Candidate Statement

Yeo Jin Jung / jjung1089@gmail.com / Statistics Ph.D.

My purpose of pursuing studies in statistics is to devise methodologies for analyzing large scale complex data that work as the key analytic tools in the development of various academic fields.

My Desire for Pursuing a Ph.D. in Statistics

Studying the mathematical side of statistical models has been the most stimulating part of my academic endeavors. This naturally led to taking relevant mathematical classes such as calculus, linear algebra, and analysis, and showing good performance in those classes. A strong mathematical background helped me tremendously when I came across new statistical ideas and later began developing ideas of my own. I was also eager to implement these ideas into different programming languages and enjoyed challenging myself on projects that required intensive programming in R, Python, and C++. The fulfillment of exploring new concepts motivated me to join the master's program in Statistics and Data Science at Yonsei University. I was introduced to the field of combining spatio-temporal and functional data analysis by my advisor Dr. Jaewoo Park. During my independent studies, I was intrigued by the applications of spatial functional models on the analysis of medical images. The concept of treating high-dimensional image data as infinite-dimensional smooth functional curves, thereby enhancing the simplicity of the given tasks, was fascinating. I was fortunate to have had the opportunity to participate in several related in-depth research projects focused on these fields. These experiences made me realize my desire to continue my academic journey in statistics. I hope to conduct my own independent research that integrates knowledge from various disciplines.

My Research Experience and Interests

My previous research projects center around analyzing spatio-temporal data, functional data, and network data with applications to the fields of epidemiology and psychometrics.

Fast Bayesian Spatial Model with Functional Covariates I developed a Functional Sparse Spatial Generalized Linear Mixed Model (FSSGLMM) with my advisors Dr. Jaewoo Park and Dr. Hyun Bin Kang. The goal of the project was to extend an existing areal model, SGLMM, by adding a functional parameter, and to apply the model on COVID-19 incidence datasets. We projected spatial random effects to lower dimensions for faster computation and reduction of spatial confounding. My main contribution was measuring the uncertainty of estimated functional parameters by constructing their credible interval bands. It was challenging because inference on functional parameters is an often overlooked topic in the related literature. I adapted the idea of simultaneous credible bounds for functions and merged credible intervals on finite grid points that were constructed from posterior samples from Markov Chain Monte Carlo (MCMC). I also implemented these MCMC algorithms using a C++ compiler in R to make computations faster. We submitted a paper based on this work and I am currently working on its revision with a focus on simulation studies under different spatial and functional settings. I presented this work at the 2021 Brain Korea Academic Conference and earned an Excellent Paper Award.

Autologistic Network Model for Progression of Depression Symptoms The goal of this project is to identify the underlying network structure of depression symptoms and predict the path of how these symptoms develop sequentially. Advised by Dr. Jaewoo Park, Dr. Ick Hoon Jin, and Dr. Minjeong Jeon, I assumed that proximity between symptoms could be modeled by a spatial model, and I fitted an auto-logistic model using Bayesian inference. To reflect the temporal dependence, I studied various existing methods on analyzing spatio-temporal binary datasets and adapted the idea that the development of a particular symptom could be contingent on the previous states of neighboring symptoms. I incorporated this concept into the auto-logistic model by dividing the predictors into previously emergent and non-emergent sets. A psychological path of the symptoms could be estimated from a full transition matrix derived from the model. I am currently working on the interpretation of the estimated auto-logistic parameters in describing how they reflect the complex relationships between different symptoms. I hope to conclude this research soon with the publication of a paper in a psychometrics journal.

I believe that my prior studies and research experiences provide me with a solid foundation in which to expand my interests in high-dimensional statistics. For instance, my work on the project with functional covariates allowed me to familiarize myself with existing literature and understand how to adapt the concept of credible intervals in order to address a deficiency found in prior research. This skill of being able to recognize gaps in previous research and suggest ways of overcoming them will be beneficial as I continue engaging in research at the doctoral level. My primary research interest is developing methodologies for high-dimensional inference that can be applied to many forms of large scale complex data, especially data that has functional and spatial information. I am currently taking a graduate level course, *Statistical Theory for High Dimensional and Big Data* to study mathematical theories related to this field. While the core part of my academic goal is to formulate concrete statistical theories, I also want to contribute to the expansion of statistics' applied work. My prior independent study of functional and spatial data analysis was closely linked with analyzing brain images, which also uses adjacency information between brain voxels. With this spark of interest, I am looking forward to formulating statistical theories for biomedical areas such as epidemiology, genetics, and medical imaging.

Why University of Chicago?

The broad range of interdisciplinary research and its distinguished training program in the Department of Statistics at University of Chicago make it the ideal place to pursue doctoral studies. I believe that the Ph.D. program aligns well with my specific interests involving the interactions between mathematics and applications of high-dimensional statistics. I am also looking forward to interacting with other brilliant researchers and faculty members at the Department and conjointly finding answers to scientific questions. Upon completing my Ph.D., I would like to remain in academia and continue to contribute to the academic community. It would be an honor to have the opportunity for further studies at University of Chicago.