Master thesis proposal
Applying Statistical properties to State Speciation and Extinction models' inference

Krzysztof Bartoszek

October 9, 2025

Background—the SSE models

The field of phylogenetic comparative methods aims at understanding trait (like body size, eye colour, presence/absence of tail, compound/simple eyes, number of chromosomes) evolution on the between species level. From a statistical point of view this implies that the sample (usually some average measurement for each species) cannot be considered to be an independent one. The relationships between the species are described by their phylogeny—a rooted binary tree (in the graph theory sense) that indicates when species diverged from each other. With a given tree one assumes some model for the evolution of the trait, e.g. a multi–state Markov chain, Brownian motion.

Most contemporary methods make a crucial assumption—the evolving trait does not affect the speciation, it only evolves "on top" of the phylogeny. This is of course in stark contrast to any biological intuition—after all something had to differentiate between the species to make them distinct. The Binary state speciation and Extinction (BissE) model [9] as the first (and still is one of the few used) exception to this. BissE models a binary trait (two states e.g. compound, simple eyes denoted as 0 and 1) and has six parameters—the speciation rate in state 0 (λ_0), in state 1 (λ_0), the extinction rate in state 0 (μ_0), in state 1 (μ_0), the transition rate from state 0 to 1 (q_{01}) and the transition rate from state 0 to 1 (q_{10}). The BiSSE model has since then been generalized to many other models, e.g., multi–state models

(MuSSE [5]), with geographical distribution (GeoSSE [6]) or hidden Markov chain models (HiSSE [2], GeoHiSSE [3]).

Thesis project

These SSE (State Speciation and Extinction models) are by now well established in the community. However, they very often only their output is reported, without, e.g., parameter uncertainty. Laws of Large Numbers (LLNs) and Central Limit Theorems (CLTs) have been developed for these models [8] and furthermore, these have been translated into a more applied setting [1, 12]. The aim of the thesis is to take studies (e.g., [4, 7, 10, 11, 13], but there are also many others in the literature) that used the BiSSE, MuSSE, HiSSE (and possibly other variations of these) and for the found parameter estimates derive LLNs, CLTs, and other asymptotic properties (similarly as in [12]). Then, as a next step calculate asymptotic confidence intervals, compare to simulated ones (e.g., parametric bootstrap) for the given sample size and deduce whether the conclusions of the studies are valid or not.

Goals

The below general goals are for an "ideal" thesis. Depending on the student they will be made more specific in the direction of the student's interests.

1. Become familiar with the modelling approach in evolutionary biology.
2. Explore different biological studies that employ these models and the conclusions they draw.
3. Apply the various asymptotic results (e.g., [8, 1, 12]) to the studies and discuss if the conclusions still hold.
4. Compare asymptotics with results (from simulations) for the given sample size and discuss how this relates to Authors' conclusions.

Data

The topic will use data from the literature and also simulated ones.

References

[1] K. Bartoszek, M. Majchrzak, S. Sakowski, A. B. Kubiak-Szeligowska, I. Kaj, and P. Parniewski. Predicting pathogenicity behavior in *E. coli* population through a state dependent model and TRS profiling. *PLOS Comput. Biol.*, 2018. doi: 10.1371/journal.pcbi.1005931.

[2] J. M. Beaulieu and B. C. O'Meara. Detecting hidden diversification shifts in models of trait–dependent speciation and extinction. *Syst. Biol.*, 65(4):583–601, 2016. doi: 10.1093/sysbio/syw022.

[3] D. S. Caetano, B. C. O'Meara, and J. M. Beaulieu. Hidden state models improve state–dependent diversification approaches, including biogeographical models. *Evolution*, 72(11):2308–2324, 2018. doi: 10.1111/evo.13602.

[4] P. Cockx, M. J. Benton, and J. N. Keating. Estimating ancestral states of complex characters: a case study on the evolution of feathers. *Syst. Biol.*, page syaf063, 2025. doi: 10.1093/sysbio/syaf063.

[5] R. G. FitzJohn. **Diversitree**: comparative phylogenetic analyses of diversification in `R`. *Methods Ecol. Evol.*, 3:1084–1092, 2012. doi: 10.1111/j.2041-210X.2012.00234.x.

[6] E. E. Goldberg, L. T. Lancaster, and R. H. Ree. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Syst. Biol.*, 60:451–465, 2011. doi: 10.1093/sysbio/syr046.

[7] B. Igic and J. W. Busch. Is self-fertilization an evolutionary dead end? *New Phytol.*, 198:386–397, 2013.

[8] S. Janson. Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stoch. Proc. Appl.*, 110:177–245, 2020. doi: 10.1016/j.spa.2003.12.002.

[9] W. P. Maddison, P. E. Midford, and S. P. Otto. Estimating a binary character's effect on speciation and extinction. *Syst. Biol.*, 56(5):701–710, 2007.

[10] M. E. Maliska, M. W. Pennell, and B. J. Swalla. Developmental mode influences diversification in ascidians. *Biol. Lett.*, 9:20130068, 2013.

[11] A. Skinner. Rate heterogeneity, ancestral character state reconstruction, and the evolution of limb morphology in lerista (scincidae, squamata). *Syst. Biol.*, 59(6):723–740, 2010. doi: 10.1093/sysbio/syq055.

[12] D. Tahir, I. Kaj, K. Bartoszek, M. Majchrzak, P. Parniewski, and S. Sakowski. Using multitype branching models to analyze bacterial pathogenicity. *Math. Applicanda*, 48(1):59–86, 2020. doi: 10.14708/ma.v48i1.6465.

[13] A. M. Wright, K. M. Lyons, M. C. Brandley, and D. M. Hillis. Which came first: The lizard or the egg? Robustness in phylogenetic reconstruction of ancestral states. *J. Exp. Zool. (Mol. Dev. Evol.)*, 324(B): 504–516, 2015. doi: 10.1093/sysbio/syaf063.