## THE UNIVERSITY OF CHICAGO

Department of Statistics

# DISSERTATION PROPOSAL PRESENTATION

YI LIU

Department of Statistics
The University of Chicago

## Efficient Likelihood Free Bayesian Inference for High Dimensional Problems with Applications to the Study of Transcriptional Control

FRIDAY, April 10, 2020, at 2:00 PM
Zoom Meeting, 5747 S. Ellis Avenue

## ABSTRACT

A central unsolved problem in genomics is understanding how DNA sequence controls the process by which genes get turned on and off. A biologically validated thermodynamic model of this process has been previously proposed. As a preliminary result for this proposal, I have shown that this model can be reformulated as a fully-interpretable neural network. The reformulated model, while highly predictive, while highly predictive, does not allow for statistical inference on its fitted parameters. To analyze the parameters of the model carefully, one would have to develop a set of inference techniques for this class of models where the likelihood function is computationally intractable, if not unknown, in the first place. Approximate Bayesian Computation (ABC) is a natural class of algorithms that comes into mind, but such algorithms are known to be slow and not able to scale up to moderately high dimensions. algorithms to consider for this purpose, but such algorithms are known to be slow and not able to scale up to moderately high dimensions. As a further preliminary result, I have developed a class of ABC algorithms for variable selection in a non-parametric setting and, non-parametric setting and, in an unpublished work, further demonstrated that it is possible to speed up this ABC algorithm using Thompson Sampling. The application of the new algorithms I have devised to the transcription model I have reimplemented has the potentially profound implications for both Bayesian computation and system biology.

In this presentation, I will start with introductions to ABC, recent progress of ABC variable selection, and Thompson sampling. Then I will switch gears to discuss the biological problem of transcriptional control of the *even-skipped* gene in embryos of the fruit fly Drosophila, and how this process can be effectively modeled by a Deep Neural Network. Finally, I will explain how methods for ABC can be applied to this class of models to conduct statistical inference.