Miaoyan Wang

Assistant Professor of Statistics
University of Wisconsin - Madison
1250B MSC, 1300 University Ave.
Wisconsin, WI 53703

Phone: 608-265-3990
Email: miaoyan.wang@wisc.edu

Research Interests

Mathematical foundations for machine learning, numerical analysis, higher-order tensors, high-dimensional statistics

Applications to population genetics, gene expression, and brain imaging analysis

Employment

Assistant Professor, Department of Statistics, UW-Madison, 2018/9 - current.

Affiliated faculty in Institute for Foundations of Data Science (IFDS), UW-Madison, 2019/1 - current.

Visiting fellow, Institute for Pure and Applied Mathematics (IPAM), UCLA (virtual), 2021/3 - 2021/6.

Education Training

Simons Math + X Postdoc at UC Berkeley (Computer Science) and University of Pennsylvania (Biology), 2016-2018. Supervisor: Yun S. Song

Ph.D in Statistics, The University of Chicago, 2010-2015.
Advisor: Mary Sara McPeek

B.S in Mathematics, Fudan University, 2006-2010.
2007 - 2010: switched to Math major (Rank: 1/172) after admitted into the China national program for talented students in fundamental sciences
2006 - 2007: majored in Computer Science (Rank: 1/105)

Selected Awards

UW-Madison nominee for Johnson and Johnson WiSTEM^{2}D scholar award in math discipline. Each University can nominate ONE candidate per discipline per year, 2019/2020 (twice).

Diversity and Inclusion Fellowship, International Conference of Machine Learning (ICML), 2020.

Institute of Mathematical Statistics (IMS) New Researcher Travel Award, 2020.

NeurIPS Junior Researcher Travel Award, 2019.

Google Cloud Platform Education Grant. Free computing credit to students in class. 2019-2020.

Vice Chancellor for Research and Graduate Education (VCRGE) Travel Fund, UW-Madison, 2018.

Simons Math+X Postdoc Fellowship. Simons Foundation, 2015-2017.

Runner-up for Department of Statistics Consulting Award (ranked as #2 among all PhD students in the departmental vote of 2014). Department of Statistics, The University of Chicago, 2014.

American Society of Human Genetics (ASHG) Charles J. Epstein Trainee Award for Excellence in Human Genetics Research – semifinalist (27 predoctoral recipients out of $\sim$ 550 candidates), 2014.

International Genetic Epidemiology Society (IGES) Williams Award for Best Platform Presentation by A Graduate Student – finalist (3 predoctoral recipients out of 156 candidates), 2013.

Paul Meier Fellowship, Department of Statistics, The University of Chicago. 2010 - 2013.

National Merit Scholarship $\times 3$ ($\sim$ top 1-2 per department per year), Ministry of Education of China. 2006 - 2007, 2007 - 2008, 2008 - 2009.

Honors Fellowship in Mathematics $\times$ 3, Fudan University, China. 2007 - 2008, 2008 - 2009, 2009 - 2010.

President's list, Top 10 role models out of $\sim 3,300$ freshmen, Fudan University, China, 2006 - 2007.

Publications

Authors underlined are students at UW-Madison.

Under review:

• C. Lee and **M. Wang**. Beyond the Signs: Nonparametric tensor completion via sign series. Under review by Neural Information Processing Systems (NeurIPS), (2021). arXiv:2102.00384.

• C. Lee, L. Li, H. Zhang, and **M. Wang**. Nonparametric trace regression in high dimensions via sign series representation. Under review by Annals of Statistics, (2021). arxiv: 2105.01783.

• R. Han, Y. Luo, **M. Wang**, and A. R. Zhang. Exact clustering in tensor block model: Statistical optimality and computational limit. Under review by Journal of the Royal Statistical Society Series B, (2020). arXiv:2012.09996.

This work won **2021 Best Student Paper Award** *from the Statistical Learning and Data Science Section of the American Statistical Association (ASA).*

Published:

• J. Hu, C. Lee and **M. Wang**. Supervised Tensor Decomposition with interactive side information. Journal of Computational and Graphical Statistics, in press, (2021).

This work won **2021 Best Student Paper Award** *from the Statistical Computing and Graphics Section of American Statistical Association (ASA).*

• C. Lee and **M. Wang**. Optimal tensor denoising and completion based on ordinal observation. International Conference on Machine Learning (ICML), Proceedings of Machine Learning Research (PMLR), 119:5778-5788, (2020).

• J. Hu, C. Lee, and **M. Wang**. Learning Multiple Networks via Supervised Tensor Decomposition. Neural Information Processing Systems 34 (NeurIPS), Workshop on Machine Learning and the Physical Sciences, (2020).

• **M. Wang** and L. Li. Learning from Binary Multiway Data: Probabilistic Tensor Decomposition and Its Statistical Optimality. Journal of Machine Learning Research, 21(154): 1-38, (2020).

• **M. Wang** and Y. Zeng. Multiway clustering via tensor block models. Neural Information Processing Systems 33 (NeurIPS), 715-725, (2019).

• **M. Wang**, J. Fischer, and Y. S. Song. Three-way Clustering of Multi-tissue Gene Expression Data Using Semi-Nonnegative Tensor Decomposition. The Annals of Applied Statistics, 13 (2), 1103-1127, (2019).

• **M. Wang,** F. Roux, C. Bartoli, C. H.-Chauveau, C. Meyer, H. Lee, D. Roby, M. S. McPeek, and J. Bergelson. Two-Way Mixed-Effects Methods for Joint Association Analyses Using Both Host and Pathogen Genomes. Proceedings of the National Academy of Sciences (PNAS direct submission), 115 (24), E5440-E5449, (2018).

Attention score in the top 5% of all research articles ever tracked by Altmetric; Higher than 89% of the research articles published in PNAS.

For this work, I was selected to present a **platform talk** *(21/156 submissions) at the 2nd Meeting for Probabilistic Modeling in Genomics, Cold Spring Harbor Laboratory, NY.*

• D. Jiang and **M. Wang**. Recent Developments in Statistical Methods for GWAS and High-throughput Sequencing Studies of Complex Traits. Biostatistics and Epidemiology, 2 (1), 132-159, (2018).

• **M. Wang** and Y. S. Song. Tensor Decomposition via Two-Mode Higher-Order SVD (HOSVD). Journal of Machine Learning Research W&CP (AISTATS track), 54, 614-622, (2017).

• **M. Wang,** K. Dao Duc, J. Fischer, and Y. S. Song. Operator Norm Inequalities Between Tensor Unfoldings on the Partition Lattice. Linear Algebra and its Applications, 520, 44-66, (2017).

• **M. Wang,** J. Jakobsdottir, A. V. Smith, and M. S. McPeek. G-STRATEGY: Optimal Selection of Individuals for Sequencing in Genetic Association Studies. Genetic Epidemiology, 40 (6), 446-60, (2016).

Highlighted as **Editor's Pick Paper** *of this issue*.

For part of this work, I was named a **semifinalist for the 2014 ASHG Charles J. Epstein Trainee Award** *(27 predoctoral recipients out of 550 candidates) for Excellence in Human Genetics Research at the Annual Meeting of American Society of Human Genetics; Also invited to give a platform talk (top 8%) in 2014 ASHG.*

For a different part of this work, I was named a **finalist for the 2013 IGES Williams Award** *(3 out of 156) for Best Platform Presentation by a Graduate Student at the Annual Meeting of the International Genetic Epidemiology Society; Also invited to give one of the six talks in Neels and Williams Awards Session in 2013 IGES.*

• B. W. Engelmann, Y. Kim, **M. Wang,** B. Peters, R. S. Rock, and P. D. Nash. The Development and Application of A Quantitative Peptide Microarray Platform to Protein Interaction Domain Specificity Space. Molecular and Cellular Proteomics, 13 (12), 3647-62, (2014).

Current Funding

As sole PI or co-PI:

National Science Foundation DMS-1915978, "Spectral methods for high-dimensional tensors", $180k, sole PI. 2019-2022.

National Science Foundation, NSF's 10 Big Ideas - Rule of Life: Emerging Network. "Does re-wilding leading to re-wiring of gene expression and species interaction network", $3 million, multi-PI (Daniel Bolnick, Tina Eliassi-Rad and Miaoyan Wang), 2022-2027.

Wisconsin Alumni Research Foundation, Fall Research Competition, "Tensor theory and methods in data science", $40k, sole PI. 2020-2021.

Wisconsin Alumni Research Foundation, Fall Research Competition, "Beyond matrices: statistical machine learning with higher-order tensors", $40k, sole PI. 2021-2022.

As co-I or senior personal:

National Science Foundation. TRIPODS: Collaborative: Institute for Foundations of Data Science. Role: senior personal (PI: Stephen J Wright from UW-Madison). 2020-2025.

U.S. Department of Defense, ARMY. Resolving the enigma of Factor Analysis. Role: co-I (PI: Karl Rohe from UW-Madison). 2020-2023.

Mentoring

PhD students

Chanwoo Lee (2019 -), Jiaxin Hu (2019 -).

Masters/Undergraduate students

Yuchen Zeng (2018 - 2020). On to CS PhD student.

Zhuoyan Xu (2018 - 2020). On to Statistics PhD student.

PhD thesis committee

Praewa Wongburi (Civil and Environmental Engineering), Luxi Cao (Statistics), Ruosi Guo (Statistics), Lili Zheng (Statistics), Rungang Han (Statistics)

Selected Invited Talks

Institute for Mathematical and Statistical Innovation (IMSI), University of Chicago, Invited talk, 08/2021.

Institute of Foundation of Data Science (IFDS), UW-Madison, 08/2021.

UC Berkeley, Laboratory for Information and System Science, 04/2021.

The Pennsylvania State University, Invited department talk, 02/2021.

Fred Hutchinson Cancer Research Center, Biostatistics department. 02/2021.

UW-Madison, Institute of Foundation of Data Science (IFDS), Brown Bag talk, 02/2021.

Neural Information Processing Systems 33 (NeurIPS) Workshop on Machine Learning and the Physical Sciences. 12/2020.

ICSA Applied Statistics Symposium. Invited conference talk, 12/2020.

Bernoulli-IMS One World Seminar. 08/2020.

International Conference on Machine Learning (ICML), 07/2020.

Eastern North American Region (ENAR). Invited conference talk, 04/2020.

Eastern North American Region (ENAR), International Biometric Society, 03/2020.

Columbia University, Department of Biostatistics, School of Public Health, 02/2020.

UW-Madison, Institute of Foundation of Data Science (IFDS), Brown Bag talk, 12/2019.

Fudan University, School of Data Science, Shanghai, China, 07/2019.

East China Normal University, Faculty of Economics and Management, Shanghai, China, 07/2019.

The University of Chicago, Department of Statistics, IL, 11/2018.

UW-Madison, Computation and Informatics in Biology and Medicine (CIBM) seminar, 11/2018.

UW-Madison, Systems, Information, Learning and Optimization (SILO) Seminar, 10/2018.

Stanford University, Department of Statistics, 08/2018.

UC Berkeley, Department of Biostatistics, 04/2018.

CMU, Department of Statistics and Data Science, 02/2018.

Columbia University, Department of Statistics, 02/2018.

University of Toronto, Department of Statistics, Canada, 02/2018.

UW-Madison, Department of Statistics, 01/2018.

Duke University, Department of Statistics, 01/2018.

Johns Hopkins University, Department of Biostatistics, 01/2018.

Queen's University, Department of Mathematics and Statistics, Canada, 01/2018.

University of Massachusetts Amherst, Department of Mathematics and Statistics,12/2017.

University of Pennsylvania, Song's Group, 09/2016.

Boston University, Department of Mathematics and Statistics, 08/2016.

International Conference on Frontiers of Data Science, Hangzhou, China, 05/2019.

European Society for Evolutionary Biology - Coevolution Workshop, Munich, Germany, 03/2019.

International Conference on Data Science, Shanghai, China, 12/2018.

Institute of Mathematical Statistics (IMS) China Meeting, Dalian, China, 07/2019.

Society for Industrial and Applied Mathematics (SIAM) Annual Meeting, Portland, OR, 08/2018.

Joint Statistical Meetings, Chicago, IL. 08/2016.

Platform talk (**top 13% of all submissions)**, the 2nd Meeting for Probabilistic Modeling in Genomics, Cold Spring Harbor Laboratory, NY. 10/2015.

Platform talk **(top 8% of all submissions)**, American Society of Human Genetics (ASHG) Annual Meeting, San Diego, CA. 11/2014.

Neel and Williams Awards Talk, International Genetic Epidemiology Society (IGES) Annual Meeting, Chicago, IL. 09/2013.

Bosch Center for Artificial Intelligence in North America, Sunnyvale, CA, 06/2018.

Global Analytic Group. Takeda Pharmaceuticals Inc., Deerfield, IL. 08/2014.

The 33rd Advances in Neural Information Processing Systems (NeurIPS), Vancouver, British Columbia, Canada, 12/2019.

Computation and Informatics in Biology and Medicine (CIBM) Training Program and the Bio-Data Science (BDS) Program, Madison, 11/2019.

2018 Berkeley Statistics Annual Research Symposium (BSTARS), Berkeley, CA, 03/2018.

The 20th International Conference on Artificial intelligence and Statistics (AISTATS), Florida, Fort Lauderdale, 04/2017.

The 7th Midwest Statistics Research Colloquium, Chicago, IL, 03/2014.

Teaching

Madison Teaching and Learning Excellence (MTLE) Fellow, 2019 -2020.

Lecturer, 2018 - current

Spring 2020. STAT 850: Theory and Application of Regression and Analysis of Variance-II (PhD core)

Fall 2019, 2020. STAT 849: Theory and Application of Regression and Analysis of Variance-I (PhD core)

Spring 2019. STAT 602: Statistical Methods–II (senior undergraduate)

Fall 2018, 2020. STAT 601: Statistical Methods–I (senior undergraduate)

Fall 2013. STAT 23400: Statistical Models and Methods (junior students in economics major)

Teaching Assistant at UChicago, 2011 - 2015

STAT 22000: Statistical Methods and Applications. Professor: Peter McCullagh

STAT 24400: Statistical Theory and Methods 1. Professor: Stephen Stigler and Debashis Mondal

STAT 24500: Statistical Theory and Methods 2. Professor: Debashis Mondal and Weibiao Wu

College Tutor at UChicago, Summer 2013: ECON 21000: Econometrics

Software

• **SignT**: Nonparametric Tensor Completion via Sign Series.
`https://cran.r-project.org/web/packages/tensorsign/index.html`

• **OrdinalT**: A set of R tools for noise reduction and completion from ordinal tensor data with possibly missing values.
`https://cran.r-project.org/web/packages/tensorordinal/index.html`

• **Tensor_regress:** R program for generalized tensor regression with covariates on multiple modes.
`https://cran.r-project.org/web/packages/tensorregress/index.html`

• **Tensor_sparse**: R program for multiway clustering via tensor block models.
`https://cran.r-project.org/web/packages/tensorsparse/index.html`

• **BinaryT**: R program for low-rank tensor estimation from binary observations.
`https://github.com/Miaoyanwang/Binary-Tensor`

• **Multi-Cluster**: Matlab/R programs for three-way clustering of gene expression tensorial data.
`https://github.com/Miaoyanwang/Multi-Cluster`

• **TM-HOSVD**: Matlab program for efficient decomposition of higher-order tensors.
`https://github.com/Miaoyanwang/Two-mode-HOSVD`

• **ATOMM**: C program for association **a**nalysis with a **t**wo-**o**rganism **m**ixed-effects **m**odel.
`https://github.com/Miaoyanwang/Two-way-MMA`

• **G-STRATEGY**: C program for optimal selection strategy of individuals for genotyping based on phenotypes and pedigrees. https://github.com/Miaoyanwang/G-STRATEGY

Services

Grant Review Panel:

Research Grants Council (RGC) of Hong Kong, Physical Science Panel, 2020.

Natural Sciences & Engineering Research Council (NSERC, NSF equivalent in Canada), Discovery Grant Competition Panel, 2020.

NSF-DMS panel, 2019 (declined due to time conflict with 2019 NeurIPS).

University of Wisconsin Madison, Institute for Clinical and Translational Research (ICTR), Pilot Program, 2019.

Contribution to Diversity:

Media coverage and feature article **"Women in STEM: 5 Thoughtful Ways to Recruit and Retain Them"** by Course Hero, July, 2020.

Member in Gooey Chocolate Cake Lunch as women faculty in the physical science in UW-Madison.

Member in Women in Probability, an NSF-funded organization for women active in probability research in North America.

Seminar Organizer:

European Society for Evolution Biology - Coevolution workshop, Munich, Germany, 2019

International Conference on Frontiers of Data Science, Hangzhou, China, 2019

Department Weekly Seminar Committee, Madison, WI, 2018 - 2019.

Department Committee for PhD admission, PhD qualification exam, MS exam. 2018 - now.

Reviewer for Journal of the American Statistical Association (JASA), Journal of the Royal Statistical Society: Series B (JRSS-B), Journal of Machine Learning Research (JMLR), International Conference of Machine Learning (ICML), IEEE Transactions on Information Theory, IEEE Transactions on Signal Processing, NeurIPS Comp Bio, Journal of Computational and Graphical Statistics, Linear Algebra and Applications, Linear and Multilinear Algebra, Statistics and Probability Letter.

Member in IMS (Institute of Mathematical Statistics), SIAM (Society for Industrial and Applied Mathematics), ASHG (American Society of Human Genetics), IGES (International Genetic Epidemiology Society)

Statistical Consultant, The University of Chicago, 2011 - 2015. Led consulting projects and supervised Masters and junior PhD students to provide statistical support for the larger university community.