Aijun Zhang

[Last updated: June 2020]

Department of Statistics & Actuarial Science
The University of Hong Kong
Pokfulam Road, Hong Kong
Hong Kong SAR, China

Mobile: (852) 5565-3308
Email: ajzhang@umich.edu
Website: www.statsoft.org
ORCID: 0000-0001-9729-901

Research Interests

My main R&D interests are Experimental Design, Statistical Machine Learning, and Quantitative Risk Management. Recent projects include Automated and Interpretable Machine Learning, Big Data Subsampling, and L0-constrained Optimization for high-dimensional and nonparametric modeling.

Education

2004 – 2009	Ph.D. in Statistics, University of Michigan in Ann Arbor, USA Dissertation: Statistical Methods in Credit Risk Modeling Co-advisors: Vijayan N. Nair and Agus Sudjianto
2002 – 2004	MPhil in Statistics, Hong Kong Baptist University, Hong Kong Thesis: Majorization Methodology for Experimental Designs Supervisor: Kai-Tai Fang
1999 – 2002	BSc in Mathematics, Hong Kong Baptist University, Hong Kong
1998	Tsinghua University (Beijing) through NJCEE/GaoKao, China
1995 – 1998	Yangzhou High School of Jiangsu Province, China

Employment

2016 –	Assistant Professor (Tenure-track), Department of Statistics & Actuarial Science, The University of Hong Kong, Hong Kong
2014 – 2016	R&D Director, Center for Big Data in Education, Hong Kong Baptist University IRACE, Shenzhen; Research Assistant Professor, Department of Mathematics, Hong Kong Baptist University, Hong Kong
2009 – 2013	Sr. Quantitative Research Associate, Global Risk, Bank of America Merrill Lynch, Hong Kong
2008 – 2009	Quantitative Research Associate, Enterprise Credit Risk, Bank of America, Charlotte, North Carolina, USA

Research Grants

Current Grants

"L0 Trend Filtering with Automatic Knot Detection". Hong Kong GRF 17306519, HK$332,261, 09/2019 – 08/2021 (PI)

"Large-scale Machine Learning for FinTech Models". SXD Collaborative Grant, HK$1,000,000, 02/2019 – 01/2022 (PI)

"Artificial-Intelligence-powered Medical Imaging Analytics: from Radiomics to Deep Learning". TCL Innovative Research Fund for Supporting HKU Faculty of Science 80th Anniversary, 09/2018 – 08/2022 (PI)

Past Grants

"Machine Learning Approach to International Large-scale Mathematical Assessments". HKU Big Data Project Fund. HK$100,500, 06/2018 – 08/2019 (PI)

"Large-Scale Statistical Learning Methods in Cancer Pharmacogenomic Study". HKU Big Data Project Fund. HK$160,000, 05/2017 – 10/2018 (Co-PI)

"Dual-time Survival Analysis for Credit Risk Modeling". HKU Seed Fund for Basic Research 201611159250. HK$145,000, 03/2017 – 08/2018 (PI)

"Big Data Analytics in Online Education with Applications to Micro-Lectures and MOOCs". Hong Kong Innovation and Technology Fund ITS/268/14FX. HKD3,266,000, 05/2015 – 04/2017 (PI)

Publications

Refereed Articles

[1] Yang, Z., Zhang, A. and Sudjianto, A.(2020). Enhancing explainability of neural networks through architecture constraints. *IEEE Trans. on Neural Networks and Learning Systems*. In press.

[2] Zhang, A., Zhang, H. and Yin, G. (2020). Adaptive iterative Hessian sketch via A-optimal subsampling. *Statistics and Computing*, **30**, 1075–1090.

[3] Wen, C., Zhang, A., Quan, S. and Wang, X. (2020). BeSS: an R package for best subset selection in linear, logistic and CoxPH models. *Journal of Statistical Software*, **94**(4), June 2020.

[4] Zhang, M., Zhang, A. and Zhou, Y. (2020). Construction of uniform designs on arbitrary domains by inverse Rosenblatt transformation. In: *Fan, J., Li, G., Li, R., Liu, M.-Q. and Pan, J. (eds.) Recent Advances in Statistics and Data Science – Festschrift in Honour of Professor Kai-Tai Fang*. Springer.

[5] Zhang, A. and Yang, Z. (2020). Hyperparameter tuning methods in automated machine learning (in Chinese). *Scientia Sinica Mathematica*, **50**(5), 695–710.

[6] Yang, Z., Lin, D.K.J. and Zhang, A. (2019). Interval-valued data prediction via regularized artificial neural network. *Neurocomputing*, **331**, 336–345.

[7] Yang, F., Zhou, Y.-D. and Zhang, A. (2019). Mixed-level column augmented uniform designs. *Journal of Complexity*, **53**, 23–39.

[8] Zhu, J., Lv, K., Zhang, A., Pan, W. and Wang, X. (2019). Two-sample test for compositional data with ball divergence. *Statistics and Its Interface*, **12**, 275–282.

[9] Tao, L., Ip, H.S., Zhang, A. and Shu, X. (2016). Exploring canonical correlation analysis with subspace and structured sparsity for web image annotation. *Image and Vision Computing*, **54**, 22–30.

[10] Tang, T., Zhang, A. and Yang, X. (2014). The current situation, issues and recommendations for microlecture development (in Chinese). Expert Advice to Science and Technology Commission of China Ministry of Education, **39**.

[11] Sudjianto, A., Nair, S., Yuan, M., Zhang, A., Kern D. and Cela-Diaz, F. (2010). Statistical methods for fighting financial crimes. *Technometrics*, **52**, 5–19.

[12] Dillard, A.J., Ubel, P. A., Smith, D. M., Zikmund-Fisher, B. J., Nair, V., Derry, H. A., Zhang, A., Pitsch, R. K., Alford, S. H., McClure, J. B., Fagerlin, A. (2011). The distinct role of comparative risk perceptions in a breast cancer prevention program. *Annals of Behavioral Medicine*. **42**(2), 262–268.

[13] Nair, V., Strecher, V., Fagerlin, A., Ubel, P., Resnicow, K., Murphy, S., Little, R., Chakraborty, B. and Zhang, A. (2008). Screening experiments and the use of fractional factorial designs in behavioral intervention research. *American Journal of Public Health*, **98**, 1354–1359.

[14] Wu, Z.-L., Zhang, A., Li, C.-H. and Sudjianto, A. (2008). Trace solution paths for SVMs via parametric quadratic programming. In: *Proceedings of KDD DMMT'2008*. ACM Press.

[15] Sudjianto, A., Cela-Diaz, F., Zhang, A., Yuan, M. (2007). Anomaly detection in high-dimensional financial databases. In: *Proceedings of MLMTA'2007*. CSREA Press.

[16] Zhang, A. (2007). One-factor-at-a-time screening designs for computer experiments. *SAE Technical Paper*, 2007-01-1660.

[17] Fang, K.-T., Zhang, A. and Li, R. (2007). An effective algorithm on generation of factorial designs with generalized minimum aberration. *Journal of Complexity*, **23**, 740–751.

[18] Zhang, A., Fang, K.-T., Li, R. and Sudjianto, A. (2005). Majorization framework for balanced lattice designs. *Annals of Statistics*, **33**, 2837–2853.

[19] Zhang, A. (2005). Schur-convex discrimination of designs using power and exponential kernels. In: *Fan, J. and Li, G. (eds.) Contemporary Multivariate Analysis and Design of Experiments – In Celebration of Professor Kai-Tai Fang's 65th Birthday*, 293–311. World Scientific Publisher.

[20] Fang, K.-T. and Zhang, A. (2004). Minimum aberration majorization for non-isomorphic saturated designs. *Journal of Statistical Planning and Inference*, **126**, 337–346.

[21] Zhang, A., Wong, R.N.S., Ha, A.W.Y., Hu, Y.H., and Fang, K-T. (2003). Authentication of traditional Chinese medicines using RAPD and functional polymorphism analysis. In: *Proceedings of the 1st Conference on Data Mining and Bioinformatics in Chemistry and Chinese Medicines*, 81–98.

[22] Zhang, A., Wu, Z.-L., Li, C.-H. and Fang, K.-T. (2003). On Hadamard-type output coding in multiclass learning. In: *Liu, et al. (eds.) Intelligent Data Engineering and Automated Learning*, 397–404. Springer-Verlag.

Manuscripts Under Review

[23] Zhang, H., Yang, Z., Sudjianto, A. and Zhang, A. (2020). A sequential Stein's method for faster training of additive index models. In submission.

[24] Yang, Z., Zhang, H., Sudjianto, A. and Zhang, A. (2020). An effective and efficient initialization scheme for training multi-layer feedforward neural networks. *Neural Networks*. In revision.

[25] Kuang, K., Zhang, H., Wu, F., Zhuang, Y. and Zhang, A. (2020). Balance-subsampled stable prediction across unknown test data. In submission.

[26] Yang, Z., Zhang, A. and Sudjianto, A. (2020). GAMI-Net: an explainable neural network based on generalized additive models with structured interactions. In submission.

[27] Zhang, M., Zhou, Y.-D. and Zhang, A. (2020). Model-free subsampling method based on uniform designs. In submission.

[28] Yang, Z. and Zhang, A. (2019). Hyperparameter optimization via sequential uniform designs. *Journal of Machine Learning Research*. In revision.

[29] Wen, C., Zhu, J., Wang, X. and Zhang, A. (2020). L0 trend filtering. In submission.

[30] Wen, C., Zhang, A., Shen, Y. and Wang, X. (2019). Copy number variation detection via a constrained least-squares problem. In submission.

Manuscripts In Preparation

[31] Zhang, H., Zhang*, A. and Li, R. (2020). Least squares approximation via deterministic leveraging. In preparation.

[32] Guo, Y., Su, Y., Yang, Z. and Zhang, A*. (2020). Explainable recommendation systems by generalized additive models with manifest and latent interactions. In preparation.

[33] Zhang*, A., Zhang, M. and Zhou, Y.-D. (2020). Quasi-Monte Carlo subsampling for large-scale machine learning. In preparation.

[34] Zhang, A. and Yang, Z. (2019). UniDOE: an R package for uniform design construction via stochastic optimization. In preparation.

Statistical Softwares

[35] Python package **GAMI-Net:** Generalized Additive Model with Structured Interactions. Available at `https://github.com/zebinyang/gaminet`.

[36] Python package **ExNN:** Enhanced Explainable Neural Networks. Available at `https://zebinyang.github.io/exnn/`.

[37] Python package **SeqUD:** Sequential Uniform Designs.
Available at `https://zebinyang.github.io/SeqUD/`.

[38] R package **UniDOE:** Uniform Design of Experiments.
Available at `https://CRAN.R-project.org/package=UniDOE`.

[39] R package **AMIAS:** Alternating Minimization Induced Active Set Algorithms.
Available at `https://CRAN.R-project.org/package=AMIAS`.

[40] R package **BeSS:** Best Subset Selection in Linear, Logistic and CoxPH Models.
Available at `https://CRAN.R-project.org/package=BeSS`.

Patents

[41] "Risk and Reward Assessment Mechanism", US Patent 7,765,139 (2010); 8,326,723 (2012); 8,577,776 (2013). Assignee: Bank of America Corp.
Inventors: Agus Sudjianto; Aijun Zhang; Tony Nobili; Arun R. Pinto; Timothy J. Breault; Kaloyan Mihaylov; etc.

Selected Invited Talks

[T1] "Explainable Artificial Intelligence in Banking and Finance". Jilin University, Changchun, Dec 2019; HSBC Quant Sympsium, Hong Kong, Dec 2019; Tsinghua University, Beijing, Nov 2019.

[T2] "Least Squares Approximation via Deterministic Leveraging". The 14th Uniform Design Workshop, Soochow University, Suzhou, Nov 2019.

[T3] "A Constructive Approach to Explainable Neural Networks". International Workshop on Complex Data and Statistical Learning, Fudan University, Shanghai, Sep 2019; Big Data Challenges for Predictive Modeling of Complex Systems, The University of Hong Kong, Nov 2018.

[T4] "Data-driven Space-filling Design". The 2019 ICSA China Conference, Nankai University, July 2019; The 27th International Workshop on Matrices and Statistics, Shanghai University of International Business and Economics, June 2019.

[T5] "Machine Learning: Automation and Interpretability". China Retail Bank Artificial Intelligence Innovation Forum, Nanjing, Mar 2019.

[T6] "Machine Learning Approach to International Large-scale Mathematical Assessments". Big Data Workshop, The University of Hong Kong, Feb 2019.

[T7] "Tuning Machine Learning Algorithms via Sequential Uniform Designs". The 2018 ISI Young Statisticians Regional Workshop,Academia Sinica, Taipei, Nov 2018; The 2018 ICSA China Conference, Qingdao, July 2018; The HKMS Annual General Meeting, City University of Hong Kong, May 2018.

[T8] "AI-powered Medical Imaging Analytics via Deep Learning". Workshop on Data Science and Deep Learning (DSDL2018), The University of Hong Kong, Apr 2018; Big Data Workshop, The University of Hong Kong, Jan 2018.

[T9] "UniDOE: An R Package for Constructing Uniform Design of Experiments". Young Forum on Experimental Design, Southern University of Science and Technology, Shenzhen, Mar 2018; Young Forum on Experimental Design, Nankai University, Tianjin, Dec 2017.

[T10] "A Generalized L0 Approach to Sparse and Piecewise Smooth Modeling". International Workshop on Complex Big Data and Statistical Machine Learning. Fudan University, Shanghai, Mar 2018.

[T11] "Experience Sharing on Teaching Data Science". Conference on Teaching Statistics for Young Teachers from South China's Colleges, Sun Yat-sen University, Guangzhou, Nov 2017; Editorial Board Meeting on Modern Statistics Textbook Series by China Higher Education Press, Zhuhai, Apr 2016.

[T12] "Interval-valued Data Analysis of International Large-scale Educational Assessments". Nankai University, Tianjin, Jun 2017.

[T13] "Design-theoretic Subdata Selection for Big Data Models". Big Data Workshop, The University of Hong Kong, Dec 2016.

[T14] "Dual-time Modeling and Forecasting in Quantitative Risk Management". Sun Yat-sen University, Guangzhou, Apr 2016; The University of Hong Kong, Mar 2016; City University of Hong Kong, Feb 2016.

[T15] "Math Talk Like TED" (Joint talk with Chinese Academician Qun Lin). Asian Technology Conference in Mathematics 2015, Leshan, Dec 2015.

[T16] "Big Data Analytics in Online Education". From Industrial Statistics to Data Science – A Conference in Honor of Vijay Nair, University of Michigan, Ann Arbor, Oct 2015; Youth Statistician Forum, Hong Kong Polytechnic University,

June 2015; Tencent Da-Jiang-Tang, Shenzhen, June 2015; Hunan First Normal University, Changsha, Aug 2017.

[T17] "Microlecture Technology and Its Applications". Chongqing Normal University, Chongqing, Oct 2014.

[T18] "Uniform Design Online Platform". International Workshop on Experimental Designs, Guangzhou University, Guangzhou, Dec 2010.

[T19] "Statistical Software-as-a-Service for Statistics Education". The 6th National Forum on University Mathematics Curriculum, Fuzhou, Nov 2010.

[T20] "Statistical Applications to Financial Risk Management". BNU-HKBU United International College, Zhuhai, June 2010; Shenzhen University, Nov 2009.

[T21] "Macro-Micro Approach to Credit Risk Modeling". Bank of America Quant Technical Colloquium, Mar 2010.

[T22] "Dual-time Credit Risk Modeling". Credit Reference Center, The People's Bank of China, Beijing, Dec 2009; Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, Sept 2009.

[T23] "Peeling Algorithm in Financial Risk Analysis". Ford Motor Credit Company, Detroit, Nov 2008.

[T24] "One-factor-at-a-time Screening Designs". SAE 2007 World Congress, Detroit, Apr 2007.

[T25] "Space-Filling Designs with Minimum Energy". SAMSI 2006 Program on the Development, Assessment and Utilization of Complex Computer Models, Raleigh, Sep 2006.

[T26] "Some General Optimality Results for Fractional Factorial Designs". Oakland University, Rochester, Nov 2005.

[T27] "Theory of Schur-Power Discrimination for Orthogonal Designs under GMA Criterion". International Conference on Statistics in Honour of Professor Kai-Tai Fang's 65th Birthday, Hong Kong, June 2005.

[T28] "Minimum Aberration Majorization for Non-isomorphic Designs". Experimental Design Workshop, Academia Sinica, Taipei, Dec 2003.

Teaching

Instructor, The University of Hong Kong

STAT3612 Statistical Machine Learning (Fall 2020, Fall 2019)

STAT3612 Data Mining (Spring 2019, Spring 2018, Spring 2017)

STAT3622 Data Visualization (Spring 2020, Fall 2018, Fall 2017, Fall 2016)

Other Taught Courses

Spring 2016, Selected Topics in Statistics (Hong Kong Baptist University)

Fall 2015, Statistics in Banking and Finance (SUSTech at Shenzhen)

Fall 2014, GCNU1025 Numbers Save the Day (Hong Kong Baptist University)

Fall 2012, SCIT1020 The Power of Statistics (BNU-HKBU United International College at Zhuhai)

Summer 2007, Statistical Methods in Credit Risk Modeling (Bank of America at Charlotte, Weekly Seminar Series)

Year 2004 – 2006, STATS350 Introduction to Statistics and Data Analysis (University of Michigan in Ann Arbor, Graduate Student Instructor)

Graduate Students

GUO Yifeng, PhD student (HKU), 2019–

ZHANG Hengtao, PhD student (HKU), 2018–

YANG Zebin, PhD student (HKU), 2017–

HAN Xiaoxue, Visiting PhD student from Nankai Univ., 2018–2019

ZHANG Mei, Visiting PhD student from Sichuan Univ., 2018, 2019

ZHU Junxian, Visiting PhD student from Sun Yat-sen Univ., 2018

WEN Canhong, Visiting PhD student from Sun Yat-sen Univ., 2016–17, now an assistant professor at University of Science and Technology of China

TAO Liang, Postdoc (HKBU), 2015, now at Wisers Information Limited

QIAO Motong, Master student (HKBU), 2010–2011, now at HSBC.

Professional Services

Founding Director, HKU BASc Programme in Applied Artificial Intelligence, 2018

Associate Director, HKU Master of Data Science Programme, 2017–2018

Member, Statistics Teaching Steering Committee in Guangdong Province, 2019–

Standing Committee Member, The Uniform Design Association of China, 2016–

Editorial Board Member, Modern Statistics Textbook Series, China Higher Education Press, 2013–2018, 2018–

Organizing Committee Member, Workshop on Data Science and Deep Learning (DSDL2018), HKU, 2018

Co-Organizer, Workshop on Big Data and Spark Programming, SUSTech, 2015

Referee for IEEE Trans. on Neural Networks and Learning Systems, Annals of Statistics, Journal of the American Statistical Association, Technometrics, Computational Statistics & Data Analysis, Applied Stochastic Models in Business and Industry, Communications in Statistics –Theory and Methods, etc.

PhD Thesis Committee Member for Yao Zheng (HKU-SAAS), Fanghu Dong (HKU-SAAS), Xifen Huang (HKU-SAAS), Zhen Li (HKU-CS).

Personal Data

Born in Taixing, Jiangsu, China.

Married, with three lovely kids.

Amateur math microblogger and vlogger.