THE UNIVERSITY OF CHICAGO | DEPARTMENT OF STATISTICS

Master's Thesis Presentation

Xiaolong Wang
Department of Statistics
The University of Chicago

"Variational Inference for Mixed Infections"

April 29, 2024, at 9:00 AM
Jones 111, 5747 S. Ellis Avenue

Abstract

Mixed infections, where multiple strains coexist within a single host, pose a significant challenge for genomic data analysis. In this paper, we develop a scalable variational inference framework for resolving mixed infections from SNP read count data. We model the underlying strain composition using a latent feature assignment matrix and a feature dictionary and place an Indian Buffet Process (IBP) prior to allow for an unbounded number of latent strains.

To perform inference, we derive a Coordinate Ascent Variational Inference (CAVI) algorithm under both finite and infinite variational approximations. The finite approach provides a tractable approximation via a truncated Beta–Bernoulli model, while the infinite approach leverages a stick-breaking construction to more flexibly capture latent structure. We address intractable expectations in the evidence lower bound (ELBO) using suitable approximations, enabling efficient optimization.

Empirical results on simulated SNP read count data demonstrate that both approaches can recover the dominant latent structure of mixed infections. The finite variational approach achieves fast and stable convergence but is limited by the truncation level, whereas the infinite variational approach provides improved reconstruction quality and better captures fine-scale heterogeneity. Overall, our framework offers an effective and scalable alternative to sampling-based methods for inferring latent strain compositions from genomic data.

For information about building access for persons with disabilities, please contact Keisha Prowoznik at 773.702-0541 or send an email to kprowoznik@uchicago.edu. If you wish to subscribe to our email list, please visit the following website: https://lists.uchicago.edu/web/info/statseminars.