SOUMIK PURKAYASTHA

E-mail: soumikp@umich.edu **Cell phone:** +1-734-881-5075
Website: soumikp.github.io **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 October, 2023; † denotes equal contribution.

Peer-reviewed publications

(P13) **Purkayastha, S.**, & Song, P. X. K. (2023). fastMI: a fast and consistent copula-based estimator of mutual information. *The Journal of Multivariate Analysis (to appear)*. Retrieved from http://arxiv.org/abs/2212.10268 doi: 10.48550/arXiv.2212.10268

(P12) Salvatore, M., **Purkayastha, S.**, Ganapathi, L., Bhattacharyya, R., Kundu, R., Zimmermann, L., ... Mukherjee, B. (2022). Lessons from SARS-CoV-2 in India: A data-driven framework for pandemic resilience. *Science Advances*, *8*(24). Retrieved from https://www.science.org/doi/full/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*, *41*(13), 2317–2337. Retrieved from https://onlinelibrary.wiley.com/doi/abs/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*, *9*(2), 137–179. 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, July). Estimating the wave 1 and wave 2 infection fatality rates from SARS-CoV-2

in India. *BMC Research Notes*, *14*(1), 262. 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., ... Mukherjee, B. (2021). A comparison of five epidemiological models for transmission of SARS-CoV-2 in India. *BMC Infectious Diseases*, *21*(1), 533. 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*, *10*(12). Retrieved from https://bmjopen.bmj.com/content/10/12/e041778 doi: 10.1136/bmjopen-2020-041778

(P06) **Purkayastha, S.**, & Song, P. (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*, *14*(1), 21–22. Retrieved from https://dx.doi.org/10.4310/20-SII652 doi: 10.4310/20-SII652

(P05) Tang, L., Zhou, Y., Wang, L., **Purkayastha, S.**, Zhang, L., He, J., ... Song, P. X.-K. (2020). A Review of Multi-Compartment Infectious Disease Models. *International Statistical Review*, *88*(2), 462–513. Retrieved from https://onlinelibrary.wiley.com/doi/abs/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., ... Song, P. (2020). A Spatiotemporal Epidemiological Prediction Model to Inform County-Level COVID-19 Risk in the United States. *Harvard Data Science Review*(1). Retrieved from https://hdsr.mitpress.mit.edu/pub/qqg19a0r/release/1 doi: 10.1162/99608f92.79e1f45e

(P02) Ray, D., Salvatore, M., Bhattacharyya, R., Wang, L., Du, J., Mohammed, S., ... 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*, *2020*(1). Retrieved from https://hdsr.mitpress.mit.edu/pub/r1qq01kw/release/8 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 structural time series model. *Regional Studies in Marine Science*, *39*, 101456. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S2352485520305843 doi: 10.1016/j.rsma.2020.101456

Under prepararation/submission

(S02) **Purkayastha, S.**, & Song, P. X. K. (2022). *Asymmetric predictability in causal discovery: an information theoretic approach.* arXiv. Retrieved from http://arxiv.org/abs/2210.14455 doi: 10.48550/arXiv.2210.14455

(S01) **Purkayastha, S.**, & Basu, A. (2020). On minimum bregman divergence inference. *Metrika (under review)*. Retrieved from https://arxiv.org/abs/2008.06987 doi: 10.48550/arXiv.2008.06987

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 Consortium. Built and maintained an **automated data-pooling and analysis pipeline** and an **RShiny**-based dashboard for faster dissemination of interactive **Plotly visualization** and **model-based** findings that is **accessible to clincians**.

(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. Built Siri Search products by implementing **semi-supervised language models** on partially labelled user data in **Python**.

(PE1) **Walmart Labs**, Bangalore, IND. **Statistician intern** *May 2018 – Jul. 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. *Jul. 2022*
Linear regression (39 students). Materials: slides and handout.
Probability theory (39 students). Materials: slides and handouts.

(TE1) Teaching assistant, BIOSTAT 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)** *2021, 2022, 2023*
  University of Michigan
- **Rackham Predoctoral Fellowship** *Sep. 2023 – Aug. 2024*
  University of Michigan
- **Best Paper** and **Runner-up Presentation** awards *Jun. 2023*
  WNAR of the International Biometrics Society

- **Rising Star Award** *Apr. 2023*
  School of Public Health, University of Michigan
- **Richard G. Cornell Fellowship** *Dec. 2020*
  Department of Biostatistics, University of Michigan
- **Michigan Data Science Challenge Winner** *Dec. 2019*
  Michigan Institute for Data Science, University of Michigan
- **Sabyasachi Roy Memorial Gold Medal** *Jul. 2019*
  Indian Statistical Institute
- **Scholarship for academic performance** *2017 – 2019*
  Indian Statistical Institute

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 - 2024); Department of Biostatistics Seminars and Brown Bag Committee (2020-2022)

Selected Press

- Bastian, H., *Women Versus Men Leaders in the Pandemic: An Update and Dig Into the Latest Data.* **PLOS Blogs**, 2022. url: 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. url: 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. url: 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. url: 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. url: https://thewire.in/health/covid-19-india-national-data-state