

Asymptotic analysis of genealogies induced by sequential Monte Carlo algorithms

Suzie Brown

March 6, 2019

1 Introduction

- what is SMC and what is it used for?
- basic SMC algorithm
- the three main tasks: prediction, filtering, smoothing
- the problem with smoothing: ancestral vs. weight degeneracy
- motivating plots
- organisation of the report

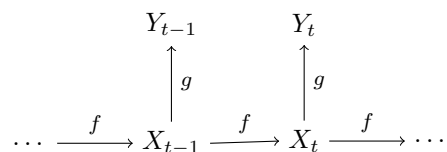
[SMC is super useful for loads of stuff...]
[Define subscript notation $(\cdot)_t$ and $(\cdot)_{0:t}$]

1.1 Class of models

For our purposes it is sufficient to describe sequential Monte Carlo (SMC) algorithms in the context of inference on a time-homogeneous state space model. Consider the following model:

$$\begin{aligned} X_0 &\sim \mu(\cdot) \\ X_{t+1} | (X_t = x_t) &\sim f(\cdot | x_t) \quad t = 0, \dots, T-1 \\ Y_t | (X_t = x_t) &\sim g(\cdot | x_t) \quad t = 0, \dots, T \end{aligned}$$

where $(X_t)_{t=0}^T$ is an unobservable discrete-time Markov process and the observables $(Y_t)_{t=0}^T$ satisfy $Y_t \perp \{Y_{-t}, X_{-t}\} | X_t$. The conditional independence structure is described by the following graphical model.



We assume that the *transition* and *emission* kernels have densities which are denoted by f and g respectively, but this is not necessary in general. We only require that we can sample from $\mu(\cdot)$ and $f(\cdot|x)$, and calculate *unnormalised* potentials $g(y|x)$, for all x, y .

1.2 Inference in state space models

Suppose we are in a Bayesian setting, where μ is our prior distribution at time 0, observations y_t arrive sequentially, and we want to infer information about the hidden states (either on- or off-line). The three main inference problems are:

Filtering (where is it now?) $p(x_t | y_{0:t})$

Prediction (where will it go next?) $p(x_{t+1} | y_{0:t})$

Smoothing (where has it been?) $p(x_{0:t} | y_{0:t})$

In the on-line setting, we take as our prior the posterior distribution from the previous time step $t - 1$, and update it using the new observation y_t . The inference must be fast enough to keep up with the rate of arrival of observations, so in particular the complexity of our inference algorithm must not increase with T . In the off-line setting, we take μ as the prior distribution, and infer the set of posteriors once all $T + 1$ observations have arrived.

Prediction and filtering are essentially equivalent, because given a filtering distribution, the corresponding predictive distribution can be obtained by applying the transition kernel f . Smoothing is considered a harder task because it requires us to infer many more parameters from the same amount of information; indeed the dimension of the problem increases linearly with T . In general it is not possible to infer the smoothing distributions on-line, because this would require using all of the previous observations at each time step, so the complexity would increase at each step. Algorithms that exist for on-line smoothing [REFs] are typically either approximate or computationally expensive.

In the case of linear Gaussian state space models (i.e. where f and g are Gaussian densities that depend only linearly on their arguments), the posterior distributions of interest are available analytically, by way of the Kalman filter [REF] and Rauch-Tung-Striebel (RTS) smoother recursions [REF]. The other analytic case occurs if the state space of $(X_t)_{t=0}^\infty$ is finite, in which case the forward-backward algorithm [REF] yields the exact posteriors.

2 SMC as a coalescent

- k-coalescent & Kingman coalescent
- pop gen literature about large population cts time limits of various models
- resampling viewed backwards in time: branching process \rightarrow coalescent process
- asymptotic properties of SMC lit review: CLT, path storage, coalescence etc.
- the gap in knowledge that we aim to fill
- (remark: although SMC has other problems in high dimension, the coalescence rate doesn't depend on the dimension...)

3 Conditional SMC

- motivation: particle MCMC, need for multiple lineages
- conditional (multinomial) SMC algorithm and its context within particle Gibbs
- result: coalescence rate etc in terms of standard multinomial one; verification of assumptions of KJJS theorem (but exile horrible calculations to appendix)
- simulations & conclusions thence

4 Alternative resampling schemes

- multinomial not really used in practice, but other schemes hard to analyse
- overview of the main variance-reducing schemes
- results: theorem for residual resampling (hopefully)
- maybe results for other schemes
- simulation comparing all of them, and conclusions thence

5 Discussion

- results so far
- impact of this work: to practitioners, to enriching the SMC literature, interpretation within pop gen.
- future directions