Soham Mukherjee

Personal information

Date of Birth	*February 6th, 1998*
Place of Birth	Kolkata, India
Work Address	Konemiehentie 2, 02150 Espoo, Finland
Mobile	+358 50 3434402
Personal Email	sohammukherjee1998@gmail.com
Work Email	soham.mukherjee@aalto.fi

Education

2022–2026 **Ph.D,** *Faculty of Science,* University of Tübingen, 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

Mar 2026–Current **Postdoctoral Researcher**, *Department of Computer Science, Aalto University*, Espoo, Finland.

- Bayesian workflow, model comparison, assessment and diagnostics

Jan 2026–Mar 2026 **Academic Scientist**, *M3 Forschungszentrum, Universitätsklinikum Tübingen, University of Tübingen*, Tübingen, Germany.

- On latent variable estimation using derivative Gaussian processes (*Ph.D Thesis*)

Jan 2024–Dec 2025 **Academic Scientist**, *Faculty of Statistics, TU Dortmund*, Dortmund, Germany.

- On latent variable estimation using derivative Gaussian processes (*Ph.D Thesis*)

Aug 2022–Dec 2023 **Academic Scientist**, *Innere Medizin I, Universitätsklinikum Tübingen, University of Tübingen*, Tübingen, Germany.

- On latent variable estimation using derivative Gaussian processes (*Ph.D Thesis*)

Feb 2021– Oct 2021 **Research Intern,** *Indian Statistical Institute,* Kolkata, India.

- On Bayesian stochastic volatility model using finite Gaussian mixtures with applications to 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, India.

- Regression analysis and prediction algorithm for various atmospheric target variables for cloud formation inference.

May 2020– Jun 2020 **Summer Intern,** *Deloitte India (Offices of US)*, Hyderabad, India.

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

Supervision

2023–2024 Marcello Zago

- **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."

2024 **International Conference on Statistics and Data Science (ICSDS 2024)**: 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.

Theses and other projects

Dec 2025 Ph.D. Thesis

- On latent variable estimation using derivative Gaussian processes.

July 2021 M.Sc Thesis

- 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

- 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

- 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

Software	R, Stan, JAGS, Pyro-PPL
Bengali	Native proficiency
English	TOEFL iBT (Sept 2021): 114/120; Reading: 30, Listening: 28, Speaking: 29, Writing: 27
Hindi	Professional proficiency
German	3 years course funded by IMPRS-IS (upto B1)

Research articles

Mukherjee S., Claassen M., & Bürkner P. C. (2026). Hilbert space methods for approximating multi-output latent variable Gaussian processes. *Stat Comput* 36, 110. `https://doi.org/10.1007/s11222-026-10869-x` [Pre-print: `https://arxiv.org/abs/2505.16919`]

Mukherjee S., Aguilar J. E., Zago M., Claassen M. & Bürkner P. C. (2025). Latent variable estimation with composite Hilbert space Gaussian processes.(*in review*)(`https://arxiv.org/abs/2510.25371`)

Bischoff S., Poličar P.G., **Mukherjee S.**, Macke J.H., Claassen M., & Schröder C. (2025). velotest: Statistical assessment of RNA velocity embeddings reveals quality differences for reliable trajectory visualizations.(*in review*)(`https://doi.org/10.1101/2025.10.26.683064`)

Mukherjee S., Claassen M., & Bürkner P. C. (2025). DGP-LVM: Derivative Gaussian process latent variable models. *Stat Comput* 35, 120. `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`)