Differentially Private Multi-Sampling from Distributions

Albert Cheu[1] Debanuj Nayak[2]

[1]Google [2]Boston University

Private Multi-Sampling

Probability distribution P from class C

$x_1, x_2, \vdots, x_n$ → DP Sampling Algorithm → $y_1, \vdots, y_m$

Input samples

Randomness of output includes
- Randomness of Input $X \sim P^{\otimes n}$
- Coins of the DP sampling algorithm

Input : n samples from a distribution P from a class C
Output : Privately generate m independent samples from $Q \approx P$
Private Single-Sampling : Special case $m = 1$ [RSSS21]

Measure of Accuracy: Total Variation Distance
- **α-Sampling** : Ensure $||P - Q||_{TV} \leq \alpha$ (when $m = 1$)
- **Weak (m, α)-Sampling** : Ensure $||P - Q||_{TV} \leq \alpha$
- **Strong (m, α)-Sampling** : Ensure $\left|\left|P^{\otimes m} - Q^{\otimes m}\right|\right|_{TV} \leq \alpha$

Baseline Techniques

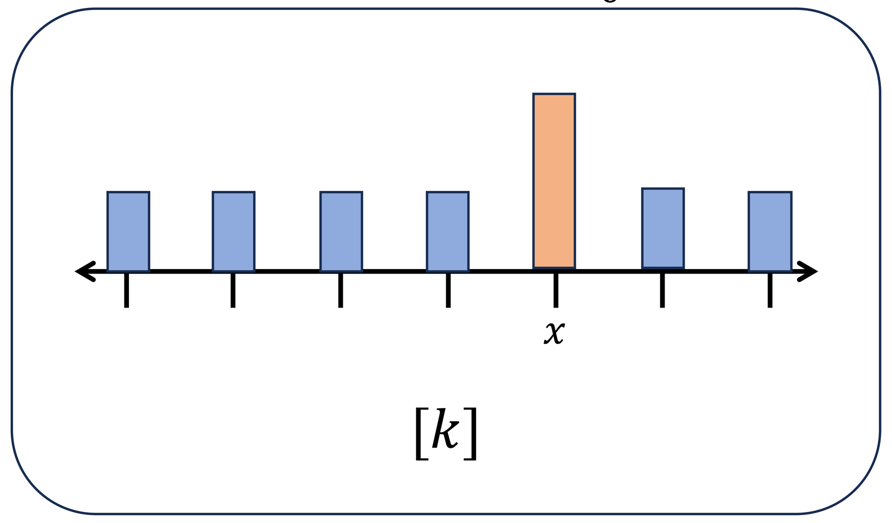
Weak to Strong Multi Sampling	Weak to Strong Multi Sampling
Running m instances of a DP single sampler on m independent datasets is a DP weak multi-sampler	$\|\|P - Q\|\|_{TV} \leq \alpha \Rightarrow \left\|\left\|P^{\otimes m} - Q^{\otimes m}\right\|\right\|_{TV} \leq \alpha m$

k-ary Randomized Response

Input : $x \in [k]$ **Parameter** : $\epsilon_0 > 0$

Output x with probability $\frac{\exp(\epsilon_0)}{\exp(\epsilon_0)+k-1}$

For every $c \in [k] \setminus \{x\}$, output c with probability $\frac{1}{\exp(\epsilon_0)+k-1}$



Sample Complexity of Private Multi-Sampling

Distributions on $[k]$	Pure DP	Approximate DP
Single sampling	$O\left(\frac{k}{\alpha\epsilon}\right)$ [RSSS21]	$\Omega\left(\frac{k}{\alpha\epsilon}\right)$ [RSSS21]
Weak multi sampling [Our Work]	$O\left(\frac{mk}{\alpha\epsilon}\right)$ [baseline technique]	$O\left(m + \frac{k}{\alpha \min(\epsilon, \epsilon^2)} \log\frac{1}{\delta}\right)$
Strong multi sampling [Our Work]	$O\left(\frac{m^2 k}{\alpha\epsilon}\right)$ [baseline technique]	$O\left(m \cdot \frac{k}{\alpha \min(\epsilon, \epsilon^2)} \log\frac{1}{\delta}\right)$ $\Omega\left(\sqrt{m} \cdot \frac{k}{\alpha\epsilon}\right)$

Euclidean-Laplace Distribution

A Probability distribution which is a generalization of Laplace distribution to d-dimensional space.

$$\mathrm{ELap}_b(x) \propto \exp\left(-\frac{\|x\|_2}{b}\right)$$

Euclidean-Laplace Mechanism

Input : $x_1, \ldots, x_n \in \mathbb{R}^d, \|x_i\|_2 \leq B$ for all i **Parameter** : ϵ, B

$x_1, x_2, \vdots, x_n$ → $\sum_{i \in [n]} x_i$ → Add z → y

Sample from $\mathrm{ELap}_b(x)$ with $b = {}^{B}/_{\epsilon}$ → z

For all B, ϵ, the Euclidean-Laplace satisfies ϵ-DP

Pure-DP Gaussian Sampler for $\mathcal{N}(\leq R, I)$

$x_1, x_2, \vdots, x_n$ → Clip length to B → $\hat{x}_1, \hat{x}_2, \vdots, \hat{x}_n$ → Euclidean-Laplace Mechanism with B, ϵ → w

$B = O(R + \sqrt{d \log(1/\alpha)})$

$\mathcal{N}\left(0, \frac{n-1}{n} \cdot I\right) \sim Z$ → Output $Z + {}^{w}/_{n}$

Sample Complexity of Private Multi-Sampling

d-dimensional Gaussians with known covariance Σ	Pure DP	Approximate DP
Single sampling	$O\left(\frac{d^{3/2}}{\alpha\epsilon} \log\left(\frac{d}{\alpha}\right)\right)$	$\tilde{\Theta}\left(\frac{\sqrt{d}}{\epsilon}\right)$ [GHKM23]
Weak multi sampling [Our Work]	$O\left(m \cdot \frac{d^{3/2}}{\alpha\epsilon} \log\left(\frac{d}{\alpha}\right)\right)$ [baseline technique]	$\tilde{\Theta}\left(m \cdot \frac{\sqrt{d}}{\epsilon}\right)$ [baseline technique]
Strong multi sampling [Our Work]	$O\left(m^2 \cdot \frac{d^{3/2}}{\alpha\epsilon} \log\left(\frac{d}{\alpha}\right)\right)$ [baseline technique]	$\tilde{\Theta}\left(m \cdot \frac{\sqrt{d}}{\epsilon}\right)$ [baseline technique]

Subsampled Randomized Response

$x_1, x_2, \vdots, x_n$ → Sample r uniformly from $[n]$ → x_r → k-ary Randomized Response ϵ_0-DP → y

Sample complexity $n = O\left(\frac{k}{\alpha\epsilon}\right)$

Shuffled Randomized Response

$x_1, x_2, \vdots, x_n$ → k-ary Randomized Response (ϵ_0-DP) → $z_1, z_2, \ldots, z_n$ → Shuffle step → $y_1, \vdots, y_m$

Privacy amplification through shuffling [FMT21]

Sample complexity $n = O\left(m + \frac{k}{\alpha \min(\epsilon, \epsilon^2)} \log\frac{1}{\delta}\right)$

Lower Bound on Strong Multi-Sampling

Theorem [K24] : For two probability distributions P , Q over finite sets, we have

$$\left\|P^{\otimes m} - Q^{\otimes m}\right\|_{TV} \geq c \cdot \sqrt{m} \cdot \|P - Q\|_{TV}$$

$$\|P - Q\|_{TV} \geq \alpha \Longrightarrow \left\|P^{\otimes m} - Q^{\otimes m}\right\|_{TV} \geq c \cdot \sqrt{m} \cdot \alpha$$

References

- **[RSSS21]** Sofya Raskhodnikova, Satchit Sivakumar, Adam D. Smith, and Marika Swanberg. Differentially private sampling from distributions, (NeurIPS 2021)
- **[GHKM23]** Badih Ghazi, Xiao Hu, Ravi Kumar, and Pasin Manurangsi. On differentially private sampling from gaussian and product distributions, (NeurIPS 2023)
- **[FMT21]** Vitaly Feldman, Audra McMillan, and Kunal Talwar. Hiding among the clones: A simple and nearly optimal analysis of privacy amplification by shuffling, (FOCS, 2021)
- **[K24]** Aryeh Kontorovich. On the tensorization of the variational distance, 2024