

Hao Chai

Research Interests

- High-dimensional data analysis and inference, variable selection, survival analysis, network analysis.
- Various applications in clinical trials, high-throughput “-omics” data in oncology, gene-environment interactions, and other complex human biology and public health data.

Education

2009–2014 **Ph.D. (Statistics)**, *University of Iowa*, Iowa City, IA, USA, GPA: 3.96/4.0.

Advisor: Dr. Jian Huang

Thesis: Statistical Inference for Treatment Effects in High Dimensional AFT Models

Courses: Design and analysis of experiments, clinical trials, computational biology, statistical genetics, survival analysis, linear models, categorical data analysis, spatial statistics, advanced inference, time series, Bayesian statistics, statistical intensive computing.

2005–2009 **B.S. (Statistics and Financial Mathematics)**, *Beijing Normal University*, Beijing, China.

Professional Experience

2014–Present **Postdoctoral Associate**, *Dept. of Biostatistics, Yale University*, New Haven, CT, USA.

Collaborate with Yale scientists to identify genetic risk factors for different types of cancers including melanoma and thyroid cancer. Perform quality control and analysis on multi-platform data of size larger than 10GB. Develop cutting edge statistical methodologies for epigenetic data and other complex biomedical data.

Spring 2009 **Intern Statistician**, *Dept. of Biometrics, Fuwai Cardiovascular Hospital*, Beijing, China.

Designed randomized clinical trials to test bioequivalence between two drugs. Created SAS macros to automatically generate Phase I clinical trial reports.

Projects

2016 **Identification of genetic risk factors for melanoma**, *Yale University*.

Collaborated with pathologists and dermatologists from Yale cancer center to identify potential genetic biomarkers for melanoma. Explored the causal relationship between germ line and somatic mutations. Analyzed the statistical association between gene expressions and number of DNA mutations under various biological assumptions using parametric and non-parametric tests. Discovered mutations that drew great interest from oncologists and pathologists.

2015 **Genome wide association study of thyroid cancer**, *Yale University*.

Refined research hypothesis with scientists from various disciplines at Yale. Performed quality control and data processing using Linux shell and Python on melanoma data sets of size over 10GB. Conducted genome wide association study for 4.5 million SNPs. Identified biomarkers that could affect tumor growth rate for future investigation.

2013 **Study of spatio-temporal models with missing data**, *University of Iowa*.

Systematically studied common missing mechanisms in spatio-temporal models. Proposed a procedure that imputes missing data and estimates model parameters simultaneously by combining the generalized and conditional EM algorithms. Established the convergence and the extensibility of the new procedure.

2012 **Analysis of time-dependent microarray data**, *University of Iowa*.

Processed and analyzed sequentially collected microarray data using robust multi-array analysis and time series analysis. Parallelized the computation by using R packages *snow* and *multicore* and substantially reduced the computational time.

2011 **Permutation-based analysis of observational data**, *University of Iowa*.

Proposed a group-wise permutation method to discover the causal relationship between a few selected key factors and multiple responses. Applied the method to ancient skull thickness data and found the relationship between skull thickness and race. Visualized significant findings in a 3-D interactive animation to anthropologists.

2009 **Analysis of online study involvement data**, *Beijing Normal University*.

Collected and processed students' weekly online study involvement data. Analyzed the data using zero-inflated Poisson regression. Developed a MATLAB scoring system to automatically evaluate students' performance.

Publications

Chai, H., Shi, X. (2017). Analysis of Cancer Gene Expression Data with an Assisted Robust Marker Identification Approach. *Genetic Epidemiology*.

Lazova R, Pornputtapong N, Halaban R, Bosenberg M, Bai Y, **Chai H**, Krauthammer M. (2017). Spitz nevi and Spitzoid melanomas: exome sequencing and comparison with conventional melanocytic nevi and melanomas. *Modern Pathology*, 30(5):640-649, <http://dx.doi.org/10.1038/modpathol.2016.237>.

Chai, H., Ma, S., and Huang, J. (2016). Statistical Inference in High Dimensional Accelerated Failure Time Model. *Submitted to Statistica Sinica*.

Zang, Y., Zhao, Y., Zhang, Q., **Chai, H.**, Zhang, S., and Ma, S. (2016). Identifying Gene-Environment Interactions with A Least Relative Error Approach. *Proceeding Book for 2015 ICSA Symposium*.

Chai, H., Zhang, Q., Jiang, Y., Wang, G., Zhang, S., Ahmed, S., and Ma, S. (2016). Identifying Gene-Environment Interactions for Prognosis Using a Robust Approach. *Econometrics and Statistics*, ISSN 2452-3062, <http://dx.doi.org/10.1016/j.ecosta.2016.10.004>.

Chai, H., Zhang, Q., Jiang, Y., Huang, J., Wang, G., Zhang, S., Jones, C, and Ma, S. (2015). Robust Identification of Gene-Environment Interactions under High-Dimensional AFT Models. *Submitted to Statistica Sinica*.

Hu, J., **Chai, H.** (2013). Adjusted regularized estimation in the accelerated failure time model with high dimensional covariates. *Journal of Multivariate Analysis*, 122, 96-114.

Conferences Presentations

June 2015 **Statistical Society of Canada 2015 Annual Meeting**, (*invited*), Halifax, NS, Canada.
Topic: Identifying Gene-Environment Interactions for Prognosis in Integrative Analysis.

August 2014 **2014 Joint Statistical Meetings**, Boston, MA, USA.
Topic: Statistical Inference in High Dimensional Accelerated Failure Time Model.

Service

Referee: BMC Bioinformatics, Annals of the Institute of Statistical Mathematics, Science China.

Teaching Experience

Spring 2014 Statistical Inference II (Graduate level), University of Iowa.

Fall 2013 Statistical Inference I (Graduate level), University of Iowa.

Spring 2013 Statistics in Engineering, University of Iowa.

Fall 2012 Statistics in Business, University of Iowa.

Awards and Honors

2014 University of Iowa Strategic Initiative Fund Scholarship.

2013 Henry L. Rietz Award for outstanding Ph.D. students.

Technical Skills

Programming Languages: C, Python, Linux Shell, OpenCL, Pascal

Statistical Software: R, SAS, MATLAB

Genetics: PLINK, Haploview

Others: Microsoft Office, MapReduce, SQL, Latex

Languages

English, Chinese.

Reference

Shuangge Steven Ma, *Associate Professor*, Dept. of Biostatistics, Yale University.

Office: (203)785-3119.

Email: shuangge.ma@yale.edu.

Address: 60 College Street, Ste 201, New Haven, CT 06510.

Jian Huang, *Professor*, Dept. of Statistics and Actuarial Science, University of Iowa.

Office: (319)335-0712.

Email: jian.huang@yale.edu.

Address: 241 Schaeffer Hall, University of Iowa, Iowa City, IA 52242.

Joseph B. Lang, *Professor*, Dept. of Statistics and Actuarial Science, University of Iowa.

Office: (319)335-3129.

Email: joseph-lang@uiowa.edu.

Address: 207 Schaeffer Hall, University of Iowa, Iowa City, IA 52242.