Chapter 1

Introduction

When faced with difficult problems, randomized approximations are a tool commonly employed by quantitative scientists which can offer tractable analysis and algorithms. In this body of work developed during the author's PhD and collected in this thesis, randomized methods are a unifying motif and and used in the first section to construct and analyze tractable approximations and later as the central object of study in density estimation, Monte-Carlo Markov Chain (MCMC), and variational inference (VI). As computational statisticians, the applications considered here sit at the intersection of randomized algorithms (DPPs, MCMC, VI) and Bayesian statistics (double descent, experimental design, Bayesian inference).

The first section uses randomized approximations in order to make closed-form non-asymptotic analysis tractable. In **??**, we consider approximately optimal Bayesian experimental design using an adaptive row sampling algorithm based on DPPs and show that it provides good approximations. Through generalizing the previous chapter's proof techniques, in **??** an extension of the DPP-based design is analyzed in closed-form for the over-parameterized $n < d$ regime and predicts a double-descent phenomenon in linear regression which closely matches empirical experiments. In **??** we isolate the part of the proof involving concentration of bilinear forms of matrix resolvents away from the DPP-based design in order to obtain bounds on expected projections $X_S^\dagger X_S$ when $X_S = SX$ is obtained by sub-Gaussian sketching matrix S. These chapters correspond to the following publications:

- **bayesian-experimental-design**
- **surrogate-design**
- **precise-expressions**

The second three chapters concern the setting of probabilistic programming, a computational tool for specifying probabilistic models and automating inference through MCMC and VI. In particular, we study approximating a target density $p(x)$ with approximations $q(x)$ for the purposes of importance sampling and variational inference. In **??**, we consider parameterizing $q(x)$ by graph neural networks which condition each node on its Markov

blanket. This reduces the conditioning sets for a node, resulting in improvements over [**le2017inference**] and run-times which depend on sparsity in the graphical model rather than the length of execution traces. When the target $p(x)$ is both multivariate and heavy tailed, **??** considers the problem of tail anisotropy through both a theoretical and practical perspective. We establish that prior fat-tailed estimators [**jaini2020tails**] are tail isotropic, propose an anisotropic approximation (fat-tailed variational inference, FTVI) where an anisotropic product base measures is pushed forwards through a bijective neural network, and confirm that in practice FTVI improves both density estimation as well as variational inference. However, we find in practice FTVI and other tail-adaptive approximations often have trouble optimizing the tail parameter. In **??**, we consider addressing this issue during static analysis of a probabilistic program's source code. We define generalized Gamma tail asymptotics for a number of elementary distributions and establish how the tail asymptotics are transformed under algebraic transformations such as sums and products. This enables a priori computation of tail parameters, which we show improves the stability and convergence of a number of inference tasks. These chapters correspond to the following publications:

- **lic**
- **ftvi**
- **gga**