Debolina Ghatak

| CONTACT INFORMATION | Computer Science and Engineering Department
Indian Institute of Technlogy Jammu
1a, NH 44
Nagrota 181221
Jammu and Kashmir, India | *Phone:* (+91) 9674645433
E-mail: deboghatak@gmail.com |

| RESEARCH INTERESTS | Data Obfuscation, Data Masking, Deconvolution Theory, Data Perturbation |

| WORK EXPERIENCE | Presently working as a **Visiting Faculty** at Indian Institute of Technlogy Jammu and the courses taught are given below |

1. Probability and Stochastic Processes (Semester III Course for BTech and Backlog Course for all UG students)

2. Introduction to Statistics (Open elective to UG students)

3. Discrete Mathematical Structures (For Semester III BTech)

4. Probability and Statistics (For MTech First year).

Presently the course I am teaching is,

1. Introduction to Statistics (For Semester VI BTech)

| EDUCATION | **PhD in Statistics** | January 2019 |
| | Indian Statistical Institute, Kolkata | Thesis Advisor: Prof. Bimal K. Roy |

PAPERS PUBLISHED

1. Ghatak, D. and Roy, B.K., Estimation of True Quantiles from Quantitative Data Obfuscated with Additive Noise, *Journal of Official Statistics*, Vol. 34, No. 3, 2018, pp. 1-24, DOI: http://dx.doi.org/10.2478/JOS-2018-0032.

2. Ghatak, D. and Roy, B.K., Conditional Masking to Numerical Data, *Journal of Statistical Theory and Practise*, 2019, DOI 10.1007/s42519-019-0042-y

3. Ghatak, D. and Roy, B.K., An Improved Bound for Security in an Identity Disclosure Problem. *International Journal of Statistics and Probability* Vol 8, No. 3, 2019, pp. 24-31

POSTER PUBLISHED

Ghatak, D. and Sakurai, K. and Roy, B.K., Can Data Obfuscation techniques be beneficial for preserving Data Utility unlike Differentially Private Algorithms? *IWSEC 2020.*

OTHER PAPER

Ghatak D., Sengupta D and Roy B. Optimal Gamma density to Obfuscate Quantitative data with Added Noise (To be submitted)

PAPER REVIEWED   Penalized Likelihood Approach for the Four-parameter Kappa Distribution *Journal of Applied Statistics*

ACADEMIC HISTORY :

COURSE	INSTITUTION	DURATION	MARKS OBTAINED (%age)
M.SC in Statistics	University Of Calcutta	July 2010-July 2012	78.75
B.SC in Statistics	Presidency College	July 2007 - July 2010	60.375
Higher Secondary	WBCHSE	2005 - 2007	75.6
Secondary	WBSE	1993 - 2005	82.8

OTHER EXPERIENCE   Worked as a **Project Linked Personnel** at Indian Statistical Institute Kolkata under the supervision of Prof. Ayanendranath Basu (July, 2012 to July, 2013)

Worked as a **Teaching Assistant** with Prof. Bimal K. Roy at Indian Statistical Institute Kolkata (Subject: Statistics for the course M.S.Q.E.)

Worked as a **Teaching Assistant** with Prof. Bimal K. Roy at Indian Statistical Institute Kolkata (Subject: Statistics for the course M. Tech in Cryptology.)

Worked as a **Supervisor for the summer internship** project of Ms. Priyal Garg at Indian Institute of Technlogy Jammu, Project Title "Q-Q plots for Noisy data-sets".

OTHER AWARD   Qualified the National Eligiblity Test (NET), 2012.

TALKS DELIVERED   1. STAT QUEST, March 2017, Organised by Department of Statistics, University of Calcutta (Under UGC sponsored DSA Phase II Programme).

2. One Week Online Workshop on Statistics and Machine Learning in Practice, July 2020, Organised by Department of Statistics with IQAC, Brahmananda Keshab Chandra College.

WORKSHOPS/
SEMINARS/
CONFERENCES
ATTENDED

1. One day seminar on Applied Statistics, March 2013, Organized by Bayesian and Interdisciplinary Research Unit, ISI Kolkata.

2. Three day workshop on emerging areas in Applied Statistics, March 2013, Organized by Bayesian and Interdisciplinary Research Unit, ISI Kolkata.

3. The Eighth Lectures in Probability and Stochastic Processes, December 2013, at ISI Bangalore.

4. Indo-Russian joint conference in Statistics and Probability, January 2015, organized by ISI/NBHM held at ISI, Delhi Centre.

5. Calcutta University Triennial, 2015, organized by Calcutta University, held at Ballygunge Science College Kolkata.

6. INDOCRYPT 2016, held at ISI Kolkata.

COMPUTER
PROFICIENCY

Comfortable in languages like **C programming**, **R programming**, MSword, MSexcel, MSpower-

point, **Latex**. Prefers to work with **R programming** and **Latex**.

LANGUAGES KNOWN English, Hindi, Bengali.

M.Sc PROJECT Worked on an econometric project entitled *A study of the economic growth in U.S.A.* (a joint work with Soumita Modak).

THESIS ABSTRACT Privacy protection and data security have received a huge amount of attention these days due to the increasing need to protect various sensitive information like credit card data, medical data etc. Statistical data is very important to make inferences in most studies but this often calls for threats to individual privacy. In order to take care of this problem, agencies release masked version of the true data-sets for statistical studies. However, this masking process results in loss of information to the statistician. The problem of data obfuscation deals with the challenge of masking the data-set in a way such that the obfuscated data may ensure individual privacy but is also useful in making statistical inferences. The data obfuscation problem is mainly of two types, (i) the identity disclosure problem, and (ii) the data disclosure problem. In this dissertation, we discuss both the problems and how one can take care of them. Mathematically, both the problems first need a measure that ensures privacy protection to the individuals and an ideal model is the one that gives the best estimation of certain statistics provided a certain logical criterion related to this measure is satisfied. First, we discuss a few existing methods of obfuscating both categorical and numerical data for dealing with the data disclosure problem, and the estimation of some commonly used statistics from the obfuscated data. Then, we share a few contributions we made in this field. We discuss the additive noise model for obfuscating numerical data and propose a parameter selection criterion for the error distribution that is ideal for quantile estimation. We apply the proposed method to both simulated and real-life data sets and see how it works. We also show why Laplace Noise is preferred over some other common distributions for estimation of quantiles. Then, we introduce a new model for obfuscating numerical data that gives an unbiased estimation of the true distribution function and performs better than Laplace noise addition in many cases. We also justify the results by simulations. We extend our work on numerical data by proposing a method to find from an additive noise model, an optimal noise for estimation of the distribution function of the true data, within the two-sided Gamma family. Finally, we discuss the identity disclosure problem, an existing model that ensures obfuscation and also how we believe the security bound to this existing model can be improved.