### Gustavo Lacerda

Quantitative Researcher, Data Scientist, Statistician

#### Contact

Gustavo Lacerda	English:	0	igodot	igodot	•	igodot
415-297-6732	Portuguese:	igodot	igodot	igodot	igodot	igodot
guscv@optimizelife.com	Dutch:	igodot	۲	igodol	igodot	0
www.optimizelife.com	German:	igodol	۲	igodot		0
	French:	igodol	۲	igodot	0	0
	Spanish:	$\circ$	igodot	igodol	0	0

#### SUMMARY

- I bring expertise in statistics and machine learning to bear on learning and decision problems.
- Extensive experience with hierarchical Bayesian modeling, generalized linear models (GLMs), Monte Carlo methods, non-parametric Bayesian modeling, non-standard (sparse, non-Gaussian) factor analysis, and design & implementation of custom statistical models. Published author on causal inference methodology with graphical models.
- Experience with a large variety of languages and systems. 10 years of experience with R. Statistical modeling with Stan. Python, Julia and Matlab come easily to me. Interactive art with Processing.
- Strong interests in cogsci, pedagogy, forecasting, and ways to integrate AI with human expertise.

### Employment History Sta

### Statistics Consultant, self-employed (Remote), 2020-present

Mostly helping a Biotech startup with epidemiological data analysis, forecasting and visualizations; Smaller project on fraud detection; Statistical theory project.

#### Data Scientist, Google (Mountain View and Sunnyvale, CA), 2014–2019

Developed tools for forecasting ad metrics (clicks, cost-per-click, conversion value, auction win rate, etc) as a function of bid. Drove the first segmentation analysis of publisher revenue on AdSense, a billion-dollar business. Forecast calibration for new ad campaigns. Research, implementation and evaluation of time series forecasts. Contributed general-purpose utilities for debugging R code. R interactively and in production, GoogleSQL + PLX macros, BCL (Borg Config), Lingo (Logs in Go) pipelines, Bash scripting, Google Sheets, PLX dashboards.

### Research Internship, Google (Mountain View, CA), Summer 2013

Exploratory statistical analysis of the webtables corpus. Wrote C++/Flume jobs to extract and transform data. Worked with sharded files. Data analysis with R.

### Teaching Assistant, Columbia University, Statistics Department (New York, NY), 2010 – 2014

Graduate courses in Data Mining, Stochastic Processes; Statistical Inference. Undergrad course on statistical literacy and critical thinking.

# Teaching Assistant, University of British Columbia (Vancouver, Canada), 2008–2009

Models of Computation, Introduction to Artificial Intelligence, Introduction to Relational Databases.

### AI Programmer, Carnegie Mellon University (Pittsburgh, Pennsylvania), 2006 – 2008 (60% time)

Developing and maintaining SimStudent, a system that learns production rule models of expert or novice skills from behavior data using Inductive Logic Programming.

# Independent Researcher, Carnegie Mellon University (Pittsburgh, Pennsylvania), 2006 – 2008 (40% time)

Main researcher on project about discovering causal graphical models from observational data, extending Shimizu et al's ICA-based method (with Peter Spirtes).

### Lisp Programming Internship, Cadence Design Systems (Munich, Germany), Jan 2006 – May 2006

Worked in the custom development team: developing, debugging, testing and re-releasing packages, using the SKILL language + Common Lisp.

### Software Engineer, Amazon Technologies, 2001–2002 (Woburn, Massachusetts) Web programming with JSP / JBuilder: front-end and back-end Java.

# Research Intern at Bell Labs, Lucent Technologies (Murray Hill, New Jersey), Summer 2000

Research intern at the Mathematics of Communications department. Implemented variations of the Viterbi algorithm to decode linear error-correcting codes over simulated noisy channels.

### EDUCATION

# PhD candidate in Statistics, Columbia University (left with Master's), 2010–2014

Research in non-parametric factor analysis, subspace dictionary learning. Core Courses: Mathematical Stat I, II, III; Applied Stat I, II; Analysis & Probability I; Stochastic Processes. Electives: Bayesian Computation; Statistical Computing; Non-parametric Bayes; Sparsity Seminar; Bayesian Models for Machine Learning.

### MSc in Computer Science, University of British Columbia, 2008–2010

Master's thesis with Dr. Jennifer Bryan about statistical methods for identifying geneand protein-interaction networks.

### MSc in Logic, Universiteit van Amsterdam, 2003–2005

Mostly Artificial Intelligence, Logic and Cognitive Science.

**B.S. in Mathematics and Computer Science, Bucknell University, 1997–2001** Also courses in Physics, Philosophy, Psychology and Linguistics.

### Non-degree

- Complex Systems Summer School, Santa Fe Institute, June 2009.
- IPAM Graduate Summer School on Bayesian methods in Cognitive Science, UCLA, July 2007.

### PUBLICATIONS

S. Carré, F. Gabriel, C. Hongler, G. Lacerda, G. Capano – Smart Proofs via Smart Contracts: Succint and Informative Mathematical Derivations via Decentralized Markets, Feb 2021

**Gustavo Lacerda** – "Identification of gene modules using a generative model for relational data", UBC Master's thesis (2010), supervised by Jennifer Bryan.

**Gustavo Lacerda**, Peter Spirtes, Joseph Ramsey, Patrik Hoyer – Discovering Cyclic Causal Models by Independent Components Analysis. In Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI-2008). (plenary talk)

Patrik Hoyer, Aapo Hyvärinen, Richard Scheines, Peter Spirtes, Joseph Ramsey, **Gustavo Lacerda**, Shohei Shimizu – Causal discovery of linear acyclic models with arbitrary distributions. In Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI-2008), in press.

Noboru Matsuda, William W. Cohen, Jonathan Sewall, **Gustavo Lacerda**, and Kenneth R. Koedinger – Why Tutored Problem Solving may be better than Example Study: Theoretical Implications from a Simulated-Student Study. In Proceedings of the International Conference on Intelligent Tutoring Systems 2008

Noboru Matsuda, William W. Cohen, Jonathan Sewall, **Gustavo Lacerda**, and Kenneth R. Koedinger – Evaluating a Simulated Student using Real Students Data for Training and Testing, *In Artificial Intelligence in Education* 

Noboru Matsuda, William W. Cohen, Jonathan Sewall, **Gustavo Lacerda**, and Kenneth R. Koedinger – Predicting Students' Performance with SimStudent: Learning Cognitive Skills from Observation. *In International Conference on User Modeling 2007.* 

S. Fissaha Adafre, W.R. van Hage, J. Kamps, G. Lacerda de Melo, and M. de Rijke – The University of Amsterdam at CLEF 2004, In C. Peters and F. Borri, editors, Working Notes for the CLEF 2004 Workshop, pages 91-98, 2004.

Community Service

- **Reviewer:** International Conference on Learning Representations (ICLR) 2019; Human Brain Mapping 2010; Cognitive Science 2008, 2009; International Conference on Computers in Education 2007.
- **Conference organization:** Neural Information Processing Systems 2009 (student volunteer), Artificial Intelligence in Education 2005 (student volunteer)