

Jyotishka Datta

CONTACT INFORMATION

Department of Mathematical Sciences
University of Arkansas at Fayetteville
SCEN 219
Fayetteville, AR 72701, USA

Mobile: +1-765-398-2914
E-mail: jd033@uark.edu
Webpage: <https://sites.uark.edu/jd033/>
Git: <https://github.com/DattaHub>

RESEARCH INTERESTS

Bayesian Methodology and Theory, Bioinformatics, Sparse Signal Recovery, Global-local Shrinkage Priors, Default Bayes, Discrete Data, High-dimensional data, Compositional Data, Applied probability, and Bayesian nonparametrics.

PAST ACADEMIC APPOINTMENTS

2014- 2016: Postdoctoral Associate.

[Department of Statistical Science, Duke University](#), Durham, NC.

[Statistical and Applied Mathematical Sciences Institute](#), Durham, NC.

- **Mentors:** Prof. David B. Dunson (Department of Statistical Science), and Prof. Sandeep S. Dave (Department of Medicine), Duke University.
- **SAMSI Program:** Beyond Bioinformatics.

EDUCATION

2009 - 2014: Ph.D. in Statistics, [Purdue University](#), West Lafayette, IN.

- **Dissertation Topic:** Some Theoretical and Methodological Aspects of Multiple Testing, Model Selection and Related Areas, **Advisor:** Prof. Jayanta K. Ghosh.

2003 - 2008: B.Stat and M. Stat, [Indian Statistical Institute](#), Kolkata, India.

PUBLICATIONS (STATISTICS)

- [1] Bhadra, **Datta**, Polson, and Willard (2017), (*alphabetical) "The Horseshoe+ Estimator of Ultra-Sparse Signals", *Bayesian Analysis online*. **12 (4)**, 1105-1131.
- [2] **Datta** and Dunson (2016), "Bayesian inference on quasi-sparse count data", *Biometrika*, **103 (4)**: 971-983.
- [3] Bhadra, **Datta**, Polson, and Willard (2016), (*alphabetical) "Default Bayesian analysis with global-local shrinkage priors", *Biometrika*, **103 (4)**: 955-969.
- [4] **Datta**, and Ghosh (2014), "Bootstrap – An Exploration." *Statistical Methodology*: **20**, 63-72.
- [5] **Datta**, and Ghosh (2013), "Asymptotic Properties of Bayes Risk for the Horseshoe Prior". *Bayesian Analysis 8.1*: 111-132.

ARTICLES UNDER REVIEW

- [6] Bhadra, **Datta**, Polson, and Willard (2018+), (*alphabetical), "Lasso Meets Horseshoe - A Survey". [arXiv](#), Minor Revision, Statistical Science.
- [7] Bhadra, **Datta**, Polson, and Willard (2018+), (*alphabetical), "Horseshoe Regularization for Feature Subset Selection". [ArXiv](#).
- [8] **Datta** and Dunson (2018+), "Nonparametric Bayes multiresolution testing for massive-dimensional rare events".

- [9] Bhadra, **Datta**, Polson, and Willard (2018+), (*alphabetical), "Prediction Risk for Global-Local Shrinkage Regression". *submitted to JMLR*.
- [10] Bhadra, **Datta**, Polson, and Willard (2018+), (*alphabetical), "Global-local mixtures". *submitted*. [Featured on Prof. Christian Robert's blog: [Global-Local Mixtures](#)]
- PUBLICATIONS (CANCER GENOMICS)
- [11] Zhang, Jenny et al. (2017) "Integrative Analysis of 1001 Diffuse Large B Cell Lymphoma Identifies Novel Oncogenic Roles for Rho-a". *Blood*, 130 (Suppl 1), 37-37.
- [12] Reddy, Anupama, et al. (2017) "Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma". *Cell*, **171.2**: 481-494. Featured on EurekAlert!, the newsletter from AAAS, [link](#).
- [13] Moffitt et al. (2017). "Enteropathy-associated T cell lymphoma subtypes are characterized by loss of function of SETD2", *Journal of Experimental Medicine*, **214(5)**, 1371-86.
- [14] McKinney, Moffitt, et al. (2017) "The Genetic Basis of Hepatosplenic T Cell Lymphoma". *Cancer Discovery*, CD-16-0330.
- [15] Healy, et al. (2016). "GNA13 loss in germinal center B cells leads to impaired apoptosis and GCB cell persistence and promotes lymphoma in vivo". *Blood*, **127 (22)**, 2723-2731.
- [16] Zhang et al. (2016). Integrative Genetic and Clinical Analysis through Whole Exome Sequencing in 1001 Diffuse Large B Cell Lymphoma (DLBCL) Patients Reveals Novel Disease Drivers and Risk Groups. *Blood*, **128 (22)**, 1087.
- [17] McKinney et al. (2016). "SETD2 Functional Loss through Mutation or Genetic Deletion Promotes Expansion of Normal and Malignant $\gamma\delta$ T Cells through Loss of Tumor Suppressor Function and Upregulation of Oncogenic Pathways. *Blood*. **128 (22)**, 1052.
- PUBLICATIONS (INTERDISCIPLINARY)
- [18] **Pediatrics:** Chaudhuri, Biswas, Datta, : : , Chakarabrty. (2016). "Evaluation of malnutrition as a predictor of adverse outcomes in febrile neutropenia associated with pediatric hematological malignancies." *Journal of Paediatrics and Child Health*, 52 (7), 704-709.
- [19] **Neuroscience:** Parthasarathy, Datta, Torres, Hopkins, and Bartlett. (2014) "Age-Related Changes in the Relationship Between Auditory Brainstem Responses and Envelope-Following Responses." *Journal of the Association for Research in Otolaryngology*. 15 (4), 649-661.
- [20] **Geology:** Libohova, , Winzeler, Lee, Schoeneberger, Datta, and Owens. (2016). "Geomorphons: Landform and property predictions in a glacial moraine in Indiana landscapes". *Catena* 2016 v.142.

- REFEREED BOOK CHAPTERS [21] **Datta** and Ghosh. "In Search of Optimal Objective Priors for Model Selection and Estimation". In S. Upadhyay, U. Singh, D. Dey, & A. Loganathan (Eds.), *Current Trends in Bayesian Methodology with Applications (2015)*, 225-239. Chapman & Hall/CRC Press.
- [22] Dasgupta, Ghosh, Chakravarty, and **Datta**. "Some Remarks on Pseudo Panel Data". In R. Dasgupta (Ed.) *Growth Curve and Structural Equation Modeling (2015)*, 25-34. Springer International.
- MANUSCRIPTS IN PREPARATION (* → Graduate student)
- [23] Abba*, Bhadra, **Datta**, Polson, and Willard (2018+), (*alphabetical), "Bayesian Square-root Lasso".
- [24] Price*, **Datta**, "An Empirical Bayes Approach to Power Calculation and Cross-Validation in Large Scale Hypotheses Testing".
- [25] Sengupta, **Datta**, Chen (2018+), "Proximity Block-models for Network Data".
- [26] Datta and Dunson (2018+), "Sparse generalized Dirichlet distributions for high-dimensional probabilities".
- [27] **Datta**, Ghosh, and Majumder. "A Statistical Method for Drawing Robust Inferences in the Presence of Local Dependence in Genome-scale Data".
- TECHNICAL REPORTS [28] Bernhardt-Barry, M.L., Datta, J., and Wood, C.M., "Predicting Soil Type from Non-destructive Geophysical Data using Bayesian Statistical Methods", Final Research Report to the Maritime Transportation Research and Education Center.
- CONFERENCE PUBLICATIONS [29] LeBow V., Bernhardt-Barry, M. L., and **Datta**, J. (2018, June), Improving Spatial Visualization Abilities Using 3D Printed Blocks Paper presented at 2018 ASEE Annual Conference & Exposition , Salt Lake City, Utah. [Permanent URL](#).
- OTHER PUBLICATIONS [30] Datta and Drawve, "Does Machine Learning Reduce Racial Disparities in Policing?", IISA Newsletter, December, 2016.
- [31] Datta and Ghosh, "Optimal Objective Priors for Linear Models", Indian Bayesian Society Newsletter, Vol XI, No. 1, May, 2014.
- AWARDS AND HONORS
- Travel Grant for attending 19th IMS Meeting of New Researchers in Statistics and Probability, Johns Hopkins University, Baltimore, Maryland.
 - **Travel Grant** for presenting at the IISA Conference in Corvallis, OR.
 - **Travel Grant** for presenting at the ASA/Kutner faculty poster session at the SRCOS Summer Research Conference in Bentonville, AR.
 - **Honorable Mention Award for Best Theoretical Poster** at the O-Bayes 2013 Meeting: The Tenth International Workshop on Objective Bayesian Statistics, December 15-19, Durham, Raleigh, USA.
 - **Travel Support Award** for attending the O-Bayes 2013 Meeting: The Tenth International Workshop on Objective Bayesian Statistics, December 15-19, Durham, Raleigh, USA.

- **William J. Studden Publication Award** for an outstanding publication in a mathematical statistics journal, 2013, Department of Statistics, Purdue University.
- Award for Academic Excellence, Indian Statistical Institute, Kolkata, 2008.
- Ranked **8th** and **10th** in State Level Joint Entrance Examination in **Engineering** and **Medicine** (out of two hundred thousand students), 2003.

FUNDING

- Robert and Sandra Connor **Endowed Faculty Fellowship** from the University of Arkansas, 2018-19. (\$ 5,000)
- Datta, Jyotishka, Grant Drawve, Casey Harris, and Shaun Thomas (*alphabetical). November 2017. "Participant Field Training with Little Rock Police Department." Provost's Collaborative Research Grant (\$ 2,000).
- Datta, Jyotishka, Mohamed Abdelkader Abba* (*graduate student). November 2016. "Multiresolution Nonparametric Bayesian Hotspot Detection." Provost's Collaborative Research Grant (\$2,000).
- **NSF Postdoctoral Fellowship**, Statistical and Applied Mathematical Sciences Institute, **2015-2016**.
- **Summer Research Grant**, Department of Statistics, Purdue University, **2011-2013**.
- **Summer Research Grant**, Department of Statistics, Purdue University, 2011.

INVITED TALKS

- August, 2019: Special Invited Session in Joint Statistical Meeting, Denver, Colorado.
- January 2019: Plenary Session in 10th International Calcutta Triennial Symposium, December 27-30, 2018, Kolkata, India.
- December 2018: Invited Session in Young Statisticians' Meet: Data Science in Action: January 4-5, 2019, Indian Statistical Institute, Kolkata, India.
- December 2017: "Horseshoe Regularization for Feature Subset Selection", 2017, IISA International Conference on Statistics at Hyderabad, India.
- December 2017: "Horseshoe Regularization for Feature Subset Selection", ERCIM WG Meeting, CMStatistics 2017 Conference at London, UK.
- August 2017: "Detecting rare mutational hotspots by multiscale BNP method", Joint Statistical Meeting, Baltimore, Maryland.
- January 2017: "Sparse signal recovery and default Bayesian analysis using global-local shrinkage priors", Applied Statistics Unit, Indian Statistical Institute, Kolkata.
- August, 2016: "Default Bayesian analysis for global-local shrinkage priors", IISA Conference, Corvallis, Oregon.
- August 2016: "Shrinkage Priors for High-Dimensional Sparse Poisson Means", Joint Statistical Meeting, Chicago, Illinois.
- February, 2016: "Shrinkage Priors for High-Dimensional Sparse Poisson Means" (STAT 701 Talk): Duke University.
- December, 2015 -January, 2016: "Sparse Signal Recovery for Discrete & Continuous Data" (Job Talk): Binghamton University, University of Arkansas at Fayetteville, and Clemson University.
- May, 2015: "Multiscale Bayesian cluster detection and testing for whole genome sequencing studies", Transition workshop for "Beyond Bioinformatics", SAMSI, North Carolina.
- August, 2014: "Sparse and Ultra-Sparse Signal Recovery: The Horseshoe and The Horseshoe+Prior", Department of Statistical Science, Duke University.

- January, 2014: “Shrinkage priors for multiple testing and model selection”, University of Texas - M. D. Anderson Cancer Center, Houston, TX.
- November, 2013: “In Search of Optimal Objective Priors for Model Selection and Estimation”, Mathematical Statistics Seminar, Purdue University.
- May, 2013: “Two-groups and One-Group Models for Multiple Testing”, National Institute of Biomedical Genomics, Kalyani, India.

CONTRIBUTED TALKS AND POSTERS

- September, 2016: “Sparse Signal Recovery for Discrete & Continuous Data” and “Detecting rare mutational hotspots by multiscale BNP method”, Departmental seminar, University of Arkansas, Fayetteville.
- September, 2015: “Shrinkage Priors for High-Dimensional Sparse Poisson Means”, *Poster presentation*, John W. Tukey 100th Birthday Celebration at Princeton University.
- September, 2015: “Shrinkage Priors for Sparse High-Dimensional Discrete or Continuous Data”. *Talk*, SAMSI postdoc seminar.
- July, 2015: “Bayesian Cluster Detection for Rare Variants”, *Poster Presentation*, SAHD (Sensing and Analysis of High-Dimensional Data Workshop), Duke University, Durham, NC.
- June, 2015: “Multiscale Bayesian cluster detection and testing for whole genome sequencing studies”, *Poster presentation*, SRCOS (Southern Research Conference), Carolina Beach, NC.
- May, 2015: “Multiresolution nonparametric Bayesian cluster detection and association testing for whole genome sequencing studies”, *Poster presentation*, CCPS (Cancer Control and Population Sciences Fair), Duke University, NC.
- May, 2015: “Multiresolution nonparametric Bayesian cluster detection and association testing for whole genome sequencing studies with applications in CVID”, *Poster presentation*, The Biology of Genomes Meeting, Cold Spring Harbor Lab, NY.
- September, 2014: “Ultra-Sparse Signal Recovery through the Horseshoe+ Prior”, *Talk*, SAMSI.
- December, 2013: “In Search of Optimal Objective Priors for Model Selection and Estimation”, *Poster presentation*, O-Bayes 2013, Duke University.
- March, 2013: “Two-groups and One-Group Models for Multiple Testing”, *Talk*, Machine Learning Seminar, Department of Computer Science, Purdue University.
- October, 2012: “Asymptotic properties of Bayes risk for the Horseshoe prior”, *Talk*, Graduate Student Organization Seminar, Department of Statistics, Purdue University.

MENTORING

- Undergraduate Students
 - Honors Thesis Committee: Vanessa Lebow, Winson Chee, Dhruva Dasgupta, Christopher Peterson.
- Graduate Students
 - Primary Advisor (MS): Mohamed Abba, Josh Price, Kai Cui.
 - Committee Member (MS): Ji Li, Michael Ellis, James Willbanks.
 - Committee Member: Gina Riggio, Md Jubaer Hossain Pantho.

TEACHING
EXPERIENCE

- Spring 2018, Instructor, Department of Mathematical Sciences, University of Arkansas.
 - STAT 5443 (Computational Statistics): Graduate Course. Audience: Advanced graduate students from quantitative disciplines. Open-source materials.
- Fall 2017, Instructor, Department of Mathematical Sciences, University of Arkansas.
 - STAT 4033, Nonparametric Statistics. Audience: Undergraduate and Graduate students from quantitative disciplines. <http://dattahub.github.io/stat4033/list.html>.
 - STAT 3013 (Introduction to Probability): Undergraduate Course.
- Spring 2017, Instructor, Department of Mathematical Sciences, University of Arkansas.
 - STAT 5443 (Computational Statistics): Graduate Course. Audience: Advanced graduate students from quantitative disciplines. Open-source materials.
 - STAT 3013 (Introduction to Probability): Undergraduate Course. Apps: Central Limit Theorem: <https://jdatta.shinyapps.io/demoivre/> and Glivenko-Cantelli lemma <https://jdatta.shinyapps.io/eCDFdemo/>
- Fall 2016, **Instructor**, Department of Mathematical Sciences, University of Arkansas.
 - Stat 4033, Nonparametric Statistics. Audience: Undergraduate and Graduate students from Quantitative disciplines.
- Summer 2014, **Instructor**, Department of Statistics, Purdue University.
 - Stat 301, Introduction to Statistics, Course Coordinator: Meghan Tooman.
 - Responsibilities: Designing and holding recitations and lab sessions for using SPSS for undergraduate students, grading homework, lab exercises, and midterm and final exams.
- Spring 2012-Spring 2013, **Teaching Assistant (Lab Instructor)**, Department of Statistics, Purdue University.
 - Stat 598Z, Introduction to Computing for Statisticians, Instructor: Prof. S. V. N. Vishwanathan.
 - Stat 598G, Introduction to Computational Statistics, Instructor: Prof. Sergey Kirshner.
 - Responsibilities: Holding lab sessions for teaching Statistics using SPSS to small groups of undergraduate students, grading homework, lab exercises, and midterm tests.
 - Lab Website: <https://learning.cs.purdue.edu/courses/sp2013/598z/lab>.
- Spring 2011-Fall 2011, **Teaching Assistant (Lab Instructor)**, Department of Statistics, Purdue University.
 - Stat 301, Introduction to Statistics, Course Coordinator: Ellen Gundlach.
 - Stat 113, Statistics for Society, Course Coordinator: Prof. John Deely.
 - Responsibilities: Teaching recitation sessions for undergraduate students, holding office hours, grading homework, lab exercises, and the midterm.

- Fall 2010, **Teaching Assistant (Grader)**, Department of Computer Science, Purdue University.
 - CS 471, Artificial Intelligence, Instructor: Prof. Alan Qi.
 - Responsibilities: Teaching recitation sessions for undergraduate students (groups of 20), holding office hours, grading homeworks, lab exercises, and the midterm.

PAST ACADEMIC EXPERIENCE [Department of Statistical Science, Duke University](#), and [Genomic and Computational Biology Center, Duke University](#), Durham, NC.

Postdoctoral Associate **August, 2014 - July, 2016.**

- Supervisors: Prof. David B. Dunson (Department of Statistical Science), Prof. Sandeep S. Dave (Department of Medicine).
- Developing scalable Bayesian methodologies for whole-genome sequencing studies that incorporates additional spatial information and adjusts for clinical and demographic covariates.
- Providing statistical consulting for genomic data analysis.

[Statistical and Applied Mathematical Sciences Institute](#), Durham, NC.

Postdoctoral Fellow **August, 2014 to present**

- Supervisors: Prof. David B. Dunson (Department of Statistical Science).
- SAMSI Program: Beyond Bioinformatics (Working Groups: High-dimensional Discrete Data, Sparse Multiway Data Integration).
- Responsibilities: Managing and leading working groups, maintaining group website, presenting new ideas and forging interdisciplinary collaborations to solve important new problems in Bioinformatics.

[Department of Statistics, Purdue University](#), West Lafayette, IN.

Statistical Consultant **August, 2011 to present**

- Experimental Design and Data Analysis Consulting, providing assistance with all phases of research projects including but not limited to design of studies, survey design, analysis of data, interpretation of results, presentation of results and other statistics or probability problems.
- Supervisors: Prof. Bruce Craig, Prof. Thomas Kuczek, Prof. Michael Zhu.
- Statistical Software Consulting, Providing assistance with the set up and running of a wide variety of statistical computing programs, including JMP, Matlab, Minitab, R, SAS, SPSS, S-Plus, Stata.

[Department of Statistics, Purdue University](#), West Lafayette, IN.

Graduate Research Assistant **August, 2009 to December, 2010**

- Supervisors: Prof. Alan Qi.
- Implementing a new Bayesian regularization method, EigenNet, which selects groups of correlated variables by utilizing eigenstructure of data and an adaptive Metropolis-within-Gibbs sampler to ensure faster convergence than regular Metropolis Hastings algorithm.

INDUSTRIAL
PROFESSIONAL
EXPERIENCE

Barclays Bank, PLC, Mumbai, India.

Graduate Emerging Manager

June 2008 to September 2009

- Developing and Testing Scorecard for evaluating potential customers, Developing performance based strategies for Approval of Credit Cards and various Segmentation Analysis for the existing population to identify Non-delinquent and delinquent sectors of customers.

Systat Softwares Asia Pacific Ltd., Bangalore, India.

Summer Intern

May 2005 to July 2005

- Supervisor: Dr. T. Krishnan.
- Worked on Markov Chain Monte Carlo Methods Using SYSTAT 11 and implementation of Transformed Density Rejection Algorithm.

SOFTWARE
SKILLS

- Languages: C, MATLAB, PYTHON, R, STAN.
- Systems: UNIX, Windows.
- Packages: SPSS, SAS, JMP, STATA, MINITAB.
- Other Softwares: L^AT_EX and other common productivity tools.

PROFESSIONAL
ACTIVITY

- Served as a reviewer for *Biometrika*, *Journal of American Statistical Association (Theory and Methods)*, *Statistica Sinica*, *Biometrika*, *Electronic Journal of Statistics*, *Sankhya Series A*, *Entropy*, *Machine Learning Conferences: NIPS, ICML, AISTats*.
- Served as a proposal reviewer for National Science Foundation (2017).
- Served as a reviewer for *Biometrika*, *Journal of American Statistical Association (Theory and Methods)*, *Statistica Sinica*.
- Organized a topic-contributed paper titled "Recent Advances in Bayesian Methodology and Computation for Ultra-High Dimensional Data" sponsored by the Section on Bayesian Statistical Science (SBSS) at Joint Statistical Meeting, 2016.
- Served as a chair for the invited paper session on "High-dimensional Bayesian statistics: spike-and-slab and global-local shrinkage" at Joint Statistical Meeting, 2016.
- Served as a chair for the invited paper session on "Bayesian Model Selection" at Joint Statistical Meeting 2017. Baltimore, MD.
- Served as a chair for the invited paper session on "Modeling Dependence in Large Systems" at IISA 2017 Annual Conference. Hyderabad, India.

MEMBERSHIPS

- International Society for Bayesian Analysis (ISBA), since January, 2012.
- American Statistical Association (ASA), since December, 2011.
- Institute of Mathematical Statistics (IMSTAT), Since April, 2011.
- Mu-Sigma-Rho, the National Statistics Honorary, 2012.

REFERENCES
AVAILABLE TO
CONTACT

Prof. David B. Dunson (e/mail: dunson@stat.duke.edu; phone: +1-919-260-6615)

- Arts and Sciences Distinguished Professor, [Department of Statistical Science, Duke University](#)

★ *Prof. Dunson is my postdoctoral advisor.*

Prof. Nicholas G. Polson (e/mail: ngp@chicagobooth.edu ; phone: +1-773-702-9298)

- Professor of Econometrics and Statistics, [Booth School of Business, University of Chicago](#)

★ *Prof. Polson is my collaborator on a few ongoing projects.*

Prof. Sandeep S. Dave (e/mail: sandeep.dave@duke.edu; phone: +1-919-681-1922)

- Associate Professor, [Department of Medicine, Duke University School of Medicine.](#)

★ *Prof. Dave is my mentor and collaborator for a few ongoing projects.*

Prof. Partha P. Majumder (e/mail: ppm@nibmg.ac.in; phone: +91-33-25892151)

- Director, [National Institute of Biomedical Genomics](#)

★ *Prof. Majumder is my collaborator on a current project and has taught me at ISI, Kolkata.*

Prof. S. V. N. Vishwanathan (e/mail: vishy@ucsc.edu; phone: +1-831-459-1721)

- Professor, [Jack Baskin School Of Engineering, UC Santa Cruz.](#)

★ *I was the lab instructor for two statistical computing courses taught by Prof. Vishwanathan.*

MORE
INFORMATION

More information can be found at <https://you.uark.edu/jd033>.

My Google Scholar Profile: <http://bit.ly/1OTdd9h>