BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Soumik Purkayastha

eRA COMMONS USER NAME (credential, e.g., agency login): soumikp

POSITION TITLE: Assistant Professor of Biostatistics and Health Data Science

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE	Completion Date MM/YYYY	FIELD OF STUDY
St. Xavier's College (Autonomous), Kolkata, INDIA	B.Sc.	07/2017	Statistics
Indian Statistical Institute, Kolkata, INDIA	M.Stat.	07/2019	Statistics
University of Michigan, Ann Arbor, USA	M.S.	04/2021	Biostatistics
University of Michigan, Ann Arbor, USA	Ph.D.	07/2024	Biostatistics

A. Personal Statement

My primary research interest lies in understanding the causes of health disparities, including those related to pharmaceutical treatments and/or non-pharmaceutical interventions such as social and environmental exposures, and public health policy. When trying to address the excessive complexity of the underlying causes of such disparities, existing methods of analysis often fail to combine flexibility, efficiency, and power. This motivates me to pursue further methodological research on non-linear, mechanistic causal relationships that advance traditional statistical models, methods, and algorithms in causal discovery and inference. Beyond my current methodological interests, I have had many opportunities to collaborate with researchers from different fields, including epidemiologists, economists, and other biomedical researchers, primarily in the areas of analyzing imaging and biomarker data, and infectious disease modeling.

I have focused on developing novel statistical methods that project complex causal structures to low- dimensional causal estimands that are free of the fundamental assumptions of causal inference. The significant contribution of this work lies in bridging Shannon's seminal work on information theory to statistical examinations of association [1] and causality [2] in observational data. I propose new analytic methods that facilitate the study of not only association but also asymmetry (which is reflective of underlying causality) beyond the scope of traditional causal analysis. Said methods can be readily applied in practical studies with intuitive conditions and interpretations of causal mechanisms.

Additionally, I have been interested in the spatiotemporal forecasting of infectious diseases, specifically, COVID-19 in India. Using epidemiological disease transmission models [3, 4, 5], I worked on generating counterfactual evidence in support of how strengthening public health interventions early would have helped control transmission in the country and significantly reduced mortality, even without harsh lockdowns.

Citations:

1. [*PREPRINT*] **Purkayastha, S.**, & Song, P. X.-K. (2025). *A Mechanistic Framework for Collider Detection in Observational Data.* ARXIV:2502.10317
2. [*PREPRINT*] **Purkayastha, S.**, & Song, P. X.-K. (2025). *Quantification and cross-fitting inference of asymmetric relations under generative exposure mapping models.* ARXIV.2311.04696.
3. **Purkayastha, S.**, & Song, P. X.-K. (2024). *fastMI: A fast and consistent copula-based nonparametric estimator of mutual information. Journal of Multivariate Analysis*, 201, 105270.

4. Salvatore, M., **Purkayastha, S.**, Ganapathi, L., Bhattacharyya, R., Kundu, R., Zimmermann, L., Ray, D., Hazra, A., Kleinsasser, M., Solomon, S., Subbaraman, R., & Mukherjee, B. (2022). *Lessons from SARS-CoV-2 in India: A data-driven framework for pandemic resilience.* Science Advances (Vol. 8, Issue 24).
5. **Purkayastha, S.,** Bhattacharyya, R., Bhaduri, R., Kundu, R., Gu, X., Salvatore, M., Ray, D., Mishra, S., & Mukherjee, B. (2021). *A comparison of five epidemiological models for transmission of SARS-CoV-2 in India.* BMC Infectious Diseases (Vol. 21, Issue 1).
6. **Purkayastha, S.**, Kundu, R., Bhaduri, R., Barker, D., Kleinsasser, M., Ray, D., & Mukherjee, B. (2021). *Estimating the wave 1 and wave 2 infection fatality rates from SARS-CoV-2 in India.* BMC Research Notes (Vol. 14, Issue 1).

B. Positions, Scientific Appointments, and Honors

Positions and Scientific Appointments

2018 – 2018	*Intern*, Data science group, Walmart Labs, Bengaluru, KA, INDIA.
2020 – 2024	*Graduate student research assistant*, University of Michigan, Ann Arbor, MI, USA.
2021 – 2021	*Intern*, AI-ML group, Apple Inc., Cupertino, CA, USA.
2024 – Present	*Assistant Professor*, Department of Biostatistics and Health Data Science, University of Pittsburgh, PA, USA.
2024 – Present	*Research Biostatistician*, US Department of Veteran Affairs, Center for Healthcare Evaluation, Research, and Promotion, Pittsburgh Healthcare System, Pittsburgh PA, USA.
2021 – Present	*Member*, International Biometric Society, Eastern North American Region. *Member*, International Biometric Society, Western North American Region. *Member*, American Statistical Association. *Member*, Institute of Mathematical Statistics.

Honors

2019	*Sabyasachi Roy Memorial Gold Medal*, Indian Statistical Institute, WB, INDIA.
2020	*Richard G. Cornell Fellowship*, University of Michigan Ann Arbor, MI, USA.
2021, 2022, 2023	*Rackham Conference Travel Award* (awarded annually), University of Michigan Ann Arbor, MI, USA.
2023	*Rising Star Award*, University of Michigan Ann Arbor, MI, USA.
2023	*Best Paper Award*, International Biometric Society, Western North American Region.
2023 – 2024	*Rackham Predoctoral Fellowship*, University of Michigan Ann Arbor, MI, USA.

C. Contributions to Science

1. An information-theoretic framework for association and causality: my methodological research proposes novel statistical methods that study complex causal structures through low-dimensional causal estimands that are free of the fundamental assumptions of causal inference. I study Shannon's seminal work on information theory in the context of statistical examinations of causality in observational data. This framework has exciting applications in modern and prominent areas of scientific research, including mediation analysis and instrumental variables.
2. Compartmental models for infectious disease modeling: Using spatiotemporal forecasting of infectious diseases, I focused on the transmission and fallout of COVID-19 in India in collaboration with epidemiologists and economists. In addition to examining the strengths and weakness of different models, I was involved in preparing an interactive online platform offering a comprehensive view of the trajectory of the pandemic in India. This resource was crucial for effective policymaking, given the massive attention on forecasting case and death counts linked with COVID-19 and to counter the spread of misinformation.

Complete List of Published Work in MyBibliography:
https://www.ncbi.nlm.nih.gov/myncbi/soumik.purkayastha.2/bibliography/public/