The Correlated Gaussian Sparse Histogram Mechanism

Christian Janos Lebeda[1], Lukas Retschmeier[2]

[1] *INRIA, France* [2] *Basic Algorithm Research Copenhagen (BARC), Copenhagen*

TL;DR

Better **Sparse Histograms with Gaussian noise** under the **add-remove** neighboring relationship!

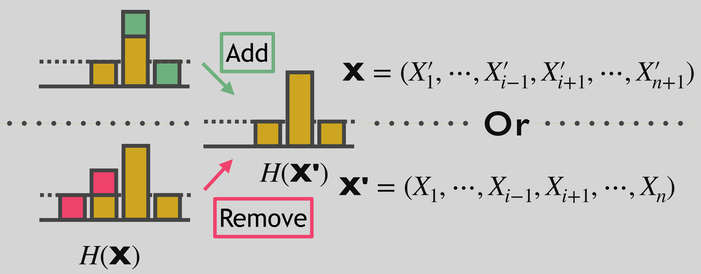
We are using correlated noise to improve accuracy by up to a factor of 2!

Preliminaries

Let $\mathbf{X} = (X_1, \cdots, X_n)$ for $X_i = \{0,1\}^d$, then the **histogram**

$$H(\mathbf{X}) = \sum_{i=1}^{n} X_i$$

Add-Remove neighboring relationship



Sensitivity Space of a function H is the set

$$\Delta_H := \{H(\mathbf{X}) - H(\mathbf{X}') \mid \mathbf{X} \sim \mathbf{X}'\}$$

k-sparse monotonic histogram

An input histogram $H(\mathbf{X})$ is **k-sparse monotonic**, if

✓ $\|H(\mathbf{X})\|_0 \leq k$ for **all datasets** $\mathbf{X}$, and

✓ $\Delta_H = \{0,1\}^d \cup \{0,-1\}^d$

The Main Algorithm

Procedure Correlated Gaussian Sparse Histogram Mechanism

Input: Parameters σ and τ, histogram $H(\mathbf{X}) \in \mathbb{N}^d$, where $\|H(\mathbf{X})\|_0 \leq k$

Let $\tilde{H} = \{0\}^d$

Correlated Sample

Sample $Z_c \sim \mathrm{N}(0, \sigma^2/\sqrt{k})$

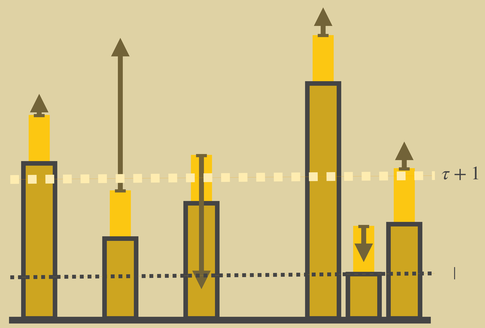
for each $i \in [d]$ where $H(\mathbf{X})_i \neq 0$ **do**

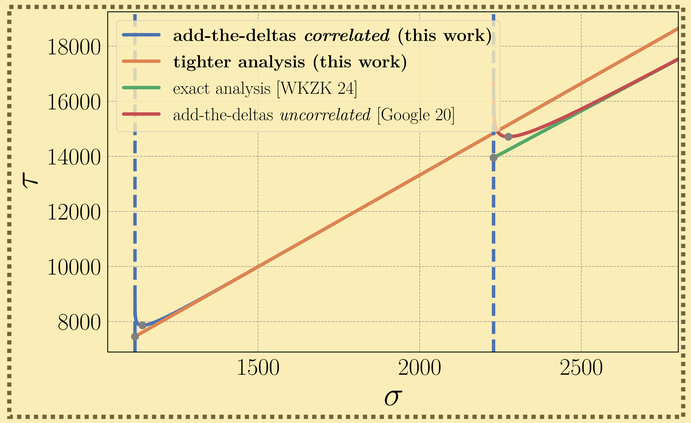
Sample $Z_i \sim \mathrm{N}(0, \sigma^2)$.

if $H(\mathbf{X})_i + Z_i + Z_c > 1 + \tau$ **then**

Set $\tilde{H}_i = H(\mathbf{X})_i + Z_i + Z_c$

end for





Numerical Evaluation. *Minimal τ for $k = 51914$ to get $(0.35, 10^{-5})$-DP guarantees for a noise level σ. Minimums are marked by a circle (•).*

Our Results

Theorem 1 ("Add-the-deltas"). The Correlated Gaussian Sparse Histogram Mechanism satisfies $(\epsilon, \delta_{\text{gauss}} + \delta_{\text{inf}})$-DP for k*−sparse monotonic* histograms where

$$\delta_{\text{gauss}} = \Phi\left(\frac{\sqrt{k+\sqrt{k}}}{4\sigma} - \frac{2\epsilon\sigma}{\sqrt{k+\sqrt{k}}}\right) - e^{\epsilon}\Phi\left(-\frac{\sqrt{k+\sqrt{k}}}{4\sigma} - \frac{2\epsilon\sigma}{\sqrt{k+\sqrt{k}}}\right)$$

(Gain 1 → 2) (Gain 2 → 4)

$$\delta_{\text{inf}} = 1 - \Phi\left(\frac{\tau}{\sigma(1+k^{-1/4})}\right)^{k+1}$$

Theorem 2 (Tighter Bound). Improvements similar to [WKZK 24]: Taking max over sensitivity space and a more intricate case distinction.

Additionally, we have extensions for **Discrete Gaussian Noise, Aggregator Functions,** and an **Additional Sparsity Threshold**

Open Questions

- **Tight** bound like in [WKK 24]?
- Optimal ratio of correlated noise for setting where $\|X_i\|_0 \leq k/2$ like in [JY24]?

Key References

[Google 20] Google Anonymization Team, "Delta for Thresholding", Technical Report

[WKZK 24] Wilkins A. and Kifer, D. and Zhang D. And Karrer B. "Exact privacy analysis of the gaussian sparse histogram mechanism", JPC 2024

[L 24] Lebeda C. "Better gaussian mechanism using correlated nosie", SOSA 25

[BW 18] Balle B. And Wang Y, "Improving the gaussian mechanism for differential privacy", PMLR 2017

[JY 24] Joseph, M and Yu, A. **"**Some Constructions of Private, Efficient, and Optimal K-Norm and Elliptic Gaussian Noise"

Research supported by the VILLUM foundation grant number 16582 (BARC), a Data Science Investigator Grant from Novo Nordisk Fonden (Providentia), and grant ANR-20-CE23-0015 (Project PRIDE).