

# Gustavo Lacerda

*Quantitative Researcher, Data Scientist, Statistician*

---

## CONTACT

Gustavo Lacerda  
415-297-6732  
guscv@optimizelife.com  
www.optimizelife.com

English: ● ● ● ● ●  
Portuguese: ● ● ● ● ●  
Dutch: ● ● ● ● ○  
German: ● ● ● ● ○  
French: ● ● ● ○ ○  
Spanish: ● ● ● ○ ○

## SUMMARY

- I bring expertise in statistics, machine learning and information theory 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, prototype and production. 10 years of experience with R. Modeling with Stan. Have done projects in Python and Julia.
- Strong interests in cogsci, pedagogy, forecasting, and ways to integrate AI with human expertise.

## EMPLOYMENT HISTORY

### **Data Science Consultant, self-employed (Remote), 2020–present**

Epidemiological forecasting, visualization; Fraud detection for anonymous client.

### **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.

**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.

**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 gene- and protein-interaction networks.

**MSc in Logic, Universiteit van Amsterdam**

Mostly Artificial Intelligence, Logic and Cognitive Science.

**B.S. in Mathematics and Computer Science, Bucknell University**

Also courses in Physics, Philosophy, Psychology and Linguistics.

**Non-degree**

- Complex Systems Summer School, Santa Fe Institute, June 2009.

SOME  
PUBLICATIONS

S. Carré, F. Gabriel, C. Hongler, **G. Lacerda**, G. Capano (2021) – “Smart Proofs via Smart Contracts: Succinct and Informative Mathematical Derivations via Decentralized Markets”

**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 Con-*

*ference 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.*

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. Artificial Intelligence in Education 2005 (student volunteer)