RUTGERS
School of Arts and Sciences

Department of Statistics

Ruobin Gong
Department of Statistics
Rutgers University

A Refreshment Stirred, Not Shaken (I): Building Blocks of Differential Privacy; and (II): Can Swapping be Differentially Private?

Wednesday, March 27, 2024
11:50 AM
96 Frelinghuysen Road, CoRE Building, Room 431

Zoom Meeting: Meeting ID: 969 0606 4706
Password: 745339
https://rutgers.zoom.us/j/96906064706?pwd=ZklvbExpRVBJQ3c5dUhhYTFuR2ZrZz09

Light refreshments will be served in Hill 452, 11:15am

Abstract: Data Swapping is a statistical disclosure control (SDC) procedure that randomly perturbs tabular data subject to common features. It is the predominant SDC method used in the 1990 through 2010 U.S. Decennial Censuses. For the 2020 Decennial, swapping was deprecated in favor of a suite of methods that "us[e] differential privacy for privacy-loss accounting" (Abowd et al., 2022) which, if one reads between the lines, is not the same as saying that the methods are differentially private. What are we missing?

This talk presents a two-part inquiry that harmonizes the conflict between the desirable principles brought to bear by formal privacy standards (notably differential privacy) on SDC, and the pragmatic constraints that an SDC implementation must respect within its context of application. Part (I) traces the etymological evolution of differential privacy from Dwork (2006) to its myriad variations seen the literature today. It reveals the constituent elements of a differential privacy specification to be choices along the dimensions of "who" (protection domain), "where" (scope of protection), "what" (protection unit), "how" (standard of protection), and "how much" (privacy-loss accounting). Part (II) delineates the privacy parameters of a randomized swapping algorithm, constructed to be compatible with known features of the 1990 through 2010 Decennial Census swapping procedure, and analyzes its privacy guarantee alongside that offered by the 2020 Decennial Census TopDown procedure.

This study makes precise the vocabulary necessary to confer benefits of formal privacy, including provability and transparency, to a wide range of SDC practices that originated outside of the differential privacy literature with minimal hinderance to the operations of statistical data curators and the usability of their data products. It underscores the deficiency in the common tendency to simplistically equate privacy guarantees with their nominal privacy loss, and offers a constructive framework to compare and modify privacy protection procedures in a principled way.

Bio: Ruobin Gong is Assistant Professor of Statistics at Rutgers University. Her research interests encompass theoretical foundations of uncertainty reasoning and statistical data privacy problems, motivated by applications that arise from the curation of official statistical databases. Ruobin is an associate editor for Harvard Data Science Review, Statistics and Public Policy, and JASA/TAS Reviews. She edits a contributed column, Sound the Gong, for the IMS Bulletin.