

# GOURAB MUKHERJEE

---

## EMPLOYMENT:

- Aug 2013 -**                      **Marshall School of Business, University of Southern California.**  
Assistant Professor of Data Sciences and Operations.
- Sept 2008 - July 2013**   **Statistics Department, Stanford University.**  
Teaching and Research Assistant.

## EDUCATION:

- 2008 - 2013 :**    **Ph.D. in Statistics**, Stanford University.  
*Dissertation Title: Sparsity and shrinkage in predictive density estimation.*  
*Advisor: Iain M. Johnstone.*
- 2006 - 2008 :**    **Master in Statistics**, Indian Statistical Institute, Kolkata.  
*Specialization: Mathematical Statistics and Probability.*
- 2003 - 2006 :**    **Bachelor in Statistics**, Indian Statistical Institute, Kolkata.

## CONTACT INFORMATION:

3670 Trousdale Parkway, Bridge Hall 300 B,  
University of Southern California, Los Angeles, California 90089-0809.

- **Tel:** (213) 740-6612    • **Email:** gourab@usc.edu
- **Webpage:** <http://www-bcf.usc.edu/~gourab/>

## RESEARCH INTERESTS:

### A) STATISTICS THEORY AND METHODS

- Predictive Inference, Empirical Bayes and Shrinkage Methods, Inference under Heterogeneity;
- Predictive Density Estimation, Sparse Estimation and Penalized Regression methods;

### B) APPLICATIONS AND STATISTICAL METHODS

- Quantitative Marketing: Mixed-effects models for pricing and promotions strategies;
- Statistical methods in Inventory planning: Quantile and Asymmetric losses;
- Virology and Infectious Diseases: Clustering, Non-parametric two-sample tests;

## AWARDS AND HONORS:

- Institute for Outlier Research in Business, Role of Digital Platforms in Increasing Access to Affordable Medication, joint with Shantanu Dutta and Dinesh Puranam; Amount: \$ 20000 for 2021-2022.
- NSF-DMS-1811866 Grant: Optimal Shrinkage and Empirical Bayes Prediction under Asymmetry, Censoring and Nonexchangeability. *Single PI, Amount: \$127,443*; 1st June 2018 - 31st Dec 2021, NSF Division of Mathematical Sciences Statistics program;
- USC Zumberge Individual Award, 2016. Project Title: High-dimensional Prediction under Asymmetric Check Loss; Total Amount Awarded: \$ 23,500.
- Institute of Mathematical Statistics Laha Travel Award, 2012.
- NSF travel award to present at the 8<sup>th</sup> World Congress on Probability and Statistics in Istanbul, 2012.
- Centennial Teaching Assistant Award from the school of Humanities and Sciences, Stanford University, 2011.
- M. R. Iyer memorial medal for outstanding performance in Bachelor of Statistics (rank 1 in the program) at the Indian Statistical Institute, 2006.

**PUBLISHED:** (Google Scholar Citations as of June 2021: 424 )

1. Mukherjee G, and Johnstone IM. *On Minimax Optimality of Sparse Bayes Predictive Density Estimates*, **Annals of Statistics** (forthcoming), (2021).
2. Karmakar B, Liu P, Mukherjee G, Che H, and Dutta S. *Improved Retention Analysis in Freemium Role-Playing Games by Jointly Modeling Players' Motivations, Progressions and Churns*, **Journal of Royal Statistical Society, Series A** (forthcoming), (2021).
3. George E, Mukherjee G, and Yano K. *Optimal Shrinkage Estimation of Predictive Densities under  $\alpha$ -divergences*. **Bayesian Analysis**, (2021) Advance Publication 1 - 17.
4. Gangopadhyay U, and Mukherjee G. *On Discrete Priors and Sparse Minimax Optimal Predictive Densities*, **Electronic Journal of Statistics** (2021), Vol. 15, No. 1, 1636-1660.
5. Banerjee T, Liu Q, Mukherjee G, and Sun W. A General Framework for Empirical Bayes Estimation in the Discrete Linear Exponential Family. **Journal of Machine Learning Research**, (2021) 22(67):146.

6. Gang B, Mukherjee G, and Sun W. *Large-Scale Shrinkage Estimation under Markovian Dependence*. **Book chapter for the IISA Series on Statistics and Data Science**, (2021), arXiv:2003.01873 [stat.ME] .
7. Sen N, **Mukherjee G**, and Arvin A. *The Use of Single Cell Mass Cytometry to Define the Molecular Mechanisms of Varicella-Zoster Virus Lymphotropism*. **Frontiers in Microbiology**, (2020), Vol 11, Article 1224.
8. Banerjee T, **Mukherjee G**, and Sun W. *Adaptive Sparse Estimation with Side Information*. **Journal of American Statistical Association**, (2020), Vol. 115, No. 532, Pages 2053-2067.
9. Banerjee T, Bhattacharya B, and **Mukherjee G**. *A nearest-neighbor based nonparametric test for viral remodeling in heterogeneous single-cell proteomic data*. **Annals of Applied Statistics**, (2020) 14(4): 1777-1805.
10. Ma T, Luo X, George A , **Mukherjee G**, Sen N, Spitzer T, Giudice L, Greene WC, and Roan N. *HIV efficiently infects T cells from the endometrium and remodels them to promote systemic viral spread*. **Elife** 9 (2020): e55487.
11. Banerjee T, **Mukherjee G**, Dutta S, and Ghosh P. *A Large-scale Constrained Joint Modeling Approach For Predicting User Activity, Engagement And Churn With Application To Freemium Mobile Games*. **Journal of American Statistical Association**, (2020), Vol. 115, No. 530, Pages 538-554.
12. George EI, Marchand E, **Mukherjee G**, and Paul D. *New and Evolving Roles of Shrinkage in Large-Scale Prediction and Inference*. **BIRS Workshop Report**, Pages 1-14, (2019).
13. Banerjee T and **Mukherjee G**. *Discussion of CARS: covariate assisted ranking and screening for large-scale two-sample inference by Cai, Sun and Wang*. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)** (2019).
14. Brown LB, **Mukherjee G**, and Weinstein A. *Empirical Bayes Estimates for a 2-Way Cross-Classified Additive Model*. **Annals of Statistics**, Vol. 46, No. 4 (2018), Pages 1693-1720.
15. Banerjee T, **Mukherjee G**, and Radchenko P. *Feature screening in Large Scale Cluster Analysis*, **Journal of Multivariate Analysis**, (2017), Vol 161, Pages 191-212.
16. Cavois M, Banerjee T, **Mukherjee G**, Raman N, Hussien R, Rodriguez BA, Vasquez J, Spitzer MH, Lazarus NH, Jones JJ, Ochsenbauer C, McCune JM, Butcher EC, Arvin AM, Sen N, Greene WC, and Roan NR. *Mass cytometric analysis of HIV entry, replication, and remodeling in tissue CD4+ T cells*. **Cell Reports**, (2017), Volume 20 , Issue 4 , 984-998.

17. Radchenko P and **Mukherjee G**. *Convex clustering via  $L_1$  fusion penalization*. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, (2017) Vol. 79, No. 5, Pages 1527-1546.
18. **Mukherjee G**. Book Review of: *Robust Methods for Data Reduction* by A. Farcomeni and L. Greco in **Journal of the American Statistical Association**, 917, Vol 111, (2016).
19. Sen N, **Mukherjee G**, and Arvin AM. *Single Cell Mass Cytometry Reveals Remodeling of Human T Cell Phenotypes by Varicella Zoster Virus*. **Methods**, (2015), Volume 90, Pages 85-94.
20. **Mukherjee G**, and Johnstone IM. *Exact minimax estimation of the predictive density in sparse Gaussian models*. **Annals of Statistics**, (2015), Vol. 43, No. 3, Pages 937-961.
21. Sen N, **Mukherjee G**, Sen A, Bendall S, Sung P, Nolan G, and Arvin AM. *Single Cell Analysis of the Remodeling of Human Tonsil T Cells by Varicella-Zoster Virus*. **Cell Reports**, (2014), Volume 8, Issue 2, 633-645.
22. Sen A, Rott L, Phan N, **Mukherjee G**, and Greenberg HB. *The rotavirus NSP1 protein inhibits IFN-mediated STAT1 activation*. **Journal of Virology**, (2014), 88.1 : 41-53.
23. Gustin SL, **Mukherjee G**, Baker VL, Westphal LM, Milki AA, and Lathi RB. *Early pregnancy testosterone after ovarian stimulation and pregnancy outcome*. **Fertility and Sterility**, (2012), Volume 97, Issue 1.
24. Sen A, Rothenberg ME, **Mukherjee G**, Feng N, Nair N, Kalisky T, Johnstone IM, Clarke MF, and Greenberg HB. *Innate immune response to homologous rotavirus infection in the small intestinal villous epithelium at single-cell resolution*. **Proceedings of the National Academy of Sciences (PNAS)** 109.50 (2012): 20667-20672.
25. Gustin SL, **Mukherjee G**, Baker VL, Westphal LM, Milki AA, Ma C, and Lathi, RB. *First Trimester Testosterone After Ovarian Stimulation and Its Effect on Pregnancy Outcomes*. **Fertility and Sterility**, (2011), Volume 95, Issue 4, .

## **SOFTWARE:**

- **emarketing**: a github repository of R codes for jointly modeling the conversion funnel
- **casp**: an R package for shrinkage prediction under spiked covariance structures
- **cezij**: a matlab toolbox for high-dimensional joint modeling with a time-to-event variable
- **asus**: an R package for sparse estimation with side information
- **fusionclust**: an R package for clustering with fusion penalty

- **slide**: an R package for testing viral remodeling by single cell linkage distance estimation

## **PRESENTATIONS:**

### CONFERENCE PRESENTATIONS:

1. Informs Annual Meeting, Anaheim, Oct 2021.
2. IISA 2021 Conference: Statistics in the Era of Evidence Based Inference, May 2021.
3. Conference on Artificial Intelligence In Management, May 2021.
4. ENAR 2021 Spring Meeting, March 2021.
5. CM Statistics, Virtual Meeting, Dec 2020.
6. Conference on Artificial Intelligence, Machine Learning, and Business Analytics, Dec 2020.
7. Informs Annual Meeting, Virtual Conference, Nov 2020.
8. CM Statistics, London, Dec 2019.
9. IISA Conference, Mumbai, India, Dec 2019.
10. Informs Annual Meeting, Seattle, September 2019.
11. 3rd International Conference on Econometrics and Statistics: EcoSta, Taiwan, June 2019.
12. Workshop in memory of Professor Lawrence D. Brown, Philadelphia, Dec 2018.
13. Conference on Digital Marketing and Machine Learning, CMU, Pittsburg, Dec 2018.
14. 5th IMS-APRM, Singapore, June 2018.
15. 2nd International Conference on Econometrics and Statistics: EcoSta, June 2018.
16. Statistical Learning and Data Science conference, New York City, June 2018.
17. IISA Conference, Gainesville, Florida, May 2018.
18. IISA Conference on Statistics and Data Science, Hyderabad, India, Dec 2017.
19. Joint Statistical Meeting, Baltimore, USA, August 2017.
20. INFORMS Applied Probability Society Conference, Evanston, IL, USA, July 2017.
21. IMS-China International Conference on Statistics and Probability, Nanning, July 2017.
22. Ist International Conference on Econometrics and Statistics: EcoSta, June 2017.
23. Mini-Conference on Big Data in Business and Data Sciences II, Feb 2017.
24. 9<sup>th</sup> International Conference of the ERCIM WG on Computational and Methodological Statistics, Seville, Dec 2016.

25. Joint Statistical Meeting, Chicago, USA, August 2016.
26. ICSA Conference on Data Science, Dali, China, July 2016.
27. 4th IMS-APRM, Hong Kong, China, June 2016.
28. 2015 International Indian Statistical Association Conference, Pune, India; Dec 2015.
29. Third International Conference on Business Analytics and Intelligence, IIM Bangalore, 2015.
30. 43rd Annual Meeting of the Statistical Society of Canada, Halifax, Canada, 2015.
31. ICORS 2015, Kolkata, India, Jan 2015.
32. 3rd IMS APRM, Taipei, Taiwan, June 2014.
33. IISA Conference on Research Innovations in Statistics, Riverside, USA, 2014.
34. Joint Statistical Meeting, Boston, USA, 2014.
35. Young Researchers Meet, Stanford University, 2012.
36. 8<sup>th</sup> World Congress in Probability and Statistics, Istanbul, 2012.
37. NSF Workshop for High-Dimensional Data, Yale University (poster presentation), 2012.
38. Joint Statistical Meeting, Miami, USA, 2011.

INVITED SEMINAR TALKS:

39. Ohio State University, Oct 2020.
40. George Mason University, Oct 2020.
41. University of Minnesota, Sept 2020.
42. University of Cambridge, May 2020.
43. University of Tokyo, July 2019.
44. University of Illinois, Urbana-Champaign, Nov 2017.
45. University of California, Los Angeles, May 2017.
46. USC Marshall School of Business, April 2017.
47. University of California, Los Angeles, April 2016.
48. University of California, Davis, March 2016.
49. University of California, Riverside, January 2014.
50. Wharton School of the University of Pennsylvania, February 2013.
51. Harvard University, February 2013.

52. University of Chicago, February 2013.
53. USC Marshall School of Business, February 2013.
54. University of Florida, February 2013.
55. Yale University, January 2013.
56. University of North Carolina at Chapel Hill, January 2013.
57. Carnegie Mellon University, January 2013.
58. Applied Statistics Unit, Indian Statistical Institute, December 2011.
59. Mahalanobis International Symposium on Statistics, Indian Statistical Institute, June 2008.

## TEACHING

- APPLIED BUSINESS STATISTICS (BUAD310g)
  - *introductory class for undergraduate business students; the format changed over years;*

Year	#. of weekly classes	#. of Students
2021	2 lectures, 4 applied sections	132
2020	2 lectures, 4 applied sections	139
2018	2 lectures, 4 applied sections	133
2016	2 lectures, 4 applied sections	117
2015	2 lectures, 6 applied sections	196
2014	8 applied sections	258
2013	4 lectures	141

- REGRESSION AND GENERALIZED LINEAR MODELS FOR BUSINESS APPLICATIONS (GSBA604)
  - *taught in fall 2018, fall 2019 and fall 2020; graduate class with an approx class size of 10.*

## SERVICE TO PROFESSION:

### CONFERENCE & WORKSHOP ORGANIZER:

1. Co-organizer of the workshop on *New and Evolving Roles of Shrinkage in Large-Scale Prediction and Inference* held in the Banff International Research Station from Apr 7- 12, 2019. The workshop had 26 talks and was attended by 36 participants.
2. Scientific Program Committee member of 14th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2021), London, December 2021.

### ORGANIZER OF CONFERENCE SESSIONS:

3. *Inference under heterogeneity*, CMStatistics, London, Dec 2021.
4. *Statistical Methods for Contemporary Business Applications*, Informs Annual Meeting, Anaheim, Nov 2021.
5. *Statistical Methods for Contemporary Applications*, IISA Conference: Statistics in the Era of Evidence Based Inference, May 2021.
6. *Recent Advances in Estimation theory*, IISA Conference: Statistics in the Era of Evidence Based Inference, May 2021.
7. *Recent Advances in Statistical Learning*, International Conference on Econometrics and Statistics (EcoSta), Seoul, South Korea, July 2021.
8. *Statistical Methods for Large-Scale Data Analysis*, Informs Annual Meeting, Maryland, Nov 2020.
9. *Large Scale Inference and Applications*: a session with 4 speakers, IISA 2019, Mumbai, Dec 2019.
10. *Statistical methods for new-age inference problems*: a session with 4 speakers, CMStatistics, London, Dec 2019.
11. *Large-scale statistical inference for contemporary business applications*: a session with 4 speakers, Informs Annual Meeting 2019, Seattle, October 2019.
12. *Contemporary inference issues in big data problems*: a session with 4 speakers, EcoSta 2019, Taiwan, June, 2018.
13. *Recent Advances in Large-scale Inference*: a session with 4 speakers, EcoSta 2018, Hong Kong, June, 2018.
14. *Statistical Methods for large-scale data analysis*: a session with 3 speakers, IISA Conference, Gainesville, Florida, May, 2018.
15. *Large-scale Inference and Applications*: a session with 3 speakers, IISA Conference on Statistics and Data Science for better life, society and science, Hyderabad, India, December 2017.
16. *Recent Advances in Predictive Inference*: a session with 5 speakers, Joint Statistical Meeting, Baltimore, USA, August 2017.(Topic Contributed Session; Sponsored by IMS)
17. *Recent Advances in Empirical Bayes Methods*: a session with 4 speakers, Joint Statistical Meeting, Chicago, USA, August 2016. (Invited Session; Primary sponsor: IISA; Co-sponsored: IMS)
18. *Recent Advances in High Dimensional Estimation Theory*: a session with 4 speakers, IMS-APRM, Hong Kong, China, June 2016.



## EDITORIAL AND REFEREE WORK:

- ☐ Associate Editor of Sankhya Series A which covers Mathematical Statistics and Probability.
- ☐ SLDS Student Paper Award Committee
- ☐ Refereed multiple papers in the following journals:
  - Annals of Statistics • Journal of the American Statistical Association • Biometrika • Journal of Econometrics • IEEE Transactions of Information Theory • Bernoulli • Electronic Journal of Statistics • Australian and New Zealand Journal of Statistics • Journal of Quantitative Analysis of Sports • Nature Communications • Statistica Sinica • Journal of Statistical Planning and Inference • Computational Statistics and Data Analysis • Journal of Multivariate Analysis • Biometrics • Big and Complex Data Analysis: Statistical Methodologies and Applications • Journal of Statistical Theory and Applications • Mathematical Reviews • Statistics • Test • Journal of Machine Learning Research • Journal of Royal Statistical Society, Series B

## PHD STUDENT ADVISING:

*includes students for whom I was a dissertation committee member or was significantly involved in a least one of the papers during their phd studies;*

### PRINCIPAL ADVISOR:

1. Trambak Banerjee, Data Sciences and Operations, University of Southern California. Graduated in 2020; Starting Position: University of Kansas, School of Business.
2. Rashmi Ranjan Bhuyian, Data Sciences and Operations, University of Southern California. Currently 1st year Phd student.

### CO-AUTHOR OR COMMITTEE MEMBER:

3. Abdullah Karakus, Department of Mathematics, University of Southern California; Currently 4th year Phd student.
4. Jiajun Luo, Department of Mathematics, University of Southern California. Graduated in 2021; Starting Position: Machine Learning researcher at DiDi, Mountain View, CA.
5. Ujan Gangopadhyay, Department of Mathematics, University of Southern California. Graduated in 2021; Starting Position: Post doc. at National University of Singapore, Department of Mathematics.
6. Brian Finley, Department of Economics, University of Southern California. Graduated in 2021; Starting Position: US Census Bureau.
7. Bowen Gang, Department of Mathematics, University of Southern California. Graduated in 2020; Starting Position: Fudan University, School of Business.

8. Luella Fu, Data Sciences and Operations, University of Southern California Graduated in 2018; Starting Position: San Francisco State University, Department of Mathematics.
9. Weinan Wang, Data Sciences and Operations, University of Southern California Graduated in 2018; Starting Position: Snapchat.
10. Sungjune Pyun, Finance and Business Economics, University of Southern California Graduated in 2017; Starting Position: National University of Singapore, School of Business.