SURAJIT RAY

Contact Details:

Department of Mathematics and Statistics	☎: +1-617-353-5209
111 Cummington Street, Rm 222	🖷: +1-617-353-8100
Boston University, Boston, MA 02215	✉: sray@math.bu.edu

Homepage: http://math.bu.edu/people/sray/

Professional Experience

2006–	Assistant Professor, Dept. of Mathematics and Statistics, Boston University, Boston.
2005–2006	Visiting Assistant Professor, Dept. of Biostatistics, University of North Carolina at Chapel Hill., Chapel Hill
2004–2005	Post Doctoral Fellow, Statistics and Applied Mathematical Sciences Institute, Research Triangle Park, Durham
2003–2004	Visiting Assistant Professor, Dept. of Biostatistics University of North Carolina at Chapel Hill.
2000–2003	Research Assistant, Dept. of Statistics, Pennsylvania State University, University Park

Education

2003 **Ph.D.** Dept of Statistics, Pennsylvania State University

Dissertation: "Distance-based Model-Selection with application to Analysis of Gene Expression Data"

Advisor: Bruce G. Lindsay.

Dates attended: Aug 1999- Aug 2003

1999 **M. Stat.**, Indian Statistical Institute, Calcutta, India.

– First Division with distinction;

– Specialization: Applied Statistics and Data Analysis.

Dates attended: Aug 1997- May 1999

1997 **B. Sc.** (Honors) in Statistics, Presidency College, Calcutta

– First Division with distinction;

– Minors: Mathematics, Economics.

Dates attended: Aug 1994- May 1997

Honors and Awards

2007	Honored by the Class of 2007 through the Class of 2007 Gift Program at Boston University.
2004	Laha Travel Award from the Institute of Mathematical Statistics
2003	"Most Outstanding Student Presentation" in Theoretical Statistics at the International Conference on Statistics, Combinatorics and Related Areas. (See *Presentations* below)
2001-2003	Several graduate student travel awards from the Eberly College of Science, PennState.
2002	Davey Graduate fellowship award from the Eberly College of Science, PennState.
2002	August and Ruth Homeyer Graduate fellowship award from the Eberly College of Science, PennState.
2000	Vollmer-Kleckner Scholarship award in Science from the Eberly College of Science, PennState, for the most outstanding performance in PhD Qualifiers.

Research Interests

- ❒ Theory and applications of finite mixture models, and detection of modes in high dimensional data, modal clustering.
- ❒ Assessment of model fit in high dimensional data and nonlinear space.
- ❒ Statistical methodology for social sciences focusing on structural equation models.
- ❒ Medical Imaging- Segmentation and characterization of anatomical objects in high dimensions and non-linear manifolds.
- ❒ Bioinformatics- focusing on classification of epitopes and model based clustering of microarray gene-expression data, with applications to "epitope-based" vaccine development.

Teaching Experience

Spr 2008	MA 576 (Generalized Linear Models) Dept. of Mathematics and Statistics, BU.
	MA 584 (Multivariate Statistical Analysis) Dept. of Mathematics and Statistics.
Fall 2007	MA 881 (Topics in High Dimensional Data Analysis) Dept. of Mathematics and Statistics, BU.
Spr 2007	MA 576 (Generalized Linear Models) Dept. of Mathematics and Statistics, BU.
Fall 2006	MA 586 (Design of Experiments) Dept. of Mathematics and Statistics, BU.
Spr 2006	BIOS 110 (Principles of Statistical Inference) Dept. of Biostatistics, UNC.
Fall 2005	BIOS 110 (Principles of Statistical Inference) Dept. of Biostatistics, UNC.
Sum 2005	Taught 3 Modules in SAMSI/CRSC Undergraduate Workshop, CRSC, North Carolina State University.
Spr 2004	BIOS 145 (Principles of Experimental Analysis), Dept. of Biostatistics, UNC.
Fall 2001	STAT 401 (Experimental Methods), Pennsylvania State University.

Publications

- ❒ Hong Huang Lin, **Ray, S.**, Songsak Tongchusak, Ellis L. Reinherz, Vladimir Brusic (2008) ▪ Evaluation of MHC class I peptide binding prediction servers: applications for vaccine research. *BMC Immunology*, **9**:8
- ❒ Lindsay, B.G., Markatou M., **Ray, S.**, Yang, K., Chen, S.C. (2008) ▪ Quadratic distances on probabilities: the foundations. *The Annals of Statistics* Vol. 36, No. 2, page 983–1006
- ❒ **Ray, S.**, Lindsay, B.(2008). ▪ Model selection in High-Dimensions: A Quadratic-risk Based Approach. *Journal of the Royal Statistical Society - Series B* Volume **70** Issue 1 (Feb), 95–118.
- ❒ **Ray, S.**, Tom Kepler (2007). ▪ Amino acid biophysical properties in the statistical prediction of peptide-MHC class I binding. *Immunome Research* Oct 29;3(1):9
- ❒ Li, J., **Ray, S.**, Bruce G Lindsay. (2007) ▪ A Nonparametric Statistical Approach to Clustering via Mode Identification *Journal of Machine Learning Research* 8(Aug):1687–1723.
- ❒ Levy, J.H, Broadhurst R.R., **Ray, S.**, Chaney, E.L., Pizer, S.M.(2007) ▪ Signaling local non-credibility in an automatic segmentation pipeline *Proceedings of the International Society for Optical Engineering meetings on Medical Imaging,* Volume 6512
- ❒ Jeong, J., Pizer, S.M., **Ray, S.** (2006)▪ Statistics on Anatomic Objects Reflecting Inter-Object Relations. *Proceedings of International Workshop on Mathematical Foundations of Computational Anatomy.*
- ❒ **Ray, S.**, Lindsay, B.(2005). The Topography of Multivariate Normal Mixtures. *The Annals of Statistics* **33**, 5, 2042–2065.
- ❒ M. Gupta, **Ray, S.** (2005). Sequence pattern discovery with applications to understanding gene regulation and vaccine design. *Handbook of Statistics* Ed. Chakraborty, R. and Rao, C.R. *Elsevier Press* [in press]
- ❒ **Ray, S.**, Lindsay, B.,(2005) Selecting the Number of Components in a Finite Mixture: A Risk-Based Approach. Proceedings of the of the 37th Symposium on the Interface, *Computing Science and Statistics* **37**.
- ❒ Basu, A., **Ray, S.**, Park, C., Basu, S. (2002) Improved Power in Multinomial Goodness-of-fit Tests, *Journal of the Royal Statistical Society Series D*, **51**, 3, 381–393.
- ❒ **Ray, S.** (2003) Distance-based Model-Selection with application to the Analysis of Gene Expression Data. Electronic Thesis.
  `http://etda.libraries.psu.edu/theses/approved/WorldWideIndex/ETD-375/`

Submitted Manuscripts

- ❒ Jeong, J,, **Ray, S.**, Han, Q., Lu X, Muller K E. and Pizer, S. M., (2008) ▪ Goodness of Prediction for Principal Components of Shape: A Correlation Measure ▪ Submitted to *International Journal of Computer Vision*

Working Papers

- ❒ **Ray, S.** ▪ Model-based bi-clustering using two-way mixtures.
- ❒ **Ray, S.**, Marron, J.S. ▪ Feature selection based on high dimensional low sample size geometry.
- ❒ **Ray, S.**, Lindsay, B.G., Li, J. ▪ Modal EM for Mixtures and its Application in Clustering.
- ❒ **Ray, S.,** Yeong, J., Pizer, S., Muller, K., Han Q. ▪ Sample size advantages of statistics on a nonlinear manifold to characterize nonlinear variation in a population.
- ❒ Lindsay, B.G., Markatou M., **Ray, S.**. ▪ Degrees of Freedom in Quadratic Goodness of Fit.
- ❒ Berger, J.O., **Ray, S.**, Visser, I, Bayarri, M.J., Jang, W. ▪ Generalization of BIC.
- ❒ Bollen, K.A., **Ray, S.**, Zavisca, J. ▪ A Scaled unit information prior approximation to the Bayes Factor.

Published software

The following softwares will be shortly available through CRAN (`http://cran.r-project.org`). For current information about the packages and downloads visit `http://math.bu.edu/people/sray/software/`

- ❒ QUADRISK: `C++` binary for calculating quadratic risk of a mixture fit and providing graphical aid to high-dimensional model selection problems.
- ❒ MODALITY: `R-package` for finding the number of modes of a multivariate normal and providing graphical and analytical representation of high-dimensional manifolds.
- ❒ MHCPROP: `R-package` for MHC binder prediction based on biophysiochemical properties of amino acids.

Skills

- ❒ Statistics: Mixture models, Model selection in high dimensions, Asymptotics of high dimensional low sample size, Bioinformatics, Immunoinformatics, Medical Image Analysis.
- ❒ Programming Languages: C/C++, Perl, Python, Java, PASCAL, CSS-HTML.
- ❒ Computing Platforms: Unix (Linux, Sun Solaris) , DOS/Windows, Mac OS-X.
- ❒ Statistical Software: Extensive experience with R/SPlus; SAS, Matlab, Mathematica, SPSS.

Recent Invited Presentations

- ❒ Data Mining and Knowledge Discovery of Land Cover and Terrestrial Ecosystem Processes from Global Remote Sensing Data *NASA conference on Intelligent Data Understanding: Presented by Mark Friedl* ▪ September 9-10, 2008
- ❒ Modal Inference and Its Application to High-Dimensional Clustering *Session on Mixture Models: A Tool for Multilayered Clustering and Dimension Reduction at the Joint Statistical Meetings* ▪ August 3-7, 2008.
- ❒ A tool for multi-layered clustering and dimension reduction. *International Conference on*

Statistical Paradigms - Recent Advances AND Reconciliations (ICSPRAR-2008), Indian Statistical Institute, Kolkata ▪ January 1-4, 2008.

- ❒ Modal Inference and Its Application to High-Dimensional Clustering. *Department of Biostatistics, University of Minnesota,* ▪ October 31, 2007.
- ❒ An Extended BIC for Model Selection. *Joint Statistical Meetings, Salt Lake City* ▪ August 1, 2007.
- ❒ Modal Inference: Building the bridge between nonparametric clustering and mixture analyses. *WNAR and IMS Meetings, Irvine* ▪ June 26, 2007.
- ❒ Modal inference and its application to high-dimensional clustering, *Department of Statistics, Harvard University, Cambridge* ▪ April 30, 2007.
- ❒ Modal Inference: Building the bridge between nonparametric clustering and mixture analyses. *Department of Electrical and Computer Engineering, Boston University* ▪ March 21, 2007.
- ❒ Modal inference and its application to high-dimensional clustering, *Department of Biotatistics, Harvard University, Boston* ▪ Feb 21, 2007.
- ❒ Hierarchical Modal Clustering based on the Topography of Multivariate Mixtures, *International Conference on Multivariate Statistical Methods, Kolkata, India* ▪ Dec 29, 2006.
- ❒ Modal inference and its application to high-dimensional clustering, *Department Statistics, University of Connecticut, Storrs* ▪ Nov 9, 2006.
- ❒ Modal EM for Mixtures and its Application in Clustering, *Department of Mathematics and Statistics, Boston University, Boston* ▪ Sep 28, 2006.
- ❒ Hierarchical Modal Clustering based on the Topography of Multivariate Mixtures, *Department of Mathematics and Statistics, Boston University, Boston* ▪ Mar 22, 2006.
- ❒ Hierarchical Modal Clustering based on the Topography of Multivariate Mixtures, *Department of Mathematics and Statistics, McGill University, Montreal, Canada* ▪ Feb 28, 2006.
- ❒ Hierarchical Modal Clustering based on the Topography of Multivariate Mixtures, *Department of Biostatistics, University of North Carolina, Chapel Hill* ▪ Feb 24, 2006.
- ❒ Hierarchical Modal Clustering based on the Topography of Multivariate Mixtures, *Department of Statistics, Yale University, New haven* ▪ Feb 13, 2006.
- ❒ Model Selection in High-Dimensions: A Quadratic risk-based approach, *Department of Probability and Statistics, National University of Singapore, Singapore* ▪ Feb 3, 2006
- ❒ The topography of multivariate mixtures and Modal clusters, *Department of Mathematics and Statistics, University of Bristol, UK*, ▪ Jan 11, 2006
- ❒ Quadratic Distance:The basis for building High-dimensional model selection tool, *Department of Statistics, University of Glasgow, UK* ▪ Dec 12, 2005
- ❒ Effective sample size and the Bayes factor, *Transition Workshop: Latent Variable Models in the Social Sciences, SAMSI*▪ Nov 11, 2005
- ❒ Model Selection in High-Dimensions: A Quadratic risk-based approach, *Department of Statistics, University of California, Davis* ▪ Oct 6, 2005

❒ Using Quadratic Risk to select models in High dimensions, *Department of Statistics, London School of Economics, London, UK* ▪ Sep 27, 2005

❒ Classification of MHC-I binding epitopes, *WNAR/IMS Annual Meeting, Fairbanks, Alaska* ▪ June 21-24, 2005

❒ On using Quadratic Risk to Select High dimensional Mixture Model, *Annual Meeting of the Statistical Society of Canada* ▪ June 12-15, 2005

❒ Selecting the Number of Components in a Finite Mixture: A Risk-based Approach, *Joint Annual Meeting of the Interface and the Classification Society of North America: Theme: Clustering and Classification Washington University School of Medicine, St. Louis, Missouri* ▪ June 8-12, 2005

❒ Bayes Factors in Structural Equation Models: Schwarz's BIC and Other Approximations, *American Sociological Association Section on Methodology: 2005 Annual Meeting, Chapel Hill* ▪ Apr 22, 2005

❒ Selecting the Number of Components in a Finite Mixture: A Risk-based Approach, *International Conference on the future of statistical theory, practice and education, Hyderabad, India* ▪ Dec 29-Jan 1, 2004,

❒ The Topography of Multivariate Normal Mixtures, *Seventh North American New Research Conference* Toronto, Canada ▪ Aug 4-7, 2004

❒ Distance-based Model-selection in Mixture Distributions, *International Conference on:Statistics in Health Sciences* , Nantes, France ▪ June 23-25, 2004.

Professional Activities

❒ Organizer and chairperson of sessions in scientific meetings.
- Joint Statistical Meetings, Denver, 2008
- WNAR/IMS Annual Meeting, Irvine,2007.
- International Conference on Multivariate Statistical Methods, Kolkata, India, 2006
- WNAR/IMS Annual Meeting, Fairbanks,2005.
- Joint Annual Meeting of the Interface and the Classification Society of North America, 2005,
- WNAR/IMS Annual Meeting, Irvine, 2007.

❒ Organizer of NSF sponsored Undergraduate workshop in statistics, North Carolina State University, 2005

❒ Reviewer of several peer reviewed journal articles.
- The Annals of Statistics
- Journal of American Statistical Association
- Journal of Royal Statistical Society (Series B)
- Multivariate Analysis
- Statistical Methodology
- Australian and New Zealand Journal of Statistics.

❒ Designed and taught distance learning courses at University of North Carolina, Chapel Hill.

Professional Memberships

2001-present	American Statistical Association (ASA)
2001-present	Institute of Mathematical Statistics (IMS)
2004-present	Mathematical Association of America (MAA)

Student Mentoring

- ❒ Burton Shank (Ph. D., Biology, Boston University) Thesis: *Spatial Variation in Coral Reef community*
  - ▪ Role: Committee member.
- ❒ Ja-Yeon Jeong (Ph.D., Computer Science, University of North Carolina at Chapel Hill) Thesis:*Estimation of probability distribution on multiple anatomical object complex*
  - ▪ Role: Committee member.
- ❒ Joshua Stough (Ph.D., Computer Science, University of North Carolina at Chapel Hill) Thesis:*Object-Relative Tissue Mixture Models for Deformable Model Segmentation*
  - ▪ Role: Committee member.

Research Funding

- ❒ NASA Carbon Cycle & Ecosystem Grant: MODIS Algorithm Refinement and Earth Science Data Record Development for Global Land Cover and Land Cover Dynamics. NNX08AE61A (P.I. Mark Friedl) 06/01/2008- Current ▪ Role: Statistician.
- ❒ NIH Program Project Grant: Medical Image Presentation, (P.I: Pizer, S.M.) 09/15/2005-06/31/2007 ▪ Role: Consultant.