

Paul Shearer

shearer.pr@gmail.com | (734) 272 - 6817

<http://shearerp.github.io>

[GitHub](#) | [LinkedIn](#) | [Twitter](#)

Education

2006 - 2013	Ph.D.	Applied and Interdisciplinary Mathematics	U. Michigan	GPA: 4.0/4.0
2002 - 2006	B.S.	Mathematics and Physics, <i>high distinction</i>	U. Minnesota	GPA: 3.98/4.0

Recent Positions

Aug 2014 - present: *Data Scientist (AVP)*, MassMutual Financial Group, Boston, MA.

- Created & backtested a risk underwriting model for a complex insurance product. Estimated that decline rates could be cut 50% while maintaining the same level of risk control as existing underwriting practice.
- Developed web-based lead prioritization tool for a \$6B/yr sales team using advanced logistic regression methods and the Shiny web framework. Led a pilot involving several dozen users, managers, and operational sales personnel.
- Improved data quality for a large-scale survival study by designing and applying a custom fuzzy record linkage method. Implementation of a scalable version in-progress using HDFS, PySpark, and Spark SQL.
- Isolated factors driving a key sales metric by analyzing a large, messy multi-level dataset.
- Solved a data quality issue impacting key performance metrics on a major insurance product.
- Improved survival analysis performance using a novel feature engineering approach.

Jun 2013 - Aug 2014: *Research Fellow*, Space Science, U. Michigan.

- Designed and implemented new statistical analysis software for a solar wind ion spectrometer.
- Used software to process 17 years of data (~100 GB), producing a new, continuously updated, and carefully validated public dataset for the space research community.

Aug 2010 - May 2013: *Ph.D. Candidate*, U. Michigan.

- Developed new optimization techniques for separable inverse problems, new methods for blind deconvolution, and applied these methods to stray light correction for the extreme ultraviolet solar imager STEREO/EUVI.
- Distributed stray light tools to astrophysicists through a community analysis package.
- Distributed camera shake correction code through GitHub.

Skills

Technical Computing & Data Manipulation

- *Proficient:* R, Python, SQL, algorithms & data structures, Unix shell scripting, MATLAB, LaTeX
- *Some Experience:* HDFS, PySpark, ODBC/JDBC, C/C++, JavaScript, Perl, C#, Excel/VBA

Math, Statistics, Machine Learning, & Signal Processing

- *Modeling:* GLMs, mixed and multilevel models, longitudinal/panel data analysis, time-to-event analysis (e.g. PH and AFT), mixture models, random forests, boosting, Bayesian methods, graphical models, robust statistics & fitting, Fourier & multiscale signal processing
- *Performance evaluation:* ROC, bootstrap, cross-validation & stratification, backtesting
- *Technical foundations:* probability & stochastic processes, linear algebra, sparse arrays, numerical optimization (convex, nonconvex, and large-scale), penalized likelihood, Monte Carlo

Visualization & Web

- *R:* Shiny, ggplot, dygraphs, rCharts (wrapper for D3.js, NVD3, etc.)
- *Python:* matplotlib, IPython/Jupyter notebook, Jekyll, Flask

Awards & Honors

2015	Script of the Week, Kaggle
2015	Leading & Strong Performance Ratings, MassMutual Financial Group
2014	Insight Data Science Fellowship (declined for MassMutual position)
2013	Outstanding Paper Award, IEEE International Conference on Image Processing.
2012 - 2013	Rackham Predoctoral Fellowship, University of Michigan.
2006 - 2009	Departmental Fellowship, University of Michigan.
2005	Top 3% Score, Putnam Mathematics Exam.
2005	Mathematics Departmental Scholarship, University of Minnesota.
2005	J. Morris Blair Scholarship in Physics, University of Minnesota.
2005	IT Merit Scholarship, University of Minnesota.

Other Positions & Projects

2014: Matrix factorization (personal project)

Designed a fast matrix factorization method based on alternating projected gradient descent. Method enforces user-specified structural constraints on the factors using a proximal operator framework. Applied to gene microarray data, delivering order-of-magnitude speedup over a published method designed for the same dataset.

2012: Summer intern, MIT Lincoln Laboratories (less than 10% applicants accepted).

Developed a simulation-based performance benchmark for image reconstruction algorithms operating on ISAR radar data.