Statistics Research Statement

Austin Brown
Postdoctoral Fellow, University of Toronto

July 13, 2024

With the rapidly increasing size and complexity of modern data sets, understanding the computational strengths and limitations of statistical algorithms is becoming an increasingly important field. Markov chain Monte Carlo (MCMC) algorithms are widely utilized in Bayesian statistics, but their computational requirements remain an important problem for practitioners. From a high-level perspective, my primary research focuses on the computational efficiency and limitations of MCMC algorithms. In particular, I focus on the computational requirements of Metropolis-Hastings algorithms in high dimensions and large data regimes.

My dissertation at the University of Minnesota consists of three research projects. My first research work considers the question: can the geometric convergence rate of Metropolis-Hastings algorithms be improved upon? In many situations, traditional analysis of MCMC algorithms with the total variation distance has proven difficult to scale to high dimensions. Modern analyses of MCMC algorithms in high dimensions consider Wasserstein distances from the theory of optimal transportation due to their potential to exhibit improved dimension scaling in some situations.

In this first research work, I focus on the simplified setting of the Metropolis-Hastings independence sampler. I show the exact geometric convergence rate in total variation is identical in many Wasserstein distances. In short, the optimal geometric convergence rate in total variation for the Metropolis-Hastings independence sampler cannot be improved upon. However, I combine ideas from optimization theory to construct a centered independent proposal for Metropolis-Hastings algorithms. Using this proposal, I develop explicit upper bounds on the convergence rate in a large dimension and sample size regime for Bayesian logistic regression. In this case, informative convergence rates are developed for practitioners which can scale to large problem sizes when the convergence rates are exact.

In my second research work, I look to aid practitioners in answering the question: how does one properly tune Metropolis-Hastings algorithms in order to avoid poor empirical performance? Metropolis-Hastings algorithms often require choosing tuning parameters, which can be an arduous task in applications. Practitioners are often left tuning these algorithms by trial and error to avoid poor empirical performance. For example, the popular random-walk and Metropolis-adjusted Langevin algorithms require carefully choosing a variance parameter in the proposal distribution. In this work, I develop general lower bounds on the geometric convergence rates of Metropolis-Hastings algorithms to study their computational complexity. If the target density concentrates with the number of data samples, I show the convergence rate can deteriorate exponentially fast if the tuning parameters are not chosen carefully in terms of both the number of data samples and the dimension of the problem. These lower bounds can show practitioners which tuning parameters to avoid, and examples are applied to Bayesian logistic regression with Zellner's g-priors and flat priors.

My third research work looks at a separate topic in Bayesian error-in-variable models. Many modeling applications involve errors in the variables (EIV) which classical linear regression does not take into account. EIV can occur in many situations such as measurement error in data collection, a discrepancy between the data distribution and the model, or purposeful adversarial attacks against the data. I consider Bayesian EIV linear regression accounting for additional additive Gaussian error in the covariates and response. Although Gibbs samplers are used in practice for these models, the reliability of estimation using these Gibbs samplers is not understood. In this work, I construct a Gibbs sampler for Bayesian error-in-variables regression and prove this Gibbs sampler converges at a geometric rate to ensure a central limit theorem for time averages from the Markov chain.

My current research at the University of Warwick continues in the same area as my primary research focus. For Bayesian linear mixed models, the likelihood involves an integral which cannot be evaluated and creates difficulty in Metropolis-Hastings algorithms. Pseudo-marginal Metropolis-Hastings algorithms provide a solution to this problem using an unbiased estimator of the intractable likelihood. Currently, I am focusing on the computational efficiency of certain pseudo-marginal MCMC algorithms with Krzysztof Latuszynski and Gareth Roberts.

My future research plan is to divide my research into two main areas. First, I plan to continue my primary research topic in studying the computational efficiency and limitations of MCMC and other sampling algorithms. Secondly, I plan to continue research in error-in-variable models which are of practical importance in areas such as epidemiology. Constructing new algorithms for error-in-variable models and analyzing real data sets would be great for undergraduates to be involved with my research. This may also be a fruitful and impactful research topic to pursue as an opportunity for external funding and collaboration with the other departments. I am fortunate the University of Minnesota and the University of Warwick provided me with the opportunity to collaborate with many talented faculty members, and I look forward to the opportunity of collaborating as a faculty member myself in the future.

Statistics Teaching Philosophy Statement

Austin Brown
Postdoctoral Fellow, University of Toronto

July 13, 2024

One of the things that captivated me while learning statistics as an undergraduate was a rather elementary example of linear regression applied to real-world data. At the time, I was a student in a regression analysis course being introduced to a more abstract linear regression model in statistics. During the lecture, I followed the mathematical reasoning and symbols of the regression model, but it was when a tangible example was discussed that really sparked my interest in statistics. The simple yet concrete example motivated the abstract statistical concept and helped me grasp the vast applicability of statistics.

As an instructor, my core teaching principle is based on what captivated me as an undergraduate, that is, motivating abstract statistical ideas and concepts through relevant examples. In my teaching instruction, I spend a significant effort providing modern, real-world examples. I put applications with real-world data at the forefront of my lectures and focus on ”why” statistics is being taught. For example, if an abstract linear regression model is introduced in mathematical symbols, I would follow this with a straightforward example using data where students can concretely see values in place of the more abstract symbols.

My goal as an educator is to create an interactive and engaging learning environment for students. Throughout my lectures, I continually review students’ knowledge of presented lecture material in an interactive way. I implement this with questions to the class during the lecture. If I formally introduce a confidence interval in lecture, I will then check the student’s understanding of certain key aspects and details of a confidence interval. As students are learning topics for the first time, asking questions and creating an interactive lecture can help maintain the attention of students and also correct my own pacing at which I am moving through the lecture material. This serves as valuable feedback and helps me understand where to improve upon parts of my lectures.

In my student assessments, I consider it beneficial to pursue multiple avenues in assessing students' knowledge throughout a course. Each individual student has their own unique academic strengths and weaknesses. My current strategy for assessing students is through a diverse combination of homework, exams, and a group project. Not every talented student excels at timed exams, so I like to incorporate at least one avenue for each student to excel in a course. When relevant, I include an ongoing group project throughout the course where real-world data is collected and analyzed. The group project is not only used to provide a way for a student to excel outside of timed examinations, but also to prepare students for professional environments. I feel strongly that the ability to analyze real data in a team environment can be a critical career skill outside of the university.

While teaching is rewarding, learning to teach has come with challenges as in my first teaching experience at the University of Minnesota. I taught a course covering multivariate regression and time series and halfway through the semester, the COVID-19 pandemic abruptly shifted the course curriculum to fully remote learning. Rapidly transitioning from an in-person curriculum and lecture format to an online format was a learning experience. However, due to this transition, I also had new opportunities. During this time, I was able to collaborate with existing faculty in shifting the course online where I created new remotely administered exams. In this transition period, it was also an experience for me to notice certain benefits to distance learning such as recorded lectures which may allow more time flexibility for some students.

Teaching, just like research, takes time, attention, and significant effort. As a graduate student, the University of Minnesota provided many opportunities to hone strengths and improve my weaknesses in my teaching ability. I am continually searching new approaches and ideas to improve my ability to teach effectively and inspire students to learn statistics. I had the opportunity to learn from talented faculty and instructors at the University of Minnesota, and I hope to continue to improve my teaching ability as a faculty member in the future.