Tianyi (Tai) Cai

Website: padtai.ca
LinkedIn: linkedin.com/in/taicai
GitHub: github.com/padtai

Updated Jul 2018
tai@mail.harvard.edu
(857) 998-8619

Summary

Graduate-level training in: Bayesian statistics (multilevel models, causal inference, model averaging, nonparametric theory), classical statistics (hypothesis testing, mixed models, regularization, asymptotic theory)

Comfortable using: Python (scikit-learn, NumPy, pandas, matplotlib, SQLAlchemy, Flask, Beautiful Soup), R (ggplot2, Markdown, Shiny), Jupyter Notebook, Amazon Web Services, LaTeX/LyX, Microsoft Excel, HTML, CSS, Javascript, git

- **Some exposure to:** Keras, Hadoop, Spark, Scala, Java, SAS, Stata

Experience

Data Scientist, BitSight Technologies, *Cambridge MA* — Oct 2017 - Present

- Modeled time series data to provide an early warning system for customers, experimenting with quantile regression forests, recurrent neural networks, and ARIMA models
- Evaluated the efficacy of various products with statistical models, and conveyed results to non-technical audiences
- Led reading groups covering forecasting methods, prediction intervals for neural networks, and model evaluation metrics
- Supported data science team members in ad hoc statistical tasks

Data Science Fellow, Insight Data Science, *Boston MA* — May 2017 - Sep 2017

- Generated idea for predicting supply/demand at Hubway bike-sharing stations in Boston, despite a lack of labeled data
- Consolidated and cleaned multiple data sources to tally labeled information for 200 stations over 11 million time points
- Predicted supply/demand for each station using tree-based machine learning models
- Created dynamically updating website to help users plan trips using Hubway

PhD Student, Research Assistant, Harvard University, *Cambridge MA* — Aug 2012 - Sep 2017

- Developed R code for novel methods in statistical genetics and Bayesian hierarchical models
- Derived distributions of statistics used in classical and Bayesian inference problems
- Taught graduate-level courses with topics ranging from introductory statistics to seminar topics in Bayesian nonparametrics, decision theory, and sequential methods

Actuarial Analyst, Biometric Research/Enterprise Risk Management, *Toronto ON* — Jan 2011 - Dec 2011

- Built cash flow projection model used for assessing the financial implications of actuarial assumption changes
- Calibrated risk management models of economic risk capital, pandemics, and calamities
- Calculated counterparty exposure used to assess credit risk before finalizing a major reinsurance proposal

Actuarial Analyst, Group Benefits/Segregated Funds, Manulife Financial, *Waterloo ON* — Sep 2009 - Aug 2010

- Conducted quantitative studies to report the effects of deterministic market and interest rate shocks on actuarial reserves
- Mapped segregated funds to clusters of existing proxies and evaluated the mapping algorithm

Pension Administrator, Hewitt Associates, *Toronto ON* — Jan 2009 - Apr 2009

- Prepared pension statements, adjustments, and valuations for defined benefit and contribution plans

Database Analyst, Logitech, *Toronto ON* — May 2008 - Aug 2008

- Reviewed database of electronic devices and produced XML code to expand the compatibility of the Harmony remote

Education

PhD, Biostatistics, Harvard University, *Cambridge MA* — Aug 2012 - Sep 2017

- Thesis: Statistical Methods for the Analysis of Observational Data with Multiple Correlated Outcomes

Bachelor of Mathematics, University of Waterloo, *Waterloo ON* — Sep 2007 - Jun 2012

- Honours Actuarial Science/Finance Option, Honours Statistics, Co-operative Program

Associate of the Society of Actuaries, Society of Actuaries, *Schaumburg IL* — Jan 2009 - Mar 2012