Bi-monthly Newsletter of April & May 2022

Founded in 1935, GW's Department of Statistics was the first statistics department in a College of Arts and Sciences in the United States. Since then, the department has continually produced high quality research in both theoretical and applied areas of statistics and probability. The department offers both M.S. and Ph.D. programs of full-time and part-time enrollments. The great location of being in the heart of Washington, D.C. gives students unique access to the research and employment opportunities in the major government agencies and private companies and connection with the distinguished statisticians from the local Washington Statistical Society.

Message from the Editor

It is a great pleasure to introduce the Newsletter to all the students in the statistics department, and we thank the professors of the department for providing constructive support to the successful launch of the Newsletter. The Newsletter is created with the intention to help students be informed and engaged with the statistical events locally or nationwide, to introduce collaboration opportunities with other researchers, to celebrate the achievements and milestones of our faculties and students. The Newsletter will be issued every two months. Meanwhile, we encourage students to promote their works and establish professional networks through online platforms such as GW blogs and Twitter. We invite students to join the editorial team and contribute to the Newsletter. Ultimately, we hope every student has an enriched and great experience in the statistics department and we all become better and stronger as a whole community.

Thank you, we hope you find the Newsletter informative.
Xiang Li (Editor/Co-Chair of PhD Student Committee)

Statistics Department Events

GWU STAT DAY. April 8-9th
The GWU statistics department will hold the STAT DAY event on April 8th and 9th. Students are invited to attend and register for the short course and presentations.

➢ **Keynote speaker** Dr. Dean Follmann will give a colloquium talk on April 8th (11am - noon). Dr. Follmann has served as Chief of the Biostatistics Research Branch at the National Institute of Allergy and Infectious Diseases (NIAID) since 2002. He is an elected Fellow of the American Statistical Association. His current research interests focus on statistical methods related to immunology and vaccinology.
   Title: Statistical Issues in COVID-19 Vaccine Trials - The Past and The Future
   Abstract: COVID-19 vaccines were developed at an amazing pace which required fast decisions about statistical approaches to design and analysis. While very high efficacy was initially demonstrated for multiple vaccine platforms, the durability of vaccination and the effect against variants remain ongoing questions. In this talk we provide an overview of the US government supported COVID-19 vaccine trials with an emphasis on emerging statistical issues. Topics include the basic design of the trials, durability of vaccine efficacy, the effect of boosting, and the role of variants.

➢ **Short course** on Friday, April 8th (2p - 5p) & Saturday, April 9th (10a - 1p). Dr Rajeshwari (Raji) Sundaram, a senior investigator of biostatistics at National Institutes of Health, will give a short course on Practical Statistical Methods to Assess Health Effects in the Context of Multi-pollutant Mixtures. Register the course before **April 4th**.

➢ **Student presentation** on Saturday, April 9th afternoon. Students are strongly encouraged to present an oral or poster presentation. Please submit your abstract before **April 2nd**.

Departmental Graduation Celebration. Thursday, May 12th, 6:30 PM. City View Room
The statistics department will hold a celebration event for the 2022 graduating class. Pre-registration is required, and more details will be sent out by email soon.

Statistics Conferences
Some international and area statistical conferences will take place in Summer, with registration deadlines occurring earlier in May. Students are encouraged to attend these conferences.

➢ 2nd Annual Graduate Student Research Conference - NISS Graduate Student Network Event will take place on May 14 and 15, 2021, from noon - 5 pm ET and will feature graduate student presentations, invited speakers and a social networking hour. Abstract submission deadline is **April 15th** at 5 pm ET.

➢ The Joint Statistical Meetings (JSM) is the largest gathering of statisticians and data scientists held in North America. Early registration opens on May 2 and closes on **May 31**. Join the ASA membership to be entitled for the student membership fees.

➢ SAE 2022: Small Area Estimation, Surveys and Data Science will be hosted by the University of Maryland between May 23rd and 27th. Early registration ends on **March 31**.

➢ NESS 2022: Statistics, The Heart of Data Science (35th New England Statistics Symposium) will be hosted by the University of Connecticut on May 22 - 25, 2022. Registration ends on **May 9th** (in-person participation) and **May 16th** (virtual participation). Student Research Awards and Student Poster Awards will be given, paper and poster need to be submitted before May 9th.

Career Events

GW career workshops and opportunities. Check the Handshake website for more events

Friday-Saturday, April 1-2, 7:00PM 10:00 PM Google Code Jam.

Monday, April 4, 7:00 PM - 8:00 PM Finding your leadership style.

Wednesday, April 6, 12:00pm - 1:00pm Publishing in your Field: Tips for navigating the publication process.

Thursday, April 7, 8:00 AM - Friday, April 8, 5:00 PM Wells Fargo CIB 2022 Women on Wall Street.

Wednesday, April 13, 7:30 PM - 8:30 PM Codepath's technical interview prep.

Events in D.C.

Data Science DC (online)

Local group Data Science DC will host two online events in April and May. Check more details and register the event on their website.

➢ Tuesday, April 12 from 6:00 PM to 7:30 PM: "Network-based Trajectory Analysis of Topic Interactions for COVID-19 Literature". Check more details and register on the webpage.

➢ Tuesday, Mayl 10 from 6:00 PM to 7:30 PM: "Scaling and Distributing Python & ML Applications with Ray". Check more details and register on the webpage.

Kudos & Milestones

✋Congratulations to Professor Fang Jin and Professor Xiaoke Zhang for their interdisciplinary team being awarded a $2.2 million grant from the National Institutes of Health.

✋Congratulations to Professor Qing Pan for winning the 2022 Gertrude Cox award and being elected as the WSS program Chair Elect.

✋Congratulations to our PhD candidate Rui Miao for winning the student paper award in the 2021 ICSA Applied Statistics Symposium.

✋Congratulations to our PhD candidate Xiang Li for winning the 2022 Washington Statistical Society Student Travel Award.

✋Congratulations to our PhD graduate Yifan Zhou and PhD candidate Yang Liu for their paper "Adaptive A/B Test on Networks with Cluster Structures" being accepted to the 2022 Artificial Intelligence and Statistics (AISTAT) conference.

⚖Ph.D. graduate Panpan Zhang will join the tenure track position at Vanderbilt University.

⚖Ph.D. graduate Xiang Peng defended his dissertation on December 2021 and joined the Merck thereafter.

⚠Ph.D. candidates Mingze Zhang, Yang Liu, Shuyang Gao, Shunyan Luo, Yufei Guo, Bingqi Han have successfully defended their dissertations in March.

Reading Recommendations

We are grateful to Professors Tapan Nayak and Xiaoke Zhang for their encouragement and reading recommendations for this Newsletter.

Professor Tapan Nayak: Professor Nayak suggests an exercise (see the appendix or the google drive) that concerns some important and familiar statistics topics - sufficiency, factorization theorem and conditional distribution. He kindly offers to hold a discussion session if students are interested.

Professor Xiaoke Zhang: Professor Zhang recommends two papers on causal inference. For more information on Professor Zhang's research works, please visit his personal website or the appendix.

> 1. Zhang, Xiaoke, Wu Xue, and Qiyue Wang. "Covariate balancing functional propensity score for functional treatments in cross-sectional observational studies." *Computational Statistics & Data Analysis* 163 (2021): 107303. https://doi.org/10.1016/j.csda.2021.107303
> **Remark:** This is the first work to study causal effect estimation in observational studies when the treatment is a stochastic process/random function. This paper addresses the non-existence of the propensity score of the functional treatment to balance confounding factors for causal effect estimation.
>
> 2. Qi, Zhengling, Rui Miao, and Xiaoke Zhang. "Proximal Learning for Individualized Treatment Regimes Under Unmeasured Confounding." *arXiv preprint arXiv:2105.01187* (2021). https://doi.org/10.48550/arXiv.2105.01187
> **Remark:** The optimal individualized treatment regime (ITR) is typically not identifiable when some covariates are unobservable. This is the first work that generalizes the recently proposed proximal causal inference framework to ITR identification and learning with unmeasured confounding.

Contact Us

American Statistical Association (ASA) Student Chapter
Students from the GWU Statistics department are actively organizing and participating in statistical events through the GWU ASA Student Chapter. The purpose of the student chapter is to provide students at the George GWU with opportunities to connect with other students interested in statistics and interact with prominent statisticians locally and at national meetings. If you are interested to be an officer or participate in the event, please contact Jiaqian Yu (jiaqian7@gwu.edu) or Lei Song (leisong@gwu.edu).

Contribute to the Newsletter

We will issue the next Newsletter on June 1st, 2022. If you would like to contribute to the sections `Research Spotlight` or `Reading Recommendations`, or want to join the editorial team for the Newsletter, please contact Xiang Li (xiangli@gwu.edu).

Full-time Faculty & Their Research Interests

Tatiyana Apanasovich
Measurement error models, Spatial statistics, Positive definite kernels, Non/Semiparametric regression, Vector valued random fields.

Srinivasan Balaji
Diffusion processes, Markov chains, Stochastic differential equations and applications.

Sudip Bose
Bayesian statistics, Bayesian robustness, Pitman closeness.

Joseph L. Gastwirth
Statistics in law and public policy, Robust statistical methods, Grouped data.

Feifang Hu
Adaptive design of clinical trials, Bioinformatics, Biostatistics, Bootstrap methods, Statistical issues in personalized medicine, Statistical methods in financial econometrics, Stochastic process.

Fang Rachel Jin
Deep learning, Natural language processing, Data mining, Machine learning, Social network analysis.

Subrata Kundu
Sequential analysis, Density estimation, Software reliability, Hypothesis testing, Nonparametric statistics.

Yinglei Lai
Bioinformatics, Computational biology, Statistical genetics.

Joshua Landon
Bayesian statistics, Stochastic processes, Markov chain, Monte Carlo methods, Decision analysis, Reliability.

Zhaohai Li
Genetic epidemiology, Statistical methods for meta analysis, Empirical Bayes, Methods for clinical trials, Clinical biostatistics.

Hua Liang
Partially linear models, High-dimensional semiparametric modeling, Model averaging and model selection, Longitudinal data analysis, Measurement error models, Nonlinear and nonparametric mixed effect models, HIV/AIDS clinical trial and dynamic modeling.

Hosam M. Mahmoud
Probabilistic analysis of algorithms, Random discrete structures, Analytic probability.

Reza Modarres
Statistical computing, Multivariate analysis, Environmental statistics, Nonparametric statistics.

Tapan Nayak
Inference, Prediction, Software reliability, Randomized response design.

Qing Pan
Survival analysis, Recurrent event data, Observational studies, Costs analysis.

Huixia Judy Wang
Quantile regression; Extreme value theory and applications; Bioinformatics; Nonparametric (Semiparametric) regression; Inference; Variable selection; Survival analysis; Longitudinal data analysis; Spatial analysis; Measurement error; Missing data

Lin Wang
Experimental design, Computer experiments, Big data reduction, Statistical learning.

Xiaoke Zhang
Functional data analysis, longitudinal data analysis, smoothing, dimension reduction, nonparametric statistics, statistical learning, and neuroimaging.

Exercise suggested by Professor Nayak.

An Exercise

1. Let $X_1, \ldots, X_n$ be a random sample from a discrete uniform distribution with pmf

$$f_\theta(x) = \frac{1}{\theta}, \quad x = 1, 2, \ldots, \theta,$$

where $\theta \in \mathbb{Z}^+ = \{1, 2, \ldots\}$ is an unknown parameter. Let $\mathbf{X} = (X_1, \ldots, X_n)$ and $T(\mathbf{X}) = \max\{X_1, \ldots, X_n\}$

(a) Prove using the Factorization Theorem that T is a sufficient statistic.

(b) Since T is a sufficient statistic, the conditional distribution of $\mathbf{X}$ given T must be independent of θ, i.e., for any fixed $\mathbf{x} = (x_1, \ldots, x_n)$ and t, with $x_i \in \mathbb{Z}^+, i = 1, \ldots, n$, and $t \in \mathbb{Z}^+$, $P_\theta(X_1 = x_1, \ldots, X_n = x_n | T = t)$, denoted $h_\theta(\mathbf{x}|t)$, must be the same for all $\theta \in \mathbb{Z}^+$. Can you prove this directly? In particular, answer the following two questions.

(i) For $n = 4, x_1 = 5, x_2 = 6, x_3 = 2, x_4 = 8$ and $t = 8$, can you show that $h_\theta(\mathbf{x}|t)$ is the same for $\theta = 100, 12$ and 5?

(ii) For $n = 4, x_1 = 5, x_2 = 8, x_3 = 2, x_4 = 6$ and $t = 6$, can you verify that $h_\theta(\mathbf{x}|t)$ for $\theta = 100, 12, 5$ are equal?

(c) In parts (i) and (ii) of (b), how did (or can) you handle the case of $\theta = 5$?

(d) Read the proof Theorem 6.2.6 (for discrete distributions) in Casella-Berger book. Try to apply the arguments in the proof to establish that $h_\theta(\mathbf{x}|t)$ is the same for all $\theta \in \mathbb{Z}^+$ in the two cases considered above. Does that work? Is the proof correct?

(e) What makes defining sufficiency and proving the Factorization Theorem difficult even for discrete distributions? What approaches can you suggest for overcoming the difficulty?

More information kindly offered by Professor Zhang.

Dr. Xiaoke Zhang joined GW in 2017 as an Assistant Professor of Statistics. Before joining GW, he was Assistant Professor of Statistics in the University of Delaware. He earned his BS from Peking University, China and PhD from UC Davis. Zhang's research areas in statistics include functional data analysis, causal inference, and individualized treatment regimes/reinforcement learning. Some of his work has been published in top statistics journals, such as the *Journal of the American Statistical Association*, *Annals of Statistics* and *Biometrika*. Dr. Zhang has also been active in interdisciplinary research with collaborators in a variety of fields, including neuroscience, animal science, sociology, psychology, and computer science among others. Zhang's research has been continuously funded by NSF, NIH and internal grants.

Here are his recent works and grants within his research areas:

1. Functional Data Analysis.
- Miao, Rui, Xiaoke Zhang, and Raymond KW Wong. "A Wavelet-Based Independence Test for Functional Data with an Application to MEG Functional Connectivity." *Journal of the American Statistical Association,* in press, (2021). https://doi.org/10.1080/01621459.2021.2020126
- Zhang, Xiaoke, Wu Xue, and Qiyue Wang. "Covariate balancing functional propensity score for functional treatments in cross-sectional observational studies." *Computational Statistics & Data Analysis* 163 (2021): 107303. https://doi.org/10.1016/j.csda.2021.107303
- Co-Investigator: NIH/NIDDK, R01DK129809, Advancing 3D Optical Body Surface Scan Technology to Assess Physiological and Psychological Effects in Highly Obese Population, $2,232,783, 09/01/2021– 08/31/2025

2. Causal Inference.
- Zhang, Xiaoke, Wu Xue, and Qiyue Wang. "Covariate balancing functional propensity score for functional treatments in cross-sectional observational studies." *Computational Statistics & Data Analysis* 163 (2021): 107303. https://doi.org/10.1016/j.csda.2021.107303
- Miao, Rui, Wu Xue, and Xiaoke Zhang. "Average Treatment Effect Estimation in Observational Studies with Functional Covariates." *Statistics and Its Interface* 15, no. 2 (2022): 237-246. https://dx.doi.org/10.4310/20-SII632
- Principal Investigator: George Washington University University Facilitating Fund (UFF), RKHS-Based Covariate Balancing for Functional Treatments in Observational Studies, $17,582, 07/01/2021–06/30/2022.

3. Individualized Treatment Regimes/Reinforcement Learning.
- Qi, Zhengling, Rui Miao, and Xiaoke Zhang. "Proximal Learning for Individualized Treatment Regimes Under Unmeasured Confounding." *arXiv preprint arXiv:2105.01187* (2021). https://doi.org/10.48550/arXiv.2105.01187