilde{\xi}_t^\ell, \tilde{\xi}_{t+1}^i)} \; , \tag{2.7}$$

et pour $0 \leq t \leq T$

$$\mathcal{F}_t^N \stackrel{\mathrm{def}}{=} \sigma\{Y_{0:T}, (\tilde{\xi}_s^i, \tilde{\omega}_s^i); 0 \leq s \leq t, 1 \leq i \leq N\} \; .$$

Les probabilités de transition définies en (2.7) engendrent une chaîne de Markov inhomogène $\{J_t\}_{t=0}^T$ évoluant dans le sens contraire du temps de la manière suivante. Au temps T, l'indice aléatoire J_T est distribué de telle sorte que la probabilité qu'il prenne la valeur i est proportionnelle à $\tilde{\omega}_T^i$. Au temps $t \leq T-1$ et conditionnellement à l'indice J_{t+1} et à $\mathcal{F}_T^N$, l'indice J_t est distribué de telle sorte que la probabilité qu'il prenne la valeur j soit donnée par $\Lambda_t^N(J_{t+1}, j)$. La distribution jointe de $J_{0:T}$ est alors donnée pour tout $j_{0:T} \in \{1,\ldots,N\}^{T+1}$ par

$$\mathbb{P}\left\{J_{0:T} = j_{0:T} \middle| \mathcal{F}_T^N\right\} = \frac{\tilde{\omega}_T^{j_T}}{\sum_{\ell=1}^N \tilde{\omega}_T^\ell} \Lambda_{T-1}^N(j_T, j_{T-1}) \cdots \Lambda_0^N(j_1, j_0) \; .$$

Il vient finalement que l'estimateur du FFBS donné en (2.6) peut s'écrire pour toute fonction mesurable h sur $\mathbb{X}^{T+1}$ avec l'espérance conditionnelle suivante :

$$\phi_{0:T|T}^{N,\mathrm{FFBS}} = \mathbb{E}\left[h(\tilde{\xi}_0^{J_0},\ldots,\tilde{\xi}_T^{J_T})\middle|\mathcal{F}_T^N\right] .$$

On en déduit la construction d'un estimateur qui consiste à simuler conditionnellement à $\mathcal{F}_T^N$, N trajectoires indépendantes $\{J_{0:T}^\ell\}_{\ell=1}^N$ de la chaîne de Markov inhomogène introduite précédemment puis à approcher la loi jointe de lissage $\phi_{0:T|T}(h)$ par

$$\phi_{0:T|T}^{N,\mathrm{FFBSi}}(h) \stackrel{\mathrm{def}}{=} N^{-1}\sum_{\ell=1}^N h(\tilde{\xi}_0^{J_0^\ell},\ldots,\tilde{\xi}_T^{J_T^\ell}) . \tag{2.8}$$

Cet estimateur introduit par Godsill et al. [2004] permet donc d'approcher la loi jointe de lissage avec une complexité en $O(N^2)$. Cette complexité peut cependant être rendue linéaire grâce à une méthode de rejet pour simuler les indices $\{J_{0:T}^\ell\}_{\ell=1}^N$ développée par Douc et al. [2010] (Algorithme 7) et le FFBSi résultant est alors donné par l'algorithme 8. La méthode de rejet suppose que le noyau de transition m est borné dans le sens où il existe $0<\sigma_+<\infty$ tel que pour tout $(x,x')\in\mathbb{X}^2$, $m(x,x')\le\sigma_+$. Alors $\Lambda_t^N(i,j)$ défini par (2.7) peut être majoré par

$$\Lambda_t^N(i,j) \le \frac{\sigma_+\sum_{\ell=1}^N\tilde{\omega}_t^\ell}{\sum_{\ell=1}^N\tilde{\omega}_t^\ell m(\tilde{\xi}_t^\ell,\tilde{\xi}_{t+1}^i)} \times \frac{\tilde{\omega}_t^j}{\sum_{\ell=1}^N\tilde{\omega}_t^\ell} ,$$

et l'algorithme 7 en découle. Il est à noter que pour garantir la linéarité, il est impératif d'avoir une méthode efficace de simulation multinomiale. Douc et al. [2010] en suggèrent une qui nécessite $O(n(1+\log(1+N/n)))$ opérations élémentaires pour simuler n variables aléatoires multinomiales parmi un ensemble de N valeurs.

Algorithm 7 Simulation linéaire des indices

$L \leftarrow (1,\ldots,N)$.
while L n'est pas vide **do**
  $n \leftarrow$ taille(L).
  Simuler $I_1,\ldots,I_n$ multinomialement avec des probabilités proportionnelles à $(\tilde{\omega}_t^i)_{i=1}^N$.
  Simuler $U_1,\ldots,U_n$ indépendamment et uniformément sur $[0,1]$.
  $nL \leftarrow \emptyset$.
  for k from 1 to n **do**
    if $U_k \le m\left(\tilde{\xi}_t^{I_k},\tilde{\xi}_{t+1}^{J_{t+1}^{L_k}}\right)$ **then**
      $J_t^{L_k} \leftarrow I_k$.
    else
      $nL \leftarrow nL \cup \{L_k\}$.
    end if
  end for
  $L \leftarrow nL$.
end while

La Figure 2.2 illustre bien la linéarité de cette version du FFBSi. Le temps CPU moyen empirique de 500 exécutions de l'algorithme y est représenté pour plusieurs nombres de particules dans le cas du modèle de volatilité stochastique (Exemple 1.2).

2.3.2 Propriétés

La passe backward de ces deux algorithmes rend leur analyse problématique, et elle n'a pu être réalisée que récemment. Dans Douc et al. [2010], il est montré que pour un horizon de temps T donné, la probabilité d'obtenir une erreur de Monte-Carlo excédant $\varepsilon>0$, donnée lorsque l'on remplace $\phi_{t:T|T}$ par $\phi_{t:T|T}^{N,\mathrm{FFBS}}$ ou

Algorithm 8 FFBSi linéaire

1: Simuler $(\tilde{\xi}_{0:T}^{1:N}, \tilde{\omega}_{0:T}^{1:N})$ grâce au filtre auxiliaire (Algorithme 2).
2: $(\omega_T^1, \ldots, \omega_T^N) \leftarrow (1/N, \ldots, 1/N)$.
3: Pour $1 \leq i \leq N$, simuler indépendamment et multinomialement J_T^i avec des probabilités proportionnelles à $\tilde{\omega}_T^i$.
4: $(\xi_T^i)_{i=1}^N \leftarrow (\tilde{\xi}_T^{J_T^i})_{i=1}^N$.
5: **for** t from $T-1$ down to 0 **do**
6: Pour tout $1 \leq i \leq N$, simuler J_t^i indépendamment et multinomialement avec les probabilités $\left(\Lambda_t^n(J_{t+1}^i, j) \stackrel{\text{def}}{=} \frac{\tilde{\omega}_t^j m(\tilde{\xi}_t^j, \tilde{\xi}_{t+1}^{J_{t+1}^i})}{\sum_{\ell=1}^N \tilde{\omega}_t^\ell m(\tilde{\xi}_t^\ell, \tilde{\xi}_{t+1}^{J_{t+1}^i})}\right)_{j=1}^N$ grâce à l'algorithme 7.
7: $(\xi_t^i)_{i=1}^N \leftarrow (\tilde{\xi}_t^{J_t^i})_{i=1}^N$.
8: **end for**

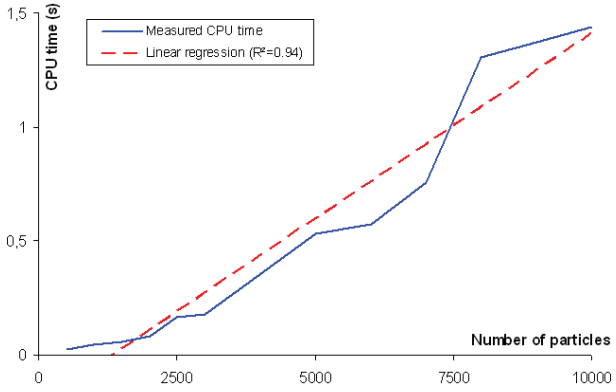


FIGURE 2.2 – Temps d'exécution du FFBSi en fonction du nombre de particules

$\phi_{t:T|T}^{N,\text{FFBSi}}$, est bornée par une quantité d'ordre $O(\exp(-cN\varepsilon^2))$ (*exponential deviation inequality*) où c est une constante positive dépendant de T et de la fonction lissée. Dans les mêmes conditions, un théorème central limite est établi de taux $\sqrt{N}$. La dépendance en T de la borne exponentielle et de la variance asymptotique est supprimée dans Douc et al. [2010] uniquement dans le cas de la loi marginale de lissage à condition que le noyau de la chaîne de Markov cachée soit *rapidement mélangeant*.

Comme détaillé dans l'introduction, on s'intéresse dans les cas les plus simples au lissage de fonctionnelles additives de la forme

$$S_{T,r}(x_{0:T}) = \sum_{t=r}^{T} h_t(x_{t-r:t}) \,, \tag{2.9}$$

où r est un entier positif ou nul et $\{h_t\}_{t=r}^T$ est une famille de fonctions bornées et mesurables de $\mathbb{X}$. En appliquant directement les résultats de Douc et al. [2010], l'erreur de Monte-Carlo serait bornée avec des termes en T^2/N. Cependant, dans Del Moral et al. [2010b], il est montré que la variance de l'estimateur FFBS d'une fonctionnelle additive est bornée par des termes ne dépendants de T et de N qu'à travers le ratio T/N mais dans Del Moral et al. [2010a], la norme L_q de cette même erreur pour $q > 2$ n'est bornée que par des termes en $T/\sqrt{N}$.

Dans l'article correspondant au Chapitre 4, nous établissons que l'erreur L_q est en fait bornée uniquement par des termes en T/N pour tout $q \geq 2$ et qu'il en est de même pour les inégalités exponentielles. Ces résultats s'appliquent également à des fonctionnelles quelconques (non additives) mais dépendent évidemment de l'échelle de la fonctionnelle. Plus précisément, plaçons-nous sous les hypothèses suivantes :

(A1)

(i) Il existe $(\sigma_-, \sigma_+) \in (0,\infty)^2$ tels que $\sigma_- < \sigma_+$ et pour tout $(x,x') \in \mathbb{X}^2$, $\sigma_- \leq m(x,x') \leq \sigma_+$ et on pose $\rho \stackrel{\text{def}}{=} 1 - \sigma_-/\sigma_+$.

(ii) Il existe $c_- \in \mathbb{R}_+^*$ tel que $\int \chi(\mathrm{d}x) g_0(x) \geq c_-$ et pour tout $t \in \mathbb{N}^*$, $\inf_{x\in\mathbb{X}} \int M(x,\mathrm{d}x') g_t(x') \geq c_-$.

(**A2**)

(i) Pour tout $t \geq 0$ et tout $x \in \mathbb{X}$, $g_t(x) > 0$.

(ii) $\sup_{t\geq 0} |g_t|_\infty < \infty$.

(**A3**) $\sup_{t\geq 1} |\vartheta_t|_\infty < \infty$, $\sup_{t\geq 0} |p_t|_\infty < \infty$ et $\sup_{t\geq 0} |\omega_t|_\infty < \infty$ où

$$\omega_0(x) \stackrel{\text{def}}{=} \frac{d\chi}{d\rho_0}(x) g_0(x), \quad \omega_t(x,x') \stackrel{\text{def}}{=} \frac{m(x,x') g_t(x')}{\vartheta_t(x) p_t(x,x')}, \forall t \geq 1 \, .$$

Enfin, définissons pour toute fonction bornée h d'un espace quelconque E dans $\mathbb{R}$ son oscillation $\mathrm{osc}(h)$ par

$$\mathrm{osc}(h) \stackrel{\text{def}}{=} \sup_{(z,z')\in \mathrm{E}^2} |h(z) - h(z')| \, .$$

Les deux principaux résultats obtenus dans le Chapitre 4 sont les Théorèmes 2.1 et 2.2 ci-dessous.

Théorème 2.1. *Supposons A1–3. Pour tout $q \geq 2$, il existe une constante C (dépendant uniquement de q, σ_-, σ_+, c_-, $\sup_{t\geq 1} |\vartheta_t|_\infty$ et $\sup_{t\geq 0} |\omega_t|_\infty$) telle que pour tout $T < \infty$, tout entier r et toutes fonctions bornées et mesurables $\{h_s\}_{s=r}^T$,*

$$\left\| \phi_{0:T|T}^{N,\mathrm{FFBS}}[S_{T,r}] - \phi_{0:T|T}[S_{T,r}] \right\|_q \leq \frac{C}{\sqrt{N}} \Upsilon_{r,T}^N \left(\sum_{s=r}^T \mathrm{osc}(h_s)^2 \right)^{1/2} ,$$

où $S_{T,r}$ est définie en (2.9) *et où*

$$\Upsilon_{r,T}^N \stackrel{\text{def}}{=} \sqrt{r+1} \left(\sqrt{1+r} \wedge \sqrt{T-r+1} + \frac{\sqrt{1+r}\sqrt{T-r+1}}{\sqrt{N}} \right) .$$

De manière similaire,

$$\left\| \phi_{0:T|T}^{N,\mathrm{FFBSi}}[S_{T,r}] - \phi_{0:T|T}[S_{T,r}] \right\|_q \leq \frac{C}{\sqrt{N}} \Upsilon_{r,T}^N \left(\sum_{s=r}^T \mathrm{osc}(h_s)^2 \right)^{1/2} .$$

Remarque 2.1. Dans le cas particulier où $r = 0$, $\Upsilon_{0,T}^N = 1 + \sqrt{T + 1/N}$. Le Théorème 2.1 donne alors

$$\left\| \phi_{0:T|T}^{N,\mathrm{FFBS}}[S_{T,0}] - \phi_{0:T|T}[S_{T,0}] \right\|_q \leq C \frac{\left(\sum_{s=0}^T \mathrm{osc}(h_s)^2\right)^{1/2}}{\sqrt{N}} \left(1 + \sqrt{\frac{T+1}{N}} \right) .$$

La preuve de ce théorème (complètement détaillée dans le Chapitre 4) consiste à décomposer l'erreur $\phi_{0:T|T}^{N,\mathrm{FFBS}}(S_{T,r}) - \phi_{0:T|T}(S_{T,r})$ en une somme d'erreurs locales qui permettent d'écrire pour tout $q \geq 2$

$$\phi_{0:T|T}^{N,\mathrm{FFBS}}(S_{T,r}) - \phi_{0:T|T}(S_{T,r}) = \sum_{t=0}^T D_{t,T}^N(S_{T,r}) + \sum_{t=0}^T C_{t,T}^N(S_{T,r}) \, , \tag{2.10}$$

où les termes $\{D_{t,T}^N(S_{T,r})\}_{t=0}^T$ sont des incréments de martingale donnant la borne en $\sqrt{(T+1)/N}$ et les termes $\{C_{t,T}^N(S_{T,r})\}_{t=0}^T$ sont des produits dont la norme L_q est bornée par $1/N$.

Cette même décomposition permet également d'obtenir des inégalités de déviations exponentielles. Le terme martingale est alors borné grâce l'inégalité de Azuma-Hoeffding tandis que le second requiert une égalité de type Hoeffding spécifique aux ratios de variables aléatoires.

Théorème 2.2. *Supposons A1–3. Il existe une constante C (dépendant uniquement de* σ_-, σ_+, c_-, $\sup_{t\geq 1}|\vartheta_t|_\infty$ *et* $\sup_{t\geq 0}|\omega_t|_\infty$*) telle que pour tout* $T<\infty$*, tout* $N\geq 1$*, tout* $\varepsilon>0$*, tout entier r, et toutes fonctions mesurables et bornées* $\{h_s\}_{s=r}^T$*,*

$$\mathbb{P}\left\{\left|\phi_{0:T|T}^{N,\mathrm{FFBS}}[S_{T,r}]-\phi_{0:T|T}[S_{T,r}]\right|>\varepsilon\right\} \leq 2\exp\left(-\frac{CN\varepsilon^2}{\Theta_{r,T}\sum_{s=r}^T \mathrm{osc}(h_s)^2}\right)+8\exp\left(-\frac{CN\varepsilon}{(1+r)\sum_{s=r}^T \mathrm{osc}(h_s)}\right),$$

où $S_{T,r}$ *est définie en* (2.9) *et où*

$$\Theta_{r,T}\stackrel{\mathrm{def}}{=}(1+r)\left\{(1+r)\wedge(T-r+1)\right\}. \tag{2.11}$$

De la même manière,

$$\mathbb{P}\left\{\left|\phi_{0:T|T}^{N,\mathrm{FFBSi}}[S_{T,r}]-\phi_{0:T|T}[S_{T,r}]\right|>\varepsilon\right\} \leq 4\exp\left(-\frac{CN\varepsilon^2}{\Theta_{r,T}\sum_{s=r}^T \mathrm{osc}(h_s)^2}\right)+8\exp\left(-\frac{CN\varepsilon}{(1+r)\sum_{s=r}^T \mathrm{osc}(h_s)}\right).$$

Conclusion Cette analyse théorique nous permet de penser que la version linéaire du FFBSi ne présente pas le problème de dégénérescence du Filter-Smoother. En effet, dans le cas du lissage de fonctionnelles additives, l'erreur d'approximation se comporte en T/N au lieu de T^2/N. Ceci se confirme numériquement (voir Figure 2.3). Les hypothèses sous lesquelles nous nous plaçons sont celles classiquement utilisées. La plus contraignante est sans doute la minoration du noyau de transition de la chaîne de Markov cachée. C'est elle qui permet d'utiliser la propriété d'oubli de la condition initiale, et elle est primordiale au bon déroulement de nos preuves. Il est souvent conjecturé que les résultats restent valables sous des conditions plus générales.

Aussi, malgré un coût en $O(N)$, cette même expérimentation numérique montre que le FFBSi reste très lent à exécuter. De plus, les particules générées lors de la passe forward ne tiennent pas compte des observations futures et pourtant la passe backward ne les replace pas pour les prendre en considération (seule la généalogie est modifiée). Nous avons donc exploité plus en-avant une méthode mise en exergue par Gilks and Berzuini [2001] qui consiste à resimuler les particules lors de plusieurs passes backward. Notre algorithme est présenté dans la Section 2.4 et plus en détails dans le Chapitre 6.

2.4 Ajout de passes MCMC

Parmi les méthodes de Monte-Carlo séquentielles décrites dans les Sections précédentes, la moins coûteuse en temps CPU pour approcher la loi jointe de lissage est le Filter-Smoother. Cependant, comme déjà mentionné, cet algorithme présente l'inconvénient majeur d'être fortement dégénéré dès que l'horizon de temps devient un peu trop grand (le nombre de particules nécessaires est de l'ordre de T^2 pour les fonctionnelles additives). En suivant l'idée de Gilks and Berzuini [2001], nous avons décrit et analysé (dans l'article correspondant au Chapitre 6) un algorithme consistant à ajouter des passes de MCMC à la population de particules produite par le Filter-Smoother. A la différence de Gilks and Berzuini [2001] où une passe de MCMC est ajoutée après chaque itération du filtre auxiliaire, nous nous fixons un horizon de temps T, appliquons l'algorithme du Filter-Smoother jusqu'au temps T puis utilisons plusieurs passes de MCMC. Ceci est plus efficace en temps de calcul car les particules replacées ne sont pas ensuite resélectionnées et leur diversité augmente progressivement.

2.4.1 Amélioration d'une population de particules

La loi cible à laquelle nous nous intéressons est $\phi_{0:T|T}$ définie par (1.2) dont on connaît la densité $\varphi_{0:T|T}$ à une constante près :

$$\varphi_{0:T|T}(x_{0:T}) \propto \chi(x_0) g(x_0, y_0) \prod_{t=1}^{T} M(x_{t-1}, \mathrm{d}x_t) g(x_t, y_t) \ ,$$

ce qui correspond exactement au cadre d'application de l'algorithme de *Metropolis-Hastings*. La chaîne de Markov résultante évolue dans un espace de grande dimension $\mathbb{X}^{T+1}$ et nécessite donc de choisir intelligemment un candidat à chaque itération de telle sorte que la probabilité d'acceptation ne soit pas trop proche de zéro. Dans ce cas de figure, la littérature MCMC propose l'utilisation de l'algorithme de *Gibbs*, et plus généralement de *Metropolis-within-Gibbs*, qui actualise chaque composante une par une. La convergence vers la loi cible sera alors d'autant plus rapide que l'algorithme est bien initialisé. L'idée est donc d'appliquer N algorithmes de Metropolis-within-Gibbs indépendants $(\xi^i_{0:T}[k])_{k\geq 0}$ pour $i \in \{1, \ldots, N\}$, initialisés avec les particules $\xi^i_{0:T}[0] = \xi^i_{0:T}$ données par le Filter-Smoother. Les composantes seront mises à jour dans le sens backward de T vers 0 afin de propager la bonne dispersion de $(\xi^i_T)_{i=1}^N$ à l'instant T vers les instants inférieurs de plus en plus dégénérés. L'estimateur résultant après K itérations MCMC s'écrit alors

$$\phi^{N,K,\mathrm{MH-IPS}}_{0:T|T} \stackrel{\mathrm{def}}{=} \left(\sum_{i=1}^{N} \omega^i_T \right)^{-1} \sum_{i=1}^{N} \omega^i_T \delta_{\xi^i_{0:T}[K]} \ . \tag{2.12}$$

Plus précisément, nous nous donnons une famille de densités de noyaux de transition $(r_t)_{0\leq t\leq T}$ telles que r_0 et r_T soient des densités de noyaux de transition sur $(\mathbb{X}, \mathcal{X})$ et pour tout $t \in \{1, \ldots, T-1\}$, r_t est la densité d'un noyau de transition sur $(\mathbb{X} \times \mathbb{X}, \mathcal{X})$. Pour $u, v, w, x \in \mathbb{X}$, nous définissons

$$\begin{aligned}
\alpha_0(v, w; x) &\stackrel{\mathrm{def}}{=} \frac{\chi(x) g(x, y_0) m(x, w)}{\chi(v) g(v, y_0) m(v, w)} \frac{r_0(w; v)}{r_0(w; x)} \wedge 1 \ , \\
\alpha_t(u, v, w; x) &\stackrel{\mathrm{def}}{=} \frac{m(u, x) g(x, y_t) m(x, w)}{m(u, v) g(v, y_t) m(v, w)} \frac{r_t(u, w; v)}{r_t(u, w; x)} \wedge 1 \ , \quad 1 \leq t \leq T-1 \ , \\
\alpha_T(u, v; x) &\stackrel{\mathrm{def}}{=} \frac{m(u, x) g(x, y_T)}{m(u, v) g(v, y_t)} \frac{r_T(u; v)}{r_T(u; x)} \wedge 1 \ .
\end{aligned}$$

A l'itération k, le nouveau chemin $\xi_{0:T}[k]$ est obtenu en actualisant dans le sens inverse du temps chaque composante $\xi_t[k]$ de la manière suivante :

(i) Simuler un candidat $X \sim r_t(\xi^i_{t-1}[k-1], \xi^i_{t+1}[k], \cdot)$,

(ii) Accepter $\xi^i_t[k] = X$ avec la probabilité $\alpha_t(\xi^i_{t-1}[k-1], \xi^i_{t+1}[k]; X)$,

(iii) Sinon, poser $\xi^i_t[k] = \xi^i_t[k-1]$.

Cette procédure, appelée *Metropolis-Hastings Improved Particle Smoother* (MH-IPS), est détaillée dans l'algorithme 9.

Lors des passes de MCMC, les particules n'interagissent plus entre elles ce qui permet d'éliminer le phénomène de dégénérescence au fur et à mesure que k augmente. Ceci est clairement confirmé par l'exemple numérique ci-dessous.

Exemple 2.1 (LGM)**.** Nous considérons une version simplifiée de l'exemple 1.1 :

$$X_{t+1} = \phi X_t + \sigma_U U_t \ , \quad Y_t = X_t + \sigma_V V_t \ ,$$

où $X_0 \sim \mathcal{N}\left(0, \sigma_U^2/(1-\phi^2)\right)$, $(U_t)_{t\geq 0}$ et $(V_t)_{t\geq 0}$ sont des suites de variables aléatoires indépendantes et iid selon la loi normale centrée réduite (et indépendante de X_0). Nous avons simulé $T+1 = 101$ observations en utilisant ce modèle avec les paramètres $\phi = 0.9$, $\sigma_U = 0.6$ et $\sigma_V = 1$. Dans ce cas précis, le fully adapted filter peut être implémenté facilement et il en est de même pour l'algorithme de Gibbs.

La diversité d'une population de particules à chaque instant t de chaque algorithme de lissage peut être mesurée par une estimation de l'*effective sample size* $N^{\mathrm{algo}}_{\mathrm{eff}}(t)$ comme défini dans Fearnhead et al.

Algorithm 9 MH-IPS

Initialisation
Simuler $(\xi^i_{0:T}, \omega^i_{0:T})_{i=1}^N$ par le Filter-Smoother (Algorithme 3).
Poser : $\forall\ 1 \leq i \leq N,\ \xi^i_{0:T}[0] = \xi^i_{0:T}$.
K passes d'amélioration
for k from 1 to K **do**
    for i from 1 to N **do**
        Simuler $X \sim r_T(\xi^i_{T-1}[k-1];\cdot)$,
        Accepter $\xi^i_T[k] = X$ avec la probabilité $\alpha_T(\xi^i_{T-1}[k-1], \xi^i_T[k-1], X)$,
        Sinon, poser $\xi^i_T[k] = \xi^i_T[k-1]$.
        for t from $T-1$ down to 1 **do**
            Simuler $X \sim r_t(\xi^i_{t-1}[k-1], \xi^i_{t+1}[k];\cdot)$,
            Accepter $\xi^i_t[k] = X$ avec la probabilité $\alpha_t(\xi^i_{t-1}[k-1], \xi^i_{t+1}[k], X)$,
            Sinon, poser $\xi^i_t[k] = \xi^i_t[k-1]$.
        end for
        Simuler $X \sim r_0(\xi^i_1[k];\cdot)$,
        Accepter $\xi^i_0[k] = X$ avec la probabilité $\alpha_0(\xi^i_0[k-1], \xi^i_1[k], X)$,
        Sinon, poser $\xi^i_0[k] = \xi^i_0[k-1]$.
    end for
end for

[2010]. En remarquant que $1/N = \mathbb{E}\left[(\bar{X}_N - \mu)^2/\sigma^2\right]$ où $X^{(1)}, \ldots, X^{(N)}$ sont iid telles que $\mathbb{E}\left[X^{(1)}\right] = \mu$, $\mathbb{V}\mathrm{ar}\left(X^{(1)}\right) = \sigma^2$ et $\bar{X}_N$ est leur moyenne empirique, on pose

$$N^{\mathrm{algo}}_{\mathrm{eff}}(t) \stackrel{\mathrm{def}}{=} \mathbb{E}\left[\left(\frac{\phi^{N,\mathrm{algo}}_{t|T}(\mathrm{Id}) - \mu_t}{\sigma_t}\right)^2\right]^{-1}, \tag{2.13}$$

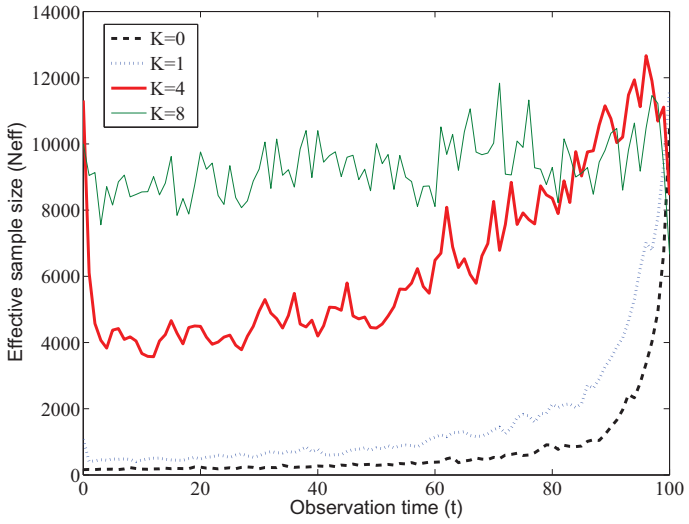
où Id est la fonction identité sur $\mathbb{R}$, μ_t et σ_t^2 sont la moyenne et la variance exactes de X_t sachant $Y_{0:T}$ obtenues par le Kalman smoother. On peut interpréter la quantité définie en (2.13) comme l'inverse de l'erreur quadratique renormalisée associée à l'estimation de $\mathbb{E}[X_t|Y_{0:T}]$. L'espérance impliquée dans (2.13) est approchée simplement par la moyenne empirique de 250 répétitions de chaque algorithme linéaire (Filter-Smoother, FFBSi, Two-Filter et MH-IPS) avec un nombre de particules choisi de telle sorte que le temps d'exécution de chaque algorithme soit le même. La comparaison de ces algorithmes est donc la plus équitable possible.

La figure 2.3.a confirme bien que la dégénérescence des particules diminue au fur et à mesure que K augmente, jusqu'à atteindre la même diversité pour tous les instants avec $K = 8$. La figure 2.3.b compare l'effective sample size des différents algorithmes de lissage linéaires. Comme attendu, le Filter-Smoother est très dégénéré pour les petites valeurs de t contrairement aux autres algorithmes. De plus, le MH-IPS surpasse clairement tous les autres pour un même temps d'exécution. La même analyse expérimentale a été conduite pour un modèle plus complexe dans le Chapitre 6 et les résultats restent en faveur du MH-IPS.

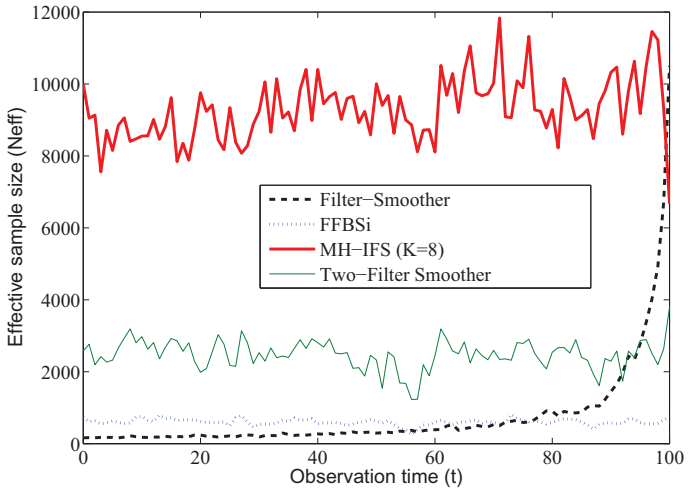
2.4.2 Analyse de la variance

La mise à jour composante par composante de l'algorithme 9 correspond à l'application d'un noyau de transition Markovien noté Q sur $(\mathbb{X}^{T+1}, \mathcal{X}^{\otimes(T+1)})$ admettant $\phi_{0:T|T}$ comme distribution invariante. Ainsi, pour tout $i \in \{1, \ldots, N\}$ on obtient

$$\begin{aligned} &\xi^i_{0:T}[0] = \xi^i_{0:T}\ , \\ &\xi^i_{0:T}[k+1] \sim Q(\xi^i_{0:T}[k], \cdot)\ , \quad k \geq 0\ . \end{aligned}$$



(a) Influence du nombre de passes K



(b) Comparaison de différents algorithmes de lissage

FIGURE 2.3 – Effective sample size moyen pour chaque étape de temps dans le LGM pour différents algorithmes de lissage à temps CPU donné.

On définit alors la norme en variation totale $||\cdot||_{\mathrm{TV}}$ par $||\mu||_{\mathrm{TV}} \stackrel{\mathrm{def}}{=} \sup_{|f|_\infty \leq 1} |\mu(f)|$ où $|f|_\infty \stackrel{\mathrm{def}}{=} \sup_{x\in\mathbb{X}} |f(x)|$ et on introduit l'hypothèse suivante

(**A4**) Pour tout $x \in \mathbb{X}^{T+1}$, $\lim_{k\to\infty} ||Q^k(x,\cdot) - \phi_{0:T|T}||_{\mathrm{TV}} = 0$.

Sous cette hypothèse, le Chapitre 6 montre que pour toute fonction mesurable et bornée h sur $\mathbb{X}$, lorsque le nombre d'itérations de l'algorithm MCMC tend vers l'infini, l'erreur quadratique

$$\mathbb{E}\left[\left(\phi_{0:T|T}^{N,k,\mathrm{MH-IPS}}(h) - \phi_{0:T|T}(h)\right)^2\right],$$

converge vers une limite qui est minimale lorsque tous les poids $(\omega_T^i)_{i=1}^N$ sont égaux. Cette situation peut être obtenue en sélectionnant les particules initiales $(\xi_T^i)_{i=1}^N$ de manière multinomiale selon des probabilités proportionnelles à $(\omega_T^i)_{i=1}^N$ avant d'appliquer l'algorithme de Metropolis-within-Gibbs. Nous supposerons

donc par la suite que cette étape a été effectuée et que tous les poids valent $1/N$. Dans ce cas, on montre que

$$\lim_{k\to\infty} \mathbb{E}\left[\left(\phi_{0:T|T}^{N,k,\mathrm{MH-IPS}}(h) - \phi_{0:T|T}(h)\right)^2\right] = \frac{\mathbb{V}\mathrm{ar}_{\phi_{0:T|T}}(h)}{N} ,$$

ce qui implique que pour N fixé et k tendant vers l'infini, l'estimateur MH-IPS ne sera pas meilleur que N trajectoires simulées de manière indépendante et exacte selon $\phi_{0:T|T}$. Une question importante est donc de donner un ordre de grandeur à k en fonction de N sans le laisser tendre vers l'infini mais en conservant cette propriété de N trajectoires simulées de manière indépendante et exacte selon $\phi_{0:T|T}$. Pour cela, nous introduisons des hypothèses supplémentaires avant d'énoncer le Théorème 2.3 démontré dans le Chapitre 6.

(**A5**) Il existe une fonction mesurable $V : \mathbb{X}^{T+1} \to [1,\infty($ telle que

(i) $\phi_{0:T|T}(V) < \infty$ et pour tout $x \in \mathbb{X}^{T+1}$ et tout $k \in \mathbb{N}$, $Q^k V(x) < \infty$,

(ii) il existe $\beta \in (0,1)$ tel que pour $h \in \mathcal{C}_V \stackrel{\text{def}}{=} \{h; |h/V|_\infty < \infty\}$ et tout $x \in \mathbb{X}^{T+1}$,

$$|Q^k h(x) - \phi_{0:T|T}(h)| < \beta^k V(x) ,$$

(iii) la suite $\{N^{-1}\sum_{i=1}^N V^2(\xi_{0:T}^i)\}_{N\geq 1}$ de variables aléatoires est bornée en probabilité.

Théorème 2.3. *Supposons (A5). Soit $(k_N)_{N\geq 1}$ une suite d'entiers telle que*

$$\lim_{N\to\infty} k_N + \ln N/(2\ln\beta) = \infty .$$

Alors, pour toute fonction $h \in \mathcal{C}_V$, on a le théorème central limite suivant

$$\sqrt{N}\left[\phi_{0:T|T}^{N,\mathrm{MH-IPS}}(h) - \phi_{0:T|T}(h)\right] \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \mathbb{V}\mathrm{ar}_{\phi_{0:T|T}}(h)\right) .$$

Ce TCL présente l'avantage indiscutable d'avoir une variance asymptotique très simple $\mathbb{V}\mathrm{ar}_{\phi_{0:T|T}}(h)$ qui peut être estimée à partir d'une seule population de particules. On peut donc obtenir un intervalle de confiance sans ré-appliquer une méthode de Monte-Carlo à l'estimateur lui-même, comme c'est le cas pour tous les lisseurs particulaires. Pour cela, il suffit d'ajouter par exemple $k_N = -\ln N/\ln\beta$ passes de MCMC.

Conclusion Nous en déduisons donc deux façons d'utiliser cet algorithme.

Il est possible de fixer arbitrairement le nombre d'itérations k pour améliorer la qualité d'une population de particules donnée mais sans chercher à obtenir des intervalles de confiance pour l'estimateur. L'algorithme est alors de complexité $O(N)$ comme montré dans la Figure 2.4 et permet d'approcher la loi de lissage numériquement de manière très efficace (voir Exemple 2.1).

Si la variance de l'estimateur doit être évaluée, on choisit $k_N \propto \ln N$ et on obtient un algorithme en $O(N\ln N)$. La figure 2.5 illustre bien le fait que dans ce cas une seule population de particules améliorée avec l'algorithme de Metropolis-within-Gibbs permet d'estimer la variance aussi bien que 250 estimateurs avec l'algorithme de Gibbs ou de Metropolis-within-Gibbs. Il s'agit d'un avantage supplémentaire à effectuer toutes les passes MCMC pour un horizon de temps fixé (contrairement à Gilks and Berzuini [2001]). C'est en effet cette caractéristique qui nous a permis d'établir le Théorème 2.3 faisant figurer une variance extrêmement simple à estimer. Nous obtenons ainsi le seul algorithme d'approximation de la loi de lissage (jointe ou marginale) incluant une estimation de la variance grâce à une seule population de particules. Une question légitime concerne le rôle du terme β défini dans l'Hypothèse 5. Il est possible de craindre en effet que cette constante décroisse exponentiellement avec l'horizon de temps et qu'ainsi le nombre k_N de passes MCMC soit très important. Nous montrons cependant numériquement que dans le cas de l'exemple 2.1, 8 passes suffisent et dans le Chapitre 6, un exemple plus complexe n'en nécessite que 4.

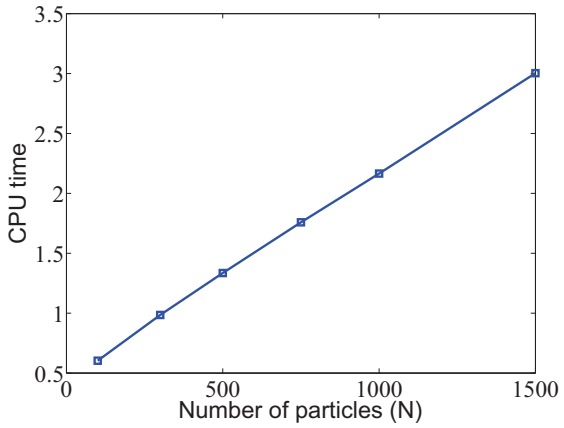


FIGURE 2.4 – Temps CPU moyen pour le lissage d'une fonctionnelle additive par l'algorithme MH-IPS dans le modèle de volatilité stochastique en fonction du nombre de particules.

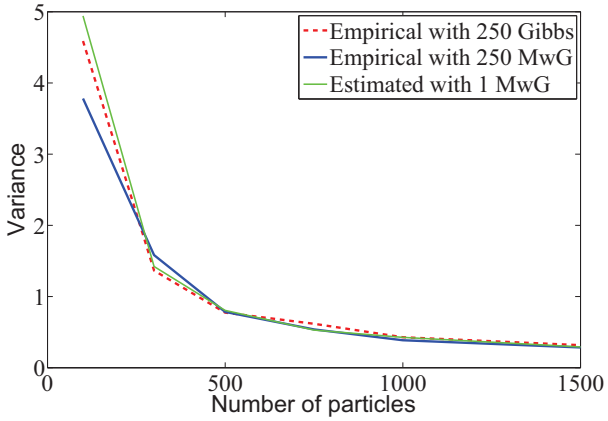


FIGURE 2.5 – Variance du MH-IPS en fonction du nombre de particules dans le modèle de volatilité stochastique.

Chapitre 3

Convergence des modèles d'îlots de Feynman-Kac

Sommaire

Les algorithmes présentés dans le Chapitre 2 peuvent se généraliser à des modèles non HMM où la fonction de vraisemblance g définie sur $\mathbb{X} \times \mathbb{Y}$ est remplacée simplement par une collection de fonctions $(g_n)_{n\geq 0}$ sur $\mathbb{X}$. Le champ d'application devient donc plus large que celui des HMM (voir Exemple A.1, Del Moral [2004] et Doucet et al. [2001]). Les notations de ce chapitre sont celles utilisées habituellement dans les modèles de Feynman-Kac et sont donc indépendantes de celles introduites dans le cas des HMM.

3.1 Introduction

L'approximation numérique des semi-groupes de Feynman-Kac par des systèmes de particules en interaction est un domaine de recherche très actif. Ces méthodes sont de plus en plus utilisées dans la simulation de distributions en grande dimension et en probabilité appliquée par exemple dans le filtrage et le lissage des HMM, l'inférence bayésienne, la biologie, ou encore la physique (voir par exemple Del Moral [2004], Del Moral and Doucet [2010-2011]).

Plus précisément, soit $(\mathbb{X}_n, \mathcal{X}_n)_{n\geq 0}$ une suite d'ensembles mesurables et désignons par $\mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$ l'espace de Banach des fonctions mesurables et bornées f sur $\mathbb{X}_n$, muni de la norme uniforme $\|f\| = \sup_{x_n \in \mathbb{X}_n} |f(x_n)|$. On considère également une suite de fonctions mesurables g_n appelées *fonctions potentielles* sur les espaces d'état $\mathbb{X}_n$, une distribution η_0 sur $\mathbb{X}_0$, et une suite de noyaux markoviens M_n de $(\mathbb{X}_{n-1}, \mathcal{X}_{n-1})$ dans $(\mathbb{X}_n, \mathcal{X}_n)$. On associe la suite de mesures de Feynman-Kac, définie pour toute fonction $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$ par l'équation

$$\eta_n(f_n) \stackrel{\text{def}}{=} \gamma_n(f_n)/\gamma_n(1) \quad \text{avec} \quad \gamma_n(f_n) \stackrel{\text{def}}{=} \int \eta_0(\mathrm{d}x_0) \prod_{0\leq p<n} g_p(x_p) M_{p+1}(x_p, \mathrm{d}x_{p+1}) f_n(x_n)\,, \tag{3.1}$$

à la paire potentiels/transitions (g_n, M_n).

Les suites de distributions $(\eta_n)_{n\geq 0}$ et $(\gamma_n)_{n\geq 0}$ sont séquentiellement approchées à l'aide d'un système de particules en interaction. Cet algorithme consiste à approcher pour tout $n \in \mathbb{N}$ la mesure de probabilité η_n par un ensemble de particules (ξ_n^i) générées itérativement. La procédure de mise à jour des particules se déroule en deux étapes. Dans l'étape de sélection, les particules sont d'abord resélectionnées selon des poids proportionnels aux fonctions potentielles. Dans l'étape de mutation, une nouvelle génération de particules (ξ_{n+1}^i) est générée à partir des particules sélectionnées grâce au noyau M_{n+1}. Le comportement asymptotique de cette approximation particulaire est maintenant bien compris (voir Del Moral [2004] et Del Moral et al. [2012]).

Récemment, le développement des méthodes de calcul en parallèle a conduit à l'étude de l'implémentation en parallèle de cette approximation particulaire. Dans ce chapitre, nous nous intéressons aux *modèles d'îlots de particules* qui consistent à exécuter N_2 systèmes de particules en interaction, chacun de taille N_1. Ce problème soulève plusieurs questions parmi lesquelles l'optimisation de la taille des îlots N_1 et du nombre d'îlots N_2 pour un coût d'exécution donné $N \stackrel{\text{def}}{=} N_1 N_2$, et l'opportunité de faire interagir les îlots.

Quand les N_2 îlots sont exécutés de manière indépendante, le biais généré par chacun d'entre eux ne dépend que de leur taille N_1. Il peut donc être intéressant d'introduire une interaction entre les îlots même si cette interaction augmente la variance de l'estimateur final à travers les étapes de resélection. Nous proposons ici d'étudier le biais et la variance asymptotiques dans chaque cas. Ainsi, pour N_1 et N_2 suffisamment grands et imposés, nous pouvons déterminer quel est le meilleur choix en terme d'erreur quadratique moyenne asymptotique.

Ce chapitre est organisé de la manière suivante. Dans la Section 3.2 nous décrivons les modèles de Feynman-Kac et leur approximation particulaire de taille N_1. Nous introduisons dans la Section 3.3 les modèles d'îlot de Feynman-Kac et l'algorithme d'îlots en interaction correspondant. Finalement le biais et la variance asymptotiques des îlots indépendants et en interaction sont donnés dans la Section 3.4.

3.2 Modèles de Feynman-Kac

3.2.1 Description du modèle

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. Tous les processus de ce chapitre seront définis sur cet espace. Soit $(X_n)_{n\geq 0}$ une chaîne de Markov non-homogène (définie par rapport à sa filtration naturelle) de distribution initiale η_0, et de noyaux de transition $(M_n)_{n\geq 1}$. On peut noter que pour tout $n \in \mathbb{N}$, $X_n \in \mathbb{X}_n$. En utilisant ces notations, nous pouvons réécrire γ_n défini en (3.1) de la manière suivante

$$\gamma_n(f_n) = \mathbb{E}\left[f_n(X_n) \prod_{0\leq p<n} g_p(X_p) \right] .$$

Description du semi-groupe Pour $\ell \in \mathbb{N}$, considérons le noyau Q_ℓ de $(\mathbb{X}_{\ell-1}, \mathcal{X}_{\ell-1})$ vers $(\mathbb{X}_\ell, \mathcal{X}_\ell)$ donné par

$$Q_\ell(x_{\ell-1}, A_\ell) \stackrel{\text{def}}{=} g_{\ell-1}(x_{\ell-1}) M_\ell(x_{\ell-1}, A_\ell) , \quad x_{\ell-1} \in \mathbb{X}_{\ell-1} , \quad A_\ell \in \mathcal{X}_\ell .$$

Pour $p \leq n$, définissons $Q_{p,n}$ comme le noyau de $(\mathbb{X}_p, \mathcal{X}_p)$ dans $(\mathbb{X}_n, \mathcal{X}_n)$ par l'équation

$$Q_{p,n} \stackrel{\text{def}}{=} Q_{p+1} Q_{p+2} \cdots Q_n .$$

Avec cette définition, on a $\gamma_n = \gamma_p Q_{p,n}$, et, pour tout $x_p \in \mathbb{X}_p$, $A_n \in \mathcal{X}_n$

$$Q_{p,n}(x_p, A_n) = \mathbb{E}\left[\mathbf{1}_{A_n}(X_n) \prod_{p\leq q<n} g_q(X_q) \middle| X_p = x_p \right] .$$

Définissons également $\mathcal{P}(\mathbb{X}_n, \mathcal{X}_n)$ comme étant l'ensemble des mesures de probabilité sur $(\mathbb{X}_n, \mathcal{X}_n)$. Il est alors facile de vérifier que la suite de mesures $(\eta_n)_{n\geq 0}$ satisfait la relation suivante

$$\eta_{n+1} = \Phi_{n+1}(\eta_n) , \tag{3.2}$$

où $\Phi_{n+1} : \mathcal{P}(\mathbb{X}_n, \mathcal{X}_n) \to \mathcal{P}(\mathbb{X}_{n+1}, \mathcal{X}_{n+1})$ est donné par

$$\Phi_{n+1}(\eta_n) \stackrel{\text{def}}{=} \Psi_n(\eta_n) M_{n+1} \,. \tag{3.3}$$

Dans l'équation ci-dessus, $\Psi_n : \mathcal{P}(\mathbb{X}_n, \mathcal{X}_n) \to \mathcal{P}(\mathbb{X}_n, \mathcal{X}_n)$ désigne la transformation de Boltzmann-Gibbs

$$\Psi_n(\eta_n)(A_n) \stackrel{\text{def}}{=} \frac{1}{\eta_n(g_n)} \int_{A_n} g_n(x_n)\, \eta_n(\mathrm{d}x_n) \,, \quad A_n \in \mathcal{X}_n \,.$$

Comme $\eta_n = \gamma_p Q_{p,n} / \gamma_p(Q_{p,n}(1))$ et $\eta_p = \gamma_p / \gamma_p(1)$, on obtient

$$\eta_n = \frac{\eta_p Q_{p,n}}{\eta_p(Q_{p,n}(1))} = \eta_p \bar{Q}_{p,n} \,, \quad \text{où} \quad \bar{Q}_{p,n} \stackrel{\text{def}}{=} \frac{Q_{p,n}}{\eta_p(Q_{p,n}(1))} = \frac{\gamma_p(1)}{\gamma_n(1)} Q_{p,n} \,.$$

Interprétation markovienne non-linéaire Soit S_{n,η_n} le noyau markovien sur $(\mathbb{X}_n, \mathcal{X}_n)$ donné pour $x_n \in \mathbb{X}_n$ et $A_n \in \mathcal{X}_n$ par

$$S_{n,\eta_n}(x_n, A_n) = \varepsilon_n\, g_n(x_n) \delta_{x_n}(A_n) + (1 - \varepsilon_n\, g_n(x_n))\, \Psi_n(\eta_n)(A_n) \,. \tag{3.4}$$

Il n'est pas difficile de vérifier que

$$\Psi_n(\eta_n) = \eta_n S_{n,\eta_n} \,. \tag{3.5}$$

Cette propriété reste valide pour tout choix de ε_n tel que $\varepsilon_n\, g_n \in [0,1]$. Par exemple, pour $\varepsilon_n = 0$, on a simplement

$$S_{n,\eta_n}(x_n, A_n) = \Psi_n(\eta_n)(A_n) \,, \quad x_n \in \mathbb{X}_n \,, \quad A_n \in \mathcal{X}_n \,.$$

Par (3.2), (3.3) et (3.5), on a

$$\eta_{n+1} = \eta_n K_{n+1,\eta_n} \quad \text{avec} \quad K_{n+1,\eta_n} \stackrel{\text{def}}{=} S_{n,\eta_n} M_{n+1} \,.$$

Modèle particulaire Nous décrivons maintenant plus précisément l'approximation particulaire des probabilités $(\eta_n)_{n\geq 0}$. Soit N_1 un entier. On définit le noyau markovien $\boldsymbol{M}_{n+1}$ de $(\mathbb{X}_n^{N_1}, \mathcal{X}_n^{\otimes N_1})$ dans $(\mathbb{X}_{n+1}^{N_1}, \mathcal{X}_{n+1}^{\otimes N_1})$ pour tout $(x_n^1, \ldots, x_n^{N_1}) \in \mathbb{X}_n^{N_1}$ et $(A_{n+1}^1, \ldots, A_{n+1}^{N_1}) \in \mathcal{X}_{n+1}^{N_1}$ par

$$\boldsymbol{M}_{n+1}(x_n^1, \ldots, x_n^{N_1}, A_{n+1}^1 \times \cdots \times A_{n+1}^{N_1}) \stackrel{\text{def}}{=} \prod_{1 \leq i \leq N_1} K_{n+1, m(x_n^1, \ldots, x_n^{N_1}, \cdot)}(x_n^i, A_{n+1}^i) \,, \tag{3.6}$$

où $m(x_n^1, \ldots, x_n^{N_1}, \cdot)$ désigne la mesure empirique des points $(x_n^1, \ldots, x_n^{N_1})$ donnée pour tout $A_n \in \mathcal{X}_n$ par

$$m(x_n^1, \ldots, x_n^{N_1}, A_n) \stackrel{\text{def}}{=} \frac{1}{N_1} \sum_{i=1}^{N_1} \delta_{x_n^i}(A_n) \,.$$

Pour $\varepsilon_n = 0$, on a $S_{n,\eta_n}(x_n, A_n) = \Psi_n(\eta_n)(A_n)$, ce qui implique que

$$K_{n+1, m(x_n^1, \ldots, x_n^{N_1}, \cdot)}(x_n^i, \cdot) = \Phi_{n+1}(m(x_n^1, \ldots, x_n^{N_1}, \cdot)) \,.$$

Nous définissons une chaîne de Markov $(\xi_n^{N_1})_{n\geq 0}$ où pour tout $n \in \mathbb{N}$, $\xi_n^{N_1} = (\xi_n^{N_1,1}, \ldots, \xi_n^{N_1,N_1}) \in \mathbb{X}_n^{N_1}$ de noyau de transition $\boldsymbol{M}_{n+1}$ et la filtration $\mathcal{F}_n^{N_1}$ associée à l'évolution des particules :

$$\mathcal{F}_n^{N_1} \stackrel{\text{def}}{=} \sigma\left(\xi_p^{N_1}, 0 \leq p \leq n\right) \,.$$

Les approximations particulaires des mesures η_n et γ_n sont définies pour $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$ par

$$\eta_n^{N_1}(f_n) \stackrel{\text{def}}{=} m(\xi_n^{N_1}, f_n) \quad \text{et} \quad \gamma_n^{N_1}(f_n) \stackrel{\text{def}}{=} \eta_n^{N_1}(f_n) \prod_{0 \leq p < n} \eta_p^{N_1}(g_p) \,. \tag{3.7}$$

Proposition 3.1. *L'estimateur $\gamma_n^{N_1}$ est non-biaisé : pour $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$*

$$\mathbb{E}\left[\eta_n^{N_1}(f_n)\prod_{0\le p<n}\eta_p^{N_1}(g_p)\right] = \mathbb{E}\left[f_n(X_n)\prod_{0\le p<n}g_p(X_p)\right]. \tag{3.8}$$

La preuve de cette Proposition repose sur le Lemme suivant.

Lemme 3.1. *Pour tout $p > 0$ et $f_p \in \mathcal{B}_b(\mathbb{X}_p, \mathcal{X}_p)$, on a*

$$\mathbb{E}\left[\eta_p^{N_1}(f_p)\middle|\mathcal{F}_{p-1}^{N_1}\right] = \frac{\eta_{p-1}^{N_1}(Q_p f_p)}{\eta_{p-1}^{N_1}(g_{p-1})}.$$

Démonstration. Par définition,

$$\mathbb{E}\left[\eta_p^{N_1}(f_p)\middle|\mathcal{F}_{p-1}^{N_1}\right] = \eta_{p-1}^{N_1}(K_{p,\eta_{p-1}^{N_1}} f_p),$$

où

$$\begin{aligned}
K_{p,\eta_{p-1}^{N_1}}(x_{p-1}, f_p) &= S_{p-1,\eta_{p-1}^{N_1}} M_p(x_{p-1}, f_p) = \int S_{p-1,\eta_{p-1}^{N_1}}(x_{p-1}, \mathrm{d}x'_{p-1}) M_p(x'_{p-1}, f_p) \\
&= \varepsilon_{p-1} g_{p-1}(x_{p-1}) M_p f_p(x_{p-1}) + (1 - \varepsilon_{p-1} g_{p-1}(x_{p-1}))\Psi_{p-1}(\eta_{p-1}^{N_1}) M_p f_p \\
&= \varepsilon_{p-1} Q_p f_p(x_{p-1}) + (1 - \varepsilon_{p-1} g_{p-1}(x_{p-1}))\Psi_{p-1}(\eta_{p-1}^{N_1}) M_p f_p,
\end{aligned}$$

impliquant

$$\mathbb{E}\left[\eta_p^{N_1}(f_p)\middle|\mathcal{F}_{p-1}^{N_1}\right] = \varepsilon_{p-1}\eta_{p-1}^{N_1}(g_{p-1})\frac{\eta_{p-1}^{N_1}(Q_p f_p)}{\eta_{p-1}^{N_1}(g_{p-1})} + (1 - \varepsilon_{p-1}\eta_{p-1}^{N_1}(g_{p-1}))\Psi_{p-1}(\eta_{p-1}^{N_1}) M_p f_p.$$

Finalement, la preuve est achevée en remarquant que

$$\Psi_{p-1}(\eta_{p-1}^{N_1}) M_p f_p = \frac{1}{N_1}\sum_{i=1}^{N_1}\frac{g_{p-1}(\xi_{p-1}^{N_1,i})}{\frac{1}{N_1}\sum_{j=1}^{N_1} g_{p-1}(\xi_{p-1}^{N_1,j})} M_p f_p(\xi_{p-1}^{N_1,i}) = \frac{\eta_{p-1}^{N_1}(g_{p-1} M_p f_p)}{\eta_{p-1}^{N_1}(g_{p-1})} = \frac{\eta_{p-1}^{N_1}(Q_p f_p)}{\eta_{p-1}^{N_1}(g_{p-1})}.$$

□

Preuve de la Proposition 3.1. Par la définition de $\gamma_n^{N_1}$ donnée en (3.7), on a

$$\mathbb{E}\left[\gamma_n^{N_1}(f_n)\right] = \mathbb{E}\left[\mathbb{E}\left[\eta_n^{N_1}(f_n)\middle|\mathcal{F}_{n-1}^{N_1}\right]\prod_{0\le p<n}\eta_p^{N_1}(g_p)\right].$$

Le Lemme 3.1 donne

$$\mathbb{E}\left[\eta_n^{N_1}(f_n)\middle|\mathcal{F}_{n-1}^{N_1}\right] = \frac{\eta_{n-1}^{N_1}(Q_n f_n)}{\eta_{n-1}^{N_1}(g_{n-1})},$$

et

$$\mathbb{E}\left[\gamma_n^{N_1}(f_n)\right] = \mathbb{E}\left[\eta_{n-1}^{N_1}(Q_n f_n)\prod_{0\le p<n-1}\eta_p^{N_1}(g_p)\right].$$

En itérant cette étape, on obtient

$$\begin{aligned}
\mathbb{E}\left[\gamma_n^{N_1}(f_n)\right] &= \mathbb{E}\left[\mathbb{E}\left[\eta_{n-1}^{N_1}(Q_n f_n)\middle|\mathcal{F}_{n-2}^{N_1}\right]\prod_{0\le p<n-1}\eta_p^{N_1}(g_p)\right] = \mathbb{E}\left[\eta_{n-2}^{N_1}(Q_{n-1}Q_n f_n)\prod_{0\le p<n-2}\eta_p^{N_1}(g_p)\right] \\
&= \mathbb{E}\left[\eta_0^{N_1}(Q_{0,n} f_n)\right] = \mathbb{E}\left[Q_{0,n} f_n(\xi_0^{N_1,1})\right] = \gamma_0 Q_{0,n} f_n = \gamma_n(f_n).
\end{aligned}$$

□

Exemple 3.1 (Call Barrière Up&Out)**.** Une application directe est le pricing d'options en finance. Le payoff d'un Call Barrière Up&Out de strike K, de barrière $H > K$, de maturité T, portant sur un actif S est défini par

$$\max(S_T - K, 0)\mathbf{1}_{A_H}\left(\max_{0\leq i\leq N_{\text{obs}}} S_{iT/N_{\text{obs}}}\right) ,$$

où $A_H \overset{\text{def}}{=} \{x; x \leq H\}$. Si les taux d'intérêt sont supposés constants et égaux à r, alors le prix de cette option est donné par

$$P = \mathrm{e}^{-rT}\mathbb{E}\left[\max(S_T - K, 0)\mathbf{1}_{A_H}\left(\max_{0\leq i\leq N_{\text{obs}}} S_{iT/N_{\text{obs}}}\right)\right] .$$

On se place dans le modèle de Black-Scholes :

$$\forall t \geq 0 , \quad S_t = S_0 \mathrm{e}^{(r-\sigma^2/2)t+\sigma W_t} ,$$

où $S_0 < H$ est la valeur initiale de l'actif, σ sa volatilité et le processus W est un mouvement brownien standard.

En posant $n = N_{\text{obs}}$, $X_p = S_{pT/N_{\text{obs}}}$, $g_p = \mathbf{1}_{A_H}$ et $f_n(x) = \max(x - K, 0)\mathbf{1}_{A_H}(x)$ on a

$$P = \mathrm{e}^{-rT}\gamma_n(f_n) .$$

L'avantage de l'estimateur $\gamma_n^{N_1}$ par rapport à celui du Monte-Carlo classique est que les trajectoires menant à un payoff nul ne sont pas abandonnées grâce à une resélection parmi celles qui n'ont toujours pas dépassé la barrière. Cette étape de resélection augmente le temps de calcul bien que les deux algorithmes soit de complexité linéaire par rapport au nombre de trajectoires simulées. Dans le cas où $K = 100$, $H = 110$, $T = 1$, $N_{\text{obs}} = 250$, $r = 0.05$, $S_0 = 100$ et $\sigma = 0.3$, la méthode de Monte-Carlo séquentielle est deux fois plus lente que la méthode de Monte-Carlo classique. Dans un souci d'équité nous avons donc utilisé deux fois moins de particules dans le cas du SMC. De plus, afin de pouvoir calculer des intervalles de confiance, les particules du SMC ont été réparties entre 150 exécutions indépendantes de l'algorithme SMC. La Figure 3.1 montre bien qu'à temps CPU donné, la méthode SMC donne un intervalle de confiance à 95% dont la demi-largeur en pourcentage du prix estimé est meilleure que celle de la méthode MC classique.

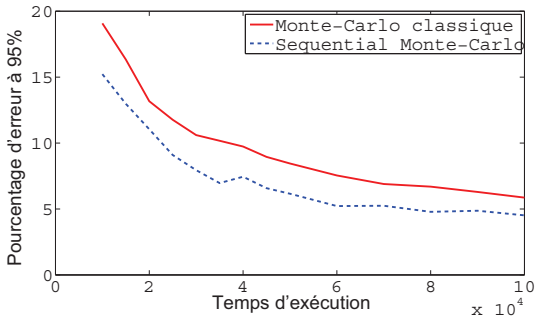


FIGURE 3.1 – Demi-largeur de l'intervalle de confiance des méthodes SMC et MC à temps CPU fixé.

3.2.2 Comportement asymptotique

L'analyse des fluctuations des mesures particulaires $\{\eta_n^{N_1}\}_{n\geq 0}$ autour de leurs valeurs limites $\{\eta_n\}_{n\geq 0}$ est essentiellement basée sur l'analyse asymptotique des erreurs de simulation locales associées aux étapes de sélection et mutation. Ces erreurs locales s'expriment comme des champs aléatoires $W_n^{N_1}$, donnés pour toute fonction $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$ par

$$W_n^{N_1}(f_n) = \sqrt{N_1}\left[\eta_n^{N_1}(f_n) - \eta_{n-1}^{N_1}(K_{n,\eta_{n-1}^{N_1}} f_n)\right] .$$

Le théorème central limite suivant pour les champs aléatoires locaux est essentiel. Sa preuve peut être trouvée dans Del Moral [2004] (Corollary 9.3.1, pp. 295-298).

Théorème 3.1. *Pour tout horizon de temps fixé $n \geq 1$, la suite $(W_p^{N_1})_{1\leq p\leq n}$ converge en loi, quand N_1 tend vers l'infini, vers une suite de n champs gaussiens indépendants et centrés $(W_p)_{1\leq p\leq n}$ tels que, pour tout $f_p, h_p \in \mathcal{B}_b(\mathbb{X}_p, \mathcal{X}_p)$, et $1 \leq p \leq n$,*

$$\mathbb{E}\left[W_p(f_p)W_p(h_p)\right] = \eta_{p-1}K_{p,\eta_{p-1}}([f_p - K_{p,\eta_{p-1}}(f_p)][h_p - K_{p,\eta_{p-1}}(h_p)]) \ ,$$

où, par convention, pour tout $x_{p-1} \in \mathbb{X}_{p-1}$

$$\begin{aligned} K_{p,\eta_{p-1}}(x_{p-1}, [f_p - K_{p,\eta_{p-1}}(f_p)][h_p - K_{p,\eta_{p-1}}(h_p)]) \\ &\stackrel{\text{def}}{=} K_{p,\eta_{p-1}}(x_{p-1}, [f_p - K_{p,\eta_{p-1}}(x_{p-1}, f_p)][h_p - K_{p,\eta_{p-1}}(x_{p-1}, h_p)(x_{p-1}]) \\ &= K_{p,\eta_{p-1}}(x_{p-1}, f_p h_p) - K_{p,\eta_{p-1}}(x_{p-1}, f_p)K_{p,\eta_{p-1}}(x_{p-1}, h_p) \ . \end{aligned}$$

Dans le cas particulier où $\varepsilon_p \equiv 0$, cette covariance devient

$$\mathbb{E}\left[W_p(f_p)W_p(h_p)\right] = \eta_p\left[(f_p - \eta_p(f_p))\,(g_p - \eta_p(g_p))\right] \ .$$

Nous considérons maintenant la décomposition de l'erreur globale en une somme d'erreurs locales :

$$\gamma_n^{N_1} - \gamma_n = \sum_{p=0}^{n}\left[\gamma_p^{N_1}Q_{p,n} - \gamma_{p-1}^{N_1}Q_{p-1,n}\right] \ ,$$

où pour $p = 0$ on a pris la convention $\gamma_{-1}^{N_1}Q_{-1,n} = \gamma_n$. Remarquons aussi que

$$\begin{aligned} \gamma_p^{N_1}Q_{p,n} - \gamma_{p-1}^{N_1}Q_{p-1,n} &= \gamma_p^{N_1}Q_{p,n} - \gamma_{p-1}^{N_1}Q_pQ_{p,n} \\ &= (\gamma_p^{N_1} - \gamma_{p-1}^{N_1}Q_p)Q_{p,n} = (\gamma_p^{N_1} - \gamma_{p-1}^{N_1}(g_{p-1})\eta_{p-1}^{N_1}K_{p,\eta_{p-1}^{N_1}})Q_{p,n} \\ &= (\gamma_p^{N_1} - \gamma_p^{N_1}(1)\eta_{p-1}^{N_1}K_{p,\eta_{p-1}^{N_1}})Q_{p,n} = \gamma_p^{N_1}(1)(\eta_p^{N_1} - \eta_{p-1}^{N_1}K_{p,\eta_{p-1}^{N_1}})Q_{p,n} \ , \end{aligned}$$

ce qui implique que pour toute fonction $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$,

$$W_n^{\gamma,N_1}(f_n) \stackrel{\text{def}}{=} \sqrt{N_1}\left[\gamma_n^{N_1} - \gamma_n\right](f_n) = \sum_{p=0}^{n}\gamma_p^{N_1}(1)W_p^{N_1}(Q_{p,n}f_n) \ . \tag{3.9}$$

Par le Theorème 7.4.2 dans Del Moral [2004] la suite aléatoire $(\gamma_p^{N_1}(1))_{0\leq p\leq n}$ converge en probabilité, quand N_1 tend vers l'infini, vers la suite déterministe $(\gamma_p(1))_{0\leq p\leq n}$. Une simple application du lemme de Slutsky implique que la suite de champs aléatoires W_n^{γ,N_1} converge en loi, quand N_1 tend vers l'infini, vers les champs gaussiens W_n^{γ} définis pour toute function $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$ par

$$W_n^{\gamma}(f_n) \stackrel{\text{def}}{=} \sum_{p=0}^{n}\gamma_p(1)W_p(Q_{p,n}f_n) = \gamma_n(1)\sum_{p=0}^{n}W_p(\bar{Q}_{p,n}f_n) \ . \tag{3.10}$$

De manière similaire, la suite de champs aléatoires

$$W_n^{\eta,N_1}(f_n) \stackrel{\text{def}}{=} \sqrt{N}\left[\eta_n^{N_1} - \eta_n\right](f_n) = \gamma_n^{N_1}(1)^{-1}W_n^{\gamma,N_1}\left(f_n - \eta_n(f_n)\right) \ , \tag{3.11}$$

converge en loi, quand N_1 tend vers l'infini, vers les champs gaussiens W_n^{η} définis pour toute fonction $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$ par

$$W_n^{\eta}(f_n) \stackrel{\text{def}}{=} \sum_{p=0}^{n}W_p(\bar{Q}_{p,n}(f_n - \eta_n(f_n))) \ , \tag{3.12}$$

et on obtient également le Lemme suivant.

Lemme 3.2. *Pour tout $f_n^1, f_n^2 \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$, la paire $(W_n^{\gamma,N_1}(f_n^1), W_n^{\eta,N_1}(f_n^2))$ converge en loi, quand N_1 tend vers l'infini, vers $(W_n^{\gamma}(f_n^1), W_n^{\eta}(f_n^2))$.*

Cette analyse permet de montrer le résultat suivant qui sera essentiel dans la preuve du Théorème 3.4.

Théorème 3.2. *Pour tout horizon de temps $n \geq 0$ et toute fonction $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$, on a*

$$\lim_{N_1 \to \infty} N_1 \mathbb{E}\left[\eta_n^{N_1}(f_n) - \eta_n(f_n)\right] = B_n(f_n) \,,$$
$$\lim_{N_1 \to \infty} N_1 \mathbb{V}\mathrm{ar}\left(\eta_n^{N_1}(f_n)\right) = V_n(f_n) \,,$$

où

$$B_n(f_n) \stackrel{\text{def}}{=} -\sum_{p=0}^{n} \mathbb{E}\left[W_p(\bar{Q}_{p,n}(1)) W_p(\bar{Q}_{p,n}(f_n - \eta_n(f_n)))\right] \,, \tag{3.13}$$

et

$$V_n(f_n) \stackrel{\text{def}}{=} \sum_{p=0}^{n} \mathbb{E}\left[W_p(\bar{Q}_{p,n}(f_n - \eta_n(f_n)))^2\right] \,. \tag{3.14}$$

Démonstration. En ce qui concerne le biais, nous décomposons l'erreur de la manière suivante

$$\begin{aligned}
\eta_n^{N_1}(f_n) - \eta_n(f_n) &= \frac{\gamma_n^{N_1}}{\gamma_n^{N_1}(1)}(f_n - \eta_n(f_n)) = \frac{\gamma_n(1)}{\gamma_n^{N_1}(1)} \frac{\gamma_n^{N_1}}{\gamma_n(1)}(f_n - \eta_n(f_n)) \\
&= \frac{\gamma_n(1)}{\gamma_n^{N_1}(1)} \left[\gamma_n^{N_1}\left(\frac{f_n - \eta_n(f_n)}{\gamma_n(1)}\right) - \gamma_n\left(\frac{f_n - \eta_n(f_n)}{\gamma_n(1)}\right)\right] \\
&= \left[\frac{\gamma_n(1)}{\gamma_n^{N_1}(1)} - 1\right] \left[\gamma_n^{N_1}\left(\frac{f_n - \eta_n(f_n)}{\gamma_n(1)}\right) - \gamma_n\left(\frac{f_n - \eta_n(f_n)}{\gamma_n(1)}\right)\right] + [\gamma_n^{N_1} - \gamma_n]\left(\frac{f_n - \eta_n(f_n)}{\gamma_n(1)}\right) \,,
\end{aligned}$$

où l'espérance du second terme est nulle d'après la Proposition 3.1. En remarquant que

$$\left[\frac{\gamma_n(1)}{\gamma_n^{N_1}(1)} - 1\right] = -\frac{\gamma_n(1)}{\gamma_n^{N_1}(1)} \left[\frac{\gamma_n^{N_1}(1)}{\gamma_n(1)} - 1\right] = -\frac{1}{\gamma_n^{N_1}(1)}[\gamma_n^{N_1} - \gamma_n](1) \,,$$

on obtient

$$\begin{aligned}
N_1 \mathbb{E}\left[\eta_n^{N_1}(f_n) - \eta_n(f_n)\right] &= -\mathbb{E}\left[\frac{1}{\gamma_n^{N_1}(1)} W_n^{\gamma,N_1}(1) W_n^{\gamma,N_1}\left(\frac{f_n - \eta_n(f_n)}{\gamma_n(1)}\right)\right] \\
&= -\mathbb{E}\left[\frac{1}{\gamma_n(1)} W_n^{\gamma,N_1}(1) W_n^{\eta,N_1}(f_n)\right] \,,
\end{aligned}$$

où W_n^{γ,N_1} et W_n^{η,N_1} sont définis respectivement en (3.11) et (3.9). D'après le Théorème 3.1 :

$$\lim_{N_1 \to \infty} N_1 \mathbb{E}\left[\eta_n^{N_1}(f_n) - \eta_n(f_n)\right] = -\mathbb{E}\left[\frac{1}{\gamma_n(1)} W_n^{\gamma}(1) W_n^{\eta}(f_n)\right] = B_n(f_n) \,, \tag{3.15}$$

par les définitions de W_n^{γ} et W_n^{η}.

Considérons maintenant la décomposition suivante de la variance :

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left(\eta_n^{N_1}(f_n)\right) &= \mathbb{E}\left[\left(\eta_n^{N_1}(f_n) - \mathbb{E}\left[\eta_n^{N_1}(f_n)\right]\right)^2\right] \\
&= \mathbb{E}\left[\left(\eta_n^{N_1}(f_n) - \eta_n(f_n)\right)^2\right] - \left\{\mathbb{E}\left[\eta_n^{N_1}(f_n) - \eta_n(f_n)\right]\right\}^2 \,.
\end{aligned}$$

Par (3.15) :

$$\left\{\mathbb{E}\left[\eta_n^{N_1}(f_n) - \eta_n(f_n)\right]\right\}^2 = O(1/N_1^2) \,.$$

D'après la définition (3.11) de W_n^{η,N_1}, il vient

$$\mathbb{E}\left[\left(\eta_n^{N_1}(f_n)-\eta_n(f_n)\right)^2\right]=\frac{1}{N_1}\mathbb{E}\left[W_n^{\eta,N_1}(f_n)^2\right]\ ,$$

ce qui implique que

$$\lim_{N_1\to\infty} N_1\mathbb{V}\mathrm{ar}\left(\eta_n^{N_1}(f_n)\right)=\mathbb{E}\left[W_n^{\eta}(f_n)^2\right]=V_n(f_n)\ ,$$

par la définition de W_n^{η}. □

Corollaire 3.1. *La variance asymptotique de $\boldsymbol{\eta}_n$ est minimale par rapport à $(\varepsilon_p)_{0\le p\le n-1}$ introduit en* (3.4) *pour*

$$\varepsilon_p=\left(\mathrm{essup}_{\eta_p}(g_p)\right)^{-1}\ ,\quad 0\le p\le n-1\ .$$

Démonstration. Le Théorème 3.2 indique que la variance asymptotique de $\boldsymbol{\eta}_n$ peut s'écrire comme la somme d'erreurs locales de la forme, pour $f_p\in\mathcal{B}_b(\mathbb{X}_p,\mathcal{X}_p)$

$$\eta_p\left(S_{p,\eta_p}\left[f_p-S_{p,\eta_p}(f_p)\right]^2\right)\ ,$$

où $S_{p,\eta_p}\left[f_p-S_{p,\eta_p}(f_p)\right]^2$ désigne la fonction

$$x\mapsto S_{p,\eta_p}\left(x,\left[f_p-S_{p,\eta_p}(f_p)\right]^2\right)=S_{p,\eta_p}\left(x,\left[f_p-S_{p,\eta_p}(x,f_p)\right]^2\right)=S_{p,\eta_p}(x,f_p^2)-S_{p,\eta_p}(x,f_p)^2\ .$$

Dans le cas particulier où $\varepsilon_p=0$, cette fonction est constante et on a simplement

$$S_{p,\eta_p}\left(x,[f-S_{n,\eta_n}(f)]^2\right)=\Psi_p(\eta_p)\left([f_p-\Psi_p(\eta_p)(f_p)]^2\right)\ .$$

Plus généralement, on a

$$\eta_p\left(S_{p,\eta_p}\left[f_p-S_{p,\eta_p}(f_p)\right]^2\right)=\Psi_p(\eta_p)\left([f_p-\Psi_p(\eta_p)(f_p)]^2\right)-\left[\eta_p\left(S_{p,\eta_p}(f_p)^2\right)-\Psi_p(\eta_p)(f_p)^2\right]\ .$$

En utilisant le fait que

$$\eta_p\left(S_{p,\eta_p}(f_p)^2\right)-\Psi_p(\eta_p)(f_p)^2=\eta_p\left[\left(S_{p,\eta_p}(f_p)-\Psi_p(\eta_p)(f_p)\right)^2\right]\ ,$$

et

$$S_{p,\eta_p}(f_p)-\Psi_p(\eta_p)(f_p)=\varepsilon_p\,g_p\left(f_p-\Psi_{f_p}(\eta_p)(f_p)\right)\ ,$$

on conclu que

$$\eta_p\left(S_{p,\eta_p}\left[f_p-S_{p,\eta_p}(f_p)\right]^2\right)=\Psi_p(\eta_p)\left([f_p-\Psi_p(\eta_p)(f_p)]^2\right)-\varepsilon_p^2\eta_p\left(g_p^2\,(f_p-\Psi_p(\eta_p)(f_p))^2\right)\ .$$

□

Exemple 3.2 (Réduction de la variance de $\eta_n^{N_1}$ dans le modèle de volatilité stochastique)**.** Dans le cadre du modèle de volatilité stochastique présenté dans l'Exemple 1.2, nous avons simulé l'estimateur $\eta_{1000}^{N_1}$ pour plusieurs valeurs de N_1 avec $\varepsilon_n=0$ et $\varepsilon_n=1/\max_{1\le i\le N_1} g_n(\xi_n^{N_1,i})$. D'une part, l'utilisation d'un ε non nul n'augmente pas le temps de calcul comme le montre la Figure 3.2. En effet, il y a une étape supplémentaire de pré-sélection des particules qui vont être resélectionnées mais elle est compensée par le fait qu'il y a justement moins de particules à resélectionner.

D'autre part, comme attendu, choisir $\varepsilon_n=1/\max_{1\le i\le N_1} g_n(\xi_n^{N_1,i})$ permet de réduire significativement la variance. La Table 3.1 montre en effet que l'on obtient une réduction de 13% à 39% de la variance par rapport au choix $\varepsilon_n=0$.

Nous prouvons enfin la Proposition suivante qui sera utile dans la prochaine section.

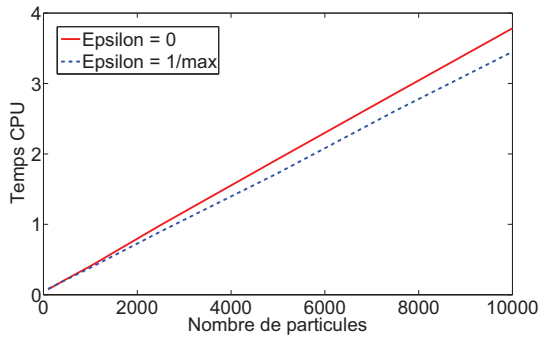


FIGURE 3.2 – Temps CPU avec ou sans l'utilisation de ε.

TABLE 3.1 – Réduction de variance en pourcentage de la variance pour $\varepsilon = 0$.

N	100	500	1000	2500	5000	7500	10000
Réduction (%)	24	31	34	21	13	39	16

Proposition 3.2. *Pour tout horizon de temps $n \geq 0$ et toutes fonctions $f_n^1, f_n^2 \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$ telles que $\eta_n(f_n^1) = 0$, on a*

$$\lim_{N_1 \to \infty} N_1 \mathbb{E}\left[\eta_n^{N_1}(f_n^1)\eta_n^{N_1}(f_n^2)\prod_{p=0}^{n-1}\eta_p^{N_1}(g_p)\right] = \gamma_n(1)\sum_{p=0}^{n}\mathbb{E}\left[W_p(\bar{Q}_{p,n}(f_n^1))W_p(\bar{Q}_{p,n}(f_n^2 - \eta_n(f_n^2)))\right] .$$

Démonstration. Par la définition (3.7) de $\gamma_n^{N_1}$ on a $\eta_n^{N_1}(f_n^1)\prod_{p=0}^{n-1}\eta_p^{N_1}(g_p) = \gamma_n^{N_1}(f_n^1)$ et d'après la Proposition 3.1, $\mathbb{E}\left[\gamma_n^{N_1}(f_n^1)\right] = \gamma_n(f_n^1) = \gamma_n(1)\eta_n(f_n^1) = 0$ de telle sorte que

$$\begin{aligned}
\mathbb{E}\left[\eta_n^{N_1}(f_n^1)\eta_n^{N_1}(f_n^2)\prod_{p=0}^{n-1}\eta_p^{N_1}(g_p)\right] &= \mathbb{E}\left[\gamma_n^{N_1}(f_n^1)\eta_n^{N_1}(f_n^2)\right] \\
&= \mathbb{E}\left[\gamma_n^{N_1}(f_n^1)\left(\eta_n^{N_1}(f_n^2) - \eta_n(f_n^2)\right)\right] + \mathbb{E}\left[\gamma_n^{N_1}(f_n^1)\right]\eta_n(f_n^2) \\
&= \mathbb{E}\left[\left(\gamma_n^{N_1}(f_n^1) - \gamma_n(f_n^1)\right)\left(\eta_n^{N_1}(f_n^2) - \eta_n(f_n^2)\right)\right] \\
&= \frac{1}{N_1}\mathbb{E}\left[W_n^{\gamma,N_1}(f_n^1)W_n^{\eta,N_1}(f_n^2)\right] .
\end{aligned}$$

Alors, le Lemme 3.2 donne

$$\lim_{N_1 \to \infty} N_1 \mathbb{E}\left[\eta_n^{N_1}(f_n^1)\eta_n^{N_1}(f_n^2)\prod_{p=0}^{n-1}\eta_p^{N_1}(g_p)\right] = \mathbb{E}\left[W_n^{\eta}(f_n^2)W_n^{\gamma}(f_n^1)\right] ,$$

où W_n^{γ} et W_n^{η} sont définis par (3.10) et (3.12). □

3.3 Modèle d'îlot de Feynman-Kac

Pour $\mathbf{x}_n \in \mathbb{X}_n^{N_1}$ on pose

$$\boldsymbol{g}_n(\mathbf{x}_n) = m(\mathbf{x}_n, g_n) , \tag{3.16}$$

et on introduit le modèle de Feynman-Kac $\{(\boldsymbol{\eta}_n, \boldsymbol{\gamma}_n)\}_{n\geq 0}$ défini par

$$\boldsymbol{\eta}_n(\boldsymbol{f}_n) = \boldsymbol{\gamma}_n(\boldsymbol{f}_n)/\boldsymbol{\gamma}_n(1) \quad \text{et} \quad \boldsymbol{\gamma}_n(\boldsymbol{f}_n) = \mathbb{E}\left[\boldsymbol{f}_n(\xi_n^{N_1})\prod_{0\leq p<n}\boldsymbol{g}_p(\xi_p^{N_1})\right] ,$$

où $\boldsymbol{f}_n \in \mathcal{B}_b(\mathbb{X}_n, \boldsymbol{\mathcal{X}}_n)$, $(\mathbb{X}_n, \boldsymbol{\mathcal{X}}_n) \stackrel{\text{def}}{=} (\mathbb{X}_n^{N_1}, \mathcal{X}_n^{\otimes N_1})$ et $(\xi_n^{N_1})_{n\geq 0}$ est la chaîne de Markov définie dans la Section 3.2.

Si $\boldsymbol{f}_n$ est de la forme $\boldsymbol{f}_n(\xi_n^{N_1}) = m(\xi_n^{N_1}, f_n) = \eta_n^{N_1}(f_n)$ pour $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$, étant donné que $\eta_p^{N_1}(g_p) = m(\xi_p^{N_1}, g_p) = \boldsymbol{g}_p(\xi_p^{N_1})$, d'après la Proposition 3.1, on a

$$\gamma_n(f_n) = \mathbb{E}\left[f_n(X_n) \prod_{0\leq p<n} g_p(X_p)\right] = \mathbb{E}\left[\boldsymbol{f}_n(\xi_n^{N_1}) \prod_{0\leq p<n} \boldsymbol{g}_p(\xi_p^{N_1})\right] = \boldsymbol{\gamma}_n(\boldsymbol{f}_n)\,,$$

ce qui implique également

$$\boldsymbol{\eta}_n(\boldsymbol{f}_n) = \eta_n(f_n)\,. \tag{3.17}$$

Description du semi-groupe Pour $\ell \in \mathbb{N}$, considérons le noyau $\boldsymbol{Q}_\ell$ de $(\mathbb{X}_{\ell-1}, \boldsymbol{\mathcal{X}}_{\ell-1})$ vers $(\mathbb{X}_\ell, \boldsymbol{\mathcal{X}}_\ell)$ défini par

$$\boldsymbol{Q}_\ell(x_{\ell-1}, \mathbf{A}_\ell) \stackrel{\text{def}}{=} \boldsymbol{g}_{\ell-1}(\mathbf{x}_{\ell-1}) \boldsymbol{M}_\ell(\mathbf{x}_{\ell-1}, \mathbf{A}_\ell)\,, \quad \mathbf{x}_{\ell-1} \in \mathbb{X}_{\ell-1}\,, \quad \mathbf{A}_\ell \in \boldsymbol{\mathcal{X}}_\ell\,,$$

où $\boldsymbol{M}_n$ est défini par (3.6) et $\boldsymbol{g}_n$ par (3.16). Pour $p \leq n$, définissons $\boldsymbol{Q}_{p,n}$ le noyau de $(\mathbb{X}_p, \boldsymbol{\mathcal{X}}_p)$ vers $(\mathbb{X}_n, \boldsymbol{\mathcal{X}}_n)$ par l'équation

$$\boldsymbol{Q}_{p,n} \stackrel{\text{def}}{=} \boldsymbol{Q}_{p+1} \boldsymbol{Q}_{p+2} \cdots \boldsymbol{Q}_n\,.$$

Avec cette définition, on a $\boldsymbol{\gamma}_n = \boldsymbol{\gamma}_p \boldsymbol{Q}_{p,n}$, et, pour tout $\mathbf{x}_p \in \mathbb{X}_p$, $\mathbf{A}_n \in \boldsymbol{\mathcal{X}}_n$

$$\boldsymbol{Q}_{p,n}(\mathbf{x}_p, \mathbf{A}_n) = \mathbb{E}\left[\mathbf{1}_{\mathbf{A}_n}(\xi_n^{N_1}) \prod_{p\leq q<n} \boldsymbol{g}_q(\xi_q^{N_1}) \middle| \xi_p^{N_1} = \mathbf{x}_p\right]\,.$$

Définissons par $\mathcal{P}(\mathbb{X}_n, \boldsymbol{\mathcal{X}}_n)$ l'ensemble des mesures de probabilités sur $(\mathbb{X}_n, \boldsymbol{\mathcal{X}}_n)$. Il est facilement vérifiable que la suite de mesures $(\boldsymbol{\eta}_n)_{n\geq 0}$ satisfait la relation suivante

$$\boldsymbol{\eta}_{n+1} = \boldsymbol{\Phi}_{n+1}(\boldsymbol{\eta}_n)\,, \tag{3.18}$$

où $\boldsymbol{\Phi}_{n+1} : \mathcal{P}(\mathbb{X}_n, \boldsymbol{\mathcal{X}}_n) \to \mathcal{P}(\mathbb{X}_{n+1}, \boldsymbol{\mathcal{X}}_{n+1})$ est donné par

$$\boldsymbol{\Phi}_{n+1}(\boldsymbol{\eta}_n) \stackrel{\text{def}}{=} \boldsymbol{\Psi}_n(\boldsymbol{\eta}_n) \boldsymbol{M}_{n+1}\,. \tag{3.19}$$

Dans l'équation ci-dessus, $\boldsymbol{\Psi}_n : \mathcal{P}(\mathbb{X}_n, \boldsymbol{\mathcal{X}}_n) \to \mathcal{P}(\mathbb{X}_n, \boldsymbol{\mathcal{X}}_n)$ désigne la transformation de Boltzmann-Gibbs

$$\boldsymbol{\Psi}_n(\boldsymbol{\eta}_n)(\mathbf{A}_n) \stackrel{\text{def}}{=} \frac{1}{\boldsymbol{\eta}_n(\boldsymbol{g}_n)} \int_{\mathbf{A}_n} \boldsymbol{g}_n(\mathbf{x})\, \boldsymbol{\eta}_n(\mathrm{d}\mathbf{x})\,, \quad \mathbf{A}_n \in \boldsymbol{\mathcal{X}}_n\,.$$

Comme $\boldsymbol{\eta}_n = \boldsymbol{\gamma}_p \boldsymbol{Q}_{p,n} / \boldsymbol{\gamma}_p(\boldsymbol{Q}_{p,n}(1))$ et $\boldsymbol{\eta}_p = \boldsymbol{\gamma}_p / \boldsymbol{\gamma}_p(1)$, on a

$$\boldsymbol{\eta}_n = \frac{\boldsymbol{\eta}_p \boldsymbol{Q}_{p,n}}{\boldsymbol{\eta}_p(\boldsymbol{Q}_{p,n}(1))} = \boldsymbol{\eta}_p \bar{\boldsymbol{Q}}_{p,n}\,, \quad \text{où} \quad \bar{\boldsymbol{Q}}_{p,n} \stackrel{\text{def}}{=} \frac{\boldsymbol{Q}_{p,n}}{\boldsymbol{\eta}_p(\boldsymbol{Q}_{p,n}(1))} = \frac{\boldsymbol{\gamma}_p(1)}{\boldsymbol{\gamma}_n(1)} \boldsymbol{Q}_{p,n} = \frac{\gamma_p(1)}{\gamma_n(1)} \boldsymbol{Q}_{p,n}\,.$$

Lemme 3.3. *Pour des fonctions liénaires $\boldsymbol{f}_n \in \mathcal{B}_b(\mathbb{X}_n, \boldsymbol{\mathcal{X}}_n)$ de la forme*

$$\boldsymbol{f}_n(\mathbf{x}_n) = m(\mathbf{x}_n, f_n)\,, \quad f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)\,,$$

on a

$$\boldsymbol{Q}_n(\mathbf{x}_{n-1}, \boldsymbol{f}_n) = m(\mathbf{x}_{n-1}, Q_n f_n)\,, \tag{3.20}$$

pour tout $p \leq n$

$$\boldsymbol{Q}_{p,n}(\mathbf{x}_p, \boldsymbol{f}_n) = m(\mathbf{x}_p, Q_{p,n} f_n)\,, \tag{3.21}$$

et

$$\bar{\boldsymbol{Q}}_{p,n}(\mathbf{x}_p, \boldsymbol{f}_n) = m(\mathbf{x}_p, \bar{Q}_{p,n} f_n)\,.$$

Démonstration. On a d'après le Lemme 3.1

$$\boldsymbol{M}_n(\mathbf{x}_{n-1},\boldsymbol{f}_n)=\mathbb{E}\left[\boldsymbol{f}_n(\xi_n^{N_1})\middle|\xi_{n-1}^{N_1}=\mathbf{x}_{n-1}\right]=\mathbb{E}\left[m(\xi_n^{N_1},f_n)\middle|\xi_{n-1}^{N_1}=\mathbf{x}_{n-1}\right]=\frac{m(\mathbf{x}_{n-1},Q_nf_n)}{m(\mathbf{x}_{n-1},g_{n-1})}\,,$$

ce qui implique

$$\boldsymbol{Q}_n(\mathbf{x}_{n-1},\boldsymbol{f}_n)=\boldsymbol{g}_{n-1}(\mathbf{x}_{n-1})\,\boldsymbol{M}_n(\mathbf{x}_{n-1},\boldsymbol{f}_n)=m(\mathbf{x}_{n-1},g_{n-1})\times\frac{m(\mathbf{x}_{n-1},Q_nf_n)}{m(\mathbf{x}_{n-1},g_{n-1})}\,,$$

montrant (3.20). La preuve de (3.21) suit par récurrence car

$$\boldsymbol{Q}_{p,n}(\mathbf{x}_p,\boldsymbol{f}_n)=\boldsymbol{Q}_{p,n-1}(\mathbf{x}_p,\boldsymbol{Q}_n\boldsymbol{f}_n)\,.$$

□

Modèle d'îlots de particules en interaction Soit N_2 un entier. Nous définissons le noyau markovien $\mathcal{K}_{n+1}^{N_2}$ de $(\mathbb{X}_n^{N_2},\boldsymbol{\mathcal{X}}_n^{\otimes N_2})$ dans $(\mathbb{X}_{n+1}^{N_2},\boldsymbol{\mathcal{X}}_{n+1}^{\otimes N_2})$ pour tout $(\mathbf{x}_n^1,\ldots,\mathbf{x}_n^{N_2})\in\mathbb{X}_n^{N_2}$ et $(\mathbf{A}_{n+1}^1,\ldots,\mathbf{A}_{n+1}^{N_2})\in\boldsymbol{\mathcal{X}}_{n+1}^{N_2}$ par

$$\mathcal{K}_{n+1}^{N_2}(\mathbf{x}_n^1,\ldots,\mathbf{x}_n^{N_2},\mathbf{A}_{n+1}^1\times\cdots\times\mathbf{A}_{n+1}^{N_2})\stackrel{\text{def}}{=}\prod_{1\leq i\leq N_2}\boldsymbol{\Phi}_{n+1}(\boldsymbol{m}(\mathbf{x}_n^1,\ldots,\mathbf{x}_n^{N_2},\cdot))(\mathbf{A}_{n+1}^i)\,,$$

où $\boldsymbol{m}(\mathbf{x}_n^1,\ldots,\mathbf{x}_n^{N_2},\cdot)$ désigne la mesure empirique associée aux points $(\mathbf{x}_n^1,\ldots,\mathbf{x}_n^{N_2})$ donnée pour tout $\mathbf{A}_n\in\boldsymbol{\mathcal{X}}_n$ par

$$\boldsymbol{m}(\mathbf{x}_n^1,\ldots,\mathbf{x}_n^{N_2},\mathbf{A}_n)\stackrel{\text{def}}{=}\frac{1}{N_2}\sum_{i=1}^{N_2}\delta_{\mathbf{x}_n^i}(\mathbf{A}_n)\,.$$

Définissons également une chaîne de Markov $(\boldsymbol{\xi}_n)_{n\geq 0}$ où pour tout $n\in\mathbb{N}$, $\boldsymbol{\xi}_n=(\boldsymbol{\xi}_n^1,\ldots,\boldsymbol{\xi}_n^{N_2})\in\mathbb{X}_n^{N_2}$ avec le noyau de transition $\mathcal{K}_{n+1}^{N_2}$.

Chaque individu $\boldsymbol{\xi}_n^i$ peut être interprété comme un îlot constitué de N_1 particules

$$\boldsymbol{\xi}_n^i=\left(\boldsymbol{\xi}_n^{i,j}\right)_{1\leq j\leq N_1}\in\mathbb{X}_n\,.$$

Dans cette interprétation, le modèle particulaire défini ci-dessus peut être vu comme un modèle de particules en interaction. L'interaction est donnée par la moyenne empirique des fonctions potentielles dans chaque îlot. Nous pouvons aussi remarquer que pour $N_1=1$, chaque îlot est en fait une seule particule. Dans ce cas, le modèle coïncide avec le modèle particulaire de taille N_2 associé aux mesures de Feynman-Kac $(\eta_n)_{n\geq 0}$. D'après (3.3), on remarque également que la transition $\boldsymbol{\xi}_n\rightsquigarrow\boldsymbol{\xi}_{n+1}$ est du type sélection-mutation

$$\boldsymbol{\xi}_n=\left(\boldsymbol{\xi}_n^i\right)_{1\leq i\leq N_2}\in\mathbb{X}_n^{N_2}\xrightarrow{\text{sélection}}\widehat{\boldsymbol{\xi}}_n=\left(\widehat{\boldsymbol{\xi}}_n^i\right)_{1\leq i\leq N_2}\in\mathbb{X}_n^{N_2}\xrightarrow{\text{mutation}}\boldsymbol{\xi}_{n+1}=\left(\boldsymbol{\xi}_{n+1}^i\right)_{1\leq i\leq N_2}\in\mathbb{X}_{n+1}^{N_2}$$

Lors de l'étape de sélection, on sélectionne aléatoirement N_2 îlots $\widehat{\boldsymbol{\xi}}_n=\left(\widehat{\boldsymbol{\xi}}_n^i\right)_{1\leq i\leq N_2}$ parmi l'ensemble des îlots $\boldsymbol{\xi}_n=\left(\boldsymbol{\xi}_n^i\right)_{1\leq i\leq N_2}\in\mathbb{X}_n^{N_2}$ avec une probabilité proportionnelle à $\boldsymbol{g}_n(\boldsymbol{\xi}_n^i)=m(\boldsymbol{\xi}_n^i,g_n)$. Lors de l'étape de mutation, les îlots sélectionnés $\widehat{\boldsymbol{\xi}}_n^i$ évoluent aléatoirement vers une nouvelle configuration $\boldsymbol{\xi}_{n+1}^i$ selon le noyau $\boldsymbol{M}_{n+1}$. Plus formellement, $\boldsymbol{\xi}_{n+1}^i$ sont des variables aléatoires indépendantes distribuées selon $\boldsymbol{M}_{n+1}\left(\widehat{\boldsymbol{\xi}}_n^i,\cdot\right)$.

Les approximations particulaires de taille N_2 des mesures $\boldsymbol{\eta}_n$ et $\boldsymbol{\gamma}_n$ sont définies pour $\boldsymbol{f}_n\in\mathcal{B}_b(\mathbb{X}_n,\boldsymbol{\mathcal{X}}_n)$ par

$$\boldsymbol{\eta}_n^{N_2}(\boldsymbol{f}_n)\stackrel{\text{def}}{=}\boldsymbol{m}(\boldsymbol{\xi}_n,\boldsymbol{f}_n)\quad\text{et}\quad\boldsymbol{\gamma}_n^{N_2}(\boldsymbol{f}_n)\stackrel{\text{def}}{=}\boldsymbol{\eta}_n^{N_2}(\boldsymbol{f}_n)\prod_{0\leq p<n}\boldsymbol{\eta}_p^{N_2}(\boldsymbol{g}_p)\,.\tag{3.22}$$

Pour des fonctions linéaires $\boldsymbol{f}_n\in\mathcal{B}_b(\mathbb{X}_n,\boldsymbol{\mathcal{X}}_n)$ de la forme $\boldsymbol{f}_n(x_n)=m(x_n,f_n)$, on observe que

$$\boldsymbol{\eta}_n^{N_2}(\boldsymbol{f}_n)=\frac{1}{N_2}\sum_{i=1}^{N_2}\boldsymbol{f}_n(\boldsymbol{\xi}_n^i)=\frac{1}{N_2}\sum_{i=1}^{N_2}m(\boldsymbol{\xi}_n^i,f_n)=\frac{1}{N_1N_2}\sum_{i=1}^{N_2}\sum_{j=1}^{N_1}f_n(\boldsymbol{\xi}_n^{i,j})\,.$$

3.4 Biais et variance asymptotiques des modèles d'îlots de particules

3.4.1 Îlots indépendants

Soit $(\boldsymbol{\zeta}_n^i)_{i=1}^{N_2}$ N_2 îlots indépendants de taille N_1 et définissons pour tout $\boldsymbol{f}_n \in \mathcal{B}_b(\mathbf{X}_n, \boldsymbol{\mathcal{X}}_n)$

$$\widetilde{\boldsymbol{\eta}}_n^{N_2}(\boldsymbol{f}_n) \stackrel{\text{def}}{=} \frac{1}{N_2} \sum_{i=1}^{N_2} \boldsymbol{f}_n(\boldsymbol{\zeta}_n^i) \, .$$

Pour des fonctions linéaires $\boldsymbol{f}_n \in \mathcal{B}_b(\mathbf{X}_n, \boldsymbol{\mathcal{X}}_n)$ de la forme $\boldsymbol{f}_n(\mathbf{x}_n) = m(\mathbf{x}_n, f_n)$, $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$, on a

$$\widetilde{\boldsymbol{\eta}}_n^{N_2}(\boldsymbol{f}_n) = \frac{1}{N_2} \sum_{i=1}^{N_2} m(\boldsymbol{\zeta}_n^i, f_n) \, .$$

Le comportement asymptotique du biais et de la variance de $m(\boldsymbol{\zeta}_n^i, f_n)$ est celui de $\eta_n^{N_1}(f_n) = m(\xi_n^{N_1}, f_n)$ pour une population de particules générique $\xi_n^{N_1}$ telle que définie dans la Section 3.2, et d'après le Théorème 3.2 on a pour toute fonction $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$, pour le biais

$$\lim_{N_1 \to \infty} N_1 \left\{ \mathbb{E}\left[\eta_n^{N_1}(f_n)\right] - \eta_n(f_n) \right\} = B_n(f_n) \, ,$$

et pour la variance

$$\lim_{N_1 \to \infty} N_1 \mathbb{V}\mathrm{ar}\left(\eta_n^{N_1}(f_n)\right) = V_n(f_n) \, ,$$

où $B_n(f_n)$ et $V_n(f_n)$ sont définis respectivement en (3.13) et (3.14). Le théorème suivant s'en déduit alors directement.

Théorème 3.3. *Pour tout horizon de temps $n \geq 0$ et toute fonction $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$, on a*

$$\lim_{N_1 \to \infty} N_1 \left\{ \mathbb{E}\left[\widetilde{\boldsymbol{\eta}}_n^{N_2}(m(\cdot, f_n))\right] - \eta_n(f_n) \right\} = B_n(f_n) \, ,$$
$$\lim_{N_1 \to \infty} N_1 N_2 \mathbb{V}\mathrm{ar}\left(\widetilde{\boldsymbol{\eta}}_n^{N_2}(m(\cdot, f_n))\right) = V_n(f_n) \, ,$$

où $B_n(f_n)$ et $V_n(f_n)$ sont définis respectivement en (3.13) *et* (3.14)*.*

On remarque que malgré une variance qui se comporte en $(N_1 N_2)^{-1}$, le biais est indépendant de N_2 et se comporte seulement en N_1^{-1}. Son carré ne sera donc pas toujours asymptotiquement négligeable devant la variance.

3.4.2 Îlots en interaction

Dans le cas des îlots en interaction décrits dans la Section 3.3, le comportement asymptotique du biais et de la variance est donné par le théorème suivant.

Théorème 3.4. *Pour tout horizon de temps $n \geq 0$ et toute fonction $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$, on a*

$$\lim_{N_1 \to \infty} \lim_{N_2 \to \infty} N_1 N_2 \mathbb{E}\left[\boldsymbol{\eta}_n^{N_2}(m(\cdot, f_n)) - \boldsymbol{\eta}_n(m(\cdot, f_n))\right] = B_n(f_n) + \widetilde{B}_n(f_n) \, ,$$
$$\lim_{N_1 \to \infty} \lim_{N_2 \to \infty} N_1 N_2 \mathbb{V}\mathrm{ar}\left(\boldsymbol{\eta}_n^{N_2}(m(\cdot, f_n))\right) = V_n(f_n) + \widetilde{V}_n(f_n) \, ,$$

où $B_n(f_n)$, $\widetilde{B}_n(f_n)$, $V_n(f_n)$, $\widetilde{V}_n(f_n)$ sont respectivement définis en (3.13)*,* (3.26)*,* (3.14) *et* (3.27)*.*

On remarque que dans ce cas la variance présente un terme additionnel à celui des îlots indépendants mais se comporte toujours en $(N_1 N_2)^{-1}$. En revanche, le biais décroît cette fois plus rapidement en $(N_1 N_2)^{-1}$ et son carré sera donc asymptotiquement négligeable devant la variance.

Démonstration. **Biais asymptotique :** Pour tout entier fixé N_1, le biais asymptotique de $\boldsymbol{\eta}_n^{N_2}(\boldsymbol{f}_n)$ est donné pour toute fonction $\boldsymbol{f}_n \in \mathcal{B}_b(\mathbb{X}_n, \boldsymbol{\mathcal{X}}_n)$ en appliquant le Théorème 3.2 au modèle d'îlots particulaires pour $\varepsilon_n \equiv 0$:

$$\lim_{N_2\to\infty} N_2 \mathbb{E}\left[\boldsymbol{\eta}_n^{N_2}(\boldsymbol{f}_n) - \boldsymbol{\eta}_n(\boldsymbol{f}_n)\right] = -\sum_{p=0}^{n} \boldsymbol{\eta}_p\left[\bar{\boldsymbol{Q}}_{p,n}(1)\bar{\boldsymbol{Q}}_{p,n}(\boldsymbol{f}_n - \boldsymbol{\eta}_n(\boldsymbol{f}_n))\right] .$$

Pour des fonctions linéaires $\boldsymbol{f}_n$ de la forme $\boldsymbol{f}_n(\cdot) = m(\cdot, f_n)$ où $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$, le Lemme 3.3 indique que

$$\bar{\boldsymbol{Q}}_{p,n}\left(\xi_p^{N_1}, \boldsymbol{f}_n - \boldsymbol{\eta}_n(\boldsymbol{f}_n)\right) = m(\xi_p^{N_1}, \bar{Q}_{p,n}(f_n - \eta_n(f_n))) = \eta_p^{N_1}(\bar{Q}_{p,n}(f_n - \eta_n(f_n))) , \tag{3.23}$$

et

$$\bar{\boldsymbol{Q}}_{p,n}\left(\xi_p^{N_1}, 1\right) = m(\xi_p^{N_1}, \bar{Q}_{p,n}(1)) = \eta_p^{N_1}(\bar{Q}_{p,n}(1)) . \tag{3.24}$$

On obtient donc

$$\begin{aligned}
\boldsymbol{\eta}_p\left[\bar{\boldsymbol{Q}}_{p,n}(1)\bar{\boldsymbol{Q}}_{p,n}(\boldsymbol{f}_n - \boldsymbol{\eta}_n(\boldsymbol{f}_n))\right] &\overset{(1)}{=} \frac{\boldsymbol{\gamma}_p\left[\bar{\boldsymbol{Q}}_{p,n}(1)\bar{\boldsymbol{Q}}_{p,n}(\boldsymbol{f}_n - \boldsymbol{\eta}_n(\boldsymbol{f}_n))\right]}{\boldsymbol{\gamma}_p(1)} \\
&\overset{(2)}{=} \frac{\mathbb{E}\left[\bar{\boldsymbol{Q}}_{p,n}(\xi_p^{N_1}, 1)\bar{\boldsymbol{Q}}_{p,n}\left(\xi_p^{N_1}, \boldsymbol{f}_n - \boldsymbol{\eta}_n(\boldsymbol{f}_n)\right)\prod_{\ell=0}^{p-1}\boldsymbol{g}_\ell^{N_1}(\xi_\ell^{N_1})\right]}{\boldsymbol{\gamma}_p(1)} \\
&\overset{(3)}{=} \frac{\mathbb{E}\left[\eta_p^{N_1}(\bar{Q}_{p,n}(f_n - \eta_n(f_n)))\eta_p^{N_1}(\bar{Q}_{p,n}(1))\prod_{\ell=0}^{p-1}\eta_\ell^{N_1}(g_\ell)\right]}{\gamma_p(1)} ,
\end{aligned} \tag{3.25}$$

où (1) est simplement la définition (3.22) de $\boldsymbol{\eta}_p$, (2) vient de la définition (3.22) de $\boldsymbol{\gamma}_p$, et (3) est une conséquence de la Proposition 3.1, de la définition (3.16) de $(\boldsymbol{g}_\ell)_{\ell\geq 0}$ et des équations (3.23) et (3.24). Comme $\eta_p(\bar{Q}_{p,n}(f_n - \eta_n(f_n))) = 0$ on peut appliquer la Proposition 3.2 et

$$\begin{aligned}
&\lim_{N_1\to\infty} N_1 \boldsymbol{\eta}_p\left[\bar{\boldsymbol{Q}}_{p,n}(1)\bar{\boldsymbol{Q}}_{p,n}(\boldsymbol{f}_n - \boldsymbol{\eta}_n(\boldsymbol{f}_n))\right] \\
&\qquad = \sum_{\ell=0}^{p} \mathbb{E}\left[W_\ell(\bar{Q}_{\ell,p}(\bar{Q}_{p,n}(1) - \eta_p\bar{Q}_{p,n}(1)))W_\ell(\bar{Q}_{\ell,p}\bar{Q}_{p,n}(f_n - \eta_n(f_n)))\right] \\
&\qquad = \sum_{\ell=0}^{p} \mathbb{E}\left[W_\ell(\bar{Q}_{\ell,n}(1) - \bar{Q}_{\ell,p}(1))W_\ell(\bar{Q}_{\ell,n}(f_n - \eta_n(f_n)))\right] ,
\end{aligned}$$

d'où nous concluons que

$$\begin{aligned}
&\lim_{N_1\to\infty}\lim_{N_2\to\infty} N_1 N_2 \mathbb{E}\left[\boldsymbol{\eta}_n^{N_2}(\boldsymbol{f}_n) - \boldsymbol{\eta}_n(\boldsymbol{f}_n)\right] \\
&\qquad = -\sum_{p=0}^{n}\sum_{\ell=0}^{p} \mathbb{E}\left[W_\ell(\bar{Q}_{\ell,n}(1) - \bar{Q}_{\ell,p}(1))W_\ell(\bar{Q}_{\ell,n}(f_n - \eta_n(f_n)))\right] \\
&\qquad = -\sum_{\ell=0}^{n}\sum_{p=\ell}^{n} \mathbb{E}\left[W_\ell(\bar{Q}_{\ell,n}(1) - \bar{Q}_{\ell,p}(1))W_\ell(\bar{Q}_{\ell,n}(f_n - \eta_n(f_n)))\right] \\
&\qquad = B_n(f_n) + \widetilde{B}_n(f_n) ,
\end{aligned}$$

où $B_n(f_n)$ est défini en (3.13) et $\widetilde{B}_n(f_n)$ est donné par

$$\begin{aligned}
\widetilde{B}_n(f_n) \overset{\text{def}}{=} &-\sum_{\ell=0}^{n}(n-\ell)\mathbb{E}\left[W_\ell(\bar{Q}_{\ell,n}(1))W_\ell(\bar{Q}_{\ell,n}(f_n - \eta_n(f_n)))\right] \\
&+ \sum_{\ell=0}^{n}\mathbb{E}\left[W_\ell\left(\sum_{p=\ell}^{n}\bar{Q}_{\ell,p}(1)\right)W_\ell(\bar{Q}_{\ell,n}(f_n - \eta_n(f_n)))\right] .
\end{aligned} \tag{3.26}$$

Variance asymptotique : Pour tout entier fixé N_1, la variance asymptotique de $\boldsymbol{\eta}_n^{N_2}(\boldsymbol{f}_n)$ est donnée pour tout $\boldsymbol{f}_n \in \mathcal{B}_b(\mathbb{X}_n, \boldsymbol{\mathcal{X}}_n)$ en appliquant le Théorème 3.2 au modèle d'îlots particulaires pour $\varepsilon_n \equiv 0$:

$$\lim_{N_2\to\infty} N_2 \mathbb{V}\mathrm{ar}\left(\boldsymbol{\eta}_n^{N_2}(\boldsymbol{f}_n)\right) = \sum_{p=0}^{n} \boldsymbol{\eta}_p \left[\bar{\boldsymbol{Q}}_{p,n}\left(\boldsymbol{f}_n - \boldsymbol{\eta}_n(\boldsymbol{f}_n)\right)^2\right] .$$

Pour des fonctions linéaires $\boldsymbol{f}_n$ de la forme $\boldsymbol{f}_n(\cdot) = m(\cdot, f_n)$ où $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$, avec un raisonnement similaire à celui utilisé en (3.25), on obtient

$$\boldsymbol{\eta}_p \left[\bar{\boldsymbol{Q}}_{p,n}\left(\boldsymbol{f}_n - \boldsymbol{\eta}_n(\boldsymbol{f}_n)\right)^2\right] = \frac{\mathbb{E}\left[\eta_p^{N_1}(\bar{Q}_{p,n}(f_n - \eta_n(f_n)))^2 \prod_{\ell=0}^{p-1} \eta_\ell^{N_1}(g_\ell)\right]}{\gamma_p(1)} .$$

Comme $\eta_p(\bar{Q}_{p,n}(\cdot, f_n - \eta_n(f_n))) = 0$ on peut appliquer la Proposition 3.2 et

$$\begin{aligned}\lim_{N_1\to\infty} N_1 \boldsymbol{\eta}_p \left[\bar{\boldsymbol{Q}}_{p,n}\left(\boldsymbol{f}_n - \boldsymbol{\eta}_n(\boldsymbol{f}_n)\right)^2\right] &= \sum_{\ell=0}^{p} \mathbb{E}\left[W_\ell(\bar{Q}_{\ell,p}\bar{Q}_{p,n}(f_n - \eta_n(f_n)))^2\right] \\ &= \sum_{\ell=0}^{p} \mathbb{E}\left[W_\ell(\bar{Q}_{\ell,n}(f_n - \eta_n(f_n)))^2\right] ,\end{aligned}$$

d'où l'on conclu que

$$\begin{aligned}\lim_{N_1\to\infty}\lim_{N_2\to\infty} N_1 N_2 \mathbb{V}\mathrm{ar}\left(\boldsymbol{\eta}_n^{N_2}(\boldsymbol{f}_n)\right) &= \sum_{p=0}^{n}\sum_{\ell=0}^{p} \mathbb{E}\left[W_\ell(\bar{Q}_{\ell,n}(f_n - \eta_n(f_n)))^2\right] \\ &= \sum_{\ell=0}^{n}\sum_{p=\ell}^{n} \mathbb{E}\left[W_\ell(\bar{Q}_{\ell,n}(f_n - \eta_n(f_n)))^2\right] = V_n(f_n) + \widetilde{V}_n(f_n) ,\end{aligned}$$

où $V_n(f_n)$ est défini en (3.14) et $\widetilde{V}_n(f_n)$ est donné par

$$\widetilde{V}_n(f_n) \stackrel{\text{def}}{=} \sum_{\ell=0}^{n} (n-\ell)\mathbb{E}\left[W_\ell(\bar{Q}_{\ell,n}(f_n - \eta_n(f_n)))^2\right] . \tag{3.27}$$

□

3.5 Conclusion

Comme montré dans les Théorèmes 3.3 et 3.4, un compromis doit être fait entre le biais et la variance afin de choisir lequel des deux estimateurs $\boldsymbol{\eta}_n^{N_2}$ et $\widetilde{\boldsymbol{\eta}}_n^{N_2}$ est le plus avantageux pour N_1 et N_2 fixés et suffisamment grands. On peut comparer les erreurs quadratiques moyennes de chacun d'eux, données par la somme de la variance et du biais au carré. Les îlots indépendants seront alors asymptotiquement plus efficaces que ceux en interaction dès lors que

$$\frac{V_n(f_n)}{N_1 N_2} + \frac{B_n(f_n)^2}{N_1^2} < \frac{V_n(f_n) + \widetilde{V}_n(f_n)}{N_1 N_2} \quad \Leftrightarrow \quad N_1 > \frac{B_n(f_n)^2}{\widetilde{V}_n(f_n)} N_2 .$$

Par conséquence, les îlots doivent être gardés indépendants pour les grandes valeurs de N_1 par rapport à N_2 pour éviter le terme de variance additionnel induit par les étapes de resélection entre les îlots en interaction. Au contraire, on favorisera le modèle d'îlots en interaction lorsque N_1 est petit par rapport à N_2 afin de compenser le biais du cas indépendant.

Exemple 3.3 (Îlots indépendants vs en interaction dans le LGM)**.** Dans le cas du LGM en dimension 1 présenté dans l'exemple 2.1, η_n peut être calculée explicitement à l'aide du filtre de Kalman. Nous avons donc utilisé ce modèle pour évaluer l'erreur quadratique générée par $\boldsymbol{\eta}_n^{N_2}$ et $\widetilde{\boldsymbol{\eta}}_n^{N_2}$. Comme annoncé précédemment, la Figure 3.3 montre bien que l'utilisation d'îlots en interaction est préconisée pour les petites valeurs de N_1 comparativement à N_2 et inversement.

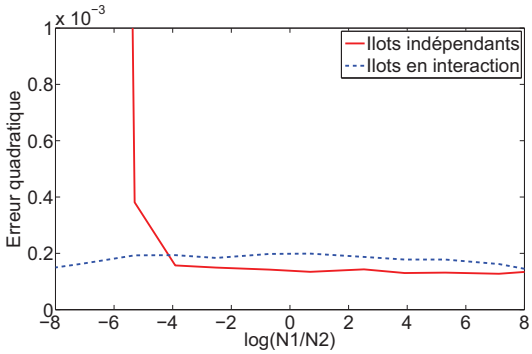


FIGURE 3.3 – Erreur quadratique moyenne avec ou sans interaction entre les îlots pour $N_1 N_2$ constant.

Chapitre 4

Inégalités de déviation non asymptotiques pour le lissage de fonctionnelles additives dans le cadre des HMM non linéaires

Sommaire

Abstract: The approximation of fixed-interval smoothing distributions is a key issue in inference for general state-space hidden Markov models (HMM). This contribution establishes non-asymptotic bounds for the Forward Filtering Backward Smoothing (FFBS) and the Forward Filtering Backward Simulation (FFBSi) estimators of fixed-interval smoothing functionals. We show that the rate of convergence of the L_q-mean errors of both methods depends on the number of observations T and the number of particles N only through the ratio T/N for additive functionals. In the case of the FFBS, this improves recent results providing bounds depending on $T/\sqrt{N}$.

Keywords: Additive functionals, Deviation inequalities, FFBS, FFBSi, Particle-based approximations, Sequential Monte Carlo methods

This work was supported by the French National Research Agency, under the program ANR-07 ROBO 0002

4.1 Introduction

State-space models play a key role in statistics, engineering and econometrics; see Cappé et al. [2005], Durbin and Koopman [2000], West and Harrison [1989]. Consider a process $\{X_t\}_{t\geq 0}$ taking values in a general state-space $\mathbb{X}$. This hidden process can be observed only through the observation process $\{Y_t\}_{t\geq 0}$ taking values in $\mathbb{Y}$. Statistical inference in general state-space models involves the computation of expectations of additive functionals of the form

$$S_T = \sum_{t=1}^{T} h_t(X_{t-1}, X_t) \,,$$

conditionally to $\{Y_t\}_{t=0}^{T}$, where T is a positive integer and $\{h_t\}_{t=1}^{T}$ are functions defined on $\mathbb{X}^2$. These smoothed additive functionals appear naturally for maximum likelihood parameter inference in hidden Markov models. The computation of the gradient of the log-likelihood function (Fisher score) or of the intermediate quantity of the Expectation Maximization algorithm involves the estimation of such smoothed functionals, see [Cappé et al., 2005, Chapter 10 and 11] and Doucet et al. [2010].

Except for linear Gaussian state-spaces or for finite state-spaces, these smoothed additive functionals cannot be computed explicitly. In this paper, we consider Sequential Monte Carlo algorithms, henceforth referred to as particle methods, to approximate these quantities. These methods combine sequential importance sampling and sampling importance resampling steps to produce a set of random particles with associated importance weights to approximate the fixed-interval smoothing distributions.

The most straightforward implementation is based on the so-called path-space method. The complexity of this algorithm per time-step grows only linearly with the number N of particles, see Del Moral [2004]. However, a well-known shortcoming of this algorithm is known in the literature as the path degeneracy; see Doucet et al. [2010] for a discussion.

Several solutions have been proposed to solve this degeneracy problem. In this paper, we consider the Forward Filtering Backward Smoothing algorithm (FFBS) and the Forward Filtering Backward Simulation algorithm (FFBSi) introduced in Doucet et al. [2000] and further developed in Godsill et al. [2004]. Both algorithms proceed in two passes. In the forward pass, a set of particles and weights is stored. In the Backward pass of the FFBS the weights are modified but the particles are kept fixed. The FFBSi draws independently different particle trajectories among all possible paths. Since they use a backward step, these algorithms are mainly adapted for batch estimation problems. However, as shown in Del Moral et al. [2010a], when applied to additive functionals, the FFBS algorithm can be implemented forward in time, but its complexity grows quadratically with the number of particles. As shown in Douc et al. [2010], it is possible to implement the FFBSi with a complexity growing only linearly with the number of particles.

The control of the L_q-norm of the deviation between the smoothed additive functional and its particle approximation has been studied recently in Del Moral et al. [2010a,b]. In an unpublished paper by Del Moral et al. [2010b], it is shown that the FFBS estimator variance of any smoothed additive functional is upper bounded by terms depending on T and N only through the ratio T/N. Furthermore, in Del Moral et al. [2010a], for any $q > 2$, a L_q-mean error bound for smoothed functionals computed with the FFBS is established. When applied to strongly mixing kernels, this bound amounts to be of order $T/\sqrt{N}$ either for

(i) uniformly bounded in time general path-dependent functionals,

(ii) unnormalized additive functionals (see [Del Moral et al., 2010a, Eq. (3.8), pp. 957]).

In this paper, we establish L_q-mean error and exponential deviation inequalities of both the FFBS and FFBSi smoothed functionals estimators. We show that, for any $q \geq 2$, the L_q-mean error for both algorithms is upper bounded by terms depending on T and N only through the ratio T/N under the strong mixing conditions for (i) and (ii). We also establish an exponential deviation inequality with the same functional dependence in T and N.

This paper is organized as follows. Section 4.2 introduces further definitions and notations and the FFBS and FFBSi algorithms. In Section 4.3, upper bounds for the L_q-mean error and exponential deviation inequalities of these two algorithms are presented. In Section 4.4, some Monte Carlo experiments are presented to support our theoretical claims. The proofs are presented in Sections 4.5 and 4.6.

4.2 Framework

Let $\mathbb{X}$ and $\mathbb{Y}$ be two general state-spaces endowed with countably generated σ-fields $\mathcal{X}$ and $\mathcal{Y}$. Let M be a Markov transition kernel defined on $\mathbb{X}\times\mathcal{X}$ and $\{g_t\}_{t\geq 0}$ a family of functions defined on $\mathbb{X}$. It is assumed that, for any $x\in\mathbb{X}$, $M(x,\cdot)$ has a density $m(x,\cdot)$ with respect to a reference measure λ on $(\mathbb{X},\mathcal{X})$. For any integers $T\geq 0$ and $0\leq s\leq t\leq T$, any measurable function h on $\mathbb{X}^{t-s+1}$, and any probability distribution χ on $(\mathbb{X},\mathcal{X})$, define

$$\phi_{s:t|T}[h] \stackrel{\text{def}}{=} \frac{\int \chi(\mathrm{d}x_0)g_0(x_0)\prod_{u=1}^{T} M(x_{u-1},\mathrm{d}x_u)g_u(x_u)h(x_{s:t})}{\int \chi(\mathrm{d}x_0)g_0(x_0)\prod_{u=1}^{T} M(x_{u-1},\mathrm{d}x_u)g_u(x_u)} , \tag{4.1}$$

where $a_{u:v}$ is a short-hand notation for $\{a_s\}_{s=u}^{v}$. The dependence on $g_{0:T}$ is implicit and is dropped from the notations.

Remark 4.1. Note that this equation has a simple interpretation in the particular case of hidden Markov models. Indeed, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and $\{X_t\}_{t\geq 0}$ a Markov chain on $(\Omega,\mathcal{F},\mathbb{P})$ with transition kernel M and initial distribution χ (which we denote $X_0\sim\chi$). Let $\{Y_t\}_{t\geq 0}$ be a sequence of observations on $(\Omega,\mathcal{F},\mathbb{P})$ conditionally independent given $\sigma(X_t,t\geq 0)$ and such that the conditional distribution of Y_u given $\sigma(X_t,t\geq 0)$ has a density given by $g(X_u,\cdot)$ with respect to a reference measure on $\mathcal{Y}$ and set $g_u(x)=g(x,Y_u)$. Then, the quantity $\phi_{s:t|T}[h]$ defined by (4.1) is the conditional expectation of $h(X_{s:t})$ given $Y_{0:T}$:

$$\phi_{s:t|T}[h] = \mathbb{E}\left[h(X_{s:t})|Y_{0:T}\right] , \quad X_0\sim\chi .$$

In its original version, the FFBS algorithm proceeds in two passes. In the forward pass, each filtering distribution $\phi_t \stackrel{\text{def}}{=} \phi_{t:t}$, for any $t\in\{0,\ldots,T\}$, is approximated using weighted samples $\left\{(\omega_t^{N,\ell},\xi_t^{N,\ell})\right\}_{\ell=1}^{N}$, where T is the number of observations and N the number of particles: all sampled particles and weights are stored. In the backward pass of the FFBS, these importance weights are then modified (see Doucet et al. [2000], Hürzeler and Künsch [1998], Kitagawa [1996]) while the particle positions are kept fixed. The importance weights are updated recursively backward in time to obtain an approximation of the fixed-interval smoothing distributions $\left\{\phi_{s:T|T}\right\}_{s=0}^{T}$. The particle approximation is constructed as follows.

Forward pass Let $\{\xi_0^{N,\ell}\}_{\ell=1}^{N}$ be i.i.d. random variables distributed according to the instrumental density ρ_0 and set the importance weights $\omega_0^{N,\ell} \stackrel{\text{def}}{=} \mathrm{d}\chi/\mathrm{d}\rho_0(\xi_0^{N,\ell})g_0(\xi_0^{N,\ell})$. The weighted sample $\{(\xi_0^{N,\ell},\omega_0^{N,\ell})\}_{\ell=1}^{N}$ then targets the initial filter ϕ_0 in the sense that $\phi_0^N[h] \stackrel{\text{def}}{=} \sum_{\ell=1}^{N}\omega_0^{N,\ell}h(\xi_0^{N,\ell})/\sum_{\ell=1}^{N}\omega_0^{N,\ell}$ is a consistent estimator of $\phi_0[h]$ for any bounded and measurable function h on $\mathbb{X}$.
Let now $\{(\xi_{s-1}^{N,\ell},\omega_{s-1}^{N,\ell})\}_{\ell=1}^{N}$ be a weighted sample targeting ϕ_{s-1}. We aim at computing new particles and importance weights targeting the probability distribution ϕ_s. Following Pitt and Shephard [1999], this may be done by simulating pairs $\{(I_s^{N,\ell},\xi_s^{N,\ell})\}_{\ell=1}^{N}$ of indices and particles from the instrumental distribution:

$$\pi_{s|s}(\ell,h) \propto \omega_{s-1}^{N,\ell}\vartheta_s(\xi_{s-1}^{N,\ell})P_s(\xi_{s-1}^{N,\ell},h) ,$$

on the product space $\{1,\ldots,N\}\times\mathbb{X}$, where $\{\vartheta_s(\xi_{s-1}^{N,\ell})\}_{\ell=1}^{N}$ are the adjustment multiplier weights and P_s is a Markovian proposal transition kernel. In the sequel, we assume that $P_s(x,\cdot)$ has, for any $x\in\mathbb{X}$, a density $p_s(x,\cdot)$ with respect to the reference measure λ. For any $\ell\in\{1,\ldots,N\}$ we associate to the particle $\xi_s^{N,\ell}$ its importance weight defined by:

$$\omega_s^{N,\ell} \stackrel{\text{def}}{=} \frac{m(\xi_{s-1}^{N,I_s^{N,\ell}},\xi_s^{N,\ell})g_s(\xi_s^{N,\ell})}{\vartheta_s(\xi_{s-1}^{N,I_s^{N,\ell}})p_s(\xi_{s-1}^{N,I_s^{N,\ell}},\xi_s^{N,\ell})} .$$

Backward smoothing For any probability measure η on $(\mathbb{X},\mathcal{X})$, denote by B_η the backward smoothing kernel given, for all bounded measurable function h on $\mathbb{X}$ and for all $x\in\mathbb{X}$, by:

$$\mathrm{B}_\eta(x,h) \stackrel{\text{def}}{=} \frac{\int \eta(\mathrm{d}x')\, m(x',x)h(x')}{\int \eta(\mathrm{d}x')\, m(x',x)} ,$$

For all $s \in \{0,\dots,T-1\}$ and for all bounded measurable function h on $\mathbb{X}^{T-s+1}$, $\phi_{s:T|T}[h]$ may be computed recursively, backward in time, according to

$$\phi_{s:T|T}[h] = \int \mathrm{B}_{\phi_s}(x_{s+1}, \mathrm{d}x_s)\, \phi_{s+1:T|T}(\mathrm{d}x_{s+1:T})\, h(x_{s:T}) \,.$$

4.2.1 The forward filtering backward smoothing algorithm

Consider the weighted samples $\left\{(\xi_t^{N,\ell}, \omega_t^{N,\ell})\right\}_{\ell=1}^N$, drawn for any $t \in \{0,\dots,T\}$ in the forward pass. An approximation of the fixed-interval smoothing distribution can be obtained using

$$\phi_{s:T|T}^N[h] = \int \mathrm{B}_{\phi_s^N}(x_{s+1}, \mathrm{d}x_s)\, \phi_{s+1:T|T}^N(\mathrm{d}x_{s+1:T})\, h(x_{s:T}) \,, \tag{4.2}$$

and starting with $\phi_{T:T|T}^N[h] = \phi_T^N[h]$. Now, by definition, for all $x \in \mathbb{X}$ and for all bounded measurable function h on $\mathbb{X}$,

$$\mathrm{B}_{\phi_s^N}(x,h) = \sum_{i=1}^N \frac{\omega_s^{N,i} m(\xi_s^{N,i}, x)}{\sum_{\ell=1}^N \omega_s^{N,\ell} m(\xi_s^{N,\ell}, x)} h\left(\xi_s^{N,i}\right) \,,$$

and inserting this expression into (4.2) gives the following particle approximation of the fixed-interval smoothing distribution $\phi_{0:T|T}[h]$

$$\phi_{0:T|T}^N[h] = \sum_{i_0=1}^N \cdots \sum_{i_T=1}^N \left(\prod_{u=1}^T \Lambda_u^N(i_u, i_{u-1})\right) \times \frac{\omega_T^{N,i_T}}{\Omega_T^N} h\left(\xi_0^{N,i_0}, \dots, \xi_T^{N,i_T}\right) \,, \tag{4.3}$$

where h is a bounded measurable function on $\mathbb{X}^{T+1}$,

$$\Lambda_t^N(i,j) \stackrel{\text{def}}{=} \frac{\omega_t^{N,j} m(\xi_t^{N,j}, \xi_{t+1}^{N,i})}{\sum_{\ell=1}^N \omega_t^{N,\ell} m(\xi_t^{N,\ell}, \xi_{t+1}^{N,i})} \,, \quad (i,j) \in \{1,\dots,N\}^2 \,, \tag{4.4}$$

and

$$\Omega_t^N \stackrel{\text{def}}{=} \sum_{\ell=1}^N \omega_t^{N,\ell} \,. \tag{4.5}$$

The estimator of the fixed-interval smoothing distribution $\phi_{0:T|T}^N$ might seem impractical since the cardinality of its support is N^{T+1}. Nevertheless, for additive functionals of the form

$$S_{T,r}(x_{0:T}) = \sum_{t=r}^T h_t(x_{t-r:t}) \,, \tag{4.6}$$

where r is a non negative integer and $\{h_t\}_{t=r}^T$ is a family of bounded measurable functions on $\mathbb{X}^{r+1}$, the complexity of the FFBS algorithm is reduced to $O(N^{r+2})$. Furthermore, the smoothing of such functions can be computed forward in time as shown in Del Moral et al. [2010a]. This forward algorithm is exactly the one presented in Doucet et al. [2010] as an alternative to the use of the path-space method. Therefore, the results outlined in Section 4.3 hold for this method and confirm the conjecture mentioned in Doucet et al. [2010].

4.2.2 The forward filtering backward simulation algorithm

We now consider an algorithm whose complexity grows only linearly with the number of particles for any functional on $\mathbb{X}^{T+1}$. For any $t \in \{1,\dots,T\}$, we define

$$\mathcal{F}_t^N \stackrel{\text{def}}{=} \sigma\left\{(\xi_s^{N,i}, \omega_s^{N,i}); 0 \le s \le t, 1 \le i \le N\right\} \,.$$

The transition probabilities $\{\Lambda_t^N\}_{t=0}^{T-1}$ defined in (4.4) induce an inhomogeneous Markov chain $\{J_u\}_{u=0}^T$ evolving backward in time as follows. At time T, the random index J_T is drawn from the set $\{1,\dots,N\}$

with probability proportional to $(\omega_T^{N,1},\ldots,\omega_T^{N,N})$. For any $t\in\{0,\ldots,T-1\}$, the index J_t is sampled in the set $\{1,\ldots,N\}$ according to $\Lambda_t^N(J_{t+1},\cdot)$. The joint distribution of $J_{0:T}$ is therefore given, for $j_{0:T}\in\{1,\ldots,N\}^{T+1}$, by

$$\mathbb{P}\left[J_{0:T}=j_{0:T}\,\middle|\,\mathcal{F}_T^N\right]=\frac{\omega_T^{N,j_T}}{\Omega_T^N}\Lambda_{T-1}^N(j_T,j_{T-1})\ldots\Lambda_0^N(j_1,j_0)\,. \tag{4.7}$$

Thus, the FFBS estimator (4.3) of the fixed-interval smoothing distribution may be written as the conditional expectation

$$\phi_{0:T|T}^N[h]=\mathbb{E}\left[h\left(\xi_0^{N,J_0},\ldots,\xi_T^{N,J_T}\right)\,\middle|\,\mathcal{F}_T^N\right]\,,$$

where h is a bounded measurable function on $\mathbb{X}^{T+1}$. We may therefore construct an unbiased estimator of the FFBS estimator given by

$$\widetilde{\phi}_{0:T|T}^N[h]=N^{-1}\sum_{\ell=1}^N h\left(\xi_0^{N,J_0^\ell},\ldots,\xi_T^{N,J_T^\ell}\right)\,, \tag{4.8}$$

where $\{J_{0:T}^\ell\}_{\ell=1}^N$ are N paths drawn independently given $\mathcal{F}_T^N$ according to (4.7) and where h is a bounded measurable function on $\mathbb{X}^{T+1}$. This practical estimator was introduced in Godsill et al. [2004] (Algorithm 1, p. 158). An implementation of this estimator whose complexity grows linearly in N is introduced in Douc et al. [2010].

4.3 Non-asymptotic deviation inequalities

In this Section, the L_q-mean error bounds and exponential deviation inequalities of the FFBS and FFBSi algorithms are established for additive functionals of the form (4.6). Our results are established under the following assumptions.

A1 (i) There exists $(\sigma_-,\sigma_+)\in(0,\infty)^2$ such that $\sigma_-<\sigma_+$ and for any $(x,x')\in\mathbb{X}^2$, $\sigma_-\le m(x,x')\le\sigma_+$ and we set $\rho\stackrel{\text{def}}{=}1-\sigma_-/\sigma_+$.

(ii) There exists $c_-\in\mathbb{R}_+^*$ such that $\int\chi(\mathrm{d}x)g_0(x)\ge c_-$ and for any $t\in\mathbb{N}^*$, $\inf_{x\in\mathbb{X}}\int M(x,\mathrm{d}x')g_t(x')\ge c_-$.

A2 (i) For all $t\ge0$ and all $x\in\mathbb{X}$, $g_t(x)>0$.

(ii) $\sup_{t\ge0}|g_t|_\infty<\infty$.

A3 $\sup_{t\ge1}|\vartheta_t|_\infty<\infty$, $\sup_{t\ge0}|p_t|_\infty<\infty$ and $\sup_{t\ge0}|\omega_t|_\infty<\infty$ where

$$\omega_0(x)\stackrel{\text{def}}{=}\frac{d\chi}{d\rho_0}(x)g_0(x),\quad \omega_t(x,x')\stackrel{\text{def}}{=}\frac{m(x,x')g_t(x')}{\vartheta_t(x)p_t(x,x')},\forall t\ge1\,.$$

Assumptions A1 and A2 give bounds for the model and assumption A3 for quantities related to the algorithm. A1(i), referred to as the strong mixing condition, is crucial to derive time-uniform exponential deviation inequalities and a time-uniform bound of the variance of the marginal smoothing distribution (see Del Moral and Guionnet [2001] and Douc et al. [2010]). For all function h from a space E to $\mathbb{R}$, $\mathrm{osc}(h)$ is defined by:

$$\mathrm{osc}(h)\stackrel{\text{def}}{=}\sup_{(z,z')\in\mathrm{E}^2}|h(z)-h(z')|\,.$$

Theorem 4.1. *Assume A1–3. For all $q\ge2$, there exists a constant C (depending only on q, σ_-, σ_+, c_-, $\sup_{t\ge1}|\vartheta_t|_\infty$ and $\sup_{t\ge0}|\omega_t|_\infty$) such that for any $T<\infty$, any integer r and any bounded and measurable functions $\{h_s\}_{s=r}^T$,*

$$\left\|\phi_{0:T|T}^N[S_{T,r}]-\phi_{0:T|T}[S_{T,r}]\right\|_q\le\frac{C}{\sqrt{N}}\Upsilon_{r,T}^N\left(\sum_{s=r}^T\mathrm{osc}(h_s)^2\right)^{1/2}\,,$$

where $S_{T,r}$ is defined by (4.6)*, $\phi^N_{0:T|T}$ is defined by* (4.3) *and where*

$$\Upsilon^N_{r,T} \stackrel{\text{def}}{=} \sqrt{r+1}\left(\sqrt{1+r}\wedge\sqrt{T-r+1}+\frac{\sqrt{1+r}\sqrt{T-r+1}}{\sqrt{N}}\right).$$

Similarly,

$$\left\|\widetilde{\phi}^N_{0:T|T}[S_{T,r}]-\phi_{0:T|T}[S_{T,r}]\right\|_q \leq \frac{C}{\sqrt{N}}\Upsilon^N_{r,T}\left(\sum_{s=r}^{T}\operatorname{osc}(h_s)^2\right)^{1/2},$$

where $\widetilde{\phi}^N_{0:T|T}$ is defined by (4.8)*.*

Remark 4.2. In the particular cases where $r=0$ and $r=T$, $\Upsilon^N_{0,T}=1+\sqrt{T+1/N}$ and $\Upsilon^N_{T,T}=\sqrt{T+1}(1+\sqrt{T+1/N})$. Then, Theorem 4.1 gives

$$\left\|\phi^N_{0:T|T}[S_{T,0}]-\phi_{0:T|T}[S_{T,0}]\right\|_q \leq C\frac{\left(\sum_{s=0}^{T}\operatorname{osc}(h_s)^2\right)^{1/2}}{\sqrt{N}}\left(1+\sqrt{\frac{T+1}{N}}\right),$$

and

$$\left\|\phi^N_{0:T|T}[S_{T,T}]-\phi_{0:T|T}[S_{T,T}]\right\|_q \leq C\sqrt{\frac{T+1}{N}}\left(1+\sqrt{\frac{T+1}{N}}\right)\operatorname{osc}(h_T)^2.$$

As stated in Section 4.1, theses bounds improve the results given in Del Moral et al. [2010a] for the FFBS estimator.

Remark 4.3. The dependence on $1/\sqrt{N}$ is hardly surprising. Under the stated strong mixing condition, it is known that the L_q-norm of the marginal smoothing estimator $\phi^N_{t-r:t|T}[h]$, $t\in\{r,\ldots,T\}$ is uniformly bounded in time by $\left\|\phi^N_{t-r:t|T}[h]\right\|_q \leq C\operatorname{osc}(h)N^{-1/2}$ (where C depends only on q, σ_-, σ_+, c_-, $\sup_{t\geq 1}|\vartheta_t|_\infty$ and $\sup_{t\geq 0}|\omega_t|_\infty$). The dependence in $\sqrt{T}$ instead of T reflects the forgetting property of the filter and the backward smoother. As for $r\leq s<t\leq T$, the estimators $\phi^N_{s-r:s|T}[h_s]$ and $\phi^N_{t-r:t|T}[h_t]$ become asymptotically independent as $(t-s)$ gets large, the L_q-norm of the sum $\sum_{t=r}^{T}\phi^N_{t-r:t|T}[h_t]$ scales as the sum of a mixing sequence (see Davidson [1997]).

Remark 4.4. It is easy to see that the scaling in $\sqrt{T/N}$ cannot in general be improved. Assume that the kernel m satisfies $m(x,x')=m(x')$ for all $(x,x')\in\mathbb{X}\times\mathbb{X}$. In this case, for any $t\in\{0,\ldots,T\}$, the filtering distribution is

$$\phi_t[h_t]=\frac{\int m(x)g_t(x)h_t(x)\mathrm{d}x}{\int m(x)g_t(x)\mathrm{d}x},$$

and the backward kernel is the identity kernel. Hence, the fixed-interval smoothing distribution coincides with the filtering distribution. If we assume that we apply the bootstrap filter for which $p_s(x,x')=m(x')$ and $\vartheta_s(x)=1$, the estimators $\{\phi^N_{t|T}[h_t]\}_{t\in\{0,\ldots,T\}}$ are independent random variables corresponding to importance sampling estimators. It is easily seen that

$$\left\|\sum_{t=0}^{T}\phi^N_t[h_t]-\phi_t[h_t]\right\|_q \leq C\max_{0\leq t\leq T}\{\operatorname{osc}(h_t)\}\sqrt{\frac{T}{N}}.$$

Remark 4.5. The independent case also clearly illustrates why the path-space methods are sub-optimal (see also Briers et al. [2010] for a discussion). When applied to the independent case (for all $(x,x')\in\mathbb{X}\times\mathbb{X}$, $m(x,x')=m(x')$ and $p_s(x,x')=m(x')$), the asymptotic variance of the path-space estimators is given in Del Moral [2004] by

$$\begin{aligned}\Gamma_{0:T|T}[S_{T,0}] \stackrel{\text{def}}{=} &\sum_{t=0}^{T-1}\frac{m(g_T^2)}{m(g_T)^2}\frac{m(g_t[h_t-\phi_t(h_t)]^2)}{m(g_t)}+\frac{m(g_T^2[h_T-\phi_T(h_T)]^2)}{m(g_T)^2}\\ &+\sum_{t=1}^{T-1}\left\{\sum_{s=0}^{t-1}\frac{m(g_t^2)}{m(g_t)^2}\frac{m(g_s[h_s-\phi_s(h_s)]^2)}{m(g_s)}+\frac{m(g_t^2[h_t-\phi_t(h_t)]^2)}{m(g_t)^2}\right\}+\frac{\chi(g_0^2[h_0-\phi_0(h_0)]^2)}{\chi(g_0)^2}.\end{aligned}$$

The asymptotic variance thus increases as T^2 and hence, under the stated assumptions, the variance of the path-space methods is of order T^2/N. It is believed (and proved in some specific scenarios) that the same scaling holds for path-space methods for non-degenerated Markov kernel (the result has been formally established for strongly mixing kernel under the assumption that σ_-/σ_+ is sufficiently close to 1).

We provide below a brief outline of the main steps of the proofs (a detailed proof is given in Section 4.5). Following Douc et al. [2010], the proofs rely on a decomposition of the smoothing error. For all $0 \leq t \leq T$ and all bounded and measurable function h on $\mathbb{X}^{T+1}$ define the kernel $\mathrm{L}_{t,T} : \mathbb{X}^{t+1} \times \mathcal{X}^{\otimes T+1} \to [0,1]$ by

$$\mathrm{L}_{t,T} h(x_{0:t}) \stackrel{\text{def}}{=} \int \prod_{u=t+1}^{T} M(x_{u-1}, \mathrm{d}x_u) g_u(x_u) h(x_{0:T}) \, .$$

The fixed-interval smoothing distribution may then be expressed, for all bounded and measurable function h on $\mathbb{X}^{T+1}$, by

$$\phi_{0:T|T}[h] = \frac{\phi_{0:t|t}\left[\mathrm{L}_{t,T} h\right]}{\phi_{0:t|t}\left[\mathrm{L}_{t,T} \mathbf{1}\right]} \, ,$$

and this suggests to decompose the smoothing error as follows

$$\begin{aligned} \Delta_T^N[h] &\stackrel{\text{def}}{=} \phi_{0:T|T}^N[h] - \phi_{0:T|T}[h] \\ &= \sum_{t=0}^{T} \frac{\phi_{0:t|t}^N\left[\mathrm{L}_{t,T} h\right]}{\phi_{0:t|t}^N\left[\mathrm{L}_{t,T} \mathbf{1}\right]} - \frac{\phi_{0:t-1|t-1}^N\left[\mathrm{L}_{t-1,T} h\right]}{\phi_{0:t-1|t-1}^N\left[\mathrm{L}_{t-1,T} \mathbf{1}\right]} \, , \end{aligned} \tag{4.9}$$

where we used the convention

$$\frac{\phi_{0:-1|-1}^N\left[\mathrm{L}_{-1,T} h\right]}{\phi_{0:-1|-1}^N\left[\mathrm{L}_{-1,T} \mathbf{1}\right]} = \frac{\phi_0\left[\mathrm{L}_{0,T} h\right]}{\phi_0\left[\mathrm{L}_{0,T} \mathbf{1}\right]} = \phi_{0:T|T}[h] \, .$$

Furthermore, for all $0 \leq t \leq T$,

$$\begin{aligned} \phi_{0:t|t}^N\left[\mathrm{L}_{t,T} h\right] &= \int \phi_{0:t|t}^N(\mathrm{d}x_{0:t}) \mathrm{L}_{t,T} h(x_{0:t}) \\ &= \int \phi_t^N(\mathrm{d}x_t) \mathrm{B}_{\phi_{t-1}^N}(x_t, \mathrm{d}x_{t-1}) \cdots \mathrm{B}_{\phi_0^N}(x_1, \mathrm{d}x_0) \mathrm{L}_{t,T} h(x_{0:t}) \\ &= \int \phi_t^N(\mathrm{d}x_t) \mathcal{L}_{t,T}^N h(x_t) \, , \end{aligned}$$

where $\mathcal{L}_{t,T}^N$ and $\mathcal{L}_{t,T}$ are two kernels on $\mathbb{X} \times \mathcal{X}^{\otimes(T+1)}$ defined for all $x_t \in \mathbb{X}$ by

$$\mathcal{L}_{t,T} h(x_t) \stackrel{\text{def}}{=} \int \mathrm{B}_{\phi_{t-1}}(x_t, \mathrm{d}x_{t-1}) \cdots \mathrm{B}_{\phi_0}(x_1, \mathrm{d}x_0) \mathrm{L}_{t,T} h(x_{0:t}) \tag{4.10}$$

$$\mathcal{L}_{t,T}^N h(x_t) \stackrel{\text{def}}{=} \int \mathrm{B}_{\phi_{t-1}^N}(x_t, \mathrm{d}x_{t-1}) \cdots \mathrm{B}_{\phi_0^N}(x_1, \mathrm{d}x_0) \mathrm{L}_{t,T} h(x_{0:t}) \, . \tag{4.11}$$

For all $1 \leq t \leq T$ we can write

$$\begin{aligned} \frac{\phi_{0:t|t}^N[\mathrm{L}_{t,T} h]}{\phi_{0:t|t}^N[\mathrm{L}_{t,T} \mathbf{1}]} - \frac{\phi_{0:t-1|t-1}^N[\mathrm{L}_{t-1,T} h]}{\phi_{0:t-1|t-1}^N[\mathrm{L}_{t-1,T} \mathbf{1}]} &= \frac{\phi_t^N[\mathcal{L}_{t,T}^N h]}{\phi_t^N[\mathcal{L}_{t,T}^N \mathbf{1}]} - \frac{\phi_{t-1}^N[\mathcal{L}_{t-1,T}^N h]}{\phi_{t-1}^N[\mathcal{L}_{t-1,T}^N \mathbf{1}]} \\ &= \frac{1}{\phi_t^N[\mathcal{L}_{t,T}^N \mathbf{1}]} \left(\phi_t^N[\mathcal{L}_{t,T}^N h] - \frac{\phi_{t-1}^N[\mathcal{L}_{t-1,T}^N h]}{\phi_{t-1}^N[\mathcal{L}_{t-1,T}^N \mathbf{1}]} \phi_t^N[\mathcal{L}_{t,T}^N \mathbf{1}] \right) \, , \end{aligned}$$

and then,

$$\Delta_T^N[h] = \sum_{t=0}^{T} \frac{N^{-1} \sum_{\ell=1}^N \omega_t^{N,\ell} G_{t,T}^N h(\xi_t^{N,\ell})}{N^{-1} \sum_{\ell=1}^N \omega_t^{N,\ell} \mathcal{L}_{t,T} \mathbf{1}(\xi_t^{N,\ell})} \, , \tag{4.12}$$

with $G^N_{t,T}$ is a kernel on $\mathbb{X}\times\mathcal{X}^{\otimes(T+1)}$ defined, for all $x_t\in\mathbb{X}$ and all bounded and measurable function h on $\mathbb{X}^{T+1}$, by

$$G^N_{t,T}h(x_t)\stackrel{\text{def}}{=}\mathcal{L}^N_{t,T}h(x_t)-\frac{\phi^N_{t-1}[\mathcal{L}^N_{t-1,T}h]}{\phi^N_{t-1}[\mathcal{L}^N_{t-1,T}\mathbf{1}]}\mathcal{L}^N_{t,T}\mathbf{1}(x_t)\;,$$

where, by the same convention as above,

$$G^N_{0,T}h(x_0)\stackrel{\text{def}}{=}\mathrm{L}_{0,T}h(x_0)-\frac{\phi_0[\mathcal{L}_{0,T}h]}{\phi_0[\mathcal{L}_{0,T}\mathbf{1}]}\mathcal{L}_{0,T}\mathbf{1}(x_0)\;.$$

Two families of random variables $\left\{C^N_{t,T}(f)\right\}_{t=0}^T$ and $\left\{D^N_{t,T}(f)\right\}_{t=0}^T$ are now introduced to transform (4.12) into a suitable decomposition to compute an upper bound for the L_q-mean error. As shown in Lemma 4.1, the random variables $\{\omega^{N,\ell}_t G^N_{t,T}f(\xi^{N,\ell}_t)\}_{\ell=1}^N$ are centered given $\mathcal{F}^N_{t-1}$. The idea is to replace $N^{-1}\sum_{\ell=1}^N\omega^{N,\ell}_t\mathcal{L}_{t,T}\mathbf{1}(\xi^{N,\ell}_t)$ in (4.12) by its conditional expectation given $\mathcal{F}^N_{t-1}$ to get a martingale difference. This conditional expectation is computed using the following intermediate result. For any measurable function h on $\mathbb{X}$ and any $t\in\{0,\dots,T\}$,

$$\mathbb{E}\left[\omega^{N,1}_t h(\xi^{N,1}_t)\middle|\mathcal{F}^N_{t-1}\right]=\frac{\phi^N_{t-1}[Mg_th]}{\phi^N_{t-1}[\vartheta_t]}\;. \tag{4.13}$$

Indeed,

$$\begin{aligned}
&\mathbb{E}\left[\omega^{N,1}_t h(\xi^{N,1}_t)\middle|\mathcal{F}^N_{t-1}\right]\\
&=\mathbb{E}\left[\frac{m(\xi^{N,I^{N,1}_t}_{t-1},\xi^{N,1}_t)g_t(\xi^{N,1}_t)}{\vartheta_t(\xi^{N,I^{N,1}_t}_{t-1})p_t(\xi^{N,I^{N,1}_t}_{t-1},\xi^{N,1}_t)}h(\xi^{N,1}_t)\middle|\mathcal{F}^N_{t-1}\right]\\
&=\left(\sum_{i=1}^N\omega^{N,i}_{t-1}\vartheta_t(\xi^{N,i}_{t-1})\right)^{-1}\sum_{i=1}^N\int\omega^{N,i}_{t-1}\vartheta_t(\xi^{N,i}_{t-1})p_t(\xi^{N,i}_{t-1},x)\frac{M(\xi^{N,i}_{t-1},\mathrm{d}x)g_t(x)}{\vartheta_t(\xi^{N,i}_{t-1})p_t(\xi^{N,i}_{t-1},x)}h(x)\\
&=\left(\sum_{i=1}^N\omega^{N,i}_{t-1}\vartheta_t(\xi^{N,i}_{t-1})\right)^{-1}\sum_{i=1}^N\int\omega^{N,i}_{t-1}M(\xi^{N,i}_{t-1},\mathrm{d}x)g_t(x)h(x)\\
&=\frac{\phi^N_{t-1}[Mg_th]}{\phi^N_{t-1}[\vartheta_t]}\;.
\end{aligned}$$

This result, applied with the function $h=\mathcal{L}_{t,T}\mathbf{1}$, yields

$$\mathbb{E}\left[\omega^{N,1}_t\mathcal{L}_{t,T}\mathbf{1}(\xi^{N,1}_t)\middle|\mathcal{F}^N_{t-1}\right]=\frac{\phi^N_{t-1}[Mg_t\mathcal{L}_{t,T}\mathbf{1}]}{\phi^N_{t-1}[\vartheta_t]}=\frac{\phi^N_{t-1}[\mathcal{L}_{t-1,T}\mathbf{1}]}{\phi^N_{t-1}[\vartheta_t]}\;.$$

For any $0\le t\le T$, define for all bounded and measurable function h on $\mathbb{X}^{T+1}$,

$$\begin{aligned}
D^N_{t,T}(h)&\stackrel{\text{def}}{=}\mathbb{E}\left[\omega^{N,1}_t\frac{\mathcal{L}_{t,T}\mathbf{1}(\xi^{N,1}_t)}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}\middle|\mathcal{F}^N_{t-1}\right]^{-1}N^{-1}\sum_{\ell=1}^N\omega^{N,\ell}_t\frac{G^N_{t,T}h(\xi^{N,\ell}_t)}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}\\
&=\frac{\phi^N_{t-1}[\vartheta_t]}{\phi^N_{t-1}\left[\frac{\mathcal{L}_{t-1,T}\mathbf{1}}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}\right]}N^{-1}\sum_{\ell=1}^N\omega^{N,\ell}_t\frac{G^N_{t,T}h(\xi^{N,\ell}_t)}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}\;,
\end{aligned} \tag{4.14}$$

$$C^N_{t,T}(h)\stackrel{\text{def}}{=}\left[\frac{1}{N^{-1}\sum_{i=1}^N\omega^{N,i}_t\frac{\mathcal{L}_{t,T}\mathbf{1}(\xi^{N,i}_t)}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}}-\frac{\phi^N_{t-1}[\vartheta_t]}{\phi^N_{t-1}\left[\frac{\mathcal{L}_{t-1,T}\mathbf{1}}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}\right]}\right]\times N^{-1}\sum_{\ell=1}^N\omega^{N,\ell}_t\frac{G^N_{t,T}h(\xi^{N,\ell}_t)}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}\;. \tag{4.15}$$

Using these notations, (4.12) can be rewritten as follows:

$$\Delta_T^N[h] = \sum_{t=0}^{T} D_{t,T}^N(h) + \sum_{t=0}^{T} C_{t,T}^N(h) \; . \tag{4.16}$$

For any $q \geq 2$, the derivation of the upper bound relies on the triangle inequality:

$$\left\|\Delta_T^N[S_{T,r}]\right\|_q \leq \left\|\sum_{t=0}^{T} D_{t,T}^N(S_{T,r})\right\|_q + \sum_{t=0}^{T} \left\|C_{t,T}^N(S_{T,r})\right\|_q \; ,$$

where $S_{T,r}$ is defined in (4.6). The proof for the FFBS estimator $\phi_{0:T|T}^N$ is completed by using Proposition 4.1 and Proposition 4.2. According to (4.16), the smoothing error can be decomposed into a sum of two terms which are considered separately. The first one is a martingale whose L_q-mean error is upper-bounded by $\sqrt{(T+1)/N}$ as shown in Proposition 4.1. The second one is a sum of products, L_q-norm of which being bounded by $1/N$ in Proposition 4.2.

The end of this section is devoted to the exponential deviation inequality for the error $\Delta_T^N[S_{T,r}]$ defined by (4.9). We use the decomposition of $\Delta_T^N[S_{T,r}]$ obtained in (4.16) leading to a similar dependence on the ratio $(T+1)/N$. The martingale term $D_{t,T}^N(S_{T,r})$ is dealt with using the Azuma-Hoeffding inequality while the term $C_{t,T}^N(S_{T,r})$ needs a specific Hoeffding-type inequality for ratio of random variables.

Theorem 4.2. *Assume A1–3. There exists a constant C (depending only on σ_-, σ_+, c_-, $\sup_{t\geq 1}|\vartheta_t|_\infty$ and $\sup_{t\geq 0}|\omega_t|_\infty$) such that for any $T < \infty$, any $N \geq 1$, any $\varepsilon > 0$, any integer r, and any bounded and measurable functions $\{h_s\}_{s=r}^T$,*

$$\mathbb{P}\left\{\left|\phi_{0:T|T}[S_{T,r}] - \phi_{0:T|T}^N[S_{T,r}]\right| > \varepsilon\right\} \leq 2\exp\left(-\frac{CN\varepsilon^2}{\Theta_{r,T}\sum_{s=r}^T \mathrm{osc}(h_s)^2}\right) + 8\exp\left(-\frac{CN\varepsilon}{(1+r)\sum_{s=r}^T \mathrm{osc}(h_s)}\right) \; ,$$

where $S_{T,r}$ is defined by (4.6)*, $\phi_{0:T|T}^N$ is defined by* (4.3) *and where*

$$\Theta_{r,T} \stackrel{\mathrm{def}}{=} (1+r)\left\{(1+r) \wedge (T-r+1)\right\} \; . \tag{4.17}$$

Similarly,

$$\mathbb{P}\left\{\left|\phi_{0:T|T}[S_{T,r}] - \widetilde{\phi}_{0:T|T}^N[S_{T,r}]\right| > \varepsilon\right\} \leq 4\exp\left(-\frac{CN\varepsilon^2}{\Theta_{r,T}\sum_{s=r}^T \mathrm{osc}(h_s)^2}\right) + 8\exp\left(-\frac{CN\varepsilon}{(1+r)\sum_{s=r}^T \mathrm{osc}(h_s)}\right) \; ,$$

where $\widetilde{\phi}_{0:T|T}^N$ is defined by (4.8)*.*

4.4 Monte-Carlo Experiments

In this section, the performance of the FFBSi algorithm is evaluated through simulations and compared to the path-space method.

4.4.1 Linear gaussian model

Let us consider the following model:

$$\begin{cases} X_{t+1} &= \phi X_t + \sigma_u U_t \; , \\ Y_t &= X_t + \sigma_v V_t \; , \end{cases}$$

where X_0 is a zero-mean random variable with variance $\frac{\sigma_u^2}{1-\phi^2}$, $\{U_t\}_{t\geq 0}$ and $\{V_t\}_{t\geq 0}$ are two sequences of independent and identically distributed standard gaussian random variables (independent from X_0). The

Table 4.1: Empirical variance for different values of T and N.

Path-space

T \ N	300	500	750	1000	1500	5000	10000	15000	20000
300	137.8	119.4	63.7	46.1	36.2	12.8	7.1	3.8	3.0
500	290.0	215.3	192.5	161.9	80.3	30.1	14.9	11.3	7.4
750	474.9	394.5	332.9	250.5	206.8	71.0	35.6	24.4	21.7
1000	673.7	593.2	505.1	483.2	326.4	116.4	70.8	37.9	34.6
1500	1274.6	1279.7	916.7	804.7	655.1	233.9	163.1	89.7	80.0

FFBSi

T \ N	300	500	750	1000	1500
300	5.1	3.1	2.3	1.4	1.0
500	9.7	5.1	3.7	2.6	2.2
750	11.2	7.1	4.9	3.7	2.6
1000	16.5	10.5	6.7	5.1	3.4
1500	25.6	14.1	7.8	6.8	5.1

parameters $(\phi, \sigma_u, \sigma_v)$ are assumed to be known. Observations were generated using $\phi = 0.9$, $\sigma_u = 0.6$ and $\sigma_v = 1$. Table 4.1 provides the empirical variance of the estimation of the unnormalized smoothed additive functional $\imath_T \stackrel{\text{def}}{=} \sum_{t=0}^{T} \mathbb{E}[X_t | Y_{0:T}]$ given by the path-space and the FFBSi methods over 250 independent Monte Carlo experiments. We display in Figure 4.1 the empirical variance for different values of N as a function of T for both estimators. These estimates are represented by dots and a linear regression (resp. quadratic regression) is also provided for the FFBSi algorithm (resp. for the path-space method).

In Figure 4.2 the FFBSi algorithm is compared to the path-space method to compute the smoothed value of the empirical mean $(T+1)^{-1}\imath_T$. For the purpose of comparison, this quantity is computed using the Kalman smoother. We display in Figure 4.2 the box and whisker plots of the estimations obtained with 100 independent Monte Carlo experiments. The FFBSi algorithm clearly outperforms the other method for comparable computational costs. In Table 4.2, the mean CPU times over the 100 runs of the two methods are given as a function of the number of particles (for $T = 500$ and $T = 1000$).

Table 4.2: Average CPU time to compute the smoothed value of the empirical mean in the LGM

$T = 500$	FFBSi	Path-space method		
N	500	500	5000	10000
CPU time (s)	4.87	0.24	2.47	4.65

$T = 1000$	FFBSi	Path-space method		
N	1000	1000	10000	20000
CPU time (s)	16.5	0.9	8.5	17.2

4.4.2 Stochastic Volatility Model

Stochastic volatility models (SVM) have been introduced to provide better ways of modeling financial time series data than ARCH/GARCH models (Hull and White [1987]). We consider the elementary SVM

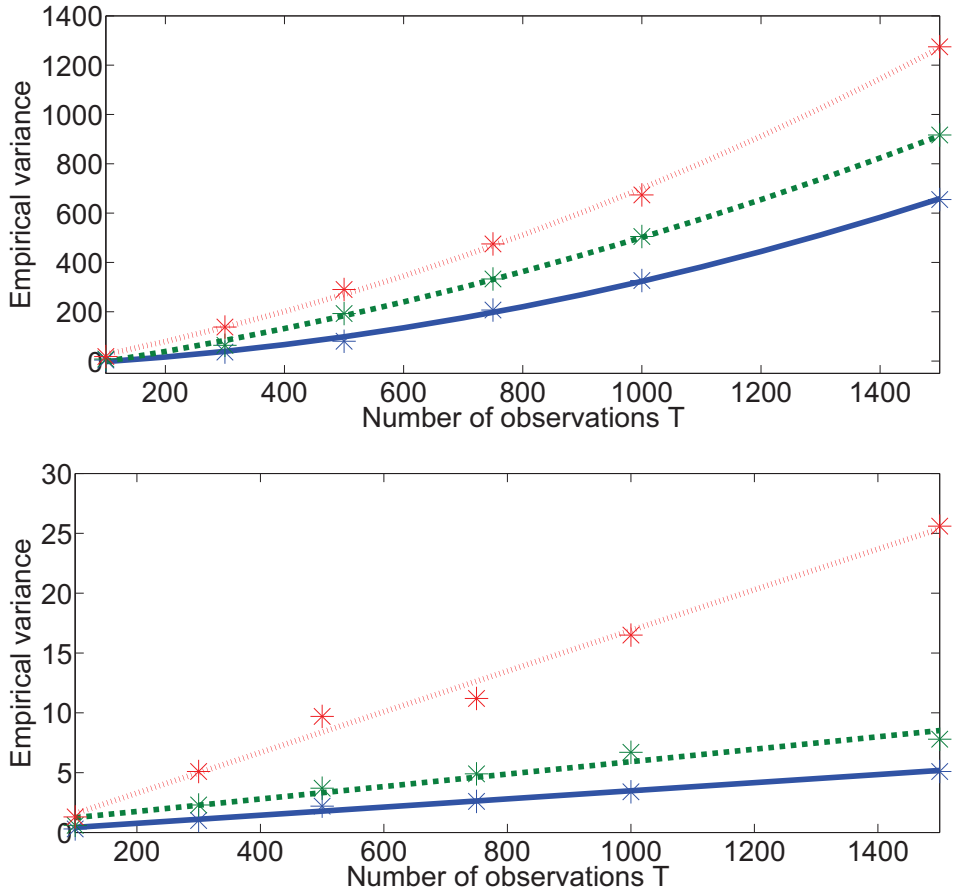


Figure 4.1: Empirical variance of the path-space (top) and FFBSi (bottom) for $N = 300$ (dotted line), $N = 750$ (dashed line) and $N = 1500$ (bold line).

model introduced by Hull and White [1987]:

$$\begin{cases} X_{t+1} = \phi X_t + \sigma U_{t+1} \,, \\ Y_t = \beta e^{\frac{X_t}{2}} V_t \,, \end{cases}$$

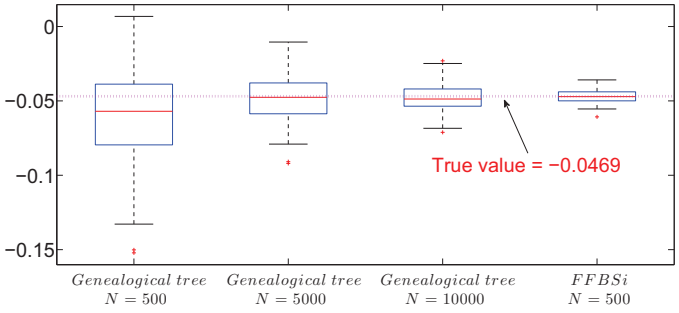
where X_0 is a zero-mean random variable with variance $\frac{\sigma_u^2}{1-\phi^2}$, $\{U_t\}_{t\geq 0}$ and $\{V_t\}_{t\geq 0}$ are two sequences of independent and identically distributed standard gaussian random variables (independent from X_0). This model was used to generate simulated data with parameters $(\phi = 0.3, \sigma = 0.5, \beta = 1)$ assumed to be known in the following experiments. The empirical variance of the estimation of $\imath_T$ given by the path-space and the FFBSi methods over 250 independent Monte Carlo experiments is displayed in Table 4.3. We display in Figure 4.3 the empirical variance for different values of N as a function of T for both estimators.

4.5 Proof of Theorem 4.1

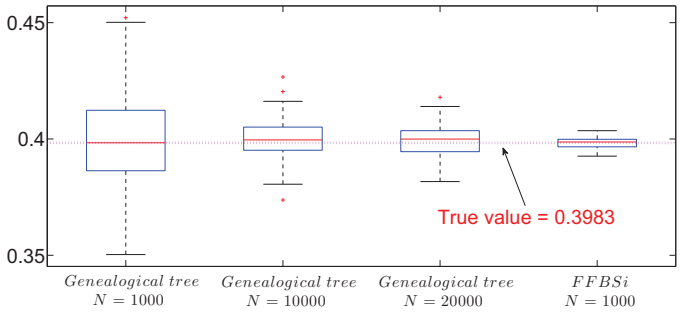
We preface the proof of Proposition 4.1 by the following Lemma:

Lemma 4.1. *Under assumptions A1–3, we have, for any $t \in \{0, \ldots, T\}$ and any measurable function h on $\mathbb{X}^{T+1}$:*

(i) The random variables $\left\{ \omega_t^{N,\ell} \dfrac{G_{t,T}^N h(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \right\}_{\ell=1}^{N}$ *are, for all $N \in \mathbb{N}$:*



(a) Time $T = 500$



(b) Time $T = 1000$

Figure 4.2: Computation of smoothed additive functionals in a linear gaussian model. The variance of the estimation given by the FFBSi algorithm is the smallest one in both cases.

Table 4.3: Empirical variance for different values of T and N in the SVM.

Path-space method

T \ N	300	500	750	1000	1500	5000	10000	15000	20000
300	52.7	33.7	22.0	17.8	12.3	3.8	2.0	1.4	1.2
500	116.3	84.8	64.8	53.5	30.7	11.4	6.8	4.1	2.8
750	184.7	187.6	134.2	120.0	65.8	29.1	12.8	7.3	7.7
1000	307.7	240.4	244.7	182.8	133.2	43.6	24.5	15.6	11.6
1500	512.1	487.5	445.5	359.9	249.5	90.9	52.0	32.6	29.3

FFBSi

T \ N	300	500	750	1000	1500
300	1.2	0.6	0.5	0.4	0.2
500	2.1	1.2	0.8	0.6	0.4
750	3.7	1.8	1.4	0.9	0.6
1000	4.0	2.7	1.8	1.3	0.9
1500	7.3	3.8	3.1	1.6	1.4

(a) conditionally independent and identically distributed given $\mathcal{F}_{t-1}^N$,

(b) centered conditionally to $\mathcal{F}_{t-1}^N$.

where $G_{t,T}^N h$ *is defined in* (4.3) *and* $\mathcal{L}_{t,T}^N$ *is defined in* (4.11).

(ii) For any integers r, t and N:

$$\left| \frac{G_{t,T}^N S_{T,r}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \right| \leq \sum_{s=r}^{T} \rho^{\max(t-s,s-r-t,0)} \operatorname{osc}(h_s) \,, \tag{4.18}$$

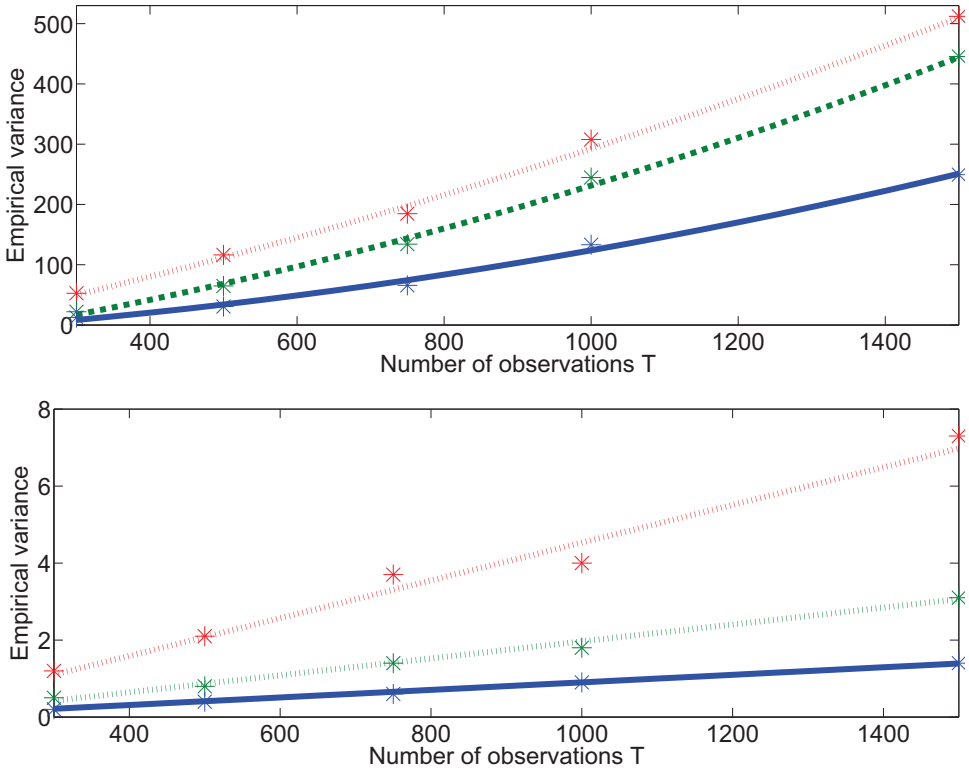


Figure 4.3: Empirical variance of the path-space (top) and FFBSi (bottom) for $N = 300$ (dotted line), $N = 750$ (dashed line) and $N = 1500$ (bold line) in the SVM.

where $S_{T,r}$ *and* ρ *are respectively defined in* (4.6) *and in A1*(i).

(iii) For all $x \in \mathbb{X}$, $\dfrac{\mathcal{L}_{t,T}\mathbf{1}(x)}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \geq \dfrac{\sigma_-}{\sigma_+}$ *and* $\dfrac{\mathcal{L}_{t-1,T}\mathbf{1}(x)}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \geq c_- \dfrac{\sigma_-}{\sigma_+}$.

Proof. The proof of (i) is given by [Douc et al., 2010, Lemma 3].

Proof of (ii). Let $\Pi_{s-r:s,T}$ be the operator which associates to any bounded and measurable function h on $\mathbb{X}^{r+1}$ the function $\Pi_{s-r:s,T}h$ given, for any $(x_0,\ldots,x_T) \in \mathbb{X}^{T+1}$, by

$$\Pi_{s-r:s,T}h(x_{0:T}) \stackrel{\text{def}}{=} h(x_{s-r:s}) .$$

Then, we may write $S_{T,r} = \sum_{s=r}^{T} \Pi_{s-r:s,T}h_s$ and $G_{t,T}^N S_{T,r} = \sum_{s=r}^{T} G_{t,T}^N \Pi_{s-r:s,T}h_s$. By (4.3), we have

$$\frac{G_{t,T}^N \Pi_{s-r:s,T}h_s(x_t)}{\mathcal{L}_{t,T}^N\mathbf{1}(x_t)} = \frac{\mathcal{L}_{t,T}^N \Pi_{s-r:s,T}h_s(x_t)}{\mathcal{L}_{t,T}^N\mathbf{1}(x_t)} - \frac{\phi_{t-1}^N[\mathcal{L}_{t-1,T}^N \Pi_{s-r:s,T}h_s]}{\phi_{t-1}^N[\mathcal{L}_{t-1,T}^N\mathbf{1}]} ,$$

and, following the same lines as in [Douc et al., 2010, Lemma 10],

$$|G_{t,T}^N \Pi_{s-r:s,T}h_s|_\infty \leq \rho^{s-r-t}\mathrm{osc}(h_s)|\mathcal{L}_{t,T}\mathbf{1}|_\infty \quad \text{if} \quad t \leq s-r ,$$

$$|G_{t,T}^N \Pi_{s-r:s,T}h_s|_\infty \leq \rho^{t-s}\mathrm{osc}(h_s)|\mathcal{L}_{t,T}\mathbf{1}|_\infty \quad \text{if} \quad t > s ,$$

where ρ is defined in A1(i). Furthermore, for any $s-r < t \leq s$,

$$|G_{t,T}^N \Pi_{s-r:s,T}h_s|_\infty \leq \mathrm{osc}(h_s)|\mathcal{L}_{t,T}\mathbf{1}|_\infty ,$$

which shows (ii).
Proof of (iii). From the definition (4.10), for all $x \in \mathbb{X}$ and all $t \in \{1,\dots,T\}$,

$$\mathcal{L}_{t,T}\mathbf{1}(x) = \int m(x,x_{t+1}) g_{t+1}(x_{t+1}) \prod_{u=t+2}^{T} M(x_{u-1},\mathrm{d}x_u) g_u(x_u) \lambda(\mathrm{d}x_{t+1}) \,,$$

hence, by assumption A1,

$$|\mathcal{L}_{t,T}\mathbf{1}|_\infty \le \sigma_+ \int g_{t+1}(x_{t+1}) \mathcal{L}_{t+1,T}\mathbf{1}(x_{t+1}) \lambda(\mathrm{d}x_{t+1})$$
$$\mathcal{L}_{t,T}\mathbf{1}(x) \ge \sigma_- \int g_{t+1}(x_{t+1}) \mathcal{L}_{t+1,T}\mathbf{1}(x_{t+1}) \lambda(\mathrm{d}x_{t+1}) \,,$$

which concludes the proof of the first statement. By construction, for any $x \in \mathbb{X}$ and any $t \in \{1,\dots,T\}$,

$$\mathcal{L}_{t-1,T}\mathbf{1}(x) = \int M(x,\mathrm{d}x') g_t(x') \mathcal{L}_{t,T}\mathbf{1}(x') \,,$$

and then, by assumption A1,

$$\frac{\mathcal{L}_{t-1,T}\mathbf{1}(x)}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} = \int M(x,\mathrm{d}x') g_t(x') \frac{\mathcal{L}_{t,T}\mathbf{1}(x')}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \ge c_- \frac{\sigma_-}{\sigma_+} \,.$$

□

Proposition 4.1. *Assume A1–3. For all $q \ge 2$, there exists a constant C (depending only on q, σ_-, σ_+, c_-, $\sup_{t\ge1}|\vartheta_t|_\infty$ and $\sup_{t\ge0}|\omega_t|_\infty$) such that for any $T < \infty$, any integer r and any bounded and measurable functions $\{h_s\}_{s=r}^T$ on $\mathbb{X}^{r+1}$,*

$$\left\| \sum_{t=0}^{T} D_{t,T}^N(S_{T,r}) \right\|_q \le \frac{C}{\sqrt{N}} \sqrt{1+r} \left(\sqrt{1+r} \wedge \sqrt{T-r+1} \right) \left(\sum_{s=r}^{T} \mathrm{osc}(h_s)^2 \right)^{1/2} , \tag{4.19}$$

where $D_{t,T}^N$ is defined in (4.14).

Proof. Since $\left\{ D_{t,T}^N(S_{T,r}) \right\}_{0\le t\le T}$ is a is a forward martingale difference and $q \ge 2$, Burkholder's inequality (see [Hall and Heyde, 1980, Theorem 2.10, page 23]) states the existence of a constant C depending only on q such that:

$$\mathbb{E}\left[\left| \sum_{t=0}^{T} D_{t,T}^N(S_{T,r}) \right|^q \right] \le C\mathbb{E}\left[\left| \sum_{t=0}^{T} D_{t,T}^N(S_{T,r})^2 \right|^{\frac{q}{2}} \right] .$$

Moreover, by application of the last statement of Lemma 4.1(iii),

$$\frac{\phi_{t-1}^N[\vartheta_t]}{\phi_{t-1}^N\left[\frac{\mathcal{L}_{t-1,T}\mathbf{1}}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}\right]} \le \frac{\sigma_+ \sup_{t\ge0}|\vartheta_t|_\infty}{\sigma_- c_-} \,,$$

and thus,

$$\mathbb{E}\left[\left| \sum_{t=0}^{T} D_{t,T}^N(S_{T,r})^2 \right|^{\frac{q}{2}} \right] \le \left(\frac{\sigma_+ \sup_{t\ge0}|\vartheta_t|_\infty}{\sigma_- c_-} \right)^q \mathbb{E}\left[\left| \sum_{t=0}^{T} \left(N^{-1} \sum_{\ell=1}^{N} a_{t,T}^{N,\ell} \right)^2 \right|^{\frac{q}{2}} \right] ,$$

where $a_{t,T}^{N,\ell} \stackrel{\text{def}}{=} \omega_t^{N,\ell} \frac{G_{t,T}^N S_{T,r}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}$. By the Minkowski inequality,

$$\left\| \sum_{t=0}^{T} D_{t,T}^N(S_{T,r}) \right\|_q \le C \left\{ \sum_{t=0}^{T} \left(\mathbb{E}\left[\left| N^{-1} \sum_{\ell=1}^{N} a_{t,T}^{N,\ell} \right|^q \right] \right)^{2/q} \right\}^{1/2} . \tag{4.20}$$

Since for any $t \geq 0$ the random variables $\left\{a_{t,T}^{N,\ell}\right\}_{\ell=1}^{N}$ are conditionally independent and centered conditionally to $\mathcal{F}_{t-1}^{N}$, using again the Burkholder and the Jensen inequalities we obtain

$$\mathbb{E}\left[\left|\sum_{\ell=1}^{N} a_{t,T}^{N,\ell}\right|^{q}\middle|\mathcal{F}_{t-1}^{N}\right] \leq CN^{q/2-1}\sum_{\ell=1}^{N}\mathbb{E}\left[\left|a_{t,T}^{N,\ell}\right|^{q}\middle|\mathcal{F}_{t-1}^{N}\right] \leq C\left[\sum_{s=r}^{T}\rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)\right]^{q} N^{q/2}\,, \quad (4.21)$$

where the last inequality comes from (4.18). Finally, by (4.20) and (4.21) we get

$$\left\|\sum_{t=0}^{T} D_{t,T}^{N}(S_{T,r})\right\|_{q} \leq CN^{-1/2}\left\{\sum_{t=0}^{T}\left(\sum_{s=r}^{T}\rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)\right)^{2}\right\}^{1/2}.$$

By the Holder inequality, we have

$$\begin{aligned}\sum_{s=r}^{T}\rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s) &\leq \left(\sum_{s=r}^{T}\rho^{\max(t-s,s-r-t,0)}\right)^{1/2}\times\left(\sum_{s=r}^{T}\rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)^2\right)^{1/2}\\ &\leq C\sqrt{1+r}\left(\sum_{s=r}^{T}\rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)^2\right)^{1/2},\end{aligned}$$

which yields

$$\left\|\sum_{t=0}^{T} D_{t,T}^{N}(S_{T,r})\right\|_{q} \leq CN^{-1/2}(1+r)\left(\sum_{s=r}^{T}\mathrm{osc}(h_s)^2\right)^{1/2}.$$

We obtain similarly

$$\left\|\sum_{t=0}^{T} D_{t,T}^{N}(S_{T,r})\right\|_{q} \leq CN^{-1/2}(1+r)^{1/2}\sum_{s=r}^{T}\mathrm{osc}(h_s)\,,$$

which concludes the proof. □

Proposition 4.2. *Assume A1–3. For all $q \geq 2$, there exists a constant C (depending only on q, σ_-, σ_+, c_-, $\sup_{t\geq 1}|\vartheta_t|_\infty$ and $\sup_{t\geq 0}|\omega_t|_\infty$) such that for any $T < +\infty$, any $0 \leq t \leq T$, any integer r, and any bounded and measurable functions $\{h_s\}_{s=r}^{T}$ on $\mathbb{X}^{r+1}$,*

$$\left\|C_{t,T}^{N}(S_{T,r})\right\|_{q} \leq \frac{C}{N}\sum_{s=r}^{T}\rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)\,, \quad (4.22)$$

where $C_{t,T}^{N}$ is defined in (4.15).

Proof. According to (4.15), $C_{t,T}^{N}(S_{T,r})$ can be written

$$C_{t,T}^{N}(S_{T,r}) = U_{t,T}^{N} V_{t,T}^{N} W_{t,T}^{N}\,, \quad (4.23)$$

where

$$\begin{aligned}U_{t,T}^{N} &= \frac{N^{-1}\sum_{\ell=1}^{N}\omega_t^{N,\ell}\frac{G_{t,T}^{N}S_{T,r}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}}{N^{-1}\Omega_t^{N}}\,,\\ V_{t,T}^{N} &= N^{-1}\sum_{\ell=1}^{N}\left(\mathbb{E}\left[\omega_t^{N,1}\frac{\mathcal{L}_{t,T}\mathbf{1}(\xi_t^{N,1})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}\middle|\mathcal{F}_{t-1}\right] - \omega_t^{N,\ell}\frac{\mathcal{L}_{t,T}\mathbf{1}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}\right),\\ W_{t,T}^{N} &= \frac{N^{-1}\Omega_t^{N}}{\mathbb{E}\left[\omega_t^{N,1}\frac{\mathcal{L}_{t,T}\mathbf{1}(\xi_t^{N,1})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}\middle|\mathcal{F}_{t-1}\right]N^{-1}\sum_{\ell=1}^{N}\omega_t^{N,\ell}\frac{\mathcal{L}_{t,T}\mathbf{1}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}}\,,\end{aligned}$$

and where Ω_t^N is defined by (4.5). Using the last statement of Lemma 4.1, we get the following bound:

$$\mathbb{E}\left[\omega_t^{N,1}\frac{\mathcal{L}_{t,T}\mathbf{1}(\xi_t^{N,1})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}\middle|\mathcal{F}_{t-1}\right]=\frac{\phi_{t-1}^N\left[\mathcal{L}_{t-1,T}\mathbf{1}/|\mathcal{L}_{t,T}\mathbf{1}|_\infty\right]}{\phi_{t-1}^N[\vartheta_t]}\geq\frac{c_-\sigma_-}{|\vartheta_t|_\infty\sigma_+}\,,$$

$$\frac{\mathcal{L}_{t,T}\mathbf{1}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}\geq\frac{\sigma_-}{\sigma_+}\,,$$

which implies

$$\left|W_{t,T}^N\right|\leq\left(\frac{\sigma_+}{\sigma_-}\right)^2\frac{|\vartheta_t|_\infty}{c_-}\,. \tag{4.24}$$

Then, $\left|C_{t,T}^N(S_{T,r})\right|\leq C\left|U_{t,T}^N\right|\left|V_{t,T}^N\right|$ and we can use the decomposition

$$U_{t,T}^NV_{t,T}^N=V_{t,T}^N\left[\frac{N^{-1}\sum_{\ell=1}^N a_{t,T}^{N,\ell}}{\mathbb{E}\left[\widetilde{\Omega}_t^N\middle|\mathcal{F}_{t-1}\right]}+\frac{N^{-1}\sum_{\ell=1}^N a_{t,T}^{N,\ell}}{\widetilde{\Omega}_t^N\mathbb{E}\left[\widetilde{\Omega}_t^N\middle|\mathcal{F}_{t-1}\right]}\left(\mathbb{E}\left[\widetilde{\Omega}_t^N\middle|\mathcal{F}_{t-1}\right]-\widetilde{\Omega}_t^N\right)\right]\,,$$

where $a_{t,T}^{N,\ell}\stackrel{\text{def}}{=}\omega_t^{N,\ell}\frac{G_{t,T}^NS_{T,r}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}$ and $\widetilde{\Omega}_t^N\stackrel{\text{def}}{=}N^{-1}\Omega_t^N$. By (4.13), $\mathbb{E}\left[\omega_t^{N,1}\middle|\mathcal{F}_{t-1}^N\right]=\frac{\phi_{t-1}^N[Mg_t]}{\phi_{t-1}^N[\vartheta_t]}$ and then, by A1(ii), A3 and (4.18),

$$\frac{1}{\mathbb{E}\left[\widetilde{\Omega}_t^N\middle|\mathcal{F}_{t-1}\right]}\leq\frac{|\vartheta_t|_\infty}{c_-}\quad\text{and}\quad\frac{N^{-1}\sum_{\ell=1}^N a_{t,T}^{N,\ell}}{\widetilde{\Omega}_t^N\mathbb{E}\left[\widetilde{\Omega}_t^N\middle|\mathcal{F}_{t-1}\right]}\leq C\frac{|\vartheta_t|_\infty}{c_-}\sum_{s=r}^T\rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)\,.$$

Therefore, $\left|C_{t,T}^N(S_{T,r})\right|\leq C\left(C_{t,T}^{1,N}+\sum_{s=r}^T\rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)C_{t,T}^{2,N}\right)$ where

$$C_{t,T}^{1,N}\stackrel{\text{def}}{=}V_{t,T}^N\cdot N^{-1}\sum_{\ell=1}^N a_{t,T}^{N,\ell}\quad\text{and}\quad C_{t,T}^{2,N}\stackrel{\text{def}}{=}V_{t,T}^N\left|\mathbb{E}\left[\widetilde{\Omega}_t^N\middle|\mathcal{F}_{t-1}\right]-\widetilde{\Omega}_t^N\right|\,.$$

The random variables $\left\{\omega_t^{N,\ell}\frac{\mathcal{L}_{t,T}\mathbf{1}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}\right\}_{\ell=1}^N$ being bounded and conditionally independent given $\mathcal{F}_{t-1}^N$, following the same steps as in the proof of Proposition 4.1, there exists a constant C (depending only on q, σ_-, σ_+, c_- and $\sup_{t\geq0}|\omega_t|_\infty$) such that $\left\|V_{t,T}^N\right\|_{2q}\leq CN^{-1/2}$. Similarly

$$\left\|N^{-1}\sum_{\ell=1}^N a_{t,T}^{N,\ell}\right\|_{2q}\leq C\frac{\sum_{s=r}^T\rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)}{N^{1/2}}\,,$$

and

$$\left\|\mathbb{E}\left[\widetilde{\Omega}_t^N\middle|\mathcal{F}_{t-1}\right]-\widetilde{\Omega}_t^N\right\|_{2q}\leq\frac{C}{N^{1/2}}\,.$$

The Cauchy-Schwarz inequality concludes the proof of (4.22). □

The proof of Theorem 4.1 is now concluded for the FFBS estimator $\phi_{0:T|T}^N[S_{T,r}]$ and we can proceed to the proof for the FFBSi estimator. We preface the proof of Theorem 4.1 for the FFBSi estimator $\widetilde{\phi}_{0:T|T}^N$ by the following Lemma. We first define the backward filtration $\left\{\mathcal{G}_{t,T}^N\right\}_{t=0}^{T+1}$ by

$$\begin{cases}\mathcal{G}_{T+1,T}^N\stackrel{\text{def}}{=}\mathcal{F}_T^N\,,\\ \mathcal{G}_{t,T}^N\stackrel{\text{def}}{=}\mathcal{F}_T^N\vee\sigma\left\{J_u^\ell,1\leq\ell\leq N,t\leq u\leq T\right\},\quad\forall\,t\in\{0,\ldots,T\}\,.\end{cases}$$

Lemma 4.2. *Assume A1–3. Let $\ell \in \{1,\dots,N\}$ and $T < +\infty$. For any bounded measurable function h on $\mathbb{X}^{r+1}$ we have,*

(i) for all u,t such that $r \le t \le u \le T$,

$$\left| \mathbb{E}\left[h\left(\xi^{N,J^{\ell}_{t-r:t}}_{t-r:t}\right) \middle| \mathcal{G}^N_{u,T} \right] - \mathbb{E}\left[h\left(\xi^{N,J^{\ell}_{t-r:t}}_{t-r:t}\right) \middle| \mathcal{G}^N_{u+1,T} \right] \right| \le \rho^{u-t} \mathrm{osc}(h) \,,$$

where ρ is defined in A1(i)*.*

(ii) for all u,t such that $t-r \le u \le t-1 \le T$,

$$\left| \mathbb{E}\left[h\left(\xi^{N,J^{\ell}_{t-r:t}}_{t-r:t}\right) \middle| \mathcal{G}^N_{u,T} \right] - \mathbb{E}\left[h\left(\xi^{N,J^{\ell}_{t-r:t}}_{t-r:t}\right) \middle| \mathcal{G}^N_{u+1,T} \right] \right| \le \mathrm{osc}(h) \,.$$

Proof. According to Section 4.2.2, for all $\ell \in \{1,\dots,N\}$, $\{J^{N,\ell}_u\}_{u=0}^T$ is an inhomogeneous Markov chain evolving backward in time with backward kernel $\{\Lambda^N_u\}_{u=0}^{T-1}$. For any $r \le t \le u \le T$, we have

$$\begin{aligned}
&\mathbb{E}\left[h\left(\xi^{N,J^{N,\ell}_{t-r:t}}_{t-r:t}\right) \middle| \mathcal{G}^N_{u,T} \right] - \mathbb{E}\left[h\left(\xi^{N,J^{N,\ell}_{t-r:t}}_{t-r:t}\right) \middle| \mathcal{G}^N_{u+1,T} \right] \\
&\qquad = \sum_{j_{t:u}} \left[\delta_{J^{N,\ell}_u}(j_u) - \left(\Lambda_u(J^{N,\ell}_{u+1}, j_u)\mathbf{1}_{u<T} + \frac{\omega^{N,j_u}_T}{\Omega_u}\mathbf{1}_{u=T} \right) \right] \\
&\qquad\qquad \times \prod_{\ell=u}^{t+1} \Lambda^N_{\ell-1}(j_\ell, j_{\ell-1}) \sum_{j_{t-r:t-1}} \prod_{\ell=t}^{t-r+1} \Lambda^N_{\ell-1}(j_\ell, j_{\ell-1}) h\left(\xi^{N,j_{t-r:t}}_{t-r:t}\right) \,.
\end{aligned}$$

The RHS of this equation is the difference between two expectations started with two different initial distributions. Under A1(i), the backward kernel satisfies the uniform Doeblin condition,

$$\forall (i,j) \in \{1,\dots,N\}^2 \quad \Lambda^N_s(i,j) \ge \frac{\sigma_-}{\sigma_+} \frac{\omega^i_s}{\Omega^N_s} \,,$$

and the proof is completed by the exponential forgetting of the backward kernel (see Cappé et al. [2005], Del Moral and Guionnet [2001]). The proof of (ii) follows exactly the same lines. □

To compute an upper-bound for the L_q-mean error of the FFBSi algorithm, we may define the difference between the FFBS and the FFBSi estimators:

$$\delta^N_T[S_{T,r}] = \widetilde{\phi}^N_{0:T|T}[S_{T,r}] - \phi^N_{0:T|T}[S_{T,r}] \,. \tag{4.25}$$

Proof of Theorem 4.1 for the FFBSi estimator. The difference between the FFBS and the FFBSi estimators, δ^N_T, defined in (4.25), can be written

$$\begin{aligned}
\delta^N_T[S_{T,r}] &= \frac{1}{N}\sum_{\ell=1}^N \sum_{t=r}^T h_t\left(\xi^{N,J^{N,\ell}_{t-r:t}}_{t-r:t}\right) - \mathbb{E}\left[h_t\left(\xi^{N,J^{N,1}_{t-r:t}}_{t-r:t}\right) \middle| \mathcal{F}^N_T \right] \\
&= \frac{1}{N}\sum_{\ell=1}^N \sum_{t=r}^T \sum_{u=t-r}^T \mathbb{E}\left[h_t\left(\xi^{N,J^{N,\ell}_{t-r:t}}_{t-r:t}\right) \middle| \mathcal{G}^N_{u,T} \right] - \mathbb{E}\left[h_t\left(\xi^{N,J^{N,\ell}_{t-r:t}}_{t-r:t}\right) \middle| \mathcal{G}^N_{u+1,T} \right] \\
&= \frac{1}{N}\sum_{\ell=1}^N \sum_{u=0}^T \zeta^{N,\ell}_u \,,
\end{aligned}$$

where

$$\zeta^{N,\ell}_u \stackrel{\mathrm{def}}{=} \sum_{t=r}^{(u+r)\wedge T} \mathbb{E}\left[h_t\left(\xi^{N,J^{N,\ell}_{t-r:t}}_{t-r:t}\right) \middle| \mathcal{G}^N_{u,T} \right] - \mathbb{E}\left[h_t\left(\xi^{N,J^{N,\ell}_{t-r:t}}_{t-r:t}\right) \middle| \mathcal{G}^N_{u+1,T} \right] \,.$$

For all $\ell \in \{1,\dots,N\}$ and all $u \in \{0,\dots,T\}$, the random variable $\zeta_u^{N,\ell}$ is $\mathcal{G}_{u,T}^N$-measurable and $\mathbb{E}\left[\zeta_u^{N,\ell} \middle| \mathcal{G}_{u+1,T}^N\right] = 0$ so that $\zeta_u^{N,\ell}$ can be seen as the increment of a backward martingale. Hence, since $q \geq 2$, using the Burkholder inequality (see [Hall and Heyde, 1980, Theorem 2.10, page 23]), there exists a constant C (depending only on q, σ_-, σ_+, c_-, $\sup_{t\geq 1}|\vartheta_t|_\infty$ and $\sup_{t\geq 0}|\omega_t|_\infty$) such that:

$$\left\|\delta_T^N[S_{T,r}]\right\|_q \leq C\left\{\sum_{u=0}^{T}\mathbb{E}\left[\left|N^{-1}\sum_{\ell=1}^{N}\zeta_u^{N,\ell}\right|^q\right]^{2/q}\right\}^{1/2}. \tag{4.26}$$

Then, since the random variables $\{\zeta_u^{N,\ell}\}_{\ell=1}^N$ are conditionally independent and centered conditionally to $\mathcal{G}_{u+1,T}^N$, using the Burkholder inequality once again implies:

$$\mathbb{E}\left[\left|\sum_{\ell=1}^{N}\zeta_u^{N,\ell}\right|^q \middle| \mathcal{G}_{u+1,T}^N\right] \leq CN^{q/2-1}\sum_{\ell=1}^{N}\mathbb{E}\left[\left|\zeta_u^{N,\ell}\right|^q \middle| \mathcal{G}_{u+1,T}^N\right]. \tag{4.27}$$

Furthermore, according to Lemma 4.2(i),

$$\begin{aligned}\left|\zeta_u^{N,\ell}\right| &\leq \sum_{t=r}^{u}\left|\mathbb{E}\left[h_t\left(\xi_{t-r:t}^{N,J_{t-r:t}^{N,\ell}}\right)\middle|\mathcal{G}_{u,T}^N\right] - \mathbb{E}\left[h_t\left(\xi_{t-r:t}^{N,J_{t-r:t}^{N,\ell}}\right)\middle|\mathcal{G}_{u+1,T}^N\right]\right| \\ &\quad + \sum_{t=u+1}^{(u+r)\wedge T}\left|\mathbb{E}\left[h_t\left(\xi_{t-r:t}^{N,J_{t-r:t}^{N,\ell}}\right)\middle|\mathcal{G}_{u,T}^N\right] - \mathbb{E}\left[h_t\left(\xi_{t-r:t}^{N,J_{t-r:t}^{N,\ell}}\right)\middle|\mathcal{G}_{u+1,T}^N\right]\right| \\ &\leq \sum_{t=r}^{u}\rho^{u-t}\mathrm{osc}(h_t) + \sum_{t=u+1}^{(u+r)\wedge T}\mathrm{osc}(h_t). \end{aligned} \tag{4.28}$$

Putting (4.26), (4.27) and (4.28) together leads to

$$\left\|\delta_T^N[S_{T,r}]\right\|_q \leq \frac{C}{\sqrt{N}}\left\{\sum_{u=0}^{T}\left(\sum_{t=r}^{(u+r)\wedge T}\rho^{(u-t)\vee 0}\mathrm{osc}(h_t)\right)^2\right\}^{1/2}.$$

Using the Holder inequality as in the proof of Proposition 4.1 yields

$$\left\|\delta_T^N[S_{T,r}]\right\|_q \leq \frac{C}{\sqrt{N}}\sqrt{1+r}\left(\sqrt{1+r}\wedge\sqrt{T-r+1}\right)\left(\sum_{s=r}^{T}\mathrm{osc}(h_s)^2\right)^{1/2},$$

and the proof of Theorem 4.1 for the FFBSi estimator is derived from the triangle inequality:

$$\left\|\phi_{0:T|T}(S_{T,r}) - \widetilde{\phi}_{0:T|T}^N(S_{T,r})\right\|_q \leq \left\|\Delta_T^N[S_{T,r}]\right\|_q + \left\|\delta_T^N[S_{T,r}]\right\|_q,$$

where $\Delta_T^N[S_{T,r}]$ is defined by (4.9) and $\delta_T^N[S_{T,r}]$ is defined by (4.25). □

4.6 Proof of Theorem 4.2

We preface the proof of the Theorem by showing that the martingale term of the error $\Delta_T^N[S_{T,r}]$ (which is defined by (4.9)) satisfies an exponential deviation inequality in the following Proposition.

Proposition 4.3. *Assume A1–3. There exists a constant C (depending only on σ_-, σ_+, c_-, $\sup_{t\geq 1}|\vartheta_t|_\infty$ and $\sup_{t\geq 0}|\omega_t|_\infty$) such that for any $T<\infty$, any $N\geq 1$, any $\varepsilon>0$, any integer r and any bounded and measurable functions $\{h_s\}_{s=r}^T$ on $\mathbb{X}^{r+1}$,*

$$\mathbb{P}\left\{\left|\sum_{t=0}^{T}D_{t,T}^N(S_{T,r})\right| > \varepsilon\right\} \leq 2\exp\left(-\frac{CN\varepsilon^2}{\Theta_{r,T}\sum_{s=r}^{T}\mathrm{osc}(h_s)^2}\right), \tag{4.29}$$

where $D_{t,T}^N$ is defined in (4.14) *and $\Theta_{r,T}$ is defined by* (4.17).

Proof. According to the definition of $D^N_{t,T}(S_{T,r})$ given in (4.14), we can write

$$\sum_{t=0}^{T} D^N_{t,T}(S_{T,r}) = \sum_{k=1}^{N(T+1)} \upsilon^N_k \ ,$$

where for all $t \in \{0,\dots,T\}$ and $\ell \in \{1,\dots,N\}$, $\upsilon^N_{Nt+\ell}$ is defined by

$$\upsilon^N_{Nt+\ell} = \frac{\phi^N_{t-1}[\vartheta_t]}{\phi^N_{t-1}\left[\frac{\mathcal{L}_{t-1,T}\mathbf{1}}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}\right]} N^{-1}\omega^{N,\ell}_t \frac{G^N_{t,T}S_{T,r}(\xi^{N,\ell}_t)}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \ ,$$

and is bounded by (see (4.18))

$$\left|\upsilon^N_{Nt+\ell}\right| \le CN^{-1}\sum_{s=r}^{T} \rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s) \ .$$

Furthermore, we define the filtration $\left\{\mathcal{H}^N_k\right\}_{k=1}^{N(T+1)}$, for all $t \in \{0,\dots,T\}$ and $\ell \in \{1,\dots,N\}$, by:

$$\mathcal{H}^N_{Nt+\ell} \stackrel{\text{def}}{=} \mathcal{F}^N_{t-1} \vee \sigma\left\{\left(\omega^{N,i}_t, \xi^{N,i}_t\right), 1 \le i \le \ell\right\} \ ,$$

with the convention $\mathcal{F}^N_{-1} = \sigma(Y_{0:T})$. Then, according to Lemma 4.1, $\{\upsilon_k\}_{k=1}^{N(T+1)}$ is martingale increment for the filtration $\{\mathcal{H}^N_k\}_{k=1}^{N(T+1)}$ and the Azuma-Hoeffding inequality completes the proof. □

Proposition 4.4. *Assume A1–3. There exists a constant C (depending only on σ_-, σ_+, c_-, $\sup_{t\ge1}|\vartheta_t|_\infty$ and $\sup_{t\ge0}|\omega_t|_\infty$) such that for any $T < \infty$, any $N \ge 1$, any $\varepsilon > 0$, any integer r and any bounded and measurable functions $\{h_s\}_{s=r}^T$ on $\mathbb{X}^{r+1}$,*

$$\mathbb{P}\left\{\left|\sum_{t=0}^{T} C^N_{t,T}(S_{T,r})\right| > \varepsilon\right\} \le 8\exp\left(-\frac{CN\varepsilon}{(1+r)\sum_{s=r}^T \mathrm{osc}(h_s)}\right) \ . \tag{4.30}$$

where $C^N_{t,T}(F)$ is defined in (4.15)*.*

Proof. In order to apply Lemma 4.4 in the appendix, we first need to find an exponential deviation inequality for $C^N_{t,T}(S_{T,r})$ which is done by using the decomposition $C^N_{t,T}(S_{T,r}) = U^N_{t,T}V^N_{t,T}W^N_{t,T}$ given in (4.23). First, the ratio $U^N_{t,T}$ is dealt with through Lemma 4.3 in the appendix by defining

$$\begin{cases} a_N \stackrel{\text{def}}{=} N^{-1}\sum_{\ell=1}^N \omega^{N,\ell}_t G^N_{t,T}S_{T,r}(\xi^{N,\ell}_t)/|\mathcal{L}_{t,T}\mathbf{1}|_\infty \ , \\ b_N \stackrel{\text{def}}{=} N^{-1}\sum_{\ell=1}^N \omega^{N,\ell}_t \ , \\ b \ \stackrel{\text{def}}{=} \mathbb{E}[\omega^1_t|\mathcal{F}^N_{t-1}] = \phi^N_{t-1}\left[Mg_t\right]/\phi^N_{t-1}[\vartheta_t] \ , \\ \beta \ \stackrel{\text{def}}{=} c_-/|\vartheta_t|_\infty \ . \end{cases}$$

Assumption A1(ii) and A3 shows that $b \ge \beta$ and (4.18) shows that $|a_N/b_N| \le C(1+r)\max_{r\le t\le T}\{\mathrm{osc}(h_t)\}$. Therefore, Condition (I) of Lemma 4.3 is satisfied. The bounds $0 < \omega^l_t \le |\omega_t|_\infty$ and the Hoeffding inequality lead to

$$\mathbb{P}[|b_N - b| \ge \varepsilon] = \mathbb{E}\left[\mathbb{P}\left[\left|N^{-1}\sum_{\ell=1}^N\left(\omega^{N,\ell}_t - \mathbb{E}[\omega^{N,1}_t|\mathcal{F}^N_{t-1}]\right)\right| \ge \varepsilon \middle| \mathcal{F}^N_{t-1}\right]\right] \le 2\exp\left(-\frac{2N\varepsilon^2}{|\omega_t|^2_\infty}\right) \ ,$$

establishing Condition (ii) in Lemma 4.3. Finally, Lemma 4.1(i) and the Hoeffding inequality imply that

$$\mathbb{P}\left[|a_N| \ge \varepsilon\right] = \mathbb{E}\left[\mathbb{P}\left[\left|N^{-1}\sum_{\ell=1}^N \omega^{N,\ell}_t G^N_{t,T}S_{T,r}(\xi^{N,\ell}_t)/|\mathcal{L}_{t,T}\mathbf{1}|_\infty\right| \ge \varepsilon \middle| \mathcal{F}^N_{t-1}\right]\right]$$
$$\le 2\exp\left(-\frac{N\varepsilon^2}{2|\omega_t|^2_\infty\left(\sum_{s=r}^T \rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)\right)^2}\right) \ .$$

Lemma 4.3 therefore yields

$$\mathbb{P}\left\{\left|U_{t,T}^N\right| \geq \varepsilon\right\} \leq 2\exp\left(-\frac{CN\varepsilon^2}{\left(\sum_{s=r}^T \rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)\right)^2}\right) .$$

Then $V_{t,T}^N$ is dealt with by using again the Hoeffding inequality and the bounds $0 < b_{t,T}^{N,\ell} \leq |\omega_t|_\infty$, where $b_{t,T}^{N,\ell} \stackrel{\text{def}}{=} \omega_t^{N,\ell} \frac{\mathcal{L}_{t,T}\mathbf{1}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}$:

$$\mathbb{P}\left[\left|N^{-1}\sum_{\ell=1}^N b_{t,T}^{N,\ell} - \mathbb{E}\left[b_{t,T}^{N,1}\middle|\mathcal{F}_{t-1}\right]\right| \geq \varepsilon\right]$$
$$= \mathbb{E}\left[\mathbb{P}\left[\left|N^{-1}\sum_{\ell=1}^N \left(b_{t,T}^{N,\ell} - \mathbb{E}\left[b_{t,T}^{N,\ell}\middle|\mathcal{F}_{t-1}^N\right]\right)\right| \geq \varepsilon\middle|\mathcal{F}_{t-1}^N\right]\right] \leq 2\exp\left(-CN\varepsilon^2\right) .$$

Finally, $W_{t,T}^N$ has been shown in (4.24) to be bounded by a constant depending only on σ_-, σ_+, c_-, $\sup_{t\geq 1}|\vartheta_t|_\infty$ and $\sup_{t\geq 0}|\omega_t|_\infty$: $\left|W_{t,T}^N\right| \leq C$ so that

$$\mathbb{P}\left\{\left|C_{t,T}^N(S_{T,r})\right| > \varepsilon\right\} \leq \mathbb{P}\left\{\left|U_{t,T}^N V_{t,T}^N\right| > \varepsilon/C\right\} \leq \mathbb{P}\left\{\left|U_{t,T}^N\right| > \varepsilon_u\right\} + \mathbb{P}\left\{\left|V_{t,T}^N\right| > \varepsilon_v\right\} ,$$

where

$$\varepsilon_u \stackrel{\text{def}}{=} \sqrt{\varepsilon\sum_{s=r}^T \rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)/C} \quad \text{and} \quad \varepsilon_u \stackrel{\text{def}}{=} \sqrt{\frac{\varepsilon}{C\sum_{s=r}^T \rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)}} .$$

Therefore,

$$\mathbb{P}\left\{\left|C_{t,T}^N(S_{T,r})\right| > \varepsilon\right\} \leq 4\exp\left(-\frac{CN\varepsilon}{\sum_{s=r}^T \rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)}\right) .$$

The proof of (4.30) is finally completed by applying Lemma 4.4 with

$$X_t = C_{t,T}^N(S_{T,r}) , \quad A = 4 , \quad B_t = \frac{CN}{\sum_{s=r}^T \rho^{\max(t-s,s-r-t,0)}\mathrm{osc}(h_s)} , \quad \gamma = 1/2 .$$

□

Proof of Theorem 4.2 for the FFBS estimator. The result is obtained by writing

$$\mathbb{P}\left\{\left|\Delta_T^N[S_{T,r}]\right| > \varepsilon\right\} \leq \mathbb{P}\left\{\left|\sum_{t=0}^T C_{t,T}^N(S_{T,r})\right| > \varepsilon/2\right\} + \mathbb{P}\left\{\left|\sum_{t=0}^T D_{t,T}^N(S_{T,r})\right| > \varepsilon/2\right\} ,$$

and using (4.29) and (4.30). □

Proof of Theorem 4.2 for the FFBSi estimator. We recall the decomposition used in the proof of Theorem 4.1 for the FFBSi estimator:

$$\delta_T^N[S_{T,r}] = \frac{1}{N}\sum_{\ell=1}^N\sum_{u=0}^T \zeta_u^{N,\ell} ,$$

where $\delta_T^N[S_{T,r}]$ is defined by (4.25). Since $\left\{\zeta_u^{N,\ell}\right\}_{\ell=1}^N$ are $\mathcal{G}_{u,T}^N$ measurable and centered conditionally to $\mathcal{G}_{u+1,T}^N$ using the same steps as in the proof of Proposition 4.3, we get

$$\mathbb{P}\left\{\left|\delta_T^N[S_{T,r}]\right| > \varepsilon\right\} \leq 2\exp\left(-\frac{CN\varepsilon^2}{\Theta_{r,T}\sum_{s=r}^T \mathrm{osc}(h_s)^2}\right) ,$$

where $\Theta_{r,T}$ is defined by (4.17). The proof is finally completed by writing

$$\phi_{0:T|T}[S_{T,r}] - \widetilde{\phi}_{0:T|T}^N[S_{T,r}] = \Delta_T^N[S_{T,r}] + \delta_T^N[S_{T,r}] ,$$

and by using Theorem 4.2 for the FFBS estimator. □

4.A Technical results

Lemma 4.3. *Assume that a_N, b_N, and b are random variables defined on the same probability space such that there exist positive constants β, B, C, and M satisfying*

(i) $|a_N/b_N| \leq M$, $\mathbb{P}$*-a.s. and* $b \geq \beta$, $\mathbb{P}$*-a.s.,*

(ii) For all $\varepsilon > 0$ *and all* $N \geq 1$, $\mathbb{P}[|b_N - b| > \varepsilon] \leq B\mathrm{e}^{-CN\varepsilon^2}$,

(iii) For all $\varepsilon > 0$ *and all* $N \geq 1$, $\mathbb{P}[|a_N| > \varepsilon] \leq B\mathrm{e}^{-CN(\varepsilon/M)^2}$.

Then,

$$\mathbb{P}\left\{\left|\frac{a_N}{b_N}\right| > \varepsilon\right\} \leq B\exp\left(-CN\left(\frac{\varepsilon\beta}{2M}\right)^2\right).$$

Proof. See [Douc et al., 2010, Lemma 4]. □

Lemma 4.4. *For $T \geq 0$, let $\{X_t\}_{t=0}^T$ be $(T+1)$ random variables. Assume that there exists a constants $A \geq 1$ and for all $0 \leq t \leq T$, there exists a constant $B_t > 0$ such that and all $\varepsilon > 0$*

$$\mathbb{P}\{|X_t| > \varepsilon\} \leq Ae^{-B_t\varepsilon}.$$

Then, for all $0 < \gamma < 1$ and all $\varepsilon > 0$, we have

$$\mathbb{P}\left\{\left|\sum_{t=0}^T X_t\right| > \varepsilon\right\} \leq \frac{A}{1-\gamma}e^{-\gamma B\varepsilon/(T+1)},$$

where

$$B \stackrel{\text{def}}{=} \left(\frac{1}{T+1}\sum_{t=0}^T B_t^{-1}\right)^{-1}.$$

Proof. By the Bienayme-Tchebychev inequality, we have

$$\mathbb{P}\left\{\left|\sum_{t=0}^T X_t\right| > \varepsilon\right\} = \mathbb{P}\left\{\exp\left[\frac{\gamma B}{T+1}\left|\sum_{t=0}^T X_t\right|\right] > e^{\gamma B\varepsilon/(T+1)}\right\} \leq e^{-\gamma B\varepsilon/(T+1)}\mathbb{E}\left[\exp\left[\frac{\gamma B}{T+1}\left|\sum_{t=0}^T X_t\right|\right]\right]. \tag{4.31}$$

It remains to bound the expectation in the RHS of (4.31) by $A(1-\gamma)^{-1}$. First, by the Minkowski inequality,

$$\mathbb{E}\left[\exp\left[\frac{\gamma B}{T+1}\left|\sum_{t=0}^T X_t\right|\right]\right] = \sum_{q=0}^\infty \frac{\gamma^q B^q}{q!(T+1)^q}\mathbb{E}\left[\left|\sum_{t=0}^T X_t\right|^q\right] \leq 1 + \sum_{q=1}^\infty \frac{\gamma^q B^q}{q!(T+1)^q}\left(\sum_{t=0}^T \|X_t\|_q\right)^q.$$

Moreover, for $q \geq 1$, $\mathbb{E}[|X_t|^q]$ can be bounded by

$$\mathbb{E}[|X_t|^q] = \int_0^\infty \mathbb{P}\{|X_t| > \varepsilon^{1/q}\}\mathrm{d}\varepsilon \leq A\int_0^\infty e^{-B_t\varepsilon^{1/q}}\mathrm{d}\varepsilon = \frac{Aq!}{B_t^q},$$

Finally,

$$\mathbb{E}\left[\exp\left[\frac{\gamma B}{T+1}\left|\sum_{t=0}^T X_t\right|\right]\right] \leq A\sum_{q=0}^\infty \gamma^q = \frac{A}{(1-\gamma)}.$$

□

Chapitre 5

Calibrer le modèle de volatilité stochastique d'Ornstein-Uhlenbeck multi-échelles

Sommaire

Abstract: This paper exhibits a tractable and efficient way of calibrating a multiscale exponential Ornstein-Uhlenbeck stochastic volatility model including a correlation between the asset and its volatility. As opposed to many contributions where this correlation is assumed to be null, this framework allows to describe the leverage effect widely observed in equity markets. The resulting model is non exponential and driven by a degenerated noise, thus requiring high carefulness about the estimation algorithm design. The way we overcome this difficulty provides guidelines concerning the development of estimation algorithm in non standard framework. We propose to use a block-type expectation maximization algorithm along with particle smoothing. This method results in an accurate calibration process able to identify up to three time scale factors. Furthermore, we introduce an intuitive heuristic which can be used to choose the number of factors.

Keywords: Multiscale stochastic volatility model, Inference, Particle smoothing, Maximum split data estimate, Expectation-Maximization algorithm

5.1 Introduction

The well-known Black-Scholes model is the first step of the financial pricing understanding. However it fails to include some statistical properties of observed financial data. This issue is partially due to the constant volatility used in the Black-Scholes model. As a consequence, more sophisticated models have been introduced to make the volatility non constant: ARCH/GARCH models for discrete time introduced in Engle [1982], Bollerslev [1986] and stochastic volatility models for continuous time [Hull and White, 1987]. In the latest model class, the exponential Ornstein-Uhlenbeck multiscale stochastic volatility model (ExpOU) is of real interest to take some financial data properties into account (see Masoliver and Perello 2006, Buchbinder and Chistilin 2007) and has been successfully used in Eisler et al. 2007 to infer the hidden volatility process in the particular case of one time scale. This model is presented below as a stochastic differential equation (SDE) system:

$$\begin{cases} \mathrm{d}S_t = S_t[\kappa_t \mathrm{d}t + \sigma_t^S \mathrm{d}B_t^S] \,, \\ \sigma_t^S = \exp\left(\frac{1}{2}\langle 1_p, \xi_t\rangle\right) \,, \\ \mathrm{d}\xi_t = \mathrm{diag}\,(a)\,(\mu - \xi_t)\mathrm{d}t + b\mathrm{d}B_t^\xi \,, \end{cases} \tag{5.1}$$

where S is the price of a financial asset, κ is a drift factor which is not dealt with in this paper, σ^S is the volatility of S driven by a p-dimensional vector ξ of Ornstein-Uhlenbeck processes of parameters $a = (a_1, \cdots, a_p)^{\mathrm{T}}, \mu = (\mu_1, \cdots, \mu_p)^{\mathrm{T}}$ and $b = (b_1, \cdots, b_p)^{\mathrm{T}}$, and $\langle \cdot, \cdot \rangle$ denotes the standard scalar product of two p-dimensional vectors. When coming to the application to real data, κ will be discarded using the detrended series defined in Kim et al. [1998]. Without loss of generality, we will restrict our attention to the case where the components of a are distinct. Moreover, we assume that the volatility process has reached its stationary regime. The process (B^S, B^ξ) is a two-dimensional brownian motion with correlation ρ i.e. $\mathrm{d} < B^S, B^\xi >= \rho \mathrm{d}t$ such that marginally, B^S and B^ξ are standard brownian motions. 1_p denotes a p-dimensional vector of ones. Typically, each component ξ^i of ξ may be associated with the timescale $1/a_i$. The same brownian motion B^ξ is driving all the components of ξ. We could have chosen to include a more general correlation matrix but in this case, the model may become not identifiable (see Fouque et al. 2008). This choice for the correlation matrix models the fact that market participants operate at different time scales not independently (intraday market makers, hedge funds or pension funds operate with quite different schedules).

A challenging issue with this model remains the estimation of its parameters (a, μ, b, ρ) given the observation of the process S over a fixed period of time. This is a tricky task since the volatility noise is degenerated and correlated to the one of the asset and thus classical algorithms cannot be used within this framework. The calibration of a similar model has been performed in Fouque et al. [2008] in the particular case of $p \in \{1, 2\}$ using a Markov chain Monte Carlo (MCMC) method. The main differences between Fouque et al. [2008] and the model introduced above is that in Fouque et al. [2008], the brownian motions driving S and ξ are independent and the components of ξ are driven by independent brownian motions, whereas in this paper, we keep the same brownian motion for all the components of ξ (making it degenerated) correlated to the one driving S. As opposed to many contributions where this correlation between the asset and its volatility is assumed to be null, this model allows to describe the leverage effect widely observed in equity markets. In fact, early studies (e.g. Black 1976, Christie 1982) stated that a decrease of the stock price implies an increase of the associated risk, ie $\rho < 0$. In this paper, we focus on a block-type variant of the expectation-maximization (EM) algorithm (introduced in Dempster et al. 1977) to perform the calibration. The expectation step is approximated using sequential Monte-Carlo methods giving an estimate of posterior distributions in non linear and non gaussian state-space models [Gordon et al., 1993, Del Moral, 2004]. An overview of these methods can be found in Doucet et al. [2001].

This technique implies working on a discretization of (5.1) which is done in section 5.2. Then, we prove a quantitative identifiability result for the discrete model in section 5.3 of which we take profit in a calibration method through a block-type expectation-maximization algorithm (section 5.4). This is completed by a detailed study of the expectation step in section 5.5. We finally apply the algorithm to simulated and real data (CAC 40 and Dow Jones indices over 10 years) in section 5.6 and give a criterium to select the number of factors. The proposed algorithm works well with up to three time scale factors ($p \in \{1, 2, 3\}$).

5.2 The discretized model

In order to discretize (5.1), we apply an Euler scheme associated to a fixed time step δ and discretization grid $\{t_k\}_{k\geq 0}$ with $t_k = k\delta$:

$$\begin{cases} \overline{S}_{t_{k+1}} - \overline{S}_{t_k} = \overline{S}_{t_k}[\overline{\kappa}_{t_k}\delta + \overline{\sigma}^S_{t_k}\sqrt{\delta}V_{k+1}] \,, \\ \overline{\sigma}^S_{t_k} = \exp\left(\frac{1}{2}\left\langle 1_p, \overline{\xi}_{t_k}\right\rangle\right) \,, \\ \overline{\xi}_{t_{k+1}} - \overline{\xi}_{t_k} = \mathrm{diag}\,(a)\,(\mu - \overline{\xi}_{t_k})\delta + b\sqrt{\delta}W_{k+1} \,, \end{cases} \tag{5.2}$$

where $\{(W_k, V_k)\}_{k\geq 1}$ is a sequence of iid two dimensional gaussian vectors such that

$$\begin{bmatrix} W_1 \\ V_1 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right) .$$

By setting $Y_{k+1} \stackrel{\text{def}}{=} \dfrac{1}{\sqrt{\delta}}\left[\dfrac{\overline{S}_{t_{k+1}} - \overline{S}_{t_k}}{S_{t_k}} - \overline{\kappa}_{t_k}\delta\right]$ and $X_k \stackrel{\text{def}}{=} \overline{\xi}_{t_k} - \mu$, (5.2) can be rewritten according to the following discretized model.

Definition 5.1. THE DISCRETIZED MODEL.

$$\begin{cases} X_{k+1} = \mathrm{diag}\,(\alpha)X_k + \sigma W_{k+1} \,, \\ Y_{k+1} = \beta e^{\langle 1_p, X_k\rangle/2}V_{k+1} \,, \end{cases} \tag{5.3}$$

with $\alpha = 1_p - \delta a, \sigma = \sqrt{\delta}b$, and $\beta = \exp(\frac{1}{2}\langle 1_p, \mu\rangle)$. We assume that for $i \in \{1,\ldots,p\}$, $0 \leq \alpha_i < 1$, $\sigma_i > 0$. In addition, since the model is not affected by any permutation of the components of α, we restrict our attention to the case where $\alpha_i < \alpha_{i+1}$ for $i \in \{1,\ldots,p-1\}$.

The previously described model is parametrized by the $(2p+2)$-dimensional parameter vector $\theta \in \Theta_p$ where

$$\Theta_p \stackrel{\text{def}}{=} \left\{(\alpha,\sigma,\beta,\rho) \in [0,1)^p \times (\mathbb{R}_+^\star)^p \times \mathbb{R}_+^\star \times (-1,1); \alpha_1 < \cdots < \alpha_p\right\} ,$$

and inference on this parameter will only be based on the observations $(Y_k)_{k\geq 1}$. At this stage, it is worthwhile to note that it is not possible to estimate separately the components of μ since this vector appears in the discretized model only through the quantity $\langle 1_p, \mu\rangle$ which will be estimated by the parameter β. In the sequel, we will assume that the distribution of X_0 is the stationary distribution of the Markov chain X:

$$X_0 \sim \mathcal{N}\left(0, \Upsilon_{\alpha,\sigma}\right) , \tag{5.4}$$

where the $p \times p$ covariance matrix satisfies $\Upsilon_{\alpha,\sigma} = \mathrm{diag}\,(\sigma)\mathbf{A}_\alpha \mathrm{diag}\,(\sigma)$, with $\mathbf{A}_\alpha = \left((1-\alpha_i\alpha_j)^{-1}\right)_{1\leq i,j\leq p}$. Using the Cholesky decomposition of $\mathbf{A}_\alpha$, $\Upsilon_{\alpha,\sigma}$ can be written $\Upsilon_{\alpha,\sigma} = \mathrm{diag}\,(\sigma)\mathbf{L}_\alpha\mathbf{L}_\alpha^{\mathrm{T}}\mathrm{diag}\,(\sigma)$ where $\mathbf{L}_\alpha$ is the lower triangular matrix defined by:

$$\forall i,j \in \{1,\ldots,p\}, \quad (\mathbf{L}_\alpha)_{i,j} = \begin{cases} 0, & j > i \,, \\ \left(\displaystyle\prod_{k=1}^{j-1}\frac{\alpha_i - \alpha_k}{1-\alpha_k\alpha_i}\right)\frac{\sqrt{1-\alpha_j^2}}{1-\alpha_j\alpha_i}, & j \leq i \,. \end{cases}$$

This decomposition allows to easily simulate X_0 and shows that its distribution is not degenerated.

5.3 Identifiability

The following theorem shows that two different parameters induce different distributions for the observation process Y. More precisely, the distribution of the $2p$-random vector $(Y_1,\ldots,Y_{2p})$ is proved to be sufficient for characterizing the parameter θ. The latest will be exploited in section 5.4 through a block-type estimation of the parameter.

Theorem 5.1. *Let $\theta^{(i)} = (\alpha^{(i)}, \sigma^{(i)}, \beta^{(i)}, \rho^{(i)})$, $i \in \{1,2\}$ be two sets of parameters in Θ_p and define two pairs of processes $(X^{(i)}, Y^{(i)})$, $i \in \{1,2\}$ such that for all $k \geq 0$,*

$$\begin{cases} X_{k+1}^{(i)} = \operatorname{diag}\left(\alpha^{(i)}\right) X_k^{(i)} + \sigma^{(i)} W_{k+1}^{(i)} , \\ Y_{k+1}^{(i)} = \beta^{(i)} \mathrm{e}^{\left\langle 1_p, X_k^{(i)} \right\rangle / 2} V_{k+1}^{(i)} , \end{cases} \tag{5.5}$$

where $\{(W_k^{(1)}, V_k^{(1)})\}_{k\geq 1}$ and $\{(W_k^{(2)}, V_k^{(2)})\}_{k\geq 1}$ are two independent sequences of iid two dimensional gaussian vectors such that

$$\begin{bmatrix} W_1^{(i)} \\ V_1^{(i)} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) ,$$

and $X_0^{(i)} \sim \mathcal{N}\left(0, \boldsymbol{\Upsilon}_{\alpha^{(i)}, \sigma^{(i)}}\right)$. Then the three following assertions are equivalent:

$$i)\ Y_{1:2p}^{(1)} \overset{\mathcal{L}}{=} Y_{1:2p}^{(2)} , \qquad ii)\ \theta^{(1)} = \theta^{(2)} , \qquad iii)\ \forall k \geq 1, \quad Y_{1:k+1}^{(1)} \overset{\mathcal{L}}{=} Y_{1:k+1}^{(2)} .$$

For ease of reading, proof of this Theorem is postponed to Appendix 5.A.1. It relies on a rewriting of the model. Indeed, it turns out that the distribution of Y conditionally to X only depends on the random quantities X_0 and $\tilde{X}_k \overset{\text{def}}{=} \langle 1_p, X_k \rangle$ rather than on X itself. This is easily seen from (5.3) when $\rho = 0$ but in the correlated case, this has to be checked carefully.

Remark 5.1. When X_0 is distributed according to (5.4), the process $\tilde{X}$ is gaussian, centered and stationary. Consequently, this process is characterized by its auto-covariance function $\gamma^{\tilde{X}}$ given for all $k \geq 0$ by (using (5.19))

$$\gamma_k^{\tilde{X}} \overset{\text{def}}{=} \mathbb{C}\mathrm{ov}\left(\tilde{X}_0, \tilde{X}_k\right) = 1_p^{\mathrm{T}} \boldsymbol{\Upsilon}_{\alpha,\sigma} \alpha^k . \tag{5.6}$$

Thus the law of the process Y can be characterized by the law of $\tilde{X}$, β and ρ. This remark will be used in section 5.6 as a criterium to select the number of time scale factors when calibrating real data.

5.4 Inference

In order to infer the parameter θ using a fixed set of observations $Y_{1:T}$, a natural idea in hidden models would be to use an expectation-maximization (EM) algorithm. This algorithm relies on a recursive sequence $(\hat{\theta}_n)$ in the parameter space Θ_p satisfying

$$\hat{\theta}_{n+1} = \operatorname{argmax}_{\theta \in \Theta_p} \mathbb{E}_{\hat{\theta}_n} \left[\log p_\theta(Y_{1:T}, U) | Y_{1:T} \right] , \tag{5.7}$$

where $\log p_\theta(Y_{1:T}, U)$ is the log-likelihood of the joint distribution of $(Y_{1:T}, U)$ for any hidden random variable U. The choice of U is usually driven by the fact that i) the log-likelihood is easily computable ii) the expectation and maximization steps in the EM algorithm can be separated. $X_{0:T}$ is not a good candidate for U: it is a degenerated Markov chain, for which $p_\theta(Y_{1:T}, X_{0:T})$ cannot be defined. However, we show that $U = (X_0, \tilde{X}_{1:T})$ meets the previous two requirements at the cost of having to estimate the smoothing distribution of all the couples $(\tilde{X}_k, \tilde{X}_\ell), 1 \leq k < \ell \leq T$. As standard smoothing algorithms for approximating such joint distributions are degenerated for huge values of $\ell - k$, we develop a block-type EM by splitting the observations into B smaller blocks to decrease the degeneracy. As will be detailed in subsection 5.4.2, this approach is closely linked to the split data likelihood technique introduced in Rydén [1994].

5.4.1 The standard EM algorithm

As mentioned above, we first need to express the log-likelihood of the joint distribution of $(Y_{1:T}, X_0, \tilde{X}_{1:T})$ where X_0 is distributed according to (5.4). This is done in the following proposition.

Proposition 5.1. *The log-likelihood of the joint distribution of the random variable $(Y_{1:T}, X_0, \tilde{X}_{1:T})$ is given by:*

$$\begin{aligned}\log p_\theta(Y_{1:T}, X_0, \tilde{X}_{1:T}) = K - T\log\left(\beta\langle\sigma, 1_p\rangle\sqrt{1-\rho^2}\right) - \frac{1}{2}\sum_{k=1}^{T-1}\tilde{X}_k \\ - \frac{1}{2}\log|\det\Upsilon_{\alpha,\sigma}| - \frac{1}{2}\sum_{i,j=1}^{p}\left(\Upsilon_{\alpha,\sigma}^{-1}\right)_{i,j}X_0^iX_0^j - \frac{1}{2\beta^2(1-\rho^2)}\sum_{k=1}^{T}Y_k^2\mathrm{e}^{-\tilde{X}_{k-1}} \\ + \frac{\rho}{\beta(1-\rho^2)}\sum_{k=1}^{T}\sum_{\ell=1}^{k}Y_k\omega_{k-\ell}\left[\mathrm{e}^{-\frac{1}{2}\tilde{X}_{k-1}}\tilde{X}_\ell - \sum_{i=1}^{p}\alpha_i^l\mathrm{e}^{-\frac{1}{2}\tilde{X}_{k-1}}X_0^i\right] \\ - \frac{1}{2(1-\rho^2)}\sum_{k=1}^{T}\sum_{\ell,m=1}^{k}\omega_{k-\ell}\omega_{k-m}\left[\tilde{X}_\ell\tilde{X}_m + \sum_{i,j=1}^{p}\alpha_i^m\alpha_j^\ell X_0^iX_0^j - 2\sum_{i=1}^{p}\alpha_i^m\tilde{X}_\ell X_0^i\right], \quad (5.8)\end{aligned}$$

where $\tilde{X}_0 = \langle 1_p, X_0\rangle$, K is a deterministic constant which does not depend on θ, and

$$\begin{aligned}\omega_0 &= 1/\langle\sigma, 1_p\rangle\ , \\ \omega_k &= -\sum_{\ell=0}^{k-1}\frac{\langle\sigma, \alpha^{k-\ell}\rangle}{\langle\sigma, 1_p\rangle}\omega_\ell\ . \end{aligned} \quad (5.9)$$

Proof of this proposition is postponed to Appendix 5.A.2.

By plugging (5.8) into (5.7) with $U = (X_0, \tilde{X}_{1:T})$, the recursive sequence $(\hat{\theta}_n)$ satisfies

$$\hat{\theta}_{n+1} = \mathrm{argmax}_{\theta\in\Theta_p}\ \ell_{n,\theta}^{1,T}(Y_{1:T})\ , \quad (5.10)$$

where for all $1 \le r < s \le T$:

$$\begin{aligned}\ell_{n,\theta}^{r,s}(Y_{r:s}) \stackrel{\text{def}}{=} -(s-r+1)\log\left(\beta\langle\sigma, 1_p\rangle\sqrt{1-\rho^2}\right) - \frac{1}{2}\log|\det\Upsilon_{\alpha,\sigma}| \\ - \frac{1}{2}\sum_{i,j=1}^{p}\left(\Upsilon_{\alpha,\sigma}^{-1}\right)_{i,j}\mathbb{E}_{\hat{\theta}_n}\left[X_{r-1}^iX_{r-1}^j\middle|Y_{r:s}\right] - \frac{1}{2\beta^2(1-\rho^2)}\sum_{k=r}^{s}Y_k^2\mathbb{E}_{\hat{\theta}_n}\left[\mathrm{e}^{-\tilde{X}_{k-1}}\middle|Y_{r:s}\right] \\ + \frac{\rho}{\beta(1-\rho^2)}\sum_{k=r}^{s}\sum_{\ell=r}^{k}Y_k\omega_{k-\ell}\left[\mathbb{E}_{\hat{\theta}_n}\left[\mathrm{e}^{-\frac{1}{2}\tilde{X}_{k-1}}\tilde{X}_\ell\middle|Y_{r:s}\right] - \sum_{i=1}^{p}\alpha_i^l\mathbb{E}_{\hat{\theta}_n}\left[\mathrm{e}^{-\frac{1}{2}\tilde{X}_{k-1}}X_{r-1}^i\middle|Y_{r:s}\right]\right] \\ - \frac{1}{2(1-\rho^2)}\sum_{k=r}^{s}\sum_{\ell,m=r}^{k}\omega_{k-\ell}\omega_{k-m}\left[\mathbb{E}_{\hat{\theta}_n}\left[\tilde{X}_\ell\tilde{X}_m\middle|Y_{r:s}\right] + \sum_{i,j=1}^{p}\alpha_i^m\alpha_j^\ell\mathbb{E}_{\hat{\theta}_n}\left[X_{r-1}^iX_{r-1}^j\middle|Y_{r:s}\right]\right. \\ \left. - 2\sum_{i=1}^{p}\alpha_i^m\mathbb{E}_{\hat{\theta}_n}\left[\tilde{X}_\ell X_{r-1}^i\middle|Y_{r:s}\right]\right]\ . \quad (5.11)\end{aligned}$$

Consequently, the expectation step of the EM algorithm (developed in details in section 5.5) is separated from the maximization step and consists of estimating the following expectations

$$\begin{aligned}&\mathbb{E}_{\hat{\theta}_n}\left[X_0^iX_0^j\middle|Y_{1:T}\right], && \forall(i,j)\in\{1,\dots,p\}^2\ , \\ &\mathbb{E}_{\hat{\theta}_n}\left[\mathrm{e}^{-\tilde{X}_{k-1}}\middle|Y_{1:T}\right], && \forall k\ge 1\ , \\ &\mathbb{E}_{\hat{\theta}_n}\left[\mathrm{e}^{-\frac{1}{2}\tilde{X}_{k-1}}\tilde{X}_\ell\middle|Y_{1:T}\right], && \forall k\ge\ell\ge 1\ , \\ &\mathbb{E}_{\hat{\theta}_n}\left[\mathrm{e}^{-\frac{1}{2}\tilde{X}_{k-1}}X_0^i\middle|Y_{1:T}\right], && \forall k\ge 1, i\in\{1,\dots,p\}\ , \\ &\mathbb{E}_{\hat{\theta}_n}\left[\tilde{X}_\ell\tilde{X}_m\middle|Y_{1:T}\right], && \forall l,m\ge 1\ , \\ &\mathbb{E}_{\hat{\theta}_n}\left[\tilde{X}_\ell X_0^i\middle|Y_{1:T}\right], && \forall\ell\ge 1, i\in\{1,\dots,p\}\ . \end{aligned} \quad (5.12)$$

The maximization step is then performed with some deterministic optimization process (e.g. conjugate gradient method) over the whole parameter space Θ.

The estimation of these expectations thus requires to get an approximation of the smoothing distribution of all the couples $(\tilde{X}_k, \tilde{X}_\ell), 1 \leq k < \ell \leq T$ which is a quite difficult task when T is large. As noted above, we overcome this difficulty by proposing a block-type EM algorithm that will be described in the following subsection.

5.4.2 The block EM algorithm

The Maximum Split Data Likelihood Estimate (MSDLE) introduced by Rydén [1994] consists in splitting the observations into blocks of fixed size η viewing these groups as independent and then maximizing the resulting likelihood. In other words, the estimator is obtained by maximizing the contrast function $\sum_{u=0}^{B-1} \log p_\theta(Y_{u\eta+1:(u+1)\eta})$ where $B = T/\eta$ instead of the log-likelihood $\log p_\theta(Y_{1:T})$. This allows to derive asymptotic properties of the MSDLE from the M-estimator theory. The identifiability result obtained in Theorem 5.1 ensures that for $\eta \geq 2p$ the condition C4 in Rydén [1994] is satisfied and thus, using a generalized version of Rydén [1994, Theorem 1 and 2] to continuous state spaces (see Lai and Tung 2003), the MSDLE is consistent and asymptotically normal.

This estimator being intractable for the ExpOU model, we use the EM algorithm to maximize the contrast function by defining the intermediary random variables $Z_{0:B-1}$ such that for all $u \in \{0, \dots, B-1\}$, $Y_{u\eta+1:(u+1)\eta}$ and Z_u have the same distribution but $(Z_u)_{u\in\{0,\dots,B-1\}}$ are independent. Exploiting these two properties, the log-likelihood of Z is exactly the proposed contrast function of Y:

$$\log p_\theta(Z_{0:B-1}) = \sum_{u=0}^{B-1} \log p_\theta(Z_u) = \sum_{u=0}^{B-1} \log p_\theta(Y_{u\eta+1:(u+1)\eta}) \ .$$

For any hidden random variable $(V_u)_{u\in\{0,\dots,B-1\}}$, the EM algorithms suggests to recursively compute the sequence $(\hat{\theta}_n)$ defined by

$$\hat{\theta}_{n+1} = \operatorname{argmax}_\theta \mathbb{E}_{\hat{\theta}_n}\left[\log p_\theta(Z_{0:B-1}, V_{0:B-1}) | Z_{0:B-1}\right] \ .$$

We choose V such that for all $u \in \{0, \dots, B-1\}$, $(Y_{u\eta+1:(u+1)\eta}, X_{u\eta}, \tilde{X}_{u\eta+1:(u+1)\eta})$ and (Z_u, V_u) have the same distribution but $(Z_u, V_u)_{u\in\{0,\dots,B-1\}}$ are independent. Then we have

$$\begin{aligned}
\mathbb{E}_{\hat{\theta}_n}\left[\log p_\theta(Z_{0:B-1}, V_{0:B-1}) | Z_{0:B-1}\right] &= \sum_{u=0}^{B-1} \mathbb{E}_{\hat{\theta}_n}\left[\log p_\theta(Z_u, V_u) | Z_{0:B-1}\right] \\
&= \sum_{u=0}^{B-1} \mathbb{E}_{\hat{\theta}_n}\left[\log p_\theta(Z_u, V_u) | Z_u\right] \\
&= \sum_{u=0}^{B-1} \mathbb{E}_{\hat{\theta}_n}\left[\log p_\theta(Y_{u\eta+1:(u+1)\eta}, X_{u\eta}, \tilde{X}_{u\eta+1:(u+1)\eta}) \middle| Y_{u\eta+1:(u+1)\eta}\right] \ .
\end{aligned}$$

Consequently, we can replace the quantity defined in (5.11) with

$$L_{n,\theta}^{B,T}(Y_{1:T}) = \sum_{u=0}^{B-1} \ell_{n,\theta}^{u\eta+1,(u+1)\eta}\left(Y_{u\eta+1:(u+1)\eta}\right) \ . \tag{5.13}$$

This splitting technique points out a new issue: the choice of η which is directly linked to the estimator efficiency. As already mentioned, the identifiability result obtained in Theorem 5.1 shows that η should be necessarily greater than $2p$, and small enough to avoid degeneracy of the smoothing algorithm.

5.5 Expectation step

In section 5.4 we have seen that inference about the parameter θ based on a fixed set of observations $Y_{1:T}$ requires to evaluate expectations of the form given in (5.12). To that purpose, we aim to approximate

the smoothing distribution of $\left(X_0, \tilde{X}_{1:T}\right)$ conditionally to $Y_{1:T}$.
For the block EM algorithm, the smoothing distributions to approximate are for all $u \in \{0, \dots, B-1\}$, $\left(X_{u\frac{T}{B}}, \tilde{X}_{u\frac{T}{B}+1:(u+1)\frac{T}{B}}\right)$ knowing $Y_{u\frac{T}{B}+1:(u+1)\frac{T}{B}}$. X_0 following the stationary distribution of the Markov chain X, we only need to focus on $\left(X_0, \tilde{X}_{1:T/B}\right)$ conditionally to $Y_{1:T/B}$. For clarity reasons, we will keep T as the time horizon.
Standard smoothing algorithms could be performed only if $\tilde{X}$ was a Markov chain. Unfortunately it is not (see Remark 5.2 in Appendix 5.A.1) and we have to replace it by a more suitable hidden variable before exhibiting two closely linked smoothing algorithms: the genealogical tree and the forward filtering backward simulation algorithms.

5.5.1 Replacing the hidden variable

An obvious idea to handle the fact that $\tilde{X}$ is not in general a Markov chain is to recall that $\tilde{X}_k = \langle 1_p, X_k \rangle$ where X is a Markov chain, such that the smoothing distribution of X could be approximated with a particle smoother. However, this Markov chain is degenerated, preventing from using most smoothing algorithms. This issue can be overcome by a decimation in time for X. To that purpose, we define the p-dimensional random vector $\overline{X}$ for all $k \geq 0$ by

$$\forall k \geq 0, \quad \overline{X}_k = X_{pk} \,,$$

and the p-dimensional random vector $\overline{Y}$ for all $k \geq 1$ by

$$\forall k \geq 1, \quad \overline{Y}_k = [Y_{p(k-1)+1}, \cdots, Y_{pk}]^{\mathrm{T}} \,.$$

Then we check that smoothing algorithms can be performed on the hidden model defined by $\overline{X}$ and $\overline{Y}$ in the following proposition.

Proposition 5.2. *$\overline{X}$ is a Markov chain, is not degenerated and*

$$\overline{X}_k | \overline{X}_{k-1} \sim \mathcal{N}\left(\mathrm{diag}\,(\alpha)^p \overline{X}_{k-1}, \mathbf{\Sigma}(p)\mathbf{\Sigma}(p)^{\mathrm{T}}\right) \,, \tag{5.14}$$

where, for all $q \in \{1, \dots, p\}$, $\mathbf{\Sigma}(q)$ is a $p \times p$ matrix defined by:

$$\mathbf{\Sigma}(q) = \begin{bmatrix} 0 & \cdots & 0 & \sigma_1 \alpha_1^0 & \cdots & \sigma_1 \alpha_1^{q-1} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & \sigma_p \alpha_p^0 & \cdots & \sigma_p \alpha_p^{q-1} \end{bmatrix} \,,$$

in particular, $\mathbf{\Sigma}(p) = \mathrm{diag}\,(\sigma)\mathrm{Vdm}\,(\alpha)^{\mathrm{T}}$ and by convention, $\mathbf{\Sigma}(0)$ is the $p \times p$ null matrix.

Furthermore, the law of $\overline{Y}_k$ conditionally to $\overline{X}_{0:k}$ only depends on $(\overline{X}_{k-1}, \overline{X}_k)$ and

$$\overline{Y}_k | \overline{X}_{k-1}, \overline{X}_k \sim \mathcal{N}\left(\rho\beta\left[\mathrm{e}^{\frac{1}{2} f_{\sigma,\alpha,q-1}(\overline{X}_{k-1}, \overline{X}_k)} g_{\sigma,\alpha,q}(\overline{X}_{k-1}, \overline{X}_k)\right]_{1 \leq q \leq p}, (1-\rho^2)\beta^2 \mathrm{diag}\left(\left[\mathrm{e}^{f_{\sigma,\alpha,q-1}(\overline{X}_{k-1}, \overline{X}_k)}\right]_{1 \leq q \leq p}\right)\right) \,, \tag{5.15}$$

where the two function families $(f_{\sigma,\alpha,q})_{0 \leq q \leq p-1}$ and $(g_{\sigma,\alpha,q})_{1 \leq q \leq p}$ are defined for all $q \in \{1, \dots, p\}$, all $x \in \mathbb{R}^p$ and all $x' \in \mathbb{R}^p$ by:

$$\begin{aligned} f_{\sigma,\alpha,q-1}(x, x') &= \left\langle 1_p, \mathrm{diag}\,(\alpha)^{q-1} x + \mathbf{\Sigma}(q-1)\mathbf{\Sigma}(p)^{-1}(x' - \mathrm{diag}\,(\alpha)^p x) \right\rangle \,, \\ g_{\sigma,\alpha,q}(x, x') &= \left[\mathbf{\Sigma}(p)^{-1}(x' - \mathrm{diag}\,(\alpha)^p x)\right]_{p+1-q} \,. \end{aligned}$$

Finally, $\tilde{X}$ can be found back from $\overline{X}$: for all $k \geq 1$ and $q \in \{0, \dots, p-1\}$,

$$\tilde{X}_{p(k-1)+q} = f_{\sigma,\alpha,q}(\overline{X}_{k-1}, \overline{X}_k) \,. \tag{5.16}$$

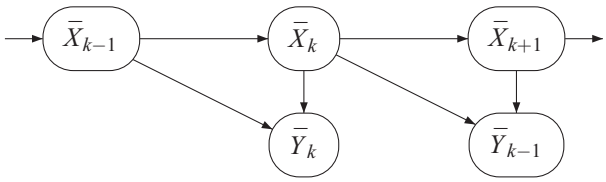


Figure 5.1

Proof. It is direct to see that

$$\bar{X}_k = \mathrm{diag}\,(\alpha)^p \bar{X}_{k-1} + \mathbf{\Sigma}(p)[W_{pk}, \cdots, W_{p(k-1)+1}]^{\mathrm{T}} . \tag{5.17}$$

This shows that $\bar{X}$ is a Markov chain and its transition law is given by (5.14). $\mathbf{\Sigma}(p)$ being invertible, $\bar{X}$ is not degenerated.

Then it is possible to compute the missing X_k through a deterministic function of $\bar{X}$. First, we invert (5.17)

$$[W_{pk}, \cdots, W_{p(k-1)+1}]^{\mathrm{T}} = \mathbf{\Sigma}(p)^{-1}(\bar{X}_k - \mathrm{diag}\,(\alpha)^p \bar{X}_{k-1}) ,$$

so that

$$X_{p(k-1)+q} = \mathrm{diag}\,(\alpha)^q \bar{X}_{k-1} + \mathbf{\Sigma}(q)\mathbf{\Sigma}(p)^{-1}(\bar{X}_k - \mathrm{diag}\,(\alpha)^p \bar{X}_{k-1}) .$$

From the two previous equations, we get

$$\begin{aligned} \tilde{X}_{p(k-1)+q} &= \langle 1_p, X_{p(k-1)+q} \rangle = f_{\sigma,\alpha,q}(\bar{X}_{k-1}, \bar{X}_k) , \\ W_{p(k-1)+q} &= g_{\sigma,\alpha,q}(X_{k-1}, X_k) . \end{aligned}$$

Finally the proof is completed by writting Y as follows

$$Y_{p(k-1)+q} = \beta e^{\frac{1}{2} f_{\sigma,\alpha,q-1}(\bar{X}_{k-1}, \bar{X}_k)} [\rho g_{\sigma,\alpha,q}(\bar{X}_{k-1}, \bar{X}_k) + \sqrt{1-\rho^2}\tilde{V}_{p(k-1)+q}] .$$

□

Consequently, (5.14) and (5.15) can be used in any smoothing algorithm to get an approximation of the law of $\bar{X}_{1:T/p}$ knowing $\bar{Y}_{1:T/p}$. Then, the expectations involved in (5.11) and (5.13) are computed from $\bar{X}_{1:T/p}$ conditionally to $\bar{Y}_{1:T/p}$ using (5.16).

5.5.2 Smoothing algorithms

We consider the model illustrated in figure 5.1. $\bar{X}$ is an unobserved Markov chain and at time k, the observation $\bar{Y}_k$ depends on $(\bar{X}_{k-1}, \bar{X}_k)$ and not only on $\bar{X}_k$ so that standard smoothing algorithms [Doucet et al., 2001] have to be adapted to this particular case. We first show that the classical bootstrap filter can be easily extended to figure 5.1 as it will be the base for most smoothing algorithms. Then we apply it to the genealogical tree method and focus carefully on how the forward smoothing backward simulation (FFBSi) algorithm is impacted by this specific dependence structure.

Adapted bootstrap filter

In the framework of the bootstrap filter, we iteratively approximate the filtering distribution $p(\bar{X}_k|\bar{Y}_{1:k})$ of $\bar{X}_k$ conditionally to $\bar{Y}_{1:k}$ (with convention $p(\bar{X}_0|\bar{Y}_{1:0}) = p(\bar{X}_0)$) by $\widehat{p}_N(\bar{X}_k|\bar{Y}_{1:k})$ using a set of N weighted particles $(\omega_k^{1:N}, \xi_k^{1:N})$ such that

$$\widehat{p}_N(\mathrm{d}\bar{X}_k|\bar{Y}_{1:k}) = \left(\sum_{i=1}^N \omega_k^i\right)^{-1} \sum_{i=1}^N \omega_k^i \delta_{\xi_k^i}(\mathrm{d}\bar{X}_k) ,$$

where δ is the Dirac measure. The way of drawing the particles and the updating formulas for the weights are naturally derived from the following equation: for all $k \geq 1$,

$$p\left(\overline{X}_{k+1}|\overline{Y}_{1:k+1}\right) = \frac{\int p\left(\overline{X}_k|\overline{Y}_{1:k}\right) p\left(\overline{X}_{k+1}|\overline{X}_k\right) p\left(\overline{Y}_{k+1}|\overline{X}_k,\overline{X}_{k+1}\right) \mathrm{d}\overline{X}_k}{\iint p\left(X_k|Y_{1:k}\right) p\left(X_{k+1}|X_k\right) p\left(Y_{k+1}|X_k,X_{k+1}\right) \mathrm{d}X_{k:k+1}} \,.$$

Each iteration of this procedure consists of a selection step (optional) and a mutation step. The selection step resamples particles $\xi_k^{1:N}$ by drawing independent indices $I_k^{1:N}$ with probability proportional to the weights $\omega_k^{1:N}$. Then, the mutation step draws particles $\xi_{k+1}^{1:N}$ independently such that for all $i \in \{1,\ldots,N\}$, $\xi_{k+1}^i \sim p\left(\xi_{k+1}^i \middle| \xi_k^{I_k^i}\right)$. Finally, the weights are updated by setting $\omega_{k+1}^i = p\left(\overline{Y}_{k+1} \middle| \xi_k^{I_k^i}, \xi_{k+1}^i\right)$. This is summarized in algorithm 10.

Algorithm 10 Adapted bootstrap filter

1: sample $(\xi_0^i)_{i=1}^N$ independently according to $p\left(\overline{X}_0\right)$
2: $(\omega_0^1,\ldots,\omega_0^N) \leftarrow (1/N,\ldots,1/N)$
3: **for** k from 0 to $T/p-1$ **do**
4: sample $I_k^{1:N}$ multinomially with probability proportional to $\omega_k^{1:N}$
5: sample $\xi_{k+1}^{1:N}$ independently such that $\xi_{k+1}^i \sim p\left(\xi_{k+1}^i \middle| \xi_k^{I_k^i}\right)$
6: $\omega_{k+1}^{1:N} \leftarrow \left(p\left(\overline{Y}_{k+1} \middle| \xi_k^{I_k^i}, \xi_{k+1}^i\right)\right)_{i=1}^N$
7: **end for**

Adapted genealogical tree

The genealogical tree algorithm [Gordon et al., 1993, Del Moral, 2004] approximates the smoothing distribution $p\left(\overline{X}_{0:T/p}|\overline{Y}_{1:T/p}\right)$ of $\overline{X}_{0:T/p}$ conditionally to $\overline{Y}_{1:T/p}$ by $\widehat{p}_N\left(\overline{X}_{0:T/p}|\overline{Y}_{1:T/p}\right)$ using the particles $\xi_{0:T/p}^{1:N}$, the weights $\omega_{T/p}^{1:N}$ and the genealogy $I_{0:T/p-1}^{1:N}$ computed by the bootstrap filter (algorithm 10) such that

$$\widehat{p}_N\left(\mathrm{d}\overline{X}_{0:T/p}|\overline{Y}_{1:T/p}\right) = \left(\sum_{i=1}^N \omega_{T/p}^i\right)^{-1} \sum_{i=1}^N \omega_{T/p}^i \delta_{\left(\xi_{0:T/p}^{J_{0:T/p}^i}\right)}\left(\mathrm{d}\overline{X}_{0:T/p}\right) \,,$$

where $J_{0:T/p}^{1:N}$ is deterministically defined from the genealogy $I_{0:T/p-1}^{1:N}$ by the following backward recursion for all $i \in \{1,\ldots,N\}$

$$\begin{cases} J_{T/p}^i = i \,, \\ J_k^i = I_k^{J_{k+1}^i}, \quad \forall k \in \{0,\ldots,T/p-1\} \,. \end{cases}$$

Adapted Forward Filtering Backward Simulation

The FFBSi algorithm (described in Doucet et al. 2000 and analyzed in Douc et al. 2010) approximates the smoothing distribution $p\left(\overline{X}_{0:T/p}|\overline{Y}_{1:T/p}\right)$ of $\overline{X}_{0:T/p}$ conditionally to $\overline{Y}_{1:T/p}$ with $\widehat{p}_N\left(\overline{X}_{0:T/p}|\overline{Y}_{1:T/p}\right)$ through two simulation passes. The first one consists of computing the weighted particles $(\omega_{0:T/p}^{1:N}, \xi_{0:T/p}^{1:N})$ using the bootstrap filter (algorithm 10). The second one draws backward N independent paths of particle indices $J_{0:T/p}^{1:N}$ from the set $\{1,\ldots,N\}^{T/p+1}$ such that

$$\widehat{p}_N\left(\mathrm{d}\overline{X}_{0:T/p}|\overline{Y}_{1:T/p}\right) = N^{-1} \sum_{i=1}^N \delta_{\left(\xi_{0:T/p}^{J_{0:T/p}^i}\right)}\left(\mathrm{d}\overline{X}_{0:T/p}\right) \,.$$

The way of drawing the indices $J_{0:T/p}^{1:N}$ is derived from the following equation: for all $k \in \{0,\dots,T/p-1\}$,

$$p\left(\bar{X}_{k:T/p}\middle|\bar{Y}_{1:T/p}\right) = p\left(\bar{X}_{k+1:T/p}\middle|\bar{Y}_{1:T/p}\right) \times \frac{p\left(\bar{X}_k|\bar{Y}_{1:k}\right) p\left(\bar{X}_{k+1}|\bar{X}_k\right) p\left(\bar{Y}_{k+1}|\bar{X}_k,\bar{X}_{k+1}\right)}{\int p\left(X_k|Y_{1:k}\right) p\left(X_{k+1}|X_k\right) p\left(Y_{k+1}|X_k,X_{k+1}\right) \mathrm{d}X_k} .$$

As a consequence, for $i \in \{1,\dots,N\}$, $(J_{0:T/p}^i)$ can be independently sampled backward in time according to

$$\begin{cases} \mathbb{P}\left\{J_{T/p}^i = j\right\} \propto \omega_{T/p}^j , \\ \mathbb{P}\left\{J_k^i = j \middle| J_{k+1}^i\right\} \propto \omega_k^j p\left(\xi_{k+1}^{J_{k+1}^i}\middle|\xi_k^j\right) p\left(\bar{Y}_{k+1}\middle|\xi_k^j,\xi_{k+1}^{J_{k+1}^i}\right), \quad 0 \le k \le T/p-1 . \end{cases}$$

This extension of the FFBSi algorithm is summarized in algorithm 11.

Algorithm 11 Adapted FFBSi

1: sample $(\omega_{0:T/p}^{1:N}, \xi_{0:T/p}^{1:N})$ using algorithm 10
2: sample $J_{T/p}^{1:N}$ multinomially with probability proportional to $\omega_{T/p}^{1:N}$
3: **for** k from $T/p-1$ down to 0 **do**
4: **for** i from 1 to N **do**
5: sample J_k^i with proba. prop. to $\omega_k^j p\left(\xi_{k+1}^{J_{k+1}^i}\middle|\xi_k^j\right) p\left(\bar{Y}_{k+1}\middle|\xi_k^j,\xi_{k+1}^{J_{k+1}^i}\right)$
6: **end for**
7: **end for**

5.6 Application

In this section, we apply the previously described algorithm with $N = 3000$ particles to simulated and real data. We first experimentally show that the proposed technique can capture well up to three time scale factors and identify a heuristic to choose the number of factors. This criterium if based on Remark 5.1. When coming to calibrating real data, we use the detrended returns [Kim et al., 1998] of the CAC 40 and Dow Jones indices over 10 years.

5.6.1 Simulated data

In order to evaluate the performance of the algorithm to identify up to three time scale factors, we have generated $T = 1500$ observations $Y_{1:T}$ for three different settings defined in table 5.1. Then, for each setting, the calibration algorithm has been run for $p \in \{1,2,3\}$ with 100 EM iterations, $N = 3000$ particles and a different block size η for each value of p given in table 5.2. As seen in Section 5.3 and Subsection 5.4.2, choosing η depending on p is crucial in order to guaranty the identifiability which ensures that $\hat{\theta}_n$ converges to the true value of the parameters. The convergence of the algorithm is shown in figures 5.2, 5.3, 5.4 for their respective values of p and in figures 5.7, 5.8, 5.9 in Appendix 5.B for other values of p. The estimated parameters are given in table 5.3.

These results show that the parameter θ is accurately estimated by our algorithm for up to three factors when knowing the value of p even though the convergence might occur after an excursion of $\hat{\theta}_n$ far from the true value as the maximization steps do not take into account any *a priori* knowledge of the parameters. Moreover, the estimation of θ with the wrong number of factors leads to similar values of (β,ρ). The estimated (α,σ) cannot be compared directly as they do not belong to the same space. Following Remark 5.1, we compare instead the auto-covariance function $\gamma^{\tilde{Y}}$ defined in (5.6) which characterizes the law of $\tilde{X}$. For any setting, we denote by $\gamma^{\tilde{Y},p}$ the auto-covariance function induced by (α,σ) estimated with p time

Table 5.1: Setting's definition

	Setting 1	Setting 2	Setting 3
p	1	2	3
α_1	0.6	0.3	0.3
α_2	-	0.98	0.6
α_3	-	-	0.98
σ_1	0.4	0.8	0.8
σ_2	-	0.2	0.6
σ_3	-	-	0.2
β	0.2	0.2	0.2
ρ	-0.1	-0.2	-0.1

Table 5.2: Block size

p	1	2	3
η	15	20	30

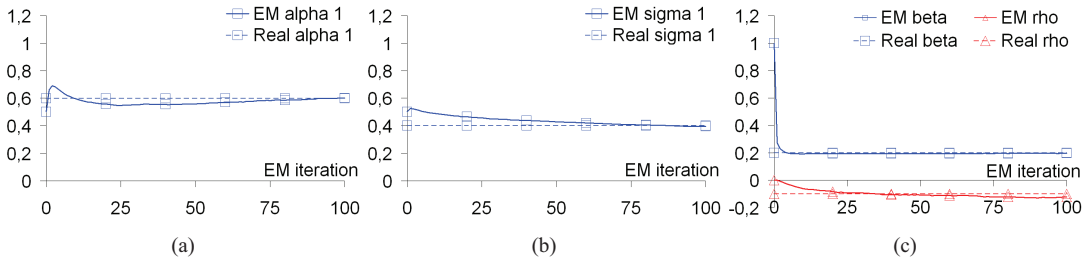


Figure 5.2: Calibration of setting 1 for $p = 1$

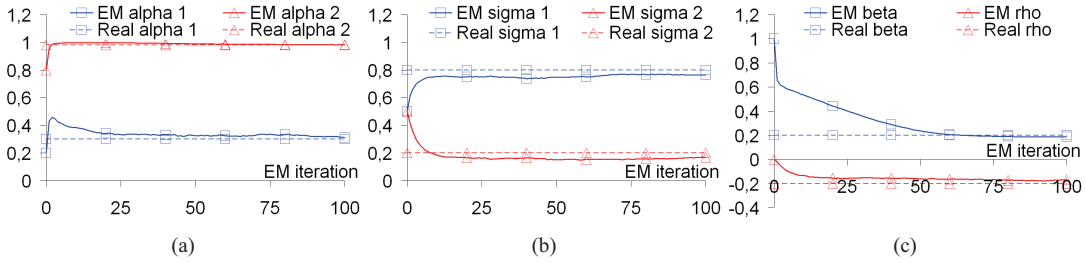
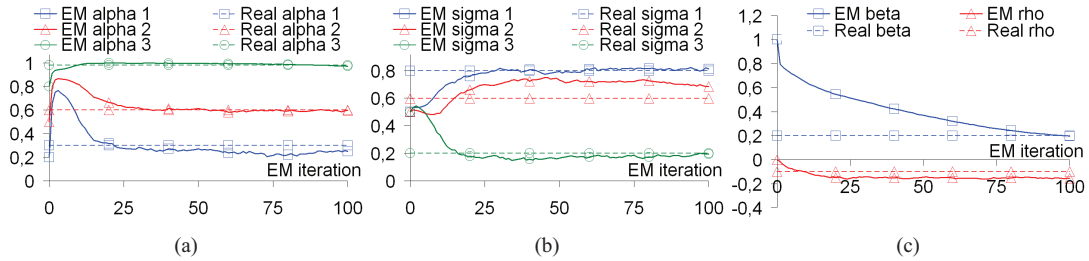


Figure 5.3: Calibration of setting 2 for $p = 2$



(a) (b) (c)

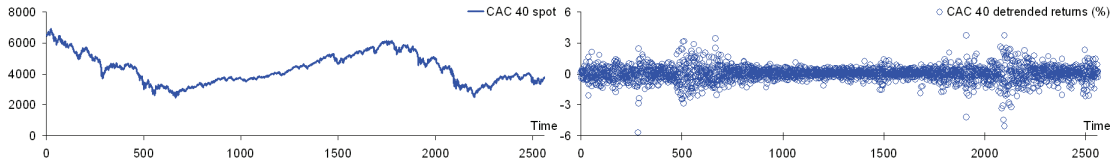
Figure 5.4: Calibration of setting 3 for $p = 3$

Table 5.3: Estimated parameters from simulated data
(real parameter between brackets when applicable)

	Setting 1			Setting 2			Setting 3		
p	1	2	3	1	2	3	1	2	3
α_1	0.60 (0.6)	0.32	0.10	0.82	0.31 (0.3)	0.19	0.62	0.38	0.25 (0.3)
α_2	-	0.90	0.59	-	0.98 (0.98)	0.67	-	0.96	0.60 (0.6)
α_3	-	-	0.997	-	-	0.99	-	-	0.97 (0.98)
σ_1	0.39 (0.4)	0.37	0.21	0.76	0.76 (0.8)	0.45	1.67	1.51	0.81 (0.8)
σ_2	-	0.08	0.23	-	0.17 (0.2)	0.38	-	0.24	0.67 (0.6)
σ_3	-	-	0.11	-	-	0.12	-	-	0.20 (0.2)
β	0.20 (0.2)	0.20	0.20	0.19	0.19 (0.2)	0.20	0.17	0.17	0.20 (0.2)
ρ	-0.12 (-0.1)	-0.13	-0.11	-0.15	-0.17 (-0.2)	-0.18	-0.15	-0.16	-0.15 (-0.1)

Table 5.4: Differences of auto-covariance functions estimated from simulated data

	Setting 1	Setting 2	Setting 3
$d_{1,2}$	0.09	4.37	3.45
$d_{2,3}$	0.24	0.78	1.12



(a) (b)

Figure 5.5: Daily CAC 40 spot (left) and detrended returns (right) for the period 01/08/2000-09/08/2010

scale factors. Consequently, a possible measure of the difference between the law estimated with p_1 and p_2 factors is

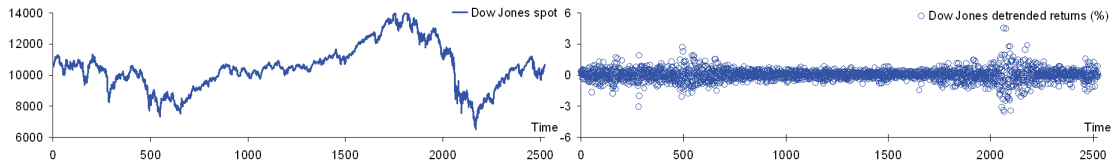
$$d_{p_1,p_2} \overset{\text{def}}{=} \left|\left|\gamma^{\tilde{X},p_1}_{0:99} - \gamma^{\tilde{X},p_2}_{0:99}\right|\right|_2 = \sqrt{\sum_{k=0}^{99}\left(\gamma^{\tilde{X},p_1}_{k} - \gamma^{\tilde{X},p_2}_{k}\right)^2}. \tag{5.18}$$

This quantity is computed for each setting in table 5.4. As expected, the difference in law when going from one to two or from two to three factors is very small for setting 1, meaning that one factor is sufficient to model the data. For setting 2, the difference is important when going from one to two factors and negligible when going from two to three factors, meaning that one factor is not enough to model the data and that two factors are sufficient to do so. Finally, for setting 3, the difference in law when going from one to two or from two to three factors is significant, meaning that at least three factors are needed.

5.6.2 Real data

The calibration algorithm has been applied to equity market data: the CAC 40 and Dow Jones indices over the past ten years. Their daily spot and returns (detrended according to Kim et al. 1998) are plotted in figures 5.5 and 5.6.

The estimated parameters with different numbers of factors are presented in table 5.5 and the convergence of the algorithm is displayed in Appendix 5.B, figures 5.10 and 5.11. An analysis similar to the one of the simulated data can be done. First, we remark that the estimated (β, ρ) are approximately the same whichever the value of p and the negative value of ρ is consistent with the leverage effect announced in the introduction [Black, 1976, Christie, 1982]. Then, we compute the auto-covariance differences (defined



(a) (b)

Figure 5.6: Daily Dow Jones spot (left) and detrended returns (right) for the period 01/08/2000-09/08/2010

Table 5.5: Estimated parameters from CAC 40 and Dow Jones

	CAC 40			Dow Jones		
p	1	2	3	1	2	3
α_1	0.96	0.00	0.00	0.97	0.00	0.00
α_2	-	0.98	0.26	-	0.99	0.28
α_3	-	-	0.98	-	-	0.99
σ_1	0.27	0.30	0.17	0.26	0.37	0.17
σ_2	-	0.19	0.15	-	0.12	0.17
σ_3	-	-	0.18	-	-	0.14
β	0.53	0.53	0.53	0.41	0.41	0.41
ρ	-0.58	-0.58	-0.54	-0.58	-0.45	-0.53

Table 5.6: Differences of auto-covariance functions estimated from CAC 40 and Dow Jones

	CAC 40	Dow Jones
$d_{1,2}$	1.11	3.45
$d_{2,3}$	0.36	0.80

in (5.18)) in table 5.6 to select the most appropriate number of factors. For both the CAC 40 and Dow Jones indices, the difference is important when going from one to two factors and negligible when going from two to three factors, meaning that one factor is not enough to model the data and that two factors are sufficient to do so. Consequently, we have captured a short time scale factor (about one day) and a long one (a few months). Identification of these two different time scales is a desired characteristic of the multiscale exponential Ornstein-Uhlenbeck model (see Chernov et al. 2003).

5.7 Conclusion

This paper exhibits a tractable and efficient way of calibrating a stationary multiscale exponential Ornstein-Uhlenbeck model including a correlation between the asset and its volatility. To do so, a precise identifiability result justifies the use of the MSDLE. This estimate is then approximated by a stochastic EM algorithm involving particle smoothing. Estimation in non exponential model with degenerated noise requires to be particularly careful about the hidden variables which will be chosen differently for each step of the estimation method design (identifiability, E-step/M-step separation and particle smoothing). More generally, this provides guidelines concerning the design of estimation algorithm in non standard framework.

Experiments show that the proposed algorithm is able to identify up to three time scale factors and an intuitive heuristic allows to select the number of factors. This criterium gives a measure of the difference in law induced when running the estimation algorithm with various number of factors. The choice of one, two or three factors can be driven by this heuristic.

5.A Technical proofs

5.A.1 Proof of identifiability

Proposition 5.3. *The model* (5.3) *can be written as follows:*

$$\begin{cases} \tilde{X}_k = \langle \alpha^k, X_0 \rangle + \sum_{\ell=1}^k \langle \sigma, \alpha^{k-\ell} \rangle W_\ell, \\ Y_k = \beta e^{\frac{1}{2}\tilde{X}_{k-1}} \left(\rho \sum_{\ell=1}^k \omega_{k-\ell} \left[\tilde{X}_\ell - \langle \alpha^\ell, X_0 \rangle \right] + \sqrt{1-\rho^2}\tilde{V}_k \right) \end{cases} \tag{5.19}$$

where $\{(W_k, \tilde{V}_k)\}_{k\geq 1}$ *is a sequence of iid two dimensional independent standard gaussian vectors,* $\alpha^k = (\alpha_1^k, \cdots, \alpha_p^k)^{\mathrm{T}}$ *and* $(\omega_k)_{k\geq 0}$ *is defined in* (5.9)*.*

Remark 5.2. By straightforward algebra, one can invert the first equation of (5.19), $W_\ell = \sum_{m=1}^\ell \omega_{\ell-m} \left[\tilde{X}_m - \langle \alpha^m, X_0 \rangle \right]$ and plug it again into (5.19) seeing that $\tilde{X}$ is not in general a Markov chain. Nevertheless, this is not an issue when proving the identifiability of the model.

Proof. By expanding (5.3), we obtain for $i \in \{1, \ldots, p\}$, $X_k^i = \alpha_i^k X_0^i + \sum_{\ell=1}^k \sigma_i \alpha_i^{k-\ell} W_\ell$ so that

$$\tilde{X}_k = \langle 1_p, X_k \rangle = \left\langle \alpha^k, X_0 \right\rangle + \sum_{\ell=1}^k \left\langle \sigma, \alpha^{k-\ell} \right\rangle W_\ell \,. \tag{5.20}$$

A direct calculation from (5.20) shows that

$$W_k = \sum_{\ell=1}^k \omega_{k-\ell} \left[\tilde{X}_\ell - \left\langle \alpha^\ell, X_0 \right\rangle \right] \,, \tag{5.21}$$

where $(\omega_k)_{k\geq 0}$ is defined in (5.9). As $\mathrm{Corr}(W_k, V_\ell) = \rho \mathbb{1}_{k=\ell}$, the proof is completed by rewritting the random variable V_k as follows

$$V_k = \rho W_k + \sqrt{1-\rho^2}\tilde{V}_k = \rho \sum_{\ell=1}^k \omega_{k-\ell} \left[\tilde{X}_\ell - \left\langle \alpha^\ell, X_0 \right\rangle \right] + \sqrt{1-\rho^2}\tilde{V}_k \,,$$

where $\{(W_k, \tilde{V}_k)\}_{k\geq 1}$ is a sequence of iid two dimensional independent standard gaussian vectors. □

Lemma 5.1. *Let* $\theta^{(i)} = (\alpha^{(i)}, \sigma^{(i)}, \beta^{(i)}, \rho^{(i)})$*,* $i \in \{1,2\}$ *be two sets of parameters in* Θ_p *and define two pairs of processes* $(X^{(i)}, Y^{(i)})$*,* $i \in \{1,2\}$ *by* (5.5)*. Then, for all* $k \geq 1$*,*

$$Y_{1:k+1}^{(1)} \overset{\mathcal{L}}{=} Y_{1:k+1}^{(2)} \Longrightarrow \begin{cases} \beta^{(1)} = \beta^{(2)} \,, \\ \rho^{(1)} \left\langle \sigma^{(1)}, (\alpha^{(1)})^m \right\rangle = \rho^{(2)} \left\langle \sigma^{(2)}, (\alpha^{(2)})^m \right\rangle, & \forall m \leq k-1 \,, \\ 1_p^{\mathrm{T}} \Upsilon_{\alpha^{(1)}, \sigma^{(1)}} (\alpha^{(1)})^m = 1_p^{\mathrm{T}} \Upsilon_{\alpha^{(2)}, \sigma^{(2)}} (\alpha^{(2)})^m, & \forall m \leq k \,, \end{cases}$$

Proof. To obtain the distribution of the Y, we substitute the expression of $\tilde{X}$ in (5.19) so that for all $k \geq 1$,

$$Y_k^{(i)} = \beta^{(i)} e^{\frac{1}{2}\left\langle (\alpha^{(i)})^{k-1}, X_0^{(i)} \right\rangle} e^{\frac{1}{2}\sum_{\ell=1}^{k-1} \langle \sigma^{(i)}, (\alpha^{(i)})^{k-1-\ell} \rangle W_\ell^{(i)}} \left(\rho^{(i)} W_k^{(i)} + \sqrt{1-(\rho^{(i)})^2}\tilde{V}_k^{(i)} \right) \,, \tag{5.22}$$

where $\{W_k^{(1)}, \tilde{V}_k^{(1)}, W_k^{(2)}, \tilde{V}_k^{(2)},\ k \geq 1\}$ are i.i.d. standard gaussian variables and $X_0^{(i)} \sim \mathcal{N}\left(0, \Upsilon_{\alpha^{(i)}, \sigma^{(i)}}\right)$.

Let $k \geq 1$ and assume that $Y_{1:k+1}^{(1)} \overset{\mathcal{L}}{=} Y_{1:k+1}^{(2)}$. Then, for all $s \in \mathbb{N}^*$, $\mathbb{E}\left[(Y_1^{(1)})^{2s}\right] = \mathbb{E}\left[(Y_1^{(2)})^{2s}\right]$ and for $i \in \{1,2\}$,

$$\mathbb{E}\left[(Y_1^{(i)})^{2s}\right] = (\beta^{(i)})^{2s} \exp\left(\frac{1}{2} s^2 1_p^{\mathrm{T}} \Upsilon_{\alpha^{(i)}, \sigma^{(i)}} 1_p \right) \mathbb{E}\left[(V_1^{(i)})^{2s}\right] \,,$$

so that $2\log\beta^{(1)} + \frac{1}{2}1_p^{\mathrm{T}}\boldsymbol{\Upsilon}_{\alpha^{(1)},\sigma^{(1)}}1_p s = 2\log\beta^{(2)} + \frac{1}{2}1_p^{\mathrm{T}}\boldsymbol{\Upsilon}_{\alpha^{(2)},\sigma^{(2)}}1_p s$ implying $\beta^{(1)} = \beta^{(2)}$ and $1_p^{\mathrm{T}}\boldsymbol{\Upsilon}_{\alpha^{(1)},\sigma^{(1)}}1_p = 1_p^{\mathrm{T}}\boldsymbol{\Upsilon}_{\alpha^{(2)},\sigma^{(2)}}1_p$. Furthermore, for all $2 \leq m \leq k+1$ and all $s \in \mathbb{N}^*$, $\mathbb{E}\left[(Y_m^{(1)})^{2s}Y_1^{(1)}\right] = \mathbb{E}\left[(Y_m^{(2)})^{2s}Y_1^{(2)}\right]$, and by noting that for $i \in \{1,2\}$,

$$((\alpha^{(i)})^{m-1})^{\mathrm{T}}\boldsymbol{\Upsilon}_{\alpha^{(i)},\sigma^{(i)}}(\alpha^{(i)})^{m-1} + \sum_{\ell=0}^{m-2}\left\langle \sigma^{(i)}, (\alpha^{(i)})^{\ell}\right\rangle^2 = 1_p^{\mathrm{T}}\boldsymbol{\Upsilon}_{\alpha^{(i)},\sigma^{(i)}}1_p\,,$$

we have

$$\mathbb{E}\left[(Y_m^{(i)})^{2s}Y_1^{(i)}\right] = (\beta^{(i)})^{2s+1}\rho^{(i)}s\left\langle \sigma^{(i)}, (\alpha^{(i)})^{m-2}\right\rangle \mathbb{E}\left[(V_m^{(i)})^{2s}\right] \times \exp\left[\left(\frac{1}{8}+\frac{s^2}{2}\right)1_p^{\mathrm{T}}\boldsymbol{\Upsilon}_{\alpha^{(i)},\sigma^{(i)}}1_p + \frac{s}{2}1_p^{\mathrm{T}}\boldsymbol{\Upsilon}_{\alpha^{(i)},\sigma^{(i)}}(\alpha^{(i)})^{m-1}\right],$$

so that

$$\rho^{(1)}\left\langle \sigma^{(1)}, (\alpha^{(1)})^{m-2}\right\rangle \exp\left[\frac{s}{2}1_p^{\mathrm{T}}\boldsymbol{\Upsilon}_{\alpha^{(1)},\sigma^{(1)}}(\alpha^{(1)})^{m-1}\right] = \rho^{(2)}\left\langle \sigma^{(2)}, (\alpha^{(2)})^{m-2}\right\rangle \exp\left[\frac{s}{2}1_p^{\mathrm{T}}\boldsymbol{\Upsilon}_{\alpha^{(2)},\sigma^{(2)}}(\alpha^{(2)})^{m-1}\right]. \quad (5.23)$$

As a consequence, if $\rho^{(1)} \neq 0$ or $\rho^{(2)} \neq 0$, then $\rho^{(1)} \neq 0$ and $\rho^{(2)} \neq 0$ and (5.23) leads to $\rho^{(1)}\left\langle \sigma^{(1)}, (\alpha^{(1)})^{m-2}\right\rangle = \rho^{(2)}\left\langle \sigma^{(2)}, (\alpha^{(2)})^{m-2}\right\rangle$ and $1_p^{\mathrm{T}}\boldsymbol{\Upsilon}_{\alpha^{(1)},\sigma^{(1)}}(\alpha^{(1)})^{m-1} = 1_p^{\mathrm{T}}\boldsymbol{\Upsilon}_{\alpha^{(2)},\sigma^{(2)}}(\alpha^{(2)})^{m-1}$.
On the other hand, if $\rho^{(1)} = 0$ or $\rho^{(2)} = 0$, then $\rho^{(1)} = \rho^{(2)} = 0$ and for $i \in \{1,2\}$

$$\mathbb{E}\left[(Y_k^{(i)}Y_1^{(i)})^2\right] = (\beta^{(i)})^4 e^{1_p^{\mathrm{T}}\boldsymbol{\Upsilon}_{\alpha^{(i)},\sigma^{(i)}}1_p + 1_p^{\mathrm{T}}\boldsymbol{\Upsilon}_{\alpha^{(i)},\sigma^{(i)}}(\alpha^{(i)})^{m-1}}\mathbb{E}\left[(V_k^{(i)}V_1^{(i)})^2\right],$$

which leads again to $1_p^{\mathrm{T}}\boldsymbol{\Upsilon}_{\alpha^{(1)},\sigma^{(1)}}(\alpha^{(1)})^{m-1} = 1_p^{\mathrm{T}}\boldsymbol{\Upsilon}_{\alpha^{(2)},\sigma^{(2)}}(\alpha^{(2)})^{m-1}$. □

Proof of Theorem 5.1. ii) $\Rightarrow$ *iii)* and *iii)* $\Rightarrow$ *i)* are obvious. Let's show *i)* $\Rightarrow$ *ii)*. If *i)* holds true, then by Lemma 5.1 we have $\beta^{(1)} = \beta^{(2)}$,

$$\forall k \in \{0,\ldots,2p-1\}, \quad 1_p^{\mathrm{T}}\boldsymbol{\Upsilon}_{\alpha^{(1)},\sigma^{(1)}}(\alpha^{(1)})^k = 1_p^{\mathrm{T}}\boldsymbol{\Upsilon}_{\alpha^{(2)},\sigma^{(2)}}(\alpha^{(2)})^k\,, \quad (5.24)$$

$$\rho^{(1)}\left\langle \sigma^{(1)}, 1_p\right\rangle = \rho^{(2)}\left\langle \sigma^{(2)}, 1_p\right\rangle. \quad (5.25)$$

(5.25) shows that if we assume $\sigma^{(1)} = \sigma^{(2)}$, then $\rho^{(1)} = \rho^{(2)}$. As a consequence it only remains to show that $\alpha^{(1)} = \alpha^{(2)}$ and $\sigma^{(1)} = \sigma^{(2)}$. To that purpose, we rewrite (5.24) into the following equivalent assertion

$$\forall k \in \{0,\ldots,p\}, \quad \begin{bmatrix} \left(\alpha_1^{(1)}\right)^k & \cdots & \left(\alpha_p^{(1)}\right)^k \\ \vdots & & \vdots \\ \left(\alpha_1^{(1)}\right)^{k+p-1} & \cdots & \left(\alpha_p^{(1)}\right)^{k+p-1} \end{bmatrix} \boldsymbol{\Upsilon}_{\alpha^{(1)},\sigma^{(1)}}1_p = \begin{bmatrix} \left(\alpha_1^{(2)}\right)^k & \cdots & \left(\alpha_p^{(2)}\right)^k \\ \vdots & & \vdots \\ \left(\alpha_1^{(2)}\right)^{k+p-1} & \cdots & \left(\alpha_p^{(2)}\right)^{k+p-1} \end{bmatrix} \boldsymbol{\Upsilon}_{\alpha^{(2)},\sigma^{(2)}}1_p\,,$$

which is equivalent to, for all $k \in \{0,\ldots,p\}$,

$$\mathrm{Vdm}\left(\alpha^{(1)}\right)\mathrm{diag}\left(\alpha^{(1)}\right)^k\boldsymbol{\Upsilon}_{\alpha^{(1)},\sigma^{(1)}}1_p = \mathrm{Vdm}\left(\alpha^{(2)}\right)\mathrm{diag}\left(\alpha^{(2)}\right)^k\boldsymbol{\Upsilon}_{\alpha^{(2)},\sigma^{(2)}}1_p\,,$$

where $\mathrm{Vdm}\left(\alpha^{(i)}\right)$ denotes the invertible $p \times p$ Vandermonde matrix of $\alpha^{(i)}$:

$$\mathrm{Vdm}\left(\alpha^{(i)}\right) = \begin{bmatrix} 1 & \cdots & 1 \\ \alpha_1^{(i)} & \cdots & \alpha_p^{(i)} \\ \vdots & & \vdots \\ \left(\alpha_1^{(i)}\right)^{p-1} & \cdots & \left(\alpha_p^{(i)}\right)^{p-1} \end{bmatrix} . \tag{5.26}$$

We now introduce a sequence $(Z_k)_{0\le k\le p}$ of $\mathbb{R}^p$-vectors defined by: for all k such that $0 \le k \le p$,

$$Z_k = \mathrm{Vdm}\left(\alpha^{(i)}\right) \mathrm{diag}\left(\alpha^{(i)}\right)^k \boldsymbol{\Upsilon}_{\alpha^{(i)},\sigma^{(i)}} 1_p ,$$

which is well defined since the rhs does not depend on $i \in \{1,2\}$. Since $\boldsymbol{\Upsilon}_{\alpha^{(i)},\sigma^{(i)}} 1_p = \mathrm{Vdm}\left(\alpha^{(i)}\right)^{-1} Z_0$, we have

$$\begin{aligned} Z_k &= \mathrm{Vdm}\left(\alpha^{(i)}\right) \mathrm{diag}\left(\alpha^{(i)}\right)^k \mathrm{Vdm}\left(\alpha^{(i)}\right)^{-1} Z_0 \\ &= \left\{ \mathrm{Vdm}\left(\alpha^{(i)}\right) \mathrm{diag}\left(\alpha^{(i)}\right) \mathrm{Vdm}\left(\alpha^{(i)}\right)^{-1} \right\}^k Z_0 \\ &= \mathrm{Vdm}\left(\alpha^{(i)}\right) \mathrm{diag}\left(\alpha^{(i)}\right) \mathrm{Vdm}\left(\alpha^{(i)}\right)^{-1} Z_{k-1} , \end{aligned}$$

so that for all $k \in \{0,\dots,p-1\}$, Z_{k+1} is equal to

$$\mathrm{Vdm}\left(\alpha^{(1)}\right) \mathrm{diag}\left(\alpha^{(1)}\right) \mathrm{Vdm}\left(\alpha^{(1)}\right)^{-1} Z_k = \mathrm{Vdm}\left(\alpha^{(2)}\right) \mathrm{diag}\left(\alpha^{(2)}\right) \mathrm{Vdm}\left(\alpha^{(2)}\right)^{-1} Z_k . \tag{5.27}$$

Assume first that $\{Z_0,\dots,Z_{p-1}\}$ forms a base of $\mathbb{R}^p$, then (5.27) for $k \in \{0,\dots,p-1\}$ implies the equality of the matrices:

$$\mathrm{Vdm}\left(\alpha^{(1)}\right) \mathrm{diag}\left(\alpha^{(1)}\right) \mathrm{Vdm}\left(\alpha^{(1)}\right)^{-1} = \mathrm{Vdm}\left(\alpha^{(2)}\right) \mathrm{diag}\left(\alpha^{(2)}\right) \mathrm{Vdm}\left(\alpha^{(2)}\right)^{-1} .$$

By uniqueness of the eigenvalues, $\alpha^{(1)} = \alpha^{(2)}$ and

$$\boldsymbol{\Upsilon}_{\alpha^{(1)},\sigma^{(1)}} 1_p = \mathrm{Vdm}\left(\alpha^{(1)}\right)^{-1} Z_0 = \mathrm{Vdm}\left(\alpha^{(2)}\right)^{-1} Z_0 = \boldsymbol{\Upsilon}_{\alpha^{(2)},\sigma^{(2)}} 1_p ,$$

which implies that for all $q \in \{1,\dots,p\}$,

$$\sum_{\ell=1}^p \frac{\sigma_q^{(1)}\sigma_\ell^{(1)}}{1-\alpha_q^{(1)}\alpha_\ell^{(1)}} = \sum_{\ell=1}^p \frac{\sigma_q^{(2)}\sigma_\ell^{(2)}}{1-\alpha_q^{(1)}\alpha_\ell^{(1)}} .$$

This can be stated as follows: for all $q \in \{1,\dots,p\}$,

$$\sum_{\ell=1}^p \frac{\sigma_q^{(1)}\sigma_\ell^{(1)}}{1-\alpha_q^{(1)}\alpha_\ell^{(1)}} \left(\frac{\sigma_q^{(2)}}{\sigma_q^{(1)}} \frac{\sigma_\ell^{(2)}}{\sigma_\ell^{(1)}} - 1 \right) = 0 . \tag{5.28}$$

By introducing $m = \mathrm{argmax}_{\ell\in\{1,\dots,p\}} \sigma_\ell^{(2)}/\sigma_\ell^{(1)}$, we get from (5.28) that for all $q \in \{1,\dots,p\}$,

$$0 \le \left(\frac{\sigma_q^{(2)}}{\sigma_q^{(1)}} \frac{\sigma_m^{(2)}}{\sigma_m^{(1)}} - 1 \right) \sum_{\ell=1}^p \frac{\sigma_q^{(1)}\sigma_\ell^{(1)}}{1-\alpha_q^{(1)}\alpha_\ell^{(1)}} ,$$

implying that for all $q \in \{1,\dots,p\}$, $\dfrac{\sigma_q^{(2)}}{\sigma_q^{(1)}} \dfrac{\sigma_m^{(2)}}{\sigma_m^{(1)}} \ge 1$. Consequently, when writing (5.28) for $q = m$, we get a sum of non-negative terms being equal to 0. This shows that each term $\dfrac{\sigma_m^{(1)}\sigma_\ell^{(1)}}{1-\alpha_m^{(1)}\alpha_\ell^{(1)}} \left(\dfrac{\sigma_m^{(2)}}{\sigma_m^{(1)}} \dfrac{\sigma_\ell^{(2)}}{\sigma_\ell^{(1)}} - 1 \right)$,

$\ell \in \{1,\dots,p\}$ is null, i.e.

$$\forall \ell \in \{1,\dots,p\}, \quad \frac{\sigma_m^{(2)}}{\sigma_m^{(1)}} \frac{\sigma_\ell^{(2)}}{\sigma_\ell^{(1)}} = 1 \, .$$

For $\ell = m$ we get $\sigma_m^{(1)} = \sigma_m^{(2)}$ and finally $\sigma^{(1)} = \sigma^{(2)}$. Thus, (iii) is shown.

It remains to prove that $\{Z_0,\dots,Z_{p-1}\}$ is a base of $\mathbb{R}^p$. To that purpose, we only need to show that this family of p vectors in $\mathbb{R}^p$ is linearly independent. For all linear combination of this family associated to the weights $\lambda = (\lambda_0,\cdots,\lambda_{p-1})^{\mathrm{T}}$, we have

$$\sum_{k=0}^{p-1} \lambda_k Z_k = \mathrm{Vdm}\left(\alpha^{(i)}\right) \mathrm{diag}\left(\Upsilon_{\alpha^{(i)},\sigma^{(i)}} 1_p\right) \left(\sum_{k=0}^{p-1} \lambda_k \left(\alpha_1^{(i)}\right)^k, \cdots, \sum_{k=0}^{p-1} \lambda_k \left(\alpha_p^{(i)}\right)^k \right)^{\mathrm{T}} .$$

Furthermore, the components of $\Upsilon_{\alpha^{(i)},\sigma^{(i)}} 1_p$ are all positive, and $\alpha_j^{(i)} \neq \alpha_k^{(i)}$ for $j \neq k$ implying that $\mathrm{Vdm}\left(\alpha^{(i)}\right)$ and $\mathrm{diag}\left(\Upsilon_{\alpha^{(i)},\sigma^{(i)}} 1_p\right)$ are invertible. The linear independence is then deduced from

$$\sum_{k=0}^{p-1} \lambda_k Z_k = 0 \Leftrightarrow \forall j \in \{1,\dots,p\}, \quad \sum_{k=0}^{p-1} \lambda_k \left(\alpha_j^{(i)}\right)^k = 0 \Leftrightarrow \lambda^{\mathrm{T}} \mathrm{Vdm}\left(\alpha^{(i)}\right) = 0 \Leftrightarrow \lambda = 0 \, ,$$

which concludes the proof. □

5.A.2 Complete data likelihood

Proof of Proposition 5.1. $p_\theta(Y_{1:T}, X_0, \tilde{X}_{1:T})$ can be decomposed as follows

$$p_\theta(Y_{1:T}, X_0, \tilde{X}_{1:T}) = p_\theta(X_0) \prod_{k=1}^{T} p_\theta(Y_k|X_0, \tilde{X}_{1:k}) p_\theta(\tilde{X}_k|X_0, \tilde{X}_{1:k-1}) \, , \tag{5.29}$$

with the convention $p_\theta(\tilde{X}_1|X_0, \tilde{X}_{1:0}) = p_\theta(\tilde{X}_k|X_0)$. The explicit computation of (5.29) is obtained from the following conditional laws. From (5.19), we have for all $k \geq 1$,

$$Y_k \big| X_0, \tilde{X}_{1:k} \sim \mathcal{N}\left(\beta\rho \mathrm{e}^{\frac{1}{2}\tilde{X}_{k-1}} \sum_{\ell=1}^{k} \omega_{k-\ell} \left[\tilde{X}_\ell - \left\langle \alpha^\ell, X_0 \right\rangle \right], \beta^2(1-\rho^2)\mathrm{e}^{\tilde{X}_{k-1}} \right) .$$

From (5.20) and (5.21), it can be shown that

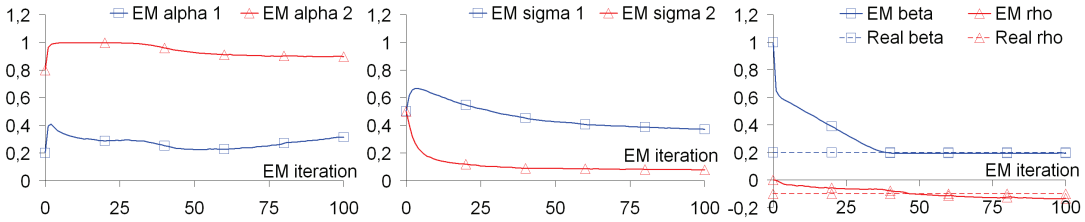
$$\tilde{X}_k = \langle \sigma, 1_p \rangle W_k + \left\langle \alpha^k, X_0 \right\rangle - \sum_{\ell=1}^{k-1} \langle \sigma, 1_p \rangle \omega_{k-\ell} \left[\tilde{X}_\ell - \left\langle \alpha^\ell, X_0 \right\rangle \right] ,$$

and for all $k \geq 1$,

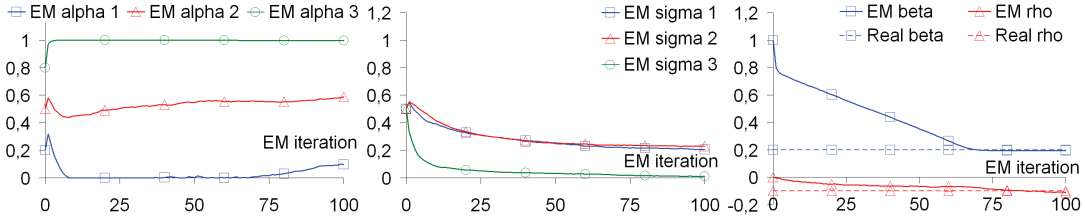
$$\tilde{X}_k \big| X_0, \tilde{X}_{1:k-1} \sim \mathcal{N}\left(\left\langle \alpha^k, X_0 \right\rangle - \sum_{\ell=1}^{k-1} \langle \sigma, 1_p \rangle \omega_{k-\ell} \left[\tilde{X}_\ell - \left\langle \alpha^\ell, X_0 \right\rangle \right], \langle \sigma, 1_p \rangle^2 \right) .$$

□

5.B Additional graphs

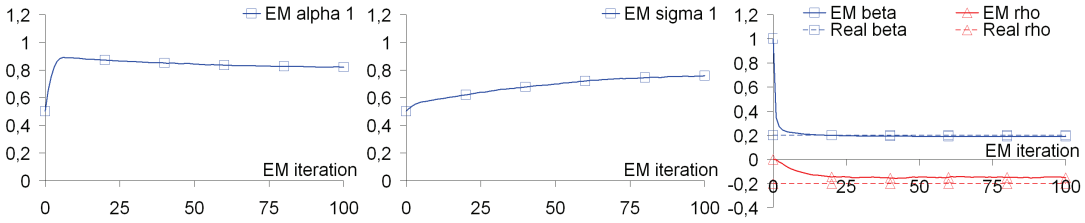


(a) α for $p = 2$ (b) σ for $p = 2$ (c) (β, ρ) for $p = 2$

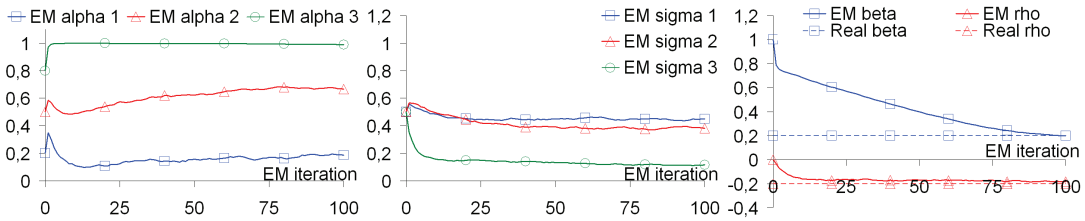


(d) α for $p = 3$ (e) σ for $p = 3$ (f) (β, ρ) for $p = 3$

Figure 5.7: Calibration of setting 1 for $p \in \{2, 3\}$

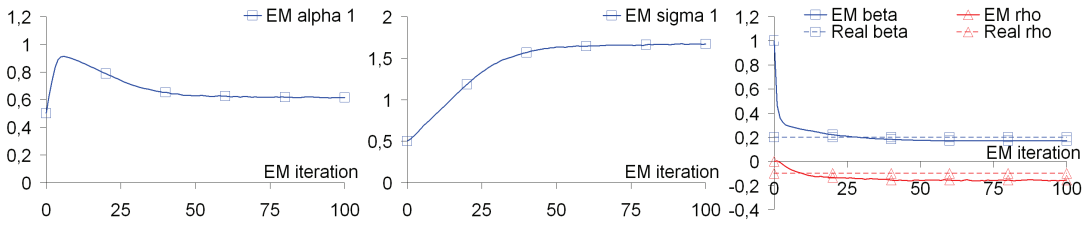


(a) α for $p = 1$ (b) σ for $p = 1$ (c) (β, ρ) for $p = 1$

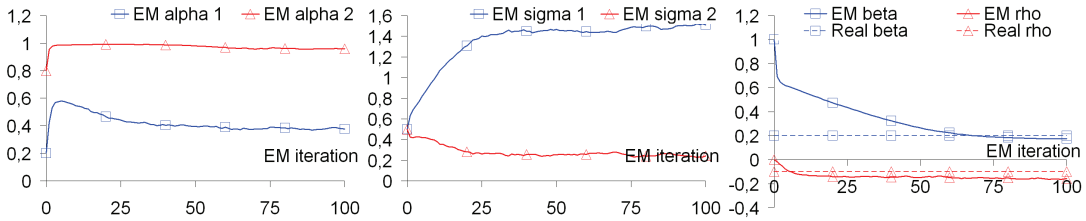


(d) α for $p = 3$ (e) σ for $p = 3$ (f) (β, ρ) for $p = 3$

Figure 5.8: Calibration of setting 2 for $p \in \{1, 3\}$

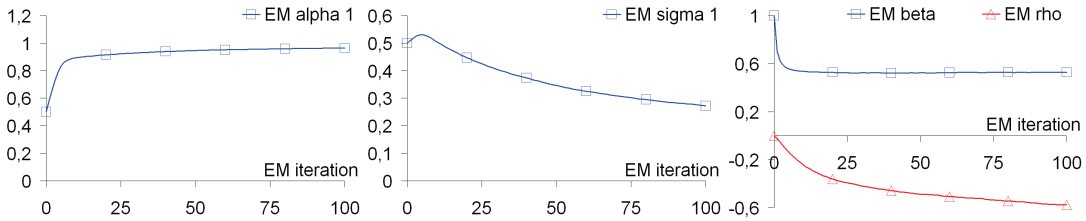


(a) α for $p = 1$ (b) σ for $p = 1$ (c) (β, ρ) for $p = 1$

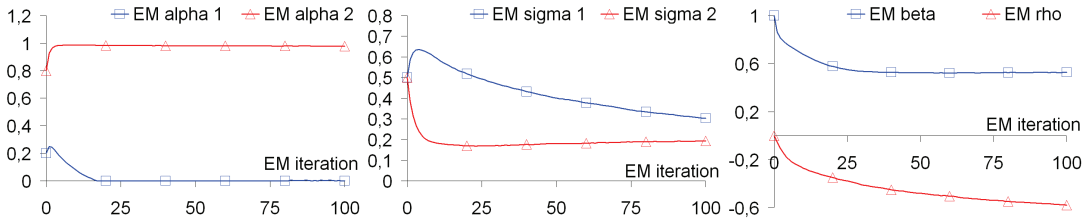


(d) α for $p = 2$ (e) σ for $p = 2$ (f) (β, ρ) for $p = 2$

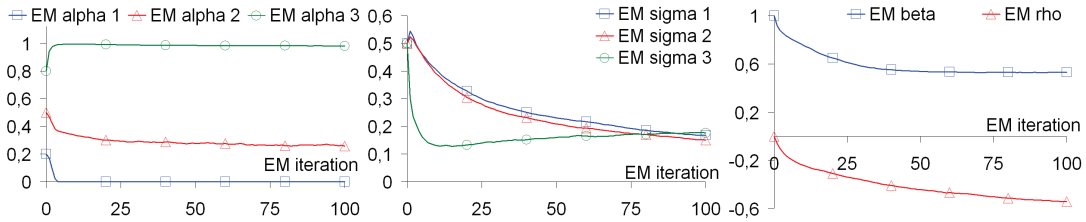
Figure 5.9: Calibration of setting 3 for $p \in \{1, 2\}$



(a) α for $p = 1$ (b) σ for $p = 1$ (c) (β, ρ) for $p = 1$

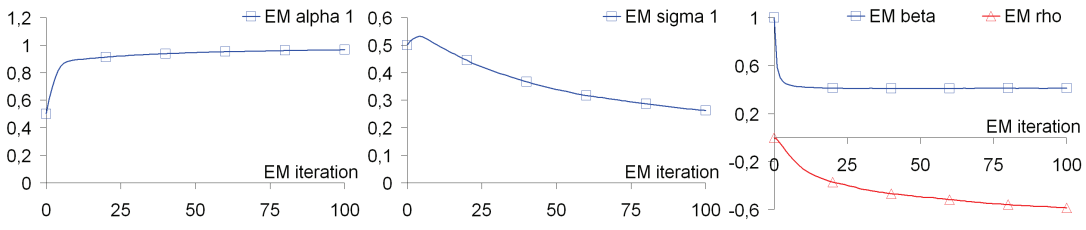


(d) α for $p = 2$ (e) σ for $p = 2$ (f) (β, ρ) for $p = 2$



(g) α for $p = 3$ (h) σ for $p = 3$ (i) (β, ρ) for $p = 3$

Figure 5.10: Calibration of CAC 40 for $p \in \{1, 2, 3\}$



(a) α for $p = 1$ (b) σ for $p = 1$ (c) (β, ρ) for $p = 1$

(d) α for $p = 2$ (e) σ for $p = 2$ (f) (β, ρ) for $p = 2$

(g) α for $p = 3$ (h) σ for $p = 3$ (i) (β, ρ) for $p = 3$

Figure 5.11: Calibration of Dow Jones for $p \in \{1, 2, 3\}$

Chapitre 6

Amélioration de l'approximation particulaire de la distribution jointe de lissage avec estimation de la variance simultanée

Sommaire

Abstract: Particle smoothers are widely used algorithms allowing to approximate the smoothing distribution in hidden Markov models. Existing algorithms often suffer from slow computational time or degeneracy. We propose in this paper a way to improve any of them with a linear complexity in the number of particles. When iteratively applied to the degenerated Filter-Smoother, this method leads to an algorithm which turns out to outperform existing linear particle smoothers for a fixed computational time. Moreover, the associated approximation satisfies a central limit theorem with a close-to-optimal asymptotic variance, which can be easily estimated by only one run of the algorithm. This last feature is totally new in the smoothing literature.

Keywords: Degeneracy, Hidden Markov model, Particle smoothing, Sequential Monte-Carlo, Variance estimation

6.1 Introduction

A *hidden Markov model* (HMM) is a doubly stochastic process where a Markov chain $\{X_t\}_{t=0}^{\infty}$ is only partially observed through a sequence of observations $\{Y_t\}_{t=0}^{\infty}$. More precisely, let $\mathbb{X}$ and $\mathbb{Y}$ be two spaces equipped with countably generated σ-fields $\mathcal{X}$ and $\mathcal{Y}$, respectively, and denote by M a Markovian transition kernel on $(\mathbb{X},\mathcal{X})$ and by G a transition kernel from $(\mathbb{X},\mathcal{X})$ to $(\mathbb{Y},\mathcal{Y})$. In our setting, the dynamics of the bivariate process $\{(X_k,Y_k)\}_{k=0}^{\infty}$ follows the Markovian transition kernel

$$P[(x,y),\mathsf{A}] \stackrel{\text{def}}{=} M\otimes G[(x,y),\mathsf{A}] = \iint M(x,\mathrm{d}x')\,G(x',\mathrm{d}y')\mathbf{1}_{\mathsf{A}}(x',y')\,, \tag{6.1}$$

where $(x,y)\in\mathbb{X}\times\mathbb{Y}$ and $\mathsf{A}\in\mathcal{X}\otimes\mathcal{Y}$.

We are interested here in estimating the expectation of a function of $(X_0,\ldots,X_T)$ conditionally on the observations $Y_0,\ldots,Y_T$ using particle smoothing algorithms. Many different implementations of the particle filters and smoothers have been proposed in the literature with different computational costs; see for example [Del Moral, 2004, Cappé et al., 2005, Doucet and Johansen, 2009]. So far, the existing particle smoothers rely on the so-called *Forward-Filter* whose complexity is linear in the number of particles N. In its simplest extension, storing the paths of the Forward-Filter allows to approximate the joint smoothing distribution as seen by [Kitagawa, 1996]. This method known as the *Filter-Smoother* unfortunately suffers from a poor representation of the states corresponding to times $t \ll T$. To circumvent this drawback, the *FFBS (Forward Filtering Backward Smoothing)* algorithm introduced by [Doucet et al., 2000] adds a backward pass to the forward filter at the cost of a quadratic complexity when used for approximating the marginal smoothing distributions. However, [Godsill et al., 2004] extended it to the *FFBSi (Forward Filtering Backward Simulation)*, an algorithm which can be implemented with a $\mathcal{O}(N)$ computational cost per time step as proposed by [Douc et al., 2010] when approximating the whole joint smoothing distribution. If we are interested only in approximations of the marginal smoothing distributions, the *Two-Filter smoother* of [Briers et al., 2010] may also be used as an alternative method. This algorithm originally suffers from a quadratic computational cost but has recently been modified in [Fearnhead et al., 2010] to get a linear one.

Whereas more and more SMC-based smoothing algorithms are linear in the number of particles, there is a recent surge of interest in mixed strategies (see [Andrieu et al., 2010, Olsson and Rydén, 2010] or [Chopin et al., 2011]) where nice properties of SMC and MCMC algorithms are conjugated to produce better approximations. Whereas these methods are developed mostly in the framework of Bayesian inference for state space models, we focus here on the quality of the approximation of the smoothing distribution associated to a fixed Hidden Markov model. This is a crucial problem to address and the hope is to exhibit the key factors that affects the quality of the estimation. More precisely, fix (once and for all) a set of observations $Y_0,\ldots,Y_T$ and try to approximate the law of $X_0,\ldots,X_T$ conditionally on the observations with a set of particles $(\xi_0^{i,N},\ldots,\xi_T^{i,N})_{i=1}^N$ associated to equal or unequal weights $(\omega_T^{i,N})_{i=1}^N$. For a fixed CPU time, how to build the best population of particles? Should we use mixed strategies? Can we obtain confidence intervals without additional Monte Carlo passes? These are some of the questions we consider in this work. Since T is fixed, the context of this work does not exactly correspond to the one of [Gilks and Berzuini, 2001] who propose to sequentially alternate SMC stages and MCMC stages as more and more observations are available. Nevertheless, the MCMC step called the Move stage by these authors is now included in the method proposed in this paper to form an efficient algorithm where some directional update of the components extends sequentially the diversity of the population from high values of t to lower values of t. In the filtering context, [Gilks and Berzuini, 2001] proposed to use these MCMCs to increase the diversity of the particles at time T. This is not ideal: the ratio between the diversity improvement and the additional computational cost may be low since the population at time T is usually already sufficiently spread. On the other hand, in the smoothing context, the population for small values of T might be substantially diversified by only a few MCMC steps. Despite its simplicity, the resulting algorithm turns out to be more than a strong competitor to existing smoothing samplers.

We propose here to improve any consistent particle approximation of the joint smoothing distribution by moving sequentially the particles according to a Metropolis-within-Gibbs iteration. Such algorithm has a linear computational cost and can be applied in particular to the Filter-Smoother to reduce the degeneracy without increasing the complexity. The paper is organized as follows: in Section 6.2, we describe the algorithm. In Section 6.3, we show that the limiting variance of the algorithm is reduced in comparison

with the original SMC-based population with a multinomial resampling stage. One major characteristic of this algorithm is the fact that, by letting the number of iterations of the Markov chains proportional to $\ln N$, the asymptotic variance is close to optimal and can be estimated using the evolution of only one population of particle paths. Up to our knowledge, this feature is totally new in the smoothing literature. Numerical experiments and comparisons with existing linear smoothers are provided in Section 6.4 for the Linear Gaussian Model (LGM) and the Stochastic Volatility Model (StoVolM).

6.2 MH-Improvement of a particle path population

We assume that there exist nonnegative σ-finite measures λ on $(\mathbb{X}, \mathcal{X})$ and μ on $(\mathbb{Y}, \mathcal{Y})$ such that for any $x \in \mathbb{X}$, $M(x,\cdot)$ and $G(x,\cdot)$ are dominated by λ and μ, respectively. This implies the existence of kernel densities

$$m(x,x') \stackrel{\text{def}}{=} \frac{\mathrm{d}M(x,\cdot)}{\mathrm{d}\lambda}(x') \text{ and } g(x,y) \stackrel{\text{def}}{=} \frac{\mathrm{d}G(x,\cdot)}{\mathrm{d}\mu}(y) .$$

In what follows, we simply write $\mathrm{d}x$ for $\lambda(\mathrm{d}x)$.

Denote for $u \leq s$, $a_{u:s} = (a_u, a_{u+1}, \ldots, a_s)$ and define the smoothing distribution $\Pi_{0:T|T}$ associated to a fixed set of observations $Y_{0:T} = y_{0:T}$ by: for any $\mathsf{A} \in \mathcal{X}^{\otimes(T+1)}$,

$$\Pi_{0:T|T}(\mathsf{A}) \stackrel{\text{def}}{=} \frac{\int \cdots \int \chi(\mathrm{d}x_0) g(x_0,y_0) \ell_{1:T}(x_{0:T}) \mathbf{1}_{\mathsf{A}}(x_{0:T}) \mathrm{d}x_{1:T}}{\int \cdots \int \chi(\mathrm{d}x_0) g(x_0,y_0) \ell_{1:T}(x_{0:T}) \mathrm{d}x_{1:T}} ,$$

where χ is a probability measure on $(\mathbb{X}, \mathcal{X})$ and $\ell_{1:T}(x_{0:T}) \stackrel{\text{def}}{=} \prod_{i=1}^{T} m(x_{i-1},x_i) g(x_i,y_i)$. The distribution $\Pi_{0:T|T}$ is thus the law of $X_{0:T}$ conditionally to $Y_{0:T} = y_{0:T}$ when X_0 follows the distribution χ. In the sequel, χ is assumed to have a density w.r.t. $\lambda(\mathrm{d}x)$, density which will be denoted by χ by abuse of notation: $\chi(\mathrm{d}x) = \chi(x)\lambda(\mathrm{d}x)$. Then, the density $\pi_{0:T|T}$ of the distribution $\Pi_{0:T|T}$ with respect to $\prod_{t=0}^{T} \lambda(\mathrm{d}x_t)$ writes

$$\pi_{0:T|T}(x_{0:T}) \propto \chi(x_0) g(x_0,y_0) \left[\prod_{i=1}^{T} m(x_{i-1},x_i) g(x_i,y_i) \right] . \tag{6.2}$$

As noted in [Gilks and Berzuini, 2001], the smoothing density $\pi_{0:T|T}$ in (6.2) is known up to a normalizing constant so that approximation of this distribution can be perfectly cast into the general framework of the Metropolis-Hastings algorithm. Given that the resulting Markov chain evolves in the path space $\mathbb{X}^{T+1}$, the candidate at each iteration should be carefully chosen to keep the acceptance rate away from zero which is a delicate task in high dimensional spaces. Considering this, an appealing approach in the MCMC literature is the Gibbs sampler and more generally the Metropolis-within-Gibbs sampler which proposes to update only one component at a time. One could also choose to update components by blocks but as will be seen in Section 6.4, moving only one component at a time is sufficient for our purpose. A key point for exploring the posterior distribution within a reasonable number of iterations is that the algorithm should be well initialized at least for the first components to be updated. We propose here to achieve this by exploiting approximation of $\Pi_{0:T|T}$ provided by SMC-based algorithms.

More precisely, suppose that we already have an approximation of $\Pi_{0:T|T}$ through a set of (normalized) weighted particle paths, $(\xi_{0:T}^{i,N}, \omega_{0:T}^{i,N})_{i=1}^{N}$ in the sense that

$$\Pi_{0:T|T}(h) \approx \sum_{i=1}^{N} \omega_{0:T}^{i,N} h(\xi_{0:T}^{i,N}) , \quad \sum_{i=1}^{N} \omega_{0:T}^{i,N} = 1 , \tag{6.3}$$

We intend here to improve this approximation by running N independent Metropolis-within-Gibbs Markov chains $(\xi_{0:T}^{i,N}[k], k \geq 0)$ for $i \in \{1,\ldots,N\}$ starting from each path $\xi_{0:T}^{i,N}$, that is, we set $\xi_{0:T}^{i,N}[0] = \xi_{0:T}^{i,N}$ for $i \in \{1,\ldots,N\}$. The resulting approximation after K iterations of the Markov chains then writes

$$\Pi_{0:T|T}(h) \approx \sum_{i=1}^{N} \omega_{0:T}^{i,N} h(\xi_{0:T}^{i,N}[K]) . \tag{6.4}$$

Let us now detail the transition of $(\xi_{0:T}^{i,N}[k],\ k \geq 0)$. For a simpler exposition, we drop here the dependence on i,N. Now, consider a family of transition kernel densities $(r_t)_{0\leq t\leq T}$ such that r_0, r_T are transition kernel densities on $(\mathbb{X}, \mathcal{X})$ whereas for $t \in \{1,\ldots,T-1\}$, r_t is a transition kernel density on $(\mathbb{X}\times\mathbb{X}, \mathcal{X})$. For $u,v,w,x \in \mathbb{X}$, set

$$\alpha_0(v,w;x) \stackrel{\text{def}}{=} \frac{\chi(x)g(x,y_0)m(x,w)}{\chi(v)g(v,y_0)m(v,w)}\frac{r_0(w;v)}{r_0(w;x)} \wedge 1\ , \tag{6.5}$$

$$\alpha_t(u,v,w;x) \stackrel{\text{def}}{=} \frac{m(u,x)g(x,y_t)m(x,w)}{m(u,v)g(v,y_t)m(v,w)}\frac{r_t(u,w;v)}{r_t(u,w;x)} \wedge 1\ ,$$
$$1 \leq t \leq T-1\ , \tag{6.6}$$

$$\alpha_T(u,v;x) \stackrel{\text{def}}{=} \frac{m(u,x)g(x,y_T)}{m(u,v)g(v,y_t)}\frac{r_T(u;v)}{r_T(u;x)} \wedge 1\ . \tag{6.7}$$

At time k, the new path $\xi_{0:T}[k]$ is obtained by updating backward in time each component $\xi_t[k]$ as follows

(i) Sample a candidate $X \sim r_t(\xi_{t-1}[k-1], \xi_{t+1}[k], \cdot)$,

(ii) Accept $\xi_t[k] = X$ with probability $\alpha_t(\xi_{t-1:t}[k-1], \xi_{t+1}[k]; X)$,

(iii) Otherwise, set $\xi_t[k] = \xi_t[k-1]$.

This procedure is valid for $t \in \{1,\ldots,T-1\}$; we skip the description of the updates for $\xi_0[k]$ and $\xi_T[k]$ since they follow the same lines under very slight modifications. The complete pseudo-code version of the Metropolis-Hastings Improved Particle Smoother (MH-IPS) is given below.

Algorithm 12 MH-IPS

Initialization
Run an SMC-algorithm targeting $\Pi_{0:T|T}$ and store $(\xi_{0:T}^{i,N}, \omega_{0:T}^{i,N})_{i=1}^N$.
Set: $\forall\ 1 \leq i \leq N,\ \xi_{0:T}^{i,N}[0] = \xi_{0:T}^{i,N}$.
K improvement passes
for k from 1 to K **do**
  for i from 1 to N **do**
    Sample $X \sim r_T(\xi_{T-1}^{i,N}[k-1]; \cdot)$,
    Accept $\xi_T^{i,N}[k] = X$ with probability $\alpha_T(\xi_{T-1:T}^{i,N}[k-1], X)$,
    Otherwise, set $\xi_T^{i,N}[k] = \xi_T^{i,N}[k-1]$.
    for t from $T-1$ down to 1 **do**
      Sample $X \sim r_t(\xi_{t-1}^{i,N}[k-1], \xi_{t+1}^{i,N}[k]; \cdot)$,
      Accept $\xi_t^{i,N}[k] = X$ with probability $\alpha_t(\xi_{t-1:t}^{i,N}[k-1], \xi_{t+1}^{i,N}[k], X)$,
      Otherwise, set $\xi_t^{i,N}[k] = \xi_t^{i,N}[k-1]$.
    end for
    Sample $X \sim r_0(\xi_1^{i,N}[k]; \cdot)$,
    Accept $\xi_0^{i,N}[k] = X$ with probability $\alpha_0(\xi_0^{i,N}[k-1], \xi_1^{i,N}[k], X)$,
    Otherwise, set $\xi_0^{i,N}[k] = \xi_0^{i,N}[k-1]$.
  end for
end for

Straightforwardly, for any $t \in \{0,\ldots,T\}$, α_t is the classical Metropolis-Hastings acceptance rate associated to the proposal kernel r_t and the target distribution $\Pi_{0:T|T}$. Due to the specific structure of $\Pi_{0:T|T}$ whose density is a product of quantities involving consecutive components, the acceptance ratios in (6.5), (6.6) and (6.7) do not depend on the path space dimension and are therefore nondegenerated. Of course, it is also possible to update each component from an arbitrary number of neighbors. Nevertheless, in the Gibbs Sampler for which all the acceptance rates are equal to one, the t-th component is updated according to the distribution of X_t conditionally on $X_{0:t-1}, X_{t+1:T}, Y_{0:T}$ which only depends on X_{t-1}, X_{t+1}, Y_t. Such dependence suggests that the candidate in the Metropolis-within-Gibbs algorithm should be proposed according to a distribution which only involves its nearest neighbors.

MH-IPS is based on a first approximation of $\Pi_{0:T|T}$ given in (6.3) whereas some SMC algorithms like the Filter-Smoother are known to suffer from a poor representation of the states close to 0 but are accurate for states close to T. As a consequence, $(\xi_t^{i,N})_{i=1}^N$ for large values of t are well-distributed and this set of particles is then propagated to the poorer ones by updating the components backward in time. In other words, instead of a random-scan procedure where components are updated at random, this determistic-scan Metropolis-Hastings algorithm extends the diversity of the particle paths to the lower values of t at each backward pass. The fact that MH-IPS uses the SMC-based approximation just once and then, keep the N Metropolis-within-Gibbs Markov chains independent from each other implies that the path degeneracy vanishes as the number of iterations increases. Strong empirical evidences of this phenomenon are provided in Section 6.4.

A last but striking particularity of MH-IPS when compared to classical MH algorithms is the fact that the approximation (6.4) only involves the states at iteration K of the N Markov chains instead of using all the history of these Markov chains. Indeed, since only one component is updated at a time, the consecutive paths are highly positively correlated so that including them into (6.4) is detrimental to the quality of the approximation. Another advantage of considering only states at iteration K is that the CLT of the approximation (6.4) which is quite easy to establish when $K \propto \ln N$ includes a very simple and close-to-optimal expression of the asymptotic variance. The estimation of this variance can be performed using the evolution of only one population of sample paths. Therefore, on the contrary to all the smoothing algorithms proposed in the literature so far, confidence intervals can be obtained without additional Monte Carlo passes.

6.3 Properties of the algorithm

In this section, since the number of observations is fixed, T is dropped for simplicity from the notation. For example, we set $\Pi = \Pi_{0:T|T}$, $\xi^{i,N} = \xi^{i,N}_{0:T|T}$, $\omega^{i,N} = \omega^{i,N}_{0:T|T}$ and so on.

The general procedure induced by MH-IPS can be described as follows. Let Q be a Markov transition kernel on $(\mathbb{X}^{T+1}, \mathcal{X}^{\otimes(T+1)})$ with invariant distribution Π. Consider a set of normalized weighted particles $(\xi^{i,N}, \omega^{i,N})_{i=1}^N$ and move the particles *independently* according to the kernel Q. To be specific, define N *independent* Markov chains $(\xi^{i,N}[k], k \geq 0)_{i=1}^N$ such that:

$$\xi^{i,N}[0] = \xi^{i,N} , \tag{6.8}$$

$$\xi^{i,N}[k+1] \sim Q(\xi^{i,N}[k], \cdot) , \quad k \geq 0 . \tag{6.9}$$

According to (6.4), Πh is approximated after k iterations of the Markov chains by:

$$\Pi h \approx \sum_{i=1}^N \omega^{i,N} h(\xi^{i,N}[k]), \quad \sum_{i=1}^N \omega^{i,N} = 1 . \tag{6.10}$$

6.3.1 A resampling step in the initialization

Let us first consider the impact of the weights on the quality of the approximation. A resampling step in the initialization consists in replacing the weighted particles $(\xi^{i,N}, \omega^{i,N})_{i=1}^N$ by the unweighted particles $(\tilde{\xi}^{i,N}, \frac{1}{N})_{i=1}^N$ such that some unbiasedness condition is fulfilled. Whereas many resampling strategies have been developed in the literature ([Liu and Chen, 1998], [Kitagawa, 1998], [Carpenter et al., 1999]; see also [Douc et al., 2005] for a brief review of their different properties), we only focus here on the most simple one, the multinomial resampling:

(i) $(\tilde{\xi}^{j,N})_{j=1}^N$ are independent conditionally on the weighted particles $(\xi^{i,N}, \omega^{i,N})_{i=1}^N$,

(ii) for all $i, j \in \{1, \ldots, N\}$, $\mathbb{P}\left[\tilde{\xi}^{j,N} = \xi^{i,N}\right] = \omega^{i,N}$.

A straightforward calculation yields:

$$\mathbb{V}\mathrm{ar}\left(\sum_{i=1}^N \omega^{i,N} h(\xi^{i,N})\right) \leq \mathbb{V}\mathrm{ar}\left(\sum_{i=1}^N h(\tilde{\xi}^{i,N})/N\right) ,$$

showing that at time 0, the particle system with equal weights is less efficient than the one with original weights. Despite this, the resampling stage discards particles with small weights and duplicates "informative" particles (with high weights). As in the particle filtering theory, our hope is that the resampling stage increases the number of Markov chains starting from interesting regions with respect to the target distribution.

Denote by $\|\cdot\|_{\mathrm{TV}}$ the total variation norm: $\|\mu\|_{\mathrm{TV}} \stackrel{\mathrm{def}}{=} \sup_{|f|_\infty \leq 1} |\mu(f)|$ where $|f|_\infty \stackrel{\mathrm{def}}{=} \sup_{x\in\mathbb{X}} |f(x)|$ and assume that

(**A1**) For any $x \in \mathbb{X}^{T+1}$, $\lim_{k\to\infty} \left\|Q^k(x,\cdot) - \Pi\right\|_{\mathrm{TV}} = 0$.

Under this assumption, it is direct that for any bounded measurable function h, $\sum_{i=1}^N \omega^{i,N} h(\xi^{i,N}[k])$ is asymptotically unbiased whatever the weights are, provided their sum is equal to one. To go further, consider the effect of the weights on the second order approximation. The following proposition shows that as the iterations of the Markov chains goes to infinity, the quadratic error tends to a limit which is minimal when all the weights are equal to $1/N$. This advocates for a particle system with equal weights in the initialization as provided by a resampling step before letting evolve the N Markov chains. This conclusion is not surprising since the resampled particles are intuitively closer to the target distribution Π than the weighted particles.

Proposition 6.1. Assume (**A1**). Then, for any bounded measurable function h,

$$\lim_{k\to\infty} \mathbb{E}\left[\left(\sum_{i=1}^N \omega^{i,N} h(\xi^{i,N}[k]) - \Pi h\right)^2\right] = \mathbb{V}\mathrm{ar}_\Pi(h)\ \mathbb{E}\left[\sum_{i=1}^N \left(\omega^{i,N}\right)^2\right],$$

where $\mathbb{V}\mathrm{ar}_\Pi(h) = \Pi h^2 - (\Pi h)^2$. Moreover, the previous limit is minimized when all the weights are equal: $\omega^{i,N} = 1/N$ for all $i \in \{1,\ldots,N\}$.

Proof. Proof is given the Appendix. □

As a consequence of this proposition, it is assumed in the sequel that the *multinomial resampling stage* has been performed in the initialization, i.e. (6.8), (6.9) and (6.10) are replaced by

$$\tilde{\xi}^{i,N}[0] = \tilde{\xi}^{i,N}, \tag{6.11}$$

$$\tilde{\xi}^{i,N}[k+1] \sim Q(\tilde{\xi}^{i,N}[k],\cdot), \quad k \geq 0, \tag{6.12}$$

$$\Pi h \approx \sum_{i=1}^N h(\tilde{\xi}^{i,N}[k])/N, \tag{6.13}$$

Then, according to Proposition 6.1,

$$\lim_{k\to\infty} \mathbb{E}\left[\left(\sum_{i=1}^N h(\tilde{\xi}^{i,N}[k])/N - \Pi h\right)^2\right] = \mathbb{V}\mathrm{ar}_\Pi(h)/N. \tag{6.14}$$

Thus, when N is fixed and k goes to infinity, (6.14) shows that the approximation cannot be better than having N independent draws from the distribution Π. A natural question is now to properly tune the number of iterations k of the Markov chains to the number N of initial points so that the unweighted particles $(\xi^{i,N}[k], 1/N)_{i=1}^N$ have properties close to iid draws according to Π *without* letting k go to infinity.

6.3.2 Central limit theorem

MH-IPS is based on a first approximation of Πh by a family of normalized weighted particles $(\xi^{i,N}, \omega^{i,N})_{i=1}^N$. For various versions of SMC methods, the asymptotic normality of $(\xi^{i,N}, \omega^{i,N})_{i=1}^N$ have already been obtained under different techniques (see for example [Del Moral and Guionnet, 1999], [Künsch, 2000], [Chopin, 2004] or [Douc and Moulines, 2008]). The following proposition now focus on the effect of the multinomial resampling on the central limit theorem: whatever SMC method is chosen, if $(\xi^{i,N}, \omega^{i,N})_{i=1}^N$ are asymptotically normal, then $(\tilde{\xi}^{i,N}, 1/N)_{i=1}^N$ are also asymptotically normal with $\mathbb{V}\mathrm{ar}_\Pi(h)$ as an *additional* term in the variance.

Proposition 6.2. Assume that $(\xi^{i,N}, \omega^{i,N})_{i=1}^N$ are asymptotically normal, in the sense that for any bounded measurable function h, there exists $0 < \sigma^2(h) < \infty$ such that

$$N^{1/2}\left[\sum_{i=1}^N \omega^{i,N} h(\xi^{i,N}) - \Pi h\right] \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(h)) .$$

Then, for any bounded measurable function h,

$$N^{1/2}\left[\sum_{i=1}^N h(\tilde{\xi}^{i,N})/N - \Pi h\right] \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathbb{V}\mathrm{ar}_\Pi(h) + \sigma^2(h)) .$$

The proof follows closely the lines of [Chopin, 2004, Theorem 1] or [Douc and Moulines, 2008, Theorem 4] and is omitted for the sake of brevity.

Proposition 6.2 shows the asymptotic normality of $(\tilde{\xi}^{i,N}[k], 1/N)_{i=1}^N$ for $k = 0$. The Markov chains are then run independently according to the transition kernel Q and we now consider the impact on the approximation given in (6.13) for $k = k_N$. To be specific, the following theorem shows that under the assumption that the kernel Q is V-geometrically ergodic, for $k_N \propto \ln N$, the unweighted particles $(\tilde{\xi}^{i,N}[k_N], 1/N)_{i=1}^N$ are asymptotically normal with a reduced asymptotic variance. Define the following set of assumptions:

(**A2**) There exists a measurable function $V : \mathbb{X}^{T+1} \to [1, \infty)$ such that

(i) $\Pi V < \infty$ and for any $x \in \mathbb{X}$ and any $k \in \mathbb{N}$, $Q^k V(x) < \infty$,

(ii) there exists $\beta \in (0,1)$ such that for any $h \in \mathcal{C}_V \stackrel{\text{def}}{=} \{h; |h/V|_\infty < \infty\}$ and any $x \in \mathbb{X}$,

$$|Q^k h(x) - \Pi h| \leq \beta^k V(x) ,$$

(iii) the sequence $\{N^{-1}\sum_{i=1}^N V^2(\tilde{\xi}^{i,N})\}_{N\geq 1}$ of random variables is bounded in probability.

(**A2**)-(i) ensures that the quantities appearing in (**A2**)-(ii) are well defined. (**A2**)-(ii) shows that Q is V-geometrically ergodic. (**A2**)-(iii) is a weak assumption concerning the initial unweighted particles $(\tilde{\xi}^{i,N}, \frac{1}{N})_{i=1}^N$. If for example, $(\tilde{\xi}^{i,N}, \frac{1}{N})_{i=1}^N$ is consistent with respect to the function V^2 in the sense that $\sum_{i=1}^N V^2(\tilde{\xi}^{i,N})/N$ converges in probability to ΠV^2, then (**A2**)-(iii) holds. Condition under which such convergence results hold for possibly unbounded functions may be found for example in [Douc and Moulines, 2008].

Theorem 6.1. *Assume (A2). Let $(k_N)_{N\geq 0}$ be a sequence of integers such that*

$$\lim_{N\to\infty} k_N + \ln N/(2\ln\beta) = \infty . \tag{6.15}$$

Then, for any h such that $h^2 \in \mathcal{C}_V$, the following central limit theorem holds:

$$N^{-1/2}\sum_{i=1}^N \left[h(\tilde{\xi}^{i,N}[k_N]) - \Pi h\right] \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathbb{V}\mathrm{ar}_\Pi(h)) .$$

Proof. Proof is given in the Appendix. □

Theorem 6.1 and Proposition 6.2 show that k_N iterations of the Markov chains reduce the asymptotic variance when compared to a sample obtained by multinomial resampling of a population issued from any SMC method. The asymptotic variance $\mathbb{V}\mathrm{ar}_\Pi(h)$ in Theorem 6.1 is close to optimal since it is the same as for i.i.d. draws with distribution Π. Moreover, the expression of $\sigma^2(h)$ in Proposition 6.2 is usually quite involved and for obtaining confidence intervals, the estimation of the asymptotic variance in Proposition 6.2 is classically obtained by adding some Monte Carlo passes. This is not at all the case in Theorem 6.1 since estimation of $\mathbb{V}\mathrm{ar}_\Pi(h)$ can be performed directly via $(\tilde{\xi}^{i,N}[k_N], 1/N)_{i=1}^N$. Finally, by adding typically $k_N = -\ln N/\ln\beta$ iterations of a transition kernel to a SMC-based population of particles, we obtain a sample with a reduced and close-to-optimal variance which can be easily approximated without additional simulations.

The fact that the CLT holds for $k_N \propto \ln N$ suggests that a good approximation of the target distribution may be achieved with only a few number of iterations of the parallel Markov chains. This will be confirmed empirically in the next section.

6.4 Experiments

The *Filter-smoother* is known to be quite easy to implement and efficient in terms of CPU time, but suffers dramatically from the degeneracy of the ancestors. We now see how only a few iterations of MH-IPS reduce the degeneracy and turn the *Filter-smoother* to a strong competitor to the existing smoother algorithms. In the sequel, denote by the *Metropolis-Hastings Improved Filter-Smoother* (MH-IFS), Algorithm 12 initialized with the Filter-Smoother. The performance of this algorithm is now compared to the other linear-in-N particle smoothers (Filter-Smoother, FFBSi, Two-Filter). In order to be as computationally fair as possible, all these algorithms are implemented in the same way as their common base, the Forward-Filter.

6.4.1 Linear Gaussian Model

We first consider the LGM defined by:

$$X_{t+1} = \phi X_t + \sigma_u U_t \ , \quad Y_t = X_t + \sigma_v V_t \ ,$$

where $X_0 \sim \mathcal{N}\left(0, \frac{\sigma_u^2}{1-\phi^2}\right)$, $\{U_t\}_{t\geq 1}$ and $\{V_t\}_{t\geq 1}$ are independent sequences of i.i.d. standard gaussian random variables (independent of X_1). $T+1 = 101$ observations were generated using the model with $\phi = 0.9$, $\sigma_u = 0.6$ and $\sigma_v = 1$. Furthermore, in this model, the fully-adapted filters are explicitly computable when needed and the Gibbs sampler may be implemented.

The diversity of the particle population at each time step for each algorithm is measured by an estimate of the *effective sample size* $N_{\text{eff}}^{\text{algo}}(t)$ as defined in [Fearnhead et al., 2010]. Motivated by the fact that $1/N = \mathbb{E}\left[\left(\bar{X}_N - \mu\right)^2 / \sigma^2\right]$, when $X^{(1)}, \ldots, X^{(N)}$ are i.i.d. with $\mathbb{E}[X^{(1)}] = \mu$, $\mathbb{V}\text{ar}(X^{(1)}) = \sigma^2$ and $\bar{X}_N$ is their sample mean, we set

$$N_{\text{eff}}^{\text{algo}}(t) \stackrel{\text{def}}{=} \mathbb{E}\left[\left(\frac{\pi_{t|T}^{\text{algo},N}(\text{Id}) - \mu_t}{\sigma_t}\right)^2\right]^{-1} , \tag{6.16}$$

where Id is the identity function on $\mathbb{R}$, μ_t and σ_t^2 are the exact mean and variance of X_t conditionally to $Y_{0:T}$ obtained from the Kalman smoother. In some sense, the weighted sample produced by a given algorithm is as accurate at estimating X_t as an "independent" sample of size $N_{\text{eff}}^{\text{algo}}(t)$. The expression of $N_{\text{eff}}^{\text{algo}}(t)$ given in (6.16) shows that it is inversely proportional to the quadratic error associated to a normalized estimator of $\mathbb{E}(X_t|Y_{0:T})$. To estimate the expectation in (6.16) we use the mean value from 250 repetitions of each algorithm with a number of particles chosen such that the computation time of each of them is the same.

Figure 6.1.a shows that when the number of improvements increases, the degeneracy of the particle population for small values of t decreases and for $K = 8$ all the time steps have the same diversity.

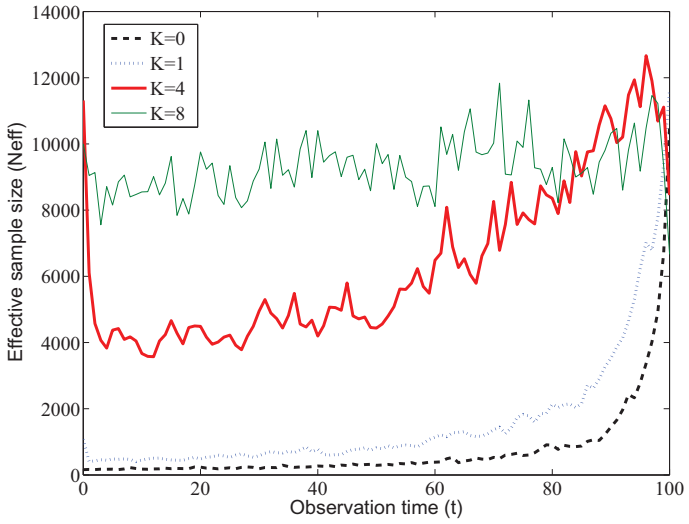
Figure 6.1.b displays the effective sample size of the four linear smoothing algorithms. As expected, the Filter-Smoother is highly degenerated for small values of t as opposed to the other algorithms. Furthermore, the MH-IFS clearly outperforms all others within a fixed computational time. In order to check that this efficiency is not due to the fact that the LGM allows to easily implement the Gibbs sampler, we now turn to a model where a rejection sampling is required.

6.4.2 Stochastic Volatility Model

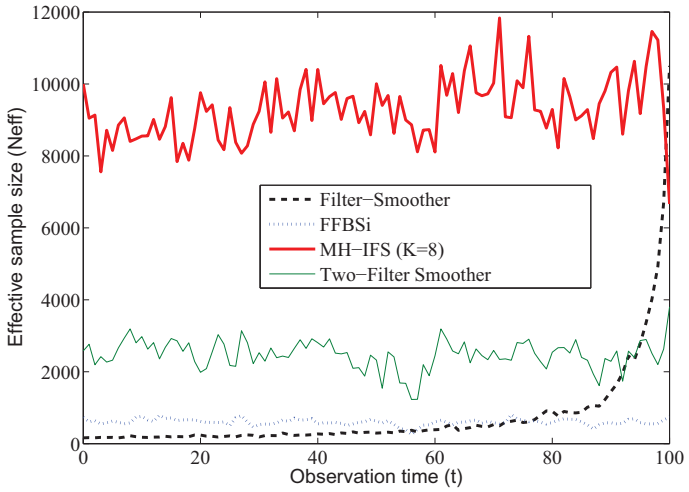
StoVolM have been introduced in financial time series modeling to capture more realistic features than the first ARCH/GARCH models ([Hull and White, 1987]). Despite its apparent simplicity, the following equations do not allow to directly simulate according to $r_t(u, w; \cdot) \propto m(u, \cdot) g(\cdot, y_t) m(\cdot, w)$:

$$X_{t+1} = \alpha X_t + \sigma U_{t+1} \ , \quad Y_t = \beta e^{\frac{X_t}{2}} V_t \ ,$$

where $X_0 \sim \mathcal{N}\left(0, \frac{\sigma^2}{1-\alpha^2}\right)$, U_t and V_t are independent standard gaussian random variables. $T+1 = 101$ observations were generated using the model with $\alpha = 0.3$, $\sigma = 0.5$ and $\beta = 1$ in order to estimate the



(a) Influence of the number of improvements K



(b) Comparison of four linear smoothing algorithms

Figure 6.1: Average effective sample size for each of the 100 time steps of the LGM using different smoothing algorithms for a fixed CPU time.

effective sample size defined in (6.16). The true values of μ_t and σ_t cannot be computed explicitly so they are estimated by running the MH-IFS with $N = 650000$.

Gibbs sampler

In the StoVolM, the Gibbs sampler requires to sample exactly from

$$r_t(u,w;x) \propto \exp\left\{-\frac{e^{-x}}{2\beta^2}y_t^2 - \frac{1+\alpha^2}{2\sigma^2}\left[x - \left(\frac{\alpha}{1+\alpha^2}(u+w) - \frac{\sigma^2/2}{1+\alpha^2}\right)\right]^2\right\}, \tag{6.17}$$

for $1 \leq t \leq T-1$ (the cases $t = 0$ and $t = T$ are dealt with in a similar way) which does not correspond to a classical distribution. However, we propose here to implement a rejection sampling. The first idea is to sam-

ple the proposal candidate $X = x$ according to the *a priori* distribution of X_t conditionally to $X_{t-1} = u$ and $X_{t+1} = w$. The corresponding ratio of acceptance is then given by $(|y_t|/\beta)\exp\left\{-(x-1)/2 - e^{-x}y_t^2/(2\beta^2)\right\}$ and will obviously lead to poor results for small values of y_t. To counterbalance the effect of y_t in the acceptance rate, the proposal distribution should also take the value of y_t into account; we then rewrite (6.17) for any $\gamma_t \geq 0$ (possibly depending on y_t):

$$r_t(u,w;x) \propto \exp\left\{-\frac{\gamma_t}{2}x - \frac{e^{-x}}{2\beta^2}y_t^2 - \frac{1+\alpha^2}{2\sigma^2} \times \left[x - \left(\frac{\alpha}{1+\alpha^2}(u+w) - \frac{\sigma^2/2}{1+\alpha^2}(1-\gamma_t)\right)\right]^2\right\}, \qquad (6.18)$$

which suggests to propose x according to

$$\mathcal{N}\left(\frac{\alpha}{1+\alpha^2}(u+w) - \frac{\sigma^2/2}{1+\alpha^2}(1-\gamma_t), \frac{\sigma^2}{1+\alpha^2}\right)$$

and to accept it with a probability given by:

$$\left(\frac{|y_t|}{\gamma_t^{1/2}\beta}\right)^{\gamma_t} \exp\left\{-\frac{\gamma_t}{2}(x-1) - \frac{e^{-x}}{2\beta^2}y_t^2\right\}. \qquad (6.19)$$

An optimal choice for γ_t would consist in maximizing the smoothed expectation of (6.19) but this quantity is intractable. An intuitive choice for γ_t is then:

$$\gamma_t = \begin{cases} (|y_t|/\beta)^2, & \text{if } |y_t| \leq \beta, \\ |y_t|/\beta, & \text{if } |y_t| > \beta. \end{cases} \qquad (6.20)$$

Indeed, for small values of y_t, (6.19) is then close to one and for bigger values, the exponential becomes very small but the first term remains non-neglectable.

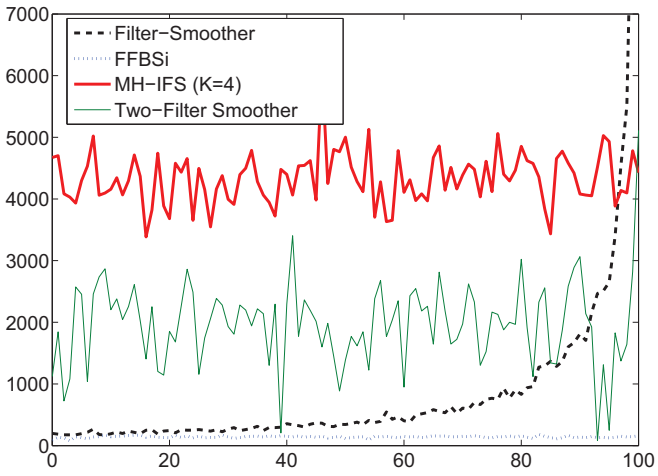


Figure 6.2: Average effective sample size for each of the 100 time steps of the StoVolM using different smoothing algorithms for a fixed CPU time.

The Improved Filter-Smoother used to generate Figure 6.2 performs simulations using the Gibbs sampler with the previous rejection sampling. We can see that this algorithm still leads to better results than the other ones within an equivalent computational time.

In many instances (for example Expectation-Maximization algorithm, score computation), it is necessary to estimate smoothed additive functionals such as $\Pi_{0:T|T}(H)$ where for all $x_{0:T} \in \mathbb{X}^{T+1}$, $H(x_{0:T}) =$

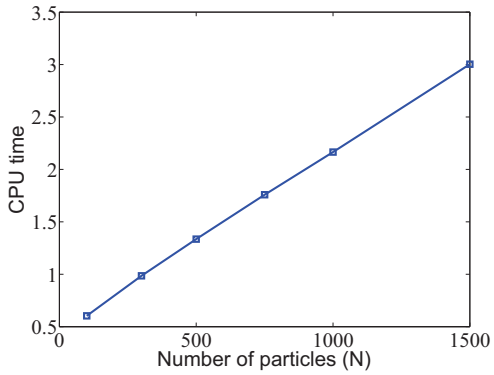


Figure 6.3: Average CPU time for computing a smoothed additive functional with the MH-IFS as a function of the number of particles.

$\sum_{t=0}^{T} x_t$. In order to assess the smoothing algorithms on this matter, $T+1 = 1001$ observations were generated. As seen before, the computational cost of the MH-IFS is linear in N which is verified by numerical experiments in Figure 6.3.

Figure 6.4.a shows that the variance vanishes quickly with the number of improvement passes and only 4 iterations of the Markov chains are sufficient to get an efficient estimator. Then, the variances displayed in Figure 6.4.b allow again to draw the conclusion that for a fixed CPU time, the MH-IFS is more efficient than the Two-Filter. Finally, one improvement pass has been applied to the particle paths given by the FFBSi. The variance reduction is again significant as shown in Figure 6.4.c.

Metropolis-within-Gibbs and confidence interval

In order to assess Algorithm 12 in the case where the Gibbs sampler could not be implemented, we now turn to the Metropolis-within-Gibbs sampler which is implemented by using again the proposal distribution:

$$r_t(u,w;\cdot) \sim \mathcal{N}\left(\frac{\alpha}{1+\alpha^2}(u+w) - \frac{\sigma^2/2}{1+\alpha^2}(1-\gamma_t), \frac{\sigma^2}{1+\alpha^2}\right) ,$$

where γ_t is defined in (6.20), and the associated acceptance rate is now given by:

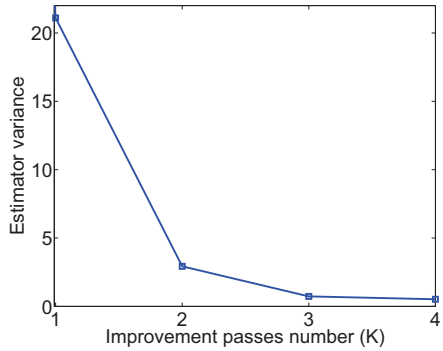
$$\alpha_t(u,v,w;x) = \exp\left\{-\frac{\gamma_t}{2}(x-v) - \frac{e^{-x}-e^{-v}}{2\beta^2}y_t^2\right\} \wedge 1 .$$

Figure 6.5 compares the empirical variance of the Gibbs and Metropolis-within-Gibbs samplers of the smoothed additive functional conditionally to the $T+1 = 1001$ observations used previously. The efficiency of both algorithms is equivalent, showing that Algorithm 12 remains a great performer even when exact *a posteriori* simulation is not possible.

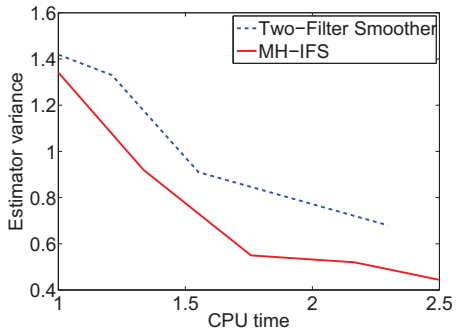
Finally, Theorem 6.1 is assessed in Figure 6.6. The empirical variance of the estimator given by Algorithm 12 run with $K_N \propto \ln N$ has been computed over 250 runs using the Gibbs and the Metropolis-within-Gibbs samplers for different number of particles N and compared to the asymptotic variance $\mathbb{V}\mathrm{ar}_\Pi(h)/N$ estimated through only one population of particles. The results show that it is possible in practice to get a confidence interval for the approximation with only one run of Algorithm 12 of complexity $O(N \ln N)$.

Influence of α

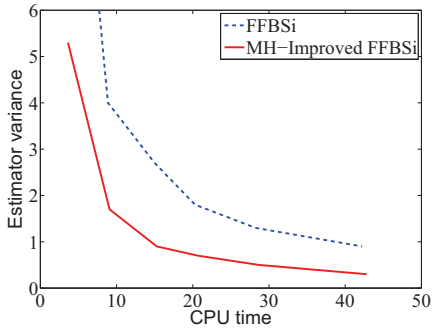
StoVolM becomes more and more challenging when the value of α gets close to 1 as the number of rejected candidates in the Metropolis-within-Gibbs algorithm tends to increase. This is easily seen from Figure 6.7 which shows that the number of improvement passes needed to reduce the variance of the Improved Filter-Smoother increases with the value of α.



(a) Variance of the Improved Filter-Smoother according to the number of improvement passes K



(b) Variance of the Two-Filter Smoother and the Improved Filter-Smoother according to the CPU time



(c) Variance of the FFBSi and its improved version according to the CPU time

Figure 6.4: Variance of different smoothed additive functional particle estimators in the StoVolM.

As a consequence, for an extreme value of $\alpha = 0.98$, the CPU time will clearly increase. However, the Improved Filter-Smoother provides an estimator for the asymptotic variance with *only one run*, whereas for the Filter-Smoother and the Two-Filter Smoother, the estimation of the variance is provided by 150 runs of the corresponding algorithm. To be fair in the comparison, we estimate both $\pi_{t|T}(\mathrm{Id})$ and the associated variance within a fixed CPU time.

According to Figure 6.8, the variance of this algorithm is better than the one of the Filter-Smoother and comparable to the one of the Two-Filter Smoother. Consequently, the Improved Filter-Smoother remains the best choice as it is marginally as good as the Two-Filter Smoother and provides in addition an approximation of the joint smoothing distribution.

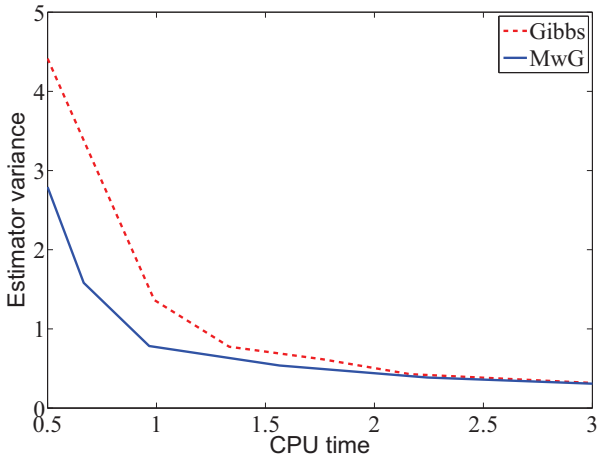


Figure 6.5: Variance of the Gibbs and Metropolis-within-Gibbs samplers according to the CPU time.

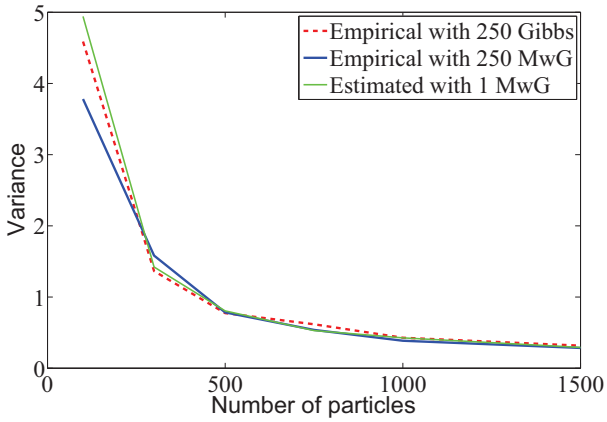


Figure 6.6: Algorithm 12 variance according to the number of observations.

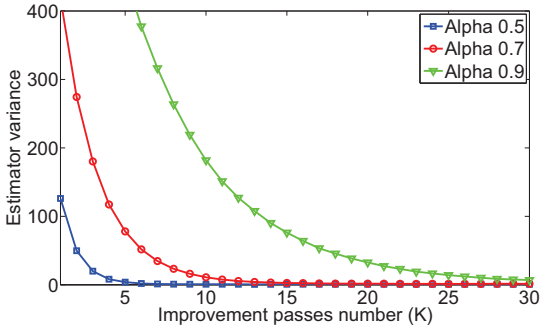


Figure 6.7: Variance reduction for different values of α.

6.5 Conclusion

At first sight, one could fear that the MH-IPS is too slow since the updates concern only one component at a time. The various comparisons performed for a *fixed* CPU time in the previous section show that this *is not* the case at all. Roughly speaking, a backward pass in the MH-IPS proposes to sequentially modify each component of the N parallel Markov chains. This can be seen as one run of N particles through $T+1$

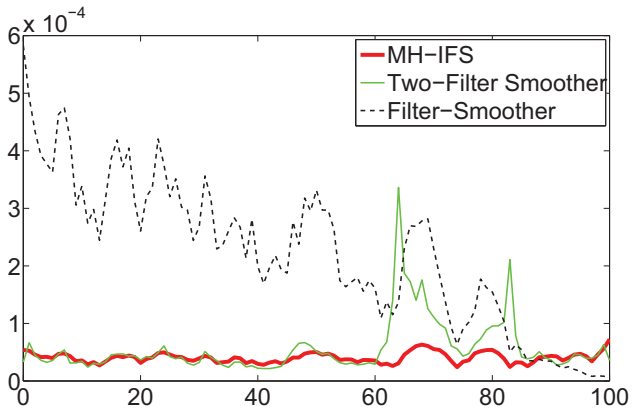


Figure 6.8: TCL variance for each of the 100 time steps of the StoVolM using different smoothing algorithms for a fixed CPU time.

observations which is, in the worse case, computationally equivalent to one pass of the bootstrap filter. By empirical evidences, we have seen that only a few backward passes ($K = 4$ or 8 in the examples) of the MH-IFS sweep out the degeneracy of the ancestors by extending backward in time the diversity of the particles.

This method is linear in N and outperforms other existing algorithms as the FFBSi or the Two-Filter within a fixed CPU time. These performance results may be explained by the fact that in the FFBSi algorithm, the points are sampled in the forward pass once and for all; the backward pass in the FFBSi only modifies the weights of the particles without moving them. On the contrary, the MH-IPS allows in the backward pass to move the particles and thus to explore interesting regions of the posterior distribution. In the Two-Filter sampler, two populations (the "forward" population and the "backward" population) evolve *independently*. At time t, a particle is sampled after choosing a couple of particles at time $t-1$ and $t+1$. The two components of these couples belong to independent populations and it is likely that even if their weights are respectively high, associating these independent particles could be detrimental to the approximation. On the contrary, in the MH-IPS, even if the Markov chains are independent, the proposed modification of the component is sampled with respect to its two neighbors which both belong to the *same* Markov chain. Note that we did not compare this algorithm to the Population Monte Carlo by Markov chains (PMCMC) samplers introduced by [Andrieu et al., 2010] since the framework here is not the Bayesian inference of parameterized Markov chains.

Another major advantage here is the fact that a CLT can be obtained with a very simple asymptotic variance which can be estimated with only one run of the Algorithm and a complexity in $O(N \ln N)$. This is totally new in comparison to all the smoothing algorithms proposed in the literature so far, where the asymptotic variances are usually particularly involved. Thus, for a fixed CPU time and only one run, this algorithm is able to produce both approximations of the smoothing distributions and confidence intervals.

Finally, we only focus here on the MH-IFS since it is efficient enough for our purpose. Of course, many other variants with different SMC-based approximations in the initialization step may be performed. In the context of the paper, the MH-IPS only uses the SMC-based approximation once before starting independent MCMC Markov chains. The empirical performances of this algorithm, namely with respect to the diversity of the population and the precision of the approximation, seem to us convincing enough to let the Markov chains evolve independently without trying to interact them again. Of course, as previously noted in [Gilks and Berzuini, 2001], in some different contexts, where for example, the observations are available sequentially whereas approximations of the smoothing distributions are needed at each time, some variants with SMC steps mixed with MCMC steps can also be elaborated. Nevertheless, in the framework of this paper, the number T of the observations is fixed and we only focus here on how the independent MCMC steps drastically improve the first approximation obtained by SMC algorithms. In this context, there is no need to interact again the Markov chains; this allows to keep the diversity of the population while approximations and confidence intervals are obtained without effort.

Appendix

6.A Proof of Proposition 6.1

For all $k \geq 0$, the bias plus variance decomposition writes

$$\mathbb{E}\left[\left(\sum_{i=1}^{N}\omega^{i,N}h(\xi^{i,N}[k]) - \Pi h\right)^2\right] = \left\{\mathbb{E}\left[\sum_{i=1}^{N}\omega^{i,N}h(\xi^{i,N}[k])\right] - \Pi h\right\}^2 + \mathbb{V}\mathrm{ar}\left(\sum_{i=1}^{N}\omega^{i,N}h(\xi^{i,N}[k])\right)$$
$$= \left\{\mathbb{E}\left[\sum_{i=1}^{N}\omega^{i,N}h(\xi^{i,N}[k])\right] - \Pi h\right\}^2 + \mathbb{V}\mathrm{ar}\left(\mathbb{E}\left[\sum_{i=1}^{N}\omega^{i,N}h(\xi^{i,N}[k])\middle|\mathcal{F}_0^N\right]\right) + \mathbb{E}\left[\mathbb{V}\mathrm{ar}\left(\sum_{i=1}^{N}\omega^{i,N}h(\xi^{i,N}[k])\middle|\mathcal{F}_0^N\right)\right], \tag{6.21}$$

where $\mathcal{F}_0^N = \sigma\left\{\xi^{i,N}, \omega^{i,N},\ i \in \{1,\dots,N\}\right\}$. Now, by definition of $\xi^{i,N}[k],\ i \in \{1,\dots,N\}$,

$$\mathbb{E}\left[\sum_{i=1}^{N}\omega^{i,N}h(\xi^{i,N}[k])\middle|\mathcal{F}_0^N\right] = \sum_{i=1}^{N}\omega^{i,N}Q^k h(\xi^{i,N}),$$

and the first term of the RHS of (6.21) is bounded by

$$\left|\mathbb{E}\left[\sum_{i=1}^{N}\omega^{i,N}h(\xi^{i,N}[k])\right] - \Pi h\right| \leq \mathbb{E}\left[\sum_{i=1}^{N}\omega^{i,N}\left|Q^k h(\xi^{i,N}) - \Pi h\right|\right].$$

The RHS goes to 0 as k tends to infinity by the Lebesgue convergence theorem since h is bounded. The same argument holds to handle the second term of the RHS of (6.21):

$$\lim_{k\to\infty}\mathbb{V}\mathrm{ar}\left(\mathbb{E}\left[\sum_{i=1}^{N}\omega^{i,N}h(\xi^{i,N}[k])\middle|\mathcal{F}_0^N\right]\right) = \lim_{k\to\infty}\mathbb{V}\mathrm{ar}\left(\sum_{i=1}^{N}\omega^{i,N}Q^k h(\xi^{i,N})\right)$$
$$= \mathbb{V}\mathrm{ar}\left(\sum_{i=1}^{N}\omega^{i,N}\Pi h\right) = \mathbb{V}\mathrm{ar}\left(\Pi h\right) = 0.$$

Finally, conditionally to $\mathcal{F}_0^N$, the random variables $(\xi^{i,N}[k])_{i=1}^N$ are independent and

$$\mathbb{V}\mathrm{ar}\left(\sum_{i=1}^{N}\omega^{i,N}h(\xi^{i,N}[k])\middle|\mathcal{F}_0^N\right) = \sum_{i=1}^{N}(\omega^{i,N})^2\mathbb{V}\mathrm{ar}\left(h(\xi^{i,N}[k])\middle|\mathcal{F}_0^N\right)$$
$$= \sum_{i=1}^{N}(\omega^{i,N})^2\left[Q^k h^2(\xi^{i,N}) - \left(Q^k h(\xi^{i,N})\right)^2\right],$$

leading to

$$\lim_{k\to\infty}\mathbb{E}\left[\mathbb{V}\mathrm{ar}\left(\sum_{i=1}^{N}\omega^{i,N}h(\xi^{i,N}[k])\middle|\mathcal{F}_0^N\right)\right] = \left[\Pi h^2 - (\Pi h)^2\right]\mathbb{E}\left[\sum_{i=1}^{N}(\omega^{i,N})^2\right].$$

This shows the first part of the proposition. Now, by the Cauchy-Schwartz inequality:

$$1 = \sum_{i=1}^{N}\omega^{i,N} \leq \left(\sum_{i=1}^{N}(\omega^{i,N})^2\right)^{1/2} N^{1/2},$$

i.e. $\sum_{i=1}^{N}(\omega^{i,N})^2 \geq 1/N$ with equality only for $\omega^{i,N} = 1/N$ for all i. The proof is completed.

6.B Proof of Theorem 6.1

Let $\gamma_N = k_N + \ln N/(2\ln\beta)$. Under the assumptions of Theorem 6.1, $\lim_{N\to\infty}\gamma_N = \infty$. Now, write

$$N^{-1/2}\sum_{i=1}^{N}\left[h(\tilde{\xi}^{i,N}[k_N]) - \Pi h\right] = N^{-1/2}\sum_{i=1}^{N}\left[Q^{k_N}h(\tilde{\xi}^{i,N}) - \Pi h\right] + N^{-1/2}\sum_{i=1}^{N}\left[h(\tilde{\xi}^{i,N}[k_N]) - Q^{k_N}h(\tilde{\xi}^{i,N})\right] . \tag{6.22}$$

Since $V \geq 1$, (**A2**)-(iii) implies that $\{N^{-1}\sum_{i=1}^{N} V(\tilde{\xi}^{i,N})\}_{N\geq 1}$ is bounded in probability. Combining this with

$$\left|N^{-1/2}\sum_{i=1}^{N}\left[Q^{k_N}h(\tilde{\xi}^{i,N}) - \Pi h\right]\right| \leq N^{-1/2}\beta^{k_N}\sum_{i=1}^{N}V(\tilde{\xi}^{i,N}) = \beta^{\gamma_N}\times N^{-1}\sum_{i=1}^{N}V(\tilde{\xi}^{i,N}) ,$$

shows that the first term of the RHS of (6.22) converges in probability to 0. Now, the second term of the RHS of (6.22) writes

$$N^{-1/2}\sum_{i=1}^{N}\left[h(\tilde{\xi}^{i,N}[k_N]) - Q^{k_N}h(\tilde{\xi}^{i,N})\right] = \sum_{i=1}^{N}\left\{U_{N,i} - \mathbb{E}\left[U_{N,i}|\mathcal{F}_{N,i-1}\right]\right\} ,$$

where

$$\begin{aligned} U_{N,i} &= N^{-1/2}h\left(\tilde{\xi}^{i,N}[k_N]\right) , \\ \mathcal{F}_{N,i} &= \sigma\left\{\tilde{\xi}^{\ell,N}, \tilde{\xi}^{j,N}[k_N], (\ell,j)\in\{1,\ldots,i\}^2\right\} . \end{aligned}$$

To apply [Douc and Moulines, 2008, Theorem A3] with $M_N = N$ and $\sigma^2 = \mathbb{V}\mathrm{ar}_\Pi(h)$, we need to check that

$$\sum_{i=1}^{N}\mathbb{V}\mathrm{ar}\left(U_{N,i}|\mathcal{F}_{N,i-1}\right) \xrightarrow{\mathbb{P}} \sigma^2 , \tag{6.23}$$

$$\sum_{i=1}^{N}\mathbb{E}\left[U_{N,i}^2\mathbf{1}_{\{|U_{N,i}|\geq\varepsilon\}}\middle|\mathcal{F}_{N,i-1}\right] \xrightarrow{\mathbb{P}} 0 , \quad \text{for any } \varepsilon > 0 . \tag{6.24}$$

We start with (6.23). Write

$$\left|\sum_{i=1}^{N}\mathbb{V}\mathrm{ar}\left(U_{N,i}|\mathcal{F}_{N,i-1}\right) - \sigma^2\right| \leq N^{-1}\sum_{j=1}^{N}\left|Q^{k_N}h^2(\tilde{\xi}^{j,N}) - \Pi h^2\right| + N^{-1}\sum_{j=1}^{N}\left|\left[Q^{k_N}h(\tilde{\xi}^{j,N})\right]^2 - (\Pi h)^2\right| . \tag{6.25}$$

As $h^2 \in \mathcal{C}_V$, the first term of the RHS is upper-bounded by

$$\beta^{k_N}\times N^{-1}\sum_{i=1}^{N}V(\tilde{\xi}^{i,N}) ,$$

which converges in probability to 0. Now, note that the functions h^2 and V are in $\mathcal{C}_V$ and $|h| \leq \max(h^2,1) \leq \max(h^2,V)$ so that $h \in \mathcal{C}_V$. By applying $|a^2-b^2| \leq |a-b|^2 + 2|b||a-b|$, the second term of (6.25) is then upper-bounded by

$$\beta^{2k_N}\times N^{-1}\sum_{i=1}^{N}\left[V(\tilde{\xi}^{i,N})\right]^2 + 2|\Pi h|\beta^{k_N}\times N^{-1}\sum_{i=1}^{N}V(\tilde{\xi}^{i,N}) ,$$

which again converges in probability to 0. This proves (6.23). Now, let $\varepsilon > 0$,

$$\begin{aligned} \sum_{i=1}^{N}\mathbb{E}\left[U_{N,i}^2\mathbf{1}_{\{|U_{N,i}|\geq\varepsilon\}}\middle|\mathcal{F}_{N,i-1}\right] &\leq \Pi\left[h^2\mathbf{1}_{\{h^2\geq\varepsilon^2 N\}}\right] + \frac{1}{N}\sum_{i=1}^{N}\left|Q^{k_N}\left[h^2(\tilde{\xi}^{i,N})\mathbf{1}_{\{h^2(\tilde{\xi}^{i,N})\geq\varepsilon^2 N\}}\right] - \Pi\left[h^2\mathbf{1}_{\{h^2\geq\varepsilon^2 N\}}\right]\right| \\ &\leq \Pi\left[h^2\mathbf{1}_{\{h^2\geq\varepsilon^2 N\}}\right] + \beta^{k_N}\times N^{-1}\sum_{i=1}^{N}V(\tilde{\xi}^{i,N}) , \end{aligned} \tag{6.26}$$

where $h^2\mathbf{1}_{\{h^2\geq\varepsilon^2 N\}} \in \mathcal{C}_V$. Since $h^2 \in \mathcal{C}_V$,(**A2**)-(i) implies that $\Pi h^2 < \infty$. Then, the RHS of (6.26) converges in probability to 0, showing (6.24). The proof is completed.

Annexe A

On the convergence of Island particle models

Sommaire

This appendix is the translation into English of Chapter 3. It displays results from a research work that I have done in collaboration with Pierre Del Moral and Eric Moulines. A journal paper will shortly be written from this Chapter.

A.1 Introduction

Numerical approximation of Feynman-Kac semigroups by systems of interacting particles is a very active research field. Such methods are increasingly used to sample complex high dimensional distributions and they find a wide range of applications in applied probability, among others filtering, smoothing for non linear state-space models, Bayesian inference of hierarchical models, branching processes in biology, absorption problems in physics (see for example Del Moral [2004], Del Moral and Doucet [2010-2011]).

More precisely, let $(\mathbb{X}_n, \mathcal{X}_n)_{n\geq 0}$ be a sequence of measurable sets, and denote $\mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$ the Banach space of all bounded and measurable functions f on $\mathbb{X}_n$, equipped with the uniform norm $\|f\| = \sup_{x_n \in \mathbb{X}_n} |f(x_n)|$. We also consider a sequence of measurable functions g_n referred to as *potential functions* on the state spaces $\mathbb{X}_n$, a distribution η_0 on $\mathbb{X}_0$, and a sequence of Markov kernels M_n from $(\mathbb{X}_{n-1}, \mathcal{X}_{n-1})$ into $(\mathbb{X}_n, \mathcal{X}_n)$. We associate the sequence of Feynman-Kac measures, defined for any $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$ by the formulae

$$\eta_n(f_n) \stackrel{\text{def}}{=} \gamma_n(f_n)/\gamma_n(1) \quad \text{with} \quad \gamma_n(f_n) \stackrel{\text{def}}{=} \int \eta_0(\mathrm{d}x_0) \prod_{0\leq p<n} g_p(x_p) M_{p+1}(x_p, \mathrm{d}x_{p+1}) f_n(x_n) \,, \tag{A.1}$$

to the pair potentials/transitions (g_n, M_n).

The sequences of distributions $(\eta_n)_{n\geq 0}$ and $(\gamma_n)_{n\geq 0}$ are approximated sequentially using interacting particle systems. Such particle approximations are often referred to as sequential Monte-Carlo (SMC)

methods. This algorithm consists in approximating for each $n \in \mathbb{N}$ the probability η_n by a set of *particles* (ξ_n^i) which are generated recursively. The procedure to update the particles is in two steps. In the *selection step* the particles are first sampled with weights proportional to the potential functions. In the *mutation step*, a new generation of particles (ξ_{n+1}^i) is generated from the selected particles using the kernel M_{n+1}. The asymptotic behavior of such particle approximation is now well understood (see Del Moral [2004] and Del Moral et al. [2012]).

Recently, the development of parallel computing methods lead to the study of parallel implementation of this particle approximation. In this paper we focus on the so-called *island particle models*, consisting in running N_2 interacting particle systems each of size N_1. This problem gives rise to several questions among which the optimization of the size of the islands N_1 compared to the number of islands N_2 for a given total computing cost $N \stackrel{\text{def}}{=} N_1 N_2$ and the opportunity of letting the islands interact.

When the N_2 islands are run independently, the bias induced in each of them only depends on their population size N_1; thus it can be interesting to introduce an interaction between the islands even though this interaction increases the final estimator variance through the sampling steps. We propose here to study the asymptotic bias and variance in each case so that when N_1 and N_2 are large and fixed, we can determine which choice is the best in terms of asymptotic mean squared error.

The rest of the paper is organized as follows. In Section A.2 we provide a description of Feynman-Kac models and of their particle approximation with size N_1. We introduce in Section A.3 island Feynman-Kac models which lead to the interacting island algorithm. Finally the asymptotic bias and variance of independent and interacting islands are given in Section A.4.

A.2 Feynman-Kac models

A.2.1 Description of the model

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. In the sequel, all the processes are defined on this probability space. Let $(X_n)_{n\geq 0}$ be a non-homogenous Markov chain (with respect to its natural filtration) with initial distribution η_0, and Markov kernels $(M_n)_{n\geq 1}$. Note that for each $n \in \mathbb{N}$, $X_n \in \mathbb{X}_n$. Using such notation, we may rewrite γ_n defined in (A.1) as follows

$$\gamma_n(f_n) = \mathbb{E}\left[f_n(X_n) \prod_{0\leq p<n} g_p(X_p) \right] .$$

Semigroup description For $\ell \in \mathbb{N}$, consider the finite kernel Q_ℓ from $(\mathbb{X}_{\ell-1}, \mathcal{X}_{\ell-1})$ into $(\mathbb{X}_\ell, \mathcal{X}_\ell)$ given by

$$Q_\ell(x_{\ell-1}, A_\ell) \stackrel{\text{def}}{=} g_{\ell-1}(x_{\ell-1}) M_\ell(x_{\ell-1}, A_\ell) , \quad x_{\ell-1} \in \mathbb{X}_{\ell-1} , \quad A_\ell \in \mathcal{X}_\ell .$$

For $p \leq n$, define by $Q_{p,n}$ the finite kernel from $(\mathbb{X}_p, \mathcal{X}_p)$ into $(\mathbb{X}_n, \mathcal{X}_n)$ by the equation

$$Q_{p,n} \stackrel{\text{def}}{=} Q_{p+1} Q_{p+2} \cdots Q_n .$$

With this definition, we get $\gamma_n = \gamma_p Q_{p,n}$, and, for any $x_p \in \mathbb{X}_p$, $A_n \in \mathcal{X}_n$ by

$$Q_{p,n}(x_p, A_n) = \mathbb{E}\left[\mathbf{1}_{A_n}(X_n) \prod_{p\leq q<n} g_q(X_q) \middle| X_p = x_p \right] .$$

Define by $\mathcal{P}(\mathbb{X}_n, \mathcal{X}_n)$ the set of probability measures on $(\mathbb{X}_n, \mathcal{X}_n)$. It is easily checked that the sequence of measures $(\eta_n)_{n\geq 0}$ satisfies the following recursion

$$\eta_{n+1} = \Phi_{n+1}(\eta_n) , \tag{A.2}$$

where $\Phi_{n+1} : \mathcal{P}(\mathbb{X}_n, \mathcal{X}_n) \to \mathcal{P}(\mathbb{X}_{n+1}, \mathcal{X}_{n+1})$ is given by

$$\Phi_{n+1}(\eta_n) \stackrel{\text{def}}{=} \Psi_n(\eta_n) M_{n+1} . \tag{A.3}$$

In the above displayed equation, $\Psi_n : \mathcal{P}(\mathbb{X}_n, \mathcal{X}_n) \to \mathcal{P}(\mathbb{X}_n, \mathcal{X}_n)$ stands for the Boltzmann-Gibbs transformation

$$\Psi_n(\eta_n)(A_n) \stackrel{\text{def}}{=} \frac{1}{\eta_n(g_n)} \int_{A_n} g_n(x_n)\, \eta_n(\mathrm{d}x_n) \,, \quad A_n \in \mathcal{X}_n \,.$$

Since $\eta_n = \gamma_p Q_{p,n} / \gamma_p(Q_{p,n}(1))$ and $\eta_p = \gamma_p / \gamma_p(1)$, we get

$$\eta_n = \frac{\eta_p Q_{p,n}}{\eta_p(Q_{p,n}(1))} = \eta_p \bar{Q}_{p,n} \,, \quad \text{where} \quad \bar{Q}_{p,n} \stackrel{\text{def}}{=} \frac{Q_{p,n}}{\eta_p(Q_{p,n}(1))} = \frac{\gamma_p(1)}{\gamma_n(1)} Q_{p,n} \,.$$

Nonlinear Markov interpretation Define by S_{n,η_n} the Markov kernel on $(\mathbb{X}_n, \mathcal{X}_n)$ given for $x_n \in \mathbb{X}_n$ and $A_n \in \mathcal{X}_n$ by

$$S_{n,\eta_n}(x_n, A_n) = \varepsilon_n\, g_n(x_n) \delta_{x_n}(A_n) + (1 - \varepsilon_n\, g_n(x_n))\, \Psi_n(\eta_n)(A_n) \,. \tag{A.4}$$

It is not difficult to check that

$$\Psi_n(\eta_n) = \eta_n S_{n,\eta_n} \,. \tag{A.5}$$

This property is valid for any choice of ε_n such that $\varepsilon_n\, g_n \in [0,1]$. For instance, if we set $\varepsilon_n = 0$, we simply have that

$$S_{n,\eta_n}(x_n, A_n) = \Psi_n(\eta_n)(A_n) \,, \quad x_n \in \mathbb{X}_n \,, \quad A_n \in \mathcal{X}_n \,.$$

Using (A.2), (A.3) and (A.5), we get

$$\eta_{n+1} = \eta_n K_{n+1,\eta_n} \quad \text{with} \quad K_{n+1,\eta_n} \stackrel{\text{def}}{=} S_{n,\eta_n} M_{n+1} \,.$$

Mean field particle models We now describe more precisely the particle approximation of the probabilities $(\eta_n)_{n \geq 0}$. Let N_1 be an integer. We define the Markov kernel $\boldsymbol{M}_{n+1}$ from $(\mathbb{X}_n^{N_1}, \mathcal{X}_n^{\otimes N_1})$ to $(\mathbb{X}_{n+1}^{N_1}, \mathcal{X}_{n+1}^{\otimes N_1})$ for any $(x_n^1, \ldots, x_n^{N_1}) \in \mathbb{X}_n^{N_1}$ and $(A_{n+1}^1, \ldots, A_{n+1}^{N_1}) \in \mathcal{X}_{n+1}^{N_1}$ by

$$\boldsymbol{M}_{n+1}(x_n^1, \ldots, x_n^{N_1}, A_{n+1}^1 \times \cdots \times A_{n+1}^{N_1}) \stackrel{\text{def}}{=} \prod_{1 \leq i \leq N_1} K_{n+1, m(x_n^1, \ldots, x_n^{N_1}, \cdot)}(x_n^i, A_{n+1}^i) \,, \tag{A.6}$$

where $m(x_n^1, \ldots, x_n^{N_1}, \cdot)$ stands for the empirical measure of the points $(x_n^1, \ldots, x_n^{N_1})$ given for any $A_n \in \mathcal{X}_n$ by

$$m(x_n^1, \ldots, x_n^{N_1}, A_n) \stackrel{\text{def}}{=} \frac{1}{N_1} \sum_{i=1}^{N_1} \delta_{x_n^i}(A_n) \,.$$

When $\varepsilon_n = 0$, we have $S_{n,\eta_n}(x_n, A_n) = \Psi_n(\eta_n)(A_n)$ which implies that

$$K_{n+1, m(x_n^1, \ldots, x_n^{N_1}, \cdot)}(x_n^i, \cdot) = \Phi_{n+1}(m(x_n^1, \ldots, x_n^{N_1}, \cdot)) \,.$$

We define a Markov chain $(\xi_n^{N_1})_{n \geq 0}$ where for each $n \in \mathbb{N}$, $\xi_n^{N_1} = (\xi_n^{N_1,1}, \ldots, \xi_n^{N_1,N_1}) \in \mathbb{X}_n^{N_1}$ with the transition kernel $\boldsymbol{M}_{n+1}$. And we define by $\mathcal{F}_n^{N_1}$ the increasing filtration associated to the particle evolution:

$$\mathcal{F}_n^{N_1} \stackrel{\text{def}}{=} \sigma\left(\xi_p^{N_1}, 0 \leq p \leq n\right) \,.$$

The N_1-particle approximations of the measures η_n and γ_n are defined for $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$ by

$$\eta_n^{N_1}(f_n) \stackrel{\text{def}}{=} m(\xi_n^{N_1}, f_n) \quad \text{and} \quad \gamma_n^{N_1}(f_n) \stackrel{\text{def}}{=} \eta_n^{N_1}(f_n) \prod_{0 \leq p < n} \eta_p^{N_1}(g_p) \,. \tag{A.7}$$

Proposition A.1. *The estimator $\gamma_n^{N_1}$ is unbiased: for $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$*

$$\mathbb{E}\left[\eta_n^{N_1}(f_n) \prod_{0 \leq p < n} \eta_p^{N_1}(g_p)\right] \;=\; \mathbb{E}\left[f_n(X_n) \prod_{0 \leq p < n} g_p(X_p)\right] \,. \tag{A.8}$$

The proof of this Proposition relies on the following Lemma.

Lemma A.1. *For any $p > 0$ and $f_p \in \mathcal{B}_b(\mathbb{X}_p, \mathcal{X}_p)$, we have*

$$\mathbb{E}\left[\eta_p^{N_1}(f_p)\middle|\mathcal{F}_{p-1}^{N_1}\right] = \frac{\eta_{p-1}^{N_1}(Q_p f_p)}{\eta_{p-1}^{N_1}(g_{p-1})} .$$

Proof. By definition,

$$\mathbb{E}\left[\eta_p^{N_1}(f_p)\middle|\mathcal{F}_{p-1}^{N_1}\right] = \eta_{p-1}^{N_1}(K_{p,\eta_{p-1}^{N_1}} f_p) ,$$

where

$$\begin{aligned} K_{p,\eta_{p-1}^{N_1}}(x_{p-1}, f_p) &= S_{p-1,\eta_{p-1}^{N_1}} M_p(x_{p-1}, f_p) = \int S_{p-1,\eta_{p-1}^{N_1}}(x_{p-1}, \mathrm{d}x'_{p-1}) M_p(x'_{p-1}, f_p) \\ &= \varepsilon_{p-1} g_{p-1}(x_{p-1}) M_p f_p(x_{p-1}) + (1 - \varepsilon_{p-1} g_{p-1}(x_{p-1})) \Psi_{p-1}(\eta_{p-1}^{N_1}) M_p f_p \\ &= \varepsilon_{p-1} Q_p f_p(x_{p-1}) + (1 - \varepsilon_{p-1} g_{p-1}(x_{p-1})) \Psi_{p-1}(\eta_{p-1}^{N_1}) M_p f_p , \end{aligned}$$

so that

$$\mathbb{E}\left[\eta_p^{N_1}(f_p)\middle|\mathcal{F}_{p-1}^{N_1}\right] = \varepsilon_{p-1}\eta_{p-1}^{N_1}(g_{p-1})\frac{\eta_{p-1}^{N_1}(Q_p f_p)}{\eta_{p-1}^{N_1}(g_{p-1})} + (1 - \varepsilon_{p-1}\eta_{p-1}^{N_1}(g_{p-1}))\Psi_{p-1}(\eta_{p-1}^{N_1}) M_p f_p .$$

Finally, the proof is concluded by noting that

$$\Psi_{p-1}(\eta_{p-1}^{N_1}) M_p f_p = \frac{1}{N_1}\sum_{i=1}^{N_1} \frac{g_{p-1}(\xi_{p-1}^{N_1,i})}{\frac{1}{N_1}\sum_{j=1}^{N_1} g_{p-1}(\xi_{p-1}^{N_1,j})} M_p f_p(\xi_{p-1}^{N_1,i}) = \frac{\eta_{p-1}^{N_1}(g_{p-1} M_p f_p)}{\eta_{p-1}^{N_1}(g_{p-1})} = \frac{\eta_{p-1}^{N_1}(Q_p f_p)}{\eta_{p-1}^{N_1}(g_{p-1})} .$$

□

Proof of Proposition A.1. By the definition of $\gamma_n^{N_1}$ given in (A.7), we have

$$\mathbb{E}\left[\gamma_n^{N_1}(f_n)\right] = \mathbb{E}\left[\mathbb{E}\left[\eta_n^{N_1}(f_n)\middle|\mathcal{F}_{n-1}^{N_1}\right] \prod_{0\le p<n} \eta_p^{N_1}(g_p)\right] .$$

We get from Lemma A.1

$$\mathbb{E}\left[\eta_n^{N_1}(f_n)\middle|\mathcal{F}_{n-1}^{N_1}\right] = \frac{\eta_{n-1}^{N_1}(Q_n f_n)}{\eta_{n-1}^{N_1}(g_{n-1})} ,$$

and

$$\mathbb{E}\left[\gamma_n^{N_1}(f_n)\right] = \mathbb{E}\left[\eta_{n-1}^{N_1}(Q_n f_n) \prod_{0\le p<n-1} \eta_p^{N_1}(g_p)\right] .$$

By iterating this step we get

$$\begin{aligned} \mathbb{E}\left[\gamma_n^{N_1}(f_n)\right] &= \mathbb{E}\left[\mathbb{E}\left[\eta_{n-1}^{N_1}(Q_n f_n)\middle|\mathcal{F}_{n-2}^{N_1}\right] \prod_{0\le p<n-1} \eta_p^{N_1}(g_p)\right] = \mathbb{E}\left[\eta_{n-2}^{N_1}(Q_{n-1} Q_n f_n) \prod_{0\le p<n-2} \eta_p^{N_1}(g_p)\right] \\ &= \mathbb{E}\left[\eta_0^{N_1}(Q_{0,n} f_n)\right] = \mathbb{E}\left[Q_{0,n} f_n(\xi_0^{N_1,1})\right] = \gamma_0 Q_{0,n} f_n = \gamma_n(f_n) . \end{aligned}$$

□

Example A.1 (Up&Out barrier call)**.** A straightforward application is the pricing of options in finance. The Up&Out barrier call payoff of strike K, barrier $H > K$, maturity T, on an asset S is defined by

$$\max(S_T - K, 0)\mathbf{1}_{A_H}\left(\max_{0\leq i\leq N_{\text{obs}}} S_{iT/N_{\text{obs}}}\right) ,$$

where $A_H \stackrel{\text{def}}{=} \{x; x \leq H\}$. If the interest rates are supposed uniformly equal to r, this option price is given by

$$P = \mathrm{e}^{-rT}\mathbb{E}\left[\max(S_T - K, 0)\mathbf{1}_{A_H}\left(\max_{0\leq i\leq N_{\text{obs}}} S_{iT/N_{\text{obs}}}\right)\right] .$$

In the framework of the Black-Scholes model:

$$\forall t \geq 0 , \quad S_t = S_0\mathrm{e}^{(r-\sigma^2/2)t+\sigma W_t} ,$$

where $S_0 < H$ is the initial value of the asset, σ is its volatility and the process W is a standard brownian motion.

By setting $n = N_{\text{obs}}$, $X_p = S_{pT/N_{\text{obs}}}$, $g_p = \mathbf{1}_{A_H}$ and $f_n(x) = \max(x - K, 0)\mathbf{1}_{A_H}(x)$ we get

$$P = \mathrm{e}^{-rT}\gamma_n(f_n) .$$

Compared to the classical Monte-Carlo estimate, $\gamma_n^{N_1}$ does not forget the paths leading to zero payoff but it uses a resampling technic to replace them with path that have not reached the barrier. Although both algorithms has a linear complexity, the resampling step increases the computational time. In the case where $K = 100$, $H = 110$, $T = 1$, $N_{\text{obs}} = 250$, $r = 0.05$, $S_0 = 100$ and $\sigma = 0.3$ the CPU time for the SMC method is twice as big as the one of the classical Monte-Carlo. Consequently, to be perfectly fair, the number of particles used for each algorithm has been adjusted. Furthermore, in order to get confidence intervals, the SMC particles has been spread into 150 runs of the SMC algorithm. Figure A.1 shows that for a fixed CPU time, the SMC method leads to a better 95% confidence interval since its half-width as a percentage of the estimated price is lower than the one the classical MC.

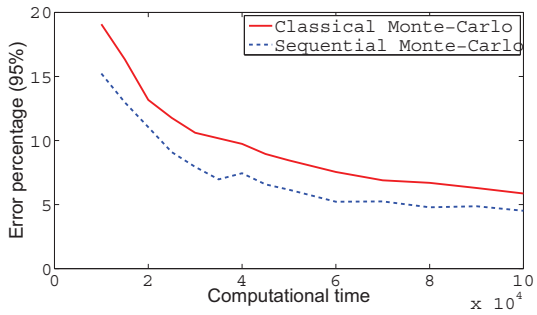


Figure A.1: Confidence interval half-width of SMC and MC methods for the same CPU time.

A.2.2 Asymptotic behavior

The fluctuation analysis of the particle measures $\{\eta_n^{N_1}\}_{n\geq 0}$ around their limiting values $\{\eta_n\}_{n\geq 0}$ is essentially based on the asymptotic analysis of the local sampling errors associated with the particle approximation sampling steps. These local errors are expressed in terms of the random fields $W_n^{N_1}$, given for any bounded function $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$ by the formula

$$W_n^{N_1}(f_n) = \sqrt{N_1}\left[\eta_n^{N_1}(f_n) - \eta_{n-1}^{N_1}(K_{n,\eta_{n-1}^{N_1}} f_n)\right] .$$

The next central limit theorem for random fields is pivotal. Its complete proof can be found in Del Moral [2004] (Corollary 9.3.1, pp. 295-298).

Theorem A.1. *For any fixed time horizon $n \geq 1$, the sequence $(W_p^{N_1})_{1\leq p\leq n}$ converges in law, as N_1 tends to infinity, to a sequence of n independent, Gaussian and centered random fields $(W_p)_{1\leq p\leq n}$ with, for any bounded functions $f_p, h_p \in \mathcal{B}_b(\mathbb{X}_p, \mathcal{X}_p)$, and $1 \leq p \leq n$,*

$$\mathbb{E}\left[W_p(f_p)W_p(h_p)\right] = \eta_{p-1}K_{p,\eta_{p-1}}([f_p - K_{p,\eta_{p-1}}(f_p)][h_p - K_{p,\eta_{p-1}}(h_p)]) \,,$$

where, by convention, for any $x_{p-1} \in \mathbb{X}_{p-1}$

$$\begin{aligned} K_{p,\eta_{p-1}}(x_{p-1}, [f_p - K_{p,\eta_{p-1}}(f_p)][h_p - K_{p,\eta_{p-1}}(h_p)]) \\ &\stackrel{\text{def}}{=} K_{p,\eta_{p-1}}(x_{p-1}, [f_p - K_{p,\eta_{p-1}}(x_{p-1}, f_p)][h_p - K_{p,\eta_{p-1}}(x_{p-1}, h_p)(x_{p-1}])) \\ &= K_{p,\eta_{p-1}}(x_{p-1}, f_p h_p) - K_{p,\eta_{p-1}}(x_{p-1}, f_p)K_{p,\eta_{p-1}}(x_{p-1}, h_p) \,. \end{aligned}$$

In the particular case of $\varepsilon_p \equiv 0$, this covariance becomes

$$\mathbb{E}\left[W_p(f_p)W_p(h_p)\right] = \eta_p\left[(f_p - \eta_p(f_p))\,(g_p - \eta_p(g_p))\right] \,.$$

We now consider the decomposition of the error as a sum of local errors:

$$\gamma_n^{N_1} - \gamma_n = \sum_{p=0}^{n} \left[\gamma_p^{N_1} Q_{p,n} - \gamma_{p-1}^{N_1} Q_{p-1,n}\right] \,,$$

where for $p = 0$ we take the convention $\gamma_{-1}^{N_1} Q_{-1,n} = \gamma_n$. Note also that

$$\begin{aligned} \gamma_p^{N_1} Q_{p,n} - \gamma_{p-1}^{N_1} Q_{p-1,n} &= \gamma_p^{N_1} Q_{p,n} - \gamma_{p-1}^{N_1} Q_p Q_{p,n} \\ &= (\gamma_p^{N_1} - \gamma_{p-1}^{N_1} Q_p) Q_{p,n} = (\gamma_p^{N_1} - \gamma_{p-1}^{N_1}(g_{p-1})\eta_{p-1}^{N_1} K_{p,\eta_{p-1}^{N_1}}) Q_{p,n} \\ &= (\gamma_p^{N_1} - \gamma_p^{N_1}(1)\eta_{p-1}^{N_1} K_{p,\eta_{p-1}^{N_1}}) Q_{p,n} = \gamma_p^{N_1}(1)(\eta_p^{N_1} - \eta_{p-1}^{N_1} K_{p,\eta_{p-1}^{N_1}}) Q_{p,n} \,. \end{aligned}$$

which implies that for any $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$,

$$W_n^{\gamma,N_1}(f_n) \stackrel{\text{def}}{=} \sqrt{N_1}\left[\gamma_n^{N_1} - \gamma_n\right](f_n) = \sum_{p=0}^{n} \gamma_p^{N_1}(1) W_p^{N_1}(Q_{p,n} f_n) \,. \tag{A.9}$$

By Theorem 7.4.2 in Del Moral [2004] the random sequence $(\gamma_p^{N_1}(1))_{0\leq p\leq n}$ converges in probability, as N_1 tends to infinity, to the deterministic sequence $(\gamma_p(1))_{0\leq p\leq n}$. A simple application of Slutsky's Lemma implies that the random fields W_n^{γ,N_1} converge in law, as N_1 tends to infinity, to the Gaussian random fields W_n^{γ} defined for any function $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$ by

$$W_n^{\gamma}(f_n) \stackrel{\text{def}}{=} \sum_{p=0}^{n} \gamma_p(1) W_p(Q_{p,n} f_n) = \gamma_n(1) \sum_{p=0}^{n} W_p(\bar{Q}_{p,n} f_n) \,. \tag{A.10}$$

In much the same way, the sequence of random fields

$$W_n^{\eta,N_1}(f_n) \stackrel{\text{def}}{=} \sqrt{N}\left[\eta_n^{N_1} - \eta_n\right](f_n) = \gamma_n^{N_1}(1)^{-1} W_n^{\gamma,N_1}\left(f_n - \eta_n(f_n)\right) \,, \tag{A.11}$$

converges in law, as N_1 tends to infinity, to the Gaussian random fields W_n^{η} defined for any function $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$ by

$$W_n^{\eta}(f_n) \stackrel{\text{def}}{=} \sum_{p=0}^{n} W_p(\bar{Q}_{p,n}(f_n - \eta_n(f_n))) \,, \tag{A.12}$$

and we get the following Lemma.

Lemma A.2. *For any $f_n^1, f_n^2 \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$, the pair $(W_n^{\gamma,N_1}(f_n^1), W_n^{\eta,N_1}(f_n^2))$ converges in law, as N_1 tends to infinity, to $(W_n^{\gamma}(f_n^1), W_n^{\eta}(f_n^2))$.*

This analysis allows to show the following results which will be the key to the proof of Theorem A.4.

Theorem A.2. *For any time horizon $n \geq 0$ and any function $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$, we have*

$$\lim_{N_1 \to \infty} N_1 \mathbb{E}\left[\eta_n^{N_1}(f_n) - \eta_n(f_n)\right] = B_n(f_n) ,$$
$$\lim_{N_1 \to \infty} N_1 \mathbb{V}\mathrm{ar}\left(\eta_n^{N_1}(f_n)\right) = V_n(f_n) ,$$

where

$$B_n(f_n) \stackrel{\text{def}}{=} -\sum_{p=0}^{n} \mathbb{E}\left[W_p(\bar{Q}_{p,n}(1)) W_p(\bar{Q}_{p,n}(f_n - \eta_n(f_n)))\right] , \tag{A.13}$$

and

$$V_n(f_n) \stackrel{\text{def}}{=} \sum_{p=0}^{n} \mathbb{E}\left[W_p(\bar{Q}_{p,n}(f_n - \eta_n(f_n)))^2\right] . \tag{A.14}$$

Proof. For the bias, we decompose the error as follows

$$\begin{aligned}
\eta_n^{N_1}(f_n) - \eta_n(f_n) &= \frac{\gamma_n^{N_1}}{\gamma_n^{N_1}(1)}(f_n - \eta_n(f_n)) = \frac{\gamma_n(1)}{\gamma_n^{N_1}(1)} \frac{\gamma_n^{N_1}}{\gamma_n(1)}(f_n - \eta_n(f_n)) \\
&= \frac{\gamma_n(1)}{\gamma_n^{N_1}(1)} \left[\gamma_n^{N_1}\left(\frac{f_n - \eta_n(f_n)}{\gamma_n(1)}\right) - \gamma_n\left(\frac{f_n - \eta_n(f_n)}{\gamma_n(1)}\right)\right] \\
&= \left[\frac{\gamma_n(1)}{\gamma_n^{N_1}(1)} - 1\right] \left[\gamma_n^{N_1}\left(\frac{f_n - \eta_n(f_n)}{\gamma_n(1)}\right) - \gamma_n\left(\frac{f_n - \eta_n(f_n)}{\gamma_n(1)}\right)\right] + [\gamma_n^{N_1} - \gamma_n]\left(\frac{f_n - \eta_n(f_n)}{\gamma_n(1)}\right) ,
\end{aligned}$$

where the expectation of the second term is zero according to Proposition A.1. By noting that

$$\left[\frac{\gamma_n(1)}{\gamma_n^{N_1}(1)} - 1\right] = -\frac{\gamma_n(1)}{\gamma_n^{N_1}(1)} \left[\frac{\gamma_n^{N_1}(1)}{\gamma_n(1)} - 1\right] = -\frac{1}{\gamma_n^{N_1}(1)} [\gamma_n^{N_1} - \gamma_n](1) ,$$

we get

$$\begin{aligned}
N_1 \mathbb{E}\left[\eta_n^{N_1}(f_n) - \eta_n(f_n)\right] &= -\mathbb{E}\left[\frac{1}{\gamma_n^{N_1}(1)} W_n^{\gamma,N_1}(1) W_n^{\gamma,N_1}\left(\frac{f_n - \eta_n(f_n)}{\gamma_n(1)}\right)\right] \\
&= -\mathbb{E}\left[\frac{1}{\gamma_n(1)} W_n^{\gamma,N_1}(1) W_n^{\eta,N_1}(f_n)\right] ,
\end{aligned}$$

where W_n^{γ,N_1} and W_n^{η,N_1} are respectively defined in (A.11) and (A.9). According to Theorem A.1:

$$\lim_{N_1 \to \infty} N_1 \mathbb{E}\left[\eta_n^{N_1}(f_n) - \eta_n(f_n)\right] = -\mathbb{E}\left[\frac{1}{\gamma_n(1)} W_n^{\gamma}(1) W_n^{\eta}(f_n)\right] = B_n(f_n) , \tag{A.15}$$

by the definitions of W_n^{γ} and W_n^{η}.

Consider now the following decomposition of the variance:

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left(\eta_n^{N_1}(f_n)\right) &= \mathbb{E}\left[\left(\eta_n^{N_1}(f_n) - \mathbb{E}\left[\eta_n^{N_1}(f_n)\right]\right)^2\right] \\
&= \mathbb{E}\left[\left(\eta_n^{N_1}(f_n) - \eta_n(f_n)\right)^2\right] - \left\{\mathbb{E}\left[\eta_n^{N_1}(f_n) - \eta_n(f_n)\right]\right\}^2 .
\end{aligned}$$

Using (A.15), note that

$$\left\{\mathbb{E}\left[\eta_n^{N_1}(f_n) - \eta_n(f_n)\right]\right\}^2 = O(1/N_1^2) .$$

From the definition (A.11) of W_n^{η,N_1}, it follows

$$\mathbb{E}\left[\left(\eta_n^{N_1}(f_n) - \eta_n(f_n)\right)^2\right] = \frac{1}{N_1} \mathbb{E}\left[W_n^{\eta,N_1}(f_n)^2\right] ,$$

implying that

$$\lim_{N_1\to\infty} N_1 \mathbb{V}\mathrm{ar}\left(\eta_n^{N_1}(f_n)\right) = \mathbb{E}\left[W_n^{\eta}(f_n)^2\right] = V_n(f_n)\,,$$

by the definition of W_n^{η}. □

Corollary A.1. *The asymptotic variance of* $\boldsymbol{\eta}_n$ *is minimal with respect to* $(\varepsilon_p)_{0\le p\le n-1}$ *introduced in* (A.4) *for*

$$\varepsilon_p = \left(\mathrm{essup}_{\eta_p}(g_p)\right)^{-1}\,, \quad 0\le p\le n-1\,.$$

Proof. Theorem A.2 states that the asymptotic variance of $\boldsymbol{\eta}_n$ can be written as a sum of the local sampling errors of the form, for $f_p \in \mathcal{B}_b(\mathbb{X}_p, \mathcal{X}_p)$

$$\eta_p\left(S_{p,\eta_p}\left[f_p - S_{p,\eta_p}(f_p)\right]^2\right)\,.$$

In the above display $S_{p,\eta_p}\left[f_p - S_{p,\eta_p}(f_p)\right]^2$ stands for the function

$$x \mapsto S_{p,\eta_p}\left(x, \left[f_p - S_{p,\eta_p}(f_p)\right]^2\right) = S_{p,\eta_p}\left(x, \left[f_p - S_{p,\eta_p}(x, f_p)\right]^2\right) = S_{p,\eta_p}(x, f_p^2) - S_{p,\eta_p}(x, f_p)^2\,.$$

In the special case $\varepsilon_p = 0$, the function is constant and we simply have

$$S_{p,\eta_p}\left(x, [f - S_{n,\eta_n}(f)]^2\right) = \Psi_p(\eta_p)\left([f_p - \Psi_p(\eta_p)(f_p)]^2\right)\,.$$

In more general cases, we have

$$\eta_p\left(S_{p,\eta_p}\left[f_p - S_{p,\eta_p}(f_p)\right]^2\right) = \Psi_p(\eta_p)\left([f_p - \Psi_p(\eta_p)(f_p)]^2\right) - \left[\eta_p\left(S_{p,\eta_p}(f_p)^2\right) - \Psi_p(\eta_p)(f_p)^2\right]\,.$$

Using the fact that

$$\eta_p\left(S_{p,\eta_p}(f_p)^2\right) - \Psi_p(\eta_p)(f_p)^2 = \eta_p\left[\left(S_{p,\eta_p}(f_p) - \Psi_p(\eta_p)(f_p)\right)^2\right]\,,$$

and

$$S_{p,\eta_p}(f_p) - \Psi_p(\eta_p)(f_p) = \varepsilon_p\, g_p\left(f_p - \Psi_{f_p}(\eta_p)(f_p)\right)\,,$$

we conclude that

$$\eta_p\left(S_{p,\eta_p}\left[f_p - S_{p,\eta_p}(f_p)\right]^2\right) = \Psi_p(\eta_p)\left([f_p - \Psi_p(\eta_p)(f_p)]^2\right) - \varepsilon_p^2\eta_p\left(g_p^2\left(f_p - \Psi_p(\eta_p)(f_p)\right)^2\right)\,.$$

□

Example A.2 (Variance reduction of $\eta_n^{N_1}$ in the stochastic volatility model)**.** In the framework of the stochastic volatility model introduced in Example 1.2, we have simulated the estimator $\eta_{1000}^{N_1}$ for different values of N_1 with $\varepsilon_n = 0$ and $\varepsilon_n = 1/\max_{1\le i\le N_1} g_n(\xi_n^{N_1,i})$. On one hand, using a non-zero ε does not increase the computational time as shown in Figure A.2.This can be explained by noting that the selection of particles that need to be resampled is compensated by the fact the final number of particles which are resampled is reduced.

On the other hand, as expected, choosing $\varepsilon_n = 1/\max_{1\le i\le N_1} g_n(\xi_n^{N_1,i})$ allows to clearly reduce the variance. Table A.1 shows that we get a reduction from 13% to 39% of the variance, compared to the choice $\varepsilon_n = 0$.

We finally prove the following Proposition which will be used in the next Section.

Proposition A.2. *For any time horizon* $n \ge 0$ *and any functions* $f_n^1, f_n^2 \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$ *such that* $\eta_n(f_n^1) = 0$*, we have*

$$\lim_{N_1\to\infty} N_1 \mathbb{E}\left[\eta_n^{N_1}(f_n^1)\eta_n^{N_1}(f_n^2)\prod_{p=0}^{n-1}\eta_p^{N_1}(g_p)\right] = \gamma_n(1)\sum_{p=0}^{n}\mathbb{E}\left[W_p(\bar{Q}_{p,n}(f_n^1))W_p(\bar{Q}_{p,n}(f_n^2 - \eta_n(f_n^2)))\right]\,.$$

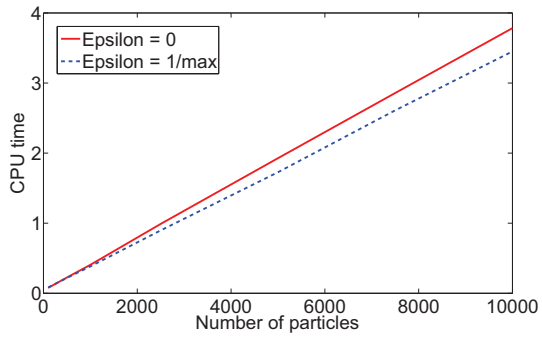


Figure A.2: CPU time with or without ε.

Table A.1: Variance reduction as a percentage of the variance for $\varepsilon = 0$.

N	100	500	1000	2500	5000	7500	10000
Reduction (%)	24	31	34	21	13	39	16

Proof. By the definition (A.7) of $\gamma_n^{N_1}$ we have $\eta_n^{N_1}(f_n^1)\prod_{p=0}^{n-1}\eta_p^{N_1}(g_p) = \gamma_n^{N_1}(f_n^1)$ and according to Proposition A.1, $\mathbb{E}\left[\gamma_n^{N_1}(f_n^1)\right] = \gamma_n(f_n^1) = \gamma_n(1)\eta_n(f_n^1) = 0$ so that we get

$$\begin{aligned}
\mathbb{E}\left[\eta_n^{N_1}(f_n^1)\eta_n^{N_1}(f_n^2)\prod_{p=0}^{n-1}\eta_p^{N_1}(g_p)\right] &= \mathbb{E}\left[\gamma_n^{N_1}(f_n^1)\eta_n^{N_1}(f_n^2)\right] \\
&= \mathbb{E}\left[\gamma_n^{N_1}(f_n^1)\left(\eta_n^{N_1}(f_n^2) - \eta_n(f_n^2)\right)\right] + \mathbb{E}\left[\gamma_n^{N_1}(f_n^1)\right]\eta_n(f_n^2) \\
&= \mathbb{E}\left[\left(\gamma_n^{N_1}(f_n^1) - \gamma_n(f_n^1)\right)\left(\eta_n^{N_1}(f_n^2) - \eta_n(f_n^2)\right)\right] \\
&= \frac{1}{N_1}\mathbb{E}\left[W_n^{\gamma,N_1}(f_n^1)W_n^{\eta,N_1}(f_n^2)\right] .
\end{aligned}$$

Then, Lemma A.2 gives

$$\lim_{N_1\to\infty} N_1\mathbb{E}\left[\eta_n^{N_1}(f_n^1)\eta_n^{N_1}(f_n^2)\prod_{p=0}^{n-1}\eta_p^{N_1}(g_p)\right] = \mathbb{E}\left[W_n^{\eta}(f_n^2)W_n^{\gamma}(f_n^1)\right] ,$$

where W_n^{γ} and W_n^{η} are given by (A.10) and (A.12). □

A.3 Interacting island Feynman-Kac models

For $\mathbf{x}_n \in \mathbb{X}_n^{N_1}$ we set

$$\boldsymbol{g}_n(\mathbf{x}_n) = m(\mathbf{x}_n, g_n) , \tag{A.16}$$

and we introduce the Feynman-Kac model $\{(\boldsymbol{\eta}_n, \boldsymbol{\gamma}_n)\}_{n\geq 0}$ defined by

$$\boldsymbol{\eta}_n(\boldsymbol{f}_n) = \boldsymbol{\gamma}_n(\boldsymbol{f}_n)/\boldsymbol{\gamma}_n(1) \quad \text{and} \quad \boldsymbol{\gamma}_n(\boldsymbol{f}_n) = \mathbb{E}\left[\boldsymbol{f}_n(\xi_n^{N_1})\prod_{0\leq p<n}\boldsymbol{g}_p(\xi_p^{N_1})\right] ,$$

where $\boldsymbol{f}_n \in \mathcal{B}_b(\mathbf{X}_n, \boldsymbol{x}_n)$, and $(\mathbf{X}_n, \boldsymbol{x}_n) \stackrel{\text{def}}{=} (\mathbb{X}_n^{N_1}, x_n^{\otimes N_1})$ and $(\xi_n^{N_1})_{n\geq 0}$ is the Markov chain defined in Section A.2.

If $\boldsymbol{f}_n$ is of the form $\boldsymbol{f}_n(\xi_n^{N_1}) = m(\xi_n^{N_1}, f_n) = \eta_n^{N_1}(f_n)$ for $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$, since $\eta_p^{N_1}(g_p) = m(\xi_p^{N_1}, g_p) = \boldsymbol{g}_p(\xi_p^{N_1})$, according to Proposition A.1, we have

$$\gamma_n(f_n) = \mathbb{E}\left[f_n(X_n) \prod_{0\le p<n} g_p(X_p)\right] = \mathbb{E}\left[\boldsymbol{f}_n(\xi_n^{N_1}) \prod_{0\le p<n} \boldsymbol{g}_p(\xi_p^{N_1})\right] = \boldsymbol{\gamma}_n(\boldsymbol{f}_n)\,,$$

which implies

$$\boldsymbol{\eta}_n(\boldsymbol{f}_n) = \eta_n(f_n)\,. \tag{A.17}$$

Semigroup description For $\ell \in \mathbb{N}$, consider the finite kernel $\boldsymbol{Q}_\ell$ from $(\mathbb{X}_{\ell-1}, \boldsymbol{\mathcal{X}}_{\ell-1})$ into $(\mathbb{X}_\ell, \boldsymbol{\mathcal{X}}_\ell)$ defined by

$$\boldsymbol{Q}_\ell(x_{\ell-1}, \mathbf{A}_\ell) \stackrel{\text{def}}{=} \boldsymbol{g}_{\ell-1}(\mathbf{x}_{\ell-1}) \boldsymbol{M}_\ell(\mathbf{x}_{\ell-1}, \mathbf{A}_\ell)\,, \quad \mathbf{x}_{\ell-1} \in \mathbb{X}_{\ell-1}\,, \quad \mathbf{A}_\ell \in \boldsymbol{\mathcal{X}}_\ell\,,$$

where $\boldsymbol{M}_n$ is defined in (A.6) and $\boldsymbol{g}_n$ in (A.16). For $p \le n$, define by $\boldsymbol{Q}_{p,n}$ the finite kernel from $(\mathbb{X}_p, \boldsymbol{\mathcal{X}}_p)$ into $(\mathbb{X}_n, \boldsymbol{\mathcal{X}}_n)$ by the equation

$$\boldsymbol{Q}_{p,n} \stackrel{\text{def}}{=} \boldsymbol{Q}_{p+1} \boldsymbol{Q}_{p+2} \cdots \boldsymbol{Q}_n\,.$$

With this definition, we get $\boldsymbol{\gamma}_n = \boldsymbol{\gamma}_p \boldsymbol{Q}_{p,n}$, and, for any $\mathbf{x}_p \in \mathbb{X}_p$, $\mathbf{A}_n \in \boldsymbol{\mathcal{X}}_n$

$$\boldsymbol{Q}_{p,n}(\mathbf{x}_p, \mathbf{A}_n) = \mathbb{E}\left[\mathbf{1}_{\mathbf{A}_n}(\xi_n^{N_1}) \prod_{p\le q<n} \boldsymbol{g}_q(\xi_q^{N_1}) \middle| \xi_p^{N_1} = \mathbf{x}_p\right]\,.$$

Define by $\mathcal{P}(\mathbb{X}_n, \boldsymbol{\mathcal{X}}_n)$ the set of probability measures on $(\mathbb{X}_n, \boldsymbol{\mathcal{X}}_n)$. It is easily checked that the sequence of measures $(\boldsymbol{\eta}_n)_{n\ge 0}$ satisfies the following recursion

$$\boldsymbol{\eta}_{n+1} = \boldsymbol{\Phi}_{n+1}(\boldsymbol{\eta}_n)\,, \tag{A.18}$$

where $\boldsymbol{\Phi}_{n+1} : \mathcal{P}(\mathbb{X}_n, \boldsymbol{\mathcal{X}}_n) \to \mathcal{P}(\mathbb{X}_{n+1}, \boldsymbol{\mathcal{X}}_{n+1})$ is given by

$$\boldsymbol{\Phi}_{n+1}(\boldsymbol{\eta}_n) \stackrel{\text{def}}{=} \boldsymbol{\Psi}_n(\boldsymbol{\eta}_n) \boldsymbol{M}_{n+1}\,. \tag{A.19}$$

In the above displayed equation, $\boldsymbol{\Psi}_n : \mathcal{P}(\mathbb{X}_n, \boldsymbol{\mathcal{X}}_n) \to \mathcal{P}(\mathbb{X}_n, \boldsymbol{\mathcal{X}}_n)$ stands for the Boltzmann-Gibbs transformation

$$\boldsymbol{\Psi}_n(\boldsymbol{\eta}_n)(\mathbf{A}_n) \stackrel{\text{def}}{=} \frac{1}{\boldsymbol{\eta}_n(\boldsymbol{g}_n)} \int_{\mathbf{A}_n} \boldsymbol{g}_n(\mathbf{x})\, \boldsymbol{\eta}_n(\mathrm{d}\mathbf{x})\,, \quad \mathbf{A}_n \in \boldsymbol{\mathcal{X}}_n\,.$$

Since $\boldsymbol{\eta}_n = \boldsymbol{\gamma}_p \boldsymbol{Q}_{p,n} / \boldsymbol{\gamma}_p(\boldsymbol{Q}_{p,n}(1))$ and $\boldsymbol{\eta}_p = \boldsymbol{\gamma}_p / \boldsymbol{\gamma}_p(1)$, we get

$$\boldsymbol{\eta}_n = \frac{\boldsymbol{\eta}_p \boldsymbol{Q}_{p,n}}{\boldsymbol{\eta}_p(\boldsymbol{Q}_{p,n}(1))} = \boldsymbol{\eta}_p \bar{\boldsymbol{Q}}_{p,n}\,, \quad \text{where} \quad \bar{\boldsymbol{Q}}_{p,n} \stackrel{\text{def}}{=} \frac{\boldsymbol{Q}_{p,n}}{\boldsymbol{\eta}_p(\boldsymbol{Q}_{p,n}(1))} = \frac{\boldsymbol{\gamma}_p(1)}{\boldsymbol{\gamma}_n(1)} \boldsymbol{Q}_{p,n} = \frac{\gamma_p(1)}{\gamma_n(1)} \boldsymbol{Q}_{p,n}\,.$$

Lemma A.3. *For linear functions $\boldsymbol{f}_n \in \mathcal{B}_b(\mathbb{X}_n, \boldsymbol{\mathcal{X}}_n)$ of the form*

$$\boldsymbol{f}_n(\mathbf{x}_n) = m(\mathbf{x}_n, f_n)\,, \quad f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)\,,$$

we have

$$\boldsymbol{Q}_n(\mathbf{x}_{n-1}, \boldsymbol{f}_n) = m(\mathbf{x}_{n-1}, Q_n f_n)\,, \tag{A.20}$$

for any $p \le n$

$$\boldsymbol{Q}_{p,n}(\mathbf{x}_p, \boldsymbol{f}_n) = m(\mathbf{x}_p, Q_{p,n} f_n)\,, \tag{A.21}$$

and

$$\bar{\boldsymbol{Q}}_{p,n}(\mathbf{x}_p, \boldsymbol{f}_n) = m(\mathbf{x}_p, \bar{Q}_{p,n} f_n)\,.$$

Proof. We have from Lemma A.1

$$\boldsymbol{M}_n(\mathbf{x}_{n-1},\boldsymbol{f}_n)=\mathbb{E}\left[\boldsymbol{f}_n(\xi_n^{N_1})\middle|\xi_{n-1}^{N_1}=\mathbf{x}_{n-1}\right]=\mathbb{E}\left[m(\xi_n^{N_1},f_n)\middle|\xi_{n-1}^{N_1}=\mathbf{x}_{n-1}\right]=\frac{m(\mathbf{x}_{n-1},Q_nf_n)}{m(\mathbf{x}_{n-1},g_{n-1})}\,,$$

which implies

$$\boldsymbol{Q}_n(\mathbf{x}_{n-1},\boldsymbol{f}_n)=\boldsymbol{g}_{n-1}(\mathbf{x}_{n-1})\,\boldsymbol{M}_n(\mathbf{x}_{n-1},\boldsymbol{f}_n)=m(\mathbf{x}_{n-1},g_{n-1})\times\frac{m(\mathbf{x}_{n-1},Q_nf_n)}{m(\mathbf{x}_{n-1},g_{n-1})}\,,$$

showing (A.20). The proof of (A.21) follows by an induction since

$$\boldsymbol{Q}_{p,n}(\mathbf{x}_p,\boldsymbol{f}_n)=\boldsymbol{Q}_{p,n-1}(\mathbf{x}_p,\boldsymbol{Q}_n\boldsymbol{f}_n)\,.$$

□

Interacting island particle models Let N_2 be an integer. We define the Markov kernel $\mathcal{K}_{n+1}^{N_2}$ from $(\mathbb{X}_n^{N_2},\boldsymbol{\mathcal{X}}_n^{\otimes N_2})$ to $(\mathbb{X}_{n+1}^{N_2},\boldsymbol{\mathcal{X}}_{n+1}^{\otimes N_2})$ for any $(\mathbf{x}_n^1,\ldots,\mathbf{x}_n^{N_2})\in\mathbb{X}_n^{N_2}$ and $(\mathbf{A}_{n+1}^1,\ldots,\mathbf{A}_{n+1}^{N_2})\in\boldsymbol{\mathcal{X}}_{n+1}^{N_2}$ by

$$\mathcal{K}_{n+1}^{N_2}(\mathbf{x}_n^1,\ldots,\mathbf{x}_n^{N_2},\mathbf{A}_{n+1}^1\times\cdots\times\mathbf{A}_{n+1}^{N_2})\stackrel{\text{def}}{=}\prod_{1\le i\le N_2}\boldsymbol{\Phi}_{n+1}(\boldsymbol{m}(\mathbf{x}_n^1,\ldots,\mathbf{x}_n^{N_2},\cdot))(\mathbf{A}_{n+1}^i)\,,$$

where $\boldsymbol{m}(\mathbf{x}_n^1,\ldots,\mathbf{x}_n^{N_2},\cdot)$ stands for the empirical measure of the points $(\mathbf{x}_n^1,\ldots,\mathbf{x}_n^{N_2})$ given for any $\mathbf{A}_n\in\boldsymbol{\mathcal{X}}_n$ by

$$\boldsymbol{m}(\mathbf{x}_n^1,\ldots,\mathbf{x}_n^{N_2},\mathbf{A}_n)\stackrel{\text{def}}{=}\frac{1}{N_2}\sum_{i=1}^{N_2}\delta_{\mathbf{x}_n^i}(\mathbf{A}_n)\,.$$

We define a Markov chain $(\boldsymbol{\xi}_n)_{n\ge0}$ where for each $n\in\mathbb{N}$, $\boldsymbol{\xi}_n=(\boldsymbol{\xi}_n^1,\ldots,\boldsymbol{\xi}_n^{N_2})\in\mathbb{X}_n^{N_2}$ with the transition kernel $\mathcal{K}_{n+1}^{N_2}$.

Every individual $\boldsymbol{\xi}_n^i$ can be interpreted as an island with N_1 particles

$$\boldsymbol{\xi}_n^i=\left(\boldsymbol{\xi}_n^{i,j}\right)_{1\le j\le N_1}\in\mathbb{X}_n\,.$$

In this interpretation, the N_2-particle model defined above can be seen as an interacting island particle model. The interaction is given by the empirical mean of the individuals in each island. Also observe that for $N_1=1$, every island has a single particle. In this situation, this model coincides with the N_2-particle model associated with the Feynman-Kac measures η_n. Using (A.3), we also observe that the transition $\boldsymbol{\xi}_n\rightsquigarrow\boldsymbol{\xi}_{n+1}$ can be interpreted as a genetic type selection-mutation transition

$$\boldsymbol{\xi}_n=\left(\boldsymbol{\xi}_n^i\right)_{1\le i\le N_2}\in\mathbb{X}_n^{N_2}\xrightarrow{\text{selection}}\widehat{\boldsymbol{\xi}}_n=\left(\widehat{\boldsymbol{\xi}}_n^i\right)_{1\le i\le N_2}\xrightarrow{\text{mutation}}\boldsymbol{\xi}_{n+1}=\left(\boldsymbol{\xi}_{n+1}^i\right)_{1\le i\le N_2}\in\mathbb{X}_{n+1}^{N_2}$$

During the selection stage, we select randomly N_2 islands $\widehat{\boldsymbol{\xi}}_n=\left(\widehat{\boldsymbol{\xi}}_n^i\right)_{1\le i\le N_2}$ among the current islands $\boldsymbol{\xi}_n=\left(\boldsymbol{\xi}_n^i\right)_{1\le i\le N_2}\in\mathbb{X}_n^{N_2}$ with probability

$$\sum_{i=1}^{N_2}\frac{\boldsymbol{g}_n(\boldsymbol{\xi}_n^i)}{\sum_{j=1}^{N_2}\boldsymbol{g}_n(\boldsymbol{\xi}_n^j)}\,\delta_{\boldsymbol{\xi}_n^i}=\sum_{i=1}^{N_2}\frac{m(\boldsymbol{\xi}_n^i,g_n)}{\sum_{j=1}^{N_2}m(\boldsymbol{\xi}_n^j,g_n)}\,\delta_{\boldsymbol{\xi}_n^i}\,.$$

During the mutation transition, selected islands $\widehat{\boldsymbol{\xi}}_n^i$ evolve randomly to a new configuration $\boldsymbol{\xi}_{n+1}^i$ according to the Markov transition $\boldsymbol{M}_{n+1}$. More formally, $\boldsymbol{\xi}_{n+1}^i$ are independent random variables with distributions $\boldsymbol{M}_{n+1}\left(\widehat{\boldsymbol{\xi}}_n^i,\cdot\right)$.

The N_2-particle approximations of the measures $\boldsymbol{\eta}_n$ and $\boldsymbol{\gamma}_n$ are defined for $\boldsymbol{f}_n\in\mathcal{B}_b(\mathbb{X}_n,\boldsymbol{\mathcal{X}}_n)$ by

$$\boldsymbol{\eta}_n^{N_2}(\boldsymbol{f}_n)\stackrel{\text{def}}{=}\boldsymbol{m}(\boldsymbol{\xi}_n,\boldsymbol{f}_n)\quad\text{and}\quad\boldsymbol{\gamma}_n^{N_2}(\boldsymbol{f}_n)\stackrel{\text{def}}{=}\boldsymbol{\eta}_n^{N_2}(\boldsymbol{f}_n)\prod_{0\le p<n}\boldsymbol{\eta}_p^{N_2}(\boldsymbol{g}_p)\,.\tag{A.22}$$

For linear functions $\boldsymbol{f}_n$ on $\mathbb{X}_n$ of the form $\boldsymbol{f}_n(x_n)=m(x_n,f_n)$, we observe that

$$\boldsymbol{\eta}_n^{N_2}(\boldsymbol{f}_n)=\frac{1}{N_2}\sum_{i=1}^{N_2}\boldsymbol{f}_n(\boldsymbol{\xi}_n^i)=\frac{1}{N_2}\sum_{i=1}^{N_2}m(\boldsymbol{\xi}_n^i,f_n)=\frac{1}{N_1N_2}\sum_{i=1}^{N_2}\sum_{j=1}^{N_1}f_n(\boldsymbol{\xi}_n^{i,j})\,.$$

A.4 Asymptotic bias and variance of island particle models

Given the number of islands N_2 and the number of particles in each of them N_1, is it better to use interacting islands as described earlier or to use the mean of N_2 independent islands? To answer this question, we study the asymptotic behavior of the bias and the variance in both cases.

A.4.1 Independent islands

Let $(\boldsymbol{\zeta}_n^i)_{i=1}^{N_2}$ be N_2 independent particle populations of size N_1 and define for any $\boldsymbol{f}_n \in \mathcal{B}_b(\mathbf{X}_n, \boldsymbol{\mathcal{X}}_n)$

$$\widetilde{\boldsymbol{\eta}}_n^{N_2}(\boldsymbol{f}_n) \stackrel{\text{def}}{=} \frac{1}{N_2} \sum_{i=1}^{N_2} \boldsymbol{f}_n(\boldsymbol{\zeta}_n^i) \, .$$

For linear functions $\boldsymbol{f}_n$ on $\mathbf{X}_n$ of the form $\boldsymbol{f}_n(\mathbf{x}_n) = m(\mathbf{x}_n, f_n)$, $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$, we have

$$\widetilde{\boldsymbol{\eta}}_n^{N_2}(\boldsymbol{f}_n) = \frac{1}{N_2} \sum_{i=1}^{N_2} m(\boldsymbol{\zeta}_n^i, f_n) \, .$$

The asymptotic behavior of the bias and variance of $m(\boldsymbol{\zeta}_n^i, f_n)$ is the one of $\eta_n^{N_1}(f_n) = m(\xi_n^{N_1}, f_n)$ for a generic particle population $\xi_n^{N_1}$ as defined in Section A.2, and according to Theorem A.2 we have for any $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$, for the bias

$$\lim_{N_1 \to \infty} N_1 \left\{ \mathbb{E}\left[\eta_n^{N_1}(f_n)\right] - \eta_n(f_n) \right\} = B_n(f_n) \, ,$$

and for the variance

$$\lim_{N_1 \to \infty} N_1 \mathbb{V}\mathrm{ar}\left(\eta_n^{N_1}(f_n)\right) = V_n(f_n) \, ,$$

where $B_n(f_n)$ and $V_n(f_n)$ are defined respectively in (A.13) and (A.14). The following theorem is then directly derived.

Theorem A.3. *For any time horizon $n \geq 0$ and any $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$, we have*

$$\lim_{N_1 \to \infty} N_1 \left\{ \mathbb{E}\left[\widetilde{\boldsymbol{\eta}}_n^{N_2}(m(\cdot, f_n))\right] - \eta_n(f_n) \right\} = B_n(f_n) \, ,$$
$$\lim_{N_1 \to \infty} N_1 N_2 \mathbb{V}\mathrm{ar}\left(\widetilde{\boldsymbol{\eta}}_n^{N_2}(m(\cdot, f_n))\right) = V_n(f_n) \, ,$$

where $B_n(f_n)$ and $V_n(f_n)$ are defined respectively in (A.13) *and* (A.14)*.*

We can remark that although the variance behaves as $(N_1 N_2)^{-1}$, the bias is independent of N_2 and only behaves as N_1^{-1}.

A.4.2 Interacting islands

In the case of interacting islands as described in Section A.3, the asymptotic behavior of the bias and the variance can also be derived as shown in the following theorem.

Theorem A.4. *For any time horizon $n \geq 0$ and any $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$, we have*

$$\lim_{N_1 \to \infty} \lim_{N_2 \to \infty} N_1 N_2 \mathbb{E}\left[\boldsymbol{\eta}_n^{N_2}(m(\cdot, f_n)) - \boldsymbol{\eta}_n(m(\cdot, f_n))\right] = B_n(f_n) + \widetilde{B}_n(f_n) \, ,$$
$$\lim_{N_1 \to \infty} \lim_{N_2 \to \infty} N_1 N_2 \mathbb{V}\mathrm{ar}\left(\boldsymbol{\eta}_n^{N_2}(m(\cdot, f_n))\right) = V_n(f_n) + \widetilde{V}_n(f_n) \, ,$$

where $B_n(f_n)$, $\widetilde{B}_n(f_n)$, $V_n(f_n)$, $\widetilde{V}_n(f_n)$ are defined respectively in (A.13)*,* (A.26)*,* (A.14) *and* (A.27)*.*

We remark here that the variance presents an additional term compared to the independent case but still behaves as $(N_1 N_2)^{-1}$ whereas this time the bias decreases much more quickly also as $(N_1 N_2)^{-1}$.

Proof. **Asymptotic bias behavior:** For any fixed N_1, the asymptotic bias behavior of $\boldsymbol{\eta}_n^{N_2}(\boldsymbol{f}_n)$ is given for any $\boldsymbol{f}_n \in \mathcal{B}_b(\mathbb{X}_n, \boldsymbol{\mathcal{X}}_n)$ by applying Theorem A.2 to the island particle model for $\varepsilon_n \equiv 0$:

$$\lim_{N_2\to\infty} N_2 \mathbb{E}\left[\boldsymbol{\eta}_n^{N_2}(\boldsymbol{f}_n) - \boldsymbol{\eta}_n(\boldsymbol{f}_n)\right] = -\sum_{p=0}^{n} \boldsymbol{\eta}_p\left[\bar{\boldsymbol{Q}}_{p,n}(1)\bar{\boldsymbol{Q}}_{p,n}(\boldsymbol{f}_n - \boldsymbol{\eta}_n(\boldsymbol{f}_n))\right] .$$

For linear functions $\boldsymbol{f}_n$ of the form $\boldsymbol{f}_n(\cdot) = m(\cdot, f_n)$ where $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$, Lemma A.3 states that

$$\bar{\boldsymbol{Q}}_{p,n}\left(\xi_p^{N_1}, \boldsymbol{f}_n - \boldsymbol{\eta}_n(\boldsymbol{f}_n)\right) = m(\xi_p^{N_1}, \bar{Q}_{p,n}(f_n - \eta_n(f_n))) = \eta_p^{N_1}(\bar{Q}_{p,n}(f_n - \eta_n(f_n))) , \tag{A.23}$$

and

$$\bar{\boldsymbol{Q}}_{p,n}\left(\xi_p^{N_1}, 1\right) = m(\xi_p^{N_1}, \bar{Q}_{p,n}(1)) = \eta_p^{N_1}(\bar{Q}_{p,n}(1)) . \tag{A.24}$$

Therefore, we get

$$\begin{aligned}
\boldsymbol{\eta}_p\left[\bar{\boldsymbol{Q}}_{p,n}(1)\bar{\boldsymbol{Q}}_{p,n}(\boldsymbol{f}_n - \boldsymbol{\eta}_n(\boldsymbol{f}_n))\right] &\overset{(1)}{=} \frac{\boldsymbol{\gamma}_p\left[\bar{\boldsymbol{Q}}_{p,n}(1)\bar{\boldsymbol{Q}}_{p,n}(\boldsymbol{f}_n - \boldsymbol{\eta}_n(\boldsymbol{f}_n))\right]}{\boldsymbol{\gamma}_p(1)} \\
&\overset{(2)}{=} \frac{\mathbb{E}\left[\bar{\boldsymbol{Q}}_{p,n}(\xi_p^{N_1}, 1)\bar{\boldsymbol{Q}}_{p,n}\left(\xi_p^{N_1}, \boldsymbol{f}_n - \boldsymbol{\eta}_n(\boldsymbol{f}_n)\right)\prod_{\ell=0}^{p-1}\boldsymbol{g}_\ell^{N_1}(\xi_\ell^{N_1})\right]}{\boldsymbol{\gamma}_p(1)} \\
&\overset{(3)}{=} \frac{\mathbb{E}\left[\eta_p^{N_1}(\bar{Q}_{p,n}(f_n - \eta_n(f_n)))\eta_p^{N_1}(\bar{Q}_{p,n}(1))\prod_{\ell=0}^{p-1}\eta_\ell^{N_1}(g_\ell)\right]}{\gamma_p(1)} ,
\end{aligned} \tag{A.25}$$

where (1) is simply the definition (A.22) of $\boldsymbol{\eta}_p$, (2) comes from the definition (A.22) of $\boldsymbol{\gamma}_p$, and (3) follows from Proposition A.1, the definition (A.16) of $(\boldsymbol{g}_\ell)_{\ell\geq 0}$ and equations (A.23) and (A.24). As $\eta_p(\bar{Q}_{p,n}(f_n - \eta_n(f_n))) = 0$ we can apply Proposition A.2 and

$$\begin{aligned}
&\lim_{N_1\to\infty} N_1 \boldsymbol{\eta}_p\left[\bar{\boldsymbol{Q}}_{p,n}(1)\bar{\boldsymbol{Q}}_{p,n}(\boldsymbol{f}_n - \boldsymbol{\eta}_n(\boldsymbol{f}_n))\right] \\
&\qquad = \sum_{\ell=0}^{p} \mathbb{E}\left[W_\ell(\bar{Q}_{\ell,p}(\bar{Q}_{p,n}(1) - \eta_p\bar{Q}_{p,n}(1)))W_\ell(\bar{Q}_{\ell,p}\bar{Q}_{p,n}(f_n - \eta_n(f_n)))\right] \\
&\qquad = \sum_{\ell=0}^{p} \mathbb{E}\left[W_\ell(\bar{Q}_{\ell,n}(1) - \bar{Q}_{\ell,p}(1))W_\ell(\bar{Q}_{\ell,n}(f_n - \eta_n(f_n)))\right] ,
\end{aligned}$$

from which we conclude that

$$\begin{aligned}
&\lim_{N_1\to\infty}\lim_{N_2\to\infty} N_1 N_2 \mathbb{E}\left[\boldsymbol{\eta}_n^{N_2}(\boldsymbol{f}_n) - \boldsymbol{\eta}_n(\boldsymbol{f}_n)\right] \\
&\qquad = -\sum_{p=0}^{n}\sum_{\ell=0}^{p} \mathbb{E}\left[W_\ell(\bar{Q}_{\ell,n}(1) - \bar{Q}_{\ell,p}(1))W_\ell(\bar{Q}_{\ell,n}(f_n - \eta_n(f_n)))\right] \\
&\qquad = -\sum_{\ell=0}^{n}\sum_{p=\ell}^{n} \mathbb{E}\left[W_\ell(\bar{Q}_{\ell,n}(1) - \bar{Q}_{\ell,p}(1))W_\ell(\bar{Q}_{\ell,n}(f_n - \eta_n(f_n)))\right] \\
&\qquad = B_n(f_n) + \widetilde{B}_n(f_n) ,
\end{aligned}$$

where $B_n(f_n)$ is defined in (A.13) and $\widetilde{B}_n(f_n)$ is given by

$$\begin{aligned}
\widetilde{B}_n(f_n) \overset{\text{def}}{=} &-\sum_{\ell=0}^{n}(n-\ell)\mathbb{E}\left[W_\ell(\bar{Q}_{\ell,n}(1))W_\ell(\bar{Q}_{\ell,n}(f_n - \eta_n(f_n)))\right] \\
&+ \sum_{\ell=0}^{n} \mathbb{E}\left[W_\ell\left(\sum_{p=\ell}^{n}\bar{Q}_{\ell,p}(1)\right)W_\ell(\bar{Q}_{\ell,n}(f_n - \eta_n(f_n)))\right] .
\end{aligned} \tag{A.26}$$

Asymptotic variance behavior: For any fixed N_1, the asymptotic variance behavior of $\boldsymbol{\eta}_n^{N_2}(\boldsymbol{f}_n)$ is given for any $\boldsymbol{f}_n \in \mathcal{B}_b(\mathbb{X}_n, \boldsymbol{\mathcal{X}}_n)$ by applying Theorem A.2 to the island particle model for $\varepsilon_n \equiv 0$:

$$\lim_{N_2 \to \infty} N_2 \mathbb{V}\mathrm{ar}\left(\boldsymbol{\eta}_n^{N_2}(\boldsymbol{f}_n)\right) = \sum_{p=0}^{n} \boldsymbol{\eta}_p \left[\bar{\boldsymbol{Q}}_{p,n}\left(\boldsymbol{f}_n - \boldsymbol{\eta}_n(\boldsymbol{f}_n)\right)^2\right] .$$

For linear functions $\boldsymbol{f}_n$ of the form $\boldsymbol{f}_n(\cdot) = m(\cdot, f_n)$ where $f_n \in \mathcal{B}_b(\mathbb{X}_n, \mathcal{X}_n)$, using the same steps as in (A.25), we get

$$\boldsymbol{\eta}_p \left[\bar{\boldsymbol{Q}}_{p,n}\left(\boldsymbol{f}_n - \boldsymbol{\eta}_n(\boldsymbol{f}_n)\right)^2\right] = \frac{\mathbb{E}\left[\eta_p^{N_1}(\bar{Q}_{p,n}(f_n - \eta_n(f_n)))^2 \prod_{\ell=0}^{p-1} \eta_\ell^{N_1}(g_\ell)\right]}{\gamma_p(1)} .$$

As $\eta_p(\bar{Q}_{p,n}(\cdot, f_n - \eta_n(f_n))) = 0$ we can apply Proposition A.2 and

$$\begin{aligned}\lim_{N_1 \to \infty} N_1 \boldsymbol{\eta}_p \left[\bar{\boldsymbol{Q}}_{p,n}\left(\boldsymbol{f}_n - \boldsymbol{\eta}_n(\boldsymbol{f}_n)\right)^2\right] &= \sum_{\ell=0}^{p} \mathbb{E}\left[W_\ell(\bar{Q}_{\ell,p}\bar{Q}_{p,n}(f_n - \eta_n(f_n)))^2\right] \\ &= \sum_{\ell=0}^{p} \mathbb{E}\left[W_\ell(\bar{Q}_{\ell,n}(f_n - \eta_n(f_n)))^2\right] ,\end{aligned}$$

from which we conclude that

$$\begin{aligned}\lim_{N_1 \to \infty} \lim_{N_2 \to \infty} N_1 N_2 \mathbb{V}\mathrm{ar}\left(\boldsymbol{\eta}_n^{N_2}(\boldsymbol{f}_n)\right) &= \sum_{p=0}^{n} \sum_{\ell=0}^{p} \mathbb{E}\left[W_\ell(\bar{Q}_{\ell,n}(f_n - \eta_n(f_n)))^2\right] \\ &= \sum_{\ell=0}^{n} \sum_{p=\ell}^{n} \mathbb{E}\left[W_\ell(\bar{Q}_{\ell,n}(f_n - \eta_n(f_n)))^2\right] = V_n(f_n) + \widetilde{V}_n(f_n) ,\end{aligned}$$

where $V_n(f_n)$ is defined in (A.14) and $\widetilde{V}_n(f_n)$ is given by

$$\widetilde{V}_n(f_n) \stackrel{\text{def}}{=} \sum_{\ell=0}^{n} (n-\ell)\mathbb{E}\left[W_\ell(\bar{Q}_{\ell,n}(f_n - \eta_n(f_n)))^2\right] . \tag{A.27}$$

□

A.5 Conclusion

As seen in Theorems A.3 and A.4, a trade-off has to be made between the bias and the variance to choose which of the two estimators $\boldsymbol{\eta}_n^{N_2}$ and $\widetilde{\boldsymbol{\eta}}_n^{N_2}$ is the best. We can compare the mean squared errors given as the sum of the variance and the squared bias, and independent islands are asymptotically better than interacting ones when

$$\frac{V_n(f_n)}{N_1 N_2} + \frac{B_n(f_n)^2}{N_1^2} < \frac{V_n(f_n) + \widetilde{V}_n(f_n)}{N_1 N_2} \quad \Leftrightarrow \quad N_1 > \frac{B_n(f_n)^2}{\widetilde{V}_n(f_n)} N_2 .$$

Consequently, the islands will be kept independent for big values of N_1 compared to N_2 to avoid the additional variance term due to the reselection between interacting islands. On the contrary, we will choose interacting islands for small values of N_1 compared to N_2 to compensate the bias of the independent case.

Example A.3 (Independent vs interacting islands in the LGM)**.** In the framework of the one-dimensional LGM introduced in Example 2.1, η_n can be computed explicitly through the Kalman filter. As a consequence, we used this model to measure the mean square error generated by $\boldsymbol{\eta}_n^{N_2}$ and $\widetilde{\boldsymbol{\eta}}_n^{N_2}$. As previously announced, Figure A.3 shows that using interacting islands is the best choice for small values of N_1 compared to N_2.

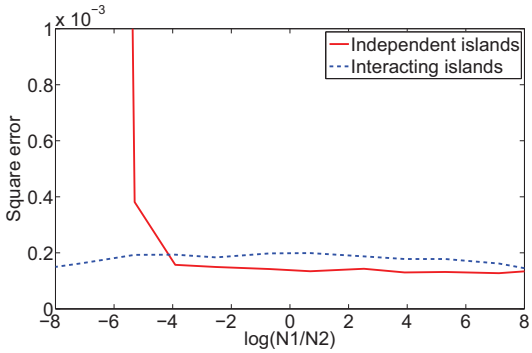


Figure A.3: Mean square error with or without interaction between islands for a constant $N_1 N_2$.

Bibliographie

C. Andrieu, N. De Freitas, A. Doucet, and M.I. Jordan. An introduction to mcmc for machine learning. *Machine learning*, 50 :5–43, 2003.

C. Andrieu, A. Doucet, and R. Holenstein. Particle markov chain monte carlo methods. *J. Roy. Statist. Soc. B*, 72(Part 3) :269–342, 2010.

F. Black. Studies of stock market volatility changes. In *Proceedings of the American Statistical Association*, pages 177–181, 1976.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *J. Econometrics*, 31 :307–327, 1986.

M. Briers, A. Doucet, and S. Maskell. Smoothing algorithms for state-space models. *Annals Institute Statistical Mathematics*, 62(1) :61–89, 2010.

G.L. Buchbinder and K.M. Chistilin. Multiple time scales and the empirical models for stochastic volatility. *Physica A : Statistical Mechanics and its Applications*, 379 :168–178, 2007.

O. Cappé, E. Moulines, and T. Rydén. *Inference in Hidden Markov Models*. Springer, 2005.

O. Cappé, S. J. Godsill, and E. Moulines. An overview of existing methods and recent advances in sequential Monte Carlo. *IEEE Proceedings*, 95(5) :899–924, 2007. doi : 10.1109/JPROC.2007.893250.

J. Carpenter, P. Clifford, and P. Fearnhead. An improved particle filter for non-linear problems. *IEE Proc., Radar Sonar Navigation*, 146 :2–7, 1999.

M. Chernov, R. Gallant, E. Ghysels, and G. Tauchen. Alternative models for stock price dynamics. *J. Econometrics*, 116 :225–257, 2003.

N. Chopin. Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.*, 32(6) :2385–2411, 2004.

N. Chopin, P.e. Jacob, and O. Papaspiliopoulos. *smc*^{2} : A sequential monte carlo algorithm with particle markov chain monte carlo updates. 2011. Preprint, arXiv :1011.1528v2.

A. A. Christie. The stochastic behavior of common stock variances : Value, leverage and interest rate effects. *Journal of Financial Economics*, 10 :407–432, 1982.

J. Cornebise, E. Moulines, and J. Olsson. Adaptive methods for sequential importance sampling with application to state space models. *Stat. Comput.*, 18(4) :461–480, 2008.

J. Davidson. *Stochastic limit theory*. Oxford university press, 1997.

P. Del Moral. *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Springer, 2004.

P. Del Moral and Arnaud Doucet. Particle methods : An introduction with applications. *Springer LNCS/LNAI Tutorial book*, (6368), 2010-2011.

P. Del Moral and A. Guionnet. Central limit theorem for nonlinear filtering and interacting particle systems. *Ann. Appl. Probab.*, 9(2) :275–297, 1999. ISSN 1050-5164.

P. Del Moral and A. Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. *Annales de l'Institut Henri Poincaré*, 37 :155–194, 2001.

P. Del Moral, A. Doucet, and S. Singh. A Backward Particle Interpretation of Feynman-Kac Formulae. *ESAIM M2AN*, 44(5) :947–975, 2010a.

P. Del Moral, A. Doucet, and S. Singh. Forward smoothing using sequential Monte Carlo methods. Preprint, 2010b.

P. Del Moral, P. Hu, and L. Wu. On the concentration properties of interacting particle processes. *Foundations and Trends in Machine Learning*, 3(3-4) :225–289, 2012.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. B*, 39(1) :1–38 (with discussion), 1977.

R. Douc and E. Moulines. Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Ann. Statist.*, 36(5) :2344–2376, 2008.

R. Douc, O. Cappé, and E. Moulines. Comparison of resampling schemes for particle filtering. In *4th International Symposium on Image and Signal Processing and Analysis (ISPA)*, Zagreb, Croatia, September 2005. arXiv : cs.CE/0507025.

R. Douc, É. Moulines, and J. Olsson. Optimality of the auxiliary particle filter. *Probab. Math. Statist.*, 29 (1) :1–28, 2009.

R. Douc, A. Garivier, E. Moulines, and J. Olsson. Sequential Monte Carlo smoothing for general state space hidden Markov models. *To appear in Ann. Appl. Probab.*, 4 2010.

A. Doucet and A.M. Johansen. A tutorial on particle filtering and smoothing : fifteen years later. *Oxford handbook of nonlinear filtering*, 2009.

A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte-Carlo sampling methods for Bayesian filtering. *Stat. Comput.*, 10 :197–208, 2000.

A. Doucet, N. De Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer, New York, 2001.

A. Doucet, G. Poyiadjis, and S.S. Singh. Particle approximations of the score and observed information matrix in state-space models with application to parameter estimation. *Biometrika*, 2010.

J. Durbin and S. J. Koopman. Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *J. Roy. Statist. Soc. B*, 62 :3–29, 2000.

Z. Eisler, J. Perello, and J. Masoliver. Volatility : a hidden markov process in financial time series. *Physical Review E 76 056105*, 2007.

R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50 :987–1007, 1982.

P. Fearnhead. Computational methods for complex stochastic systems : A review of some alternatives to mcmc. *Stat. Comput.*, 18 :151–171, 2008.

P. Fearnhead, D. Wyncoll, and J. Tawn. A sequential smoothing algorithm with linear computational cost. *Biometrika*, 97(2) :447–464, 2010.

J.-P. Fouque, C.-H. Han, and G. Molina. Mcmc estimation of multiscale stochastic volatility models. In C.F. Lee and A.C. Lee, editors, *Handbook of Quantitative Finance and Risk Management*. Springer, 2008.

Walter R. Gilks and Carlo Berzuini. Following a moving target—Monte Carlo inference for dynamic Bayesian models. *J. Roy. Statist. Soc. B*, 63(1) :127–146, 2001.

S. J. Godsill, A. Doucet, and M. West. Monte Carlo smoothing for non-linear time series. *J. Am. Statist. Assoc.*, 99 :156–168, 2004.

N. Gordon, D. Salmond, and A. F. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc. F, Radar Signal Process.*, 140 :107–113, 1993.

P. Hall and C. C. Heyde. *Martingale Limit Theory and its Application*. Academic Press, New York, London, 1980.

J. Handschin and D. Mayne. Monte Carlo techniques to estimate the conditionnal expectation in multi-stage non-linear filtering. In *Int. J. Control*, volume 9, pages 547–559, 1969.

J. Hull and A. White. The pricing of options on assets with stochastic volatilities. *J. Finance*, 42 :281–300, 1987.

M. Hürzeler and H. R. Künsch. Monte Carlo approximations for general state-space models. *J. Comput. Graph. Statist.*, 7 :175–193, 1998.

M. Isard and A. Blake. CONDENSATION – conditional density propagation for visual tracking. *Int. J Computer Vision*, 29(1) :5–28, 1998.

R. E. Kalman and R. Bucy. New results in linear filtering and prediction theory. *J. Basic Eng., Trans. ASME, Series D*, 83(3) :95–108, 1961.

S. Kim, N. Shephard, and S. Chib. Stochastic volatility : Likelihood inference and comparison with ARCH models. *Rev. Econom. Stud.*, 65 :361–394, 1998.

G. Kitagawa. Monte-Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graph. Statist.*, 1 :1–25, 1996.

G. Kitagawa. A self-organizing state-space model. *J. Am. Statist. Assoc.*, 93(443) :1203–1215, 1998.

H. R. Künsch. State space and hidden Markov models. In O. E. Barndorff-Nielsen, D. R. Cox, and C. Kluppelberg, editors, *Complex Stochastic Systems*. CRC Press, 2000.

T. L. Lai and J. Tung. Parameter estimation in hidden Markov models with general state spaces. Technical Report 2003-9, Standford University, 2003.

J. Liu and R. Chen. Sequential Monte-Carlo methods for dynamic systems. *J. Am. Statist. Assoc.*, 93(443) : 1032–1044, 1998.

J. Masoliver and J. Perello. Multiple time scales and the exponential ornstein-uhlenbeck stochastic volatility model. *Quant. Finance*, 6 :423–433, 2006.

D. Q. Mayne. A solution of the smoothing problem for linear dynamic systems. *Automatica*, 4 :73–92, 1966.

J. Olsson and T. Rydén. Metropolising forward particle filtering backward sampling and rao-blackwellisation of metropolised particle smoothers. 2010. Preprint, arXiv :1011.2153v1.

M. K. Pitt and N. Shephard. Filtering via simulation : Auxiliary particle filters. *J. Am. Statist. Assoc.*, 94 (446) :590–599, 1999.

D. B. Rubin. A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when the fraction of missing information is modest : the SIR algorithm (discussion of Tanner and Wong). *J. Am. Statist. Assoc.*, 82 :543–546, 1987.

Tobias Rydén. Consistent and asymptotically normal parameter estimates for hidden Markov models. *Ann. Statist.*, 22(4) :1884–1895, 1994.

G. C. G. Wei and M. A. Tanner. A Monte-Carlo implementation of the EM algorithm and the poor man's Data Augmentation algorithms. *J. Am. Statist. Assoc.*, 85 :699–704, 1991.

M. West and J. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer, 1989.