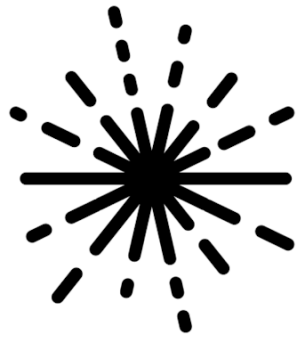


CS2080: Applied Privacy for Data Science
The Local Model: Foundations

School of Engineering & Applied Sciences
Harvard University

April 14, 2025

History of RR

Figure 1: An Example Warner Randomized Response Item

Please spin the following spinner in private and observe if it lands in section "**A**" or "**B**." Do not tell me where it lands.

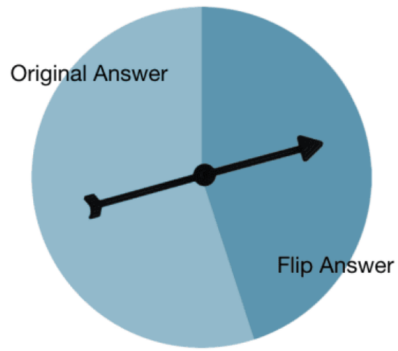
If the arrow landed in section "**A**," respond True or False *only* with respect to statement **A**. If the arrow landed in section "**B**," respond True or False *only* with respect to statement **B**.

A: I would be willing to pay a bribe to a police officer in order to avoid a traffic ticket. (True = I would be willing…, False = I would NOT be willing…)

B: I would NOT be willing to pay a bribe to a police officer in order to avoid a traffic ticket. (True = I would NOT be willing…, False = I would be willing…)

Please respond **TRUE** or **FALSE** now

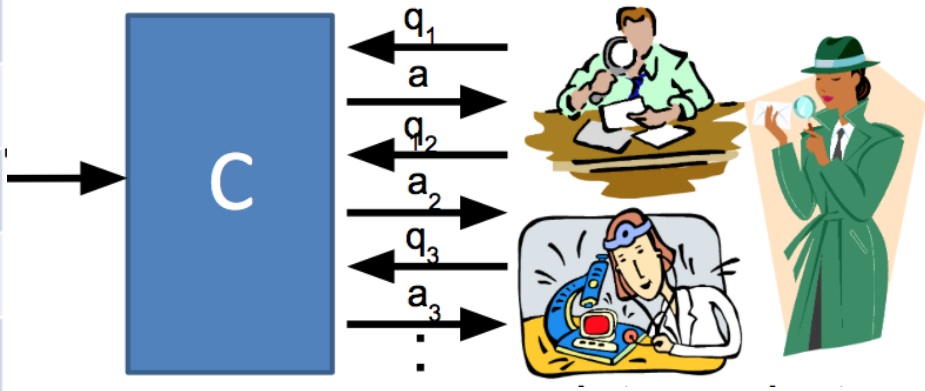
Remember: Your answers are completely confidential. Since only you know if the spinner landed in section A or B, no one else can know to which statement your answer corresponds.



See on NB: Gingerich 2015 "Randomized Response: Foundations and New Developments"

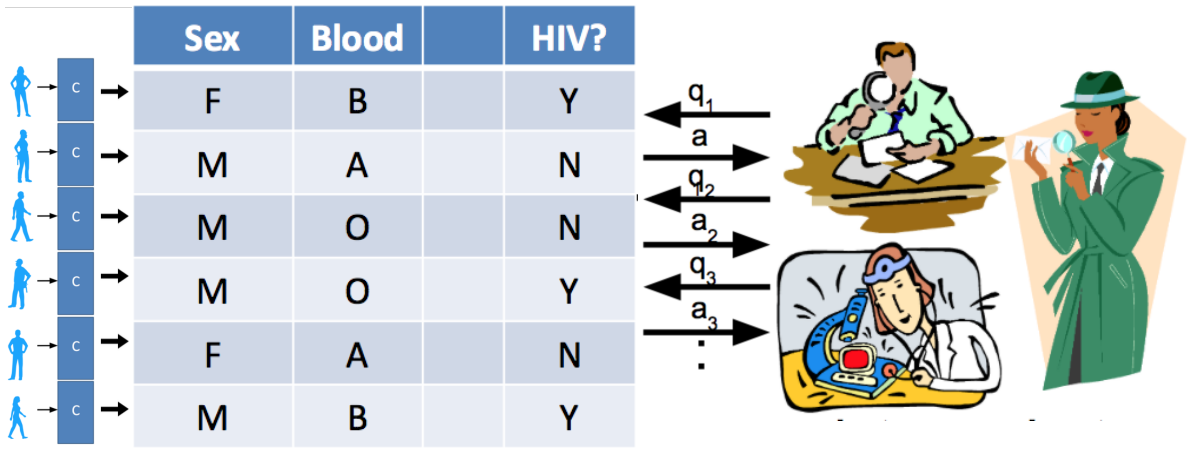
Central Model

Sex	Blood		HIV?
F	B		Y
M	A		N
M	O		N
M	O		Y
F	A		N
M	B		Y



Data Repository Curator Analysts

Local Model



Subjects Repository Analysts

Example: Randomized response

- Each person has data $x_i \in \mathcal{X}$
  - Analyst wants to know average of $f: \mathcal{X} \to \{-1,1\}$ over $\boldsymbol{x}$
  - E.g. "what is the fraction of diabetics"?
- Randomization operator takes $y \in \{-1,1\}$:

$$Q(y) = \begin{cases} +y & w.p. \ \frac{e^{\epsilon}}{e^{\epsilon}+1} \\ -y & w.p. \ \frac{1}{e^{\epsilon}+1} \end{cases}$$

ratio is e^{ϵ}

- Observe:
  - If $c_{\epsilon} = \frac{e^{\epsilon}+1}{e^{\epsilon}-1}$, then $\boldsymbol{E}(c_{\epsilon} \cdot Q(y)) = y$
- How can we estimate a proportion?
  - $A(x_1, \ldots, x_n) = \frac{1}{n}\sum_i c_{\epsilon} \cdot Q(f(x_i))$

Contrast with $\frac{1}{n\epsilon}$ in central model (via Laplace noise)

- **Proposition:** $\boldsymbol{E}\left|A(\boldsymbol{x}) - \frac{1}{n}\sum_i f(x_i)\right| \leq \frac{c_{\epsilon}}{2\sqrt{n}} \approx \frac{1}{\epsilon\sqrt{n}}$.

Slide from Adam Smith

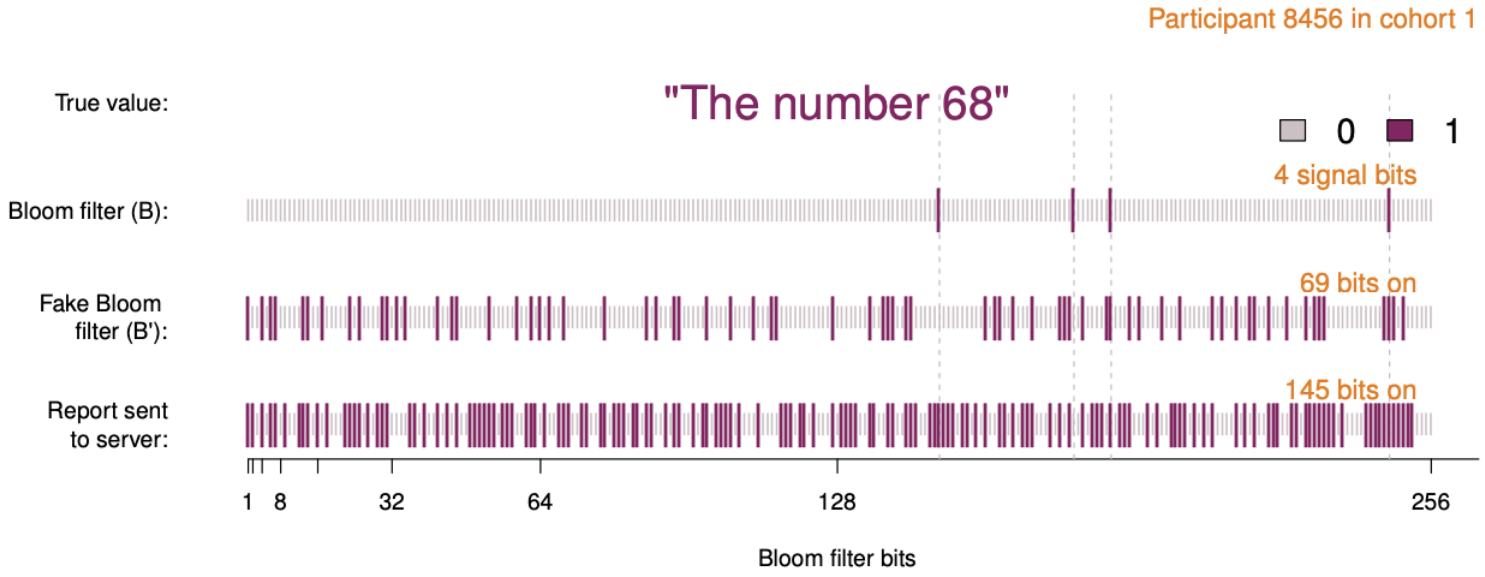


Figure 1: Life of a RAPPOR report: The client value of the string "The number 68" is hashed onto the Bloom filter B using h (here 4) hash functions. For this string, a Permanent randomized response B' is produces and memoized by the client, and this B' is used (and reused in the future) to generate Instantaneous randomized responses S (the bottom row), which are sent to the collecting service.

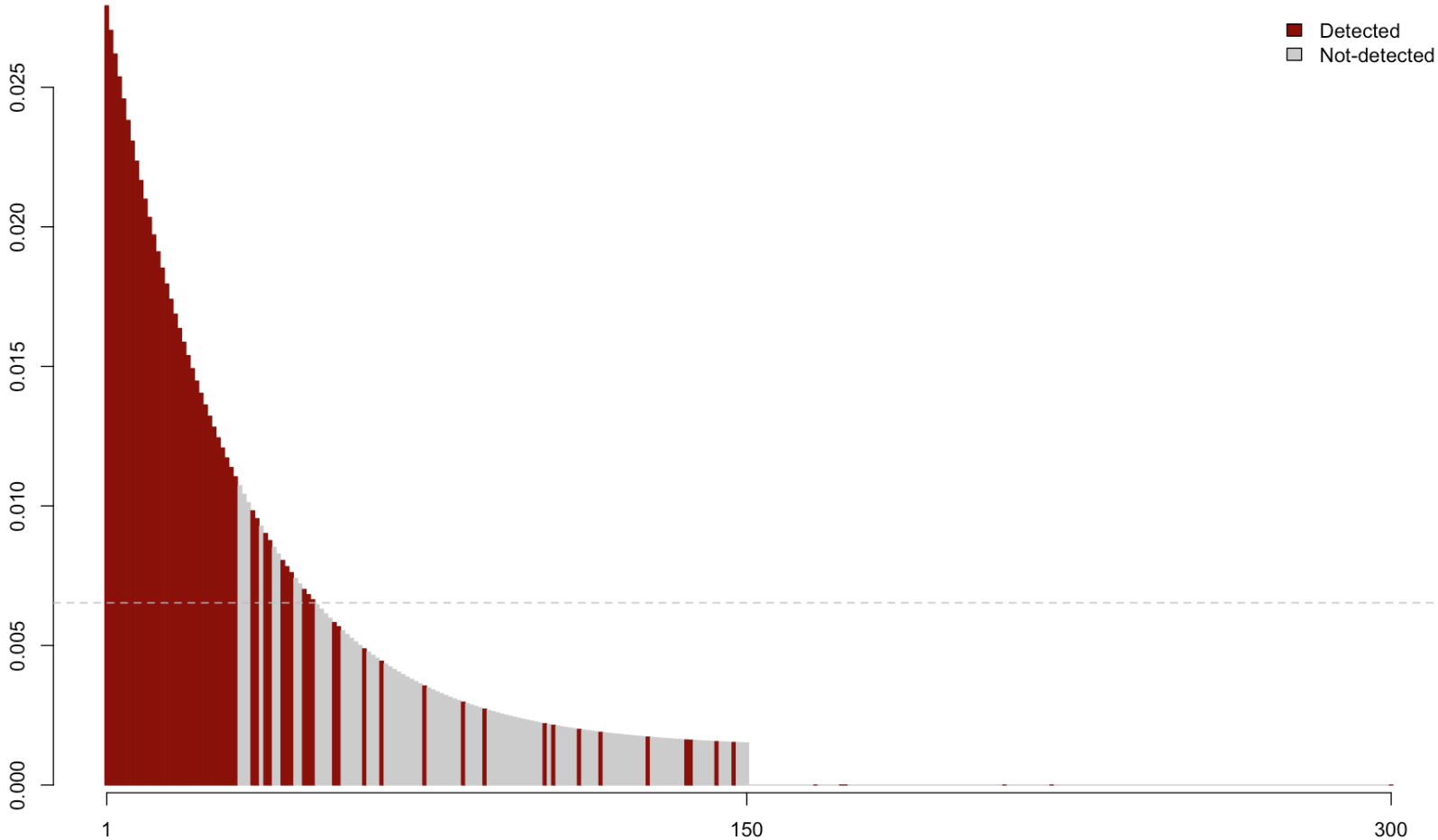
RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response

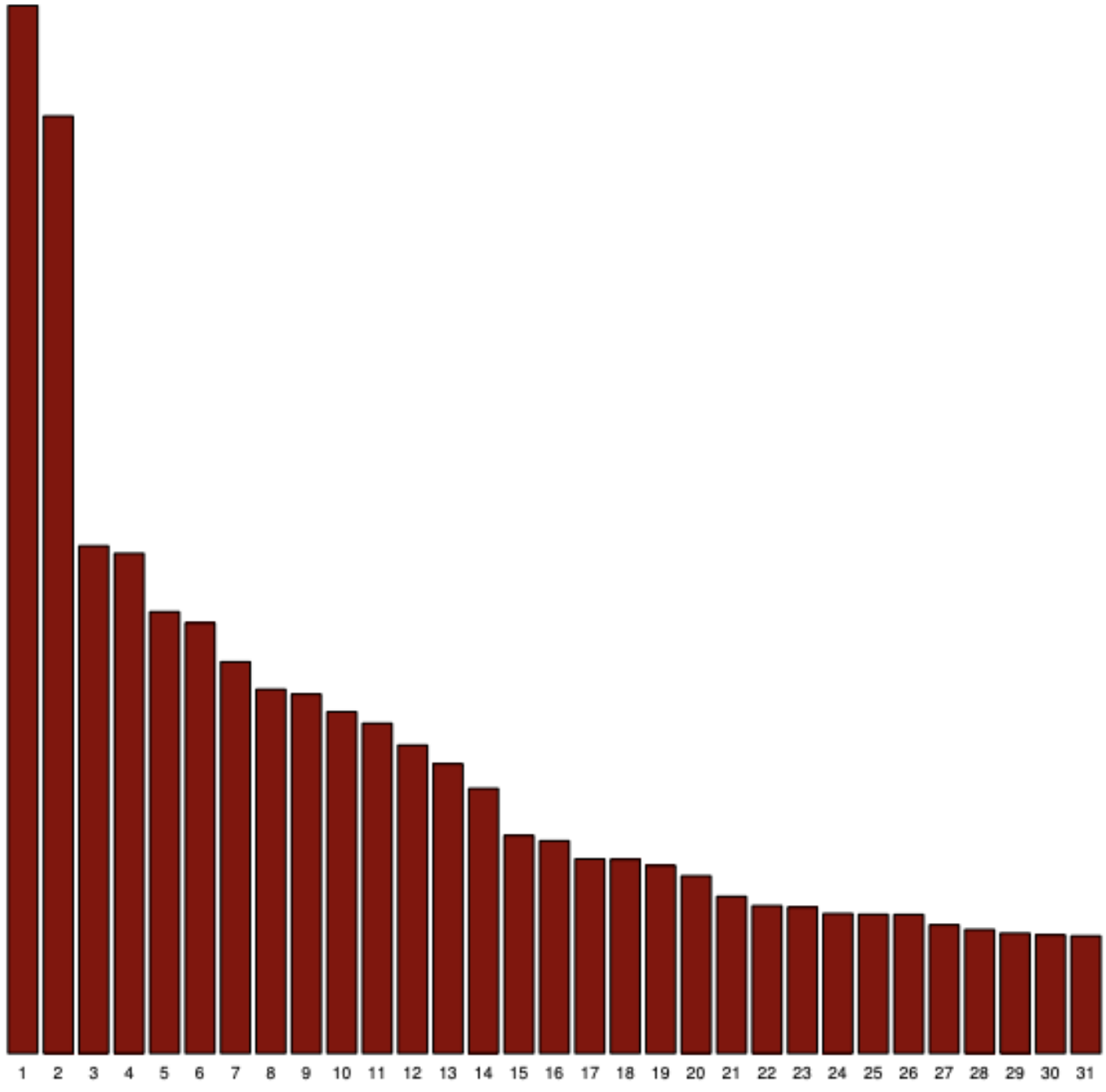
Úlfar Erlingsson
Google, Inc.
ulfar@google.com

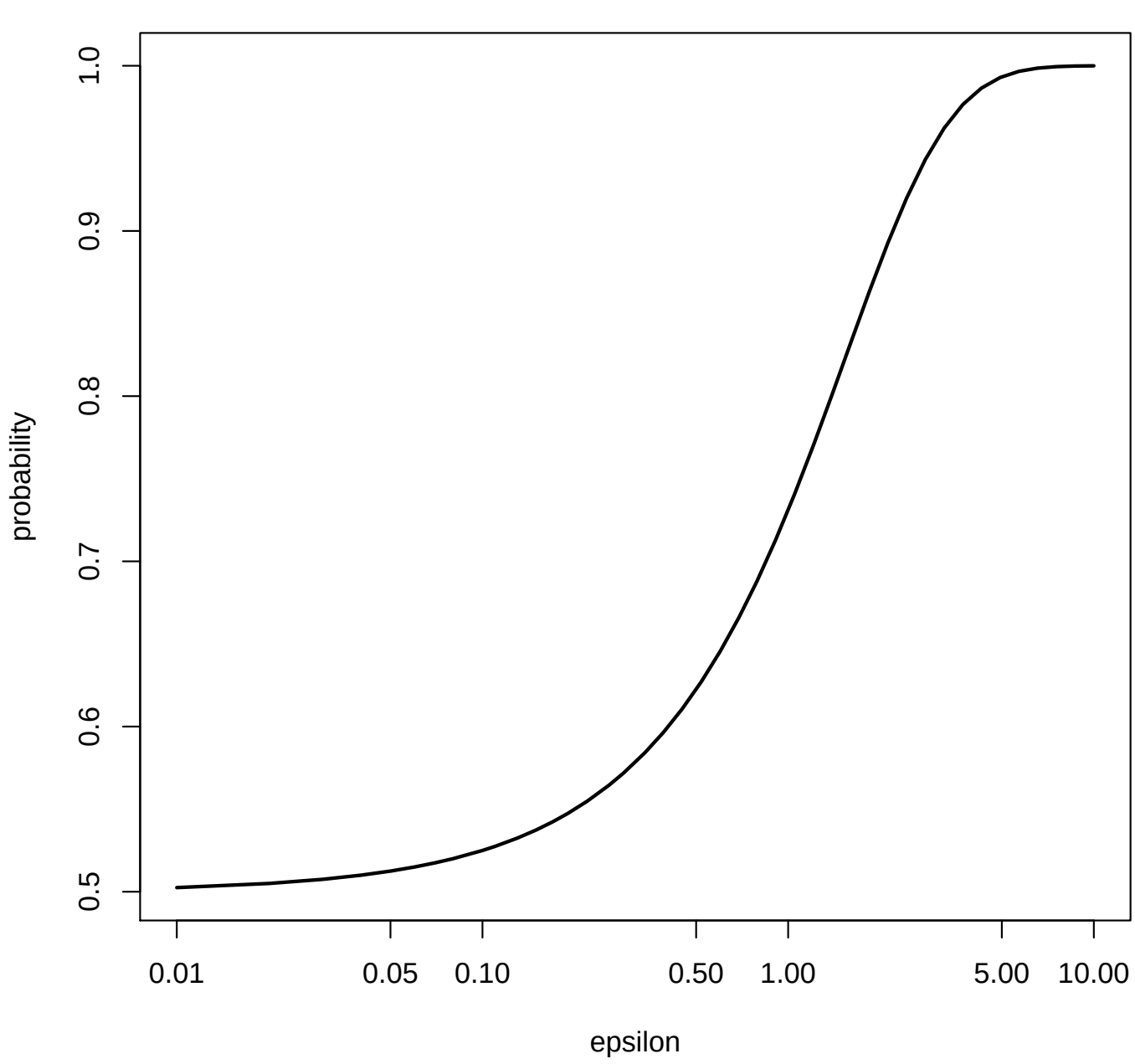
Vasyl Pihur
Google, Inc.
vpihur@google.com

Aleksandra Korolova
University of Southern California
korolova@usc.edu

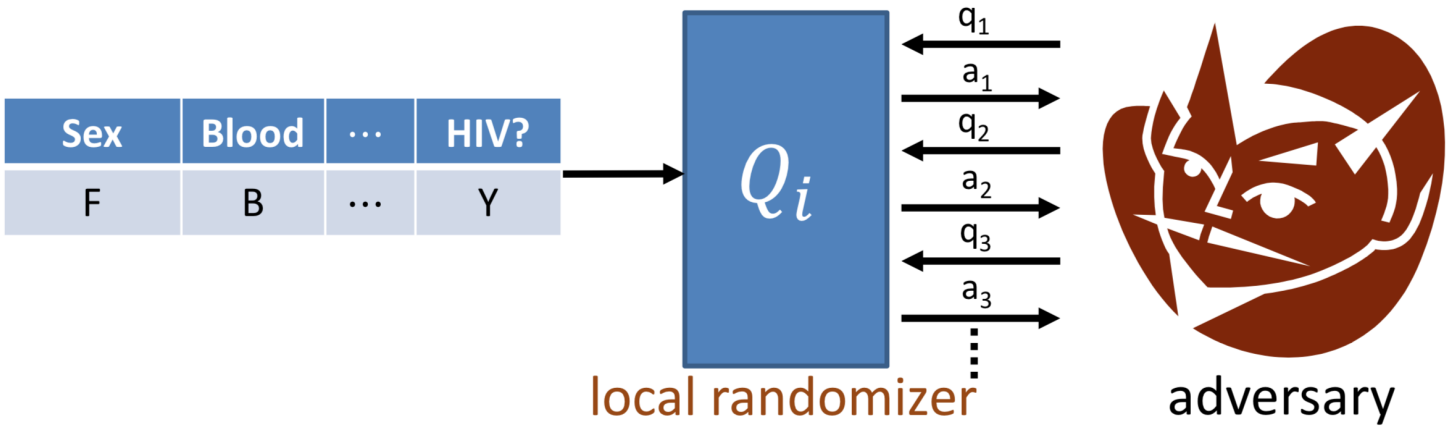
2014


Detected
Not-detected
0.025
0.020
0.015
0.010
0.005
0.000
1
150
300


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

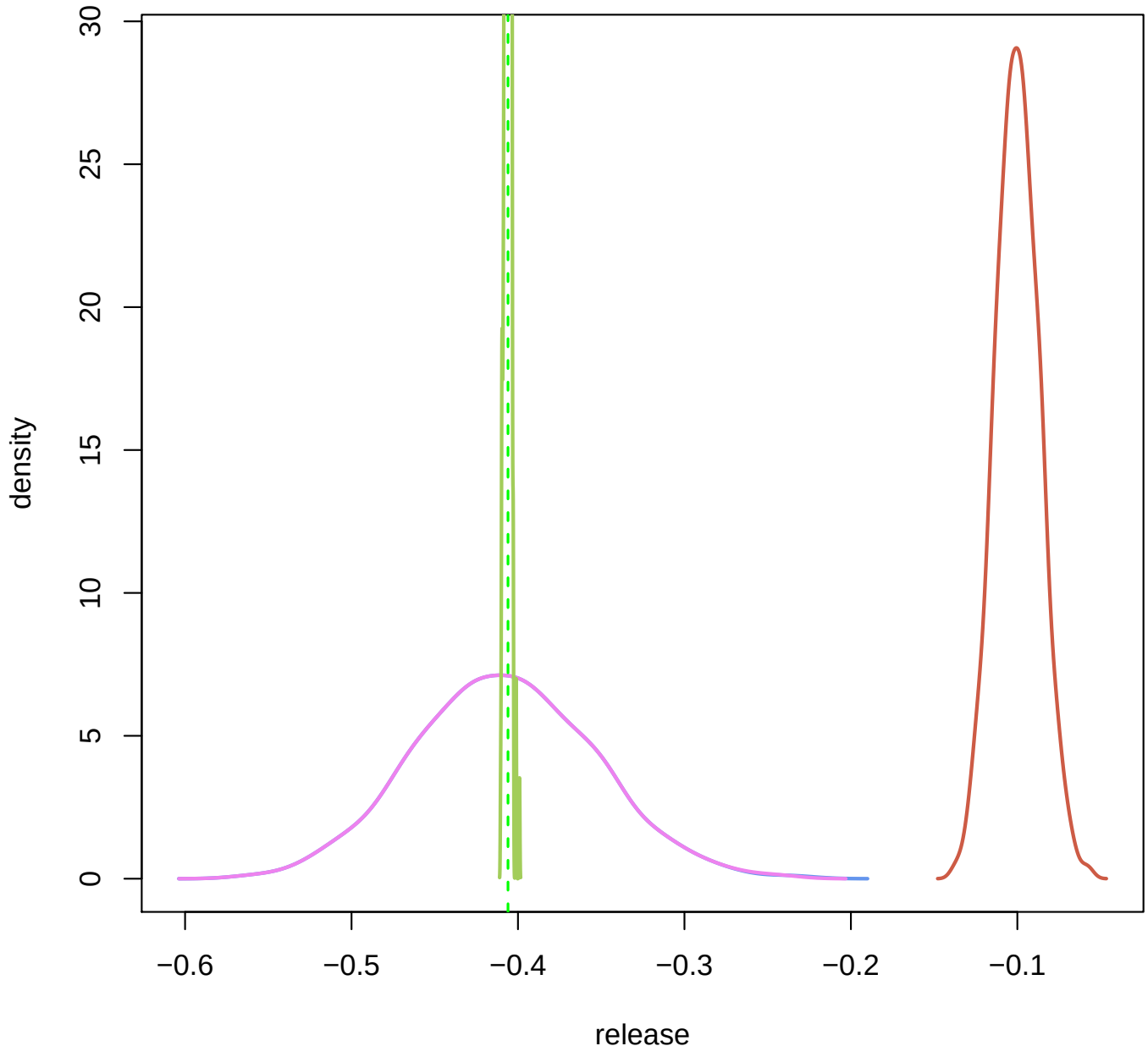

Probability Truth Revealed in Local Model
probability
0.5
0.6
0.7
0.8
0.9
1.0
0.01
0.05
0.10
0.50
1.00
5.00
10.00
epsilon

Local DP



Require: for all i, x_i, x_i' ~~differing on one row,~~ all strategies A

$$\Pr[A \text{ outputs YES after interacting w}/Q_i(x_i)]$$
$$\leq e^{\varepsilon} \cdot \Pr[A \text{ outputs YES after interacting w}/Q_i(x_i')] + \delta$$

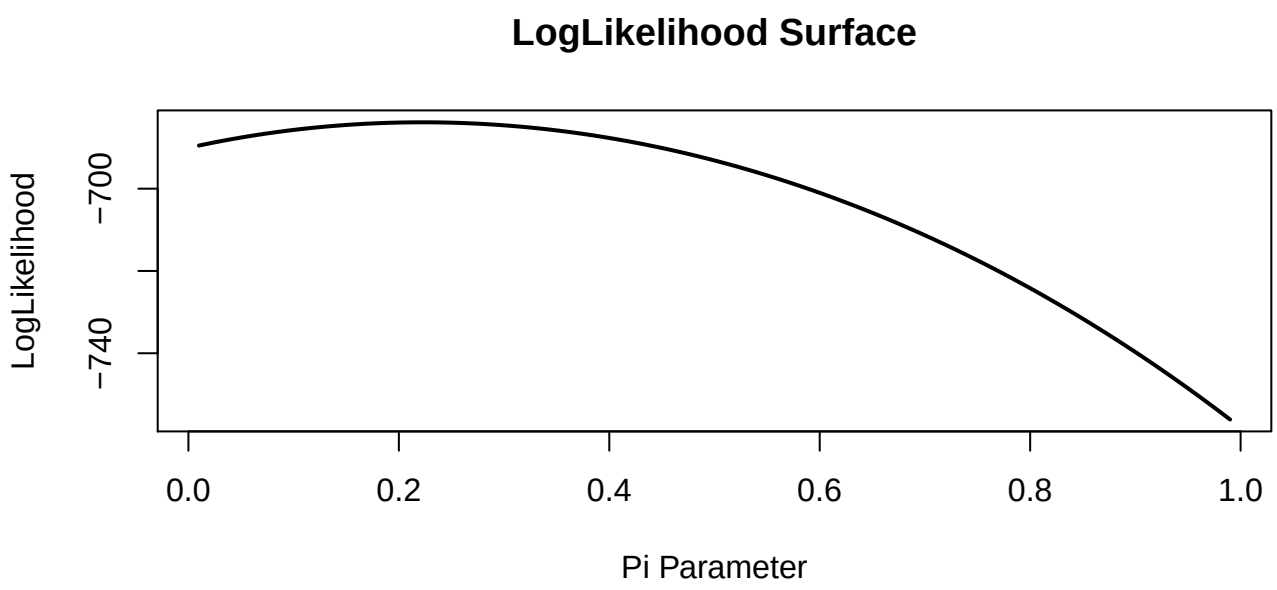

density
0
5
10
15
20
25
30
−0.6
−0.5
−0.4
−0.3
−0.2
−0.1
release

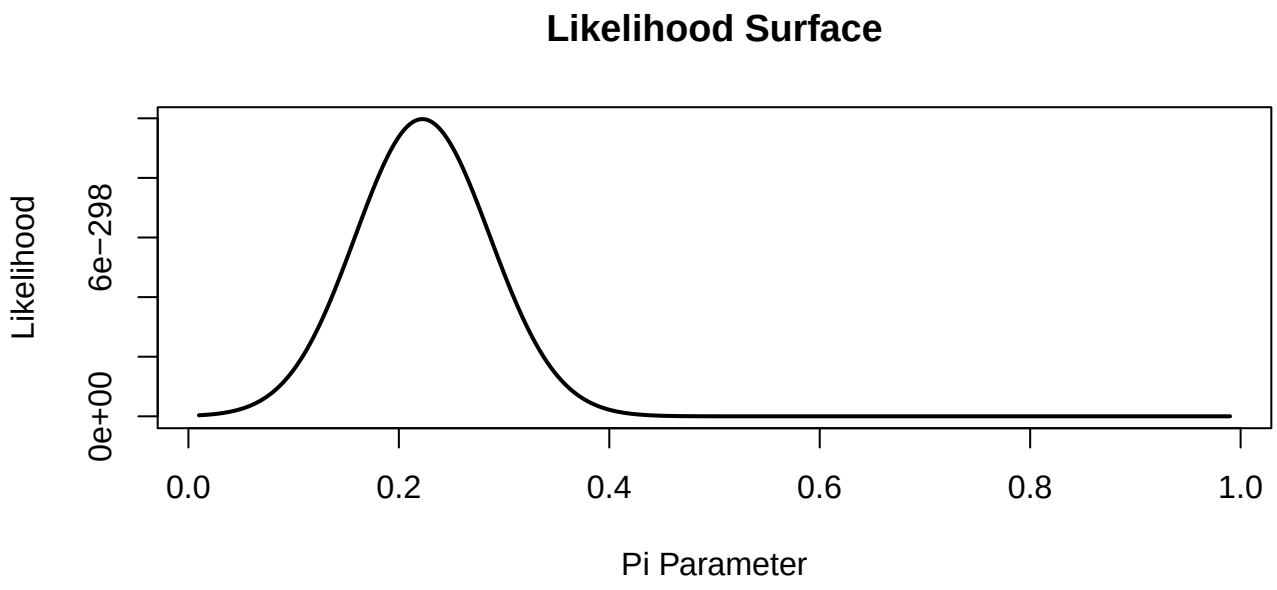
Likelihood Based Correction

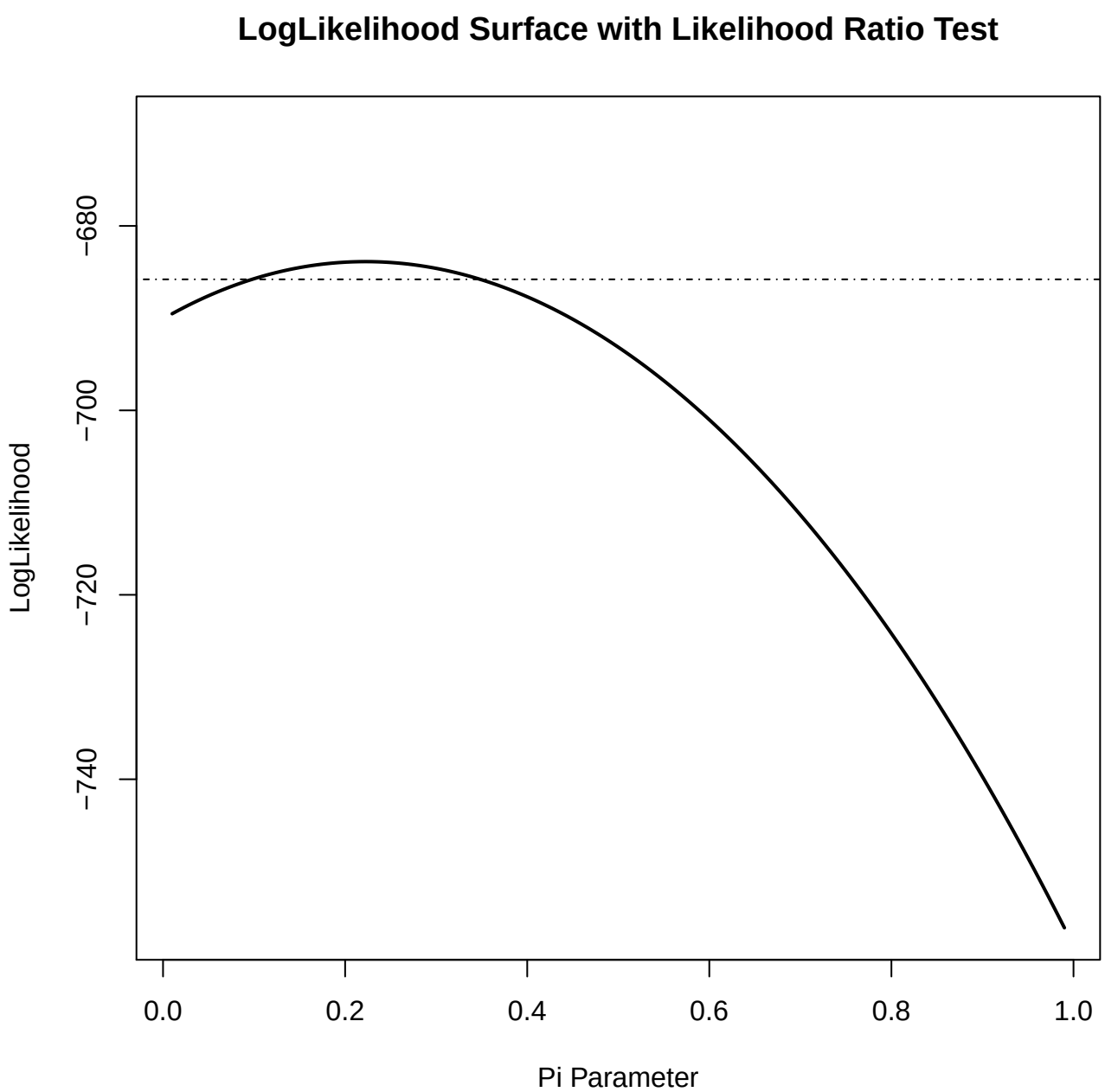
$$L(\theta|Y) \propto Pr(Y|\theta)$$

For $Y \sim \text{Bernoulli}(\pi)$:

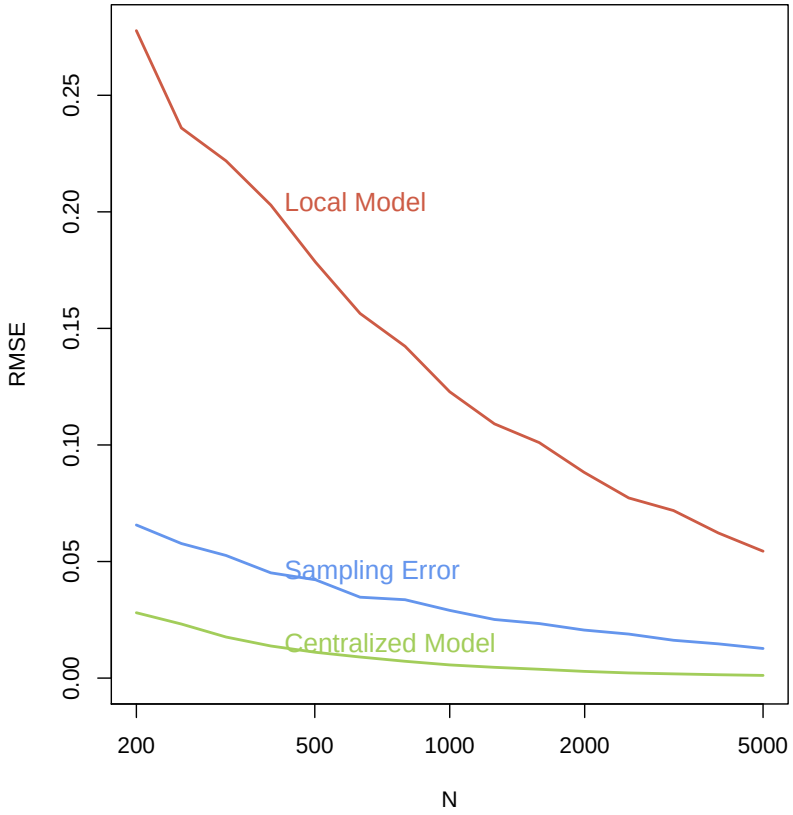
$$L(\pi|Y) \propto \prod_{i=1}^{N} \pi^{y_i}(1-\pi)^{1-y_i}$$

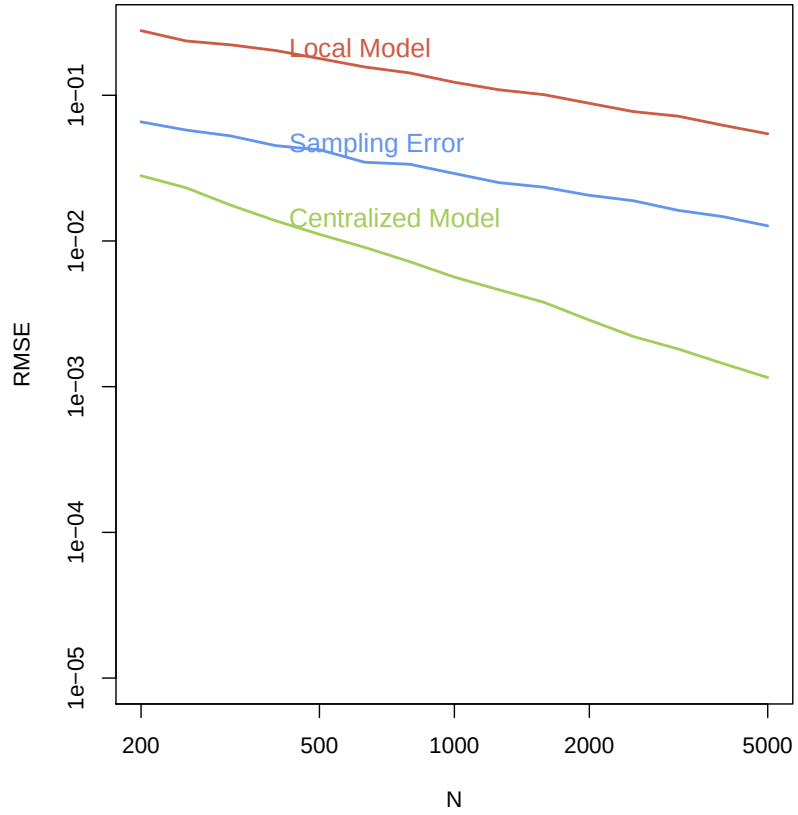

LogLikelihood Surface
LogLikelihood
−700
−740
0.0
0.2
0.4
0.6
0.8
1.0
Pi Parameter


Likelihood Surface
Likelihood
6e−298
0e+00
0.0
0.2
0.4
0.6
0.8
1.0
Pi Parameter


LogLikelihood Surface with Likelihood Ratio Test
LogLikelihood
−680
−700
−720
−740
0.0
0.2
0.4
0.6
0.8
1.0
Pi Parameter

Asymptotics of Error





$\epsilon = 0.5$

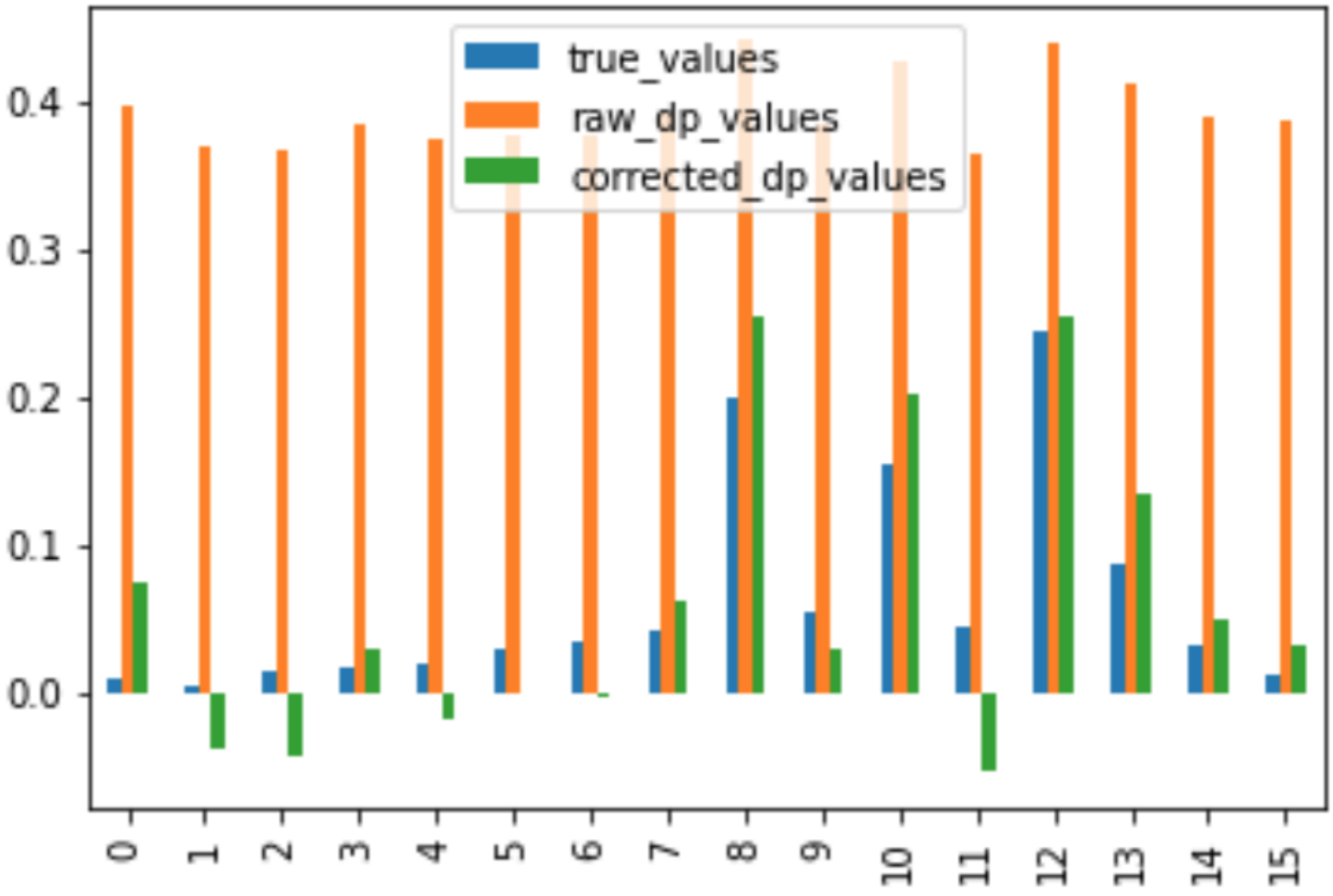
Recap: DP Histograms

- Local randomizer $Q(x_i)$ for $x_i \in \{1, \dots, D\}$
  1. Construct "1-hot" vector $e_{x_i} = (0,0, \dots, 0,1,0, \dots, 0) \in \{0,1\}^D$.
  2. Apply $(\varepsilon/2)$-DP RR to each coordinate to get $y_i \in \{0,1\}^D$:

$$y_i[j] = \begin{cases} e_{x_i}[j] & \text{w.p. } e^{\varepsilon/2}/(1+e^{\varepsilon/2}) \\ 1 - e_{x_i}[j] & \text{o.w.} \end{cases}$$

  3. Send y_i to server.
- Server uses $(y_1, \dots, y_n)$ to estimate histogram $f = \sum_i e_{x_i}$.
- Error per bin $O(\sqrt{n}/\varepsilon)$.

Local Model for Integers/Histograms



$\epsilon = 1, N = 2,000$