Biometrika (2018), **xx**, x, *pp*. 1–17
Printed in Great Britain

Variance estimation in the particle filter

BY A. LEE, N. WHITELEY
School of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, U.K.
anthony.lee@bristol.ac.uk nick.whiteley@bristol.ac.uk

SUMMARY

This paper concerns numerical assessment of Monte Carlo error in particle filters. We show that by keeping track of certain key features of the genealogical structure arising from resampling operations, it is possible to estimate variances of a number of Monte Carlo approximations that particle filters deliver. All our estimators can be computed from a single run of a particle filter. We establish that as the number of particles grows, our estimators are weakly consistent for asymptotic variances of the Monte Carlo approximations and some of them are also non-asymptotically unbiased. The asymptotic variances can be decomposed into terms corresponding to each time step of the algorithm, and we show how to estimate each of these terms consistently. When the number of particles may vary over time, this allows approximation of the asymptotically optimal allocation of particle numbers.

Some key words: Allocation; Particle filter; Sequential Monte Carlo; Simulation; Variance estimation.

1. INTRODUCTION

Particle filters, or sequential Monte Carlo methods, provide approximations of integrals with respect to sequences of measures. In popular statistical inference applications, these measures arise naturally from conditional distributions in hidden Markov models, or are constructed artificially to bridge between target distributions in Bayesian analysis. The number of particles used controls the tradeoff between computational complexity and accuracy. Theoretical properties of this relationship have been the subject of intensive research; the literature includes central limit theorems (Del Moral & Guionnet, 1999; Chopin, 2004; Künsch, 2005; Douc & Moulines, 2008) and a variety of refined asymptotic (Douc et al., 2005; Del Moral et al., 2007) and non-asymptotic (Del Moral & Miclo, 2001; Cérou et al., 2011) results. These studies provide a wealth of insight into the mathematical behaviour of particle filter approximations and validate them theoretically, but considerably less is known about how, in practice, to extract information from a realization of a single particle filter in order to report numerical measures of Monte Carlo error. This is in notable contrast to other families of Monte Carlo techniques, especially Markov chain Monte Carlo, for which an extensive literature on variance estimation exists. Our main aim is to address this gap.

We introduce particle filters via a framework of Feynman–Kac models (Del Moral, 2004). This allows us to identify the key ingredients of particle filters and the measures they approximate. Based on a single realization of a particle filter, we provide unbiased estimators of the variance and individual asymptotic variance terms for a class of unnormalized particle approximations. No estimators of these quantities based on a single run of a particle filter have previously appeared in the literature, and all of our estimators ultimately arise from particle approximations of quantities appearing in a non-asymptotic second-moment expression. Upon suitable rescaling, we establish

that our estimators are weakly consistent for asymptotic variances associated with a larger class of particle approximations. One of these re-scaled estimators is closely related to that of Chan & Lai (2013), which is the only other consistent asymptotic variance estimator based on a single realization of a particle filter in the literature. We also demonstrate how one can use the estimators to inform the choice of algorithm parameters in an attempt to improve performance.

2. PARTICLE FILTERS

2·1. *Notation and conventions*

For a generic measurable space $(\mathsf{E}, \mathcal{E})$, we denote by $\mathcal{L}(\mathcal{E})$ the set of $\mathbb{R}$-valued, $\mathcal{E}$-measurable and bounded functions on E. For $\varphi \in \mathcal{L}(\mathcal{E})$, μ a measure and K an integral kernel on $(\mathsf{E}, \mathcal{E})$, we write $\mu(\varphi) = \int_{\mathsf{E}} \varphi(x)\mu(\mathrm{d}x)$, $K(\varphi)(x) = \int_{\mathsf{E}} K(x, \mathrm{d}x')\varphi(x')$ and $\mu K(A) = \int_{\mathsf{E}} \mu(\mathrm{d}x)K(x, A)$. Constant functions $x \in E \mapsto c \in \mathbb{R}$ are denoted simply by c. For $\varphi \in \mathcal{L}(\mathcal{E})$, $\varphi^{\otimes 2}(x, x') = \varphi(x)\varphi(x')$. The Dirac measure located at x is denoted δ_x. For any sequence $(a_n)_{n\in\mathbb{Z}}$ and $p \leq q$, $a_{p:q} = (a_p, \ldots, a_q)$ and by convention $\prod_{p=0}^{-1} a_p = 1$. For any $m \in \mathbb{N}$, $[m] = \{1, \ldots, m\}$. For any $c \in \mathbb{R}$, $\lceil c \rceil$ is the smallest integer greater than or equal to c. For a vector of positive values $(a_1, \ldots, a_m)$, we denote by $\mathcal{C}(a_1, \ldots, a_m)$ the categorical distribution over $\{1, \ldots, m\}$ with probabilities $(a_1/\sum_{i=1}^m a_i, \ldots, a_m/\sum_{i=1}^m a_i)$. When a random variable is indexed by a superscript N, a sequence of such random variables is implicitly defined by considering each value $N \in \mathbb{N}$, and limits will always be taken along this sequence.

2·2. *Discrete time Feynman–Kac models*

On a measurable space $(\mathsf{X}, \mathcal{X})$ with n a non-negative integer, let M_0 be a probability measure, $M_1, \ldots, M_n$ a sequence of Markov kernels and $G_0, \ldots, G_n$ a sequence of $\mathbb{R}$-valued, strictly positive, upper-bounded functions. We assume throughout that X does not consist of a single point. We define a sequence of measures by $\gamma_0 = M_0$ and, recursively,

$$\gamma_p(S) = \int_{\mathsf{X}} \gamma_{p-1}(\mathrm{d}x)G_{p-1}(x)M_p(x, S), \qquad p \in [n], \quad S \in \mathcal{X}. \tag{1}$$

Since $\gamma_p(\mathsf{X}) \in (0, \infty)$ for each p, the following probability measures are well-defined:

$$\eta_p(S) = \frac{\gamma_p(S)}{\gamma_p(\mathsf{X})}, \qquad p \in \{0, \ldots, n\}, \quad S \in \mathcal{X}. \tag{2}$$

The representation $\gamma_n(\varphi) = E\{\varphi(X_n)\prod_{p=0}^{n-1} G_p(X_p)\}$, where the expectation is taken with respect to the Markov chain with initial distribution $X_0 \sim M_0$ and transitions $X_p \sim M_p(X_{p-1}, \cdot)$, establishes the connection to Feynman–Kac formulae. Measures with the structure in (1)–(2) arise in a variety of statistical contexts.

2·3. *Motivating examples of Feynman–Kac models*

As a first example, consider a hidden Markov model: a bivariate Markov chain $(X_p, Y_p)_{p=0,\ldots,n}$ where $(X_p)_{p=0,\ldots,n}$ is itself Markov with initial distribution M_0 and transitions $X_p \sim M_p(X_{p-1}, \cdot)$, and such that each Y_p is conditionally independent of $(X_q, Y_q; q \neq p)$ given X_p. If the conditional distribution of Y_p given X_p admits a density $g_p(X_p, \cdot)$ and one fixes a sequence of observed values $y_0, \ldots, y_{n-1}$, then with $G_p(x_p) = g_p(x_p, y_p)$, η_n is the conditional distribution of X_n given $y_0, \ldots, y_{n-1}$. Hence, $\eta_n(\varphi)$ is a conditional expectation and $\gamma_n(\mathsf{X}) = \gamma_n(1)$ is the marginal likelihood of $y_0, \ldots, y_{n-1}$

As a second example, consider the following sequential simulation setup. Let π_0 and π_1 be two probability measures on $(\mathsf{X}, \mathcal{X})$ such that $\pi_0(dx) = \bar{\pi}_0(x)dx/Z_0$ and $\pi_1(dx) = \bar{\pi}_1(x)dx/Z_1$,

where $\bar{\pi}_0$ and $\bar{\pi}_1$ are unnormalized probability densities with respect to a common dominating measure dx and $Z_i = \int_{\mathsf{X}} \bar{\pi}_i(x)dx$, $i \in \{0, 1\}$ are integrals unavailable in closed form. In Bayesian statistics π_1 may arise as a badly-behvaved posterior distribution from which one wishes to sample, π_0 is a more benign distribution from which sampling is feasible, and calculating Z_1/Z_0 allows assessment of model fit. Introducing a sequence $0 = \beta_0 < \cdots < \beta_n = 1$ and taking $G_p(x) = \{\bar{\pi}_1(x)/\bar{\pi}_0(x)\}^{\beta_{p+1}-\beta_p}$, $M_0 = \pi_0$, and, for each $p = 1, \ldots, n$, taking M_p as a Markov kernel invariant with respect to the distribution with density proportional to $\bar{\pi}_0(x)^{1-\beta_p}\bar{\pi}_1(x)^{\beta_p}$, elementary manipulations yield

$$\gamma_p(S) = \frac{1}{Z_0}\int_S \bar{\pi}_0(x)^{1-\beta_p}\bar{\pi}_1(x)^{\beta_p}dx, \quad \eta_n = \pi_1, \quad \gamma_n(\mathsf{X}) = \frac{Z_1}{Z_0},$$

so that $\eta_1, \ldots, \eta_{n-1}$ forms a sequence of intermediate distributions between π_0 and π_1. This type of construction appears in Del Moral et al. (2006) and references therein.

2·4. *Particle approximations*

We now introduce particle approximations of the measures in (1)–(2). Let $c_{0:n}$ be a sequence of positive real numbers and let $N \in \mathbb{N}$. We define a sequence of particle numbers $N_{0:n}$ by $N_p = \lceil c_p N \rceil$ for $p \in \{0, \ldots, n\}$. To avoid notational complications, we shall assume throughout that $c_{0:n}$ and N are such that $\min_p N_p \geq 2$. The particle system consists of a sequence $\zeta = \zeta_{0:n}$, where for each p, $\zeta_p = (\zeta_p^1, \ldots, \zeta_p^{N_p})$ and each ζ_p^i is valued in X. To describe the resampling operation we also introduce random variables denoting the indices of the ancestors of each random variable ζ_p^i. That is, for each $i \in [N_p]$, A_{p-1}^i is a $[N_{p-1}]$-valued random variable and we write $A_{p-1} = (A_{p-1}^1, \ldots, A_{p-1}^{N_p})$ for $p \in [n]$ and $A = A_{0:n-1}$.

A simple description of the particle system is given in Algorithm 1. An important and non-standard feature is that we keep track of a collection of indices $E_{0:n}$ with $E_p = (E_p^1, \ldots, E_p^{N_p})$ for each p, which will be put to use in our variance estimators. We call these Eve indices because E_p^i represents the index of the time 0 ancestor of ζ_p^i . The fact that N_p may vary with p is also atypical, and allows us to address asymptotically optimal particle allocation in Section 5·1. On a first reading, one may wish to assume that $N_{0:n}$ is not time-varying, i.e., $c_p = 1$ so $N_p = N$ for all $p \in \{0, \ldots, n\}$. Figure 1 is a graphical representation of a realization of a small particle system.

Algorithm 1. The particle filter.

1. At time 0: for each $i \in [N_0]$, sample $\zeta_0^i \sim M_0(\cdot)$ and set $E_0^i \leftarrow i$.
2. At each time $p = 1, \ldots, n$: for each $i \in [N_p]$,
   a. sample $A_{p-1}^i \sim \mathcal{C}\{G_{p-1}(\zeta_{p-1}^1), \ldots, G_{p-1}(\zeta_{p-1}^{N_{p-1}})\}$.
   b. sample $\zeta_p^i \sim M_p(\zeta_{p-1}^{A_{p-1}^i}, \cdot)$ and set $E_p^i \leftarrow E_{p-1}^{A_{p-1}^i}$.

The particle approximations to η_n and γ_n are defined respectively by the random measures

$$\eta_n^N = \frac{1}{N_n}\sum_{i\in[N_n]} \delta_{\zeta_n^i}, \qquad \gamma_n^N = \left\{\prod_{p=0}^{n-1} \eta_p^N(G_p)\right\}\eta_n^N,$$

and we observe that, similar to (2), $\eta_n^N = \gamma_n^N/\gamma_n^N(1)$. To simplify presentation, the dependence of γ_n^N and η_n^N on $c_{0:n}$ is suppressed from the notation. The following proposition establishes basic properties of the particle approximations, which validate their use.

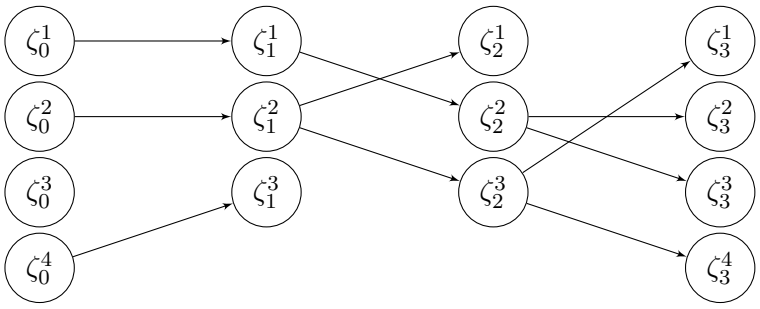


Figure 1: A particle system with $n = 3$ and $N_{0:3} = (4, 3, 3, 4)$. An arrow from ζ_{p-1}^i to ζ_p^j indicates that the ancestor of ζ_p^j is ζ_{p-1}^i, i.e. $A_{p-1}^j = i$. In the realization shown, the ancestral indices are $A_0 = (1, 2, 4)$, $A_1 = (2, 1, 2)$ and $A_2 = (3, 2, 2, 3)$, while $E_0 = (1, 2, 3, 4)$, $E_1 = (1, 2, 4)$, $E_2 = (2, 1, 2)$ and $E_3 = (2, 1, 1, 2)$.

PROPOSITION 1. *There exists a map $\sigma_n^2 : \mathcal{L}(\mathcal{X}) \to [0, \infty)$ such that for any $\varphi \in \mathcal{L}(\mathcal{X})$:*

1. $E\left\{\gamma_n^N(\varphi)\right\} = \gamma_n(\varphi)$*, for all* $N \geq 1$*;*
2. $\gamma_n^N(\varphi) \to \gamma_n(\varphi)$ *almost surely and* $N\mathrm{var}\left\{\gamma_n^N(\varphi)/\gamma_n(1)\right\} \to \sigma_n^2(\varphi)$*;*
3. $\eta_n^N(\varphi) \to \eta_n(\varphi)$ *almost surely and* $NE\left[\left\{\eta_n^N(\varphi) - \eta_n(\varphi)\right\}^2\right] \to \sigma_n^2\{\varphi - \eta_n(\varphi)\}$.

In the case that the number of particles is constant over time, $N_p = N$, these properties are well known and can be deduced, for example, from various results of Del Moral (2004). The arguments used to treat the general $N_p = \lceil c_p N \rceil$ case are not substantially different, but since they seem not to have been published anywhere in exactly the form we need, we include a proof of Proposition 1 in the supplement.

2·5. *A variance estimator*

For $\varphi \in \mathcal{L}(\mathcal{X})$, consider the quantity

$$V_n^N(\varphi) = \eta_n^N(\varphi)^2 - \left(\prod_{p=0}^{n} \frac{N_p}{N_p - 1}\right) \frac{1}{N_n^2} \sum_{i,j:E_n^i \neq E_n^j} \varphi(\zeta_n^i)\varphi(\zeta_n^j) \tag{3}$$

$$= \eta_n^N(\varphi)^2 \left(1 - \prod_{p=0}^{n} \frac{N_p}{N_p - 1}\right) + \left(\prod_{p=0}^{n} \frac{N_p}{N_p - 1}\right) \frac{1}{N_n^2} \sum_{i \in [N_0]} \sum_{j:E_n^j = i} \varphi(\zeta_n^j)^2, \tag{4}$$

which is readily computable as a byproduct of Algorithm 1. The following theorem is the first main result of the paper. We state it here to make some of the practical implications of our work accessible to the reader before entering into more technical details; it shows that via (3), the variables E_n^i can be used to estimate the Monte Carlo errors associated with $\gamma_n^N(\varphi)$ and $\eta_n^N(\varphi)$.

THEOREM 1. *The following hold for any $\varphi \in \mathcal{L}(\mathcal{X})$, with $\sigma_n^2(\cdot)$ as in Proposition* 1*:*

1. $E\left\{\gamma_n^N(1)^2 V_n^N(\varphi)\right\} = \mathrm{var}\left\{\gamma_n^N(\varphi)\right\}$ *for all* $N \geq 1$*;*
2. $NV_n^N(\varphi) \to \sigma_n^2(\varphi)$ *in probability;*
3. $NV_n^N\{\varphi - \eta_n^N(\varphi)\} \to \sigma_n^2\{\varphi - \eta_n(\varphi)\}$ *in probability.*

Remark 1. Since $\eta_n^N\{\varphi - \eta_n^N(\varphi)\} = 0$, the estimator $NV_n^N\{\varphi - \eta_n^N(\varphi)\}$ simplifies to

$$NV_n^N\left\{\varphi - \eta_n^N(\varphi)\right\} = N\left(\prod_{p=0}^{n}\frac{N_p}{N_p - 1}\right)\frac{1}{N_n^2}\sum_{i\in[N_0]}\sum_{j:E_n^j=i}\left\{\varphi(\zeta_n^j) - \eta_n^N(\varphi)\right\}^2.$$

This estimator is a deterministic and asymptotically negligible modification of Chan & Lai (2013)'s weakly consistent estimator of $\sigma_n^2\{\varphi - \eta_n(\varphi)\}$, given by

$$\hat{\sigma}_{\mathrm{CL}}^2\{\varphi - \eta_n(\varphi)\} = \frac{1}{N}\sum_{i\in[N]}\sum_{j:E_n^j=i}\left\{\varphi(\zeta_n^j) - \eta_n^N(\varphi)\right\}^2,$$

when N is not time-varying. Our estimator is larger than Chan and Lai's due to the factor $\prod_{p=0}^{n} N_p/(N_p - 1)$; we find in the examples that there is little difference in the regime where both are nearly unbiased. Our main contributions, therefore, are the estimators proposed for which there are no existing alternatives in the literature: those with properties 1 or 2 of Theorem 1, and those developed in the sequel to estimate individual asymptotic variance terms arising from a natural decomposition of $\sigma_n^2(\varphi)$.

The proof of Theorem 1, given in the Appendix, relies on a number of intermediate results concerning moment properties of the particle approximations which we shall develop. Before embarking on this, we discuss how $V_n^N(\varphi)$ may be interpreted. Consider independent, identically distributed random variables $X_1, \ldots, X_N$ with sample mean $\bar{X}$. The unbiased estimator of the variance of $\bar{X}$ is

$$\frac{1}{N(N-1)}\sum_i (X_i - \bar{X})^2 = \bar{X}^2\left(1 - \frac{N}{N-1}\right) + \left(\frac{N}{N-1}\right)\frac{1}{N^2}\sum_{i=1}^{N} X_i^2. \tag{5}$$

Observe the resemblance between the right–hand sides of (4) and (5): the role of X_i^2 is played by $\sum_{j:E_n^j=i}\varphi(\zeta_n^j)^2$, the sum of φ^2 evaluated at the descendants of ζ_0^i. This change, and the product term $\prod_{p=0}^{n} N_p/(N_p - 1)$ replacing $N/(N-1)$, arise from the non-trivial dependence structure associated with $\zeta_0, \ldots, \zeta_n$. One of the main difficulties we face is to develop a suitable mathematical perspective from which to account for this dependence and establish Theorem 1.

The main statistical implication of Theorem 1 is that the variance estimators are weakly consistent as $N \to \infty$ with n fixed. In the opposite regime, where N is fixed and $n \to \infty$, the estimators degenerate because the resampling operations cause $E_n^1, \ldots, E_n^N$ to eventually become equal. Using results reported here, Olsson & Douc (2018) address the degeneracy issue by modifying $\hat{\sigma}_{\mathrm{CL}}^2$ so that ancestries are traced only over a fixed time horizon.

3. Moment properties of the particle approximations

3·1. *Genealogical tracing variables*

Our next step is to introduce some auxiliary random variables associated with the genealogical structure of the particle system. These variables are introduced only for purposes of analysis: they will assist in deriving and justifying our variance estimators. Given (A, ζ), the first collection of variables, $K^1 = (K_0^1, \ldots, K_n^1)$, is conditionally distributed as follows: K_n^1 is uniformly distributed on $[N_n]$ and for each $p = n-1, \ldots, 0$, $K_p^1 = A_p^{K_{p+1}^1}$. Given (A, ζ) and K^1, the second collection of variables, $K^2 = (K_0^2, \ldots, K_n^2)$, is conditionally distributed as follows:

K_n^2 is uniformly distributed on $[N_n]$ and for each $p = n-1, \ldots, 0$ we have $K_p^2 = A_p^{K_{p+1}^2}$ if $K_{p+1}^2 \neq K_{p+1}^1$ and $K_p^2 \sim \mathcal{C}\{G_p(\zeta_p^1), \ldots, G_p(\zeta_p^{N_p})\}$ if $K_{p+1}^2 = K_{p+1}^1$. The interpretation of K^1 is that it traces backwards in time the ancestral lineage of a particle chosen randomly from the population at time n. The interpretation of K^2 is slightly more complicated: it traces backwards in time a sequence of broken ancestral lineages, where breaks occur when components of K^1 and K^2 coincide.

3·2. *Lack of bias and second moment of* $\gamma_n^N(\varphi)$

We now give expressions for the first two moments of $\gamma_n^N(\varphi)$.

LEMMA 1. *For any* $\varphi \in \mathcal{L}(\mathcal{X})$, $E\left\{\gamma_n^N(1)\varphi(\zeta_n^{K_n^1})\right\} = \gamma_n(\varphi)$ *and* $E\left\{\gamma_n^N(\varphi)\right\} = \gamma_n(\varphi)$.

The proof is in the Supplementary Material. The lack-of-bias property $E\left\{\gamma_n^N(\varphi)\right\} = \gamma_n(\varphi)$ is well-known and a martingale proof for the $N_p = N$ case can be found in Del Moral (2004, Ch. 9).

In order to present an expression for the second moment of $\gamma_n^N(\varphi)$, we now introduce a collection of measures on $\mathcal{X}^{\otimes 2}$, denoted $\{\mu_b : b \in B_n\}$ where $B_n = \{0,1\}^{n+1}$ is the set of binary strings of length $n+1$. The measures are constructed as follows. For a given $b \in B_n$, let $(X_p, X_p')_{p=0,\ldots,n}$ be a Markov chain with state-space X^2, distributed according to the following recipe. If $b_0 = 0$ then $X_0 \sim M_0$ and $X_0' \sim M_0$ independently, while if $b_0 = 1$ then $X_0' = X_0 \sim M_0$. Then, for $p = 1, \ldots, n$, if $b_p = 0$ then $X_p \sim M_p(X_{p-1}, \cdot)$ and $X_p' \sim M_p(X_{p-1}', \cdot)$ independently, while if $b_p = 1$ then $X_p' = X_p \sim M_p(X_{p-1}, \cdot)$. Letting E_b denote expectation with respect to the law of this Markov chain we then define

$$\mu_b(S) = E_b\left[\mathbb{I}\left\{(X_n, X_n') \in S\right\} \prod_{p=0}^{n-1} G_p(X_p) G_p(X_p')\right], \qquad S \in \mathcal{X}^{\otimes 2}, \quad b \in B_n.$$

Recalling that $\gamma_n(\varphi) = E\{\varphi(X_n)\prod_{p=0}^{n-1} G_p(X_p)\}$ for $\varphi \in \mathcal{L}(\mathcal{X})$, we write $\mu_b(\phi) = E_b\left\{\phi(X_n, X_n')\prod_{p=0}^{n-1} G_p(X_p)G_p(X_p')\right\}$ for $\phi \in \mathcal{L}(\mathcal{X}^{\otimes 2})$ and $b \in B_n$, and can view μ_b as defining a Feynman–Kac model on $\mathcal{X}^{\otimes 2}$.

Remark 2. Observe that with $0_n \in B_n$ denoting the zero string, $\mu_{0_n}(\varphi^{\otimes 2}) = \gamma_n(\varphi)^2$.

In order to succinctly express the second moment of $\gamma_n^N(\varphi)$, we define appropriate sets of pairs of strings of length $n+1$. Letting $[N_{0:n}] = [N_0] \times \cdots \times [N_n]$, and for any $b \in B_n$,

$$\mathcal{I}(b) = \{(k^1, k^2) \in [N_{0:n}]^2 \,:\, \text{for each } p,\ k_p^1 = k_p^2 \iff b_p = 1\},$$

we have that $\mathcal{I}(b)$ contains strings which coincide in their p-th coordinate exactly when $b_p = 1$.

LEMMA 2. *For any* $\phi \in \mathcal{L}(\mathcal{X}^{\otimes 2})$, $\varphi \in \mathcal{L}(\mathcal{X})$ *and* $b \in B_n$,

$$E\left[\mathbb{I}\left\{(K^1, K^2) \in \mathcal{I}(b)\right\}\gamma_n^N(1)^2\phi(\zeta_n^{K_n^1}, \zeta_n^{K_n^2})\right] = \prod_{p=0}^{n}\left\{\left(\frac{1}{N_p}\right)^{b_p}\left(1 - \frac{1}{N_p}\right)^{1-b_p}\right\}\mu_b(\phi) \tag{6}$$

and

$$E\left\{\gamma_n^N(\varphi)^2\right\} = \sum_{b \in B_n}\prod_{p=0}^{n}\left(\frac{1}{N_p}\right)^{b_p}\left(1 - \frac{1}{N_p}\right)^{1-b_p}\mu_b(\varphi^{\otimes 2}). \tag{7}$$

The proof of Lemma 2uses an argument involving the law of a doubly conditional sequential Monte Carlo algorithm (see also Andrieu et al., 2018). The identity (7) was first proved by Cérou et al. (2011) in the case where $N_p = N$. Our proof technique is different: we obtain (7) as a consequence of (6). The appearance of K^1, K^2 in (6) is also central to the justification of our variance estimators below.

3·3. *Asymptotic variances*

For each $p \in \{0, \ldots, n\}$, let $e_p \in B_n$ denote the vector with a 1 in position p and zeros elsewhere. As in Remark 2, 0_n denotes the zero string in B_n. The following result builds upon Lemmas 1–2. It shows that a particular subset of the measures $\{\mu_b : b \in B_n\}$, namely μ_{0_n} and $\{\mu_{e_p} : p = 0, \ldots, n\}$, appear in the asymptotic variances; its proof is in the Supplementary Material.

LEMMA 3. *Let, for any $\varphi \in \mathcal{L}(\mathcal{X})$,*

$$v_{p,n}(\varphi) = \frac{\mu_{e_p}(\varphi^{\otimes 2}) - \mu_{0_n}(\varphi^{\otimes 2})}{\gamma_n(1)^2}, \qquad p \in \{0, \ldots, n\}. \tag{8}$$

Then $N\mathrm{var}\left\{\gamma_n^N(\varphi)/\gamma_n(1)\right\} \to \sum_{p=0}^n c_p^{-1} v_{p,n}(\varphi)$ *and*

$$NE\left[\left\{\eta_n^N(\varphi) - \eta_n(\varphi)\right\}^2\right] \to \sum_{p=0}^n c_p^{-1} v_{p,n}\{\varphi - \eta_n(\varphi)\}. \tag{9}$$

Remark 3. The map in Proposition 1 satisfies $\sigma_n^2(\varphi) = \sum_{p=0}^n c_p^{-1} v_{p,n}(\varphi)$. We observe that if $Q_p(x_{p-1}, \mathrm{d}x_p) = G_{p-1}(x_{p-1}) M_p(x_{p-1}, \mathrm{d}x_p)$ for $p \in [n]$ and $Q_{n,n} = Id$, $Q_{p,n} = Q_{p+1} \cdots Q_n$ for $p \in \{0, \ldots, n-1\}$, then $\mu_{e_p}(\varphi^{\otimes 2}) = \gamma_p(1)\gamma_p\{Q_{p,n}(\varphi)^2\}$. With Remark 2, we obtain

$$v_{p,n}(\varphi) = \frac{\gamma_p(1)\gamma_p\{Q_{p,n}(\varphi)^2\}}{\gamma_n(1)^2} - \eta_n(\varphi)^2 = \frac{\eta_p\{Q_{p,n}(\varphi)^2\}}{\eta_p Q_{p,n}(1)^2} - \eta_n(\varphi)^2. \tag{10}$$

This particular decomposition of $\sigma_n^2(\varphi)$ is also prominent in the limiting variance for the Central Limit Theorem for $\gamma_n^N(\varphi)$ in Del Moral (2004, Chapter 9).

4. ESTIMATORS

4·1. *Particle approximations of each μ_b*

We now introduce particle approximations to the measures $\{\mu_b : b \in B_n\}$, from which we shall subsequently derive the variance estimators. For each $b \in B_n$, and $\phi \in \mathcal{L}(\mathcal{X}^{\otimes 2})$ we define

$$\mu_b^N(\phi) = \left[\prod_{p=0}^n (N_p)^{b_p} \left(\frac{N_p}{N_p - 1}\right)^{1-b_p}\right] \gamma_n^N(1)^2 E\left[\mathbb{I}\left\{(K^1, K^2) \in \mathcal{I}(b)\right\} \phi(\zeta_n^{K_n^1}, \zeta_n^{K_n^2}) \mid A, \zeta\right]. \tag{11}$$

Recalling from Section 3·1 that given A and ζ, K_n^1 and K_n^2 are conditionally independent and uniformly distributed on $[N_n]$, it follows from (11) that

$$\begin{aligned}\gamma_n^N(\varphi)^2 &= \gamma_n^N(1)^2 \frac{1}{N_n^2} \sum_{i,j\in[N_n]} \varphi(\zeta_n^i)\varphi(\zeta_n^j) \\ &= \gamma_n^N(1)^2 \sum_{b\in B_n} E\left[\mathbb{I}\left\{(K^1,K^2)\in\mathcal{I}(b)\right\}\varphi(\zeta_n^{K_n^1})\varphi(\zeta_n^{K_n^2}) \mid A,\zeta\right] \\ &= \sum_{b\in B_n}\left\{\prod_{p=0}^{n}\left(\frac{1}{N_p}\right)^{b_p}\left(1-\frac{1}{N_p}\right)^{1-b_p}\right\}\mu_b^N(\varphi^{\otimes 2}), \end{aligned} \tag{12}$$

mirroring (7). This identity is complemented by the following result.

THEOREM 2. *For any $b\in B_n$ and $\phi\in\mathcal{L}(\mathcal{X}^{\otimes 2})$,*

1. $E\left\{\mu_b^N(\phi)\right\} = \mu_b(\phi)$ *for all* $N\geq 1$,
2. $\sup_{N\geq 1} NE\left[\left\{\mu_b^N(\phi)-\mu_b(\phi)\right\}^2\right] < \infty$ *and hence* $\mu_b^N(\phi)\to\mu_b(\phi)$ *in probability.*

The proof of Theorem 2 is in the Supplementary Material. Although (11) can be computed in principle from the output of Algorithm 1 without the need for any further simulation, the conditional expectation in (11) involves a summation over all binary strings in $\mathcal{I}(b)$, so calculating $\mu_b^N(\varphi^{\otimes 2})$ in practice may be computationally expensive. Fortunately, relatively simple and computationally efficient expressions are available for $\mu_b^N(\varphi^{\otimes 2})$ in the cases $b=0_n$ and $b=e_p$ (see Lemma 7), and those are the only ones required to construct our variance estimators.

4·2. *Variance estimators*

Our next objective is to explain how (3) is related to the measures μ_b^N and to introduce another family of estimators associated with the individual terms (10).

LEMMA 4. *The following identity of events holds:* $\left\{E_n^{K_n^1}\neq E_n^{K_n^2}\right\} = \left\{(K^1,K^2)\in\mathcal{I}(0_n)\right\}$.

The proof is in the Appendix. Combined with the fact that given (A,ζ), K_n^1, K_n^2 are independent and identically distributed according to the uniform distribution on $[N_n]$, we have

$$E\left[\mathbb{I}\left\{(K^1,K^2)\in\mathcal{I}(0_n)\right\}\phi(\zeta_n^{K_n^1},\zeta_n^{K_n^2}) \mid A,\zeta\right] = N_n^{-2}\sum_{i,j:E_n^i\neq E_n^j}\phi(\zeta_n^i,\zeta_n^j), \tag{13}$$

and therefore we arrive at the following equivalent of (3), written in terms of $\mu_{0_n}^N$,

$$V_n^N(\varphi) = \eta_n^N(\varphi)^2 - \frac{\mu_{0_n}^N(\varphi^{\otimes 2})}{\gamma_n^N(1)^2}. \tag{14}$$

Detailed pseudocode for computing $V_n^N(\varphi)$ in $\mathcal{O}(N)$ time and space upon running Algorithm 1 is provided in the Supplementary Material. Mirroring (8), we now define

$$v_{p,n}^N(\varphi) = \frac{\mu_{e_p}^N(\varphi^{\otimes 2}) - \mu_{0_n}^N(\varphi^{\otimes 2})}{\gamma_n^N(1)^2}, \quad p\in\{0,\ldots,n\}, \qquad v_n^N(\varphi) = \sum_{p=0}^{n} c_p^{-1} v_{p,n}^N(\varphi),$$

and these estimators also satisfy lack-of-bias and weak consistency properties.

THEOREM 3. *For any $\varphi\in\mathcal{L}(\mathcal{X})$,*

1. $E\{\gamma_n^N(1)^2 v_{p,n}^N(\varphi)\} = \gamma_n(1)^2 v_{p,n}(\varphi)$ *for all* $N \geq 1$;
2. $v_{p,n}^N(\varphi) \to v_{p,n}(\varphi)$ *and* $v_{p,n}^N\{\varphi - \eta_n^N(\varphi)\} \to v_{p,n}\{\varphi - \eta_n(\varphi)\}$, *both in probability;*
3. $E\{\gamma_n^N(1)^2 v_n^N(\varphi)\} = \gamma_n(1)^2 \sigma_n^2(\varphi)$ *for all* $N \geq 1$ *and* $v_n^N(\varphi) \to \sigma_n^2(\varphi)$ *in probability.*

Pseudocode for computing each $v_{p,n}^N(\varphi)$ and $v_n^N(\varphi)$ with time and space complexity in $\mathcal{O}(Nn)$ time upon running Algorithm 1 is provided in the Supplementary Material. The time complexity is the same as that of running Algorithm 1, but the space complexity is larger. Empirically, we have found that $NV_n^N(\varphi)$ is very similar to $v_n^N(\varphi)$ as an estimator of $\sigma_n^2(\varphi)$ when N is large enough that they are both accurate, and hence may be preferable due to its reduced space complexity. On the other hand, the estimators $v_{p,n}^N(\varphi)$ and $v_{p,n}^N\{\varphi - \eta_n^N(\varphi)\}$ are the first of their kind to appear in the literature, and may be used to gain insight into the underlying Feynman–Kac model.

5. Use of the estimators to tune the particle filter

5·1. *Asymptotically optimal allocation*

The variance estimators can be used to report Monte Carlo error alongside particle approximations, but may also be useful in algorithm design and tuning. Here and in Section 5·2 we provide simple examples to illustrate this point. To simplify presentation, we focus on performance in estimating $\gamma_n^N(\varphi)$, but the ideas can easily be modified to deal instead with $\eta_n^N(\varphi)$.

The following well known result is closely related to Neyman's optimal allocation in stratified random sampling (Tschuprow, 1923; Neyman, 1934). A short proof using Jensen's inequality can be found in Glasserman (2004, Section 4.3).

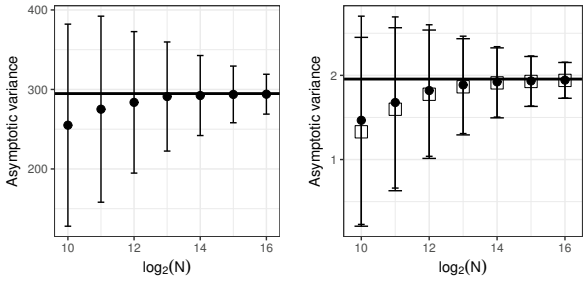
LEMMA 5. *Let* $a_0, \ldots, a_n \geq 0$. *The function* $(c_0, \ldots, c_n) \mapsto \sum_{p=0}^n c_p^{-1} a_p$ *is minimized, subject to* $\min_p c_p > 0$ *and* $\sum_{p=0}^n c_p = n + 1$, *at* $(n+1)^{-1}(\sum_{p=0}^n a_p^{1/2})^2$ *when* $c_p \propto a_p^{1/2}$.

As a consequence, we can in principle minimize $\sigma_n^2(\varphi)$ by choosing $c_p \propto v_{p,n}(\varphi)^{1/2}$. An approximation of this optimal allocation can be obtained by the following two-stage procedure. First run a particle filter with $N_p = N$ to obtain the estimates $v_{p,n}^N(\varphi)$ and then define $c_{0:n}$ by $c_p = \max\{v_{p,n}^N(\varphi), g(N)\}^{1/2}$, where g is some positive but decreasing function with $\lim_{N\to\infty} g(N) = 0$. Then run a second particle filter with each $N_p = \lceil c_p N \rceil$, and report the quantities of interest, e.g., $\gamma_n^N(\varphi)$. The function g is chosen to ensure that $c_p > 0$ and that for large N we permit small values of c_p. The quantity $\sum_{p=0}^n v_{p,n}^N(\varphi)/\{\sum_{p=0}^n c_p^{-1} v_{p,n}^N(\varphi)\}$, obtained from the first run, is an indication of the improvement in variance using the new allocation.

Approximately optimal allocation has previously been addressed by Bhadra & Ionides (2016), who introduced a meta-model to approximate the distribution of the Monte Carlo error associated with $\log \gamma_n^N(1)$ in terms of an autoregressive process, the objective function to be minimized then being the variance under this meta-model. They provide only empirical evidence for the fit of their meta-model, whereas our approach targets the true asymptotic variance $\sigma_n^2(\varphi)$ directly.

5·2. *An adaptive particle filter*

Monte Carlo errors of particle filter approximations can be sensitive to N, and an adequate value of N to achieve a given error may not be known a priori. The following procedure increases N until $V_n^N(\varphi)$ is in a given interval. Given an initial number of particles $N^{(0)}$ and a threshold $\delta > 0$, one can run successive particle filters, doubling the number of particles each time, until the associated random variable $V_n^{N^{(\tau)}}(\varphi) \in [0, \delta]$. Finally, one runs a final particle filter with



(a) $\varphi \equiv 1$ (b) $\varphi = Id - \eta_n^N(Id)$

Figure 2: Estimated asymptotic variances $NV_n^N(\varphi)$ (dots and error bars for the mean $\pm$ one standard deviation from 10^4 replicates) against $\log_2 N$ for the linear Gaussian example. The horizontal lines correspond to the true asymptotic variances. The sample variances of $\gamma_n^N(1)/\gamma_n(1)$ and $\eta_n^N(Id)$, scaled by N, were close to their asymptotic variances. Corresponding results for the estimator of Chan & Lai (2013) are overlaid with boxes instead of dots and wider tick marks on the error bars.

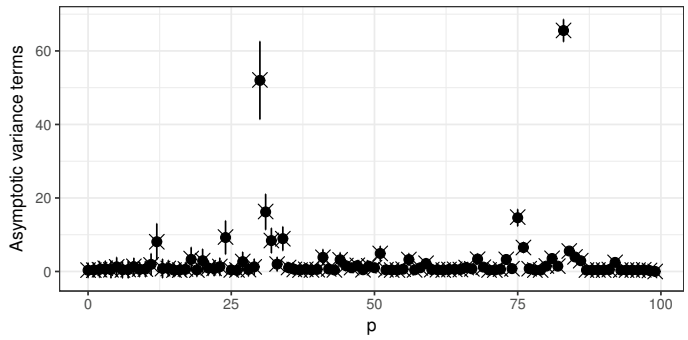


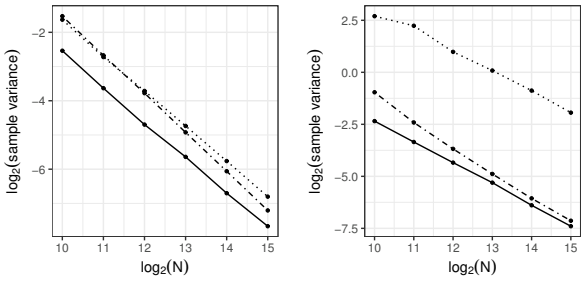
Figure 3: Plot of $v_{p,n}^N(1)$ (dots and error bars for the mean $\pm$ one standard deviation from 10^3 replicates) and $v_{p,n}(1)$ (crosses) at each $p \in \{0, \ldots, n\}$ for the Linear Gaussian example, with $N = 10^5$.

$N^{(\tau)}$ particles, and returns the estimate of interest. In Section 6 we provide empirical evidence that this procedure can be effective in some applications.

6. Applications and illustrations

6·1. *Linear Gaussian hidden Markov model*

This model is specified by $M_0(\cdot) = \mathcal{N}(\cdot; 0, 1)$, $M_p(x_{p-1}, \cdot) = \mathcal{N}(\cdot; 0.9x_{p-1}, 1)$ and $G_p(x_p) = \mathcal{N}(y_p; x_p, 1)$. The measures η_n and γ_n are available in closed form via the Kalman filter, and $v_{p,n}(\varphi)$ can be computed exactly and very accurately for $\varphi \equiv 1$ and $\varphi = Id$ respectively, allowing us to assess the accuracy of our estimators. We used a synthetic dataset, simulated according to the model with $n = 99$. A Monte Carlo study with 10^4 replicates of $V_n^N(\varphi)$ for each value of N and $c_p \equiv 1$ was used to measure the accuracy of the estimate $NV_n^N(\varphi)$ as N grows; results are displayed in Figure 2 and for this data $\sigma_n^2(1) = 294.791$ and



(a) (b)

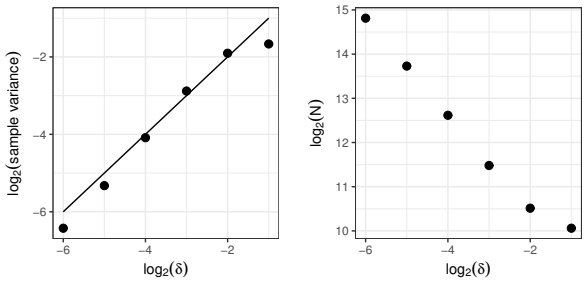
Figure 4: Logarithmic plots of sample variance for 10^4 replicates of $\gamma_n^N(1)/\gamma_n(1)$ against N for the linear Gaussian example, using a constant N particle filter (dotted), the approximately asymptotically optimal particle filter (dot-dash), and the asymptotically optimal particle filter (solid). In Figure 4b, the observation sequence is $y_p = 0$ for $p \in \{0, \ldots, 99\} \setminus \{49\}$ and $y_{49} = 8$.

$\sigma_n^2\{Id - \eta_n(Id)\} \approx 1.95$. The Chan & Lai (2013) estimator of $\sigma_n^2\{Id - \eta_n(Id)\}$ is fairly similar for N large enough that the variance estimator is approximately unbiased. The estimates $v_n^N(\varphi)$ differed very little from $NV_n^N(\varphi)$, and so are not shown. We then tested the accuracy of the estimates $v_{p,n}^N(1)$; results are displayed in Figure 3. The estimates $v_{p,n}^N\{Id - \eta_n^N(Id)\}$ are very close to 0 for $p < 95$ with values $(0.009, 0.07, 0.39, 1.48)$ for $p \in \{96, 97, 98, 99\}$; this behaviour is in keeping with time-uniform bounds on asymptotic variances (Whiteley, 2013, and references therein) .

We also compared a constant N particle filter, the asymptotically optimal particle filter where the asymptotically optimal allocation is computed exactly, and its approximation described in Section 5·1 for different values of N using a Monte Carlo study with 10^4 replicates. We took $g(N) = 2/\log_2 N$ in defining the approximation, and the results in Figure 4a indicate that the approximation reduces the variance. The improvement is fairly modest for this particular model, and indeed the exact asymptotic variances associated with the constant N and asymptotically optimal particle filters differ by less than a factor of 2. In contrast, Figure 4b shows that the improvement can be fairly dramatic in the presence of outlying observations; the improvement in variance there is by a factor of close to 40. Finally, we tested the adaptive particle filter described in Section 5·2 using 10^4 replicates for each value of δ; results are displayed in Figure 5, and indicate that the variances are close to their prescribed thresholds.

6·2. *Stochastic volatility hidden Markov model*

A stochastic volatility model is defined by $M_0(\cdot) = \mathcal{N}\left\{\cdot\,; 0, \sigma^2/(1-\rho^2)\right\}$, $M_p(x_{p-1}, \cdot) = \mathcal{N}(\cdot\,; \rho x_{p-1}, \sigma^2)$ and $G_p(x_p) = \mathcal{N}(y_p; 0, \beta^2 \exp(x_p))$. We used the pound/dollar daily exchange rates for 100 consecutive weekdays ending on 28th June, 1985, a subset of the well-known dataset analyzed in Harvey, Ruiz and Shephard (1994). Our results are obtained by choosing the parameters $(\rho, \sigma, \beta) = (0.95, 0.25, 0.5)$. We provide in the supplement plots of the accuracy of the estimate $NV_n^N(\varphi)$ as N grows using 10^4 replicates for each value of N; the asymptotic variances $\sigma_n^2(1)$ and $\sigma_n^2(Id - \eta_n(Id))$ are estimated as being approximately 347 and 1.24 respectively. In the Supplementary Material we plot the estimates of $v_{p,n}(\varphi)$. We found modest improvement for the approximation of the asymptotically optimal particle filter, as one could infer from the estimated $v_{p,n}(\varphi)$ and Lemma 5. For the simple adaptive N particle filter, results in the Supplementary Material indicate that the variances are close to their prescribed thresholds.



(a) (b)

Figure 5: Logarithmic plots for the simple adaptive N particle filter estimates of $\gamma_n(1)$ for the linear Gaussian example. Figure (a) plots the sample variance of $\gamma_n^N(1)/\gamma_n(1)$ against δ, with the straight line $y = x$. Figure (b) plots N against δ, where N is the average number of particles used by the final particle filter.

6·3. *A Sequential Monte Carlo sampler*

We consider a sequential simulation problem, as described in Section 2·3, with $\mathsf{X} = \mathbb{R}$, $\bar{\pi}_0(x) = \mathcal{N}(0, 10^2)$ and $\bar{\pi}_1(x) = 0.3\mathcal{N}(x; -10, 0.1^2) + 0.7\mathcal{N}(x; 10, 0.2^2)$. The distribution π_1 is bi-modal with well-separated modes. With $n = 11$, and the sequence of tempering parameters

$$\beta_{0:n} = (0, 0.0005, 0.001, 0.0025, 0.005, 0.01, 0.025, 0.05, 0.1, 0.25, 0.5, 1),$$

we let each Markov kernel $M_p, p \in \{1, \ldots, n\}$ be an η_p-invariant random walk Metropolis kernel iterated $k = 10$ times with proposal variance τ_p^2, where $\tau_{1:n} = (10, 9, 8, 7, 6, 5, 4, 3, 2, 1, 1)$.

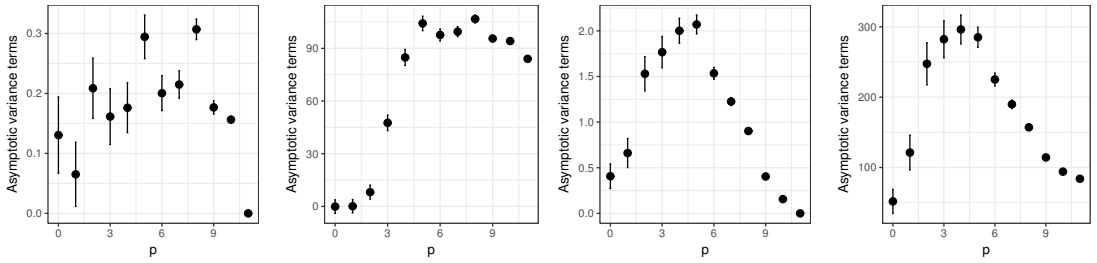
One striking difference between the estimates for this model and those for the hidden Markov models above is that the asymptotic variance $\sigma_n^2\{Id - \eta_n(Id)\} \approx 822$ is considerably larger than $\sigma_n^2(1) \approx 2.1$; the variability of the estimates $NV_n^N(\varphi)$ is shown in the Supplementary Material. Inspection of the estimates of $v_{p,n}(\varphi)$ in Figures 6 allows us to investigate both this difference and the dependence of $v_{p,n}(\varphi)$ on k in greater detail.

Figure 6(a)–(b) shows that while $v_{p,n}(1)$ is small for all p, the values of $v_{p,n}\{Id - \eta_n(Id)\}$ are larger for large p than for small p; this could be due to the inability of the Metropolis kernels $(M_q)_{q \geq p}$ to mix well due to the separation of the modes in $(\eta_q)_{q \geq p}$ when p is large. In Figure 6(c)–(d), $k = 1$, that is each M_p consists of only a single iterate of a Metropolis kernel, and we see that the values of $v_{p,n}(\varphi)$ associated with small p are much larger than when $k = 10$, indicating that the larger number of iterates does improve the asymptotic variance of the particle approximations. However, the impact on $v_{p,n}(\varphi)$ is less pronounced for large p. Results for the simple adaptive N particle filter approximating $\eta_n(Id)$ are provided in the supplement, which again show that the estimates are close to their prescribed thresholds.

7. Discussion

7·1. *Alternatives to the bootstrap particle filter*

In the hidden Markov model examples above, we have constructed the Feynman–Kac measures taking $M_0, \ldots, M_n$ to be the initial distribution and transition probabilities of the latent process and defining $G_0, \ldots, G_n$ to incorporate the realized observations. This is only one, albeit important, way to construct particle approximations of η_n, and the algorithm itself is usually referred to as the bootstrap particle filter. Alternative specifications of $(M_p, G_p)_{0 \leq p \leq n}$ lead to dif-



(a) $\varphi \equiv 1, k = 10$ (b) $\varphi = Id - \eta_n(Id), k = 10$ (c) $\varphi \equiv 1, k = 1$ (d) $\varphi = Id - \eta_n(Id), k = 1$

Figure 6: Plot of $v_{p,n}^N(\varphi)$ (dots and error bars for the mean $\pm$ one standard deviation) at each $p \in \{0, \ldots, n\}$ with $k = 10$ iterations (a)–(b) and $k = 1$ iteration (c)–(d) for each Markov kernel in the sequential Monte Carlo sampler example and $N = 10^5$.

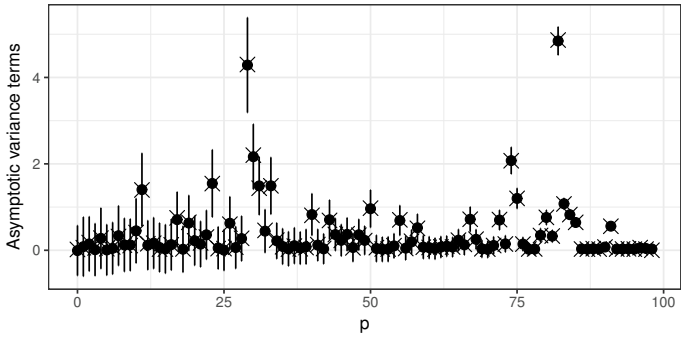


Figure 7: Plot of $\check{v}_{p,n}^N(1)$ (dots and error bars for the mean $\pm$ one standard deviation) and $\check{v}_{p,n}(1)$ (crosses) at each $p \in \{0, \ldots, n\}$ in the linear Gaussian example.

ferent Feynman–Kac models, as discussed in Del Moral (2004, Section 2.4.2), and the variance estimators introduced here are applicable to these models as well.

One particular specification corresponds to the fully adapted auxiliary particle filter of Pitt & Shephard (1999), as discussed by Doucet & Johansen (2008). Specifically, we define $\check{M}_0(\mathrm{d}x_0) = M_0(\mathrm{d}x_0)G_0(x_0)/M_0(G_0)$, and

$$\check{M}_p(x_{p-1}, \mathrm{d}x_p) = \frac{M_p(x_{p-1}, \mathrm{d}x_p)G_p(x_p)}{M_p(G_p)(x_{p-1})}, \qquad p \in [n],$$

and then $\check{G}_0(x_0) = M_0(G_0)M_1(G_1)(x_0)$ and $\check{G}_p(x_p) = M_{p+1}(G_{p+1})(x_p)$, $p \geq 1$. If we denote by $\check{\gamma}_n$ and $\check{\eta}_n$ the Feynman–Kac measures associated with $(\check{M}_p, \check{G}_p)_{0 \leq p \leq n}$, we obtain $\check{\gamma}_{n-1}(1) = \gamma_n(1)$. Moreover, the variance of $\check{\gamma}_{n-1}^N(1)$ is often smaller than the variance of $\gamma_n^N(1)$. In Figure 7, we plot the corresponding $\check{v}_{p,n-1}(1)$ and their approximations for the same linear Gaussian example in Section 6·1. Here, the asymptotic variance of $\check{\gamma}_{n-1}^N(1)/\check{\gamma}_{n-1}(1)$ is 40.679, more than 7 times smaller than $\sigma_n^2(1)$.

7·2. *Estimators based on independent, identically distributed replicates*

It is clearly possible to consistently estimate the variance of $\gamma_n^N(\varphi)/\gamma_n(1)$ by using independent identically distributed replicates of γ_n^N. Such estimates necessarily entail simulation

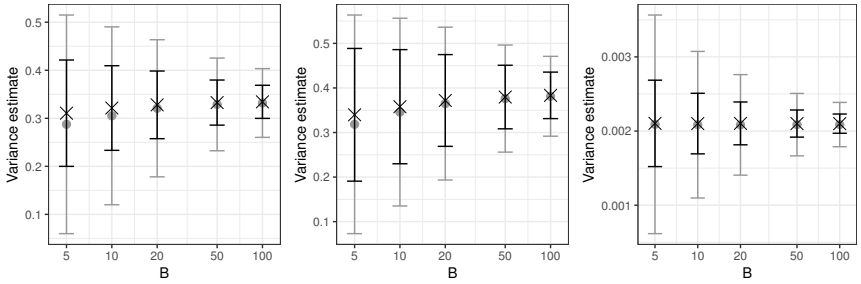


Figure 8: Plot of the standard estimate of $\mathrm{var}\left\{\gamma_n^N(\varphi)/\gamma_n(1)\right\}$ (gray dots and error bars) and the alternative estimate using $V_n^N(1)$ (black crosses and error bars) against B in (left to right) the examples of Sections 6·1–6·3.

of multiple particle filters. We now compare the accuracy of such estimates with those based on independent, identically distributed replicates of $V_n^N(\varphi)$. For some $\varphi \in \mathcal{L}(\mathcal{X})$ and $B \in \mathbb{N}$, let $\gamma_{n,i}^N(\varphi)$ and $V_{n,i}^N(\varphi)$ be i.i.d. replicates for $i \in [B]$, and define $M = N^{-1}\sum_{i\in[B]}\gamma_{n,i}^N(1)$. A standard estimate of $\mathrm{var}\left\{\gamma_n^N(\varphi)/\gamma_n(1)\right\}$ is obtained by calculating the sample variance of $\{M^{-1}\gamma_{n,i}^N(\varphi)\,;\, i \in [B]\}$. Noting the lack-of-bias of $\gamma_n^N(1)^2 V_n^N(\varphi)$, an alternative estimate of $\mathrm{var}\left\{\gamma_n^N(\varphi)/\gamma_n(1)\right\}$ can be obtained as $B^{-1}\sum_{i\in[B]}\left\{M^{-1}\gamma_{n,i}^N(1)\right\}^2 V_{n,i}^N(\varphi)$. Both these estimates can be seen as ratios of simple Monte Carlo estimates of $\mathrm{var}\left\{\gamma_n^N(\varphi)\right\}$ and $\gamma_n(1)^2$, and are therefore consistent as $B \to \infty$. We show in Figure 8 a comparison between these estimates for the three models discussed in Section 6 with $N = 10^3$ and $\varphi \equiv 1$, and we can see that the alternative estimate based on $V_n^N(1)$ is empirically more accurate for these examples.

7·3. *Final remarks*

The particular approximations developed here provide a natural way to estimate the terms appearing in the non-asymptotic second moment expression (7). We have also provided the first generally applicable, consistent estimators of $v_{p,n}(\varphi)$. The expression (7) does not apply to particle approximations with resampling schemes other than multinomial; one possible avenue of future research is to investigate these other settings. Whilst we have emphasized variances and asymptotic variances, the measures μ_b also appear in expressions which describe propagation of chaos properties of the particle system. For instance, in the situation $N_p \equiv N$, the asymptotic bias formula of Del Moral et al. (2007, p.7.) can be expressed as

$$NE\left\{\eta_n^N(\varphi) - \eta_n(\varphi)\right\} \to -\sum_{p=0}^{n-1}\frac{\eta_p\left\{Q_{p,n}(1)Q_{p,n}(\varphi - \eta_n(\varphi))\right\}}{\eta_p Q_{p,n}(1)^2} \equiv -\sum_{p=0}^{n-1}\frac{\mu_{e_p}\left\{1\otimes(\varphi-\eta_n(\varphi))\right\}}{\gamma_n(1)^2},$$

which can be consistently estimated using $\mu_{e_p}^N$ and γ_n^N. Finally, the technique used to prove Lemma 2 can be generalized to arbitrary positive integer moments of $\gamma_n^N(\varphi)$.

In many applications, particularly in the context of hidden Markov models, particle filters are used to approximate conditional expectations with respect to updated Feynman–Kac measures. We define these, their approximations, and provide corresponding variance estimators in the supplement. Of some interest is the updated estimator $\hat{\gamma}_{n-1}^N(1) = \gamma_n^N(1)$ whose variance estimator is $\hat{V}_{n-1}^N(1) = V_{n-1}^N(G_{n-1})/\eta_{n-1}^N(G_{n-1})^2 \neq V_n^N(1)$. In fact, $V_n^N(1)$ is an unbiased, noisy approxi-

mation of $\hat{V}_{n-1}^N(1)$, due to using E_n instead of G_{n-1} and ζ_{n-1}. However, empirical investigations indicate that the difference in variance is practically negligible.

ACKNOWLEDGEMENT

This paper was first submitted when Anthony Lee was at the University of Warwick. He was also partially supported by the Alan Turing Institute.

SUPPLEMENTARY MATERIAL

The supplementary material includes algorithms for efficient computation of the variance estimators, results for estimators associated with updated Feynman–Kac measures, figures and proofs.

APPENDIX

Proof of Theorem 1. Throughout the proof, $\to$ denotes convergence in probability. For part 1., the fact $\mu_{0_n}(\varphi^{\otimes 2}) = \gamma_n(\varphi)^2$ and Theorem 2 together give

$$E\left\{\gamma_n^N(1)^2 V_n^N(\varphi)\right\} = E\left\{\gamma_n^N(\varphi)^2 - \mu_{0_n}^N(\varphi^{\otimes 2})\right\} = E\left\{\gamma_n^N(\varphi)^2\right\} - \gamma_n(\varphi)^2 = \mathrm{var}\left\{\gamma_n^N(\varphi)\right\}.$$

For part 2., combining the identity (12), $\mu_b^N(\varphi^{\otimes 2}) \to \mu_b(\varphi^{\otimes 2})$ by Theorem 2, and the fact that for any $b \in B_n$ other than 0_n and $e_0, \ldots, e_n$, $\prod_{p=0}^n \left(\frac{1}{N_p}\right)^{b_p} \left(1 - \frac{1}{N_p}\right)^{1-b_p}$ is in $\mathcal{O}(N^{-2})$, we obtain

$$\gamma_n^N(\varphi)^2 - \mu_{0_n}^N(\varphi^{\otimes 2}) = \left\{\sum_{p=0}^n \frac{\mu_{e_p}^N(\varphi^{\otimes 2}) - \mu_{0_n}^N(\varphi^{\otimes 2})}{\lceil c_p N \rceil}\right\} + \mathcal{O}_p(N^{-2}). \tag{A1}$$

Also noting that by Proposition 1 $\gamma_n^N(1)^2 \to \gamma_n(1)^2$, from (8) that $\gamma_n(1)^2 v_{p,n}(\varphi) = \mu_{e_p}(\varphi^{\otimes 2}) - \mu_{0_n}(\varphi^{\otimes 2})$ and again using $\mu_b^N(\varphi^{\otimes 2}) \to \mu_b(\varphi^{\otimes 2})$, we then have

$$NV_n^N(\varphi) = \frac{N}{\gamma_n^N(1)^2}\left\{\gamma_n^N(\varphi)^2 - \mu_{0_n}^N(\varphi^{\otimes 2})\right\} \to \sum_{p=0}^n \frac{v_{p,n}(\varphi)}{c_p} = \sigma_n^2(\varphi). \tag{A2}$$

For part 3., first note that by Theorem 2 and Proposition 1, for any $b \in B_n$,

$$\begin{aligned}\mu_b^N[\{\varphi - \eta_n^N(\varphi)\}^{\otimes 2}] &= \mu_b^N(\varphi^{\otimes 2}) - \eta_n^N(\varphi)\{\mu_b^N(\varphi \otimes 1) + \mu_b^N(1 \otimes \varphi)\} + \eta_n^N(\varphi)^2 \mu_b^N(1^{\otimes 2}) \\ &\to \mu_b[\{\varphi - \eta_n(\varphi)\}^{\otimes 2}],\end{aligned}$$

□

from which it follows that (A1) also holds with φ replaced by $\varphi - \eta_n^N(\varphi)$, and similarly to (A2),

$$NV_n^N\{\varphi - \eta_n^N(\varphi)\} \to \sum_{p=0}^n \frac{v_{p,n}\{\varphi - \eta_n(\varphi)\}}{c_p} = \sigma_n^2\{\varphi - \eta_n(\varphi)\}.$$

Proof of Lemma 4. For $i \in [N_n]$ define $B_{n-1}^i = A_{n-1}^i$ and $B_{p-1}^i = A_{p-1}^{B_p^i}$ for $p \in [n-1]$. Since in Algorithm 1, $E_p^i = E_{p-1}^{A_{p-1}^i}$ for all $p \in [n], i \in [N_p]$, a simple inductive argument then shows that

$$E_n^i = E_p^{B_p^i}, \quad p \in \{0, \ldots, n\},\ i \in [N_n]. \tag{A3}$$

We shall now prove $(K^1, K^2) \in \mathcal{I}(0_n) \Rightarrow E_n^{K_n^1} \neq E_n^{K_n^2}$. Recall from Section 3·1 that when $(K^1, K^2) \in \mathcal{I}(0_n)$, we have $A_{p-1}^{K_p^1} = K_{p-1}^1 \neq K_{p-1}^2 = A_{p-1}^{K_p^2}$ for all $p \in [n]$, hence $B_0^{K_n^1} = K_0^1 \neq K_0^2 = B_0^{K_n^2}$. Applying (A3) with $p = 0$ and using the fact that in Algorithm 1, $E_0^i = i$ for all $i \in [N_n]$, we have

$E_n^i = E_0^{B_0^i} = B_0^i$, hence $E_n^{K_n^1} = B_0^{K_n^1} \neq B_0^{K_n^2} = E_n^{K_n^2}$ as required. It remains to prove $(K^1, K^2) \notin \mathcal{I}(0_n) \Rightarrow E_n^{K_n^1} = E_n^{K_n^2}$. Assuming $(K^1, K^2) \notin \mathcal{I}(0_n)$, consider $\tau = \max\{p : K_p^1 = K_p^2\}$. If $\tau = n$ then clearly $E_n^{K_n^1} = E_n^{K_n^2}$, so suppose $\tau < n$. It follows from Section 3·1 that $B_\tau^{K_n^1} = K_\tau^1 = K_\tau^2 = B_\tau^{K_n^2}$, so taking $p = \tau$ and $i = K_n^1, K_n^2$ in (A3) gives $E_n^{K_n^1} = E_n^{K_n^2}$. □

Proof of Theorem 3. For part 1., Theorem 2 gives

$$E\left\{\gamma_n^N(1)^2 v_{p,n}^N(\varphi)\right\} = E\left\{\mu_{e_p}^N(\varphi^{\otimes 2}) - \mu_{0_n}^N(\varphi^{\otimes 2})\right\} = \mu_{e_p}(\varphi^{\otimes 2}) - \mu_{0_n}(\varphi^{\otimes 2}) = \gamma_n(1)^2 v_{p,n}(\varphi).$$

For the remainder of the proof, $\to$ denotes convergence in probability. For part 2., $\mu_{e_p}^N(\varphi^{\otimes 2}) - \mu_{0_n}^N(\varphi^{\otimes 2}) \to \gamma_n(1)^2 v_{p,n}(\varphi)$ by Theorem 2, and $\gamma_n^N(1)^2 \to \gamma_n(1)^2$ by Proposition 1, so $v_{p,n}^N(\varphi) = \{\mu_{e_p}^N(\varphi^{\otimes 2}) - \mu_{0_n}^N(\varphi^{\otimes 2})\}/\gamma_n^N(1)^2 \to v_{p,n}(\varphi)$; as in the proof of Theorem 1, $\mu_b^N[\{\varphi - \eta_n^N(\varphi)\}^{\otimes 2}] \to \mu_b[\{\varphi - \eta_n(\varphi)\}^{\otimes 2}]$ gives $v_{p,n}^N\{\varphi - \eta_n^N(\varphi)\} \to v_{p,n}\{\varphi - \eta_n(\varphi)\}$. Part 3. follows from parts 1. and 2. □

BIBLIOGRAPHY

ANDRIEU, C., LEE, A. & VIHOLA, M. (2018). Uniform ergodicity of the iterated conditional SMC and geometric ergodicity of particle Gibbs samplers. *Bernoulli* **24**, 842–872.

BHADRA, A. & IONIDES, E. L. (2016). Adaptive particle allocation in iterated sequential Monte Carlo via approximating meta-models. *Stat. Comput.* **26**, 393–407.

CÉROU, F., DEL MORAL, P. & GUYADER, A. (2011). A nonasymptotic theorem for unnormalized Feynman–Kac particle models. *Ann. Inst. H. Poincaré Probab. Statist.* **47**, 629–649.

CHAN, H. P. & LAI, T. L. (2013). A general theory of particle filters in hidden Markov models and some applications. *Ann. Statist.* **41**, 2877–2904.

CHOPIN, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.* **32**, 2385–2411.

DEL MORAL, P. (2004). *Feynman–Kac formulae: Genealogical and Interacting Particle Systems with Applications*. New York: Springer Verlag.

DEL MORAL, P., DOUCET, A. & JASRA, A. (2006). Sequential Monte Carlo samplers. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68**, 411–436.

DEL MORAL, P., DOUCET, A. & PETERS, G. W. (2007). Sharp propagation of chaos estimates for Feynman–Kac particle models. *Theory Probab. Appl.* **51**, 459–485.

DEL MORAL, P. & GUIONNET, A. (1999). Central limit theorem for nonlinear filtering and interacting particle systems. *Ann. Appl. Probab.* **9**, 275–297.

DEL MORAL, P. & MICLO, L. (2001). Genealogies and increasing propagation of chaos for Feynman–Kac and genetic models. *Ann. Appl. Probab.* **11**, 1166–1198.

DOUC, R., GUILLIN, A. & NAJIM, J. (2005). Moderate deviations for particle filtering. *Ann. Appl. Probab.* **15**, 587–614.

DOUC, R. & MOULINES, E. (2008). Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Ann. Statist.* **36**, 2344–2376.

DOUCET, A. & JOHANSEN, A. M. (2008). A tutorial on particle filtering and smoothing: fifteen years later. In *Handbook of Nonlinear Filtering*, D. Crisan & B. Rozovsky, eds. Oxford University Press.

GLASSERMAN, P. (2004). *Monte Carlo Methods in Financial Engineering*. New York: Springer.

KÜNSCH, H. (2005). Recursive Monte Carlo filters: algorithms and theoretical analysis. *Ann. Statist.* **33**, 1983–2021.

NEYMAN, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *J. Roy. Statist. Soc.* **97**, 558–625.

OLSSON, J. & DOUC, R. (2018). Numerically stable online estimation of variance in particle filters. *Bernoulli.* To appear.

PITT, M. K. & SHEPHARD, N. (1999). Filtering via simulation: auxiliary particle filters. *J. Amer. Statist. Assoc.* **94**, 590–599.

TSCHUPROW, A. A. (1923). On the mathematical expectation of the moments of frequency distributions in the case of correlated observations. *Metron* **2**, 461–493, 646–683.

WHITELEY, N. (2013). Stability properties of some particle filters. *Ann. Appl. Prob.* **23**, 2500–2537.