

Stochastic properties of generalised Yule models, with biodiversity applications

Tanja Gernhard · Klaas Hartmann · Mike Steel

Received: 4 March 2008 / Revised: 22 April 2008 / Published online: 29 May 2008
© Springer-Verlag 2008

Abstract The Yule model is a widely used speciation model in evolutionary biology. Despite its simplicity many aspects of the Yule model have not been explored mathematically. In this paper, we formalise two analytic approaches for obtaining probability densities of individual branch lengths of phylogenetic trees generated by the Yule model. These methods are flexible and permit various aspects of the trees produced by Yule models to be investigated. One of our methods is applicable to a broader class of evolutionary processes, namely the Bellman–Harris models. Our methods have many practical applications including biodiversity and conservation related problems. In this setting the methods can be used to characterise the expected rate of biodiversity loss for Yule trees, as well as the expected gain of including the phylogeny in conservation management. We briefly explore these applications.

T. Gernhard and K. Hartmann contributed equally.

T. Gernhard
Kombinatorische Geometrie, Technische Universität München, München, Germany
e-mail: gernhard@ma.tum.de

K. Hartmann
Biomathematics Research Centre, University of Canterbury, Christchurch, New Zealand

Present Address:
K. Hartmann (✉)
ICT Centre, CSIRO, Hobart, Australia
e-mail: klaas.hartmann@csiro.au

M. Steel
Allan Wilson Centre for Molecular Ecology and Evolution, Biomathematics Research Centre,
University of Canterbury, Christchurch, New Zealand
e-mail: m.steel@math.canterbury.ac.nz

Keywords Yule · Phylogenetic diversity · Tree · Null model · Biodiversity · Extinction · Bellman Harris

Mathematics Subject Classification (2000) 92-08 · 60J80 · 05C05

1 Introduction

The Yule model is an evolutionary model for phylogenetic trees [10,25]. Under this model each species has an equal probability of undergoing a speciation event at any given point in time. The instantaneous probability of speciation is referred to as the rate of speciation and is assumed to be constant over time. The Yule model has been widely used as a null model with which to compare real phylogenetic trees and explore evolutionary hypotheses [1,16]. Despite its wide application the Yule model has many mathematical aspects that remain uncharacterised. In this paper, we provide two approaches for calculating edge length probability distributions.

The first approach we present applies to a more general class of evolutionary models based on Bellman–Harris (BH) processes which we describe here and refer to as BH models. For trees produced by the BH model we provide methods for calculating (i) the probability density of the tree shape, (ii) of any edge length for a given tree shape and (iii) of the longest pendant edge length for a given tree shape. For Yule models, analytic solutions are obtainable, however for some BH models it may be necessary to solve the required integrals numerically. These methods extend the results in [23] and have been applied in [19]. Other related properties can be readily explored by extending our methodology.

The second approach we present utilises rank functions to obtain edge length probability densities for the Yule model. This approach was first introduced in [9] for expectations; here it is extended to give distributions and to permit a known age of the tree to be incorporated.

Our methods can be useful in many contexts including testing evolutionary hypotheses, constructing phylogenetic trees, and biodiversity conservation. In this paper, we illustrate our methods by applying them to two biodiversity conservation problems.

Phylogenetic diversity (PD) is a measure of biodiversity introduced in [6] that has subsequently been used in many biodiversity conservation applications. Biodiversity conservation planning should aim to protect as much PD as possible [3]. In this paper, we use our methods to consider the expected PD increase when the phylogeny (tree) produced by the Yule model is used to select species to protect.

Understanding the rate at which PD loss occurs as species become extinct is important for deciding when conservation interventions will be most effective—now or in the future. PD loss under a process of random extinction has been considered in several studies [12,18]. Here we combine these earlier results with the methods described in this paper to investigate the expected PD loss for trees produced by the Yule model.

2 Evolutionary models

Throughout this paper, we consider rooted binary trees (see Fig. 1). The root represents the ancestral species from which all other species are descendant. Internal nodes (with

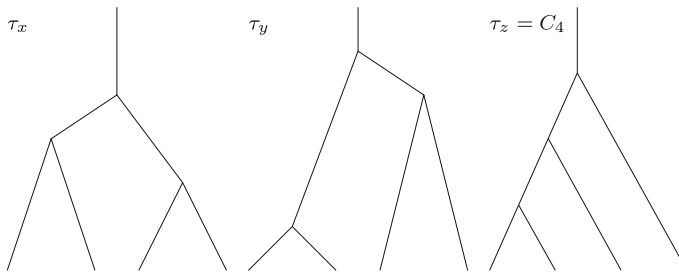


Fig. 1 Two trees have the same shape if they are indistinguishable when branch lengths and leaf labels are disregarded, thus the two trees on the left (τ_x and τ_y) have the same shape whereas the tree on the right (τ_z) has a different shape. The latter is an example of a caterpillar tree (C_n , $n = 4$) where each internal node has one species as a direct descendant

degree three) are ancestral species and the leaves (nodes of degree one) correspond to their modern descendants. The edges between any two nodes have associated lengths which may be interpreted as the time between speciation events or the genetic difference between the species corresponding to those nodes; this interpretation will depend on the data from which the tree was derived.

The Yule model makes the simple assertion that each species is equally likely to undergo a speciation event at any given point in time. Speciation can therefore be considered a Poisson process on any given lineage and the time between speciation events on a lineage is exponentially distributed with rate β , in various examples throughout this paper, we set $\beta = 1$. In a Bellman–Harris (BH) process an individual has a random lifespan, u , described by a probability distribution, $g(u)$, after which the individual is replaced by a random number of new individuals. Note that every species speciates according to the same distribution g . The Yule model is therefore analogous to a BH process where the “lifespan” of an individual is the time between speciation events on a given lineage (which is exponentially distributed) and each species is replaced by two new species (only binary trees are considered here).

This connection between BH processes and the Yule model suggests that it may be worthwhile to consider the larger class of BH evolutionary models. The BH models proposed here retain the constraint that each species is replaced by two new species, however the time between speciation events on a given lineage may be distributed according to an arbitrary probability density, $g(u)$. BH processes have been considered extensively in the mathematical branching process literature [14,20], particularly as applied to birth and death processes, however they have seen little application to phylogenetic trees [1].

The motivation for introducing BH models here is simply that our first method applies to the entire class of BH models. As such we do not discuss the implications of BH models further (or investigate different probability densities for $g(u)$). It should be noted that analytic solutions for all the applications presented here exist for the Yule model but no such guarantee exists for other BH models. Solutions for these models may need to be found numerically, which introduces additional complications due to the nested nature of some of the integrals.

The BH model we consider is restricted to binary trees. Approach 1 can readily be adapted to multifurcations, however the biological motivation for such processes seems

limited. An interesting extension would be to consider the more general Crump-Mode-Jagers models [4,5] which allow a speciation to occur without replacing the original species.

3 Approach 1: using nested integrals

The first approach describes the probability of a tree recursively in the form of nested integrals. These integrals will be nested to the same order as the depth of the tree. Our method applies to all BH models, however for some models the integrals may need to be solved numerically. Fortunately for the Yule model we can show that analytic solutions to these integrals exist.

3.1 Probability density of a BH-tree

Let τ denote the shape of the a tree, that is the tree without the associated edge lengths (see Fig. 1). The two trees descendant from the root of τ are denoted by τ_a and τ_b ; the number of species (leaves) in a tree is given by $|\tau|$. A tree, τ , may have edge lengths associated with it; the set of all edge lengths is denoted by λ_τ and the length of an individual edge, e , is denoted by λ_e . The root edge is denoted by r and its descendants are a and b , thus their edge lengths are λ_r, λ_a and λ_b respectively. For BH models the distance between the root node and any leaf node is the same for all leaves (the tree is ultrametric) and is denoted by t .

Using this notation the probability density for a tree, τ , with specified edge lengths under a BH model can be stated recursively as the product of the probability density of the root edge and the probability density for the trees descendant from the root:

$$h(\tau, \lambda_\tau) = \begin{cases} v_\lambda g(\lambda_r)h(\tau_a, \lambda_{\tau_a})h(\tau_b, \lambda_{\tau_b}) & |\tau| > 1 \\ 1 - \int_0^{\lambda_r} g(u)du & |\tau| = 1 \end{cases}$$

where v_λ equals two if $\lambda_a \neq \lambda_b$ and one otherwise. This factor of two arises as the side on which each descendant tree occurs is irrelevant. If the tree is of size one, the probability of obtaining it is simply one minus the probability of a speciation event occurring too soon (resulting in a tree with more than one species). Note that the edge lengths are continuous variables, hence $h(\tau, \lambda_\tau)$ is a probability density.

The shape of a tree depends only on the number of species descendant from each internal node, thus two trees have the same shape if they are indistinguishable after the edge lengths and leaf labels are disregarded (see Fig. 1 for an example). The probability of obtaining a particular tree shape after time t from a BH model can also be stated recursively (see also [23]) :

$$p(\tau|t) = \begin{cases} v_\tau \int_0^t g(\lambda_r)p(\tau_a|t - \lambda_r)p(\tau_b|t - \lambda_r)d\lambda_r & |\tau| > 1 \\ 1 - \int_0^t g(u)du & |\tau| = 1, \end{cases} \tag{1}$$

where v_τ equals two if τ_a and τ_b are different and one if they are equal. Note that there are a discrete number of tree shapes, hence $p(\tau|t)$ is a probability mass.

For trees with more than one species ($|\tau| > 1$) all possible lengths of the root edge are integrated over. The probability of obtaining the tree for a given λ_r is the product of the probability of the speciation event on the root lineage and the probabilities of obtaining the tree shapes descendant from the root in the remaining time. This is multiplied by two (v_τ) if the two descendant tree shapes differ as it does not matter which descendant tree shape occurs on which lineage descendant from the root. As before the probability of obtaining a tree with a single species is simply one minus the probability of a speciation event on the root lineage occurring “too soon”.

For any BH model, given a number of species the probability of obtaining a particular tree shape at time t is simply found by normalising over the set of all tree shapes of that size, Y_n :

$$p(\tau|t, n) = \frac{p(\tau|t)}{\sum_{\gamma \in Y_n} p(\gamma|t)}, \tag{2}$$

where it should be noted that τ is also in Y_n .

For Yule trees (where $g(u)$ is an exponential distribution) solutions to Eq. (1) can be found analytically. For example it is easy to show (using induction) that the probability of obtaining a caterpillar tree (see Fig. 1) with n leaves, $p(C_n|t)$ is:

$$p(C_n|t) = \begin{cases} 2^{n-2}e^{-t} (1 - e^{-t})^{n-1} / (n - 1)! & n > 1 \\ e^{-t} & n = 1 \end{cases} \tag{3}$$

Furthermore under the Yule model the tree shape probabilities ($p(\tau|n)$) are well known and independent of time [21]:

$$p(\tau|n) = \frac{2^{n-1-s(\tau)}}{\prod_{e, c_e > 2} (c_e - 1)},$$

where $s(\tau)$ is the number of internal edges for which the two descendant trees have the same tree shape, and c_e is the number of leaf descendants of edge e . The time dependent probability for any tree shape under the Yule model can therefore be obtained as follows:

$$\begin{aligned} p(\tau|t) &= \frac{p(\tau|t)}{p(C_n|t)} p(C_n|t) \\ &= \frac{p(\tau|n)}{p(C_n|n)} p(C_n|t) \\ &= 2^{n-1-s(\tau)} e^{-t} (1 - e^{-t})^{n-1} / \prod_{e, c_e > 2} (c_e - 1). \end{aligned}$$

It is interesting to note that for a given tree size, n , the tree shape probability, $p(\tau|t)$, has a maximum at $t = \log(n)$.

For other BH models the relative tree shape probabilities for a given number of species ($p(\tau|n, t)$) may not be independent of time. To remove the time dependency

contained in the relative tree shape probabilities (Eq. 2) these probabilities must be weighted by the probability distribution of the age of a tree given that it has n leaves, which we denote by $\phi(t|n)$. Assuming a uniform prior, $p(t)$, on the age of the tree between 0 and T we can obtain the following using Bayes theorem:

$$\begin{aligned}\phi(t|n) &= \frac{p(n|t)p(t)}{p(n)} \\ &= \frac{p(n|t)p(t)}{\int_0^T p(n|u)p(u)du} \\ &= \frac{p(n|t)/T}{\int_0^T p(n|u)/T du} \\ &= \frac{\sum_{\gamma \in Y_n} p(\gamma|t)}{\int_0^T \sum_{\gamma \in Y_n} p(\gamma|u)du}.\end{aligned}$$

If any age is possible we can take the limit of $\phi(t|n)$ as $T \rightarrow \infty$:

$$\lim_{T \rightarrow \infty} \phi(t|n) := \phi_\infty(t|n) = \frac{\sum_{\gamma \in Y_n} p(\gamma|t)}{\int_0^\infty \sum_{\gamma \in Y_n} p(\gamma|u)du}.$$

Obviously we require the denominator to be finite, in Appendix A it is shown that this holds as long as the mean of the speciation probability density, $g(t)$, is finite (otherwise $p(\tau|n) = \lim_{t \rightarrow \infty} p(\tau|t, n)$ if the limit exists). Using $\phi_\infty(t|n)$ the time dependence in Eq. (2) can be integrated out giving the time independent relative tree shape probability:

$$\begin{aligned}p(\tau|n) &= \int_0^\infty p(\tau|t, n)\phi_\infty(t|n)dt \\ &= \frac{\int_0^\infty p(\tau|t)dt}{\int_0^\infty \sum_{\gamma \in Y_n} p(\gamma|u)du}.\end{aligned}$$

This should be interpreted as the relative probability of observing a particular tree shape given that there are n species and speciation occurred according to the BH model (and the associated density, $g(t)$). One method for testing “real” trees against such a model is to compare the distribution of tree shapes for the real trees with those predicted by the model, this is the approach taken in [2].

3.2 Individual edge probability densities

Recently, [9] developed a method for calculating the expected length of any edge under a Yule model. Here an alternative approach is used to give the full probability distribution of that edge length, not just for the Yule model but for any BH model.

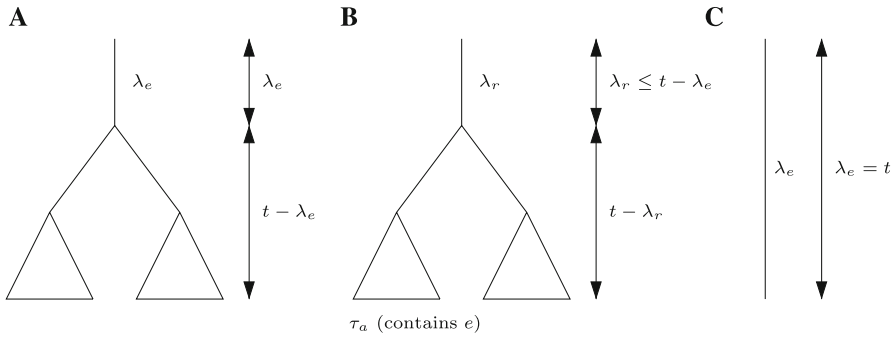


Fig. 2 The three possible scenarios for the location of the desired edge. In Scenarios *A* and *B* the tree contains more than one species and the edge is either the root (Scenario *A*) or in one of the subtrees (Scenario *B*). In Scenario *C* the tree contains only one species and the pendant edge belonging to this species is the desired edge. The probability of obtaining each of these trees is considered further in the main text

The probability density of the length of a particular edge, *e*, for a given tree shape, can be expressed recursively by integrating over the possible lengths of all other edges. To do so it is necessary to consider three possible positions of the desired edge as illustrated in Fig. 2. Denote the probability of obtaining a tree shape, τ , at time *t* with a specified edge having length λ_e as $\theta(\lambda_e, \tau, t)$, making use of three possible positions of the specified edge this can be stated recursively as:

$$\theta(\lambda_e, \tau|t) = \begin{cases} v_\tau g(\lambda_e) p(\tau_a|t - \lambda_e) p(\tau_b|t - \lambda_e) & \text{A: root is } e \text{ and } |\tau| > 1 \\ v_\tau \int_0^{t-\lambda_e} g(\lambda_r) \theta(\lambda_e, \tau_a|t - \lambda_r) p(\tau_b|t - \lambda_r) d\lambda_r & \text{B: } e \text{ in } \tau_a \text{ and } |\tau| > 1 \\ \delta(t - \lambda_e) \int_{\lambda_e}^\infty g(u) du & \text{C: } |\tau| = 1 \end{cases} \quad (4)$$

In scenario *A* the tree (τ) contains more than one species and the desired edge is the root. The probability of obtaining the tree in this scenario is therefore simply the product of the probability of a speciation event at time λ_e on the root ($g(\lambda_e)$) and the probability of each of the daughter trees having the appropriate shape $p(\tau_a|t - \lambda_e)$ and $p(\tau_b|t - \lambda_e)$. If the two daughter tree shapes differ then there are two possible ways of obtaining the final tree shape, this introduces the factor of two (v_τ).

In scenario *B* the desired edge is in one of the daughter trees which we refer to as τ_a without loss of generality. In this scenario the probability of obtaining the tree is obtained by integrating over all possible root edge lengths, λ_r . The root edge can range in length from 0 to $t - \lambda_e$ as this is the longest it can be and still “leave” sufficient time for edge *e* to obtain its desired length. For a given root edge length the probability of obtaining the tree is the product of the probability of the speciation event on the root lineage ($g(\lambda_r)$), the probability that τ_a will have the appropriate shape and edge length ($\theta(\lambda_e, \tau_a|t - \lambda_r)$) and the probability that τ_b will have the appropriate shape ($p(\tau_b|t - \lambda_r)$).

In scenario *C* τ contains only one species which must be the desired edge. For the desired edge to have length λ_e we must have $t = \lambda_e$ hence the dirac delta function

$\delta(t - \lambda_e)$. Furthermore the speciation event on the lineage must take place after time λ_e , this gives the integral from λ_e to ∞ of the density $g(u)$ (this is simply one minus the cumulative density function of $g(u)$ at λ_e).

The probability density $\theta(\lambda_e, \tau|t)$ is not conditioned on the tree shape or the number of species. Without normalisation it should therefore be interpreted as the probability of obtaining that particular tree shape and edge length out of all possible trees of age t . In many cases it is desirable to condition $\theta(\lambda_e, \tau|t)$ on the tree shape, this is achievable by simple normalisation:

$$\theta(\lambda_e|\tau, t) = \frac{\theta(\lambda_e, \tau|t)}{\int_0^t \theta(u, \tau|t)du},$$

yielding the probability density of edge length e given the tree shape τ and age of the tree.

3.3 Density of the longest pendant edge

Many other interesting properties of the trees created by a BH model can be considered. In this paper one final situation is considered where we wish to find the probability density of the longest pendant edge length for a given tree shape. The motivation for this came from a study where similarities between different indices for biodiversity conservation were being considered [19]. These indices were highly dependent on the lengths of pendant edges, consequently it became necessary to develop a good understanding of the distribution of pendant edge lengths.

A similar method to that employed in the preceding sections can be used to investigate this situation. This method is less obvious as (depending on the tree shape) there may be several edges which could be the longest pendant edge. Let τ_a and τ_b respectively be the smaller and larger daughter trees of τ . We define $\phi(\lambda_l, \tau|t)$ as the probability of obtaining a tree shape τ with a longest pendant edge with length λ_l given its age t , this can be stated recursively:

$$\begin{aligned} \phi(\lambda_l, \tau|t) &= v_\tau \\ &\times \begin{cases} \int_0^{t-\lambda_l} g(\lambda_r) [\Psi(\lambda_l, t - \lambda_r, \tau_a, \tau_b) + \Psi(\lambda_l, t - \lambda_r, \tau_b, \tau_a)] d\lambda_r & |\tau_a| > 1 \\ g(t - \lambda_l) p(\tau_b|\lambda_l) \int_{\lambda_l}^\infty g(u) du & |\tau_a| = 1 \end{cases} \quad (5) \end{aligned}$$

$$\Psi(\lambda_l, t, \tau_a, \tau_b) = \phi(\lambda_l, \tau_a|t) \int_0^{\lambda_l} \phi(m, \tau_b|t) dm$$

To gain some insight into Eq. (5) we give further consideration to the two cases illustrated in Fig. 3.

$|\tau_a| = 1$. A pendant edge is directly descendant from the root. This edge is guaranteed to be the longest pendant edge in τ . The probability, $\phi(\lambda_l, \tau|t)$, is therefore the product of the probability of the speciation event on the root lineage ($g(t - \lambda_l)$), the probability of obtaining the right tree shape for τ_b and the probability that no

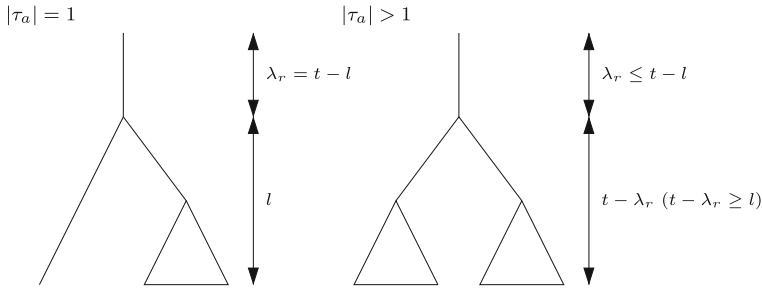


Fig. 3 The two possible scenarios considered for the location of the longest pendant edge. Either a pendant edge is adjacent to the root, in which case it must be the longest, or it may be located in either of the trees descendant from the root

speciation event will take place on the pendant edge in τ_a before time t . As before the possible factor of two (v_τ) represents the fact that it is irrelevant which set of events takes place on which of the lineages descendant from the root in τ .

$|\tau_a| > 1$. Both trees descendant from the root have more than one species. The longest pendant edge may occur in either of the trees descendant from the root, both possibilities must therefore be taken into account. $\Psi(\lambda_l, t, \tau_a, \tau_b)$ is the probability of obtaining the tree shapes τ_a and τ_b in time t with τ_a having the longest pendant edge of length λ_l . This is found by integrating over the possible longest pendant edge lengths of τ_b which can range from 0 to λ_l . Using $\Psi(\lambda_l, t, \tau_a, \tau_b)$ the probability of obtaining the tree τ with longest pendant edge λ_l is expressed as the integral over all possible root lengths, λ_r , of the product of the probability of obtaining a speciation event at time r and a longest pendant λ_l from one of the trees descendant from the root.

3.4 Analytic and numerical considerations

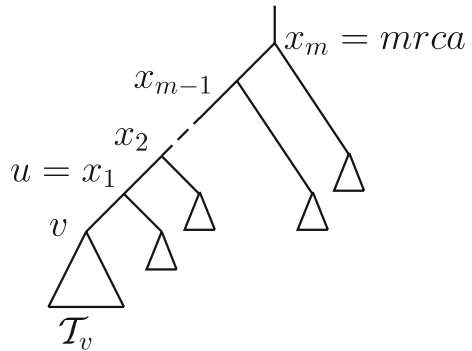
For the Yule model the equations for $p(\tau|t)$, $p(\tau|n)$, $\theta(\lambda_e, \tau|t)$ and $\phi(\lambda_l, \tau|t)$ possess analytic solutions. This is easily proven using induction on a case by case basis. The solutions for the Yule model are analytic because at each stage in the recursive approach a sum of exponential terms with linear exponents are produced, these expressions are easily combined and integrated to obtain an expression with the same form.

A symbolic algebra package was used to solve the recursive equations we have presented for the Yule model. Unfortunately not all BH models possess analytic solutions to the recursive equations. For instance if $g(t)$ is a distribution that does not possess an analytic cumulative density function then even the simplest case for $p(\tau, t)$ with $|\tau| = 1$ will not possess an analytic solution. To obtain solutions for the methods presented in this paper it may therefore be necessary to resort to numerical methods; this is further complicated by the nested nature of the integrals.

4 Approach 2: using rank functions for Yule trees

The second approach we present utilises rank functions for Yule models as introduced in [7,9]. No work to date has been done on rank functions of BH models in general,

Fig. 4 Labeling of the tree for calculating the probability for the rank of a vertex



hence at present this approach is strictly for Yule models. Using rank functions we derive a closed form equation for the density of an edge length in a tree which evolved under the Yule model, $\theta(\lambda_e | \tau, t)$.

For the concept of rank functions, we need to consider a tree shape with leaf labels, a so called phylogenetic tree \mathcal{T} . In the following, if a tree shape is given, we label the leaves in an arbitrary way to obtain a phylogenetic tree. Let \mathring{V} be the set of vertices in \mathcal{T} of degree > 1 . So the set \mathring{V} consists of all vertices in \mathcal{T} except of leaves and the root of the tree. A rank function [21] on a phylogenetic tree is a bijection from \mathring{V} to $\{1, 2, \dots, |\mathring{V}|\}$ with the property that the ranks are increasing on any path from the root to a leaf. We call a phylogenetic tree with a rank function a ranked phylogenetic tree.

The Yule model induces a uniform distribution on the ranked phylogenetic trees on n species [1]. In [7,9], polynomial time algorithms for calculating the probability of the rank of a vertex are provided for the uniform distribution on ranked phylogenetic trees. In the following we will explain the idea of the algorithms and adjust them to the application in this paper.

4.1 Calculating the rank distribution

Let r be a rank function on the phylogenetic tree \mathcal{T} . Define $p_u := (\mathbb{P}[r(u) = i])_{i=1, \dots, n-1}$. In [9], a formula for calculating p_u is given: Label the vertices on the path from the vertex u to the most recent common ancestor $mrca$ with $u = x_1, x_2, \dots, x_m = mrca$, see Fig. 4. Define λ_j as the number of leaves below x_j minus 1. With that notation, we get from [9] that

$$p_u = \frac{M_{m-1} M_{m-2} \dots M_1 e_1}{|M_{m-1} M_{m-2} \dots M_1 e_1|_1} \tag{6}$$

where $|\cdot|_1$ is the 1-norm, $e_1 = (1, 0, 0, \dots, 0)^T$ and the matrix M_j is defined as follows,

$$(M_k)_{i,j} = \begin{cases} 0 & \text{if } j < i - 1 - (\lambda_{k+1} - \lambda_k), \\ 0 & \text{if } j > i - 1, \\ \binom{\lambda_{k+1}-i}{\lambda_{k+1}-\lambda_k-i+j+1} \binom{i-2}{i-j-1} & \text{else.} \end{cases}$$

The algorithm RANKPROB in [9] calculates p_u according to Eq. (6).

For an edge $e = (u, v)$ in \mathcal{T} , we want to obtain the probability $p_{u,v}(i, j) := \mathbb{P}[r(u) = i, r(v) = j]$. First, let e be an interior edge. In [9], we calculate $p_{u,v}(i, j)$, $1 \leq i < j \leq n - 1$ by running RANKPROB on different subtrees of \mathcal{T} . In the following, we give an expression to calculate $p_{u,v}(i, j)$ directly from $p_u(i)$ which makes the calculations faster. Let \mathcal{T}_v be the smallest subtree induced by the leave descendants of v , see Fig. 4. The subtree \mathcal{T}_v has n_v leaves. Let $r(\mathcal{T})$ be the set of rank functions on \mathcal{T} .

The number of rank functions where $r(u) = i$ is $p_u(i) \cdot |r(\mathcal{T})|$. Assume we fix the first i interior nodes, with u being the i th node. There are $\binom{n-1-i}{n_v-1}$ possibilities to shuffle the interior vertices in \mathcal{T}_v with the remaining interior vertices. Only $\binom{n-1-j}{n_v-2}$ of those shuffles assign rank j to vertex v . Overall, we therefore get for the number of rank functions with $r(u) = i$ and $r(v) = j$:

$$p_u(i) \cdot |r(\mathcal{T})| \frac{\binom{n-1-j}{n_v-2}}{\binom{n-1-i}{n_v-1}}$$

For the probability $p_{u,v}(i, j)$, we have to divide the previous equation by the number of rank functions. Therefore

$$p_{u,v}(i, j) = p_u(i) \cdot \frac{\binom{n-1-j}{n_v-2}}{\binom{n-1-i}{n_v-1}}$$

This is equivalent to

$$p_{u,v}(i, j) = \begin{cases} p_u(i) \frac{n_v-1}{n-n_v-i+1} \prod_{k=1}^{n_v-2} \frac{n-j-k}{n-i-k}, & \text{if } n - j + 1 \geq n_v, 1 \leq i < j < n; \\ 0, & \text{else.} \end{cases} \tag{7}$$

We will extend the distribution $p_{u,v}$ for leaves. Since the leaves are after the $(n - 1)$ st speciation event, we can assume that all leaves have rank n . So for pendant edges, we have

$$p_{u,v}(i, n) = \begin{cases} p_u(i), & \text{if } v \text{ is a leaf;} \\ 0, & \text{else.} \end{cases} \tag{8}$$

Further, we will define $p_r, p_{r,v}$ for the root r . The root is always the very first vertex, the most recent common ancestor (*mrca*) is its descendant. Therefore, we define,

$$p_r(0) = 1, \quad p_{r,v}(0, 1) = \begin{cases} 1, & \text{if } v \text{ is the } mrca; \\ 0, & \text{else.} \end{cases} \tag{9}$$

4.2 Calculating densities under the Yule model

In the following we want to calculate the edge length density for an edge $e = (u, v)$ in a tree τ . Let Λ_e be the random variable “edge length of e with density function θ ”. Let Λ_u be the random variable “time of the speciation event u ” with density function f_{Λ_u} . In a tree with n species, let $\Lambda_{i,j}$ be the random variable “time between the i th and the j th speciation event”; with density function $f_{\Lambda_{i,j}}$. Let Λ_i be the random variable “time of the i th speciation event”; with density function f_{Λ_i} . With $i = 0$, we denote the root of the tree. Time is measured between today and the speciation event.

Under the Yule model, the waiting time between the $(k - 1)$ st speciation event and the k th speciation event, X_k , is exponentially distributed with rate k . We have

$$\Lambda_{i,j} = \sum_{k=i+1}^j X_k. \tag{10}$$

The present is between the $(n - 1)$ st speciation event and the n th speciation event. However, it has been shown in [13] that the time between the $(n - 1)$ st speciation event and the present has the exponential (rate n) distribution, this is X_n . So the present can be considered as the n th speciation event when not conditioning on the age of the tree, t . Later we will see that the same holds with conditioning on t .

We will derive the density and expectation of the random variables $\Lambda_i, \Lambda_{i,j}, \Lambda_u, \Lambda_e$. Recall that u is some interior vertex and e any edge. For the density, we have,

$$f_{\Lambda_u}(\lambda_u|\tau) = \sum_{i=0}^{n-1} f_{\Lambda_i}(\lambda_u|n)p_u(i);$$

$$\theta(\lambda_e|\tau) = \sum_{i=0}^{n-1} \sum_{j=i+1}^n f_{\Lambda_{i,j}}(\lambda_e|n)p_{u,v}(i, j).$$

Note that for interior edges, $p_{u,v}(i, n) = 0$, for pendant edges, $p_{u,v}(i, j) = 0$ for $j < n$ and for the root edge, $p_{u,v}(0, 1) = 1$.

For the expectation, we obtain from $\mathbb{E}[\Lambda_i|n]$,

$$\mathbb{E}[\Lambda_u|\tau] = \sum_{i=0}^{n-1} \mathbb{E}[\Lambda_i|n]p_u(i);$$

$$\mathbb{E}[\Lambda_{i,j}|n] = \mathbb{E}[\Lambda_i|n] - \mathbb{E}[\Lambda_j|n], \quad 0 \leq i < j \leq n;$$

$$\mathbb{E}[\Lambda_e|\tau] = \mathbb{E}[\Lambda_u|\tau] - \mathbb{E}[\Lambda_v|\tau] = \sum_{i=0}^n \mathbb{E}[\Lambda_i|n](p_u(i) - p_v(i));$$

with Λ_n being the present (which is, as explained above, equivalent to Λ_n being the time of the n th speciation event). Hence, $\Lambda_n = 0$ and therefore $\mathbb{E}[\Lambda_n|n] = 0$.

We can calculate the probabilities $p_u(i), p_{u,v}(i, j)$ as described in Sect. 4.1, so it is left to calculate $f_{\Lambda_{i,j}}(\lambda_{i,j}|n), f_{\Lambda_i}(\lambda_i|n), \mathbb{E}[\Lambda_i|n]$. The three values will be computed both conditioning and not conditioning on the age of the tree.

4.2.1 Unknown age of the tree

If we do not condition on the age of the tree, t , but assume a uniform prior for the age of the tree, the distribution of $\Lambda_i, \Lambda_{i,j}$ was calculated by [8] as follows,

$$f_{\Lambda_i}(\lambda_i|n) = (i + 1) \binom{n}{i + 1} e^{-n\lambda_i} (e^{\lambda_i} - 1)^{n-i-1}, \quad 0 \leq i \leq n - 1; \quad (11)$$

$$\begin{aligned} \mathbb{E}[\Lambda_i|n] &= \sum_{k=i+1}^n \frac{1}{k}, \quad 0 \leq i \leq n - 1; \\ f_{\Lambda_{i,j}}(\lambda_{i,j}) &= f_{\Lambda_{i,j}}(\lambda_{i,j}|n) \\ &= (i + 1) \binom{j}{i + 1} e^{-j\lambda_{i,j}} (e^{\lambda_{i,j}} - 1)^{j-i-1}, \quad 0 \leq i < j \leq n. \end{aligned} \quad (12)$$

Note that $\Lambda_{i,j}$ and $\Lambda_{k,l}, i < j \leq k < l$ are independent since the X_k are independent.

4.2.2 Known age of the tree

Next we state the density and expectation of Λ_i and the density of $\Lambda_{i,j}$ conditioned on the age of the tree. The proofs of these three theorems are provided in Appendix B.

Theorem 1 Each random variable $\Lambda_i (1 \leq i \leq n - 1)$ has the density function:

$$f_{\Lambda_i}(\lambda_i|n, t) = i \binom{n-1}{i} (1 - e^{-t})^{1-n} e^{-i\lambda_i} (1 - e^{-\lambda_i})^{n-i-1} (1 - e^{-(t-\lambda_i)})^{i-1}.$$

For $i = 0$, we have $f_{\Lambda_0}(\lambda_0|n, t) = \delta(\lambda_0 - t)$.

Theorem 2 The expectation of Λ_i (for $1 \leq i \leq n - 1$) is:

$$\mathbb{E}[\Lambda_i|n, t] = \sum_{k_1=0}^{n-i-1} \sum_{k_2=0}^{i-1} B_{k_1,k_2} (1 - e^{-t})^{1-n} (e^{-k_2t} - ((i + k_1 - k_2)t + 1)e^{-(i+k_1)t})$$

with $B_{k_1,k_2} := i \binom{n-1}{i} \binom{n-i-1}{k_1} \binom{i-1}{k_2} (-1)^{k_1+k_2} (i + k_1 - k_2)^{-2}$.

For $i = 0$, we have $\mathbb{E}[\Lambda_0|n, t] = t$.

Theorem 3 Each random variable $\Lambda_{i,j} (1 \leq i < j \leq n - 1)$ has the density function:

$$\begin{aligned} f_{\Lambda_{i,j}}(\lambda_{i,j}|n, t) &= \sum_{k_1=0}^{i-1} \sum_{k_2=0}^{n-j-1} B_{k_1,k_2} e^{(n-j)\lambda_{i,j}} \\ &\quad \times \frac{(e^{\lambda_{i,j}} - 1)^{j-i-1}}{(e^t - 1)^{n-1}} \left(e^{(n-i+k_1)(t-\lambda_{i,j})} - e^{k_2(t-\lambda_{i,j})} \right) \end{aligned}$$

with $B_{k_1,k_2} = i(i + 1) \binom{j}{i+1} \binom{n-1}{j} \binom{i-1}{k_1} \binom{n-j-1}{k_2} \frac{(-1)^{n+i-j-k_1-k_2}}{n-i+k_1-k_2}$.

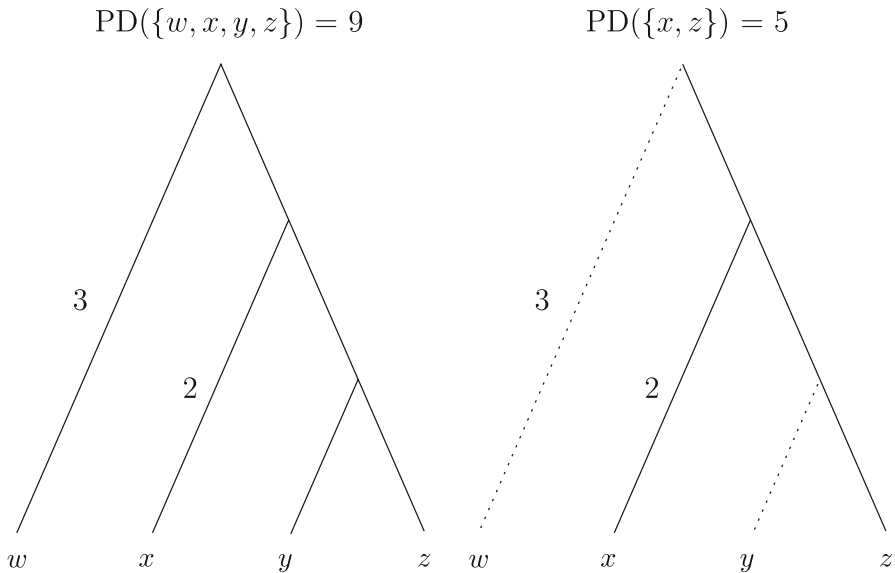


Fig. 5 Left: The PD of the set of all species in this tree is found by adding all edge lengths (which are length one unless otherwise indicated). The PD of a subset of species is the sum of all edges spanned by the subset and the root. Right: For a subset $Y = \{x, z\}$, $PD(Y)$ is found by adding the lengths of the solid lines of the tree on the right

For $\Lambda_{i,n}$, $i < n$, we have $f_{\Lambda_{i,n}}(\lambda_{i,n}|n, t) = f_{\Lambda_i}(\lambda_{i,n}|n, t)$, i.e. today can be interpreted as the n th speciation event.

For $\Lambda_{0,j}$ we have $f_{\Lambda_{0,j}}(\lambda_{0,j}|n, t) = f_{\Lambda_1}(t - \lambda_{0,j}|n, t)$.

5 Applications to biodiversity conservation

Phylogenetic diversity (PD ; [6]) is a commonly used index for measuring the biodiversity of a set of species contained in a phylogenetic tree. The PD of a subset of species, S , in a given tree is denoted by $PD(S)$ and is simply the sum of the edge lengths of the minimal spanning tree connecting the species in S and the root of the tree being considered (see Fig. 5).

Maximising PD has been widely considered as an appropriate aim for conservation management [3, 17]. In this section we address two important questions that naturally arise in this setting—how important are phylogenies for conservation and what is the characteristic rate of loss of PD ?

5.1 The value of phylogenetic trees for conservation management

An important question that arises is how much additional PD can be represented by picking the best set of k species when compared to a random set of k species? In the spirit of [24] we call this quantity the Expected Value of Perfect Choice ($EVPC$).

Mathematically for a give phylogenetic tree with edge lengths, \mathcal{T} , this quantity is given by:

$$\begin{aligned} EVPC(\mathcal{T}|k) &= PD(\text{best set with } k \text{ species}) - \mathbb{E}[PD(\text{any set with } k \text{ species})] \\ &= \max_S PD(S) - \mathbb{E}_S[PD(S)], \end{aligned}$$

where S ranges over subsets of k species. If the $EVPC$ is low then the phylogenetic tree may be unimportant for conservation, if the $EVPC$ is high then including the phylogeny may result in a much greater representation of biodiversity.

Here we explore the $EVPC$ values expected for a Yule model using the methods developed in this paper. This will give an indication of the importance of including phylogenies in conservation management if the evolutionary process can be approximated by a Yule model. In particular we will calculate the expected value of $EVPC(\mathcal{T}|k)$ for trees generated by the Yule model conditional on their size and age:

$$\mathbb{E}_{\mathcal{T}}[EVPC(\mathcal{T})] = \mathbb{E}_{\mathcal{T}}[\max_S PD(S)] - \mathbb{E}_{\mathcal{T}}[\mathbb{E}_S[PD(S)]],$$

where the expectation is over all trees with n species that are of age t .

The maximum PD of a set of size k depends only on the timing of the first k speciation events. Until the k th speciation event all edges will be spanned by an optimal set, after this point only k edges extant at a given time may be conserved. Hence we have:

$$\max_S PD(S) = \sum_{i=1}^k i(\Lambda_{i-1} - \Lambda_i) + k(\Lambda_k) = t + \sum_{i=1}^{k-1} \Lambda_i$$

From Theorem 1 we have the expectation of the speciation events, it is therefore straightforward to find the expected maximum PD .

The expected PD over all sets of species depends on the tree shape as well as the timing of the speciation events. For a given tree shape, τ , the probability with which an edge will be spanned by a random set depends only on the number of descendants it has. Introducing C_i as the number of descendants of an edge i we obtain:

$$\mathbb{E}_{\mathcal{T}}[\mathbb{E}_S[PD(S)]] = \sum_{\tau} p(\tau|n) \sum_e \mathbb{E}[\lambda_e|\tau] \left(1 - \frac{\binom{n-C_e}{k}}{\binom{n}{k}} \right).$$

This can be calculated using either of the methods presented in this paper.

Figure 6 depicts the expected maximum PD and the expected PD of a random set for various set sizes; the difference between these curves is the expected $EVPC(\mathcal{T}|k)$ for trees of age t with n species that are evolved according to the Yule model.

Many factors will influence whether the possible biodiversity gains suggested by this figure are significant enough to warrant the inclusion of phylogenies in conservation management. One of the most influential factors will be the extent of other restrictions that are imposed on the conservation manager. The $EVPC$ and related

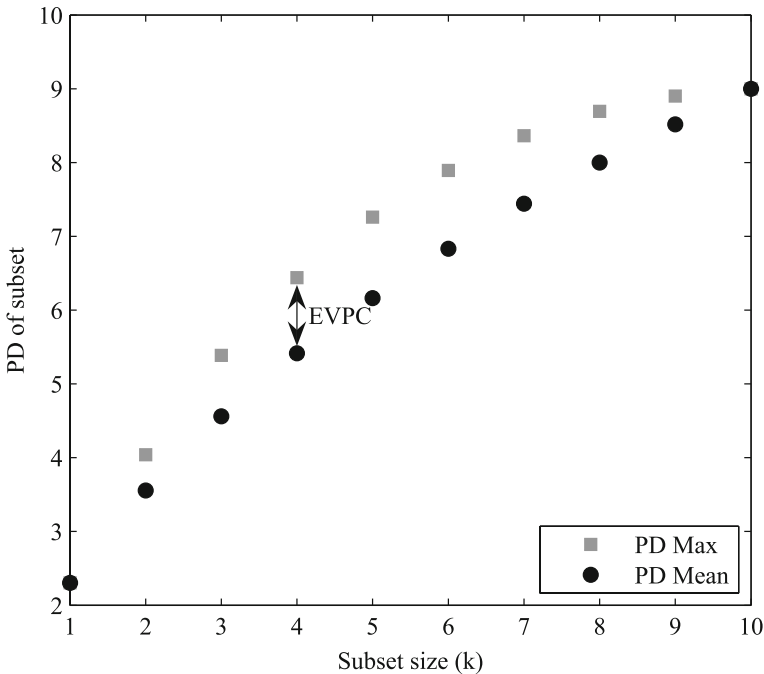


Fig. 6 Phylogenetic trees produced by the Yule model with ten species are considered in this figure. The expected *PD* represented by optimal subsets (squares) and random subsets (circles) of species from each tree is depicted. The difference between these points is the expected *EVPC*—the expected *PD* gain obtainable by using a phylogeny to select a set of species. The trees are conditioned on having the most likely age for their size ($t = \log(10)$)

indices will be considered extensively in forthcoming work [11] using the methods presented in this paper, of particular interest is the distribution of *EVPC*.

5.2 Loss of phylogenetic diversity over time

Understanding how the rate of *PD* loss due to extinction changes over time is crucial for predicting when conservation intervention will be most important—now or in the future. Nee and May [18] and Steel [22] considered the random extinction of species and showed that each new extinction causes a greater loss of *PD* than any previous extinction events. This suggests that conservation will gain importance as further extinctions occur; mathematically *PD* is a concave function of the number of extinctions that have occurred.

In [12] each species was considered to have the same probability of becoming extinct per unit time. *PD* loss was therefore considered as a function of time instead of the number of extinctions. It was shown that this curve is eventually convex – after some time, \hat{t} , the rate of expected *PD* loss reduces. This suggests that conservation intervention may be more important now than in the future. The time after which convexity is guaranteed (\hat{t}) was shown to be quite late, at a point in time

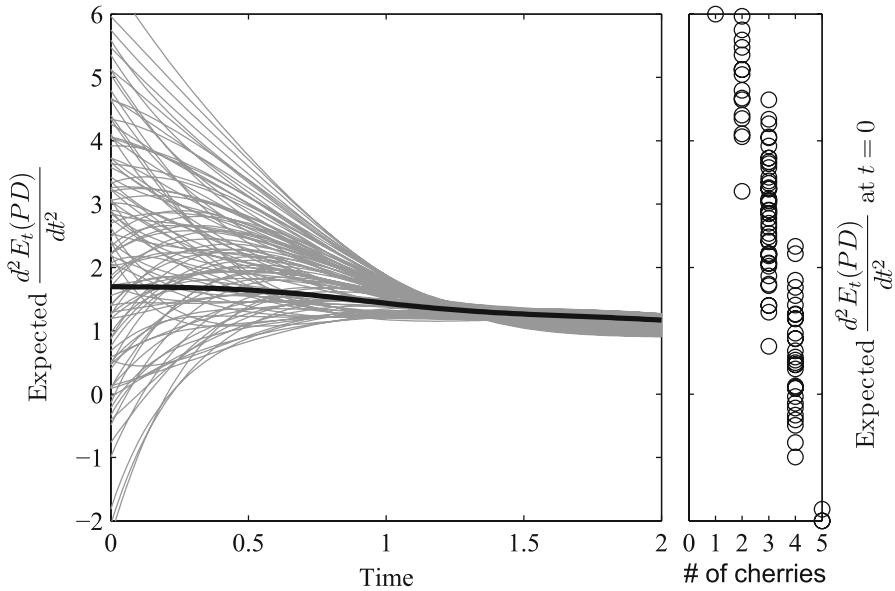


Fig. 7 The expected second derivative of the *PD* loss curve is shown here for trees with ten species generated by the Yule model. The gray lines in **a** are conditional on individual tree shapes and the solid line is the weighted mean over all tree shapes. Only 20% of tree shapes are expected to have a negative derivative at some time, the remainder are expected to be convex for all times. **b** shows the number of cherries for each tree shape as a function of the initial expected second derivative for that tree shape (as shown in **a**). As suggested in [12] this indicates that a large number of cherries is required for initial concavity. The trees are conditioned on having the most likely age for their size ($t = \log(10)$) and the time to extinction is exponentially distributed with rate 1

when between one and two species are expected to remain. The time to guaranteed convexity, \hat{t} , is an overestimate of the true time as this result applies to all possible phylogenetic tree shapes and edge lengths. For most trees it is expected that convexity will be reached much earlier. It is therefore of interest to explore the second derivative of the *PD* loss function for trees produced by the Yule model since these give some approximation of the trees expected to be found in nature. From [12] the second derivative of the expected *PD* loss curve (which must be positive for convexity) is:

$$\frac{d^2 \mathbb{E}_t(PD)}{dt^2} = r_e^2 e^{-r_e t} \left(\alpha_1 + \sum_{j=2}^n \alpha_j j (1 - j e^{-r_e t}) (1 - e^{-r_e t})^{j-2} \right) \quad (13)$$

where r_e is the risk of extinction and α_j is the sum of all edge lengths with exactly j descendants.

Using either of the methods developed in this paper, we can easily derive the expectation of Eq. (13) for trees produced by a Yule model. This is straightforward due to the linearity (in α_j) of Eq. (13), the result is shown in Fig. 7a. This indicates that for most tree shapes the expected *PD* loss will be convex for all times t . It should

be noted that for any tree shape we can choose edge lengths (which may be very improbably under the Yule model) such that the PD loss will be concave at some time.

Whenever two species/leaves are directly descendant from the same ancestor we refer to them as a cherry [15]. As discussed in [12] the number of cherries contributes to initial convexity and Fig. 7b shows that this also holds for Yule trees in our situation. Unless a ten species tree has at least four cherries it is expected to exhibit initial convexity. Approximately 20% of observed tree shapes are expected to exhibit initial concavity.

The extinction process considered here is very simple—all species are at the same risk of becoming extinct throughout time. In reality it is expected that these risks will not be independent, for example if a species becomes extinct other species dependent on it will have an increased risk of extinction or its competitors may have a decreased risk of extinction. This may also have a temporal effect on the extinction process, after many extinction events have occurred the extinction risk for the remaining species may increase due to their interdependency. The effect of interdependent extinction risks on the loss of PD is worthy of further exploration for which our methods may help yield greater insight.

6 Concluding comments

In this paper, we have studied the class of Bellman Harris (BH) evolutionary models, a class that includes the widely used Yule model. A method for calculating various probability distributions of tree shapes and edge lengths of trees produced under BH models has been presented. For Yule models analytic solutions exist for the proposed method, however for other BH models it may be necessary to resort to numerical methods.

A second method for calculating edge densities using rank functions has also been presented. This method only applies to the Yule model, however this limited scope makes this method conceptually easier to work with and implement. Reassuringly for the Yule model our methods (which are conceptually independent) give identical results.

We applied our methods to two problems related to biodiversity conservation. This highlights how evolutionary models can be used to characterise the expected outcome or behaviour for particular conservation scenarios. In the PD loss scenario we showed that the situations where the PD loss curves are concave are relatively uncommon and that even in these situations the curves quickly become convex.

Obtaining analytic solutions for properties of BH models in general or Yule models in particular can be complicated. However results often exist—particularly for Yule models—and it is worthwhile pursuing these. The approaches presented here are adaptable to a wide range of questions and scenarios. If it is necessary or desirable to simulate phylogenetic trees instead we caution that this should be done with some care using approaches such as those discussed in [13].

Acknowledgments We would like to thank Alan Hastings and two anonymous referees for their helpful comments on this manuscript.

Appendix A

Theorem 4 Consider the probability of obtaining a particular tree shape after time t , $p(\tau|t)$, under any BH evolutionary model with speciation probability density $g(t)$. If $g(t)$ has a finite mean, then the integral of $p(\tau|t)$ over all possible times is finite for any tree shape. More concisely, for any tree shape, τ , $\int_0^\infty p(\tau|t)dt$ is finite if $\int_0^\infty ug(u)du$ is finite.

Proof To prove Theorem 4, $p(\tau|t)$ is expressed in a new form and its integral is shown to have an upper bound of $\int_0^\infty ug(u)du$, hence if this upper bound is finite, the integral must also be finite.

Let $\phi(\tau, u)$ be the probability density that a tree shape, τ , is obtained during an evolutionary process and first occurs at time u . Let $\theta(\tau, u)$ be the probability that the tree shape exists for at least some time u . Making use of these two quantities the probability of obtaining a particular tree shape, τ , of age t can be expressed as:

$$p(\tau|t) = \int_0^t \phi(\tau, u)\theta(\tau, t - u)du. \tag{14}$$

This is simply the product of the probability of obtaining τ before time t and then retaining τ until time t (no further speciation events may occur).

The probability $\theta(\tau, u)$ is complicated to derive, however for this proof an upper bound will suffice. Note that $\theta(\tau, u)$ can be interpreted as the probability that no speciation events take place in a period of length u . It will therefore be bounded by the probability that no speciation event takes place on the last lineage that speciated. Fortunately we know the time of this speciation event (at the start of the period of length u) so we have:

$$\theta(\tau, u) \leq \int_u^\infty g(v)dv.$$

Substituting in Eq. (14) gives:

$$p(\tau|t) \leq \int_0^t \phi(\tau, u) \int_{t-u}^\infty g(v)dvdu. \tag{15}$$

To prove Theorem 4 we need to consider the integral of $p(\tau|t)$ over all possible times, t . Integrating both sides of Eq. (15) and changing the order of integration we obtain the required condition:

$$\int_0^\infty p(\tau|t)dt \leq \int_0^\infty \int_0^t \phi(\tau, u) \int_{t-u}^\infty g(v)dvdu dt$$

$$\begin{aligned}
&= \int_0^\infty \int_u^\infty \phi(\tau, u) \int_{t-u}^\infty g(v) dv dt du \\
&= \int_0^\infty \phi(\tau, u) \int_u^\infty \int_{t-u}^\infty g(v) dv dt du \\
&= \int_0^\infty \phi(\tau, u) du \int_0^\infty \int_i^\infty g(v) dv d\hat{t}, \quad \text{where } \hat{t} = t - u \\
&= \int_0^\infty \int_0^v g(v) d\hat{t} dv \\
&= \int_0^\infty v g(v) dv.
\end{aligned}$$

Appendix B

In the following we provide proofs for the results of Sect. 4.2.2.

Proof of Theorem 1 The distribution for $i = 0$ is obvious, since we condition on the age t of the tree. So now let $i > 0$. Note that $\Lambda_{0,i}$ and $\Lambda_{i,n}$ are independent since the X_k are independent (see Eq. 10). Denote the joint density function of $\Lambda_{0,i}$, $\Lambda_{i,n}$ as $f_{\Lambda_{0,i}, \Lambda_{i,n}}$, and apply Eq. (12) to obtain:

$$\begin{aligned}
f_{\Lambda_i}(\lambda_i | n, t) &= f_{\Lambda_{i,n}}(\lambda_i | n, \Lambda_{0,n} = t) \\
&= \frac{f_{\Lambda_{i,n}, \Lambda_{0,i}}(\lambda_i, t - \lambda_i | n)}{f_{\Lambda_{0,n}}(t | n)} \\
&= \frac{f_{\Lambda_{i,n}}(\lambda_i | n) f_{\Lambda_{0,i}}(t - \lambda_i | n)}{f_{\Lambda_{0,n}}(t | n)} \\
&= \frac{(i+1) \binom{n}{i+1} e^{-n\lambda_i} (e^{\lambda_i} - 1)^{n-i-1} i e^{-i(t-\lambda_i)} (e^{t-\lambda_i} - 1)^{i-1}}{n e^{-nt} (e^t - 1)^{n-1}} \\
&= i \binom{n-1}{i} (1 - e^{-t})^{1-n} e^{-i\lambda_i} (1 - e^{-\lambda_i})^{n-i-1} (1 - e^{-(t-\lambda_i)})^{i-1}. \quad (16)
\end{aligned}$$

Proof of Theorem 2 The expectation for $i = 0$ is obvious since we condition on the age t of the tree. Now let $i > 0$. We have

$$\begin{aligned} \mathbb{E}[\Lambda_i | n, t] &= \int_0^t \lambda_i f_{\Lambda_i}(\lambda_i | n, t) d\lambda_i \\ &= i \binom{n-1}{i} (1 - e^{-t})^{1-n} \sum_{k_1=0}^{n-i-1} \sum_{k_2=0}^{i-1} \binom{n-i-1}{k_1} \binom{i-1}{k_2} \\ &\quad \times (-1)^{k_1+k_2} e^{-k_2 t} \int_0^t \lambda_i e^{-(i+k_1-k_2)\lambda_i} d\lambda_i. \end{aligned}$$

Using integration by parts,

$$\int_a^b \lambda_i e^{-c\lambda_i} d\lambda_i = \left[-\frac{\lambda_i}{c} e^{-c\lambda_i} \right]_a^b + \frac{1}{c} \int_a^b e^{-c\lambda_i} d\lambda_i = -\frac{1}{c^2} [(\lambda_i c + 1) e^{-c\lambda_i}]_a^b.$$

Therefore, with $B_{k_1, k_2} := i \binom{n-1}{i} \binom{n-i-1}{k_1} \binom{i-1}{k_2} (-1)^{k_1+k_2} (i + k_1 - k_2)^{-2}$, we have:

$$\mathbb{E}[\Lambda_i | n, t] = \sum_{k_1=0}^{n-i-1} \sum_{k_2=0}^{i-1} B_{k_1, k_2} (1 - e^{-t})^{1-n} \left(e^{-k_2 t} - ((i + k_1 - k_2)t + 1) e^{-(i+k_1)t} \right)$$

which establishes the theorem.

Proof of Theorem 3 First, consider $j < n$. We can write $f_{\Lambda_{i,j}}(\lambda_{i,j} | n, t)$ as

$$\begin{aligned} f_{\Lambda_{i,j}}(\lambda_{i,j} | n, t) &= f_{\Lambda_{i,j}}(\lambda_{i,j} | n, \Lambda_{0,n} = t) \\ &= \frac{f_{\Lambda_{i,j}, \Lambda_{0,n}}(\lambda_{i,j}, t | n)}{f_{\Lambda_{0,n}}(t | n)} \\ &= \frac{f_{\Lambda_{i,j}, \Lambda_{0,i} + \Lambda_{j,n}}(\lambda_{i,j}, t - \lambda_{i,j} | n)}{f_{\Lambda_{0,n}}(t | n)} \\ &= \frac{f_{\Lambda_{i,j}}(\lambda_{i,j} | n) f_{\Lambda_{0,i} + \Lambda_{j,n}}(t - \lambda_{i,j} | n)}{f_{\Lambda_{0,n}}(t | n)}. \end{aligned} \tag{17}$$

The last equality holds since $\Lambda_{i,j}$ and $\Lambda_{0,i} + \Lambda_{j,n}$ are independent, because the X_k in Eq. (10) are independent. We will now obtain an expression for $f_{\Lambda_{0,i} + \Lambda_{j,n}}(t - \lambda_{i,j} | n)$. The random variables $\Lambda_{0,i}, \Lambda_{j,n}, j \geq i$ are independent, because the X_k in Eq. (10)

are independent. So $f_{\Lambda_{0,i}+\Lambda_{j,n}}(t - \lambda_{i,j}|n)$ is the convolution of $\Lambda_{0,i}$, $\Lambda_{j,n}$,

$$\begin{aligned}
 f_{\Lambda_{0,i}+\Lambda_{j,n}}(t - \lambda_{i,j}|n) &= \int_0^{t-\lambda_{i,j}} f_{\Lambda_{0,i}}(u|n) f_{\Lambda_{j,n}}(t - \lambda_{i,j} - u|n) du \\
 &= \int_0^{t-s} i e^{-iu} (e^u - 1)^{i-1} (j+1) \\
 &\quad \times \binom{n}{j+1} e^{-n(t-\lambda_{i,j}-u)} (e^{t-\lambda_{i,j}-u} - 1)^{n-j-1} du \\
 &= i(j+1) \binom{n}{j+1} e^{-n(t-\lambda_{i,j})} \sum_{k_1=0}^{i-1} \sum_{k_2=0}^{n-j-1} \binom{i-1}{k_1} \binom{n-j-1}{k_2} \\
 &\quad \times \frac{(-1)^{n+i-j-k_1-k_2}}{n-i+k_1-k_2} e^{k_2(t-\lambda_{i,j})} (e^{(n-i+k_1-k_2)(t-\lambda_{i,j})} - 1).
 \end{aligned} \tag{18}$$

Equation (17) combined with Eqs. (12) and (18) gives the formula described in Theorem 3.

For $j = n$, we can write

$$\begin{aligned}
 f_{\Lambda_{i,n}}(\lambda_{i,n}|n, t) &= f_{\Lambda_{i,n}}(\lambda_{i,n}|n, \Lambda_{0,n} = t) \\
 &= \frac{f_{\Lambda_{i,n}, \Lambda_{0,i}}(\lambda_{i,n}, t - \lambda_{i,n}|n)}{f_{\Lambda_{0,n}}(t|n)} \\
 &= \frac{f_{\Lambda_{i,n}}(\lambda_{i,n}|n) f_{\Lambda_{0,i}}(t - \lambda_{i,n}|n)}{f_{\Lambda_{0,n}}(t|n)} \\
 &= f_{\Lambda_i}(\lambda_{i,n}|n, t)
 \end{aligned}$$

where the last equality follows from Eq. (16).

The time between $i = 0$ and j is the age of the tree minus the time from today to the j th speciation event. This is $t - \Lambda_j$, which completes the proof.

References

1. Aldous, D.J.: Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. *Stat. Sci.* **16**(1), 23–34 (2001)
2. Blum, M., Francois, O.: Which random processes describe the Tree of Life? A large-scale study of phylogenetic tree imbalance. *Syst. Biol.* **55**(4), 685–691 (2006)
3. Crozier, R.H.: Preserving the information content of species: Genetic diversity, phylogeny, and conservation worth. *Annu. Rev. Ecol. Syst.* **28**, 243–268 (1997)
4. Crump, K.S., Mode, C.J.: A general age-dependent branching process I. *J. Math. Anal. Appl.* **24**, 494–508 (1968)

5. Crump, K.S., Mode, C.J.: A general age-dependent branching process II. *J. Math. Anal. Appl.* **25**, 8–17 (1969)
6. Faith, D.P.: Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10 (1992)
7. Gernhard, T.: Stochastic models of speciation events in phylogenetic trees. Diplom Thesis, Technical University of Munich, Germany (2006)
8. Gernhard, T.: The conditioned reconstructed process. *J. Theor. Biol.* (2008, in press)
9. Gernhard, T., Ford, D., Vos, R., Steel, M.: Estimating the relative order of speciation or coalescence events on a given phylogeny. *Evol. Bioinform. Online* **2**, 309–317 (2006)
10. Harding, E.F.: The probabilities of rooted tree-shapes generated by random bifurcation. *Adv. Appl. Probab.* **3**, 44–47 (1971)
11. Hartmann, K., Mooers, A.O.: When should phylogenies guide conservation management? (in preparation)
12. Hartmann, K., Steel, M.: Phylogenetic diversity: from combinatorics to ecology. In: Gascuel, O., Steel, M. (eds.) *Reconstructing Evolution—New Mathematical and Computational Advances*. Oxford University Press, New York (2007)
13. Hartmann, K., Gernhard, T., Wong, D.: Sampling trees from evolutionary models. *Systematic Biology* (submitted)
14. Jagers, P.: *Branching processes with biological applications*. Wiley, New York (1975)
15. McKenzie, A., Steel, M.: Distributions of cherries for two models of trees. *Math. Biosci.* **164**(1), 81–92 (2000)
16. Mooers, A.Ø., Heard, S.B.: Inferring evolutionary process from phylogenetic tree shape. *Q. Rev. Biol.* **72**(1), 31–54 (1997)
17. Mooers, A.O., Heard, S.B., Chrostowski, E.: Evolutionary heritage as a metric for conservation. In: Purvis, A., Brooks, T., Gittleman, J. (eds.) *Phylogeny and Conservation*, pp. 120–138. Cambridge University Press, Cambridge (2005)
18. Nee, S., May, R.M.: Extinction and the loss of evolutionary history. *Science* **278**(5338), 692–694 (1997)
19. Redding, D., Hartmann, K., Mimoto, A., Bokal, D., DeVos, M., Mooers, A.O.: The most “original species” often capture more phylogenetic diversity than expected. *J. Theor. Biology*, in Press **251**, 606–615 (2008)
20. Sankaranarayanan, G.: *Branching Processes and its Estimation Theory*. Wiley, New York (1989)
21. Semple, C., Steel, M.: *Phylogenetics*. Oxford University Press, New York (2003)
22. Steel, M.: Tools to construct and study big trees: A mathematical perspective. In: Hodkinson, T.R., Parnell, J.A. (eds.) *Reconstructing the Tree of Life: Taxonomy and Systematics of Species Rich Taxa*. CRC Press, USA (2006)
23. Steel, M., McKenzie, A.: Properties of phylogenetic trees generated by Yule-type speciation models. *Math. Biosci.* **170**, 91–112 (2001)
24. Walters, C.J.: *Adaptive Management of Renewable Resources*. Macmillan, New York (1986)
25. Yule, G.U.: A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis, F.R.S. *Philos. Trans. R. Soc. Lond. Ser. B* **213**, 21–87 (1924)