# Jun Yu

 $15650709378 \mid \underline{yujunbeta@bit.edu.cn} \mid https://math.bit.edu.cn/szdw/jgml/sxfxx/yj/index.htm$ 

#### EDUCATION

Peking University	Beijing, China
Ph.D. in Statistics	2014/09– $2019/07$
• Supervisor: Prof. Mingyao Ai	
• Dissertation: Design of Medical Experiments and Subsampling Techniques for Big Data	
Nankai University	Tianjin, China
B.S. in Mathematics and Applied Mathematics	2010/09-2014/07
Nankai University	Tianjin, China
B.E. in Finance (Second major)	2010/09-2014/07

#### RESEARCH INTERESTS

**Statistical theory and methodology:** Design of experiments; Sampling theory and subsampling techniques; Data reduction methods of big data

Applied statistics: Analysis the network data and solving the real-world problem via networks.

#### WORKING EXPERIENCE

Associate Professor	$2025/02-\mathrm{Present}$
School of Mathematics and Statistics, Beijing Institute of Technology	Beijing, China
Assistant Professor	2019/09 - 2025/01
School of Mathematics and Statistics, Beijing Institute of Technology	Beijing, China
Ph.D. Graduate Supervisor of Statistics	$2021/11-\mathrm{Present}$
School of Mathematics and Statistics, Beijing Institute of Technology	Beijing, China
Visiting Scholar	2019/12 - 2021/01
Department of Statistics, University of Georgia.	GA, USA.
• Co-mentor: Prof. Ping Ma and Prof. Wenxuan Zhong	

#### PUBLICATIONS

Remarks: \* corresponding author, # first and co-first author, students/trainees underlined.

#### **Refereed Journal Papers**

Sampling and subsampling techniques

- 1. Chen, Haolin, Holger Dette, and **Jun Yu**<sup>\*</sup>. "Multi-resolution subsampling for linear classification with massive data." Journal of the Royal Statistical Society: Series B. (2025): DOI:10.1093/jrsssb/qkaf017.
- Yu, Jun, Haiying Wang, Mingyao Ai, and Huiming Zhang. "Optimal distributed subsampling for maximum quasi-likelihood estimators with massive data." *Journal of the American Statistical Association* 117 (2022): 265-276. (2024 ESI highly cited paper)
- Li, Mengyu, Jun Yu<sup>#</sup>, and Cheng Meng. "Importance sparsification for Sinkhorn algorithm." Journal of machine learning research, 24 (2023), 1-44.
- 4. Yu, Jun, Haiying Wang, and Mingyao Ai. "A Subsampling Strategy for AIC-based Model Averaging with Generalized Linear Models." *Technometrics*, (2025), 67, 122–132.
- Yu, Jun, Zhiqiang Ye, Mingyao Ai, and Ping Ma. "Optimal subsampling for online data streams with categorical responses." *Journal of Computational and Graphical Statistics*, (2024), 1–19, DOI: 10.1080/10618600.2024.2421990.
- Li, Mengyu, Jun Yu<sup>#</sup>, Tao Li, and Cheng Meng. "Core-elements for Least Squares Estimation." Statistics and Computing, (2024), 34, 190.

- Yu, Jun, Mingyao Ai, and Zhiqiang Ye. "A review on design inspired subsampling for big data." Statistical Papers, (2024), 65, 467–510. (2023 early access version ESI highly cited paper)
- 8. Ai, Mingyao, **Jun Yu**, Huiming Zhang, and Haiying Wang. "Optimal subsampling algorithms for big data regressions." *Statistica Sinica* 31 (2021): 749-772. (2020-2024 ESI highly cited paper)
- 9. Deng, Jiayi, Xiaodong Yang, **Jun Yu**, Jun Liu, Zhaiming Shen, Danyang Huang, and Huimin Cheng. "Network Tight Community Detection." International Conference on Machine Learning (ICML 2024)
- 10. Ye, Zhiqiang, **Jun Yu**<sup>#</sup>, and Mingyao Ai. "Optimal subsampling for multinomial logistic models with big data." *Statistica Sinica*, (2023), DOI: 10.5705/ss.202022.0277.
- Han, Yixin, Jun Yu<sup>#</sup>, Nan Zhang, Cheng Meng, Ping Ma, Wenxuan Zhong, and Changliang Zou. "Leverage classifier: another look at support vector machine." *Statistica Sinica*, (2023), DOI: 10.5705/ss.202023.0124.
- Li, Mengyu, Jun Yu<sup>#</sup>, Hongteng Xu, and Cheng Meng. "Efficient approximation of Gromov Wasserstein distance using importance sparsification." *Journal of Computational and Graphical Statistics*, (2023), DOI: 10.1080/10618600.2023.2165500.
- Yu, Jun, Jiaqi Liu, and Haiying Wang. "Information-based optimal subdata selection for non-linear models." *Statistical Papers*, 64 (2023), 1069–1093.
- 14. Lv, Shurui, **Jun Yu**, Yan Wang, and Jiang Du. "Fast calibration for computer models with massive physical observations." *SIAM/ASA Journal on Uncertainty Quantification*, 11 (2023): 1069-1104.
- Diao, Huaimin, Mengtong Ai, Yubin Tian, and Jun Yu<sup>\*</sup>. "Efficient basis selection for smoothing splines via rotated lattices." STAT, 12 (2023), e581.
- Yu, Jun and Haiying Wang. "Subdata selection algorithm for linear model discrimination." Statistical Papers 63 (2022), 1883–1906.
- Zhang, Jingyi, Cheng Meng, Jun Yu, Mengrui Zhang, Wenxuan Zhong, and Ping Ma. "An optimal transport approach for selecting a representative subsample with application in efficient kernel density estimation." Journal of Computational and Graphical Statistics, 32 (2022), 329-339.
- Meng, Cheng, Jun Yu, Yongkai Chen, Wenxuan Zhong, and Ping Ma. "Smoothing splines approximation using Hilbert curve basis selection." *Journal of Computational and Graphical Statistics*, 31 (2022), 802-812.
- Ai, Mingyao, Fei Wang, Jun Yu<sup>\*</sup>, and Huiming Zhang. "Optimal subsampling for large-scale quantile regression." Journal of Complexity 62 (2021): 101512. (2025 ESI highly cited paper)

Experimental design and its application

- 1. Ai, Mingyao, Holger Dette, Zhengfu Liu, and **Jun Yu**<sup>\*</sup>. "A new approach to optimal design under model uncertainty motivated by multi-armed bandits." *Journal of the American Statistical Association* (2025).
- Li, Tao, Cheng Meng, Hongteng Xu, and Jun Yu<sup>#</sup>. "Hilbert curve projection distance for distribution comparison." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2024), DOI: 10.1109/TPAMI.2024.3363780.
- Ai, Mingyao, Zhiqiang Ye, Jun Yu<sup>\*</sup>. "Locally D-optimal designs for hierarchical response experiments." Statistica Sinica, 33 (2023), 381-399.
- 4. Yu, Jun, Xiran Meng, Yaping Wang. "Optimal designs for semi-parametric does-response models under random contamination." Computational Statistics and Data Analysis 178 (2023): 107615.
- 5. Ai, Mingyao, Yimin Huang, and **Jun Yu**<sup>\*</sup>. "A non-parametric solution to the multi-armed bandit problem with covariates." *Journal of Statistical Planning and Inference* 211 (2021): 402-413.
- Yu, Jun, Xiangshun Kong, Mingyao Ai, and Kwok Leung Tsui. "Optimal designs for dose–response models with linear effects of covariates." *Computational Statistics and Data Analysis* 127 (2018): 217-228.
- 7. Yu, Jun, Mingyao Ai, and Yaping Wang. "Optimal designs for linear models with Fredholm-type errors." *Journal of Statistical Planning and Inference* 194 (2018): 65-74.

Other topics

- 1. <u>Miao, Zhuoyi</u>, **Jun Yu**<sup>\*</sup>. "A Robust Learning Framework for Smart Grids in Defense of False Data Injection Attacks." *ACM Transactions on Sensor Networks*, (2023), DOI: 10.1145/3588439.
- 2. Li, Tao, **Jun Yu**, Cheng Meng. "Scalable model-free feature screening via sliced-Wasserstein dependency." *Journal of Computational and Graphical Statistics*, 32 (2023), 1501-1511.
- 3. Yu, Jun, Huimin Cheng, Jinan Zhang, Qi Li, Shushan Wu, Wenxuan Zhong, Jin Ye, Wen Zhan Song, and Ping Ma. "CONGO<sup>2</sup>: scalable online anomaly detection and localization in power electronics networks." *IEEE Internet of Things Journal*, 9 (2022), 13862-13875.
- 4. Cheng, Huimin, **Jun Yu**, Zhen Wang, Ping Ma, Cunlan Guo, Bin Wang, Wenxuan Zhong, and Bingqian Xu. "Details of single-molecule force spectroscopy data decoded by a network-based automatic clustering algorithm" *The Journal of Physical Chemistry B* 125 (2021), 9660-9667 (Cover paper)
- 5. Wang, Sili, Shengjie Min, **Jun Yu**, Huimin Cheng, Zion Tse, and WenZhan Song, "Contact-less home activity tracking system with floor seismic sensor network" 2021 IEEE 7th World Forum on Internet of Things (WF-IoT) (2021), pp. 13-18.
- 6. Meng, Cheng, **Jun Yu**, Jingyi Zhang, Ping Ma, and Wenxuan Zhong. "Sufficient dimension reduction for classification using principal optimal transport direction." *Advances in Neural Information Processing Systems 33* (2020).
- Ai, Mingyao, Yimin Huang, and Jun Yu. "Data-based priors for Bayesian model averaging." In Contemporary Experimental Design, Multivariate Analysis and Data Mining, Springer, Cham, (2020): pp. 357-372.

# Books

 Zhao, Ying, Dianpeng Wang, Xiangshun Kong, and Jun Yu. Mathmematical Statistics. (in Chinese) (2024). China Machine Press.

## Papers preprinted

- 1. Du, Chengshuo, Han Hao, Mengyu Li, Tao Li, Cheng Meng, and Jun Yu $^{\#}.$  "Ensemble pruning using optimal transport."
- 2. Dang, Huiping, Jun Yu, and Fasheng Sun. "D-optimal designs for ordinal response experiments."
- 3. Cheng, Huimin, Shushan Wu, **Jun Yu**, Haoran Lu, and Wenxuan Zhong. "Graphon convolutional network: A highly efficient learner for random graph."

## GRANTS, HONORS AND AWARDS

2025/01–2028/12, National Natural Science Foundation of China (Grant No. 12001042), General program, Principal Investigator

2025/01–2029/12, National Key R&D Plan of China (Grant No.2024YFA1016200), Young scholar program, co-Principal Investigator

2023/01-2025/12,Beijing Municipal Natural Science Foundation (Grant No. 1232019), General program, Principal Investigator

2021/01-2023/12,National Natural Science Foundation of China (Grant No. 12001042), Young scholar program, Principal Investigator

2019/09-2025/12, Beijing Institute of Technology research fund program for young scholars, Started-up program, Principal Investigator

2024/06 Honorable Mention in 2024 ICSA China Conference Junior Researcher Award.

2023/07 Excellent Paper Award at the 2nd International Workshop on Statistical Theory and Related Fields

2023/06 Excellent Paper Award at 2023 Academic Annual Meeting of the Uniform design (1st class)

2019/06 Excellent Poster Award at Peking-Tsinghua Joint Statistics Colloquium

2019/06 Zhong Jiaqing Outstanding Paper Award at the 23rd Beijing-Tianjin-Hebei Youth Probability Statistics Academic Conference

2018/11 Third National Forum for Ph.D. Students in Statistics Excellent Paper Award (1st class)

#### TEACHING AND MENTORING

#### Graduate courses

Advanced Theory of Statistics. 2021 Spring, 2022 Spring, 2023 Spring, 2024 Spring, 2025 Spring. General guide for data analysis. 2025 Spring.

#### Undergraduate courses

Bayesian data analysis: theory and examples. 2023 Spring, 2024 Spring, 2025 Spring. Mathematical Statistics (MOOC)

• This course is built up together with Zhao, Ying; Tian, Yubin; Wang, Dianpeng; Kong, Xiangshun.

- The corresponding textbook has been published by China machine press in 2024.
- The first batch of exemplary courses under the 2024 "Tuo Jin Plan" by the Online Education Research Center of the Ministry of Education.
- The corresponding textbook is the Textbooks under the '14th Five-Year Plan' of Beijing Institute of Technology and the Development Series for First-Class Undergraduate Majors and Courses.

### Students enrolled in Ph.D. Program

Liu, Zhengfu (2023) from BIT. Chen, Haolin (2025) from BIT.

## Graduated Master Students in Applied Statistics

Zhang, Yangyang; Yang, Xiaodan (2020)
Cao, Haoze; Lei, Kaifeng; Li, Hengli; Liu, Zhihan; Liu, Zhengfu (2021)
Peng, Hanying; Miao, Zhuoyi; Yao, Yunbao; Zeng Rui (2022)
Chen, Haolin; He, Junwen, Wang, Zixi (2023)
\* Liu, Zhengfu award 2023 outstanding graduates of Beijing Institute of Technology.

Liu, Zhengiu awaid 2020 outstanding graduates of Deljing Institute of Teenhology.

 $\ast$ Yao, Yunbao award 2024 outstanding graduates of Beijing Institute of Technology.

#### PROFESSIONAL SERVICES

#### Standing council member

2023-present Committee of Statistical Survey, CAAS.

## Council member

2023–present Committee of Multivariate anlysis, CAAS.

2021–present Committee of Uniform Design, Chinese Math Society.

2021–present Committee of Design of Experiments, CAAS.

2019–present Beijing Society for Applied Statistics.

2019–present Beijing Association for Big Data

# Organizing committee member

Symposium on design of experiments, Beijing Institute of Technology, 2021/05

Conference on design of experiments and its application, Beijing Institute of Technology (ZhuHai Campus), 2024/10

# INVITED TALKS AND PRESENTATIONS

## Invited Talks

- 1. The 2nd Joint Conference on Statistics and Data Science in China, Yunnan, China, 2024/07
- 7th International Conference on Econometrics and Statistics (EcoSta 2024) Hybrid Conference, Beijing Normal University, Beijing, China, 2024/07
- 3. International Chinese Statistical Association (ICSA) 2024 China Conference, Wuhan, China, 2024/06
- 4. 16th International conference of the ERCIM WG on computational and methodological statistics and 17th International conference on computational and financial econometrics, HTW Berlin, University of Applied Sciences, Berlin, Germany, 2023/12
- 5. Hangzhou international conference on frontiers of data science, Zhejiang University, Hangzhou, China, 2023/08
- 6. The 9th international forum on statistics, Renmin University of China, Beijing, China, 2023/07
- 7. The 2nd international workshop on statistical theory and related fields, Nankai University, Tianjin, 2023/07
- 8. Symposium on big data and intelligent processing, Peking University, Beijing, China, 2021/12
- 14th International conference of the ERCIM WG on computational and methodological statistics and 15th International conference on computational and financial econometrics, King's College London, UK, 2021/12
- 10. Symposium for Ph.D students in statistics, Guangzhou University, Guangzhou, China, 2018/11

# Contributed Talks

- 1. 2024 The 16th national symposium on survival analysis and applied statistics, Zhejiang Gongshang University, Hangzhou, China, 2024/09
- 2. 2024 Academic annual meeting of the statistical survey, Shaanxi University of Finance and Economics, Shanxi, China, 2024/09
- 3. 2023 Academic annual meeting of the uniform design, Jishou University, Hunan, China, 2023/06
- 4. 2023 Academic annual meeting of the statistical survey, Shanxi University of Finance and Economics, Shanxi, China, 2023/04
- 5. Conference on design of experiments and its application, Lanzhou University of Finance and Economics, Lanzhou, China, 2022/11, online
- 6. Conference on design of experiments and its application, Qufu Normal University, Rizhao, China, 2021/10
- 7. Conference on design of experiments and its application, Guangxi Normal University, Guilin, China, 2019/10
- 8. Peking-Tsinghua joint statistics colloquium, Tsinghua University, Beijing, China, 2019/06
- 9. Beijing-Tianjin-Hebei youth probability statistics academic conference, University of International Business and Economics, Beijing, China, 2019/06
- 10. Conference on design of experiments and its application, Beijing Institute of Technology, Beijing, China, 2018/10
- 11. Conference on design of experiments and its application, Lanzhou University of Finance and Economics, Lanzhou, China, 2017/10
- 12. Conference on design of experiments and its application, Qufu Normal University, Qufu, China, 2016/10