

BOYANG SHANG

✉ BoyangS@u.northwestern.edu ☎ +852 3917-2466 🌐 <https://boyangshang.github.io/>

RESEARCH INTERESTS

Statistical modeling; design of computer experiments; subsampling and large-scale data reduction; copula modeling and estimation; variable selection.

ACADEMIC APPOINTMENTS

The University of Hong Kong, Hong Kong SAR, China
Postdoctoral Fellow, Department of Statistics and Actuarial Science Mar 2024 – present
Advisors: Prof. Marius J. Hofert

City University of Hong Kong, Hong Kong SAR, China
Postdoctoral Fellow, Department of Data Science Nov 2022 – Nov 2023
Advisors: Prof. Lishuai Li and Prof. S. Joe Qin

EDUCATION

Northwestern University, Evanston, IL, USA Sep 2022
Ph.D. in Industrial Engineering and Management Sciences
Dissertation: “Space-Filling Designed Sampling from Databases”
Advisors: Prof. Daniel W. Apley and Prof. Sanjay Mehrotra
Selected coursework: Mathematical Statistics; Predictive Analytics; Stochastic Processes and Simulation; Mathematical Optimization; Stochastic Optimization

Northwestern University, Evanston, IL, USA Jun 2016
M.S. in Engineering Sciences and Applied Mathematics
Selected coursework: Differential Equations of Mathematical Physics; Asymptotic and Perturbation Methods; High-Performance Scientific Computing; Numerical Solutions of Partial Differential Equations; Models in Applied Mathematics

Beihang University, Beijing, China Jun 2015
B.S. in Mathematics and System Science

PUBLICATIONS

Peer-Reviewed Journal Articles

Shang, B. and Apley, D. (2020). Fully-sequential space-filling design algorithms for computer experiments. *Journal of Quality Technology*, 53(2), pp. 173–196. DOI: 10.1080/00224065.2019.1705207.

Shang, B., Apley, D. and Mehrotra, S. (2023). Diversity subsampling: Custom subsamples from large data sets. *INFORMS Journal on Data Science*, 2(2), pp. 161-182. DOI: 10.1287/ijds.2022.00017.

Preprints and Working Papers

Shang, B. (2023). On convergence rate of the diversity subsampling algorithm. *arXiv preprint* arXiv:2309.00636. Available at <http://arxiv.org/abs/2309.00636>.

Shang, B. and Hofert, M. MMD-based estimation of copula densities by mixture-induced copulas. *Manuscript in preparation*.

Patents

Shang, B., J. Fan, Q. Ye, Y. Yao and X. Ding. A system of human-machine emotional interaction based on facial expression recognition. *National Patent of China*, No. ZL 2013 1 0413648.7.

Software Packages

Shang, B. and Apley, D. FSSF: Fully-sequential space-filling designs for computer experiments. R package on CRAN, <https://CRAN.R-project.org/package=FSSF>.

Shang, B., Apley, D. and Mehrotra, S. FADS: Custom subsampling from large data sets. Python package on PyPI, <https://pypi.org/project/FADS/>.

RESEARCH EXPERIENCE

The University of Hong Kong, Hong Kong SAR, China
MMD-based Estimation of Copula Densities by Mixture-Induced Copulas Mar 2024 – present
Joint work with Prof. Marius J. Hofert.

- Proposed a method to estimate the copula density corresponding to the unknown probability distribution underlying an identically and independently distributed sample, assuming the true model is the copula of a mixture model and using maximum mean discrepancy (MMD) as the fitting criterion.

- Showed theoretically that, under either MMD or Kullback–Leibler divergence, the copula of a mixture model can approximate a general copula density up to an error term that shrinks as the number of mixture components grows.

City University of Hong Kong, Hong Kong SAR, China

Time-Dependent Mixture Probabilistic Principal Component Analysis

Nov 2022 – Nov 2023

Joint work with Jicheng Li, Prof. Lishuai Li and Prof. S. Joe Qin.

- Developed a time-dependent Mixture Probabilistic Principal Component Analysis (MPPCA) model for fault detection in multivariate time-series data.

Northwestern University, Evanston, IL, USA

Diversity Subsampling: Custom Subsamples from Large Data Sets

Sep 2016 – May 2022

Joint work with Prof. Daniel W. Apley and Prof. Sanjay Mehrotra.

- Proposed a novel density-based subsampling algorithm for selecting diverse subsamples from large databases, achieving roughly 20-fold speedup over its most competitive competitor.
- Generalized the approach to select subsamples following user-specified distributions known up to a normalizing constant.
- Investigated the computational efficiency, estimation accuracy and robustness of about 10 multivariate density estimation methods.
- Results published in *INFORMS Journal on Data Science* and implemented in the Python package **FADS**.

Large-Scale Fully-Sequential Space-Filling Algorithms for Computer Experiments

Apr 2016 – Mar 2018

Joint work with Prof. Daniel W. Apley.

- Proposed a heuristic for the minimax design problem with quadratic-time complexity.
- Reduced the theoretical time complexity of an existing maximin design algorithm from quadratic to near-linear (“linearithmic”), achieving about 100-fold speedup in practice.
- Demonstrated numerically that the proposed minimax heuristic yields designs with smaller integrated mean squared error (IMSE) under a Gaussian process model.
- Showed mathematically that, under a Gaussian process model, solutions to the minimax design problem approximate minimum-IMSE designs as the design size increases and the correlation function approaches a delta function.
- Results published in the *Journal of Quality Technology* and implemented in the R package **FSSF**.

Chinese National Undergraduate Training Program for Innovation and Entrepreneurship, Beijing, China

Human-Machine Emotional Interaction

Jun 2012 – Dec 2013

Supervised by Prof. Xilun Ding; joint work with Jie Fan, Yongqiang Yao and Qianwen Ye.

INDUSTRY EXPERIENCE

Amazon.com, Seattle, WA, USA

Data Scientist Intern

Jun 2019 – Sep 2019

- Proposed an algorithm to predict summary statistics of simulation outputs using a limited number of simulation runs.

TEACHING EXPERIENCE

Northwestern University, Evanston, IL, USA

Teaching Assistant

Courses: Applied Mathematical Statistics; Predictive Analytics; Introduction to Health Systems Engineering; Quality Improvement by Experimental Design; Service Operations Management.

TECHNICAL SKILLS

Programming: Python, R, C++, C

High-Performance & Parallel Computing: MPI (C, Python); CUDA (C)

Mathematical & Optimization Software: AMPL, Matlab, Mathematica

Scientific Writing & Typesetting: L^AT_EX

Data Management: MySQL

REFERENCES

Prof. Daniel W. Apley
Northwestern University
Evanston, IL, USA
Ph.D. supervisor
☎ +1 847 491 2397
✉ apley@northwestern.edu

Prof. Sanjay Mehrotra
Northwestern University
Evanston, IL, USA
Ph.D. supervisor
☎ +1 847 491 3155
✉ mehrotra@northwestern.edu

Prof. Marius J. Hofert
The University of Hong Kong
Hong Kong SAR, China
Postdoctoral supervisor
☎ +852 3917 8138
✉ mhofert@hku.hk