A Central Limit Theorem for Punctuated Equilibrium

K. Bartoszek *

Abstract

Current evolutionary biology models usually assume that a phenotype undergoes gradual change. This is in stark contrast to biological intuition, which indicates that change can also be punctuated — the phenotype can jump. Such a jump can especially occur at speciation, i.e. dramatic change occurs that drives the species apart. Here we derive a central limit theorem for punctuated equilibrium. We show that, if adaptation is fast, for weak convergence to hold, dramatic change has to be a rare event.

Keywords : Branching diffusion process, Conditioned branching process, Central Limit Theorem, Lévy process, Punctuated equilibrium, Yule–Ornstein–Uhlenbeck with jumps process

AMS subject classification : 60F05, 60J70, 60J85, 62P10, 92B99

1 Introduction

A long–standing debate in evolutionary biology is whether changes take place at times of speciation (punctuated equilibrium Eldredge and Gould [16], Gould and Eldredge [20]) or gradually over time (phyletic gradualism, see references in Eldredge and Gould [16]). Phyletic gradualism is in line with Darwin's original envisioning of evolution (Eldredge and Gould [16]). On the other hand, the theory of punctuated equilibrium was an answer to what fossil data was indicating (Eldredge and Gould [16], Gould and Eldredge [19, 20]). A complete unbroken fossil series was rarely observed, rather

*bartoszekkj@gmail.com, Department of Mathematics, Uppsala University, 751 06, Uppsala, Sweden

distinct forms separated by long periods of stability (Eldredge and Gould [16]). Darwin saw "the fossil record more as an embarrassment than as an aid to his theory" (Eldredge and Gould [16]) in the discussions with Falconer at the birth of the theory of evolution. Evolution with jumps was proposed under the name "quantum evolution" (Simpson [32]) to the scientific community. However, only later (Eldredge and Gould [16]) was punctuated equilibrium re–introduced into contemporary mainstream evolutionary theory. Mathematical modelling of punctuated evolution on phylogenetic trees seems to be still in its infancy (but see Bokma [8, 10, 11], Mattila and Bokma [23], Mooers and Schluter [25], Mooers et al. [26]). The main reason is that we do not seem to have sufficient understanding of the stochastic properties of these models. An attempt was made in this direction (Bartoszek [5]) — to derive the tips' mean, variance, covariance and interspecies correlation for a branching Ornstein–Uhlenbeck (OU) process with jumps at speciation, alongside a way of quantitatively assessing the effect of both types of evolution.

Combining jumps with an OU process is attractive from a biological point of view. It is consistent with the original motivation behind punctuated equilibrium. At branching, dramatic events occur that drive species apart. But then stasis between these jumps does not mean that no change takes place, rather that during it "fluctuations of little or no accumulated consequence" occur (Gould and Eldredge [20]). The OU process fits into this idea because if the adaptation rate is large enough, then the process reaches stationarity very quickly and oscillates around the optimal state. This then can be interpreted as stasis between the jumps — the small fluctuations. Mayr [24] supports this sort of reasoning by hypothesizing that "The further removed in time a species from the original speciation event that originated it, the more its genotype will have become stabilized and the more it is likely to resist change."

In this work we build up on previous results (Bartoszek [5], Bartoszek and Sagitov [6]) and study in detail the asymptotic behaviour of the average of the tip values of a branching OU process with jumps at speciation points evolving on a pure birth tree. To the best of our knowledge the work here is the first to analytically consider the effect of jumps on a branching OU process in a phylogenetic context. It is possible that some of the results could be special subcases of general results on branching Markov processes (Ren et al. [29, 30]). However, these studies use a very

heavy functional analysis apparatus, which unlike the direct one here, is completely inaccessible to the applied reader. We can observe the same competition between the tree's speciation and OU's adaptation (drift) rates, resulting in a phase transition when the latter is half the former (the same as in the no jumps case Adamczak and Miłoś [1, 2], Bartoszek and Sagitov [6]). We show here that if large jumps are rare enough, then the contemporary sample mean will be asymptotically normally distributed. Otherwise the weak limit can be characterized as a "normal distribution with a random variance". Such probabilistic characterizations are important as there is a lack of tools for punctuated phylogenetic models. This is partially due to an uncertainty of what is estimable, especially whether the contribution of gradual and punctuated change may be disentangled (but Bokma [11] indicates that they should be distinguishable). Large sample size distributional approximations will allow for choosing seeds for numerical maximum likelihood procedures and sanity checks if the results of numerical procedures make sense. Most importantly the understanding of the limiting behaviour of evolutionary models with jumps will allow us to see the effects of these jumps, especially how much do they push the populations out of stationarity.

In Section 2 we introduce the considered probabilistic model. Then in Section 3 we present the main results. Section 4 is devoted to a series of technical convergence lemmata that characterize the speed of decay of the effect of jumps on the variance and covariance of tip species. Finally in Section 5 we calculate the first two moments of a scaled sample average, introduce a submartingale related to the model and put this together with the previous convergence lemmata to prove the Central Limit Theorems of this paper.

2 A model for punctuated stabilizing selection

2.1 Phenotype model

Stochastic differential equations (SDEs) are today the standard language to model continuous traits evolving on a phylogenetic tree. The general framework is that of a diffusion process

$$\mathrm{d}X(t) = \mu(t, X(t))\mathrm{d}t + \sigma_a \mathrm{d}B_t. \tag{1}$$

The trait follows Eq. (1) along each branch of the tree (with possibly branch specific parameters). At speciation times this process divides into two processes evolving independently from that point. The full generality of Eq. (1) is not implemented in contemporary phylogenetic comparative methods (PCMs). Currently they are focused on the OU processes

$$\mathrm{d}X(t) = -\alpha(X(t) - \theta(t))\mathrm{d}t + \sigma_a \mathrm{d}B_t, \tag{2}$$

where $\theta(t)$ can be piecewise linear, with different values assigned to different regimes (see e.g. Bartoszek et al. [7], Butler and King [12], Hansen [21]). There have been a few attempts to go beyond the diffusion framework into Lévy process, including Laplace motion, (Bartoszek [4, 5], Duchen et al. [14], Landis et al. [22]) and jumps at speciation points (Bartoszek [5], Bokma [9, 10]). We follow in the spirit of the latter and consider that just after a branching point, with a probability p, independently on each daughter lineage, a jump can occur. We assume that the jump random variable is normally distributed with mean 0 and variance $\sigma_c^2 < \infty$. In other words, if at time t there is a speciation event, then just after it, independently for each daughter lineage, the trait process $X(t^+)$ will be

$$X(t^+) = (1-Z)X(t^-) + Z(X(t^-) + Y), \tag{3}$$

where $X(t^{-/+})$ means the value of $X(t)$ respectively just before and after time t, Z is a binary random variable with probability p of being 1 (i.e. jump occurs) and $Y \sim \mathcal{N}(0, \sigma_c^2)$. The parameters p and σ_c^2 can, in particular, differ between speciation events.

2.2 The branching phenotype

In this work we consider a fundamental model of phylogenetic tree growth — the conditioned on number of tip species pure birth process. We first introduce some notation, illustrated in Fig. 1 (see also Bartoszek [5], Bartoszek and Sagitov [6], Sagitov and Bartoszek [31]). We consider a tree that has n tip species. Let $U^{(n)}$ be the tree height, $\tau^{(n)}$ the time from today (backwards) to the coalescent of a pair of randomly chosen tip species, $\tau_{ij}^{(n)}$ the time to coalescent of tips i, j, $\Upsilon^{(n)}$ the number of

speciation events on a random lineage, $\upsilon^{(n)}$ the number of common speciation events for a random pair of tips bar the splitting them event and $\upsilon_{ij}^{(n)}$ the number of common speciation events for a tips i, j bar the splitting them event. Furthermore let T_k be the time between speciation events k and $k+1$, p_k and $\sigma_{c,k}^2$ be respectively the probability and variance of the jump just after the k–th speciation event on each daughter lineage.

The above intuitive descriptions can be made more precise. We first introduce two separate labellings for the tip and internal nodes of the tree. Let the origin of the tree have label "0". Next we label from "1" to "$n-1$" the internal nodes of the tree in their temporal order of appearance. The root is "1", the node corresponding to the second speciation event is "2" and so on. We label the tips of the tree from "1" to "n" in an arbitrary fashion. This double usage of the numbers "1" to "$n-1$" does not generate any confusion as it will always be clear whether one refers to a tip or internal node.

Definition 2.1

$$N_{\text{Tip}}(t) = \{\text{set of tip nodes at time } t\}$$

Definition 2.2

$$U^{(n)} = \inf\{t \geq 0 : N_{\text{Tip}}(t) = n\}$$

Definition 2.3 *For $i \in N_{\text{Tip}}(U^{(n)})$,*

$$\Upsilon^{(i,n)} = \{\text{number of nodes on the path from the root (internal node 1, including it) to tip node } i\}$$

Definition 2.4 *For $i \in N_{\text{Tip}}(U^{(n)})$,*

$$\mathrm{I}^{(i,n)} = \left(\mathrm{I}_j^{(i,n)} : \mathrm{I}_j^{(i,n)} \text{is a node on the root to tip node } i \text{ path and } \mathrm{I}_j^{(i,n)} < \mathrm{I}_k^{(i,n)} \text{ for } 1 \leq j < k \leq \Upsilon^{(i,n)}\right)_{j=1}^{\Upsilon^{(i,n)}}$$

Definition 2.5 *For $i \in N_{\text{Tip}}(U^{(n)})$ and $r \in \{1, \ldots, n-1\}$, let $\mathbf{1}_r^{(i,n)}$ be a binary random variable such*

that

$$\mathbf{1}_r^{(i,n)} = 1 \text{ iff } r \in \mathrm{I}^{(i,n)}.$$

Definition 2.6 *For $i \in N_{\mathrm{Tip}}(U^{(n)})$ and $r \in \{1,\ldots,\Upsilon_i^{(n)}\}$, let $J_r^{(i,n)}$ be a binary random variable equalling 1 iff a jump took place just after the r–th speciation event in the sequence $\mathrm{I}^{(i,n)}$.*

Definition 2.7 *For $i \in N_{\mathrm{Tip}}(U^{(n)})$ and $r \in \{1,\ldots,n-1\}$, let $Z_r^{(i,n)}$ be a binary random variable equalling 1 iff $\mathbf{1}_r^{(i,n)} = 1$ and $J_k^{(i,n)} = 1$, where $\mathrm{I}_k^{(i,n)} = r$.*

Definition 2.8 *For $i, j \in N_{\mathrm{Tip}}(U^{(n)})$,*

$$\mathrm{I}^{(i,j,n)} = \mathrm{I}^{(i,n)} \cap \mathrm{I}^{(j,n)}$$

Definition 2.9 *For $i, j \in N_{\mathrm{Tip}}(U^{(n)})$,*

$$\upsilon^{(i,j,n)} = |\mathrm{I}^{(i,j,n)}| - 1$$

Remark 2.10 *We have the -1 in the above definition of $\upsilon^{(i,j,n)}$ as we are interested in counting the speciation events that could have a jump common to both lineages. As the jump occurs after a speciation event, the jumps connected to the coalescent node of tip nodes i and j cannot affect both of these tips.*

Definition 2.11 *For $i, j \in N_{\mathrm{Tip}}(U^{(n)})$ and $r \in \{1,\ldots,\max\{I^{(i,j,n)}\}-1\}$, let $\mathbf{1}_r^{(i,j,n)}$ be a binary random variable such that*

$$\mathbf{1}_r^{(i,j,n)} = 1 \text{ iff } r \in \mathrm{I}^{(i,j,n)}.$$

Definition 2.12 *For $i, j \in N_{\mathrm{Tip}}(U^{(n)})$,*

$$\tau^{(i,j,n)} = U^{(n)} - \inf\{t \geq 0 : N(t) = \max\left(\mathrm{I}^{(i,j,n)}\right)\}$$

Definition 2.13 *For $i,j \in N_{\text{Tip}}(U^{(n)})$ and $r \in \{1,\dots,\upsilon_{i,j}^{(n)}\}$, let $J_r^{(i,j,n)}$ be a binary random variable equalling 1 iff $J_r^{(i,n)} = 1$ and $J_r^{(j,n)} = 1$.*

Definition 2.14 *For $i,j \in N_{\text{Tip}}(U^{(n)})$ and $r \in \{1,\dots,n-1\}$, let $Z_r^{(i,j,n)}$ be a binary random variable equalling 1 iff $Z_r^{(i,n)} = 1$ and $Z_r^{(j,n)} = 1$.*

Definition 2.15 *Let R be uniformly distributed on $\{1,\dots,n\}$ and (R,K) be uniformly distributed on the set of ordered pairs drawn from $\{1,\dots,n\}$ (i.e. $\text{Prob}((R,K)=(r,k)) = \binom{n}{2}^{-1}$, for $1 \le r < k \le n$)*

$$\tau^{(n)} = \tau^{(R,K,n)},\ \ \Upsilon^{(n)} = \Upsilon^{(R,n)},\ \ \upsilon^{(n)} = \upsilon^{(R,K,n)},\ \ \mathrm{I}^{(n)} = \mathrm{I}^{(R,n)},\ \ \tilde{\mathrm{I}}^{(n)} = \mathrm{I}^{(R,K,n)},$$
$$\mathbf{1}_i = \mathbf{1}_i^{(R,n)},\ \ \tilde{\mathbf{1}}_i = \mathbf{1}_i^{(R,K,n)},\ \ J_i = J_i^{(R,n)},\ \ \tilde{J}_i = J_i^{(R,K,n)},\ \ Z_i = Z_i^{(R,no)},\ \ \tilde{Z}_i = Z_i^{(R,K,n)}.$$

Remark 2.16 *For the sequences $\mathrm{I}^{(n)}$, $\mathrm{I}^{(r,n)}$, $\mathrm{I}^{(R,n)}$, $\tilde{\mathrm{I}}^{(n)}$, $\mathrm{I}^{(r,k,n)}$, $\mathrm{I}^{(R,K,n)}$ the i–th element is naturally indicated as $\mathrm{I}_i^{(n)}$, $\mathrm{I}_i^{(r,n)}$, $\mathrm{I}_i^{(R,n)}$, $\tilde{\mathrm{I}}_i^{(n)}$, $\mathrm{I}_i^{(r,k,n)}$, $\mathrm{I}_i^{(R,K,n)}$ respectively.*

Remark 2.17 *We drop the n in the superscript for the random variables $\mathbf{1}_i$, $\tilde{\mathbf{1}}_i$, J_i, $\tilde{J}_i$, Z_i and $\tilde{Z}_i$ as their distribution does not depend on n. In fact, in principle, there is no need to distinguish between the version with and without the tilde. However, such a distinction will make it more clear to what one is referring to in the derivations of this work.*

The following simple, yet very powerful, Lemma comes from the uniformity of the choice of pair to coalesce at the i–th speciation event in the backward description of the Yule process. The full proof can be found in Bartoszek [5] on p. 45.

Lemma 2.18 *For all $i \in \{1,\dots,n-1\}$*

$$\mathrm{E}\left[\tilde{\mathbf{1}}_i\right] = \mathrm{E}\left[\mathbf{1}_i\right] = \text{Prob}(\mathbf{1}_i = 1) = \frac{2}{i+1}.$$

We called the model a conditioned one. By conditioning we consider stopping the tree growth just before the $n+1$ species occurs, or just before the n–th speciation event. Therefore, the tree's height $U^{(n)}$ is a random stopping time. The asymptotics considered in this work are when $n \to \infty$.

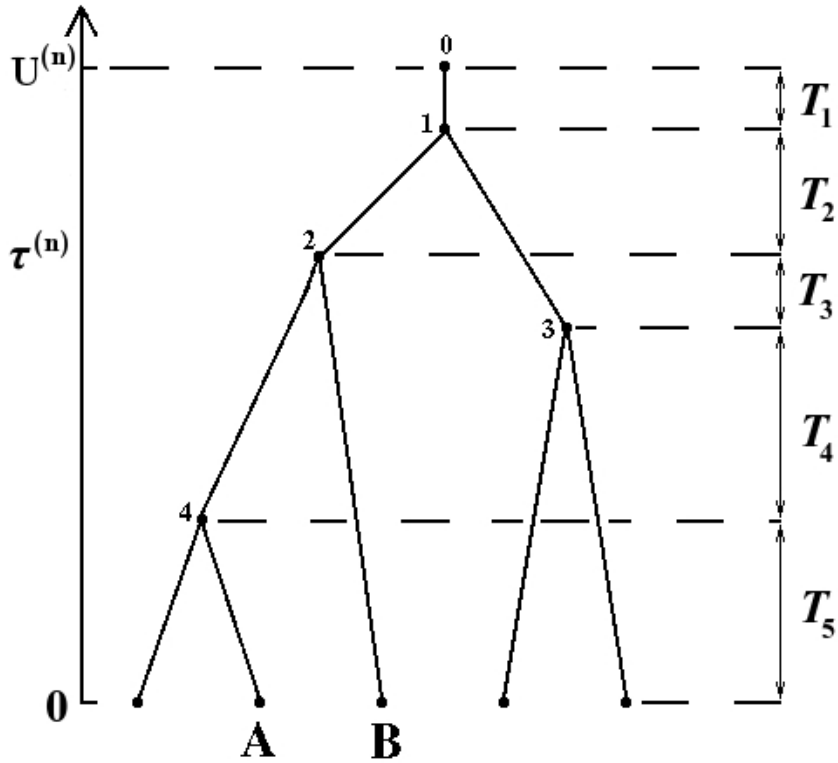

Figure 1: A pure–birth tree with the various time components marked on it. If we "randomly sample" node "A", then $\Upsilon^{(n)} = 3$ and the indexes of the speciation events on this random lineage are $\mathrm{I}_3^{(n)} = 4$ $\mathrm{I}_2^{(n)} = 2$ and $\mathrm{I}_1^{(n)} = 1$. Notice that $\mathrm{I}_1^{(n)} = 1$ always. The between speciation times on this lineage are T_1, T_2, $T_3 + T_4$ and T_5. If we "randomly sample" the pair of extant species "A" and "B", then $\upsilon^{(n)} = 1$ and the two nodes coalesced at time $\tau^{(n)}$. The random index of their joint speciation event is $\tilde{\mathrm{I}}_1 = 1$. See also Bartoszek [5]'s Fig. A.8. for a more detailed discussion on relevant notation. The internal node labellings 0–4 are marked on the tree.

The key model parameter describing the tree component is λ, the birth rate. At the start, the process starts with a single particle and then splits with rate λ. Its descendants behave in the same manner. Without loss generality we take $\lambda = 1$, as this is equivalent to rescaling time.

In the context of phylogenetic methods this branching process has been intensively studied (e.g. Bartoszek and Sagitov [6], Crawford and Suchard [13], Edwards [15], Gernhard [17, 18], Mulder and Crawford [27], Sagitov and Bartoszek [31], Steel and McKenzie [35]), hence here we will just describe its key property. The time between speciation events k and $k+1$ is exponential with

parameter k. This is immediate from the memoryless property of the process and the distribution of the minimum of exponential random variables. From this we obtain some important properties of the process. Let $H_n = 1 + 1/2 + \ldots + 1/n$ be the n–th harmonic number, $x > 0$ and then their expectations and Laplace transforms are (Bartoszek and Sagitov [6], Sagitov and Bartoszek [31])

$$\begin{aligned}
\mathrm{E}\left[U^{(n)}\right] &= H_n, \\
\mathrm{E}\left[e^{-xU^{(n)}}\right] &= b_{n,x}, \\
\mathrm{E}\left[\tau^{(n)}\right] &= \tfrac{n+1}{n-1}H_n - \tfrac{2}{n-1}, \\
\mathrm{E}\left[e^{-x\tau^{(n)}}\right] &= \begin{cases} \frac{2-(n+1)(x+1)b_{n,x}}{(n-1)(x-1)} & x \neq 1, \\ \frac{2}{n-1}\left(H_n - 1\right) - \frac{1}{n+1} & x = 1, \end{cases}
\end{aligned}$$

where

$$b_{n,x} = \frac{1}{x+1}\cdots\frac{n}{n+x} = \frac{\Gamma(n+1)\Gamma(x+1)}{\Gamma(n+x+1)} \sim \Gamma(x+1)n^{-x},$$

$\Gamma(\cdot)$ being the gamma function.

Now let $\mathscr{Y}_n$ be the σ–algebra that contains information on the Yule tree and jump pattern. Bartoszek [5] previously studied the branching OU with jumps model and it was shown (but, therein for constant p_k and $\sigma^2_{c,k}$ and therefore there was no need to condition on the jump pattern) that, conditional on the tree height and number of tip species the mean and variance of an arbitrary tip species, $X_r^{(n)}$, are

$$\begin{aligned}
\mathrm{E}\left[X^{(n)_r}|\mathscr{Y}_n\right] &= \theta + e^{-\alpha U^{(n)}}(X_0 - \theta) \\
\mathrm{Var}\left[X^{(n)_r}|\mathscr{Y}_n\right] &= \tfrac{\sigma_a^2}{2\alpha}(1 - e^{-2\alpha U^{(n)}}) + \sum_{i=1}^{\Upsilon^{(r,n)}} \sigma^2_{c,\mathrm{I}_i^{(r,n)}} J_i^{(r,n)} e^{-2\alpha(T_n + \ldots + T_{\mathrm{I}_i^{(r,n)}+1})}.
\end{aligned} \tag{4}$$

A key difference that the phylogeny brings in is that the tip measurements are correlated through the tree structure. One can easily show that conditional on $\mathscr{Y}_n$, the covariance between an arbitrary pair of extant traits, $X_r^{(n)}$ and $X_k^{(n)}$ is

$$\mathrm{Cov}\left[X_r^{(n)}, X_k^{(n)} | \mathscr{Y}_n\right] = \frac{\sigma_a^2}{2\alpha}(e^{-2\alpha\tau^{(r,k,n)}} - e^{-2\alpha U^{(n)}}) + \sum_{i=1}^{\upsilon^{(r,k,n)}} \sigma^2_{c,\mathrm{I}_i^{(r,k,n)}} J_i^{(r,k,n)} e^{-2\alpha(\tau^{(r,k,n)}+\ldots+T_{\mathrm{I}_i^{(r,k,n)}+1})}. \tag{5}$$

We will call, the considered model the Yule–Ornstein–Uhlenbeck with jumps (YOUj) process.

Remark 2.19 *Keeping the parameter θ constant on the tree is not as simplifying as it might seem. Varying θ models have been considered since the introduction of the OU process to phylogenetic methods (Hansen [21]). However, it can very often happen that the θ parameter is constant over whole clades, as these species share a common optimum. Therefore, understanding the model's behaviour with a constant θ is a crucial first step. Furthermore if constant θ clades are far enough apart one could think of them as independent samples and attempt to construct a test (based on normality of the species' averages) if jumps have a significant effect (compare Thms. 3.1 and 3.5).*

2.3 Martingale formulation

Our main aim is to study the asymptotic behaviour of the sample average and it actually turns out to be easier to work with scaled trait values, $Y^{(n)} = (X^{(n)} - \theta)/\sqrt{\sigma_a^2/2\alpha}$. Denoting $\delta = (X_0 - \theta)/\sqrt{\sigma_a^2/2\alpha}$ we have

$$\mathrm{E}\left[Y^{(n)}\right] \quad = \quad \delta b_{n,\alpha}. \tag{6}$$

The initial condition of course will be $Y_0 = \delta_0$. Just as was done by Bartoszek and Sagitov [6] we may construct a martingale related to the average

$$\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i^{(n)}.$$

Then (cf. Lemma 10 of Bartoszek and Sagitov [6]), we define

$$H_n := (n+1)e^{(\alpha-1)U^{(n)}}\overline{Y}_n, \;\; n \geq 0.$$

This is a martingale with respect to $\mathscr{F}_n$, the σ–algebra containing information on the Yule n–tree and the phenotype's evolution.

3 Asymptotic regimes — main results

Branching Ornstein–Uhlenbeck models commonly have three asymptotic regimes (Adamczak and Miłoś [1, 2], Ané et al. [3], Bartoszek [5], Bartoszek and Sagitov [6], Ren et al. [29, 30]). The dependency between the adaptation rate α and branching rate $\lambda = 1$ governs in which regime the process is. If $\alpha > 1/2$, then the contemporary sample is similar to an i.i.d. sample, in the critical case, $\alpha = 1/2$, we can, after appropriate rescaling, still recover the "near" i.i.d. behaviour and if $0 < \alpha < 1/2$, then the process has "long memory" ("local correlations dominate over the OU's ergodic properties", Adamczak and Miłoś [1, 2]). In the OU process with jumps setup the same asymptotic regimes can be observed, even though Adamczak and Miłoś [1, 2], Ren et al. [29, 30]) assume that the tree is observed at a given time point, t, with n_t being random. In what follows here, the constant C may change between (in)equalities. It may in particular depend on α. We illustrate the below Theorems in Fig. 2.

Theorem 3.1 *Assume that the jump probabilities and jump variances are constant equalling p and $\sigma_c^2 < \infty$ respectively. Let $\overline{Y}_n = (\overline{X}_n - \theta)/\sqrt{\sigma_a^2/2\alpha}$ be the normalized sample mean of the YOUj process with $\overline{Y}_0 = \delta_0$. The process $\overline{Y}_n$ has the following, depending on α, asymptotic with n behaviour.*

(I) If $0.5 < \alpha$, then the conditional variance of the scaled sample mean $\sigma_n^2 := n\,\mathrm{Var}\left[\overline{Y}_n|\mathscr{Y}_n\right]$ converges in $\mathbb{P}$ to a random variable σ_∞^2 with mean

$$\mathrm{E}\left[\sigma_\infty^2\right] \quad = \quad 1 + \tfrac{2p\sigma_c^2}{\sigma_a^2} + \tfrac{4p\sigma_c^2}{2\alpha(2\alpha-1)\sigma_a^2}.$$

The scaled sample mean, $\sqrt{(n)}\,\overline{Y}_n$ converges weakly to random variable whose characteristic function can be expressed in terms of the Laplace transform of σ_∞^2

$$\phi_{\sqrt{(n)}\,\overline{Y}_n}(x) = \mathscr{L}(\sigma_\infty^2)(x^2/2).$$

(II) If $0.5 = \alpha$*, then the conditional variance of the scaled sample mean* $\sigma_n^2 := n \ln^{-1} n \operatorname{Var}\left[\overline{Y}_n | \mathscr{Y}_n\right]$ *converges in* $\mathbb{P}$ *to a random variable* σ_∞^2 *with mean*

$$\mathrm{E}\left[\sigma_\infty^2\right] \quad = \quad 2 + 4p\sigma_c^2(\sigma_a^2/(2\alpha))^{-1}.$$

The scaled sample mean, $\sqrt{(n/\ln n)}\,\overline{Y}_n$ *converges weakly to random variable whose characteristic function can be expressed in terms of the Laplace transform of* σ_∞^2

$$\phi_{\sqrt{(n/\ln n)}\,\overline{Y}_n}(x) = \mathscr{L}(\sigma_\infty^2)(x^2/2).$$

(III) If $0 < \alpha < 0.5$*, then* $n^\alpha \overline{Y}_n$ *converges almost surely and in* L^2 *to a random variable* $Y_{\alpha,\delta}$ *with first two moments*

$$\begin{aligned} \mathrm{E}\left[Y_{\alpha,\delta}\right] &= \delta\Gamma(1+\alpha), \\ \mathrm{E}\left[Y_{\alpha,\delta}^2\right] &= \left(2p\sigma_c^2(\sigma_a^2/(2\alpha))^{-1}\tfrac{1+2\alpha}{1-2\alpha} - (1-\delta^2)\right)\Gamma(1+2\alpha). \end{aligned}$$

Remark 3.2 *For the a.s. and* L^2 *convergence to hold in Part (III), it suffices that the sequence of jump variances is bounded. Of course, the first two moments will differ if the jump variance is not constant.*

Definition 3.3 *A subset* $E \subset \mathscr{N}$ *of positive integers is said to have density* 0 *(e.g. Petersen [28]) if*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_E(k) = 0,$$

where $\chi_E(\cdot)$ *is the indicator function of the set* E*.*

Definition 3.4 *A sequence* a_n *converges to* 0 *with density* 1 *if there exists a subset* $E \subset \mathscr{N}$ *of density* 0 *such that*

$$\lim_{n\to\infty, n\notin E} a_n = 0.$$

Theorem 3.5 *If $\sigma_{c,n}^4 p_n$ is bounded and goes to 0 with density 1, then depending on α the process $\overline{Y}_n$ has the following asymptotic with n behaviour.*

(I) If $0.5 < \alpha$, then $\sqrt{(n)}\ \overline{Y}_n$ is asymptotically normally distributed with mean 0 and variance $(2\alpha+1)/(2\alpha-1)$.

(II) If $0.5 = \alpha$, then $\sqrt{(n/\ln n)}\ \overline{Y}_n$ is asymptotically normally distributed with mean 0 and variance 2.

Remark 3.6 *The assumption $\sigma_{c,n}^4 p_n \to 0$ with density 1 is an essential one for the limit to be a normal distribution, when $\alpha \geq 0.5$. This is visible from the proof of Lemma 4.5. In fact, this is the key difference that the jumps bring in — if they occur too often (or with too large magnitude), then they will disrupt the weak convergence.*

One natural way is to keep $\sigma_{c,n}^2$ constant and allow $p_n \to 0$, the chance of jumping becomes smaller relative to the number of species. Alternatively $\sigma_{c,n}^2 \to 0$, which could mean that with more and more species — smaller and smaller jumps occur at speciation. Actually, this could be biologically more realistic — as there are more and more species, then there is more and more competition and smaller and smaller differences in phenotype drive the species apart. Specialization occurs and tinier and tinier niches are filled.

4 Key convergence lemmata

We will now prove a series of technical lemmata describing the asymptotics of driving components of the considered YOUj process.

Lemma 4.1 *(Lemma 11 of Bartoszek and Sagitov [6])*

$$\mathrm{Var}\left[\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right]\right] = \begin{cases} O(n^{-4\alpha}) & 0 < \alpha < 0.75, \\ O(n^{-3}\ln n) & \alpha = 0.75, \\ O(n^{-3}) & 0.75 < \alpha. \end{cases} \tag{7}$$

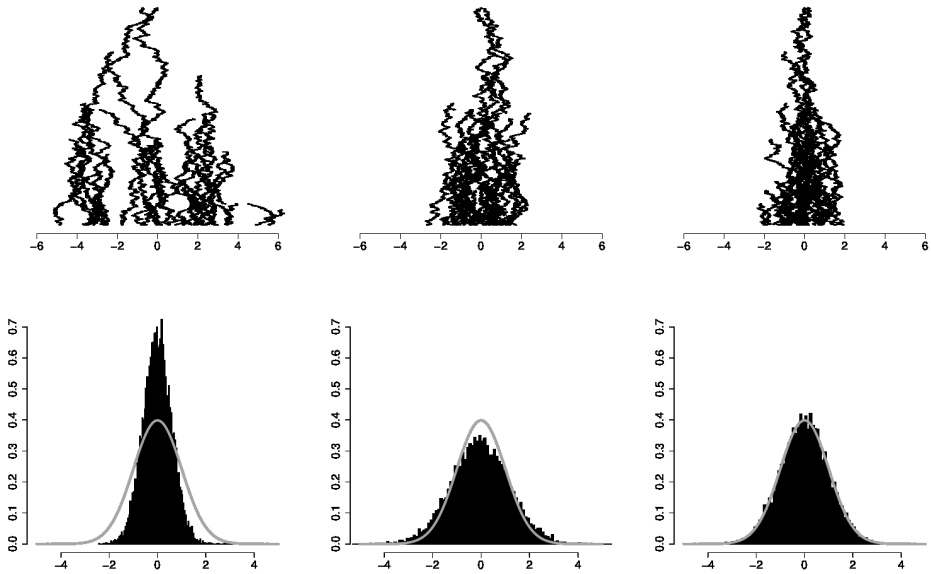

Figure 2: Left: $\alpha = 0.25$ centre: $\alpha = 0.5$ and right: $\alpha = 1$. Top row: examples of simulated YOUj process trajectories, bottom row: histograms of sample averages, left: scaled by $1/\Gamma(3/2)$, centre: scaled by $\sqrt{n \ln^{-1} n/2}$, right: scaled by $\sqrt{n/3}$. In all three cases, $p = 0.5$, $\sigma_c^2 = 1$, $\sigma_a^2 = 1$, $X_0 = \theta = 0$. The phylogenetic trees are pure birth trees with $\lambda = 1$ conditioned on number of tips, $n = 30$ for the trajectory plots and $n = 200$ for the histograms. The histograms are based on 10000 simulated trees. The sample mean and variances of the scaled data in the histograms are left: $(-0.006, 0.360)$, centre: $(-0.033, 1.481)$ and right: $(0.004, 1.008)$. The gray curve painted on the histograms is the standard normal distribution. The phylogenies are simulated by the TreeSim R package (Stadler [33, 34]) and simulations of phenotypic evolution and trajectory plots are done by newly implemented functions of the, available on CRAN, mvSLOUCH R package (Bartoszek et al. [7]).

PROOF For a given realization of the Yule n-tree we denote by $\tau_1^{(n)}$ and $\tau_2^{(n)}$ two independent versions of $\tau^{(n)}$ corresponding to two independent choices of pairs of tips out of n available. We have,

$$\mathrm{E}\left[\left(\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right]\right)^2\right] = \mathrm{E}\left[\mathrm{E}\left[e^{-2\alpha(\tau_1^{(n)}+\tau_2^{(n)})}|\mathscr{Y}_n\right]\right] = \mathrm{E}\left[e^{-2\alpha(\tau_1^{(n)}+\tau_2^{(n)})}\right].$$

Let $\pi_{n,k}$ be the probability that two randomly chosen tips coalesced at the k–th speciation event. We know that (cf. Bartoszek and Sagitov [6], Steel and McKenzie [35])

$$\pi_{n,k} = 2\frac{n+1}{n-1}\frac{1}{(k+1)(k+2)}.$$

Writing

$$f(a,k,n) := \frac{k+1}{a+k+1}\cdots\frac{n}{a+n} = \frac{\Gamma(n+1)\Gamma(a+k+1)}{\Gamma(k+1)\Gamma(a+n+1)}$$

and as the times between speciation events are independent and exponentially distributed we obtain

$$\mathrm{E}\left[\left(\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right]\right)^2\right] = \sum_{k=1}^{n-1} f_{4\alpha}(k,n)\pi_{n,k}^2 + 2\sum_{k_1<k_2} f_{2\alpha}(k_1,k_2)f_{4\alpha}(k_2,n)\pi_{n,k_1}\pi_{n,k_2}.$$

On the other hand,

$$\left(\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}\right]\right)^2 = \left(\sum_{k_1} f_{2\alpha}(k_1,n)\pi_{n,k_1}\right)\left(\sum_{k_2} f_{2\alpha}(k_2,n)\pi_{n,k_2}\right).$$

Taking the difference between the last two expressions we find

$$\begin{aligned}\mathrm{Var}\left[\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right]\right] = \quad & \sum_k \left(f_{4\alpha}(k,n) - f_{2\alpha}(k,n)^2\right)\pi_{n,k}^2 \\ & +2\sum_{k_1=1}^{n-1}\sum_{k_2=k_1+1}^{n-1} f_{2\alpha}(k_1,k_2)\left(f_{4\alpha}(k_2,n) - f_{2\alpha}(k_2,n)^2\right)\pi_{n,k_1}\pi_{n,k_2}.\end{aligned}$$

Using the simple equality

$$a_1\cdots a_n - b_1\cdots b_n = \sum_{i=1}^{n} b_1\cdots b_{i-1}(a_i - b_i)a_{i+1}\cdots a_n$$

we see that it suffices to study the asymptotics of,

$$\sum_{k=1}^{n-1} A_{n,k}\pi_{n,k}^2 \text{ and } \sum_{k_1=1}^{n-1}\sum_{k_2=k_1+1}^{n-1} f_{2\alpha}(k_1,k_2)A_{n,k_2}\pi_{n,k_1}\pi_{n,k_2},$$

where

$$A_{n,k} := \sum_{j=k+1}^{n} f_{2\alpha}(k,j)^2\Big(\frac{4\alpha^2}{j(j+4\alpha)}\Big)f_{4\alpha}(j,n).$$

To consider these two asymptotic relations we observe that for large n

$$A_{n,k} \lesssim 4\alpha^2\frac{b_{n,4\alpha}}{b_{k,2\alpha}^2}\sum_{j=k+1}^{n}\frac{b_{j,2\alpha}^2}{b_{j,4\alpha}}\frac{1}{j(4\alpha+j)} \lesssim C\frac{b_{n,4\alpha}}{b_{k,2\alpha}^2}\sum_{j=k+1}^{n} j^{-2} \lesssim C\frac{b_{n,4\alpha}}{b_{k,2\alpha}^2}k^{-1}.$$

Now since $\pi_{n,k} = \frac{2(n+1)}{(n-1)(k+2)(k+1)}$, it follows

$$\sum_{k=1}^{n-1} A_{n,k}\pi_{n,k}^2 \lesssim Cb_{n,4\alpha}\sum_{k=1}^{n-1}\frac{1}{k^5 b_{k,2\alpha}^2} \lesssim Cn^{-4\alpha}\sum_{k=1}^{n} k^{4\alpha-5} \lesssim C\begin{cases} n^{-4\alpha} & 0<\alpha<1 \\ n^{-4}\ln n & \alpha=1 \\ n^{-4} & 1<\alpha \end{cases}$$

and

$$\begin{aligned}
\sum_{k_1=1}^{n-1}\sum_{k_2=k_1+1}^{n-1} f_{2\alpha}(k_1,k_2)A_{n,k_2}\pi_{n,k_1}\pi_{n,k_2} &\lesssim Cb_{n,4\alpha}\sum_{k_1=1}^{n-1}\sum_{k_2=k_1+1}^{n-1}\tfrac{1}{b_{k_1,2\alpha}b_{k_2,2\alpha}}\tfrac{1}{k_1^2k_2^3} \\
&\lesssim n^{-4\alpha}\sum_{k_1=1}^{n-1} k_1^{2\alpha-2}\sum_{k_2=k_1+1}^{n-1} k_2^{2\alpha-3} \lesssim C\begin{cases} n^{-4\alpha}\sum_{k_1=1}^{n-1} k_1^{4\alpha-4} & 0<\alpha<1 \\ n^{-4}\sum_{k_2=2}^{n} k_2^{-1}\sum_{k_1=1}^{k_2} 1 & \alpha=1 \\ n^{-4\alpha}\sum_{k_2=2}^{n} k_2^{4\alpha-4} & 1<\alpha \end{cases} \\
&\lesssim C\begin{cases} n^{-4\alpha} & 0<\alpha<0.75 \\ n^{-3}\ln n & \alpha=0.75 \\ n^{-3} & 0.75<\alpha<1 \\ n^{-4}\sum_{k_2=2}^{n} 1 & \alpha=1 \\ n^{-3} & 1<\alpha \end{cases} \lesssim C\begin{cases} n^{-4\alpha} & 0<\alpha<0.75 \\ n^{-3}\ln n & \alpha=0.75 \\ n^{-3} & 0.75<\alpha<1 \\ n^{-3} & \alpha=1 \\ n^{-3} & 1<\alpha. \end{cases}
\end{aligned}$$

Summarizing

$$\sum_{k_1=1}^{n-1}\sum_{k_2=k_1+1}^{n-1} f_{2\alpha}(k_1,k_2)A_{n,k_2}\pi_{n,k_1}\pi_{n,k_2} \lesssim C\begin{cases} n^{-4\alpha} & 0<\alpha<0.75 \\ n^{-3}\ln n & \alpha=0.75 \\ n^{-3} & 0.75<\alpha<1. \end{cases}$$

Notice that obviously $\mathrm{Var}\left[\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right]\right] = \mathrm{Var}\left[\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right]\right]$.

□

Remark 4.2 *The above Lemma 4.1 is a corrected version of Bartoszek and Sagitov [6]'s Lemma* 11. *There it is wrongly stated that* $\mathrm{Var}\left[\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right]\right] = O(n^{-3})$ *for all* $\alpha>0$. *From the above we can see that this holds only for* $\alpha>3/4$. *This does not however change Bartoszek and Sagitov [6]'s main results. If one inspects the proof of Theorem* 1 *therein, then one can see that for* $\alpha>0.5$ *it is required that* $\mathrm{Var}\left[\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right]\right] = O(n^{-(2+\varepsilon)})$, *where* $\varepsilon>0$. *This by Lemma 4.1 holds. Bartoszek and Sagitov [6]'s Thm.* 2 *does not depend on the rate of convergence, only that* $\mathrm{Var}\left[\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right]\right]\to 0$ *with n. This remains true, just with a different rate.*

Lemma 4.3 *For random variables* $(\Upsilon^{(n)},\mathrm{I}^{(n)},(J_i)_{i=1}^{\Upsilon^{(n)}})$ *derived from the same random lineage and a fixed jump probability p we have*

$$\mathrm{Var}\left[\mathrm{E}\left[\sum_{i=1}^{\Upsilon^{(n)}} J_i e^{-2\alpha(T_n+\ldots+T_{\mathrm{I}_i^{(n)}+1})}|\mathscr{Y}_n\right]\right] \lesssim p\begin{cases} n^{-4\alpha} & 0<\alpha<0.25 \\ n^{-1}\ln n & \alpha=0.25 \\ n^{-1} & 0.25<\alpha. \end{cases} \tag{8}$$

PROOF We introduce the random variables

$$\Psi^{*(n)} := \sum_{i=1}^{\Upsilon^{(n)}} J_i e^{-2\alpha(T_n+\ldots+T_{\mathrm{I}_i^{(n)}+1})}$$

and

$$\phi_i^* := Z_i e^{-2\alpha(T_n+\ldots+T_{i+1})}\,\mathrm{E}\left[\mathbf{1}_i|\mathscr{Y}_n\right].$$

Obviously

$$\mathrm{E}\left[\Psi^{*^{(n)}}|\mathscr{Y}_n\right] = \sum_{i=1}^{n-1} \phi_i^*.$$

Immediately (for $i < j$)

$$\begin{aligned}
\mathrm{E}\left[\phi_i^*\right] &= \tfrac{2p}{i+1}\tfrac{b_{n,2\alpha}}{b_{i,2\alpha}}, \\
\mathrm{E}\left[\phi_i^*\phi_j^*\right] &= \tfrac{4p^2}{(i+1)(j+1)}\tfrac{b_{n,4\alpha}}{b_{j,4\alpha}}\tfrac{b_{j,2\alpha}}{b_{i,2\alpha}}, \\
\mathrm{E}\left[\phi_i^{*2}\right] &= p\tfrac{b_{n,4\alpha}}{b_{i,4\alpha}}\mathrm{E}\left[\left(\mathrm{E}\left[\mathbf{1}_i|\mathscr{Y}_n\right]\right)^2\right].
\end{aligned}$$

The term $\mathrm{E}\left[\left(\mathrm{E}\left[\mathbf{1}_i|\mathscr{Y}_n\right]\right)^2\right]$ can be [see Lemma 11 in 6] expressed as $\mathrm{E}\left[\mathbf{1}_i^{(1)}\mathbf{1}_i^{(2)}\right]$ where $\mathbf{1}_i^{(1)}$ and $\mathbf{1}_i^{(2)}$ are two independent copies of $\mathbf{1}_i$, i.e. we sample two lineages and ask if the i–th speciation event is on both of them. This will occur if these lineages coalesced at a speciation event $k \geq i$. Therefore

$$\begin{aligned}
&\mathrm{E}\left[\mathbf{1}_i^{(1)}\mathbf{1}_i^{(2)}\right] = \tfrac{2}{i+1}\sum_{k=i+1}^{n-1}\pi_{k,n} + \pi_{i,n} = \tfrac{n+1}{n-1}\tfrac{2}{i+1}\left(\sum_{k=i+1}^{n-1}\tfrac{2}{(k+1)(k+2)} + \tfrac{1}{i+2}\right) \\
&= \tfrac{n+1}{n-1}\tfrac{2}{i+1}\left(\tfrac{2}{i+2} - \tfrac{2}{n+1} + \tfrac{1}{i+2}\right) = \tfrac{n+1}{n-1}\tfrac{6}{(i+1)(i+2)} - \tfrac{2}{n-1}\tfrac{2}{i+1}.
\end{aligned}$$

Together with the above

$$\mathrm{E}\left[\phi_i^{*2}\right] = p\frac{b_{n,4\alpha}}{b_{i,4\alpha}}\left(\frac{n+1}{n-1}\frac{6}{(i+1)(i+2)} - \frac{1}{n-1}\frac{4}{i+1}\right).$$

Now

$$
\begin{aligned}
\operatorname{Var}\left[\sum_{i=1}^{n-1}\phi_i^*\right] &= \sum_{i=1}^{n-1}\left(\mathrm{E}\left[\phi_i^{*2}\right]-(\mathrm{E}[\phi_i^*])^2\right)+2\sum_{i<j}^{n-1}\left(\mathrm{E}\left[\phi_i^*\phi_j^*\right]-\mathrm{E}[\phi_i^*]\,\mathrm{E}\left[\phi_j^*\right]\right) \qquad (9)\\
&= \sum_{i=1}^{n-1}\left(p\frac{b_{n,4\alpha}}{b_{i,4\alpha}}\left(\frac{n+1}{n-1}\frac{6}{(i+1)(i+2)}-\frac{1}{n-1}\frac{4}{i+1}\right)-\frac{4p^2}{(i+1)^2}\left(\frac{b_{n,2\alpha}}{b_{i,2\alpha}}\right)^2\right)\\
&\quad +2\sum_{i<j}^{n-1}\left(\frac{4p^2}{(i+1)(j+1)}\frac{b_{n,4\alpha}}{b_{j,4\alpha}}\frac{b_{j,2\alpha}}{b_{i,2\alpha}}-\frac{4p^2}{(i+1)(j+1)}\frac{b_{n,2\alpha}}{b_{i,2\alpha}}\frac{b_{n,2\alpha}}{b_{j,2\alpha}}\right)\\
&\lesssim 2p\sum_{i=1}^{n-1}\frac{1}{(i+1)^2}\left(3\frac{b_{n,4\alpha}}{b_{i,4\alpha}}-2p\left(\frac{b_{n,2\alpha}}{b_{i,2\alpha}}\right)^2\right) \quad \mathrm{I}\\
&\quad +4p(n-1)^{-1}\sum_{i=1}^{n-1}\frac{b_{n,4\alpha}}{b_{i,4\alpha}}\left(\frac{3}{(i+1)^2}-\frac{1}{i+1}\right) \quad \mathrm{II}\\
&\quad +8p^2\sum_{i<j}^{n-1}\left(\frac{1}{(i+1)(j+1)}\frac{b_{j,2\alpha}}{b_{i,2\alpha}}\left(\frac{b_{n,4\alpha}}{b_{j,4\alpha}}-\left(\frac{b_{n,2\alpha}}{b_{j,2\alpha}}\right)^2\right)\right). \quad \mathrm{III}
\end{aligned}
$$

We use the equality [cf. Lemma 11 in 6]

$$
a_1\cdots a_m-b_1\cdots b_m=\sum_{i=1}^{m}b_1\cdots b_{i-1}(a_i-b_i)a_{i+1}\cdots a_m
$$

and consider the three parts in turn.

I

$$
\begin{aligned}
&\sum_{i=1}^{n-1}\tfrac{1}{(i+1)^2}\left(3\tfrac{b_{n,4\alpha}}{b_{i,4\alpha}}-2p\left(\tfrac{b_{n,2\alpha}}{b_{i,2\alpha}}\right)^2\right)\\
&=\sum_{i=1}^{n-1}\tfrac{1}{(i+1)^2}\left(\left(\tfrac{b_{n-1,2\alpha}}{b_{i,2\alpha}}\right)^2\left(\tfrac{3n}{n+4\alpha}-\tfrac{2pn^2}{(n+2\alpha)^2}\right)+3\sum_{k=i+1}^{n-1}\left(\tfrac{b_{k-1,2\alpha}}{b_{i,2\alpha}}\right)^2\left(\tfrac{k}{k+4\alpha}-\tfrac{k^2}{(k+2\alpha)^2}\right)\tfrac{b_{n,4\alpha}}{b_{k,4\alpha}}\right)
\end{aligned}
$$

$$
\begin{aligned}
&=\sum_{i=1}^{n-1}\tfrac{1}{(i+1)^2}\left(\left(\tfrac{b_{n-1,2\alpha}}{b_{i,2\alpha}}\right)^2\tfrac{n^2}{(n+2\alpha)^2}\tfrac{(3-2p)n+(3-2p)4\alpha+n^{-1}12\alpha^2}{n+4\alpha}+3\sum_{k=i+1}^{n-1}\left(\tfrac{b_{k-1,2\alpha}}{b_{i,2\alpha}}\right)^2\tfrac{k^2}{(k+2\alpha)^2}\tfrac{4\alpha^2}{k(k+4\alpha)}\tfrac{b_{n,4\alpha}}{b_{k,4\alpha}}\right)\\
&\lesssim(3-2p)n^{-4\alpha}\sum_{i=1}^{n}i^{4\alpha-2}+12\alpha^2n^{-4\alpha}\sum_{i=1}^{n}i^{4\alpha-3}
\end{aligned}
$$

$$\sim \begin{cases} Cn^{-4\alpha} & 0<\alpha<0.25 \\ Cn^{-1}\ln n & \alpha=0.25 \\ (3-2p)(4\alpha-1)^{-1}n^{-1} & 0.25<\alpha. \end{cases}$$

II

$$n^{-1}\sum_{i=1}^{n-1}\frac{b_{n,4\alpha}}{b_{i,4\alpha}}\left(\frac{6}{(i+1)^2}-\frac{1}{i+1}\right)\lesssim 6n^{-4\alpha-1}\sum_{i=1}^{n}i^{4\alpha-2}-n^{-4\alpha-1}\sum_{i=1}^{n}i^{4\alpha-1}\sim\begin{cases} -(4\alpha)^{-1}n^{-1} & 0<\alpha<0.25 \\ -n^{-1} & \alpha=0.25 \\ -(4\alpha)^{-1}n^{-1} & 0.25<\alpha. \end{cases}$$

III

$$\sum_{i<j}^{n-1}\left(\frac{1}{(i+1)(j+1)}\frac{b_{j,2\alpha}}{b_{i,2\alpha}}\left(\frac{b_{n,4\alpha}}{b_{j,4\alpha}}-\left(\frac{b_{n,2\alpha}}{b_{j,2\alpha}}\right)^2\right)\right)=b_{n,4\alpha}\sum_{i<j}^{n-1}\frac{1}{(i+1)(j+1)}\frac{1}{b_{i,2\alpha}b_{j,2\alpha}}\sum_{k=j+1}^{n}\frac{b_{k,2\alpha}^2}{b_{k,4\alpha}}\frac{4\alpha^2}{k(k+4\alpha)}$$

$$\lesssim Cn^{-4\alpha}\sum_{i<j}^{n}i^{-1+2\alpha}j^{-2+2\alpha}\sim\begin{cases} Cn^{-4\alpha} & 0<\alpha<0.25 \\ Cn^{-1}\ln n & \alpha=0.25 \\ (1-2\alpha)^{-1}(4\alpha-1)^{-1}n^{-1} & 0.25<\alpha<0.5 \\ n^{-1} & \alpha=0.5 \\ (2\alpha-1)^{-1}(4\alpha-1)^{-1}n^{-1} & 0.5<\alpha. \end{cases}$$

Putting these together we obtain

$$\mathrm{Var}\left[\sum_{i=1}^{n-1}\phi_i^*\right] \quad\lesssim\quad pC\begin{cases} n^{-4\alpha} & 0<\alpha<0.25 \\ n^{-1}\ln n & \alpha=0.25 \\ n^{-1} & 0.5<\alpha. \end{cases}$$

On the other hand the variance is bounded from below by III. Its asymptotic behaviour is tight as the calculations there are accurate up to a constant (independent of p). This is further illustrated by

graphs in Fig. 3.

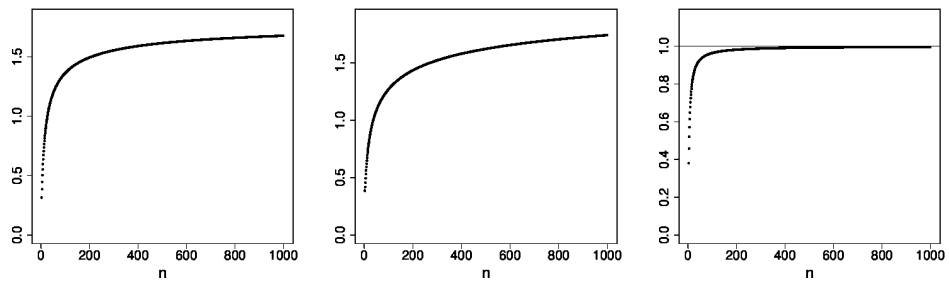

Figure 3: Numerical evaluation of scaled Eq. (9) for different values of α. The scaling for left: $\alpha = 0.1$ equals $n^{-4\alpha}$, centre: $\alpha = 0.25$ equals $n^{-1}\log n$ and right $\alpha = 1$ equals $(2p(3-2p)/(4\alpha-1)-4p/(4\alpha)+32p^2\alpha^2(1/(8\alpha^2)+1/(2\alpha(2\alpha-1))-1/(4\alpha^2)^{-1}-1/((2\alpha-1)(4\alpha-1))))n^{-1}$. In all cases, $p = 0.5$.

□

Corollary 4.4 *Let p_n and $\sigma^2_{c,n}$ be respectively the jump probability and variance at the n–th speciation event, such that the sequence $\sigma^4_{c,n}p_n$ is bounded. We have*

$$\begin{aligned} n\ln^{-1} n\,\mathrm{Var}\left[\sum_{i=1}^{n-1}\sigma^2_{c,i}\phi^*_i\right] &\to 0 \quad \text{for } \alpha = 0.25, \\ n\,\mathrm{Var}\left[\sum_{i=1}^{n-1}\sigma^2_{c,i}\phi^*_i\right] &\to 0 \quad \text{for } 0.25 < \alpha. \end{aligned}$$

iff $\sigma^4_{c,n}p_n \to 0$ with density 1.

PROOF We consider the case, $\alpha > 0.25$. Notice that in the proof of Lemma 4.3 $\mathrm{Var}\left[\sum_{i=1}^{n-1}\phi^*_i\right] \lesssim pn^{-4\alpha}\sum_{i=1}^{n-1} i^{4\alpha-2}$. If the jump probability and variance are not constant, but as in the Corollary, then $\mathrm{Var}\left[\sum_{i=1}^{n-1}\sigma^2_{c,i}\phi^*_i\right] \lesssim n^{-4\alpha}\sum_{i=1}^{n-1} p_i\sigma^4_{c,i}i^{4\alpha-2}$.

The Corollary is a consequence of a more general ergodic property that if $u > 0$ and the sequence $a_i \to 0$ with density 1, then

$$n^{-u}\sum_{i=1}^{n-1} a_i i^{u-1} \to 0.$$

For $1 \leq u$ the above is a direct consequence of Petersen [28]'s Lemma 6.2 (p. 65). For $0 < u < 1$ take a subsequence $N_0(n) \leq n^u$, such that $N_0(n) \to \infty$, $N_0(n)n^{-u} \to 0$ and $a_{N_0(n)} \searrow 0$. Then

$$n^{-u} \sum_{i=1}^{n-1} a_i i^{u-1} \leq N_0(n) n^{-u} + a_{N_0(n)} n^{-u} \sum_{i=N_0(n)+1}^{n-1} i^{u-1} \to 0.$$

On the other hand if a_i does not go to 0 with density 1, then $\limsup_n n^{-u} \sum_{i=1}^{n-1} a_i i^{u-1} > 0$.

When $\alpha = 0.25$ we obtain the Corollary using the same ergodic argumentation for $\ln^{-1} n \sum_{i=1}^{n-1} p_i \sigma_{c,i}^4 i^{-1}$.

□

Lemma 4.5 *For random variables* $(\upsilon^{(n)}, \tilde{\mathbf{I}}^{(n)}, (\tilde{J}_i)_{i=1}^{\upsilon^{(n)}})$ *derived from the same random pair of lineages and a fixed jump probability* p

$$\mathrm{Var}\left[\mathrm{E}\left[\sum_{i=1}^{\upsilon^{(n)}} \tilde{J}_i e^{-2\alpha(\tau^{(n)}+\ldots+T_{\tilde{\mathrm{I}}_i^{(n)}+1})} | \mathscr{Y}_n\right]\right] \lesssim p \begin{cases} n^{-4\alpha} & 0 < \alpha < 0.5, \\ n^{-2} \ln n & \alpha = 0.5, \\ n^{-2} & 0.5 < \alpha. \end{cases} \tag{10}$$

PROOF We introduce the notation

$$\Psi^{(n)} := \sum_{i=1}^{\upsilon^{(n)}} \tilde{J}_i e^{-2\alpha(\tau^{(n)}+\ldots+T_{\tilde{\mathrm{I}}_i^{(n)}+1})}$$

and by definition we have

$$\mathrm{Var}\left[\mathrm{E}\left[\sum_{i=1}^{\upsilon^{(n)}} \tilde{J}_i e^{-2\alpha(\tau^{(n)}+\ldots+T_{\tilde{\mathrm{I}}_i^{(n)}})}\right]\right] = \mathrm{E}\left[\left(\mathrm{E}\left[\Psi^{(n)} | \mathscr{Y}_n\right]\right)^2\right] - \left(\mathrm{E}\left[\Psi^{(n)} | \mathscr{Y}_n\right]\right)^2.$$

We introduce the random variable

$$\phi_i = \tilde{Z}_i \tilde{\mathbf{I}}_i e^{-2\alpha(T_n+\ldots+T_{i+1})}$$

and obviously (for $i_1 < i_2$)

$$
\begin{aligned}
\mathrm{E}\left[\phi_i\right] &= \tfrac{2p}{i+1} b_{n,2\alpha}/b_{i,2\alpha}, \\
\mathrm{E}\left[\phi_i^2\right] &= \tfrac{2p}{i+1} b_{n,4\alpha}/b_{i,4\alpha}, \\
\mathrm{E}\left[\phi_{i_1}\phi_{i_2}\right] &= \tfrac{4p^2}{(i_1+1)(i_2+1)} \tfrac{b_{n,4\alpha}}{b_{i_2,4\alpha}} \tfrac{b_{i_2,2\alpha}}{b_{i_1,2\alpha}}.
\end{aligned}
$$

As usual let $(\tau_1^{(n)}, \upsilon_1^{(n)}, \Psi_1^{(n)})$ and $(\tau_2^{(n)}, \upsilon_2^{(n)}, \Psi_2^{(n)})$ be two independent copies of $(\tau^{(n)}, \upsilon^{(n)}, \Psi^{(n)})$ [cf. Lemma 11 in 6] and now

$$
\mathrm{E}\left[\left(\mathrm{E}\left[\Psi^{(n)}|\mathscr{Y}_n\right]\right)^2\right] = \mathrm{E}\left[\mathrm{E}\left[\Psi_1^{(n)}|\mathscr{Y}_n\right]\mathrm{E}\left[\Psi_2^{(n)}|\mathscr{Y}_n\right]\right] = \mathrm{E}\left[\mathrm{E}\left[\Psi_1^{(n)}\Psi_2^{(n)}|\mathscr{Y}_n\right]\right] = \mathrm{E}\left[\Psi_1^{(n)}\Psi_2^{(n)}\right].
$$

Writing out

$$
\begin{aligned}
\mathrm{Var}\left[\mathrm{E}\left[\Psi^{(n)}|\mathscr{Y}_n\right]\right] &= \mathrm{E}\left[\Psi_1^{(n)}\Psi_2^{(n)}\right] - \left(\mathrm{E}\left[\Psi^{(n)}\right]\right)^2 \qquad (11) \\
&= \sum_{k=1}^{n-1} \pi_{k,n}^2 \Bigg(\underbrace{\sum_{i=1}^{k-1}\left(\mathrm{E}\left[\phi_i^2\right] - \mathrm{E}\left[\phi_i\right]^2\right)}_{\mathrm{I}} + \underbrace{2\sum_{1=i_1<i_2}^{k-1}\left(\mathrm{E}\left[\phi_{i_1}\phi_{i_2}\right] - \mathrm{E}\left[\phi_{i_1}\right]\mathrm{E}\left[\phi_{i_2}\right]\right)}_{\mathrm{II}} \Bigg) \\
&\quad + 2\sum_{1=k_1<k_2}^{n-1} \pi_{k_1,n}\pi_{k_2,n} \Bigg(\underbrace{\sum_{i=1}^{k_1-1}\left(\mathrm{E}\left[\phi_i^2\right] - \mathrm{E}\left[\phi_i\right]^2\right)}_{\mathrm{III}} \\
&\quad + \underbrace{2\sum_{1=i_1<i_2}^{k_1-1}\left(\mathrm{E}\left[\phi_{i_1}\phi_{i_2}\right] - \mathrm{E}\left[\phi_{i_1}\right]\mathrm{E}\left[\phi_{i_2}\right]\right)}_{\mathrm{IV}} + \underbrace{2\sum_{i_1=1}^{k_1-1}\sum_{i_2=1}^{k_2-1}\left(\mathrm{E}\left[\phi_{i_1}\phi_{i_2}\right] - \mathrm{E}\left[\phi_{i_1}\right]\mathrm{E}\left[\phi_{i_2}\right]\right)}_{\mathrm{V}} \Bigg).
\end{aligned}
$$

We first observe

$$
\begin{aligned}
&\mathrm{E}\left[\phi_i^2\right] - \mathrm{E}\left[\phi_i\right]^2 = \tfrac{2p}{i+1}\left(\tfrac{b_{n,4\alpha}}{b_{i,4\alpha}} - \tfrac{2p}{i+1}\left(\tfrac{b_{n,2\alpha}}{b_{i,2\alpha}}\right)^2\right) = \tfrac{2p}{i+1}\Bigg(\tfrac{(i+1)^2}{(i+1+2\alpha)^2}\tfrac{(i+1)+(4\alpha-1)+(i+1)^{-1}4\alpha(\alpha-1)}{(i+1+4\alpha)}\tfrac{b_{n,4\alpha}}{b_{i+4\alpha}} \\
&+4\alpha^2\tfrac{b_{n,4\alpha}}{b_{i,2\alpha}^2}\sum_{j=i+2}^{n-1}\tfrac{b_{j,2\alpha}^2}{b_{j,4\alpha}}\tfrac{1}{j(j+4\alpha)} + \left(\tfrac{b_{n,2\alpha}}{b_{i,2\alpha}}\right)^2\tfrac{n(1-2p)+4\alpha(1-2p)+n^{-1}4\alpha^2}{n+4\alpha}\Bigg)
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathrm{E}\left[\phi_{i_1}\phi_{i_2}\right]-\mathrm{E}\left[\phi_{i_1}\right]\mathrm{E}\left[\phi_{i_2}\right]=\tfrac{4p^2}{(i_1+1)(i_2+1)}\left(\tfrac{b_{n,4\alpha}}{b_{i_2,4\alpha}}\tfrac{b_{i_2,2\alpha}}{b_{i_1,2\alpha}}-\left(\tfrac{b_{n,2\alpha}}{b_{i_1,2\alpha}}\right)\left(\tfrac{b_{n,2\alpha}}{b_{i_2,2\alpha}}\right)\right)\\
&=\tfrac{4p^2}{(i_1+1)(i_2+1)}\tfrac{b_{n,4\alpha}b_{i_2,2\alpha}}{b_{i_1,2\alpha}b_{i_2,2\alpha}^2}\left(\sum_{j=i_2+1}^{n}\tfrac{b_{j,2\alpha}^2}{b_{j,4\alpha}}\tfrac{4\alpha^2}{j(j+4\alpha)}\right).
\end{aligned}
$$

Using the above we consider each of the five components in this sum separately.

I

$$
\begin{aligned}
&\sum_{k=1}^{n-1}\pi_{k,n}^2\sum_{i=1}^{k-1}\left(\mathrm{E}\left[\phi_i^2\right]-\mathrm{E}\left[\phi_i\right]^2\right)\\
\lesssim\;& 4pn^{-4\alpha}\sum_{i=1}^{n}\left(i^{4\alpha-1}+(4\alpha-1)i^{4\alpha-2}+4\alpha(\alpha-1)i^{4\alpha-3}+4\alpha^2i^{4\alpha-2}+(1-2p)i^{4\alpha-1}\right)\sum_{k=i+1}^{n}k^{-4}\\
\lesssim\;& pC\begin{cases} n^{-4\alpha} & 0<\alpha<0.75\\ n^{-3}\ln n & \alpha=0.75\\ n^{-3} & 0.75<\alpha\end{cases}
\end{aligned}
$$

II

$$
\begin{aligned}
&\sum_{k=1}^{n-1}\pi_{k,n}^2\sum_{1=i_1<i_2}^{k-1}\left(\mathrm{E}\left[\phi_{i_1}\phi_{i_2}\right]-\mathrm{E}\left[\phi_{i_1}\right]\mathrm{E}\left[\phi_{i_2}\right]\right)\lesssim 64\alpha^2p^2n^{-4\alpha}\sum_{k=1}^{n}k^{-4}\sum_{i_1=1}^{k}i_1^{2\alpha-1}\sum_{i_2=i_1+1}^{k}i_2^{2\alpha-2}\\
\lesssim\;& Cp^2\begin{cases} n^{-4\alpha}\sum_{i_1=1}^{n}i_1^{4\alpha-2}\sum_{k=i_1+1}^{n}k^{-4} & 0<\alpha<0.5\\ n^{-2}\sum_{k=1}^{n}k^{-4}\sum_{i_2=2}^{k}1 & \alpha=0.5\\ n^{-4\alpha}\sum_{i_1=1}^{n}i_1^{4\alpha-2}\sum_{k=i_1+1}^{n}k^{-4} & 0.5<\alpha\end{cases}\lesssim Cp^2\begin{cases} n^{-4\alpha} & 0<\alpha<1\\ n^{-4}\ln n & \alpha=1\\ n^{-4} & 1<\alpha\end{cases}
\end{aligned}
$$

III

$$
\begin{aligned}
&\sum_{1=k_1<k_2}^{n-1}\pi_{k_1,n}\pi_{k_2,n}\sum_{i=1}^{k_1-1}\left(\mathrm{E}\left[\phi_i^2\right]-\mathrm{E}\left[\phi_i\right]^2\right)\\
\lesssim\;& 8pn^{-4\alpha}\sum_{i=1}^{n}\left(i^{4\alpha-1}+(4\alpha-1)i^{4\alpha-2}+4\alpha(\alpha-1)i^{4\alpha-3}+4\alpha^2i^{4\alpha-2}+(1-2p)i^{4\alpha-1}\right)\sum_{k_1=i+1}^{n}k_1^{-3}\\
\lesssim\;& pC\begin{cases} n^{-4\alpha} & 0<\alpha<0.5\\ n^{-2}\ln n & \alpha=0.5\\ n^{-2} & 0.5<\alpha\end{cases}
\end{aligned}
$$

IV

$$
\begin{aligned}
&\sum_{1=k_1<k_2}^{n-1} \pi_{k_1,n}\pi_{k_2,n} \sum_{1=i_1<i_2}^{k_1-1} \left(\mathrm{E}\left[\phi_{i_1}\phi_{i_2}\right] - \mathrm{E}\left[\phi_{i_1}\right]\mathrm{E}\left[\phi_{i_2}\right]\right) \\
\lesssim\ & 64\alpha^2 p^2 n^{-4\alpha} \sum_{1=k_1<k_2}^{n} k_1^{-2}k_2^{-2} \sum_{1=i_1<i_2}^{k_1} \left(i_1^{2\alpha-1} i_2^{2\alpha-2}\right) \lesssim p^2 C \begin{cases} n^{-4\alpha} & 0<\alpha<0.75 \\ n^{-4\alpha}\ln n & \alpha = 0.75 \\ n^{-3} & 0.75<\alpha \end{cases}
\end{aligned}
$$

V

$$
\begin{aligned}
&\sum_{1=k_1<k_2}^{n-1} \pi_{k_1,n}\pi_{k_2,n} \sum_{i_1=1}^{k_1-1}\sum_{i_2=1}^{k_2-1} \left(\mathrm{E}\left[\phi_{i_1}\phi_{i_2}\right] - \mathrm{E}\left[\phi_{i_1}\right]\mathrm{E}\left[\phi_{i_2}\right]\right) \\
\lesssim\ & 64\alpha^2 p^2 n^{-4\alpha} \sum_{1=k_1<k_2}^{n} k_1^{-2}k_2^{-2} \sum_{i_1=1}^{k_1}\sum_{i_2=1}^{k_2} \left(i_1^{2\alpha-1} i_2^{2\alpha-2}\right) \\
\lesssim\ & 64\alpha^2 p^2 n^{-4\alpha} \begin{cases} \sum_{i_1=1}^{n} i_1^{2\alpha-1} \sum_{k_1=i_1+1}^{n} k_1^{-2} \left(\sum_{i_2=1}^{n} i_2^{2\alpha-2} \sum_{k_2=i_2+1}^{n} k_2^{-2}\right) & \alpha \notin \{0.5, 1\} \\ \sum_{k_1=1}^{n} k_1^{-1} \left(\sum_{k_2=k_1+1}^{n} k_2^{-2} H_{k_2}\right) & \alpha = 0.5 \\ \frac{1}{2} \sum_{1=k_1<k_2}^{n} k_2^{-1} & \alpha = 1 \end{cases} \\
\lesssim\ & p^2 C \begin{cases} n^{-2} & \alpha = 0.5 \\ n^{-4\alpha} \sum_{i_1=1}^{n} i_1^{2\alpha-1} \sum_{k_1=i_1+1}^{n} k_1^{-2} & \alpha \in (0,1)\setminus\{0.5\} \\ n^{-3} & \alpha = 1 \\ n^{-2\alpha-2} \sum_{i_1=1}^{n} i_1^{2\alpha-1} \sum_{k_1=i_1+1}^{n} k_1^{-2} & 1<\alpha \end{cases} \lesssim p^2 C \begin{cases} n^{-4\alpha} & 0<\alpha\le 0.5 \\ n^{-2\alpha-1} & 0.5<\alpha<1 \\ n^{-3} & 1\le\alpha. \end{cases}
\end{aligned}
$$

Putting I–V together we obtain

$$
\mathrm{Var}\left[\Psi^{(n)}\right] \le pC \begin{cases} n^{-4\alpha} & 0<\alpha<0.5 \\ n^{-2}\ln n & \alpha = 0.5 \\ n^{-2} & 0.5<\alpha. \end{cases}
$$

The variance is bounded from below by III and as these derivations are correct up to a constant (independent of p) the variance behaves as above. This is further illustrated by graphs in Fig. 4.

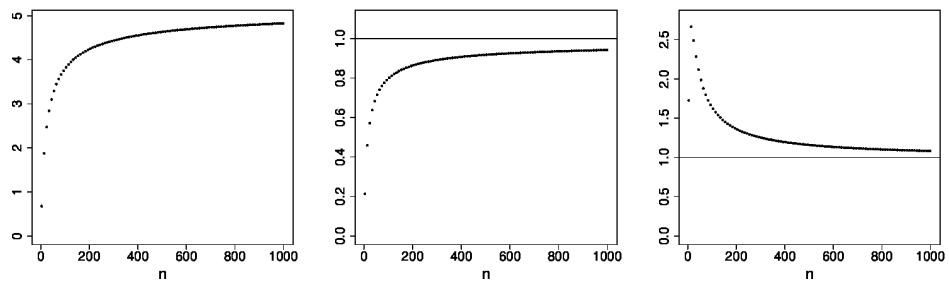

Figure 4: Numerical evaluation of scaled Eq. (11) for different values of α. The scaling for left: $\alpha = 0.35$ equals $n^{-4\alpha}$, centre: $\alpha = 0.5$ equals $16p(1-p)n^{-2}\log n$ and right $\alpha = 1$ equals $(32p(1-p)(1/(4\alpha-2)-1/(4\alpha-1))/(4\alpha))n^{-2}$. In all cases, $p = 0.5$.

□

The proof of the next Corollary, 4.6, is exactly the same as of Corollary 4.4.

Corollary 4.6 *Let p_n and $\sigma^2_{c,n}$ be respectively the jump probability and variance at the n–th speciation event, such that the sequence $\sigma^4_{c,n}p_n$ is bounded. We have*

$$\begin{aligned} n\ln^{-1} n\,\mathrm{Var}\left[\sum_{i=1}^{n-1}\sigma^2_{c,i}\phi_i\right] &\to 0 \quad \text{for } \alpha = 0.5, \\ n\,\mathrm{Var}\left[\sum_{i=1}^{n-1}\sigma^2_{c,i}\phi_i\right] &\to 0 \quad \text{for } 0.5 < \alpha. \end{aligned}$$

iff $\sigma^4_{c,n}p_n \to 0$ with density 1.

5 Proof of the Central Limit Theorems 3.1 and 3.5

To avoid unnecessary notation it will be always assumed that under a given summation sign the random variables $(\Upsilon^{(n)}, \mathrm{I}^{(n)}, (J_i)_{i=1}^{\Upsilon^{(n)}})$ are derived from the same random lineage and also $(\upsilon^{(n)}, \tilde{\mathrm{I}}^{(n)}, (\tilde{J}_i)_{i=1}^{\upsilon^{(n)}})$ are derived from the same random pair of lineages

Lemma 5.1 *Conditional on $\mathscr{Y}_n$ the first two moments of the scaled sample average are*

$$
\begin{aligned}
\mathrm{E}\left[\overline{Y}_n|\mathscr{Y}_n\right] &= \delta e^{-\alpha U^{(n)}} \\
\mathrm{E}\left[\overline{Y}_n^2|\mathscr{Y}_n\right] &= n^{-1}-(1-\delta^2)e^{-2\alpha U^{(n)}}+(1-n^{-1})\,\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right] \\
&\quad +n^{-1}(\sigma_a^2/(2\alpha))^{-1}\,\mathrm{E}\left[\sum_{k=1}^{\Upsilon^{(n)}}\sigma^2_{c,\mathrm{I}_k^{(n)}}J_k e^{-2\alpha(T_n+\ldots+T_{\mathrm{I}_k^{(n)}+1})}|\mathscr{Y}_n\right] \\
&\quad +(1-n^{-1})(\sigma_a^2/(2\alpha))^{-1}\,\mathrm{E}\left[\sum_{k=1}^{\upsilon^{(n)}}\sigma^2_{c,\tilde{\mathrm{I}}_k^{(n)}}\tilde{J}_k e^{-2\alpha(\tau^{(n)}+\ldots+T_{\tilde{\mathrm{I}}_k+1})}|\mathscr{Y}_n\right], \\
\mathrm{Var}\left[\overline{Y}_n|\mathscr{Y}_n\right] &= n^{-1}-e^{-2\alpha U^{(n)}}+(1-n^{-1})\,\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right] \\
&\quad +n^{-1}(\sigma_a^2/(2\alpha))^{-1}\,\mathrm{E}\left[\sum_{k=1}^{\Upsilon^{(n)}}\sigma^2_{c,\mathrm{I}_k^{(n)}}J_k e^{-2\alpha(T_n+\ldots+T_{\mathrm{I}_k+1})}|\mathscr{Y}_n\right] \\
&\quad +(1-n^{-1})(\sigma_a^2/(2\alpha))^{-1}\,\mathrm{E}\left[\sum_{k=1}^{\upsilon^{(n)}}\sigma^2_{c,\tilde{\mathrm{I}}_k^{(n)}}\tilde{J}_k e^{-2\alpha(\tau^{(n)}+\ldots+T_{\tilde{\mathrm{I}}_k^{(n)}+1})}|\mathscr{Y}_n\right],
\end{aligned}
$$

PROOF The first equality is immediate. The variance follows from

$$
\begin{aligned}
\mathrm{Var}\left[Y_1+\ldots+Y_n|\mathscr{Y}_n\right] &= n(1-e^{-2\alpha U^{(n)}})+(\sigma_a^2/(2\alpha))^{-1}\sum_{i=1}^{n}\sum_{k=1}^{\Upsilon^{(i,n)}}\sigma^2_{c,\mathrm{I}_k^{(i,n)}}J_k^{(i,n)}e^{-2\alpha(T_n+\ldots+T_{\mathrm{I}_k^{(i,n)}})} \\
&\quad +2\sum_{i<j}^{n}\Bigg((e^{-2\alpha\tau^{(i,j,n)}}-e^{-2\alpha U^{(n)}})+ \\
&\quad (\sigma_a^2/(2\alpha))^{-1}\sum_{k=1}^{\upsilon^{(i,j,n)}}\sigma^2_{c,\mathrm{I}_k^{(i,j,n)}}J_k^{(i,j,n)}e^{-2\alpha(\tau^{(i,j,n)}+\ldots+T_{\mathrm{I}_k^{(i,j,n)}})}\Bigg) \\
&= n-n^2e^{-2\alpha U^{(n)}}+n(n-1)\,\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right] \\
&\quad +n(\sigma_a^2/(2\alpha))^{-1}\,\mathrm{E}\left[\sum_{k=1}^{\Upsilon^{(n)}}\sigma^2_{c,\mathrm{I}_k^{(n)}}J_k e^{-2\alpha(T_n+\ldots+T_{\mathrm{I}_k^{(n)}+1})}|\mathscr{Y}_n\right] \\
&\quad +n(n-1)(\sigma_a^2/(2\alpha))^{-1}\,\mathrm{E}\left[\sum_{k=1}^{\upsilon^{(n)}}\sigma^2_{c,\tilde{\mathrm{I}}_k^{(n)}}\tilde{J}_k e^{-2\alpha(\tau^{(n)}+\ldots+T_{\mathrm{I}_k^{(n)}})}|\mathscr{Y}_n\right]
\end{aligned}
$$

This immediately entails the second moment.

□

Lemma 5.2 *Assume that the jump probability is constant, equalling p, at every speciation event. Let*

$$a_n(\alpha) = \begin{cases} n^{2\alpha} & 0 < \alpha < 0.5, \\ n \ln^{-1} n & 0.5 = \alpha, \\ n & 0.5 < \alpha \end{cases}$$

and then for all $\alpha > 0$ and n greater than some $n(\alpha)$

$$W_n := a_n(\alpha) \mathrm{E}\left[\Psi^{(n)} | \mathscr{Y}_n\right],$$

is a submartingale and furthermore W_n converges a.s. and in L^1 to a random variable W_∞ with expectation

$$\mathrm{E}[W_\infty] = \begin{cases} 2p(2\alpha+1)/(2\alpha(1-2\alpha)) & 0 < \alpha < 0.5, \\ 4p & 0.5 = \alpha, \\ 4p/(2\alpha(2\alpha-1)) & 0.5 < \alpha. \end{cases}$$

PROOF PROOF FOR $\alpha > 0.5$

$$\begin{aligned} W_{n+1} &= (n+1)\tfrac{2}{(n+1)n} \sum_{i<j}^{n+1} \Psi_{ij}^{(n+1)} \\ &= e^{-2\alpha T_{n+1}} \left(\tfrac{n-1}{n} W_n + \tfrac{2}{n} \sum_{i=1}^{n} \xi_i \sum_{k=1}^{\Upsilon^{(i,n)}} J_k^{(i,n)} e^{-2\alpha(T_n + \ldots + T_{\mathrm{I}_k^{(i,n)}+1})} \right. \\ &\quad \left. + \tfrac{2}{n} \sum_{i=1}^{n} \xi_i \sum_{j \neq i}^{n} \sum_{k=1}^{\upsilon^{(i,j,n)}} J_k^{(i,j,n)} e^{-2\alpha(\tau^{(i,j,n)} + \ldots + T_{\mathrm{I}_k^{(i,j,n)}+1})} \right) \\ &= e^{-2\alpha T_{n+1}} \left(\tfrac{n-1}{n} W_n + \tfrac{2}{n} \sum_{i=1}^{n} \xi_i \Psi_i^{*(n)} + \tfrac{2}{n} \sum_{i=1}^{n} \xi_i \sum_{j \neq i}^{n} \Psi_{ij}^{(n)} \right), \end{aligned}$$

where ξ_i is a binary random variable indicating whether it is the i–th lineage that split (see Fig 5).

Obviously the distribution of the vector $(\xi_1, \ldots, \xi_n)$ is uniform on the n–element set $\{(1,0,\ldots,0),\ldots,(0,\ldots,0,1)\}$. In particular note

Figure 5: The situation of the process between the n–th and $n+1$–st split. Node m split so $\xi_m = 1$ and $\xi_i = 0$ for $i \neq m$. The time between the splits is $T_{n+1} \sim \exp(n+1)$.

$$
\begin{aligned}
\mathrm{E}\left[W_{n+1}|\mathscr{Y}_n\right] &= \tfrac{n+1}{n+1+2\alpha}\left(\tfrac{n-1}{n}W_n + \tfrac{2}{n^2}\,\mathrm{E}\left[\sum_{i=1}^{n}\Psi_i^{*(n)}|\mathscr{Y}_n\right] + \tfrac{2}{n^2}\,\mathrm{E}\left[\sum_{i=1}^{n}\sum_{j\neq i}^{n}\Psi_{ij}^{(n)}|\mathscr{Y}_n\right]\right)\\
&= \tfrac{n+1}{n+1+2\alpha}\left(\tfrac{n-1}{n}W_n + \tfrac{2}{n}\,\mathrm{E}\left[\Psi^{*(n)}|\mathscr{Y}_n\right] + \tfrac{2n(n-1)}{n^2}\,\mathrm{E}\left[\Psi^{(n)}|\mathscr{Y}_n\right]\right)\\
&= \tfrac{n+1}{n+1+2\alpha}\left(\tfrac{n-1}{n}W_n + \tfrac{2}{n}\,\mathrm{E}\left[\Psi^{*(n)}|\mathscr{Y}_n\right] + \tfrac{n-1}{n}W_n\right)\\
&= \tfrac{n+1}{n+1+2\alpha}\left(2\tfrac{n-1}{n}W_n + \tfrac{2}{n}\,\mathrm{E}\left[\Psi^{*(n)}|\mathscr{Y}_n\right]\right)\\
&= 2\tfrac{(n-1)(n+1)}{n(n+1+2\alpha)}W_n + \tfrac{2(n+1)}{n(n+1+2\alpha)}\,\mathrm{E}\left[\Psi^{*(n)}|\mathscr{Y}_n\right]\\
&\geq W_n \quad \text{for } n \text{ large enough.}
\end{aligned}
$$

Therefore W_n will be a submartingale with respect to $\mathscr{Y}_n$ from a certain n. We know that $\mathrm{E}[W_n] < C_E$ for some constant C_E, as $\mathrm{E}[W_n] \to 4p/(2\alpha(2\alpha-1))$ [Appendix A.2. 5] and hence by the martingale convergence theorem $W_n \to W_\infty$ a.s. for some random variable W_∞. As all expectations are finite $\mathrm{E}[W_\infty] < \infty$. Furthermore $\mathrm{Var}[W_n] < C_V$, for some constant C_V, as, by Lemma 4.5, $\mathrm{Var}[W_n] \to 8p(1-p)/2\alpha$, indicating $\mathrm{E}[W_n] \to \mathrm{E}[W_\infty]$ and therefore $\mathrm{E}[W_\infty] = 4p/(2\alpha(2\alpha-1))$. Also the previous implies uniform integrability of $\{W_n\}$ and hence L^1 convergence.

PROOF FOR $\alpha = 0.5$ is the same as the proof for $\alpha > 0.5$ except that we will have for n large enough

$$
\mathrm{E}\left[W_{n+1}|\mathscr{Y}_n\right] \geq 2\frac{(n-1)(n+1)\ln n}{n(n+2)\ln(n+1)}W_n + \frac{2(n+1)}{n(n+2)\ln(n+1)}\,\mathrm{E}\left[\Psi^{*(n)}|\mathscr{Y}_n\right] \geq W_n.
$$

Now following again Bartoszek [5]'s Appendix A.2. we have $\mathrm{E}[W_n] \to 4p$ and $\mathrm{Var}[W_n] \to 8p(1-p)$ (Lemma 4.5).

PROOF FOR $0 < \alpha < 0.5$ is the same as the proof for $\alpha > 0.5$ except that we will have for n large enough

$$\mathrm{E}\left[W_{n+1}|\mathscr{Y}_n\right] \geq 2\frac{(n+1)^{2\alpha}(n-1)}{n^{2\alpha}(n+1+2\alpha)}W_n + \frac{2(n+1)^{2\alpha+1}}{n(n+1+2\alpha)}\mathrm{E}\left[\Psi^{*^{(n)}}|\mathscr{Y}_n\right] \geq W_n.$$

Following Bartoszek [5]'s Appendix A.2. we obtain $\mathrm{E}\left[W_n\right] \to 2p(2\alpha+1)/(1-2\alpha)$ and $\mathrm{Var}\left[W_n\right]$ converges to a constant by Lemma 4.5.

□

PROOF OF THEOREM 3.1, PART (I), $\alpha > 0.5$

We will show convergence in probability of the conditional mean and variance

$$\begin{aligned}
\mu_n &:= \sqrt{n}\mathrm{E}\left[\overline{Y}_n|\mathscr{Y}_n\right] \overset{\mathbb{P}}{\to} 0 \quad n\to\infty \\
\sigma_n^2 &:= n\,\mathrm{Var}\left[\overline{Y}_n|\mathscr{Y}_n\right] \overset{\mathbb{P}}{\to} \sigma_\infty^2 \quad n\to\infty,
\end{aligned}$$

for a finite mean and variance random variable σ_∞^2. Then due to the conditional normality of $\overline{Y}_n$ this will give the convergence of characteristic functions and the desired weak convergence, i.e.

$$\mathrm{E}\left[e^{ix\sqrt{n}\cdot\overline{Y}_n}\right] = \mathrm{E}\left[e^{i\mu_n x - \sigma_n^2 x^2/2}\right] \to \mathrm{E}\left[e^{-\sigma_\infty^2 x^2/2}\right].$$

Using Lemma 5.1 and that the Laplace transform of the average coalescent time [Lemma 3 in 6] is

$$\mathrm{E}\left[e^{-2\alpha\tau_{ij}^{(n)}}\right] = \frac{2-(n+1)(2\alpha+1)b_{n,2\alpha}}{(n-1)(2\alpha-1)} = \frac{2}{2\alpha-1}n^{-1} + O(n^{-2\alpha}) \tag{12}$$

we can calculate

$$\begin{aligned}
\mathrm{E}\left[\mu_n\right] &= \delta\,\mathrm{E}\left[e^{-\alpha U^{(n)}}\right] = \delta b_{n,\alpha} = O(n^{-\alpha}), \\
\mathrm{Var}\left[\mu_n\right] &= n\left(\mathrm{E}\left[\mu_n^2\right] - \left(\mathrm{E}\left[\mu_n\right]\right)^2\right) = \delta^2 n\left(\mathrm{E}\left[e^{-2\alpha U^{(n)}}\right] - \left(\mathrm{E}\left[e^{-\alpha U^{(n)}}\right]\right)^2\right) = \delta^2 n\left(b_{n,2\alpha} - b_{n,\alpha}^2\right) \\
&= \delta^2\alpha n b_{n,2\alpha}\sum_{j=1}^{n}\frac{b_{j,\alpha}^2}{b_{j,2\alpha}}\frac{1}{j(j+2\alpha)} = O(n^{-2\alpha+1}).
\end{aligned}$$

Therefore we have $\mu_n \to 0$ in L^2 and hence in $\mathbb{P}$.

Lemma 5.1 states that

$$\begin{aligned}\sigma_n^2 &= 1 - ne^{-2\alpha U^{(n)}} + n(1-n^{-1})\,\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right] \\ &\quad +(\sigma_a^2/(2\alpha))^{-1}\,\mathrm{E}\left[\sum_{k=1}^{\Upsilon^{(n)}} \sigma^2_{c,\mathrm{I}_k^{(n)}} J_k e^{-2\alpha(T_n+\ldots+T_{\mathrm{I}_k^{(n)}+1})}|\mathscr{Y}_n\right] \\ &\quad +n(1-n^{-1})(\sigma_a^2/(2\alpha))^{-1}\,\mathrm{E}\left[\sum_{k=1}^{\upsilon^{(n)}} \sigma^2_{c,\tilde{\mathrm{I}}_k^{(n)}} \tilde{J}_k e^{-2\alpha(\tau^{(n)}+\ldots+T_{\mathrm{I}_k^{(n)}+1})}|\mathscr{Y}_n\right]\end{aligned}$$

Using Lemmata 4.1, 4.3, 4.5, 5.2, Bartoszek [5]'s Appendix A.2 and remembering that p_k, $\sigma^2_{c,k}$ were assumed constant (equalling p, σ_c^2 respectively), by looking at the individual components we can see that

$$\sigma_n^2 \xrightarrow{\mathbb{P}} 1 + \frac{2p\sigma_c^2}{\sigma_a^2} + \frac{\sigma_c^2 W_\infty}{\sigma_a^2} =: \sigma_\infty^2.$$

By Lemma 5.2

$$\mathrm{E}\left[\sigma_\infty^2\right] = 1 + \frac{2p\sigma_c^2}{\sigma_a^2} + \frac{4p\sigma_c^2}{2\alpha(2\alpha-1)\sigma_a^2}.$$

PROOF OF PART (II), $\alpha = 0.5$

This is proved in the same way as PART (I) except with the normalizing factor of the order of $n\ln^{-1} n$.

PROOF OF PART (III), $0 < \alpha < 0.5$

We notice that the martingale $H_n = (n+1)e^{(\alpha+1)U^{(n)}}\overline{Y}_n$ has uniformly bounded second moments. Namely by boundedness of σ_c^2, Lemma 5.1 and Cauchy–Schwarz

$$
\begin{aligned}
&\mathrm{E}\left[H_n^2\right] = (n+1)^2\,\mathrm{E}\left[e^{2(\alpha-1)U^{(n)}}\,\mathrm{E}\left[\overline{Y}_n^2|\mathscr{Y}_n\right]\right] \le Cn^2\left(n^{-1}\,\mathrm{E}\left[e^{-2(1-\alpha)U^{(n)}}\right] + \mathrm{E}\left[e^{-2(1-\alpha)U^{(n)}-2\alpha\tau^{(n)}}\right]\right.\\
&\left.+n^{-1}(\sigma_a^2/(2\alpha))^{-1}\,\mathrm{E}\left[e^{-2(1-\alpha)U^{(n)}}\Psi^{*(n)}\right] + (\sigma_a^2/(2\alpha))^{-1}\,\mathrm{E}\left[e^{-2(1-\alpha)U^{(n)}}\Psi^{(n)}\right]\right)\\
&\le Cn^2\left(n^{-1}n^{-2(1-\alpha)} + n^{-2(1-\alpha)}n^{-2\alpha} + n^{-1}n^{-2(1-\alpha)}n^{-4\alpha} + n^{-2(1-\alpha)}n^{-4\alpha}\right)\\
&\le C\left(n^{-1+2\alpha} + 1 + n^{-2\alpha-1} + n^{-2\alpha}\right) \to C < \infty.
\end{aligned}
$$

Hence, $\sup_n \mathrm{E}\left[H_n^2\right] < \infty$ and by the martingale convergence theorem, $H_n \to H_\infty$ a.s. and in L^2. As was done by Bartoszek and Sagitov [6] we obtain (Bartoszek and Sagitov [6]'s Lemma 9) $n^\alpha \overline{Y}_n \to V^{(\alpha-1)}H_\infty$ a.s. as in L^2. Notice that for the convergence to hold in this regime, it is not required that $\sigma_{c,n}^2$ is constant, only bounded. We may also obtain directly the first two moments of $n^\alpha \overline{Y}_n$ (however, for these formulae to hold, $\sigma_{c,n}^2$ has to be constant)

$$
\begin{aligned}
n^\alpha\,\mathrm{E}\left[\overline{Y}_n\right] &= \delta n^\alpha b_{n,\alpha} \to \delta\Gamma(1+\alpha)\\
n^{2\alpha}\,\mathrm{E}\left[\overline{Y}_n^2\right] &= n^{2\alpha-1} - (1-\delta^2)n^{2\alpha}b_{n,2\alpha} + n^{2\alpha}(1-n^{-1})\,\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}\right]\\
&\quad + n^{2\alpha-1}\sigma_c^2(\sigma_a^2/(2\alpha))^{-1}\,\mathrm{E}\left[\Psi^{*(n)}\right] + n^{2\alpha}\sigma_c^2(\sigma_a^2/(2\alpha))^{-1}\,\mathrm{E}\left[\Psi^{(n)}\right]\\
&\to -(1-\delta^2)\Gamma(2\alpha+1) + \tfrac{1+2\alpha}{1-2\alpha}\Gamma(1+2\alpha)2p\sigma_c^2(\sigma_a^2/(2\alpha))^{-1}.
\end{aligned}
$$

PROOF OF THEOREM 3.5, PART (I), $\alpha > 0.5$

From the proof of Part (I), Theorem 3.1 we know that $\mu_n \to 0$ in probability. By the same ergodic argument as in Corollary 4.4 we obtain

$$
\begin{aligned}
\mathrm{E}\left[\sigma_n^2\right] &= n\Bigg(n^{-1} - \mathrm{E}\left[e^{-2\alpha U^{(n)}}\right] + (1-n^{-1})\,\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}\right] \\
&\quad + n^{-1}(\sigma_a^2/(2\alpha))^{-1}\,\mathrm{E}\left[\sum_{k=1}^{\Upsilon^{(n)}} \sigma^2_{c,\mathrm{I}_k^{(n)}} J_k e^{-2\alpha(T_n+\ldots+T_{\mathrm{I}_k^{(n)}+1})}\right] \\
&\quad + (1-n^{-1})(\sigma_a^2/(2\alpha))^{-1}\,\mathrm{E}\left[\sum_{k=1}^{\upsilon^{(n)}} \sigma^2_{c,\tilde{\mathrm{I}}_k^{(n)}} \tilde{J}_k e^{-2\alpha(\tau^{(n)}+\ldots+T_{\tilde{\mathrm{I}}_k^{(n)}+1})}\right]\Bigg) \\
&\to \tfrac{2\alpha+1}{2\alpha-1}.
\end{aligned}
$$

Furthermore Corollaries 4.4 and 4.6 imply

$$
\begin{aligned}
\mathrm{Var}\left[\sigma_n^2\right] &= n^2\,\mathrm{Var}\left[\mathrm{Var}\left[\overline{Y}_n|\mathscr{Y}_n\right]\right] = n^{-2}\,\mathrm{Var}\left[\mathrm{Var}\left[Y_1+\ldots+Y_n|\mathscr{Y}_n\right]\right] \\
&\le C\Bigg(n^2\,\mathrm{Var}\left[e^{-2\alpha U^{(n)}}\right] + (n-1)^2\,\mathrm{Var}\left[\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n\right]\right] \\
&\quad + (\sigma_a^2/(2\alpha))^{-2}\,\mathrm{Var}\left[\mathrm{E}\left[\sum_{k=1}^{\Upsilon^{(n)}} \sigma^2_{c,\mathrm{I}_k^{(n)}} J_k e^{-2\alpha(T_n+\ldots+T_{\mathrm{I}_k^{(n)}+1})}|\mathscr{Y}_n\right]\right] \\
&\quad + (n-1)^2(2\sigma_a^2/(2\alpha))^{-2}\,\mathrm{Var}\left[\mathrm{E}\left[\sum_{k=1}^{\upsilon^{(n)}} \sigma^2_{c,\tilde{\mathrm{I}}_k^{(n)}} \tilde{J}_k e^{-2\alpha(\tau^{(n)}+\ldots+T_{\tilde{\mathrm{I}}_k^{(n)}+1})}|\mathscr{Y}_n\right]\right]\Bigg) \\
&\to 0.
\end{aligned}
$$

Therefore we obtain that $\sigma_n^2 \to (2\alpha+1)/(2\alpha-1)$ in probability and by convergence of characteristic functions

$$
\mathrm{E}\left[e^{ix\sqrt{n}\cdot\overline{Y}_n}\right] = \mathrm{E}\left[e^{i\mu_n x - \sigma_n^2 x^2/2}\right] \to \mathrm{E}\left[e^{-((2\alpha+1)/(2\alpha-1))x^2/2}\right]
$$

we obtain the asymptotic normality. Notice that on the other hand

$$
\begin{aligned}
&\mathrm{Var}\left[\sigma_n^2\right] \ge n^{-2}\,\mathrm{Var}\left[-n^2 e^{-2\alpha U^{(n)}} + n(n-1)(2\sigma_a^2/(2\alpha))^{-1}\,\mathrm{E}\left[\sum_{k=1}^{\upsilon^{(n)}} \sigma^2_{c,\tilde{\mathrm{I}}_k^{(n)}} \tilde{J}_k e^{-2\alpha(\tau^{(n)}+\ldots+T_{\tilde{\mathrm{I}}_k^{(n)}+1})}|\mathscr{Y}_n\right]\right] \\
&\ge \left(1-\tfrac{1}{\sqrt{2}}\right)\Bigg(n^2\,\mathrm{Var}\left[e^{-2\alpha U^{(n)}}\right] + (n-1)^2(2\sigma_a^2/(2\alpha))^{-2}\,\mathrm{Var}\left[\mathrm{E}\left[\sum_{k=1}^{\upsilon^{(n)}} \sigma^2_{c,\tilde{\mathrm{I}}_k^{(n)}} \tilde{J}_k e^{-2\alpha(\tau^{(n)}+\ldots+T_{\tilde{\mathrm{I}}_k^{(n)}+1})}|\mathscr{Y}_n\right]\right]\Bigg)
\end{aligned}
$$

implying that the convergence $p_n\sigma_{c,n}^4 \to 0$ with density 1 is a necessary assumption for the asymptotic normality.

PROOF OF PART (II), $\alpha = 0.5$ This is proved in the same way as PART (I) except with the normalizing factor of the order of $n \ln^{-1} n$.

Acknowledgements

KB would like to acknowledge Olle Nerman for his suggestion on adding jumps to the branching OU process, Wojciech Bartoszek for helpful suggestions concerning ergodic arguments, Venelin Mitov and Tanja Stadler for many discussions. KB was supported by the Knut and Alice Wallenberg Foundation.

References

[1] R. Adamczak and P. Miłoś. U–statistics of Ornstein–Uhlenbeck branching particle system. *J. Th. Probab.*, 27(4):1071–1111, 2014.

[2] R. Adamczak and P. Miłoś. CLT for Ornstein–Uhlenbeck branching particle system. *Elect. J. Probab.*, 20(42):1–35, 2015.

[3] C. Ané, L. S. T. Ho, and S. Roch. Phase transition on the convergence rate of parameter estimation under an Ornstein–Uhlenbeck diffusion on a tree. *ArXiv e-prints*, 2014.

[4] K. Bartoszek. The Laplace motion in phylogenetic comparative methods. In *Proceedings of the Eighteenth National Conference on Applications of Mathematics in Biology and Medicine, Krynica Morska*, pages 25–30, 2012.

[5] K. Bartoszek. Quantifying the effects of anagenetic and cladogenetic evolution. *Math. Biosci.*, 254:42–57, 2014.

[6] K. Bartoszek and S. Sagitov. Phylogenetic confidence intervals for the optimal trait value. *J. App. Prob.*, 52:1115–1132, 2015.

[7] K. Bartoszek, J. Pienaar, P. Mostad, S. Andersson, and T. F. Hansen. A phylogenetic comparative method for studying multivariate adaptation. *J. Theor. Biol.*, 314:204–215, 2012.

[8] F. Bokma. Detection of punctuated equilibrium from molecular phylogenies. *J. Evol. Biol.*, 15:1048–1056, 2002.

[9] F. Bokma. Testing for equal rates of cladogenesis in diverse taxa. *Evolution*, 57(11):2469–2474, 2003.

[10] F. Bokma. Detection of "punctuated equilibrium" by Bayesian estimation of speciation and extinction rates, ancestral character states, and rates of anagenetic and cladogenetic evolution on a molecular phylogeny. *Evolution*, 62(11):2718–2726, 2008.

[11] F. Bokma. Time, species and seperating their effects on trait variance in clades. *Syst. Biol.*, 59 (5):602–607, 2010.

[12] M. A. Butler and A. A. King. Phylogenetic comparative analysis: a modelling approach for adaptive evolution. *Am. Nat.*, 164(6):683–695, 2004.

[13] F. W. Crawford and M. A. Suchard. Diversity, disparity, and evolutionary rate estimation for unresolved Yule trees. *Syst. Biol.*, 62(3):439–455, 2013.

[14] P. Duchen, C. Leuenberger, S. M. Szilàgyi, L. Harmon, J. Eastman, M. Schweizer, and D. Wegmann. Inference of evolutionary jumps in large phylogenies using Lévy processes. *Syst. Biol.*, in press 2016.

[15] A. W. F. Edwards. Estimation of the branch points of a branching diffusion process. *J. Roy. Stat. Soc. B*, 32(2):155–174, 1970.

[16] N. Eldredge and S. J. Gould. Punctuated equilibria: an alternative to phyletic gradualism. In T. J. M. Schopf and J. M. Thomas, editors, *Models in Paleobiology*, pages 82–115. Freeman Cooper, San Francisco, 1972.

[17] T. Gernhard. The conditioned reconstructed process. *J. Theor. Biol.*, 253:769–778, 2008.

[18] T. Gernhard. New analytic results for speciation times in neutral models. *B. Math. Biol.*, 70: 1082–1097, 2008.

[19] S. J. Gould and N. Eldredge. Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology*, 3(2):115–151, 1977.

[20] S. J. Gould and N. Eldredge. Punctuated equilibrium comes of age. *Nature*, 366:223–227, 1993.

[21] T. F. Hansen. Stabilizing selection and the comparative analysis of adaptation. *Evolution*, 51 (5):1341–1351, 1997.

[22] M. J. Landis, J. G. Schraiber, and M. Liang. Phylogenetic analysis using Lévy processes: finding jumps in the evolution of continuous traits. *Syst. Biol.*, 62(2):193–204, 2013.

[23] T. M. Mattila and F. Bokma. Extant mammal body masses suggest punctuated equilibrium. *Proc. R. Soc. B*, 275:2195–2199, 2008.

[24] E. Mayr. Speciation and macroevolution. *Evolution*, 36:1192–1132, 1982.

[25] A. Ø. Mooers and D. Schluter. Fitting macroevolutionary models to phylogenies: an example using vertebrate body sizes. *Contrib. Zool.*, 68:3–18, 1998.

[26] A. Ø. Mooers, S. M. Vamosi, and D. Schluter. Using phylogenies to test macroevolutionary hypotheses of trait evolution in Cranes (Gruinae). *Am. Nat.*, 154:249–259, 1999.

[27] W. H. Mulder and F. W. Crawford. On the distribution of interspecies correlation for Markov models of character evolution on Yule trees. *J. Theor. Biol.*, 364:275–283, 2015.

[28] K Petersen. *Ergodic Theory*. Cambridge University Press, Cambridge, 1983.

[29] Y-X Ren, R. Song, and R. Zhang. Central limit theorems for supercritical branching Markov processes. *J. Func. Anal.*, 266:1716–1756, 2014.

[30] Y-X Ren, R. Song, and R. Zhang. Central limit theorems for supercritical superprocesses. *Stoch. Proc. Appl.*, 125:428–457, 2015.

[31] S. Sagitov and K. Bartoszek. Interspecies correlation for neutrally evolving traits. *J. Theor. Biol.*, 309:11–19, 2012.

[32] G. G. Simpson. *Tempo and mode in evolution*. Columbia biological series, 15, New York, 1947.

[33] T. Stadler. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. *J. Theor. Biol.*, 261(1):58–68, 2009.

[34] T. Stadler. Simulating trees with a fixed number of extant species. *Syst. Biol.*, 60(5):676–684, 2011.

[35] M. Steel and A. McKenzie. Properties of phylogenetic trees generated by Yule–type speciation models. *Math. Biosci.*, 170:91–112, 2001.