

Filtrage particulaire optimal et filtrage particulaire auxiliaire adapté : Une analyse non asymptotique

François DESBOUVRIES, Yohan PETETIN, Emmanuel MONFRINI

¹Télécom SudParis / Département CITI & CNRS UMR 5157, 9 rue Charles Fourier, 91011 Evry, France
{Francois.Desbouvries, yohan.petetin, emmanuel.monfrini}@it-sudparis.eu

Résumé – Le filtrage particulaire (FP) et le filtrage particulaire auxiliaire (FPA) sont deux méthodes classiques de Monte Carlo séquentielles (MCS) pour estimer la densité de filtrage $p_{n|n}$ dans une chaîne de Markov cachée. Sur un plan théorique ces algorithmes ont été analysés du point de vue de la statistique asymptotique. Dans cet article nous proposons une analyse comparative *non asymptotique, pour un nombre fini de particules* de deux algorithmes MCS : l’algorithme SIR optimal, et l’algorithme FPA totalement adapté (FA). Partant d’un nuage commun de N particules, nous calculons la moyenne et la variance d’estimateurs empiriques d’un moment de $p_{n|n}$. Les deux algorithmes ont la même moyenne, mais en cas de rééchantillonnage la variance de l’algorithme SIR est toujours supérieure à celle de l’algorithme FA. Nous soulignons finalement le rôle du nombre de particules efficaces dans l’excès de variance.

Abstract – Particle filters (PF) and auxiliary particle filters (APF) are widely used SMC techniques for estimating the filtering pdf $p_{n|n}$ in a hidden Markov chain. These algorithms have been theoretically analysed from an asymptotical statistics perspective. In this paper we provide a *non asymptotical, finite number of samples* comparative analysis of two SMC algorithms : the Sampling Importance Resampling (SIR) PF with optimal importance distribution, and the fully adapted APF (FA). Starting from a common set of N particles, we compute the mean and variance of empirical estimators of a moment of $p_{n|n}$. Both algorithms have the same mean, but if resampling is used, the variance of the SIR algorithm always exceeds that of the FA one. We finally outline the role of the number of efficient particles in the excess of variance.

1 Introduction

Soit X_n (resp. Y_n) un processus caché (resp. observé). Soit $p(x_n|\mathbf{y}_{0:n})$ (ou $p_{n|n}$) la densité de x_n sachant $\mathbf{y}_{0:n} = \{y_i\}_{i=0}^n$. On suppose que $\{X_n, Y_n\}_{n \in \mathbb{N}}$ est une chaîne de Markov cachée : $p(\mathbf{x}_{0:n}, \mathbf{y}_{0:n}) = p(x_0) \prod_{i=1}^n p(x_i|x_{i-1}) \prod_{i=0}^n p(y_i|x_i)$. Le filtrage bayésien consiste à calculer $p(x_n|\mathbf{y}_{0:n})$, ou au moins une approximation de la mesure $p(dx_n|\mathbf{y}_{0:n})$ de densité $p_{n|n}$. $p_{n|n}$ peut être calculée à partir de $p_{n-1|n-1}$ via

$$p(x_n|\mathbf{y}_{0:n}) = \frac{p(y_n|x_n) \int p(x_n|x_{n-1})p(x_{n-1}|\mathbf{y}_{0:n-1})dx_{n-1}}{p(y_n|\mathbf{y}_{0:n-1})} \quad (1)$$

Si le calcul exact de (1) est impossible on recourt à des approximations, parmi lesquelles le FP et le FPA (cf. par ex. [1]-[3]) sont des méthodes MCS qui propagent une approximation discrète $\hat{p}(dx_n|\mathbf{y}_{0:n})$ de $p(dx_n|\mathbf{y}_{0:n})$. D’un point de vue théorique, des résultats de convergence de certains algorithmes de FP ont été établis, voir par ex. [4]-[7] et leurs références. Ces résultats ont récemment été étendus au FPA [8].

Tous ces résultats sont asymptotiques, c’est-à-dire qu’ils sont vérifiés lorsque le nombre de particules tend vers l’infini. Dans cet article nous proposons une analyse comparative *non asymptotique* d’un algorithme de FP : l’algorithme SIR avec densité d’importance optimale [9], et l’algorithme FPA totalement adapté (FA) [2]. Notre article est organisé de la façon suivante. Une analyse au second ordre des algorithmes SIR et FA est proposée au §2. Nous en discutons les résultats au §3 (et analysons en

particulier le rôle du nombre efficace de particules dans l’excès de variance), et nous effectuons des simulations au §4.

2 Une analyse non asymptotique des algorithmes SIR et FA

2.1 Algorithmes SIR et FA

Rappelons tout d’abord brièvement les algorithmes SIR et FA. L’algorithme SIR avec densité d’importance optimale [9] consiste en la succession d’une étape d’échantillonnage S , une réactualisation des poids W et une étape (optionnelle) de rééchantillonnage R :

Algorithme SIR (avec densité optimale et rééchantillonnage à chaque pas).

Soit $\hat{p}(dx_{n-1}|\mathbf{y}_{0:n-1}) = \sum_{i=1}^N \frac{1}{N} \delta_{x_{n-1}^{\text{SIR},i}}$.

S . pour $1 \leq i \leq N$, tirer $x_n^{\text{IS},i} \sim p(x_n|x_{n-1}^{\text{SIR},i}, y_n)$;

W . Pour $1 \leq i \leq N$, calculer $w_n^i \propto p(y_n|x_{n-1}^{\text{SIR},i})$, $\sum_{i=1}^N w_n^i = 1$;

R . Pour $1 \leq i \leq N$, tirer $x_n^{\text{SIR},i} \sim \sum_{i=1}^N w_n^i \delta_{x_n^{\text{IS},i}}$.

Alors $\hat{p}(dx_n|\mathbf{y}_{0:n}) = \sum_{i=1}^N \frac{1}{N} \delta_{x_n^{\text{SIR},i}} \simeq p(dx_n|\mathbf{y}_{0:n})$.

Venons en au FPA [2] [3]. Remplaçons dans (1) $p(dx_{n-1}|\mathbf{y}_{0:n-1})$ par une approximation discrète $\sum_{i=1}^N \frac{1}{N} \delta_{x_{n-1}^i}$. On obtient une

approximation continue π_n de $p_{n|n}$:

$$\pi_n(x_n) = \sum_{i=1}^N \frac{p(y_n|x_{n-1}^i)}{\sum_{i=1}^N p(y_n|x_{n-1}^i)} p(x_n|x_{n-1}^i, y_n). \quad (2)$$

Il reste à obtenir des échantillons x_n^i . S'il est impossible de tirer selon π_n on peut utiliser l'échantillonnage d'importance, ce qui fournit la classe des algorithmes de FPA. Si en revanche on peut tirer directement des échantillons selon π_n on est dans le cas *totalemt adapté*, selon la terminologie de [2] :

Algorithme FA

Soit $\hat{p}(dx_{n-1}|y_{0:n-1}) = \sum_{i=1}^N \frac{1}{N} \delta_{x_{n-1}^{\text{FA},i}}$.

W. Pour $1 \leq i \leq N$, calculer $w_n^i \propto p(y_n|x_{n-1}^{\text{FA},i})$, $\sum_{i=1}^N w_n^i = 1$;

R. Pour $1 \leq i \leq N$, tirer $\tilde{x}_{n-1}^i \sim \sum_{i=1}^N w_n^i \delta_{x_{n-1}^{\text{FA},i}}$;

S. Pour $1 \leq i \leq N$, tirer $x_n^{\text{FA},i} \sim p(x_n|\tilde{x}_{n-1}^i, y_n)$.

Alors $\hat{p}(dx_n|y_{0:n}) = \sum_{i=1}^N \frac{1}{N} \delta_{x_n^{\text{FA},i}} \simeq p(dx_n|y_{0:n})$.

Comme on le constate l'algorithme FA est un *réordonnement* de l'algorithme SIR. Dans l'algorithme SIR les étapes successives sont $S \rightarrow W \rightarrow R$ alors que la boucle de l'algorithme FA est constituée des étapes successives $W \rightarrow R \rightarrow S$.

2.2 Moments du second ordre d'estimateurs d'un moment d'intérêt

Soit désormais $\mathbf{x}_n^{\{i\}} \stackrel{\text{def}}{=} \{x_n^i\}_{i=1}^N$, $\alpha_i \stackrel{\text{def}}{=} \frac{p(y_n|x_{n-1}^i)}{\sum_{i=1}^N p(y_n|x_{n-1}^i)}$, $p_i(\cdot) \stackrel{\text{def}}{=} p(\cdot|x_{n-1}^i, y_n)$, $\mu_i(f) \stackrel{\text{def}}{=} E_{p_i}(f(X))$ et $\sigma_i(f) \stackrel{\text{def}}{=} \text{var}_{p_i}(f(X))$. Supposons que le but est de calculer un moment $\Theta_n = E_{p_{n|n}}(f(X))$ (où $f(\cdot)$ est une fonction d'intérêt) de la vraie densité $p_{n|n}$. À $n-1$ on dispose de N points $\mathbf{x}_{n-1}^{\{i\}}$ (tirés approximativement selon $p_{n-1|n-1}$). Partant de cet ensemble commun de points, l'algorithme FA (resp. SIR) produit un nouveau nuage de points $\mathbf{X}_n^{\text{FA},\{i\}}$ (resp. $\mathbf{X}_n^{\text{IS},\{i\}}$ et $\mathbf{X}_n^{\text{SIR},\{i\}}$) qui à son tour permet de calculer les estimateurs suivants de Θ_n :

$$\hat{\Theta}_n^{\text{FA}}(\mathbf{X}_n^{\text{FA},\{i\}}) = \frac{1}{N} \sum_{i=1}^N f(X_n^{\text{FA},i}), \quad (3)$$

$$\hat{\Theta}_n^{\text{IS}}(\mathbf{X}_n^{\text{IS},\{i\}}) = \sum_{i=1}^N \alpha_i f(X_n^{\text{IS},i}), \quad (4)$$

$$\hat{\Theta}_n^{\text{SIR}}(\mathbf{X}_n^{\text{SIR},\{i\}}) = \frac{1}{N} \sum_{i=1}^N f(X_n^{\text{SIR},i}); \quad (5)$$

dans le cas de l'algorithme SIR $\hat{\Theta}_n^{\text{IS}}$ (resp. $\hat{\Theta}_n^{\text{SIR}}$) est l'estimateur construit avant (resp. après) l'étape de rééchantillonnage. Le résultat suivant compare, pour un nombre fini N de particules, la moyenne et la variance des estimateurs (3), (4) et (5).

Théorème 1 Soit $\hat{\Theta}_n^{\text{FA}}$, $\hat{\Theta}_n^{\text{IS}}$ et $\hat{\Theta}_n^{\text{SIR}}$ définis en (3)-(5), et soit $\pi_n(\cdot)$ le mélange (2). Alors conditionnellement à $\mathbf{x}_{n-1}^{\{i\}}$ et y_n ,

$$E(\hat{\Theta}_n^{\text{FA}}) = E(\hat{\Theta}_n^{\text{IS}}) = E(\hat{\Theta}_n^{\text{SIR}}) = E_{\pi_n}(f(X)), \text{ et}$$

$$\text{var}(\hat{\Theta}_n^{\text{FA}}) = \frac{1}{N} \text{var}_{\pi_n}(f(X)), \quad (6)$$

$$\text{var}(\hat{\Theta}_n^{\text{IS}}) = \sum_{i=1}^N \alpha_i^2 \sigma_i(f), \quad (7)$$

$$\text{var}(\hat{\Theta}_n^{\text{SIR}}) = \frac{1}{N} \text{var}_{\pi_n}(f(X)) + \frac{N-1}{N} \sum_{i=1}^N \alpha_i^2 \sigma_i(f). \quad (8)$$

3 Discussion et rôle de N_{eff}

Conditionnellement à $\mathbf{x}_{n-1}^{\{i\}}$ et y_n , les trois estimateurs $\hat{\Theta}_n^{\text{FA}}$, $\hat{\Theta}_n^{\text{IS}}$ et $\hat{\Theta}_n^{\text{SIR}}$ ont même moyenne, et nous les comparons donc en fonction de leur variance conditionnelle.

3.1 Algorithmes SIR et FA.

En comparant (6) et (8) on voit que pour tout N , $\text{var}(\hat{\Theta}_n^{\text{SIR}}|\mathbf{x}_{n-1}^{\{i\}}, y_n) > \text{var}(\hat{\Theta}_n^{\text{FA}}|\mathbf{x}_{n-1}^{\{i\}}, y_n)$. Partant d'un ensemble commun de particules, il est donc préférable d'effectuer les étapes (W,R,S) dans cet ordre, plutôt que d'échantillonner d'abord, de réactualiser les poids puis de rééchantillonner.

3.2 Algorithmes SIR et IS.

Même si (8) décompose $\text{var}(\hat{\Theta}_n^{\text{SIR}}|\mathbf{x}_{n-1}^{\{i\}}, y_n)$ comme la somme de deux termes positifs, il ne faut pas confondre (8) avec l'égalité de Rao-Blackwell qui ici s'écrit (cf. par ex. [7, p. 213])

$$\begin{aligned} \text{var}(\hat{\Theta}_n^{\text{SIR}}|\mathbf{x}_{n-1}^{\{i\}}, y_n) &= \text{var}(\hat{\Theta}_n^{\text{IS}}|\mathbf{x}_{n-1}^{\{i\}}, y_n) + \Delta_n, \quad (9) \\ \Delta_n &= E((\hat{\Theta}_n^{\text{SIR}} - \hat{\Theta}_n^{\text{IS}})^2|\mathbf{x}_{n-1}^{\{i\}}, y_n), \end{aligned}$$

où Δ_n est dû au seul rééchantillonnage. Il est donc bien connu qu'il est préférable de calculer $\hat{\Theta}_n$ à partir de l'ensemble de points pondérés $\{x_n^{\text{IS},i}, \alpha_i\}_{i=1}^N$ plutôt qu'à partir de l'ensemble non pondéré $\{x_n^{\text{SIR},i}, 1/N\}_{i=1}^N$.

Rappelons cependant qu'à l'origine cette étape de rééchantillonnage multinomial a été introduite afin de limiter la dégénérescence des poids, et effectivement il est pertinent d'éliminer des particules de poids faible. Cependant si tous les poids sont à peu près égaux, rééchantillonner fait courir le risque d'éliminer de bonnes particules. On s'attend à ce que Δ_n soit d'autant plus grand que le nombre de particules efficaces $N_{eff} \stackrel{\text{def}}{=} (\sum_{i=1}^N \alpha_i^2)^{-1}$ [10] est grand. Nous précisons maintenant cette idée; deux arguments sont disponibles.

Supposons tout d'abord que \tilde{N}_{eff} particules ont un poids non nul et que la masse totale est distribuée uniformément sur ces \tilde{N}_{eff} particules. Soit N' le nombre de particules après rééchantillonnage. Alors $E(N') = g(\tilde{N}_{eff}) = \tilde{N}_{eff} - \tilde{N}_{eff} \times (1 - \frac{1}{\tilde{N}_{eff}})^N$. $g' \geq 0$ et $g'' \leq 0$ donc $E(N')$ s'éloigne d'autant plus de \tilde{N}_{eff} que \tilde{N}_{eff} augmente (cf. Fig. 1).

- Si $\tilde{N}_{eff} = 1$, on est certain de retirer l'unique particule, $\hat{\Theta}_n^{\text{SIR}}$ et $\hat{\Theta}_n^{\text{IS}}$ coïncident, et $\Delta_n = 0$;

- $g'''(\cdot)$ s'annule en $\tilde{N}_{eff} = \frac{N+1}{3}$: tant que $\tilde{N}_{eff} \leq \frac{N+1}{3}$ le nombre moyen de particules perdues est faible, mais augmente de façon significative après ce point;
- Si $\tilde{N}_{eff} = N$, $E(N') \approx \frac{2N}{3}$: rééchantillonner fait perdre (en moyenne) un tiers des particules.

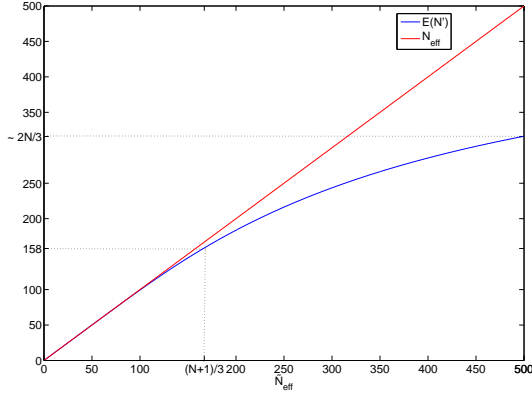


FIG. 1: Taille (moyenne) du support après rééchantillonnage, en fonction de \tilde{N}_{eff} ($N = 500$).

Lorsque les particules sont à peu près équilibrées la restriction du support après rééchantillonnage est donc sévère; on s'attend alors à ce que la différence de variances Δ_n dans (9) soit significative. Même si *stricto sensu* Δ_n n'est pas une fonction croissante de N_{eff} , nous avons le résultat suivant :

Proposition 1 Soit $\gamma = \min_{i=1 \dots N} \sigma_i(f)$, $\Gamma = \max_{i=1 \dots N} \sigma_i(f)$, $\lambda = \min_{1 \leq i < j \leq N} [\mu_i(f) - \mu_j(f)]^2$, $\Lambda = \max_{1 \leq i < j \leq N} [\mu_i(f) - \mu_j(f)]^2$. Alors

$$\left(1 - \frac{1}{N_{eff}}\right)\left(\gamma + \frac{\lambda}{2}\right) \leq \Delta_n \leq \left(1 - \frac{1}{N_{eff}}\right)\left(\Gamma + \frac{\Lambda}{2}\right). \quad (10)$$

Δ_n est borné inférieurement et supérieurement par deux fonctions croissantes de N_{eff} , donc le terme de variance induit par l'étape de rééchantillonnage aura tendance à croître si N_{eff} croît; en quelque sorte Δ_n mesure si le rééchantillonnage aurait dû avoir lieu ou non.

3.3 Algorithmes IS et FA.

Comparons maintenant $\hat{\Theta}_n^{IS}$ et $\hat{\Theta}_n^{FA}$. Des résultats de statistique asymptotique [8] montrent qu'aucun de ces deux estimateurs n'est meilleur que l'autre dans tous les cas de figure. Dans ce paragraphe nous essayons d'expliquer ce phénomène en mettant en évidence les paramètres qui jouent un rôle important dans le signe de $\Delta'_n \stackrel{\text{def}}{=} \text{var}(\hat{\Theta}_n^{FA}) - \text{var}(\hat{\Theta}_n^{IS})$. On a

$$\Delta'_n = \sum_{i=1}^N \alpha_i \left(\frac{1}{N} - \alpha_i\right) \sigma_i(f) + \frac{1}{N} \sum_{1 \leq i < j \leq N} \alpha_i \alpha_j (\mu_i(f) - \mu_j(f))^2. \quad (11)$$

En général, le signe de Δ'_n dépend de la fonction f d'intérêt et du modèle, sauf dans les deux cas extrêmes suivants :

- Si $\alpha_i = \frac{1}{N} \forall i$, $\Delta'_n \geq 0$ et $\hat{\Theta}_n^{IS}$ est préférable. Cela n'est pas surprenant car l'étape de rééchantillonnage de l'algorithme FA élimine environ un tiers des particules, alors que toutes les particules sont prises en compte dans l'algorithme IS.
- Si il existe j tel que $\alpha_j = 1$ et $\alpha_i = 0$ pour $i \neq j$, alors $\Delta'_n = \frac{1-N}{N} \sigma_j(f) \leq 0$ et $\hat{\Theta}_n^{FA}$ est préférable. Cela n'est pas surprenant car dans ce cas $\hat{\Theta}_n^{IS}$ ne tient compte que d'une particule ($\hat{\Theta}_n^{IS} = f(X_n^j)$ où $X_n^j \sim p_j$), alors que $\hat{\Theta}_n^{FA} = \frac{1}{N} \sum_{i=1}^N f(X_n^i)$ où $X_n^i \sim p_j$.

Considérons maintenant les cas intermédiaires où la distribution des poids n'est ni uniforme ni réduite à une seule masse non nulle :

Proposition 2 Soit γ , Γ , λ et Λ définis en Proposition 1. On a

$$\frac{\gamma + \frac{\lambda}{2}}{N} - \left(\Gamma + \frac{\lambda}{2N}\right) \frac{1}{N_{eff}} \leq \Delta'_n \leq \frac{\Gamma + \frac{\Lambda}{2}}{N} - \left(\gamma + \frac{\Lambda}{2N}\right) \frac{1}{N_{eff}}. \quad (12)$$

Corollaire 1 Si $1 \leq N_{eff} \leq a_n = \frac{2\gamma + \lambda}{2\Gamma + \Lambda}$ alors $\hat{\Theta}_n^{FA}$ est préférable à $\hat{\Theta}_n^{IS}$. Si $b_n = \frac{2\Gamma N + \lambda}{2\gamma + \lambda} \leq N_{eff} \leq N$ alors $\hat{\Theta}_n^{IS}$ est préférable à $\hat{\Theta}_n^{FA}$.

Remarque 2 Les résultats du Corollaire 1 sont paramétrés par γ , Γ , λ , Λ and N . Commentons quelques cas spécifiques :

- Pour γ , Γ et N fixés, $a_n \rightarrow \frac{\gamma}{\Gamma} N$ si $\Lambda \rightarrow 0$, et si $\gamma \approx \Gamma$ $a_n \approx N$: $\hat{\Theta}_n^{FA}$ est préférable même si $N_{eff} \approx N$. Λ petit et $\gamma \approx \Gamma$ signifient que $\mu_i(f)$ et $\sigma_i(f)$ sont quasi constants donc indépendants de i . Dans ce cas rééchantillonner est sans risque, car quelle que soit la particule \tilde{x}_{n-1}^i sélectionnée à l'étape i , la moyenne et la variance de $f(X_n^i)$, où $X_n^i \sim p_i$, ne dépendent pas de \tilde{x}_{n-1}^i .
- Pour γ , Γ et N fixés, $b_n \rightarrow 1$ si $\lambda \rightarrow \infty$: dans ce cas $\hat{\Theta}_n^{IS}$ est préférable même si N_{eff} est petit. Rappelons que dans l'algorithme FA $X_n^i \sim p_i$, où \tilde{x}_{n-1}^i est la $i^{\text{ème}}$ particule rééchantillonnée. Puisqu'ici Γ est petit par rapport à λ , $f(X_n^i)$ et $f(X_n^j)$ sont bien séparés si $\tilde{x}_{n-1}^i \neq \tilde{x}_{n-1}^j$, dans un même voisinage sinon. Rééchantillonner crée donc des amas de points (les points dans un amas proviennent d'une particule sélectionnée plusieurs fois), et des régions vides, car si \tilde{x}_{n-1}^i est écartée lors du rééchantillonnage, une compensation provenant des particules qui ont été rééchantillonnées est très peu probable, et il n'y aura donc pas de point au voisinage de $\mu_i(f)$.
- En pratique on peut dans certains cas calculer pour quelles valeurs de N_{eff} $\hat{\Theta}_n^{IS}$ est préférable à $\hat{\Theta}_n^{FA}$. Considérons

par exemple le cas où $\gamma \approx \Gamma$. Alors si $\lambda = 2\Gamma b_n = \frac{N+1}{2}$; pour ces valeurs des paramètres (qui signifient que les points $f(X_n^i)$ sont bien séparés), rééchantillonner est dangereux même si N_{eff} est faible ($\approx \frac{N}{2}$).

4 Simulations

Soit le modèle ARCH

$$\begin{cases} X_{n+1} &= \sqrt{\beta_0 + \beta_1 X_n^2} \times U_n \\ Y_n &= X_n + V_n \end{cases}, \quad (13)$$

où $\{U_n\}$ et $\{V_n\}$ sont i.i.d, mutuellement indépendants et indépendants de X_0 , avec X_0 et $U_n \sim \mathcal{N}(0, 1)$, $V_n \sim \mathcal{N}(0, R)$, $\beta_0 = 9, \beta_1 = 5$ et $R = 1$; $p(y_n|x_{n-1}) = \mathcal{N}(0, R + \beta_0 + \beta_1 x_{n-1}^2)$ et $p(x_n|x_{n-1}, y_n) = \mathcal{N}(\frac{\beta_0 + \beta_1 x_{n-1}^2}{\beta_0 + \beta_1 x_{n-1}^2 + R} y_n, \frac{(\beta_0 + \beta_1 x_{n-1}^2)R}{\beta_0 + \beta_1 x_{n-1}^2 + R})$.

Nous faisons tourner l'algorithme FA (avec $N = 1000$) sur $P = 400$ réalisations du modèle (13); pour chacune, en tout n nous utilisons le nuage de points commun $\{x_{n-1}^i\}_{i=1}^N$ produit par cet algorithme FA pour calculer $\hat{\Theta}_n^{\text{FA}}, \hat{\Theta}_n^{\text{SIR}}$ ou $\hat{\Theta}_n^{\text{IS}}$ avec $f(x) = x$. Dans la Fig. 2 nous traçons $P(n) \stackrel{\text{def}}{=} \frac{1}{P} \sum_{i=1}^P (\hat{\Theta}_n - E_{\pi_n}(f))^2$, $\hat{\Theta}_n$ représentant ici $\hat{\Theta}_n^{\text{FA}}, \hat{\Theta}_n^{\text{SIR}}$ ou $\hat{\Theta}_n^{\text{IS}}$.

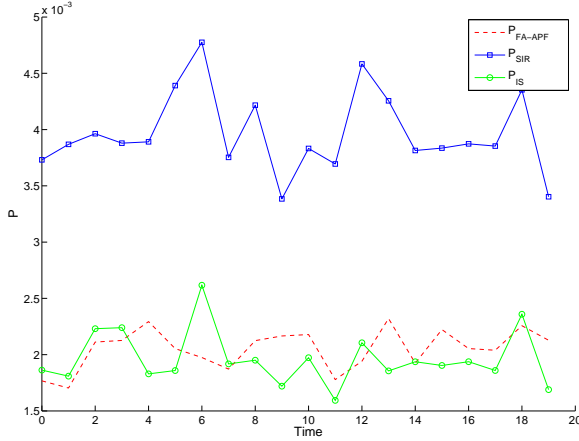


FIG. 2: $P(n)$, modèle ARCH, $\beta_0 = 9, \beta_1 = 5, R = 1$

On vérifie (cf (8)) que $P_{\text{SIR}}(n) \approx P_{\text{FA}}(n) + \frac{N-1}{N} \times P_{\text{IS}}(n)$, et que $P_{\text{FA}}(n) - P_{\text{IS}}(n)$ n'est pas de signe constant.

La figure 3 visualise le rôle de N_{eff} . Ici $\beta_0 = 5, \beta_1 = 3$ et $R = 5$, et $\text{MSE}(n) \stackrel{\text{def}}{=} \frac{1}{P} \sum_{i=1}^P (\hat{\Theta}_n - E_{p_{n|n}}(f))^2$, où $E_{p_{n|n}}(f)$ est calculé avec un filtre bootstrap avec 5×10^5 particules. Ces e.q.m. sont normalisées par rapport à celle de l'algorithme FA de façon à comparer le comportement des algorithmes IS et FA en fonction de $E(N_{eff}(n))$ ($\text{MSE}_{\text{FA}}(n) \approx 2 \times 10^{-3}$ pour tout n). Quand $E(N_{eff})$ n'est pas proche de N $\hat{\Theta}_n^{\text{FA}}$ est meilleur que $\hat{\Theta}_n^{\text{IS}}$. C'est le cas à $n = 2$ où $E(N_{eff}) = \frac{N}{2}$; $\text{var}(\hat{\Theta}_n^{\text{IS}})$ est alors proche de $\text{var}(\hat{\Theta}_n^{\text{SIR}})$, ce qui illustre que les effets

du rééchantillonage sont moins sévères si $E(N_{eff})$ est faible. $\hat{\Theta}_n^{\text{IS}}$ semble préférable si $E(N_{eff})$ est proche de N .

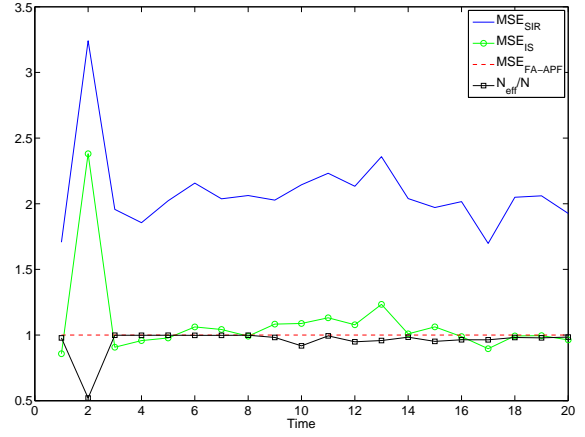


FIG. 3: e.q.m. relative et moyenne de N_{eff} - modèle ARCH, $\beta_0 = 5, \beta_1 = 3, R = 5$.

Références

- [1] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, Springer, 2001.
- [2] M. K. Pitt and N. Shephard, "Filtering via simulation: Auxiliary particle filter," *J. Am. Stat. Ass.*, 1999.
- [3] P. Fearnhead, "Computational methods for complex stochastic systems: A review of some alternatives to MCMC," *Statistics and Computing*, 2008.
- [4] H. R. Künsch, "State space and hidden Markov models," in *Complex Stochastic Systems*, O. E. Barndorff-Nielsen, D. R. Cox, and C. Klüppelberg, Eds., vol. 87 of *Monographs on Statistics and Applied Probability*, pp. 109–73. Chapman and Hall / CRC, New York, 2001.
- [5] D. Crisan and A. Doucet, "A survey of convergence results on particle filtering methods for practitioners," *IEEE Tr. Sig. Proc.*, vol. 50, no. 3, 2002.
- [6] P. del Moral, *Feynman-Kac formulae. Genealogical and interacting particle systems with applications*, Probability and its applications. Springer, New-York, 2004.
- [7] O. Cappé, E. Moulines, and T. Rydén, *Inference in Hidden Markov Models*, Springer-Verlag, 2005.
- [8] A. M. Johansen and A. Doucet, "A note on the auxiliary particle filter," *Stat. and Prob. Letters*, 2008.
- [9] A. Doucet, S. J. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, pp. 197–208, 2000.
- [10] J. S. Liu A. Kong and W. H. Wong, "Sequential imputations and bayesian missing data problems," *J. Am. Stat. Ass.*, vol. 89, no. 425, pp. 278–88, March 1994.