

Manuscript version: Author's Accepted Manuscript

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:

<http://wrap.warwick.ac.uk/136101>

How to cite:

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

Sequential Monte Carlo: Particle Filters and Beyond

Adam M. Johansen

University of Warwick, UK

March 18, 2020

Abstract

Sequential Monte Carlo methods are a family of computational algorithms which use an ensemble of weighted samples to approximate, in turn, each of a sequence of distributions and their associated normalizing constants. These algorithms first came to prominence as efficient methods for approximating the optimal filter in the context of hidden Markov models on general state spaces, online as observations become available, but are very much more widely applicable. This article seeks to provide a high level overview of these methods, introducing the methods themselves and some of their key theoretical properties before discussing their application in a range of settings which include inference for hidden Markov models; Bayesian inference more generally; maximum likelihood estimation and rare event estimation.

Keywords: approximate Bayesian computation; Bayesian inference; estimation; hidden Markov models; sequential Monte Carlo.

1 Introduction

Sequential Monte Carlo (SMC) methods are a broad class of algorithms for approximating distributions of interest, integrals with respect to those distributions and their normalizing constants. They employ an ensemble of weighted samples which targets each of a sequence of distributions in turn. In some settings this sequence arises naturally from the problem being addressed and in others it is specified as a design choice. This chapter presents a generic framework in which these methods can be described, arguing that the vast majority of SMC approaches admit an interpretation directly within this framework and that the remainder require only small extensions of it, before dedicating some space to a number of major statistical applications of these methods. A high-level view is taken, with many details left to references so that a broad overview of this area can be provided. This allows us to showcase a number of the areas in which SMC finds natural applications, not just the particle filtering setting in which it has particular prominence, but also in many other contexts including Bayesian inference, approximate Bayesian computation and rare event estimation.

2 Sequential Importance Sampling and Resampling

We will be interested in providing weighted sample approximations to each of a sequence of related distributions in turn. In some settings each of these distributions will be interesting in its own right and might arise naturally from a problem at hand; this is the case in the filtering context, explored in Section 3.1, for example. In other cases, the sequence of distributions is a computational device with only the final distribution in the sequence being of independent interest but with the others used to allow it to be approximated efficiently, typically by constructing a sequence which moves from a tractable distribution to that of interest. We shall consider throughout probability distributions defined on Euclidean spaces which admit (Lebesgue) densities; the generalization to arbitrary Polish spaces is essentially direct but significantly complicates the required notation.

Consider a sequence of probability distributions, $\{\pi_n\}_{n \in \mathbb{N}}$ defined on an increasing sequence of state spaces, $\tilde{E}_n = \otimes_{p=1}^n E_p$ with $E_p = \mathbb{R}^{d_p}$ so that, for each n , π_n is a density over $\mathbb{R}^{d_1 + \dots + d_n}$. Assume that this sequence of densities may be evaluated up to a possibly unknown normalising constant, i.e. for each n , $\pi_n = \gamma_n / Z_n$ where $\gamma_n : \tilde{E}_n \rightarrow (0, \infty)$ is an unnormalized probability density and $Z_n := \int_{\tilde{E}_n} \gamma_n(x_{1:n}) dx_{1:n}$ may not be available.

A simple importance sampling solution to the problem of approximating both π_n and Z_n would be to draw some number, N , of independent samples from a proposal distribution Q_n with respect to which π_n was absolutely continuous and to use it to approximate both of these quantities via the standard importance sampling identities:

$$\hat{\pi}_n^N(\varphi) := \frac{1}{N \cdot \hat{Z}_n^N} \sum_{i=1}^N \frac{\gamma_n(X_{1:n}^i)}{Q_n(X_{1:n}^i)} \varphi(X_{1:n}^i) \quad \hat{Z}_n^N = \frac{1}{N} \sum_{i=1}^N \frac{\gamma_n(X_{1:n}^i)}{Q_n(X_{1:n}^i)},$$

where $\varphi : \tilde{E}_n \rightarrow \mathbb{R}$ is any suitably integrable test function and $\hat{\pi}_n^N(\varphi)$ denotes the N -particle approximation of the expectation of $\varphi(X_{1:n})$ with $X_{1:n}$ distributed according to π_n .

However, if one seeks to approximate each distribution in turn such an approach seems wasteful. It is natural in this context to consider $Q_n(x_{1:n})$ which can be decomposed as

$$Q_n(x_{1:n}) = q_1(x_1)q_2(x_2|x_1) \dots q_n(x_n|x_{1:n-1}).$$

In this case, given an importance weighted ensemble of samples, $\{W_{n-1}^i, X_{1:n-1}^i\}_{i=1}^N$, which targets π_{n-1} (and which were drawn from Q_{n-1}) one can extend the sample to approximate π_n by drawing $X_n^i \sim q_n(\cdot | X_{1:n-1}^i)$ independently for $i = 1, \dots, N$ and updating the weights accordingly, setting

$$W_n^i = \frac{W_{n-1}^i G_n(X_{1:n}^i)}{\sum_{j=1}^N W_{n-1}^j G_n(X_{1:n}^j)},$$

where $G_n(x_{1:n}) := \gamma_n(x_{1:n}) / \gamma_{n-1}(x_{1:n-1}) q_n(x_n | x_{1:n-1})$ is termed the incremental weight function. (In most settings in which SMC methods find application, further simplification arises via a Markovian decomposition of the unnormalized target, $\gamma_n(x_{1:n}) = \gamma_{n-1}(x_{1:n-1}) \gamma_n(x_n | x_{n-1})$, and proposal, $q_n(x_n | x_{1:n-1}) = q_n(x_n | x_{n-1})$, distributions which means that $G_n(x_{1:n}) = G_n(x_{n-1:n}) = \gamma_n(x_n | x_{n-1}) / q_n(x_n | x_{n-1})$.) This

gives a mechanism by which each distribution can be approximated in turn, at a computational cost per iteration which does not increase with n . However, this *sequential importance sampling* (SIS) strategy is of limited direct usefulness because the variance of the importance weights and associated estimators will grow with the length of the sequence^[1], typically exponentially^[2], and only sequences of modest length can be approximated adequately using a reasonable sample size.

In order to make further progress, it is necessary to constrain the class of problems which we hope to solve a little. In particular, approximating $\pi_n(x_{1:n})$ is a task which becomes harder and harder as n increases because the dimension of the space on which these distributions are defined is growing. If we instead settle for approximating only the final time marginal of these distributions $\pi_n(x_n) = \int \pi_n(x_{1:n}) dx_{1:n-1}$ then we arrive at a sequence of problems which are of comparable difficulty. Within this regime, an approach known as resampling can be combined with the sequential importance sampling strategy described above. Resampling is a process by which a weighted sample $\{W_n^i, X_{1:n}^i\}$ is replaced with the equally-weighted population $\{1/N, \tilde{X}_{1:n}^i\}$ in such a way that the expected number of copies of each member of the original ensemble is proportional to its weight. A simple algorithmic description of this generic *Sequential Importance Resampling* (SIR) scheme is provided in Algorithm 1. In practice, in the sequential setting described above, in which $G_n(x_{1:n})$ is dependent on only $x_{n-1:n}$, it is not necessary to store the entire history of the surviving particles as a direct implementation of this algorithm would suggest, a feature which is important in the filtering context described in Section 3.1.1, for example; when one does need to store the entire history of every surviving particle space-efficient methods for doing so exist^[3].

Algorithm 1 The generic sequential importance resampling algorithm.

Initialization: $n = 1$

Sample $X_1^1, \dots, X_1^N \sim q_1$.

Compute $W_1^i = G_1(X_1^i) / \sum_{j=1}^N G_1(X_1^j)$ for $i = 1, \dots, N$.

Iteration: $n \leftarrow n + 1$

Resample $(W_{n-1}^i, X_{1:n-1}^i)_{i=1}^N$ to obtain $(1/N, \tilde{X}_{1:n-1}^i)_{i=1}^N$.

Sample $X_n^i \sim q_n(\cdot | \tilde{X}_{1:n-1}^i)$ for $i = 1, \dots, N$.

Concatenate $X_{1:n}^i \leftarrow (\tilde{X}_{1:n-1}^i, X_n^i)$ for $i = 1, \dots, N$.

Compute $W_n^i = G_n(X_{1:n}^i) / \sum_{j=1}^N G_n(X_{1:n}^j)$ for $i = 1, \dots, N$.

Where, to keep notation light, we slightly abusively allow $X_{1:n-1}^i$ to be overwritten with new values in the concatenation step.

The simplest approach to resampling is known as multinomial resampling — because one can view the number of copies made of each particle in the present generation under this scheme as a multinomial random variable with N trials and categorical probabilities given by the vector of particle weights. Multinomial resampling often features in theoretical work on sequential Monte Carlo. In practical implementations there are often advantages to employing lower variance resampling schemes. One comparative review of common resampling schemes^[4] compares the properties of a number of simple schemes; more recently the properties of a broad class of algorithms have been studied in detail^[5] — and in both cases there is evidence that better performance can be obtained by using more sophisticated schemes than the simple multinomial one. In the particular case of finite state spaces, a specialised resampling scheme can be shown to outperform generic techniques^[6].

Resampling is often viewed as a *selection* mechanism in which the “fittest” particles survive and are replicated and the least fit produce no offspring. The act of resampling clearly introduces additional variance into the estimators associated with SMC algorithms in the sense that one obtains better lower variance estimates immediately before a resampling operation than immediately after it; however, the immediate increase in variance is justified by the stability that it provides to the system in the future. Consequently, it may be desirable to avoid resampling more often than is necessary — particularly if a simple scheme, such as the multinomial one is being used. Occasional resampling, for example when the effective sample size^[1] falls below a threshold, is one way to limit the number of resampling events. This approach is widespread and intuitive, but was only shown to inherit many favourable convergence properties from standard sequential importance resampling schemes rather more recently^[7]. Resampling only at some iterations makes no fundamental change to the algorithm, but doing so at iterations which are selected based upon the properties of the collection of particles introduces some additional considerations which require additional steps to justify theoretically (one successful strategy^[7] being, essentially, to demonstrate that for large enough sample sizes the times at which resampling occurs converge, almost surely, to a deterministic limit).

A number of estimators can be associated with this algorithm, one of the normalizing constant, Z_n , of γ_n :

$$Z_n^N = \prod_{p=1}^n \frac{1}{N} \sum_{i=1}^N G_p(X_p^i), \quad (1)$$

and one of expectations with respect to each of the target distributions in turn:

$$\pi_n^N(\varphi) := \sum_{i=1}^N W_n^i \varphi(X_n^i).$$

There is now a considerable literature on the theoretical properties of these algorithms, the rigorous analysis of which dates back to the mid 1990s^[8]. Methods within this broad general class can be profitably interpreted as mean field approximations of Feynman-Kac formulæ^[9;10] which provides an elegant framework within which central limit theorem and law of large number results among many others have been obtained. Direct analysis of these methods is, of course, also possible^[11;12;13;14;15].

Although a detailed theoretical survey is beyond the scope of this article, it is convenient to sketch some of the most prominent results as these provide a formal justification for the use of SMC methods. We present below three results taken from a single monograph^[9] by way of illustration; in each case variants whose proofs hold under different (and often weaker) assumptions can also be found in the literature. The unbiasedness of normalizing constant estimates is a consequence of Theorem 7.4.2 of the monograph and holds under minimal conditions, although some care is required if the potential functions can take the value zero. Corollary 7.4.2 of the monograph provides a strong law of large numbers for particle approximations. The development in Chapter 9 of the monograph provides a central limit theorem.

Proposition 1 (Unbiasedness). *If the potential functions are uniformly bounded above, $\sup_{p \leq n, x} G_p(x) < \infty$, and the system has not become extinct (i.e. the associated weights have never all simultaneously taken the value zero), then: $\mathbb{E}[Z_n^N] = Z_n$.*

Proposition 2 (Strong Law of Large Numbers). *Provided G_p is bounded above and away from zero for all $p \leq n$, or other technical conditions are met, for bounded measurable $\varphi : E_n \rightarrow \mathbb{R}$, $\pi_n^N(\varphi) \xrightarrow{a.s.} \pi_n(\varphi)$.*

Proposition 3 (Central Limit Theorem). *If the potential functions are uniformly bounded above and away from zero, $\sup_{p \leq n, x, y} G_p(x)/G_p(y) < \infty$, then, for bounded measurable φ ,*

$$\sqrt{N} \left[\pi_n^N(\varphi) - \pi_n(\varphi) \right] \rightarrow \mathcal{N}(0, \sigma_n^2(\varphi))$$

where \rightarrow denotes convergence in distribution and the asymptotic variance $\sigma_n^2(\varphi)$ can be written either recursively or as a sum of integral expressions. Explicit forms for the asymptotic variance can be found for particle filters^[12], auxiliary particle filters^[16] and SMC Samplers^[17], for example.

2.1 Extended State Spaces and SMC Samplers

It is often the case that we are interested in a sequence of distributions over a common space (rather than distributions defined on spaces of increasing dimension) or even a single distribution, π . In order to use SMC in the first of these cases it is necessary to define a sequence of distributions of the correct form to allow the sequential importance resampling paradigm to be deployed while retaining the distributions of interest as marginals; the second case can be handled by constructing an artificial sequence of distributions which lead from a tractable distribution to that of interest. Examples of both cases in the context of Bayesian inference are provided in Section 3.2.

An explicit technique for doing exactly this in some degree of generality was introduced by Del Moral, Doucet and Jasra in 2006^[17]. Given a sequence of target distributions over some space, E , π_1, \dots, π_T , one can define a sequence of distributions over $\tilde{E}_n = E^n$, say, $\tilde{\pi}_1, \dots, \tilde{\pi}_T$ such that $\tilde{\pi}_n$ is a distribution over vectors of n elements of E , in such a way that they satisfy the requirements for the deployment of SMC and such that they admit the distributions of interest as a marginal (and, in particular, the final time marginal which SMC algorithms are best able to approximate). In order to do this, it is convenient to introduce a sequence of Markov kernels L_1, \dots, L_{T-1} which operate *backwards* in time so that we can define $\tilde{\pi}_n(x_{1:n}) = \pi_n(x_n) \prod_{p=n-1}^1 L_p(x_{p+1}, x_p)$. If one denotes the proposal distribution at iteration n of such an algorithm q_n , then one arrives at importance weights:

$$\begin{aligned} G_n(x_{1:n}) &= \frac{\tilde{\pi}_n(x_{1:n})}{\tilde{\pi}_{n-1}(x_{1:n-1})q_n(x_n|x_{n-1})} \\ &= \frac{\pi_n(x_n) \prod_{p=n-1}^1 L_p(x_{p+1}, x_p)}{\pi_{n-1}(x_{n-1}) \prod_{p=n-2}^1 L_p(x_{p+1}, x_p)q_n(x_n|x_{n-1})} = \frac{\pi_n(x_n)L_{n-1}(x_n, x_{n-1})}{\pi_{n-1}(x_{n-1})q_n(x_n|x_{n-1})}. \end{aligned}$$

The simple form of these weights and the lack of dependence on any but the final two coordinates is a consequence of the Markovian approach to extending these distributions. The remaining question of how to choose the backward kernels, L_n , can be (partially) answered by considering the variance of the resulting importance weights^[17]. The optimal choice for finite sample sizes is intractable as it depends on the actual marginal sampling distribution of the particles at iteration n which is hard to characterise as a

consequence of the resampling mechanism; but asymptotic arguments suggest that neglecting the departure of the approximation from the target at time n is a reasonable way to proceed. This suggests a near optimal strategy would be to choose

$$L_p(x_p, x_{p-1}) = \frac{\pi_{p-1}(x_{p-1})q_p(x_p|x_{p-1})}{\int \pi_{p-1}(x'_{p-1})q_p(x_p|x'_{p-1})dx'_{p-1}},$$

but in general this will lead to intractable importance weights (loosely speaking it can be seen as an attempt to integrate out the history of the particle system). In the case in which q_p is a π_p -invariant Markov kernel¹ a small departure from the optimal expression gives rise to the time-reversal of q_p with respect to its invariant distribution:

$$L_p(x_p, x_{p-1}) = \frac{\pi_p(x_{p-1})q_p(x_p|x_{p-1})}{\int \pi_{p-1}(x'_{p-1})q_p(x_p|x'_{p-1})dx'_{p-1}} = \frac{\pi_p(x_{p-1})}{\pi_{p-1}(x_p)},$$

which can be more readily used. This is a rather natural choice when one uses π_p -invariant Markov kernels in the proposal mechanism; indeed, this auxiliary kernel appears in the proof of a central limit theorem for the resample-move algorithm^[18]. In this particular setting one can also arrive at the same importance weights using more direct arguments^[19].

The ease with which adaptation can be incorporated within SMC methods is one of their great strengths and several strategies have been proposed^[20;21;22;23] and theoretical support provided^[24]. There are two areas in which adaptation is most commonly employed: first within the sequence of target distributions (in settings in which a single distribution is of ultimate interest) and in the parameters of the proposal distribution employed at each step. Appropriate adaptive methods naturally vary between contexts, but at least in contexts in which one expects consecutive distributions within the sequence of targets to be broadly “similar” and Metropolis-Hastings kernels are used there are two common approaches to tuning the proposal distribution: using the particle population at time $n - 1$ to estimate the moments of the target at that time and to employ a proposal at time n which would be optimal, for example, for a Gaussian target with those moments; or, to adjust the proposal scale whenever the acceptance rate falls outside some target region (motivated by optimal scaling considerations). In settings in which one builds an artificial sequence of distributions in order to reach a single distribution of interest it is common practice to specify a sequence of distributions which differ from one another by approximately equal amounts; strategies which control the effective sample size (or variants in the case of occasional resampling^[25]) aim, essentially, to control the χ -squared discrepancy between consecutive distributions, which is intuitively appealing if not, in general, optimal.

2.2 Particle MCMC and Related Methods

The particle MCMC^[26] approach essentially employs SMC algorithms *within* MCMC algorithms — in some sense the counterpart of the use of MCMC moves within SMC algorithms — in order to provide good proposal distributions. It is intuitive that as SMC provides good approximations to its target distributions

¹Including those, like those arising from Metropolis-like accept-reject mechanisms, which do not admit Lebesgue densities; a more careful treatment allows it to be established that absolute continuity of $\pi_p(x_p)L_{p-1}(x_p, x_{p-1})$ with respect to $\pi_{p-1}(x_{p-1})q_p(x_p|x_{p-1})$ is all that is really required and the time reversal kernel described here readily satisfies that requirement.

that it could provide good approximations to, for example, block-Gibbs sampler proposals and intractable marginal distributions but considerable care is required to justify this: one cannot simply use approximations naively within an MCMC algorithm and expect to obtain the correct invariant distribution.

In order to justify this type of algorithm it is necessary to characterize the distribution of all of the random variables generated in the running of an SMC algorithm and to do this it is convenient to reinterpret the resampling slightly as sampling an *ancestor* for each member of the resulting population. Having done this, the joint distribution over the variables simulated in the proposal step and in the selection of ancestors can be characterized straightforwardly allowing for a variety of MCMC algorithms which make use of SMC as a constituent part to be justified by an extended state space construction in which the distribution of interest is admitted as a marginal and the additional variables involved in the SMC algorithm can be viewed as auxiliary variables.

More precisely, let $\mathbf{a}_p = (a_p^1, \dots, a_p^N)$ denote the vector of time p ancestors of the particles at time $p+1$ so that, for example, x_{p+1}^i is an offspring of $x_p^{a_p^i}$. Similarly, let $\mathbf{x}_p = (x_p^1, \dots, x_p^N)$. For simplicity, consider the case in which $q_p(x_p|x_{1:p-1}) = q_p(x_p|x_{p-1})$ and $G_p(x_{1:p}) = G_p(x_{p-1}, x_p)$; the general case follows by identical arguments but with somewhat more cumbersome notational requirements. The random variables simulated in the course of running Algorithm 1 up to time n are the states $\mathbf{x}_{1:n} \in \prod_{p=1}^n E_p^N$ and $\mathbf{a}_{1:n-1} \in \{1, \dots, N\}^{Nn}$ and have the joint distribution

$$\psi_n(\mathbf{x}_{1:n}, \mathbf{a}_{1:n-1}) := \prod_{i=1}^N q_1(x_1^i) \cdot \prod_{p=2}^n \left[r(\mathbf{a}_{p-1}|\mathbf{w}_{p-1}) \prod_{i=1}^N q_p(x_p^i|x_{p-1}^{a_{p-1}^i}) \right]$$

where $r(\cdot|\mathbf{w})$ denotes the conditional distribution of ancestors arising from a resampling operation with weight vector \mathbf{w} and the weight vectors are included to simplify the notation but are formally redundant as $\mathbf{w}_p = (w_p^1, \dots, w_p^N)$ is a deterministic function of $\mathbf{x}_{1:p}$ and $\mathbf{a}_{1:p-1}$ with $w_p^i = G_p(x_{p-1}^{a_{p-1}^i}, x_p^i) / \sum_{j=1}^N G_p(x_{p-1}^{a_{p-1}^j}, x_p^j)$ in the context described. For a concrete example of such a construction, consider the case in which multinomial resampling is used, in which case

$$r(\mathbf{a}_{p-1}|\mathbf{w}_{p-1}) = \prod_{i=1}^N w_{p-1}^{a_{p-1}^i}.$$

Two broad categories of algorithms arise from the use of this construction within an MCMC context. Algorithms within the first category mimic a marginal form of Metropolis-Hastings algorithm in settings in which a completed likelihood is tractable but the marginal one is not; such *particle Marginal Metropolis Hastings* (PMMH) algorithms can be justified directly as pseudo-marginal algorithms^[27], noting that the normalizing constant estimates provided by SMC algorithms are unbiased, or via the type of auxiliary variable construction described above. Algorithms in the second category mimic an idealised block-Gibbs sampler in which the full vector of random variables being updated are drawn from their joint conditional distribution; these algorithms are a little more complex requiring the introduction of so-called *conditional SMC* (cSMC) algorithms and admitting a justification as partially collapsed Gibbs samplers^[28]. The cSMC algorithm corresponds essentially to an SMC algorithm which is modified so one particular particle trajectory is fixed in advance and guaranteed to survive through resampling steps; although notationally awkward to describe

in full generality, such algorithms are simple to implement and enjoy good mixing properties, potentially justifying a little additional implementation effort.

cSMC algorithms warrant a little discussion in their own right; it can be shown that running a cSMC algorithm and drawing a single particle trajectory from the resulting weighted ensemble corresponds to a Markov kernel which is invariant with respect to a particular distribution (the smoothing distribution in the context of hidden Markov models as described in Section 3.1.2) and can enjoy uniform ergodicity^[29]. The basic algorithm can be further improved in many cases by sampling not from the population of particle trajectories obtained naturally by the resampling mechanism but employing a backward simulation approach^[30] reminiscent of the backward simulation smoother described in Section 3.1.2, or a forward-only representation of the same known as ancestor sampling^[31] — and with these modifications it can be possible to employ very modest population sizes.

The SMC² algorithm^[32] embeds the particle MCMC approach within a sequential Monte Carlo sampler and, to some extent, allows for online parameter estimation within state space models. Roughly speaking, a data-tempered SMC sampler is used to approximate the distribution over the parameter space with the importance weights associated with this algorithm being obtained from an ensemble of particle filters approximating the distribution in the latent variable space although, of course, some care is needed in dealing with the details.

3 SMC in Statistical Contexts

3.1 SMC for Hidden Markov Models

Perhaps the most widely known application of SMC methods is to Bayesian inference for general state space hidden Markov models (HMMs) or state-space models (SSMs) as they are sometimes known. This approach dates back at least to the early 1990s^[33;34] along with the term bootstrap filter^[34] and interacting particle filter^[8]. One fairly recent survey of SMC^[2] in the HMM context demonstrates that almost all particle filtering methods can be viewed within the simple SIR framework described above, which is also the perspective which we take here.

Consider an \mathbb{R}^{d_x} -valued discrete-time Markov process $\{X_n\}_{n \geq 1}$ such that

$$X_1 \sim \mu(x_1) \text{ and } X_n | (X_{n-1} = x_{n-1}) \sim f(x_n | x_{n-1}) \quad (2)$$

where “ \sim ” means distributed according to, $\mu(x)$ is a probability density function and $f(x|x')$ denotes the probability density associated with moving from x' to x . We are interested in estimating $\{X_n\}_{n \geq 1}$ but only have access to the \mathbb{R}^{d_y} -valued process $\{Y_n\}_{n \geq 1}$. We assume that, given $\{X_n\}_{n \geq 1}$, the observations $\{Y_n\}_{n \geq 1}$ are statistically independent and their marginal densities are given by

$$Y_n | (X_n = x_n) \sim g(y_n | x_n). \quad (3)$$

For the sake of simplicity, we have considered only the homogeneous case here; that is, the transition and

observation densities are independent of the time index n . The extension to the inhomogeneous case is straightforward.

There are several inferential problems associated with this class of models: *filtering* corresponds to the sequential characterization of the law of the latent state, X_n at time n given observations $y_{1:n}$ for each n as observations become available; *smoothing* to the characterization of the law of the entire vector of latent states $X_{1:n}$ up until time n given observations $y_{1:n}$ again often sequentially as observations become available; *prediction* to the characterization of the law of X_{n+p} for $p \geq 1$ given observations $y_{1:n}$ for each n and can often be treated as a straightforward extension of filtering; and parameter estimation corresponds to the estimation of static model parameters which do not evolve over time. Until Section 3.1.3 it will be assumed that any model parameters are known.

3.1.1 Filtering

Perhaps the most natural approach to filtering within the SMC framework described above is to simply set

$$\gamma_n(x_{1:n}) = p(x_{1:n}, y_{1:n}) = \mu(x_1)g(y_1|x_1) \prod_{p=2}^n f(x_p|x_{p-1})g(y_p|x_p),$$

where p denotes the joint density of the latent and observation processes over the time horizon indicated by its arguments as well as associated conditional and marginal distributions as is common in this literature. In this case, $Z_n = p(y_{1:n})$ and $\pi_n(x_{1:n}) = p(x_{1:n}|y_{1:n})$. If one also sets $q_n(x_n|x_{1:n-1}) \equiv f(x_n|x_{n-1})$ one arrives at a particularly simple algorithm known as the bootstrap particle filter.

There are numerous strategies to improve the performance of SMC in the context of filtering problems; a number of the more prominent strategies are summarised below; for more details and a demonstration that all of these methods can be viewed as SIR algorithms (sometimes on suitably extended state spaces), see [2].

Alternative Proposals can improve the performance of the algorithm; the *locally optimal* proposal $q_n(x_n|x_{1:n-1}) = p(x_n|x_{n-1}, y_n)$ minimizes the conditional variance of the importance weights within the class of algorithms being considered here [35].

Auxiliary Particle Filters [36;16;37] attempt to further improve performance by deferring resampling until after the influence of the next observation has been (partially) incorporated into the importance weights.

Lookahead methods [38] / block-sampling [39] techniques extend these ideas further into the future, albeit at the expense of immediacy. They do this either by modifying the target distribution to approximately incorporate the influence of several subsequent observations, or by sampling new values for the most recently estimated states (using an extended state space construction similar to that employed within SMC samplers) during each iteration to allow for the influence of the most recent observations to be incorporated. Some recent work [40] attempts to address the difficulty of designing good high-dimensional proposals via an iterative scheme appropriate only *outside* the online filtering framework; this idea was recently explored more extensively outside the HMM context [41].

MCMC Moves can be included within particle filters. There are two broad approaches to the inclusion of MCMC-based innovations within SMC algorithms. So-called resample-move [18] based approaches

add a Markov kernel with respect to which the target distribution is invariant to each iteration of the algorithm; this provides a mechanism to improve sample diversity, but doesn't fundamentally change the structure of the algorithm. Another approach, often termed sequential-MCMC replaces the simulation of a collection of conditionally independent samples during each iteration with the simulation of a Markov chain with an appropriate invariant distribution; such approaches have been present in the literature for some time^[42] and good empirical performance has been observed^[43;44], although convergence results appear to have become available only recently^[45]. The ensemble HMM method^[46], in which a grid of points is obtained at each time via the simulation of a Markov chain of an appropriate invariant distribution, prior to the performance of inference using that grid as a discrete state space can be shown to be closely connected with sequential MCMC methods^[47] combined with particle MCMC.

There is also considerable work on the use of SMC for filtering in the continuous time setting, good recent surveys^[48;49] and references therein provide a good overview, but a detailed survey falls outside the scope of this chapter.

3.1.2 Smoothing

In principle, Algorithm 1 applied to a sequence of target distributions coinciding with $p(x_{1:n}|y_{1:n})$ provides an approximation of each smoothing distribution in turn. However, this naive approach sometimes known as the “smoother mode” of the particle filter is doomed to fail eventually as it corresponds to an importance sampling like approach on a space of ever increasing dimension. In fact, the situation is a little worse as every resampling step reduces the number of unique paths at earlier times and eventually $p(x_1|y_{1:n})$ is approximated by only a single surviving path. There has been considerable attention in the literature to the problem of better approximating smoothing distributions.

Fixed-lag methods provide one simple approximate scheme^[50] which allows for smoothing in an online fashion as observations become available. Rather than attempting to approximate the distribution of x_p given all of the observations received, one settles for an approximation given all of the observations obtained up until a time some fixed time after p , i.e. making the approximation $p(x_p|y_{1:n}) \approx p(x_p|y_{1:\min(p+L,n)})$ which is intuitively reasonable for sufficiently large L provided that the process under study is sufficiently ergodic. The resulting approximation error can be controlled under mixing assumptions, at least for the estimation of additive functionals^[51].

Several more sophisticated methods are possible, see^[52]. In particular the forward-filtering backward-simulation (FFBSi) approach which revolves around the decomposition of the smoothing distribution as

$$p(x_{1:n}|y_{1:n}) = p(x_n|y_{1:n}) \prod_{p=1}^{n-1} p(x_p|y_{1:p}, x_{p+1}),$$

with

$$p(x_p|x_{p+1}, y_{1:p}) = \frac{p(x_p|y_{1:p})f(x_p|x_{p-1})}{p(x_{p+1}|y_{1:p})}.$$

This allows us to write:

$$p(x_{1:n}|y_{1:n}) = p(x_n|y_{1:n}) \prod_{p=1}^{n-1} \frac{p(x_p|y_{1:p})f(x_p|x_{p-1})}{p(x_p|y_{1:p})},$$

and within the SMC framework one can obtain a sample approximation of the smoothing distribution by first running a standard particle filter forwards to the final time computing and storing all of the marginal filtering distributions along the way and then to run a backward pass using a the resulting particle approximation of $p(x_p|x_{p+1}, y_{1:p})$. A theoretical analysis of this and related approaches is provided by^[53].

The FFBSi approach has a computational cost of $O(Nn)$ per backward sample path (where N is the number of particles used in the forward filtering phase and n is the length of the time series) and hence a cost of $O(N^2n)$ if one wishes to obtain an N -particle approximation. Some work has been done to mitigate this in the literature, including a slightly different approximation of the distribution which can reduce the cost to something linear in the sample size if one is interested in only marginal smoothing distributions^[54] and methods which allow efficient estimation of smoothing expectations of additive functionals^[55;53;56].

Offline approaches to smoothing via particle MCMC, or iterated conditional SMC have recently been developed^[57] and are closely related to the problem of static parameter estimation which is discussed in the next section.

3.1.3 Parameter Estimation

Estimating static parameters, i.e. those parameters which take a single value which is common to all time points, is a challenging problem in the HMM context — particularly in online contexts. Online, here, means providing an estimate each time a new observation is obtained which incorporates the influence of all observations received to date at an iterative cost which is bounded in time. The particular difficulties arise from the non-trivial dependence structure in which the static parameter and the entire latent state vector has complex dependencies; the path degeneracy problem of the particle filter makes dealing with the full joint distribution challenging.

Broadly speaking methods can be characterized as online or offline and make use of either maximum likelihood or Bayesian approaches to parameter estimation. Offline inference, a competitor to MCMC for the same problem is generally easier and likelihood-based methods are less computationally demanding than fully Bayesian ones, especially in the online setting in which it is possible to leverage ideas based around Fisher scoring on stochastic expectation maximisation algorithms. Good approaches to these problems are somewhat specialised, but an excellent recent survey exists^[58].

3.2 SMC for Bayesian Inference

There are many ways in which SMC finds application in the context of Bayesian inference; there is a good recent review of methods applicable in the context of graphical models^[59]. One common application of SMC in the statistical literature is in the approximation of the Bayesian posterior distribution for some parameter θ for which one has prior distribution $p(\theta)$ and a likelihood $p(y|\theta)$ where y denotes the full set of data available.

Approaches to this problem date back in the statistics literature have existed for approximately two

decades^[19], with related ideas to be found in the earlier literature^[60;61], and a framework incorporating this and many other algorithms is provided by^[17]. Two common approaches, widely identified as data tempering and likelihood tempering, to the specification of a suitable sequence of distributions are widespread. In the data tempering setting, one defines a sequence of distributions by adding additional observations at each step, arriving at a sequence of partial posteriors of the form:

$$\pi_n(\theta) \propto \gamma_n(\theta) = p(\theta)p(y_{1:m_n}|\theta)$$

for some sequence (m_n) of data sizes increasing from zero to the actual size of the data set; whereas in likelihood tempering

$$\pi_n(\theta) \propto \gamma_n(\theta) = p(\theta)p(y|\theta)^{\alpha_n}$$

for some monotonically increasing real-valued sequence, (α_n) , which increases from zero to one. Both m_n and α_n can be specified adaptively.

In the context of Bayesian inference for static parameters with either of these sequences of target distributions, it is natural to employ π_n -invariant Markov kernels as the mutation element of the SMC algorithm giving rise to incremental importance weights at time n of the form $p(y_{m_{n-1}+1:m_n}|\theta)$ and $p(y|\theta)^{\alpha_n - \alpha_{n-1}}$, respectively, if one operates within the SMC sampler framework using the time reversal of these invariant Markov kernels as the associated auxiliary kernels.

Of course, the SMC framework provides very considerable flexibility and we need not be constrained to sequences of distributions which temper from prior to posterior. In the context of generalised linear mixed models, for example, it has been found that starting with a distribution motivated by quasilielihood arguments and moving from that to the posterior leads to somewhat better performance^[62].

3.2.1 SMC for Model Comparison

Like parameter estimation in HMMs, Bayesian model comparison centres around some computation of the marginal likelihoods; i.e. the marginal probability under a given model of observing the data actually observed, with unknown model parameters marginalized out. In the context of any sequence of distributions which begins with a properly normalized distribution over the space of unknown parameters and finishes with the posterior characterized as the product of the complete likelihood and parameter priors divided by an unknown normalizing constant, that normalizing constant corresponds exactly with the marginal likelihood and is estimated unbiasedly by the associated SMC scheme via Equation (1) (i.e. $Z_n = \int \gamma_n(\theta)d\theta = \int p(\theta)p(y|\theta)d\theta$ when n is the final distribution within either the data- or likelihood-tempering schemes described in Section 3.2), so that either m_n corresponds to the size of the data set or $\alpha_n = 1$.

As the estimation of normalizing constants and marginal likelihoods is somewhat natural in the SMC setting, these algorithms lend themselves to this problem. A number of different approaches to this problem have been explored and found to perform well in many settings^[25;63]. These approaches include simultaneously addressing model and parameter inference in a similar manner to reversible jump MCMC methods^[64], explicitly approximating the marginal likelihoods of each of a family of competing models and directly computing the ratio of marginal likelihoods of pairs of competing models, the so-called *Bayes factor*.

3.2.2 SMC for ABC

Approximate Bayesian Computation (ABC; introduced in ^[65]; recent survey ^[66]) is another area in which SMC has been widely applied ^[67;21]. ABC is a technique for performing computational inference in settings in which the likelihood cannot be evaluated but it is possible to simulate from the associated data generating model for given parameter values. A detailed survey of ABC methods is outside the scope of this chapter, but in essence the fit of a parameter value to a given data set is assessed by simulating a data set from the generative model for that parameter value and comparing it with the actually observed data, typically by determining the distance between summary statistics computed using the real and simulated datasets. For example, by considering a target distribution of the form

$$\pi_\epsilon(\theta, y) = p(\theta)f(y|\theta)\mathbb{I}_{[0,\epsilon]}(d(S(y), S(y_{\text{obs}})))$$

where ϵ denotes a tolerance, θ the unknown parameters of interest, $p(\theta)$ a prior distribution, y the auxiliary simulated data, $f(y|\theta)$ the modelled generative relationship between parameters and data, S a mapping from the data space to a low-dimensional summary statistic space, d is some appropriate distance and y_{obs} the actually-observed data. In the SMC context it is natural to make use of a sequence of distributions which require an increasing degree of fidelity between the observed and simulated data, i.e. considering a (possibly adaptive) decreasing sequence of values of ϵ .

In an ABC context, the need to resimulate synthetic data whenever a new parameter value is proposed limits the ability for SMC to benefit from local exploration as it does in standard Bayesian inferential settings; one remedy to this is to adopt an appropriate non-centred parametrisation when this is possible ^[68].

It is also possible to compute estimates of model evidence within the ABC framework using SMC ^[69]; although considerable caution is required in doing so, particularly in the selection of summary statistics, and interpreting the conclusions ^[69;70].

3.3 SMC for Maximum Likelihood Estimation

It is worthwhile noting that, although SMC like many Monte Carlo methods is widely used within the Bayesian domain, it also finds application in other statistical paradigms.

Maximum likelihood estimation (MLE) is, at heart, an optimization problem and it is no surprise that simulated-annealing like methods can be used in this context; within the marginal MLE setting SMC samplers and data cloning provide one natural approach to this problem ^[71]. A more direct use of a simulated annealing strategy was explored by ^[72], and a pseudomarginal ^[27] variant also shows promise ^[73]. All of these approaches essentially involve the construction of a sequence of distributions which become progressively more concentrated on the set of maximisers of the likelihood function and targetting this sequence using SMC sampler algorithms.

3.4 SMC for Rare Event Estimation

Estimating the probabilities of rare events (i.e. those with small probability of occurrence) is a natural application of SMC methods; in this context one can begin from the law of some underlying random variable and move via a sequence of intermediate distributions to the restriction of that law to the rare event of interest obtaining both an approximation of the probability of this event (via the normalizing constant of this restriction) and also an approximation of the law of the random variable restricted to that set (via the final particle set).

Sequential Monte Carlo provides natural approaches to the so-called dynamic rare event problem, in which one is interested in establishing the probability that a Markov process hits a specified rare set of interest before its next entrance into some recurrent set^[74] and the static rare event problem in which the question is whether a random variable/process takes a value within some set which has small probability under its law^[75;76;77]. In the dynamic case it is common to employ a sequence of intermediate distributions in order to characterize the probability of hitting each of a sequence of increasingly rare sets before the recurrent set; in the latter one simply needs to construct a sequence of distributions which begins with the law of the random quantity of interest and becomes increasingly concentrated on the rare set of interest.

4 Selected Recent Developments

This chapter concludes with a brief summary of some exciting emerging topics within the field of SMC.

One, perhaps surprising, recent development is the emergence of methodology which permits the consistent estimation of the variance and asymptotic variance of SMC algorithms using the output from a single realisation of the particle system^[78;79]. This has recently been extended to the case of a class of adaptive algorithms^[80]. In the context of online inference in state space models a “fixed lag” approach was explored by^[81]. These methods provide an avenue to the characterization of the quality of estimates obtained from SMC algorithms *without* recourse to multiple costly runs of those algorithms.

Considering the “genealogical properties” of SMC algorithms (i.e the trees which one obtains by tracing back particles surviving until the current generation and producing a tree containing all particles in previous generations which are ancestors to surviving particles) has provided another avenue to understanding the behaviour of these algorithms. Both bounds on properties of these trees^[3] and a characterization of the limiting tree^[82] have been obtained and provide information about storage costs of algorithms as well as efficient data structures for storing the entire history of the currently surviving particles.

Efficient distributed implementation via modifications of the fundamentally synchronous resampling operation^[83;84;85] or via more fundamental changes to the methodology suitable for offline inference^[86] has been the subject of substantial recent research and further developments in this direction are to be expected in the future.

Quasi-Monte Carlo (QMC) methods eschew the use of random numbers in favour of low discrepancy sequences which seek, in a suitable sense, to fill space as regularly as possible. Leveraging these techniques in an SMC setting is challenging, in part because of the increasing-state space justification of most SMC

methods and in part due to complications arising from resampling, but substantial progress in this direction was made in the form of Sequential Quasi-Monte Carlo^[87], which employs quasi-Monte Carlo within a marginal framework, at iteration n sampling (a_{n-1}^i, x_n^i) jointly according to $r(a_{n-1}^i | \mathbf{w}_{n-1})q_n(x_n^i | x_{n-1}^{a_{n-1}^i})$, in the notation of Section 2.2, via quasi-Monte Carlo methods. It shows particularly substantial performance gains in relatively low-dimensional filtering type problems.

Acknowledgments

The author’s research is partially supported by the Alan Turing Institute–Lloyd’s Register Foundation programme on Data-Centric Engineering and the Engineering and Physical Sciences Council Grant EP/R034710/1.

5 References

- [1] 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.
- [2] A. Doucet and A. M. Johansen. A tutorial on particle filtering and smoothing: Fiteen years later. In D. Crisan and B. Rozovskii, editors, *The Oxford Handbook of Nonlinear Filtering*, pages 656–704. Oxford University Press, 2011.
- [3] P. E. Jacob, L. Murray, and S. Rubenthaler. Path storage in the particle filter. *Statistics and Computing*, 25(2):487–496, 2015.
- [4] R. Douc, O. Cappé, and E. Moulines. Comparison of resampling schemes for particle filters. In *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis*, volume I, pages 64–69. IEEE, 2005.
- [5] M. Gerber, N. Chopin, and N. Whiteley. Negative association, ordering and convergence of resampling methods. *Annals of Statistics*, 47(4):2236–2260, 2019.
- [6] P. Fearnhead and P. Clifford. On-line inference for hidden Markov models via particle filters. *Journal of the Royal Statistical Society B*, 65(4):887–899, 2003.
- [7] P. Del Moral, A. Doucet, and A. Jasra. On adaptive resampling procedures for sequential Monte Carlo methods. *Bernoulli*, 18(1), 2012.
- [8] P. Del Moral. Nonlinear filtering using random particles. *Theory of Probability and Its Applications*, 40(4):690–701, 1995.
- [9] P. Del Moral. *Feynman-Kac formulae: genealogical and interacting particle systems with applications*. Probability and Its Applications. Springer Verlag, New York, 2004.
- [10] P. Del Moral. *Mean Field Integration*. Chapman Hall, 2013.
- [11] D. Crisan and A. Doucet. A survey of convergence results on particle filtering methods for practitioners. *IEEE Transactions on Signal Processing*, 50(3):736–746, March 2002.
- [12] 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.
- [13] H. R. Künsch. Recursive Monte Carlo filters: Algorithms and theoretical analysis. *Annals of Statistics*, 33(5):1983–2021, 2005.
- [14] O. Cappé, E. Moulines, and T. Ryden. *Inference in Hidden Markov Models*. Springer Verlag, New York, 2005.
- [15] R. Douc and E. Moulines. Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Annals of Statistics*, 36(5):2344–2376, 2008.
- [16] A. M. Johansen and A. Doucet. A note on the auxiliary particle filter. *Statistics and Probability Letters*, 78(12):1498–1504, September 2008.
- [17] P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society B*, 63(3):411–436, 2006.
- [18] W. R. Gilks and C. Berzuini. Following a moving target – Monte Carlo inference for dynamic Bayesian models. *Journal of the Royal Statistical Society B*, 63(1):127–146, 2001.
- [19] N. Chopin. A sequential particle filter method for static models. *Biometrika*, 89(3):539–551, 2002.
- [20] A. Jasra, D. A. Stephens, A. Doucet, and T. Tsagaris. Inference for Lévy-Driven Stochastic Volatility Models via Adaptive Sequential Monte Carlo. *Scandinavian Journal of Statistics*, 38(1):1–22, December 2010.

- [21] P. Del Moral, A. Doucet, and A. Jasra. An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Statistics and Computing*, 22(5):1009–1020, 2012.
- [22] C. Schäfer and N. Chopin. Sequential Monte Carlo on large binary sampling spaces. *Statistics and Computing*, 23(2):163–184, March 2013.
- [23] P. Fearnhead and B. Taylor. An adaptive sequential Monte Carlo sampler. *Bayesian Analysis*, 8(2):411–438, 2013.
- [24] A. Beskos, A. Jasra, N. Kantas, and A. H. Thiéry. On the convergence of adaptive sequential Monte Carlo methods. *Annals of Applied Probability*, 26(2):1111–1146, 2016.
- [25] Y. Zhou, A. M. Johansen, and J. A. D. Aston. Towards automatic model comparison: An adaptive sequential Monte Carlo approach. *Journal of Computational and Graphical Statistics*, 25(3):701–726, 2016. doi: 10.1080/10618600.2015.1060885.
- [26] C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte Carlo. *Journal of the Royal Statistical Society B*, 72(3):269–342, 2010.
- [27] C. Andrieu and G. O. Roberts. The pseudo-marginal approach for efficient Monte Carlo computations. *Annals of Statistics*, 37(2):697–725, 2009.
- [28] D. A. Van Dyk and T. Park. Partially collapsed Gibbs samplers: Theory and methods. *Journal of the American Statistical Association*, 103(482):790–796, 2008.
- [29] C. Andrieu, A. Lee, and M. Vihola. Uniform ergodicity of the iterated conditional SMC and geometric ergodicity of particle Gibbs samplers. *Bernoulli*, 24(2):842–872, 2018.
- [30] N. Whiteley. Contribution to the discussion on ‘Particle Markov chain Monte Carlo methods’ by Andrieu, C., Doucet, A., and Holenstein, R. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):306–307, 2010.
- [31] F. Lindsten, M. I. Jordan, and T. B. Schön. Ancestor sampling for particle Gibbs. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Proceedings of the 2012 Conference on Neural Information Processing Systems (NIPS)*, pages 2591–2599, Lake Tahoe, NV, USA, 2012. Curran Associates, Inc.
- [32] N. Chopin, P. Jacob, and O. Papaspiliopoulos. SMC²: an efficient algorithm for sequential analysis of state space models. *Journal of the Royal Statistical Society B*, 75(3):397–426, 2013.
- [33] L. Stewart and P. McCarty. The use of Bayesian belief networks to fuse continuous and discrete information for target recognition, tracking and situation assessment. In *Proceedings of SPIE Signal Processing, Sensor Fusion and Target Recognition*, volume 1699, pages 177–185, 1992.
- [34] N. J. Gordon, S. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F*, 140(2):107–113, April 1993.
- [35] A. Doucet, S. Godsill, and C. Andrieu. On sequential simulation-based methods for Bayesian filtering. *Statistics and Computing*, 10(3):197–208, 2000.
- [36] M. K. Pitt and N. Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94(446):590–599, 1999.
- [37] R. Douc, E. Moulines, and J. Olsson. Optimality of the auxiliary particle filter. *Probability and Mathematical Statistics*, 29(1):1–28, 2009.
- [38] M. Lin, R. Chen, and J. S. Liu. Lookahead strategies for sequential Monte Carlo. *Statistical Science*, 28(1):69–94, 2013.
- [39] A. Doucet, M. Briers, and S. Sénécal. Efficient block sampling strategies for sequential Monte Carlo methods. *Journal of Computational and Graphical Statistics*, 15(3):693–711, 2006.
- [40] P. Guarniero, A. M. Johansen, and A. Lee. The iterated auxiliary particle filter. *Journal of the American Statistical Association*, 112(520):1636–1647, 2017.
- [41] J. Heng, A. N. Bishop, G. Deligiannidis, and A. Doucet. Controlled sequential Monte Carlo. *Annals of Statistics*, 2020. In press.
- [42] C. Berzuini, N. G. Best, W. R. Gilks, and C. Larizza. Dynamic conditional independence models and Markov chain Monte Carlo. *Journal of the American Statistical Association*, 92(440):1403–1412, December 1997.
- [43] François Septier, Sze Kim Pang, Avishy Carmi, and Simon Godsill. On MCMC-based particle methods for Bayesian filtering: Application to multitarget tracking. In *3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 360–363. IEEE, 2009.
- [44] François Septier and Gareth W Peters. Langevin and Hamiltonian based sequential MCMC for efficient Bayesian filtering in high-dimensional spaces. *IEEE Journal of Selected Topics in Signal Processing*, 10(2):312–327, 2016.
- [45] A. Finke, A. Doucet, and A. M. Johansen. Limit theorems for sequential MCMC methods. *Advances in Applied Probability*, 52(2), June 2020. In press.
- [46] A. Y. Shestopaloff and R. M. Neal. MCMC for non-linear state space models using ensembles of latent sequences. *ArXiv e-prints*, May 2013.
- [47] A. Finke, A. Doucet, and A. M. Johansen. On embedded hidden Markov models and particle Markov chain Monte Carlo methods. ArXiv mathematics e-print 1610.08962, ArXiv Mathematics e-prints, October 2016.
- [48] A. Bain and D. Crisan. *Fundamentals of Stochastic Filtering*. Stochastic Modelling and Applied Probability. Springer, 2009.

- [49] D. Crisan and B. Rozovskiĭ, editors. *The Oxford Handbook of Nonlinear Filtering*. Oxford University Press, Great Clarendon Street, Oxford, UK, 2011.
- [50] G. Kitigawa and S. Sato. Monte carlo smoothing and self-organising state-space model. In 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.
- [51] J. Olsson, O. Cappé, R. Douc, and E. Moulines. Sequential Monte Carlo smoothing with application to parameter estimation in non-linear state space models. *Bernoulli*, 14(1):155–179, 2008.
- [52] M. Briers, A. Doucet, and S. Maskell. Smoothing algorithms for state space models. *Annals of the Institute of Statistical Mathematics*, 62(1):61–89, 2010.
- [53] R. Douc, A. Garivier, E. Moulines, and J. Olsson. Sequential Monte Carlo smoothing for general state space hidden Markov models. *Annals of Applied Probability*, 21(6):2109–2145, 2011.
- [54] P. Fearnhead, D. Wyncoll, and J. Tawn. A sequential smoothing algorithm with linear computational cost. *Biometrika*, 97(2), 2010.
- [55] P. Del Moral, A. Doucet, and S. S. Singh. Forward smoothing using sequential Monte Carlo. ArXiv Mathematics e-prints 1012.5390, ArXiv Mathematics e-prints, 2010.
- [56] J. Olsson and J. Westerborn. Efficient particle-based online smoothing in general hidden markov models: the PaRIS algorithm. *Bernoulli*, 23:1951–1996, 2017.
- [57] P. Jacob, F. Lindsten, and T. Schön. Smoothing with couplings of conditional particle filters. *Journal of the American Statistical Association*, 2019. In press.
- [58] N. Kantas, A. Doucet, S. S. Singh, J. M. Maciejowski, and N. Chopin. On particle methods for parameter estimation in general state-space models. *Statistical Science*, 30(3):328–351, 2015.
- [59] A. Doucet and A. Lee. Sequential Monte Carlo methods. In M. Maathuis, M. Drton, S. L. Lauritzen, and M. Wainwright, editors, *Handbook of Graphical Models*, pages 165–189. CRC Press, 2018.
- [60] R. M. Neal. Annealed importance sampling. Technical Report 9805, University of Toronto, Department of Statistics, 1998.
- [61] S. N. MacEachern, M. Clyde, and J. S. Liu. Sequential importance sampling for nonparametric Bayes models: The next generation. *Canadian Journal of Statistics*, 27(2):251–267, 1999.
- [62] Y. Fan, D. Leslie, and M. P. Wand. Generalized linear mixed model analysis via sequential Monte Carlo sampling. *Electronic Journal of Statistics*, 2:916–938, 2008.
- [63] A. Jasra, A. Doucet, D. A. Stephens, and C. C. Holmes. Interacting sequential Monte Carlo samplers for trans-dimensional simulation. *Computational Statistics and Data Analysis*, 52(4):1765–1791, January 2008.
- [64] P. J. Green. Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–732, 1995.
- [65] S. Tavaré, D. J. Balding, R. C. Griffiths, and P. Donnelly. Inferring Coalescence Times From DNA Sequence Data. *Genetics*, 145(2):505–518, February 1997.
- [66] S. A. Sisson, Y. Fan, and M. Beaumont. *Handbook of approximate Bayesian computation*. Chapman and Hall/CRC, 2018.
- [67] S. A. Sisson, Y. Fan, and M. M. Tanaka. Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Science, USA*, 104(4):1760–1765, February 2007.
- [68] C. Andrieu, A. Doucet, and A. Lee. Discussion of “constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation” by Fearnhead and Prangle. *Journal of the Royal Statistical Society B*, 74(3):451–452, 2012.
- [69] X. Didelot, R. G. Everitt, A. M. Johansen, and D. J. Lawson. Likelihood-free estimation of model evidence. *Bayesian Analysis*, 6(1):49–76, 2011.
- [70] J.-M. Marin, N. Pillai, C. P. Robert, and J. Rousseau. Relevant statistics for Bayesian model choice. *Journal of the Royal Statistical Society B*, 76(5):833–859, 2014.
- [71] A. M. Johansen, A. Doucet, and M. Davy. Particle methods for maximum likelihood parameter estimation in latent variable models. *Statistics and Computing*, 18(1):47–57, March 2008.
- [72] S. Rubenthaler, T. Rydén, and M. Wiktorsson. Fast simulated annealing in \mathbb{R}^d with an application to maximum likelihood estimation in state-space models. *Stochastic Processes and their Applications*, 119(6):1912–1931, 2009.
- [73] A. Finke. *On Extended State-Space Constructions for Monte Carlo Methods*. Ph.D. thesis, University of Warwick, 2015.
- [74] F. Cérou, P. Del Moral, F. Le Gland, and P. Lezaud. Genetic genealogical models in rare event analysis. *ALEA: Latin American Journal of Probability and Mathematical Statistics*, 1:181–203, 2006.
- [75] P. Del Moral and J. Garnier. Genealogical particle analysis of rare events. *Annals of Applied Probability*, 15(4):2496–2534, November 2005.
- [76] A. M. Johansen, P. Del Moral, and A. Doucet. Sequential Monte Carlo samplers for rare events. In *Proceedings of the 6th International Workshop on Rare Event Simulation*, pages 256–267, Bamberg, Germany, October 2006.
- [77] F. Cérou, P. Del Moral, T. Furon, and A. Guyader. Sequential Monte Carlo for rare event estimation. *Statistics and Computing*, 22(3):795–808, 2012.

- [78] H. P. Chan and T. L. Lai. A general theory of particle filters in hidden Markov models and some applications. *Annals of Statistics*, 41(6):2877–2904, 2013.
- [79] A. Lee and N. Whiteley. Variance estimation in the particle filter. *Biometrika*, 105(1):609–625, 2018.
- [80] Q. Du and A. Guyader. Variance estimation in adaptive sequential Monte Carlo. ArXiv Mathematics e-prints 1909.13602, ArXiv Mathematics e-prints, 2019.
- [81] J. Olsson and R. Douc. Numerically stable online estimation of variance in particle filters. *Bernoulli*, 25(2):1504–1535, 2019.
- [82] J. Koskela, P. Jenkins, A. M. Johansen, and D. Spanò. Asymptotic genealogies of interacting particle systems with an application to sequential Monte Carlo. *Annals of Statistics*, 2020. In press.
- [83] L. Murray, A. Lee, and P. Jacob. Parallel resampling in the particle filter. *Journal of Computational and Graphical Statistics*, 25(3):789–805, 2016.
- [84] A. Lee and N. Whiteley. Forest resampling for distributed sequential Monte Carlo. *Statistical Analysis and Data Mining*, 9(4):230–248, 2016.
- [85] N. Whiteley, A. Lee, and K. Heine. On the role of interaction in sequential Monte Carlo algorithms. *Bernoulli*, 22(1):494–429, 2016.
- [86] F. Lindsten, A. M. Johansen, C. A. Naesseth, B. Kirkpatrick, T. Schön, J. A. D. Aston, and A. Bouchard-Côté. Divide and conquer with sequential Monte Carlo samplers. *Journal of Computational and Graphical Statistics*, 26(2):445–458, 2017.
- [87] M. Gerber and N. Chopin. Sequential quasi Monte Carlo. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(3):509–579, 2015.