Hellinger Differential Privacy

Hellinger Differential Privacy

Fengnan Deng

College of Engineering and Computing
STATISTICS
George Mason University

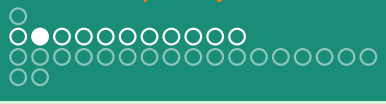
Department of Statistics, George Mason University

Table of Contents

Current Section

Differential privacy

Motivation and backgrounds

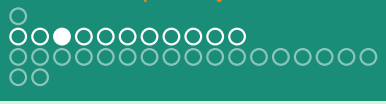


Problem description

1. **Data collection**: Many entities collect individually identifiable data to provide personalized services and **share** them with other organizations to improve and enhance the quality of the service.

2. **Risks:** Sharing sensitive data may release personal information, which is illegal.

3. **Privacy protection:** Statistical methods, such as **differential privacy**, can be used to prevent re-identification while remain statistical structure of the data.



Differential privacy

1. **Goal:** Reduce the risk of re-identification of personal information based on a statistic.

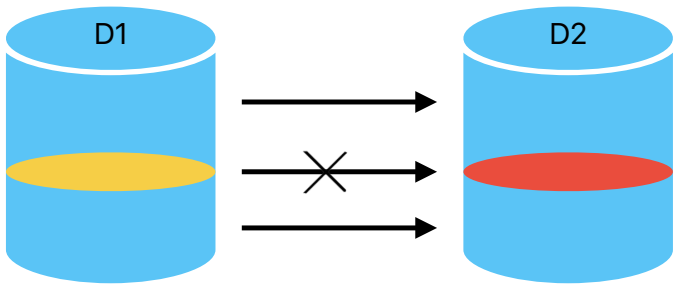
2. **Method:** Introduce randomness to the statistic, for example: adding an appropriate noise.

3. **Applications:** US Census Bureau, IT companies such as Google, Apple, Healthcare organizations such as Epic, Cardinal Health, etc.

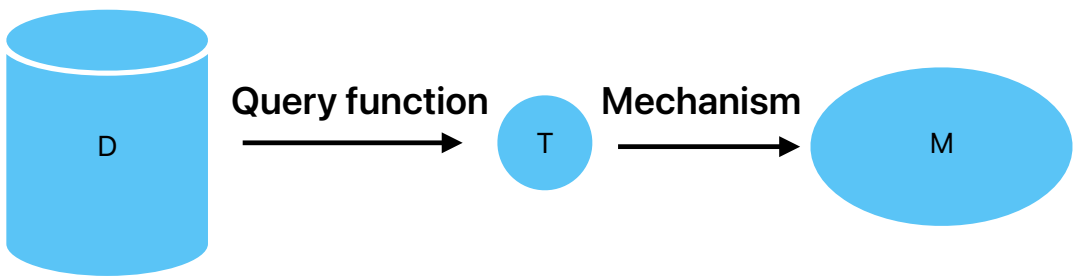
4. **Question:** How to decide the "**correct**" amount of randomness?

Basic terminology

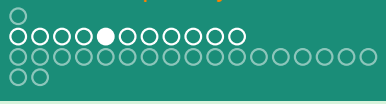
- Adjacent datasets:



- Query function and mechanism:



- A **mechanism** is a perturbation of the **query function**.

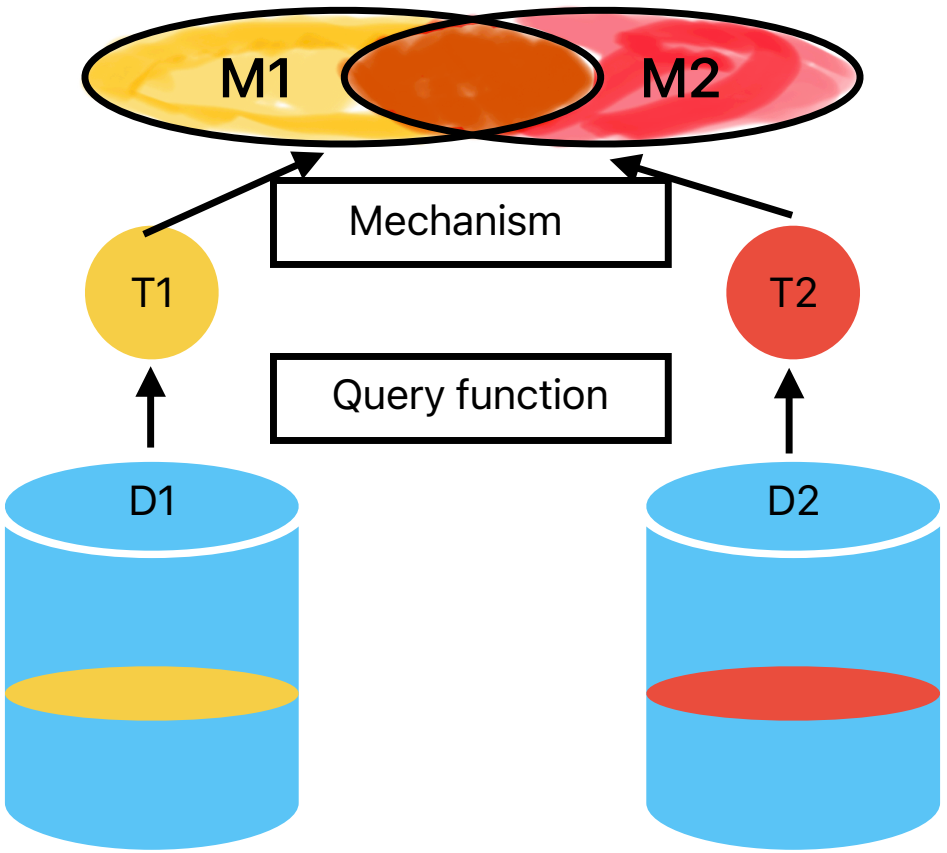


Basic Idea of DP

The basic ideas behind DP are the following:

- Start with a query and identify the **sensitivity** of the query using **adjacent** datasets.
- Identify an appropriate **mechanism** to perturb the output from the query.
- Release the perturbed output.

Basic Idea of DP



Mechanism

- We focus on additive mechanism: $M(T) = T + \epsilon$.
  - Gaussian mechanism: $\epsilon \sim N(0, \sigma^2)$.
  - Laplace mechanism: $\epsilon \sim Lap(0, b)$.

Divergence

- Divergence between 2 distributions, $D(M_1||M_2)$, can be used to measure the difficulty distinguishing M_1 from M_2.

- Definition of divergence: For 2 distributions with density $p(x), q(x)$,

  - $D(p||q) \geq 0$.
  - $D(p||q) = 0 \iff p(x) = q(x)$.

  Note that the divergence may not be a metric.

Types of differential privacy

- Max divergence: $\epsilon - DP$. [Dwork et al. (2006)]

$$D_\infty(M_1||M_2) = \max_{S \in Supp(M_2)} \log \frac{\mathbf{P}(M_1 \in S)}{\mathbf{P}(M_2 \in S)} \leq \epsilon$$

- A relaxation of $\epsilon - DP$: $(\epsilon, \delta)-$ differential privacy.

$$D_\infty^{(\delta)}(M_1||M_2) = \max_{S \in Supp(M_2)} \log \frac{\mathbf{P}(M_1 \in S) - \delta}{\mathbf{P}(M_2 \in S)} \leq \epsilon$$

Types of differential privacy

A common way to write $\epsilon - DP$ and $(\epsilon, \delta) - DP$ is as follows:

- $\epsilon - DP$:

$$\mathbf{P}(M_1 \in S) \leq e^{\epsilon} \cdot \mathbf{P}(M_2 \in S).$$

- $(\epsilon, \delta) - DP$:

$$\mathbf{P}(M_1 \in S) \leq e^{\epsilon} \cdot \mathbf{P}(M_2 \in S) + \delta.$$

Types of differential privacy

- Rényi divergence: $(\alpha, \epsilon)-$Rényi differential privacy ($\alpha > 1$). [Mironov (2017)]

$$D_\alpha(M_1||M_2) = \frac{1}{\alpha - 1} \log \int \left(\frac{f_{M_1}(x)}{f_{M_2}(x)} \right)^\alpha f_{M_2}(x) \mathrm{d}\mu(x) \leq \epsilon$$

- Hypothesis testing approach: μ-Gaussian differential privacy. [Dong et al. (2022)],[Avella-Medina et al. (2023)].

$$H : x \sim f_{M_1}(x) \quad v.s. \quad K : x \sim f_{M_2}(x)$$

$$H : x \sim N(0, 1) \quad v.s. \quad K : x \sim N(\mu, 1)$$

Differential privacy

Hellinger differential privacy

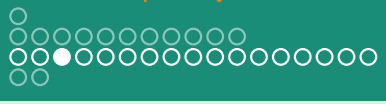
Hellinger differential privacy

For two density function p, q, squared Hellinger distance is defined as

$$D_{HD}(p, q) = \int_{\mathbb{R}} \left(\sqrt{p(x)} - \sqrt{q(x)} \right)^2 \mathrm{d}x$$

- $\epsilon-$HDP [Deng and Vidyashankar (2025b)]: A mechanism M is said to satisfy $\epsilon-$Hellinger differential privacy if

$$D_{HD}(M_1, M_2) \leq \epsilon$$



Power Divergence Privacy

- The power divergence between two densities p and q is

$$D_\lambda(p,q) = \frac{1}{\lambda(\lambda+1)}\mathbf{E}_q\left[\left(\frac{p(X)}{q(X)}\right)^{\lambda+1} - 1\right].$$

- A mechanism M is said to satisfy $(\lambda,\epsilon)-$Power divergence differential privacy (PDP) if

$$D_\lambda(M_1, M_2) \leq \epsilon.$$

- When $\lambda = -\frac{1}{2}$, we obtain $2HD^2(p,q)$.

Power Divergence Privacy

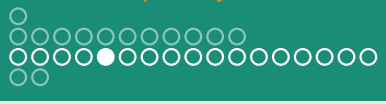
- If M is additive Gaussian mechanism,

$$D_\lambda(M_1, M_2) = \frac{1}{\lambda(\lambda+1)} \left[e^{\frac{\lambda(\lambda+1)||W(D)-W(D')||_2^2}{2\sigma^2}} - 1 \right].$$

- Set $\Delta_{L_2} W = \sup_{D,D'} ||W(D) - W(D')||_2$.
- Then to obtain $(\lambda, \epsilon)-$ PDP the variance of the perturbation is given by

$$\sigma^2 = \sigma^2_{\lambda,\epsilon} = \begin{cases} (\Delta_{L_2} W)^2 \cdot \frac{\lambda(\lambda+1)}{2\log(1+\lambda(\lambda+1)\epsilon)} & \text{if } \lambda(\lambda+1) \neq 0 \\ (\Delta_{L_2} W)^2 \cdot \frac{1}{2\epsilon} & \text{otherwise.} \end{cases}$$

- Under additional calculations, one can show $\sigma^2_{\lambda,\epsilon}$ is minimized at $\lambda = -\frac{1}{2}$.



Hellinger differential privacy

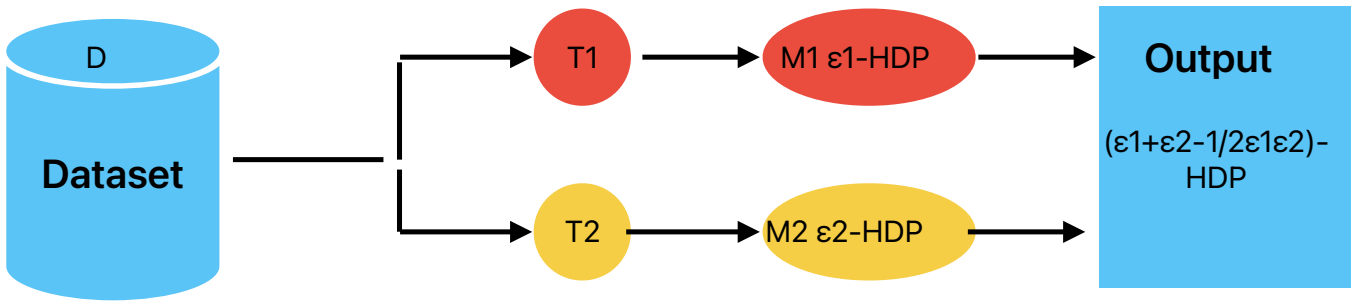
- **Relationship with other types of differential privacy**:
  - $\epsilon - HDP$ implies $(\epsilon', \delta') - DP$, $\epsilon' = 0, \delta' = \epsilon$.
  - $\epsilon - HDP$ implies $\mu - GDP$, $\mu = 2\Phi^{-1}(\frac{\epsilon+1}{2})$.

- **Post processing guarantee**

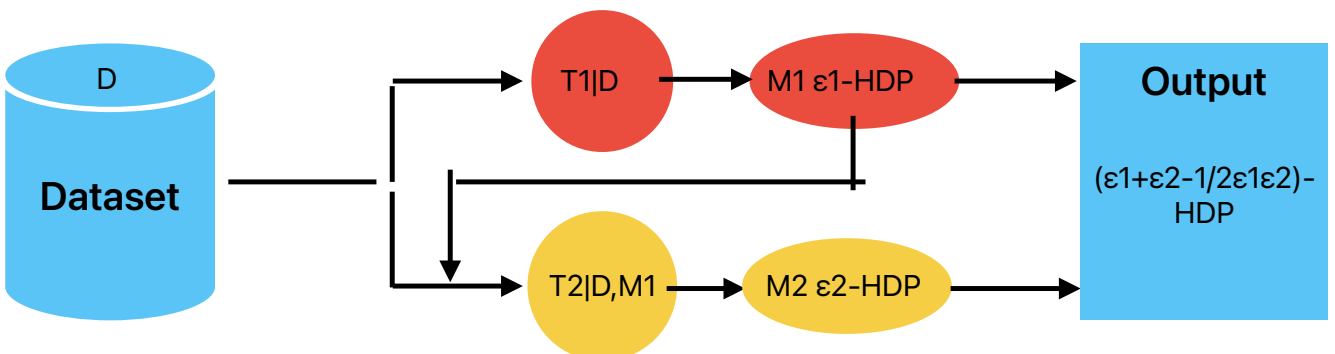
$$M(T) \sim \epsilon - HDP \Longrightarrow g(M(T)) \sim \epsilon - HDP.$$

Hellinger differential privacy

- Composition rule
  - Sequential composition



  - Adaptive composition



Minimum Hellinger distance estimation

- Let $X_1, \cdots, X_n$ be i.i.d. random variables postulated to be distributed according to $f_{\theta}(\cdot)$, $\theta \in \Theta \subset \mathbb{R}^d$. Let $g(\cdot)$ denote the true density. The minimum Hellinger distance estimator of θ is the minimizer of

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} D_{HD}(g_n, f_{\theta}),$$

  where $g_n(\cdot)$ is a nonparametric estimator of $g(\cdot)$.

- Under family regularity and identifiability conditions, the minimizer exists and equals θ_0 if $g(\cdot) = f_{\theta_0}$.

Computing the estimator

- Two possible algorithms:
  1. Gradient descent method
  2. Newton-Raphson method

- How to make the algorithm private?

Private optimization methods

- **Question**: If we adopt the additive mechanism, how do we add noise to the estimator?

- **Loss function:** $L_n(\theta)$.
- **Statistic:** $\hat{\theta}_n = \arg\min_{\theta \in \Theta} L_n(\theta)$

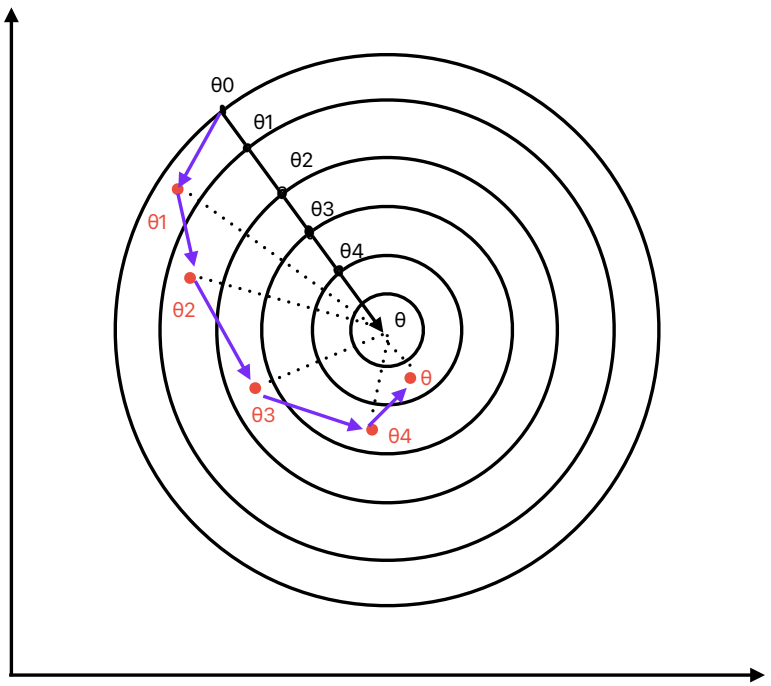
- **Answer**:
  - Inject noise directly into $\hat{\theta}$. This approach requires knowledge of the explicit distribution of $\hat{\theta}$.
  - Integrate noise into the algorithm computing $\hat{\theta}$ from the loss function $L_n(\theta)$.

Private gradient descent

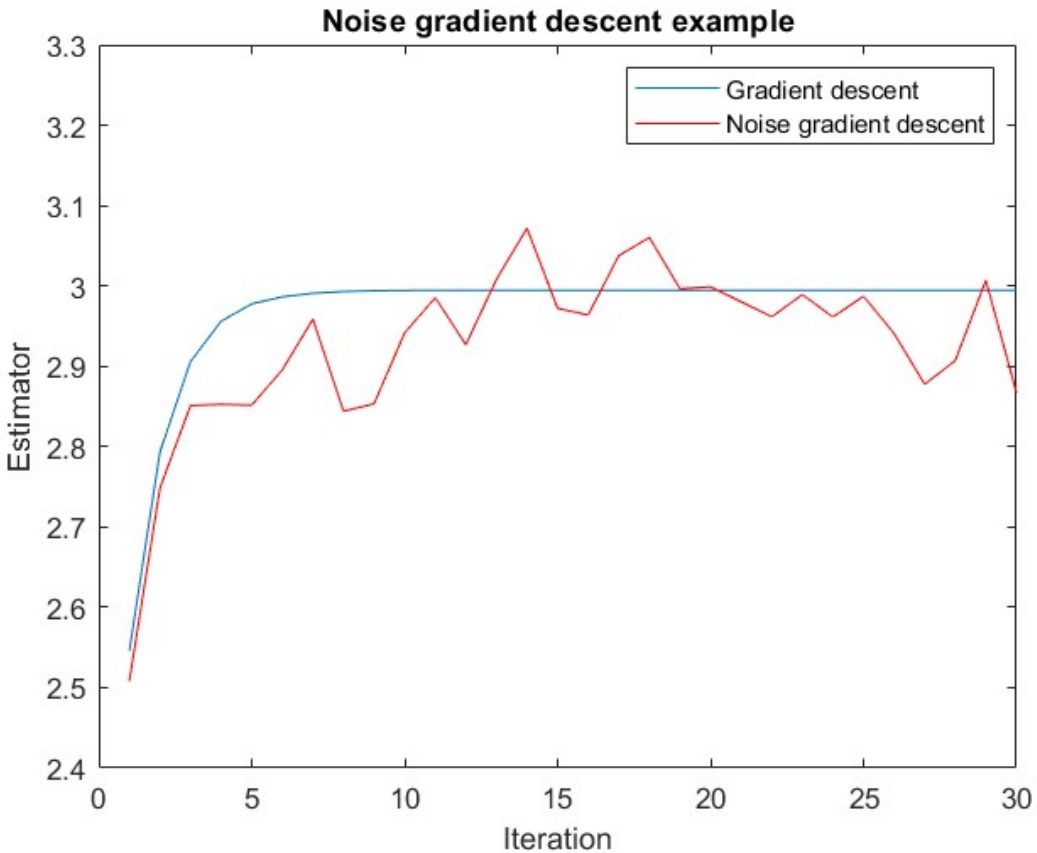
- Choose the mechanism.
- Predetermined number of iterations K.
- Identify the query and in every iteration calculate σ using the HDP property.
- Return the updated estimator.

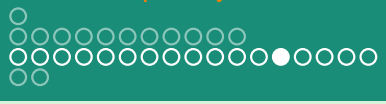
Private gradient descent (contd.)

$$\hat{\theta}_n^{(k+1)} = \hat{\theta}_n^{(k)} - \underbrace{\eta}_{\text{learning rate}} \left(\nabla L_n(\hat{\theta}_n^{(k)}) + \sigma Z_k \right)$$



Private gradient descent (contd.)

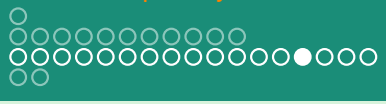




Privacy budget allocation

- If our overall privacy budget is ϵ, then we would like to distribute it over the K iterations. The *composition rule* plays a significant role in this part.

- Distributing the budget equally between iterations, namely, each iteration satisfies $\frac{\epsilon}{K} - HDP$, then $\hat{\theta}_n^{(K)}$ satisfies $\epsilon - HDP$.



Utilty

- We next ask, how close is the private estimator to the non-private estimator?

- The main result is:

$$||\hat{\theta}_n^{(K)} - \hat{\theta}_n||_2 \leq \Delta_n \cdot R(K, \epsilon, m) \text{ with high probability.}$$

- In the above,
  1. Δ_n is L_2 sensitivity of gradient (Hessian) of $L_n(\theta)$.
  2. $\Delta_n \to 0$ as $n \to \infty$.
  3. m is the dimension of θ.
  4. $R(K, \epsilon, m) \to \infty$ as $K \to \infty$.

Utility

- For M estimator, $\Delta_n \sim \frac{1}{n}$. [Avella-Medina et al. (2023)].
- For minimum Hellinger distance estimator, $\Delta_n \leq \frac{1}{\sqrt{n}}$.
  - $\{x_1, \cdots, x_n\} \overset{i.i.d.}{\sim} f_\theta(x)$.
  - $g_n(x)$ is kernel density estimator.
  - $L_n = \int \left(\sqrt{f_\theta(x)} - \sqrt{g_n(x)}\right)^2 \mathrm{d}x$.
  - Sharper bound $\Delta_n \leq \frac{1}{\sqrt{n}} \cdot \frac{1}{n^\eta}$ has been studied in Deng and Vidyashankar (2025a) to achieve efficiency. $\eta > 0$ is a positive constant.

Oracle Property

- The sharper bound plays a crucial role when establishing the limit distribution of the private MHDE. Specifically, we show that as $n \to \infty$,

$$\sqrt{n}(\hat{\theta}_n^{(K_n)} - \theta_g) \xrightarrow{d} N(\mathbf{0}, \Sigma_g)$$

and $K_n \sim \log n$ for gradient descent and $K_n \sim \log \log n$ for Newton Raphson method.

Numerical Experiments

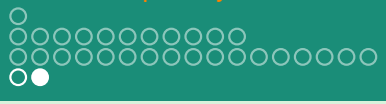
The simulated dataset contains 1000 observations with $N(5, 2^2)$. We are doing 5000 repetitions.

		ϵ		
		2.00	0.60	0.20
Estimator	μ: Mean (Std.)	4.991 (0.083)	4.989 (0.2)	4.996 (0.349)
	σ: Mean (Std.)	1.984 (0.058)	2.002 (0.144)	2.043 (0.256)
CI coverage	μ: 95%	0.861	0.836	0.824
	σ: 95%	0.819	0.933	0.927

Table 1: Results for different values of epsilon (Gradient descent)

Differential privacy

Concluding Remarks



Concluding Remarks

- We described a new differential privacy metric called $\epsilon-$ HDP.
- We discussed the private version of gradient-descent and Newton Raaphson methods for obtaining private MHDE.
- We established the utility and efficiency of our estimator.

Bibliographic References i

Avella-Medina, M., C. Bradshaw, and P.-L. Loh (2023). Differentially private inference via noisy optimization. *The Annals of Statistics 51*(5), 2067–2092.

Deng, F. and A. Vidyashankar (2025a). Conditional and unconditional sharp large deviations for branching processes with immigration. *To be submitted*.

Deng, F. and A. N. Vidyashankar (2025b). Private minimum hellinger distance estimation via hellinger distance differential privacy.

Dong, J., A. Roth, and W. J. Su (2022). Gaussian differential privacy. *Journal of the Royal Statistical Society Series B: Statistical Methodology 84*(1), 3–37.

Bibliographic References ii

Dwork, C., F. McSherry, K. Nissim, and A. Smith (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3*, pp. 265–284. Springer.

Mironov, I. (2017). Rényi differential privacy. In *2017 IEEE 30th computer security foundations symposium (CSF)*, pp. 263–275. IEEE.