Privately Answering Counting Queries with Generalized Gaussian Mechanisms

Arun Ganesh (UC Berkeley)
Jiazheng Zhao (Stanford)

FORC 2021

Counting Queries

Given vector $d \in \boldsymbol{R}^k$, design mechanism that outputs distribution $M(d)$ of noisy version of d, $\tilde{d}$. Such that:

- Satisfies (ϵ, δ)-differential privacy. Two vectors adjacent if $\|d - d'\|_\infty \leq 1$.
- Minimizes some function of error $\tilde{d} - d$.

Error is $\mathbf{E}\left[\left\|\tilde{d} - d\right\|_1\right]$: Gaussian mechanism has optimal error $OPT \coloneqq \sqrt{k \log \frac{1}{\delta}}\,/\epsilon$.

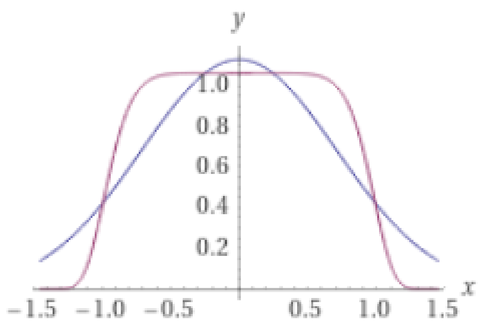
Error is $\mathbf{E}\left[\left\|\tilde{d} - d\right\|_\infty\right]$: Gaussian's error is $OPT \cdot \sqrt{\log k}$.

Generalized Gaussians

Gaussian pdf is $\propto \exp\left(-\left(\frac{\|x\|_2}{\sigma}\right)^2\right)$.

Generalized Gaussians (with shape p) have pdf $\propto \exp\left(-\left(\frac{\|x\|_p}{\sigma}\right)^p\right)$.

Generalized Gaussian with shape 8 vs Gaussian:



High-dimensional Generalized Gaussian's coordinates concentrate better but have higher ℓ_1-norm.

Main result: "Generalized Gaussian mechanism" has expected ℓ_∞-error $OPT \cdot O\left(\sqrt{p} \log^{1/p} k\right)$.

Composition with sparse vector: $OPT \cdot O\left(\sqrt{p} (\log \log k)^{1/p}\right)$.

Privacy Analysis

For (ϵ, δ)-differential privacy, suffices if

$$\Pr_{\tilde{d} \sim M(d)}\left[\log \frac{\Pr[M(d)=\tilde{d}]}{\Pr[M(d')=\tilde{d}]} > \epsilon\right] \leq \delta.$$

$$\log \frac{\Pr[M(d)=\tilde{d}]}{\Pr[M(d')=\tilde{d}]} = \frac{\|x+\mathbf{1}\|_p^p - \|x\|_p^p}{\sigma^p}$$

Just need to tail bound on $\|x + \mathbf{1}\|_p^p - \|x\|_p^p$, then can choose σ appropriately.

$$\|x + \mathbf{1}\|_p^p - \|x\|_p^p \approx p \sum_j x_j^{p-1}$$

Lemma: For x sampled from Generalized Gaussian, x_j^{p-1} are independent Generalized Gammas, which are sub-gamma.

Sub-gamma → $p \sum_j x_j^{p-1} \leq O\left(\sqrt{kp \log 1/\delta}\, \sigma^{p-1}\right)$ w.p. $1 - \delta$.

In turn, $\sigma = OPT \cdot O\left(\sqrt{p}\right)$ suffices.

Error Analysis

$\left(\frac{\|x\|_p}{\sigma}\right)^p$ has Gamma distribution with mean k/p.

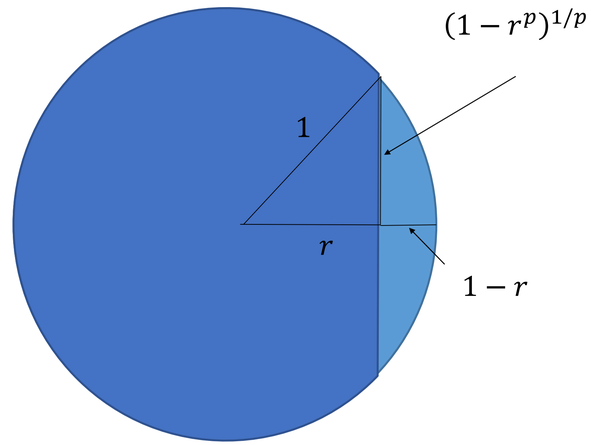
→$\mathbf{E}\left[\|x\|_p\right] = O(k^{1/p}\sigma)$.

Lemma: $\mathbf{E}[\|x\|_\infty] = O\left(\frac{\log^{1/p} k}{k^{1/p}}\right) \mathbf{E}\left[\|x\|_p\right]$

→ $\mathbf{E}[\|x\|_\infty] = O\left(\sigma \log^{1/p} k\right) = OPT \cdot O(\sqrt{p} \log^{1/p} k)$.

Lemma follows from tail bound on each coordinate and union bound.

To tail bound each coordinate, upper bound volume of "sphere cap" on ℓ_p-sphere where that coordinate is $\geq r$.



Future Directions

Optimizing constants: We did not attempt to optimize constants. In practice we can empirically estimate the multiplicative constant needed for privacy, but having a tighter theoretical bound would be nice.

Other notions of privacy: Can we prove the Generalized Gaussian mechanism satisfies e.g. RDP, zCDP? Our analysis crucially uses that we can assume $\delta \geq 2^{-k/p}$, which makes this tricky.

Other applications: Where else do we use the Gaussian mechanism where a Generalized Gaussian mechanism might do better, in theory or practice?