Locally Private Gaussian Estimation

Matthew Joseph, Janardhan Kulkarni, Jieming Mao, Zhiwei Steven Wu
majos@cis.upenn.edu, jakul@microsoft.com, maojm@google.com, zsw@umn.edu

Motivation

Estimate mean of $\mathcal{N}(\mu, \sigma)$ while guaranteeing local differential privacy.

Problem Setup

Users $i \in [n]$ draw i.i.d. samples $x_i \sim N(\mu, \sigma)$. Analyst wants to communicate with users and estimate μ. However, this communication protocol must satisfy *local differential privacy* (LDP).

Definition: $Q \colon X \to Y$ is an (ε, δ)-local randomizer if for all $x, x' \in X$ and $S \subset Y$,

$$\mathbb{P}\left[Q(x) \in S\right] \leq e^{\varepsilon}\mathbb{P}\left[Q(x') \in S\right] + \delta.$$

Communication protocol $\mathcal{A}$ is (ε, δ)-**LDP** if all users send data through (ε, δ)-local randomizers. Here, we use **sequentially interactivity**: each user outputs ≤ 1 message.

LDP adds randomness so that public communication reveals little about any one user, but not so much that the analyst can't learn. Expensive communication $\Rightarrow$ should minimize sample and round complexity.

Related Work

Central DP [3]: $O(\sigma\sqrt{\frac{\log(1/\beta)}{n}} + \frac{\text{polylog}(1/\beta)}{\varepsilon n})$.

Local DP [1]: If σ known, 2 rounds. If σ unknown, $\Omega(\log(\mu))$ rounds. $O(\frac{\sigma}{\varepsilon}\sqrt{\frac{\log(1/\beta)\log(n/\beta)\log(1/\delta)}{n}})$ accuracy. $\Omega(\frac{\sigma}{\varepsilon}\sqrt{\frac{\log(1/\beta)}{n}})$ lower bound for sequentially interactive (ε, δ)-locally private protocols.

Our Results

Suppose σ is known and $\frac{n}{\log(n)} = \Omega(\frac{\log(\mu)\log(1/\beta)}{\varepsilon^2})$.

Theorem 1: Protocol is $(\varepsilon, 0)$-LDP, takes 2 rounds of interaction and, w.p. $1-\beta$ estimates μ to $O(\frac{\sigma}{\varepsilon}\sqrt{\frac{\log(1/\beta)}{n}})$ accuracy.

Suppose σ lies in known interval $[\sigma_{\min}, \sigma_{\max}]$ and $\frac{n}{\log(n)} = \Omega(\frac{[\log(\frac{\sigma_{\max}}{\sigma_{\min}}+1)+\log(\mu)]\log(1/\beta)}{\varepsilon^2})$.

Theorem 2: Protocol is $(\varepsilon, 0)$-LDP, takes 2 rounds of interaction and, w.p. $1-\beta$ estimates μ to $O(\frac{\sigma}{\varepsilon}\sqrt{\frac{\log(1/\beta)\log(n)}{n}})$ accuracy.

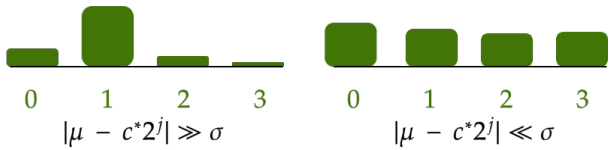
Lower bound for SI locally private protocols.

Theorem 3: Given σ, no SI (ε, δ)-locally private protocol estimates μ to $o(\frac{\sigma}{\varepsilon}\sqrt{\frac{1}{n}})$ accuracy w.p. $\geq 15/16$.

Step 1: Coarse Estimation of μ

Binary search: split users into $L = \Theta(\frac{n\varepsilon^2}{\log(n/\beta)})$ groups. Each one estimates one bit of μ.

Key trick: each user i in group j randomized responds on $\lfloor x_i/2^j \rfloor$ mod 4. Uses Gaussian concentration to binary search in single round.



After first round: $O(\sigma)$-accurate estimate $\hat{\mu}_1$.

(Maybe) Step 1.5: Estimation of σ

Relies on the same randomized user responses on $\lfloor x_i/2^j \rfloor$. When $2^j \gg \sigma$, responses concentrate. When $2^j \ll \sigma$, responses spread out. Examine transition to estimate σ to $O(\sigma)$ accuracy.

Step 2: Fine Estimation of μ

If σ known, have users randomized respond on $\text{sgn}(\frac{\mu - x_i}{\sigma})$. Compare bias to that of standard normal CDF. Requires very accurate knowledge of σ, yields $\hat{\mu}_2$.

If σ estimated, use $\hat{\mu}_1$ and $\hat{\sigma}$ to generate large confidence interval I that w.h.p. contains μ. Each user i clips x_i to I to produce x_i', reports $y_i = x_i' + \text{Lap}\left(\frac{|I|}{\varepsilon}\right)$. Take empirical mean $\hat{\mu}_2 = \frac{2}{n}\sum_i y_i$.

More Stuff in Paper

Can use [2] to extend lower bound to full interactivity with some assumptions.

Nonadaptive and 1-round (but higher-error) versions of both Theorems 1 and 2. Idea: parallelize and do all steps simultaneously. (Some groups will do Step 1, others will do Step 2 for different possible values from Step 1. Only one of these Step 2 groups will be useful. More groups $\Rightarrow$ fewer people in "right" eventual group $\Rightarrow$ higher error.)

References

[1] "Locally Private Mean Estimation: Z-test and Tight Confidence Intervals". Marco Gaboardi, Ryan Rogers, and Or Sheffet. AISTATS 2019.

[2] "The Role of Interactivity in Local Differential Privacy". Matthew Joseph, Jieming Mao, Seth Neel, and Aaron Roth. FOCS 2019.

[3] "Finite Sample Differentially Private Confidence Intervals". Vishesh Karwa and Salil Vadhan. ITCS 2017.