DAI, DEHAO

3869 Miramar Street #4232, La Jolla, CA, 92037 ·(858)-291-3647 · ddai@ucsd.edu

EDUCATION

University of California, San Diego — La Jolla, CA
Master of Science, Statistics — **GPA**: 3.881/4.0 — 10/2020-Present

- **Qualifying Exam Results**: **MATH 270** Numerical Analysis, **MATH 281** Statistics **(All PhD pass)**

Donghua University — Shanghai, China
Bachelor of Science, Mathematics and Applied Mathematics — **GPA**: 4.33/5.0 — Graduated: 07/2020

- **Honors & Awards**: Excellent Graduate in Shanghai, China (Top 1%); National Scholarship (Top 1%); Third Prize in the 2018 National College Student Mathematics Competition

RESEARCH EXPERIENCES

Reading and Research
Graduate Research MATH 299 — *Supervisor:* **Prof. Jelena Bradic** *and* **Prof. Ronghui Xu** — 06/2021-Present
Topic: ***Causal Inference under High-Dimensional Data using Semiparametric Theory***
Codes: https://github.com/DehaoDai/Notebook-of-paper-by-Tian

- Stimulated the data with continuous and binary response by R, modified the original iteration condition of LASSO to reduce the variance of Spearman's rank correlation coefficient under the models with two difference responses
- Studied some important approaches in the Semiparametric Theory like Augmented inverse probability weighting (AIPW) estimator and corresponding applications
- Explored the consistency and normality of estimators using Central Limit Theorem, Cramer-Wald Device and some important inequalities from the course Mathematics Statistics (MATH 281)
- Considered time-varying factors integrated to estimate the average causal excursion treatment effect and tried to obtain an estimator for truncated-time effect by calculating the proposed estimating function

Comparison among Different Manifold Learning Algorithms and Application on Classification
Graduate Course MATH 287B Assignment Report — *Supervisor:* **Prof. Ery Arias-Castro** — 01/2021-03/2021
Codes: https://github.com/DehaoDai/Manifold-Learning

- Generated three-dimensional data following from normal distribution and t-distribution via R, compared the results and running time of dimensional reduction to two dimensions under different algorithms in Manifold Learning and traditional Principle Component Algorithm (PCA) respectively
- Collected about 2000 image data of one person's face under different angles from websites via Python and reduced the original dimension to two and three under different algorithms in Manifold Learning and traditional Principle Component Algorithm (PCA) respectively
- Interpreted the two-dimensional or three-dimensional components through image itself, classified faces under different angles and compared the error and Running time

Law of Large Numbers for the Linear Self-interacting Diffusion Driven by α-Stable Motion
Undergraduate Capstone Project — *Advisor:* **Prof. Litan Yan** — 09/2019-06/2020
Description: In this research, the specific asymptotic behavior of the linear self-interacting diffusion was discussed and extended to $\theta > 0$ and $\theta < 0$ with two rates of convergences.

- Awarded Excellent Graduate Capstone (Thesis) in 2020 Donghua University (Top 3%)
- Extended the linear self-interacting diffusion driven by Brownian motion, and considered the linear self-interacting diffusion by α-stable motion based on a stochastic differential model for the shape of a growing polymer
- Introduced law of large numbers associated with the process with $1 < \alpha < 2$ and $\theta > 0$ based on the result provide by Xichao Sun and Litan Yan, and obtained the results under different convergence speeds
- Constructed the almost surely and L^p convergence under the situation $\theta < 0$ and extended the studied stochastic processes from X. Sun and L. Yan

Parameter Estimation of Yield Distribution of a Class of Stocks Based on the Characteristic Function Method
Undergraduate Research Project — *Team Member, Advisor:* **Prof. Zhenzhong Zhang** — 09/2019-07/2020

- Proposed to treat the stocks' daily rate of return as a type of censored truncated stable random variable
- Discussed the properties of non-truncated stable distribution and simulated the non-truncated stable random number by Chambers, Mallows and Stuck Algorithm (CMS) via MATLAB
- Constructed the characteristic function expression of truncated stable distribution, estimated four parameters of interest, deduced the skewness and kurtosis function expression by inverse Fourier Transform method and estimated the truncated stable distribution of stocks' daily rate of return by Kernel Density Estimate Method (KDE)

- Compared the result under non-truncated stable distribution and truncated one through comparing P-P plots, calculating the total variation and KS(Kolmogorov-Smirnov) test method, and concluded the suitability of the truncated stable distribution in the financial market

Seminar on Probability Theory
Undergraduate Summer Seminar *Supervisor:* **Prof. Litan Yan** *and* **Prof. Zhenzhong Zhang** 07/2019-08/2019
Topic1: ***Parameter Estimation of some Stochastic Differential Equations (SDE)*** *host by Prof. Litan Yan*
Topic2: ***Properties of Stable Distribution and its Applications*** *host by Prof. Zhenzhong Zhang*

- Invited two professors to organize the bi-weekly undergraduate summer seminars
- Discussed about similarities and difference between the equivalent Partial Differential Equations forms of SDEs driven by different processes like Brownian motion, fractional Brownian motion and α-stable motion
- Studied how to solve the SDEs driven by (fractional) Brownian motion by Itô integral and developed Itô integral method to solve the SDEs driven by α-stable motion
- Considered the existence of n^{th} moment of stable distribution under different parameters, and proved proprieties of stable distribution and simulated the random number via MATLAB

The Driving Behavior Evaluation Based on K-means Classification Method
2019 DHUMCM Contest *Team Leader, Supervisor:* **Prof. Liangjian Hu** 03/2019-05/2019

- Won the First Prize of 2018 Donghua University Contest in Mathematical Modeling
- Collected 450 sets of driving data provided by Transportation Research Center via Python, corrected the inaccurate data such as the longitude and latitude and normalized the main characteristics
- Established the Evaluation Index System including the indicators selected in the characteristic of vehicle: longitude, latitude, time, speed, mileage and the weather conditions by Principal Component Analysis (PCA)
- Proposed the K-means Clustering Algorithm to change the parameters, group number k, clustered the 450 drivers into 4 different groups based on the score and perfected the evaluation system with the Structural Equation Model

Energy Usage Prediction and Analysis, and its Feasible Method
2018 MCM Contest *Team Leader, Supervisor:* **Prof. Liangjian Hu** 02/2018

- Won the Honorable Mention in 2018 American Mathematical Contest in Modeling
- Established an energy profile and a Back Propagation Neural Network with time series to predict the energy usage situation in 4 different states in 2025 and 2050, and proposed the entropy weight method to compute the parameter to evaluate the best energy profile based on the criteria of the energy profile for cleaner and renewable energy
- Proposed to transport different kinds of the energy among these states via Lingo in order to balance the energy usage and production after discussion of a linear optimization, and discovered the model stability with the Sensitivity Analysis (SA) based on the changing 1% of production and consumption

Statistical Analysis on Human Body Sleep Quality
Undergraduate Course Assignment Project *Leader, Supervisor:* **Prof. Liangjian Hu** 11/2017

- Computed the correlations between the indicators and sleep quality by SPSS, and concluded Age, Psychoticism, Nervousness and Character related to sleep quality
- Qualified the diagnosis to the number from 1 to 43, used the techniques of Discriminant Analysis to establish the relationship between the diagnosis results and sleep, and predicted 6 patients' diagnosis based on the given data
- Developed a sleep program based on the indicators affected the sleeping quality and other sleeping indicators to decrease the morbidity and combined Discriminant Analysis with the machine learning to decrease the misjudgment

INTERNSHIP EXPERIENCES

Shanghai Smallville Asset Management Co., Ltd.
Investment Assistant 07/2018-08/2018

- Separated the target customers' predilection into three parts: the customer scale (including age, city, and mobile phone system), gender structure and using period by Python's Correlation Analysis, and scored their preference to 1-10 to propose that game companies should cater to the taste of the young males by changing the game feature
- Discovered the relationship between users' downloads, companies' cost, equity, and PE ratio with one year after report publication to infer the game companies' release threshold
- Predicted the uptrend in the stock prices based on the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model in order to demonstrate the shareholders' confidence in the market and improved the accuracy by simulating the stock prices with Ornstein-Uhlenbeck Process