

Curriculum Vitae

Shounak Chattopadhyay

Phone: (919)-641-4283

Email: shounak.chattopadhyay@duke.edu

GitHub: <https://github.com/shounakchattopadhyay>

EDUCATION:

Doctor of Philosophy, Statistical Science 2018 – 2023 (expected)

Duke University

Dissertation Topic: Scalable and structured Bayesian nonparametric modeling
(Under the supervision of Dr. David Dunson)

Master of Statistics (M. Stat) 2016 – 2018

Indian Statistical institute, Kolkata, India

Dissertation Topic: On the time series modelling of Directional Data
(Under the supervision of Dr. Atanu Biswas)

Bachelor of Statistics (B. Stat with Honors) 2013 – 2016

Indian Statistical institute, Kolkata, India

RESEARCH INTERESTS:

Nonparametric Bayesian inference, Scalable Bayesian inference, Shape constrained function estimation, Latent factor models, Environmental health, Drug interaction modeling

RESEARCH EXPERIENCE:

Combination Drug Interaction modeling guided by structural covariate information (2021+)

with Han, R., Tansey, W., and Dunson, D. B.

- Motivated by discovering effective combination drug therapies in the context of high-throughput cancer data.
- Incorporate relevant drug and cell-line covariate information in the model to identify existing associations between cell-lines and drug combinations.
- Focus on developing an interpretable model with proper uncertainty quantification to identify synergistic drug interactions.

The Blessing of Dimensionality in Bayesian Latent Factor Modeling (2020+)

with Reeves, G. and Dunson, D. B.

- Motivated by the need for Bayesian inference in latent factor models with growing sample size and dimension.
- Develop an embarrassingly parallel algorithm to obtain a pseudo-Bayesian estimate of the underlying covariance matrix without the need for Markov chain Monte Carlo (MCMC).
- Application in the context of joint species distribution modeling.

Synergistic Interaction Modeling

(2020+)

with Dunson, D. B.

- Motivated by the need to account for possible synergistic or antagonistic drug interactions between toxic chemicals in an epidemiological setting.
- Develop the Synergistic Interaction Modeling (SIM) framework for identifying such interactions by constructing a novel prior distribution to estimate non-negative or non-positive functions.
- Focus on both interpretability of the model and improved predictive performance in comparison to existing methods by incorporating such interactions in the model.

Nearest Neighbor Dirichlet Mixtures

(2018-2020)

with Chakraborty, A. and Dunson, D. B.

- Motivated by the need to develop fast and scalable algorithms for multivariate density estimation with accurate uncertainty quantification.
- Proposed a class of Nearest Neighbor Dirichlet Mixtures for embarrassingly parallel density estimation using Bayesian models in local neighborhoods and combining them to produce a global estimate.
- The proposed method showcased robust performance with increasing dimension and efficient computation with increasing sample size.
- Desirable asymptotic properties are derived providing frequentist justification of the method in terms of consistency and uncertainty quantification.

On the time series modeling of Directional Data

(2017-2018)

with Biswas, A.

- Motivated by the need to develop interpretable models for analysis of directional data.
- Explored Pegram's operator to propose a conditional distribution of one categorical variable given another categorical variable.
- An induced autoregressive model of order 1 using Wrapped Symmetric distributions and Asymmetric Geometric distributions was studied.

Comparison of Horseshoe priors with other shrinkage estimators

(2017)

with Bhattacharya, A. and Mallick, B. K.

- Studied the performance of the Horseshoe (HS) prior in comparison to other well-known shrinkage priors such as the LASSO, MCP and SCAD.
- Investigated performance under model misspecification to assess robustness of methods.
- A modification of the Horseshoe prior to capture weak signals was explored.
- We compared the methods under a three-cluster data generating model where the signals arrive in two categories of strength, instead of the usual spike-and-slab model.

Comparison of Difficulty Levels of different examinations

(2017)

with Sengupta, D.

- Compared difficulty levels of two different sorts of questions in a questionnaire.
- Explored Rasch models and proposed a method to compare between the maximum, minimum, and mean difficulty levels of two questionnaires.

Image Segmentation Using Mixture Models

(2016)

with Parui, S.K.

- Motivated by modeling pixel distribution in images by a discrete mixture model.
- Developed a mixture model of multivariate Laplace distributions as a robust alternative to existing models of pixel distribution densities.
- Application to standard data sets provided improvements over the Gaussian mixture model.

Analysis of Survey Methodology adopted in Inflation-Expectation Survey of Households

(2016)

with Sengupta, D.

- Analysis of the Inflation-Expectation Survey of Households (IESH) carried out by the Reserve Bank of India (RBI).
- Proposed a variance estimator for current, 3-month ahead and 1-year ahead inflation expectations using bootstrap methods.
- Used precise GPS locations of each investigator to detect anomalies in data collection.

Data Classification using Logistic Regression in Asymmetric Cost scenario

(2015)

with Sarkar, U.K.

- Motivated by the presence of asymmetric cost functions in classification problems.
- Implemented a logistic regression model with a variable cutoff parameter determined using the magnitudes of the asymmetric misclassification costs.
- Compared with other methods such as neural networks and decision trees.

TEACHING EXPERIENCE:

- Teaching Assistant, STA 325 (Data Mining and Machine Learning), Duke University, Fall 2021.
- Teaching Assistant, STA 532 (Theory of Statistical Inference), Duke University, Spring 2021.
- Teaching Assistant, STA 111 (Probability and Statistical Inference), Duke University, Spring 2020.

PUBLICATIONS:

- **Chattopadhyay, S.**, Chakraborty, A., and Dunson, D.B. "Nearest Neighbor Dirichlet Mixtures." (*under revision*, ***Journal of Machine Learning Research***, 2021+) [[Arxiv](#)]

WORKING PAPERS:

- **Chattopadhyay, S.** and Dunson, D.B. "Synergistic Interaction Modeling."
- **Chattopadhyay, S.**, Reeves, G., and Dunson, D.B. "The Blessing of Dimensionality in Bayesian Latent Factor Modeling."
- **Chattopadhyay, S.**, Han, R., Tansey, W., and Dunson, D.B. "Combination Drug Interaction modeling guided by structural covariate information."

AWARDS:

- Winner of Capital One Data Challenge, Indian Statistical Institute, Kolkata (2017).
- Winner of Capital One Data Challenge, Indian Statistical Institute, Kolkata (2016).
- Awards in form of book grants for outstanding academic performance, Indian Statistical Institute, Kolkata (2016-18).
- Recipient of the **INSPIRE** scholarship from The Department of Science and Technology, India (2013 – 2018).

CONTRIBUTED TALKS AND POSTERS:

- Synergistic Interaction Detection. *Contributed Session at World Meeting of the International Society for Bayesian Analysis*, June 2021.
- Nearest Neighbor Dirichlet Process. *Contributed Poster Presentation at Joint Statistical Meetings*, August 2020.
- Scalable Bayesian Methods. *PCM Gold Medal Award Presentation at Indian Statistical Institute, Kolkata*, July 2018.

TECHNICAL SKILLS:

- R package **NNDM**: Implementation of the Nearest Neighbor Dirichlet Mixtures. [\[Link\]](#)
- Proficient in R, Python, C++, LaTeX, and Microsoft Office.