

Jin-Hong Du

School Address

Department of Statistics and Actuarial Science &
Musketeers Foundation Institute of Data Science
The University of Hong Kong
Hong Kong

Contact Information

jinhongd@hku.hk
+852-56630493



[Personal website](#)

PROFESSIONAL POSITIONS

Assistant Professor (tenure-track)

Aug. 2025 - Present

The University of Hong Kong

Department of Statistics and Actuarial Science & Musketeers Foundation Institute of Data Science

EDUCATION

Doctor of Philosophy, Statistics & Machine Learning

Carnegie Mellon University, Pittsburgh, PA 15213

Aug. 2021 - Aug. 2025

THESIS - Towards Reliable and Robust Causal Inference with High-dimensional Outcomes

Master of Science, Statistics

The University of Chicago, Chicago, IL 60637

Aug. 2019 - Mar. 2021

THESIS - Spectral Nearness Problems as Multiparameter Singular Value Problems

Bachelor of Science, Statistics

Sun Yat-sen University, Guangzhou, China

Sept. 2015 - Jun. 2019

THESIS - Blind Deconvolution on Graphs with Noisy Observations

Exchange Student

The University of California, Berkeley, Berkeley, CA 94720

Aug. 2017 - Dec. 2017

PROFESSIONAL EXPERIENCE

- Research Engineer Intern
Tencent, TEG (Technology Engineering Group) - IDC (Internet Data Center) Shenzhen, China
May-Aug. 2019
- Algorithm Engineer Intern
Perception Vision Medical Technologies - Nasopharynx Carcinoma Project Group Guangzhou, China
Jul. 2018-Jan. 2019

PUBLICATIONS

* Equally Contributed ^a Alphabetical Order ‡ Corresponding Author

Preprint & Under Review

- [1] Yan Chen, Weijing Tang, and **Jin-Hong Du**. “AI-Augmented Statistical Network Estimation with Proxy Gene Embeddings”. 2026. DOI: [10.48550/arXiv.2607.05774](#).
- [2] Shibo Yu, Yingzhou Wang, Yan Chen, Guodong Li, and **Jin-Hong Du**. “Heterogeneous Judge-Aware Ranking with Sensitivity, Disagreement, and Confidence”. 2026. DOI: [10.48550/arXiv.2605.05073](#).
- [3] **Jin-Hong Du**, Kathryn Roeder, and Larry Wasserman. “Disentangled feature importance”. 2025. DOI: [10.48550/arXiv.2507.00260](#).

- [4] Pratik Patil, **Jin-Hong Du**, and Ryan J. Tibshirani. “Revisiting optimism and model complexity in the wake of overparameterized learning”. 2024. DOI: [10.48550/arXiv.2410.01259](https://doi.org/10.48550/arXiv.2410.01259).

Journal

- [1] **Jin-Hong Du**, Maya Shen, Hansruedi Mathys, and Kathryn Roeder. “Uncovering causal relationships in single cell omic studies with causarray”. In: *Briefings in Bioinformatics* (2026). [\[bioRxiv\]](https://doi.org/10.1093/bib/bbag175). DOI: [10.1093/bib/bbag175](https://doi.org/10.1093/bib/bbag175).
- [2] **Jin-Hong Du**, Kathryn Roeder, and Larry Wasserman. “Assumption-lean post-integrated inference with surrogate control outcomes”. In: *Biometrika* (2026). [\[arXiv\]](https://doi.org/10.1093/biomet/asag004). DOI: [10.1093/biomet/asag004](https://doi.org/10.1093/biomet/asag004).
- [3] Takuya Koriyama, Pratik Patil, **Jin-Hong Du**, Kai Tan, and Pierre C. Bellec. “Precise asymptotics of bagging regularized M-estimators”. In: *Annals of Statistics* 54.2 (2026). [\[arXiv\]](https://doi.org/10.1214/25-AOS2590), pp. 796–833. DOI: [10.1214/25-AOS2590](https://doi.org/10.1214/25-AOS2590).
- [4] **Jin-Hong Du**, Larry Wasserman, and Kathryn Roeder. “Simultaneous inference for generalized linear models with unmeasured confounders”. In: *Journal of the American Statistical Association* 120.551 (2025). [\[arXiv\]](https://doi.org/10.1080/01621459.2025.2485379), pp. 1945–1959. DOI: [10.1080/01621459.2025.2485379](https://doi.org/10.1080/01621459.2025.2485379).
- [5] Yaoming Zhen and **Jin-Hong Du**. “Network-based neighborhood regression”. In: *Journal of the American Statistical Association* 120.552 (2025). [\[arXiv\]](https://doi.org/10.1080/01621459.2025.2485342), pp. 2658–2671. DOI: [10.1080/01621459.2025.2485342](https://doi.org/10.1080/01621459.2025.2485342).
- [6] **Jin-Hong Du**, Zhenghao Zeng, Edward H Kennedy, Larry Wasserman, and Kathryn Roeder. “Causal inference for genomic data with multiple heterogeneous outcomes”. In: *Journal of the American Statistical Association* 120.552 (2025). [\[arXiv\]](https://doi.org/10.1080/01621459.2025.2468014), pp. 2484–2497. DOI: [10.1080/01621459.2025.2468014](https://doi.org/10.1080/01621459.2025.2468014).
- [7] Haeun Moon, **Jin-Hong Du**, Jing Lei, and Kathryn Roeder. “Augmented doubly robust post-imputation inference for proteomic data”. In: *The Annals of Applied Statistics* 19.2 (2025). [\[arXiv\]](https://doi.org/10.1214/25-AOAS2012), pp. 1006–1027. DOI: [10.1214/25-AOAS2012](https://doi.org/10.1214/25-AOAS2012).
- [8] ^aPierre C. Bellec, **Jin-Hong Du**, Takuya Koriyama, Pratik Patil, and Kai Tan. “Corrected generalized cross-validation for finite ensembles of penalized estimators”. In: *Journal of the Royal Statistical Society: Series B* 87.2 (2025). [\[arXiv\]](https://doi.org/10.1093/jrsssb/qkae092), pp. 289–318. DOI: [10.1093/jrsssb/qkae092](https://doi.org/10.1093/jrsssb/qkae092).
- [9] **Jin-Hong Du**^{*}, Tianyu Chen^{*}, Ming Gao, and Jingshu Wang. “Joint trajectory inference for single-cell genomics using deep learning with a mixture prior”. In: *Proceedings of the National Academy of Sciences* 121.37 (2024). [\[bioRxiv\]](https://doi.org/10.1073/pnas.2316256121), e2316256121. DOI: [10.1073/pnas.2316256121](https://doi.org/10.1073/pnas.2316256121).
- [10] **Jin-Hong Du**, Pratik Patil, Kathryn Roeder, and Arun Kumar Kuchibhotla. “Extrapolated cross-validation for randomized ensembles”. In: *Journal of Computational and Graphical Statistics* 33.3 (2024). [\[arXiv\]](https://doi.org/10.1080/10618600.2023.2288194), pp. 1061–1072. DOI: [10.1080/10618600.2023.2288194](https://doi.org/10.1080/10618600.2023.2288194).
- [11] Pratik Patil^{*}, **Jin-Hong Du**^{*}, and Arun Kumar Kuchibhotla. “Bagging in overparameterized learning: Risk characterization and risk monotonicization”. In: *Journal of Machine Learning Research* 24.319 (2023). [\[arXiv\]](https://doi.org/10.1080/10618600.2023.2288194), pp. 1–113. URL: <http://jmlr.org/papers/v24/23-0887.html>.
- [12] **Jin-Hong Du**, Zhanrui Cai, and Kathryn Roeder. “Robust probabilistic modeling for single-cell multimodal mosaic integration and imputation via scVAEIT”. In: *Proceedings of the National Academy of Sciences* 119.49 (2022). [\[bioRxiv\]](https://doi.org/10.1073/pnas.2214414119), e2214414119. DOI: [10.1073/pnas.2214414119](https://doi.org/10.1073/pnas.2214414119).
- [13] **Jin-Hong Du**^{*}, Yifeng Guo^{*}, and Xueqin Wang. “High Dimensional Portfolio Selection with Cardinality Constraints”. In: *Journal of the American Statistical Association* (2022). [\[arXiv\]](https://doi.org/10.1080/01621459.2022.2133718), pp. 1–13. DOI: [10.1080/01621459.2022.2133718](https://doi.org/10.1080/01621459.2022.2133718).
- [14] Jose Israel Rodriguez, **Jin-Hong Du**, Yiling You, and Lek-Heng Lim. “Fiber product homotopy method for multiparameter eigenvalue problems”. In: *Numerische Mathematik* 148 (2021). [\[arXiv\]](https://doi.org/10.1007/s00211-021-01215-6), pp. 853–888. DOI: [10.1007/s00211-021-01215-6](https://doi.org/10.1007/s00211-021-01215-6).

Conference & Workshop

- [1] Yifeng Guo, **Jin-Hong Du**, and Xiang Chen. “Beyond Importance: Interchange-Sobol Sensitivity Reveals Task-Specific Content Channels in Transformer Components”. In: *The First Conference on Statistics and Trustworthy AI for Cross (X)-Domain Acceleration*. [arXiv]. 2026.
- [2] Takuya Koriyama, Pratik Patil, **Jin-Hong Du**, Kai Tan, and Pierre C. Bellec. “Precise asymptotics of bagging regularized M-estimators”. In: *International Conference on Machine Learning (Journal-to-Conference track)*. [arXiv]. 2026.
- [3] Qiyuan Liu, Qirui Zhang, **Jin-Hong Du**[†], Siming Zhao[‡], and Jingshu Wang[‡]. “Effects of distance metrics and scaling on the perturbation discrimination score”. In: *International Conference on Learning Representations 2026 Workshop on Generative AI in Genomics (Gen²): Barriers and Frontiers*. [OpenReview] [arXiv]. 2026.
- [4] **Jin-Hong Du** and Pratik Patil. “Implicit regularization paths of weighted neural representations”. In: *Thirty-eighth Conference on Neural Information Processing Systems*. [OpenReview] [arXiv]. 2024.
- [5] Wenbin Zhou and **Jin-Hong Du**. “Distance-preserving generative modeling of spatial transcriptomics”. In: *23rd International Workshop on Data Mining in Bioinformatics*. [arXiv]. 2024.
- [6] Pratik Patil, **Jin-Hong Du**, and Ryan J. Tibshirani. “Optimal ridge regularization for out-of-distribution prediction”. In: *Proceedings of the 41st International Conference on Machine Learning*. Vol. 235. [PMLR] [arXiv]. PMLR, 21–27 Jul 2024, pp. 39908–39954.
- [7] Pratik Patil and **Jin-Hong Du**. “Generalized equivalences between subsampling and ridge regularization”. In: *Thirty-seventh Conference on Neural Information Processing Systems*. Vol. 36. [OpenReview] [arXiv]. 2023.
- [8] **Jin-Hong Du**^{*}, Pratik Patil^{*}, and Arun Kumar Kuchibhotla. “Subsample ridge ensembles: equivalences and generalized cross-validation”. In: *Proceedings of the 40th International Conference on Machine Learning*. Vol. 202. Proceedings of Machine Learning Research. [PMLR] [arXiv]. 23–29 Jul 2023, pp. 8585–8631.

Interdisciplinary Research

- [1] Yifeng Guo, **Jin-Hong Du**[†], and Xueqin Wang. “Conditional Value-at-Risk Portfolio Optimization in High Dimensions”. In: *Statistics and Its Interface* (2026). DOI: [10.4310/SII.260602011232](https://doi.org/10.4310/SII.260602011232).
- [2] Wenbin Zhou and **Jin-Hong Du**. “Distance-Preserving Representations for Genomic Spatial Reconstruction”. In: *IEEE Transactions on Computational Biology and Bioinformatics* (2025). [arXiv]. DOI: [10.1109/TCBBIO.2025.3628794](https://doi.org/10.1109/TCBBIO.2025.3628794).
- [3] Zhen Yang, **Jin-Hong Du**, Yiting Lin, Zhen Du, Li Xia, Qianchuan Zhao, and Xiaohong Guan. “Increasing the Energy Efficiency of a Data Center Based on Machine Learning”. In: *Journal of Industrial Ecology* (2021). DOI: [10.1111/jieec.13155](https://doi.org/10.1111/jieec.13155).
- [4] Rutong Zeng, Xiang Zhang, Chushan Zheng, **Jin-Hong Du**, Zixiong Gao, Jun Wei, Jun Shen, and Yao Lu. “Decoupling Convolution Network for Characterizing the Metastatic Lymph Nodes of Breast Cancer Patients”. In: *Medical Physics* 48.7 (2021), pp. 3679–3690. DOI: [10.1002/mp.14876](https://doi.org/10.1002/mp.14876).
- [5] Jing-Xian Tang^{*}, **Jin-Hong Du**^{*}, Yiting Lin, and Qing-Shan Jia. “Predictive Maintenance of VRLA Batteries in UPS towards Reliable Data Centers”. In: *IFAC-PapersOnLine* 53.2 (2020), pp. 13607–13612. ISSN: 2405-8963. DOI: [10.1016/j.ifacol.2020.12.854](https://doi.org/10.1016/j.ifacol.2020.12.854).
- [6] XingYu Fu, **Jin-Hong Du**, YiFeng Guo, MingWen Liu, Tao Dong, and XiuWen Duan. “A machine learning framework for stock selection”. In: (2018). DOI: [10.48550/arXiv.1806.01743](https://doi.org/10.48550/arXiv.1806.01743).

MENTORSHIP

◦ Postdoctoral Researchers

– Dr. Yan Chen (2025 - Present) – PhD from USTC, co-advised with Prof. Guodong Li

◦ PhD Students

- Yingqi Fan (IDS, incoming Fall 2026)
- Manxi Sun (SAAS, incoming Fall 2026)

TEACHING EXPERIENCE

- Instructor Fall 2026
DATA 8020: Advanced Causal Inference
- Teaching Assistant (Carnegie Mellon University) Fall 2021
STAT 36315: Statistical Graphics and Visualization
- Teaching Assistant (The University of Chicago) Autumn 2020
STAT 32940/CAAM 32940/FINM 33180: Multivariate Data Analysis via Matrix Decompositions
- Grader (The University of Chicago) Autumn 2020
STAT 30900/CMSC 37810: Mathematical Computation I
- Grader (The University of Chicago) Spring 2020, Winter 2021
STAT 28000/CAAM 28000: Optimization.

ACADEMIC ACTIVITIES

- 2026/07/13 Invited talk on “AI-Augmented Statistical Network Estimation with Proxy Gene Embeddings”, at 2026 International Biometric Conference, Seoul, South Korea.
- 2026/07/07 Invited talk on “Seeing through correlations: Disentangled feature importance”, at 2026 IMS Annual Meeting, Salzburg, Austria.
- 2026/06/28 Invited talk on “Seeing through correlations: Disentangled feature importance”, at 2026 ICSA China Conference, Southern University of Science and Technology, Shenzhen, China.
- 2026/06/16 Invited talk on “Seeing through correlations: Disentangled feature importance”, at IMS Asia Pacific Rim Meeting 2026, Chinese University of Hong Kong, Hong Kong, China.
- 2026/04/15 Contributed talk on “Assumption-lean post-integrated inference with surrogate control outcomes”, at European Causal Inference Meeting 2026, University of Oxford, Oxford, UK.
- 2026/01/13 Internal talk on “Predicting unseen perturbation effects using data similarity and global structure”, at [AI for Science Mini-Symposium](#), Hong Kong, China.
- 2025/12/15 Invited talk on “Assumption-lean post-integrated inference with surrogate control outcomes”, at 2025 IMS International Conference on Statistics and Data Science, Seville, Spain.
- 2025/12/10 Internal talk on “Seeing through correlations: Disentangled feature importance”, at 2025 IDS seminar, Hong Kong, China.
- 2025/11/01 Invited talk on “Assumption-lean post-integrated inference with surrogate control outcomes”, at 2025 Shanghai-Hong Kong-Macao-Singapore Symposium, Shanghai, China.
- 2025/08/04 Contributed talk on “Assumption-lean post-integrated inference with surrogate control outcomes”, at 2025 Joint Statistical Meeting, Nashville, TN, USA.
- 2025/07/31-2025/08/02 Contributed talk and poster presentation on “Assumption-lean post-integrated inference with surrogate control outcomes”, at 2025 IMS New Researchers Conference, Nashville, TN, USA.
- 2025/05/12-2025/05/16 American Institute of Mathematics (AIM) workshop on [Algorithmic stability: mathematical foundations for the modern era](#), at California Institute of Technology, Pasadena, CA, USA.
- 2024/12/12 Poster presentation on “Implicit regularization paths of weighted neural representations”, at Thirty-Eighth Annual Conference on Neural Information Processing Systems, Vancouver, Canada.

- 2024/08/05 Contributed talk on “Simultaneous inference for generalized linear models with unmeasured confounders”, at 2024 Joint Statistical Meeting, Portland, OR, USA.
- 2024/07/23 Spotlight poster presentation (**335/9473=3.5%**) on “Optimal ridge regularization for out-of-distribution prediction”, at Forty-First International Conference on Machine Learning, Vienna, Austria.
- 2024/05/02 Contributed talk on “Methods for removing unmeasured confounding effects in high throughput screens, such as CRISPR and single-cell differential expression”, at STATGEN 2024, the ASA SSGG first section conference, University of Pittsburgh, Pittsburgh, PA, USA.
- 2024/04/12 Invited talk on “Extrapolated cross-validation for randomized ensembles”, at Speaker Series “Hot Topics in Data Science”, University at Buffalo, remote.
- 2023/11/20 Invited talk on “Simultaneous inference for generalized linear models with unmeasured confounders”, at Institute for Mathematical Sciences, National University of Singapore, Singapore.
- 2023/08/06 Contributed talk on “Extrapolated cross-validation for randomized ensembles”, at 2023 Joint Statistical Meeting, Toronto, ON, Canada.
- 2023/07/28 Oral presentation (**158/6538=2.4%**) on “Subsample ridge ensembles: equivalences and generalized cross-validation”, at Fortieth International Conference on Machine Learning, Honolulu, HI, USA.
- 2021/04/26 Contributed talk on “Model-based trajectory inference for single-cell RNA sequencing using deep learning with a mixture prior”, at MCBIOS & MAQC 2021 Joint Conference, remote.

ACADEMIC SERVICE

- Session chair of “DL05 — Modern Inference and Optimization for Machine Learning” at IMS Asia Pacific Rim Meeting 2026, Chinese University of Hong Kong, Hong Kong, China.
- Session chair of “Single cell and spatial” at STATGEN 2024, the ASA SSGG first section conference, University of Pittsburgh, Pittsburgh, PA, USA.
- Reviewer for Journal of Machine Learning Research, Electronic Journal of Statistics, Journal of the Royal Statistical Society Series B, Annals of Applied Statistics, Statistics in Medicine.
- Reviewer for Proceedings of the National Academy of Sciences, Nature Communications, Nature Structural & Molecular Biology.
- Reviewer for IEEE Transactions on Neural Networks and Learning Systems.
- Reviewer for Conference on Neural Information Processing Systems, International Conference on Machine Learning, International Conference on Learning Representations, International Conference on Artificial Intelligence and Statistics.

AWARDS & HONORS

- IMS Lawrence D. Brown Ph.D. Student Award 2026
- Team leader of *Outlier*, winner (3rd place) of [The First Virtual Cell Challenge](#) Dec. 2025
- HKU-100 Scholar Aug. 2025
- IMS New Researchers Travel Award Aug. 2025
- Young Scientists Excellence Awards - Finalist (MCBIOS & MAQC 2021 Joint Conference) Apr. 2021
- Tuition Scholarship Award (The University of Chicago) 2019-2021
- Outstanding Undergraduate Thesis Award (Sun Yat-sen University) Jun. 2019
- Outstanding Graduates Award (Sun Yat-sen University) Jun. 2019

