This is an electronic appendix to the paper by Pinelis 2003 Evolutionary models of phylogenetic trees. *Proc. R. Soc. Lond.* B **270**, 1425-1431. (DOI 10.1098/rspb.2003.2374.)

Electronic Appendix A

In this Appendix, one can find more details to the abovementioned paper, *Evolutionary models of phylogenetic trees*, referred here to as EMPT. Notation and assumptions introduced in EMPT are used here as well.

1. ON THE DEFINITION OF THE PDA MODEL

The usual definition of the PDA model (Mooers & Heard 1997; Rosen 1978) can be formalized as follows.

First, let us define the notion of a *distinguishable arrangement* (DA) $\mathcal{A}$ of a finite non-empty set A (of species), by induction:

- the only DA of any singleton set $A = \{a\}$ is $\mathcal{A} = a$;
- the DA's $\mathcal{A}$ of any finite non-singleton non-empty set A are all the non-ordered pairs of the form $\llbracket \mathcal{B}, \mathcal{C} \rrbracket$, where $\mathcal{B}$ and $\mathcal{C}$ are, respectively, DA's of some disjoint non-empty sets B and C whose union is A.

The DA's are also referred to as the *labelled shapes* (Harding 1971).

The (tree) *type* (also referred to as the "(unlabelled) shape") of a DA $\mathcal{A}$ of a set A is obtained by replacing all the entries of the elements of set A in the representation of $\mathcal{A}$ by $*$.

The size of a DA $\mathcal{A}$ of a set A can be naturally defined as the size of the tree type of DA $\mathcal{A}$ or, equivalently, as the number of elements of the set A.

For example, the DA's of the 3-element set $\{a, b, c\}$ are $\llbracket a, \llbracket b, c \rrbracket \rrbracket$, $\llbracket b, \llbracket a, c \rrbracket \rrbracket$, and $\llbracket c, \llbracket a, b \rrbracket \rrbracket$; all of them are of type $\llbracket *, \llbracket *, * \rrbracket \rrbracket$ and of size 3. There are exactly 15 DA's of any 4-element set.

The usual definition of the PDA model is that all DA's of a given set of species are equally probable. This condition determines a certain probability distribution on the set of all tree types of a given size w, so that the probability of a tree type $\mathcal{T}$ of size $|\mathcal{T}| = w$ is proportional to the number $a(\mathcal{T})$ of those DA's of set $\{1, \ldots, w\}$ that have tree type $\mathcal{T}$.

For any tree type $\mathcal{T}$, let $|\mathcal{T}|_{\mathsf{even}}$ and $|\mathcal{T}|_{\mathsf{uneven}}$ denote, respectively, the numbers of even and uneven splits. These numbers can be defined recursively by the requirements that $|*|_{\mathsf{even}} = |*|_{\mathsf{uneven}} = 0$, and

$$|\mathcal{T}|_{\mathsf{even}} = \mathsf{I}\{\mathcal{P} = \mathcal{D}\} + |\mathcal{P}|_{\mathsf{even}} + |\mathcal{D}|_{\mathsf{even}} \quad \text{and} \quad |\mathcal{T}|_{\mathsf{uneven}} = \mathsf{I}\{\mathcal{P} \neq \mathcal{D}\} + |\mathcal{P}|_{\mathsf{uneven}} + |\mathcal{D}|_{\mathsf{uneven}} \text{ for } \mathcal{T} = \llbracket \mathcal{P}, \mathcal{D} \rrbracket, \tag{1}$$

where the so-called indicator $\mathsf{I}\{\mathit{condition}\}$ equals 1 if *condition* holds and 0 otherwise. It is easy to see by induction that the total number of splits, $|\mathcal{T}|_{\mathsf{even}} + |\mathcal{T}|_{\mathsf{uneven}}$, coincides with $|\mathcal{T}| - 1$.

Now it is not hard to show that, for any tree type $\mathcal{T}$ of size $|\mathcal{T}| = w$, the number of the DA's of tree type $\mathcal{T}$ is given by the formula

$$a(\mathcal{T}) = \frac{w!}{2^{|\mathcal{T}|_{\mathsf{even}}}} = \frac{w!}{2^{w-1}} 2^{|\mathcal{T}|_{\mathsf{uneven}}}. \tag{2}$$

This can be either seen directly or checked by induction, using the relations $a(*) = 1$ and

$$a(\mathcal{T}) = \frac{1}{2^{\mathsf{I}\{\mathcal{P}=\mathcal{D}\}}} \binom{w}{u} a(\mathcal{P}) a(\mathcal{D})$$

if $\mathcal{T} = \llbracket \mathcal{P}, \mathcal{D} \rrbracket$, $|\mathcal{T}| = w$, and $|\mathcal{P}| = u$; cf. Harding 1971.

On the other hand, let $\mathit{tr}(\mathcal{T})$ denote the number of the trees of type $\mathcal{T}$. Then it is easy to see by induction that

$$\mathit{tr}(\mathcal{T}) = 2^{|\mathcal{T}|_{\mathsf{uneven}}}, \tag{3}$$

so that, in view of relation (2), the numbers $a(\mathcal{T})$ and $\mathit{tr}(\mathcal{T})$ of, respectively, the DA's and the trees of type $\mathcal{T}$ are proportional to each other given the size $|\mathcal{T}| = w$:

$$a(\mathcal{T}) = \frac{w!}{2^{w-1}} \mathit{tr}(\mathcal{T}). \tag{4}$$

One concludes that, given the size $|\mathcal{T}| = w$, the two assumptions (the assumption that all DA's of a set of size w are equally probable and the assumption that all trees of size w are equally probable) result in the same probability distribution on the set of all tree types of size w.

The proportionality coefficient $\frac{w!}{2^{w-1}}$ in Eq. (4) is no less than 1 for all $w \geqslant 1$, so that the number $a(\mathcal{T})$ of the DA's is no less than the number $\mathit{tr}(\mathcal{T})$ of the trees of type $\mathcal{T}$. However, in general the DA's of a set of a given size cannot be put into a correspondence with the trees of the same size so that the same number of DA's correspond to every tree. The obvious reason is that the proportionality coefficient $\frac{w!}{2^{w-1}}$ is not an integer in general.

2. ON THE CONDITIONALLY STATIONARY DISTRIBUTION

For any time moment $t \geqslant 0$ and any states i and j in U, let $p_{ij}(t)$ stand for the conditional probability – given one species in state i at time moment 0 – that the following will hold: (i) the species will not switch to a state *not in* U before time moment t; (ii) the species will not give birth to a separate species before time t; and (iii) the state of the species at time t will be j.

Let $P(t)$ stand for the nonnegative matrix whose entries are $p_{ij}(t)$ for i and j in U. Then it is not hard to verify the semigroup property: $P(t_1)P(t_2) = P(t_1 + t_2)$ for all nonnegative t_1 and t_2. It follows that $P(t) = e^{tA}$ for all $t \geqslant 0$, where the matrix A is defined by the relation $P(dt) = I + (dt)A$ for an infinitely small dt; here I is the identity matrix, whose rows and columns are labeled by i and j in U. By the definition of the MR model, the ij-entry of the matrix $P(dt)$ is $\mu_{ij}dt$ if $j \neq i$ and $1 - \lambda_i dt$ if $j = i$ (assuming again that $\mu_{ii} = 0$ for all i). Therefore, the ij-entry of A is μ_{ij} if $j \neq i$ and $-\lambda_i$ if $j = i$. Note also that all entries of the matrix $P(t)$ are strictly positive for all $t > 0$, because all unstable states were assumed to communicate with one another.

Now well-known results by Perron and Frobenius on nonnegative matrices (see e.g. Karlin 1992, Volume I, Section 8.2, Theorem 8.2.1) imply that there exist a real number ρ and a unique probability distribution $(\pi_i)_{i\in U}$ on the set U such that

$$\sum_{i\in U} \pi_i p_{ij}(t) = e^{-\rho t}\pi_j \tag{5}$$

for all j in U and all $t \geqslant 0$. In other words, $e^{-\rho t}$ is an eigenvalue of matrix $P(t)$ and $(\pi_i)_{i\in U}$ is the corresponding left eigenvector:

$$\boldsymbol{\pi} P(t) = e^{-\rho t}\boldsymbol{\pi} \tag{6}$$

where $\boldsymbol{\pi} := (\pi_i)_{i\in U}$ is the row of the π_i's. Moreover, the eigenvalue $e^{-\rho t}$ is greater than the modulus of any other eigenvalue of matrix $P(t)$. Furthermore, one has

$$\pi_i > 0 \quad \text{for all } i \text{ in } U.$$

(Note that the matrix $P(t)$ is aperiodic for all $t \geqslant 0$, because $P(t) \to I$ as $t \downarrow 0$.)

It will be seen presently that $\rho \geqslant 0$ and the corresponding distribution $\boldsymbol{\pi} = (\pi_i)_{i\in U}$ is conditionally stationary.

Indeed, assume that condition (MR2) of Section 5 of EMPT holds, so that the state distribution of the species at time moment 0 coincides with $\boldsymbol{\pi}$.

Summing now equations (5) in j, one sees that $e^{-\rho t}$ is the probability that before time moment t the species will not switch to a quasi-stable state and will not become extinct and will not give birth to a separate species. Therefore, $\rho \geqslant 0$, and what equations (5) express is the following: given that before time t the species does not switch to a quasi-stable state and does not become extinct and does not give birth to a separate species – the conditional distribution of the state of the species at time t will still be $\boldsymbol{\pi}$. Thus, $\boldsymbol{\pi} = (\pi_i)_{i\in U}$ is indeed a conditionally stationary distribution on the set U of all unstable states.

Therefore, this conditionally stationary distribution, $\boldsymbol{\pi} = (\pi_i)_{i\in U}$, will be the conditional probability distribution of the parent species of the random tree $\mathfrak{T}$ at the time of giving birth given the event of ever giving birth.

On the other hand, if condition (MR3) of EMPT holds and, for all i in U, one has $\sigma_{iQ} = \sigma_{iD} = 0$, then the conditional probability that the daughter species will be in state $j \in U$ upon its birth given that the birth occurs will be

$$\sum_{i\in U} \pi_i \frac{\sigma_{ij}}{\sigma_{iS}} = \sum_{i\in U} \pi_i \frac{\sigma_{ij}}{\sigma_{iU}} = \sum_{i\in U} \pi_i \pi_j = \pi_j;$$

thus, just as for the parent species, the conditional probability distribution of the daughter species upon its birth (given the event of birth) will be the same as the conditionally stationary distribution $(\pi_i)_{i\in U}$.

Moreover, even without conditions (MR2) and (MR3) of EMPT, one has

$$p_{ij}(t)/e^{-\rho t} \to \nu_i \pi_j \quad \text{as } t \to \infty \text{ for all } i \text{ and } j \text{ in } U,$$

where the (necessarily positive) numbers ν_i form a column matrix $\boldsymbol{\nu}$ that is the right eigenvector of matrix $P(t)$ corresponding to eigenvalue $e^{-\rho t}$ and normalized so that $\boldsymbol{\pi\nu} = \sum_{i\in U} \pi_i \nu_i = 1$ (while $P(t)\boldsymbol{\nu} = e^{-\rho t}\boldsymbol{\nu}$) – see e.g. Karlin 1992, Volume I, Section 8.2, especially Theorem 8.2.2 there. Hence,

$$\frac{p_{ij}(t)}{\sum_{k\in U} p_{ik}(t)} \to \pi_j \tag{7}$$

as $t \to \infty$ for all i and j in U; that is, the conditional distribution of the state of the species at time t – given that before time moment t the species has not switched to a quasi-stable state and has not become extinct and has not not given birth to a separate species – converges to the unique conditionally stationary distribution $\boldsymbol{\pi} = (\pi_j)_{j\in U}$ as $t \to \infty$, for any initial (at time $t = 0$) state i in U.

In particular, this shows that assumptions (MR2) and (MR3) of EMPT – that the initial state distribution of both the root and daughter species is conditionally stationary – are not unreasonable. Indeed, if the rates $\mu_{iQ} + \mu_{iD} + \sigma_{iS}$ of essential change are small enough (compared to μ_{iU}) for all i, then the time to an essential change (that is the time to the first exit from the set U) will be large with a probability close to 1. Hence, in view of (7), the state distribution of the parent species at the time of giving birth will be close to the conditionally stationary distribution $(\pi_i)_{i\in U}$, for *any* given state distribution of the species at time $t = 0$. Thus, the probability distribution of the random tree $\mathfrak{T}$ will be close to that under conditions (MR2) and (MR3) of EMPT.

Differentiating both sides of (6) at $t = 0$, one sees that $\boldsymbol{\pi}$ is a left eigenvector of the matrix A corresponding to its eigenvalue $(-\rho)$: $\boldsymbol{\pi}A = -\rho\boldsymbol{\pi}$; that is, ρ and $\boldsymbol{\pi} = (\pi_i)_{i\in U}$ satisfy the following system of equations:

$$\sum_{i\in U} \pi_i \mu_{ij} = (\lambda_j - \rho)\pi_j \quad \text{for all } j \in U. \tag{8}$$

Summing these equations in j and recalling that $\mu_{iS} + \sigma_{iS} = \lambda_i$, one has

$$\rho = \sum_{i\in U} \pi_i(\mu_{iQ} + \mu_{iD} + \sigma_{iS}). \tag{9}$$

For any infinitesimal interval $(t, t{+}dt)$, the quantity $(\mu_{iQ} + \mu_{iD} + \sigma_{iS})\,dt$ can be interpreted as the probability that during this time interval a species in the state i undergoes an essential phylogenetic change, that is a change other than a transition from the unstable state i to another unstable state. Therefore, the term $\mu_{iQ} + \mu_{iD} + \sigma_{iS}$ can be referred to as the rate of essential change (from state i).

Thus, ρ can be referred to as the average rate of essential change. More specifically, the probability that in a time interval $(t, t + dt)$ the unstable species will switch to a quasi-stable state or become extinct or give birth to a separate species is $\rho\,dt$.

Because $\sigma_{iS} > 0$ for some i in U and $\pi_i > 0$ for all i in U, it follows that

$$\rho > 0.$$

Hence, $e^{-\rho t} \to 0$ as $t \to \infty$. Thus, the probability that the root species (i.e. the one at the root of the random tree $\mathfrak{T}$) will ever switch to a quasi-stable state or will ever become extinct or will ever give birth to a separate species is 1. Introduce now the corresponding probabilities:

- q, that the root species will ever switch to a *quasi-stable* state;
- d, that the root species will ever become *dead* (i.e. extinct);
- b, that the root species will ever give *birth* to a separate species.

For any $t > 0$ these probabilities coincide, respectively, with the conditional probabilities that in an infinitesimal time interval $(t, t + dt)$ the root species

- will switch to a *quasi-stable* state;
- will become *dead* (i.e. extinct);
- will give *birth* to a separate species

– given that before time moment t the root species does not switch to a quasi-stable state and does not become extinct and does not give birth to a separate species but during the time interval $(t, t + dt)$ undergoes one of the three kinds of essential change.

It follows that

$$q = \frac{\sum_{i\in U} \pi_i \mu_{iQ}}{\rho}; \quad d = \frac{\sum_{i\in U} \pi_i \mu_{iD}}{\rho}; \quad b = \frac{\sum_{i\in U} \pi_i \sigma_{iS}}{\rho} = \frac{\sum_{i\in U} \pi_i (\sigma_{iU} + \sigma_{iQ})}{\rho}. \tag{10}$$

Thus, in view of Eq. (9), one can see once again that

$$q + d + b = 1.$$

At that, one has

$$b > 0, \tag{11}$$

since $\sigma_{iS} > 0$ for some i in U. Introduce also

$$\gamma := \frac{\sum_{i\in U} \pi_i \sigma_{iQ}}{\sum_{i\in U} \pi_i \sigma_{iS}} = \frac{\sum_{i\in U} \pi_i \sigma_{iQ}}{\sum_{i\in U} \pi_i (\sigma_{iU} + \sigma_{iQ})}. \tag{12}$$

Then $0 \leqslant \gamma \leqslant 1$ and, moreover, γ is the conditional probability that the state of the daughter species upon its birth will be quasi-stable given the event of birth. Respectively,

$$1 - \gamma = \frac{\sum_{i\in U} \pi_i \sigma_{iU}}{\sum_{i\in U} \pi_i \sigma_{iS}} = \frac{\sum_{i\in U} \pi_i \sigma_{iU}}{\sum_{i\in U} \pi_i (\sigma_{iU} + \sigma_{iQ})}$$

is the conditional probability that the state of the daughter species upon its birth will be unstable given the event of birth.

One can now see that – under conditions (MR2) and (MR3) of EMPT – the probability distribution of the random tree $\mathfrak{T}$ in *any* MR model will be the same as that in the following simple case of a MR model.

Namely, in such a mathematically reduced MR model one has just one state (say, labeled by 0) for the "extinct species", one state (labeled by 1) for the "quasi-stable" (or, in this case, one can simply say "stable") "species", and one state (say, labeled by 2) for the "unstable species". In other words, here $\tilde{S} = \{0, 1, 2\}$, $\tilde{D} = \{0\}$, $\tilde{Q} = \{1\}$, and $\tilde{U} = \{2\}$; the tilde is used at this point to distinguish the attributes of the reduced MR model from the original ones. (Note that in the reduced MR model the "species" or their possible "states" (0, 1, and 2) do not have to be in a one-to-one correspondence with any real species and their states. Rather, the "species" and "states" in the reduced MR model are mathematical abstractions, which are convenient for the purposes of verification of claims made in Section 5 of EMPT.) At that one has the following expressions of the parameters of the reduced MR model in terms of the parameters of the original one:

- $\tilde{\mu}_{0\tilde{S}} = \tilde{\sigma}_{0\tilde{S}} = 0;$
- $\tilde{\mu}_{1\tilde{S}} = \tilde{\sigma}_{1\tilde{S}} = 0;$

- $\tilde\mu_{20} = \rho d$, $\tilde\mu_{21} = \rho q$, $\tilde\mu_{22} = 0$, $\tilde\sigma_{20} = 0$, $\tilde\sigma_{21} = \rho b\gamma$, $\tilde\sigma_{22} = \rho b(1-\gamma)$.

In words the reduced MR model can be described as follows. The evolutionary tree begins with one root "species" in the "unstable" state, 2. The probability that a "species" that was in state 2 at time $t = 0$ undergoes an essential change of state in any infinitesimal time interval $(t, t + dt)$ is $\rho\, dt$; this probability is the sum of the following probabilities:

- $\rho\, d\, dt$, the probability that within this time interval the "species" will become "extinct" ;
- $\rho\, q\, dt$, the probability that within this time interval the "species" will become "stable";
- $\rho\, b\,(1-\gamma)\, dt$, the probability that within this time interval the "species" will give birth to an "unstable species";
- $\rho\, b\, \gamma\, dt$, the probability that within this time interval the "species" will give birth to a "stable species".

Neither "stable" nor "extinct" "species" ever change their state. However, the "stable" "species" are always included into the tree, together with their lineages; in contrast, the "extinct" "species" (together with their lineages) may or may not be included. All current "species" develop independently of one another and of the prehistory.

3. ON MR INTERPRETATIONS OF THE ERM AND (S)PDA MODELS

In this section, conditions (MR2) and (MR3) of EMPT are assumed, which makes it possible to reduce the MR model as described just above. Therefore, here all the reasoning is shown for the reduced MR model. Yet, for simplicity, let us use here the terms "species", "state(s)", "stable", "unstable", and "extinct" without quotation marks, even though they pertain to the reduced mathematical model and are mere mathematical abstractions.

Condition (MR1) of EMPT also is assumed in this section. However, all the actual reasoning here is done only for the case when the random tree $\mathfrak{T}$ is D-incomplete in the original MR model, which corresponds to the $\{0\}$-incompleteness in the reduced MR model. The case of a complete random tree $\mathfrak{T}$ is only simpler; moreover, the reduced complete MR model can be treated mathematically as a special case of the $\{0\}$-incomplete reduced model, with the "death" probability $d = 0$.

3.1. On MR interpretations of the ERM model

In this subsection, in accordance with Section 5 of EMPT, it is assumed the MR-ERM conditions in Table 1 of EMPT are satisfied, as well as conditions (MR1), (MR2), and (MR3).

In view of Eqs. (10) and (12), conditions $\mu_{iQ} = \sigma_{iQ} = 0$ mean that $q = 0$ and $\gamma = 0$. Therefore, the $\{0\}$-incomplete reduced MR model here is mathematically equivalent to the BDP process with the birth and death rates ρb and $\rho d = \rho(1-b)$, respectively, and with only the live particles and their lineages included into the random tree.

Hence, according to Slowinski & Guyer 1989 and Rogers 1994, the resulting evolutionary model is equivalent to the ERM model in the following sense: the probability distribution of the random tree grown to a given size is the same as that in the ERM model.

However, in real evolutionary processes, the size of the tree is not pre-determined. It seems more natural to consider the process stopped at a finite non-random time $T > 0$, rather than at the random time T_w when the tree first reaches size w.

Let us denote by $\mathfrak{T}_T$ the random tree observed over the time interval $[0, T]$. In this subsection, the probability distribution of $\mathfrak{T}_T$ will be obtained for the MR-ERM model described here. That is, for every given non-random finite tree $\mathbf{t}$, an explicit expression will be found for the probability, denoted here by $p_T(\mathbf{t})$, that the random tree $\mathfrak{T}_T$ happens to coincide with $\mathbf{t}$.

In particular, it will follow that the conditional probability distribution of the tree $\mathfrak{T}_T$ given that its size $|\mathfrak{T}_T|$ equals w coincides with the conditional probability distribution of the tree $\mathfrak{T}_{T_w}$ given $T_w < \infty$, for every natural number w.

Let $p_T(w)$ denote the probability that the random tree $\mathfrak{T}_T$ happens to have size w, so that

$$p_T(w) = \sum_{\mathbf{t}:\ |\mathbf{t}|=w} p_T(\mathbf{t}).$$

Expressions for $p_T(w)$ were given by Kendall 1948 (Eq. (16) there), which he ascribed to Palm. In our terms,

$$p_T(0) = \frac{(1-b)\cdot\bigl(e^{\rho\cdot(2b-1)T}-1\bigr)}{b\,e^{\rho\cdot(2b-1)T}-1+b} \quad\text{and}\quad p_T(w) = (1-p_T(0))(1-\eta_T)\eta_T^{w-1} \text{ for } w \geqslant 1,$$

where

$$\eta_T = \frac{b}{1-b}\,p_T(0).$$

In fact, the above expression for $p_T(0)$ is good only for $b \neq 1/2$; for $b = 1/2$, one has, as if by continuity, $p_T(0) = \frac{\rho T/2}{1+\rho T/2}$. Note the important fact the the conditional distribution of the size $|\boldsymbol{\mathfrak{T}}_T|$ given that it is nonzero is geometric; in other words, for any given natural $w \geqslant 2$, the product $p_T(u)p_T(w-u)$ is the same for all $u = 1, \dots, w-1$.

It will be shown presently by induction on the size $|\boldsymbol{\mathfrak{t}}|$ that for any non-random finite tree $\boldsymbol{\mathfrak{t}}$

$$p_T(\boldsymbol{\mathfrak{t}}) = \kappa(\mathcal{T}_{\boldsymbol{\mathfrak{t}}})\,p_T(|\boldsymbol{\mathfrak{t}}|), \tag{13}$$

where $\mathcal{T}_{\boldsymbol{\mathfrak{t}}}$ stands for the type of tree $\boldsymbol{\mathfrak{t}}$ and the numbers $\kappa(\mathcal{T})$ are defined recursively: $\kappa(\emptyset) = 1$, $\kappa(*) = 1$, and

$$\kappa(\mathcal{T}) = \frac{\kappa(\mathcal{P})\kappa(\mathcal{D})}{|\mathcal{T}|-1} \quad \text{if} \quad \mathcal{T} = [\![\mathcal{P},\mathcal{D}]\!]. \tag{14}$$

Indeed, Eq. (13) is obvious if the size $|\boldsymbol{\mathfrak{t}}|$ is 0 or 1. Let now $|\boldsymbol{\mathfrak{t}}| = w \geqslant 2$. Then

$$\begin{aligned} p_T(w) &= \int_0^T e^{-\rho t}\rho\,dt\cdot b\Bigl(2p_{T-t}(0)p_{T-t}(w) + \sum_{u=1}^{w-1} p_{T-t}(u)p_{T-t}(w-u)\Bigr) \\ &= \int_0^T e^{-\rho t}\rho\,dt\cdot b\Bigl(2p_{T-t}(0)p_{T-t}(w) + (w-1)p_{T-t}(u_0)p_{T-t}(w-u_0)\Bigr) \end{aligned} \tag{15}$$

for any $u_0 = 1, \dots, w-1$, in view of the mentioned geometric property.

Similarly and also using the induction, one has for any tree of the form $\boldsymbol{\mathfrak{t}} = (\boldsymbol{\mathfrak{p}}, \boldsymbol{\mathfrak{d}})$ with $|\boldsymbol{\mathfrak{t}}| = w$

$$\begin{aligned} p_T(\boldsymbol{\mathfrak{t}}) &= \int_0^T e^{-\rho t}\rho\,dt\cdot b\Bigl(2p_{T-t}(\emptyset)p_{T-t}(\boldsymbol{\mathfrak{t}}) + p_{T-t}(\boldsymbol{\mathfrak{p}})p_{T-t}(\boldsymbol{\mathfrak{d}})\Bigr) \\ &= \int_0^T e^{-\rho t}\rho\,dt\cdot b\Bigl(2p_{T-t}(0)\kappa(\mathcal{T}_{\boldsymbol{\mathfrak{t}}})p_{T-t}(|\boldsymbol{\mathfrak{t}}|) + \kappa(\mathcal{T}_{\boldsymbol{\mathfrak{p}}})\kappa(\mathcal{T}_{\boldsymbol{\mathfrak{d}}})p_{T-t}(|\boldsymbol{\mathfrak{p}}|)p_{T-t}(|\boldsymbol{\mathfrak{d}}|)\Bigr) \\ &= \kappa(\mathcal{T}_{\boldsymbol{\mathfrak{t}}})\int_0^T e^{-\rho t}\rho\,dt\cdot b\Bigl(2p_{T-t}(0)p_{T-t}(|\boldsymbol{\mathfrak{t}}|) + (w-1)p_{T-t}(|\boldsymbol{\mathfrak{p}}|)p_s(|\boldsymbol{\mathfrak{d}}|)\Bigr) \\ &= \kappa(\mathcal{T}_{\boldsymbol{\mathfrak{t}}})\,p_T(|\boldsymbol{\mathfrak{t}}|), \end{aligned}$$

in view of Eq. (15). This shows that formula (13) is true.

Now one can find the probability distribution of the type of the random tree $\boldsymbol{\mathfrak{T}}_T$. For any given finite tree type $\mathcal{T}$, let $p_T(\mathcal{T})$ be the probability that the type of the random tree $\boldsymbol{\mathfrak{T}}_T$ happens to coincide with $\mathcal{T}$. Then Eqs. (3) and (13) yield

$$p_T(\mathcal{T}) = 2^{|\mathcal{T}|_{\mathsf{uneven}}}\kappa(\mathcal{T})p_T(|\mathcal{T}|)$$

for any tree type $\mathcal{T}$.

Thus, the conditional probability of the tree type $\mathcal{T} = \mathcal{T}_{\boldsymbol{\mathfrak{T}}_T}$ given its size $|\mathcal{T}|$ is

$$p_T\left(\mathcal{T}\middle||\mathcal{T}|\right) = \frac{p_T(\mathcal{T})}{p_T(|\mathcal{T}|)} = 2^{|\mathcal{T}|_{\mathsf{uneven}}}\kappa(\mathcal{T}).$$

In view of Eqs. (1) and (14), it follows now that for any tree type of the form $\mathcal{T} = [\![\mathcal{P},\mathcal{D}]\!]$

$$p_T\left(\mathcal{T}\middle||\mathcal{T}|\right) = 2^{\mathbf{1}\{\mathcal{P}\neq\mathcal{D}\}}\frac{p_T\left(\mathcal{P}\middle||\mathcal{P}|\right)p_T\left(\mathcal{D}\middle||\mathcal{D}|\right)}{|\mathcal{T}|-1},$$

which coincides with the recursive relation that determines the conditional probability distribution of the type of the random tree $\mathfrak{T}_{T_w}$ given $T_w < \infty$, for all sizes w in the ERM model (see e.g. Harding 1971, Theorem 1 on page 64, and – with extinction taken into account – Rogers 1994).

Thus, it has been demonstrated that the conditional probability distribution of the tree $\mathfrak{T}_T$ given that its size $|\mathfrak{T}_T|$ equals w coincides with the conditional probability distribution of the tree $\mathfrak{T}_{T_w}$ given $T_w < \infty$, for every natural number w.

3.2. On MR interpretations of the PDA model

In this subsection, in accordance with Section 5 of EMPT, it is assumed the MR-PDA conditions in Table 1 of EMPT are satisfied, as well as conditions (MR1), (MR2), and (MR3). In particular, the condition that $\sigma_{iQ} = 0$ for all i in U implies $\gamma = 0$. Recall also the natural assumption that $\sigma_{iD} = 0$ for all i in U stated in Section 5 of EMPT. Also, here one has

$$q > 0,$$

in view of Eq. (10) above and the MR-PDA conditions in Table 1 of EMPT.

Let here

$$p(\mathfrak{t}) := p_\infty(\mathfrak{t}).$$

In Section 3 of EMPT, Eq. (3.1) was deduced from conditions (PDA1) and (PDA2). However, it is more convenient here not to verify conditions (PDA1) and (PDA2) but to derive Eq. (3.1) of EMPT directly.

$\Big($ On the other hand, it is not hard to see that conditions (PDA1) and (PDA2) follow from Eq. (3.1) of EMPT for any $s > 0$, provided also that $p(*) > 0$. Indeed, then $p(\mathfrak{t}) = \mathsf{P}(\mathfrak{T} = \mathfrak{t}) > 0$, in view of Eq. (3.2) of EMPT (which follows from Eq. (3.1) of EMPT). Next, recall that $\text{FNE} = \{1 \leqslant |\mathfrak{T}| < \infty\}$ and $\text{FB} = \{2 \leqslant |\mathfrak{T}| < \infty\}$. Let $\mathfrak{P}_{\mathfrak{T}}$ and $\mathfrak{D}_{\mathfrak{T}}$ denote, respectively, the parent and daughter branches of the random tree $\mathfrak{T}$ given FB. Let $\mathfrak{p}$ and $\mathfrak{d}$ denote be any non-random finite non-empty trees. One has

$$\mathsf{P}(\text{FB}) = \mathsf{P}(2 \leqslant |\mathfrak{T}| < \infty) = \sum_{\mathfrak{t}:\ 2 \leqslant |\mathfrak{t}| < \infty} p(\mathfrak{t}) > 0,$$

so that condition (PDA1) of EMPT is satisfied. Also, Eq. (3.1) of EMPT implies that

$$\mathsf{P}(\text{FB}) = s \sum_{\mathfrak{p},\mathfrak{d}:\ 1 \leqslant |\mathfrak{p}|,|\mathfrak{d}| < \infty} p(\mathfrak{p})p(\mathfrak{d}) = s\,\mathsf{P}(1 \leqslant |\mathfrak{T}| < \infty)^2 = s\,\mathsf{P}(\text{FNE})^2.$$

Hence, Eq. (3.1) of EMPT also implies that

$$\mathsf{P}\left(\mathfrak{P}_{\mathfrak{T}} = \mathfrak{p}, \mathfrak{D}_{\mathfrak{T}} = \mathfrak{d} | \text{FB}\right) = \frac{sp(\mathfrak{p})p(\mathfrak{d})}{s\,\mathsf{P}(\text{FNE})^2} = \mathsf{P}\left(\mathfrak{T} = \mathfrak{p} | \text{FNE}\right)\,\mathsf{P}\left(\mathfrak{T} = \mathfrak{d} | \text{FNE}\right),$$

which means that, given FB, the parent and daughter branches $\mathfrak{P}_{\mathfrak{T}}$ and $\mathfrak{D}_{\mathfrak{T}}$ of the random tree $\mathfrak{T}$ are conditionally independent of each other. In turn, it follows that

$$\mathsf{P}\left(\mathfrak{P}_{\mathfrak{T}} = \mathfrak{p} | \text{FB}\right) = \sum_{\mathfrak{d}:\ 1 \leqslant |\mathfrak{d}| < \infty} \mathsf{P}\left(\mathfrak{P}_{\mathfrak{T}} = \mathfrak{p}, \mathfrak{D}_{\mathfrak{T}} = \mathfrak{d} | \text{FB}\right) = \mathsf{P}\left(\mathfrak{T} = \mathfrak{p} | \text{FNE}\right)$$

similarly, $\mathsf{P}\left(\mathfrak{D}_{\mathfrak{T}} = \mathfrak{d} | \text{FB}\right) = \mathsf{P}\left(\mathfrak{T} = \mathfrak{d} | \text{FNE}\right)$. Thus, the conditional distribution of each branch of $\mathfrak{T}$ given FB is the same as the conditional distribution of the entire tree $\mathfrak{T}$ given that $\mathfrak{T}$ is finite and non-empty; that is, one has condition (PDA2) of EMPT. $\Big)$

Concerning the derivation of Eq. (3.1) of EMPT, let us first consider the simpler case when $\mu_{iD} = 0$, i.e., $d = 0$ (no extinction). Then Eq. (3.1) of EMPT obviously holds, with $s = b > 0$ (by Eq. (11)) and $p(*) = q = 1 - b > 0$, so that one has an MR-PDA model.

In the general case (with d possibly nonzero), taking into account that $T = \infty$, one has the equation

$$p(\emptyset) = d + bp(\emptyset)^2$$

(for $p(\emptyset)$), whose only root that is no greater than 1 is

$$p(\emptyset) = \frac{1-\sqrt{1-4bd}}{2b}.$$

Moreover, one has

$$p(\mathbf{t}) = b\big(p(\mathbf{p})p(\boldsymbol{\mathfrak{d}}) + 2p(\mathbf{t})p(\emptyset)\big) \quad \text{if} \quad \mathbf{t} = (\mathbf{p}, \boldsymbol{\mathfrak{d}}),$$

which yields Eq. (3.1) of EMPT with

$$s := \frac{b}{1-2bp(\emptyset)} = \frac{b}{\sqrt{1-4bd}} > 0, \tag{16}$$

by Eq. (11). (Note that, if $q \approx 0$ and $d = b$, then $d = b \approx 1/2$, and so, s is much greater than 1; in particular, it follows that in general s cannot be interpreted as a probability.)

Furthermore,

$$p(*) = q + b \cdot 2p(*)p(\emptyset),$$

whence

$$p(*) = \frac{q}{1-2bp(\emptyset)} = \frac{q}{\sqrt{1-4bd}} > 0.$$

Thus, one still has Eq. (3.1) of EMPT with strictly positive s and $p(*)$, so that indeed one has an instance of the PDA model.

¿From the more general results of the next subsection, it will follow that the random tree $\boldsymbol{\mathfrak{T}}$ in any MR-PDA model is finite with probability 1 if and only if $b \leqslant 1/2$.

3.3. On MR interpretations of the SPDA models, including the CU models

The treatment here is more general than that in the previous subsection, in that it is no longer assumed that $\sigma_{iQ} = 0$ for all i in U; i.e., it is no longer assumed that $\gamma = 0$. At that, all the other MR-PDA conditions in Table 1 of EMPT are still assumed to be satisfied, as well as conditions (MR1), (MR2), and (MR3).

Equations of the previous subsection are generalized here as follows:

$$p(\emptyset) = d + b(1-\gamma)p(\emptyset)^2; \tag{17}$$

$$p(\emptyset) = \frac{1-\sqrt{1-4b(1-\gamma)d}}{2b(1-\gamma)} \quad (\text{if } \gamma < 1); \tag{18}$$

$$p(\mathbf{t}) = b\Big((1-\gamma)\Big(p(\mathbf{p})p(\boldsymbol{\mathfrak{d}}) + 2p(\mathbf{t})p(\emptyset)\Big) + \gamma p(\mathbf{p})\,\mathsf{I}\{\boldsymbol{\mathfrak{d}} = *\}\Big) \quad \text{if} \quad \mathbf{t} = (\mathbf{p}, \boldsymbol{\mathfrak{d}}); \tag{19}$$

$$p(\mathbf{t}) = s\,p(\mathbf{p})\,p(\boldsymbol{\mathfrak{d}}) \quad \text{if} \quad \mathbf{t} = (\mathbf{p}, \boldsymbol{\mathfrak{d}}),$$

where

$$s := b\frac{1-\gamma+\gamma\,\mathsf{I}\{\boldsymbol{\mathfrak{d}} = *\}/p(*)}{1-2b(1-\gamma)p(\emptyset)} = b\frac{1-\gamma+\gamma\,\mathsf{I}\{\boldsymbol{\mathfrak{d}} = *\}/p(*)}{\sqrt{1-4b(1-\gamma)d}} = b(1+\varepsilon)^{\mathsf{I}\{\boldsymbol{\mathfrak{d}}=*\}}\,B,$$

$$\varepsilon := \frac{\gamma/p(*)}{1-\gamma} \geqslant 0 \quad (\text{if } \gamma < 1), \qquad B := b\,\frac{1-\gamma}{\sqrt{1-4b(1-\gamma)d}};$$

$$p(*) = q + b\big((1-\gamma)\cdot 2p(*)p(\emptyset) + \gamma p(\emptyset)\big);$$

$$p(*) = \frac{q + b\gamma p(\emptyset)}{1-2b(1-\gamma)p(\emptyset)} = \frac{q + b\gamma p(\emptyset)}{\sqrt{1-4b(1-\gamma)d}}. \tag{20}$$

It follows by induction that for any non-random finite non-empty tree $\mathbf{t}$

$$p(\mathbf{t}) = (1+\varepsilon)^{|\mathbf{t}|_{\mathsf{uds}}}B^{|\mathbf{t}|-1}p(*)^{|\mathbf{t}|} = \alpha^{|\mathbf{t}|_{\mathsf{uds}}}\beta^{|\mathbf{t}|_{\mathsf{nuds}}}p(*), \tag{21}$$

where

$$\beta = Bp(*), \quad \alpha = (1+\varepsilon)\beta,$$

and $|\boldsymbol{\mathfrak{t}}|_{\mathsf{uds}}$ and $|\boldsymbol{\mathfrak{t}}|_{\mathsf{nuds}}$ denote, respectively, the numbers of unitary-daughter and non-unitary-daughter splits in a finite tree $\boldsymbol{\mathfrak{t}}$. These numbers can be defined recursively, as follows: $|\emptyset|_{\mathsf{uds}} = |\emptyset|_{\mathsf{nuds}} = |*|_{\mathsf{uds}} = |*|_{\mathsf{nuds}} = 0$; if $\boldsymbol{\mathfrak{t}} = (\boldsymbol{\mathfrak{p}}, \boldsymbol{\mathfrak{d}})$, then

$$|\boldsymbol{\mathfrak{t}}|_{\mathsf{uds}} = \mathsf{I}\{\boldsymbol{\mathfrak{d}} = *\} + |\boldsymbol{\mathfrak{p}}|_{\mathsf{uds}} + |\boldsymbol{\mathfrak{d}}|_{\mathsf{uds}} \quad \text{and} \quad |\boldsymbol{\mathfrak{t}}|_{\mathsf{nuds}} = \mathsf{I}\{\boldsymbol{\mathfrak{d}} \neq *\} + |\boldsymbol{\mathfrak{p}}|_{\mathsf{nuds}} + |\boldsymbol{\mathfrak{d}}|_{\mathsf{nuds}};$$

it follows by induction that for any finite non-empty tree $\boldsymbol{\mathfrak{t}}$

$$|\boldsymbol{\mathfrak{t}}|_{\mathsf{uds}} + |\boldsymbol{\mathfrak{t}}|_{\mathsf{nuds}} = |\boldsymbol{\mathfrak{t}}| - 1,$$

the total numbers of splits in $\boldsymbol{\mathfrak{t}}$.

By Eq. (21), the probability $p(\boldsymbol{\mathfrak{t}})$ is an increasing function of $|\boldsymbol{\mathfrak{t}}|_{\mathsf{uds}}$, the number of unitary-daughter splits.

However, $p(\boldsymbol{\mathfrak{t}})$ here is not a function of the tree type of $\boldsymbol{\mathfrak{t}}$; in general, the probability $p((\boldsymbol{\mathfrak{p}}, \boldsymbol{\mathfrak{d}}))$ of tree $(\boldsymbol{\mathfrak{p}}, \boldsymbol{\mathfrak{d}})$ differs from $p((\boldsymbol{\mathfrak{d}}, \boldsymbol{\mathfrak{p}}))$.

Define the reflection $\tilde{\boldsymbol{\mathfrak{t}}}$ of any finite non-empty tree $\boldsymbol{\mathfrak{t}}$ recursively, as follows: $\tilde{*} = *$, and $\tilde{\boldsymbol{\mathfrak{t}}} = (\tilde{\boldsymbol{\mathfrak{d}}}, \tilde{\boldsymbol{\mathfrak{p}}})$ if $\boldsymbol{\mathfrak{t}} = (\boldsymbol{\mathfrak{p}}, \boldsymbol{\mathfrak{d}})$. In words, the roles of all corresponding parent and daughter branches are mutually interchanged. E.g., the reflection of tree $(*, ((*, *), *))$ is $((*, (*, *)), *)$.

By induction, it is not difficult to check that for any finite non-empty tree $\boldsymbol{\mathfrak{t}}$

$$|\boldsymbol{\mathfrak{t}}|_{\mathsf{uds}} + |\tilde{\boldsymbol{\mathfrak{t}}}|_{\mathsf{uds}} = |\boldsymbol{\mathfrak{t}}|\, \mathsf{I}\{\boldsymbol{\mathfrak{t}} \neq *\}.$$

Let us now pair all finite trees of size $\geqslant 2$, each such tree $\boldsymbol{\mathfrak{t}}$ with its reflection $\tilde{\boldsymbol{\mathfrak{t}}}$. For any such pair, let $u := \max\big(|\boldsymbol{\mathfrak{t}}|_{\mathsf{uds}}, |\tilde{\boldsymbol{\mathfrak{t}}}|_{\mathsf{uds}}\big)$ and $w := |\boldsymbol{\mathfrak{t}}| = |\tilde{\boldsymbol{\mathfrak{t}}}|$, so that $u \geqslant w/2$. Then, in view of Eq. (21), with the size $w \geqslant 2$ of the tree $\boldsymbol{\mathfrak{t}}$ fixed, it is not difficult to check that the sum

$$p(\boldsymbol{\mathfrak{t}}) + p(\tilde{\boldsymbol{\mathfrak{t}}}) = \big((1+\varepsilon)^u + (1+\varepsilon)^{w-u}\big)\, B^{w-1} p(*)^w$$

of the probabilities of $\boldsymbol{\mathfrak{t}}$ and $\tilde{\boldsymbol{\mathfrak{t}}}$ is an increasing function of $u = \max\big(|\boldsymbol{\mathfrak{t}}|_{\mathsf{uds}}, |\tilde{\boldsymbol{\mathfrak{t}}}|_{\mathsf{uds}}\big)$. In particular, the sum of the probabilities $p(\boldsymbol{\mathfrak{t}})$ and $p(\tilde{\boldsymbol{\mathfrak{t}}})$ of the pair of completely unbalanced trees $\boldsymbol{\mathfrak{t}} = (*, ((*, *), *))$ and $\tilde{\boldsymbol{\mathfrak{t}}} = ((*, (*, *)), *)$ in any MR-SPDA model is greater than twice the probability of the completely balanced tree $((*, *), (*, *))$ of the same size 4.

It is intuitively clear that in the case $\gamma = 1$ one has an MR-CU model. The same conclusion follows from Eqs. (17), (19), and (20), which for $\gamma = 1$ turn into the equations

$$p(\emptyset) = d, \quad p(\boldsymbol{\mathfrak{t}}) = b\, p(\boldsymbol{\mathfrak{p}})\, \mathsf{I}\{\boldsymbol{\mathfrak{d}} = *\} \text{ if } \boldsymbol{\mathfrak{t}} = (\boldsymbol{\mathfrak{p}}, \boldsymbol{\mathfrak{d}}), \quad \text{and} \quad p(*) = q + bd = (1-d)(1-b),$$

respectively. Hence, at $\gamma = 1$,

$$p(\emptyset) = d \quad \text{and} \quad p(\boldsymbol{\mathfrak{t}}) = \begin{cases} 0 & \text{if } \boldsymbol{\mathfrak{t}} \neq \emptyset \ \& \ |\boldsymbol{\mathfrak{t}}|_{\mathsf{nuds}} \neq 0, \\ (1-d)(1-b)b^{|\boldsymbol{\mathfrak{t}}|-1} & \text{if } \boldsymbol{\mathfrak{t}} \neq \emptyset \ \& \ |\boldsymbol{\mathfrak{t}}|_{\mathsf{nuds}} = 0. \end{cases}$$

Thus, at $\gamma = 1$, the random tree $\boldsymbol{\mathfrak{T}}$ is completely unbalanced with probability 1. (At that, if $d \neq 1$, then the conditional distribution of the size $|\boldsymbol{\mathfrak{T}}|$ of $\boldsymbol{\mathfrak{T}}$ given that $\boldsymbol{\mathfrak{T}} \neq \emptyset$ is geometric.)

Moreover, for every pair of positive values b and d such that $b + d \leqslant 1$, letting parameter γ vary from 0 to 1, one obtains a continuous one-parameter spectrum of MR-SPDA models interpolating between MR-PDA (at $\gamma = 0$) and MR-CU (at $\gamma = 1$) models. Such spectra may be referred to as γ-spectra.

Next, let us obtain a necessary and sufficient condition for the random tree $\boldsymbol{\mathfrak{T}}$ in an (S)PDA model to be finite with probability 1. Let $N_{u,v}$ denote the number of the finite *non-empty* trees $\boldsymbol{\mathfrak{t}}$ with $|\boldsymbol{\mathfrak{t}}|_{\mathsf{uds}} = u$ and $|\boldsymbol{\mathfrak{t}}|_{\mathsf{nuds}} = v$, where $u, v = 0, 1, \dots$. Then, in view of Eq. (21),

$$\mathsf{P}(|\boldsymbol{\mathfrak{T}}| < \infty) = p(\emptyset) + \sum_{\boldsymbol{\mathfrak{t}}\colon 0 < |\boldsymbol{\mathfrak{t}}| < \infty} p(\boldsymbol{\mathfrak{t}}) = p(\emptyset) + f(\alpha, \beta) p(*), \tag{22}$$

where

$$f(x, y) := \sum_{u,v \geqslant 0} N_{u,v} x^u y^v$$

is the generating function of the $N_{u,v}$'s. In particular, it follows that $f(x,y) < \infty$ for all small enough positive x and y.

On the other hand, for all $u, v = 1, 2, \dots$ one has

$$N_{u,v} = N_{0,0} N_{u-1,v} + \sum_{\substack{i+j \geqslant 1 \\ 0 \leqslant i \leqslant u \\ 0 \leqslant j \leqslant v-1}} N_{i,j} N_{u-i,v-1-j}.$$

Multiplying both sides of these equations by $x^u y^v$ and then summing in all $u, v = 1, 2, \dots$ and taking into account that $N_{u,0} = 1$ and $N_{0,v} = \mathsf{I}\{v = 0\}$ for all u and v, one obtains the quadratic equation

$$yf^2 - (1 + y - x)f + 1 = 0$$

for $f = f(x,y)$. It follows that (say) for all small enough positive x and y

$$f(x,y) = \frac{1 + y - x - \sqrt{(1 + y - x)^2 - 4y}}{2y}; \tag{23}$$

the minus sign before the radical is chosen here because the function f is analytic in a neighbourhood of $(x,y) = (0,0)$ and $f(0,y) = \sum_{v \geqslant 0} N_{0,v} y^v = 1$ for all y.

Now, using Eqs. (22) and $q = 1 - b - d$, one obtains

$$\mathsf{P}(|\mathfrak{T}| < \infty) - 1 = -\frac{|1 - (2 - \gamma)b| - (1 - (2 - \gamma)b)}{2(1 - \gamma)b}$$

(this calculation is better done using Mathematica or similar software). Hence, in an MR-SPDA model the random tree

$\mathfrak{T}$ is finite with probability 1 if and only if $(2 - \gamma)b \leqslant 1$.

Moreover, using Eq. (21), one can find an expression for the joint generating function of the numbers $|\mathfrak{T}|_{\mathsf{uds}}$ and $|\mathfrak{T}|_{\mathsf{nuds}}$ of the unitary-daughter and non-unitary-daughter splits in the random tree $\mathfrak{T}$:

$$G(x,y) = \mathsf{E} x^{|\mathfrak{T}|_{\mathsf{uds}}} y^{|\mathfrak{T}|_{\mathsf{nuds}}} = p(\emptyset) + p(*) f(\alpha x, \beta y),$$

for all x and y in $(0,1]$, where $f(x,y)$ is given by Eq. (23). This formula, together with (20) and (18), determines the joint distribution of $|\mathfrak{T}|_{\mathsf{uds}}$ and $|\mathfrak{T}|_{\mathsf{nuds}}$, and hence the distribution of the size

$$|\mathfrak{T}| = |\mathfrak{T}|_{\mathsf{uds}} + |\mathfrak{T}|_{\mathsf{nuds}} + 1.$$

The corresponding expressions are simpler if $d = 0$; in particular, then one has

$$\mathsf{E}|\mathfrak{T}|_{\mathsf{uds}} = G'_x(1,1) = b\,\frac{1 - b + \gamma b}{1 - 2b + \gamma b} \quad \text{and} \quad \mathsf{E}|\mathfrak{T}|_{\mathsf{nuds}} = G'_y(1,1) = b\,\frac{b - \gamma b}{1 - 2b + \gamma b},$$

whence

$$\mathsf{E}|\mathfrak{T}| = 1 + \mathsf{E}|\mathfrak{T}|_{\mathsf{uds}} + \mathsf{E}|\mathfrak{T}|_{\mathsf{nuds}} = \frac{1 - b + \gamma b}{1 - 2b + \gamma b},$$

so that

$$\mathsf{E}|\mathfrak{T}|_{\mathsf{uds}} = b\,\mathsf{E}|\mathfrak{T}|.$$

4. CONTINUOUS ONE-PARAMETER SPECTRA OF MR MODELS BRIDGING THE ERM AND PDA MODELS

It is clear from Section 5 of EMPT that there are infinitely many such spectra. In this section, just two simple kinds of them are considered, which are called here the α- and β-spectra. As such, these spectra provide answers to the problem, posed by Aldous 1996 & 2001, to find a mathematically simple spectrum of evolutionary models interpolating between the ERM and PDA models.

For these purposes, it is enough to have just two states, say 1 and 2, so that the state space $S = \{1, 2\}$. Assume also that all the MR models considered in this section are complete and $D = \emptyset$. Moreover, assume that $\mu_{11} = \mu_{12} = \mu_{22} = \sigma_{12} = \sigma_{21} = 0$, $\mu_{21} = \mu$, $\sigma_{11} = \sigma_1$, and $\sigma_{22} = \sigma_2$, where μ, σ_1, and σ_2 are nonnegative numbers.

Let $\mathfrak{T}_T$ denote the random tree in such an MR model observed over the time interval $[0, T]$. For every given non-random finite tree $\mathfrak{t}$ and $i = 1, 2$, let $p_{i;T}(\mathfrak{t})$ denote the conditional probability that the random tree $\mathfrak{T}_T$ happens to coincide with $\mathfrak{t}$ given that the initial state was i.

Then for all $T > 0$

$$p_{1;T}(*) = e^{-\sigma_1 T}; \tag{24}$$

$$p_{2;T}(*) = e^{-(\mu+\sigma_2)T} + \int_0^T e^{-(\mu+\sigma_2)t} \mu \, dt \, p_{1;T-t}(*) = \frac{\mu e^{-\sigma_1 T} + (\sigma_2 - \sigma_1) e^{-(\mu+\sigma_2)T}}{\mu + \sigma_2 - \sigma_1} \tag{25}$$

(if $\mu + \sigma_2 - \sigma_1 \neq 0$).

Next, for $\mathfrak{t} = (\mathfrak{p}, \mathfrak{d})$ one has the recursive equations

$$p_{1;T}(\mathfrak{t}) = \int_0^T e^{-\sigma_1 t} \sigma_1 \, dt \, p_{1;T-t}(\mathfrak{p}) p_{1;T-t}(\mathfrak{d}); \tag{26}$$

$$p_{2;T}(\mathfrak{t}) = \int_0^T e^{-(\mu+\sigma_2)t} \Big(\sigma_2 \, dt \, p_{2;T-t}(\mathfrak{p}) p_{2;T-t}(\mathfrak{d}) + \mu \, dt \, p_{1;T-t}(\mathfrak{p}) p_{1;T-t}(\mathfrak{d}) \Big). \tag{27}$$

These equations allow one to obtain by induction expressions for the probabilities $p_{1;T}(\mathfrak{t})$ and $p_{2;T}(\mathfrak{t})$ for any finite tree $\mathfrak{t}$. These expressions will be polynomials in $e^{-\sigma_1 t}$, $e^{-(\mu+\sigma_2)T}$, and T, with coefficients being rational functions of μ, σ_1, and σ_2.

Similarly, for the conditional expected size $\mathsf{E}_i|\mathfrak{T}_T|$ given an initial state $i = 1, 2$, one has the integral equations

$$\mathsf{E}_1|\mathfrak{T}_T| = e^{-\sigma_1 T} + \int_0^T e^{-\sigma_1 t} \, \sigma_1 \, dt \cdot 2\mathsf{E}_1|\mathfrak{T}_{T-t}|; \tag{28}$$

$$\mathsf{E}_2|\mathfrak{T}_T| = e^{-(\mu+\sigma_2)T} + \int_0^T e^{-(\mu+\sigma_2)t} \Big(\mu \, dt \cdot \mathsf{E}_1|\mathfrak{T}_{T-t}| + \sigma_2 \, dt \cdot 2\mathsf{E}_2|\mathfrak{T}_{T-t}| \Big). \tag{29}$$

Making the substitution $u = T - t$ in the integrals in Eqs. (28) and (29), then multiplying both sides of the equations by $e^{\sigma_1 T}$ and $e^{(\mu+\sigma_2)T}$ (respectively), and finally differentiating in T, one has the differential equations

$$\frac{d \, \mathsf{E}_1|\mathfrak{T}_T|}{dT} = \sigma_1 \mathsf{E}_1|\mathfrak{T}_T|;$$

$$\frac{d \, \mathsf{E}_2|\mathfrak{T}_T|}{dT} = (\sigma_2 - \mu)\mathsf{E}_2|\mathfrak{T}_T| + \mu \mathsf{E}_1|\mathfrak{T}_T|.$$

Together with the initial conditions $\mathsf{E}_1|\mathfrak{T}_0| = \mathsf{E}_2|\mathfrak{T}_0| = 1$, these differential equations yield

$$\mathsf{E}_1|\mathfrak{T}_T| = e^{\sigma_1 T};$$

$$\mathsf{E}_2|\mathfrak{T}_T| = \frac{\mu e^{\sigma_1 T} + (\sigma_1 - \sigma_2) e^{(\sigma_2 - \mu)T}}{\mu + \sigma_1 - \sigma_2}; \tag{30}$$

the latter expression, for $\mathsf{E}_2|\mathfrak{T}_T|$, holds only if $\mu + \sigma_1 - \sigma_2 \neq 0$; otherwise, as if by continuity, $\mathsf{E}_2|\mathfrak{T}_T| = e^{\sigma_1 T}(1 + \mu T)$.

Let us now further specify the parameters, as follows:

$$\mu > 1, \ \sigma_1 = \alpha, \ \sigma_2 = 1 - \alpha, \tag{31}$$

where α is a parameter, varying from 0 to 1. At that, let the initial state be 2 with probability 1 and let the stopping time $T > 0$ continuously depend on the value of the parameter α, so that $T = T(\alpha)$. Moreover, assume that $T(0) = \infty$ while $T(\alpha) < \infty$ if $0 < \alpha \leqslant 1$. The continuity of $T(\alpha)$ at $\alpha = 0$ is taken to mean here that $T(\alpha) \to \infty$ as $\alpha \downarrow 0$. Let here $\mathfrak{T}_{\alpha,T}$ denote the random tree observed over the time interval $[0, T]$ when the value of the parameter is α. For any non-random tree $\mathbf{t}$, let $p_{i;\alpha,T}(\mathbf{t})$ stand for the conditional probability that the random tree $\mathfrak{T}_{\alpha,T}$ happens to coincide with $\mathbf{t}$ given the initial state i.

Then, in view of results stated in Section 5 of EMPT, one has an MR-PDA model (with $Q = \{1\}$ and $U = \{2\}$) at $\alpha = 0$.

Moreover, one has an MR-ERM model (now with $Q = \emptyset$ and $U = \{1, 2\}$) at $\alpha = 1/2$ (when $\sigma_1 = \sigma_2$), according to the remark at the end of Section 5 of EMPT.

Furthermore, here one has an MR-ERM model at $\alpha = 1$ (when $\sigma_2 = 0$) as well. Indeed, in this case the root species will eventually (with probability 1) switch from the initial state 2 to state 1, and after that the evolution will proceed as in the pure-birth process with rate $\sigma_1 = 1$. In this sense, one has a "delayed" ERM-type random tree. However, the conditional distribution of the tree given any size will be the same as that in the standard, "non-delayed" pure-birth process.

Thus, for every $\mu > 1$ and every appropriately chosen stopping-time function $T(\alpha)$, one has a one-parameter spectrum of MR models interpolating from an MR-PDA model (at $\alpha = 0$) to an MR-ERM one (at $\alpha = 1/2$) to another MR-ERM one (at $\alpha = 1$). Let us refer to such a spectrum as an α-spectrum.

Note that the set Q of quasi-stable states is empty for all values of α in the interval $[0, 1]$ – except for $\alpha = 0$, when $Q = \{1\}$.

One must be somewhat careful here so as not to let $T(\alpha)$ grow to ∞ too fast as $\alpha \downarrow 0$. Indeed, in view of Eqs. (25) and (31), for any fixed $\alpha \in (0, 1)$ one has $p_{2;\alpha,T}(*) \to 0$ as $T \to \infty$. Therefore, one would have $p_{2;\alpha,T(\alpha)}(*) \to 0$ as $\alpha \to 0$ if $T(\alpha)$ were allowed to grow to ∞ too fast as $\alpha \downarrow 0$. On the other hand, $p_{2;\alpha,T(\alpha)}(*)$ is strictly positive at $\alpha = 0$, as it must be in any MR-PDA model. Thus, one has a somewhat subtle problem here with the continuity of an α-spectrum at $\alpha = 0$.

One way to obtain a *continuous* α-spectrum is to fix any finite value $\mathit{es} > 1$ of the expected size of the tree, so that the stopping time $T = T(\alpha)$ is given by the equation $\mathsf{E}_2|\mathfrak{T}_{\alpha,T}| = \mathit{es}$. Then a unique solution $T = T(\alpha) > 0$ exists for all α in the interval $(0, 1]$, because $\mathsf{E}_2|\mathfrak{T}_{\alpha,T}|$ is continuously increasing in T from 1 to ∞ as T increases from 1 to ∞. In view of (30),

$$\mathsf{E}_2|\mathfrak{T}_{\alpha,T}| = \frac{\mu e^{\alpha T} + (2\alpha - 1)e^{(1-\alpha-\mu)T}}{\mu + 2\alpha - 1}. \tag{32}$$

Hence and because $\mu > 1$, one must have

$$\mathit{es} = \mathsf{E}_2|\mathfrak{T}_{0,\infty}| = \frac{\mu}{\mu - 1}.$$

Moreover, if $\alpha \downarrow 0$ while $T(\alpha)$ stays bounded (say by some finite positive number T_0), then the limit superior of $\mathsf{E}_2|\mathfrak{T}_{\alpha,T}|$ cannot exceed

$$\frac{\mu - e^{(1-\mu)T_0}}{\mu - 1},$$

which is strictly less than $\frac{\mu}{\mu-1} = \mathit{es}$. This shows that indeed $T(\alpha) \to \infty$ as $\alpha \downarrow 0$. Therefore, $e^{(1-\alpha-\mu)T(\alpha)} \to 0$ as $\alpha \downarrow 0$. Hence, the equations

$$\frac{\mu e^{\alpha T(\alpha)} + (2\alpha - 1)e^{(1-\alpha-\mu)T(\alpha)}}{\mu + 2\alpha - 1} = \mathit{es} = \frac{\mu}{\mu - 1}$$

imply that $\alpha T(\alpha) \to 0$ as $\alpha \downarrow 0$. Now Eqs. (25) and (31) show that

$$p_{2;\alpha,T(\alpha)}(*) = \frac{\mu e^{-\alpha T(\alpha)} + (1 - 2\alpha)e^{-(\mu+1-\alpha)T(\alpha)}}{\mu + 1 - 2\alpha} \to \frac{\mu}{\mu + 1} = p_{2;0,\infty}(*) = p_{2;0,T(0)}(*).$$

It follows by induction using Eqs. (24), (26), and (27) that, for all finite trees $\mathbf{t}$,

$$p_{1;\alpha,T(\alpha)}(\mathbf{t}) \to \mathsf{I}\{\mathbf{t} = *\} = p_{1;0,\infty}(\mathbf{t}) = p_{1;0,T(0)}(\mathbf{t}) \quad \text{and} \quad p_{2;\alpha,T(\alpha)}(\mathbf{t}) \to p_{2;0,T(0)}(\mathbf{t})$$

as $\alpha \downarrow 0$.

Thus, it has been verified that the α-spectrum with any fixed expected tree size $es > 1$ is continuous, even at $\alpha = 0$.

Another kind of spectra interpolating between MR-PDA and MR-ERM models is obtained if conditions (31) are replaced by

$$\sigma_1 = 0, \ \sigma_2 = \beta, \ \mu = 1 - \beta,$$

with the parameter β varying from some fixed value $\beta_{\min}$ to 1, where $\beta_{\min}$ is a number strictly between 0 and 1. The stopping time here can be given by any continuous function $T = T(\beta)$ on the interval $[\beta_{\min}, 1]$ such that $T(\beta) < \infty$ if $\beta_{\min} < \beta \leqslant 1$ and $T(\beta) \to T(\beta_{\min}) = \infty$ as $\beta \downarrow \beta_{\min}$.

Then one has an MR-PDA model at $\beta = \beta_{\min}$ and an MR-ERM model at $\beta = 1$. Such a spectrum can be referred to as a β-spectrum.

The problem of a possible discontinuity at the left endpoint does not arise for a β-spectrum because here $p_{1;\beta,T}(\mathbf{t}) = \mathsf{I}\{\mathbf{t} = *\}$ for all $T \geqslant 0$ and all β and, by Eq. (25),

$$p_{2;\beta,T}(*) \to 1 - \beta_{\min} = p_{2;\beta_{\min},\infty}(*) = p_{2;\beta_{\min},T(\beta_{\min})}(*)$$

as $\beta \downarrow \beta_{\min}$ and $T \to \infty$ in whatever manner. In fact, the MR models comprising a β-spectrum are mathematically equivalent to BDP's with the dead particles and their lineages included into the tree.

5. OTHER SPECIAL CASES OF THE GENERAL MR MODEL

One may try to further extend the MR family of evolutionary models by making the rates μ_{ij} and σ_{ij} dependent on the age (say a), of the species, so that $\mu_{ij} = \mu_{ij}(a)$ and $\sigma_{ij} = \sigma_{ij}(a)$. There are two distinct possibilities here:

(i) the age of the species is reset to 0 every time it gives birth to another species and

(ii) it is never reset.

Let us call the corresponding models multi-rate *reset-age-dependent (MRRAD)* and multi-rate *age-dependent (MRAD)*, respectively.

In the simplest case of the MRAD and MRRAD models, there is only one state (say labeled by 1) and given positive numbers A and λ so that $\mu_{11}(a) = 0$ for all values $a \geqslant 0$ of the age of the species, $\sigma_{11}(a) = 0$ for $a < A$, and $\sigma_{11}(a) = \lambda$ for $a \geqslant A$. These special cases of the MRAD and MRRAD models correspond to the models proposed by Losos & Adler 1995 and will be referred here to as the LA-AD and LA-RAD models. The number A was referred to by Losos & Adler as the length of the "refractory period".

On the other hand, $1/\lambda$ has the meaning of the expected time to speciation – counted from the end of the refractory period $[0, A)$. The case when the length A of the refractory period is much greater than the expected time $1/\lambda$ to speciation is especially instructive here, and then the LA-AD and LA-RAD models are extremely different from each other. In fact, it is not difficult to see that then the LA-AD and LA-RAD models can be however close to the CU and CB models, respectively.

Indeed, let $\lambda \to \infty$ while $A > 0$ is kept fixed. Then, in the limit, in the LA-RAD model every species splits into two exactly in time A after its birth. Thus, the random tree in the LA-RAD model with $0 < A < \infty$ and $\lambda = \infty$ is completely balanced with probability 1. Moreover, the LA-RAD model continuously varies from an ERM model to a CB one as λA varies from 0 to ∞.

In contrast, in the LA-AD model with $0 < A < \infty$ and $\lambda = \infty$, the root species speciates infinitely often in any time interval after the refractory period $[0, A)$; hence, within the time interval $[A, 2A)$ (say) it produces infinitely many daughter species, none of which can speciate within the time interval $[A, 2A)$, because their refractory periods do not expire before time moment $2A$. Hence, the random tree in the LA-AD model with $A < \infty$ and $\lambda = \infty$ is completely unbalanced with probability 1. Moreover, the LA-AD model continuously varies from an ERM model to a CU one as λA varies from 0 to ∞.

On the other hand, one can see that the MRAD and MRRAD models are not really extensions of the MR family; rather, they are parts of that same family. Indeed, as was pointed out in Section 4 of EMPT, the age of the species can be considered as just another attribute of the state of the species.

Let us illustrate this approach by showing that both the LA-AD and LA-RAD models belong to the (original, not ostensibly "extended") MR family. Because of the (inessential) assumption that the state space S is countable, let us discretize time by considering only the time moments that are the integer multiples of an infinitesimal time period dt.

To obtain an MR representation of the LA-AD model, let us identify the state i of the species with its age a. Then, for any two states i and j, set

$$\sigma_{ij}\,dt = \lambda\,dt\,\mathsf{I}\{i \geqslant A, j = 0\}; \tag{33}$$

$$\mu_{ij}\,dt = (1 - \sigma_{i0}\,dt)\,\mathsf{I}\{j = i + dt\}, \tag{34}$$

where I stands for the indicator defined in Section 1 of this Appendix. Thus, $\mu_{ij} = 0$ if $j \neq i + dt$ and $\mu_{ij} = \frac{1-\sigma_{i0}\,dt}{dt} \approx \infty$ if $j = i + dt$. Hence, formula (34) says the following: for any given time moment t, if the age of a species at time t was i, then with probability $1 - \sigma_{i0}\,dt \approx 1$ the age of the species at time $t + dt$ will be $j = i + dt$. Similarly, formula (33) says the following: for any given time moment t, if the age of a species at time t was i, then

(i) with probability $\lambda\,dt$ the species will give birth to another, daughter species at time $t + dt$ and the age of the daughter species at time $t + dt$ (of its birth) will be $j = 0$ – all this provided that $i \geqslant A$;

(ii) if, however, $i < A$, then it is certain that no speciation occurs at time $t + dt$.

Thus, it is seen that the LA-AD model is a particular instance of the MR family.

To obtain an MR representation of the LA-RAD model, let us identify the state i of the species with the ordered pair (a, ω), where a is again the age and ω is the indicator of giving birth; that is, $\omega = 1$ if the species gives birth at the given time moment and $\omega = 0$ otherwise. Then for any two states $i = (a_i, \omega_i)$ and $j = (a_j, \omega_j)$, set

$$\sigma_{ij}\,dt = \mathsf{I}\{a_i \geqslant A, \omega_i = 1, a_j = 0, \omega_j = 0\}; \tag{35}$$

$$\begin{aligned} \mu_{ij}\,dt = {} & \mathsf{I}\{a_i < A, a_j = a_i + dt, \omega_j = 0\} \\ & + \mathsf{I}\{a_i \geqslant A, \omega_i = 0\}\Big(\lambda\,dt\,\mathsf{I}\{\omega_j = 1, a_j = 0\} + (1 - \lambda\,dt)\,\mathsf{I}\{\omega_j = 0, a_j = a_i + dt\}\Big). \end{aligned} \tag{36}$$

¿From this, it is seen that the LA-RAD model, too, is a particular instance of the MR family.

Moreover, if one mutually interchanges the conditions $a_i < A$ and $a_i \geqslant A$ in Eqs. (35) and (36) (and also allows λ to depend on a_i), then one obtains an MR representation of the model that was considered by Steel & McKenzie 2001 and mentioned in EMPT.

Heard 1996 (Table 1 therein) considered 9 distinguishable models of evolution, in which the speciation rate depends on a particular trait of the species. Let us present formulas similar to (33)–(36) just for two of those 9 models, to show that they too belong to the same MR family; the other 7 models can be treated similarly. Here, identify the state i of the species with the ordered pair (ℓ, ε), where ℓ and ε are real numbers, interpreted as the (base-10) logarithms of the trait value and the noise value, respectively. Because of the (inessential) assumption that the state space S is countable, let us consider only the discrete values ℓ and ε that are the integer multiples of infinitesimal positive numbers $d\ell$ and $d\varepsilon$, respectively. Fix also two positive numbers, denoted here by s_ℓ and s_ε, which will be used to scale the values of ℓ and ε. Let $\varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, the density function of the standard normal distribution.

Then for any two states $i = (\ell_i, \varepsilon_i)$ and $j = (\ell_j, \varepsilon_j)$ in the *log-Brownian punctuated intermittent one-daughter-change* model in Heard 1996 set

$$\begin{aligned} \sigma_{ij}\,dt &= 10^{\ell_i + \varepsilon_i} dt \cdot \varphi\left(\frac{\ell_j - \ell_i}{s_\ell}\right) \frac{d\ell}{s_\ell} \cdot \varphi\left(\frac{\varepsilon_j}{s_\varepsilon}\right) \frac{d\varepsilon}{s_\varepsilon}; \\ \mu_{ij}\,dt &= 0\,dt. \end{aligned}$$

To obtain a similar MR representation of the *log-Brownian gradual* model in Heard 1996, set

$$\begin{aligned} \sigma_{ij}\,dt &= 10^{\ell_i + \varepsilon_i} dt \cdot \varphi\left(\frac{\ell_j - \ell_i}{s_\ell \sqrt{dt}}\right) \frac{d\ell}{s_\ell \sqrt{dt}} \cdot \varphi\left(\frac{\varepsilon_j}{s_\varepsilon}\right) \frac{d\varepsilon}{s_\varepsilon}; \\ \mu_{ij}\,dt &= \varphi\left(\frac{\ell_j - \ell_i}{s_\ell \sqrt{dt}}\right) \frac{d\ell}{s_\ell \sqrt{dt}} \cdot \varphi\left(\frac{\varepsilon_j}{s_\varepsilon}\right) \frac{d\varepsilon}{s_\varepsilon}. \end{aligned}$$

REFERENCES

Aldous, D. J. 1996 Probability distributions on cladograms. In *Random Discrete Structures* (eds. D. Aldous & R. Pemantle), pp. 1–18. Springer.

Aldous, D. J. 2001 Stochastic models and descriptive statistics for phylogenetic trees. *Statist. Sci.* **16,** 23–34.

Harding, E. F. 1971 The probabilities of rooted tree-shapes generated by random bifurcation. *Adv. Appl. Prob.* **3,** 44–77.

Heard, S. B. 1996 Patterns in phylogenetic tree balance with variable and evolving speciation rates. *Evolution* **50,** 2141–2148.

Kendall, D. G. 1948 On the generalized "birth-and-death" process. *Ann. Math. Statist.* **19,** 1–15.

Karlin, S. 1992 *Mathematical Methods and Theory in Games, Programming, and Economics.* Dover, New York.

Losos, J. B. & Adler, F. R. 1995 Stumped by trees? A generalized null model for patterns of organismal diversity. *Amer. Nat.* **145,** 329–342.

Mooers, A. Ø. & Heard, S. B. 1997 Inferring evolutionary process from phylogenetic tree shape. *Q. Rev. Biol.* **72,** 31–54.

Rogers, J. S. 1994 Central moments and probability distribution of Colless's coefficient of tree imbalance. *Evolution* **48,** 2026–2036.

Rosen, D. E. 1978 Vicariant patterns and historical explanation in biogeography. *Syst. Zool.* **27,** 159–188.

Slowinski, J. B. & Guyer, C. 1989 Testing the stochasticity of patterns of organismal diversity: an improved null model. *Am. Nat.* **134,** 907–920.

Steel, M. & McKenzie, A. 2001 Properties of phylogenetic trees generated by Yule-type speciation models. *Math. Biosci.* **170,** 91–112.