

THE UNIVERSITY OF CHICAGO

MONTE CARLO METHODS AND DEEP VARIATIONAL INFERENCE FOR FAST
BAYESIAN ANALYSIS OF HIGH DIMENSIONAL BIOLOGICAL SYSTEMS

A THESIS SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

BY
CLAYTON W. SEITZ

CHICAGO, ILLINOIS
SPRING 20XX

Copyright © 2022 by Clayton W. Seitz
All Rights Reserved

TABLE OF CONTENTS

ABSTRACT	iv
1 PRIMER ON EXACT BAYESIAN METHODS AND VARIATIONAL INFERENCE	1
1.1 Markov Chain Monte Carlo	1
1.1.1 Metropolis-Hastings and Gibbs sampling	1
1.1.2 Langevin Monte Carlo	1
1.1.3 Hamiltonian Monte Carlo	1
1.1.4 Stochastic Gradient Langevin Dynamics	1
1.2 Variational Inference	1
1.2.1 Neural networks represent probability distributions	1
1.2.2 Training criteria for neural networks	1
1.2.3 The evidence lower bound	1
2 DECONVOLVING IMMUNOGENIC TUMOR SUBSTRUCTURE WITH VARIATIONAL INFERENCE	2
3 A BAYESIAN APPROACH FOR INFERRING NEURONAL CONNECTIVITY FROM CA2+ IMAGING DATA AND MONTE CARLO SIMULATIONS	3
4 BAYESIAN INFERENCE OF THE KINETIC PARAMETERS OF INTERFERON-GAMMA INDUCED TRANSCRIPTION	4

ABSTRACT

CHAPTER 1

PRIMER ON EXACT BAYESIAN METHODS AND VARIATIONAL INFERENCE

1.1 Markov Chain Monte Carlo

The invention of fast digital computers gives us the ability to simulate random processes and perform inference

1.1.1 Metropolis-Hastings and Gibbs sampling

1.1.2 Langevin Monte Carlo

The metropolis adjusted Langevin algorithm (MALA) aka Langevin Monte Carlo

1.1.3 Hamiltonian Monte Carlo

1.1.4 Stochastic Gradient Langevin Dynamics

1.2 Variational Inference

1.2.1 Neural networks represent probability distributions

1.2.2 Training criteria for neural networks

1.2.3 The evidence lower bound

CHAPTER 2

DECONVOLVING IMMUNOGENIC TUMOR SUBSTRUCTURE WITH VARIATIONAL INFERENCE

Single cell variational inference. Using deep generative modeling for batch correction, differential expression, imputation, etc. Establish how that latent structure relates to immunogenicity. In particular, I am interested in the relationship between tumor substructure i.e. heterogeneity and how it relates to the degree of T-cell inflammation. We can be confident that certain malignant clusters are more immunogenic than others, which, to a first approximation, can be understood by measuring T-cell quantity in the sample. This *could* then be supplemented by using T-cell signatures documented in the literature, differential expression, and clustering methods/mixture models to design fluorescent biomarkers. These will then be added to T-cell markers. In principle, this would give a more complete association of gene expression to the inflammatory state, and would be a useful method in trying to understand changes during the evolution of cancer and following pharmacological treatment.

to For the implementation, I will use Seurat for preprocessing and visualization, as these plotting functions are more visually appealing than Python functions. We can port the preprocessed data over to Python for the deep learning step. Malignant cells are isolated first using standard methods e.g, filtering, normalization, and UMAP clustering For training, I will try to place emphasis on physically inspired algorithms for optimization in variational methods

CHAPTER 3

**A BAYESIAN APPROACH FOR INFERRING NEURONAL
CONNECTIVITY FROM Ca^{2+} IMAGING DATA AND
MONTE CARLO SIMULATIONS**

CHAPTER 4

**BAYESIAN INFERENCE OF THE KINETIC PARAMETERS
OF INTERFERON-GAMMA INDUCED TRANSCRIPTION**