Computational Statistics and Data Analysis 111 (2017) 145–165

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Residual and stratified branching particle filters

Michael A. Kouritzin

Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton (Alberta), Canada T6G 2G1

ARTICLE INFO

Article history:
Received 19 February 2016
Received in revised form 28 January 2017
Accepted 2 February 2017
Available online 21 February 2017

Keywords:
Particle filter
Branching process
Bayesian model selection

ABSTRACT

A class of discrete-time branching particle filters is introduced with individual resampling: If there are $\mathbb{N}_n$ particles alive at time n, $\mathbb{N}_0 = N$, $a_n \le 1 \le b_n$, $\widehat{\mathbb{L}}^i_{n+1}$ is the current unnormalized importance weight for particle i and $\mathbb{A}_{n+1} = \frac{1}{N}\sum_{i=1}^{\mathbb{N}_n} \widehat{\mathbb{L}}^i_{n+1}$, then weight is preserved when $\widehat{\mathbb{L}}^i_{n+1} \in (a_n \mathbb{A}_{n+1}, b_n \mathbb{A}_{n+1})$. Otherwise, $\left\lfloor \frac{\widehat{\mathbb{L}}^i_{n+1}}{\mathbb{A}_{n+1}} \right\rfloor + \rho^i_n$ offspring are produced and assigned weight $\mathbb{A}_{n+1}$, where ρ^i_n is a Bernoulli of parameter $\frac{\widehat{\mathbb{L}}^i_{n+1}}{\mathbb{A}_{n+1}} - \left\lfloor \frac{\widehat{\mathbb{L}}^i_{n+1}}{\mathbb{A}_{n+1}} \right\rfloor$. The algorithms are shown to be stable with respect to the number of particles and perform better than the bootstrap algorithm as well as other popular resampled particle filters on both tracking problems considered here. Moreover, the new branching filters run significantly faster than these other particle filters on tracking and Bayesian model selection problems.

1. Introduction

Nonlinear filtering deals with determining the distribution of the current state of a non-observable, random, dynamic signal X given the history of a distorted, corrupted partial observation process Y living on the same probability space $(\Omega, \mathcal{F}, P)$ as X. Bayesian model selection, sometimes done while filtering, deals with determining which of a class of signal models $\{X^{(i)}\}_{i\in I}$ best fits the observed values of Y by pairwise Bayes' factor comparison. For many practical problems each potential signal is a time-homogeneous discrete-time Markov process $\{X_n,\ n = 0, 1, 2, \ldots\}$, living on some complete, separable metric space (E, ρ), with initial distribution π_0 and transition probability kernel K. The observation process takes the form ($Y_0 = 0$ and) $Y_n = h(X_{n-1}) + V_n$ for $n \in \mathbb{N}$, where $\{V_n\}_{n=1}^{\infty}$ are independent random vectors with common *strictly positive, bounded* density g that are independent of X, and the sensor function h is a mapping from E to $\mathbb{R}^d$. Then, the objective of filtering is to compute the conditional probabilities $\pi_n(A) = P\left(X_n \in A \,\middle|\, \mathcal{F}^Y_n\right)$, $n = 1, 2, \ldots,$ for all Borel sets A or, equivalently, the conditional expectations $\pi_n(f) = E^P\left(f(X_n) \,\middle|\, \mathcal{F}^Y_n\right)$ for bounded functions $f : E \to \mathbb{R}$, where $\mathcal{F}^Y_n \doteq \mathcal{B}\{Y_l, l = 1, \ldots, n\}$ is the information obtained (meaning the σ-algebra generated) from the back observations $\{Y_l, l = 1, \ldots, n\}$. The objective of Bayes factor model selection is to compare the ratio B^{12}_n of marginal likelihoods between potential signal models $X^{(1)}$ and $X^{(2)}$ with respect to some reference probability measure Q.

To do both filtering and model selection, a reference probability measure Q is introduced under which the signal, observation process $\{(X_n, Y_{n+1}),\ n = 0, 1, \ldots\}$ has the same distribution as the signal, noise process $\{(X_n, V_{n+1}),\ n = 0, 1, \ldots\}$ does under P. Hence, the observations are i.i.d. random vectors with strictly positive bounded density g and are independent of X under measure Q. All the observation information is absorbed into the likelihood process $\{L_n,\ n =$

E-mail address: michaelk@ualberta.ca.

http://dx.doi.org/10.1016/j.csda.2017.02.003

$1, 2, \ldots\}$ transforming Q back to P, which in our case has the form

$$\frac{dP}{dQ}\Big|_{\mathcal{F}^X_\infty \vee \mathcal{F}^Y_n} = L_n = \prod_{j=1}^n \alpha_j(X_{j-1}), \quad \text{with } \alpha_j(x) = \frac{g\left(Y_j - h(x)\right)}{g\left(Y_j\right)}, \tag{1.1}$$

so $L_n = \alpha_n(X_{n-1})L_{n-1}$ and $L_0 = 1$. (Here and in the sequel, $\mathcal{F}^X_n = \mathcal{B}(X_j, j \leq n)$ and $\mathcal{F}^X_\infty = \mathcal{B}(X_j, j \geq 0)$ are the σ-algebras generated by $\{X_j, 0 \leq j \leq n\}$ and $X_j, j \geq 0$ respectively.) The following (well-known) discrete Girsanov's theorem constructs the real probability P from the reference Q.

Theorem 1. *Suppose that $\Omega = (E \times \mathbb{R}^d)^\infty$, $\mathcal{F} = \mathcal{B}((E \times \mathbb{R}^d)^\infty)$, $\{X_n, n = 0, 1, \ldots\}$ and $\{Y_n, n = 1, 2, \ldots\}$ are independent processes on $(\Omega, \mathcal{F}, Q)$, the $\{Y_n\}$ are i.i.d. with strictly-positive, bounded density g on $\mathbb{R}^d$ and $V_n \doteq Y_n - h(X_{n-1})$ for all $n = 1, 2, \ldots$. Then, there exists a probability measure P such that (1.1) holds, $\{V_n, n = 1, 2, \ldots\}$ are i.i.d. on $(\Omega, \mathcal{F}, P)$ with density g and $\{X_n\}$ is independent of $\{V_n\}$ with the same law as on $(\Omega, \mathcal{F}, Q)$.*

The *unnormalized filters* are then

$$\sigma_n(f) = E^Q\left(L_n f(X_n)\,\middle|\,\mathcal{F}^Y_n\right), \tag{1.2}$$

so $\sigma_0 = \pi_0$, as $L_0 = 1$ and $\mathcal{F}^Y_0 = \{\emptyset, \Omega\}$ and the filter satisfies $\pi_n(f) = \frac{\sigma_n(f)}{\sigma_n(1)}$ by Bayes rule. Moreover, the Bayes factor satisfies $B^{12}_n = \frac{\sigma^{(1)}_n(1)}{\sigma^{(2)}_n(1)}$, where $\sigma^{(i)}_n(f) = E^Q\left(L^{(i)}_n f\left(X^{(i)}_n\right)\middle|\mathcal{F}^Y_n\right)$, with $L^{(i)}_n = \prod_{j=1}^n \alpha_j(X^{(i)}_{j-1})$, is the unnormalized filter for signal model $X^{(i)}$. Therefore, we can combine Bayesian model selection *and* filtering (for each potential signal) by constructing approximations (denoted $\mathbb{S}^N_n$ below) to the unnormalized filter for each candidate signal model. As words of caution, our setting is certainly not the most general possible for our unnormalized and branching particle filter approach as we do not want to over complicate the setting and conditions while introducing new methods. Indeed, it is anticipated that with some work other observation models can be used and some form of Lookahead Sequential Monte Carlo strategy related to those considered in Lin et al. (2013) could be developed based upon our branching particle algorithms. However, there are also interesting situations like rare event importance sampling (see Le Gland and Oudjane, 2006) that appear ill-suited for our approach, even with a more general setting.

1.1. Background

Particle filters are utilized widely and the original (resampled) interacting particle filters have been intensely studied (see e.g. Del Moral and Miclo, 2000 and Cappe et al., 2007 for an overview and historical account). However, particle filters performance depends heavily upon at least two factors:

- The importance density proposals used for sampling, and
- The resampling method used,

with both being active areas of investigation and the later claim being justified in e.g. Del Moral et al. (2001), Douc et al. (2005) and Hol et al. (2006). Moreover, resampling is the most difficult and critical step to parallelization as is pointed out in Murray et al. (2016) so effective replacement by branching may be even more valuable in parallel implementations. Furthermore, resampled particle filters approximate the actual filter π_n so prior filter estimates must be stored to perform Bayes factor model selection. On the other hand, the weighted particle filter (credited to Handschin, 1970; Handschin and Mayne, 1969) approximates the unnormalized particle filter σ_n, is the most basic particle filter and is embarrassingly computer parallelizable. More generally, branching particle filters, like those introduced by Crisan and Lyons (1997), can have model selection capabilities, effective resampling and be highly parallelizable. However, branching particle filters suffer from dramatic particle swings and difficult analysis—or do they? Herein, we introduce and analyze branching particle filters that avoid the weighted-particle-filter particle spread problems yet still have immediate model selection capabilities. They include the weighted particle filter as the extreme zero-resampling case and a branching variation of the better algorithm in Del Moral et al. (2001) as the fully-resampled case. They are stable with respect to particle number swings and can be analyzed using exchangeability (in lieu of independence). In order to focus just on our branching scheme, we ignore possible (large, problem-dependent) gains attainable by using alternative importance sampling density proposals and stick to sampling from the signal dynamics.

There are many approaches to reducing resampling noise in the basic bootstrap filter. For example, researchers brought in importance sampling and delayed bulk resampling methods (see e.g. Del Moral et al., 2012). Others have introduced less noisy types of resampling, which we discuss below. However, there are few studies like Ballantyne et al. (2000) of the practical partially-resampled algorithms where decisions are made on a particle-by-particle basis with the aim of only removing the poor particles and splitting the best particles (in an unbiased manner). Kouritzin and Sun (2005) do obtain L_2-rates of convergence for a partially-resampled algorithm in a specific setting. Our present work introduces new classes of branching particle filters, motivates their use and sets up a framework for studying them. We refer the reader to standard