RETHINKING DISCLOSURE PREVENTION WITH POINTWISE MAXIMAL LEAKAGE

SARA SAEIDIAN, GIULIA CERVIA, TOBIAS J. OECHTERING, AND MIKAEL SKOGLUND

KTH Royal Institute of Technology, 100 44 Stockholm, Sweden
e-mail address: saeidian@kth.se

IMT Nord Europe, Centre for Digital Systems, F-59000 Lille, France
e-mail address: giulia.cervia@imt-nord-europe.fr

KTH Royal Institute of Technology, 100 44 Stockholm, Sweden
e-mail address: oech@kth.se

KTH Royal Institute of Technology, 100 44 Stockholm, Sweden
e-mail address: skoglund@kth.se

Abstract. This paper introduces a paradigm shift in the way privacy is defined, driven by a novel interpretation of the fundamental result of Dwork and Naor about the impossibility of absolute disclosure prevention. We propose a general model of utility and privacy in which utility is achieved by disclosing the value of *low-entropy* features of a secret X, while privacy is maintained by hiding the value of *high-entropy* features of X. Adopting this model, we prove that, contrary to popular opinion, it is possible to provide meaningful *inferential* privacy guarantees. These guarantees are given in terms of an operationally-meaningful information measure called *pointwise maximal leakage* (PML) and prevent privacy breaches against a large class of adversaries regardless of their prior beliefs about X. We show that PML-based privacy is compatible with and provides insights into existing notions such as differential privacy. We also argue that our new framework enables highly flexible mechanism designs, where the randomness of a mechanism can be adjusted to the entropy of the data, ultimately, leading to higher utility.

1. Introduction

Research in the privacy domain continuously evolves as novel notions of privacy aim to address challenges emerging in applications of data science. Arguably, one of the most successful notions of privacy is *differential privacy* (Dwork et al., 2006b, 2014). Introduced by Dwork et al. (2006b), differential privacy guarantees that an individual partaking in a data processing scheme will not face substantially increased risks due to their participation. This guarantee is achieved by ensuring that the outcome of the data processing is not much affected by whether or not each person participates. On the other hand, differential privacy, by design, does not rule out the possibility of privacy violations by *association*. That is, an

Key words and phrases: Disclosure Prevention, Information Leakage, Inferential Privacy, Pointwise Maximal Leakage.

adversary can still exploit *correlations* among pieces of data to uncover sensitive information about an individual from the outcome of a differentially private mechanism. To account for potential privacy violations by association, Tschantz et al. (2020) argue that differential privacy should be understood as a *causal* property of an algorithm. That is, differential privacy simply ensures that an algorithm produces similar outputs when supplied with inputs that differ in a single parameter. From the causal standpoint, (an unintended) inference about an individual is considered to be a privacy breach only if it is specifically caused by the inclusion of the individual's information in a dataset (Kifer et al., 2022).

On the other hand, the above causal interpretation no longer applies if we adopt a Bayesian perspective and assume that databases are sampled from an underlying probability distribution. In particular, several works argue that from the Bayesian point of view, differential privacy either (implicitly) assumes a product distribution on the database or restricts itself to *informed adversaries* (Kifer and Machanavajjhala, 2011, 2012; He et al., 2014; Liu et al., 2016; Yang et al., 2015; Li et al., 2013; Zhu et al., 2014).[1] These works usually provide examples and attack scenarios involving databases containing highly correlated data points and then argue that differential privacy falls short of providing sufficient protection in these cases. For instance, Kifer and Machanavajjhala (2011) give an example about a medical database in which Bob's data is perfectly correlated with the data of a large number of other patients. Then, they argue that the Laplace mechanism (Dwork et al., 2006b) does not provide sufficient protection in this case since the effect of Bob's data is amplified by the other data points. These works then often focus on developing tools to guarantee privacy particularly in the case of correlated datasets (Zhu et al., 2014; Liu et al., 2016).

A privacy guarantee that can rule out the possibility of privacy breaches due to association must be *inferential* in nature, that is, it must ensure that an adversary's knowledge about the world after interacting with a mechanism does not change much from her prior knowledge.[2] However, inferential guarantees are generally considered to be impossible to achieve by the negative results of Dwork and Naor (2010) and Kifer and Machanavajjhala (2011) (see also (Kifer et al., 2022, Sec. 7.1)). Particularly, Dwork and Naor (2010) prove that (under certain assumptions), no mechanism providing *non-trivial utility* can prevent disclosures against adversaries who may possess *auxiliary information* about a secret X. This is because an adversary may exploit auxiliary information to disclose more information than what a privacy mechanism intended to release. As an illustrating example, suppose each person's exact height is a secret, and consider a database containing height measurements of people with different nationalities. Assume that the average heights of women of different nationalities are released. Then, an adversary who observes the released values and has the auxiliary information "Terry Gross is two inches shorter than the average Lithuanian woman" learns Terry Gross' exact height (Dwork and Naor, 2010). Here, if we adopt an inferential view of privacy naively we may conclude that Terry Gross' privacy rights are violated.

1.1. **Overview and Contributions.**

At a high level, Dwork and Naor (2010) demonstrate that to provide utility a privacy mechanism necessarily has to disclose some information. To

[1] An informed adversary knows all the entries in a database except for one (Dwork et al., 2006b).

[2] We call a privacy notion *inferential* if it is defined by comparing an adversary's posterior knowledge with her prior knowledge. This includes definitions such as maximal leakage (Alvim et al., 2014; Issa et al., 2019), pointwise maximal leakage (Saeidian et al., 2022b), and (local) information privacy (Calmon and Fawaz, 2012; Jiang et al., 2021) but excludes frameworks that simply assume an underlying distribution on the data, e.g., Pufferfish privacy (Kifer and Machanavajjhala, 2014) or Bayesian differential privacy (Yang et al., 2015).

account for this result, differential privacy was designed to distinguish between data that is part of a dataset X and data that is not part of X but correlated with it, where the former is protected but the latter may be disclosed. We call this distinction the *in/out dichotomy*. In this paper, we argue that the in/out dichotomy is not the only way of distinguishing between information that should be protected through privacy guarantees and information that may be disclosed. In particular, we present an alternative distinction termed the *local/global dichotomy*. The concept of the local/global dichotomy yields a fresh perspective on privacy which is compatible with the Bayesian view rather than the causal one required by differential privacy.

The key to enabling our paper's findings is a fundamental and application-agnostic examination of what constitutes meaningful privacy and what we may consider as utility in privacy-preserving analytics. Roughly speaking, we define privacy as the ability of a mechanism to hide properties that are unique to each realization of the secret $x \in \mathcal{X}$. We call these properties *local* features of X. Conversely, we define utility as the ability of a mechanism to disclose properties of the entire population of X, that is, properties that X satisfies with high probability. These properties are called *global* features of X. We formally characterize local and global features of X using the concept of *min-entropy* of a probability distribution. Then, we prove that a recently proposed information measure called *pointwise maximal leakage* (PML) (Saeidian et al., 2022b,a, 2023) can be used to protect local features of X but disclose global features of it. That is, we prove that PML achieves privacy according to the local/global dichotomy. Most notably, we establish that to avoid privacy breaches, it is sufficient to make privacy guarantees based solely on assumptions about the true data-generating distribution, and without the need to assess the subjective information leaked to each adversary. It should be emphasized that PML measures information leakage through posterior-to-prior comparisons. Hence, our results indicate that, contrary to popular belief, an inferential perspective on privacy is not all at odds with the results of Dwork and Naor (2010).

Since PML and the local/global dichotomy introduce a new perspective on privacy, we are motivated to ask if existing *indistinguishability*-based definitions (Dwork et al., 2006b) can be understood and discussed from our inferential viewpoint. We demonstrate that (pure) differential privacy (Dwork et al., 2006b) and free-lunch privacy (Kifer and Machanavajjhala, 2011) admit several equivalent formulations in terms of PML. These formulations also show that existing privacy-preserving methods such as the Laplace mechanism (Dwork et al., 2006b) or randomized response (Warner, 1965) can be used out of the box to achieve privacy in the sense of PML. On top of that, our Bayesian view has the advantage that the calculated privacy parameter takes the data-generating distribution into account. More precisely, we show that when the data-generating distribution has large entropy, the privacy cost associated with the Laplace mechanism can be significantly lower than the differential privacy parameter.

Our contributions are briefly summarized as follows:

- We characterize privacy and utility and formally define disclosure in terms of min-entropy (Sections 3.1 and 3.2). Using these concepts, we argue that the results of Dwork and Naor (2010) can be retrieved (Proposition 3.6) and re-interpreted in our framework in a way that is consistent with the inferential perspective on privacy.
- We show that PML provides privacy guarantees according to the local/global dichotomy by relying solely on assumptions about the underlying distribution of the data (Theorem 3.5).

We also argue that the privacy parameter of a PML-based guarantee is easily interpretable and admits meaningful upper bounds (Section 3.3).

- We show how the inferential view can be used to understand the no-free-lunch theorem of Kifer and Machanavajjhala (2011) about the inconsistency of utility with privacy under all possible data-generating distributions (Theorem 3.7).
- We show that pivotal definitions such as pure differential privacy and free-lunch privacy admit several equivalent formulations in terms of PML (Theorems 4.2 and 4.4); hence, they can be interpreted from the inferential standpoint.
- We argue that the inferential perspective offers a significantly flexible design paradigm, where it may even be safe to answer highly general questions about the data deterministically (Example 3.11). We also argue that existing mechanisms can be used more efficiently when privacy is guaranteed in the sense of PML. This is because the privacy cost we pay for answering queries can be adjusted to the entropy of the underlying distribution on the data. We demonstrate this in the case of a counting query answered by the Laplace mechanism (Section 4.1).

2. Preliminaries

2.1. **Notation and Terminology.** We use uppercase letters to describe random variables and calligraphic letters to describe sets. Specifically, X denotes some data that contains sensitive information (i.e., the *secret*) and takes values in the finite set $\mathcal{X}$. We use P_X to represent the (true) probability distribution of X, p_X to represent the probability mass function (pmf) of X, and $\operatorname{supp}(P_X) \coloneqq \{x \in \mathcal{X} : p_X(x) > 0\}$ to represent the support set of P_X. Without loss of generality, we assume that $\mathcal{X} = \operatorname{supp}(P_X)$. Let $\mathcal{P}_{\mathcal{X}}$ denote the set of all distributions with full support on $\mathcal{X}$. We use $Q_X \in \mathcal{P}_{\mathcal{X}}$ to represent an adversary's (prior) belief about X.[3] Note that Q_X may be different from the true distribution P_X on X, but we assume that Q_X and P_X are mutually absolutely continuous. We use q_X to denote the pmf of Q_X.

Let $P_{Y|X}$ be a *privacy mechanism* (i.e., a conditional probability kernel) that answers queries about X and let Y represent the public query responses. Suppose Y takes values in the set $\mathcal{Y}$. The set $\mathcal{Y}$ may be finite (e.g. if Y is the outcome of the randomized response mechanism (Warner, 1965)) or infinite (e.g. if Y is the outcome of the Laplace mechanism (Dwork et al., 2006b)). Given $x \in \mathcal{X}$, we use $p_{Y|X=x}$ to denote the density of $P_{Y|X=x}$ with respect to a suitable σ-finite measure on $\mathcal{Y}$. For example, when $\mathcal{Y}$ is a countable set then we use the counting measure and when $\mathcal{Y}$ is a Euclidean space then we use the Lebesgue measure. Similarly, P_Y denotes the distribution of Y induced by $P_{Y|X}$ and P_X and p_Y denotes the density of P_Y with respect to a suitable measure on $\mathcal{Y}$.

Let P_{XY} denote the joint distribution of X and Y. We write $P_{XY} = P_{Y|X} \times P_X$ to imply that $p_{XY}(x, y) = p_{Y|X=x}(y)p_X(x)$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$, where p_{XY} is the density of P_{XY} with respect to a suitable σ-finite measure on $\mathcal{X} \times \mathcal{Y}$. Furthermore, we use $P_Y = P_{Y|X} \circ P_X$ to denote marginalization over X, i.e., to imply that $p_Y(y) = \sum_{x \in \mathcal{X}} p_{Y|X=x}(y)p_X(x)$ for all $y \in \mathcal{Y}$. These notations can be extended to more than two random variables in a natural way.

Given random variables U, X and Y, we say that the Markov chain $U - X - Y$ holds if U and Y are conditionally independent given X, that is, if $P_{UY|X} = P_{U|X} \times P_{Y|X}$. This

[3] For convenience, we identify adversaries with their prior beliefs.

implies that Y depends on U only through X and vice versa. We call a U satisfying the Markov chain $U - X - Y$ an *attribute* or *feature* of X, which is induced by the probability kernel $P_{U|X}$. Following the terminology of Dwork and Naor (2010, p. 96), we call $P_{U|X}$ a piece of *auxiliary information* which describes how U depends on the secret X. We assume that U takes values in a finite set $\mathcal{U}$.

Finally, given a positive integer n, $[n] \coloneqq \{1, \dots, n\}$ describes the set of all positive integers smaller than or equal to n, and $\log(\cdot)$ denotes the natural logarithm.

2.2. **Min-entropy and Rényi Divergence of Order Infinity.** We use *min-entropy*, i.e., Rényi entropy of order infinity (Rényi, 1961), as a measure of the uncertainty of a probability distribution.

Definition 2.1 (Min-entropy)**.** Suppose X is a (finite) random variable distributed according to P_X. The min-entropy $H_\infty(P_X)$ of X is

$$H_\infty(P_X) = -\log\left(\max_{x \in \mathcal{X}} p_X(x)\right).$$

Note that $H_\infty(P_X)$ is maximized when P_X is uniform over $\mathcal{X}$, and becomes zero when P_X is degenerate, i.e., when $\operatorname{supp}(P_X)$ is a singleton. Henceforth, we use the terms "entropy" and "min-entropy" interchangeably.

We now recall the definition of Rényi divergence of order infinity (Rényi, 1961; van Erven and Harremoës, 2014), which we then use to define PML.

Definition 2.2 (Rényi divergence of order ∞ (van Erven and Harremoës, 2014, Thm. 6))**.** Let P and Q be probability measures on a measurable space. Let p and q denote the densities of P and Q with respect to a dominating σ-finite measure. The Rényi divergence of order ∞ of P from Q is

$$D_\infty(P\|Q) = \log\left(\operatorname*{ess\,sup}_P \frac{p}{q}\right),$$

where $\operatorname{ess\,sup}_P f = \sup\{c \in \mathbb{R} \colon P(f > c) > 0\}$ for all measurable functions f.

2.3. **Pointwise Maximal Leakage.** Pointwise maximal leakage (PML) (Saeidian et al., 2022b,a, 2023) is an operationally meaningful privacy measure that quantifies the amount of information leaking about a secret random variable X to a single outcome of a privacy mechanism $P_{Y|X}$. Saeidian et al. (2022b) defined PML by considering two different threat models: the *randomized function view* and the *gain function view*. According to the randomized function view (first introduced by Issa et al. (2019)), PML is defined as the largest increase in the posterior probability of correctly guessing the value of an arbitrary attribute of X compared to the prior probably of correctly guessing the value of that attribute. That is, the adversary of this model is assumed to possess all possible auxiliary information about X, making PML a particularly suitable privacy notion for discussing the results of Dwork and Naor (2010). Moreover, according to the gain function view (first introduced by Alvim et al. (2012)), PML is defined as the largest increase in the posterior expected gain of an adversary compared to her prior gain. Saeidian et al. (2022b) then prove that both definitions of PML yield a simple expression which we state below in Definition 2.3. It is worth emphasizing that these threat models make explicit the type of privacy captured by PML; thus, unlike most other definitions, PML need not rely on

natural-language descriptions of a mathematical quantity to be interpreted. Furthermore, PML satisfies a post-processing inequality and increases linearly under composition (Saeidian et al., 2022b, Lemma 1).

Below, we define PML and conditional PML which are used extensively in the paper. To define the conditional form, it is assumed that the adversary has some *side information* about X already before interacting with the mechanism $P_{Y|X}$, for instance, a subset of the entries in a database. This information is modeled as the outcome of a random variable correlated with X and Y.

Definition 2.3 (PML (Saeidian et al., 2022b, Thm. 1))**.** Let P_{XY} be a distribution on the set $\mathcal{X} \times \mathcal{Y}$ with the marginal distribution P_X on $\mathcal{X}$. The pointwise maximal leakage from X to $y \in \mathcal{Y}$ is[4]

$$\ell_{P_{XY}}(X \to y) = D_\infty(P_{X|Y=y} \| P_X),$$

where $P_{X|Y=y}$ denotes the posterior distribution of X given $y \in \mathcal{Y}$.

When the joint distribution used to measure the information leakage is clear from context, we do not specify it as a subscript and write $\ell(X \to y)$. Note that PML is non-negative and bounded above by $-\log\left(\min_{x\in\mathcal{X}} P_X(x)\right)$. It also satisfies a pre-processing inequality indicating that a privacy mechanism leaks less information about attributes of X compared to X itself. Formally, if the Markov chain $U - X - Y$ holds, then $\ell(U \to y) \leq \ell(X \to y)$ for all $y \in \mathcal{Y}$ (Saeidian et al., 2022b, Lemma 1).

Definition 2.4 (Conditional PML (Saeidian et al., 2022b, Def. 3))**.** Let P_{XYZ} be a distribution on the set $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. Given $z \in \operatorname{supp}(P_Z)$, the conditional pointwise maximal leakage from X to $y \in \mathcal{Y}$ is

$$\ell_{P_{XYZ}}(X \to y \mid z) = D_\infty(P_{X|Y=y,Z=z} \| P_{X|Z=z}).$$

Saeidian et al. (2022b) define several privacy guarantees by restricting PML in various ways. The simplest definition, called ϵ-PML, bounds the information leaking through the mechanism by $\epsilon \geq 0$ with probability one. Here, we extend the definition of ϵ-PML to encompass scenarios where P_X is not precisely known, but is assumed to belong to a subset of $\mathcal{P}_\mathcal{X}$.

Definition 2.5 ($(\epsilon, \mathcal{P})$-PML)**.** Suppose X is distributed according to $P_X \in \mathcal{P} \subseteq \mathcal{P}_\mathcal{X}$. Given $\epsilon \geq 0$, we say that the mechanism $P_{Y|X}$ satisfies $(\epsilon, \mathcal{P})$-PML if

$$P_Y\left(\left\{y \in \mathcal{Y} : \ell_{P_{Y|X}\times P_X}(X \to y) \leq \epsilon\right\}\right) = 1,$$

for all $P_X \in \mathcal{P}$, or equivalently, if

$$\sup_{P_X \in \mathcal{P}} D_\infty(P_{Y|X} \times P_X \| P_Y \times P_X) \leq \epsilon.$$

For simplicity, we assume that the density $p_{Y|X=x}(y)$ is continuous on $\mathcal{Y}$ for all $x \in \mathcal{X}$.[5] In this case, $P_{Y|X}$ satisfies $(\epsilon, \mathcal{P})$-PML if

$$\sup_{P_X \in \mathcal{P}} \sup_{y \in \mathcal{Y}} \ell_{P_{Y|X}\times P_X}(X \to y) \leq \epsilon.$$

[4]To be able to define PML for all $y \in \mathcal{Y}$, we use the convention that $P_{X|Y=y} = P_X$ if $p_Y(y) = 0$. That is, conditioning on outcomes with density zero equals no conditioning.

[5]See (Rudin, 1986, Remark 3.15) for a discussion on replacing the essential supremum by the actual supremum of a function.

2.4. **Leakage Capacity.** Now, we define the notion of the *leakage capacity* of a privacy mechanism which, according to (Issa et al., 2019, Thm. 14), describes the largest amount of information that can leak through a mechanism $P_{Y|X}$.

Definition 2.6 (Leakage Capacity)**.** The leakage capacity of a privacy mechanism $P_{Y|X}$ is

$$C(P_{Y|X}) \coloneqq \log \sup_{y\in\mathcal{Y}} \max_{x,x'\in\mathcal{X}} \frac{p_{Y|X=x}(y)}{p_{Y|X=x'}(y)}.$$

Note that $C(P_{Y|X})$ is infinite if there exists $(x,y) \in \mathcal{X}\times\mathcal{Y}$ such that $p_{Y|X=x}(y) = 0$ but $p_Y(y) > 0$. In (Fernandes et al., 2022), the quantity $\exp\big(C(P_{Y|X})\big)$ is called *lift capacity* and is used to establish a connection between a privacy measure called *max-case g-leakage* and local differential privacy (Duchi et al., 2013). Definition 2.6 is also related to the notion of *indistinguishability* (Dwork et al., 2006b) as well as differential privacy.

Theorem 2.7 ((Issa et al., 2019, Thm. 14))**.** *Given a privacy mechanism $P_{Y|X}$ it holds that*

$$C(P_{Y|X}) = \sup_{P_X\in\mathcal{P}_\mathcal{X}} \sup_{y\in\mathcal{Y}} \ell_{P_{Y|X}\times P_X}(X\to y) = \sup_{P_X\in\mathcal{P}_\mathcal{X}} D_\infty(P_{Y|X}\times P_X \| P_Y \times P_X).$$

In Section 4, we characterize (pure) differential privacy and free-lunch privacy (Kifer and Machanavajjhala, 2011) in terms of PML. These results can be considered as consequences of Theorem 2.7.

3. Impossibility of Absolute Disclosure Prevention

Any aspiring inferential privacy framework should first be reconciled with the results of Dwork and Naor (2010) and Kifer and Machanavajjhala (2011), and this is the subject we take up in this section. We mainly discuss the results of Dwork and Naor (2010) as they correspond more directly to the ideas laid out in this paper, but we also draw connections to (Kifer and Machanavajjhala, 2011).[6]

Dwork (2006) proves a fundamental result that marks the beginning of the developments in the area at of differential privacy. This result, dubbed the *impossibility result*, proves that no mechanism providing "non-trivial utility" can prevent disclosures against adversaries who may possess arbitrary *auxiliary information* about a secret X.[7] Roughly speaking, Dwork and Naor (2010) demonstrate that an adversary can exploit auxiliary information to make unintended inferences about quantities correlated with X, as illustrated by the example about Terry Gross' height. Thus, privacy guarantees that ensure neither X nor any quantity correlated with X is disclosed can be achieved only at the cost of destroying all utility, because it is not feasible to control for the adversary's auxiliary information.

The impossibility result states that to provide utility, one necessarily has to disclose some information. This raises the question: What information can we (and should we) protect through privacy guarantees, and what information will we inevitably disclose? The answer differential privacy gives to this question is that privacy guarantees should be limited

[6]We emphasize that Dwork and Naor (2010) and Kifer and Machanavajjhala (2011) prove conceptually different results. Specifically, Dwork and Naor (2010) prove that absolute disclosure prevention is impossible due to the auxiliary information that may be available to an adversary, even if we assume a single fixed and publicly known prior distribution. On the other hand, Kifer and Machanavajjhala (2011) prove that guaranteeing privacy under all possible prior distributions severely restricts utility.

[7]The impossibility result is somewhat extended in (Dwork and Naor, 2010) and we mostly refer to ideas from this later version.

to information that is directly included in X. So, when X is a database, the individuals who have contributed their data to the database should be protected, but no such guarantee is provided to individuals whose data may be correlated with X in other ways. That is, a distinction is made between information that is directly included in X and information that is not part of X but may be correlated with it. We call this distinction the *in/out dichotomy*. As the basis for differential privacy, the in/out dichotomy has proved to be a very useful idea for addressing the impossibility result.

Nevertheless, the in/out dichotomy is not the only way we can distinguish between the information that we protect and the information that we allow to be disclosed. Below, we present an alternative distinction which we call the *local/global dichotomy*. The idea behind the local/global dichotomy is that we protect features of X that have large entropy (i.e., local features) while we allow disclosing features of X with small entropy (i.e., global features).[8] This view is motivated by how we define utility and what we consider to be a privacy breach. In particular, we argue that features of the data that capture properties of the population as a whole have small entropy and may be disclosed for the sake of utility, whereas instance-dependent features of the data have large entropy and should remain secret. Then, similar to how differential privacy provides guarantees according to the in/out dichotomy, we show that PML's guarantees are based on the local/global dichotomy. In short, the local/global dichotomy allows us to reconcile the results of Dwork and Naor (2010) with the guarantees of an effective inferential privacy framework. The main advantage of this view is that *those features of X that may be revealed by the privacy mechanism are exactly those population-level features of the data that we would anyway want to be able to disclose to provide utility.* On top of that, this view is directly applicable to many types of secrets and not just private databases.

In what follows, we assume that X is any type of data containing sensitive information, for example, a database or a piece of information belonging to a single individual. All the results presented in this section are proved in Appendix A.

3.1. **What Is Privacy and What Is Utility?** Our results and discussions throughout the paper depend crucially on definitions of utility and privacy formulated in terms of entropy. As such, in this subsection, we recall the definitions and assumptions of Dwork and Naor (2010), in particular, the notions of utility and privacy posited there. We then present our own definitions and assumptions and discuss how they differ from that of Dwork and Naor (2010).

Suppose X is distributed according to $P_X \in \mathcal{P}_{\mathcal{X}}$. Dwork and Naor (2010) assume that P_X is publicly known, that is, P_X also represents the prior belief of an adversary who interacts with a privacy mechanism $P_{Y|X}$. To define utility, Dwork and Naor (2010) posit a random variable U satisfying the Markov chain $U - X - Y$ whose value represents the answer to a question posed about X. It is assumed that the value of U cannot be *a priori* predicted from its distribution $P_U = P_{U|X} \circ P_X$, that is, the entropy $H_\infty(P_U)$ is large. However, to provide utility, the mechanism must either disclose the value of U exactly or allow estimating U with high accuracy, i.e., it is assumed that there exists $y \in \mathcal{Y}$ such that the entropy $H_\infty(P_{U|Y=y})$ is either very small or zero. Furthermore, to define privacy, Dwork and Naor (2010) suppose the existence of a random variable W satisfying the Markov chain $W - X - Y$ whose value must remain secret. That is, the value of W must be difficult to

[8]These entropies are calculated using the true prior distribution P_X on the data.

guess with or without access to the mechanism, but it is assumed that W has smaller entropy compared to U. Formally speaking, $H_\infty(P_W)$ and $H_\infty(P_{W|Y=y})$ are both large for all $y \in \mathcal{Y}$, but $H_\infty(P_W) < H_\infty(P_U)$. It is important to note that the condition $H_\infty(P_W) < H_\infty(P_U)$[9] is indispensable in the proof of the impossibility result because Dwork and Naor (2010) assume that it is possible to extract enough randomness from U to mask the value of W.

Our setup differs from (Dwork and Naor, 2010) in several key aspects. We let distribution $Q_X \in \mathcal{P}_\mathcal{X}$ represent the prior belief of an adversary who observes the outcome of the privacy mechanism. This distribution may or may not be equal to P_X, but P_X and Q_X are mutually absolutely continuous. To provide utility, the mechanism $P_{Y|X}$ releases some *global* information about the secret X, and releasing this information is *not* considered to be a privacy breach. We define global information as the value of any attribute of X that can be accurately predicted by an analyst who knows the true distribution P_X and possibly some auxiliary information but without access to the privacy mechanism $P_{Y|X}$. Formally, we posit a Markov chain $U - X - Y$, where U is an attribute of X and the kernel $P_{U|X}$ is the analyst's auxiliary information. If U contains global information about X, then the entropy $H_\infty(P_U)$ must be *small* since the value of U should be predictable using the distribution P_U alone (where $P_U = P_{U|X} \circ P_X$) and without access to $P_{Y|X}$. Heuristically, such attributes describe properties of the population of X and are largely instance-independent. Hence, they may be disclosed to provide utility. By contrast, to maintain privacy, we wish to protect instance-dependent and *local* properties of X, which are represented by those attributes of X that have *large* entropy. Consider an attribute W of X satisfying the Markov chain $W - X - Y$. If $H_\infty(P_W)$ is large (where $P_W = P_{W|X} \circ P_X$), then even an analyst who knows the true underlying distribution P_X and the auxiliary information $P_{W|X}$ cannot reliably estimate W; hence, it is only through the mechanism $P_{Y|X}$ that the value of W can be disclosed. Accordingly, we consider it to be a privacy breach if the value of any high-entropy attribute of X is disclosed.

The above distinction between high-entropy local features of X and low-entropy global features of it is what was earlier called the local/global dichotomy. This is further illustrated by the examples below, where the second example is inspired by (Kasiviswanathan and Smith, 2014).

Example 3.1. Suppose the database $X = (D_1, \ldots, D_n)$ is i.i.d, where each entry D_i is drawn according to a distribution P_D defined over a finite set of real numbers in the interval $[a, b)$. Our goal is to estimate the expectation $\mu = \mathbb{E}_{P_D}[D_i]$. We may aim to disclose one of the following two estimates: the quantized sample mean $\hat{\mu}_1 = q_m\left(\frac{\sum_{i=1}^n D_i}{n}\right)$, or the first row of the database $\hat{\mu}_2 = D_1$. The quantization $q_m(.)$ can be described as follows: Fix a large integer m, and values $c_1, \ldots, c_{m-1}$ satisfying $a = c_0 < c_1 < \ldots < c_m = b$. Let $\mathcal{C} = \{\frac{c_0+c_1}{2}, \ldots, \frac{c_{m-1}+c_m}{2}\}$. Then, $q_m : [a, b) \to \mathcal{C}$ denotes a quantizer that maps real numbers in the interval $[c_j, c_{j+1})$ to $\frac{c_j+c_{j+1}}{2}$.[10]

By the law of large numbers, as $n \to \infty$ the sample mean converges in probability to μ; thus, $H_\infty(\hat{\mu}_1) \to 0$. In contrast, the distribution of $\hat{\mu}_2$ does not depend on n; hence, $\hat{\mu}_2$ has

[9] This condition is implied by the lower bound on the entropy of the utility vector in terms of the length of the privacy breach in (Dwork and Naor, 2010, Assumption 1).

[10] By the central limit theorem, the sample mean converges in distribution to a Gaussian random variable as $n \to \infty$. Thus, we use the quantization to ensure that the entropy of our estimator remains well-defined as $n \to \infty$. The quantization introduces some bias, which can be made arbitrarily small by taking m sufficiently large.

larger entropy compared to $\hat{\mu}_1$. Therefore, a privacy mechanism is allowed to disclose the value of $\hat{\mu}_1$ for the sake of utility but $\hat{\mu}_2$ must be kept secret for the sake of privacy.

The above example also sheds light on Terry Gross' case: If the average height of Lithuanian women is released using a low-entropy accurate estimator with suitable convergence properties (e.g. $\hat{\mu}_1$), then we do not consider the disclosure of her height as a privacy breach. This is because an adversary who knows the distribution of women's height can predict her height even without access to the mechanism.

Example 3.2. An insurance company has access to an i.i.d medical database X of size n and queries it through $P_{Y|X}$ to obtain (quantized) relative frequencies $\hat{p}_s$ and $\hat{p}_{ns}$ describing the empirical probabilities of developing lung disease for smokers and non-smokers, respectively. Let p_s and p_{ns} denote the true probabilities of developing lung disease for smokers and non-smokers, which can be calculated from the prior distribution P_X. If n is large, then the estimates $\hat{p}_s$ and $\hat{p}_{ns}$ have small entropies and well-approximate the true probabilities.

Now, suppose based on $\hat{p}_s$ and $\hat{p}_{ns}$ the company draws some conclusions about Bob's probability of developing lung disease, and adjusts his insurance premium accordingly. Assuming that $\hat{p}_s$ and $\hat{p}_{ns}$ well-approximate the true probabilities, we do not consider this to be in violation of Bob's privacy (regardless of his participation in the database). This is because the insurance company could have drawn the same conclusions about Bob from the prior P_X even without access to the privacy mechanism.

In essence, the differences between our setup and (Dwork and Naor, 2010) stem from the fundamental principle that if an analyst knows the true distribution P_X on the data, then they should be granted no further utility. Interestingly, the local/global dichotomy also allows us to distinguish between *adversarial* and *non-adversarial* analysts. The non-adversarial analyst Alice is only interested in the value of low-entropy attributes of X, which reflect properties of the population as a whole. If Alice knows P_X, then she gains no further value from interacting with the mechanism $P_{Y|X}$. On the other hand, the adversarial analyst Eve even equipped with P_X is motivated to query X through $P_{Y|X}$ to uncover the value of high-entropy, instance-dependent, and local features of X which she cannot *a priori* predict, even if she possesses arbitrary auxiliary information.

3.2. **Entropy-based Disclosure Prevention.** Equipped with our definitions of privacy and utility, in this subsection, we state the main results of the paper: that (a) disclosing a piece of information (in the sense of Definition 3.3) to one adversary in $\mathcal{P}_{\mathcal{X}}$ is tantamount to disclosing that information to all adversaries in $\mathcal{P}_{\mathcal{X}}$ (Theorem 3.4), and (b) PML provides privacy guarantees according to the local/global dichotomy (Theorem 3.5). In particular, we show that if a mechanism $P_{Y|X}$ satisfies (ϵ, P_X)-PML, then it cannot disclose the value of any attribute of X with entropy greater than ϵ to any adversary with prior belief in the set $P_{\mathcal{X}}$. Afterward, in the spirit of the impossibility result, we prove that when a mechanism discloses the value of an attribute U of X, then it also discloses another attribute of X with smaller prior entropy compared to U. Finally, toward the end of this subsection, we discuss *absolute disclosure prevention*, i.e., we examine the condition ensuring that no attribute of X is disclosed by a privacy mechanism.

We begin by formally defining a notion of *disclosure*. Consider an adversary with prior belief $Q_X \in \mathcal{P}_{\mathcal{X}}$, and let U be an attribute of X. Then, the adversary's prior belief about U is $Q_U = P_{U|X} \circ Q_X$. We may define disclosure as the event that the adversary's belief

about U changes after observing an outcome of the privacy mechanism.[11] That is, disclosure is the event that $Q_U \neq Q_{U|Y=y}$ for some $y \in \mathcal{Y}$, where $Q_{U|Y=y} = P_{U|X} \circ Q_{X|Y=y}$ denotes the adversary's posterior belief about U after observing y. Thus, disclosure prevention requires that Y and U be independent. Clearly, this is a very stringent requirement and may necessitate the independence of X and Y,[12] e.g, if $U = X$. Hence, we instead postulate the following weaker but more intuitive definition that also matches the notions of disclosure considered in (Dwork and Naor, 2010) and (Kifer and Machanavajjhala, 2011).

Definition 3.3 (Disclosure)**.** Let U be an attribute of X. We say that the privacy mechanism $P_{Y|X}$ discloses the value of U to adversary $Q_X \in \mathcal{P}_{\mathcal{X}}$ if $\inf_{y \in \mathcal{Y}} H_\infty(Q_{U|Y=y}) = 0$.

Henceforth, we use the terms "disclosure" and "disclose" in the sense of Definition 3.3. The following theorem asserts that, in fact, we do not need to specify to which adversary a piece of information has been disclosed. This is because disclosures are ubiquitous across $\mathcal{P}_{\mathcal{X}}$.

Theorem 3.4 (Ubiquity of Disclosures)**.** *Let U be an attribute of X. If the privacy mechanism $P_{Y|X}$ discloses the value of U to an adversary $Q_X \in \mathcal{P}_{\mathcal{X}}$, then it also discloses the value of U to all other adversaries in $\mathcal{P}_{\mathcal{X}}$.*

We now exploit Theorem 3.4 to prove that PML-based privacy guarantees prevent disclosing high-entropy attributes of X to all adversaries in $\mathcal{P}_{\mathcal{X}}$.

Theorem 3.5 (Disclosure prevention via PML)**.** *Suppose X is distributed according to P_X, and let U be an attribute of X with entropy $H_\infty(P_U) > \epsilon$, where $\epsilon \geq 0$ and $P_U = P_{U|X} \circ P_X$. If the privacy mechanism $P_{Y|X}$ satisfies (ϵ, P_X)-PML, then $P_{Y|X}$ cannot disclose the value of U to any adversary $Q_X \in \mathcal{P}_{\mathcal{X}}$.*

The above theorem contains a powerful idea: It states that if we protect the data under its true distribution, then we are simultaneously preventing privacy breaches against all adversaries in $\mathcal{P}_{\mathcal{X}}$. Furthermore, Theorem 3.5 demonstrates that the two goals of privacy and utility are not inherently at odds with each other. This is because while PML imposes lower bounds on the remaining uncertainty in the value of high-entropy local attributes of X, it does not directly restrict the remaining uncertainty in the value of low-entropy global attributes of X. Indeed, when the answer to a query describes a feature of X that has very small entropy, it may even be safe to answer it precisely and without any randomness. We give an example of a query answered deterministically in Section 3.3.

It is worth emphasizing that Theorem 3.5 does *not* mean that mechanism $P_{Y|X}$ leaks the same amount of information to all adversaries. In fact, an attribute U of X that has small entropy under the true distribution P_X may have very large entropy according to the belief of adversary Q_X. In this case, a privacy mechanism that discloses the value of U leaks a large amount of information to adversary Q_X, and this leakage is captured by $\ell_{Q_{XY}}(X \to y)$, where $Q_{XY} = P_{Y|X} \times Q_X$. Nevertheless, Theorem 3.5 asserts that we need not be alarmed by the large value of $\ell_{Q_{XY}}(X \to y)$ because despite this large leakage, adversary Q_X will not be able to infer the value of any local features of X. Put differently, while we may use PML *subjectively* to calculate the amount of information leaked to each adversary, the parameter

[11] This is often called Dalenius' desideratum in the literature.

[12] Rassouli and Gündüz (2021) show that under certain conditions it is possible to design $P_{Y|X}$ such that Y is independent of U but correlated with X.

ϵ of the privacy guarantee should be determined and interpreted *objectively* according to our assumptions about the true underlying distribution on the data.

As a converse to Theorem 3.5, we now show that when the mechanism $P_{Y|X}$ discloses the value of an attribute U of X, then we can no longer guarantee privacy for attributes of X with entropies smaller than $H_\infty(P_U)$. In fact, disclosing U inevitably leads to disclosing another attribute of X with a smaller entropy compared to U.

Proposition 3.6. *Suppose X is distributed according to P_X. Assume that the privacy mechanism $P_{Y|X}$ discloses the value of an attribute of X, denoted by U. Then, there exists an attribute of X, denoted by W, satisfying $H_\infty(P_W) < H_\infty(P_U)$ whose value is also disclosed.*

Proposition 3.6 is conceptually similar to the impossibility result (specifically, (Dwork and Naor, 2010, Thm. 3)); yet, it is interpreted differently in our framework: If U is disclosed to provide utility, then U has small entropy and can be estimated accurately using its distribution P_U alone. Since $H_\infty(P_W) < H_\infty(P_U)$, then W can too be estimated accurately using its distribution P_W, even without access to the privacy mechanism. Thus, if disclosing U is not considered as a privacy breach, then disclosing W should not be considered as a privacy breach either. It is also worth mentioning that the proof of Proposition 3.6 requires no assumptions about the mechanism $P_{Y|X}$ other than the fact that it discloses U. Hence, the result holds even if we assume that $P_{Y|X}$ satisfies (ϵ, P_X)-PML with $\epsilon \geq H_\infty(P_U)$.

As the final topic in this subsection, we discuss *absolute disclosure prevention*, i.e., we investigate conditions ensuring that *no* attribute of X can be disclosed by the mechanism $P_{Y|X}$. We show that absolute disclosure prevention can be achieved by mechanisms that have finite leakage capacity (see Definition 2.6). Moreover, we prove that these mechanisms guarantee a lower bound on the remaining uncertainty in the value of all (non-constant) deterministic attributes of X for all adversaries in $\mathcal{P}_\mathcal{X}$.

Theorem 3.7 (Absolute disclosure prevention)**.** *If $P_{Y|X}$ satisfies $C(P_{Y|X}) < \infty$, then for all $P_X \in \mathcal{P}_\mathcal{X}$ no attribute of X can be disclosed by $P_{Y|X}$. Furthermore, given an arbitrary (non-constant) deterministic function of X, denoted by V, the remaining uncertainty in the value of V for adversary $Q_X \in \mathcal{P}_\mathcal{X}$ is at least*

$$H_\infty(Q_{V|Y=y}) \geq \log\left(1 + \frac{\min_x q_X(x)}{1 - \min_x q_X(x)} e^{-C(P_{Y|X})}\right),$$

for all $y \in \mathcal{Y}$.

By Theorem 2.7, a privacy mechanism $P_{Y|X}$ has finite leakage capacity if and only if it satisfies $(\epsilon, \mathcal{P}_\mathcal{X})$-PML with some finite value of ϵ. As such, the above result contains a similar idea to the no-free-lunch theorem of (Kifer and Machanavajjhala, 2011). More precisely, (Kifer and Machanavajjhala, 2011, Thm. 2.1) states that it is not possible to discriminate between different instances of the secret X if we guarantee privacy under all possible distributions on the data. That is, utility is essentially destroyed when we make no assumptions about the data-generating distribution. Here, however, we may have a different take on Theorem 3.7 when viewed through the lens of the local/global dichotomy: Guaranteeing privacy under all possible distributions entails that we no longer can distinguish between local and global features of the data. For example, an attribute U of X may have small entropy under distribution $P_X^{(1)} \in \mathcal{P}_\mathcal{X}$ but large entropy under another distribution $P_X^{(2)} \in \mathcal{P}_\mathcal{X}$. Since no non-trivial attribute of X can have consistently small entropy under

all possible distributions in $\mathcal{P}_{\mathcal{X}}$, then no attribute of X can be considered to capture a property of the whole population. Hence, we inevitably protect all features of X. In other words, when we make no assumptions about the data-generating distribution, then a privacy mechanism provides no utility because there is no utility to be provided.

3.3. **How to Pick ϵ?** According to Theorem 3.5, if a mechanism $P_{Y|X}$ satisfies (ϵ, P_X)-PML, then it cannot disclose the value of any attribute of X with entropy larger than ϵ to any adversary in $\mathcal{P}_{\mathcal{X}}$. Essentially, ϵ describes where (in terms of entropy) we draw the line between global and local features of X, and smaller ϵ implies stricter privacy requirements. We may select ϵ by asking: Which features of X do we consider to be sufficiently easy to guess by an analyst who knows P_X such that they may be disclosed without causing a privacy breach? Conversely, we may ask: Which features of X do we wish to keep secret even from an analyst who knows P_X and what is the entropy of those features? In this subsection, we give a few concrete examples of attributes of X that are disclosed at different values of ϵ. We also argue that ϵ should always remain below the entropy of the data $H_\infty(P_X)$.

First, we establish the existence of an attribute of X which can be disclosed at the smallest ϵ compared to all other attributes of X. Let $p_{\min} \coloneqq \min_x p_X(x)$ and $x_{\min} \in \mathcal{X}$ be a realization of X with probability $p_{\min}$.

Proposition 3.8. *Suppose X is distributed according to P_X. If the privacy mechanism $P_{Y|X}$ satisfies (ϵ, P_X)-PML with $\epsilon < \log \frac{1}{1-p_{\min}}$, then $C(P_{Y|X}) < \infty$. Conversely, for each $\epsilon \geq \log \frac{1}{1-p_{\min}}$ there exists an attribute U of X and a privacy mechanism $P_{Y|X}$ satisfying (ϵ, P_X)-PML that discloses the value of U.*

The second statement in Proposition 3.8 is proved by constructing an attribute of X which requires the smallest privacy cost (i.e., $\epsilon = \log \frac{1}{1-p_{\min}}$) to be disclosed. This attribute describes a binary random variable that determines whether or not X has value $x \neq x_{\min}$. Thus, Proposition 3.8 essentially states that giving an affirmative answer (deterministically) to the query "Is $X \in \mathcal{X} \setminus \{x_{\min}\}$?" induces the smallest privacy cost. Note that an analyst who possesses P_X can correctly predict the answer to this query with probability $1 - p_{\min}$ even without access to the mechanism. On the other hand, Issa et al. (2019, Thm. 1) construct an attribute of X which takes the largest privacy cost to be disclosed. Roughly speaking, (Issa et al., 2019, Thm. 1) shows that an affirmative answer can be given to the query "Is $X \in \{x_{\min}\}$?" when $\epsilon \geq \log \frac{1}{p_{\min}}$. Note that the answer to this query is correctly guessed (without access to the mechanism) with the small probability of $p_{\min}$, and that at $\epsilon = \log \frac{1}{p_{\min}}$ a mechanism is allowed to answer all possible queries about X error-free.

Of course, ϵ should be picked such that no realization of X can be disclosed. That is, $P_{Y|X}$ should not be able to deterministically give an affirmative answer to any query of the form "Is $X \in \{x\}$?" for any $x \in \mathcal{X}$. We call this particularly pernicious type of disclosure *singling out*. When $\epsilon < H_\infty(P_X)$, $P_{Y|X}$ cannot single out the value of X.

Definition 3.9 (Singling out)**.** Suppose X is distributed according to P_X. We say that a privacy mechanism $P_{Y|X}$ singles out the value of X if $\inf_{y \in \mathcal{Y}} H_\infty(P_{X|Y=y}) = 0$.

By noting that X is an attribute of X, we obtain the following corollary of Theorem 3.5.

Corollary 3.10. *Suppose X is distributed according to P_X. If the privacy mechanism $P_{Y|X}$ satisfies (ϵ, P_X)-PML with $\epsilon < H_\infty(P_X)$, then it cannot single out the value of X.*

Thus, when $P_{Y|X}$ has infinite leakage capacity, $H_\infty(P_X)$ must be treated as a strict upper bound on ϵ. In practice, however, $H_\infty(P_X)$ will likely be very large and we should opt for much smaller values of ϵ. We examine this in the example below about a query that could be answered deterministically under favorable conditions.

Example 3.11. Consider a database $X = (D_1, \dots, D_n)$ containing n i.i.d entries. Suppose we want to answer the query "Are there more than m individuals in the database who identify as female?" as accurately as possible but without disclosing the gender of any individual in the database. When $m \ll n$ or $n - m \ll n$ it may be safe to answer this query deterministically and with no randomness at all. To see why, suppose the individuals in this population identify as female with probability $p \in [0.3, 0.7]$. Let S_i be a binary random variable that describes whether or not individual $i \in [n]$ identifies as female, and note that the Markov chain $S_i - D_i - X - Y$ holds. Let $y = 1$ denote an affirmative answer to the query and $y = 0$ denote a negative answer to the query.

First, suppose $\frac{m}{n} \leq p$. In this case, answering deterministically with $y = 1$ causes the information leakage

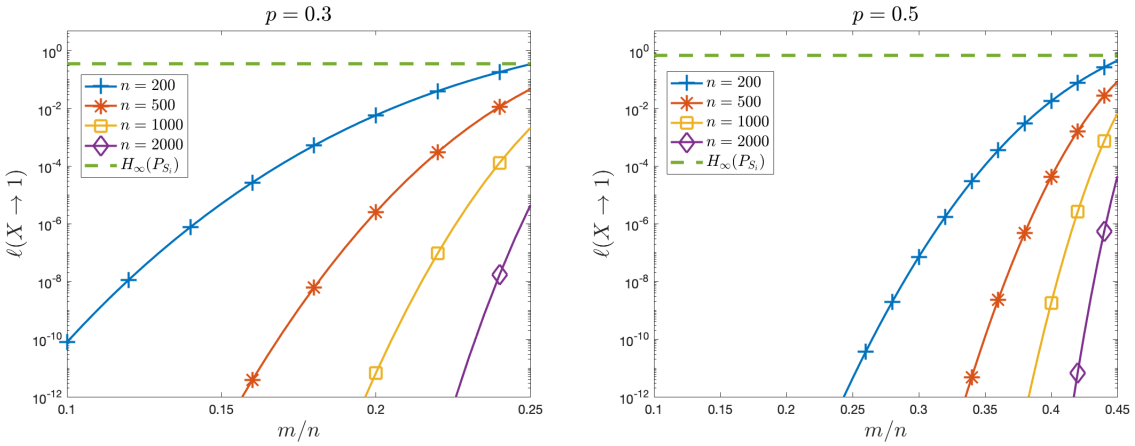
$$\begin{aligned}
\ell_{P_{XY}}(X \to 1) &= \log \frac{\max_{x \in \mathcal{X}} p_{Y|X=x}(1)}{p_Y(1)} \\
&= \log \frac{1}{1 - P_X\left(\{x : x \text{ contains less than or equal to } m \text{ females}\}\right)} \\
&= -\log \left(1 - \sum_{k=0}^{m} \binom{n}{k} p^k \cdot (1-p)^{n-k}\right) \\
&\leq -\log \left(1 - \exp\Big(-nD_{\mathrm{KL}}(\frac{m}{n} \| p)\Big)\right),
\end{aligned}$$

where the last inequality follows from a Chernoff bound on the tail of the Binomial distribution (Hagerup and Rüb, 1990), and $D_{\mathrm{KL}}(q\|r) = q \log \frac{q}{r} + (1-q) \log \frac{1-q}{1-r}$ denotes the KL-divergence between two Bernoulli distributions with parameters $q, r \in (0, 1)$. In Figure 1, we have plotted the above upper bound on $\ell_{P_{XY}}(X \to 1)$ for different values of p and n. It can be observed that when $\frac{m}{n}$ is small, the amount of information leaked by the deterministic query response is several orders of magnitude smaller than $H_\infty(P_{S_i})$. Note that by Theorem 3.5, the gender of no individual will be disclosed by the query response as long as $\ell_{P_{XY}}(X \to 1) < \min_{p \in [0.3, 0.7]} H_\infty(P_{S_i}) = 0.36$. Similarly, when $\frac{m+1}{n} \geq p$, answering the query deterministically with $y = 0$ causes the information leakage

$$\ell(X \to 0) \leq -\log \left(1 - \exp\Big(-nD_{\mathrm{KL}}(1 - \frac{m+1}{n} \| 1 - p)\Big)\right),$$

which is very small when n is large and m is close to n.

In conclusion, ϵ in a PML guarantee is a data-dependent parameter that is easily interpretable in terms of the entropy of the features of X that we allow to be disclosed. This interpretability is a big advantage over many other privacy definitions, including differential privacy, where no clear guidelines exist that explain how small the privacy parameter should be in order to maintain meaningful privacy guarantees (Dwork et al., 2019).



(A) Leakage bounds when $p = 0.3$.

(B) Leakage bounds when $p = 0.5$.

FIGURE 1. Upper bounds on $\ell_{P_{XY}}(X \to 1)$ in Example 3.11 when $p \in \{0.3, 0.5\}$ and $n \in \{200, 500, 1000, 2000\}$.

4. Inferential Database Privacy

In the previous section, we addressed the main criticisms levied against the inferential perspective on privacy and argued that the inferential view can provide a solid foundation for a consistent privacy framework. In this section, we explore how the inferential view fits in with existing notions of database privacy. Undoubtedly, the most widely adopted measure of database privacy is differential privacy, which is formulated based on the notion of indistinguishability (Dwork et al., 2006b). Roughly speaking, a privacy mechanism is differentially private if all possible outcomes are produced with similar probabilities by two databases that agree on all but one entry. Here, we show that even though differential privacy was proposed as a response to the idea that useful inferential privacy guarantees are impossible (Dwork, 2006), it is in fact compatible with the inferential perspective through the paradigm of PML. More precisely, we prove that a privacy mechanism satisfies (pure) differential privacy (a) if and only if conditioned on all but one entry, the mechanism releasing information about the remaining entry satisfies ϵ-PML under all possible (product) distributions on X, or (b) if and only if the mechanism releasing information about each entry satisfies ϵ-PML under all possible product distributions on X.

Next, we proceed to discuss *free-lunch privacy* and its relationship with PML. Kifer and Machanavajjhala (2011) argue that differential privacy implicitly assumes that the entries in a database are independent and propose free-lunch privacy as an alternative definition that avoids assumptions about the underlying data-generating distribution. Here, we establish that a mechanism $P_{Y|X}$ satisfies free-lunch privacy (a) if and only if $P_{Y|X}$ satisfies ϵ-PML under all possible (product) distributions on X, or (b) if and only if the mechanism releasing information about each entry satisfies ϵ-PML under all possible distributions on X.

Given that both differential privacy and free-lunch privacy can be expressed as PML constraints, it follows naturally that mechanisms satisfying either of these definitions also guarantee privacy in the sense of PML. To illustrate this point, we examine the counting query and the Laplace mechanism (Dwork et al., 2006b) through the lens of PML. In doing so, we demonstrate that when the data-generating distribution has large entropy, the privacy

cost of using the Laplace mechanism can be considerably smaller than (up to half of) the differential privacy parameter. Thus, PML-based analysis also has the advantage of precisely characterizing the privacy cost by taking into account the data-generating distribution.

Suppose X is a random variable representing a database containing n entries. Given $i \in [n]$, let D_i be the random variable corresponding to the i-th entry, which takes values in a finite alphabet $\mathcal{D}$. Then, each database (realization) $x = (d_1, \dots, d_n) \in \mathcal{D}^n$ is an n-tuple and X is a sequence of n random variables. Suppose $P_X = P_{D_1,\dots,D_n}$ denotes the distribution according to which databases are drawn from $\mathcal{D}^n$. To obtain the probability distribution describing the i-th entry we marginalize over the remaining $n-1$ entries, that is, for each $d_i \in \mathcal{D}$ and $i \in [n]$ we have

$$p_{D_i}(d_i) = \sum_{d_{-i} \in \mathcal{D}^{n-1}} p_{D_i | D_{-i} = d_{-i}}(d_i)\, p_{D_{-i}}(d_{-i}),$$

where $d_{-i} \coloneqq (d_1, \dots, d_{i-1}, d_{i+1}, \dots, d_n) \in \mathcal{D}^{n-1}$ is a tuple describing the database with its i-th entry removed. Note that this setup is very general in the sense that the entries can be arbitrarily correlated.

Suppose an analyst poses a query to the database whose answer is returned by the privacy mechanism $P_{Y|X}$. Below, we define ϵ-differential privacy[13] in our notation.

Definition 4.1 (Differential privacy)**.** Given $\epsilon \geq 0$, we say that the privacy mechanism $P_{Y|X}$ satisfies ϵ-differential privacy if

$$\sup_{y \in \mathcal{Y}} \max_{\substack{d_i, d_i' \in \mathcal{D}: \\ i \in [n]}} \max_{d_{-i} \in \mathcal{D}^{n-1}} \log \frac{p_{Y | D_i = d_i, D_{-i} = d_{-i}}(y)}{p_{Y | D_i = d_i', D_{-i} = d_{-i}}(y)} \leq \epsilon.$$

Let $\mathcal{P}_{\mathcal{X}}$ denote the set of all distributions with full support on $\mathcal{X} = \mathcal{D}^n$. Note that assuming the prior distributions on X belong to the set $\mathcal{P}_{\mathcal{X}}$ ensures that all conditional probabilities given subsets of the database are well-defined (by making sure that we do not condition on events with probability zero). Furthermore, let $\mathcal{Q}_{\mathcal{X}}$ denote the set of product distributions in $\mathcal{P}_{\mathcal{X}}$, that is, $\mathcal{Q}_{\mathcal{X}} \coloneqq \{P_X \in \mathcal{P}_{\mathcal{X}} \colon P_X = \prod_{i=1}^n P_{D_i}\}$. We now show that differential privacy admits multiple different but equivalent formulations in terms of PML. All the results presented in this section are proved in Appendix B.

Theorem 4.2 (Differential privacy as a PML constraint)**.** *Given $\epsilon \geq 0$, the privacy mechanism $P_{Y|X}$ satisfies ϵ-differential privacy if and only if*

(1) $\sup\limits_{y \in \mathcal{Y}} \sup\limits_{P_X \in \mathcal{P}_{\mathcal{X}}} \max\limits_{\substack{d_{-i} \in \mathcal{D}^{n-1}: \\ i \in [n]}} \ell(D_i \to y \mid d_{-i}) \leq \epsilon$, *or,*

(2) $\sup\limits_{y \in \mathcal{Y}} \sup\limits_{P_X \in \mathcal{Q}_{\mathcal{X}}} \max\limits_{\substack{d_{-i} \in \mathcal{D}^{n-1}: \\ i \in [n]}} \ell(D_i \to y \mid d_{-i}) \leq \epsilon$, *or,*

(3) $\sup\limits_{y \in \mathcal{Y}} \sup\limits_{P_X \in \mathcal{Q}_{\mathcal{X}}} \max\limits_{i \in [n]} \ell(D_i \to y) \leq \epsilon$.

The first formulation of differential privacy in the above theorem is similar to a result of Dwork et al. (2006b). Specifically, (Dwork et al., 2006b, Claim 3) shows that differential privacy is equivalent to *semantic security* (Dwork et al., 2006b, Def. 6), where semantic

[13] Technically, Definition 4.1 describes indistinguishability but it is often taken as the definition of differential privacy. This definition is sometimes called *bounded differential privacy* in works such as (Kifer and Machanavajjhala, 2011).

security is defined by imposing both an upper bound and a lower bound on the posterior-prior ratio of all binary predicates of the data. The above result can then be considered as a generalization of (Dwork et al., 2006b, Claim 3) because it only requires an upper bound on the posterior-prior ratio and D_i is not restricted to be binary.

Next, we define free-lunch privacy in our notation and show how it can be expressed in terms of PML.

Definition 4.3 (Free-lunch privacy (Kifer and Machanavajjhala, 2011, Def. 2.3))**.** Given $\epsilon \geq 0$, we say that the privacy mechanism $P_{Y|X}$ satisfies ϵ-free-lunch privacy if

$$\sup_{y\in\mathcal{Y}} \max_{d^n,\tilde{d}^n\in\mathcal{D}^n} \log \frac{p_{Y|X=d^n}(y)}{p_{Y|X=\tilde{d}^n}(y)} \leq \epsilon.$$

Theorem 4.4 (Free-lunch privacy as a PML constraint)**.** *Given $\epsilon \geq 0$, the privacy mechanism $P_{Y|X}$ satisfies ϵ-free-lunch privacy if an only if*

(1) $\sup_{y\in\mathcal{Y}} \sup_{P_X\in\mathcal{P}_\mathcal{X}} \ell(X \to y) \leq \epsilon$, *or*
(2) $\sup_{y\in\mathcal{Y}} \sup_{P_X\in\mathcal{Q}_\mathcal{X}} \ell(X \to y) \leq \epsilon$, *or*
(3) $\sup_{y\in\mathcal{Y}} \sup_{P_X\in\mathcal{P}_\mathcal{X}} \max_{i\in[n]} \ell(D_i \to y) \leq \epsilon$.

We highlight a few points about the above results. First, note that by the Markov chain $D_i - X - Y$ and the pre-processing inequality for PML (Saeidian et al., 2022b, Lemma 1), $\ell(D_i \to y) \leq \ell(X \to y)$ for all $i \in [n]$, $y \in \mathcal{Y}$ and $P_X \in \mathcal{P}_\mathcal{X}$. Theorem 4.4 then implies that under certain distributions, the amount of information leaking about a single entry can be as large as the information leaking about the whole database. Roughly speaking, this happens when the entropy of the whole dataset is concentrated on a single entry. Second, by comparing (3) in Theorem 4.4 and (1) in Theorem 4.2 we arrive at a similar conclusion to Kifer and Machanavajjhala (2011) and Yang et al. (2015) that the informed adversary assumption may lead to underestimating the information leaking about the entries in the dataset. Nevertheless, this can happen only when the entries in the database are highly correlated. Indeed, if we restrict our attention to product distributions, then by (2) and (3) in Theorem 4.2 the conditional and unconditional leakages become equal. Third, in neither of the above results the supremum is ever actually attained by any distribution in $\mathcal{P}_\mathcal{X}$ or $\mathcal{Q}_\mathcal{X}$ (see Remark B.1). Instead, the proofs construct a sequence of distributions with decreasing (conditional) entropy under which PML converges to the corresponding log-likelihood ratio in the definition of differential privacy or free-lunch privacy. Therefore, when the dataset has large entropy, the amount of information leaking through a privacy mechanism can be noticeably smaller than the ϵ reported by differential privacy or free-lunch privacy. Below, we use this observation to show how incorporating knowledge about the data-generating distribution into our analysis results in a more accurate privacy risk assessment of the counting query and the Laplace mechanism.

4.1. **Laplace Mechanism and the Counting Query.** Here, we discuss a concrete example demonstrating that existing mechanisms are compatible with the type of privacy discussed in this paper. We also show that incorporating assumptions about the prior distribution can lead to tighter bounds on the privacy parameter associated with a mechanism. This is because, as discussed earlier, privacy is easier to achieve when the distribution P_X has large entropy compared to when it has small entropy. Note that certain datasets such as financial

data for fraud detection or health data for studying rare diseases may naturally contain features with very small entropy. However, in many everyday applications, one encounters high-entropy datasets with more balanced probabilities. In these cases, we can save on the privacy cost paid, and ultimately, achieve more utility. Below, we illustrate this for the archetypical example of a counting query that is answered by the Laplace mechanism (Dwork et al., 2006b).

We consider the third characterization of differential privacy in Theorem 4.2 and restrict the set of product distributions from which X may be drawn. Suppose X is an i.i.d database containing n entries. Consider a predicate $f : \mathcal{D} \to \{0, 1\}$ and suppose we want to answer the counting query "What fraction of the entries in the database satisfy $f(d_i) = 1$?". Let $0 \leq c < \frac{1}{2}$ be a constant and assume $P_X \in \mathcal{P}_c^f$, where

$$\mathcal{P}_c^f = \Big\{ P_X \in \mathcal{Q}_{\mathcal{X}} : P_{D_i}(\{d \in \mathcal{D} : f(d) = 1\}) = p \text{ for all } i \in [n] \text{ and } p \in (c, 1-c) \Big\}.$$

That is, we assume that each entry in the database satisfies the predicate f with probability $p \in (c, 1-c)$. Let $\mathrm{Lap}(\mu, b)$ denote the Laplace distribution with mean $\mu \in \mathbb{R}$ and scale parameter $b > 0$. To answer the counting query, the Laplace mechanism returns an outcome according to the distribution $Y \mid X = (d_1, \ldots, d_n) \sim \mathrm{Lap}(\frac{f(d_1) + \ldots + f(d_n)}{n}, b)$ (Dwork et al., 2006b).

Proposition 4.5. *Consider the predicate $f : \mathcal{D} \to \{0, 1\}$. Suppose X is a database of size n drawn according to a distribution $P_X \in \mathcal{P}_c^f$. Let $P_{Y|X}$ denote the Laplace mechanism with scale parameter $b > 0$ answering the counting query corresponding to f. Then, the information leaking about each entry in the database is upper bounded by*

$$\sup_{P_X \in \mathcal{P}_c^f} \sup_{y \in \mathbb{R}} \ell(D_i \to y) \leq \frac{1}{nb} - \log\left((1-c) + c \exp\left(\frac{1}{nb} \right) \right),$$

for all $i \in [n]$.

When nb is large we may use $e^x \geq 1 + x$ and $\log(1 + x) \geq x - \frac{x^2}{2}$ for $x \geq 0$ to obtain the simplified bound

$$\sup_{P_X \in \mathcal{P}_c^f} \sup_{y \in \mathbb{R}} \ell(D_i \to y) \leq \frac{1-c}{nb} + \frac{c^2}{2n^2b^2},$$

for all $i \in [n]$. Observe that $\frac{1}{nb}$ corresponds to the well-known differential privacy parameter of the Laplace mechanism returning the answer to a query with global sensitivity $\frac{1}{n}$ (Dwork et al., 2006b). As expected, the above leakage bound also reduces to $\frac{1}{nb}$ when $c = 0$, describing the situation where P_X can be any i.i.d distribution in $\mathcal{Q}_{\mathcal{X}}$ with arbitrarily small entropy. On the other hand, when c is close to $\frac{1}{2}$, then the privacy parameter is reduced by almost a factor of $\frac{1}{2}$. Hence, this approach allows us to adjust the privacy cost we pay based on the entropy of the data, and ultimately, achieve higher utility.

5. Other Related Works

Definitional works. Apart from differential privacy and its extensions (e.g., (Dwork et al., 2006a; Mironov, 2017; Dwork and Rothblum, 2016; Bun and Steinke, 2016; Dong et al., 2022)), a large number of privacy definitions have been proposed in the literature, e.g.,

differential identifiability (Lee and Clifton, 2012), membership privacy (Li et al., 2013), and Pufferfish privacy (Kifer and Machanavajjhala, 2014). Yang et al. (2015) introduced *Bayesian differential privacy* which generalizes the informed adversary assumption of differential privacy and considers adversaries who *a priori* know an arbitrary subset of the dataset. Yang et al. (2015) also show that differential privacy and Bayesian differential privacy are equivalent when the prior is a product distribution. Bassily et al. (2013) introduced the framework of *coupled-worlds privacy* which, similarly to PML, is a prior-dependent notion of privacy. Coupled-worlds privacy relaxes differential privacy by requiring that neighboring databases remain indistinguishable under a predefined set of priors instead of all possible priors. A similar definition to coupled-worlds privacy is *noiseless privacy* (Bhaskar et al., 2011) whose goal is to provide noise-free answers to certain queries by leveraging the intrinsic uncertainty in the value of a database as described by a prior distribution. This is a similar idea to what we have considered in our paper as we have shown that some attributes (functions) of the sensitive data X can be disclosed error-free by a mechanism satisfying ϵ-PML, while others will be distorted to avoid revealing too much information about X.

We stress that none of the above notions have as clear an operational meaning as PML. In fact, most definitions have been obtained by formalizing some intuitive understanding of privacy, which then may lead to misunderstandings in what they do or do not guarantee.

'Semantics' of differential privacy. Several works have interpreted the guarantees of differential privacy or cast it as a constraint in terms of familiar quantities such as total variation distance or mutual information. For example, Kasiviswanathan and Smith (2014) provided a Bayesian interpretation of differential privacy by showing that the posterior belief of an adversary about the input data does not change much (in terms of total variation distance) whether or not each individual's data is included. Wasserman and Zhou (2010) considered a hypothesis test on the value of a single entry in a database and showed that differential privacy imposes a tradeoff between the Type I and Type II error probabilities. Ghosh and Kleinberg (2016) defined *inferential privacy* as a constraint on how much an analyst's posterior belief can diverge from her prior belief, and study the inferential privacy guarantees of differentially private mechanisms assuming a certain class of prior distributions. Cuff and Yu (2016) showed that differential privacy is equivalent to a constraint on the conditional mutual information of a privacy mechanism; however, this equivalence is in a weaker sense compared to the one we have established using PML. Finally, Kifer et al. (2022) gave an account of frequentist and Bayesian semantics of multiple variants of differential privacy. Their Bayesian semantics, however, rely solely on posterior-to-posterior comparisons, and posterior-to-prior comparisons are deemed unsuitable due to the results of Dwork and Naor (2010) and Kifer and Machanavajjhala (2011). This is exactly the point of view challenged in this paper.

6. Conclusions

In summary, this paper describes a paradigm shift in how privacy is defined that follows from a novel interpretation of the fundamental result of Dwork and Naor (2010) about the impossibility of absolute disclosure prevention. According to the definition of privacy presented here, we must distinguish between the properties of the secret X that are predictable using the prior P_X alone and those properties that can be obtained only through the mechanism $P_{Y|X}$. This is an important distinction to make especially in applications where

the priors are publicly known. For example, it is well-known that a big portion of the human DNA is largely predictable, and intuitively, a mechanism should be able to release this information without it being considered a privacy breach. The advantages of our new paradigm are briefly summarized as follows:

- The view of privacy presented here is inherently Bayesian and inferential. This allows devising privacy-preserving solutions that are adapted to each dataset in terms of entropy, correlations among data points, and so on.
- Privacy guarantees are adapted to the underlying distribution of the data but do not depend on each adversary's perceived information leakage. Essentially, it is possible to distinguish between the useful and necessary information leakage that allows analysts (with inaccurate priors) to learn about the data and the harmful information leakage causing privacy breaches.
- Privacy is rendered an actionable goal and is improved by constructing high-quality estimators of the features of the data with suitable convergence properties. From this point of view, the two goals of privacy and utility actually coincide with each other.
- The framework's central privacy notion, PML, is operationally meaningful and precisely defined, with explicit assumptions about adversaries and the general setup.
- The privacy parameter in PML guarantees is easily interpretable, providing clear guidelines for parameter selection.
- The framework is compatible with and provides insights into existing privacy definitions, including differential privacy.
- The framework enables more flexible mechanism designs. Existing mechanisms can be used efficiently to provide meaningful PML-based guarantees and novel mechanisms can be conceived that adjust their randomness to the entropy of the data for increased utility.

References

M. S. Alvim, K. Chatzikokolakis, C. Palamidessi, and G. Smith. Measuring information leakage using generalized gain functions. In *2012 IEEE 25th Computer Security Foundations Symposium*, pages 265–279, 2012.

M. S. Alvim, K. Chatzikokolakis, A. McIver, C. Morgan, C. Palamidessi, and G. Smith. Additive and multiplicative notions of leakage, and their capacities. In *2014 IEEE 27th Computer Security Foundations Symposium*, pages 308–322, 2014.

R. Bassily, A. Groce, J. Katz, and A. Smith. Coupled-worlds privacy: Exploiting adversarial uncertainty in statistical data privacy. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 439–448. IEEE, 2013.

R. Bhaskar, A. Bhowmick, V. Goyal, S. Laxman, and A. Thakurta. Noiseless database privacy. In *International Conference on the Theory and Application of Cryptology and Information Security*, pages 215–232. Springer, 2011.

M. Bun and T. Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography Conference*, pages 635–658. Springer, 2016.

F. P. Calmon and N. Fawaz. Privacy against statistical inference. In *2012 50th annual Allerton conference on communication, control, and computing (Allerton)*, pages 1401–1408. IEEE, 2012.

P. Cuff and L. Yu. Differential Privacy as a Mutual Information Constraint. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 43–54, Vienna Austria, Oct. 2016. ACM. ISBN 978-1-4503-4139-4.

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

J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 429–438. IEEE, 2013.

C. Dwork. Differential privacy. In *Automata, Languages and Programming*, pages 1–12. Springer Berlin Heidelberg, 2006. ISBN 978-3-540-35908-1.

C. Dwork and M. Naor. On the difficulties of disclosure prevention in statistical databases or the case for differential privacy. *Journal of Privacy and Confidentiality*, 2(1), 2010.

C. Dwork and G. N. Rothblum. Concentrated differential privacy. *arXiv preprint arXiv:1603.01887*, 2016.

C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. Our Data, Ourselves: Privacy Via Distributed Noise Generation. In S. Vaudenay, editor, *Advances in Cryptology - EUROCRYPT 2006*, volume 4004, pages 486–503. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006a. ISBN 978-3-540-34546-6 978-3-540-34547-3.

C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006b.

C. Dwork, A. Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

C. Dwork, N. Kohli, and D. Mulligan. Differential privacy in practice: Expose your epsilons! *Journal of Privacy and Confidentiality*, 9(2), 2019.

N. Fernandes, A. McIver, and P. Sadeghi. Explaining epsilon in differential privacy through the lens of information theory. *arXiv preprint arXiv:2210.12916*, 2022.

A. Ghosh and R. Kleinberg. Inferential privacy guarantees for differentially private mechanisms. *arXiv preprint arXiv:1603.01508*, 2016.

T. Hagerup and C. Rüb. A guided tour of chernoff bounds. *Information processing letters*, 33(6):305–308, 1990.

X. He, A. Machanavajjhala, and B. Ding. Blowfish privacy: Tuning privacy-utility trade-offs using policies. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, pages 1447–1458, 2014.

I. Issa, A. B. Wagner, and S. Kamath. An operational approach to information leakage. *IEEE Transactions on Information Theory*, 66(3):1625–1657, 2019.

B. Jiang, M. Seif, R. Tandon, and M. Li. Context-aware local information privacy. *IEEE Transactions on Information Forensics and Security*, 2021.

S. P. Kasiviswanathan and A. Smith. On the 'semantics' of differential privacy: A Bayesian formulation. *Journal of Privacy and Confidentiality*, 6(1), 2014.

D. Kifer and A. Machanavajjhala. No free lunch in data privacy. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, pages 193–204, 2011.

D. Kifer and A. Machanavajjhala. A rigorous and customizable framework for privacy. In *Proceedings of the 31st ACM SIGMOD-SIGACT-SIGAI symposium on Principles of Database Systems*, pages 77–88, 2012.

D. Kifer and A. Machanavajjhala. Pufferfish: A framework for mathematical privacy definitions. *ACM Transactions on Database Systems*, 39(1):1–36, Jan. 2014. ISSN 0362-5915, 1557-4644.

D. Kifer, J. M. Abowd, R. Ashmead, R. Cumings-Menon, P. Leclerc, A. Machanavajjhala, W. Sexton, and P. Zhuravlev. Bayesian and frequentist semantics for common variations of differential privacy: Applications to the 2020 census. *arXiv preprint arXiv:2209.03310*,

2022.

J. Lee and C. Clifton. Differential identifiability. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '12*, page 1041, Beijing, China, 2012. ACM Press. ISBN 978-1-4503-1462-6.

N. Li, W. Qardaji, D. Su, Y. Wu, and W. Yang. Membership privacy: A unifying framework for privacy definitions. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security - CCS '13*, pages 889–900, Berlin, Germany, 2013. ACM Press. ISBN 978-1-4503-2477-9.

C. Liu, S. Chakraborty, and P. Mittal. Dependence makes you vulnberable: Differential privacy under dependent tuples. In *NDSS*, volume 16, pages 21–24, 2016.

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

B. Rassouli and D. Gündüz. On perfect privacy. *IEEE Journal on Selected Areas in Information Theory*, 2(1):177–191, 2021.

A. Rényi. On measures of entropy and information. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, volume 1. Berkeley, California, USA, 1961.

W. Rudin. *Real and Complex Analysis*. McGraw-Hill, May 1986. ISBN 0070542341.

S. Saeidian, G. Cervia, T. J. Oechtering, and M. Skoglund. Pointwise maximal leakage. In *2022 IEEE International Symposium on Information Theory (ISIT)*, pages 626–631, 2022a. doi: 10.1109/ISIT50566.2022.9834814.

S. Saeidian, G. Cervia, T. J. Oechtering, and M. Skoglund. Pointwise maximal leakage. *arXiv preprint arXiv:2205.04935*, 2022b.

S. Saeidian, G. Cervia, T. J. Oechtering, and M. Skoglund. Pointwise maximal leakage on general alphabets. *arXiv preprint arXiv:2304.07722*, 2023.

M. C. Tschantz, S. Sen, and A. Datta. SoK: Differential Privacy as a Causal Property. In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 354–371, San Francisco, CA, USA, May 2020. IEEE. ISBN 978-1-72813-497-0.

T. van Erven and P. Harremoës. Rényi divergence and Kullback-Leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014.

S. L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.

L. Wasserman and S. Zhou. A Statistical Framework for Differential Privacy. *Journal of the American Statistical Association*, 105(489):375–389, Mar. 2010. ISSN 0162-1459, 1537-274X.

B. Yang, I. Sato, and H. Nakagawa. Bayesian Differential Privacy on Correlated Data. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pages 747–762, Melbourne Victoria Australia, May 2015. ACM. ISBN 978-1-4503-2758-9.

T. Zhu, P. Xiong, G. Li, and W. Zhou. Correlated differential privacy: Hiding information in non-iid data set. *IEEE Transactions on Information Forensics and Security*, 10(2): 229–242, 2014.

Appendix A. Proofs for Section 3

A.1. **Proof of Theorem 3.4.** Consider the Markov chain $U - X - Y$ and suppose $P_{Y|X}$ discloses the value of U to adversary $Q_X \in \mathcal{P}_{\mathcal{X}}$. Fix $R_X \in \mathcal{P}_{\mathcal{X}}$. First, we argue that since Q_X and R_X are mutually absolutely continuous, then the posterior distributions $Q_{X|Y=y}$ and $R_{X|Y=y}$ are also mutually absolutely continuous for all $y \in \mathcal{Y}$. Let $f(x) = \frac{r_X(x)}{q_X(x)}$ denote the Radon-Nikodym derivate of R_X with respect to Q_X and observe that $f(x) > 0$ for all $x \in \mathcal{X}$. Fix an arbitrary $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. We have

$$\begin{aligned} r_{X|Y=y}(x) &= \frac{p_{Y|X=x}(y) \cdot r_X(x)}{r_Y(y)} \\ &= \frac{p_{Y|X=x}(y) \cdot f(x) \cdot q_X(x)}{r_Y(y)} \\ &= \frac{q_{X|Y=y}(x) \cdot f(x) \cdot q_Y(y)}{r_Y(y)} \\ &= q_{X|Y=y}(x) \cdot g(x,y), \end{aligned}$$

where $g(x,y) \coloneqq \frac{f(x) \cdot q_Y(y)}{r_Y(y)}$ is strictly positive for all $x \in \mathcal{X}$ and all $y \in \mathcal{Y}$. Thus, if $q_{X|Y=y}(x)$ is positive, then so is $r_{X|Y=y}(x)$ and vice versa, proving that the posterior distributions $Q_{X|Y=y}$ and $R_{X|Y=y}$ are mutually absolutely continuous for all $y \in \mathcal{Y}$. Next, we note that $g(x,y)$ is bounded above because

$$\begin{aligned} \max_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} g(x,y) &= \left(\max_{x \in \mathcal{X}} f(x) \right) \sup_{y \in \mathcal{Y}} \frac{q_Y(y)}{r_Y(y)} \\ &= \left(\max_{x \in \mathcal{X}} f(x) \right) \exp\Big(D_\infty(Q_Y \| R_Y) \Big) \\ &\leq \left(\max_{x \in \mathcal{X}} f(x) \right) \exp\Big(D_\infty(Q_X \| R_X) \Big), \end{aligned}$$

where the inequality is due to the data-processing inequality for Rényi divergence (van Erven and Harremoës, 2014, Thm. 9). The Rényi divergence $D_\infty(Q_X \| R_X)$ is also finite since Q_X and R_X are mutually absolutely continuous. Let $c > 0$ be a constant satisfying $\max_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} g(x,y) < c$.

Fix a small $\varepsilon > 0$ and an outcome $y \in \mathcal{Y}$ with $H_\infty(Q_{U|Y=y}) < \varepsilon$. Then, there exists $u^* \in \mathcal{U}$ such that $q_{U|Y=y}(u^*) > e^{-\varepsilon}$, which in turn implies that $q_{U|Y=y}(u) < 1 - e^{-\varepsilon}$ for all $u \neq u^*$. Now, for all $u \neq u^*$ we can write

$$\begin{aligned} r_{U|Y=y}(u) &= \sum_{x \in \mathcal{X}} p_{U|X=x}(u) \cdot r_{X|Y=y}(x) \\ &= \sum_{x \in \mathcal{X}} p_{U|X=x}(u) \cdot g(x,y) \cdot q_{X|Y=y}(x) \\ &\leq \Big(\max_{x \in \mathcal{X}} g(x,y) \Big) \sum_{x \in \mathcal{X}} p_{U|X=x}(u) \cdot q_{X|Y=y}(x) \\ &= \Big(\max_{x \in \mathcal{X}} g(x,y) \Big) \, q_{U|Y=y}(u) \end{aligned}$$

$$< \Big(\max_{x\in\mathcal{X}} g(x,y)\Big)\,(1-e^{-\varepsilon})$$
$$< c\cdot(1-e^{-\varepsilon}).$$

Thus, we get $r_{U|Y=y}(u^*) = 1 - \sum_{u\neq u^*} r_{U|Y=y}(u) > 1 - \Big(|\mathcal{U}|-1\Big)\cdot c\cdot(1-e^{-\varepsilon})$. Finally, taking $\varepsilon\to 0$ yields $r_{U|Y=y}(u^*)\to 1$ and we conclude that $\inf_{y\in\mathcal{Y}} H_\infty(R_{U|Y=y}) = 0$. In other words, $P_{Y|X}$ discloses the value of U to adversary $R_X\in\mathcal{P}_\mathcal{X}$.

A.2. **Proof of Theorem 3.5.** Fix some U satisfying the Markov chain $U-X-Y$ with entropy $H_\infty(P_U)>\epsilon$, where $P_U = P_{U|X}\circ P_X$. First, consider an adversary with prior belief P_X. Let $P_{UY} = (P_{U|X}\times P_{Y|X})\circ P_X$ denote the joint distribution of U and Y. Fix an arbitrary $y\in\mathcal{Y}$. We can write

$$\begin{aligned}
\ell_{P_{UY}}(U\to y) &= \log \max_{u\in\mathrm{supp}(P_U)} \frac{p_{U|Y=y}(u)}{p_U(u)}\\
&\geq \log \max_{u\in\mathrm{supp}(P_U)} p_{U|Y=y}(u) + \log\frac{1}{\max_{u\in\mathrm{supp}(P_U)} p_U(u)}\\
&\geq \log \max_{u\in\mathrm{supp}(P_{U|Y=y})} p_{U|Y=y}(u) + H_\infty(P_U) \qquad \text{(A.1a)}\\
&= H_\infty(P_U) - H_\infty(P_{U|Y=y}),
\end{aligned}$$

where (A.1a) is due to the fact that $\mathrm{supp}(P_{U|Y=y})\subseteq\mathrm{supp}(P_U)$ for all $y\in\mathcal{Y}$. That is, we have

$$\begin{aligned}
H_\infty(P_{U|Y=y}) &\geq H_\infty(P_U) - \ell_{P_{UY}}(U\to y)\\
&\geq H_\infty(P_U) - \ell_{P_{XY}}(X\to y),
\end{aligned} \qquad \text{(A.2)}$$

where the second inequality follows from the pre-processing lemma for PML (Saeidian et al., 2022b, Lemma 1). Now, assuming that $P_{Y|X}$ satisfies (ϵ, P_X)-PML, taking the supremum over $y\in\mathcal{Y}$ yields

$$\inf_{y\in\mathcal{Y}} H_\infty(P_{U|Y=y}) \geq H_\infty(P_U) - \sup_{y\in\mathcal{Y}} \ell_{P_{XY}}(X\to y) > 0. \qquad \text{(A.3)}$$

Therefore, $P_{Y|X}$ cannot disclose the value of U to adversary P_X. Finally, by Theorem 3.4, $P_{Y|X}$ cannot disclose the value of U to any adversary in $\mathcal{P}_\mathcal{X}$.

A.3. **Proof of Proposition 3.6.** Suppose U is a random variable taking values in the set $\mathcal{U}=\{1,\dots,k\}$. Fix a small $\varepsilon>0$ and an outcome $y\in\mathcal{Y}$ satisfying $H_\infty(P_{U|Y=y})<\varepsilon$. Then, there exists $u\in\mathcal{U}$ with $p_{U|Y=y}(u) > e^{-\varepsilon}$. For simplicity, let this be $u=1$. We now construct an attribute of X with entropy smaller than $H_\infty(P_U)$ whose value is also disclosed by $P_{Y|X}$. Let W be a random variable with alphabet $\mathcal{W}=\mathcal{U}$ defined by the conditional pmf

$$p_{W|U=1}(w) = \begin{cases} 1, & \text{if } w=1,\\ 0, & \text{if } w\neq 1,\end{cases}$$

and,

$$p_{W|U=i}(w) = \begin{cases} \lambda, & \text{if } w=1,\\ 1-\lambda, & \text{if } w=i,\\ 0, & \text{otherwise,}\end{cases} \qquad \text{for } i=2,\dots,k,$$

where $0 < \lambda < 1$. Let $P_W = P_{W|U} \circ P_U$ and $P_{W|Y=y} = P_{W|U} \circ P_{U|Y=y}$. Observe that $p_W(1) = \lambda + (1-\lambda)p_U(1)$, and $p_W(i) = (1-\lambda)p_U(i)$ for $i = 2, \ldots, k$. Thus, if

$$\lambda > \frac{\max_{u\in[k]} p_U(u) - p_U(1)}{1 - p_U(1)},$$

then $p_W(1) > \max_{u\in[k]} p_U(u)$, which in turn, yields $H_\infty(P_W) < H_\infty(P_U)$. Finally, we have

$$\begin{aligned} p_{W|Y=y}(1) &= \sum_{u\in\mathcal{U}} p_{W|U=u}(1)\, p_{U|Y=y}(u) \\ &\geq p_{W|U=1}(1)\, p_{U|Y=y}(1) \\ &> e^{-\varepsilon}, \end{aligned} \tag{A.4}$$

which implies that $H_\infty(P_{W|Y=y}) < \varepsilon$. Taking $\varepsilon \to 0$, we conclude that $P_{Y|X}$ discloses the value of W.

A.4. **Proof of Theorem 3.7.** Consider an adversary $Q_X \in \mathcal{P}_\mathcal{X}$ and suppose $C(P_{Y|X}) < \infty$. We prove the theorem by contradiction. Suppose $P_{Y|X}$ discloses the value of an attribute of X, denoted by U. Then, for each $\varepsilon > 0$ there exists $y \in \mathcal{Y}$ such that $H_\infty(Q_{U|Y=y}) < \varepsilon$, or equivalently, $q_{U|Y=y}(u) > e^{-\varepsilon}$ for some $u \in \mathcal{U}$. Denote this outcome by u_1. By Bayes' theorem, we have

$$q_{Y|U=u_1}(y) = \frac{q_{U|Y=y}(u_1) q_Y(y)}{q_U(u_1)} > \frac{q_Y(y)}{q_U(u_1)} \cdot e^{-\varepsilon}.$$

On the other hand, we also have

$$q_{Y|U=u_1}(y) = \sum_{x\in\mathcal{X}} p_{Y|X=x}(y) q_{X|U=u_1}(x) \leq \max_x p_{Y|X=x}(y),$$

hence, we get $\max_x p_{Y|X=x}(y) > \frac{q_Y(y)}{q_U(u_1)} \cdot e^{-\varepsilon}$. Furthermore, for all $u \neq u_1$ we have $q_{U|Y=y}(u) < 1 - e^{-\varepsilon}$. Let u_2 be one such outcome. Once again, Bayes' theorem yeilds

$$\sum_{x\in\mathcal{X}} p_{Y|X=x}(y) q_{X|U=u_2}(x) = q_{Y|U=u_2}(y) = \frac{q_{U|Y=y}(u_2) q_Y(y)}{q_U(u_2)} < \frac{q_Y(y)}{q_U(u_2)}(1 - e^{-\varepsilon})$$

which, in turn, implies that $p_{Y|X=x}(y) q_{X|U=u_2}(x) < \frac{q_Y(y)}{q_U(u_2)}(1 - e^{-\varepsilon})$ for all $x \in \mathcal{X}$. Now, since $\sum_x q_{X|U=u_2}(x) = 1$, $q_{X|U=u_2}(x)$ must be strictly positive for at least one $x \in \mathcal{X}$. Let $x^* \in \mathcal{X}$ be one such outcome. Hence, we get

$$p_{Y|X=x^*}(y) < \frac{q_Y(y)}{q_U(u_2) \cdot q_{X|U=u_2}(x^*)}(1 - e^{-\varepsilon}).$$

Finally, we get

$$\exp\big(C(P_{Y|X})\big) > \frac{\max_x p_{Y|X=x}(y)}{p_{Y|X=x^*}(y)} > \frac{q_U(u_2) \cdot q_{X|U=u_2}(x^*)}{q_U(u_1)} \cdot \frac{e^{-\varepsilon}}{1 - e^{-\varepsilon}} > c \cdot \frac{e^{-\varepsilon}}{1 - e^{-\varepsilon}},$$

where $c > 0$ is a suitably small constant. Then, by letting $\varepsilon \to 0$, we conclude that the capacity $C(P_{Y|X})$ is infinite which is a contradiction. This proves the first statement.

To prove the second statement, suppose V is a deterministic function of X which is induced by the kernel $P_{V|X}$ and takes values in the set $\mathcal{V}$. Fix an arbitrary $v \in \mathcal{V}$ and define $\mathcal{X}_v \coloneqq \{x \in \mathcal{X} : p_{V|X=x}(v) = 1\}$. Note that $p_{V|X=x}(v) = 0$ for all $x \notin \mathcal{X}_v$. Fix an arbitrary $y \in \mathcal{Y}$ and let $r_{\min} = \min_x p_{Y|X=x}(y)$ and $r_{\max} = \max_x p_{Y|X=x}(y)$. Observe that

$\exp(C(P_{Y|X})) \geq \frac{r_{\max}}{r_{\min}}$. Let $Q_{VY} = (P_{V|X} \times P_{Y|X}) \circ Q_X$ denote the joint distribution of V and Y. We can write

$$\begin{aligned} q_{V|Y=y}(v) = \frac{q_{VY}(v,y)}{q_Y(y)} &= \frac{\sum_{x\in\mathcal{X}} p_{V|X=x}(v)\, p_{Y|X=x}(y)\, q_X(x)}{\sum_{x\in\mathcal{X}} p_{Y|X=x}(y)\, q_X(x)} \\ &= \frac{\sum_{x\in\mathcal{X}_v} p_{Y|X=x}(y) q_X(x)}{\sum_{x\in\mathcal{X}} p_{Y|X=x}(y) q_X(x)} \\ &= \frac{1}{1+\frac{\sum_{x\notin\mathcal{X}_v} p_{Y|X=x}(y) q_X(x)}{\sum_{x\in\mathcal{X}_v} p_{Y|X=x}(y) q_X(x)}} \\ &\leq \frac{1}{1+\frac{r_{\min}\ (1-Q_X(\mathcal{X}_v))}{r_{\max}\ Q_X(\mathcal{X}_v)}} \\ &\leq \frac{1}{1+\frac{\min_x q_X(x)}{\exp(C(P_{Y|X}))(1-\min_x q_X(x))}}. \end{aligned}$$

Thus, we get

$$\begin{aligned} H_\infty(Q_{V|Y=y}(v)) &= \log \frac{1}{\max_v q_{V|Y=y}(v)} \\ &\geq \log\left(1+\frac{\min_x q_X(x)}{1-\min_x q_X(x)} e^{-C(P_{Y|X})}\right). \end{aligned}$$

A.5. **Proof of Proposition 3.8.** Suppose $C(P_{Y|X}) = \infty$. Then, for each $\varepsilon > 0$, there exists $y_\varepsilon \in \mathcal{Y}$ such that

$$\frac{\max_x p_{Y|X=x}(y_\varepsilon)}{\min_x p_{Y|X=x}(y_\varepsilon)} \geq \frac{1}{\varepsilon}.$$

Let $\bar{x}_\varepsilon \in \arg\max_x p_{Y|X=x}(y_\varepsilon)$ and $\underline{x}_\varepsilon \in \arg\min_x p_{Y|X=x}(y_\varepsilon)$. We have

$$\begin{aligned} \sup_{y\in\mathcal{Y}} \ell_{P_{XY}}(X\to y) &= \sup_{y\in\mathcal{Y}} \log \frac{\max_{x\in\mathcal{X}} p_{Y|X=x}(y)}{p_Y(y)} \\ &\geq \sup_{\varepsilon>0} \log \frac{p_{Y|X=\bar{x}_\varepsilon}(y_\varepsilon)}{p_{Y|X=\underline{x}_\varepsilon}(y_\varepsilon)\, p_X(\underline{x}_\varepsilon) + \sum_{x\neq\underline{x}_\varepsilon} p_{Y|X=x}(y_\varepsilon)\, p_X(x)} \\ &\geq \sup_{\varepsilon>0} \log \frac{p_{Y|X=\bar{x}_\varepsilon}(y_\varepsilon)}{\varepsilon\, p_{Y|X=\bar{x}_\varepsilon}(y_\varepsilon)\, p_X(\underline{x}_\varepsilon) + \sum_{x\neq\underline{x}_\varepsilon} p_{Y|X=\bar{x}_\varepsilon}(y_\varepsilon)\, p_X(x)} \\ &= \sup_{\varepsilon>0} \log \frac{1}{\varepsilon\ p_X(\underline{x}_\varepsilon) + \sum_{x\neq\underline{x}_\varepsilon} p_X(x)} \\ &\geq \log \frac{1}{1-p_{\min}}. \end{aligned}$$

Therefore, no mechanism with infinite leakage capacity can satisfy (ϵ, P_X)-PML with $\epsilon < \log\frac{1}{1-p_{\min}}$.

To prove the second part of the statement, it suffices to construct a mechanism $P_{Y|X}$ satisfying $\log \frac{1}{1-p_{\min}}$-PML, and an attribute U of X which is disclosed by an outcome of $P_{Y|X}$. Consider the binary random variable $U(X) = \mathbb{1}_{\mathcal{X}\setminus\{x_{\min}\}}(X)$ which is a deterministic function of X.

The posterior distribution $P_{X|U}$ is given by

$$p_{X|U=0}(x) = \begin{cases} 1, & \text{if } x = x_{\min}, \\ 0, & \text{if } x \neq x_{\min}, \end{cases}$$

and

$$p_{X|U=1}(x) = \begin{cases} 0, & \text{if } x = x_{\min}, \\ \frac{P_X(x)}{1-p_{\min}}, & \text{if } x \neq x_{\min}. \end{cases}$$

Let $\alpha > 0$ be a small constant. Suppose Y be a binary random variable induced by the privacy mechanism $P_{Y|X}$ defined as

$$p_{Y|X=x}(0) = \begin{cases} 0, & \text{if } x = x_{\min}, \\ \alpha, & \text{if } x \neq x_{\min}, \end{cases}$$

and,

$$p_{Y|X=x}(1) = \begin{cases} 1, & \text{if } x = x_{\min}, \\ 1-\alpha, & \text{if } x \neq x_{\min}. \end{cases}$$

Then, we have

$$\ell(X \to 0) = \log \frac{1}{1-p_{\min}},$$
$$\ell(X \to 1) = \log \frac{1}{1-\alpha(1-p_{\min})}.$$

Note that for small enough α we have $\ell(X \to 0) > \ell(X \to 1)$. Hence, $P_{Y|X}$ satisfies $\log \frac{1}{1-p_{\min}}$-PML. We now verify that $P_{Y|X}$ discloses the value of U. Let $P_{Y|U} = P_{Y|X} \circ P_{X|U}$. We have

$$p_{Y|U=u}(0) = \sum_x p_{Y|X=x}(0) p_{X|U=u}(x) = \begin{cases} 0, & \text{if } u = 0, \\ \alpha, & \text{if } u = 1. \end{cases}$$

That is, if the adversary observes $y = 0$ she will be certain that U has value $u = 1$. Hence, the privacy mechanism $P_{Y|X}$ discloses the value of U, which completes the proof.

Appendix B. Proofs for Section 4

B.1. **Proof of Theorem 4.2.** Fix an arbitrary $i \in [n]$, $y \in \mathcal{Y}$, and $P_X \in \mathcal{P}_{\mathcal{X}}$. We have

$$\max_{d_{-i} \in \mathcal{D}^{n-1}} \exp\Big(\ell(D_i \to y \mid d_{-i})\Big)$$
$$= \max_{d_{-i} \in \mathcal{D}^{n-1}} \max_{d_i \in \operatorname{supp}(P_{D_i|D_{-i}=d_{-i}})} \frac{p_{Y|D_{-i}=d_{-i},D_i=d_i}(y)}{p_{Y|D_{-i}=d_{-i}}(y)} \tag{B.1a}$$
$$= \max_{d_{-i} \in \mathcal{D}^{n-1}} \max_{d_i \in \mathcal{D}} \frac{p_{Y|D_{-i}=d_{-i},D_i=d_i}(y)}{p_{Y|D_{-i}=d_{-i}}(y)} \tag{B.1b}$$

$$
\begin{aligned}
&= \max_{d_{-i}\in\mathcal{D}^{n-1}} \max_{d_i\in\mathcal{D}} \frac{p_{Y|D_{-i}=d_{-i},D_i=d_i}(y)}{\sum_{d_i'\in\mathcal{D}} p_{Y|D_{-i}=d_{-i},D_i=d_i'}(y)\; p_{D_i|D_{-i}=d_{-i}}(d_i')} \\
&\le \max_{d_{-i}\in\mathcal{D}^{n-1}} \max_{d_i\in\mathcal{D}} \frac{p_{Y|D_{-i}=d_{-i},D_i=d_i}(y)}{\Big(\min_{d_i'\in\mathcal{D}} p_{Y|D_{-i}=d_{-i},D_i=d_i'}(y)\Big) \sum_{d_i'\in\mathcal{D}} p_{D_i|D_{-i}=d_{-i}}(d_i')} \qquad \text{(B.1c)} \\
&= \max_{d_{-i}\in\mathcal{D}^{n-1}} \max_{d_i\in\mathcal{D}} \frac{p_{Y|D_{-i}=d_{-i},D_i=d_i}(y)}{\min_{d_i'\in\mathcal{D}} p_{Y|D_{-i}=d_{-i},D_i=d_i'}(y)} \\
&= \max_{d_{-i}\in\mathcal{D}^{n-1}} \max_{d_i,d_i'\in\mathcal{D}} \frac{p_{Y|D_{-i}=d_{-i},D_i=d_i}(y)}{p_{Y|D_{-i}=d_{-i},D_i=d_i'}(y)},
\end{aligned}
$$

where (B.1a) follows from Definition 2.4, and (B.1b) uses the fact that $\operatorname{supp}(P_{D_i|D_{-i}=d_{-i}}) = \mathcal{D}$ for each $P_X \in \mathcal{P}_{\mathcal{X}}$.

Next, we show that the above inequality holds with equality for a product distribution $P_X^* \in \mathcal{Q}_{\mathcal{X}}$. This then proves that (1) and (2) in the statment of the theorem are equivalent to each other and to differential privacy. Let $\varepsilon > 0$ be a small constant. Suppose $P_X^* = \Pi_{i=1}^n P_{D_i}^*$, where

$$
p_{D_i}^*(d_i') \coloneqq \begin{cases} 1-\varepsilon, & \text{for some } d_i' \in \arg\min_{\tilde{d}_i\in\mathcal{D}} p_{Y|D_{-i}=d_{-i},D_i=\tilde{d}_i}(y), \\ \frac{\varepsilon}{|\mathcal{D}|-1}, & \text{otherwise.} \end{cases}
$$

Then, $\sum_{d_i'\in\mathcal{D}} p_{Y|D_{-i}=d_{-i},D_i=d_i'}(y)\; p_{D_i}^*(d_i') \to \min_{d_i'\in\mathcal{D}} p_{Y|D_{-i}=d_{-i},D_i=d_i'}(y)$ as $\varepsilon \to 0$. Thus, inequality (B.1c) holds with equality for P_X^*.

Now, we show that (3) in the statement of the theorem is also equivalent to differential privacy. Fix an arbitrary $i \in [n]$ and $y \in \mathcal{Y}$. Note that each $P_X \in \mathcal{Q}_{\mathcal{X}}$ can be written as $P_X = P_{D_i} \times P_{D_{-i}}$; hence, we can optimize over P_{D_i} and $P_{D_{-i}}$ separately:

$$
\begin{aligned}
\sup_{P_{D_{-i}}} \sup_{P_{D_i}} \exp\Big(\ell(D_i \to y)\Big) &= \sup_{P_{D_{-i}}} \max_{d_i,d_i'} \frac{p_{Y|D_i=d_i}(y)}{p_{Y|D_i=d_i'}(y)} \qquad \text{(B.2a)} \\
&= \max_{d_i,d_i'} \sup_{P_{D_{-i}}} \frac{\sum_{d_{-i}} p_{Y|D_i=d_i,D_{-i}=d_{-i}}(y)\; p_{D_{-i}}(d_{-i})}{\sum_{d_{-i}} p_{Y|D_i=d_i',D_{-i}=d_{-i}}(y)\; p_{D_{-i}}(d_{-i})} \\
&\le \max_{d_i,d_i'} \max_{d_{-i}} \frac{p_{Y|D_i=d_i,D_{-i}=d_{-i}}(y)}{p_{Y|D_i=d_i',D_{-i}=d_{-i}}(y)}, \qquad \text{(B.2b)}
\end{aligned}
$$

where (B.2a) is due to Theorem 2.7. To show that inequality (B.2b) can be attained, for fixed d_i and d_i' let

$$
d_{-i}^* = (d_1^*, \dots, d_{i-1}^*, d_{i+1}^*, \dots, d_n^*) \in \arg\max_{\tilde{d}_{-i}} \frac{p_{Y|D_i=d_i,D_{-i}=\tilde{d}_{-i}}(y)}{p_{Y|D_i=d_i',D_{-i}=\tilde{d}_{-i}}(y)}.
$$

Consider the pmf $q_{D_j}^*$ defined by

$$
q_{D_j}^*(d_j) \coloneqq \begin{cases} 1-\varepsilon, & d_j = d_j^*, \\ \frac{\varepsilon}{|\mathcal{D}|-1}, & \text{otherwise,} \end{cases} \qquad \text{(B.3)}
$$

for $j \neq i$. Let $q^*_{D_{-i}} = \prod_{j\neq i} q^*_{D_j}$ which satisfies $q^*_{D_{-i}}(d^*_{-i}) = (1-\varepsilon)^{n-1}$, and $q^*_{D_{-i}}(d_{-i}) \leq \frac{\varepsilon}{|\mathcal{D}|-1}(1-\varepsilon)^{n-2}$ for all $d_{-i} \neq d^*_{-i}$. Then, for fixed n,

$$\frac{\sum_{d_{-i}} p_{Y|D_i=d_i,D_{-i}=d_{-i}}(y)\; q^*_{D_{-i}}(d_{-i})}{\sum_{d_{-i}} p_{Y|D_i=d'_i,D_{-i}=d_{-i}}(y)\; q^*_{D_{-i}}(d_{-i})} \to \max_{d_{-i}} \frac{p_{Y|D_i=d_i,D_{-i}=d_{-i}}(y)}{p_{Y|D_i=d'_i,D_{-i}=d_{-i}}(y)},$$

as $\varepsilon \to 0$. Thus, inequality (B.2b) holds with equality for distribution $Q^*_{D_{-i}}$, which completes the proof.

□

Remark B.1. It is important to note that in all of the formulations above the supremum is never actually attained by any distribution in $\mathcal{P}_\mathcal{X}$ or $\mathcal{Q}_\mathcal{X}$. For example, consider statement (1). Fix $i \in [n]$ and $d_{-i} \in \mathcal{D}^{n-1}$, and suppose there exists $Q^*_{D_i|D_{-i}=d_{-i}}$ such that

$$\sum_{d'_i\in\mathcal{D}} p_{Y|D_{-i}=d_{-i},D_i=d'_i}(y)\; q^*_{D_i|D_{-i}=d_{-i}}(d'_i) = \min_{\tilde{d}_i\in\mathcal{D}} p_{Y|D_{-i}=d_{-i},D_i=\tilde{d}_i}(y). \tag{B.4}$$

This equality holds if and only if $p_{Y|D_{-i}=d_{-i},D_i=d'_i}(y) = \min_{\tilde{d}_i\in\mathcal{D}} p_{Y|D_{-i}=d_{-i},D_i=\tilde{d}_i}(y)$ for all $d'_i \in \mathcal{D}$. Since (B.4) must hold for all i and all d_{-i}, $p_{Y|D_{-i}=d_{-i},D_i=d_i}(y)$ must be a constant that does not depend on d_i and d_{-i} for all y. However, this implies that X and Y are independent.

B.2. **Proof of Theorem 4.4.** The proof is fairly similar to the proof of Theorem 4.2; thus, some details are removed. First, note that it follows directly from Theorem 2.7 that (1) in the statement of the theorem is equivalent to ϵ-free-lunch privacy. To prove that (2) is equivalent to (1) we show that $\sup_{P_X\in\mathcal{Q}_\mathcal{X}} \ell(X\to y) \geq \sup_{P_X\in\mathcal{P}_\mathcal{X}} \ell(X\to y)$ for all $y \in \mathcal{Y}$ since the reverse inequality holds trivially. Consider the database $x^* = (d^*_1,\dots,d^*_n) \in \arg\min_x P_{Y|X=x}(y)$. We can use a construction similar to (B.3) to obtain a product distribution Q^*_X that satisfies $q^*_X(x^*) = (1-\varepsilon)^n$ while $q^*_X(x) \leq \frac{\varepsilon}{|\mathcal{D}|-1}(1-\varepsilon)^{n-1}$ for all $x \neq x^*$. Then, we get

$$\begin{aligned}
\sup_{P_X\in\mathcal{Q}_\mathcal{X}} \exp\Big(\ell(X\to y)\Big) &\geq \exp\Big(\ell_{P_{Y|X}\times Q^*_X}(X\to y)\Big)\\
&= \frac{\max_x p_{Y|X=x}(x)}{\sum_{x'} p_{Y|X=x'}(y) q^*_X(x)}\\
&\geq \frac{\max_x p_{Y|X=x}(x)}{(1-\varepsilon)^n p_{Y|X=x^*}(y) + \frac{\varepsilon}{|\mathcal{D}|-1}(1-\varepsilon)^{n-1}\sum_{x'\neq x^*} p_{Y|X=x'}(y)}.
\end{aligned}$$

For fixed n, letting $\varepsilon \to 0$ yields

$$\begin{aligned}
\sup_{P_X\in\mathcal{Q}_\mathcal{X}} \exp\Big(\ell(X\to y)\Big) &\geq \frac{\max_x p_{Y|X=x}(x)}{\min_{x'} p_{Y|X=x'}(y)}\\
&= \sup_{P_X\in\mathcal{P}_\mathcal{X}} \exp\Big(\ell(X\to y)\Big),
\end{aligned}$$

as desired.

Finally, we show that (3) is equivalent to (1). By the pre-processing inequality for PML (Saeidian et al., 2022b, Lemma 1) we have $\ell(D_i\to y) \leq \ell(X\to y)$ for all $i \in [n]$, $y \in \mathcal{Y}$,

and $P_X \in \mathcal{P}_{\mathcal{X}}$. So, we show that $\sup_{P_X \in \mathcal{P}_{\mathcal{X}}} \max_{i\in[n]} \ell(D_i \to y) \geq \sup_{P_X \in \mathcal{P}_{\mathcal{X}}} \ell(X \to y)$ for all $y \in \mathcal{Y}$. Fix an arbitrary $i \in [n]$. We write $P_X = P_{D_i} \times P_{D_{-i}|D_i}$ and optimize over P_{D_i} and $P_{D_{-i}|D_i}$ separately:

$$\begin{aligned}
\sup_{P_{D_{-i}|D_i}} \sup_{P_{D_i}} \exp\Big(\ell(D_i \to y)\Big) &= \sup_{P_{D_{-i}|D_i}} \max_{d_i} \sup_{P_{D_i}} \frac{p_{Y|D_i=d_i}(y)}{p_Y(y)} \\
&= \sup_{P_{D_{-i}|D_i}} \max_{d_i,d_i'} \frac{p_{Y|D_i=d_i}(y)}{p_{Y|D_i=d_i'}(y)} \qquad \text{(B.5a)} \\
&= \max_{d_i,d_i'} \sup_{P_{D_{-i}|D_i}} \frac{\sum_{d_{-i}} p_{Y|D_i=d_i,D_{-i}=d_{-i}}(y)\; p_{D_{-i}|D_i=d_i}(d_{-i})}{\sum_{d'_{-i}} p_{Y|D_i=d_i',D_{-i}=d'_{-i}}(y)\; p_{D_{-i}|D_i=d_i'}(d'_{-i})},
\end{aligned}$$

where (B.5a) follows from Theorem 2.7.

Consider the kernel $P^*_{D_{-i}|D_i}$ described by

$$p^*_{D_{-i}|D_i=d_i}(d_{-i}) \coloneqq \begin{cases} 1-\varepsilon, & \text{for some } d_{-i} \in \arg\max_{\tilde{d}_{-i}} p_{Y|D_{-i}=\tilde{d}_{-i},D_i=d_i}(y), \\ \frac{\varepsilon}{|\mathcal{D}|^{n-1}-1}, & \text{otherwise,} \end{cases}$$

and

$$p^*_{D_{-i}|D_i=d_i'}(d_{-i}) \coloneqq \begin{cases} 1-\varepsilon, & \text{for some } d_{-i} \in \arg\min_{\tilde{d}_{-i}} p_{Y|D_{-i}=\tilde{d}_{-i},D_i=d_i'}(y), \\ \frac{\varepsilon}{|\mathcal{D}|^{n-1}-1}, & \text{otherwise.} \end{cases}$$

Then, we get

$$\begin{aligned}
&\sup_{P_{D_{-i}|D_i}} \sup_{P_{D_i}} \exp\Big(\ell(D_i \to y)\Big) \\
&\geq \max_{d_i,d_i'} \frac{\sum_{d_{-i}} p_{Y|D_i=d_i,D_{-i}=d_{-i}}(y)\; p^*_{D_{-i}|D_i=d_i}(d_{-i})}{\sum_{d'_{-i}} p_{Y|D_i=d_i',D_{-i}=d'_{-i}}(y)\; p^*_{D_{-i}|D_i=d_i'}(d'_{-i})} \\
&= \max_{d_i,d_i'} \frac{\max_{d_{-i}} p_{Y|D_i=d_i,D_{-i}=d_{-i}}(y)}{\min_{d'_{-i}} p_{Y|D_i=d_i',D_{-i}=d'_{-i}}(y)} \qquad \text{(B.6a)} \\
&= \max_{d_i,d_i'} \max_{d_{-i},d'_{-i}} \frac{p_{Y|D_i=d_i,D_{-i}=d_{-i}}(y)}{p_{Y|D_i=d_i',D_{-i}=d'_{-i}}(y)},
\end{aligned}$$

where (B.6a) follows by letting $\varepsilon \to 0$.

□

B.3. **Proof of Proposition 4.5.** Given $i \in [n]$, let $B_i = f(D_i)$ be a binary random variable that determines whether or not entry D_i satisfies the predicate f. Since the outcome of the Laplace mechanism depends on D_i only through B_i the Markov chain $D_i - B_i - Y$ holds and $\ell(D_i \to y) \leq \ell(B_i \to y)$ for all outcomes $y \in \mathcal{Y}$. Thus, we may without loss of generality assume that $D_i = B_i$, that is, we assume that the database is binary.

For notational simplicity suppose $i = 1$. We write

$$\sup_{y\in\mathcal{Y}} \ell(D_1 \to y) = \sup_{y\in\mathcal{Y}} \log \frac{\max_{d_1\in\{0,1\}} p_{Y|D_1=d_1}(y)}{p_Y(y)}$$

$$= \sup_{y\in\mathcal{Y}} \log \frac{\max\limits_{d_1\in\{0,1\}} \mathbb{E}_{D_{-1}}\Big[\exp(-\frac{|y-\frac{d_1}{n}-\frac{S_{-1}}{n}|}{b})\Big]}{\mathbb{E}_X\Big[\exp(-\frac{|y-\frac{S_X}{n}|}{b})\Big]},$$

where $S_{-1} \coloneqq \sum_{i=2}^n D_i$ and $S_X \coloneqq \sum_{i=1}^n D_i$. We argue that it is sufficient to consider $y > 1$ and $y < 0$. This is because in the numerator we have

$$\mathbb{E}_{D_{-1}}\Big[\exp(-\frac{|y-\frac{d_1}{n}-\frac{S_{-1}}{n}|}{b})\Big] \le \min\left\{\mathbb{E}_{D_{-1}}\Big[\exp(-\frac{y-\frac{d_1}{n}-\frac{S_{-1}}{n}}{b})\Big], \mathbb{E}_{D_{-1}}\Big[\exp(\frac{y-\frac{d_1}{n}-\frac{S_{-1}}{n}}{b})\Big]\right\}.$$

Furthermore, the mapping $y \mapsto \mathbb{E}_X\Big[\exp(-\frac{|y-\frac{S_X}{n}|}{b})\Big]$ in the denominator is increasing in $(-\infty, p]$ and decreasing in $[p, \infty)$ since S_X is a Binomial random variable with success probability p.

Now, if $y > 1$, then

$$\begin{aligned}
\ell(D_1 \to y) &= \log \frac{\max\limits_{d_1\in\{0,1\}} \mathbb{E}_{D_{-1}}\Big[\exp(-\frac{y-\frac{d_1}{n}-\frac{S_{-1}}{n}}{b})\Big]}{\mathbb{E}_X\Big[\exp(-\frac{y-\frac{S_X}{n}}{b})\Big]} \\
&= \log \frac{\max\limits_{d_1\in\{0,1\}} \mathbb{E}_{D_{-1}}\Big[\exp(\frac{d_1}{nb}+\frac{S_{-1}}{nb})\Big]}{\mathbb{E}_X\Big[\exp(\frac{S_X}{nb})\Big]} \\
&= \frac{1}{nb} + \log \frac{\mathbb{E}_{D_{-1}}[\exp(\frac{D_2+\ldots+D_n}{nb})]}{\mathbb{E}_X[\exp(\frac{D_1+\ldots+D_n}{nb})]} \\
&= \frac{1}{nb} + \log \frac{\Pi_{j=2}^n \mathbb{E}[\exp(\frac{D_j}{nb})]}{\Pi_{j=1}^n \mathbb{E}[\exp(\frac{D_j}{nb})]} \\
&= \frac{1}{nb} - \log \left((1-p) + p\exp\left(\frac{1}{nb}\right)\right) \\
&\le \frac{1}{nb} - \log \left((1-c) + c\exp\left(\frac{1}{nb}\right)\right),
\end{aligned}$$

where the inequality is due to the fact that the mapping $p \mapsto (1-p) + p\exp(\frac{1}{nb})$ is increasing in p. Similarly, if $y < 0$, then

$$\ell(D_1 \to y) = \log \frac{\max\limits_{d_1\in\{0,1\}} \mathbb{E}_{D_{-1}}\Big[\exp(\frac{y-\frac{d_1}{n}-\frac{S_{-1}}{n}}{b})\Big]}{\mathbb{E}_X\Big[\exp(\frac{y-\frac{S_X}{n}}{b})\Big]}$$

$$
\begin{aligned}
&= \log \frac{\max\limits_{d_1\in\{0,1\}} \mathbb{E}_{D_{-1}}\Big[\exp(-\frac{d_1}{nb}-\frac{S_{-1}}{nb})\Big]}{\mathbb{E}_X\Big[\exp(-\frac{S_X}{nb})\Big]}\\
&= \log \frac{\mathbb{E}_{D_{-1}}[\exp(-\frac{D_2+\cdots+D_n}{nb})]}{\mathbb{E}_X[\exp(-\frac{D_1+\ldots+D_n}{nb})]}\\
&= \log \frac{\Pi_{j=2}^n \mathbb{E}[\exp(-\frac{D_j}{nb})]}{\Pi_{j=1}^n \mathbb{E}[\exp(-\frac{D_j}{nb})]}\\
&= \frac{1}{nb} - \log\left(p+(1-p)\exp\left(\frac{1}{nb}\right)\right)\\
&\le \frac{1}{nb} - \log\left((1-c)+c\exp\left(\frac{1}{nb}\right)\right),
\end{aligned}
$$

where the last inequality is due to the fact that the mapping $p \mapsto p+(1-p)\exp(\frac{1}{nb})$ is decreasing in p. We conclude that

$$
\sup_{P_X\in\mathcal{P}_c^f} \sup_{y\in\mathbb{R}} \ell(D_1\to y) = \frac{1}{nb} - \log\left((1-c)+c\exp\left(\frac{1}{nb}\right)\right).
$$