

Education

May 2018 **Ph.D. Statistics**, *The Ohio State University*, Columbus, Ohio

Advisor: Yoonkyung Lee

January 2010 **M.S. Statistics**, *The Ohio State University*, Columbus, Ohio

May 2007 **B.A. Mathematics, Computer Science**, *Capital University*, Columbus, Ohio

Experience

Director, Data Science and Analytics

OCLC, *Columbus, OH*

August 2023-present Leads three teams comprising the Data Science and Analytics group having the collective mission to unlock the value contained in the company's extensive data ecosystem in service of our academic, public, and global library customers.

Develops the vision and execution of OCLC's approach to data science and analytics, working closely with product management and technology leaders to influence and align the business needs to technical strategy. Drives growth and maturation of OCLC's new internal data science practice, from identifying business opportunities to leveraging data science and machine learning to building predictive models and leading their implementation and deployment. Fosters a company-wide culture of data literacy by acting as an advocate and visionary leader in data visualization and storytelling within the technology organization as well as across OCLC's product portfolio.

Additionally leads two technical teams; one of these teams owns the development of two of OCLC's software products that serve as tools for library collection analysis and inventory management. The last team is an analytics enablement team which is responsible for providing self-service data capabilities for internal OCLC stakeholders, developing enterprise data models, and monitoring and measuring data quality.

Leads hiring and career development of engineers and data scientists within the team and across the organization to foster and nurture a culture of innovation and teamwork throughout Global Technology at OCLC.

Founding Data Scientist, Privacy Engineer

Ketch, *San Francisco, CA*

September 2021-March 2023 Lead research in the quantification of disclosure risk and served as engineering's privacy expert. As one of the first members of the company's product and technology organization and the first data science hire, lead the development of Ketch's implementation of privacy enhancing technologies including k -anonymization and mechanisms guaranteeing differential privacy, in addition to the development of several components of Ketch's core data governance product including its native integration with Snowflake, AWS Redshift, Oracle, MySQL, Postgres, and MariaDB database technologies. Spearheaded the data science practice within the engineering organization, and owned the data science effort backing the classification model underlying Ketch's software product for the discovery of sensitive data and PII (personally identifying data).

Senior Research Scientist

Immuta, Columbus, OH

September 2020-September 2021 Senior member of research team driving product feature development for data access and governance software. As Immuta's expert in re-identification risk quantification, lead research and product development of a comprehensive policy-agnostic disclosure risk and utility measurement framework for facilitation of automating privacy policy recommendations for maximum data utility for given disclosure risk tolerance levels.

Contributed to Immuta's digital catalogue of technical and white of ebooks and white papers on topics in statistical disclosure control including methods in risk quantification and recovering basic statistical models with anonymized inputs. Lead efforts in designing and prototyping performance testing for the Immuta/Databricks native integration.

Senior Data Scientist

Redjack, Columbus, OH

September 2019-September 2020 Lead consulting data scientist supporting federal government cyber defense efforts. Leveraged supervised machine learning techniques to identify spam email campaigns, specifically exploratory data analysis techniques for random forests to understand the best features for discriminating between phishing and non-phishing emails.

Data Scientist

Root Insurance, Columbus, OH

April 2019-September 2019 Data scientist supporting teams responsible for customer acquisition and development of pricing plans. Designed and analyzed A/B tests for assessing the impact of experimental treatment on customer behavior, including acquisition and retention. Developed conversion and retention models for predicting the impact of pricing changes on KPIs including loss ratio and total bound premium. Integrated with product teams and actuaries at every stage of the data science life cycle, from problem formulation and data collection to model deployment and communication of insights generated from analysis.

Senior Consulting Data Scientist

Information Control Company, Columbus, OH

January 2017-March 2019 As a consulting senior data scientist, lead data science team members including junior analysts, data engineers, and visualization specialists in planning, designing, and delivering data science solutions addressing a variety of business questions for clients across several industries including retail, food and beverage, insurance, and marketing.

Internally at ICC, contributed to Advanced Analytics development seminars and coordinated the Advanced Analytics journal review, curating and leading discussions about current literature in statistics, data science, and machine learning and its application to ICC client problems.

Machine Learning Specialist

Pillar Technology, Columbus, OH

February 2016-December 2016 Machine learning specialist on a team responsible for designing and developing adaptive software for an IoT product to be embedded in a line of luxury vehicles. Responsible for setting expectations with both internal teams and with the client given the available resources such as data, in-device storage space, and computational constraints. Challenged to design and implement machine learning models in the absence of data a priori, leveraging regularization methods and programmatic model selection techniques. Owned data science efforts from initial exploratory analysis to production deployment of models in Java and Scala using Agile development best practices.

Spearheaded components of project planning pertaining to data and the corresponding infrastructure necessary for data collection, storage, and modeling at scale. Lead efforts to establish client trust in data modeling and algorithms, a new operating space for Pillar.

Data Scientist

Store Development, Starbucks Coffee Company, Seattle, WA

May 2014 - November 2015 Data scientist on a team serving the company in market planning and strategy. Utilized a variety of statistical and machine learning methods, both supervised and unsupervised such as penalized regression and classification, generalized linear models, ensemble methods including boosting, bagging, and random forests, with applications of the latter to both classification and clustering with application to a product recommendation engine, predicting consumer total lifetime value, sales forecasting, and estimating cannibalization effect of newly launched locations on existing stores.

Using heterogeneous spatial and temporal data from disparate sources, estimated the causal impact of a variety of interventions, such as competitor store openings, pricing changes, and new product launches on store performance using Bayesian structural time series models and summarizes comparisons between this methodology and prior approaches, including the traditional difference-in-differences estimators.

Adjunct Statistics Instructor

Department of Mathematics, Columbus State Community College, Columbus, Ohio

August 2013 - January 2014 Lecturer for an introductory statistics course for undergraduate students, Statistics 1350. Non-instruction responsibilities included curriculum and assessment development, including lecture presentations and online learning tools and learning assessments.

Graduate Research Assistant

Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio

October 2011 - August 2012 Responsibilities included analysis of large microarray data sets, in particular utilizing data mining and dimension reduction techniques to find genetic markers in leukemic patients, sharing results and collaborating with medical professionals to both direct further laboratory investigation as well as further statistical investigation.

Graduate Research Assistant

Nationwide Center for Advanced Customer Insights, The Ohio State University, Nationwide Insurance, Columbus, Ohio

June 2010 - June 2011 Responsibilities included work on projects modeling agency behavior using high dimensional demographic and marketing data, specifically by modeling survival times using Cox proportional hazards models with both static and time-varying coefficients. Modeling was done with an emphasis on building parsimonious, interpretable models. I was responsible for presenting results in a corporate setting to high level company executives with motive to encourage and motivate business decisions and action.

Research Interests

My current research interests lie in the quickly emerging space of privacy enhancing technologies and statistical disclosure control. I am particularly interested in how we assess the quality of statistical disclosure control methods and, specifically, how to quantify the risk of re-identification associated with data to which these methods have been applied. Quantifying this risk is challenging for a number of reasons, including the breadth of the mathematical mechanisms underlying these techniques and how to justify the assumptions about the resources available to any potential adversary. My most recent research proposes a framework for estimating the probability of re-identification that is agnostic to privacy model under the most conservative assumptions about an attacker's background knowledge and available resources.

Previous research interests include nonparametric function estimation, particularly utilizing reproducing kernel Hilbert space methods; my dissertation work proposed a novel method for the unconstrained, nonparametric estimation of large covariance matrices which permits irregularly spaced and sparsely sampled data. A specific parameterization of the covariance matrix permits unconstrained optimization while rendering covariance estimation as the estimation of a varying coefficient regression model. This formulation allows one to leverage the extensive machinery typically used in the classic function estimation setting to estimate a positive definite covariance matrix.

Computing Skills

Machine Learning and Data Modeling:

R, Python (Jupyter & JupyterLab, Sci-kit Learn, Sci-Py, Pandas, Numpy, Matplotlib, Plotly), PySpark, SparkML, Snowflake, Databricks, MLFlow, Tableau, PowerBi, dbt

General Software Development:

Golang, Python, PySpark, SQL (Postgres, MySQL), Unix shell, Flask, Docker, AWS (EC2, RDS, Redshift, Elastic Beanstalk, S3, CLI), Protocol Buffers (protobuf), basic knowledge of C++, Java and HTML

Other:

Git/version control, L^AT_EX, project management software including Jira and Confluence

Presentations

- December 2022 **Toward a Unifying Information-Theoretic Framework for Re-identification Risk Quantification**, *IMS International Conference on Statistics and Data Science*
- August 2022 **Statistical Privacy: No Free Lunch**, *Superset Super Summit*
- August 2018 **Smoothing spline ANOVA models for nonparametric covariance estimation for longitudinal data**, *Joint Statistical Meetings*
- August 2017 **Nonparametric covariance estimation for longitudinal data via tensor product smoothing**, *Joint Statistical Meetings*
- October 2016 **The Machines Are Coming: Will Algorithms Replace Designers in the UX World?**, *Wards Auto User Experience Conference*
- August 2012 **Nonparametric Covariance Estimation for Functional Data with Shrinkage Toward Stationary Models**, *Joint Statistical Meetings*

Honors and Awards

- June 2010 Ohio State University Department of Statistics Teaching Assistant of the Year Nominee
- June 2009 Ohio State University Department of Statistics Teaching Assistant of the Year
- June 2008 Ohio State University Department of Statistics Teaching Assistant of the Year Nominee

References

Available upon Request