Coalescence in the diffusion limit of a Bienaymé-Galton-Watson branching process

Conrad J. Burden*a, Albert C. Soewongsono*a

*aMathematical Sciences Institute, Australian National University, Canberra, Australia

Abstract

We consider the problem of estimating the elapsed time since the most recent common ancestor of a finite random sample drawn from a population which has evolved through a Bienaymé-Galton-Watson branching process. More specifically, we are interested in the diffusion limit appropriate to a super-critical process in the near-critical limit evolving over a large number of time steps. Our approach differs from earlier analyses in that we assume the only known information is the mean and variance of the number of offspring per parent, the observed total population size at the time of sampling, and the size of the sample. The main result is a joint likelihood surface defined as a function of two parameters: (1) the population size at the time of the most recent common ancestor, and (2) the time elapsed since the existence of the most recent common ancestor.

Keywords: Coalescent, Diffusion process, Branching process, Most recent common ancestor

1. Introduction

Suppose one observes the current size of a population which has been evolving through a Bienaymé-Galton-Watson (BGW) branching process since some unknown time in the past. Is it possible to infer the time that has elapsed since the most recent common ancestor (MRCA) of the current population, and of a random sample taken from the current population?

This paper is a continuation of an earlier paper (Burden and Simon, 2016) which gives a partial answer to this question in the form of a joint estimate.

Email addresses: conrad.burden@anu.edu.au (Conrad J. Burden), albert.soewongsono@anu.edu.au (Albert C. Soewongsono)
of the elapsed time since coalescence to the MRCA and the corresponding size of the entire population at the time of the MRCA. Our new results extend the previous calculation to the MRCA of finite random samples of the population, and to obtaining a likelihood surface in the joint parameter space of the time since coalescence and the population size at the time of coalescence.

There have been many studies of coalescence in BGW process [Lambert et al., 2013; Pardoux, 2016; Grosjean and Huillet, 2018], specifically for critical (Durrett, 1978; Athreya, 2012) sub-critical (Lambert, 2003; Le, 2014) and supercritical processes. In the current paper we concentrate on supercritical processes in the near-critical limit, or, equivalently, the continuous-time diffusion limit. Closest to our treatment is that of O’Connell (1995), who derives a formula (with a minor correction in Kimmel and Axelrod (2015)) for the coalescent time for sample of size $n = 2$ as a fraction of the time since initialisation of the BGW process with a known initial population size. Numerical simulation of an ensemble of BGW processes initiated from a single founder by Cyran and Kimmel (2010) is consistent with this formula. The ultimate aim of Cyran and Kimmel’s paper was to compare the effectiveness of various population genetics models in dating the time since mitochondrial Eve (mtE). More recently O’Connell’s formula has been generalised to samples of size $n > 2$ by Harris et al. (2019).

Our approach differs from previous treatments in that we initialise the BGW process at the (a priori unknown) time of coalescence of a random sample of the current population, treat the time since coalescence and initial population size as parameters to be estimated, and determine a likelihood surface in the two-dimensional parameter space defined by these two parameters. Central to our approach is the Feller (1951a) solution to the forward Kolmogorov equation defining the diffusion limit of a BGW process. This approach is closer in spirit to the intention of Cyran and Kimmel’s aim of determining the time since mtE without prior knowledge of the time at which the process was initiated from a single founding individual. It also enables a direct comparison with the analogous solution by Slatkin and Hudson (1991) to the problem of determining the distribution of the time since coalescence of a sample of size 2 in a Wright-Fisher (WF) population with exponential growth.

The layout of this paper is as follows. The diffusion limit of a BGW process, Feller’s solution for a specified initial population, and notational conventions employed in subsequent sections are summarised in Section 2.
Likelihood contours for the event that a sample of the current population is descended from a single ancestor and for the event that the BGW process undershoots the currently observed population are derived in Section 3. These likelihoods are functions of two scaled parameters; (1) the population size at the time of initialisation of the process, denoted $\kappa_0$, and (2) the time elapsed from initialisation of the process to observation of the current population, denoted $s$. This section is a generalisation of analogous derivations in (Burden and Simon, 2016), which concentrated on the entire population as opposed to a finite random sample. In Section 4 we derive from these contours a likelihood surface in the ($s, \kappa_0$)-plane for starting co-ordinates corresponding to the MRCA of an observed finite sample, given the observed current size of the population. We also derive a marginal likelihood for the time $s$ since the MRCA, and compare with the probability distributions of Slatkin and Hudson (1991) for a sample of size $n = 2$. In Section 5 we obtain an asymptotic form for the likelihood surface in the limit of large observed current population sizes. This enables a comparison of our analytic formulae with a numerical simulation, which is also in Section 5. Our results are discussed and conclusions drawn in Section 6. Two technical appendices are devoted to details of calculations arising in Section 4.

2. Definitions and Notation

Consider Bienaymé-Galton-Watson (BGW) branching process consisting of a population of $M(t)$ haploid individuals reproducing in discrete, non-overlapping generations $t = 0, 1, 2, \ldots$, with initial population size $M(0) = m_0$. It will prove convenient to subdivide the population at any generation into two allele types with population sizes $Y_1(t)$ and $Y_2(t)$, so that $Y_1(t) + Y_2(t) = M(t)$. The two alleles are assumed to be neutral with respect to selection and no mutation between types is included in the model in the current paper. Individuals of each type reproduce according to an independent BGW process whereby the number of offspring per individual of allele type $i$ is given by a set of identically and independently distributed (i.i.d.) random variables $(S^{(i)}_\alpha, \alpha = 1, \ldots, Y_i(t))$, whose common distribution is denoted by a generic non-negative integer valued random variable $S$ with finite moments to all higher orders and with mean and variance

$$\mathbb{E}(S) = \lambda, \quad \text{Var}(S) = \sigma^2.$$ (1)
Thus
\[ Y_i(t + 1) = \sum_{\alpha=1}^{Y_i(t)} S^{(i)}_\alpha, \quad i = 1, 2. \] (2)

If the \( Y_i(0) \) are mutually independent, then \( Y_i(t) \) are mutually independent at all subsequent times \( t \).

We are interested in the supercritical case of the diffusion limit studied by Feller (1951a) in which the initial population \( m_0 \) becomes infinite while the growth rate simultaneously approaches the critical value \( \lambda \to 1 \) in such a way that the parameter
\[ \kappa_0 = \frac{2m_0 \log \lambda}{\sigma^2}, \] (3)
remains fixed. As argued in Burden and Simon (2016), this parameter has a straightforward physical meaning: Suppose the initial population is divided into two subpopulations of roughly equal size. If the initial population size \( m_0 \) is such that \( \kappa_0 >> 1 \), then with high probability descendant lineages of both subpopulations will survive as \( t \to \infty \). On the other hand, if \( \kappa_0 << 1 \) and the population does not go extinct, the entire surviving population as \( t \to \infty \) will, with high probability, be descended from an individual in one of the two subpopulations. Thus \( \kappa_0 \approx 1 \) sets a scale for inferring coalescent events.

With this in mind, we introduce the following, slightly unconventional, continuum quantities adapted from Burden and Simon (2016). Define a continuum time \( s \) and continuous random population sizes \( K(s) \), \( K_1(s) \) and \( K_2(s) \) by
\[ s = t \log \lambda, \quad K(s) = \frac{2M(t) \log \lambda}{\sigma^2}, \quad K_i(s) = \frac{2Y_i(t) \log \lambda}{\sigma^2}, \quad i = 1, 2. \] (4)
The corresponding forward Kolmogorov equation for each of the densities \( f_{K_i}(\kappa_i; s) \) is
\[ \frac{\partial f_{K_i}}{\partial s} = -\frac{\partial (\kappa_i f_{K_i})}{\partial \kappa_i} + \frac{\partial^2 (\kappa_i f_{K_i})}{\partial \kappa_i^2}, \] (5)
with a similar equation for the probability density \( f_K(\kappa; s) \) of the total population. Note that with our scaling choice all parameters of the problem occur in the initial conditions and not in the differential equation.

The initial conditions are
\[ K(0) = \kappa_0, \quad K_1(0) = x_0 \kappa_0, \quad K_2(0) = (1 - x_0) \kappa_0, \quad 0 \leq x_0 \leq 1, \] (6)
or equivalently

\[ f_K(\kappa; 0|\kappa_0) = \delta(\kappa - \kappa_0), \]
\[ f_{K_1}(\kappa_1; 0|\kappa_0, x_0) = \delta(\kappa_1 - x_0\kappa_0), \]
\[ f_{K_2}(\kappa_2; 0|\kappa_0, x_0) = \delta(\kappa_2 - (1 - x_0)\kappa_0), \]  

(7)

where \( x_0 = Y_1(0)/m_0 \) is the proportion of the initial population which is of allele type-1.

This problem was first solved, albeit with a different continuum scaling, by Feller (1951a,b) by means of a Laplace transform. A summary of the method of solution is given in Section 4 of Burden and Wei (2018) for the forward Kolmogorov equation corresponding to the random variable \( Z(s) = e^{-s}K(s)/\kappa_0 \). The density for \( K(s) \) is easily obtained either from that of \( Z(s) \) by a transformation or by setting parameters appropriately in the solution stated on p235 of Feller (1951a). The resulting solution is

\[ f_K(\kappa; s|\kappa_0) = \delta(\kappa) \exp \left( -\frac{\kappa_0}{1 - e^{-s}} \right) + \]
\[ \frac{1}{e^s - 1} \left( \frac{\kappa_0 e^s}{\kappa} \right)^\frac{3}{2} \exp \left\{ -\frac{\kappa_0 e^s + \kappa}{e^s - 1} \right\} I_1 \left( \frac{2(\kappa_0 e^s)^\frac{1}{2}}{e^s - 1} \right), \]  

(8)

where \( I_1(\cdot) \) is a modified Bessel function. In the first term the coefficient of the delta-function is the probability of the population becoming extinct up to time \( s \) (See, for instance, p206 of Bailey, 1964). This solution can be written in terms of a 1-parameter family of functions\(^1\)

\[ f_{\text{Feller}}(z; \xi) = \delta(z)e^{-\xi} + \xi z^{-\frac{1}{2}} e^{-\xi(1+z)} I_1 \left( 2\xi z^\frac{1}{2} \right), \]  

(9)

where \( z, \xi \geq 0 \), as

\[ f_K(\kappa; s|\kappa_0) = \frac{1}{\kappa_0 e^s} f_{\text{Feller}} \left( \frac{\kappa}{\kappa_0 e^s}; \frac{\kappa_0}{1 - e^{-s}} \right). \]  

(10)

The densities corresponding to the individual allele populations \( K_1 \) and

\(^1\)Note that Eq. (31) of Burden and Wei (2018) defining the function \( f_{\text{Feller}}(\cdot) \) contains an error: the argument of the modified Bessel function should be \( 2\kappa_0 z^{\frac{1}{2}} \), not \( 2\kappa_0 z^{-\frac{1}{2}} \).
\( K_2 \) follow immediately from Eq. (8) as

\[
f_{K_1}(\kappa_1; s|\kappa_0, x_0) = f_K(\kappa_1; s|x_0\kappa_0) \\
= \delta(\kappa_1) \exp \left( -\frac{x_0\kappa_0}{1 - e^{-s}} \right) + \\
\frac{1}{e^s - 1} \left( \frac{x_0\kappa_0e^s}{\kappa_1} \right)^{\frac{1}{2}} \exp \left\{ -\frac{x_0\kappa_0e^s + \kappa_1}{e^s - 1} \right\} I_1 \left( \frac{2(\kappa_1 x_0\kappa_0 e^s)^{\frac{1}{2}}}{e^s - 1} \right),
\]

(11)

and

\[
f_{K_2}(\kappa_2; s|\kappa_0, x_0) = f_K(\kappa_2; s|(1 - x_0)\kappa_0) \\
= \delta(\kappa_2) \exp \left( -\frac{(1 - x_0)\kappa_0}{1 - e^{-s}} \right) + \\
\frac{1}{e^s - 1} \left( \frac{\kappa_0(1 - x_0)e^s}{\kappa_2} \right)^{\frac{1}{2}} \exp \left\{ -\frac{\kappa_0(1 - x_0)e^s + \kappa_2}{e^s - 1} \right\} \times \\
I_1 \left( \frac{2(\kappa_2\kappa_0(1 - x_0)e^s)^{\frac{1}{2}}}{e^s - 1} \right),
\]

(12)

respectively.

3. Coalescence of a sample of \( n \) individuals

Our aim is to estimate the time of the most recent common ancestor (MRCA) of a sample of \( n \) individuals chosen randomly and uniformly from a population of scaled size \( \kappa = (2m\log \lambda)/\sigma^2 \), where \( m \) is the population size at the “current” time \( s = t \log \lambda \) since the coalescent event, which we set to be at time 0. The population is assumed to have evolved via a BGW process with constant values of the parameters \( \lambda \) and \( \sigma^2 \), which are assumed to be given. Thus \( \kappa \) is to be thought of as a known input parameter, while \( s \) is an unknown parameter which seek to estimate. In the process, the scaled population size \( \kappa_0 = 2m_0 \log \lambda/\sigma^2 \) at the time of the MRCA will also be estimated. The analysis follows the same reasoning as earlier work of Burden and Simon (2016) on estimating the time of the MRCA of an entire population, known as mtE in the case of the existing human population.
As a preliminary step, consider the population to consist of two allele types, as in the previous section, and consider joint distribution of the two random variables

\begin{align*}
K(s) &= K_1(s) + K_2(s), \\
X(s) &= \frac{K_1(s)}{K_1(s) + K_2(s)}.
\end{align*}

(13)

Here \(K(s)\) is the total population size and \(X(s)\) is the fraction of the population which is of the first allele type at time \(s\). Since \(K_1(s)\) and \(K_2(s)\) are independent, the joint density corresponding to the initial conditions \(K(0) = \kappa_0, X(0) = x_0\) is

\begin{equation}
 f_{K,X}(\kappa, x; \kappa_0, x_0) = \kappa f_{K_1}(x\kappa; s|\kappa_0, x_0) f_{K_2}((1-x)\kappa; s|\kappa_0, x_0),
\end{equation}

(14)
defined over the range \(0 \leq \kappa < \infty, 0 \leq x \leq 1, 0 \leq s < \infty\). The factor of \(\kappa\) arises from the Jacobean of the transformation Eq. (13).

Define \(E_n(\kappa_0, s)\) to be the event that a random uniform sample of \(n\) individuals taken at time \(s\) are descended from a single individual in the original population at time 0, given that the initial scaled population size was \(\kappa_0\). Descendants of this individual are defined to be of allele type-1, while members of the population who are not descended from this individual are type-2. We claim that the probability of the joint event that

1. \(E_n(\kappa_0, s)\) happens; and
2. the population size at time \(s\) is in the range \([\kappa, \kappa + d\kappa)\)

is

\begin{equation}
 \text{Prob} \left( E_n(\kappa_0, s), K(s) \in [\kappa, \kappa + d\kappa)|K(0) = \kappa_0 \right) = m_0 \int_0^1 x^n f_{K,X} \left( \kappa, x; s|\kappa_0, \frac{1}{m_0} \right) dx d\kappa,
\end{equation}

(15)

where the function \(f_{K,X}\) is defined by Eq. (14). The factor \(m_0\) accounts for the \(m_0\) possibilities for the common ancestor of the sample and the integral ranges over all possible fractions \(x\) of the final population that can be descended from that ancestor. From Eqs. (11), (12) and (14), and using the property of the
modified Bessel function that $I_1(2w) = w + O(w^2)$ as $w \to 0$, we obtain
\begin{equation}
 f_{K,X}(\kappa,x;|\kappa_0, \frac{1}{m_0}) =
 \left\{ \delta(x) \left( 1 - \frac{\kappa_0}{m_0(1 - e^{-s})} \right) + \frac{1}{m_0} \frac{\kappa_0 e^s}{(e^s - 1)^2} \exp \left( \frac{-x\kappa}{e^s - 1} \right) + O \left( \frac{1}{m_0^2} \right) \right\} \times
 \left\{ \frac{1}{\kappa} \delta(1 - x) \exp \left( \frac{-\kappa_0}{1 - e^{-s}} \right) + \frac{1}{e^s - 1} \left( \frac{\kappa_0 e^s}{(1 - x)\kappa} \right)^{\frac{1}{2}} \exp \left( \frac{-\kappa_0 e^s + (1 - x)\kappa}{e^s - 1} \right) \times I_1 \left( \frac{2(1 - x)\kappa_0 e^s}{e^s - 1} \right) + O \left( \frac{1}{m_0} \right) \right\}. \tag{16}
\end{equation}

Substituting back into Eq. (15), noting that the $\delta(x)$ term in the first factor of Eq. (16) does not contribute because of the factor $x^n$ in the integrand, integrating out the $\delta(1 - x)$ in the second factor, and taking the limit $m_0 \to \infty$ gives, after some straightforward algebra,
\begin{equation}
 \text{Prob} (\mathcal{E}_n(\kappa_0, s), K(s) \in [\kappa, \kappa + d\kappa)|K(0) = \kappa_0) \nonumber
 = \frac{\kappa_0 e^s}{(e^s - 1)^2} \exp \left( -\frac{\kappa_0 + \kappa e^{-s}}{1 - e^{-s}} \right) \left( 1 + w \int_0^1 \frac{x^n}{(1 - x)^{\frac{1}{2}}} I_1 \left( 2w(1 - x)^{\frac{1}{2}} \right) dx \right) d\kappa, \tag{17}
\end{equation}
where
\begin{equation}
 w = \frac{(\kappa_0 e^s)^{\frac{1}{2}}}{e^s - 1} = \frac{(\kappa_0)^{\frac{1}{2}}}{2 \sinh \frac{1}{2}s}. \tag{18}
\end{equation}

The integral can be evaluated with the aid of Wolfram alpha [Wolfram Research, Inc. 2019] as
\begin{equation}
 \int_0^1 \frac{x^n}{(1 - x)^{\frac{1}{2}}} I_1 \left( 2w(1 - x)^{\frac{1}{2}} \right) dx = 2 \int_0^1 (1 - z^2)^n I_1(2wz) dz
 = \frac{1}{w} \left( _0F_1(; n + 1; w^2) - 1 \right)
 = \frac{1}{w} \left( n! I_n(2w) \frac{w^n}{w^n} - 1 \right), \tag{19}
\end{equation}
\[
\text{Prob} \left( E_n(\kappa_0, s), K(s) \in [\kappa, \kappa + d\kappa] | K(0) = \kappa_0 \right) = \frac{\kappa_0 e^s}{(e^s - 1)^2} \frac{n!}{w^n} \frac{I_n(2w)}{I_1(2w)} \exp \left( -\frac{\kappa_0 + \kappa e^{-s}}{1 - e^{-s}} \right) d\kappa. \tag{20}
\]

From Eq. (20) we can calculate the probability, given that the process starts with a scaled population \( \kappa_0 \) and arrives at a scaled population \( \kappa \) at a time \( s \) later, that the initial population contains an individual who is the common ancestor of a sample of size \( n \) in the final population. This probability is

\[
u_n(s, \kappa_0 | \kappa) := \frac{\text{Prob} \left( E_n(\kappa_0, s), K(s) \in [\kappa, \kappa + d\kappa] | K(0) = \kappa_0 \right)}{\text{Prob} \left( K(s) \in [\kappa, \kappa + d\kappa] | K(0) = \kappa_0 \right)}. \tag{21}
\]

The denominator, obtained from the continuous part of Eq. (8), is

\[
\text{Prob} \left( K(s) \in [\kappa, \kappa + d\kappa] | K(0) = \kappa_0 \right) = \frac{1}{e^s - 1} \left( \frac{\kappa_0 e^s}{\kappa} \right)^{\frac{1}{2}} \exp \left\{ -\frac{\kappa_0 + \kappa e^{-s}}{1 - e^{-s}} \right\} I_1(2w) d\kappa. \tag{22}
\]

Dividing Eq. (20) by Eq. (22) gives finally

\[
u_n(s, \kappa_0 | \kappa) = \frac{n!}{w^{n-1}} \frac{I_n(2w)}{I_1(2w)}. \tag{23}
\]

From the ascending series for modified Bessel function (see p375 of Abramowitz and Stegun [1965]) we have \( I_n(2w) = (w^n/n!) (1 + O(1/n)) \) as \( n \to \infty \). Hence we recover the result of Burden and Simon [2016] for the probability that the entire population at time \( s \) is descended from a single individual given the initial and final population sizes are \( \kappa_0 \) and \( \kappa \) respectively, namely

\[
u_\infty(s, \kappa_0 | \kappa) = \frac{w}{I_1(2w)}. \tag{24}
\]

Figures 1(a) and 1(b) show a contour map of the density in Eq. (23) in the \( s-\kappa_0 \) plane for a random sample of size \( n = 2 \) taken from an observed final scaled population of size \( \kappa = 3 \) and \( \kappa = 2000 \) respectively. The map grades from a probability near 0 on the left edge of the plot corresponding
Figure 1: Contours (black lines) of the probability $u_n(s, \kappa_0|\kappa)$ in Eq. (23) that a random sample of $n = 2$ individuals in the “current” population is descended from a single individual in the “initial” population. Axes refer to the scaled time $s$ at which the current population is measured and the scaled initial population at time zero, $\kappa_0$. Superimposed in blue are contours of the likelihood function $v(s, \kappa_0|\kappa)$ defined in Eq. (25) as probability that the population at time $s$ will not exceed the observed population $\kappa$ for a given starting position in the plane. The red curves are the contours of the likelihood surface $L(s, \kappa_0|\kappa)$ described in §4.

to recent initialisation times to a probability near 1 at earlier times. If the BGW process were initiated at a time $s$ in the past with a population of size $\kappa_0$, the contour passing through the point $(s, \kappa_0)$ gives the probability that the ancestral coalescence of the sample occurred after the initialisation of the process. For larger values of $\kappa$ the coalescent event appears to be fixed within a relatively narrow range.

Superimposed in blue in Figure 1 are contours of the likelihood function that the population at time $s$ will not exceed the observed population $\kappa$ for a given starting population $\kappa_0$,

$$v(s, \kappa_0|\kappa) := \text{Prob} \left( K(s) \leq \kappa | K(0) = \kappa_0 \right) = \int_0^\kappa f_K(\eta; s|\kappa_0) d\eta,$$

(25)
where the function $f_K$ is defined by Eq. (8). Note that the integrand includes the point mass at $\eta = 0$ which accounts for the possibility of extinction of the entire population. The contour marked 0.1 connects starting configurations for which there is a 10% probability the population will undershoot the observed population at time $s$, and the contour marked 0.9 connects starting configurations for which there is a 10% chance of overshooting. Thus these two contours define the edges of an 80% confidence interval on starting configurations $(s, \kappa_0)$ for the observed final population $\kappa$. The small circle in each plot is a “median estimate” of the coalescence point which we define as the (scaled) starting time and population size for which there is a 50% probability that the coalescent event is yet to occur, and a 50% probability that the final population will overshoot the observed current observed population size.

4. Locating the coalescence point

Naturally one would like to devise a more intuitive description of the location of the ancestral coalescent point of a random sample. In this section we propose a likelihood surface over the $(s, \kappa_0)$-plane corresponding to initial conditions $K(0) = \kappa_0$ coinciding with the instant of the MCRA of an observed sample at time $s$, given that the current population is observed to be $K(s) = \kappa$.

Consider two neighbouring contours in the $(s, \kappa_0)$-plane corresponding to the probability defined in Eq. (23) or (24) taking values $u$ and $u + du$ respectively, as shown in Fig. 2. Now $u(s, \kappa_0|\kappa)$ is the probability that a BGW process initiated at any point on the first contour will result in a sample of size $n$ at time $s$ having a single ancestor in the initial population, conditional on $K(0) = \kappa_0$ and $K(s) = \kappa$. Similarly $u(s, \kappa_0|\kappa) + du(s, \kappa_0|\kappa)$ is the corresponding probability for the second contour. Then for a fixed value of $\kappa_0$, the random confidence interval

$$I(u, u + du|\kappa_0, K) := \{s : u < u_n(s, \kappa_0|K) < u + du\}, \quad (26)$$

has a probability $du$ of containing the time since the MRCA of the sample, given that the current population is observed to be of size $K$.

Furthermore, by the argument set out in Appendix A at a given value of $s$ the random confidence interval delimited by the contours $v$ and $v + dv$ shown in Fig. 2 namely

$$I(v, v + dv|s, K) := \{\kappa_0 : v < v(s, \kappa_0|K) < v + dv\}, \quad (27)$$

11
has a probability $dv$ of containing the initial population $K(0)$ given an observed population size $K(s)$.

We now map the $(s,\kappa_0)$-plane onto the $(u,v)$-plane and conjecture that, if (1) a BGW process is initiated with an initial population $\kappa_0$, (2) at time $s$ the population is observed to be $K$, and (3) the ancestry of a sample of the population happens to converge at a MRCA at time zero, then the region shown in red in Fig. 2 has a probability $du\,dv$ of containing the unobserved parameters $s,\kappa_0$. Equivalently, we conjecture that the likelihood density is uniformly distributed over the $(u,v)$-plane. Establishing the equivalent likelihood density over the $(s,\kappa_0)$-plane is then a matter of mapping a uniform distribution over the range $(u,v) \in [0,1] \times [0,1]$ back to the $(s,\kappa_0)$-plane. This is effected by the Jacobean of the transformation, giving the density

$$L(s,\kappa_0|\kappa) = \left| \frac{\partial u_n(s,\kappa_0|\kappa)}{\partial s} \frac{\partial v(s,\kappa_0|\kappa)}{\partial \kappa_0} - \frac{\partial u_n(s,\kappa_0|\kappa)}{\partial \kappa_0} \frac{\partial v(s,\kappa_0|\kappa)}{\partial s} \right|, \quad (28)$$

where $K = \kappa$ is the observed current population.

Contours of the density function Eq. (28) are marked in red in Figure 1 for $\kappa = 3$ and $\kappa = 2000$. Details of the calculation of the required derivatives are set out in Appendix B. These contour lines have the property that they bound a minimal area confidence region in the $(s,\kappa_0)$-plane for a given integrated likelihood. Immediately noticeable is that the in both cases the
confidence regions are extremely skewed, and the median estimate defined at the end of the previous section is well separated from the maximum of the likelihood surface.

The marginal likelihood of the scaled time since coalescence, which we define from Eq. (28) as

$$L(s|\kappa) = \int_0^\infty L(s,\kappa_0|\kappa)d\kappa_0,$$

is plotted in Fig. 3 for a sample of \(n = 2\) individuals and various values of the currently observed scaled population \(\kappa\). The numerical integration is tricky to evaluate for large values of \(\kappa\) and small values of \(s\) because of the difficulty in numerically evaluating \(\partial v/\partial \kappa_0\) along the ridge in the joint density \(L(s,\kappa_0|\kappa)\) running up to the point \((s,\kappa_0) = (0,\kappa)\) (see Fig. 1(b), for instance). To produce the plots in Fig. 3(b) we avoided this problem by
integrating by parts to obtain

$$L(s|\kappa) = \int_0^\infty \left[ \frac{\partial u_n(s, \kappa_0|\kappa)}{\partial \kappa_0} \frac{\partial v(s, \kappa_0|\kappa)}{\partial s} + \frac{\partial^2 u_n(s, \kappa_0|\kappa)}{\partial s \partial \kappa_0} v(s, \kappa_0|\kappa) \right] d\kappa_0,$$

where the second partial derivative of $u_n$ is given in Appendix B, Eq. (B.7).

Since the marginal likelihood integrates to 1, a comparison can be made with the probability distribution of times since coalescence for a sample of size $n = 2$ taken from an exponentially growing WF population, as determined by Slatkin and Hudson (1991). Their distribution is quoted in terms of a single parameter $\alpha = N_0r$ where $N_0$ is the final observed haploid population size and $r$ is the exponential growth rate. In terms of our notation, the expected number of offspring per individual per generation is $\lambda = e^r$, and the variance of the number of offspring per individual per generation for a WF model in the diffusion limit is $\sigma^2 = 1 + O(1/M(t))$. Thus the condition $K(s) = 2M(t)\log \lambda / \sigma^2 = \kappa$ translates to $\kappa = 2\alpha$, and Slatkin and Hudson’s Eq. (6) translates to

$$f_{SH}^S(s|\kappa) = \frac{2e^s}{\kappa} \exp \left( -\frac{2(e^s - 1)}{\kappa} \right).$$

This distribution is plotted as dashed curves in Fig. 3. Note that the parameters chosen in Fig. 3(b) match those of Fig. 4 in Slatkin and Hudson (1991).

Figure 4 shows the marginal likelihood of the time since coalescence for samples of size $n = 2, 3, \ldots$. The choice of scaled current population, $\kappa = 4500$, matches the naive model of human population growth during the upper Paleolithic period employed by Burden and Simon (2016) to illustrate an application of BGW processes to mitochondrial Eve. In common with the predictions of the WF model for an exponentially growing population (Slatkin and Hudson 1991), we observe that MRCAs for samples of any size are likely to be restricted to a limited time range, consistent with a “star” shaped phylogeny.

5. Asymptotic behaviour for large $\kappa$

Plots of the marginal likelihood and of the Slatkin-Hudson density in Fig. 3 suggest that these functions each converge to an asymptotic shape as
\( \kappa \to \infty \). Specifically, by defining a shifted time scale
\[
s' = s - \log \kappa,
\]
one easily checks from Eq. (31) that
\[
L_{\text{SH}}(s'|\log \kappa) = 2e^{s'}e^{-2e^{s'}}(1 + O(\kappa^{-1})),
\]
as \( \kappa \to \infty \). Thus the Slatkin-Hudson density rapidly takes on a shape related to the Gumbel distribution for large \( \kappa \). As we next show, it turns out that each of the likelihood functions defined above for a BGW process also converges to an asymptotic shape depending only on \( s' \) (and, where relevant, \( \kappa_0 \)) for large \( \kappa \).

To check this it is sufficient to demonstrate that the functions \( u_n \) and \( v \) whose contours are plotted in Fig. 1 each take on a universal form as functions of \( s' \) and \( \kappa_0 \) as \( \kappa \to \infty \). From Eqs. (23), (24) and (18) we have that \( u_n \) and \( u_\infty \) are functions only of \( w \) which behaves asymptotically as
\[(\kappa_0 e^{-s'})^{\frac{1}{2}} (1 + O(\kappa^{-1})).\] It immediately follows that for large \(\kappa\),

\[
u_n(s, \kappa_0 | \kappa) \approx \tilde{u}_n(s - \log \kappa, \kappa_0), \quad u_\infty(s, \kappa_0 | \kappa) \approx \tilde{u}_\infty(s - \log \kappa, \kappa_0),
\]

where

\[
\begin{align*}
\tilde{u}_n(s', \kappa_0) &= \frac{n!}{w^{n-1}} \frac{I_n(2w')}{I_1(2w')}, \\
\tilde{u}_\infty(s', \kappa_0) &= \frac{w'}{I_1(2w')}, \quad w' = (\kappa_0 e^{-s'})^{\frac{1}{2}}.
\end{align*}
\]

The asymptotic form of \(v\) is obtained from Eqs. (25) and (10) by substituting Eq. (32) and discarding terms of \(O(\kappa^{-1})\) to obtain

\[
v(s, \kappa_0 | \kappa) \approx \tilde{v}_n(s - \log \kappa, \kappa_0),
\]

where

\[
\begin{align*}
\tilde{v}_n(s', \kappa_0) &= F_{\text{Feller}} \left( \frac{1}{\kappa_0 e^{s'}; \kappa_0} \right).
\end{align*}
\]

Here \(F_{\text{Feller}}(z, \kappa_0) = \int_0^z f_{\text{Feller}}(\zeta, \kappa_0) d\zeta\) is the cumulative distribution corresponding to the density defined by Eq. (9).

Figure 5 shows contours of the functions \(u_2, u_\infty, v\), and contours of associated likelihood surfaces, \(L(s' + \log \kappa, \kappa_0 | \kappa)\) defined by Eq. (28), for relatively large values \(\kappa = 50\) and 1000 of the observed current population. On the scale shown, the \(\kappa = 1000\) contours are indistinguishable from those corresponding to the universal functions \(\tilde{u}_n, \tilde{u}_\infty, \text{and} \tilde{v}\) (not shown). Note that the universal functions are approached more rapidly for larger values of \(n\).

Contours of the limiting likelihood \(\tilde{L}(s', \kappa_0) := \lim_{\kappa \to \infty} L(s' + \log \kappa, \kappa_0 | \kappa)\) and the limiting marginal likelihood \(\tilde{L}(s') := \lim_{\kappa \to \infty} L(s' + \log \kappa | \kappa)\), calculated from \(\tilde{u}_n(s', \kappa_0)\) and \(\tilde{v}(s', \kappa_0)\) are plotted in Fig. 6. We have observed that the numerically calculated limiting marginal likelihood for \(n = \infty\), that is, the red curve in Figures 6(d) and (f), is a very close fit to a shifted Gumbel distribution. So far we have been unable to verify this analytically.

Also plotted in Fig. 6 are simulated data produced as follows. A set of trees was generated, each starting with a single ancestor and evolving as a BGW process, with each parent in the process independently producing a Poisson number of offspring with \(\lambda = \sigma^2 = 1.01\). Trees were generated until a dataset of 40000 family trees, each surviving for \(t_{\text{sim}} = 1000\) generations, was accumulated. For each tree \(n = 2\) individuals were randomly chosen.
Figure 5: Contours of the probability \( u_n(s' + \log \kappa, \kappa_0 | \kappa) \) that a random sample of \( n \) individuals is descended from a single individual (black lines), the probability \( v(s' + \log \kappa, \kappa_0 | \kappa) \) that the population at time \( s \) will not exceed the observed population \( \kappa \) (blue lines), and the corresponding likelihood surface (red contours) as functions of the shifted time \( s' \) and currently observed population \( \kappa_0 \).

from the final generation and their ancestry traced to locate the time of their MRCA, and to record the population size at that time. Each sample was taken from a separate tree to avoid correlations between the MRCAs of samples with a common history (Ball et al., 1990; Slatkin and Hudson, 1991). For each family tree the time and population size corresponding to the MRCA of the entire population in the final generation was also recorded. These coalescence times and population sizes were then transformed to the scaled quantities \( s' \) and \( \kappa_0 \).

For consistency with the \( \kappa \to \infty \) limit, a small number of trees with final scaled population sizes \( \kappa < 50 \) were discarded to produce the scatter plots in Figures 6(a) and (b) and the histograms of \( s' \) values in Figures 6(c) and (d). It is clear that the right hand tails of the histograms fall short of the theoretical marginal likelihood curves. To appreciate the cause of this discrepancy, note that the marginal likelihoods in Figures 6(c) and (d) drop to almost zero outside a finite range, \( s'_{\text{lower}} < s' < s'_{\text{upper}} \). However, for
Figure 6: Plots (a) and (b): Contours of the asymptotic universal functions $\tilde{u}_n(s', \kappa_0)$ (black), $\tilde{v}(s', \kappa_0)$ (blue) and likelihood surface $\tilde{L}(s', \kappa_0)$ (red), together with a scatterplot of simulated data as described in the text. Plots (c) and (d): Asymptotic marginal likelihood $\tilde{L}(s')$ together with a histogram of the data set of scaled and shifted coalescent times $s'$ with data corresponding to $\kappa < 50$ culled. Plots (e) and (f): the same, but with more severe culling to include only trees corresponding to earlier starting times, as described in the text.
practical reasons, a numerical simulation is restricted to be of finite time, namely $t_{\text{sim}}$ generations, and because of this most of the trees in the sample have not existed sufficiently far in the past since the initial founder to cover this entire range. In an attempt to remedy this in Figures 6(e) and (f) we have also culled from the data all trees for which

$$t_{\text{sim}} \log \lambda < s'_{\text{upper}} + \log \kappa.$$  \hspace{1cm} (38)

This second culling is more severe, and removes the majority of generated trees. Setting $s'_{\text{upper}} = 5$, the dataset is reduced to trees satisfying $50 < \kappa < 141.2$, which reduces the original set to 192 trees. Nevertheless, the 2 final histograms are in closer agreement with the right hand tails of the theoretical marginal likelihoods.

6. Conclusions

We have derived joint likelihood functions for the scaled time $s = t \log \lambda$ since coalescence to a MRCA of a random sample and the scaled population size $\kappa_0 = (2m_0 \log \lambda)/\sigma^2$ at the time of the MRCA under the assumption that the current observed population is the result of a BGW process that has been evolving since some unknown time in the past. Here $t$ is the number of time steps since the MRCA and $m_0$ is the corresponding population size. The scalings are defined in terms of the mean $\lambda$ and variance $\sigma^2$ of the number of offspring per parent. The mean $\lambda$ is assumed to be close to and slightly larger than 1, and $t$ is assumed to be large, so that the diffusion limit can be applied. The only input into the likelihood functions is the currently observed scaled population $\kappa = 2m \log \lambda/\sigma^2$, where $m$ is the currently observed population.

The approach differs markedly from earlier analyses of the MRCA in a BGW diffusion (O’Connell 1995; Harris et al. 2019) in that the BGW process is not initiated from a pre-specified point in the past. This is not a problem for tracing ancestry in WF-like models (Slatkin and Hudson 1991) which are inherently backward-looking in time. In such models Kingman’s coalescent is immediately applicable, and the concept of a probability density for the time since a MRCA is meaningful. For a BGW process, however, causality runs forward in time, and to be on firm ground we have couched our results in terms of likelihoods. Nevertheless some interpretation is in order.
To return to the question posed in the opening paragraph of this paper, suppose we are confronted with a population which we are told has evolved through a BGW process. Although the process is stochastic and its history unknown to us, our ontological commitment is to a single realised path through the \((s,\kappa_0)\)-plane leading to the the current observed scaled population \(\kappa\), and at some point that unknown path will have passed through co-ordinates corresponding to the MRCA of the current population. Given that a BGW process is Markovian, the current population can be considered to be the result of a BGW process initiated from any point along the path, and in particular, from those co-ordinates corresponding to the MRCA. Our claim is that Eq. \[(28)\] calculated from the observed value of \(\kappa\) defines a likelihood surface in the sense that any subset \(\Omega\) of the \((s,\kappa_0)\)-plane has a probability \(\int_{\Omega} \mathcal{L}(s,\kappa_0|\kappa)dsd\kappa_0\) of containing the unknown but extant co-ordinates corresponding to the MRCA of the current population.

Having made this claim, we mention a couple of final caveats. Firstly, note that the paragraph containing Eq. \[(28)\] includes a conjecture that essentially says that the likelihood surface over the \((u,v)\)-plane is uniform. Although the marginal likelihoods in \(u\) and \(v\) are uniform, this is not necessarily the case for the joint likelihood, and without further mathematical proof, remains a conjecture. Secondly, if the likelihood is well defined, one feels it should be possible, at least in principle if not in practice, to devise a numerical simulation which will confirm its interpretation. The simulation in Section \[5\] attempts to do this, but as we have seen, there is some ambiguity about how match this particular simulation with theory. Devising an appropriate simulation which is guaranteed to include the MRCA of the final population but does not impose any prior constraint on the population size at the time of the MRCA remains an open problem.

**Acknowledgements**

CB would like to thank Robert Griffiths for interesting and helpful discussions.

**Appendix A. Confidence interval for \(\kappa_0\) given an observed \(K(s)\)**

This appendix is devoted to establishing a confidence interval for estimating the initial population \(K(0) = \kappa_0\) given that a population generated by a BGW process is observed to be of size \(K(s)\) at a known time \(s\).
The function \( v(s, \kappa_0|\kappa) \) was defined in Eq. (25) to be the cumulative distribution function of \( K(s) \) given an initial population \( \kappa_0 \). Consider an inverse function \( v^{-1}(p|s, \kappa) \) mapping the interval \([0, 1]\) back to the initial condition \( \kappa_0 \) in the interval \([0, \infty)\), thus,

\[
v^{-1}(v(s, \kappa_0|\kappa)|s, \kappa) = \kappa_0.
\] (A.1)

From Fig. 1 it is clear that \( v^{-1} \) is in general a decreasing function of \( p \), with \( v^{-1}(0|s, \kappa) = \infty \) and \( v^{-1}(1|s, \kappa) = 0 \). We claim that for any \( 0 \leq p_1 < p_2 \leq 1 \), the random interval

\[
I(p_1, p_2|K(s)) := [v^{-1}(p_2|s, K(s)), v^{-1}(p_1|s, K(s))],
\] (A.2)

has a probability \( p_2 - p_1 \) of containing the true starting population \( \kappa_0 \).

**Proof.**

\[
\text{Prob} \left( I(p_1, p_2|K(s)) \ni \kappa_0 \right) = \text{Prob} \left( v^{-1}(p_1|s, K(s)) \geq \kappa_0 \geq v^{-1}(p_2|s, K(s)) \right) \\
= \text{Prob} \left( p_1 \leq v(s, \kappa_0|K(s)) \leq p_2 \right) \\
= p_2 - p_1. \tag{A.3}
\]

The second last line follows because \( v^{-1} \) is a decreasing function of \( p \), and the final line follows because the cumulative distribution \( v(s, \kappa_0|K(s)) \) is uniformly distributed on \([0, 1]\). \(\square\)

**Appendix B. Derivatives of \( u_n(s, \kappa_0|\kappa) \) and \( v(s, \kappa_0|\kappa) \)**

Derivatives of \( u_n(s, \kappa_0|\kappa) \) and \( v(s, \kappa_0|\kappa) \) are needed for the numerical determination of the likelihood surface defined by Eq. (28) and marginal likelihood defined by Eq. (30). Derivatives of \( u_n(s, \kappa_0|\kappa) \) are straightforward to calculate from Eqs. (23) and (18). Using the identity (Abramowitz and Stegun, 1965, p376)

\[
\frac{d}{dz} \left( \frac{I_n(z)}{z^n} \right) = \frac{I_{n+1}(z)}{z^n}, \tag{B.1}
\]

we obtain

\[
\frac{\partial u_n(s, \kappa_0|\kappa)}{\partial s} = - \frac{n!}{w^{n-2}} \coth \left( \frac{1}{2} s \right) \Phi_n(w), \quad \frac{\partial u_n(s, \kappa_0|\kappa)}{\partial \kappa_0} = \frac{n!}{\kappa_0 w^{n-2}} \Phi_n(w), \tag{B.2}
\]
and
\[
\frac{\partial^2 u_n(s, \kappa_0|\kappa)}{\partial s \partial \kappa_0} = \frac{-n!}{\kappa_0 w^{n-2}} (\Phi_n(w) + w \Psi_n(w)) \coth(\frac{1}{2}s),
\]  
(B.3)

where
\[
\Phi_n(w) = \frac{I_1(2w)I_{n+1}(2w) - I_2(2w)I_n(2w)}{I_1(2w)^2},
\]  
(B.4)

and
\[
\Psi_n(w) = \frac{1}{I_1(2w)^3} \left\{ I_1(2w)[I_1(2w)I_{n+2}(2w) - I_3(2w)I_n(2w)] 
- 2I_2(2w)[I_1(2w)I_{n+1}(2w) - I_2(2w)I_n(2w)] \right\}. 
\]  
(B.5)

From Eq. (24) the analogous derivatives for \(n = \infty\) are
\[
\frac{\partial u_\infty(s, \kappa_0|\kappa)}{\partial s} = w^2 \coth(\frac{1}{2}s) \Phi_\infty(w), \quad \frac{\partial u_\infty(s, \kappa_0|\kappa)}{\partial \kappa_0} = -\frac{w^2}{\kappa_0} \Phi_\infty(w),
\]  
(B.6)

and
\[
\frac{\partial^2 u_\infty(s, \kappa_0|\kappa)}{\partial s \partial \kappa_0} = \frac{w^2}{\kappa_0} (\Phi_\infty(w) + w \Psi_\infty(w)) \coth(\frac{1}{2}s),
\]  
(B.7)

where
\[
\Phi_\infty(w) = \frac{I_2(2w)}{I_1(2w)^2},
\]  
(B.8)

and
\[
\Psi_\infty(w) = \frac{I_1(2w)I_3(2w) - 2I_2(2w)^2}{I_1(2w)^3}. 
\]  
(B.9)

Derivatives of \(v(s, \kappa_0|\kappa)\) are less straightforward. The derivative with respect to \(s\) can be calculated from the forward-Kolmogorov equation, Eq. (5), as follows;
\[
\frac{\partial v(s, \kappa_0|\kappa)}{\partial s} = -\int^\infty_\kappa \frac{\partial f_K(\eta; s|\kappa_0)}{\partial s} d\eta
\]  
\[= \int^\infty_\kappa \left[ \frac{\partial}{\partial \eta}(\eta f_K(\eta; s|\kappa_0)) - \frac{\partial^2}{\partial \eta^2}(\eta f_K(\eta; s|\kappa_0)) \right] d\eta
\]  
\[= -\kappa f_K(\kappa; s|\kappa_0) + \frac{\partial}{\partial \kappa}(\kappa f_K(\kappa; s|\kappa_0))
\]  
\[= w \left[ \left( \frac{1 - \kappa}{\kappa} - \frac{1}{e^s - 1} \right) I_1(2w) + \frac{w}{\kappa} I_2(2w) \right] \exp \left\{ -\frac{\kappa_0 e^s + \kappa}{e^s - 1} \right\}, 
\]  
(B.10)
where Eqs. (8), (18) and (B.1) have been used in the last line. Evaluation of the derivative of \( v \) with respect to \( \kappa_0 \) involves a numerical integration. Some straightforward but lengthy algebra gives

\[
\frac{\partial}{\partial \kappa_0} f_K(\kappa; s|\kappa_0) = -\frac{\delta(\kappa)}{1 - e^{-s}} \exp \left\{ -\frac{\kappa_0}{1 - e^{-s}} \right\} \\
+ \frac{1}{2 \cosh s - 1} \exp \left\{ -\frac{\kappa_0 e^s + \kappa}{e^s - 1} \right\} \times \\
\left[ I_2(2w) + \left( 1 - \frac{\kappa_0 e^s}{e^s - 1} \right) \frac{I_1(2w)}{w} \right]. 
\]

Then from Eq. (25)

\[
\frac{\partial v(s, \kappa_0|\kappa)}{\partial \kappa_0} = \int_0^\kappa \frac{\partial}{\partial \kappa_0} f_K(\eta; s|\kappa_0) \, d\eta \\
= -\frac{1}{1 - e^{-s}} \exp \left\{ -\frac{\kappa_0}{1 - e^{-s}} \right\} \\
+ \frac{1}{2 \cosh s - 1} \int_0^\kappa \exp \left\{ -\frac{\kappa_0 e^s + \eta}{e^s - 1} \right\} \times \\
\left[ I_2(2\omega) + \left( 1 - \frac{\kappa_0 e^s}{e^s - 1} \right) \frac{I_1(2\omega)}{\omega} \right] \, d\eta, 
\]

where

\[
\omega = \frac{(\eta \kappa_0 e^s)^{\frac{1}{2}}}{e^s - 1}. 
\]

References

References

Abramowitz, M., Stegun, I. A., 1965. Handbook of mathematical functions: with formulas, graphs, and mathematical tables. Dover Publications, New York.

Athreya, K. B., 2012. Coalescence in critical and subcritical Galton-Watson branching processes. Journal of Applied Probability 49 (3), 627–638.

Bailey, N. T. J., 1964. The elements of stochastic processes with applications to the natural sciences. Wiley, New York.
Ball, R. M., Neigel, J. E., Avise, J. C., 1990. Gene genealogies within the organismal pedigrees of random-mating populations. Evolution 44 (2), 360–370.

Burden, C. J., Simon, H., 2016. Genetic drift in populations governed by a Galton–Watson branching process. Theoretical Population Biology 109, 63–74.

Burden, C. J., Wei, Y., 2018. Mutation in populations governed by a Galton–Watson branching process. Theoretical Population Biology 120, 52–61.

Cyran, K. A., Kimmel, M., 2010. Alternatives to the Wright–Fisher model: The robustness of mitochondrial Eve dating. Theoretical population biology 78 (3), 165–172.

Durrett, R., 1978. The genealogy of critical branching processes. Stochastic Processes and their Applications 8 (1), 101–116.

Feller, W., 1951a. Diffusion processes in genetics. In: Proc. Second Berkeley Symp. Math. Statist. Prob. Vol. 227. p. 246.

Feller, W., 1951b. Two singular diffusion problems. Annals of Mathematics 54 (1), 173–182.

Grosjean, N., Huillet, T., 2018. On the genealogy and coalescence times of Bienaymé-Galton-Watson branching processes. Stochastic Models 34 (1), 1–24.

Harris, S. C., Johnston, S. G., Roberts, M. I., 2019. The coalescent structure of continuous-time Galton-Watson trees. arXiv preprint arXiv:1703.00299v4.

Kimmel, M., Axelrod, D., 2015. Branching Processes in Biology. Interdisciplinary Applied Mathematics. Springer New York. URL https://books.google.com.au/books?id=-ZK3BqAAQBAJ

Lambert, A., 2003. Coalescence times for the branching process. Advances in Applied Probability 35 (4), 1071–1089.

Lambert, A., Popovic, L., et al., 2013. The coalescent point process of branching trees. The Annals of Applied Probability 23 (1), 99–144.
Le, V., 2014. Coalescence times for the Bienaymé-Galton-Watson process. Journal of Applied Probability 51 (1), 209–218.

O’Connell, N., 1995. The genealogy of branching processes and the age of our most recent common ancestor. Advances in applied probability, 418–442.

Pardoux, É., 2016. Probabilistic models of population evolution. Mathematical Biosciences Institute Lecture Series. Stochastics in Biological Systems. Springer, Berlin.

Slatkin, M., Hudson, R. R., Oct 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics 129 (2), 555–62.

Wolfram Research, Inc., February 2019.
URL https://www.wolframalpha.com