STATS 700-002 Class 11.
Final project suggestions

Aaron King and Edward Ionides

November 6, 2025

Suggestions I

1. Andréoletti et al. (2022). The occurrence birth–death process for combined-evidence analysis in macroevolution and epidemiology. *Systematic Biology* syac037. https://doi.org/10.1093/sysbio/syac037. Summarize the key methodological ideas, compare them with standard birth–death and coalescent approaches, and discuss their implications for macroevolutionary and epidemiological inference. Try reproducing a small example from the paper.
2. Experiment with phylodeep. This package relates to Voznica et al. (2022). Deep learning from phylogenies to uncover the transmission dynamics of epidemics. *Nat Commun* 13:3896 https://doi.org/10.1038/s41467-022-31511-0. Three others proposed studying this paper. In particular, in the new Compact Bijective Ladderized Vector (CBLV) representation developed to facilitate deep learning.

Suggestions II

3. Reproducing results from Ki, C., & Terhorst, J. (2022). Variational phylodynamic inference using pandemic-scale data. *Molecular Biology and Evolution*, 39(8), msac154. https://doi.org/10.1093/molbev/msac154.
4. Reproducing results from King, A. A., Lin, Q., & Ionides, E. L. (2025). Exact phylodynamic likelihood via structured Markov genealogy processes. *ArXiv:2405.17032*. https://doi.org/10.48550/arXiv.2405.17032.
5. Reproducing results from King, A. A., Lin, Q., & Ionides, E. L. (2022). Markov genealogy processes. *Theoretical Population Biology*, 143, 77-91. https://doi.org/10.1016/j.tpb.2021.11.003.
6. Reproducing results from Vaughan, T. G., & Stadler, T. (2025). Bayesian Phylodynamic Inference of Multitype Population Trajectories Using Genomic Data. *Molecular Biology and Evolution*, 42(6), msaf130. https://doi.org/10.1093/molbev/msaf130.

Suggestions III

7. Read Thompson, A, Liebeskind, BJ, Scully, EJ, & Landis, MJ. (2024). Deep learning and likelihood approaches for viral phylogeography converge on the same answers whether the inference model is right or wrong. *Systematic Biology* 73:183–206. https://doi.org/10.1093/sysbio/syad074. (Suggested three times) Perhaps try modifying the code to assess the effects of model misspecifcation, or varying amounts of data, or details of the BEAST pipeline.

8. Phylogeography. Inferring movement of humans, viruses, or other species from genetic data. Two people proposed this. In addition to (7), a suggested paper is De Maio, N, Wu, CH, O'Reilly, KM, & Wilson, D. (2015). New routes to phylogeography: a Bayesian structured coalescent approximation. *PLOS Genetics* 11:e1005421. https://doi.org/10.1371/journal.pgen.1005421

Suggestions IV

9. Sequential Markov coalescent to address recombination. There are relationships between demographic history of a population, time to most recent common ancestor at a given site in the genome, and linkage between sites. The Terhorst group has investigated such questions. What can be learned by comparison with other phylodynamic tools?

10. Zika as a study in sampling-aware phylodynamics with Nextstrain. Nextstrain build and example data: https://nextstrain.org/zika Repository with inputs and configuration: https://github.com/nextstrain/zika

11. Dengue to look at the role of competing serotypes. Nextstrain dengue portal with serotype-specific views (DENV1–DENV4): https://nextstrain.org/dengue Repository with workflows and curated inputs: https://github.com/nextstrain/dengue

Other brainstorming ideas

12. Explain phylopomp (https://github.com/kingaa/phylopomp) to the class, perhaps in the context of the code for King et al., 2025.
13. Extend or improve the analysis in Peter Yang's honors thesis on Phylodynamic Inference of MERS-CoV Using Structured Markov Genealogy Processes, which has code on GitHub.