Journal of Mathematical Biology manuscript No.
(will be inserted by the editor)

Calculations for multi-type age-dependent binary branching processes

Graham Jones

Received: date / Accepted: date

Abstract This article provides a method for calculating the joint probability density for the topology and the node times of a tree which has been produced by an multi-type age-dependent binary branching process and then sampled at a given time. These processes are a generalization, in two ways, of the constant rate birth-death process. There are a finite number of types of particle instead of a single type: each particle behaves in the same way as all others of the same type, but different types can behave differently. Secondly, the lifetime of a particle (before it either dies, changes to another type, or splits into two) follows an arbitrary distribution, instead of the exponential lifetime in the constant rate case. Two applications concern models for macroevolution: the particles represent species, and the extant species are randomly sampled. In one application, 1-type and 2-type models for macroevolution are compared. The other is aimed at Bayesian phylogenetic analysis where the models considered here can provide a more realistic and more robust prior distribution over trees than is usually used. A third application is in the study of cell proliferation, where various types of cell can divide and differentiate.

Keywords Tree · Tree shape · Branching process · Bellman-Harris · Bayesian phylogenetic analysis · Macroevolution · Cell proliferation

Mathematics Subject Classification (2000) 92-08 · 60J85 · 60J80 · 45D05

1 Introduction

Branching processes have a long history as models for various biological processes and structures [26]. The models considered here have been studied by mathematicians for over 50 years [7,8]. Their asymptotic properties are well understood, but in biological applications, the time scales are often such that the behavior of the processes is

---

G. Jones
21e Balnakeil, Durness, Lairg, Sutherland, IV27 4PT, United Kingdom
E-mail: art@gjones.name

quite different from the asymptotic behavior. Also, for some applications, notably the evolutionary process, detailed information about the process is required. For example, in applications in macroevolution and phylogenetic analysis, the probabilities of different tree topologies may be required. (The tree topology is the branching pattern ignoring the node times.) This article is aimed at providing a practical computational method for finding the probability that a particular tree is generated by such a process. Three motivating examples are described next. After dealing with the mathematical and computational aspects in Sect. 2 and Sect. 3, the examples are revisited in Sect. 4.

1.1 Macroevolution

The Tree of Life has many nodes in which one of the two subtrees rooted at that node is much larger than the other. A well-known example is the 22 Crocodylla (crocodiles, alligators and gharials) versus the 10,000 or so birds. A node like this is called 'unbalanced' and there are various statistics (Colless, Sackin, etc) which measure the degree of imbalance in a tree or a collection of trees. Since the constant rate birth-death model produces trees which are too balanced to match empirical phylogenetic trees, there has been considerable interest in more complex models [2,4,10,16,20,28, 27,35,39]. Age-dependent models (also known as Bellman-Harris processes) were considered in [1,19,24]. In [19] some formulas for tree probabilities are given, but restricted to the case where there are no extinctions or sampling, and only one type.

It is unclear in most cases whether the imbalance at a node is due to specific biological reasons (for example a trait such as flight possessed by one of the two groups of species) or whether evolution is just a particular kind of random process which regularly produces this type of imbalance. A suitable 'null' model is desirable in order to test hypotheses of this kind.

Another major issue is that of obtaining estimates for rates of extinctions. Sometimes fossil evidence for species duration is available, so that the extinction rate can be estimated directly. If there is only fossil evidence for the age of a clade then the diversification rate can be estimated, but without a model for the evolutionary process, the speciation rates and extinction rates cannot be separated [12, Chapter 12]. If a model of the kind discussed here is assumed, a distribution for the node times in the tree can be found, and then the relative rates of speciation and extinction can be estimated along the lines of [42,43].

Models for macroevolution can broadly be classified in terms of how much of the rest of the tree influences the behavior of a given species. The constant rate birth-death process is the simplest, where every species at every point in time behaves in the same way.

An age-dependent process with a single type means that the behavior of a species depends on the time since it diverged, but on nothing else. In particular, nothing is inherited from its ancestors. Imbalance can arise with a suitable distribution for lifetimes. Consider the tree process in Fig. 1. Each of the random variables Y_i is independently drawn from the same distribution. There are currently three tips, and the balance of the tree depends on which of the three tips divides next. If one of the two branches on the left divides, then there will be a 3:1 (unbalanced) split at the root, otherwise a 2:2 (balanced) split. In the constant rate birth process, where the Y_i follow an exponential distribution, each of the branches has an equal probability of dividing, so the probability of a 2:2 split is 1/3. If the Y_i follow another distribution, this may be different. In general, the probability of a 2:2 split is

$$\Pr[\min(Y_1, Y_2) + \min(Y_3, Y_4) > \max(Y_1, Y_2)]$$

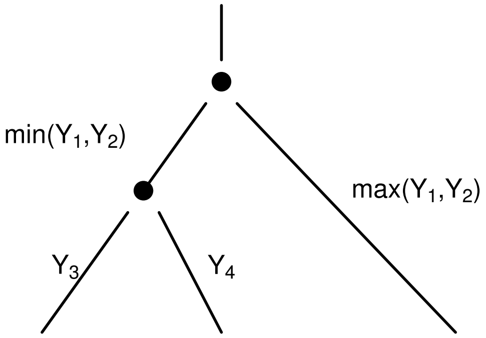

Fig. 1 A tree with 3 tips generated by a pure birth process. The Y_i are random variables giving the lengths of the branches.

If the Y_i follow a gamma distribution with shape parameter 0.5, this probability is about 0.24 instead of 1/3 if Y_i follows an exponential distribution. In general, distributions which have a higher probability of producing very small and very large values, result in more unbalanced trees [24]. The term 'L-shaped' is used to describe such distributions. Examples include a gamma distribution with shape parameter less than one, a mixture of two exponential distributions with different scales, and a Pareto distribution which will be illustrated later.

By introducing different types of species, the behavior of a species can depend on its ancestors. For example, a fast diversifying type could model a species with a key innovation, while a slow diversifying type represents an 'ordinary' species. Sect. 4.1 compares a 1-type model and a 2-type model.

1.2 Phylogenetic Analysis

A closely related application is in Bayesian phylogenetic analysis. Two well-known programs for performing Bayesian phylogenetic analysis are MrBayes [21] and BEAST [14]. The former uses unrooted trees in which branches have a length measured in terms of substitution rates. This allows the topology to be estimated, but does not provide dates of nodes. On the other hand BEAST uses rooted trees with a strict or relaxed molecular clock and does estimate node times measured in years.

In both cases a prior for the distribution of trees is required. MrBayes uses the uniform or 'proportional to distinguishable arrangements' (PDA) prior for topologies. The term PDA applies both to the unrooted trees used by MrBayes and to those trees after they have been rooted within any branch. The PDA distribution produces unbalanced rooted trees. For example, in the 4-tip case, it gives a probability of 1/5 to a 2:2 split, and 2/5 to a 1:3 or 3:1 split. BEAST uses a prior based on the constant rate birth-death model. This results in a joint distribution for topology and node times. The distribution over tree topologies is the same as for the constant rate pure birth model [4], thus the probability of a 2:2 split in the 4-tip case is 1/3.

The choice of prior has been controversial ([15, Chapter 18], [11,33,34,44]). However, the distribution of topologies in real phylogenetic trees is known quite well and is surprisingly homogeneous across many different parts of the Tree of Life [10, 24]. In particular, the empirical trees are more balanced than the PDA prior used in MrBayes, and less balanced than the prior used in BEAST. It makes sense to use this information when undertaking a phylogenetic analysis, and this should not be controversial, even to those who object to Bayesian methods.

It is easier and more natural to deal with the case of rooted trees with a strict or relaxed molecular clock (as used in [13,30] for example) and so this article concentrates on rooted trees with node times (divergence times) measured in years. I will call the duration of a branch an 'internode time' to avoid confusion with branch lengths measured in terms of substitution rates.

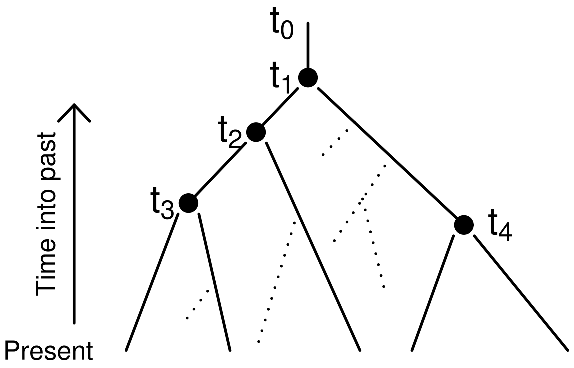

Fig. 2 A tree with 5 species appearing in the sample. Dotted lines indicate branches which either go extinct before the present, or reach the present but are not sampled. The solid lines constitute the reconstructed tree. The t_i are node times.

Researchers almost never see the entire tree. Extinctions and sampling of extant species mean that the tree that a phylogenetic analysis attempts to infer (the reconstructed tree) is only a part of the full tree as shown in Fig. 2. It is difficult to model mathematically the way researchers choose taxa, but it is important to note that even unbiased sampling affects the distribution of trees. If species are sparsely sampled, this will change the distribution for node times [42,43,38], and may well affect the distribution over topologies [20,24], when compared to reconstructed trees arising from densely sampled species. Thus, despite the difficulty, it is important to make an attempt at modeling the sampling.

Currently, the amount of reliable empirical data about the distribution of node times is small, and until more good quality data is available, it seems best to make the prior robust in the sense that it should not make extreme values (very short or very long internode times) too unlikely. A distribution more L-shaped than the exponential distribution for species lifetime seems appropriate. See [9, Section 4.7] for a general discussion of robustness in Bayesian analysis, and [44] for an investigation into the effects of priors on branch lengths in the context of unrooted trees.

A good prior distribution on trees for Bayesian phylogenetic analysis would (a) provide a distribution over topologies which matches the observed degree of balance, (b) provide a suitably robust (L-shaped) distribution over node times for varying rates of extinction and varying rates of sampling, and (c) be mathematically tractable in the sense that the probability of observing a particular tree given the prior can be readily calculated. There is currently no distribution satisfying all these criteria. Models such as those of Aldous [2], Ford [16] and Blum and François [10] satisfy (a) and (c), but the first two provide no node times, and none of them model extinctions. The much-studied constant rate birth-death model [3,32,37,42,43] satisfies (c) and mostly satisfies (b) (though a more L-shaped distribution seems preferable), but produces trees which are too balanced to match empirical trees. Many models which simulate the macroevolutionary process in computer programs can satisfy (a) and (b) with suitable choice of parameters, but there is no sensible way of calculating the probability that a particular tree could be generated by the program. The processes studied here can satisfy (a) and (b), and Sect. 4.2 tackles the calculation of probabilities for them. These calculations are still slow for use in phylogenetic analysis with large trees, but not impossibly slow.

1.3 Cell proliferation

Multi-type age-dependent binary branching processes have often been used as models of the kinetics of populations of dividing and differentiating cells [23,26,31,41]. In general the methods used have been based on simple models where analytic results are available; asymptotic results for more complex models; or computer simulations. All of these have disadvantages [22]. A numerical method based on saddlepoint approximations is developed in [22], which provides a way of calculating moments (such as means, variances, and covariances) of the numbers of cells of different types. The method presented here (Sect. 4.3) provides an alternative. The two approaches are complementary: the present approach provides more detailed information at the cost of a greater amount of calculation. It can answer questions such as “What is the probability that a single cell of type A gives rise to more than n cells of type B in a given time t?".

2 The Main Equations

In the models considered here, each particle of a given type has the same lifetime distribution, and at the end of its lifetime (an event I will call *termination*) a particle either produces two particles (*division*), or none (*extinction*). These restrictions help to simplify the formulas, and more complex situations such as those considered in [23] can be modeled in this framework by using unobserved types. For example, in some applications one might want to allow the possibility that a particle changes type without dividing. This can be modeled as a special kind of division which produces one particle of a different type, and another particle of a special type which either goes extinct immediately, or which, by definition, is never observed. In other applications one might want to have different lifetime distributions for particles which divide and for those which go extinct. This too can be dealt with using extra types.

Assume there are r types of particles (e.g. species or cells) labeled $1, \ldots, r$. A single particle of some type i is born at some time t in the past. I will call this particle the *ancestor*. Time is measured backwards from the present which is taken to be time zero. The particle lives for a random period x determined by a distribution $F^{(i)}(x)$ with support $[0, \infty)$, and then at time $t - x$ it either divides with probability $p^{(i)}$, or becomes extinct with probability $1 - p^{(i)}$. The types of a new pair of particles is chosen according to the discrete probabilities $\xi^{(i)}(j,k)$ $(1 \leq i, j, k \leq r)$. Here $\xi^{(i)}(j,k)$ is the probability that when a particle of type i divides, the 'left' descendant has type j and the 'right' descendant has type k. The labels 'left' and 'right' are just to facilitate counting and have no biological significance; $\xi^{(i)}(j,k) = \xi^{(i)}(k,j)$ always. This means that the trees are 'oriented' trees as described in [17,18]. At any point in time, all particles behave independently of one another. The only dependence between particles is captured in the $\xi^{(i)}(j,k)$. At time zero, each particle that is still alive is randomly and independently sampled with a probability q. Together, $F^{(i)}$, $p^{(i)}$, $\xi^{(i)}(j,k)$ $(1 \leq i, j, k \leq r)$, and q determine the model. See Fig. 3.

In the multi-type case, the $\xi^{(i)}(j,k)$ can be chosen to allow particles to inherit properties from their parent particles. For example, a fast diversifying species may tend to have fast diversifying descendants. The random sampling at time zero models the choice by researchers of some species from a clade, as discussed in Sect. 1.2.

It will be assumed that $F^{(i)}(0) = 0$ for all i, that is, there is zero probability of a two or more simultaneous divisions. It will also be assumed that each $F^{(i)}(x)$ has a derivative $f^{(i)}(x)$ for $x > 0$. It is possible to develop the theory with more general distributions, but this is not likely to be of biological interest. When there is only one type, the superscripts will be dropped, so that $F = F^{(1)}$, and so on. Note that the assumptions on $F^{(i)}(x)$ mean that the probability of two or more events occuring in $[s, s + \delta s)$ is of order δs^2.

The following functions give the probabilities of certain events that the ancestor may produce. $P_n^{(i)}(t)$ is the probability that exactly n descendants of the ancestor appear in the sample. For convenience set $X^{(i)}(t) = P_0^{(i)}(t)$ and $L^{(i)}(t) = P_1^{(i)}(t)$.

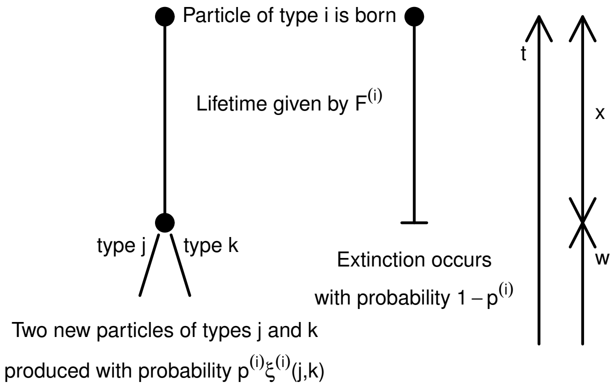

Fig. 3 The model with the main parameters, and the variables t, x, and w.

Next, functions $b^{(i)}(s,t)$ for $0 < s \leq t$ provide the probability that some descendant of the ancestor of type i divides at a time s, and that the ancestor produces no other particles which appear in the sample. More formally, let $\Delta(x)$ be the set of particles extant at time x which are descendants of the ancestor. Let δs be a short interval of time. Then in the limit as $\delta s \to 0$, $b^{(i)}(s,t)\delta s$ is the probability that one member of $\Delta(x)$ divides with $x \in [s, s+\delta s)$ and that no other member of $\Delta(x)$ leaves a particle in the sample.

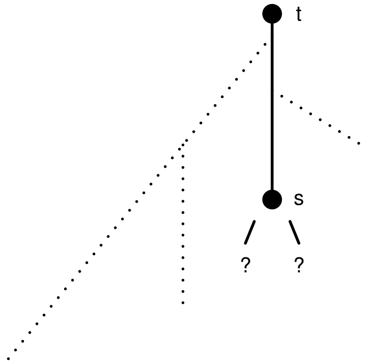

Fig. 4 Illustration for $b^{(i)}(s,t)$.

Fig 4 illustrates the meaning of $b^{(i)}(s,t)$. For each i, the function $b^{(i)}(s,t)$, regarded as a function of s for fixed t, can be seen as the density of 'candidate' nodes in the reconstructed tree which may be produced by the ancestor. A candidate node is a node which is produced as the tree grows, and which will appear in the reconstructed tree as the most recent common ancestor of the species in the sample, if and only if both its descendants leave species in the sample. Whether a candidate node at s becomes a node in the reconstructed tree depends on the fate of the two particles born at s, but not on any other part of the tree. Regarded in this way, $b^{(i)}(s,t)$ is not a probability density since the event may happen more than once at different values of s, or not at all.

The next step is to find integral equations satisfied by these functions, starting with $X^{(i)}(t)$. It may help to refer to Fig. 3. The particle born at t may live without terminating until the present, and then fail to be sampled, which happens with probability $(1-q)(1-F^{(i)}(t))$. Otherwise there is a termination, and the first one is at some time $w \in [0,t]$. If this termination is an extinction, there will be nothing appearing in the sample, and the probability of this is

$$\int_0^t (1-p^{(i)}) f^{(i)}(t-w)\mathrm{d}w = (1-p^{(i)})F^{(i)}(t).$$

Finally, if the termination at w is a division, then both new particles must produce zero particles in the sample. If the two new particles, distinguished as ‘left’ and ‘right’, are of types j and k, the probability that nothing appears in the sample is $X^{(j)}(w)X^{(k)}(w)$. Thus

$$\begin{aligned} X^{(i)}(t) = {} & 1-q+(q-p^{(i)})F^{(i)}(t)+ \\ & p^{(i)} \int_0^t f^{(i)}(t-w) \sum_{j=1}^{r} \sum_{k=1}^{r} \xi^{(i)}(j,k) X^{(j)}(w) X^{(k)}(w)\mathrm{d}w \end{aligned} \tag{1}$$

The integral equations for the $L^{(i)}(t)$ are found in a similar way. The ancestor may live without terminating until the present, and then be sampled, which happens with probability $q(1-F^{(i)}(t))$. Otherwise there is a termination, the first one is at some time $w \in [0,t]$, and this must be a division. Then one particle born at w must produce zero particles in the sample, and the other must produce exactly one (but it does not matter which particle does what). Thus

$$\begin{aligned} L^{(i)}(t) = {} & q(1-F^{(i)}(t))+ \\ & 2p^{(i)} \int_0^t f^{(i)}(t-w) \sum_{j=1}^{r} \sum_{k=1}^{r} \xi^{(i)}(j,k) X^{(j)}(w) L^{(k)}(w)\mathrm{d}w \end{aligned} \tag{2}$$

For $P_n^{(i)}(t)$ when $n \geq 2$ at least one speciation must occur in $[0,t]$ and so

$$P_n^{(i)}(t) = p^{(i)} \int_0^t f^{(i)}(t-w) \sum_{m=0}^{n} \sum_{j=1}^{r} \sum_{k=1}^{r} \xi^{(i)}(j,k) P_m^{(j)}(w) P_{n-m}^{(k)}(w)\mathrm{d}w. \tag{3}$$

This can be rewritten as

$$
\begin{aligned}
P_n^{(i)}(t) = p^{(i)} \int_0^t f^{(i)}(t-w) \sum_{m=1}^{n-1} \sum_{j=1}^{r} \sum_{k=1}^{r} \xi^{(i)}(j,k) P_m^{(j)}(w) P_{n-m}^{(k)}(w) \mathrm{d}w \; + \\
2p^{(i)} \int_0^t f^{(i)}(t-w) \sum_{j=1}^{r} \sum_{k=1}^{r} \xi^{(i)}(j,k) X^{(k)}(w) P_n^{(j)}(w) \mathrm{d}w. \qquad (4)
\end{aligned}
$$

which shows that once the $P_m^{(i)}$ are known for $m < n$, this is an equation for $P_n^{(i)}$ of similar form to that for $L^{(i)}$.

It is also possible to write down equations for the probabilities of numbers of different types in a similar fashion. For example, if there are two types and $P_{ab}^{(i)}(t)$ is the probability that there are a of type 1 and b of type 2 produced by the ancestor, then there are equations for all $a, b \geq 0$ and all types i.

Finally consider $b^{(i)}(s,t)$. The particle born at t may live without terminating until s, and then divide at s. Otherwise there is a termination before s, the first one is at some time $w \in [0,s)$, which must be a division. Then one particle born at w must produce zero particle in the sample and the other must lead to a division at s. Thus

$$
\begin{aligned}
b^{(i)}(s,t) = p^{(i)} f^{(i)}(t-s) + \\
2p^{(i)} \int_s^t f^{(i)}(t-w) \sum_{j=1}^{r} \sum_{k=1}^{r} \xi^{(i)}(j,k) X^{(j)}(w) b^{(k)}(s,w) \mathrm{d}w \qquad (5)
\end{aligned}
$$

3 Solving the Integral Equations

3.1 Some General Observations

Equations (2), (3), and (5) are of a type known as Volterra integral equations. Equation (1) is more complex since products of the $X^{(i)}$ appear in the integral. The key feature that makes all these equations relatively easy to solve numerically is that all the integrals are convolutions, which means that they can be carried out efficiently using Fourier transforms.

Equations (1), (2), (3), and (5) each have a unique solution as made precise by the following theorem.

Theorem 1 *Equations (1), (2), and (3) each have a solution (that is, a set of solutions for $1 \leq i \leq r$) for $X^{(i)}$, $L^{(i)}$, and $P_n^{(i)}$ respectively. Under the further assumption that the $f^{(i)}$ are bounded, (5) has a solution for $b^{(i)}$. All these solutions are unique among bounded functions.*

In the single-type case, the solution to equation (1) has $X(0) = X^{(1)}(0) = 1 - q$, the probability that a particle which reaches the present is not included in the sample. The asymptotic behaviour of $X(t)$ is described by the following theorem.

Theorem 2 *In the single-type case with* $p > 1/2$, $X(t) \to (1-p)/p$ *as* $t \to \infty$. *Furthermore,* X *is monotonic, either increasing from* $1-q$ *to* $(1-p)/p$ *if* $1-q < (1-p)/p$, *decreasing from* $1-q$ *to* $(1-p)/p$ *if* $1-q > (1-p)/p$, *or remaining constant if* $1-q = (1-p)/p$. *If* $p \leq 1/2$ *then* $X(t)$ *increases monotonically to 1 as* $t \to \infty$.

3.2 A Special Case: The Constant Rate Birth-Death Process

In the constant rate birth-death process, there is only one type, and $F(t) = 1 - \exp(-\alpha t)$ for some $\alpha > 0$ so that the division rate is αp and the extinction rate is $\alpha(1-p)$ for all species at all times.

Theorem 3 *Under the constant rate birth-death model equations (1), (2), (3), and (5) have solutions* $\tilde{X}$, $\tilde{L}$, $\tilde{P}_n$, $\tilde{b}$ *for* X, L, P_n, b *respectively, as follows. In the non-critical case, when* $p \neq 1/2$, *then*

$$\begin{aligned}
\tilde{X}(t) &= \frac{q(1-p) + (2p-1-qp)e^{(1-2p)\alpha t}}{D(t)} \\
\tilde{L}(t) &= \frac{q(2p-1)^2 e^{(1-2p)\alpha t}}{D(t)^2} \\
\tilde{P}_n(t) &= \tilde{L}(t)\left(\frac{pq(1-e^{(1-2p)\alpha t})}{D(t)}\right)^{n-1} \\
\tilde{b}(s,t) &= \alpha p \tilde{L}(t)/\tilde{L}(s)
\end{aligned}$$

where $D(t) = qp + (2p-1-qp)e^{(1-2p)\alpha t}$. *In the critical case when* $p = 1/2$, *then*

$$\begin{aligned}
\tilde{X}(t) &= 1 - \frac{q}{1+pq\alpha t} \\
\tilde{L}(t) &= \frac{q}{(1+pq\alpha t)^2} \\
\tilde{P}_n(t) &= \tilde{L}(t)\left(\frac{pq\alpha t}{1+pq\alpha t}\right)^{n-1} \\
\tilde{b}(s,t) &= \alpha p \tilde{L}(t)/\tilde{L}(s)
\end{aligned}$$

3.3 Numerical solutions

Equations (1), (2), (3), and (5) can be solved numerically by an iterative procedure which uses the same idea as the proof of Theorem 1 (see Sect. 6). For example, for $L^{(i)}(t)$, set $L_0^{(i)}(t) = q(1 - F^{(i)}(t))$, and define

$$L_{m+1}^{(i)}(t) = L_0^{(i)}(t) + 2p^{(i)} \int_0^t f(t-w) \sum_{j=1}^{r} \sum_{k=1}^{r} \xi^{(i)}(j,k) X^{(j)}(w) L_m^{(k)}(w) \mathrm{d}w \quad (6)$$

for $m \geq 0$ and $1 \leq i \leq r$. Thus all the $L_1^{(i)}$ are found, then the $L_2^{(i)}$, and so on. The proof of Theorem 1 guarantees that this method converges to a unique solution, but does not give bounds on the speed of convergence.

Equations (5) can be regarded as a set of equations in one variable, one for each fixed s. To make this explicit, let $c_s^{(i)}(x) = b^{(i)}(s, s+x)$ so that $b^{(i)}(s,t) = c_s^{(i)}(t-s)$ and $X_s^{(i)}(x) = X^{(i)}(s+x)$, equation (5) can be written as

$$\begin{aligned} c_s^{(i)}(x) = {} & p^{(i)} f^{(i)}(x) + \\ & 2p^{(i)} \int_0^x f^{(i)}(x-w) \sum_{j=1}^{r} \sum_{k=1}^{r} \xi^{(i)}(j,k) X_s^{(j)}(w) c_s^{(k)}(w) \mathrm{d}w \end{aligned}$$

where the $c_s^{(i)}$ are functions of one variable.

The integrals in all cases are convolutions, and the FFT can therefore be used to do these, but this does mean that time must be discretized into equal steps. Experience so far indicates that the method usually converges to double-precision accuracy in around one hundred iterations if the process is not close to critical (either sub-critical or super-critical). For processes that are close to critical, more are needed. In general, the discretization of time produces a larger error, so the main speed-accuracy trade-off is in choosing the length of the time step. For most of the results presented here 2^{14} time steps were found to be enough to achieve errors of one percent or less. Equation (3) must be solved successively for $P_2^{(i)}, P_3^{(i)}, \ldots$. This might be expected to lead to numerical problems, since errors can accumulate. In practice this does not seem to happen.

4 Illustrations of Use

4.1 Models for macroevolution

When considering tree topologies only, with no node times, the balance can be described in terms of splitting probabilities [2]. These are probabilities $s(n,m)$ for each $n = 2,3,4,\ldots$, and each m with $1 \leq m \leq n-1$ which give the probability that a node with n tips splits as $m : n-m$, so that the ‘left’ descendant has m tips and the ‘right’ descendant has $n-m$ tips. Thus for the example in Sect. 1.1 with 4 tips, $s(1,3) = s(2,2) = s(3,1) = 1/3$ for the case of the exponential distribution, while $s(1,3) = s(3,1) \approx .38$ and $s(2,2) \approx .24$ for the gamma distribution with shape parameter 0.5. In the branching processes considered here, these probabilities are also functions of time: it is possible that an ancient node with n tips has different splitting probabilities to a recent node of the same size. For the one-type case, the functions $P_m(t)$ were found for $1 \leq m \leq n$, and the following formula holds when assuming a uniform prior for t.

$$s(n,m) = \frac{\int_0^\infty P_m(t) P_{n-m}(t) \mathrm{d}t}{\sum_{k=1}^{n-1} \int_0^\infty P_k(t) P_{n-k}(t) \mathrm{d}t}$$

For the two-type case, the following formula holds when assuming a uniform prior for t.

$$s(n,m) = \eta_1 \frac{\int_0^\infty P_m^{(1)}(t) P_{n-m}^{(1)}(t)\mathrm{d}t}{\sum_{k=1}^{n-1}\int_0^\infty P_k^{(1)}(t) P_{n-k}^{(1)}(t)\mathrm{d}t} + \eta_2 \frac{\int_0^\infty P_m^{(2)}(t) P_{n-m}^{(2)}(t)\mathrm{d}t}{\sum_{k=1}^{n-1}\int_0^\infty P_k^{(2)}(t) P_{n-k}^{(2)}(t)\mathrm{d}t}$$

The values η_1 and η_2 are the asymptotic proportions of the two types, described later. The resulting splitting probabilities summarize the shape of the topologies produced by the process.

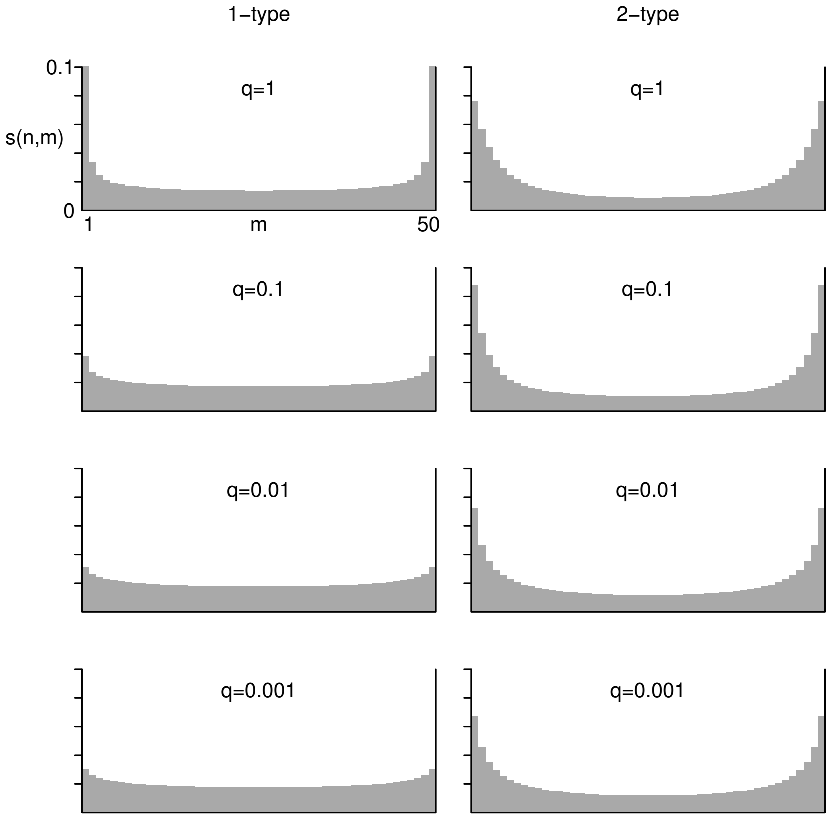

Fig. 5 Splitting probabilities $s(n,m)$ for two macroevolution models with $n = 51$ tips, for several values of q, the sampling probability for extant species. The '1-type' models are age-dependent but single type, while the '2-type' models have two types but individual species behave in an age-independent manner. The vertical scale in each graph runs from 0 to 0.1 with ticks at intervals of 0.02. For comparison, note that in the constant rate birth-death model $s(n,m) = 1/(n-1) = 0.02$ for all q.

Fig. 5 shows the splitting probabilities for two models. The parameters are chosen as follows. Since apart from mass extinctions and their aftermath, the total number of species increases slowly, models for macroevolution should be super-critical but close to critical. If they are too close to being critical, the expected time for a large tree to be produced becomes unrealistically long. [12, Table 12.1] has a list of about 60 groups of species together with estimated diversification rates. There is a large range but the median value is about 0.17 My^{-1}. An extinction rate of around 0.25, corresponding to an expected species duration of 4 My can be estimated from [29,25]. Together this gives a ratio of extinction rate to speciation rate of $0.25/(0.25+0.17) \approx 0.6$. The data in [25, Fig. 4] for 12 groups of species gives a ratio of about 0.65. Others have estimated higher values, often around 0.9 (e.g. [6,5]). A compromise is chosen here: the probabilities for speciation at termination for the models were set to 0.55, giving a ratio of extinction rate to speciation rate of $0.45/0.55 \approx 0.8$. The other parameters were chosen to approximately match the degree of imbalance observed in empirical trees.

The single-type models have a lifetime distribution given by a Pareto distribution with cumulative distribution function $F(x) = 1 - (1+x)^{-2}$ and density $f(x) = 2(1+x)^{-3}$. As mentioned above, $p = 0.55$.

The 2-type models have types 1 and 2 ('fast diversifying' and 'slow diversifying') both of which have exponential lifetime distributions. There are 6 parameters. p_1 and p_2 are the respective probabilities of division upon termination. v_1 and v_2 are parameters for the lifetimes. c_{12} gives the probability that a parent of type 1 produces a descendant of type 2 (independently for both descendants) and similarly for c_{21}. In the notation of Sect. 2, $p^{(i)} = p_i$, $f^{(i)}(t) = v_i \exp(-v_i t)$, $\xi^{(1)}(1,1) = (1-c_{12})^2$, $\xi^{(1)}(1,2) = \xi^{(1)}(2,1) = c_{12}(1-c_{12})$, $\xi^{(1)}(2,2) = c_{12}^2$, with similar expressions for the $\xi^{(2)}(i,j)\ (1 \le i,j \le 2)$. The values $p_1 = p_2 = 0.55$, $c_{12} = c_{21} = 0.1$, $v_1 = 3.0$ and $v_2 = 0.3$ were used for the 2-type model. Thus the fast type is ten times as fast as the slow type.

From results in [7, Section V.7] the asymptotic proportion of the two types 1 and 2 can be found from a matrix R. R_{ij} is the expected rate of production of particles of type j from a single particle of type i. In this case,

$$R = \begin{pmatrix} [2p_1(1-c_{12})-1]v_1 & 2p_1 c_{12} v_1 \\ 2p_2 c_{21} v_2 & [2p_2(1-c_{21})-1]v_2 \end{pmatrix}$$

The left eigenvector of R corresponding to the largest eigenvalue λ_{max} gives the asymptotic proportion (η_1, η_2) of the two types, and the root was assumed to have a type chosen at random from the two types according to this proportion. The process is critical if $\lambda_{max} = 0$, and supercritical if $\lambda_{max} > 0$.

The results show that the balance of the single-type age-dependent models increase as the trees are more sparsely sampled, that is, as q tends to zero. The same trend for the 1-type case was observed with results using other L-shaped lifetime distributions (results not shown) and is accordance with results derived from simulations in [24]. In the 2-type case, balance initially decreases slightly as q is reduced, and then increases slightly. This initial decrease followed by an increase seems typical for 2-type models for other values of the parameters, although the amounts of the change and the values of q at which they occur can vary considerably (results not shown).

Currently, there is insufficient data to be certain about what behavior is typical of empirical species trees. However it does seem clear that balance does not decrease with sampling, so that the 1-type models are unsuitable, under the assumption that sampling is uniformly at random.

4.2 Phylogenetic Analysis

Once $L(t)$ and $b(s,t)$ are known, it is possible to calculate the probability of observing a reconstructed tree. Denote the tree topology by τ, the number of tips by $|\tau|$, and the internal node times by $T = t_1, ...t_{|\tau|-1}$. The formulas below provide the conditional joint density $\pi(\tau, T \mid i, t)$ for the complete process of a multi-type age-dependent branching process initiated with a species of type i born at $t_0 = t$, and then sampled uniformly at time 0. The density of the tree is found recursively, as in [19]. When $|\tau| > 1$ the topologies and node times of two subtrees are denoted by (τ_a, T_a) and (τ_b, T_b). For $|\tau| > 1$ the result is

$$\pi(\tau, T|i, t) = \nu(\tau) b^{(i)}(t, s) \sum_{j=1}^{r} \sum_{k=1}^{r} \xi^{(i)}(j, k) \pi(\tau_a, T_a|j, s) \pi(\tau_b, T_b|k, s)) \qquad (7)$$

and for $|\tau| = 1$ it is

$$\pi(\tau, T|i, t) = L^{(i)}(t) \qquad (8)$$

where s is the time of the first speciation, and $\nu(\tau)$ is 2 if the two subtrees are distinguishable, and 1 otherwise. Note that the existence of the derivatives $f^{(i)}$ of $F^{(i)}$ implies that the probability of two nodes having exactly the same time is zero. This means that the two subtrees at each node are distinguishable except in the case of a cherry (ie where both subtrees are single branches). Even in the case of a cherry, $\nu(\tau)$ is 2 if there is more than one type and they are distinguishable.

If there is only one type, the result simplifies to

$$\pi(\tau, T|t) = 2^{n-1-m} \prod_{i \in E} L(a_i) \prod_{i \in I} b(d_i, a_i) \qquad (9)$$

where m is the number of cherries, E is the set of external branches, I is the set of internal branches, and branch i begins at time a_i and ends at time d_i (each of which is equal to one of $t_1, ...t_{n-1}$).

The formulas above are given for a tree initiated by a single particle. The birth of this particle is called the tree origin [18]. In phylogenetic analysis, the tree is usually modeled as starting at the first division, that is, the tree root. One can obtain the probability for a tree started at the root by multiplying the probabilities for the 'left' and 'right' subtrees which both have origins at the root.

4.3 Cell proliferation

A model for the generation of oligodendrocytes (brain cells) from their immediate progenitor cells (stem cells) is used as an illustration. The model is the 'full model' taken from [22], using parameter values as estimated there. There are 4 cell types, which I label A_2, A_1, A_0, and B, where A_2, A_1, A_0 represent oligodendrocyte-type 2 astrocyte progenitor cells, and B represents oligodendrocytes. The subscripts on the As are the number of cell cycles the cell must go through before it is competent to produce a oligodendrocyte. Type A_2 always divides, and produces two A_1 cells after a lifetime distributed as a gamma G_A. Type A_1 is very similar: it always divides, and produces two A_0 cells after the same lifetime distribution G_A. Type A_0 has more complex behaviour. With probability p, an A_0 cell produces two more type A_0 cells after a lifetime distributed as G_A. With probability $1 - p$ it differentiates into a type B after a lifetime distributed as a different gamma G_B. Type B cells have an infinite lifetime. At the start of the experiment, a mixture of types A_2, A_1, A_0 is present.

An alternative but equivalent way of describing the behaviour of the A_0 cells is to regard them as being of two types A_{0A}, and A_{0B}. These are produced by type A_1 cells in proportion $p : 1 - p$. A_{0A} cells always produce more A_{0A} and A_{0B} cells, again in proportion $p : 1 - p$, while A_{0B} cells always produce B cells. This way of looking at the situation makes the model fit into the scheme described in Sect. 2. However, I will use the former description where there is a single complex type A_0 in the following.

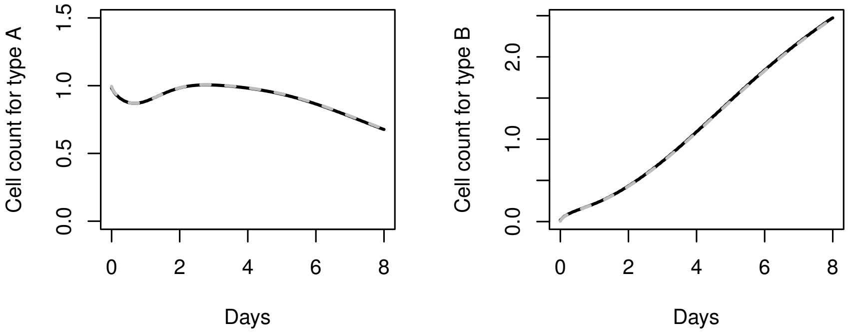

Fig. 6 These graphs show the predicted mean cell counts for cell types A and B. In both cases, results using 512 and 4096 time steps are superimposed, demonstrating that 512 time steps provides good accuracy.

There are no cell deaths, and in the notation of Sect. 2, $q = 1$, so that $X^{(i)}(t) = 0$ for all types, which simplifies the situation. The following probabilities can be defined. Let $P_{ab}(t)$ be the probability that a single type A_0 cell produces a cells of type A_0 and b of type B after a time t. Let $P_{jab}(t)$ be the probability that a single type A_1 cell produces j cells of type A_1, a cells of type A_0 and b of type B. Let $P_{ijab}(t)$

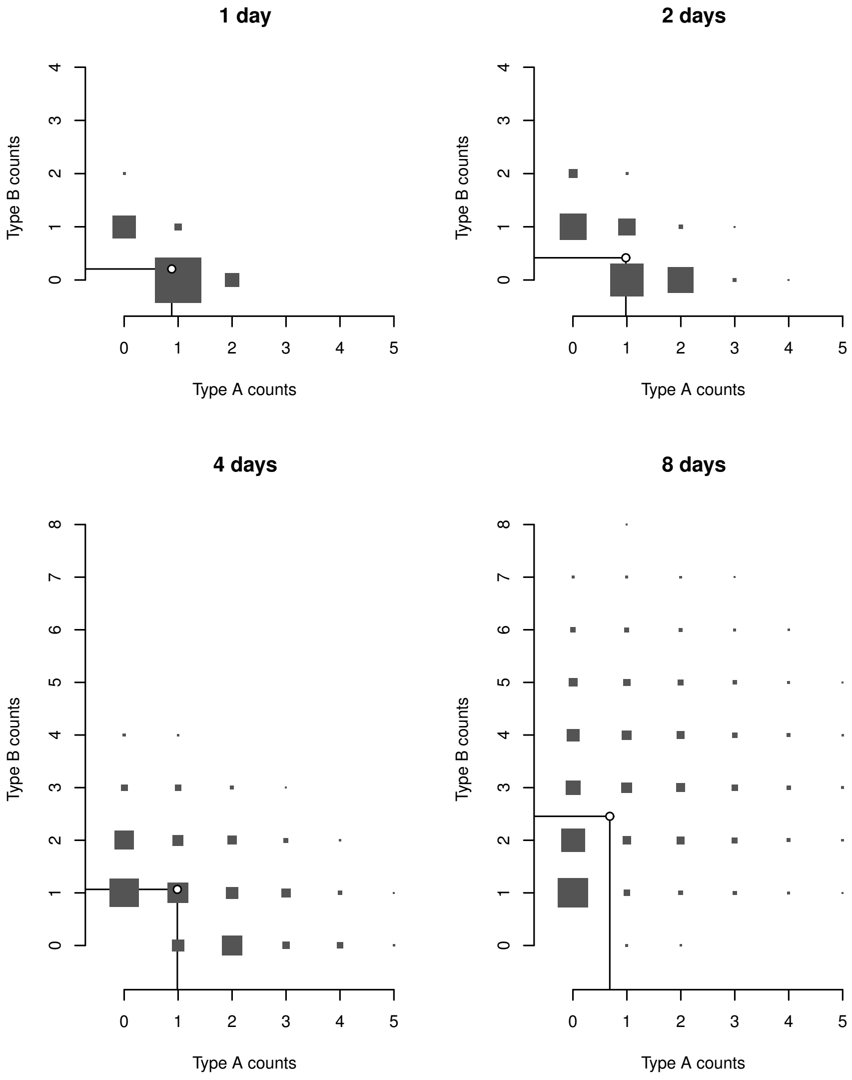

Fig. 7 The predicted probabilities for the behavior of a single type A cell over 1,2,4 and 8 days. The area of each rectangle shows the proportion of cells predicted to produce the numbers of type A and B cells. The mean values are also shown.

be the probability that a single type A_2 cell produces i cells of type A_2, j cells of type A_1, a cells of type A_0 and b of type B. It can be seen that these probabilities are zero except for the following cases.

$$P_{10}(t) = p(1 - G_A(t)) + (1 - p)(1 - G_B(t))$$

$$P_{01}(t) = (1-p)G_B(t)$$

$$P_{ab}(t) = p\int_0^t g_A(t-w)\sum_{i=0}^{a}\sum_{j=0}^{b} P_{ij}(w)P_{a-i\ b-j}(w)\mathrm{d}w \quad (a+b>1)$$

$$P_{100}(t) = 1 - G_A(t)$$

$$P_{0ab}(t) = \int_0^t g_A(t-w)\sum_{i=0}^{a}\sum_{j=0}^{b} P_{ij}(w)P_{a-i\ b-j}(w)\mathrm{d}w \quad (a+b>1)$$

$$P_{1000}(t) = 1 - G_A(t)$$

$$P_{0200}(t) = \int_0^t g_A(t-w)(1-G_A(w))^2\mathrm{d}w$$

$$P_{01ab}(t) = 2\int_0^t g_A(t-w)(1-G_A(w))P_{0ab}(w)\mathrm{d}w \quad (a+b>0)$$

$$P_{00ab}(t) = \int_0^t g_A(t-w)\sum_{i=0}^{a}\sum_{j=0}^{b} P_{0ij}(w)P_{0\ a-i\ b-j}(w)\mathrm{d}w \quad (a+b>1)$$

These can be calculated for a range $0 \leq a + b \leq M$. Approximately $3M^2/2$ convolutions are required. Fig. 6 shows the mean counts for type A (types A_2, A_1, A_0 combined) and type B over a period of 8 days. These correspond to the lower two panels of Figure 2 in [22]. Fig. 7 shows the more detailed information that is available from the current approach.

5 Discussion

This paper has introduced new methods for calculating quantities of interest in biological processes and structures which can be modeled using branching processes. It is hoped that they will prove useful in various areas of biology. The code to carry out the calculations in Sect. 4.1 and Sect. 4.3 was written in R [36]. The R scripts are available from the author on request.

Sect. 4.1 is a preliminary investigation into multi-type age-dependent branching processes as models for macroevolution. There is clearly scope for much further work here. Sect. 4.2 provides a method which could be used to provide a prior probability for trees in a phylogenetic analysis program like BEAST [14].

The example of cell proliferation in 4.3 illustrates how the same method can be used in other areas. It is complementary to other approaches such as simulation and saddlepoint methods.

The equations in Sect. 4.2 could also be useful in studies of macroevolution. So far as I am aware, these provide the most practical method currently available for calculating the joint probability for topology and node times for multi-type age-dependent branching processes. As described in Sect. 1.1 there has been much interest in the balance of phylogenetic trees. A recent study [40] investigates the distribution of branch lengths in a collection of phylogenetic trees. I am not aware of any study that has used *both* topology and node times to match models to data, which should provide more stringent tests of models. As more data becomes available, especially data which includes estimates for the dates of the nodes in a tree, I believe that such studies will become common.

6 Proofs

Theorem 1 can be established using a technique similar to that used in [7, Section IV.2] for generating functions.

Lemma 1 *Let $F^{(i)}$ $(1 \leq i \leq r)$ be distributions on $[0,\infty)$ with $F^{(i)}(0+) = 0$ for all i. Let S_n be the sum of n independent random variables, each of which has a distribution equal to one of the $F^{(i)}$. Let $\gamma > 0$ be a constant. Then there is a $\theta > 0$, depending on γ, but not depending on n nor on which of the $F^{(i)}$ are involved in S_n, such that*

$$\gamma^n \Pr(S_n \leq t) \leq (1/2)^n e^{\theta t}$$

Proof For $1 \leq i \leq r$, let Z_i be a random variable with distribution $F^{(i)}$. Since $F^{(i)}(0+) = 0$, a number $\epsilon_i > 0$ can be chosen so that $F^{(i)}(\epsilon_i) < 1/(4\gamma)$. For each i, choose $\theta_i > 0$ large enough to ensure that $e^{-\theta_i \epsilon_i} < 1/(4\gamma)$. Then

$$\begin{aligned}
E(e^{-\theta_i Z_i}) &\leq \Pr(Z_i < \epsilon_i) E(e^{-\theta_i Z_i} | Z_i < \epsilon_i) \; + \; \Pr(Z_i \geq \epsilon_i) E(e^{-\theta_i Z_i} | Z_i \geq \epsilon_i) \\
&\leq F^{(i)}(\epsilon_i) E(1 | Z_i < \epsilon_i) \; + \; E(1/(4\gamma) | Z_i \geq \epsilon_i) \\
&< 1/(4\gamma) + 1/(4\gamma) = 1/(2\gamma).
\end{aligned}$$

Let $\theta = \max_i \theta_i$, so that $\gamma E(e^{-\theta Z_i}) < 1/2$ for all i. Then

$$\begin{aligned}
\gamma^n \Pr(S_n \leq t) &\leq \gamma^n \Pr(e^{-\theta S_n} \geq e^{-\theta t}) \\
&\leq \gamma^n e^{\theta t} E(e^{-\theta S_n}) \\
&\leq e^{\theta t} \prod_{j=1}^{n} \gamma E(e^{-\theta Y_j})
\end{aligned}$$

where each Y_j has distribution equal to one of the $F^{(i)}$. The second line in the above derivation uses Markov's inequality, and the third line uses the independence of the Y_j. Hence $\gamma^n \Pr(S_n \leq t) \leq (1/2)^n e^{\theta t}$. $\square$

Proof of Theorem 1. Only the case of equation (1) for $X^{(i)}$ will be proved in detail. Define $X_0^{(i)}(t) = 1 - q + (q - p^{(i)})F^{(i)}(t)$ and for $m \geq 0$ let

$$X_{m+1}^{(i)}(t) = X_0^{(i)}(t) + p^{(i)} \int_0^t f^{(i)}(t-w) \sum_{j=1}^{r} \sum_{k=1}^{r} \xi^{(i)}(j,k) X_m^{(j)}(w) X_m^{(k)}(w) \mathrm{d}w. \tag{10}$$

Firstly we show by induction that $0 \leq X_m^{(i)}(t) \leq 1$ for all i and m. Since $X_0^{(i)}(t)$ is linear in $F^{(i)}(t)$, it always lies between $1-q$ and $1-p^{(i)}$, so $0 \leq X_0^{(i)}(t) \leq 1$ for all i. It follows by induction that $X_m^{(i)}(t) \geq 0$ for all i and m, since $X_{m+1}^{(i)}(t)$ is a sum of non-negative terms. If $X_m^{(i)}(t) \leq 1$ for all i, then using $\sum_{j=1}^{r} \sum_{k=1}^{r} \xi^{(i)}(j,k) = 1$,

$$\begin{aligned}
X_{m+1}^{(i)} &= 1 - q + (q - p^{(i)})F^{(i)}(t) + \\
&\qquad p^{(i)} \int_0^t f^{(i)}(t-w) \sum_{j=1}^{r} \sum_{k=1}^{r} \xi^{(i)}(j,k) X^{(j)}(w) X^{(k)}(w) \mathrm{d}w \\
&\leq 1 - q + (q - p^{(i)})F^{(i)}(t) + p^{(i)} \int_0^t f^{(i)}(t-w) \mathrm{d}w \\
&= 1 - q + (q - p^{(i)})F^{(i)}(t) + p^{(i)} F^{(i)}(t) \\
&= 1 - q + qF^{(i)}(t) \leq 1,
\end{aligned}$$

so it follows by $0 \leq X_m^{(i)}(t) \leq 1$ for all i and m.

Now we show by induction that for all i

$$|X_{m+1}^{(i)}(t) - X_m^{(i)}(t)| \leq (2r^2)^{m+1} \Pr(S_{m+1} \leq t) \tag{11}$$

for some random variable S_{m+1} of the type in Lemma 1. For $m = 0$, equation (10) gives

$$X_1^{(i)}(t) - X_0^{(i)}(t) = p^{(i)} \int_0^t f^{(i)}(t-w) \sum_{j=1}^{r} \sum_{k=1}^{r} \xi^{(i)}(j,k) X_0^{(j)}(w) X_0^{(k)}(w) \mathrm{d}w.$$

Now, for every i, j, k and all w, the numbers $p^{(i)}$, $\xi^{(i)}(j,k)$ and $X_0^{(j)}(w)$ are all in the interval $[0,1]$, so

$$|X_1^{(i)}(t) - X_0^{(i)}(t)| \leq \int_0^t f^{(i)}(t-w) r^2 \mathrm{d}w \leq r^2 F^{(i)}(t) \leq (2r)^2 F^{(i)}(t)$$

establishing (11) for $m = 0$. If $m > 0$,

$$|X_{m+1}^{(i)}(t) - X_m^{(i)}(t)| \leq$$

$$p^{(i)} \int_0^t f^{(i)}(t-w) \sum_{j=1}^{r} \sum_{k=1}^{r} \xi^{(i)}(j,k) |X_m^{(j)}(w) X_m^{(k)}(w) - X_{m-1}^{(j)}(w) X_{m-1}^{(k)}(w)| \mathrm{d}w$$

and since

$$\begin{aligned}
&|X_m^{(j)}(w)X_m^{(k)}(w) - X_{m-1}^{(j)}(w)X_{m-1}^{(k)}(w)| \\
&= |\ \big(X_m^{(j)}(w) - X_{m-1}^{(j)}(w)\big)\ X_m^{(k)}(w)\ +\ X_{m-1}^{(j)}(w)\ \big(X_m^{(k)}(w) - X_{m-1}^{(k)}(w)\big)\ | \\
&\le |X_m^{(j)}(w) - X_{m-1}^{(j)}(w)| + |X_m^{(k)}(w) - X_{m-1}^{(k)}(w)| \\
&\le 2\max_j |X_m^{(j)}(w) - X_{m-1}^{(j)}(w)|
\end{aligned}$$

and the numbers $p^{(i)}$, $\xi^{(i)}(j,k)$ and $X_0^{(j)}(w)$ are all in $[0,1]$, it follows using induction that

$$\begin{aligned}
|X_{m+1}^{(i)}(t) - X_m^{(i)}(t)| &\le 2r^2 \int_0^t f^{(i)}(t-w) \max_j |X_m^{(j)}(w) - X_{m-1}^{(j)}(w)| \mathrm{d}w \\
&\le 2r^2 \int_0^t f^{(i)}(t-w)(2r^2)^m \Pr(S_m \le w) \mathrm{d}w \\
&\le (2r^2)^{m+1} \int_0^t f^{(i)}(t-w) \Pr(S_m \le w) \mathrm{d}w.
\end{aligned}$$

The integral in the last line can be recognized as a convolution, representing the cumulative distribution function for the sum of two independent random variables S_m and Z_i, where Z_i has distribution $F^{(i)}$. Thus

$$|X_{m+1}^{(i)}(t) - X_m^{(i)}(t)| \le (2r^2)^{m+1} \Pr(S_m + Z_i \le t)$$

which establishes (11). Now Lemma 1 with $\gamma = 2r^2$ shows that

$$|X_{m+1}^{(i)}(t) - X_m^{(i)}(t)| \le (1/2)^{m+1} e^{\theta t}$$

for some $\theta > 0$. The sequence of functions $X_m^{(i)}$ is therefore a Cauchy sequence and so converges to some function $X_\omega^{(i)}$. Now let $m \to \infty$ in equation (10). Since the convergence of $X_m^{(i)}$ to $X_\omega^{(i)}$ is uniform on any finite interval $[0, t_0]$, limits can be taken through the integral to show that the $X_\omega^{(i)}$ form a solution.

Uniqueness follows by considering $V^{(i)}(t) = |Y^{(i)}(t) - X_\omega^{(i)}(t)|$ where $Y^{(i)}$ is another bounded solution to (1). Details are omitted.

The other equations (2), (3), and (5) can be dealt with similarly. For (3), the proof uses induction on n using equation (4). In the case of (5), the proof establishes existence and uniqueness for a set of one-dimensional equations in t, one for each value of s. □

Proof of Theorem 2. The proof uses general results from Athreya and Ney's book [7]. Let $N_n(t)$ be the probability that there are exactly n particles surviving at time zero given an ancestor at time t. $N_n(t)$ is thus like $P_n(t)$ but does not take into account the sampling at time zero. Let

$$F(s,t) = \sum_{n=0}^{\infty} N_n(t)s^n. \tag{12}$$

$F(s,t)$ is known as the generating function for the process [7, p138].

Let $g(s)$ represent the generating function governing particle production as described in [7, p137] and let ρ be the smallest solution of the equation $g(s) = s$ as described in [7, Section 1.3]. (Note that $g(s)$ is $f(s)$ in Athreya and Ney's notation, and ρ is q in their notation.) Restating [7, Theorem 3.1, p143], we have

Theorem 4 *If* $0 \leq s \leq \rho$*, then* $F(s,t)$ *is nondecreasing in* t*, and* $F(s,t) \to \rho$ *as* $t \to \infty$*. If* $\rho \leq s < 1$*, then* $F(s,t)$ *is nonincreasing in* t*, and* $F(s,t) \to \rho$ *as* $t \to \infty$*.*

In the case of a binary tree, where a particle can only produce zero or two descendants, $g(s) = (1-p) + ps^2$, and it follows that $\rho = (1-p)/p$ if $p > 1/2$ and $\rho = 1$ if $p \leq 1/2$. Theorem 2 now follows from the observation that

$$X(t) = \sum_{n=0}^{\infty} N_n(t)(1-q)^n = F(1-q,t).$$

□

Proof of Theorem 3. In the constant rate birth-death process, equations (1), (2), (3), and (5) simplify to

$$X(t) = 1 - p + (p-q)e^{-\alpha t} + p\int_0^t \alpha e^{-\alpha(t-w)}X(w)^2 \mathrm{d}w \tag{13}$$

$$L(t) = qe^{-\alpha t} + 2p\int_0^t \alpha e^{-\alpha(t-w)}X(w)L(w)\mathrm{d}w \tag{14}$$

$$P_n(t) = p\int_0^t \alpha e^{-\alpha(t-w)} \sum_{m=0}^{n} P_m(w)P_{n-m}(w)\mathrm{d}w. \tag{15}$$

$$b(s,t) = p\alpha e^{-\alpha(t-s)} + 2p\int_s^t \alpha e^{-\alpha(t-w)}X(w)b(s,w)\mathrm{d}w \tag{16}$$

The critical and non-critical cases can be handled together. For equation (13), let $g(x) = e^{\alpha x}(\tilde{X}(x) - (1-p))$. Then $g'(x)$ can be calculated as $\alpha p e^{\alpha x}\tilde{X}(x)^2$, so

$$\begin{aligned} p\int_0^t \alpha e^{-\alpha(t-w)}\tilde{X}(w)^2 \mathrm{d}w &= e^{-\alpha t}[g(t) - g(0)] \\ &= \tilde{X}(t) - (1-p) - e^{-\alpha t}(p-q) \end{aligned}$$

as required to satisfy the equation (13).

Next, for equation (14), let $h(x) = e^{\alpha x}\tilde{L}(x)$. Then $h'(x) = 2\alpha p e^{\alpha x}\tilde{L}(x)\tilde{X}(x)$, so

$$\begin{aligned} 2p\int_0^t \alpha e^{-\alpha(t-w)}\tilde{X}(w)\tilde{L}(w)\mathrm{d}w &= e^{-\alpha t}[h(t) - h(0)] \\ &= \tilde{L}(t) - e^{-\alpha t}q \end{aligned}$$

as required to satisfy the equation. Next, for equation (15), write

$$\tilde{P}_n(t) = \tilde{L}(t)\beta(t)^{n-1}$$

for $n \geq 1$, where

$$\beta(t) = \frac{pq(1 - e^{(1-2p)\alpha t})}{qp + (2p - 1 - qp)e^{(1-2p)\alpha t}}$$

if $p \neq 1/2$ and

$$\beta(t) = \frac{pq\alpha t}{1 + pq\alpha t}$$

when $p = 1/2$. Note that the solution is already proved for $\tilde{P}_1(x) = \tilde{L}(x)$. Assume $n \geq 2$ and that the result is true for $\tilde{P}_i(x)$ with $i < n$. Let

$$j(x) = e^{\alpha x}\tilde{P}_n(x) = e^{\alpha x}\tilde{L}(x)\beta(x)^{n-1} = h(x)\beta(x)^{n-1}$$

Then, using $\beta'(x) = \alpha p\tilde{L}(x)$,

$$j'(x) = 2\alpha p e^{\alpha x}\tilde{L}(x)\tilde{X}(x)\beta(x)^{n-1} \ + \ \alpha p(n-1)e^{x}\tilde{L}(x)^2\beta(x)^{n-2}$$

so

$$\begin{aligned} & p\int_0^t \alpha e^{-\alpha(t-w)}\sum_{i=0}^{n}\tilde{P}_i(w)\tilde{P}_{n-i}(w)\mathrm{d}w \\ &= \alpha e^{-\alpha t}\int_0^t pe^{w}\sum_{i=1}^{n-1}\tilde{P}_i(w)\tilde{P}_{n-i}(w) + 2pe^{\alpha w}\tilde{P}_0(w)\tilde{P}_n(w)\mathrm{d}w \\ &= \alpha e^{-\alpha t}\int_0^t pe^{\alpha w}(n-1)\tilde{L}(w)^2\beta(w)^{n-2} + 2pe^{w}X(w)\tilde{L}(w)\beta(w)^{n-1}\mathrm{d}w \\ &= e^{-\alpha t}\int_0^t j'(w)\mathrm{d}w = e^{-\alpha t}(j(t) - j(0)) = \tilde{P}_n(t) \end{aligned}$$

as required to satisfy equation (15), so the result follows by induction.

Lastly, for equation (16), with s regarded as fixed but arbitrary, let $k(x) = e^{\alpha x}\tilde{b}(s,x) = pe^{\alpha x}\tilde{L}(x)/\tilde{L}(s)$. Then $k'(x) = 2\alpha p e^{\alpha x}\tilde{X}(x)\tilde{b}(s,x)$ so

$$\begin{aligned}
&2p\int_s^t \alpha e^{-\alpha(t-w)}\tilde{X}(w)\tilde{b}(s,w)\mathrm{d}w\\
&= e^{-\alpha t}\int_s^t k'(w)\mathrm{d}w = e^{-\alpha t}(k(t)-k(s))\\
&= \tilde{b}(s,t) - e^{-\alpha t}pe^{\alpha s}\tilde{L}(s)/\tilde{L}(s)\\
&= \tilde{b}(s,t) - \alpha pe^{\alpha(s-t)}
\end{aligned}$$

as required to satisfy the equation. □

References

1. Agapow, P.M., Purvis, A.: Power of eight tree shape statistics to detect nonrandom diversification: A comparison by simulation of two models of cladogenesis. Syst Biol. **51**, 866–872 (2002)
2. Aldous, D.: Probability distributions on cladograms. In: Random discrete structures (Minneapolis, MN, 1993), *IMA Vol. Math. Appl.*, vol. 76, pp. 1–18. Springer, New York (1996). URL http://www.stat.berkeley.edu/users/aldous/Papers/me69.pdf
3. Aldous, D., Popovic, L.: A critical branching process model for biodiversity. Adv. in Appl. Probab. **37**(4), 1094–1115 (2005)
4. Aldous, D.J.: Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. Statist. Sci. **16**, 23–34 (2001)
5. Alroy, J.: Dynamics of origination and extinction in the marine fossil record. Natl Acad Sci USA **105**(1), 11,536–11,542 (2008)
6. Arrizon, V., Tepichin, E., Ortiz-Gutierrez, M., Lohmann, A.W., Alroy, J.: Constant extinction, constrained diversification, and uncoordinated stasis in north american mammals. Palaeogeography, Palaeoclimatology, Palaeoecology **127**(1), 285–311 (1996)
7. Athreya, K.B., Ney, P.E.: Branching Processes. Springer-Verlag (1972)
8. Bellman, R., Harris, T.E.: On age-dependent binary branching processes. Ann. of Math. **55**, 280–295 (1952)
9. Berger, J.O.: Statistical Decision Theory and Bayesian Analysis, 2nd Edition. Springer (1985)
10. Blum, M.G.B., François, O.: Which random processes describe the tree of life? a large-scale study of phylogenetic tree imbalance. Syst Biol. **55**(4), 685–691 (2007)
11. Brandley, M.C., Leach, A.D., Warren, D.L., McGuire, J.A.: Are unequal clade priors problematic for Bayesian phylogenetics? Syst. Biol. **55**(1), 138–146 (2006)
12. Coyne, J.A., Orr, H.A.: Speciation. Sinauer Associates (2004)
13. Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A.: Relaxed phylogenetics and dating with confidence. PLoS Biol **4**(5), e88 (2006 May). DOI 10.1371/journal.pbio.0040088. URL http://dx.doi.org/10.1371/journal.pbio.0040088
14. Drummond, A.J., Rambaut, A.: BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology **7**(214) (2007)
15. Felsenstein, J.: Inferring Phylogenies. Sinauer Associates (2003). DOI http://dx.doi.org/10.1016/S0022-0000(02)00003-X
16. Ford, D.: Probabilities on cladogram: Introduction to the alpha model (2005). URL http://arxiv.org/abs/math/0511246
17. Ford, D., Matsen, E., Stadler, T.: A method for investigating relative timing information on phylogenetic trees. Syst. Biol. **58**, 167–183 (2009)
18. Gernhard, T.: The conditioned reconstructed process. J. Theo. Biol. **253**, 769–778 (2008)
19. Gernhard, T., Hartmann, K., Steel, M.: Stochastic properties of generalised Yule models, with biodiversity applications. J. Math. Biol. **57**, 713–735 (2008)
20. Heath, T.A., Zwickl, D.J., Kim, J., Hillis, D.M.: Taxon sampling affects inferences of macroevolutionary processes from phylogenetic trees. Syst. Biol. **57**(1), 160–166 (2008)

21. Huelsenbeck, J.P., Ronquist, F.: MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17**(8), 754–755 (2001). URL http://dblp.uni-trier.de/db/journals/bioinformatics/bioinformatics17.html
22. Hyrien, O., Chen, R., Mayer-Proöschel, M., Noble, M.: Saddlepoint approximations to the moments of multitype age-dependent branching processes, with applications. Biometrics **66**(2), 567–577 (2010)
23. Hyrien, O., Mayer-Proschel, M., Noble, M., Yakovlev, A.: A stochastic model to analyze clonal data on multi type cell populations. Biometrics **61**, 199–207 (2005)
24. Jones, G.: Tree models for macro-evolution and phylogenetic analysis. To appear in Syst. Biol. (2009)
25. Jr, J.J.S.: Rates of speciation in the fossil record. Philos. Trans. R. Soc. London B **353**, 315–326 (1998)
26. Kimmel, M., Axelrod, D.: Branching Processes in Biology. Springer (2002)
27. Kirkpatrick, M., Slatkin, M.: Searching for evolutionary patterns in the shape of a phylogenetic tree. Evolution **47**, 1171–1181 (1993)
28. Kontoleon, N.: The Markovian binary tree: A model of the macroevolutionary process. PhD Thesis, PhD Thesis, The University of Adelaide. (2006)
29. Lawton, J.H., May, R.M.: Extinction Rates. Oxford University Press, USA (1995)
30. Lepage, T., Bryant, D., Philippe, H., Lartillot, N.: A general comparison of relaxed molecular clock models. Mol Biol Evol (2007). DOI 10.1093/molbev/msm193. URL http://dx.doi.org/10.1093/molbev/msm193
31. Nedelman, J., Downs, H., Pharr, P.: Inference for an age-dependent, multitype branching-process model of mast cells. Journal of Mathematical Biology **25**, 203–226 (1987)
32. Nee, S., May, R.M., Harvey, P.H.: The reconstructed evolutionary process. Philosophical Transactions: Biological Sciences **344**(1309), 305=–311 (1994)
33. Pickett, K.M., Randle, C.P.: Strange Bayes indeed: uniform topological priors imply non-uniform clade priors. Molecular Phylogenetics and Evolution **34**, 203–211 (2005)
34. Pickett, K.M., Randle, C.P.: Are nonuniform clade priors important in Bayesian phylogenetic analysis? a response to Brandley et al. Syst. Biol. **55**(1), 147–151 (2006)
35. Pinelis, I.: Evolutionary models of phylogenetic trees. Proc Biol Sci **270**(1522), 1425–31 (2003). DOI 10.1098/rspb.2003.2374. URL http://dx.doi.org/10.1098/rspb.2003.2374
36. R Development Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2008). URL http://www.R-project.org. ISBN 3-900051-07-0
37. Stadler, T.: Evolving trees - models for speciation and extinction in phylogenetics. PhD Thesis, Technische Universität München, Zentrum Mathematik (2008)
38. Stadler, T.: On incomplete sampling under birth-death models and connections to the sampling-based coalescent. J. Theo. Biol. **261**, 58–66 (2009)
39. Steel, M., McKenzie, A.: Properties of phylogenetic trees generated by Yule-type speciation models. Math Biosci **170**(1), 91–112 (2001)
40. Venditti, C., Meade, A., Pagel, M.: Phylogenies reveal new interpretation of speciation and the red queen. Nature **463**, 349–352 (2009)
41. Yakovlev, A.Y., Yanev, N.M.: Transient Processes in Cell Proliferation Kinetics. Springer, Berlin (1989)
42. Yang, Z., Rannala, B.: Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference. J Mol Evol **43**, 304–311 (1996)
43. Yang, Z., Rannala, B.: Bayesian phylogenetic inference using DNA sequences: A Markov chain Monte Carlo method. Mol. Biol. Evol. **14**(7), 717–724 (1997)
44. Yang, Z., Rannala, B.: Branch-length prior influences Bayesian posterior probability of phylogeny. Syst. Biol. **54**(3), 455–470 (2005)