Rudradev Sengupta

Curriculum Vitae

Contact Information

Address: E101, Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Center for Statistics, Hasselt University, 3590 Diepenbeek, Belgium.

Email: rudradev.sengupta@uhasselt.be
rudradevsengupta@gmail.com
Tel: +32-478013794
DOB: 06.06.1990

Research Interests

- **Statistical Bioinformatics**: Microbiome data analysis, Data mining and machine learning techniques, Multi-source data integration, Exploration and analysis of high-dimensional datasets, Identification of molecular biomarkers, Clustering and biclustering analysis, Time-series analysis.
- **Biostatistics**: Multivariate data analysis, Longitudinal and clustered data analysis, Bayesian data analysis, Nonlinear modeling, Predictive modeling, Survival data analysis.
- **High Performance Computing**: Parallel computing, Distributed computing, Cloud computing.

Education

Hasselt University (UHasselt), Martelarenlaan 42, 3500 Hasselt, Belgium
- *Ph.D. in Statistics* (to be graduated: June 2018)

Indian Statistical Institute (ISI), Kolkata, West Bengal, India
- *M.Stat* (graduated with Distinction: June 2013)
- *B.Stat (Hons.)* (graduated: June 2011)

Current Position

Ph.D. Student in I-BioStat **June, 2014 to Present**

- **Research Lines**:
  - **Integrated Data Analysis in Drug Discovery:**
    - Model the association between genes and bioactivity data, taking into account the effect of different chemical structures.
    - Using Bayesian Variable Selection to model the monotone relationship between dose and response.
    - High dimensional surrogacy in drug development: single, multiple and partial surrogacy.
    - Different approaches to model dose-response data.
  - **Microbiome Research:**
    - Parametric joint model for microbiome and immunological data, taking care of the treatment effect.
    - Using non-parametric models to take care of zero-inflation in microbiome data.
    - Development of high dimensional microbiome biomarkers for an immune response: hierarchical bayesian approach.
    - Non-linear models and longitudinal analysis of microbiome data.
    - Modeling interaction between OTUs / Families.
  - **High Performance Computing**
    - Using VSC cluster in Leuven (Belgium) for faster implementation of different statistical models which helped to do further analysis by reducing the total computational time significantly.
    - Use publicly available resources (e.g. Amazon cloud, Azure etc.) for statistical analysis.

Previous Position

Software Engineer/Analyst in Polycom **August, 2013 to May, 2014**
Role: Software development (using Java) and data analysis (demand forecasting, using R and SQL).

Awards

By Govt. of India
- INSPIRE Scholar, 2008-2011.
- INSPIRE Scholar, 2011-2013.

By Hasselt University Doctoral School
- International and intersectoral mobility grant award, 2016.

Publications and Ongoing Research

Journal Articles:

- IsoGeneGUI: Multiple Approaches for Dose-Response Analysis of Microarray Data Using R (2017)
  Otava, M., Sengupta, R.[1], *Shkedy, Z., Lin, D., Pramana, S., Verbeke, T., Haldermans, P., Hothorn, L.A., Gerhard, D., Kuiper, R.M., Klinglmueller, F. and Kasim, A.*
- Surrogacy and Biomarker detection in Big Data Environment: An Integrated Analysis of Chemical Structure, Bioactivity and Genetic Information in Drug Discovery (*To be submitted*)
  Sengupta, R., Shkedy, Z. and Biecek, P.
- High Dimensional Surrogacy in Drug Development: single, multiple and partial surrogacy - a joint modeling approach (*To be submitted*)
  Sengupta, R. and Shkedy, Z.
- Development of High Dimensional Microbiome Biomarkers for an Immune Response: Hierarchical Bayesian Approach (*To be submitted*)
  Sengupta, R., Shkedy, Z., Bijnens, L., Ruiz, V.E., Battaglia, T. and Blaser, M.
- Development of High-dimensional Microbiome Biomarkers in Small-sample Studies (*To be submitted*)
  *P*erualila-Tan, N.J., Sengupta, R., Shkedy, Z., Bijnens, L., Ruiz, V.E., Battaglia, T. and Blaser, M.
- Some observations on HC-128 (2011)
  Maitra, S., Paul, G., Raizada, S., Sen, S. and Sengupta, R.

Book Chapters:

Applied Biclustering Methods for Big and High Dimensional Data Using R (2016)

- Identification of Local Patterns in the NBA Performance Indicators
  Shkedy, Z., Sengupta, R. and Perualila, N.J.
- Biclustering for Cloud Computing
  Sengupta, R., Trelles, O., Tirado, O.T. and Shkedy, Z.
- The biclustGUI Shiny App
  De Troyer, E., Sengupta, R., Otava, M., Zhang, J.D., Kaiser, S., Culhane, A., Gusenleitner, D., Gestraud, P., Csardi, G., Hochreiter, S., Klambauer, G., Clevert, D.A., Perualila, N.J., Kasim, A. and Shkedy, Z.

Applied Surrogate Endpoint Evaluation Methods with SAS and R (2016)

- High Dimensional Biomarkers in Drug Discovery: The QSTAR Framework
  Perualila, N.J., Shkedy, Z., Sengupta, R., Bigirumurame, T., Bijnens, L., Talloen, W., Verbist, B., Göhlmann, H.W.H., QSTAR Consortium and Kasim, A.

Other Research and Professional Internships

Research Visit - Institute of Bioinformatics, Johannes Kepler University
Linz, Austria. **14 - 27 June, 2015**

- Working with Prof. Oswaldo Trelles' group about using R in Amazon EC2 instance as well as in EMR cluster with EC2 instances as nodes.
- Automated the cluster creation in Amazon cloud to run R from the browser - no need to have R installed in the laptop.
- Parallel programing using the computer cluster: faster implementation using the Amazon EMR cluster as compared to the implementation in Laptop and later more things were investigated.

Research Visit - Blaser Lab Group, Langone Medical Center, New York School of Medicine
New York, USA. **22 - 24 February, 2016**

- Working with Prof. Martin Blaser's group on microbiome data analysis.
- Ongoing collaboration for future publications.

Research Visit - Unilever Center for Molecular Informatics, Department of Chemistry, University of Cambridge
Cambridge, UK. **3 - 7 October, 2016**

- Working with Prof. Andreas Bender's group about analysing gene-expression data and dose-response analysis of L1000 data.
- Ongoing collaboration for future publications.

Master's Thesis - Indian Statistical Institute
Kolkata, India. **April, 2013 - June, 2013**

- Working with Prof. Mandar Mitra's group on evaluation methods for information retrieval systems.
- Suggested new statistical approaches to estimate the number of relevant documents corresponding to a particular query.

[1] Joint First Author

Internship - Mzaya Pvt. Ltd.
Bangalore, India. **May, 2012 - July, 2012**

- Day-ahead exchange price forecasting and solving the Unit Commitment problem.
- Both *Time Series models* and *Neural Network approach* were adopted to perform day-ahead exchange-price forecasting of power and specifically, the Neural Network based model, produced much better results than general regression models.
- The electrical unit commitment problem is the problem of optimally deciding which electricity generation units should be running in each period to satisfy a predictably varying demand for electricity. *Genetic Algorithm* was used to deal with this optimization problem.

Internship - GE Global Research
Bangalore, India. **May, 2011 - July, 2011**

- Classification of data on change of impedance collected via RFID sensors, using Functional Data Analysis.
- We managed to find a concrete mathematical function so that better classification, as compared to Principal Component Analysis, can be done, solely based on its parameters.

Software Development

Software Development with R:

- R Packages: *IntegratedJM*, *ORCME*, *IsoGene*, *IsoGeneGUI*.
- *Biclustering AMI* in Amazon Cloud.
- Several *ShinyR web applications* in Amazon Cloud.

Teaching Experience

Hasselt University:

- **Full Course:** Business Statistics for Master of Management; 2016-2017 and 2017-2018.
- **Lessons:** Computer Intensive Methods, Master of Biostatistics/ Master of Bioinformatics/ Master of Statistics: Epidemiology & Public Health Methodology; 2015-2016 and 2016-2017.
- **Co-promoter:** Aimé Lambert Uwimana, Masters thesis, Master of Statistics, 2016-2017 - *Microbiome Data Analysis: Development of high-dimensional microbiome biomarkers.*
- **Supervision:** Pieter Giesen, Project for Analysis of Gene Expression course, Master of Statistics, 2015-2016 - *Analysis of microbiomics data.*

Presentations and Conferences

Conferences:

- Model Based Clustering using Bayesian Variable Selection: Modeling and Computational Issues
  PRACE Winter School, Ostrava, Czech Republic; January 12-15, 2015
- Development of Microbiome Biomarkers for IgA: a dynamic joint modeling approach
  The First Workshop on Challenges in Microbiome Data Analysis - Simons Foundation, New York, USA; February 24-25, 2016
- Model Based Classification of Monotone Gene Profiles using Bayesian Variable Selection
  Bayes, Leuven, Belgium; May 17-20, 2016
- Evaluation of Surrogate Endpoint in High Dimensional Data: A Longitudinal Analysis of Family-level Microbial Diversity and Immunoglobulin A
  Non-Clinical Statistics Conference, Cambridge, UK; October 4-6, 2016
- Development of High Dimensional Biomarkers
  IBS Channel Network Conference, Hasselt, Belgium; April 24-26, 2017
- IntegratedJM - an R package to Jointly Model the Gene-Expression and Bioassay Data, Taking Care of the Fingerprint Feature effect
  useR Conference, Brussels, Belgium; July 4-7, 2017
- Development of Microbiome Biomarkers for Immunoglobulin A (IgA): A Joint Modeling Approach
  6th Exploring Human Host-Microbiome Interactions in Health and Disease Meeting, Hinxton, Cambridge, UK; September 13-15, 2017

Other Presentations:

- Cloud Computing and Parallel Programming in Statistical Bioinformatics, using R: Different Case Studies
  ExaScience Project Meeting - IMEC, Belgium; 16.09.2015

Key Skills

Statistical Modeling and Implementation:

- Proficient with mathematical and statistical models as well as computational aspects of modeling.
- Programming languages: Matlab, R, Python, C, C++, Java, SPSS and SAS.
- Prompt in picking up a new computer language or software.
- Good logical reasoning skills.
- Good team player.

	Affiliation	Collaborators
Collaborations	Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Belgium	Ziv Shkedy, Ewoud de Troyer.
	Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge, United Kingdom	Andreas Bender, Avid Afzal.
	Institute of Bioinformatics, Johannes Kepler University, Linz, Austria	Sepp Hochreiter, Günter Klambauer, Djork-Arné Clevert.
	Department of Mathematical Modeling, Statistics and Bioinformatics, Ghent University,Belgium	Olivier Thas, Stijn Hawinkel.
	Wolfson Research Institute, Durham University, United Kingdom	Adetayo Kasim.
	Janssen Pharmaceutical Companies of Johnson & Johnson, Beerse, Belgium	Luc Bijnens, Martin Otava, Nolen Joy Perualila-Tan, Bie Verbist, Willem Talloen, Hinrich W.H. Göhlmann,
	Computer Architecture Department, University of Malaga, Spain	Oswaldo Trelles.
	Advanced Computing Technologies Unit, RISC Software GmBH, Hagenberg, Austria	Oscar Torreno Tirado.
	School of Medicine, Pharmacy and Health and Research Design Service (RDS), Durham University, United Kingdom	Theophile Bigirumurame.
	Department of Statistics - Faculty of Mathematics, Informatics and Statistics, Ludwig-Maximilians, Munich University, Germany	Sebastian Kaiser.
	Institute of Biostatistics, Leibniz University Hannover, Germany	Ludwig A. Hothorn.
	GlaxoSmithKline, Belgium	Dan Lin.
	Sekolah Tinggi Ilmu Statistik/Institute of Statistics, Indonesia and Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden	Setia Pramana.
	Open Analytics NV, Belgium	Tobias Verbeke.
	PXL-IT, PXL University College, Hasselt, Belgium	Philippe Haldermans.
	School of Mathematics and Statistics, University of Canterbury, New Zealand	Daniel Gerhard.
	Department of Methodology & Statistics, Utrecht University, The Netherlands	Rebecca M. Kuiper.
	Section for Medical Statistics, Center for Medical Statistics and Informatics, Medical University of Vienna, Austria	Florian Klinglmueller.
	Pharmaceutical Sciences, Translational Technologies and Bioinformatics (PS-TTB), Roche Pharmaceutical Research and Early Development (pRED), Roche Innovation Center, Basel, Switzerland	Jitao David Zhang.
	Computational Biology and Functional Genomics Laboratory, Harvard School of Public Health, Dana-Farber Cancer Institute, USA	Aedin Culhane.
	Bioinformatics Program, Boston University, USA	Daniel Gusenleitner.
	Institut Curie, INSERM U900 and Mines ParisTech, Fontainebleau and Paris, France	Pierre Gestraud.
	Department of Statistics, Harvard University, USA	Gabor Csardi.
	Applied Statistics Unit, Indian Statistical Institute, India	Subhamoy Maitra, Goutam Paul.
	Microsoft Research and The Department of Mathematics, Massachusetts Institute of Technology, USA	Subhabrata Sen.