The Annals of Statistics
2013, Vol. 41, No. 6, 2877–2904
DOI: 10.1214/13-AOS1172

A GENERAL THEORY OF PARTICLE FILTERS IN HIDDEN MARKOV MODELS AND SOME APPLICATIONS

BY HOCK PENG CHAN[1] AND TZE LEUNG LAI[2]

National University of Singapore and Stanford University

By making use of martingale representations, we derive the asymptotic normality of particle filters in hidden Markov models and a relatively simple formula for their asymptotic variances. Although repeated resamplings result in complicated dependence among the sample paths, the asymptotic variance formula and martingale representations lead to consistent estimates of the standard errors of the particle filter estimates of the hidden states.

1. Introduction. Let $\mathbf{X} = \{X_t, t \geq 1\}$ be a Markov chain and let $Y_1, Y_2, \ldots$ be conditionally independent given $\mathbf{X}$, such that

$$X_t \sim p_t(\cdot|X_{t-1}), \qquad Y_t \sim g_t(\cdot|X_t), \tag{1.1}$$

in which p_t and g_t are density functions with respect to some measures ν_X and ν_Y, and $p_1(\cdot|X_0)$ denotes the initial density $p_1(\cdot)$ of X_1. Gordon, Salmond and Smith [11] introduced *particle filters* for Monte Carlo estimation of $E[\psi(X_T)|Y_1, \ldots, Y_T]$. More generally, letting $\mathbf{X}_t = (X_1, \ldots, X_t)$, $\mathbf{Y}_t = (Y_1, \ldots, Y_t)$, and ψ be a measurable real-valued function of $\mathbf{X}_T$, we consider estimation of $\psi_T := E[\psi(\mathbf{X}_T)|\mathbf{Y}_T]$. The density function of $\mathbf{X}_T$ conditional on $\mathbf{Y}_T$ in the above hidden Markov model (HMM) is

$$\tilde{p}_T(\mathbf{x}_T|\mathbf{Y}_T) \propto \prod_{t=1}^{T} \bigl[p_t(x_t|x_{t-1}) g_t(Y_t|x_t)\bigr]. \tag{1.2}$$

However, this conditional distribution is often difficult to sample from and the normalizing constant is also difficult to compute for high-dimensional or complicated state spaces. Particle filters that use sequential Monte Carlo (SMC) methods involving importance sampling and resampling have been developed to circumvent this difficulty. Asymptotic normality of the particle filter estimate of ψ_T when the number of simulated trajectories becomes infinite has also been established [6, 8, 9, 12]. However, no explicit formulas are available to estimate the standard error of $\hat{\psi}_T$ consistently.

In this paper, we provide a comprehensive theory of the SMC estimate $\hat{\psi}_T$, which includes asymptotic normality and consistent standard error estimation. The

Received May 2012; revised September 2013.

[1] Suported by the National University of Singapore Grant R-155-000-120-112.

[2] Supported by NSF Grant DMS-11-06535.

MSC2010 subject classifications. Primary 60F10, 65C05; secondary 60J22, 60K35.

Key words and phrases. Particle filter, importance sampling and resampling, standard error.

main results are stated in Theorems 1 and 2 in Section 2 for the case in which bootstrap resampling is used, and extensions are given in Section 4 to residual Bernoulli (instead of bootstrap) resampling. The proof of Theorem 1 for the case where bootstrap resampling is used at every stage, given in Section 3.2, proceeds in two steps. First we assume that the normalizing constants [i.e., constants of proportionality in (1.2)] are easily computable. We call this the "basic prototype," for which we derive in Section 3.1 martingale representations of $m(\tilde{\psi}_T - \psi_T)$ for novel SMC estimates $\tilde{\psi}_T$ that involve likelihood ratios and general resampling weights. We first encountered this prototype in connection with rare event simulation using SMC methods in [5]. Although traditional particle filters also use similar sequential importance sampling procedures and resampling weights that are proportional to the likelihood ratio statistics, the estimates of ψ_T are coarser averages that do not have this martingale property. Section 3.2 gives an asymptotic analysis of $m(\hat{\psi}_T - \psi_T)$ as $m \to \infty$ in this case, by making use of the results in Section 3.1.

In contrast to Gilks and Berzuini [10] who use particle set sizes m_t which increase with t and approach ∞ in a certain way to guarantee consistency of their standard error estimate which differs from ours, we use the same particle set size m for every t, where m is the number of SMC trajectories. As noted in page 132 of [10], "the mode of convergence of the theorem is not directly relevant to the practical context" in which $m_1 = m_2 = \cdots = m$. A major reason why we are able to overcome this difficulty in developing a consistent standard error estimate lies in the martingale approximation for our SMC estimate $\hat{\psi}_T$. We note in this connection that for the basic prototype, although both martingale representations of $m(\tilde{\psi}_T - \psi_T)$ developed in Section 3.1 can be used to prove the asymptotic normality of $\tilde{\psi}_T$ as the number m of SMC trajectories approaches ∞, only one of them leads to an estimable expression for the asymptotic variance. Further discussion of the main differences between the traditional and our particle filters is given in Section 5.

2. Standard error estimation for bootstrap filters.

Since $Y_1, Y_2, \ldots$ are the observed data, we will treat them as constants in the sequel. Let $q_t(\cdot|\mathbf{x}_{t-1})$ be a conditional density function with respect to ν_X such that $q_t(x_t|\mathbf{x}_{t-1}) > 0$ whenever $p_t(x_t|x_{t-1}) > 0$. In the case $t = 1$, $q_t(\cdot|\mathbf{x}_{t-1}) = q_1(\cdot)$.

2.1. *Bootstrap resampling at every stage.*

Let

$$w_t(\mathbf{x}_t) = p_t(x_t|x_{t-1}) g_t(Y_t|x_t) / q_t(x_t|\mathbf{x}_{t-1}). \tag{2.1}$$

To estimate ψ_T, a particle filter first generates m conditionally independent random variables $\widetilde{X}_t^1, \ldots, \widetilde{X}_t^m$ at stage t, with $\widetilde{X}_t^i$ having density $q_t(\cdot|\mathbf{X}_{t-1}^i)$, to form $\widetilde{\mathbf{X}}_t^i = (\mathbf{X}_{t-1}^i, \widetilde{X}_t^i)$ and then use normalized resampling weights

$$W_t^i = w_t(\widetilde{\mathbf{X}}_t^i) \Big/ \sum_{j=1}^m w_t(\widetilde{\mathbf{X}}_t^j) \tag{2.2}$$

to draw m sample paths $\mathbf{X}_t^j$, $1 \le j \le m$, from $\{\widetilde{\mathbf{X}}_t^i, 1 \le i \le m\}$. Note that W_t^i are the importance weights attached to $\widetilde{\mathbf{X}}_t^i$ and that after resampling the $\mathbf{X}_t^i$ have equal weights. The following recursive algorithm can be used to implement particle filters with bootstrap resampling. It generates not only the $\widetilde{\mathbf{X}}_T^i$ but also the ancestral origins A_{T-1}^i that are used to compute the standard error estimates. It also computes $\widetilde{H}_t^i$ and H_t^i recursively, where

$$\begin{aligned} \widetilde{H}_t^i &= \bar{w}_1 \cdots \bar{w}_t h_t(\widetilde{\mathbf{X}}_t^i), \qquad H_t^i = \bar{w}_1 \cdots \bar{w}_t h_t(\mathbf{X}_t^i), \\ \bar{w}_k &= \sum_{j=1}^m w_k(\widetilde{\mathbf{X}}_k^j)/m, \qquad h_t(\mathbf{x}_t) = 1 \Big/ \prod_{k=1}^t w_k(\mathbf{x}_k). \end{aligned} \tag{2.3}$$

Initialization: let $A_0^i = i$ and $H_0^i = 1$ for all $1 \le i \le m$.

Importance sampling at stage $t = 1, \dots, T$: generate conditionally independent $\widetilde{X}_t^i$ from $q_t(\cdot|\mathbf{X}_{t-1}^i)$, $1 \le i \le m$.

Bootstrap resampling at stage $t = 1, \dots, T-1$: generate i.i.d. random variables $B_t^1, \dots, B_t^m$ such that $P(B_t^1 = j) = W_t^j$ and let

$$(\mathbf{X}_t^j, A_t^j, H_t^j) = (\widetilde{\mathbf{X}}_t^{B_t^j}, A_{t-1}^{B_t^j}, \widetilde{H}_t^{B_t^j}) \qquad \text{for } 1 \le j \le m.$$

The SMC estimate is defined by

$$\hat{\psi}_T = (m\bar{w}_T)^{-1} \sum_{i=1}^m \psi(\widetilde{\mathbf{X}}_T^i) w_T(\widetilde{\mathbf{X}}_T^i). \tag{2.4}$$

Let $E_{\tilde{p}_T}$ denote expectation with respect to the probability measure under which $\mathbf{X}_T$ has density (1.2), and let E_q denote that under which $X_t|\mathbf{X}_{t-1}$ has the conditional density function q_t for all $1 \le t \le T$. In Section 3.2 we prove the following theorem on the asymptotic normality of $\hat{\psi}_T$ and the consistency of its standard error estimate. Define

$$\begin{aligned} \eta_t &= E_q\left[\prod_{k=1}^t w_k(\mathbf{X}_k)\right], \\ h_t^*(\mathbf{x}_t) &= \eta_t \Big/ \prod_{k=1}^t w_k(\mathbf{x}_k), \qquad \zeta_t = \eta_{t-1}/\eta_t, \\ \Gamma_t(\mathbf{x}_t) &= \prod_{k=1}^t [w_k(\mathbf{x}_k) + w_k^2(\mathbf{x}_k)], \end{aligned} \tag{2.5}$$

with the convention $\eta_0 = 1$ and $h_0^* \equiv \Gamma_0 \equiv 1$. Let

$$L_T(\mathbf{x}_T) = \tilde{p}_T(\mathbf{x}_T|\mathbf{Y}_T) \Big/ \prod_{t=1}^T q_t(x_t|\mathbf{x}_{t-1}). \tag{2.6}$$

Let $f_0 = 0$ and define for $1 \le t \le T$,

$$f_t(\mathbf{x}_t)[= f_{t,T}(\mathbf{x}_t)] = E_q\{[\psi(\mathbf{X}_T) - \psi_T]L_T(\mathbf{X}_T)|\mathbf{X}_t = \mathbf{x}_t\}. \tag{2.7}$$

THEOREM 1. *Assume that* $\sigma^2 < \infty$, *where* $\sigma^2 = \sum_{k=1}^{2T-1} \sigma_k^2$ *and*

$$\begin{aligned} \sigma_{2t-1}^2 &= E_q\{[f_t^2(\mathbf{X}_t) - f_{t-1}^2(\mathbf{X}_{t-1})]h_{t-1}^*(\mathbf{X}_{t-1})\}, \\ \sigma_{2t}^2 &= E_q[f_t^2(\mathbf{X}_t)h_t^*(\mathbf{X}_t)]. \end{aligned} \tag{2.8}$$

Then $\sqrt{m}(\hat{\psi}_T - \psi_T) \Rightarrow N(0, \sigma^2)$. *Moreover, if* $E_q\Gamma_T(\mathbf{X}_T) < \infty$, *then* $\hat{\sigma}^2(\hat{\psi}_T) \xrightarrow{p} \sigma^2$, *where for any real number* μ,

$$\hat{\sigma}^2(\mu) = m^{-1} \sum_{j=1}^{m} \left(\sum_{i \,:\, A_{T-1}^i = j} \frac{w_T(\widetilde{\mathbf{X}}_T^i)}{\bar{w}_T} [\psi(\widetilde{\mathbf{X}}_T^i) - \mu] \right)^2. \tag{2.9}$$

2.2. *Extension to occasional resampling*. Due to the computational cost of resampling and the variability introduced by the resampling step, *occasional resampling* has been recommended, for example, by Liu (Chapter 3.4.4 of [13]), who propose to resample at stage k only when cv_k, the coefficient of variation of the resampling weights at stage k, exceeds some threshold. Let $\tau_1 < \tau_2 < \cdots < \tau_r$ be the successive resampling times and let $\tilde{\tau}(k)$ be the most recent resampling time before stage k, with $\tilde{\tau}(k) = 0$ if no resampling has been carried out before time k. Define

$$v_k(\mathbf{x}_k) = \prod_{t=\tilde{\tau}(k)+1}^{k} w_t(\mathbf{x}_t), \qquad V_k^i = v_k(\widetilde{\mathbf{X}}_k^i)/(m\bar{v}_k), \tag{2.10}$$

where $\bar{v}_k = m^{-1} \sum_{j=1}^{m} v_k(\widetilde{\mathbf{X}}_k^j)$. Note that if bootstrap resampling is carried out at stage k, then V_k^i, $1 \le i \le m$, are the resampling weights. Define the SMC estimator

$$\hat{\psi}_{\mathrm{OR}} = \sum_{i=1}^{m} V_T^i \psi(\widetilde{\mathbf{X}}_T^i). \tag{2.11}$$

In Section 4.3, we show that the resampling times, which resample at stage k when cv_k^2 exceeds some threshold c, satisfy

$$r \xrightarrow{p} r^* \quad \text{and} \quad \tau_s \xrightarrow{p} \tau_s^* \qquad \text{for } 1 \le s \le r^* \text{ as } m \to \infty, \tag{2.12}$$

for some nonrandom positive integers r^*, $\tau_1^*, \ldots$ and $\tau_{r^*}^*$.

In Section 4.2 we prove the following extension of Theorem 1 to $\hat{\psi}_{\mathrm{OR}}$ and give an explicit formula (4.6), which is assumed to be finite, for the limiting variance σ_{OR}^2 of $\sqrt{m}(\hat{\psi}_{\mathrm{OR}} - \psi_T)$.

THEOREM 2. *Under* (2.12), $\sqrt{m}(\hat{\psi}_{\mathrm{OR}} - \psi_T) \Rightarrow N(0, \sigma^2_{\mathrm{OR}})$ *as* $m \to \infty$. *Moreover, if* $E_q \Gamma_T(\mathbf{X}_T) < \infty$, *then* $\hat{\sigma}^2_{\mathrm{OR}}(\hat{\psi}_{\mathrm{OR}}) \xrightarrow{p} \sigma^2_{\mathrm{OR}}$, *where for any real number* μ,

$$\hat{\sigma}^2_{\mathrm{OR}}(\mu) = m^{-1} \sum_{j=1}^{m} \biggl(\sum_{i\,:\,A^i_{\tau_r}=j} \frac{v_T(\widetilde{\mathbf{X}}^i_T)}{\bar{v}_T} [\psi(\widetilde{\mathbf{X}}^i_T) - \mu] \biggr)^2.$$

2.3. *A numerical study.* Yao [15] has derived explicit formulas for $\psi_T = E(X_T|\mathbf{Y}_T)$ in the normal mean shift model $\{(X_t, Y_t) : t \geq 1\}$ with the unobserved states X_t generated recursively by

$$X_t = \begin{cases} X_{t-1}, & \text{with prob. } 1-\rho, \\ Z_t, & \text{with prob. } \rho, \end{cases}$$

where $0 < \rho < 1$ is given and $Z_t \sim N(0, \xi)$ is independent of $\mathbf{X}_{t-1}$. The observations are $Y_t = X_t + \varepsilon_t$, in which the ε_t are i.i.d. standard normal, $1 \leq t \leq T$. Instead of $\mathbf{X}_T$, we simulate using the bootstrap filter the change-point indicators $\mathbf{I}_T := (I_1, \ldots, I_T)$, where $I_t = \mathbf{1}_{\{X_t \neq X_{t-1}\}}$ for $t \geq 2$ and $I_1 = 1$; this technique of simulating $E(X_T|\mathbf{Y}_T)$ via realizations of $E(X_T|\mathbf{I}_T, \mathbf{Y}_T)$ is known as Rao–Blackwellization. Let $C_t = \max\{j \leq t : I_j = 1\}$ denote the most recent change-point up to time t, $\lambda_t = (t - C_t + 1 + \xi^{-1})^{-1}$ and $\mu_t = \lambda_t \sum_{i=C_t}^{t} Y_i$. It can be shown that $E(X_T|\mathbf{I}_T, \mathbf{Y}_T) = \mu_T$ and that for $t \geq 2$, $P\{I_t = 1|\mathbf{I}_{t-1}, \mathbf{Y}_t\} = a(\rho)/\{a(\rho) + b(\rho)\}$, where $a(\rho) = \rho\phi_{0,1+\xi}(Y_t)$ and $b(\rho) = (1-\rho)\phi_{\mu_{t-1},1+\lambda_{t-1}}(Y_t)$, in which $\phi_{\mu,V}$ denotes the $N(\mu, V)$ density function.

The results in Table 1 are based on a single realization of $\mathbf{Y}_{1000}$ with $\xi = 1$ and $\rho = 0.01$. We use particle filters to generate $\{\mathbf{I}^i_{1000} : 1 \leq i \leq m\}$ for $m = 10{,}000$ and compute $\hat{\psi}_{\mathrm{OR}}$ for $T = 200, 400, 600, 800, 1000$, using bootstrap resampling when $\mathrm{cv}^2_t \geq c$, for various thresholds c. The values of $\hat{\sigma}_{\mathrm{OR}}(\hat{\psi}_{\mathrm{OR}})$ in Table 1 suggest that

TABLE 1
Simulated values of $\hat{\psi}_{\mathrm{OR}} \pm \hat{\sigma}_{\mathrm{OR}}(\hat{\psi}_{\mathrm{OR}})/\sqrt{m}$. *An italicized value denotes that* ψ_T *is outside the interval but within two standard errors of* $\hat{\psi}_{\mathrm{OR}}$; *a boldfaced value denotes that* ψ_T *is more than two standard errors away from* $\hat{\psi}_{\mathrm{OR}}$

c	**T = 200**	**T = 400**	**T = 600**	**T = 800**	**T = 1000**
∞	*−0.3718 ± 0.0067*	*−0.1791 ± 0.0281*	*−0.3022 ± 0.0136*	0.5708 ± 0.0380	−0.5393 ± 0.0180
50	−0.3592 ± 0.0019	−0.1453 ± 0.0056	*−0.2768 ± 0.0020*	0.6004 ± 0.0030	−0.5314 ± 0.0024
10	*−0.3622 ± 0.0019*	−0.1421 ± 0.0032	−0.2802 ± 0.0029	0.6013 ± 0.0024	−0.5287 ± 0.0015
5	*−0.3558 ± 0.0018*	*−0.1382 ± 0.0027*	−0.2777 ± 0.0026	0.6004 ± 0.0022	*−0.5321 ± 0.0012*
2	−0.3584 ± 0.0013	−0.1421 ± 0.0024	−0.2798 ± 0.0018	0.6030 ± 0.0020	*−0.5316 ± 0.0008*
1	*−0.3615 ± 0.0013*	−0.1402 ± 0.0027	*−0.2761 ± 0.0019*	*0.5997 ± 0.0019*	−0.5291 ± 0.0014
0.5	*−0.3610 ± 0.0016*	**−0.1488 ± 0.0036**	−0.2785 ± 0.0020	*0.6057 ± 0.0021*	−0.5315 ± 0.0014
0	−0.3570 ± 0.0032	*−0.1485 ± 0.0067*	−0.2759 ± 0.0033	0.6033 ± 0.0031	−0.5301 ± 0.0047
ψ_T	−0.3590	−0.1415	−0.2791	0.6021	−0.5302

TABLE 2
Fraction of confidence intervals $\hat{\psi}_{\text{OR}} \pm \hat{\sigma}/\sqrt{m}$ *containing the true mean* ψ_T

	$\hat{\sigma}^2 = \hat{\sigma}^2_{\text{OR}}(\hat{\psi}_{\text{OR}})$		$\hat{\sigma}^2 = \widehat{V}_T$	
T	**1 se**	**2 se**	**1 se**	**2 se**
200	0.644	0.956	0.980	1.000
400	0.652	0.948	0.998	1.000
600	0.674	0.958	1.000	1.000
800	0.716	0.974	1.000	1.000
1000	0.650	0.958	1.000	1.000

variability of $\hat{\psi}_{\text{OR}}$ is minimized when $c = 2$. Among the 40 simulated values of $\hat{\psi}_{\text{OR}}$ in Table 1, 24 fall within one estimated standard error [$= \hat{\sigma}_{\text{OR}}(\hat{\psi}_{\text{OR}})/\sqrt{m}$] and 39 within two estimated standard errors of ψ_T. This agrees well with the corresponding numbers 27 and 39, respectively, given by the central limit theorem in Theorem 2. Table 2 reports a larger simulation study involving 500 realizations of $\mathbf{Y}_{1000}$, with an independent run of the particle filter for each realization. The table shows confidence intervals for each T, with resampling threshold $c = 2$. The results for $\hat{\sigma}_{\text{OR}}(\hat{\psi}_{\text{OR}})$ agree well with the coverage probabilities of 0.683 and 0.954 given by the limiting standard normal distribution in Theorem 2. Table 2 also shows that the Gilks–Berzuini estimator $\widehat{V}_T$ described in Section 5 is very conservative.

3. Martingales and proof of Theorem 1. In this section we first consider in Section 3.1 the basic prototype mentioned in the second paragraph of Section 1, for which $m(\tilde{\psi}_T - \psi_T)$ can be expressed as a sum of martingale differences for a new class of particle filters. We came up with this martingale representation, which led to a standard error estimate for $\tilde{\alpha}_T$, in [5] where we used a particle filter $\tilde{\alpha}_T$ to estimate the probability α of a rare event that $\mathbf{X}_T$ belongs to Γ. Unlike the present setting of HMMs, the $\mathbf{X}_T$ is not a vector of hidden states in [5], where we showed that $m(\tilde{\alpha}_T - \alpha)$ is a martingale, thereby proving the unbiasedness of $\tilde{\alpha}_T$ and deriving its standard error. In Section 3.1 we extend the arguments for $m(\tilde{\alpha}_T - \alpha)$ in [5] to $m(\tilde{\psi}_T - \psi_T)$ under the assumption that the constants of proportionality in (1.2) are easily computable, which we have called the "basic prototype" in Section 1. We then apply the results of Section 3.1 to develop in Section 3.2 martingale approximations for traditional particle filters and use them to prove Theorem 1.

3.1. *Martingale representations for the basic prototype.* We consider here the basic prototype, which requires T to be fixed, in a more general setting than HMM. Let $(\mathbf{X}_T, \mathbf{Y}_T)$ be a general random vector, with only $\mathbf{Y}_T$ observed, such that the conditional density $\tilde{p}_T(\mathbf{X}_T|\mathbf{Y}_T)$ of $\mathbf{X}_T$ given $\mathbf{Y}_T$ is readily computable. As in Section 2, since $\mathbf{Y}_T$ represents the observed data, we can treat it as constant and simply denote $\tilde{p}_T(\mathbf{x}_T|\mathbf{Y}_T)$ by $\tilde{p}_T(\mathbf{x}_T)$. The bootstrap resampling scheme in Section 2.1 can be applied to this general setting and we can use general weight functions $w_t(\mathbf{x}_t) > 0$, of which (2.1) is a special case for the HMM (1.1). Because $\tilde{p}_T(\mathbf{x}_T)$ is readily computable, so is the likelihood ratio $L_T(\mathbf{x}_T) = \tilde{p}_T(\mathbf{x}_T)/\prod_{t=1}^T q_t(x_t|\mathbf{x}_{t-1})$.

This likelihood ratio plays a fundamental role in the theory of the particle filter estimate $\tilde{\alpha}_T$ of the rare event probability α in [5]. An unbiased estimate of ψ_T under the basic prototype is

$$\tilde{\psi}_T = m^{-1} \sum_{i=1}^{m} L_T(\widetilde{\mathbf{X}}_T^i)\psi(\widetilde{\mathbf{X}}_T^i)H_{T-1}^i, \tag{3.1}$$

where H_t^i is defined in (2.3). The unbiasedness follows from the fact that $m(\tilde{\psi}_T - \psi_T)$ can be expressed as a sum of martingale differences. There are in fact two martingale representations of $m(\tilde{\psi}_T - \psi_T)$. One is closely related to that of Del Moral (Chapter 9 of [8]); see (3.11) below. The other is related to the variance estimate $\hat{\sigma}^2(\tilde{\psi}_T)$ in Theorem 1 and is given in Lemma 1. Recall the meaning of the notation E_q introduced in the second paragraph of Section 2.1. Let $\#_k^i$ denote the number of copies of $\widetilde{\mathbf{X}}_k^i$ generated from $\{\widetilde{\mathbf{X}}_k^1, \dots, \widetilde{\mathbf{X}}_k^m\}$ to form the m particles in the kth generation. Then, conditionally, $(\#_k^1, \dots, \#_k^m) \sim \text{Multinomial}(m, W_k^1, \dots, W_k^m)$.

LEMMA 1. *Let $\tilde{f}_0 = \psi_T$ and define for $1 \le t \le T$,*

$$\tilde{f}_t(\mathbf{x}_t) = E_q[\psi(\mathbf{X}_T)L_T(\mathbf{X}_T)|\mathbf{X}_t = \mathbf{x}_t]. \tag{3.2}$$

Then $\tilde{f}_t(\mathbf{x}_t) = E_q[\tilde{f}_T(\mathbf{X}_T)|\mathbf{X}_t = \mathbf{x}_t]$ and

$$m(\tilde{\psi}_T - \psi_T) = \sum_{j=1}^{m} (\varepsilon_1^j + \cdots + \varepsilon_{2T-1}^j), \tag{3.3}$$

where

$$\begin{aligned} \varepsilon_{2t-1}^j &= \sum_{i\,:\,A_{t-1}^i = j} [\tilde{f}_t(\widetilde{\mathbf{X}}_t^i) - \tilde{f}_{t-1}(\mathbf{X}_{t-1}^i)]H_{t-1}^i, \\ \varepsilon_{2t}^j &= \sum_{i\,:\,A_{t-1}^i = j} (\#_t^i - mW_t^i)[\tilde{f}_t(\widetilde{\mathbf{X}}_t^i)\widetilde{H}_t^i - \tilde{f}_0]. \end{aligned} \tag{3.4}$$

Moreover, for each fixed j, $\{\varepsilon_k^j, \mathcal{F}_k, 1 \le k \le 2T-1\}$ is a martingale difference sequence, where

$$\begin{aligned} \mathcal{F}_{2t-1} &= \sigma(\{\widetilde{X}_1^i : 1 \le i \le m\} \cup \{(\mathbf{X}_s^i, \widetilde{\mathbf{X}}_{s+1}^i, A_s^i) : 1 \le s < t, 1 \le i \le m\}), \\ \mathcal{F}_{2t} &= \sigma(\mathcal{F}_{2t-1} \cup \{(\mathbf{X}_t^i, A_t^i) : 1 \le i \le m\}) \end{aligned} \tag{3.5}$$

are the σ-algebras generated by the random variables associated with the m particles just before and just after the tth resampling step, respectively.

PROOF. Recalling that the "first generation" of the m particles consists of $\widetilde{X}_1^1, \dots, \widetilde{X}_1^m$ (before resampling), $A_t^i = j$ if the first component of $\mathbf{X}_t^i$ is $\widetilde{X}_1^j$. Thus,

A_t^i represents the "ancestral origin" of the "genealogical particle" $\mathbf{X}_t^i$. It follows from (2.2), (2.3) and simple algebra that for $1 \le i \le m$,

$$mW_t^i = H_{t-1}^i / \widetilde{H}_t^i, \tag{3.6}$$

$$\sum_{i:A_t^i=j} \tilde{f}_t(\mathbf{X}_t^i) H_t^i = \sum_{i:A_{t-1}^i=j} \#_t^i \tilde{f}_t(\widetilde{\mathbf{X}}_t^i) \widetilde{H}_t^i, \tag{3.7}$$

noting that $\widetilde{\mathbf{X}}_t^i$ has the same first component as $\mathbf{X}_{t-1}^i$. Multiplying (3.6) by $\tilde{f}_t(\widetilde{\mathbf{X}}_t^i)\widetilde{H}_t^i$ and combining with (3.7) yield

$$\begin{aligned}
&\sum_{t=1}^{T} \sum_{i:A_{t-1}^i=j} [\tilde{f}_t(\widetilde{\mathbf{X}}_t^i) - \tilde{f}_{t-1}(\mathbf{X}_{t-1}^i)] H_{t-1}^i \\
&\quad + \sum_{t=1}^{T-1} \sum_{i:A_{t-1}^i=j} (\#_t^i - mW_t^i) \tilde{f}_t(\widetilde{\mathbf{X}}_t^i) \widetilde{H}_t^i \\
&\quad = \sum_{i:A_{T-1}^i=j} \tilde{f}_T(\widetilde{\mathbf{X}}_T^i) H_{T-1}^i - \tilde{f}_0,
\end{aligned} \tag{3.8}$$

recalling that $A_0^i = i$ and $H_0^i = 1$. Since $\tilde{f}_0 = \psi_T$ and $\tilde{f}_T(\mathbf{x}_T) = \psi(\mathbf{x}_T) L_T(\mathbf{x}_T)$ by (3.2), it follows from (3.1) that

$$m(\tilde{\psi}_T - \psi_T) = \sum_{i=1}^{m} \tilde{f}_T(\widetilde{\mathbf{X}}_T^i) H_{T-1}^i - m\tilde{f}_0. \tag{3.9}$$

Since $\sum_{i=1}^m \#_t^i = \sum_{i=1}^m mW_t^i = m$ for $1 \le t \le T-1$, combining (3.9) with (3.8) yields (3.3).

Let E_m denote expectation under the probability measure induced by the random variables in the SMC algorithm with m particles. Then

$$E_m(\varepsilon_2^j | \mathcal{F}_1) = E_m\{(\#_1^j - mW_1^j) | \widetilde{X}_1^1, \dots, \widetilde{X}_1^m\} [\tilde{f}_1(\widetilde{X}_1^j) \widetilde{H}_1^j - \tilde{f}_0] = 0$$

since $E_m(\#_1^j | \mathcal{F}_1) = mW_1^j$. More generally, the conditional distribution of $(\mathbf{X}_t^1, \dots, \mathbf{X}_t^m)$ given $\mathcal{F}_{2t-1}$ is that of m i.i.d. random vectors which take the value $\widetilde{\mathbf{X}}_t^j$ with probability W_t^j. Moreover, the conditional distribution of $(\widetilde{X}_t^1, \dots, \widetilde{X}_t^m)$ given $\mathcal{F}_{2(t-1)}$ is that of m independent random variables such that $\widetilde{X}_t^i$ has density function $q_t(\cdot | \mathbf{X}_{t-1}^i)$ and, therefore, $E_m[\tilde{f}_t(\widetilde{\mathbf{X}}_t^i) | \mathcal{F}_{2(t-1)}] = \tilde{f}_{t-1}(\mathbf{X}_{t-1}^i)$ by (3.2) and the tower property of conditional expectations. Hence, in particular,

$$E_m(\varepsilon_3^j | \mathcal{F}_2) = \sum_{i:A_1^i=j} \{E_m[\tilde{f}_2(\widetilde{\mathbf{X}}_2^i) | X_1^i] - \tilde{f}_1(X_1^i)\} H_1^i = 0.$$

Proceeding inductively in this way shows that $\{\varepsilon_k^j, \mathcal{F}_k, 1 \le k \le 2T-1\}$ is a martingale difference sequence for all j. $\square$

Without tracing their ancestral origins as in (3.4), we can also use the successive generations of the m particles to form martingale differences directly and thereby obtain another martingale representation of $m(\tilde{\psi}_T - \psi_T)$. Specifically, the preceding argument also shows that $\{(Z_k^1, \dots, Z_k^m), \mathcal{F}_k, 1 \le k \le 2T-1\}$ is a martingale difference sequence, where

$$\begin{aligned} Z_{2t-1}^i &= [\tilde{f}_t(\widetilde{\mathbf{X}}_t^i) - \tilde{f}_{t-1}(\mathbf{X}_{t-1}^i)]H_{t-1}^i, \\ Z_{2t}^i &= \tilde{f}_t(\mathbf{X}_t^i)H_t^i - \sum_{j=1}^m W_t^j \tilde{f}_t(\widetilde{\mathbf{X}}_t^j)\widetilde{H}_t^j. \end{aligned} \tag{3.10}$$

Moreover, $Z_k^1, \dots, Z_k^m$ are conditionally independent given $\mathcal{F}_{k-1}$. It follows from (3.1), (3.6), (3.10) and an argument similar to (3.8) that

$$m(\tilde{\psi}_T - \psi_T) = \sum_{k=1}^{2T-1} (Z_k^1 + \cdots + Z_k^m). \tag{3.11}$$

This martingale representation yields the limiting normal distribution of $\sqrt{m}(\tilde{\psi}_T - \psi_T)$ for the basic prototype in the following.

COROLLARY 1. *Let* $\sigma_C^2 = \tilde{\sigma}_1^2 + \cdots + \tilde{\sigma}_{2T-1}^2$, *where*

$$\tilde{\sigma}_k^2 = \begin{cases} E_q\{[\tilde{f}_t^2(\mathbf{X}_t) - \tilde{f}_{t-1}^2(\mathbf{X}_{t-1})]h_{t-1}^*(\mathbf{X}_{t-1})\}, & \text{if } k = 2t-1, \\ E_q\{[\tilde{f}_t(\mathbf{X}_t)h_t^*(\mathbf{X}_t) - \tilde{f}_0]^2/h_t^*(\mathbf{X}_t)\}, & \text{if } k = 2t. \end{cases} \tag{3.12}$$

Assume that $\sigma_C^2 < \infty$. *Then as* $m \to \infty$,

$$\sqrt{m}(\tilde{\psi}_T - \psi_T) \Rightarrow N(0, \sigma_C^2). \tag{3.13}$$

The proof of Corollary 1 is given in the Appendix. The main result of this section is consistent estimation of σ_C^2 in the basic prototype. Define for every real number μ,

$$\begin{aligned} \tilde{\sigma}^2(\mu) = m^{-1} \sum_{j=1}^m \Bigg\{ & \sum_{i:A_{T-1}^i = j} L_T(\widetilde{\mathbf{X}}_T^i)\psi(\widetilde{\mathbf{X}}_T^i)H_{T-1}^i \\ & - \Bigg[1 + \sum_{t=1}^{T-1} \sum_{i:A_{t-1}^i = j} (\#_t^i - mW_t^i)\Bigg]\mu\Bigg\}^2, \end{aligned} \tag{3.14}$$

and note that by (3.4) and (3.8),

$$\tilde{\sigma}^2(\psi_T) = m^{-1} \sum_{j=1}^m (\varepsilon_1^j + \cdots + \varepsilon_{2T-1}^j)^2.$$

We next show that $\tilde{\sigma}^2(\psi_T) \xrightarrow{p} \sigma_C^2$ by making use of the following two lemmas which are proved in the Appendix. Since for every fixed T as $m \to \infty$, $\tilde{\psi}_T = \psi_T + O_p(1/\sqrt{m})$ by (3.13), it then follows that $\tilde{\sigma}^2(\tilde{\psi}_T)$ is also consistent for σ_C^2.

LEMMA 2. *Let $1 \le t \le T$ and let G be a measurable real-valued function on the state space. Define h_t^*, η_t and ζ_t as in* (2.5).

(i) *If $E_q[|G(\mathbf{X}_t)|/h_{t-1}^*(\mathbf{X}_{t-1})] < \infty$, then as $m \to \infty$,*

$$m^{-1}\sum_{i=1}^m G(\widetilde{\mathbf{X}}_t^i) \xrightarrow{p} E_q[G(\mathbf{X}_t)/h_{t-1}^*(\mathbf{X}_{t-1})]. \tag{3.15}$$

(ii) *If $E_q[|G(\mathbf{X}_t)|/h_t^*(\mathbf{X}_t)] < \infty$, then as $m \to \infty$,*

$$m^{-1}\sum_{i=1}^m G(\mathbf{X}_t^i) \xrightarrow{p} E_q[G(\mathbf{X}_t)/h_t^*(\mathbf{X}_t)]. \tag{3.16}$$

In particular, for the special case $G = w_t$, (3.15) *yields $\bar{w}_t \xrightarrow{p} \zeta_t^{-1}$ and, hence,*

$$\frac{\widetilde{H}_t^i}{h_t^*(\widetilde{\mathbf{X}}_t^i)} = \frac{H_t^i}{h_t^*(\mathbf{X}_t^i)} = \eta_t^{-1}\bar{w}_1\cdots\bar{w}_t \xrightarrow{p} 1 \qquad \text{as } m \to \infty. \tag{3.17}$$

Moreover, if $E_q[|G(\mathbf{X}_t)|/h_{t-1}^*(\mathbf{X}_{t-1})] < \infty$, then applying (3.15) *to $|G(\cdot)| \times \mathbf{1}_{\{|G(\cdot)|>M\}}$ and letting $M \to \infty$ yield*

$$m^{-1}\sum_{i=1}^m |G(\widetilde{\mathbf{X}}_t^i)|\mathbf{1}_{\{|G(\widetilde{\mathbf{X}}_t^i)|>\varepsilon\sqrt{m}\}} \xrightarrow{p} 0 \qquad \text{as } m \to \infty \text{ for every } \varepsilon > 0. \tag{3.18}$$

LEMMA 3. *Let $1 \le t \le T$ and let G be a measurable nonnegative valued function on the state space.*

(i) *If $E_q[G(\mathbf{X}_t)/h_{t-1}^*(\mathbf{X}_{t-1})] < \infty$, then as $m \to \infty$,*

$$m^{-1}\max_{1\le j\le m}\sum_{i:A_{t-1}^i=j} G(\widetilde{\mathbf{X}}_t^i) \xrightarrow{p} 0. \tag{3.19}$$

(ii) *If $E_q[G(\mathbf{X}_t)/h_t^*(\mathbf{X}_t)] < \infty$, then as $m \to \infty$,*

$$m^{-1}\max_{1\le j\le m}\sum_{i:A_t^i=j} G(\mathbf{X}_t^i) \xrightarrow{p} 0. \tag{3.20}$$

COROLLARY 2. *Suppose $\sigma_C^2 < \infty$. Then $\tilde{\sigma}^2(\psi_T) \xrightarrow{p} \sigma_C^2$ as $m \to \infty$.*

PROOF. By (3.4) and (3.8),

$$\sum_{k=1}^{2T-1}\varepsilon_k^j = \sum_{i:A_{T-1}^i=j} \tilde{f}_T(\widetilde{\mathbf{X}}_T^i)H_{T-1}^i - \left[1 + \sum_{t=1}^{T-1}\sum_{i:A_{t-1}^{(i)}=j}(\#_t^i - mW_t^i)\right]\tilde{f}_0.$$

We make use of Lemmas 2 and 3, (3.12) and mathematical induction to show that for all $1 \le k \le 2T-1$,

$$m^{-1}\sum_{j=1}^{m}(\varepsilon_1^j + \cdots + \varepsilon_k^j)^2 \xrightarrow{p} \sum_{\ell=1}^{k}\tilde{\sigma}_\ell^2. \tag{3.21}$$

By the weak law of large numbers, (3.21) holds for $k=1$. Next assume that (3.21) holds for $k=2t$ and consider the expansion

$$\begin{aligned} &m^{-1}\sum_{j=1}^{m}[(\varepsilon_1^j + \cdots + \varepsilon_{k+1}^j)^2 - (\varepsilon_1^j + \cdots + \varepsilon_k^j)^2] \\ &\quad = m^{-1}\sum_{j=1}^{m}(\varepsilon_{k+1}^j)^2 + 2m^{-1}\sum_{j=1}^{m}(\varepsilon_1^j + \cdots + \varepsilon_k^j)\varepsilon_{k+1}^j. \end{aligned} \tag{3.22}$$

Let $C_t^j = \{(i,\ell): i \ne \ell \text{ and } A_t^i = A_t^\ell = j\}$. Suppressing the subscript t, let $U_i = [\tilde{f}_{t+1}(\widetilde{\mathbf{X}}_{t+1}^i) - \tilde{f}_t(\mathbf{X}_t^i)]H_t^i$. Then $\varepsilon_{2t+1}^j = \sum_{i:A_t^i=j} U_i$ and

$$m^{-1}\sum_{j=1}^{m}(\varepsilon_{2t+1}^j)^2 = m^{-1}\sum_{i=1}^{m}U_i^2 + m^{-1}\sum_{j=1}^{m}\sum_{(i,\ell)\in C_t^j}U_iU_\ell. \tag{3.23}$$

By (3.12), (3.17) and Lemma 2(i),

$$m^{-1}\sum_{i=1}^{m}U_i^2 \xrightarrow{p} \tilde{\sigma}_{2t+1}^2. \tag{3.24}$$

Since $U_1, \ldots, U_m$ are independent mean zero random variables conditioned on $\mathcal{F}_{2t}$, it follows from (3.17), Lemmas 2(ii) and 3(ii) that

$$\begin{aligned} &m^{-2}\operatorname{Var}_m\Bigg(\sum_{j=1}^{m}\sum_{(i,\ell)\in C_t^j}U_iU_\ell\Bigg|\mathcal{F}_{2t}\Bigg) \\ &\quad = m^{-2}\sum_{j=1}^{m}\sum_{(i,\ell)\in C_t^j}E_m(U_i^2U_\ell^2|\mathcal{F}_{2t}) \le m^{-2}\sum_{j=1}^{m}\Bigg\{\sum_{i:A_t^i=j}E_m(U_i^2|\mathcal{F}_{2t})\Bigg\}^2 \\ &\quad \le \Bigg[m^{-1}\sum_{i=1}^{m}E_m(U_i^2|\mathcal{F}_{2t})\Bigg]\Bigg[m^{-1}\max_{1\le j\le m}\sum_{i:A_t^i=j}E_m(U_i^2|\mathcal{F}_{2t})\Bigg] \xrightarrow{p} 0. \end{aligned} \tag{3.25}$$

By (3.23)–(3.25),

$$m^{-1}\sum_{j=1}^{m}(\varepsilon_{2t+1}^j)^2 \xrightarrow{p} \tilde{\sigma}_{2t+1}^2. \tag{3.26}$$

We next show that

$$m^{-1}\sum_{j=1}^{m}(\varepsilon_1^j+\cdots+\varepsilon_{2t}^j)\varepsilon_{2t+1}^j \xrightarrow{p} 0. \tag{3.27}$$

Since $\varepsilon_1^j,\dots,\varepsilon_{2t}^j$ are measurable with respect to $\mathcal{F}_{2t}$ for $1\le j\le m$ and $\varepsilon_{2t+1}^1,\dots,\varepsilon_{2t+1}^m$ independent conditioned on $\mathcal{F}_{2t}$, by the induction hypothesis and by (3.17) and Lemma 3(ii), it follows that

$$\begin{aligned}
&m^{-2}\sum_{j=1}^{m}\mathrm{Var}_m\big((\varepsilon_1^j+\cdots+\varepsilon_{2t}^j)\varepsilon_{2t+1}^j|\mathcal{F}_{2t}\big)\\
&\quad=m^{-2}\sum_{j=1}^{m}\Big\{(\varepsilon_1^j+\cdots+\varepsilon_{2t}^j)^2\sum_{i:A_t^i=j}E_m(U_i^2|\mathcal{F}_{2t})\Big\}\\
&\quad\le\Big[m^{-1}\sum_{j=1}^{m}(\varepsilon_1^j+\cdots+\varepsilon_{2t}^j)^2\Big]\Big[m^{-1}\max_{1\le j\le m}\sum_{i:A_t^i=j}E_m(U_i^2|\mathcal{F}_{2t})\Big]\xrightarrow{p}0,
\end{aligned}$$

and therefore (3.27) holds. By (3.22), (3.26), (3.27) and the induction hypothesis, (3.21) holds for $k=2t+1$.

Next assume (3.21) holds for $k=2t-1$. Suppressing the subscript t, let $S_i=(\#_t^i-mW_t^i)[\tilde f_t(\widetilde{\mathbf{X}}_t^i)\widetilde H_t^i-\tilde f_0]$. Then

$$m^{-1}\sum_{j=1}^{m}(\varepsilon_{2t}^j)^2=m^{-1}\sum_{i=1}^{m}S_i^2+m^{-1}\sum_{j=1}^{m}\sum_{(i,\ell)\in C_{t-1}^j}S_iS_\ell. \tag{3.28}$$

From Lemma 2(i), (3.12) and (3.17), it follows that

$$\begin{aligned}
&m^{-1}\sum_{i=1}^{m}E_m(S_i^2|\mathcal{F}_{2t-1})\\
&\quad=m^{-1}\sum_{i=1}^{m}\mathrm{Var}_m(\#_t^i|\mathcal{F}_{2t-1})[\tilde f_t(\widetilde{\mathbf{X}}_t^i)\widetilde H_t^i-\tilde f_0]^2\\
&\quad=m^{-1}\sum_{i=1}^{m}\frac{w_t(\widetilde{\mathbf{X}}_t^i)}{\bar w_t}\Big(1-\frac{w_t(\widetilde{\mathbf{X}}_t^i)}{m\bar w_t}\Big)[\tilde f_t(\widetilde{\mathbf{X}}_t^i)\widetilde H_t^i-\tilde f_0]^2\xrightarrow{p}\tilde\sigma_{2t}^2.
\end{aligned} \tag{3.29}$$

Technical arguments given in the Appendix show that

$$m^{-2}\,\mathrm{Var}_m\Big(\sum_{i=1}^{m}S_i^2\Big|\mathcal{F}_{2t-1}\Big)\xrightarrow{p}0, \tag{3.30}$$

$$m^{-2}E_m\Big[\Big(\sum_{j=1}^{m}\sum_{(i,\ell)\in C_{t-1}^j}S_iS_\ell\Big)^2\Big|\mathcal{F}_{2t-1}\Big]\xrightarrow{p}0 \tag{3.31}$$

and

$$m^{-2} E_m \left[\left\{ \sum_{j=1}^{m} (\varepsilon_1^j + \cdots + \varepsilon_{2t-1}^j) \varepsilon_{2t}^j \right\}^2 \middle| \mathcal{F}_{2t-1} \right] \xrightarrow{p} 0 \tag{3.32}$$

under (3.21) for $k = 2t - 1$. By (3.22) and (3.28)–(3.32), (3.21) holds for $k = 2t$. The induction proof is complete and Corollary 2 holds. $\square$

Let $\Psi(\mathbf{x}_T) = \psi(\mathbf{x}_T) - \psi_T$ and define $\widetilde{\Psi}_T$ as in (3.1) with Ψ in place of ψ. Then replacing $\tilde{\psi}_T$ by $\widetilde{\Psi}_T$ in Corollaries 1 and 2 shows that as $m \to \infty$,

$$\sqrt{m}\widetilde{\Psi}_T \Rightarrow N(0, \sigma^2), \qquad \tilde{\sigma}_*^2(0) \xrightarrow{p} \sigma^2, \tag{3.33}$$

where $\tilde{\sigma}_*^2(\mu)$ is defined by (3.14) with ψ replaced by Ψ, noting that replacing ψ by Ψ in the definition of σ_C^2 in Corollary 1 gives σ^2 defined in (2.8) as $f_0 = 0$.

3.2. *Approximation of $\hat{\psi}_T - \psi_T$ by $\widetilde{\Psi}_T$ and proof of Theorem 1.* We now return to the setting of Section 2.1, in which we consider the HMM (1.1) and use weight functions of the form (2.1) and the particle filter (2.4) in lieu of (3.1). We make use of the following lemma to extend (3.33) to $\hat{\psi}_T - \psi_T$.

LEMMA 4. *Let w_t be of the form (2.1) for $1 \le t \le T$ and define η_T and ζ_T by (2.5) and $\widetilde{\Psi}_T$ by (3.1) with ψ replaced by $\Psi = \psi - \psi_T$. Then*

$$\tilde{p}_T(\mathbf{x}_T) [= \tilde{p}_T(\mathbf{x}_T | \mathbf{Y}_T)] = \eta_T^{-1} \prod_{t=1}^{T} [p_t(x_t | x_{t-1}) g_t(Y_t | x_t)], \tag{3.34}$$

$$L_T(\mathbf{x}_T) = \eta_T^{-1} \prod_{t=1}^{T} w_t(\mathbf{x}_t), \tag{3.35}$$

$$\hat{\psi}_T - \psi_T = (m\bar{w}_T)^{-1} \sum_{i=1}^{m} \Psi(\widetilde{\mathbf{X}}_T^i) w_T(\widetilde{\mathbf{X}}_T^i) = (\bar{w}_1 \cdots \bar{w}_T)^{-1} \eta_T \widetilde{\Psi}_T. \tag{3.36}$$

PROOF. By (2.1) and (2.5),

$$\eta_T = \int \prod_{t=1}^{T} [w_t(\mathbf{x}_t) q_t(x_t | \mathbf{x}_{t-1})] \, d\nu_X(\mathbf{x}_T) = \int \prod_{t=1}^{T} [p_t(x_t | x_{t-1}) g_t(Y_t | x_t)] \, d\nu_X(\mathbf{x}_T),$$

and (3.34) follows from (1.2). Moreover, (3.35) follows from (2.1), (2.6) and (3.34). The first equality in (3.36) follows from (2.3) and (2.4). By (2.3) and (3.35),

$$L_T(\widetilde{\mathbf{X}}_T^i) H_{T-1}^i = \left(\frac{\bar{w}_1 \cdots \bar{w}_T}{\eta_T} \right) \frac{w_T(\widetilde{\mathbf{X}}_T^i)}{\bar{w}_T}. \tag{3.37}$$

Hence, the second equality in (3.36) follows from (3.1). $\square$

The following lemma, proved in the Appendix, is used to prove Theorem 1. Although it resembles Lemma 3, its conclusions are about the square of the sums in Lemma 3 and, accordingly, it assumes finiteness of second (instead of first) moments.

LEMMA 5. *Let $1 \le t \le T$, G be a measurable function on the state space, and $\Gamma_t(\mathbf{x}_t)$ be defined as in* (2.5).

(i) *If $E_q[G^2(\mathbf{X}_t)\Gamma_{t-1}(\mathbf{X}_{t-1})] < \infty$, then*

$$m^{-1}\sum_{j=1}^{m}\biggl[\sum_{i:A_{t-1}^i=j} G(\widetilde{\mathbf{X}}_t^i)\biggr]^2 = O_p(1) \qquad \text{as } m \to \infty.$$

(ii) *If $E_q[G^2(\mathbf{X}_t)\Gamma_t(\mathbf{X}_t)] < \infty$, then*

$$m^{-1}\sum_{j=1}^{m}\biggl[\sum_{i:A_t^i=j} G(\mathbf{X}_t^i)\biggr]^2 = O_p(1) \qquad \text{as } m \to \infty.$$

PROOF OF THEOREM 1. By (3.17), (3.33) and (3.36),

$$\sqrt{m}(\hat{\psi}_T - \psi_T) = \bigl(1 + o_p(1)\bigr)\sqrt{m}\widetilde{\Psi}_T \quad \Rightarrow \quad N\bigl(0, \sigma^2\bigr). \tag{3.38}$$

Since $\tilde{\sigma}_*^2(0) = m^{-1}\sum_{j=1}^m\{\sum_{i:A_{T-1}^i=j} L_T(\widetilde{\mathbf{X}}_T^i)H_{T-1}^i\Psi(\widetilde{\mathbf{X}}_T^i)\}^2 \xrightarrow{p} \sigma^2$ by (3.14) and (3.33), and since $\tilde{\sigma}_*^2(0) = (1 + o_p(1))\hat{\sigma}^2(\psi_T)$ in view of (2.9), (3.17) and (3.37),

$$\hat{\sigma}^2(\psi_T) \xrightarrow{p} \sigma^2. \tag{3.39}$$

Letting $c_j = \sum_{i:A_{T-1}^i=j}\frac{w_T(\widetilde{\mathbf{X}}_T^i)}{\bar{w}_T}[\psi(\widetilde{\mathbf{X}}_T^i) - \psi_T]$, $b_j = \sum_{i:A_{T-1}^i=j} w_T(\widetilde{\mathbf{X}}_T^i)/\bar{w}_T$ and $a = \hat{\psi}_T - \psi_T$, it follows from (2.9) that

$$\hat{\sigma}^2(\hat{\psi}_T) = \frac{1}{m}\sum_{j=1}^m (c_j + ab_j)^2 = \frac{1}{m}\sum_{j=1}^m c_j^2 + \frac{2a}{m}\sum_{j=1}^m b_j c_j + \frac{a^2}{m}\sum_{j=1}^m b_j^2. \tag{3.40}$$

Since $a = O_p(m^{-1/2})$ by (3.38), $m^{-1}\sum_{j=1}^m c_j^2 = \hat{\sigma}^2(\psi_T) \xrightarrow{p} \sigma^2$ by (3.39), and since $m^{-1}\sum_{j=1}^m b_j^2 = O_p(1)$ by Lemma 5 (with $G = w_t$),

$$\left|\frac{2a}{m}\sum_{j=1}^m b_j c_j\right| \le |a|\left(m^{-1}\sum_{j=1}^m b_j^2 + m^{-1}\sum_{j=1}^m c_j^2\right) \xrightarrow{p} 0.$$

Hence, by (3.40), $\hat{\sigma}^2(\hat{\psi}_T) \xrightarrow{p} \sigma^2$. $\square$

4. Extensions and proof of Theorem 2. In this section we first extend the results of Section 2.1 to the case where residual Bernoulli resampling is used in lieu of bootstrap resampling at every stage. We then consider occasional bootstrap resampling and prove Theorem 2, which we also extend to the case of occasional residual Bernoulli resampling.

4.1. *Residual Bernoulli resampling.* The *residual resampling* scheme introduced in [2, 3] often leads to smaller asymptotic variance than that of bootstrap resampling; see [6, 14]. We consider here the residual Bernoulli scheme given in [7], which has been shown in Section 2.4 of [9] to yield a consistent and asymptotically normal particle filtering estimate of ψ_T. To implement residual Bernoulli resampling, we modify bootstrap resampling at stage t as follows: let $M(1) = m$ and let $\xi_t^1, \dots, \xi_t^{M(t)}$ be independent Bernoulli random variables conditioned on $(M(t), W_t^i)$ satisfying $P\{\xi_t^i = 1 | \mathcal{F}_{2t-1}\} = M(t)W_t^i - \lfloor M(t)W_t^i \rfloor$. For each $1 \le i \le M(t)$ and $t \le T - 1$, let $\widetilde{H}_t^i = H_{t-1}^i / (M(t)W_t^i)$ and make $\#_t^i := \lfloor M(t)W_t^i \rfloor + \xi_t^i$ copies of $(\widetilde{\mathbf{X}}_t^i, A_{t-1}^i, \widetilde{H}_t^i)$. These copies constitute an augmented sample $\{(\mathbf{X}_t^j, A_t^j, H_t^j) : 1 \le j \le M(t+1)\}$, where $M(t+1) = \sum_{i=1}^{M(t)} \#_t^i$. Estimate ψ_T by

$$\hat{\psi}_{T,\mathrm{R}} := [M(T)\bar{w}_T]^{-1} \sum_{i=1}^{M(T)} \psi(\widetilde{\mathbf{X}}_T^i) w_T(\widetilde{\mathbf{X}}_T^i).$$

Let $\widetilde{\Psi}_{T,\mathrm{R}}$ be (3.1) with $\Psi = \psi - \psi_T$ in place of ψ and $M(T)$ in place of m. Since H_{t-1}^i is still given by (2.3) and Lemma 2 still holds for residual Bernoulli resampling, it follows from (3.17) and (3.37) that

$$\begin{aligned} M(T)(\hat{\psi}_{T,\mathrm{R}} - \psi_T) &= \bar{w}_T^{-1} \sum_{i=1}^{M(T)} \Psi(\widetilde{\mathbf{X}}_T^i) w_T(\widetilde{\mathbf{X}}_T^i) \\ &= \left(\frac{\eta_T}{\bar{w}_1 \cdots \bar{w}_T}\right) \sum_{i=1}^{M(T)} \Psi(\widetilde{\mathbf{X}}_T^i) L_T(\widetilde{\mathbf{X}}_T^i) h_{t-1}(\mathbf{X}_{t-1}^i) \qquad (4.1) \\ &= (1 + o_p(1)) M(T) \widetilde{\Psi}_{T,\mathrm{R}}. \end{aligned}$$

Because replacing ψ by Ψ in (3.2) modifies $\tilde{f}_t$ to f_t as defined in (2.7), we have for residual Bernoulli resampling the following analog of (3.11):

$$M(T)\widetilde{\Psi}_{T,\mathrm{R}} = \sum_{k=1}^{2T-1} (Z_{k,\mathrm{R}}^1 + \cdots + Z_{k,\mathrm{R}}^{M(\lceil k/2 \rceil)}), \qquad (4.2)$$

where $Z_{2t-1,\mathrm{R}}^i = [f_t(\widetilde{\mathbf{X}}_t^i) - f_{t-1}(\mathbf{X}_{t-1}^i)] H_{t-1}^i$, $Z_{2t,\mathrm{R}}^i = (\#_t^i - M(t)W_t^i) f_t(\widetilde{\mathbf{X}}_t^i) \widetilde{H}_t^i$.

Let $\sigma_{\mathrm{R}}^2 = \sum_{k=1}^{2T-1} \sigma_{k,\mathrm{R}}^2$, with

$$\sigma_{2t-1,\mathrm{R}}^2 = \sigma_{2t-1}^2, \qquad \sigma_{2t,\mathrm{R}}^2 = E_q[\gamma(w_t^*(\mathbf{X}_t)) f_t^2(\mathbf{X}_t) h_t^*(\mathbf{X}_t)],$$

in which $\gamma(x) = (x - \lfloor x \rfloor)(1 - x + \lfloor x \rfloor)/x$, $w_t^*(\mathbf{x}_t) = h_{t-1}^*(\mathbf{x}_{t-1})/h_t^*(\mathbf{x}_t) = \zeta_t w_t(\mathbf{x}_t)$ and σ_{2t-1}^2 is defined in Theorem 1. Since $\mathrm{Var}(\xi) = \gamma(x)x$ when ξ is a Bernoulli$(x - \lfloor x \rfloor)$ random variable,

$$\mathrm{Var}\left(\sum_{i=1}^{M(t)} Z_{2t,\mathrm{R}}^i \middle| \mathcal{F}_{2t-1}\right) = \sum_{i=1}^{M(t)} \gamma\big(M(t)W_t^i\big)M(t)W_t^i f_t^2(\widetilde{\mathbf{X}}_t^i)(\widetilde{H}_t^i)^2$$

and, therefore, it follows from Lemma 2 that

$$m^{-1}\,\mathrm{Var}\left(\sum_{i=1}^{M(\lceil k/2 \rceil)} Z_{k,\mathrm{R}}^i \middle| \mathcal{F}_{k-1}\right) \xrightarrow{p} \sigma_{k,\mathrm{R}}^2, \qquad 1 \le k \le 2T - 1. \tag{4.3}$$

By (4.1)–(4.3) and a similar modification of (3.3) for residual Bernoulli resampling, Theorem 1 still holds with $\hat{\psi}_T$ replaced by $\hat{\psi}_{T,\mathrm{R}}$; specifically,

$$\sqrt{m}(\hat{\psi}_{T,\mathrm{R}} - \psi_T) \Rightarrow N(0, \sigma_{\mathrm{R}}^2), \qquad \hat{\sigma}^2(\hat{\psi}_{T,\mathrm{R}}) \xrightarrow{p} \sigma_{\mathrm{R}}^2, \tag{4.4}$$

where $\hat{\sigma}^2(\mu)$ is defined in (2.9). The following example shows the advantage of residual resampling over bootstrap resampling.

EXAMPLE. Consider the bearings-only tracking problem in [11], in which $X_t = (X_{t1}, \ldots, X_{t4})'$ and (X_{t1}, X_{t3}) represents the coordinates of a ship and (X_{t2}, X_{t4}) the velocity at time t. A person standing at the origin observes that the ship is at angle Y_t relative to him. Taking into account measurement errors and random disturbances in velocity of the ship, and letting Φ be a 4×4 matrix with $\Phi_{12} = \Phi_{34} = 1 = \Phi_{ii}$ for $1 \le i \le 4$ and all other entries 0, Γ be a 4×2 matrix with $\Gamma_{11} = \Gamma_{32} = 0.5$, $\Gamma_{21} = \Gamma_{42} = 1$ and all other entries 0, the dynamics can be described by state-space model

$$X_t = \Phi X_{t-1} + \Gamma z_t, \qquad Y_t = \tan^{-1}(X_{t3}/X_{t1}) + u_t,$$

where $X_{11} \sim N(0, 0.5^2)$, $X_{12} \sim N(0, 0.005^2)$, $X_{13} \sim N(0.4, 0.3^2)$, $X_{14} \sim N(-0.05, 0.01^2)$, $z_{t+1} \sim N(0, 0.001^2\mathbf{I}_2)$ and $u_t \sim N(0, 0.005^2)$ are independent random variables for $t \ge 1$. The quantities of interest are $E(X_{T1}|\mathbf{Y}_T)$ and $E(X_{T3}|\mathbf{Y}_T)$. To implement the particle filters, we let $q_t = p_t$ for $t \ge 2$. Unlike [11] in which $q_1 = p_1$ as well, we use an importance density q_1 that involves p_1 and Y_1 to generate the initial location (X_{11}, X_{13}). Let ξ and ζ be independent standard normal random variables, and $r = \tan(Y_1 + 0.005\xi)$. Since u_1 has small variance, we can estimate X_{13}/X_{11} well by r. This suggests choosing q_1 to be degenerate bivariate normal with support $y = rx$, with (y, x) denoting (X_{13}, X_{11}), so that its density function on this line is proportional to $\exp\{-(x-\mu)^2/(2\tau)\}$, where $\mu = 0.4r/(0.36 + r^2)$, $\tau = 0.09/(0.36 + r^2)$. Thus, q_1 generates $X_{11} = \mu + \sqrt{\tau}\zeta$ and $X_{13} = rX_{11}$, but still follows p_1 to generate (X_{12}, X_{14}). By (2.1),

$$w_1(x_1) = |x_{11}|\sqrt{\tau}(1 + r^2)\exp\left\{-\frac{x_{11}^2}{2(0.5)^2} - \frac{(x_{13} - 0.4)^2}{2(0.3)^2} + \frac{\zeta^2}{2}\right\},$$

TABLE 3
Simulated values of $\hat{\psi}_T \pm \hat{\sigma}(\hat{\psi}_T)/\sqrt{m}$ (for bootstrap resampling) and $\hat{\psi}_{T,\mathrm{R}} \pm \hat{\sigma}(\hat{\psi}_{T,\mathrm{R}})/\sqrt{m}$ (for residual Bernoulli resampling)

	$E(X_{T1}\mid\mathbf{Y}_T)$			$E(X_{T3}\mid\mathbf{Y}_T)$		
T	**boot(P)**	**boot**	**resid**	**boot(P)**	**boot**	**resid**
4	−0.515	−0.536	−0.519	0.091	0.098	0.095
	±0.026	±0.009	±0.007	±0.007	±0.002	±0.001
8	−0.505	−0.515	−0.497	−0.118	−0.121	−0.117
	±0.043	±0.012	±0.011	±0.008	±0.003	±0.003
12	−0.503	−0.532	−0.508	−0.327	−0.346	−0.331
	±0.042	±0.014	±0.010	±0.027	±0.009	±0.007
16	−0.506	−0.538	−0.516	−0.538	−0.572	−0.549
	±0.044	±0.015	±0.010	±0.046	±0.016	±0.011
20	−0.512	−0.537	−0.532	−0.747	−0.783	−0.777
	±0.045	±0.016	±0.011	±0.066	±0.023	±0.017
24	−0.512	−0.551	−0.540	−0.950	−1.023	−1.002
	±0.047	±0.016	±0.012	±0.088	±0.029	±0.022

noting that $0.005|x_{11}|\sqrt{\tau}(1+r^2)$ is the Jacobian of the transformation of (ξ,ζ) to (x_{11},x_{13}). For $t\ge 2$, (2.1) yields the resampling weights $w_t(\mathbf{x}_t)=\exp\{-[Y_t-\tan^{-1}(x_{t3}/x_{t1})]^2/[2(0.005)^2]\}$. We use $m=10{,}000$ particles to estimate $E((X_{T1},X_{T3})|\mathbf{Y}_T)$ by particle filters, using bootstrap (boot) or residual (resid) Bernoulli resampling at every stage, for different values of T. Table 3, which reports results based on a single realization of $\mathbf{Y}_T$, shows that residual Bernoulli resampling has smaller standard errors than bootstrap resampling. Table 3 also considers bootstrap resampling that uses $q_t=p_t$ for $t\ge 1$, as in [11] and denoted by boot(P), again with $m=10{,}000$ particles. Although boot and boot(P) differ only in the choice of the importance densities at $t=1$, the standard error of boot is substantially smaller than that of boot(P) over the entire range $4\le T\le 24$. Unlike Section 2.3, the standard error estimates in Table 3 use a sample splitting refinement that is described in the first paragraph of Section 5.

4.2. *Proof of Theorem* 2. First consider the following modification of $\widetilde{\Psi}_T$:

$$\widetilde{\Psi}_{\mathrm{OR}} := m^{-1}\sum_{i=1}^{m} L_T(\widetilde{\mathbf{X}}_T^i)\Psi(\widetilde{\mathbf{X}}_T^i)H^i_{\tilde{\tau}(T)} = \eta_T^{-1}\bar{v}_{\tau_1}\cdots\bar{v}_{\tau_{r+1}}(\hat{\psi}_{\mathrm{OR}}-\psi_T). \tag{4.5}$$

Analogous to (3.11), $m\widetilde{\Psi}_{\mathrm{OR}}=\sum_{k=1}^{2r+1}(Z_k^1+\cdots+Z_k^m)$, where

$$Z^i_{2s-1}=[f_{\tau_s}(\widetilde{\mathbf{X}}^i_{\tau_s})-f_{\tau_{s-1}}(\mathbf{X}^i_{\tau_{s-1}})]H^i_{\tau_{s-1}},$$

$$Z^i_{2s}=f_{\tau_s}(\widetilde{\mathbf{X}}^i_{\tau_s})\widetilde{H}^i_{\tau_s}-\sum_{j=1}^{m}V^j_{\tau_s}f(\widetilde{\mathbf{X}}^j_{\tau_s})\widetilde{H}^j_{\tau_s}.$$

Let $\tau_0^* = 0$. Since $\bar{v}_{\tau_s} \xrightarrow{p} \prod_{t=\tau_{s-1}^*+1}^{\tau_s^*} \zeta_t$ by (2.12) and Lemma 2, it follows from (2.5) and (4.5) that

$$\sqrt{m}(\hat{\psi}_{\mathrm{OR}} - \psi_T) = \big(1 + o_p(1)\big)\sqrt{m}\widetilde{\Psi}_{\mathrm{OR}} \quad \Rightarrow \quad N\big(0, \sigma_{\mathrm{OR}}^2\big),$$

similarly to (3.38), where

$$\begin{aligned}\sigma_{\mathrm{OR}}^2 = {} & \sum_{s=1}^{r^*+1} E_q\{[f_{\tau_s^*}^2(\mathbf{X}_{\tau_s^*}) - f_{\tau_{s-1}^*}^2(\mathbf{X}_{\tau_{s-1}^*})]h_{\tau_{s-1}^*}^*(\mathbf{X}_{\tau_{s-1}^*})\} \\ & + \sum_{s=1}^{r^*} E_q[f_{\tau_s^*}^2(\mathbf{X}_{\tau_s^*})h_{\tau_s^*}^*(\mathbf{X}_{\tau_s^*})].\end{aligned} \tag{4.6}$$

Moreover, analogous to (3.3), $m\widetilde{\Psi}_{\mathrm{OR}} = \sum_{j=1}^m (\varepsilon_1^j + \cdots + \varepsilon_{2r+1}^j)$, where

$$\varepsilon_{2s-1}^j = \sum_{i\,:\, A_{\tau_{s-1}}^i = j} [f_{\tau_s}(\widetilde{\mathbf{X}}_{\tau_s}^i) - f_{\tau_{s-1}}(\mathbf{X}_{\tau_{s-1}}^i)]H_{\tau_{s-1}}^i,$$

$$\varepsilon_{2s}^j = \sum_{i\,:\, A_{\tau_{s-1}}^i = j} (\#_{\tau_s}^i - mV_{\tau_s}^i) f_{\tau_s}(\widetilde{\mathbf{X}}_{\tau_s}^i)\widetilde{H}_{\tau_s}^i,$$

therefore, the proof of $\hat{\sigma}_{\mathrm{OR}}^2(\hat{\psi}_{\mathrm{OR}}) \xrightarrow{p} \sigma_{\mathrm{OR}}^2$ is similar to that of Theorem 1.

4.3. *Occasional residual resampling and assumption* (2.12). In the case of occasional residual resampling, Theorem 2 still holds with $\hat{\psi}_{\mathrm{OR}}$ replaced by $\hat{\psi}_{\mathrm{ORR}}$ and with σ_{OR}^2 replaced by

$$\begin{aligned}\sigma_{\mathrm{ORR}}^2 = {} & \sum_{s=1}^{r^*+1} E_q\{[f_{\tau_s^*}^2(\mathbf{X}_{\tau_s^*}) - f_{\tau_{s-1}^*}^2(\mathbf{X}_{\tau_{s-1}^*})]h_{\tau_{s-1}^*}^*(\mathbf{X}_{\tau_{s-1}^*})\} \\ & + \sum_{s=1}^{r^*} E_q\left[\gamma\left(\prod_{t=\tau_{s-1}^*+1}^{\tau_s^*} w_t^*(\mathbf{X}_t)\right) f_{\tau_s^*}^2(\mathbf{X}_{\tau_s^*})h_{\tau_s^*}^*(\mathbf{X}_{\tau_s^*})\right].\end{aligned}$$

In particular, $\sqrt{m}(\hat{\psi}_{\mathrm{ORR}} - \psi_T) \Rightarrow N(0, \sigma_{\mathrm{ORR}}^2)$ and $\hat{\sigma}_{\mathrm{OR}}^2(\hat{\psi}_{\mathrm{ORR}}) \xrightarrow{p} \sigma_{\mathrm{ORR}}^2$.

Assumption (2.12) is often satisfied in practice. In particular, if one follows Liu [13] and resamples at stage t whenever

$$\mathrm{cv}_t^2 := m^{-1}\sum_{i=1}^m (mV_t^i)^2 - 1 = \frac{\sum_{i=1}^m [v_t(\widetilde{\mathbf{X}}_t^i) - \bar{v}_t]^2}{m\bar{v}_t^2} \ge c, \tag{4.7}$$

where $c > 0$ is a prespecified threshold for the coefficient of variation, then (2.12) can be shown to hold by making use of the following lemma.

LEMMA 6. *Let resampling be carried out at stage t whenever* $\mathrm{cv}_t^2 \ge c$. *Then* (2.12) *holds with*

$$\tau_s^* = \inf\left\{t > \tau_{s-1}^* : E_q\left(\frac{[\prod_{k=\tau_{s-1}^*+1}^{t} w_k^*(\mathbf{X}_k)]^2}{h_{\tau_{s-1}^*}^*(\mathbf{X}_{\tau_{s-1}^*})}\right) - 1 \ge c\right\} \quad \text{for } 1 \le s \le r^*, \tag{4.8}$$

where $\tau_0^* = 0$ *and* $r^* = \max\{s : \tau_s^* < T\}$, *provided that*

$$E_q\left(\frac{[\prod_{k=\tau_{s-1}^*+1}^{\tau_s^*} w_k^*(\mathbf{X}_k)]^2}{h_{\tau_{s-1}^*}^*(\mathbf{X}_{\tau_{s-1}^*})}\right) - 1 \ne c \qquad \text{for } 1 \le s \le r^*. \tag{4.9}$$

PROOF. Let $\tau_{s-1}^* + 1 \le \ell \le \tau_s^*$ and $G(\mathbf{x}_\ell) = \prod_{k=\tau_{s-1}^*+1}^{\ell} w_k^*(\mathbf{x}_k)$. Apply Lemma 2(i) to G and G^2, with $t-1$ in the subscript replaced by the most recent resampling time τ_{s-1}^*. It then follows from (4.7) that

$$\mathrm{cv}_\ell^2 \xrightarrow{p} E_q\big[G^2(\mathbf{X}_\ell)/h_{\tau_{s-1}^*}^*(\mathbf{X}_{\tau_{s-1}^*})\big] - 1 \qquad \text{on } \{\tau_{s-1} = \tau_{s-1}^*\}. \tag{4.10}$$

In view of (4.8) and (4.9), it follows from (4.10) with $\ell = \tau_s^*$ that $P_m\{\tau_s > \tau_s^*, \tau_{s-1} = \tau_{s-1}^*\} \xrightarrow{p} 0$, for $1 \le s \le r^*$, and from (4.10) with $\tau_{s-1}^* + 1 \le \ell < \tau_s^*$ that $P_m\{\tau_s = \ell, \tau_{s-1} = \tau_{s-1}^*\} \xrightarrow{p} 0$. □

5. Discussion and concluding remarks. The central limit theorem for particle filters in this paper and in [6, 8, 9, 12] considers the case of fixed T as the number m of particles becomes infinite. In addition to the standard error estimates (2.9), one can use the following modification that substitutes ψ_T in $\hat{\sigma}^2(\psi_T)$ by "out-of-sample" estimates of ψ_T. Instead of generating additional particles to accomplish this, we apply a sample-splitting technique to the m particles, which is similar to k-fold cross-validation. The standard error estimates in the example in Section 4.1 use this technique with $k = 2$. Divide the m particles into k groups of equal size $r = \lfloor m/k \rfloor$ except for the last one that may have a larger sample size. For definiteness consider bootstrap resampling at every stage; the case of residual resampling or occasional resampling can be treated similarly. Denote the particles at the tth generation, before resampling, by $\{\widetilde{\mathbf{X}}_t^{ij} : 1 \le i \le r, 1 \le j \le k\}$, and let $\mathbf{X}_t^{1j}, \dots, \mathbf{X}_t^{rj}$ be sampled with replacement from $\{\widetilde{\mathbf{X}}_t^{1j}, \dots, \widetilde{\mathbf{X}}_t^{rj}\}$, with weights $W_t^{ij} = w_t(\widetilde{\mathbf{X}}_t^{ij})/(m\bar{w}_t^j)$, where $\bar{w}_t^j = r^{-1}\sum_{i=1}^r w_t(\widetilde{\mathbf{X}}_t^{ij})$. With these "stratified" resampling weights, we estimate ψ_T by

$$\hat{\psi}_T = k^{-1}\sum_{j=1}^k \hat{\psi}_T^j \qquad \text{where } \hat{\psi}_T^j = (r\bar{w}_T^j)^{-1}\sum_{i=1}^r \psi(\widetilde{\mathbf{X}}_T^{ij}) w_T(\widetilde{\mathbf{X}}_T^{ij}). \tag{5.1}$$

Similarly, letting $\hat{\psi}_T^{-j} = (k-1)^{-1} \sum_{\ell:\ell \neq j} \hat{\psi}_T^{\ell}$, we estimate σ^2 by

$$\hat{\sigma}^2_{\mathrm{SP}} = m^{-1} \sum_{j=1}^{k} \sum_{i=1}^{r} \left\{ \sum_{\ell : A_{T-1}^{\ell j} = i} \frac{w_T(\widetilde{\mathbf{X}}_T^{\ell j})}{\bar{w}_T^j} [\psi(\widetilde{\mathbf{X}}_T^{\ell j}) - \hat{\psi}_T^{-j}] \right\}^2. \tag{5.2}$$

Chopin (Section 2.1 of [6]) summarizes a general framework of traditional particle filter estimates of ψ_T, in which the resampling (also called "selection") weights $w_t(\mathbf{X}_t)$ are chosen to convert a weighted approximation of the posterior density of $\mathbf{X}_t$ given $\mathbf{Y}_t$, with likelihood ratio weights associated with importance sampling (called "mutation" in [6]), to an unweighted approximation (that has equal weights 1), so that the usual average $\bar{\psi}_T := m^{-1} \sum_{i=1}^m \psi(\mathbf{X}_T^i)$ converges a.s. to ψ_T as $m \to \infty$. He uses induction on $t \le T$ to prove the central limit theorem for $\sqrt{m}(\bar{\psi}_T - \psi_T)$: "conditional on past iterations, each step generates independent (but not identically distributed) particles, which follow some (conditional) central limit theorem" (page 2392 of [6]), and he points out that the particle filter of Gilks and Berzuini [10] is a variant of this framework.

Let V_t be the variance of the limiting normal distribution of $\sqrt{m_t}(\bar{\psi}_T - \psi_T)$ in the Gilks–Berzuini particle filter (pages 132–134 of [10]), that assumes

$$m_1 \to \infty, \qquad \text{then } m_2 \to \infty, \ldots \text{ and then } m_T \to \infty, \tag{5.3}$$

which means that $m_1 \to \infty$ and $m_t - m_{t-1} \to \infty$ for $1 < t \le T$. The estimates of V_t proposed in [10] use the idea of "ancestor" instead of the ancestral origin we use. For $s < t$, particle $\widetilde{\mathbf{X}}_s^i$ is called an ancestor of $\widetilde{\mathbf{X}}_t^k$ if $\widetilde{\mathbf{X}}_t^k$ descends from $\widetilde{\mathbf{X}}_s^i$. Thus, the ancestral origin is the special case corresponding to $s = 1$ (the first generation of the particles). Let

$$N_t^{k,\ell} = \sum_{s=1}^{t} \sum_{i=1}^{m_s} \mathbf{1}_{\{\widetilde{\mathbf{X}}_s^i \text{ is an ancestor of } \widetilde{\mathbf{X}}_t^k \text{ and of } \widetilde{\mathbf{X}}_t^\ell\}}, \tag{5.4}$$

$$\widehat{V}_t = \frac{1}{m_t^2} \sum_{k=1}^{m_t} \sum_{\ell=1}^{m_t} N_t^{k,\ell} \{\psi(\widetilde{\mathbf{X}}_t^k) - \bar{\psi}_T\}\{\psi(\widetilde{\mathbf{X}}_t^\ell) - \bar{\psi}_T\}. \tag{5.5}$$

Theorem 3 of Gilks and Berzuini (Appendix A of [10]) shows that $\widehat{V}_t / V_t \xrightarrow{p} 1$ under (5.3). The basic idea is to use the law of large numbers for each generation conditioned on the preceding one and an induction argument which relies on the assumption (5.3) that "is not directly relevant to the practical context." In contrast, our Theorem 1 or 2 is based on a much more precise martingale approximation of $m(\hat{\psi}_T - \psi_T)$ or $m(\hat{\psi}_{\mathrm{OR}} - \psi_T)$. While we have focused on importance sampling in this paper, another approach to choosing the proposal distribution is to use Markov chain Monte Carlo iterations for the mutation step, as in [10], and important advances in this approach have been developed recently in [1, 4].

APPENDIX: LEMMAS 2, 3, 5, COROLLARY 1 AND (3.30)–(3.32)

PROOF OF LEMMA 2. We use induction on t and show that if (ii) holds for $t-1$, then (i) holds for t, and that if (i) holds for t, then (ii) also holds for t. Since $G = G^+ - G^-$, we can assume without loss of generality that G is nonnegative. By the law of large numbers, (i) holds for $t = 1$, noting that $h_0^* \equiv 1$. Let $t > 1$ and assume that (ii) holds for $t-1$. To show that (i) holds for t, assume $\mu_t = E_q[G(\mathbf{X}_t)/h_{t-1}^*(\mathbf{X}_{t-1})]$ to be finite and suppose that, contrary to (3.15), there exist $0 < \varepsilon < 1$ and $m_1 < m_2 < \cdots$ such that

$$P_m\left\{\left|m^{-1}\sum_{i=1}^m G(\widetilde{\mathbf{X}}_{t,m}^i) - \mu_t\right| > 2\varepsilon\right\} > \varepsilon(2+\mu_t) \qquad \text{for all } m \in \mathcal{M}, \tag{A.1}$$

in which $\mathcal{M} = \{m_1, m_2, \ldots\}$ and we write $\widetilde{\mathbf{X}}_{t,m}^i (= \widetilde{\mathbf{X}}_t^i)$ to highlight the dependence on m. Let $\delta = \varepsilon^3$. Since $x = (x \wedge y) + (x-y)^+$, we can write $G(\widetilde{\mathbf{X}}_{t,m}^i) = U_{t,m}^i + V_{t,m}^i + S_{t,m}^i$, where

$$\begin{aligned} U_{t,m}^i &= G(\widetilde{\mathbf{X}}_{t,m}^i) \wedge (\delta m) - E_m[G(\widetilde{\mathbf{X}}_{t,m}^i) \wedge (\delta m) | \mathcal{F}_{2t-2}], \\ V_{t,m}^i &= E_m[G(\widetilde{\mathbf{X}}_{t,m}^i) \wedge (\delta m) | \mathcal{F}_{2t-2}], \qquad S_{t,m}^i = [G(\widetilde{\mathbf{X}}_{t,m}^i) - \delta m]^+. \end{aligned} \tag{A.2}$$

Since $E_m(U_{t,m}^i | \mathcal{F}_{2t-2}) = 0$ and $\mathrm{Cov}_m(U_{t,m}^i, U_{t,m}^\ell | \mathcal{F}_{2t-2}) = 0$ for $i \neq \ell$, it follows from Chebyshev's inequality, $(U_{t,m}^i)^2 \le \delta m G(\widetilde{\mathbf{X}}_{t,m}^i)$ and $\delta = \varepsilon^3$ that

$$\begin{aligned} P_m&\left\{\left|m^{-1}\sum_{i=1}^m U_{t,m}^i\right| > \varepsilon \middle| \mathcal{F}_{2t-2}\right\} \\ &\le (\varepsilon m)^{-2}\,\mathrm{Var}_m\left(\sum_{i=1}^m U_{t,m}^i \middle| \mathcal{F}_{2t-2}\right) \\ &= (\varepsilon m)^{-2}\sum_{i=1}^m E_m(|U_{t,m}^i|^2 | \mathcal{F}_{2t-2}) \\ &\le \varepsilon m^{-1}\sum_{i=1}^m E_m[G(\widetilde{\mathbf{X}}_{t,m}^i) | \mathcal{F}_{2t-2}] \xrightarrow{p} \varepsilon \mu_t, \end{aligned} \tag{A.3}$$

as $m \to \infty$, by (3.16) applied to $G^*(\mathbf{X}_{t-1}) = E_q[G(\mathbf{X}_t) | \mathbf{X}_{t-1}]$, noting that $E_q[G^*(\mathbf{X}_{t-1})/h_{t-1}^*(\mathbf{X}_{t-1})] = \mu_t$. Application of (3.16) to $G^*(\mathbf{X}_{t-1})$ also yields

$$m^{-1}\sum_{i=1}^m V_{t,m}^i \xrightarrow{p} \mu_t. \tag{A.4}$$

From (A.3), it follows that

$$P_m\left\{\left|m^{-1}\sum_{i=1}^m U_{t,m}^i\right| > \varepsilon\right\} \le \varepsilon(1+\mu_t) \qquad \text{for all large } m. \tag{A.5}$$

Since $\mu_t < \infty$, we can choose n_k such that

$$E_q[G(\mathbf{X}_t)\mathbf{1}_{\{G(\mathbf{X}_t)>n_k\}}/h^*_{t-1}(\mathbf{X}_{t-1})] < k^{-1}.$$

Hence, by applying (3.16) to $G_k(\mathbf{X}_{t-1}) = E_q[G(\mathbf{X}_t)\mathbf{1}_{\{G(\mathbf{X}_t)>n_k\}}|\mathbf{X}_{t-1}]$, we can choose a further subsequence m_k that satisfies (A.1), $m_k \geq n_k/\delta$ and

$$P_m\left\{m^{-1}\sum_{i=1}^m E_m[G(\widetilde{\mathbf{X}}^i_{t,m})\mathbf{1}_{\{G(\widetilde{\mathbf{X}}^i_{t,m})>n_k\}}|\mathcal{F}_{2t-2}] \geq 2k^{-1}\right\} \leq k^{-1} \qquad \text{for } m = m_k.$$

Then for $m = m_k$, with probability at least $1 - k^{-1}$,

$$\begin{aligned}\sum_{i=1}^m P_m\{S^i_{t,m} \neq 0|\mathcal{F}_{2t-2}\} &= \sum_{i=1}^m P_m\{G(\widetilde{\mathbf{X}}^i_{t,m}) > \delta m|\mathcal{F}_{2t-2}\}\\ &\leq (\delta m)^{-1}\sum_{i=1}^m E_m[G(\widetilde{\mathbf{X}}^i_{t,m})\mathbf{1}_{\{G(\widetilde{\mathbf{X}}^i_{t,m})>\delta m\}}|\mathcal{F}_{2t-2}]\\ &\leq 2\delta^{-1}k^{-1}.\end{aligned}$$

Therefore, $P_m\{S^i_{t,m} \neq 0$ for some $1 \leq i \leq m\} \to 0$ as $m = m_k \to \infty$. Combining this with (A.4) and (A.5), we have a contradiction to (A.1). Hence, we conclude (3.15) holds for t. The proof that (ii) holds for t whenever (i) holds for t is similar. In particular, since (i) holds for $t = 1$, so does (ii). □

PROOF OF LEMMA 3. We again use induction and show that if (ii) holds for $t-1$, then (i) holds for t, and that if (i) holds for t, then so does (ii). First (i) holds for $t = 1$ because $A^i_0 = i$, $h^*_0 \equiv 1$ and $E_q G(X_1) < \infty$ implies that for any $\varepsilon > 0$,

$$P_m\left\{m^{-1}\max_{1\leq i\leq m} G(\widetilde{X}^i_1) \geq \varepsilon\right\} \leq m P_q\{G(X_1) \geq \varepsilon m\} \to 0 \qquad \text{as } m \to \infty.$$

Let $t > 1$ and assume that (ii) holds for $t-1$. To show that (i) holds for t, suppose $\mu_t(= E_q[G(\mathbf{X}_t)/h^*_{t-1}(\mathbf{X}_{t-1})]) < \infty$ and that, contrary to (3.19), there exist $0 < \varepsilon < 1$ and $\mathcal{M} = \{m_1, m_2, \ldots\}$ with $m_1 < m_2 < \cdots$ such that

$$P_m\left\{m^{-1}\max_{1\leq j\leq m}\sum_{i:A^i_{t-1}=j} G(\widetilde{\mathbf{X}}^i_{t,m}) > 2\varepsilon\right\} > \varepsilon(2+\mu_t) \qquad \text{for all } m \in \mathcal{M}. \tag{A.6}$$

As in (A.2), let $\delta = \varepsilon^3$ and $G(\widetilde{\mathbf{X}}^i_{t,m}) = U^i_{t,m} + V^i_{t,m} + S^i_{t,m}$. For $1 \leq j \leq m$,

$$P_m\left\{\left|m^{-1}\sum_{i:A^i_{t-1}=j} U^i_{t,m}\right| > \varepsilon\Big|\mathcal{F}_{2t-2}\right\} \leq \varepsilon m^{-1}\sum_{i:A^i_{t-1}=j} E_m[G(\widetilde{\mathbf{X}}^i_{t,m})|\mathcal{F}_{2t-2}],$$

by an argument similar to (A.3). Summing over j and then taking expectations, it then follows from Lemma 2 that

$$\sum_{j=1}^m P_m\left\{\left|m^{-1}\sum_{i:A^i_{t-1}=j} U^i_{t,k}\right| > \varepsilon\right\} \leq \varepsilon(1+\mu_t) \qquad \text{for all large } m. \tag{A.7}$$

By (3.20) applied to the function $G^*(\mathbf{X}_{t-1}) = E_q[G(\mathbf{X}_t)|\mathbf{X}_{t-1}]$,

$$m^{-1} \max_{1\le j\le m} \sum_{i\,:\,A^i_{t-1}=j} V^i_{t,m} \xrightarrow{p} 0 \qquad \text{as } m\to\infty. \tag{A.8}$$

As in the proof of Lemma 2, select a further subsequence m'_k such that (A.6) holds and $P_m\{\sum_{i=1}^m S^i_{t,m} = 0\} \to 1$ as $m = m'_k \to \infty$. Combining this with (A.7) and (A.8) yields a contradiction to (A.6). Hence, (3.19) holds for t.

Next let $t \ge 1$ and assume that (i) holds for t. To show that (ii) holds for t, assume $\tilde{\mu}_t = E_q[G(\mathbf{X}_t)/h^*_t(\mathbf{X}_t)]$ to be finite and note that $\sum_{i\,:\,A^i_t=j} G(\mathbf{X}^i_{t,k}) = \sum_{i\,:\,A^i_{t-1}=j} \#^i_{t,k} G(\widetilde{\mathbf{X}}^i_{t,k})$. Suppose that, contrary to (3.20), there exist $0<\varepsilon<1$ and $\mathcal{M} = \{m_1, m_2, \ldots\}$ with $m_1 < m_2 < \cdots$ such that

$$P_m\Biggl\{m^{-1} \max_{1\le j\le m} \sum_{i\,:\,A^i_{t-1}=j} \#^i_{t,k} G(\widetilde{\mathbf{X}}^i_{t,k}) > 2\varepsilon\Biggr\} > \varepsilon(2+\tilde{\mu}_t) \qquad \text{for all } m\in\mathcal{M}. \tag{A.9}$$

Let $\delta = \varepsilon^3$ and write $\#^i_{t,m} G(\widetilde{\mathbf{X}}^i_{t,m}) = \widetilde{U}^i_{t,m} + \widetilde{V}^i_{t,m} + \#^i_{t,m} S^i_{t,m}$, where

$$\widetilde{U}^i_{t,m} = (\#^i_{t,m} - mW^i_{t,m})[G(\widetilde{\mathbf{X}}^i_{t,m}) \wedge (\delta m)],$$
$$\widetilde{V}^i_{t,m} = mW^i_{t,m}[G(\widetilde{\mathbf{X}}^i_{t,m}) \wedge (\delta m)]$$

and $S^i_{t,m}$ is defined in (A.2). Since $E_m(\#^i_{t,m}|\mathcal{F}_{2t-1}) = mW^i_{t,m}$, $\mathrm{Var}_m(\#^i_{t,m}|\mathcal{F}_{2t-1}) \le mW^i_{t,m} = w_t(\widetilde{\mathbf{X}}^i_{t,m})/\bar{w}_{t,m}$ and since $\#^1_{t,m}, \ldots, \#^m_{t,m}$ are pairwise negatively correlated conditioned on $\mathcal{F}_{2t-1}$, we obtain by Chebyshev's inequality that for $1\le j\le m$, $P_m\{|m^{-1}\sum_{i\,:\,A^i_{t-1}=j} \widetilde{U}^i_{t,m}| > \varepsilon|\mathcal{F}_{2t-1}\}$ is bounded above by

$$\begin{aligned} &\frac{1}{(\varepsilon m)^2} \sum_{i\,:\,A^i_{t-1}=j} \mathrm{Var}_m(\#^i_{t,m}|\mathcal{F}_{2t-1})[G(\widetilde{\mathbf{X}}^i_{t,m}) \wedge (\delta m)]^2 \\ &\quad\le \frac{\varepsilon}{m\bar{w}_{t,m}} \sum_{i\,:\,A^i_{t-1}=j} G^*(\widetilde{\mathbf{X}}^i_{t,m}), \end{aligned}$$

where $G^*(\cdot) = w_t(\cdot)G(\cdot)$. Hence, applying Lemma 2 to G^* and noting that, by (2.5), $E_q[G^*(\mathbf{X}_t)/h^*_{t-1}(\mathbf{X}_{t-1})] = \zeta_t^{-1}\tilde{\mu}_t$ and $\bar{w}_{t,m} = m^{-1}\sum_{i=1}^m w_m(\widetilde{\mathbf{X}}^i_{t,m}) \xrightarrow{p} \zeta_t^{-1}$ by Lemma 2, we obtain

$$\sum_{j=1}^m P_m\Biggl\{\Biggl|m^{-1} \sum_{i\,:\,A^i_{t-1}=j} \widetilde{U}^i_{t,m}\Biggr| > \varepsilon\Biggr\} \le \varepsilon(1+\tilde{\mu}_t) \qquad \text{for all large } m. \tag{A.10}$$

By (3.19) applied to G^* and Lemma 2,

$$m^{-1} \max_{1\le j\le m} \sum_{i\,:\,A^i_{t-1}=j} \widetilde{V}^i_{t,m} \le \bar{w}^{-1}_{t,m}\Biggl[m^{-1} \max_{1\le j\le m} \sum_{i\,:\,A^i_{t-1}=j} G^*(\widetilde{\mathbf{X}}^i_{t,m})\Biggr] \xrightarrow{p} 0. \tag{A.11}$$

Finally, by Lemma 2 applied to $G_m(\mathbf{X}_t) = w_t(\mathbf{X}_t)G(\mathbf{X}_t)\mathbf{1}_{\{G(\mathbf{X}_t)>\delta m\}}$,

$$\begin{aligned}
&\sum_{i=1}^m P_m\{\#_{t,m}^i S_{t,m}^i \neq 0|\mathcal{F}_{2t-1}\}\\
&\qquad = \sum_{i=1}^m P_m\{\#_{t,m}^i > 0|\mathcal{F}_{2t-1}\}\mathbf{1}_{\{S_{t,m}^i \neq 0\}}\\
&\qquad \leq \frac{1}{\delta m}\sum_{i=1}^m E_m(\#_{t,m}^i|\mathcal{F}_{2t-1})\frac{G_m(\widetilde{\mathbf{X}}_{t,m}^i)}{w_t(\widetilde{\mathbf{X}}_{t,m}^i)}\\
&\qquad = \frac{1}{\delta \bar{w}_{t,m}}\left[\frac{1}{m}\sum_{i=1}^m G_m(\widetilde{\mathbf{X}}_{t,m}^i)\right] \xrightarrow{p} 0
\end{aligned} \tag{A.12}$$

as $m = m_k \to \infty$. Since (A.10)–(A.12) contradicts (A.9), (3.20) follows. □

PROOF OF LEMMA 5. The proof of Lemma 5 is similar to that of Lemma 3, and in fact is simpler because the finiteness assumption on the second moments of G dispenses with truncation arguments of the type in (A.2). First (i) holds for $t = 1$ because $A_0^i = i$, $h_0^* = 1$ and

$$P_m\left\{m^{-1}\sum_{i=1}^m G^2(\widetilde{X}_1^i) \geq K\right\} \leq K^{-1}E_q G^2(X_1) \to 0 \qquad \text{as } K \to \infty.$$

As in the proof of Lemma 3, let $t > 1$ and assume that (ii) holds for $t-1$. To show that (i) holds for t, suppose $\mu_t := E_q[G^2(\mathbf{X}_t)\Gamma_{t-1}(\mathbf{X}_{t-1})] < \infty$. By expressing $G = G^+ - G^-$, we may assume without loss of generality G is nonnegative. Write $G(\widetilde{\mathbf{X}}_t^i) = U_t^i + V_t^i$, where $V_t^i = E_m[G(\widetilde{\mathbf{X}}_t^i)|\mathbf{X}_{t-1}^i]$. By the induction hypothesis, $m^{-1}\sum_{j=1}^m(\sum_{i:A_{t-1}^i=j} V_t^i)^2 = O_p(1)$. To show that $m^{-1}\sum_{j=1}^m(\sum_{i:A_{t-1}^i=j} U_t^i)^2 = O_p(1)$, note that $E_m(U_t^i U_t^\ell|\mathcal{F}_{2t-2}) = 0$ for $i \neq \ell$. Hence, it follows by an argument similar to (A.3) that

$$\begin{aligned}
&P_m\left\{m^{-1}\sum_{j=1}^m\left(\sum_{i:A_{t-1}^i=j} U_t^i\right)^2 \geq K\Big|\mathcal{F}_{2t-2}\right\}\\
&\quad \leq (Km)^{-1}\sum_{i=1}^m E_m[(U_t^i)^2|\mathcal{F}_{2t-2}]\\
&\quad \xrightarrow{p} K^{-1}E_q\{|G(\mathbf{X}_t) - E_q[G(\mathbf{X}_t)|\mathbf{X}_{t-1}]|^2/h_{t-1}^*(\mathbf{X}_{t-1})\}.
\end{aligned}$$

Next assume that (i) holds for t and show that (ii) holds for t if $\tilde{\mu}_t := E_q[G^2(\mathbf{X}_t)\Gamma_t(\mathbf{X}_t)]$ is finite. Note that $\sum_{i:A_t^i=j} G(\mathbf{X}_t^i) = \sum_{i:A_{t-1}^i=j} \#_t^i G(\widetilde{\mathbf{X}}_t^i)$. Write $\#_t^i G(\widetilde{\mathbf{X}}_t^i) = \widetilde{U}_t^i + \widetilde{V}_t^i$, where $\widetilde{V}_t^i = mW_t^i G(\widetilde{\mathbf{X}}_t^i) = w_t(\widetilde{\mathbf{X}}_t^i)G(\widetilde{\mathbf{X}}_t^i)/\bar{w}_t$. Since

$\bar{w}_t \xrightarrow{p} \zeta_t^{-1}$ and since $E_q\{[w_t(\mathbf{X}_t)G(\mathbf{X}_t)]^2\Gamma_{t-1}(\mathbf{X}_{t-1})\} \le \tilde{\mu}_t < \infty$, it follows from the induction hypothesis that $m^{-1}\sum_{j=1}^m(\sum_{i:A_{t-1}^i=j}\widetilde{\widetilde{V}}_t^i)^2 = O_p(1)$. Moreover, since $\mathrm{Var}_m(\#_t^i|\mathcal{F}_{2t-1}) \le mW_t^i$ and $\mathrm{Cov}_m(\#_t^i, \#_t^\ell|\mathcal{F}_{2t-1}) \le 0$ for $i \ne \ell$, it follows from an argument similar to (A.3) that

$$P_m\left\{m^{-1}\sum_{j=1}^m\left(\sum_{i:A_{t-1}^i=j}\widetilde{U}_t^i\right)^2 \ge K\middle|\mathcal{F}_{2t-1}\right\} \le (Km)^{-1}\sum_{i=1}^m mW_t^i G^2(\widetilde{\mathbf{X}}_t^i)$$

$$\xrightarrow{p} K^{-1}E_q[G^2(\mathbf{X}_t)/h_t^*(\mathbf{X}_t)],$$

noting that $E_q[G^2(\mathbf{X}_t)/h_t^*(\mathbf{X}_t)] \le \eta_t^{-1}\tilde{\mu}_t < \infty$. Hence,

$$m^{-1}\sum_{j=1}^m\left(\sum_{i:A_{t-1}^i=j}\widetilde{U}_t^i\right)^2 = O_p(1).$$

This shows that (ii) holds for t, completing the induction proof. □

PROOF OF COROLLARY 1. Recall that $\{(Z_k^1, \dots, Z_k^m), \mathcal{F}_k, 1 \le k \le 2T-1\}$ is a martingale difference sequence and that $Z_k^1, \dots, Z_k^m$ are conditionally independent given $\mathcal{F}_{k-1}$, where $\mathcal{F}_0$ is the trival σ-algebra; moreover,

$$\sqrt{m}(\tilde{\psi}_T - \psi_T) = \sum_{k=1}^{2T-1}\left(\sum_{i=1}^m Z_k^i/\sqrt{m}\right). \tag{A.13}$$

Fix k. Conditioned on $\mathcal{F}_{k-1}$, $Y_{mi}^k := Z_k^i/\sqrt{m}$ is a triangular array of row-wise independent (i.i.d. when $k = 2t$) random variables. Noting that $E_m[\tilde{f}_t(\widetilde{\mathbf{X}}_t^i)|\mathbf{X}_{t-1}^i] = \tilde{f}_{t-1}(\mathbf{X}_{t-1}^i)$, we obtain from (3.10) that

$$E_m[(Z_{2t-1}^i)^2|\mathcal{F}_{2t-2}] = \{E_m[\tilde{f}_t^2(\widetilde{\mathbf{X}}_t^i)|\mathbf{X}_{t-1}^i] - \tilde{f}_{t-1}^2(\mathbf{X}_{t-1}^i)\}(H_{t-1}^i)^2,$$

$$E_m[(Z_{2t}^i)^2|\mathcal{F}_{2t-1}] = \sum_{j=1}^m W_t^j\tilde{f}_t^2(\widetilde{\mathbf{X}}_t^j)(\widetilde{H}_t^j)^2 - \left[\sum_{j=1}^m W_t^j\tilde{f}_t(\widetilde{\mathbf{X}}_t^j)\widetilde{H}_t^j\right]^2.$$

Let $\varepsilon > 0$. Since $W_t^j = w_t(\widetilde{\mathbf{X}}_t^j)/(m\bar{w}_t)$ and $\bar{w}_t \xrightarrow{p} \zeta_t^{-1}$,

$$\sum_{i=1}^m E_m[(Y_{mi}^k)^2|\mathcal{F}_{k-1}] \xrightarrow{p} \tilde{\sigma}_k^2, \qquad \sum_{i=1}^m E_m[(Y_{mi}^k)^2\mathbf{1}_{\{|Y_{mi}^k|>\varepsilon\}}|\mathcal{F}_{k-1}] \xrightarrow{p} 0,$$

by (2.5) and Lemma 2. Hence, by Lindeberg's central limit theorem for triangular arrays of independent random variables, the conditional distribution of $\sum_{i=1}^m Z_k^i/\sqrt{m}$ given $\mathcal{F}_{k-1}$ converges to $N(0, \tilde{\sigma}_k^2)$ as $m \to \infty$. This implies that for any real u,

$$E_m(e^{\mathbf{i}u\sum_{j=1}^m Z_k^j/\sqrt{m}}|\mathcal{F}_{k-1}) \xrightarrow{p} e^{-u^2\tilde{\sigma}_k^2/2}, \qquad 1 \le k \le 2T-1, \tag{A.14}$$

where $\mathbf{i}$ denotes the imaginary number $\sqrt{-1}$. Let $S_t = \sum_{k=1}^t (\sum_{j=1}^m Z_k^j/\sqrt{m})$. Since $\{\sum_{j=1}^m Z_k^j/\sqrt{m}, \mathcal{F}_k, 1 \le k \le 2T-1\}$ is a martingale difference sequence, it follows from (A.13), (A.14) and mathematical induction that

$$\begin{aligned}
&E_m(e^{\mathbf{i}u\sqrt{m}(\tilde{\psi}_T-\psi_T)}) \\
&\quad = E_m[e^{\mathbf{i}uS_{2T-2}} E_m(e^{\mathbf{i}u\sum_{j=1}^m Z_{2T-1}^j/\sqrt{m}} | \mathcal{F}_{2T-2})] \\
&\quad = e^{-u^2\tilde{\sigma}_{2T-1}^2/2} E_m(e^{\mathbf{i}uS_{2T-2}}) + o(1) = \cdots = \exp(-u^2\sigma_{\mathrm{C}}^2/2) + o(1).
\end{aligned}$$

Hence, (3.13) holds. $\square$

PROOF OF (3.30)–(3.32). For distinct i, j, k, ℓ, define $c_{ij} = (mW_t^i)(mW_t^j)$, $c_{ijk\ell} = (mW_t^i)(mW_t^j)(mW_t^k)(mW_t^\ell)$, $c_{iijk} = (mW_t^i)^2(mW_t^j)(mW_t^k)$ and define $c_{ijk}, c_{iij}, c_{ijj}$ similarly. Since $\#_t^i = \sum_{h=1}^m \mathbf{1}_{\{B_t^h=i\}}$, where B_t^h are i.i.d. conditioned on $\mathcal{F}_{2t-1}$ such that $P\{B_t^h = i | \mathcal{F}_{2t-1}\} = W_t^i$, combinatorial arguments show that

$$\begin{aligned}
&E_m[(\#_t^i - mW_t^i)(\#_t^j - mW_t^j)(\#_t^k - mW_t^k)(\#_t^\ell - mW_t^\ell) | \mathcal{F}_{2t-1}] \\
&\quad = m^{-4}[m(m-1)(m-2)(m-3) \\
&\qquad - 4m^2(m-1)(m-2) + 6m^3(m-1) - 4m^4 + m^4] c_{ijk\ell} \\
&\quad = (3m^{-2} - 6m^{-3}) c_{ijk\ell},
\end{aligned} \tag{A.15}$$

$$\begin{aligned}
&E_m[(\#_t^i - mW_t^i)^2(\#_t^j - mW_t^j)(\#_t^k - mW_t^k) | \mathcal{F}_{2t-1}] \\
&\quad = (3m^{-2} - 6m^{-3}) c_{iijk} + (-m^{-1} + 2m^{-2}) c_{ijk},
\end{aligned} \tag{A.16}$$

$$\begin{aligned}
&E_m[(\#_t^i - mW_t^i)^2(\#_t^j - mW_t^j)^2 | \mathcal{F}_{2t-1}] \\
&\quad = (3m^{-2} - 6m^{-3}) c_{iijj} + (-m^{-1} + 2m^{-2})(c_{iij} + c_{ijj}) \\
&\qquad + (1 - m^{-1}) c_{ij},
\end{aligned} \tag{A.17}$$

$$\begin{aligned}
&E_m[(\#_t^i - mW_t^i)^4 | \mathcal{F}_{2t-1}] \\
&\quad = m[W_t^i(1 - W_t^i)^4 + (1 - W_t^i)(W_t^i)^4] \\
&\qquad + 3m(m-1)[(W_t^i)^2(1 - W_t^i)^4 \\
&\qquad\qquad + 2(W_t^i)^3(1 - W_t^i)^3 + (W_t^i)^4(1 - W_t^i)^2],
\end{aligned} \tag{A.18}$$

$$E_m\{(\#_t^i - mW_t^i)(\#_t^j - mW_t^j) | \mathcal{F}_{2t-1}\} = -m^{-1} c_{ij}. \tag{A.19}$$

To prove (3.30), let $T_i = mW_t^i[\tilde{f}_t(\widetilde{\mathbf{X}}_t^i)\widetilde{H}_t^i - \tilde{f}_0]^2$, $\overline{T} = m^{-1}\sum_{i=1}^m T_i$ and $T_t^* = m^{-1}\max_{1\le j\le m}\sum_{i:A_{t-1}^i=j} T_i$. We can use (A.17) and (A.18) to show that

$m^{-2}E_m[(\sum_{i=1}^m S_i^2)^2|\mathcal{F}_{2t-1}]$ is equal to

$$m^{-2}\sum_{i=1}^{m} E_m\big(S_i^4|\mathcal{F}_{2t-1}\big) + m^{-2}\sum_{i\neq j} E_m\big(S_i^2 S_j^2|\mathcal{F}_{2t-1}\big) = \overline{T}^{\,2} + o_P(1)$$

as $m \to \infty$. By Lemma 2(i), $\overline{T}^{\,2} \xrightarrow{p} \tilde{\sigma}_{2t}^4$. From this and (3.29), (3.30) follows.

To prove (3.31), let $D_i = mW_t^i[\tilde{f}_t(\widetilde{\mathbf{X}}_t^i)\widetilde{H}_t^i - \tilde{f}_0]$, $\overline{D} = m^{-1}\sum_{i=1}^m |D_i|$ and $D_t^* = m^{-1}\max_{1\le j\le m}\sum_{i\,:\,A_{t-1}^i=j}|D_i|$. By (A.15)–(A.17), the left-hand side of (3.31) is equal to

$$\begin{aligned}
&\frac{1}{m^2}\left(\frac{3}{m^2} - \frac{6}{m^3}\right)\left\{\sum_{j=1}^m \sum_{(i,\ell)\in C_{t-1}^j} D_i D_\ell\right\}^2 + \frac{1}{m^2}\left(1 - \frac{1}{m}\right)\sum_{j=1}^m \sum_{(i,\ell)\in C_{t-1}^j} T_i T_\ell \\
&\quad - \frac{4}{m^2}\left(\frac{1}{m} - \frac{2}{m^2}\right)\sum_{j=1}^m\left\{\sum_{i\,:\,A_{t-1}^i=j} T_i\left(\sum_{\ell\neq i\,:\,A_{t-1}^\ell=j} D_\ell\right)^2\right\} \\
&\quad \le 3\big(\overline{D}D_t^*\big)^2 + \overline{T}T_t^* \qquad \text{for } m \ge 2.
\end{aligned}$$

By Lemmas 2 and 3, the upper bound in the above inequality converges to 0 in probability as $m \to \infty$, proving (3.31).

By (A.19), $E_m(S_iS_j|\mathcal{F}_{2t-1}) = -m^{-1}D_iD_j$ and, similarly, $E_m(S_i^2|\mathcal{F}_{2t-1}) = T_i - m^{-1}D_i^2$. Therefore, by (3.4), the left-hand side of (3.32) is equal to

$$m^{-1}\sum_{j=1}^m (\varepsilon_1^j + \cdots + \varepsilon_{2t-1}^j)^2\left\{m^{-1}\sum_{i\,:\,A_{t-1}^i=j} T_i - \left(m^{-1}\sum_{i\,:\,A_{t-1}^i=j} D_i\right)^2\right\} \xrightarrow{p} 0,$$

since $T_t^* + (D_t^*)^2 \xrightarrow{p} 0$ by Lemma 3 and since (3.21) holds for $k = 2t-1$ by the induction hypothesis. □

REFERENCES

[1] ANDRIEU, C., DOUCET, A. and HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 269–342. MR2758115

[2] BAKER, J. E. (1985). Adaptive selection methods for genetic algorithms. In *Proc. International Conference on Genetic Algorithms and Their Applications* (J. Grefenstette, ed.) 101–111. Erlbaum, Mahwah, NJ.

[3] BAKER, J. E. (1987). Reducing bias and inefficiency in the selection algorithm. In *Genetic Algorithms and Their Applications* (J. Grefenstette, ed.) 14–21. Erlbaum, Mahwah, NJ.

[4] BROCKWELL, A., DEL MORAL, P. and DOUCET, A. (2010). Sequentially interacting Markov chain Monte Carlo methods. *Ann. Statist.* **38** 3387–3411. MR2766856

[5] CHAN, H. P. and LAI, T. L. (2011). A sequential Monte Carlo approach to computing tail probabilities in stochastic models. *Ann. Appl. Probab.* **21** 2315–2342. MR2895417

[6] CHOPIN, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.* **32** 2385–2411. MR2153989

[7] CRISAN, D., DEL MORAL, P. and LYONS, T. (1999). Discrete filtering using branching and interacting particle systems. *Markov Process. Related Fields* **5** 293–318. MR1710982

[8] DEL MORAL, P. (2004). *Feynman–Kac Formulae: Genealogical and Interacting Particle Systems With Applications*. Springer, New York. MR2044973

[9] DOUC, R. and MOULINES, E. (2008). Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Ann. Statist.* **36** 2344–2376. MR2458190

[10] GILKS, W. R. and BERZUINI, C. (2001). Following a moving target—Monte Carlo inference for dynamic Bayesian models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **63** 127–146. MR1811995

[11] GORDON, N. J., SALMOND, D. J. and SMITH, A. F. M. (1993). A novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEEE Proc. Radar Signal Processing* **140** 107–113.

[12] KÜNSCH, H. R. (2005). Recursive Monte Carlo filters: Algorithms and theoretical analysis. *Ann. Statist.* **33** 1983–2021. MR2211077

[13] LIU, J. S. (2008). *Monte Carlo Strategies in Scientific Computing*. Springer, New York. MR2401592

[14] LIU, J. S. and CHEN, R. (1998). Sequential Monte Carlo methods for dynamic systems. *J. Amer. Statist. Assoc.* **93** 1032–1044. MR1649198

[15] YAO, Y.-C. (1984). Estimation of a noisy discrete-time step function: Bayes and empirical Bayes approaches. *Ann. Statist.* **12** 1434–1447. MR0760698

DEPARTMENT OF STATISTICS
AND APPLIED PROBABILITY
NATIONAL UNIVERSITY OF SINGAPORE
6 SCIENCE DRIVE 2
SINGAPORE 117546
SINGAPORE
E-MAIL: stachp@nus.edu.sg

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
SEQUOIA HALL
STANFORD, CALIFORNIA 94305-4065
USA
E-MAIL: lait@stat.stanford.edu