Annesh Pal

✉ | in | 🌐

About Me

I am currently pursuing a doctoral degree in Biostatistics at the University of Bordeaux. My primary interests lie in Bayesian Statistics, Mathematical modeling and Machine Learning and their applications on biological and medical fields. I am also a self-taught artist and musician, with a keen interest in sports.

Education

Integrated Bachelor's and Master's of Science | ***Major: Mathematics*** — Aug. 2017 – May 2022
Indian Institute of Science Education and Research (IISER) Kolkata — CGPA: 7.88 (out of 10)
Academic Transcript

Higher Secondary Education — 2017
Central Board of Secondary Education — 94.2 %

Research

Doctoral Thesis — December 2023 – ongoing
Dr. Boris Hejblum, Prof. Dr. Rodolphe Thiébaut — University of Bordeaux, France

- **Non-parametric Bayesian mixture modeling applied to cellular deconvolution from bulk RNA-seq data**: Our objective is to develop a novel semi-supervised method to infer cellular proportion from bulk transcriptomics, predict outcomes, and evaluate surrogacy for high-dimensional data sets. Working in a Bayesian non-parametric paradigm, we propose a Dirichlet process (DP) mixture model for kernel distributions of RNA-seq data (Gaussian, Negative-Binomial) to allow the discovery of new mixture components, while using variational inference to tackle the underlying model complexity and the high dimensionality of such data. Additionally, we plan to incorporate weakly informative prior on the covariance matrix, accompanied by efficient software development in R, and apply in immunological substudies embedded within COVERAGE and PREVACUP clinical trials.

Independent Project (remote) — March 2023 – May 2023
Dr. Boris Hejblum — University of Bordeaux, France

- **Deconvoluting gene expression data using Bayesian methods**: Based on the Bayesian Log-Normal Deconvolution (BLADE) method to estimate both cell fraction and gene expression for each cell type from bulk gene expression data (*Barbosa et al., 2021*), we worked on developing a BLADE package in R to enable further refinement and practical applications.

Independent Project (remote) — November 2022 – December 2022
Dr. Minerva Mukhopadhyay — Indian Institute of Technology (IIT) Kanpur, India

- **Bayesian Variable Selection for Clustering**: Bayesian variable selection models have been developed to address clustering and variable selection simultaneously for high-dimensional gene expression data. Our work focused on comprehending the Dirichlet Process Prior (DPP) and Mixture of Finite Mixture (MFM) models and their applications on DNA microarray data, with particular emphasis on the choice of prior selection.

Master's Thesis — May 2021 – May 2022
Supervisor: Dr. Ashok K. Bansal, Co-Supervisor: Dr. Satyaki Mazumdar — IISER Kolkata, India

- **Estimating Fish Population using Capture-Recapture Method**: Population estimation plays a key role in ecological and epidemiological studies. We worked on developing Bayesian approaches for Capture-recapture methods for estimating fish population. The emphasis was on the subjective nature of Bayesian paradigm where we illustrated the sensitivity of estimate of the population parameter with respect to different Non-Informative priors from a decision-theoretic point of view (by considering a variety of loss functions).
Thesis blueprint

Term Project August 2021 – May 2022
Dr. Kripabandhu Ghosh IISER Kolkata, India

- **Text classification using Machine Learning techniques**: Text mining using four supervised learning methods (Naive-Bayes, KNN, SentenceBERT and Universal Sentence Encoder) techniques on the Artificial Intelligence for Legal Assistance (AILA) 2019 dataset to identify relevant prior cases and statutes based on 50 queries, and comparing each method using Mean Average Precision (MAP) scores.

Summer Research Project May 2021 – August 2021
Dr. Koel Das IISER Kolkata, India

- **Attitudes towards vaccines and intention to vaccinate against COVID-19**: Analyzing nominal data from the Indian population on attitudes toward COVID-19 vaccination using multinomial regression, one-way ANOVA, and Cronbach's Alpha, with additional focus on correlating attitudes toward self and child vaccination.

Summer Research Project May 2020 – July 2020
Dr. Bidisha Sinha IISER Kolkata, India

- **Modeling protein folding from Primary structure to Tertiary structure**: Quantifying the protein folding Landscape statistically by using Probability Models and Markov chains. ***(Link)***

WORKSHOP

Vijyoshi National Science Camp December 2017
Kishore Vaigyanik Protsahan Yojana (KVPY), INSPIRE and DST, India

Approximate Inference in Theory and Practice June 2024
ESSEC, Paris, France

HADACA3 Health Data challenge December 2024
PEPR Santé Numérique M4DI project, Aussois, France

SCHOLARSHIP

Innovation in Science Pursuit for Inspired Research (INSPIRE) Scholarship August 2017-May 2022
The Department of Science and Technology (**DST**), Government of India.

SKILLS

Languages: English (C2), Bengali (Native, C2), Hindi (C2), Sanskrit (C1) and Kokborok (A2)
Programming: R, MATLAB, Python (NumPy, SciPy, Matplotlib, Pandas)
Digital: Github, LaTex, Microsoft Office Suite, Canva, Photoshop