Theoretical Population Biology 133 (2020) 38–39

Contents lists available at ScienceDirect

Theoretical Population Biology

journal homepage: www.elsevier.com/locate/tpb

Commentary

Lambert and Stadler (2013): A unifying framework for modelling evolutionary trees

Mike Steel

Biomathematics Research Centre, University of Canterbury, Christchurch, New Zealand

ARTICLE INFO

Article history:
Received 8 March 2019
Available online 4 July 2019

Stochastic models of trees generated by birth and death processes are central to modern phylogenetics. Here, 'birth' and 'death' refer to speciation and extinction. The application of these models dates back to a classic 1925 paper by George Udny Yule, who showed how a simple birth model could help explain the long-tailed distribution of the species richness across genera (Yule, 1925). Birth–death processes were further developed in the 1940s (Kendall, 1948); by the 1990s, their phylogenetic relevance came to the fore on a number of fronts: modelling macroevolution (Harvey et al., 1994; Nee et al., 1994), studying gene trees (Rannala, 1997), population dynamics, and as priors for Bayesian phylogenetics.

The growing availability of large phylogenies inferred using molecular data, from the 1990s onward, led to a further opportunity to extend Yule's pioneering work; namely, to use these trees to estimate speciation and extinction rates, and to test different model hypotheses. Slowinski and Guyer (1989) had already noted that the discrete probability distribution on the shape of the *reconstructed tree* (i.e. the evolutionary tree once extinct lineages are ignored) is remarkably robust to model variation. This discrete probability distribution on tree shapes (often called 'Yule-Harding') arises whenever there is a uniform distribution on ranked trees (URT). Later, Aldous (1996) explained how this URT property holds under general exchangeability assumptions on speciation and extinction events, even when the rates of these events change with time and with the number of species present (and, in the case in which extinction is allowed, past history).

With this stage set, Amaury Lambert and Tanja Stadler, in their 2013 paper (Lambert and Stadler, 2013), provided a new and unified way to view and apply macroevolutionary models in phylogenetics. First, their paper explored in more detail the model assumptions under which the URT property describes the shape of the reconstructed tree, making precise what constraints need to be imposed on the speciation and extinction rates. This exploration led to the second main contribution of this paper. Rather than viewing birth–death trees as evolving 'forward in time', the authors showed how the reconstructed tree can be described much more effectively by a simple 'horizontal' procedure called a 'Coalescent Point Process' (CPP).

The basic idea of the CPP is as follows: suppose we wish to sample a reconstructed tree grown for time t under a birth–death process, and conditioned on it having n leaves at time t. The direct (but inefficient) way to do this would be to simulate trees for time t, form the reconstructed tree, and discard all the samples that do not have exactly n leaves. By contrast, for the CPP, one simply draws $n-1$ i.i.d. samples $x_1, x_2, \ldots, x_{n-1}$ from a fixed density f, from which the tree is reconstructed in a geometrically simple way (roughly speaking, by linking vertical line segments of lengths $x_1, x_2, \ldots, x_{n-1}$ to a vertical line of length t by adding linking horizontal line segments). The CPP also applies if we do not condition on having n leaves at time t; one simply stops the process at the first length x_i that is larger than t (which nicely explains why the number of leaves in a tree grown for time t follows a geometric distribution).

Although the CPP notion had been explored earlier (Popovic, 2004; Aldous and Popovic, 2005; Lambert, 2010), the Lambert–Stadler paper provided a general treatment, describing precise conditions under which it would hold, and showing how it gives a more effective means of sampling and performing likelihood calculations. The CPP applies to a wide class of models beyond the simple constant birth–death model (though not quite as wide as the processes that lead to the URT property, since birth/death rates that depend on the number of lineages are problematic for applying the CPP). The CPP should not be confused with Kingman's coalescent process (or its phylogenetic incarnation, the 'multispecies coalescent process'); indeed, Lambert and Stadler showed that the Kingman coalescent *cannot* be exactly described by a CPP, except for a finite number of values of n (they conjecture the only possible value of n for which this occurs is $n = 2$).

One challenge for the future is to explain why real phylogenetic trees tend to be a little less 'balanced' than the URT property predicted by the CPP approach. Aldous described a one-parameter

E-mail address: mike.steel@canterbury.ac.nz.

https://doi.org/10.1016/j.tpb.2019.07.002

model (the β-splitting model), for which the URT property corresponds to $\beta = 0$, whereas real phylogenetic trees tend to be often cluster around $\beta = -1$ (Aldous, 2001). Despite a number of attempts (e.g. Hagen et al. (2015)), a compelling and simple explanation for this phenomenon has so far proved elusive.

The CPP representation holds particular promise for phylodynamics (modelling the shape of evolutionary trees and thereby using reconstructed trees to test macroevolutionary hypotheses). It has already been applied to give an exact description of how 'phylogenetic diversity' is lost in large trees as species become extinct (Lambert and Steel, 2013), and to quantify 'age-dependent' extinction, in which the extinction rate of a species depends on the time since it split from another species (Alexander et al., 2016). An extension of the CPP has provided a new way to quantify protracted evolution (where a new lineage takes time to develop into a separate species) based on maximum likelihood calculations (Lambert et al., 2015), and to predict genetic diversity in branching populations.

References

Aldous, D., 1996. Probability distributions on cladograms. In: Aldous, D., Pemantle, R.E. (Eds.), Random Discrete Structures, IMA Volumes in Mathematics and Its Applications, vol. 76. Springer, pp. 1–18.

Aldous, D.J., 2001. Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. Stat. Sci. 16 (1), 23–34.

Aldous, D., Popovic, L., 2005. A critical branching process model for biodiversity. Adv. Appl. Probab. 37 (4), 1094–1115.

Alexander, H., Lambert, A., Stadler, T., 2016. Quantifying age-dependent extinction from species phylogenies. Syst. Biol. 65 (1), 35–50.

Hagen, O., Hartmann, K., Steel, M., Stadler, T., 2015. Age-dependent speciation explains empirical tree shape distribution. Syst. Biol. 64 (3), 432–440.

Harvey, P.H., May, R.M., Nee, S., 1994. Phylogenies without fossils. Evolution 48 (3), 523–529.

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

Lambert, A., 2010. The contour of splitting trees is a Lévy process. Ann. Probab. 38 (1), 348–395.

Lambert, A., Morlon, H., Etienne, R.S., 2015. The reconstructed tree in the lineage-based model of protracted speciation. J. Math. Biol. 70 (1), 367–397.

Lambert, A., Stadler, T., 2013. Birth-death models and coalescent point processes: The shape and probability of reconstructed phylogenies. Theor. Popul. Biol. 90, 113–128.

Lambert, A., Steel, M., 2013. Predicting the loss of phylogenetic diversity under non-stationary diversification models. J. Theoret. Biol. 337, 111–124.

Nee, S., May, R.M., Harvey, P.H., 1994. The reconstructed evolutionary process. Phil. Trans. R. Soc. B 344, 305–311.

Popovic, L., 2004. Asymptotic genealogy of a critical branching process. Ann. Appl. Probab. 14 (4), 2120–2148.

Rannala, B., 1997. Gene genealogy in a population of variable size. Heredity 78, 417–423.

Slowinski, J.B., Guyer, C., 1989. Testing the stochasticity of patterns of organismal diversity: An improved null model. Amer. Natur. 134, 907–921.

Yule, G.U., 1925. A mathematical theory of evolution: Based on the conclusions of Dr. J.C. Willis F.R.S.. Phil. Trans. R. Soc. B 213, 21–87.