Introductions for the Data Privacy and Confidentiality Event
September 11, 2021, 9:00 a.m. to 12:00 p.m., PST

Speaker Biographies and Topics

#1 Dr. Darren Toh

Biography: Darren Toh, ScD is Professor in the Department of Population Medicine at Harvard Medical School and Harvard Pilgrim Health Care Institute. He is a pharmacoepidemiologist with an interest in the comparative safety and effectiveness research of medical products.

His research focuses on 1) assessing the risks and benefits of medical products using electronic data collected as part of routine healthcare delivery, and 2) developing and applying privacy-protecting analytic methods to conduct multi-center studies in distributed data networks.

Darren is Chief Scientist at the Operations Center of the FDA-funded Sentinel System, a congressionally mandated national medical product safety surveillance system. He is Principal Investigator of projects funded by the National Institutes of Health, the Agency for Healthcare Research and Quality, the Patient-Centered Outcomes Research Institute, and the Food and Drug Administration. Darren received his doctoral degree in Epidemiology from the Harvard School of Public Health.

Title: Protecting data privacy in multi-center studies: Experience from an epidemiologist.

#2 Dr. Thomas Kent

Biography: Dr. Thomas Kent received his Ph.D. in mathematics from the University of Wisconsin – Madison in 2006. He also received a master’s degree in Computer Science from the University of Wisconsin – Madison and a Master’s degree in statistics from Pennsylvania State University.

After Graduate School, he was a visiting assistant professor at Brightman Young University, and then spent two years at the University of Siena, Italy as a European Union Marie Curie Fellow.

He was at Marywood University in Scranton, Pennsylvania for 9 years where he earned tenure and was the chair of the Department of Mathematics and Computer Science. Deciding he wanted to try something new, he moved to State College, Pennsylvania for 3 years where he studied statistics and worked on the project which is the subject of his talk. He is currently a statistician for the USDA in Ames, Iowa.

Title: An R package to Implement Secure Multiparty Distributed Regression on Vertically Partitioned Data.

Abstract: We present the VDRA R package which is available on Github and on CRAN. This package implements linear, logistic, and cox regression on vertically partitioned data across several data partners with the aid of a trusted third party (analysis center). Data is not shared.
between data partners, or the analysis center and the computations can be considered secure. Three different protocols have been implemented: 2-Party, 2^T-Party, and K^T-Party. In 2-Party, two data partners which communicate directly without an intermediate analysis center; In 2T-Party, two data partners communicate indirectly via an analysis center, and KParty: two or more data partners plus an analysis center are all allowed to communicate directly. 2-Party and 2^T-Party use a form of secure multiplication as found in Karr, et. Al (2009) and Slavkovic et. al. (2007). Full details can be found in Samizo (In preparation). This project was funded by the National Institutes of Health Grant U01EB023683.

---

#3. Dr. Aleksandra (Seša) Slavković

Biography: Aleksandra (Seša) Slavković is a Professor of Statistics and Associate Dean for Graduate Education in Eberly College of Science at Penn State. Her research focuses on methodological developments in data privacy and confidentiality in the context of small- and large-scale surveys, health, genomic, and network data, including differential privacy, and broad data access that guarantees the level of accurate statistical inference needed to support reliable science and policy.

Other research interests include evaluation methods for human performance in virtual environments, application of statistics to information sciences and social sciences, algebraic statistics, and causal inference.

She earned her PhD in Statistics from Carnegie Mellon University. She is a fellow of the American Statistical Association, Institute of Mathematical Statistics, and the International Statistical Institute. Dr. Slavković is an associate editor of the Annals of Applied Statistics and Journal of Privacy and Confidentiality and an editor-elect for Statistics and Policy. She served as a chair of the ASA Privacy and Confidentiality committee and is Chair-Elect 2021 for the ASA Social Statistics Section, and currently serves on a half dozen advisory committees.

**Topic: Valid statistical inference with privacy constraints**

Abstract: Limiting the disclosure risk of sensitive data and statistical analyses is a long-standing problem in statistics. Differential privacy (DP) provides a framework for a strong provable privacy protection against arbitrary adversaries while allowing the release of summary statistics and potentially synthetic data. DP methods/mechanisms require the introduction of randomness which reduces the utility of the results especially in finite samples. In this talk we give an overview of statistical data privacy and its links to DP. We also describe a general framework, built on sound statistical principles from measurement error, robustness, and the likelihood-based inference, and give specific examples of how to achieve optimal statistical inference under formal privacy, focused on survey and census data.

---

#4 Dr. Jordan Awan

Biography: Dr. Jordan Awan is currently an Assistant Professor of Statistics at Purdue University. Before joining Purdue, he completed his PhD in Statistic at Penn State University under the advisement of Drs. Aleksandra Slavkovic and Matthew Reimherr. Jordan's research
interests are in differential privacy, synthetic data, and functional data analysis, as well as applied problems in the analysis of physiological signals. Dr. Awan primarily works in differential privacy, which is the primary method currently being used by the U.S. Census, Apple, and Google.

**Title: Canonical Noise Distributions and Private Hypothesis Tests**

Abstract: $f$-DP has recently been proposed as a generalization of classical definitions of differential privacy allowing a lossless analysis of composition, post-processing, and privacy amplification via subsampling. In the setting of $f$-DP, we propose the concept **canonical noise distribution** (CND) which captures whether an additive privacy mechanism is appropriately tailored for a given $f$, and give a construction that produces a CND given an arbitrary tradeoff function $f$. We show that private hypothesis tests are intimately related to CNDs, allowing for the release of private $p$-values at no additional privacy cost as well as the construction of uniformly most powerful (UMP) tests for binary data.

We apply our techniques to the problem of difference of proportions testing and construct a UMP unbiased “semi-private" test which upper bounds the performance of any DP test. Using this as a benchmark we propose a private test, based on the inversion of characteristic functions, which allows for optimal inference for the two population parameters and is nearly as powerful as the semi-private UMPU. When specialized to the case of $(\epsilon,0)$-DP, we show empirically that our proposed test is more powerful than any $(\epsilon/\sqrt{2})$-DP test and has more accurate type I errors than the classic normal approximation test.