CS2080: Applied Privacy for Data Science
Machine Learning under DP

School of Engineering & Applied Sciences
Harvard University

March 10, 2025

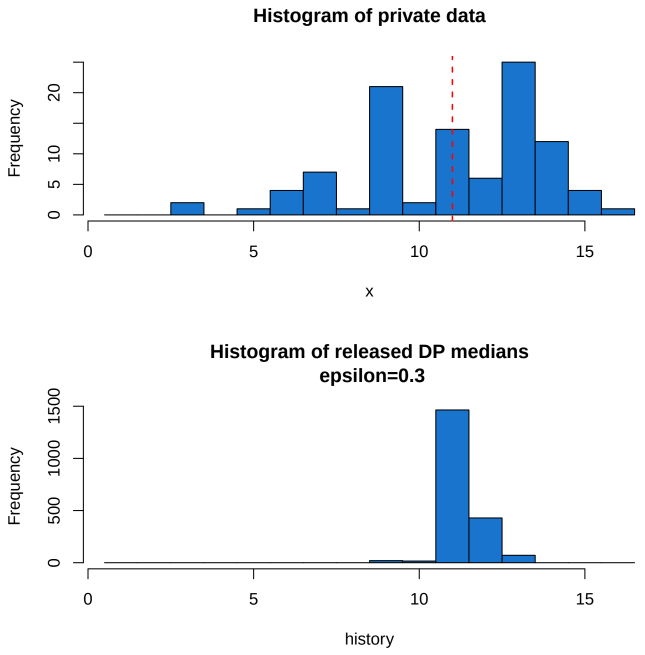
Exponential Mechanism for the Median

- Say $\mathcal{X} = \{1, 2, \cdots, M\}$.
- $M(x):$ output $y \in \mathcal{X}$ with prob $\propto \exp(\epsilon \cdot u(x, y)/2)$
  Where $u(x, y) = \min\{\#\{i : x_i \leq y\}, \#\{i : x_i \geq y\}\}$.
- Note that true median y^* has $u(x, y^*) \geq n/2$.
- Can show that for all x, with high probability,

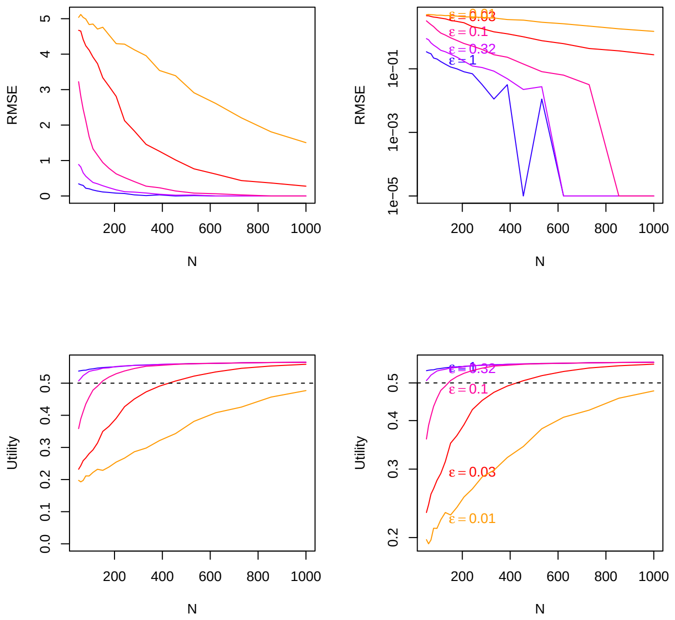
$$u(x, M(x)) \geq n/2 - O(log(M)/\epsilon)$$

Education Values

Codebook for Census PUMS 5 Percent CS208 Datasets		
`educ`	1:	No schooling completed,
	2:	Nursery school to 4th grade,
	3:	5th grade or 6th grade,
	4:	7th grade or 8th grade,
	5:	9th grade,
	6:	10th grade,
	7:	11th grade,
	8:	12th grade, no diploma,
	9:	High school graduate,
	10:	Some college, but less than 1 year,
	11:	One or more years of college, no degree,
	12:	Associate degree,
	13:	Bachelor's degree,
	14:	Master's degree,
	15:	Professional degree,
	16:	Doctorate degree.



Histogram of private data
Frequency
0
5
10
20
0
5
10
15
x
Histogram of released DP medians
epsilon=0.3
Frequency
0
500
1000
1500
0
5
10
15
history

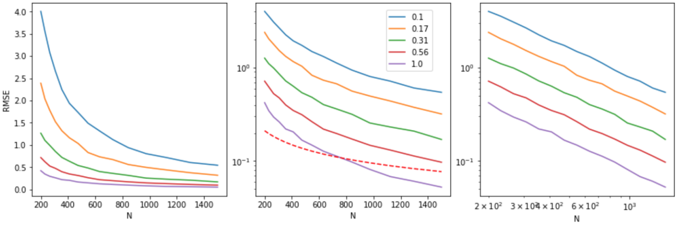

RMSE
N
0
1
2
3
4
5
200
400
600
800
1000
RMSE
1e−05
1e−03
1e−01
ε = 0.01
ε = 0.03
ε = 0.1
ε = 0.32
ε = 1
Utility
0.0
0.1
0.2
0.3
0.4
0.5
Utility
0.2
0.3
0.4
0.5
ε = 1
ε = 0.32
ε = 0.1
ε = 0.03
ε = 0.01

Discussion

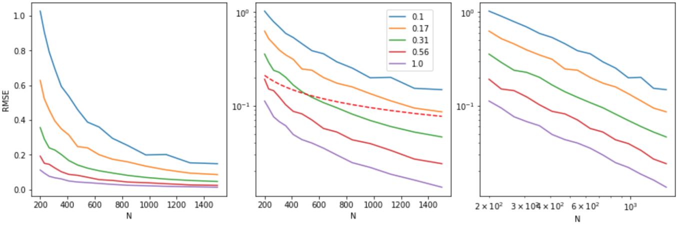
1. Why is the coverage of the population mean failing? Why particularly at low ϵ?

2. What is the sensitivity of the (sample estimate) of standard error of the mean?

$$\mathrm{SE} = \frac{1}{\sqrt{N}} \frac{\sqrt{\sum (x_i - \bar{x})^2}}{N}$$



Gaussian Mechanism



Laplace Mechanism

Correcting Coverage in Confidence Intervals

$$\tilde{M} = \bar{X} + Z; \quad Z \sim \mathcal{N}(0, \Delta^2/2\rho)$$

$$\bar{X} = \mu + Y; \quad Y \sim \mathcal{N}(0, \sigma^2/N)$$

$$\tilde{M} = \bar{X} + Z = \mu + Y + Z; \quad (Y + Z) \sim \mathcal{N}(0, \sigma^2/N + \Delta^2/2\rho)$$

$$CI_{(1-\alpha)} = \tilde{M} \pm z_{(\alpha/2)} S; \qquad S = \sqrt{\mathrm{Var}(Y + X)} = \sqrt{\sigma^2/N + \Delta^2/2\rho}$$

Following slides from:

Practical Method to Reduce Privacy Loss when Disclosing Statistics Based on Small Samples

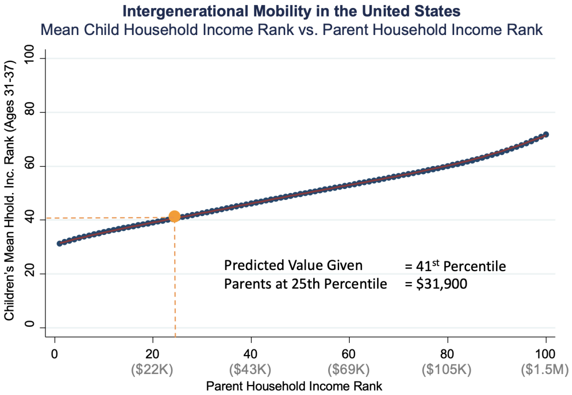
Raj Chetty, Harvard University and NBER

John N. Friedman, Brown University and NBER

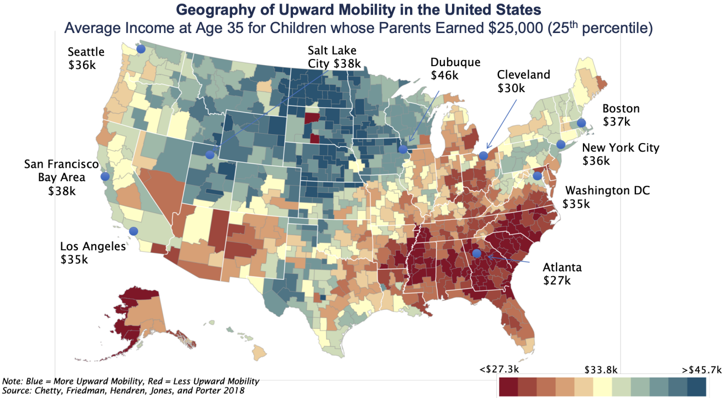
March 2019

Publishing Statistics Based on Small Cells

- Social scientists increasingly use confidential data to publish statistics based on cells with a small number of observations
- Causal effects of schools or hospitals [e.g., Angrist et al. 2013, Hull 2018]
- Local area statistics on health outcomes or income mobility [e.g., Cooper et al. 2015, Chetty et al. 2018]



Source: Chetty, Friedman, Hendren, Jones, Porter (2018)


Geography of Upward Mobility in the United States
Average Income at Age 35 for Children whose Parents Earned $25,000 (25th percentile)
Seattle
$36k
Salt Lake
City $38k
Dubuque
$46k
Cleveland
$30k
Boston
$37k
New York City
$36k
San Francisco
Bay Area
$38k
Washington DC
$35k
Los Angeles
$35k
Atlanta
$27k
<$27.3k
$33.8k
>$45.7k
Note: Blue = More Upward Mobility, Red = Less Upward Mobility
Source: Chetty, Friedman, Hendren, Jones, and Porter 2018

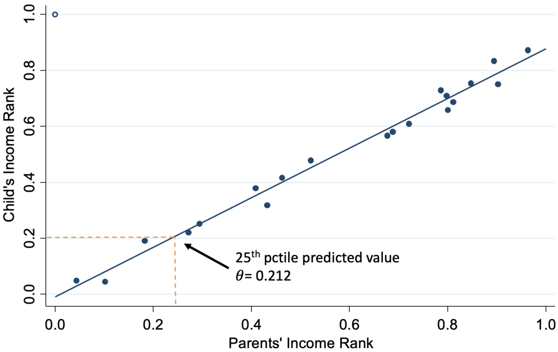
Controlling Privacy Loss

- Problem with releasing such estimates at smaller geographies (e.g., Census tract): risk of disclosing an individual's data
- Literature on differential privacy has developed practical methods to protect privacy for simple statistics such as means and counts [Dwork 2006, Dwork et al. 2006]
- But methods for disclosing more complex estimates, e.g. regression or quasiexperimental estimates, are not feasible for many social science applications [Dwork and Lei 2009, Smith 2011, Kifer et al. 2012]

This Paper: A Practical Method to Reduce Privacy Loss

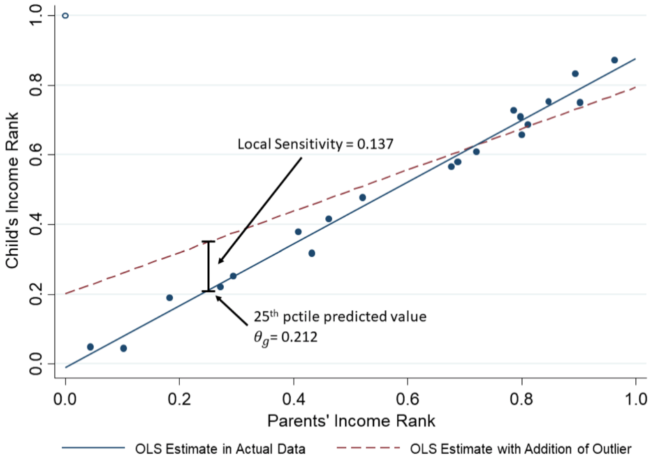
- We develop and implement a simple method of controlling privacy loss when disclosing arbitrarily complex statistics in small samples
  - The “Maximum Observed Sensitivity” (MOS) algorithm
- Method outperforms widely used methods such as cell suppression both in terms of privacy loss and statistical accuracy
  - Does not offer a formal guarantee of privacy, but potential risks occur only at more aggregated levels (e.g., the state level)

Example Regression from One Small Cell



Source: Authors' simulations.

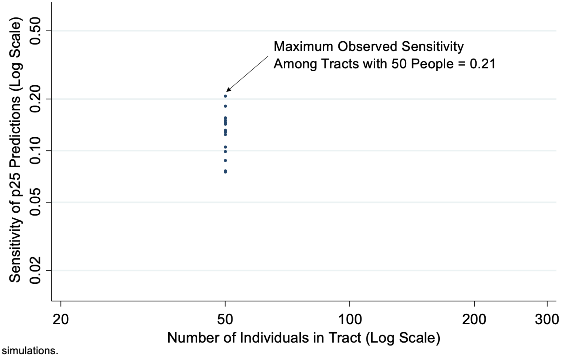
Figure 1: Calculation of local sensitivity



Maximum Observed Sensitivity

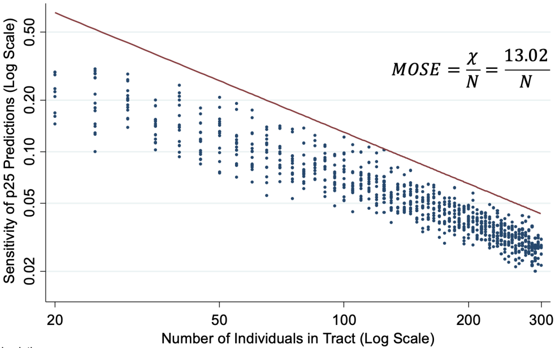
- Our method: use the maximum observed local sensitivity across all cells in the data
  - In geography of opportunity application, calculate local sensitivity in every tract
  - Then use the maximum observed sensitivity (MOS) across all tracts within a given state as the sensitivity parameter for every tract in that state
- Analogous to Empirical Bayes approach of using actual data to construct prior on possible realizations rather than considering all possible priors

Maximum Observed Sensitivity Envelope



Source: Authors' simulations.

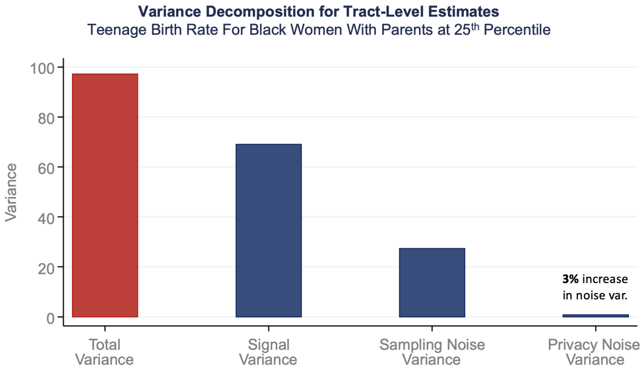
Computing Maximum Observed Sensitivity



Source: Authors' simulations.

Producing Noise-Infused Estimates for Public Release

- Main lesson: tools from differential privacy literature can be adapted to control privacy loss while improving statistical inference
  - Opportunity Atlas has been used by half a million people, by housing authorities to help families move to better neighborhoods, and in downstream research [Creating Moves to Opportunity Project; Morris et al. 2018]
  - The MOS algorithm can be practically applied to any empirical estimate
- Example: difference-in-differences or regression discontinuity
  - Even when there is only one quasi-experiment, pretend that a similar change occurred in other cells of the data and compute MOS across all cells



Source: Chetty, Friedman, Hendren, Jones, Porter (2018)

Conclusion

- Use max observed sensitivity χ, tract counts, and exogenously specified privacy parameter ϵ to add noise and construct public estimates:

$$\tilde{\theta}_g = \theta_g + L\left(0, \frac{\chi}{\epsilon N_g}\right) \quad \tilde{N}_g = N_g + L\left(0, \frac{1}{\epsilon}\right)$$

  - This method not "provably private," but it reduces privacy risk to release of the single max observed sensitivity parameter (!)
  - Privacy loss from release of regression statistics themselves is controlled below risk tolerance threshold ϵ.
- Critically, χ can be computed at a sufficiently aggregated level that disclosure risks are considered minimal ex-ante