Optimal Estimation of Discrete Distribution under Local Differential Privacy

Alexander Barg[1]

(joint work with Min Ye[2])

[1] University of Maryland, College Park
[2] Princeton University

SPCOM 2018

Motivation of the problem

- Responding to a survey: Y/N reveals individual preferences

Motivation of the problem

- Responding to a survey: Y/N reveals individual preferences
- Goals: Private response/accurate statistics

Motivation of the problem

- Responding to a survey: Y/N reveals individual preferences
- Goals: Private response/accurate statistics
- Classic example: Randomized Response

Suppose we would like to learn a Bernoulli distribution $(p, 1-p)$
Toss a fair coin, and

Outcome	Answer
H	Y
T	Say truth

$$\Pr(\text{No}) = \frac{1}{2}(1-p)$$

Motivation of the problem

- Responding to a survey: Y/N reveals individual preferences
- Goals: Private response/accurate statistics
- Classic example: Randomized Response

  Suppose we would like to learn a Bernoulli distribution $(p, 1-p)$
  Toss a fair coin, and

Outcome	Answer
H	Y
T	Say truth

$$\Pr(\text{No}) = \frac{1}{2}(1-p)$$

- ***Problem:*** Private estimation of discrete distribution

Motivation of the problem

- Responding to a survey: Y/N reveals individual preferences
- Goals: Private response/accurate statistics
- Classic example: Randomized Response

Suppose we would like to learn a Bernoulli distribution $(p, 1-p)$
Toss a fair coin, and

Outcome	Answer
H	Y
T	Say truth

$$\Pr(\text{No}) = \frac{1}{2}(1-p)$$

- ***Problem:*** Private estimation of discrete distribution
- Privacy degrades if the survey is repeated

Estimation of discrete distribution: Classical setup

- $\mathcal{X} = \{1, 2, \ldots, k\}$

Estimation of discrete distribution: Classical setup

- $\mathcal{X} = \{1, 2, \dots, k\}$
- $\boldsymbol{p} = (p_1, p_2, \dots, p_k), p_i = P(X = i)$

Estimation of discrete distribution: Classical setup

- $\mathcal{X} = \{1, 2, \ldots, k\}$
- $\boldsymbol{p} = (p_1, p_2, \ldots, p_k), p_i = P(X = i)$
- Observation: n i.i.d. samples $X^n := (X^{(1)}, X^{(2)}, \ldots, X^{(n)})$ drawn according to $\boldsymbol{p}$

Estimation of discrete distribution: Classical setup

- $\mathcal{X} = \{1, 2, \ldots, k\}$
- $\boldsymbol{p} = (p_1, p_2, \ldots, p_k), p_i = P(X = i)$
- Observation: n i.i.d. samples $X^n := (X^{(1)}, X^{(2)}, \ldots, X^{(n)})$ drawn according to $\boldsymbol{p}$
- Our task is to estimate $\boldsymbol{p}$ from X^n

Estimation of discrete distribution: Classical setup

- $\mathcal{X} = \{1, 2, \dots, k\}$
- $\boldsymbol{p} = (p_1, p_2, \dots, p_k), p_i = P(X = i)$
- Observation: n i.i.d. samples $X^n := (X^{(1)}, X^{(2)}, \dots, X^{(n)})$ drawn according to $\boldsymbol{p}$
- Our task is to estimate $\boldsymbol{p}$ from X^n
  1. Need to find an estimator $\hat{\boldsymbol{p}} : \mathcal{X}^n \to \mathbb{R}^k$

Estimation of discrete distribution: Classical setup

- $\mathcal{X} = \{1, 2, \ldots, k\}$
- $\boldsymbol{p} = (p_1, p_2, \ldots, p_k), p_i = P(X = i)$
- Observation: n i.i.d. samples $X^n := (X^{(1)}, X^{(2)}, \ldots, X^{(n)})$ drawn according to $\boldsymbol{p}$
- Our task is to estimate $\boldsymbol{p}$ from X^n
  1. Need to find an estimator $\hat{\boldsymbol{p}} : \mathcal{X}^n \to \mathbb{R}^k$
  2. Assess the quality of the estimator $\hat{\boldsymbol{p}}$

Private estimation

- The raw samples X^n are not accessible

Private estimation

- The raw samples X^n are not accessible
- Instead, we observe an n-tuple of *privatized samples* $Y^n = (Y^{(1)}, Y^{(2)}, \ldots, Y^{(n)})$

$$X^{(i)} \xrightarrow{\boldsymbol{Q}} Y^{(i)}$$

where the $Y^{(i)}$ take values in a set $\mathcal{Y}$

Private estimation

- The raw samples X^n are not accessible
- Instead, we observe an n-tuple of *privatized samples* $Y^n = (Y^{(1)}, Y^{(2)}, \dots, Y^{(n)})$

$$X^{(i)} \xrightarrow{\boldsymbol{Q}} Y^{(i)}$$

  where the $Y^{(i)}$ take values in a set $\mathcal{Y}$

- $\boldsymbol{Q}(y|x)$: conditional (one-dimensional) distribution from $\mathcal{X}$ to $\mathcal{Y}$
  (We can view it as a DMC)

Private estimation

- The raw samples X^n are not accessible
- Instead, we observe an n-tuple of *privatized samples* $Y^n = (Y^{(1)}, Y^{(2)}, \ldots, Y^{(n)})$

$$X^{(i)} \xrightarrow{\boldsymbol{Q}} Y^{(i)}$$

  where the $Y^{(i)}$ take values in a set $\mathcal{Y}$
- $\boldsymbol{Q}(y|x)$: conditional (one-dimensional) distribution from $\mathcal{X}$ to $\mathcal{Y}$
  (We can view it as a DMC)
- $\boldsymbol{Q}$ is called a *privatization scheme*

Private estimation

- The raw samples X^n are not accessible
- Instead, we observe an n-tuple of *privatized samples* $Y^n = (Y^{(1)}, Y^{(2)}, \ldots, Y^{(n)})$

$$X^{(i)} \xrightarrow{\boldsymbol{Q}} Y^{(i)}$$

  where the $Y^{(i)}$ take values in a set $\mathcal{Y}$
- $\boldsymbol{Q}(y|x)$: conditional (one-dimensional) distribution from $\mathcal{X}$ to $\mathcal{Y}$
  (We can view it as a DMC)
- ***Q*** is called a ***privatization scheme***
- $\mathcal{Y}$ isn't necessarily the same as $\mathcal{X}$

Private estimation

- The raw samples X^n are not accessible
- Instead, we observe an n-tuple of *privatized samples* $Y^n = (Y^{(1)}, Y^{(2)}, \dots, Y^{(n)})$

$$X^{(i)} \xrightarrow{\boldsymbol{Q}} Y^{(i)}$$

  where the $Y^{(i)}$ take values in a set $\mathcal{Y}$
- $\boldsymbol{Q}(y|x)$: conditional (one-dimensional) distribution from $\mathcal{X}$ to $\mathcal{Y}$ (We can view it as a DMC)
- $\boldsymbol{Q}$ is called a *privatization scheme*
- $\mathcal{Y}$ isn't necessarily the same as $\mathcal{X}$

Problem: Estimate $\boldsymbol{p}$ from Y^n

Local Differential Privacy

Definition (DUCHI, JORDAN, AND WAINWRIGHT, FOCS '13)

For a given $\epsilon > 0$, a privatization scheme $\boldsymbol{Q} : \mathcal{X} \to \mathcal{Y}$ is said to be ϵ-*locally differentially private* (ϵ-*LDP*) if

$$\ln \frac{\boldsymbol{Q}(Y = y|X = x)}{\boldsymbol{Q}(Y = y|X = x')} \leq \epsilon \text{ for all } x, x' \in \mathcal{X} \text{ and all } y \in \mathcal{Y}$$

where we assume that $|\mathcal{Y}| < \infty$.

Local Differential Privacy

Definition (Duchi, Jordan, and Wainwright, FOCS '13)

For a given $\epsilon > 0$, a privatization scheme $\boldsymbol{Q} : \mathcal{X} \to \mathcal{Y}$ is said to be *ϵ-locally differentially private (ϵ-LDP)* if

$$\ln \frac{\boldsymbol{Q}(Y = y|X = x)}{\boldsymbol{Q}(Y = y|X = x')} \leq \epsilon \text{ for all } x, x' \in \mathcal{X} \text{ and all } y \in \mathcal{Y}$$

where we assume that $|\mathcal{Y}| < \infty$.

- Large ϵ ($\epsilon \approx k$) means low privacy
  Small ϵ ($\epsilon \approx 1$) means high privacy

Local Differential Privacy

Definition (Duchi, Jordan, and Wainwright, FOCS '13)

For a given $\epsilon > 0$, a privatization scheme $\boldsymbol{Q} : \mathcal{X} \to \mathcal{Y}$ is said to be ϵ-*locally differentially private (ϵ-LDP)* if

$$\ln \frac{\boldsymbol{Q}(Y = y | X = x)}{\boldsymbol{Q}(Y = y | X = x')} \leq \epsilon \text{ for all } x, x' \in \mathcal{X} \text{ and all } y \in \mathcal{Y}$$

where we assume that $|\mathcal{Y}| < \infty$.

- Large ϵ ($\epsilon \approx k$) means low privacy
  Small ϵ ($\epsilon \approx 1$) means high privacy
- This definition extends to arbitrary $\mathcal{Y}$, but this yields no gain over finite $\mathcal{Y}$

Minimax estimation

- Throughout this talk we focus on minimax estimation

Minimax estimation

- Throughout this talk we focus on minimax estimation
- Assess $\hat{\boldsymbol{p}}$ by its worst-case estimation loss:

$$r_{k,n}^{\ell}(\hat{\boldsymbol{p}}) = \sup_{\boldsymbol{p}\in\Delta_k} \mathbb{E}_{X^n\sim \boldsymbol{p}^n}\ell(\hat{\boldsymbol{p}}(X^n), \boldsymbol{p})$$

where

- ℓ is a loss function, e.g., $\ell = \ell_2^2$ (Mean Square Error) or $\ell = \ell_1$
- $\Delta_k := \{\boldsymbol{p} \in (\mathbb{R}_+)^k : \sum_{i=1}^k p_i = 1\}$ is the *probability simplex*

Minimax estimation

- Throughout this talk we focus on minimax estimation
- Assess $\hat{\boldsymbol{p}}$ by its worst-case estimation loss:

$$r_{k,n}^{\ell}(\hat{\boldsymbol{p}}) = \sup_{\boldsymbol{p} \in \Delta_k} \mathbb{E}_{X^n \sim \boldsymbol{p}^n} \ell(\hat{\boldsymbol{p}}(X^n), \boldsymbol{p})$$

where
  - ℓ is a loss function, e.g., $\ell = \ell_2^2$ (Mean Square Error) or $\ell = \ell_1$
  - $\Delta_k := \{\boldsymbol{p} \in (\mathbb{R}_+)^k : \sum_{i=1}^k p_i = 1\}$ is the *probability simplex*

Define

$$r_{k,n}^{\ell} := \inf_{\hat{\boldsymbol{p}}} r_{k,n}^{\ell}(\hat{\boldsymbol{p}})$$

Minimax estimation

- Throughout this talk we focus on minimax estimation
- Assess $\hat{\boldsymbol{p}}$ by its worst-case estimation loss:

$$r_{k,n}^{\ell}(\hat{\boldsymbol{p}}) = \sup_{\boldsymbol{p} \in \Delta_k} \mathbb{E}_{X^n \sim \boldsymbol{p}^n} \ell(\hat{\boldsymbol{p}}(X^n), \boldsymbol{p})$$

where

- ℓ is a loss function, e.g., $\ell = \ell_2^2$ (Mean Square Error) or $\ell = \ell_1$
- $\Delta_k := \{\boldsymbol{p} \in (\mathbb{R}_+)^k : \sum_{i=1}^k p_i = 1\}$ is the *probability simplex*

Define

$$r_{k,n}^{\ell} := \inf_{\hat{\boldsymbol{p}}} r_{k,n}^{\ell}(\hat{\boldsymbol{p}})$$

We say that an estimator $\hat{\boldsymbol{p}}$ is order optimal if

$$\lim_{n \to \infty} \frac{r_{k,n}^{\ell}(\hat{\boldsymbol{p}})}{r_{k,n}^{\ell}} = \text{Const}$$

and asymptotically optimal if Const=1

Estimation under Local Differential Privacy

- Estimator: $\hat{\boldsymbol{p}} : \mathcal{Y}^n \to \mathbb{R}^k$

Estimation under Local Differential Privacy

- Estimator: $\hat{\boldsymbol{p}} : \mathcal{Y}^n \to \mathbb{R}^k$
- Given the LDP constraint ϵ, we need to construct both $\boldsymbol{Q}$ and $\hat{\boldsymbol{p}}$

Estimation under Local Differential Privacy

- Estimator: $\hat{\boldsymbol{p}} : \mathcal{Y}^n \to \mathbb{R}^k$
- Given the LDP constraint ϵ, we need to construct both $\boldsymbol{Q}$ and $\hat{\boldsymbol{p}}$

MINIMAX PROBLEM

Define Minimax Risk

$$r^{\ell}_{k,n}(\boldsymbol{Q}) = \inf_{\hat{\boldsymbol{p}}} \sup_{\boldsymbol{p} \in \Delta_k} \mathbb{E}_{Y^n \sim (\boldsymbol{p}\boldsymbol{Q})^n} \ell(\hat{\boldsymbol{p}}(Y^n), \boldsymbol{p})$$

$$r^{\ell}_{\epsilon,k,n} = \inf_{\boldsymbol{Q} \in \mathcal{D}_\epsilon} r^{\ell}_{k,n}(\boldsymbol{Q})$$

Restrictions on $\mathcal{Y}$ and $\boldsymbol{Q}$

- It suffices to consider finite output alphabet $\mathcal{Y}$.

  Namely, let $\mathcal{D}_{\epsilon,F}$ be the set of all $\boldsymbol{Q} : \mathcal{X} \to \mathcal{Y}$ with $|\mathcal{Y}| < \infty$. For $\ell = \ell_2^2$ or ℓ_1,

$$r^{\ell}_{\epsilon,k,n} = \inf_{\boldsymbol{Q} \in \mathcal{D}_{\epsilon,F}} r^{\ell}_{k,n}(\boldsymbol{Q})$$

Restrictions on $\mathcal{Y}$ and $\boldsymbol{Q}$

- It suffices to consider finite output alphabet $\mathcal{Y}$.

  Namely, let $\mathcal{D}_{\epsilon,F}$ be the set of all $\boldsymbol{Q} : \mathcal{X} \to \mathcal{Y}$ with $|\mathcal{Y}| < \infty$. For $\ell = \ell_2^2$ or ℓ_1,

$$r^{\ell}_{\epsilon,k,n} = \inf_{\boldsymbol{Q} \in \mathcal{D}_{\epsilon,F}} r^{\ell}_{k,n}(\boldsymbol{Q})$$

Theorem

Let

$$\mathcal{D}_{\epsilon,E} = \left\{ \boldsymbol{Q} \in \mathcal{D}_{\epsilon,F} : \frac{\boldsymbol{Q}(y|x)}{\min_{x' \in \mathcal{X}} \boldsymbol{Q}(y|x')} \in \{1, e^{\epsilon}\} \text{ for all } x \in \mathcal{X} \text{ and all } y \in \mathcal{Y} \right\}.$$

For $\ell = \ell_2^2$ *and* ℓ_1,

$$r^{\ell}_{\epsilon,k,n} = \inf_{\boldsymbol{Q} \in \mathcal{D}_{\epsilon,E}} r^{\ell}_{k,n}(\boldsymbol{Q}).$$

Known results: Randomized Response $\boldsymbol{Q}_{\text{RR}}$ (Warner, '65)

- Output alphabet $\mathcal{Y}_{\text{RR}} = \mathcal{X}$

$$\boldsymbol{Q}_{\text{RR}}(y|x) = \begin{cases} \dfrac{e^{\epsilon}}{e^{\epsilon}+k-1}, & \text{if } y = x \\ \dfrac{1}{e^{\epsilon}+k-1}, & \text{if } y \neq x \end{cases}$$

Known results: Randomized Response $\boldsymbol{Q}_{\text{RR}}$ (Warner, '65)

- Output alphabet $\mathcal{Y}_{\text{RR}} = \mathcal{X}$

$$\boldsymbol{Q}_{\text{RR}}(y|x) = \begin{cases} \frac{e^{\epsilon}}{e^{\epsilon}+k-1}, & \text{if } y = x \\ \frac{1}{e^{\epsilon}+k-1}, & \text{if } y \neq x \end{cases}$$

- Matrix form:

$$\boldsymbol{Q}_{\text{RR}} = \frac{1}{e^{\epsilon}+k-1} \begin{bmatrix} e^{\epsilon} & 1 & \dots & 1 \\ 1 & e^{\epsilon} & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \dots & e^{\epsilon} \end{bmatrix}$$

Known results: Randomized Response $\boldsymbol{Q}_{\mathrm{RR}}$ (Warner, '65)

- Output alphabet $\mathcal{Y}_{\mathrm{RR}} = \mathcal{X}$

$$\boldsymbol{Q}_{\mathrm{RR}}(y|x) = \begin{cases} \frac{e^{\epsilon}}{e^{\epsilon}+k-1}, & \text{if } y = x \\ \frac{1}{e^{\epsilon}+k-1}, & \text{if } y \neq x \end{cases}$$

- Matrix form:

$$\boldsymbol{Q}_{\mathrm{RR}} = \frac{1}{e^{\epsilon}+k-1} \begin{bmatrix} e^{\epsilon} & 1 & \dots & 1 \\ 1 & e^{\epsilon} & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \dots & e^{\epsilon} \end{bmatrix}$$

- Clearly, $\boldsymbol{Q}_{\mathrm{RR}}$ is ϵ-LDP

Known results: Randomized Response $\boldsymbol{Q}_{\text{RR}}$ (Warner, '65)

- Output alphabet $\mathcal{Y}_{\text{RR}} = \mathcal{X}$

$$\boldsymbol{Q}_{\text{RR}}(y|x) = \begin{cases} \dfrac{e^{\epsilon}}{e^{\epsilon}+k-1}, & \text{if } y = x \\ \dfrac{1}{e^{\epsilon}+k-1}, & \text{if } y \neq x \end{cases}$$

- Matrix form:

$$\boldsymbol{Q}_{\text{RR}} = \frac{1}{e^{\epsilon}+k-1} \begin{bmatrix} e^{\epsilon} & 1 & \dots & 1 \\ 1 & e^{\epsilon} & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \dots & e^{\epsilon} \end{bmatrix}$$

- Clearly, $\boldsymbol{Q}_{\text{RR}}$ is ϵ-LDP

- When $\epsilon \to \infty$, $\boldsymbol{Q}_{\text{RR}} \to I_k$. $\boldsymbol{Q}_{\text{RR}}$ and its empirical estimator are order-optimal in the low privacy regime

Empirical estimator for $\boldsymbol{Q}_{\text{RR}}$

- For $i = 1, 2, \ldots, k$, let $m_i = P(Y = i)$.

Empirical estimator for $\boldsymbol{Q}_{\mathrm{RR}}$

- For $i = 1, 2, \ldots, k$, let $m_i = P(Y = i)$.

$$
\begin{aligned}
m_i = \sum_{j=1}^{k} p_j \boldsymbol{Q}_{\mathrm{RR}}(i|j) &= \frac{1}{e^\epsilon + k - 1}\Big(\sum_{j \neq i} p_j\Big) + \frac{e^\epsilon}{e^\epsilon + k - 1} p_i \\
&= \frac{1}{e^\epsilon + k - 1}(1 - p_i) + \frac{e^\epsilon}{e^\epsilon + k - 1} p_i \\
&= \frac{e^\epsilon - 1}{e^\epsilon + k - 1} p_i + \frac{1}{e^\epsilon + k - 1}
\end{aligned}
$$

Empirical estimator for $\boldsymbol{Q}_{\mathsf{RR}}$

- For $i = 1, 2, \ldots, k$, let $m_i = P(Y = i)$.

$$
\begin{aligned}
m_i = \sum_{j=1}^{k} p_j \boldsymbol{Q}_{\mathsf{RR}}(i|j) &= \frac{1}{e^\epsilon + k - 1}\Big(\sum_{j \neq i} p_j\Big) + \frac{e^\epsilon}{e^\epsilon + k - 1} p_i \\
&= \frac{1}{e^\epsilon + k - 1}(1 - p_i) + \frac{e^\epsilon}{e^\epsilon + k - 1} p_i \\
&= \frac{e^\epsilon - 1}{e^\epsilon + k - 1} p_i + \frac{1}{e^\epsilon + k - 1}
\end{aligned}
$$

- Therefore,

$$
p_i = \frac{e^\epsilon + k - 1}{e^\epsilon - 1} m_i - \frac{1}{e^\epsilon - 1}
$$

k-RAPPOR

- $\mathcal{Y} = \{0, 1\}^k$

k-RAPPOR

- $\mathcal{Y} = \{0,1\}^k$
- Privatization scheme:

$$i \mapsto \boldsymbol{y}_i$$

where $\boldsymbol{y}_i$ is obtained from $\boldsymbol{e}_i$ by randomly flipping each coordinate with probability $1/(1+e^{\epsilon/2})$

k-RAPPOR

- $\mathcal{Y} = \{0,1\}^k$
- Privatization scheme:
$$i \mapsto \boldsymbol{y}_i$$
where $\boldsymbol{y}_i$ is obtained from $\boldsymbol{e}_i$ by randomly flipping each coordinate with probability $1/(1+e^{\epsilon/2})$
- Empirical estimator
$$\hat{p}_i = \left(\frac{e^{\epsilon/2}+1}{e^{\epsilon/2}-1}\right)\frac{T_i}{n} - \frac{1}{e^{\epsilon/2}-1}$$
where T_i is the Hamming weight of Y_i.

ERLINGSSON ET AL., '14;
DUCHI, JORDAN, AND WAINWRIGHT, '13;
KAIROUZ ET AL., *Journ. Machine Learning Research* '16

Known results about $r_{\epsilon,k,n}^{\ell_2^2}$

Low privacy regime (large ϵ)

$$r_{k,n}^{\ell_2^2} \leq r_{\epsilon,k,n}^{\ell_2^2} \leq r_{k,n}^{\ell_2^2}(\boldsymbol{Q}_{\mathrm{RR}}, \hat{\boldsymbol{p}})$$

Theorem (KAIROUZ, BONAWITZ, AND RAMAGE, '16)

$$\frac{1-\frac{1}{k}}{(\sqrt{n}+1)^2} \leq r_{\epsilon,k,n}^{\ell_2^2} \leq \left(\frac{e^\epsilon + k - 1}{e^\epsilon - 1}\right)^2 \frac{1-\frac{1}{k}}{n}$$

Known results about $r^{\ell_2^2}_{\epsilon,k,n}$

Low privacy regime (large ϵ)

$$r^{\ell_2^2}_{k,n} \le r^{\ell_2^2}_{\epsilon,k,n} \le r^{\ell_2^2}_{k,n}(\boldsymbol{Q}_{\mathrm{RR}}, \hat{\boldsymbol{p}})$$

Theorem (Kairouz, Bonawitz, and Ramage, '16)

$$\frac{1-\frac{1}{k}}{(\sqrt{n}+1)^2} \le r^{\ell_2^2}_{\epsilon,k,n} \le \left(\frac{e^\epsilon + k - 1}{e^\epsilon - 1}\right)^2 \frac{1-\frac{1}{k}}{n}$$

High privacy regime ($\epsilon \approx 0$)

Theorem (Duchi, Jordan, and Wainwright, 16')

For all $\epsilon \le 1$,

$$C_l \frac{k}{n\epsilon^2} \le r^{\ell_2^2}_{\epsilon,k,n} \le C_u \frac{k}{n\epsilon^2},$$

where $0 < C_l < C_u < 5$ are constants independent of ϵ, k and n.

Our results and recent developments

- We propose a family of new privatization schemes and corresponding estimators that are
  - *order-optimal* for medium to high-privacy regime for ℓ_1 loss[1]
  - *asymptotically optimal* in all regimes for ℓ_2 loss[2]
- We prove lower bounds[1] on $r^{\ell}_{\epsilon,k,n}$ for $\ell = \ell_1$ and ℓ_2 in the medium privacy regime (previously such bounds were known only for low- and high-privacy regimes (KAIROUZ ET AL., '16, DUCHI ET AL., '16)
- We prove a tight lower bound[2] on $r^{\ell_2}_{\epsilon,k,n}$
- Recently ACHARYA-SUN-ZHANG (2018) proposed a *Hadamard response mechanism* which is order-optimal in all regimes

[1] *IEEE Trans. IT*, no. 8, 2018; arXiv:1702.00059
[2] arXiv 1708.00610

$\boldsymbol{Q}_{k,\epsilon,d}$

- Given ϵ and k, we define a family of schemes $\boldsymbol{Q}_{k,\epsilon,d}$ parameterized by $d \in \{1, 2, \ldots, k-1\}$.

$\boldsymbol{Q}_{k,\epsilon,d}$

- Given ϵ and k, we define a family of schemes $\boldsymbol{Q}_{k,\epsilon,d}$ parameterized by $d \in \{1, 2, \ldots, k-1\}$.

- Output alphabet $\mathcal{Y}_{k,d} = \{y \in \{0,1\}^k : \sum_{i=1}^k y_i = d\}$; $|\mathcal{Y}_{k,d}| = \binom{k}{d}$.

$\boldsymbol{Q}_{k,\epsilon,d}$

- Given ϵ and k, we define a family of schemes $\boldsymbol{Q}_{k,\epsilon,d}$ parameterized by $d \in \{1, 2, \ldots, k-1\}$.
- Output alphabet $\mathcal{Y}_{k,d} = \{y \in \{0,1\}^k : \sum_{i=1}^k y_i = d\}$; $|\mathcal{Y}_{k,d}| = \binom{k}{d}$.
- Define the Subset Selection Mechanism

$$\boldsymbol{Q}_{k,\epsilon,d}(y|i) = \begin{cases} \dfrac{e^\epsilon}{\binom{k-1}{d-1}e^\epsilon + \binom{k-1}{d}} & \text{if } y_i = 1 \\[2ex] \dfrac{1}{\binom{k-1}{d-1}e^\epsilon + \binom{k-1}{d}} & \text{if } y_i = 0 \end{cases}$$

$\boldsymbol{Q}_{k,\epsilon,d}$

- Given ϵ and k, we define a family of schemes $\boldsymbol{Q}_{k,\epsilon,d}$ parameterized by $d \in \{1, 2, \ldots, k-1\}$.

- Output alphabet $\mathcal{Y}_{k,d} = \{y \in \{0,1\}^k : \sum_{i=1}^k y_i = d\}$; $|\mathcal{Y}_{k,d}| = \binom{k}{d}$.

- Define the Subset Selection Mechanism

$$\boldsymbol{Q}_{k,\epsilon,d}(y|i) = \begin{cases} \dfrac{e^\epsilon}{\binom{k-1}{d-1}e^\epsilon + \binom{k-1}{d}} & \text{if } y_i = 1 \\ \dfrac{1}{\binom{k-1}{d-1}e^\epsilon + \binom{k-1}{d}} & \text{if } y_i = 0 \end{cases}$$

- $\boldsymbol{Q}_{k,\epsilon,d}$ is ϵ-LDP

$\boldsymbol{Q}_{k,\epsilon,d}$

- Given ϵ and k, we define a family of schemes $\boldsymbol{Q}_{k,\epsilon,d}$ parameterized by $d \in \{1, 2, \ldots, k-1\}$.
- Output alphabet $\mathcal{Y}_{k,d} = \{y \in \{0,1\}^k : \sum_{i=1}^{k} y_i = d\}$; $|\mathcal{Y}_{k,d}| = \binom{k}{d}$.
- Define the Subset Selection Mechanism

$$\boldsymbol{Q}_{k,\epsilon,d}(y|i) = \begin{cases} \dfrac{e^\epsilon}{\binom{k-1}{d-1}e^\epsilon + \binom{k-1}{d}} & \text{if } y_i = 1 \\ \dfrac{1}{\binom{k-1}{d-1}e^\epsilon + \binom{k-1}{d}} & \text{if } y_i = 0 \end{cases}$$

- $\boldsymbol{Q}_{k,\epsilon,d}$ is ϵ-LDP
- When $d = 1$, $\boldsymbol{Q}_{k,\epsilon,1}$ is the same as $\boldsymbol{Q}_{\text{RR}}$

$\boldsymbol{Q}_{k,\epsilon,d}$

- Given ϵ and k, we define a family of schemes $\boldsymbol{Q}_{k,\epsilon,d}$ parameterized by $d \in \{1, 2, \ldots, k-1\}$.
- Output alphabet $\mathcal{Y}_{k,d} = \{y \in \{0,1\}^k : \sum_{i=1}^{k} y_i = d\}$; $|\mathcal{Y}_{k,d}| = \binom{k}{d}$.
- Define the Subset Selection Mechanism

$$\boldsymbol{Q}_{k,\epsilon,d}(y|i) = \begin{cases} \dfrac{e^{\epsilon}}{\binom{k-1}{d-1}e^{\epsilon} + \binom{k-1}{d}} & \text{if } y_i = 1 \\[2ex] \dfrac{1}{\binom{k-1}{d-1}e^{\epsilon} + \binom{k-1}{d}} & \text{if } y_i = 0 \end{cases}$$

- $\boldsymbol{Q}_{k,\epsilon,d}$ is ϵ-LDP
- When $d = 1$, $\boldsymbol{Q}_{k,\epsilon,1}$ is the same as $\boldsymbol{Q}_{\text{RR}}$
- SS mechanism was also proposed in
  S. Wang, L. Huang, P. Wang, Y. Nie, H. Xu, W. Yang, X. Li, and C. Qiao
  "Mutual information optimally local private discrete distribution estimation"
  arXiv:1607.08025

An example: $\boldsymbol{Q}_{4,\epsilon,2}$ $(k = 4, d = 2)$

- Output alphabet $\mathcal{Y}_{4,2}$: all the binary vectors of length 4 and Hamming weight 2.

An example: $\boldsymbol{Q}_{4,\epsilon,2}$ $(k = 4, d = 2)$

- Output alphabet $\mathcal{Y}_{4,2}$: all the binary vectors of length 4 and Hamming weight 2.

	$(1,1,0,0)$	$(1,0,1,0)$	$(1,0,0,1)$	$(0,1,1,0)$	$(0,1,0,1)$	$(0,0,1,1)$
1	e^ϵ	e^ϵ	e^ϵ	1	1	1
2	e^ϵ	1	1	e^ϵ	e^ϵ	1
3	1	e^ϵ	1	e^ϵ	1	e^ϵ
4	1	1	e^ϵ	1	e^ϵ	e^ϵ

An example: $\boldsymbol{Q}_{4,\epsilon,2}$ $(k = 4, d = 2)$

- Output alphabet $\mathcal{Y}_{4,2}$: all the binary vectors of length 4 and Hamming weight 2.

	$(1,1,0,0)$	$(1,0,1,0)$	$(1,0,0,1)$	$(0,1,1,0)$	$(0,1,0,1)$	$(0,0,1,1)$
1	e^ϵ	e^ϵ	e^ϵ	1	1	1
2	e^ϵ	1	1	e^ϵ	e^ϵ	1
3	1	e^ϵ	1	e^ϵ	1	e^ϵ
4	1	1	e^ϵ	1	e^ϵ	e^ϵ

- Normalize by $\dfrac{1}{3e^\epsilon + 3}$

Empirical estimator for $\boldsymbol{Q}_{4,\epsilon,2}$

	$(1,1,0,0)$	$(1,0,1,0)$	$(1,0,0,1)$	$(0,1,1,0)$	$(0,1,0,1)$	$(0,0,1,1)$
1	e^ϵ	e^ϵ	e^ϵ	1	1	1
2	e^ϵ	1	1	e^ϵ	e^ϵ	1
3	1	e^ϵ	1	e^ϵ	1	e^ϵ
4	1	1	e^ϵ	1	e^ϵ	e^ϵ

Empirical estimator for $\boldsymbol{Q}_{4,\epsilon,2}$

	$(1,1,0,0)$	$(1,0,1,0)$	$(1,0,0,1)$	$(0,1,1,0)$	$(0,1,0,1)$	$(0,0,1,1)$
1	e^ϵ	e^ϵ	e^ϵ	1	1	1
2	e^ϵ	1	1	e^ϵ	e^ϵ	1
3	1	e^ϵ	1	e^ϵ	1	e^ϵ
4	1	1	e^ϵ	1	e^ϵ	e^ϵ

$$\begin{aligned} P(Y_1 = 1) &= P(Y \in \{(1,1,0,0),(1,0,1,0),(1,0,0,1)\}) \\ &= \frac{3e^\epsilon p_1 + (e^\epsilon + 2)(p_2 + p_3 + p_4)}{3e^\epsilon + 3} \\ &= \frac{3e^\epsilon p_1 + (e^\epsilon + 2)(1 - p_1)}{3e^\epsilon + 3} \\ &= \frac{(2e^\epsilon - 2)p_1 + e^\epsilon + 2}{3e^\epsilon + 3} \end{aligned}$$

Empirical estimator for $\boldsymbol{Q}_{4,\epsilon,2}$

	$(1,1,0,0)$	$(1,0,1,0)$	$(1,0,0,1)$	$(0,1,1,0)$	$(0,1,0,1)$	$(0,0,1,1)$
1	e^ϵ	e^ϵ	e^ϵ	1	1	1
2	e^ϵ	1	1	e^ϵ	e^ϵ	1
3	1	e^ϵ	1	e^ϵ	1	e^ϵ
4	1	1	e^ϵ	1	e^ϵ	e^ϵ

$$\begin{aligned} P(Y_1 = 1) &= P(Y \in \{(1,1,0,0),(1,0,1,0),(1,0,0,1)\}) \\ &= \frac{3e^\epsilon p_1 + (e^\epsilon + 2)(p_2 + p_3 + p_4)}{3e^\epsilon + 3} \\ &= \frac{3e^\epsilon p_1 + (e^\epsilon + 2)(1 - p_1)}{3e^\epsilon + 3} \\ &= \frac{(2e^\epsilon - 2)p_1 + e^\epsilon + 2}{3e^\epsilon + 3} \end{aligned}$$

$$p_1 = \frac{3e^\epsilon + 3}{2e^\epsilon - 2} P(Y_1 = 1) - \frac{e^\epsilon + 2}{2e^\epsilon - 2}$$

Empirical estimator for $\boldsymbol{Q}_{4,\epsilon,2}$

$$p_1 = \frac{3e^\epsilon + 3}{2e^\epsilon - 2} P(Y_1 = 1) - \frac{e^\epsilon + 2}{2e^\epsilon - 2}$$

Empirical estimator for $\boldsymbol{Q}_{4,\epsilon,2}$

$$p_1 = \frac{3e^\epsilon + 3}{2e^\epsilon - 2} P(Y_1 = 1) - \frac{e^\epsilon + 2}{2e^\epsilon - 2}$$

- Let T_1 be the number of privatized samples whose first coordinate is 1:

$$T_1 := \sum_{j=1}^{n} Y_1^{(j)}$$

Empirical estimator for $\boldsymbol{Q}_{4,\epsilon,2}$

$$p_1 = \frac{3e^\epsilon + 3}{2e^\epsilon - 2} P(Y_1 = 1) - \frac{e^\epsilon + 2}{2e^\epsilon - 2}$$

- Let T_1 be the number of privatized samples whose first coordinate is 1:

$$T_1 := \sum_{j=1}^{n} Y_1^{(j)}$$

- T_1/n is the empirical estimator of $P(Y_1 = 1)$. Therefore

$$\hat{p}_1 = \frac{3e^\epsilon + 3}{2e^\epsilon - 2} \frac{T_1}{n} - \frac{e^\epsilon + 2}{2e^\epsilon - 2}$$

Empirical estimator for $\boldsymbol{Q}_{4,\epsilon,2}$

$$p_1 = \frac{3e^\epsilon + 3}{2e^\epsilon - 2} P(Y_1 = 1) - \frac{e^\epsilon + 2}{2e^\epsilon - 2}$$

- Let T_1 be the number of privatized samples whose first coordinate is 1:

$$T_1 := \sum_{j=1}^{n} Y_1^{(j)}$$

- T_1/n is the empirical estimator of $P(Y_1 = 1)$. Therefore

$$\hat{p}_1 = \frac{3e^\epsilon + 3}{2e^\epsilon - 2} \frac{T_1}{n} - \frac{e^\epsilon + 2}{2e^\epsilon - 2}$$

- Similarly, for $i = 2, 3, 4$,

$$T_i \triangleq \sum_{j=1}^{n} Y_i^{(j)}, \quad \hat{p}_i = \frac{3e^\epsilon + 3}{2e^\epsilon - 2} \frac{T_i}{n} - \frac{e^\epsilon + 2}{2e^\epsilon - 2}$$

Performance of $\boldsymbol{Q}_{k,\epsilon,d}$ and its empirical estimator

- Empirical estimator for $\boldsymbol{Q}_{k,\epsilon,d}$:

$$\hat{p}_i = \left(\frac{(k-1)e^{\epsilon} + \frac{(k-1)(k-d)}{d}}{(k-d)(e^{\epsilon}-1)} \right) \frac{T_i}{n} - \frac{(d-1)e^{\epsilon} + k - d}{(k-d)(e^{\epsilon}-1)}$$

Performance of $\boldsymbol{Q}_{k,\epsilon,d}$ and its empirical estimator

- Empirical estimator for $\boldsymbol{Q}_{k,\epsilon,d}$:

$$\hat{p}_i = \Big(\frac{(k-1)e^{\epsilon} + \frac{(k-1)(k-d)}{d}}{(k-d)(e^{\epsilon}-1)}\Big)\frac{T_i}{n} - \frac{(d-1)e^{\epsilon}+k-d}{(k-d)(e^{\epsilon}-1)}$$

- $r_{k,n}^{\ell_1}(\boldsymbol{Q}_{k,\epsilon,d}, \hat{\boldsymbol{p}})$ is estimated by showing that worst-case performance is achieved for $\boldsymbol{p} \sim \text{Unif}(\mathcal{X})$:

$$r_{k,n}^{\ell_2^2}(\boldsymbol{Q}_{k,\epsilon,d}, \hat{\boldsymbol{p}}) = \frac{(k-1)^2}{nk(e^{\epsilon}-1)^2}\frac{(de^{\epsilon}+k-d)^2}{d(k-d)}.$$

Performance of $\boldsymbol{Q}_{k,\epsilon,d}$ and its empirical estimator

- Empirical estimator for $\boldsymbol{Q}_{k,\epsilon,d}$:

$$\hat{p}_i = \Big(\frac{(k-1)e^\epsilon + \frac{(k-1)(k-d)}{d}}{(k-d)(e^\epsilon - 1)}\Big)\frac{T_i}{n} - \frac{(d-1)e^\epsilon + k - d}{(k-d)(e^\epsilon - 1)}$$

- $r^{\ell_1}_{k,n}(\boldsymbol{Q}_{k,\epsilon,d}, \hat{\boldsymbol{p}})$ is estimated by showing that worst-case performance is achieved for $\boldsymbol{p} \sim \text{Unif}(\mathcal{X})$:

$$r^{\ell_2^2}_{k,n}(\boldsymbol{Q}_{k,\epsilon,d}, \hat{\boldsymbol{p}}) = \frac{(k-1)^2}{nk(e^\epsilon - 1)^2}\frac{(de^\epsilon + k - d)^2}{d(k-d)}.$$

- Optimal choice of d:

$$d^* = \underset{1\le d\le k-1}{\text{argmin}} \frac{(de^\epsilon + k - d)^2}{d(k-d)} = \underset{1\le d\le k-1}{\text{argmin}} \Big(\frac{d}{k-d}e^{2\epsilon} + \frac{k-d}{d}\Big)$$

$$d^* = \lceil k/(e^\epsilon + 1)\rceil \text{ or } \lfloor k/(e^\epsilon + 1)\rfloor$$

Performance of $\boldsymbol{Q}_{k,\epsilon,d}$ and its empirical estimator

- Empirical estimator for $\boldsymbol{Q}_{k,\epsilon,d}$:

$$\hat{p}_i = \left(\frac{(k-1)e^\epsilon + \frac{(k-1)(k-d)}{d}}{(k-d)(e^\epsilon - 1)}\right)\frac{T_i}{n} - \frac{(d-1)e^\epsilon + k - d}{(k-d)(e^\epsilon - 1)}$$

- $r^{\ell_1}_{k,n}(\boldsymbol{Q}_{k,\epsilon,d}, \hat{\boldsymbol{p}})$ is estimated by showing that worst-case performance is achieved for $\boldsymbol{p} \sim \text{Unif}(\mathcal{X})$:

$$r^{\ell_2^2}_{k,n}(\boldsymbol{Q}_{k,\epsilon,d}, \hat{\boldsymbol{p}}) = \frac{(k-1)^2}{nk(e^\epsilon - 1)^2}\frac{(de^\epsilon + k - d)^2}{d(k-d)}.$$

- Optimal choice of d:

$$d^* = \operatorname*{argmin}_{1 \le d \le k-1} \frac{(de^\epsilon + k - d)^2}{d(k-d)} = \operatorname*{argmin}_{1 \le d \le k-1} \left(\frac{d}{k-d}e^{2\epsilon} + \frac{k-d}{d}\right)$$

$$d^* = \lceil k/(e^\epsilon + 1)\rceil \text{ or } \lfloor k/(e^\epsilon + 1)\rfloor$$

- If $k/(e^\epsilon + 1) \le 1$ then $d^* = 1$, and our scheme turns into k-RR

Minimax risk for ℓ_1 loss

Theorem

For $e^\epsilon \ll k$ and n large enough,

$$r_{\epsilon,k,n}^{\ell_1} = \Theta\left(\frac{k\sqrt{e^\epsilon}}{(e^\epsilon - 1)\sqrt{n}}\right)$$

Minimax risk for ℓ_1 loss

Theorem

For $e^\epsilon \ll k$ and n large enough,

$$r^{\ell_1}_{\epsilon,k,n} = \Theta\left(\frac{k\sqrt{e^\epsilon}}{(e^\epsilon - 1)\sqrt{n}}\right)$$

- Upper bound by a calculation from the expected loss $\mathbb{E}_{Y^n \sim (\boldsymbol{Q}\,\mathbf{p})^n}\, \ell_1(\boldsymbol{p}(\hat{Y}^n), \boldsymbol{p})$

Minimax risk for ℓ_1 loss

Theorem

For $e^\epsilon \ll k$ and n large enough,

$$r^{\ell_1}_{\epsilon,k,n} = \Theta\left(\frac{k\sqrt{e^\epsilon}}{(e^\epsilon - 1)\sqrt{n}}\right)$$

- Upper bound by a calculation from the expected loss $\mathbb{E}_{Y^n \sim (\boldsymbol{Q}\,\mathbf{p})^n}\, \ell_1(\boldsymbol{p}(\hat{Y}^n), \boldsymbol{p})$
- Lower bound by an application of Assouad's method: Reduction to hypothesis testing for distributions of the form

$$\boldsymbol{p}_\nu := \boldsymbol{p}_U + \frac{\delta}{k}\begin{bmatrix} \nu \\ -\nu \end{bmatrix} \in \Delta_k.$$

where $\nu = (\nu_1, \ldots, \nu_{k/2}) \in \{-1, 1\}^{k/2}$, $\delta \in (0, 1)$.

Minimax risk for ℓ_2^2 loss

- For $k \geq \max(4, e^\epsilon - 1)$, $\boldsymbol{Q} = \boldsymbol{Q}_{k,\epsilon,d}$ we have

$$\mathbb{E}_{Y^n \sim (\boldsymbol{Q}\,\mathbf{p})^n}\, \ell_2^2(\boldsymbol{p}(\hat{Y}^n), \boldsymbol{p}) < \frac{4ke^\epsilon}{n(e^\epsilon - 1)^2}\left(1 + \frac{2e^\epsilon + 3}{4k}\right)$$

- At the same time, for $e^\epsilon \geq 3$ we have

$$r_{\epsilon,k,n}^{\ell_2^2} \geq \frac{(k-1)}{64n(e^\epsilon - 1)}$$

- Overall, in the medium privacy regime,

$$r_{\epsilon,k,n}^{\ell_2^2} = \Theta\left(\frac{k}{ne^\epsilon}\right)$$

Better lower bound for mean square loss

Theorem

For every k and ϵ,

$$r_{\epsilon,k,n}^{\ell_2^2} = \frac{(k-1)^2}{nk(e^\epsilon - 1)^2} \frac{(d^* e^\epsilon + k - d^*)^2}{d^*(k-d^*)} - O(n^{-14/13}),$$

where

$$d^* = \lceil k/(e^\epsilon + 1) \rceil \text{ or } \lfloor k/(e^\epsilon + 1) \rfloor$$

Lower bound on $r^{\ell_2^2}_{\epsilon,k,n}$

- Objective:

$$r^{\ell_2^2}_{\epsilon,k,n} \geq r^{\ell_2^2}_{k,n}(\boldsymbol{Q}_{k,\epsilon,d^*}, \hat{\boldsymbol{p}}) - O\left(n^{-14/13}\right).$$

Lower bound on $r_{\epsilon,k,n}^{\ell_2^2}$

- Objective:

$$r_{\epsilon,k,n}^{\ell_2^2} \geq r_{k,n}^{\ell_2^2}(\boldsymbol{Q}_{k,\epsilon,d^*}, \hat{\boldsymbol{p}}) - O\left(n^{-14/13}\right).$$

- Equivalently, prove the following: For any given ϵ-LDP mechanism $\boldsymbol{Q}$

$$r_{k,n}^{\ell_2^2}(\boldsymbol{Q}) \geq r_{k,n}^{\ell_2^2}(\boldsymbol{Q}_{k,\epsilon,d^*}, \hat{\boldsymbol{p}}) - O\left(n^{-14/13}\right).$$

Lower bound on $r_{\epsilon,k,n}^{\ell_2^2}$

- Bound $r_{\epsilon,k,n}^{\ell_2^2}$ below by a Bayes estimation loss

Lower bound on $r^{\ell_2^2}_{\epsilon,k,n}$

- Bound $r^{\ell_2^2}_{\epsilon,k,n}$ below by a Bayes estimation loss
- Assume that $\boldsymbol{p}$ is drawn uniformly from $\mathcal{P}$, a small neighborhood of $\boldsymbol{p}_U = (1/k, 1/k, \dots, 1/k)$

Lower bound on $r_{\epsilon,k,n}^{\ell_2^2}$

- Bound $r_{\epsilon,k,n}^{\ell_2^2}$ below by a Bayes estimation loss
- Assume that $\boldsymbol{p}$ is drawn uniformly from $\mathcal{P}$, a small neighborhood of $\boldsymbol{p}_U = (1/k, 1/k, \dots, 1/k)$
- Let $\boldsymbol{P}$ be the random variable (vector) corresponding to $\boldsymbol{p}$

Lower bound on $r^{\ell_2^2}_{\epsilon,k,n}$

- Bound $r^{\ell_2^2}_{\epsilon,k,n}$ below by a Bayes estimation loss
- Assume that $\boldsymbol{p}$ is drawn uniformly from $\mathcal{P}$, a small neighborhood of $\boldsymbol{p}_U = (1/k, 1/k, \ldots, 1/k)$
- Let $\boldsymbol{P}$ be the random variable (vector) corresponding to $\boldsymbol{p}$
- Local Asymptotic Normality of the posterior distribution
  (Hajek '72, Le Cam '70; Ibragimov and Has'minskii '81)

Lower bound on $r^{\ell_2^2}_{\epsilon,k,n}$

- Bound $r^{\ell_2^2}_{\epsilon,k,n}$ below by a Bayes estimation loss
- Assume that $\boldsymbol{p}$ is drawn uniformly from $\mathcal{P}$, a small neighborhood of $\boldsymbol{p}_U = (1/k, 1/k, \ldots, 1/k)$
- Let $\boldsymbol{P}$ be the random variable (vector) corresponding to $\boldsymbol{p}$
- Local Asymptotic Normality of the posterior distribution (HAJEK '72, LE CAM '70; IBRAGIMOV AND HAS'MINSKII '81)
- When the radius of $\mathcal{P}$ is of order $n^{-1/2}$, the posterior distribution of $\boldsymbol{P}$ given Y^n is very close to a jointly Gaussian distribution

Lower bound on $r^{\ell_2^2}_{\epsilon,k,n}$

- Bound $r^{\ell_2^2}_{\epsilon,k,n}$ below by a Bayes estimation loss
- Assume that $\boldsymbol{p}$ is drawn uniformly from $\mathcal{P}$, a small neighborhood of $\boldsymbol{p}_U = (1/k, 1/k, \ldots, 1/k)$
- Let $\boldsymbol{P}$ be the random variable (vector) corresponding to $\boldsymbol{p}$
- Local Asymptotic Normality of the posterior distribution (HAJEK '72, LE CAM '70; IBRAGIMOV AND HAS'MINSKII '81)
- When the radius of $\mathcal{P}$ is of order $n^{-1/2}$, the posterior distribution of $\boldsymbol{P}$ given Y^n is very close to a jointly Gaussian distribution
- The covariance matrix $\Sigma(n, \boldsymbol{Q})$ of this Gaussian distribution is independent of the value of Y^n.

Lower bound on $r_{\epsilon,k,n}^{\ell_2^2}$

- Trace of $\Sigma(n, \boldsymbol{Q})$: Sum of variances of $P_1, P_2, \ldots, P_k$ given Y^n

Lower bound on $r^{\ell_2^2}_{\epsilon,k,n}$

- Trace of $\Sigma(n, \boldsymbol{Q})$: Sum of variances of $P_1, P_2, \ldots, P_k$ given Y^n
- $\mathrm{tr}\,(\Sigma(n, \boldsymbol{Q}))$ is an asymptotic lower bound of $r^{\ell_2^2}_{k,n}(\boldsymbol{Q})$:

$$r^{\ell_2^2}_{k,n}(\boldsymbol{Q}) \geq \mathrm{tr}\,(\Sigma(n, \boldsymbol{Q})) - o\left(\frac{1}{n}\right)$$

Lower bound on $r^{\ell_2^2}_{\epsilon,k,n}$

- Trace of $\Sigma(n, \boldsymbol{Q})$: Sum of variances of $P_1, P_2, \ldots, P_k$ given Y^n
- $\operatorname{tr}(\Sigma(n, \boldsymbol{Q}))$ is an asymptotic lower bound of $r^{\ell_2^2}_{k,n}(\boldsymbol{Q})$:

$$r^{\ell_2^2}_{k,n}(\boldsymbol{Q}) \geq \operatorname{tr}(\Sigma(n, \boldsymbol{Q})) - o\left(\frac{1}{n}\right)$$

- If $\boldsymbol{Q}$ is ϵ-LDP, then

$$\operatorname{tr}(\Sigma(n, \boldsymbol{Q})) \geq r^{\ell_2^2}_{k,n}(\boldsymbol{Q}_{k,\epsilon,d^*}, \hat{\boldsymbol{p}}),$$

Lower bound on $r^{\ell_2^2}_{\epsilon,k,n}$

- Trace of $\Sigma(n, \boldsymbol{Q})$: Sum of variances of $P_1, P_2, \ldots, P_k$ given Y^n
- $\text{tr}\,(\Sigma(n, \boldsymbol{Q}))$ is an asymptotic lower bound of $r^{\ell_2^2}_{k,n}(\boldsymbol{Q})$:

$$r^{\ell_2^2}_{k,n}(\boldsymbol{Q}) \geq \text{tr}\,(\Sigma(n, \boldsymbol{Q})) - o\left(\frac{1}{n}\right)$$

- If $\boldsymbol{Q}$ is ϵ-LDP, then

$$\text{tr}\,(\Sigma(n, \boldsymbol{Q})) \geq r^{\ell_2^2}_{k,n}(\boldsymbol{Q}_{k,\epsilon,d^*}, \hat{\boldsymbol{p}}),$$

- Therefore,

$$r^{\ell_2^2}_{\epsilon,k,n} \geq r^{\ell_2^2}_{k,n}(\boldsymbol{Q}_{k,\epsilon,d^*}, \hat{\boldsymbol{p}}) - o\left(\frac{1}{n}\right)$$

Summary of known results for ℓ_1 loss

- Minimax risk for $\ell = \ell_1$

ϵ	RR	RAPPOR	SS	HR
$(0,1)$	$\frac{k^{3/2}}{\epsilon\sqrt{n}}$	$\frac{k}{\epsilon\sqrt{n}}$	$\frac{k}{\epsilon\sqrt{n}}$	$\frac{k}{\epsilon\sqrt{n}}$
$(1, \log k)$	$\frac{k^{3/2}}{e^\epsilon\sqrt{n}}$	$\frac{k}{\sqrt{e^{\epsilon/2}n}}$	$\frac{k}{\sqrt{e^{\epsilon}n}}$	$\frac{k}{\sqrt{e^{\epsilon}n}}$
$(\log k, 2\log k)$	$\sqrt{\frac{k}{n}}$	$\frac{k}{\sqrt{e^{\epsilon/2}n}}$	$\sqrt{\frac{k}{n}}$	$\sqrt{\frac{k}{n}}$
$(2\log k, \infty)$	$\sqrt{\frac{k}{n}}$	$\sqrt{\frac{k}{n}}$	$\sqrt{\frac{k}{n}}$	$\sqrt{\frac{k}{n}}$

Summary of known results for ℓ_1 loss

- Sample complexity: Let ℓ_1-risk $= \delta$

ϵ	RR	RAPPOR	SS	HR
$(0,1)$	$\frac{k^3}{\epsilon^2\delta^2}$	$\frac{k^2}{\epsilon^2\delta^2}$	$\frac{k^2}{\epsilon^2\delta^2}$	$\frac{k^2}{\epsilon^2\delta^2}$
$(1,\log k)$	$\frac{k^3}{e^{2\epsilon}\delta^2}$	$\frac{k^2}{e^{\epsilon/2}\delta^2}$	$\frac{k^2}{e^{\epsilon}\delta^2}$	$\frac{k^2}{e^{\epsilon}\delta^2}$
$(\log k, 2\log k)$	$\frac{k}{\delta^2}$	$\frac{k^2}{e^{\epsilon/2}\delta^2}$	$\frac{k}{\delta^2}$	$\frac{k}{\delta^2}$
$(2\log k, \infty)$	$\frac{k}{\delta^2}$	$\frac{k}{\delta^2}$	$\frac{k}{\delta^2}$	$\frac{k}{\delta^2}$

Summary of known results for ℓ_1 loss

- Communication complexity

ϵ	RR	RAPPOR	SS	HR
$(0, 1)$	$\log k$	k	k	$\log k$
$(1, \log k)$	$\log k$	$\frac{k}{e^{\epsilon/2}}$	$\frac{k}{e^{\epsilon}}$	$\log k$
$(\log k, 2\log k)$	$\log k$	$\frac{k}{e^{\epsilon/2}}$	$\log k$	$\log k$
$(2\log k, \infty)$	$\log k$	$\log k$	$\log k$	$\log k$

Summary of known results for ℓ_1 loss

- Communication complexity

ϵ	RR	RAPPOR	SS	HR
$(0, 1)$	$\log k$	k	k	$\log k$
$(1, \log k)$	$\log k$	$\frac{k}{e^{\epsilon/2}}$	$\frac{k}{e^{\epsilon}}$	$\log k$
$(\log k, 2\log k)$	$\log k$	$\frac{k}{e^{\epsilon/2}}$	$\log k$	$\log k$
$(2\log k, \infty)$	$\log k$	$\log k$	$\log k$	$\log k$

- Implementation complexity of HR is $n + k$, while for SS it is $n \cdot \max\{1, \frac{k}{e^{\epsilon}}\} + k$

Summary of known results for ℓ_1 loss

- Communication complexity

ϵ	RR	RAPPOR	SS	HR
$(0, 1)$	$\log k$	k	k	$\log k$
$(1, \log k)$	$\log k$	$\frac{k}{e^{\epsilon/2}}$	$\frac{k}{e^{\epsilon}}$	$\log k$
$(\log k, 2\log k)$	$\log k$	$\frac{k}{e^{\epsilon/2}}$	$\log k$	$\log k$
$(2\log k, \infty)$	$\log k$	$\log k$	$\log k$	$\log k$

- Implementation complexity of HR is $n + k$, while for SS it is $n \cdot \max\{1, \frac{k}{e^{\epsilon}}\} + k$

Thank you!