Differential Privacy in Theory and Practice, with a focus on the 2020 US Census

Dr. Abraham Flaxman
Associate Professor
Department of Global Health and Health Metrics Sciences
University of Washington, Seattle

Abstract: In the United States, the Decennial Census is an important part of democratic governance. Every ten years, the US Census Bureau is constitutionally required to count the "whole number of persons in each State," and in 2020 this effort is likely to cost over 15 billion dollars. The confidentiality of information in the decennial census is also required by law, and the 2020 US Census will use a Differential Privacy to protect respondents' data.

Differential Privacy is a mathematical definition of privacy that has been developed over the last decade and a half in theoretical computer science and cryptography communities. Although the new approach allows a more precise accounting of the variation introduced by the process, it also risks reducing the utility of census data—it may produce counts that are substantially less accurate than the previous disclosure avoidance system.

Graduate student Samantha Petti and I recently used code, preprints, and data files from the Census Bureau to figure out just what they have done in their test run of the system, and to quantify the error introduced by differential privacy when the Census Bureau applied it to 1940 census data. In this talk, I'll tell you what we found, and highlight how constrained convex optimization ("non-negative least squares" as some have been calling it) plays a central role.

Bio: Abraham Flaxman is an Associate Professor of Global Health and Health Metrics Sciences at the University of Washington. But his PhD is in Algorithms, Combinatorics, and Optimization, so has been very excited to work on understanding how differential privacy might be used in the 2020 census, since it brings together his roots in optimization and his daily focus on data for global health metrics.