

AARÓN SONABEND

872 Mass. Ave. Apt. 607. Cambridge MA 02139
(617) 751 8872 asonabend@g.harvard.edu

EDUCATION

Harvard University

2016 - 2021

Ph.D., Biostatistics (expected 2021)

A.M., Biostatistics (May 2018)

Advisor: Professor Tianxi Cai, Dissertation Committee: Peter Szolovits, Rajarshi Mukherjee

Instituto Tecnológico Autónomo de México

2010 - 2014

B.S., Applied Mathematics, and B.A., Economics

My research has been motivated by the biological field, specifically sequential settings where outcomes are expensive to measure and depend on all previous history, such as chronic diseases. I'm interested in developing theory and methods for: 1) **reinforcement learning in semi-supervised, non-Markovian settings**, 2) **unsupervised classification methods using natural language processing**. These methods are used for optimizing disease treatments, and phenotyping with electronic health records.

SELECTED PAPERS

Sonabend A, M. Pellegrini A, Chan S, E. Brown H, N. Rosenquist J, H. Perlis R, Cai T, *Integrating questionnaire measures for transdiagnostic psychiatric phenotyping using word2vec*. PLOS ONE, 2020.

Sonabend A, Cai W, Ahuja Y, Ananthakrishnan A, Xia Z, Hong C, Yu S *Automated ICD coding via unsupervised knowledge integration (UNITE)*. International Journal of Medical Informatics, 2020.

Sonabend, Adam M.; Zacharia, Brad E.; Cloney, Michael B.; **Sonabend, A** et. al. *Defining Glioblastoma Resectability Through the Wisdom of the Crowd : A Proof-of-Principle Study*. Neurosurgery, 2016.

Sonabend A, Mukherjee R, Cai T *Semi-supervised Q-Learning and Off-Policy Value function Estimation for Dynamical Treatment Regimes*. (Submitted)

Sonabend A, Lu J, Celi L, Szolovits P, Cai T *Expert-Supervised Reinforcement Learning for Offline Policy Learning and Evaluation*. (Submitted)

Hong C, Sun J, Lu J, **Sonabend A**, Liao K, Cai T, et. al. *Large Scale Code Embedding with Applications to Feature Selection and Knowledge Discovery in Electronic Health Records*. (Submitted)

Sonabend A, Laha N, Mukherjee R, Cai T *Convex Loss Relaxation for Value Function Optimization*. (In preparation)

Sonabend A, Zhang J, Lu J, Coull B, Schwartz *Median Posterior Sampling for Fast Bayesian Kernel Machine Regression*. (In preparation)

ACADEMIC EXPERIENCE

PhD Student

2016-2021

Department of Biostatistics - Harvard University

Cambridge, USA

- Dissertation: Reinforcement Learning Methods for learning Dynamical Treatment Regimes using Electronic Health Records

Visiting Scholar

2019

Department of Statistics in Systems Biology - University of Bordeaux

Bordeaux, France

- Developed a natural language transfer-learning tool for French and US EHR data
- Implemented Automated ICD Coding via Unsupervised Knowledge Integration method for phenotyping based on clinical notes from the *Centre Hospitalier Universitaire de Bordeaux*

Research Analyst

2014-2016

Mathematics Department-Instituto Tecnológico Autónomo de México

Mexico City, Mexico

- Investigated the impact of centrality on the evolution of the economic development of US counties using a US railway dynamical multiplex network

Research Analyst

General Directorate of Economic Research, Mexico's Central Bank

2013-2014
Mexico City, Mexico

- Developed dynamic macroeconomic risk indicator models, for financial risk prediction
- Published monthly macroeconomic memorandum for the Monetary Policy Committee

Research Collaborator

National Institute of Perinatology Genomics Department

2013-2014
Mexico City, Mexico

- Implemented ensemble models for prediction of embryonic congenital malformations and miscarriage in high-risk pregnancies

Researcher

Aalto Science Institute - Applied Physics Laboratory

2013
Helsinki, Finland

- Derived an analytical model of Random Sequential Adsorption (RSA) for a high-dimensional lattice
- Programmed and simulated multidimensional RSA dynamics on a single dimension space

PROFESSIONAL EXPERIENCE

Senior Financial Analyst

Evercore Private Equity Fund

2014-2016
Mexico City, Mexico

- Diagnosed macroeconomic impact on different industries where the Fund is invested
- Performed analysis and screening of acquisitions opportunities for the Fund
- Lead team in charge of the debt restructuring and merger process of a portfolio company

Staff writer

Economics Research and Analysis Center

2012-2014
Mexico City, Mexico

- Wrote International Economics monthly section

TALKS AND CONFERENCES

Expert-Supervised Reinforcement Learning for Offline Policy Learning and Evaluation

Virtual Conference on Reinforcement Learning for Real Life

2020

Natural Language Processing: Embeddings and Use Case Examples

Department of Statistics in Systems Biology - University of Bordeaux

2019
Bordeaux, France

Interpretable Q-learning for Optimal Dynamic Treatment Regimes with Observational Data

Department of Statistics in Systems Biology - University of Bordeaux

2019
Bordeaux, France

Interpretable Q-learning for Optimal Dynamic Treatment Regimes with Observational Data

Eastern North American Region. International Biometric Society Conference poster

2019
Philadelphia, PA.

Smartphone-Based Digital Phenotyping: Analysis of Social Behavior in a Brain and Spine Tumor Cohort

Harvard T.H. Chan School of Public Health, Department of Biostatistics

2017
Boston, MA

Network Centrality as a Precursor of Growth: Evidence from the US 1840-1900

Network Stress Testing for Financial Stability and Macprudential Policy Design

2015
Mexico City, Mexico

GRANTS AND AWARDS

Rose Traveling Fellowship Program in Chronic Disease Epidemiology and Biostatistics

Harvard T.H. Chan School of Public Health

2019

Certificate of Distinction in Teaching <i>Harvard T.H. Chan School of Public Health, Department of Biostatistics</i>	2018
Highest Honors on Bachelor's Thesis <i>Instituto Tecnológico de México</i>	2014

TEACHING EXPERIENCE

International Meeting on Artificial Intelligence and its Applications <i>Taught summer course: Reinforcement Learning & OpenAI</i>	2019 <i>Mexico City, Mexico</i>
Harvard T.H. Chan School of Public Health <i>Designed and taught 10 day summer course: Data Science in Action: CNN for Self-Driving Cars</i>	2019, 2020 <i>Boston, MA</i>
Clubes de Ciencia Mexico <i>Designed and taught summer course: Smart automatons with machine learning</i>	2018 <i>Guadalajara, México</i>
Harvard T.H. Chan School of Public Health. <i>Teaching Assistant to the following Courses:</i>	2016-2019 <i>Boston, MA</i>
<ul style="list-style-type: none"> · Data Science II, Professor: Heather Mattie · Practice and Culminating Experience for Quantitative Methods, Professor: Marcia Testa · Introductory Genomics & Bioinformatics for Health Research, Professor: John Quackenbush, · Applied Regression Analysis, Professor: Robert J. Glynn 	
Economics Department -Instituto Tecnológico Autonomo de Mexico <i>Teaching Assistant to the following Courses:</i>	2013-2014 <i>Mexico City, Mexico</i>
<ul style="list-style-type: none"> · Industrial Organization, Professor: Ricardo Enriquez Frola · Microeconomics I, Professor: Magdalena Barba 	

LEADERSHIP & TEAMWORK

Clubes de Ciencias <i>Fund Raiser and Harvard liason</i>	2018-2019 <i>Boston, MA</i>
<ul style="list-style-type: none"> · Manage collaboration between Clubes de Ciencias Mexico and Harvard University, currently securing funding 	
Harvard TH Chan School of Public Health Biostatistics Student Consulting Center <i>Student Consultant</i>	2018-2019 <i>Boston, MA</i>
<ul style="list-style-type: none"> · Provided statistical consulting to Master and Doctoral students for their dissertation and research projects 	
University Council of Honor and Excellence, ITAM <i>President</i>	2012-2013 <i>Mexico City, Mexico</i>
<ul style="list-style-type: none"> · Coordinated student unions, and reformed election system to incentivize competition between representatives · Pioneered a Diversity Group to promote tolerance towards the LGBTQ community 	
Economical and Financial council, Model United Nations <i>Council President</i>	2011-2012 <i>Mexico City, Mexico</i>
<ul style="list-style-type: none"> · Moderated a Model United Nations, lead delegates to write and present resolutions to the General Assembly 	

SKILLS

- Computational: Python (TensorFlow, Pytorch, Keras), R, Matlab, Stan, Stata, Microsoft Office
- Language: Fluent in English, Fluent in Spanish (Native language), Conversational Hebrew