Abstract—This work examines a novel question: how much randomness is needed to achieve local differential privacy (LDP)? A motivating scenario is providing multiple levels of privacy to multiple analysts, either for distribution or for heavy hitter estimation, using the same (randomized) output. We call this setting successive refinement of privacy, as it provides hierarchical access to the raw data with different privacy levels. For example, the same randomized output could enable one analyst to reconstruct the input, while another can only estimate the distribution subject to LDP requirements. This extends the classical Shannon (wiretap) security setting to local differential privacy. We provide (order-wise) tight characterizations of privacy-utility-randomness trade-offs in several cases distribution estimation, including the standard LDP setting under a randomness constraint. We also provide a non-trivial privacy mechanism for multi-level privacy. Furthermore, we show that we cannot reuse random keys over time while preserving privacy of each user.

Index Terms—Data privacy, local differential privacy, privacy and learning, privacy-utility-randomness trade-off, multiple levels of privacy.

I. INTRODUCTION

DIFFERENTIAL privacy [1] – a cryptographically motivated notion of privacy – has recently emerged as the gold standard in privacy-preserving data analysis. Privacy is provided by guaranteeing that the participation of a single person in a dataset does not change the probability of any outcome by much; this is ensured by randomness – either by adding noise to (or randomizing) the raw data itself or to a function or statistic computed directly on the data. If the randomization is large enough relative to the change caused by a single person’s data, then their participation is indistinguishable, and privacy is attained. An underlying assumption in the body of work on differential privacy has long been that an underlying assumption in the vast majority of the literature has focused on achieving better privacy-utility trade-offs – see, for example, [2], [3] for surveys. In this article, we ask: how much randomness do we need to achieve a desired level of privacy and utility, and study privacy-utility-randomness trade-offs instead. Answering this question both contributes to our theoretical understanding, and also could support specific emerging applications that we discuss later in the section.

We consider local differential privacy (LDP) – a privacy model that has recently seen use in industrial applications, [4], [RAPPOR], [5]. Here, an untrusted analyst acquires already-privatized pieces of information from a number of users, and aggregates them into a statistic or a machine learning model. Concretely, there are $n$ users who observe i.i.d. inputs $X_1, X_2, \ldots, X_n$ (user $i$ observes $X_i$) from a finite alphabet $\mathcal{X}$ of size $k$, where each $X_i$ is distributed according to a probability distribution $p$. Each user has a certain amount of randomness, measured in Shannon entropy, to randomize her input, that she then publicly shares. Our general setup also includes $d$ analysts who would like to use the users’ public outputs to estimate $p$, each at a different level of privacy $\epsilon_1, \ldots, \epsilon_d$, where smaller $\epsilon$ means higher privacy. Each analyst may or may not share some common randomness with the users. We call this general setup successive refinement of privacy, in which each user shares a public output with highest privacy level. Then, each analyst uses a shared random key to partially undo the randomization of the public output to get less privacy and higher utility.

This general formulation includes several interesting special cases, for which we study the trade-offs between privacy, utility, and randomness. These are:

(i) There is a single analyst ($d = 1$), who shares no randomness with the users and estimates $p$ with privacy level $\epsilon$. This setting directly generalizes the classical setup of LDP to the case of limited randomness.

(ii) There are two analysts ($d = 2$), who observe the same public outputs from the users; the first analyst who shares common randomness with the users has permission to perfectly recover the original inputs (i.e., privacy level $\epsilon_1 \to \infty$), while the second analyst who shares no randomness with the users estimates $p$ with privacy level $\epsilon_2$. This setting is an adaptation of the classical perfect secrecy setup of Shannon [6] to the differential privacy world. In Shannon’s setup, Alice (users) wants to send a secret to Bob (the first analyst), which must remain perfectly private from Eve (the second analyst); whereas, in our setting, instead of complete independence, we only want that the secret remains hidden from Eve in the sense of differential privacy. We call this setup private-recoverability.
(iii) There are $d > 1$ analysts, who share some common randomness with the users. Analyst $i$ would like to estimate $p_i$ with privacy level $\epsilon_i$, where $\epsilon_1 > \ldots > \epsilon_d$.

A. Motivation

In general, designing private mechanisms with a small amount of randomness can be translated into communication efficiency and/or storage efficiency. For instance, when there are multiple privacy levels, each user needs to send additional information to some analysts, that is a function of the randomness used in the mechanism. Hence, using a smaller amount of randomness implies delivering a smaller number of bits to each analyst.

The private-recoverability setup ($d = 2$) can be useful in applications such as census surveys, [7], that collect large amounts of data and are prohibitively expensive to repeat. Using our approach, we can store the randomized data on a public database (second analyst) without compromising the privacy of individuals; we can also give to the first analyst (e.g., the government, who may wish to exactly calculate the population count, or verify the validity of census results) a secret key, that can be used to “de-randomize” the publicly stored data and perfectly reconstruct the user inputs. An alternative approach would be to store the data twice (once randomized in a public database and once in a secure government database), which would incur an additional storage cost, as also shown in Section IV. Another alternative would be to use a cryptographic scheme to encode the user inputs; in this case, the resulting outputs may not allow public use in an efficient manner.

The multi-level privacy $d > 1$ illustrates a new technical capability of hierarchical access to the raw data that might inspire and support a variety of applications. For example, given data collected from a fleet of autonomous cars, we could imagine different privacy access levels provided to the car manufacturer itself, to police departments, to applications interested in online traffic regulation, to applications interested in long-term traffic predictions or road planning. Essentially, this capability enables providing the desired utility needed for each application while maintaining the maximum possible amount of privacy.

B. Contributions

Our contributions are as follows.

- For the single analyst case ($d = 1$), we characterize the trade-off between randomness and utility for a fixed privacy level $\epsilon$, by proving an information-theoretic lower bound and a matching upper bound for a minimax private estimation problem.

- For private-recoverability ($d = 2$), we derive an information-theoretic lower bound on the minimum randomness required to achieve it, and prove that the Hadamard scheme proposed in [8] is order optimal. We also show that we cannot reuse random keys over time while preserving privacy of each user. Hence, to preserve privacy of $T$ samples, any $\epsilon$-DP mechanism has to use an amount of randomness equal to $T$ times the amount of randomness used for a single data sample. We also extend this result to estimating heavy hitters.

- In the multi-level privacy ($d > 1$) setting, a trivial scheme is to use the $d = 1$ scheme multiple times, separately for each analyst. We propose instead a non-trivial scheme that uses a smaller amount of randomness with no sacrifice in utility. Our scheme publicly announces the users’ outputs, and allows each analyst to remove an appropriate amount of (shared) randomness with the help of an associated key. This approach enables efficient hierarchical access to the data (for example, when analysts have different levels of authorized access).

Overall, our investigation into privacy-utility-randomness trade-offs for LDP yields (optimal) privacy mechanisms that use randomness more economically. These include new guarantees for existing schemes such as the Hadamard mechanism, as well as new multi-user and multi-level mechanisms that allow for hierarchically private data access.

C. Related Work

To the best of our knowledge, the role of limited randomness has not been previously explored either in the context of local or global differential privacy. In this work, we consider local differential privacy in the context of distribution estimation and heavy hitter estimation for reasons of simplicity.

Popular local differentially private mechanisms for distribution estimation include RAPPOR [4], randomized response (RR) [10], subset selection (SS) [11], [12], and the Hadamard response (HR) [8]. The randomized response mechanism is known to be order optimal in the low privacy regime, and the RAPPOR scheme in the high privacy regimes [13], [14]. Subset selection and the Hadamard mechanisms are order optimal in utility for all privacy regimes; additionally, the Hadamard mechanism has the advantage of communication and computational efficiency for all privacy regimes [8]. We build on this extensive literature, and show that the Hadamard mechanism is also near-optimal in terms of the amount of randomness used.

Heavy hitter estimation under local differential privacy has been studied in [15]–[19], again with unrestricted randomness.

1We can assume, without loss of generality, that $\epsilon_j > \epsilon_{j+1}$, $\forall j \in [d-1]$; otherwise, we can group the equal $\epsilon_j$’s together and the corresponding analysts can use the same privatized data that the users share with them.

2In principle, we could use homomorphic encryption that allows to compute a function on the encrypted data without decrypting it explicitly; however, such encryption schemes are computationally inefficient and expensive to deploy.

3Except for a notable exception of [9], which showed that imperfect source of randomness allows efficient protocols with global differential privacy. This is different from our problem, where our goal is to quantify the amount of randomness required (measured in terms of Shannon entropy) in local differential privacy and give privacy-utility-randomness trade-offs.
Our work adds to this line of work by showing that the Hadamard mechanism is capable of achieving order-optimal accuracy for heavy hitter estimation while using an order-optimal amount of randomness.

Local differential privacy in a multi-user setting where the users and the server may have some shared randomness has also been looked at in prior work – see [15], [20], [21] among others. These works however investigate other orthogonal aspects of such multi-user protocols. Local differentially private mechanisms with bounded communication have also been studied by [20]; in their setup, multiple agents transmit their data in a locally private manner to an aggregator, and communication is measured by the number of bits transmitted by each user. They consider both private and public coin mechanisms, and show that the Hadamard mechanism is near optimal in terms of communication for both distribution and heavy-hitter estimation; however, unlike ours, their mechanisms do not impose any randomness constraints.

Our results in the multiple analyst setting are also related to privacy amplification by stochastic postprocessing [22] – which analyzes the privacy risk achieved by applying a (stochastic) post-processing mechanism to the output of a differentially private algorithm. While these methods might also be used to provide multi-level privacy to multiple analysts, our work is different from [22] in the following aspect. First, their privacy amplification methodology does not apply to pure DP and applies instead to approximate DP, while our work focuses on pure DP. Second, the work in [22] does not include a randomness constraint, and finally, a closer look at their mechanism reveals that it does not use the optimal amount of randomness.

Finally, a line of work on locally differentially private estimation considers the case when the inputs comprise of i.i.d. samples from the same distribution. References [23], [24] derive lower and upper bounds for estimation under LDP in this setting – their work considers that all users observe i.i.d. samples from the same distribution, and the goal for each user is to preserve privacy of its raw sample. Our work is different from this setting in that we focus on designing private mechanisms with finite randomness.

D. Article Organization

Section II formally defines LDP mechanisms under randomness constraints and presents the distribution and heavy hitter estimation problem formulations. Section III states our main results for the single-level privacy, private-recoverability, and multi-level privacy settings. Section IV presents numerical evaluations on the effect of parameters such as \( n, \epsilon, d \) on the estimation error and the required randomness. Section V derives an information-theoretic lower bound and an upper bound (achievability scheme) on the minimax risk estimation under randomness and privacy constraints for a single analyst. Section VI proposes a new LDP mechanism for the multi-level privacy \( d > 1 \). Section VII presents the necessary and sufficient conditions on the randomness to design an \( (\epsilon, R) \)-LDP mechanism with input recoverability requirement. Section VIII introduces the necessary and sufficient conditions on the randomness to preserve privacy of a sequence of samples per user. Omitted proofs from this article appear in appendices, which are provided as part of the supplementary material.
Note that the output $Y$ is a function of $(X, U)$. Therefore, we have $U_{i,t} \cap U_{j,t} = \phi$ for $i' \neq j$, since there is only one output for each input. In addition, if we want (3) to satisfy the privacy condition (1), we also have $\cup_{j \in [k]} U_{i,t} = U$ for each $x \in X$. We will leverage this representation of randomness in LDP mechanisms to design multi-level privacy mechanisms. Figure 4 shows an example of designing a private mechanism with binary inputs $X = \{0, 1\}$, binary random keys $U = \{0, 1\}$, and binary outputs $Y = \{0, 1\}$. In this example, we can represent the output of the mechanism as a function of $(X, U)$ by $Y = X \oplus U$, where $\oplus$ denotes the XOR operation. If the random key $U$ is drawn from a distribution $q = \left[\frac{1}{2}, \frac{1}{2}\right]$, then it is easy to show that the mechanism is $\epsilon$-LDP.

**C. Problem Formulation**

We consider $n$ users who observe i.i.d. inputs $X_1, X_2, \ldots, X_n$ (user $i$ observes input $X_i$), drawn from an unknown discrete distribution $p \in \Delta_k$, where $\Delta_k = \{p \in \mathbb{R}^k | \sum_{j=1}^k p_j = 1, p_j \geq 0, \forall j \in [k]\}$ is the probability simplex over $X$. The $i$'th user has a random key $U_i$ with $H(U_i) \leq R$; we assume that $U^n = (U_1, \ldots, U_n)$ are independent random variables, unless otherwise stated. The $i$'th user generates (and publicly shares) an output $Y_i$, using an $(\epsilon, R)$-LDP mechanism $Q_i$ and her random key $U_i$. The output $Y_i$ has a marginal distribution given by

$$M_i(y|p) = \sum_{x \in X} Q_i(y|x)p_x \quad \forall y \in \mathcal{Y}_i,$$

where $\mathcal{X}$ and $\mathcal{Y}_i$ are the input and output alphabets. We also have $d$ analysts who want to use the users’ public outputs $Y^n = [Y_1, \ldots, Y_n]$ to estimate $p$, each at a different level of privacy $\epsilon_1 > \ldots > \epsilon_d$. The system model is shown in Figure 1.

**Risk Minimization:** For simplicity of exposition, consider for now a single analyst, and let $\hat{p} = [\hat{p}_1, \ldots, \hat{p}_k]$ denote the analyst’s estimator (this is a function $\hat{p} : Y^n \rightarrow \mathbb{R}^k$ that maps the outputs $Y^n$ to a distribution in the simplex $\Delta_k$).\(^5\) For given private mechanisms $Q^n = [Q_1, \ldots, Q_n]$, the estimator $\hat{p}$ is obtained by solving the problem

$$r_{\epsilon,R,n,k}^\ell(Y^n) = \inf_{\hat{p}} \sup_{p \in \Delta_k} \mathbb{E}[\ell(\hat{p}(Y^n), p)],$$

where $r_{\epsilon,R,n,k}^\ell$ is the minimax risk, the expectation is taken over the randomness in the outputs $Y^n = [Y_1, \ldots, Y_n]$ with

\(^4\)Otherwise we can distinguish inputs causing $\epsilon \rightarrow \infty$.

\(^5\)Observe that it is sufficient to consider a deterministic estimator $\hat{p}$, since for any randomized estimator, there exists a deterministic estimator that dominates the performance of the randomized one.

$Y_i \sim M_i$, and $\ell : \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}_+$ is a loss function that measures the distance between two distributions in $\Delta_k$. Unless otherwise stated, we adopt as loss function the $1$-norm, namely $\ell = \|\cdot\|_1$ and the squared $2$-norm, namely $\ell = \|\cdot\|_2^2$. Our task is to design private mechanisms $Q_1, \ldots, Q_n$ that minimize the minimax risk estimation, namely,

$$r_{\epsilon,R,n,k}^\ell = \inf_{Q \in \mathcal{Q}(\epsilon,R)} \inf_{p \in \Delta_k} \mathbb{E}[\ell(\hat{p}(Y^n), p)],$$

where $\mathcal{Q}(\epsilon,R)$ denotes the set of mechanisms that satisfy $(\epsilon, R)$-LDP. Observe that when $R \rightarrow \infty$, the problem (6) is reduced to the standard LDP distribution estimation studied previously in [8], [11], [13], [25]. The difference in the formulation in (6) is the randomness constraint.

**LDP heavy hitter estimation:** In heavy hitter estimation, the input samples $X^n = [X_1, \ldots, X_n]$ do not have an associated distribution. Furthermore, the analyst is interested in estimating the frequency of each element $x \in X$ with the infinity norm being the loss function (i.e., $\ell = \ell_\infty$). Frequency of each element $x \in X$ is defined by $f(x) = \frac{\sum_{t \in T} 1(x_{i,t} = x)}{n}$. We then want to calculate

$$\ell_{\text{hh},R,n,k}^{\ell_\infty} = \inf_{Q \in \mathcal{Q}(\epsilon,R)} \inf_{\hat{p}} \mathbb{E} \left[ \max_{x \in X} |\hat{p}_k(x^n) - f(x)| \right],$$

where the expectation is taken over the randomness in the outputs $Y^n = [Y_1, \ldots, Y_n]$ and $\hat{p}$ denotes the estimator of the analyst. Note, again, that in this case we do not make any distributional assumptions on $X_1, \ldots, X_n$.

**Multi-level privacy:** Consider now the general case of $d$ analysts operating at a different level of privacy $\epsilon_1 > \ldots > \epsilon_d$. All analysts observe the users’ public outputs $Y^n$; additionally, analyst $j$ may also observe some side information on the user randomness. The question we ask is: what is the minimum amount of randomness $U$ per user required to maintain the privacy of each user while achieving the minimum risk estimation for each analyst?

**Sequence of distribution (or heavy hitter) estimation:** We assume that each user $i$ has a random key $U_i$ to preserve the privacy of a sequence of $T$ independent samples $X_i^{(1)}, \ldots, X_i^{(T)}$, where the $i$'th samples for $t \in [T]$ at all users are drawn i.i.d. from an unknown distribution $p_i^{(t)}$.\(^6\) At time $t$, the $i$'th user generates an output $Y_i^{(t)}$ that may be a function of the random key $U_i$ and all input samples $[X_i^{(m)}]_{m=1}^t$. Each of the $d$ analysts uses the outputs $Y_i^{(t)}$, $i \in [n]$, $t \in [T]$ to estimate $T$ distributions $p_i^{(1)}, \ldots, p_i^{(T)}$ (or estimate the heavy hitters).

A private mechanism $Q$ with a sequence of inputs $X_T = (X^{(1)}, \ldots, X^{(T)})$ and a sequence of outputs $Y_T = (Y^{(1)}, \ldots, Y^{(T)})$ is said to satisfy $\epsilon$-DP; if for every neighboring databases $x, x' \in \mathcal{X}^T$, we have

$$\sup_{y \in [T]} \frac{Q(y|x)}{Q(y|x')} \leq \exp(\epsilon),$$

where $Q(y|x) = \Pr[Y_T = y|X_T = x]$; and we say that two databases, $x = (x^{(1)}, \ldots, x^{(T)})$ and $x' = (x^{(1)}, \ldots, x^{(T)}) \in \mathcal{X}^T \times \mathcal{X}^T$ do not have an associated distribution.

\(^6\)As mentioned earlier, for heavy hitter estimation, the samples $X_i^{(1)}, \ldots, X_i^{(T)}$ do not have an associated distribution.
\(X^T\) are neighboring, if there exists an index \(t \in [T]\), such that \(x^{(t)} \neq x^{(l)}\) and \(x^{(t)} = x^{(l)}\) for \(l \neq t\). Observe that when \(T = 1\), the definition of \(\epsilon\)-DP in (7) coincides with the definition of \(\epsilon\)-LDP in (1). We are interested in the question: Is there a private mechanism that uses a smaller amount of randomness than \(T\) times the amount of randomness used for a single data sample? In other words, can we perhaps reuse the randomness over time while preserving privacy?

### III. MAIN RESULTS

This section formally presents our main results. First, we characterize the minimax risk estimation under randomness and privacy constraints in Theorems 1 and 2 for single-level privacy \((d = 1)\). Then, we propose in Theorem 3 a new lower bound on the minimax risk of samples under a recoverability constraint. Formulation of a non-convex optimization problem to bound the minimax risk estimation without randomness by \(r\) \(\epsilon\), factors. In Remark 1, we derive a lower bound on the minimax risk under privacy and randomness constraints, \((\epsilon, \beta)\), our lower bound from Theorem 1 gives \(r\) \(\epsilon\) \(\beta\) factors.

### A. Single-Level Privacy, \(d = 1\)

We here study the fundamental trade-off between randomness and utility for a fixed privacy level \(\epsilon\). In the following theorem, we derive a lower bound on the minimax risk estimation \(r_{\epsilon,R,n,k}\) and \(r_{\epsilon,R,n,k}^{(1)}\) defined in (6).

**Theorem 1:** For every \(\epsilon, R \geq 0\) and \(n, k \in \mathbb{N}\), the minimax risk under \(\ell_2\)-norm loss is bounded by

\[
r_{\epsilon,R,n,k}^{(2)} \geq \tau = \begin{cases} \frac{k(\epsilon^2 + 1)^2}{16n^2\epsilon^2(\epsilon^2 - 1)^2} & \text{if } R \geq H_2\left(\frac{\epsilon^2}{\epsilon^2 + 1}\right), \\ \frac{ke^2}{16n^2\epsilon^2(\epsilon^2 - 1)^2} & \text{if } R < H_2\left(\frac{\epsilon^2}{\epsilon^2 + 1}\right). \end{cases}
\]

where \(pR \leq 0.5\) is the inverse of the binary entropy function \(pR = H_2^{-1}(R)\). The minimax risk under 1-norm loss is bounded by \(r_{\epsilon,R,n,k}^{(1)} \geq \sqrt{\kappa \tau/8}\).

The main contribution in our proof (see Section V-A) is a formulation of a non-convex optimization problem to bound the minimax risk under privacy and randomness constraints, and obtaining a tight bound on its solution for every value of privacy level \(\epsilon\) and randomness \(R\).

**Remark 1:** In [11], the authors derive the following lower bound on the minimax risk estimation without randomness constraints \((R \to \infty)\)

\[
r_{\epsilon,\infty,n,k}^{(2)} \geq \frac{k(\epsilon^2 + 1)^2}{512n^2(\epsilon^2 - 1)^2} \quad \text{for } \epsilon^2 < 3,
\]

\[
\geq \frac{6k(\epsilon^2 + 1)^2}{512n^2\epsilon^2(\epsilon^2 - 1)^2} \quad \text{for } \epsilon^2 \geq 3.
\]

For \(\epsilon = \mathcal{O}(1)\) and \(R \geq H_2\left(\frac{\epsilon^2}{\epsilon^2 + 1}\right)\) (which includes \(R \to \infty\) as well), our lower bound from Theorem 1 gives \(r_{\epsilon,R,n,k}^{(2)} = \Omega(\frac{1}{\epsilon^2 + 1})\), which coincides with (9). However, our lower bound is tighter for all values of \(\epsilon \in [0, \infty)\) with smaller constant factors.

We next show that there exists an achievable scheme for all values of \(\epsilon, R \geq 0\) that matches (up to a constant factor) the lower bound given in Theorem 1 for \(\epsilon = \mathcal{O}(1)\) and \(R \geq 0\).

**Theorem 2:** For any \(\epsilon, R \geq 0\), there exists an \((\epsilon, R)\)-LDP mechanisms \(Q_1, \ldots, Q_n\) and an estimator \(\hat{p}\) such that the error \(\mathcal{E} := \sup_{p \in \Delta_k} \mathbb{E}[\|\hat{p}(Y^n) - p\|_2^2]\) is bounded by

\[
\mathcal{E} \leq \eta = \begin{cases} \frac{2k(\epsilon^2 + 1)^2}{n(\epsilon^2 - 1)^2} & \text{if } R \geq H_2\left(\frac{\epsilon^2}{\epsilon^2 + 1}\right), \\ \frac{2k^2n}{n^2(\epsilon^2 - 1)^2} & \text{if } R < H_2\left(\frac{\epsilon^2}{\epsilon^2 + 1}\right). \end{cases}
\]

The error under \(\ell_1\)-norm loss is bounded by \(\sup_{p \in \Delta_k} \mathbb{E}[\|\hat{p}(Y^n) - p\|_1] \leq \sqrt{k\eta}\).

We prove Theorem 2 constructively in Section V-B, by adapting the Hadamard response scheme given in [20] to our setting of limited randomness. Note that the value \(H_2\left(\frac{\epsilon^2}{\epsilon^2 + 1}\right)\) in both the lower and upper bounds is an exact threshold for randomness that determines the value of the minimax risk. Furthermore, we can see that the multiplicative gap between the lower bound presented in Theorem 1 and the achievable scheme in Theorem 2 is exactly \(32(\epsilon + 1)^2\) for all randomness regimes. We plot this gap as a function of \(\epsilon\) in Figure 3. We can see that this gap becomes smaller with more privacy (smaller \(\epsilon\)). For example, the Gap is equal to 52 for \(\epsilon = 0.5\). Theorems 1 and 2 together imply the following characterization for \(r_{\epsilon,R,n,k}^{(2)}\) and \(r_{\epsilon,R,n,k}^{(1)}\) for the case when \(\epsilon = \mathcal{O}(1)\).

**Corollary 1:** For \(\epsilon = \mathcal{O}(1)\) and \(R \geq 0\), we have

\[
r_{\epsilon,R,n,k}^{(2)} = \begin{cases} \Theta\left(\frac{k}{n\epsilon^2}\right) & \text{if } R \geq H_2\left(\frac{\epsilon^2}{\epsilon^2 + 1}\right), \\ \Theta\left(\frac{k}{n^2\epsilon^2}\right) & \text{if } R < H_2\left(\frac{\epsilon^2}{\epsilon^2 + 1}\right). \end{cases}
\]

We next provide a comparison between well-known mechanisms from randomness perspective. Table I describe the amount of randomness required to implement different \(\epsilon\)-LD mechanism: RAPPOR [4], Randomized Response (RR) [10], Hadamard Response (HR) [8], and Binary Hadamard (BH) [20].

Observe that all private mechanisms are order optimal in the high privacy regime except for the RR scheme. However, only the BH scheme uses the smallest amount of randomness \(R = H_2\left(\frac{\epsilon^2}{\epsilon^2 + 1}\right)\) per user, while the other mechanisms require a larger amount of randomness. Table I considers only the regime of randomness \(R \geq H_2\left(\frac{\epsilon^2}{\epsilon^2 + 1}\right)\), since the privacy-utility trade-off when the amount of randomness \(R < H_2\left(\frac{\epsilon^2}{\epsilon^2 + 1}\right)\) has not been studied before. Corollary 1 characterizes the privacy-utility trade-offs for all regions of randomness \(R\).
Remark 2: Observe that when \( R < H_2(\frac{\epsilon^2}{e^2+1} \) ), there exists a trade-off between \( R \) and \( \epsilon^2 \) \( \epsilon \) as \( R \) increases, \( \epsilon^2 \) decreases proportionally to \( 1/e^2 \). However, when \( R \geq H_2(\frac{\epsilon^2}{e^2+1} \), the minimax risk is not affected by \( R \). Hence, \( R = H_2(\frac{\epsilon^2}{e^2+1} \) is a critical point that defines the minimum amount of randomness required for each user to generate an \( \epsilon \)-LDP mechanism, while achieving the optimal utility at the analyst.

Remark 3: Corollary 1 also characterizes the number of users \( n \) (sample complexity) required to estimate the distribution \( p \) with estimation error at most \( \alpha \) for given privacy level \( \epsilon \) and randomness \( R \) bits per user is (where \( k \) is the input alphabet size):

\[
\begin{align*}
n = \left\{ \begin{array}{ll}
\Theta\left(\frac{k^2}{\alpha^2} \right) & \text{if } R \geq H_2\left(\frac{\epsilon^2}{e^2+1} \right), \\
\Theta\left(\frac{k}{\alpha} \right) & \text{if } R < H_2\left(\frac{\epsilon^2}{e^2+1} \right).
\end{array} \right.
\end{align*}
\]

A remark analogous to Remark 2 also holds here.

**B. Multi-Level Privacy, \( d > 1 \)**

Here, we study the case of \( d \) different analysts, with privacy levels \( \epsilon_1 > \cdots > \epsilon_d \), and \( \epsilon_j = O(1) \) for \( j \in [d] \) (See Section I-A for the motivation of this setup). A trivial scheme is to use the \( d = 1 \) scheme multiple times, separately for each analyst: each user \( i \in [n] \) generates \( d \) samples \( (Y_i^1, \ldots, Y_i^d) \) from its input sample \( X_i \). The \( j \)th sample \( Y_i^j \) is delivered privately to the \( j \)th analyst. Note that the \( j \)th sample must be generated from an \( \epsilon_j \)-LDP. It then follows from Corollary 1 that the minimum risk for the \( j \)th analyst is given by \( \epsilon_j \), which requires each user to have \( R_j \geq H_2\left(\frac{\epsilon_j^2}{e^2+1} \right) \) bits of randomness, and results in a total amount of randomness

\[
R_{\text{total}} = \sum_{j=1}^{d} H_2\left(\frac{\epsilon_j^2}{e^2+1} \right).
\]

We propose a new solution for this problem, in which each user generates a single output that is publicly accessible by all analysts; each analyst is given a part of the random key that was used to privatize the data, and leverages this key to reduce the perturbation of the public output. The next theorem is proved in Section VI.

**Theorem 3:** There exists a private mechanism using a total amount of randomness given by \( R_{\text{total}} = \sum_{j=1}^{d} H_2(q_j) \), such that the \( j \)th analyst achieves the minimum risk estimation \( \epsilon_j^2 \) \( \epsilon_j \) of \( \epsilon_j \), while preserving privacy of each user with privacy level \( \epsilon_j \) for \( j \in [d] \). Here, for every \( j \in [d] \), \( q_j \) is defined as follows (where \( z_j = \frac{1}{e^2+1} \)):

\[
q_j = \begin{cases} 
\frac{z_j}{2z_j-1} & \text{if } j = 1, \\
\frac{z_{j-1}}{2z_{j-1}} & \text{if } j > 1.
\end{cases}
\]

Our results demonstrate that the proposed scheme in Theorem 3 achieves exactly the same privacy and minimax risk as the trivial scheme, but with a much lower amount of randomness (See Eq. (40) in Section VI for more details) as we elaborate in the next remark. The reason is that by doing the hierarchical randomization, each analyst can recover the same output as in the trivial scheme using the shared private key. Hence, we get the same privacy and minimax risk as the trivial scheme.

**Remark 4:** Note that \( z_j > 1 \), \( \epsilon_j > \epsilon_1 \), and \( \epsilon_j = O(1) \) for \( j \in [d] \) as \( \epsilon_j = O(1) \). Moreover, we also have \( z_j = 1/(\epsilon_j + 1) < 0.5 \) for all \( j \in [d] \). As a result, we can show that for \( j > 1 \), we have

\[
q_j = \frac{z_j - z_{j-1}}{1 - 2z_{j-1}} = z_j - \frac{z_{j-1}(1 - 2z_j)}{1 - 2z_{j-1}} < z_j.
\]

Hence, we get that \( H_2(q_j) < H_2(z_j) \) holds for all \( j > 1 \). Therefore, our proposed scheme uses a strictly smaller amount of randomness than the trivial scheme.

**C. Private-Recoverability, \( d = 2 \)**

We here consider a legitimate analyst with permission to access the data \( (X_i)_{i=1}^n \), i.e., \( \epsilon_1 \to \infty \), and an untrusted analyst with privacy level \( \epsilon_2 < \infty \). The ith user uses a random private key \( U_i \) and her mechanism \( Q_i \) to generate an output \( Y_i \) that is publicly accessible by both analysts.

**Definition 1 (LDP-Rec Mechanisms):** We say that a private mechanism \( Q \) is \( \epsilon \)-LDP-Rec, if it is an \( \epsilon \)-LDP mechanism and it is possible to recover the input \( X \) from output \( Y \) and the key \( U \).

We derive necessary and sufficient conditions on the random keys \( \{U_j\} \) and the mechanisms \( \{Q_j\} \), such that the legitimate analyst can recover \( X_i \) from observing \( U_i \) and \( Y_i \), while preserving privacy level \( \epsilon_2 \) against the untrusted analyst who does not have access to the keys.

We first consider a simplified setting as shown in Figure 4. Alice (an arbitrary user)\(^7\) has a sample \( X \in X \). Alice wants to send her sample \( X \) to Bob (the legitimate analyst) while keeping her sample \( X \) private against Eve (the untrusted analyst) with differential privacy level \( \epsilon \). Eve has access to the message between Alice and Bob. However, Alice has a random key \( U \) shared with Bob that Eve does not have access to. Let \( Y \) be

\(^7\)Since the input samples \( X_1, \ldots, X_n \) are i.i.d., and the random keys \( U_1, \ldots, U_n \) are independent random variables, it is sufficient to study the private-recoverable mechanism for any single user.
the output of the private mechanism \( Q \) used by Alice. The following theorem (which we prove in Section VII-A) provides necessary and sufficient conditions on the random key \( U \) and the privatized output \( Y \) to generate an \( \epsilon \)-LDP-Rec mechanism.

**Remark 5:** Observe that in the simplified model in Figure 4, we do not impose any assumptions on the input \( X \). Furthermore, we do not impose any assumptions about the task for Eve. Hence, our model and results in Theorem 4 are applicable to any task for Eve including distribution estimation, heavy hitter estimation, or learning from sample \( X \).

**Theorem 4:** Let \( Q \) be an \( \epsilon \)-LDP-Rec mechanism that uses a random key \( U \in \mathcal{U} \) and an input \( X \in \mathcal{X} \) to produce a privatized output \( Y \in \mathcal{Y} \). The following conditions are necessary and sufficient to allow recovery of \( X \) from \( (U, Y) \):

1. \( |U| \geq |Y| \geq |X| \).
2. The entropy of the random key must satisfy \( H(U) \geq H(U^{s^{\epsilon}}) \), where \( s^{\epsilon} = \arg \min \{|l|: H(U^{s^{\epsilon}}) \geq l \} \) for \( l = k^{\epsilon/(e-1)} \) and \( U^{s^{\epsilon}} \) is a random variable with support size equal to \( |X| = k \) and has the following distribution:

\[
q_{\min}^{s^{\epsilon}} = \left[1/t, \ldots, 1/t, e^{\epsilon}/t, \ldots, e^{\epsilon}/t\right],
\]

where \( t = (se^{\epsilon} + k - s) \), the first \( k - s \) terms are equal to \( 1/t \) and the remaining \( s \) terms are equal to \( e^{\epsilon}/t \).

We now discuss the effect of \( \epsilon \) on the structure of optimal distribution \( q_{\min}^{s^{\epsilon}} \) for \( U^{s^{\epsilon}} \): (i) When \( \epsilon \gg \log(k) \), the optimal \( s^{\epsilon} = 1 \), and the corresponding \( q_{\min}^{1} \) has its first \( k - 1 \) terms equal to \( 1/(e^{\epsilon} + k - 1) \) and the last term equal to \( e^{\epsilon}/(e^{\epsilon} + k - 1) \). This distribution is equivalent to the one used in the Randomized Response (RR) model proposed in [10]. (ii) When \( \epsilon \to 0 \), the optimal \( s^{\epsilon} \) is around \( k/2 \), and the corresponding \( q_{\min}^{k/2} \) has its first \( k/2 \) terms equal to \( 1/k(e^{\epsilon} + 1) \) and the remaining \( k/2 \) terms equal to \( e^{\epsilon}/k(e^{\epsilon} + 1) \). (iii) When \( \epsilon = 0 \), the distribution \( q_{\min}^{s^{\epsilon}} \) becomes uniform (irrespective of the value of \( s \)). Thus, when \( \epsilon < 0 \), the distribution \( q_{\min}^{s^{\epsilon}} \) approaches to the uniform distribution. On the other hand, when \( \epsilon \) increases, the distribution \( q_{\min}^{s^{\epsilon}} \) becomes skewed. It turns out that the minimum randomness required to generate an \( \epsilon \)-LDP-Rec mechanism for input recoverability is a non-increasing function of \( \epsilon \). In other words, more privacy requires more randomness.

**Remark 6:** Consider the cryptosystem introduced by Shannon in [6], where Alice wants to send a secure message \( X \) to Bob using a shared random key \( U \). Let \( Y \) be the encrypted message sent to Bob. Eve eavesdrops the channel between Alice and Bob and observes \( Y \). This cryptosystem achieves perfect secrecy if and only if \( I(X; Y) = 0 \). Shannon showed that perfect secrecy requires \( H(U) \geq H(X) \). Since the distribution of \( X \) is not known to any node (Alice, Bob, and Eve), this implies \( H(U) \geq \max_{X \in \mathcal{X}} H(X) = \log k \). We can easily verify that the \( \epsilon \)-LDP-Rec mechanism satisfies a cryptosystem with secrecy measure \( \max_{X \in \mathcal{X}} I(X; Y) \leq \epsilon \). Hence, a perfect secrecy system with unknown input distribution is a \( \epsilon \)-LDP-Rec mechanism, which is a special case of our problem. Moreover, the \( \epsilon \)-LDP-Rec mechanism with data recovery is a cryptosystem leaking an amount of information measured by \( \max_{X \in \mathcal{X}} I(X; Y) \leq \epsilon \).

Observe that Theorem 4 does not provide performance guarantees for Eve, it only guarantees privacy for Alice with respect to Eve, and recoverability for Bob. Hence, we can ask the question: does there exist an \( \epsilon \)-LDP-Rec mechanism using the smallest amount of randomness and guaranteeing the smallest error for distribution estimation or heavy hitter estimation for Eve (the untrusted analyst)? In the following theorem (which we prove in Section VII-B), we show that such a mechanism exists.

**Theorem 5:** The Hadamard Response mechanism from [8] satisfies private-recoverability, and is utility-wise order-optimal for distribution estimation and heavy hitter estimation while using an order-optimal amount of randomness.

**D. Sequence of Distribution (or Heavy Hitter) Estimation**

We again start from the setting in Figure 4, but with the modification that Alice (an arbitrary user) wants to send to Bob (a legitimate analyst) \( T \) independent samples \( X^T = (X^{(1)}, \ldots, X^{(T)}) \), where \( X^{(t)} \in \mathcal{X} \), while keeping them private against Eve (an untrusted analyst) with differential privacy level \( \epsilon \). Eve has access to the sequence of outputs \( Y^T = (Y^{(1)}, \ldots, Y^{(T)}) \) that Alice produces, but not to the random key \( U \) that Alice and Bob share. Note that each output \( Y^{(t)} \) might be a function of all input samples \( X^{(t)} = (X^{(1)}, \ldots, X^{(T)}) \) and the key \( U \). Furthermore, the output \( Y^{(t)} \) can take values from a set \( \mathcal{Y}^{(t)} \) that is not required to be the same as \( \mathcal{Y}^{(t)} \) for \( t \neq t' \). Let \( \mathcal{Y}^{(T)} = \mathcal{Y}^{(1)} \times \cdots \times \mathcal{Y}^{(T)} \). The following theorem is proved in Section VIII.

We can define \( \epsilon \)-LDP-Rec mechanisms in the same way as we defined \( \epsilon \)-LDP-Rec mechanisms in Definition 1: A mechanism \( Q \) is \( \epsilon \)-DP-Rec, if it satisfies (7), and allows the recovery of input \( X \) from the output \( Y \) and the key \( U \).

**Theorem 6:** Let \( Q \) be an \( \epsilon \)-DP-Rec mechanism that uses a random key \( U \in \mathcal{U} \) and an input database \( X^T \in \mathcal{X}^T \) to create an output \( Y^T \in \mathcal{Y}^T \). The following conditions are necessary and sufficient to allow recovery of the input \( X^T \) from \( (U, Y^T) \):

1. \( |U| \geq |Y^T| \geq |X^T| \).
2. The entropy of the random key must satisfy \( H(U) \geq \)
In Figure 5, we plot the estimation error for the performance of the estimator presented in Theorem 2 for a \( R \) min.

In Figure 7, we compare our privacy scheme proposed in Theorem 3 and the trivial scheme for two privacy levels \( \epsilon_1 = 1 \) and \( \epsilon_2 = 0.01:1 \).

\[
T \min_{\epsilon \in \{1, 2\}} H(U_{\min}^\epsilon), \text{ where } U_{\min}^\epsilon \text{ is the same random variable with support size } |X| = k, \text{ as defined in Theorem 4.}
\]

Theorem 6 shows that the minimum amount of randomness required to preserve privacy of \( T \) samples is equal to \( T \) times the amount of randomness required to preserve privacy of a single sample. That is, for \( \epsilon \)-DP-Rec, it is optimal to use an \( \epsilon \)-LDP-Rec mechanism \( T \) times.

Remark 7: Observe that Theorem 6 is applicable in a \( n \)-user setting (by setting \( T = n \)), where user \( i \) has a single sample \( X^{(i)} \), and all users have access to a shared random key \( U \). So we have that shared randomness among users does not help in reducing the overall required amount of randomness.

**IV. NUMERICAL EVALUATION**

In this section, we numerically validate our theoretical results through simulation.

**Single-level privacy:** In this part, we investigate the performance of the estimator presented in Theorem 2 for a single-level privacy. Each point is obtained by averaging over 20 runs. In Figure 5, we plot the estimation error for the \( \ell = \ell_1 \) loss function (\( \|p - \hat{p}(Y^n)\|_1 \)) for estimating a discrete distribution \( p \in \Delta_k \). The input size is \( k = 1000 \), the number of users is \( n \in [10^5:10^6] \), and the privacy level is \( \epsilon = 1 \) for two values of randomness \( R \in [0.7, 1] \) bits per user. The input samples are drawn from a Geometric distribution with parameter \( q = 0.8 \) (Geo(0.8)), in which \( p_i = C q^{i-1} (1 - q) \) for \( i \in [k] \), where \( C \) is a normalization term. Figure 5 shows that the number of users required to achieve a certain estimation error increases as the amount of randomness per user decreases. For instance, to achieve an \( \ell_1 \)-error equal to 1.4, we need \( n \approx 150,000 \) users if \( R = 1 \) bits per user, while we need \( n \approx 850,000 \) users if \( R = 0.7 \) bits per user.

Figure 6 depicts the \( \ell_1 \) estimation error as a function of the privacy level \( \epsilon \) for input size \( k = 1000 \) and number of users \( n = 500000 \) for two different values of randomness \( R \in \{1, 0.6\} \) bits per user. As we discussed in Theorem 1, for each privacy level \( \epsilon \), there is a critical point of randomness \( R = H(\epsilon^\epsilon/(\epsilon^\epsilon + 1)) \). When each user has \( R < H(\epsilon^\epsilon/(\epsilon^\epsilon + 1)) \) bits of randomness, then the \( \ell_1 \) estimation loss increases as the randomness \( R \) decreases. While when each user has \( R \geq H(\epsilon^\epsilon/(\epsilon^\epsilon + 1)) \) bits of randomness, the estimation error is not affected by the amount of randomness \( R \). In Figure 6, we find that the \( \ell_1 \) error depends on the randomness \( R \) for all \( \epsilon < 0.8 \), since we have \( R = 0.9 < H(\epsilon^\epsilon/(\epsilon^\epsilon + 1)) \) for all \( \epsilon < 0.8 \).

**Multi-level privacy:** Figure 7 and Figure 8 compare our proposed scheme in Theorem 3 with the trivial scheme with respect to the total amount of randomness used. In the trivial scheme, each user generates \( d \) different privatized samples, one for each analyst. In Figure 7 we consider two privacy levels \( \epsilon_1 = 1 \) and \( \epsilon_2 \leq \epsilon_1 \). We find that when \( \epsilon_1 - \epsilon_2 \) is small, then the trivial scheme requires approximately twice the total amount of randomness used in our scheme. However, when \( \epsilon_1 - \epsilon_2 \) is large, then our scheme and the trivial scheme use similar amounts of randomness. In Figure 8, we consider \( d \in [1:10] \), \( \epsilon_1 = 2 \) and \( \epsilon_2 = \epsilon_1 - 0.1j \), for \( j \in \{2: \ldots , d\} \). We find that the gap between the amount of randomness used in our scheme and the trivial scheme increases with \( d \).

**Private-recoverability:** Observe that each user needs \( \log(k) \) bits to store her input sample \( X \in [k] \), since she does not know the distribution \( X \sim p \). In private-recoverability, we can recover \( X \) from observing \( Y \) and \( U \); hence, we only need to store \( U \). Figure 9 plots the number of bits required to store \( U \) (see Theorem 4) as a function of the privacy level \( \epsilon \) and different values of input size \( k \in [10, 100, 1000] \). The black lines represent the \( \log(k) \) bits required to store \( X \) (an additional secure copy). Note that the amount of bits needed to store \( U \) is strictly smaller than \( \log(k) \) for \( \epsilon > 0 \), and decreases as the privacy level \( \epsilon \) increases. Observe that the gain in Figure 9 is per user. Hence, the total amount of saving in storage would be considerable when the number of users is large and \( \epsilon > 0 \). For example, when \( \epsilon = 5 \), alphabet size \( k = 2, 4, 10 \), we get gain in efficiency \( \frac{\log(k) - H(U)}{\log(k)} \) of 94.2%, 91.4%, and 85% respectively.

**V. SINGLE-LEVEL PRIVACY (PROOFS OF THEOREM 1 AND THEOREM 2)**

**A. Lower Bound on the Minimax Risk Estimation Using the Assouad’s Method**

Now we prove the lower bound on the minimax risk given in Theorem 1 (see page 749). We first follow similar steps as in [11], [23] to reduce the minimax problem into multiple binary testing problems using Assouad’s method. We note
that [11], [23] do not consider a randomness constraint. Hence, we formulate an optimization problem to obtain a lower bound on the minimax risk estimation with a randomness constraint. Finding a tight bound on the solution of this problem is the main step in our proof. We also provide an alternative proof of Theorem 1 by using Fisher information, which leads to a tight bound for $\ell = \ell^2$ with smaller constant factors (see Appendix A in the supplementary material).

Let $|X| = k$ be the input alphabet size. Let $\{p^v\}$ be a set of distributions parameterized by $v = (v_1, \ldots, v_{k/2}) \in V = \{-1, 1\}^{k/2}$. The distribution $p^v = (p^v_1, \ldots, p^v_k)$ is given by:

$$p^v_j = \begin{cases} \frac{1}{2} + \frac{\delta}{2}v_j & \text{if } j \in \{1, \ldots, k/2\} \\ \frac{1}{2} - \frac{\delta}{2}v_{j+k/2} & \text{if } j \in \{k/2 + 1, \ldots, k\} \\ \end{cases}$$

where $0 \leq \delta \leq 1/k$ is a parameter that will be chosen later. Let $Y^n = (Y_1, \ldots, Y_n)$ and $\nu^n = \nu_1 \times \cdots \times \nu_n$. Following [23], for any loss function $\ell(\hat{p}, p) = \sum_{i=1}^k \phi(\hat{p}_i - p_i)$, where $\phi : \mathbb{R} \rightarrow \mathbb{R}_+$ is a symmetric function, we have

$$\ell(\hat{p}(y^n), p^v) = \sum_{j=1}^k \phi(\hat{p}_j(y^n) - p^v_j) \geq \phi(\delta) \sum_{j=1}^{k/2} \left( \text{sgn}(\hat{p}_j(y^n) - \frac{1}{k}) - v_j \right),$$

where $\text{sgn}(x) = 1$ if $x \geq 0$ and $\text{sgn}(x) = 0$ otherwise. Suppose that user $i$ chooses a private mechanism $Q_i \in Q_{(\epsilon,R)}$ that generates an output $Y_i \in Y_i$. Let $\nu_i^+$ be the output distribution on $Y_i$ for an input distribution $p^v_i$ on $X_i$ defined by

$$\nu_i^+(y) = \sum_{j=1}^k Q_i(y | X_i = j)p^v_i.$$  \hfill (17)

Let $\nu_i^{+\epsilon}$ and $\nu_i^{-\epsilon}$ denote the marginal distribution on $Y^n$ conditioned on $v_j = +1$ and $v_j = -1$, respectively, where

$$\nu_i^{\epsilon}(y^n) = \frac{1}{|V|} \sum_{\nu_{-\epsilon}y = +1} \prod_{i=1}^n \nu_i^+(y_i)$$

$$\nu_i^{-\epsilon}(y^n) = \frac{1}{|V|} \sum_{\nu_{-\epsilon}y = -1} \prod_{i=1}^n \nu_i^+(y_i).$$

Thus, the minimax risk can be bounded using the following lemma whose proof is presented in Appendix B in the supplementary material.

**Lemma 1:** For the family of distributions $\{p^v : v \in V = \{-1, 1\}^{k/2}\}$, and a loss function $\ell(\hat{p}, p) = \sum_{i=1}^k \phi(\hat{p}_i - p_i)$ defined above, we have

$$r^\ell_{\epsilon,R,n,k} \geq \phi(\delta) \frac{t^2}{2} \left(1 - \sqrt{\frac{2}{\pi e} \sup_{j \in \{k/2\}} \sup_{v : v_j = +1} \sup_{Q_i \in Q_{(\epsilon,R)}} D_{KL}(\nu_i^+(y^n) \| \nu_i^{-\epsilon}(y^n))} \right).$$

where $\phi(x) = x^2$, and for loss function $\ell = \ell_1$, we have $\phi(x) = |x|$.

8Observe that for loss function $\ell = \ell_2^2$, we have $\phi(x) = x^2$, and for loss function $\ell = \ell_1$, we have $\phi(x) = |x|$.

Fig. 9. Comparison between storage required for $X$ and a random key $U$, for input alphabet sizes $k \in \{10, 100, 1000\}$. The black lines represent log($k$).

Fix arbitrary $i \in [n], j \in [k/2]$ and $v \in V$. We have

$$D_{KL}(\nu_i^+(y^n) \| \nu_i^{-\epsilon}(y^n)) \leq \frac{|V|}{2} \sum_{j=1}^{k/2} (Q_i(y | j) - Q_i(y | j + k/2))^2 \sum_{j=1}^{k/2} (Q_i(y | j) - Q_i(y | j + k/2))^2 \leq 2|V| e^\epsilon.$$
this function (21) is obtained when the output of the mechanism $Q \in Q(\epsilon, R)$ is binary. Then, we obtain a tight bound numerically for the binary output.

**Proof of Lemma 2:** Without loss of generality assume that $Y = \{y_1, \ldots, y_m\}$ with $|Y| = m$. For ease of notation, we write $Q(y_i) = q_{i,j}$ and $Q(y_i | j + k/2) = q_{i,j+k/2}$. The problem (21) can be formulated as follows

$$\textbf{P1:} \max_{\{q_{ij}, q_{ij+k/2}\}^m_{i=1}} \sum_{i=1}^m \frac{(q_{ij} - q_{ij+k/2})^2}{q_{ij} + q_{ij+k/2}} \quad (22)$$

s.t. $H(q_{ij}, \ldots, q_{mj}) \leq R$,

$H(q_{ij+k/2}, \ldots, q_{mj+k/2}) \leq R$

$e^{-\epsilon} \leq q_{ij} / q_{ij+k/2} \leq e^\epsilon$, $\forall i \in [m]$

$q_{ij} \geq 0$, $q_{ij+k/2} \geq 0$, $\forall i \in [m]$

$m \sum_{i=1}^m q_{ij} = 1$, $m \sum_{i=1}^m q_{ij+k/2} = 1 \quad (23)$

Note that the objective function (22) is jointly convex in both $\{q_{ij}\}^m_{i=1}$ and $\{q_{ij+k/2}\}^m_{i=1}$. However, the optimization problem $\textbf{P1}$ is non-convex due to two reasons. First, we maximize a convex function, and second the entropy constraints (23) are sub-level sets of a concave function and are non-convex constraints. However, we can solve the optimization problem $\textbf{P1}$ by exploiting the results of Lemma 3 below.

**Lemma 3:** The optimal solution of the non-convex optimization problem $\textbf{P1}$ is obtained when the output size is $m = 2$.

The proof of Lemma 3 is presented in Appendix C in the supplementary material. Since the output alphabet is binary, we can efficiently plot the feasible region of $\textbf{P1}$ for $m = 2$ as depicted in Figure 10. Since we maximize a convex function, the optimal solution is at the boundary of the feasible set. Furthermore, the objective function (22) is symmetric on $q_{1,j}$, $q_{1,j+k/2}$ for $m = 2$. As a result, the optimal solution is given by

$q^*_{1,j} = \begin{cases} e^\epsilon / (\epsilon + 1) & \text{if } R \geq H_2(e^\epsilon / (\epsilon + 1)) \\ p_R & \text{if } R < H_2(e^\epsilon / (\epsilon + 1)) \end{cases}$

$q^*_{1,j+k/2} = \begin{cases} 1 / (\epsilon + 1) & \text{if } R \geq H_2(1 / (\epsilon + 1)) \\ p_R / e^\epsilon & \text{if } R < H_2(1 / (\epsilon + 1)) \end{cases} \quad (24)$

where $q^*_{2,j} = 1 - q^*_{1,j}$ and $q^*_{2,j+k/2} = 1 - q^*_{1,j+k/2}$. Substituting from (24) into the objective function (22), we get

$$\sum_{i=1}^m \frac{(q_{ij} - q_{ij+k/2})^2}{q_{ij} + q_{ij+k/2}} \leq \begin{cases} 2(e^\epsilon - 1)^2 / (e^\epsilon + 1)^2 - e^\epsilon & \text{if } R \geq H_2(e^\epsilon / (\epsilon + 1)) \\ 2p_R(e^\epsilon - 1)^2 / e^\epsilon & \text{if } R < H_2(e^\epsilon / (\epsilon + 1)) \end{cases} \quad (25)$$

Hence, the proof is completed for Lemma 2.

Using the bound from Lemma 2 in (19) and taking supremum over all $Q_i \in Q(\epsilon, R)$, we get

$$\sup_{Q_i \in Q(\epsilon, R)} D_{KL}(M_i^u || M_i^{y(2)}) \leq 2s^2 e^\epsilon \sup_{Q_i \in Q(\epsilon, R)} \sum_{y \in Y_i} (Q_i(y) - Q_i(y + j + k/2))^2 \quad (26)$$

Substituting from (26) into (18), we get

$$r_{\ell, \epsilon, R, n, k} \geq \begin{cases} \phi(\delta) b^2 \left( 1 - \sqrt{2s^2 n e^\epsilon (e^\epsilon - 1)^2 / (e^\epsilon + 1)^2} \right) & \text{if } R \geq H_2(e^\epsilon / (\epsilon + 1)) \\ \phi(\delta) b^2 \left( 1 - \sqrt{2s^2 n p_R e^\epsilon (e^\epsilon - 1)^2 / e^\epsilon} \right) & \text{if } R < H_2(e^\epsilon / (\epsilon + 1)) \end{cases} \quad (27)$$

By setting $s^2 = \frac{(e^\epsilon + 1)^2}{8n e^\epsilon (e^\epsilon - 1)^2}$ if $R \geq H_2(e^\epsilon / (\epsilon + 1))$ and $\delta^2 = \frac{e^\epsilon}{8p_R g(e^\epsilon - 1)^2}$ if $R < H_2(e^\epsilon / (\epsilon + 1))$, we get

$$r_{\ell, \epsilon, R, n, k} \geq \begin{cases} \phi(\delta) b^2 \left( \sqrt{\frac{(e^\epsilon + 1)^2}{8n e^\epsilon (e^\epsilon - 1)^2}} \right) & \text{if } R \geq H_2(e^\epsilon / (\epsilon + 1)) \\ \phi(\delta) b^2 \left( \sqrt{\frac{e^\epsilon}{8p_R g(e^\epsilon - 1)^2}} \right) & \text{if } R < H_2(e^\epsilon / (\epsilon + 1)) \end{cases} \quad (28)$$

For the loss function $\ell = \ell_2^2$, we set $\phi(x) = x^2$ and for $\ell = \ell_1$, we set $\phi(x) = |x|$. This completes the proof of Theorem 1 with a slightly worse constant of 32 instead of 16 in the denominator. We provide a different proof of Theorem 1 in Appendix A in the supplementary material using Fisher information that gives the exact bound as stated in Theorem 1.

**B. Upper Bound on the Minimax Risk Estimation Using the Hadamard Response**

In this section, we prove Theorem 2 (see page 749) by proposing a private mechanism by adapting the Hadamard response given in [20], where each user answers to a yes-no
question such that the probability of telling the truth depends on the amount of randomness $R$. Each user $i \in [n]$ has a binary output $Y_i \in \{0, 1\}$. The $(ε, R)$-LDP mechanism of the $i$-th user is defined by

$$Q(Y_i = 1|X) = \begin{cases} q & \text{if } X \in B_i \\ \frac{q}{2} & \text{if } X \notin B_i \end{cases} \quad (29)$$

where $B_i \subset [K]$ is a subset of inputs, and $q$ is a probability value that will be determined later such that $H_2(q) \leq R$. Let $K = 2^{\lceil \log(K) \rceil}$ denote the smallest power of 2 larger than $K$, and $H_K$ be the $K \times K$ Hadamard matrix. In the following, we assume an extended distribution $\tilde{p}$ over the set $X = [K]$ with $|X| = K$ that is obtained by zero-paddning the original distribution $p$ with $(K-k)$ zeros, i.e., $\tilde{p} = [p_1, \ldots, p_K] = [p_1, \ldots, p_k, 0, \ldots, 0]$. For $j \in [K]$, let $B_j$ be a set of row indices that have 1 in the $j$-th column of the Hadamard matrix $H_K$. For example, when $K = 4$, the Hadamard matrix is given by

$$H_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \quad (30)$$

Hence, $B_1 = \{1, 2, 3, 4\}$, $B_2 = \{1, 3\}$, $B_3 = \{1, 2\}$, and $B_4 = \{1, 4\}$. We divide the users into $K$ sets $(US_1, \ldots, US_K)$, where each set contains $n/K$ users. For each user $i \in US_j$, we set $B_i = B_j$. Let $p(B^i) = Pr[X \in B^i] = \sum_{x \in B^i} p_x$, and $s_j = Pr[Y_i = 1]$ for $i \in Uj$. Then, we can easily see that

$$s_j = p(B^j)q + (1 - p(B^j))\frac{q}{e^\epsilon} = p(B^j)q\left(\frac{e^\epsilon - 1}{e^\epsilon}\right) + \frac{q}{e^\epsilon} \quad (31)$$

Let $\hat{s}_j = \frac{1}{|US_j|} \sum_{i \in US_j} 1{Y_i = 1}$ denote the estimate of $s_j$. Then, we can estimate $p(B^j)$ as $\hat{p}(B^j) = \frac{e^\epsilon}{q(e^\epsilon - 1)}(\hat{s}_j - \frac{q}{e^\epsilon})$. Observe that the relation between the distribution $\tilde{p}$ and $p(B) = \{p(B^1), \ldots, p(B^K)\}$ is given by [20, Eq. 13]

$$p(B) = \frac{H_K \tilde{p} + 1_K}{2} \quad (32)$$

where $1_K$ denotes a vector of $K$ ones. Hence, we can estimate the distribution $\tilde{p}$ as

$$\hat{p} = H_K^{-1}(2\hat{p}(B) - 1_K) = \frac{1}{K}H_K(2\hat{p}(B) - 1_K). \quad (33)$$

**Lemma 4:** For arbitrary $p \in \Delta_k$, we have

$$E[\|p - \hat{p}\|^2] \leq \frac{2ke^{2\epsilon}}{nq^2(e^\epsilon - 1)^2} \quad (34)$$

The proof is exactly the same as the proof in [20, Th. 5]. By setting $q = \frac{e^\epsilon}{\epsilon + 1}$ if $R \geq H_2(\frac{e^\epsilon}{\epsilon + 1})$ and $q = \frac{p_R}{R}$ if $R < H_2(\frac{e^\epsilon}{\epsilon + 1})$, we get

$$\epsilon_{\ell,R,n,k}^j \leq \begin{cases} \frac{2ke^{\epsilon + 1}}{nq^2(e^\epsilon - 1)} & \text{if } R \geq H_2(\frac{e^\epsilon}{\epsilon + 1}) \\ \frac{2ke^{\epsilon}}{np_R(e^\epsilon - 1)^2} & \text{if } R < H_2(\frac{e^\epsilon}{\epsilon + 1}) \end{cases} \quad (35)$$

The difference in our mechanism is that we design the private mechanism (29) for all values of randomness $R$. This completes the proof of Theorem 2.

**VI. MULTI-LEVEL PRIVACY (PROOF OF THEOREM 3)**

This section proves Theorem 3 (see page 750) by establishing a new technique using a smaller amount of randomness than the trivial scheme mentioned in Section III-B while achieving the minimum risk estimation for each analyst. Our proposed mechanism for multi-level privacy (where $\epsilon_1 > \cdots > \epsilon_d$) is a cascading mechanism, where in each step, we add a random key to the output of the previous step (see Figure 11, for example). The common output of the mechanism is accessible by all analysts. However, each analyst would have a different privacy level depending on the amount of randomness shared with it. Thus, each analyst uses the shared random key to partially undo the randomization of the common output to get less privacy and higher utility. Let $z_j = \frac{1}{\epsilon_j + 1}$ for $j \in [d]$. For $i \in [n]$, let $(U^i_1, \ldots, U^i_d)$ be a set of $d$ Bernoulli random variables, where $U^j_i$ has a parameter $q_j = Pr[U^j_i = 1]$ given by

$$q_j = \begin{cases} z_j & \text{if } j = 1, \\ \frac{z_j - z_{j-1}}{1 - z_{j-1}} & \text{if } j > 1. \end{cases} \quad (36)$$

We first use the Hadamard response proposed in [20] for getting the first step of our mechanism (see Section V-B for more details). Let $H_K$ be the $K \times K$ Hadamard matrix. Let $B^j$ be a set of the row indices that have 1 in the $j$-th column of Hadamard matrix $H_K$ for $l \in [K]$. We divide the users into $K$ sets $(US_1, \ldots, US_K)$, where each set contains $n/K$ users. We assign a set $B_i = B^j$ representing a subset of inputs for each user $i \in US_l$. Then, user $i$ generates a virtual output $Y^j_i \in \{0, 1\}$ as follows

$$Y^j_i = \begin{cases} 1 & \text{if } (X_i \in B_j & U^j_i = 0) \text{ or } (X \notin B_i & U^j_i = 1), \\ 0 & \text{otherwise}. \end{cases} \quad (37)$$

Observe that the representation of $Y^j_i$ in (37) is exactly the same as in (29) by setting $q = Pr[U^j_i = 0] = \frac{e^\epsilon}{\epsilon + 1}$. We represent $Y^j_i$ with this form to explicitly show the random keys used to design the Hadamard scheme presented in Section V-B. Let $Y^d_i$ be the virtual output generated by user $i$ for the $j$th analyst, which is given by

$$Y^d_i = Y^j_i \oplus U^j_{i+1} \oplus \cdots \oplus U^d_{i} \quad (38)$$

where $\oplus$ denotes the bitwise XOR. Hence, we add randomization to the first step of the Hadamard scheme. User $i$ transmits the output $Y^d_i$ to all analysts. The private scheme is shown in Figure 11.
Lemma 5: The jth output of user i satisfies $\epsilon_j$-LDP, i.e.,
\[
\sup_{y_i \in [0,1]} \sup_{y_j': \in [0,1]} \Pr[ y_i' = y_i' \mid X_i = x_i] \leq \epsilon_j
\]  
(39)

We prove Lemma 5 in Appendix E in the supplementary material. Note that each analyst has access to the public outputs \( Y_1, \ldots, Y_d \) which is $\epsilon_i$-LDP. Additionally, user $i$ sends a random key $L_i^d = U_i^d \oplus \ldots \oplus U_i^{d+1}$ to the $i$th analyst. Using the random keys \( L_i^1, \ldots, L_i^d \), the $i$th analyst can construct the private outputs \( Y_i^1, \ldots, Y_i^d \) which are $\epsilon_j$-LDP, where $Y_i^j = Y_i^j \oplus L_i^j$. Observe that the privatized output $Y_i^j$ has a conditional distribution given by
\[
Q_i(Y_i^j \mid X_i) = \left\{ \begin{array}{ll}
\frac{\epsilon_j}{\epsilon_j + 1} & \text{if } X_i \in B_i \\
\frac{1}{\epsilon_j + 1} & \text{if } X_i \not\in B_i
\end{array} \right.
\]  
(40)

which coincides with the privacy mechanism given in (29) with $q = \frac{\epsilon_j}{\epsilon_j + 1}$. Thus, the $j$th analyst can recover the same output as in the trivial scheme using the shared private key. Hence, we get the same privacy and minimax risk as the trivial scheme.

From Lemma 4, for privacy level $\epsilon_j = O(1)$, we get that
\[
\epsilon_{ij} = O\left( \frac{k}{n\epsilon_j^2} \right),
\]  
(41)

for analyst $j$, which coincides with the lower bound stated in Corollary 1. Observe that the total amount of randomness per user in the proposed mechanism is given by
\[
R_{\text{total}}^{\text{proposed}} = \sum_{j=1}^{d} H(U_j) = \sum_{j=1}^{d} H_2(q_j) \leq R_{\text{total}}^{\text{trivial}},
\]  
(42)

where $q_j$ is defined in (36). Note that the last inequality is strict for $d > 1$, which follows from the argument presented in Section III-B. This completes the proof of Theorem 3.

VII. PRIVATE-RECOVERABILITY (PROOF OF THEOREM 4 AND THEOREM 5)

In this section, we prove Theorem 4 (see page 751) and Theorem 5 (see page 751).

A. Proof of Theorem 4

This section proves the necessary and sufficient conditions on the random key $U$ and the privatized output $Y$ to design an $\epsilon$-LDP-Rec mechanism. We first prove that $|\mathcal{Y}| \geq |\mathcal{X}|$ is necessary to recover $X$ from $Y$ and $U$. We then prove that each input $x \in \mathcal{X}$ should be mapped with non-zero probability to every output $y \in \mathcal{Y}$; hence, we get $|U| \geq |\mathcal{X}|$, since each input $x \in \mathcal{X}$ can be mapped with non-zero probability to at most $|\mathcal{U}|$ outputs. The main part of our proof is bounding the randomness of the key $U$ in the second condition. We first prove in Lemma 7 that for any $\epsilon$-LDP-Rec mechanism designed using a random key of size greater than the input size, there exists another $\epsilon$-LDP-Rec mechanism designed using a random key of size equal to the input size with the same or smaller amount of randomness. Thus, we can assume that $|\mathcal{U}| = |\mathcal{X}|$ and minimize the entropy of the random key $U$ over all possible distributions and under the $\epsilon$-LDP constraint. Since entropy is a concave function of the distribution, we get a non-convex problem. However, we can obtain an exact solution for the problem due to the structure of the privacy constraints that form a closed polytope. For the sufficiency part, we prove in Lemma 6 that we can construct an $\epsilon$-LDP-Rec mechanism using the random key $U_{\text{min}}$ defined in Theorem 4 that satisfies the two necessary conditions.

Before we proceed into the proof of Theorem 4, we first present the following two lemmas whose proofs are given in Appendix F and Appendix G, in the supplementary material, respectively.

Lemma 6: For a random key $U \in \mathcal{U}$ with size $|\mathcal{U}| = k$ having a distribution $q = \{q_1, \ldots, q_k\}$ such that $\max_q q_{\max} \leq \epsilon^e$, there exists an $\epsilon$-LDP-Rec mechanism using input $X \in [k]$ and an output $Y \in [k]$ designed using $U$.

This lemma shows that we can design an $\epsilon$-LDP mechanism with output size equal to the input size if we have a random key with size equal to the input size and having a distribution such that $\min_q q_{\min} \leq \epsilon^e$.

Lemma 7: Suppose that an $\epsilon$-LDP-Rec mechanism with an input $X \in [k]$ and an output $Y \in [k]$ is designed using a random key $U \in \mathcal{U}$ with size $|\mathcal{U}| = m > k$. Then there exists an $\epsilon$-LDP-Rec mechanism with an input $X \in [k]$ and an output $Y \in [k]$ designed using a random key $U' \in [k]$ such that $H(U) \geq H(U')$.

Now, we are ready to prove Theorem 4. We prove the first necessary condition of Theorem 4 in two parts: We can show $|\mathcal{Y}| \geq |\mathcal{X}|$ using the recoverability constraint and $|\mathcal{U}| \geq |\mathcal{Y}|$ using the privacy constraint. We prove these in Appendix H in the supplementary material.

From Lemma 7 and the first necessary condition, we see that the $\epsilon$-LDP-Rec mechanism with the smallest amount of randomness is obtained when $|\mathcal{U}| = |\mathcal{Y}| = |\mathcal{X}|$. Hence, we restrict our attention to this case only. Let $U \in [k]$ be a random key having a distribution $q = \{q_1, \ldots, q_k\}$. Without loss of generality, we assume that $q_1 \leq q_2 \leq \ldots \leq q_k$. Before we prove the necessity of the second condition, we claim that $q_k/q_1 \leq \epsilon^e$. We prove this using both privacy and recoverability constraints in Appendix H in the supplementary material.

Now, we are ready to prove the necessity of the second condition. Our objective is to find the minimum entropy of the random key $U$ with size $|\mathcal{U}| = k$ such that the private mechanism is $\epsilon$-LDP and the sample $X$ can be recovered from observing $Y$ and the random key $U$. The problem can be formulated as follows
\[
\min_{q = \{q_1, \ldots, q_k\}} H(U) = -\sum_{j=1}^{k} q_j \log(q_j)
\]  
(43)

s.t.,
\[
1 \leq q_j \leq \epsilon^e \quad \forall j \in [k] \quad (44)
\]
\[
\sum_{j=1}^{k} q_j = 1, \quad q_j \geq 0 \quad \forall j \in [k] \quad (45)
\]

where the constraint (44) is obtained from the claim proved above. Observe that the constraints (44)-(45) form a closed
polytope. Furthermore, the objective function (43) is a concave function on $q$. Since we minimize a concave function over a polytope, the global optimum point is one of the vertices of the polytope [26]. Since we have a single equality constraint, a vertex has to satisfy at least $k - 1$ inequality constraints with equality. Observe that none of the inequalities in (45) can be satisfied with equality, otherwise the privacy constraints in (44) would be violated. Thus, the optimal vertex is of the form

$$q = \left[ q_1, \ldots, q_k, e^\epsilon q_1, \ldots, e^\epsilon q_k \right]_{k-s \text{ terms } s \text{ terms}},$$

such that $s$ of inequalities from $\frac{q_j}{q_i} \leq e^\epsilon$ are satisfied with equality and $(k-s-1)$ of inequalities from $1 \leq \frac{q_j}{q_i}$ are satisfied with equality, where $s$ is a variable to be optimized. Hence, the optimal distribution has the form

$$q^* = \left[ q_1, \ldots, q_s, e^\epsilon q_1, \ldots, e^\epsilon q_s \right]_{k-s \text{ terms } s \text{ terms}},$$

where $q_s = \frac{1}{e^{s+1} - s}$, and $s$ is an integer parameter chosen to minimize the entropy as follows

$$s^* = \arg \min_{s \in \mathbb{N}} \sum_{j=1}^k q_j^* \log \left( \frac{1}{q_j^*} \right) = \arg \min_{s \in \mathbb{N}} \log(s(e^\epsilon - 1) + k) - \frac{se^\epsilon}{s(e^\epsilon - 1) + k} = \arg \min_{s \in \mathbb{N}} \log(s(e^\epsilon - 1) + k) + \frac{e^\epsilon k}{(e^\epsilon - 1)(s(e^\epsilon - 1) + k)} - \frac{e^\epsilon}{e^\epsilon - 1}. \tag{47}$$

In order to solve the optimization problem (47), we relax the problem by assuming $s$ is a real number taking values in $[0, k]$. The optimization problem in (47) is non-convex in for general values of $\epsilon$ and $k$. Thus, we get all local minima by setting the derivative to zero along with the boundary points $s \in [0, k]$. Then we check all these critical points to obtain the global minimum point. However, we can see that at the boundary points $s \in [0, k]$, the objective function is equal to $\log(k)$ which is the maximum entropy for any random variable with support size $k$. Hence, the optimal solution is one of the local minima. We can verify that the objective function has only one local minimum point by setting the derivative with respect to $s$ to zero. Thus, we get

$$s^* = k \frac{e^\epsilon (e^\epsilon - 1) + 1}{(e^\epsilon - 1)^2}, \tag{48}$$

where $s^*$ denotes the local minimum point. Since (47) is a continuous function in the real variable $s$, the optimal discrete point $s^*$ is within the local minimum $s^*$. Hence, we get the closest integer to the real value in (48). As a result, we get

$$H(U) \geq H \left( U^*_{\min} \right),$$

where $s^* = \arg \min_{s \in [\lfloor U \rfloor, \lfloor U \rfloor]} H(U^*_{\min})$ for $l = k \frac{e^\epsilon (e^\epsilon - 1) + 1}{(e^\epsilon - 1)^2}$, and $U^*_{\min}$ is a random variable having a distribution $q^*$ given in (46). Hence, the proof of the necessary part is completed.

The sufficiency part is straightforward: Note that the random key $U^*_{\min}$ defined in Theorem 4 satisfies the necessary conditions, and Lemma 6, we can construct an $\epsilon$-LDP-Rec mechanism using the random key $U^*_{\min}$. Thus, these conditions are sufficient.

B. Proof of Theorem 5

In this section, we show that the Hadamard response (HR) scheme proposed in [8] is, in fact, an $\epsilon$-LDP-Rec mechanism, where it is possible to recover the input $X$ from the output $Y$ and randomness $U$. Furthermore, we show that it is order optimal from a randomness perspective.9

We briefly describe the HR mechanism, and then analyze its performance. We refer to [8] for more details. The HR mechanism is parameterized by two parameters: $K$ denotes the support size of the private mechanism output ($Y = [K]$), and $s \leq K$ is a positive integer. For each $x \in X$, let $C_x \subseteq [K]$ be a subset of outputs of size $|C_x| = s$. The private mechanism for HR is defined by

$$Q(y|X) = \begin{cases} e^{\epsilon} & \text{if } y \in C_x \\ \frac{e^{\epsilon}}{se^{\epsilon} + k - s} & \text{if } y \notin C_x \end{cases} \tag{49}$$

We can easily show that this is a symmetric mechanism, i.e., it can be represented using a private key $U$ of size $|K|$ that is independent of the mechanism input $X$. Furthermore, the distribution of the private key $U$ is given by

$$q^{HR} = \left[ q_1, \ldots, q_k, e^\epsilon q_1, \ldots, e^\epsilon q_k \right]_{K-s \text{ terms } s \text{ terms}},$$

where $q = \frac{1}{se^{\epsilon} + k - s}$. It remains to choose $K$, $s$, and $\{C_x\}_{x \in X}$ for fixed $\epsilon$ and input size $|X| = k$. In [8, Sec. 5], the authors proposed $K = B \times b$ and $s = b/2$, where $B = 2^{\lfloor \log_2(\min(e^\epsilon, 2^{b})\rfloor - 1}$, and $b = 2^{\lfloor \log_2(k+1) \rfloor}$. Furthermore, each set $C_x$ is a subset of rows indices of the Hadamard matrix. These parameters are chosen such that $s$ is close to $\max\left(\frac{k}{e^\epsilon}, 1\right)$, and $K$ is approximately the smallest power of 2 greater than $k$. The reason behind using values that are powers of 2 is to exploit the structure of the Hadamard matrix. In [8, Th. 7], the authors proved that the minimax risk of HR for $\ell_2^2$ loss function is given by

$$\ell_2^2 \left( e_{n,k} \right) \leq \begin{cases} O \left( \frac{k}{e^\epsilon} \right) & \text{for } \epsilon < 1 \\ O \left( \frac{k}{e^\epsilon} \right) & \text{for } 1 \leq \epsilon \leq \log(k) \\ O \left( \frac{1}{n} \right) & \text{for } \epsilon > \log(k) \end{cases} \tag{50}$$

which is order optimal for all privacy levels. In addition, the authors in [20] have shown that the HR scheme is order optimal for heavy hitter estimation in the high privacy regime ($\epsilon = O(1)$). In the following, we analyze the performance of HR with respect to the randomness of the private mechanism. Observe that for fixed $\epsilon$ and $k$, the parameters $K$, $B$, and $b$ of HR is bounded by $\min(e^\epsilon, 2^{k}) \leq B \leq \min(e^\epsilon, 2k)$, $\min(e^\epsilon, 2^{k}) \leq b \leq \min(e^\epsilon, 2^{k-1})$, and $K \leq K \leq 4k$. Hence, the

9We mention that the Hadamard mechanism in [8] is symmetric with non-binary outputs, while the Hadamard response in [20] has only binary outputs.
entropy of the private key used to generate the HR private mechanism is bounded by
\[
H^{HR}(U) = \log \left( \frac{b}{2} e^\epsilon + K - \frac{b}{2} \right) - \frac{e^\epsilon b}{2 e^\epsilon + K - b}.
\]
\[
\leq \log \left( \frac{2k}{\min\{e^\epsilon, 2k\}} (e^\epsilon - 1) + 4k \right) - \frac{e^\epsilon}{e^\epsilon - 1 + 2 \min\{e^\epsilon, 2k\}}
\]
\[
= \begin{cases} 
\log \left( \frac{2k e^\epsilon - 1}{e^\epsilon - 1} \right) - \frac{e^\epsilon}{e^\epsilon + 1} & \text{if } \epsilon \leq \log(k) + 1, \\
\log(2k + 1) - \frac{e^\epsilon}{e^\epsilon + 1} & \text{if } \epsilon > \log(k) + 1.
\end{cases}
\]
(51)

The minimum entropy of the private key to generate an ε-LDP-Rec mechanism is bounded by (Theorem 4)
\[
H^{\min}(U) = \log(s^* e^\epsilon + k - s^*) - \frac{e^\epsilon s^*}{s^* e^\epsilon + k - s^*}
\]
\[
\geq \begin{cases} 
\log \left( \frac{e^\epsilon}{e^\epsilon + 1} \right) - \frac{e^\epsilon}{e^\epsilon + 1} & \text{if } \epsilon \leq \log(k), \\
\log(e^\epsilon + 1) - \frac{e^\epsilon}{e^\epsilon + 1} & \text{if } \epsilon > \log(k).
\end{cases}
\]
(52)

From (51) and (52), we can verify that HR is randomness-order-optimal for all privacy levels ε.

VIII. SEQUENCE OF DISTRIBUTION ESTIMATION (PROOF OF THEOREM 6)

In this section, we prove Theorem 6 (see page 751). The main idea of our proof is as follows. The first condition is obtained in a similar manner as in the proof of Theorem 4. For the second condition, we relate the minimum amount of randomness required to preserve privacy of $T - 1$ samples to the minimum amount of randomness required to preserve privacy of $T$ samples. In particular, we prove that $H(U) \geq H(U_{\min,T-1}) + H(U_{\min,1})$, where $H(U_{\min,i})$ is the minimum amount of randomness of a key when we have a database of $i$ input samples.

Definition 2: Let $U \in U$ be a random key drawn from a discrete distribution $q = [q_1, \ldots, q_T]$ with a support size $|U| = k^T$, where $q_u = \Pr[U = u]$. We say that the distribution $q$ satisfies ε-DP, if there exists a bijective function $f : \mathcal{X}^T \to [1 : k^T]$ from the dataset $\mathcal{X}^T$ to integers $[1 : k^T]$, such that for every neighboring databases $x, x' \in [k]^T$, we have
\[
\frac{q(x)}{q(x')} \leq e^\epsilon.
\]

We begin our proof with the following lemma which is a generalized version of Lemma 6. We prove it in Appendix I in the supplementary material.

Lemma 8: Consider an input database $x = (x^{(1)}, \ldots, x^{(T)}) \in [k]^T$, and a random key $U \in U = \{u_1, \ldots, u_T\}$ distributed according to an ε-DP distribution $q = [q_1, \ldots, q_T]$. Then, there exists an ε-DP-Rec mechanism $Q:[k]^T \to [k]^T$ that uses $U$ to create an output $Y^T \in [k]^T$, such that we can recover the input database $X^T$ from $(U, Y^T)$.

We can prove the first necessary condition of Theorem 6 (which is to show $|U| \geq |Y^T| \geq |X^T|$) in the same way as we proved that for Theorem 4. For completeness, we provide a proof of it in Appendix I in the supplementary material. Now we prove the necessity of the second condition. Consider an arbitrary ε-DP-Rec mechanism $Q$ with output $Y^T \in [k]^T$ using a random key $U \in U$, where $|Y^T| = m \geq k^T$ and $|U| = l \geq m$. Let $U \sim q$, where $q = [q_1, \ldots, q_l]$ such that $q_u = \Pr[U = u]$ for $u \in U$. Let $U_{X^T} \subset U$ be a subset of key values such that the input $X^T = x$ is mapped to $Y^T = y$ when $U \in U_{X^T}$. Thus, the private mechanism $Q$ can be represented as
\[
Q(y|x) = \sum_{u \in U_{X^T}} q_u.
\]
(54)

Observe that $\sum_{y \in [k]^T} Q(y|x) = 1$, since $Q(y|x)$ is a conditional distribution for any given $x \in [k]^T$. Since $Q$ is an ε-DP-Rec mechanism, it follows from the recoverability constraint that each input $x$ is mapped to $y$ using a different set of key values $(U_{X^T} \cap U_{Y^T}) = \phi$. Thus, for each $y \in [k]^T$, we have $s_y = \sum_{x \in [k]^T} Q(y|x) \leq 1$. Furthermore, we get $\sum_{x \in [k]^T} s_y = k^T$. We sort the $k^T$ databases in $X^T$ in lexicographic order by arranging them in increasing order of $x^{(1)}$ and so on. For example, database $x = (x^{(1)}, \ldots, x^{(i)}, x^{(i+1)}, \ldots, x^{(T)})$ will appear before the database $\tilde{x} = (x^{(1)}, \ldots, x^{(i)}, \tilde{x}^{(i+1)}, \ldots, \tilde{x}^{(T)})$ when $x^{(i+1)} < \tilde{x}^{(i+1)}$. Furthermore, we denote $x_i$ as the $i$th database in the lexicographic order for $i \in [k]^T$. Observe that $s_y = \sum_{x \in [k]^T} Q(y|x)$ for given $y \in [k]^T$. Thus, the probabilities $P^0 = \{P^0_1, \ldots, P^0_{k^T}\}$ construct a valid distribution with support size $k^T$, where $P^0_j = \frac{Q(y|x)}{s_y}$ for $j \in [k]^T$. Furthermore, for every neighboring databases $x, x' \in [k]^T$, we have
\[
\frac{Q(y|x)}{Q(y|x')} \leq e^\epsilon,
\]
where step (a) follows from the fact that $Q$ is an ε-DP-Rec mechanism. Hence, the distribution $P^0$ is ε-DP distribution.

The proof of the following lemma is presented in Appendix J in the supplementary material.

Lemma 9: For every output $y \in [k]^T$, we have $H(P^0) \geq H(U_{\min,T-1}) + H(U_{\min,1})$, where $H(U_{\min,i})$ denotes the minimum randomness of a private key when we have a database of $i$ samples for $i \in \{1, \ldots, T\}$. Using Lemma 9, we can prove Theorem 6 as follows.

\[
H(U) = \frac{1}{k^T} \sum_{x \in [k]^T} H(U) = \frac{1}{k^T} \sum_{x \in [k]^T} H(Y^T | X^T = x)
\]
\[
= \frac{1}{k^T} \sum_{y \in [k]^T} \sum_{x \in [k]^T} -Q(y|x) \log(Q(y|x))
\]
\[
= \frac{1}{k^T} \sum_{y \in [k]^T} \left[ s_y \left( \sum_{x \in [k]^T} \frac{Q(y|x)}{s_y} \log \left( \frac{Q(y|x)}{s_y} \right) \right) - s_y \log(s_y) \right]
\]
\[
= \frac{1}{k^T} \sum_{y \in [k]^T} [s_y H(P^0) - s_y \log(s_y)]
\]
(56)
\[
\begin{align*}
\sum_{y \in \mathcal{Y}^T} & \left[ s_y (H(U_{\min,T-1}) + H(U_{\min,1})) - s_y \log(s_y) \right] \\
\geq & \frac{1}{k_T} \left[ \sum_{y \in \mathcal{Y}^T} s_y (H(U_{\min,T-1}) + H(U_{\min,1})) - s_y \log(s_y) \right] \\
\geq & H(U_{\min,T-1}) + H(U_{\min,1}).
\end{align*}
\]

where step (a) follows from the fact that \( Q(y|x) \) is a function of \( U \). Step (b) follows from Lemma 9. The inequality (c) follows from solving the problem

\[
\begin{align*}
\min_{s_y} & \sum_{y \in \mathcal{Y}^T} s_y [H(U_{\min,T-1}) + H(U_{\min,1})] - s_y \log(s_y) \\
\text{s.t.} & \sum_{y \in \mathcal{Y}^T} s_y = k_T \quad \text{and} \quad 0 \leq s_y \leq 1, \forall y \in \mathcal{Y}^T
\end{align*}
\]

Note that \( f(x) = -x \log(x) \) is a concave function on \( 0 \leq x \leq 1 \). Therefore, the objective function in (59) is concave in \( \{s_y\} \). The minimum value of a concave function is one of the vertices which is obtained when all the inequalities are satisfied by equalities. By setting \( k_T \) of the \( s_y \)’s to be one and setting the remaining \( |\mathcal{Y}^T| - k_T \) of \( s_y \)’s to be zero, the objective value in (59) becomes \( k_T \), which gives inequality (c).

Now, from (58), we conclude that \( H(U) \geq TH(U_{\min,1}) \), where \( H(U_{\min,1}) \) is the minimum amount of randomness required to design an \( \epsilon \)-LDP-Rec mechanism given in Theorem 4. This completes the proof of Theorem 6.

REFERENCES

[1] C. Dwork, F. McSherry, K. Nissim, and A. Smith, “Calibrating noise to sensitivity in private data analysis,” in Proc. Theory Cryptogr. Conf., 2006, pp. 265–284.

[2] C. Dwork and A. Roth, “The algorithmic foundations of differential privacy,” Found. Trends Theor. Comput. Sci., vol. 9, nos. 3–4, pp. 211–407, 2014.

[3] A. D. Sarwate and K. Chaudhuri, “Signal processing and machine learning with differential privacy: Algorithms and challenges for continuous data,” IEEE Signal Process. Mag., vol. 30, no. 5, pp. 86–94, Sep. 2013.

[4] Ü. Erlingsson, V. Pihur, and A. Korolova, “RAPPOR: Randomized aggregatable privacy-preserving ordinal response,” in Proc. ACM SIGSAC Conf. Comput. Commun. Security, 2014, pp. 1054–1067.

[5] Apple. (2017). Differential Privacy. [Online]. Available: https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf

[6] C. E. Shannon, “Communication theory of secrecy systems,” Bell Syst. Tech. J., vol. 28, no. 4, pp. 656–715, 1949.

[7] C. Dwork, “Differential privacy and the U.S. census,” in Proc. ACM SIGMOD-SIGACT-SIGAI Symp. Principles Database Syst., 2019, pp. 2879–2887.

[8] J. Acharya, Z. Sun, and H. Zhang, “Hadamard response: Estimating distributions privately, efficiently, and with limited communication,” in Proc. Int. Conf. Artif. Intell. Statist. (AISTATS), 2019, pp. 1120–1129.

[9] Y. Dodis, A. López-Alt, I. Mironov, and S. P. Vadhan, “Differential privacy with imperfect randomness,” in Advances in Cryptology (CRYPTO), vol. 7417, R. Safavi-Naini and R. Canetti, Eds. Berlin, Germany: Springer, 2012, pp. 497–516.

[10] S. L. Warner, “Randomized response: A survey technique for eliminating evasive answer bias,” J. Amer. Stat. Assoc., vol. 60, no. 309, pp. 63–69, 1965.

[11] M. Ye and A. Barg, “Optimal schemes for discrete distribution estimation under locally differential privacy,” IEEE Trans. Inf. Theory, vol. 64, no. 8, pp. 5662–5676, Aug. 2018.

[12] S. Wang et al., “Mutual information optimally local private discrete distribution estimation,” 2016. [Online]. Available: arXiv:1607.08025.

[13] P. Kairouz, K. Bonawitz, and D. Ramage, “Discrete distribution estimation under local privacy,” 2016. [Online]. Available: arXiv:1602.07387.

[14] P. Kairouz, S. Oh, and P. Viswanath, “Extremal mechanisms for local differential privacy,” in Proc. Adv. Neural Inf. Process. Syst., 2014, pp. 2879–2887.

[15] R. Bassily and A. Smith, “Local, private, efficient protocols for succinct histograms,” in Proc. 47th Annu. ACM Symp. Theory Comput. (STOC), 2015, pp. 127–135.

[16] Z. Qin, Y. Yang, T. Yu, I. Khalil, X. Xiao, and K. Ren, “Heavy hitter estimation over set-valued data with local differential privacy,” in Proc. ACM SIGSAC Conf. Comput. Commun. Security (CSS), 2016, pp. 192–203.

[17] J. Hsu, S. Khanna, and A. Roth, “Distributed private heavy hitters,” in Proc. Int. Colloqium Automata Lang. Program., 2012, pp. 461–472.

[18] R. Bassily, K. Nissim, U. Stemmer, and A. G. Thakurta, “Practical locally private heavy hitters,” in Proc. Adv. Neural Inf. Process. Syst., 2017, pp. 2288–2296.

[19] M. Bun, J. Nelson, and U. Stemmer, “Heavy hitters and the structure of local privacy,” in Proc. ACM SIGMOD-SIGACT-SIGA Symp. Principles Database Syst., 2018, pp. 435–447.

[20] J. Acharya and Z. Sun, “Communication complexity in locally private distribution estimation and heavy hitters,” in Proc. Int. Conf. Mach. Learn. (ICML), 2019, pp. 51–60.

[21] J. Acharya, C. L. Canonne, C. Freitag, and H. Tyagi, “Test without trust: Optimal locally private distribution testing,” 2018. [Online]. Available: arXiv:1808.02174.

[22] B. Balle, G. Barthe, M. Gaboardi, and J. Geumleek, “Privacy amplification by mixing and diffusion mechanisms,” in Proc. Adv. Neural Inf. Process. Syst., 2019, pp. 13277–13287.

[23] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, “Minimax optimal procedures for locally private estimation,” J. Amer. Stat. Assoc., vol. 113, no. 521, pp. 182–201, 2018.

[24] J. Duchi and R. Rogers, “Lower bounds for locally private estimation via communication complexity,” 2019. [Online]. Available: arXiv:1902.00582.

[25] J. Duchi, M. J. Wainwright, and M. I. Jordan, “Local privacy and min-max bounds: Sharp rates for probability estimation,” in Proc. Adv. Neural Inf. Process. Syst., 2013, pp. 1529–1537.

[26] J. B. Rosen, “Global minimization of a linearly constrained concave function by partition of feasible domain,” Math. Oper. Res., vol. 8, no. 2, pp. 215–230, 1983.

Antonious M. Girgis received the B.Sc. degree in electrical engineering from Cairo University, Egypt, in 2014, and the M.Sc. degree in electrical engineering from Nile University, Egypt, in 2018. He is currently pursuing the Ph.D. degree in the Electrical Engineering Department, University of California at Los Angeles (UCLA), Los Angeles, CA, USA, and was an Exchange Research Assistant with Sabanci University, Turkey, from 2016 to 2017. He received the Masters Fellowship and a Graduate Research Assistantship from Nile University, from 2014 to 2017. He received the Electrical and Computer Engineering Department Fellowship from the UCLA for the year 2018 and 2019. His research interests include privacy, machine learning, information theory, and optimization.

Deepesh Data received the B.Tech. degree in computer science and engineering from the International Institute of Information Technology Hyderabad, Hyderabad, India, in 2011, and the M.Sc. and Ph.D. degrees from the School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai, India, in 2017. After that, he joined the Indian Institute of Technology Bombay as a Postdoctoral Fellow. Since March 2018, he has been with the University of California at Los Angeles, Los Angeles, USA, as a Postdoctoral Scholar. He has received the Microsoft Research India Ph.D. Fellowship. His research interests are in distributed optimization, machine learning, differential privacy, cryptography, algorithms, and information theory. He has received the ACM India Doctoral Dissertation Award (Honorable Mention), and the TIFR-Sasken Best Ph.D. Thesis Award in Technology and Computer Sciences.

Kamalika Chaudhuri, photograph and biography not available at the time of publication.

Christina Fragouli, photograph and biography not available at the time of publication.

Suhas N. Diggavi, photograph and biography not available at the time of publication.