Huizi Zhang

+44(0)7568208152 | zhanghuizi98@outlook.com

EDUCATION

The University of Edinburgh 09/2021 – Present

PhD in Statistics

- Supervisors: Dr Natalia Bochkina, Dr Sara Wade.
- Research Thesis: Bayesian Inference for Model-based Clustering and RNA Velocity in Single-cell Data.
- Developed a Bayesian nonparametric clustering model to identify cell subtypes using single-cell RNA sequencing data. Extending the hierarchical Dirichlet process, the method enables clustering across related groups and incorporation of external covariates flexibly using kernel methods.
- Constructed a Bayesian hierarchical model to infer RNA velocity, facilitating understanding of cellular dynamics and cell ordering with uncertainty.
- Designed scalable Bayesian methods for large sequencing data and compared with MCMC methods.
- Created custom R packages to implement tailored statistical algorithms.

Imperial College London 10/2020 – 10/2021

MSc in Statistics (Data Science)

- GPA: 83.41/100 (Distinction).
- Key Modules: Biomedical Statistics, Statistical Genetics and Bioinformatics, Machine Learning.

University of Glasgow 09/2018 – 06/2020

BSc in Statistics

- GPA: 21.7/22.0 (First Class Honours).
- Key Modules: Statistical Genetics, Biostatistics, Data Analysis, Multivariate Methods.

Zhongnan University of Economics and Law 09/2016 – 06/2018

Bachelor of Economics in Statistics

- GPA: 3.92/4.0.

PUBLICATIONS

Under Review:

- Zhang, H., Bochkina, N. and Wade, S., 2024. Quantifying uncertainty in RNA velocity.

In Preparation:

- Zhang, H., Wade, S. and Bochkina, N., 2024. Covariate-dependent hierarchical Dirichlet processes.

Published:

- Zhang, H., Swallow, B. and Gupta, M., 2022. Bayesian hierarchical mixture models for detecting non-normal clusters applied to noisy genomic and environmental datasets. Australian & New Zealand Journal of Statistics, 64(2), pp.313-337.

CONFERENCES

- July 2024, International Society for Bayesian Analysis, Venice, Italy. Poster presentation. "Covariate-dependent hierarchical Dirichlet process".
- June 2024, Bayesian Young Statisticians Meeting, Venice, Italy. Contributed talk. "Bayesian modelling of RNA velocity from single-cell RNA sequencing data".
- June 2024, Centre for Statistics, Edinburgh, United Kingdom. Poster presentation. "Covariate-dependent hierarchical Dirichlet process".
- June 2024, Early Career Researchers Day, Centre for Statistics, Edinburgh, United Kingdom. Contributed talk. "Covariate-dependent hierarchical Dirichlet process for cluster analysis".
- June 2024, BioInference, Warwick, United Kingdom. Contributed talk. "Bayesian modelling of RNA velocity from single-cell RNA sequencing data".

- July 2023, European meeting of Statisticians, Warsaw, Poland. Poster presentation. "Incorporating covariates in single-cell clustering with dependent hierarchical Dirichlet processes".
- November 2022, CSHL biological data science meeting, online. Poster presentation. "Bayesian modelling of transcriptional dynamics".

TEACHING/WORK EXPERIENCES

Tutor 09/2022 – Present

The University of Edinburgh

- Provided tutoring for various courses: Probability, Bayesian Data Analysis, Statistical Methodology, Multivariate Data Analysis, receiving positive feedback from most students.
- Assisted students in understanding key concepts in statistics, learning programming languages and popular R packages (rjags, Rstan and INLA), and preparing for coursework and exams.

Data Centre Intern 07/2020 – 08/2020

Industrial and Commercial Bank of China

- Worked on setting self-learning strategy during network maintenance period and awarded “Excellent Intern”.
- Employed statistical models for predicting daily maximum network traffic in R and Python.
- Implemented unsupervised methods for anomaly detection, identifying abrupt changes in network traffic data.

PROJECTS

Implementing Non-Convex Optimization Algorithms 07/2024

Generative Modelling Hackathon

- Implemented data augmentation and optimization algorithms for improved generalization to unseen data using PyTorch.
- Trained the MobileNetV2 model on image classification task and achieved an accuracy of 80.51% on the Food101 dataset under limited time, computational power and number of parameters.

Modelling Soil and Crop Nutrient with Earth Observation/Satellite Data 04/2023

IntelliDigest, Modelling Camp

- Performed feature engineering, combining raw soil data and satellite data.
- Employed machine learning methods to predict the pH value of the soil.
- Informed decision-making in bio-nutrient food production.
- Key achievement: the group project was chosen as the winning project among 5 projects.

Investigating Protein Localization in Single-Cell Data with Machine Learning 05/2021 – 09/2021

MSc Project, Imperial College London

- Conducted thorough exploratory data analysis using kernel trick methods.
- Analyzed the effects of drugs on cell shape and proliferation, and localization of a protein “YAP” to nucleus.
- Developed classification and regression models to identify crucially affected cellular features.

Model-Based Tight Clustering for Large Datasets 01/2020 – 03/2020

BSc Project, University of Glasgow

- Developed Bayesian clustering algorithms to obtain tight and stable clusters in the presence of noise.
- Compared the proposed method with state-of-the-art clustering methods using HPC.
- Received higher accuracy from the proposed model on image segmentation and genomic data.

SKILLS

Programming Languages and Software:

- Proficient in R (dplyr, tidyverse, coda).
- Advanced intermediate in Python (numpy, pandas, scikit-learn) and JAGS.
- Some experience with SQL, STAN, SAS.

Frameworks and Tools:

- Linux Bash, LaTeX, Git, R Markdown, Rcpp, R Shiny, Hadoop, Spark, Microsoft Office Software.

Machine Learning and Data Analysis:

- Confident user of clustering methods (K-means, Nonparametric Model-Based Clustering), ensemble methods (Random Forests), generalized linear models.
- Thorough knowledge of Bayesian methods, Markov chain Monte Carlo, uncertainty quantification.
- Proficient in data visualization tools including ggplot2, matplotlib.

Languages:

- English: Fluent (IELTS 8.0, CET 6).
- Chinese: Native.

AWARDS & GRANTS

Laura Wisewell Travel Fund, The University of Edinburgh	2024
Laura Wisewell Travel Fund, The University of Edinburgh	2023
School of Mathematics Studentship, The University of Edinburgh	2021 – 2025
Royal Statistical Society Prize for academic performance, Imperial College London	2021
Royal Statistical Society Prize for academic performance, University of Glasgow	2020
SAS Prize for academic performance in the 3rd year, University of Glasgow	2019
China National Scholarship	2018