Adam M. Johansen and Arnaud Doucet

Auxiliary Variable Sequential Monte Carlo Methods

University of Bristol, Statistics Groups

Technical Report 07:09

12th July, 2007

Adam M. Johansen
Department of Mathematics
University of Bristol, UK
Adam.Johansen@bristol.ac.uk

Arnaud Doucet
Departments of Statistics & Computer Science
University of British Columbia
Vancouver, B.C., Canada
Arnaud@stat.ubc.ca

Summary

The Auxiliary Particle Filter (APF) introduced by Pitt and Shephard (1999) is a very popular alternative to Sequential Importance Sampling / Resampling (SISR) algorithms to perform inference in state-space models. We propose a novel interpretation of the APF as an SISR algorithm. This interpretation allows us to present simple guidelines to ensure good performance of the APF and the first convergence results for this algorithm. Additionally, we show that, contrary to popular belief, the asymptotic variance of APF-based estimators is not always smaller than those of the corresponding SISR – even in the 'perfect adaptation' scenario. We also explain how similar concepts can be applied to general Sequential Monte Carlo Samplers and provide similar results in this context.

Key words: Optimal filtering, Particle filtering, State-space models.

1. Introduction

The principle purpose of this report is to reinterpret the auxiliary particle filter (APF) of [15, 16] in a manner which makes no use of auxiliary variables.

We begin in section 2 by presenting a novel interpretation of the APF and showing that this interpretation allows us to employ now standard analysis techniques to demonstrate convergence and other asymptotic results easily. This is illustrated with a central limit theorem together with an easy to interpret asymptotic variance decomposition. Such results have until now been absent from the literature.

In section 3 we show that this interpretation allows us to extend the technique employed by the APF from the domain of filtering to the more general sampling regime of the sequential Monte Carlo sampler of [4] and, again, the analysis is straightforward; a general central limit theorem is provided for this class of algorithms which includes those considered in section 2 as particular cases.

Sections 2 and 3 are each intended to be largely self contained. The intention is that it should be possible to read either independently whilst duplication is minimised. Section 2 contains interesting particular cases of the results with direct proofs. The results presented in section 3 cover a broader range of algorithms and some methodological extensions are proposed; this section illustrates that is is possible to analyse a wide range of algorithms within a common framework but introduces a few additional complications in doing so. Both sections contain a central limit theorem and variance expression for the APF.

2. Sequential Monte Carlo Filtering

2.1 Introduction

Let $t = 1, 2, ...$ denote a discrete-time index. Consider an unobserved $\mathcal{X}-$valued Markov process $\{X_t\}_{t\geq 1}$ such that $X_1 \sim \mu(\cdot)$ and $X_t|\,(X_{t-1} = x_{t-1}) \sim f(\cdot|\,x_{t-1})$ where $f(\cdot|\,x_{t-1})$ is the homogeneous transition density with respect to a suitable dominating measure. The observations $\{Y_t\}_{t\geq 1}$ are conditionally independent given $\{X_t\}_{t\geq 1}$ and distributed according to $Y_t|\,(X_t = x_t) \sim g(\cdot|\,x_t)$.

For any sequence $\{z_t\}_{t\geq 1}$, we use the notation $z_{i:j} = (z_i, z_{i+1}, ..., z_j)$. In numerous applications, we are interested in estimating recursively in time the sequence of posterior distributions $\{p(x_{1:t}|\,y_{1:t})\}_{t\geq 1}$ given by

$$p(x_{1:t}|\,y_{1:t}) \propto \mu(x_1)\, g(y_1|\,x_1) \prod_{k=2}^{t} f(x_k|\,x_{k-1})\, g(y_k|\,x_k). \tag{2.1}$$

When the model is linear Gaussian, the posterior distributions are Gaussian and their statistics can be computed using Kalman techniques. For non-linear non-Gaussian methods, these distributions do not typically admit a closed-form and it is necessary to employ numerical approximations. Recently, the class of Sequential Monte Carlo (SMC) methods - also known as particle filters - has emerged to solve this problem; see [7, 14] for a review of the literature. Two classes of methods are primarily used: Sequential Importance Sampling / Resampling (SISR) algorithms [3, 14, 8] and Auxiliary Particle Filters (APF) [15, 1, 16].

In the literature, the APF methodology is always presented as significantly different from the SISR methodology. It was originally introduced in [15] using auxiliary variables hence its name. Several improvements were proposed to reduce its variance [1, 16]. In [11], the APF is presented without introducing any auxiliary variable and also reinterpreted as an SISR algorithm. However, this SISR algorithm is non-standard as it relies on a proposal distribution at time t on the path space $\mathcal{X}^t$ which is dependent on all the paths sampled previously.

We study here the version of the APF presented in [1] which only includes one resampling step at each time instance. Experimentally this version outperforms the original two stage resampling algorithm proposed in [15] and is widely used; see [1] for a comparison of both approaches and [9] for an application to partially-observed diffusions. We propose a novel interpretation of this APF as a *standard* SISR algorithm which we believe has two principal advantages over previous derivations/interpretations. First, it allows us to give some simple guidelines to ensure good performance of the APF. These guidelines differ from many practical implementations of the APF and explain some of the poor performance reported in the literature. Second, there is no convergence result available for the APF in the literature whereas there are numerous results available for SISR algorithms; see [3] for a thorough treatment. Via this novel interpretation, we can easily adapt these results to the APF. We present here the asymptotic variance associated with APF-based estimators and show that this asymptotic variance is not necessarily lower than that of the corresponding standard SISR-based estimators – even in the 'perfect adaptation' case which is discussed further below.

2.2 SISR and APF

2.2.1 A Generic SISR algorithm

Consider an arbitrary sequence of probability distributions $\{\pi_t(x_{1:t})\}_{t\geq 1}$. To sample sequentially from these distributions, the SISR algorithm introduces at time t an importance distribution $q_t(x_t|\, x_{t-1})$ to impute X_t (and $q_1(x_1)$ at time 1). Note that it is possible to use a distribution $q_t(x_t|\, x_{1:t-1})$ but this additional freedom is not useful for the optimal filtering applications discussed here. The SISR algorithm proceeds as follows; see for example [7], [14, chapter 3] for variations:

At time 1.

Sampling Step

For $i = 1 : N$, sample $X_{1,1}^{(i)} \sim q_1(\cdot)$.

Resampling Step

For $i = 1 : N$, compute $w_1\left(X_{1,1}^{(i)}\right) = \frac{\pi_1\left(X_{1,1}^{(i)}\right)}{q_1\left(X_{1,1}^{(i)}\right)}$ and $W_1^{(i)} = \frac{w_1\left(X_{1,1}^{(i)}\right)}{\sum_{j=1}^N w_1\left(X_{1,1}^{(j)}\right)}$.

For $i = 1 : N$, sample $\check{X}_{1,1}^{(i)} \sim \sum_{j=1}^N W_1^{(j)} \delta_{X_{1,1}^{(j)}}(dx_1)$.

At time t, $t \geq 2$.

Sampling Step

For $i = 1 : N$, sample $X_{t,t}^{(i)} \sim q_t\left(\cdot|\, \check{X}_{t-1,t-1}^{(i)}\right)$.

Resampling Step

For $i = 1 : N$, compute $w_t\left(\check{X}_{1:t-1,t-1}^{(i)}, X_{t,t}^{(i)}\right) = \frac{\pi_t\left(\check{X}_{1:t-1,t-1}^{(i)}, X_{t,t}^{(i)}\right)}{\pi_{t-1}\left(\widetilde{X}_{1:t-1,t-1}^{(i)}\right) q_t\left(X_{t,t}^{(i)}\middle|\check{X}_{t-1,t-1}^{(i)}\right)}$

and $W_t^{(i)} = \frac{w_t\left(\widetilde{X}_{1:t-1,t-1}^{(i)}, X_{t,t}^{(i)}\right)}{\sum_{j=1}^N w_t\left(\check{X}_{1:t-1,t-1}^{(j)}, X_{t,t}^{(j)}\right)}$.

For $i = 1 : N$, sample $\check{X}_{1:t,t}^{(i)} \sim \sum_{j=1}^N W_t^{(j)} \delta_{\left(\check{X}_{1:t-1,t-1}^{(j)}, X_{t,t}^{(j)}\right)}(dx_{1:t})$.

The empirical measure

$$\rho_t^N(dx_{1:t}) = \frac{1}{N}\sum_{i=1}^N \delta_{\left(\check{X}_{1:t-1,t-1}^{(i)}, X_{t,t}^{(i)}\right)}(dx_{1:t})$$

is an approximation of $\pi_{t-1}(x_{1:t-1})\, q_t(x_t|\, x_{t-1})$ whereas

$$\pi_t^N(dx_{1:t}) = \sum_{i=1}^N W_t^{(i)} \delta_{\left(\check{X}_{1:t-1,t-1}^{(i)}, X_{t,t}^{(i)}\right)}(dx_{1:t})$$

is an approximation of $\pi_t(x_{1:t})$.

Whilst, in practice, one may also wish to employ a lower variance resampling strategy such as residual resampling and to use it only when some criterion indicates that it is necessary, results of the sort presented here are sufficient to guide the design of particular algorithms and the additional complexity involved in considering more general scenarios serves largely to produce substantially more complex expressions which obscure the important points.

2.2.2 APF as an SISR algorithm

The standard SISR algorithm for filtering corresponds to the case in which we set $\pi_t(x_{1:t}) = p(x_{1:t}|\, y_{1:t})$. In this case, for any test function $\varphi_t : \mathcal{X}^t \to \mathbb{R}$, we estimate $\overline{\varphi}_t = \int \varphi_t(x_{1:t})\, p(x_{1:t}|\, y_{1:t})\, dx_{1:t}$ by

$$\widehat{\varphi}^N_{t,SISR} = \int \varphi_t(x_{1:t})\, \pi^N_t(dx_{1:t}) = \sum_{i=1}^N W^{(i)}_t \varphi\left(\check{X}^{(i)}_{1:t-1,t-1}, X^{(i)}_{t,t}\right). \tag{2.2}$$

The APF described in [1] corresponds to the case where we select

$$\pi_t(x_{1:t}) = \widehat{p}(x_{1:t}|\, y_{1:t+1}) \propto p(x_{1:t}|\, y_{1:t})\, \widehat{p}(y_{t+1}|\, x_t) \tag{2.3}$$

with $\widehat{p}(y_{t+1}|\, x_t)$ an approximation of

$$p(y_{t+1}|\, x_t) = \int g(y_{t+1}|\, x_{t+1})\, f(x_{t+1}|\, x_t)\, dx_{t+1}$$

if $p(y_{t+1}|\, x_t)$ is not known analytically. As the APF does not approximate directly $p(x_{1:t}|\, y_{1:t})$, we need to use importance sampling to estimate $\overline{\varphi}_t$. We use the importance distribution $\pi_{t-1}(x_{1:t-1})\, q_t(x_t|\, x_{t-1})$ whose approximation $\rho^N_t(dx_{1:t})$ is obtained after the sampling step. The resulting estimate is given by

$$\widehat{\varphi}^N_{t,APF} = \sum_{i=1}^N \widetilde{W}^{(i)}_t \varphi_t\left(\check{X}^{(i)}_{1:t-1,t-1}, X^{(i)}_{t,t}\right) \tag{2.4}$$

where

$$\widetilde{W}^{(i)}_t = \frac{\widetilde{w}_t\left(\check{X}^{(i)}_{t-1,t-1}, X^{(i)}_{t,t}\right)}{\sum_{j=1}^N \widetilde{w}_t\left(\check{X}^{(j)}_{t-1,t-1}, X^{(j)}_{t,t}\right)}$$

and

$$\widetilde{w}_t(x_{t-1:t}) = \frac{p(x_{1:t}|\, y_{1:t})}{\pi_{t-1}(x_{1:t-1})\, q_t(x_t|\, x_{t-1})} \propto \frac{g(y_t|\, x_t)\, f(x_t|\, x_{t-1})}{\widehat{p}(y_t|\, x_{t-1})\, q_t(x_t|\, x_{t-1})}. \tag{2.5}$$

In both cases, we usually select $q_t(x_t|\, x_{t-1})$ as an approximation to

$$p(x_t|\, y_t, x_{t-1}) = \frac{g(y_t|\, x_t)\, f(x_t|\, x_{t-1})}{p(y_t|\, x_{t-1})}.$$

This distribution is often referred to as the optimal importance distribution [7]. When it is possible to select $q_t(x_t|\, x_{t-1}) = p(x_t|\, y_t, x_{t-1})$ and $\widehat{p}(y_t|\, x_{t-1}) = p(y_t|\, x_{t-1})$, we obtain the so-called 'perfect adaptation' case [15]. In this case, the APF takes a particularly simple form as the importance weights (2.5) are all equal. This can be interpreted as a standard SISR algorithm where the order of the sampling and resampling steps is interchanged. It is widely believed that this strategy yields estimates with a necessarily smaller variance as it increases the number of distinct particles at time t. We will show further that this is not necessarily the case.

2.2.3 APF Settings

It is well-known in the literature that we should select $q_t(x_t|x_{t-1})$ as a distribution with thicker tails than $p(x_t|y_t, x_{t-1})$. However, this simple reinterpretation of the APF also shows that we should select a distribution $\widehat{p}(x_{1:t-1}|y_{1:t})$ with thicker tails than $p(x_{1:t-1}|y_{1:t})$ as $\widehat{p}(x_{1:t-1}|y_{1:t})$ is used as an importance distribution to estimate $p(x_{1:t-1}|y_{1:t})$. Thus $\widehat{p}(y_t|x_{t-1})$ should be more diffuse than $p(y_t|x_{t-1})$. It has been suggested in the literature to set $\widehat{p}(y_t|x_{t-1}) = g(y_t|\mu(x_{t-1}))$ where $\mu(x_{t-1})$ corresponds to the mode, mean or median of $f(x_t|x_{t-1})$. However, this simple approximation will often yield an importance weight function (2.5) which is not upper bounded on $\mathcal{X} \times \mathcal{X}$ and could lead to an estimator with a large/infinite variance. An alternative, and preferable approach consists of selecting an approximation $\widehat{p}(y_t, x_t|x_{t-1}) = \widehat{p}(y_t|x_{t-1})\widehat{p}(x_t|y_t, x_{t-1})$ of the distribution $p(y_t, x_t|x_{t-1}) = p(y_t|x_{t-1})p(x_t|y_t, x_{t-1}) = g(y_t|x_t)f(x_t|x_{t-1})$ such that the ratio (2.5) is upper bounded on $\mathcal{X} \times \mathcal{X}$ and such that it is possible to compute $\widehat{p}(y_t|x_{t-1})$ pointwise and to sample from $\widehat{p}(x_t|y_t, x_{t-1})$.

2.2.4 Convergence Results

There is a wide range of sharp convergence results available for SISR algorithms [3]. We present here a Central Limit Theorem (CLT) for the SISR and the APF estimates (2.2) and (2.4), giving the asymptotic variances of these estimates. The asymptotic variance of the CLT for the SISR estimate (2.2) has been established several times in the literature. We present here a new representation which we believe clarifies the influence of the ergodic properties of the optimal filter on the asymptotic variance.

Proposition 2.2.1. *Under the regularity conditions given in [2, Theorem 1] or [3, Section 9.4, pp. 300-306], we have*

$$\sqrt{N}\left(\widehat{\varphi}^N_{t,SISR} - \overline{\varphi}_t\right) \to \mathcal{N}\left(0, \sigma^2_{SISR}(\varphi_t)\right),$$
$$\sqrt{N}\left(\widehat{\varphi}^N_{t,APF} - \overline{\varphi}_t\right) \to \mathcal{N}\left(0, \sigma^2_{APF}(\varphi_t)\right)$$

where '$\to$' denotes convergence in distribution and $\mathcal{N}\left(0,\sigma^2\right)$ is the zero-mean normal of variance σ^2. Moreover, at time $t = 1$ we have

$$\sigma^2_{SISR}(\varphi_1) = \sigma^2_{APF}(\varphi_1) = \int \frac{p(x_1|y_1)^2}{q_1(x_1)}\left(\varphi_1(x_1) - \overline{\varphi}_1\right)^2 dx_1$$

whereas for $t > 1$

$$\begin{aligned}\sigma^2_{SISR}(\varphi_t) = &\int \frac{p(x_1|y_{1:t})^2}{q_1(x_1)}\left(\int \varphi_t(x_{1:t})\,p(x_{2:t}|y_{2:t}, x_1)\,dx_{2:t} - \overline{\varphi}_t\right)^2 dx_1 \\ &+ \sum_{k=2}^{t-1}\int \frac{p(x_{1:k}|y_{1:t})^2}{p(x_{1:k-1}|y_{1:k-1})\,q_k(x_k|x_{k-1})}\left(\int \varphi_t(x_{1:t})\,p(x_{k+1:t}|y_{k+1:t}, x_k)\,dx_{k+1:t} - \overline{\varphi}_t\right)^2 dx_{1:k} \\ &+ \int \frac{p(x_{1:t}|y_{1:t})^2}{p(x_{1:t-1}|y_{1:t-1})\,q_t(x_t|x_{t-1})}\left(\varphi_t(x_{1:t}) - \overline{\varphi}_t\right)^2 dx_{1:t},\end{aligned} \tag{2.6}$$

and

$$\begin{aligned}\sigma_{APF}^2(\varphi_t) =& \int \frac{p(x_1|y_{1:t})^2}{q_1(x_1)}\left(\int \varphi_t(x_{1:t})p(x_{2:t}|y_{2:t},x_1)dx_{2:t}-\bar{\varphi}_t\right)^2 dx_1 \\ &+\sum_{k=2}^{t-1}\int \frac{p(x_{1:k}|y_{1:t})^2}{\hat{p}(x_{1:k-1}|y_{1:k})q_k(x_k|x_{k-1})}\left(\int \varphi_t(x_{1:t})p(x_{k+1:t}|y_{k+1:t},x_k)dx_{k+1:t}-\bar{\varphi}_t\right)^2 dx_{1:k} \\ &+\int \frac{p(x_{1:t}|y_{1:t})^2}{\hat{p}(x_{1:t-1}|y_{1:t})q_t(x_t|x_{t-1})}\left(\varphi_t(x_{1:t})-\bar{\varphi}_t\right)^2 dx_{1:t}.\end{aligned} \tag{2.7}$$

Sketch of Proof. Expression (2.6) follows from a straightforward but tedious rewriting of the expression given in [3, Section 9.4, pp. 300-306]. We defer these lengthy calculations to appendix A.

The variance of the estimate $\sum_{i=1}^N W_t^{(i)}\varphi_t\left(\check{X}_{1:t-1,t-1}^{(i)}, X_{t,t}^{(i)}\right)$ when $\pi_t\left(x_{1:t}\right)$ is given by (2.3) is given by an expression similar to (2.6) but with the terms $\widehat{p}\left(x_{1:k}\middle| y_{1:t+1}\right)$, $\widehat{p}\left(x_{1:k-1}\middle| y_{1:k}\right)$ and $\widehat{p}\left(x_{k+1:t-1}\middle| y_{k+1:t+1}, x_k\right)$ replacing $p\left(x_{1:k}\middle| y_{1:t}\right)$, $p\left(x_{1:k-1}\middle| y_{1:k-1}\right)$ and $p\left(x_{k+1:t}\middle| y_{k+1:t}, x_k\right)$, respectively (and with $\bar{\varphi}_t$ replaced by $\int \varphi_t(x_{1:t})\hat{p}(x_{1:t}|y_{1:t+1})dx_{1:t}$)). Then by the same argument as [2, Lemma A2] the variance $\sigma_{APF}^2\left(\varphi_t\right)$ is equal to the variance of $\sum_{i=1}^N W_t^{(i)}\varphi_t'\left(\check{X}_{1:t-1,t-1}^{(i)}, X_{t,t}^{(i)}\right)$ where

$$\varphi_t'\left(x_{1:t}\right)=\frac{p\left(x_{1:t}\middle| y_{1:t}\right)}{\widehat{p}\left(x_{1:t}\middle| y_{1:t+1}\right)}\left[\varphi_t\left(x_{1:t}\right)-\bar{\varphi}_t\right]$$

and the expression (2.7) follows directly. Full details can be found in appendix A.

Corollary. In the perfect adaptation scenario where $\widehat{p}\left(y_t\middle| x_{t-1}\right)=p\left(y_t\middle| x_{t-1}\right)$ and $q_t\left(x_t\middle| x_{t-1}\right)=p\left(x_t\middle| y_t, x_{t-1}\right)$, we have

$$\begin{aligned}\sigma_{APF}^2(\varphi_t) =& \int \frac{p(x_1|y_{1:t})^2}{p(x_1|y_1)}\left(\int \varphi_t(x_{1:t})p(x_{2:t}|y_{2:t},x_1)dx_{2:t}-\bar{\varphi}_t\right)^2 dx_1 \\ &+\sum_{k=2}^{t-1}\int \frac{p(x_{1:k}|y_{1:t})^2}{p(x_{1:k}|y_{1:k})}\left(\int \varphi_t(x_{1:t})p(x_{k+1:t}|y_{k+1:t},x_k)dx_{k+1:t}-\bar{\varphi}_t\right)^2 dx_{1:k} \\ &+\int p(x_{1:t}|y_{1:t})\left(\varphi_t(x_{1:t})-\bar{\varphi}_t\right)^2 dx_{1:t}.\end{aligned}$$

Remark. The asymptotic bias for the APF can also be established by a simple adaptation of [6, Theorem 1.1]. Both the bias and variance associated to $\varphi_t\left(x_{1:t}\right)=\varphi_t\left(x_t\right)$ can be uniformly bounded in time using [6, Proposition 4.1.]; see also [2, Theorem 5].

One may interpret these variance expressions via a local error decomposition such as that of [3, Chapters 7 & 9]. The error of the particle system estimate at time t may be decomposed as a sum of differences, specifically, the difference in the estimate due to propagating forward the particle system rather than the exact solution from that time-step to the next. Summing over all such terms gives the difference between the particle system estimate and the truth. These variance expressions illustrate that, asymptotically at least, the variance follows a similar decomposition.

Each term in the variance expressions matches an importance sampling variance. Loosely, it is the variance of estimating the integral of a function under the smoothing distribution $p(x_{1:k}|y_{1:t})$ using as an importance distribution the last resampling distribution propagated forward according to the proposal; the functions being integrated correspond to propagating the system forward to time t using all remaining observations and then estimating the integral of φ_t. Thus, for ergodic systems in which some

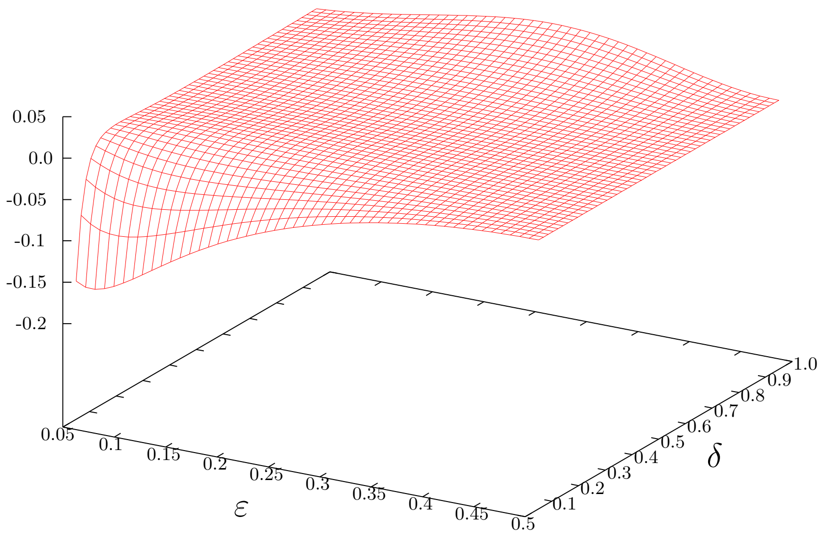

Fig. 2.1. Asymptotic variance difference, $\sigma^2_{APF}(\varphi_2) - \sigma^2_{SISR}(\varphi_2)$ for the example. This is negative wherever the APF outperforms SISR.

forgetting property holds, the early terms in this sum will decay (at least when φ_t depends only upon the final time marginal) and the system will remain well behaved over time.

2.3 Example

To illustrate the implications of these results, we employ the following binary state-space model with common state and observation spaces:

$$\mathcal{X} = \{0,1\} \quad p(x_1 = 0) = 0.5 \quad p(x_t = x_{t-1}) = 1 - \delta \quad y_t \in \mathcal{X} \quad p(y_t = x_t) = 1 - \varepsilon.$$

This is an extremely simple state-space model and one would could obtain the exact solution without difficulty. However, the evolution of this system from $t = 1$ to $t = 2$ provides sufficient structure to illustrate the important points and the simplicity of the model enables us to demonstrate concepts which generalise to more complex scenarios.

We consider the estimation of the function $\varphi_2(x_{1:2}) = x_2$ during the second iteration of the algorithms when $y_1 = 0, y_2 = 1$. The optimal importance distributions and the true predictive likelihood are available in this case. Additionally, the model has two parameters which are simple to interpret: δ determines how informative the dynamic model is (when δ is close to 0.5 the state at time t is largely unrelated to that at time $t-1$; when it approaches 0 or 1 the two become extremely highly correlated) and ε determines how informative the observations are (when ε reaches zero, the observation at time t specifies the state deterministically, and as it approaches 0.5 it provides no information about the state).

Figure 2.1 shows the difference between the asymptotic variance of the APF and SISR algorithms in this setting; note that the function plotted is negative whenever

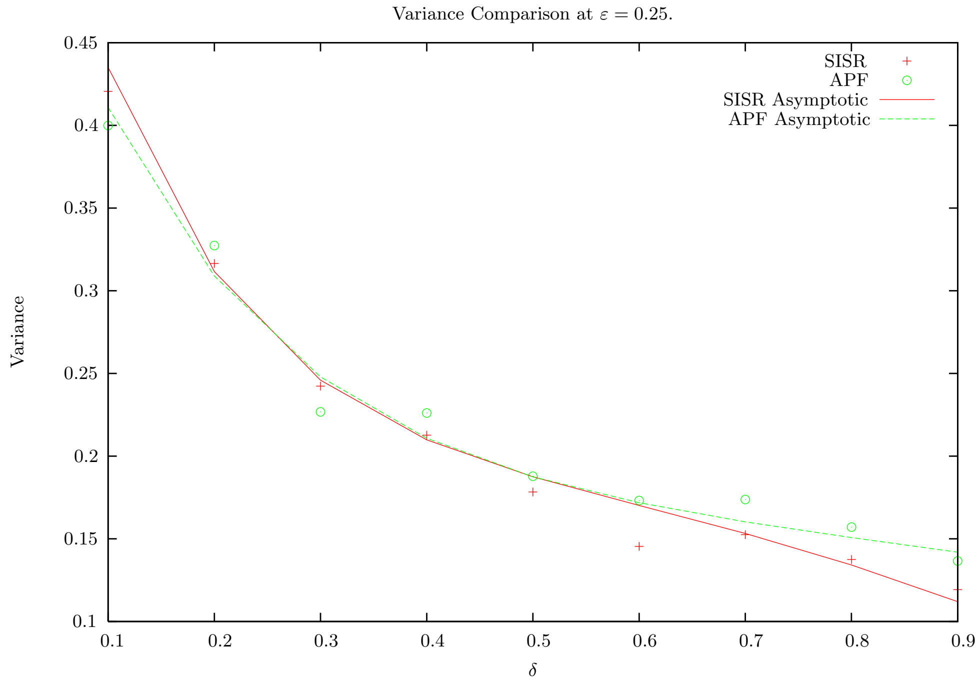

Fig. 2.2. Comparative variance graph: empirical and asymptotic results for the example.

the APF outperforms SISR in terms of the asymptotic variance of its estimates. A number of interesting features can be discerned. Particularly, the APF provides better estimates when δ is small, but exhibits poorer performance when $\delta \sim 1$ and $\varepsilon \sim 0.25$. When $\delta < 0.5$ the observation sequence has low probability under the prior, the APF ameliorates the situation by taking into account the predictive likelihood. The case in which ϵ and δ are both small in, unsurprisingly that in which the APF performs best: the prior probability of the observation sequence is low, but the predictive likelihood is very concentrated.

Whilst it may appear counter-intuitive that the APF can be outperformed by SIR even in the perfect adaptation case, this can perhaps be understood by noting that perfect adaptation is simply a one-step-ahead process. The variance decomposition contains terms propagated forward from all previous times and whilst the adaptation may be beneficial at the time which it is performed it may have a negative influence on the variance at a later point. We also note that, although the APF approach does not dominate SIR, it seems likely to provide better performance in most scenarios.

Figure 2.2 shows experimental and asymptotic variances for the two algorithms. The displayed experimental variances were calculated as N times the empirical variance of 500 runs of each algorithm with $N = 3,000$ particles. This provides an illustration that the asymptotic results provided above do provide a useful performance guide.

2.4 Discussion and Extension

The main idea behind the APF, that is modifying the original sequence of targets distributions to guide particles in promising regions, can be extended outside the filtering framework. Assume we are interested in a sequence of distributions $\{\pi_t(x_{1:t})\}$.

Instead of using the SISR algorithm to sample from it, we use the SISR algorithm on a sequence of distributions $\{\widehat{\pi}_{t+1}(x_{1:t})\}$ where $\widehat{\pi}_{t+1}(x_{1:t})$ is an approximation of

$$\pi_{t+1}(x_{1:t}) = \int \pi_{t+1}(x_{1:t+1})\, dx_{t+1}.$$

We then perform inference with respect to $\pi_t(x_{1:t})$ by using importance sampling with the importance distribution $\widehat{\pi}_{t-1}(x_{1:t-1})\, q_t(x_t|\, x_{1:t-1})$ obtained after the sampling step at time t. This is discussed in more detail in section 3.

We also note that it has been recommended in the literature by a few authors (e.g. [14, pp. 73-74]) to resample the particles not according to their normalized weights associated to $w_t^{\mathrm{SISR}}(x_{1:t}) = \frac{\pi_t(x_{1:t})}{\pi_{t-1}(x_{1:t-1})q_t(\,x_t|x_{t-1})}$ but according to a generic score function $w_t(x_{1:t}) > 0$ at time t

$$w_t(x_{1:t}) = g\left(w_t^{\mathrm{SISR}}(x_{1:t})\right)$$

where $g : \mathbb{R}^+ \to \mathbb{R}^+$ is a monotone function; a common choice being $g(x) = x^\alpha$ where $0 < \alpha \leq 1$. To the best of our knowledge, it has never been specified clearly in the literature that this approach simply corresponds to a standard SISR algorithm for the sequence of distributions

$$\pi_t'(x_{1:t}) \propto g\left(w_t^{\mathrm{SISR}}(x_{1:t})\right) \pi_{t-1}(x_{1:t-1})\, q_t(x_t|\, x_{t-1}).$$

The estimates of expectations with respect to $\pi_t(x_{1:t})$ can then be computed using importance sampling. This approach is rather similar to the APF and could also be easily studied.

3. Auxiliary Sequential Monte Carlo Samplers

In this section we introduce a general framework for a class of algorithms which incorporates a number of previously proposed filtering and sampling procedures, as well as the auxiliary SMC sampler which will be introduced in section 3.4.2.

In section 2, we demonstrated that, in order to allow standard analysis techniques to be applied to the auxiliary particle filter, it is useful to consider it from an angle slightly different to that from which it is usually viewed. Rather than appealing to an auxiliary *variable* technique, we view it as a technique which invokes a sequence of auxiliary *distributions* which are related to those of interest but which do not correspond to them. This sequence of distributions is then used as a sequence of *importance distributions*: a collection of weighted samples from these distributions is reweighted to provide a collection of weighted samples from the filtering distributions.

We further note that, as we are interested in the integrals of test functions φ_t under a sequence of measures π_t which admit a density with respect to μ_t which we term $\widetilde{W}_t$, we may view the integral of φ_t under π_t as being the integral of $\widetilde{W}_t\varphi_t$ under μ_t. Thus we may view estimating the integral of φ_t from a collection of samples from auxiliary distribution μ_t as estimating the integral of the transformed function $\widetilde{W}_t\varphi_t$ under the sampling distribution. This leaves us obtaining a collection of samples from a sequence of distributions via a sequential importance resampling strategy and then using this sequence of samples to estimate the integral of functions.

With this interpretation we are able to treat the APF in very much the same was as SIR and to apply results from the field of Feynman-Kac formulae, it will also allow us to incorporate the APF into the general analytic framework which is introduced in section 3.2. The section has a number of purposes:

– To illustrate that this technique may be extended to the more general SMC samplers framework.
– To show that results obtained for SISR algorithms or SMC samplers can easily be transferred to the other system by noting that SMC samplers are simply SISR algorithms on a larger space, whilst any SISR algorithm may be embedded within the SMC samplers framework.

3.1 Background

The SMC samplers framework of [4] is a very general method for obtaining a set of samples from a sequence of distributions which can exist on the same or different spaces. This is a generalisation of the standard SMC method (commonly referred to as particle filtering and summarised by [7]) in which the target distribution exists on a space of strictly increasing dimension and no mechanism exists for updating the estimates of the state at earlier times after receiving new data.

Given a sequence of distributions $(\mu_t)_{t\geq 1}$ on a sequence of measurable spaces $(E_t, \mathcal{E})_{t\geq 1}$ from which we wish to obtain sets of weighted samples, we construct a sequence of distributions on a sequence of spaces of increasing dimension which admit the distributions of interest as marginals, by defining:

$$\widetilde{\mu}_t(x_{1:t}) = \mu_t(x_t) \prod_{s=t-1}^{1} L_s(x_{s+1}, x_s)$$

where L_s is an arbitrary Markov kernel from space E_{s+1} to E_s (these act, in some sense, backwards in time). It is clear that standard SMC methods can now be applied on this space, by propagating samples forward from one distribution to the next according to a sequence of Markov kernels, $(K_t)_{t\geq 2}$, and correcting for the discrepancy between the proposal and the target distribution by importance sampling. As always it is important to ensure that a significant fraction of the particle set have non-negligible weights. The effective sample size (ESS), introduced by [13], is an approximation obtained by Taylor expansion of a quantity which describes the effective number of iid samples to which the set corresponds. The ESS is defined as $ESS = \left[\sum_{i=1}^{N} {W^{(i)}}^{-2}\right]^{-1}$ where $\left\{W^{(i)}\right\}$ are the normalized weights. This approximation, of course, fails if the particle set does not accurately represent the support of the distribution of interest. Resampling should be carried out after any iteration which causes the ESS to fall below a reasonable threshold (typically around half of the total number of particles), to prevent the sample becoming degenerate with a small number of samples having very large weights.

The rapidly increasing dimension raises the concern that the variance of the importance weights will be extremely high. It can be shown (again, see [4]) that the optimal form for the Markov kernels L_s – in the sense that they minimise the variance of the importance weights if resampling occurs at every time step – depends upon the distributions of interest and the importance sampling proposal kernels K_t in the following way:

$$L_t^{opt}(x_{t+1}, x_t) = \frac{\mu_t(x_t) K_{t+1}(x_t, x_{t+1})}{\int \mu_t(x) K_{t+1}(x, x_{t+1}) dx} \tag{3.1}$$

In practice it is important to choose a sequence of kernels which are as close to the optimal case as possible to prevent the variance of the importance weights from becoming extremely large.

3.2 A Class of Sequential Samplers

We consider a broad class of sequential samplers which we shall term auxiliary SMC (ASMC) samplers. We will show that as well as being of some interest in its own right, studying the general properties of this class of samplers will allow statements to be made about several particularly interesting classes of samplers which are widely used at present, using a strong law of large numbers and central limit theorem as exemplar results.

Definition 3.2.1 (Auxiliary SMC Samplers). *An* auxiliary SMC sampler *for a sequence of distributions π_t on $(E_t, \mathcal{E}_t)$ consists of:*

- *a "conventional" SMC sampler targeting some auxiliary sequence of distributions μ_t from which a sequence of samples are obtained at each iteration.*
- *a sequence of importance weights $\widetilde{W}_t$ proportional to the density of π_t with respect to μ_t.*

Given an N-particle set of weighted samples from the ASMC sampler, we identify the following pair of random measures with the two weighted empirical distributions of interest:

$$\mu_t^N(\cdot) = \frac{\sum\limits_{i=1}^{t} W_t^{(i)} \delta_{X_t^{(i)}}(\cdot)}{\sum\limits_{i=1}^{t} W_t^{(i)}} \qquad \pi_t^N(\cdot) = \frac{\sum\limits_{i=1}^{t} \widetilde{W}_t^{(i)} W_t^{(i)} \delta_{X_t^{(i)}}(\cdot)}{\sum\limits_{i=1}^{t} \widetilde{W}_t^{(i)} W_t^{(i)}},$$

noting that for measurable function φ_n, we may write:

$$\int \varphi_n(x_n)\pi_t^N(x_n)dx_n = \frac{\int \widetilde{W}_t(x_n)\varphi_n(x_n)\mu_t^N(x_n)dx_n}{\int \widetilde{W}_t(x_n)\mu_t^N(x_n)dx_n}.$$

Although this appears complex, in practice this amounts to the machinery for constructing a SMC sampler for the auxiliary sequence of distributions μ_t – which can be done according to the guidelines for kernel selection in [4]. The only additional complication with this method is choosing the sequence of auxiliary distributions μ_t for which to construct the sampler. This is precisely the problem which is addressed, in the filtering case, by the APF and we present a general analogue below. Algorithm 1 illustrates the minimal additional complexity introduced, relative to the SMC sampler.

Algorithm 1 ASMC Sampler

1: $t = 1$
2: **for** $i = 1$ to N **do**
3: $\quad X_1^{(i)} \sim K_1(\cdot)$
4: $\quad W_1^{(i)} \propto \frac{\mu_1(X_1^{(i)})}{K_1(X_1^{(i)})}$
5: **end for**
6: Resample
7: $\int \varphi_t(x_t)\pi_t(dx_t) \approx \frac{\sum_{i=1}^{t} W_t^{(i)}\widetilde{W}_t^{(i)}\varphi_t(X_t^{(i)})}{\sum_{i=1}^{t} W_t^{(i)}\widetilde{W}_t^{(i)}}$.
8: $t \leftarrow t + 1$
9: **for** $i = 1$ to N **do**
10: $\quad X_t^{(i)} \sim K_t(X_{t-1}^{(i)}, \cdot)$
11: $\quad W_t^{(i)} \propto W_{t-1}^{(i)} \frac{\mu_t(X_t^{(i)})L_{t-1}(X_t^{(i)}, X_{t-1}^{(i)})}{\mu_{t-1}(X_{t-1}^{(i)})K_t(X_{t-1}^{(i)}, X_t^{(i)})}$
12: **end for**
13: Go to step 6

3.3 Convergence Results

Here we provide convergence results for ASMC sampler estimates of integrals, before showing in section 3.4 that this sampler is sufficiently general that it may be used to describe the other sampling frameworks of interest. The approach which is taken is to make use of an appropriate Feynman-Kac flow (one for which the particle system of interest may be considered a mean field approximation), thereby allowing us to apply more or less directly the pioneering work of [3]. Although the mathematical novelty is limited (indeed, we take essentially the same approach as that used by [5] in the analysis of the SMC sampler) our principle contribution lies predominantly within the interpretation of the auxiliary variable technique and the results obtained here should have real practical implications. In this section we simply present the results of interest together with an intuitive explanation. The proofs are deferred to appendix B.

The following collection of assumptions is made in order to obtain these convergence results:

Assumption 1. Resampling is conducted at every step according to the multinomial resampling scheme which corresponds to sampling, with replacement, from the weighted collection of particles to obtain an unweighted collection with the same distribution.

Assumption 2. The importance weight is such that the estimator is always well defined. In the simplest case, one may assume that:

$$\begin{aligned} \forall x_{t-1} : \text{ess inf}\, W_t(x_{t-1}, \cdot) &> 0 \\ \text{ess sup}\, W_t(x_{t-1}, \cdot) &< \infty \end{aligned}$$

where the essential infimum and supremum is taken with respect to $K(x_{t-1}, \cdot)$.

Assumption 3. The auxiliary weight is bounded: $0 < \widetilde{W}_t < \infty$.

Assumption 4. φ_t satisfies the regularity conditions given in [3, Chapter 9].

3.3.1 Strong Law of Large Numbers

In order to simplify the presentation and to present results of the form usually seen within statistics, we provide a strong law of large numbers for the integrals of bounded measurable functions. We note, however, that the result which we utilise in order to do this is somewhat more general and we could, in fact, present the convergence of the empirical distribution itself under a broad class of Zolotarev semi-norms with no additional complications.

Theorem 3.3.1 (SLLN). *Given an ASMC sampler, and a bounded measurable function $\varphi : E_t \to \mathbb{R}$, at all iterations $t \geq 1$ the following holds, providing that assumptions 1, 2 and 3 are satisfied:*

$$\lim_{N\to\infty} \left(\int \varphi_t(x_t)\pi_t^N(dx_t) - \int \varphi_t(x_t)\pi_t(dx_t) \right) \overset{a.s.}{\to} 0.$$

Whilst it is convenient and appealing to obtain this result for a broad class of algorithms within a unified framework, it is the next theorem which illustrates the real power of the approach. Via the central limit theorem we obtain a quantified estimate of the variation of the estimator about the true value for ASMC samplers.

3.3.2 Central Limit Theorem

Here we are able to show that, under weak regularity conditions, asymptotically, as the number of particles employed in the sampler tends to infinity, the estimate of the integral of bounded measurable functions provided by an ASMC sampler obeys a central limit theorem with a particular asymptotic variance expression.

Theorem 3.3.2 (CLT). *Given an ASMC sampler, the following central limit theorem holds under weak regularity conditions, for bounded measure* $\varphi_t : E_t \to \mathbb{R}$ *:*

$$\sqrt{N}\left(\int \varphi_t(x_t)\pi_t^N(dx_t) - \int \varphi_t(x_t)\pi_t(dx_t)\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{ASMC}\left(\varphi_t\right)\right), \tag{3.2}$$

where

$$\begin{aligned}\sigma^2_{ASMC}(\varphi_t) = &\int \frac{\widetilde{\pi}_t(x_1)^2}{K_1(x_1)}\left[\int \varphi_t(x_t)\widetilde{\pi}_{x_t|1}(t|x_1)dx_t - \bar{\varphi}_t^{\pi_t}\right]^2 dx_1 + \\ &\sum_{k=2}^{t-1}\int \frac{\left[\widetilde{\pi}_t(x_k)L_{k-1}(x_k,x_{k-1})\right]^2}{\mu_{k-1}(x_{k-1})K_k(x_{k-1},x_k)}\left[\int \varphi_t(x_t)\widetilde{\pi}_{t|k}(x_t|x_k)dx_t - \bar{\varphi}_t^{\pi_t}\right]^2 dx_{k-1:k}, \\ &+\int \frac{\left[\pi_t(x_t)L_{t-1}(x_t,x_{t-1})\right]^2}{\mu_{t-1}(x_{t-1})K_t(x_{t-1},x_t)}\left[\varphi_t(x_t) - \bar{\varphi}_t^{\pi_t}\right]^2 dx_{t-1:t}.\end{aligned} \tag{3.3}$$

where, for the sake of definiteness:

$$\begin{aligned}\widetilde{\pi}_t(x_{1:t}) &= \pi_t(x_t)\prod_{k=1}^{t-1} L_(x_{k+1},x_k) \\ \widetilde{\pi}_t(x_k) &= \int \widetilde{\pi}_t(x_{1:t})dx_{1:k-1}dx_{k+1:t} \\ \widetilde{\pi}_{t|k}(x_t|x_k) &= \frac{1}{\widetilde{\pi}_t(x_k)}\int \widetilde{\pi}_t(x_{1:t})dx_{1:k-1}dx_{k+1:t-1},\end{aligned}$$

and throughout this section[1] *we define* $\bar{\varphi}_t^{\pi_t} = \int \varphi_t(x_t)\pi_t(dx_t)$*, and we will subsequently make use of* $\bar{\varphi}_t^{\mu_t} = \int \varphi_t(x_t)\mu_t(dx_t)$*.*

This variance expression has a very clear interpretation. Indeed, each term in the summation corresponds to precisely the variance of an importance sampling estimator (see, for example, [10]). Thus the variance at any time in the algorithm may be decomposed into a collection of random variables each of which corresponds to the difference obtained by propagating forward the existing, sampled particle set rather than the true measure; each of which may asymptotically be viewed as an importance sampling estimator. This is a particular case of a very general result for interacting particle systems which evolve under the action of sampling and resampling (see [3, Chapters 7-9]).

3.4 Examples

The formulation and convergence results presented above allow us to provide an analysis of the asymptotic behaviour of a class of sampling algorithms.

[1] There is a very slight formal difference between this definition and that adopted in appendix 2 which arises from the slightly different formulation of the problem employed in this section. We prefer to retain this notation as it retains the intention of the notation and the symbol fulfills the same rôle in both places.

We consider two applications of the suggested sampling framework: the first is particle filtering by various techniques, including the auxiliary particle filter. The second is the more recent SMC sampler, for which we present an enhancement which is to the SMC sampler as the APF is to the SIR algorithm. We provide asymptotic variance expressions for all of the algorithms consider, obtaining these directly from the central limit theorem for the ASMC class of samplers.

3.4.1 Application 1: Particle Filtering

It is possible, if apparently overly complicated, to characterise both SIR and the APF as particular cases of the framework described here. The benefit of doing this is that it is possible to analyse both of these approaches within the same framework, allowing their asymptotic variances to be compared, for example. This provides a clear characterisation of the ease with which both SMC samplers and particle filtering algorithms can be analysed concurrently. Note that the following identifications allow us to recover the asymptotic variance expressions obtained directly in section 2.

Sequential Importance Resampling. It is reasonably apparent that the SIR algorithm for approximate filtering can be interpreted within the SMC samplers framework by making suitable identifications. Any SMC sampler then has an interpretation as an ASMC sampler, as explained in section 3.4.2. Consequently, with the following identifications, we obtain an ASMC sampler which corresponds precisely to the SIR algorithm described above (note that, formally, we operate on an increasing sequence of spaces which parameterise the full path of the filtering distribution at each time):

$$\begin{aligned}
\pi_t(x_{1:t}) &= \mu_t(x_{1:t}) = p(x_{1:t}|y_{1:n}) \\
K_t(x_{1:t-1}, x'_{1:t}) &= \delta_{x_{1:t-1}}(x'_{1:t-1}) q_t(x'_t|x_{t-1}) \\
L_{t-1}(x'_{1:t}, x_{1:t-1}) &= \frac{\pi_{t-1}(x_{1:t-1}) K_t(x_{1:t-1}, x'_{1:t})}{\int \pi_{t-1}(z_{1:t-1}) K_t(z_{1:t-1}, x'_{1:t}) dz_{1:t}} \\
W_t(x_{1:t-1}, x'_{1:t}) &= \frac{\pi_t(x'_{1:t}) \delta_{x_{1:t-1}}(x'_{1:t-1})}{\pi_{t-1}(z_{1:t-1}) K_t(z_{1:t-1}, x'_{1:t}) dz_{1:t}} = \frac{g(y_t|x'_t) f(x'_t|x'_{t-1})}{q_t(x_t|x_{t-1})} \\
\widetilde{W}_t(x_{1:t}) &= \mathbf{1}(x_{1:t}),
\end{aligned}$$

we note that in this case, as in all cases in which the proposal kernel simply extends an estimate of the state vector up to time $t-1$ to one up to time t, it is possible to use the optimal form for the auxiliary kernel of the SMC sampler, and this reduces the effective space upon which importance sampling is performed to that of the time marginal.

If we further employ the conditional prior as the proposal distribution, setting $q_t(x_t|x_{t-1}) = f(x_t|x_{t-1})$ then we obtain the bootstrap filter.

Auxiliary Particle Filters. In the case of the auxiliary particle filter, if we once again allow $\pi_t(x_{1:t}) = p(x_{1:t}|y_{1:n})$ but construct an SMC sampler for the sequence of auxiliary distributions μ_n defined as follows, we obtain the interpretation which we require with the following definitions:

$$\begin{aligned}
\mu_t(x_{1:t}) &= \pi_t(x_{1:t})\hat{p}_{t+1}(y_{t+1}|x_t)\\
K_t(x_{1:t-1}, x'_{1:t}) &= \delta_{x_{1:t-1}}(x'_{1:t-1})q_t(x'_t|x_{t-1})\\
L_{t-1}(x'_{1:t}, x_{1:t-1}) &= \frac{\mu_{t-1}(x_{1:t-1})K_t(x_{1:t-1}, x'_{1:t})}{\int \mu_{t-1}(z_{1:t-1})K_t(z_{1:t-1}, x'_{1:t})dz_{1:t}}\\
W_t(x_{1:t-1}, x'_{1:t}) &= \frac{\mu_t(x'_{1:t})\delta_{x_{1:t-1}}(x'_{1:t-1})}{\mu_{t-1}(z_{1:t-1})K_t(z_{1:t-1}, x'_{1:t})dz_{1:t}} \propto \frac{g(y_t|x'_t)f(x'_t|x'_{t-1})\hat{p}_{t+1}(y_{t+1}|x'_t)}{q_t(x'_t|x_{t-1})\hat{p}_t(x_t|y_{t-1})}\\
\widetilde{W}_t(x'_{1:t}) &\propto \hat{p}_{t+1}(y_{t+1}|x'_t)^{-1},
\end{aligned}$$

again, it is possible to use the optimal auxiliary kernel.

3.4.2 Application 2: SMC Samplers

In comparison with section 3.4.1 this application seems relatively natural. SMC samplers, and an extension thereof, both fit directly into the proposed framework, allowing us to reproduce the familiar expression for the variance of the SMC sampler (this is essentially the same result as that presented in [4]) and to provide an asymptotic variance expression for a novel algorithm which we term the auxiliary variable SMC sampler, which combines elements of the SMC sampler with the auxiliary variable strategy of the APF.

SMC Samplers. In order to cast the standard SMC sampler into the present framework, one simply selects $\mu_t = \pi_t$ and the additional weight function is then the unit function, $\widetilde{W}_t(x_t) = 1$.

This provides, as one would expect, the usual asymptotic variance expression:

$$\begin{aligned}
\sigma^2_{SMC,t}(\varphi_t) = &\int \frac{\widetilde{\pi}_t(x_1)^2}{K_1(x_1)}\left(\int \varphi_t(x_t)\widetilde{\pi}_{t|1}(x_t|x_1)dx_t - \bar{\varphi}_t^{\pi_t}\right)^2 dx_1 +\\
&\sum_{k=2}^{t-1}\int \frac{\widetilde{\pi}_t(x_k)^2 L_{k-1}(x_k, x_{k-1})^2}{\pi_{k-1}(x_{k-1})K_k(x_{k-1}, x_k)}\left(\int \varphi_t(x_t)\widetilde{\pi}_{t|k}(x_t|x_k)dx_t - \bar{\varphi}_t^{\pi_t}\right)^2 dx_{k-1,k}\\
&+ \int \frac{\widetilde{\pi}_t(x_t)^2 L_{t-1}(x_t, x_{t-1})^2}{\pi_{t-1}(x_{t-1})K_t(x_{t-1}, x_t)}\left(\varphi_t(x_t) - \bar{\varphi}_t^{\pi_t}\right)^2 dx_{t-1,t}.
\end{aligned}$$

Auxiliary Variable SMC Samplers. An obvious extension of the SMC sampler follows by considering how to emulate the approach of the APF within a more general framework. Essentially one constructs an SMC sampler for an auxiliary sequence of distributions which is defined implicitly by a pre-weighting at time $t-1$ which, after resampling, is intended to place more mass in regions which lead to good solutions at time t.

A Simple Example. As has been previously noted, [4], in a setting in which one has a fixed state space, $E_t = E$ at every iteration, and employs a MCMC kernel of invariant distribution π_t as the proposal, and makes use of the auxiliary kernel:

$$L_{t-1}(x_t, x_{t-1}) = \frac{\pi_t(x_{t-1})K_t(x_{t-1}, x_t)}{\pi_t(x_t)},$$

the importance weights are simply $W_t(x_{t-1}, x_t) = \pi_t(x_{t-1})/\pi_{t-1}(x_{t-1})$ which is independent of the proposed state, x_t.

Consequently, it is intuitively clear that one should resample *before* proposing new states in the interests of maximising sample diversity. This has been observed previously, [12]. Here we are able to incorporate this approach into our general framework and provide convergence results and an asymptotic variance expression. By making the following identifications, we obtain an ASMC sampler which coincides exactly with this approach:

$$\begin{aligned}
\mu_t(x_t) &= \pi_{t+1}(x_t) \\
L_{t-1}(x_t, x_{t-1}) &= \frac{\mu_{t-1}(x_{t-1})K_t(x_{t-1}, x_t)}{\mu_{t-1}(x_t)} = \frac{\pi_t(X_{t-1})K_t(X_{t-1}, X_t)}{\pi_t(X_t)} \\
W_t(x_{t-1}, x_t) &= \frac{\mu_t(x_t)}{\mu_{t-1}(x_t)} = \frac{\pi_{t+1}(x_t)}{\pi_t(x_t)} \\
\widetilde{W}_t(x_t) &= \mu_{t-1}(x_t)/\mu_t(x_t) = \pi_t(x_t)/\pi_{t+1}(x_t),
\end{aligned}$$

note that in this particular special case $W_t(x_{t-1}, x_t)\widetilde{W}_t(x_t) = 1$ and the weights used in the importance sampling step of the algorithm are uniform (of course, this is not the case for the resampling weights, $W_t(x_{t-1}, x_t)$). We have, in this case, that:

$$\begin{aligned}
\sigma^2_{ASMC,t}(\varphi_t) = &\int \frac{\widetilde{\pi}_t(x_1)}{K_1(x_1)} \left[\int \varphi_t(x_t)\widetilde{\pi}_{t|1}(x_t|x_1)dx_t - \bar{\varphi}_t^{\pi_t} \right]^2 dx_1 + && (3.4) \\
&\sum_{k=2}^{t-1} \int \frac{\widetilde{\pi}_t(x_k)^2}{\pi_k(x_k)} \left[\int \varphi_t(x_t)\widetilde{\pi}_{t|k}(x_t|x_k)dx_t - \bar{\varphi}_t^{\pi_t} \right]^2 dx_k && (3.5) \\
&+ \int \pi_t(x_t) \left[\varphi_t(x_t) - \bar{\varphi}_t^{\pi_t}\right]^2 dx_t. && (3.6)
\end{aligned}$$

where the independence of the importance weights from the state after the mutation step is exhibited in the decoupling of the terms in the asymptotic variance expression.

A More General Approach:. In general one seeks a sequence of auxiliary distributions μ_t and associated proposal and auxiliary kernels K_t and L_{t-1} which are such that the importance sampling weights which are obtained:

$$W_t(x_{t-1}, x_t)\widetilde{W}_t(x_t) = \frac{\mu_t(x_t)L_{t-1}(x_t, x_{t-1})}{\mu_{t-1}(x_{t-1})K_t(x_{t-1}, x_t)} \frac{\pi_t(x_t)}{\mu_t(x_t)} = \frac{\pi_t(x_t)L_{t-1}(x_t, x_{t-1})}{\mu_{t-1}(x_{t-1})K_t(x_{t-1}, x_t)},$$

have lower variance than those which would be obtained by using a standard SMC sampler targeting π_t directly.

To distinguish these approaches from the ASMC sampler introduced above, we term it the auxiliary variable SMC (AvSMC) sampler, and it may be interpreted in very much the same manner as the auxiliary variable particle filter. The asymptotic variance expression for the fully general case cannot be given in a simpler form than that presented in theorem 3.3.2.

References

[1] J. Carpenter, P. Clifford, and P. Fearnhead. An improved particle filter for non-linear problems. *IEEE Proceedings on Radar, Sonar and Navigation*, 146(1):2–7, 1999.

[2] N. Chopin. Central limit theorem for sequential Monte Carlo methods and its applications to Bayesian inference. *Annals of Statistics*, 32(6):2385–2411, December 2004.

[3] P. Del Moral. *Feynman-Kac formulae: genealogical and interacting particle systems with applications*. Probability and Its Applications. Springer Verlag, New York, 2004.

[4] P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo methods for Bayesian Computation. In *Bayesian Statistics 8*. Oxford University Press, 2006.

[5] P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society B*, 63(3):411–436, 2006.

[6] P. Del Moral, A. Doucet, and G. Peters. Sharp propagation of chaos estimates for feynman-kac particle models. *Teoriya Veroyatnostei i ee Primeneniya*, 51, 2006. (to be reprinted in SIAM *Theory of Probability and Its Applications*.

[7] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Statistics for Engineering and Information Science. Springer Verlag, New York, 2001.

[8] A. Doucet, S. Godsill, and C. Andrieu. On sequential simulation-based methods for Bayesian filtering. *Statistics and Computing*, 10(3):197–208, 2000.

[9] P. Fearnhead, O. Papaspiliopoulos, and G. O. Roberts. Particle filters for partially-observed diffusion. Technical report, University of Lancaster, 2007.

[10] J. Geweke. Bayesian inference in econometrics models using Monte Carlo integration. *Econometrica*, 57(6):1317–1339, November 1989.

[11] S. Godsill and T. Clapp. Improvement strategies for Monte Carlo particle filters. In Doucet et al. [7], pages 139–158.

[12] A. M. Johansen. *Some Non-Standard Sequential Monte Carlo Methods With Applications*. Ph.D. thesis, University of Cambridge Department of Engineering, 2006.

[13] A. Kong, J. S. Liu, and W. H. Wong. Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association*, 89(425):278–288, March 1994.

[14] J. S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer Series in Statistics. Springer Verlag, New York, 2001.

[15] M. K. Pitt and N. Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94(446):590–599, 1999.

[16] M. K. Pitt and N. Shephard. Auxiliary variable based particle filters. In Doucet et al. [7], chapter 13, pages 273–293.

A. Inductive Proofs

We present proofs of the results presented in section 2 obtained by applying the results of [2]. We decompose the proof of proposition 2.2.1 into two parts: Initially, we show that the recursive variance expression obtained in [2] may be written as an explicit sum and that this recursion holds for any sufficiently regular sequences of sampling distributions. We then show that the importance sampling estimator provided by the APF also obeys a central limit theorem whose variance can be obtained by the same techniques.

A.1 SIR Variance

We begin by illustrating that the variance recursion of [2] may be rewritten explicitly in a form reminiscent of that obtained by the direct study of the underlying Feynman-Kac flow.

Given an SIR algorithm targeting the sequence of distributions $\{\pi_t\}$ using q_1 as the initial proposal distribution and $\{q_t\}_{t\geq 2}$ as the transition kernel proposals, [2, Section 2.3] illustrates that the following three sequences of empirical measures obey central limit theorems:

$$\rho_t^N(dx_{1:t}) = \frac{1}{N}\sum_{i=1}^N \delta_{\left(\check{X}_{1:t-1,t-1}^{(i)}, X_{t,t}^{(i)}\right)}(dx_{1:t})$$

$$\pi_t^N(dx_{1:t}) = \sum_{i=1}^N W_t^{(i)} \delta_{\left(\check{X}_{1:t-1,t-1}^{(i)}, X_{t,t}^{(i)}\right)}(dx_{1:t})$$

$$\widetilde{\pi}_t^N(dx_{1:t}) = \frac{1}{N}\sum_{i=1}^N \delta_{\left(\check{X}_{1:t,t}^{(i)}\right)}(dx_{1:t})$$

corresponding to the particle system after mutation (sampling from the importance distribution), correction (importance weighting) and selection (resampling), respectively. Both π_t^N and $\widetilde{\pi}_t^N$ provide approximations to the target measure π_t, whilst $\rho_t^N(dx_{1:t})$ approximates $\rho_t(dx_{1:t}) := \pi_{t-1}(dx_{1:t-1})q_t(dx_t|x_{t-1})$.

The central limit theorem of [2] tells us that under mild regularity conditions, the following central limit theorem holds for a broad class of test functions $\varphi_t : E_{1:t} \to \mathbb{R}$:

$$\sqrt{N}\left[\int \varphi_t(x_{1:t})\pi_t^N(dx_{1:t}) - \int \varphi_t(x_{1:t})\pi_t(dx_{1:t})\right] \xrightarrow{d} \mathcal{N}\left(0, V_t(\varphi_t)\right)$$

where the variance expression is obtained from the following recursion[1]:

$$\begin{aligned}
\widetilde{V}_t(\varphi_t) &= \hat{V}_{t-1}\left(\mathbb{E}_{\rho_t}\left[\varphi_t \,|\sigma(X_{1:t-1})\right]\right) + \mathbb{E}_{\rho_t}\left[\mathsf{Var}_{\rho_t}\left[\varphi_t \,|\sigma(X_{1:t-1})\right]\right] \\
V_t(\varphi_t) &= \widetilde{V}_t\left(W_t\left(\varphi_t - \mathbb{E}_{\pi_t}\left[\varphi_t\right]\right)\right) \\
\hat{V}_t(\varphi_t) &= V_t(\varphi_t) + \mathsf{Var}_{\pi_t}\left[\varphi_t\right].
\end{aligned}$$

Proposition A.1.1. *We will show, inductively, that $\widetilde{W}_t(\varphi_t)$ may be written in a closed form as:*

$$\begin{aligned}
V_t(\varphi_3) = &\sum_{k=1}^{t-1} \mathsf{Var}_{\rho_k}\left[W_k\left(\mathbb{E}_{\pi_t}\left[\varphi_t \,|\sigma(X_{1:k})\right] - \bar{\varphi}_t^{\pi_t}\right)\right] + \\
&\mathsf{Var}_{\rho_t}\left[W_t(\varphi_t - \bar{\varphi}_t^{\pi_t})\right].
\end{aligned} \tag{A.1}$$

where it has been convenient to define $\bar{\varphi}_t^{\pi_t} = \int \varphi_t(x_{1:t})\pi_t(dx_{1:t})$ (we will further define $\bar{\varphi}_t^{\rho_t} = \int \varphi_t(x_{1:t})\rho_t(dx_{1:t})$).

Proof. We begin by considering the initialisation. At time 1, we have a self-normalised importance sampling estimator and so (see [10]):

$$V_1(\varphi_1) = \mathsf{Var}_{\rho_1}\left[W_1(\varphi_1 - \bar{\varphi}_1^{\pi_1})\right],$$

and after resampling we have:

$$\hat{V}_1(\varphi_1) = \mathsf{Var}_{\rho_1}\left[W_1(\varphi_1 - \bar{\varphi}_1^{\pi_1})\right] + \mathsf{Var}_{\pi_1}\left[\varphi_1\right].$$

At iteration two, we obtain:

$$\begin{aligned}
\widetilde{V}_2(\varphi_2) =& \hat{V}_1\left(\mathbb{E}_{\rho_2}\left[\varphi_2 \,|\sigma(X_1)\right]\right) + \mathbb{E}_{\rho_2}\left[\mathsf{Var}_{\rho_2}\left[\varphi_2 \,|\sigma(X_1)\right]\right] \\
=& \mathsf{Var}_{\rho_1}\left[W_1\left(\mathbb{E}_{\rho_2}\left[\varphi_2 \,|\sigma(X_1)\right] - \bar{\varphi}_2^{\rho_2}\right)\right] + \\
& \mathsf{Var}_{\pi_1}\left[\mathbb{E}_{\rho_2}\left[\varphi_2 \,|\sigma(X_1)\right]\right] + \mathbb{E}_{\rho_2}\left[\mathsf{Var}_{\rho_2}\left[\varphi_2 \,|\sigma(X_1)\right]\right] \\
=& \mathsf{Var}_{\rho_1}\left[W_1\left(\mathbb{E}_{\rho_2}\left[\varphi_2 \,|\sigma(X_1)\right] - \bar{\varphi}_2^{\rho_2}\right)\right] + \mathsf{Var}_{\rho_2}\left[\varphi_2\right]
\end{aligned}$$

where the final line follows from two observations:

- The marginal distribution of X_1 under the target measure at time 1 and the uncorrected measure at time two are identical: $\rho_2(dx_1) = \pi_1(dx_1)$.
- As is well known, given any σ-algebra, $\mathcal{F}$, and any function φ one has $\mathsf{Var}\left[\varphi\right] = \mathbb{E}\left[\mathsf{Var}\left[\varphi|\mathcal{F}\right]\right] + \mathsf{Var}\left[\mathbb{E}\left[\varphi|\mathcal{F}\right]\right]$.

After correction, noting that $\mathbb{E}_{\rho_2}\left[W_2(\varphi_2 - \bar{\varphi}_2^{\pi_2})\right] = 0$, we obtain the asymptotic variance of interest at $t = 2$:

$$\begin{aligned}
V_2(\varphi_2) &= \widetilde{V}_2\left(W_2\left(\varphi_2 - \bar{\varphi}_2^{\pi_2}\right)\right) \\
&= \mathsf{Var}_{\rho_1}\left[W_1 \mathbb{E}_{\rho_2}\left[W_2\left(\varphi_2 - \bar{\varphi}_2^{\pi_2}\right) |\sigma(X_1)\right]\right] + \mathsf{Var}_{\rho_2}\left[W_2\left(\varphi_2 - \bar{\varphi}_2^{\pi_2}\right)\right] \\
&= \mathsf{Var}_{\rho_1}\left[W_1 \mathbb{E}_{\pi_2}\left[\varphi_2 - \bar{\varphi}_2^{\pi_2} \,|\sigma(X_1)\right]\right] + \mathsf{Var}_{\rho_2}\left[W_2\left(\varphi_2 - \bar{\varphi}_2^{\pi_2}\right)\right].
\end{aligned}$$

And, after resampling:

[1] This has been written in a slightly different form to the original to emphasize the rôle of ρ_t for reasons which will become apparent.

$$\hat{V}_2(\varphi_2) = \mathsf{Var}_{\rho_1}\left[W_1 \mathbb{E}_{\pi_2}\left[\varphi_2 - \bar{\varphi}_2^{\pi_2} \,|\sigma(X_1)\right]\right] + \mathsf{Var}_{\rho_2}\left[W_2\left(\varphi_2 - \bar{\varphi}_2^{\pi_2}\right)\right] + \mathsf{Var}_{\pi_2}\left[\varphi_2\right].$$

As a final preliminary we will obtain the variance expression for $t = 3$ directly as this will simplify the induction process.

$$\begin{aligned}
\widetilde{V}_3(\varphi_3) =& \hat{V}_2\left(\mathbb{E}_{\rho_3}\left[\varphi_3 \,|\sigma(X_{1:2})\right]\right) + \mathbb{E}_{\rho_3}\left[\mathsf{Var}_{\rho_3}\left[\varphi_3 \,|\sigma(X_{1:2})\right]\right] \\
=& \mathsf{Var}_{\rho_1}\left[W_1 \mathbb{E}_{\pi_2}\left[\mathbb{E}_{\rho_3}\left[\varphi_3 \,|\sigma(X_{1:2})\right] - \bar{\varphi}_3^{\rho_3} \,|\sigma(X_1)\right]\right] + \\
& \mathsf{Var}_{\rho_2}\left[W_2\left(\mathbb{E}_{\rho_3}\left[\varphi_3 \,|\sigma(X_{1:2})\right] - \bar{\varphi}_3^{\rho_3}\right)\right] + \\
& \mathsf{Var}_{\pi_2}\left[\mathbb{E}_{\rho_3}\left[\varphi_3 \,|\sigma(X_{1:2})\right]\right] + \mathbb{E}_{\rho_3}\left[\mathsf{Var}_{\rho_3}\left[\varphi_3 \,|\sigma(X_{1:2})\right]\right] \\
=& \mathsf{Var}_{\rho_1}\left[W_1\left(\mathbb{E}_{\rho_3}\left[\varphi_3 \,|\sigma(X_1)\right] - \bar{\varphi}_3^{\rho_3}\right)\right] + \\
& \mathsf{Var}_{\rho_2}\left[W_2\left(\mathbb{E}_{\rho_3}\left[\varphi_3 \,|\sigma(X_{1:2})\right] - \bar{\varphi}_3^{\rho_3}\right)\right] + \\
& \mathsf{Var}_{\rho_3}\left[\varphi_3\right],
\end{aligned}$$

which just leaves the correction step (noting that $\mathbb{E}_{\rho_3}\left[W_3(\varphi_3 - \bar{\varphi}_3^{\pi_3})\right] = 0$):

$$\begin{aligned}
V_3(\varphi_3) =& \widetilde{V}_3\left(W_3\left(\varphi_3 - \bar{\varphi}_3^{\pi_3}\right)\right) \\
=& \mathsf{Var}_{\rho_1}\left[W_1\left(\mathbb{E}_{\pi_3}\left[\varphi_3 \,|\sigma(X_1)\right] - \bar{\varphi}_3^{\pi_3}\right)\right] + \\
& \mathsf{Var}_{\rho_2}\left[W_2\left(\mathbb{E}_{\pi_3}\left[\varphi_3 \,|\sigma(X_{1:2})\right] - \bar{\varphi}_3^{\pi_3}\right)\right] + \\
& \mathsf{Var}_{\rho_3}\left[W_3(\varphi_3 - \bar{\varphi}_3^{\pi_3})\right].
\end{aligned}$$

Now, assume that equation (A.1) holds at t, then:

$$\begin{aligned}
\hat{V}_t(\varphi_t) =& \sum_{k=1}^{t-1} \mathsf{Var}_{\rho_k}\left[W_k\left(\mathbb{E}_{\pi_t}\left[\varphi_t \,|\sigma(X_{1:k})\right] - \bar{\varphi}_t^{\pi_t}\right)\right] + \\
& \mathsf{Var}_{\rho_t}\left[W_t(\varphi_t - \bar{\varphi}_t^{\pi_t})\right] + \mathsf{Var}_{\pi_t}\left[\varphi_t\right]
\end{aligned}$$

and after the proposal step at time $t + 1$, we obtain:

$$\begin{aligned}
\widetilde{V}_{t+1}(\varphi_{t+1}) =& \hat{V}_t\left(\mathbb{E}_{\rho_{t+1}}\left[\varphi_{t+1} \,|\sigma(X_{1:t})\right]\right) + \mathbb{E}_{\rho_{t+1}}\left[\mathsf{Var}_{\rho_{t+1}}\left[\varphi_{t+1} \,|\sigma(X_{1:t})\right]\right] \\
=& \sum_{k=1}^{t-1} \mathsf{Var}_{\rho_k}\left[W_k\left(\mathbb{E}_{\pi_t}\left[\mathbb{E}_{\rho_{t+1}}\left[\varphi_{t+1} \,|\sigma(X_{1:t})\right] |\sigma(X_{1:k})\right] - \bar{\varphi}_{t+1}^{\rho_{t+1}}\right)\right] + \\
& \mathsf{Var}_{\rho_t}\left[W_t\left(\mathbb{E}_{\rho_{t+1}}\left[\varphi_{t+1} \,|\sigma(X_{1:t})\right] - \bar{\varphi}_{t+1}^{\rho_{t+1}}\right)\right] + \\
& \mathsf{Var}_{\pi_t}\left[\mathbb{E}_{\rho_{t+1}}\left[\varphi_{t+1} \,|\sigma(X_{1:t})\right]\right] + \mathbb{E}_{\rho_{t+1}}\left[\mathsf{Var}_{\rho_{t+1}}\left[\varphi_{t+1} \,|\sigma(X_{1:t})\right]\right] \\
=& \sum_{k=1}^{t} \mathsf{Var}_{\rho_k}\left[W_k\left(\mathbb{E}_{\rho_{t+1}}\left[\varphi_{t+1} \,|\sigma(X_{1:k})\right] - \bar{\varphi}_{t+1}^{\rho_{t+1}}\right)\right] + \\
& \mathsf{Var}_{\rho_{t+1}}\left[\varphi_{t+1}\right]
\end{aligned}$$

all that remains now is the correction step:

$$\begin{aligned}
V_{t+1}(\varphi_{t+1}) =& \widetilde{V}_{t+1}\left(W_{t+1}\left(\varphi_{t+1} - \bar{\varphi}_{t+1}^{\pi_{t+1}}\right)\right) \\
=& \sum_{k=1}^{t} \mathsf{Var}_{\rho_k}\left[W_k \mathbb{E}_{\rho_{t+1}}\left[W_{t+1}\left(\varphi_{t+1} - \bar{\varphi}_{t+1}^{\pi_{t+1}}\right) |\sigma(X_{1:k})\right]\right] + \\
& \mathsf{Var}_{\rho_{t+1}}\left[W_{t+1}(\varphi_{t+1} - \bar{\varphi}_{t+1}^{\pi_{t+1}})\right] \\
=& \sum_{k=1}^{t} \mathsf{Var}_{\rho_k}\left[W_k\left(\mathbb{E}_{\pi_{t+1}}\left[\varphi_{t+1} \,|\sigma(X_{1:k})\right] - \bar{\varphi}_{t+1}^{\pi_{t+1}}\right)\right] + \\
& \mathsf{Var}_{\rho_{t+1}}\left[W_{t+1}(\varphi_{t+1} - \bar{\varphi}_{t+1}^{\pi_{t+1}})\right].
\end{aligned}$$

This completes the induction argument.

A.1.1 Standard SIR

Note that the estimator used in the standard SIR algorithm is precisely that studied in proposition A.1.1 in which:

$$\pi_t(dx_{1:t}) = p(dx_{1:t}|y_{1:t})$$
$$\rho_t(dx_{1:t}) = p(dx_{1:t-1}|y_{1:t-1})q(dx_t|x_{t-1})$$

and, as always, $W_t(x_{1:t}) = \frac{d\pi_t}{d\rho_t}(x_{1:t})$. Substituting these expressions into equation (A.1), and noting that all of the variances are of centred quantities, we obtain:

$$\begin{aligned}
&\sigma^2_{SISR}(\varphi_t) = V_t(\varphi_t) \\
=&\sum_{k=1}^{t-1} \mathsf{Var}_{\rho_k}\left[W_k\left(\mathbb{E}_{\pi_t}\left[\varphi_t\left|\sigma(X_{1:k})\right.\right] - \bar{\varphi}_t^{\pi_t}\right)\right] + \mathsf{Var}_{\rho_t}\left[W_t(\varphi_t - \bar{\varphi}_t^{\pi_t})\right]. \\
=&\int \frac{p(x_1|y_{1:t})^2}{q(x_1)} \int \left(\varphi_t(x_{1:t}) - \bar{\varphi}_t\right)^2 p(x_{2:t}|y_{2:t}, x_1) dx_{2:t} dx_1 + \\
&\sum_{k=2}^{t-1} \int \frac{p(x_{1:k}|y_{1:t})^2}{p(x_{1:k-1}|y_{1:k-1})q_k(x_k|x_{k-1})} \int \left(\varphi_t(x_{1:t}) - \bar{\varphi}_t\right)^2 p(x_{k+1:t}|y_{k+1:t}, x_k) dx_{k+1:t} dx_{1:k} + \\
&\int \frac{p(x_{1:t}|y_{1:t})^2}{p(x_{1:t-1}|y_{1:t-1})q_t(x_t|x_{t-1})} \left(\varphi_t(x_{1:t} - \bar{\varphi}_t\right)^2 dx_{1:t},
\end{aligned}$$

which is the first part of proposition 2.2.1

A.1.2 APF via SIR

In the case of the APF, the underlying SIR flow is constructed with:

$$\pi_t(dx_{1:t}) = \hat{p}(dx_{1:t}|y_{1:t+1})$$
$$\rho_t(dx_{1:t}) = \hat{p}(dx_{1:t-1}|y_{1:t})q(dx_t|x_{t-1}),$$

and so we obtain from proposition A.1.1,

$$\begin{aligned}
V_t(\varphi_t) =& \int \frac{\hat{p}(x_1|y_{1:t+1})^2}{q_1(x_1)} \left(\int \varphi_t(x_{1:t}) \hat{p}(x_{2:t}|y_{2:t+1}, x_1) dx_{2:t} - \hat{\varphi}_t \right)^2 dx_1 \\
&+ \sum_{k=2}^{t-1} \int \frac{\hat{p}(x_{1:k}|y_{1:t+1})^2}{\hat{p}(x_{1:k-1}|y_{1:k})q_k(x_k|x_{k-1})} \left(\int \varphi_t(x_{1:t}) \hat{p}(x_{k+1:t}|y_{k+1:t+1}, x_k) dx_{k+1:t} - \hat{\varphi}_t \right)^2 dx_{1:k} \\
&+ \int \frac{\hat{p}(x_{1:t}|y_{1:t+1})^2}{\hat{p}(x_{1:t-1}|y_{1:t})q_t(x_t|x_{t-1})} \left(\varphi_t(x_{1:t}) - \hat{\varphi}_t\right)^2 dx_{1:t}.
\end{aligned}$$

We can obtain the corresponding asymptotic variance for the importance sampling estimate under p by an additional application of Chopin's correction lemma (A2) as this importance sampling step takes precisely the same form as the standard one. If we do this, noting that in our case this importance weight function is proportional to $\widetilde{W}_t(x_{1:t}) = p(x_{1:t}|y_{1:t})/\hat{p}(x_{1:t}|y_{1:t+1})$, we note that:

$$\sigma^2_{APF}(\varphi_t) = V_t\left(\widetilde{W}_t[\varphi_t - \bar{\varphi}_t]\right), \tag{A.2}$$

where $\bar{\varphi}_t = \int \varphi_t(x_{1:t})p(x_{1:t}|y_{1:t})$. Substituting $\varphi_t \leftarrow \widetilde{W}_t[\varphi_t - \bar{\varphi}_t]$ into the appropriate variance expression, noting that in this instance $\hat{\varphi}_t = 0$, we obtain for the first of the three terms in this variance expression:

$$
\begin{aligned}
&\int \frac{\hat{p}(x_1|y_{1:t+1})^2}{q_1(x_1)} \left(\int [\varphi_t(x_{1:t}) - \bar{\varphi}_t] \frac{p(x_{1:t}|y_{1:t})}{\hat{p}(x_{1:t}|y_{1:t+1})} \hat{p}(x_{2:t}|y_{2:t+1}, x_1) dx_{2:t} \right)^2 dx_1 \\
=& \int \frac{\hat{p}(x_1|y_{1:t+1})^2}{q_1(x_1)} \left(\int [\varphi_t(x_{1:t}) - \bar{\varphi}_t] \frac{p(x_1|y_{1:t}) p(x_{2:t}|y_{2:t}, x_1)}{\hat{p}(x_1|y_{1:t+1}) \hat{p}(x_{2:t}|y_{2:t+1}, x_1)} \hat{p}(x_{2:t}|y_{2:t+1}, x_1) dx_{2:t} \right)^2 dx_1 \\
=& \int \frac{p(x_1|y_{1:t})^2}{q_1(x_1)} \left(\int [\varphi_t(x_{1:t}) - \bar{\varphi}_t] \frac{p(x_{2:t}|y_{2:t}, x_1)}{\hat{p}(x_{2:t}|y_{2:t+1}, x_1)} \hat{p}(x_{2:t}|y_{2:t+1}, x_1) dx_{2:t} \right)^2 dx_1 \\
=& \int \frac{p(x_1|y_{1:t})^2}{q_1(x_1)} \left(\int [\varphi_t(x_{1:t}) - \bar{\varphi}_t] p(x_{2:t}|y_{2:t}, x_1) dx_{2:t} \right)^2 dx_1.
\end{aligned}
$$

The elements of the second term becomes

$$
\begin{aligned}
&\int \frac{\hat{p}(x_{1:k}|y_{1:t+1})^2}{\hat{p}(x_{1:k-1}|y_{1:k}) q_k(x_k|x_{k-1})} \left(\int \frac{p(x_{1:t}|y_{1:t})}{\hat{p}(x_{1:t}|y_{1:t+1})} [\varphi_t(x_{1:t}) - \bar{\varphi}_t] \hat{p}(x_{k+1:t}|y_{k+1:t+1}, x_k) dx_{k+1:t} \right)^2 dx_{1:k} \\
=& \int \frac{\hat{p}(x_{1:k}|y_{1:t+1})^2}{\hat{p}(x_{1:k-1}|y_{1:k}) q_k(x_k|x_{k-1})} \\
&\left(\int \frac{p(x_{1:k}|y_{1:t}) p(x_{k+1:t}|y_{k+1:t}, x_k)}{\hat{p}(x_{1:k}|y_{1:t+1}) \hat{p}(x_{k+1:t}|y_{k+1:t+1}, x_k)} [\varphi_t(x_{1:t}) - \bar{\varphi}_t] \hat{p}(x_{k+1:t}|y_{k+1:t+1}, x_k) dx_{k+1:t} \right)^2 dx_{1:k} \\
=& \int \frac{p(x_{1:k}|y_{1:t})^2}{\hat{p}(x_{1:k-1}|y_{1:k}) q_k(x_k|x_{k-1})} \left(\int [\varphi_t(x_{1:t}) - \bar{\varphi}_t] p(x_{k+1:t}|y_{k+1:t}, x_k) dx_{k+1:t} \right)^2 dx_{1:k}.
\end{aligned}
$$

Finally, the third term is:

$$
\begin{aligned}
&\int \frac{\hat{p}(x_{1:t}|y_{1:t+1})^2}{\hat{p}(x_{1:t-1}|y_{1:t}) q_t(x_t|x_{t-1})} \left(\frac{p(x_{1:t}|y_{1:t})}{\hat{p}(x_{1:t}|y_{1:t+1})} [\varphi_t(x_{1:t}) - \bar{\varphi}_t] \right)^2 dx_{1:t} \\
=& \int \frac{p(x_{1:t}|y_{1:t})^2}{\hat{p}(x_{1:t-1}|y_{1:t}) q_t(x_t|x_{t-1})} \left(\varphi_t(x_{1:t}) - \bar{\varphi}_t \right)^2 dx_{1:t}.
\end{aligned}
$$

This gives as a variance expression:

$$
\begin{aligned}
\sigma_{APF}(\varphi_t) =& \int \frac{p(x_1|y_{1:t})^2}{q_1(x_1)} \left(\int [\varphi_t(x_{1:t}) - \bar{\varphi}_t] p(x_{2:t}|y_{2:t}, x_1) dx_{2:t} \right)^2 dx_1 \\
&+ \sum_{k=2}^{t-1} \int \frac{p(x_{1:k}|y_{1:t})^2}{\hat{p}(x_{1:k-1}|y_{1:k}) q_k(x_k|x_{k-1})} \left(\int [\varphi_t(x_{1:t}) - \bar{\varphi}_t] p(x_{k+1:t}|y_{k+1:t}, x_k) dx_{k+1:t} \right)^2 dx_{1:k} \\
&+ \int \frac{p(x_{1:t}|y_{1:t})^2}{\hat{p}(x_{1:t-1}|y_{1:t}) q_t(x_t|x_{t-1})} \left(\varphi_t(x_{1:t}) - \bar{\varphi}_t \right)^2 dx_{1:t}.
\end{aligned}
$$

This completes the proof of proposition 2.2.1 and we obtain the corollary by direct substitution: in the case of *perfect adaptation* we have:

$$
\begin{aligned}
\sigma_{APF}(\varphi_t) =& \int \frac{p(x_1|y_{1:t})^2}{p(x_1|y_1)} \left(\int [\varphi_t(x_{1:t}) - \bar{\varphi}_t] p(x_{2:t}|y_{2:t}, x_1) dx_{2:t} \right)^2 dx_1 \\
&+ \sum_{k=2}^{t-1} \int \frac{p(x_{1:k}|y_{1:t})^2}{p(x_{1:k}|y_{1:k})} \left(\int [\varphi_t(x_{1:t}) - \bar{\varphi}_t] p(x_{k+1:t}|y_{k+1:t}, x_k) dx_{k+1:t} \right)^2 dx_{1:k} \\
&+ \int p(x_{1:t}|y_{1:t}) \left(\varphi_t(x_{1:t}) - \bar{\varphi}_t \right)^2 dx_{1:t}.
\end{aligned}
$$

B. Feynman-Kac Proofs

We present proofs of the general results presented in section 3 obtained by applying the results of [3]. It is, in fact, possible to use the same technique employed in the proof of the filtering results in appendix A to obtain the more general results obtained here. However, the approach employed here illustrates an elegant and powerful alternative.

The general proof strategy is to consider the sequence of distributions of the historical process $X_{1:t}$ under the action of a sequence of proposals and importance weights as a Feynman-Kac flow in which the proposals correspond to the mutation kernel and the importance weights to potential functions. The particle approximation proposed here may then be treated as a mean field approximation to a suitable McKean interpretation of that flow. This is not a new idea and has been widely used in the study of interacting particle systems. See [3] for further details. Having made this connection, the proofs which we present below require little more than that we are able to verify the conditions of various general results on the particle interpretations of Feynman-Kac flows for the particular case of interest.

B.1 Strong Law of Large Numbers

Proof (Strong Law of Large Numbers). Theorem 3.3.1 follows by a direct analogue (simplified slightly as, by construction, the particle system constructed here never "dies") of [3, Theorem 7.4.3 and Corollary 7.4.2], using the decomposition:

$$\begin{aligned}
&\int \left(\pi_t^N(x_t) - \pi_t(x_t)\right) \varphi_t(x_t) dx_t \\
=& \int \left(\frac{\mu_t^N(x_t)}{\int \mu_t^N(x_t')\widetilde{W}_t(x_t')dx_t'} - \frac{\mu_t(x_t)}{\int \mu_t(x_t')\widetilde{W}_t(x_t')dx_t'} \right) \widetilde{W}_t(x_t)\varphi_t(x_t)dx_t \\
=& \frac{\int \mu_t(x_t')\widetilde{W}_t(x_t')dx_t'}{\int \mu_t^N(x_t')\widetilde{W}_t(x_t')dx_t'} \int \frac{\mu_t^N(x_t)\widetilde{W}_t(x_t)}{\int \widetilde{W}_t(x_t')\mu_t(x_t')dx_t'} \left(\varphi_t(x_t) - \bar{\varphi}_t^{\pi_t}\right) dx_t,
\end{aligned} \tag{B.1}$$

which follows by noting that the normalising constant associated with the importance weight function cancels in this ratio and hence we may use the normalised version for simplicity of calculation.

B.2 Central Limit Theorem

Proof (Theorem 3.3.2: Central Limit Theorem). The approach of the proof is closely related to that of [3, 4]. We begin by considering decomposition (B.1), noting that μ_t corresponds to a normalised time marginal of a Feynman-Kac flow[1], and hence allowing us to use the result that:

$$\frac{\int \mu_t(x_t')\widetilde{W}_t(x_t')dx_t'}{\int \mu_t^N(x_t')\widetilde{W}_t(x_t')dx_t'} \xrightarrow{p} 1,$$

together with a simple application of the central limit theorem [3, Proposition 9.4.1] which tells us that:

$$\int \frac{\mu_t^N(x_t)\widetilde{W}_t(x_t)}{\int \widetilde{W}_t(x_t')\mu_t(x_t')dx_t'} \left(\varphi_t(x_t) - \bar{\varphi}_t^{\pi_t}\right) dx_t \xrightarrow{d} \tau_t^2(\widetilde{W}_t\varphi_t),$$

where the variance function is defined by the expression:

$$\begin{aligned}\tau_t^2(\varphi_t) =& \mathbb{E}_{K_1}\left[W_1^2 Q_{2:t}(\varphi_t - \bar{\varphi}_t^{\mu_t})\right] + \\ & \sum_{k=2}^{t-1} \mathbb{E}_{\mu_{k-1}\otimes K_k}\left[W_k^2 \times Q_{k+1:t}(\varphi_t - \bar{\varphi}_t^{\mu_t})^2\right] + \\ & \mathbb{E}_{\mu_{t-1}\otimes K_t}\left[W_t^2 \times (\varphi_t - \bar{\varphi}_t^{\mu_t})^2\right].\end{aligned}$$

where, given some function $W_t : E_{t-1} \times E_t \to \mathbb{R}$ and a second function $\psi_t : E_t \to \mathbb{R}$ we write $W_t \times \psi$ to denote the function $(W_t \times \psi_t)(x_{t-1}, x_t) = W_t(x_{t-1}, x_t)\psi_t(x_t)$; the product measure $(\mu_{k-1}\otimes K_p)(x_{k-1}, x_p) = \mu_{k-1}(x_{k-1})K_p(x_{k-1}, x_p)$ and the semigroup, $Q_{k+1:t}$ is defined by:

$$\begin{aligned}Q_{k+1:t}(\varphi_t) &= Q_{k+1} \circ \cdots \circ Q_t(\varphi_t) \\ Q_t(\varphi_t)(x_{t-1}) &= \mathbb{E}_{K_t(x_{t-1},\cdot)}\left[(W_t \times \varphi_t)(x_{t-1}, \cdot)\right].\end{aligned}$$

Before obtaining the final result, it is useful to rewrite this expression in a form which is somewhat easier to interpret. Beginning by noting that:

$$\begin{aligned}Q_t(\varphi_t)(x_{t-1}) &= \int K_t(x_{t-1}, x_t)\frac{\mu_t(x_t)L_{t-1}(x_t, x_{t-1})}{\mu_{t-1}(x_{t-1})K_t(x_{t-1}, x_t)}dx_t \\ &= \frac{\widetilde{\mu}_t(x_{t-1})}{\mu_{t-1}(x_{t-1})}\int \varphi_t(x_t)\widetilde{\mu}_t(x_t|x_{t-1}),\end{aligned}$$

and proceeding inductively we obtain:

$$Q_{k+1:t}(\varphi_t)(x_k) = \frac{\widetilde{\mu}_t(x_k)}{\mu_k(x_k)}\int \varphi_t(x_t)\widetilde{\mu}_t(x_t|x_k).$$

And this allows us to write each of the central terms in the variance decomposition in the form:

$$\begin{aligned}&\mathbb{E}_{\mu_{k-1}\otimes K_k}\left[W_k^2 \times Q_{k+1:t}(\varphi_t - \bar{\varphi}_t^{\mu)t})^2\right] \\ &= \int \frac{\mu_k(x_k)^2 L_{k-1}(x_k, x_{k-1})^2}{\mu_{k-1}(x_{k-1})K_k(x_{k-1}, x_k)}\frac{\widetilde{\mu}_t(x_k)^2}{\mu_k(x_k)^2}\left(\int [\varphi_t(x_t) - \bar{\varphi}_t^{\mu_t}]\widetilde{\mu}_t(x_t|x_k)dx_t\right)^2 dx_{k-1:k} \\ &= \int \frac{\widetilde{\mu}_t(x_k)^2 L_{k-1}(x_k, x_{k-1})^2}{\mu_{k-1}(x_{k-1})K_k(x_{k-1}, x_k)}\left(\int [\varphi_t(x_t) - \bar{\varphi}_t^{\mu_t}]\widetilde{\mu}_t(x_t|x_k)dx_t\right)^2 dx_{k-1:k}.\end{aligned}$$

[1] In fact, a number of the results follow here from noting that the ratio of two integrals under the normalised marginals corresponds to the ratio of the same functions integrated under the *unnormalised* marginals. However, for simplicity, we avoid introducing the unnormalized measures here.

Applying Slutzky's theorem to these two convergence results, and noting that the first and last terms in the variance decomposition may be handled similarly, tells us that the quantity of interest obeys a central limit theorem, with variance given by:

$$
\begin{aligned}
&\sigma_t^2(\varphi_t) = \tau_t^2(\widetilde{W}_t \varphi_t) \\
&= \int \frac{\widetilde{\mu}_t(x_1)^2}{K_1(x_1)} \left(\int \left[\widetilde{W}_t(x_t)\varphi_t(x_t) - \mathbb{E}_{\mu_t}\left[\tilde{W}_t \varphi_t\right]\right] \widetilde{\mu}_t(x_t|x_1) dx_t \right)^2 dx_1 + \\
&\sum_{k=2}^{t-1} \int \frac{\widetilde{\mu}_t(x_k)^2 L_{k-1}(x_k, x_{k-1})^2}{\mu_{k-1}(x_{k-1}) K_k(x_{k-1}, x_k)} \left(\int \left[\widetilde{W}_t(x_t)\varphi_t(x_t) - \mathbb{E}_{\mu_t}\left[\tilde{W}_t \varphi_t\right]\right] \widetilde{\mu}_t(x_t|x_k) dx_t \right)^2 dx_{k-1:k} \\
&+ \int \frac{\mu_t(x_t)^2 L_{t-1}(x_t, x_{t-1})^2}{\mu_{t-1}(x_{t-1}) K_t(x_{t-1}, x_t)} \left(\widetilde{W}_t(x_t)\varphi_t(x_t) - \mathbb{E}_{\mu_t}\left[\tilde{W}_t \varphi_t\right]\right)^2 dx_{k-1:k} \\
&= + \int \frac{\widetilde{\pi}_t(x_1)^2}{K_1(x_1)} \left(\int \left[\varphi_t(x_t) - \bar{\varphi}_t^{\pi_t}\right] \widetilde{\pi}_t(x_t|x_1) dx_t \right)^2 dx_1 + \\
&\sum_{k=2}^{t-1} \int \frac{\widetilde{\pi}_t(x_k)^2 L_{k-1}(x_k, x_{k-1})^2}{\mu_{k-1}(x_{k-1}) K_k(x_{k-1}, x_k)} \left(\int \left[\varphi_t(x_t) - \bar{\varphi}_t^{\pi_t}\right] \widetilde{\pi}_t(x_t|x_k) dx_t \right)^2 dx_{k-1:k} \\
&+ \int \frac{\pi_t(x_t)^2 L_{t-1}(x_t, x_{t-1})^2}{\mu_{t-1}(x_{t-1}) K_t(x_{t-1}, x_t)} \left(\varphi_t(x_t) - \bar{\varphi}_t^{\pi_t}\right)^2 dx_{k-1:k},
\end{aligned}
$$

which is precisely the result.