

# STAT548 PH.D. QUALIFYING COURSE PAPERS

Last updated: Sept 2023

I'm generally interested in large-scale, complex, and streaming data, with theoretically rigorous analysis and efficient computation as primary goals. In particular, my research usually falls at the intersection of Bayesian (sometimes nonparametric) modelling, statistical theory, computation, and optimization. I'd also be happy to discuss any other paper you find interesting that you think might fall in my area of expertise.

The first section of the report should provide **a critical analysis** (max 5 pages) of the paper. It should summarize the problem the paper addresses in the context of previous work, limitations of previous work, the solution technique, important results, *why* they are important, and limitations of the paper. Papers listed below may have multiple contribution areas (theory, modelling, computation), and the summary should reflect that. The goal of this portion of the report is to show that you can take a complex body of work (one paper and earlier relevant work), digest it, and present a concise summary of the important points, and comment on them. Do not provide a rote summary of the paper.

The remainder of the report (no page limit) will be devoted to a **small paper-specific project** that we decide on together (come to our first meeting with ideas). Your grade will not be affected by how good the results look, whether your approach improves on past work, or whether you achieve the initial goal of the project. The aim of this exercise, from my perspective, is to evaluate your research potential. This includes taking a problem and formulating it precisely, whiteboarding some ideas with me and on your own, thinking creatively and independently to develop a solution, asking questions when necessary, selecting ways to investigate / justify claims, choosing appropriate metrics, reflecting on and communicating results, choosing the next course of action / reformulating the problem when things don't turn out well, thinking of future directions, etc. The report should be typeset in L<sup>A</sup>T<sub>E</sub>X (ask me for my style file, it'll save you time). Any coding tasks should be done in python or Julia.

## Please submit:

- Your PDF report
- Your LaTeX source
- Any computational code you wrote

## Papers:

1. Doucet et al. "Score-Based Diffusion meets Annealed Importance Sampling," NeurIPS 2022.  
**Summary:** Better proposals for AIS via score-based generative modelling  
**URL:** <https://arxiv.org/pdf/2208.07698.pdf>
2. Arya et al. "Differentiating Metropolis–Hastings to optimize intractable densities," ICML 2023 Differentiating Almost Everything Workshop  
**Summary:** Differentiating expectations estimated via nondifferentiable MCMC schemes  
**URL:** <https://arxiv.org/pdf/2306.07961.pdf>
3. Biswas et al. "Estimating convergence of Markov chains with L-Lag couplings," NeurIPS 2019.  
**Summary:** Bounds on total variation/Wasserstein convergence of MCMC chains using coupling  
**URL:** <https://arxiv.org/abs/1905.09971>
4. Gabrié et al. "Adaptive Monte Carlo augmented with normalizing flows," PNAS 2022  
**Summary:** Improved MCMC proposals with variational flows  
**URL:** <https://arxiv.org/abs/2105.12603>