Zifan (Fred) Yu

66 N. Pauline Street, Suite 303 – UTHSC – Memphis, TN
+1 2406443948 • zyu20@uthsc.edu • learningMalanya.github.io

Current scientific research programmer at the University of Tennessee with educational backgrounds in Mathematics and Statistics. I am passionate about developing methods using theories in statistics and modern scientific computing to uncover the truths of real-life problems.

Education

Master of Science in Biostatistics 3.7/4.0
University of Washington (UW), Seattle *2020-2022*

Bachelar of Science in Mathematics 3.8/4.0
University of Maryland (UMD), College Park *2016-2020*

Research Experience

Fast Multiple Trait Genome Scans Using a Linear Mixed Model Approach with Permutation Tests
Mentor: Saunak Sen, University of Tennessee Health Science Center (UTHSC) *Ongoing*

- Developed a linear mixed effects model framework for modeling the non-independence of gene-expression data due to genetic relatedness.
- Programmed the method in to efficient *Julia* code; employed techniques of linear algebra theories, vectorized operations and multi-threading for speeding up the computations.
- Based on our experiments on the BXD mouse strains data, our program performs **real-time (0.05s)** genome scans for a single trait with 1000 permutations; for multivariate traits GWAS, our program approximates the approach and finishes the scans of all 35k traits in **15s** with multithreading.
- Investigated and introduced a **Bayesian regularization approach** to fix a numerical instability issue due to the unexpected nature of the optimizer near one boundary of the parameter-space.
- Authoring the package ***BulkLMM*** which will soon be registered on GitHub; preparing the manuscript and will be presenting in ENAR 2023.

Applications of Nowcasting Methods to Notifiable Disease Surveillance
Mentor: Ian Painter, Washington Department of Health *Sep. 2021 - Mar. 2022*

- Conducted a literature review on the nowcasting methods to predict daily hospitalization count in Washington State due to COVID-19 and constructed a statistical framework suitable for our problem.
- Adapted an existing Bayesian hierarchical model for modeling the variable timings of reporting delays.
- By comparing the predicted count retrospectively in time with the actual case count, our proposed method precedes the current WADOH approach for about **eight days** to give accurate predictions.

Sparse Matrix Linear Models (MLM) with the Elastic-net Regularization
Mentor: Gregory Farage & Saunak Sen, UTHSC *June 2021 - Sep. 2021*

- Investigated the proximal gradient methods and derived the formulas for the proximal estimators of MLM under the Elastic-net regularization.
- Implemented the proximal solutions in the Julia package ***MatrixLMnet*** and conducted the performance analysis of our program under different simulation settings.
- Confirmed based on simulation results that the Elastic-net regularized model outperformed the LASSO regularized model in selections of important variables and predictions, under the cases of high multicollinearity and large dimensionality of the covariates.

Variable Importance for Fixed Effects in Linear Mixed Models (Presented at WNAR 2022)
Student paper co-authored with Yongzhe Wang & Lingbo Ye, UW *Spring 2022*

- Adapted the Shapley values concept from "Explainable AI" for the variable importance measurements of fixed effects in linear mixed models.
- Implemented our method in R and conducted a sensitivity analysis to study the performance of our method under various simulation settings.
- Concluded from the simulation study results that our proposed measurements captured the **true variable importance**, and was robust for the **high-muticollinearity** case where the traditional p-values performed undesirably.

Review of Approximation Methods for Large Sparse Precision Matrices
Independent reading project for the Stochastic processes class, UW *Spring 2022*

- Reviewed estimation methods for large precision matrices of Gaussian-Markov Random Field random variables proposed in *Efficient Covariance Approximations for Large Precision Matrices* by P. Siden, et. al. .
- Analyzed the algorithmic complexity of each method and discussed the improvements of the proposed Blocked Rao-Blackwellized Monte Carlo method over the traditional Takahashi Equations.

Differential Equations Modeling of Fish Migration Patterns (MCM Meritorious Award)
Team project, Mathematical Contest in Modeling (MCM) *Spring 2020*

- Modeled the fish-environment interactions of Scottish economic fish species using differential equations.
- Proposed a fishing business schedule based on the fish migration patterns learnt from our model.
- Our team was ranked among the **top 8%** of a total 3851 teams.

Software

- **BulkLMM.jl**: https://github.com/senresearch/BulkLMM.jl
- **MatrixLMnet.jl**: https://github.com/senresearch/MatrixLMnet.jl

Other Experience

Survival Analysis Working Group
Department of Biostatistics, UW *Summer 2021*

- Attended bi-weekly discussion sessions on Survival Analysis following the lecture notes of Prof.Chan, Gary.
- Led and presented at the session about the EM algorithm for estimating the Keplan-Meier as non-parametric MLE, and the Greenwood's formula estimation of Keplan-Meier estimator's standard error.

Linear Algebra Tutor
Department of Mathematics, University of Maryland *Fall 2019*

- Held weekly office hours to assisted students with their conceptual understanding and Matlab programming skills for two undergraduate-level Linear Algebra classes.

Directed Reading on Markov Chains
Directed Reading Program (DRP), Department of Mathematics, University of Maryland *Spring 2019*

- Read about discrete-time Markov chains (*Theory of probability and random processes* by L. Koralov, et. al.).
- Discussed questions at weekly meetings with the mentor and presented findings at the DRP seminar.

Skills

- **Programming Languages:** Julia, R, Matlab, Java, Python
- **Others:** Distributed version control (Git), parallel computing, Linux environments, LaTeX