Information Leakage Envelopes

Anonymous Authors

***Abstract*—The abstract goes here.**

I. Introduction

The introduction of *differential privacy* (DP) [1] marked a major advance in privacy-preserving technologies. Rather than focusing on defenses against specific privacy attacks, such as re-identification or attribute disclosure, DP frames privacy as an inherent property of a data-processing system, one that can be quantified through analytical arguments. Central to this perspective is the concept of *privacy loss* [2], defined as

$$L_{x,x'}(Y) \coloneqq \log \frac{P_{Y|X=x}(Y)}{P_{Y|X=x'}(Y)},$$

where x and x' denote two possible values of the sensitive input (the *secret*) and $P_{Y|X}$ is the mechanism that produces the released information Y. The privacy loss random variable captures how much evidence an observation Y provides for distinguishing between x and x'. The goal of DP is then to allow making useful inferences about the input X while controlling the privacy loss.

The original formulation of DP, known as ε-DP, imposes a uniform upper bound on the privacy loss by a constant $\varepsilon > 0$ across all relevant pairs of inputs, that is, $L_{x,x'}(Y) \leq \varepsilon$ [1]. This strong guarantee often requires adding noise with Laplace or geometric distributions. While these mechanisms provide effective privacy protection, Gaussian noise is often more appealing from a statistical standpoint. The faster tail decay of the Gaussian distribution typically translates into improved utility, and properties such as closure under convolution simplify the theoretical analysis (since the sum of Gaussians remains Gaussian). These properties motivated the search for relaxations of ε-DP that allow the use of Gaussian noise.

A natural way to relax ε-DP is to allow the privacy loss to exceed ε with a small probability. This idea led to the notion of *probabilistic DP* [3], which requires

$$\mathbb{P}\{L_{x,x'}(Y) > \varepsilon\} \leq \delta$$

for some $\delta \in (0, 1)$. Despite its intuitive appeal, this definition did not gain widespread use due to a fundamental limitation: probabilistic DP is not *closed under post-processing*. That is, an adversary may apply a function to the output of a mechanism satisfying probabilistic DP (without access to the secret) to produce a result that no longer satisfies the original guarantee. Closure under post-processing is often treated as an axiom for privacy definitions and is motivated by the data-processing inequality in information theory [4]. Intuitively, more processing cannot increase the information available about the secret, and therefore should preserve privacy guarantees. As a result, probabilistic DP was largely set aside in favor of *approximate differential privacy* (ADP), also known as (ε, δ)-DP. A mechanism satisfies (ε, δ)-DP if

$$P_{Y|X=x}(\mathcal{A}) \leq e^{\varepsilon} P_{Y|X=x'}(\mathcal{A}) + \delta, \tag{1}$$

for all (measurable) sets $\mathcal{A}$. This definition is closed under post-processing, accommodates the use of Gaussian noise, and also enables *advanced composition* [5].

The parameter $\delta \in (0, 1)$ in approximate differential privacy is often informally interpreted as restricting the *failure probability* at level ε, i.e., $\mathbb{P}\{L_{x,x'}(Y) > \varepsilon\}$. This interpretation is, however, somewhat loose. In particular, one can show that δ merely provides a *lower bound* on this probability, and in fact, [6] constructed an example in which $\mathbb{P}\{L_{x,x'}(Y) > \varepsilon\}$ was much larger than the δ in ADP. Instead, the parameter δ in ADP upper bounds the failure probability at larger thresholds, such as 2ε or, more generally, $k\varepsilon$ (see (4) for a precise statement). Note that while a constant-factor increase in δ is generally harmless (since δ is chosen to be cryptographically small), a constant-factor increase in ε exponentially weakens the privacy guarantee as (1) depends on e^{ε}. That said, this difficulty in interpreting δ as a failure probability has not posed major problems in practice. In most modern systems, privacy loss is not directly tracked using ADP. Instead, one typically relies on accounting frameworks based on Rényi DP [7], concentrated DP [8], or Gaussian DP [9], which control the distribution of the privacy loss random variable. The resulting guarantees are often converted to (ε, δ)-DP, but this is primarily done for reporting and comparison purposes.

Interestingly, the difficulties associated with defining relaxed privacy guarantees based on a probability of failure are not unique to DP, but have been observed in a different framework based on a privacy measure called *pointwise maximal leakage* (PML) [10]. PML was recently introduced as a privacy measure with several attractive properties. Privacy guarantees based on PML allow the release of population-level information while protecting more nuanced, instance-dependent features of the data [11]. The framework is built around a central random quantity, the *information leakage random variable*, and different notions of privacy correspond to placing suitable restrictions on this quantity. In particular, the guarantee known as ε-PML imposes a uniform upper bound on the information leakage, in direct analogy with ε-DP.

In [10], the authors also proposed a relaxation of ε-PML based on bounding the failure probability at level ε, that is, the probability that the information leakage exceeds ε. However, much like probabilistic DP, this relaxed definition is not closed under post-processing. In this paper, we undertake the task of defining a relaxation of ε-PML that satisfies two desirable properties:

(i) the parameter δ should provide an *upper bound* on the failure probability at level ε, and

(ii) the resulting privacy guarantee should be closed under post-processing.

A. Contributions and Outline

After covering some background material in Section II, in Section III, we examine two relaxations of ε-PML inspired by ADP. Although ADP itself does not satisfy property (i), this investigation is nevertheless informative, as it reveals the types of PML-based definitions that arise when one follows the ADP design philosophy. Of the two candidate definitions, the first again fails to be post-processing (Theorem 1). The second is closed under post-processing, but does not admit a clear or consistent relationship with the failure probability: depending on the setting, either quantity may dominate the other. Consequently, neither definition satisfies both properties (i) and (ii).

Section IV contains our main contribution, namely the introduction of the *PML envelope*. The idea is to "close" the failure probability over all possible post-processings of Y. Concretely, for $\varepsilon > 0$, we define

$$\delta_c(\varepsilon) \coloneqq \sup_{Z} \mathbb{P}\{\ell(X \to Z) > \varepsilon\},$$

where the supremum is taken over all deterministic or randomized functions of Y, denoted by Z. The quantity $\delta_c(\varepsilon)$ represents the largest probability with which the information leakage can exceed ε after arbitrary downstream transformations.

Instead of fixing a leakage threshold ε and determining $\delta_c(\varepsilon)$, we may alternatively fix $\delta \in (0,1)$ and seek the smallest threshold ε that holds uniformly over all post-processings. This leads to the definition of the PML envelope

$$\varepsilon_c(\delta) \coloneqq \inf\{\varepsilon > 0 : \delta_c(\varepsilon) \leq \delta\}.$$

We provide equivalent characterizations of $\varepsilon_c(\delta)$ in terms of the cumulative distribution function (CDF) of the information leakage random variable $\ell(X \to Y)$ (Theorem 2). We also establish basic properties of ε_c, including monotonicity and continuity (Corollary 1). Note that by definition, the PML envelope satisfies both properties (i) and (ii) discussed above.

Next, we derive general upper and lower bounds on ε_c. The upper bound (Theorem 3) is expressed in terms of the *multiplicative Bayes capacity* [12] (also known as *maximal leakage* [13]), which is a well-studied quantity in the quantitative information flow literature. Lower bounds are obtained by restricting the class of admissible post-processings. In particular, we derive a lower bound based on *binary* post-processings, which is simple and efficiently computable.

Finally, we illustrate how the PML envelope can be computed and bounded by analyzing two canonical mechanisms. The first is the class of *PML-extremal* mechanisms in the high-privacy regime [14]. These mechanisms are the utility-optimal solutions to a broad class of optimization problems under the ε-PML constraint with sufficiently small $\varepsilon > 0$. For these mechanisms, we characterize the envelope exactly (Theorem 4) and show that $\varepsilon_c(\delta) = \varepsilon$ for all $\delta \in (0,1)$.

The second is the *randomized response* mechanism [15], a standard and widely used tool for guaranteeing local differential privacy [16]. Randomized response is a natural benchmark due to its simplicity, in particular, its symmetric structure, as well as its widespread use. For this mechanism, we derive upper and lower bounds on the PML envelope. Comparing these bounds to the corresponding ADP guarantees highlights that the two frameworks can behave quite differently, with no consistent relationship in general. This is not surprising because in the PML envelope, δ represents the failure probability, whereas in ADP it appears as an additive slack parameter.

II. Background

A. Notation

Uppercase letters denote random variables, lowercase letters denote their realizations, and calligraphic letters denote sets. All sets are assumed to be finite. We use X to denote a random variable containing sensitive information, also referred to as the *secret*. Its probability distribution is denoted by P_X, and its domain by $\mathcal{X}$. A privacy mechanism (or simply, a mechanism) is specified by a conditional probability distribution $P_{Y|X}$, which takes X as input and produces an output Y with domain $\mathcal{Y}$ and (marginal) distribution P_Y. The joint distribution of (X, Y) is denoted by P_{XY}.

Random variables X, Y, and Z are said to form a Markov chain $X - Y - Z$ if X and Z are conditionally independent given Y, that is, $P_{XZ|Y} = P_{X|Y} \times P_{Z|Y}$. We write $P_{Z|X} = P_{Z|Y} \circ P_{Y|X}$ to denote *marginalization*, meaning that

$$P_{Z|X=x}(z) = \sum_{y \in \mathcal{Y}} P_{Z|Y=y}(z)\, P_{Y|X=x}(y), \quad x \in \mathcal{X}, z \in \mathcal{Z}.$$

For a set $\mathcal{E}$, $\mathbf{1}_{\mathcal{E}}$ denotes its indicator function. For a positive integer k, we write $[k] = \{1, \ldots, k\}$.

B. Differential Privacy

Differential privacy (DP) was introduced as a framework for releasing statistics about databases while protecting the privacy of individual entries [1, 2]. Informally, the original definition requires that an adversary should not be able to distinguish between *neighboring* datasets, i.e., databases that differ in a single record, based on the statistics released from each.

Later, DP was extended to decentralized settings, where data is perturbed before collection. This variant, known as *local differential privacy* (LDP) [17, 18], eliminates the need for a trusted data curator, as privacy is enforced at the data source. In this work, we adopt the local model for all DP definitions. This choice is made without loss of generality, since switching between the two models amounts to redefining what it means for two inputs to be "neighbors." By working in the local model, we abstract away the structure of the secret and instead focus on our main objective: to analyze how various definitions behave under post-processing.

Given $x, x' \in \mathcal{X}$, let

$$L_{x,x'}(Y) \coloneqq \log \frac{P_{Y|X=x}(Y)}{P_{Y|X=x'}(Y)},$$

denote the *privacy loss random variable* of DP [2]. The simplest DP definition imposes a uniform bound on the privacy loss.

Definition 1 (Pure DP [1])**.** *Let $\varepsilon > 0$. A privacy mechanism $P_{Y|X}$ is said to satisfy ε-DP if $L_{x,x'}(y) \leq \varepsilon$ for all $x, x' \in \mathcal{X}$ and all $y \in \mathcal{Y}$.*

A standard relaxation of pure DP, known as approximate DP (ADP) [19], allows an additive slack $\delta \in (0,1)$ in the event-wise comparison of the conditional distributions.

Definition 2 (Approximate DP [2])**.** *Let $\varepsilon > 0$ and $\delta \in (0,1)$. A privacy mechanism $P_{Y|X}$ is said to satisfy (ε,δ)-DP if for all $x, x' \in \mathcal{X}$ and all sets $\mathcal{E} \subset \mathcal{Y}$ we have*

$$P_{Y|X=x}(\mathcal{E}) \leq e^{\varepsilon} P_{Y|X=x'}(\mathcal{E}) + \delta. \tag{2}$$

The common interpretation of (ε,δ)-DP is that it allows the privacy guarantees of ε-DP to fail with probability δ. However, as pointed out in [6], this interpretation is misleading: the parameter δ is, in fact, a *lower bound* on the worst-case failure probability. To demonstrate this, we find the smallest δ that satisfies (2) for a fixed ε. This function, called the *privacy profile* [20] and denoted by $\delta^*(\varepsilon, P_{Y|X})$, can be expressed as

$$\delta^*(\varepsilon, P_{Y|X}) = \max_{x,x' \in \mathcal{X}} \max_{\mathcal{E} \subset \mathcal{Y}} \Big(P_{Y|X=x}(\mathcal{E}) - e^{\varepsilon} P_{Y|X=x'}(\mathcal{E}) \Big).$$

Then, $P_{Y|X}$ satisfies (ε,δ)-DP for all $\delta^*(\varepsilon) \leq \delta < 1$. It is not difficult to see that δ^* can also be expressed in the following more intuitive form [21, Lemma 2.8]:

$$\begin{aligned} &\delta^*(\varepsilon, P_{Y|X}) \\ &= \max_{x,x'} \mathbb{E}_{Y \sim P_{Y|X=x}} \left[\max\left\{ 0, 1 - \frac{\exp(\varepsilon)}{\exp(L_{x,x'}(Y))} \right\} \right] \qquad (3) \\ &= \max_{x,x'} \mathbb{E}_{Y \sim P_{Y|X=x}} \left[\mathbf{1}_{\{L_{x,x'} > \varepsilon\}}(Y) \left(1 - \frac{\exp(\varepsilon)}{\exp(L_{x,x'}(Y))} \right) \right]. \end{aligned}$$

Intuitively, the term

$$1 - \frac{\exp(\varepsilon)}{\exp(L_{x,x'}(y))},$$

is a penalty applied to outcomes y with privacy loss larger than ε. The greater the deviation of privacy loss from ε, the larger this penalty becomes, reflecting the idea that such outcomes contribute more significantly to the overall privacy risk. Observe that $0 \leq 1 - \frac{\exp(\varepsilon)}{\exp(L_{x,x'}(y))} \leq 1$ on the set $\{L_{x,x'} > \varepsilon\}$, so we have

$$\begin{aligned} \delta^*(\varepsilon, P_{Y|X}) &\leq \max_{x,x' \in \mathcal{X}} \mathbb{E}_{Y \sim P_{Y|X=x}} \left[\mathbf{1}_{\{L_{x,x'} > \varepsilon\}}(Y) \right] \\ &= \max_{x,x' \in \mathcal{X}} P_{Y|X=x} \left\{ L_{x,x'}(Y) > \varepsilon \right\}. \end{aligned}$$

The quantity $\max_{x,x'} P_{Y|X=x} \{L_{x,x'}(Y) > \varepsilon\}$ represents the worst-case failure probability of DP, that is, the (maximum) probability that the privacy loss exceeds ε. Hence, the above derivation shows that the value of δ in ADP merely lower bounds this quantity. It is nevertheless true that (ε,δ)-DP implies the weaker tail bound

$$P_{Y|X=x}\big\{L_{x,x'}(Y) > 2\varepsilon\big\} \leq \frac{\delta}{1 - e^{-\varepsilon}},$$

for all x, x' [22, Lemma 3.3] (see also [23]). More generally, (ε,δ)-DP implies

$$\max_{x,x'} P_{Y|X=x}\big\{L_{x,x'}(Y) > k\varepsilon\big\} \leq \frac{\delta}{1 - e^{-(k-1)\varepsilon}}, \tag{4}$$

for all integers $k \geq 2$. While some recent works have taken care to be more precise in their description of ADP, for instance, by stating that ADP restricts the failure probability up to a scaling of the parameters, the misconception remains widespread, especially in more applied contexts where the focus is not on the theoretical nuances.

The above discussion naturally raises the question of whether we could define a DP variant by explicitly upper bounding the failure probability. Such a definition was, in fact, proposed by Machanavajjhala *et al.* [3] and is known as *probabilistic DP*.

Definition 3 (Probabilistic DP [3])**.** *Let $\varepsilon > 0$ and $\delta \in (0,1)$. A privacy mechanism $P_{Y|X}$ is said to satisfy (ε,δ)-probabilistic DP if for all $x, x' \in \mathcal{X}$ we have $P_{Y|X=x}\{L_{x,x'}(Y) > \varepsilon\} \leq \delta$.*

While probabilistic DP is arguably more intuitive than ADP, it is rarely used in practice. This is primarily because probabilistic DP may not be preserved after post-processing [6, 24], in contrast to ADP, which is post-processing safe [2]. This property has contributed to ADP becoming the de facto standard in the literature. To illustrate what can go wrong for probabilistic DP, let us recall the following example from [6].

Example 1. Consider a privacy mechanism $P_{Y|X}$ with a binary input space $\mathcal{X} = \{0,1\}$ and a quaternary output space $\mathcal{Y} = [4]$. Given fixed $\varepsilon > 0$ and $\delta \in (0,1)$, define

$$P_{Y|X=0}(y) = \begin{cases} \delta & \text{if } y = 1, \\ \frac{e^{\varepsilon}}{1+e^{\varepsilon}}(1-\delta) & \text{if } y = 2, \\ \frac{1}{1+e^{\varepsilon}}(1-\delta) & \text{if } y = 3, \\ 0 & \text{if } y = 4, \end{cases}$$

$$P_{Y|X=1}(y) = \begin{cases} 0 & \text{if } y = 1, \\ \frac{1}{1+e^{\varepsilon}}(1-\delta) & \text{if } y = 2, \\ \frac{e^{\varepsilon}}{1+e^{\varepsilon}}(1-\delta) & \text{if } y = 3, \\ \delta & \text{if } y = 4. \end{cases}$$

We have the following values for the privacy loss:

$$\begin{aligned} L_{0,1}(1) = L_{1,0}(4) = \infty, &\quad L_{0,1}(2) = L_{1,0}(3) = \varepsilon, \\ L_{0,1}(3) = L_{1,0}(2) = -\varepsilon, &\quad L_{0,1}(4) = L_{1,0}(1) = -\infty. \end{aligned}$$

When $X = 0$, the "bad" set on which the privacy loss exceeds ε is $\{1\}$ and when $X = 1$, the bad set is $\{4\}$.

Since $P_{Y|X=0}\{1\} = P_{Y|X=1}\{4\} = \delta$, the mechanism satisfies (ε, δ)-probabilistic DP. Now, let $Z = h(Y)$, where h is

$$h(y) = \begin{cases} \perp & \text{if } y \in \{1, 2\}, \\ y & \text{otherwise.} \end{cases}$$

This transformation collapses part of the output space into a new symbol $\perp$, and merges a "bad" output ($y = 1$ for $X = 0$) with a "good" one ($y = 2$). The mechanism $P_{Z|X} = P_{Z|Y} \circ P_{Y|X}$ from X to Z is

$$P_{Z|X=0}(z) = \begin{cases} \delta + \frac{e^\varepsilon}{1+e^\varepsilon}(1-\delta) & \text{if } z = \perp, \\ \frac{1}{1+e^\varepsilon}(1-\delta) & \text{if } z = 3, \\ 0 & \text{if } z = 4, \end{cases}$$

$$P_{Z|X=1}(z) = \begin{cases} \frac{1}{1+e^\varepsilon}(1-\delta) & \text{if } z = \perp, \\ \frac{e^\varepsilon}{1+e^\varepsilon}(1-\delta) & \text{if } z = 3, \\ \delta & \text{if } z = 4. \end{cases}$$

Observe that

$$\begin{aligned} P_{Z|X=0}(\perp) &= \delta + \frac{e^\varepsilon}{1+e^\varepsilon}(1-\delta) \\ &> e^\varepsilon \cdot \frac{1}{1+e^\varepsilon}(1-\delta) \\ &= e^\varepsilon \cdot P_{Z|X=1}(\perp), \end{aligned}$$

so $P_{Z|X}$ does *not* satisfy (ε, δ)-probabilistic DP.

In the following section, we discuss similar challenges with post-processing in the framework of pointwise maximal leakage.

C. Pointwise Maximal Leakage

Pointwise maximal leakage (PML) [10] is a recent notion of privacy defined using concepts from *quantitative information flow* [12]. PML quantifies the inference risk posed by a broad class of adversaries. Its threat model can be described as follows: consider an adversary who seeks to maximize a non-negative gain function g by producing a guess W of the private variable X. The gain function encodes the adversary's objective and can capture a wide range of privacy attacks, including membership and attribute inference [10]. For an output $y \in \mathcal{Y}$, PML measures information leakage as the ratio between the adversary's expected gain after observing y and the expected gain before observing y. Then, to obtain a robust and attack-agnostic notion of leakage, this ratio is maximized over all nonnegative gain functions. Formalizing this idea leads to the following definition for PML.

Definition 4 (PML [10])**.** *Suppose $X \sim P_X$ and let Y be the random variable induced by the mechanism $P_{Y|X}$. The pointwise maximal leakage from X to $y \in \mathcal{Y}$ is defined as*

$$\ell_{P_{XY}}(X \to y) := \log \sup_g \frac{\sup_{P_{W|Y}} \mathbb{E}[g(X, W) \mid Y = y]}{\max_{w' \in \mathcal{W}} \mathbb{E}[g(X, w')]}, \quad (5)$$

where $P_{W|Y}$ is the conditional distribution of the adversary's guess W given Y. The supremum is over all and non-negative measurable functions g.

In this work, both X and Y are assumed to be finite-valued. Under this assumption, it was shown in [10] that PML takes the simpler form

$$\begin{aligned} \ell_{P_{XY}}(X \to y) &= \log \max_{x \in \mathcal{X}} \frac{P_{X|Y=y}(x)}{P_X(x)} \\ &= \log \max_{x \in \mathcal{X}} \frac{P_{Y|X=x}(y)}{P_Y(y)}, \end{aligned}$$

where $P_{X|Y}$ denotes the posterior distribution of X given Y. It is straightforward to see that PML satisfies the bounds

$$0 \le \ell_{P_{XY}}(X \to y) \le \log \frac{1}{\min_{x \in \mathcal{X}} P_X(x)}, \quad (6)$$

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

In the information theory literature, the quantity

$$i_{P_{XY}}(x; y) = \log \frac{P_{XY}(x, y)}{P_X(x) P_Y(y)}, \quad x \in \mathcal{X}, y \in \mathcal{Y},$$

is commonly referred to as the *information density* of P_{XY}. PML can also be expressed as

$$\ell_{P_{XY}}(X \to y) = \max_{x \in \mathcal{X}} i_{P_{XY}}(x; y).$$

Note that, unlike DP, PML depends on the prior distribution P_X and is therefore a property of the joint distribution P_{XY}. When the joint distribution is clear from context, we omit the subscript and write $i(x; y)$ for information density and $\ell(X \to y)$ or simply $\ell(y)$ for PML.

The joint distribution P_{XY} is said to satisfy ε-PML with $\varepsilon > 0$ if $\ell(X \to y) \le \varepsilon$ for all $y \in \mathcal{Y}$. In [10], the authors also introduced a relaxation of ε-PML by imposing an upper bound on the tail of $\ell(X \to Y)$.[1]

Definition 5 (Probabilistic PML)**.** *Let $\varepsilon > 0$ and $\delta \in (0, 1)$. The joint distribution P_{XY} is said to satisfy (ε, δ)-probabilistic PML if*

$$P_Y\{\ell(X \to Y) > \varepsilon\} \le \delta.$$

Much like probabilistic LDP, probabilistic PML is *not* closed under post-processing. To illustrate this, below we give an example similar to [10, Example 7].

Example 2. Let X be uniformly distributed on $\mathcal{X} = [4]$. Consider the privacy mechanism

$$P_{Y|X} = \begin{bmatrix} 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0.2 & 0.4 & 0.4 \\ 0.2 & 0 & 0.4 & 0.4 \end{bmatrix}, \quad (7)$$

where $(P_{Y|X})_{ij} = P_{Y|X=i}(j)$. The outcomes have information leakage

$$\begin{aligned} \ell_{P_{XY}}(X \to 1) &= \ell_{P_{XY}}(X \to 2) = \log 4, \\ \ell_{P_{XY}}(X \to 3) &= \ell_{P_{XY}}(X \to 4) = \log \frac{10}{9}. \end{aligned}$$

[1] The function $\ell(X \to y)$ is defined pointwise for each $y \in \mathcal{Y}$. Consequently, $\ell(X \to Y)$ is a random variable induced by Y.

Since $P_Y(1) = P_Y(2) = \frac{1}{20}$, P_{XY} satisfies $(\log\frac{10}{9}, 0.1)$-probabilistic PML.

Now, let $Z = h(Y)$, where

$$h(y) = \begin{cases} 1 & \text{if } y \in \{1,3\}, \\ 2 & \text{if } y \in \{2,4\}, \end{cases}$$

The outcomes of Z are equiprobable and have information leakage $\ell_{P_{XZ}}(X \to 1) = \ell_{P_{XZ}}(X \to 2) = \log\frac{6}{5}$. Since $\frac{6}{5} > \frac{10}{9}$, $P_{XZ} = P_{Z|Y} \circ P_{XY}$ does *not* satisfy the original guarantee of $(\log\frac{10}{9}, 0.1)$-probabilistic PML.

In light of Example 2, we ask: What alternative definitions can reconcile PML-based privacy guarantees with the post-processing requirement? We explore answers to this question in the subsequent sections.

III. Two Candidate Definitions of Approximate PML

As a first step, we examine two candidate definitions of approximate PML, inspired by ADP. In particular, we define analogues of (3) by replacing the DP privacy loss with either the information leakage random variable or the information density.

Given $\varepsilon > 0$, define

$$\begin{aligned} \psi_1(\varepsilon, P_{XY}) &:= \mathbb{E}_{Y \sim P_Y}\left[\max\left\{0, 1 - \frac{\exp(\varepsilon)}{\exp(\ell(Y))}\right\}\right] \\ &= \mathbb{E}_{Y \sim P_Y}\left[\mathbf{1}_{\{\ell > \varepsilon\}}(Y)\left(1 - \frac{\exp(\varepsilon)}{\exp(\ell(Y))}\right)\right], \end{aligned}$$

and

$$\begin{aligned} &\psi_2(\varepsilon, P_{XY}) \\ &:= \max_x \mathbb{E}_{Y \sim P_{Y|X=x}}\left[\max\left\{0, 1 - \frac{\exp(\varepsilon)}{\exp(i(x;Y))}\right\}\right] \\ &= \max_x \mathbb{E}_{Y \sim P_{Y|X=x}}\left[\mathbf{1}_{\{i(x;\cdot) > \varepsilon\}}(Y)\left(1 - \frac{\exp(\varepsilon)}{\exp(i(x;Y))}\right)\right]. \end{aligned}$$

Both definitions follow the same general pattern: they assign a penalty to outcomes where the information leakage $\ell(y)$ or the information density $i(x;y)$ exceeds the threshold ε. Note that $0 \leq 1 - \frac{\exp(\varepsilon)}{\exp(\ell(y))} \leq 1$ when $\ell(y) > \varepsilon$, so ψ_1 lower bounds the PML failure probability $P_Y\{\ell(Y) > \varepsilon\}$.

Despite the parallelism in their expressions, these two definitions behave differently under post-processing: ψ_1 is not post-processing safe, whereas ψ_2 is.

Theorem 1. *Let $\varepsilon > 0$.*

(i) There exist random variables X, Y, Z satisfying the Markov chain $X - Y - Z$ such that

$$\psi_1(\varepsilon, P_{Z|Y} \circ P_{XY}) > \psi_1(\varepsilon, P_{XY}).$$

(ii) For all random variables X, Y, Z satisfying the Markov chain $X - Y - Z$ we have

$$\psi_2(\varepsilon, P_{Z|Y} \circ P_{XY}) \leq \psi_2(\varepsilon, P_{XY}).$$

Proof:

(i) It suffices to construct an example where $\psi_1(\varepsilon, P_{Z|Y} \circ P_{XY}) > \psi_1(\varepsilon, P_{XY})$. Recall the setup of Example 2. Setting $\varepsilon = \log\frac{10}{9}$, a direct calculation yields

$$\psi_1(\varepsilon, P_{XY}) = \frac{13}{180} < \frac{2}{27} = \psi_1(\varepsilon, P_{XZ}).$$

Hence, ψ_1 can increase under post-processing.

(ii) Our argument mirrors the standard proof of post-processing for ADP. Consider the Markov chain $X - Y - Z$. Fix an arbitrary $x \in \mathcal{X}$ and observe that

$$\begin{aligned} &\mathbb{E}_{Z \sim P_{Z|X=x}}\left[\mathbf{1}_{\{i(x;\cdot) > \varepsilon\}}(Z)\left(1 - \frac{\exp(\varepsilon)}{\exp(i(x;Z))}\right)\right] \\ &= \sum_{z: i(x;z) > \varepsilon}\left(1 - \frac{\exp(\varepsilon)}{\exp(i(x;z))}\right) P_{Z|X=x}(z) \\ &= \sum_{z: i(x;z) > \varepsilon} P_{Z|X=x}(z) - e^{\varepsilon} \sum_{z: i(x;z) > \varepsilon} P_Z(z) \\ &= P_{Z|X=x}\{z : i(x;z) > \varepsilon\} - e^{\varepsilon} P_Z\{z : i(x;z) > \varepsilon\}. \end{aligned}$$

Thus, to prove that ψ_2 does not increase under post-processing, it suffices to show that

$$P_{Z|X=x}(\mathcal{A}) - e^{\varepsilon} P_Z(\mathcal{A}) \leq \psi_2(\varepsilon, P_{XY}),$$

for all $x \in \mathcal{X}$ and arbitrary sets $\mathcal{A} \subseteq \mathcal{Z}$. Indeed, for each set $\mathcal{A} \subseteq \mathcal{Z}$ we have

$$\begin{aligned} &P_{Z|X=x}(\mathcal{A}) - e^{\varepsilon} P_Z(\mathcal{A}) \\ &= \sum_{y \in \mathcal{Y}} P_{Z|Y=y}(\mathcal{A})\Big(P_{Y|X=x}(y) - e^{\varepsilon} P_Y(y)\Big) \\ &\leq \sum_{y:\, i(x;y) > \varepsilon} P_{Z|Y=y}(\mathcal{A})\Big(P_{Y|X=x}(y) - e^{\varepsilon} P_Y(y)\Big) \\ &\leq \sum_{y:\, i(x;y) > \varepsilon} P_{Y|X=x}(y) - e^{\varepsilon} P_Y(y) \\ &= \sum_{y:\, i(x;y) > \varepsilon}\left(1 - \frac{\exp(\varepsilon)}{\exp(i(x;y))}\right) P_{Y|X=x}(y) \\ &= \mathbb{E}_{Y \sim P_{Y|X=x}}\left[\mathbf{1}_{\{i(x;\cdot) > \varepsilon\}}(Y)\left(1 - \frac{\exp(\varepsilon)}{\exp(i(x;Y))}\right)\right] \\ &\leq \psi_2(\varepsilon, P_{XY}). \end{aligned}$$

■

Thus, Theorem 1 establishes that ε-PML can be relaxed with an additive parameter such that the resulting definition is closed under post-processing. It is straightforward to see that ψ_2 can be written as

$$\psi_2(\varepsilon, P_{XY}) = \max_{x \in \mathcal{X}} \max_{\mathcal{E} \subset \mathcal{Y}} \Big(P_{Y|X=x}(\mathcal{E}) - e^{\varepsilon} P_Y(\mathcal{E})\Big).$$

Although ψ_2 is closed under post-processing, it does not provide a suitable proxy for the failure probability $P_Y\{\ell(Y) > \varepsilon\}$, since neither quantity bounds the other one in general. To illustrate this, consider again the mechanism $P_{Y|X}$ in (7). We have

$$P_Y\{\ell(Y) > \log \tfrac{10}{9}\} = P_Y\{\ell(Y) > \log 3\} = 0.1,$$

while at the same time,

$$\psi_2(\log 3, P_{XY}) = 0.05 \;<\; 0.1 \;<\; \frac{13}{90} = \psi_2\big(\log \tfrac{10}{9}, P_{XY}\big)\,.$$

This example shows that ψ_2 is not comparable to the tail probability $P_Y\{\ell(Y) > \varepsilon\}$. As such, in the next sections, we take a different approach and focus on directly "closing" the tail probability $P_Y\{\ell(Y) > \varepsilon\}$.

IV. Closing the Probability of Failure and the PML Envelope

We now turn to a natural approach to the post-processing question, one that directly addresses the core issue of probabilistic PML. Instead of defining an additive relaxation or adjusting penalties, we examine the worst-case probability of failure across all possible post-processing mechanisms.

Formally, given a joint distribution P_{XY} and $\varepsilon > 0$, define

$$\delta_c(\varepsilon) := \sup_{Z : X - Y - Z} \mathbb{P}\{\ell(Z) > \varepsilon\}, \tag{8}$$

where the supremum is taken over all finite random variables Z satisfying the Markov chain $X - Y - Z$, alternatively, all conditional distributions $P_{Z|Y}$. Observe that δ_c quantifies the largest probability that PML exceeds ε under any downstream transformation, so by definition, it is post-processing safe.

In (8), we fix ε and find the largest failure probability. Alternatively, we could fix $\delta \in (0,1)$ and find the *smallest* ε that holds with probability at least $1-\delta$ after arbitrary post-processing. Let Z denote a (possibly randomized) function of Y, and consider its corresponding leakage random variable $\ell(Z)$. Let

$$C_Z(t) = \mathbb{P}\{\ell(Z) \le t\}, \quad t \ge 0,$$

denote the *cumulative distribution function* (CDF) of $\ell(Z)$, and for $s \in (0,1)$, define

$$\mathrm{Quant}_Z^{\leftarrow}(s) := \inf\big\{t \ge 0 : C_Z(t) \ge s\big\}.$$

to be the *left-continuous quantile function* of $\ell(Z)$ at level s. Then, for $\delta \in (0,1)$, we define

$$\varepsilon_Z(\delta) := \mathrm{Quant}_Z^{\leftarrow}(1-\delta) = \inf\big\{t \ge 0 : C_Z(t) \ge 1-\delta\big\}, \tag{9}$$

which captures the smallest threshold t such that PML is bounded by t with probability at least $1-\delta$. Observe that the mapping $s \mapsto \mathrm{Quant}_Z^{\leftarrow}(s)$ is non-decreasing and left-continuous. Consequently, $\delta \mapsto \varepsilon_Z(\delta)$ is non-increasing and right-continuous. The function $\varepsilon_Z(\delta)$ also admits the equivalent, and somewhat more explicit formulation

$$\varepsilon_Z(\delta) = \min_{\substack{\mathcal{A} \subset \mathcal{Z} \\ P_Z(\mathcal{A}) \ge 1-\delta}} \max_{z \in \mathcal{A}} \ell(z), \qquad \delta \in (0,1). \tag{10}$$

That is, $\varepsilon_Z(\delta)$ tells us how small we can make the worst-case leakage, up to ignoring a set of probability δ. The equivalence between (9) and (10) is proved in Lemma 2 in the appendix for completeness.

We now define the *PML envelope* of P_{XY}, denoted by ε_c, as the supremum of ε_Z over all post-processings of Y, i.e.,

$$\varepsilon_c(\delta) := \sup_{Z : X - Y - Z} \varepsilon_Z(\delta), \qquad \delta \in (0,1).$$

In words, $\varepsilon_c(\delta)$ captures the tightest privacy guarantee that survives arbitrary downstream transformations, up to a failure probability of δ. By definition, the map $\delta \mapsto \varepsilon_c(\delta)$ is non-increasing.

Before we characterize and compute the PML envelope, let us first express it in an alternative form. Given a random variable Z and $s \in (0,1)$, let

$$\mathrm{Quant}_Z^{\rightarrow}(s) := \sup\big\{t \ge 0 : C_Z(t) \le s\big\},$$

be the *right-continuous quantile function* of $\ell(Z)$ at level s. Then, for $\delta \in (0,1)$ define

$$\begin{aligned} \bar{\varepsilon}_Z(\delta) &:= \mathrm{Quant}_Z^{\rightarrow}(1-\delta) = \sup\{t \ge 0 : C_Z(t) \le 1-\delta\} \\ &= \max_{\substack{\mathcal{A} \subset \mathcal{Z} \\ P_Z(\mathcal{A}) \ge \delta}} \min_{z \in \mathcal{A}} \ell(z). \end{aligned}$$

Observe that the mapping $\delta \mapsto \bar{\varepsilon}_Z(\delta)$ is non-increasing and left-continuous.

Heuristically, the difference between $\bar{\varepsilon}_Z(\delta)$ and $\varepsilon_Z(\delta)$ can be understood as follows: $\varepsilon_Z(\delta)$ is the smallest upper bound on the worst-case PML over all the "good" sets of outputs (i.e., sets with probability at least $1-\delta$). In contrast, $\bar{\varepsilon}_Z(\delta)$ is the largest lower bound on the worst-case PML over all "bad" sets of outputs (i.e., sets with probability at least δ). If C_Z is strictly increasing, then $\varepsilon_Z(\delta) = \bar{\varepsilon}_Z(\delta)$ for all $\delta \in (0,1)$, since in that case the inverse of C_Z is well defined and coincides with both the left-continuous and right-continuous quantile functions. More generally, we have $\bar{\varepsilon}_Z(\delta) \ge \varepsilon_Z(\delta)$ for all $\delta \in (0,1)$ and the inequality may be strict, particularly when Z is a discrete random variable.

The difference between ε_Z and $\bar{\varepsilon}_Z$ is further illustrated in the following example.

Example 3. Recall X and Y from Example 2, and fix $\delta = 0.1$. The leakage random variable $\ell(Y)$ takes on two distinct values: a low leakage value of $\log(10/9)$ with probability 0.9, and a high leakage value of $\log(4)$ with probability 0.1. For this distribution, the left-continuous quantile is $\varepsilon_Y(\delta) = \log(10/9)$, while the right-continuous quantile is $\bar{\varepsilon}_Y(\delta) = \log 4$. Figure 1 shows the distribution function C_Y of $\ell(Y)$, together with two vertical lines marking $\varepsilon_Y(\delta)$ and $\bar{\varepsilon}_Y(\delta)$.

Below, we show that maximizing over all post-processings of Y eliminates the gap between $\bar{\varepsilon}_Z(\delta)$ and $\varepsilon_Z(\delta)$. Thus, both quantities can be used to define the PML envelope.

Theorem 2. *Suppose X and Y are finite random variables. For all $\delta \in (0,1)$, we have*

$$\varepsilon_c(\delta) = \sup_{Z : X - Y - Z} \varepsilon_Z(\delta) = \sup_{Z : X - Y - Z} \bar{\varepsilon}_Z(\delta).$$

Proof: See Appendix C. ■

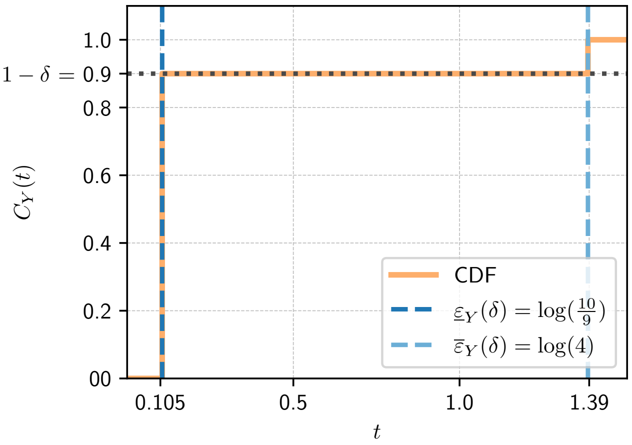


Fig. 1: An example of the CDF $C_Y(t)$ together with $\underline{\varepsilon}_Y(\delta)$ and $\bar{\varepsilon}_Y(\delta)$ at $\delta = 0.1$. Since C_Y is not strictly increasing, there is a gap between $\underline{\varepsilon}_Y(\delta)$ and $\bar{\varepsilon}_Y(\delta)$.

Importantly, Theorem 2 explains *why* probabilistic PML is not closed under post-processing. It states that the idea of simply ignoring a small "bad" set does not work because, through post-processing, an adversary can increase the information leakage all the way up to the smallest value in the bad set. To illustrate this, let us revisit Example 2. For $\delta = 0.1$, the good set consists of two outcomes with PML $\log(10/9)$ and the bad set consists of two outcomes with PML $\log(4)$. By Theorem 2, the adversary can increase the leakage from $\log(10/9)$ all the way up to $\log(4)$. Note that by (6), $\log(4)$ is the largest amount of information *any* mechanism can leak about a uniformly distributed quaternary secret X. As $\delta \mapsto \varepsilon_c(\delta)$ is non-increasing, it follows that $\varepsilon_c(\delta) = \log(4)$ for all $\delta \in (0, 0.1]$.

Theorem 2 can also be used to prove that the map $\delta \mapsto \varepsilon_c(\delta)$ is continuous.

Corollary 1. Suppose X and Y are finite random variables. The function ε_c is continuous on $(0, 1)$.

In the following sections, we characterize and bound ε_c in various examples.

A. Upper Bounding the PML envelope

Computing the PML envelope is, in general, a challenging task, since it involves a maximization over all possible post-processings of Y. This motivates us to derive bounds on the envelope. Below, we present a general upper bound.

Theorem 3. *Suppose the joint distribution P_{XY} satisfies ε-PML with $\varepsilon > 0$. Then, for and all $\delta \in (0, 1)$ we have*

$$\varepsilon_c(\delta) \leq \min\left\{\mathcal{L}(X \to Y) + \log\frac{1}{\delta},\ \varepsilon\right\}, \tag{11}$$

where

$$\mathcal{L}(X \to Y) = \log \mathbb{E}\big[e^{\ell(Y)}\big] = \log \sum_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} P_{Y|X=x}(y),$$

denotes multiplicative Bayes capacity [25] or maximal leakage [13].

Proof: To prove the first term of the bound, fix some Z satisfying the Markov chain $X - Y - Z$, and $t \geq 0$. We write

$$\begin{aligned} S_Z(t) &= \mathbb{P}\{\ell(Z) > t\} \\ &= \mathbb{P}\Big\{e^{\ell(Z)} > e^t\Big\} \\ &\leq \mathbb{E}\big[e^{\ell(Z)}\big]e^{-t} && \text{(12a)} \\ &\leq \mathbb{E}\Big[e^{\ell(Y)}\Big]e^{-t}, && \text{(12b)} \end{aligned}$$

where (12a) is due to Markov's inequality and (12b) is due to the data-processing inequality for maximal leakage [13, Lemma 1]. Thus,

$$\{t \geq 0 : S_Z(t) \geq \delta\} \subseteq \Big\{t \geq 0 : \mathbb{E}\big[e^{\ell(Y)}\big]e^{-t} \geq \delta\Big\},$$

which implies that

$$\begin{aligned} \bar{\varepsilon}_Z(\delta) &= \sup\,\{t \geq 0 : S_Z(t) \geq \delta\} \\ &\leq \sup\,\Big\{t \geq 0 : \mathbb{E}\big[e^{\ell(Y)}\big]e^{-t} \geq \delta\Big\} \\ &= \mathcal{L}(X \to Y) + \log\left(\frac{1}{\delta}\right). \end{aligned}$$

To prove the second term, we use the following four facts:

(i) P_{XY} satisfies ε-PML if

$$\max_{y \in \mathcal{Y}} \ell(X \to y) \leq \varepsilon,$$

(ii) when $\mathcal{Y}$ is a finite set, there exists $\delta_0 > 0$ such that for all $0 < \delta \leq \delta_0$ we have

$$\underline{\varepsilon}_Y(\delta) = \bar{\varepsilon}_Y(\delta) = \max_{y \in \mathcal{Y}} \ell(X \to y),$$

(iii) the post-processing inequality for PML [10, Lemma 1] states that if the Markov chain $X - Y - Z$ holds, then

$$\max_{z \in \mathcal{Z}} \ell_{P_{XZ}}(X \to z) \leq \max_{y \in \mathcal{Y}} \ell_{P_{XY}}(X \to y).$$

and

(iv) $\bar{\varepsilon}_Z(\delta)$, $\underline{\varepsilon}_Z(\delta)$, and $\varepsilon_c(\delta)$ are all non-increasing in δ for all Z.

Thus, for all $\delta \in (0, 1)$, it holds that

$$\begin{aligned} \varepsilon_c(\delta) &\leq \sup_{\delta' \in (0,1)} \varepsilon_c(\delta') \\ &= \sup_{Z : X - Y - Z}\ \sup_{\delta' \in (0,1)} \bar{\varepsilon}_Z(\delta') \\ &= \sup_{Z : X - Y - Z}\ \max_{z \in \mathcal{Z}} \ell_{P_{XZ}}(X \to z) \\ &\leq \max_{y \in \mathcal{Y}} \ell_{P_{XY}}(X \to y) \\ &\leq \varepsilon. \end{aligned}$$

■

We make a few remarks on Theorem 3. First, when X is a finite random variable, (6) implies that PML is uniformly bounded for all outcomes $y \in \mathcal{Y}$. As a consequence, regardless of the mechanism $P_{Y|X}$, the joint distribution P_{XY} satisfies ε-PML for some finite ε. Second, observe that maximal leakage $\mathcal{L}(X \to Y)$, and consequently, the first term in the bound, depends only on the mechanism $P_{Y|X}$ and is independent of the prior P_X.

B. Lower Bounding the PML envelope

To obtain lower bounds on ε_c, we may restrict attention to specific classes of post-processings and compute the PML δ-quantile of the resulting outputs. This can be viewed as restricting the computational power of the adversary. In general, the choice of post-processings used to derive lower bounds is mechanism dependent (see Section V-B for an illustration). Nevertheless, it is instructive to consider two simple and broadly applicable instances: (i) taking $Z = Y$, i.e., no post-processing, which yields $\bar{\varepsilon}_Y$; and (ii) restricting attention to binary post-processings of Y.

Let

$$\varepsilon_b(\delta) \coloneqq \sup_{Z: X - Y - Z,\ \mathcal{Z}=\{0,1\}} \bar{\varepsilon}_Z(\delta), \quad \delta \in (0,1), \tag{13}$$

denote the *binary* envelope of P_{XY}. Then, for all $\delta \in (0,1)$, we have the lower bound

$$\varepsilon_c(\delta) \geq \max\left\{\bar{\varepsilon}_Y(\delta), \varepsilon_b(\delta)\right\}.$$

One may obtain sharper lower bounds by extending the analysis to ternary, quaternary, and other higher-order k-ary post-processings. As we will see below, the advantage of ε_b is that it is simple to calculate.

The quantity ε_b previously appeared in the work of Saeidian *et al.* [10] (in a slightly different form) as a stand-alone privacy definition. In contrast, here, we use ε_b only as a lower bound on the PML envelope. Algorithm 1 provides a procedure for computing ε_b based on the proof of [10, Thm. 3]. The main idea underlying both this proof and Algorithm 1 is to generalize the notion of PML from individual outcomes to *events*. In particular, given an event $\mathcal{E} \subseteq \mathcal{Y}$ with $P_Y(\mathcal{E}) > 0$, and $Z = \mathbf{1}_{\mathcal{E}}(Y)$, we may define event-wise leakage as

$$\ell_{P_{XY}}(X \to \mathcal{E}) \coloneqq \ell_{P_{XZ}}(X \to 1).$$

Technically, this is not a new concept since event-wise leakage is simply the PML of the affirmative outcome of an indicator function. Nevertheless, it provides a convenient shorthand for reasoning about post-processing. Note that extending PML to events is natural in this context, since any binary post-processing corresponds to selecting an event in $\mathcal{Y}$ and revealing whether or not the outcome Y lies in that event.

Heuristically, Algorithm 1 works as follows: For each $x \in \mathcal{X}$, we compute the largest value of

$$\frac{P_{Y|X=x}(\mathcal{E})}{P_Y(\mathcal{E})},$$

over events $\mathcal{E} \subseteq \mathcal{Y}$ with $P_Y(\mathcal{E}) = \delta$ (with possible randomization at the boundary). The quantity ε_b is then obtained by taking the logarithm of the maximum of this value over all $x \in \mathcal{X}$. Further discussion about the information leakage of events and their connection to post-processing is provided in Appendix A.

Algorithm 1 Computing $\varepsilon_b(\delta)$ for a joint distribution P_{XY}

```
Require: $P_{Y|X}$, $P_X$ and $\delta \in (0,1)$
Ensure: $\varepsilon_b(\delta)$
Compute the marginal $P_Y$
Initialize $M \leftarrow 0$
for all $x \in \mathcal{X}$ do
    for all $y \in \mathcal{Y}$ do
        $s_x(y) \leftarrow \frac{P_{Y|X=x}(y)}{P_Y(y)}$
    end for
    Sort $\mathcal{Y}$ as $(y_1, \ldots, y_{|\mathcal{Y}|})$ so that
        $s_x(y_1) \geq s_x(y_2) \geq \cdots \geq s_x(y_{|\mathcal{Y}|})$
    Find the smallest index $k^\star$ such that $\sum_{j=1}^{k^\star} P_Y(y_j) \geq \delta$
    $p \leftarrow \sum_{j=1}^{k^\star - 1} P_Y(y_j)$
    $\zeta \leftarrow \frac{\delta - p}{P_Y(y_{k^\star})}$
    $v \leftarrow \frac{1}{\delta}\left(\sum_{j=1}^{k^\star-1} P_{Y|X=x}(y_j) + \zeta P_{Y|X=x}(y_{k^\star})\right)$
    $M \leftarrow \max(M, v)$
end for
return $\varepsilon_b \leftarrow \log M$
```

V. Applications

We now calculate and bound the PML envelope in two canonical settings. These are: the *PML–extremal mechanisms* in the high-privacy regime [14] and the *randomized response* mechanism [15, 16].

A. PML-extremal Mechanisms

In [14], Grosse *et al.* investigated the design of utility-optimal privacy mechanisms under the ε–PML constraint for secrets with finite alphabets. They formulated a linear program for maximizing a broad class of convex and sublinear utility functions, and the resulting optimal mechanisms were termed *PML–extremal mechanisms*. Fix a prior distribution P_X on $\mathcal{X} = [k]$ with $k \geq 2$. Grosse *et al.* [14] showed that in the *high-privacy regime*, corresponding to

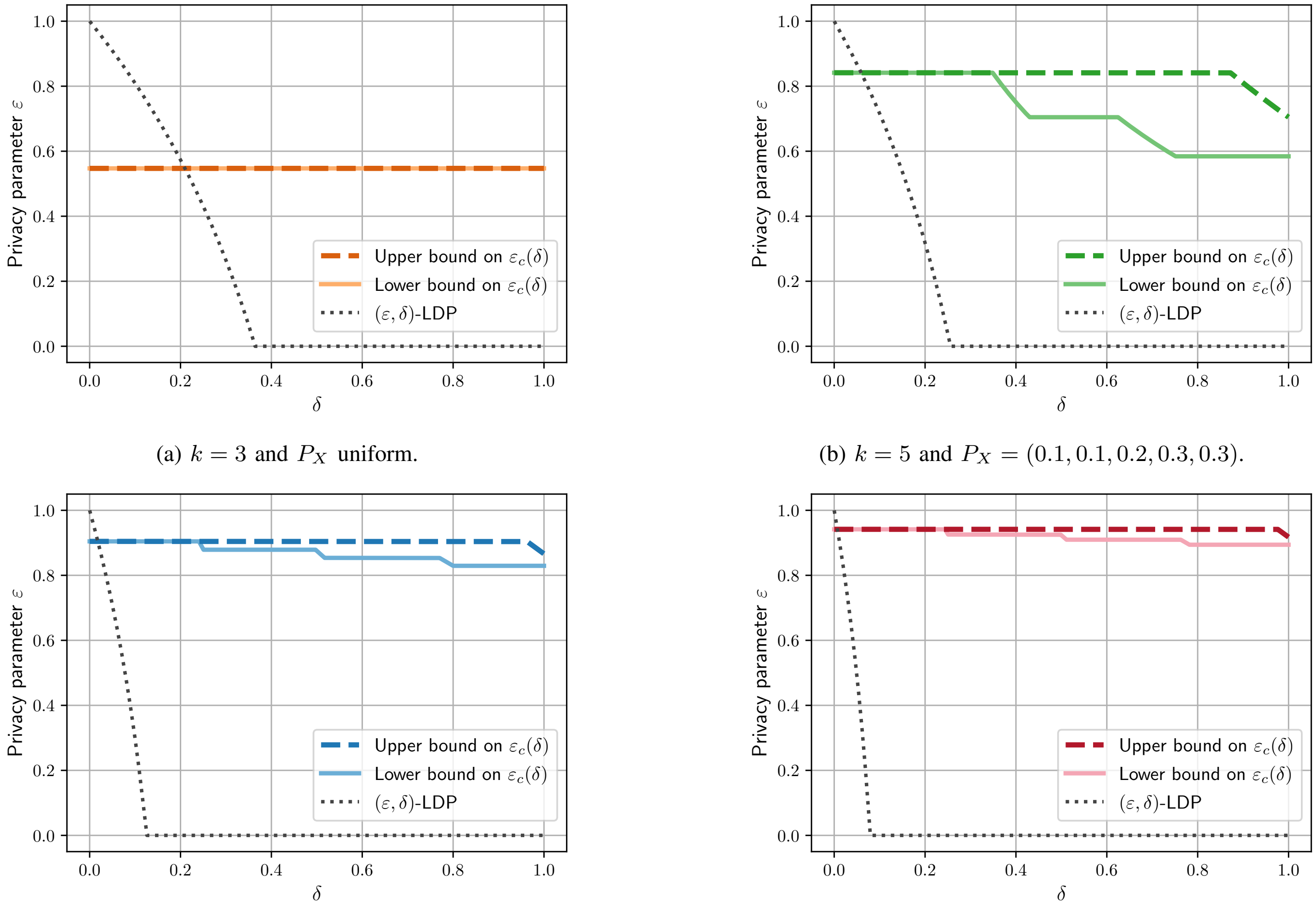
$$0 \leq \varepsilon < \log\left(\frac{1}{1 - \min_{x \in \mathcal{X}} P_X(x)}\right),$$

the optimal mechanism has the form

$$P^*_{Y|X=i}(j) = \begin{cases} 1 - e^{\varepsilon}\big(1 - P_X(i)\big), & \text{if } i = j, \\ e^{\varepsilon} P_X(j), & \text{if } i \neq j, \end{cases}$$

where $\mathcal{Y} = \mathcal{X}$, and $i, j \in [k]$. Note that the outcomes of this mechanism have the PML $\ell(X \to j) = \varepsilon$ for all $j \in [k]$.

Theorem 4. *Let P_X be a distribution on $\mathcal{X} = [k]$, and fix $0 \leq \varepsilon < -\log\left(1 - \min_{x \in \mathcal{X}} P_X(x)\right)$. Let $P^*_{Y|X}$ denote the PML-extremal mechanism. Then, for all $\delta \in (0,1)$, the PML envelope is $\varepsilon_c(\delta) = \varepsilon$.*



(a) $k = 3$ and P_X uniform.

(b) $k = 5$ and $P_X = (0.1, 0.1, 0.2, 0.3, 0.3)$.

(c) $k = 12$ and P_X is the four-level prior with $\rho = 0.1$.

(d) $k = 20$ and P_X is the four-level prior with $\rho = 0.1$.

Fig. 2: The PML envelope and (ε, δ)-LDP guarantees for the k-RR mechanism with $\varepsilon_r = 1.0$. Each sub-figure plots upper and lower bounds on $\varepsilon_c(\delta)$, together with the corresponding (ε, δ)-LDP curve for different alphabet sizes k and priors P_X.

Proof: By Theorem 3, any mechanism satisfying ε-PML has $\varepsilon_c(\delta) \leq \varepsilon$ for all $\delta \in (0, 1)$. We argue that $P^*_{Y|X}$ achieves this upper bound.

Let $\mathcal{A} \subset \mathcal{Y}$ be an arbitrary event with probability $P_Y(\mathcal{A}) \geq \delta$. Since every $j \in \mathcal{Y}$ has leakage ε, we have $\min_{j \in \mathcal{A}} \ell(X \to j) = \varepsilon$. Hence, we get the lower bound

$$\begin{aligned}\varepsilon_c(\delta) \geq \bar{\varepsilon}_Y(\delta) &= \max_{\substack{\mathcal{E} \subseteq \mathcal{Y} \\ P_Y(\mathcal{E}) \geq \delta}} \min_{j \in \mathcal{E}} \ell(X \to j) \\ &\geq \min_{j \in \mathcal{A}} \ell(X \to j) = \varepsilon.\end{aligned}$$

Combining this with the upper bound establishes the claim. ■

B. Randomized Response Mechanism

Given an integer $k \geq 2$, let $\mathcal{X} = \mathcal{Y} = [k]$. The *$k$-randomized response* ($k$-RR) mechanism with parameter $\varepsilon_r > 0$ is defined as

$$P_{Y|X=i}(j) = \begin{cases} \frac{e^{\varepsilon_r}}{e^{\varepsilon_r}+k-1}, & j = i, \\ \frac{1}{e^{\varepsilon_r}+k-1}, & j \neq i, \end{cases} \quad i, j \in [k]. \tag{14}$$

The k-RR mechanism satisfies ε_r-LDP. For simplicity, let

$$\begin{aligned}\alpha &:= \frac{e^{\varepsilon_r}}{e^{\varepsilon_r}+k-1}, \\ \beta &:= \frac{1}{e^{\varepsilon_r}+k-1}, \\ p_i &:= P_X(i), \quad i \in [k], \\ q_j &:= P_Y(j), \quad j \in [k],\end{aligned}$$

and observe that $q_i = \beta + (\alpha - \beta) p_i$ for all $i \in [k]$. Furthermore, each outcome of the mechanism has the PML

$$\ell(X \to j) = \log \frac{\alpha}{q_j}, \quad j \in [k].$$

For simplicity, let us assume that $p_1 \leq p_2 \leq \cdots \leq p_k$, which implies that $q_1 \leq q_2 \leq \cdots \leq q_k$ and $\ell(X \to 1) \geq \ell(X \to 2) \geq \cdots \geq \ell(X \to k)$.

Theorem 5. *Let P_X be a distribution on $\mathcal{X} = [k]$ and let $P_{Y|X}$ be the k-RR mechanism. The PML envelope satisfies the following:*

(i) If $\delta \in (0, q_1]$, then $\varepsilon_c(\delta) = \log \frac{\alpha}{q_1}$.

(ii) For $\delta \in (q_1, 1)$, the PML envelope admits the upper bound

$$\varepsilon_c(\delta) \leq \min \left\{ \log \frac{k\alpha}{\delta}, \log \frac{\alpha}{q_1} \right\}.$$

(iii) For $\delta \in (q_1, 1)$, let $N \in \{2, \ldots, k\}$ be the unique index such that $\sum_{j=1}^{N-1} q_j < \delta \leq \sum_{j=1}^{N} q_j$. Let $\theta := \frac{\delta - \sum_{j=1}^{N-1} q_j}{q_N} \in (0, 1]$, and suppose the prior satisfies

$$p_N \leq \frac{\alpha \sum_{j=1}^{N-1} p_j + \beta}{(N-2)\alpha + \beta}.$$

Then, the PML envelope admits the lower bound

$$\varepsilon_c(\delta) \geq h_\delta(\theta),$$

where

$$h_\delta(\theta) = \begin{cases} \ell(X \to N-1) & \text{if } 0 < \theta \leq \theta_1, \\ \log \frac{(N-1)\alpha + \theta\beta}{\delta} & \text{if } \theta_1 < \theta \leq \theta_2, \\ \ell(X \to N) & \text{if } \theta_2 < \theta \leq 1, \end{cases}$$

and

$$\theta_1 = \frac{\alpha\Big((N-2)q_{N-1} - \sum_{j=1}^{N-2} q_j\Big)}{\alpha q_N - \beta q_{N-1}},$$

$$\theta_2 = \frac{\alpha\Big((N-1)q_N - \sum_{j=1}^{N-1} q_j\Big)}{q_N(\alpha - \beta)}.$$

Proof:

(i) The k-RR mechanism satisfies $\log \frac{\alpha}{q_1}$-PML, so we have the general upper bound $\varepsilon_c(\delta) \leq \log \frac{\alpha}{q_1}$ which holds for all $\delta \in (0, 1)$. Suppose $\delta \in (0, q_1]$, and let $\mathcal{A} = \{1\}$ which satisfies $P_Y(\mathcal{A}) = q_1$. Then, we have

$$\begin{aligned} \varepsilon_c(\delta) \geq \bar{\varepsilon}_Y(\delta) &= \max_{\substack{\mathcal{E} \subseteq \mathcal{Y} \\ P_Y(\mathcal{E}) \geq \delta}} \min_{j \in \mathcal{E}} \ell(X \to j) \\ &\geq \min_{j \in \mathcal{A}} \ell(X \to j) = \log \frac{\alpha}{q_1}. \end{aligned}$$

(ii) We use Theorem 3 and calculate the maximal leakage of the k-RR mechanism:

$$\mathcal{L}(X \to Y) = \log \sum_{j \in [k]} \max_{i \in [k]} P_{Y|X=i}(j) = \log(k\alpha).$$

This yields the upper bound

$$\varepsilon_c(\delta) \leq \mathcal{L}(X \to Y) + \log \frac{1}{\delta} = \log \frac{k\alpha}{\delta}.$$

(iii) To prove the lower bounds, we construct two post-processings of Y, denoted by W and Z, and pick the one providing the tighter bound.

Given N and θ, let $P_{W|Y}$ be a channel with output alphabet $\mathcal{W} = [N+1]$ expressed as

$$P_{W|Y=j}(w) = \begin{cases} 1, & \text{if } j \in [N-1] \text{ and } w = j, \\ \theta, & \text{if } j = w = N, \\ 1, & \text{if } j > N \text{ and } w = N+1, \\ 0, & \text{otherwise.} \end{cases}$$

Consider the subset $\mathcal{A} = [N]$ of outcomes of W and observe that

$$P_W(\mathcal{A}) = \sum_{w=1}^{N} P_W(w) = \sum_{w=1}^{N-1} q_w + \theta q_N = \delta.$$

In addition, each $w \in \mathcal{A}$ has PML

$$\ell_{P_{XW}}(X \to w) = \log \frac{\max_x P_{W|X}(w \mid x)}{P_Z(w)} = \log \frac{\alpha}{q_w}.$$

Thus, using W, we obtain the lower bound

$$\varepsilon_c(\delta) = \bar{\varepsilon}_W(\delta) \geq \min_{w \in \mathcal{A}} \ell_{P_{XW}}(X \to w) = \log \frac{\alpha}{q_N}.$$

Next, we define a post-processing channel $P_{Z|Y}$ with the output alphabet $\mathcal{Z} = [N]$. Let $\eta = (\eta_1, \ldots, \eta_N)$ be a tuple satisfying

$$\begin{aligned} \eta_z \geq 0, \quad &\text{for all } z \in [N], \\ \sum_{z=1}^{N-1} \eta_z &= \theta, \\ \sum_{z=1}^{N} \eta_z &= 1, \end{aligned} \tag{15}$$

and let

$$P_{Z|Y=j}(z) = \begin{cases} 1, & \text{if } j \in [N-1] \text{ and } z = j, \\ \eta_z, & \text{if } j = N \text{ and } z \in [N], \\ 1, & \text{if } j > N \text{ and } z = N, \\ 0, & \text{otherwise.} \end{cases}$$

In words, each k-RR outcome $j \in [N-1]$ is deterministically mapped to $z \in [N-1]$, all $j > N$ are mapped to a single catch-all symbol $z = N$, and $j = N$ is split across $z = 1, \ldots, N$ with weights $\eta_1, \ldots, \eta_N$.

Consider the event $\mathcal{B} = [N-1]$ of outcomes of Z, and note that by construction, we have

$$\begin{aligned} P_Z(\mathcal{B}) &= \sum_{z=1}^{N-1} P_Z(z) \\ &= \sum_{z=1}^{N-1} q_z + \eta_z q_N \\ &= \sum_{z=1}^{N-1} q_z + \theta q_N = \delta. \end{aligned}$$

Thus, we may use the set $\mathcal{B}$ to obtain lower bounds on the PML envelope:

$$\varepsilon_c(\delta) = \bar{\varepsilon}_Z(\delta) \geq \min_{z \in \mathcal{B}} \ell_{P_{XZ}}(X \to z).$$

Our goal is to optimize the weights $\{\eta_i\}_{i=1}^N$ in order to obtain the tightest possible lower bound. Note that if $N = 2$, then (15) forces $\eta_1 = \theta$, so $P_{Z|Y}$ is fully specified. Therefore, for the rest of the proof assume that $N > 2$.

We begin by calculating the PML for symbols in the set $\mathcal{B}$. For each $i \in [N-1]$, using the structure of the k-RR mechanism and $P_{Z|Y}$, we observe that

$$\max_x P_{Z|X}(i \mid x) = P_{Z|X}(i \mid i) = \alpha + \eta_i \beta,$$
$$P_Z(i) = q_i + \eta_i q_N.$$

Hence, the PML is

$$\ell_{P_{XZ}}(X \to i) = \log \frac{\max_x P_{Z|X}(i \mid x)}{P_Z(i)} = \log \frac{\alpha + \eta_i \beta}{q_i + \eta_i q_N},$$

Let

$$M_i(\eta_i) \coloneqq \frac{\alpha + \eta_i \beta}{q_i + \eta_i q_N}, \quad i = 1, \dots, N-1,$$

and consider the optimization problem:

$$\begin{aligned} &\max_{\eta_1, \dots, \eta_{N-1}} \min_{i \in [N-1]} M_i(\eta_i), \\ &\text{subject to} \sum_{i=1}^{N-1} \eta_i = \theta, \\ &\eta_i \geq 0, \ i = 1, \dots, N-1. \end{aligned}$$

It is easy to verify that $M_i''(\eta_i) \geq 0$, so M_i is convex. Thus, the above optimization problem is *not* a convex one (since the minimum of a collection of convex functions need not be convex). Nevertheless, we can solve it by inspection.

First regime. Let us start by noting that $M_i'(\eta_i) < 0$ for $\eta_i \geq 0$, implying that $M_i(\eta_i) \leq M_i(0) = \frac{\alpha}{q_i}$ for all $i \in [N-1]$. This yields the upper bound on the objective

$$\max_{\eta_1, \dots, \eta_{N-1} \geq 0} \min_{i \in [N-1]} M_i(\eta_i) \leq \min_{i \in [N-1]} \frac{\alpha}{q_i} = \frac{\alpha}{q_{N-1}}, \tag{16}$$

since $q_1 \leq \dots \leq q_k$. This bound is achievable at $\theta = 0$ since $\eta_1 = \dots = \eta_{N-1} = 0$ is feasible at this point.[2] Next, we argue that there exists $\theta_1 \geq 0$ such that the upper bound in (16) is achievable for $\theta \in (0, \theta_1]$. This is because in order to achieve $\frac{\alpha}{q_{N-1}}$, all we really need is to have $\eta_{N-1} = 0$ and $M_i(\eta_i) \geq \frac{\alpha}{q_{N-1}}$ for $i = 1, \dots, N-2$.

Let η_i^* be such that $M_i(\eta_i^*) = \frac{\alpha}{q_{N-1}}$, that is,

$$\begin{aligned} &M_i(\eta_i^*) = \frac{\alpha + \eta_i^* \beta}{q_i + \eta_i^* q_N} = \frac{\alpha}{q_{N-1}} \iff \\ &\eta_i^* = \frac{\alpha(q_{N-1} - q_i)}{\alpha q_N - \beta q_{N-1}} \geq 0, \quad i = 1, \dots, N-2, \end{aligned}$$

and also $\eta_{N-1}^* = 0$. This choice of the parameters yields

$$\theta_1 = \sum_{i=1}^{N-1} \eta_i^* = \frac{\alpha\Big((N-2)q_{N-1} - \sum_{i=1}^{N-2} q_i\Big)}{\alpha q_N - \beta q_{N-1}}.$$

Therefore, assuming that $\theta_1 > 0$, for $\theta \in (0, \min\{\theta_1, 1\}]$, we have the first piece of the lower bound

$$\begin{aligned} \varepsilon_c(\delta) &\geq \min_{z \in \mathcal{A}} \ell_{P_{XZ}}(X \to z) \\ &= \min_{i \in [N-1]} \log M_i(\eta_i^*) = \log \frac{\alpha}{q_{N-1}}. \end{aligned}$$

Note that in this regime, we use the lower bound obtained from Z and not W, since $\frac{\alpha}{q_{N-1}} \geq \frac{\alpha}{q_N}$.

Second regime. Next, suppose $\theta_1 < 1$ and $\theta > \theta_1$. In the second regime, we are forced to increase at least one η_i beyond η_i^*, so the objective falls below $\frac{\alpha}{q_{N-1}}$. Let $\{\tilde{\eta}_i\}$ denote the optimal parameters. There exists a common threshold $\tau \in [\frac{\alpha+\beta}{q_1+q_N}, \frac{\alpha}{q_{N-1}}]$ such that

$$M_i(\tilde{\eta}_i) = \frac{\alpha + \tilde{\eta}_i \beta}{q_i + \tilde{\eta}_i q_N} = \tau \iff \tilde{\eta}_i(\tau) = \frac{\alpha - \tau q_i}{\tau q_N - \beta} > 0,$$

for $i = 1, \dots, N-1$.[3] Hence, θ can be expressed as

$$\theta(\tau) = \sum_{i=1}^{N-1} \tilde{\eta}_i(\tau) = \frac{(N-1)\alpha - \tau \sum_{i=1}^{N-1} q_i}{\tau q_N - \beta}.$$

Solving for τ gives

$$\tau(\theta) = \frac{(N-1)\alpha + \theta\beta}{\sum_{i=1}^{N-1} q_i + \theta q_N} = \frac{(N-1)\alpha + \theta\beta}{\delta}, \quad \theta > \theta_1.$$

Therefore, for $\theta \in (\theta_1, \theta_2]$ (with θ_2 specified below), we get the middle piece of the lower bound

$$\varepsilon_c(\delta) \geq \min_{i \in [N-1]} \log M_i(\tilde{\eta}_i) = \log \frac{(N-1)\alpha + \theta\beta}{\delta}. \tag{17}$$

Third regime. The point θ_2 is the value where any further increase in θ would make the lower bound in (17) drop below the lower bound obtained from W, i.e.,

$$\tau(\theta_2) = \frac{\alpha}{q_N} \iff \theta_2 = \frac{\alpha\Big((N-1)q_N - \sum_{i=1}^{N-1} q_i\Big)}{q_N(\alpha - \beta)}.$$

Thus, in the third regime $\theta \in (\theta_2, 1]$ we have

$$\varepsilon_c(\delta) \geq \bar{\varepsilon}_W(\delta) \geq \log \frac{\alpha}{q_N}.$$

Finally, we find the condition on the prior distribution ensuring that $\theta_2 \leq 1$. Two conditions need to be satisfied for this: We require $\theta_1 \leq 1$ (ensuring that we enter the second regime) and also $\tau(1) \leq \frac{\alpha}{q_N}$ (ensuring that we enter the third regime). By using $q_i = \beta + (\alpha - \beta)p_i$, and after some algebra, we obtain the following conditions:

$$p_{N-1} \leq \frac{\alpha \sum_{i=1}^{N} p_i + \beta}{(N-1)\alpha + \beta}, \tag{18}$$

$$p_N \leq \frac{\alpha \sum_{i=1}^{N-1} p_i + \beta}{(N-2)\alpha + \beta}. \tag{19}$$

[2] Technically, we assume that $\theta > 0$, but we may consider the limiting value of the objective as $\theta \downarrow 0$ since the M_i's are continuous.

[3] Note that each M_i is continuous and strictly decreasing. Therefore, if $M_i(\eta_i) > M_j(\eta_j)$, then there exists $\zeta > 0$ such that $M_j(\eta_j) < M_j(\eta_j - \zeta) = M_i(\eta_i + \zeta)$. Thus, the optimal parameters must yield a common value for all M_i's.

Observe that (19) can be written in the form

$$(N-2)\alpha+\beta \leq \frac{\alpha \sum_{i=1}^{N-1} p_i + \beta}{p_N},$$

which implies that

$$\frac{\alpha \sum_{i=1}^{N} p_i + \beta}{(N-1)\alpha+\beta} \geq \frac{\alpha \sum_{i=1}^{N} p_i + \beta}{\alpha + \frac{\alpha \sum_{i=1}^{N-1} p_i + \beta}{p_N}} = p_N \geq p_{N-1},$$

so if (19) is satisfied, (18) is also automatically satisfied.

■

The lower bounds in Theorem 5 are stated under the assumption that $0 < \theta_1 < \theta_2$. The following degenerate cases are handled by interpreting the piecewise definition in the natural way: (i) If $N = 2$ or $q_1 = \cdots = q_{N-1}$, then $\theta_1 = 0$, so the first regime is vacuous and omitted; (ii) If $q_{N-1} = q_N$, then $\theta_1 = \theta_2$, so the second regime is vacuous and omitted. In this case, the bound equals $\ell(X \to N-1) = \ell(X \to N)$ for all $\theta \in (0, 1]$.

To simplify the prior distribution on larger alphabets, we use the following *four-level* construction obtained by partitioning the alphabet into four blocks of equal size (assuming k is divisible by 4). Fix $\rho \in (0, 1/3)$ and define

$$m_0 = 1-3\rho, \quad m_1 = 1-\rho, \quad m_2 = 1+\rho, \quad m_3 = 1+3\rho.$$

For $r \in \{0, 1, 2, 3\}$, set

$$p_i = \frac{m_r}{k}, \quad i \in \{\frac{k}{4}r+1, \ldots, \frac{k}{4}(r+1)\}.$$

Then, P_X is constant on each block and satisfies $\sum_{i=1}^{k} p_i = 1$.

VI. Conclusion

The conclusion goes here.

References

[1] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of cryptography conference*, Springer, 2006, pp. 265–284. DOI: 10.1007/11681878_14.

[2] C. Dwork and A. Roth, "The Algorithmic Foundations of Differential Privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, no. 3-4, pp. 211–407, Aug. 2014, ISSN: 1551-305X. DOI: 10.1561/0400000042.

[3] A. Machanavajjhala, D. Kifer, J. Abowd, J. Gehrke, and L. Vilhuber, "Privacy: Theory meets practice on the map," in *2008 IEEE 24th international conference on data engineering*, IEEE, 2008, pp. 277–286.

[4] D. Kifer and B.-R. Lin, "An axiomatic view of statistical privacy and utility," *Journal of Privacy and Confidentiality*, vol. 4, no. 1, 2012.

[5] C. Dwork, G. N. Rothblum, and S. Vadhan, "Boosting and Differential Privacy," in *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, Las Vegas, NV, USA: IEEE, Oct. 2010, pp. 51–60, ISBN: 978-1-4244-8525-3. DOI: 10.1109/FOCS.2010.12.

[6] S. Meiser, "Approximate and probabilistic differential privacy definitions," *Cryptology ePrint Archive*, 2018. [Online]. Available: https://eprint.iacr.org/2018/277.

[7] I. Mironov, "Rényi Differential Privacy," in *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, IEEE Computer Society, Aug. 2017, pp. 263–275, ISBN: 978-1-5386-3217-8. DOI: 10.1109/CSF.2017.11.

[8] M. Bun and T. Steinke, "Concentrated Differential Privacy: Simplifications, Extensions, and Lower Bounds," in *Proceedings, Part I, of the 14th International Conference on Theory of Cryptography - Volume 9985*, Berlin, Heidelberg: Springer-Verlag, Oct. 2016, pp. 635–658, ISBN: 978-3-662-53640-7. DOI: 10.1007/978-3-662-53641-4_24.

[9] J. Dong, A. Roth, and W. J. Su, "Gaussian Differential Privacy," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 84, no. 1, pp. 3–37, Feb. 2022, ISSN: 1369-7412. DOI: 10.1111/rssb.12454.

[10] S. Saeidian, G. Cervia, T. J. Oechtering, and M. Skoglund, "Pointwise maximal leakage," *IEEE Transactions on Information Theory*, vol. 69, no. 12, pp. 8054–8080, 2023. DOI: 10.1109/TIT.2023.3304378.

[11] S. Saeidian, G. Cervia, T. J. Oechtering, and M. Skoglund, "Rethinking disclosure prevention with pointwise maximal leakage," *Journal of Privacy and Confidentiality*, vol. 15, no. 1, Mar. 2025. DOI: 10.29012/jpc.893.

[12] M. S. Alvim, K. Chatzikokolakis, A. McIver, C. Morgan, C. Palamidessi, and G. Smith, *The Science of Quantitative Information Flow*. Springer Cham, 2020.

[13] I. Issa, A. B. Wagner, and S. Kamath, "An operational approach to information leakage," *IEEE Transactions on Information Theory*, vol. 66, no. 3, pp. 1625–1657, 2019. DOI: 10.1109/TIT.2023.3341148.

[14] L. Grosse, S. Saeidian, and T. J. Oechtering, "Extremal Mechanisms for Pointwise Maximal Leakage," *IEEE Transactions on Information Forensics and Security*, vol. 19, pp. 7952–7967, 2024, ISSN: 1556-6021. DOI: 10.1109/TIFS.2024.3449556.

[15] S. L. Warner, "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 63–69, 1965, ISSN: 0162-1459. DOI: 10.2307/2283137. JSTOR: 2283137.

[16] P. Kairouz, S. Oh, and P. Viswanath, "Extremal mechanisms for local differential privacy," *Journal of Machine Learning Research*, vol. 17, no. 1, pp. 492–542, Jan. 2016, ISSN: 1532-4435.

[17] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, "What can we learn privately?" In *49th Annual IEEE Symposium on Foundations of Computer Science*, 2008, pp. 531–540.

[18] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy and statistical minimax rates," in *IEEE 54th Annual Symposium on Foundations of Computer Science*, 2013, pp. 429–438.

[19] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *Proceedings of the 25th Annual International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT)*, vol. 4004, Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 486–503. DOI: 10.1007/11761679_29.

[20] B. Balle, G. Barthe, and M. Gaboardi, "Privacy amplification by subsampling: Tight analyses via couplings and divergences," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, ser. NIPS'18, Red Hook, NY, USA: Curran Associates Inc., 2018, pp. 6280–6290.

[21] C. L. Canonne, G. Kamath, and T. Steinke, "The Discrete Gaussian for Differential Privacy," *Journal of Privacy and Confidentiality*, vol. 12, no. 1, 2022. DOI: 10.29012/jpc.784.

[22] S. P. Kasiviswanathan and A. Smith, "On the 'Semantics' of Differential Privacy: A Bayesian Formulation," *Journal of Privacy and Confidentiality*, vol. 6, no. 1, 2014, ISSN: 2575-8527. DOI: 10.29012/jpc.v6i1.634.

[23] S. Jonany, *Correcting (ϵ, δ) misconception in differential privacy*, 2022. [Online]. Available: https://medium.com/@sjonany/correcting-%CF%B5-%CE%B4-misconception-in-differential-privacy-e830dbdce0ab.

[24] D. Kifer and B.-R. Lin, "An Axiomatic View of Statistical Privacy and Utility," *Journal of Privacy and Confidentiality*, vol. 4, no. 1, Jul. 2012, ISSN: 2575-8527. DOI: 10.29012/jpc.v4i1.610.

[25] 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*, 2014, pp. 308–322. DOI: 10.1109/CSF.2014.29.

[26] A. McIver, C. Morgan, G. Smith, B. Espinoza, and L. Meinicke, "Abstract Channels and Their Robust Information-Leakage Ordering," in *Principles of Security and Trust*, Berlin, Heidelberg: Springer, 2014, pp. 83–102, ISBN: 978-3-642-54792-8. DOI: 10.1007/978-3-642-54792-8_5.

Appendix A
Measuring the Information Leaked to Events

Given an event $\mathcal{E} \subseteq \mathcal{Y}$ with $P_Y(\mathcal{E}) > 0$ and a random variable $Z = \mathbf{1}_{\mathcal{E}}(Y)$, we define

$$\ell_{P_{XY}}(X \to \mathcal{E}) \coloneqq \ell_{P_{XZ}}(X \to 1).$$

This definition is motivated by the observation that both deterministic and randomized post-processings can be naturally expressed in terms of indicator functions. For deterministic mappings, this is immediate: If $Z = h(Y)$ and $z \in \mathcal{Z}$, then

$$\ell_{P_{XZ}}(X \to z) = \ell_{P_{XY}}(X \to \mathcal{E}_z),$$

where $\mathcal{E}_z = \{y \in \mathcal{Y} : h(y) = z\}$ denotes the pre-image of z under h.

We can also cast randomized mappings as deterministic ones by adopting a few formalisms from [26] and [10]. Given a privacy mechanism $P_{Y|X}$, we say that two outcomes y, y' are *similar* if there exists a constant $c > 0$ such that $P_{Y|X=x}(y) = cP_{Y|X=x}(y')$ for all $x \in \mathcal{X}$. Similar outcomes have the same information density $i(x; y) = i(x; y')$ for all $x \in \mathcal{X}$, and induce the same posterior distributions $P_{X|Y=y} = P_{X|Y=y'}$. Consequently, “merging” similar outcomes (i.e., mapping similar outcomes to the same symbol) does not alter the distribution of $\ell(X \to Y)$. In [26], the mechanism obtained by merging all similar outcomes is called the *reduced mechanism*. As an example, for the mechanism $P_{Y|X}$ in (7), outcomes 3 and 4 are similar, and its reduced form is

$$P_{Y_r|X} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0.2 & 0.8 \\ 0.2 & 0 & 0.8 \end{bmatrix}.$$

Next, we define an equivalence relation that unifies all mechanisms with the same reduced form. Let $[P_{Y|X}]$ denote the equivalence class of $P_{Y|X}$. A key advantage of introducing such equivalence classes is that a randomized post-processing applied to the outputs of $P_{Y|X}$ can alternatively be viewed as a deterministic post-processing applied to some mechanism in $[P_{Y|X}]$. While this idea can be established formally, we illustrate it with a simple example that readily extends to a general proof. Fix a mechanism $P_{Y|X}$ with binary output alphabet $\mathcal{Y} = \{0, 1\}$, and consider a randomized post-processing $P_{Z|Y}$ of the form

$$P_{Z|Y} = \begin{bmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{bmatrix},$$

with $0 < \alpha, \beta < 1$. Define a mechanism $P_{\tilde{Y}|X}$ with the output space $\tilde{\mathcal{Y}} = \{00, 01, 10, 11\}$ by

$$\begin{aligned} P_{\tilde{Y}|X=x}(00) &= \alpha P_{Y|X=x}(0), \\ P_{\tilde{Y}|X=x}(01) &= (1-\alpha) P_{Y|X=x}(0), \\ P_{\tilde{Y}|X=x}(10) &= (1-\beta) P_{Y|X=x}(1), \\ P_{\tilde{Y}|X=x}(11) &= \beta P_{Y|X=x}(1), \end{aligned}$$

for all x. Then, $P_{Y|X}$ and $P_{\tilde{Y}|X}$ belong to the same equivalence class, and we have

$$\begin{aligned} \ell_{P_{XZ}}(X \to 0) &= \ell_{P_{X\tilde{Y}}}(X \to \{00, 10\}), \\ \ell_{P_{XZ}}(X \to 1) &= \ell_{P_{X\tilde{Y}}}(X \to \{01, 11\}), \end{aligned}$$

that is, Z is a deterministic function of $\tilde{Y}$. Extending this construction to general post-processings shows that any randomized post-processing of Y can be treated as a deterministic mapping applied to some mechanism in the equivalence class $[P_{Y|X}]$.

Below, we establish some elementary properties of the map $\mathcal{E} \mapsto \ell_{P_{XY}}(X \to \mathcal{E})$. As a preliminary step, let us express $\ell_{P_{XY}}(X \to \mathcal{E})$ in terms of the information density:

$$\begin{aligned} \ell_{P_{XY}}(X \to \mathcal{E}) &= \log \max_{x \in \mathcal{X}} \frac{P_{Y|X=x}(\mathcal{E})}{P_Y(\mathcal{E})} \\ &= \log \max_{x \in \mathcal{X}} \frac{\sum_{y \in \mathcal{E}} P_{Y|X=x}(y)}{P_Y(\mathcal{E})} \\ &= \log \max_{x \in \mathcal{X}} \frac{\sum_{y \in \mathcal{E}} \exp\big(i(x; y)\big) P_Y(y)}{P_Y(\mathcal{E})} \\ &= \log \max_{x \in \mathcal{X}} \mathbb{E}_{Y \sim Q_{\mathcal{E}}}\Big[\exp\big(i(x; Y)\big)\Big], \end{aligned} \tag{20}$$

where $Q_{\mathcal{E}}$ is the conditional distribution of Y given $\mathcal{E}$, that is,

$$Q_{\mathcal{E}}(y) = \begin{cases} \frac{P_Y(y)}{P_Y(\mathcal{E})} & \text{if } y \in \mathcal{E}, \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 1. The function $\ell(X \to \mathcal{E})$ satisfies the following properties:

(i) $0 \le \ell(X \to \mathcal{E}) \le \log \frac{1}{P_Y(\mathcal{E})}$ for all $\mathcal{E} \subseteq \mathcal{Y}$.

(ii) $\ell(X \to \mathcal{Y}) = 0$.

(iii) If $\mathcal{E} \cap \mathcal{E}' = \emptyset$, then

$$\ell(X \to \mathcal{E} \cup \mathcal{E}') \le \max\Big\{\ell(X \to \mathcal{E}), \ell(X \to \mathcal{E}')\Big\}.$$

(iv) Suppose $P_Y(\mathcal{E}) = \theta > 0$. For each $0 < \theta' < \theta$, there exists an event $\mathcal{E}' \subset \mathcal{E}$ with probability $P_Y(\mathcal{E}') = \theta'$ such that

$$\ell(X \to \mathcal{E}') \ge \ell(X \to \mathcal{E}).$$

Proof:

(i) These bounds follow immediately from the definition of $\ell(X \to \mathcal{E})$ and were also noted in [10].

(ii)

$$\ell_{P_{XY}}(X \to \mathcal{Y}) = \log \max_{x \in \mathcal{X}} \frac{P_{Y|X=x}(\mathcal{Y})}{P_Y(\mathcal{Y})} = \log \frac{1}{1} = 0.$$

(iii) Suppose $\mathcal{E} \cap \mathcal{E}' = \emptyset$. We have

$$\begin{aligned} \ell(X \to \mathcal{E} \cup \mathcal{E}') &= \log \max_x \frac{P_{Y|X=x}(\mathcal{E} \cup \mathcal{E}')}{P_Y(\mathcal{E} \cup \mathcal{E}')} \\ &= \log \max_x \frac{P_{Y|X=x}(\mathcal{E}) + P_{Y|X=x}(\mathcal{E}')}{P_Y(\mathcal{E}) + P_Y(\mathcal{E}')} \\ &\le \log \max_x \max\left\{\frac{P_{Y|X=x}(\mathcal{E})}{P_Y(\mathcal{E})}, \frac{P_{Y|X=x}(\mathcal{E}')}{P_Y(\mathcal{E}')}\right\} \end{aligned}$$

$$= \max\Big\{\ell(X \to \mathcal{E}), \ell(X \to \mathcal{E}')\Big\}.$$

(iv) Fix $x \in \mathcal{X}$ satisfying

$$\log \frac{P_{Y|X=x}(\mathcal{E})}{P_Y(\mathcal{E})} \geq \ell(X \to \mathcal{E}).$$

Choose $\tau > 0$ so that

$$\begin{aligned} P_Y\{y \in \mathcal{E} : \exp\big(i(x;y)\big) > \tau\} &\leq \theta' \\ &\leq P_Y\{y \in \mathcal{E} : \exp\big(i(x;y)\big) \geq \tau\}. \end{aligned}$$

Let $\mathcal{A} = \{y \in \mathcal{E} : \exp\big(i(x;y)\big) > \tau\}$. If $P_Y(\mathcal{A}) < \theta'$, choose a set $\mathcal{B} \subseteq \{y \in \mathcal{E} : \exp\big(i(x;y)\big) = \tau\}$ so that $P_Y(\mathcal{A}) + P_Y(\mathcal{B}) = \theta'$ and let $\mathcal{E}' = \mathcal{A} \cup \mathcal{B}$.[4] If $P_Y(\mathcal{A}) = \theta'$, then let $\mathcal{E}' = \mathcal{A}$. By construction, we have

$$\begin{aligned} &\exp\big(i(x;y)\big) \geq \tau, \quad y \in \mathcal{E}', \\ &\exp\big(i(x;y)\big) \leq \tau, \quad y \in \mathcal{E} \setminus \mathcal{E}'. \end{aligned}$$

Now, we can write

$$\begin{aligned} \ell(X \to \mathcal{E}) &\leq \log \frac{P_{Y|X=x}(\mathcal{E})}{P_Y(\mathcal{E})} \\ &= \log \frac{P_{Y|X=x}(\mathcal{E}') + P_{Y|X=x}(\mathcal{E} \setminus \mathcal{E}')}{P_Y(\mathcal{E}') + P_Y(\mathcal{E} \setminus \mathcal{E}')} \\ &\leq \log \max\left\{\frac{P_{Y|X=x}(\mathcal{E}')}{P_Y(\mathcal{E}')}, \frac{P_{Y|X=x}(\mathcal{E} \setminus \mathcal{E}')}{P_Y(\mathcal{E} \setminus \mathcal{E}')}\right\} \\ &= \log \frac{P_{Y|X=x}(\mathcal{E}')}{P_Y(\mathcal{E}')} \qquad (21a) \\ &\leq \ell(X \to \mathcal{E}'), \end{aligned}$$

where (21a) follows because

$$\begin{aligned} \frac{P_{Y|X=x}(\mathcal{E}')}{P_Y(\mathcal{E}')} &= \mathbb{E}_{Y \sim Q_{\mathcal{E}'}}\Big[\exp\big(i(x;Y)\big)\Big] \geq \tau, \\ \frac{P_{Y|X=x}(\mathcal{E} \setminus \mathcal{E}')}{P_Y(\mathcal{E} \setminus \mathcal{E}')} &= \mathbb{E}_{Y \sim Q_{\mathcal{E} \setminus \mathcal{E}'}}\Big[\exp\big(i(x;Y)\big)\Big] \leq \tau. \end{aligned}$$

■

We make some further remarks about the event-wise leakage. First, in general, the reverse of Lemma 1(iii) need not hold, i.e., one cannot claim that $\ell(X \to \mathcal{E} \cup \mathcal{E}')$ upper bounds $\min\Big\{\ell(X \to \mathcal{E}), \ell(X \to \mathcal{E}')\Big\}$.

Example 4. Let X be an unbiased Bernoulli random variable, and let $\mathcal{Y} = \{1,2,3\}$. Consider the mechanism

$$P_{Y|X} = \begin{bmatrix} 0.9 & 0 & 0.1 \\ 0 & 0.9 & 0.1 \end{bmatrix},$$

which induces the marginal distribution P_Y with

$$P_Y(1) = 0.45, \qquad P_Y(2) = 0.45, \qquad P_Y(3) = 0.1.$$

Let $\mathcal{E} = \{1\}$ and $\mathcal{E}' = \{2\}$. Then, we have

$$\begin{aligned} \ell(X \to \mathcal{E}) &= \ell(X \to \mathcal{E}') \\ &= \log \max_{x \in \{0,1\}} \frac{P_{Y|X=x}(\{1\})}{P_Y(\{1\})} \\ &= \log \frac{0.9}{0.45} \\ &= \log 2. \end{aligned}$$

However, for their union $\mathcal{E} \cup \mathcal{E}' = \{1,2\}$ we have

$$\ell(X \to \mathcal{E} \cup \mathcal{E}') = \log \max_{x \in \{0,1\}} \frac{P_{Y|X=x}(\{1,2\})}{P_Y(\{1,2\})} = \log 1 = 0.$$

Thus, $\ell(X \to \mathcal{E} \cup \mathcal{E}') = 0 < \min\{\ell(X \to \mathcal{E}), \ell(X \to \mathcal{E}')\} = \log 2$.

Second, the event-wise leakage is, in general, not monotone. That is, given events $\mathcal{E} \subset \mathcal{E}'$, either $\ell(X \to \mathcal{E})$ or $\ell(X \to \mathcal{E}')$ can be larger.

Example 5. Recall the setup of Example 4. Let $\mathcal{E} = \{3\}$ and $\mathcal{E}' = \{1,3\}$, so that $\mathcal{E} \subset \mathcal{E}'$. Then, we have

$$\ell(X \to \mathcal{E}) = \log \max_x \frac{P_{Y|X=x}(\{3\})}{P_Y(\{3\})} = \log \frac{0.1}{0.1} = 0,$$

whereas

$$\ell(X \to \mathcal{E}') = \log \max_x \frac{P_{Y|X=x}(\{1,3\})}{P_Y(\{1,3\})} = \log \frac{1}{0.55} > 0.$$

Now, let $\mathcal{F} = \{1\}$ and $\mathcal{F}' = \{1,3\}$, so that $\mathcal{F} \subset \mathcal{F}'$. We have

$$\ell(X \to \mathcal{F}) = \log 2,$$

while, as computed above,

$$\ell(X \to \mathcal{F}') = \log\left(\frac{1}{0.55}\right) < \log 2.$$

A. *Connection to Prior Work and Derivation of Algorithm 1*

Our definition of the event-wise leakage, as well as the procedure in Algorithm 1 are closely related to a privacy guarantee studied in [10]. In particular, Saeidian *et al.* [10] introduced a guarantee based on the worst-case PML of post-processed outcomes with probability at least δ, namely the quantity

$$\sup_{Z : X - Y - Z} \; \max_{z \in \mathcal{Z} : P_Z(z) \geq \delta} \ell_{P_{XZ}}(X \to z). \qquad (22)$$

Then, they showed that (22) admits an equivalent formulation in terms of the worst-case information leaked to events $\mathcal{E} \subseteq \mathcal{Y}$ with probability $P_Y(\mathcal{E}) \geq \delta$ [10, Thm. 3]. It follows immediately from this equivalence (and the connection between events and indicator functions discussed above) that the quantity in (22) coincides with the binary PML envelope ε_b defined in (13).

Saeidian *et al.* [10] also characterized the solution to (22). Fix $x \in \mathcal{X}$ and order the outputs $y \in \mathcal{Y}$ in decreasing information density $i(x;y)$. Let $\mathcal{F}_k$ denote the set consisting of the first k outputs in this ordering, and let $k^\star$ be the smallest index such that $P_Y(\mathcal{F}_{k^\star}) \geq \delta$. If $P_Y(\mathcal{F}_{k^\star}) > \delta$, the optimal construction uses randomization at the boundary output so that the selected event has probability exactly δ. For each $x \in \mathcal{X}$, this yields the value

$$\kappa(x) = \frac{1}{\delta}\Big(P_{Y|X=x}(\mathcal{F}_{k^\star - 1}) + \zeta \, P_{Y|X=x}(y_{k^\star})\Big), \qquad (23)$$

[4] To select such $\mathcal{B}$, we might need to use some other mechanism in $[P_{Y|X}]$.

where $\zeta \in (0,1]$ is chosen so that $P_Y(\mathcal{F}_{k^\star-1})+\zeta P_Y(y_{k^\star}) = \delta$. Finally, the binary envelope is obtained as

$$\varepsilon_b(\delta) = \log \max_{x\in\mathcal{X}} \kappa(x).$$

Algorithm 1 formalizes this procedure.

Appendix B
Proof of

Lemma 2. For all $\delta \in (0,1)$, it holds that

$$\inf\{t \geq 0 : C_Z(t) \geq 1-\delta\} = \min_{\substack{\mathcal{A}\subset\mathcal{Z}\\ P_Z(\mathcal{A})\geq 1-\delta}} \max_{z\in\mathcal{A}} \ell(z).$$

Proof. Let

$$r_1 \coloneqq \min_{\substack{\mathcal{A}\subset\mathcal{Z}\\ P_Z(\mathcal{A})\geq 1-\delta}} \max_{z\in\mathcal{A}} \ell(z),$$

$$r_2 \coloneqq \inf\{t\geq 0 : P_Z\{z : \ell(z) \leq t\} \geq 1-\delta\}.$$

Given $r \geq 0$, let $\mathcal{J}_r = \big\{z : \ell(z) \leq r\big\}$, and note that $\max_{z\in\mathcal{J}_r} \ell(z) = r$. By definition, if $r < r_2$, then $P_Z(\mathcal{J}_r) < 1-\delta$, therefore $P_Z(\mathcal{J}_r) \geq 1-\delta$ for all $r \geq r_2$. Therefore, we have

$$\begin{aligned} r_1 &= \min_{\substack{\mathcal{A}\subset\mathcal{Z}\\ P_Z(\mathcal{A})\geq 1-\delta}} \max_{z\in\mathcal{A}} \ell(z) \\ &\leq \inf_{\mathcal{J}_r : r\geq r_2} \max_{z\in\mathcal{J}_r} \ell(z) \\ &\leq \inf_{r\geq r_2} r = r_2. \end{aligned}$$

Next, we argue that the strict inequality $r_1 < r_2$ would lead to a contradiction. To see this, suppose $r_1 < r_2$. This means that there exists a set $\mathcal{E} \subset \mathcal{Z}$ with $P_Z(\mathcal{E}) \geq 1-\delta$ such that $r_1 \leq r_3 = \max_{z\in\mathcal{E}} \ell(z) < r_2$. On the other hand, note that

$$\mathcal{E} \subseteq \big\{z : \ell(z) \leq r_3\big\},$$

therefore,

$$P_Z\big\{z : \ell(z) \leq r_3\big\} \geq P_Z(\mathcal{E}) \geq 1-\delta.$$

Hence,

$$r_2 = \inf\{t \geq 0 : P_Z\{z : \ell(z) \leq t\} \geq 1-\delta\} \leq r_3,$$

which is a contradiction. We conclude that $r_1 = r_2$. □

Appendix C
Proof of Theorem 2

It is well-known that the right-continuous quantile function upper bounds the left-continuous quantile function. Nevertheless, we include a proof for completeness. Fix some random variable Z with PML $\ell(Z)$, and $\delta \in (0,1)$. Let $0 \leq \alpha < \underline{\varepsilon}_Z(\delta)$ and consider the set

$$\mathcal{B}_\alpha = \{z \in \mathcal{Z} : \ell(z) \geq \alpha\}.$$

Then, $P_Z(\mathcal{B}_\alpha) > \delta$,[5] and we get

$$\bar{\varepsilon}_Z(\delta) = \max_{\substack{\mathcal{A}\subset\mathcal{Z}:\\ P_Z(\mathcal{A})\geq\delta}} \min_{z\in\mathcal{A}} \ell(z) \geq \min_{z\in\mathcal{B}_\alpha} \ell(z) \geq \alpha.$$

Letting $\alpha \to \underline{\varepsilon}_Z(\delta)$ yields

$$\bar{\varepsilon}_Z(\delta) \geq \underline{\varepsilon}_Z(\delta).$$

Next, we prove the opposite inequality, i.e.,

$$\sup_{Z: X-Y-Z} \underline{\varepsilon}_Z(\delta) \geq \sup_{Z: X-Y-Z} \bar{\varepsilon}_Z(\delta).$$

Fix $\delta \in (0,1)$ and suppose

$$\underline{c} = \underline{\varepsilon}_Y(\delta) < \bar{\varepsilon}_Y(\delta) = \bar{c},$$

where $\bar{c}, \underline{c} > 0$. This happens if and only if there exists a subset $\mathcal{B} \subset \mathcal{Y}$ with probability $P_Y(\mathcal{B}) = \delta$ and $\min_{y\in\mathcal{B}} \ell(y) = \bar{c}$ and $\mathcal{G} = \mathcal{Y}\setminus\mathcal{B}$ and $\max_{y\in\mathcal{G}} \ell(y) = \underline{c}$. Therefore, there exist outcomes $y_1 \in \mathcal{B}, y_2 \in \mathcal{G}$ such that

$$\begin{aligned} \ell(X \to y_1) &\geq \bar{c}, \\ \ell(X \to y_2) &\leq \underline{c}. \end{aligned} \tag{24}$$

Fix a parameter $\eta \in (0,1)$ and let $B \sim \text{Bernoulli}(\eta)$ be independent of (X,Y). Define $Z = h_\eta(Y,B)$ where

$$h_\eta(Y,B) \;=\; \begin{cases} \perp, & \text{if } Y = y_1, \\ \perp, & \text{if } Y = y_2 \text{ and } B = 1, \\ \diamond, & \text{if } Y = y_2 \text{ and } B = 0, \\ Y, & \text{otherwise.} \end{cases}$$

Note that Z is a randomized function of Y since it also depends on B. Now, observe that

$$\begin{aligned} P_Z(\perp) &= P_Y(y_1) + \eta P_Y(y_2), \\ P_{Z|X=x}(\perp) &= P_{Y|X=x}(y_1) + \eta P_{Y|X=x}(y_2). \end{aligned}$$

Then, for fixed x we have

$$\begin{aligned} &\frac{P_{Z|X=x}(\perp)}{P_Z(\perp)} = \frac{P_{Y|X=x}(y_1) + \eta P_{Y|X=x}(y_2)}{P_Y(y_1) + \eta P_Y(y_2)} \\ &= \frac{\frac{P_{Y|X=x}(y_1)}{P_Y(y_1)} + \eta\left(\frac{P_{Y|X=x}(y_2)}{P_Y(y_2)}\right)\left(\frac{P_Y(y_2)}{P_Y(y_1)}\right)}{1 + \eta\frac{P_Y(y_2)}{P_Y(y_1)}} \\ &= \left(\frac{P_{Y|X=x}(y_1)}{P_Y(y_1)} + \eta\left(\frac{P_{Y|X=x}(y_2)}{P_Y(y_2)}\right)\cdot\left(\frac{P_Y(y_2)}{P_Y(y_1)}\right)\right)\cdot \\ &\qquad\qquad \left(1 - \eta\left(\frac{P_Y(y_2)}{P_Y(y_1)}\right) + O(\eta^2)\right) \\ &= \frac{P_{Y|X=x}(y_1)}{P_Y(y_1)} - \eta\left(\frac{P_Y(y_2)}{P_Y(y_1)}\right)\cdot \\ &\qquad \left(\frac{P_{Y|X=x}(y_1)}{P_Y(y_1)} - \frac{P_{Y|X=x}(y_2)}{P_Y(y_2)}\right) + O(\eta^2). \end{aligned}$$

[5] This can be shown by contradiction: If $P_Z(\mathcal{B}_\alpha) \leq \delta$, then the set $\mathcal{B}_\alpha^{\mathsf{c}} = \mathcal{Z}\setminus\mathcal{B}_\alpha$ has probability $P_Z(\mathcal{B}_\alpha^{\mathsf{c}}) \geq 1-\delta$ and also satisfies $\max_{z\in\mathcal{B}_\alpha^{\mathsf{c}}} \ell(z) \leq \alpha < \underline{\varepsilon}_Z(\delta)$. This contradicts the definition of $\underline{\varepsilon}_Z(\delta)$ as the smallest upper bound on the PML of all sets with probability at least $1-\delta$.

Taking the logarithm and maximum over $x \in \mathcal{X}$ on both sides gives

$$\begin{aligned}
\ell_{P_{XZ}}(X \to \perp) &= \log \max_x \frac{P_{Z|X=x}(\perp)}{P_Z(\perp)} \\
&= \log \max_x \Bigg(\frac{P_{Y|X=x}(y_1)}{P_Y(y_1)} - \eta \left(\frac{P_Y(y_2)}{P_Y(y_1)} \right) \cdot \\
&\qquad \left(\frac{P_{Y|X=x}(y_1)}{P_Y(y_1)} - \frac{P_{Y|X=x}(y_2)}{P_Y(y_2)} \right) + O(\eta^2) \Bigg) \\
&\geq \log \Bigg(\max_x \frac{P_{Y|X=x}(y_1)}{P_Y(y_1)} - \eta \left(\frac{P_Y(y_2)}{P_Y(y_1)} \right) \cdot \\
&\qquad \max_{x'} \left(\frac{P_{Y|X=x'}(y_1)}{P_Y(y_1)} - \frac{P_{Y|X=x'}(y_2)}{P_Y(y_2)} \right) + O(\eta^2) \Bigg), \\
&\geq \log \left(e^{\bar{c}} - \eta\beta + O(\eta^2) \right) \\
&= \bar{c} + \log\left(1 - \eta\beta e^{-\bar{c}} + O(\eta^2)\right), \qquad (25)
\end{aligned}$$

where

$$\beta = \left(\frac{P_Y(y_2)}{P_Y(y_1)} \right) \max_{x'} \left(\frac{P_{Y|X=x'}(y_1)}{P_Y(y_1)} - \frac{P_{Y|X=x'}(y_2)}{P_Y(y_2)} \right).$$

Note that $\beta > 0$ because

$$\begin{aligned}
&\max_{x'} \left(\frac{P_{Y|X=x'}(y_1)}{P_Y(y_1)} - \frac{P_{Y|X=x'}(y_2)}{P_Y(y_2)} \right) \\
&\qquad \geq \max_{x'} \frac{P_{Y|X=x'}(y_1)}{P_Y(y_1)} - \max_x \frac{P_{Y|X=x}(y_2)}{P_Y(y_2)} \\
&\qquad \geq \exp(\bar{c}) - \exp(\underline{c}) > 0.
\end{aligned}$$

Now, using the elementary bound $\log(1-t) \geq -t - t^2$ for $0 < t < \frac{1}{2}$ in (25) yields

$$\ell_{P_{XZ}}(X \to \perp) \geq \bar{c} - \eta\beta e^{-\bar{c}} + O(\eta^2) = \bar{c} - \eta\gamma + O(\eta^2),$$

with $\gamma = \beta e^{-\bar{c}}$. Thus, by taking $\eta \to 0$, we can bring $\ell_{P_{XZ}}(X \to \perp)$ arbitrarily close to $\bar{c}$.

The final step is to argue that $\underline{\varepsilon}_Z(\delta) \geq \bar{c}$. To show this, let $\mathcal{B}' = \mathcal{B} \setminus \{y_1\}$ and $\mathcal{G}' = \mathcal{G} \setminus \{y_2\}$ so that the alphabet of Z can be represented by $\mathcal{Z} = \mathcal{B}' \cup \mathcal{G}' \cup \{\perp, \diamond\}$. Let $\mathcal{A}$ be an arbitrary subset of $\mathcal{Z}$ with probability $P_Z(\mathcal{A}) \geq 1 - \delta$. Since $P_Z(\mathcal{G}' \cup \{\diamond\}) = P_Y(\mathcal{G}) - \eta P_Y(y_2) < 1 - \delta$, any such set $\mathcal{A}$ either intersects with $\mathcal{B}'$ or contains $\perp$. If $\mathcal{A}$ contains elements from $\mathcal{B}'$, then

$$\max_{z \in \mathcal{A}} \ell_{P_{XZ}}(X \to z) \geq \min_{z \in \mathcal{B}'} \ell_{P_{XZ}}(X \to z) \geq \bar{c},$$

and if $\perp \in \mathcal{A}$, then

$$\max_{z \in \mathcal{A}} \ell_{P_{XZ}}(X \to z) \geq \ell_{P_{XZ}}(X \to \perp) \geq \bar{c} - \eta\gamma + O(\eta^2),$$

and taking $\eta \to 0$ yields $\max_{z \in \mathcal{A}} \ell_{P_{XZ}}(X \to z) \geq \bar{c}$. Hence, we have proved that

$$\underline{\varepsilon}_Z(\delta) = \min_{\mathcal{A}: P_Z(\mathcal{A}) \geq 1-\delta} \max_{z \in \mathcal{A}} \ell(X \to z) \geq \bar{c} = \bar{\varepsilon}_Y(\delta).$$

This, in turn, implies that

$$\sup_{Z: X - Y - Z} \underline{\varepsilon}_Z(\delta) \geq \sup_{Z: X - Y - Z} \bar{\varepsilon}_Z(\delta),$$

as desired.