NATIONAL STATISTICS DAY 2025 CELEBRATION

ONE DAY SYMPOSIUM AT IIT BOMBAY

NATIONAL STATISTICS DAY 2025 CELEBRATION
ONE DAY SYMPOSIUM AT IIT BOMBAY

Venue: Ramanujan Hall, Department of Mathematics, IIT Bombay

Date: 30th June 2025 (Monday)

Schedule:

09:45 AM - 10:00 AM: *Inauguration*

10:00 AM - 10:50 AM: *Entwined Paths: The Evolutionary Bond Between Statistics and Genetics*, Analabha Basu, National Institute of Biomedical Genomics

11:00 AM - 11:50 AM: *Climate, Data Science and Applications*, Subimal Ghosh, Indian Institute of Technology Bombay

12:00 PM - 12:50 PM: *Testing for the Rank of a Stochastic Process*, Anirvan Chakraborty, Indian Institute of Science Education and Research Kolkata

1:00 PM - 2:50 PM: *Lunch Break*

3:00 PM - 3:50 PM: *MCMC Importance Sampling via Moreau-Yosida Envelopes*, Dootika Vats, Indian Institute of Technology Kanpur

4:00 PM - 4:50 PM: *High-dimensional Central Limit Theorem*, Debraj Das, Indian Institute of Technology Bombay

5:00 PM - 5:50 PM: *Conclusion and High Tea*

NATIONAL STATISTICS DAY 2025 CELEBRATION

ONE DAY SYMPOSIUM AT IIT BOMBAY

Venue: Ramanujan Hall, Department of Mathematics, IIT Bombay

Date: 30th June 2025 (Monday)

Schedule:

09:45 AM - 10:00 AM: *Inauguration*

10:00 AM - 10:50 AM: *Entwined Paths: The Evolutionary Bond Between Statistics and Genetics,* Analabha Basu, National Institute of Biomedical Genomics

Abstract: The relationship between statistics and genetics is foundational to modern biological research. From the earliest days of Mendelian segregation analysis and Fisher's biometric models, to the complex inference frameworks underpinning genome-wide association studies (GWAS), population structure estimation, and quantitative trait mapping, the fields have co-evolved in deep synergy. Advances in one have continuously driven progress in the other: the high-dimensionality and noise inherent in genetic data have posed statistical challenges that inspired new models and computational tools, while statistical innovation has enabled robust interpretation of genetic variation, inheritance patterns, and evolutionary dynamics. This abstract reflects on the historical trajectory, mutual influence, and future outlook of this enduring partnership—arguing that the convergence of these disciplines is not incidental but essential, and will remain so as genetics increasingly integrates with large-scale data science, causal inference, and personalized medicine.

11:00 AM - 11:50 AM: *Climate, Data Science and Applications,* Subimal Ghosh, Indian Institute of Technology Bombay

Abstract: As climate change intensifies, the demand for accurate, scalable, and actionable climate information grows rapidly. This talk explores the convergence of physics-based climate models and modern data science techniques, focusing on recent advances in machine learning, including GraphCast and Neural GCMs. It highlights the growing role of physics-informed AI in downscaling, emulation, and causal analysis, while questioning longstanding assumptions such as the universality of regional models and bias correction techniques. The presentation also emphasizes real-world applications, showcasing India's first operational urban flood forecasting system built by IIT Bombay in collaboration with NCCR, Chennai. Drawing on case studies from Mumbai's 2024 monsoon, we illustrate the integration of radar data, sensors, crowdsourced information, and AI-based forecasting. The talk aims to raise critical questions about the pace and direction of climate data science, the role of foundation models, and the imperative for systems that are not just intelligent but also interpretable and physically grounded.

NATIONAL STATISTICS DAY 2025 CELEBRATION

ONE DAY SYMPOSIUM AT IIT BOMBAY

12:00 PM - 12:50 PM: *Testing for the Rank of a Stochastic Process,* Anirvan Chakraborty, Indian Institute of Science Education and Research Kolkata

Abstract: How can we determine whether a stochastic process has a finite rank, and if so, what this precise rank is? And how can we do so at a given level of confidence? This question is central to a great deal of methods for functional data, which require low-dimensional representations whether by functional PCA or other methods. The difficulty is that the determination is to be made on the basis of iid replications of the process observed discretely and with measurement error contamination. This adds a ridge to the empirical covariance, obfuscating the underlying dimension. We build a matrix-completion inspired test statistic that circumvents this issue by measuring the best possible least square fit of the empirical covariance's off-diagonal elements, optimized over covariances of given finite rank. For a fixed grid of sufficient size, we determine the statistic's asymptotic null distribution as the number of replications grows. We then use it to construct a bootstrap implementation of a step-wise testing procedure controlling the family-wise error rate corresponding to the collection of hypotheses formalizing the question at hand. Under minimal regularity assumptions, the procedure is consistent and that its bootstrap implementation is valid. The procedure involves no tuning parameters or pre-smoothing, is indifferent to the homoskedasticity or lack of it in the measurement errors, and does not assume a low-noise regime. This talk is based on joint work with Victor M. Panaretos (EPFL, Switzerland).

1:00 PM - 2:50 PM: *Lunch Break*

3:00 PM - 3:50 PM: *MCMC Importance Sampling via Moreau-Yosida Envelopes,* Dootika Vats, Indian Institute of Technology Kanpur

Abstract: Markov chain Monte Carlo (MCMC) is the workhorse computational algorithm employed for inference in Bayesian statistics. Gradient-based MCMC algorithms are known to yield faster converging Markov chains. In modern parsimonious models, the use of non-differentiable priors is fairly standard, yielding non-differentiable posteriors. Without differentiability, gradient-based MCMC algorithms cannot be employed effectively. Recently proposed proximal MCMC approaches, however, can partially remedy this limitation. These approaches employ the Moreau-Yosida (MY) envelope to smooth the nondifferentiable prior enabling sampling from an approximation to the target posterior. In this work, we leverage properties of the MY envelope to construct an importance sampling paradigm to correct for this approximation error. We establish asymptotic normality of the importance sampling estimators with an explicit expression for the asymptotic variance which we use to derive a practical metric of sampling efficiency. Numerical studies show that the proposed scheme can yield lower variance estimators compared to existing proximal MCMC alternatives.

NATIONAL STATISTICS DAY 2025 CELEBRATION

ONE DAY SYMPOSIUM AT IIT BOMBAY

4:00 PM - 4:50 PM: *High-dimensional Central Limit Theorem,* Debraj Das, Indian Institute of Technology Bombay

Abstract: Central Limit Theorem (CLT) is one of the oldest as well as remarkable results of classical probability theory. In most simplest words, CLT is a statement about the convergence of properly centered and scaled sample mean of a sequence of random vectors to the Gaussian random vector in distribution. The recent interest lies in establishing CLT when the dimension also grows with the sample size. I will shed light on recent developments as well as describe some necessary and sufficient results related to critical growth rate of dimension. I will state a characterisation of the Normal distribution in terms of high dimensional CLT. I will also try to highlight the benefits of self-normalization/studentization in reducing the requirement of existence of exponential moments to some polynomial moments.

5:00 PM - 5:50 PM: *Conclusion and High Tea*