Peter Krafft (PhD Applicant): Statement of Purpose

In statistics, flexible models are at the heart of the principle of letting datasets speak for themselves. In machine learning, flexible models are necessary if we want to reach the ideal of programs that can teach themselves. In both cases, Bayesian statistics offers a principled framework for developing these flexible models, and much recent work in Bayesian statistics such as in developing hierarchical Bayesian models and nonparametric Bayesian models, has focused on this goal. I intend to further advance the capabilities of Bayesian statistics.

This interest originated in part from my experiences with Prof. Michael Lavine, whose steadfast dedication to the principles of Bayesian statistics inspired me in the graduate-level mathematical statistics courses he taught and in the two research projects I did with him—one on modeling the light that reaches the floor of Harvard Forest and one on diagnosing an estimation issue in a model from spatial statistics.[1] In these classes and projects, I learned about how to design Bayesian models, how to use Monte Carlo methods to infer posterior distributions, and how to diagnose unexpected properties of complicated models both through close consideration of their mathematical form and through simulation.

At the same time I was learning about methodology, I was interested in applications involving cognitive science. I combined these interests in my undergraduate thesis advised by Prof. Andrew Barto and his student, Dr. George Konidaris. Hierarchical representations offer the ability to organize data into meaningful patterns at multiple levels of abstraction, but whether these structures are necessary is not clear. The game of Go is a grand challenge to artificial intelligence, so showing that hierarchical representations are useful in Go would provide powerful evidence for hierarchies being important in representing complex data. We hypothesized that a deep belief network, a model that extracts hierarchically organized latent features from data, would be able to represent the concept of *shape* in Go, a component of human strategy, by using its lowest level features to capture meaningful configurations of pieces, and higher levels to capture common interactions between those configurations. Since neither Prof. Barto nor Dr. Konidaris had prior experience with deep belief networks, I had to understand and implement this method with only their high-level guidance. Although the results of my experiments were inconclusive, this project was a valuable introduction to machine learning.

Recently, I started focusing on model development. I am now doing my MS project with Prof. Hanna Wallach (a computer scientist and an expert in Bayesian statistics) and Prof. Bruce Desmarais (a political scientist and an expert in network analysis). In this project we are using the email inboxes and outboxes from the county department managers of New Hanover County, North Carolina to infer attributes of managers who are particularly effective at getting their own agendas on their county's legislative docket. This project could help bureaucrats design more effective communication strategies. However, this research poses a major modeling challenge. Email data contain both text attributes (the subject and body of each email) and network attributes (the author and recipients of each email), but there has been very little work in joint text and network modeling. After brainstorming and discussing many possible models, I decided on one to investigate first. By extending a model from the network literature to a framework that is appropriate for email data and combining

[1] For full papers from these and other projects, please see `www.cs.umass.edu/~pkrafft/papers/`.

it with a standard probabilistic model of text, the model I developed can identify who each managers communicates with about particular topics. This model will allow us to discover whether managers with broader communication strategies are more effective than managers with more targeted communication strategies.[2]

For my PhD I want to focus on developing Bayesian methodology while continuing to work on interesting applications. The University of California at Berkeley would be ideal for this goal. In the Computer Science Division of the Department of Electrical Engineering and Computer Science at UC Berkeley, Prof. Dan Klein and Prof. Michael Jordan both do excellent work related to my interests. Their recent paper with Prof. Percy Liang, "Learning Programs: A Hierarchical Bayesian Approach", is fascinating. An example of a project I would be interested in pursuing would be extending this work to probabilistic programs, which are a recent flexible tool for expressing Bayesian models. I am also interested in developing nonparametric priors over new data structures. An example of this type of work is the paper by Prof. Jordan and his collaborators, "Nonparametric Combinatorial Sequence Models". If I followed that direction, I would also be interested in collaborating with Prof. Thomas Griffiths from the Department of Psychology who has done great work in nonparametric Bayesian modeling as well as in topic modeling, which is a subject of my current research.

In addition to doing research, I also enjoy teaching. I started teaching in my second semester at university, and I held an undergraduate teaching assistant position in every remaining semester of my undergraduate program. Furthermore, this semester I am the teaching assistant for the graduate machine learning course in my department, and I am mentoring my second undergraduate student. In all of these positions I have tried to teach passionately and effectively, and I believe this has benefited both my students' understanding and my own. By studying the methods of the lecturers I most admire, notably Prof. Farshid Hajir and Prof. Michael Lavine, I decided on four characteristics important to good teaching. I try to *balance* giving extra help to students who struggle and giving challenging problems to students who excel. I achieve *clarity* by using simple examples, giving visual explanations of abstract concepts, and using mnemonics when memorization is necessary. But I will not sacrifice *detail*, so I show all the background and logical steps for an idea while alluding to more advanced topics. Finally, by *experimenting* with different explanations and presentation styles, I constantly improve the quality of my teaching.

I am committed to the path of getting a faculty position at a university and joining the academic research community. My previous projects have introduced me to the steps of conceptualizing, formalizing, applying, implementing, analyzing, and advertising my research, and my teaching and mentorship positions have helped prepare me for the full duties of professorship. I have also begun to define my research strategy. I intend to make fundamental contributions to an emerging area with high potential impact such as probabilistic programming, and to collaborate with experts from areas such as cognitive science or political science so I can work on substantial applications. Given the opportunity, I would look forward to pursuing these goals with the support of the faculty and students at UC Berkeley.

Thank you for your time and consideration.

[2] I am currently writing up our preliminary results in the form of a tech report, which we plan to submit to the next International Conference on Machine Learning.