Achieving Differential Privacy (DP) Using a Two-Tailed Geometric "Bottom-Up" Mechanism

Steve Clark and Aref Dajani, Mathematical Statisticians
Center for Enterprise Dissemination – Disclosure Avoidance
U.S. Census Bureau

Presented at Challenges and New Approaches for
Protecting Privacy in Federal Statistical Programs: A Workshop
National Academy of Sciences, June 7, 2019

Acknowledgments: Nelson Chung, Philip Leclerc, and William Sexton (Census) and Chris Clifton (Purdue)

This presentation was approved under Disclosure Review Board delegated authority #: CDDRB-FY19-CED002-B0014

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

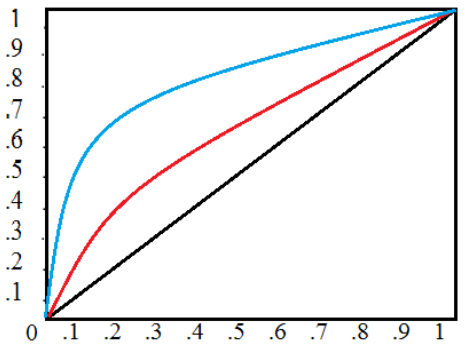
This presentation is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S Census Bureau.

Method (Slide 1 of 2)

- The Two-Tailed Geometric Mechanism effectively injects noise for count data.
  - ε is the parameter that in DP serves as the privacy budget: the tradeoff between privacy loss and accuracy.
  - The lower the value of ε, the more noise that will be injected.
  - $f(x) = \begin{cases} (1-p)^{x-1}p, & 0 < p < 1, x = 1,2, \ldots \\ 1, & p = 1, x = 1 \end{cases}$
  - $p = 1 - e^{-\varepsilon}$
  - For all $\varepsilon > 0$, $0 < p < 1$.
  - By generating two random numbers and taking their difference, the distribution becomes two-tailed.
- The Bottom-Up Approach applies noise to each element in the histogram.
  - The "histogram" in DP literature is the highest-order interaction of all factors such as age, race, and sex.
  - We generated uniform random numbers with a cryptographically secure random number generator.
  - We used an inverse cumulative distribution function for the two-tailed geometric mechanism to generate noise for every highest-order interaction.
  - We aggregate the noise-injected values to form the margins and the entire table.

Method (Slide 2 of 2)

- Sensitivity affects the value of ε.
  - Sensitivity is defined as the maximum impact from the addition or deletion of a record.
  - The sensitivity of count queries for unweighted datasets is one.
- Senior disclosure avoidance leaders select the value of ε to construct the noise-injected information product
  - We replicate the "bottom-up" approach for several values of ε.
  - We calculate L1 and L2 Norms: the average absolute and squared distance, respectively, between the actual counts and calibrated noise-injected counts for each value of ε.
  - We create a Receiver Operating Characteristic (ROC) curve to guide senior disclosure leaders to select the appropriate value of ε.
  - Privacy budgets are calculated for all DP products released by Census.
  - To the right is a sample ROC curve (no Census data were used).



Data, findings, challenges, and future directions

- A special tabulation was conducted on the 1960 Long Form on gender by group quarters for all counties in the United States.
- Population Invariants: Some were not present and some were inconsistent.
  - We calibrated our noise-injected counts to published state totals that were consistent with our state counts.
  - We did not calibrate for unpublished state counts.
  - We did not report geographies that were inconsistent with published state counts.
- The long form was a sample of the 1960 Census, yielding weighted estimates.
  - Weighted estimates increase the sensitivity.
  - There are many forms of sensitivity. We used global sensitivity to report conservative results.
- The L2 norm reached asymptote of one faster than the L1 norm due to the high population value.
- We will consider “secrecy of the sample” in the future to calculate a more accurate level of sensitivity and rein in the privacy budget.
- This method is easy to code and implement for any tabular product within or outside Census.
- For more information, including references, contact Steve Clark at the Census Bureau.
  - Stephen.Clark@census.gov
  - Work #: 301-763-3793