

LINQUAN MA

+1 608 736 0567 ✧ mlqmlq1997@gmail.com

1300 University Avenue, Madison, WI, 53706, U.S.

Personal website: <https://mlqmlq.github.io>

EDUCATION

Department of Statistics, University of Wisconsin-Madison

Aug. 2017 - May 2019

M.S. in Data Science.

GPA: **4.0/4.0**; Rank: **1/42** (out of all students in my program).

Core Courses: Statistical Inference I, II, Linear Models, Design and Analysis of Experiments, Time Series Analysis, Non-parametric Statistics, Classification and Regression Trees.

Ongoing Courses: Mathematical Statistics I (Ph.D. level), Data Science Computing Project.

School of Gifted Young, University of Science and Technology of China Aug. 2014 - May 2018

B.S. in Statistics.

Major GPA: **3.9/4.3**; Rank: **Top 5/51** (out of all students studying statistics at School of Gifted Young).

Core Courses: Regression Analysis, Applied Statistical Software, Probability Theory, Mathematical Statistics, Stochastic Processes, Real Analysis, Complex Analysis, Functional Analysis, Mathematical Analysis I, II, III, Linear Algebra I, II.

RESEARCH INTERESTS

Sufficient dimension reduction, multivariate regression, causal inference and missing data analysis.

PUBLICATIONS

- **Ma, L.**, Yin, Y., Liu, L., Geng, Z. (2017). On the individual surrogate paradox. *Invited to submit revision, Biostatistics*. Preprint version: <http://arxiv.org/abs/1712.08732>
- **Ma, L.**, Liu, L., Yang, W. (2018). Envelope method with ignorable missing data. *Submitted to JASA*. Slides: <https://mlqmlq.github.io/envelope.pdf>
- Shi, Y., **Ma, L.**, Liu, L.. Efficient multivariate mixed regression in task fMRI study. *In manuscript*.

RESEARCH EXPERIENCE

On the Individual Surrogate Paradox

Apr. 2017 - Dec. 2017

Adviser: Lan Liu, School of Statistics, University of Minnesota

- Proposed a new method to evaluate the surrogate paradox based on an individual perspective.
- Proved the individual surrogate paradox can be excluded under certain criteria.
- Derived the sharp bounds of heterogeneity effect by applying simplex algorithm on the dual problem.
- Performed real data analysis on a clinical trial data, and built an R package called *HRsurrogate*.

Envelope Method with Ignorable Missing Data

Oct. 2017 - Nov. 2018

Adviser: Lan Liu, School of Statistics, University of Minnesota

- Developed a method for envelope estimation with both the predictors and the responses missing at random.
- Derived the parameter updates in the EM algorithm using advanced linear algebra techniques.
- Proved the superiority of the EM envelope over the standard EM in the sense of asymptotic variance.
- Confirmed the efficiency gain of our method in the simulations.
- Applied our method to the Chronic Renal Insufficiency Cohort (CRIC) study. Discovered two new significant biomarkers and got shorter confidence intervals using bootstrap.

Efficient Multivariate Mixed Regression in Task fMRI Study

Aug. 2018

Adviser: Lan Liu, School of Statistics, University of Minnesota

- Realized the proposed method in R code, and discussed the the selection of envelope dimension.
- Helped with performing real data analysis on the fMRI data set.

Max Independent Component Analysis (In-progress)

May 2018 - Present

Adviser: Zhengjun Zhang, Department of Statistics, University of Wisconsin-Madison

- Recovering independent components in the existence of max operator in the linear system.
- Applied fast Fourier transformation to analyze signals in the frequency domain.
- Proposed a method of recovering the mixing matrix $\mathbf{A}_{p \times p}$ by the joint distribution when $p = 2$.
- Working on generalizing the method when $p \geq 3$.

COURSE PROJECTS

Machine Learning with Large Scale Data Set Containing Missing Value Oct. 2017 - Dec. 2017

Adviser: Wei-Yin Loh, Department of Statistics, University of Wisconsin-Madison

- Handled high-dimensional National Birth Rate data set (several GB) through the server to predict whether a new born infant is underweight.
- Imputed missing values using MICE and GUIDE.
- Assessed various machine learning algorithms: logistic regression, regression trees, random forest, SVM, LDA, GUIDE, XgBoost and ensemble learning.
- Selected the best model by cross validation, and showed that ensemble learning outperforms other methods.

Statistical Inference on the Gene Expression

Dec. 2017

Adviser: Zhengjun Zhang, Department of Statistics, University of Wisconsin-Madison

- Applied cross validation for tuning the lasso penalty parameter.
- Utilized Generalized Measures of Correlation (GMC) method for model selection.
- Detected the global maxima of the non-convex objective function by grid search of the starting points.

AWARDS

Student Academic Excellence Award, UW-Madison

2017, 2018

Outstanding Student Scholarship, USTC

2014, 2015, 2016, 2017

SKILLS & LANGUAGE

Programming: R, \LaTeX , Matlab, Bash, Python, C++.

TOEFL: 28(R) + 30(L) + 24(S) + 27(W) = 109.

GRE: Verbal 152; Quantitative 166; Writing 3.0.

MISCELLANEOUS

Talk: Presented the envelope method in group meeting.

TA Experience: Hold office hours for the graduate level course *Statistical Inference*.

Fall 2018

Seminars: Got familiar with high-dimensional statistics, randomized linear algebra, non-linear optimization, graphical models, network analysis and deep learning.