Jeremiah Zhe Liu

Department of Biostatistics
Harvard University

zhl112@mail.harvard.edu
jereliu.info
github.com/jereliu
+1 (319) 594-4694

Education

Harvard University (Boston, MA) *PhD Biostatistics, Minor in Computer Science* Expected 2018

Research Keyword: Gaussian Process, Ensemble Learning, MCMC & VI, Robust Inference for ML Models

GPA: 3.94/4.00.

University of Iowa (Iowa City, IA) *BS Statistics, Mathematics, Minor Computer Science* May 2013

magna cum laude, GPA: 3.96/4.00.

Professional Experience

Martinos Center for Biomedical Imaging, Mass General Hospital 2016-Pres

Research Fellow / Machine Learning Scientist

- Building LSTM/DNC-based meta-learning optimizer to aid automatic discovery of novel fMRI configurations.
- Participated in design of manifold-inspired architecture for deep learning system for fMRI image reconstruction.

Harvard Clean Air Research Center 2013-2015

Assistant Statistician

- Built spatiotemporal prediction system for heavy-metal air pollutants by integrating information from various sources (air monitoring records, meteorological information, etc) under Random Forrest and Kernel Regression.
- Implemented automated feature selection for GIS features using a combination of measurement error-based weighting and Ridge-type penalization. Conducted stratified cross validation to assess the model's out-of-sample prediction and the influence of prediction error on the risk estimation in second-stage association studies.

Technical Skills

- **Analysis & Modelling:** Python (tensorflow, pytorch, pyMC3), R, Matlab
- **Graphics & Documents:** ggplot2, OpenGL, Shiny, ArcGIS, LaTeX
- **High Performance Computing:** C (CUDA, OpenCL, OpenMP)
- **Software Development:** Python, C++, Java, Bash

Thesis Research

Robust Hypothesis Test for Functional Effect with Gaussian Process, *NIPS 2017* 2015-2017
Advisor: Dr. Brent Coull & Dr. Xihong Lin

- Theme: Enable classical inference on ML models
- Proposed an efficient hypothesis test to detect nonlinear feature effects under Gaussian Process.
- Proposed a cross-validated ensemble estimator for null model to guarantee robust estimation in small sample.
- Work revealed unique connection between model generalizability and the performance of the statistical test.

On Convergence of Modern Sampling Methods for Bayesian Latent Factor Models 2016-Present.
Advisor: Dr. Lorenzo Trippa & Dr. Sergio Bacallado

- Theme: Discern the efficacy of modern sampling methods in high-dimension, nonlinear factor models.
- Introduced an modelling framework that recast exponential-family factor models as a regularized Bayesian NN.
- Under proposed framework, investigating the convergence speed of HMC, ADVI, and Operatator VI under different settings of model dimension and outcome distribution.
- Devised a diagnostic method to detect algorithm convergence in high-dimension parameter space for intractible posteriors. Proved the validity of proposed method throgh weak convergence arguments.

Open Source Project

GURLS_MKL: Fast Multiple Kernel Learning Library for GURLS Package 2015

- Independently developed multiple kernel learning functionality for *Grand Unified Regularized Least Squares* (GURLS), an state-of-art supervised-learning package developed at MIT
- Extended fast Proximal Forward-Backward Splitting (PFBS) optimization algorithm to allow memory-efficient iteration update with parallel support. Derived boundary conditions on algorithm parameters to guarantee model convergence.

GPU-Accelerated Sampling for Bayesian Normal Conditional Autoregressive Models 2012
PI: Dr. Kate Cowles

- Designed and implemented parallel algorithms in OpenCL for new model computation strategy proposed by Cowles et al.(2012) for Bayesian Normal CAR model.
- Implementation incorporated into R package *CARrampsOcl*.

Coursework

Statistical Methods
- Theory of Reproducing Kernels • Advanced Regression & Learning
- Semi-parametric Inference • Computation Intensive Statistics

Theoretical Statistics
- Probability Theory • Advanced Bayesian Inference • Theory of Hierarchical Linear Models
- Environmental & Spatial Statistics • Analysis of Genetic Association Studys
- Causal Inference

Mathematics
- Linear Algebra & Multivariate Calculus • Real Analysis & Measure Theory • Matrix Theory
- Numeric Analysis • Ordinary Differential Equation • Nonlinear Optimization

Computer Science
- High Performance & Parallel Computing • Data Structure • Algorith • Foundation in GIS

Consulting Experience

Causal Networks for Retarded Bone Growth in HIV-infected Adolescents 2014-2015.
Advisor: Dr. Brent Coull, Dr. Jane Lindsey & Dr. Denise Jacobson

- Trained regularized probabilistic network to model the association between biomarkers and bone growth measures in prenatally HIV-infected adolescents in PACTG 1045 study.
- Wrote interface between M-plus and R to deploy computation-intensive hypothesis generation and testing.
- Identified sub-collection of biomarkers robustly associated with retarded bone growth in HIV-affected teenagers.

High-level Air Pollution and Health in Taiyuan, China: from Birth to Death 2011-2012

- **Objective** To assess the association between air pollution (PM10, SO2, NO2, CO) and series of health endpoints (Negative Birth Outcome, Children Repiratory Symptom, Mortality) in highly air polluted city in China.
- **Role** Modelled Concentration-response Functions (CRF) between exposure and health outcome through Generalized Additive Model with cubic splines. Wrote C++ program to detect change points in nonlinear CRFs in order to estimate the dose-specific relative risk of air pollutants.

Standards for Birthweight by Gestational Age in Northern China, 2010 2010

- **Objective** To construct centile charts of birthweight, body length, head and chest circumference from 2006-2010 Taiyuan birth surveillance data containing 200,000 newborn records. (Negative Birth Outcome, Children Repiratory Symptom, Mortality) in highly air polluted city in China.
- **Role** Combined Two-compononent Mixture Models with LOWESS to obtain robust percentile regression for noisy data.
- **Status** Submitted to local Dept. of Health to release as diagnosis reference for Small-for- Gestational-Age (SGA) Newborn, and journal paper accepted by *Early Human Developement*

Publications

Machine Learning, Theory & Method

Liu JZ, Coull B. *Robust Hypothesis Test for Nonlinear Effect with Gaussian Processes*. Proceedings of Neural Information Processing Systems (NIPS) 2017

Machine Learning, Application

Zhu B, **Liu JZ**, Rosen B, Rosen M *Image reconstruction by domain transform manifold learning*. Nature. *Accepted*

Public Health

Hswen Y, Brownstein J, **Liu JZ**, Hawkins J *Use of a Digital Health Application for Influenza Surveillance in China*. American Journal of Public Health, 2017; e1 DOI: 10.2105/AJPH.2017.303767

Liu JZ, Lindsey J, Coull B, Jacobson D. *Biomarkers and bone growth across Tanner stages in perinatally HIV- exposed youth in PACTG 1045*. AIDS *In Progress*

Wang Z, Zheng Y, Zhao B, Zhang Y, **Liu Z**, Xu J, Chen Y, Yang Z, Wang F, Wang H, He J, Zhang R, Abliz Z. *Human Metabolic Responses to Chronic Environmental Polycyclic Aromatic Hydrocarbon Exposure by a Metabolomic Approach*. Journal of Proteome Research, 2015, 14 (6), pp 2583 - 2593

Liu Z, Zhang J, Zhao B, et al. *Population-based reference for birth weight for gestational age in northern China*. Early Human Development 2014;90(4):177-87.