Electronic Journal of Statistics
Vol. 13 (2019) 4102–4120
ISSN: 1935-7524
https://doi.org/10.1214/19-EJS1625

Optimal locally private estimation under ℓ_p loss for $1 \le p \le 2$

Min Ye

Department of Electrical Engineering
Princeton University
Princeton, NJ, 08544
e-mail: yeemmi@gmail.com

Alexander Barg[*]

Department of Electrical and Computer Engineering
and Institute for Systems Research
University of Maryland
College Park, MD 20742
e-mail: abarg@umd.edu

Abstract: We consider the minimax estimation problem of a discrete distribution with support size k under locally differential privacy constraints. A privatization scheme is applied to each raw sample independently, and we need to estimate the distribution of the raw samples from the privatized samples. A positive number ϵ measures the privacy level of a privatization scheme.

In our previous work (*IEEE Trans. Inform. Theory*, 2018), we proposed a family of new privatization schemes and the corresponding estimator. We also proved that our scheme and estimator are order optimal in the regime $e^\epsilon \ll k$ under both ℓ_2^2 (mean square) and ℓ_1 loss. In this paper, we sharpen this result by showing asymptotic optimality of the proposed scheme under the ℓ_p^p loss for all $1 \le p \le 2$. More precisely, we show that for any $p \in [1, 2]$ and any k and ϵ, the ratio between the worst-case ℓ_p^p estimation loss of our scheme and the optimal value approaches 1 as the number of samples tends to infinity. The lower bound on the minimax risk of private estimation that we establish as a part of the proof is valid for any loss function $\ell_p^p, p \ge 1$.

AMS 2000 subject classifications: 62G05.
Keywords and phrases: Minimax estimation, local differential privacy.

Received August 2018.

1. Introduction

This paper continues our work [28]. The context of the problem that we consider is related to a major challenge in the statistical analysis of user data, namely, the conflict between learning accurate statistics and protecting sensitive information about the individuals. As in [28], we rely on a particular formalization

[*] A. Barg is also with Institute for Problems of Information Transmission (IITP), Russian Academy of Sciences, 127051 Moscow, Russia. His research was partially supported by NSF grants CCF1814487 and CCF1618603.

of user privacy called *differential privacy*, introduced in [9, 8]. Generally speaking, differential privacy requires that the adversary not be able to reliably infer an individual's data from public statistics even with access to all the other users' data. The concept of differential privacy has been developed in two different contexts: the *global privacy* context (for instance, when institutions release statistics related to groups of people) [12], and the *local privacy* context when individuals disclose their personal data [6].

In this paper, we consider the minimax estimation problem of a discrete distribution with support size k under locally differential privacy. This problem has been studied in the non-private setting [18, 20], where we can learn the distribution from the raw samples. In the private setting, we need to estimate the distribution of raw samples from the privatized samples which are generated independently from the raw samples according to a conditional distribution $\boldsymbol{Q}$ (also called a *privatization scheme*). Given a privacy parameter $\epsilon > 0$, we say that $\boldsymbol{Q}$ is ϵ-locally differentially private if the probabilities of the same output conditional on different inputs differ by a factor of at most e^ϵ. Clearly, smaller ϵ means that it is more difficult to infer the original data from the privatized samples, and thus leads to higher privacy. For a given ϵ, our objective is to find the optimal ϵ-private scheme that minimizes the expected estimation loss for the worst-case distribution. In this paper, we are mainly concerned with the scenario where we have a large number of samples, which captures the modern trend toward "big data" analytics.

1.1. Existing results

The following two privatization schemes are the most well-known in the literature: the k-ary Randomized Aggregatable Privacy-Preserving Ordinal Response (k-RAPPOR) scheme [5, 10], and the k-ary Randomized Response (k-RR) scheme [26, 17]. The k-RAPPOR scheme is order optimal in the high privacy regime where ϵ is very close to 0, and the k-RR scheme is order optimal in the low privacy regime where $e^\epsilon \approx k$ [16]. Very recently, a family of privatization schemes and the corresponding estimators were proposed independently by Wang et al. [25] and the present authors [28]. In [28], we further showed that under both ℓ_2^2 (mean square) and ℓ_1 loss, these privatization schemes and the corresponding estimators are order-optimal in the medium to high privacy regimes when $e^\epsilon \ll k$. Subsequent to our work, [3] proposed another privatization scheme and proved that it is order optimal in all regimes for ℓ_1 loss. At the same time, prior to this paper, no schemes were shown to be asymptotically optimal in the literature.

Duchi et al. [7] gave an order-optimal lower bound on the minimax private estimation loss for the high privacy regime where ϵ is very close to 0. In [28], we proved a stronger lower bound which is order-optimal in the whole region $e^\epsilon \ll k$. This lower bound implies that the schemes and the estimators proposed in [25, 28] are order optimal in this regime. Here order-optimal means that the ratio between the true value and the lower bound is upper bounded by a constant (larger than 1) when n and k/e^ϵ both become large enough.

1.2. Our contributions

In this paper, we study the private estimation problem under the ℓ_p^p loss for $1 \le p \le 2$, which in particular includes the widely used ℓ_1 and ℓ_2^2 loss. We prove an asymptotically tight lower bound on the ℓ_p^p loss of the minimax private estimation for all values of k, ϵ and $1 \le p \le 2$. This improves upon the lower bounds in [28] and [7] for the following three reasons: First, although the lower bounds in [28] and [7] are order-optimal, they differ from the true value by a factor of several hundred. In practice, an improvement of several percentage points is already considered as a substantial advance (see for instance, [16]), so tighter bounds are of interest. Second, the bounds in [28] and [7] only hold for certain regions of k and ϵ while the lower bound in this paper holds for all values of k and ϵ. Finally, previous results were limited to ℓ_1 and ℓ_2^2 loss functions while the results in this paper hold for all ℓ_p^p loss functions, where $1 \le p \le 2$.

Furthermore, as an immediate consequence of our lower bound, we show that the schemes and the estimators proposed in [25, 28] are universally optimal under the ℓ_p^p loss for all $1 \le p \le 2$ in the sense that the ratio between the lower bound and the worst-case estimation loss of these schemes and estimators goes to 1 when n goes to infinity.

In this paper we both generalize the results, and shorten the proofs in the preprint [27] which addressed only the case of mean square loss.

1.3. Related work

While in this paper we consider only the sample complexity, a recent work by Acharya et al. [3] took communication complexity into consideration and proposed a new privatization scheme with reduced communication complexity while maintaining the optimal order of sample complexity for the ℓ_1 loss function. Apart from the ℓ_p loss measures considered in this paper, significant attention in the literature was devoted to the ℓ_∞ estimation of a discrete distribution (also called the heavy hitters problem) under local differential privacy [21, 14, 4]. Although we only consider the case where the same privatization scheme is applied to each raw sample in this paper, one can also construct privatization schemes that depend on the values of previously observed privatized samples. Such interactive privatization schemes are important for online and sequential procedures in private learning [22, 23, 7]. A recent work [1] addresses the private estimation problem of distributional properties when the support size k is not known to the estimator. Other estimation-related problems that were studied under local differential privacy constraints include the problem of testing identity and closeness of discrete distributions [2] and hypothesis testing [11].

1.4. Organization of the paper

In Section 2, we formulate the problem and give a more detailed review of the existing results. Section 3 is devoted to an overview of the main results of this paper. The proofs of the main results are given in Sections 4-5.

2. Problem formulation and existing results

Notation:
Let $\mathcal{X} = \{1, 2, \ldots, k\}$ be the source alphabet and let $\boldsymbol{p} = (p_1, p_2, \ldots, p_k)$ be a probability distribution on $\mathcal{X}$. Denote by $\Delta_k = \{\boldsymbol{p} \in \mathbb{R}^k : p_i \ge 0 \text{ for } i = 1, 2, \ldots, k, \sum_{i=1}^k p_i = 1\}$ the k-dimensional probability simplex. Let X be a random variable (RV) that takes values on $\mathcal{X}$ according to $\boldsymbol{p}$, so that $p_i = P(X = i)$. Denote by $X^n = (X^{(1)}, X^{(2)}, \ldots, X^{(n)})$ the vector formed of n independent copies of the RV X. Denote the uniform distribution as $\boldsymbol{p}_U = (1/k, 1/k, \ldots, 1/k)$.

2.1. Problem formulation

In the classical (non-private) distribution estimation problem, we are given direct access to i.i.d. samples $\{X^{(i)}\}_{i=1}^n$ drawn according to some unknown distribution $\boldsymbol{p} \in \Delta_k$. Our goal is to estimate $\boldsymbol{p}$ based on the samples [20]. We define an estimator $\hat{\boldsymbol{p}}$ as a function $\hat{\boldsymbol{p}} : \mathcal{X}^n \to \mathbb{R}^k$, and assess its quality in terms of the worst-case risk (expected loss)

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

where ℓ is some loss function. The minimax risk is defined as the solution of the following saddlepoint problem:

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

In the private distribution estimation problem, we can no longer access the raw samples $\{X^{(i)}\}_{i=1}^n$. Instead, we estimate the distribution $\boldsymbol{p}$ from the privatized samples $\{Y^{(i)}\}_{i=1}^n$, obtained by applying a privatization mechanism $\boldsymbol{Q}$ independently to each raw sample $X^{(i)}$. A *privatization mechanism* (also called privatization scheme) $\boldsymbol{Q} : \mathcal{X} \to \mathcal{Y}$ is simply a conditional distribution $\boldsymbol{Q}_{Y|X}$. The privatized samples $Y^{(i)}$ take values in a set $\mathcal{Y}$ (the "output alphabet") that does not have to be the same as $\mathcal{X}$.

The quantities $\{Y^{(i)}\}_{i=1}^n$ are i.i.d. samples drawn according to the marginal distribution $\boldsymbol{m}$ given by

$$\boldsymbol{m}(S) = \sum_{i=1}^k \boldsymbol{Q}(S|i) p_i \tag{1}$$

for any $S \in \sigma(\mathcal{Y})$, where $\sigma(\mathcal{Y})$ denotes an appropriate σ-algebra on $\mathcal{Y}$. In accordance with this setting, the estimator $\hat{\boldsymbol{p}}$ is a measurable function $\hat{\boldsymbol{p}} : \mathcal{Y}^n \to \mathbb{R}^k$. We assess the quality of the privatization scheme $\boldsymbol{Q}$ and the corresponding estimator $\hat{\boldsymbol{p}}$ by the worst-case risk

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

where $\boldsymbol{m}^n$ is the n-fold product distribution and $\boldsymbol{m}$ is given by (1). Define the *minimax risk* of the privatization scheme $\boldsymbol{Q}$ as

$$r_{k,n}^{\ell}(\boldsymbol{Q}) := \inf_{\hat{\boldsymbol{p}}} r_{k,n}^{\ell}(\boldsymbol{Q}, \hat{\boldsymbol{p}}). \tag{2}$$

Definition 2.1. *For a given $\epsilon > 0$, a privatization mechanism $\boldsymbol{Q} : \mathcal{X} \to \mathcal{Y}$ is said to be* ϵ-locally differentially private *if for all $x, x' \in \mathcal{X}$*

$$\sup_{S \in \sigma(\mathcal{Y})} \log \frac{\boldsymbol{Q}(Y \in S | X = x)}{\boldsymbol{Q}(Y \in S | X = x')} \le \epsilon. \tag{3}$$

Denote by $\mathcal{D}_\epsilon$ the set of all ϵ-locally differentially private mechanisms. Given a privacy level ϵ and a loss function ℓ, we seek to find the optimal $\boldsymbol{Q} \in \mathcal{D}_\epsilon$ with the smallest possible minimax risk $r_{k,n}^{\ell}(\boldsymbol{Q})$ among all the ϵ-locally differentially private mechanisms. As already mentioned, in this paper we will consider[1] $\ell = \ell_u^u$ for $1 \le u \le 2$, where for $x = (x_1, x_2, \dots, x_k) \in \mathbb{R}^k$

$$\ell_u^u(x) := \sum_{i=1}^k |x_i|^u.$$

It is easy to see that for any valid privatization scheme $\boldsymbol{Q}$, the order of its ℓ_u^u minimax estimation risk is $\Theta(n^{-u/2})$, and $\lim_{n\to\infty} r_{k,n}^{\ell_u^u}(\boldsymbol{Q}) n^{u/2}$ is the coefficient of the dominant term, which measures the performance of $\boldsymbol{Q}$ when n is large.

Main Problem: *Suppose that the cardinality k of the source alphabet is known to the estimator. For a given privacy level ϵ, we would like to find the optimal (smallest possible) value of* $\lim_{n\to\infty} r_{k,n}^{\ell_u^u}(\boldsymbol{Q}) n^{u/2}$ *among all $\boldsymbol{Q} \in \mathcal{D}_\epsilon$ and to construct a privatization mechanism and a corresponding estimator to achieve this optimal value.*

It is this problem that we address—and resolve—in this paper. Specifically, we prove a lower bound on $\lim_{n\to\infty} r_{k,n}^{\ell_u^u}(\boldsymbol{Q}) n^{u/2}$ for $\boldsymbol{Q} \in \mathcal{D}_\epsilon$, which implies that the mechanism and the corresponding estimator proposed in [28] are universally optimal for all loss functions $\ell_u^u, 1 \le u \le 2$.

2.2. Previous results

In this section we briefly review known results that are relevant to our problem. In Sect. 1.1 we mentioned several papers that have considered it, viz., [26, 5, 10, 17, 16, 25, 7, 3]. In this section we focus on the results of [28] because they are stated in the form convenient for our presentation.

[1] The standard notation for the loss function should be ℓ_p^p, as we used in the Introduction. However, in order to avoid confusion with the notation for probability distribution, we will use ℓ_u^u from now on.

Let $\mathcal{D}_{\epsilon,F}$ be the set of ϵ-locally differentially private schemes with finite output alphabet. Let

$$\mathcal{D}_{\epsilon,E} = \left\{ \boldsymbol{Q} \in \mathcal{D}_{\epsilon,F} : \frac{\boldsymbol{Q}(y|x)}{\min_{x' \in \mathcal{X}} \boldsymbol{Q}(y|x')} \in \{1, e^{\epsilon}\} \text{ for all } x \in \mathcal{X} \text{ and all } y \in \mathcal{Y} \right\}. \tag{4}$$

In [28, Theorem 13], we have shown that

$$r_{k,n}^{\ell_u^u}(\boldsymbol{Q}) \ge \inf_{\boldsymbol{Q}' \in \mathcal{D}_{\epsilon,E}} r_{k,n}^{\ell_u^u}(\boldsymbol{Q}') \quad \text{for all } \boldsymbol{Q} \in \mathcal{D}_{\epsilon}. \tag{5}$$

As a result, below we limit ourselves to schemes $\boldsymbol{Q} \in \mathcal{D}_{\epsilon,E}$ in this paper. For such schemes, since the output alphabet is finite, we can write the marginal distribution $\boldsymbol{m}$ in (1) as a vector $\boldsymbol{m} = (\sum_{j=1}^{k} p_j \boldsymbol{Q}(y|j), y \in \mathcal{Y})$. We will also use the shorthand notation $\boldsymbol{m} = \boldsymbol{p}\boldsymbol{Q}$ to denote this vector.

In [28], we introduced a family of privatization schemes which are parameterized by the integer $d \in \{1, 2, \dots, k-1\}$. Given k and d, let the output alphabet be $\mathcal{Y}_{k,d} = \{y \in \{0,1\}^k : \sum_{i=1}^{k} y_i = d\}$, so $|\mathcal{Y}_{k,d}| = \binom{k}{d}$.

Definition 2.2 ([28])**.** *Consider the following privatization scheme:*

$$\boldsymbol{Q}_{k,\epsilon,d}(y|i) = \frac{e^{\epsilon} y_i + (1 - y_i)}{\binom{k-1}{d-1} e^{\epsilon} + \binom{k-1}{d}} \tag{6}$$

for all $y \in \mathcal{Y}_{k,d}$ and all $i \in \mathcal{X}$. The corresponding empirical estimator of $\boldsymbol{p}$ under $\boldsymbol{Q}_{k,\epsilon,d}$ is defined as follows: For $y^n = (y^{(1)}, y^{(2)}, \dots, y^{(n)}) \in \mathcal{Y}_{k,d}^n$,

$$\hat{p}_i(y^n) = \left(\frac{(k-1)e^{\epsilon} + \frac{(k-1)(k-d)}{d}}{(k-d)(e^{\epsilon} - 1)} \right) \frac{t_i(y^n)}{n} - \frac{(d-1)e^{\epsilon} + k - d}{(k-d)(e^{\epsilon} - 1)}, \quad i \in [k] \tag{7}$$

where $t_i(y^n) = \sum_{j=1}^{n} y_i^{(j)}$ is the number of privatized samples whose i-th coordinate is 1.

Some papers [3] call $\boldsymbol{Q}_{k,\epsilon,d}$ the *Subset Selection* mechanism. It is easy to verify that $\boldsymbol{Q}_{k,\epsilon,d}$ is ϵ-locally differentially private. The worst-case estimation loss under $\boldsymbol{Q}_{k,\epsilon,d}$ and the empirical estimator is calculated in the following proposition.

Proposition 2.3. [28, Prop. 4-5] *Let $\boldsymbol{Q} = \boldsymbol{Q}_{k,\epsilon,d}$ and suppose that the empirical estimator $\hat{\boldsymbol{p}}$ is given by (7). Let $\boldsymbol{m} = \boldsymbol{p}\boldsymbol{Q}_{k,\epsilon,d}$. The estimation loss $\underset{Y^n \sim \boldsymbol{m}^n}{\mathbb{E}} \ell_2^2(\hat{\boldsymbol{p}}(Y^n), \boldsymbol{p})$ is maximized for the uniform distribution $\boldsymbol{p}_U$, and*

$$r_{k,n}^{\ell_2^2}(\boldsymbol{Q}_{k,\epsilon,d}, \hat{\boldsymbol{p}}) = \underset{Y^n \sim \boldsymbol{m}_U^n}{\mathbb{E}} \ell_2^2(\hat{\boldsymbol{p}}(Y^n), \boldsymbol{p}_U) = \frac{(k-1)^2}{nk(e^{\epsilon} - 1)^2} \frac{(de^{\epsilon} + k - d)^2}{d(k-d)}, \tag{8}$$

where $\boldsymbol{m}_U = \boldsymbol{p}_U \boldsymbol{Q}_{k,\epsilon,d}$.

It is clear that the smallest value of the risk $\boldsymbol{r}$ is obtained by optimizing on d in (8). Namely, given k and ϵ, let

$$d^* = d^*(k, \epsilon) := \underset{1 \le d \le k-1}{\operatorname{argmin}} \frac{(de^{\epsilon} + k - d)^2}{d(k-d)}, \tag{9}$$

where the ties are resolved arbitrarily. We find that d^* takes one the following two values:

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

Therefore, when $k/(e^\epsilon+1)\le 1$, $d^*=1$, and when $k/(e^\epsilon+1)>1$, the value of d^* can be determined by simple comparison.

As a consequence of Prop. 2.3 we find that

$$r_{k,n}^{\ell_2^2}(\boldsymbol{Q}_{k,\epsilon,d^*},\hat{\boldsymbol{p}}) = \min_{1\le d\le k-1} r_{k,n}^{\ell_2^2}(\boldsymbol{Q}_{k,\epsilon,d},\hat{\boldsymbol{p}}).$$

While in [28] we proved the above results for the mean-square loss (and a similar claim for $\ell=\ell_1$), in this paper we show that they apply more universally. Namely, let

$$M(k,\epsilon) := \frac{(k-1)^2}{k^2(e^\epsilon-1)^2}\frac{(d^*e^\epsilon+k-d^*)^2}{d^*(k-d^*)}, \tag{10}$$

and note that $r_{k,n}^{\ell_2^2}(\boldsymbol{Q}_{k,\epsilon,d^*},\hat{\boldsymbol{p}}) = \frac{k}{n}M(k,\epsilon)$. In this paper we show that the quantity $M(k,\epsilon)$ bounds below the main term of the minimax risk for all loss functions $\ell_u^u, u\ge 1$.

3. Main result of the paper

Our main result is that the scheme $\boldsymbol{Q}_{k,\epsilon,d^*}$ and the empirical estimator $\hat{\boldsymbol{p}}$ defined by (7) are universally optimal for all loss functions $\ell_u^u, 1\le u\le 2$. Namely, the following is true.

Theorem 3.1. *Let $k=|\mathcal{X}|$, let $\epsilon>0, 1\le u\le 2$. Then*

$$\lim_{n\to\infty}\frac{r_{k,n}^{\ell_u^u}(\boldsymbol{Q})}{r_{k,n}^{\ell_u^u}(\boldsymbol{Q}_{k,\epsilon,d^*},\hat{\boldsymbol{p}})}\ge 1 \quad \textit{for all } \boldsymbol{Q}\in\mathcal{D}_\epsilon.$$

This theorem is a consequence of two results which we state next.

Let $X\sim\mathcal{N}(0,1)$ and define the constant

$$C_u := E|X|^u = 2^{u/2}\Gamma((u+1)/2)/\sqrt{\pi} \quad \text{for } u>0.$$

Theorem 3.2. *For any $\epsilon>0$, any $u\ge 1$, and any mechanism $\boldsymbol{Q}\in\mathcal{D}_\epsilon$*

$$\lim_{n\to\infty} r_{k,n}^{\ell_u^u}(\boldsymbol{Q})n^{u/2}\ge kC_uM(k,\epsilon)^{u/2}. \tag{11}$$

Note that this lower bound holds for any loss function $\ell_u^u, u\ge 1$. The proof of this theorem is given in Section 4.

Theorem 3.3. *Consider the privatization scheme $\boldsymbol{Q}=\boldsymbol{Q}_{k,\epsilon,d^*}$ and let $\hat{\boldsymbol{p}}$ be the empirical estimator given by (7). For every k and ϵ and every $0<u\le 2$,*

$$r_{k,n}^{\ell_u^u}(\boldsymbol{Q}_{k,\epsilon,d^*},\hat{\boldsymbol{p}}) = \frac{k}{n^{u/2}}C_uM(k,\epsilon)^{u/2}+o(n^{-u/2}).$$

The proof of this theorem is given in Section 5. Note that, unlike Theorem 3.2, the claim that we make here allows the values of $u \in (0,1)$. The special cases of Theorem 3.3 for $u = 1$ and $u = 2$ were addressed in our previous paper [28], see in particular Theorem 10.

The crux of our argument is in the proof of Theorem 3.2, where we reduce the estimation problem in the k-dimensional space to a one-dimensional problem. Generally, it is well known that the local minimax risk can be calculated from the inverse of the Fisher information matrix. However, it is difficult to obtain the exact expression of the inverse of a large-size matrix, and without it, the path to the desired estimates is not so clear. To work around this complication, we view a ball in a high-dimensional space as a union of parallel line segments with a certain direction $\boldsymbol{v}_i$. We first consider the estimation problem on each line segment individually. Since this is a one-dimensional problem, its minimax rate can be easily calculated from the Fisher information of the corresponding parameter. For the estimation of each component p_i of the probability distribution, we choose a suitable direction vector $\boldsymbol{v}_i$. In this way, we reduce the original k-dimensional estimation problem to k one-dimensional estimation problems and then rely on the additivity of the loss function for the final result.

4. Proof of Theorem 3.2

4.1. Bayes estimation loss

In light of (5), to prove Theorem 3.2, it suffices to show that for every $u \ge 1$,

$$\lim_{n\to\infty} r_{k,n}^{\ell_u^u}(\boldsymbol{Q}) n^{u/2} \ge k C_u M(k,\epsilon)^{u/2} \quad \text{for all } \boldsymbol{Q} \in \mathcal{D}_{\epsilon,E}. \tag{12}$$

Since the worst-case estimation loss is always lower bounded by the average estimation loss, the minimax risk $r_{k,n}^{\ell_u^u}(\boldsymbol{Q})$ can be bounded below by the Bayes estimation loss. More specifically, we assume that $\boldsymbol{p} := \{p_1, p_2, \dots, p_k\}$ is drawn uniformly from

$$\mathcal{P} := \left\{ \boldsymbol{p} \in \Delta_k : \|\boldsymbol{p} - \boldsymbol{p}_U\|_2 \le \frac{D}{\sqrt{n}} \right\}, \tag{13}$$

where $D \gg 1$ is a constant. Let $\boldsymbol{P} = (P_1, P_2, \dots, P_k)$ denote the random vector that corresponds to $\boldsymbol{p}$. For a given privatization scheme $\boldsymbol{Q}$ and the corresponding estimator $\hat{\boldsymbol{p}} := (\hat{p}_1, \hat{p}_2, \dots, \hat{p}_k)$, the ℓ_u^u Bayes estimation loss is defined as

$$\begin{aligned} r_{\text{Bayes}}^{\ell_u^u}(\boldsymbol{Q}, \hat{\boldsymbol{p}}) &:= \mathop{\mathbb{E}}_{\boldsymbol{P}\sim \text{Unif}(\mathcal{P})} \left[\mathop{\mathbb{E}}_{Y^n \sim (\boldsymbol{PQ})^n} \ell_u^u(\hat{\boldsymbol{p}}(Y^n), \boldsymbol{P}) \right] \\ &= \sum_{i=1}^k \left(\mathop{\mathbb{E}}_{\boldsymbol{P}\sim \text{Unif}(\mathcal{P})} \left[\mathop{\mathbb{E}}_{Y^n \sim (\boldsymbol{PQ})^n} |\hat{p}_i(Y^n) - P_i|^u \right] \right), \end{aligned}$$

and the optimal Bayes estimation loss for $\boldsymbol{Q}$ is

$$r_{\text{Bayes}}^{\ell_u^u}(\boldsymbol{Q}) := \inf_{\hat{\boldsymbol{p}}} r_{\text{Bayes}}^{\ell_u^u}(\boldsymbol{Q}, \hat{\boldsymbol{p}}).$$

We further define component-wise Bayes estimation loss for $\boldsymbol{Q}$ and $\hat{\boldsymbol{p}}$

$$r_{i,\text{Bayes}}^{\ell_u^u}(\boldsymbol{Q},\hat{p}_i) := \underset{\boldsymbol{P}\sim\text{Unif}(\mathcal{P})}{\mathbb{E}}\Big[\underset{Y^n\sim(\boldsymbol{P}\boldsymbol{Q})^n}{\mathbb{E}}|\hat{p}_i(Y^n)-P_i|^u\Big], \quad i\in[k],$$

and the optimal component-wise Bayes estimation loss for $\boldsymbol{Q}$

$$r_{i,\text{Bayes}}^{\ell_u^u}(\boldsymbol{Q}) := \inf_{\hat{p}_i} r_{i,\text{Bayes}}^{\ell_u^u}(\boldsymbol{Q},\hat{p}_i), \quad i\in[k].$$

Therefore,

$$r_{\text{Bayes}}^{\ell_u^u}(\boldsymbol{Q},\hat{\boldsymbol{p}}) = \sum_{i=1}^k r_{i,\text{Bayes}}^{\ell_u^u}(\boldsymbol{Q},\hat{p}_i), \quad r_{\text{Bayes}}^{\ell_u^u}(\boldsymbol{Q}) = \sum_{i=1}^k r_{i,\text{Bayes}}^{\ell_u^u}(\boldsymbol{Q}).$$

As mentioned above,

$$r_{k,n}^{\ell_u^u}(\boldsymbol{Q}) \ge r_{\text{Bayes}}^{\ell_u^u}(\boldsymbol{Q}) = \sum_{i=1}^k r_{i,\text{Bayes}}^{\ell_u^u}(\boldsymbol{Q}).$$

We will prove (12) by showing that

$$\sum_{i=1}^k r_{i,\text{Bayes}}^{\ell_u^u}(\boldsymbol{Q}) \ge \frac{k}{n^{u/2}} C_u M(k,\epsilon)^{u/2} - o(n^{-u/2}) \quad \text{for all } \boldsymbol{Q}\in\mathcal{D}_{\epsilon,E}. \tag{14}$$

4.2. Lower bound on one-dimensional Bayes estimation loss

Below we will prove a lower bound on $r_{i,\text{Bayes}}^{\ell_u^u}(\boldsymbol{Q})$. To this end, in this section we consider a one-dimensional Bayes estimation problem. Define the following vectors:

$$\boldsymbol{v}_i := \Big(-\frac{1}{k-1},\dots,-\frac{1}{k-1},1,-\frac{1}{k-1},\dots,-\frac{1}{k-1}\Big), \quad i\in[k], \tag{15}$$

where the 1 is in the ith position and all the other coordinates are $-\frac{1}{k-1}$. Let $\boldsymbol{p}^* := (p_1^*,p_2^*,\dots,p_k^*)\in\Delta_k$ be a probability distribution and let $S_i(\boldsymbol{p}^*)$ be a line segment with midpoint $\boldsymbol{p}^*$ and direction vector $\boldsymbol{v}_i$:

$$S_i(\boldsymbol{p}^*) := \Big\{\boldsymbol{p}^*+s\boldsymbol{v}_i : |s|\le\frac{D'}{\sqrt{n}}\Big\}, \quad i\in[k], \tag{16}$$

where $D'\gg 1$ is a constant. Let $\boldsymbol{p}=(p_1,\dots,p_k)$ be a PMF in the segment $S_i(\boldsymbol{p}^*)$. Given the value p_i, we can find all the other components of $\boldsymbol{p}$ as follows:

$$p_v = p_v^* - \frac{1}{k-1}(p_i-p_i^*) \quad \text{for all } v\neq i. \tag{17}$$

Assume that $\boldsymbol{p}=(p_1,p_2,\dots,p_k)$ is drawn uniformly from $S_i(\boldsymbol{p}^*)$, and we consider the Bayes estimation of p_i from the privatized samples Y^n obtained from applying $\boldsymbol{Q}$ to the raw samples. More precisely, for an estimator $\hat{p}_i$, we define its Bayes estimation loss

$$r^{\ell_u^u}_{i,S_i(\boldsymbol{p}^*)}(\boldsymbol{Q}, \hat{p}_i) := \mathop{\mathbb{E}}_{\boldsymbol{P}\sim\mathrm{Unif}(S_i(\boldsymbol{p}^*))}\Big[\mathop{\mathbb{E}}_{Y^n\sim(\boldsymbol{P}\boldsymbol{Q})^n}|\hat{p}_i(Y^n) - P_i|^u\Big], \quad i \in [k],$$

then the optimal estimation loss is

$$r^{\ell_u^u}_{i,S_i(\boldsymbol{p}^*)}(\boldsymbol{Q}) := \inf_{\hat{p}_i} r^{\ell_u^u}_{i,S_i(\boldsymbol{p}^*)}(\boldsymbol{Q}, \hat{p}_i), \quad i \in [k].$$

Our approach to obtain the lower bound on this Bayes estimation loss relies on a classical method in asymptotic statistics, namely, local asymptotic normality (LAN) of sequences of statistical models [19, 13], see also [15, 24]. The exact formulation of the results that pertain to the method involves several technical concepts; we will limit ourselves to explaining the general idea and the implications for our problem. We will also confine ourselves to the one-dimensional case as opposed to the general formulation of LAN. Let p_θ be the density function of a distribution P_θ, where the parameter θ takes values in an open subset $\Theta \subset \mathbb{R}$. For every fixed x we have the following Taylor expansion:

$$\log\frac{p_{\theta+h}}{p_\theta}(x) = h\frac{\partial}{\partial\theta}\log p_\theta(x) + \frac{1}{2}h^2\frac{\partial^2}{\partial\theta^2}\log p_\theta(x) + o(h^2).$$

Suppose that X^n are n i.i.d. samples drawn from the distribution P_θ. It follows that

$$\log\prod_{i=1}^n\frac{p_{\theta+h/\sqrt{n}}}{p_\theta}(X_i) = \frac{h}{\sqrt{n}}\sum_{i=1}^n\frac{\partial}{\partial\theta}\log p_\theta(X_i) + \frac{1}{2}\frac{h^2}{n}\sum_{i=1}^n\frac{\partial^2}{\partial\theta^2}\log p_\theta(X_i) + o(1).$$

Under some mild smoothness conditions, we have

$$\mathbb{E}_{X\sim P_\theta}\frac{\partial}{\partial\theta}\log p_\theta(X) = 0,$$
$$\mathbb{E}_{X\sim P_\theta}\Big(\frac{\partial}{\partial\theta}\log p_\theta(X)\Big)^2 = -\mathbb{E}_{X\sim P_\theta}\frac{\partial^2}{\partial\theta^2}\log p_\theta(X) = I_\theta,$$

where I_θ is the Fisher information of θ, which is assumed to be nonzero. Therefore, by central limit theorem, $\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{\partial}{\partial\theta}\log p_\theta(X_i)$ is asymptotically normal with mean zero and variance I_θ. Furthermore, $\frac{1}{n}\sum_{i=1}^n\frac{\partial^2}{\partial\theta^2}\log p_\theta(X_i)$ converges to I_θ, by the law of large numbers. Consequently, under suitable conditions we have

$$\log\prod_{i=1}^n\frac{p_{\theta+h/\sqrt{n}}}{p_\theta}(X_i) = hX - \frac{1}{2}I_\theta h^2 + o(1), \quad \text{where } X\sim\mathcal{N}(0, I_\theta)$$

The quadratic form on the right-hand side is very similar to the exponent of the Gaussian distribution, and one can derive a normal approximation from this similarity. More precisely, if T_n is a sequence of statistics in the experiments

$(P_{\theta+h/\sqrt{n}} : h \in \mathbb{R})$ such that T_n converges in distribution for every h, then there exists a (randomized) statistic T in the experiment $\mathcal{N}(h, I_\theta^{-1}), h \in \mathbb{R}$ such that T_n converges in distribution to T for every h; in other words, every converging sequence of statistics in the local experiments $(P_{\theta+h/\sqrt{n}} : h \in \mathbb{R})$ approaches in distribution the statistic of a single normal experiment. We refer in particular to [24, Ch.7] for a detailed, accessible account of the above informal discussion.

The implications of the general LAN results for our problem can be stated as follows. When the constant D' in (16) is large enough, the conditional distribution of P_i given $Y^n = y^n$ is approximately a Gaussian distribution with variance $(I_{p_i^*})^{-1}$ for almost all[2] $y^n \in \mathcal{Y}^n$ as n goes to infinity, where I_{p_i} is the Fisher information of the parameter p_i. Before we calculate the value of $I_{p_i^*}$, let us recall a simple fact about Gaussian distribution: Suppose that X is a Gaussian random variable, then one can easily verify[3] that for any $u \geq 1$,

$$\mathbb{E}X = \operatorname*{argmin}_a \mathbb{E}|X-a|^u. \tag{18}$$

Therefore, the estimator $\hat{p}_i(y^n) = \mathbb{E}(P_i|Y^n = y^n)$ is asymptotically optimal for this Bayes estimation problem under the ℓ_u^u loss function for all $u \geq 1$. Since the variance of P_i given $Y^n = y^n$ is $(I_{p_i^*})^{-1}$ for almost all $y^n \in \mathcal{Y}^n$, the Bayes estimation loss of this asymptotically optimal estimator is

$$C_u (I_{p_i^*})^{-u/2}(1-o(1)).$$

Thus we conclude that

$$r_{i,S_i(\boldsymbol{p}^*)}^{\ell_u^u}(\boldsymbol{Q}) \geq C_u (I_{p_i^*})^{-u/2}(1-o(1)) \quad \text{for all } u \geq 1. \tag{19}$$

Now we are left to calculate the value of $I_{p_i^*}$. To this end, we introduce some notation. For a given privatization scheme $\boldsymbol{Q} \in \mathcal{D}_{\epsilon,E}$ with output size L, we write its output alphabet as $\mathcal{Y} = \{1, 2, \dots, L\}$, and we use the shorthand notation

$$q_{jv} := \boldsymbol{Q}(j|v) \tag{20}$$

for all $j \in [L]$ and $v \in [k]$. For $j \in [L]$ and $y^n = (y^{(1)}, y^{(2)}, \dots, y^{(n)}) \in \mathcal{Y}^n$, define $w_j(y^n) := \sum_{v=1}^n \mathbb{1}[y^{(v)} = j]$ to be the number of times that symbol j appears in y^n. Let $\mathbb{P}(y^n; p_i)$ be the probability mass function of a random vector Y^n formed of i.i.d. samples drawn according to the distribution $\boldsymbol{m} = \boldsymbol{p}\boldsymbol{Q}$, where the other components of $\boldsymbol{p}$ are calculated from p_i according to (17). The random variables $w_j(Y^n)$ follow the multinomial distribution, and $\mathbb{E}w_j(Y^n) = n\boldsymbol{m}(j), j \in [L]$. Therefore,

$$\log \mathbb{P}(y^n; p_i) = \sum_{j=1}^L w_j(y^n) \log\Big(\sum_{v=1}^k p_v q_{jv}\Big)$$

[2] More precisely, for any $\epsilon_1, \epsilon_2 > 0$ there is N such that for any $n > N$ there is a subset $E \subseteq \mathcal{Y}^n$ such that (1) $\mathbb{P}(E) > 1 - \epsilon_1$, and (2) for all $y^n \in E$ the relative difference between the pdf of conditional distribution of P_i given $Y^n = y^n$ and the Gaussian pdf is at most ϵ_2.

[3] Let $\phi(x)$ be the pdf of X and note that $\phi(x) = \phi(2\mathbb{E}X - x)$ for all real x. By convexity of $|\cdot|^u, u \geq 1$ we have $|x - \mathbb{E}X|^u \leq (1/2)(|a-x|^u + |2\mathbb{E}X - x - a|^u)$ for all a. Integrating against $\phi(x)$ and using the symmetry condition, we obtain that $\mathbb{E}|X - \mathbb{E}X|^u \leq \mathbb{E}|X-a|^u$ for all $u \geq 1, a \in \mathbb{R}$.

$$= \sum_{j=1}^{L} w_j(y^n) \log\Big(p_i q_{ji} + \sum_{v\neq i}\Big(p_v^* - \frac{1}{k-1}(p_i - p_i^*)\Big) q_{jv}\Big),$$

and the Fisher information of p_i is

$$\begin{aligned}
I(p_i) &= - \mathop{\mathbb{E}}_{Y^n\sim(\boldsymbol{p}\boldsymbol{Q})^n}\left[\frac{d^2}{dp_i^2}\log \mathbb{P}(y^n;p_i)\right] \\
&= \sum_{j=1}^{L} \frac{(q_{ji} - \frac{1}{k-1}\sum_{v\neq i} q_{jv})^2}{\Big(p_i q_{ji} + \sum_{v\neq i}\Big(p_v^* - \frac{1}{k-1}(p_i - p_i^*)\Big)q_{jv}\Big)^2} \mathop{\mathbb{E}}_{Y^n\sim(\boldsymbol{p}\boldsymbol{Q})^n} w_j(Y^n) \\
&= \sum_{j=1}^{L} \frac{(q_{ji} - \frac{1}{k-1}\sum_{v\neq i} q_{jv})^2}{\Big(\sum_{v=1}^{k} p_v q_{jv}\Big)^2} \mathop{\mathbb{E}}_{Y^n\sim(\boldsymbol{p}\boldsymbol{Q})^n} w_j(Y^n) \\
&= n \sum_{j=1}^{L} \frac{(q_{ji} - \frac{1}{k-1}\sum_{v\neq i} q_{jv})^2}{\sum_{v=1}^{k} p_v q_{jv}} \\
&= \frac{nk^2}{(k-1)^2} \sum_{j=1}^{L} \frac{(q_{ji} - \frac{1}{k}\sum_{v=1}^{k} q_{jv})^2}{\sum_{v=1}^{k} p_v q_{jv}},
\end{aligned}$$

where p_v's on the last line are given by (17). In particular,

$$I_{p_i^*} = \frac{nk^2}{(k-1)^2} \sum_{j=1}^{L} \frac{(q_{ji} - \frac{1}{k}\sum_{v=1}^{k} q_{jv})^2}{\sum_{v=1}^{k} p_v^* q_{jv}}.$$

Combining this with (19), we have that for all $u \ge 1$,

$$r_{i,S_i(\boldsymbol{p}^*)}^{\ell_u^u}(\boldsymbol{Q}) \ge C_u \Big(\frac{nk^2}{(k-1)^2} \sum_{j=1}^{L} \frac{(q_{ji} - \frac{1}{k}\sum_{v=1}^{k} q_{jv})^2}{\sum_{v=1}^{k} p_v^* q_{jv}}\Big)^{-u/2} - o(n^{-u/2}).$$

For $j \in [L]$, define

$$q_j := \frac{1}{k}\sum_{v=1}^{k} q_{jv}. \tag{21}$$

It is clear that when $\boldsymbol{p}^*$ is in the neighborhood of the uniform distribution $\boldsymbol{p}_U$, i.e., when $p_v^* = 1/k + o_n(1)$ for all $v \in [k]$, we have

$$r_{i,S_i(\boldsymbol{p}^*)}^{\ell_u^u}(\boldsymbol{Q}) \ge C_u \Big(\frac{nk^2}{(k-1)^2} \sum_{j=1}^{L} \frac{(q_{ji} - q_j)^2}{q_j}\Big)^{-u/2} - o(n^{-u/2}) \quad \text{for all } u \ge 1. \tag{22}$$

4.3. Proof of (14)

Our first step in this section will be to prove a lower bound on $r_{i,\mathrm{Bayes}}^{\ell_u^u}(\boldsymbol{Q})$. Let us phrase the claim in (22) in a more detailed form: For any $\delta > 0$, there exists

$D_0 > 0$ such that whenever the constant D' in the definition of $S_i(\boldsymbol{p}^*)$ is larger than D_0,

$$r_{i,S_i(\boldsymbol{p}^*)}^{\ell_u^u}(\boldsymbol{Q}) \ge C_u\Big(\frac{nk^2}{(k-1)^2}\sum_{j=1}^{L}\frac{(q_{ji}-q_j)^2}{q_j}\Big)^{-u/2} - \delta n^{-u/2} \quad \text{for all } u \ge 1. \tag{23}$$

The constant D' is required to be large for the local asymptotic normality arguments to hold (refer again to [15, Chapter 2, Theorem 1.1] and [24, Ch. 7]).

Proposition 4.1. *Let $\mathcal{P}$ be the Euclidean ball around $\boldsymbol{p}_U$ defined in (13). For a sufficiently large constant D and any $u \ge 1$ we have*

$$r_{i,\text{Bayes}}^{\ell_u^u}(\boldsymbol{Q}) \ge C_u\Big(\frac{nk^2}{(k-1)^2}\sum_{j=1}^{L}\frac{(q_{ji}-q_j)^2}{q_j}\Big)^{-u/2} - o(n^{-u/2}). \tag{24}$$

Proof. We can view $\mathcal{P}$ as a union of (uncountably many) parallel line segments with direction vector $\boldsymbol{v}_i$ defined in (15). Each of these line segments can be written as $S_i(\boldsymbol{p}^*)$ (see (16)), with a suitably chosen midpoint $\boldsymbol{p}^* \in \mathcal{P}$. Since the midpoints of all the line segments lie inside $\mathcal{P}$, which is a neighborhood of the uniform distribution, by (23) we have that for any estimator $\hat{p}_i$, the average ℓ_u^u estimation loss $r_{i,S_i(\boldsymbol{p}^*)}^{\ell_u^u}(\boldsymbol{Q}, \hat{p}_i)$ on any of these line segments $S_i(\boldsymbol{p}^*)$ with $D' \ge D_0$ is lower bounded by

$$r_{i,S_i(\boldsymbol{p}^*)}^{\ell_u^u}(\boldsymbol{Q}, \hat{p}_i) \ge r_{i,S_i(\boldsymbol{p}^*)}^{\ell_u^u}(\boldsymbol{Q}) \ge C_u\Big(\frac{nk^2}{(k-1)^2}\sum_{j=1}^{L}\frac{(q_{ji}-q_j)^2}{q_j}\Big)^{-u/2} - \delta n^{-u/2}$$

for $u \ge 1$. To compute the average estimation loss $r_{i,\text{Bayes}}^{\ell_u^u}(\boldsymbol{Q}, \hat{p}_i)$ on $\mathcal{P}$ we need to average over all the segments with weight proportional to the length of the segment. Given D_0, we can choose D in (13) large enough so that the proportion of the segments $S_i(\boldsymbol{p}^*)$ with $D' \ge D_0$ out of all the segments in $\mathcal{P}$ is arbitrarily close to one (formally, denote the union of such segments as $\mathcal{P}_0$, then $\operatorname{Vol}(\mathcal{P}_0)/\operatorname{Vol}(\mathcal{P})$ can be made arbitrarily close to 1 as long as we set D/D_0 to be large enough). The average estimation loss along each of these segments is uniformly bounded below as in (23), and thus the average loss on $\mathcal{P}_0$ is lower bounded by the same quantity. Combining the fact that $\operatorname{Vol}(\mathcal{P}_0)/\operatorname{Vol}(\mathcal{P}) = 1 - o(1)$, we have

$$r_{i,\text{Bayes}}^{\ell_u^u}(\boldsymbol{Q}, \hat{p}_i) \ge C_u\Big(\frac{nk^2}{(k-1)^2}\sum_{j=1}^{L}\frac{(q_{ji}-q_j)^2}{q_j}\Big)^{-u/2} - o(n^{-u/2}) \quad \text{for all } u \ge 1.$$

This lower bound holds for any estimator $\hat{p}_i$, and this implies the claimed lower bound (24). □

We will need the following lemma.

Lemma 4.2. *For every $\boldsymbol{Q} \in \mathcal{D}_{\epsilon,E}$ with output alphabet $\mathcal{Y} = \{1, 2, \dots, L\}$ we have*

$$\sum_{i=1}^{k} \frac{q_{ji}^2}{q_j^2} \le k\Big(1 + (e^\epsilon - 1)^2 \frac{d^*(k-d^*)}{(d^* e^\epsilon + k - d^*)^2}\Big) \qquad \textit{for all } j \in [L].$$

Proof. Let $m_j := \min_{i \in [k]} q_{ji}$. According to the definition of $\mathcal{D}_{\epsilon,E}$ in (4), the coordinates of the vector $(q_{ji}, i \in [k])$ are either $m_j e^\epsilon$ or m_j. Let d be the number of $m_j e^\epsilon$ entries, then

$$q_j = \frac{m_j}{k}(de^\epsilon + k - d),$$
$$\sum_{i=1}^{k} q_{ji}^2 = m_j^2 (de^{2\epsilon} + k - d).$$

We obtain

$$\begin{aligned}
\sum_{i=1}^{k} \frac{q_{ji}^2}{q_j^2} &= \frac{k^2(de^{2\epsilon} + k - d)}{(de^\epsilon + k - d)^2} = k\frac{(de^{2\epsilon} + k - d)(d + k - d)}{(de^\epsilon + k - d)^2} \\
&= k\frac{d^2 e^{2\epsilon} + (k-d)^2 + d(k-d)(e^{2\epsilon} + 1)}{(de^\epsilon + k - d)^2} \\
&= k\frac{d^2 e^{2\epsilon} + 2d(k-d)e^\epsilon + (k-d)^2 + d(k-d)(e^{2\epsilon} - 2e^\epsilon + 1)}{(de^\epsilon + k - d)^2} \\
&= k\frac{(de^\epsilon + k - d)^2 + d(k-d)(e^\epsilon - 1)^2}{(de^\epsilon + k - d)^2} \\
&= k\Big(1 + (e^\epsilon - 1)^2 \frac{d(k-d)}{(de^\epsilon + k - d)^2}\Big) \\
&\le k\Big(1 + (e^\epsilon - 1)^2 \frac{d^*(k-d^*)}{(d^* e^\epsilon + k - d^*)^2}\Big),
\end{aligned}$$

where the last inequality follows from the definition of d^* in (9). □

Now we are ready to prove (14). Using the obvious relations $\sum_{j=1}^{L} q_{ji} = \sum_{j=1}^{L} q_j = 1$, we can simplify the right-hand side of (24) as follows:

$$\begin{aligned}
\sum_{j=1}^{L} \Big(\frac{(q_{ji} - q_j)^2}{q_j}\Big) &= \sum_{j=1}^{L} \Big(\sum_{j=1}^{L} \frac{q_{ji}^2}{q_j} - 2\sum_{j=1}^{L} q_{ji} + \sum_{j=1}^{L} q_j\Big)\Big) \\
&= \sum_{j=1}^{L} \frac{q_{ji}^2}{q_j} - 1.
\end{aligned}$$

Now let us sum (24) over $i \in [k]$ on both sides and use the simplification above:

$$\sum_{i=1}^{k} r_{i,\text{Bayes}}^{\ell_u^u}(\boldsymbol{Q}) \ge C_u \sum_{i=1}^{k} \Big(\frac{nk^2}{(k-1)^2}\Big(\sum_{j=1}^{L} \frac{q_{ji}^2}{q_j} - 1\Big)\Big)^{-u/2} - o(n^{-u/2}). \tag{25}$$

Since for $u > 0$, $x^{-u/2}$ is a convex function for $x > 0$, we can further bound below the right-hand side of (25):

$$\begin{aligned}
\sum_{i=1}^{k}\Big(\frac{nk^2}{(k-1)^2}\Big(\sum_{j=1}^{L}\frac{q_{ji}^2}{q_j}-1\Big)\Big)^{-u/2} &\ge k\Big(\frac{1}{k}\sum_{i=1}^{k}\frac{nk^2}{(k-1)^2}\Big(\sum_{j=1}^{L}\frac{q_{ji}^2}{q_j}-1\Big)\Big)^{-u/2}\\
&= k\Big(\frac{nk}{(k-1)^2}\sum_{j=1}^{L}\sum_{i=1}^{k}\frac{q_{ji}^2}{q_j}-\frac{nk^2}{(k-1)^2}\Big)^{-u/2}\\
&= k\Big(\frac{nk}{(k-1)^2}\sum_{j=1}^{L}\Big(q_j\sum_{i=1}^{k}\frac{q_{ji}^2}{q_j^2}\Big)-\frac{nk^2}{(k-1)^2}\Big)^{-u/2}\\
&\ge k\Big(\frac{nk^2}{(k-1)^2}\Big(1+(e^\epsilon-1)^2\frac{d^*(k-d^*)}{(d^*e^\epsilon+k-d^*)^2}\Big)\sum_{j=1}^{L}q_j-\frac{nk^2}{(k-1)^2}\Big)^{-u/2}\\
&= k\Big(\frac{nk^2(e^\epsilon-1)^2}{(k-1)^2}\frac{d^*(k-d^*)}{(d^*e^\epsilon+k-d^*)^2}\Big)^{-u/2}\\
&= \frac{k}{n^{u/2}}M(k,\epsilon)^{u/2} \qquad\qquad \text{for all } \boldsymbol{Q}\in\mathcal{D}_{\epsilon,E}
\end{aligned}$$

where the second inequality follows by Lemma 4.2 (note the inverted inequality of the Lemma because of the negative power $-u/2$). Combining this with (25), we conclude that

$$\sum_{i=1}^{k} r_{i,\text{Bayes}}^{\ell_u^u}(\boldsymbol{Q}) \ge \frac{k}{n^{u/2}}C_u M(k,\epsilon)^{u/2} - o(n^{-u/2}) \qquad \text{for all } \boldsymbol{Q}\in\mathcal{D}_{\epsilon,E}.$$

Thus we have established (14), and this completes the proof of Theorem 3.2.

5. Proof of Theorem 3.3

We begin with showing that for the privatization scheme $\boldsymbol{Q}_{k,\epsilon,d}$ defined in (6) and the estimator (7), the ℓ_u^u estimation loss is maximized for the uniform distribution $\boldsymbol{p}_U$ for all $0 < u \le 2$ when n is large. To shorten the notation, rewrite (7) as

$$\hat{p}_i(y^n) = A\frac{t_i(y^n)}{n} - B, \quad i \in [k],$$

where

$$A := \frac{(k-1)e^\epsilon + \frac{(k-1)(k-d)}{d}}{(k-d)(e^\epsilon-1)}, \qquad B := \frac{(d-1)e^\epsilon + k - d}{(k-d)(e^\epsilon-1)}.$$

In [28] we have shown that the estimator $\hat{p}_i(y^n)$ is unbiased, i.e.,

$$p_i = A\underset{Y^n\sim(\boldsymbol{p}\boldsymbol{Q}_{k,\epsilon,d})^n}{\mathbb{E}}\Big(\frac{t_i(Y^n)}{n}\Big) - B, \quad i \in [k].$$

By definition,

$$t_i(Y^n) = \sum_{j=1}^{n} \mathbb{1}[Y_i^{(j)} = 1]$$

is the sum of n i.i.d. Bernoulli random variables with parameter

$$\mathbb{P}[Y_i^{(j)} = 1] = \mathbb{E}\frac{t_i(Y^n)}{n} = \frac{p_i}{A} + \frac{B}{A}.$$

Therefore the variance of $\frac{t_i(Y^n)}{n}$ is $\frac{1}{n}(\frac{p_i}{A} + \frac{B}{A})(1 - \frac{p_i}{A} - \frac{B}{A})$, and the variance of $\hat{p}_i(Y^n)$ is

$$\operatorname{Var} \hat{p}_i(Y^n) = A^2 \frac{1}{n}\Big(\frac{p_i}{A} + \frac{B}{A}\Big)\Big(1 - \frac{p_i}{A} - \frac{B}{A}\Big) = \frac{1}{n}(p_i + B)(A - p_i - B).$$

Using the Central Limit Theorem, we then obtain for the absolute moment of $\hat{p}_i(Y^n)$ around p_i the following approximation:

$$\mathop{\mathbb{E}}_{Y^n \sim (\boldsymbol{p}\boldsymbol{Q}_{k,\epsilon,d})^n} |\hat{p}_i(Y^n) - p_i|^u = C_u \Big(\frac{1}{n}(p_i + B)(A - p_i - B)\Big)^{u/2} + o(n^{-u/2}),$$

where C_u is the absolute moment of the $\mathcal{N}(0,1)$ RV; see Section 3. Therefore,

$$\begin{aligned}\mathop{\mathbb{E}}_{Y^n \sim (\boldsymbol{p}\boldsymbol{Q}_{k,\epsilon,d})^n} \ell_u^u(\hat{\boldsymbol{p}}(Y^n), \boldsymbol{p}) &= \sum_{i=1}^{k} C_u \Big(\frac{1}{n}(p_i + B)(A - p_i - B)\Big)^{u/2} + o(n^{-u/2}) \\ &\le k C_u n^{-u/2} \Big(\frac{1}{k}\sum_{i=1}^{k}(p_i + B)(A - p_i - B)\Big)^{u/2} + o(n^{-u/2}) \\ &= k C_u n^{-u/2} \Big(\frac{A}{k} - \frac{2B}{k} + AB - B^2 - \frac{1}{k}\sum_{i=1}^{k} p_i^2\Big)^{u/2} + o(n^{-u/2}) \\ &\le k C_u n^{-u/2} \Big(\frac{A}{k} - \frac{2B}{k} + AB - B^2 - \frac{1}{k^2}\Big)^{u/2} + o(n^{-u/2}),\end{aligned}$$

where the first inequality follows from the fact that $x^{u/2}$ is a concave function of x on $(0, +\infty)$ for all positive $0 < u \le 2$, and the last line uses the Cauchy–Schwarz inequality. Both inequalities hold with equality if and only if $\boldsymbol{p}$ is the uniform distribution. Thus when n is large, for all $0 < u \le 2$ and all $1 \le d \le k - 1$, we have

$$r_{k,n}^{\ell_u^u}(\boldsymbol{Q}_{k,\epsilon,d}, \hat{\boldsymbol{p}}) = \mathop{\mathbb{E}}_{Y^n \sim (\boldsymbol{p}_U \boldsymbol{Q}_{k,\epsilon,d})^n} \ell_u^u(\hat{\boldsymbol{p}}(Y^n), \boldsymbol{p}_U).$$

In particular, it also holds for $d = d^*$. Next we calculate the estimation loss at the uniform distribution. By symmetry, it is clear that

$$\mathop{\mathbb{E}}_{Y^n \sim (\boldsymbol{p}_U \boldsymbol{Q}_{k,\epsilon,d^*})^n} \Big|\hat{p}_i(Y^n) - \frac{1}{k}\Big|^2 = \frac{1}{k}\Big(\mathop{\mathbb{E}}_{Y^n \sim (\boldsymbol{p}_U \boldsymbol{Q}_{k,\epsilon,d^*})^n} \ell_2^2(\hat{\boldsymbol{p}}(Y^n), \boldsymbol{p}_U)\Big)$$

$$= \frac{1}{k} r^{\ell_2^2}_{k,n}(\boldsymbol{Q}_{k,\epsilon,d^*}, \hat{\boldsymbol{p}}) = \frac{M(k,\epsilon)}{n}.$$

Therefore when the input distribution is uniform, $\hat{p}_i(Y^n)$ can be approximated for large n by a Gaussian random variable with mean $1/k$ and variance $\frac{M(k,\epsilon)}{n}$. Thus,

$$\underset{Y^n \sim (\boldsymbol{p}_U \boldsymbol{Q}_{k,\epsilon,d^*})^n}{\mathbb{E}} \left| \hat{p}_i(Y^n) - \frac{1}{k} \right|^u = C_u \Big(\frac{M(k,\epsilon)}{n} \Big)^{u/2} + o(n^{-u/2}),$$

so for $0 < u \le 2$,

$$\begin{aligned} r^{\ell_u^u}_{k,n}(\boldsymbol{Q}_{k,\epsilon,d^*}, \hat{\boldsymbol{p}}) &= \underset{Y^n \sim (\boldsymbol{p}_U \boldsymbol{Q}_{k,\epsilon,d^*})^n}{\mathbb{E}} \ell_u^u(\hat{\boldsymbol{p}}(Y^n), \boldsymbol{p}_U) \\ &= \frac{k}{n^{u/2}} C_u M(k,\epsilon)^{u/2} + o(n^{-u/2}). \end{aligned}$$

This completes the proof of Theorem 3.3.

References

[1] Acharya, J., Kamath, G., Sun, Z. and Zhang, H. (2018). INSPECTRE: privately estimating the unseen. In *Proceedings of the 35th International Conference on Machine Learning* **80** 30–39.

[2] Acharya, J., Sun, Z. and Zhang, H. (2018). Differentially private testing of identity and closeness of discrete distributions. In *Advances in Neural Information Processing Systems 31* 6878–6891. Curran Associates, Inc.

[3] Acharya, J., Sun, Z. and Zhang, H. (2019). Hadamard response: estimating distributions privately, efficiently, and with little communication. In *Proceedings of Machine Learning Research* **89** 1120–1129.

[4] Bassily, R. and Smith, A. (2015). Local, private, efficient protocols for succinct histograms. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing* 127–135. ACM. MR3388190

[5] Duchi, J., Wainwright, M. J. and Jordan, M. I. (2013). Local privacy and minimax bounds: Sharp rates for probability estimation. In *Advances in Neural Information Processing Systems* 1529–1537. MR3727612

[6] Duchi, J. C., Jordan, M. I. and Wainwright, M. J. (2013). Local privacy and statistical minimax rates. In *54th Annual IEEE Symposium on the Foundations of Computer Science (FOCS)* 429–438. MR3246246

[7] Duchi, J. C., Jordan, M. I. and Wainwright, M. J. (2018). Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association* **113** 182–201. MR3803452

[8] Dwork, C. (2008). Differential privacy: A survey of results. In *International Conference on Theory and Applications of Models of Computation* 1–19. Springer.

[9] Dwork, C., McSherry, F., Nissim, K. and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography Conference* 265–284. Springer. MR2241676

[10] Erlingsson, Ú., Pihur, V. and Korolova, A. (2014). RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* 1054–1067. ACM.

[11] Gaboardi, M. and Rogers, R. (2018). Local private hypothesis testing: chi-square tests. In *Proc. 35th ICML, Stockholm, Sweden, 2018* **80** 1626–1635.

[12] Ghosh, A., Roughgarden, T. and Sundararajan, M. (2012). Universally utility-maximizing privacy mechanisms. *SIAM Journal on Computing* **41** 1673–1693. MR3029267

[13] Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* **1** 175–194. MR0400513

[14] Hsu, J., Khanna, S. and Roth, A. (2012). Distributed private heavy hitters. In *International Colloquium on Automata, Languages, and Programming* 461–472. Springer. MR2995330

[15] Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation*. Springer.

[16] Kairouz, P., Bonawitz, K. and Ramage, D. (2016). Discrete distribution estimation under local privacy. In *Proc. 33rd Int. Conf. Machine Learning* **48** 2436–2444.

[17] Kairouz, P., Oh, S. and Viswanath, P. (2016). Extremal mechanisms for local differential privacy. *Jounral of Machine Learning Research* **17** 1–51. MR3491111

[18] Kamath, S., Orlitsky, A., Pichapati, V. and Suresh, A. T. (2015). On learning distributions from their samples. *Jounral of Machine Learning Research: Workshop and Conference Proceedings* **40** 1–35.

[19] Le Cam, L. (2012). *Asymptotic Methods in Statistical Decision Theory*. Springer Science & Business Media. MR0856411

[20] Lehmann, E. L. and Casella, G. (2006). *Theory of Point Estimation*. Springer Science & Business Media. MR1639875

[21] Mishra, N. and Sandler, M. (2006). Privacy via pseudorandom sketches. In *Proceedings of the Twenty-Fifth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* 143–152. ACM.

[22] Smith, A. (2011). Privacy-preserving statistical estimation with optimal convergence rates. In *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing* 813–822. ACM. MR2932032

[23] Thakurta, A. G. and Smith, A. (2013). (Nearly) optimal algorithms for private online learning in full-information and bandit settings. In *Advances in Neural Information Processing Systems* 2733–2741.

[24] van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Univesity Press. MR1652247

[25] Wang, S., Huang, L., Wang, P., Nie, Y., Xu, H., Yang, W., Li, X. and Qiao, C. (2016). Mutual information optimally local private discrete distribution estimation. arXiv:1607.08025.

[26] Warner, S. L. (1965). Randomized response: A survey technique for elim-

inating evasive answer bias. *Journal of the American Statistical Association* **60** 63–69.

[27] Ye, M. and Barg, A. (2017). Asymptotically optimal private estimation under mean square loss. arXiv:1708.00059.

[28] Ye, M. and Barg, A. (2018). Optimal schemes for discrete distribution estimation under locally differential privacy. *IEEE Trans. Inform. Theory* **64** 5662–5676. MR3832328