Irina Gaynanova

Assistant Professor, Department of Statistics, Texas A&M University

irinag@stat.tamu.edu

irinagain.github.io

February 20, 2020

Research interests

Statistical learning, high-dimensional data, multivariate analysis, classification, data integration, computational statistics, machine learning.

Education

- Ph.D, Statistics **05/2015**
  Cornell University, Ithaca, NY
  Advisors: James Booth and Martin Wells
- M.S., Statistics **05/2013**
  Cornell University, Ithaca, NY
- Diploma with honors (M.S.), Applied Mathematics and Computer Science **06/2009**
  Lomonosov Moscow State University, Moscow, Russia

Professional Positions

- *Assistant Professor* **since 07/2015**
  Department of Statistics, Texas A&M University
- *Senior Specialist* **06/2009 - 07/2010**
  Balancing Market Division, OJSC Trading System Administrator, Moscow, Russia
- *Junior Statistician* **06/2008 - 05/2009**
  Census Division, AC Nielsen, Moscow, Russia

Funding

- NSF DMS-1712943 **Gaynanova I. (PI)** **07/2017 - 06/2020**
  Scalable methods for classification of high-dimensional heterogeneous data
  Awarded amount: $162,539
- Johns Hopkins University, Subcontract **Gaynanova I. (PI)** **07/2019 - 06/2020**
  Statistical Analysis of CGM Data
  Awarded amount: $30,063
- TAMU Institute of Data Science Postdoctoral Project Program **Ni Y. (PI)** **07/2020 - 06/2022**
  Studying Microbial Interactions and Host Heterogeneity via Data Integration
  Role: Co-PI
  Awarded amount: 50% of postdoctoral researcher support
- NSF CCF-1934904 **Mallick B. (PI)** **10/2019 - 09/2022**
  HDR Tripods: Texas A&M Research Institute for Interdisciplinary Foundations of Data Science
  Role: Co-Investigator

Awards and Honors

- David P. Byar Young Investigator Award, ASA Biometrics Section   **2018**
- Cornelia Ye Outstanding Teaching Assistant Award, Cornell University   **2014**
- EducationUSA Opportunity Award   **2009**
- Study Abroad Scholarship, Technical University of Munich, Germany   **2007**

Competitive Travel Awards

- IMS New Researchers Conference   **2017**
- NISS Writing Workshop for Junior Researchers   **2016**
- SAMSI LDHD Summer School Program   **2013**
- Building Future Faculty Program, NC State University   **2013**
- Diversity and Mentoring Program at JSM   **2012**

Publications and Submitted Manuscripts

Note: the sign * at the beginning of a paper indicates alphabetical order of authorships; mentored student and post-doctoral associate co-authors are underlined; the sign ⊠ denotes corresponding author.

Submitted Manuscripts:

18. Zhang Y.⊠ and **Gaynanova, I.** (2020+) "Joint association and classification analysis of multi-view data." Preprint available on arXiv:1811.08511 [stat.ML]

17. **Gaynanova, I.**⊠, Punjabi, N. and Crainiceanu, C. (2020+) "Modeling continuous glucose monitoring (CGM) data during sleep."

16. Lapanowski A.⊠ and **Gaynanova, I.** (2020+) "Compressing large-sample data for discriminant analysis."

Peer-reviewed Publications:

15. Yoon, G., Carroll, R. and **Gaynanova, I.**⊠ (2020+) "Sparse semiparametric canonical correlation analysis for data of mixed types." *Biometrika,* accepted.

14. **Gaynanova, I.**⊠ (2020) "Prediction and estimation consistency of sparse multi-class penalized optimal scoring." *Bernoulli,* Vol. 26, No. 1, 286-322.

13. **Gaynanova, I.**⊠ and Li, G. (2019). "Structural learning and integrative decomposition of multi-view data." *Biometrics,* Vol. 75, No. 4, 1121-1132.

12. * Bien, J., **Gaynanova, I.**, Müller, C. and Lederer, J.⊠(2019). "Prediction error bounds for linear regression with the TREX." *TEST,* Vol. 28, No. 2, 451-474.

11. Yoon G., **Gaynanova, I.** and Müller, C.⊠(2019) "Microbial networks in SPRING - Semi-parametric rank-based correlation and partial correlation estimation for quantitative microbiome data." *Frontiers in Genetics,* Vol. 10, 516.

10. Lapanowski A.⊠ and **Gaynanova, I.** (2019) "Sparse feature selection in kernel discriminant analysis via optimal scoring." *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS) ,* PMLR 89, 1704-1713.

9. Lederer, J.[✉], Lu Y. and **Gaynanova, I.** (2019).   "Oracle inequalities for high-dimensional prediction." *Bernoulli,* Vol. 25, No. 2, 1225-1255.

8. **Gaynanova, I.**[✉] and Wang, T. (2019) "Sparse quadratic classification rules via linear dimension reduction." *Journal of Multivariate Analysis,* Vol. 169, 278-299.

7. Li, G.[✉]and **Gaynanova, I.** (2018). "A general framework for association analysis of heterogeneous data." *Annals of Applied Statistics,* Vol. 12, No. 3, 1700-1726.

6. Hokamp J.[✉], Leidy S., **Gaynanova, I.**, Cianciolo R. and Nabity M. (2018) "Correlation of electrophoretic urine protein banding patterns with severity of renal damage in dogs with proteinuric chronic kidney disease." *Veterinary Clinical Pathology,* Vol. 47, No. 3, 424-434.

5. **Gaynanova, I.**[✉], Urbanek J. and Punjabi N. (2018). "Letter to the Editor: Corrections of equations on glycemic variability and quality of glycemic control." *Diabetes Technology & Therapeutics,* Vol. 20, No. 4, 317.

4. * Bien, J., **Gaynanova, I.**[✉], Müller, C. and Lederer, J. (2018). "Non-convex global minimization and false discovery rate control for the TREX." *Journal of Computational and Graphical Statistics,* Vol. 27, No. 1, 23-33.

3. **Gaynanova, I.**[✉], Booth, J. and Wells, M. (2017). "Penalized versus constrained generalized eigenvalue problems." *Journal of Computational and Graphical Statistics,* Vol. 26, No. 2, 379-387.

2. **Gaynanova, I.**[✉], Booth, J. and Wells, M. (2016). "Simultaneous sparse estimation of canonical vectors in the $p \gg n$ setting." *Journal of the American Statistical Association,* Vol. 111, No. 514, 696-706.

1. **Gaynanova, I.**[✉] and Kolar, M. (2015). "Optimal variable selection in multi-group sparse discriminant analysis." *Electronic Journal of Statistics,* Vol. 9, No. 2, 2007-2034.

Unrefereed Manuscripts:

2. **Gaynanova, I.** (2015). "Estimation of sparse low-dimensional linear projections." Ph.D. Thesis, Cornell University.

1. **Gaynanova, I.**, Booth, J. and Wells, M. (2013).   "Supervised classification using sparse Fisher's LDA." Technical report, arXiv:1301.4976 [stat.ML].

Presentations

Invited Presentations:

42. Department of Biostatistics, University of Washington, May 2020

41. Department of Statistics, University of Washington, May 2020

40. Department of Biostatistics, Emory University, April 2020

39. Banff International Research Station, workshop on the "Use of Wearable and Implantable Devices in Health Research", Banff, Canada, February 2020

38. Department of Statistics, Oregon State University, February 2020

37. The Gatsby Computational Neuroscience Unit, University College London, UK, December 2019

36. The Fourth Workshop on Higher-Order Asymptotics and Post-Selection Inference (WHOA-PSI), St. Louis, MO, August 2019

35. Topic-Contributed session, Joint Statistical Meetings, Denver, CO, July 2019

34. ICSA Applied Statistics Symposium, Raleigh, NC, June 2019

33. Econometrics and Statistics Colloquium, Booth School of Business, University of Chicago, May 2019

32. Department of Biostatistics, University of Michigan, April 2019

31. Big Data Working group, College of Veterinary Medicine, Texas A&M University, March 2019

30. Friday Science Seminar, College of Science, Texas A&M University, February 2019

29. Department of Biostatistics, University of Minnesota, February 2019

28. 11th International Conference of the ERCIM WG on Computational and Methodological Statistics (CM Statistics 2018), Pisa, Italy, December 2018

27. Bioinformatics and Cancer Symposium, Texas A&M University, September 2018

26. Topic-Contributed session, Joint Statistical Meetings, Vancouver, Canada, August 2018

25. IMS Asia Pacific Rim Meeting 2018, Singapore, June 2018

24. ICSA Applied Statistics Symposium, New Brunswick, NJ, June 2018

23. Statistical Learning and Data Science Conference, New York, NY, June 2018

22. Department of Statistics, Indiana University, April 2018

21. SAMSI Operator Splitting Workshop, Research Triangle Park, NC, March 2018

20. Department of Biostatistics, UT MD Anderson Cancer Center, March 2018

19. 10th International Conference of the ERCIM WG on Computational and Methodological Statistics (CM Statistics 2017), London, UK, December 2017

18. Department of Statistical Science, Baylor University, November 2017

17. Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, September 2017

16. Data-Driven Model Reduction Workshop, Texas A&M University, April 2017

15. Conference of Texas Statisticians (COTS), Dallas, TX, March 2017

14. ENAR, Washington DC, March 2017

13. 9th International Conference of the ERCIM WG on Computational and Methodological Statistics (CM Statistics 2016), Seville, Spain, December 2016

12. Bio-seminar, Department of Electrical and Computer Engineering, Texas A&M University, October 2016

11. Topic-Contributed session, Joint Statistical Meetings, Chicago, IL, August 2016

10. Southern Regional Council on Statistics Summer Research Conference, Bentonville, AR, June 2016

9. Structured Multivariate Data Workshop, Texas A&M University, January 2016

8. Department of Statistics, Rice University, February 2015

7. Department of Statistical Science, Cornell University, February 2015

6. Department of Statistics, Indiana University, January 2015

5. Department of Statistics, Texas A&M University, January 2015

4. Department of Mathematical and Statistical Sciences, University of Colorado, Denver, January 2015

3. Department of Statistics and Actuarial Science, University of Iowa, January 2015

2. Department of Statistics, University of California, Davis, January 2015

1. Department of Biostatistics, University of Iowa, December 2014

Contributed Presentations:

12. Poster, ENAR, Philadelphia, PA, March 2019

11. IMS New Researchers Conference, Baltimore, MD, August 2017

10. Poster, International Conference on Machine Learning in New York City, NY, June 2016

9. Contributed session, Joint Statistical Meetings in Boston, MA, August 2014

8. Statistics Student Seminar, Cornell University, April 2014

7. Poster, SAMSI LDHD Workshop, February 2014

6. Contributed session, Joint Statistical Meetings in Montreal, Canada, August 2013

5. Poster, SAMSI LDHD Summer School, August 2013

4. Statistics Student Seminar, Cornell University, March 2013

3. Contributed session, Joint Statistical Meetings in San Diego, CA, August 2012

2. Cross-Campus Collaborative Colloquium, Cornell University, December 2011

1. Statistics Student Seminar, Cornell University, September 2011

Software Developed

Mentored student and post-doctoral associate co-authors are underlined.

7. SPRING: an R package for estimation of sparse microbial association networks using rank-based correlation, available from Github.
   Authors: Yoon G., **Gaynanova I.**, Müller, C.

6. sparseKOS: an R package for nonlinear binary classification using sparse kernel optimal scoring, available from Github.
   Authors: Lapanowski A., **Gaynanova I.**

5. JACA: an R package for joint association and classification analysis of multi-view data, available from Github.
   Authors: Zhang Y., **Gaynanova I.**

4. mixedCCA: an R package for semiparametric sparse canonical correlation analysis for data of mixed types (continuous/ binary/ zero-inflated), available from Github.
   Authors: Yoon G., **Gaynanova I.**

3. DAP: an R package to perform discriminant analysis via projections, available from Github and CRAN.
   Authors: Wang T., **Gaynanova I.**

2. MGSDA: an R package to perform sparse multi-group discriminant analysis, available from CRAN.
   Authors: **Gaynanova I.**

1. TREX: a Matlab package to perform sparse linear regression using TREX, available from github.
   Authors: Müller C., Bien J., **Gaynanova I.**, Combettes P.

Teaching Experience

Primary instructor (at Texas A&M University):

STAT 211: Principles of Statistics I   **Fall 2015/2016/2017, Spring 2016**
    Calculus-based introduction to probability and probability distributions; sampling and descriptive measures; inference and hypothesis testing; analysis of variance; linear regression.

STAT 610: Distribution Theory   **Fall 2017/2018/2019**
    Graduate-level introduction to probability theory; distributions and expectations of random variables, transformations of random variables and order statistics; generating functions and basic limit concepts.

STAT 689: Statistical Learning with Sparsity   **Spring 2017**
    Graduate-level class covering penalized empirical loss minimization methods with sparsity-inducing penalties. The course also includes brief introduction to convex optimization and duality.

STAT 600: Computational Statistics   **Spring 2018/Fall 2019**
    Graduate-level course on computational statistics and optimization. Topics include version control with Git and Github, code vectorization and profiling, writing R packages, introduction to convex optimization and optimization algorithms.

Table 1: Summary of teaching evaluations at Texas A&M University, Q2: On the whole, this was a good instructor. Scale 1(Strongly Disagree) to 5(Strongly Agree).

Course	Semester	Total class size	Mean (median) response
STAT 211	Fall 2015	77	4.68 (5.00)
STAT 211	Spring 2016	77	4.21 (4.00)
STAT 211	Fall 2016	97	4.57 (5.00)
STAT 211	Fall 2017	97	4.44 (5.00)
STAT 610	Fall 2017	35	4.82 (5.00)
STAT 610	Fall 2018	34	5.00 (5.00)
STAT 610	Fall 2019	39	4.54 (5.00)
STAT 600	Spring 2019	7	4.84 (5.00)
STAT 600	Fall 2019	31	4.73 (5.00)
STAT 689	Spring 2017	12	5.00 (5.00)

Primary instructor (at Cornell University):

ILRST 2100: Introductory Statistics   **Winter 2013, 2014 ,2015**
    Non-calculus-based introduction to statistics; sampling and descriptive measures; inference and hypothesis testing; analysis of variance; linear regression.

Teaching Assistant (at Cornell University):

ILRST 2100: Introductory Statistics   **Summer 2013, 2014**

Non-calculus-based introduction to statistics; sampling and descriptive measures; inference and hypothesis testing; analysis of variance; linear regression.

ENGRD 2700: Basic Engineering Probability and Statistics (recitation leader)   **Fall 2013**
Calculus-based introduction to probability and probability distributions; sampling and descriptive measures; inference and hypothesis testing; analysis of variance; linear regression.

MATH 4720: Statistics   **Spring 2015**
Introduction to mathematical statistics based in calculus, linear algebra, and probability theory; inference for proportions, inference for means, contingency tables and linear regression.

BTRY 6010: Statistical Methods I (TA/Lab instructor)   **Fall 2009, 2010, 2014**
Graduate level introduction to statistical methods for analyzing data; descriptive statistics and data visualization; analysis of variance; linear regression.

BTRY 6020: Statistical Methods II (TA/Lab instructor)   **Spring 2010, 2011, 2012**
Continuation of BTRY 6010. Emphasizes the use of multiple regression analysis, experimental design and generalized linear models.

Mentoring

Postdoctoral Researchers

1. Grace Yoon   **since 08/2017**
   Postdoctoral Trainee of DHHS-NIH National Cancer Institute T32, Texas A&M University, Department of Statistics

PhD Committee Chair (Statistics)

1. Dongbang Yuan   **2022 (expected)**
2. Alex Lapanowski   **2020 (expected)**
3. Yunfeng Zhang   **2020 (expected)**
4. Tianying Wang (co-chair with Raymond Carroll)   **2018**
   Position upon graduation: Postdoctoral Research Scientist in the Department of Biostatistics of the Mailman School of Public Health at Columbia University.

PhD Committee Member (Statistics)

1. Krystin Pantoja (chair David Jones)   **2022 (expected)**
2. Jiayi Wang (chair Raymond Wong)   **2022 (expected)**
3. Allyson Souris (chair Anirban Bhattacharya)   **2020 (expected)**
4. Sangyoon Yi (chair Xianyang Zhang)   **2020 (expected)**

PhD Committee Member (Other majors)

1. Nida Obatake (chair Anne Shiu, Major: Mathematics)   **2022 (expected)**

2. Jianling Wang (chair James Caverlee, Major: Computer Science and Engineering)   **2022 (expected)**

MS Committee Member (Statistics)

1. Yiwen Zhou   **2020**
2. Eric Riley   **2019**

MS Committee Member (Other majors)

1. Terin Thomas (Major: Industrial and Systems Engineering)   **2020**
2. Yinsong Wang (Major: Industrial and Systems Engineering)   **2020**
3. Aaron Knodell (Major: Computer Science)   **2019**
4. Haley Pichler (Major: Mathematics)   **2019**
5. Jung Sim Hyun (Major: Mathematics)   **2019**

Undergraduate Student Research

1. Steven Broll, STAT major   **06/2019 - present**
2. David Buchanan, STAT major   **01/2020 - present**
3. Mary Martin, STAT major   **01/2020 - present**
4. Elizabeth Chun, BMCB major   **01/2020 - present**
5. Sangaman Senthil, INEN major   **01/2020 - present**
6. Eric Wang, ENGE major   **01/2020 - present**

Other formal mentoring

1. Brittany Segundo   **01/2019 - present**
   Dr. Gaynanova is a teaching mentor through the Academy for Future Faculty (AFF) at Texas A&M University

Departmental and University Service

At Texas A&M University:

Committees:

- Faculty Evaluation and Promotion Guidelines Committee, Department of Statistics   **since 02/2020**
- Graduate Committee, Department of Statistics   **since 08/2019**
- Colloquium Chair, Department of Statistics   **06/2018 - 12/2019**
- Name Change Committee, Department of Statistics   **07/2019 - 08/2019**
- Hiring Committee, Department of Statistics   **07/2018 - 02/2019**
- Undergraduate Committee, Department of Statistics   **09/2017 - 08/2018**

Panelist:

- 3rd Annual TX-LA Undergraduate Mathematics conference   **10/2019**
- TAMU College of Science, Lunch and Learn series   **09/2019**
- TAMU Symposium for Faculty, Staff, Graduate Students and PostDocs in the Sciences **02/2017**
- TAMU AWM Chapter Panel "Career in Mathematics"   **03/2016**

Other:

- Friday Science Seminar, College of Science, Texas A&M University **02/2019**

At Cornell University:

- President, STATS graduate student organization **11/2013 - 05/2015**
- Organiser, Statistics Student Seminars **08/2011 - 12/2013**

Professional Activities

Editorial service:

- Associate Editor, *Journal of Computational and Graphical Statistics* **since 2018**

Journal refereeing:

- Bioinformatics
- Biometrics
- Biometrika
- Journal of the American Statistical Association
- Journal of Computational and Graphical Statistics
- Journal of Machine Learning Research
- Journal of Multivariate Analysis
- IEEE Transactions on Network Science and Engineering
- PLOS One
- Statistical Analysis and Data Mining
- Statistical Modeling: an International Journal
- Stat
- Statistics in Biosciences
- Statistics in Medicine
- Statistics & Probability Letters
- Statistica Sinica
- Technometrics

Book proposals revfereeing:

- Springer **2018**
- CRC/Chapman & Hall **2017, 2018**
- Wiley **2017**
- MacMillan Education **2015**

Grant review panels:

- Joint NSF/NIH panel **2016**

Elected Positions:

- President, Southeast Texas Chapter of the American Statistical Association **02/2019-01/2020**
- Vice-president, Southeast Texas Chapter of the American Statistical Association **10/2017-01/2019**

Professional Committees Chair:

- Co-chair (chair Genevera Allen) of ASA SLDS Student Paper Award Committee **01/2020**

Professional Committees Member:

- SLDS conference Student Paper Award Committee **03/2020**
- ASA SLDS JSM Poster Competition Committee **07/2019**
- ASA SLDS Student Paper Award Committee **01/2019**
- ASA SCS John M. Chambers Statistical Software Award Committee **01/2016**

Conference Program Committees:

- 13th International Conference of the ERCIM WG on Computational and Methodological Statistics (CM Statistics 2020) in London, UK, Scientific Program committee **2020**
- ENAR 2019 program committee, Statistical Learning and Data Mining section representative **2019**

Sessions Organized:

- Invited Session at CM Statistics in London, UK: "Statistical methods for time-varying multivariate data" **12/2019**
- Topic-Contributed Session at Joint Statistical Meetings: "Integrative approaches for statistical analysis of data from multiple sources" **08/2019**
- Invited Session at Joint Statistical Meetings: "Discovering homology in multi-view data: new statistical methods for data integration" **08/2018**
- Topic-Contributed Session at Joint Statistical Meetings: "Exploiting Low-Dimensional Structures: Recent Advances of Statistical Learning Methods in Genetics and Genomics" **08/2016**

Sessions Chaired:

- CM Statistics, London, UK **12/2019**
- Joint Statistical Meetings, Denver, CO **07/2019**
- ENAR, Philadelphia, PA **03/2019**
- CM Statistics, Pisa, Italy **12/2018**
- Joint Statistical Meetings, Vancouver, Canada **08/2018**
- IMS Asian Pacific Rim Conference, Singapore **06/2018**

Membership:

- American Statistical Association (ASA) **since 2011**
- Institute of Mathematical Statistics (IMS) **since 2012**
- ENAR **since 2015**

Other:

- Docent at JSM in Boston **08/2014**