

# Ray Bai

LeConte College 207  
1523 Greene St  
Columbia, SC 29208

Email: [rbai@mailbox.sc.edu](mailto:rbai@mailbox.sc.edu)  
Homepage: <http://www.raybai.net>  
Cell: (617) 953-3266

## Academic Employment

- 2020- Assistant Professor of Statistics, University of South Carolina
- 2018-2020 Postdoctoral Researcher in Biostatistics and Informatics, University of Pennsylvania  
Supervisors: Yong Chen and Mary Boland

## Education

- 2014-2018 PhD in Statistics, University of Florida  
Dissertation: Bayesian High-Dimensional Models with Scale-Mixture Shrinkage Priors  
Advisor: Malay Ghosh
- 2014-2016 Master of Statistics, University of Florida  
Project: Irrigation Forecasting in Southwest Florida  
Advisor: Nikolay Bliznyuk
- 2010-2012 MS in Applied Mathematics, University of Massachusetts Amherst
- 2003-2007 BA in Economics and Government, Cornell University

## Refereed Publications

( \_\_\_\_ = student author, \* = co-first author )

1. Wang, S., Shin, M., and **Bai, R.** (2024+). Generative quantile regression with variability penalty. *Journal of Computational and Graphical Statistics* (in press).
2. Wang, S., Shin, M., and **Bai, R.** (2024). Fast bootstrapping nonparametric maximum likelihood for latent mixture models. *IEEE Signal Processing Letters*, **31**: 870-874.
3. **Bai, R.**, Boland, M. R., and Chen, Y. (2023). Scalable high-dimensional Bayesian varying coefficient models with unknown within-subject covariance. *Journal of Machine Learning Research*, **24**(259): 1-49.
4. **Bai, R.\***, Moran, G. E.\*, Antonelli, J. L.\*, Chen, Y., and Boland, M. R. (2022). Spike-and-slab group lassos for grouped regression and sparse generalized additive models. *Journal of the American Statistical Association*, **117**(537): 184-197.
5. Meeker, J. R., Burris, H. H., **Bai, R.**, Levine, L. D., and Boland, M. R. (2022). Neighborhood deprivation increases the risk of post-induction cesarean delivery. *Journal of the American Medical Informatics Association*, **29**(2): 329-334.
6. **Bai, R.** and Ghosh, M. (2021). On the beta prime prior for scale parameters in high-dimensional Bayesian regression models. *Statistica Sinica*, **31**(2): 843-865.

7. Meeker, J. R., Canelón, S. P., **Bai, R.**, Levine, L. D., and Boland, M. R. (2021). Individual- and neighborhood-level risk factors for severe maternal morbidity. *Obstetrics & Gynecology*, **137**(5): 847-854.
8. Boland, M. R., Liu, J., Balocchi, C., Meeker, J., **Bai, R.**, Mellis, I., Mowery, D. L., and Herman, D. (2021). Association of neighborhood-level factors and COVID-19 infection patterns in Philadelphia using spatial regression. *AMIA Annual Symposium Proceedings*, **2021**: 545-554.
9. **Bai, R.** and Ghosh, M. (2019). Large-scale multiple hypothesis testing with the normal-beta prime prior. *Statistics*, **53**(6): 1210-1233.
10. **Bai, R.** and Ghosh, M. (2018). High-dimensional multivariate posterior consistency under global-local shrinkage priors. *Journal of Multivariate Analysis*, **167**: 157-170.
11. Duerr, I., Merrill, H. R., Wang, C., **Bai, R.**, Boyer, M., Dukes, M. D., and Bliznyuk, N. (2018). Forecasting urban water demand with statistical and machine learning methods using large space-time data: A comparative study. *Environmental Modelling and Software*, **102**: 29-38.

## Chapters in Edited Volumes

12. **Bai, R.**, Ročková, V., and George, E. I. (2021). Spike-and-slab meets LASSO: A review of the spike-and-slab LASSO. In Tadesse, M. G. and Vannucci, M. (Eds.), *Handbook of Bayesian Variable Selection*, 81-108. Chapman & Hall/CRC Press.

## Technical Reports

( \_\_\_\_ = student author, \* = co-first author )

1. Liu, Q., Huang, X., and **Bai, R.** (2024+). Bayesian modal regression based on mixture distributions. arXiv:2211.10776. *Under revision*.
2. Wang, S.-H., **Bai, R.**, and Huang, H.-H. (2024+). Two-step mixed-type multivariate Bayesian sparse variable selection with shrinkage priors. arXiv:2201.12839. *Under revision*.
3. Deshpande, S. K., **Bai, R.**, Balocchi, C., Starling, J. E., and Weiss, J. (2024+). VCBART: Bayesian trees for varying coefficients. arXiv:2003.06416. *Under revision*.
4. Zgodic, A., **Bai, R.**, Zhang, J., Olejua, P., and McLain, A. C. (2024+). Sparse high-dimensional linear mixed modeling with a partitioned empirical Bayes ECM algorithm. arXiv:2310.12285. *Under revision*.
5. **Bai, R.** (2024+). Bayesian group regularization in generalized linear models with a continuous spike-and-slab prior. arXiv:2007.07021. *Under revision*.
6. **Bai, R.**, Lin, L., Boland, M. R., and Chen, Y. (2024+). A robust Bayesian Copas selection model for correcting and quantifying the impact of publication bias. arXiv:2005.02930. *Under revision*.
7. **Bai, R.\***, Liu, X.\*, Lin, L., Liu, Y., Kimmel, S. E., Chu, H., and Chen, Y. (2024+). A Bayesian selection model for correcting outcome reporting bias with application to a meta-analysis on heart failure interventions. arXiv:2110.08849. *Under revision*.
8. Zgodic, A., **Bai, R.**, Zhang, J., Wang, Y., Rorden, C., and McLain, A. C. (2024+). Quantifying predictive uncertainty of aphasia severity in stroke patients with sparse heteroscedastic Bayesian high-dimensional regression. arXiv:2309.08783. *Submitted*.

9. Wang, S., Chakraborty, S., Qin, Q., and **Bai, R.** (2024+). Neural-g: A deep learning framework for mixing density estimation. arXiv:2406.05986. *Submitted*.
10. Zhao, Z., Li, Y., Luo, X., and **Bai, R.** (2024+). A unified three-state model framework for analysis of treatment crossover in survival trials. arXiv:2401.17008. *Submitted*.
11. Balocchi, C.\* , **Bai, R.\***, Liu, J., Canelón, S. P., George, E. I., Chen, Y., and Boland, M. R. (2024+). Quantifying patient and neighborhood risks for stillbirth and preterm birth in Philadelphia with a Bayesian spatial model. arXiv:2105.04981. *Submitted*.
12. Zhao, Z., Srivastava, S., Bandyopadhyay, D., and **Bai, R.** (2024+). Bayesian semiparametric joint modeling of cluster size and survival time for kidney transplantation. *Working paper*.
13. Fan, S., Xiong, L., Wang, D., Cai, G., and **Bai, R.** (2024+). Simultaneous drug discovery and drug clustering with spike-and-slab lasso matrix completion. *Working paper*.
14. **Bai, R.** (2024+). Adaptive posterior contraction for high-dimensional Bayesian varying coefficient models under shrinkage priors. *Working paper*.

## Software Development

(\* I am the primary author and maintainer of the following software products)

SSGL (R package): Spike-and-Slab Group Lasso for Group-Regularized Generalized Linear Models

MBSP (R package): Multivariate Bayesian Model with Shrinkage Priors

NVCSSL (R package): Nonparametric Varying Coefficient Spike-and-Slab Lasso

MtMBSP (R package): Mixed-type Multivariate Bayesian Regression with shrinkage Priors

NormalBetaPrime (R package): Normal Beta Prime Prior

## Advising

### *Past PhD Students*

Shijie Wang (PhD 2024)

*Dissertation:* New Deep Learning Approaches to Classical Statistical Problems

*First job:* Applied Scientist, Gauss Labs (Palo Alto, CA)

**Winner of James D. Lynch Graduate Student Research Award**

### *Current PhD Students*

Zile Zhao (expected 2024)

*Dissertation:* Methods and Applications for Bayesian Semiparametric Survival Analysis

Sijian Fan (expected 2025)

*Member of Doctoral Committee***Department of Statistics**

Qingyuan Hong (PhD expected 2025), Jihyun Kim (PhD expected 2024), Xin Zhi (PhD expected 2024), Yuchen Mao (PhD 2024), Zehao Yu (PhD 2024), Qingyang Liu (PhD 2023), Yang He (PhD 2023), Shan Zhong (PhD 2022), Zhen Yang (PhD 2022)

**Other Departments**

Anja Zgodic (PhD in Biostatistics, 2023), Chrisogonas Odhiambo (PhD in Computer Science, 2022)

*Undergraduate Senior Thesis*

Second reader for Harrison Cassel (BS in Statistics, 2024)

*Other Student Supervisions\**

(\* I mentored these students on research projects but was not their primary supervisor)

Qingyang Liu (PhD in Statistics, 2023, University of South Carolina)

Anja Zgodic (PhD in Biostatistics, 2023, University of South Carolina)

**Winner of ENAR Distinguished Student Paper Award**

Hung-Tien Huang (BS in Computer Science, 2023, University of South Carolina)

Xin Zhi (PhD candidate in Statistics at University of South Carolina)

Liyan Xiong (PhD candidate in Biostatistics at University of South Carolina)

Dayuan Wang (PhD candidate in Biostatistics at University of Florida)

**Funding***External Grants*

National Science Foundation, Division of Mathematical Sciences, Award 2015528 (May 2022-May 2024). Collaborative Research: Bayesian and Semi-Bayesian Methods for Detecting Relationships in High Dimensions. Role: **PI**. Amount: \$93,831.

National Science Foundation, Division of Mathematical Sciences, Award 2038080 (August 2021-July 2026). RTG: Mathematical Foundation of Data Science at University of South Carolina. PI: Linyuan Lu, University of South Carolina. Role: **Senior Personnel**. Amount awarded to RB: \$10,932.

*Internal Grants*

University of South Carolina McCausland Fellowship Research Award (August 2023-August 2024). New Methods for Bayesian Semiparametric Survival Analysis. Role: **PI**. Amount: \$10,000.

University of South Carolina Big Data Health Science Center Pilot Project grant (August 2023-August 2024). Leveraging Side Information for Improved Prediction and Inference in Computational Drug Repositioning. Role: **PI**. Amount: \$20,000.

University of South Carolina ASPIRE-I, Track 1 (July 2022-September 2023). Scalable Bayesian Survival Analysis with Informative Cluster Size. Role: **PI**. Amount: \$13,583.

## Unfunded Grant Submissions

National Science Foundation, Division of Mathematical Sciences. New Directions for Bayesian Inference in High Dimensions with Highly Structured Data. Role: **PI**. Funding timeline: August 2024-August 2027. Amount requested: \$239,308.

National Institutes of Health R01. DREAM: Data-driven Drug Repurposing for AD/ADRD Medications. PIs: Cui Tao, University of Texas Health Science Center at Houston, and Yong Chen, University of Pennsylvania. Role: **Subcontract PI**. Funding timeline: July 2023-June 2028. Amount requested for RB: \$250,000.

## Honors and Awards

McCausland Faculty Fellow, University of South Carolina College of Arts and Sciences, August 2023-May 2026

Graduate School Fellow, University of Florida, August 2014-August 2018

Student Paper Competition Award, Section on Bayesian Statistical Science, Joint Statistical Meetings, January 2018

Anderson Scholars Faculty Honoree, University of Florida, November 2016

## Teaching

*University of South Carolina*

### Undergraduate Courses Taught

STAT 515: Statistical Methods I, Fall 2024

STAT 517: Advanced Statistical Models, Fall 2022, Fall 2024

### Graduate Courses Taught

STAT 714: Linear Statistical Models, Fall 2020, Fall 2021, Fall 2022

STAT 718: High-Dimensional Data, Spring 2021, Spring 2023

STAT 721: Stochastic Processes, Spring 2022, Spring 2024

### Independent Studies Supervised

Kaniz Fatema (graduate student in Statistics), "Introduction to Causal Inference," Summer 2024

Dip Das (graduate student in Statistics), "Advanced Algorithms in Bayesian Computing," Spring 2024

Hung-Tien Huang (undergraduate student in Computer Science), "Advanced Topics in Stochastic Processes," Spring 2022

**Courses Developed and Revised**

STAT 718: High-Dimensional Data

New special topics class covering statistical methodology and algorithms for supervised learning, unsupervised learning, and deep learning

STAT 721: Stochastic Processes

Course revised to place more emphasis on contemporary topics and applications such as Bayesian nonparametrics, spatial statistics, reinforcement learning, and financial mathematics

*University of Florida***Courses Taught**

STA 3024: Introduction to Statistics II, Spring 2016

**Courses Served as Teaching Assistant**

STAT 2023: Introduction to Statistics I, Fall 2015

**Presentations***Invited Seminar Talks and Guest Lectures*

1. Department of Statistics & Operations Research, University of North Carolina at Chapel Hill, March 2024
2. South Carolina SmartState Center for Healthcare Quality, University of South Carolina, March 2023
3. Department of Statistics, University of Georgia, February 2023
4. School of Mathematics & Statistics, University of Glasgow, December 2022
5. Department of Statistics, Virginia Tech, November 2022
6. School of Mathematical and Statistical Sciences, Clemson University, October 2022
7. Department of Mathematical Sciences, University of Cincinnati, April 2022
8. Department of Biostatistics, Virginia Commonwealth University, November 2021
9. Department of Statistics, University of South Carolina, October 2021
10. School of Statistics, University of Minnesota, October 2021
11. Department of Statistics, University of California, Davis, April 2021
12. School of Mathematical and Statistical Sciences, Arizona State University, January 2020
13. Department of Statistics, Florida State University, January 2020
14. Department of Mathematics & Statistics, San Diego State University, January 2020
15. Department of Statistics, University of California, Santa Cruz, January 2020
16. Department of Statistics, University of South Carolina, January 2020

17. Graduate Group in Genomics and Computational Biology, University of Pennsylvania, December 2019
18. Statistics Student Seminar, University of Florida, April 2019
19. PennCIL Lab, University of Pennsylvania, March 2018

### *Invited Conference Talks*

1. 2024 ICSA Applied Statistics Symposium, Nashville, TN, June 2024
2. 2024 ICSA-Canada Chapter Symposium, Niagara Falls, Canada, June 2024
3. 2024 SRCOS Summer Research Conference, Clemson, SC, June 2024
4. 2023 Joint Statistical Meetings, Toronto, ON, Canada, August 2023
5. EcoSta 2023, Tokyo, Japan, August 2023
6. Invited poster session, 2022 Joint Statistical Meetings, Washington, DC, August 2022
7. EURO 2022 Conference, Espoo, Finland, July 2022 (canceled due to illness)
8. 2022 ICSA Applied Statistics Symposium, Gainesville, FL, June 2022
9. UP-STAT 2022 Hybrid Conference, Buffalo, NY, May 2022
10. CFE-CMStatistics 2021, London, UK, December 2021
11. Fifth EAC-ISBA Conference: A Satellite Meeting of the 2020 ISBA World Meeting in Celebrating James O Berger's 70th Birthday (virtual), November 2021
12. 2021 ICSA Applied Statistics Symposium (virtual), September 2021

### *Contributed Talks*

- 2024 ISBA World Meeting, Venice, Italy, July 2024
- 2021 Joint Statistical Meetings (virtual), August 2021
- 2021 ISBA World Meeting (virtual), June 2021
- 2019 Joint Statistical Meetings, Denver, CO, July 2019
- 2018 Joint Statistical Meetings, Vancouver, BC, Canada, July 2018

### *Contributed Conference Posters*

- Bayes Comp 2020, Gainesville, FL, January 2020
- DBEI and CCEB Research Day, University of Pennsylvania, March 2019
- O-Bayes 2017 Meeting, Austin, TX, December 2017

## Departmental Service

Colloquium Chair, 2023-2024, 2024-2025

Member of Graduate Committee, 2021-2022, 2022-2023, 2023-2024

Member of Hiring Committee, 2021-2022

Member of Data Science Degree Planning Committee, 2022

Member of PhD Qualification Exam Committee, 2021

## Service to the University of South Carolina

Member, ASPIRE Grant Review Committee, 2023

Panelist, LGBTQ+ Graduate Student Association's Queer in Higher Ed Panel, 2023

Member, SPARC Grant Review Committee, 2022

Lunch Ambassador for New Faculty Orientation, 2022

Member, Top Scholar Selection Committee, 2021-2022, 2022-2023

## Professional Service

### *Editorial Activities*

Journal reviewer for 51 papers: *Journal of the American Statistical Association* (5), *Journal of the Royal Statistical Society: Series B* (3), *Annals of Applied Statistics* (2), *Biometrics* (2), *Journal of Machine Learning Research* (2), *Journal of Computational and Graphical Statistics* (1), *Bernoulli* (1), *Statistica Sinica* (2), *IEEE Transactions on Information Theory* (1), *Electronic Journal of Statistics* (1), *Scandinavian Journal of Statistics* (3), *Bayesian Analysis* (3), *Statistics in Medicine* (4), *Journal of the Royal Statistical Society: Series C* (1), *Computational Statistics & Data Analysis* (1), *Statistics and Computing* (1), *Statistical Methods in Medical Research* (4), *Journal of Statistical Planning and Inference* (2), *Journal of Nonparametric Statistics* (1), *The American Statistician* (2), *Lifetime Data Analysis* (1), *Metrika* (1), *Statistical Modelling* (1), *Journal of Biopharmaceutical Statistics* (1), *The R Journal* (1), *Research Synthesis Methods* (1), *Science China Mathematics* (1), *Annals of the Brazilian Academy of Sciences* (1), *BMJ Open* (1)

Judge for SBSS Student Paper Competition, 2020, 2024

### *Grant Reviewing*

External grant reviewer for the National Science Foundation, 2020, 2021, 2022

Panelist for the National Science Foundation, 2022

### *Conference Activities*

Session chair, "Junior advances in Bayesian treed regression," ISBA 2022 World Meeting

Session chair, Section on Bayesian Statistical Science, Joint Statistical Meetings, 2021, 2022

Member of Program Committee, "Your Model is Wrong: Robustness and misspecification in probabilistic modeling," NeurIPS 2021 Workshop



## Professional Society Memberships

American Statistical Association

International Society for Bayesian Analysis

International Biometric Society

## Professional Development

### *Certificates*

New Faculty Academy Certificate of Completion, University of South Carolina, 2022

Graduate Certificate in Business Administration, Northeastern University, 2009

### *Workshop Attendance*

Propel Research Mentorship Program, University of South Carolina, 2022-2023

Perspectives in Statistical Modeling and Inference: A Workshop in Honor of Ed George's 70th Birthday, Philadelphia, PA, 2021

## Industry Experience

2012-2014 Systems Engineer, General Dynamics Mission Systems

2007-2010 Business Analyst, State Street Bank & Trust

## Computer Skills

R, C/C++, Python, MATLAB, Octave, Julia, JAGS, Stan, Windows, Linux