

SOUMIK PURKAYASTHA

E-mail: soumikp@umich.edu

Website: soumikp.github.io

Cell phone: +1-734-881-5075

ORCID: 0000-0002-3619-2804

Education

University of Michigan, Dept. of Biostatistics

Sep. 2019 – Aug. 2024 (expected)

Ph.D. in Biostatistics, Advisor: [Peter X. K. Song](#)

University of Michigan, Dept. of Biostatistics

Sep. 2019 – Apr. 2021

M.S. in Biostatistics.

Indian Statistical Institute

July 2017 – June 2019

M.S. in Statistics, with specialization in Biostatistics.

St. Xavier's College (Autonomous), Kolkata

July 2014 – June 2017

B.Sc. (Hons) in Statistics and minors in Mathematics and Computer Science.

Publications and pre-prints

h-index: 9 ([Google scholar](#)) as of November, 2023; † denotes equal contribution.

Peer-reviewed publications

- (P13) **Purkayastha, S.** & Song, P. X. K. (2022). fastMI: a fast and consistent copula-based estimator of mutual information. *The Journal of Multivariate Analysis (to appear)*. Retrieved from <https://doi.org/10.48550/ARXIV.2212.10268>. doi: 10.48550/arXiv.2212.10268.
- (P12) 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)*. American Association for the Advancement of Science (AAAS). Retrieved from <https://doi.org/10.1126/sciadv.abp8621>. doi: 10.1126/sciadv.abp8621.
- (P11) Bhaduri, R., Kundu, R., **Purkayastha, S.**, Kleinsasser, M., Beesley, L. J., Mukherjee, B. & Datta, J. (2022). Extending the susceptible-exposed-infected-removed (SEIR) model to handle the false negative rate and symptom-based administration of COVID-19 diagnostic tests: SEIR-fansy. *Statistics in Medicine (Vol. 41, Issue 13, pp. 2317–2337)*. Wiley. Retrieved from <https://doi.org/10.1002/sim.9357>. doi: 10.1002/sim.9357.
- (P10) Zimmermann, L., Bhattacharya, S., **Purkayastha, S.**, Kundu, R., Bhaduri, R., Ghosh, P. & Mukherjee, B. (2021). SARS-CoV-2 Infection Fatality Rates in India: Systematic Review, Meta-analysis and Model-based Estimation. *Studies in Microeconomics (Vol. 9, Issue 2, pp. 137–179)*. SAGE Publications. Retrieved from <https://doi.org/10.1177/23210222211054324>. doi: 10.1177/23210222211054324.
- (P09) **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). Springer Science and Business Media LLC. Retrieved from <https://doi.org/10.1186/s13104-021-05652-2>. doi: 10.1186/s13104-021-05652-2.
- (P08) **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). Springer Science and Business Media LLC. Retrieved from <https://doi.org/10.1186/s12879-021-06077-9>. doi: 10.1186/s12879-021-06077-9.
- (P07) Salvatore, M., Basu, D., Ray, D., Kleinsasser, M., **Purkayastha, S.**, Bhattacharyya, R. & Mukherjee, B. (2020). Comprehensive public health evaluation of lockdown as a non-pharmaceutical intervention on COVID-19 spread in India: national trends masking state-level variations. *BMJ Open* (Vol. 10, Issue 12, p. e041778). BMJ. Retrieved from <https://doi.org/10.1136/bmjopen-2020-041778>. doi: 10.1136/bmjopen-2020-041778.
- (P06) **Purkayastha, S.** & Song, P. X. K. (2021). Discussion on “The timing and effectiveness of implementing mild interventions of COVID-19 in large industrial regions via a synthetic control method” by Tian et al. *Statistics and Its Interface* (Vol. 14, Issue 1, pp. 21–22). International Press of Boston. Retrieved from <https://doi.org/10.4310/20-sii652>. doi: 10.4310/20-SII652.
- (P05) Tang, L., Zhou, Y., Wang, L., **Purkayastha, S.**, Zhang, L., He, J., Wang, F. & Song, P. X. K. (2020). A Review of Multi-Compartment Infectious Disease Models. *International Statistical Review* (Vol. 88, Issue 2, pp. 462–513). Wiley. Retrieved from <https://doi.org/10.1111/insr.12402>. doi: 10.1111/insr.12402.
- (P04) **Purkayastha, S.**, Salvatore, M. & Mukherjee, B. (2020). Are women leaders significantly better at controlling the contagion during the COVID-19 pandemic? *Journal of health and social sciences*, 5(2), 231–240. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7457824>.
- (P03) Zhou, Y., Wang, L., Zhang, L., Shi, L., Yang, K., He, J., Zhao, B., Overton, W., **Purkayastha, S.** & Song, P. X. K. (2020). A Spatiotemporal Epidemiological Prediction Model to Inform County-Level COVID-19 Risk in the United States. *Harvard Data Science Review*. Retrieved from <https://doi.org/10.1162/99608f92.79e1f45e>. doi: 10.1162/99608f92.79e1f45e.
- (P02) Ray, D., Salvatore, M., Bhattacharyya, R., Wang, L., Du, J., Mohammed, S., **Purkayastha, S.**, Halder, A., Rix, A., Barker, D., Kleinsasser, M., Zhou, Y., Bose, D., Song, P. X. K., Banerjee, M., Baladandayuthapani, V., Ghosh, P. & Mukherjee, B. (2020). Predictions, Role of Interventions, and Effects of a Historic National Lockdown in India’s Response to the COVID-19 Pandemic: Data Science Call to Arms. *Harvard Data Science Review, (Special Issue 1)*. Retrieved from <https://doi.org/10.1162/99608f92.60e08ed5>. doi: 10.1162/99608f92.60e08ed5.
- (P01) Giri, S., **Purkayastha, S.**, Hazra, S., Chanda, A., Das, I. & Das, S. (2020). Prediction of monthly Hilsa (*Tenualosa ilisha*) catch in the Northern Bay of Bengal using Bayesian struc-

tural time series model. *Regional Studies in Marine Science* (Vol. 39, p. 101456). Retrieved from <https://doi.org/10.1016/j.rsma.2020.101456>. doi: 10.1016/j.rsma.2020.101456.

Under preparation/submission

- (S04) **Purkayastha, S.** & Song, P. X. K. (2023). Generative causality: using Shannon’s information theory to infer underlying asymmetry in causal relations. *arXiv*. Retrieved from <https://doi.org/10.48550/ARXIV.2311.04696>. (*Under submission*)
- (S03) Zhang, L., **Purkayastha, S.**, Kirsner, R., Spino, C. & Song, P. X. K. (2023) Determinants of Enrollment in 284 Clinical Trials for Healing Diabetic Foot Ulcers: A Systematic Review. (*Under preparation*)
- (S02) **Purkayastha, S.** & Song, P. X. K. (2023). Asymmetric predictability in causal discovery: an information theoretic approach. *arXiv*. Retrieved from <https://doi.org/10.48550/ARXIV.2210.14455>. (*Technical report*)
- (S01) **Purkayastha, S.** & Basu, A. (2022+). On minimum Bregman divergence inference. *arXiv*. Retrieved from <https://doi.org/10.48550/ARXIV.2008.06987>. (*Under review*)

Professional Experience

- (PE3) Michigan Medicine, Ann Arbor, USA. **Research Assistant** *May 2020 – Aug. 2023*
Performed statistical analyses in SAS and R for the NIH-funded [Diabetes Foot Ulcer Consortium](#). Built an automated data-pooling and analysis pipeline and an RShiny-based dashboard for fast and accessible dissemination of interactive Plotly visualization and model findings.
- (PE2) **Apple Inc.**, Cupertino, USA. **AI-ML Intern** *May 2021 – Aug. 2021*
Developed Pytorch-based natural language models to analyze user speech patterns. Built multi-level predictors of user search intent in Python to improve data quality for algorithm training and evaluation.
- (PE1) **Walmart Labs**, Bangalore, IND. **Statistician Intern** *May 2018 – July 2018*
Worked on data query and analysis of very large data sets and improved existing online grocery forecasting models in R and C++.

Teaching Experience

- (TE2) Lecturer, Summer Institute in Biostatistics Program, University of Michigan. *July 2022*
Linear regression (39 students). Materials: [slides](#) and [handout](#).
Probability theory (39 students). Materials: [slides](#) and [handouts](#).
- (TE1) Teaching assistant, BIOSSTAT 802, University of Michigan. *Jan. – Apr. 2021*
Advanced inference II (25 students).

Honors and Awards

Excellence in Research (Honorable Mention) *Oct., 2023*
Department of Biostatistics, University of Michigan

Rackham Conference Travel Award (Annual) <i>University of Michigan</i>	<i>2021, 2022, 2023</i>
Rackham Predoctoral Fellowship <i>University of Michigan</i>	<i>Sep. 2023 – Aug. 2024</i>
Best Paper and Runner-up Presentation Awards <i>WNAR of the International Biometrics Society</i>	<i>June 2023</i>
Rising Star Award <i>School of Public Health, University of Michigan</i>	<i>Apr. 2023</i>
Richard G. Cornell Fellowship <i>Department of Biostatistics, University of Michigan</i>	<i>Dec. 2020</i>
Michigan Data Science Challenge Winner <i>Michigan Institute for Data Science, University of Michigan</i>	<i>Dec. 2019</i>
Sabyasachi Roy Memorial Gold Medal <i>Indian Statistical Institute</i>	<i>July 2019</i>
Scholarship for academic performance <i>Indian Statistical Institute</i>	<i>2017 – 2019</i>

Professional Service

Journal peer review: Annals of Applied Statistics (2022), New England Journal of Statistics in Data Science (2022), and PLOS One (2021).

Professional affiliations: International Biometric Society, Western North American Region (WNAR) (2022+), American Statistical Association (2021+), Institute of Mathematical Statistics (2021+), International Biometric Society, Eastern North American Region (ENAR) (2021+)

University service: [STATCOM](#): Statistics in the community, Co-president (2022-23); Leadership team member (2022 +); [Michigan Institute of Data Science \(MIDAS\)](#) Student Organizations Council: Leadership team member (2023+); Department of Biostatistics Seminars and Brown Bag Committee (2020-2022)

Conferences & Workshops

Conferences

- (C03) Joint Statistical Meetings (August 2023). “Asymmetric predictability in causal discovery: an information theoretic approach.”
- (C02) Western North American Region (WNAR) of the International Biometric Society Spring Meeting (June 2023). “Asymmetric predictability in causal discovery: an information theoretic approach.”
- (C01) Eastern North American Region (ENAR) of the International Biometric Society Spring Meeting (March 2023). “An information-theoretic framework for causal discovery in epigenetic data”

Workshops

- (W01) Foundations of Causal Graphical Models and Structure Discovery. *Travel award provided by the National Science Foundation (NSF DMS-2227849) and Texas A&M Institute for Applied Mathematics and Computational Science (IAMCS).*

Selected Press

Bastian, H., *Women Versus Men Leaders in the Pandemic: An Update and Dig Into the Latest Data*. **PLOS Blogs**, 2022. Retrieved from <https://absolutelymaybe.plos.org/2022/06/28/women-versus-men-leaders-in-the-pandemic-an-update-and-dig-into-the-latest-data/>.

Laguipo, A., *A study of the COVID fatality rates in India during waves 1 and 2*. **News-Medical.Net**, 2021. Retrieved from <https://www.news-medical.net/news/20210603/A-study-of-the-COVID-fatality-rates-in-India-during-waves-1-and-2.aspx>.

Mukherjee, B., **Purkayastha, S.**, Salvatore, M., Mishra, S. *Underreporting does hurt the COVID fight*. **The Hindu**, 2021. Retrieved from <https://www.thehindu.com/opinion/lead/under-reporting-does-hurt-the-covid-fight/article34474676.ece>.

Ellis-Petersen, H. *India's 1.3bn population locked down to beat coronavirus*. **The Guardian**, 2021. Retrieved from <https://www.theguardian.com/world/2020/mar/24/indias-13bn-population-locked-down-to-beat-coronavirus>.

Basu, D., Salvatore, M., Kleinsasser, M., Purkayastha, S., Bhattacharya, R., and Mukherjee, B., *We're Focusing on National Data on COVID-19 When We Should Be Looking at State-Level Trends* **The Wire**, 2020. Retrieved from <https://thewire.in/health/covid-19-india-national-data-state>.