A Central Limit Theorem for Punctuated Equilibrium

KRZYSZTOF BARTOSZEK

February 16, 2016

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

1 A model for punctuated stabilizing selection

1.1 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 Ornstein–Uhlenbeck (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. 7, 10, 16]. There have been a few attempts to go beyond the diffusion framework into Laplace motion [4, 5, 17] and jumps at speciation points [5, 8, 9]. 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 $Y \sim \mathcal{N}(0, \sigma_c^2)$.

Combining jumps with an Ornstein–Uhlenbeck 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 [15]. 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 [18] 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."

We first introduce some notation, illustrated in Fig. 1 [see also 5, 6, 22]. 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 $k+1$ and t_{k+1} be the time between speciation events k and $k+1$ on a randomly chosen lineage.

Let $\mathscr{Y}_n$ be the σ–algebra that contains information on the Yule tree. The described above model was studied previously [5] where I showed that, conditional on the tree height and number of tip species (the n index on p_n and $\sigma_{c,n}^2$ will be discussed in the next section), the mean and variance of a tip species, $X^{(n)}$, are

$$\begin{array}{rcl} \mathrm{E}\left[X^{(n)} | \mathscr{Y}_n\right] & = & \theta + e^{-\alpha U_n}(X_0 - \theta) \\ \mathrm{Var}\left[X^{(n)} | \mathscr{Y}_n\right] & = & \frac{\sigma_a^2}{2\alpha}(1 - e^{-2\alpha U_n}) + \sigma_{c,n}^2 p_n \sum\limits_{i=2}^{\Upsilon+1} e^{-2\alpha(t_{\Upsilon+1} + \ldots + t_i)}. \end{array} \tag{4}$$

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 the tree, the covariance between a pair of extant traits, $X_1^{(n)}$ and $X_2^{(n)}$ is

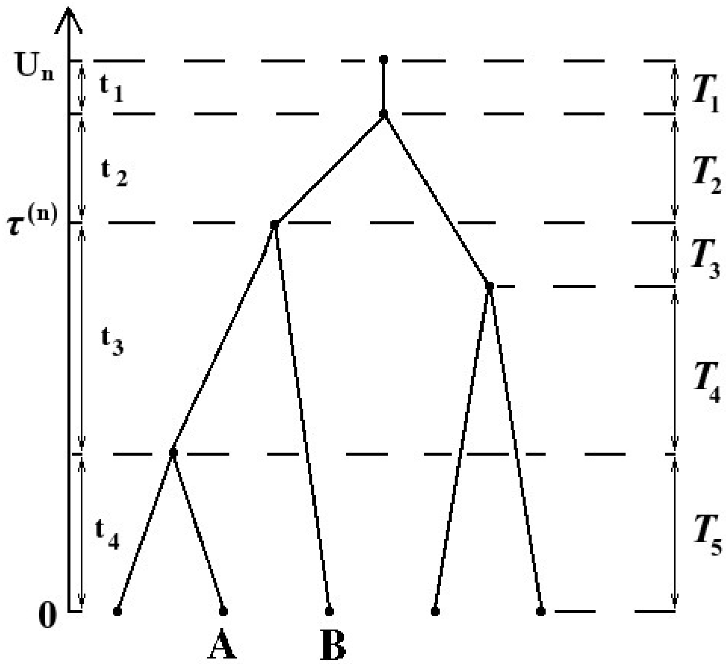

Figure 1: A pure–birth tree with the various time components marked on it. If we "randomly sample" node "A" then $\Upsilon^{(n)} = 3$ and the between speciation times on this lineage are $t_1 = T_1$, $t_2 = T_2$, $t_3 = T_3 + T_4$ and $t_4 = T_5$. If we "randomly sample" the pair of extant species "A" and "B" then $\upsilon^{(n)} = 1$ and the two nodes coalesced at time $\tau^{(n)}$. See also Bartoszek [5]'s Fig. A.8. for more detailed discussion on relevant notation.

$$\mathrm{Cov}\left[X_1^{(n)}, X_2^{(n)} | \mathscr{Y}_n\right] = \frac{\sigma_a^2}{2\alpha}(e^{-2\alpha\tau_{12}^{(n)}} - e^{-2\alpha U_n}) + \sigma_{c,n}^2 p_n \sum_{i=2}^{\upsilon+1} e^{-2\alpha(\tau_{12}^{(n)} + t_{\upsilon+1} + \ldots + t_i)}. \quad (5)$$

1.2 Tree model

In this work we consider a fundamental model of phylogenetic tree growth — the conditioned on number of tip species pure birth process. 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 $n \to \infty$, we stop the process when there are more and more species. The two parameters p_n and $\sigma_{c,n}^2$ are fixed for a given n, but may vary with n. In fact we will later see that for a central limit theorem to hold a product of them must decay to 0 with n. This is in fact the key characteristic that jumps bring in.

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 phylogenetics methods this branching process has been intensively studied [6, 11, 12, 13, 14, 19, 22, 23], hence 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 U_n be the height of the tree and $\tau^{(n)}$ the time (counted from today) till coalescent of a random pair tips. 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 [6, 22]

$$\begin{array}{rcl}
\mathrm{E}\left[U_n\right] & = & H_n, \\
\mathrm{E}\left[e^{-xU_n}\right] & = & b_{n,x}, \\
\mathrm{E}\left[\tau^{(n)}\right] & = & \frac{n+1}{n-1}H_n - \frac{2}{n-1}, \\
\mathrm{E}\left[e^{-x\tau^{(n)}}\right] & = & \left\{ \begin{array}{ll} \frac{2-(n+1)(x+1)b_{n,x}}{(n-1)(x-1)} & x \neq 1, \\ \frac{2}{n-1}\left(H_n - 1\right) - \frac{1}{n+1} & x = 1, \end{array} \right.
\end{array}$$

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.

We will call the considered model the Yule–Ornstein–Uhlenbeck with jumps (YOUj) process. For it I calculated [5] the mean, variance, covariance and interspecies correlation for random tips. We recall, denoting $\kappa_n = 2p_n\sigma_{c,n}^2/(2p_n\sigma_{c,n}^2 + \sigma_a^2)$, $\delta = (X_0 - \theta)/\sqrt{\sigma_a^2/2\alpha}$

$$\begin{array}{rcl}
\mathrm{E}\left[X^{(n)}\right] & = & b_{n,\alpha}X_0 + (1-b_{n,\alpha})\theta, \\
\mathrm{Var}\left[X^{(n)}\right] & = & \frac{\sigma_a^2+2p_n\sigma_{c,n}^2}{2\alpha}\left((1-\kappa_n)V_a^{(n)}(\alpha,\delta) + \kappa_n V_c^{(n)}(\alpha)\right), \\
\mathrm{Cov}\left[X_1^{(n)}, X_2^{(n)}\right] & = & \frac{\sigma_a^2+2p_n\sigma_{c,n}^2}{2\alpha}\left((1-\kappa_n)C_a^{(n)}(\alpha,\delta) + \kappa_n C_c^{(n)}(\alpha)\right), \\
\rho_n & = & \frac{(1-\kappa)C_a^{(n)}(\alpha,\delta)+\kappa_n C_c^{(n)}(\alpha)}{(1-\kappa_n)V_a^{(n)}(\alpha,\delta)+\kappa_n V_c^{(n)}(\alpha)},
\end{array} \tag{6}$$

where,

$$\begin{array}{rcl}
C_a^{(n)}(\alpha,\delta) & = & \left\{ \begin{array}{ll} \frac{2-(n+1)(2\alpha+1)b_{n,2\alpha}}{(n-1)(2\alpha-1)} - b_{n,2\alpha} + \delta^2\left(b_{n,2\alpha} - b_{n,\alpha}^2\right), & 0 < \alpha \neq 0.5, \\ \frac{2}{n-1}\left(H_n - 1\right) - \frac{2}{n+1} + \delta^2\left(\frac{1}{n+1} - b_{n,0.5}^2\right), & \alpha = 0.5, \end{array} \right. \\
C_c^{(n)}(\alpha) & = & \left\{ \begin{array}{ll} \frac{2-(2\alpha n-2\alpha+2)(2\alpha+1)b_{n,2\alpha}}{(n-1)(2\alpha-1)}, & 0 < \alpha \neq 0.5, \\ \frac{2}{n-1}\left(H_n - \frac{5n-1}{2(n+1)}\right), & \alpha = 0.5, \end{array} \right. \\
V_a^{(n)}(\alpha,\delta) & = & 1 - b_{n,2\alpha} + \delta^2(b_{n,2\alpha} - b_{n,\alpha}^2), \\
V_c^{(n)}(\alpha) & = & 1 - (1+2\alpha)b_{n,2\alpha}.
\end{array}$$

For large n these behave as

$$
\begin{array}{rcl}
\mathrm{E}\left[X^{(\infty)}\right] & = & \theta + O(n^{-\alpha}), \\
\mathrm{Var}\left[X^{(\infty)}\right] & = & \frac{\sigma_a^2+2p_\infty\sigma_{c,\infty}^2}{2\alpha} + O(n^{-2\alpha}), \\
\mathrm{Cov}\left[X_1^{(\infty)}, X_2^{(\infty)}\right] & \sim & \frac{\sigma_a^2+2p_\infty\sigma_{c,\infty}^2}{2\alpha} \cdot \begin{cases} \left((1-\kappa_\infty)C_a^{(\infty)}(\alpha,\delta) + \kappa_\infty C_c^{(\infty)}(\alpha)\right) n^{-2\alpha}, & 0<\alpha<0.5, \\ 2n^{-1}\ln n, & \alpha = 0.5, \\ \frac{2}{2\alpha-1}n^{-1}, & \alpha > 0.5, \end{cases} \\
\rho_n & \sim & \begin{cases} \left((1-\kappa_\infty)C_a^{(\infty)}(\alpha,\delta) + \kappa_\infty C_c^{(\infty)}(\alpha)\right) n^{-2\alpha}, & 0<\alpha<0.5, \\ 2n^{-1}\ln n, & \alpha = 0.5, \\ \frac{2}{2\alpha-1}n^{-1}, & \alpha > 0.5\ . \end{cases}
\end{array}
\tag{7}
$$

1.3 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, $Y^{(n)} = (X^{(n)} - \theta)/\sqrt{\gamma_n}$, where $\gamma_n := (\sigma_a^2 + 2p_n\sigma_{c,n}^2)/2\alpha$ and then denoting $\delta_n^* = (X_0 - \theta)/\sqrt{\gamma_n}$

$$
\begin{array}{rcl}
\mathrm{E}\left[Y^{(n)}\right] & = & \delta_n^* b_{n,\alpha}, \\
\mathrm{Var}\left[Y^{(n)}\right] & = & (1-\kappa_n)V_a^{(n)}(\alpha,\delta) + \kappa_n V_c^{(n)}(\alpha), \\
\mathrm{Cov}\left[Y_1^{(n)}, Y_2^{(n)}\right] & = & (1-\kappa_n)C_a^{(n)}(\alpha,\delta) + \kappa_n C_c^{(n)}(\alpha).
\end{array}
\tag{8}
$$

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

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

Then [cf. Lemma 10 in 6] we define

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

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.

2 Asymptotic regimes — main results

Branching Ornstein–Uhlenbeck models commonly have three asymptotic regimes. [1, 2, 3, 5, 6, 20, 21] . 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 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", 1, 2]. In the OU process with jumps setup the same asymptotic regimes can be observed, even though in [1, 2, 20, 21] the tree is observed at a given time point, t, with n_t being random. In what follows in the paper the constant C may change between (in)equalities. It may in particular depend on α.

Theorem 2.1 *Let $\overline{Y}_n = (\overline{X}_n - \theta)/\sqrt{\gamma_n}$ be the normalized sample mean of the YOUj process with $\overline{Y}_0 = \delta_0^*$. The process $\overline{Y}_n$ has the following asymptotic with n behaviour depending on α.*

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

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

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

$$\begin{aligned}
\mathrm{E}\left[Y_{\alpha,\delta_\infty^*,\kappa_\infty}\right] &= \delta_\infty^* \Gamma(1+\alpha), \\
\mathrm{E}\left[Y^2_{\alpha,\delta_\infty^*,\kappa_\infty}\right] &= \left(\delta_\infty^{*2} + (1-\kappa_\infty)\tfrac{4\alpha}{1-2\alpha}\right)\Gamma(1+2\alpha).
\end{aligned}$$

Remark 2.2 *The assumption $\sigma_{c,n}^2 \kappa_n \to 0$ is an essential one for $\alpha \geq 0.5$. This is visible from the proof of Lemma 3.4. 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 of the process.*

One natural way of achieving this desired limit is keeping $\sigma_{c,n}^2$ constant and allowing $p_n \to 0$, 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.

3 Key convergence lemmata

We will now prove a series of lemmata describing the asymptotics of driving components of the considered YOUj process. Let $\mathscr{Y}_n^*$ denote the σ–algebra that contains information on the Yule tree and jump pattern.

Lemma 3.1 *[Lemma 11 in 6]*

$$\mathrm{Var}\left[\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}|\mathscr{Y}_n^*\right]\right] = \begin{cases} O(n^{-4\alpha}) & 0<\alpha<0.75, \\ O(n^{-4\alpha}\ln n) & \alpha = 0.75, \\ O(n^{-3}) & 0.75<\alpha. \end{cases} \tag{9}$$

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,

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

Writing

$$\pi_{n,k} = 2\frac{n+1}{n-1}\frac{1}{(k+1)(k+2)}, \qquad f(a,k,n) := \frac{k+1}{a+k+1}\cdots\frac{n}{a+n}$$

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

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

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

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

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

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

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

$$\sum_{k_1=1}^{n-1} \sum_{k_2=k_1+1}^{n-1} f_{2\alpha}(k_1,k_2) A_{n,k_2}\pi_{n,k_1}\pi_{n,k_2} \lesssim Cb_{n,4\alpha} \sum_{k_1=1}^{n-1} \sum_{k_2=k_1+1}^{n-1} \frac{b_{k_2,2\alpha}}{b_{k_1,2\alpha} b_{k_2,4\alpha}} \frac{1}{k_1^2 k_2^3}$$

$$\lesssim n^{-4\alpha} \sum_{k_1=1}^{n-1} k_1^{2\alpha-2} \sum_{k_2=k_1+1}^{n-1} k_2^{2\alpha-3} \lesssim C \begin{cases} n^{-4\alpha} \sum_{k_1=1}^{n-1} k_1^{4\alpha-4} & 0<\alpha<1 \\ n^{-4} \sum_{k_2=2}^{n} k_2^{-1} \sum_{k_1=1}^{k_2} 1 & \alpha=1 \\ n^{-4\alpha} \sum_{k_2=2}^{n} k_2^{4\alpha-4} & 1<\alpha \end{cases}$$

$$\lesssim C \begin{cases} n^{-4\alpha} & 0<\alpha<0.75 \\ n^{-3}\ln n & \alpha=0.75 \\ n^{-3} & 0.75<\alpha<1 \\ n^{-4}\sum_{k_2=2}^{n} 1 & \alpha=1 \\ n^{-3} & 1<\alpha \end{cases} \lesssim C \begin{cases} n^{-4\alpha} & 0<\alpha<0.75 \\ n^{-3}\ln n & \alpha=0.75 \\ n^{-3} & 0.75<\alpha<1 \\ n^{-3} & \alpha=1 \\ n^{-3} & 1<\alpha. \end{cases}$$

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 3.2 *The above Lemma 3.1 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 3.1 holds. Theorem* 2 *[6] 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 is true, just with a different rate.*

Lemma 3.3 *Let J_i be a binary random variable indicating if a jump took place on the i–th (counting from the origin of the tree) speciation event of a randomly sampled lineage. For a fixed jump probability p we have*

$$\mathrm{Var}\left[\mathrm{E}\left[\sum_{i=2}^{\Upsilon^{(n)}+1} J_i e^{-2\alpha(t_{\Upsilon+1}+\ldots+t_i)}|\mathscr{Y}_n^*\right]\right] \lesssim p \begin{cases} n^{-4\alpha} & 0<\alpha<0.25 \\ n^{-1}\ln n & \alpha=0.25 \\ n^{-1} & 0.25<\alpha. \end{cases} \tag{10}$$

PROOF We introduce the random variables

$$\Psi^{*(n)} := \sum_{i=2}^{\Upsilon^{(n)}+1} J_i e^{-2\alpha(t_{\Upsilon+1}+\ldots+t_i)}$$

and

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

Recall that Z_i is the binary random variable if a jump took place at the i–th speciation event of the tree and $\mathbf{1}_i$ the indicator random variable if the i–th speciation event is on the randomly sampled lineage. 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 this

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

Now

$$
\begin{aligned}
\operatorname{Var}\left[\sum_{i=1}^{n-1}\phi_i^*\right] &= \sum_{i=1}^{n-1}\left(\mathrm{E}\left[\phi_i^{*2}\right]-(\mathrm{E}[\phi_i^*])^2\right)+2\sum_{i<j}^{n-1}\left(\mathrm{E}\left[\phi_i^*\phi_j^*\right]-\mathrm{E}[\phi_i^*]\,\mathrm{E}\left[\phi_j^*\right]\right) \qquad (11)\\
&= \sum_{i=1}^{n-1}\left(p\frac{b_{n,4\alpha}}{b_{i,4\alpha}}\left(\frac{n+1}{n-1}\frac{6}{(i+1)(i+2)}-\frac{1}{n-1}\frac{4}{i+1}\right)-\frac{4p^2}{(i+1)^2}\left(\frac{b_{n,2\alpha}}{b_{i,2\alpha}}\right)^2\right)\\
&\quad +2\sum_{i<j}^{n-1}\left(\frac{4p^2}{(i+1)(j+1)}\frac{b_{n,4\alpha}}{b_{j,4\alpha}}\frac{b_{j,2\alpha}}{b_{i,2\alpha}}-\frac{4p^2}{(i+1)(j+1)}\frac{b_{n,2\alpha}}{b_{i,2\alpha}}\frac{b_{n,2\alpha}}{b_{j,2\alpha}}\right)\\
&\lesssim 2p\sum_{i=1}^{n-1}\frac{1}{(i+1)^2}\left(3\frac{b_{n,4\alpha}}{b_{i,4\alpha}}-2p\left(\frac{b_{n,2\alpha}}{b_{i,2\alpha}}\right)^2\right) \quad \mathrm{I}\\
&\quad +4p(n-1)^{-1}\sum_{i=1}^{n-1}\frac{b_{n,4\alpha}}{b_{i,4\alpha}}\left(\frac{3}{(i+1)^2}-\frac{1}{i+1}\right) \quad \mathrm{II}\\
&\quad +8p^2\sum_{i<j}^{n-1}\left(\frac{1}{(i+1)(j+1)}\frac{b_{j,2\alpha}}{b_{i,2\alpha}}\left(\frac{b_{n,4\alpha}}{b_{j,4\alpha}}-\left(\frac{b_{n,2\alpha}}{b_{j,2\alpha}}\right)^2\right)\right). \quad \mathrm{III}
\end{aligned}
$$

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

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

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

II

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

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. 2.

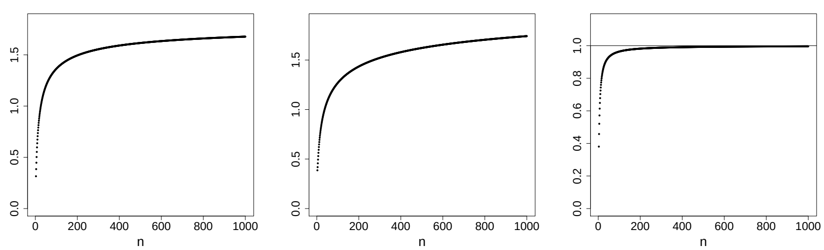

Figure 2: Numerical evaluation of scaled Eq. (11) for different values of α. The scaling for left: $\alpha=0.1$ equals $n^{-4\alpha}$, centre: $\alpha=0.25$ equals $n^{-1}\log n$ and right $\alpha=1$ equals $(2p(3-2p)/(4\alpha-1)-4p/(4\alpha)+32p^2\alpha^2(1/(8\alpha^2)+1/(2\alpha(2\alpha-1))-1/(4\alpha^2)^{-1}-1/((2\alpha-1)(4\alpha-1))))n^{-1}$. In all cases $p=0.5$.

$\square$

Lemma 3.4 *Using the same notation as in Lemma 3.3 we have for a fixed jump probability p*

$$\mathrm{Var}\left[\mathrm{E}\left[\sum_{i=2}^{\upsilon^{(n)}+1}J_ie^{-2\alpha\left(\tau^{(n)}+t_\upsilon+\ldots+t_{\upsilon_i+1}\right)}|\mathscr{Y}_n^*\right]\right]\lesssim p\begin{cases} n^{-4\alpha} & 0<\alpha<0.5,\\ n^{-2}\ln n & \alpha=0.5,\\ n^{-2} & 0.5<\alpha.\end{cases}\qquad(12)$$

PROOF We introduce the notation

$$\Psi^{(n)} := \sum_{i=2}^{\upsilon^{(n)}} J_i e^{-2\alpha\left(\tau^{(n)}+t_\upsilon+\ldots+t_{\upsilon_i+1}\right)}$$

and we obviously have

$$\mathrm{Var}\left[\mathrm{E}\left[\sum_{i=2}^{\upsilon^{(n)}} J_i e^{-2\alpha\left(\tau^{(n)}+t_\upsilon+\ldots+t_{\upsilon_i+1}\right)}|\mathscr{Y}_n^*\right]\right] = \mathrm{E}\left[\left(\mathrm{E}\left[\Psi^{(n)}|\mathscr{Y}_n\right]\right)^2\right] - \left(\mathrm{E}\left[\Psi^{(n)}|\mathscr{Y}_n\right]\right)^2.$$

We introduce the random variable

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

and obviously (for $i_1 < i_2$)

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

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 $(\tau^{(n)}, \upsilon^{(n)}, \Psi^{(n)})$ 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

$$\begin{aligned} \mathrm{Var}\left[\mathrm{E}\left[\Psi^{(n)}|\mathscr{Y}_n^*\right]\right] &= \mathrm{E}\left[\Psi_1^{(n)}\Psi_2^{(n)}\right] - \left(\mathrm{E}\left[\Psi^{(n)}\right]\right)^2 \qquad (13) \\ &= \sum_{k=1}^{n-1} \pi_{k,n}^2 \left(\underbrace{\sum_{i=1}^{k-1} \left(\mathrm{E}\left[\phi_i^2\right] - \mathrm{E}\left[\phi_i\right]^2\right)}_{\mathrm{I}} + 2 \underbrace{\sum_{1=i_1<i_2}^{k-1} \left(\mathrm{E}\left[\phi_{i_1}\phi_{i_2}\right] - \mathrm{E}\left[\phi_{i_1}\right]\mathrm{E}\left[\phi_{i_2}\right]\right)}_{\mathrm{II}} \right) \\ &\quad + 2\sum_{1=k_1<k_2}^{n-1} \pi_{k_1,n}\pi_{k_2,n} \left(\underbrace{\sum_{i=1}^{k_1-1} \left(\mathrm{E}\left[\phi_i^2\right] - \mathrm{E}\left[\phi_i\right]^2\right)}_{\mathrm{III}} \right. \\ &\quad \left. + 2 \underbrace{\sum_{1=i_1<i_2}^{k_1-1} \left(\mathrm{E}\left[\phi_{i_1}\phi_{i_2}\right] - \mathrm{E}\left[\phi_{i_1}\right]\mathrm{E}\left[\phi_{i_2}\right]\right)}_{\mathrm{IV}} + 2 \underbrace{\sum_{i_1=1}^{k_1-1}\sum_{i_2=1}^{k_2-1} \left(\mathrm{E}\left[\phi_{i_1}\phi_{i_2}\right] - \mathrm{E}\left[\phi_{i_1}\right]\mathrm{E}\left[\phi_{i_2}\right]\right)}_{\mathrm{V}} \right). \end{aligned}$$

We first observe

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

and

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

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

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

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. 3.

□

4 Proof of the central limit theorem, Theorem 2.1

Lemma 4.1 *Conditional on $\mathscr{Y}_n^*$, for given $p_n, \kappa_n > 0$ and denoting $v_n = \alpha p_n^{-1}\kappa_n$ the first two moments of the scaled sample average are*

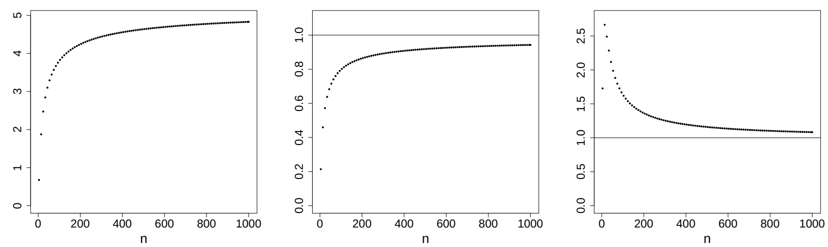

Figure 3: Numerical evaluation of scaled Eq. (13) for different values of α. The scaling for left: $\alpha = 0.35$ equals $n^{-4\alpha}$, centre: $\alpha = 0.5$ equals $16p(1-p)n^{-2}\log n$ and right $\alpha = 1$ equals $(32p(1-p)(1/(4\alpha-2)-1/(4\alpha-1))/(4\alpha))n^{-2}$. In all cases $p = 0.5$.

$$
\begin{array}{rcl}
\mathrm{E}\left[\overline{Y}_n|\mathscr{Y}_n^*\right] & = & \delta_n^* e^{-\alpha U_n} \\
\mathrm{E}\left[\overline{Y}_n^2|\mathscr{Y}_n^*\right] & = & n^{-1}(1-\kappa_n) - (1-\kappa_n-\delta_n^{*2})e^{-2\alpha U_n} + (1-n^{-1})(1-\kappa_n)\,\mathrm{E}\left[e^{-2\alpha\tau_{ij}^{(n)}}|\mathscr{Y}_n^*\right] \\
& & +n^{-1}v_n\,\mathrm{E}\left[\sum_{k=2}^{\Upsilon+1} J_k e^{-2\alpha(t_{\Upsilon+1}+\ldots+t_k)}|\mathscr{Y}_n^*\right] \\
& & +(1-n^{-1})v_n\,\mathrm{E}\left[\sum_{k=2}^{\upsilon+1} J_k e^{-2\alpha(\tau^{(n)}+t_{\upsilon+1}+\ldots+t_k)}|\mathscr{Y}_n^*\right], \\
\mathrm{Var}\left[\overline{Y}_n|\mathscr{Y}_n^*\right] & = & n^{-1}(1-\kappa_n) - (1-\kappa_n)e^{-2\alpha U_n} + (1-n^{-1})(1-\kappa_n)\,\mathrm{E}\left[e^{-2\alpha\tau_{ij}^{(n)}}|\mathscr{Y}_n^*\right] \\
& & +n^{-1}v_n\,\mathrm{E}\left[\sum_{k=2}^{\Upsilon+1} J_k e^{-2\alpha(t_{\Upsilon+1}+\ldots+t_k)}|\mathscr{Y}_n^*\right] \\
& & +(1-n^{-1})v_n\,\mathrm{E}\left[\sum_{k=2}^{\upsilon+1} J_k e^{-2\alpha(\tau^{(n)}+t_{\upsilon+1}+\ldots+t_k)}|\mathscr{Y}_n^*\right],
\end{array}
$$

where J_k is the binary random variable indicating if the jump took place at the k–th speciation event on the randomly sampled path.

PROOF The first equality is immediate. The variance follows from

$$\begin{aligned}
\mathrm{Var}\left[Y_1+\ldots+Y_n|\mathscr{Y}_n^*\right] &= n(1-\kappa_n)(1-e^{-2\alpha U_n})+v_n\sum_{i=1}^{n}\sum_{k=2}^{\Upsilon_i+1}J_{k^{(i)}}e^{-2\alpha(t^{(i)}_{\Upsilon_i+1}+\ldots+t^{(i)}_k)} \\
&\quad +2\sum_{i<j}^{n}\Bigg((1-\kappa_n)(e^{-2\alpha\tau_{ij}^{(n)}}-e^{-2\alpha U_n})+ \\
&\quad v_n\sum_{k=2}^{\upsilon_{(i,j)}+1}J_{k^{(i,j)}}e^{-2\alpha(\tau_{ij}^{(n)}+t^{(i,j)}_{\upsilon_{(i,j)}+1}+\ldots+t^{(i,j)}_k)}\Bigg) \\
&= n(1-\kappa_n)-n^2(1-\kappa_n)e^{-2\alpha U_n}+n(n-1)(1-\kappa_n)\,\mathrm{E}\left[e^{-2\alpha\tau_{ij}^{(n)}}|\mathscr{Y}_n^*\right] \\
&\quad +nv_n\,\mathrm{E}\left[\sum_{k=2}^{\Upsilon+1}J_ke^{-2\alpha(t_{\Upsilon+1}+\ldots+t_k)}|\mathscr{Y}_n^*\right] \\
&\quad +n(n-1)v_n\,\mathrm{E}\left[\sum_{k=2}^{\upsilon+1}J_ke^{-2\alpha(\tau^{(n)}+t_{\upsilon+1}+\ldots+t_k)}|\mathscr{Y}_n^*\right]
\end{aligned}$$

This immediately entails the second moment.

□

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

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

$$(\mu_n,\sigma_n^2) := \left(\sqrt{n}\,\mathrm{E}\left[\overline{Y}_n|\mathscr{Y}_n^*\right], n\,\mathrm{Var}\left[\overline{Y}_n|\mathscr{Y}_n^*\right]\right) \xrightarrow{P} \left(0,\frac{2\alpha+1}{2\alpha-1}\right) \quad n\to\infty,$$

as due to the conditional normality of $\overline{Y}_n$ this will give the convergence of characteristic functions and the desired weak convergence, i.e.

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

Using Lemma 4.1, the Laplace transform of the average coalescent time [Lemma 3 in 6]

$$\mathrm{E}\left[e^{-2\alpha\tau_{ij}^{(n)}}\right] = \frac{2-(n+1)(2\alpha+1)b_{n,2\alpha}}{(n-1)(2\alpha-1)} = \frac{2}{2\alpha-1}n^{-1}+O(n^{-2\alpha}) \qquad (14)$$

and the following expectations [Appendix A.2 in 5]

$$\begin{aligned}
\mathrm{E}\left[\sum_{i=2}^{\Upsilon+1}e^{-2\alpha(t_{\Upsilon+1}+\ldots+t_i)}\right] &= \tfrac{2}{2\alpha}\left(1-(1+2\alpha)b_{n,2\alpha}\right) = \tfrac{2}{2\alpha}+O(n^{-2\alpha}) \\
\mathrm{E}\left[\sum_{i=2}^{\upsilon+1}e^{-2\alpha(\tau+t_{\upsilon+1}+\ldots+t_i)}\right] &= \tfrac{2}{2\alpha}\left(\tfrac{2-(2\alpha+1)(2\alpha n-2\alpha+2)b_{n,2\alpha}}{(n-1)(2\alpha-1)}\right) = \tfrac{4}{2\alpha(2\alpha-1)}n^{-1}+O(n^{-2\alpha})
\end{aligned} \qquad (15)$$

we obtain

$$
\begin{aligned}
\mathrm{E}[\mu_n] &= \delta_n^* \mathrm{E}\left[e^{-\alpha U_n}\right] = \delta_n^* b_{n,\alpha} = O(n^{-\alpha}),\\
\mathrm{Var}[\mu_n] &= n\left(\mathrm{E}\left[\mu_n^2\right] - (\mathrm{E}[\mu_n])^2\right) = {\delta_n^*}^2 n\left(\mathrm{E}\left[e^{-2\alpha U_n}\right] - \left(\mathrm{E}\left[e^{-\alpha U_n}\right]\right)^2\right) = {\delta_n^*}^2 n\left(b_{n,2\alpha} - b_{n,\alpha}^2\right)\\
&= {\delta_n^*}^2 \alpha n b_{n,2\alpha} \sum_{j=1}^{n} \frac{b_{j,\alpha}^2}{b_{j,2\alpha}} \frac{1}{j(j+2\alpha)} = O(n^{-2\alpha+1}),\\
\mathrm{E}\left[\sigma_n^2\right] &= n\Bigg(n^{-1}(1-\kappa_n) - (1-\kappa_n)\mathrm{E}\left[e^{-2\alpha U_n}\right] + (1-n^{-1})(1-\kappa_n)\mathrm{E}\left[e^{-2\alpha\tau_{ij}^{(n)}}\right]\\
&\quad + n^{-1} v_n \mathrm{E}\left[\sum_{k=2}^{\Upsilon+1} J_k e^{-2\alpha(t_{\Upsilon+1}+\ldots+t_k)}\right]\\
&\quad + (1-n^{-1}) v_n \mathrm{E}\left[\sum_{k=2}^{\upsilon+1} J_k e^{-2\alpha(\tau^{(n)}+t_{\upsilon+1}+\ldots+t_k)}\right]\Bigg)\\
&= (1-\kappa_n) + (1-\kappa_n)\frac{2}{2\alpha-1} + \kappa_n\frac{2}{2\alpha} + \kappa_n\frac{4}{2\alpha(2\alpha-1)} + O(n^{-2\alpha+1})\\
&= (1-\kappa_n)\frac{2\alpha+1}{2\alpha-1} + \kappa_n\frac{2(2\alpha+2)}{2\alpha(2\alpha-1)} + O(n^{-2\alpha+1}).
\end{aligned}
$$

We now write out, where ε is an appropriate constant [see Lemma 3.1 and Lemma 11 in 6]

$$
\begin{aligned}
&\mathrm{Var}\left[\sigma_n^2\right] = n^2 \mathrm{Var}\left[\mathrm{Var}\left[\overline{Y}_n | \mathscr{Y}_n^*\right]\right] = n^{-2}\mathrm{Var}\left[\mathrm{Var}\left[Y_1+\ldots+Y_n|\mathscr{Y}_n^*\right]\right]\\
&= n^{-2}\mathrm{Var}\Bigg[n(1-\kappa_n) - n^2(1-\kappa_n)e^{-2\alpha U_n} + n(n-1)(1-\kappa_n)\mathrm{E}\left[e^{-2\alpha\tau_{ij}^{(n)}}|\mathscr{Y}_n^*\right]\\
&+ n v_n \mathrm{E}\left[\sum_{k=2}^{\Upsilon+1} J_k e^{-2\alpha(t_{\Upsilon+1}+\ldots+t_k)}|\mathscr{Y}_n^*\right] + n(n-1)v_n\mathrm{E}\left[\sum_{k=2}^{\upsilon+1} J_k e^{-2\alpha(\tau^{(n)}+t_{\upsilon+1}+\ldots+t_k)}|\mathscr{Y}_n^*\right]\Bigg]\\
&= n^{-2}\mathrm{Var}\Bigg[-n^2(1-\kappa_n)e^{-2\alpha U_n} + n(n-1)(1-\kappa_n)\mathrm{E}\left[e^{-2\alpha\tau_{ij}^{(n)}}|\mathscr{Y}_n^*\right]\\
&+ n v_n \mathrm{E}\left[\sum_{k=2}^{\Upsilon+1} J_k e^{-2\alpha(t_{\Upsilon+1}+\ldots+t_k)}|\mathscr{Y}_n^*\right] + n(n-1)v_n\mathrm{E}\left[\sum_{k=2}^{\upsilon+1} J_k e^{-2\alpha(\tau^{(n)}+t_{\upsilon+1}+\ldots+t_k)}|\mathscr{Y}_n^*\right]\Bigg]\\
&\leq 4n^{-2}\Bigg(n^4(1-\kappa_n)^2\mathrm{Var}\left[e^{-2\alpha U_n}\right] + n^2(n-1)^2(1-\kappa_n)^2\mathrm{Var}\left[\mathrm{E}\left[e^{-2\alpha\tau_{ij}^{(n)}}|\mathscr{Y}_n^*\right]\right]\\
&+ n^2 v_n^2\mathrm{Var}\left[\mathrm{E}\left[\sum_{k=2}^{\Upsilon+1} J_k e^{-2\alpha(t_{\Upsilon+1}+\ldots+t_k)}|\mathscr{Y}_n^*\right]\right] + n^2(n-1)^2 v_n^2\mathrm{Var}\left[\mathrm{E}\left[\sum_{k=2}^{\upsilon+1} J_k e^{-2\alpha(\tau^{(n)}+t_{\upsilon+1}+\ldots+t_k)}|\mathscr{Y}_n^*\right]\right]\Bigg)\\
&\leq C\left(n^2 n^{-4\alpha} + n^2 n^{-2-\varepsilon} + \sigma_{c,n}^2\kappa_n n^{-1} + n^2\sigma_{c,n}^2\kappa_n n^{-2}\right) = O(n^{-2(2\alpha-1)} + n^{-\varepsilon} + \sigma_{c,n}^2\kappa_n).
\end{aligned}
$$

Therefore as we assumed $p_n \to 0$ we have $\mathrm{Var}\left[\sigma_n^2\right] \to 0$. Together this implies the desired L^2 and hence in P convergence of $(\mu_n, \sigma_n^2) \to (0, (2\alpha+1)/(2\alpha-1))$.
On the other hand

$$
\begin{aligned}
&\mathrm{Var}\left[\sigma_n^2\right] \geq n^{-2}\mathrm{Var}\left[-n^2(1-\kappa_n)e^{-2\alpha U_n} + n(n-1)v_n\mathrm{E}\left[\sum_{k=2}^{\upsilon+1} J_k e^{-2\alpha(\tau^{(n)}+t_{\upsilon+1}+\ldots+t_k)}|\mathscr{Y}_n^*\right]\right]\\
&\geq \left(1-\frac{1}{\sqrt{2}}\right)n^{-2}\left(n^4(1-\kappa_n)^2\mathrm{Var}\left[e^{-2\alpha U_n}\right] + n^2(n-1)^2 v_y^2\mathrm{Var}\left[\mathrm{E}\left[\sum_{k=2}^{\upsilon+1} J_k e^{-2\alpha(\tau^{(n)}+t_{\upsilon+1}+\ldots+t_k)}|\mathscr{Y}_n^*\right]\right]\right)\\
&= \Omega(n^{-2(2\alpha-1)} + \sigma_{c,n}^2\kappa_n).
\end{aligned}
$$

From these we obtain that the assumption that $\sigma_{c,n}^2\kappa_n \to 0$ is a necessary one for a CLT to hold.

PROOF OF PART ((II))
This is proved in the same way as PART (I) except with the normalizing constant equalling $n\ln^{-1} n$.

PROOF OF PART ((III))
We notice that the martingale $H_n = (n+1)e^{(\alpha+1)U_n}\overline{Y}_n$ has uniformly bounded second moments. Namely by Lemma 4.1 and Cauchy–Schwarz

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

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. Notice that this convergence result does not depend on $\kappa_n \to 0$. In fact in this regime κ_n can be constant (by definition $\kappa_n \le 1$). As in Bartoszek and Sagitov [6] we obtain [Lemma 9 in 6] $n^{\alpha}\overline{Y}_n \to V^{(\alpha-1)}H_\infty$ a.s. as in L^2. We may also obtain directly the first two moments of $n^{\alpha}\overline{Y}_n$,

$$\begin{array}{rcl}
n^{\alpha}\,\mathrm{E}\left[\overline{Y}_n\right] & = & \delta_n^* n^{\alpha} b_{n,\alpha} \to \delta_\infty^*\Gamma(1+\alpha)\\
n^{2\alpha}\,\mathrm{E}\left[\overline{Y}_n^2\right] & = & (1-\kappa_n)n^{2\alpha-1}-(1-\kappa_n-\delta_n^{*^2})n^{2\alpha}b_{n,2\alpha}+n^{2\alpha}(1-n^{-1})(1-\kappa_n)\,\mathrm{E}\left[e^{-2\alpha\tau^{(n)}}\right]\\
& & +n^{2\alpha-1}\nu_n\,\mathrm{E}\left[\Psi^{*^{(n)}}\right]+n^{2\alpha}\nu_n\,\mathrm{E}\left[\Psi^{(n)}\right]\\
& \to & -(1-\kappa_\infty-\delta_\infty^{*^2})\Gamma(2\alpha+1)+(1-\kappa_\infty)\frac{1+2\alpha}{1-2\alpha}\Gamma(1+2\alpha)\\
& = & \left(\delta_\infty^{*^2}+(1-\kappa_\infty)\frac{4\alpha}{1-2\alpha}\right)\Gamma(1+2\alpha).
\end{array}$$

Notice that again in this regime the convergence of the moments does not depend on how κ_n behaves with n.

Acknowledgments

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

[9] 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.

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

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

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

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

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

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

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

[17] 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.

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

[19] 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.

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

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

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

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