Adv. Appl. Prob. **48,** 69–87 (2016)
*doi:*10.1017/apr.2015.7

THEORY OF SEGMENTED PARTICLE FILTERS

HOCK PENG CHAN,* **
CHIANG-WEE HENG * *** AND
AJAY JASRA,* **** *National University of Singapore*

Abstract

We study the asymptotic behavior of a new particle filter approach for the estimation of hidden Markov models. In particular, we develop an algorithm where the latent-state sequence is segmented into multiple shorter portions, with an estimation technique based upon a separate particle filter in each portion. The partitioning facilitates the use of parallel processing, which reduces the wall-clock computational time. Based upon this approach, we introduce new estimators of the latent states and likelihood which have similar or better variance properties compared to estimators derived from standard particle filters. We show that the likelihood function estimator is unbiased, and show asymptotic normality of the underlying estimators.

Keywords: Central limit theorem; parallel processing; sequential Monte Carlo; standard error estimation

2010 Mathematics Subject Classification: Primary 64C05
Secondary 62F15

1. Introduction

Hidden Markov models (HMMs) are a flexible class of statistical models that are applied in a wide variety of applications such as bioinformatics, economics, engineering, and finance; see [5] for an introduction. Mathematically, a HMM corresponds to a pair of discrete-time processes $X_t \in \mathsf{X}$, $Y_t \in \mathsf{Y}$, with the observed Y_t conditionally independent given X_t and the unobserved X_t obeying a first-order Markov chain

$$(X_t \mid X_{t-1} = x) \sim \mathbb{P}_\theta(\cdot \mid x), \qquad (Y_t \mid X_t = x) \sim G_\theta(\cdot \mid x), \quad t \geq 1 \tag{1.1}$$

with densities, p_θ and g_θ with respect to dominating measures on their state-spaces and θ is a static parameter.

From an inferential perspective, we are interested in the conditional distribution of X_t given all the observations up to and perhaps after time t. This has a wide-range of interpretations, particularly in real-time applications. In addition there is much practical interest in the calculation of the likelihood of the observations, for model comparison and parameter estimation. The difficulty with the aforementioned objectives is that the exact computation of the conditional distribution or likelihood is typically not possible, as often the high-dimensional integrals that

Received 16 September 2014; revision received 10 March 2015.
* Postal address: Department of Statistics and Applied Probability, National University of Singapore, 6 Science Drive 2, 117546, Singapore.
** Email address: stachp@nus.edu.sg
*** Email address: a0068363@nus.edu.sg
**** Email address: staja@nus.edu.sg
Supported by MOE grants R-155-000-119-133 and R-155-000-120-112.

it depends on are often difficult to evaluate. In practice Monte Carlo-based numerical methods are adopted, in particular the method of particle filters or, equivalently, sequential Monte Carlo (SMC); see [12] for an overview.

SMC methods can be described as a collection of techniques that approximate a sequence of distributions, known up to normalizing constants, of increasing dimensions, and are often applied to HMMs. SMC methods combine importance sampling and resampling to approximate distributions. The idea is to introduce a sequence of proposal densities and sequentially simulate a collection of $K > 1$ samples, termed particles, in parallel from these proposals. In most scenarios it is not possible to use the distribution of interest as a proposal. Therefore, one must correct for the discrepancy between proposal and target via importance weights. As the variance of these importance weights can potentially increase exponentially with algorithmic time, resampling is applied to control it. Resampling consists of sampling with replacement from the current samples using the weights and then resetting them to K^{-1}. The theoretical properties of SMC with regards to their convergence as K grows are well studied; see [6], [8], [10], [11], and [15].

In recent years, the applicability of SMC techniques has been enhanced by parallel computation; see [17]. One of the main bottlenecks in the application of parallel computation to SMC methods is the resampling step, a major requirement for the method to be efficient. This has led to a number of researchers investigating methodologies that reduce the degree of interaction in SMC algorithms; see [16] and [18]. This work is complementary to the aforementioned references, and is a methodology designed to assist in the parallelization of SMC algorithms, while attempting to retain their attractive properties. Our objective is to study the asymptotic behavior of HMM estimators when the latent-state sequence is segmented into multiple shorter portions, by applying an estimation technique based upon a separate particle filter in each portion. The partitioning facilitates the use of parallel processing. Based upon this approach, we introduce new SMC-based estimators of the latent states (that is, expectations with respect to the filter and smoother) and likelihood with similar or better variance properties compared to standard SMC estimators, but due to parallelization can be calculated in less wall-clock computational time. In particular we show the following:

(i) unbiasedness of our likelihood estimate;

(ii) central limit convergences of the likelihood and latent-state estimates;

(iii) consistent estimation of asymptotic variances.

Our likelihood estimates can be used in conjunction with recent advances in SMC methodology in which particle filter processing is just one component of a two-layered process when learning θ in a Bayesian manner: particle MCMC (PMCMC) [2], SMC^2 [9], and MCMC substitution [7]. That is, our procedure can be used routinely in the context of these works. In principle, there is no need to break up the observation sequence into strictly disjoint segments; it can be advantageous to include additional observations at the edges of the segments to smooth out the joining of the sample paths. This latter interesting idea is still subject to further study, however.

We describe the algorithm and estimators in Section 2 and the asymptotic theory in Section 3, with an illustration of variance reduction in smoothed latent-state estimators. We discuss refinements in Section 4. The technical proofs are consolidated in Appendix A.

2. Independent particle filters for segmented data

Let $\boldsymbol{Y}_U = (Y_1, \ldots, Y_U)$ for some $U > 1$. As the observation sequence is conditionally independent given the latent-state sequence, the density of $\boldsymbol{X}_U := (X_1, \ldots, X_U)$ conditioned on $\boldsymbol{Y}_U$ is given by

$$p_\theta(\boldsymbol{x}_U \mid \boldsymbol{Y}_U) = \frac{\prod_{t=1}^{U}[p_\theta(x_t \mid x_{t-1}) g_\theta(Y_t \mid x_t)]}{\lambda(\theta)}, \tag{2.1}$$

where $\lambda(\theta)[= \lambda(\boldsymbol{Y}_U \mid \theta)]$ is the likelihood function that normalizes $p_\theta(\cdot \mid \boldsymbol{Y}_U)$ so that it integrates to 1, and $p_\theta(x_t \mid x_{t-1})$ for $t = 1$ denotes $p_\theta(x_1)$.

Let $\boldsymbol{x}_t = (x_1, \ldots, x_t)$ and let $q_t(\cdot \mid \boldsymbol{x}_{t-1})$ be an importance density of X_t, with $q_t(\cdot \mid \boldsymbol{x}_{t-1})$ for $t = 1$ denoting $q_1(\cdot)$. We shall require that $q_t(x_t \mid \boldsymbol{x}_{t-1}) > 0$ whenever $p_\theta(x_t \mid x_{t-1}) > 0$. For notational simplicity, we assume that $U = MT$ for positive integers M and T, so that the latent-state sequence can be partitioned neatly into M subsequences of equal length T. We shall operate M particle filters independently, with the mth particle filtering generating sample paths of $\boldsymbol{X}_{m,mT}$, where $\boldsymbol{X}_{m,t} = (X_{(m-1)T+1}, \ldots, X_t)$. Due to the independent nature of the particle filters, we require that for $(m-1)T < t \le mT$, $q_t(\cdot \mid \boldsymbol{x}_{t-1})$ does not depend on $\boldsymbol{x}_{(m-1)T}$. We can thus express $q_t(\cdot \mid \boldsymbol{x}_{t-1})$ as $q_t(\cdot \mid \boldsymbol{x}_{m,t-1})$.

Let $w_t(\boldsymbol{x}_t)$ be the positive resampling weights of a sample path $\boldsymbol{x}_t$, and again due to the independent nature of the particle filters, we shall require that for $(m-1)T < t \le mT$, $w_t(\boldsymbol{x}_t)$ does not depend on $\boldsymbol{x}_{(m-1)T}$, and express $w_t(\boldsymbol{x}_t)$ also as $w_t(\boldsymbol{x}_{m,t})$.

2.1. Approach

We shall apply standard multinomial resampling at every stage, as proposed in the seminal paper [14]; this is not necessary from a methodological point of view, but we will analyze this case. In the case of a single particle filter, it is common to adopt $q_t(x_t \mid \boldsymbol{x}_{t-1}) = p_\theta(x_t \mid x_{t-1})$ and $w_t(x_t) = g_\theta(Y_t \mid x_t)$, but it need not be the case, and we can, in general, let

$$w_t(\boldsymbol{x}_t) = \frac{g_\theta(Y_t \mid x_t) p_\theta(x_t \mid x_{t-1})}{q_t(x_t \mid \boldsymbol{x}_{t-1})}. \tag{2.2}$$

Therefore, the single particle filter targets $\prod_{u=1}^{t}[g_\theta(Y_u \mid x_u) p_\theta(x_u \mid x_{u-1})]$, up to proportionality, after resampling at stage t.

For the parallel particle filters, the mth particle filter, after resampling at stage $(m-1)T < t \le mT$, targets

$$\begin{aligned}\pi_{m,t}(\boldsymbol{x}_{m,t}) \propto{}& r_m(x_{(m-1)T+1}) g_\theta(Y_{(m-1)T+1} \mid x_{(m-1)T+1}) \\ &\times \prod_{u=(m-1)T+2}^{t} [g_\theta(Y_u \mid x_u) p_\theta(x_u \mid x_{u-1})],\end{aligned} \tag{2.3}$$

where $r_m(\cdot)$ is a positive probability on X which can be evaluated up to a constant, and is independent from the output of the others filters; sensible choices of $r_m(\cdot)$ are suggested in Section 3.3.2. For $m = 1$, we can simply let $r_m(x_1) = p_\theta(x_1)$. The forms of the target and proposal mean that the particle filters can be run in parallel with each other. Therefore, w_t has the form given in (2.2), with the exceptions that when $t = (m-1)T + 1$, $p_\theta(x_t \mid x_{t-1})$ is replaced by $r_m(x_t)$.

The particle filter approach is given below. It is remarked that some of the notation, for example H_t^k, is not needed to run the particle filter but will help to facilitate the subsequent theoretical analysis.

https://doi.org/10.1017/apr.2015.7 Published online by Cambridge University Press

Particle filter m (PFm). For $1 \le m \le M$. Recursively at stages $t = (m-1)T+1, \dots, mT$ satisfying the following conditions.

(i) *Importance sampling.* For $1 \le k \le K$, sample $\widetilde{X}_t^k \sim q_t(\cdot \mid \boldsymbol{X}_{m,t-1}^k)$ and let $\widetilde{\boldsymbol{X}}_{m,t}^k = \widetilde{X}_t^k$ if $t = (m-1)T+1$, $\widetilde{\boldsymbol{X}}_{m,t}^k = (\boldsymbol{X}_{m,t-1}^k, \widetilde{X}_t^k)$ otherwise.

(ii) *Resampling.* Generate independent and identically distributed $B_t^1, \dots, B_t^K$ [$B(1), \dots, B(K)$ for short] such that

$$\mathbb{P}\{B(1) = j\} = \frac{w_t(\widetilde{\boldsymbol{X}}_{m,t}^j)}{(K\bar{w}_t)(= W_t^j)}, \tag{2.4}$$

where $\bar{w}_t = K^{-1}\sum_{k=1}^K w_t(\widetilde{\boldsymbol{X}}_{m,t}^k)$.

(iii) *Updating.* Let $(\boldsymbol{X}_{m,t}^k, A_{m,t}^k) = (\widetilde{\boldsymbol{X}}_{m,t}^{B(k)}, A_{m,t-1}^{B(k)})$,

$$\widetilde{H}_{m,t}^j = \frac{H_{m,t-1}^j}{(KW_t^j)}, \qquad H_{m,t}^k = \widetilde{H}_{m,t}^{B(k)}, \quad 1 \le k \le K, \tag{2.5}$$

with the conventions $A_{m,(m-1)T}^k = k$, $H_{m,(m-1)T}^k = 1$.

Remark 2.1. There are other procedures that use two or more particle filters to perform estimations. For instance, [19] introduced a method based upon generalized two-filter smoothing. However, that approach is restricted to two particle filters that run forwards and backwards, and requires the choice of pseudo-densities, which may be more difficult than the choice of $r_m(\cdot)$. The approach of [20] used multiple particle filters to perform estimations, but is different from the ideas in this paper. Typically, that approach will run filters in parallel on the same target and allow the filters themselves to interact. In our approach, we are able to reduce the variability (relative to one particle filter) of estimates by segmentation, which was possibly not achieved in [20].

2.2. Notation

Set $\eta_0 = 1$ and assume that

$$\eta_t := \mathbb{E}_q\bigg[\prod_{u=1}^t w_u(\boldsymbol{X}_u)\bigg] < \infty \quad \text{for } 1 \le t \le U, \tag{2.6}$$

where for (integrable) $\varphi\colon \mathsf{X}^t \to \mathbb{R}$,

$$\mathbb{E}_q[\varphi(\boldsymbol{X}_t)] = \int_{\mathsf{X}^t} \varphi(\boldsymbol{x}_t)\bigg[\prod_{u=1}^t q_u(x_u \mid \boldsymbol{x}_{u-1})\bigg]\,\mathrm{d}\boldsymbol{x}_t, \tag{2.7}$$

where $\mathbb{E}_q$ denotes expectation with respect to the importance densities q_t.

Consider $(m-1)T < t \le mT$. Define $\eta_{m,t} = \eta_t/\eta_{(m-1)T}$, and let

$$h_t(\boldsymbol{x}_t) = \frac{\eta_t}{\prod_{u=1}^t w_u(\boldsymbol{x}_u)}, \qquad h_{m,t}(\boldsymbol{x}_{m,t}) = \frac{\eta_{m,t}}{\prod_{u=(m-1)T+1}^t w_u(\boldsymbol{x}_{m,u})}. \tag{2.8}$$

By (2.4)–(2.8),

$$H_{m,t}^k = \bigg(\frac{\bar{w}_{(m-1)T+1}\cdots\bar{w}_t}{\eta_{m,t}}\bigg)h_{m,t}(\boldsymbol{X}_{m,t}^k). \tag{2.9}$$

https://doi.org/10.1017/apr.2015.7 Published online by Cambridge University Press

Let $\mathsf{Z}^m (= \mathsf{Z}^m_K) = \{(k(1), \ldots, k(m)) \colon 1 \le k(n) \le K \text{ for } 1 \le n \le m\}$. For $\boldsymbol{k} \in \mathsf{Z}^m$, let

$$\widetilde{\boldsymbol{X}}^{\boldsymbol{k}}_t = (\boldsymbol{X}^{k(1)}_{1,T}, \ldots, \boldsymbol{X}^{k(m-1)}_{m-1,(m-1)T}, \widetilde{\boldsymbol{X}}^{k(m)}_{m,t}), \qquad \widetilde{H}^{\boldsymbol{k}}_t = \bigg(\prod_{n=1}^{m-1} H^{k(n)}_{n,nT}\bigg)\widetilde{H}^{k(m)}_{m,t}, \tag{2.10}$$

$$\boldsymbol{X}^{\boldsymbol{k}}_t = (\boldsymbol{X}^{k(1)}_{1,T}, \ldots, \boldsymbol{X}^{k(m-1)}_{m-1,(m-1)T}, \boldsymbol{X}^{k(m)}_{m,t}), \qquad H^{\boldsymbol{k}}_t = \bigg(\prod_{n=1}^{m-1} H^{k(n)}_{n,nT}\bigg)H^{k(m)}_{m,t}.$$

Thus, analogous to (2.9),

$$H^{\boldsymbol{k}}_t = \bigg(\frac{\bar{w}_1 \cdots \bar{w}_t}{\eta_t}\bigg) h_t(\boldsymbol{X}^{\boldsymbol{k}}_t).$$

The notation $A^k_{m,t}$ refers to the first-generation ancestor of $\boldsymbol{X}^k_{m,t}$ (or $\widetilde{\boldsymbol{X}}^k_{m,t+1}$). That is, $A^k_t = j$ if the first component of $\boldsymbol{X}^k_{m,t}$ is $\widetilde{X}^j_{(m-1)T+1}$. This ancestor tracing is exploited in Section 3.3.3 for standard error approximations of the estimates. Finally, $N(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2.

3. Estimation theory

We are interested in the estimation of the likelihood $\lambda(\theta)$, and also of $\psi_U := \mathbb{E}_p[\psi(\boldsymbol{X}_U) \mid \boldsymbol{Y}_U]$ for some real-valued measurable function ψ. Here, $\mathbb{E}_p$ denotes expectation under the HMM (1.1). The estimation of $\lambda(\theta)$ falls under the canonical case; the theory is given in Section 3.2. The estimation of ψ_U falls under the noncanonical case; the theory is given in Section 3.3.

3.1. Estimates and remarks

3.1.1. *Canonical case.* Define the function

$$L(\boldsymbol{x}_U) = \frac{p_\theta(\boldsymbol{x}_U \mid \boldsymbol{Y}_U)}{\prod_{t=1}^{U} q_t(x_t \mid \boldsymbol{x}_{t-1})}, \tag{3.1}$$

where $p_\theta(\boldsymbol{x}_U \mid \boldsymbol{Y}_U)$ is as in (2.1). The estimator of ψ_U in the canonical case, which we will prove is unbiased, is given by

$$\widehat{\psi}_U = K^{-M} \sum_{\boldsymbol{k} \in \mathsf{Z}^M} L(\boldsymbol{X}^{\boldsymbol{k}}_U)\psi(\boldsymbol{X}^{\boldsymbol{k}}_U)H^{\boldsymbol{k}}_U. \tag{3.2}$$

By (2.1) and (3.1), $\lambda(\theta)$ appears in the denominator on the right-hand side of (3.2). This does not pose a problem in the estimation of $\lambda(\theta)$, as by setting $\psi \equiv \lambda(\theta)$, we cancel out $\lambda(\theta)$. We define $\widehat{\lambda}(\theta)$ to be the estimator obtained this way, that is

$$\frac{\widehat{\lambda}(\theta)}{\lambda(\theta)} = K^{-M} \sum_{\boldsymbol{k} \in \mathsf{Z}^M} L(\boldsymbol{X}^{\boldsymbol{k}}_U)H^{\boldsymbol{k}}_U. \tag{3.3}$$

To further understand this estimate, we write (3.3) as

$$\widehat{\lambda}(\theta) = \bigg(\prod_{t=1}^{U} \bar{w}_t\bigg) \prod_{m=2}^{M} \bigg(K^{-2} \sum_{k=1}^{K} \sum_{\ell=1}^{K} \frac{p_\theta(X^\ell_{(m-1)T+1} \mid X^k_{(m-1)T})}{r_m(X^\ell_{(m-1)T+1})}\bigg), \tag{3.4}$$

where $X^\ell_{(m-1)T+1}$ here refers to the first component of $X^\ell_{m,mT}$. A heuristic justification is as follows. For expositional purposes, let us consider the simplest case of $M = 2$. The final term on the right-hand side of (3.4), the double summation, is an SMC estimate of the ratio, up to a constant, of the actual target of interest (2.1), and the normalized target (2.3) that is sampled

https://doi.org/10.1017/apr.2015.7 Published online by Cambridge University Press

by the two particle filters. That is as $K \to \infty$, we would like to obtain convergence to

$$\int_{\mathsf{X}^U} \frac{p_\theta(x_{T+1} \mid x_T)}{r_2(x_{T+1})} \pi_{1,T}(\boldsymbol{x}_T)\pi_{2,U}(\boldsymbol{x}_{2,U})\,\mathrm{d}\boldsymbol{x}_U.$$

The term $(\prod_{t=1}^U \bar{w}_t)$ will converge in probability to the normalizing constants of $\pi_{2,U}(\boldsymbol{x}_{2,U})$.

The expression (3.4) also suggests a good choice of $r_2(\cdot)$. If we take $r_2(x) = K^{-1}\sum_{k=1}^K p_\theta(x \mid x_T^k)$, then the double sum on the right-hand side of (3.4) is exactly 1; that is, it does not contribute to the variance of the estimate. The choice above is exactly the SMC approximation of the predictor. However, the choice suggested above is not reasonable in that it circumvents the parallel implementation of the two filters. However, we should thus choose $r_2(\cdot)$ to approximate the predictor. This will be illustrated in Section 3.3.2. We will also discuss in Section 4 how subsampling can be used to reduce the computational cost of calculating $\widehat{\lambda}(\theta)$ to $\mathcal{O}(K)$.

3.1.2. *Noncanonical case.* In the case of latent-state estimation under the noncanonical case, the unknown $\lambda(\theta)$ inherent in (3.2) is replaced by $\widehat{\lambda}(\theta)$, that is, we divide the right-hand side of (3.2) and (3.3) to obtain the estimator

$$\widetilde{\psi}_U = \frac{\sum_{\boldsymbol{k}\in\mathsf{Z}^M} L(\boldsymbol{X}_U^{\boldsymbol{k}})\psi(\boldsymbol{X}_U^{\boldsymbol{k}})H_U^{\boldsymbol{k}}}{\sum_{\boldsymbol{k}\in\mathsf{Z}^M} L(\boldsymbol{X}_U^{\boldsymbol{k}})H_U^{\boldsymbol{k}}}. \tag{3.5}$$

We can write the above estimate in a standard form seen in the literature, and reduce the cost of computation to $\mathcal{O}(K)$. For example, if there is only one particle filter and we select w_t to satisfy (2.2), then the estimate reduces to $K^{-1}\sum_{k=1}^K \psi(\boldsymbol{X}_U^k)$, which is the standard estimate in the literature.

3.2. Unbiasedness and central limit theorem (CLT) under the canonical case

Let $f_0 = \psi_U$ and define

$$f_t(\boldsymbol{x}_t) = \mathbb{E}_q[\psi(\boldsymbol{X}_U)L(\boldsymbol{X}_U) \mid \boldsymbol{X}_t = \boldsymbol{x}_t], \qquad 1 \le t \le U, \tag{3.6}$$

There is no resampling involved under $\mathbb{E}_q$. Let $\#_t^k$ denote the number of copies of $\widetilde{\boldsymbol{X}}_t^k$ generated from $(\widetilde{\boldsymbol{X}}_t^1, \dots, \widetilde{\boldsymbol{X}}_t^K)$ to form $(\boldsymbol{X}_t^1, \dots, \boldsymbol{X}_t^K)$. Thus conditionally, $(\#_t^1, \dots, \#_t^K) \sim \operatorname{multinomial}(K, W_t^1, \dots, W_t^K)$. Let $\mathcal{F}_{2t-1}$ and $\mathcal{F}_{2t}$ denote the σ-algebras generated by all random variables just before and just after resampling, respectively, at the tth stage. In the case of $(m-1)T < t \le mT$ for $m > 1$, these σ-algebras include all random variables generated by PF1 to PF$(m-1)$. Let $\mathbb{E}_K$ denote expectation with respect to K sample paths generated in each particle filter.

Theorem 3.1. *Define, for $(m-1)T < t \le mT$ and $1 \le j \le K$,*

$$\varepsilon_{2t-1}^j = K^{-m+1} \sum_{\boldsymbol{k}\in\mathsf{Z}^m\colon A_{m,t-1}^{k(m)}=j} [f_t(\widetilde{\boldsymbol{X}}_t^{\boldsymbol{k}}) - f_{t-1}(\boldsymbol{X}_{t-1}^{\boldsymbol{k}})]H_{t-1}^{\boldsymbol{k}}, \tag{3.7}$$

$$\varepsilon_{2t}^j = K^{-m+1} \sum_{\boldsymbol{k}\in\mathsf{Z}^m\colon A_{m,t-1}^{k(m)}=j} (\#_t^{k(m)} - KW_t^{k(m)})f_t(\boldsymbol{X}_t^{\boldsymbol{k}})\widetilde{H}_t^{\boldsymbol{k}}.$$

Then for each j and m, $\{\varepsilon_u^j, \mathcal{F}_u, 2(m-1)T < u \le 2mT\}$ is a martingale difference sequence, and

$$K(\widehat{\psi}_U - \psi_U) = \sum_{m=1}^M \sum_{j=1}^K (\varepsilon_{2(m-1)T+1}^j + \cdots + \varepsilon_{2mT}^j). \tag{3.8}$$

Therefore, $\mathbb{E}_K\widehat{\psi}_U = \psi_U$.

https://doi.org/10.1017/apr.2015.7 Published online by Cambridge University Press

Proof. Since $\#_t^k \sim \text{binomial}(K, W_t^k)$ when conditioned on $\mathcal{F}_{2t-1}$, by the tower law of conditional expectations

$$\mathbb{E}_K(\varepsilon_{2t-1}^j \mid \mathcal{F}_{2t-2}) = K^{-m+1} \sum_{\boldsymbol{k}\in Z^m:\ A_{t-1}^{k(m)}=j} \{\mathbb{E}_K[f_t(\widetilde{\boldsymbol{X}}_t^{\boldsymbol{k}} \mid \mathcal{F}_{2t-2}] - f_{t-1}(\boldsymbol{X}_{t-1}^{\boldsymbol{k}})\} H_{t-1}^{\boldsymbol{k}} = 0,$$

$$\mathbb{E}_K(\varepsilon_{2t}^j \mid \mathcal{F}_{2t-1}) = K^{-m+1} \sum_{\boldsymbol{k}\in Z^m:\ A_{t-1}^{k(m)}=j} [\mathbb{E}_K(\#_t^{k(m)} \mid \mathcal{F}_{2t-1}) - K W_t^{k(m)}] f_t(\widetilde{\boldsymbol{X}}_t^{\boldsymbol{k}}) \widetilde{H}_t^{\boldsymbol{k}} = 0;$$

therefore, $\{\varepsilon_u^j, \mathcal{F}_u, 2(m-1)T < u \le 2mT\}$ are indeed martingale difference sequences.

It follows from (2.5) and (2.10) that

$$\begin{aligned}\sum_{\boldsymbol{k}\in Z^m:\ A_{t-1}^{k(m)}=j} &(\#_t^{k(m)} - K W_t^{k(m)}) f_t(\widetilde{\boldsymbol{X}}_t^{\boldsymbol{k}}) \widetilde{H}_t^{\boldsymbol{k}} \\ &= \sum_{\boldsymbol{k}\in Z^m:\ A_t^{k(m)}=j} f_t(\boldsymbol{X}_t^{\boldsymbol{k}}) H_t^{\boldsymbol{k}} - \sum_{\boldsymbol{k}\in Z^m:\ A_{t-1}^{k(m)}=j} f_t(\widetilde{\boldsymbol{X}}_t^{\boldsymbol{k}}) H_{t-1}^{\boldsymbol{k}};\end{aligned}$$

therefore, by (3.7) and the cancellation of terms in a telescoping series, we obtain

$$\begin{aligned}\sum_{u=2(m-1)T+1}^{2mT} \varepsilon_u^j = K^{-m+1} &\sum_{\boldsymbol{k}\in Z^m:\ A_{m,mT}^{k(m)}=j} f_{mT}(\boldsymbol{X}_{mT}^{\boldsymbol{k}}) H_{mT}^{\boldsymbol{k}} \\ &- K^{-m+1} \sum_{\boldsymbol{k}\in Z^{m-1}} f_{(m-1)T}(\boldsymbol{X}_{(m-1)T}^{\boldsymbol{k}}) H_{(m-1)T}^{\boldsymbol{k}}.\end{aligned}$$

Therefore,

$$\begin{aligned}\sum_{j=1}^{K}\Bigg(\sum_{u=2(m-1)T+1}^{2mT} \varepsilon_u^j\Bigg) = K^{-m+1} &\sum_{\boldsymbol{k}\in Z^m} f_{mT}(\boldsymbol{X}_{mT}^{\boldsymbol{k}}) H_{mT}^{\boldsymbol{k}} \\ &- K^{-m+2} \sum_{\boldsymbol{k}\in Z^{m-1}} f_{(m-1)T}(\boldsymbol{X}_{(m-1)T}^{\boldsymbol{k}}) H_{(m-1)T}^{\boldsymbol{k}}.\end{aligned} \tag{3.9}$$

By (3.2), (3.8) follows from adding (3.9) over $1 \le m \le M$, keeping in mind that $f_0 = \psi_U$ and $f_U(\boldsymbol{x}_U) = L(\boldsymbol{x}_U)\psi(\boldsymbol{X}_U)$. □

The martingale difference expansion (3.8) is for the purpose of standard error estimation; see Section 3.3.3. An alternative expansion, for the purpose of CLT theory in the spirit of [10, Chapter 8], is formed from the martingale difference sequence $\{(Z_u^1, \dots, Z_u^K)\colon 2(m-1)T < u \le 2mT\}$, where

$$Z_{2t-1}^k = K^{-m+1} \sum_{\boldsymbol{k}\in Z^m:\ k(m)=k} [f_t(\widetilde{\boldsymbol{X}}_t^{\boldsymbol{k}}) - f_{t-1}(\boldsymbol{X}_{t-1}^{\boldsymbol{k}})] H_{t-1}^{\boldsymbol{k}}, \tag{3.10}$$

$$Z_{2t}^k = K^{-m+1} \sum_{\boldsymbol{k}\in Z^m::\ k(m)=k} f_t(\boldsymbol{X}_t^{\boldsymbol{k}}) H_t^{\boldsymbol{k}} - K^{-m+1} \sum_{\boldsymbol{k}\in Z^m} W_t^k f_t(\widetilde{\boldsymbol{X}}_t^{\boldsymbol{k}}) \widetilde{H}_t^{\boldsymbol{k}}.$$

Analogous to (3.8), we have

$$K(\widehat{\psi}_U - \psi_U) = \sum_{u=1}^{2U} (Z_u^1 + \dots + Z_u^K), \tag{3.11}$$

from which we can also conclude that $\widehat{\psi}_U$ is unbiased.

https://doi.org/10.1017/apr.2015.7 Published online by Cambridge University Press

The technical difficulties in working with (3.10) to prove the CLT are considerably more involved compared to the standard single particle filter, as there is now a sum over a multi-dimensional space. Therefore, let us provide some intuition first, focusing on the key arguments in the extension of the CLT to $M = 2$ segments.

For $t > T$, let $f_{2,t}(\boldsymbol{x}_{2,t}) = \mathbb{E}_q[f_t(\boldsymbol{X}_t) \mid \boldsymbol{X}_{2,t} = \boldsymbol{x}_{2,t}]$. By a 'law of large numbers' argument, see Lemma A.2 in Appendix A.1, we have

$$K^{-1} \sum_{k=1}^{K} f_t(\widetilde{\boldsymbol{X}}_t^{k\ell}) H_T^k \doteq f_{2,t}(\widetilde{\boldsymbol{X}}_{2,t}^{\ell}), \qquad K^{-1} \sum_{k=1}^{K} f_t(\boldsymbol{X}_t^{k\ell}) H_T^k \doteq f_{2,t}(\boldsymbol{X}_{2,t}^{\ell}).$$

Therefore, by (2.10) and (3.10), $Z_u^\ell \doteq Z_{2,u}^\ell$ for $2T < u \le 2U$, where

$$\begin{aligned} Z_{2,2t-1}^{\ell} &= [f_{2,t}(\widetilde{\boldsymbol{X}}_{2,t}^{\ell}) - f_{2,t-1}(\boldsymbol{X}_{2,t-1}^{\ell})] H_{2,t-1}^{\ell}, \\ Z_{2,2t}^{\ell} &= f_{2,t}(\boldsymbol{X}_{2,t}^{\ell}) H_{2,t}^{\ell} - \sum_{j=1}^{K} W_t^j f_{2,t}(\widetilde{\boldsymbol{X}}_{2,t}^{j}) \widetilde{H}_{2,t}^{j}. \end{aligned} \tag{3.12}$$

We now have a martingale difference sequence $\{(Z_u^1, \ldots, Z_u^K), \mathcal{F}_v, 1 \le u \le 2T\}$ that depends on the outcomes of PF1 only, and another sequence $\{(Z_{2,u}^1, \ldots, Z_{2,u}^K), \mathcal{G}_u, 2T < u \le 2U\}$ that depends on the outcomes of PF2 only, where $\mathcal{G}_{2t-1}$ and $\mathcal{G}_{2t}$ denote the σ-algebras generated by random variables in PF2 only, just before and just after resampling, respectively, at stage t. Moreover,

$$K(\widehat{\psi}_U - \psi_U) \doteq \sum_{u=1}^{2T} \biggl(\sum_{k=1}^{K} Z_u^k \biggr) + \sum_{u=2T+1}^{2U} \biggl(\sum_{\ell=1}^{K} Z_{2,u}^{\ell} \biggr). \tag{3.13}$$

Therefore, subject to negligible error in (3.13), $\sqrt{K}(\widehat{\psi}_U - \psi_U)$ is asymptotically normal, with variance the sum of the variance components due to each particle filter.

More generally in the case of M independent particle filters, define

$$f_{m,t}(\boldsymbol{x}_{m,t}) = \mathbb{E}_q[f_t(\boldsymbol{X}_t) \mid \boldsymbol{X}_{m,t} = \boldsymbol{x}_{m,t}], \tag{3.14}$$

and recall the definition of $h_{m,t}$ in (2.8).

Theorem 3.2. *Let $\sigma^2 = \sum_{u=1}^{2U} \sigma_u^2$, where for $(m-1)T < t \le mT$,*

$$\begin{aligned} \sigma_{2t-1}^2 &= \mathbb{E}_q\{[f_{m,t}^2(\boldsymbol{X}_{m,t}) - f_{m,t-1}^2(\boldsymbol{X}_{m,t-1})] h_{m,t-1}(\boldsymbol{X}_{m,t-1})\}, \\ \sigma_{2t}^2 &= \mathbb{E}_q\biggl\{ \frac{[f_{m,t}(\boldsymbol{X}_{m,t}) h_{m,t}(\boldsymbol{X}_{m,t}) - f_0]^2}{h_{m,t}(\boldsymbol{X}_{m,t})} \biggr\}. \end{aligned}$$

Assume that $\mathbb{E}_q\{f_t^2(\boldsymbol{X}_t)[h_t(\boldsymbol{X}_t) + h_{t-1}(\boldsymbol{X}_t)]\} < \infty$ for $1 \le t \le U$. Then $\sigma^2 < \infty$ and

$$\sqrt{K}(\widehat{\psi}_U - \psi_U) \implies N(0, \sigma^2) \quad \text{as } K \to \infty.$$

3.3. Asymptotic theory in the noncanonical case

In the noncanonical case, the estimator $\widetilde{\psi}_U$, see (3.5), can be approximated by $\widehat{\psi}_U^{\mathrm{c}}$, an unbiased estimator under the canonical case of the centered function $\psi^{\mathrm{c}}(\boldsymbol{x}_U) := \psi(\boldsymbol{x}_U) - \psi_U$. Therefore, analogous to (3.6) and (3.14), we define $f_{m,(m-1)T}^{\mathrm{c}}(\boldsymbol{x}_t) = 0$ and

$$f_{m,t}^{\mathrm{c}}(\boldsymbol{x}_t) = \mathbb{E}_q[\psi^{\mathrm{c}}(\boldsymbol{X}_U) L(\boldsymbol{X}_U) \mid \boldsymbol{X}_{m,t} = \boldsymbol{x}_{m,t}], \qquad (m-1)T < t \le mT. \tag{3.15}$$

The corollary below then follows from Theorem 3.2.

https://doi.org/10.1017/apr.2015.7 Published online by Cambridge University Press

Corollary 3.1. *Let $\sigma_c^2 = \sum_{u=1}^{2U} \sigma_{c,u}^2$, where for $(m-1)T < t \leq mT$,*

$$\sigma_{c,2t-1}^2 = \mathbb{E}_q(\{[f_{m,t}^{\mathrm{c}}(\boldsymbol{X}_{m,t})]^2 - [f_{m,t-1}^{\mathrm{c}}(\boldsymbol{X}_{m,t-1})]^2\}h_{m,t-1}(\boldsymbol{X}_{m,t-1})), \tag{3.16}$$
$$\sigma_{c,2t}^2 = \mathbb{E}_q\{[f_{m,t}^{\mathrm{c}}(\boldsymbol{X}_{m,t})]^2 h_{m,t}(\boldsymbol{X}_{m,t})\}.$$

Under the assumptions of Theorem 3.2,

$$\sqrt{K}(\widetilde{\psi}_U - \psi_U) \quad \Longrightarrow \quad N(0, \sigma_c^2) \quad \textit{as } K \to \infty. \tag{3.17}$$

In Section 3.3.3 we show how σ_c^2 can be estimated in-sample, and discuss the implications in particle size allocation. Before that, we shall illustrate, in Sections 3.3.1 and 3.3.2, the advantage of segmentation in providing stability to smoothed latent-state estimations, that is, the estimation of $\mathbb{E}_p(X_u \mid \boldsymbol{Y}_U)$ for $u < U$, with u fixed as $U \to \infty$.

3.3.1. *Example.* Consider the linear time-series

$$X_t = aX_{t-1} + \varepsilon_t, \qquad Y_t = X_t + \eta_t, \tag{3.18}$$

with $0 < a < 1$, $\varepsilon_t \sim N(0, (1-a^2)\sigma_X^2)$ and $\eta_t \sim N(0, \sigma_Y^2)$. Let $\theta = (a, \sigma_X^2, \sigma_Y^2)$.

We shall illustrate on this simple example the advantage of parallel particle filters in smoothed estimation of X_t. Consider firstly the segmented particle filter with $T = 1$ and $q_t(x_t \mid \boldsymbol{x}_{t-1}) = p_\theta(x_t \mid \boldsymbol{Y}_U)$. Let $w_t \equiv 1$; therefore, $h_{t,t} \equiv 1$ (recall also that $h_{t,t-1} \equiv 1$) for all t. Consider $\psi(\boldsymbol{x}_U) = x_u$ for some $1 \leq u < U$. By (3.1) and (3.15), $f_{t,t}^{\mathrm{c}}(\boldsymbol{x}_t) = \mathbb{E}_p(X_u \mid X_t = x_t, \boldsymbol{Y}_U) - \mathbb{E}_p(X_u \mid \boldsymbol{Y}_U)$. Since $f_{t,t-1}^{\mathrm{c}} \equiv 0$; therefore, by (3.16),

$$\sigma_{c,2t}^2 = \sigma_{c,2t-1}^2 = \operatorname{var}_p(\mathbb{E}_p(X_u \mid X_t, \boldsymbol{Y}_U) \mid \boldsymbol{Y}_U) \leq \operatorname{var}_p(\mathbb{E}_p(X_u \mid X_t)) = a^{|u-t|}\sigma_X^2, \tag{3.19}$$

with the inequality in (3.19) following from, say, [1, Equation (2.5.3)]. Therefore,

$$\sigma_c^2 = \sum_{t=1}^{U}(\sigma_{c,2t-1}^2 + \sigma_{c,2t}^2) \leq \left(\frac{4}{1-a} - 2\right)\sigma_X^2.$$

Corollary 3.1 says that, in this example, σ_c^2 is uniformly bounded and, hence, the estimation of X_u is stable as $U \to \infty$.

For the standard particle filter with no latent-state sequence segmentation, that is in the case $M = 1$ and $T = U$, we consider $q_t(x_t \mid \boldsymbol{x}_{t-1}) = p_\theta(x_t \mid \boldsymbol{x}_{t-1}, \boldsymbol{Y}_U)$ and $w_t \equiv 1$. By (3.15), if $t \geq u$ then $f_{2t}^{\mathrm{c}}(\boldsymbol{x}_t)[= f_{1,2t}^{\mathrm{c}}(\boldsymbol{x}_t)] = \mathbb{E}_p(X_u \mid \boldsymbol{X}_t = \boldsymbol{x}_t, \boldsymbol{Y}_U) - \mathbb{E}_p(X_u \mid \boldsymbol{Y}_U) = X_u - \mathbb{E}_p(X_u \mid \boldsymbol{Y}_U)$. Since $\operatorname{var}_p(X_u \mid \boldsymbol{Y}_U) \geq \operatorname{var}_p(X_u \mid X_{u-1}, X_{u+1}, Y_u) > 0$; therefore, in this example, $\sigma_{c,2t}^2$ is bounded away from 0 for $t \geq u$, and, consequently, $\sigma_c^2 \to \infty$ as $U \to \infty$.

Intuitively, in the case of the standard particle filter, the estimation of $\mathbb{E}_p(X_u \mid \boldsymbol{Y}_U)$ is unstable as $U \to \infty$ simply due to the degeneracy caused by repeated resamplings at $t \geq u$. On the other hand, the repeated resamplings does not cause instability in the segmented method because resampling in one segment does not result in sample depletion of another segment. There is a vast literature on smoothed latent-state estimators; see, for example, [5]. We do not go into detail here as our main motivation for looking at parallel particle filters is to achieve wall-clock computation time savings; the variance reductions in smoothed estimates can be viewed as an added benefit.

https://doi.org/10.1017/apr.2015.7 Published online by Cambridge University Press

TABLE 1: Mean-square error ($\times 10^{-2}$) of $\widehat{\mathbb{E}}_p(X_u \mid \boldsymbol{Y}_{50})$ for the standard particle filter, the segmented particle filter (Seg.) initialized at $N(0,1)$, and the segmented particle filter initialized at $N(\widehat{\mu}_m, \widehat{\sigma}_m^2)$, with $(\widehat{\mu}_m, \widehat{\sigma}_m^2)$ estimated from past observations. In (a) resampling is performed at every stage, in (b) at every two stages.

		Mean-square error ($\times 10^{-2}$)					
		Standard		Seg. $N(0,1)$		Seg. $N(\widehat{\mu}_m, \widehat{\sigma}_m^2)$	
u	$\mathbb{E}_p(X_u \mid \boldsymbol{Y}_{50})$	(a)	(b)	(a)	(b)	(a)	(b)
10	0.45	6.0	5.7	0.4	0.3	0.3	0.3
20	0.95	4.9	5.6	1.1	0.9	0.8	1.0
30	2.81	2.7	1.7	2.2	2.0	1.0	1.3
40	−1.10	1.4	0.9	1.5	1.1	0.9	0.3
50	0.38	0.2	0.2	0.2	0.3	0.2	0.2

3.3.2. *Numerical study.* As in Section 3.3.1, consider the linear time-series (3.18) and the estimation of $\psi(\boldsymbol{x}_U) = x_u$ for $0 < u \le U$, conditioned on $\boldsymbol{Y}_U$. Kalman updating formulas are applied in order to compute $\mathbb{E}_p(X_u \mid \boldsymbol{Y}_{u-1})$ and $\mathbb{E}_p(X_u \mid \boldsymbol{Y}_u)$ analytically, and the Rauch–Tung–Striebel smoother is applied to compute $\mathbb{E}_p(X_u \mid \boldsymbol{Y}_t)$ for $t > u$. The mean-squared errors (MSE) of the particle filter methods can then be computed using Monte Carlo methods.

The first method we consider is the standard particle filter that performs bootstrap resampling. We select q_1 as $N(0, \sigma_X^2)$ and $q_t(\cdot \mid \boldsymbol{x}_{t-1})$ as $N(ax_{t-1}, (1-a^2)\sigma_X^2)$ for $t > 1$; hence, if resampling is performed at every stage, then

$$w_t(\boldsymbol{x}_t) = \exp\left[\frac{-(Y_t - x_t)^2}{(2\sigma_Y^2)}\right], \qquad t \ge 1. \tag{3.20}$$

We next consider parallel particle filters with $\boldsymbol{X}_U$ segmented into M portions of equal length $T = U/M$. The resampling weights are as in (3.20), and like the standard particle filter, $q_{m,t}(\cdot \mid \boldsymbol{x}_{t-1}) = N(ax_{t-1}, (1-a^2)\sigma_X^2)$ for $(m-1)T + 1 < t \le mT$.

We consider two versions of $q_{m,(m-1)T+1}$ ($= r_m$ here), the initial sampling distribution of $\widetilde{X}^k_{m,(m-1)T+1}$. In the first version, we simply let $q_{m,(m-1)T+1} = N(0, \sigma_X^2)$. In the second version, we let $q_{m,(m-1)T+1} = N(\widehat{\mu}_m, \widehat{\sigma}_m^2)$, with $(\widehat{\mu}_1, \widehat{\sigma}_1^2) = (0, \sigma_X^2)$ and for $m \ge 2$, $(\widehat{\mu}_m, \widehat{\sigma}_m^2)$ simulated using $(Y_{(m-1)T-r}, \dots, Y_{(m-1)T})$ for some $r \ge 0$. More specifically, we let

$$\mu_m = \mathbb{E}_p(X_{(m-1)T+1} \mid Y_{(m-1)T-r}, \dots, Y_{(m-1)T}),$$
$$\sigma_m^2 = \operatorname{var}_p(X_{(m-1)T+1} \mid Y_{(m-1)T-r}, \dots, Y_{(m-1)T}),$$

and estimate them by sampling $(X_{(m-1)T-r}, \dots, X_{(m-1)T+1})$ using particle filters that resample at every stage.

In our simulation study, we select $a = 0.8$, $\sigma_X^2 = \sigma_Y^2 = 1$, $U = 50$, and $M = 5$. We apply $K = 500$ particles in each filter, and for the second version of the segmented method, we consider $r = 4$. For each method, there are 100 repetitions for resampling at every stage, and another 100 repetitions for resampling at every two stages denoted by (a) and (b) in Table 1, respectively. We see from Table 1 substantial MSE reductions for segmented particle filters over standard particle filters, especially when $U - u$ is big, agreeing with the discussions in Section 3.3.1. In addition, we see that applying estimation of (μ_m, σ_m^2) improves upon the performance of segmented particle filters.

3.3.3. *In-sample variance estimation and particle size allocation.* Let

$$\sigma^2_{\mathrm{PF}m} = \sum_{u=2(m-1)T+1}^{2mT} \sigma^2_{c,u}$$

be the variability attributed to the mth particle filter. If K_m particles are allocated to particle filter m with K_m large, then analogous to (3.17), we have

$$\mathrm{var}(\widetilde{\psi}_U) \doteq \sum_{m=1}^{M} \frac{\sigma^2_{\mathrm{PF}m}}{K_m}. \tag{3.21}$$

Therefore, being able to estimate $\sigma^2_{\mathrm{PF}m}$ in-sample would allow us to optimally allocate the particle sizes in the M particle filters so as to minimize (3.21). The estimation can be performed in the following manner.

Consider $1 \leq m \leq M$ and let $C^j_m = \{k \colon A^k_{m,mT} = j\}$, noting that $A^k_{m,mT} = j$ if and only if $\boldsymbol{X}^k_{m,mT}$ is descended from $\widetilde{X}^j_{(m-1)T+1}$. Let

$$Q_j(\mu) = \sum_{\boldsymbol{k} \in \mathsf{Z}^M \colon k(m) \in C^j_m} L(\boldsymbol{X}^{\boldsymbol{k}}_U)[\psi(\boldsymbol{X}^{\boldsymbol{k}}_U) - \mu] H^{\boldsymbol{k}}_U.$$

Theorem 3.3. *Under the assumptions of Corollary 3.1,*

$$\widehat{\sigma}^2_{\mathrm{PF}m}(\psi_U) := K^{-1} \sum_{j=1}^{K} Q^2_j(\psi_U) \xrightarrow{\mathbb{P}} \sigma^2_{\mathrm{PF}m}, \qquad 1 \leq m \leq M. \tag{3.22}$$

Since $\widetilde{\psi}_U \xrightarrow{\mathbb{P}} \psi_U$ by Corollary 3.1, it follows that $\widehat{\sigma}^2_{\mathrm{PF}m}(\widetilde{\psi}_U)$ is consistent for $\sigma^2_{\mathrm{PF}m}$. Besides particle size allocation, being able to estimate $\sigma^2_{\mathrm{PF}m}$, $1 \leq m \leq M$, and, hence, σ^2_c allows us to assess the level of accuracy of $\widetilde{\psi}_U$ in estimating ψ_U.

Continuing upon the numerical study in Section 3.3.2, assume, hypothetically, that we are most interested in estimating $\mathbb{E}_p(X_{30} \mid \boldsymbol{Y}_{50})$. For a fixed $K_{\mathrm{tot}} = \sum_{m=1}^{M} K_m$, the right-hand side of (3.21) is minimized by setting $K_m \propto \sigma_{\mathrm{PF}m}$. Based upon a preliminary run with 500 particles in each segment, approximations for $\sigma^2_{\mathrm{PF}m}$ were computed and it was determined that the optimal allocation under the constraint $K_{\mathrm{tot}} = 25\,000$ is $(K_1, \ldots, K_5) = (3500, 2900, 12\,600, 5900, 100)$. In addition, based on 20 repetitions each of optimal and equal allocations ($K_m = 5000$ for each m), the conclusion is that the MSE for optimal allocation (0.6×10^{-3}) is significantly smaller compared to equal allocation (1.8×10^{-3}).

4. Discussion

We discuss here the subsampling approach, proposed in [4] and [13], that can be used to reduce the $\mathcal{O}(K^2)$ computational cost of our estimates. We make the discussions more concrete here by considering $M = 2$ and focusing on the estimation of the likelihood $\lambda(\theta)$.

The actual computational cost of the double sum in (3.4) may be less expensive than it seems, given that this operation is performed only once, and that time-savings can be achieved if we bother to first group the segments $\boldsymbol{X}^{\ell}_{2,U}$ having a common first component. However, asymptotically, we do have a larger computational complexity due to the double sum.

https://doi.org/10.1017/apr.2015.7 Published online by Cambridge University Press

Let $\{(k(v), \ell(v)) \colon 1 \le v \le V\}$ be selected i.i.d. from β, a positive distribution on Z_K^2, and estimate $\lambda(\theta)$ by

$$\widehat{\lambda}^*(\theta) = \Big(\prod_{t=1}^{U} \bar{w}_t\Big)(K^2V)^{-1} \sum_{v=1}^{V} \frac{p_\theta(X_{T+1}^{\ell(v)} \mid X_T^{k(v)})}{r_2(X_{T+1}^{\ell(v)})\beta(k(v), \ell(v))}.$$

Since $\widehat{\lambda}(\theta)$ is unbiased for $\lambda(\theta)$, then so is $\widehat{\lambda}^*(\theta)$. For example, we can apply stratification sampling so that 'good' pairs are chosen more frequently. The choice of $V \sim K^s$ for $s = 1$ would give us a $\mathcal{O}(K)$ algorithm, though we may have to select $s > 1$ in order to maintain the asymptotic variance of $\widehat{\lambda}(\theta)$. As the computation of $\widehat{\lambda}^*(\theta)$ is separate from the execution of the parallel particle filters and can be done offline, improving $\widehat{\lambda}^*(\theta)$ with more sampling does not require rerunning of the particle filters.

Appendix A. Proofs

We preface the proofs of the main results of Section 3 with two supporting lemmas in Appendix A.1 below. Lemma A.1 is a weak law of large number for sums of segmented sequences. Lemma A.2 provides finer approximations of such sums.

A.1. Asymptotics and finer approximations for sums of segmented sequences

Lemma A.1. *Let G be a real-valued measurable function of $\boldsymbol{x}_t$ for some $(m-1)T < t \le mT$ with $1 \le m \le M$.*

(i) *If $\widetilde{\mu}_t := \mathbb{E}_q[G(\boldsymbol{X}_t)/h_{t-1}(\boldsymbol{X}_{t-1})]$ exists and is finite, then as $K \to \infty$,*

$$K^{-m} \sum_{\boldsymbol{k} \in \mathsf{Z}^m} G(\widetilde{\boldsymbol{X}}_t^{\boldsymbol{k}}) \xrightarrow{\mathbb{P}} \widetilde{\mu}_t.$$

(ii) *If $\mu_t := \mathbb{E}_q[G(\boldsymbol{X}_t)/h_t(\boldsymbol{X}_t)]$ exists and is finite, then as $K \to \infty$,*

$$K^{-m} \sum_{\boldsymbol{k} \in \mathsf{Z}^m} G(\boldsymbol{X}_t^{\boldsymbol{k}}) \xrightarrow{\mathbb{P}} \mu_t.$$

(iii) *For each $\boldsymbol{k} \in \mathsf{Z}^m$,*

$$\frac{\widetilde{H}_t^{\boldsymbol{k}}}{h_t(\widetilde{\boldsymbol{X}}_t^{\boldsymbol{k}})} = \frac{H_t^{\boldsymbol{k}}}{h_t(\boldsymbol{X}_t^{\boldsymbol{k}})} = \frac{\bar{w}_1 \cdots \bar{w}_t}{\eta_t} \xrightarrow{\mathbb{P}} 1. \tag{A.1}$$

Proof. Since $G = G^+ - G^-$, we can assume without loss of generality that G is nonnegative. The proofs of Lemma A.1(i) and A.1(ii) for $t \le T$ follow from standard induction arguments; see [6, Lemma 2]. For $t > T$, induction arguments are again used, but the framework is now considerably more complicated with summation on a multi-dimensional instead of a one-dimensional space. Unlike in [6], characteristic functions are now needed in the induction proof.

Let $T < u \le 2T$ and assume that $\widetilde{\mu}_u$ exists and is finite, and that Lemma A.1(ii) holds for $t = u - 1$. Let $V_u^\ell = K^{-1} \sum_{k=1}^K G(\widetilde{\boldsymbol{X}}_u^{k\ell})$ and consider the decomposition $V_u^\ell = R_{u,c}^\ell + S_{u,c}^\ell$, where

$$R_{u,c}^\ell = K^{-1} \sum_{k=1}^K G(\widetilde{\boldsymbol{X}}_u^{k\ell}) \mathbf{1}_{\{G(\tilde{\boldsymbol{X}}_u^{k\ell}) \le c\}}, \qquad S_{u,c}^\ell = K^{-1} \sum_{k=1}^K G(\widetilde{\boldsymbol{X}}_u^{k\ell}) \mathbf{1}_{\{G(\tilde{\boldsymbol{X}}_u^{k\ell}) > c\}}.$$

https://doi.org/10.1017/apr.2015.7 Published online by Cambridge University Press

Let $\bar{V}_u = K^{-1}\sum_{\ell=1}^K V_u^\ell$, and define $\bar{R}_{u,c}$, $\bar{S}_{u,c}$ in a similar fashion. Let $\mathrm{i} = \sqrt{-1}$ and define

$$\widetilde{\mu}_{u,c} = \mathbb{E}_q\bigg[\frac{G(\boldsymbol{X}_u)\mathbf{1}_{\{G(\boldsymbol{X}_u)>c\}}}{h_{u-1}(\boldsymbol{X}_{u-1})}\bigg], \tag{A.2}$$

$$\varphi^\ell_{2u-1,c}(\theta \mid \mathcal{F}_{2u-2}) = \mathbb{E}_K[\exp(\mathrm{i}\theta K^{-1}R^\ell_{u,c}) \mid \mathcal{F}_{2u-2}].$$

Let $\delta > 0$. Since $R^1_{u,\delta K}, \dots, R^K_{u,\delta K}$ are independent conditioned on $\mathcal{F}_{2u-2}$, it follows that

$$\mathbb{E}_K[\exp(\mathrm{i}\theta \bar{V}_u) \mid \mathcal{F}_{2u-2}] = \prod_{\ell=1}^K \varphi^\ell_{2u-1,\delta K}(\theta \mid \mathcal{F}_{2u-2}) + r_K, \tag{A.3}$$

$$r_K = \mathbb{E}_K\{\exp(\mathrm{i}\theta\bar{R}_{u,\delta K})[\exp(\mathrm{i}\theta\bar{S}_{u,\delta K}) - 1] \mid \mathcal{F}_{2u-2}\}.$$

Since $|\mathrm{e}^{\mathrm{i}z} - 1| \le |z|$, for $K \ge c/\delta$,

$$|r_K| \le |\theta| K^{-1}\sum_{\ell=1}^K \mathbb{E}_K(S^\ell_{u,c} \mid \mathcal{F}_{2u-2}). \tag{A.4}$$

By the induction hypothesis applied on $G_c(\boldsymbol{x}_{u-1}) := \mathbb{E}_q[G(\boldsymbol{X}_u)\mathbf{1}_{\{G(\boldsymbol{X}_u)>c\}} \mid \boldsymbol{X}_{u-1} = \boldsymbol{x}_{u-1}]$, we obtain

$$K^{-1}\sum_{\ell=1}^K \mathbb{E}_K(S^\ell_{u,c} \mid \mathcal{F}_{2u-2}) \xrightarrow{\mathbb{P}} \mathbb{E}_q\bigg[\frac{G_c(\boldsymbol{X}_{u-1})}{h_{u-1}(\boldsymbol{X}_{u-1})}\bigg](= \widetilde{\mu}_{u,c}). \tag{A.5}$$

Since $\widetilde{\mu}_{u,c} \to 0$ as $c \to \infty$, therefore by (A.4) and (A.5),

$$r_K \xrightarrow{\mathbb{P}} 0. \tag{A.6}$$

Let $\widetilde{R}_{u,c} = R^\ell_{u,c} - \mathbb{E}_K(R^\ell_{u,c} \mid \mathcal{F}_{2u-2})$ and

$$\widetilde{\varphi}^\ell_{2u-1,c}(\theta \mid \mathcal{F}_{2u-2}) = \mathbb{E}_K[\exp(\mathrm{i}\theta K^{-1}\widetilde{R}^\ell_{u,c}) \mid \mathcal{F}_{2u-2}]. \tag{A.7}$$

Since $|\mathbb{E}[\mathrm{e}^{\mathrm{i}\theta Z} - (1 + \mathrm{i}\theta Z - \theta^2 Z^2/2)]| \le \theta^2\mathbb{E}Z^2$ (see [3, Equation (26.5)]) and $(R^\ell_{u,\delta K})^2 \le \delta K R^\ell_{u,\delta K}$,

$$\begin{aligned}
&\bigg|\widetilde{\varphi}^\ell_{2u-1,\delta K}(\theta \mid \mathcal{F}_{2u-2}) - \bigg\{1 - \bigg[\frac{\theta^2}{(2K^2)}\bigg]\operatorname{var}_K(R^\ell_{u,\delta K} \mid \mathcal{F}_{2u-2})\bigg\}\bigg| \\
&\quad\le \bigg(\frac{\theta}{K}\bigg)^2 \operatorname{var}_K(R^\ell_{u,\delta K} \mid \mathcal{F}_{2u-2}) \\
&\quad\le \bigg(\frac{\theta}{K}\bigg)^2 \mathbb{E}_K[(R^\ell_{u,\delta K})^2 \mid \mathcal{F}_{2u-2}] \\
&\quad\le \bigg(\frac{\delta\theta^2}{K}\bigg)\mathbb{E}_K(R^\ell_{u,\delta K} \mid \mathcal{F}_{2u-2}).
\end{aligned} \tag{A.8}$$

Since $|\prod_{\ell=1}^K z_\ell - \prod_{\ell=1}^K y_\ell| \le \sum_{\ell=1}^K |z_\ell - y_\ell|$ whenever $|z_\ell| \le 1$ and $|y_\ell| \le 1$ for all ℓ (see [3, Lemma 1, Section 2.7]), by the induction hypothesis applied on

$$G'_c(\boldsymbol{x}_u) = \mathbb{E}_q[G(\boldsymbol{X}_u)\mathbf{1}_{\{G(\boldsymbol{X}_u)\le c\}} \mid \boldsymbol{X}_{u-1} = \boldsymbol{x}_{u-1}]$$

https://doi.org/10.1017/apr.2015.7 Published online by Cambridge University Press

and then letting $c \to \infty$, we have

$$\begin{aligned}
&\Big|\prod_{\ell=1}^{K} \widetilde{\varphi}^{\ell}_{2u-1,\delta K}(\theta \mid \mathcal{F}_{2u-2}) - \prod_{\ell=1}^{K}\Big\{1 - \Big[\frac{\theta^2}{(2K^2)}\Big]\mathrm{var}_K(R^{\ell}_{u,\delta K} \mid \mathcal{F}_{2u-2})\Big\}\Big| \\
&\qquad \le \delta\theta^2 K^{-1}\sum_{\ell=1}^{K}\mathbb{E}_K(R^{\ell}_{u,\delta K} \mid \mathcal{F}_{2u-2}) \\
&\qquad \xrightarrow{\mathbb{P}} \delta\theta^2\widetilde{\mu}_u.
\end{aligned} \tag{A.9}$$

Let $\delta_0 > 0$ be such that $\log(1-y) \ge -2y$ for $0 < y < (\theta\delta_0)^2$. Therefore, by the inequalities in (A.8), for $0 < \delta \le \delta_0$,

$$\begin{aligned}
\prod_{\ell=1}^{K}\Big\{1 - \Big[\frac{\theta^2}{(2K^2)}\Big]\mathrm{var}_K(R^{\ell}_{u,\delta K} \mid \mathcal{F}_{2u-2})\Big\} &\ge \prod_{\ell=1}^{K}\Big\{1 - \Big[\frac{\delta\theta^2}{(2K)}\Big]\mathbb{E}_K[(R^{\ell}_{u,\delta K})^2 \mid \mathcal{F}_{2u-2}]\Big\} \\
&\ge \exp\Big[-\delta\theta^2 K^{-1}\sum_{\ell=1}^{K}\mathbb{E}_K(R^{\ell}_{u,\delta K} \mid \mathcal{F}_{2u-2})\Big] \\
&\xrightarrow{\mathbb{P}} \exp(-\delta\theta^2\widetilde{\mu}_u).
\end{aligned} \tag{A.10}$$

By the definitions of $\varphi^{\ell}_{2u-1,c}$ and $\widetilde{\varphi}^{\ell}_{2u-1,c}$ in (A.2) and (A.7),

$$\prod_{\ell=1}^{K}\Big[\frac{\varphi^{\ell}_{2u-1,\delta K}(\theta \mid \mathcal{F}_{2u-2})}{\widetilde{\varphi}^{\ell}_{2u-1,\delta K}(\theta \mid \mathcal{F}_{2u-2})}\Big] = \exp\Big[\mathrm{i}\theta K^{-1}\sum_{\ell=1}^{K}\mathbb{E}_K(R^{\ell}_{u,\delta K} \mid \mathcal{F}_{2u-1})\Big] \xrightarrow{\mathbb{P}} \exp(\mathrm{i}\theta\widetilde{\mu}_u). \tag{A.11}$$

It follows from (A.3), (A.6), and (A.9)–(A.11), with $\delta \to 0$, that $\mathbb{E}_K[\exp(\mathrm{i}\theta\bar{V}_u) \mid \mathcal{F}_{2u-2}] \xrightarrow{\mathbb{P}} \exp(\mathrm{i}\theta\widetilde{\mu}_u)$. Therefore, $\mathbb{E}_K \exp(\mathrm{i}\theta\bar{V}_u) \xrightarrow{\mathbb{P}} \exp(\mathrm{i}\theta\widetilde{\mu}_u)$, equivalently $\bar{V}_u \xrightarrow{\mathbb{P}} \widetilde{\mu}_u$. Hence, Lemma A.1(i) holds for $t = u$ whenever Lemma A.1(ii) holds for $t = u - 1$. By similar arguments, Lemma A.1(ii) holds for $t = u$ whenever Lemma A.1(i) holds for $t = u$. The induction arguments to show Lemma A.1(i) and A.1(ii) for $T < t \le 2T$ are now complete. Similar induction arguments can be used to show Lemma A.1(i) and A.1(ii) for $(m-1)T < t \le mT$ for $m = 3, \dots, M$.

The identities in (A.1) follow from multiplying (2.9) over '(m, t)' $= (1, T), \dots, (m-1, (m-1)T), (m, t)$. By (2.6) and (2.8), applying Lemma A.1(i) on $G = w_t$ we obtain $\bar{w}_t \xrightarrow{\mathbb{P}} \eta_t/\eta_{t-1}$. Therefore Lemma A.1(iii) holds. □

Lemma A.2. *Let G_u be a measurable function of $\boldsymbol{x}_u$ with $(m-1)T < u \le mT$ and define $G_{m,u}(\boldsymbol{x}_{m,u}) = \mathbb{E}_q[G_u(\boldsymbol{X}_u) \mid \boldsymbol{X}_{m,u} = \boldsymbol{x}_{m,u}]$.*

(i) *If $\mathbb{E}_q[G_u^2(\boldsymbol{X}_u)h_{u-1}(\boldsymbol{X}_{u-1})] < \infty$ then*

$$K^{-1}\sum_{\ell=1}^{K}\Big[K^{-M+1}\sum_{\boldsymbol{k}\in Z^{m-1}} G_u(\widetilde{\boldsymbol{X}}^{\boldsymbol{k}\ell}_u)H^{\boldsymbol{k}\ell}_{u-1} - G_{m,u}(\widetilde{\boldsymbol{X}}^{\ell}_{m,u})H^{\ell}_{m,u-1}\Big]^2 \xrightarrow{\mathbb{P}} 0, \tag{A.12}$$

$$\sum_{\ell=1}^{K}\Big[K^{-M+1}\sum_{\boldsymbol{k}\in Z^{m-1}} G_u(\widetilde{\boldsymbol{X}}^{\boldsymbol{k}\ell}_u)H^{\boldsymbol{k}\ell}_{u-1} - G_{m,u}(\widetilde{\boldsymbol{X}}^{\ell}_{m,u})H^{\ell}_{m,u-1}\Big] = o_p(K^{1/2}), \tag{A.13}$$

where $\boldsymbol{k}\ell = (k(1), \dots, k(m-1), \ell)$.

https://doi.org/10.1017/apr.2015.7 Published online by Cambridge University Press

(ii) *If* $\mathbb{E}_q[G_u^2(\boldsymbol{X}_u)h_u(\boldsymbol{X}_u)] < \infty$ *then*

$$K^{-1}\sum_{\ell=1}^{K}\bigg[K^{-M+1}\sum_{\boldsymbol{k}\in Z^{m-1}} G_u(\boldsymbol{X}_u^{\boldsymbol{k}\ell})H_u^{\boldsymbol{k}\ell} - G_{m,u}(\boldsymbol{X}_{m,u}^{\ell})H_{m,u}^{\ell}\bigg]^2 \xrightarrow{\mathbb{P}} 0, \tag{A.14}$$

$$\sum_{\ell=1}^{K}\bigg[K^{-M+1}\sum_{\boldsymbol{k}\in Z^{m-1}} G_u(\boldsymbol{X}_u^{\boldsymbol{k}\ell})H_u^{\boldsymbol{k}\ell} - G_{m,u}(\boldsymbol{X}_{m,u}^{\ell})H_{m,u}^{\ell}\bigg] = o_p(K^{1/2}). \tag{A.15}$$

Proof. The $m = 1$ case is trivial, so let us first consider $m = 2$. For $1 \le t \le T$, let $G_{t,u}(\boldsymbol{x}_t, \boldsymbol{x}_{2,u}) = \mathbb{E}_q[G_u(\boldsymbol{X}_u) \mid \boldsymbol{X}_t = \boldsymbol{x}_t, \boldsymbol{X}_{2,u} = \boldsymbol{x}_{2,u}]$. By Lemma A.1(iii),

$$K^{-1}\sum_{k=1}^{K} G_u(\widetilde{\boldsymbol{X}}_u^{k\ell})H_{u-1}^{k\ell} - G_{2,u}(\widetilde{\boldsymbol{X}}_{2,u}^{\ell})H_{2,u-1}^{\ell} = [1 + o_p(1)]D^{\ell}$$

uniformly over ℓ, where

$$D^{\ell} = \bigg[K^{-1}\sum_{k=1}^{K} G_u(\widetilde{\boldsymbol{X}}_u^{k\ell})H_T^{k} - G_{2,u}(\widetilde{\boldsymbol{X}}_{2,u}^{\ell})\bigg]h_{2,u-1}(\boldsymbol{X}_{2,u-1}^{\ell}). \tag{A.16}$$

By (A.16), we have the expansion

$$D^{\ell} = K^{-1}\sum_{s=1}^{2T}(d_s^{1\ell} + \cdots + d_s^{K\ell})h_{2,u-1}(\boldsymbol{X}_{2,u-1}^{\ell}),$$

where

$$d_{2t-1}^{k\ell} = [G_{t,u}(\widetilde{\boldsymbol{X}}_t^{k}, \widetilde{\boldsymbol{X}}_{2,u}^{\ell}) - G_{t-1,u}(\boldsymbol{X}_{t-1}^{k}, \widetilde{\boldsymbol{X}}_{2,u}^{\ell})]H_{t-1}^{k},$$
$$d_{2t}^{k\ell} = G_{t,u}(\boldsymbol{X}_t^{k}, \widetilde{\boldsymbol{X}}_{2,u}^{\ell})H_t^{k} - \sum_{j=1}^{K} W_t^{j} G_{t,u}(\widetilde{\boldsymbol{X}}_t^{k}, \widetilde{\boldsymbol{X}}_{2,u}^{\ell})H_{t-1}^{k},$$

with the convention that for $t = 1$, $H_{t-1}^{k} = 1$ and $G_{t-1,u}(\boldsymbol{X}_{t-1}^{k}, \widetilde{\boldsymbol{X}}_{2,u}^{\ell}) = G_{2,u}(\widetilde{\boldsymbol{X}}_{2,u}^{\ell})$.

Let $D_v^{\ell} = K^{-1}\sum_{s=1}^{v}(d_s^{1\ell} + \cdots + d_s^{K\ell})h_{2,u-1}(\boldsymbol{X}_{2,u-1}^{\ell})$. We shall show inductively that uniformly over ℓ,

$$\sum_{\ell=1}^{K}\mathbb{E}_K[(D_v^{\ell})^2 \mid \mathcal{G}_{2u-2}] = O_p(1), \qquad v = 1, \ldots, 2T, \tag{A.17}$$

where $\mathcal{G}_{2u-2}$ denotes the σ-algebra for all random variables generated in the second particle filter up to and including resampling at the $(u-1)$th stage. Since $d_1^{1\ell}, \ldots, d_1^{K\ell}$ are uncorrelated with mean 0 conditioned on $\mathcal{G}_{2u-2}$, by Lemma A.1(i) and A.1(iii),

$$\sum_{\ell=1}^{K}\mathbb{E}_K[(D_1^{\ell})^2 \mid \mathcal{G}_{2u-2}] = K^{-2}\sum_{k=1}^{K}\sum_{\ell=1}^{K}\mathbb{E}_K[(d_1^{k\ell})^2 \mid \mathcal{G}_{2u-2}]h_{2,u-1}(\boldsymbol{X}_{2,u-1}^{\ell})^2 = O_p(1).$$

Therefore (A.17) holds for $v = 1$. Consider next $v > 1$. Let $\mathcal{H}_v = \mathcal{F}_v \cup \mathcal{G}_{2u-2}$. Since $D_v^{\ell} = D_{v-1}^{\ell} + K^{-1}(d_v^{1\ell} + \cdots d_v^{K\ell})h_{2,u-1}(\boldsymbol{X}_{2,u-1}^{\ell})$, and $d_v^{1\ell}, \ldots, d_v^{K\ell}$ are conditionally independent

https://doi.org/10.1017/apr.2015.7 Published online by Cambridge University Press

with mean 0 given $\mathcal{H}_{v-1}$, by Lemma A.1(a) and (c),

$$\sum_{\ell=1}^{K} \mathbb{E}_K[(D_v^\ell)^2 \mid \mathcal{H}_{v-1}] = \sum_{\ell=1}^{K} (D_{v-1}^\ell)^2 + K^{-2} \sum_{k=1}^{K} \sum_{\ell=1}^{K} \mathbb{E}_K[(d_v^{k\ell})^2 \mid \mathcal{H}_{v-1}] h_{2,u-1}^2(\boldsymbol{X}_{2,u-1}^\ell)$$
$$= \sum_{\ell=1}^{K} (D_{v-1}^\ell)^2 + O_p(1).$$

Therefore, (A.17) for v follows from (A.17) for $v-1$. By induction, (A.17) holds for $1 \le v \le 2T$. In particular, since $D_{2T}^\ell = D^\ell$, (A.12) holds for $m = 2$.

By (A.17) for $v = 2T$, and noting that $D^1, \dots, D^K$ are conditionally independent with mean 0 given $\mathcal{F}_{2u-2}$, and that $\mathcal{G}_{2u-2} \subset \mathcal{F}_{2u-2}$, we have

$$\mathbb{E}_K\left[\left(\sum_{\ell=1}^{K} D^\ell\right)^2 \,\middle|\, \mathcal{F}_{2u-2}\right] = \sum_{\ell=1}^{K} \mathbb{E}_K[(D^\ell)^2 \mid \mathcal{F}_{2u-2}] = O_p(K). \tag{A.18}$$

Therefore, $K^{-1}(\sum_{\ell=1}^{K} D^\ell)^2 \xrightarrow{\mathbb{P}} 0$, and (A.13) holds for $m = 2$. The extension of the proof to $m > 2$ and the proofs of (A.14) and (A.15) apply similar arguments to those of (A.12) and (A.13). □

A.2. Proofs of Theorem 3.2, Corollary 3.1, and Theorem 3.3

Proof of Theorem 3.2. Let $S = \sum_{m=1}^{M} S_m$, where

$$S_m = \sum_{u=2(m-1)T+1}^{2mT} (Z_{m,u}^1 + \cdots + Z_{m,u}^K),$$
$$Z_{m,2t-1}^k = [f_{m,t}(\widetilde{\boldsymbol{X}}_{m,t}^k) - f_{m,t-1}(\boldsymbol{X}_{m,t-1}^k)] H_{m,t}^k,$$
$$Z_{m,2t}^k = f_{m,t}(\boldsymbol{X}_{m,t}^k) H_{m,t}^k - \sum_{j=1}^{K} W_t^j f_{m,t}(\widetilde{\boldsymbol{X}}_{m,t}^j) \widetilde{H}_{m,t}^j.$$

By the CLT for particle filters on unsegmented HMM sequences, see, e.g. [6], $K^{-1/2} S_m \Rightarrow N(0, \sigma_{\mathrm{Pm}}^2)$, $1 \le m \le M$. Since the particle filters operate independently and S_m depends only on the mth particle filter,

$$K^{-1/2} S \quad \Longrightarrow \quad N(0, \sigma^2), \tag{A.19}$$

where $\sigma^2 = \sum_{m=1}^{M} \sigma_{\mathrm{PF}m}^2$. By (3.11),

$$\sqrt{K}(\widehat{\psi}_U - \psi_U) = K^{-1/2} S + K^{-1/2} \sum_{u=1}^{2mU} \sum_{\ell=1}^{K} (Z_u^\ell - Z_{m,u}^\ell). \tag{A.20}$$

Therefore, by (3.10), (3.12), and (A.13) applied on $G_t(\boldsymbol{x}_t) = f_t(\boldsymbol{x}_t) - f_{t-1}(\boldsymbol{x}_{t-1})$, we have

$$K^{-1/2} \sum_{k=1}^{K} (Z_u^k - Z_{m,u}^k) \xrightarrow{\mathbb{P}} 0 \quad \text{for } u = 2t-1. \tag{A.21}$$

By (3.10), (3.12), and (A.15) applied on $G_t(\boldsymbol{x}_t) = f_t(\boldsymbol{x}_t)$ and (A.15) applied on $G_t(\boldsymbol{x}_t) = w_t(\boldsymbol{x}_t) f_t(\boldsymbol{x}_t)$, (A.21) holds for $u = 2t$. We conclude the proof of Theorem 3.2 from (A.19)–(A.21). □

https://doi.org/10.1017/apr.2015.7 Published online by Cambridge University Press

Proof of Corollary 3.1. By (3.5),

$$\widetilde{\psi}_U - \psi_U = \bigg[K^{-M} \sum_{\boldsymbol{k}\in Z^M} L(\boldsymbol{X}_U^{\boldsymbol{k}}) H(\boldsymbol{X}_U^{\boldsymbol{k}})\bigg]^{-1} \widehat{\psi}_U^{\mathrm{c}}, \tag{A.22}$$

where $\widehat{\psi}_U^{\mathrm{c}} = K^{-M} \sum_{\boldsymbol{k}\in Z^M} L(\boldsymbol{X}_U^{\boldsymbol{k}}) \psi^{\mathrm{c}}(\boldsymbol{X}_U^{\boldsymbol{k}}) H_U^{\boldsymbol{k}}$, is (3.2) with ψ replaced by $\psi^{\mathrm{c}}(\boldsymbol{x}_U) = \psi(\boldsymbol{x}_U) - \psi_U$. By Theorem 3.2,

$$\sqrt{K}\widehat{\psi}_U^{\mathrm{c}} \quad \Longrightarrow \quad N(0, \sigma_c^2). \tag{A.23}$$

By Lemma A.1(ii) and A.1(iii), $K^{-M} \sum_{\boldsymbol{k}\in Z^M} L(\boldsymbol{X}_U^{\boldsymbol{k}}) H(\boldsymbol{X}_U^{\boldsymbol{k}}) \xrightarrow{\mathbb{P}} 1$, and Corollary 3.1 therefore follows from (A.22) and (A.23). □

Proof of Theorem 3.3. We shall show the proof of Theorem 3.3 in detail for the $M = m = 2$ case, assuming, without loss of generality, that $\psi_U = 0$. It follows from [6, Corollary 2] that

$$K^{-1} \sum_{j=1}^{K} (e^j)^2 \xrightarrow{\mathbb{P}} \sigma_{\mathrm{PF2}}^2, \tag{A.24}$$

where $e^j = \sum_{\ell : A_{2,U}^\ell = j} f_{2,U}(\boldsymbol{X}_{2,U}^\ell) H_{2,U}^\ell$. By (3.10), (3.12), and (3.22),

$$\widehat{\sigma}_{\mathrm{PF2}}^2(0) = K^{-1} \sum_{j=1}^{K} \bigg(e^j + \sum_{v=2T+1}^{2U} \sum_{\ell:\, A_{2,t}^\ell = j} \zeta_v^\ell\bigg)^2,$$

where $t = \lfloor v/2 \rfloor$, with $\lfloor \cdot \rfloor$ denoting the greatest integer function,

$$\zeta_{2u-1}^\ell = Z_{2u-1}^\ell - Z_{2,2u-1}^\ell, \qquad \zeta_{2u}^\ell = (\#_u^\ell - K W_u^\ell)\chi^\ell,$$

$$\chi^\ell = K^{-1} \sum_{k=1}^{K} f_u(\widetilde{\boldsymbol{X}}^{k\ell}) \widetilde{H}_u^{k\ell} - f_{2,u}(\boldsymbol{X}_{2,u}^\ell) \widetilde{H}_{2,u}^\ell.$$

Therefore, by (A.24), in order to show (3.22), it suffices to show that

$$K^{-1} \sum_{j=1}^{K} \bigg(\sum_{v=2T+1}^{2U} \sum_{\ell:\, A_{2,t}^\ell = j} \zeta_v^\ell\bigg)^2 \xrightarrow{\mathbb{P}} 0.$$

We shall apply induction to show that

$$K^{-1} \sum_{j=1}^{K} \bigg(\sum_{v=2T+1}^{s} \sum_{\ell:\, A_{2,t}^\ell = j} \zeta_v^\ell\bigg)^2 \xrightarrow{\mathbb{P}} 0, \qquad s = 2T+1, \dots, 2U. \tag{A.25}$$

By (3.10), (3.12), and (A.16), it follows that $\zeta_{2u-1}^\ell = [1 + o_p(1)] D^\ell$ uniformly in ℓ, for $G_u(\boldsymbol{x}_u) = f_u(\boldsymbol{x}_u) - f_{u-1}(\boldsymbol{x}_{u-1})$. Therefore, by (A.18),

$$\sum_{\ell=1}^{K} \mathbb{E}_K[(\zeta_{2u-1}^\ell)^2 \mid \mathcal{F}_{2u-2}] \xrightarrow{\mathbb{P}} 0. \tag{A.26}$$

https://doi.org/10.1017/apr.2015.7 Published online by Cambridge University Press

Since $A^{\ell}_{2,T} = \ell$, $1 \le \ell \le K$, (A.26) for $u = T+1$ implies (A.25) for $s = 2T+1$. Since $\zeta^1_{2u-1}, \dots, \zeta^K_{2u-1}$ are independent with mean 0 conditioned on $\mathcal{F}_{2u-2}$, and ζ^{ℓ}_v are measurable with respect to $\mathcal{F}_{2u-2}$ for $v \le 2u-2$, it follows that

$$K^{-1}\sum_{j=1}^{K}\mathbb{E}_K\bigg[\bigg(\sum_{v=2T+1}^{2u-1}\sum_{\ell:\, A^{\ell}_{2,t}=j}\zeta^{\ell}_v\bigg)^2 \,\bigg|\, \mathcal{F}_{2u-2}\bigg]$$
$$= K^{-1}\sum_{j=1}^{K}\bigg(\sum_{v=2T+1}^{2u-2}\sum_{\ell:\, A^{\ell}_{2,t}=j}\zeta^{\ell}_v\bigg)^2 + K^{-1}\sum_{\ell=1}^{K}\mathbb{E}_K[(\zeta^{\ell}_{2u-1})^2 \mid \mathcal{F}_{2u-2}].$$

Therefore, by (A.26), (A.25) for $s = 2u-1$ follows from (A.25) for $s = 2u-2$.

Since $\mathrm{var}_K(\#^{\ell}_u \mid \mathcal{F}_{2u-1}) = KW^{\ell}_u(1-W^{\ell}_u)$, $\mathrm{cov}_K(\#^i_u, \#^{\ell}_u \mid \mathcal{F}_{2u-1}) = -KW^i_uW^{\ell}_u$, and χ^{ℓ} are measurable with respect to $\mathcal{F}_{2u-1}$, we obtain

$$K^{-1}\sum_{j=1}^{K}\bigg[\bigg(\sum_{v=2T+1}^{2u}\sum_{\ell:\, A^{\ell}_{2,T}=j}\zeta^{\ell}_v\bigg)^2 \,\bigg|\, \mathcal{F}_{2u-1}\bigg]$$
$$= K^{-1}\sum_{j=1}^{K}\bigg(\sum_{v=2T+1}^{2u-1}\sum_{\ell:\, A^{\ell}_{2,T}=j}\zeta^{\ell}_v\bigg)^2 + \sum_{\ell=1}^{K}W^{\ell}_u(\chi^{\ell})^2 - K\bigg(\sum_{\ell=1}^{K}W^{\ell}_u\chi^{\ell}\bigg)^2. \qquad \text{(A.27)}$$

It follows from an induction argument similar to that used to show (A.26) that $\sum_{\ell=1}^{k} W^{\ell}_u(\chi^{\ell})^2 \xrightarrow{\mathbb{P}} 0$. Therefore, by (A.27), (A.25) for $s = 2u$ follows from (A.25) for $s = 2u-1$. The induction arguments are complete and we have shown that (A.25) holds for $2T+1 \le s \le 2U$, and Theorem 3.3 holds. □

Acknowledgement

A. Jasra was supported by an MOE Singapore grant R-155-000-119-133 and is affiliated with the RMI and CQF at the National University of Singapore. We thank the anonymous referee for comments which have greatly improved the paper.

References

[1] Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. 2nd edn. John Wiley, New York.
[2] Andrieu, C., Doucet, A. and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Statist. Soc. B* **72,** 269–342.
[3] Billingsley, P. (1986). *Probability and Measure*, 2nd edn. John Wiley, New York.
[4] Briers, M., Doucet, A. and Singh, S. S. (2005). Sequential auxiliary particle belief propagation. In *Proc. 8th Internat. Conf. Information Fusion,* IEEE, New York.
[5] Cappé, O., Moulines, É. and Ryden, T. (2005). *Inference in Hidden Markov Models*. Springer, New York.
[6] Chan, H. P. and Lai, T. L. (2013). A general theory of particle filters in hidden Markov models and some applications. *Ann. Statist.* **41,** 2877–2904.
[7] Chan, H. P. and Lai, T. L. (2014). A new approach to Markov chain Monte Carlo with applications to adaptive particle filters. Unpublished manuscript.
[8] Chopin, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.* **32,** 2385–2411.
[9] Chopin, N., Jacob, P. E. Papaspiliopoulos, O. (2013). SMC^2: an efficient algorithm for sequential analysis of state space models. *J. R. Statist. Soc. B* **75**, 397–426.
[10] Del Moral, P. (2004). *Feynman–Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer, New York.

[11] Douc, R. and Moulines, E. (2008). Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Ann. Statist.* **36,** 2344–2376.

[12] Doucet, A. and Johansen, A. M. (2011). A tutorial on particle filtering and smoothing: fifteen years later. In *The Oxford Handbook of Nonlinear Filtering*, Oxford University Press, pp. 656–704.

[13] Fearnhead, P., Wyncoll, D. and Tawn, J. (2010). A sequential smoothing algorithm with linear computatonal cost. *Biometrika* **97,** 447–464.

[14] Gordon, N. J., Salmond, D. J. and Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proc. Radar Signal Process.* **140,** 107–113.

[15] Künsch, H. R. (2005). Recursive Monte Carlo filters: algorithms and theoretical analysis. *Ann. Statist.* **33,** 1983–2021.

[16] Lee, A. and Whiteley, N. (2014). Forest resampling for distributed sequential Monte Carlo. Preprint. Available at http://arxiv.org/abs/1406.6010.

[17] Lee, A. *et al.* (2010) On the utility of graphics cards to perform massively parallel simulation of advanced Monte Carlo methods. *J. Comp. Graph. Statist.* **19,**, 769–789.

[18] Lindsten, F. *et al.* (2014). Divide-and-conquer with sequential Monte Carlo. Preprint. Available at http://arxiv.org/abs/1406.4993.

[19] Persing, A. and Jasra, A. (2013). Likelihood computation for hidden Markov models via generalized two-filter smoothing. *Statist. Prob. Lett.* **83,** 1433–1442.

[20] Vergé, C., Dubarry, C., Del Moral, P. and Moulines, É. (2015). On parallel implementation of sequential Monte Carlo methods: the island particle model. *Statist. Comput.* **25,** 243–260.

https://doi.org/10.1017/apr.2015.7 Published online by Cambridge University Press