

PURVASHA CHAKRAVARTI

Department of Statistics and Data Science
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA, USA 15213

Email: pchakrav@stat.cmu.edu
Phone: +1 (412) 478-1908
Homepage: stat.cmu.edu/~pchakrav

EDUCATION

Department of Statistics and Data Science, Carnegie Mellon University *May 2020*
PhD in Statistics *(expected)*

Thesis Title: Gaussian Mixture Clustering Using Relative Tests of Fit

Thesis committee: Larry Wasserman (Chair), Sivaraman Balakrishnan, Mikael Kuusela,
Andrew Nobel, Rebecca Nugent, Alessandro Rinaldo

Machine Learning Department, Carnegie Mellon University *May 2018*
Secondary Masters in ML

Indian Statistical Institute, Kolkata *May 2014*
Master of Statistics
Specialization: Mathematical Statistics and Probability

Indian Statistical Institute, Kolkata *May 2012*
Bachelor of Statistics (*Hons.*)

RESEARCH INTERESTS

Methodologies: Clustering; Hypothesis Testing; Density Estimation; Optimization; Neural Networks; Graphical Models.

Applications: Particle Physics; Astrostatistics; Neuroscience; Image Processing; Signal Processing.

PUBLICATIONS

Gaussian Mixture Clustering Using Relative Tests of Fit
Purvasha Chakravarti, Sivaraman Balakrishnan and Larry Wasserman

- Submitted to Journal of the American Statistical Association (JASA) Theory and Methods (2019).
- Preprint - [arXiv:1910.02566](https://arxiv.org/abs/1910.02566)

A Generalization of Convolutional Neural Networks to Graph-Structured Data
Yotam Hechtlinger, **Purvasha Chakravarti** and Jining Qin

- Preprint - [arXiv:1704.08165](https://arxiv.org/abs/1704.08165)

Spatially Adaptive Kernel Regression Using Risk Estimation
Sunder Ram Krishnan, Chandra Sekhar Seelamantula and **Purvasha Chakravarti**

- Published in [IEEE Signal Processing Letters](https://doi.org/10.1109/LSP.2014.2358444), 2014

Gaussian Mixture Clustering Using Relative Tests of Fit*Joint with Sivaraman Balakrishnan and Larry Wasserman*

- We develop a test (RIFT) for whether a mixture of Gaussians provides a better fit to the data relative to a single Gaussian, without assuming that either model is correct. We then use this test to develop a clustering algorithm that comes equipped with significance guarantees. We show how the test can be used in a hierarchical as well as a sequential manner for clustering.

Model-Independent Detection of New Physics Signals Using Semi-Supervised Classifiers*Joint with Mikael Kuusela and Larry Wasserman*

- In high energy physics, an important problem is to detect if there is any significant difference between the distribution of just background events (generated from an assumed Monte Carlo model) and the distribution of the actual observations, which could be a mixture of background and signal events. We propose two tests, one based on the performance of a classifier in differentiating between two distributions and the other that is based on Mixture Models.

A Generalization of Convolutional Neural Networks to Graph-Structured Data*Joint with Yotam Hechtlinger and Jining Qin*

- We propose a novel spatial convolution that can be applied to graph-structured data. We use a random walk to quantify a general notion of neighborhood for graph-structured data, analogous to the way the standard convolution uses the spatial neighborhood of a pixel on the grid. The convolution has an intuitive interpretation, is efficient and scalable and can also be used on data with varying graph structure.

Statistical Analysis of the Chikungunya Fever*Joint with William F. Eddy and Virginia Dato*

- This research studies the spread of Chikungunya in the Americas. We propose two models for modeling the number of infected cases of Chikungunya: a multi-country SIR model and a multi-country ARIMA model, which is a multivariate time series model. We use the former model to explain the rate of spread of the disease and the latter model to forecast the number of new infected cases of Chikungunya.

Spatially Adaptive Kernel Regression Using Risk Estimation*Joint with Sunder Ram Krishnan and Chandra Sekhar Seelamantula*

- An important question in kernel regression is one of estimating the order and bandwidth parameters from available noisy data. We propose to solve the problem within a risk estimation framework. Considering an independent and identically distributed (i.i.d.) Gaussian observations model, we use Stein's unbiased risk estimator (SURE) to estimate a weighted mean-square error (MSE) risk, and optimize it with respect to the order and bandwidth parameters. We consider the problem of image restoration from uniform/non-uniform data, and show that the SURE approach to spatially adaptive kernel regression results in better quality estimation compared with its spatially non-adaptive counterparts.

SELECTED TALKS

- Gaussian Mixture Clustering Using Relative Tests of Fit** 2019
Contributed Talk at Joint Statistical Meetings, Denver, Colorado.
- A Generalization of Convolutional Neural Networks to Graph-Structured Data** 2019
Poster at The Science of Deep Learning, National Academy of Sciences Arthur M. Sackler Colloquium, Washington, D.C.
- Hierarchical Significance Testing for Gaussian Mixture Clustering** 2018
Contributed Talk at Joint Statistical Meetings, Vancouver, Canada.
- Gaussian Mixture Clustering Using Relative Tests of Fit (RIFTs)** 2018
Working Group on Model-Based Clustering Summer Session, Ann Arbor, Michigan.
- Statistical Significance of k-Means Clustering** 2017
Contributed Talk at Joint Statistical Meetings, Baltimore, Maryland.
- Statistical Analysis of the Chikungunya Fever** 2016
Women in Statistics and Data Science Conference, Charlotte, North Carolina.
- Women in Statistics at Carnegie Mellon University** 2016
Women in Statistics and Data Science Conference, Charlotte, North Carolina.
Jointly presented with Shannon Gallagher.

TEACHING EXPERIENCE

• Instructor

- CMU 36-225 Introduction to Probability Theory** Summer 2019
9 credits undergraduate course with 43 students
Evaluation (5.00) (Response Rate: 91%): Overall = 4.1; Teaching = 4.15
- CMU 36-200 Reasoning with Data** Summer 2018
9 credits undergraduate course with 8 students
Evaluation (5.00) (Response Rate: 25%): Overall = 5; Teaching = 3.5
- CMU 36-226 Introduction to Statistical Inference** Summer 2017, 2016
9 credits undergraduate course with 31 students and 26 students respectively.
Evaluation (5.00) (Response Rate: 57%, 67%): Overall = 4.69, 4.42; Teaching = 4.75, 4.42

• Teaching Assistant

- CMU 36-303 Sampling, Survey and Society** Spring 2019
- CMU 36-401 Modern Regression** Fall 2018, 2014
- CMU 36-402 Advanced Methods for Data Analysis** Spring 2018, 2017, 2016
- CMU 36-705 Intermediate Statistics** Fall 2017, 2016
- CMU 36-225 Introduction to Probability Theory** Fall 2015
- CMU 36-625 Probability and Mathematical Statistics (Hons.)** Spring 2015

AWARDS AND HONORS

- **Honorable Mention for the 2019 Do-Bui Travel Award.**
Received an Honorable Mention for Do-Bui Travel Award given by the Caucus for Women in Statistics (CWS), Joint Statistical Meetings, 2019.
- **National Level Scholarship**
Obtained the INSPIRE Scholarship offered by the Department of Science and Technology (DST), Government of India from 2009 - 2014.
- **Cyber Olympiad**
Secured All India rank 19 in 5th National Cyber Olympiad held on 19th February, 2006.
- **Summer Fellowship**
Received Indian Academy of Science Fellowship, 2012.

SKILLS

Programming languages: Proficient in R, Matlab, and C. Experience in Python and Mathematica.

PROFESSIONAL AFFILIATION AND SERVICE

Member: CMU Women in Statistics, American Statistical Association, Institute of Electrical and Electronics Engineers.

Volunteer for Women in Data Science Pittsburgh @CMU 2018, 2019
Women in Data Science Pittsburgh @CMU Conference, Pittsburgh, PA

Panelist for Women in Statistics at Carnegie Mellon University 2016
Women in Statistics and Data Science Conference, Charlotte, NC

Gave an Outreach Talk on Opportunities in Statistics 2016
Winchester Thurston, Pittsburgh, PA

Cultural Chair (Music), Indian Graduate Student Association 2016-present
Carnegie Mellon University, Pittsburgh, PA

Convenor of Irene J.Curie Hall 2011-2013
Indian Statistical Institute, Kolkata, India

REFERENCES

Professor Larry Wasserman

UPMC Professor of Statistics and Data Science,
Department of Statistics and Data Science and Machine Learning Department,
Carnegie Mellon University.

Professor Sivaraman Balakrishnan

Assistant Professor,
Department of Statistics and Data Science and Machine Learning Department,
Carnegie Mellon University.

Professor Mikael Kuusela

Assistant Professor,
Department of Statistics and Data Science,
Carnegie Mellon University.

Professor Rebecca Nugent

Fienberg Professor of Statistics & Data Science,
Associate Department Head, and Co-Director of Undergraduate Studies
Department of Statistics and Data Science,
Carnegie Mellon University.