Subhabrata Majumdar

Address:
716 SW 16th Ave Apt 313
Gainesville, FL 32601

e-mail: smajumdar@ufl.edu
Web: https://shubhobm.github.io

Research interests

Methodology: Statistical machine learning- complex graphical models, variational inference; High dimensional multivariate models- sparse regression, robust statistics; Data depth and its inferential applications- variable selection, dimension reduction.

Application: Statistical models for drug discovery, Data Science for social good, genetic mapping of complex behavioral traits.

Education

PhD Statistics, University of Minnesota - Twin Cities, 2017. Advisor: Snigdhansu Chatterjee;

M.S. Statistics, Indian Statistical Institute, 2012. Specialization: Actuarial science and Genetics;

B.S. Statistics, Indian Statistical Institute, 2010.

Publications

Book chapters:

- Basak, S.C., **Majumdar, S.** Current Landscape of Hierarchical QSAR Modeling and its Applications: Some Comments on the Importance of Mathematical Descriptors as well as Rigorous Statistical Methods of Model Building and Validation. In *Advances in Mathematical Chemistry and Applications: Vol. 1*, **2015**, 251-281, published by Elsevier and Bentham e-Books.

Journal articles:

- **Majumdar, S.** and Basak, S. C. Beware of external validation! – A Comparative Study of Several Validation Techniques used in QSAR Modelling. *Curr. Comput. Aided Drug Des.*, **2018**, *in press*;
- **Majumdar, S.** and Chatterjee, S. Nonconvex penalized multitask regression using data depth-based penalties, *Stat*, **2018**, *7*, e174, http://arxiv.org/abs/1610.07540;
- **Majumdar, S.** and Basak, S. C. Exploring intrinsic dimensionality of chemical spaces for robust QSAR model development: A comparison of several statistical approaches. *Curr. Comput. Aided Drug Des.*, **2016**, *12*, 294–301;
- Basak, S. C. and **Majumdar, S.** Prediction of Mutagenicity of Chemicals from Their Calculated Molecular Descriptors: A Case Study with Structurally Homogeneous versus Diverse Datasets. *Curr. Comput. Aided Drug Des.*, **2015**, *11*, 117–123;
- **Majumdar, S.**, Basak, S. C. and Grunwald, G. D. Adapting Interrelated Two-Way Clustering Method for Quantitative Structure-Activity Relationship (QSAR) Modeling of Mutagenicity/ Non-Mutagenicity of a Diverse Set of Chemicals. *Curr. Comput. Aided Drug Des.*, **2013**, *9*, 463–471.

Conference proceedings:

- **Majumdar, S.**, Dietz, L. and Chatterjee, S. Identifying Driving Factors Behind Indian Monsoon Precipitation using Model Selection based on Data Depth. *Proc. Fifth International Workshop on Climate Informatics: CI 2015. J. G. Dy, J. Emile-Geay, V. Lakshmanan, Y. Liu (Eds.)*, ISBN: 978-0-9973548-0-5, **2015**;
- With 9 authors. Predictive Modeling for Public Health: Preventing Childhood Lead Poisoning. *KDD Proceedings*, **2015**, 2039–2047;

- Mukherjee, U., **Majumdar, S.** and Chatterjee, S. Fast and Robust Supervised Learning in High Dimensions Using the Geometry of the Data. In: *Advances in Data Mining: Applications and Theoretical Aspects*, ser. *Lecture Notes in Computer Science*, **2015**, *9165*, 109–123.

Others:

- (Submitted) **Majumdar, S.**, Basak, S. C., Lungu, C. N., Diudea, M. V., Grunwald, G. D. Mathematical structural descriptors and mutagenicity assessment: A study with congeneric and diverse data sets;
- (Submitted) Han, B., **Majumdar, S.** and others. Predicting primate sources of long-term Zika virus spillover infection;
- (Submitted) Ghosh, A. and **Majumdar, S.** Ultrahigh-dimensional Robust and Efficient Sparse Regression using Non-Concave Penalized Density Power Divergence, `http://arxiv.org/abs/1802.04906`;
- (Submitted) **Majumdar, S.** and Michailidis, G. Joint Estimation and Inference for Data Integration Problems based on Multiple Multi-layered Gaussian Graphical Models, `http://arxiv.org/abs/1803.03348`;
- (Submitted) **Majumdar, S.**, Basu, S., McGue, M. and Chatterjee, S. Simultaneous Selection of Multiple Important Single Nucleotide Polymorphisms in Familial Genome Wide Association Studies data, `http://arxiv.org/abs/1802.01141`;
- (Submitted) **Majumdar, S.** and Chatterjee, S. Fast and General Best Subset Selection using Data Depth and Resampling, `https://arxiv.org/abs/1706.02429`;
  Winner of 2016 IISA conference Best Student Paper in theory and methods Award;
- (Tech. report) **Majumdar, S.**, Chatterjee, S. Robust estimation of principal components from depth-based multivariate rank covariance matrix, `http://arxiv.org/abs/1502.07042`;

Google scholar profile: `https://scholar.google.com/citations?user=wED36bwAAAAJ&hl=en`

Invited presentations

- (Upcoming) IMS New Researchers Conference, Burnaby, BC, Canada, July 2018;
- 2018 International Indian Statistical Association Conference, Gainesville, FL, May 2018;
- Savvysherpa, Inc., Minneapolis, MN, May 2018;
- 2017 International Indian Statistical Association Conference, Hyderabad, India, December 2017;
- Indian Statistical Institute, Kolkata, India, December 2017;
- (Student paper) 2016 International Indian Statistical Association Conference, Corvallis, OR, August 2016;
- (Poster) 9th International Triennial Calcutta Symposium. Kolkata, India, Dec 2015;
- (Poster) 5th International Workshop on Climate Informatics, Boulder, CO, Sep 2015.

Awards

- Best Student Paper, International Indian Statistical Association (IISA) conference, Corvallis, OR, 2016;
- School of Statistics Martin Award in Statistics, 2016-17;
- IISA Conference student travel award, 2016;
- University of Minnesota Interdisciplinary Doctoral Fellowship, 2016-17;
- 5th International Workshop on Climate Informatics travel award, 2015;
- University of Minnesota School of Statistics travel award, 2014 – 2016;
- Debesh-Kamal Scholarship for Higher Studies Abroad, Ramakrishna Mission Institute of Culture, Kolkata, India, 2012;

- KVPY national fellowship, Department of Science and Technology, Govt. of India, 2008 – 2012;
- National scholar, National Council of Educational Research and Training, Govt. of India, 2005 – 2008;
- Best Project award in state-level conference of the Association of Surgeons in India, 2011.

Academic and professional experience

- **University of Florida Informatics Institute**, Gainesville, FL: Postdoctoral Researcher, July 2017 – present;
- **University of Minnesota Twin Cities**, Minneapolis, MN: Graduate Assistant, Sep 2012 – May 2017;
- **IBM T. J. Watson Research Center**, Yorktown Heights, NY: Research intern, May 2016 – Aug 2016;
- **Santander Consumer USA**, Dallas, TX: Data Science intern, May 2015 – Aug 2015;
- **Data Science for Social Good fellowship**, University of Chicago, Chicago, IL. June 2014 – Aug 2014;
- **National Marrow Donor Program**, Minneapolis, MN: Statistician Intern, Bioinformatics division, June 2013 – Aug 2013;
- **Educational Initiatives**, Bangalore, India: Summer Intern, June 2011 – July 2011;
- **Saha Institute of Nuclear Physics**, Kolkata, India: Undergraduate Research Associate in biophysical sciences, Jan 2008 – June 2010.

Teaching experience

Teaching Assistant at School of Statistics, Univ. of Minnesota, Fall 2012 – Fall 2014;

STAT 8051 - Advanced Regression Techniques; Fall 2014;
STAT 3022 - Data Analysis; Spring 2014;
STAT 5021 - Statistical Analysis, STAT 5031 - Statistical Methods for Quality Improvement; Fall 2013;
STAT 5303 - Designing Experiments, STAT 5401 - Applied Multivariate Methods; Spring 2013;
STAT 3011 - Introduction to Statistical Analysis; Fall 2012.

Professional activities

Journal Referring: Statistica Sinica, Current Computer-Aided Drug Design, Australasian Medical Journal;

Statistical Consulting;

Affiliations: Member of the Institute of Mathematical Statistics, ENAR and International Indian Statistical Association.

Last updated: June 7, 2018