

Subha Guha

Department of Biostatistics
University of Florida

Phone: 352.294.5921
E-mail: s.guha@ufl.edu

RESEARCH INTERESTS Biostatistical methods, cancer genomics, causal inference, clustering and classification, computational methods for Big Data, connectomics, development and testing of novel medical devices, dental and craniofacial data, electronic health records, generalized linear models, health disparity research, hidden Markov models, high-dimensional inference, image processing, independent component analysis, MCMC simulation, microbiome data analysis, neuroscience, nonparametric Bayesian methods, observational studies, quality control and process monitoring, spatial statistics.

EDUCATION **Ohio State University**, Columbus, OH
Ph.D. in Statistics, June 2004
Indian Institute of Technology, Kanpur, India
M.Sc. in Statistics, June 1997

PROFESSIONAL AND ACADEMIC EXPERIENCE **University of Florida**, Department of Biostatistics, Gainesville, FL
Associate Professor **2018 - present**
University of Missouri, Department of Statistics, Columbia, MO
Associate Professor **2013 - 2017**
Assistant Professor **2007 - 2013**
Harvard School of Public Health, Department of Biostatistics, Boston, MA
Postdoctoral Research Fellow **2004 - 2007**
Ohio State University, Department of Statistics, Columbus, OH
Graduate Teaching Assistant **1999 - 2004**

RESEARCH GRANTS **National Science Foundation/National Institutes of Health**, *Collaborative Research: New Bayesian Nonparametric Paradigms of Personalized Medicine for Lung Cancer*, \$1,600,000, 9/1/2015 to 8/31/2020.

Principal Investigators: Subha Guha of University of Florida, Veera Baladandayuthapani of University of Michigan, and Ignacio Wistuba of M.D. Anderson Cancer Center. This body of work leverages the computing expertise of Marc Suchard of UCLA.

National Science Foundation, *Bayesian Mixture Models: Unified Theoretical Frameworks and MCMC Methods*, \$150,000, 7/1/2009 to 6/30/2013. Role: Principal Investigator.

RESEARCH
GRANTS
(CONTINUED)

Department of Health and Human Services, *Statistical Informatics for Cancer Research*, one month's summer salary, 2009 to 2013. Role: Co-investigator.

University of Missouri Research Council Committee, \$7,000, Summer Research Fellowship 2008.

HONORS AND
AWARDS

Top Faculty Achiever of the University of Missouri in February, 2016.

Faculty Fellowship Award of the Department of Statistics, University of Missouri in August, 2014.

Travel support awarded by organizers of the Seminar on Bayesian Inference in Econometrics and Statistics held in St. Louis, Missouri in 2005 and 2009; Valencia/ISBA World Meeting on Bayesian Statistics held in Alicante, Spain in 2006; 2nd IMS-ISBA Joint Meeting held in Bormio, Italy in 2005; Fourth International Workshop on Objective Prior Methodology held in Aussois, France in 2003.

Travel support awarded by Department of Biostatistics, Harvard School of Public Health, for attending ENAR Spring Meeting 2005, Joint Statistical Meetings 2005 and Joint Statistical Meetings 2006; ASA Section on Bayesian Statistical Science for attending Joint Statistical Meetings 2004; Graduate School and Department of Statistics, Ohio State University for attending Joint Statistical Meetings 2003.

Craig Cooley Memorial Prize for scholarly excellence and leadership qualities, 2004, Department of Statistics, Ohio State University.

Thomas and Jean Powers Teaching Award, 2003, Department of Statistics, Ohio State University.

Distinguished teaching assistant, Department of Statistics, Ohio State University, for four consecutive quarters.

Winner of the Student Paper Competition 2002 organized by ASA Statistical Computing and Graphics Sections.

EDITORIAL
WORK

Associate Editor, *Bayesian Analysis*, International Society for Bayesian Analysis, 2020–current.

ACADEMIC
SERVICE

Conference Organization

Organized a session of invited talks at the conference titled, *Ordered Data Analysis, Models and Health Research Methods: An International Conference in Honor of H.N. Nagaraja for His 60th Birthday*, held at University of Texas at Dallas, March 2014.

Grant Review

National Institutes of Health, Biostatistical Methods and Research Design (BMRD) Study Section

National Security Agency, Mathematical Sciences Grant Program

National Science Foundation

American Statistical Association

Organizer and chair, Invited Session titled *Innovative Approaches for High-dimensional Omics and Neuroimaging Data*, Joint Statistical Meetings 2019

Chair, Contributed Session on *Modern Statistical Methods for Multi-Scale and Time Series Data*, Joint Statistical Meetings 2017

Committee Member, Student Paper Award of the Statistics in Genetics and Genomics Section, Joint Statistical Meetings 2016

Committee Member, Student Paper Award of the Section on Bayesian Statistical Science, Joint Statistical Meetings 2011 and Joint Statistical Meetings 2013

International Society for Bayesian Analysis

Organizer, Invited Session titled *Innovative Bayesian Approaches for High-dimensional Variable Selection* at 4th Eastern Asia Meeting On Bayesian Statistics (EAC-ISBA), Kobe University, Japan, July 2019

Member, Savage Award Committee, 2016 and 2018

International Indian Statistical Association

Organizer and chair, Invited Session titled *Some Innovative Applications in High-dimensional Data Analysis*, Innovations in Data and Statistical Sciences held in Mumbai, India, December 2019

Organizer, Invited Session titled *Novel Approaches for Genomics and Censored Outcome Data* at Conference on Current Trends in Survey Statistics held at Institute of Mathematical Sciences, National University of Singapore, August 2019

Organizer, Invited Session titled *Some Innovative Applications in High-dimensional Data Analysis* at INDSTATS 2019 – Innovations in Data and Statistical Sciences, Mumbai, India, December 2019

Computational and Methodological Statistics

Organizer, Invited Session titled *Flexible Bayesian approaches in high-dimensional applications* at CMStatistics 2020 – 13th International Conference of the ERCIM WG on Computational and Methodological Statistics, London, UK, December 2020

International Biometric Society

Organizer, Invited Session titled *Innovative Statistical Approaches for High-dimensional Omic and Microbiomic Data*, ENAR 2020 Spring Meeting held in Nashville, Tennessee, March 2020

University of Florida

2020–current Director, Biostatistics Consulting Lab
2019–current Member, Departmental Awards Committee
2018–current Chair of a Ph.D. Committee
2018–current Member of 3 Ph.D. Committees (Biostatistics and Astronomy)
2020 Poster judge, PPHP Annual Research Day
2019 UF Marshal for Spring commencement

University of Missouri

2012–2017 Chair of 4 Ph.D. Committees
2007–2017 Member of 15 Ph.D. and 4 Master’s Thesis Committees
2016–2017 Chair, Qualifying Exam Committee
2016–2017 Chair, Computing Committee
2012–2013 University of Missouri Research Board
2011–2015 Graduate Admissions Committee
2011 Chair, Bayesian Discussion Group
2011–2014 Faculty Search Committee
2010–2012 Advisor, Minor in Psychological Statistics and Methods
2009–2014 Member, Qualifying/Preliminary Exam Committee
2008–2011 Seminar Series Chair
2008 Co-organizer, Winemiller Conference

PROFESSIONAL MEMBERSHIPS American Statistical Association
International Biometric Society
International Indian Statistical Association
International Society for Bayesian Analysis
Institute of Mathematical Statistics

JOURNAL REVIEWER FOR Australian and New Zealand Journal of Statistics, Bayesian Analysis, Bioinformatics, BMC Bioinformatics, Biometrics, Biostatistics, Communications in Statistics - Simulation and Computation, Computational Statistics and Data Analysis, Journal of the American Statistical Association, Journal of Applied Statistics, Journal of Computational and Graphical Statistics, Journal of the Royal Statistical Society Series B, Journal of Statistical Computation and Simulation, Quality Technology and Quantitative Management, Statistical Methods and Applications, Statistica Sinica, Statistics in Medicine

PUBLICATIONS 1. Anyaso-Samuel, S., Sachdeva, A., **Guha, S.**, and Datta, S. (2021) Metagenomic Geolocation Prediction Using an Adaptive Ensemble Classifier. *Frontiers*

in Genetics, section Computational Genomics, 12:642282.

2. Anyaso-Samuel, S., Sachdeva, A., **Guha, S.**, and Datta, S. (2021) Bioinformatics Pre-processing of Microbiome Data with an Application to Metagenomic Forensics. *Statistical Analysis of Microbiome Data* (eds. Datta, S. and Guha, S.), Springer Nature, Switzerland.
3. **Guha, S.**, and Datta, S. (2021) A Bayesian Approach to Restoring the Duality between Principal Components of a Distance Matrix and Operational Taxonomic Units in Microbiome Analyses. *Statistical Analysis of Microbiome Data* (eds. Datta, S. and Guha, S.), Springer Nature, Switzerland.
4. **Guha, S.** and Ghosh, S. K. (2020) Probabilistic Detection and Estimation of Conic Sections from Noisy Data. *Journal of Computational and Graphical Statistics*, 29, 513–522.
5. Yan, D., **Guha, S.**, Ahn, C., and Tiwari, R. (2020) Semiparametric Bayesian Markov Analysis of Personalized Benefit-Risk Assessment. *Annals of Applied Statistics*, 14, 768–788.
6. Drusbosky, L.M, Singh, N. K., Hawkin, K. E., et al (2019). A genomics-informed computational biology platform prospectively predicts treatment responses in AML and MDS patients. *Blood Advances*, 3(12), 1837–1847.
7. Jha, C., Li, Y. and **Guha, S.** (2017). Semiparametric Bayesian Analysis of High-Dimensional Censored Outcome Data. *Statistical Theory and Related Fields*, 1, 194–204.
8. **Guha, S.** and Baladandayuthapani, V. (2016). A Nonparametric Bayesian Technique for High-Dimensional Regression. *Electronic Journal of Statistics*, 10, 3374–3424.
9. **Guha, S.**, Banerjee, S., Gu, C. and Baladandayuthapani, V. (2015). Nonparametric Variable Selection, Clustering and Prediction for Large Biological Datasets. *Nonparametric Bayesian Inference in Biostatistics* (eds. Mitra, R. and Müller, P.), Springer International Publishing.
10. Cui, S., **Guha, S.**, Ferreira, M. A. R. and Tegge, A. N. (2015). A Hidden Markov Model for Detecting Differentially Expressed Genes from RNA-Seq Data. *Annals of Applied Statistics*, 9, 901–925.

11. **Guha, S.**, Ji, Y., and Baladandayuthapani, V. (2014). Bayesian Disease Classification using Copy Number Data. *Cancer Informatics*, 13(S2), 83–91.
12. **Guha, S.** (2011). Discussion of *Sampling schemes for generalized linear Dirichlet process random effects models* by Kyung, Gill and Casella. *Statistical Methods & Applications*, 20, 291–293.
13. MacEachern, S. N. and **Guha, S.** (2010). Parametric and Semiparametric Hypotheses in the Linear Model. *The Canadian Journal of Statistics*, 39, 165–180.
14. **Guha, S.** (2010). Posterior Simulation in Countable Mixture Models for Large Datasets. *Journal of the American Statistical Association*, 105, 775–786.
15. **Guha, S.** (2010). Bayesian Hidden Markov Modeling of Array CGH Data. *Bayesian Modeling in Bioinformatics* (eds. Dey, D. K., Ghosh, S. and Mallick, B.), Chapman & Hall/CRC.
16. **Guha, S.**, Ryan, L. and Morara, M. (2009). Gauss-Seidel Estimation of Generalized Linear Mixed Models with Application to Poisson Modeling of Spatially Varying Disease Rates. *Journal of Computational and Graphical Statistics*, 18, 818–837.
17. **Guha, S.** (2008). Posterior Simulation in the Generalized Linear Mixed Model with Semiparametric Random Effects. *Journal of Computational and Graphical Statistics*, 17, 410–425.
18. **Guha, S.**, Li, Y. and Neuberg, D. (2008). Bayesian Hidden Markov Modeling of Array CGH Data. *Journal of the American Statistical Association*, 103, 485–497.
19. **Guha, S.** and MacEachern, S. N. (2006). Generalized Post-stratification and Importance Sampling for Subsampled Markov Chain Monte Carlo Estimation. *Journal of the American Statistical Association*, 101, 1175–1184.
20. Li, Y., Tiwari R., and **Guha, S.** (2006). Mixture Cure Survival Models with Dependent Censoring. *Journal of the Royal Statistical Society - Series B*, 69, 285–306.
21. Burden, S., **Guha, S.**, Morgan, G., Ryan, L. Sparks, G. and Young, L. (2005). Spatio-temporal Analysis of Ischemic Heart Disease in NSW, Australia. *Environmental and Ecological Statistics*, 12, 427–448.

22. **Guha, S.**, MacEachern, S. N. and Peruggia, M. (2004). Benchmark Estimation for Markov Chain Monte Carlo Samples. *Journal of Computational and Graphical Statistics*, 13, 683–701.
23. MacEachern, S. N., Peruggia, M. and **Guha, S.** (2003). Discussion of *A theory of statistical models for Monte Carlo integration* by Kong, McCullagh, Nicolae, Tan and Meng. *Journal of the Royal Statistical Society - Series B*, 65, 612.

EDITED
BOOKS

Datta, S. and **Guha, S.** (Eds.) (2021). *Statistical Analysis of Microbiome Data*, Frontiers in Probability and the Statistical Sciences, Springer Nature, Switzerland.

SUBMITTED
PAPERS

1. Gu, C., Baladandayuthapani, V., and **Guha, S.** Nonparametric Bayes Differential Analysis for Dependent Multigroup Data with Application to DNA Methylation Analyses in Cancer. Invited revision by *Journal of the American Statistical Association*.
– **Recognized through the 2018 SBSS Student Paper Competition Award to the first author, Chiyu Gu.**
2. Sachdeva, A., Tiwari, R. and **Guha, S.** A Novel Approach to Augment Single-Arm Clinical Studies with Real-World Data. Invited revision by *Journal of Biopharmaceutical Statistics*.
3. **Guha, S.**, Jung, R. and Dunson, D. Predicting Phenotypes from Brain Connection Structure. Invited revision by *Journal of the Royal Statistical Society: Series C*.

WORK IN
PROGRESS

1. **Guha, S.** and Qiu, P. Bayesian Pairwise Comparison of High-Dimensional Images.
2. Sachdeva, A., Wu, M. and **Guha, S.** Bayesian Identification of Phenotypic Signatures using Microbial Balances.
3. Yan, D. and **Guha, S.** Flexible Prediction and Clustering Models for Integrating Information in Multi-Domain Genomic Data.
4. **Guha, S.** and Ghosh, S. A Nonparametric Bayesian Approach to Nonlinear Independent Component Analysis.

5. Ghosh, K. and **Guha, S.** Bayesian Analysis of Diffusion Tractography with Multiple Fibre Orientations.

CONFERENCE
ABSTRACTS

1. Gu, C., Baladandayuthapani, V., and **Guha, S.** (2020). Bayesian Nonparametric Differential Analysis for Dependent Multigroup Data with Application to DNA Methylation Analyses in Cancer. *ENAR 2020 Spring Meeting (virtual), March 2020.*
2. **Guha, S.** (2020). Scalable Methods for Posterior Inference from Big Data. *Bayes Comp 2020, Gainesville, Florida, January 2020.*
3. Gu, C., Baladandayuthapani, V., and **Guha, S.** (2019). Bayesian Nonparametric Differential Analysis for Dependent Multigroup Data with Application to DNA Methylation Analyses in Cancer. *Innovations in Data and Statistical Sciences, Mumbai, India, December 2019.*
4. Gu, C., Baladandayuthapani, V., and **Guha, S.** (2019). Bayesian Nonparametric Differential Analysis for Dependent Multigroup Data with Application to DNA Methylation Analyses in Cancer. *Current Trends in Survey Statistics, Singapore, August 2019.*
5. Gu, C., Baladandayuthapani, V., and **Guha, S.** (2019). Bayesian Nonparametric Differential Analysis for Dependent Multigroup Data with Application to DNA Methylation Analyses in Cancer. *4th Eastern Asia Meeting On Bayesian Statistics (EAC-ISBA), Kobe, Japan, July 2019.*
6. Gu, C., Baladandayuthapani, V., and **Guha, S.** (2019). Bayesian Nonparametric Differential Analysis for Dependent Multigroup Data with Application to DNA Methylation Analyses in Cancer. *ENAR Spring Meeting, Philadelphia, Pennsylvania, March 2019.*
7. Gu, C., Baladandayuthapani, V., and **Guha, S.** (2019). Bayesian Nonparametric Differential Analysis for Dependent Multigroup Data with Application to DNA Methylation Analyses in Cancer. *Tenth International Triennial Calcutta Symposium on Probability and Statistics, Kolkata, India, December 2018.*
8. **Guha, S.** and Baladandayuthapani, V. (2018). A Nonparametric Bayesian Technique For High-Dimensional Regression. *12th International Interdisciplinary Conference MSAST 2018, Kolkata, India, December 2018.*
9. Gu, C., Baladandayuthapani, V., and **Guha, S.** (2018). Bayesian Nonparametric Differential Analysis for Dependent Multigroup Data with Application to DNA Methylation Analyses in Cancer. *11th International Conference of the ERCIM WG on Computational and Methodological Statistics, Pisa, Italy, December 2018.*
10. Gu, C., Baladandayuthapani, V., and **Guha, S.** (2018). Bayesian Nonparametric Differential Analysis for Dependent Multigroup Data with Application to DNA Methylation Analyses in Cancer. *IISA International Conference in Statistics held in Gainesville, Florida.*

11. **Guha, S.** and Baladandayuthapani, V. (2017). A Nonparametric Bayesian Technique For High-Dimensional Regression. *NBER-NSF Seminar on Bayesian Inference in Econometrics and Statistics, Washington University in St. Louis, Missouri.*
12. Gu, C., **Guha, S.** and Baladandayuthapani, V. (2017). Scalable Bayesian Nonparametric Learning for High-Dimensional Lung Cancer Genomics Data. *ENAR Spring meeting held in Washington, DC.*
13. Jha, C., Li, Y. and **Guha, S.** (2016). Semiparametric Bayesian Analysis of High-Dimensional Censored Outcome Data: Discovering Spatial Variation of Breast Cancer Mortality Rates in New Mexico. *Joint Statistical Meetings held in Chicago, Illinois.*
14. Gu, C., **Guha, S.** and Baladandayuthapani, V. (2015). Scalable Bayesian Nonparametric Learning for High-Dimensional Lung Cancer Genomics Data. *International Chinese Statistical Association (ICSA) Applied Statistical Symposium held in Atlanta, GA.*
15. Gu, C., **Guha, S.** and Baladandayuthapani, V. (2015). Scalable Bayesian Nonparametric Learning for High-Dimensional Lung Cancer Genomics Data. *Joint Statistical Meetings held in Seattle, Washington.*
16. **Guha, S.** and Baladandayuthapani, V. (2015). Nonparametric Variable Selection, Clustering and Prediction for High-Dimensional Regression. *10th Conference on Bayesian Nonparametrics held in Raleigh, North Carolina.*
17. Gu, C., **Guha, S.** and Baladandayuthapani, V. (2015). Scalable Bayesian Nonparametric Learning for High-Dimensional Lung Cancer Genomics Data. *ENAR Spring Meeting held in Miami, Florida.*
18. **Guha, S.** and Baladandayuthapani, V. (2014). Variable selection in Bayesian high-dimensional regression of survival outcomes. *Joint Statistical Meetings held in Boston, Massachusetts.*
19. **Guha, S.** and Baladandayuthapani, V. (2014). Variable selection in Bayesian high-dimensional regression of survival outcomes. *International Conference in Honor of H.N. Nagaraja for His 60th Birthday.*
20. **Guha, S.** and Baladandayuthapani, V. (2013). Survival Predictor Detection by Dirichlet Processes. *Tenth Annual Conference on Frontiers in Applied and Computational Mathematics held at the New Jersey Institute of Technology in Newark, New Jersey.*
21. **Guha, S.** (2011). Posterior Simulation in Countable Mixture Models for Large Datasets. *Eighth Workshop on Bayesian Nonparametrics held at Veracruz, Mexico.*
22. **Guha, S.** (2010). Posterior Simulation in Countable Mixture Models for Large Datasets. *SAMSI Program on Semiparametric Bayesian Inference: Applications in Pharmacokinetics and Pharmacodynamics held at Research Triangle Park, North Carolina.*

23. **Guha, S.** (2010). Markov Chain Monte Carlo Methods for Mixture Models. *Ninth Valencia International Meeting on Bayesian Statistics / ISBA 2010 World Meeting held in Alicante, Spain.*
24. **Guha, S.,** Li Y. and Neuberger D. (2009). Bayesian Hidden Markov Modeling of Array CGH Data. *Seminar on Bayesian Inference in Econometrics and Statistics, held in St. Louis, Missouri.*
25. **Guha, S.** (2009). Markov Chain Monte Carlo Methods for Mixture Models. *Joint Statistical Meetings held in Washington D.C.*
26. **Guha, S.,** Li Y. and Neuberger D. (2008). Bayesian Hidden Markov Modeling of Array CGH Data. *Southern Regional Council on Statistics held in Charleston, South Carolina.*
27. **Guha, S.** (2008). Posterior Simulation in Countable Mixture Models for Large Datasets. *Joint Statistical Meetings held in Denver, Colorado.*
28. **Guha, S.,** Li Y. and Neuberger D. (2008). Bayesian Hidden Markov Modeling of Array CGH Data. *Conference on Statistical Paradigms: Recent Advances and Reconciliations, held in Kolkata, India.*
29. **Guha, S.,** Li Y. and Neuberger D. (2008). Bayesian Hidden Markov Modeling of Array CGH Data. *Session on Bayesian Computational Methods for Biomedical Applications, Joint Statistical Meetings, held in Seattle, Washington.*
30. **Guha, S.** and MacEachern, S. N. (2005). Generalized Post-stratification and Importance Sampling for Subsampled Markov Chain Monte Carlo Estimation. *Seminar on Bayesian Inference in Econometrics and Statistics, held in St. Louis, Missouri.*
31. **Guha, S.,** MacEachern, S. N. and Peruggia, M. (2002). Benchmark Estimation for Markov Chain Monte Carlo Samples. *Joint Statistical Meetings held in New York City.*

INVITED
ACADEMIC
PRESENTATIONS

Barretos Cancer Hospital, Brazil; Department of Environmental Health, Harvard TH Chan School of Public Health; Department of Mathematical Sciences, University of Nevada, Las Vegas; ENAR 2020 Spring Meeting; Bayes Comp 2020 held in Gainesville, Florida; Conference on Innovations in Data and Statistical Sciences held in Mumbai, India; Conference on Current Trends in Survey Statistics held in Singapore; 4th Eastern Asia Meeting On Bayesian Statistics (EAC-ISBA) held in Kobe, Japan; ENAR Spring Meeting held in Philadelphia, Pennsylvania; Tenth International Triennial Calcutta Symposium on Probability and Statistics held in Kolkata, India; 12th International Interdisciplinary Conference MSAST 2018 held in Kolkata, India; 11th International Conference of the ERCIM WG on Computational and Methodological Statistics held in Pisa, Italy; IISA International Conference in Statistics held in Gainesville, Florida, from May 2018 - present.

Seminar on Bayesian Inference in Econometrics and Statistics (SBIES) conference held in St. Louis, Missouri; Seminar talks at the Epidemiology and Biostatistics Di-

vision, University of Illinois at Chicago; Department of Statistics, Temple University; Department of Statistics, North Carolina State University; Department of Statistics and Applied Mathematics, University of California at Santa Cruz; Statistical and Applied Mathematical Sciences Institute (SAMSI) in RTP, North Carolina, from October 2015 - April 2018.

Joint Statistical Meetings held in Seattle, Washington; 10th Conference on Bayesian Nonparametrics, held in Raleigh, North Carolina; ENAR Spring Meeting held in Miami, Florida; Joint Statistical Meetings held in Boston, Massachusetts; Tenth Annual Conference on Frontiers in Applied and Computational Mathematics held at the New Jersey Institute of Technology in Newark, New Jersey; Eighth Workshop on Bayesian Nonparametrics held at Veracruz, Mexico; SAMSI Program on Semiparametric Bayesian Inference: Applications in Pharmacokinetics and Pharmacodynamics held at Research Triangle Park, North Carolina, from July 2010 - September 2015.

Environmental Statistics Retreat 2010 organized by the Department of Biostatistics, Harvard School of Public Health in Rockport, Massachusetts. Seminar talks at the Department of Mathematics and Statistics, University of Missouri - Kansas City; Division of Biostatistics, Washington University in St. Louis; Department of Biostatistics, Harvard School of Public Health; Department of Statistics, Western Michigan University; Department of Biostatistics, MD Anderson Cancer Center, from October 2009 - June 2010.

Seminar talks at the Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health; Department of Statistics, University of Virginia; Department of Statistics, Iowa State University; Department of Epidemiology & Biostatistics, University of Florida; Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison; Department of Statistics, University of California, Davis; Department of Applied Mathematics and Statistics, University of California, Santa Cruz, from January - March 2007.

Seminar talks at the Department of Ambulatory Care and Prevention, Harvard Medical School; Department of Statistics, Temple University; Department of Statistics, University of California, Irvine; Department of Biostatistics, University of California, Los Angeles; Department of Epidemiology and Biostatistics, University of South Carolina; Department of Health Care Policy, Harvard Medical School; Statistics Department, University of Pennsylvania, from January - March 2007.

Racebrook Environmental Statistics Retreat, held in Sheffield, Massachusetts. Seminar talks at the Department of Statistics, Texas A&M University; Statistics Department, Harvard University; Department of Statistics, Texas A&M University; Department of Mathematical Sciences, University of Arkansas; Department of Statistics, Oklahoma State University, from January 2004 - November 2006.

CURRENT AND
PAST PH.D.
STUDENTS

Shiqi Cui. Dissertation topic: *Bayesian Mixture Models for High-Throughput Bioinformatics Applications*. Co-directed with Marco Ferreira. Graduated in December 2014. Working for Google LLC.

Chiyu Gu. Dissertation topic: *Scalable Bayesian Nonparametric Learning for Biomedical Big Data*. Graduated in May 2018. Currently working for Bayer Crop Science.

Chetkar Jha. Dissertation topic: *Bayesian Nonparametric Analysis of Multivariate Unordered Categorical data*. Graduated in May 2019. Postdoctoral researcher at Pennsylvania State University.

Dongyan Yan. Dissertation topic: *Scalable Computational Algorithms and Massively Parallel Computing for Bayesian Mixture Models*. Graduated in June 2019. Working for Elli Lilly and Company.

Archie Sachdeva. Dissertation topic: *Bayesian Analytical Methods for Microbiome Data*.

SOFTWARE

2020 R code for generating simulated data and fitting the Bayesian Connectomics (BaCon) model class of Guha, Jung and Dunson (2020) is publicly available at <https://github.com/sguha-lab/BaCon/>

2020 Code for fitting microbiome datasets using the Bayesian SVD-type decomposition technique of Guha and Datta (2021) is available at <https://github.com/sguha-lab/Microbiome-SVD>

2020 Code for implementing the Bayesian technique of Guha and Ghosh (2020) for inferring unknown conic sections on the basis of noisy data is available at <https://github.com/sguha-lab/BayesConics>

2017 R package *NPCluster* implementing VariScan, a nonparametric Bayesian technique for clustering, variable selection, and prediction in high-throughput regression settings. The methodology is developed in Guha and Baladandayuthapani (EJS, 2016). Available at <https://github.com/suchard-group/NPCluster>.

We are currently working on implementing the package's MCMC procedure in a parallel computing framework using graphical processing units. Co-developed with Veera Baladandayuthapani, Chiyu Gu, and Marc Suchard.

2015 R package *hmmSeq* implementing the Bayesian technique for analyzing RNA-Seq data proposed in Cui, Guha, Ferreira and Tegge (AOAS, 2015). Joint work with the paper's co-authors.

- 2011 R package *glmmGS* for fitting generalized linear mixed models to massive datasets. Publicly available from CRAN at <http://r-forge.r-project.org/projects/glmmgs>. Co-developed with Michele Morara, Louise Ryan, and Christopher Paciorek.
- 2009 The Bayesian hidden Markov model strategy proposed by Guha, Li and Neuberg (JASA, 2008) for array CGH genomics data is implemented in Bioinformatics Toolbox 3.2. It is available at <http://www.mathworks.com>
- 2007 Created software for generating the sample paths of a number of common stochastic processes, verifying their theoretical properties by simulation and visualizing abstract results like the martingale central limit theorem. Used in the Department of Biostatistics, Harvard School of Public Health to teach the graduate level courses, *Analysis of Failure Time Data* and *Probability Theory and Applications I*.

TEACHING
EXPERIENCE

PHC 6937 – Frontiers in Biostatistics. *Spring 2018–2020.*

Level: Graduate. Introduces biostatistics Masters and Ph.D. students to current issues and methods in modern biostatistics research. Current faculty present selected topics from their current research.

PHC 6937 – Bayesian Biostatistical Methods. *Fall 2019.*

Level: Graduate. Equips students with an understanding of the basics of Bayesian statistics, with special emphasis on practical implementation.

PHC 6937 – Analysis of Multivariate Data. *Fall 2018, 2020.*

Level: Graduate. Covers linear models methodology including simple and multiple regression and analysis of variance including factorial and block designs. The course covers regression for categorical data, random effects models for correlated data, and nonparametric and semiparametric regression.

PHC 7090 – Advanced Biostatistical Methods I. *Fall 2020.*

Level: Graduate. This is the first course of a two-course sequence providing students with advanced knowledge and mathematical details on statistical inference and linear models. Covers Frequentist and Bayesian estimation and inference. It also encompasses estimation and inference for linear models and fitting of regression models.

TEACHING EXPERIENCE AT UNIVERSITY OF MISSOURI **STAT 9710 – Mathematical Statistics I.** *Fall 2011–2014.*
Level: Graduate. Theory of estimation and tests of hypotheses including sufficiency, completeness and exponential families. Neyman-Pearson lemma, most powerful tests, similarity and invariance. Bayes and minimum variance unbiased estimates. Confidence intervals and ellipsoids.

STAT 9720 – Mathematical Statistics II. *Spring 2012–2017.*
Level: Graduate. Asymptotic distributions of maximum likelihood estimators, chi-square and likelihood ratio test statistics. EM algorithm, bootstrap, and introduction to generalized linear models.

STAT 4710/7710 – Introduction to Mathematical Statistics. *Spring 2008–2010, Spring 2012, Spring 2014, Spring 2017, Fall 2010, Fall 2013, Fall 2016, Fall 2017.*
Level: Graduate/Undergraduate. Introduction to theory of probability and statistics using concepts and methods of calculus.

STAT 4510/7510 – Applied Statistical Models I. *Spring 2011, Spring 2013, Spring 2015, Fall 2011–2012, Fall 2014, Fall 2016.*
Level: Graduate/Advanced Undergraduate. Introduction to applied linear models including regression (simple and multiple, subset selection, estimation and testing) and analysis of variance (fixed and random effects, multifactor models, contrasts, multiple testing).

STAT 4750/7750 – Introduction to Probability Theory. *Spring 2009–2011, Fall 2009–2010, Fall 2017.*
Level: Graduate/Advanced Undergraduate. Probability spaces; random variables and their distributions; repeated trials; probability limit theorems.

STAT 4410/7410 – Biostatistics. *Fall 2009.*
Level: Graduate/Advanced Undergraduate. Study of statistical techniques for the design and analysis of clinical trials, laboratory studies and epidemiology. Topics include randomization, power and sample size calculation, sequential monitoring, Carcinogenicity bioassay and case-cohort designs.

STAT 4830/7830 – Categorical Data Analysis. *Fall 2008.*
Level: Graduate/Advanced Undergraduate. Discrete distributions, frequency data, multinomial data, chi-square and likelihood ratio tests, logistic regression, log linear models, rates, relative risks, random effects, case studies.

OTHER TEACHING EXPERIENCE **Harvard School of Public Health**
BIO503 – Programming and Statistical Modeling in R (co-taught with Christopher Paciorek). *Spring 2007.*

BIO283 – Spatial Statistics for Health Research (co-taught with Louise Ryan, Yi Li, Christopher Paciorek). *Fall 2004*.

Ohio State University

STAT 428 – Introduction to Probability and Statistics for Engineering and the Sciences II. *Fall 2002, Winter 2003, Spring 2003*.