

# CURRICULUM VITAE

## PERSONAL INFORMATION

---

**Name:** Xin Bing **English first name:** Mike  
**Email:** xin.bing@utoronto.edu  
**Phone:** (607) 697 3443  
**Current position:** Assistant Professor at the University of Toronto

## EDUCATION

---

**Cornell University** 09/2016 – 06/2021  
Ph.D. in Statistics with minor in Mathematics  
*Advisors: Florentina Bunea and Marten Wegkamp*

**University of Washington – Seattle** 09/2014 – 06/2016  
M.S. in Statistics

**Shandong University** 09/2013 – 06/2016  
M.F.E. in Financial Mathematics

**Shandong University** 09/2009 – 06/2013  
B.S. in Mathematics and Statistics

## EMPLOYMENT

---

**Assistant Professor:** University of Toronto, Department of Statistical Sciences. 06/2022 – present

**Computational Systems Immunology Research Associate:** jointly between the University of Pittsburgh, School of Medicine (with Jishnu Das), and Cornell University, Department of Statistics and Data Science. 01/2020 – 08/2020

**Teaching Assistant:** Cornell University, Department of Statistics and Data Science. 09/2016 – 06/2021

## RESEARCH OF INTEREST

---

High-dimensional statistics, low-rank matrix estimation, multivariate analysis, model-based clustering, latent factor model, topic models, minimax estimation, high-dimensional inference, statistical and computational trade-offs, applications of the Wasserstein distance in statistics, applications to genetics, neuroscience, immunology and other areas.

## HONORS AND AWARDS

---

- 2021 Hsien and Daisy Yen Wu Scholarships
- 2021 IMS Lawrence D. Brown Ph.D. Student Award
- Honor graduate of Shandong University in 2013
- Excellent student scholarship grantee for meticulous academic performance from 2010 to 2012

## JOURNAL PUBLICATIONS

---

1. Xin Bing, Florentina Bunea and Marten Wegkamp.  
Detecting approximate replicate components of a high-dimensional random vector with latent structure.  
*Bernoulli*, 29(2): 1368–1391, 2023. [Link to the paper.](#)

2. Dian Jin, Xin Bing and Yuqian Zhang.  
Unique sparse decomposition of low rank matrices.  
*IEEE Transactions on Information Theory*, 2022. [Link to the paper.](#)
3. Xin Bing, Florentina Bunea, Seth Strimas-Mackey and Marten Wegkamp.  
Likelihood estimation of sparse topic distributions in topic models and its applications to Wasserstein document distance calculations.  
*Ann. Statist.* 50(6): 3307-3333, December 2022. [Link to the paper.](#)
4. Xin Bing, Florentina Bunea, Seth Strimas-Mackey and Marten Wegkamp.  
Prediction in latent factor regression: Adaptive PCR and beyond.  
*The Journal of Machine Learning Research*, 22(177):150, 2021. [Link to the paper.](#)
5. Xin Bing, Florentina Bunea and Marten Wegkamp.  
Inference in interpretable latent factor regression models.  
*Bernoulli*, 28(2): 997–1020, May 2022. [Link to the paper.](#)
6. Xin Bing, Yang Ning and Yaosheng Xu.  
Adaptive estimation of multivariate regression with hidden variables.  
*Ann. Statist.* 50(2): 640-672, 2022. [Link to the paper.](#)
7. Xin Bing, Florentina Bunea and Marten Wegkamp.  
Optimal estimation of sparse topic models.  
*The Journal of Machine Learning Research* **21** (2020), no. 177, 1–45. [Link to the paper.](#)
8. Xin Bing, Florentina Bunea and Marten Wegkamp.  
A fast algorithm with minimax optimal guarantees for topic models with an unknown number of topics.  
*Bernoulli* 26 (2020), no. 3, 1765–1796. [Link to the paper.](#)
9. Xin Bing, Florentina Bunea, Yang Ning and Marten Wegkamp.  
Adaptive estimation in structured factor models with applications to overlapping clustering.  
*Ann. Statist.* 48 (2020), no. 4, 2055–2081. [Link to the paper.](#)
10. Xin Bing and Marten Wegkamp.  
Adaptive estimation of the rank of the coefficient matrix in high-dimensional multivariate response regression models.  
*Ann. Statist.* 47 (2019), no. 6, 3157–3184. [Link to the paper.](#)

## CONFERENCE PUBLICATION

---

1. Dian Jin, Xin Bing and Yuqian Zhang.  
Unique sparse decomposition of low rank matrices.  
*Neural Information Processing Systems*, 2021. [Link to the paper.](#)

## COLLABORATIVE PUBLICATIONS AND DISCUSSIONS

---

1. Xin Bing, Tyler Lovelace, Florentina Bunea, Marten Wegkamp, Harinder Singh, Panayiotis Benos, Jishnu Das.  
Essential Regression – a generalizable framework for inferring causal latent factors from multi-omic human datasets.  
*Patterns* (Cell press), 3(5): 100473, 2022. [Link to the paper.](#)
2. Xin Bing, Florentina Bunea, Martin Royer and Jishnu Das.  
Latent model-based clustering for biological discovery.  
*iScience* 14 (2019), 125–135. [Link to the paper.](#)

3. Xin Bing and Marten Wegkamp.  
Discussion of random-projection ensemble classification by Timothy I. Cannings and Richard J. Samworth.  
*J. R. Statist. Soc. B* 79 (2017), no. 4, 1006–1007.

## PAPERS UNDER REVIEW

---

1. Optimal Discriminant Analysis in High-Dimensional Latent Factor Models. Joint work with Florentina Bunea and Marten Wegkamp. [arXiv link](#). (under major revision in the *Annals of Statistics*)
2. Inference in High-dimensional Multivariate Response Regression with Hidden Variables. Joint work with Wei Cheng, Huijie Feng and Yang Ning. [arXiv link](#). (under major revision in the *Journal of American Statistical Association*)
3. Interpolating Discriminant Functions in High-Dimensional Gaussian Latent Mixtures. [arXiv link](#). (under major revision in *Biometrika*)
4. Xin Bing, Florentina Bunea and Jonathan Niles-Weed. The Sketched Wasserstein Distance for mixture distributions. [arXiv link](#). (under revision in the *Annals of Statistics*)
5. Javad Rahimikollu, Hanxi Xiao, Anna E. Rosengart, Tracy Tabib, Paul Zdinak, Kun He, Xin Bing, Florentina Bunea, Marten Wegkamp, Amanda C. Poholek, Alok V Joglekar, Robert A Lafyatis, Jishnu Das. SLIDE: Significant Latent Factor Interaction Discovery and Exploration across biological domains. [bioRxiv link](#). (under major revision in *Nature Methods*)

## WORK IN PROGRESS

---

1. Provable Recovery of Low-rank and Sparse Decomposition via Deflations. Joint work with Dian Jin and Yuqian Zhang. (Preprint available upon request)
2. High-dimensional Hypothesis Testing on Multivariate Gaussianity. Joint work with Derek Latremouille. (Preprint available upon request)
3. Minimax Optimal Discriminant Analysis in Multi-class Classification. Joint work with Marten Wegkamp. (Preprint available upon request)
4. Interpretable Latent Factor Regression Models with Generalized Linear Links. Joint work with Florentina Bunea and Marten Wegkamp. (Preprint available upon request)

## INVITED TALKS AND PRESENTATIONS

---

- International Chinese Statistical Association (ICSA): 2023 applied statistics symposium. Date: 11–14 Jun, 2023. Ann Arbor, Michigan.
- Statistics seminar in the Department of Mathematics and Statistics at Auburn University. April 19, 2023.
- Statistics seminar in the Department of Statistics and Actuarial Science at the University of Waterloo. March 08, 2023.
- 15th International Conference of the ERCIM WG on Computational and Methodological Statistics (CM-Statistics 2022). King’s College London, UK. Date: 17–19 December 2022
- International Chinese Statistical Association (ICSA): 2022 applied statistics symposium. Date: 19–22 Jun, 2022. Gainesville, FL.
- Oberwolfach Workshop: Re-thinking High-dimensional Mathematical Statistics. Date: 15–21 May, 2022. Oberwolfach Research Institute for Mathematics, Germany.
- Statistics seminar in the Institute of Data Analysis and Modeling in economics and statistics, UCLouvain. Nov 26, 2021.

- Joint Statistical Meetings / IMS Annual Meeting. Seattle, Washington, USA, August 7–12, 2021.
- International Chinese Statistical Association (ICSA): 2020 applied statistics symposium. Houston, TX. (Postponed)
- Workshop at Institute for Advanced Study (IAS, Princeton University): Missing Data Challenges in Computation, Statistics, and Applications. March 16–18, 2020. (Postponed)
- International Chinese Statistical Association (ICSA): 2019 applied statistics symposium. Date: 09 – 12 Jun, 2019. Raleigh, NC.
- International Center for Mathematical Science (ICMS) workshop: Computational strategies for large-scale statistical data analysis. Date: 02–07 July 2018. Edinburgh, UK.
- Oberwolfach Workshop: Matrix Estimation Meets Statistical Network Analysis: Extracting low-dimensional structures in high dimension. Date: 17–23 Jun 2018. Oberwolfach Research Institute for Mathematics, Germany.
- Poster sessions: Cornell Day of Statistics 2017 (2018). Date: 08 Sep 2017 (07 Sep 2018). Cornell University.
- Poster session: A Cornell Collaboration: Machine Learning in Medicine Symposium. Date: 27–28 Sep 2018. Cornell University.
- Student seminar talks in the School of Operations Research and Information Engineering, Cornell University.
- Student seminar talks in the Department of Statistics and Data Science, Cornell University.

## TEACHING

---

Teaching at the University of Toronto

- STA314H: Statistical Methods for Machine Learning I

Supervised courses at Cornell University

- STSCI 5080: Probability Models and Inference
- BTRY 3520: Statistical Computing
- ILRS 3110: Probability Models and Inference for the Social Sciences
- MATH 4720: Statistics
- STSCI 4740: Machine Learning and Data Mining
- BTRY/STSCI 4090: Theory of Statistics

## SERVICE TO THE ACADEMIC COMMUNITY

---

Referee for the Electronic Journal of Statistics, the Annals of Statistics, Biometrika, Statistica Sinica, Journal of the Royal Statistical Society - Series B, Journal of American Statistical Association, Journal of Computational and Graphical Statistics, Journal of Machine Learning Research, the SIAM Journal on Imaging Sciences, the Journal of Econometrics