# Tianyu Zhang

Department of Statistics & Data Science
Carnegie Mellon University
Pittsburgh, PA 15213

Phone: (+1) 206-488-2242
Email: tianyuz3@andrew.cmu.edu

## Education

**Ph.D.** Biostatistics, University of Washington 9/2017-8/2022

Dissertation Title *Modern Sieve Estimators for Nonparametric Problems: Streaming Data and High-dimensional Data.*
Committee Members *Noah Simon (Chair), Alex Luedtke Marco Carone, Rekha Thomas*

**B.S.** Life Science & Mathematics (Double Major), Peking University 9/2013-7/2017

## Employment

**Post Doctoral Researcher**
Department of Statistics & Data Science, Carnegie Mellon University 9/2022-Present
Hosting Faculty *Jing Lei and Kathryn Roeder*

**Research Assistant**
Department of Biostatistics, University of Washington 3/2018-8/2022

**Statistical Consultant**
FOXO Technologies Inc. 9/2021-8/2022

**Research Scientist Intern**
Modeling and Optimization team, Amazon 6-9/2020 & 6-9/2021

## Research Publication (* indicates co-first author)

### *Methodology Research*

1. **Tianyu Zhang**, Hao Lee, and Jing Lei. "Winners with Confidence: Discrete Argmin Inference with an Application to Model Selection" arXiv:2408.02060 (2024).
2. Kenta Takatsu, **Tianyu Zhang**, and Arun Kumar Kuchibhotla. "From isotonic to Lipschitz regression: a new interpolative perspective on shape-restricted estimation" arXiv:2307.05732 (2024).
3. **Tianyu Zhang**, Jing Lei, and Kathryn Roeder. "Debiased Projected Two-Sample Comparisons for Single-Cell Expression Data." arXiv preprint arXiv:2403.05679 (2024).

4. **Tianyu Zhang** and Jing Lei. "Online Estimation with Rolling Validation: Adaptive Nonparametric Estimation with Streaming Data." arXiv preprint arXiv:2310.12140 (2023). (being revised for *the Annals of Statistics*)
5. **Tianyu Zhang**, Geyu Zhou, Lambertus Klei, Peng Liu, Alexandra Chouldechova, Hongyu Zhao, Kathryn Roeder, Max G'Sell, and Bernie Devlin. "Evaluating and improving health equity and fairness of polygenic scores." *Human Genetics and Genomics Advances* 5, no. 2 (2024).
6. **Tianyu Zhang**, Noah Simon "Regression in Tensor Product Spaces by the Method of Sieves," *Electronic Journal of Statistics*, 17(2), 3660-3727, (2023)
7. Yunhua Xiang*****, **Tianyu Zhang***, Xu Wang, Ali Shojaie, and Noah Simon. "On the Optimality of Nuclear-norm-based Matrix Completion for Problems with Smooth Non-linear Structure." *Journal of Machine Learning Research* 24, no. 228 (2023): 1-38.
8. **Tianyu Zhang** and Noah Simon. "An Online Projection Estimator for Nonparametric Regression in Reproducing Kernel Hilbert Spaces." *Statistica Sinica* 33.1 (2023): 127.
9. **Tianyu Zhang** and Noah Simon. "A Sieve Stochastic Gradient Descent Estimator for Online Nonparametric Regression in Sobolev Ellipsoids." *The Annals of Statistics* 50, no. 5 (2022): 2848-2871.

## *Interdisciplinary and Collaborative Research*

1. **Tianyu Zhang***, Chris A Gentry*, Nicole M Kuderer, Gary H Lyman, Bernard Ng, Despina Michailidou. "Association of Selective Serotonin and Serotonin-Norepinephrine Reuptake Inhibitor Use with Incidence of Cardiovascular Events in Veterans with Giant Cell Arteritis and Polymyalgia Rheumatica in the United States " (in submission, available upon request)
2. Conghao Zhou, Hao-Yi Wu, Andrés N. Salcedo, Sebastian Grandis, Tesla Jeltema, Alexie Leauthaud, Matteo Costanzi, Tomomi Sunayama, David H. Weinberg, **Tianyu Zhang**, Eduardo Rozo, Chun-Hao To, Sebastian Bocquet, Tamas Varga, Matthew Kwiecien. "Forecasting the constraints on optical selection bias and projection effects of galaxy cluster lensing with multiwavelength data." *Physical Review D* 110, no. 10 (2024): 103508.
3. Yihang Shen, Lingge Yu, Yutong Qiu, **Tianyu Zhang** and Carl Kingsford. "Improving Hi-C contact matrices using genome graphs." *RECOMB* 2024
4. Despina Michailidou, **Tianyu Zhang***, Nicole M. Kuderer, Gary H. Lyman, Andreas P. Diamantopoulos, Pavlos Stamatis, and Bernard Ng. "Predictive Models for Thromboembolic Events in Giant Cell Arteritis: A US Veterans Health Administration Population-based Study." *Frontiers in Immunology* 13 (2022): 997347.

5. Despina Michailidou, **Tianyu Zhang**, Pavlos Stamatis, and Bernard Ng. "Risk of venous and arterial thromboembolism in patients with giant cell arteritis and/or polymyalgia rheumatica: A Veterans Health Administration population‐based study in the United States." *Journal of Internal Medicine* 291, no. 5 (2022): 665-675.
6. Yiming Wang, Weikaixin Kong, Liang Wang, **Tianyu Zhang**, Boyue Huang, Jia Meng, Baoxue Yang, Zhengwei Xie, and Hong Zhou. "Multiple-purpose Connectivity Map Analysis Reveals the Benefits of Esculetin to Hyperuricemia and Renal Fibrosis." *International Journal of Molecular Sciences* 21, no. 20 (2020): 7695.
7. Zhengwei Xie*****, **Tianyu Zhang***, and Qi Ouyang. "Genome-scale Fluxes Predicted under the Guidance of Enzyme Abundance using a Novel Hyper-cube Shrink Algorithm." *Bioinformatics* 34, no. 3 (2018): 502-510.

## Teaching Experience

**Guest Lecturer**

Computational Biology Department, Carnegie Mellon University — Fall 2024
Introduction to Statistical Genetics (02704), Instructor: *Martin Zhang*
(a lecture on polygenic risk score for computational biology students)

**Teaching Assistant**

Department of Biostatistics, University of Washington
BIOST 514/517 Biostatistics, Instructor: *Ken Rice* — Fall 2020
BIOST 546 Machine Learning for Biomedical and Public Health Big Data — Winter 2019
Instructor: *Daniela Witten*
(leading discussion sessions, office hours, homework, and exam grading)

**Teaching Assistant**

School of Life Sciences, Peking University
Calculus I Biostatistics, Instructor: *Jiazhong Yang* — Summer 2016 & 2017
(leading discussion sessions that cover new content and exercise, homework grading)

## Honors and Recognition

Featured in the *HGG Advances* monthly author interview series — 2024
Best Student Oral Presentation, *WNAR* — 2022
POSCO Scholarship, Peking University — 2015-2016
Suzhou Industrial Park Scholarship, Peking University — 2014-2015

## Professional Activities

**Peer Reviewer:** Annals of Statistics, Biometrika, Biostatistics, Computational Statistics, Journal of Computational and Applied Mathematics, Journal of Machine Learning Research (x4), Journal of the Royal Statistical Society Series B (x2), PLOS genetics

**Open-source software:** Sieve(R-CRAN), HMC(R-CRAN), Joint-Lassosum(GitHub)

**Statistical Learning Applied to Biostatistics (SLAB) Lab Seminar Organizer** 9/2020-9/2021
Department of Biostatistics, University of Washington

**Peer Mentor** 9/2019-9/2020
Department of Biostatistics, University of Washington

## Selected Presentations

1. Seminar, Columbia University, Department of Biostatistics, 2025. Title: *Adaptive and Scalable Nonparametric Estimation via Stochastic Optimization.*
2. Seminar, University of North Carolina at Chapel Hill, Department of Biostatistics, 2025. Title: *Adaptive and Scalable Nonparametric Estimation via Stochastic Optimization.*
3. Seminar, University of California, Santa Barbara, Department of Statistics and Applied Probability, 2025. Title: *Adaptive and Scalable Nonparametric Estimation via Stochastic Optimization.*
4. Seminar, University of Texas at Austin, Department of Statistics and Data Sciences, 2025. Title: *Adaptive and Scalable Nonparametric Estimation via Stochastic Optimization.*
5. Seminar, University of Texas at Dallas, Department of Mathematical Sciences, 2024. Title: *Adaptive and Scalable Nonparametric Estimation via Stochastic Optimization.*
6. Seminar, University of Michigan, Department of Statistics, 2024. Title: *Adaptive and Scalable Nonparametric Estimation via Stochastic Optimization.*
7. Seminar, Iowa State University, Department of Statistics, 2024. Title: *Adaptive and Scalable Nonparametric Estimation via Stochastic Optimization.*
8. Invited seminar presentation at the Heart Institute (InCor), University of São Paulo Medical School, Brazil (remote), 2024. Title: *Evaluating and Improving Health Equity and Fairness of Polygenic Scores*
9. Invited talk at the *American Society of Human Genetics* monthly journal club, 2024. Title: *Evaluating and Improving Health Equity and Fairness of Polygenic Scores*
10. Joint Statistical Meeting 2024. Topic-Contributed Sessions. Title: *Online Estimation with Rolling Validation: Adaptive Nonparametric Estimation with Streaming Data.*
11. American Statistical Association STATGEN 2024. Title: *Debiased Projected Two-Sample Comparisons for Single-Cell Expression Data.*

12. WNAR 2022. Title: *Regression in Tensor Product Spaces by the Method of Sieves*

## Skills

Computation: R, SQL, Python, Shell script, C++, MATLAB.
Language: Chinese (native); English (proficient); Japanese (beginner).