BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Ren, Boyu

eRA COMMONS USER NAME (credential, e.g., agency login): BOYUREN

POSITION TITLE: Instructor in Psychiatry (Biostatistics)

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.*)

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	FIELD OF STUDY
Peking University	BS	07/2012	Physics
Harvard University	PhD	05/2017	Biostatistics
Dana-Farber Cancer Institute	Post Doc	06/2020	Biostatistics

A. Personal Statement

As an academic biostatistician, I have a broad background in statistics, with specific training and expertise in Bayesian statistics and machine learning. My methodological research includes multi-study learning approaches and Bayesian nonparametrics. Specifically, I have developed novel statistical tools for the analysis of high-dimensional count data, to harmonize heterogeneous studies into robust prediction rules and to estimate causal effects of continuous exposure. To date, a major focus of my collaborative research has been on studies of the microbiome, cancer, and environmental health. Since joining the faculty at McLean Hospital in July 2020, I have begun collaborating on a range of projects in mental health research broadly defined. My expertise in statistical methods, coupled with my prior experience collaborating in a broad range of research projects, place me in an ideal position to ensure that appropriate statistical methods are leveraged in the design, analysis, and reporting of results of the proposed studies.

B. Positions and Honors

Positions and Employment

2017–2020 Postdoctoral Fellow, Department of Data Sciences, Dana-Farber Cancer Institute, MA
2020– Assistant Biostatistician, Laboratory for Psychiatric Biostatistics, McLean Hospital, MA
2020– Instructor, Department of Psychiatry, Harvard Medical School, MA

Other Experience and Professional Memberships

2015– Member, The International Society for Bayesian Analysis
2018– Member, American Statistical Association

Honors

2015 Robert Reed Award, Harvard T.H. Chan School of Public Health, MA
2016 Distinction in Teaching, Harvard T.H. Chan School of Public Health, MA
2016 Rose Traveling Fellowship, Harvard T.H. Chan School of Public Health, MA

C. Contributions to Science

1. My early publications focused on the construction of Bayesian models for high-dimensional count data. I developed and published two Bayesian nonparametric models based on dependent Dirichlet processes. Through collaborations with a research group on the human microbiome, I applied these two models to verify the latent structure of a vaginal microbiome community and to recover the effects of observed covariates on microbial profiles of infants that are susceptible to Type I diabetes.
    a. **Ren, B.,** Bacallado, S., Favaro, S., Holmes, S., & Trippa, L. (2017). Bayesian nonparametric ordination for the analysis of microbial communities. Journal of the American Statistical Association 112(520), 1430-1442.
    b. **Ren, B.,** Bacallado, S., Favaro, S., Vatanen, T., Huttenhower, C., & Trippa, L. (2020). Bayesian mixed effects models for zero-inflated compositions in microbiome data analysis. Annals of Applied Statistics, 14(1), 494-517.

2. During my postdoctoral training, I developed a machine learning framework for predictions based on multistudy datasets. This research focused on two types of prediction tasks: constructing predictors with replicable performance across studies (generalists) and predictors targeting a specific study (specialists). I have recently proposed a stacking method that combines study-specific predictors to accomplish both tasks, which also addresses with high computational efficiency the issue of data reuse in previously proposed methods. I have established the asymptotic optimality of our generalist predictors and have proved their finite-sample superiority in cross-study replicability over existing stacking predictors.
    a. **Ren, B.,** Patil, P., Dominici, F., Parmigiani, G., & Trippa, L. (2020). Cross-study learning for generalist and specialist predictions. arXiv preprint arXiv:2007.12807.

3. An additional area of my research is on causal exposure response curve (ERC) estimation. Existing ERC estimation methods often do not come with readily available uncertainty estimates. To resolve this difficulty, I proposed Bayesian approaches for ERC estimation, which are equipped with automatic uncertainty characterization. I have constructed a nearest neighbor Gaussian process (NNGP) model that is asymptotically equivalent to a recently proposed doubly-robust matching algorithm and have verified the performance of this NNGP model with simulation studies. I have also incorporated the ERC estimation into multi-study learning framework based on the stacking approach for generalist predictions to ensure the replicability of the estimated ERC across different studies. A manuscript describing this work is in progress.

Complete List of Published Work in MyBibliography:
https://www.ncbi.nlm.nih.gov/myncbi/1huS_a8H98g1n3/bibliography/public/

D. Additional Information: Research Support and/or Scholastic Performance

Ongoing Research Support

DMS181082 Parmigiani and Trippa (PI and Co-PI) 07/01/18-08/31/20

NSF

Statistical Methods for Multi-Study Predictions

The goal of this project is to make progress in the area of cross-study replication of predictions.
Role: Postdoctoral Researcher