Jenna Landy

Email: jlandy@g.harvard.edu
Phone: (831) 588-6073

GitHub: github.com/jennalandy
LinkedIn: linkedin.com/in/jenna-landy
Website: www.jennalandy.com/

I am interested in both statistical and deep learning models of molecular biology. My Ph.D. research focuses on cancer mutational signatures, which model underlying mutational processes with latent factor models, and I produce publicly available software to encourage the reproducibility and usability of my work. In my next role, I aim to expand further into deep learning and biologically informed modeling with applications to cancer or immune genomics.

Research interests	**Bayesian Modeling:** Bayesian factor modeling, Gibbs sampling, semi-supervised learning with informative priors, biologically-motivated models **Deep Learning:** graph neural networks (GNNs) as informative priors, deep algorithm unfolding, deep representation learning **Model Selection & Ensembling**: ensemble methods for unsupervised learning, Bayesian model averaging, selective inference **Cancer and Immune Genomics:** cancer mutational signatures analysis, liquid biopsy and fragmentomics for cancer detection, Immune receptor sequencing, T cell repertoire modeling, integration of clinical and omics data

Education

Harvard Graduate School of Arts and Sciences — Cambridge, MA
PhD in **Biostatistics** (Masters awarded May 2022) — August 2020 – May 2025
Advisor: Giovanni Parmigiani — *GPA: 3.95*

Selected coursework

- Computer Science and Data Science: Computational Biology and Bioinformatics, Cancer Genome Data Science, Deep Learning in NLP, Data Structures and Algorithms, Decision Theory
- Statistics: Bayesian Statistics and Computation, Probability Theory, Inference, Linear Models, Unsupervised Learning, Statistical Genetics, Causal Inference

California Polytechnic State University, SLO — San Luis Obispo, CA
BS in **Statistics**, minor in **Data Science** — September 2016 – June 2020
Mentors: Hunter Glanz, Prince Afriye, Brian Granger — *GPA: 3.95*

Selected coursework

- Computer Science and Data Science: Distributed Computing, Object Oriented Programming, Machine Learning and Data Science, Database Systems, Ethics in Technology
- Math and Statistics: Multivariate Statistics, Survival Analysis, Calculus, Linear Algebra, Methods of Proof in Mathematics
- Biology: Human Genetics, Cell and Molecular Biology

Skills

Mathematics and Statistics: Probability Theory, Inference, Regression, Bayesian Methods and Computation, Representation Learning, Statistical Genetics, Machine Learning

Machine Learning and Deep Learning: Model-Based Deep Learning, Deep Algorithm Unrolling, Autoencoders, RNN, LSTM, CNN, GAN

Programming: Object-Oriented and Functional Programming, Data Manipulation (Pandas, tidyverse), Simulation, Databases, APIs, Machine Learning (PyTorch, Scikit-Learn, R), Visualization (matplotlib, ggplot, RShiny), Cluster Resources, Bash, Software Development, Git/GitHub, Docker, AWS, Single Cell Analysis (Cell Ranger, scanpy, scirpy, anndata, muon)

Languages: Python, R, Java, JavaScript (TypeScript, React), MongoDB, Hadoop, PySpark, SQL (JDBC), SAS, Stata

Publications & Preprints

Landy, J.M., Zorzetto, D., De Vito, R., and Parmigiani, G. (2025). ***Causal Inference for Latent Outcomes Learned with Factor Models.*** arXiv preprint arXiv:2506.20549.
*Open source software **causalLFO***: R package to perform causal inference for latent outcomes learned with non-negative matrix factorization with uncertainty quantification.

Lang, I., **Landy, J.M.**, and Parmigiani, G. (2025). ***Bayesian Non-Negative Matrix Factorization with Correlated Mutation Type Probabilities for Mutational Signatures.*** arXiv preprint arXiv:2506.15855.

Zorzetto, D., **Landy, J.**, Zigler, C., Parmigiani, G., and De Vito, R. (2025). ***Multivariate Causal Effects: a Bayesian Causal Regression Factor Model.*** arXiv preprint arXiv:2504.03480.

Landy, J.M., Basava, N. and Parmigiani, G. (2025). ***bayesNMF: Fast Bayesian Poisson NMF with Automatically Learned Rank Applied to Mutational Signatures.*** arXiv preprint arXiv:2502.18674.
*Open source software **bayesNMF***: R package implementing a computationally efficient Gibbs sampler for Bayesian Poisson Non-Negative Matrix Factorization (NMF) which automatically determines the number of latent factors while still quantifying uncertainties of the learned matrices.

Landy, J.M., and Parmigiani, G. (2024). ***Gridsemble: Selective Ensembling for False Discovery Rates.*** arXiv preprint arXiv:2401.12865.
*Open source software **gridsemblefdr***: R package for estimating local (fdr) and tail-end (Fdr) false discovery rates in large-scale multiple hypothesis testing.

Self, B. P., **Landy, J.**, Widmann, J. M., Chen, J., Kerfs, M. (2021, July), **The Mechanics of SUCCESS: How Non-Cognitive and Affective Factors Relate to Academic Performance in Engineering Mechanics.** ASEE Virtual Annual Conference Content Access. https://strategy.asee.org/37876

Chen, J., **Landy, J.M.**, Scheidt, M., Major, J. C., Ge, J., Chambers, C. E., Grigorian, C., Kerfs, M., Berger, E. J., Godwin, A., Self, B. P., Widmann, J. M. (2020, June), **Learning in Clusters: Exploring the Association Between Noncognitive and Affective Profiles of Engineering Students and Academic Performance.** ASEE Virtual Annual Conference Content Access. https://peer.asee.org/34901, DOI: 10.18260/1-2−34901

Widmann, J., Self, B., Chen, J., Chambers, C., Kusakabe, K., **Landy, J.**, Berger, E., Ge, J., Godwin, J., and Scheidt, M. (2019, July), **Academic SUCCESS: An Analysis of How Non-Cognitive Profiles Vary by Discipline for Engineering and Computer Science Students.** Research in Engineering Education Symposium (REES). https://www.sasee.org.za/wp-content/uploads/REES-2019-proceedings.pdf, pages 540 - 548.

Invited sessions	**New England Statistics Symposium**	2024
	Mutational Signatures in Practice and an Introduction to bayesNMF	

Contributed sessions	**Joint Statistical Meetings**	2025 (planned)
	bayesNMF: Fast Bayesian Poisson NMF with automatically learned rank applied to mutational signatures analysis	
	The Bayesian Young Statisticians Meeting	2025
	Mutational signatures analysis with bayesNMF: fast Bayesian Poisson NMF with automatically learned rank	
	New England Student Research Symposium on Statistics and Data Science	2024
	Getting started with mutational signatures	
	Women in Statistics and Data Science	2023
	Gridsemblefdr: model selection and ensembling for false discovery rates with application to differential expression analysis	
	Joint Statistical Meetings	2022
	Gridsemblefdr: ensembling and hyperparameter optimization for unsupervised learning with application to false discovery rates	

Poster Sessions	**Cal Poly Honors Thesis Seminar**	2020

Walking through an intuitive, collaborator-friendly R Shiny UI to predict hospital readmission due to sepsis in collaboration with Dignity Health.

Jupytercon 2018
Three JupyterLab Extensions for Enhanced User Experience

Blog Post: Getting Started with Mutational Signatures 2024
This article introduces the biological motivation for mutational signatures, the statistical methods used, and key literature to get started in this field.

Tutorial: Document Creation with Rmd and Quarto 2023
An introduction to using Rmd and Quarto files in quantitative research. Presented at the Dana Farber Data Science workshop series.

Tutorial: Remodel your Rmd 2022
Tips and tricks for experienced quantitative researchers to boost their research workflow with R package development, parameterized reports, Rmd websites, and Rmd customizations. Presented at the Harvard Biostatistics Student Seminar.

Textbook: Introduction to Databases and API 2020
A Bookdown and RShiny app introducing statistics students to the use of databases and APIs in R, Python, SAS, and Julia. This work is aimed at lowering technical barriers that keep statistics students from using these tools. All materials are public on GitHub. Worked in a team of two advised by Hunter Glanz, PhD and Rebecca Ottesen, PhD.

Moderna Cambridge, MA
Title: Computational Immunology Co-Op July - December 2024
Manager: Roger Vargas
Key Skills: Machine learning, single-cell data analysis, Bayesian modeling, software engineering
Tools Used: Git/GitHub, Scipy, AnnData, Muon, CellRanger, AWS S3 and boto3, Docker
Applied machine learning and statistical modeling to investigate immune responses, vaccine reactogenicity, and vaccine uptake using single-cell RNA-seq, immune receptor sequencing, flow cytometry, and clinical data. Designed and documented reproducible Python tools for the team, including a `scirpy`-based package for standardized single-cell processing, a `sklearn`-based package for an in-house model selection and hyperparameter tuning framework, and a wrapper to run `Cell Ranger` directly in Python.

Delfi Diagnostics   Baltimore, MD

Title: Data Science Research Intern   June - August 2022

Mentor: Laurel Keefer

Project: Data Visualization and Automated Reports R Software Package

Developed and documented an in-house R package for visualizations, modeling, and automated Rmd reports for use by the Data Science Research team. Followed R tidyverse style guide and used git/GitHub for version control.

Amazon Web Services   Seattle, WA

Title: Data Science Intern   June - September 2019

Mentor: Steve Loeppky

Project: Public GitHub Notebook Corpus Research Collaboration

Extracted and analyzed all Jupyter Notebooks public on GitHub to understand AWS Sagemaker and Jupyter users, their processes, and their struggles in order to inform big-picture user experience questions. Used the GitHub API and AWS EC2 and S3 instances. All code and results are public on GitHub.

Project Jupyter   San Luis Obispo, CA

Title: Software Engineering Intern   March – December 2018

Mentor: Brian Granger

Project: Developing JupyterLab extensions to enhance user experience

Contributed to JupyterLab, an open-source interactive development environment. Worked in a small team to design and implement multiple extensions, including a visualization toolkit for machine learning, an interface to view and edit keyboard shortcuts, git/GitHub integration, and a minimal extension for basic version control without needing to learn git. Participated in iterative design and user testing, and presented the work at JupyterCon 2018. *Open-source tools* plyto, jupyterlab-shortcutui, jupyterlab-git, and easygit.

Teaching experience

Tutor, Biostatistics Department, Harvard   2021 – 2024

Tutored one-on-one and group sessions for the Biostatistics Ph.D. Core Coursework: Probability, Inference, Methods, Data Structure, and Algorithms. Prepared lessons, answered questions, and provided resources.

Teaching Assistant, Biostatistics Department, Harvard

BST 221: Introduction to Data Structures and Algorithms   2022

Taught lab section, hosted weekly office hours, and graded assignments for a PhD level course in design and analysis of algorithms.
Average student rating: 5/5

BST 201: Introduction to Statistical Methods   2021

Taught weekly lab section, hosted weekly office hours, graded assignments and exams for a Masters level course covering basic statistical techniques.
Average student rating: 4.8/5

Teaching Assistant, Statistics Department, Cal Poly

STAT 427: Mathematical Statistics   2020

Hosted weekly office hours and graded assignments for an upper-division course on the theory of hypothesis testing and its applications.

STAT 426: Estimation and Sampling Theory   2020

Hosted weekly office hours and graded assignments for an upper-division course on properties of statistics obtained from samples and asymptotics.

STAT 425: Probability Theory   2019

Hosted weekly office hours, assisted in lab section, and graded assignments for an upper-division course on the rigorous development of probability theory.

Competitions   **DataFest Hackathon, UCLA**   2018, 2019, 2020

ASA sponsored annual data science hackathon. In 2019, my team won the data visualization category out of ~40 student teams.

RShiny App Competition, Cal Poly Statistics Department   2020

Built an RShiny web app for Medium data science article recommendations and visualizing article popularity through web scraping and topic modeling. I won the competition and was awarded funding to attend the RStudio conference in January 2020.

Honors & Awards   **Biostatistics Department, TH Chan School of Public Health**

Certificate of Distinction in Teaching   2021-2022
Robert B. Reed Prize for Excellence in Biostatistics for the highest
score on the written qualifying exam   2021

California Polytechnic State University, San Luis Obispo

Summa Cum Laude   2020
Graduated with Honors (Cal Poly Honors Program)   2020
Award for Contribution to the Objective and Public Image of the
College of Science and Mathematics   2020
Academic Merit Award (Department of Statistics)   2020
Humanitarian Award (Department of Statistics)   2020
Dean's List and President's Honors List   2016-2020
Merit-based full-ride Frost scholarship (Cal Poly Frost Fund)   2016-2020

Service and outreach   **Stat Start**   2022

Volunteered teaching courses on basic statistics, R, and data analysis for Stat Start, a computational summer program for high school students.

Cal Poly College of Science and Math Student Council 2017 – 2020
This is a committee of student leaders and the Dean that serves as a line
of communication between students and faculty in the college. I initiated a
college-wide peer mentoring group, volunteered at the college's annual re-
search conference, and contributed to a campaign to have faculty discuss
campus climate in their classrooms. As president in the 2018 – 2019 aca-
demic year, I planned events, scheduled and ran meetings, and organized
guest speakers from across campus.

Professional
Memberships

Caucus for Women in Statistics Student member 2023 –
American Statistical Association Student member 2017 –