Nabarun Deb

Curriculum Vitae/Resume

547 Riverside Drive, Apartment 1F
NY - 10027, New York, USA
+1 9174037243
nabrog.nabarun2@gmail.com/nd2560@columbia.edu

Education

2017- **Doctor of Philosophy**, *Columbia University*, New York, USA.
-*Department of Statistics.*

2015-2017 **Master of Statistics (M. Stat)**, *Indian Statistical Institute*, Kolkata, India.
-Specialization: Theoretical Statistics.
- First Division with Distinction.

2012-15 **Bachelor of Statistics (B. Stat)(Hons)**, *Indian Statistical Institute*, Kolkata, India.
- First Division with Distinction

2012 **Indian School Certificate (ISC), Std 12**, *Julien Day School*, Ganganagar, India.
- Total Percentage Score (Best 4 including English) : 96.75%
- Courses: English, Mathematics, Physics, Chemistry, Computer Application.

2010 **Indian Certificate of Secondary Education (ICSE), Std 10**, *Julien Day School*, Ganganagar, India.
-Total Percentage Score (Best 5 including English) : 95.8%
-Courses: Bengali, English, Mathematics, Physics, Chemistry, Biology, History, Geography, Computer Science.

Professional Experience

May-July 2015 Research Internship under *Prof. Jayanta K. Ghosh*, Department of Statistics, Purdue University, Indiana, USA.

May-July 2016 Summer Internship with *Polaris Networks*, Dynamic Digital Technology, Sector-V, Kol-091

Awards and Achievements

2016 **Got the opportunity to pursue dissertation in M. Stat 2^{nd} year**.
Only 7 other students of our batch have been given this prestigious opportunity.

2010,2011 **Successful in *Regional Mathematical Olympiad* twice in as many appearances for consecutive years by virtue of being within the top 35 in the state.** .

2012 **An INMO *Indian National Mathematical Olympiad* merit certificate holder for being among the top 75 students in the country.**.

2012- **A recipient of the award of Scholarship for Higher Education (SHE) under Innovation in Science Pursuit for Inspired Research *INSPIRE***.
It is awarded to the top 1% students in the ISC Examination, 2012

2012 **Selected for admission in *Indian Statistical Institute Kolkata, Chennai Mathematical Institute, Institute of Mathematics And Applications, Orissa* in the year 2012.**.

2012- **Awarded prize money in the form of book grants**, for excellent academic performance in five out of eight semesters (first, fifth, sixth in B. Stat, all two in M.Stat) by Indian Statistical Institute during the Bachelor of Statistics and Master of Statistics Programmes (so far). .

2006 **Awarded distinction in the International competitions and Assessments for Schools (*ICAS*) organised by the University of New South Wales for Mathematics, Science and English.**.

2007,2009 **Ranked 128^{th} and 87^{th} in the final round of the Sixth and Eighth National Cyber Olympiads (respectively) organised by the *Science Olympiad Foundation.***.

2010-2012 **A school level topper throughout including a 100% score in Mathematics**, throughout classes 11, 12, ICSE and ISC Board Examinations..

2011 **The 2^{nd} runner-up in the TTIS Math Mania Finals.** .

Final Year Dissertation

Title **Likelihood ratio tests for monotone functions in non-separated models.** *(In progress)*

Supervisor *Prof. Moulinath Banerjee*, Department of Statistics, University of Michigan and *Prof. Anil K. Ghosh*, Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, Kolkata).

Abstract The project investigates the behavior of the likelihood ratio statistic for making inference on a monotone function of interest at a point in statistical models where the likelihood does not naturally separate in terms of the parameters of interest. Separated models on the other hand are fairly well-understood based on the work of Banerjee and Wellner (2001), Banerjee (2007) and Groeneboom and Jongbloed (2015).
The goal is to computationally study the likelihood ratio in a class of non-separated models of considerable interest in statistical applications, like the Case 2 and mixed-case interval censoring models (which are of particular interest in epidemiological studies), as well as counting process models observed at random observation times. The computation of such likelihoods involve convex optimization algorithms. At this point we conjecture that the likelihood ratio statistic in these models should have the same asymptotic distribution in the non-separated case as that demonstrated in Banerjee and Wellner (2001).
Another goal is to study the behavior of such likelihood ratio statistics analytically, where we anticipate that theoretical tools of greater complexity will need to be employed than those for non-separated models.

We have already addressed the convex optimisation problem for non-separated models, typically in the Case 2 Interval Censoring situation using a variant of Jongbloed's Iterative Convex Minorant Algorithm (ICMA). Also, we have computationally analysed the asymptotics for the likelihood ratio statistic which does fall in line with the existing conjectures.

Project Experience

Mar 2016-Present **Joint analysis of multiple longitudinal traits for detecting anomalous patients in clinical studies.** *(Manuscript in progress)*

Supervisor *Prof. Kiranmoy Das*, Interdisciplinary Statistical Research Unit, Indian Statistical Institute, Kolkata.

Position Project member jointly with Debarghya Mukherjee.

Description Joint analysis of multiple traits has become the recent focus of the biomedical studies. Different traits, though appear to function separately, can jointly control a dynamic system. For example, the cardiovascular system of a patient is jointly controlled by blood pressure, blood sugar, total cholesterol, Na level, K level, oxygen in the blood, and B-type natriuretic peptide. Thus, for monitoring the cardiovascular system of patients we need to model these traits jointly. The superiority of a joint model compared to a separate model for each trait has already been established in the statistics literature.

In this work, we propose a dynamic joint model for multiple longitudinal traits, some of which are binary and/or count and rest are continuous. Multiple comparison procedures are implemented for controlling the overall type-I error rate. The practical usefulness of our model and estimation method is assessed through extensive simulation studies. Our model can accurately detect anomalous patients (those under critical condition) at the correct time point and we also estimate the average sample number needed for detecting an anomalous patient. Our model will be extremely useful in biomedical and/or clinical studies on humans, animals and plants.
The manuscript is under preparation now and will be submitted shortly to the *Statistical Methods in Medical Research.*

May 2015-July 2015 **Inference and penalties under model misspecifications**

Supervisor *Prof. Jayanta K. Ghosh*, Department of Statistics, Purdue University, Indiana, USA.

Description Read about the problems of standard inferential procedures in misspecified models. For instance, in *Jim Berger's* paper from *Encyclopedia of Statistical Sciences*, an interesting problem that has been addressed is as follows: *Given a sample of i.i.d observations* $X_1, X_2, \cdots X_n$, *how would you conclude whether it comes from a Gaussian* (μ, σ^2) *or a Cauchy* (μ, σ) *distribution?* A number of problems come up while approaching this problem from a classical view. A bayesian perspective simplifies the task, although the solution is not all that satisfactory when the sample size is small. Also, in *Statistical Models and Causal Inference*, David Freedman has argued that in social life data, nearly all standard assumptions are violated and usual statistical modelling falls apart.
During my stay, I have made attempts to analyse the penalties caused by model misspecifications using Kulback-Leibler Divergence. Some very intuitive methods of classification (like characterizing the jumps that are prevalent in data generated from the Cauchy distribution) yielded good results. Further, attempts were made to distinguish between high-dimensional distributions after projecting them into lower dimensional spaces.

July 2015-Oct 2015 **Credit scoring using classification trees and a nonhomogeneous poisson regression model.**

Supervisor *Prof. Diganta Mukherjee*, Sampling and Official Statistics Unit, Indian Statistical Institute, Kolkata.

Description Credit scoring is an important problem in the modern world. Standard methods used do not take into account the macroeconomic indicators which are bound to have an impact even at an indiviual level. Our work involves classification trees to obtain the scores as opposed to the standard methods of logistic and poisson regression, and has *yielded marginally better results*. Another step in this direction is to incorporate the macro-economic factors that vary with time. To do this, we worked on a nonhomogeneous poisson process model. A similar model was studied in Massey, Parker and Whitt(1995) - Estimating the parameters of a non homogeneous Poisson process with linear rate .
This project was a part of the Statistics Comprehensive Course in which I was the highest scorer.

May 2016 - July 2016 **Cyber Security and dynamic anomaly detection**

Position Summer intern at Dynamic Digital Technology, *Polaris Networks*.

Description The purpose of this project was to detect intrusion by a foreign user/BOT in real time. A dataset with regular users and malicious BOTs was provided by the Indian Concern Dynamic Digital Technology. It featured attributes like upload bytes, download bytes, request and referrer header content lengths, HTTP exchanges between a browser and a remote host etc. of the employees collected over a period of 1 month involving 20 users. We used Bayesian estimation method to determine anomalous web sessions. This approach can help in detecting new attacks without having to know their characteristics in advance (Friedman and Singer 1999). This method yielded significantly lesser error rates compared to those implemented by the company, and hence was integrated into the AEGIS software. Two new methods gave improved results. One based on quantile based binning followed by bimodal probability mass estimation, and the other a likelihood based approach with a log concave mixture fit.

Jan 2016 - Mar 2016 **Malicious node detection in a sensor network.**

Supervisor *Prof. Kiranmoy Das*, Interdisciplinary Statistical Research Unit, Indian Statistical Institute, Kolkata.

Description We have implemented Systematic Information Sampling (SIS) and Metropolis Hastings Algorithm to estimate the different model parameters in a sensor network. The target was to be able to model a mobile malicious node as it travels across the network and to sequentially estimate the probability of detection within time point t. This work made way for the work on joint analysis of multiple longitudinal traits which followed soon.

May 2014 - July 2014 **Regression techniques, mixture distributions and iterative algorithms.**

Supervisor *Dr. Saurabh Ghosh*, Human Genetics Unit, Indian Statistical Institute, Kolkata) .

Description The purpose of this project was to explore different existing statistical machineries. For instance, regression analysis (under different loss functions), the EM algorithm and other iterative algorithms like Newton-Raphson, Fisher Scoring (rates of convergence, choice of initial values, etc). We have also analysed a data set pertaining to the incidence of Coronary Artery Disease in individuals. Among other things, missing value imputation was carried out.

May 2013 - July 2013 **Analysing the nature of cycles in RC4 (Rivest Cipher 4)**

Supervisor *Dr. Subhamoy Maitra*, Applied Statistics Unit, Indian Statistical Institute, Kolkata .

Description RC4 is generally used to generate a pseudorandom stream of bits (a keystream) for encryption. However very little is known about the possible cycle lengths for different permutations and for different values of index pointers. We could make out that the Pseudo-random generating Algorithm (PRGA) was one-one and we had some progress on predicting the periodicity of cycle lengths for the identity permutation.

Computer skills

Software Packages R, Matlab, Mathematica, Geogebra LaTeX, HTML, Macromedia Flash

Programming Languages BASIC, JAVA, C, C++,

A recurrent member of the website designing committee for *Integration*, the annual techno-cultural-sports fest of Indian Statistical Institute, Kolkata.

Miscellaneous information

Communication Skills Fluent in speaking and writing Bengali, English, Hindi. I have also given a number of presentations in Regression Techniques, Introduction to Markov Chains and Statistics Comprehensive courses which have all been appreciated.

Extracurricular Achievements Best Speaker in the *Frank Anthony* Memorial All India Inter School Debate Competition, 2012 (Senior group).

Best speaker in Inter Branch Debate competition and multiple times winner in elocution and extempore in talent contests (organised by Julien Day School).

Hobbies An avid sports lover, particularly interested in following cricket, lawn tennis and table tennis. I like playing table tennis as well. I was among the top 4 in the annual table tennis tourney in the Boys' hostel, Indian Statistical Institute, 2013.

A voracious reader and poetry lover. I am into writing prose passages myself and some of the writings have also been published in school magazines.

I enjoy watching movies, listening to music and playing the synthesizer at leisure

Positions of Responsibility Head of the web designing committee, Technical Affairs Committee and also a core-committee member of the annual techno-cultural-sports fest (*Integration*) of Indian Statistical Institute, Kolkata in 2013 and 2014.

Event head for the table tennis tourney, band-e-thlon contest (musical event), MystiQ (puzzle hunt), Technical workshops, creative writing, extempore and the Mathematical Talent Reward Programme multiple times.

Many of the students I taught were successful in Mathematics Olympiads and the Indian Statistical Institute Entrance Examination.

Member of the Mess Committee, Literary Affairs Committee, Sports Quiz Committee and the "Treasure Hunt" committee (an annual game show) in the Boys' hostel .

I hereby do certify that all the above information is true to the best of my knowledge.
Nabarun Deb