



Jugowice, 11<sup>th</sup>–15<sup>th</sup> September 2017

## TRAIT EVOLUTION WITH JUMPS: ILLUSIONARY NORMALITY

Krzysztof Bartoszek<sup>1</sup>

<sup>1</sup>Department of Computer and Information Science, Linköping University,  
Linköping 581 83, Sweden,  
<sup>1</sup>bartoszekkj@gmail.com

### ABSTRACT

Phylogenetic comparative methods for real-valued traits usually make use of stochastic process whose trajectories are continuous. This is despite biological intuition that evolution is rather punctuated than gradual. On the other hand, there has been a number of recent proposals of evolutionary models with jump components. However, as we are only beginning to understand the behaviour of branching Ornstein–Uhlenbeck (OU) processes the asymptotics of branching OU processes with jumps is an even greater unknown. In this work we build up on a previous study concerning OU with jumps evolution on a pure birth tree. We introduce an extinction component and explore via simulations, its effects on the weak convergence of such a process. We furthermore, also use this work to illustrate the simulation and graphic generation possibilities of the mvSLOUCH package.

### INTRODUCTION

Contemporary stochastic differential equation (SDE) models of continuous trait evolution are focused around the Ornstein–Uhlenbeck process [15]

$$dX(t) = -\alpha(X(t) - \theta(t))dt + \sigma_a dB(t), \quad (1)$$

where  $\theta(t)$  can be piecewise linear. These models take into account the phylogenetic structure between the contemporary species. The trait follows the SDE 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.

However, the fossil record indicates [10, 13, 14] that change is not as gradual as Eq. (1) suggests. Rather, that jumps occur and that the framework of Lévy processes could be more appropriate. There has been some work in the direction of phylogenetic Laplace motion [2, 9, 16] and jumps at speciation points [3, 4, 7, 8]. In this work we describe recent asymptotic results on such a jump model (developed in [4]) and develop them by including an extinction component.

It is worth pointing out that OU with jumps (OUj) models are actually very attractive from a biological point of view. They seem to capture a key idea behind the theory of punctuated equilibrium (i.e. the theory of evolution with jumps [13]). At a branching event something dramatic (the jump) could have occurred that drove species apart. But then “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.” [17]. Therefore, between branching events (and jumps) we could expect stasis—“fluctuations of little or no accumulated consequence”

taking place [14]. This fits well with an OUj model. If  $\alpha$  is large enough, then the process approaches its stationary distribution rapidly and the stationary oscillations around the (constant) mean can be interpreted as stasis between jumps.

In applications the phylogeny is given (from molecular sequences) but when the aim is to study large sample properties some model of growth has to be assumed. A typical one is the constant rate birth–death process conditioned on  $n$  contemporary tips (cf. [11, 12, 19]).

Regarding the phenotype the Yule–Ornstein–Uhlenbeck with jumps (YOUj) model was recently considered [4]. The phylogeny is a pure birth process (no extinction) and the trait follows an OU process. However, just after the  $k$ -th branching point ( $k = 1, \dots, n - 1$ , counting from the root with the root being the first branching point) on the phylogeny, with a probability  $p_k$ , independently on each daughter lineage, a jump can occur. The jump is assumed to be normally distributed with mean 0 and variance  $\sigma_{c,k}^2 < \infty$ . Just after a speciation event at time  $t$ , independently for each daughter lineage, the trait value  $X(t)$  will be

$$X^+(t) = (1 - Z)X(t^-) + Z(X(t^-) + Y). \quad (2)$$

In the above Eq. (2)  $X(t^{-/+})$  means the value of  $X(t)$  respectively just before and after time  $t$ ,  $Z$  is a binary random variable with probability  $p_k$  of being 1 (i.e. jump occurs) and  $Y \sim \mathcal{N}(0, \sigma_{c,k}^2)$ .

### THE PURE BIRTH CASE

In the case of the pure birth tree central limit theorems (CLTs) can be explicitly derived thanks to a key property of the pure birth process. The time between speciation events  $k$  and  $k + 1$  is exponential with parameter  $\lambda k$  due to the memoryless property of the process and the law of the minimum of i.i.d. exponential random variables. This allows us to calculate Laplace transforms of relevant speciation times and count speciation events on lineages [3, 4, 6, 19].

We first remind the reader about the mathematical concept of sequence convergence with density 1 (see e.g. [18]) and then summarize the previously obtained CLTs.

**Definition 1.** A subset  $E \subset \mathbb{N}$  of positive integers is said to have density 0 if

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

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

**Definition 2.** A sequence  $a_n$  converges to 0 with density 1 if there exists a subset  $E \subset \mathbb{N}$  of density 0 such that

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

Let  $\bar{Y}_n = (\bar{X}_n - \theta) / \sqrt{\sigma_a^2 / 2\alpha}$  be the normalized sample mean of the YOUj process with  $\bar{Y}_0 = \delta$ . Denote by  $\mathcal{Y}_n$  the  $\sigma$ -algebra containing information on the tree and jump pattern ( $\bar{Y}_n$  conditional on  $\mathcal{Y}_n$  is normal). Assume also that  $\lambda = 1$ . The restriction of  $\lambda = 1$  is a mild one, changing  $\lambda$  is equivalent to rescaling time.

**Theorem 1** ([4]). Assume that the jump probabilities and jump variances are constant equalling  $p$  and  $\sigma_c^2 < \infty$  respectively. The process  $\bar{Y}_n$  has the following, depending on  $\alpha$ , asymptotic with  $n$  behaviour:

(I) If  $0.5 < \alpha$ , then the conditional variance of the scaled sample mean  $\sigma_n^2 := n \text{Var} [\bar{Y}_n | \mathcal{Y}_n]$  converges in  $\mathbb{P}$  to a random variable  $\sigma_\infty^2$  with mean

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

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

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

- (II) If  $0.5 = \alpha$ , then the conditional variance of the scaled sample mean  $\sigma_n^2 := n \ln^{-1} n \text{Var} [\bar{Y}_n | \mathcal{Y}_n]$  converges in  $\mathbb{P}$  to a random variable  $\sigma_\infty^2$  with mean

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

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

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

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

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

**Theorem 2** ([4]). If  $\sigma_{c,k}^4 p_k$  is bounded and goes to 0 with density 1, then depending on  $\alpha$  the process  $\bar{Y}_n$  has the following asymptotic with  $n$  behaviour.

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

**Remark 1.** Notice that Thm. 2 immediately implies the CLTs when there are no jumps, i.e.  $p_k = 0$  for all  $k$  [6].

## EXTINCTION PRESENT

The no extinction assumption is difficult to defend biologically, unless one considers extremely young clades. Therefore, it is desirable to generalize Thms. 1 and 2 to the case when the extinction rate,  $\mu$ , is non-zero. However, there are a number of intrinsic difficulties associated with such a generalization. We do not have the Laplace transforms of the time to coalescent of a random pair of tips (its expectation seems involved enough, [19]). More importantly we seem to be unable to say much about the number of hidden speciation events on a random lineage. We have to remember that a jump can be due to any speciation event, including those that lead to extinct lineages. A lineage that survived till today can have multiple branches stemming from it that faded away in the past, see Fig. 1.

Furthermore, for the OU model of trait evolution, we need to know the distribution (or Laplace transform) of the times between the speciation events. Without extinction,  $\mu = 0$ , this was simple. The time between speciation events  $k$  and  $k + 1$  was exponential with rate  $\lambda k$ , as the minimum of  $k$  independent rate  $\lambda$  exponential random variables. However, when  $\mu > 0$  we not only need to know the number of hidden speciation events between two non-hidden (i.e. leading to contemporary tips) speciation events but also the law of the time between the hidden speciation. We do not know this law and furthermore as we are conditioning on  $n$  it is not entirely clear if times between speciation events will be independent (like they are in the pure birth case).

Our question is whether we can expect counterparts of Thms. 1 and 2 to hold when extinction is present, i.e.  $0 < \mu \leq 1 = \lambda$ . We are not aware of analytical results on the issues raised in the previous paragraph. Hence, we will approach answering the problem by simulations. Based on the results in [1] we should expect a phase transition to occur for  $1 - \mu = \alpha/2$  (remember  $\lambda = 1$ ).

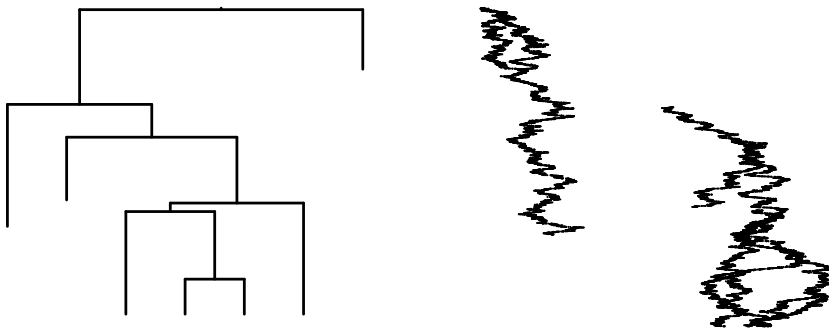


Figure 1. Left: birth–death tree with a clade of 4 contemporary species. Right: OU process evolving on this tree (graphic by mvSLOUCH). There is a single jump in the trait process at the first speciation event after the root. The OU process is a slowly adapting one with large jump variance ( $\alpha = 0.3$ ,  $\sigma_a^2 = \sigma_c^2 = 4$ , tree height: 2.427).

We simulate 1000 birth–death trees for  $\mu = 0.25, 0.5, 0.75$  conditioned on  $n = 500$  contemporary tip species with the TreeSim [20, 21] R package. On each phylogeny we simulate an OU process with  $\alpha = (1 - \mu)$ ,  $(1 - \mu)/2$ ,  $(1 - \mu)/4$  using the mvSLOUCH R package [5]. In all simulations  $\sigma_a^2 = \sigma_c^2 = 1$ ,  $p_k = 0.5$  (for all internal nodes) and  $X_0 = \theta = 0$ . For a given phylogeny, OU simulation pair we calculate the scaled sample average,

$$\bar{V}_n = (\sqrt{n})\bar{X}_n / \sqrt{\sigma_a^2 / (2\alpha)}. \quad (3)$$

We report the results of the simulations by plotting histograms with a mean 0 and variance equalling the scaled sample variance normal curve.

Table 1. Summary of simulated samples. The mean, variance, skewness and excess kurtosis refer to the sample moments of  $\bar{V}_n$ . The 95% bootstrap confidence intervals for the excess kurtosis are based on 10000 bootstrap replicates (of the  $\bar{V}_n$  values) and calculated by the R package boot (basic bootstrap confidence intervals are reported).

$\mu$	$\alpha$	$\alpha/(1 - \mu)$	mean	variance	skewness	excess kurtosis
0.25	0.75	1	-0.117	7.173	0.114	0.046 (-0.195, 0.27)
0.5	0.5	1	0.116	10.016	-0.084	0.280 (-0.297, 0.803)
0.75	0.25	1	-0.082	17.158	0.058	-0.058 (-0.318, 0.186)
0.25	0.375	0.5	0.344	24.227	0.038	0.110 (-0.185, 0.392)
0.5	0.25	0.5	-0.141	34.432	0.012	0.121 (-0.186, 0.412)
0.75	0.125	0.5	-0.056	52.357	0.087	0.3 (-0.064, 0.646)
0.25	0.1875	0.25	0.212	77.016	0.072	0.316 (-0.107, 0.722)
0.5	0.125	0.25	0.575	88.128	0.102	0.149 (-0.14, 0.416)
0.75	0.0625	0.25	0.3	114.806	-0.024	1.169 (0.328, 1.964)

### ILLUSIONARY NORMALITY?

We report the result of our simulations in Fig. 2 and Tab. 1. The histograms are not conclusive but they can be interpreted as indicating a similar as in Thms. 1 and 2 phase transition. In the fast adaptation case,  $\alpha = (1 - \mu) > (1 - \mu)/2$ , the histograms do not seem to deviate much from the normal curve. On the other hand, it is a bit surprising that the “worst” looking histogram is

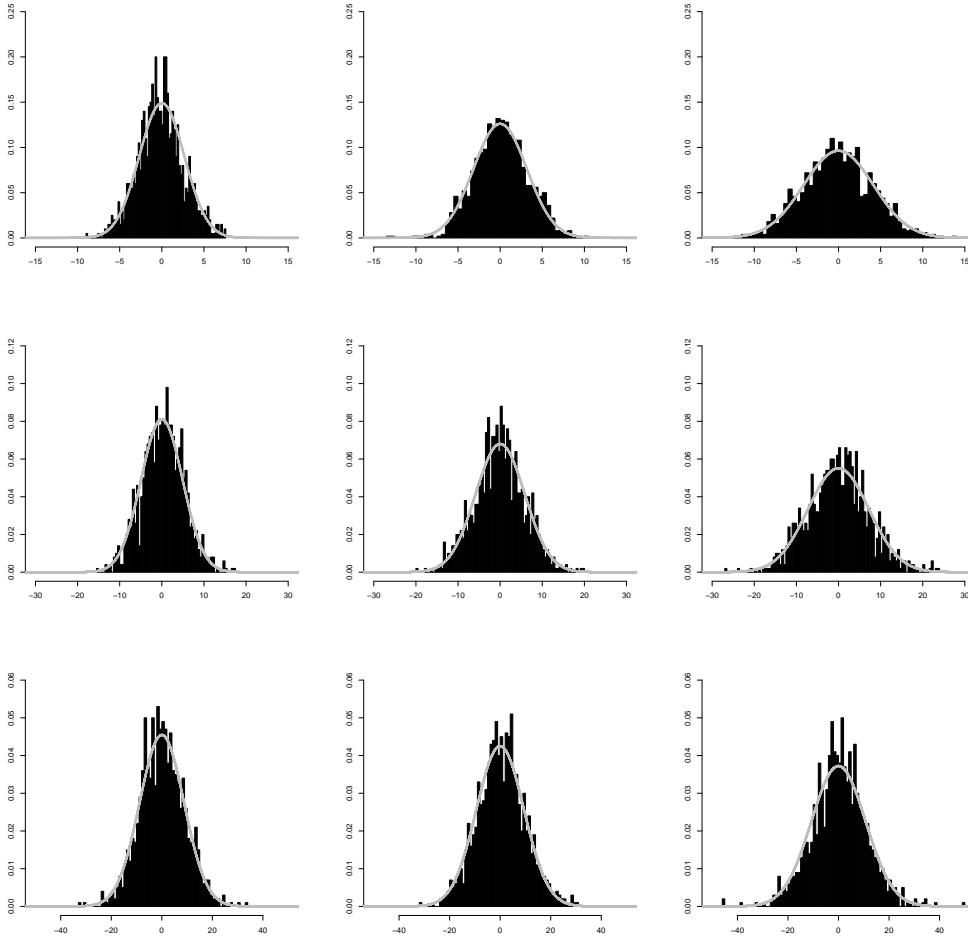


Figure 2. Simulations of Eq. (3). Left column  $\mu = 0.25$ , centre column  $\mu = 0.5$ , right column  $\mu = 0.75$ , top row  $\alpha = (1 - \mu)$ , centre row  $\alpha = (1 - \mu)/2$  and bottom row  $\alpha = (1 - \mu)/4$ . The gray curve is the density curve of the normal distribution with mean 0 and variance equalling the sample variance. Each histogram is constructed from 1000 simulated birth–death trees (birth rate 1, death rate  $\mu$ ) with an OUj process ( $\theta = X_0 = 0$ ,  $\sigma_a^2 = \sigma_c^2 = 1$ ) evolving on top of the tree. Notice that the  $x$  and  $y$  axes differ between rows.

when  $\mu = 0.25$ , the one we would expect to be closest to “normality”. As the ratio of  $\alpha$  to  $1 - \mu$  decreases we can see that the histograms of  $\bar{V}_n$  deviate more from the normal curve. Furthermore, when  $\alpha = (1 - \mu)/4 < (1 - \mu)/2$  we can start to see heavier than normal tails.

The analysis of the first four moments in Tab. 1 does point to two things. Firstly, the scaling by  $(\sqrt{n})/(\sqrt{\sigma_a^2/(2\alpha)})$  is incorrect (cf. Thm. 3.3 [1]). In all setups the sample variance of  $\bar{V}_n$  is much greater than 1. However, based on the estimates of skewness and excess kurtosis we cannot reject normality outright. Only in the most extreme case ( $\mu = 0.75$ ,  $\alpha/(1 - \mu) = 0.25$ ) do the 95% bootstrap confidence intervals of the excess kurtosis not cover 0. When  $\alpha = (1 - \mu)/2$  we should actually expect to have a logarithmic correction in the scaling,  $\sqrt{n \ln n}$ , however we present here the histograms without it. Such a logarithmic correction did not bring in any qualitative changes to the results and hence, for easiness of comparison in Tab. 1 and Fig. 2 we refrain from using it.

Based on the histograms and analysis of the first four moments we would not suspect that with a constant jump probability  $p$  we do not have a nearly classical CLT, i.e. weak convergence to a normal limit after scaling by  $\sqrt{(n)}$ . All that we would think would remain, would be to find the correct variance of the limit. In fact if we look at Fig. 2 in [4] we would be under a similar illusion. The histogram for  $\alpha = 1$ ,  $\lambda = 1$ ,  $\mu = 0$ ,  $p = 0.5$  and  $\sigma_c^2 = 1$  nearly perfectly fits into a normal density curve. However, Thms. 1 and 2 are very clear that for a constant product of  $p_k \sigma_{c,k}^4$  we will not have a normal limit. Therefore, with  $\mu > 0$  we cannot expect a change of the situation. On the one hand a non-zero extinction rate does cause the (conditional on  $n$  contemporary tips) tree to be higher. But on the other hand, with greater height comes to opportunity for more jumps. And this variability of jump occurrences seems to be the force pushing the limit away from normality.

The simulations and results presented here and in [4] should also serve as a warning. Visual inspection of histograms from simulated data is never sufficient for drawing conclusions about a model's underlying distribution. A bare minimum are goodness-of-fit tests but their conclusions should be supported by analytical derivations.

### ACKNOWLEDGEMENTS

KB's research was supported by the Knut and Alice Wallenberg Foundation. KB's conference participation was supported by the Wenner-Gren Foundation (grant nr. RSh2016-0078).

### REFERENCES

- [1] R. Adamczak and P. Miłoś: *CLT for Ornstein–Uhlenbeck branching particle system*, *Elect. J. Probab.* **20** (2015), 1-35.
- [2] K. Bartoszek: *The Laplace Motion in Phylogenetic Comparative Methods*, Proceedings of the Eighteenth National Conference on Applications of Mathematics in Biology and Medicine, Krynica Morska, 2012, pp. 25–30.
- [3] ———: *Quantifying the effects of anagenetic and cladogenetic evolution*, *Math. Biosci.* **254** (2014), 42-57.
- [4] ———: *A Central Limit Theorem for punctuated equilibrium*, ArXiv e-prints (2016), 1602.05189.
- [5] K. Bartoszek, J. Pienaar, P. Mostad, S. Andersson, and T. F. Hansen: *A phylogenetic comparative method for studying multivariate adaptation*, *J. Theor. Biol.* **314** (2012), 204-215.
- [6] K. Bartoszek and S. Sagitov: *Phylogenetic confidence intervals for the optimal trait value*, *J. App. Prob.* **52** (2015), 1115-1132.
- [7] F. Bokma: *Testing for equal rates of cladogenesis in diverse taxa*, *Evolution* **57** (2003), 2469-2474.
- [8] ———: *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** (2008), 2718-2726.
- [9] P. Duchon, C. Leuenberger, S. M. Szilágyi, L. Harmon, J. Eastman, M. Schweizer, and D. Wegmann: *Inference of evolutionary jumps in large phylogenies using Lévy processes*, *Syst. Biol.* (in press 2017).
- [10] N. Eldredge and S. J. Gould: *Punctuated equilibria: an alternative to phyletic gradualism*, *Models in Paleobiology* (T. J. M. Schopf and J. M. Thomas, eds.), Freeman Cooper, 1972, pp. 82-115.
- [11] T. Gernhard: *The conditioned reconstructed process*, *J. Theor. Biol.* **253** (2008), 769-778.
- [12] ———: *New Analytic Results for Speciation Times in Neutral Models*, *B. Math. Biol.* **70** (2008), 1082-1097.
- [13] S. J. Gould and N. Eldredge: *Punctuated equilibria: the tempo and mode of evolution reconsidered*, *Paleobiology* **3** (1977), 115-151.
- [14] ———: *Punctuated equilibrium comes of age*, *Nature* **366** (1993), 223-227.
- [15] T. F. Hansen: *Stabilizing selection and the comparative analysis of adaptation*, *Evolution* **51** (1997), 1341-1351.
- [16] 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** (2013), 193-204.
- [17] E. Mayr: *Speciation and macroevolution*, *Evolution* **36** (1982), 1119-1132.
- [18] K Petersen: *Ergodic Theory*, Cambridge University Press, 1983.
- [19] S. Sagitov and K. Bartoszek: *Interspecies correlation for neutrally evolving traits*, *J. Theor. Biol.* **309** (2012), 11-19.
- [20] T. Stadler: *On incomplete sampling under birth-death models and connections to the sampling-based coalescent*, *J. Theor. Biol.* **261** (2009), 58-68.
- [21] ———: *Simulating Trees with a Fixed Number of Extant Species*, *Syst. Biol.* **60** (2011), 676-684.