William Cipolli
55 University Ave.
Hamilton, NY 13346
+1 (203) 848 5643
will@cipolli.com
www.cipolli.com

Research Statement

The purpose of this statement is to describe my multidisciplinary research activities in the five semesters I have been an Assistant Professor at Colgate University. My research interests can be roughly categorized into four main areas: Bayesian nonparametrics, statistical collaboration, multidisciplinary collaboration, and statistical education. My main research line deals with creating more flexible statistical solutions that extend the Bayesian nonparametric literature, placing high importance on efficient computation. My research is guided by an overarching interest in issues that humanists and social scientists can solve by asking questions statisticians can answer. As of January 2019, I am an author on thirteen papers. Seven are published, four are accepted, one is in revision, and one is submitted. I have received one grant and I am currently collaborating on a second.

The organization of this section starts by discussing my publications in scholarly journals and research grants, as well as the venues where my work has been presented and how it is independent of my doctoral thesis. I note my service to the discipline as a conference organizer and journal referee and I conclude by projecting where my research program will take me during the next three years.

Peer-reviewed Publications in Scholarly Journals of Statistics

I now summarize the contributions my publications have made to the statistics literature. I have co-authored five papers that contribute to the statistics and probability literature. Three of these papers are published, and two are accepted. I begin with two papers published with my dissertation advisor, Timothy Hanson – one in *Computational Statistics and Data Analysis* (2016) and one in *Statistics and Computing* (2016). We also have a paper that was accepted at *Advances in Data Analysis and Classification* that stems from an idea explored in my dissertation but was largely developed during my time at Colgate University.

The *Computational Statistics and Data Analysis* (2016) paper proposes a Bayesian nonparametric approach to multiple testing problems, which often require testing several means with the assumption of rejecting infrequently, as motivated by the need to analyze DNA microarray data. In this setting, we test multiple hypotheses $H_0 : \theta_i = 0$, where θ_i is the mean expression change for genes $i = 1, 2, ...j$, with the goal of detecting the genes for which the hypothesis should be rejected. This paper generalizes the current Bayesian approach (Scott and Berger, 2006) to a nonparametric setting by replacing the Gaussian assumption and additionally to the heteroskedastic case, which refers to a setting where the variance can fluctuate across subsets of observations. We developed a framework that

employs a discrete approximation to a Polya tree prior that enjoys fast, conjugate updating, centered at the usual Gaussian distribution that can incorporate the information of heteroskedastic errors for improved simultaneous inference. This results in a computationally efficient technique that keeps the false discovery rate low in a variety of multiple testing scenarios, including cases where the Gaussian assumption doesn't hold. We demonstrate these extensions by exploring real data to discover genes that have mean expression changes significantly larger in magnitude than other genes in cancerous (versus noncancerous) tissue.

The *Statistics and Computing* (2016) paper proposes computationally efficient methodologies for density estimation, modeling regression error, and fitting random intercept generalized linear mixed models (GLMMs) via two novel nonparametric priors based on Gaussian-centered Polya trees. The first prior is a discrete approximation that simply replaces the sets at each level of the tree with point mass, and the second prior smooths this discrete approximation for use with continuous data. The resulting density model has similarities to the penalized B spline approximation of Komárek and Lesaffre (2006) on a fixed set of knots, but with random degree. The resulting Markov Chain Monte Carlo (MCMC) scheme, a technique for estimating by simulation from probability distributions, enjoys conjugate updating for almost all parameters and excellent mixing. The resulting methodology performs similarly to the Dirichlet Process Mixture, which usually outperforms Polya tree models, by ameliorating the spikiness usually associated with Polya tree density estimates while retaining valuable interpretation of the model's key values. The versatility of these methodologies makes them relevant to a number of applications as shown in a meta-analysis exploring alcohol as a risk factor in breast cancer as well as a survival analysis about how attributes of patients (e.g., marital status, race, location, and age) affect their survival time post-diagnosis.

The paper accepted at *Advances in Data Analysis and Classification* extends Quadratic Discriminant Analysis (QDA) (Cox, 1958) and Linear Discriminant Analysis (LDA) (Fisher, 1936; Rao, 1948) to the Bayesian nonparametric setting, providing a competitor to MclustDA (Fraley and Raftery, 2002). This approach moves classification beyond the assumption that the data can be modeled by a Gaussian or mixture of Gaussians, by employing a multivariate mixture of finite Polya trees and a p-dimensional location-scale mixture that is a direct generalization of the multivariate Gaussian. The proposed method is quite fast compared with other supervised classifiers and very simple to implement as there are no kernel tricks or initialization steps, which may make it one of the most user-friendly approaches to supervised learning. The flexibility and performance of this methodology are demonstrated for a variety of machine learning benchmarks, including the diagnosis of diabetes among women of Pima Indian heritage living near Phoenix, Arizona (Smith et al., 1988).

I have collaborated on two papers with my colleague Roy Bower (Department of Mathematics, Xavier University) that present new testing procedures for the independence of bivariate Poisson data, which provides researchers a way of examining the relationship between two counts (e.g., the number of doctors visits and number of diagnosed

conditions). Our first paper was published in *Communications in Statistics: Case Studies, Data Analysis and Applications* (2018), and our second paper has been accepted for publication in the same journal. On both projects, I designed and coded the simulation studies that displayed the effectiveness of these models as compared to existing alternatives in terms of power and significance and I assisted with the derivations.

Peer-reviewed Publications in Scholarly Journals in Other Disciplines

I have co-authored eight papers for refereed journals in other disciplines. Four of these papers are published, two are accepted, one is in revision, and one is submitted. Such multidisciplinary collaboration is helpful both to me and to the collaborator(s) from other disciplines, as we employ complicated quantitative analyses that extend qualitative hypotheses which I often play a role in crafting. Through these collaborations, I expand my experience, improving my background for teaching while aiding others with my specialized knowledge and expertise. I also find motivation for new questions in statistics; e.g., my focus on biological health science applications has led to flexible approaches that aid model fitting and improve prognosis and diagnostic abilities. I now summarize my contributions to collaborative multidisciplinary research.

I have published one peer-reviewed paper in collaboration with my colleague John Fowler (Department of Orthopaedic Surgery, University of Pittsburgh) in the *Journal of Bone and Joint Surgery* (2015). I completed Bayesian latent class analysis to determine the true diagnostic accuracy of ultrasound, nerve conduction velocity test, and Carpal Tunnel Syndrome 6 symptom scale for the diagnosis of carpal tunnel syndrome without a gold standard. This allowed for the assessment and confirmation of Dr. Fowler's hypothesis that nerve conduction velocity tests, which are uncomfortable for patients, are no longer necessary as ultrasound measurement provides similar results.

I have published one peer-reviewed paper in collaboration with my colleague Derek Michael Da Silva (Department of Sociology, King's University College at Western University) at the *Sociology of Sport Journal* (2017). We employed a classification scheme to demonstrate that "Scouts," so-called expert analysts of high school athletes, promote a system of scrutiny that is arbitrary and unrelated to objective measurements. "Scouts" thus promote a system of scrutiny that contributes to the systemization of socially dominant relations between the observers and the observed. A second paper, accepted for publication at the same journal, explores public responses to the 2018 Humboldt Broncos bus crash that killed thirteen members of the Saskatchewan-based junior hockey team. We explore how a sporting tragedy generates conversation, mobilizes communities, and constructs expressions of national and regional identity through public displays of grief, sadness, and collective mourning. I worked with Dr. Da Silva to extend his qualitative hypotheses through natural language processing, latent Dirichlet allocation (Blei et al., 2003), and sentiment analysis of over 125,000 Twitter responses that I extracted through Twitter's Premium API using the Python programming language. This quantitative work provided large-scale

evidence extending the qualitative work that indicates expressions of community-building and collective bonding, particularly around notions of masculinity and youth, after the tragedy reinforce hegemonic ideals of what it means to be "properly" Canadian.

I have published one peer-reviewed paper with my colleague Erin Cooley (Department of Psychological and Brain Sciences, Colgate University), a second is in revision and a third has been submitted. Our first paper was published in the *Journal of Experimental Psychology: General* (2017) and is currently a first round finalist for the 2019 Prize for Research in Public Interest Communications awarded by the *frank* gathering at the University of Florida. If our paper is selected as a top-three paper, I will present this research in a variety of settings at the 2019 *frank* conference which focuses on the social impacts driven by science, justice, and expression. Using mixed-effect approaches to regression, in three studies, we model how subtle shifts in framing can alter the perception of groups. For example, people may be less willing to help "a camp full of refugees" than "refugees in a camp." The second paper has been submitted to *Social Psychological and Personality Science*. This paper explores the hypothesis that racial biases may be amplified when Black[1] subjects are stopped in groups versus alone. Through classification in complex moderated mediation analyses for over one million stops in New York City, we provide compelling evidence that the perception of groups amplifies racial biases in such stops and that early steps in the escalation of policing may provide a critical opportunity for intervention. The third paper was the result of discussions about the intersection of racial bias (Dr. Cooley's research area) and bias based on socioeconomic status that stemmed from my experience as a poor White student with a parent who suffered from opioid addiction. Through moderated mediation, we show that while social liberals are more sympathetic to poor people than social conservatives, reading about White privilege decreased their sympathy for a poor White (versus Black) person. A careful experimental design and regression in a moderated mediation analysis lead us to conclude that speaking of racial privilege in isolation, without also considering economic inequality, might further divisions between groups – divisions amplified by political ideologies.

I have published one peer-reviewed paper in collaboration with my colleague Ana Jimenez (Department of Biology, Colgate University) in *PLOS One* (2018). This paper employs exploratory data analysis and feature selection techniques to elucidate the mechanism that may allow for small breeds of dogs to age more slowly compared with large breeds in the context of cellular metabolism and oxidative stress. These analyses indicated that large breed dogs may have higher glycolytic rates, and DNA damage, suggesting a potential mechanism for their decreased lifespan compared with small breed dogs.

I have one paper in collaboration with my colleague Aaron Robertson (Department of Mathematics, Colgate University) that has been accepted for publication at *Experimental Mathematics*. Here, we approximate the distribution of the number of monochromatic complete subgraphs over edgewise 2-colorings of complete graphs

and monochromatic arithmetic progressions. While researchers have shown the distributions to be asymptotically Poisson, there is no such result or approximation for a graph with a finite number of vertices, which shows the Poisson approximation to be underdispersed. Through simulation, we provide convincing evidence that these distributions are well approximated by the Delaporte distribution. These implications lead to questioning how the Delaporte approximation might be used to provide better estimates about the boundaries of Ramsey numbers which can lead to further research.

Research Grants

I have submitted and received a new initiatives grant for $3000 from the Central New York Library Resources Council (CLRC) titled, "The Data Analysis and Collaboration Network," with Josh Finnell (Head of Research and Instruction, Colgate University). The work associated with this grant addresses the problem of the gap between institution-based expertise as research interest increases at smaller colleges and universities. Leveraging the New York 6 Consortium as a starting point, our intention is to build an inter-institutional statistics collaboratory of "isolated" statisticians and data librarians from community colleges and regional universities to provide resources and collaborative support for researchers unaffiliated with a large research institution. Regionally, we envision using this pilot project to apply for an Academic Research Enhancement Award (AREA; NIH-R15) grant that would help solidify, expand, and sustain the cohort of experts with the goal of developing the next generation of data scientists through the experience of quality multidisciplinary collaboration. In fact, I'm currently planning to have students from my MATH 354 class complete summer research in developing statistical software for standard statistical approaches that will contribute to the collaboratory's resource page.

I am also part of a team of colleagues, including Dr. Cooley, Lauren Philbrook (Department of Psychological and Brain Sciences, Colgate University), Jazmin Brown-Iannuzzi (Department of Psychology, University of Kentucky), and Ryan Lei (Department of Psychology, Haverford College), applying for a major research grant from the Picker Interdisciplinary Science Institute after our pre-proposal was invited for a full proposal. This work aims to explore the psychological processes that link race with social class, extending the ideas from the third paper with Dr. Cooley. Through this multidisciplinary project, we hope to better understand – and thus intervene to mitigate – distress and discrimination experienced by U.S. citizens.

I am also working with Dr. Da Silva on two grants to continue our collaborative research line in using quantitative methods to extend and provide evidence for qualitative hypotheses. We have submitted one grant titled "Identifying sentiment-based indicators of extremism within India's political landscape" for Public Safety Canada Grant. We are also collaborating on a second grant titled "How radicalization became the dominant framework for understanding terrorism," which we will submit for an Insight Development Grant through the Social Science and Humanities Research Council of Canada. These projects will leverage methods in natural language processing to explore

transitions away from conventional political, religious, or otherwise ideological beliefs and activities towards extreme views and violence. To accomplish these objectives, we will conduct a two-phased mixed-methods project utilizing computational tools of social network analysis, and natural language processing to detect a diffusion of discourses and communications to document how radicalization has emerged as the dominant governmental logic for governing issues related to terrorism.

Working with my colleagues in psychology and sociology, I will help develop resources and applications that will be made freely available on the resources page of the collaboratory mentioned above. These applications will allow any scientist to reproduce our findings and will make complicated discipline-specific methodology more accessible in `RShiny` – perhaps the most widely available venue for the dissemination of newer statistical methods. Researchers of all backgrounds will be able to utilize this methodology without the steep learning curve of a new programming language. We hope this will be the start of an arXiv-like resource that provides standard and advanced statistical procedures for bench and social scientists alike, making Colgate the place where data science is accessible to everyone.

Invited and Contributed Presentations at Professional Meetings or Conferences

I have given eleven research presentations – five contributed talks and six posters at professional meetings by consistently giving presentations at our profession's largest meetings: the Joint Statistical Meetings and the Joint Mathematical Meetings, as well as specialized conferences such as Bayesian Nonparametrics and Objective Bayes. I have also given an invited colloquium talk at SUNY Geneseo.

Establishing a Research Program Independent of the Candidate's Doctoral Dissertation

My research program has flourished well beyond my time as a graduate student. My recent advances in Bayesian nonparametrics toward statistical machine learning methodology has been tied to my experience with my advisor, Timothy Hanson, but has served as a conduit toward independent research as he has left academia, taking a position in industry. My current and future work in developing Bayesian nonparametric solutions will be independent of my dissertation and my productivity in statistics and multidisciplinary collaborations have materialized independently of my dissertation.

Organizing Research Conferences/Special Sessions at Professional Meetings

I am participating as a Mathematical Association of America Project NExT fellow. This program addresses improving the teaching and learning of mathematics, engaging in research and scholarship, as well as performing service to the discipline. As a member of this organization, I have planned a session at the 2019 Joint Mathematics Meetings entitled "Standards-based Grading: Tools and Tips for Successful Implementation." This session gives

implementation-focused information from how to craft syllabi and course design that include standards-based grading.

Editorial Service

I have served as a referee for manuscripts submitted to *Biometrics* (2); *26th Annual European Symposium on Algorithms* (1); *Computational Statistics and Data Analysis* (2); *Problems, Resources, and Issues in Mathematics Undergraduate Studies* (1); *Journal of Statistics Education* (2); and *Interdisciplinary Sciences: Computational Life Sciences* (2). I have also served as a referee on a mathematical statistics textbook for *Wiley.*

In addition to this traditional service as a reviewer, in recognition of my multidisciplinary work, I have been invited to participate as a Scientific Review Officer on a panel for *The Small Business Disease Prevention and Management, Risk Reduction and Health Behavior Change* for the National Institutes of Health (NIH) and I remain an active reviewer available for future panels.

A Projection of My Continued research Interests and Endeavors

I have spent five successful semesters as an assistant professor and, as this statement describes, I have been very active in both my methodological contributions and applications of statistics in other disciplines. I have maintained an active program of research that I look to extend as I look to the future. My goal is to continue to develop new Bayesian nonparametric methods to address the need for more flexible modeling procedures as well as collaborating on multidisciplinary projects where statistics can be used to address real problems arising in sociology, psychology, biology, and medicine.

More specifically, my personal research goals for the next three years are to average 1-2 research publications per year in reputable statistics journals; to average an additional 1-2 research publications per year in reputable journals in other disciplines; to build a new collaborative relationships with statisticians and scientists from other areas and institutions; and to continue to be active in research activities by refereeing papers, participating in conferences, etc. Below, I delineate current work and future projects aimed to achieve these stated goals.

Bayesian Nonparametrics: I am currently working to create computationally efficient and flexible approaches for prediction and feature selection with future work on new approaches to unsupervised learning. I believe this will be a fertile avenue for research productivity due to the relative infancy of Bayesian nonparametrics research in terms of its methodological application and its increased use in scientific research.

Specifically, I am currently working to extend the paper accepted at *Advances in Data Analysis and Classification* to the continuous-response case. Doing so requires a significantly improved approach to smoothing. While some work in doing this has been completed by sampling from the Haar distribution, essentially a uniform distribution

on orthogonal matrices, I believe an alternative approach can provide a more efficient but equally effective solution. This new work will propose candidate orthogonal matrices using Householder reflections which leads to a more computationally efficient MCMC scheme for smoothing. This smoothing will be paramount to the extension while retaining the computational efficiency of the original result.

The approaches to unsupervised learning would involve models similar to the paper accepted at *Advances in Data Analysis and Classification* and the current work to extend it but requires a completely new framework for these models to be employed. I have some ideas for hard clustering via Polya trees based on the framework of Chen et al. (2016), which would require deriving a variety of attributes about a multivariate Polya tree. These derivations should require an MCMC scheme with additional hyperparameters to guarantee the Polya tree is not only continuous but also differentiable. This is necessary because the local modes are related to the gradient vector of the estimated density curve.

An additional future project exists in an extension to the *Statistics and Computing* paper by reworking the computationally tractable smoothed approximate Polya trees to the heteroskedastic case. This would allow us to create an accelerated failure time model for survival data that would add substantial modeling flexibility to the current approach, likely decreasing the impact of outliers and likely decreasing the average size of prediction intervals.

As I complete my current and future work, I am also working to make new and previous research available to other statisticians and researchers as an `R` package and via `RShiny` applications.

Statistical Collaboration: I am currently working with Dr. Bower to develop two score tests for testing independence based on different specifications of the bivariate Pareto distribution, as motivated by questions about the independence of gestation time and brain weight for placental mammals. Future work includes further collaboration in developing score tests for independence due to their importance in biological and social sciences. We've discussed several projects including expanding our results to the more flexible negative binomial and Delaporte distributions, and to zero-inflated models that are important for social and health science experiments where data contain both structural (e.g., subjects that abstain from drug use) and random zeros (e.g., those that happened not use during the study) which may yield very different outcomes (e.g., depression or heart health).

I am also working with Joshua Tebbs (Department of Statistics, University of South Carolina), Chris McMahan (Department of Mathematical Sciences, Clemson University), and Christopher Bilder (Department of Statistics, University of Nebraska, Lincoln) on NIH-funded research involving pooled testing, with a focus on sexually transmitted diseases. This testing protocol allows doctors to quickly screen populations for low-prevalence infectious diseases by testing pools of specimens once. If the test comes back negative, the whole pool is diagnosed as disease-free, and if the test is positive, algorithms can help divide the pool into smaller groups for

retesting to efficiently pinpoint the infected individual. By leveraging previous Bayesian regression techniques and devising a new framework on which to apply supervised learning, we hope to create a methodology that accurately identifies misdiagnosed patients. Currently, we are reaching out to clinics in New York State to acquire de-identified data that will help us evaluate existing methods and develop a new approach that uses the risk-factors of patients to assess the likelihood of misdiagnoses in a pooled-testing setting on the patient level.

Multidisciplinary Collaboration: I am currently working on another project with Dr. Da Silva. We propose a quantitative model of "squeeze theory" on ISIS; e.g., we expect more events in the U.S. and Europe when we bomb, take more land back, etc. This project was the result of sharing an interesting dataset with Dr. Da Silva, who researches terrorism through a sociological lens and conversing about the possible implications. While sociologists and other researchers have completed qualitative research, very little has been done to quantitatively measure this theory. Many researchers expect that the squeeze theory should be quite clearly quantifiable and are excited about the results.

I recently started a multidisciplinary research project with Dr. Bower and Annie Ray (Department of Biology, Xavier University). Through a carefully designed experiment, we quantified how the presence and amount of a pheromone affect the number of beetles captured in traps. Preliminary results suggested that high pheromone levels attracted beetles to the trap and the biologists are collecting additional data, under varying conditions, which will be analyzed with negative binomial count regression models and analysis of post-hoc contrasts.

I have also started working on a project with a colleague Don Sull (Sloan School of Management, Massachusetts Institute of Technology). This project aims to define corporate culture using posts on an employee review website by creating a statistical model that predicts the ordinal culture score (1-5) given to a company by an employee-reviewer based on the features created with text from the free-response questions. This collaborative project stemmed from discussions about my work on data mining and applying natural language processing to tweets. Using Bayesian ordinal logistic regression, preliminary results suggest that health care positions (e.g., pharmacist, pharmacy technician, physician, physician advisor, or therapist) and front-line employees (e.g., driver, maintenance workers or unskilled laborer) report the lowest culture scores, on average.

I am working on a project that will be presented at the *Public Policy and Public Administration section for the 50th-anniversary meeting of the Northeastern Political Science Association* in collaboration with my colleagues Andrew Pattison (Program of Environmental Studies, Colgate University) and Jose Marichal (Department of Political Science, California Lutheran University). We expanded upon current work by applying the approach of Gupta et al. (2018) to tweets that mention fracking policy in New York State. This type of discourse framework is employed for the first time using advanced methods in natural language processing, allowing us to explore a much larger corpus of tweets (263,000) over a much longer time frame (6 years) to examine whether policy

changes (e.g., fracking bans) impacted the types of narratives used by elites to sway elite/mass opinion.

I have a project on the horizon with Kris Macomber (Department of Sociology, Meredith College) and Juliette Grimmett (Founder of the Chrysalis Network). This project is an inquiry into sexual assault on female-only college campuses for which there is little-to-no quantitative research. This project was the result of my surprise in hearing about the dearth of research in this area during a session at a Sociology conference Dr. Macomber and I presented at, and subsequent conversation. We are currently collaborating on laying the groundwork for an NIH grant to help fund conducting a nationwide survey across thirty-seven women's colleges and universities in the U.S. This inquiry is designed to illuminate the campus climate and start a new line of research about sexual violence as findings are likely to require further exploration.

Statistical Education: I have spearheaded the effort to revise and develop the statistics curriculum at Colgate by redesigning several courses. Numerous students have told me how much they appreciated the rigor of these courses and the clear connections made to other disciplines through application.

A major contribution I make in the area of instruction is to create course notes that are available to students online. At Colgate, I have created notes for MATH 105, MATH 316, and MATH 416 and I have started creating notes for the new sequence of courses MATH 354 and MATH 454. Preparing these notes is a huge time commitment, but the response to them has been positive while allowing me to cover graduate-level material in the undergraduate setting. Even students not enrolled in my courses tell me they have downloaded my notes for future use in preparing for actuarial exams, interviews, and their graduate courses. My course notes are not merely PowerPoint slides; rather, they are textbook-quality summaries of examples, ideas, and theory written using LaTeX.

The notes for MATH 105 are close to submission-ready and I plan to submit them as a textbook after I teach that course again in Fall 2019. The MATH 316 and MATH 416 notes will require another year of work. These texts push the pedagogical aspects of statistics beyond algebra that can be completed with a pad and paper. The aim of these texts is to provide challenging material to the modern student who is capable of more, having learned in a time when technology and computing are ubiquitous. For examples, the introductory statistics textbook synthesizes modern statistical approaches to classical problems by teaching the methodology R uses to calculate statistics and how they are interpreted, not the simplified versions seen in traditional introductory texts; and the probability and mathematical statistics notes use R to visualize theory so that students can see and understand complicated mathematical concepts and application usually left for a graduate-level course by discovering results graphically before derivation.

Service

In five semesters at Colgate, I have hosted Commons events and participated in many recruitment endeavors, like April Visit Days Events, opening my classrooms to visiting students, and meeting prospective students particularly interested in mathematics. I have led student trips to conferences and competitions; e.g., the DataFest17 competition and Hudson Valley Undergraduate Math Conference. I also proctored the Putnam exam in Fall 2016.

I have discussed meeting tenure-stream candidates with Tracey Hucks and have been an eager participant in that new initiative, meeting with candidates for chemistry last year. I also regularly attend White Eagle, teaching tables hosted by the Center for Learning, Teaching and Research (CLTR), orientation events for new faculty, and I helped plan the last NASC retreat, which I also presented at. I am a member of two committees – I am in my first year of a three-year appointment to the Faculty Development Council as well as a member of the Research Computing Committee from which the Inter-Institutional Statistics Collaboratory was born. I look forward to contributing to the Colgate community, both current and prospective students and faculty.

Conclusions

My liberal arts view of statistics leads me to be naturally involved with faculty and student research across campus. This has proved to be valuable for several colleagues in publishing complicated quantitative analyses that extend their qualitative hypotheses, and myself as my interests extend beyond the mathematical aspects of statistics. This multidisciplinary work has informed both my theoretical work in developing flexible and accessible extensions to many statistical models as well as my teaching.

Bibliography

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of machine Learning research*, 3:993–1022.

Chen, Y.-C., Genovese, C. R., Wasserman, L., et al. (2016). A comprehensive approach to mode clustering. *Electronic Journal of Statistics*, 10(1):210–241.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*, 20:215–242.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188.

Fraley, C. and Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97:611–631.

Gupta, K., Ripberger, J., and Wehde, W. (2018). Advocacy group messaging on social media: Using the narrative policy framework to study twitter messages about nuclear energy policy in the united states. *Policy Studies Journal*, 46(1):119–136.

Komárek, A. and Lesaffre, E. (2006). Bayesian semiparametric accelerated failure time model for paired doubly-interval-censored data. *Statistical Modelling*, 6:3–2.

Rao, C. R. (1948). The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society - Series B*, 10:159–203.

Scott, J. G. and Berger, J. O. (2006). An exploration of aspects of Bayesian multiple testing. *Journal of Statistical Planning and Inference*, 136(07):2144–2162.

Smith, J. W., Everhart, J., Dickson, W., Knowler, W., and Johannes, R. (1988). Using the adap learning algorithm to forecast the onset of diabetes mellitus. In *Proceedings of the Annual Symposium on Computer Application in Medical Care*, page 261. American Medical Informatics Association.