Calibrating Noise to Sensitivity in Private Data Analysis

Theory of Cryptography Conference, 2006

Cynthia Dwork[1], Frank McSherry[1], Kobbi Nissim[2], and Adam Smith[3]

[1]Microsoft Research, Silicon Valley
[2]Ben-Gurion University
[3]Weizmann Institute of Science

content

Motivation

Motivation

Consider the privacy problem in a statistical database $\in D^n$.
Previous work:

- specific function: noisy sums $f = \sum_i g(x_i)$ and g maps rows to $[0, 1]$

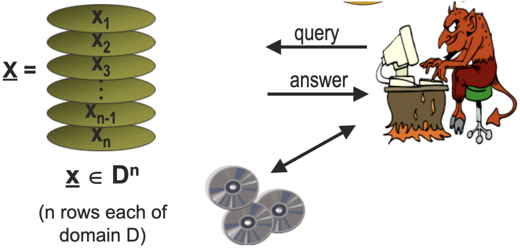
This work:

- general function f
- new definition: $\epsilon-$indistinguishability
- general method: **sensitivity**-based perturbation

New Definition

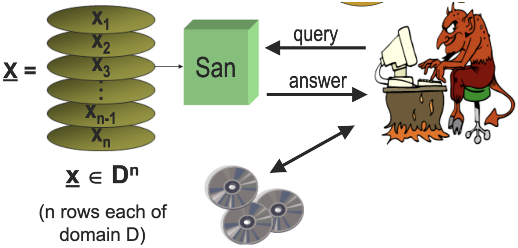
The Setting

- Statistical database $\boldsymbol{x} \in D^n$: n rows, each row $x_i \in D$.
  - D can be $\{0,1\}^d$ or $\mathbb{R}^d$.
- User/Adversary $\mathcal{A}$: a probabilistic interactive Turing machine.



The Setting

- Database access protocol San.
- Transcript $\mathcal{T}_{\text{San},\mathcal{A}}(\mathbf{x})$: $[Q_1, a_1, \ldots, Q_d, a_d]$.
- The Hamming distance $d_H(\cdot,\cdot)$ over D^n: #entries in which two databases differ.



ϵ − indistinguishability

Definition (ϵ − indistinguishability)

A mechanism is ϵ-indistinguishable if for all pairs $\mathbf{x}, \mathbf{x}' \in D^n$ which differ in only one entry, for all adversaries $\mathcal{A}$, and for all transcripts t :

$$\left|\ln\left(\frac{\Pr\left[\mathcal{T}_{\mathcal{A}}(\mathbf{x})=t\right]}{\Pr\left[\mathcal{T}_{\mathcal{A}}\left(\mathbf{x}'\right)=t\right]}\right)\right| \leq \epsilon. \tag{1}$$

Remark:

- ϵ is called *leakage*.
- $\epsilon \to 0$, $\ln(1+\epsilon) \approx \epsilon$. Then $\frac{\Pr[\mathcal{T}_{\mathcal{A}}(\mathbf{x})=t]}{\Pr[\mathcal{T}_{\mathcal{A}}(\mathbf{x}')=t]} \in 1 \pm \epsilon$.

More discussions on $\epsilon -$ indistinguishability

Example (Noisy Sum)

Suppose $\mathbf{x} \in \{0,1\}^n$, and the user wants to learn $f(\mathbf{x}) = \sum_i x_i$, the total number of 1 's in the database. Consider adding noise to $f(\mathbf{x})$ according to a Laplace distribution: $\mathcal{T}(x_1, \ldots, x_n) = \sum_i x_i + Y$, where $Y \sim \text{Lap}(1/\epsilon)$ This mechanism is ϵ-indistinguishable.

Proof of Noisy Sum

Proof.

$\mathrm{Lap}(\lambda)$ has density function $h(y) \propto \exp(-|y|/\lambda)$.
For $y, y' \in \mathbb{R}$, $\frac{h(y)}{h(y')} = \exp(\frac{1}{\lambda}(|y'| - |y|)) \leq e^{\epsilon|y-y'|}$.
Since $\mathbf{x}, \mathbf{x}' \in \{0,1\}^n$ differ in a single entry,

$$|f(\mathbf{x}) - f(\mathbf{x}')| = 1\,.$$

Thus, for $t \in \mathbb{R}$,

$$\frac{\Pr(\mathcal{T}(\mathbf{x}) = t)}{\Pr(\mathcal{T}(\mathbf{x}') = t)} = \frac{h(t - f(\mathbf{x}))}{h(t - f(\mathbf{x}'))} \leq e^{\epsilon|f(\mathbf{x}) - f(\mathbf{x}')|} = e^{\epsilon}\,,$$

which concludes the proof. $\square$

More discussions on $\epsilon -$ indistinguishability

A more common metric for cryptography:

Definition (total variation distance/statistical difference (SD))

The total variation distance between two probability measures P and Q on a sigma-algebra $\mathcal{F}$ of subsets of the sample space Ω is defined via

$$\delta(P, Q) = \sup_{A \in \mathcal{F}} |P(A) - Q(A)| \, .$$

However, $\epsilon -$ indistinguishability is more stringent.

- Example:$p_P(a) \neq 0$ and $p_Q(a) = 0$. Ratio in Eq. (1) is infinite while SD could be small.

More discussions on $\epsilon-$ indistinguishability

Example (Candidate Sanitization)

Consider the candidate sanitization

$$\mathcal{T}(x_1, \ldots, x_n) = (i, x_i) \quad \text{where } i \in_R \{1, \ldots, n\}.$$

If $\mathbf{x}$ and $\mathbf{x}'$ differ in a single position,

- $\mathrm{SD}(\mathcal{T}(\mathbf{x}), \mathcal{T}(\mathbf{x}')) = 1/n$.
- Every transcript reveals individual private information.
- no $\epsilon-$ indistinguishability: Say $\mathbf{x}$ and $\mathbf{x}'$ differ in the ith coordinate.

  $$\Pr(\mathcal{T}(\mathbf{x}') = (i, x_i)) = 0$$

  can not satisfy

  $$\left| \ln \left(\frac{\Pr[\mathcal{T}_{\mathcal{A}}(\mathbf{x}) = t]}{\Pr[\mathcal{T}_{\mathcal{A}}(\mathbf{x}') = t]} \right) \right| \leq \epsilon\,.$$

Sensitivity and Privacy

L_1 Sensitivity

Definition (L_1 Sensitivity)

The L_1 sensitivity of a function $f : D^n \to \mathbb{R}^d$ is the smallest number $S(f)$ such that for all $\mathbf{x}, \mathbf{x}' \in D^n$ which differ in a single entry,

$$\left\| f(\mathbf{x}) - f\left(\mathbf{x}'\right) \right\|_1 \leq S(f)\,.$$

Remark:

- Sensitivity is a Lipschitz condition on f : for all pairs of databases $\mathbf{x}, \mathbf{x}' \in D^n$:

$$\frac{\| f(\mathbf{x}) - f\left(\mathbf{x}'\right) \|_1}{\mathrm{d}_H\left(\mathbf{x}, \mathbf{x}'\right)} \leq S(f)\,.$$

- Can we use other distance metrics?

Examples about L_1 Sensitivity

Example (Sums)

if $D = \{0, 1\}$ and $f(\mathbf{x}) = \sum_{i=1}^{n} x_i$ (viewed as a real number), $S_{L_1}(f) = 1$.

Examples about L_1 Sensitivity

Example (Histograms)

Consider an arbitrary domain D, partitioned into d disjoint bins $B_1, \ldots, B_d$.
$f : D^n \rightarrow \mathbb{Z}^d$, computing the number of database points which fall into each bin, is called a histogram for $B_1, \ldots, B_d$,.
We have $S_{L_1}(f) = 2$, **independent of** d.

To see why:

- Changing one point in the database can change at most two of these counts: one bin loses a point, another bin gains one.

Calibrating Noise According to $S(f)$

Proposition (Non-interactive Output Perturbation)

For all $f : D^n \to \mathbb{R}^d$, the following mechanism is ϵ-indistinguishable: $\text{San}_f(\mathbf{x}) = f(\mathbf{x}) + (Y_1, \ldots, Y_d)$ where the Y_i are drawn i.i.d. from $\text{Lap}(S(f)/\epsilon)$.

Calibrating Noise According to $S(f)$

Proof.

Recall: if $y, y' \sim \text{Lap}(\lambda)$, then $h(y)/h(y') \leq e^{|y-y'|/\lambda}$.
Extend to high dimensions: if Y is a vector of d independent Laplace variables, the density function at y is proportional to $\exp\left(-\|y\|_1/\lambda\right)$.
For all $t \in \mathbb{R}^d$,

$$\frac{\Pr(z+Y=t)}{\Pr\left(z'+Y=t\right)} = \frac{\Pr(Y=t-z)}{\Pr\left(Y=t-z'\right)} \in \exp\left(\pm\frac{\|z-z'\|_1}{\lambda}\right).$$

Let $\frac{S(f)}{\lambda} = \epsilon$, we have $\lambda = \frac{S(f)}{\epsilon}$.
Then,

$$\frac{\Pr(z+Y=t)}{\Pr\left(z'+Y=t\right)} \leq \exp(\epsilon).$$

$\square$

Adaptive Query

What if the ith query can depend on $1, \ldots, i-1$th queries?

Adaptive Query

More notations:

- A transcript $t = [Q_1, a_1, Q_2, a_2 \ldots, Q_d, a_d]$ is a sequence of questions and answers.
- Assume that Q_i is a well defined function of $a_1, \ldots, a_{i-1}$, and that we can therefore truncate our transcripts to be only a vector $t = [a_1, a_2, \ldots, a_d]$[6] .
- For any transcript t, we will let $f_t : D^n \to R^d$ be the function whose i th coordinate reflects the query Q_i, determined by the first $i - 1$ components of t.

Privacy Guarantee on Adaptive Query

Consider a trusted server, holding $\mathbf{x}$.

- receive an adaptive sequence of queries $f_1, f_2, f_3, \ldots, f_d$.
- each $f_i : D^n \to \mathbb{R}$.

For each query, the server San

- either refuses to answer,
- or answers $f_i(\mathbf{x}) + \text{Lap}(\lambda)$.

Theorem (Privacy Guarantee on Adaptive Query)

For an arbitrary adversary $\mathcal{A}$, let $f_t(\mathbf{x}) : D^n \to \mathbb{R}^d$ be its query function as parameterized by a transcript t. If $\lambda = \max_t S(f_t)/\epsilon$, the mechanism above is ϵ-indistinguishable.

Privacy guarantee on Adaptive Query

Proof.

Using conditional probability and writing t_i for the indices of t,

$$\frac{\Pr\left[\mathrm{San}_f(\mathbf{x}) = t\right]}{\Pr\left[\mathrm{San}_f\left(\mathbf{x}'\right) = t\right]} = \prod_i \frac{\Pr\left[\mathrm{San}_f(\mathbf{x})_i = t_i \mid t_1, \ldots, t_{i-1}\right]}{\Pr\left[\mathrm{San}_f\left(\mathbf{x}'\right)_i = t_i \mid t_1, \ldots, t_{i-1}\right]} .$$

For the ith term,

$$\frac{\Pr\left[\mathrm{San}_f(\mathbf{x})_i = t_i \mid t_1, \ldots, t_{i-1}\right]}{\Pr\left[\mathrm{San}_f\left(\mathbf{x}'\right)_i = t_i \mid t_1, \ldots, t_{i-1}\right]} \leq \exp\left(\left|f_t(\mathbf{x})_i - f_t\left(\mathbf{x}'\right)_i\right| / \lambda\right) .$$

Thus,

$$\begin{aligned}\prod_i \frac{\Pr\left[\mathrm{San}_f(\mathbf{x})_i = t_i \mid t_1, \ldots, t_{i-1}\right]}{\Pr\left[\mathrm{San}_f\left(\mathbf{x}'\right)_i = t_i \mid t_1, \ldots, t_{i-1}\right]} &\leq \prod_i \exp\left(\left|f_t(\mathbf{x})_i - f_t\left(\mathbf{x}'\right)_i\right| / \lambda\right) \\ &= \exp\left(\left\|f_t(\mathbf{x}) - f_t\left(\mathbf{x}'\right)\right\|_1 / \lambda\right) .\end{aligned}$$

We complete the proof using the bound $\forall t, S\left(f_t\right) \leq \lambda\epsilon$. $\square$

Sensitivity in General Metric Spaces

What if the distance metric is not L_1?

Sensitivity in General Metric Spaces

We extend L_1 distance to the general distance metric $\mathrm{d}_{\mathcal{M}}$ on the output $f(\mathbf{x})$.

- Symmetry: $\mathrm{d}_{\mathcal{M}}(x,y) = \mathrm{d}_{\mathcal{M}}(y,x)$.
- The triangle inequality: $\mathrm{d}_{\mathcal{M}}(x,y) \leq \mathrm{d}_{\mathcal{M}}(x,z) + \mathrm{d}_{\mathcal{M}}(z,y)$.

Definition (Sensitivity in General Metric Spaces)

Let $\mathcal{M}$ be a metric space with a distance function $\mathrm{d}_{\mathcal{M}}(\cdot,\cdot)$. The sensitivity $S_{\mathcal{M}}(f)$ of a function $f : D^n \to \mathcal{M}$ is the amount that the function value varies when a single entry of the input is changed.

$$S_{\mathcal{M}}(f) \stackrel{\text{def}}{=} \sup_{\mathbf{x},\mathbf{x}' :\ \mathrm{d}_H(\mathbf{x},\mathbf{x}')=1} \mathrm{d}_{\mathcal{M}}\left(f(\mathbf{x}), f\left(\mathbf{x}'\right)\right) .$$

New Mechanism on New Sensitivity

Lap(λ) **only applies to L_1 sensitivity!**

Given a point $z \in \mathcal{M}$, (and a measure on $\mathcal{M}$) we define a probability density function

$$h_{z,\epsilon}(y) \propto \exp\left(\frac{-\epsilon \cdot \mathrm{d}_{\mathcal{M}}(y,z)}{2 \cdot S_{\mathcal{M}}(f)}\right).$$

To reveal an approximate version of $f(\mathbf{x})$ with sensitivity S, one can sample a value according to $h_{f(\mathbf{x}),\epsilon/S}()$.

$$\Pr[\mathcal{T}(\mathbf{x}) = y] = \frac{\exp\left(\frac{-\epsilon}{2S_{\mathcal{M}}(f)} \cdot d_{\mathcal{M}}(y, f(\mathbf{x}))\right)}{\int_{y \in \mathcal{M}} \exp\left(\frac{-\epsilon}{2S_{\mathcal{M}}(f)} \cdot d_{\mathcal{M}}(y, f(\mathbf{x}))\right) dy}.$$

Theorem (Privacy Guarantee of New Mechanism)

In a metric space where $h_{f(\mathbf{x}),\epsilon}()$ is well-defined, adding noise to $f(\mathbf{x})$ as above yields an ϵ-indistinguishable scheme.

Proof of New Mechanism

Proof.

Let x and x' be two databases differing in one entry.
First, $\mathrm{d}_{\mathcal{M}}\left(f(\mathbf{x}), f\left(\mathbf{x}'\right)\right) \leq S(f)$.
For any y,

$$\begin{aligned}\frac{\exp\left(\mathrm{d}_{\mathcal{M}}(y, f(\mathbf{x}))\right)}{\exp\left(\mathrm{d}_{\mathcal{M}}\left(y, f\left(\mathbf{x}'\right)\right)\right)} &= \exp(\mathrm{d}_{\mathcal{M}}(y, f(\mathbf{x})) - \mathrm{d}_{\mathcal{M}}(y, f(\mathbf{x}'))) \\ &\leq \exp\left(\mathrm{d}_{\mathcal{M}}(f(\mathbf{x}'), f(\mathbf{x}))\right) \leq e^{S(f)}.\end{aligned}$$

Similarly, $\dfrac{\exp\left(\frac{-\epsilon}{2S(f)} \cdot d_{\mathcal{M}}(y, f(\mathbf{x}))\right)}{\exp\left(\frac{-\epsilon}{2S(f)} \cdot d_{\mathcal{M}}(y, f(\mathbf{x}'))\right)} \leq e^{\epsilon/2}$.

Finally, the normalization constant $\int_{y \in \mathcal{M}} \exp\left(\frac{-\epsilon \cdot d_{\mathcal{M}}(y, f(\mathbf{x}))}{2S(f)}\right) dy$ also differs by a factor of at most $e^{\epsilon/2}$ between $\mathbf{x}$ and $\mathbf{x}'$.
Thus,

$$h_{f(\mathbf{x}),\epsilon}(y) / h_{f(\mathbf{x}'),\epsilon}(y) \leq e^{\epsilon/2} \cdot e^{\epsilon/2} = e^{\epsilon}.$$

Discussions on New Mechanism

We denote the new mechanism as $\mathcal{G}$.

Comparison between Lap and $\mathcal{G}$:

- Sensitivity.
  - Lap uses L_1 sensitivity.
  - $\mathcal{G}$ uses general distance metrics.
- Method.
  - Lap draws a noise and adds onto the output.
  - $\mathcal{G}$ directly draws the output from the distribution.
- Distribution.
  - Lap: $h(y) \propto \exp[-\epsilon \|y\|_1 / S_{L_1}(f)]$.
  - $\mathcal{G}$: $h(y) \propto \exp\left(-\epsilon \cdot \mathrm{d}_{\mathcal{M}}(y, f(\boldsymbol{x}))/2S_{\mathcal{M}}(f)\right)$.

Discussions on New Mechanism: Transform $\mathcal{G}$ to Lap

Let us see how $\mathcal{G}$ can be equal to Lap when $d_{\mathcal{M}}$ is L_1 distance.
Let $d_{\mathcal{M}}$ be L_1 distance.
Recap:

$$\mathcal{G} : h(y) \propto \exp\left(-\epsilon \cdot \mathrm{d}_{\mathcal{M}}(y, f(\boldsymbol{x}))/2S_{\mathcal{M}}(f)\right) .$$

We can view y as $f(x) + \eta$, where η is a noise we draw. Then, we will get the distribution of η similar to Lap,

$$\mathcal{G} : h(\eta) \propto \exp\left(-\epsilon \cdot \|\eta\|_1/2S_{L_1}(f)\right) .$$

We can actually git rid of the factor of 2.
How?

Discussions on New Mechanism: Transform $\mathcal{G}$ to Lap

Recap how we prove the privacy guarantee previously, where we bound the influence of the normalization factor,

$$\frac{\int_{y\in\mathcal{M}} \exp\left(\frac{-\epsilon\cdot d_{\mathcal{M}}(y,f(\mathbf{x}))}{2S(f)}\right) dy}{\int_{y\in\mathcal{M}} \exp\left(\frac{-\epsilon\cdot d_{\mathcal{M}}(y,f(\mathbf{x}'))}{2S(f)}\right) dy} \leq e^{\epsilon/2}.$$

If the normalization factor does not depend on $f(\boldsymbol{x})$, this equation is equal to 1 and further we can use a relatively smaller noise, by removing the factor 2, which is exactly

$$\text{Lap}: h(\eta) \propto \exp[-\epsilon\|\eta\|_1/S_{L_1}(f)].$$

Non-interactive Mechanisms

Interactive and Non-interactive

Interactive setting:

- answer queries of the form $f_g(\mathbf{x}) = \sum_{i=1}^{n} g(i, x_i)$ where $g : [n] \times D \to [0, 1]$.
- $S_{L_1}(f_g) = 1$.

Interactive and Non-interactive

Suppose the domain D is $\{0,1\}^d$.
For non-interactive ϵ-indistinguishable mechanism San:

- Many functions f_g "cannot be answered" by $\mathcal{T}_{\text{San}}$. which means, it is not possible to distinguish
  - the sanitization of a database where all entries satisfy $g(i, x_i) = 0$
  - a database where all entries satisfy $g(i, x_i) = 1$
- Unless the database consists of at least $2^{\Omega(d)}$ points.

Interactive and Non-interactive

Consider Boolean functions $g_{\mathbf{r}}$ of a specific form.

- n non-zero binary strings $\mathbf{r} = (r_1, r_2, \ldots, r_n)$, $r_i \in \{0,1\}^d$
- $g_{\mathbf{r}}(i,x)$: the inner product, modulo 2, of r_i and x, that is $g_{\mathbf{r}}(i,x) = \bigoplus_j x^{(j)} r_i^{(j)}$, denoted $r_i \odot x$, written as g.

Theorem (Non-interactive Schemes Require Large Databases)

Suppose that San is an ϵ-indistinguishable non-interactive mechanism with domain $D = \{0,1\}^d$. For at least $2/3$ of the functions of the form $f_g(\mathbf{x}) = \sum_i g(i, x_i)$, the following two distributions have statistical difference $O\left(n^{4/3}\epsilon^{2/3}2^{-d/3}\right)$:

Distribution 0: $\mathcal{T}_{San}(\mathbf{x})$ where $\mathbf{x} \in_R \{\mathbf{x} \in D^n : f_g(\mathbf{x}) = 0\}$

Distribution 1: $\mathcal{T}_{San}(\mathbf{x})$ where $\mathbf{x} \in_R \{\mathbf{x} \in D^n : f_g(\mathbf{x}) = n\}$

Prove Separation Results

For any r, partition the domain D into two sets:

- $D_r = \left\{x \in \{0,1\}^d : r \odot x = 0\right\}$
- $\bar{D}_r = D \backslash D_r = \left\{x \in \{0,1\}^d : r \odot x = 1\right\}$
- We abuse notation and let D_r also stand for a random vector chosen uniformly from that set (similarly for D and $\bar{D}_r$).

Prove RSAOD

Lemma (Random Subsets Approximate the Output Distribution (RSAOD))

Let $Z : D \to \{0,1\}^*$ be a randomized map such that for all pairs $x, x' \in D$, and all outputs z, $\frac{\Pr[Z(x)=z]}{\Pr[Z(x')=z]} \in \exp(\pm\epsilon)$. For all $\alpha > 0$: with probability at least $1-\alpha$ over $r \in \{0,1\}^d \setminus \{0^d\}$,

$$\mathbf{SD}\left(Z\left(D_r\right), Z(D)\right) \leq O\left(\frac{\epsilon^2}{\alpha \cdot 2^d}\right)^{1/3}$$

The same statement holds for $\bar{D}_r$.

Prove RSAOD

Proof.

Let $p(z|x)$ denote the probability that $Z(x) = z$. If x is chosen uniformly in $\{0,1\}^d$, then

$$p(z) = \sum_x p(z|x)p(x) = \frac{1}{2^d}\sum_x p(z|x)\,.$$

For symmetry and simplification, we pick an offset bit b, and look at the set $D_{r,b} = \left\{x \in \{0,1\}^d : r \odot x = b\right\}$.

Prove RSAOD

Then,

- Let $\hat{p}(z) = \Pr\left[Z\left(D_{r,b}\right) = z\right]$, where the probability is taken over the coin flips of Z and the choice of $x \in D_{r,b}$.
- For a fixed z, $\hat{p}(z)$ is a random variable depending on the choice of r, b.
- $\mathbb{E}_{r,b}[\hat{p}(z)] = p(z)$.

We want that

- $\mathrm{Var}_{r,b}[\hat{p}(z)]$ is constrained.

Claim

$\mathrm{Var}_{r,b}[\hat{p}(z)] \leq \frac{2 \cdot \tilde{\epsilon}^2 \cdot p(z)^2}{2^d}$, where $\tilde{\epsilon} = e^{\epsilon} - 1$.

Prove RSAOD

We say a value z is $\delta - good$ for a pair (r, b) if

$$\hat{p}(z) - p(z) \le \delta \cdot p(z) .$$

By the Chebyshev bound $\Pr(|X - \mu| \ge k\sigma) \le \frac{1}{k^2}$, with $k^2 = \frac{\delta^2 p(z)^2}{\mathrm{Var}[\hat{p}(z)]}$, for all z,

$$\Pr_{r,b}[z \text{ is not } \delta\text{-good for } (r, b)] \le \frac{\mathrm{Var}[\hat{p}(z)]}{\delta^2 p(z)^2} \le \frac{2\tilde{\epsilon}^2}{\delta^2 2^d} = \beta .$$

If we take the distribution on z given by $p(z)$, then with probability at least $1 - \alpha$ over pairs (r, b), the fraction of z 's (under $p(\cdot)$) which are good is at least $1 - \beta\alpha$.

Prove RSAOD

Finally, if a $1-\beta\alpha$ fraction of the z 's are δ-good for a particular pair (r, b), set $\delta=\sqrt[3]{\frac{2\tilde{\epsilon}^2\alpha}{2^d}}$ and we have

$$\mathbf{SD}(\hat{p}(z), p(z)) \leq 2(1-\beta\alpha)\delta + 2\beta\alpha \leq 2(\beta\alpha+\delta) \leq 4\delta\,.$$

Since $\tilde{\epsilon} < 2\epsilon$ for $\epsilon \leq 1$,

$$4\delta \leq 4\sqrt[3]{12\epsilon^2 2^{-d}}\,,$$

for at least a $1-\alpha$ fraction of the pairs (r, b).
The bit b is unimportant here,since it only switches D_r and its complement $\bar{D}_r$.
We also have

$$\mathbf{SD}(Z\left(D_r\right), Z(D)) = \mathbf{SD}(Z\left(\bar{D}_r\right), Z(D))\,,$$

since $Z(D)$ is the mid point between the two.
Q.E.D.

Prove Separation Results

Lemma (Random Subsets Approximate the Output Distribution (RSAOD))

Let $Z : D \to \{0,1\}^*$ be a randomized map such that for all pairs $x, x' \in D$, and all outputs z, $\frac{\Pr[Z(x)=z]}{\Pr[Z(x')=z]} \in \exp(\pm\epsilon)$. For all $\alpha > 0$: with probability at least $1-\alpha$ over $r \in \{0,1\}^d \setminus \{0^d\}$,

$$\mathbf{SD}\left(Z\left(D_r\right), Z(D)\right) \leq O\left(\frac{\epsilon^2}{\alpha \cdot 2^d}\right)^{1/3}$$

The same statement holds for $\bar{D}_r$.

Prove Separation Results

"Distribution 0" in the statement is $\mathcal{T}_{\mathrm{San}}(D_{r_1}, \ldots, D_{r_n})$.
We want to show: With high probability over the choice of the r_i 's,

$$\mathcal{T}_{\mathrm{San}}(D_{r_1}, \ldots, D_{r_n}) \text{ is close to } \mathcal{T}(D, \ldots, D)$$

We proceed by a hybrid argument, adding one constraint at a time. For each i, we want to show

$$\mathcal{T}_{\mathrm{San}}\left(D_{r_1}, \ldots, D_{r_i}, \quad D \quad, D, \ldots, D\right) \quad \text{is close to}$$
$$\mathcal{T}_{\mathrm{San}}\left(D_{r_1}, \ldots, D_{r_i}, \quad D_{r_{i+1}}, D, \ldots, D\right)$$

Prove Separation Results

Suppose we have chosen $r_1, \ldots, r_i$ already.
For any $x \in \{0,1\}^d$, consider the randomized map where the $(i+1)$-th coordinate is fixed to x :

$$Z(x) = \mathcal{T}_{\mathrm{San}} \left(D_{r_1}, \ldots, D_{r_i}, \quad x \quad, D, \ldots, D\right)$$

Note that $Z(D)$ is equal to the i-th step in the hybrid, and $Z\left(D_{r_{i+1}}\right)$ is equal to the $(i+1)$th step.

Prove Separation Results

By ϵ-indistinguishability of San,

$$\frac{\Pr[Z(x)=z]}{\Pr\left[Z\left(x'\right)=z\right]} \in \exp(\pm\epsilon)\,.$$

Use Lemma RSAOD and set $\alpha = \frac{1}{6n}$. With $\Pr \geq 1 - \frac{1}{6n}$,

$$\mathbf{SD}\left(Z\left(D_{r_i}\right), Z(D)\right) \leq O\left(\frac{n\epsilon^2}{2^d}\right)^{1/3} = O(\sigma)\,,$$

denoted as event A_i.

Prove Separation Results

By a union bound,

$$\Pr(\cup_{i=1}^{n} \bar{A}_i) \leq \sum_{i=1}^{n} \Pr(\bar{A}_i) \leq \frac{1}{6n} \cdot n = \frac{1}{6}.$$

Thus,

$$\Pr(\cap_{i=1}^{n} A_i) \geq \frac{5}{6}.$$

In this case, the total distance is $n\sigma$. Denote $A = \cap_{i=1}^{n} A_i$. Similarly, for Distribution 1, with probability at least $\frac{5}{6}$, the total distance is $n\sigma$, denoted as B.

Prove Separation Results

Again by a union bound,

$$\Pr(\bar{A} \cup \bar{B}) \leq \frac{1}{3}.$$

Thus,

$$\Pr(A \cap B) \geq \frac{2}{3},$$

and the distance between Distributions 0 and 1 is at most $2n\sigma = O\left(n^{4/3}\epsilon^{2/3}2^{-d/3}\right)$.

Thanks.
Questions?

Appendices

Prove Claim

Claim

$\mathrm{Var}_{r,b}[\hat{p}(z)] \leq \frac{2\cdot\tilde{\epsilon}^2\cdot p(z)^2}{2^d}$, where $\tilde{\epsilon} = e^{\epsilon} - 1$.

Proof.

Recall:

- $\hat{p}(z) = \Pr[Z(D_{r,b}) = z]$.
- $p(z) = \frac{1}{2^d}\sum_x p(z|x)$.
- $\mathbb{E}_{r,b}[\hat{p}(z)] = p(z)$.

Let:

- p^* be the minimum over x of $p(z \mid x)$.
- $q_x = p(z|x) - p^*$ and $\bar{q} = p(z) - p^*$.

We can write:

$$\hat{p}(z) - p^* = \frac{2}{2^d}\sum_x q_x \chi_0(x),$$

where $\chi_0(x)$ is 1 if $x \in D_{r,b}$. And $\mathbb{E}[\hat{p}(z) - p^*] = \bar{q} = \frac{1}{2^d}\sum_x q_x$.

Prove Claim

$$
\begin{aligned}
\mathrm{Var}_{r,b}[\hat{p}(z)] &= \mathrm{Var}_{r,b}[\hat{p}(z) - p^*] \\
&= \mathop{\mathbb{E}}_{r,b}\left[\left(\frac{2}{2^d}\sum_x q_x \chi_0(x) - \frac{1}{2^d}\sum_x q_x\right)^2\right] \\
&= \mathop{\mathbb{E}}_{r,b}\left[\left(\frac{1}{2^d}\sum_x q_x\left(2\chi_0(x) - 1\right)\right)^2\right]
\end{aligned}
$$

Now $(2\chi_0(x) - 1) = (-1)^{r\odot x\oplus b}$. Thus,

$$\mathbb{E}_{r,b}[2\chi_0(x) - 1] = 0\,.$$

Moreover, for $x \neq y$,

$$\mathbb{E}_{r,b}[(2\chi_0(x) - 1)(2\chi_0(y) - 1)] = 1/2^d\,.$$

(if we chose r with no restriction it would be 0, but we have the restriction that $r \neq 0^d$) .

Prove Claim

Expanding the square of the variance,

$$\begin{aligned}\operatorname*{Var}_{r,b}[\hat{p}(z)] &= \frac{1}{2^{2d}}\sum_{x} q_x^2 + \frac{1}{2^{3d}}\sum_{x\neq y} q_x q_y \\ &= \frac{1-\frac{1}{2^d}}{2^{2d}}\sum_{x} q_x^2 + \frac{1}{2^d}\left(\frac{1}{2^d}\sum_{x} q_x\right)^2 \\ &\leq \frac{1}{2^d}\left(\max_x q_x^2 + \bar{q}^2\right) .\end{aligned}$$

By the indistinguishability condition,

$$\left(\max_x q_x\right) \leq \left(e^{\epsilon}-1\right) p^* \leq \tilde{\epsilon}\cdot p(z) ,$$

$$\bar{q} \leq \left(e^{\epsilon}-1\right) p^* \leq \tilde{\epsilon}\cdot p(z) .$$

Plugging this into the last equation proves Claim 1 .

Reference

1. Dwork, Cynthia, et al. "Calibrating noise to sensitivity in private data analysis." Theory of cryptography conference. Springer, Berlin, Heidelberg, 2006.
2. https://slideplayer.com/slide/5223315/