Optimal Point Process Filtering of the Coalescent Process

Kris V Parag, Oliver G Pybus,
Department of Zoology, University of Oxford, Oxford, UK
E-mail: kris.parag@zoo.ox.ac.uk

Abstract

This manuscript explores the viability of using optimal point process (Snyder) filters in order to estimate the underlying parameters of the variable population size coalescent process. Estimating these population parameters is an important area of research in phylodynamics, especially given the widespread use of the coalescent in modelling the relationship between the genetic diversity and epidemiological dynamics of human pathogens. A variety of coalescent estimators based on a diverse set of techniques, such as skyline plots, Bayesian and Markov Monte Carlo approaches, already exist. However, at times these methods are inflexible, or difficult to use and there is a need to explore new estimation techniques.

This Snyder filter is proposed here as a new alternative for optimal coalescent inference and parameter estimation. Through its application, first to a canonical set of demographic models and then to empirical data from the Hepatitis C epidemic in Egypt, the filter is shown both useful and capable. The Snyder filter is exact (makes no process approximations) and is optimal in mean square error. Since its implementation is simple and it was originally developed to estimate stochastic parameters, Snyder filtering holds much potential for coalescent estimation.

Introduction

The Coalescent Process

The coalescent process is a dominant theory in phylodynamics that links the genetic diversity of a sampled population to its demographic history. Capable of describing a wide variety of biological phenomena [1], the coalescent has been found to be the convergent process of many fundamental neutral evolutionary models such as the Wright-Fisher, Moran, and their variants. Initially developed for a constant sized population by Kingman [2] the coalescent has been extended to account for time varying populations [3], geographically structured populations [4] and for populations sampled at different times [5].

Coalescent theory has been successfully applied to a wide variety of phylodynamical problems. It has been used to model and infer the discontinuous growth of the Hepatitis C epidemic in Egypt [6], the oscillating behaviour of Dengue in Vietnam [7] and to even calculate the generation time of rapidly mutating viruses like HIV-1 [5]. Given its usefulness and popularity, it is no surprise that the coalescent is an important and well studied process. While the constant population size model is already completely specified, work on the variable population generalisation is still ongoing [8].

In all variants of the coalescent, it describes how a sample of lineages of size n, from a population of size $N(t)$, converges to its most recent common ancestor. Coalescent events indicate when two lineages from that sample have converged to a common parent (in which case the lineage count falls by 1). Such events continue occurring until only one common parent or ancestor is attained. In this setting $t = 0$ indicates the present and $t = \text{TMRCA}$ is the process stopping time (time to most recent common ancestor). As a descriptor of the genetic diversity of a population, the coalescent is predicated on two main assumptions:

- Neutral evolution: there is no selection pressure so the genealogical and mutation processes are independent and separable. This not only allows one to introduce mutations later via an independent Poisson process or nucleotide substitution model, but also ensures one can explore the genealogy of a sample from a population without worrying about the rest of the population [1]. Consequently, this assumption makes inference simpler and more efficient.

- Small sample size: the lineage sample size n is much smaller than the underlying population $N(t)$. This assumption is important for the coalescent to maintain its biological significance as the convergent process of population genetics models like the Wright-Fisher. Mathematically the coalescent involves an approximation from a discrete geometrical process to a continuous exponential one. A key condition for this diffusion approximation is that $\frac{n}{N(t)} \to 0$.

The original forms of the coalescent also included auxiliary assumptions such as a constant, panmitic population, and no recombination [9]. However, depending on the variant of the process studied these can be relaxed. The following work maintains the main assumptions and further assumes a panmitic, isochronously sampled haploid population with no recombination.

This manuscript initially analyses the original constant population size coalescent process from an information theoretic and point process filtering perspective before shifting focus to the estimation of complex, deterministically varying populations. No treatment of the other variants beyond a mention of how to account for stochastic demographic functions, is provided. Such extensions are the likely topics of future research. The main results presented show that optimal Snyder filtering can be used to achieve effective, minimum mean squared estimation of demographic histories, that can be easily implemented.

Existing Inference Schemes for the Coalescent

Coalescent inference has been a major topic of study in the fields of phylodynamics and population genetics [10]. Consequently a plethora of parametric and non-parametric methods exist for estimating its underlying demographic history. This study focuses on parametric methods since non-parametric schemes should be used more as a tool for model selection [11]. This work therefore assumes that a suitable model for the population $N(t, \vec{x})$ has already been chosen and that its parameters $\vec{x}$ or a function of the parameters are to be estimated optimally in some way from the phylogenetic data.

Phylodynamic inference methods can be split into those that use maximum likelihood (ML) or Bayesian analysis. This work focuses on the latter since comparisons until now have found that Bayesian approaches generally outperform or are at least as good as corresponding ML ones [12]. Moreover likelihood functions are difficult to compute for complex demographic functions [13]. Existing Bayesian methods often use Markov Chain Monte Carlo and importance sampling [14]. These approaches, while capable and able to account for genealogical uncertainty, can be complex, inflexible or difficult to implement [15], especially when they involve integration over multidimensional spaces.

This manuscript will show how the Snyder filter can be used as an alternative and useful parametric estimator. While the developed form does not yet account for genealogical uncertainty and instead focuses on data from a single coalescent tree, the Snyder filter presents a unique, easily adaptable and different perspective on Bayesian coalescent inference that very naturally and directly treats the coalescent data in its full richness as a point process stream.

As a method of estimation this technique has remained largely unknown to the biological sciences with its only applications so far being to neuronal spiking by Bobrowski *et al* [16] and to invertebrate visual phototransduction by Parag and Vinnicombe [17]. In both cases new and interesting results emerged by taking this analytical viewpoint to estimation. It will be seen that the Snyder filter presents a fresh Bayesian approach to inference with much potential for phylodynamics.

Defining the Coalescent as a Point Process

The standard coalescent with n lineages and effective population $N(t)$, $\forall t \geq 0$ can be described as an inhomogeneous Poisson process with a maximum count of $n - 1$. In this interpretation the $n - 1$ point coalescent event stream is separated into inter-event intervals and the coalescent rate is described for each interval by conditioning on the currently existing number of lineages [3]. All the information for estimation is then compactly contained in the timing of the coalescent events. Such a treatment, while not new to the literature [9] has only recently been used as a starting point for inference (previous schemes started with genealogies and integrations about a tree space) [8].

This work extends this description by explicitly incorporating the fall in lineages, due to stochastic coalescent events, into the rate. This contrasts the usual coalescent description since it removes the need for direct conditioning. This allows the coalescent to then be redefined as a self-correcting inhomogeneous Poisson process. Thus the salient difference between this description and those in the literature is simply an understanding that lineage deaths are equivalent to process feedback.

A self-correcting process is dependent on past events in a manner which hinders the occurrence of future events. This feature results from the death process description of the coalescent [2], which means that the rate of producing coalescent events falls with the number of events. This dependence is summed up by noting that the process has negative autocovariance over distinct time intervals, as do its converging Wright-Fisher and Moran models over distinct generations [9].

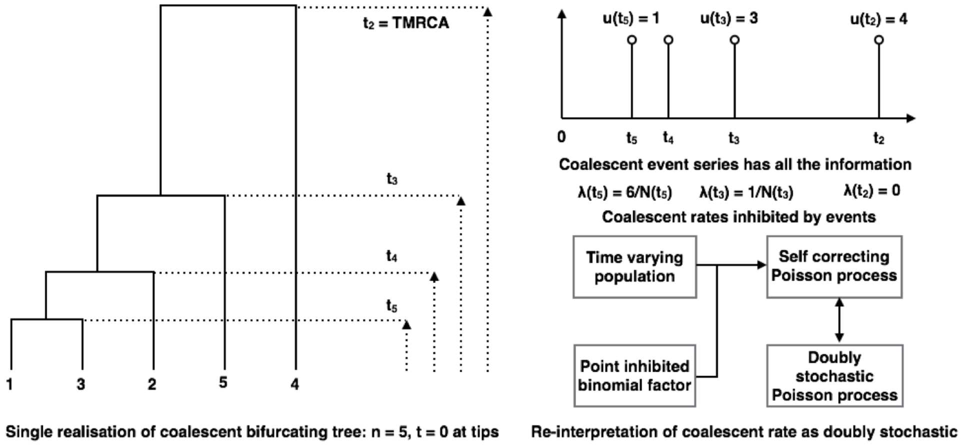
Define $\{\mathcal{D}(s) : 0 \leq s \leq t\} := \mathcal{D}_0^t$ be the counting process with points at coalescent times: $\{t_k\}$, $\forall k \in \mathbb{Z}_2^n : t_k \leq t$ and let $u(t) = |\mathcal{D}(t)|$ be the number of points up to time t. Here t_k is the coalescent time for the $k \rightarrow k-1$ transition in lineages, which means $t_k < t_{k-1}$, $\forall k$. Further $\lambda\left(t, \mathcal{D}_0^t\right)$ is the feedback dependent rate of this Poisson process. Setting $\mathcal{D}(0) = 0$, the coalescent in self correcting form is:

$$\mathcal{D}(t) \sim \text{Poiss}\left(\lambda\left(t, \mathcal{D}_0^t\right)\right) \tag{1}$$

$$\lambda := \lambda\left(t, \mathcal{D}_0^t\right) = \binom{n - u(t)}{2} \frac{1}{N(t)} \tag{2}$$

$$\text{cov}\left(\mathcal{D}(t-s), \mathcal{D}(t)\right) < 0 \tag{3}$$

Since $\lambda \geq 0$, $\forall t \geq 0$ then $\max_t(u(t)) = n-1$, which is the number of coalescent events for n lineages, with $\frac{n}{N(t)} \ll 1$, $\forall t \geq 0$. The fact that self-correcting inhomogeneous Poisson processes can be reinterpreted as doubly stochastic Poisson processes (Poisson processes with stochastic intensities) is the key insight that allows the application of point process filtering techniques developed by Snyder to the coalescent [18]. A summary of the notation used and the coalescent - Poisson process reinterpretation are given in figure 1.


t₂ = TMRCA
t₃
t₄
t₅
1
3
2
5
4
u(t₅) = 1
u(t₃) = 3
u(t₂) = 4
0
t₅
t₄
t₃
t₂
Coalescent event series has all the information
λ(t₅) = 6/N(t₅)
λ(t₃) = 1/N(t₃)
λ(t₂) = 0
Coalescent rates inhibited by events
Time varying population
Self correcting Poisson process
Point inhibited binomial factor
Doubly stochastic Poisson process

Single realisation of coalescent bifurcating tree: n = 5, t = 0 at tips

Re-interpretation of coalescent rate as doubly stochastic

Figure 1. Coalescent process and definition of notation. A sample bifurcating coalescent tree is shown in the main left panel for 5 samples lineages. The coalescent times $t_5, \ldots t_2$ are labelled and their event stream shown in the top right. The counting function $u(t)$ and the rate $\lambda(t)$ are also calculated. The bottom left panel presents the coalescent process interpretation for Snyder filtering.

Methods

Optimal Snyder Filtering

The Snyder filter is an exact Bayesian filter that provides the optimal causal reconstruction of the underlying intensity of a Poisson process given observations of its points $\mathcal{D}_0^t$ and a model with priors [18]. If $\mathcal{F}_t$ represents all the information carrying data then $\mathcal{F}_t = \mathcal{D}_0^t$. The parameter estimates are optimal according to a mean squared error criterion. Originally developed for doubly stochastic Poisson processes or Poisson process with stochastic rates, the filter continuously solves ordinary differential equations for the model posteriors between observed event times with discontinuous updates at event times.

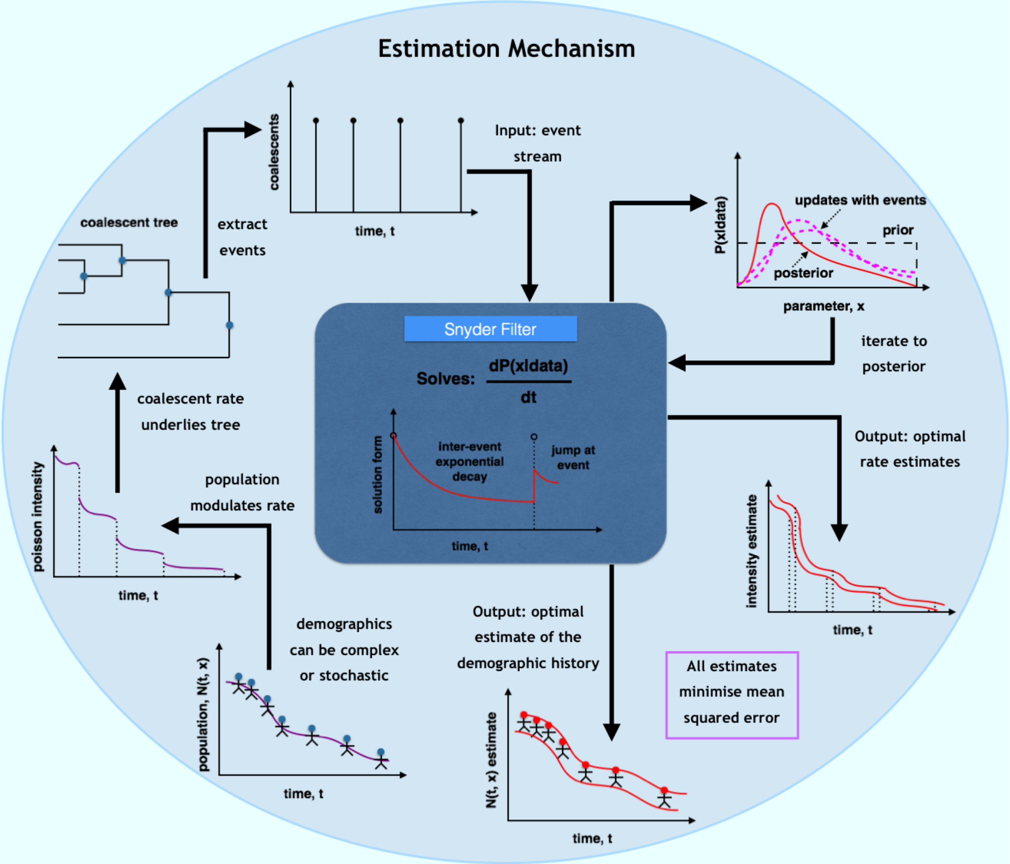


Figure 2. Summary of the Snyder filter mechanism for the coalescent. From the bottom left and moving clockwise: an underlying demographic function which is known for simulated datasets but unknown for empirical ones is to be estimated. It modulates the coalescent rate which has discontinuities due to lineage deaths (coalescent events). This rate results in a bifurcating tree from which the coalescent event stream is obtained and input to the filter. The filter solves posterior differential equations which iteratively results in the priors converging to data conditioned posteriors for the model parameters. This results in opimal MMSE estimates of the coalescent rate and the underlying population.

Since, as noted in the introduction, the coalescent process is a self-correcting Poisson process, which falls within the doubly stochastic framework, the filter can be adapted to estimate demographic history.

The mechanics of this adaptation are summarised in figure 2. The filter only requires inputs of coalescent times and parameter priors and outputs posterior distributions from which MMSE estimates are obtained. The computational complexity of this method depends on the number of parameters to estimate, l, the number of lineages, n and the dimension of the filter, m (defined in the following equations). Since it requires only the solution of linear differential equations it is quite easy to implement and requires virtually no tweaking. The filter is, however, dependent on the quality of the coalescent time data. The mathematical development of the filter for the coalesent follows.

Let a general multivariate population function with l parameters $\vec{x} = [x_1, \, x_2, \, \dots x_l]$ be denoted as $N(\vec{x}, \, t)$ with parameter x_i defined on the domain $\mathcal{X}_i$ with cardinality $|\mathcal{X}_i| = m_i$ discretised points. Then the joint prior, $\mathbb{P}(\vec{x})$ is defined on a space of $m = \prod_{i=1}^{l} m_i$ points. As the filter solves a differential equation on each possible point dimension, its dimension is m . In this framework the coalescent rate function is defined as:

$$\lambda = \lambda(\vec{x}, \, t, \, \mathcal{D}_0^t) : \otimes_{i=1}^{l} \mathcal{X}_i \, \otimes \, \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+ = \binom{n - u(t)}{2} \frac{1}{N(\vec{x}, \, t)} \tag{4}$$

In the above the function mapping is also shown with $\otimes$ indicating a Kronecker ordering. As an example of the Kronecker ordering for $l = 2$, if $\mathcal{X}_1 = \{a_1, \, a_2\}$ and $\mathcal{X}_2 = \{b_1, \, b_2, \, b_3\}$ then the probability vectors have 6 values pertaining to the sets $\{a_1, \, b_1\}$, $\{a_2, \, b_1\}$, $\{a_1, \, b_2\}$, $\{a_2, \, b_2\}$, $\{a_1, \, b_3\}$ and $\{a_2, \, b_3\}$ respectively.

If the posterior $\mathbb{P}\left(\vec{x}|\mathcal{F}_t\right)$ is denoted in normalised and unnormalised form as $q(t)$ and $q^*(t)$ respectively, with the posterior joint vector Kronecker ordered on the space of $\vec{x}$, then the Snyder filter can be written as below. The use of non-normalised probabilities, as developed in [19], converts the original non-linear Snyder differential equations into a linear set that need only be normalised at every event. Here $q_j = \mathbb{P}\left(\vec{x} = \vec{x}_j|\mathcal{F}_t\right)$ with q_j^* as the corresponding unnormalised form and $\vec{x}_j$ as the j^{th} configuration of the parameter vector which has m configurations. For convenience of notation the rate and population are sometimes referred to as simply $\lambda(t)$ and $N(t)$. The diagonal matrix $\Lambda_{u(t)} = \text{diag}\left(\{\lambda_j(t), \forall j \in \mathbb{Z}_1^m\}\right)$ is called the rate matrix, is of dimension $m \times m$ with $\Lambda_{u(t)}[j]$ as its j^{th} diagonal component. It changes at event times due to $u(t)$.

$$\frac{\mathrm{d}q_j^*(t)}{\mathrm{d}t} = -q_j^*(t)\Lambda_{u(t)}[j] \tag{5}$$

$$\Lambda_{u(t)}[j] = \frac{1}{N\left(\vec{x}_j, \, t\right)} \binom{n - u(t)}{2} = \lambda_j(t) \tag{6}$$

$$q_j(t) = \frac{q_j^*(t)}{\sum_{j=1}^{m} q_j^*(t)} \tag{7}$$

$$q_j(t_k^+) = \frac{q_j(t_k^-)\Lambda_k[j]}{\sum_{j=1}^{m} q_j(t_k^-)\Lambda_k[j]} \tag{8}$$

Equations 5 - 7 describe the evolution of the posterior between event times. Since these equations are linear the solution is a matrix exponential on the probability components. At event times this solution is discontinuously renormalised by the rate matrix as in equation 8. Here t_k^- and t_k^+ are infinitesimally before and after the event $\implies u(t_k^-) = n - k$ and $u(t_k^+) = u(t_k^-) + 1$.

The conditional estimate of the effective population, coalescent rate and the parameters are obtained from the posterior as follows. Note: A^{T} indicates the transpose of A.

$$\hat{x}_i(t) = \mathbb{E}\left[x_i(t)|\mathcal{F}_t\right] = \left[\sum_{\mathcal{X}_j,\,\forall j\neq i} q(t)\right] \mathcal{X}_i^{\mathrm{T}} \tag{9}$$

$$\hat{\lambda}(t) = \mathbb{E}\left[\lambda(t)|\mathcal{F}_t\right] = \sum_{j=1}^{m} q_j(t)\Lambda_{u(t)}[j] \tag{10}$$

$$\hat{N}(t) = \mathbb{E}\left[N(t)|\mathcal{F}_t\right] = \frac{\binom{n-u(t)}{2}}{\sum_{j=1}^{m} q_j(t)\Lambda_{u(t)}[j]} \tag{11}$$

The power of the this filter is demonstrated in this work on both standard simulated and empirical data sets. In these test cases the joint and marginal priors are uniformly defined as follows together with a posterior bound, $\mathcal{B}$ which is forced to be non-negative (negative values are meaningless). Assume the m_i values of x_i are uniformly and independently distributed between some minima and maxima, so that for $1 \leq j_i \leq m_i$, $i \in \{1,\, 2,\, \ldots l\}$, then:

$$\mathbb{P}(\vec{x} = \vec{j}) = \prod_{i=1}^{l} \mathbb{P}(x_i = x_i(j_i)) = \prod_{i=1}^{l} \frac{1}{m_i} = \frac{1}{m} \tag{12}$$

$$\mathcal{B} = \max\left(0,\, \mathbb{E}\left[x_i|\mathcal{F}_t\right] \pm 2\sqrt{\left(\mathrm{var}\left(x_i|\mathcal{F}_t\right)\right)}\right) \tag{13}$$

Thus the filter takes a coalescent time series $\mathcal{F}_t$ and parameter priors $\mathbb{P}(\vec{x})$ as its input, and iterates to a posterior $\mathbb{P}\left(\vec{x}|\mathcal{F}_t\right)$. The conditional mean estimates above minimise mean squared error. If $y(t)$ is some function of interest (example $y(t) = N(t)$ or $y(t) = x_1$) then the conditional mean estimate $\hat{y}(t)$ achieves the minimum mean squarer error (MMSE) which is defined:

$$R = \mathbb{E}\left[\left(y(t) - \mathbb{E}\left[y(t)|\mathcal{F}_t\right]\right)^2\right] = \mathbb{E}\left[\left(y(t) - \hat{y}(t)\right)^2\right] \tag{14}$$

In simulations two kinds of MMSE are used as variants of R above. The first, $R_t(y(t))$, is across the time for a run, T and is more a measure of the filter performance across a $n-1$ event set. The second, $R_r(y(t))$ is across M repeated runs for the $n-1$ events with the estimate taken at the end of each run, T_i. Here $T:\, u(T) = n-1$ and T_i is such a stopping time for the i^{th} run. Percentage relative MMSEs, J_t and J_r are also calculated. These indices will be used to quantify estimation performance in the following results section.

$$R_t(y(t)) = \frac{1}{T}\int_0^T \left(y(t) - \hat{y}(t)\right)^2 \,\mathrm{dt},\; J_t(y(t)) = \frac{1}{T}\int_0^T 100\left(1 - \frac{\hat{y}(t)}{y(t)}\right)^2 \,\mathrm{dt} \tag{15}$$

$$R_r(y(t)) = \frac{1}{M}\sum_{i=1}^{M} \left(y(T_i) - \hat{y}(T_i)\right)^2,\; J_r(y(t)) = \frac{1}{M}\sum_{i=1}^{M} 100\left(1 - \frac{\hat{y}(T_i)}{y(T_i)}\right)^2 \tag{16}$$

It is worth noting that the filter mechanism remains largely unchanged even if it is extended to more complex demographic functions than those explicitly dealt with in this script. For example, if $N(t)$ was stochastic and describable as a continuous time Markov process with infinitesimal generator Q then the only alteration would involve replacing $\Lambda_{u(t)}$ with $\Lambda_{u(t)} - Q$ in equation 5 (with appropriate changes to prior definitions). Additionally, the filter can be simplified to perform similar inference for normal inhomogeneous processes as well as combined into a more involved form to handle complex Poisson processes which have feedback dependent and stochastic multimodal rates [16].

Results and Analysis

Constant Intensity Coalescent

Consider the original constant population Kingman coalescent with $N(t) = N_0$. The time interval for this process to fall from k to $k-1$ lineages, δ_k is known to follow the exponential distribution:

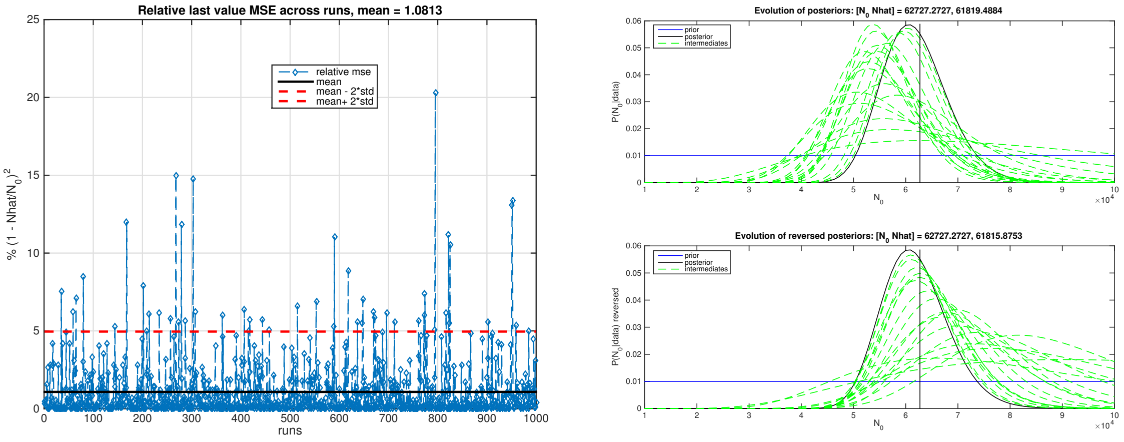
$$\delta_k = t_{k-1} - t_k \sim \exp\left(\binom{k}{2}\frac{1}{N_0}\right) \tag{17}$$

$$\mathbb{E}\left[\delta_k\right] = \frac{2N_0}{k(k-1)} \tag{18}$$

The causal estimation problem is then to find $\hat{N}_0(t) = \mathbb{E}\left[N_0(t)|\mathcal{F}_t\right]$. Since this involves estimating a single parameter on which the filter converges, then for a given evolution time, $T : u(T) = n-1$, the best estimate of the parameter N_0 is $\hat{N}_0(T)$.

For this problem the uniform prior is constructed by assuming that the minimum and maximum value possible are: $N_{\min} \gg n$ and $N_{\max} > N_{\min}$ and then that N_0 can take 1 of m values uniformly within this range with equal probability. Thus: $\mathbb{P}\left(N_0 = a_i = N_{\min} + (i-1)\frac{N_{\max}-N_{\min}}{m-1}\right) = \frac{1}{m}, \forall i \in \{1, 2, \ldots m\}$. As an examination of its behaviour, the filter is also run on the data in reverse. In this case the coalescent times from a standard run are reversed (the inter-event times are summed in reverse order) and the rows of $\Lambda_{u(t)}$ reordered appropriately so that the first event has rate $\lambda(t) = \frac{1}{N_0}$ and the last event $\lambda(t) = \binom{n}{2}\frac{1}{N_0}$.

The simulation results are for $N_{\min} = 100n$, $N_{\max} = 1000n$ with $n = 100$ and $m = 100$. Figure 3a plots the square error between N_0 and $\hat{N}_0$ from the filter across $M = 1000$ runs. The mean of these points would be J_r. While not shown, as expected, estimates were found to improve with n (J_r decreases). Figure 3b shows that reversed and forward time posteriors are essentially convergent and suggests that data can be equally well processed in either time direction.



(a) Square errors with mean J_r (normal data) **(b)** Snyder posteriors $q(t)$ for normal and reversed data

Figure 3. MMSE estimators and performance for constant size population model. Panel 3a gives the relative square errors between N_0 and its estimate across 1000 runs. Panel 3b shows for a single simulated coalescent dataset that the normal and reversed filter converge to the same posterior.

Comparing Independent Trees (Multiple Loci) to a Single Coalescent Tree: the Snyder filter treats the constant size coalescent as a self-correcting Poisson process. It will be seen that this can be simplified, after first examining the relative efficiency of the different types of constant population data that can be used. Let $\eta = \frac{1}{N_0}$ and initially consider the data from a single tree. For n lineages a single tree provides $n-1$ coalescent times, with inter-event intervals δ_k, $2 \leq k \leq n$ and distribution given in equation 17. As in [20], the likelihood function, $L_1(\eta)$, can be written as below, by using the independence of the coalescent intervals.

$$L_1(\eta) = \left[\prod_{i=2}^{n} \binom{i}{2}\right] \eta^{n-1} e^{-\eta \sum_{i=2}^{n} \binom{i}{2}\delta_i} = [h_1(n)] g_1(\eta,\, T_1(\delta)) \tag{19}$$

$$T_1(\delta) = \sum_{i=2}^{n} \binom{i}{2}\delta_i,\ \binom{i}{2}\delta_i \sim \exp(\eta) \tag{20}$$

The sufficient statistic for this estimation problem is $T_1(\delta)$, in accordance with the Fisher-Neyman factorisation. The $\binom{i}{2}\delta_i$ distribution comes from the scaling property of exponentials.

Now assume a single r^{th} coalescent event is observed from each of $n-1$ independent trees. The data are now $\tau_r(i)$ for $i = 2, 3, \ldots n$ where each $\tau_r(i) \sim \exp(\eta)$. Factorising the likelihood, $L_2(\eta)$ gives:

$$L_2(\eta) = \left[\binom{r}{2}^{n-1}\right] \eta^{n-1} e^{-\eta \sum_{i=2}^{n} \binom{r}{2}\tau_r(i)} = [h_2(n)] g_2(\eta,\, T_1(\delta)) \tag{21}$$

$$T_2(\tau_r) = \sum_{i=2}^{n} \binom{r}{2}\tau_r(i),\ \binom{r}{2}\tau_r(i) \sim \exp(\eta) \tag{22}$$

For both cases, the maximum likelihood estimator (MLE) can be obtained:

$$\frac{\partial L_1}{\partial \eta} = 0 \implies \frac{\partial L_1}{\partial N_0} = -\frac{1}{N_0^2}\frac{\partial L_1}{\partial \eta} = 0 \implies \hat{N}_{1\text{mle}} = \frac{1}{\hat{\eta_1}} = \frac{T_1(\delta)}{n-1} \tag{23}$$

$$\frac{\partial L_2}{\partial \eta} = 0 \implies \frac{\partial L_2}{\partial N_0} = -\frac{1}{N_0^2}\frac{\partial L_2}{\partial \eta} = 0 \implies \hat{N}_{2\text{mle}} = \frac{1}{\hat{\eta_2}} = \frac{T_2(\tau_r)}{n-1} \tag{24}$$

Generally $\hat{N}_0 \neq \frac{1}{\hat{\eta}}$. It applies in this case because the same value of $\hat{N}_0$ results from solving $\frac{\partial L_1}{\partial \eta} = 0$ and $\frac{\partial L_1}{\partial N_0} = 0$. Since both sufficient statistics break down into a sum of $n-1$ independent exponential variables with rate η and both MLEs only depend on the sufficient statistic and the sample number, then it is equally efficient to sample from a single tree or multiple trees. Here X_i is a sample from the $\exp(\eta)$ distribution.

$$\text{var}\left(\hat{N}_{1\text{mle}}\right) = \text{var}\left(\hat{N}_{2\text{mle}}\right) = \frac{\text{var}\left(\sum_{i=1}^{n-1} X_i\right)}{(n-1)^2} = \frac{N_0^2}{n-1} \tag{25}$$

This observation means that there is no more information in the coalescent times of one tree than in the last coalescent time from independent trees, for a constant population. This can be formally shown by noting that expression 26 implies equality 27 by the properties of sufficient statistics.

$$\mathbb{P}(T_1(\delta)) \overset{d}{=} \mathbb{P}(T_2(\tau_r)) \text{ and } \mathbb{P}(T_1(\delta)|\eta) \overset{d}{=} \mathbb{P}(T_2(\tau_r)|\eta) \tag{26}$$

$$\mathcal{I}(\eta; \mathcal{D}_1) = \mathcal{I}(\eta; T_1(\delta)) = \mathcal{I}(\eta; T_2(\tau_r)) = \mathcal{I}(\eta; \mathcal{D}_2) \tag{27}$$

Here the $\mathcal{D}_i$ indicate different forms of coalescent data, $\overset{d}{=}$ means equivalent in distribution and $\mathcal{I}(X;\, Y)$ is the mutual information between random variables X and Y. A key result of this analysis is that since information is preserved, one can generate $n-1$ variables from a $\exp\left(\frac{1}{N_0}\right)$ distribution and reformulate the coalescent inference from that for a self-correcting homogeneous Poisson process to that for a simple homogeneous Poisson process.

Available Analytical Results: given previously shown information equalities, the Snyder filter can be reformulated so that the Poisson process is homogeneous instead of self-correcting. This only affects the rate matrix since $\Lambda_{u(t)}$ is replaced with $\Lambda = \text{diag}\left(\{a_i : \forall i \in \mathbb{Z}_1^m\}\right)$. This transformation allows one to use the explicit result obtained in [21] for the posterior, $\mathbb{P}(\eta = x|\mathcal{F}_t)$ when estimating a constant intensity Poisson process with prior $\mathbb{P}(\eta = x)$.

$$\mathbb{P}(\eta = x|\mathcal{D}_0^t) = \frac{x^{u(t)}e^{-xt}\mathbb{P}(\eta = x)}{\int_0^\infty x^{u(t)}e^{-xt}\mathbb{P}(\eta = x)\,\mathrm{d}x} \tag{28}$$

This expression allows for continuous priors (in practice a discrete form would be simulated). Assuming a uniform prior then: $\hat{N}_0 = \mathbb{E}[N_0|\mathcal{D}_0^t]$ and $\mathbb{P}(\eta = \frac{1}{x}|\mathcal{D}_0^t) = \mathbb{P}(N_0 = x|\mathcal{D}_0^t)$ and:

$$\hat{\eta}(t) = \int_0^\infty x\mathbb{P}(\eta = x|\mathcal{D}_0^t)\,\mathrm{d}x = \frac{\int_0^\infty x^{u(t)+1}e^{-xt}\,\mathrm{d}x}{\int_0^\infty x^{u(t)}e^{-xt}\,\mathrm{d}x} \tag{29}$$

$$\hat{N}_0(t) = \int_0^\infty x\mathbb{P}(N_0 = x|\mathcal{D}_0^t)\,\mathrm{d}x = \frac{\int_0^\infty x^{-u(t)+1}e^{-\frac{1}{x}t}\,\mathrm{d}x}{\int_0^\infty x^{-u(t)}e^{-\frac{1}{x}t}\,\mathrm{d}x} \tag{30}$$

An informative comparison can be drawn between the posterior for η in equation 28 and the single tree likelihood expression of equation 19 when $n = 2$. In this case there is only one coalescent event. The final Snyder posterior will then occur at the point when this single coalescent event occurs which will be defined as $t = T$, $u(T) = 1$. The expression of 28 therefore collapses to:

$$\mathbb{P}(\eta|\mathcal{D}_0^T) = \frac{\eta e^{-\eta T}\mathbb{P}(\eta)}{\int_0^\infty \eta e^{-\eta T}\mathbb{P}(\eta)\,\mathrm{d}\eta} \tag{31}$$

The likelihood function gives $\mathbb{P}(\mathcal{D}_0^T|\eta)$. Using this function with event time $\delta_2 = T$ and $n = 2$ gives $L_1(\eta) = \eta e^{-\eta T}$. Applying Bayes theorem:

$$\mathbb{P}(\eta|\mathcal{D}_0^T) = \frac{L_1(\eta)\mathbb{P}(\eta)}{\int_0^\infty L_1(\eta)\mathbb{P}(\eta)\,\mathrm{d}\eta} = \frac{\eta e^{-\eta T}\mathbb{P}(\eta)}{\int_0^\infty \eta e^{-\eta T}\mathbb{P}(\eta)\,\mathrm{d}\eta} \tag{32}$$

The convergence of the results is therefore proven in the trivial single event case.

Unfortunately, similar analytic transformations and expressions like equation 28 are not available in the time varying $N(t)$ case. For these inhomogeneous rate functions it is not easy to scale the exponential parameters with the self-correcting binomial factors as before to remove the count dependent component of the process. This makes the estimation problem more difficult as much of the intensity variation may be lineage dependent instead of due to effective population changes. Consequently, the complete Snyder formulation must be used. Application of the complete Snyder to such demographic functions forms the focus of the next section.

Simulation and Estimation of Standard Phylodynamic Population Models

The power of the multivariate Snyder filter is demonstrated below on several canonical population examples often found in the phylodynamic literature. Simulations are performed according to the following procedure:

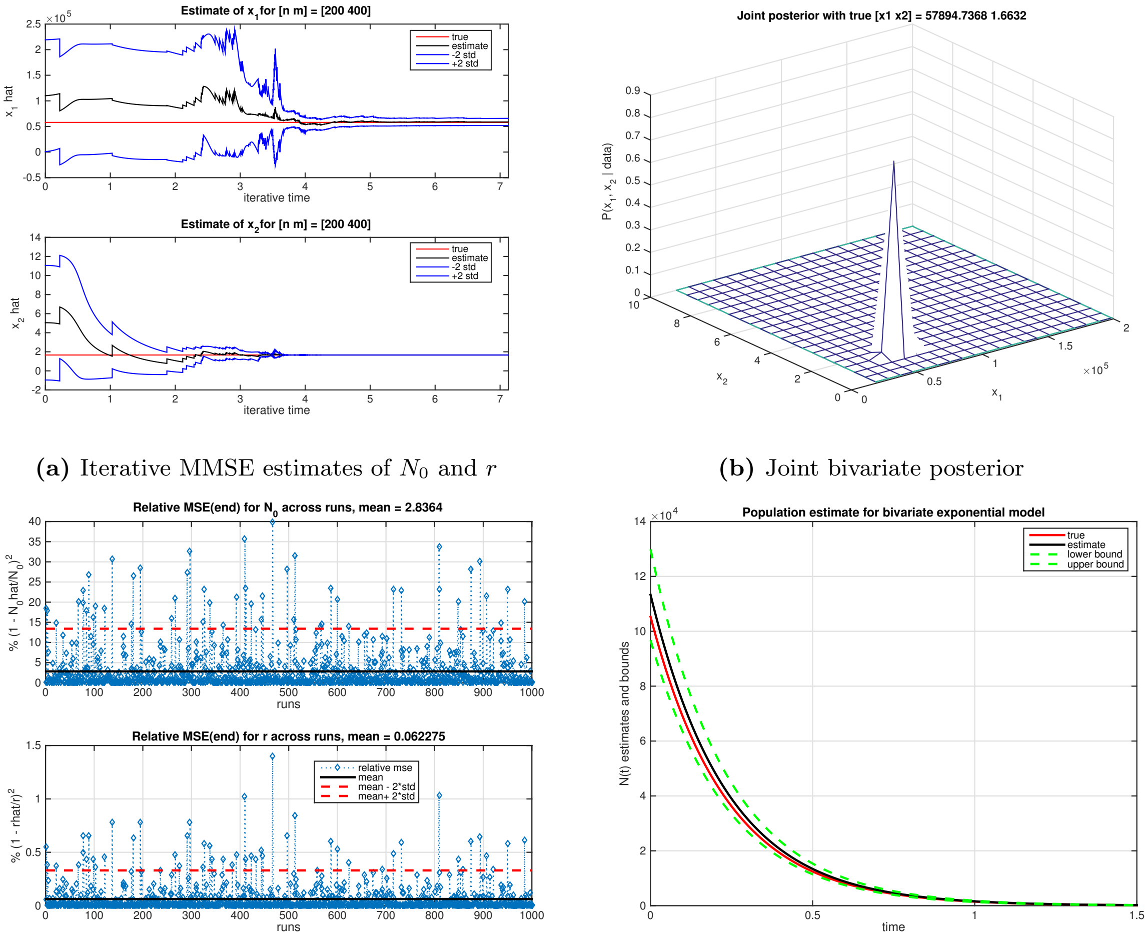
- An l parameter population model is chosen and maxima and minima specified for each parameter.
- Each parameter is discretised into m_i values within its extrema. This defines the filter dimension $m = \prod_i m_i$ and the parameter spaces $\mathcal{X}_i$.
- A true value for each parameter, x_i, is randomly selected from this set and the true demographic model defined as $N\left(t,\, \{x_i\}_{i \in \{1,\, 2,\, \ldots l\}}\right) = N(t,\, \vec{x})$.
- A coalescent stream with $n-1$ coalescent events is simulated by incorporating the true model into the coalescent rate function $\lambda(t)$. This rate is then used to produce a time series according to the appropriate inhomogeneous Poisson process.
- This simulation uses either a time rescaling [8] or standard rejection sampling algorithm. The points are generated one at a time so that the rate can be properly adjusted to account for the self-correcting nature of the process.
- The coalescent stream is then fed into the Snyder filter and the prior iteratively updated into the final joint posterior $q(T)$ where T is such that $u(T) = n - 1$.
- MMSE estimates of the parameters $\hat{x}_i$ and the demographic history $\hat{N}(t)$ are obtained by either marginalising or evolving the posterior through the appropriate function and calculating the appropriate conditional mean.

The population models and specific details related to their importance, simulation and estimation are described subsequently.

Exponential Growth Model: $N(t) = N_0 e^{-rt}$ [22] with $x_1 = N_0$ and $x_2 = r$ set for notational consistency with the literature. This function is often used to describe explosive epidemic growth in forward time. The relevant inhomogeneous self-correcting Poisson process is generated using the time-rescaling algorithm [8] which when solved for this $N(t)$ gives:

$$\int_{t_{k+1}}^{t_k} \binom{k}{2} \frac{1}{N_0} e^{rs}\, \mathrm{d}s = z \implies t_k = \frac{1}{r} \log\left(e^{r t_{k+1}} + \frac{r z N_0}{\binom{k}{2}} \right),\; z \sim \exp(1) \tag{33}$$

Using a uniform joint prior, simulations are performed at $N_{0\text{min}} = 100n$, $N_{0\text{max}} = 1000n$, $r_{\text{min}} = 0.1$, $r_{\text{max}} = 10$, $[m_1,\, m_2] = [20,\, 20] \implies m = 400$, $n = 200$ and $M = 1000$. The resulting estimates over 1000 and a single run are given in panels 4c and 4d. These types of plot will form the standard description for the subsequent mutivariate models. Figures 4a and 4b illustrate the filter convergence to the conditional mean estimates of the parameters, and the final joint posterior. Analogous illustrations are not given for the higher dimensional multivariate models that follow.



(a) Iterative MMSE estimates of N_0 and r

(b) Joint bivariate posterior

(c) Square errors across runs with mean J_r

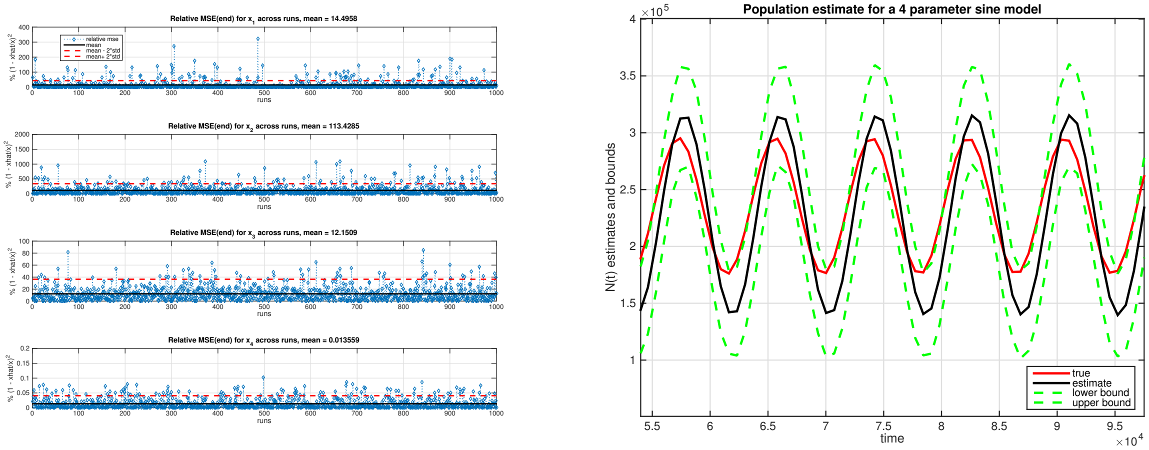
(d) Bivariate exponential reconstruction

Figure 4. Snyder estimate of exponential growth model. Panel 4a compares the mean of the marginal parameter posteriors, from the filter as it iterates across the coalescent time, to the true values. The final value is the optimal estimate. Sub-figure 4b plots the joint final posterior for the parameters. Panels 4c and 4d respectively present the relative square errors across 1000 runs and an illustrative reconstruction from a single run for the demographic function.

Sinusoidally Oscillating Population: this function models a seasonal population with $N(t) = x_1 \sin(x_2 t + x_3) + x_4$. Test data was generated under a rejection sampling algorithm applied to the appropriate inhomogeneous Poisson process with acceptance probability:

$$N(t) \geq x_4 - x_1 \implies \lambda_{\max} \leq \binom{n}{2} \frac{1}{x_4 - x_1} \implies p_{(k,\,t)} = \frac{\binom{k}{2}\frac{1}{N(t)}}{\lambda_{\max}} \tag{34}$$

Simulations were done at $[m_i,\, m,\, n,\, M] = [10,\, 10^4,\, 200,\, 1000]$ with $\max(\vec{x}) = [1000n,\, 10,\, \frac{\pi}{2},\, 1200n]$ and $\min(\vec{x}) = [100n,\, 0.1,\, 0,\, 1100n]$. Single parameter performance across 1000 runs and overally demographic reconstruction for a single run are given in figures 5a and 5b respectively.



(a) Square errors across runs with mean $J_r(x_i)$

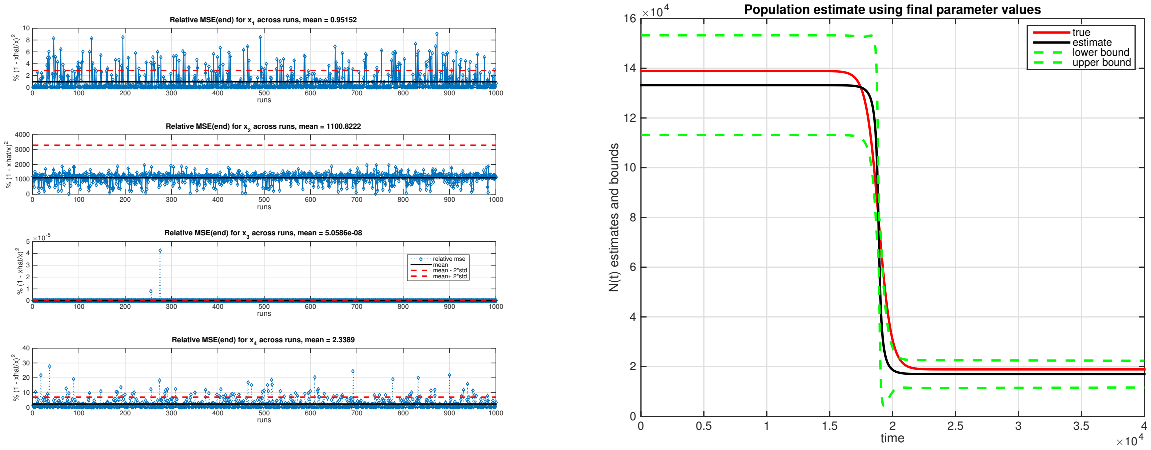
(b) Four parameter sinusoidal reconstruction

Figure 5. Snyder estimation for a multivariate sinusoidally varying population. Subplot 5a gives the relative square errors for each parameter across 1000 runs while panel 5b illustrates a reconstruction from a single run of the entire demographic history.

Logistic Growth Population Model: this function is based on the sigmoidal solution of the SIS transmission equations, derived in [23] and originally developed in [24] as $N(t) = N_0 \frac{1+e^{-rt_{50}}}{1+e^{-r(t_{50}-t)}}$. To prevent $N(t) = 0$ occurring, an offset parameter is added so that the function used is: $N(t) = x_1 \frac{1+e^{-x_2 x_3}}{1+e^{-x_2(x_3-t)}} + x_4$ with $x_1 = N_0$ as the expanding population component (decays in reverse time), $x_2 = r$ as the the exponential rate, $x_3 = t_{50}$ as the time for half decay and x_4 as the background offset. With this setup $x_4 \leq N(t) \leq x_4 + x_1$. The probability of acceptance in the rejection algorithms is:

$$\lambda_{\max} \leq \binom{n}{2} \frac{1}{x_4} \implies p_{(k,\,t)} = \frac{\binom{k}{2}\frac{1}{N(t)}}{\lambda_{\max}} \geq \frac{k(k-1)}{n(n-1)} \frac{x_4}{x_1 + x_4} \tag{35}$$

Simulations were done at $[m_i,\, m,\, n,\, M] = [20,\, 20^4,\, 200,\, 1000]$ with $\max(\vec{x}) = [1000n,\, 10,\, 5000,\, 100n]$ and $\min(\vec{x}) = [100n,\, 0.1,\, 1000,\, 50n]$. For the batch simulation $m_i = 10$ was used. Parameter estimation performance is shown in figure 6a and demographic estimation in figure 6b.



(a) Square errors across runs with mean $J_r(x_i)$

(b) Four parameter logistic reconstruction

Figure 6. Snyder estimation of a multivariate logistic demographic function. Panels 6a and 6b show batch relative square error estimation performance for each parameter across 1000 runs and an example single run reconstruction of the full population function.

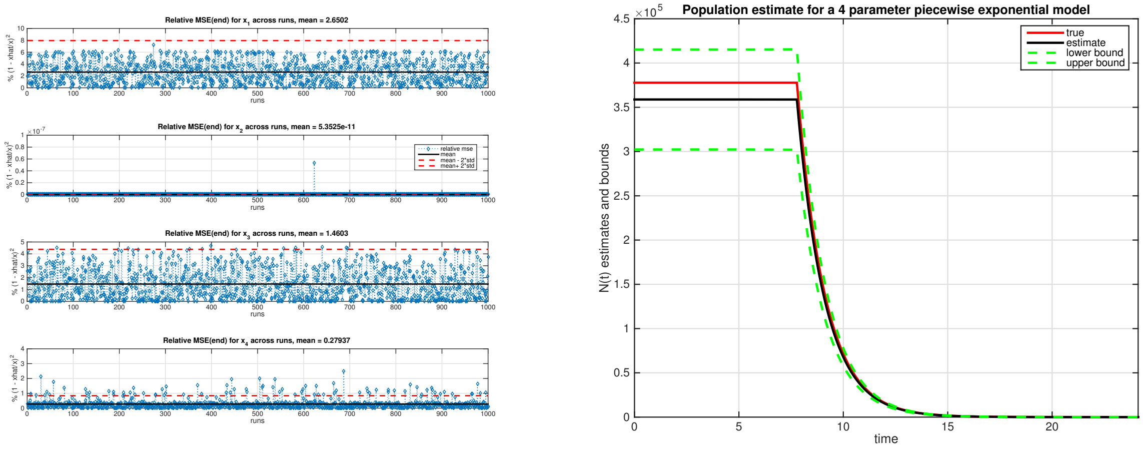
Constant-Exponential-Constant model: related to the logistic function, this population model focuses on the times at which the population rapidly changes from constant to exponential growth and then back to constant (in forward time). It was first applied to the coalescent by Pybus *et al* [6] for Egyptian HCV, which will be described more in the following section on empirical application. The population equation and the acceptance probability are given below. The indicator function $\mathbb{I}(A) = 1$ only if condition A is true.

$$N(t) = x_1 \mathbb{I}(t \leq x_3) + x_1 e^{-x_2(t-x_3)} \mathbb{I}(x_3 < t < x_4) + x_1 e^{-x_2(x_4-x_3)} \mathbb{I}(t \geq x_4) \tag{36}$$

$$x_1 e^{-x_2(x_4-x_3)} \leq N(t) \leq x_1 \implies \lambda_{\max} \leq \binom{n}{2} \frac{1}{x_1 e^{-x_2(x_4-x_3)}} \tag{37}$$

$$p_{(k,\,t)} = \frac{\binom{k}{2}\frac{1}{N(t)}}{\lambda_{\max}} \geq \frac{\binom{k}{2}\frac{1}{x_1}}{\binom{n}{2}\frac{1}{x_1 e^{-x_2(x_4-x_3)}}} = \frac{k(k-1)}{n(n-1)} e^{-x_2(x_4-x_3)} \tag{38}$$

Simulations were performed using $[m_i,\, m,\, n,\, M] = [10,\, 10^4,\, 200,\, 1000]$ with $\max(\vec{x}) = [2000n,\, 0.75,\, 100,\, 150]$ and $\min(\vec{x}) = [1000n,\, 0.1,\, 10,\, 50]$. Figure 7a examines the individual parameter batch performance. The individual run used $\max(\vec{x}) = [2000n,\, 0.1,\, 200,\, 1000]$ and $\min(\vec{x}) = [1000n,\, 0.01,\, 100,\, 500]$ for better illustration. The demographic reconstruction for this case is in panel 7b below.



(a) Square errors across runs with mean $J_r(x_i)$

(b) Constant-exponential-constant reconstruction

Figure 7. Snyder estimation of a multivariate constant-exponential-constant demographic function. Subplots 7a and 7b give the relative square error estimation performance for each parameter across 1000 runs and an illustrative single run reconstruction of the discontinuous population function.

Sub-figures 4d, 5b, 6b and 7b above compare the MMSE estimated demographic functions with their true value across time. These plots are obtained by using the final joint posterior $\mathbb{P}(\vec{x}|\mathcal{F}_T)$. Estimate uncertainty is shown on the plots by heuristic bounds at twice the posterior standard deviation. Panels 4c, 5a, 6a and 7a, however, give the relative MMSE for each parameter in each model across 1000 runs. In some of these batch comparisons, although the shape of the demographic function $N(t)$ is well estimated, some parameters appear to have high relative MMSE values and others very low ones. This could be explained by either or both of two known phenomena:

- Some parameters are more fundamentally difficult to individually estimate. The idea was hinted at by Slatkin *et al* [22]. This can be reasonably expected since a parameter that has a much larger effect on the coalescent rate would likely be more easily estimated than one that contributed relatively negligibly to the observed events.

- Coalescent data is naturally limited in the information it can encode due to inherent Poisson noise. The coalescent process in fact may be thought of as a communication over a Poisson channel. Such channels have known capacities which bound the ability to reconstruct the driving intensities [25].

Estimation of the Hepatitis C (HCV) epidemic in Egypt

Having looked at simulated data, this section now applies the filter to empirical phylodynamic data. The Egyptian HCV dataset is a standard isochronous set that is often used to evaluate coalescent inference techniques. Based on independent epidemiological data it is highly suspected that the high prevalence of HCV in Egypt is due to poor sterilisation during a parenteral antischistosomal therapy (PAT) campaign between 1920 and 1980. Consequently, a good benchmark for phylogenetic analysis is the ability to reproduce this expected rise in infection during the PAT period from genetic data. The data, sampled in 1993, consist of 63 type 4 and 5 subtype 1g sequences of length 411bp. To maintain consistency with [6] the data was separated into set A which contained all 68 samples and B which had only the 63 type 4 cases. The aim is to fit the constant-exponential-constant demographic model, used in [6] and defined below, to these datasets.

$$N(t) = \begin{cases} N_C, & \text{if } t \geq x \\ N_C e^{-r(t-x)}, & \text{if } x < t < y \\ N_A = N_C e^{-r(y-x)}, & \text{if } t \geq y \end{cases}$$

Using the Snyder filter notation, the 4 parameters to be estimated will be defined as $x_1 = N_C$, $x_2 = r$, $x_3 = x$ and $x_4 = y - x$. The demographic model can be written as below with indicator functions $\mathbb{I}(...)$ and $t > 0$ describing time in the past from 1993.

$$N(t) = x_1 \mathbb{I}(t \leq x_3) + x_1 e^{-x_2(t-x_3)} \mathbb{I}(x_3 < t < x_3 + x_4) + x_1 e^{-x_2 x_4} \mathbb{I}(t \geq x_3 + x_4) \tag{39}$$

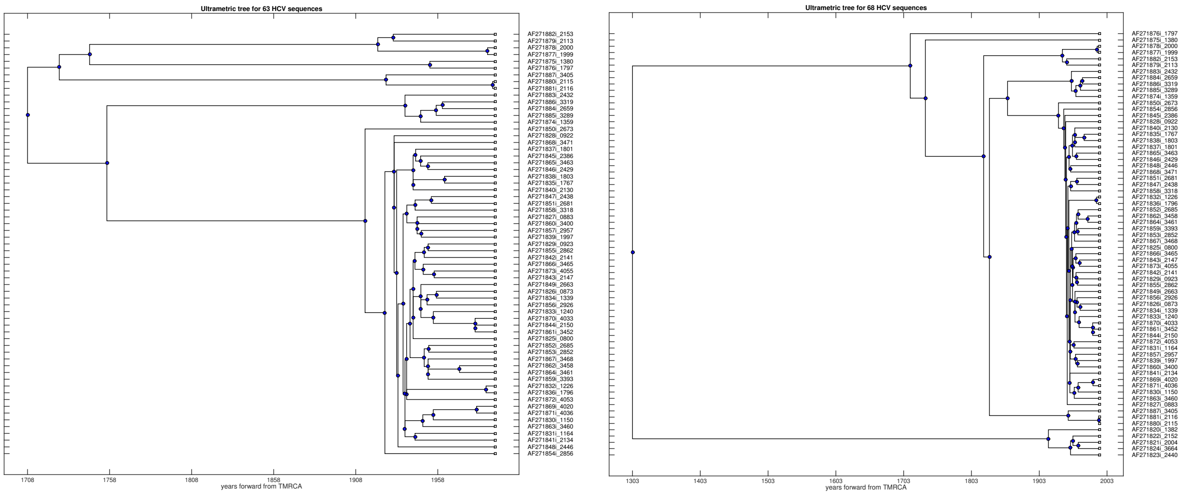
As the Snyder filter requires coalescent times, it is necessary to convert the HCV sequence data into a phylogenetic tree and then, under a molecular clock assumption, obtain a time scaled tree from which event times can be extracted. The software Garli [26] was used to estimate a ML tree for each dataset. This was done under a GTR substitution model with gamma rate heterogeneity, which is consistent with the work of Pybus *et al* [6]. The ML trees were then converted into the ultrametric time trees of figures 8a and 8b. This was done with the software R8s [27] via a Langley-Fitch clock method. The TMRCA of each root was constrained to lie within the same range reported in [6], to ensure sensible ultrametric trees. Further, the clock assumption was set to allow 3 rates (local molecular clocks).

Coalescent times were then extracted from these trees and the Snyder filter run to achieve the curves of figures 9a and 9b. After removing duplicate sequences the coalescent trees were found to possess $n = 54$ and $n = 64$ for the 63 and 68 datasets respectively. The filter simulations were done with $m_i = 20$ and $m = 20^4$ with priors set to match those used in [6] as closely as possible.

Comparison of the estimates with those obtained in [6] are given in table 1. Note that the optimal estimates of the parameters given in table 1 are not used to construct the optimal population reconstruction shown in figures 9a and 9b. Instead the final joint posterior is evolved through the population model to obtain these demographic curves. Mathematically, the difference results because $\mathbb{E}\left[N(t, \vec{x})|\mathcal{F}_t\right] \neq N\left(t, \mathbb{E}\left[\vec{x}|\mathcal{F}_t\right]\right)$. The marginal posteriors for each parameter are shown in figures 9c and 9d. The estimates from both methods are similar. The differences in values are likely due to different optimisations and differences in the coalescent trees used in both studies.

Parameter Estimates of Demographic Function $N(t)$				
Parameter	Pybus 63	Snyder 63	Pybus 68	Snyder 68
N_C	8779 (3323, 15780)	7639 (943, 14334)	10310 (4095, 18960)	8205 (1376, 15034)
r (yr^{-1})	0.237 (0.072, 0.564)	0.2678 (0, 0.5438)	0.264 (0.075, 0.620)	0.2935 (0.076, 0.5084)
x (yr)	1953 (1941, 1966)	1965 (1956, 1974)	1953 (1941, 1966)	1970 (1963, 1978)
y (yr)	1932 (1922, 1940)	1949 (1936, 1962)	1934 (1924, 1943)	1957 (1947, 1966)

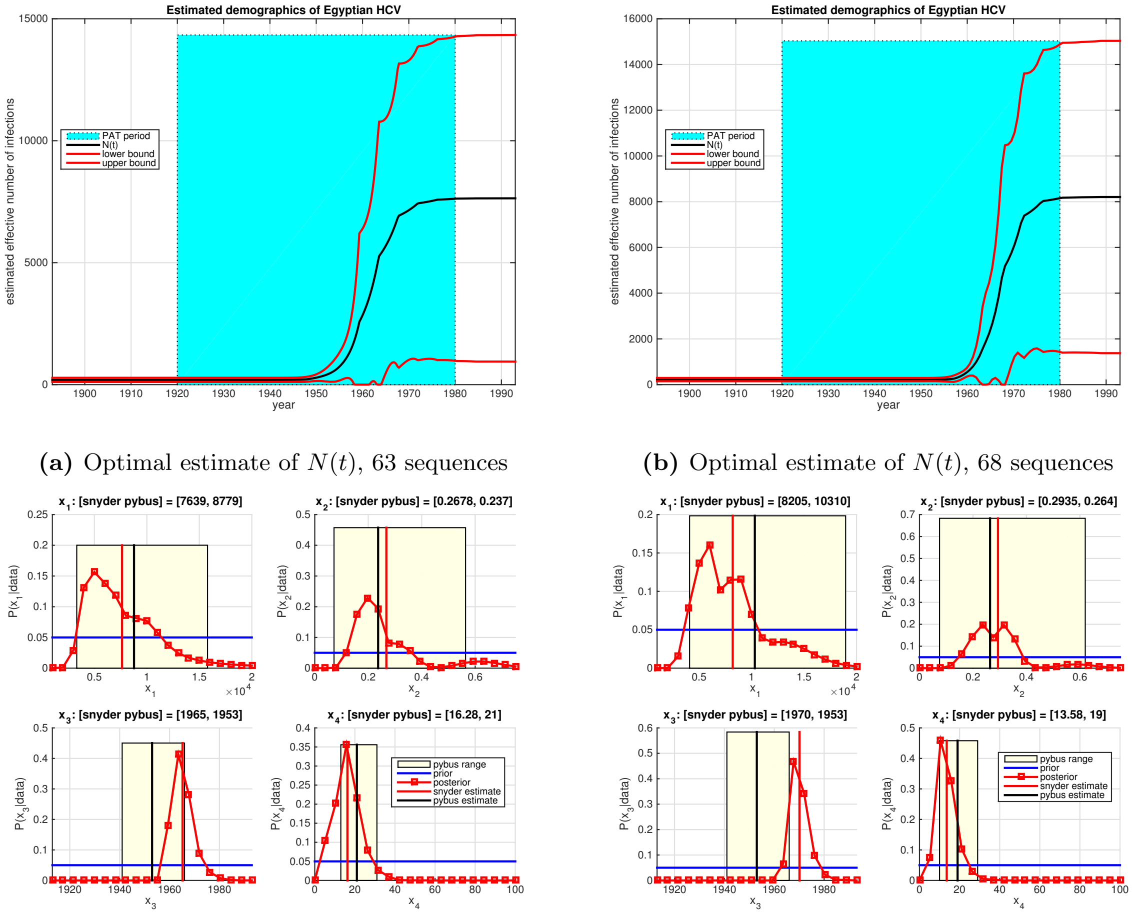
Table 1. Parameter estimates for Egyptian HCV. The Snyder and Pybus estimates are compared across both the 63 and 68 sequence datasets. The conditional means are used by both methods with 95% bounds given in the Pybus scheme and 2 standard deviations provided in the Snyder case.



(a) Time scaled tree for 63 sequences

(b) Time scaled tree for 68 sequences

Figure 8. Ultrametric trees estimated from sequence datasets. Panels 8a and 8b show time scaled trees derived using Garli and R8s on the 63 and 68 sequence datasets respectively. The extra 5 sequences in the latter form an outgroup which appears at the bottom of the tree in panel 8b. This explains why the parameter estimates are similar for both datasets.



(a) Optimal estimate of $N(t)$, 63 sequences

(b) Optimal estimate of $N(t)$, 68 sequences

(c) Posterior comparisons, 63 sequences

(d) Posterior comparisons, 68 sequences

Figure 9. Snyder estimates of the HCV demographic history showing expansion in the PAT period. Panels 9a and 9b are, respectively, optimal reconstructions of the infected population for the 63 and 68 sequence sets. Exponential growth within the PAT period is clear in both cases. Subplots 9c and 9c compare the posteriors derived from the filter for each parameter with Pybus' estimates on both datasets.

Discussion

This manuscript shows that the Snyder filter is a viable technique for coalescent parameter estimation under a range of demographic models. The filter was initially introduced and formulated in full complexity, in the methods section, for a multivariate time varying population. It was then applied to the original Kingman coalescent and found capable of optimal estimation even when the data is provided in reverse order. Additionally, it was proven that, for this constant population size case, there is an information equivalence between coalescent data across one tree and any given coalescent inter-event time from several independent trees (multiple loci). This suggested a transformation which resulted in the constant size self-correcting coalescent inference problem converging to that for a simple homogeneous Poisson process. This allowed an analytical optimal solution for the constant population model to be presented.

The filter was then applied to non-trivial multivariate non-homogeneous self-correcting coalescent processes. It was found to achieve good estimates of the demographic history for several simulated canonical phylodynamic population models. The filter performance was then tested on empirical Hepatitis C data from Egypt. Not only did it reproduce the expected infection expansion during the PAT period, but it also achieved estimates that compared well with those from Pybus *et al* [6].

Thus, the Snyder filter seems promising as an alternative inference measure for coalescent estimation. Since the filter only involves the solution of a set of linear differential equations, it is easy to implement. Moreover, the filter was originally developed for doubly stochastic Poisson processes [18]. Consequently, it can handle stochastic demographic functions and may even be able to handle the birth-death model approach which is the main competitor to coalescent theory. Thus the flexibility, exactness and robustness of this filter coupled with the initially favourable results found in this manuscript, suggest that the Snyder approach has much potential in the field of phylodynamics.

Future iterations of this work will develop the filter for heterochronously sampled data (for which information is also contained in sampling times), assess its ability to deal with birth-death models and also appraise stochastic demographic performance.

Acknowledgments

This work was supported by the European Research Council under the European Commission Seventh Framework Programme (FP7/2007-2013)/European Research Council grant agreement 614725-PATHPHYLODYN.

References

1. Nordberg M (2001) Handbook of Statistical Genetics: Coalescent Theory. John Wiley and Sons, 197-208 pp.

2. Kingman J (1982) On the Genealogy of Large Populations. Journal of Applied Probability 19: 27-43.

3. Griffiths R, Tavare S (1994) Sampling Theory for Neutral Alleles in a Varying Environment. Phil Trans R Soc B 344: 403-10.

4. Notohara M (1990) The Coalescent and the Genealogical Process in Geographically Structured Population. J Math Biol 29: 59-75.

5. Rodrigo A, Shpaer U, Delwart E, et al. (1990) Coalescent Estimates to HIV-1 Generation Time in vivo. PNAS 96: 2187-91.

6. Pybus O, Drummond A, Nakano T, et al. (2003) The Epidemiology and Iatrogenic Transmission of Hepatitis C Virus in Egypt: A Bayesian Coalescent Approach. Mol Biol Evol 20: 381-7.

7. Rasmussen D, Boni M, Koelle K (31) Reconciling Phylodynamics with Epidemiology: the case of Dengue Virus in Southern Vietnam. Mol Biol Evol 2: 258-71.

8. Palacios J, Minin V (2013) Gaussian Process-Based Bayesian Nonparametric Inference of Population Trajectories from Gene Genealogies. Biometrics 69: 8-18.

9. Wakeley J (2008) Coalescent Theory: An Introduction. Roberts and Company Publishers.

10. Kingman J (2000) Origins of the Coalescent: 1974–1982. Genetics 156: 1461-3.

11. Strimmer K, Pybus O (2001) Exploring the Demographic History of DNA Sequences using the Generalized Skyline Plot. Mol Biol Evol 18: 2298-305.

12. Beerli P (2005) Comparison of Bayesian and Maximum-Likelihood Inference of Population Genetic Parameters. Bioinformatics 22: 341-5.

13. Volz E, Koelle K, Bedford T (2013) Viral Phylodynamics. PLoS Computl Biol 9: e1002947.

14. Kuhner M, Yamato J, Felsenstein J (1995) Estimating Effective Population Size and Mutation Rate from Sequence Data using Metropolis-Hastings Sampling. Genetics 140: 1421-30.

15. Kim J, E M, Racz M, Ross N (2015) Can one Hear the Shape of a Population History? Theoretical Population Biology 100: 26-38.

16. Bobrowski O, Meir R, Eldar Y (2008) Bayesian Filtering in Spiking Neural Networks; Noise, Adaptation and Multisensory Integration. Neural Computation 21: 1277-1320.

17. Parag K, Vinnicombe G (2014) Point Process Noise in Fundamental Molecular Reactions and Invertebrate Vision. Ph.D. thesis, University of Cambridge.

18. Snyder D, Miller M (1991) Random Point Procresses in Time and Space. Springer-Verlag, 2 edition.

19. Rudemo M (1972) Doubly-Stochastic Poisson Processes and Process Control. Advances in Applied Probability 2: 318-338.

20. Felsenstein J (1992) Estimating Effective Population Size from Samples of Sequences: Inefficiency of Pairwise and Segregating Sites as compared to Phylogenetic Estimates. Genet Res 59: 139-47.

21. Snyder D (1972) Filtering and Detection for Doubly Stochastic Poisson Processes. IEEE Transactions on Information Theory 18: 91-102.

22. Slatkin M, Hudson R (1991) Pairwise Comparisons of Mitochondrial DNA Sequences in Stable and Exponentially Growing Populations. Genetics 129: 555-62.

23. Dearlove B, Wilson D (2013) Coalescent Inference for Infectious Disease: Meta-analysis of Hepatitis C. Phil Trans R Soc B 368: 2012031.

24. Pybus O, Rambaut A, Harvey P (2000) An Integrated Framework for the Inference of Viral Population History from Reconstructed Genealogies. Genetics 155: 1429-37.

25. Lestas I, Vinnicombe G, Paulsson J (2010) Fundamental Limits on the Supression of Molecular Fluctuations. Nature 467: 174-8.

26. Zwickl D (2006) Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets under the Maximum Likelihood Criterion. Ph.D. thesis, University of Texas at Austin.

27. Sanderson M (2003) R8s: Inferring Absolute Rates of Molecular Evolution and Divergence Times in the Absence of a Molecular Clock. Bioinformatics 19: 301-2.