Kentaro Hoffman

khoffm3@uw.edu

EMPLOYMENT

University of Washington, Seattle — **Seattle, WA**
Postdoctoral Scholar — *June 1st 2023- Present*

- Advised by: Tyler McCormick

Johns Hopkins University, Baltimore — **Baltimore, MD**
Postdoctoral Scholar — *2022-2023*

- Advised by: Peter Searson and Scott Zeger

AFFILIATIONS

Center for Statistics and the Social Sciences (CSSS) at UW *Postdoctoral Scientist*
eScience Institute at UW, *UW Data Science Postdoctoral Fellow*

EDUCATION

University of North Carolina, Chapel Hill — **Chapel Hill, NC**
PhD in Statistics and Operations Research — *2017- 2022*

- Dissertation: Borrowing from Your Neighbors: Three Statistical Techniques from Nontraditional Sources
- Advised by: Kai Zhang and Cynthia Rudin

Rice University — **Houston, TX**
BA in Mathematics and Statistics — *2013- 2017*

RESEARCH INTERESTS

Inference using Synthetic Data, Domain Adaptation, Data Science, Rashomon Sets, Generative AI, Decision Theory, Global Health

PUBLICATIONS

Preprint/Under Review

Stephen Salerno, Jiacheng Miao, Awan Afiaz, Kentaro Hoffman, Anna Neufeld, Qiongshi Lu, Tyler H. McCormick, Jeffrey T. Leek (2024) **ipd: An R Package for Conducting Inference on Predicted Data** *Under Review at Biostatistics*

Stephen Salerno, Kentaro Hoffman, Awan Afiaz, Anna Neufeld, Tyler McCormick, and Jeffrey T. Leek **Sample Size Considerations for Post-Prediction Inference** *Under Review at AISTATS 2025*

Kentaro Hoffman, Stephen Salerno, Jeff T. Leek, Tyler McCormick (2024) **Some models are useful, but for how long?: A decision theoretic approach to choosing when to refit large-scale prediction models** *In preparation for the Journal of the American Statistical Association 2024 Special issue on AI*

Kentaro Hoffman, Qirui Zhao, Tyler McCormick (2024) **Two phase sampling with econoimc consid-**

erations using Post-Prediction Inference *In preparation for the Journal of the American Statistical Association 2024 Special issue on AI*

Kentaro Hoffman, Stephen Salerno, Awan Afiaz, Jeffrey T. Leek, Tyler H. McCormick (2024) **Do We Really Even Need Data?** *In preparation for Harvard Data Science Review*

Journal/Conference Publications

Kentaro Hoffman, Tyler McCormick (2024) **Bayesian Optimal Experimental Design of Streaming Data Incorporating Machine Learning Generated Synthetic Data** *Accepted at NeurIPS 2024 Workshop on Bayesian Decision-making and Uncertainty*

Simon Dovan Nguyen, Kentaro Hoffman, and Tyler McCormick **Using Rashomon Sets for Robust Active Learning** (2024) *Accepted at NeurIPS 2024 Workshop on Bayesian Decision-making and Uncertainty*

Adam Visokay, Trinity Fan, Kentaro Hoffman, Stephen Salerno, Jeffrey T. Leek, Li Liu, Tyler H. McCormick (2024) **From Narratives to Numbers: Valid Inference Using Language Model Predictions from Verbal Autopsy Narratives** *Conference on Language Models 2024*

Harsh Parikh*, Kentaro Hoffman*, Haoqi Sun*, Wendong Ge, Rajesh Amerineni, Lin Liu, Alexander Volfovsky, Sahar Zafar, Cynthia Rudin, M. Brandon Westover. (2023) **Effects of epileptiform activity on discharge outcome in critically ill patients in the USA: a retrospective cross-sectional study** *Lancet Digital Health* ,Vol. 5, Issue 8, pp. 495–e502

Hoffman, K. Lees, Johnathan, and Zhang, Kai. (2023) Local **Change Point Detection and Signal Cleaning using EEMD with applications to Acoustic Shockwaves** *Circuits Systems and Signal Processing* , Vol 42, Number 8, pp. 4669–4690

Hoffman, K. Babichev, A. and Dabaghian, Y. (2016) **A model of topological mapping of space in bat hippocampus.**. *Hippocampus* 26: 1345-1353.

Peer Reviewed short comment

Ruobin Gong, Kentaro Hoffman, Yifan Cui, and Jan Hannig. **Technical Comment on "Policy impacts of statistical uncertainty and privacy"** *Science* DOI:10.1126/science.adf9724

Hoffman, K., Hannig, J. and Zhang, Kai. (2021) **Comments on "A Gibbs sampler for a class of random convex polytopes**, *Journal of the American Statistical Association* 116:535, 1206-1210

Technical Report

Kentaro Hoffman, Tyler McCormick, and Jan Hannig. (2024) **Dempster-Shafer P-values: Thoughts on an Alternative Approach for Multinomial Inference** `https://arxiv.org/abs/2402.17070`

Press

Fitzgerald, S. (2023, July 20). New study shows association between Epileptiform Activity and Poor ICU Outcomes. *Neurology Today*

HONORS AND AWARDS

eScience Institute Azure Compute Funding ($ 6,500 Award) *2023-2024*
eScience Institute Postdoctoral Fellowship ($ 2,000 Award) *2023-2024*
SAMSI Research Assistant Fellowship *2022*
NIH BD2K Biomedical Graduate Fellow *2017-2022*

SOFTWARE

IPD (co-owner) `https://cran.r-universe.dev/ipd/doc/manual.html`
The Inference on Predicted Data is a package to allow one to perform valid statistical inference when some of the data is generated from a black-box AI model. *Avaiable from CRAN*

PRESENTATIONS

When Many Models Are Accurate: Exploring Interpretability and Uncertainty Quantification
Joint Statistical Meeting 2025 (Invited Session) *2025*

Bayesian Optimal Experimental Design of Streaming Data
NeurIPS Workshop on Bayesian Decision Making and Uncertainty *2024*

Valid Inference Using Language Model Predictions from Verbal Autopsy Narratives
eScience Institute (UW) *2024*

Causal Estimation of Seizure-Like Brain Activity
UNC BIOS/STOR Joint Seminar *2021*

Multi-Resolution Inference for Multinomial Tests of Uniformity using Dempster-Shafer
Universiy of Liverpool Institute for Risk and Uncertainty. *2021*

Multi-Resolution Inference for Multinomial Tests of Uniformity using Dempster-Shafer
UNC STOR Graduate Seminar *2021*

Causal Estimation of Seizure-Like Brain Activity
Joint Statistical Meeting, Seattle *2021*

TEACHING EXPERIENCE

CS&SS 594 A: AI and the Social Sciences (Co-instructor) **Seattle, WA**
University of Washington *2024*

PhD Qualifying Exam Recitation (Instructor) **Chapel Hill, NC**
University of North Carolina Chapel Hill *2021*

STOR 320: Introduction to Data Science (Instructional Assistant) **Chapel Hill, NC**
University of North Carolina Chapel Hill *2021*

STOR 120: Foundations of Statistics and Data Science (Primary Instructor) **Chapel Hill, NC**
University of North Carolina Chapel Hill *2020*

STOR 155: Introduction to Data Models and Inference (Primary Instructor) **Chapel Hill, NC**
University of North Carolina Chapel Hill *2019-2020*

STAT 601: Neural Machine Learning (Instructional Assistant) **Houston, TX**
Rice University *2017*

MENTORSHIP

Simon Nugyen, UW Statistics	Currently working on PhD Thesis
Qirui Zhao, UW Statistics	Currently working on MS Thesis
Adam Visokay, UW Sociology	Currently working on PhD Thesis
Trinity Fan, UW Statistics	Currently working on PhD Thesis
Awan Afiaz, UW Biostatisitics	Currently working on PhD Thesis
Jizhou Tian, JHU, Biostatistics Master Thesis,	Now at Brown Biostatistics
Mikhal Ben-Joseph, B.A UNC STOR Undergraduate Research	Now at Harvard Law

PROFESSIONAL ACTIVITIES

Programming Committee *NeurIPS workshop on BDU* *NeurIPS 2024*
Student Aide to Hiring Committee *UW 2023*
Chair "Recent developments in methods for digital Brain Health data" *JSM 2023*
Graduate Mentor *ASA DataFest 2020*
Referee Services *Management Science, NeurIPS BDU Workshop (4 times), Stat*