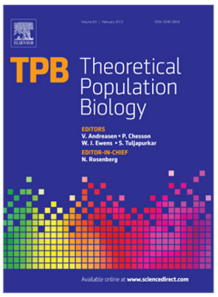
Accepted Manuscript



Exact limits of inference in coalescent models

James E. Johndrow, Julia A. Palacios

PII: S0040-5809(18)30024-8
DOI: https://doi.org/10.1016/j.tpb.2018.11.004
Reference: YTPBI 2665

To appear in: *Theoretical Population Biology*

Received date : 11 February 2018

Please cite this article as: J.E. Johndrow and J.A. Palacios, Exact limits of inference in coalescent models. *Theoretical Population Biology* (2018), https://doi.org/10.1016/j.tpb.2018.11.004

Exact Limits of Inference in Coalescent Models

James E. Johndrow

390 Serra Mall, Stanford, California 94305 USA

Julia A. Palacios

390 Serra Mall, Stanford, California 94305 USA

Abstract

Recovery of population size history from molecular sequence data is an important problem in population genetics. Inference commonly relies on a coalescent model linking the population size history to genealogies. The high computational cost of estimating parameters from these models usually compels researchers to select a subset of the available data or to rely on insufficient summary statistics for statistical inference. We consider the problem of recovering the true population size history from two possible alternatives on the basis of coalescent time data previously considered by Kim et al. (2015). We improve upon previous results by giving exact expressions for the probability of correctly distinguishing between the two hypotheses as a function of the separation between the alternative size histories, the number of individuals, loci, and the sampling times. In more complicated settings we estimate the exact probability of correct recovery by Monte Carlo simulation. Our results give considerably more pessimistic inferential limits than those previously reported. We also extended our analyses to pairwise SMC and SMC' models of recombination. This work is relevant for optimal design when the inference goal is to test scientific hypotheses about population size trajectories in coalescent models with and without recombination.

Keywords: Bayes error rates, coalescent, effective population size, sequentially Markov coalescent.

1. Introduction

Estimation of historical effective population size trajectories from genetic data provides insight into how genetic diversity evolves over time. Availability of molecular sequence data from different organisms living today and from ancient DNA samples together with the development of evolutionary probabilistic models (Wakeley, 2008),

Email addresses: johndrow@stanford.edu (James E. Johndrow), juliapr@stanford.edu (Julia A. Palacios)

has enabled reconstruction of effective population size trajectories of human populations over the past 300,000 years (Gattepaille et al., 2016; Palacios et al., 2015), Ebola virus over the course of the 2014 epidemic in Sierra Leone (Tong et al., 2015) and Egyptian hepatitis C virus for over 100 years (Iles et al., 2014).

Until recently, inference of effective population size trajectories was limited by scarcity of molecular sequence data such as single nucleotide polymorphisms (SNPs) and microsatellites. The increase in the total amount of genetic data obtained at different time points from a large number of individuals over large genomic segments (loci), and the development of more realistic probabilistic models, has led to a situation in which computationally tractable statistical inference is only available from insufficient summary statistics such as the site frequency spectrum (SFS) (Sainudiin et al., 2011), from small numbers of samples, or from short genomic regions (Drummond et al., 2012; Griffiths and Tavaré, 1994; Kuhner et al., 1995; Li and Durbin, 2011; Stephens and Donnelly, 2000). Gao and Keinan (2016) give an extensive list of methods.

Accurate detection of change points in the effective population size trajectory is of scientific relevance in many applications such as the timing and length of the human expansion out-of-Africa (Gao and Keinan, 2016) and extinctions of large mammals at the end of the Pleistocene epoch often attributed to the depredations of humans (Shapiro et al., 2004). Rather than studying the statistical properties of different estimators, we build upon previous work by Kim et al. (2015) and consider how increasing the amount of genetic data increases our ability to distinguish between two alternative hypotheses about population history under different evolutionary models. We evaluate the ability to detect change points by asking what the lowest achievable error rate is for classification between alternative hypotheses about population history with different change points –the Bayes error rate. Calculation of Bayes error rates allows us to answer such questions without considering a particular estimator. We give analytic expressions, and in more complicated settings numerical approximations, of the exact probability that the optimal procedure can identify the truth, given coalescence data. This is a significant difference between the current work and Kim et al. (2015), which emphasized the use of inequalities that, as we show in extensive comparisons, typically give quite optimistic results on the limits of inference in coalescent models. We study the effect of adding more sequences and more loci under the coalescent model with independent loci. We also study the effect of adding more loci in models with recombination for pairwise coalescent data under different demographic scenarios. We also consider the more realistic scenario where the coalescence times are observed with noise, as would result when they are estimated from sequence data, and show how the probability of correct classification is affected by the presence of noise.

2. Coalescent evolutionary models

The standard n-coalescent (Kingman, 1982) is a generative model of molecular sequence of n individuals sampled at random from a population of interest. In the single-locus neutral model, observed variation is the result of a stochastic process of mutations along the branches of the sample's genealogy; the genealogy is a timed bifurcating tree (Figure 1A) that represents the ancestral relationships among samples. When moving back in time, two individuals find a common ancestor (coalesce) in the past with rate inversely proportional to the effective population size $N(t)$. Initially, the standard (homogeneous) n-coalescent assumed constant population size $N(t) = N$ and that sequences were sampled at the same time (t=0). Assuming a global mutation rate μ, the parameter of interest is $\tau = 2N\mu$. The standard neutral coalescent has been extended to variable population size $N(t)$ (Slatkin and Hudson, 1991), varying sampling times (heterochronous coalescent (Felsenstein and Rodrigo, 1999)), and to account for population structure (Beerli and Felsenstein, 2001) and recombination (Griffiths and Marjoram, 1997).

Formally, the coalescent with variable effective population size $N(t)$ (Slatkin and Hudson, 1991) is an inhomogeneous Markov point process of $n-1$ coalescent times denoted by $x_{n-1}, \ldots, x_1$. The process starts with n individuals (lineages) at fixed time $x_n = 0$ until x_{n-1} when two of the n lineages meet their most recent common ancestor. The process continues merging (coalescing) pairs of lineages until time x_1 when the remaining two lineages reach a common ancestor. The resulting realization is a genealogy with $n-1$ coalescent times like the one depicted in Figure 1A. The conditional density of coalescent time x_{k-1} is

$$f(x_{k-1} \mid x_k, N(t)) = \frac{C_k}{N(x_{k-1})} \exp\left\{ -\int_{x_k}^{x_{k-1}} \frac{C_k}{N(t)} dt \right\}$$

where $C_k = \binom{k}{2}$ is the combinatorial factor depending on the number of possible ways that two lineages can coalesce given that there are k lineages, and $N(t)$ is the effective population size, a positive function of time. It follows that the complete likelihood is given by

$$\begin{aligned} L(x_1, \ldots, x_n \mid N(t)) &= f(x_n) \prod_{k=n}^{2} f(x_{k-1} \mid x_k, N(t)) \\ &= \prod_{k=n}^{2} \frac{C_k}{N(x_{k-1})} \exp\left\{ -\int_{x_k}^{x_{k-1}} \frac{C_k}{N(t)} dt \right\} \end{aligned} \tag{2.1}$$

where again $x_n \equiv 0$ by definition.

In the coalescent model with recombination (Griffiths and Marjoram, 1997) looking backwards in time, lineages can either coalesce or recombine at a random position along the chromosome. When a lineage undergoes recombination, the lineage is split

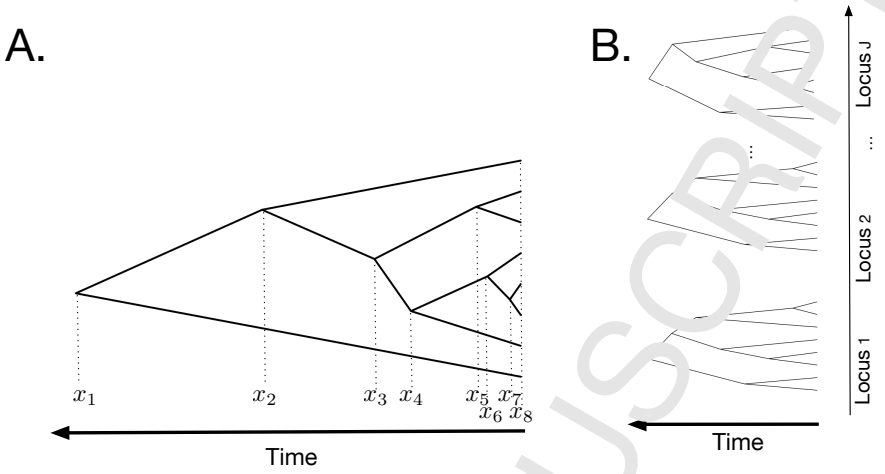


Figure 1: **A)** Genealogy of $n = 8$ sampled individuals. x_i is the time when two of $i + 1$ extant lineages coalesce. **B)** Multiple genealogies along a chromosomal region.

into two. The structure representing coalescent and recombination events is the ancestral selection graph (ARG). In the ARG, different chromosomal segments (loci) can have different genealogies and these genealogies are correlated (Figure 1B). McVean and Cardin (2005) and Marjoram and Wall (2006) introduced Markovian approximations to the ARG called SMC and SMC' respectively. In the SMC, two genealogies at different segments separated by a recombination event are necessarily different, while in the SMC', these two genealogies are not necessarily different. Figure 1B shows an example realization of the SMC or SMC' process. In this paper, we analyze these approximations to the ARG from pairwise coalescent times. Derivation for larger sample sizes involve complicated likelihoods that are beyond the scope of this manuscript.

For $n = 2$, let x^i denote the pairwise coalescent time at the i-th locus; and let J be the number of recombination events. In models with recombination, loci are contiguous chromosomal segments delineated by recombination, so we will also use J to represent the number of completely linked loci. Under the SMC process, the transition density from x^i to x^{i+1}, conditioned on a recombination event at locus $i+1$ is

$$f_{SMC}(x^{i+1} \mid x^i) = \frac{1}{x^i} \int_0^{x^{i+1} \wedge x^i} \frac{1}{N(x^{i+1})} q_1(u, x^{i+1}) du, \tag{2.2}$$

where

$$q_k(a, b) = \exp\left\{ - \int_a^b \frac{k dt}{N(t)} \right\}.$$

Given the current coalescent time x^i, a recombination breakpoint u is sampled uniformly along the height of the tree x^i. At time u, one of the two branches is pruned with equal probability, and a new coalescent time x^{i+1} is drawn with the standard coalescent rate.

Under the SMC' process, the transition density from x^i to x^{i+1}, conditioned on a recombination event at locus $i+1$ is

$$f_{SMC'}(x^{i+1} \mid x^i) = \begin{cases} \frac{1}{x^i}\int_0^{x^i}\int_u^{x^i} \frac{1}{N(t)} q_2(u,t)dtdu & \{x^{i+1} = x^i\} \\ \frac{1}{x^i N(x^{i+1})}\int_0^{x^{i+1}} q_2(u, x^{i+1})du & x^{i+1} < x^i \\ \frac{1}{x^i N(x^{i+1})} q_1(x^i, x^{i+1}) \int_0^{x^i} q_2(u, x^i)du & x^{i+1} > x^i. \end{cases} \tag{2.3}$$

Given the current coalescent time x^i, a recombination breakpoint u is sampled uniformly along the height of the tree x^i. At time u, one of the two branches is selected with equal probability and split into two; one of the emanating branches follows the same trajectory back in time (old branch), while the other emanating branch can coalesce further back in time with any of the remaining two branches in the time interval (u, x^i) with rate $2/N(t)$. Conditional on failing to coalesce with any of the remaining branches in $[u, x^i]$, it coalesces with a branch emanating from the root at rate $1/N(t)$ at some time $x^{i+1} > x^i$. The old branch is then removed. When the new branch coalesces back with the old branch, the resulting tree is the same as the original tree with coalescent time x^i. This event corresponds to the first case in equation 2.3. The likelihood under SMC is given by

$$L(x^1, \ldots, x^J \mid N(t)) = \frac{1}{N(x^1)} \exp\left\{ -\int_0^{x^1} \frac{dt}{N(t)} \right\} \prod_{i=1}^{J-1} f_{SMC}(x^{i+1} \mid x^i),$$

and the likelihood under SMC' for $n = 2$ is obtained from the previous expression replacing f_{SMC} by $f_{SMC'}$.

Given a current coalescent time x^i, the distribution of the length S_i of the current locus under both SMC and SMC' models is exponential with rate ρl_i:

$$f(s_i \mid g_i, \rho) = \rho l_i \exp(-\rho l_i s_i), \tag{2.4}$$

where ρ is the recombination rate per site per generation and l_i is the total tree length, that is $l_i = 2x^i$, the sum of the two branch lengths when $n = 2$. Since expression 2.4 does not depend on $N(t)$, it does not factor in the likelihood. However, it is important to note that older coalescent times will occur at shorter completely linked loci.

3. Bayes error rates in the standard coalescent

We will start with the simple hypothesis testing setting considered previously by Kim et al. (2015) in which the two hypotheses are:

$$H_1 : N(t) = aN_0, \quad T \le t \le T + S \tag{3.1}$$

$$H_2 : N(t) = bN_0, \quad T \leq t \leq T + S$$

with $N(t)$ equal under H_1 and H_2 outside the interval $[T, T + S]$; a, b and S are positive constants and $T \geq 0$ (Figure 5). Our goal is to determine which hypothesis represents the true state of nature under which the data were generated. For simplicity of notation, we associate the state of nature with a parameter $\vartheta \in \{1, 2\}$ such that $H_1 : \vartheta = 1$ and $H_2 : \vartheta = 2$. A (binary) Bayes classifier or decision rule $\vartheta(x)$ has the form

$$\vartheta(x) = \begin{cases} 1 & \mathrm{BF}_{12}(x) > 1 \\ 2 & \mathrm{BF}_{12}(x) < 1 \\ \xi & \mathrm{BF}_{12}(x) = 1 \end{cases}$$

where $\mathrm{BF}_{12}(x)$ is the Bayes factor for H_1 vs H_2, x is an observation of a random variable X, and $\xi \sim \text{Bernoulli}(1/2) + 1$. In the sequel, we drop the explicit argument and simply write BF_{12} in place of $\mathrm{BF}_{12}(x)$. Thus, if $\vartheta(x)$ returns 1, we infer that the data were generated under H_1, whereas if $\vartheta(x)$ returns 2, we infer that the data were generated under H_2. In the case where each hypothesis is assigned prior probability of one half, the Bayes factor is exactly the likelihood ratio, and the probability of selecting H_1 is the probability that $\mathrm{BF}_{12} > 1$ plus half the probability that $\mathrm{BF}_{12} = 1$. In this case, the probability of correct classification is

$$\begin{aligned} \mathbf{P}[\vartheta(X) = \vartheta] = \; & \tfrac{1}{2}\left[\mathbf{P}(\log \mathrm{BF}_{12} > 0 \mid H_1) + \tfrac{1}{2}\mathbf{P}(\log \mathrm{BF}_{12} = 0 \mid H_1)\right] \\ & + \tfrac{1}{2}\left[\mathbf{P}(\log \mathrm{BF}_{12} < 0 \mid H_2) + \tfrac{1}{2}\mathbf{P}(\log \mathrm{BF}_{12} = 0 \mid H_2)\right] \end{aligned}$$

When the prior is correct, the Bayes classifier is the optimal classifier, so that the probability of correct classification using the Bayes classifier is the maximum achievable probability. The Bayes error rate is $1 - \mathbf{P}[\vartheta(X) = \vartheta]$. As such, by studying the properties of the Bayes classifier, we obtain general limitations on inference for any classifier or test.

We first define some notation. Let $X = (X_1, X_2, \ldots, X_{n-1})$ be the random vector of coalescent times with distribution given by (2.1). When multiple genealogies are available, we will denote the random variable corresponding to the collection of all J genealogies by X^J. Throughout, we abuse notation by writing $\mathbf{P}[\vartheta(X) = \vartheta]$ – or $\mathbf{P}[\vartheta(X^J) = \vartheta]$ when $J > 1$ – to represent the probability of correctly identifying the true state of nature.

The following theorems provide exact expressions for the probability of distinguishing between two hypotheses of the form 3.1 from pairwise coalescent data under the coalescent with variable population size (2.1).

Theorem 3.1. *Consider the simple hypothesis testing problem of the form 3.1 when a single pairwise coalescent time is observed ($n = 2$) and assign equal prior probabilities*

to both hypotheses. Then the probability of correctly distinguishing between the two hypotheses is:

$$\mathbf{P}[\vartheta(X) = \vartheta] = \frac{1}{2} + \frac{1}{2}e^{-\Lambda(T)}\left(e^{-\frac{\delta \wedge S}{(a \vee b)N_0}} - e^{-\frac{\delta \wedge S}{(a \wedge b)N_0}}\right)$$

where

$$\delta \equiv \frac{abN_0}{b-a}\log\frac{b}{a} = \frac{abN_0}{a-b}\log\frac{a}{b} \geq 0.$$

and

$$\Lambda(T) \equiv \int_0^T \frac{1}{N(t)}dt.$$

Proofs of all results can be found in the Appendix. Type I and type II error rates can be obtained from the conditional probability expressions derived in the proof, and by modifying our proof to consider a classifier that thresholds $\mathrm{BF}_{12}(x)$ at $\zeta(\alpha)$ for which $\mathbf{P}[\mathrm{BF}_{12}(X) > \zeta(\alpha) \mid H_1] = 1 - \alpha$, one can perform power calculations for testing at level α where H_1 is designated as the null. We mention this as an obvious extension of our results, but do not pursue power calculations in the current work.

Theorem 3.1 and most of the forthcoming results assume that coalescence times are observed directly. In this sense, the results are optimistic regarding the inferential limitations for recovery of historical effective population sizes. In applications, one observes genomic variation from which the coalescence times must be inferred. We can obtain some insight into the implications of this additional estimation step by assuming that coalescence times are observed with noise, so that we observe $X + \epsilon$ rather than X. To avoid settings in which the likelihood is not defined, we assume that $\epsilon > 0$. The following result shows that for small values of ϵ, the probability of correct recovery is equal to the probability of correct recovery in the noiseless case plus a perturbation that is linear in ϵ. We also give an expression for the perturbation in the case where ϵ is exponentially distributed additive noise.

Theorem 3.2. *Consider the simple hypothesis testing problem of the form (3.1) when a noisy version of a single pairwise coalescent time is observed. Suppose $Y = X + \epsilon$ and Bayes factors are computed using y in lieu of x. Without loss of generality, take $a > b$. Then if $0 < \epsilon < (S \wedge \delta)$*

$$\mathbf{P}[\vartheta(Y) = \vartheta] = \mathbf{P}[\vartheta(X) = \vartheta] + \frac{\epsilon}{2}e^{-\Lambda(T)}\left(e^{-\frac{(S\wedge\delta)}{aN_0}}\frac{1}{aN_0} - e^{-\frac{(S\wedge\delta)}{bN_0}}\frac{1}{bN_0}\right) + \mathcal{O}(\epsilon^2).$$

Further, if $\epsilon \sim$ Exponential(λ) and is independent of X we obtain

$$\mathbf{P}[\vartheta(Y) = \vartheta] = \mathbf{P}[\vartheta(X) = \vartheta] + \frac{1}{2}e^{-\Lambda(T)}\left(\frac{1 - aN_0\lambda e^{\frac{S\wedge\delta}{aN_0} - (S\wedge\delta)\lambda}}{aN_0\lambda - 1}e^{-\frac{(S\wedge\delta)}{aN_0}} - \frac{1 - bN_0\lambda e^{\frac{S\wedge\delta}{bN_0} - (S\wedge\delta)\lambda}}{bN_0\lambda - 1}e^{-\frac{(S\wedge\delta)}{bN_0}}\right),$$

which converges to zero at a linear rate in $\mathbf{E}[\epsilon] = \lambda^{-1}$.

We now extend Theorem 3.1 to the case when J independent genealogies from (2.1) are available. When multiple loci or chromosomal segments are either coming from different chromosomes or from the same chromosome at distant locations, genealogies at those locations can be assumed to be independent. When $n = 2$ and J independent genealogies with likelihood (2.1) are available, the sample configuration $L = (L_1, L_2, L_3)$ of the $J = L_1 + L_2 + L_3$ pairwise coalescent times is $L \sim \text{Multinomial}(J, \mathbf{p} = (p_1, p_2, p_3))$, where L_1 is the number of pairwise coalescent times that fall in the interval $(0, T)$, L_2 is the number of pairwise coalescent times that fall in the interval $[T, T + S]$, L_3 is the number of pairwise coalescent times that are greater than $T + S$, and

$$\begin{aligned} p_1 &= \mathbf{P}[X \le T] = 1 - e^{-\Lambda(T)} \\ p_2 &= \mathbf{P}[T < X \le T + S] = e^{-\Lambda(T)} - e^{-\Lambda(T+S)} \\ p_3 &= \mathbf{P}[X > T + S] = e^{-\Lambda(T+S)} \end{aligned}$$

For this setting we have the following result

Theorem 3.3. *Consider the simple hypothesis testing problem of the form (3.1) when J independent pairwise coalescent times are observed ($n = 2$, $J \ge 1$). The probability of correctly distinguishing between the two hypotheses is*

$$\begin{aligned} \mathbf{P}[\vartheta(X^J) = \vartheta] = \frac{1}{2}\mathbf{P}(L_2 = 0 \mid H_1) + \frac{1}{2} \sum_{(\ell_2,\ell_3):\ell_2>0} \mathbf{P}(L = \ell \mid H_1)\mathbf{P}[W^*(\ell_2) > \ell_2\delta - \ell_3 S \mid H_1, L = \ell] \\ + \frac{1}{2} \sum_{(\ell_2,\ell_3):\ell_2>0} \mathbf{P}(L = \ell \mid H_2)\mathbf{P}[W^*(\ell_2) < \ell_2\delta - \ell_3 S \mid H_2, L = \ell] \end{aligned}$$

where $W^*(\ell_2) = \sum_{j=1}^{\ell_2} X_*^j$ is the sum of ℓ_2 independent truncated coalescent times $X_*^j \in [0, S]$, each exponentially distributed with rate $(aN_0)^{-1}$ under H_1, and rate $(bN_0)^{-1}$ under H_2; δ is defined as in Theorem 3.1, and

$$\mathcal{L} = \Big\{\ell = (\ell_1, \ell_2, \ell_3) : \ell_j \in \mathbb{N}, \sum_j \ell_j = J\Big\}.$$

is an element of the support of $\text{Multinomial}(J, \mathbf{p} = (p_1, p_2, p_3))$.

To obtain numerical results, we approximate the distribution function $\mathbf{P}[W^*(\ell_2) < t]$ by Monte Carlo. In the next section we apply these results to the problem of distinguishing between two hypotheses about the human expansion.

4. Human expansion

Many of the statistical methods proposed over the last 15 years to infer effective population sizes from genetic data have been applied to human whole genomes (Li

and Durbin, 2011; Schiffels and Durbin, 2014; Sheehan et al., 2013; Palacios et al., 2015; Terhorst et al., 2017). Several studies agree that non-African populations have experienced two severe bottlenecks, one attributed to the expansion out-of-Africa and the other attributed to the separation of Asian and European populations. There is, however, disagreement in the timing and length of such events.

Figure 2A shows a population history compatible with a human population history recovered from autosomal DNA in standard coalescent units (Li and Durbin, 2011). In order to convert coalescent parameters into real time and size, time and $N(t)$ need to be divided by the mutation rate per generation. Time need be further multiplied by the generation time. To make our results comparable to previous studies (Li and Durbin, 2011; Kim et al., 2015), we will assume a generation time of 25 years and that effective population size is expressed in units of 2.732×10^4. That is, one unit in the y-axis of Figure 2A corresponds to 2.732×10^4 and one unit in the x-axis of the same plot corresponds to 68.3×10^4 years. In our analysis, we compare a population trajectory whose second bottleneck starts at time $T = 102.45$kya (0.15 in standard units) versus a population trajectory whose second bottleneck starts earlier at time $T+S$ with S ranging from 30kyr to 150kyr. Our results from theorem 3.3 are depicted in Figure 2B. In order to correctly differentiate between the two hypotheses with $S = 130$kyr with probability of at least 0.95, we need at least 35 loci. A correct classification with probability of at least 0.95 is achievable with at least 50 loci when $S = 60$kyr, that is, when the bottleneck started around 162kya versus 102kya.

Our results differ from previously published bounds based on coalescent Bayes error rates. Kim et al. (2015) indicate that the minimal J such that any classifier can distinguish between H_1 and H_2 with probability at least 0.95 and $S = 130$kyr is at least $J = 10$; while for $S = 60$kyr it is $J \approx 20$. These numbers can be compared directly with our $J = 35$ and $J = 50$, respectively. Thus, our results indicate that 2.5-3.5 times more data are required to make inference feasible in this scenario compared to the results of Kim et al. (2015). A detailed analysis of the differences between our results and previously published bounds of (Kim et al., 2015) – which reflect the fact that we give exact expressions instead of upper bounds on $\mathbf{P}[\vartheta(X^J) = \vartheta]$ – can be found in section 8.

4.1. Value of incorporating ancient samples

Thus far, we have not considered the effect of incorporating samples at different sampling times, and have implicitly assumed that all samples are obtained at present. Results from Theorems 3.1 and 3.3 can be directly applied for the case when the two samples are obtained some time in the past. In particular, we assess the change in $\mathbf{P}[\vartheta(X^J) = \vartheta]$ when samples are obtained at the end of the bottleneck event at 102.45kya and when samples are obtained at 50kya. These scenarios are equivalent to putting $\Lambda(T) = 0$ and $\Lambda(T) = 1.54$, respectively. Ancient DNA (aDNA) corresponding to $T = 50$kya is available from ancient genomes (Fu et al., 2016). Obtaining coalescent data from the population immediately after the end of the event of interest

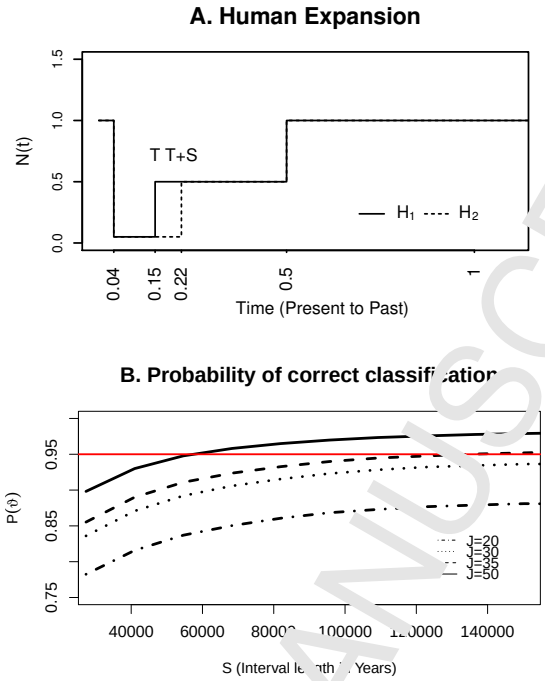


Figure 2: **A.** Human population history in coalescent units compatible with previous findings from whole genomes (Li and Durbin, 2011). One unit in the y-axis of Figure 2A corresponds to 2.732×10^4 and one unit in the x-axis of the same plot corresponds to 68.3×10^4 years. **B.** Probability of correct classification $\mathbf{P}[\vartheta(X^J) = \vartheta]$ as a function of the interval length S in years for several values of J loci corresponding to the two hypotheses depicted in A. Red line indicates probability of 0.95.

is in some sense the optimal strategy for statistical inference on that event, and can have an enormous positive effect on inference. However, it is important to emphasize that DNA degrades over time and inference of coalescent data from ancient samples is more challenging. We do not attempt to quantify this effect, and thus the results given here are in some sense optimistic regarding the benefits of using ancient samples. Figure 3A shows $\mathbf{P}[\vartheta(X^J) = \vartheta]$ for $J = 2, 3, 5, 10, 15$ when coalescent data are available free of error. For all but $J = 2$, $\mathbf{P}[\vartheta(X^J) = \vartheta] \geq 0.95$ can be achieved for S greater than about 115kya. For $J = 15$, it is possible to achieve $\mathbf{P}[\vartheta(X^J) = \vartheta] \geq 0.95$ with S larger than about 15kyr. When the samples are available from 50kya, it is possible to achieve $\mathbf{P}[\vartheta(X^J) = \vartheta] \geq 0.95$ with at least $J = 20$ loci. Thus, neglecting the effect of degradation of ancient samples, their use can significantly reduce the amount of data required to reach reasonable certainty regarding the true population size history.

5. Increasing the number of samples

Now we consider the case where $n > 2$. The following theorem gives an exact expression for the probability of correct classification in (3.1) from a single locus

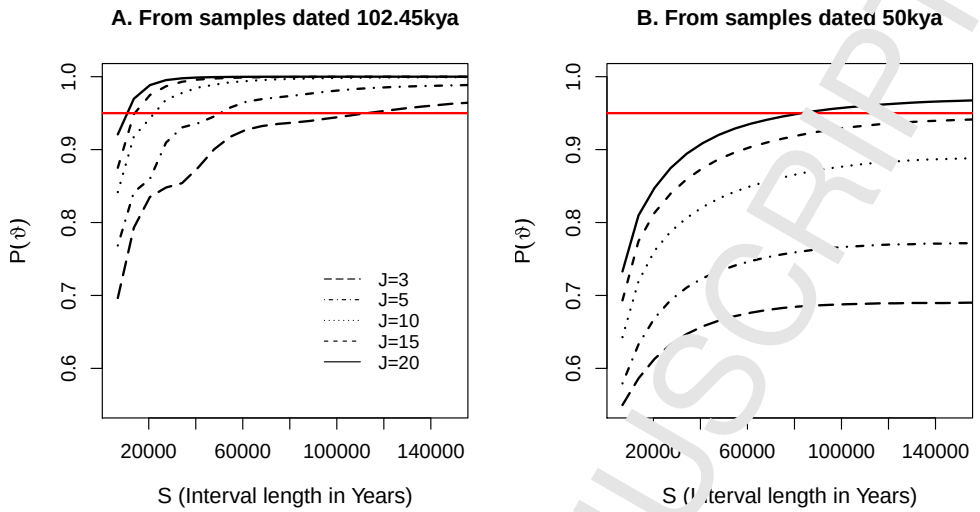


Figure 3: **Value of incorporating ancient samples.** Probability of correct classification $\mathbf{P}[\vartheta(X^J) = \vartheta]$ as a function of S in years for the human bottleneck example when samples are obtained before the bottleneck (A) and when samples are obtained around 50kya (B). Curves for different number of loci (J) is indicated by the line patterns. Red line indicates probability of 0.95.

$(J = 1)$ when $n = 3$.

Theorem 5.1. *Consider the simple hypothesis testing problem of the form 3.1 when a single genealogy of $n = 3$ individuals is observed. Define δ as in theorem 3.1, then the success rate of the optimal classifier is*

$$\mathbf{P}[\vartheta(X^J) = \vartheta] = \frac{1}{2} + \frac{1}{4}e^{-3\Lambda(T)}\xi(a, b, N_0, T, S)$$

where

$$\begin{aligned}\xi(a,b,N_0,T,S) = {} & 3e^{2\Lambda(T)}\left[e^{-\frac{\delta\wedge S}{aN_0}} - e^{-\frac{\delta\wedge S}{bN_0}}\right] \\ & + 3\left[e^{-\frac{2\left(0\vee\left(\frac{\delta-S}{2}\wedge S\right)\right)+S}{aN_0}} - e^{-\frac{2\left(0\vee\left(\frac{\delta-S}{2}\wedge S\right)\right)+S}{bN_0}}\right] - 3\left[e^{-\frac{\delta\wedge S}{aN_0}} - e^{-\frac{\delta\wedge S}{bN_0}}\right] \\ & + \begin{cases} 0 & S < \frac{2}{3}\delta \\ e^{-\frac{2\delta}{bN_0}}\left(1 + \frac{2\delta-3S}{bN_0}\right) - e^{-\frac{2\delta}{aN_0}}\left(1 + \frac{2\delta-3S}{aN_0}\right) & \frac{2}{3}\delta < S < 2\delta \\ e^{-\frac{2\delta}{aN_0}}\left(\frac{4\delta}{aN_0} + 2\right) + e^{-\frac{2\delta}{bN_0}} - 3e^{\frac{T-2\delta-S}{bN_0}} + e^{-\frac{2\delta+S}{bN_0}}\frac{3T-3S-4\delta}{bN_0} & S > 2\delta \end{cases}\end{aligned}$$

The proof is located in Appendix D. We can use this result to assess how much an additional sample helps in identifying the true population size history. Figure 4 shows

four examples of $\mathbf{P}[\vartheta(X) = \vartheta]$ as a function of a while fixing $b = 1$; increasing a is equivalent to increasing the separation between the two hypothetical population size histories. In two of the examples, $N(t) = 1$ outside the interval $[T, T + S]$, and in the other two $N(t) = e^t$ outside this interval. In both cases, the probability of identifying the true effective population size function is considerably higher with $n = 3$ than $n = 2$ when $|a - b|$ is not too close to zero. While the magnitude of increase differs for the two scenarios, additional coalescent times can help considerably to distinguish between alternative histories.

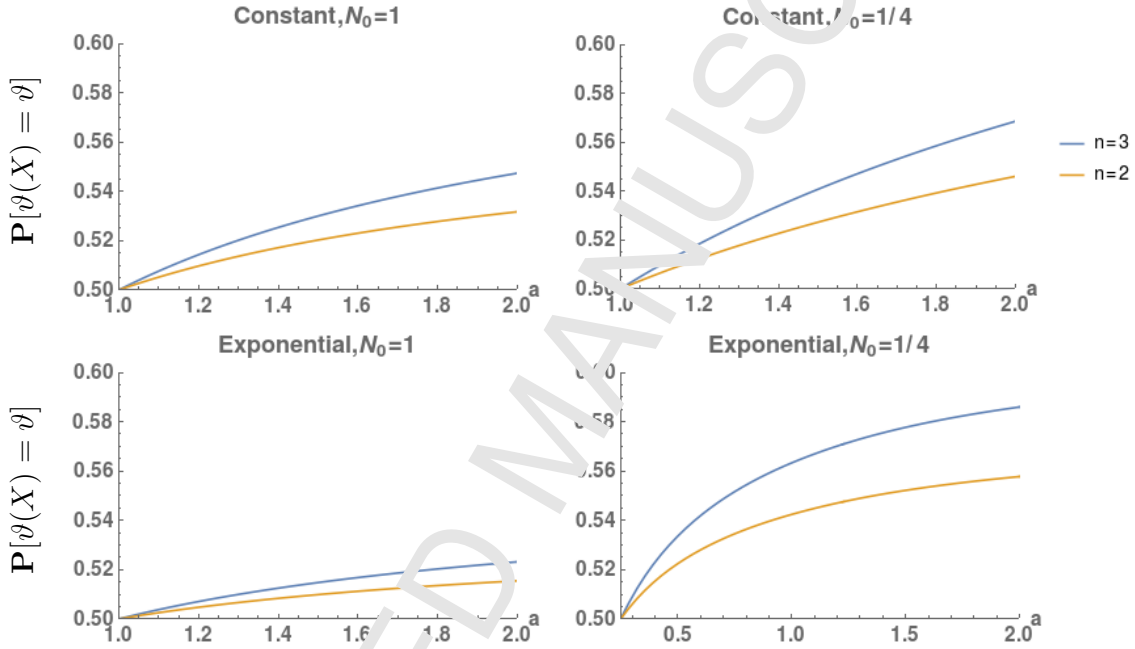


Figure 4: **Effect of adding an additional sample.** Probability of correct classification $\mathbf{P}[\vartheta(X) = \vartheta]$ as a function of a for classification problem 3.1. In each case we put $T = 1, S = 1/2$, and $b = 1$. We compare the effect of adding one more sample ($n = 2$ vs $n = 3$) for constant and exponential growth population trajectories.

It is clear from the proof of Theorem 5.1 that while it is possible to obtain exact expressions for $n > 2$, the number of cases that must be treated will grow exponentially in n. Of course it is still possible to approximate $\mathbf{P}[\vartheta(X) = \vartheta]$ by simulation for arbitrary n. Here, we re-analyze the human expansion classification problem considered in Section 4 for $n = 10$ and $J = 1, 5, 10, 20$ as a function of interval length S and compare to $n = 2$. The value of $\mathbf{P}[\vartheta(X^J) = \vartheta]$ is approximated by 10,000 Monte Carlo samples. Results are shown in Figure 5. In contrast to the case of $n = 2$, where $J = 50$ was required to achieve $\mathbf{P}[\vartheta(X^J) = \vartheta] = 0.95$ for $S = 60$Kyr, when $n = 10$ it is possible to achieve the same success probability with $J = 20$. Thus, increasing the number of contemporaneous sequences or loci gives sharper inference on the duration of the second expansion, but the effect is highly sublinear in n.

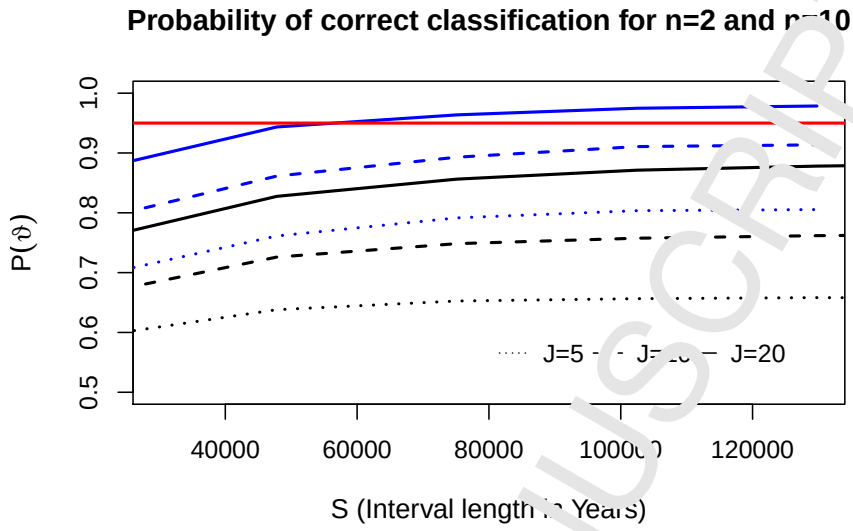


Figure 5: **Effect of adding more samples in the human expansion scenario** Blue lines represent $\mathbf{P}[\vartheta(X) = \vartheta]$ as a function of bottleneck length S for $n = 10$ and black lines represent $\mathbf{P}[\vartheta(X) = \vartheta]$ for $n = 2$. Different number of loci are distinguished by line patterns. Red line indicates probability of 0.95.

6. Other scenarios

We now consider a more general classification problem when pairwise coalescent data is available at a single locus or multiple loci :

$$\begin{aligned} H_1 &: N(t) = N_1(t) \\ H_2 &: N(t) = N_2(t). \end{aligned} \tag{6.1}$$

We consider the case where $N_2(t) = cN_1(t)$ for $c \in (0, 1)$, where analytic expressions for $\mathbf{P}[\vartheta(X) = \vartheta]$ are available even when $N_1(t)$ is not piecewise constant.

Theorem 6.1. *Consider the simple hypothesis testing problem of the form (6.1) such that $N_2(t) = cN_1(t)$ with $0 < c < 1$ when a single pairwise coalescent time is observed ($n = 2$) and assign equal prior probabilities to both hypotheses. Then the probability of correct classification is:*

$$\mathbf{P}[\vartheta(X) = \vartheta] = \frac{1}{2}c^{\frac{c}{1-c}} + \frac{1}{2}\left(1 - c^{\frac{1}{1-c}}\right).$$

Theorem 6.2. *Consider the conditions of Theorem 6.1 for J independent loci. The probability of correct classification is*

$$\mathbf{P}[\vartheta(X^J) = \vartheta] = \frac{1}{2}\left(1 - \frac{1}{\Gamma(J)}\left[\gamma\left(J, \frac{-Jc\log c}{1-c}\right) - \gamma\left(J, \frac{J\log c}{c-1}\right)\right]\right). \tag{6.2}$$

where $\gamma(a, b)$ is the lower incomplete gamma function.

Figure 6 shows (6.2) as a function of c for different values of J. As expected, the larger J, the larger the value of c at which high probability of identifying the true population size history can be achieved. However, even for $J = 100$, we must have $c \approx 0.75$ or smaller to give probability 0.95 of selecting the true population size history. Our results from Theorems 6.1 and 6.2 differ from previously published Bayes error rates bounds (Kim et al., 2015). In section 8, we present a more detailed analysis of the differences between our exact expressions and the bounds (Kim et al., 2015).

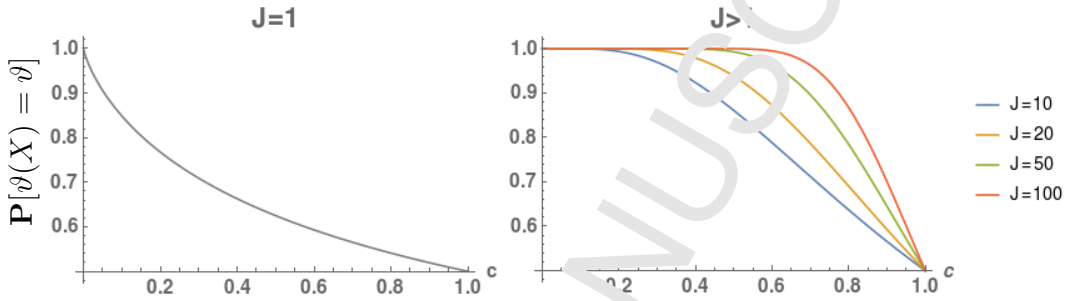


Figure 6: **Left.** Probability of correct classification $\mathbf{P}[\vartheta(X) = \vartheta]$ as in (6.2) when $N_2(t) = cN_1(t)$ and $J = 1$ (Theorem 6.1). **Right.** $\mathbf{P}[\vartheta(X^J) = \vartheta]$ as a function of c for several values of J.

7. Bayes error rates in the sequentially Markov coalescent

We now consider the same classification problem as in (3.1) from $J \geq 2$ consecutive loci along a chromosomal region. We assume the ideal scenario in which we observe the $J \geq 2$ pairwise coalescent times ($n = 2$) corresponding to each of these J loci separated by $J-1$ recombination events. Further, we assume that effective population size trajectories under H_1 and H_2 are piece-wise constant functions over time such as the human expansion scenario in Figure 2A. We then approximate $\mathbf{P}[\vartheta(X^J) = \vartheta]$ by Monte Carlo from 10,000 simulations generated from each hypothesis under the two coalescent models with recombination: SMC (2.2) and SMC' (2.3).

We re-analyze the human expansion classification problem considered in section 4 for $n = 2$ and $J = 2, 5, 10, 20, 30, 35$ as function of interval length S under independent loci (3.3), SMC' (2.3) and SMC (2.2). Our results are depicted in Figure 7. We show that either under SMC' or independent loci, $\mathbf{P}[\vartheta(X^J) = \vartheta] = 0.95$ is achievable with $J = 35$ loci. For $2 \leq J < 20$, the Bayes error rate in SMC is smaller than the other two alternatives. The significance of this is that it is not necessary to have many independently segregating loci to make inference on features of the historical population size. Instead, for the hypotheses considered, the same number of non-independent loci separated by recombination events will suffice. The set of all dependent loci is of course considerably larger than the largest set of independent loci, so the result suggests optimism in the potential to reconstruct features of the population size trajectory in the relatively distant past. We note that in the more general

setting of theorem 6.2, the probability of correct classification from independent loci is higher than from correlated loci under the SMC' model (Figure 9).

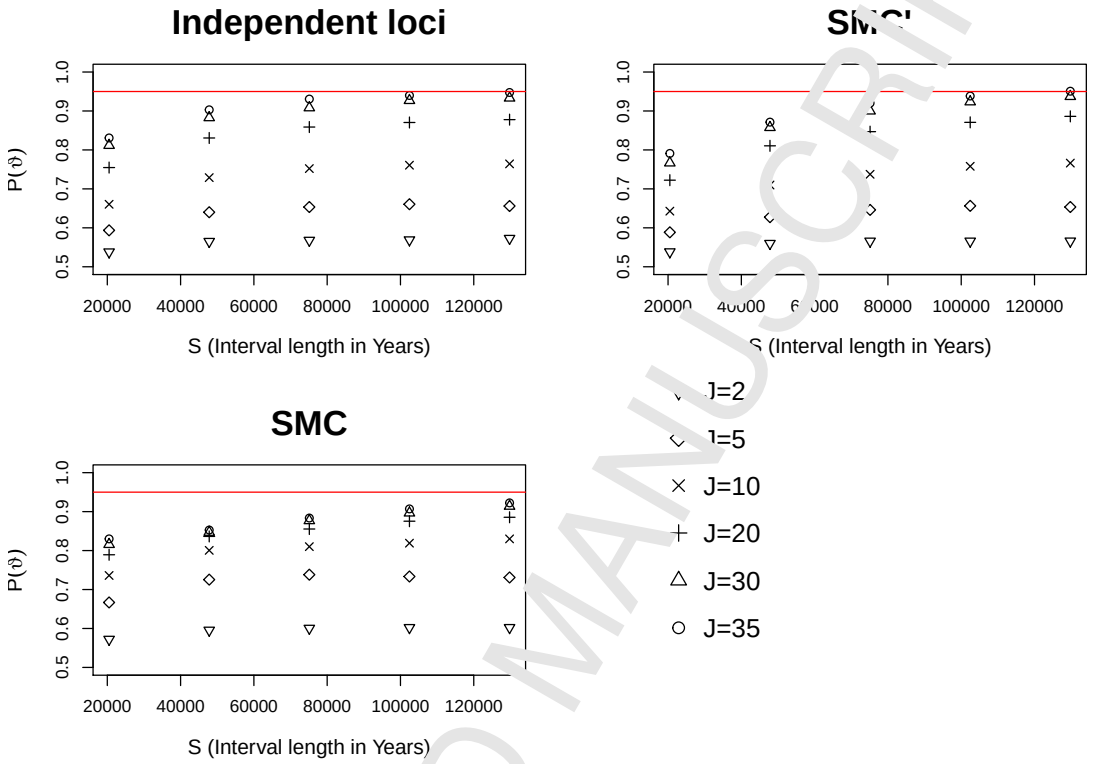


Figure 7: **Sequentially Markov coalescent in the human expansion scenario** Probability of correct classification under independent sampling, SMC' and SMC. Different patterns represent different number of loci. Red line indicates probability of 0.95.

Bayes error rates in the renewal approximation of sequentially Markov coalescent

Carmi et al. (2014) proposed a *renewal approximation* to the sequentially Markov coalescent models SMC and SMC' for a pair of chromosomes, in which the pairwise coalescent times at contiguous segments separated by a recombination event are independent and distributed according to a position-independent stationary distribution. The stationary distribution of the pairwise coalescent time under SMC and SMC' (Carmi et al., 2014) is

$$\pi(x) = \frac{\frac{x}{N(x)} \exp\left\{-\int_0^x \frac{dt}{N(t)}\right\}}{\int_0^\infty \exp\left\{-\int_0^u \frac{dt}{N(t)}\right\} du} = \frac{x}{\mu N(x)} \exp\left\{-\int_0^x \frac{dt}{N(t)}\right\}, \tag{7.1}$$

where μ is the expected pairwise coalescent time under the initial marginal density:

$$\mu = \int_0^\infty \frac{x}{N(x)} \exp\left\{-\int_0^x \frac{dt}{N(t)}\right\} dx.$$

The stationary density $\pi(x)$ may be interpreted as the density of pairwise coalescent time in a randomly chosen chromosomal segment. Convergence to the stationary distribution under SMC and SMC' is achieved after $J = 9$ recombinations when the effective population size is 1. Figure 3 in Carmi et al. (2014) shows the convergence under the SMC model and Figure 8 shows the convergence under the SMC' model with $N_e = 1$ (left) and the population size trajectory for H_2 in the human expansion scenario (right).

The log Bayes factor $\log \mathrm{BF}^{\pi}_{12}$ assuming the stationary distribution $\pi(x)$ of equation 7.1 corresponds to the Bayes factor under independent loci (BF_{12}) plus an extra constant that is a function of the expected pairwise coalescent times under both hypotheses. That is

$$\log \mathrm{BF}^{\pi}_{12} = \log \mathrm{BF}_{12} + \log(\mu_2/\mu_1), \tag{7.2}$$

where μ_i is the expected pairwise coalescent time under H_i. For the particular case when $N_2(t) = cN_1$ and $N_1(t) = N_1$, $\mu_2 = c\mu_1$ and $\log \mathrm{BF}^{\pi}_{12} = \log \mathrm{BF}_{12} + \log(c)$. The following theorems state the probability of correct classification from pairwise coalescent data under the renewal approximation.

Theorem 7.1. *Consider the simple hypothesis testing problem of the form (6.1) such that $N_1(t) = N_1$ and $N_2(t) = cN_1$ with $0 < c < 1$ when a single pairwise coalescent time is observed ($n = 2$) from the stationary distribution $\pi(x)$ of equation 7.1 and assign equal prior probabilities to both hypotheses. Then the probability of correct classification is:*

$$\mathbf{P}[\vartheta(X) = \vartheta] = \frac{1}{2} c^{\frac{2c}{1-c}} + \frac{1}{2}\left(1 - c^{\frac{2}{1-c}}\right) - \log(c) c^{\frac{1+c}{1-c}}.$$

Theorem 7.2. *Consider the conditions of Theorem 7.1 for J independent loci. The probability of correct classification is*

$$\mathbf{P}_\pi[\vartheta(X^J) = \vartheta] = \frac{1}{2}\left(1 - \frac{\gamma\left(2J, 2J\frac{c}{1-c}\log\frac{1}{c}\right) + \gamma\left(2J, 2J\frac{1}{c-1}\log c\right)}{\Gamma(2J)}\right). \tag{7.3}$$

where $\gamma(a, b)$ is the lower incomplete gamma function.

Comparing our results from theorems 6.2 and 7.2, the probability of correct classification under the renewal approximation corresponds to the probability of correct classification under independent loci from twice the number of loci. We find this result counterintuitive, however, the renewal approximation, SMC' and SMC are all approximations to the ancestral recombination graph. The renewal approximation

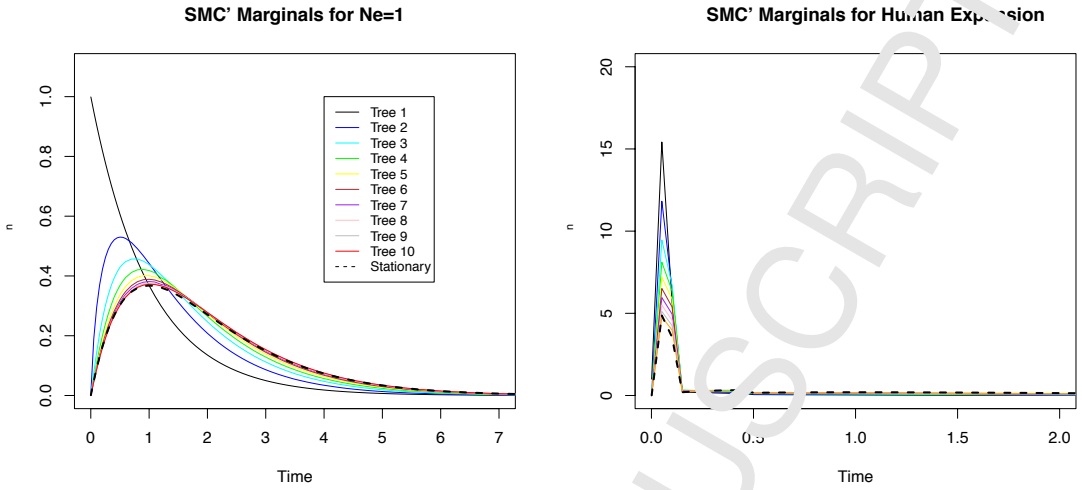


Figure 8: **Convergence to stationarity.** Marginal density of the pairwise coalescent time under the SMC' model after $J = 0, 1, \ldots, 9$ recombination events when the population size is $N_e = 1$ (left) and when the population size trajectory corresponds to H_1 in Figure 2. In both cases, the marginal distribution after 9 recombinations (Tree 10) converges to the stationary distribution.

is another model of recombination and not necessarily a good approximation to the SMC or SMC'. Figure 9 shows a comparison of the probability of correct classification for different values of c under the SMC' model, independent loci model and the renewal approximation. In the three cases: SMC', independent loci and renewal approximation, the probability of correct classification increases logarithmically with the number of loci.

We note that previous analysis is valid when $n = 2$. If we assume a constant recombination rate per site per generation of $\rho = 2 \times 10^{-9}$, then when $n = 2$ and $N(t) = N_0 = 2.732 \times 10^4$, the expected length of a locus is $1/\rho l = 1/(2 \times 10^{-9} \times 2 \times 2.732 \times 10^4) \approx 9,151$ base pairs. In a genome of 3 billion base pairs, we would expect around 330,000 completely linked loci. Increasing the number of samples increases the tree length and therefore the overall rate of recombination increases (Eq. 2.4) and the expected number of completely linked loci increases for the same chromosomal region. In addition, the number of loci required to distinguish between two hypotheses greatly depends on the hypotheses considered. The number of independent loci required to obtain a probability of 0.95 when the two hypotheses are similar to each other (c=.9 in Figure 9) is of the order of 980 loci (number obtained by direct application of Theorem 5.2) and for c=.95, the number of independent loci required is around 4,150. We want to point out that these numbers are very conservative since we are assuming that coalescent data are available and therefore our results should be interpreted as upper bounds on the achievable probability of recovering the truth.

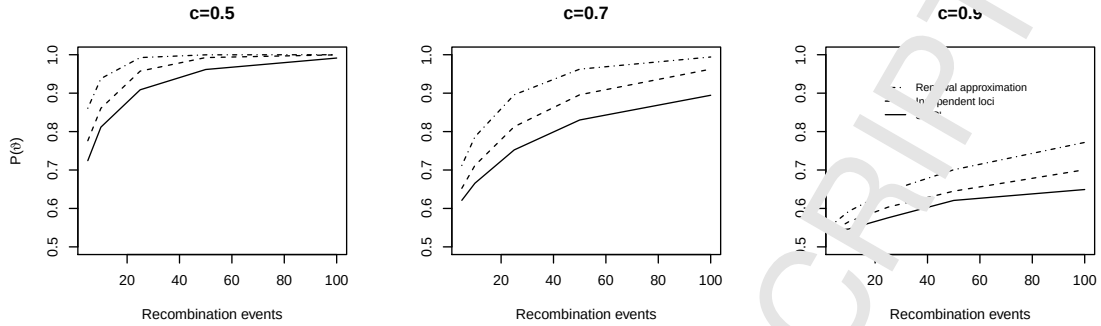


Figure 9: **Comparison between SMC', independent loci and renewal approximation from pairwise coalescent data.** Probability of correct classification in the setting of theorems 6.2 and 7.2 for three values of c of 0.5, 0.7 and 0.9. The renewal approximation achieves the greatest probability of correct classification.

8. Comparison to results of Kim et al. (2015)

Kim et al. (2015) provided lower bounds on Bayes error rates from pairwise coalescent data. We now provide a comparison of some of our results to these previously published bounds. In this section, we will let Y denote a random coalescence time generated under H_1 and Z denote a random coalescence time generated under H_2. Assuming a classification problem of the form (3.1) and prior probability 1/2 on H_1 and H_2, the Bayes error rate for any classifier is at least $(1 - \Upsilon)/2$ where

$$\Upsilon = \mathrm{d_{TV}}(Y, Z) = \mathrm{d_{TV}}(\mu, \nu) \equiv \sup_A |\mu(A) - \nu(A)|$$

is the total variation distance between probability measures μ, ν, such that $Y \sim \mu, Z \sim \nu$. The authors then apply the inequality

$$\frac{1}{2}\mathrm{d}_{\mathrm{TV}}^2 \leq \mathrm{d}_{\mathrm{H}}^2$$

where $\mathrm{d_H}$ is the Hellinger distance. Let P and Q be probability measures that are absolutely continuous with respect to some dominating measure λ, and let $f_P = \frac{dP}{d\lambda}$, $f_Q = \frac{dQ}{d\lambda}$ be their respective Radon-Nikodym derivatives. The Hellinger distance between P and Q is defined by

$$\mathrm{d}_{\mathrm{H}}^2(P, Q) = \frac{1}{2}\int (\sqrt{f_P} - \sqrt{f_Q})^2 d\lambda.$$

In the case where λ is Lebesgue measure, f_P and f_Q are the densities of P and Q. The main result of Kim et al. (2015) is

Theorem 2.1 (Kim et al. (2015), Theorem 1)**.** *Suppose $n = 2$ in (2.1). Then*

$$\mathrm{d}_{\mathrm{H}}^2(Y, Z) = e^{-\int_0^T \frac{1}{N(t)}dt}\left(1 - e^{-\frac{(a+b)S}{2abN_0}}\right)\frac{(\sqrt{a} - \sqrt{b})^2}{a+b}.$$

We give a proof in the appendix that fills in some additional details of the proof in Kim et al. (2015). Rather than obtaining bounds on the Bayes error rate using the Hellinger distance, we compute the probability of correct inference on ϑ.

In Figure 10, we compare our results to the Hellinger bounds of Kim et al. (2015) for different values of a, b, N_0. The upper bound based on the Hellinger distance from Kim et al. (2015) is given by

$$\frac{1}{2} + \frac{1}{2}\sqrt{2H^2(f_1, f_2)}$$

with $H^2(f_1, f_2)$ as in (J.3). Evidently the Hellinger bound is quite loose when $|a - b|$ is not near zero.

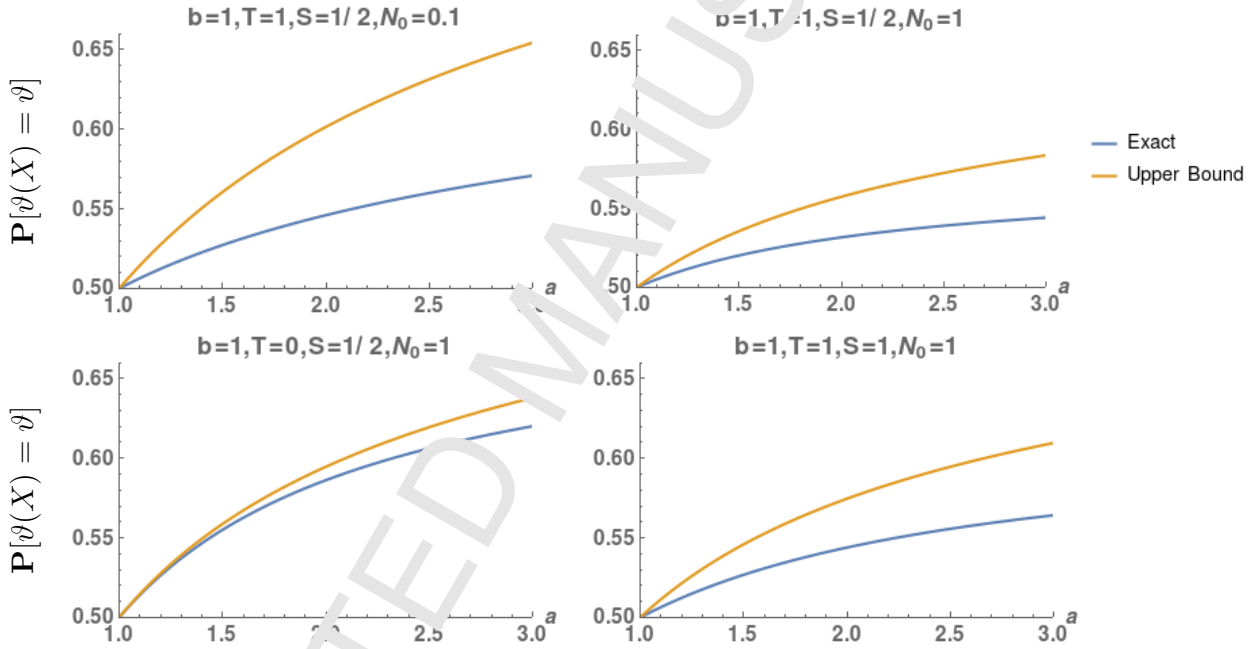


Figure 10: Exact $\mathbf{P}[\vartheta(X) = \vartheta]$ (blue) and upper bound on this quantity from Kim et al. (2015) (yellow) for different values of T, S, and N_0.

Kim et al. (2015) use the inequality

$$\mathrm{d}_H^2(Y^J, Z^J) \leq J\mathrm{d}_H^2(Y, Z), \tag{8.1}$$

which holds when the J genealogies are independent, in combination with Theorem 8.1, to obtain lower bounds on the error rate for J independent loci. They then use these lower bounds to calculate quantities like bounds on the minimal S such that the correct hypothesis will be selected with probability 0.95 for several examples.

It is worth noting that the existence of the inequality in (8.1) is not a special feature of the Hellinger distance. Indeed, the total variation distance obeys the inequalities

$$\mathrm{d}_{\mathrm{TV}}(Y^J, Z^J) \leq J\mathrm{d}_{\mathrm{TV}}(Y, Z) \tag{8.2}$$

$$\mathrm{d_{TV}}(Y^J, Z^J) \leq \sqrt{2J}\sqrt{\mathrm{d_{TV}}(Y, Z)}.$$

Although these inequalities are well-known, we provide a proof sketch in section Appendix K of the appendix. Thus there is no interpretability advantage in bounding the total variation for $J = 1$ by the Hellinger and then applying (8.1). We emphasize that an important message of this work is that bounds of the form (8.1) or (8.2) have significant limitations for understanding the behavior of $\mathrm{d_{TV}}(Y^J, Z^J)$ (equivalently, $\mathbf{P}[\vartheta(X^J) = \vartheta]$) for either large or small J. The quantity of interest $\mathbf{P}[\vartheta(X^J) = \vartheta]$ lies in the interval $[0.5, 1]$, so errors of size 0.1 or 0.05, or even 0.01 in some cases, are significant. This is reflected in the fact that roughly 3 times larger J is necessary to obtain $\mathbf{P}[\vartheta(X^J) = \vartheta] = 0.95$ in the human expansion scenario as the inequalities in Kim et al. (2015) suggest. The reason for this is straightforward. When J is small, the inequalities in (8.2) or (8.1) – particularly the latter, since there is also error due to bounding the total variation by the Hellinger – can be quite loose. When J grows large, the resulting bound on $\mathbf{P}[\vartheta(X^J) = \vartheta]$ quickly becomes 1, so the inequality is trivial. By contrast, $\mathbf{P}[\vartheta(X^J) = \vartheta]$ is never identically 1 for any finite J, and only approaches 1 in the limit as $J \to \infty$. When quantities such as $\min_J\{J : \mathbf{P}[\vartheta(X^J) = \vartheta] > 1 - \alpha\}$ are of interest, the answers obtained using the inequality (8.1) can differ substantially from the exact value.

This motivates our preference throughout the paper of giving the exact value of $\mathbf{P}[\vartheta(X^J) = \vartheta]$, which allows us to compute exactly the value of S to achieve the desired Bayes error rate for any J. It is preferable to do this numerically using Monte Carlo when the exact expression is unavailable than it is to use the upper bounds in (8.2) or (8.1) when seeking sharp results, which is our focus here. The results on the minimal number of loci J necessary to achieve a fixed error rate differ substantially from the results in Kim et al. (2015). The looseness of the bound on $\mathbf{P}[\vartheta(X) = \vartheta]$ obtained using the Hellinger distance is clear from Figure 10, but as we now show, additional looseness is introduced by relying on (8.1).

The expression in (6.2) can be directly compared with Theorem 3.2 of Kim et al. (2015). Translated into our notation and conventions, this result states that

$$\mathbf{P}[\vartheta(X^J) = \vartheta] \leq \frac{1}{2} + \frac{1}{4}\sqrt{J(n-1)}\left(\frac{1}{c} - 1\right). \tag{8.3}$$

Figure 11 shows the bound from (8.3) along with the exact probability of identifying the true $N(t)$ as a function of c for $n = 2$ and different values of J. The bound is apparently quite loose when c is not close to 1. It becomes trivial (greater than 1) for $c \approx 0.7$ when $J = 1$ and $c \approx 0.7$ when $J = 10$. The differences can be extremely large. Note that the quantity $\mathbf{P}[\vartheta(X^J) = \vartheta] \in [0.5, 1]$, so if one knows nothing at all about the problem and just approximates $\mathbf{P}[\vartheta(X^J) = \vartheta]$ by 0.75, it is never possible to make an absolute error of more than 0.25. When $c \approx 0.65$ and $J = 10$, the upper bound from Kim et al. (2015) given in (8.3) gives $\mathbf{P}[\vartheta(X^J) = \vartheta] = 1$, while the exact value is $\mathbf{P}[\vartheta(X^J) = \vartheta] \approx 0.75$ (see the right panel of Figure 11. This is an absolute

error of 0.25, which is the largest error one can ever make by using the naïve estimate $\mathbf{P}[\vartheta(X^J) = \vartheta] = 0.75$. Thus, the bound in (8.3) is sometimes no better than guessing, even in very simple settings.

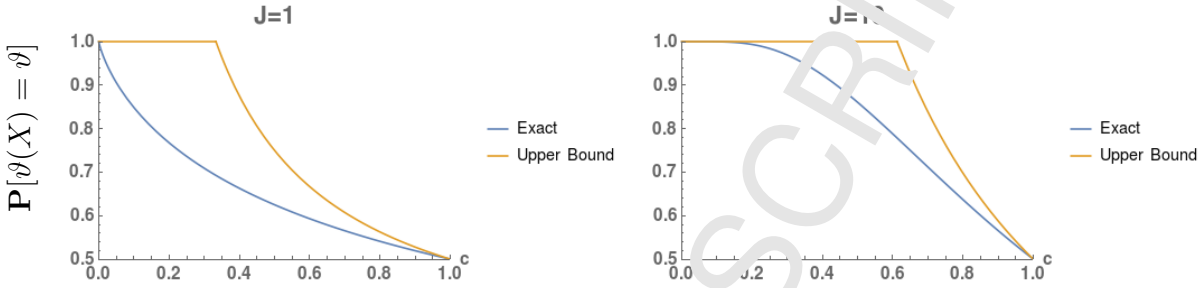


Figure 11: Exact $\mathbf{P}[\vartheta(X^J) = \vartheta]$ compared to the upper bound from Theorem 3.2 of Kim et al. (2015) as a function of c for two different values of J.

9. Risk of point estimates under conjugate priors

Although our focus has been on inferential limits for distinguishing among two states of nature, we briefly consider estimation of a constant population size trajectory. We asses the risk of estimators of the function $\Lambda(x)$ in the case of $n = 2$ and $N(t) = \frac{1}{c}$ with conjugate priors on c. In this setting, the coalescent time x is Exponential(c) with conjugate prior $c \sim \text{Gamma}(\alpha, \beta)$ and for a sample of J independent pairwise coalescent times we have

$$c \mid x^{(1)}, \ldots, x^{(J)} \sim \text{Gamma}(\alpha + J, \beta + J\bar{x})$$

with posterior expectation

$$\hat{c} = \mathbf{E}[c \mid x^{(1)}, \ldots, x^{(J)}] = \frac{\alpha + J}{\beta + J\bar{x}}.$$

Note that

$$Z \equiv J\bar{X} \mid c \sim \text{Gamma}(J, c)$$

so the Frequentist squared error risk of the posterior expectation of c is

$$R(\hat{c}, c) := \int_0^\infty \left(\frac{\alpha + J}{\beta + z} - c \right)^2 \frac{c^J}{\Gamma(J)} z^{J-1} e^{-cz} dz. \tag{9.1}$$

Taking $\alpha = \beta = 1$ – the unit-rate exponential prior on c – the risk can be expressed as

$$R(\hat{c}, c) = c^J (J+1)^2 \Psi(J, J-1, c) - 2(J+1) c^{J+1} \Psi(J, J, c) + c^2 \tag{9.2}$$

where Ψ is the Tricomi confluent hypergeometric function (see (Gradshteyn and Ryzhik, 1996, 9.211)) defined by

$$\Psi(a,b,c) \equiv \int_0^\infty z^{a-1}e^{-cz}(1+z)^{b-a-1}dz;$$

details are given in the Appendix. Figure 12 shows the square root of risk as a function of the number of loci J for values of $J \in \{1,\ldots,100\}$ with $c=1$. The root risk decreases logarithmically in J; it is approximately 0.1 for $J=100$, and about 0.24 for $J=20$. Thus, if one wants the root risk to be small relative to the truth, it is necessary to have J rather large. In this example, in order to have the root risk be about 10 percent the magnitude of the truth, we need $J \approx 100$.

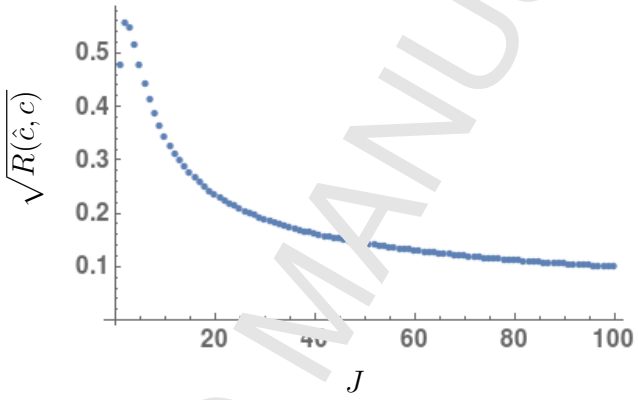


Figure 12: Root risk of Bayes estimator with $\alpha=\beta=1$ and $c=1$.

10. Discussion

Availability of ancient and present-day DNA samples from a population allows statistical reconstruction of the effective population size trajectory. The effective population size is a measure of relative genetic diversity whose actual magnitude is not easily interpreted in units of census population size (Wakeley and Sargsyan, 2009). However, changes of effective population size over time are informative about the genetic history of the population. In this manuscript, we assess the ability to differentiate or classify between alternative hypotheses about the effective population size.

Assessment of inferential limits in population genetic studies is becoming important in the face of ongoing large-scale studies of genetic variation. Statistical methods are usually restricted to small samples or rely on approximations and insufficient summary statistics. As such, choosing the optimal subset of data with which to perform statistical inference is of great interest. Aspects of the data and adequacy of the model will affect the ability to draw meaningful conclusions. For most of our results, we have

eliminated the effect of factors such as data quality, sample selection and sequence alignment and concentrated on the ideal scenario of having a complete realization of the genealogical process free of errors. In practice, genealogies are not available and instead we observe DNA sequence variation; therefore our results are upper bounds on the achievable probability of recovering the true population size history in population genetic studies. These results provide guidance to practitioners in choosing a sampling design subject to computational constraints. In particular, they give insight into the key questions of which scientific hypotheses can be assessed, and the optimal choice of the number of loci, sampling times, and number of individuals to include in a sample. They also offer a possible explanation for disagreement in the literature over timing and duration of historical genetic events such as the out-of-Africa human population bottleneck, suggesting that some studies may simply not have sufficient data to distinguish between the hypotheses of interest with high probability.

Fu and Li (1993); Pluzhnikov and Donnelly (1996) and Felsenstein (2006) argued that in the constant population model ($\theta = 2N\mu$), accuracy of estimators of θ increases linearly in the number of independent loci, logarithmically in the number of samples, and is unaffected by sequence length. In the coalescent with variable population size, Myers et al. (2008) showed that estimators based on the SFS cannot distinguish between two alternative hypotheses. Terhorst and Song (2015) showed that estimators of $N(t)$, based on the same statistic SFS, have minimax rate of convergence that is logarithmic in the number of independent loci and independent of the number of individuals sampled. Kim et al. (2015) provided lower bounds on Bayes error rates from pairwise coalescent data from independent loci and show that the Bayes error rate goes to zero with the squared root of the number of loci. Our work is closely related to the work of Kim et al. (2015). In this work, we investigate the number of loci and samples needed to correctly differentiate between alternative hypotheses about the effective population size (one minus the Bayes error rate) when genealogical data are available. We consider cases under independent loci and under some models of recombination. Our calculations from pairwise coalescent data and independent loci differ to Kim et al. (2015) in that we provide exact calculations instead of bounds. We show that for some cases, the difference between the bounds and the exact calculations is significant. Our results support a complex view of the value of additional samples or loci. While in general, the improvement in the probability of recovering the true population history appears to be sublinear in both J and n, the improvement from adding an additional sample or locus depends greatly on the details of the two hypotheses being considered and the independence assumption across loci. For example, increasing from $n = 2$ to $n = 3$ samples can in some cases double the excess probability of recovering the truth $\mathbf{P}[\vartheta(X) = \vartheta] - 1/2$ (the probability is always lower bounded by 1/2). In general, smaller improvements are seen from increasing J, but we have demonstrated that high probability of recovering the true population size history in the human expansion example is attainable using values of n and J that are available from modern datasets and for which exact computation

is feasible. In addition, our results suggest that incorporation of ancient genomes is the optimal strategy to improve inferential performance in the human expansion problem, which is of significant interest in human population genetics.

Pluzhnikov and Donnelly (1996) considered the constant population model with recombination and argued that when the recombination rate is high, increasing the sequence length effectively increases the number of independent loci. Indeed, when two genomic segments are separated by a recombination event, individuals at these two segments (loci) derive from two different but correlated genealogies. As the number of recombination events increases, the correlation between the two genealogies becomes weaker, and hence increasing the length of sequenced segments increases the opportunity to observe a larger number of realizations from genealogically independent loci (Palacios et al., 2015; Griffiths and Marjoram, 1997). Our results for the pairwise SMC' model of recombination support the conclusion that loci separated by recombination events have nearly the statistical value as the same number of independent loci in the human expansion example. For a different testing scenario of constant population size and pairwise coalescent data, our results show that independent loci offers better statistical value in some cases. This suggests that pairwise SMC' is a very powerful framework for inference of population size trajectories. An interesting future area of research is to analytically explore the effect of the number of loci and samples under the SMC' in more general settings that the ones explored in this manuscript.

Acknowledgments

The authors thank the editor and three anonymous referees whose suggestions considerably improved the manuscript. In particular, the idea of proving Theorem 3.2 was suggested by a referee. JAP acknowledges support from NIH R01GM131404-01.

References

Beerli, P., Felsenstein, J., 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proceedings of the National Academy of Sciences 98 (8), 4563–4568.

Carmi, S., Wilton, P. R., Wakeley, J., Pe'er, I., 2014. A renewal theory approach to IBD sharing. Theoretical Population Biology 97, 35 – 48.

Drummond, A., Suchard, M., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution 29, 1969–1973.

Felsenstein, J., 2006. Accuracy of coalescent likelihood estimates: Do we need more sites, more sequences, or more loci? Molecular Biology and Evolution 23 (3), 691–700.

Felsenstein, J., Rodrigo, A. G., 1999. Coalescent Approaches to HIV Population Genetics. In: The Evolution of HIV. Johns Hopkins University Press, pp. 233–272.

Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., Nickel, B., Peltzer, A., Rohland, N., Slon, V., Talamo, S., Lazaridis, I., Lipson, M., Mathieson, I., Schiffels, S., Skoglund, P., Derevianko, A. P., Drozdov, N., Slavinsky, V., Tsybankov, A., Cremonesi, R. G., Mallegni, F., Gély, B., Vacca, E., Morales, M. R. G., Straus, L. G., Neugebauer-Maresch, C., Teschler-Nicola, M., Constantin, S., Moldovan, O. T., Benazzi, S., Peresani, M., Coppola, D., Lari, M., Ricci, S., Ronchitelli, A., Valentin, F., Thevenet, C., Wehrberger, K., Grigorescu, D., Rougier, H., Crevecoeur, I., Flas, D., Semal, P., Mannino, M. A., Cupillard, C., Bocherens, H., Conard, N. J., Harvati, K., Moiseyev, V., Drucker, D. G., Svoboda, J., Richards, M. P., Caramelli, D., Pinhasi, R., Kelso, J., Patterson, N., Krause, J., Pääbo, S., Reich, D., 2016. The genetic history of ice age europe. Nature 534, 200–205.
URL http://dx.doi.org/10.1038/nature17993

Fu, Y. X., Li, W. H., 1993. Statistical tests of neutrality of mutations. Genetics 133 (3), 693–709.

Gao, F., Keinan, A., 2016. Explosive genetic evidence for explosive human population growth. Current Opinion in Genetics and Development 41 (Supplement C), 130 – 139, genetics of human origin.

Gattepaille, L., Günther, T., Jakobsson, M., 2016. Inferring past effective population size from distributions of coalescent times. Genetics.

Gradshteyn, I. S., Ryzhik, I. M., 1996. Table of integrals, series, and products. Academic press.

Griffiths, R. C., Marjoram, P., 1997. An ancestral recombination graph. In: Donnelly, P., Tavaré, S. (Eds.), Progress in population genetics and human evolution. Vol. 87 of IMA Volumes in Mathematics and Its Applications. Springer Verlag, New York, pp. 257–270.

Griffiths, R. C., Tavaré, S., Jun. 1994. Sampling theory for neutral alleles in a varying environment. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 344, 403–410.

Iles, J. C., Raghwani, J., Harrison, G. A., Pepin, J., Djoko, C. F., Tamoufe, U., Le-Breton, M., Schneider, B. S., Fair, J. N., Tshala, F. M., Kayembe, P. K., Muyembe, J. J., Edidi Basepeo, S., Wolfe, N. D., Simmonds, P., Klenerman, P., Pybus, O. G., 09 2014. Phylogeography and epidemic history of hepatitis C virus genotype 4 in Africa. Virology 464-465 (100), 233–243.

Kim, J., Mossel, E., Rácz, M. Z., Ross, N., 2015. Can one hear the shape of a population history? Theoretical population biology 100, 26–38.

Kingman, J. F. C., 1982. The coalescent. Stochastic Processes and Their Applications 13 (3), 235–248.

Kuhner, M., Yamato, J., Felsenstein, J., 1995. Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. Genetics 140 (4), 1421–1430.

Li, H., Durbin, R., Jul. 2011. Inference of human population history from individual whole-genome sequences. Nature 475 (7357), 493–496.

Marjoram, P., Wall, J., 2006. Fast "coalescent" simulation. BMC Genetics 7 (1).

McVean, G., Cardin, N., Jul 2005. Approximating the coalescent with recombination. Philos Trans R Soc Lond B Biol Sci 360 (1459), 1387–1393.

Myers, S., Fefferman, C., Patterson, N., 2008. Can one learn history from the allelic spectrum? Theoretical Population Biology 73 (3), 342 – 348.

Palacios, J. A., Wakeley, J., Ramachandran, S., 2015. Bayesian nonparametric inference of population size changes from sequential genealogies. Genetics 201 (1), 281–304.

Pluzhnikov, A., Donnelly, P., 1996. Optimal sequencing strategies for surveying molecular genetic diversity. Genetics 144, 1247–1262.

Sainudiin, R., Thornton, K., Harlow, J., Booth, J., Stillman, M., Yoshida, R., Griffiths, R., McVean, G., Donnelly, P., 2011. Experiments with the site frequency spectrum. Bulletin of Mathematical Biology 73 (4), 829–872.

Schiffels, S., Durbin, R., 2014. Inferring human population size and separation history from multiple genome sequences. Nature Genetics 46 (8), 919–925.

Shapiro, B., Drummond, A. J., Rambaut, A., Wilson, M. C., Matheus, P. E., Sher, A. V., Pybus, O. G., Gilbert, M. T. P., Barnes, I., Binladen, J., Willerslev, E., Hansen, A. J., Baryshnikov, G. F., Burns, J. A., Davydov, S., Driver, J. C., Froese, D. G., Harington, C. R., Keddie, G., Kosintsev, P., 2004. Rise and fall of the Beringian steppe bison. Science 306 (5701), 1561–1565.

Sheehan, S., Harris, K., Song, Y. S., 2013. Estimating variable effective population sizes from multiple genomes: A sequentially Markov conditional sampling distribution approach. Genetics 194 (3), 647–662.

Slatkin, M., Hudson, R., 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics 129 (2), 555–562.

Stephens, M., Donnelly, P., 2000. Inference in molecular population genetics. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62 (4), 605–635.

Terhorst, J., Kamm, J. A., Song, Y. S., 2017. Robust and scalable inference of population history from hundreds of unphased whole genomes. Nature Genetics 49, 303–309.

Terhorst, J., Song, Y. S., 2015. Fundamental limits on the accuracy of demographic inference based on the sample frequency spectrum. Proceedings of the National Academy of Sciences 112 (25), 7677–7682.

Tong, Y.-G., Shi, W.-F., LiuDi, Qian, J., Liang, L., Bo, X. C., Liu, J., Ren, H.-G., Fan, H., Ni, M., Sun, Y., Jin, Y., Teng, Y., Li, Z., Kargbo, D., Dafae, F., Kanu, A., Chen, C.-C., Lan, Z.-H., Jiang, H., Luo, Y., Lu, H.-J., Zhang, X.-G., Yang, F., Hu, Y., Cao, Y.-X., Deng, Y.-Q., Su, H.-X., Su, Y., Liu, W.-S., Wang, Z., Wang, C.-Y., Bu, Z.-Y., Guo, Z.-D., Zhang, L.-B., Jie, W.-M., Bai, C.-Q., Sun, C.-H., An, X.-P., Xu, P.-S., Zhang, X.-L.-L., Huang, Y., Mi, Z.-Q., Yu, D., Yao, H.-W., Feng, Y., Xia, Z.-P., Zheng, X.-X., Yang, S.-T., Lu, B., Jiang, J.-F., Kargbo, B., He, F.-C., Gao, G. F., Cao, W.-C., in Sierra Leone, T. C. M. L. T. T., 2015. Genetic diversity and evolutionary dynamics of Ebola virus in Sierra Lhateone. Nature 524 (7563), 93–96.

Wakeley, J., Jun. 2008. Coalescent Theory: An Introduction. Roberts & Company Publishers.

Wakeley, J., Sargsyan, O., 2009. Extensions of the coalescent effective population size. Genetics 181, 341–345.

Appendix A. Proof of theorem 3.1

Proof. Define

$$\Lambda(w, x) \equiv \int_w^x \frac{1}{N(t)} dt.$$

For shorthand we write $\Lambda(x) = \Lambda(0, x)$. $\Lambda : \mathbb{R}_+ \to \mathbb{R}_+$ is a monotone strictly increasing function, which is enough to guarantee the existence of an inverse

$$\Lambda^{-1}(t) = x \iff \Lambda(x) = t,$$

The likelihood ratio for H_1 vs H_2 (3.1) can be expressed by

$$\log \mathrm{BF}_{12}(x) = \begin{cases} 0 & x < T \\ \log \frac{b}{a} - \frac{x-T}{aN_0} + \frac{x-T}{bN_0} & T \le x < T + S \\ \frac{S}{bN_0} - \frac{S}{aN_0} & T + S \le x \end{cases} \tag{A.1}$$

Notice that the waiting time until the coalescent event has survival function

$$\mathbf{P}[X > x] = e^{-\Lambda(x)}.$$

Now we want to calculate $\mathbf{P}[\vartheta(X) = 1 \mid H_1]$. Assume that if $\log \mathrm{BF}_{12}(x) = 0$ we select either H_1 or H_2 by flipping a fair coin. If $a > b$ then

$$\log \mathrm{BF}_{12}(x) > 0, T \le x \le T + S \iff x > \delta + T,$$

and if $b > a$

$$\log \mathrm{BF}_{12}(x) > 0, T \le x \le T + S \iff x < \delta + T.$$

Assuming $a > b$ and denoting $f_i(x)$ the density under H_i for $i = 1, 2$, we have

$$\begin{aligned} \mathbf{P}[\vartheta(X) = 1 \mid H_1] &= \frac{1}{2}\mathbf{P}[X < T \mid H_1] + \int_T^{T+S} \mathbf{1}\left\{x > \frac{ab}{b-a} N_0 \log \frac{b}{a} + T\right\} f_1(x) dx \\ &\quad + \mathbf{1}\{b < a\}\mathbf{P}[X > T + S \mid H_1] \\ &= \frac{1}{2}(1 - e^{-\Lambda(T)}) + \int_{T+(\delta \wedge S)}^{T+S} e^{-\Lambda(T)} \frac{1}{aN_0} e^{-\frac{x-T}{aN_0}} dx + e^{-\Lambda(T) - \frac{S}{aN_0}} \\ &= \frac{1}{2}(1 - e^{-\Lambda(T)}) + e^{-\Lambda(T)}\left[e^{-\frac{\delta \wedge S}{aN_0}} - e^{-\frac{S}{aN_0}}\right] + e^{-\Lambda(T) - \frac{S}{aN_0}}, \end{aligned}$$

and

$$\begin{aligned} \mathbf{P}[\vartheta(X) = 2 \mid H_2] &= \frac{1}{2}\mathbf{P}[X < T \mid H_2] + \int_T^{T+S} \mathbf{1}\left\{x < \frac{ab}{b-a} N_0 \log \frac{b}{a} + T\right\} f_2(x) dx \\ &\quad + \mathbf{1}\{a < b\}\mathbf{P}[X > T + S \mid H_2] \end{aligned}$$

$$= \frac{1}{2}(1 - e^{-\Lambda(T)}) + \int_T^{T+(\delta\wedge S)} e^{-\Lambda(T)} \frac{1}{bN_0} e^{-\frac{x-T}{bN_0}} dx$$
$$= \frac{1}{2}(1 - e^{-\Lambda(T)}) + e^{-\Lambda(T)} \left[1 - e^{-\frac{\delta\wedge S}{bN_0}}\right]$$

Assuming equal prior probability of H_1 and H_2 we get

$$\begin{aligned}\mathbf{P}[\vartheta(X) = \vartheta] &= \frac{1}{2}(1 - e^{-\Lambda(T)} + e^{-\Lambda(T) - \frac{S}{(a\vee b)N_0}}) \\ &+ \frac{1}{2}e^{-\Lambda(T)} \left[e^{-\frac{\delta\wedge S}{aN_0}} - e^{-\frac{S}{aN_0}}\right] + \frac{1}{2}e^{-\Lambda(T)} \left[1 - e^{-\frac{\delta\wedge S}{bN_0}}\right] \\ &= \frac{1}{2} + \frac{1}{2}e^{-\Lambda(T)} \left(e^{-\frac{S}{(a\vee b)N_0}} - e^{-\frac{S}{aN_0}}\right) + \frac{1}{2}e^{-\Lambda(T)} \left(e^{-\frac{\delta\wedge S}{aN_0}} - e^{-\frac{\delta\wedge S}{bN_0}}\right) \\ &= \frac{1}{2} + \frac{1}{2}e^{-\Lambda(T)} \left(e^{-\frac{\delta\wedge S}{aN_0}} - e^{-\frac{\delta\wedge S}{bN_0}}\right).\end{aligned}$$

This assumed $a > b$. If instead $b > a$ then the inequalities in the integrand when we integrate between T and $T + S$ would be reversed, so the exact expression for any $a > 0, b > 0$ is

$$\mathbf{P}[\vartheta(X) = \vartheta] = \frac{1}{2} + \frac{1}{2}e^{-\Lambda(T)} \left(e^{-\frac{\delta\wedge S}{(a\vee b)N_0}} - e^{-\frac{\delta\wedge S}{(a\wedge b)N_0}}\right). \tag{A.2}$$

□

Appendix B. Proof of Theorem 3.2

Fix an unknown constant $\epsilon > 0$ and assume that we observe $Y = X + \epsilon$. The likelihood ratio for H_1 vs H_2 (3.1) can be expressed by

$$\log \mathrm{BF}_{12}(y) = \begin{cases} 0 & 0 < y < T \\ \log\frac{b}{a} - \frac{y-T}{aN_0} + \frac{y-T}{bN_0} & T \le y < T + S \\ \frac{S}{bN_0} - \frac{S}{aN_0} & T + S \le y \end{cases} \tag{B.1}$$

Assume first $a > b$. Then

$$\log \mathrm{BF}_{12}(y) > 0, T \le x + \epsilon \le T + S \Leftrightarrow y > \delta + T$$

just as before. The main difference here is that the distribution of $Y = X + \epsilon$ differs from that of X. If we assume $\epsilon > 0$ then y has density $f_i(y - \epsilon)$ under hypothesis H_i (the point of assuming $\epsilon > 0$ is so that we can avoid the problem of the density being zero on the negative half-line). So we have

$$\begin{aligned}\mathbf{P}[\vartheta(Y) = 1 \mid H_1] &= \frac{1}{2}\mathbf{P}[0 < Y < T \mid H_1] + \int_T^{T+S} \mathbf{1}\{y > \delta + T\} f_1(y - \epsilon)dy \\ &+ \mathbf{1}\{b < a\}\mathbf{P}[Y > T + S \mid H_1]\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}\mathbf{P}[0 < X < T-\epsilon \mid H_1] + \int_{T-\epsilon}^{T+S-\epsilon} \mathbf{1}\left\{x > \delta + T - \epsilon\right\} f_1(x)dx \\
&+ \mathbf{P}[X > T + S - \epsilon \mid H_1] \\
&= \frac{1}{2}(1 - e^{-\Lambda(T-\epsilon)}) \\
&+ \int_{T-\epsilon}^{T} \mathbf{1}\left\{x > \delta + T - \epsilon\right\} f_1(x)dx + \int_{T}^{T+S-\epsilon} \mathbf{1}\left\{x > \delta + T - \epsilon\right\} f_1(x)dx \\
&+ \int_{T+S-\epsilon}^{T+S} f_1(x)dx + \int_{T+S}^{\infty} f_1(x)dx
\end{aligned}
$$

Performing the last two integrals and rearranging terms we obtain

$$
\begin{aligned}
&= \frac{1}{2}(1 - e^{-\Lambda(T-\epsilon)}) + e^{-\Lambda(T)-\frac{S-\epsilon}{aN_0}} - e^{-\Lambda(T)-\frac{S}{aN_0}} + e^{-\Lambda(T)-\frac{S}{aN_0}} \\
&+ \int_{(T+\delta-\epsilon)\wedge T}^{T} f_1(x)dx + \int_{T+(((S\wedge\delta)-\epsilon)\vee 0)}^{T+S-\epsilon} \mathbf{1}\left\{x > \delta + T - \epsilon\right\} f_1(x)dx \\
&= \frac{1}{2}(1 - e^{-\Lambda(T-\epsilon)}) + e^{-\Lambda(T)-\frac{S}{aN_0}} + e^{-\Lambda((T+\delta-\epsilon)\wedge T)} - e^{-\Lambda(T)} \\
&+ \int_{T+(((S\wedge\delta)-\epsilon)\vee 0)}^{T+S-\epsilon} \mathbf{1}\left\{x > \delta + T - \epsilon\right\} f_1(x)dx \\
&= \frac{1}{2}(1 - e^{-\Lambda(T-\epsilon)}) + e^{-\Lambda(T)-\frac{S-\epsilon}{aN_0}} + e^{-\Lambda((T+\delta-\epsilon)\wedge T)} - e^{-\Lambda(T)} \\
&+ e^{-\Lambda(T)-\frac{((S\wedge\delta)-\epsilon)\vee 0}{aN_0}} - e^{-\Lambda(T)-\frac{S-\epsilon}{aN_0}} \\
&= \frac{1}{2}(1 - e^{-\Lambda(T-\epsilon)}) + e^{-\Lambda((T+\delta-\epsilon)\wedge T)} - e^{-\Lambda(T)} + e^{-\Lambda(T)-\frac{((S\wedge\delta)-\epsilon)\vee 0}{aN_0}}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{P}[\vartheta(Y) = 2 \mid H_2] &= \frac{1}{2}\mathbf{P}[X < T - \epsilon \mid H_2] + \int_{T-\epsilon}^{T+S-\epsilon} \mathbf{1}\left\{x < \delta + T - \epsilon\right\} f_2(x)dx \\
&+ \mathbf{1}\{a < b\}\mathbf{P}[X > T + S - \epsilon \mid H_2] \\
&= \frac{1}{2}(1 - e^{-\Lambda(T-\epsilon)}) + \int_{T-\epsilon}^{T\wedge(T+\delta-\epsilon)} f_2(x)dx + \int_{T}^{T+S-\epsilon} \mathbf{1}\left\{x < \delta + T - \epsilon\right\} f_2(x)dx \\
&= \frac{1}{2}(1 - e^{-\Lambda(T-\epsilon)}) + e^{-\Lambda(T-\epsilon)} - e^{-\Lambda(T\wedge(T+\delta-\epsilon))} + \int_{T}^{T+((S\wedge\delta)-\epsilon)\vee 0)} f_2(x)dx \\
&= \frac{1}{2}(1 - e^{-\Lambda(T-\epsilon)}) + e^{-\Lambda(T-\epsilon)} - e^{-\Lambda(T\wedge(T+\delta-\epsilon))} + e^{-\Lambda(T)} - e^{-\Lambda(T)-\frac{((S\wedge\delta)-\epsilon)\vee 0}{bN_0}}
\end{aligned}
$$

and again using the equal prior probability and combining

$$
\mathbf{P}[\vartheta(Y) = \vartheta] = \frac{1}{2} + \frac{1}{2}e^{-\Lambda(T)}\left(e^{-\frac{((S\wedge\delta)-\epsilon)\vee 0}{aN_0}} - e^{-\frac{((S\wedge\delta)-\epsilon)\vee 0}{bN_0}}\right).
$$

Assuming $S \wedge \delta > \epsilon$ we have

$$\begin{aligned}
\mathbf{P}[\vartheta(Y) = \vartheta] &= \frac{1}{2} + \frac{1}{2}e^{-\Lambda(T)}\left(e^{-\frac{(S\wedge\delta)-\epsilon}{aN_0}} - e^{-\frac{(S\wedge\delta)-\epsilon}{bN_0}}\right) \\
&= \frac{1}{2} + \frac{1}{2}e^{-\Lambda(T)}\left(e^{-\frac{(S\wedge\delta)}{aN_0}}\left(1 + \frac{\epsilon}{aN_0} + \mathcal{O}(\epsilon^2)\right) - e^{-\frac{(S\wedge\delta)}{bN_0}}\left(1 + \frac{\epsilon}{bN_0} + \mathcal{O}(\epsilon^2)\right)\right) \\
&= \mathbf{P}[\vartheta(X) = \vartheta] + \frac{\epsilon}{2}e^{-\Lambda(T)}\left(e^{-\frac{(S\wedge\delta)}{aN_0}}\frac{1}{aN_0} - e^{-\frac{(S\wedge\delta)}{bN_0}}\frac{1}{bN_0}\right) + \mathcal{O}(\epsilon^2).
\end{aligned}$$

If $\epsilon \sim \text{Exponential}(\lambda)$ then since

$$\int_0^{S\wedge\delta} \lambda e^{-\lambda\epsilon} e^{-\frac{((S\wedge\delta)-\epsilon)\vee 0}{aN_0}} d\epsilon = e^{-\frac{(S\wedge\delta)}{aN_0}} \int_0^{S\wedge\delta} e^{\frac{\epsilon}{aN_0}} \lambda e^{-\lambda\epsilon} d\epsilon = \frac{\lambda N_0}{1 - a\lambda N_0}\left(e^{-\lambda(S\wedge\delta)} - e^{-\frac{(S\wedge\delta)}{aN_0}}\right)$$

we obtain

$$\begin{aligned}
\mathbf{P}[\vartheta(Y) = \vartheta] &= \mathbf{E}[\mathbf{P}[\vartheta(Y) = \vartheta \mid \epsilon]] = \frac{1}{2} + \mathbf{E}\left[\frac{1}{2}e^{-\Lambda(T)}\left(e^{-\frac{((S\wedge\delta)-\epsilon)\vee 0}{aN_0}} - e^{-\frac{((S\wedge\delta)-\epsilon)\vee 0}{bN_0}}\right)\right] \\
&= \frac{1}{2} + \frac{1}{2}e^{-\Lambda(T)}\mathbf{E}\left[\left(e^{-\frac{((S\wedge\delta)-\epsilon)\vee 0}{aN_0}} - e^{-\frac{((S\wedge\delta)-\epsilon)\vee 0}{bN_0}}\right)\mathbf{1}(\epsilon < (S\wedge\delta))\right] \\
&= \frac{1}{2} + \frac{1}{2}e^{-\Lambda(T)}\left(\frac{aN_0\lambda(1 - e^{\frac{S\wedge\delta}{aN_0} - (S\wedge\delta)\lambda})}{aN_0\lambda - 1}e^{-\frac{(S\wedge\delta)}{aN_0}} - \frac{bN_0\lambda(1 - e^{\frac{S\wedge\delta}{bN_0} - (S\wedge\delta)\lambda})}{bN_0\lambda - 1}e^{-\frac{(S\wedge\delta)}{bN_0}}\right) \\
&= \mathbf{P}[\vartheta(X) = \vartheta] + \frac{1}{2}e^{-\Lambda(T)}\left(\frac{1 - aN_0\lambda e^{\frac{S\wedge\delta}{aN_0} - (S\wedge\delta)\lambda}}{aN_0\lambda - 1}e^{-\frac{(S\wedge\delta)}{aN_0}} - \frac{1 - bN_0\lambda e^{\frac{S\wedge\delta}{bN_0} - (S\wedge\delta)\lambda}}{bN_0\lambda - 1}e^{-\frac{(S\wedge\delta)}{bN_0}}\right)
\end{aligned}$$

Appendix C. Proof of Theorem 3.3

Proof. Fix an integer $J \geq 1$ and define $a_1 = a, a_2 = b$ for ease of notation. Without loss of generality, take $a > b$. We first define the following auxiliary functions

$$\begin{aligned}
Q_i(T) &\equiv e^{-\int_0^T \frac{dt}{N_i(t)}}, \quad Q_i(T, T+S) \equiv e^{-\int_T^{T+S} \frac{dt}{N_i(t)}} \\
q_i(T) &\equiv \frac{1}{N_i(T)} e^{-\int_0^T \frac{dt}{N_i(t)}}, \quad q_i(T, T+S) \equiv \frac{1}{N_i(T+S)} e^{-\int_T^{T+S} \frac{dt}{N_i(t)}}
\end{aligned}$$

The coalescent density for a coalescent time with effective population size trajectory N for the intervals $(0, T]$ and $(T+S, \infty)$ and N_i for the interval $(T, T+S]$ is

$$f_i(t) = \begin{cases} q(t) & 0 < t < T \\ Q(T)q_i(T, t) & T \leq t < T+S \\ Q(T)Q_i(T, T+S)q(T+S, t) & t \geq T+S \end{cases}$$

so that the likelihood ratio for a single time point can be expressed as

$$\frac{f_1(x^j)}{f_2(x^j)} = \left[\frac{q_1(T, x^j)}{q_2(T, x^j)}\right]^{\mathbf{1}\{T \leq x^j < T+S\}} \left[\frac{Q_1(T, T+S)}{Q_2(T, T+S)}\right]^{\mathbf{1}\{x^j \geq T+S\}}$$

$$= \left[\frac{b}{a} e^{-\frac{(b-a)(x^j - T)}{abN_0}}\right]^{\mathbf{1}\{T \le x^j < T+S\}} \left[e^{-S\frac{(b-a)}{abN_0}}\right]^{\mathbf{1}\{x^j \ge T+S\}},$$

giving

$$\log \prod_{j=1}^{J} \frac{f_1(x^j)}{f_2(x^j)} = \sum_{j=1}^{J} \mathbf{1}\{T < x^j \le T+S\} \left[\log \frac{b}{a} - (x^j - T)\frac{(b-a)}{abN_0}\right] - \sum_{j=1}^{J} \mathbf{1}\{x^j > T+S\} \frac{S(b-a)}{abN_0}.$$

Defining

$$\ell_1 = \sum_{j=1}^{J} \mathbf{1}\{x^j \le T\}, \quad \ell_2 = \sum_{j=1}^{J} \mathbf{1}\{T < x^j \le T+S\}, \quad \ell_3 = \sum_{j=1}^{J} \mathbf{1}\{x^j \ge T+S\},$$

we have that $\log \mathrm{BF}_{12} > 0$ when

$$\sum_{j=1}^{J} \mathbf{1}\{T < x^j \le T+S\} \left[\log \frac{b}{a} - (x^j - T)\frac{(b-a)}{abN_0}\right] > \sum_{j=1}^{J} \mathbf{1}\{x^j > T+S\} \frac{S(b-a)}{abN_0}$$

$$\ell_2 \left(\log \frac{b}{a} + T\frac{(b-a)}{abN_0}\right) - \ell_2 T\frac{(b-a)}{abN_0} > \frac{(b-a)}{abN_0} \sum_{j:x^j \in [T,T+S]} x^j$$

$$\ell_2 \left(\frac{abN_0}{a-b} \log \frac{a}{b} + T\right) - \ell_3 S < \sum_{j:x^j \in [T,T+S]} x^j,$$

$$\ell_2 (\delta + T) - \ell_3 S < \sum_{j:x^j \in [T,T+S]} x^j,$$

where the inequality reversed since $(b-a)/(abN_0)$ is negative.

Denote by $L = (L_1, L_2, L_3)$ the random vector whose observed entries are $\ell = (\ell_1, \ell_2, \ell_3)$. Now, $\log \mathrm{BF}_{12} = 0$ only if $x^j < T$ for all $j = 1, \ldots, J$. In this case, we flip a fair coin and accept H_1 if it shows heads. Moreover, if $L_2 = 0$ and $L_3 > 0$, then $\log \mathrm{BF}_{12} > 0$. Notice that for a generic coalescent time X

$$\begin{aligned} L \mid H_i &\sim \text{Multinomial}(J, \mathbf{p}) \\ p_1 &= \mathbf{P}[X \le T] = (1 - e^{-\Lambda(T)}) \\ p_2 &= \mathbf{P}[T < X \le T+S] = (e^{-\Lambda(T)} - e^{-\Lambda(T) - \frac{S}{a_i N_0}}) \\ p_3 &= \mathbf{P}[X > T+S] = e^{-\Lambda(T) - \frac{S}{a_i N_0}}, \end{aligned}$$

and we have

$$\mathbf{P}[\vartheta(X^J) = 1 \mid H_1] = \frac{1}{2}\mathbf{P}(L_1 = J \mid H_1) + \mathbf{P}(L_2 = 0, L_3 > 0 \mid H_1)$$

$$+ \sum_{(\ell_2,\ell_3):\ell_2>0} \mathbf{P}(L=\ell \mid H_1)\mathbf{P}(\mathrm{BF}_{12}(X^J)>0 \mid L=\ell, H_1)$$

with

$$\begin{aligned}\mathbf{P}[\mathrm{BF}_{12}(X^J)>0 \mid L=\ell, H_1] &= \mathbf{P}\left[\sum_{j:X^j\in[T,T+S]} X^j > \ell_2(\delta+T) - \ell_3 T \,\middle|\, L=\ell\right] \\ &= \mathbf{P}\left[\sum_{j:X^j\in[T,T+S]} X^j > \ell_2(\delta+T) - \ell_3 S \,\middle|\, T<X^j\le T+S\right] \\ &= \mathbf{P}\left[\sum_{j=1}^{\ell_2} X_*^j > \ell_2\delta - \ell_3 S \,\middle|\, X_*^j < S\right]\end{aligned}$$

for X_*^j independent exponential random variables with rate $(aN_0)^{-1}$. So letting $W^*(\ell_2) = \sum_{j=1}^{\ell_2} X_*^j$, the relevant probabilities involve the CDF of the sum of ℓ_2 many independent exponentials with rate $(aN_0)^{-1}$ truncated to the interval $[0,S]$, and we have

$$\begin{aligned}\mathbf{P}[\vartheta(X^J)=1 \mid H_1] = \frac{1}{2}\mathbf{P}(L_1=J \mid H_1) &+ \mathbf{P}(L_2=0, L_3>0 \mid H_1) \\ &+ \sum_{(\ell_2,\ell_3):\ell_2>0} \mathbf{P}(L=\ell \mid H_1)\mathbf{P}[W^*(\ell_2)>\ell_2\delta-\ell_3 S \mid H_1].\end{aligned}$$

It follows then that since $\mathbf{P}(L_1=J \mid H_1) = \mathbf{P}(L_1=J \mid H_2)$, the Bayes error rate can be written as

$$\begin{aligned}\mathbf{P}[\vartheta(X^J)=\vartheta] = \frac{1}{2}\mathbf{P}(L_1=J \mid H_1) &+ \frac{1}{2}\sum_{(\ell_2,\ell_3):\ell_2>0} \mathbf{P}(L=\ell \mid H_1)\mathbf{P}[W^*(\ell_2)>\ell_2\delta-\ell_3 S \mid H_1] \\ &+ \frac{1}{2}\sum_{(\ell_2,\ell_3):\ell_2>0} \mathbf{P}(L=\ell \mid H_2)\mathbf{P}[W^*(\ell_2)<\ell_2\delta-\ell_3 S \mid H_2].\end{aligned} \tag{C.1}$$

□

Appendix D. Proof of Theorem 5.1

Recall we are studying the case where $H_1 : N = N_1(t)$ and $H_2 : N = N_2(t)$ and

$$N_1(t) = \begin{cases} N(t) & 0 \le t \le T \\ aN_0 & T \le t \le T+S \\ N(t) & t > T+S \end{cases}$$

$$N_2(t) = \begin{cases} N(t) & 0 \le t \le T \\ bN_0 & T \le t \le T+S \\ N(t) & t > T+S \end{cases}$$

for $N(t)$ any bounded, strictly non-negative function.

1. Case 1: $0 < x_2 < x_1 < T$. In this case the likelihood under either H_1 or H_2 is the same

$$L(x_2, x_1 \mid N(t)) = \frac{3}{N(x_2)N(x_1)} e^{-2\Lambda(x_2)-\Lambda(x_1)}$$

and so

$$\log \mathrm{BF}_{12}(x) = 0.$$

2. Case 2: $0 < x_2 < T < x_1 < T+S$. In this case the likelihood under H_i is

$$L(x_2, x_1 \mid N(t)) = \frac{3}{N(x_2)} \frac{1}{a_i N_0} e^{-2\Lambda(x_2)-\Lambda(T)-\frac{x_1-T}{a_i N_0}}$$

so designating $a_1 = a, a_2 = b$ as before

$$\begin{aligned} \log \mathrm{BF}_{12}(x) &= \log \frac{b}{a} - \frac{x_1 - T}{aN_0} + \frac{x_1 - T}{bN_0} \\ &= \log \frac{b}{a} + \frac{(a-b)(x_1 - T)}{abN_0}. \end{aligned}$$

3. Case 3: $0 < x_2 < T < T + S < x_1$. In this case the likelihood under H_i is

$$L(x_2, x_1 \mid N(t)) = \frac{3}{N(x_2)} \frac{1}{N(x_1)} e^{-2\Lambda(x_2)-\Lambda(T)-\frac{S}{a_i N_0}-\Lambda(T+S,x_1)}$$

so

$$\log \mathrm{BF}_{12}(x) = \frac{S}{bN_0} - \frac{S}{aN_0} = \frac{(a-b)S}{abN_0}.$$

4. Case 4: $0 < T < x_2 < x_1 < T+S$. In this case the likelihood under H_i is

$$\begin{aligned} L(x_2, x_1 \mid N(t)) &= \frac{3}{a_i N_0} \frac{1}{a_i N_0} e^{-3\Lambda(T)-\frac{3(x_2-T)}{a_i N_0}-\frac{x_1-x_2}{a_i N_0}} \\ &= \frac{3}{a_i^2 N_0^2} e^{-3\Lambda(T)} e^{-\frac{2x_2+x_1-3T}{a_i N_0}} \end{aligned}$$

so

$$\begin{aligned} \log \mathrm{BF}_{12}(x) &= 2\log \frac{b}{a} - \frac{2x_2 + x_1 - 3T}{aN_0} + \frac{2x_2 + x_1 - 3T}{bN_0} \\ &= 2\log \frac{b}{a} + \frac{(a-b)(2x_2 + x_1 - 3T)}{abN_0} \end{aligned}$$

5. Case 5: $0 < T < x_2 < T + S < x_1$. In this case the likelihood under H_i is

$$\begin{aligned} L(x_2, x_1 \mid N(t)) &= \frac{3}{a_i N_0} \frac{1}{N(x_1)} e^{-2\Lambda(T) - \frac{2(x_2 - T)}{a_i N_0} - \Lambda(T) - \frac{S}{a_i N_0} - \Lambda(T+S, x_1)} \\ &= \frac{3}{a_i N_0} \frac{1}{N(x_1)} e^{-3\Lambda(T) - \Lambda(T+S, x_1)} e^{-\frac{2(x_2 - T) + S}{a_i N_0}} \end{aligned}$$

so

$$\begin{aligned} \log \mathrm{BF}_{12}(x) &= \log \frac{b}{a} - \frac{2(x_2 - T) + S}{a N_0} + \frac{2(x_2 - T) + S}{b N_0} \\ &= \log \frac{b}{a} + \frac{(a-b)(2(x_2 - T) + S)}{abN_0} \end{aligned}$$

6. Case 6: $0 < T < T + S < x_2 < x_1$. In this case the likelihood under H_i is

$$\begin{aligned} L(x_2, x_1 \mid N(t)) &= \frac{3}{N(x_2)} \frac{1}{N(x_1)} e^{-2\Lambda(T) - \frac{2S}{a_i N_0} - 2\Lambda(T+S, x_2) - \Lambda(T) - \frac{S}{a_i N_0} - \Lambda(T+S, x_1)} \\ &= \frac{3}{N(x_2)} \frac{1}{N(x_1)} e^{-3\Lambda(T) - 2\Lambda(T+S, x_2) - \Lambda(T+S, x_1)} e^{-\frac{3S}{a_i N_0}} \end{aligned}$$

so

$$\log \mathrm{BF}_{12}(x) = -\frac{3S}{aN_0} + \frac{3S}{bN_0} = \frac{3(a-b)S}{abN_0}$$

$$\log \mathrm{BF}_{12}(x) = \begin{cases} 0 & 0 < x_2 < x_1 < T \\ \log \frac{b}{a} + \frac{(a-b)(x_1 - T)}{abN_0} & 0 < x_2 < T < x_1 < T + S \\ \frac{(a-b)S}{abN_0} & 0 < x_2 < T < T + S < x_1 \\ 2 \log \frac{b}{a} + \frac{(a-b)(x_1 + 2x_2 - 3T)}{abN_0} & 0 < T < x_2 < x_1 < T + S \\ \log \frac{b}{a} + \frac{(a-b)(2x_2 - 2T + S)}{abN_0} & 0 < T < x_2 < T + S < x_1 \\ \frac{3(a-b)S}{abN_0} & 0 < T < T + S < x_2 < x_1 \end{cases}$$

We go line by line calculating the components of $\mathbf{P}[\vartheta(X) = \vartheta \mid H_1]$. Designate each of the six pieces of the expression by Q_j, $j = 1, 2, \ldots, 6$.

$$\begin{aligned} Q_1 = \frac{1}{2}\mathbf{P}[Y_1 < T] &= \frac{1}{2} \int_0^T \int_{x_2}^T \frac{3}{N(x_2) N(x_1)} e^{-2\Lambda(x_2) - \Lambda(x_1)} \\ &= \frac{1}{2} \int_0^T (e^{-\Lambda(x_2)} - e^{-\Lambda(T)}) \frac{3}{N(x_2)} e^{-2\Lambda(x_2)} dx_2 \\ &= \frac{1}{2} \int_0^T \frac{3}{N(x_2)} e^{-3\Lambda(x_2)} dx_2 - e^{-\Lambda(T)} \int_0^T \frac{3}{N(x_2)} e^{-2\Lambda(x_2)} dx_2 \end{aligned}$$

$$= \frac{1}{2}\left(1 - e^{-3\Lambda(T)} - e^{-\Lambda(T)}\frac{3}{2}\int_0^T \frac{2}{N_1(x_2)}e^{-2\Lambda(x_2)}dx_2\right)$$
$$= \frac{1}{2}\left(1 - e^{-3\Lambda(T)} - e^{-\Lambda(T)}\frac{3}{2}(1 - e^{-2\Lambda(T)})\right)$$
$$= \frac{1}{2} + \frac{1}{4}e^{-3\Lambda(T)} - \frac{3}{4}e^{-\Lambda(T)}$$

Now define

$$\delta = \frac{abN_0}{a-b}\log\frac{a}{b}$$

then we have

$$Q_2 = \int_0^T \int_T^{T+S} \frac{3}{N(x_2)N(x_1)}e^{-2\Lambda(x_2)-\Lambda(x_1)}\mathbf{1}\left\{\log\frac{b}{a} + \frac{(a-b)(x_1-T)}{abN_0} > 0\right\}dx_1dx_2$$
$$= \int_0^T \int_{T+(\delta\wedge S)}^{T+S} \frac{3}{N(x_2)N(x_1)}e^{-2\Lambda(x_2)-\Lambda(x_1)}dx_1dx_2$$
$$= (e^{-\Lambda(T)-\frac{\delta\wedge S}{aN_0}} - e^{-\Lambda(T)-\frac{S}{aN_0}})\frac{3}{2}\int_0^T \frac{2}{N(x_2)}e^{-2\Lambda(x_2)}dx_2$$
$$= (e^{-\frac{\delta\wedge S}{aN_0}} - e^{-\frac{S}{aN_0}})e^{-\Lambda(T)}\frac{3}{2}(1 - e^{-2\Lambda(T)})$$

For case 3

$$Q_3 = \mathbf{1}\{a > b\}\int_0^T \int_{T+S} \frac{3}{N(x_2)N(x_1)}e^{-2\Lambda(x_2)-\Lambda(x_1)}dx_1dx_2$$
$$= \mathbf{1}\{a > b\}e^{-\Lambda(T)-\frac{S}{aN_0}}\frac{3}{2}\int_0^T \frac{2}{N(x_2)}e^{-2\Lambda(x_2)}dx_2$$
$$= \frac{3}{2}\mathbf{1}\{a > b\}e^{-\Lambda(T)-\frac{S}{aN_0}}(1 - e^{-2\Lambda(T)})$$

Case 4

$$Q_4 = \int_T^{T+S} \int_{T+S}^{T+S} \frac{3}{N(x_2)N(x_1)}e^{-2\Lambda(x_2)-\Lambda(x_1)}\mathbf{1}\left\{x_1 > T + 2(T - x_2 + \delta)\right\}dx_1dx_2$$
$$= \int_T^{T+S} \int_{x_2}^{T+S} \frac{3}{a^2N_0^2}e^{-3\Lambda(T)}e^{-\frac{2x_2+x_1-3T}{aN_0}}\mathbf{1}\left\{x_1 > T + 2(T - x_2 + \delta)\right\}dx_1dx_2.$$

The inequalities

$$0 < T < x_2 < x_1 < T + S, \quad x_1 > T + 2(T - x_2 + \delta)$$

reduce to

$$\frac{2\delta}{3} < S < 2\delta, \quad \frac{1}{3}(3T + 2\delta) < x_1 < S + T, \quad \frac{1}{2}(3T - x_1 + 2\delta) < x_2 < x_1$$

or

$$S > 2\delta$$

and either

$$\frac{1}{3}(3T + 2\delta) < x_1 < T + 2\delta, \quad \frac{1}{2}(3T - x_1 + 2\delta) < x_2 < x_1$$

or

$$T + 2\delta < x_1 < S + T, \quad T < x_2 < x_1.$$

So then we can express Q_4 as

$$Q_4 = \begin{cases} 0 & S < \frac{2}{3}\delta \\ Q_{41} & \frac{2}{3}\delta < S < 2\delta \\ Q_{42} & S > 2\delta \end{cases}$$

where

$$\begin{aligned} Q_{41} &= \int_{\frac{1}{3}(3T+2\delta)}^{T+S} \int_{\frac{1}{2}(3T-x_1+2\delta)}^{x_1} \frac{3}{a^2 N_0^2} e^{-3\Lambda(T)} e^{-\frac{2x_2+x_1-3T}{aN_0}} dx_2 dx_1 \\ &= \frac{1}{2} e^{-3\Lambda(T)} \left(e^{-\frac{3S}{aN_0}} - e^{-\frac{2\delta}{aN_0}} \frac{N_0 - 3S + 2\delta}{aN_0} \right) \end{aligned}$$

and

$$\begin{aligned} Q_{42} &= \int_{\frac{1}{3}(3T+2\delta)}^{T+2\delta} \int_{\frac{1}{2}(3T-x_1+2\delta)}^{x_1} \frac{3}{a^2 N_0^2} e^{-3\Lambda(T)} e^{-\frac{2x_2+x_1-3T}{aN_0}} dx_2 dx_1 \\ &\quad + \int_{T+2\delta}^{T+S} \int_{x_1}^{x_1} \frac{3}{a^2 N_0^2} e^{-3\Lambda(T)} e^{-\frac{2x_2+x_1-3T}{aN_0}} dx_2 dx_1 \\ &= \frac{1}{2} e^{-3\Lambda(T)} \left(e^{-\frac{3\delta}{aN_0}} + e^{-\frac{2\delta}{aN_0}} \frac{4\delta - aN_0}{aN_0} \right) \\ &\quad + \frac{1}{2} e^{-3\Lambda(T)} \left(e^{-\frac{3S}{aN_0}} - 3e^{-\frac{S}{aN_0}} - e^{-\frac{6\delta}{aN_0}} + 3e^{-\frac{2\delta}{aN_0}} \right) \\ &= \left(\frac{1}{2} e^{-3\Lambda(T)} \left(e^{-\frac{2\delta}{aN_0}} \left(\frac{4\delta}{aN_0} + 2 \right) + e^{-\frac{3S}{aN_0}} - 3e^{-\frac{S}{aN_0}} \right) \right) \end{aligned}$$

And now for case 5

$$\begin{aligned} Q_5 &= \int_T^{T+S} \int_{T+S}^{x_1} f_1(x_1, x_2) \mathbf{1} \left\{ \log \frac{b}{a} + \frac{(a-b)(2(x_2 - T) + S)}{abN_0} > 0 \right\} dx_1 dx_2 \\ &= \int_T^{T+S} \int_{T+S}^{\infty} \frac{3}{N(x_2) N(x_1)} e^{-2\Lambda(x_2) - \Lambda(x_1)} \mathbf{1} \left\{ x_2 > T + \frac{\delta}{2} - \frac{S}{2} \right\} dx_1 dx_2 \\ &= \int_{T + \{0 \vee ((\frac{\delta}{2} - \frac{S}{2}) \wedge S)\}}^{T+S} \frac{3}{N(x_2)} e^{-2\Lambda(x_2)} dx_2 \int_{T+S}^{\infty} \frac{1}{N(x_1)} e^{-\Lambda(x_1)} dx_1 \end{aligned}$$

$$= \frac{3}{2} e^{-\Lambda(T)-\frac{S}{aN_0}} \int_{T+\{0\vee((\frac{\delta}{2}-\frac{S}{2})\wedge S)\}}^{T+S} \frac{2}{N(x_2)} e^{-2\Lambda(x_2)} dx_2$$
$$= \frac{3}{2} e^{-\Lambda(T)-\frac{S}{aN_0}} \left(e^{-2\Lambda(T+\{0\vee((\frac{\delta}{2}-\frac{S}{2})\wedge S)\})} - e^{-2\Lambda(T+S)}\right)$$
$$= \frac{3}{2} e^{-3\Lambda(T)-\frac{S}{aN_0}} \left(e^{-\frac{2\{0\vee((\frac{\delta}{2}-\frac{S}{2})\wedge S)\}}{aN_0}} - e^{-\frac{2S}{aN_0}}\right)$$

Finally case 6

$$Q_6 = \mathbf{1}\{a > b\} \int_{T+S}^{\infty} \int_{x_2}^{\infty} \frac{3}{N(x_2)} \frac{1}{N(x_1)} e^{-2\Lambda(x_2)-\Lambda(x_1)} dx_1 dx_2$$
$$= \mathbf{1}\{a > b\} \int_{T+S}^{\infty} \frac{3}{N(x_2)} e^{-3\Lambda(x_2)} dx_2$$
$$= \mathbf{1}\{a > b\} e^{-3\Lambda(T)-\frac{3S}{aN_0}}$$

Now we can get the other component fairly easily. We repeat the calculations conditioning on H_2

$$Q_1 = \frac{1}{2}\left(1 + \frac{1}{2} e^{-3\Lambda(T)} - \frac{3}{2} e^{-\Lambda(T)}\right)$$

case 2

$$Q_2 = \int_0^T \int_T^{T+S} \frac{3}{N(x_2)N(x_1)} e^{-2\Lambda(x_2)-\Lambda(x_1)} \mathbf{1}\left\{\log\frac{b}{a} + \frac{(a-b)(x_1-T)}{abN_0} < 0\right\} dx_1 dx_2$$
$$= \int_0^T \int_T^{T+(\delta\wedge S)} \frac{3}{N(x_2)N(x_1)} e^{-2\Lambda(x_2)-\Lambda(x_1)} dx_1 dx_2$$
$$= (e^{-\Lambda(T)} - e^{-\Lambda(T)-\frac{\delta\wedge S}{bN_0}}) \frac{3}{2} \int_0^T \frac{2}{N(x_2)} e^{-2\Lambda(x_2)} dx_2$$
$$= (1 - e^{-\frac{\delta\wedge S}{bN_0}}) e^{-\Lambda(T)} \frac{3}{2} (1 - e^{-2\Lambda(T)})$$

For case 3

$$Q_3 = \mathbf{1}\{b > a\} \int_0^T \int_{T+S}^{\infty} \frac{3}{N(x_2)N(x_1)} e^{-2\Lambda(x_2)-\Lambda(x_1)} dx_1 dx_2$$
$$= 0$$

Case 4

$$Q_4 = \int_T^{T+S} \int_{x_2}^{T+S} \frac{3}{N(x_2)N(x_1)} e^{-2\Lambda(x_2)-\Lambda(x_1)} \mathbf{1}\left\{x_1 < T + 2(T - x_2 + \delta)\right\} dx_1 dx_2$$
$$= \int_T^{T+S} \int_{x_2}^{T+S} \frac{3}{b^2 N_0^2} e^{-3\Lambda(T)} e^{-\frac{2x_2+x_1-3T}{bN_0}} \mathbf{1}\left\{x_1 < T + 2(T - x_2 + \delta)\right\} dx_1 dx_2.$$

The inequalities

$$0 \le T \le x_2 \le x_1 \le T+S, \quad x_1 \le T + 2(T - x_2 + \delta), \quad \delta > 0$$

reduce to

$$0 \le S \le \frac{2\delta}{3}, \quad T \le x_1 \le S+T, \quad T \le x_2 \le x_1, \quad \text{or}$$

$$\frac{2\delta}{3} \le S \le 2\delta, \quad \begin{cases} T < x_1 \le \frac{1}{3}(3T+2\delta), & T \le x_2 \le x_1 \text{ or} \\ \frac{1}{3}(3T+2\delta), \le x_1 \le T+S & T \le x_2 \le \frac{1}{2}(3T - x_1 + 2\delta) \end{cases}$$

or

$$S > 2\delta, \quad \begin{cases} T < x_1 < \frac{1}{3}(3T+2\delta), & T < x_2 \le x_1 \text{ or} \\ \frac{1}{3}(3T+2\delta) \le x_1 \le T+2\delta, & T < x_2 \le \frac{1}{2}(3T - x_1 + 2\delta) \end{cases}$$

so

$$Q_4 = \begin{cases} Q_{41} & 0 < S < \frac{2\delta}{3} \\ Q_{42} & \frac{2\delta}{3} \le S \le 2\delta \\ Q_{43} & S > 2\delta \end{cases},$$

where

$$Q_{41} = \frac{1}{2} e^{-3\Lambda(T)} \left(e^{-\frac{3S}{bN_0}} - 3e^{-\frac{S}{bN_0}} + 2 \right)$$

$$Q_{42} = \frac{1}{2} e^{-3\Lambda(T)} \left(\frac{e^{-\frac{2\delta}{bN_0}}(bN_0 + 2\delta - 3S)}{bN_0} - 3e^{-\frac{S}{bN_0}} + 2 \right)$$

$$Q_{43} = \frac{1}{2bN_0} e^{-3\Lambda(T)} e^{-\frac{2\delta+S}{bN_0}} \left(bN_0 \left(2e^{\frac{2\delta+S}{bN_0}} + e^{\frac{S}{bN_0}} - 3e^{\frac{T}{bN_0}} \right) + e^{\frac{S}{bN_0}}(-4\delta - 3S + 3T) \right)$$

Case 5

$$\begin{aligned} Q_5 &= \int_T^{T+S} \int_{T+S}^{\infty} f_{1,2}(x_1, x_2) \mathbf{1} \left\{ \log \frac{b}{a} + \frac{(a-b)(2(x_2 - T) + S)}{abN_0} < 0 \right\} dx_1 dx_2 \\ &= \int_T^{T+S} \int_{T+S}^{\infty} \frac{3}{N(x_2)N(x_1)} e^{-2\Lambda(x_2) - \Lambda(x_1)} \mathbf{1} \left\{ x_2 < T + \frac{\delta}{2} - \frac{S}{2} \right\} dx_1 dx_2 \\ &= \int_T^{T+\{0 \vee ((\frac{\delta}{2} - \frac{S}{2}) \wedge S)\}} \frac{3}{N(x_2)} e^{-2\Lambda(x_2)} dx_2 \int_{T+S}^{\infty} \frac{1}{N(x_1)} e^{-\Lambda(x_1)} dx_1 \\ &= \frac{3}{2} e^{-\Lambda(T) - \frac{S}{bN_0}} \int_T^{T+\{0 \vee ((\frac{\delta}{2} - \frac{S}{2}) \wedge S)\}} \frac{2}{N(x_2)} e^{-2\Lambda(x_2)} dx_2 \\ &= \frac{3}{2} e^{-\Lambda(T) - \frac{S}{bN_0}} (e^{-2\Lambda(T)} - e^{-2\Lambda(T + \{0 \vee ((\frac{\delta}{2} - \frac{S}{2}) \wedge S)\})}) \\ &= \frac{3}{2} e^{-3\Lambda(T) - \frac{S}{bN_0}} (1 - e^{-\frac{2\{0 \vee ((\frac{\delta}{2} - \frac{S}{2}) \wedge S)\}}{bN_0}}) \end{aligned}$$

Case 6

$$\begin{aligned} Q_6 &= \mathbf{1}\{b > a\} \int_{T+S}^{\infty} \int_{x_2}^{\infty} \frac{3}{N(x_2)} \frac{1}{N(x_1)} e^{-2\Lambda(x_2)-\Lambda(x_1)} dx_1 dx_2 \\ &= 0 \end{aligned}$$

Appendix E. Proof of theorem 6.1

Proof. Define $\Lambda_i(t) = \int_0^t \frac{1}{N_i(s)} ds$, we then have

$$\begin{aligned} \mathbf{P}[\vartheta(X) = 1 \mid H_1] &= \int_0^{\infty} \mathbf{1}\left\{ \left(\frac{1}{c} - 1\right) \Lambda_1(x) > \log \frac{1}{c} \right\} \frac{1}{N_1(x)} e^{-\int_0^x \frac{1}{N_1(t)} dt} dx \\ &= \mathbf{P}\left[X > \Lambda_1^{-1}\left(\frac{c}{1-c} \log \frac{1}{c} \right) \mid H_1 \right] \\ &= e^{-\Lambda_1\left(\Lambda_1^{-1}\left(\frac{c}{1-c}\log\frac{1}{c}\right)\right)} = e^{\frac{c}{1-c}\log c} \\ &= c^{\frac{c}{1-c}}, \end{aligned}$$

which implicitly assumed that $c < 1$. Similarly

$$\begin{aligned} \mathbf{P}[\vartheta(X) = 2 \mid H_2] &= \int_0^{\infty} \mathbf{1}\left\{ \Lambda_2(x) - \Lambda_1(x) < \log \frac{N_1(x)}{N_2(x)} \right\} \frac{1}{N_2(x)} e^{-\int_0^x \frac{1}{N_2(t)} dt} dx \\ &= \int_0^{\infty} \mathbf{1}\left\{ \Lambda_2(x)(c-1) > \log c \right\} \frac{1}{N_2(x)} e^{-\int_0^x \frac{1}{N_2(t)} dt} dx \\ &= \mathbf{P}\left[X < \Lambda_2^{-1}\left(\frac{1}{c-1} \log c \right) \mid H_2 \right] = 1 - e^{-\Lambda_2\left(\Lambda_2^{-1}\left(\frac{1}{c-1}\log c\right)\right)} \\ &= 1 - c^{\frac{1}{1-c}} \end{aligned}$$

so then

$$\mathbf{P}[\vartheta(X) = \vartheta] = \frac{1}{2} c^{\frac{c}{1-c}} + \frac{1}{2}\left(1 - c^{\frac{1}{1-c}}\right).$$

□

Appendix F. Proof of theorem 6.2

Proof. Define $\Lambda_i(t) = \int_0^t \frac{1}{N_i(s)} ds$ and notice that

$$\begin{aligned} \mathrm{BF}_{12} &= \prod_{j=1}^{J} \frac{\frac{1}{N_1(x^j)} e^{-\int_0^{x^j} \frac{1}{N_1(t)} dt}}{\frac{1}{N_2(x^j)} e^{-\int_0^{x^j} \frac{1}{N_2(t)} dt}} = \prod_{j=1}^{J} \frac{N_2(x^j) e^{-\int_0^{x^j} \frac{1}{N_1(t)} dt}}{N_1(x^j) e^{-\int_0^{x^j} \frac{1}{N_2(t)} dt}} \\ &= c^J e^{-\sum_{j=1}^J \Lambda_1(x^j) - \Lambda_2(x^j)} = c^J e^{-\sum_{j=1}^J \Lambda_1(x^j)\left(1-\frac{1}{c}\right)} \end{aligned}$$

$$\log \mathrm{BF}_{12} = J \log c - \left(1 - \frac{1}{c}\right) \sum_{j=1}^{J} \Lambda_1(x^j)$$

so then

$$\begin{aligned}
\mathbf{P}\left[\log \mathrm{BF}_{12} > 0 \mid H_1\right] &= \mathbf{P}\left[J \log c > \left(1 - \frac{1}{c}\right) \sum_{j=1}^{J} \Lambda_1(X^j) \mid H_1\right] \\
&= \mathbf{P}\left[\left(\frac{1}{c} - 1\right) \sum_{j=1}^{J} \Lambda_1(X^j) > J \log \frac{1}{c} \mid H_1\right] \\
&= \mathbf{P}\left[\sum_{j=1}^{J} \Lambda_1(X^j) > J \frac{c}{1-c} \log \frac{1}{c} \mid H_1\right].
\end{aligned}$$

Since

$$\mathbf{P}[\Lambda_1(X) > s \mid H_1] = \mathbf{P}[X > \Lambda_1^{-1}(s)] = e^{-s},$$

we have

$$\mathbf{P}\left[\log \mathrm{BF}_{12} > 0 \mid H_1\right] = \mathbf{P}\left[W > J \frac{c}{1-c} \log \frac{1}{c}\right],$$

where W is the sum of J independent unit rate exponentials, so $W \sim \mathrm{Gamma}(J, 1)$ and

$$\mathbf{P}\left[\log \mathrm{BF}_{12} > 0 \mid H_1\right] = 1 - \frac{1}{\Gamma(J)} \gamma\left(J, J \frac{c}{1-c} \log \frac{1}{c}\right),$$

where $\gamma(\alpha, \beta)$ is the lower incomplete Gamma function. Similar calculations give us that

$$\begin{aligned}
\mathbf{P}\left[\log \mathrm{BF}_{12} < 0 \mid H_2\right] &= \mathbf{P}\left[J \log c < (c-1) \sum_{j=1}^{J} \Lambda_2(X^j) \mid H_2\right] \\
&= \mathbf{P}\left[\sum_{i=1}^{J} \Lambda_2(X^j) < J \frac{1}{c-1} \log c \mid H_2\right] \\
&= \mathbf{P}\left[W < J \frac{1}{c-1} \log c\right] \\
&= \frac{1}{\Gamma(J)} \gamma\left(J, J \frac{1}{c-1} \log c\right)
\end{aligned}$$

giving us

$$\mathbf{P}[\vartheta(X^J) = \vartheta] = \frac{1}{2}\left(1 - \frac{1}{\Gamma(J)} \gamma\left(J, J \frac{c}{1-c} \log \frac{1}{c}\right) + \frac{1}{\Gamma(J)} \gamma\left(J, J \frac{1}{c-1} \log c\right)\right),$$

as claimed. □

Appendix G. Proof of theorem 7.1

Proof. The log Bayes factor under the stationary distribution is

$$\log \mathrm{BF}_{12}^{\pi} = 2\log(c) + \frac{(1-c)x}{Nc},$$

then

$$\begin{aligned}\mathbf{P}_{\pi}[\vartheta(X) = 1 \mid H_1] &= \mathbf{P}_{\pi}\left[X > \frac{2Nc}{1-c}\log\frac{1}{c} \mid H_1\right] \\ &= \int_{\frac{2Nc}{1-c}\log\frac{1}{c}}^{\infty} \frac{x}{N^2} e^{-\frac{x}{N}} dx \\ &= c^{\frac{2c}{1-c}} - \frac{2}{1-c} c^{\frac{1+c}{1-c}} \log(c),\end{aligned}$$

for $0 < c < 1$. Similarly

$$\begin{aligned}\mathbf{P}_{\pi}[\vartheta(X) = 2 \mid H_2] = \mathbf{P}_{\pi}\left[X < \frac{2Nc}{1-c}\log\frac{1}{c} \mid H_2\right] &= \int_0^{\frac{2Nc}{1-c}\log\frac{1}{c}} \frac{x}{c^2N^2} e^{-\frac{x}{cN}} dx \\ &= 1 - c^{\frac{2}{1-c}}\left(1 - \frac{2}{1-c}\log c\right)\end{aligned}$$

so then

$$\mathbf{P}_{\pi}[\vartheta(X) = \vartheta] = \frac{1}{2}c^{\frac{2c}{1-c}} + \frac{1}{2}\left(1 - c^{\frac{2}{1-c}}\right) - c^{\frac{1+c}{1-c}}\log c.$$

□

Appendix H. Proof of theorem 7.2

Proof. Consider J independent loci, then

$$\mathrm{BF}_{12}^{\pi} = \prod_{j=1}^{J} \frac{\frac{x^j}{N^2}e^{-\frac{x^j}{N}}}{\frac{x^j}{c^2N^2}e^{-\frac{x^j}{cN}}} = c^{2J} e^{\left(\frac{1-c}{cN}\right)\sum_{j=1}^{J} x^j}$$

$$\log \mathrm{BF}_{12}^{\pi} = 2J\log c + \left(\frac{1-c}{cN}\right)\sum_{j=1}^{J} x^j$$

Note that when the effective population size is constant N, the stationary density is Gamma with shape parameter $\alpha = 2$ and rate parameter $\beta = 1/N$, and the sum of J independent Gamma random variables with parameters α_j and $\beta = 1/N$, $j = 1, \ldots, J$ is Gamma with parameters $\alpha = \sum_{j=1}^{J} \alpha_j = 2J$ and $\beta = 1/N$, then

$$\mathbf{P}_{\pi}\left[\log \mathrm{BF}_{12}^{\pi} > 0 \mid H_1\right] = \mathbf{P}_{\pi}\left[\sum_{j=1}^{J} X^j > 2J\frac{cN}{1-c}\log\frac{1}{c} \mid H_1\right].$$

Then,

$$\mathbf{P}_\pi\left[\log \mathrm{BF}_{12}^\pi > 0 \mid H_1\right] = 1 - \frac{1}{\Gamma(2J)}\gamma\left(2J, 2J\frac{c}{1-c}\log\frac{1}{c}\right),$$

where $\gamma(\alpha,\beta)$ is the lower incomplete Gamma function. Similar calculations give us that and

$$\begin{aligned}\mathbf{P}_\pi\left[\log \mathrm{BF}_{12}^\pi < 0 \mid H_2\right] &= \mathbf{P}_\pi\left[\sum_{j=1}^{J} X^j < 2J\frac{c}{1-c}\log\frac{1}{c} \mid H_2\right] \\ &= \frac{1}{\Gamma(2J)}\gamma\left(2J, 2J\frac{1}{1-c}\log\frac{1}{c}\right),\end{aligned}$$

giving us

$$\mathbf{P}_\pi[\vartheta(X^J) = \vartheta] = \frac{1}{2}\left(1 - \frac{1}{\Gamma(2J)}\gamma\left(2J, 2J\frac{c}{1-c}\log\frac{1}{c}\right) + \frac{1}{\Gamma(2J)}\gamma\left(2J, 2J\frac{1}{c-1}\log c\right)\right),$$

as claimed. □

Appendix I. Derivation of Equation 9.2

Begin by expanding the square in (3.1)

$$\int_0^\infty \left(\frac{(1+J)^2}{(1+z)^2} - 2c\frac{1+J}{1+z} + c^2\right)\frac{c^J}{\Gamma(J)}z^{J-1}e^{-cz}dz$$

and perform the integration for each term separately. We have

$$\begin{aligned}\int_0^\infty \frac{(1+J)^2}{(1+z)^2}\frac{c^J}{\Gamma(J)}z^{J-1}e^{-cz}dz &= c^J(J+1)^2\Psi(J, J-1, c) \\ -2c\int_0^\infty \frac{1+J}{1+z}\frac{c^J}{\Gamma(J)}z^{J-1}e^{-cz}dz &= -2(J+1)c^{J+1}\Psi(J, J, c) \\ c^2\int_0^\infty \frac{c^J}{\Gamma(J)}z^{J-1}e^{-cz}dz &= c^2.\end{aligned}$$

So we get

$$\begin{aligned}&\int_0^\infty \left(\frac{1+J}{1+z} - c\right)^2 \frac{c^J}{\Gamma(J)}z^{J-1}e^{-cz}dz \\ &= c^J(J+1)^2\Psi(J, J-1, c) - 2(J+1)c^{J+1}\Psi(J, J, c) + c^2. \qquad \text{(I.1)}\end{aligned}$$

Appendix J. Proof of Theorem 8.1

We have

$$f_i(x) = \begin{cases} \frac{1}{N(x)} e^{-\int_0^x \frac{1}{N(t)} dt} & x < T \\ \frac{1}{a_i N_0} e^{-\int_0^T \frac{1}{N(t)} dt} e^{-\frac{x-T}{a_i N_0}} & T \le x < T+S \\ \frac{1}{N(x)} e^{-\int_0^T \frac{1}{N(t)} dt} e^{-\frac{S}{a_i N_0}} e^{-\int_{T+S}^x \frac{1}{N(t)} dt} & T+S \le x \end{cases} \tag{J.1}$$

where $f_i(x)$ is the density of a single coalescent time under H_i. Define

$$\Delta_{12}(x) \equiv (\sqrt{f_1}(x) - \sqrt{f_2}(x))^2.$$

So now we calculate

$$\int (\sqrt{f_1(x)} - \sqrt{f_2(x)})^2 dx = \int_0^T \Delta_{12}(x)dx + \int_T^{T+S} \Delta_{12}(x)dx + \int_{T+S}^\infty \Delta_{12}(x)dx$$

clearly the first term on the right is zero so

$$\int \Delta_{12}(x)dx = \int_T^{T+S} \Delta_{12}(x)dx + \int_{T+S}^\infty \Delta_{12}(x)dx.$$

Observe

$$\begin{aligned} \int_T^{T+S} \Delta_{12}(x)dx &= \int_T^{T+S} \left(\frac{1}{\sqrt{aN_0}} e^{-\frac{1}{2}\int_0^T \frac{1}{N(t)}dt} e^{-\frac{1}{2}\frac{x-T}{aN_0}} - \frac{1}{\sqrt{bN_0}} e^{-\frac{1}{2}\int_0^T \frac{1}{N(t)}dt} e^{-\frac{1}{2}\frac{x-T}{bN_0}} \right)^2 dx \\ &= e^{-\int_0^T \frac{1}{N(t)}dt} \int_T^{T+S} \left(\frac{1}{\sqrt{aN_0}} e^{-\frac{1}{2}\frac{x-T}{aN_0}} - \frac{1}{\sqrt{bN_0}} e^{-\frac{1}{2}\frac{x-T}{bN_0}} \right)^2 dx, \end{aligned}$$

then

$$e^{\int_0^T \frac{1}{N(t)}dt} \int_T^{T+S} \Delta_{12}(x)dx = 2 - e^{-\frac{S}{aN_0}} - e^{-\frac{S}{bN_0}} - \frac{4b(1 - e^{-\frac{(a+b)S}{2abN_0}})\sqrt{a}}{(a+b)\sqrt{b}}, \tag{J.2}$$

and now

$$\begin{aligned} \int_{T+S}^\infty \Delta_{12}(x)dx &= \int_{T+S}^\infty \left(\frac{1}{\sqrt{N(x)}} e^{-\frac{1}{2}\int_0^T \frac{1}{N(t)}dt} e^{-\frac{1}{2}\frac{S}{aN_0}} e^{-\frac{1}{2}\int_{T+S}^x \frac{1}{N(t)}dt} \right. \\ &\quad \left. \frac{1}{\sqrt{N(x)}} e^{-\frac{1}{2}\int_0^T \frac{1}{N(t)}dt} e^{-\frac{1}{2}\frac{S}{bN_0}} e^{-\frac{1}{2}\int_{T+S}^x \frac{1}{N(t)}dt} \right)^2 dx \\ &= \int \left(e^{-\frac{1}{2}\frac{S}{aN_0}} - e^{-\frac{1}{2}\frac{S}{bN_0}} \right)^2 \left(\frac{1}{\sqrt{N(x)}} e^{-\frac{1}{2}\int_0^T \frac{1}{N(t)}dt} e^{-\frac{1}{2}\int_{T+S}^x \frac{1}{N(t)}dt} \right)^2 dx \\ &= \left(e^{-\frac{1}{2}\frac{S}{aN_0}} - e^{-\frac{1}{2}\frac{S}{bN_0}} \right)^2 e^{-\int_0^T \frac{1}{N(t)}dt} \int_{T+S}^\infty \frac{1}{N(x)} e^{-\int_{T+S}^x \frac{1}{N(t)}dt} dx \end{aligned}$$

$$= \left(e^{-\frac{1}{2}\frac{S}{aN_0}} - e^{-\frac{1}{2}\frac{S}{bN_0}}\right)^2 e^{-\int_0^T \frac{1}{N(t)}dt} \left(-e^{-\int_{T+S}^{x} \frac{1}{N(t)}dt}\Big|_{T+S}^{\infty}\right)$$
$$= \left(e^{-\frac{1}{2}\frac{S}{aN_0}} - e^{-\frac{1}{2}\frac{S}{bN_0}}\right)^2 e^{-\int_0^T \frac{1}{N(t)}dt}$$

and adding this to (J.2)

$$H^2(f_1, f_2) = \int \Delta_{12}(x)dx = e^{-\int_0^T \frac{1}{N(t)}dt}\left(1 - e^{-\frac{(a+b)S}{2abN}}\right)\frac{(a+b-2\sqrt{ab})}{a+b}$$
$$= e^{-\int_0^T \frac{1}{N(t)}dt}\left(1 - e^{-\frac{(a+b)S}{2abN_0}}\right)\frac{(\sqrt{a}-\sqrt{b})^2}{a+b}, \qquad \text{(J.3)}$$

which is the same as the last displayed equation on Kim et al. (2015, p 11).

Appendix K. Bounds on total variation between product measures

Let P^J, Q^J be product measures on a space $\mathbf{X} = \bigtimes_{j=1}^J \mathbf{X}_j$ of dimension J. Suppose P^J, Q^J are absolutely continuous with respect to some dominating measure ν on $\mathbf{X}$, and write the densities with respect to ν as $p^J(x) = \frac{dP^J}{d\nu}(x)$, $q^J(x) = \frac{dQ^J}{d\nu}(x)$, respectively. Notice we can write $p^J(x) = \prod_{j=1}^J p(x_j)$ for some density on $\mathbf{X}_1$, and similarly for q^J. Then the total variation distance can be expressed as

$$\begin{aligned}
\mathrm{d_{TV}}(P^J, Q^J) &= \int_{\mathbf{X}} |p^J(x) - q^J(x)|dx \\
&= \int_{\mathbf{X}} \left|\prod_{j=1}^J p(x_j) - \prod_{j=1}^J q(x_j)\right| dx \\
&= \int_{\mathbf{X}} \left|p(x_1)\prod_{j=2}^J p(x_j) - q(x_1)\prod_{j=2}^J p(x_j) + q(x_1)\prod_{j=2}^J p(x_j) - q(x_1)\prod_{j=2}^J q(x_j)\right| dx \\
&\leq \int_{\mathbf{X}} |p(x_1) - q(x_1)|\prod_{j=2}^J p(x_j)dx + \int_{\mathbf{X}} q(x_1)\left|\prod_{j=2}^J p(x_j) - \prod_{j=2}^J q(x_j)\right| dx \\
&= \int_{\mathbf{X}_1} |p(x_1) - q(x_1)|dx_1 + \int_{\mathbf{X}_{-1}} \left|\prod_{j=2}^J p(x_j) - \prod_{j=2}^J q(x_j)\right| dx_{-1} \\
&= \mathrm{d_{TV}}(P, Q) + \int_{\mathbf{X}_{-1}} \left|\prod_{j=2}^J p(x_j) - \prod_{j=2}^J q(x_j)\right| dx_{-1}
\end{aligned}$$

where $\mathbf{X}_{-1}$ denotes the $(J-1)$ dimensional subspace corresponding to the coordinates $(x_2, x_3, \ldots, x_J)$. Now an obvious inductive argument gives

$$\mathrm{d_{TV}}(P^J, Q^J) \leq J\mathrm{d_{TV}}(P, Q).$$

To obtain the other inequality, use the general double-sided bound for any measures μ_1, μ_2

$$\mathrm{H}^2(\mu_1, \mu_2) \leq \mathrm{d_{TV}}(\mu_1, \mu_2) \leq \sqrt{2}\mathrm{H}(\mu_1, \mu_2). \tag{K.1}$$

Using the inequality (8.1)

$$\mathrm{d_{TV}}(P^J, Q^J) \leq \sqrt{2}\sqrt{J}\mathrm{H}(P, Q)$$

and now using the lower bound in (K.1)

$$\mathrm{H}^2(P, Q) \leq \mathrm{d_{TV}}(P, Q) \Longrightarrow \mathrm{H}(P, Q) \leq \sqrt{\mathrm{d_{TV}}(P, Q)}$$

so

$$\mathrm{d_{TV}}(P^J, Q^J) \leq \sqrt{2J}\sqrt{\mathrm{d_{TV}}(P, Q)}.$$