Daniel **Iong**

STATISTICS PH.D CANDIDATE

☎ (+1) 510-816-8686 | ✉ daniong@umich.edu | ⊙ danieliong | in danieliong | 🐦 @IongDaniel | US Citizen

Education

University of Michigan, Ann Arbor (UM)   *Ann Arbor, MI*

PH.D IN STATISTICS   *2017 – Present*

- Rackham Merit Fellow

University of California, Davis (UCD)   *Davis, CA*

B.S. STATISTICS (HIGH HONORS), B.A. ECONOMICS (HONORS), CHINESE MINOR   *2013 – 2017*

- Outstanding Academic Performance Citation (Dept. of Statistics), Dean's List

Experience

NASA Goddard Space Flight Center (Solar Physics Laboratory)   *Virtual*

RESEARCH INTERN (SUPERVISOR: CHARLES N. ARGE)   *Jun. – Aug. 2021 (10 weeks)*

- Identified novel application of dynamic time warping for model evaluation of a theoretical solar wind model in collaboration with space physicists.
- Created web app in Python using Dash, Plotly to communicate results (`https://wsa-dtw.herokuapp.com/`).

Department of Statistics, UM   *Ann Arbor, MI*

GRADUATE STUDENT INSTRUCTOR   *Sept. 2018 – Apr. 2020 (3 semesters)*

- **Courses**: Intro. to Statistics & Data Analysis (undergraduate), Bayesian Data Analysis (undergraduate), Bayesian Modeling & Computation (graduate)
- Prepared lectures on advanced topics in Bayesian modeling (mixture models, model checking, etc).
- Advised students on extensive projects in applied Bayesian analysis.
- Taught weekly labs on basic statistical concepts, held weekly office hours to answer homework questions.

Projects

Explainable machine learning for space weather forecasting

SOLAR STORMS & TERRESTIAL IMPACTS (SOLSTICE) CENTER, UM   *Feb. – Sept. 2021*

- Trained gradient boosted trees (XGBoost) to predict high resolution geomagnetic index (SYM-H) several hours ahead in Python, resulting in a 10% lower RMSE compared to the best existing forecasting methods in the space weather literature.
- Collaborated with space scientists to explain predictions using explainable ML methods, leading to novel insights about underlying physics.
- Created web app in Python using Dash, Plotly to visualize results (`https://geomag-interpret.herokuapp.com/`).
- **Paper**: `https://doi.org/10.1002/essoar.10508063.1` (*submitted to Space Weather*)

Modeling heterogenous causal mechanisms in epidemiology with observational data

DEPARTMENT OF STATISTICS, UM   *May 2019 – Aug. 2020*

- Developed novel probabilistic clustering method to model causal mechanisms between HDL cholesterol and coronary heart disease.
- Implemented Monte-Carlo EM algorithm in R/C++ (Rcpp) to perform statistical inference (parameter estimation, confidence intervals, model selection).
- **R package**: `https://github.com/danieliong/MRPATH`
- **Paper**: `https://arxiv.org/abs/2007.06476` (*submitted to Annals of Applied Statistics*)

MagNet - Model the Geomagnetic Field

SOLSTICE CENTER, UM   *Jan. – Feb. 2021*

- Ranked **top 5%** (32/623) in competiton hosted by NOAA (1st place prize: $15,000) to forecast the Disturbance Storm-Time (DST) index under operationally viable constraints (`https://www.drivendata.org/competitions/73/noaa-magnetic-forecasting/leaderboard/`).
- Preprocessed real-time solar wind data in collaboration with space physics Ph.D student.
- Trained various models including gradient boosted trees, feed-forward/long-short term memory neural networks in Python.

Predicting dynamics for functional data with applications in economics

DEPARTMENT OF STATISTICS, UCD   *Mar. – Dec. 2016*

- Implemented extension of existing empirical dynamics model to include covariates in R, which improved coefficient of determination for predicting quarterly US GDP by 40%.

Time series analysis of pesticide concentrations

DEPARTMENT OF STATISTICS, UCD   *Sept. – Dec. 2016*

- Implemented EM algorithm in R to fit state-space model to pesticide concentrations data containing missing and censored values.

Skills

Programming Languages	Python, R (Rcpp), Julia, C++, SQL
Data Science Tools	Numpy, Pandas, Scikit-learn, XGBoost, Tensorflow, PyTorch, Stan, Cython, ggplot, Matplotlib, Plotly, Dash
Data Science Methods	Bayesian/probabilistic modeling, time series forecasting, causal inference, statistical computing, deep learning
Computing Tools	Shell scripting, Linux (Ubuntu, Arch), Version Control (Git), High Performance Computing (Slurm)
Languages	English (native), Cantonese, Mandarin