

Soham Mukherjee

Personal Information

Date of Birth *February 6th, 1998*
Place of Birth *Kolkata, India*
Work Address *Raum 711, Vogelpothsweg 87, 44227 Dortmund, Germany*
Mobile *+49-15730990431*
Email *soham.mukherjee@tu-dortmund.de*

Education

2022–Current **Ph.D, Faculty of Science (Department of Computer Science)**, University of Tübingen, Germany (International Max Planck Research School for Intelligent Systems).
2019–2021 **Master of Science in Statistics, School of Mathematics and Statistics**, University of Hyderabad, Hyderabad, India.
CGPA (out of 10): 8.56 Grade: First Division with Distinction
2016–2019 **Bachelor of Science (Honours) in Statistics, Bidhannagar College, WBSU, Kolkata, India.**
Percentage: 69.88% Grade: First Class Honours

Work Experience

Jan 2024–Current **Academic Scientist, Faculty of Statistics, TU Dortmund**
o Derivative Gaussian processes latent variable models with applications to Single-Cell RNA sequencing data (*Ph.D Thesis*)
Aug 2022–Dec 2023 **Academic Scientist, Innere Medizin I, Universitätsklinikum Tübingen, University of Tübingen**
o Derivative Gaussian processes latent variable models with applications to Single-Cell RNA sequencing data (*Ph.D Thesis*)
Feb 2021– Oct 2021 **Research Intern, Indian Statistical Institute, Kolkata.**
o Research and development on Bayesian stochastic volatility model using finite Gaussian mixtures with methodology illustration on high-frequency NSE-Nifty stock returns.
Jul 2020– Aug 2020 **Summer Research Trainee, Council Of Scientific And Industrial Research–North East Institute Of Science And Technology, Jorhat, Assam.**
o Research and development of atmospheric models using Regression analysis and prediction algorithm for various atmospheric target variables to draw inference on cloud formation.
May 2020– Jun 2020 **Summer Intern, Deloitte India (Offices of US), Hyderabad.**

- o Project based on text analysis, text pre-processing and sentiment analysis using Natural Language Processing algorithms.
- o Project based on exploratory data analysis and predictive models using Linear Models, CART and Random Forest.

Research Articles

Mukherjee, S., Claassen, M., & Bürkner, P. C. (2025). Hilbert space methods for approximating multi-output latent variable Gaussian processes. (*in review*) (<https://arxiv.org/abs/2505.16919>)

Mukherjee, S., Claassen, M. & Bürkner, PC. DGP-LVM: Derivative Gaussian process latent variable models. *Stat Comput* 35, 120 (2025). <https://doi.org/10.1007/s11222-025-10644-4> [Pre-print: <https://arxiv.org/abs/2404.04074>]

Mukherjee S. (2021) Bayesian Analysis of Stochastic Volatility Model using Finite Gaussian Mixtures with Unknown Number of Components (<https://arxiv.org/abs/2110.12824>)

Supervision

2023–2024 Marcello Zago

- o **M.Sc. thesis:** Derivative Gaussian processes for Single-Cell RNA sequencing data using combined covariance functions.

Teaching

- 2025 Seminar on model comparison: Summer semester (TU Dortmund)
- 2024 Seminar on multilevel models: Winter semester (TU Dortmund)
- 2024 Intensive course in Statistics: Summer semester (TU Dortmund)

Conference/Workshops

- 2025 **Royal Statistical Society International Conference 2025 (RSS 2025):** contributed talk on "DGP-LVM: Derivative Gaussian process latent variable models." (upcoming)
- 2024 **International Conference on Statistics and Data Science (ICSDS):** contributed talk on "DGP-LVM: Derivative Gaussian process latent variable models."
- 2024 **International Society for Bayesian Analysis:** attended
- 2023 **Cambridge Ellis Unit Summer School on Probabilistic Machine Learning 2023:** poster presentation on "Derivative Gaussian processes latent variable models with applications to Single-Cell RNA sequencing data."
- 2021 Lectures in Statistics, Indo-French Workshop organized by IFCAM and ISI
- 2020 Prof. C.R. Rao Birth Centenary Conference on Statistics and Applications, 2020 funded by DST-SERB, Govt. of India.
- 2017 MTUSS-2017 (Mathematics Training for Undergraduate Statistics Students) programme comprising courses in Foundations, Real Analysis, Linear Algebra and Probability funded by NBHM.

Independent Projects

July 2021 M.Sc Thesis

- o Research and modelling of Supermassive black hole mass using Bayesian generalized linear mixed models and Bayesian missing value methods.
- May 2020 Worldwide COVID-19 data modelling and forecasting
 - o Team project for visualization and forecast of the time series data with ARIMA, Holt-Winters Exponential Smoothing, Polynomial Regression and Support Vector Machines on the effect of the pandemic. <https://www.kaggle.com/soham6298/covid-19-uoh-d5380e>
- Feb 2019 B.Sc. Thesis
 - o Study of forecasting methods for Stock Market price index: A comparative analysis of ARIMA and Exponential Smoothing methods for stock market price data.

Technical and Language Skills

Programming and Software R, Stan, JAGS, Pyro-PPL

Bengali Native proficiency

English TOEFL iBT (Sept 2021): 114/120; Reading: 30, Listening: 28, Speaking: 29, Writing: 27

German 3 years course funded by IMPRS-IS (upto B1)

Hindi Professional proficiency