SMC genealogies

Suzie Brown

May 25, 2018

1 Introduction

2 Background

2.1 Sequential Monte Carlo

2.1.1 Conditional SMC

2.2 Population genetics

The Wright-Fisher model, popular in the analysis of population genetics, bears some simplifying assumptions that render it unrealistic for that application. Namely, the population size is assumed to remain constant, and the generations non-overlapping. While these constraints may hamper the model's applicability to population genetics, it is pleasing to note that when re-purposed for the analysis of SMC genealogies, both of these rather restrictive assumptions apply automatically. At least in the most common SMC algorithms, the number of particles (population size) remains constant at each iteration, and the resampling (reproduction) procedure is applied to all particles at once.

It seems then that the myriad tools and results developed for this model by population geneticists should be seamlessly transferable to the analysis of SMC genealogies. In reality, a significant alteration must be made to the standard Wright-Fisher model before it can be of use in this context: the offspring distributions of each particle are not exchangeable, because broadly speaking, offspring of parents with high weight will tend to have high weight themselves. This follows intuitively from the notion that propagating a particle from a high density region at time t will result in a new state that is close to the high density region at time $t + 1$.

This effect can naturally be interpreted in the population genetics framework as the existence of a hereditary trait whose value affects fertility. A typical and well-studied question in the population genetics literature addresses the proliferation of such a trait. For example, in the Wright-Fisher model it is possible to study the dynamics of the proportion of the population having the trait of interest, ultimately in the limit as the number of generations goes to infinity. However, the question of how prevalent the "high density" trait may be in our population of particles is tangential to our focus (the resampling step is constructed in such a way as to maintain the desired dynamics, i.e. consistent estimation of the target distribution). Our interests lie instead with the genealogy of the particles, regardless of the properties they thus inherit.

3 Previous work

4 Theoretical results

5 Simulation study

5.1 Rauch-Tung-Striebel smoother

6 Conclusions