

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

CALIBRATION DU MODÈLE DE HESTON AVEC FILTRES À PARTICULES
ET RÉ-ÉCHANTILLONNAGE PAR BRANCHEMENT

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN MATHÉMATIQUES

PAR
NICOLAS VELLONE-SCOTT

SEPTEMBRE 2020

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.10-2015). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

J'aimerais tout d'abord remercier mes deux superviseurs de mémoire Anne MacKay et Clarence Simard. C'est à eux que je dois mon intérêt pour la recherche. J'aimerais les remercier pour leurs conseils et leur support. Ils m'ont vraiment aidé en m'amenant à me questionner, ce qui a permis au projet de fortement progresser.

Je voudrais aussi remercier l'Institut Canadien des Actuaires et le Fonds de recherche Nature et Technologies Québec (FRQNT) pour leur soutien financier tout au long de ce projet.

Je voudrais finalement adresser un dernier remerciement tout particulier à mes très chers parents (Nathalie et Domenic). Sans leur support inconditionnel, je n'aurais jamais été capable de me rendre aussi loin. Leur amour et leurs encouragements n'ont vraiment pas de prix et je leur en suis vraiment reconnaissant d'en être aussi généreux.

TABLE DES MATIÈRES

LISTE DES FIGURES	v
LISTE DES TABLEAUX	viii
RÉSUMÉ	ix
INTRODUCTION	1
CHAPITRE I FILTRES À PARTICULES	4
1.1 Modèle de la représentation d'état	4
1.1.1 Généralisation de la représentation d'état	6
1.2 Construction du filtre à particules	7
1.2.1 Échantillonnage préférentiel	11
1.2.2 Échantillonnage d'importance séquentiel	14
1.3 Échantillonnage d'importance séquentiel avec ré-échantillonnage	21
1.3.1 Algorithme de Bootstrap	24
1.3.2 Stabilité du filtre à particules	25
1.4 Amélioration du filtre d'échantillonnage d'importance séquentiel avec ré-échantillonnage (ISR)	26
1.4.1 Ré-échantillonnage stratifié	26
1.4.2 Ré-échantillonnage résiduel	30
1.4.3 Ré-échantillonnage combiné	32
CHAPITRE II ALGORITHMES AVEC BRANCHEMENT	36
2.1 Principe des filtres à particules avec branchement	36
2.1.1 Sélection de modèle	41
2.2 Modification de l'algorithme avec branchement	42
2.2.1 Algorithme avec branchement combiné	42
2.2.2 Algorithme avec branchement effectif	43
CHAPITRE III MODÈLE DE HESTON	46
3.1 Description du modèle	46

3.2	Discrétisation d'Euler-Maruyama	48
3.2.1	Principe de la discrétisation d'Euler-Maruyama	48
3.2.2	Application au modèle d'Heston	50
3.3	Discrétisation de Milstein	51
3.3.1	Principe	51
3.3.2	Application du filtre à particules au modèle d'Heston	55
3.4	Application du filtre à particules au modèle de Heston	56
3.4.1	Définition des variables	56
3.4.2	Discrétisation du modèle de Heston	56
3.4.3	Les probabilités de transitions	57
	CHAPITRE IV RÉSULTATS NUMÉRIQUES	60
4.1	Modèle simulé	60
4.1.1	Paramètres du modèle	60
4.2	Comparaison des performances des différents filtres à particules	63
4.2.1	Filtres à particules avec ré-échantillonnage	63
4.2.2	Erreur de prédiction	68
4.2.3	Filtres à particules avec branchement	70
4.3	Optimisation des paramètres à l'aide des filtres à particules	79
4.3.1	Technique d'optimisation du modèle de Heston à l'aide des filtres à particules	79
4.3.2	Filtres à particules avec ré-échantillonnage	81
4.3.3	Filtres à particules avec branchement	86
4.4	Comparaison des résultats des différents filtres	91
	CONCLUSION	93

LISTE DES FIGURES

Figure	Page
1.1 Schéma du modèle de représentation d'état	5
1.2 Schéma du modèle de représentation d'état généralisée	7
2.1 Exemple de branchement	39
4.1 Log-rendements quotidiens	62
4.2 Volatilité stochastique	62
4.3 Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec ré-échantillonnage standard avec la volatilité réelle de la trajectoire, ($N = 10\ 000$)	64
4.4 Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec ré-échantillonnage stratifié avec la volatilité réelle de la trajectoire, ($N = 10\ 000$)	65
4.5 Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec ré-échantillonnage résiduel avec la volatilité réelle de la trajectoire, ($N = 10\ 000$)	65
4.6 Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec ré-échantillonnage combiné avec la volatilité réelle de la trajectoire, ($N = 10\ 000$)	66
4.7 Évolution quotidienne des quantiles 95% (en bleu) et 5% (en rouge) de la densité estimée par un filtre à particules avec ré-échantillonnage standard ($N = 10\ 000$) de la volatilité stochastique à travers le temps.	67
4.8 Erreur de prédiction moyenne des différents filtres à particules avec ré-échantillonnage par rapport au temps d'exécution de 50 filtres.	69
4.9 Écart-type de l'erreur de prédiction des différents filtres à particules avec ré-échantillonnage par rapport au temps d'exécution de 50 filtres.	69
4.10 Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement combiné avec un taux de contagion de 95% avec la volatilité réelle de la trajectoire, ($N_0 = 10\ 000$)	72

4.11	Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement combiné avec un taux de contagion de 70% avec la volatilité réelle de la trajectoire, ($N_0 = 10\,000$)	72
4.12	Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement combiné avec un taux de contagion de 30% avec la volatilité réelle de la trajectoire, ($N_0 = 10\,000$)	73
4.13	Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement effectif avec un taux de contagion de 95% avec la volatilité réelle de la trajectoire, ($N_0 = 10\,000$)	73
4.14	Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement effectif avec un taux de contagion de 70% avec la volatilité réelle de la trajectoire, ($N_0 = 10\,000$)	74
4.15	Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement effectif avec un taux de contagion de 30% avec la volatilité réelle de la trajectoire, ($N_0 = 10\,000$)	74
4.16	Erreur de prédiction moyenne des différents filtres à particules avec branchement par rapport au temps d'exécution de 50 filtres.	76
4.17	Écart-type de l'erreur de prédiction des différents filtres à particules avec branchement par rapport au temps d'exécution de 50 filtres.	77
4.18	Comparaison des erreurs de prédiction moyenne des meilleurs filtres à particules avec ré-échantillonnage et avec branchement par rapport au temps d'exécution de 50 filtres.	78
4.19	Comparaison des écart-type de l'erreur de prédiction des meilleurs filtres à particules avec ré-échantillonnage et avec branchement par rapport au temps d'exécution de 50 filtres.	78
4.20	Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec ré-échantillonnage standard avec les paramètres initiaux et estimés par rapport à la volatilité réelle de la trajectoire, ($N = 10\,000$)	82
4.21	Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec ré-échantillonnage stratifié avec les paramètres initiaux et estimés par rapport à la volatilité réelle de la trajectoire, ($N = 10\,000$)	83

4.22	Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec ré-échantillonnage résiduel avec les paramètres initiaux et estimés par rapport à la volatilité réelle de la trajectoire, ($N = 10\,000$)	83
4.23	Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec ré-échantillonnage combiné avec les paramètres initiaux et estimés par rapport à la volatilité réelle de la trajectoire, ($N = 10\,000$)	84
4.24	Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement combiné avec un taux de contagion de 95% avec les paramètres initiaux et estimés par rapport à la volatilité réelle de la trajectoire, $N_0 = 10\,000$	87
4.24	Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement combiné avec un taux de contagion de 70% avec les paramètres initiaux et estimés par rapport à la volatilité réelle de la trajectoire, $N_0 = 10\,000$	88
4.25	Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement combiné avec un taux de contagion de 30% avec les paramètres initiaux et estimés par rapport à la volatilité réelle de la trajectoire, $N_0 = 10\,000$	88
4.26	Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement effectif avec un taux de contagion de 95% avec les paramètres initiaux et estimés par rapport à la volatilité réelle de la trajectoire, $N_0 = 10\,000$	89
4.27	Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement effectif avec un taux de contagion de 70% avec les paramètres initiaux et estimés par rapport à la volatilité réelle de la trajectoire, $N_0 = 10\,000$	89
4.28	Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement effectif avec un taux de contagion de 30% avec les paramètres initiaux et estimés par rapport à la volatilité réelle de la trajectoire, $N_0 = 10\,000$	90

LISTE DES TABLEAUX

Tableau	Page
4.1 Paramètres initiaux	61
4.2 Constante multiplicative des bornes de l'intervalle pour les filtres avec branchement	70
4.3 Paramètres estimés du modèles de Heston par les filtres à particules avec ré-échantillonnage	81
4.4 Erreur de prédiction par les filtres à particules avec ré-échantillonnage avec des paramètres optimisés et son écart-type, $N = 10\,000$	85
4.5 Paramètres estimés du modèles de Heston par les filtres à particules avec branchement	86
4.6 Erreur de prédiction par les filtres à particules avec branchement avec des paramètres optimisés et son écart-type, $N_0 = 10\,000$	90

RÉSUMÉ

Les filtres à particules permettent, à partir d'une série d'observations, d'estimer a posteriori la densité d'une variable d'état servant à générer ses observations. Cette estimation peut par la suite être utilisée afin de prédire la valeur des prochaines observations. Leur application est très répandue notamment dans le domaine financier. Ceux-ci peuvent être très utiles pour estimer la valeur d'une variable non-observable, par exemple la volatilité stochastique. Un des travaux récents est celui de Kouritzin (2016) qui a introduit des algorithmes avec branchement permettant un ré-échantillonnage partiel. Dans ces algorithmes de branchement, les particules ayant un poids s'approchant du poids moyen sont conservées et les particules dont le poids s'en éloigne trop sont ré-échantillonnées. Ces algorithmes permettent de réduire le nombre de particules ré-échantillonnées ce qui diminue les temps de calcul et permet une meilleure propagation du poids des particules d'une période à l'autre. Ces algorithmes n'ont toutefois été que très peu testés.

Dans ce mémoire, nous comparons les filtres à particules avec ré-échantillonnage et les filtres à particules avec branchement dans leur application au modèle de Heston (1993). Dans ce mémoire, il est démontré que l'application de ces nouveaux algorithmes proposés par Kouritzin (2016) permet, lorsque l'intervalle de poids à l'intérieur duquel les particules sont conservées est choisi de façon adéquate, de réduire l'erreur de prédiction des filtres à particules ainsi que son écart-type par rapport au filtre à particules avec ré-échantillonnage. Ces algorithmes montrent aussi une bonne capacité à estimer les paramètres du modèle de Heston par maximum de vraisemblance.

Mot clés : Filtre à particules, filtre à particules avec branchement, représentation d'état, volatilité stochastique, Heston, maximum de vraisemblance.

INTRODUCTION

Le filtre à particules ou méthode de Monte-Carlo séquentielle est une méthode d'estimation de la densité d'une variable aléatoire non observable par inférence bayésienne. En supposant que nous connaissons la série temporelle des observations, le principe général de l'algorithme est de simuler des particules représentant des valeurs possibles de cette variable non observable à partir d'une densité équivalente à la densité réelle de la fonction. Ensuite, la crédibilité de chacune de ces particules par rapport aux observations passées est calculée à l'aide d'inférence bayésienne. Un second échantillonnage s'effectue par la suite sur les particules afin de conserver celles qui sont les plus crédibles.

L'ancêtre des filtres à particules, l'échantillonnage préférentiel, a été développé pour la première fois par Kahn et Marshall (1953). Dans ce modèle, la série temporelle de la variable non observable est simulée au complet et la crédibilité de la série temporelle simultanément est ensuite calculée. Cette méthode est surtout adaptée pour estimer des variables non observable sur une seule période lorsque la densité réelle est connue, mais difficilement simulable. Lorsque le nombre de périodes augmente, cette méthode entraîne toutefois des problèmes majeurs puisqu'en général la densité réelle de la série temporelle de la variable non observable est difficilement calculable. Elle nécessite une analyse récursive qui entraîne souvent de très longs calculs.

Pour remédier à cette faiblesse, l'échantillonnage d'importance séquentiel a été développé par Hammersley et Morton (1954). Cette méthode propose une approche similaire à l'échantillonnage préférentiel, mais dans laquelle les particules sont simulées et leurs crédibilités sont calculées à chaque période. Cette méthode a permis de nettement réduire la lourdeur des calculs puisqu'il n'était plus nécessaire de faire une analyse récursive sur toutes les périodes précédentes pour calculer la crédibilité

des particules. Un défaut de cet algorithme est qu'il est fréquent que la crédibilité ou le poids de certaines particules deviennent trop élevés. Dans ces cas, la majorité du poids se retrouve concentré en un très petit nombre de particules ce qui biaise la densité estimée. Pour avoir une estimation précise de la variable, il devient alors nécessaire d'avoir un très grand nombre de particules.

Les filtres à particules ont finalement été conçus par Gordon *et al.* (1993) pour contrer ce défaut. Le filtre à particules ajoute un ré-échantillonnage à l'échantillonnage d'importance séquentiel. Ce ré-échantillonnage permet de rééquilibrer le poids de chaque particule après chaque période. De cette façon, il est possible d'éviter une trop grande concentration des poids. Depuis, leur utilisation s'est popularisée. Dans le domaine financier, sa première utilisation a été faite par Kim *et al.* (1998). Leur objectif était de trouver la densité de la volatilité stochastique de modèle de type ARCH.

Un des problèmes du filtre à particules standard est qu'il a très peu de mémoire, c'est-à-dire qu'après chaque période, chaque particule a le même poids. Ceci fait en sorte que très peu d'informations sont conservées sur chacune des particules. Kouritzin (2016) propose un nouvel algorithme, les filtres à particules avec branchement, afin de permettre de conserver de la mémoire de certaines particules. Le principe est simple : les particules dont le poids est dans un certain intervalle près de la moyenne restent intactes et les autres particules sont ré-échantillonnées aléatoirement.

Cet algorithme est assez récent et n'a été testé que sur un nombre limité de modèles. Dans le contexte de ce mémoire, notre objectif est de vérifier si cet algorithme peut être utilisé pour des modèles à volatilité stochastique tels que le modèle de Heston (1993). Pour ce faire, nous comparons sa performance avec celle de plusieurs variantes du filtre à particules standard pour voir si son efficacité peut surpasser celle des filtres à particules de base.

Le premier chapitre de ce mémoire sert essentiellement de revue de la littérature pour les filtres à particules. Dans la section 1.1, un modèle général sous lequel il

est possible d'utiliser les filtres à particules est présenté. Dans les sections 1.2 et 1.3, nous reconstruisons le filtre à particules en présentant la problématique et en montrant l'évolution des méthodes d'approximation de la densité de la variable latente ayant mené au développement du filtre à particules et y présentons aussi les caractéristiques du filtre à particules. La section 1.4 présente plusieurs modèles permettant de réduire la variance du ré-échantillonnage.

Le deuxième chapitre décrit les filtres à particules avec branchement développés par Kouritzin (2016). La section 2.1 décrit les filtres à particules avec branchement et donne ses caractéristiques. La section 2.2 propose des modifications permettant de mieux contrôler le nombre de particules du filtre et d'effectuer le branchement de façon plus efficace.

Le troisième chapitre présente le modèle de Heston (1993). Dans la section 3.1, le modèle et ses différentes dynamiques sont présentés. Les sections 3.2 et 3.3 décrivent les discrétisations d'Euler-Maruyama par Maruyama (1954) et de Milstein par Milstein (1975) ainsi que la façon de les appliquer au modèle de Heston. Dans la section 3.3.2, nous présentons une technique pour appliquer le filtre à particules à une série d'observations générées à partir du modèle de Heston.

Le quatrième chapitre présente les résultats numériques que nous avons obtenus en appliquant les filtres à particules au modèle de Heston. Dans ce mémoire, l'objectif est de comparer les filtres à particules avec branchement avec ceux avec ré-échantillonnage pour voir si ces nouveaux filtres sont mieux adaptés à ce type de problème. La section 4.1 décrit le modèle et la trajectoire utilisés afin de comparer les différents filtres à particules. La section 4.2 compare la performance des filtres à particules en utilisant les paramètres utilisés pour générer l'observation. La section 4.3 compare la performance de l'optimisation des divers filtres à particules. La section 4.4 offre un sommaire des résultats et compare la performance générale des filtres à particules avec branchement avec celle des filtres à particules avec ré-échantillonnage.

CHAPITRE I

FILTRES À PARTICULES

Pour de nombreux phénomènes aléatoires, les modèles statiques ne semblent pas adéquats pour décrire leur distribution. Afin de mieux estimer la dynamique de ces systèmes, certains modèles mathématiques développés récemment font appel à des variables d'état.

Les variables d'état sont des variables aléatoires X , pouvant être multidimensionnelles, sur lesquelles se base le modèle afin de décrire la dynamique des variables observables Y . Un des problèmes souvent rencontrés avec ce type de modèles est que les variables d'état, aussi appelées variables latentes, ne sont pas observables. Il devient donc nécessaire de pouvoir estimer la distribution de ces variables d'état afin d'être capable d'utiliser le modèle à des fins prédictives.

Le filtre à particules est une méthode d'estimation bayésienne permettant d'estimer la distribution a posteriori de la variable d'état.

1.1 Modèle de la représentation d'état

Soit un processus stochastique observable étudié uniquement à des temps discrets $Y = \{Y_k, k = 0, 1, 2, \dots\}$. Il est supposé qu'il soit impossible d'observer le processus stochastique d'état X . Toutefois, nous cherchons à en estimer la valeur à des temps discrets $X = \{X_k, k = 0, 1, 2, \dots\}$.

Le modèle de représentation d'état est un modèle dans lequel la valeur du processus

stochastique d'état dépend uniquement de la valeur de ce processus à la période précédente et où la valeur du processus stochastique observable dépend uniquement de la valeur du processus stochastique d'état. Un schéma de ce modèle est présenté à la figure 1.1.

Ce modèle est représenté par les équations

$$Y_k = h_k(X_k, \epsilon_k) \quad (1.1)$$

et

$$X_k = g_k(X_{k-1}, \zeta_k). \quad (1.2)$$

Dans ce modèle, h_k représente la fonction liant la variable observable Y_k à la variable latente X_k . $\{\epsilon_k\}_{k=1}^{\infty}$ représente une série de variables aléatoires indépendantes et identiquement distribuées. Ces variables aléatoires sont aussi indépendantes de X_k . Elles servent à modéliser la composante aléatoire de la variable observable. g_k représente la fonction liant le vecteur de variables non observables à ses valeurs passées. $\{\zeta_k\}_{k=1}^{\infty}$ est une série de vecteurs de variables aléatoires indépendantes et identiquement distribuées. Elle représente la composante aléatoire dans la variable latente.

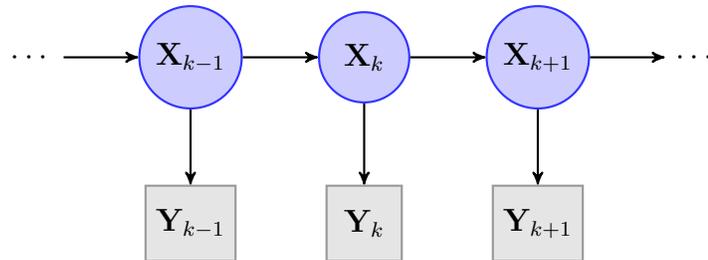


Figure 1.1: Schéma du modèle de représentation d'état

Soit θ un vecteur représentant les paramètres du modèle. Les fonctions de densité de transition pour l'équation de la variable observable et celle de la variable d'état

sont dénotées par $f(Y_k|X_k; \theta)$ et $f(X_k|X_{k-1}; \theta)$.

Dans les modèles de représentation d'état, le processus stochastique d'état est un processus markovien puisque sa valeur dépend uniquement de sa valeur à la période précédente. De plus, il est possible d'observer que la valeur de la variable observable ne dépend que de la valeur de la variable latente à la même période.

1.1.1 Généralisation de la représentation d'état

Il est possible de généraliser ce modèle afin que la fonction g dépende aussi de la dernière observation du processus stochastique Y . Ceci peut être observé notamment lorsque les composantes aléatoires de la variable observable ϵ_k et de la variable latente ζ_k sont corrélées. Dans ces cas, il est généralement possible d'exprimer la variable latente en termes d'une composante aléatoire $\tilde{\zeta}_k$ indépendante de celle de la variable observable ϵ_k . Toutefois, la fonction g sera forcément dépendante de la dernière observation Y_{k-1} .

Sous cette généralisation, l'équation (1.1) décrivant la dynamique de la variable observable demeure inchangée. Par contre, il faut ajuster l'équation (1.2) afin d'inclure une dépendance à la dernière observation Y_{k-1} dans la dynamique de la variable d'état. Le modèle sera plutôt représenté par les équations

$$Y_k = h_k(X_k, \epsilon_k) \tag{1.3}$$

et

$$X_k = g_k(X_{k-1}, Y_{k-1}, \tilde{\zeta}_k). \tag{1.4}$$

Les fonctions de transitions de la variable observable et de la variable latente seront maintenant exprimées par $f(Y_k|X_k; \theta)$ et $f(X_k|X_{k-1}, Y_{k-1}; \theta)$.

Dans cette généralisation de la représentation d'état, la variable observable dépend

toujours exclusivement de la valeur de la variable d'état à la même période. Toutefois, la variable d'état devient maintenant dépendante de la valeur de la variable observable à la période précédente en plus de la variable d'état à la période précédente. Un schéma de ce modèle est présenté à la figure 1.2.

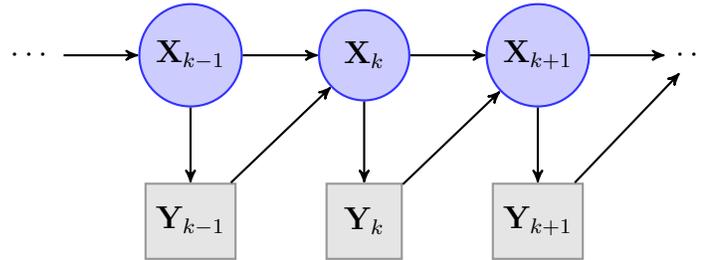


Figure 1.2: Schéma du modèle de représentation d'état généralisée

1.2 Construction du filtre à particules

Supposons, maintenant, que nous avons observé la série temporelle $Y_{1:n}$ représentant le processus stochastique Y observé sur les périodes allant de 1 à n . Notons, ici, que $Y_{1:n}$ dénote le vecteur des observations Y_1, Y_2, \dots, Y_n . Il est impossible d'observer la série temporelle représentant l'état $X_{1:n}$. Notre objectif consiste à estimer la distribution de $X_{1:n}$ et éventuellement avoir un estimateur pour son espérance conditionnelle aux observations $Y_{1:n}$. Ceci permettrait de mieux prédire la valeur des prochaines observations et faciliterait la création d'outils d'aide à la décision.

Pour ce faire, il faut trouver la fonction de densité de la série temporelle des variables d'état conditionnelles à la série temporelle de la variable mesurable $f(X_{1:n}|Y_{1:n})$. La construction faite dans ce mémoire s'inspire de celles effectuées par Creal (2012) et Arulampalam *et al.* (2002). Par contre, notre approche considère l'utilisation de la représentation d'état généralisée alors que les deux articles cités emploient plutôt des représentations d'état, où g_k est indépendant de Y_{k-1} , afin d'appliquer les filtres à particules. Par souci de simplicité, nous allons dénoter la densité conditionnelle recherchée par d_n .

Par la définition d'une densité conditionnelle, d_n vaut

$$d_n = \frac{f(X_{1:n}, Y_{1:n})}{f(Y_{1:n})},$$

où $f(X_{1:n}, Y_{1:n})$ est la fonction de densité jointe des séries temporelles des processus stochastiques de X et Y et où $f(Y_{1:n})$ est la fonction de densité de la série temporelle des observations $Y_{1:n}$. $f(Y_{1:n})$ peut aussi être interprétée comme la vraisemblance de la série temporelle des observations, que nous dénoterons $\mathcal{L}_{1:n}$. En appliquant le théorème de Bayes au numérateur et au dénominateur, l'équation précédente se décompose en

$$d_n = \frac{f(Y_n|X_{1:n}, Y_{1:n-1}) f(X_{1:n}, Y_{1:n-1})}{f(Y_{1:n-1}) f(Y_n|Y_{1:n-1})}.$$

Par les propriétés bayésiennes, il est possible de remplacer la fonction conjointe $f(X_{1:n}, Y_{1:n-1})$ par $f(X_{1:n}|Y_{1:n-1}) f(Y_{1:n-1})$ et de décomposer l'équation précédente comme étant

$$d_n = \frac{f(Y_n|X_{1:n}, Y_{1:n-1}) f(X_{1:n}|Y_{1:n-1}) f(Y_{1:n-1})}{f(Y_{1:n-1}) f(Y_n|Y_{1:n-1})}.$$

En utilisant le théorème de Bayes une dernière fois, $f(X_{1:n}|Y_{1:n-1})$ se décompose par $f(X_{1:n}|Y_{1:n-1}) = f(X_n|X_{1:n-1}, Y_{1:n-1}) f(X_{1:n-1}|Y_{1:n-1})$. De cette façon, l'équation précédente devient

$$d_n = f(X_{1:n-1}|Y_{1:n-1}) \frac{f(Y_n|X_{1:n}, Y_{1:n-1}) f(X_n|X_{1:n-1}, Y_{1:n-1}) f(Y_{1:n-1})}{f(Y_{1:n-1}) f(Y_n|Y_{1:n-1})}.$$

Il est possible de reconnaître la densité de la série temporelle d'état conditionnelle à la série temporelle des observations à la période précédente d_{n-1} dans l'équation précédente. En remplaçant cette définition par d_{n-1} , l'équation précédente devient

$$d_n = d_{n-1} \frac{f(Y_n|X_{1:n}, Y_{1:n-1}) f(X_n|X_{1:n-1}, Y_{1:n-1})}{f(Y_n|Y_{1:n-1})}.$$

Par les propriétés de la représentation d'état généralisée, l'équation précédente se simplifie à

$$d_n = d_{n-1} \frac{f(Y_n|X_n) f(X_n|X_{n-1}, Y_{n-1})}{f(Y_n|Y_{1:n-1})}. \quad (1.5)$$

Dans cette équation, les différents termes du numérateur sont facilement calculables analytiquement, car il s'agit uniquement des densités de transition de la variable observable et de la variable d'état qu'il est possible de déduire à partir des équations (1.3) et (1.4) décrivant le modèle. L'autre terme au numérateur est la densité de la série temporelle à la période précédente supposant que tous les autres facteurs sont calculables ; celui-ci s'obtient facilement par récursion. Le problème se situe toutefois au niveau du dénominateur.

Le facteur $f(Y_n|Y_{1:n-1})$ qui représente la contribution de la n^e période à la vraisemblance du modèle est très difficilement calculable analytiquement. Ce facteur sera noté \mathcal{L}_n à partir de maintenant.

Pour obtenir cette contribution à la vraisemblance du modèle \mathcal{L}_n , il est possible, à l'aide de la définition d'une densité conditionnelle, de réécrire $f(Y_n|Y_{1:n-1})$ en termes de l'intégrale

$$\mathcal{L}_n = \int f(Y_n|X_n) f(X_n|Y_{1:n-1}) dX_n. \quad (1.6)$$

La fonction de densité $f(Y_n|X_n)$ est définie par le modèle. Pour résoudre l'intégrale (1.6), il est nécessaire de connaître la densité $f(X_n|Y_{1:n-1})$. Pour la calculer, il faut intégrer la densité de $X_n|Y_{1:n-1}$ conditionnée par rapport à X_{n-1} .

$$f(X_n|Y_{1:n-1}) = \int f(X_n|X_{n-1}, Y_{1:n-1}) f(X_{n-1}|Y_{1:n-1}) dX_{n-1}$$

Par les propriétés de la représentation d'état généralisée, l'intégrale précédente se simplifie à

$$f(X_n|Y_{1:n-1}) = \int f(X_n|X_{n-1}, Y_{n-1}) f(X_{n-1}|Y_{1:n-1}) dX_{n-1}. \quad (1.7)$$

Pour résoudre cette intégrale, il faut connaître $f(X_{n-1}|Y_{1:n-1})$. Par la définition d'une densité conditionnelle, ce terme équivaut à

$$f(X_{n-1}|Y_{1:n-1}) = \frac{f(X_{n-1}, Y_{1:n-1})}{f(Y_{1:n-1})}.$$

En utilisant les propriétés bayésiennes, cette équation se développe en

$$f(X_{n-1}|Y_{1:n-1}) = \frac{f(X_{n-1}, Y_{n-1}|Y_{1:n-2}) f(Y_{1:n-2})}{f(Y_{n-1}|Y_{1:n-2}) f(Y_{1:n-2})}.$$

En éliminant le terme commun au numérateur et au dénominateur et en développant le numérateur à l'aide du théorème de Bayes, l'équation précédente se simplifie à

$$f(X_{n-1}|Y_{1:n-1}) = \frac{f(Y_{n-1}|X_{n-1}) f(X_{n-1}|Y_{1:n-2})}{f(Y_{n-1}|Y_{1:n-2})}. \quad (1.8)$$

À partir de l'équation (1.8), il est possible de calculer $f(X_{n-1}|Y_{1:n-1})$ récursivement puisque les différents termes ont tous été définis préalablement dans le problème. Le premier terme au numérateur est défini par la représentation d'état généralisée et le second terme se calcule à l'aide de l'équation (1.7). Finalement, le terme au dénominateur est la contribution à la vraisemblance qui se calcule à partir de l'équation (1.6). Ceci permettra de calculer (1.6) et, finalement, nous permettra d'estimer d_n récursivement à l'aide de l'équation (1.5).

En supposant que les fonctions de transitions h_k et g_k sont des fonctions linéaires et que les variables aléatoires ϵ_k et ζ_k sont normalement distribuées, le dénominateur peut se calculer analytiquement et le problème se résout par un filtre de Kalman.

Toutefois, pour une grande variété de problèmes de représentation d'état, trouver

une solution analytique au problème n'est pas possible. Pour ces problèmes, il devient nécessaire de développer de nouvelles techniques afin d'estimer la distribution de la variable latente. L'idée générale du filtre à particules est de s'inspirer de la technique d'échantillonnage préférentiel et de la généraliser aux problèmes de représentation d'état.

1.2.1 Échantillonnage préférentiel

1.2.1.1 Définition

L'échantillonnage préférentiel est une technique d'approximation d'intégrale par des simulations de Monte-Carlo. Cette technique a été inventée par Kahn et Marshall (1953).

Soit une intégrale I définie comme

$$I = \int m(x)f(x) dx, \quad (1.9)$$

où $m(x)$ est une fonction quelconque d'une variable aléatoire X et $f(x)$ est la fonction de densité de la variable aléatoire X .

I peut être considérée comme l'espérance de la fonction $m(X)$ sous la densité f , c'est-à-dire que

$$I = \mathbb{E}[m(X)].$$

L'estimateur de Monte-Carlo de l'intégrale I est défini par

$$\hat{I} = \frac{1}{N} \sum_{i=1}^N m(X^{(i)}),$$

où les $X^{(i)}$ sont des variables aléatoires indépendantes distribuées selon la fonction de densité f .

Nous supposons maintenant qu'il soit difficile de simuler les variables aléatoires $X^{(i)}$

à partir de la fonction f . Dans un tel cas, l'approximation par des simulations de Monte-Carlo est beaucoup plus difficile à calculer. Il faut modifier l'intégrale (1.9) pour retrouver une fonction à partir de laquelle il est possible de simuler les $X^{(i)}$. Définissons la fonction q comme une fonction de densité ayant le même support que la densité f . q et f ont le même support si

$$f(x) > 0 \Leftrightarrow q(x) > 0.$$

Si cette condition est respectée et que f et q sont des fonctions de densités, alors

$$E^q \left[\frac{f(X)}{q(X)} \right] = 1. \quad (1.10)$$

Les densités f et q seront alors des mesures équivalentes. Une façon alternative d'écrire l'intégrale I , telle que définie à l'équation (1.9), est de multiplier et diviser la fonction à intégrer par la fonction q qui a été définie précédemment. L'intégrale I peut donc s'écrire

$$I = \int m(x) \frac{f(x)}{q(x)} q(x) dx.$$

Cette intégrale est maintenant définie comme l'espérance de la transformation m multipliée par le poids corrigé $L(x) = \frac{f(x)}{q(x)}$ sous la mesure q

$$I = E^q [m(X)L(X)].$$

Le poids corrigé est aussi fréquemment appelé ratio de vraisemblance ou dérivée de Radon-Nikodym. À l'aide du ratio de vraisemblance et de la fonction q , il devient possible d'approximer l'intégrale I par

$$\hat{I} = \frac{1}{N} \sum_{i=1}^N m(\tilde{X}^{(i)})L(\tilde{X}^{(i)}),$$

où les $\tilde{X}^{(i)}$ sont des réalisations d'une variable aléatoire générées à l'aide de la fonction de densité q .

1.2.1.2 Échantillonnage préférentiel appliqué à la représentation d'état

Pour obtenir la distribution de la variable latente conditionnellement aux observations d_n , il est possible d'utiliser un échantillonnage préférentiel. Pour ce faire, il faut simuler la variable non observée sur la période complète d'observation $X_{1:n}^{(i)}$ à l'aide d'une fonction de densité équivalente $q_{1:n} = f^*(X_{1:n}|Y_{1:n})$. La fonction de densité équivalente $q_{1:n}$ est une fonction de densité choisie arbitrairement par l'utilisateur. Elle doit respecter les critères suivants :

- Cette fonction de densité équivalente doit permettre de simuler la variable latente sur toutes les périodes simultanément.
- Cette fonction de densité équivalente doit avoir le même support que la distribution réelle que la série temporelle des variables d'état.

Idéalement, il est aussi recommandé que cette fonction équivalente soit une densité se rapprochant de la densité réelle d_n . Ceci permet d'obtenir une estimation plus pertinente de la densité en utilisant un plus petit nombre de simulations. Ces simulations de la série temporelle de la variable latente $X_{1:n}^{(i)}$ seront aussi appelées des particules. Chacune de ces particules se verra aussi attribuer un poids.

Pour obtenir un estimateur pour une transformation quelconque $m(X_{1:n})$ du processus stochastique X , il faut premièrement simuler N particules à partir de la fonction de distribution équivalente appliquée sur la période complète de l'observation. Par la suite, il faut calculer le ratio entre la fonction de densité de la variable latente et la densité équivalente :

$$\omega_n^{(i)} = L(X_{1:n}^{(i)}|Y_{1:n}) = \frac{f(X_{1:n}|Y_{1:n})}{f^*(X_{1:n}|Y_{1:n})}$$

Par les définitions de d_n et de $q_{1:n}$, ceci est équivalent à

$$\omega_n^{(i)} = \frac{d_n}{q_{1:n}}.$$

Ce ratio de vraisemblance est aussi généralement considéré comme étant le « poids »

de la particule.

Il est maintenant possible d'obtenir un estimateur pour une transformation de la série temporelle de la variable latente conditionnellement aux observations $\alpha_n = E [m(X_{1:n})]$ par

$$\alpha_n = E [m(X_{1:n})],$$

$$\widehat{\alpha}_n = \frac{1}{N} \sum_{i=1}^N m(X_{1:n}^{(i)}) \omega_n^{(i)}.$$

Puisque $\omega_n^{(i)}$ représente un ratio de mesures équivalentes, il est possible de conclure que l'espérance de $\omega_n^{(i)}$ est de 1 à cause de l'équation (1.10). Ceci permet d'approximer l'équation précédente par

$$\widehat{\alpha}_n \approx \sum_{i=1}^N m(X_{1:n}^{(i)}) \widehat{\omega}_n^{(i)},$$

où $\widehat{\omega}_n^{(i)}$ représente le poids normalisé tel que $\widehat{\omega}_n^{(i)} = \frac{\omega_n^{(i)}}{\sum_{j=1}^N \omega_n^{(j)}}$.

Un des défauts de l'échantillonnage préférentiel est que pour calculer les différents poids, il est nécessaire d'utiliser une fonction récursive. Ceci fait en sorte que les calculs sont relativement lourds et que le temps de calcul pour l'évaluation d'une distribution par échantillonnage préférentiel augmente de façon exponentielle lorsque le nombre de périodes sur lesquelles la représentation d'état est modélisée augmente.

1.2.2 Échantillonnage d'importance séquentiel

Une des propositions pour réduire les temps de calcul de l'échantillonnage préférentiel est d'utiliser l'échantillonnage d'importance séquentiel. Cette technique a été proposée par Hammersley et Morton (1954). Le principe de ce processus est de simuler les particules à la période k , $X_k^{(i)}$, à partir d'une fonction conditionnelle

dépendant de la particule à la période précédente $k - 1$. Nous dénotons donc

$$q_{k|k-1} = f^*(X_k | X_{1:k-1}, Y_{1:k}).$$

La distribution de la fonction $q_{1:n}$ est maintenant

$$\begin{aligned} q_{1:n} &= f^*(X_{1:n} | X_{1:n-1}, Y_{1:n}), \\ &= f^*(X_n | X_{1:n-1}, Y_{1:n}) f^*(X_{1:n-1} | X_{1:n-2}, Y_{1:n-1}), \\ &= q_{n|n-1} q_{1:n-1}. \end{aligned} \tag{1.11}$$

Par récursion, $q_{1:n}$ se développe comme

$$q_{1:n} = q_0 \prod_{k=1}^n q_{k|k-1} \tag{1.12}$$

où q_0 est la distribution initiale de la variable latente.

Le poids de la simulation i à la période n , $\omega_n^{(i)}$, se calcule comme étant proportionnel à $\frac{d_n^{(i)}}{q_{1:n}^{(i)}}$, c'est-à-dire que

$$\omega_n^{(i)} \propto \frac{d_n^{(i)}}{q_{1:n}^{(i)}},$$

où \propto dénote que les termes des deux côtés de l'équation sont proportionnels un à l'autre.

En développant d_n et $q_{1:n}$ à l'aide des équations (1.5) et (1.12), le poids de la particule i $\omega_n^{(i)}$ vaut

$$\omega_n^{(i)} \propto \frac{d_{n-1}^{(i)}}{q_{1:n-1}^{(i)}} \frac{f(Y_n | X_n^{(i)}) f(X_n^{(i)} | X_{n-1}^{(i)}, Y_{n-1})}{q_{n|n-1}^{(i)} f(Y_n | Y_{1:n-1})}.$$

En reconnaissant la définition du poids de la i^e particule à la période précédente, elle se simplifie à

$$\omega_n^{(i)} \propto \omega_{n-1}^{(i)} \frac{f(Y_n|X_n) f(X_n|X_{n-1})}{q_{n|n-1} f(Y_n|Y_{1:n-1})}.$$

Bien que le poids des particules ait un certain intérêt, ce poids est surtout utile afin de définir une fonction de densité conditionnelle aux observations pour la série temporelle de la variable latente. Pour créer cette fonction de densité, les poids normalisés sont plus intéressants que la vraie valeur des poids. Pour cette raison, il est possible d'éliminer les termes ne dépendant pas de la variable latente simulée $X_{1:n}^{(i)}$ puisque ces termes n'offrent aucun avantage comparatif aux poids des particules. Sachant cela, le poids de la particule i est proportionnel à

$$\begin{aligned} \omega_n^{(i)} &\propto \omega_{n-1}^{(i)} \frac{f(Y_n|X_n) f(X_n|X_{n-1})}{q_{n|n-1}}, \\ \omega_n^{(i)} &\propto \omega_{n-1}^{(i)} \tilde{\omega}_n^{(i)}, \end{aligned} \tag{1.13}$$

où $\tilde{\omega}_n^{(i)} = \frac{f(Y_n|X_n) f(X_n|X_{n-1})}{q_{n|n-1}}$ représente la contribution de la n^e période sur le poids de la particule i . Les fonctions de densité $f(Y_n|X_n)$ et $f(X_n|X_{n-1})$ sont connues dans le modèle et nous choisissons $q_{n|n-1}$ afin que la fonction $q_{1:n}$ respecte les critères que nous avons fixés préalablement. Il devient alors très facile de calculer $\tilde{\omega}_n^{(i)}$.

L'espérance de la transformation $m(X_{1:n})$ s'approxime par

$$\hat{\alpha} = \sum_{i=1}^N m(X_{1:n}^{(i)}) \hat{\omega}_n^{(i)}, \tag{1.14}$$

où $\hat{\omega}_n^{(i)}$ est le poids normalisé de la particule i à la n^e période.

En utilisant la même technique d'intégration, la distribution de la fonction de la densité de la série temporelle de la variable latente \hat{d}_n s'approxime à

$$\hat{d}_n = \sum_{i=1}^N \hat{\omega}_n^{(i)} \delta_{X_{1:n}^{(i)}}(x_{1:n}) \tag{1.15}$$

où $\delta_{x_{1:n}}$ est la fonction de Dirac qui vaut 1 lorsque la série temporelle $x_{1:n}$ est égale à $X_{1:n}^{(i)}$ et qui vaut 0 ailleurs.

De façon similaire, la fonction de densité de la variable latente à la période n est de

$$\hat{f}(X_n|Y_{1:n}) = \sum_{i=1}^N \hat{\omega}_n^{(i)} \delta_{x_n}, \quad (1.16)$$

où δ_{X_n} est une fonction de Dirac qui vaut 1 lorsque la valeur de la variable latente à la n^e période x_n correspond à celle d'une des particules $X_n^{(i)}$ et qui vaut 0 ailleurs.

Soit $\mathcal{L}_{1:n}$ la vraisemblance des observations à l'aide de l'échantillonnage d'importance séquentiel. Cette vraisemblance est définie par

$$\mathcal{L}_{1:n} = f(Y_{1:n}).$$

Par Bayes, l'équation précédente devient

$$\mathcal{L}_{1:n} = f(Y_n|Y_{1:n-1})f(Y_{1:n-1}).$$

Il est possible d'exprimer récursivement cette vraisemblance par

$$\mathcal{L}_{1:n} = \prod_{k=1}^n f(Y_k|Y_{1:k-1}).$$

Ce terme, dans le produit, a été précédemment défini et représente la contribution à la vraisemblance de la k^e période \mathcal{L}_k

$$\mathcal{L}_{1:n} = \prod_{k=1}^n \mathcal{L}_k. \quad (1.17)$$

Pour pouvoir approximer la vraisemblance, il faut d'abord être capable d'approximer la contribution à la vraisemblance de la k^e période, \mathcal{L}_k . En se rappelant que

$$\mathcal{L}_k = f(Y_k|Y_{1:k-1}).$$

Ce résultat a été démontré à l'équation (1.6) comme étant

$$\mathcal{L}_k = \int f(Y_k|X_k) f(X_k|Y_{1:k-1}) dX_k.$$

C'est l'espérance de la probabilité de transition $f(Y_k|X_k)$ par rapport à la variable aléatoire X_k conditionnellement aux observations passées $Y_{1:k-1}$. L'équation précédente se réécrit comme

$$\mathcal{L}_k = E[f(Y_k|X_k) | Y_{1:k-1}].$$

Cette probabilité de transition est définie par le modèle. Afin d'approximer cette espérance, il suffit d'utiliser l'échantillonnage d'importance séquentiel afin de simuler la distribution de la densité empirique de $f(X_k|Y_{1:k-1})$. La vraie valeur de cette densité a été décrite à l'équation (1.7). Il est assez difficile de simuler X_k directement à partir de cette équation. Par contre, une distribution empirique pour $f(X_{k-1}|Y_{k-1})$ a été définie à l'équation (1.16). Il est possible de simuler la densité $f(X_k|Y_{1:k-1})$ en simulant les X_k par la fonction de transition $f(X_k|X_{k-1}, Y_{k-1})$ ou avec la probabilité de transition équivalente $q_{k|k-1}$. La contribution à la vraisemblance de la k^e période s'approxime maintenant par

$$\hat{\mathcal{L}}_k = \sum_{i=1}^N f(Y_k|X_k^{(i)}) \frac{f(X_k^{(i)}|X_{k-1}^{(i)}, Y_{k-1})}{q_{k|k-1}^{(i)}} \hat{\omega}_{k-1}^{(i)}.$$

En reconnaissant dans l'équation précédente la contribution de la k^e période sur le poids de la particule, l'équation précédente se simplifie à

$$\hat{\mathcal{L}}_k = \sum_{i=1}^N \hat{\omega}_{k-1}^{(i)} \tilde{\omega}_k^{(i)}. \quad (1.18)$$

où $\hat{\omega}_0^{(i)} = \frac{1}{N}$.

En insérant le résultat de l'équation (1.18) dans l'équation (1.17), la vraisemblance

estimée par échantillonnage séquentiel devient

$$\hat{\mathcal{L}}_{1:n} = \prod_{k=1}^n \sum_{i=1}^N \hat{\omega}_{k-1}^{(i)} \tilde{\omega}_k^{(i)}. \quad (1.19)$$

Cette méthode pour calculer la vraisemblance est surtout utilisée lorsqu'il devient impossible de suivre le poids total des particules.

Toutefois, lorsque le poids total des particules est facile à retracer, une méthode équivalente peut être considérée. À partir de l'équation (1.19), il est possible de décomposer les poids pondérés en utilisant leur définition. De cette façon, la vraisemblance devient

$$\hat{\mathcal{L}}_{1:n} = \prod_{k=1}^n \sum_{i=1}^N \frac{\bar{\omega}_{k-1}^{(i)}}{\sum_{j=1}^N \bar{\omega}_{k-1}^{(j)}} \tilde{\omega}_k^{(i)}$$

où $\bar{\omega}_k^{(i)} = \bar{\omega}_{k-1}^{(i)} \tilde{\omega}_k^{(i)}$ et $\bar{\omega}_0^{(i)} = 1$. Notons que ceci ne change pas l'égalité puisque $\bar{\omega}_k^{(i)} \propto \omega_k^{(i)}$ tel que défini par (1.13). Avec cette définition de $\bar{\omega}_k^{(i)}$, la vraisemblance peut être définie comme

$$\hat{\mathcal{L}}_{1:n} = \prod_{k=1}^n \frac{\sum_{i=1}^N \bar{\omega}_k^{(i)}}{\sum_{j=1}^N \bar{\omega}_{k-1}^{(j)}}.$$

En décomposant le produit et en éliminant les termes se trouvant au numérateur et au dénominateur, la vraisemblance devient

$$\hat{\mathcal{L}}_{1:n} = \frac{\sum_{i=1}^N \bar{\omega}_n^{(i)}}{\sum_{j=1}^N \bar{\omega}_0^{(j)}}.$$

Comme $\bar{\omega}_0^{(i)} = 1$, la vraisemblance vaut finalement

$$\hat{\mathcal{L}}_{1:n} = \frac{1}{N} \sum_{i=1}^N \bar{\omega}_n^{(i)}. \quad (1.20)$$

Cette deuxième version de la vraisemblance peut être plus simple à calculer puisqu'il

suffit de calculer la moyenne du poids total à la dernière période pour obtenir la vraisemblance. Toutefois, il est possible que ces poids ne soient pas conservables d'une période à l'autre. Dans un tel cas, il faut alors utiliser l'équation (1.19).

L'algorithme d'échantillonnage d'importance séquentiel, présenté à l'algorithme 1, consiste à générer N particules représentant différentes trajectoires possibles de la variable non-observable à l'aide d'une fonction de densité équivalente à la densité réelle qui est recherchée $q_{k|k-1}$. Ces particules sont ensuite pondérées afin que les particules les plus crédibles est un plus grand poids que celle prédisant moins bien la fonction de densité de la variable aléatoire non observable. À l'aide de ces particules, il devient possible d'estimer la densité conditionnelle de la variable non-observable à chaque période.

Algorithme 1. *Algorithme d'échantillonnage d'importance séquentiel ou "Sequential importance sampling, SIS"*

- Initialiser X_0 en simulant N particules à l'aide de la densité q_0 .
- Pour chaque période k allant de 1 jusqu'à n , répéter les étapes suivantes :
 1. Générer aléatoirement $X_k^{(i)}$ à partir de $X_{k-1}^{(i)}$ et Y_{k-1} en utilisant la fonction de densité équivalente $q_{k|k-1}^{(i)}$ pour tous les i allant de 1 à N .
 2. Calculer les poids des particules $\bar{\omega}_k^{(i)} = \bar{\omega}_{k-1}^{(i)} \tilde{\omega}_k^{(i)}$ pour tous les i allant de 1 à N .
 3. Normaliser le poids de toutes les particules à $\hat{\omega}_k^{(i)} = \frac{\omega_k^{(i)}}{\sum_{j=1}^N \omega_k^{(j)}}$.

1.2.2.1 Dégénérescence des poids

Un des problèmes de l'échantillonnage d'importance séquentiel est que, plus le nombre de périodes augmente, plus le poids total se retrouve concentré sur un nombre réduit de particules. C'est-à-dire qu'un petit nombre de particules auront un poids normalisé nettement supérieur à $\frac{1}{N}$ alors que la grande majorité des particules auront un poids normalisé négligeable. Ceci a pour conséquence que l'estimation de la fonction de densité de la variable latente est concentrée en quelques points ayant

une forte masse de probabilité. Cette concentration des poids en quelques particules augmente considérablement la variance de la distribution puisque celle-ci est maintenant basé sur un très petit nombre de particules. Ces particules ayant une forte masse de probabilité sont celles contrôlant la densité prédite. En général, elles prédisent de façon médiocre la densité dans les dernières périodes puisque les particules les plus lourdes peuvent générer de mauvaises valeurs pour la variable non-observable sans que cela n'affecte trop leurs poids. Cela fausse toutefois la distribution réelle de la variable latente à ces périodes.

1.3 Échantillonnage d'importance séquentiel avec ré-échantillonnage

Une des solutions proposées pour améliorer la densité prédite de la variable latente est de ré-échantillonner les particules à chaque période. Ceci permet de contrer les problèmes liés à la dégénérescence des poids, d'obtenir une densité empirique plus représentative de la fonction de densité réelle et de stabiliser les résultats dans les dernières périodes. Cette solution se nomme l'échantillonnage d'importance séquentiel avec ré-échantillonnage ou le filtre à particules.

Le principe général du ré-échantillonnage est d'appliquer un échantillonnage multinomial pour choisir N particules parmi l'échantillon de particules déjà existantes. Chaque nouvelle particule est générée à partir d'un échantillonnage multinomial pondéré où chaque particule initiale i est associée à une valeur de la variable latente $X_k^{(i)}$ et à un poids pondéré de $\hat{\omega}_k^{(i)}$. Après le ré-échantillonnage, les nouvelles particules i se voient toutes attribuer un poids de $\hat{\omega}_k^{(i)*} = \frac{1}{N}$ et une nouvelle valeur de la variable latente $X_k^{(i)*}$.

Pour estimer, à l'aide du filtre à particules, la densité conditionnelle aux observations de la variable latente à la période n , X_n , ou celle de sa série temporelle $X_{1:n}$, ou pour trouver un estimateur pour une transformation de la variable aléatoire $m(X)_{1:n}$ conditionnel aux observations, il faut utiliser les mêmes estimateurs que pour l'échantillonnage d'importance séquentiel. Pour estimer ces valeurs, il faut

donc utiliser les équations (1.15), (1.16) et (1.14) respectivement.

Lorsque nous cherchons à calculer la vraisemblance de l'échantillonnage d'importance séquentiel avec ré-échantillonnage à la période n , il est possible d'utiliser le résultat (1.18) pour approximer la contribution de la période k sur la vraisemblance à

$$\begin{aligned}\mathcal{L}_k &= f(Y_k | Y_{1:k-1}), \\ \mathcal{L}_k &\approx \sum_{i=1}^N \hat{\omega}_{k-1}^{(i)*} \tilde{\omega}_k^{(i)}.\end{aligned}\tag{1.21}$$

Il faut noter à l'équation (1.21) que les poids pondérés à la période précédente sont calculés après le ré-échantillonnage. Ils sont choisis ainsi puisque la contribution à la vraisemblance de la k^e période représente la somme des contributions à la vraisemblance de chaque particule pondérées par leur poids. Comme cette vraisemblance est évaluée à la période k , leur poids à la période précédente est celui après le ré-échantillonnage puisque ces particules n'ont plus de lien direct avec les particules avant le ré-échantillonnage. Pour cette raison, le poids pondéré de chaque particule devient

$$\mathcal{L}_k \approx \sum_{i=1}^N \frac{1}{N} \tilde{\omega}_k^{(i)}.\tag{1.22}$$

Tel que mentionné précédemment, les nouvelles particules se sont toutes vues attribuer un poids de $\hat{\omega}_k^{(i)*} = \frac{1}{N}$.

Tel que démontré précédemment, la vraisemblance totale du filtre est le produit des contributions à la vraisemblance des périodes allant de 1 jusqu'à n . En utilisant le résultat de l'équation (1.22) dans l'équation (1.19), la vraisemblance du filtre à

particules devient donc

$$\begin{aligned}
\mathcal{L}_{1:n} &= f(Y_{1:n}), \\
\mathcal{L}_{1:n} &= \prod_{k=1}^n f(Y_k | Y_{1:k-1}), \\
\mathcal{L}_{1:n} &\approx \prod_{k=1}^n \sum_{i=1}^N \frac{1}{N} \tilde{\omega}_k^{(i)}. \tag{1.23}
\end{aligned}$$

L'algorithme d'échantillonnage d'importance séquentiel avec ré-échantillonnage est similaire à l'algorithme 1. L'étape de l'initialisation et les étapes 1 à 4 de l'échantillonnage restent essentiellement identiques. Il faut par contre faire un ajustement pour tenir compte du fait que les particules sont ré-échantillonnées. Il est ensuite nécessaire d'ajouter quelques étapes pour le ré-échantillonnage. L'algorithme des filtres à particules standard est présenté dans l'algorithme 2. Il faut noter ici que les valeurs initialisées sont considérées comme étant des variables obtenues après le ré-échantillonnage puisqu'elles servent uniquement de point de départ pour les calculs aux périodes subséquentes et qu'il y a peu d'intérêt à les initialiser ailleurs.

Algorithme 2. *Algorithme d'échantillonnage d'importance séquentiel avec ré-échantillonnage ISR*

- *Initialiser les particules en simulant la valeur initiale de leur variable latente $X_0^{(i)*}$ à l'aide de la densité q_0 et initialiser leur poids pondéré $\hat{\omega}_0^{(i)*}$ par $\hat{\omega}_0^{(i)*} = \frac{1}{N}$.*
- *Pour chaque période k allant de 1 jusqu'à n , répéter les étapes suivantes :*
 1. *Générer aléatoirement $X_k^{(i)}$ à partir de $X_{k-1}^{(i)*}$ et Y_{k-1} en utilisant la fonction de densité équivalente $q_{k|k-1}^{(i)}$ pour tous les i allant de 1 à N .*
 2. *Calculer les poids des particules $\tilde{\omega}_k^{(i)}$ pour tous les i allant de 1 à N .*
 3. *Normaliser le poids de toutes les particules à $\hat{\omega}_k^{(i)} = \frac{\tilde{\omega}_k^{(i)}}{\sum_{j=1}^N \tilde{\omega}_k^{(j)}}$.*
 4. *Calculer la contribution à la vraisemblance de l'observation au temps k sachant les observations précédentes $\hat{L}_k \approx \sum_{i=1}^N \frac{1}{N} \tilde{\omega}_k^{(i)}$*
 5. *Ré-échantillonner les particules en effectuant les étapes suivantes :*

- Définir une variable de poids cumulatif des particules $\phi_i = \sum_{j=1}^i \hat{\omega}_k^{(j)}$ pour i allant de 1 jusqu'à N et où $\phi_0 = 0$.
- Générer N lois uniformes U entre 0 et 1.
- Pour $j = 1, 2, \dots, N$, trouver s_j tel que s_j est le plus grand entier tel que

$$s_j = \sup_{i \in \{0, 1, \dots, N-1\}} (i + 1) \mathbf{1}_{\{\phi_i < U^{(j)}\}}.$$

- Fixer $X_k^{(j)*}$ la particule j après le ré-échantillonnage comme étant $X_k^{(j)*} = X_k^{(s_j)}$ et son poids après ré-échantillonnage à $\omega_k^{(j)*} = \frac{1}{N}$.

1.3.1 Algorithme de Bootstrap

L'algorithme de Bootstrap dans les filtres à particules est une simplification du modèle qui est fréquemment utilisé. Il est communément reconnu comme étant le premier filtre à particules à avoir été développé dans la littérature par Gordon *et al.* (1993). Le principe général de l'algorithme est que la fonction équivalente choisie pour générer les particules est

$$q_{k|k-1} = f(X_k | X_{k-1}, Y_{k-1}; \theta).$$

Ceci modifie la contribution de la n^e période sur le poids de la particule, qui devient

$$\tilde{\omega}_n^{(i)} = f(Y_n | X_n). \quad (1.24)$$

Ceci facilite grandement les calculs puisque les fonctions $f(X_k | X_{k-1}, Y_{k-1}; \theta)$ et $f(Y_n | X_n)$ sont définies par le modèle. Il devient plus facile de calculer les poids et de générer les particules.

Une des faiblesses de cette simplification est que les valeurs des variables aléatoires X et Y aux périodes précédant la période $n - 1$ ne sont pas prises en considération pour générer les particules. Ceci fait en sorte que la distribution des particules générée par la fonction q n'est possiblement pas aussi représentative de la probabilité

conditionnelle qu'elle pourrait l'être.

1.3.2 Stabilité du filtre à particules

Le ré-échantillonnage, puisqu'il est aléatoire, a pour défaut qu'il vient ajouter du bruit à la densité de la variable latente. En effet, certaines particules seront représentées plus souvent qu'elles ne le devraient et d'autres pourraient être détruites alors qu'elles sont représentatives de la densité réelle du filtre.

Soit $M_k^{(i)}$ le nombre de fois que la particule i est représentée dans l'algorithme après le ré-échantillonnage. Lors du ré-échantillonnage, la particule i a une probabilité $\hat{\omega}_k^{(i)}$ d'être ré-échantillonnée à chaque variable uniforme générée aléatoirement. $M_k^{(i)}$ suit donc une loi binomiale définie comme

$$M_k^{(i)} \sim \text{Bin}(n = N, p_k^{(i)} = \hat{\omega}_k^{(i)}).$$

$M_k^{(i)}$ a une espérance de

$$E [M_k^{(i)}] = N\hat{\omega}_k^{(i)},$$

et une variance de

$$\text{Var} [M_k^{(i)}] = N\hat{\omega}_k^{(i)}(1 - \hat{\omega}_k^{(i)}).$$

Comme N est très grand, la distribution binomiale de $M_k^{(i)}$ converge vers une loi de Poisson de paramètre $\lambda_k^{(i)} = N\hat{\omega}_k^{(i)}$.

Ces lois binomiales sont dépendantes les unes des autres et elles ont une corrélation négative. Ceci s'explique par le fait qu'à chaque tirage, une seule particule est sélectionnée. Comme $M_k^{(i)}$ représente le nombre de tirages dans lesquels la particule est sélectionnée afin d'être représentée dans l'échantillon final et comme le nombre de particules est fixe d'une période à l'autre, plus une particule est sélectionnée dans l'échantillon final, plus la probabilité qu'une autre particule soit représentée dans l'échantillon final diminue.

1.4 Amélioration du filtre d'échantillonnage d'importance séquentiel avec ré-échantillonnage (ISR)

Dans cette section, nous présentons des algorithmes réduisant le bruit causé par le caractère aléatoire du ré-échantillonnage du filtre à particules standard. Ceci lisse également la fonction de vraisemblance, ce qui en réduit le nombre de maximums locaux, utilisée pour calibrer les modèles étudiés. Ces modèles sont développés notamment pour réduire la variance du nombre de fois que la particule i soit représentée dans le filtre après le ré-échantillonnage $M_k^{(i)}$.

1.4.1 Ré-échantillonnage stratifié

Dans l'algorithme du filtre à particules avec ré-échantillonnage standard, les N particules de l'échantillon final sont sélectionnées à l'aide de distributions uniformes générées entre 0 et 1. Soit une fonction ϕ_i représentant le poids cumulé normalisé des particules. ϕ_i est défini comme

$$\phi_i = \sum_{j=1}^i \hat{\omega}_k^{(j)},$$

où $\phi_0 = 0$.

Dans l'algorithme avec ré-échantillonnage standard, une particule i est choisie dans l'algorithme final au j^e tirage si lors de ce tirage uniforme, le résultat U_j se situe entre ϕ_{i-1} et ϕ_i . Toutefois, comme chaque résultat dépend de lois uniformes générées entre 0 et 1, ceci peut mener à de larges bruits dans la représentation des particules initiales dans l'échantillon final. L'algorithme de ré-échantillonnage stratifié a été introduit pour la première fois par Kitagawa (1996) pour compenser pour cette faiblesse.

Le principe de cet algorithme est de segmenter la répartition de ces lois uniformes en N strates de taille $\frac{1}{N}$. De cette façon, une seule particule par strate est sélectionnée. Dans cet algorithme, pour que la particule i soit choisie lors du j^e tirage, il faut que

le résultat stratifié U_j^{strat} défini comme

$$U_j^{strat} = \frac{j-1}{N} + \frac{U_j}{N}$$

se retrouve entre ϕ_{i-1} et ϕ_i et où U_j est le résultat d'une loi uniforme générée aléatoirement entre 0 et 1.

Cette stratification des lois uniformes réduit le bruit créé par le ré-échantillonnage en réduisant la taille de l'intervalle sur lequel les lois uniformes sont générées. Les lois uniformes générées par le filtre à particules avec ré-échantillonnage standard sont générées entre 0 et 1. Leur variance est donc de $\frac{1}{12}$. Comme la taille de l'intervalle sur lequel les lois uniformes servant à ré-échantillonner les particules du filtre à particules avec ré-échantillonnage stratifié sont générées est de $\frac{1}{N}$, la variance de ces lois uniformes est de $\frac{1}{12N^2}$. Comme la variance de ces lois uniformes est nettement plus petite, le processus par lequel les particules sont ré-échantillonnées devient moins aléatoire.

La i^e particule est représentée au moins $m^{(i)}$ fois, où $m^{(i)}$ vaut

$$m^{(i)} = \lfloor \max(N\hat{\omega}_k^{(i)} - 1, 0) \rfloor,$$

où $\lfloor x \rfloor = \max\{j \in \mathbb{Z} | j \leq x\}$.

$m^{(i)}$ représente le nombre minimal de strates que le poids de la particule couvre complètement. Le résidu

$$v_k^{(i)} = N\hat{\omega}_k^{(i)} - m^{(i)}$$

représente la somme des proportions des strates qui ne sont pas nécessairement complètement remplies par le poids de la particule i . $v_k^{(i)}$ peut prendre des valeurs comprises entre 0 et 2. Le nombre de fois que chaque particule est représentée dans l'échantillon final $M_k^{(i)}$ dépend alors d'une ou deux lois de Bernoulli.

Un des défauts de ce type de ré-échantillonnage est que le rang de la particule dans

l'échantillon initial peut influencer la variance du nombre de fois que la particule est représentée dans l'échantillon final $M_k^{(i)}$.

La variance du nombre de fois que la particule est représentée dans l'échantillon final est minimisée lorsque sa représentation dépend d'une seule loi Bernoulli. Dans ce cas, la particule est représentée $M_k^{(i)}$ fois où $M_k^{(i)}$ vaut

$$M_k^{(i)} = \begin{cases} \lfloor N\hat{\omega}_k^{(i)} \rfloor & \text{avec une probabilité de } 1 - (N\hat{\omega}_k^{(i)} - \lfloor N\hat{\omega}_k^{(i)} \rfloor), \\ \lfloor N\hat{\omega}_k^{(i)} \rfloor + 1 & \text{avec une probabilité de } N\hat{\omega}_k^{(i)} - \lfloor N\hat{\omega}_k^{(i)} \rfloor. \end{cases}$$

Dans ce cas, la variance du nombre de fois que la particule est représentée dans l'échantillon final est de

$$\text{Var} [M_k^{(i)}] = (N\hat{\omega}_k^{(i)} - \lfloor N\hat{\omega}_k^{(i)} \rfloor)(1 - (N\hat{\omega}_k^{(i)} - \lfloor N\hat{\omega}_k^{(i)} \rfloor)).$$

La variance de $M_k^{(i)}$ est maximisée lorsque sa représentation dépend de deux lois de Bernoulli de mêmes paramètres. Dans ce cas, la particule est représentée $M_k^{(i)}$ fois, avec

$$M_k^{(i)} = \begin{cases} \lfloor \max(N\hat{\omega}_k^{(i)} - 1, 0) \rfloor & \text{avec probabilité } \left(1 - \frac{v_k^{(i)}}{2}\right)^2, \\ \lfloor \max(N\hat{\omega}_k^{(i)} - 1, 0) \rfloor + 1 & \text{avec probabilité } v_k^{(i)} \left(1 - \frac{v_k^{(i)}}{2}\right), \\ \lfloor \max(N\hat{\omega}_k^{(i)} - 1, 0) \rfloor + 2 & \text{avec probabilité } \left(\frac{v_k^{(i)}}{2}\right)^2. \end{cases}$$

Dans ce cas, la variance du nombre de fois que la particule est représentée dans l'échantillon final est de

$$\text{Var} [M_k^{(i)}] = v_k^{(i)} \left(1 - \frac{v_k^{(i)}}{2}\right).$$

La différence entre ces deux variances peut aller jusqu'à 0.5 dans le cas où $v_k^{(i)} = 1$.

Dans tous les cas, cette variance est fortement réduite, car elle ne peut excéder 0.5.

L'espérance du nombre de fois que la particule i est représentée dans l'échantillon final $E \left[M_k^{(i)} \right]$ demeure inchangée par rapport au filtre à particules standard.

Le filtre à particules avec ré-échantillonnage stratifié tend aussi à être l'algorithme le plus lent de ceux présentés dans ce mémoire. Une autre caractéristique de ce ré-échantillonnage est qu'il augmente fortement la corrélation négative du nombre de fois que deux particules côte à côte seront représentées dans l'échantillon puisqu'elles auront toujours une partie de leur poids dans la même strate commune. Toutefois, la corrélation entre deux particules très éloignées les unes des autres peut être considérée comme nulle puisque ces particules ne partageront jamais la même strate.

L'algorithme d'importance séquentiel avec ré-échantillonnage stratifié est similaire à l'algorithme 2. La seule différence se trouve à l'étape 5 qui est remplacée par l'algorithme 3 ci-dessous.

Algorithme 3. *Algorithme d'échantillonnage d'importance séquentiel avec ré-échantillonnage stratifié*

5. Ré-échantillonner en effectuant les étapes suivantes :

- Définir une variable de poids cumulatif des particules $\phi_i = \sum_{j=1}^i \hat{\omega}_k^{(j)}$ pour i allant de 1 jusqu'à N et où $\phi_0 = 0$.
- Générer N lois uniformes U entre 0 et 1.
- Générer N lois uniformes stratifiées U_{str} à l'aide du vecteur de loi uniforme U . Pour $j = 1, 2, \dots, N$, la j^e loi uniforme stratifiée est de

$$U_{str}^{(j)} = \frac{U^{(j)} + j - 1}{N}$$

- Pour $j = 1, 2, \dots, N$, trouver s_j tel que s_j est le plus grand entier tel que

$$s_j = \sup_{i \in \{0, 1, \dots, N-1\}} (i + 1) \mathbb{1}_{\{\phi_i < U_{str}^{(j)}\}}.$$

- Fixer $X_k^{(j)*}$ la particule j après le ré-échantillonnage comme étant $X_k^{(j)*} = X_k^{(s_j)}$ et son poids après ré-échantillonnage à $\omega_k^{(j)*} = \frac{1}{N}$.

1.4.2 Ré-échantillonnage résiduel

L'algorithme de ré-échantillonnage résiduel a été introduit par Liu et Chen (1998). Le principe de l'algorithme est de s'assurer que toutes les particules ayant un poids pondéré supérieur à $\frac{1}{N}$ soient représentées au moins $\lfloor N\hat{\omega}_k^{(i)} \rfloor$ fois dans l'échantillon au temps k de façon déterministe. L'idée derrière ce ré-échantillonnage est de réduire le bruit aléatoire en assignant certaines particules de façon déterministe dans l'échantillon final. Les résidus $\nu_k^{(i)} = \hat{\omega}_k^{(i)} - \frac{\lfloor N\hat{\omega}_k^{(i)} \rfloor}{N}$ sont ensuite pondérés. Par la suite, un échantillon de $R = N - \sum_{i=1}^N \lfloor N\hat{\omega}_k^{(i)} \rfloor$ particules sur les résidus pondérés $\hat{\nu}_k^{(i)} = \frac{\nu_k^{(i)}}{\sum_{j=1}^N \nu_k^{(j)}}$ est simulé pour que l'échantillon final soit toujours composé de N particules. L'algorithme de ré-échantillonnage résiduel tend aussi à améliorer la vitesse des calculs puisque l'échantillonnage final s'effectue sur un nombre de particules réduit et il s'agit de la partie de l'algorithme prenant le plus de temps.

Dans cet algorithme, la particule est représentée $M_k^{(i)}$ fois, avec

$$M_k^{(i)} = \lfloor N\hat{\omega}_k^{(i)} \rfloor + \psi_k^{(i)},$$

où

$$\psi_k^{(i)} \sim \text{Bin}(n = R, p_k^{(i)} = \hat{\nu}_k^{(i)}).$$

Comme R est très grand, la distribution binomiale de $\psi_k^{(i)}$ converge vers une loi de Poisson de paramètre $\lambda_k^{(i)} = R\hat{\nu}_k^{(i)}$.

La variance du nombre de fois que la particule i est représentée dans l'algorithme final est

$$\text{Var} [M_k^{(i)}] = R\hat{\nu}_k^{(i)}(1 + \hat{\nu}_k^{(i)}).$$

Pour le filtre à particules avec ré-échantillonnage résiduel, la variance du nombre de

fois que la particule i est représentée dans l'échantillon final diminue très peu lorsque le poids pondéré de la particule est inférieur à $\frac{1}{N}$. Toutefois, une nette réduction de cette variance s'observe lorsque le poids pondéré est supérieur à $\frac{1}{N}$ puisque cette variance est diminuée d'environ $\lfloor N\hat{\omega}_k^{(i)} \rfloor$. Cette variance est toutefois toujours supérieure à celle pour les filtres à particules avec ré-échantillonnage stratifié. L'espérance du nombre de fois que la particule i est représentée dans l'algorithme final demeure inchangée.

L'algorithme d'importance séquentiel avec ré-échantillonnage résiduel est similaire à l'algorithme 2. Il suffit de remplacer l'étape 5 du filtre à particules standard avec les étapes 5 et 6 du filtre à particules avec ré-échantillonnage résiduel.

Algorithme 4. *Algorithme d'échantillonnage d'importance séquentiel avec ré-échantillonnage résiduel*

5. *Ré-échantillonner systématiquement certaines particules en effectuant les étapes suivantes :*

- *Définir un compteur du nombre de particules ayant été ré-échantillonnées de façon déterministe $K = 0$.*
- *Pour tout i allant de 1 jusqu'à N , compter m le nombre de fois que la particule doit être ré-échantillonnée de façon déterministe.*

$$m^{(i)} = \lfloor N\hat{\omega}_k^{(i)} \rfloor$$

- *Pour i allant de 1 jusqu'à N , si $m^{(i)}$ est supérieur à 0, fixer $X_k^{(K+f)*}$ la valeur de la variable latente pour la particules $K+f$ après le ré-échantillonnage comme étant $X_k^{(K+f)*} = X_k^{(i)}$ et son poids après ré-échantillonnage à $\omega_k^{(K+f)*} = \frac{1}{N}$, pour tout f où f est un nombre naturel tel que*

$$1 \leq f \leq m^{(i)}$$

et incrémenter le compteur $K = K + m^{(i)}$.

6. Ré-échantillonner aléatoirement les particules restantes en effectuant les étapes suivantes :

— Pour $i = 1, 2, \dots, N$, définir des poids résiduels $\nu_k^{(i)}$

$$\nu_k^{(i)} = \hat{\omega}_k^{(i)} - \frac{\lfloor N\hat{\omega}_k^{(i)} \rfloor}{N}$$

et normaliser ces poids $\hat{\nu}_k^{(i)}$

$$\hat{\nu}_k^{(i)} = \frac{\nu_k^{(i)}}{\sum_{j=1}^N \nu_k^{(j)}}.$$

— Définir une variable de poids cumulatif des particules $\phi_i = \sum_{j=1}^i \hat{\nu}_k^{(j)}$ pour i allant de 1 jusqu'à N et où $\phi_0 = 0$.

— Générer R lois uniformes U entre 0 et 1.

— Pour $j = 1, 2, \dots, R$, trouver s_j tel que s_j est le plus grand entier tel que

$$s_j = \sup_{i \in \{0, 1, \dots, N-1\}} (i + 1) \mathbb{1}_{\{\phi_i < U^{(j)}\}}.$$

— Fixer $X_k^{(K+j)*}$ la valeur de la variable latente pour la particule $K + j$ après le ré-échantillonnage comme étant $X_k^{(K+j)*} = X_k^{(s_j)}$ et son poids après ré-échantillonnage à $\omega_k^{(i)*} = \frac{1}{N}$

1.4.3 Ré-échantillonnage combiné

Dans cette section, nous présentons le ré-échantillonnage combiné. Cette méthode a été proposée par Douc et Cappe (2005). Le principe de cet algorithme est de combiner les caractéristiques de l'algorithme de ré-échantillonnage résiduel et de l'algorithme de ré-échantillonnage stratifié. Sachant cela, les particules sont assurées d'être représentées au moins $\lfloor N\hat{\omega}_k^{(i)} \rfloor$ fois. Un ré-échantillonnage stratifié est ensuite effectué pour générer les R particules restantes à l'aide de R lois uniformes stratifiées appliquées sur les résidus pondérés.

Bien que certaines des faiblesses du ré-échantillonnage stratifié demeurent présentes, l'algorithme combiné permet de nettement améliorer la vitesse d'exécution par rapport au filtre à particules stratifié tout en conservant la variance des $M_k^{(i)}$ très près de celle du ré-échantillonnage stratifié. Avec ce ré-échantillonnage, le nombre minimal de fois que la particule i est représentée $m^{(i)}$ est

$$m^{(i)} = \lfloor N\hat{\omega}_k^{(i)} \rfloor$$

comme pour le ré-échantillonnage stratifié.

Le résidu

$$\nu_k^{(i)} = N\hat{\omega}_k^{(i)} - \lfloor N\hat{\omega}_k^{(i)} \rfloor$$

représente la somme des proportions des strates qui ne sont pas nécessairement complètement remplies par le poids de la particule i . $\nu_k^{(i)}$ peut prendre des valeurs comprises entre 0 et 2. Le nombre de fois que chaque particule est représentée dans l'échantillon final $M_k^{(i)}$ dépend alors d'une ou deux lois de Bernoulli.

La variance du nombre de fois que la particule est représentée dans l'échantillon final est minimisée lorsque sa représentation ne dépend que d'une loi Bernoulli. Dans ce cas, la particule est représentée $M_k^{(i)}$ fois, avec

$$M_k^{(i)} = \begin{cases} \lfloor N\hat{\omega}_k^{(i)} \rfloor & \text{avec une probabilité de } 1 - \nu_k^{(i)}, \\ \lfloor N\hat{\omega}_k^{(i)} \rfloor + 1 & \text{avec une probabilité de } \nu_k^{(i)}. \end{cases}$$

La variance de $M_k^{(i)}$ est alors de

$$\text{Var} [M_k^{(i)}] = \nu_k^{(i)}(1 - \nu_k^{(i)}).$$

La variance de $M_k^{(i)}$ est maximisée lorsque sa représentation dépend de loi Bernoulli

de mêmes paramètres. Dans ce cas, la particule est représentée $M_k^{(i)}$ fois, avec

$$M_k^{(i)} = \begin{cases} \lfloor N\hat{\omega}_k^{(i)} \rfloor & \text{avec probabilité } \left(1 - \frac{\nu_k^{(i)}}{2}\right)^2, \\ \lfloor N\hat{\omega}_k^{(i)} \rfloor + 1 & \text{avec probabilité } \nu_k^{(i)} \left(1 - \frac{\nu_k^{(i)}}{2}\right), \\ \lfloor N\hat{\omega}_k^{(i)} \rfloor + 2 & \text{avec probabilité } \left(\frac{\nu_k^{(i)}}{2}\right)^2. \end{cases}$$

La variance de $M_k^{(i)}$ est alors de

$$\text{Var} [M_k^{(i)}] = \nu_k^{(i)} \left(1 - \frac{\nu_k^{(i)}}{2}\right).$$

La différence entre ces deux variances peut aller jusqu'à 0.5 dans le cas où $\nu_k^{(i)} = 1$. La différence majeure entre la variance de $M_k^{(i)}$ pour les filtres à particules avec ré-échantillonnage stratifié et combiné est que le résidu est limité à 1 dans le filtre avec ré-échantillonnage combiné alors qu'il est limité à 2 pour les filtres avec ré-échantillonnage stratifié. Cela fait en sorte que les particules ayant un résidu supérieur à 1 dans le filtre à ré-échantillonnage (toutes les particules ayant un poids pondéré supérieur à $\frac{1}{N}$) ont un résidu réduit de 1 dans le filtre à particules avec un ré-échantillonnage combiné. Ce changement n'a pas d'impact sur les particules ayant un poids pondéré inférieur à $\frac{1}{N}$. Toutefois, pour les particules dont le poids est supérieur à $\frac{1}{N}$, la variance du nombre de fois que la particule est représentée dans l'échantillon final sera réduite sous le filtre à particules avec ré-échantillonnage combiné pour les particules dont le résidu est inférieur à 0.5. Inversement, si le résidu est supérieur à 0.5, cette variance est augmentée sous le filtre à particules avec ré-échantillonnage combiné. Comme la majorité des particules ont des poids pondérés s'approchant de $\frac{1}{N}$, le filtre à particules avec ré-échantillonnage combiné tend à réduire la variance des $M_k^{(i)}$. $E [M_k^{(i)}]$ demeure inchangée par rapport au filtre à particules standard.

L'algorithme de ré-échantillonnage combiné est très similaire à l'algorithme 4. La seule différence est que la sixième étape est ajustée pour que les lois uniformes soient

stratifiées. Cette modification de l'algorithme des filtres à particules est présentée dans l'algorithme 5.

Algorithme 5. *Algorithme d'échantillonnage d'importance séquentiel avec ré-échantillonnage combiné*

6. Ré-échantillonner aléatoirement les particules restantes en effectuant les étapes suivantes :

— Définir des poids résiduels $\nu_k^{(i)}$

$$\nu_k^{(i)} = \hat{\omega}_k^{(i)} - \frac{\lfloor N\hat{\omega}_k^{(i)} \rfloor}{N}$$

et normaliser ces poids $\hat{\nu}_k^{(i)}$

$$\hat{\nu}_k^{(i)} = \frac{\nu_k^{(i)}}{\sum_{j=1}^N \nu_k^{(j)}}.$$

— Définir une variable de poids cumulatif des particules $\phi_i = \sum_{j=1}^i \hat{\nu}_k^{(j)}$ pour i allant de 1 jusqu'à N et où $\phi_0 = 0$.

— Générer R lois uniformes U entre 0 et 1.

— Générer R lois uniformes stratifiées U_{str} . Pour $j = 1, 2, \dots, R$, la j^e loi uniforme stratifiée est de

$$U_{str}^{(j)} = \frac{U^{(j)} + j - 1}{R}$$

— Pour $j = 1, 2, \dots, R$, trouver s_j tel que s_j est le plus grand entier tel que

$$s_j = \sup_{i \in \{0, 1, \dots, N-1\}} (i + 1) \mathbf{1}_{\{\phi_i < U^{(j)}\}}.$$

— Fixer $X_k^{(K+j)*}$ la valeur de la variable latente pour la particule $K + j$ après le ré-échantillonnage comme étant $X_k^{(K+j)*} = X_k^{(s_j)}$ et son poids après ré-échantillonnage à $\omega_k^{(i)*} = \frac{1}{N}$

CHAPITRE II

ALGORITHMES AVEC BRANCHEMENT

Dans la section précédente, nous avons introduit les filtres à particules et nous avons présenté une classe de filtre à particules où toutes les particules sont ré-échantillonnées à chaque période. Dans cette section, nous introduisons une nouvelle classe de filtre à particules : les filtres à particules avec branchement. Ces filtres ont pour caractéristique que les particules ne sont plus nécessairement toutes ré-échantillonnées. À la place, un branchement est effectué. Ce branchement peut être décrit comme la sélection de certaines particules qui restent identiques et conservent leur poids. Les autres sont ré-échantillonnées à l'aide d'une composante déterministe et d'une loi de Bernoulli. Ceci fait en sorte que le nombre de particules peut varier d'une période à l'autre. Ce branchement a l'effet d'un ré-échantillonnage partiel. Ceci permet une meilleure propagation des poids tout en s'assurant d'éviter la dégénérescence des poids.

2.1 Principe des filtres à particules avec branchement

Les filtres à particules avec branchement présentés dans cette section ont été introduits par Kouritzin (2016). Ce concept n'est pas nouveau en soi. Différents filtres avec branchement ont préalablement été proposés. Les premiers remontent à Crisan et Lyons (1997). Le poids moyen des particules à la période k sera noté A_k . Le principe général de ce filtre à particules est que, si les particules ont un poids suffisamment près du poids moyen, la particule est conservée dans l'algorithme avec son

poids. Sinon, la particule i est conservée $\left\lfloor \frac{\bar{\omega}_k^{(i)}}{A_k} \right\rfloor$ fois avec un poids de A_k et une loi de Bernoulli est utilisée pour déterminer si une copie additionnelle de cette particule, à laquelle est assignée un poids de A_k , est créée. Dans cet algorithme, le nombre de particules incluses dans l'échantillon final dépend du résultat de plusieurs lois de Bernoulli. Il est donc possible que le nombre de particules ne soit pas constant d'une période à l'autre.

Soit N_k le nombre de particules à la période k et N le nombre initial de particules, le poids moyen des particules à la période k , A_k , est défini comme

$$A_k = \frac{1}{N} \sum_{i=1}^{N_k} \bar{\omega}_k^{(i)}. \quad (2.1)$$

Il peut être remarqué que la somme des poids des particules est divisée par N pour nous assurer que le nombre de particules reste relativement stable. Ceci permet aussi qu'une particule ayant le poids moyen représente toujours la même proportion du poids total.

Dans les algorithmes avec branchement, les particules dont le poids est assez proche du poids moyen sont propagées à la période suivante. Une particule ayant un poids trop éloigné de ce poids moyen est séparée en de multiples branches auxquelles est attribué le poids moyen. Soit un intervalle c_k défini autour du poids moyen; une particule est conservée si son poids se retrouve dans l'intervalle.

L'intervalle c_k est déterminé de façon à ce que les bornes supérieures et inférieures de l'intervalle soient un multiple du poids moyen. L'intervalle c_k est donc défini par

$$c_k \in \left(c_k^{min} A_k, c_k^{max} A_k \right), \quad (2.2)$$

où le multiple de la borne inférieure de l'intervalle c_k^{min} est inférieure à 1 et le multiple de la borne supérieure de l'intervalle c_k^{max} est supérieure à 1.

Les particules ayant un poids $\bar{\omega}_k^{(i)}$ se trouvant à l'intérieur de l'intervalle c_k sont conservées et leur poids est préservé. Pour les particules ayant un poids hors de cet

intervalle, la particule i est représentée dans l'échantillon de la période suivante $M_k^{(i)}$ fois. La valeur de $M_k^{(i)}$ est définie par

$$M_k^{(i)} = \left\lfloor \frac{\bar{\omega}_k^{(i)}}{A_k} \right\rfloor + \rho_k^{(i)}, \quad (2.3)$$

où $\rho_k^{(i)}$ suit une loi de Bernoulli de paramètre $\frac{\bar{\omega}_k^{(i)}}{A_k} - \left\lfloor \frac{\bar{\omega}_k^{(i)}}{A_k} \right\rfloor$. Le poids attribué à ces nouvelles particules est le poids moyen $\bar{\omega}_k^{(i)} = A_k$.

Dans l'exemple avec branchement présenté à la figure 2.1, les particules en vert et en rouge sont à l'extérieur de l'intervalle. Le poids accordé aux particules créées à partir de celles-ci est le poids moyen. Lorsque le poids est à l'intérieur de l'intervalle, le poids reste constant.

Il existe deux cas extrêmes à cet algorithme. Si $c_k^{min} = c_k^{max} = 1$, l'algorithme devient un algorithme avec branchement systématique. Aucune particule n'est dans l'intervalle c_k . Toutes les particules sont donc ré-échantillonnées et sont représentées $M_k^{(i)}$ fois où $M_k^{(i)}$ est défini par l'équation (2.3). Le deuxième cas extrême est celui où $c_k^{min} = 0$ et $c_k^{max} = \infty$. Dans ce cas, les particules sont toujours dans l'intervalle et ne sont jamais ré-échantillonnées. Cet algorithme est l'échantillonnage d'importance séquentiel présenté dans la sous-section 1.2.2.

Dans cet algorithme, la variance de $M_k^{(i)}$ est fortement réduite par rapport aux filtres à particules avec ré-échantillonnage puisque la variabilité dépend seulement d'une loi de Bernoulli. Toutefois, l'algorithme a la faiblesse que, comme le poids des particules n'est pas constant et que le nombre de particules varie, le poids total des particules peut varier après le ré-échantillonnage partiel.

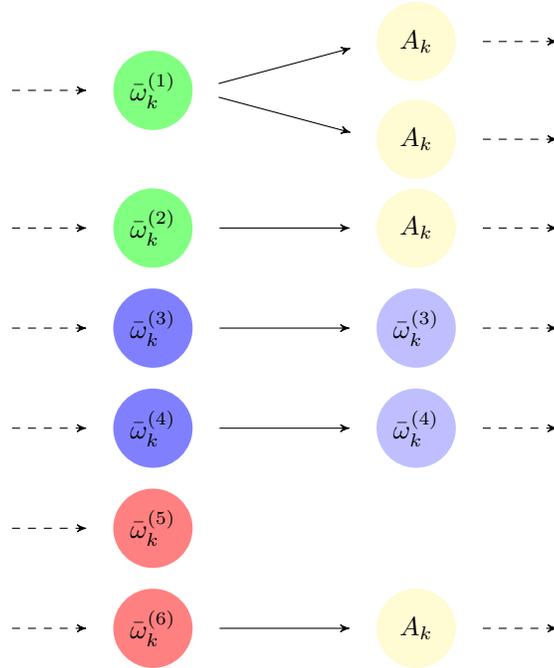


Figure 2.1: Exemple de branchement

Notre objectif lorsque nous calibrons les bornes de l'intervalle c_k est de s'assurer que le nombre de particules reste à peu près constant, c'est-à-dire que

$$E[N_k] \approx N.$$

Il est assez difficile de trouver des intervalles c_k tels que cette condition soit parfaitement respectée. En effet, cette calibration requiert un ajustement particulier à chaque modèle et il est nécessaire de connaître la densité de la variable aléatoire X_k , que nous cherchons, afin de pouvoir établir une relation adéquate entre c_k^{max} et c_k^{min} . Dans ce chapitre et les suivants, c_k^{max} est défini de façon à avoir $c_k^{max} = \frac{1}{c_k^{min}}$. Selon nos observations et celles de Kouritzin (2016), cet intervalle dit géométrique est un des intervalles offrant les résultats les plus concluants pour obtenir un intervalle qui stabilise le nombre de particules dans le filtre. Dans l'algorithme avec branchement de base, les multiplicateurs de la borne supérieure et inférieure de l'intervalle sont constants d'une période à l'autre. Sous cet algorithme, pour tout k , $c_k^{max} = c$ et $c_k^{min} = \frac{1}{c}$, avec $c \geq 1$.

Algorithme 6. *Algorithme d'échantillonnage d'importance séquentiel avec branchement*

- Initialiser X_0 en simulant N particules à l'aide de la densité q_0 .
- Pour chaque période k allant de 1 jusqu'à n , répéter les étapes suivantes :
 1. Générer aléatoirement $X_k^{(i)}$ à partir de $X_{k-1}^{(i)}$ et Y_{k-1} en utilisant la fonction de densité équivalente $q_{k|k-1}^{(i)}$ pour tous les i allant de 1 à N .
 2. Calculer les poids des particules $\bar{\omega}_k^{(i)} = \bar{\omega}_{k-1}^{(i)} \tilde{\omega}_k^{(i)}$ pour tous les i allant de 1 à N .
 3. Calculer la vraisemblance des observations précédentes $\hat{\mathcal{L}}_{1:k} \approx A_k$.
 4. Calculer R le nombre de particules à l'extérieur de l'intervalle c_k .

$$R = N_k - \sum_{i=1}^{N_k} \mathbf{1}_{\{\bar{\omega}_k^{(i)} \in c_k\}}.$$

5. Générer R variables aléatoires à partir d'une loi uniforme U .
6. Pour $i = 1, 2, \dots, N_k$, effectuer les étapes suivantes :
 - Initialiser N_{k+1} et j à 0.
 - Si le poids de la particule i est dans l'intervalle, $\bar{\omega}_k^{(i)} \in c_k$, suivre les étapes suivantes :
 - Ajouter une particule à la période suivante $N_{k+1} = N_{k+1} + 1$.
 - Accorder à cette particule le poids $\bar{\omega}_k^{(N_{k+1})*}$ de la particule i

$$\bar{\omega}_k^{(N_{k+1})*} = \bar{\omega}_k^{(i)}$$

- Accorder à la valeur de la variable latente de cette particule $\hat{X}_k^{(N_{k+1})}$ la valeur de la variable latente de la particule i .

$$X_k^{(N_{k+1})*} = X_k^{(i)}$$

- *Sinon, suivre les étapes suivantes :*
- *Calculer $M_k^{(i)}$ le nombre de fois que la particule devrait être représentée dans l'échantillon final. $M_k^{(i)}$ vaut*

$$M_k^{(i)} = \left\lfloor \frac{\omega_k^{(i)}}{A_k} \right\rfloor + p_j,$$

où p_j vaut 1 si

$$U^{(j)} < \frac{\omega_k^{(i)}}{A_k} - \left\lfloor \frac{\omega_k^{(i)}}{A_k} \right\rfloor$$

- *Si $M_k^{(i)}$ est supérieur à 0, pour tout j entre N_{k+1} exclusivement et $N_{k+1} + M_k^{(i)}$ inclusivement*
- *Accorder à la particule $\omega_k^{(j)}$ le poids moyen.*

$$\omega_k^{(j)} = A_k$$

- *Accorder à la valeur de la variable latente de cette particule $X_k^{(j)*}$ la valeur de la variable latente de la particule i .*

$$X_k^{(j)*} = X_k^{(i)}$$

- *Incrémenter N_{k+1} à $N_{k+1} = N_{k+1} + M_k^{(i)}$*

2.1.1 Sélection de modèle

Pour calculer la vraisemblance du modèle du filtre à particules avec branchement, il est possible d'utiliser les résultats de l'équation (1.20). Il faut seulement l'adapter pour qu'il s'ajuste au nombre de particules variables dans les filtres à particules avec branchement. En se rappelant que

$$L_{1:n} = f(Y_{1:n}),$$

$$\hat{\mathcal{L}}_{1:n} = \frac{1}{N} \sum_{i=1}^{N_k} \bar{\omega}_k^{(i)},$$

Il est possible de reconnaître la définition du poids moyen. Il s'en conclut donc que $\hat{\mathcal{L}}_{1:n} = A_n$.

(2.4)

2.2 Modification de l'algorithme avec branchement

Dans cette section, nous allons présenter des modifications à l'algorithme avec branchement proposées par Kouritzin (2016) permettant de réduire certaines faiblesses de l'algorithme avec branchement.

2.2.1 Algorithme avec branchement combiné

Sous les algorithmes avec branchement, le nombre de particules n'est pas stable d'une période à l'autre. Ceci peut amener une forte variabilité dans les résultats du filtre. Une façon de modifier l'algorithme de base afin de contrôler le nombre de particules est d'utiliser l'algorithme avec branchement combiné.

Il s'agit essentiellement du même algorithme. Il suffit de remplacer la cinquième étape par des lois uniformes stratifiées et randomisées par la suite. Ceci permet que le ré-échantillonnage partiel soit plus représentatif de l'échantillon initial. L'algorithme vient aussi réduire les chances que le poids total après le ré-échantillonnage partiel diffère de celui de l'échantillon initial. Ceci permet aussi de contrôler le nombre de particules.

L'algorithme avec branchement combiné a les mêmes étapes que l'algorithme 6. Il faut seulement remplacer l'étape 5 par celle de l'algorithme 7.

Algorithme 7. *Algorithme d'échantillonnage d'importance séquentiel avec branchement combiné*

5. Utiliser les étapes suivantes pour générer un vecteur U comprenant R lois

uniformes stratifiées et randomisées.

- Générer R loi uniforme $U_b^{(j)}$
- Stratifier les R lois uniformes stratifiées $U_{str}^{(j)}$

$$U_{str}^{(j)} = \frac{U_b^{(j)} + j - 1}{R}$$

- Définir $v(j)$ une permutation aléatoire des entiers entre 1 et R .
- Définir le vecteur U comme

$$U^{(j)} = U_{str}^{(v(j))}$$

2.2.2 Algorithme avec branchement effectif

Un autre des problèmes de l'algorithme avec branchement standard est que les multiplicateurs c_k^{max} et c_k^{min} sont constants dans le temps. Lorsque la variance dans le poids des particules est faible, les particules sont presque toutes incluses dans l'intervalle lorsque le multiplicateur est constant. Inversement, lorsque la variance dans le poids est forte, une très forte proportion des particules se retrouve à l'extérieur de l'intervalle. Ceci implique qu'il y a une forte perte d'information lorsque la volatilité du poids des particules est forte. La variabilité de l'échantillon final devient ainsi fortement accrue puisqu'il dépend d'un très grand nombre de lois binomiales si la volatilité du poids est forte.

Une solution proposée par Kouritzin (2016) pour résoudre ce problème est de faire varier le multiplicateur de l'intervalle d'une période à l'autre. Le principe est d'utiliser un multiplicateur qui augmente lorsque la variance des poids est élevée et utiliser un multiplicateur plus petit lorsque la variance des poids est faible. Ceci permet de mieux stabiliser le nombre de particules ré-échantillonnées à chaque période. Ainsi, même lorsque la variance est forte, un plus grand nombre de particules sont transférées telles quelles, ce qui leur donne la possibilité de «survivre» sans ré-échantillonnage partiel.

Dans l'algorithme proposé par Kouritzin (2016) pour résoudre cette problématique, nous calculons d'abord le nombre de particules effectives N_{eff} . Il s'agit d'une mesure qui calcule le niveau de dégénérescence des poids de l'échantillon. N_{eff} représente un nombre de particules autour desquelles se concentre la majeure proportion de la masse totale. Cet estimateur permet d'analyser le niveau de dégénérescence des poids. Cette mesure a été introduite par Kong *et al.* (1994). En adaptant cet estimateur aux filtres à particules avec branchement, le nombre de particules effectives peut être estimé par

$$N_k^{eff} = \frac{N^2 \hat{L}_k^2}{\sum_{i=1}^{N_k} (\omega_k^{(i)})^2}, \quad N_k^{neff} = N_k - N_k^{eff}. \quad (2.5)$$

Il est supposé que toutes les particules qui ne sont pas des particules effectives sont des particules non effectives N_k^{neff} .

c^{eff} le multiplicateur effectif représente le multiplicateur utilisé si le nombre de particules effectives est $N_k^{eff} = N_k$. C'est-à-dire si le poids des particules est parfaitement réparti entre chaque particule. Inversement, c^{neff} , le multiplicateur non effectif, représente le multiplicateur qui sera utilisé si le poids est essentiellement concentré en une seule particule. Le multiplicateur c_k^m que nous utiliserons pour l'algorithme avec branchement effectif sera calculé de façon à accorder un poids proportionnel au multiplicateur effectif et au multiplicateur non effectif en fonction de la proportion de particules effectives et non effectives sur le nombre de particules totales.

$$c_k^m = c^{eff} \frac{N_k^{eff}}{N_k} + c^{neff} \frac{N_k^{neff}}{N_k}$$

Comme $N_k^{neff} = N_k - N_k^{eff}$, ceci peut se simplifier par

$$c_k^m = c^{neff} + (c^{eff} - c^{neff}) \frac{N_k^{eff}}{N_k}. \quad (2.6)$$

Dans ce contexte, si nous voulons nous assurer de réduire le nombre de particules à ré-échantillonner lorsque la variance des poids augmente, il faut s'assurer que c^{neff}

soit supérieur à c^{eff} . De cette façon, les particules qui sont plus éloignées du poids moyen sont conservées plus souvent lorsque le poids est surtout concentré en un petit nombre de particules ce qui impliquerait une forte variance des poids.

Dans cet algorithme, toutes les étapes de l'algorithme 6 sont conservées sauf l'étape 5 qui est remplacée par l'étape 5 de l'algorithme 7. Il faut par la suite ajouter une étape entre la troisième et la quatrième étape. Les autres en sont décalées d'une étape.

Algorithme 8. *Algorithme d'échantillonnage d'importance séquentiel avec branchement effectif*

4. Calculer le multiplicateur c_k à l'aide des étapes suivantes :
 - Calculer le nombre de particules effectives à l'aide de la formule (2.5).
 - Calculer le multiplicateur c_k à l'aide de la formule (2.6).

CHAPITRE III

MODÈLE DE HESTON

Dans ce chapitre, nous décrivons le modèle de Heston ainsi que deux approximations qui nous seront utiles afin de discrétiser le modèle. Nous expliquons par la suite comment il est possible d'utiliser les filtres à particules afin d'estimer la distribution de la volatilité stochastique ou afin de trouver les paramètres du modèle.

3.1 Description du modèle

Dans le modèle de Heston, la variance stochastique est caractérisée comme un processus Cox, Ingersoll, Ross de Cox *et al.* (1985). Le processus de Cox, Ingersoll, Ross est un processus racine carrée, c'est-à-dire qu'il contient une dérive qui revient vers la moyenne et un coefficient de diffusion dépendant de la racine du processus.

$$dV_t = \kappa (\theta - V_t) dt + \sigma \sqrt{V_t} dB_t, \quad V_0 = v_0 > 0 \quad (3.1)$$

Dans cette équation, V_t représente la variance au temps t , B est un mouvement brownien, κ représente la vitesse du retour à la moyenne, θ représente la volatilité vers laquelle le processus tend à toujours revenir et σ est l'amplitude de la variance stochastique.

Il faut noter que la variance stochastique est une variable aléatoire qui ne devrait être négative. La dynamique de la volatilité stochastique est telle que la volatilité stochastique n'est jamais négative. Pour s'assurer que la volatilité stochastique ne

soit jamais nulle, la condition de Feller (1951) doit être ajoutée. Cette condition stipule que, pour éviter d'avoir une variance nulle, il faut que les paramètres respectent l'équation

$$2\kappa\theta > \sigma^2. \quad (3.2)$$

Le modèle peut être défini même lorsque cette condition n'est pas respectée. Toutefois, il peut devenir plus difficile de simuler le modèle lorsqu'elle ne l'est pas (voir Bégin (2012) pour plus de détails sur le sujet). Pour cette raison, nous allons nous restreindre aux scénarios où cette contrainte est respectée dans le cadre de ce mémoire.

Dans le modèle de Heston, cette variance stochastique est, par la suite, utilisée pour déterminer le prix du titre au temps t en utilisant l'équation stochastique suivante

$$dS_t = \mu S_t dt + \sqrt{V_t} S_t dZ_t, \quad S_0 = s_0 > 0, \quad (3.3)$$

où μ est le taux de rendement et Z est un mouvement brownien.

Dans ce modèle, il est fréquent d'ajouter une corrélation entre les mouvements browniens Z et B . En pratique, il est fréquent d'observer un effet de levier entre le rendement de l'action et sa variance. L'effet de levier signifie que lorsque la variance d'un titre augmente, son rendement moyen tend à diminuer. Inversement, lorsque la variance d'un titre diminue, son taux de rendement tend à augmenter.

Ceci peut se modéliser par une corrélation négative entre les mouvements browniens B et Z . Soit ρ la corrélation entre Z et B .

$$d\langle B_t, Z_t \rangle = \rho dt.$$

Lorsque ρ est négatif, la corrélation entre Z et B est négative. La dynamique du modèle aura donc l'effet de levier désiré.

Soit \tilde{Z} un mouvement brownien indépendant de Z . Par la décomposition de Cholesky, la dynamique du mouvement brownien B en fonction de celui de Z est

$$dB_t = \sqrt{1 - \rho^2} d\tilde{Z}_t + \rho dZ_t. \quad (3.4)$$

Ce modèle est souvent utilisé en pratique, car, tel que montré par Heston (1993), il est possible d'obtenir une forme semi-fermée pour évaluer certaines options vanilles. Ce modèle donne aussi une description plus précise de la réalité que le modèle de Black Scholes. En effet, les modèles à volatilité stochastiques offrent plus de flexibilité et peuvent mieux tenir compte du regroupement de la volatilité observée sur les marchés.

Il est possible d'utiliser l'équation (3.3) et le lemme d'Itô afin de déterminer le processus du log-rendement $y = \{y_t\}_{t \geq 0}$ défini par $y_t = \log(S_t)$ où

$$dy_t = \left(\mu - \frac{V_t}{2} \right) dt + \sqrt{V_t} dB_t \quad (3.5)$$

Bien que Broadie et Özgür Kaya (2006) ont démontré qu'il soit possible de le faire, ce modèle à volatilité stochastique est assez difficile à simuler directement puisque les incréments de la variance doivent être simulés simultanément à ceux du rendement. Nous sommes souvent contraints d'utiliser des approximations afin de simuler le modèle de Heston.

3.2 Discrétisation d'Euler-Maruyama

3.2.1 Principe de la discrétisation d'Euler-Maruyama

Cette discrétisation a été introduite par Maruyama (1954). Supposons une équation différentielle de la forme

$$dH_t = a(H_t, t)dt + b(H_t, t)dW_t,$$

où $a(H_t, t)$ est le coefficient de dérive et $b(H_t, t)$ est le coefficient de diffusion. Pour appliquer cette discrétisation, il faut que le coefficient de dérive et celui de diffusion soient différentiables. $W = \{W_t\}_{t \geq 0}$ est un mouvement brownien.

Une façon alternative d'écrire cette équation différentielle est d'utiliser sa forme intégrale.

$$H_{t+\Delta t} - H_t = \int_t^{t+\Delta t} a(H_u, u) du + \int_t^{t+\Delta t} b(H_u, u) dW_u. \quad (3.6)$$

Le principe général de la discrétisation d'Euler-Maruyama est de supposer que, sur un intervalle de temps suffisamment petit, la valeur du processus H_t changera très peu. Sous cette hypothèse, comme le changement de valeur du processus H_t est très petit et que la période de temps Δt est très petite, les coefficients de dérive et de diffusion seront plus ou moins constants sur cette période de temps. L'hypothèse utilisée dans la méthode d'Euler-Maruyama est que la valeur des coefficients de dérive et celle du coefficient de diffusion seront constantes sur cet intervalle et seront égales à leur valeur au temps t . Ceci se justifie en considérant tout changement de valeur des coefficients de dérive et de diffusion comme étant négligeable.

Sous cette hypothèse, l'intégrale du processus de dérive peut être approximée par

$$\int_t^{t+\Delta t} a(H_u, u) du \approx \int_t^{t+\Delta t} a(H_t, t) du.$$

Comme le processus de dérive au temps t est une constante par rapport à u , il faut l'isoler pour le sortir de l'intégrale

$$\begin{aligned} \int_t^{t+\Delta t} a(H_u, u) du &\approx a(H_t, t) \int_t^{t+\Delta t} du, \\ &= a(H_t, t) \Delta t. \end{aligned}$$

En utilisant le même principe, l'intégrale du processus de diffusion s'approxime par

$$\int_t^{t+\Delta t} b(H_u, u) dW_u \approx \int_t^{t+\Delta t} b(H_t, t) dW_u.$$

Comme le processus de diffusion au temps t est une constante par rapport à u , il est possible de le sortir de l'intégrale

$$\int_t^{t+\Delta t} b(H_u, u) dW_u \approx b(H_t, t) \int_t^{t+\Delta t} dW_u,$$

ce qui est égal en distribution à

$$\stackrel{d}{=} b(H_t, t) \sqrt{\Delta t} W^c,$$

où W^c suit une loi normale centrée et réduite et où $\stackrel{d}{=}$ dénote une égalité en distribution.

En joignant le résultat des deux intégrales ensemble, l'approximation de l'équation différentielle est donnée par

$$H_{t+\Delta t} \stackrel{d}{=} H_t + a(H_t, t) \Delta t + b(H_t, t) \sqrt{\Delta t} W^c, \quad (3.7)$$

pour $\Delta t > 0$.

3.2.2 Application au modèle d'Heston

En appliquant (3.7) à l'équation stochastique de la variance, il est possible discrétiser la variance stochastique avec la méthode d'Euler-Maruyama par

$$V_{t+\Delta t} = V_t + \kappa (\theta - V_t) \Delta t + \sigma \sqrt{V_t \Delta t} B^c \quad (3.8)$$

où B^c suit une loi normale centrée réduite.

En appliquant cette discrétisation aux log-rendements, il s'obtient que

$$y_{t+\Delta t} = y_t + \left(\mu - \frac{1}{2} \sqrt{V_t} \right) \Delta t + \sqrt{V_t \Delta t} Z^c, \quad (3.9)$$

où Z^c suit une loi normale centrée réduite.

B^c et Z^c suivent des lois normales centrées réduites avec une corrélation ρ . Pour que cette corrélation soit respectée, il est possible de se baser sur la décomposition de Cholesky définie à l'équation (3.4).

$$B^c = \sqrt{(1 - \rho^2)} \tilde{Z}^c + \rho Z^c. \quad (3.10)$$

La méthode d'Euler-Maruyama peut s'avérer problématique à utiliser pour le modèle de Heston puisque, lorsque la volatilité est faible, il y a une probabilité non nulle que la volatilité à la période suivante tombe sous zéro même lorsque la condition de Feller est respectée. Ceci peut être limité par une réduction de l'intervalle de temps entre les périodes. Sinon, il existe plusieurs façons de traiter les variances négatives dans le modèle de Heston. Les deux plus courantes méthodes sont la troncation, qui consiste à remplacer $V_{t+\Delta t}$ par $\max(0, V_{t+\Delta t})$, et la réflexion, qui consiste à remplacer $V_{t+\Delta t}$ par $|V_{t+\Delta t}|$.

3.3 Discrétisation de Milstein

3.3.1 Principe

La discrétisation de Milstein est une version un peu plus complexe du schéma d'Euler-Maruyama. Cette discrétisation a été introduite par Milstein (1975). Dans le contexte du modèle de Heston, la discrétisation de Milstein réduit fortement la probabilité que la valeur de la variance devienne négative. La dérivation effectuée dans le cadre de ce mémoire se base sur celle de Rouah (2011).

Pour pouvoir appliquer la discrétisation de Milstein, le coefficient de dérive et le

coefficient de diffusion doivent être choisis de façon à ne varier qu'en fonction du processus H_t . Le modèle de Heston respecte cette contrainte. De cette façon, le coefficient de dérive devient $a_t = a(H_t)$ et le coefficient de diffusion devient $b_t = b(H_t)$. Sous cette hypothèse, l'équation différentielle de H_t est maintenant de

$$dH_t = a_t dt + b_t dW_t.$$

En se basant sur le lemme d'Itô, il est possible d'écrire le coefficient de dérive et le coefficient de diffusion sous la forme des équations différentielles suivantes

$$\begin{aligned} da_t &= \frac{\partial a_t}{\partial H_t} dH_t + b_t \frac{\partial^2 a_t}{\partial H_t^2} dt \\ &= \left(a_t \frac{\partial a_t}{\partial H_t} + \frac{1}{2} b_t^2 \frac{\partial^2 a_t}{\partial H_t^2} \right) dt + b_t \frac{\partial a_t}{\partial H_t} dW_t \\ db_t &= \frac{\partial b_t}{\partial H_t} dH_t + b_t \frac{\partial^2 b_t}{\partial H_t^2} dt \\ &= \left(a_t \frac{\partial b_t}{\partial H_t} + \frac{1}{2} b_t^2 \frac{\partial^2 b_t}{\partial H_t^2} \right) dt + b_t \frac{\partial b_t}{\partial H_t} dW_t \end{aligned}$$

Notons que les deux équations précédentes sont vraies uniquement parce que $\frac{\partial a_t}{\partial t} = 0$ et $\frac{\partial b_t}{\partial t} = 0$. C'est pourquoi il était nécessaire que le coefficient de dérive et celui de diffusion ne dépendent pas de t .

En connaissant les équations différentielles du coefficient de dérive a_t et du coefficient de diffusion b_t , l'équation (3.6) peut être utilisée pour déduire leurs équations intégrales.

$$a_s = a_t + \int_t^s \left(a_u \frac{\partial a_u}{\partial H_u} + \frac{1}{2} b_u^2 \frac{\partial^2 a_u}{\partial H_u^2} \right) du + \int_t^s b_u \frac{\partial a_u}{\partial H_u} dW_u \quad (3.11)$$

$$b_s = b_t + \int_t^s \left(a_u \frac{\partial b_u}{\partial H_u} + \frac{1}{2} b_u^2 \frac{\partial^2 b_u}{\partial H_u^2} \right) du + \int_t^s b_u \frac{\partial b_u}{\partial H_u} dW_u \quad (3.12)$$

En insérant les deux équations (3.11) et (3.12) dans l'équation (3.6), cette équation

devient

$$H_{t+\Delta t} = H_t + \int_t^{t+\Delta t} \left(a_t + \int_t^s \left(a_u \frac{\partial a_u}{\partial H_u} + \frac{1}{2} b_u^2 \frac{\partial^2 a_u}{\partial H_u^2} \right) du + \int_t^s b_u \frac{\partial a_u}{\partial H_u} dW_u \right) ds + \\ \int_t^{t+\Delta t} \left(b_t + \int_t^s \left(a_u \frac{\partial b_u}{\partial H_u} + \frac{1}{2} b_u^2 \frac{\partial^2 b_u}{\partial H_u^2} \right) du + \int_t^s b_u \frac{\partial b_u}{\partial H_u} dW_u \right) dW_s$$

Selon les hypothèses de Milstein, tous les termes de degrés supérieurs à 1 dans l'expansion stochastique de la série Taylor peuvent être éliminés (voir Kloeden et Platen (2011) pour plus de détails sur le sujet), car l'impact de ceux-ci sur H_t est négligeable. De cette façon, les termes variant en fonction de dt^2 ou de $dW_t dt$ sont considérés comme nuls. Ceci permet de simplifier l'équation précédente à

$$H_{t+\Delta t} = H_t + \int_t^{t+\Delta t} a_t ds + \int_t^{t+\Delta t} \left(b_t + \int_t^s b_u \frac{\partial b_u}{\partial H_u} dW_u \right) dW_s.$$

Les facteurs a_t et b_t sont constants par rapport à s et W_s respectivement. Il est possible de les sortir de l'intégrale. Cette équation devient

$$H_{t+\Delta t} = H_t + a_t \int_t^{t+\Delta t} ds + b_t \int_t^{t+\Delta t} dW_s + \int_t^{t+\Delta t} \int_t^s b_u \frac{db_u}{dH_u} dW_u dW_s.$$

La seconde hypothèse utilisée par Milstein est qu'il est possible d'utiliser l'approximation d'Euler-Maruyama pour résoudre l'intégrale de droite dans l'équation précédente. L'équation précédente se simplifie donc à

$$H_{t+\Delta t} = H_t + a_t \int_t^{t+\Delta t} ds + b_t \int_t^{t+\Delta t} dW_s + b_t \frac{db_t}{dH_t} \int_t^{t+\Delta t} \int_t^s dW_u dW_s.$$

À partir de cette étape, l'intégrale des deux premiers termes a été calculée dans la section précédente. Ces deux termes représenteront exactement l'approximation d'Euler-Maruyama,

$$H_{t+\Delta t} = H_t + a_t \Delta t + b_t \sqrt{\Delta t} W^c + b_t \frac{db_t}{dH_t} \int_t^{t+\Delta t} \int_t^s dW_u dW_s,$$

où W^c suit une loi normale centrée réduite.

En intégrant une première fois, l'équation précédente devient

$$\begin{aligned} H_{t+\Delta t} &= H_t + a_t \Delta t + b_t \sqrt{\Delta t} W^c + b_t \frac{db_t}{dH_t} \int_t^{t+\Delta t} (W_s - W_t) dW_s, \\ &= H_t + a_t \Delta t + b_t \sqrt{\Delta t} W^c + b_t \frac{db_t}{dH_t} \left(\int_t^{t+\Delta t} W_s dW_s - W_t \int_t^{t+\Delta t} dW_s \right). \end{aligned} \quad (3.13)$$

Pour pouvoir résoudre l'intégrale précédente, il est possible d'appliquer le lemme d'Itô sur W_t^2 par rapport à l'équation différentielle de dW_t . La dynamique de ce processus stochastique est de

$$\begin{aligned} dW_t^2 &= \frac{\partial W_t^2}{\partial W_t} dW_t + \frac{1}{2} \frac{\partial^2 W_t^2}{\partial W_t^2} dt \\ &= 2W_t dW_t + dt. \end{aligned}$$

En utilisant sa forme intégrale, l'équation précédente est équivalente à

$$W_{t+\Delta t}^2 - W_t^2 = \int_t^{t+\Delta t} 2W_s dW_s + \int_t^{t+\Delta t} ds$$

En isolant le terme entre les parenthèses de l'équation (3.13), cette équation devient

$$\int_t^{t+\Delta t} W_s dW_s - W_t \int_t^{t+\Delta t} dW_s = \frac{W_{t+\Delta t}^2 - W_t^2}{2} - W_t \int_t^{t+\Delta t} dW_s - \frac{1}{2} \int_t^{t+\Delta t} dt$$

Les intégrales peuvent se résoudre par

$$\int_t^{t+\Delta t} W_s dW_s - W_t \int_t^{t+\Delta t} dW_s = \frac{W_{t+\Delta t}^2 - W_t^2}{2} - W_t (W_{t+\Delta t} - W_t) - \frac{1}{2} \Delta t.$$

Ce qui se simplifie par

$$\int_t^{t+\Delta t} W_s dW_s - W_t \int_t^{t+\Delta t} dW_s = \frac{(W_{t+\Delta t} - W_t)^2}{2} - \frac{1}{2} \Delta t.$$

Comme $W_{t+\Delta t} - W_t$ suit une loi normale où la moyenne vaut 0 et l'écart-type vaut Δt , l'équation précédente est équivalente à

$$\int_t^{t+\Delta t} W_s dW_s - W_t \int_t^{t+\Delta t} dW_s \stackrel{d}{=} \frac{(W^c)^2 - \Delta t}{2}, \quad (3.14)$$

où W^c suit une loi normale centrée réduite et où $\stackrel{d}{=}$ dénote une égalité en distribution.

En remplaçant l'intérieur de la parenthèse de l'équation (3.13) par l'égalité en distribution de ces intégrales trouvées à l'équation (3.14), cette équation devient

$$H_{t+\Delta t} \stackrel{d}{=} H_t + a_t \Delta t + b_t \sqrt{\Delta t} W^c + \frac{1}{2} b_t \Delta t \frac{db_t}{dH_t} \left((W^c)^2 - 1 \right). \quad (3.15)$$

3.3.2 Application du filtre à particules au modèle d'Heston

À partir de l'équation (3.15), il est possible d'approximer l'équation différentielle stochastique de la variance stochastique par

$$V_{t+\Delta t} = V_t + \kappa (\theta - V_t) \Delta t + \sigma \sqrt{V_t \Delta t} B^c + \frac{1}{4} \sigma^2 \Delta t (B^c - 1) \quad (3.16)$$

En utilisant l'équation (3.15), l'équation différentielle stochastique du log-rendement peut aussi s'approximer. Comme la dérivée $\frac{\partial b_t}{\partial H_t} = 0$ pour le log-rendement, la discrétisation de Milstein de y retourne la même formule que la discrétisation d'Euler-Maruyama, soit

$$y_{t+\Delta t} = y_t + \left(\mu - \frac{1}{2} \sqrt{V_t} \right) \Delta t + \sqrt{V_t \Delta t} Z^c. \quad (3.17)$$

La relation entre B^c et Z^c est la même que pour la discrétisation d'Euler-Maruyama et est décrite à l'équation (3.10).

3.4 Application du filtre à particules au modèle de Heston

Dans cette section, nous étudierons comment il est possible d'appliquer les différents filtres à particules au modèle de Heston.

3.4.1 Définition des variables

Le modèle de Heston, comme vu précédemment dans le chapitre, est un modèle dans lequel la dynamique du prix de l'action varie en fonction de la volatilité instantanée. Il est possible à tout moment d'observer le prix d'un titre sur les marchés financiers. Toutefois, il est impossible d'observer la volatilité instantanée du titre. Cette volatilité peut être implicitement estimée à l'aide des prix observés.

Le filtre à particules s'avère une méthode d'estimation intéressante pour modéliser la densité de la volatilité stochastique. Dans ce contexte, il est possible de concevoir le modèle de Heston comme une représentation d'état généralisée où la variable observable Y_k est le log-rendement quotidien de l'action et où la variable latente X_k est la valeur de la volatilité instantanée.

Comme il est difficile de simuler exactement le modèle de Heston, nous utiliserons la discrétisation de Milstein du modèle de Heston pour appliquer le filtre à particules sur les log-rendements d'un titre. Ceci permet d'estimer de façon adéquate la distribution de la volatilité stochastique. La discrétisation de Milstein a été favorisée dans ce mémoire puisqu'elle minimise la probabilité que la volatilité descende sous 0 lorsque la condition de Feller est respectée.

3.4.2 Discrétisation du modèle de Heston

Pour pouvoir discrétiser le modèle de Heston de façon à pouvoir l'appliquer aux filtres à particules, il est nécessaire que les intervalles entre les périodes soient égaux. La durée T , sous laquelle le titre est étudié, est séparée en n intervalles égaux. La durée de ces intervalles est de $\Delta_t = \frac{T}{n}$.

3.4.3 Les probabilités de transitions

Pour déterminer les probabilités de transitions du modèle de Heston pour le filtre à particules, il existe plusieurs méthodes. Dans ce mémoire, nous allons utiliser l'algorithme et allons nous inspirer de l'application du filtre à particules au modèle de Heston de Christoffersen *et al.* (2007). Nous allons toutefois adapter leur méthode afin d'utiliser une discrétisation de Milstein plutôt que celle d'Euler-Maruyama.

3.4.3.1 Échantillonnage de la volatilité stochastique

Pour pouvoir appliquer le filtre à particules au modèle de Heston, nous utilisons l'algorithme de Bootstrap présenté au chapitre I. De cette façon, pour simuler aléatoirement les volatilités stochastiques, il faut utiliser la probabilité de transition $q_{k|k-1} = f(X_k | X_{k-1}, Y_{k-1}; \theta)$.

En se basant sur la discrétisation de Milstein de l'équation (3.16), la volatilité stochastique à la période k peut être exprimée dans la façon suivante

$$V_k = V_{k-1} + \kappa(\theta - V_{k-1}) \frac{T}{n} + \sigma \sqrt{V_{k-1} \frac{T}{n}} B_k^c + \frac{1}{4} \sigma^2 \frac{T}{n} (B_k^c - 1). \quad (3.18)$$

où B_k^c suit une loi normale centrée et réduite.

En se basant sur l'équation (3.17), le log-rendement périodique R_k à la k^e période est donné par

$$\begin{aligned} R_k &= y_k - y_{k-1}, \\ R_k &= \left(\mu - \frac{1}{2} \sqrt{V_{k-1}} \right) \frac{T}{n} + \sqrt{V_{k-1} \frac{T}{n}} Z_k^c. \end{aligned} \quad (3.19)$$

où Z_k^c suit une loi normale centrée et réduite.

Les variables aléatoires B_k^c et Z_k^c sont corrélées par l'équation (3.10).

Comme R_k est une observation et que pour la i^e particule, la volatilité stochas-

tique précédente $V_{k-1}^{(i)}$ est connue, il devient possible de simuler la volatilité de la i^e particule à la période actuelle $V_k^{(i)}$.

Il faut d'abord calculer $Z_k^{c(i)}$, la variable aléatoire qui aurait été simulée afin de générer l'observation R_k si la volatilité précédente était de $V_{k-1}^{(i)}$ en l'isolant dans l'équation (3.19).

$$Z_k^{c(i)} = \frac{R_k - \left(\mu - \frac{1}{2} \sqrt{V_{k-1}^{(i)}} \right) \frac{T}{n}}{\sqrt{V_{k-1}^{(i)}} \frac{T}{n}}.$$

Par la suite, $\tilde{Z}_k^{c(i)}$ est simulée à l'aide d'une loi normale centrée réduite indépendante de $Z_k^{c(i)}$ et il devient possible de calculer $B_k^{c(i)}$ à l'aide de l'équation (3.10). Cette variable vaut

$$B_k^{c(i)} = \sqrt{(1 - \rho^2)} \tilde{Z}_k^{c(i)} + \rho Z_k^{c(i)}.$$

Enfin, il ne reste plus qu'à calculer $V_k^{(i)}$ à l'aide de

$$V_k^{(i)} = V_{k-1}^{(i)} + \kappa \left(\theta - V_{k-1}^{(i)} \right) \frac{T}{n} + \sigma \sqrt{V_{k-1}^{(i)}} \frac{T}{n} B_k^{c(i)} + \frac{1}{4} \sigma^2 \frac{T}{n} \left(B_k^{c(i)} - 1 \right).$$

3.4.3.2 Calcul des poids

Pour pouvoir appliquer le filtre à particules au modèle de Heston, nous utilisons l'algorithme de Bootstrap présenté au chapitre 1.

Sous cette hypothèse, le poids de la contribution de la période k au poids de la particule est défini par l'équation (1.24). Pour pouvoir calculer le poids des particules, il devient donc nécessaire de trouver la distribution conditionnelle $f(R_k | V_k^{(i)})$.

En se basant sur l'équation (3.19), il est possible de déduire que, comme $Z_k^{c(i)}$ suit une loi normale centrée réduite, alors

$$R_k | V_{k-1}^{(i)} \sim \text{Normale} \left(\left(\mu - \frac{1}{2} \sqrt{V_{k-1}^{(i)}} \right) \frac{T}{n}, \sqrt{V_{k-1}^{(i)}} \frac{T}{n} \right).$$

À l'aide de cette information, il devient possible de calculer $\tilde{\omega}_k^{(i)}$

$$\begin{aligned}\tilde{\omega}_k^{(i)} &= f(R_k | V_k^{(i)}) \\ &= \frac{1}{\sqrt{2\pi V_{k-1}^{(i)} \frac{T}{n}}} \exp\left(-\frac{\left(R_k - \left(\mu - \frac{1}{2}\sqrt{V_{k-1}^{(i)}}\right) \frac{T}{n}\right)^2}{2V_{k-1}^{(i)} \frac{T}{n}}\right).\end{aligned}$$

CHAPITRE IV

RÉSULTATS NUMÉRIQUES

Dans cette section, nous cherchons à observer quels filtres à particules sont les mieux adaptés pour le modèle de Heston et quels filtres à particules performant le mieux pour calibrer le modèle de Heston par maximum de vraisemblance.

4.1 Modèle simulé

Pour comparer ces différents filtres à particules, nous utiliserons une trajectoire du modèle de Heston simulée à partir d'une discrétisation de Milstein sur une période fixe. Dans ce chapitre, la trajectoire utilisée suit les log-rendements d'un titre sur une période de dix années $T = 10$ et chaque période a une longueur de $\Delta_t = \frac{1}{250}$ année. Considérant que la bourse est traditionnellement ouverte 250 jours par année, ces rendements peuvent être considérés comme des rendements quotidiens. Le titre est donc étudié sur un total de $n = 2500$ périodes.

4.1.1 Paramètres du modèle

Le jeu de paramètres utilisé afin de simuler notre trajectoire a été choisi de façon à obtenir un rendement annuel moyen d'environ $\mu = 0.03$ et à s'assurer que la volatilité du titre ait une moyenne à long terme de $\theta = 0.2$ qui est relativement proche de la dynamique des actifs observés dans la réalité. κ et σ ont été choisis afin de s'assurer qu'il puisse y avoir un certain mouvement dans la variance du titre et une relativement forte tendance du titre à revenir à sa valeur d'équilibre. Sous ces

paramètres, la dérive instantanée de V_t est de $\kappa(\theta - V_t)$. L'écart-type instantané de la volatilité stochastique est de $\sigma\sqrt{V_t}$. ρ a été sélectionné afin que l'effet de levier entre la variance et le rendement soit facilement observable. Les paramètres utilisés pour tester notre modèle sont définis dans le tableau 4.1.

Tableau 4.1: Paramètres initiaux

Paramètre	Valeur
μ	0.03
κ	6.0
θ	0.2
σ	0.5
ρ	-0.7

Cet ensemble de paramètres se notera β_0 .

À partir de ces paramètres, il est possible de simuler pour les log-rendements quotidiens R_k et la volatilité stochastique V_k . La série des log-rendements quotidiens $R_{1:n}$ et celle de la volatilité stochastique $V_{1:n}$ sont présentées ci-dessous dans les figures 4.1 et 4.2 respectivement. Dans ce modèle, les log-rendements quotidiens de la figure 4.1 serviront d'observations. La série temporelle de la volatilité stochastique qui est représentée à la figure 4.2 représente le signal dont nous souhaitons estimer la distribution.

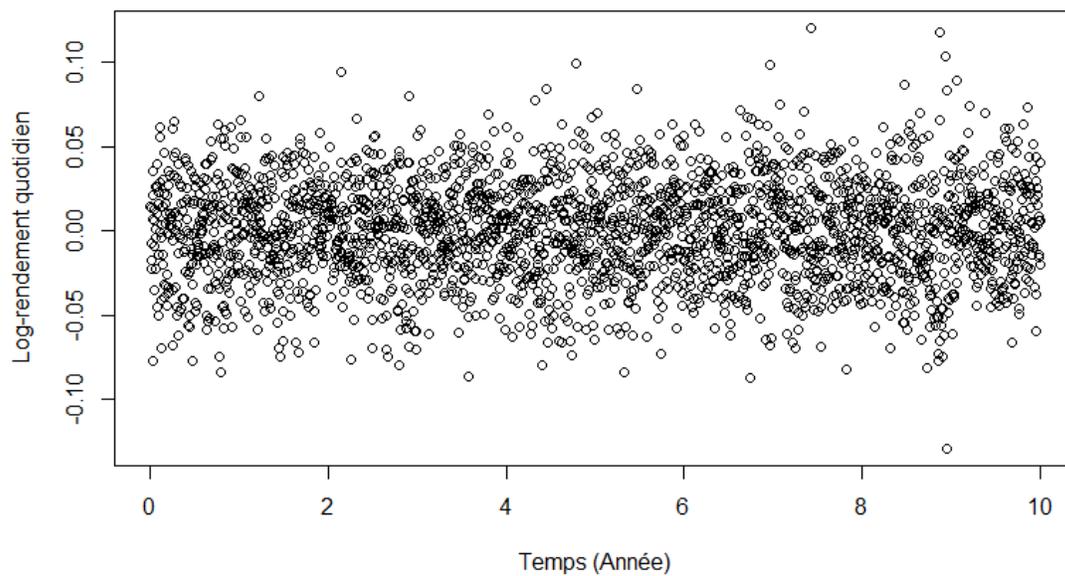


Figure 4.1: Log-rendements quotidiens

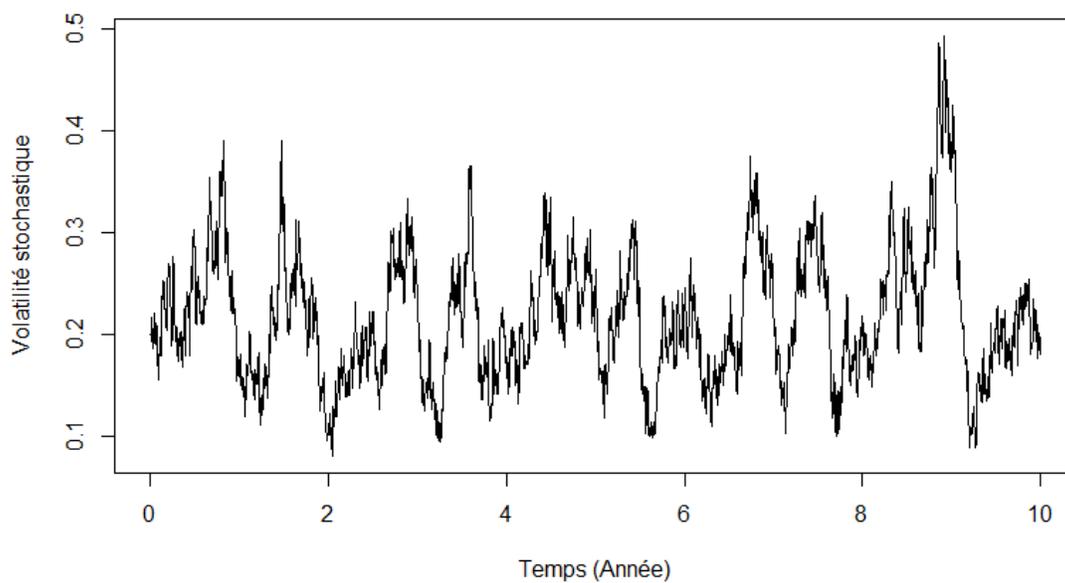


Figure 4.2: Volatilité stochastique

4.2 Comparaison des performances des différents filtres à particules

4.2.1 Filtres à particules avec ré-échantillonnage

À l'aide de la simulation effectuée dans la section précédente, nous allons comparer la performance des différents filtres à particules. Pour effectuer cette comparaison, nous devons comparer la fonction de densité de la volatilité stochastique estimée par les différents filtres à particules.

Une des premières mesures pouvant servir à évaluer cette fonction de densité est l'espérance de la volatilité stochastique estimée par le filtre à particules $\hat{E}[V_k | R_{1:k}, \beta_0]$. Il est possible de trouver la valeur de cet estimateur à l'aide de la volatilité stochastique de chaque particule $V_k^{(i)}$ et de la fonction de densité des filtres à particules présentée à l'équation (1.16). Cette espérance est calculée par

$$\hat{E}[V_k | R_{1:k}, \beta_0] = \sum_{i=1}^N \tilde{\omega}_k^{(i)} V_k^{(i)}. \quad (4.1)$$

Les poids et les volatilités utilisés pour calculer l'espérance empirique sont ceux avant le ré-échantillonnage, car bien que le ré-échantillonnage ait de nombreux avantages, la densité prédite après le ré-échantillonnage tend à être un peu moins précise que celle prédite avant le ré-échantillonnage. Ceci est causé par l'aléa ajouté par le ré-échantillonnage.

Tel que montré par les figures 4.3, 4.4, 4.5 et 4.6, l'espérance de la volatilité stochastique générée par les filtres à particules conditionnelle aux observations capte relativement bien la trajectoire de la volatilité stochastique du modèle de Heston. Pour un grand nombre de particules ($N = 10\,000$), il aussi est observé que la trajectoire des différents filtres converge vers les mêmes estimations de la valeur de la volatilité stochastique. Ceci indique que, pour la série de rendement quotidien indiqué, l'espérance de la distribution empirique de V_t converge toujours vers les mêmes valeurs pour tous les filtres à particules utilisant le même jeu de paramètres.

Toutefois, le filtre à particules a de la difficulté à capter les fortes hausses de la volatilité lorsque celles-ci ne sont pas couplées avec des log-rendements extrêmes. Inversement, lorsque la volatilité diminue et que les log-rendements restent très loin des log-rendements moyens, les filtres à particules ne captent pas très bien la baisse.

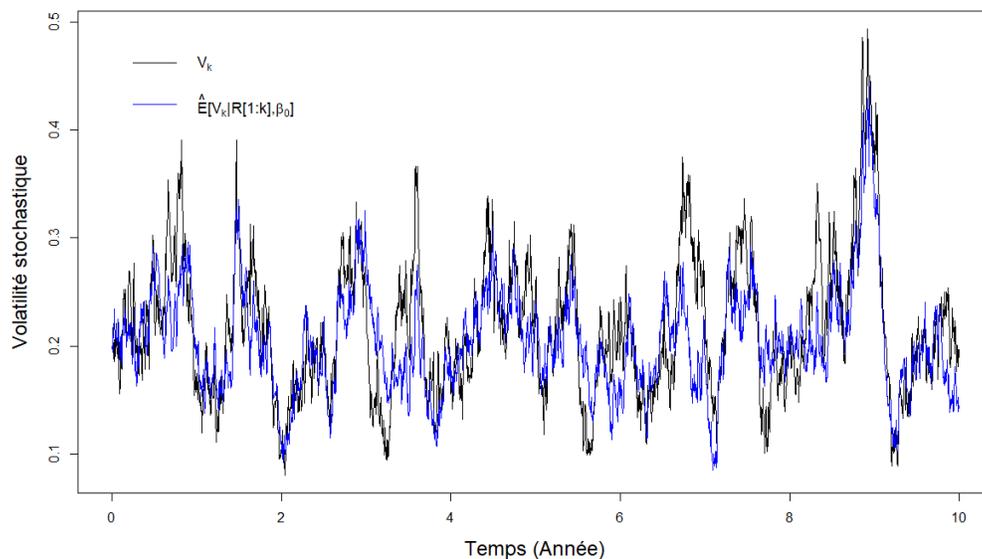


Figure 4.3: Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec ré-échantillonnage standard avec la volatilité réelle de la trajectoire, ($N = 10\,000$)

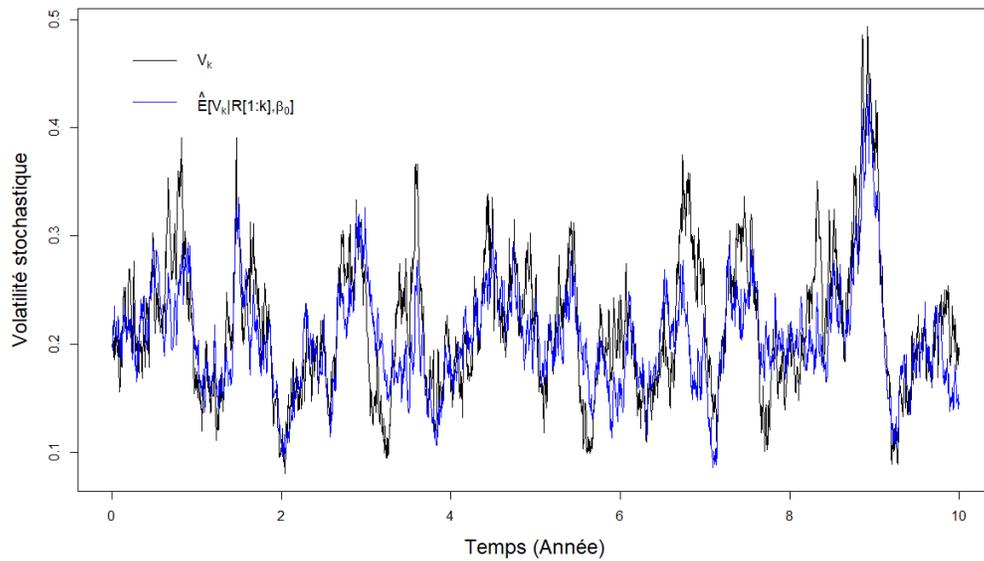


Figure 4.4: Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec ré-échantillonnage stratifié avec la volatilité réelle de la trajectoire, ($N = 10\,000$)

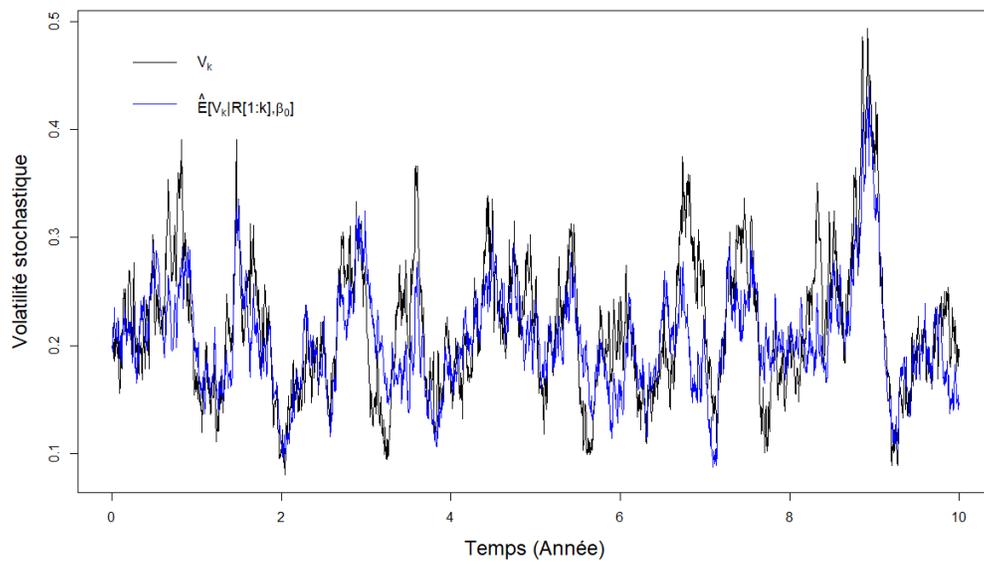


Figure 4.5: Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec ré-échantillonnage résiduel avec la volatilité réelle de la trajectoire, ($N = 10\,000$)

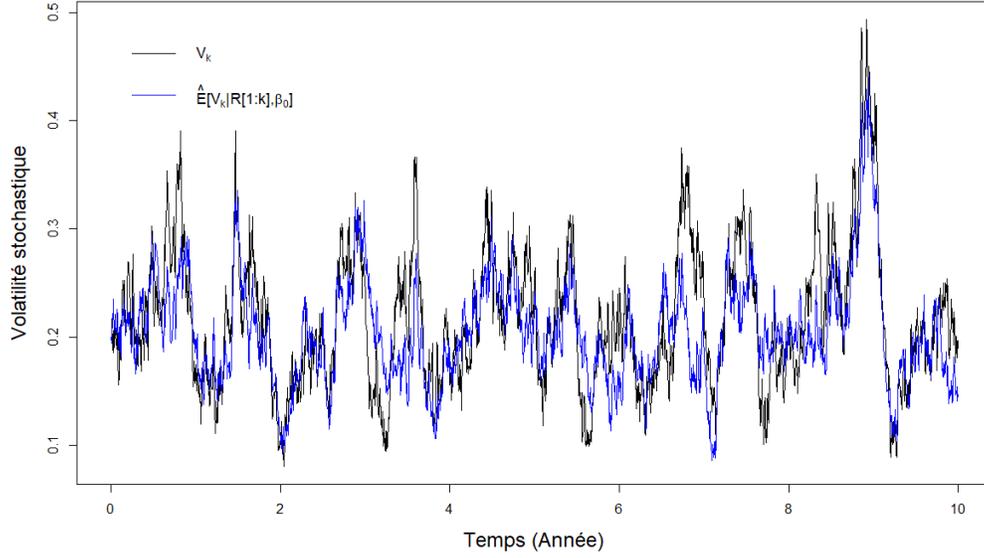


Figure 4.6: Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec ré-échantillonnage combiné avec la volatilité réelle de la trajectoire, ($N = 10\,000$)

Soit $r(i)$ le rang des particules classées en ordre croissant en fonction de leur volatilité stochastique à la période k . De cette façon, la volatilité stochastique de rang $r(i)$ est définie comme

$$V_k^{r(1)} \leq \dots \leq V_k^{r(i-1)} \leq V_k^{r(i)} \leq V_k^{r(i+1)} \leq \dots \leq V_k^{r(N)}.$$

Définissons le poids pondéré de la particule de rang $r(i)$ comme $\tilde{\omega}_k^{r(i)}$. La fonction de répartition empirique de la volatilité stochastique conditionnelle aux observations passées $F(v|R_{1:k-1})$ se définit

$$F(v|R_{1:k}, \beta_0) = \begin{cases} 0 & \text{pour } v < V_k^{r(1)}, \\ \sum_{i=1}^s \tilde{\omega}_k^{r(i)} & \text{pour } V_k^{r(s)} \leq v < V_k^{r(s+1)}, \\ 1 & \text{pour } v \geq V_k^{r(N)}. \end{cases} \quad (4.2)$$

À l'aide de cette fonction de répartition, il est possible de calculer les quan-

tiles empiriques Q_j . Ces quantiles empiriques Q_j représentent la valeur telle que la probabilité que la volatilité stochastique soit inférieure à ce quantile soit au moins de $j\%$. Pour calculer ce quantile, il faut premièrement évaluer le rang de ce quantile s_j

$$s_j = \inf_{s \in \{1, 2, \dots, N\}} \sum_{i=1}^s \tilde{\omega}_k^{r(i)} > 0.01j.$$

À partir de ce rang, évaluer le quantile Q_j peut se faire à l'aide de l'équation

$$Q_j = V_k^{r(s_j)}.$$

Hormis quelques exceptions, la figure 4.7 montre très clairement que la volatilité stochastique du titre est presque toujours entre le 5^e et le 95^e quantile de la densité estimée par le filtre à particules. Ceci indique que le filtre à particules est capable de bien s'adapter au modèle de Heston.

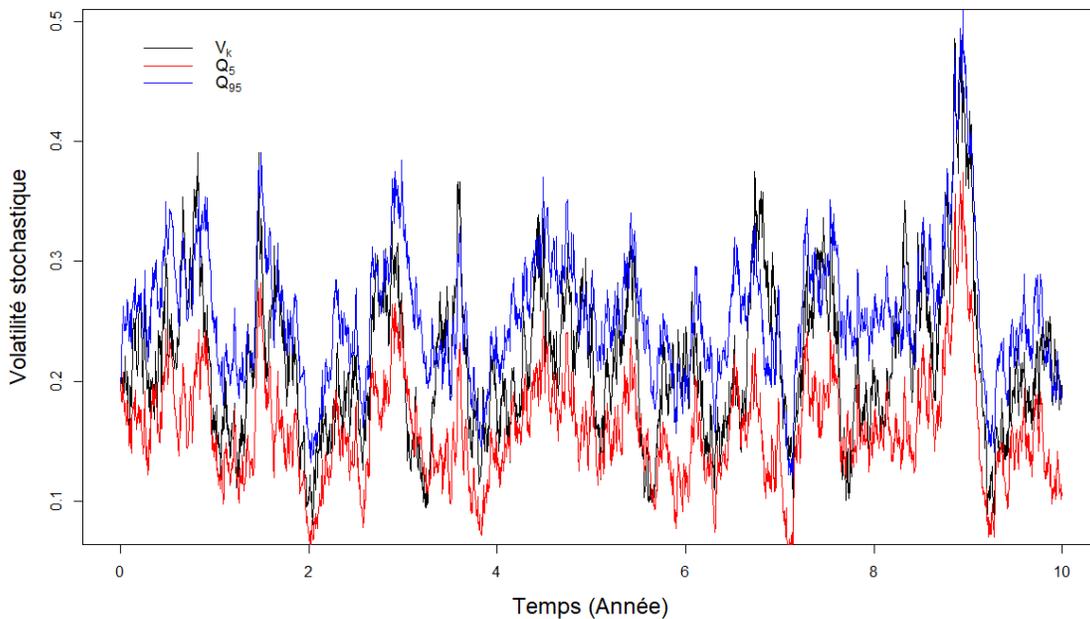


Figure 4.7: Évolution quotidienne des quantiles 95% (en bleu) et 5% (en rouge) de la densité estimée par un filtre à particules avec ré-échantillonnage standard ($N = 10\,000$) de la volatilité stochastique à travers le temps.

4.2.2 Erreur de prédiction

Dans cette sous-section, nous introduisons les erreurs de prédiction et nous comparons les différents filtres à particules par rapport à cette mesure. L'erreur de prédiction est définie par

$$Erreur = \sqrt{\frac{1}{n} \sum_{k=1}^n \left(\hat{E} [V_k | R_{1:k}, \beta_0] - V_k \right)^2},$$

où $\hat{E} [V_k | R_{1:k}, \beta_0]$ est l'espérance de la volatilité conditionnelle aux observations estimée par le filtre à particules définie par l'équation (4.1).

En se basant sur les figures 4.8 et 4.9, nous observons que lorsque nous utilisons 50 filtres distincts pour évaluer l'erreur de prédiction, l'erreur de prédiction moyenne est plus élevée lorsque le nombre de particules utilisées est très petit et que le temps d'exécution de ces filtres est très court. Toutefois, en augmentant la taille de l'échantillon de ces filtres à particules, non seulement l'erreur de prédiction est réduite, mais l'écart-type de l'erreur l'est pour tous les filtres à particules avec ré-échantillonnage. Ceci nous laisse croire que si le nombre de particules tend vers l'infini, tous les filtres vont converger vers la même fonction de densité.

En comparant les différents filtres à particules, il est possible d'observer que le filtre à ré-échantillonnage standard est celui engrangeant les plus grandes erreurs pour un même temps d'exécution, en plus d'être celui ayant la plus faible stabilité. C'est-à-dire que, pour un même temps d'exécution, l'écart-type de l'erreur de prédiction sera plus élevé ce qui indique une convergence plus lente de ce type de filtre vers la densité optimale. Bien qu'il améliore légèrement la performance du filtre à particules avec ré-échantillonnage standard, le filtre à particules avec ré-échantillonnage résiduel ne fait guère mieux. Les filtres à particules avec ré-échantillonnage combiné et stratifié sont les filtres performant le mieux. Toutefois, pour un nombre de particules égal, le filtre combiné a un temps d'exécution plus court, mais une erreur de prédiction et un écart-type sur l'erreur de prédiction plus élevés. Avec un nombre de particules

égal, le filtre à particules stratifié est aussi plus stable.

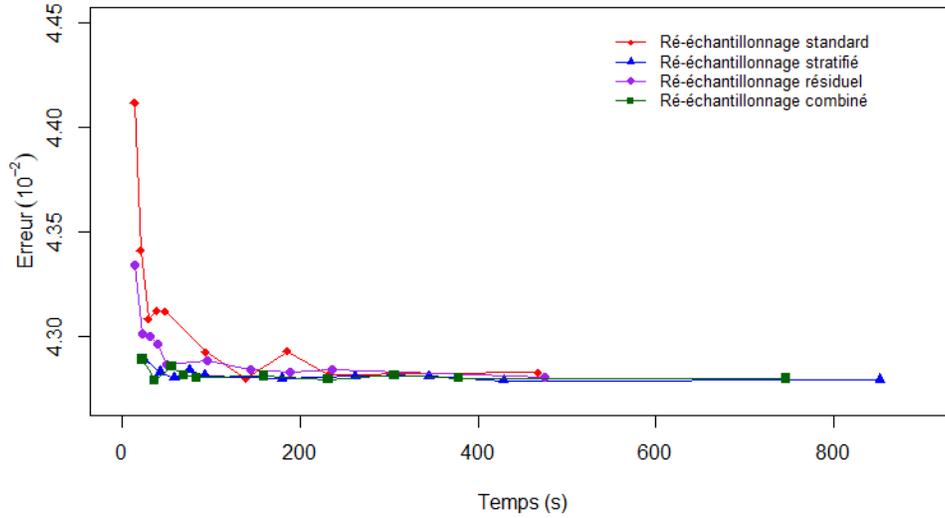


Figure 4.8: Erreur de prédiction moyenne des différents filtres à particules avec ré-échantillonnage par rapport au temps d'exécution de 50 filtres.

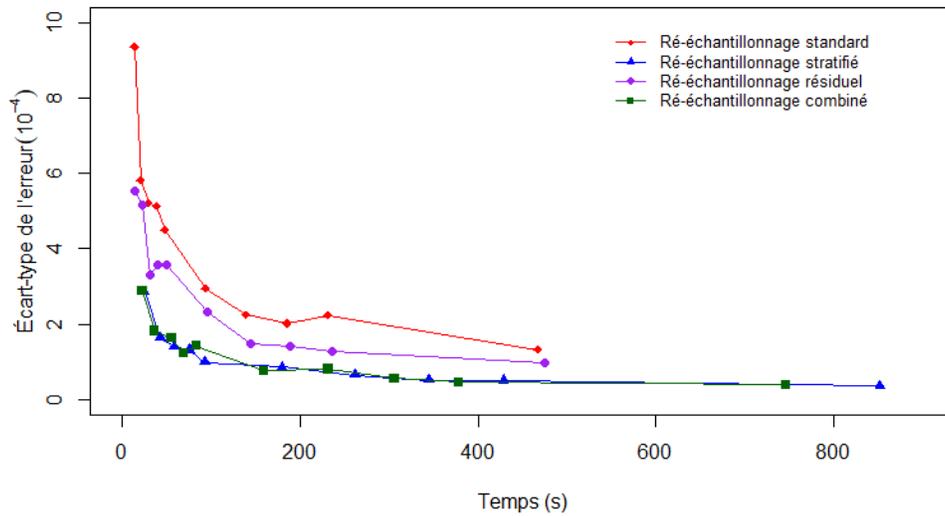


Figure 4.9: Écart-type de l'erreur de prédiction des différents filtres à particules avec ré-échantillonnage par rapport au temps d'exécution de 50 filtres.

4.2.3 Filtres à particules avec branchement

Pour les filtres à particules avec branchement, nous analyserons deux types d’algorithmes : le filtre à particules avec branchement combiné et le filtre à particules avec branchement effectif. Pour ces deux filtres, nous considérons trois niveaux de contagion. Le niveau de contagion du filtre à particules avec branchement est essentiellement la proportion des particules qui sont contaminées. Les particules qui sont contaminées sont celles qui se retrouveront dans l’intervalle c_k , c’est-à-dire que ce sont les particules dont l’information complète (poids et volatilité stochastique) seront transmis à une seule particule. Les niveaux de contagion choisis sont de 95% pour la contagion forte, 70% pour la contagion moyenne et 30% pour la contagion faible. Pour l’algorithme avec branchement combiné, c est choisi de façon à respecter ces niveaux de contamination. Alternativement, pour les algorithmes avec branchement effectif, c^{eff} et c^{neff} sont choisis de façon à respecter ces niveaux de contamination tout en stabilisant légèrement le taux de contagion d’une période à l’autre.

Tableau 4.2: Constante multiplicative des bornes de l’intervalle pour les filtres avec branchement

	Contagion faible (30%)	Contagion moyenne (70%)	Contagion forte (95%)
c	1.018	1.083	1.450
c^{eff}	1.005	1.020	1.300
c^{neff}	1.020	1.094	1.580

c se retrouve entre c^{eff} et c^{neff} . Ce résultat n’est pas surprenant puisque c’est la seule façon pour que les deux filtres aient des niveaux de contaminations similaires. Toutefois, il peut aussi être remarqué que c est plus proche de c^{neff} que de c^{eff} ce qui indiquerait que c^{eff} n’a pas un très grand impact sur le niveau de contagion. Ceci peut s’expliquer par le fait qu’à la plupart des périodes, la dégénérescence des particules est très faible si bien que c_k peut être très petit et inclure un grand pourcentage des particules à la plupart des périodes. Toutefois, il existe un nombre

de périodes relativement petit où il y a une forte dégénérescence des poids et où la contamination est considérablement plus faible si c_k n'est pas beaucoup plus grand. C'est pour cela que c doit être d'un niveau comparable à c^{neff} pour que les deux algorithmes aient des niveaux de contamination similaire.

Pour ces filtres, nous souhaitons comparer les différents estimateurs calculés à la sous-section précédente. Pour estimer l'espérance de la variable latente conditionnelle aux observations $\hat{E}[V_k | R_{1:k}, \beta_0]$, il est nécessaire, ici, d'adapter la formule (4.1) pour tenir en compte le fait que le nombre de particules est variable. La valeur de cet estimateur pour les filtres avec branchement est de

$$\hat{E}[V_k | R_{1:k}, \beta_0] = \sum_{i=1}^{N_k} \tilde{\omega}_k^{(i)} V_k^{(i)}. \quad (4.3)$$

En regardant les figures 4.10, 4.11, 4.12, 4.13, 4.14 et 4.15, il est possible de constater que les filtres à particules avec branchement obtiennent sensiblement les mêmes valeurs de l'espérance de la variable latente conditionnelle aux observations $E[V_k | R_{1:k}, \beta_0]$ que les filtres à particules avec ré-échantillonnage.

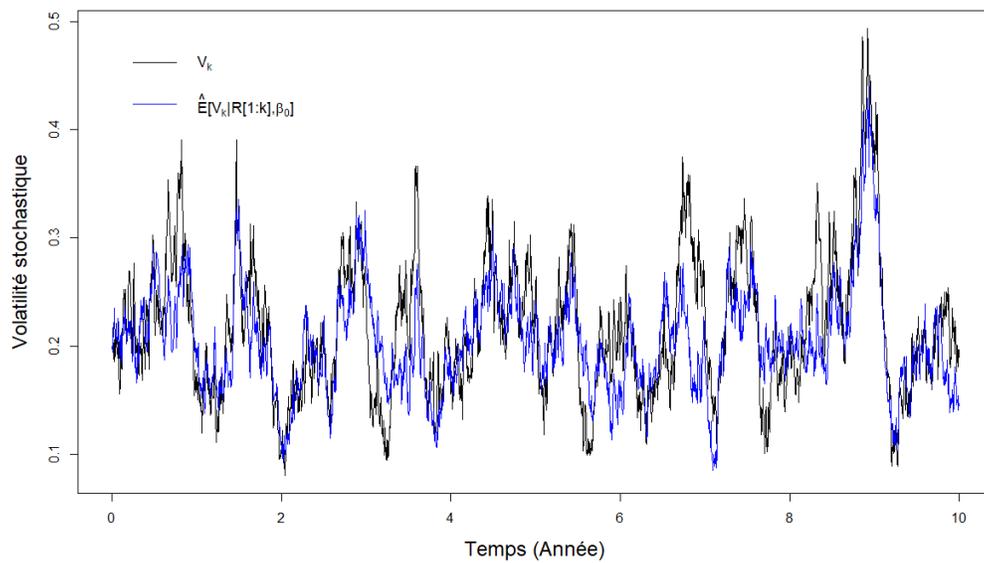


Figure 4.10: Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement combiné avec un taux de contagion de 95% avec la volatilité réelle de la trajectoire, ($N_0 = 10\,000$)

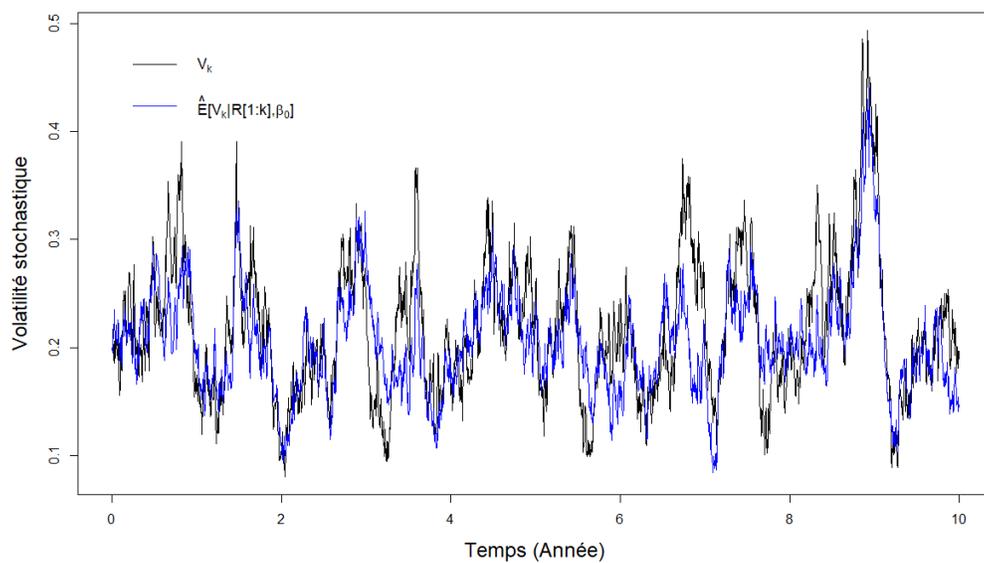


Figure 4.11: Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement combiné avec un taux de contagion de 70% avec la volatilité réelle de la trajectoire, ($N_0 = 10\,000$)

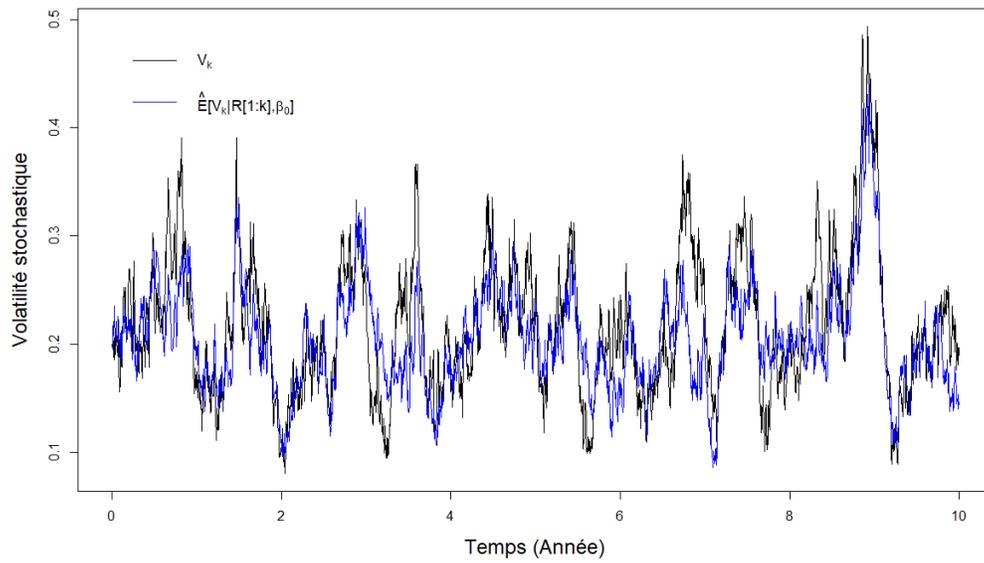


Figure 4.12: Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement combiné avec un taux de contagion de 30% avec la volatilité réelle de la trajectoire, ($N_0 = 10\,000$)

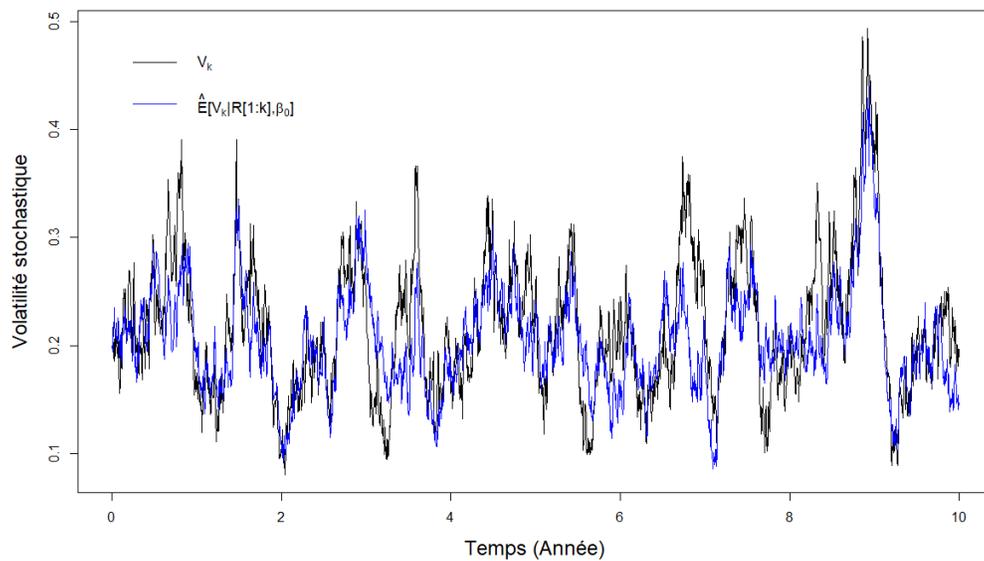


Figure 4.13: Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement effectif avec un taux de contagion de 95% avec la volatilité réelle de la trajectoire, ($N_0 = 10\,000$)

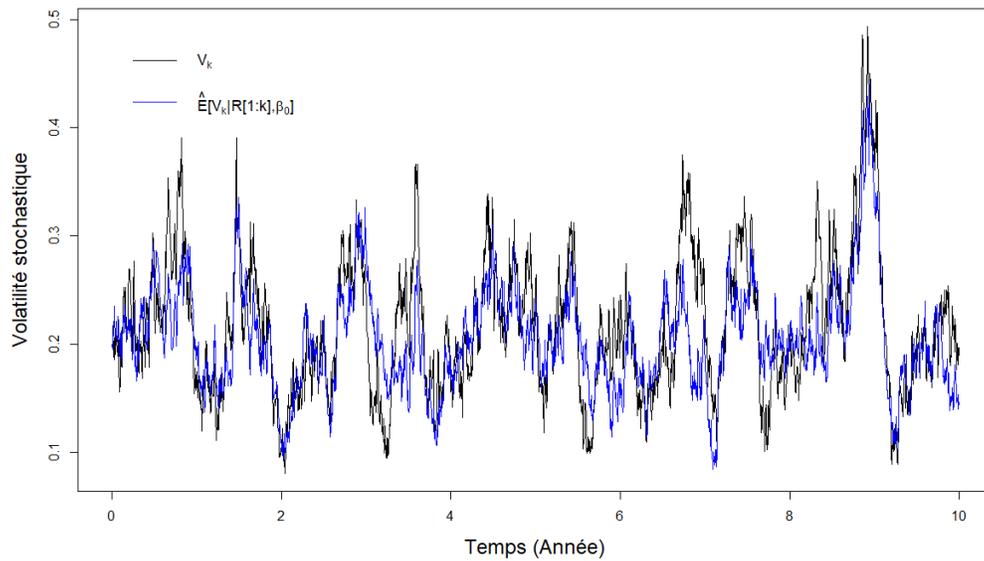


Figure 4.14: Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement effectif avec un taux de contagion de 70% avec la volatilité réelle de la trajectoire, ($N_0 = 10\,000$)

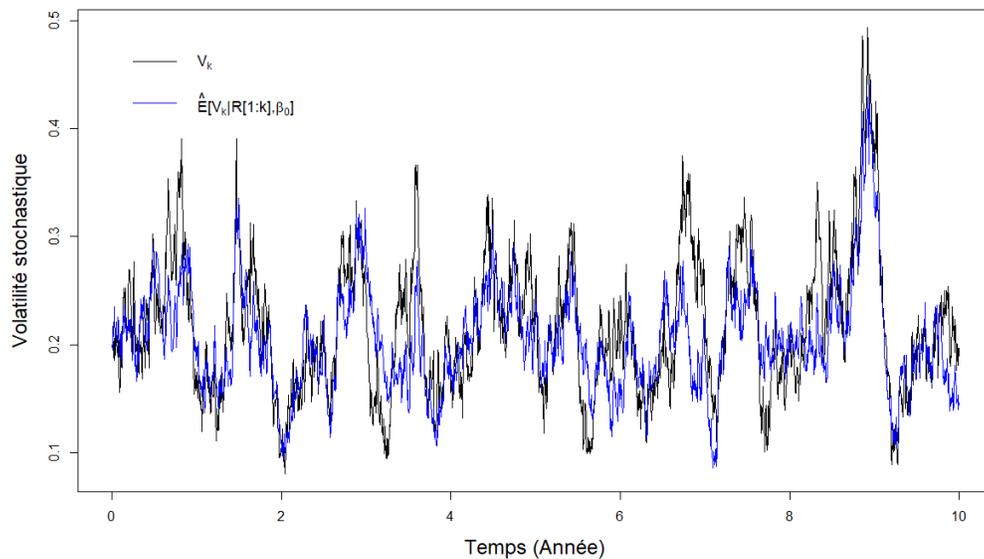
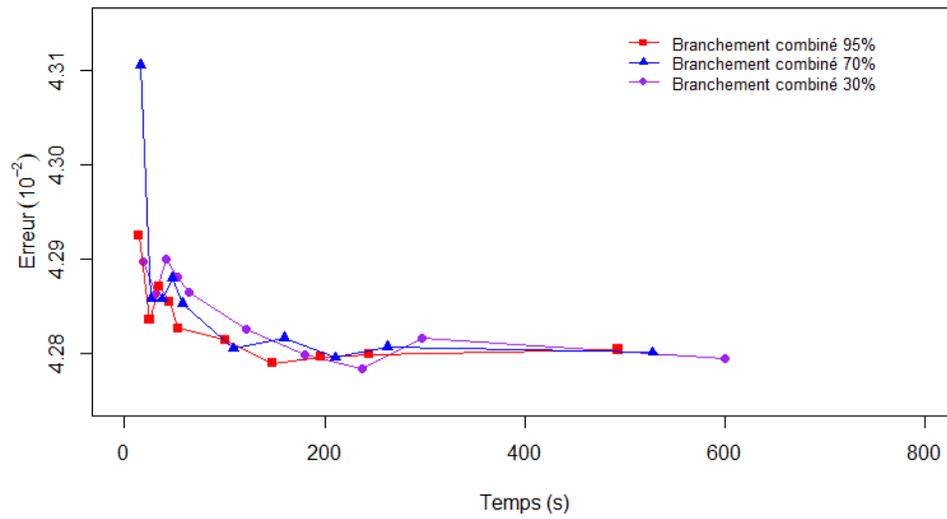


Figure 4.15: Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement effectif avec un taux de contagion de 30% avec la volatilité réelle de la trajectoire, ($N_0 = 10\,000$)

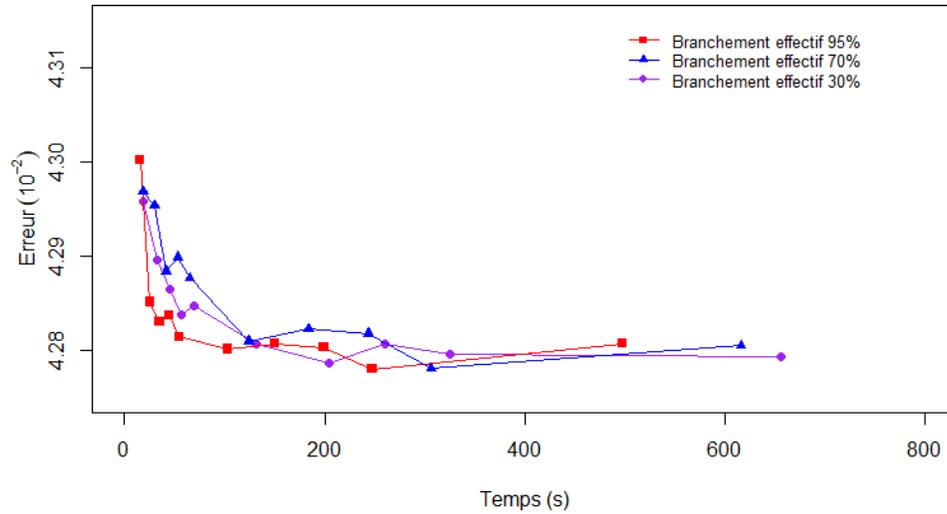
En analysant les figures 4.16 et 4.17, la même tendance pour les filtres à particules

avec branchement que pour les filtres à particules avec ré-échantillonnage peut être observée, c'est-à-dire que lorsque le temps d'exécution et le nombre de particules du filtre augmentent, l'erreur de prédiction moyenne et son écart-type diminuent. Les courbes de l'erreur de prédiction des filtres avec branchement et celles de leurs écarts-types sont beaucoup plus rapprochées les unes des autres. Les filtres à particules avec un faible taux de contamination ont tendance à avoir une erreur moyenne plus grande et un écart-type plus élevé pour un même temps d'exécution. Ceci s'explique surtout par le fait que les filtres à particules avec des taux de contagion plus élevés sont nettement plus rapides que ceux avec des taux de contagion faibles pour un nombre de particules initiales identique. Toutefois, pour un nombre de particules initiales identique, l'erreur de prédiction et son écart-type sont très similaires, bien que celle des filtres avec des taux de contamination plus élevés est légèrement plus faible.

Lorsque les filtres à particules avec branchement et ceux avec ré-échantillonnage sont comparés, il est possible de remarquer que les filtres avec branchement ont en général des temps d'exécution plus court que les filtres à particules avec ré-échantillonnage stratifié et combiné, mais plus long que ceux avec ré-échantillonnage standard et résiduel. Pour un même nombre de particules, les filtres à particules avec branchement sont légèrement moins précis que le filtre à particules avec ré-échantillonnage stratifié et combiné, mais sont nettement plus précis que ceux avec ré-échantillonnage standard et résiduel. Pour un même temps d'exécution, les filtres à contagion faible (30%) et moyenne (70%) ne sont pas aussi précis que les filtres à particules avec ré-échantillonnage combiné et stratifié. Toutefois, pour un même temps d'exécution, les filtres à particules avec branchement avec un taux de contagion élevé (95%) sont plus précis que tous les filtres à particules avec ré-échantillonnage. C'est-à-dire que pour un temps d'exécution similaire, les filtres à particules avec branchement avec des taux de contamination élevés peuvent se permettre d'utiliser un nombre de particules initiales plus élevé que les filtres à particules avec ré-échantillonnage stratifié ou combiné, ce qui leur permet d'être plus précis que ceux-ci.

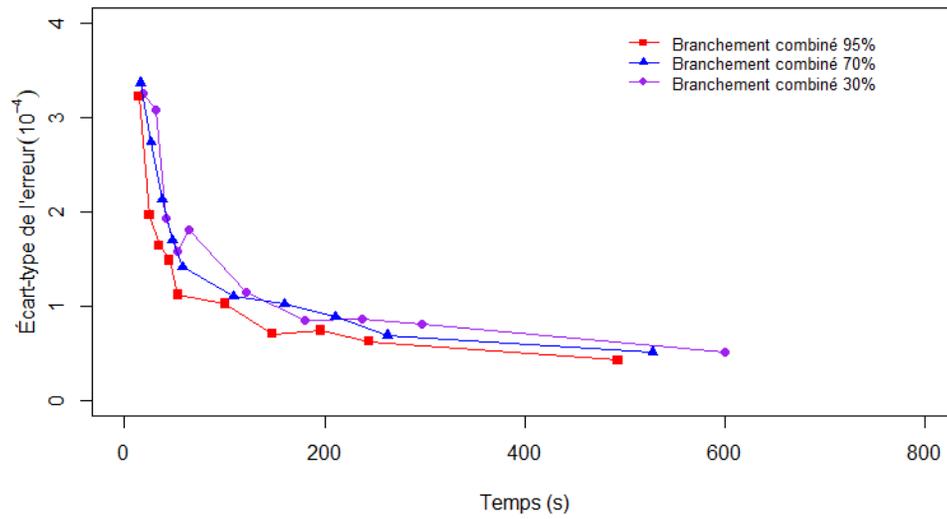


(a) Branchement combiné

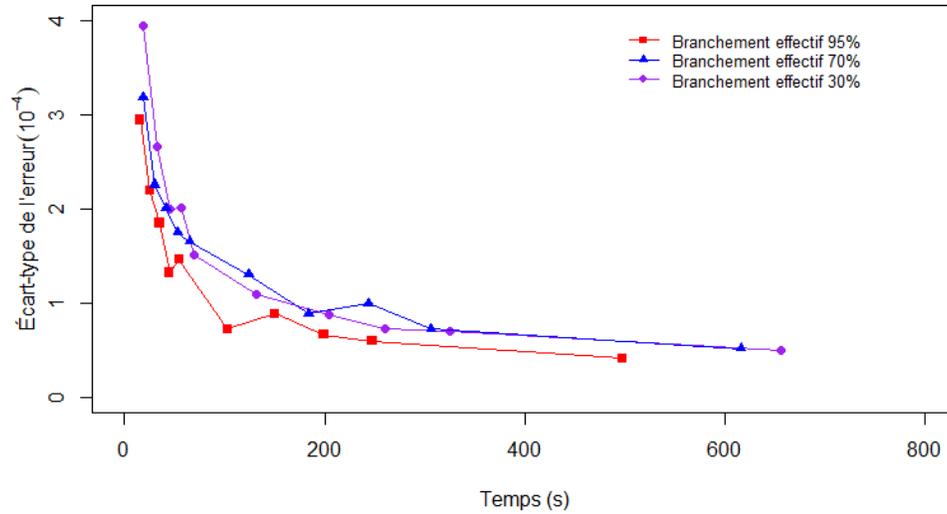


(b) Branchement effectif

Figure 4.16: Erreur de prédiction moyenne des différents filtres à particules avec branchement par rapport au temps d'exécution de 50 filtres.



(a) Branchement combiné



(b) Branchement effectif

Figure 4.17: Écart-type de l'erreur de prédiction des différents filtres à particules avec branchement par rapport au temps d'exécution de 50 filtres.

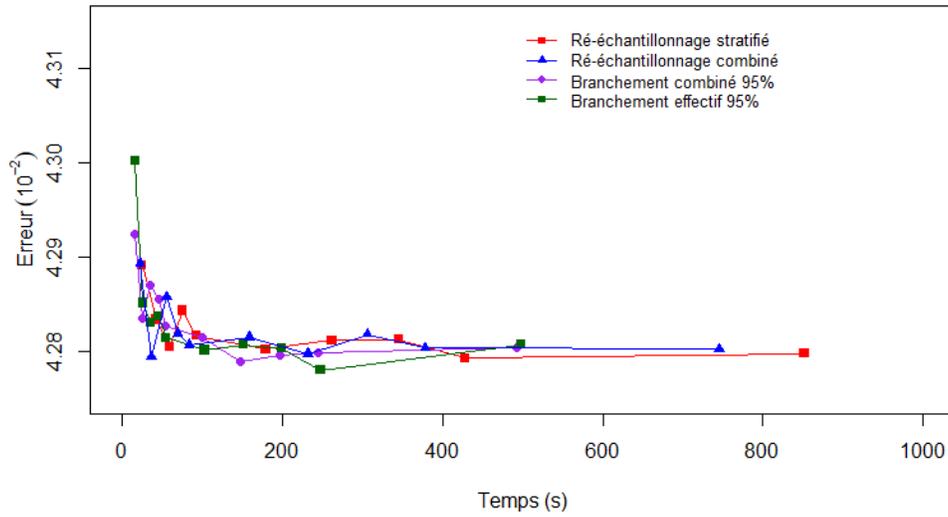


Figure 4.18: Comparaison des erreurs de prédiction moyenne des meilleurs filtres à particules avec ré-échantillonnage et avec branchement par rapport au temps d'exécution de 50 filtres.

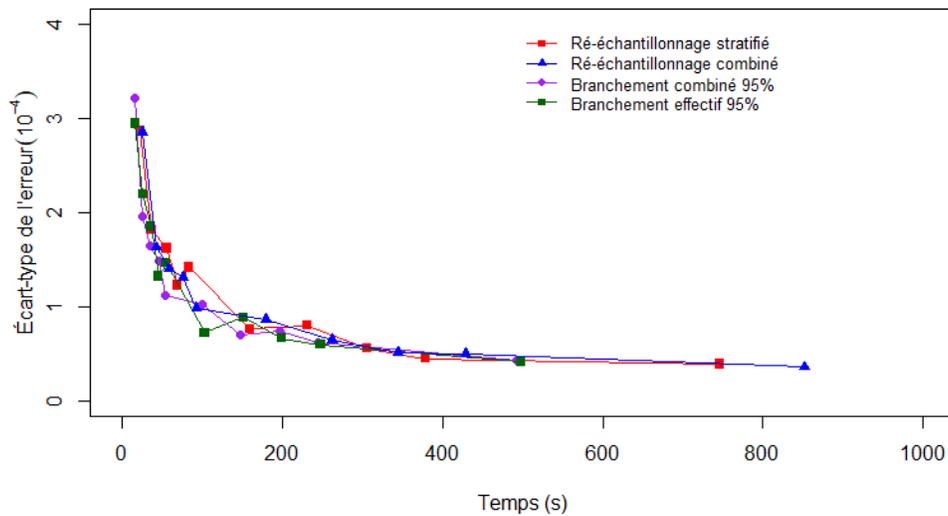


Figure 4.19: Comparaison des écart-type de l'erreur de prédiction des meilleurs filtres à particules avec ré-échantillonnage et avec branchement par rapport au temps d'exécution de 50 filtres.

En analysant les figures 4.18 et 4.19, nous pouvons constater que les filtres à parti-

cules avec ré-échantillonnage stratifié et combiné et ceux avec branchement ayant un taux de contamination de 95% semblent mener à des erreurs de prédiction de même ordre pour des temps de calcul identiques. Nous observons aussi que ces quatre filtres suivent presque exactement la même courbe pour l'écart-type de cette erreur. Ce résultat est intéressant puisqu'il démontre la capacité des filtres à particules avec branchement, dont l'intervalle c_k est choisi de façon adéquate, à performer au moins aussi bien que les filtres à particules avec ré-échantillonnage dans le cadre de ce problème.

4.3 Optimisation des paramètres à l'aide des filtres à particules

4.3.1 Technique d'optimisation du modèle de Heston à l'aide des filtres à particules

Pour estimer les paramètres du modèle de Heston par maximum de vraisemblance, nous fixons un germe aléatoire (*random seed*) et nous sélectionnons les paramètres du modèle par maximum de vraisemblance.

Nous avons utilisé la technique d'optimisation de Nelder et Mead (1965) afin de trouver les paramètres optimaux du modèle pour ce germe aléatoire. Bien que l'algorithme de Nelder-Mead est considéré lent, il est assez robuste et s'adapte bien à ce problème puisqu'il permet l'optimisation de fonctions objectives non différentiables. Cette caractéristique est nécessaire à l'optimisation par filtres à particules puisque la vraisemblance est obtenue par simulations. Comme la vitesse d'exécution n'était pas un but principal de ce mémoire, l'algorithme de Nelder-Mead s'est avéré approprié.

Ces optimisations ont été effectuées à l'aide du logiciel de programmation R. Par rapport à un langage de programmation plus complexe comme Rcpp, R simplifie la génération des variables aléatoires et rend les germes aléatoires plus faciles à fixer. Ceci fait en sorte qu'il n'est pas nécessaire d'emmagasiner toutes les variables aléatoires à l'avance afin d'effectuer l'optimisation ce qui alourdit nettement les calculs et vient annuler les bénéfices qui devraient être observés en utilisant Rcpp.

Toutefois, il serait possible d'améliorer les temps de calcul en utilisant ce progiciel si les deux failles mentionnées étaient comblées.

Un des problèmes majeurs qu'il est possible de constater lorsque nous tentons d'optimiser les paramètres d'un modèle à l'aide du filtre à particules est que la fonction de vraisemblance n'est pas lisse. Ceci est dû au fait que la fonction de vraisemblance dépend d'un grand nombre de variables simulées. Les paramètres initiaux change le poids de toutes les particules et fortement modifier celles qui seront branchées ou ré-échantillonnées. Une des solutions qui peut légèrement améliorer le lissage de la fonction de vraisemblance est l'augmentation du nombre de particules. Cette solution permet d'observer une plus grande diversité dans l'échantillon des particules et permet d'observer une fonction de répartition de la volatilité stochastique plus stable. Ceci fait en sorte que l'impact causé par le caractère aléatoire du filtre à particules sur la fonction de vraisemblance est diminué ce qui la rend plus lisse.

Dans cette optique, nous avons choisi d'effectuer l'optimisation pour $N = 5\,000$ particules pour les filtres à particules avec ré-échantillonnage et $N_0 = 5\,000$ particules initiales pour les filtres à particules avec branchement. Ce nombre de particules a été sélectionné parce qu'il permettait d'obtenir les résultats par maximum de vraisemblance relativement rapidement tout en offrant une fonction de vraisemblance relativement plus lisse que lorsque le nombre de particules initiales est inférieur. Malgré que la vraisemblance reste très peu lisse à ce nombre de particules, l'ajout de particules a très peu d'impact sur le lissage tout en ajoutant un fort coût temporel. Il semblait alors peu justifiable d'augmenter le nombre de particules davantage. De plus ce choix du nombre de particules semble cohérent avec le reste de la littérature.

Pour obtenir un intervalle de confiance sur notre optimisation pour les paramètres obtenus par maximisation de la vraisemblance, nous avons aussi décidé de répéter l'optimisation pour 50 germes aléatoires différents. De plus, le fait de répéter l'expérience une 50 fois pour différents germes aléatoires nous permet de confirmer la robustesse des paramètres prédits par l'optimisation.

Les estimations obtenues pour chacun des paramètres sont présentées dans le tableau 4.3 pour les filtres avec ré-échantillonnage et dans le tableau 4.5 pour les filtres avec branchement. Les écarts-types de chaque estimation sont donnés entre parenthèses et représentent l'écart-type des paramètres prédits par maximum de vraisemblance sur les 50 germes.

4.3.2 Filtres à particules avec ré-échantillonnage

Tableau 4.3: Paramètres estimés du modèles de Heston par les filtres à particules avec ré-échantillonnage

	κ	θ	σ	ρ	$LogL(\beta)$
Paramètres réels	6.000	0.20000	0.5000	-0.7000	—
Filtre ISR	5.923	0.22572	0.4486	-0.6334	5257.63
	(0.277)	(0.00192)	(0.0155)	(0.0172)	(0.26)
Filtre ISR résiduel	5.974	0.22604	0.4496	-0.6311	5257.52
	(0.293)	(0.00194)	(0.0155)	(0.0150)	(0.14)
Filtre ISR stratifié	7.199	0.22521	0.4817	-0.6247	5257.26
	(0.063)	(0.00017)	(0.0028)	(0.0019)	(0.09)
Filtre ISR combiné	5.965	0.22595	0.4396	-0.6295	5257.26
	(0.271)	(0.00119)	(0.0106)	(0.0101)	(0.07)

Les paramètres estimés par les filtres à particules avec ré-échantillonnage standard, résiduel et combiné convergent tous vers les mêmes valeurs. Toutefois, le filtre à particules stratifié estime des paramètres très différents pour les paramètres κ et σ . Ceci peut s'expliquer par le fait qu'il existe probablement plusieurs jeux de paramètres permettant de générer la série des log-rendements montrée à la figure 4.1. L'estimation des paramètres à l'aide du filtre à particules stratifié a une incertitude beaucoup plus faible que les estimations faites à l'aide des autres filtres. Ceci peut s'expliquer par le fait que pour un même nombre de particules, le filtre à particules avec ré-échantillonnage stratifié est celui obtenant le plus petit écart-type sur les erreurs.

Ceci signifie qu'il est moins influencé par le germe utilisé et que la vraisemblance est probablement plus lisse. Sachant cela, il n'est pas surprenant que l'incertitude sur des paramètres optimisés soit plus petite.

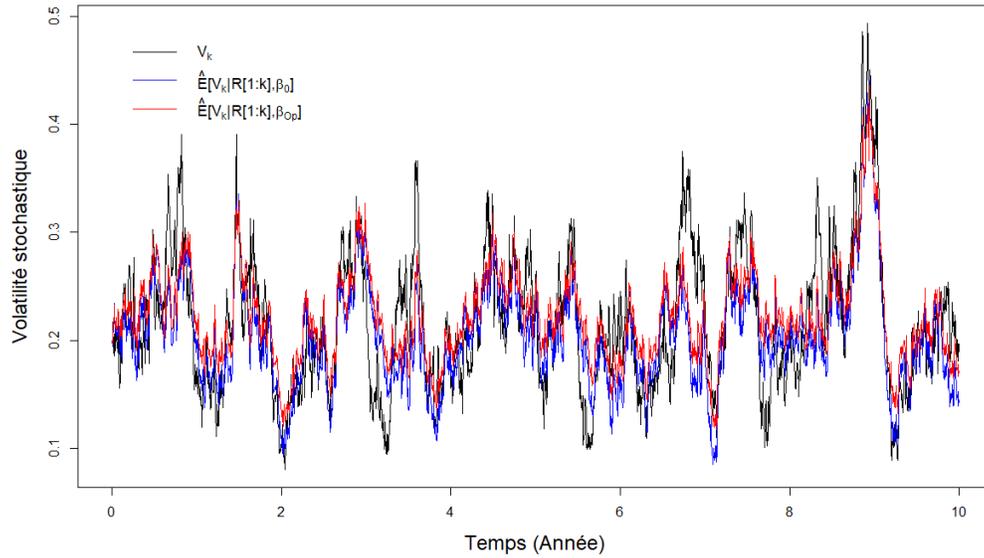


Figure 4.20: Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec ré-échantillonnage standard avec les paramètres initiaux et estimés par rapport à la volatilité réelle de la trajectoire, ($N = 10\,000$)

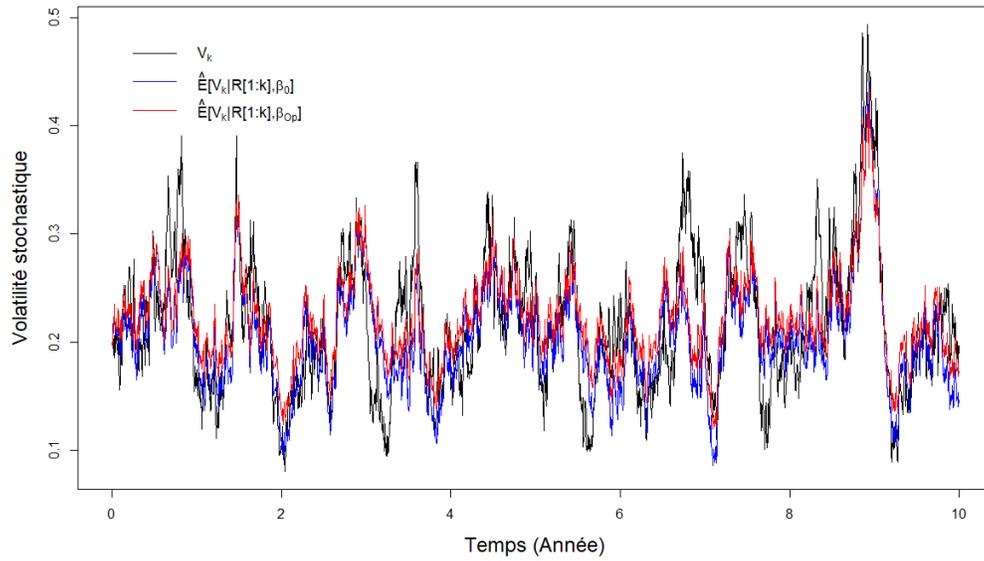


Figure 4.21: Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec ré-échantillonnage stratifié avec les paramètres initiaux et estimés par rapport à la volatilité réelle de la trajectoire, ($N = 10\,000$)

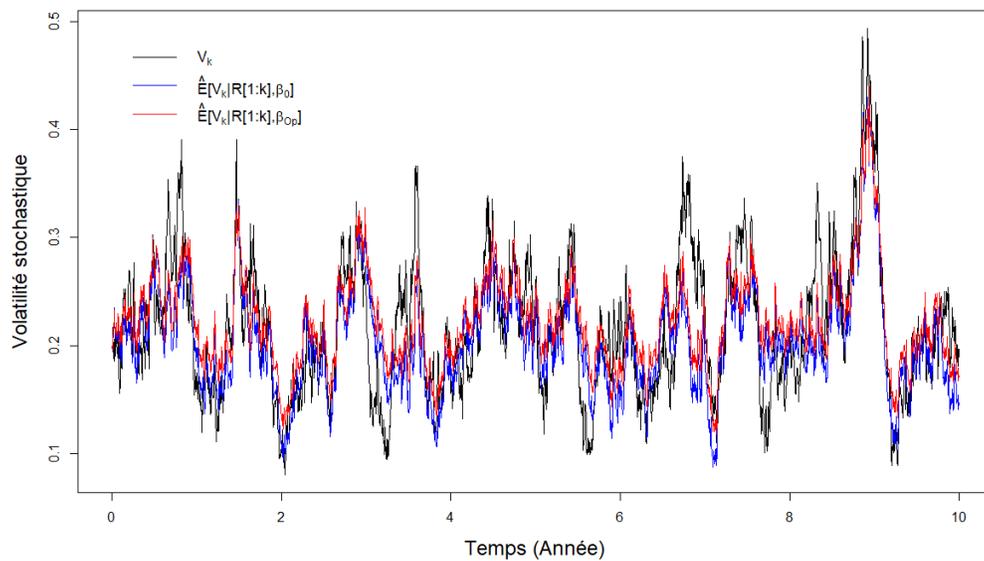


Figure 4.22: Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec ré-échantillonnage résiduel avec les paramètres initiaux et estimés par rapport à la volatilité réelle de la trajectoire, ($N = 10\,000$)

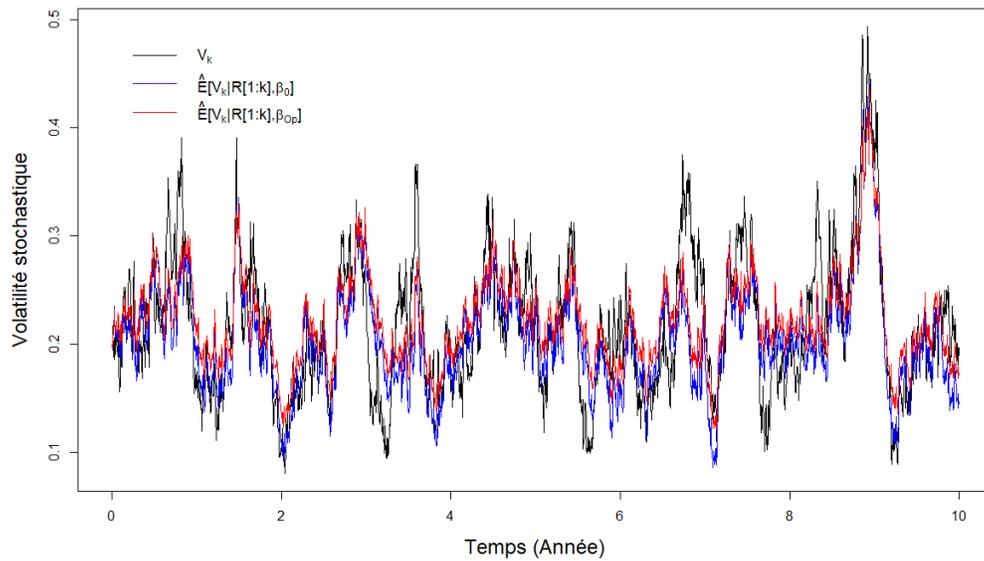


Figure 4.23: Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec ré-échantillonnage combiné avec les paramètres initiaux et estimés par rapport à la volatilité réelle de la trajectoire, ($N = 10\,000$)

Maintenant que nous avons estimé les paramètres du modèle de Heston, nous voulons évaluer si l'utilisation de ces paramètres estimés permet d'améliorer la prédiction de la densité de la variable latente. Pour ce faire, nous utiliserons les paramètres estimés, que nous noterons β_{Op} , pour appliquer les différents filtres à particules et nous comparerons leurs performances pour voir lesquels sont les plus pertinents afin de déterminer la volatilité stochastique du titre.

Tel que montré par les figures 4.20, 4.21, 4.22 et 4.23, lorsque nous comparons les estimations de l'espérance de la volatilité stochastique pour les filtres à particules générés à partir des paramètres estimés $E[V_k | R_{1:k}, \beta_{Op}]$, nous observons que ceux-ci sont, la plupart du temps, légèrement au-dessus des estimations de la volatilité stochastique calculées avec les paramètres initiaux $E[V_k | R_{1:k}, \beta_0]$. Ceci permet de mieux estimer les hausses abruptes de la volatilité stochastique, mais, inversement, diminue la capacité du filtre à estimer la baisse de la volatilité stochastique. Ces estimations sont, en général, plus proches de la valeur réelle que les estimations faites avec les paramètres initiaux. Lorsque les paramètres obtenus par optimisation

sont utilisés, le filtre à particules avec ré-échantillonnage stratifié performe aussi bien que les autres filtres afin d'estimer la volatilité stochastique, malgré le fait que ses paramètres estimés diffèrent de ceux des autres filtres. Ceci pourrait indiquer la présence de multiples maximums dans la fonction de vraisemblance du modèle.

Tableau 4.4: Erreur de prédiction par les filtres à particules avec ré-échantillonnage avec des paramètres optimisés et son écart-type, $N = 10\,000$

	Erreur (10^{-2})	Écart-type (10^{-4})
Filtre ISR	4.14120	1.31
Filtre ISR résiduel	4.14769	0.83
Filtre ISR stratifié	4.15893	0.30
Filtre ISR combiné	4.16207	0.29

4.3.3 Filtres à particules avec branchement

Tableau 4.5: Paramètres estimés du modèles de Heston par les filtres à particules avec branchement

	κ	θ	σ	ρ	$LogL(\beta)$
Paramètres réels	6.000	0.20000	0.5000	-0.7000	—
Branchement combiné (95%)	5.942	0.22638	0.4421	-0.6340	5257.28
	(0.206)	(0.00165)	(0.0109)	(0.0093)	(0.09)
Branchement combiné (70%)	5.952	0.22571	0.4442	-0.6285	5257.43
	(0.283)	(0.00200)	(0.0153)	(0.0126)	(0.11)
Branchement combiné (30%)	5.986	0.22594	0.4429	-0.6309	5257.47
	(0.268)	(0.00211)	(0.0185)	(0.0118)	(0.10)
Branchement effectif (95%)	5.942	0.22599	0.4408	-0.6304	5257.32
	(0.236)	(0.00184)	(0.0156)	(0.0128)	(0.10)
Branchement effectif (70%)	5.991	0.22593	0.4431	-0.6367	5257.50
	(0.316)	(0.00196)	(0.0137)	(0.0136)	(0.13)
Branchement effectif (30%)	6.012	0.22569	0.4396	-0.6318	5257.47
	(0.264)	(0.00213)	(0.0153)	(0.0125)	(0.12)

Les filtres à particules avec branchements convergent tous vers les mêmes paramètres (plus ou moins l'écart-type) que les filtres à particules avec ré-échantillonnage standard, avec ré-échantillonnage résiduel et avec ré-échantillonnage combiné. Par contre, il est possible de remarquer que, en général, les paramètres estimés par filtres à particules avec branchement ont des écarts-types plus petits que ceux des filtres à particules standard et résiduel, mais plus élevé que ceux combiné et stratifié. Toutefois, le filtre à particules avec branchement combiné à taux de contagion de 90% peut être considéré comme une exception puisque ses écarts-types sont du même ordre que ceux du filtre à particules avec ré-échantillonnage combiné.

D'après nos observations et les figures 4.24, 4.24, 4.25, 4.26, 4.27 et 4.28, les valeurs

prédites de V_k à l'aide du filtre à particules avec branchement utilisant les paramètres optimisés semblent avoir la même dynamique que celles prédites à l'aide des paramètres optimisés des filtres à particules avec ré-échantillonnage.

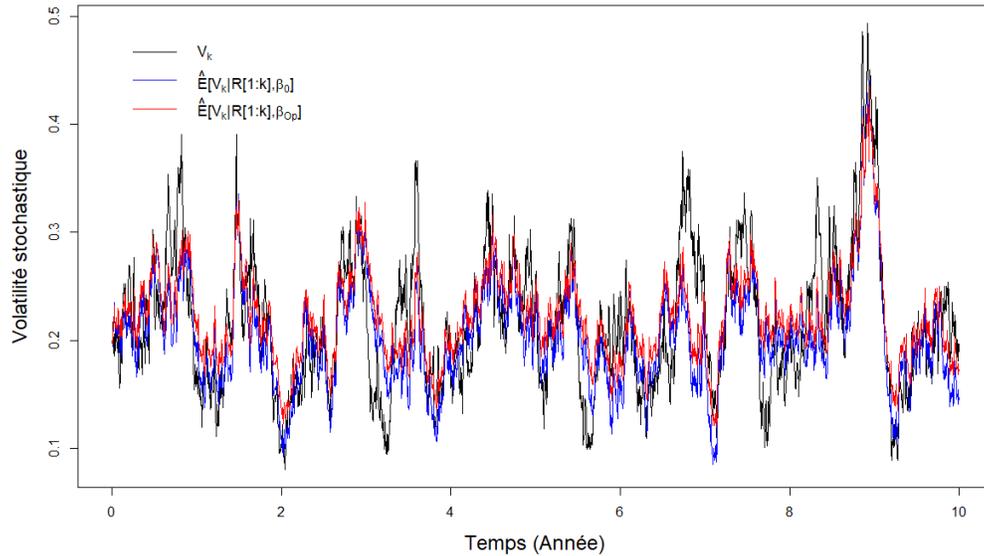


Figure 4.24: Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement combiné avec un taux de contagion de 95% avec les paramètres initiaux et estimés par rapport à la volatilité réelle de la trajectoire, $N_0 = 10\,000$

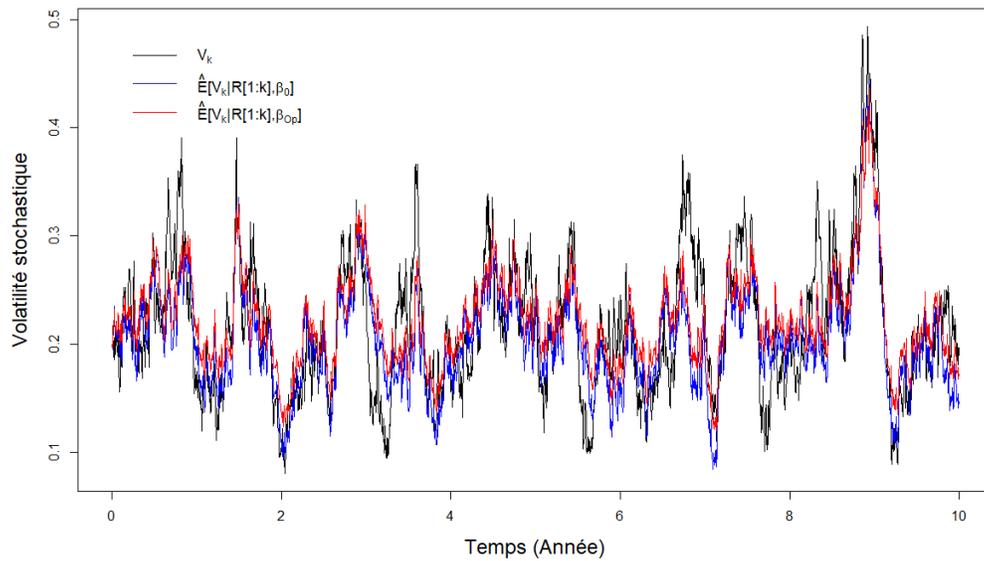


Figure 4.24: Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement combiné avec un taux de contagion de 70% avec les paramètres initiaux et estimés par rapport à la volatilité réelle de la trajectoire, $N_0 = 10\,000$

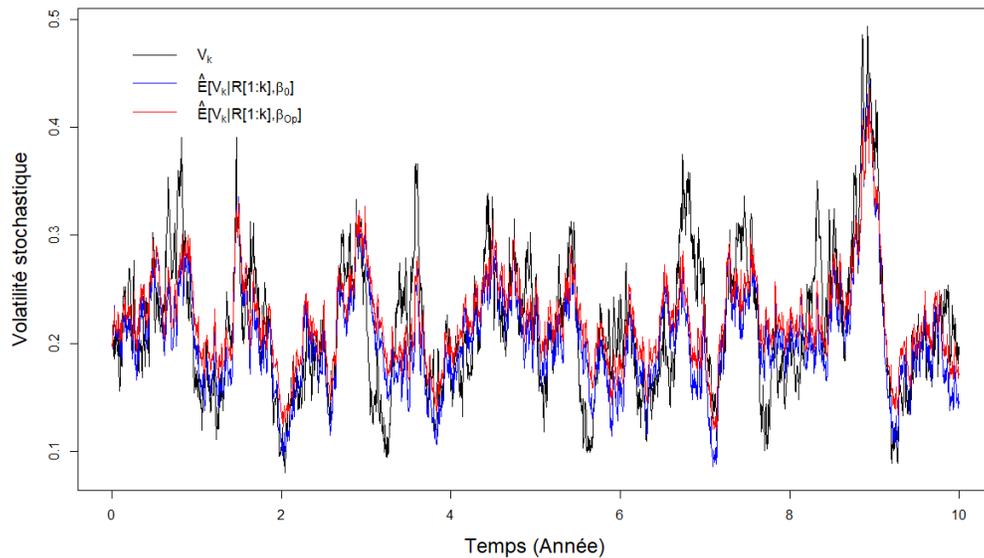


Figure 4.25: Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement combiné avec un taux de contagion de 30% avec les paramètres initiaux et estimés par rapport à la volatilité réelle de la trajectoire, $N_0 = 10\,000$

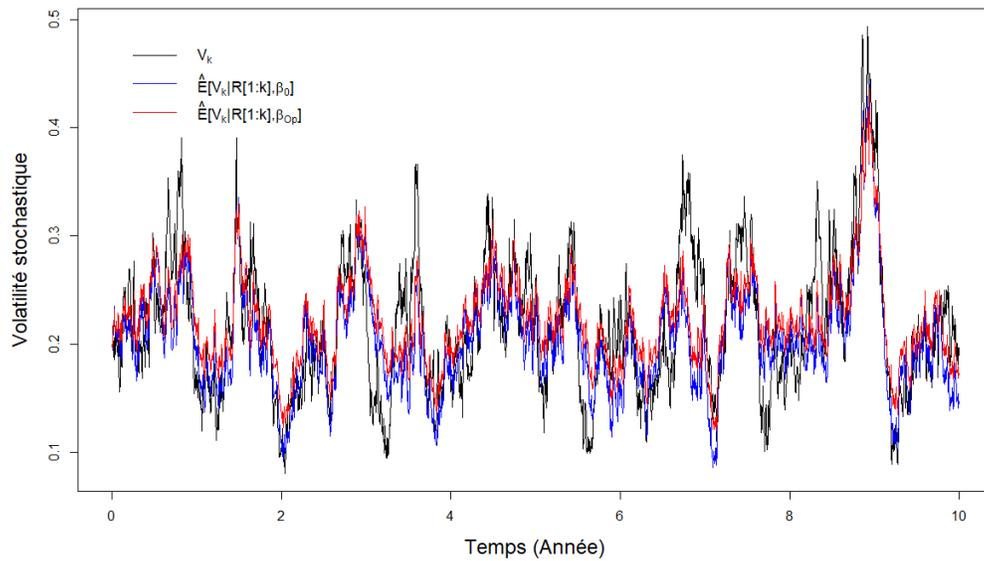


Figure 4.26: Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement effectif avec un taux de contagion de 95% avec les paramètres initiaux et estimés par rapport à la volatilité réelle de la trajectoire, $N_0 = 10\,000$

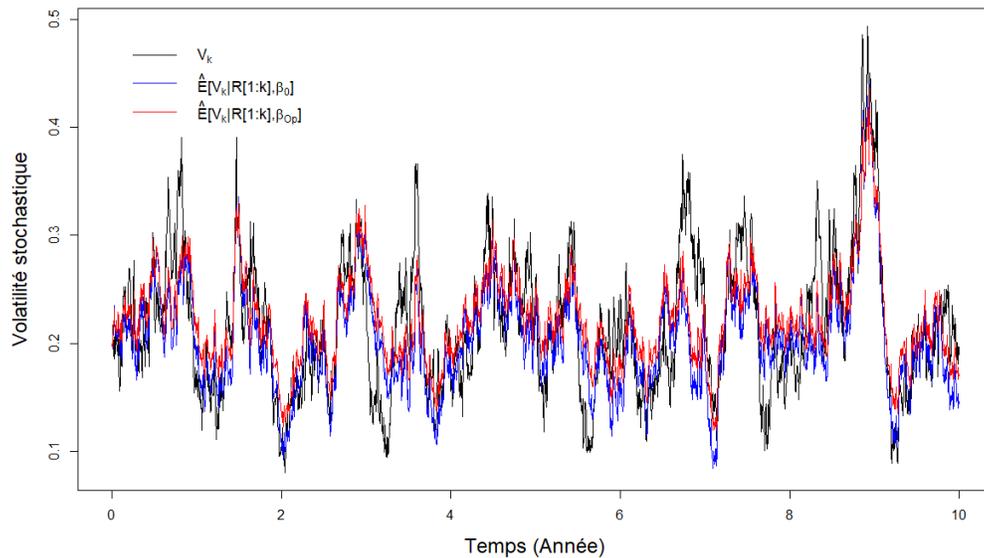


Figure 4.27: Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement effectif avec un taux de contagion de 70% avec les paramètres initiaux et estimés par rapport à la volatilité réelle de la trajectoire, $N_0 = 10\,000$

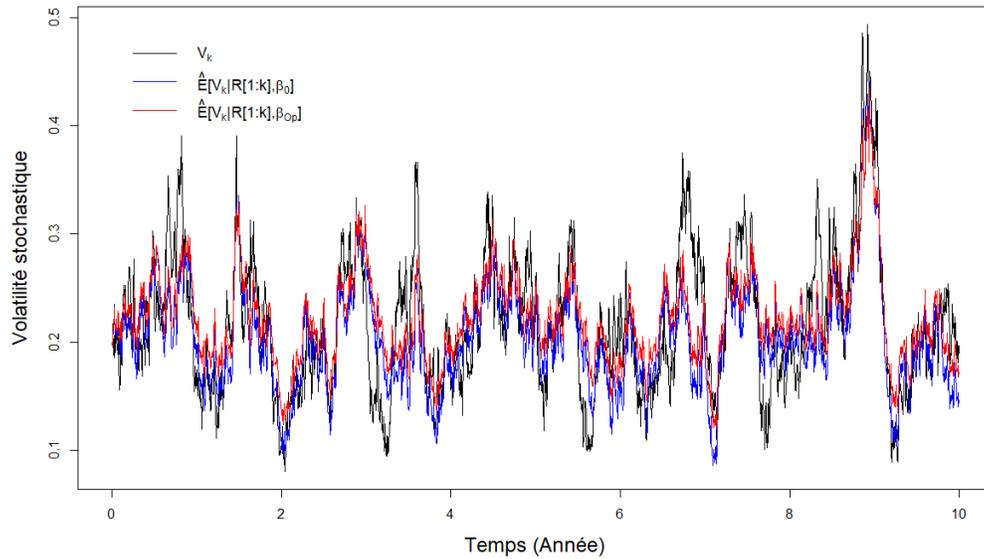


Figure 4.28: Comparaison de l'espérance de la volatilité stochastique estimée par différents filtres à particules avec branchement effectif avec un taux de contagion de 30% avec les paramètres initiaux et estimés par rapport à la volatilité réelle de la trajectoire, $N_0 = 10\,000$

En étudiant le tableau 4.6, il peut être constaté que les filtres à particules avec branchement ont des erreurs de prédictions similaires à celles des filtres à particules avec ré-échantillonnage. Toutefois, leur écart-type se retrouve sous celle du filtre avec ré-échantillonnage standard et résiduel, mais au-dessus de stratifié et combiné.

Tableau 4.6: Erreur de prédiction par les filtres à particules avec branchement avec des paramètres optimisés et son écart-type, $N_0 = 10\,000$

	Erreur (10^{-2})	Écart-type (10^{-4})
Branchement combiné (95%)	4.1634	0.44
Branchement combiné (70%)	4.1485	0.54
Branchement combiné (30%)	4.1577	0.49
Branchement effectif (95%)	4.1592	0.41
Branchement effectif (70%)	4.1620	0.49
Branchement effectif (30%)	4.1655	0.56

4.4 Comparaison des résultats des différents filtres

D'après notre analyse, nous remarquons que les filtres à particules avec branchement ayant un taux de contagion élevé performant mieux que ceux ayant un taux faible. En effet, ces filtres génèrent des erreurs de prédiction plus petites et ont des écarts-types plus petits sur cette erreur. Ceux-ci ont aussi des temps d'exécution plus courts. Les filtres à particules avec branchement effectif performant de façon similaire à ceux à branchement combiné.

Par rapport aux filtres à particules avec ré-échantillonnage, les filtres à particules avec branchement prédisent des paramètres similaires aux filtres à particules avec ré-échantillonnage standard, résiduel et combiné. Il est intéressant de noter que malgré ceci, les paramètres obtenus par l'optimisation du filtre à particules avec ré-échantillonnage stratifié ne sont pas à éliminer puisque ceux-ci performant aussi bien que les autres pour estimer la volatilité stochastique. Les filtres à particules avec branchement sont légèrement plus lents que les filtres à particules avec ré-échantillonnage standard et résiduel pour un même nombre de particules initiales. Toutefois, les filtres à particules avec branchement ont, pour un même temps d'exécution, une plus petite erreur de prédiction et un plus petit écart-type sur cette erreur. Par contre, ces filtres sont plus rapides, mais performant moins bien que les filtres à particules avec ré-échantillonnage stratifié et combiné sur ces deux dernières métriques pour un même nombre de particules initiales. Malgré cela, il est possible d'observer que lorsque le taux de contagion est plus élevé, les filtres à particules avec branchement peuvent performer au moins aussi bien que même les filtres à particules avec ré-échantillonnage combiné ou stratifié pour un même temps d'exécution.

Il devrait être possible, avec des méthodes appropriées pour calibrer les modèles avec branchement, de trouver un filtre performant mieux que les filtres à particules avec ré-échantillonnage. Toutefois, si les bornes de l'intervalle c_k ne sont pas choisies de façon à optimiser la performance du filtre, les filtres à particules avec branchement n'apportent que peu d'avantages par rapport aux filtres à particules

avec ré-échantillonnage dans le cadre de ce problème.

CONCLUSION

Dans ce mémoire, nous avons décrit les filtres à particules et l'algorithme avec branchement dont nous voulions évaluer la performance. Ce nouvel algorithme a comme caractéristique que toutes les particules dont le poids est dans un intervalle autour du poids moyen sont conservées telles quelles et les autres sont ré-échantillonnées de façon aléatoire. Cet algorithme a pour avantage qu'il permet de réduire le nombre de particules qui sont ré-échantillonnées et de simplifier la formule du ré-échantillonnage à une partie déterministe et une loi de Bernoulli. Ceci permet de réduire le bruit causé par le ré-échantillonnage, de réduire les temps de calcul et de conserver la mémoire de certaines particules. Nous avons ensuite décrit le modèle de Heston (1993) et nous avons décrit certaines discrétisations du modèle ainsi qu'une démarche pour appliquer le filtre à particules sur ce modèle. Par la suite, nous avons comparé l'application des filtres à particules avec ré-échantillonnage complet et des filtres à particules avec branchement au modèle de Heston. Nous avons estimé par maximum de vraisemblance les paramètres d'une série d'observations à l'aide de chacun des filtres et nous avons comparé l'erreur de prédiction de chacun des filtres avec les paramètres réels et les paramètres estimés. Nous avons aussi démontré que les filtres à particules avec branchement peuvent, lorsque l'intervalle est choisi de façon optimale, mieux performer que les filtres à particules avec ré-échantillonnage. Pour la trajectoire de prix analysée, il est possible, en augmentant le taux de contagion, de réduire l'erreur de prédiction et son écart-type pour les algorithmes avec branchement ayant des temps de calcul équivalents. Il est même possible d'observer des filtres à particules avec branchement qui performant mieux que les meilleurs filtres à particules avec ré-échantillonnage. Leur utilisation permet aussi de bien estimer les paramètres du modèle de Heston par maximum de vraisemblance.

Dans de futurs travaux, il serait intéressant d'analyser des modèles avec branchement

en changeant la forme du multiplicateur définissant les bornes de l'intervalle de branchement. En effet, cette borne a un fort impact sur la performance des filtres à particules avec branchement et il pourrait être intéressant de trouver les intervalles de branchement permettant d'optimiser la performance de ces filtres à particules.

RÉFÉRENCES

- Arulampalam, M. S., Maskell, S., Gordon, N. et Clapp, T. (2002). A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. *IEEE Transactions on Signal Processing*, 50(2), 174–188.
- Bégin, J.-F. (2012). *New Simulation Schemes for the Heston Model*. (Thèse de doctorat). Université de Montréal.
- Broadie, M. et Özgür Kaya (2006). Exact Simulation of Stochastic Volatility and Other Affine Jump Diffusion Processes. *Operations Research*, 54(2), 217–231.
- Christoffersen, P., Jacobs, K. et Mimouni, K. (2007). Volatility Dynamics for the S&P500 : Evidence from Realized Volatility, Daily Returns and Option Prices. *Review of Financial Studies*, 23.
- Cox, J. C., Ingersoll, J. E. et Ross, S. A. (1985). A Theory of the Term Structure of Interest Rates. *Econometrica*, 53(2), 385–407.
- Creal, D. (2012). A Survey of Sequential Monte Carlo Methods for Economics and Finance. *Econometric Reviews*, 31(3), 245–296.
- Crisan, D. et Lyons, T. (1997). Nonlinear Filtering and Measure-Valued Processes. *Probability Theory and Related Fields*, 109, 217–244.
- Douc, R. et Cappe, O. (2005). Comparison of Resampling Schemes for Particle Filtering. Dans *ISPA 2005. Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis, 2005.*, 64–69.
- Feller, W. (1951). Two Singular Diffusion Problems. *Annals of Mathematics*, 54(1), 173–182.

- Gordon, N. J., Salmond, D. J. et Smith, A. F. M. (1993). Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation. *IEE Proceedings F - Radar and Signal Processing*, 140(2), 107–113.
- Hammersley, J. M. et Morton, K. W. (1954). Poor Man's Monte Carlo. *Journal of the Royal Statistical Society. Series B (Methodological)*, 16(1), 23–38.
- Heston, S. L. (1993). A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Review of Financial Studies*, 6(2), 327–43.
- Kahn, H. et Marshall, A. W. (1953). Methods of Reducing Sample Size in Monte Carlo Computations. *Journal of the Operations Research Society of America*, 1(5), 263–278.
- Kim, S., Shephard, N. et Chib, S. (1998). Stochastic Volatility : Likelihood Inference and Comparison with ARCH Models. *The Review of Economic Studies*, 65(3), 361–393.
- Kitagawa, G. (1996). Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models. *Journal of Computational and Graphical Statistics*, 5(1), 1–25.
- Kloeden, P. et Platen, E. (2011). *The Numerical Solution of Stochastic Differential Equations*, volume 23. Springer.
- Kong, A., Liu, J. S. et Wong, W. H. (1994). Sequential Imputations and Bayesian Missing Data Problems. *Journal of the American Statistical Association*, 89(425), 278–288.
- Kouritzin, M. (2016). Residual and Stratified Branching Particle Filters. *Computational Statistics & Data Analysis*, 111.
- Liu, J. S. et Chen, R. (1998). Sequential Monte Carlo Methods for Dynamic Systems. *Journal of the American Statistical Association*, 93(443), 1032–1044.

- Maruyama, G. (1954). On the Transition Probability Functions of the Markov Process. *Nat. Sci. Rep. Ochanomizu Univ.*, 5, 10–20.
- Milstein, G. N. (1975). Approximate Integration of Stochastic Differential Equations. *Theory of Probability & Its Applications*, 19(3), 557–562.
- Nelder, J. et Mead, R. (1965). A Simplex Method for Function Minimization. *Computer Journal*, 7, 308–313.
- Rouah, F. D. (2011). Euler and Milstein Discretization. *Documento de trabajo, Sapient Global Markets, Estados Unidos. Recuperado de www.frouah.com.*