Peng Xu

px2132@columbia.edu • +1 (217) 419-9268
Homepage: francis-hsu.github.io

EDUCATION

Columbia University in the City of New York

M.A. in Statistics   Sep 2017 – Dec 2018
- Cumulative GPA: 3.999

University of Illinois at Urbana-Champaign

M.S. in Financial Engineering   Aug 2015 – Dec 2016
- Cumulative GPA: 3.91

B.S. in Finance   Aug 2010 – Jun 2015

B.S. in Mathematics and Statistics
- Cumulative GPA: 3.74
- Triple-major
- Graduated with Honors in Finance, High Distinction in Mathematics, and High Distinction in Statistics

RESEARCH INTERESTS

High-dimensional statistics; Non-parametric methods; Optimal transportation; Machine learning algorithms; Application of statistics in finance.

RECENT RESEARCH PROJECTS

- **Multivariate Ranks and Quantiles using Optimal Transportation**   Apr 2019 –
  - Mentor: Bodhisattva Sen
  - We studied the properties of multivariate ranks and quantiles defined through semi-discrete optimal transportation, and proposed applications in goodness-of-fit testing based on these notions. We further created a software package for computing and visualizing the statistics we defined.
  - Associated R-package: testOTM.

- **Approximate Leave-One-Out Cross-Validation for High Dimensional Models**   Jul 2018 –
  - Mentor: Arian Maleki
  - Leave-one-out cross-validation (LOOCV) is a popular choice of tuning method in machine learning. Its high computational cost makes it less feasible in application however. We proposed efficient approximation methods of LOOCV for popular classes of learning scheme, and implemented them for standard models, such as kernel SVM and generalized elastic net.
  - Associated R-package: alocv.

- **Counting Process Based Dimension Reduction for Survival Analysis**   Apr 2017 –
  - Mentor: Ruoqing Zhu
  - We studied a dimension reduction scheme based on counting process for right-censored survival data. The dimension reduction subspace is recovered through a semi-parametric estimating equations. We implemented the proposed formula with constraint optimization algorithm on Stiefel manifold, and investigated the effect of regularization in improving the solution sparsity and overall performance.
  - Associated R-package: orthoDr.

PUBLICATIONS

JOURNALS

- R. Zhu, J. Zhang, R. Zhao, **P. Xu**, W. Zhou, and X. Zhang. *orthoDr: semiparametric dimension reduction via orthogonality constrained optimization*. The R Journal. Jul 2019.

CONFERENCES AND WORKSHOPS

- W. Zhou, S. Wang, **P. Xu**, H. Lu, A. Maleki, and V. Mirrokni. *Approximate Leave-One-Out for Fast Parameter Tuning in High Dimensions*. *2019 INFORMS Data Mining and Decision Analysis Workshop*, Oct 2019. (Runner-up of the Best Paper Competition, Theoretical Track).

PREPRINTS

- A.L. Turner, A. Uppal, and **P. Xu**. *Spacing Distribution of a Bernoulli Sampled Sequence*. Submitted to *INTEGERS*. Oct 2015, arXiv:1510.03500 [math.PR].

SOFTWARE

- testOTM: an R package that computes multivariate ranks and quantiles defined through the theory of optimal transports. It also provides several applications of these statistics, most notably the two-sample multivariate goodness-of-fit testing. The core components are written in C++.
- alocv: an R package that implement the approximate leave-one-out cross validation strategy for common regressors in an efficient way.

OTHER ACADEMIC PROJECTS

- **Course Development: FinTech and Data Driven Innovation**   Jul 2019 –
  - Mentor: Margaret Holen
  - FinTech and Data Driven Innovation is a classto be offered at Columbia. As an RA, I helped in overhauling the statistics and machine learning related components of the class. My duties includes periodic meeting with the instructor, writing relevant lecture notes, problem sets, as well as developing python scripts and Jupyter notebooks to improve class content deliveries. Topic discussed in this class includes basic finance, estimation and inference, classification models, e.g. logistic regression and ensemble trees, and fairness.

- **Option Pricing with Variance-Gamma Model**   Aug 2015 – Dec 2016
  - Mentor: Liming Feng
  - We implemented a Hilbert transformation based algorithm for evaluating options under Variance-Gamma model. We further analyzed the effect of additional diffusion term in the Variance-Gamma process and its implication on asymptotic behavior.

- **Random Subsets of Discrete Sets**   Jan 2014 – Dec 2015
  - Mentor: Jayadev S. Athreya, Francesco Cellarosi
  - This is a research project from the Illinois Geometry Lab. In this project we inspected the properties and asymptotic behaviors of random subsets of the Farey sequence, and determined the limiting distribution of spacings of a Bernoulli sampled equidistributed sequences.

- Forecasting Mortgage Default Rate with Statistical Learning Methods
  - Sponsor: AXIS Re
  - This is a Practicum research project, and part of the MSFE program. In this project we collected and cleansed large data-set published by Fannie Mae and Freddie Mac on individual mortgage loans, and created predictive models for default prediction. Multiple classifiers were investigated, such as Bayesian penalized regression, random forests, and linear SVM.

WORK EXPERIENCES

Columbia University, New York, New York

- Research Associate, Department of Statistics   Jun 2019 –
- Teaching Assistant, Department of Statistics   Aug 2018 – Dec 2018, Jun 2019 –
  - Instructor: Banu Baydil

CITIC Securities Company Limited, Shenzhen, China

- Intern, Research Department   Jul 2015 – Aug 2015

China Merchants Bank New York Branch, New York, New York

- Intern, Banking Department   Jun 2013 – Jul 2013

COMPUTER SKILLS

- Languages: C++, Python, R.
- Software: LaTeX, Mathematica, MATLAB, Microsoft Office.