

Statistical Efficiency of Birth-Death Tree Parameter Estimation

Stat 700

Santosh Desai, Matt McAnear, Urvi Mehta

Department of Statistics, University of Michigan

December 3, 2025

The Phylodeep Package

Phylodeep Paper is by J. Voznica et al [1]

- ▶ Parameter estimation is difficult and MLE method can be challenging for larger trees.
- ▶ Major components of the paper to address this
 - ▶ CBLV representation of trees
 - ▶ Pre-trained neural networks for pre-specified tree types
- ▶ The goal is to speed up inference

Usage

- ▶ Input: Phylogenetic tree in Newick format
- ▶ Output: Estimated parameters (e.g., birth rate (λ), death rate (μ), R_0)
- ▶ Given a tree, Phylodeep selects the pre-trained model with the highest probability
- ▶ Outputs actually come with uncertainty estimates from a parametric bootstrap

Maximum Likelihood Estimation

URVI TO FILL IN

Phylodeep Methodology

SANTOSH TO FILL IN

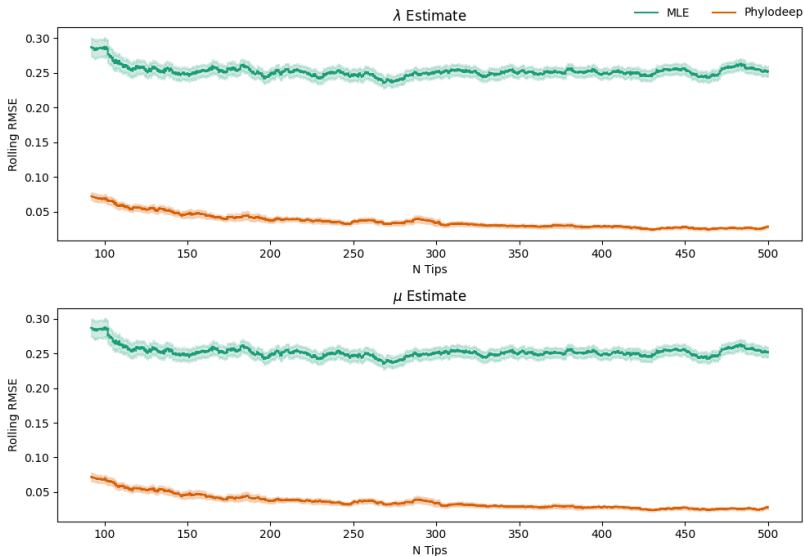


Figure: Rolling bootstrapped RMSE estimates of μ and λ by tree size and resulting 95% CI. Window size is 500.

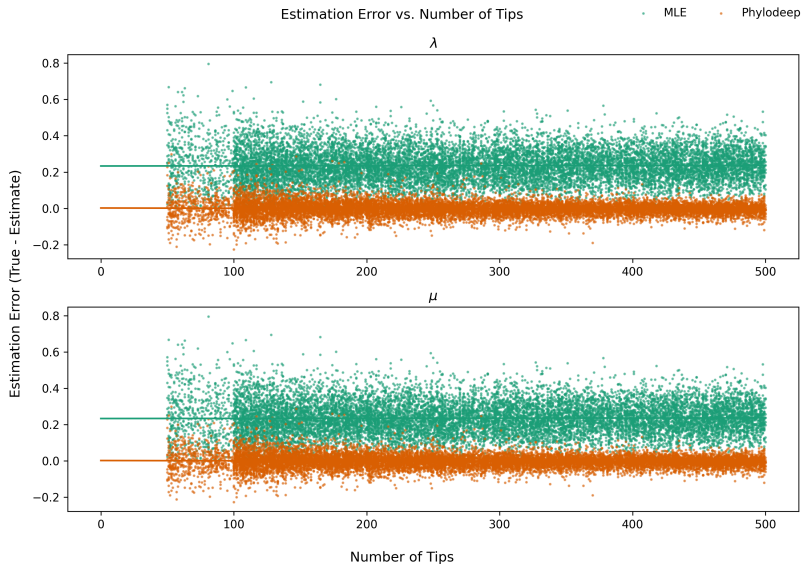


Figure: Raw plot of the absolute errors in λ, μ by tree size.

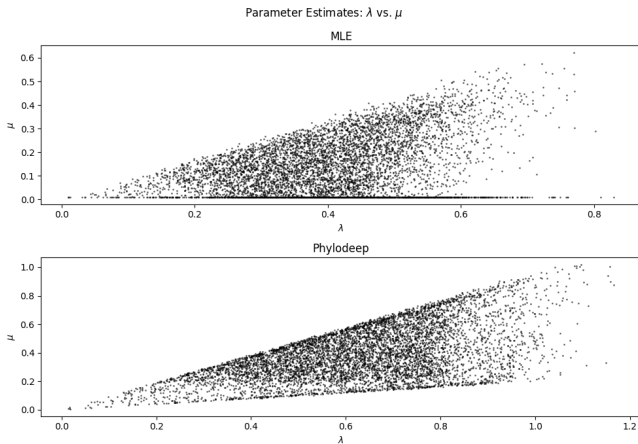


Figure: Plot of λ vs μ estimates to show distribution of estimates.

Main takeaways:

- ▶ RMSE decreases with increasing tree size for both methods.
- ▶ MLE can be used on smaller trees - Phylodeep requires a minimum tip size of around 50.
- ▶ Otherwise, Phylodeep has lower RMSE than MLE, probably due to being trained on 3.9M trees.

References

- [1] J. Voznica et al. “Deep learning from phylogenies to uncover the epidemiological dynamics of outbreaks”. en. In: *Nature Communications* 13.1 (July 2022). Publisher: Nature Publishing Group, p. 3896. ISSN: 2041-1723. DOI: 10.1038/s41467-022-31511-0. URL: <https://www.nature.com/articles/s41467-022-31511-0> (visited on 11/04/2025).