An Operational Approach to Information Leakage via Generalized Gain Functions

Gowtham R. Kurri, *Member, IEEE*, Lalitha Sankar, *Senior Member, IEEE*, and Oliver Kosut, *Senior Member, IEEE*

***Abstract*— We introduce a *gain function* viewpoint of information leakage by proposing *maximal g-leakage*, a rich class of operationally meaningful leakage measures that subsumes recently introduced leakage measures — maximal leakage and maximal α-leakage. In maximal g-leakage, the gain of an adversary in guessing an unknown random variable is measured using a gain function applied to the probability of correctly guessing. In particular, maximal g-leakage captures the multiplicative increase, upon observing Y, in the expected gain of an adversary in guessing a randomized function of X, maximized over all such randomized functions. We also consider the scenario where an adversary can make multiple attempts to guess the randomized function of interest. We show that maximal leakage is an upper bound on maximal g-leakage under multiple guesses, for any non-negative gain function g. We obtain a closed-form expression for maximal g-leakage under multiple guesses for a class of concave gain functions. We also study maximal g-leakage measure for a specific class of gain functions related to the α-loss, that interpolates log-loss ($\alpha = 1$) and (soft) 0-1 loss ($\alpha = \infty$). In particular, we first completely characterize the minimal expected α-loss under multiple guesses and analyze how the corresponding leakage measure is affected with the number of guesses. We show that a new measure of divergence that belongs to the class of Bregman divergences captures the relative performance of an arbitrary adversarial strategy with respect to an optimal strategy in minimizing the expected α-loss. Finally, we study two variants of maximal g-leakage depending on the type of adversary and obtain closed-form expressions for them, which do not depend on the particular gain function considered as long as it satisfies some mild regularity conditions. We do this by developing a variational characterization for the Rényi divergence of order infinity which naturally generalizes the definition of pointwise maximal leakage to incorporate arbitrary gain functions.**

Manuscript received 28 September 2022; revised 24 November 2023; accepted 24 November 2023. Date of publication 8 December 2023; date of current version 22 January 2024. This work was supported in part by NSF under Grant CIF-1901243, Grant CIF-1815361, Grant CIF-2007688, Grant CIF-2134256, Grant CIF-2031799, and Grant CIF-2312666. An earlier version of this paper was presented in part at the 2021 International Symposium on Information Theory [DOI: 10.1109/ISIT45174.2021.9517733] and in part at the 2022 International Symposium on Information Theory [DOI: 10.1109/ISIT50566.2022.9834358]. *(Corresponding author: Gowtham R. Kurri.)*

Gowtham R. Kurri was with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85281 USA. He is now with the Signal Processing and Communications Research Centre, International Institute of Information Technology, Hyderabad, Telangana 500032, India (e-mail: gowtham.kurri@iiit.ac.in).

Lalitha Sankar and Oliver Kosut are with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85281 USA (e-mail: lalithasankar@asu.edu; okosut@asu.edu).

Communicated by M. Skoglund, Associate Editor for Security and Privacy.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2023.3341148.

Digital Object Identifier 10.1109/TIT.2023.3341148

***Index Terms*— Privacy leakage, maximal leakage, gain function, Sibson mutual information, Rényi divergence, multiple guesses.**

I. Introduction

A FUNDAMENTAL question in many privacy and secrecy problems is — how much information does a random variable Y that represents observed data by an adversary leak about a correlated random variable X that represents sensitive data? For example, X may be a secret that must be kept confidential and the observation Y could be an inevitable consequence of a system design, for instance, in exchange for certain services. A number of approaches to quantify such information leakage have been proposed in both computer science [1], [2], [3], [4], [5], [6], [7] and information theory [8], [9], [10], [11], [12], [13], [14], [15], [16].

Leakage measures with an associated operational meaning are of specific interest in the literature since an upper bound on such a leakage measure allows the system designer to ensure certain guarantees on the system. Recently, Issa et al. [13] proposed one such leakage measure in the *guessing* framework. In particular, Issa et al. [13] consider an adversary interested in a (possibly randomized) function of X and study the logarithm of the multiplicative increase, upon observing Y, of the probability of correctly guessing a randomized function of X, say, U. Moreover, this quantity is maximized over all the random variables U such that $U - X - Y$ forms a Markov chain capturing the scenario that the function of interest U is unknown to the system designer. The resulting quantity is referred to as *maximal leakage* (MaxL) and an upper bound on it limits the amount of information leakage of any arbitrary randomized function of X through Y. Liao et al. [15] later generalized MaxL to a family of leakages, *maximal α-leakage* (Max-αL), for $\alpha \in (1, \infty)$, that allows tuning the measure to specific applications. In particular, similar to MaxL, Max-αL quantifies the maximal logarithmic increase in a monotonically increasing power function (dependent on α) applied to the probability of correctly guessing. We remark that Max-αL provides an operational interpretation to mutual information (for $\alpha = 1$) in the context of privacy leakage, which was an open problem earlier [13], [17]. Saeidian et al. [16] introduced a variant of maximal leakage, called *pointwise maximal leakage*, capturing the amount of information leaked about X due to disclosing a single outcome $Y = y$ rather than focusing on the *average outcome* as in maximal leakage.

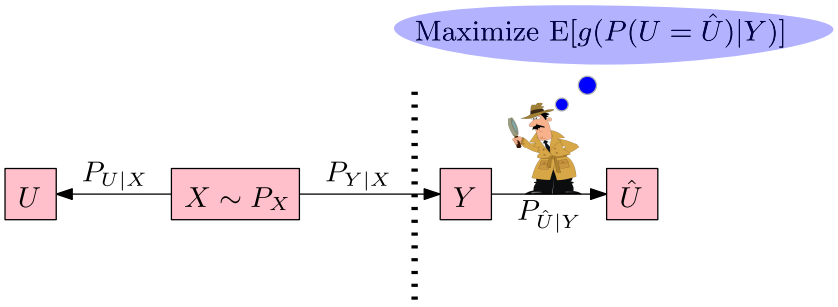

Fig. 1. A gain function viewpoint of leakage. An adversary observes Y and wants to maximize, on an average, the gain function applied to the probability of correctly guessing a randomized function of X, denoted by U.

These operationally motivated leakage measures find applications in many areas such as in privacy-utility trade-offs [15], private information retrieval [18], hypothesis testing [19], source coding [20], membership inference [21], and age of information [22].

We extend the aforementioned line of work by focusing on arbitrary *gain functions* $g : [0,1] \to [0,\infty)$ applied to the probability of correctness (see Fig. 1). In particular, we define *maximal g-leakage* as the maximal logarithmic increase in the gain (applied to the probability of correctness) of an adversary and study its properties for arbitrary g. We also consider *maximal g-leakage under multiple guesses* where the adversary is allowed to make multiple guesses. Further, we introduce and study variants of maximal g-leakage depending on the type of adversary by developing a new variational characterization for Rényi divergence of order ∞ [23]. A variational characterization for a divergence transforms its definition into an optimization problem. Variational characterizations for Rényi divergences of order $\alpha \in \mathbb{R} \setminus \{0\}$ [23] are studied in the literature [24], [25], [26], [27], [28]. We also study information leakage when the adversary is allowed to make multiple attempts to guess the randomized function of interest by focusing on a specific gain function related to the α-loss [15], [29], [30], [31] interpolating log-loss ($\alpha = 1$) and (soft) 0-1 loss ($\alpha = \infty$).

A. Related Work

Most of the approaches in the literature start with a particular metric developed in other contexts and study the properties that follow from the definition. It is difficult to associate an operational meaning to such a leakage measures. In fact, such approaches can label evidently insecure systems as secure (see [17, Section 3.6]). An alternative approach is via a specific threat model of a *guessing* adversary giving an operationally meaningful interpretation to information leakage [2], [3], [5], [13], [15], [16]. Smith [2] defines *min-entropy leakage* as the logarithm of the multiplicative increase, upon observing Y, of the probability of correctly guessing X. Braun et al. [3] consider the maximization of min-entropy leakage over all prior distributions on X resulting in a leakage measure (later shown to be equal to maximal leakage [13]) known as *min-capacity* that is dependent only on the conditional distribution (or channel) $P_{Y|X}$. Alvim et al. [5], [7] defined *g-leakage* and *g-capacity* by introducing a gain function $g : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}$ instead of looking at the probability of correctly guessing X. Alvim et al. [5] showed that maximizing g-capacity over all gain functions g yields maximal leakage. Note that the threat model in all these works focuses on guessing X itself rather than a (potentially) randomized function U of X. Issa et al. [13] considered maximal gain leakage based on the gain functions and g-leakage introduced by Alvim et al. [5], where the adversary is interested in a (potentially) randomized function of X and maximum is taken over all gain functions $g : \mathcal{U} \times \hat{\mathcal{U}} \to [0,1]$. Similar to Alvim et al. [5], they showed that maximal gain leakage is equal to maximal leakage. As we show later, our definition of gain function where a real-valued function is applied to the probability of correctly guessing can be seen as a special case of the gain function definition of Alvim et al. [5], [7]. However, the information leakage measures we study in this work differ substantially from that of the works [5], [7], [13]. In particular, Alvim et al. [5], [7] and Issa et al. [13] consider leakage measures with a maximum taken over all the gain functions while the problem of computing leakage measures with specific gain functions (other than the identity gain function that corresponds to maximal leakage) remains open and is conjectured to be challenging [7, Section VI.A]. In this work, we study leakage measures with specific gain functions and obtain their closed-form expressions. We present more details on the distinction from Alvim et al. [5], [7] and Issa et al. [13] in Section III (in particular, see Remark 3).

Instead of multiplicative increase, the notion of additive increase has also been considered in quantifying privacy leakage. For example, an additive version of g-leakage is studied by Alvim et al. [7]. Also, the notion of *semantic security* in cryptography defines 'advantage' as the additive increase, upon observing the encrypted message, of the probability of correctly guessing the value of the function. Calmon et al. [32] and Li and El Gamal [33] use maximal correlation as a measure of information leakage. Mutual information has been used as a privacy measure in many works, see e.g., [9], [12], [34], [35], [36], [37], [38], and [39]. Asoodeh et al. [40] use the probability of correctly guessing as a privacy measure. Li et al. [41] use hypothesis test performance of an adversary to measure leakage. Total variation distance is used as privacy measure by Rassouli and Gündüz [14].

Another line of work in privacy leakage is based on indistinguishability, i.e., whether an adversary can distinguish between two items of interest. Differential privacy [1], proposed in the context of querying databases, ensures that all databases that differ only in one entry produce an output to a query with *almost* equal probabilities. Similar to differential privacy and pointwise maximal leakage, Calmon and Fawaz [9], Issa et al. [13], and Jiang et al. [42] study privacy leakage providing worst-case guarantees, note that Issa et al. [13] study an average case leakage measure (maximal leakage) also. An equivalent notion of differential privacy using (conditional) mutual information is studied in [43]. The notion of differential privacy is known to be very strict and has limited applicability [44], [45]. Approximate differential privacy [46] and Rényi differential privacy [47] are proposed as relaxations of differential privacy to allow data releases with higher utility. For an extensive list of leakage measures see the surveys by Wagner and Eckhoff [48], Bloch et al. [49], and Hsu et al. [50].

B. Main Contributions

The main contributions of this paper are as follows:

- We show that maximal leakage is an upper bound on maximal g-leakage under $k \geq 1$ guesses for any arbitrary non-negative gain function g (Proposition 1). We obtain closed-form expression for maximal g-leakage under $k \geq 1$ guesses for a class of gain functions g. Specifically, when g is a non-negative concave function with a finite non-zero derivative at 0, we show that maximal g-leakage under $k \geq 1$ guesses is equal to the Sibson mutual information of order infinity [51], i.e., the maximal leakage (Theorem 1). In addition, we obtain a closed-form expression for maximal g-leakage when $g(t) = 1 + t$, for binary X and Y (Theorem 2). An interesting aspect of this expression is that it is fundamentally different in structure from that of maximal leakage and maximal α-leakage.
- We then focus on maximal g-leakage under multiple guesses for a specific class of gain functions related to the α-loss, $\alpha \in (0, \infty)$ [15], [29], [30], [31], which interpolates log-loss ($\alpha = 1$) and 0-1 loss ($\alpha = \infty$). We define two leakage measures, the α-leakage and the maximal α-leakage under multiple guesses. We show that α-leakage does not change with the number of guesses for a class of probability distributions (Theorem 4) [52]. We prove that maximal α-leakage under multiple guesses is at least that of with a single guess (Theorem 5).
- To prove these, we completely characterize the minimal expected α-loss under k guesses (Theorem 3) [52], thereby recovering the known results for $\alpha = \infty$ [13]. We illustrate an optimal guessing strategy of the adversary through Examples 1 and 2. To the best of our knowledge, such a result even for log-loss under multiples guesses was not explored earlier. We derive a technique for transforming the optimization problem over probability simplex associated with multiple random variables to that of with a single random variable using tools drawn from duality in linear programming, which may be of independent interest.
- We introduce and study two variants of maximal g-leakage, namely, opportunistic maximal g-leakage (when the adversary could choose the function of interest depending on the realization of Y) and maximal realizable g-leakage (when the adversary is interested in *maximum* guessing performance over all the realizations of Y instead of *average* performance). We obtain closed-form expressions for these leakage measures in terms of Sibson mutual information of order ∞ and Rényi divergence of order ∞, respectively (Corollary 4).
- We do this by devising a new variational characterization for Rényi divergence of order ∞ expressed in terms of the ratio of maximal expected gains in guessing a randomized function of X, which may be of independent interest (Theorem 6) [53]. An important aspect of our characterization is that it remains agnostic to the particular gain function considered as long as it satisfies some mild regularity conditions. Even though the variational characterizations mentioned earlier are presented for any finite order α, one can obtain such characterizations for ∞-Rényi divergence by applying a limiting argument (see the discussion above Proposition 2). Our characterization differs from those characterizations in view of its connection to guessing, and more importantly because of its robustness to the gain function. We also show that our variational characterization naturally extends the notion of pointwise maximal leakage [16] to incorporate arbitrary gain functions under mild regularity assumptions retaining the same closed-form expression (Corollary 5).

C. Organization of the Paper

The remainder of this paper is organized as follows. We introduce the gain function viewpoint and review maximal leakage and maximal α-leakage in Section II. In Section III, we present our results on maximal g-leakage under multiple guesses. In Section IV, we present our results on maximal g-leakage under multiple guesses by focusing on a class of specific gain functions related to the α-loss [29], [30], [31], parameterized by $\alpha \in (0, 1) \cup (1, \infty)$. In Section V, we develop a variational characterization for ∞-Rényi divergence and show how it can be employed to obtain closed-form expressions for variants of maximal g-leakage depending on the type of adversary. The proofs of all the results are presented in Appendix.

II. Preliminaries

Notation. We use capital letters to denote random variables, e.g., X, and capital calligraphic letters to denote their corresponding alphabet, e.g., $\mathcal{X}$. We consider only finite alphabets in this work. We use $\mathbb{E}_X[\cdot]$ to denote expectation with respect to P_X and $U - X - Y$ to denote that the random variables form a Markov chain. We use $\text{supp}(X) := \{x : P_X(x) > 0\}$ to denote the support set of X. We use $H(X)$, $I(X;Y)$, and $D(P_X \| Q_X)$ to denote entropy, mutual information, and relative entropy, respectively. Given two probability distributions

P_X and Q_X over an alphabet $\mathcal{X}$, we write $P_X \ll Q_X$ to denote that P_X is absolutely continuous with respect to Q_X. Finally, we use $\log$ to denote the natural logarithm.

We begin by defining the maximal expected gain of an adversary in guessing an unknown random variable, which is an impetus for this work. A *gain function* is defined as a function $g : [0,1] \to [0,\infty)$, and can be interpreted as a function that is applied to the probability of correctly guessing an unknown random variable.

Definition 1 (Maximal Expected Gain): Given a gain function $g : [0,1] \to \mathbb{R}$ and a probability distribution P_X on a finite alphabet $\mathcal{X}$, the maximal expected gain is defined as

$$\sup_{P_{\hat{X}}} \mathbb{E}_X \left[g(P_{\hat{X}}(X)) \right], \tag{1}$$

where $\hat{X}$ represents an estimator of X with same support as X.

Though maximal expected gain in (1) is defined for real-valued gain functions, we restrict our attention to non-negative gain functions for most of the time in this paper. However, some of our results hold for negative gain functions also, e.g., $g(t) = \log t$, $t \in [0,1]$ (we discuss this in Remark 7 in Section V). The objective function in (1) is the expected value of gain function g applied to the probability of correctly guessing an unknown random variable X. Alvim et al. [5], [7] consider a similar gain function viewpoint of information leakage with gain functions $g : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}$. It is worth mentioning that the gain function of Alvim et al. [5], [7] subsumes the gain function used in (1). In particular, if we consider the alphabet $\hat{\mathcal{X}}$ to be equal to the simplex of probability distributions $P_{\hat{X}}$ on $\mathcal{X}$, we can define a gain function $g(x, P_{\hat{X}}) = g'(P_{\hat{X}}(x))$ for a function $g' : [0,1] \to \mathbb{R}$. However, our gain function viewpoint is conceptually different from that of Alvim et al. [5], [7] in that the maximal expected gain in (1) is expressed in terms of the probability of correctly guessing. Moreover, as mentioned earlier in Section I-A, our approach to study information leakage measures differs substantially from that of Alvim et al. [5], [7] (see Section III for more details). We note that the notion of maximal expected gain in (1) for specific gain functions, $g(t) = t$ and $g(t) = \frac{\alpha}{\alpha-1} t^{\frac{\alpha-1}{\alpha}}$, $\alpha \in (1,\infty]$, plays a crucial role in the definitions of maximal leakage [13] and maximal α-leakage [15] (see the bulleted list near (26) for optimal guessing strategies for these gain functions in addition to the logarithmic gain function $g(t) = \log t$).

Definition 2 (Maximal α-Leakage [15], [30]): Given a joint distribution P_{XY} on a finite alphabet $\mathcal{X} \times \mathcal{Y}$, the maximal α-leakage from X to Y is defined as, for $\alpha \in (0,1) \cup (1,\infty)$,

$$\mathcal{L}_\alpha^{\max}(X \to Y) = \sup_{U - X - Y} \frac{\alpha}{\alpha - 1} \log \frac{\max_{P_{\hat{U}|Y}} \mathbb{E}_{UY} \left[\frac{\alpha}{\alpha-1} P_{\hat{U}|Y}(U|Y)^{\frac{\alpha-1}{\alpha}} \right]}{\max_{P_{\hat{U}}} \mathbb{E}_U \left[\frac{\alpha}{\alpha-1} P_{\hat{U}}(U)^{\frac{\alpha-1}{\alpha}} \right]}, \tag{2}$$

where U represents any randomized function of X that the adversary is interested in guessing and takes values in an arbitrary finite alphabet. Moreover, $\hat{U}$ is an estimator of U with the same support as U.

Maximal α-leakage captures the information leaked about *any function* of the random variable X to an adversary that observes a correlated random variable Y. Maximal α-leakage is proposed as a generalization of maximal leakage [13], where the former recovers the latter when $\alpha \to \infty$, to allow tuning the measure to specific applications. The ratio inside the logarithm in (2) is the multiplicative increase, upon observing Y, of the maximal expected gain of an adversary in guessing a randomized function of X, with gain function $g(t) = \frac{\alpha}{\alpha-1} t^{\frac{\alpha-1}{\alpha}}$, $\alpha \in (0,1) \cup (1,\infty)$. In this work, we study maximal expected gain (1) and the associated leakage measures for arbitrary gain functions $g : [0,1] \to (0,\infty)$. Maximal leakage, maximal α-leakage, and the new leakage measures we study in this work can be expressed in terms of Sibson mutual information [51] and Rényi divergence [23].

Definition 3 (Sibson Mutual Information of Order α [51]): For a given joint distribution P_{XY} on finite alphabet $\mathcal{X} \times \mathcal{Y}$, the Sibson mutual information of order $\alpha \in (0,1) \cup (1,\infty)$ is

$$I_\alpha^S(X;Y) = \frac{\alpha}{\alpha - 1} \log \sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{X}} P_X(x) P_{Y|X}(y|x)^\alpha \right)^{\frac{1}{\alpha}}.$$

It is defined by its continuous extension for $\alpha = 1$ and $\alpha = \infty$, respectively, and is given by

$$I_1^S(X;Y) = I(X;Y) \text{ (Shannon mutual information)}, \tag{3}$$

$$I_\infty^S(X;Y) = \log \sum_{y \in \mathcal{Y}} \max_{x : P_X(x) > 0} P_{Y|X}(y|x). \tag{4}$$

Definition 4 (Rényi Divergence of Order α [23]): The Rényi divergence of order $\alpha \in (0,1) \cup (1,\infty)$ between two probability distributions P_X and Q_X on a finite alphabet $\mathcal{X}$ is defined as

$$D_\alpha(P_X || Q_X) = \frac{1}{\alpha - 1} \log \left(\sum_{x \in \mathcal{X}} P_X(x)^\alpha Q_X(x)^{1-\alpha} \right). \tag{5}$$

It is defined by its continuous extension for $\alpha = 1$ and $\alpha = \infty$, respectively, and is given by

$$D_1(P_X || Q_X) = \sum_{x \in \mathcal{X}} P_X(x) \log \frac{P_X(x)}{Q_X(x)}, \tag{6}$$

$$D_\infty(P_X || Q_X) = \max_{x \in \mathcal{X}} \log \frac{P_X(x)}{Q_X(x)}. \tag{7}$$

Notice that for $\alpha \geq 1$, $D_\alpha(P_X \| Q_X)$ is finite if and only if P_X is absolutely continuous with respect to Q_X. Issa et al. [13] showed that maximal leakage is equal to Sibson mutual information of order ∞, i.e.,

$$\mathcal{L}_\infty^{\max}(X \to Y) = I_\infty^{\mathrm{S}}(X;Y) \tag{8}$$

$$= \log \sum_{y \in \mathcal{Y}} \max_{x : P_X(x) > 0} P_{Y|X}(y|x). \tag{9}$$

Generalizing this, Liao et al. [15] proved that maximal α-leakage is given by the Sibson mutual information of order α, i.e.,

$$\mathcal{L}_\alpha^{\max}(X \to Y) = \sup_{P_{\tilde{X}}} I_\alpha^{\mathrm{S}}(\tilde{X};Y), \tag{10}$$

where the supremum is over all probability distributions $P_{\tilde{X}}$ on the support of P_X.

III. Maximal g-Leakage Under Multiple Guesses

The definitions of maximal leakage and maximal α-leakage consider the adversaries interested in maximizing the expected values of specific gain functions, in particular, maximal leakage uses the gain function $g(t) = t$ and maximal α-leakage uses the gain function $g(t) = \frac{\alpha}{\alpha-1}t^{\frac{\alpha-1}{\alpha}}$, $\alpha \in (1, \infty)$. These leakage measures can be seen as special cases of a more general leakage measure incorporating an adversary interested in maximizing an arbitrary gain function g.

Definition 5 (Maximal g-Leakage): Given a gain function $g : [0, 1] \to [0, \infty)$ and a joint probability distribution P_{XY} on a finite alphabet $\mathcal{X} \times \mathcal{Y}$, the maximal g-leakage is defined as

$$\mathcal{L}_g^{\max}(X \to Y) = \sup_{U:U-X-Y} \log \frac{\sup_{P_{\hat{U}|Y}} \mathbb{E}_{UY}\left[g(P_{\hat{U}|Y}(U|Y))\right]}{\sup_{P_{\hat{U}}} \mathbb{E}_U\left[g(P_{\hat{U}}(U))\right]}. \tag{11}$$

Maximal g-leakage captures how much information an adversary can learn about any randomized function of a random variable X from a correlated random variable Y when a single guess is allowed. We now define *maximal g-leakage under k guesses* which captures the information an adversary can learn when k guesses are allowed. This definition is also related to maximal leakage under k guesses [13, Definition 4].

Definition 6 (Maximal g-Leakage Under k Guesses): Given a gain function $g : [0, 1] \to [0, \infty)$ and a joint probability distribution P_{XY}, the maximal g-leakage from X to Y under k guesses is defined as

$$\mathcal{L}_g^{(k)-\max}(X \to Y) = \sup_{U:U-X-Y} \log \frac{\max_{P_{\hat{U}_{[1:k]}|Y}} \mathbb{E}\left[g\left(\mathrm{P}\left(\bigcup_{i=1}^{k}(\hat{U}_i = U)|Y\right)\right)\right]}{\max_{P_{\hat{U}_{[1:k]}}} \mathbb{E}\left[g\left(\mathrm{P}\left(\bigcup_{i=1}^{k}(\hat{U}_i = U)\right)\right)\right]}, \tag{12}$$

where $\hat{U}_1, \hat{U}_2, \ldots, \hat{U}_k$ represent k estimators of U with the same support as U.

We interpret $\mathrm{P}(\bigcup_{i=1}^{k}(\hat{U}_i = u)|Y = y)$ as the probability of correctly estimating $U = u$ given $Y = y$ in k guesses. Due to the fact that gain function used in maximal g-leakage is a special case of the gain function of Alvim et al. [5] (as discussed in the paragraph after Definition 1), an upper bound on maximal g-leakage directly follows from Issa et al. [13, Theorem 5]. In particular, Issa et al. [13, Theorem 5] showed that, for a given joint distribution P_{XY},

$$\sup_{\substack{U:U-X-Y \\ \hat{\mathcal{U}}, g:\mathcal{U}\times\hat{\mathcal{U}}\to[0,\infty): \\ \sup_{\hat{u}\in\hat{\mathcal{U}}}\mathbb{E}_U[g(U,\hat{u})]>0}} \log \frac{\sup_{\hat{u}(\cdot)} \mathbb{E}_{UY}[g(U, \hat{u}(Y))]}{\sup_{\hat{u}\in\hat{\mathcal{U}}} \mathbb{E}_U[g(U, \hat{u})]} = I_\infty^{\mathrm{S}}(X;Y). \tag{13}$$

Thus, it follows from from (13) and Issa et al. [13, Theorem 1] that

$$\mathcal{L}_g^{\max}(X \to Y) \le I_\infty^{\mathrm{S}}(X;Y) \tag{14}$$

with equality if g is the identity gain function, $g(t) = t$, for $t \in [0, 1]$.

In the following proposition, we show that the upper bound in (14) holds for maximal g-leakage under k guesses also.

Proposition 1 (Upper Bound on Maximal g-Leakage Under k Guesses): Given a gain function $g : [0, 1] \to [0, \infty)$ and a joint probability distribution P_{XY}, we have that maximal leakage is an upper bound on maximal g-leakage under k guesses, i.e.,

$$\mathcal{L}_g^{(k)-\max}(X \to Y) \le I_\infty^{S}(X;Y), \tag{15}$$

with equality if g is the identity gain function, $g(t) = t$, for $t \in [0, 1]$

Remark 1: Proposition 1 follows from (13) together with an observation that the gain function definition used by Alvim et al. [5] (and Issa et al. [13, Theorem 5]) captures not only our gain function definition but also the multiple guesses scenario [7, Section III-C] simultaneously. In particular, let $\hat{\mathcal{U}}$ be the simplex of all probability distributions $P_{\hat{U}_{[1:k]}}$ with each U_i taking values in $\mathcal{U}$ and define

$$g'(u, P_{\hat{U}_{[1:k]}}) := g\left(\mathrm{P}\left(\bigcup_{i=1}^{k}(\hat{U}_i = u)\right)\right), \tag{16}$$

for $u \in \mathcal{U}$, $P_{\hat{U}_{[1:k]}} \in \hat{\mathcal{U}}$. Then, Proposition 1 follows from (13) together with (16). We remark that this observation also provides a simpler proof to the upper bound part in [13, Proof of Theorem 4], i.e., maximal leakage upper bounds maximal leakage under k guesses.

Proposition 1 shows that maximal g-leakage under multiple guesses is maximized when $g(t) = t$, for $t \in [0, 1]$, which corresponds to maximal leakage under multiple guesses [13, Theorem 4]. Also, we note that closed-form expressions for even maximal g-leakage are known only for few gain functions, in particular, when $g(t) = \frac{\alpha}{\alpha-1}t^{\frac{\alpha-1}{\alpha}}$, $\alpha \in (0, 1)(1, \infty)$ [13], [15], [30], that corresponds to maximal α-leakage (and maximal leakage when $\alpha \to \infty$). In the following theorem, we obtain closed-form expression for maximal g-leakage under multiple guesses for a class of concave gain functions.

Theorem 1 (Maximal g-Leakage Under Multiple Guesses): Let P_{XY} be a joint probability distribution on a finite alphabet $\mathcal{X} \times \mathcal{Y}$ and $g : [0, 1] \to [0, \infty)$ be a gain function satisfying the following assumptions:

- g is a concave function,
- $g(0) = 0$ and $0 < g'(0) < \infty$.

Then we have that maximal g-leakage under $k \ge 1$ guesses is exactly equal to maximal leakage, i.e.,

$$\mathcal{L}_g^{(k)-\max}(X \to Y) = I_\infty^{S}(X;Y).$$

Remark 2: An important consequence of Theorem 1 in the design of privacy mechanisms is the following. Suppose the inferential capability of an adversary in guessing about X from Y is measured via a function g of the probability of correctly guessing and we use maximal g-leakage as the privacy measure. The system designer needs to find an optimal

privacy mechanism $P_{Y|X}$ minimizing the leakage $\mathcal{L}_g(X \to Y)$ while maintaining a minimum level of utility measured by, say, $\mathcal{U}(X,Y)$. In many practical situations, we may have only a limited knowledge about the inferential capability of adversary. Specifically, assume that all we know about the function g is that it belongs to a class of non-negative concave gain functions g that satisfy $g(0) = 0$ with a finite positive derivative at 0 (see the paragraph below Remark 4 for an interpretation of this class). Now, as a result of Theorem 1, it suffices for the system designer to find the optimal mechanism that minimizes the Sibson mutual information of order infinity subject to utility constraints without worrying about further details of g, i.e.,

$$\inf_{P_{Y|X}:\mathcal{U}(X,Y)\geq k} I^S_\infty(X;Y). \tag{17}$$

Thus, notice that the effective situation is same as that of knowing the exact form of the gain function g as the optimal privacy mechanism with a particular g remains the same as that of (17) as long as g belongs to the aforementioned class.

Remark 3: In this remark, using Theorem 1 we outline the distinction between maximal g-leakage and the leakage measures of Alvim et al. [5] and Issa et al. [13] based on gain functions. Alvim et al. [5, Definition 3.4] define g-capacity of a channel $P_{Y|X}$ as

$$\mathcal{ML}_g(X \to Y) := \sup_{P_X} \log \frac{\sup_{\hat{x}(\cdot)} \mathbb{E}_{XY}[g(X,\hat{x}(Y))]}{\sup_{\hat{x}\in\hat{\mathcal{X}}} \mathbb{E}_X[g(X,\hat{x})]}, \tag{18}$$

for any $g : \mathcal{X}\times\hat{\mathcal{X}} \to [0,\infty)$, and showed that [5, Theorem 5.1]

$$\sup_{\substack{\hat{\mathcal{X}},g:\mathcal{X}\times\hat{\mathcal{X}}\to[0,\infty):\\ \sup_{\hat{x}\in\hat{\mathcal{X}}}\mathbb{E}_X[g(X,\hat{x})]>0}} \mathcal{ML}_g(X \to Y) = \log \sum_{y\in\mathcal{Y}} \max_{x\in\mathcal{X}} P_{Y|X}(y|x), \tag{19}$$

which considers the worst-case scenario over all g (note that the expression in RHS of (19) is equal to that of (13) if X has a full support). Thus, Alvim et al. [5] and Issa et al. [13] (see (13)) obtained conclusive results for leakage measures with a maximum taken over all the gain functions rather than focusing on leakage measures with specific gain functions. Theorem 1 provides closed-form expression for maximal g-leakage for a specific class of concave gain functions.

Remark 4: Note that the gain function that corresponds to maximal α-leakage (for $\alpha \in (0,1)\cup(1,\infty)$), i.e., $g(t) = \frac{\alpha}{\alpha-1}t^{\frac{\alpha-1}{\alpha}}$, does not belong to the class of gain functions for which Theorem 1 holds even though the gain function is concave; this is because $g'(0) = \infty$. However, as $\alpha \to \infty$, we have $g(t) = t$ for which $g'(0) = 1 < \infty$. So, Theorem 1 recovers [13, Theorem 1] when $g(t) = t$. Some other examples of gain functions for which Theorem 1 holds are $g(t) = 1-(1-t)^2, \sin t, \min\{ct, \frac{1}{2}\}$, for some constant $c > 0$.

To interpret the conditions on g in Theorem 1, we first note that the concavity of g is natural in that we are examining the optimization problems involving maximization of gain functions in the definition of maximal g-leakage. The condition $g(0) = 0$ suggests that the minimum possible value of the gain function be assigned to adversary when the probability of correctly guessing is zero. When the minimum possible value has been assigned to $g(0)$, it has to be the case that g is increasing at 0, in fact, we need g to be strictly increasing at 0 with a finite derivative in Theorem 1. Extending this intuition, we may impose an additional constraint that the maximum possible value of the gain function be assigned when the probability of correctly guessing is 1, i.e., $g(1) = \sup_{t\in[0,1]} g(t)$ (though it is not required for Theorem 1). Then because of concavity, the function g needs to be non-decreasing in the probability of correctly guessing,[1] which may be more relevant in practical scenarios. A detailed proof of Theorem 1 is in Appendix A-A. A consequence of Theorem 1 is that the maximum in (15) is achieved by a class of concave gain functions with identity gain function (that corresponds to maximal leakage) being one of them.

Notice that maximal g-leakage in Theorem 1 and maximal α-leakage depends on P_X only through its support. The following theorem presents a closed-form expression for maximal g-leakage that corresponds to the gain function $g(t) = 1 + t$, $t \in [0,1]$, for binary X and Y. Interestingly, it turns out that this expression is fundamentally different in its structure from that of maximal leakage and maximal α-leakage. In particular, this leakage explicitly depends on P_X (i.e., not just through its support).

Theorem 2: Let P_{XY} be a joint probability distribution on $\mathcal{X}\times\mathcal{Y}$ with $|\mathcal{X}| = |\mathcal{Y}| = 2$. Then, for $g(t) = 1+t$, $t \in [0,1]$, we have

$$\mathcal{L}_g^{\max}(X \to Y) = \log \frac{1 + p^* \sum_{y\in\mathcal{Y}} \max_{x\in\mathcal{X}} P_{Y|X}(y|x)}{1+p^*}, \tag{20}$$

where $p^* = \min_{x:P_X(x)>0} P_X(x)$.
A detailed proof of Theorem 2 is in Appendix A-B.

IV. Evaluating Multiple Guesses via a Tunable Loss Function

In this section, we focus on a specific class of gain functions, related to the α-loss [29], [30], [31], parameterized by $\alpha \in (0,1)\cup(1,\infty)$ and the corresponding maximal g-leakage measures under multiple guesses. We first characterize the minimal expected α-loss (or equivalently, the associated maximal expected gain) under multiples guesses, and then study maximal g-leakage for the corresponding class of gain functions parameterized by $\alpha \in (0,1)\cup(1,\infty)$.

Consider a setup where an adversary is interested in guessing the unknown value of a random variable X on observing another correlated random variable Y, where X and Y are jointly distributed according to P_{XY} over the finite support $\mathcal{X}\times\mathcal{Y}$. The adversary can make a fixed number of guesses, say k, to estimate X. We focus on evaluating the adversary's success using loss functions that in turn can measure the information leaked by Y about X. To this end, we model the adversary's strategy using α-loss, a class of tunable loss functions parameterized by $\alpha \in (0,\infty]$ [30], [31]. This class

[1] To see this, suppose that there exist x, y such that $x < y$ and $g(x) > g(y)$. Since $y \in [x,1]$, there exists $\lambda \in [0,1]$ with $y = \lambda x + (1-\lambda)$. Then, concavity of g implies that $g(y) = g(\lambda x + (1-\lambda) \geq \lambda g(x) + (1-\lambda)g(1) > \lambda g(y) + (1-\lambda)g(1) \geq \lambda g(y) + (1-\lambda)g(y) = g(y)$, which is a contradiction.

captures the well-known exponential loss ($\alpha = 1/2$) [54], log-loss ($\alpha = 1$) [55], [56], [57], and the 0-1 loss ($\alpha = \infty$) [56], [58]. The adversary then seeks to find the optimal (possibly randomized) guessing strategy that minimizes the expected α-loss over k guesses. We first review α-loss and then define maximal expected α-loss under k guesses.

Definition 7 (α-Loss [29], [30], [31]): For $\alpha \in (0,1) \cup (1,\infty)$, the α-loss is a function defined from $[0,1]$ to $\mathbb{R}_+$ as

$$\ell_\alpha(p) := \frac{\alpha}{\alpha-1}\left(1 - p^{\frac{\alpha-1}{\alpha}}\right). \tag{21}$$

It is defined by continuous extension for $\alpha = 1$ and $\alpha = \infty$, respectively, and is given by

$$\ell_1(p) = \log\frac{1}{p},\ \ell_\infty(p) = 1 - p. \tag{22}$$

Notice that $\ell_\alpha(p)$ is decreasing in p.

Definition 8 (Minimal Expected α-Loss Under k Guesses): Consider random variables $(X,Y) \sim P_{XY}$ and an adversary that makes k guesses $\hat{X}_{[1:k]} = \hat{X}_1, \hat{X}_2, \ldots, \hat{X}_k$ on observing Y such that $X - Y - \hat{X}_{[1:k]}$ is a Markov chain. Let $P_{\hat{X}_{[1:k]}|Y}$ be a strategy for estimating X from Y in k guesses. For $\alpha \in (0,\infty]$, the minimal expected α-loss under k guesses is defined as

$$\mathcal{ME}_\alpha^{(k)}(P_{XY}) := \min_{P_{\hat{X}_{[1:k]}|Y}} \sum_{x,y} P_{XY}(x,y)\ell_\alpha\left(\mathrm{P}\left(\bigcup_{i=1}^{k}(\hat{X}_i = x)|Y = y\right)\right). \tag{23}$$

$\mathrm{P}(\bigcup_{i=1}^{k}(\hat{X}_i = x)|Y = y)$ is the probability of correctly estimating $X = x$ given $Y = y$ in k guesses. An adversary seeks to find the optimal guessing strategy in (23). Note that the optimization problem in (23) was solved for a special case of $k = 1$ by Liao et al. [30, Lemma 1]. Notice that

$$\mathcal{ME}_\alpha^{(k)}(P_{XY}) = \sum_y P_Y(y)\mathcal{ME}_\alpha^{(k)}(P_{X|Y=y}), \tag{24}$$

where we have slightly abused the notation in the R.H.S. of (24). Hence, in view of (24), in order to solve the optimization problem in (23), it suffices to solve for a case where $Y = \emptyset$, i.e.,

$$\mathcal{ME}_\alpha^{(k)}(P_X) := \min_{P_{\hat{X}_{[1:k]}}} \sum_x P_X(x)\ell_\alpha\left(P\left(\bigcup_{i=1}^{k}(\hat{X}_i = x)\right)\right). \tag{25}$$

Also, in the sequel, it suffices to consider the optimization problem in (25) only for the case where $k < n$, where P_X is supported on $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ because if $k \geq n$, we have $\mathcal{ME}_\alpha^{(k)}(P_X) = 0$, since a strategy $P^*_{\hat{X}_{[1:k]}}$ such that $P^*_{\hat{X}_{[1:n]}}(x_1, x_2, \ldots, x_n) = 1$ is optimal. We review some simple special cases of (25) that are well known in the literature.

- Consider (25) for the case of log-loss ($\alpha = 1$) and $k = 1$ [57]. It is well known that the expected log-loss can be expanded as
$$\mathbb{E}\left[\log\frac{1}{P_{\hat{X}}(X)}\right] = H(X) + D(P_X\|P_{\hat{X}}), \tag{26}$$
which implies that the optimal guessing strategy is equal to the original distribution, P_X, and the minimal expected log-loss is given by its entropy, $H(X)$. Moreover, the relative entropy $D(P_X\|P_{\hat{X}})$ quantifies the relative performance of an arbitrary adversarial guessing strategy $P_{\hat{X}}$ with respect to the optimal guessing strategy P_X. To the best of our knowledge, minimal expected loss under multiple guesses was not explored even for log-loss ($\alpha = 1$) earlier.
- At the other extreme, Issa et al. [13] studied maximal expected probability of correctness in a fixed number of guesses which corresponds to minimal expected α-loss for the special case of $\alpha = \infty$. The optimal guessing strategy here is to guess the k most likely outcomes according to the distribution P_X.
- Liao et al. [15, Lemma 1] solved the optimization problem in (25) for a special case of $k = 1$ and for arbitrary $\alpha \in (0,\infty]$, where they showed that the optimal guessing strategy is a tilted probability distribution given by $P_X^{(\alpha)}(x) := \frac{P_X(x)^\alpha}{\sum_x P_X(x)^\alpha}$.

We completely characterize the minimal expected α-loss under k guesses, for $\alpha \in (0,\infty]$. We first express the objective function in (25), i.e., the expected α-loss, in a way similar to (26) where it turns out that Bregman divergence [59], a generalization of relative entropy, naturally arises. For a continuously-differentiable convex function $F : \Omega \to \mathbb{R}$, the associated Bregman divergence between p and q in Ω is defined as $B_F(p,q) = F(p) - F(q) - \langle \nabla F(q), p - q\rangle$.

Lemma 1 (Expected α-Loss Under k Guesses): For a fixed probability distribution P_X and an arbitrary guessing strategy $P_{\hat{X}_{[1:k]}}$, we have

$$\mathbb{E}_X\left[\ell_\alpha(\mathrm{P}(\cup_{i=1}^{k}(\hat{X}_i = X)))\right] = \frac{\alpha}{\alpha-1}\left(1 - k^{\frac{\alpha-1}{\alpha}}\mathrm{e}^{\frac{1-\alpha}{\alpha}H_\alpha(X)}\right) + k^{\frac{\alpha-1}{\alpha}}B_F(P_X, P_{\tilde{X}}^{(\frac{1}{\alpha})}), \tag{27}$$

where $P_{\tilde{X}}(x) := \frac{\mathrm{P}(\cup_{i=1}^{k}(\hat{X}_i = x))}{k}$, $P_{\tilde{X}}^{(\frac{1}{\alpha})}(x) := \frac{P_{\tilde{X}}(x)^{\frac{1}{\alpha}}}{\sum_x P_{\tilde{X}}(x)^{\frac{1}{\alpha}}}$, and $B_F(\cdot,\cdot)$ is the Bregman divergence associated with the function $F(P_X) = \frac{\alpha}{\alpha-1}\left((\sum_x P_X(x)^\alpha)^{\frac{1}{\alpha}} - 1\right)$ given by $B_F(P_X, P_{\tilde{X}}^{(\frac{1}{\alpha})}) = k^{\frac{\alpha-1}{\alpha}}(\sum_x P_X(x)^\alpha)^{\frac{1}{\alpha}}\left(1 - \mathrm{e}^{\frac{1-\alpha}{\alpha}D_{\frac{1}{\alpha}}(P_X^{(\alpha)}\|P_{\tilde{X}})}\right)$. Moreover, the minimal expected α-loss is given by $\mathcal{ME}_\alpha^{(k)}(P_X) = \frac{\alpha}{\alpha-1}(1 - k^{\frac{\alpha-1}{\alpha}}\mathrm{e}^{\frac{1-\alpha}{\alpha}H_\alpha(X)})$ if and only if $P_X^{(\alpha)}(x) \leq \frac{1}{k}$, for all $x \in \mathcal{X}$.

A detailed proof of Lemma 1 is in Appendix B-A. Note that $P_{\tilde{X}}(x)$ defined in Lemma 1 may not be even a probability distribution from the way it is defined since $\sum_x \mathrm{P}(\cup_{i=1}^{k}(\hat{X}_i = x)) \leq k$, in general. However, as we prove later (see Lemma 5 and Remark 9), it suffices to consider the guessing strategies

$P_{\hat{X}_{[1:k]}}$ that satisfy the condition $\sum_x \mathrm{P}(\cup_{i=1}^k(\hat{X}_i = x)) = k$ in order to solve the optimization problem in (25), thereby making it sufficient to limit $P_{\hat{X}}$ to be a probability distribution. Note that Lemma 1 characterizes the minimal expected α-loss under k guesses only for the case when $P_X^{(\alpha)}(x) \leq \frac{1}{k}$, for all $x \in \mathcal{X}$. This condition is trivially satisfied by any P_X when $k = 1$. Thus, when $P_X^{(\alpha)}(x) \leq \frac{1}{k}$, for all $x \in \mathcal{X}$, Bregman divergence B_F (with the function F in Lemma 1) captures the relative performance of an arbitrary guessing strategy with respect to an optimal strategy in minimizing the expected α-loss. This generalizes the observation below (26) that relative entropy quantifies the relative performance of an arbitrary adversarial guessing strategy with respect to the optimal guessing strategy in minimizing the expected log-loss. We completely characterize the minimal expected α-loss under k guesses in the next theorem, i.e., comprising the case when $P_X^{(\alpha)}(x) > \frac{1}{k}$, for some $x \in \mathcal{X}$ also.

Theorem 3 (Minimal Expected α-Loss Under k Guesses): Let P_X be a probability distribution supported on $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ such that $p_1 \geq p_2 \geq \cdots \geq p_n$, where $p_i := P_X(x_i)$, for $i \in [1 : n]$. Then the minimal expected α-loss under k guesses is given by

$$\mathcal{ME}_\alpha^{(k)}(P_X) = \frac{\alpha}{\alpha - 1} \sum_{i=s^*}^{n} p_i \left(1 - \left(\frac{(k - s^* + 1)p_i^\alpha}{\sum_{j=s^*}^n p_j^\alpha}\right)^{\frac{\alpha-1}{\alpha}}\right), \tag{28}$$

where

$$s^* = \min\left\{r \in \{1, 2, \ldots, k\} : \frac{(k - r + 1)p_r^\alpha}{\sum_{i=r}^n p_i^\alpha} \leq 1\right\}. \tag{29}$$

Remark 5: Theorem 3 implies that the minimal expected α-loss under k guesses induces a polychotomy on the simplex of probability distributions P_X on $\mathcal{X}$ depending on the value of the parameter s^*. Also, in an optimal guessing strategy, the adversary always guesses each of the $s^* - 1$ most likely outcomes in one of the k guesses with probability 1, i.e., $\mathrm{P}(\cup_{j=1}^k(\hat{X}_j = x_i)) = 1$, for $i \in [1 : s^* - 1]$, and guesses the remaining outcomes with a probability proportional their tilted probability values, i.e., $\mathrm{P}(\cup_{j=1}^k(\hat{X}_j = x_i)) = \frac{(k-s^*+1)p_i^\alpha}{\sum_{j=s^*}^n p_j^\alpha}$, for $i \in [s^* : n]$. It may not be immediately clear if there indeed exists an optimal guessing strategy $P_{\hat{X}_{[1:k]}}$ consistent with these value assignments to $\mathrm{P}(\cup_{j=1}^k(\hat{X}_j = x))$, $x \in \mathcal{X}$. However, as we show in the proof of Theorem 3 (see Lemma 6 and the discussion above it), it follows from Farkas' lemma [60, Proposition 6.4.3] that an arbitrary value assignment to $\mathrm{P}(\cup_{j=1}^k(\hat{X}_j = x))$, for each $x \in \mathcal{X}$, will in fact guarantee the existence of a consistent probability distribution $P_{\hat{X}_{[1:k]}}$ as long as the assignment is such that $\sum_{x\in\mathcal{X}} \mathrm{P}(\cup_{j=1}^k(\hat{X}_j = x)) = k$, which is true for the case above. For the special case when $k = s^* = 2$, this optimal strategy is exactly the same as that of a seemingly different guessing problem considered in [61, Section II-B].

Remark 6: Notice that whenever $s^* = 1$ in (29), the expression in (28) simplifies to

$$\frac{\alpha}{\alpha - 1}\left(1 - k^{\frac{\alpha-1}{\alpha}} \mathrm{e}^{\frac{1-\alpha}{\alpha}H_\alpha(X)}\right), \tag{30}$$

where $H_\alpha(X) = \frac{1}{1-\alpha}\log\left(\sum_{i=1}^n p_i^\alpha\right)$ is the Rényi entropy of order α [23]. Also, note that for the special case of $k = 1$, we always have $s^* = 1$, thereby recovering [30, Lemma 1].

A detailed proof of Theorem 3 is in Appendix B-C. The proof builds up on a careful decomposition of the simplex of probability distributions P_X and then solving the desired optimization problem in each case using tools from convex optimization. We present a simpler and a more descriptive proof for the special case of log-loss ($\alpha = 1$) and $k = 2$ in Appendix B-B. We remark that the proof of this special case does not directly generalize to arbitrary $\alpha \in (0, \infty]$ (even for the case of $k = 2$). However, we remark that it essentially captures the intuition and the main tools involved in the relatively complex analysis in the proof for the more general case with $\alpha \in (0, \infty]$ and $k \in \mathbb{N}$.

We illustrate the optimal guessing strategies to achieve the minimal expected α-loss in (28) through Examples 1 and 2.

Example 1 (Optimal Guessing Strategy With $k = 2$ Guesses): Consider a probability distribution P_X supported on $\mathcal{X} = \{x_1, x_2, x_3\}$. Let $p_i = P_X(x_i)$ and $p_i^{(\alpha)} = \frac{p_i^\alpha}{\sum_{j=1}^3 p_j^\alpha}$, for $i \in [1 : 3]$ and fix $\alpha = 2$. Optimal guessing strategy depends on the value of the parameter s^* in (29). We present an optimal guessing strategy for each case specified by $s^* \in [1 : 2]$. Notice that $\frac{p_r^\alpha}{\sum_{i=r}^n p_i^\alpha} = \frac{p_r^{(\alpha)}}{\sum_{i=r}^n p_i^{(\alpha)}}$.

With $s^* = 1$: Let $p_1 = \frac{3}{8}$, $p_2 = \frac{3}{8}$, and $p_3 = \frac{1}{4}$. This gives us $p_1^{(\alpha)} = \frac{9}{22}$, $p_2^{(\alpha)} = \frac{9}{22}$, and $p_3^{(\alpha)} = \frac{2}{11}$. It can be verified that $s^* = 1$ for this P_X. In particular, we have $p_i^{(\alpha)} \leq \frac{1}{2}$, for all $i \in [1 : 3]$. From (28), an optimal guessing strategy $P^*_{\hat{X}_1\hat{X}_2}$ is such that the adversary always guesses x_i in one of the two guesses with a probability proportional to $p_i^{(\alpha)}$, i.e., $\mathrm{P}^*(\hat{X}_1 = x_1 \text{ or } \hat{X}_2 = x_1) = 2p_1^{(\alpha)} = \frac{9}{11}$, $\mathrm{P}^*(\hat{X}_1 = x_2 \text{ or } \hat{X}_2 = x_2) = 2p_2^{(\alpha)} = \frac{9}{11}$, and $\mathrm{P}^*(\hat{X}_1 = x_3 \text{ or } \hat{X}_2 = x_3) = 2p_3^{(\alpha)} = \frac{4}{11}$.

With $s^* = 2$: Let $p_1 = \frac{2}{3}$, $p_2 = \frac{1}{4}$, and $p_3 = \frac{1}{12}$. This gives us $p_1^{(\alpha)} = \frac{32}{37}$, $p_2^{(\alpha)} = \frac{9}{74}$, and $p_3^{(\alpha)} = \frac{1}{74}$. It can be verified that $s^* = 2$ for this P_X. In particular, we have $p_1^{(\alpha)} > \frac{1}{2}$. From (28), an optimal guessing strategy $P^*_{\hat{X}_1\hat{X}_2}$ is such that the adversary always guesses x_1 in one of the two guesses, i.e., $\mathrm{P}^*(\hat{X}_1 = x_1 \text{ or } \hat{X}_2 = x_1) = 1$, and guesses x_i with a probability proportional to $p_i^{(\alpha)}$, for $i \in [2 : 3]$, i.e., $\mathrm{P}^*(\hat{X}_1 = x_2 \text{ or } \hat{X}_2 = x_2) = \frac{p_2^{(\alpha)}}{\sum_{j=2}^3 p_j^{(\alpha)}} = \frac{9}{10}$ and $\mathrm{P}^*(\hat{X}_1 = x_3 \text{ or } \hat{X}_2 = x_3) = \frac{p_3^{(\alpha)}}{\sum_{j=2}^3 p_j^{(\alpha)}} = \frac{1}{10}$.

Example 2 (Optimal Guessing Strategy With $k = 3$ Guesses): Let P_X be supported on $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$. Let $p_i = P_X(x_i)$ and $p_i^{(\alpha)} = \frac{p_i^\alpha}{\sum_{j=1}^4 p_j^\alpha}$, for $i \in [1 : 4]$ and fix $\alpha = 2$. We present an optimal guessing strategy for each case specified by $s^* \in [1 : 3]$.

With $s^* = 1$: Notice that specifying a tilted distribution $P_X^{(\alpha)}$ uniquely determines the original distribution P_X. Let P_X be such that $p_1^{(\alpha)} = p_2^{(\alpha)} = \frac{1}{4}$, $p_3^{(\alpha)} = \frac{1}{5}$, and $p_4^{(\alpha)} = \frac{3}{10}$. It can be verified that $s^* = 1$ for this P_X. In particular, we have $p_1^{(\alpha)} \leq \frac{1}{3}$, for all $i \in [1 : 4]$. From (28), an optimal guessing strategy $P^*_{\hat{X}_1\hat{X}_2\hat{X}_3}$ is such that the adversary guesses x_i in one of the three guesses with a probability proportional to $p_i^{(\alpha)}$, for $i \in [1 : 3]$, i.e., $\mathrm{P}^*(\cup_{j=1}^3(\hat{X}_j = x_1)) = 3p_1^{(\alpha)} =$

$\frac{3}{4}$, $\mathrm{P}^*(\cup_{j=1}^3(\hat{X}_j = x_2)) = 3p_2^{(\alpha)} = \frac{3}{5}$, and $\mathrm{P}^*(\cup_{j=1}(\hat{X}_j = x_3)) = 3p_3^{(\alpha)} = \frac{9}{10}$.

With $s^* = 2$: Let P_X be such that $p_1^{(\alpha)} = \frac{3}{8}$, $p_2^{(\alpha)} = \frac{1}{4}$, $p_3^{(\alpha)} = \frac{3}{16}$, and $p_4^{(\alpha)} = \frac{3}{16}$. It can be verified that $s^* = 2$ for this P_X. In particular, we have $p_1^{(\alpha)} > \frac{1}{3}$ and $\frac{2p_i^{(\alpha)}}{\sum_{j=2}^4 p_j^{(\alpha)}} \leq 1$, for $i \in [2:4]$. From (28), an optimal guessing strategy $P^*_{\hat{X}_1\hat{X}_2\hat{X}_3}$ is such that the adversary always guesses x_1 in one of the three guesses with a probability 1, i.e., $\mathrm{P}^*(\cup_{j=1}^3(\hat{X}_j = x_1)) = 1$, and guesses x_i with a probability proportional to $p_i^{(\alpha)}$, for $i \in [2:4]$, i.e., $\mathrm{P}^*(\cup_{j=1}^3(\hat{X}_j = x_2)) = \frac{2p_2^{(\alpha)}}{\sum_{j=2}^4 p_j^{(\alpha)}} = \frac{4}{5}$, and $\mathrm{P}^*(\cup_{j=1}(\hat{X}_j = x_i)) = \frac{2p_i^{(\alpha)}}{\sum_{j=2}^4 p_j^{(\alpha)}} = \frac{6}{10}$, for $i \in [3:4]$.

With $s^* = 3$: Let P_X be such that $p_1^{(\alpha)} = \frac{2}{3}$, $p_2^{(\alpha)} = \frac{1}{4}$, and $p_3^{(\alpha)} = p_4^{(\alpha)} = \frac{1}{24}$. It can be verified that $s^* = 3$ for this P_X. In particular, we have $p_1^{(\alpha)} > \frac{1}{3}$, $\frac{2p_2^{(\alpha)}}{\sum_{j=2}^4 p_j^{(\alpha)}} > 1$, and $\frac{p_3^{(\alpha)}}{\sum_{j=3}^4 p_j^{(\alpha)}} \leq 1$. From (28), an optimal guessing strategy $P^*_{\hat{X}_1\hat{X}_2\hat{X}_3}$ is such that the adversary always guesses x_i in one of the three guesses with a probability 1, i.e., $\mathrm{P}^*(\cup_{j=1}^3(\hat{X}_j = x_i)) = 1$, for $i \in [1:2]$, and guesses x_i with a probability proportional to $p_i^{(\alpha)}$, i.e., $\mathrm{P}^*(\cup_{j=1}^3(\hat{X}_j = x_i)) = \frac{p_i^{\alpha}}{\sum_{i=3}^4 p_i^{(\alpha)}} = \frac{1}{2}$, for $i \in [3:4]$.

The following two corollaries of Theorem 3 give the expressions for the minimal expected log-loss ($\alpha = 1$) for k guesses and the minimal expected 0-1 loss ($\alpha = \infty$) for k guesses, respectively.

Corollary 1 (Minimal expected log-loss {$\alpha = 1$} for k guesses): Under the notations of Theorem 3, the minimal expected log-loss for k guesses is given by

$$\mathcal{ME}_1^{(k)}(P_X) = H(X) - H_{s^*}\left(p_1, p_2, \ldots, p_{s^*-1}, \sum_{i=s^*}^{n} p_i\right) - \left(\sum_{i=s^*}^{n} p_i\right)\log(k - s^* + 1), \tag{31}$$

where $s^* = \min\left\{r \in \{1,2,\ldots,k\} : \frac{(k-r+1)p_r}{\sum_{i=r}^n p_i} \leq 1\right\}$ and $H_{s^*}(q_1, q_2, \ldots, q_{s^*}) := \sum_{i=1}^{s^*} q_i \log \frac{1}{q_i}$ is the entropy function.

The proof of Corollary 1 follows by taking limit $\alpha \to 1$ using L'Hôpital's rule in the result of Theorem 3 and rearranging the terms.

Corollary 2 (Minimal expected 0-1 loss {$\alpha = \infty$} for k guesses): Under the notations of Theorem 3, the minimal expected 0-1 loss for k guesses is given by

$$\mathcal{ME}_\infty^{(k)}(P_X) = 1 - \sum_{i=1}^{k} p_i = 1 - \max_{\substack{a_1,a_2,\ldots,a_k: \\ a_l \neq a_m, l \neq m}} \sum_{i=1}^{k} P_X(a_i). \tag{32}$$

The proof of Corollary 2 follows by taking limit $\alpha \to \infty$ in Theorem 3. Interestingly, the polychotomy induced by minimal expected α-loss for $\alpha \in (0,\infty)$ collapses for the case of $\alpha = \infty$ [13] as clear from Corollary 2.

A. α-Leakage and Maximal α-Leakage Under Multiple Guesses

Notice that minimizing the expected α-loss in (25) amounts to maximizing the expected gain for the gain function $g(t) = t^{\frac{\alpha-1}{\alpha}}$, $t \in (0,1) \cup (1,\infty)$. Motivated by α-leakage [30, Definition 5] (that is defined based on this gain function) which captures how much information an adversary can learn about a random variable X from a correlated random variable Y when a single guess is allowed, we define α-leakage under multiple guesses that captures the information an adversary can learn when k guesses are allowed.

Definition 9 (α-leakage under k guesses): Given a joint distribution P_{XY}, the α-leakage from X to Y under k guesses is defined as

$$\mathcal{L}_\alpha^{(k)}(X \to Y) \triangleq \frac{\alpha}{\alpha-1}\log\frac{\max\limits_{P_{\hat{X}_{[1:k]}|Y}} \mathbb{E}\left[\frac{\alpha}{\alpha-1}\mathrm{P}\left(\bigcup\limits_{i=1}^{k}(\hat{X}_i = X)|Y\right)^{\frac{\alpha-1}{\alpha}}\right]}{\max\limits_{P_{\hat{X}_{[1:k]}}} \mathbb{E}\left[\frac{\alpha}{\alpha-1}\mathrm{P}\left(\bigcup\limits_{i=1}^{k}(\hat{X}_i = X)\right)^{\frac{\alpha-1}{\alpha}}\right]}, \tag{33}$$

where $\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_k$ represent k estimators of X with the same support as X, for $\alpha \in (0,1) \cup (1,\infty)$, and by the continuous extension of (33) for $\alpha = 1$ and $\alpha = \infty$.

We call maximal g-leakage under k guesses when specialized to the gain function $g(t) = t^{\frac{\alpha-1}{\alpha}}$, $t \in (0,1) \cup (1,\infty)$, as maximal α-leakage under k guesses.

Definition 10 (Maximal α-leakage under k guesses): Given a joint distribution P_{XY}, the maximal α-leakage from X to Y under k guesses is defined as

$$\mathcal{L}_\alpha^{(k)-\max}(X \to Y) \triangleq \sup_{U-X-Y} \mathcal{L}_\alpha^{(k)}(U \to Y), \tag{34}$$

for $\alpha \in (0,1) \cup (1,\infty)$.

Let $P^{(\alpha)}_{X|Y=y}$ denote the tilted distribution of $P_{X|Y=y}$, i.e., $P^{(\alpha)}_{X|Y}(x|y) = \frac{P_{X|Y}(x|y)^\alpha}{\sum_x P_{X|Y}(x|y)^\alpha}$.

Theorem 4 (Robustness of α-leakage to number of guesses): Consider a P_{XY} such that $P^{(\alpha)}_{X|Y}(x|y) \leq \frac{1}{k}$ and $P_X^{(\alpha)}(x) \leq \frac{1}{k}$, for all x, y, and for $\alpha \in (0,1) \cup (1,\infty)$. Then

$$\mathcal{L}_\alpha^{(k)}(X \to Y) = \mathcal{L}_\alpha^{(1)}(X \to Y). \tag{35}$$

For $k = 1$, recall from Theorem 3 that in the optimal guessing strategy, the adversary guesses x with a probability that is equal to the tilted probability value $P_X^{(\alpha)}(x)$, $x \in \mathcal{X}$. For $k > 1$, when the condition $P_X^{(\alpha)}(x) \leq \frac{1}{k}$, for all x, holds, it follows from Theorem 3 that the adversary can apply essentially the same guessing strategy as with $k = 1$ in the following sense. In particular, when this condition holds, in the optimal guessing strategy, the adversary can still guess x in one of the k guesses with a probability that is proportional to tilted probability $P_X^{(\alpha)}(x)$, i.e., $\mathrm{P}^*(\cup_{j=1}^k(\hat{X}_j = x)) = kP_X^{(\alpha)}(x)$, $x \in \mathcal{X}$. Moreover, for the existence of a strategy to satisfy this equality, the condition $P_X^{(\alpha)}(x) \leq \frac{1}{k}$, for all x, is a necessary and sufficient condition. This is because if $P_X^{(\alpha)}(x) > \frac{1}{k}$,

for some $x \in \mathcal{X}$, then $\mathrm{P}^*(\cup_{j=1}^k(\hat{X}_j = x))$ cannot be equal to $kP_X^{(\alpha)}(x) > 1$. A detailed proof of Theorem 4 is in Appendix B-D. It is shown in [13, Theorem 4] that maximal leakage, i.e., maximal α-leakage with $\alpha = \infty$, does not change with the number of guesses, i.e., $\mathcal{L}_\infty^{(k)-\max}(X \to Y) = \mathcal{L}_\infty^{(1)-\max}(X \to Y)$. The proof of this result does not directly generalize to any arbitrary $\alpha \in (0, \infty]$. This is mainly due to the fact that unlike the case for $\alpha = \infty$, the minimal expected α-loss for arbitrary α ($\neq \infty$) induces a polychotomy on the simplex of probability distributions. However, we are able to show that maximal α-leakage under k guesses is at least that of under a single guess where the proof relies on a non-trivial observation about the optimizer in (34) and it also uses Theorem 4.

Theorem 5 (Lower Bound on Maximal α-Leakage Under k Guesses): Given a joint probability distribution P_{XY} on finite alphabet $\mathcal{X} \times \mathcal{Y}$, we have

$$\mathcal{L}_\alpha^{(k)-\max}(X \to Y) \geq \mathcal{L}_\alpha^{(1)-\max}(X \to Y), \tag{36}$$

for $\alpha \in (0, 1) \cup (1, \infty)$.

A detailed proof of Theorem 5 is in Appendix B-E. We do not know if the reverse direction of (36) is true, i.e., if $\mathcal{L}_\alpha^{(k)-\max}(X \to Y) \leq \mathcal{L}_\alpha^{(1)-\max}(X \to Y)$, in general.

V. A Variational Formula for ∞-Rényi Divergence With Applications to Information Leakage

In the previous two sections, we examined information leakage measures, in particular, maximal g-leakage, where the adversary is allowed to make multiple guesses. In this section, we study some variants of maximal g-leakage depending on the type of adversary by obtaining a variational characterization of Rényi divergence of order ∞. Such a characterization adds to a growing set of information measures written in a variational form, which may be of independent interest. We focus on the scenario where the adversary can make only a single guess in this section. In particular, we consider maximal expected gains of an adversary in separately guessing randomized functions of X and our variational characterization is in terms of the ratio of these maximal expected gains.

Theorem 6 (A variational characterization for $D_\infty(\cdot||\cdot)$): Let P_X and Q_X be two probability distributions on a finite alphabet $\mathcal{X}$ and $g : [0, 1] \to [0, \infty)$ be a gain function satisfying the following assumptions:

- $g(0) = 0$ and g is continuous at 0,
- $0 < \sup_{p \in [0,1]} g(p) < \infty$.

Then, we have

$$D_\infty(P_X||Q_X) = \sup_{P_{U|X}} \log \frac{\sup_{P_{\hat{U}}} \mathbb{E}_{U \sim P_U}\left[g(P_{\hat{U}}(U))\right]}{\sup_{P_{\hat{U}}} \mathbb{E}_{U \sim Q_U}\left[g(P_{\hat{U}}(U))\right]}, \tag{37}$$

where $P_U(u) = \sum_x P_X(x) P_{U|X}(u|x)$ and $Q_U(u) = \sum_x Q_X(x) P_{U|X}(u|x)$.

Remark 7: Interestingly, there are non-positive gain functions also for which while the conditions in Theorem 6 are not satisfied, (37) still holds. For example, $g(t) = \log t$ is one such function (see Appendix D for details).

A detailed proof of Theorem 6 is in Appendix C-A. The numerator and the denominator in the ratio in (37) capture the maximal expected gains in guessing an unknown random variable U distributed according to P_U or Q_U, respectively. In a way, this ratio compares the distributions P_X and Q_X and is certainly dependent on the gain function g. However, our variational characterization in (37) shows that this ratio when optimized over all the channels $P_{U|X}$ remains constant irrespective of the gain function used, i.e., (37) holds for a broad class of gain functions. Some examples of gain function g that satisfy the conditions in Theorem 6 are

$$g(p) = p^2, \ \mathbb{1}\{p = 1/2\}, \ \frac{\alpha}{\alpha - 1} p^{\frac{\alpha-1}{\alpha}} \text{ where } \alpha \in (1, \infty).$$

We obtain the following corollary from Theorem 6 by substituting the latter gain function $g_\alpha(t) = \frac{\alpha}{\alpha-1} t^{\frac{\alpha-1}{\alpha}}$, where $\alpha \in (1, \infty)$ (related to a class of adversarial loss functions, namely, α-loss [30]) and using [30, Lemma 1] which gives closed-form expressions for the corresponding optimization problems in the numerator and the denominator in (37).

Corollary 3: Given two probability distributions P_X and Q_X on a finite alphabet $\mathcal{X}$, we have, for $\alpha \in (1, \infty)$,

$$D_\infty(P_X||Q_X) = \sup_{P_{U|X}} \log \frac{\left(\sum_u P_U(u)^\alpha\right)^{\frac{1}{\alpha}}}{\left(\sum_u Q_U(u)^\alpha\right)^{\frac{1}{\alpha}}}, \tag{38}$$

where $P_U(u) = \sum_x P_X(x) P_{U|X}(u|x)$ and $Q_U(u) = \sum_x Q_X(x) P_{U|X}(u|x)$.

As mentioned earlier, we note that the existing variational characterizations for $D_\alpha(\cdot\|\cdot)$ (with finite α) also give rise to variational characterizations for $D_\infty(\cdot\|\cdot)$ by taking limit $\alpha \to \infty$. Shayevitz [24] and Birrell et al. [28] proved that

$$D_\alpha(P_X\|Q_X) = \sup_{R_X : R_X \ll P_X} \left(D(R_X\|Q_X) - \frac{\alpha D(R_X\|P_X)}{\alpha - 1}\right), \tag{39}$$

for $\alpha > 1$, and

$$D_\alpha(P_X\|Q_X) = \sup_{g : \mathcal{X} \to \mathbb{R}} \left(\frac{\alpha \log \mathbb{E}_{X \sim Q_X} \mathrm{e}^{(\alpha-1)g(X)}}{\alpha - 1} - \log \mathbb{E}_{X \sim P_X} \mathrm{e}^{\alpha g(X)}\right), \tag{40}$$

for $\alpha \in \mathbb{R} \setminus \{0, 1\}$, respectively. More general forms of (39) and an equivalent form of (40) appear in [25], [26], [27], and [62], respectively. One can obtain the variational characterizations for $D_\infty(\cdot\|\cdot)$ by taking limit $\alpha \to \infty$ in (39) and assuming interchangeability of the limit and the supremum; one can similarly do so, in (40), using a change of variable $f = \mathrm{e}^{\alpha g}$ and assuming interchangeability of the limit and the supremum. For the sake of completeness and rigor, we summarize the resulting variational forms for $D_\infty(\cdot\|\cdot)$ in the following proposition and present a proof in Appendix C-B.

Proposition 2: Given two probability distributions P_X and Q_X on a finite alphabet $\mathcal{X}$, we have

$$D_\infty(P_X\|Q_X) = \sup_{R_X : R_X \ll P_X} \left(D(R_X\|Q_X) - D(R_X\|P_X)\right), \tag{41}$$

$$D_\infty(P_X\|Q_X) = \sup_{f:\mathcal{X}\to[0,\infty)} \log \frac{\mathbb{E}_{X\sim P_X}[f(X)]}{\mathbb{E}_{X\sim Q_X}[f(X)]}. \tag{42}$$

Motivated by Issa et al. [13, Definitions 2 and 8], we define the following variants of maximal g-leakage (see Definition 5) depending on the type of adversary. In particular, note that the definition of maximal g-leakage corresponds to an adversary interested in a *single* randomized function of X. However, in some scenarios, the adversary could choose the guessing function depending on the realization of Y, leading to the following definition.

Definition 11 (Opportunistic Maximal g-Leakage): Given a gain function $g : [0,1] \to [0,\infty)$ and a joint probability distribution P_{XY} on a finite alphabet $\mathcal{X}\times\mathcal{Y}$, the opportunistic maximal g-leakage is defined as

$$\tilde{\mathcal{L}}_g^{\max}(X\to Y) = \log \sum_{y\in supp(Y)} P_Y(y) \sup_{U:U-X-Y} \frac{\sup_{P_{\hat{U}|Y=y}} \mathbb{E}_{U|Y=y}\left[g(P_{\hat{U}|Y}(U|y))\right]}{\sup_{P_{\hat{U}}} \mathbb{E}\left[g(P_{\hat{U}}(U))\right]}. \tag{43}$$

Maximal g-leakage captures the *average* guessing performance of the adversary over all the realizations of Y. In some contexts, it might be more relevant to consider *maximum* instead of the average.

Definition 12 (Maximal realizable g-leakage): Given a gain function $g : [0,1] \to [0,\infty)$ and a joint probability distribution P_{XY} on a finite alphabet $\mathcal{X}\times\mathcal{Y}$, the opportunistic maximal g-leakage is defined as

$$\mathcal{L}_g^{\mathrm{r-max}}(X\to Y) = \sup_{U:U-X-Y} \log \frac{\max_{y\in supp(Y)} \sup_{P_{\hat{U}|Y=y}} \mathbb{E}_{U|Y=y}\left[g(P_{\hat{U}|Y}(U|y))\right]}{\sup_{P_{\hat{U}}} \mathbb{E}\left[g(P_{\hat{U}}(U))\right]}. \tag{44, 45}$$

When $g(t) = t$, note that the Definitions 11 and 12 simplify to those of opportunistic maximal leakage [13, Definition 2] and maximal realizable leakage [13, Definition 8], respectively. Unlike the expressions for maximal g-leakage (e.g., for $g(t) = t$, $g(t) = \frac{\alpha}{\alpha-1}t^{\frac{\alpha-1}{\alpha}}$), interestingly, it turns out that the closed-form expressions for the opportunistic maximal g-leakage and maximal realizable g-leakage do not depend on the particular gain function g as long as it satisfies some mild regularity conditions. This is a consequence of the robustness of our variational characterization to gain function (Theorem 6).

Corollary 4 (Opportunistic maximal, and maximal realizable g-leakages): Let $g : [0,1] \to [0,\infty)$ be a function satisfying the following assumptions:

- $g(0) = 0$ and g is continuous at 0,
- $0 < \sup_{p\in[0,1]} g(p) < \infty$.

Then the opportunistic maximal g-leakage and maximal realizable g-leakage defined in (43) and (44), respectively, simplify to

$$\tilde{\mathcal{L}}_g^{\max}(X\to Y) = I_\infty^{\mathrm{S}}(X;Y), \tag{46}$$

$$\mathcal{L}_g^{\mathrm{r-max}}(X\to Y) = D_\infty(P_{XY}\|P_X\times P_Y). \tag{47}$$

Note that the intuitive interpretation given for $g(0) = 0$ below Theorem 1 holds for Corollary 4 also. A detailed proof of Corollary 4 is in Appendix C-C. When $g(t) = t$, Corollary 4 recovers the expressions for opportunistic maximal leakage and maximal realizable leakage [13, Theorems 2 and 13]. As another concrete example, consider Theorem 4 for the corresponding variants of maximal α-leakage with the gain function $g(t) = \frac{\alpha}{\alpha-1}t^{\frac{\alpha-1}{\alpha}}$, wherein the optimization problems in the numerators and the denominators of (43) and (44) have closed-form expressions [30, Lemma 1]. In particular, for $\alpha \in (1,\infty)$, the opportunistic maximal α-leakage is given by

$$\frac{\alpha}{\alpha-1}\log \sum_{y\in\mathrm{supp}(Y)} P_Y(y) \sup_{U:U-X-Y} \frac{\left(\sum_u P_{U|Y}(u|y)^\alpha\right)^{\frac{1}{\alpha}}}{\left(\sum_u P_U(u)^\alpha\right)^{\frac{1}{\alpha}}} = \frac{\alpha}{\alpha-1} I_\infty^{\mathrm{S}}(X;Y) \tag{48}$$

and the maximal realizable α-leakage is given by

$$\sup_{U:U-X-Y} \frac{\alpha}{\alpha-1}\log \frac{\max_{y\in\mathrm{supp}(Y)}\left(\sum_u P_{U|Y}(u|y)^\alpha\right)^{\frac{1}{\alpha}}}{\left(\sum_u P_U(u)^\alpha\right)^{\frac{1}{\alpha}}} = \frac{\alpha}{\alpha-1} D_\infty(P_{XY}\|P_X\times P_Y). \tag{49}$$

Note that when X and Y are independent, the opportunistic maximal $(\alpha = 1)$-leakage and the maximal realizable $(\alpha = 1)$-leakage defined by taking limit $\alpha \to 1$ are both equal to zero. When X and Y are not independent, it can be inferred from the above expressions that opportunistic maximal $(\alpha = 1)$-leakage and maximal realizable $(\alpha = 1)$-leakage are both equal to ∞. However, note that maximal α-leakage as $\alpha \to 1$ is equal to Shannon channel capacity or Shannon mutual information depending on whether we define it using the supremum first and the limit next, or the limit first and the supremum next [30, Theorem 2]. We show in Appendix E that the latter way of defining the opportunistic maximal 1-leakage and the maximal realizable 1-leakage also yields ∞.

Another interesting consequence of our variational characterization is in its natural connection to the definition of pointwise maximal leakage, another measure of information leakage studied by Saeidian et al. [16]. Pointwise maximal leakage captures the maximum multiplicative increase in the probability of correctly guessing *any* randomized function of X upon observing a single outcome $Y = y$.

Definition 13 (Pointwise maximal leakage [16]): Given a joint distribution P_{XY} on a finite alphabet $\mathcal{X}\times\mathcal{Y}$ and a $y \in \mathcal{Y}$, the pointwise maximal leakage is defined as

$$\mathcal{L}^{\text{pw-max}}(X\to y) = \sup_{U:U-X-Y} \log \frac{\sup_{P_{\hat{U}|Y=y}} \mathbb{E}_{U|Y=y}\left[P_{\hat{U}|Y}(U|y)\right]}{\sup_{P_{\hat{U}}} \mathbb{E}\left[P_{\hat{U}}(U)\right]}. \tag{50}$$

Notice that pointwise maximal leakage is also defined in terms of an adversary interested in maximizing the expected value of the gain function $g(t) = t$ but it differs from

the maximal leakage because the latter captures the average performance of the guessing adversary over all the realizations of Y. Saeidian et al. [16] showed that

$$\mathcal{L}^{\text{pw-max}}(X \to y) = D_\infty(P_{X|Y=y} \| P_X) \tag{51}$$

following the techniques in the proof of the closed-form expression for maximal realizable leakage [13, Theorem 13].

An immediate consequence of our variational characterization is that the definition of the pointwise maximal leakage can be generalized by incorporating an adversary interested in maximizing the expected values of an arbitrary gain function g that satisfies the mild regularity conditions mentioned in Corollary 6.

Corollary 5 (Pointwise Maximal g-Leakage): Let P_{XY} be a probability distribution on a finite alphabet $\mathcal{X} \times \mathcal{Y}$ and $g : [0,1] \to [0,\infty)$ be a gain function satisfying the following conditions:

- $g(0) = 0$ and g is continuous at 0,
- $0 < \sup_{p\in[0,1]} g(p) < \infty$.

Then, for $y \in \mathcal{Y}$, we have

$$\sup_{U:U-X-Y} \log \frac{\sup_{P_{\hat{U}|Y=y}} \mathbb{E}_{U|Y=y}\left[g(P_{\hat{U}|Y}(U|y))\right]}{\sup_{P_{\hat{U}}} \mathbb{E}\left[g(P_{\hat{U}}(U))\right]} = D_\infty(P_{X|Y=y} \| P_X). \tag{52}$$

The proof of this corollary follows directly from the proof of Corollary 4. However, this result may be of separate interest in view of the following observations. The expression on the LHS in (52) can be thought of as a new leakage measure, namely, pointwise maximal g-leakage (denoted by $\mathcal{L}_g^{\text{pw-max}}(X \to y)$) and Theorem 5 implies that pointwise maximal g-leakage is exactly the same as pointwise maximal leakage, i.e., $\mathcal{L}_g^{\text{pw-max}}(X \to y) = \mathcal{L}^{\text{pw-max}}(X \to y)$, for all g satisfying some mild conditions. Moreover, adopting the gain function approach of Alvim et al. [7], Saeidian et al. [16] showed that the multiplicative increase in the maximal expected value of the gain function (a function of the true value and the guessed value) in guessing X after observing an outcomne $Y = y$ when maximized over all gain functions $g(x, \hat{x})$ is equal to the pointwise maximal leakage ([16, Theorem 2 and Corollary 1]). On the other hand, Theorem 5 implies that such a multiplicative increase with gain function applied to the probability of correctly guessing a (randomized) function of X when maximized over all the functions of X is equal to pointwise maximal leakage, for *any* gain function g satisfying some mild conditions. Finally, we note that all the properties and privacy guarantees of pointwise maximal leakage studied in [16] (e.g., *composition* and *data-processing*) will straightaway generalize to pointwise maximal g-leakage.

VI. Conclusion

We have proposed a gain function viewpoint of information leakage using arbitrary non-negative functions applied to the probability of correctly guessing. The primary benefit of restricting the gain functions considered by [7] and [13] to the ones using the probability of correctness (as in Definition 1) is that it allows us to obtain closed-form expressions for maximal g-leakage for a class of concave gain functions (Theorem 1). We have shown that maximal g-leakage under multiple guesses is equal to Sibson mutual information of order infinity for a class of concave gain functions. This is in contrast to the corresponding results of [7] and [13] where the worst-case scenario over all the gain functions is considered. Moreover, obtaining a closed-form expression for such leakage measures with a fixed gain function was conjectured to be challenging in [7, Section VI-A] even when the threat model focuses on guessing X itself (rather than a possibly randomized function of X).

Even though the closed-form expression for maximal leakage is equal to that of min-capacity (i.e., the maximum min-entropy leakage) [3], maximal leakage is important mainly because of the relaxation of assumptions about the adversary. In particular, in the threat model considered for maximal leakage, the adversary is interested in the (possibly randomized) function of X. In contrast, in the setup of min-capacity, the adversary is interested in X itself. In the same vein, we extended the framework of maximal leakage to incorporate arbitrary functions of probability of correctness in the performance measure of the adversary. Thus, our setup actually leads to considering a larger class of adversaries because the gain functions in Theorem 1 recover the probability of correctness as a special case. We also presented a variational characterization of Rényi divergence of order infinity, which is naturally related to the pointwise version of maximal g-leakage.

We also studied the setting in which the adversary is allowed multiple attempts in guessing by focusing on a specific tunable gain function. Such a setting has not been explored earlier even for log-loss (a special case of the tunable loss function considered here) which is extensively used in machine learning. We have proved that a new measure of divergence that belongs to the class of Bregman divergences captures the relative performance of an arbitrary adversarial strategy with respect to an optimal strategy in minimizing the expected α-loss.

All these results strengthen the connection between privacy leakage and the information measures – Sibson mutual information and Rényi divergence of infinite orders. We believe that these results are beneficial in applications where the adversary's performance is measured via generalized gain functions. In terms of privacy-utility tradeoff, a consequence of our results is that we can obtain the same utility even when we consider various privacy leakage measures with potentially different operational interpretations as per the adversary's performance measure of interest. Motivated by the results in this work, we anticipate that closed-form expressions for maximal g-leakage measures for more gain functions depending on specific applications will be explored further. There are many questions to be further studied. One limitation of our study is that we have considered only the gain functions that are non-negative (though we studied a particular non-positive gain function, $g(t) = \log t$ in Appendix D). It would be interesting to characterize maximal g-leakage for any general class of gain functions. We have shown that maximal α-leakage under multiple guesses is at least that of with a single guess. It would

be worth studying if the reverse direction is also true as in $\alpha = \infty$ (maximal leakage) case.

APPENDIX A
PROOFS FOR SECTION III

A. Proof of Theorem 1

The upper bound $\mathcal{L}_g^{(k)-\max}(X \to Y) \le I_\infty^S(X;Y)$ follows from Proposition 1. We prove the lower bound now. For this, we use the "shattering" conditional distribution $P_{U|X}$ [13, Proof of Theorem 1], [15, Proof of Theorem 5]. Let $\mathcal{U} = \cup_{x\in\mathcal{X}}\mathcal{U}_x$ (a disjoint union) and $|\mathcal{U}_x| = m_x$. Then define

$$P_{U|X}(u|x) = \begin{cases} \frac{1}{m_x}, & u \in \mathcal{U}_x, \\ 0, & \text{otherwise.} \end{cases} \tag{53}$$

This gives

$$P_U(u) = P_X(x)/m_x, u \in \mathcal{U}_x, \tag{54}$$
$$P_{U|Y}(u|y) = P_{X|Y}(x|y)/m_x, u \in \mathcal{U}_x. \tag{55}$$

To simplify the notation, let

$$\hat{P}_k(u) := \mathrm{P}\left(\bigcup_{i=1}^k (\hat{U}_i = u)\right), \tag{56}$$
$$\hat{P}_k(u|y) := \mathrm{P}\left(\bigcup_{i=1}^k (\hat{U}_i = u)|Y = y\right). \tag{57}$$

We upper bound the denominator of

$$\frac{\sup_{P_{\hat{U}_{[1:k]}|Y}} \mathbb{E}_{UY}\left[g(\hat{P}_k(U|Y))\right]}{\sup_{P_{\hat{U}_{[1:k]}}} \mathbb{E}_U\left[g(\hat{P}_k(U))\right]} \tag{58}$$

as

$$\sup_{P_{\hat{U}_{[1:k]}}} \mathbb{E}_U\left[g(\hat{P}_k(U))\right]$$
$$= \sup_{P_{\hat{U}_{[1:k]}}} \sum_u g(\hat{P}_k(u))P_U(u) \tag{59}$$
$$\le \sup_{P_{\hat{U}_{[1:k]}}} \sum_u (g(0) + g'(0)\hat{P}_k(u))P_U(u) \tag{60}$$
$$= g'(0) \sup_{P_{\hat{U}_{[1:k]}}} \sum_u \hat{P}_k(u)P_U(u) \tag{61}$$
$$= g'(0) \max_{u_1,u_2,\ldots,u_k : u_i \ne u_j, i\ne j} \sum_{i=1}^k P_U(u_i) \tag{62}$$
$$= kg'(0) \max_u P_U(u), \tag{63}$$

where (60) follows because $g(s) \le g(t) + g'(t)(s-t)$, for all $s,t \in [0,1]$ since g is a concave function, (61) follows because $g(0) = 0$, and (63) follows from (54) if $k \le m_x$, for all $x \in \mathcal{X}$.

We now lower bound the numerator. Since the function g is differentiable at 0, there exists a function $h(x)$ such that

$$g(x) = g(0) + g'(0)x + xh(x), \lim_{x\to 0} h(x) = 0. \tag{64}$$

Notice that

$$\sup_{P_{\hat{U}_{[1:k]}|Y}} \mathbb{E}_{UY}\left[g(\hat{P}_k(U|Y))\right]$$
$$= \sum_y P_Y(y) \sup_{P_{\hat{U}_{[1:k]}|Y=y}} \mathbb{E}_{U|Y=y}\left[g(\hat{P}_k(U|y))\right]. \tag{65}$$

Consider, for a fixed $y \in \mathcal{Y}$,

$$\sup_{P_{\hat{U}_{[1:k]}|Y=y}} \mathbb{E}_{U|Y=y}\left[g(\hat{P}_k(U|y))\right] \tag{66}$$
$$= \sup_{P_{\hat{U}_{[1:k]}|Y=y}} \sum_u g(\hat{P}_k(u|y))P_{U|Y}(u|y) \tag{67}$$
$$= \sup_{P_{\hat{U}_{[1:k]}|Y=y}} \sum_x P_{X|Y}(x|y) \sum_{u\in\mathcal{U}_x} \frac{1}{m_x} g(\hat{P}_k(u|y)) \tag{68}$$
$$\ge \sup_{\substack{P_{\hat{U}_{[1:k]}|Y=y}: \forall u_1,u_2\in\mathcal{U}_x, x\in\mathcal{X} \\ \hat{P}_k(u_1|y)=\hat{P}_k(u_2|y)}} \left[\sum_x P_{X|Y}(x|y) \sum_{u\in\mathcal{U}_x} \frac{1}{m_x} g(\hat{P}_k(u|y))\right] \tag{69}$$
$$= \sup_{\substack{P_{\hat{U}_{[1:k]}|Y=y}: \forall u_1,u_2\in\mathcal{U}_x, x\in\mathcal{X} \\ \hat{P}_k(u_1|y)=\hat{P}_k(u_2|y)}} \left[\sum_x P_{X|Y}(x|y)\, g\left(\frac{1}{m_x}\sum_{u\in\mathcal{U}_x} \hat{P}_k(u|y)\right)\right] \tag{70}$$
$$= \sup_{P_{\hat{X}|Y=y}} \sum_x P_{X|Y}(x|y) g\left(\frac{k}{m_x} P_{\hat{X}|Y}(x|y)\right) \tag{71}$$
$$\ge \sup_{P_{\hat{X}|Y=y}} \sum_x P_{X|Y}(x|y)(g(0) + \frac{k}{m_x}P_{\hat{X}|Y}(x|y)(g'(0) - \epsilon)) \tag{72}$$
$$= (g'(0) - \epsilon) \sup_{P_{\hat{X}|Y=y}} \sum_x \frac{k}{m_x} P_{X|Y}(x|y)P_{\hat{X}|Y}(x|y) \tag{73}$$
$$= k(g'(0) - \epsilon) \max_x (\frac{1}{m_x} P_{X|Y}(x|y)) \tag{74}$$
$$= k(g'(0) - \epsilon) \max_x \max_{u\in\mathcal{U}_x} P_{U|Y}(u|y) \tag{75}$$
$$= k(g'(0) - \epsilon) \max_u P_{U|Y}(u|y), \tag{76}$$

where (69) follows because $\sup_{x\in A} f(x) \ge \sup_{x\in B} f(x)$ when $A \supseteq B$ and (70) follows because $\sup_{x\in B} f(x) = \sup_{x\in B} g(x)$ when $f(x) = g(x)$, for $x \in B$. We remark that it suffices to consider $P_{\hat{U}_{[1:k]}|Y=y}$ such that $\sum_u \hat{P}_k(u|y) = k$ in all the optimizations problems in (66)-(70) (see Lemma 5 and Remark 9 in Appendix B-C). Then, it follows that (70) $\le$ (71) by defining $P_{\hat{X}|Y}(x|y) := \frac{1}{k}\sum_{u\in\mathcal{U}_x} \hat{P}_k(u|y)$ which is a probability distribution noticing that $\mathcal{U}_x$, $x \in \mathcal{X}$, are disjoint and that $\sum_u \hat{P}_k(u|y) = k$. It follows that (70) $\ge$ (71) by defining $\hat{P}_k(u|y) := \frac{k}{m_x} P_{\hat{X}|Y}(x|y)$, for $u \in \mathcal{U}_x$, and using Lemma 6 (see also the discussion above it) in Appendix B-C. Thus, (70) = (71). The inequality (72) follows from (64) by choosing m_x appropriately large enough, for $x \in \mathcal{X}$ and for $0 < \epsilon \le g'(0)$. This gives

$$\sup_{P_{\hat{U}_{[1:k]}|Y}} \mathbb{E}_{UY}\left[g(\hat{P}_k(U|Y))\right]$$
$$\ge k(g'(0) - \epsilon) \sum_y P_Y(y) \max_u P_{U|Y}(u|y). \tag{77}$$

Putting together the bounds in (62) and (77) and using (65), we have

$$\sup_{U:U-X-Y} \frac{\sup_{P_{\hat{U}_{[1:k]}|Y}} \mathbb{E}_{UY}\left[g(\hat{P}_k(U|Y))\right]}{\sup_{P_{\hat{U}_{[1:k]}}} \mathbb{E}_U\left[g(\hat{P}_k(U))\right]} \tag{78}$$

$$\geq \sup_{0<\epsilon\leq g'(0)} \sup_{\substack{U:U-X-Y,\\ P_{U|X} \text{ in (53)}}} \frac{k(g'(0)-\epsilon)\sum_y P_Y(y)\max_u P_{U|Y}(u|y)}{kg'(0)\max_u P_U(u)} \tag{79}$$

$$= \sup_{0<\epsilon\leq g'(0)} \frac{(g'(0)-\epsilon)}{g'(0)} \sum_y \max_x P_{Y|X}(y|x) \tag{80}$$

$$= \sum_y \max_x P_{Y|X}(y|x), \tag{81}$$

where (80) follows because maximal leakage is achieved by the shattering $P_{U|X}$ (defined in (53)) with sufficiently large m_x [15] and (81) follows because $0 < g'(0) < \infty$.

B. Proof of Theorem 2

Note that

$$\mathcal{L}_g^{\max}(X \to Y)$$

$$= \sup_{U:U-X-Y} \log \frac{1+\sup_{P_{\hat{U}|Y}} \mathbb{E}_{UY}\left[P_{\hat{U}|Y}(U|Y)\right]}{1+\sup_{P_{\hat{U}}} \mathbb{E}_U\left[P_{\hat{U}}(U)\right]} \tag{82}$$

$$= \sup_{U:U-X-Y} \log \frac{1+\sum_y \max_u P_{UY}(u,y)}{1+\max_u P_U(u)}. \tag{83}$$

The lower bound follows directly by using the "Shattering" conditional distribution $P_{U|X}$ used in [13, Proof of Theorem 1]. In particular, let $p^* = \min_{x:P_X(x)>0}$, $k(x) = \frac{P_X(x)}{p^*}$, for each $x \in \text{supp}(X)$, and $\mathcal{U} = \cup_{x\in\text{supp}(X)}\{(x,1),\ldots,(x,\lceil k(x)\rceil)\}$. For each $u = (i_u, j_u) \in \mathcal{U}$ and $x \in \text{supp}(X)$, define $P_{U|X}$ as

$$P_{U|X}(i_u,j_u|X) = \begin{cases} \frac{p^*}{P_X(x)}, & i_u = x,\ j_u \in [1:\lfloor k(x)\rfloor], \\ 1-\frac{\lfloor k(x)\rfloor p^*}{P_X(x)}, & i_u = x,\ j_u = \lceil k(x)\rceil, \\ 0, & \text{otherwise.} \end{cases} \tag{84}$$

By substituting this into the objective function of (83), we get the lower bound

$$\mathcal{L}_g(X \to Y) \geq \log \frac{1+p^*\sum_{y\in\mathcal{Y}} \max_{x\in\text{supp}(X)} P_{Y|X}(y|x)}{1+p^*}. \tag{85}$$

We note that this lower bound holds for $\mathcal{X}$ and $\mathcal{Y}$ of arbitrary cardinalities. The binary assumption on $\mathcal{X}$ and $\mathcal{Y}$ is required for the upper bound. We prove the upper bound now. We have

$$\mathcal{L}_g^{\max}(X \to Y)$$

$$= \sup_{U:U-X-Y} \log \frac{\sup_{P_{\hat{U}|Y}} \mathbb{E}_{UY}\left[g(P_{\hat{U}|Y}(U|Y))\right]}{\sup_{P_{\hat{U}}} \mathbb{E}_U\left[g(P_{\hat{U}}(U))\right]} \tag{86}$$

$$= \sup_{U:U-X-Y} \log \frac{1+\sup_{P_{\hat{U}|Y}} \mathbb{E}_{UY}\left[P_{\hat{U}|Y}(U|Y)\right]}{1+\sup_{P_{\hat{U}}} \mathbb{E}_U\left[P_{\hat{U}}(U)\right]} \tag{87}$$

$$= \sup_{U:U-X-Y} \log \frac{1+\sum_y \max_u P_{UY}(u,y)}{1+\max_u P_U(u)} \tag{88}$$

$$= \sup_{q\in[0,1]} \sup_{\substack{U:U-X-Y,\\ \max_u P_U(u)\leq q}} \log \frac{1+\sum_y \max_u P_{UY}(u,y)}{1+q} \tag{89}$$

$$= \sup_{q\in[0,1]} \log \frac{1+\sup_{\substack{U:U-X-Y,\\ \max_u P_U(u)\leq q}} \sum_y \max_u P_{UY}(u,y)}{1+q}. \tag{90}$$

Let us consider just the optimization in the numerator in (90):

$$f(q) = \sup_{\substack{U:U-X-Y,\\ \max_u P_U(u)\leq q}} \sum_y \max_u P_{U,Y}(u,y). \tag{91}$$

We may upper bound it by

$$f(q)$$

$$= \sup_{\substack{U:U-X-Y,\\ \max_u P_U(u)\leq q}} \sum_y \max_u \sum_x P_U(u)P_{X|U}(x|u)P_{Y|X}(y|x) \tag{92}$$

$$\leq \sup_{\substack{U:U-X-Y,\\ \max_u P_U(u)\leq q}} q\sum_y \max_u \sum_x P_{X|U}(x|u)P_{Y|X}(y|x) \tag{93}$$

$$\leq q\sum_y \max_x P_{Y|X}(y|x). \tag{94}$$

We will also need another upper bound on $f(q)$ which uses the assumption that $\mathcal{X}$ and $\mathcal{Y}$ are binary. Let $\mathcal{X} = \{x_1, x_2\}$ and $\mathcal{Y} = \{y_1, y_2\}$. Assume without loss of generality that $P_X(x_1) \leq P_X(x_2)$. Also, assume without loss of generality that $x_1 = \arg\max_x P_{Y|X}(y_1|x)$. That is,

$$P_{Y|X}(y_1|x_1) \geq P_{Y|X}(y_1|x_2). \tag{95}$$

Since $\mathcal{Y}$ is binary, a consequence is that

$$P_{Y|X}(y_2|x_2) \geq P_{Y|X}(y_2|x_1). \tag{96}$$

That is, $x_2 = \arg\max_x P_{Y|X}(y_2|x)$. We may write (91) as

$$f(q) = \sup_{\substack{U:U-X-Y,\\ \max_u P_U(u)\leq q}} \max_u P_{U,Y}(u,y_1) + \max_u P_{U,Y}(u,y_2). \tag{97}$$

Let $u_i = \arg\max_u P_{U,Y}(u,y_i)$ for $i = 1,2$. If $u_1 = u_2$, then

$$f(q) \leq \sup_{\substack{U:U-X-Y,\\ \max_u P_U(u)\leq q}} P_{U,Y}(u_1,y_1) + P_{U,Y}(u_1,y_2)$$
$$= P_U(u_1) \leq q. \tag{98}$$

Now consider the case where $u_1 \neq u_2$. We construct an upper bound using weak duality as follows. Consider the constraints

$$\sum_{j=1}^{2} P_X(x_j)P_{U|X}(u_i|x_j) \leq q,\ \ i = 1,2, \tag{99}$$

$$\sum_{i=1}^{2} P_{U|X}(u_i|x_j) \leq 1,\ \ j = 1,2, \tag{100}$$

$$P_{U|X}(u_i|x_j) \geq 0,\ \ i = 1,2, j = 1,2. \tag{101}$$

We may upper bound $f(q)$ by

$$\begin{aligned} f(q) \leq & \min_{\substack{\lambda_i \geq 0,\ i=1,2,\\ \alpha_j \geq 0,\ j=1,2,\\ \nu_{ij} \geq 0,\ i=1,2,j=1,2}} \sup_{P_{U|X}} \sum_{i,j=1}^{2} P_{X,Y}(x_j,y_i) P_{U|X}(u_i|x_j) \\ & - \sum_{i=1}^{2} \lambda_i \left(\sum_{j=1}^{2} P_X(x_j) P_{U|X}(u_i|x_j) - q \right) \\ & - \sum_{j=1}^{2} \alpha_j \left(\sum_{i=1}^{2} P_{U|X}(u_i|x_j) - 1 \right) \\ & + \sum_{i,j=1}^{2} \nu_{ij} P_{U|X}(u_i|x_j) \qquad (102) \\ = & \min_{\substack{\lambda_i \geq 0,\ i=1,2,\\ \alpha_j \geq 0,\ j=1,2,\\ \nu_{ij} \geq 0,\ i=1,2,j=1,2}} \sup_{P_{U|X}} \sum_{i,j=1}^{2} \Big[P_{U|X}(u_i|x_j) \\ & \times (P_{X,Y}(x_j,y_i) - \lambda_i P_X(x_j) - \alpha_j + \nu_{ij}) \\ & + \sum_{i=1}^{2} \lambda_i q + \sum_{j=1}^{2} \alpha_j \Big]. \qquad (103) \end{aligned}$$

This leads to the upper bounding dual program

$$\begin{aligned} \text{minimize } & \sum_{i=1}^{2} \lambda_i q + \sum_{j=1}^{2} \alpha_j \\ \text{subject to } & \lambda_i \geq 0,\ i=1,2, \\ & \alpha_j \geq 0,\ j=1,2, \\ & \nu_{ij} \geq 0,\ i=1,2, j=1,2, \\ & P_{X,Y}(x_j,y_i) - \lambda_i P_X(x_j) - \alpha_j + \nu_{ij} = 0, \\ & i=1,2, j=1,2. \end{aligned}$$

Using the equality constraint to eliminate ν_{ij}, this can be further simplified to

$$\begin{aligned} \text{minimize } & \sum_{i=1}^{2} \lambda_i q + \sum_{j=1}^{2} \alpha_j \\ \text{subject to } & \lambda_i \geq 0,\ i=1,2, \\ & \alpha_j \geq 0,\ j=1,2, \\ & -P_{X,Y}(x_j,y_i) + \lambda_i P_X(x_j) + \alpha_j \geq 0, \\ & i=1,2, j=1,2. \end{aligned}$$

Noting that, for fixed α_j, $j \in \{1,2\}$, the optimal value of λ_i is $\max_j \max\{0, P_{Y|X}(y_i|x_j) - \frac{\alpha_j}{P_X(x_j)}\}$, the dual program can again be simplified to

$$\begin{aligned} \text{minimize } & q \sum_{i=1}^{2} \max_j \max \left\{ 0, P_{Y|X}(y_i|x_j) - \frac{\alpha_j}{P_X(x_j)} \right\} \\ & + \sum_{j=1}^{2} \alpha_j \\ \text{subject to } & \alpha_j \geq 0,\ j=1,2. \end{aligned}$$

We can now form a further upper bound on $f(q)$ by choosing α_j as we like. With some hindsight, we set

$$\alpha_1 = P_X(x_1)(P_{Y|X}(y_1|x_1) - P_{Y|X}(y_1|x_2)), \quad \alpha_2 = 0. \qquad (104)$$

By the assumption in (95), α_1 is non-negative. Now we have the upper bound

$$\begin{aligned} f(q) \leq & \, q \sum_{i=1}^{2} \max\{0, P_{Y|X}(y_i|x_1) \\ & - (P_{Y|X}(y_1|x_1) - P_{Y|X}(y_1|x_2)), P_{Y|X}(y_i|x_2)\} \\ & + P_X(x_1)(P_{Y|X}(y_1|x_1) - P_{Y|X}(y_1|x_2)) \qquad (105) \\ = & \, q \left(P_{Y|X}(y_1|x_2) + P_{Y|X}(y_2|x_2) \right) \\ & + P_X(x_1)(P_{Y|X}(y_1|x_1) - P_{Y|X}(y_1|x_2)) \qquad (106) \\ = & \, q + P_X(x_1)(P_{Y|X}(y_1|x_1) - P_{Y|X}(y_1|x_2)) \qquad (107) \end{aligned}$$

where in (106) we have used the assumption in (96). Note that the bound in (107) is larger than the bound $f(q) \leq q$ for the case where $u_1 = u_2$. Thus (107) is an upper bound on $f(q)$ in all cases. Let us now consider the quantity $\frac{1+f(q)}{1+q}$, separately for $q \leq P_X(x_1)$ and for $q \geq P_X(x_1)$. For $q \leq P_X(x_1)$, by the bound on $f(q)$ in (94),

$$\begin{aligned} \frac{1+f(q)}{1+q} & \leq \frac{1+q\sum_y \max_x P_{Y|X}(y|x)}{1+q} \qquad (108) \\ & \leq \frac{1+P_X(x_1)\sum_y \max_x P_{Y|X}(y|x)}{1+P_X(x_1)} \qquad (109) \end{aligned}$$

where the inequality (109) holds because the right-hand side of (108) is non-decreasing in q as $\sum_y \max_x P_{Y|X}(y|x) \geq 1$. For $q \geq P_X(x_1)$, by the bound on $f(q)$ in (107),

$$\begin{aligned} & \frac{1+f(q)}{1+q} \\ & \leq \frac{1+q+P_X(x_1)(P_{Y|X}(y_1|x_1) - P_{Y|X}(y_1|x_2))}{1+q} \qquad (110) \\ & \leq \frac{1+P_X(x_1)+P_X(x_1)(P_{Y|X}(y_1|x_1) - P_{Y|X}(y_1|x_2))}{1+P_X(x_1)} \qquad (111) \\ & = \frac{1+P_X(x_1)(P_{Y|X}(y_1|x_1) + P_{Y|X}(y_2|x_2))}{1+P_X(x_1)} \qquad (112) \\ & = \frac{1+P_X(x_1)\sum_y \max_x P_{Y|X}(y|x)}{1+P_X(x_1)} \qquad (113) \end{aligned}$$

where (111) holds because the right-hand side of (110) is non-increasing in q since $P_X(x_1)(P_{Y|X}(y_1|x_1) - P_{Y|X}(y_1|x_2)) \geq 0$. This proves that

$$\begin{aligned} \mathcal{L}_g^{\max}(X \to Y) & = \sup_{q \in [0,1]} \log \frac{1+f(q)}{1+q} \qquad (114) \\ & \leq \log \frac{1+P_X(x_1)\sum_y \max_x P_{Y|X}(y|x)}{1+P_X(x_1)}. \qquad (115) \end{aligned}$$

By the assumption that $P_X(x_1) = \min_x P_X(x) = p^*$, we have proven an upper bound on $\mathcal{L}_g^{\max}(X \to Y)$ matching the lower bound in (85).

APPENDIX B
PROOFS FOR SECTION IV

A. Proof of Proposition 1

Let $P_{\tilde{X}}(x) := \frac{\mathrm{P}(\cup_{i=1}^{k}(\hat{X}_i=x))}{k}$. Consider the following simplification to the expected α-loss under k guesses.

$$\begin{aligned}
&\mathbb{E}_X\left[\ell_\alpha(\mathrm{P}(\cup_{i=1}^{k}(\hat{X}_i=X)))\right]\\
&=\mathbb{E}\left[\ell_\alpha(kP_{\tilde{X}}(X))\right]\\
&=\frac{\alpha}{\alpha-1}\sum_x P_X(x)\left(1-(kP_{\tilde{X}}(x))^{\frac{(\alpha-1)}{\alpha}}\right) &(116)\\
&=\frac{\alpha}{\alpha-1}(1-\sum_x P_X(x)(kP_{\tilde{X}}(x))^{\frac{(\alpha-1)}{\alpha}}) &(117)\\
&=\frac{\alpha}{\alpha-1}\Bigg(1-k^{\frac{\alpha-1}{\alpha}}(\sum_x P_X(x)^\alpha)^{\frac{1}{\alpha}}+k^{\frac{\alpha-1}{\alpha}}(\sum_x P_X(x)^\alpha)^{\frac{1}{\alpha}}\\
&\quad-\sum_x P_X(x)(kP_{\tilde{X}}(x))^{\frac{(\alpha-1)}{\alpha}}\Bigg) &(118)\\
&=\frac{\alpha}{\alpha-1}\Bigg(1-k^{\frac{\alpha-1}{\alpha}}\mathrm{e}^{(\frac{1-\alpha}{\alpha}H_\alpha(X))}+k^{\frac{\alpha-1}{\alpha}}(\sum_x P_X(x)^\alpha)^{\frac{1}{\alpha}}\\
&\quad-\sum_x P_X(x)(kP_{\tilde{X}}(x))^{\frac{(\alpha-1)}{\alpha}}\Bigg) &(119)\\
&=\frac{\alpha}{\alpha-1}\Bigg(1-k^{\frac{\alpha-1}{\alpha}}\mathrm{e}^{(\frac{1-\alpha}{\alpha}H_\alpha(X))}+k^{\frac{\alpha-1}{\alpha}}(\sum_x P_X(x)^\alpha)^{\frac{1}{\alpha}}\\
&\quad-(\sum_x P_X(x)^\alpha)^{\frac{1}{\alpha}}\sum_x\Big(\frac{P_X(x)^\alpha}{\sum_{x'}P_X(x')^\alpha}\Big)^{\frac{1}{\alpha}}(kP_{\tilde{X}}(x))^{(1-\frac{1}{\alpha})}\Bigg) &(120)\\
&=\frac{\alpha}{\alpha-1}\Bigg(1-k^{\frac{\alpha-1}{\alpha}}\mathrm{e}^{(\frac{1-\alpha}{\alpha}H_\alpha(X))}+k^{\frac{\alpha-1}{\alpha}}(\sum_x P_X(x)^\alpha)^{\frac{1}{\alpha}}\\
&\quad\times\Big(1-\sum_x\Big(\frac{P_X(x)^\alpha}{\sum_{x'}P_X(x')^\alpha}\Big)^{\frac{1}{\alpha}}P_{\tilde{X}}(x)^{(1-\frac{1}{\alpha})}\Big)\Bigg) &(121)\\
&=\frac{\alpha}{\alpha-1}\Bigg(1-k^{\frac{\alpha-1}{\alpha}}\mathrm{e}^{(\frac{1-\alpha}{\alpha}H_\alpha(X))}+k^{\frac{\alpha-1}{\alpha}}(\sum_x P_X(x)^\alpha)^{\frac{1}{\alpha}}\\
&\quad\times\Big(1-\sum_x P_X^{(\alpha)}(x)^{\frac{1}{\alpha}}P_{\tilde{X}}(x)^{(1-\frac{1}{\alpha})}\Big)\Bigg) &(122)\\
&=\frac{\alpha}{\alpha-1}\Bigg(1-k^{\frac{\alpha-1}{\alpha}}\mathrm{e}^{\frac{1-\alpha}{\alpha}H_\alpha(X)}+k^{\frac{\alpha-1}{\alpha}}(\sum_x P_X(x)^\alpha)^{\frac{1}{\alpha}}\\
&\quad\times\Big(1-\mathrm{e}^{\frac{1-\alpha}{\alpha}D_{\frac{1}{\alpha}}(P_X^{(\alpha)}\|P_{\tilde{X}})}\Big)\Bigg) &(123)\\
&=\frac{\alpha}{\alpha-1}\left(1-k^{\frac{\alpha-1}{\alpha}}\mathrm{e}^{\frac{1-\alpha}{\alpha}H_\alpha(X)}\right)+k^{\frac{\alpha-1}{\alpha}}B_F(P_X,P_{\hat{X}}^{(\frac{1}{\alpha})}), &(124)
\end{aligned}$$

where (124) follows from Appendix F with Bregman divergence B_F associated with $F(P_X)=\frac{\alpha}{\alpha-1}\left((\sum_x P_X(x)^\alpha)^{\frac{1}{\alpha}}-1\right)$. In view of Lemma 5 and Remark 9 in Appendix B-C, we can consider $P_{\tilde{X}}$ to be a probability distribution, i.e., $\sum_x P_{\tilde{X}}(x)=1$, for optimizing the expected α-loss. Now from the non-negativity of the Rényi divergence, it can be seen that the last term inside the brackets in (123) is non-negative and is equal to zero if and only if there exists a guessing strategy $P_{\hat{X}_{[1:k]}}$ such that $P_{\tilde{X}}=P_X^{(\alpha)}$. There always exists such a guessing strategy for the special case of $k=1$, in particular, $P_{\hat{X}_1}=P_X^{(\alpha)}$, thereby giving an alternative proof of [30, Lemma 1] in addition to quantifying the relative performance of an arbitrary guessing strategy with respect to the optimal guessing strategy. However, this is not always possible when $k>1$. In particular, it follows from Lemma 6 and the discussion above it (in Appendix B-C) that there exists a guessing strategy $P_{\hat{X}_{[1:k]}}$ such that $P_{\tilde{X}}=P_X^{(\alpha)}$ if and only if $P_X^{(\alpha)}(x)\leq\frac{1}{k}$, for all $x\in\mathcal{X}$. Then the minimal expected α-loss is given by $\mathcal{ME}_\alpha^{(k)}(P_X)=\frac{\alpha}{\alpha-1}(1-k^{\frac{\alpha-1}{\alpha}}\mathrm{e}^{\frac{1-\alpha}{\alpha}H_\alpha(X)})$ when $P_X^{(\alpha)}(x)\leq\frac{1}{k}$, for all $x\in\mathcal{X}$.

B. Minimal Expected Log-Loss Under 2 Guesses

Here we present a proof of Theorem 3 for the special case of $\alpha=1$ and $k=2$ that essentially captures the intuition and the main tools involved in the relatively complex analysis in the proof of Theorem 3.

Theorem 7 (Minimal Expected log-loss {$\alpha=1$} under 2 guesses): Let P_X be a probability distribution supported on $\mathcal{X}=\{x_1,x_2,\ldots,x_n\}$, and let $p_{\max}=\max_{i\in[1:n]}P_X(x_i)=P_X(x_{\max})$. Then the minimal expected log-loss under 2 guesses is given by

$$\begin{aligned}
\min_{P_{\hat{X}_1\hat{X}_2}}\mathbb{E}\left[\log\frac{1}{\mathrm{P}\left(\bigcup_{i=1}^{2}(\hat{X}_i=X)\right)}\right]\\
=H(X)-h_b\left(\max\{\frac{1}{2},p_{\max}\}\right), &(125)
\end{aligned}$$

where $h_b(\cdot)$ is the binary entropy function defined by $h_b(t):=-t\log t-(1-t)\log(1-t)$ and the optimal guessing strategy is given by $P_{\hat{X}_1,\hat{X}_2}$ such that

$$\begin{cases}
\mathrm{P}(\hat{X}_1=x\ \textit{or}\ \hat{X}_2=x)=2P_X(x),\\
\quad\textit{if}\ P_X(x_{\max})\leq\frac{1}{2},\\
P_{\hat{X}_1,\hat{X}_2}(x_{\max},x)+P_{\hat{X}_1,\hat{X}_2}(x,x_{\max})=\frac{P_X(x)}{\sum_{x':x'\neq x_{\max}}P_X(x')},\\
\quad x\neq x_{\max},\textit{if}\ P_X(x_{\max})>\frac{1}{2}.
\end{cases} \tag{126}$$

Remark 8: In words, the optimal guessing strategy in (126) is to guess x in either of the guesses with a probability proportional to $P_X(x)$, for all x, if $p_{\max}\leq\frac{1}{2}$. If $p_{\max}>\frac{1}{2}$, the optimal strategy is to guess $x_{\max}$ with probability 1 (i.e., deterministically) and randomly guess the other x with a probability proportional to $P_X(x)$, for $x\neq x_{\max}$.

It can be inferred from (125) that the minimal expected log-loss under 2 guesses induces a dichotomy on the simplex of probability distributions P_X on $\mathcal{X}$. *Proof:* [Proof of Theorem 7] We first state some useful lemmas which will be needed in the proof.

Lemma 2 (Non-Negativity): If $\tilde{P}_X$ is a probability distribution and $\tilde{Q}_X$ is such that $\tilde{Q}_X(x)\geq 0$ for all x and $\sum_x\tilde{Q}_X(x)\leq 1$ with the same support set as $\tilde{P}_X$. Then $D(\tilde{P}_X||\tilde{Q}_X)\geq 0$.

Lemma 2 is proved in Appendix G.

Lemma 3 (Non-Equality): If $P^*_{\hat{X}_1,\hat{X}_2}$ is an optimal strategy for the optimization problem in (125), then

$$P^*_{\hat{X}_1,\hat{X}_2}(x,x) = 0, \text{ for all } x. \tag{127}$$

Lemma 3 follows from Lemma 5 for the special case of $k = 2$.

Lemma 4: There exists $P_{\hat{X}_1\hat{X}_2}$ such that $\mathrm{P}\left(\hat{X}_1 = x \text{ or } \hat{X}_2 = x\right) = 2P_X(x)$, for all x if and only if $P_X(x) \leq \frac{1}{2}$, for all x.

Lemma 4 follows from Lemma 6 for the special case of $k = 2$. A consequence of Lemma 3 is that if $P^*_{\hat{X}_1\hat{X}_2}$ is an optimal strategy, then we have

$$\sum_x \mathrm{P}^*\left(\hat{X}_1 = x \text{ or } \hat{X}_2 = x\right) = 2, \tag{128}$$

where the probability P^* is taken with respect to the optimal strategy $P^*_{\hat{X}_1\hat{X}_2}$. So, it suffices to consider the optimization in (125) over the strategies $P_{\hat{X}_1\hat{X}_2}$ satisfying (128). Notice that for such strategies, $P_{\tilde{X}}(x) := \frac{\Pr(\hat{X}_1 = x \text{ or } \hat{X}_2 = x)}{2}$ is a probability distribution. Now we are ready to prove Theorem 7. Let $p_{\max} \leq \frac{1}{2}$. Then we have

$$\begin{aligned}
&\mathbb{E}\left[\ell_1(\mathrm{P}(\hat{X}_1 = x \text{ or } \hat{X}_2 = x))\right] \\
&= \sum_x P_X(x) \log \frac{1}{\mathrm{P}(\hat{X}_1 = x \text{ or } \hat{X}_2 = x)} && (129)\\
&= \sum_x P_X(x) \log \frac{1}{P_X(x)} \\
&\quad + \sum_x P_X(x) \log \frac{P_X(x)}{\mathrm{P}(\hat{X}_1 = x \text{ or } \hat{X}_2 = x)} && (130)\\
&= H(X) - 1 + \sum_x P_X(x) \log \frac{2P_X(x)}{\mathrm{P}(\hat{X}_1 = x \text{ or } \hat{X}_2 = x)} && (131)\\
&= H(X) - 1 + D(P_X \| P_{\tilde{X}}) && (132)\\
&\geq H(X) - 1 && (133)
\end{aligned}$$

where we have defined $P_{\tilde{X}}(x) = \frac{\Pr(\hat{X}_1 = x \text{ or } \hat{X}_2 = x)}{2}$ in (132), and (133) follows from the non-negativity of the relative entropy. We remark that (132) can be obtained from (27) also by substituting $k = 2$ and taking limit $\alpha \to 1$. Notice that the inequality in (133) can be tight if and only if there exists $P_{\hat{X}_1\hat{X}_2}$ such that $P_{\tilde{X}}(x) = 2P_X(x)$, for all x, which is possible if and only if $p_{\max} \leq \frac{1}{2}$ using Lemma 4.

Now, consider the case when $p_{\max} > \frac{1}{2}$. Without loss of generality, suppose that $p_{\max} = P_X(x_1)$. Let $t_x = \mathrm{P}(\hat{X}_1 = x \text{ or } \hat{X}_2 = x)$. For example, $t_{x_1} = \sum_{j=2}^{n} p_{1j} + \sum_{j=2}^{n} p_{j1} = 2\sum_{j=2}^{n} p_{1j}$, where $P_{\hat{X}_1,\hat{X}_2}(x_i, x_j) = p_{ij}$. Also, note that $\sum_{i=1}^{n} t_x = 2$ in view of Lemma 3. Then we have

$$\begin{aligned}
&\mathbb{E}\left[\ell_{\log}(X, P_{\hat{X}_1,\hat{X}_2})\right] \\
&= P_X(x_1) \log \frac{1}{t_{x_1}} + \sum_{i=2}^{n} P_X(x_i) \log \frac{1}{t_{x_i}} && (134)\\
&= (P_X(x_1) - \sum_{i=2}^{n} P_X(x_i)) \log \frac{1}{t_{x_1}} + (\sum_{i=2}^{n} P_X(x_i)) \log \frac{1}{t_{x_1}} \\
&\quad + \sum_{i=2}^{n} P_X(x_i) \log \frac{1}{(\frac{P_X(x_i)}{\sum_{j=2}^{n} P_X(x_j)})} \\
&\quad + \sum_{i=2}^{n} P_X(x_i) \log \frac{(\frac{P_X(x_i)}{\sum_{j=2}^{n} P_X(x_j)})}{t_{x_i}} && (135)\\
&= \sum_{i=2}^{n} P_X(x_i) \log \frac{1}{(\frac{P_X(x_i)}{\sum_{j=2}^{n} P_X(x_j)})} \\
&\quad + (P_X(x_1) - \sum_{i=2}^{n} P_X(x_i)) \log \frac{1}{t_{x_1}} \\
&\quad + \sum_{i=2}^{n} P_X(x_i) \log \frac{(\frac{P_X(x_i)}{\sum_{j=2}^{n} P_X(x_j)})}{t_{x_1} t_{x_i}} && (136)
\end{aligned}$$

The second term in (136) is non-negative since $P_X(x_1) > 0.5$ is equivalent to $P_X(x_1) > \sum_{i=2}^{n} P_X(x_i)$. Consider the third term in (136).

$$\begin{aligned}
&\sum_{i=2}^{n} P_X(x_i) \log \frac{(\frac{P_X(x_i)}{\sum_{j=2}^{n} P_X(x_j)})}{t_{x_1} t_{x_i}} \\
&= (\sum_{j=2}^{n} P_X(x_j)) \sum_{i=2}^{n} (\frac{P_X(x_i)}{\sum_{j=2}^{n} P_X(x_j)}) \log \frac{(\frac{P_X(x_i)}{\sum_{j=2}^{n} P_X(x_j)})}{t_{x_1} t_{x_i}} && (137)\\
&= (\sum_{j=2}^{n} P_X(x_j)) D(\tilde{P}_X \| \tilde{Q}_X) && (138)\\
&\geq 0, && (139)
\end{aligned}$$

where (138) follows by defining $\tilde{P}_X(x_i) = \frac{P_X(x_i)}{\sum_{j=2}^{n} P_X(x_j)}$ and $\tilde{Q}_X(x_i) = t_{x_1} t_{x_i}$, $i = [2:n]$, (139) follows from Lemma 2 noticing that $\sum_{i=2}^{n} \tilde{Q}_X(x_i) = t_{x_1}(\sum_{i=2}^{n} t_{x_i}) = t_{x_1}(2 - t_{x_1}) \leq 1$. Equality in (139) is attained if and only if $t_{x_1} = 1$ and $p_{1i} + p_{i1} = \frac{P_X(x_i)}{\sum_{j=2}^{n} P_X(x_j)}$, $i \in [2:n]$. Under this condition, note that the second term in (136) is also zero and the first term simplifies to $H(X) - h_{\mathrm{b}}(p_{\max})$. □

C. Proof of Theorem 3

We begin with the following lemmas which will be useful in the proof of Theorem 3. It is intuitive to expect that an optimal strategy, $P^*_{\hat{X}_{[1:k]}}$, puts zero weight on ordered tuples $(a_1, a_2, \ldots, a_k)$ (denoted as $a_{[1:k]}$ in the sequel) whenever $a_i = a_j$ for some $i \neq j$, since there is no advantage in guessing

the same estimate more than once. The following lemma based on the monotonicity of the α-loss formalizes this.

Lemma 5: If $P^*_{\hat{X}_{[1:k]}}$ is an optimal strategy for the optimization problem in (25), then

$$P^*_{\hat{X}_{[1:k]}}(a_{[1:k]})=0, \text{ for all } a_{[1:k]} \text{ s.t. } a_i=a_j, \text{ for some } i\neq j.$$

The proof of Lemma 5 is deferred to Appendix H.

Remark 9: An important consequence of Lemma 5 is that, if $P^*_{\hat{X}_{[1:k]}}$ is an optimal strategy for the optimization problem in (25), then we have

$$\sum_x \mathrm{P}^*\left(\bigcup_{i=1}^{k}(\hat{X}_i=x)\right)=k, \tag{140}$$

where the probability P^* is taken with respect to an optimal strategy $P^*_{\hat{X}_{[1:k]}}$. Hence, it suffices to consider the optimization in (25) over all the strategies $P_{\hat{X}_{[1:k]}}$ satisfying (140).

A vector $(t_1,t_2,\ldots,t_n)$ such that $\sum_{i=1}^n t_i=k$ is said to be *admissible* if there exists a strategy $P_{\hat{X}_{[1:k]}}$ satisfying

$$t_i=\mathrm{P}\left(\bigcup_{j=1}^{k}(\hat{X}_j=x_i)\right), \text{ for all } i\in[1:n]. \tag{141}$$

Equivalently, (141) can be written as the following system of linear equations.

$$t_i=\sum_{a_{[1:k]}:\bigcup_{j=1}^{k}(a_j=x_i)} P_{\hat{X}_{[1:k]}}(a_{[1:k]}), \text{ for all } i\in[1:n]. \tag{142}$$

In general, in order to determine whether a vector $(t_1,t_2,\ldots,t_n)$ is admissible or not, we need to solve a linear programming problem (LPP) with number of variables and constraints that are polynomial in the support size of P_X, i.e, n. Nonetheless, the following lemma based on Farkas' lemma [60, Proposition 6.4.3] completely characterizes the necessary and sufficient conditions for the admissibility of a vector $(t_1,t_2,\ldots,t_n)$.

Lemma 6: A vector $(t_1,t_2,\ldots,t_n)$ such that $\sum_{i=1}^n t_i=k$ is admissible if and only if $0\leq t_i\leq 1$, for all $i\in[1:n]$.

The proof of Lemma 6 is deferred to Appendix I. We now prove Theorem 7. From the definition of the minimal expected α-loss for k guesses in (25), we have

$$\mathcal{ME}^{(k)}_\alpha(P_X)$$
$$=\min_{P_{\hat{X}_{[1:k]}}}\frac{\alpha}{\alpha-1}\left[\sum_{i=1}^n p_i\left(1-\mathrm{P}\left(\bigcup_{j=1}^{k}(\hat{X}_j=x_i)\right)^{\frac{\alpha-1}{\alpha}}\right)\right] \tag{143}$$
$$=\min_{P_{\hat{X}_{[1:k]}}}\frac{\alpha}{\alpha-1}\left[\sum_{i=1}^n p_i\left(1-\mathrm{P}\left(\bigcup_{j=1}^{k}(\hat{X}_j=x_i)\right)^{\frac{\alpha-1}{\alpha}}\right)\right]$$
$$\text{s.t. } \sum_{i=1}^n \mathrm{P}\left(\bigcup_{j=1}^{k}(\hat{X}_j=x_i)\right)=k \tag{144}$$
$$=\min_{t_1,\ldots,t_n}\frac{\alpha}{\alpha-1}\left[\sum_{i=1}^n p_i(1-t_i^{\frac{\alpha-1}{\alpha}})\right]$$
$$\text{s.t. } \sum_{i=1}^n t_i=k,$$
$$0\leq t_i\leq 1,\ i\in[1:n], \tag{145}$$

where (144) follows from Lemma 5 and Remark 9, and (145) follows from the change of variable $t_i=\mathrm{P}\left(\bigcup_{j=1}^{k}(\hat{X}_j=x_i)\right)$ and Lemma 6. Consider the Lagrangian

$$\mathcal{L}=\frac{\alpha}{\alpha-1}\left[\sum_{i=1}^n p_i(1-t_i^{\frac{\alpha-1}{\alpha}})\right]+\lambda\left(\sum_{i=1}^n t_i-k\right)$$
$$+\sum_{i=1}^n \mu_i(t_i-1) \tag{146}$$

The Karush-Kuhn-Tucker (KKT) conditions [63, Chapter 5.5.3] are given by

$$\text{(Stationarity): } \frac{\partial\mathcal{L}}{\partial t_i}=0, i\in[1:n],$$
$$\text{i.e., } t_i=\left(\frac{p_i}{\lambda+\mu_i}\right)^\alpha, i\in[1:n], \tag{147}$$
$$\text{(Primal feasibility): } \sum_{i=1}^n t_i=k, 0\leq t_i\leq 1, i\in[1:n], \tag{148}$$
$$\text{(Dual feasibility): } \mu_i\geq 0, i\in[1:n], \tag{149}$$
$$\text{(Complementary slackness): } \mu_i(t_i-1)=0, i\in[1:n]. \tag{150}$$

Notice that for $\alpha>1$, $t^{\frac{\alpha-1}{\alpha}}$ is a concave function of t, meaning the overall objective function in (145) is convex. For $\alpha<1$, $t^{\frac{\alpha-1}{\alpha}}$ is a convex function of t, but since $\frac{\alpha}{\alpha-1}$ is negative, the overall function is again convex. Thus (145) amounts to a convex optimization problem. Now since KKT conditions are necessary and sufficient conditions for optimality in a convex optimization problem, it suffices to find values of t_i, $i\in[1:n]$, λ, μ_i, $i\in[1:n]$ satisfying (147)–(150) in order to solve the optimization problem (145).

First we simplify the KKT conditions (147)–(150) in the following manner.

- For i such that $\left(\frac{p_i}{\lambda}\right)^\alpha\leq 1$, we take $\mu_i=0$ and $t_i=\left(\frac{p_i}{\lambda}\right)^\alpha$.
- For i such that $\left(\frac{p_i}{\lambda}\right)^\alpha>1$, we take $\mu_i=p_i-\lambda$ and $t_i=1$. Notice that for such i, we have $\mu_i>0$, since $p_i>\lambda$.

This is equivalent to choosing $t_i=\min\left\{\left(\frac{p_i}{\lambda}\right)^\alpha,1\right\}$ and $\mu_i=0$ or $\mu_i=p_i-\lambda$ depending on whether $t_i=\left(\frac{p_i}{\lambda}\right)^\alpha$ or $t_i=1$, respectively, for each $i\in[1:n]$. Notice that this choice is consistent with the KKT conditions (147)–(150) except for that λ has to be chosen appropriately satisfying $\sum_{i=1}^n t_i=k$ also. In effect, we have essentially reduced the KKT conditions (147)–(150) to the following equations by eliminating μ_i's:

$$t_i=\min\left\{\left(\frac{p_i}{\lambda}\right)^\alpha,1\right\}, i\in[1:n], \tag{151}$$
$$\sum_{i=1}^n t_i=k. \tag{152}$$

We solve the equations (151) and (152) by considering the following k mutually exclusive and exhaustive cases (clarified later) based on P_X.

Case 1: $\left(\frac{p_1^\alpha}{\sum_{i=1}^n p_i^\alpha} \le \frac{1}{k}\right)$:
Consider the choice

$$\lambda = \left(\frac{\sum_{i=1}^n p_i^\alpha}{k}\right)^{\frac{1}{\alpha}}, \; t_i = \frac{k p_i^\alpha}{\sum_{j=1}^n p_j^\alpha}, i \in [1:n]. \tag{153}$$

This choice satisfies (151) and (152) since $\frac{k p_1^\alpha}{\sum_{i=1}^n p_i^\alpha} \le 1$ and $p_1 \ge p_2 \cdots \ge p_n$.

Case 's': $(2 \le s \le k)$ $\left(\frac{(k-s+2)p_{s-1}^\alpha}{\sum_{i=s-1}^n p_i^\alpha} > 1, \frac{(k-s+1)p_s^\alpha}{\sum_{i=s}^n p_i^\alpha} \le 1\right)$:
Consider the choice

$$\lambda = \left(\frac{\sum_{i=s}^n p_i^\alpha}{k-s+1}\right)^{\frac{1}{\alpha}}, \tag{154}$$

$$t_i = 1, i \in [1:s-1], t_i = \frac{(k-s+1)p_i^\alpha}{\sum_{j=s}^n p_j^\alpha}, i \in [s:n]. \tag{155}$$

This choice satisfies (151)

- for $i \in [1:s-1]$ because $\frac{(k-s+2)p_{s-1}^\alpha}{\sum_{i=s-1}^n p_i^\alpha} > 1$ and $p_1 \ge p_2 \ge \cdots \ge p_{s-1}$, and
- for $i \in [s:n]$ because $\frac{(k-s+1)p_s^\alpha}{\sum_{i=s}^n p_i^\alpha} \le 1$ and $p_s \ge p_{s+1} \ge \cdots \ge p_n$.

Also, this choice clearly satisfies (152). Finally, notice that the condition for Case 's', $2 \le s \le n$, can be written as

$$\frac{(k-i+1)p_i^\alpha}{\sum_{j=i}^n p_j^\alpha} > 1, \text{ for } i \in [1:s-1], \frac{(k-s+1)p_s^\alpha}{\sum_{i=s}^n p_i^\alpha} \le 1 \tag{156}$$

since $\frac{(k-s+2)p_{s-1}^\alpha}{\sum_{i=s-1}^n p_i^\alpha} > 1$ and $p_1 \ge p_2 \ge \cdots \ge p_{s-1}$. This proves that the cases considered above are mutually exclusive and exhaustive, and together with the case-wise analysis gives the expression for the minimal expected α-loss for k guesses as presented in Theorem 3.

D. Proof of Theorem 4

From the definition of α-leakage with k guesses in (33), we have

$$\begin{aligned}
&\mathcal{L}_\alpha^{(k)}(X \to Y) \\
&= \frac{\alpha}{\alpha-1} \log \frac{\max\limits_{P_{\hat{X}_{[1:k]}|Y}} \mathbb{E}\left[\frac{\alpha}{\alpha-1}\mathrm{P}\left(\bigcup_{i=1}^k (\hat{X}_i = X)|Y\right)^{\frac{\alpha-1}{\alpha}}\right]}{\max\limits_{P_{\hat{X}_{[1:k]}}} \mathbb{E}\left[\frac{\alpha}{\alpha-1}\mathrm{P}\left(\bigcup_{i=1}^k (\hat{X}_i = X)\right)^{\frac{\alpha-1}{\alpha}}\right]} && (157)\\
&= \frac{\alpha}{\alpha-1} \log \frac{\frac{\alpha}{\alpha-1} k^{\frac{\alpha-1}{\alpha}} \mathrm{e}^{\frac{1-\alpha}{\alpha} H_\alpha^A(X|Y)}}{\frac{\alpha}{\alpha-1} k^{\frac{\alpha-1}{\alpha}} \mathrm{e}^{\frac{1-\alpha}{\alpha} H_\alpha(X)}} && (158)\\
&= \frac{\alpha}{\alpha-1} \log \frac{\mathrm{e}^{\frac{1-\alpha}{\alpha} H_\alpha^A(X|Y)}}{\mathrm{e}^{\frac{1-\alpha}{\alpha} H_\alpha(X)}} && (159)\\
&= \mathcal{L}_\alpha^{(1)}(X \to Y), && (160)
\end{aligned}$$

where (158) follows from Theorem 3, in particular from the case when $s^* = 1$ since $P_{X|Y}^{(\alpha)}(x|y) \le \frac{1}{k}$, for all x, y and $P_X^{(\alpha)}(x) \le \frac{1}{k}$, for all x, and $H_\alpha^A(X|Y)$ in (158) is the Arimoto conditional entropy [64] defined as $H_\alpha^A(X|Y) = \frac{\alpha}{1-\alpha} \log \sum_y \left(\sum_x P_{XY}(x,y)^\alpha\right)^{\frac{1}{\alpha}}$.

E. Proof of Theorem 5

Consider

$$\mathcal{L}_\alpha^{(1)-\max}(X \to Y) = \sup_{U-X-Y} I_\alpha^A(U;Y). \tag{161}$$

Then we have the following lemma proved later.

Lemma 7: There exists an optimizer $P_{U^*|X}$ for the optimization problem in the RHS of (161) such that $P_{U^*|Y}^{(\alpha)}(u|y)$, $P_{U^*}^{(\alpha)}(u) \le \frac{1}{k}$, for all u, y, where $P_{U^*|Y}^{(\alpha)}(u|y) = \frac{P_{U^*|Y}^\alpha(u|y)}{\sum_u P_{U^*|Y}^\alpha(u|y)}$ and $P_{U^*}^{(\alpha)}(u) = \frac{P_{U^*}^\alpha(u)}{\sum_u P_{U^*}^\alpha(u)}$.

Then for $P_{U^*|X}$ in the Lemma 7, by definition we have

$$\begin{aligned}
&\mathcal{L}_\alpha^{(k)-\max}(X \to Y) \\
&\ge \frac{\alpha}{\alpha-1} \log \frac{\max_{P_{\hat{U}_{[1:k]}|Y}} \mathbb{E}[\frac{\alpha}{\alpha-1}\mathrm{P}(\bigcup_{i=1}^k (\hat{U}_i = U^*)|Y)^{\frac{\alpha-1}{\alpha}}]}{\max_{P_{\hat{U}_{[1:k]}}} \mathbb{E}[\frac{\alpha}{\alpha-1}\mathrm{P}(\bigcup_{i=1}^k \hat{U}_i = U^*)^{\frac{\alpha-1}{\alpha}}]} && (162)\\
&= \frac{\alpha}{\alpha-1} \log \frac{\frac{\alpha}{\alpha-1} k^{\frac{\alpha-1}{\alpha}} \mathrm{e}^{\frac{1-\alpha}{\alpha} H_\alpha^A(U^*|Y)}}{\frac{\alpha}{\alpha-1} k^{\frac{\alpha-1}{\alpha}} \mathrm{e}^{\frac{1-\alpha}{\alpha} H_\alpha(U^*)}} && (163)\\
&= I_\alpha^A(U^*;Y) && (164)\\
&= \mathcal{L}_\alpha^{(1)-\max}(X \to Y), && (165)
\end{aligned}$$

where (163) follows from Theorem 3, (165) follows because $P_{U^*|X}$ is an optimizer in (161). It remains to prove Lemma 7.

Proof of Lemma 7: Note that it suffices to prove that for every $P_{U|X}$, there exists $P_{\tilde{U}|U}$ such that $I_\alpha^A(U;Y) = I_\alpha^A(\tilde{U};Y)$ and $P_{\tilde{U}|Y}^{(\alpha)}(\tilde{u}|y), P_{\tilde{U}}^{(\alpha)}(\tilde{u}) \le \frac{1}{k}$, for all u, y. To that end, we use the "shattering" conditional distribution $P_{U|X}$ [13, Proof of Theorem 1], [15, Proof of Theorem 5]. Let us first define $\tilde{\mathcal{U}} = \bigcup_{u \in \mathcal{U}} \tilde{\mathcal{U}}_u$, with $\tilde{\mathcal{U}}_u = \{(u,1),(u,2),\ldots,(u,m)\}$ for some m to be fixed later. Let $P_{\tilde{U}|U}(\tilde{u}|u) = \frac{1}{m}$, for $\tilde{u} \in \tilde{\mathcal{U}}_u$. This gives

$$\begin{aligned}
&P_{\tilde{U}|Y}(\tilde{u}|y) = \sum_u P_{U|Y}(u|y) P_{\tilde{U}|U}(\tilde{u}|u) = \frac{P_{U|Y}(u|y)}{m}, \\
&\text{for } \tilde{u} \in \tilde{\mathcal{U}}_u, && (166)\\
&P_{\tilde{U}}(\tilde{u}) = \sum_u P_U(u) P_{\tilde{U}|U}(\tilde{u}|u) = \frac{P_U(u)}{m} \text{ for } \tilde{u} \in \tilde{\mathcal{U}}_u, && (167)\\
&P_{U|\tilde{U}}(u|\tilde{u}) = 1 \text{ for } \tilde{u} \in \tilde{\mathcal{U}}_u, && (168)\\
&P_{Y|\tilde{U}}(y|\tilde{u}) = P_{Y|U}(y|u) \text{ for } \tilde{u} \in \tilde{\mathcal{U}}_u. && (169)
\end{aligned}$$

Now we have

$$\begin{aligned}
&I_\alpha^A(\tilde{U};Y) \\
&= \frac{\alpha}{\alpha-1} \log \frac{\sum_y \left(\sum_u \sum_{\tilde{u} \in \tilde{\mathcal{U}}_u} P_{Y|\tilde{U}}(y|\tilde{u})^\alpha P_{\tilde{U}}(\tilde{u})^\alpha\right)^{\frac{1}{\alpha}}}{\left(\sum_u \sum_{\tilde{u} \in \tilde{\mathcal{U}}_u} P_{\tilde{U}}(\tilde{u})^\alpha\right)^{\frac{1}{\alpha}}} && (170)
\end{aligned}$$

$$= \frac{\alpha}{\alpha-1}\log\frac{\sum_y\left(\sum_u P_{Y|U}(y|u)^\alpha \sum_{\tilde{u}\in\tilde{\mathcal{U}}_u} P_{\tilde{U}}(\tilde{u})^\alpha\right)^{\frac{1}{\alpha}}}{\left(\sum_u \sum_{\tilde{u}\in\tilde{\mathcal{U}}_u} P_{\tilde{U}}(\tilde{u})^\alpha\right)^{\frac{1}{\alpha}}} \tag{171}$$

$$= \frac{\alpha}{\alpha-1}\log\frac{\sum_y\left(\sum_u P_{Y|U}(y|u)^\alpha \frac{P_U(u)^\alpha}{m^{\alpha-1}}\right)^{\frac{1}{\alpha}}}{\left(\sum_u \frac{P_U(u)^\alpha}{m^{\alpha-1}}\right)^{\frac{1}{\alpha}}} \tag{172}$$

$$= \frac{\alpha}{\alpha-1}\log\frac{\sum_y\left(\sum_u P_{Y|U}(y|u)^\alpha P_U(u)^\alpha\right)^{\frac{1}{\alpha}}}{\left(\sum_u P_U(u)^\alpha\right)^{\frac{1}{\alpha}}} \tag{173}$$

$$= I_\alpha^A(U;Y), \tag{174}$$

where (171) follows from (169), and (172) follows from (167). Now consider

$$P_{\tilde{U}|Y}^{(\alpha)}(\tilde{u}|y) = \frac{P_{\tilde{U}|Y}(\tilde{u}|y)^\alpha}{\sum_u\sum_{\tilde{u}\in\tilde{\mathcal{U}}_u} P_{\tilde{U}|Y}(\tilde{u}|y)^\alpha} \tag{175}$$

$$= \frac{\frac{P_{U|Y}(u|y)^\alpha}{m^\alpha}}{\sum_u m\cdot\frac{P_U(u|y)^\alpha}{m^\alpha}} \tag{176}$$

$$= \frac{1}{m}\cdot\frac{P_{U|Y}(u|y)^\alpha}{\sum_u P_U(u|y)^\alpha} \tag{177}$$

$$\le \frac{1}{m}, \tag{178}$$

where (176) follows from (166). Now we choosing any m such that $m \ge k$ guarantees that $P_{\tilde{U}|Y}^{(\alpha)}(\tilde{u}|y) \le \frac{1}{k}$. Similarly, $P_{\tilde{U}}^{(\alpha)}(\tilde{u}) \le \frac{1}{k}$.

APPENDIX C
PROOFS FOR SECTION V

A. Proof of Theorem 6

We first prove the lower bound LHS $\ge$ RHS. Consider

$$\sup_{P_{U|X}} \log\frac{\sup_{P_{\hat{U}}}\mathbb{E}_{U\sim P_U} g(P_{\hat{U}}(U))}{\sup_{P_{\hat{U}}}\mathbb{E}_{U\sim Q_U} g(P_{\hat{U}}(U))}$$

$$= \sup_{P_{U|X}} \log\frac{\sup_{P_{\hat{U}}}\sum_{x,u} P_X(x)P_{U|X}(u|x)g(P_{\hat{U}}(u))}{\sup_{P_{\hat{U}}}\sum_{x,u} Q_X(x)P_{U|X}(u|x)g(P_{\hat{U}}(u))} \tag{179}$$

$$= \sup_{P_{U|X}}\sup_{P_{\hat{U}}}\inf_{Q_{\hat{U}}} \log\frac{\sum_{x,u} P_X(x)P_{U|X}(u|x)g(P_{\hat{U}}(u))}{\sum_{x,u} Q_X(x)P_{U|X}(u|x)g(Q_{\hat{U}}(u))} \tag{180}$$

$$\le \sup_{P_{U|X}}\sup_{P_{\hat{U}}} \log\frac{\sum_{x,u} P_X(x)P_{U|X}(u|x)g(P_{\hat{U}}(u))}{\sum_{x,u} Q_X(x)P_{U|X}(u|x)g(P_{\hat{U}}(u))} \tag{181}$$

$$\le \sup_{P_{U|X}}\sup_{P_{\hat{U}}}\max_{x:P_X(x)>0} \log\frac{P_X(x)\sum_u P_{U|X}(u|x)g(P_{\hat{U}}(u))}{Q_X(x)\sum_u P_{U|X}(u|x)g(P_{\hat{U}}(u))} \tag{182}$$

$$= \max_{x:P_X(x)>0}\log\frac{P_X(x)}{Q_X(x)} \tag{183}$$

$$= D_\infty(P_X\|Q_X). \tag{184}$$

where (182) follows because $\frac{\sum_i a_i}{\sum_i b_i} \le \max_i \frac{a_i}{b_i}$, for $b_i > 0$, $\forall i$.

Now we prove the upper bound LHS $\le$ RHS. We lower bound the RHS of (37) by using the "shattering" $P_{U|X}$ [13, Proof of Theorem 1], [15, Proof of Theorem 5]. We pick a letter $x^\star$, and let $\mathcal{U} = \{x^\star\} \uplus \biguplus_{x\neq x^\star}\mathcal{U}_x$, where $|\mathcal{U}_x| = m$ for each $x \neq x^\star$. Then define

$$P_{U|X}(u|x) = \begin{cases} 1 & u = x = x^\star, \\ 1/m & u\in\mathcal{U}_x, x\neq x^\star, \\ 0 & \text{otherwise.}\end{cases} \tag{185}$$

Note that

$$P_U(u) = \begin{cases} P_X(x^\star) & u = x^\star \\ P_X(x)/m, & u\in\mathcal{U}_x, x\neq x^\star\end{cases}, \tag{186}$$

and

$$Q_U(u) = \begin{cases} Q_X(x^\star) & u = x^\star \\ Q_X(x)/m, & u\in\mathcal{U}_x, x\neq x^\star\end{cases}. \tag{187}$$

Consider the numerator of the objective function in the RHS of (37). We have

$$\sup_{P_{\hat{U}}}\mathbb{E}_{U\sim P_U}\left[g(P_{\hat{U}}(U))\right] = \sup_{P_{\hat{U}}}\sum_u P_U(u)g(P_{\hat{U}}(u)) \tag{188}$$

$$\ge \sup_{P_{\hat{U}}} P_X(x^\star)g(P_{\hat{U}}(x^\star)) \tag{189}$$

$$= P_X(x^\star)\sup_{q\in[0,1]} g(q). \tag{190}$$

Note that the expression in (190) is finite because of the assumption on g that $\sup_{q\in[0,1]} g(q) < \infty$.

To bound the denominator of the objective function in the RHS of (37), we will need the upper concave envelope of g, denoted g^{**}. Since g is a function of a scalar, its upper concave envelope can be written as

$$g^{**}(q) = \sup_{a,b,\lambda\in[0,1]:a\lambda+b(1-\lambda)=q} \lambda g(a) + (1-\lambda)g(b). \tag{191}$$

We claim that $g^{**}(0) = 0$ and g^{**} is continuous at 0. Fix some $\epsilon > 0$. It suffices to show that there exists $\delta > 0$ where $g^{**}(q) \le \epsilon$ for all $q \in [0,\delta]$. By the assumption that $g(0) = 0$ and g is continuous at 0, there exists a δ small enough so that $g(q) \le \epsilon/2$ for all $q \in [0,\sqrt{\delta}]$. Now, for any $q\in[0,\delta]$, consider any a, b, λ where $a\lambda + b(1-\lambda) = q$. We assume without loss of generality that $a \le q \le b$. If $b \le \sqrt{\delta}$, then we have $\lambda g(a) + (1-\lambda)g(b) \le \epsilon/2$. If $b > \sqrt{\delta}$, then we have

$$q = a\lambda + b(1-\lambda) \ge b(1-\lambda) > \sqrt{\delta}(1-\lambda). \tag{192}$$

So we get $1-\lambda < \frac{q}{\sqrt{\delta}} \le \frac{\delta}{\sqrt{\delta}} \le \sqrt{\delta}$. Thus

$$\lambda g(a) + (1-\lambda)g(b) \le \epsilon/2 + \sqrt{\delta}\sup_{q\in[0,1]} g(q) \le \epsilon, \tag{193}$$

where (193) holds for sufficiently small δ, and again we have used the assumption that $\sup_{q\in[0,1]} g(q) < \infty$. This proves that $g^{**}(q) \le \epsilon$ whenever $q \in [0,\delta]$. In particular, for sufficiently large m,

$$\sup_{q\in[0,1/m]} g^{**}(q) \le \epsilon. \tag{194}$$

Now the denominator in (37) can be upper bounded as

$$\sup_{P_{\hat{U}}}\mathbb{E}_{U\sim Q_U}\left[g(P_{\hat{U}}(U))\right]$$

$$= \sup_{P_{\hat{U}}}\sum_u Q_U(u)g(P_{\hat{U}}(u)) \tag{195}$$

$$= \sup_{P_{\hat{U}}}(Q_X(x^\star)g(P_{\hat{U}}(x^\star)) + \sum_{x\neq x^\star} Q_X(x)\sum_{u\in\mathcal{U}_x}\frac{1}{m}g(P_{\hat{U}}(u))) \quad (196)$$

$$\leq \sup_{P_{\hat{U}}} Q_X(x^\star)g(P_{\hat{U}}(x^\star)) + \sum_{x\neq x^\star} Q_X(x)g^{**}(\sum_{u\in\mathcal{U}_x}\frac{1}{m}P_{\hat{U}}(u)) \quad (197)$$

$$\leq \sup_{q\in[0,1]} Q_X(x^\star)g(q) + \sum_{x\neq x^\star} Q_X(x)\sup_{q\in[0,1/m]} g^{**}(q) \quad (198)$$

$$\leq \sup_{q\in[0,1]} Q_X(x^\star)g(q) + (1-Q_X(x^\star))\epsilon, \quad (199)$$

where (197) follows from the definition of the upper concave envelope and (199) follows from (194) for sufficiently large m.

Putting together the bounds in (190) and (199), we have

$$\sup_{P_{U|X}} \log \frac{\sup_{P_{\hat{U}}} \mathbb{E}_{U\sim P_U} g(P_{\hat{U}}(U))}{\sup_{P_{\hat{U}}} \mathbb{E}_{U\sim Q_U} g(P_{\hat{U}}(U))} \geq \log \max_{x^\star} \sup_{\epsilon>0} \frac{\sup_{q\in[0,1]} P_X(x^\star)g(q)}{\sup_{q\in[0,1]} Q_X(x^\star)g(q) + (1-Q_X(x^\star))\epsilon} \quad (200)$$

$$= \log \max_{x^\star} \frac{\sup_{q\in[0,1]} P_X(x^\star)g(q)}{\sup_{q\in[0,1]} Q_X(x^\star)g(q)} \quad (201)$$

$$= \log \max_{x^\star} \frac{P_X(x^\star)}{Q_X(x^\star)} \quad (202)$$

$$= D_\infty(P_X\|Q_X), \quad (203)$$

where (202) uses the assumption that $\sup_{q\in[0,1]} g(q) < \infty$. Finally, we note that the assumption $\sup_{q\in[0,1]} g(q) > 0$ is to ensure that the objective function in (37) is well-defined. In particular, for any $P_{U|X}$, fix a u' such that $P_U(u') > 0$. Then we have

$$\sup_{P_{\hat{U}}} \mathbb{E}_{U\sim P_U}\left[g(P_{\hat{U}}(U))\right] \geq \sup_{P_{\hat{U}}} g(P_{\hat{U}}(u'))P_U(u') \quad (204)$$

$$= P_U(u')\sup_{q\in[0,1]} g(q) > 0. \quad (205)$$

Similarly, $\sup_{P_{\hat{U}}} \mathbb{E}_{U\sim Q_U}\left[g(P_{\hat{U}}(U))\right] > 0$.

B. Proof of Proposition 2

For any $R_X \ll P_X$, consider

$$D(R_X\|Q_X) - D(R_X\|P_X) = \sum_x R_X(x)\log\frac{R_X(x)}{Q_X(x)} - \sum_x R_X(x)\log\frac{R_X(x)}{P_X(x)} \quad (206)$$

$$= \sum_x R_X(x)\log\frac{P_X(x)}{Q_X(x)} \quad (207)$$

$$\leq \sum_x R_X(x)\left(\max_{x'}\log\frac{P_X(x')}{Q_X(x')}\right) \quad (208)$$

$$= \max_{x'}\log\frac{P_X(x')}{Q_X(x')} \quad (209)$$

$$= D_\infty(P_X\|Q_X). \quad (210)$$

Moreover, for R_X such that $R_X(x^*) = 1$ for a fixed $x^* \in \arg\max \frac{P_X(x)}{Q_X(x)}$, (208) is tight. This proves (41).

To prove (42), for the upper bound, we give a choice of the function f for which the objective function in the RHS of (42) is equal to $D_\infty(P_X||Q_X)$. In particular, fix an $x^* \in \arg\max \frac{P_X(x)}{Q_X(x)}$ and consider a function $\tilde{f}$ defined by

$$\tilde{f}(x) = \begin{cases} 1, & \text{if } x = x^*, \\ 0, & \text{otherwise} \end{cases}. \quad (211)$$

Clearly, we have

$$\log\frac{\mathbb{E}_{X\sim P_X}[\tilde{f}(X)]}{\mathbb{E}_{X\sim Q_X}[\tilde{f}(X)]} = \log\frac{P_X(x^*)}{Q_X(x^*)} = D_\infty(P_X||Q_X). \quad (212)$$

For the lower bound, consider

$$\log\frac{\mathbb{E}_{X\sim P_X}[f(X)]}{\mathbb{E}_{X\sim Q_X}[f(X)]} = \log\frac{\sum_x P_X(x)f(x)}{\sum_x Q_X(x)f(x)} \quad (213)$$

$$\leq \log\max_x \frac{P_X(x)f(x)}{Q_X(x)f(x)} \quad (214)$$

$$= \max_x \log\frac{P_X(x)}{Q_X(x)} = D_\infty(P_X\|Q_X), \quad (215)$$

where (214) follows from the fact that $\frac{\sum_i a_i}{\sum_i b_i} \leq \max_i \frac{a_i}{b_i}$, for $b_i > 0$, $\forall i$. Taking supremum over all f, we get

$$\sup_{f:\mathcal{X}\to[0,\infty)} \log\frac{\mathbb{E}_{X\sim P_X}[f(X)]}{\mathbb{E}_{X\sim Q_X}[f(X)]} \leq \log\max_{x\in\mathcal{X}}\frac{P_X(x)}{Q_X(x)} \quad (216)$$

$$= D_\infty(P_X||Q_X). \quad (217)$$

This proves (42).

C. Proof of Corollary 4

The expression for opportunistic maximal g-leakage in (43) can be simplified as

$$\tilde{\mathcal{L}}_g^{\max}(X\to Y) = \log\sum_{y\in\text{supp}(Y)} P_Y(y) \sup_{U:U-X-Y} \frac{\sup_{P_{\hat{U}|Y=y}} \mathbb{E}_{U|Y=y}[g(P_{\hat{U}|Y}(U|y))]}{\sup_{P_{\hat{U}}} \mathbb{E}_U[g(P_{\hat{U}}(U))]} \quad (218)$$

$$= \log\sum_{y:P_Y(y)>0} P_Y(y)\max_{x:P_{X|Y}(x|y)>0}\frac{P_{X|Y}(x|y)}{P_X(x)} \quad (219)$$

$$= \log\sum_{y:P_Y(y)>0} P_Y(y)\max_{x:P_{X|Y}(x|y)>0}\frac{P_{Y|X}(y|x)}{P_Y(y)} \quad (220)$$

$$= \log\sum_{y:P_Y(y)>0}\max_{x:P_{X|Y}(x|y)>0} P_{Y|X}(y|x) \quad (221)$$

$$= \log\sum_{y:P_Y(y)>0}\max_{x:P_X(x)>0} P_{Y|X}(y|x) \quad (222)$$

$$= I_\infty^{\text{S}}(X;Y), \quad (223)$$

where (219) follows from Theorem 6 and (220) follows from the Bayes' Rule.

The expression for maximal realizable g-leakage in (44) can be simplified as

$$\begin{aligned}
&\mathcal{L}_{\alpha}^{\mathrm{r-max}}(X \to Y) \\
&= \sup_{U:U-X-Y} \log \max_{y \in \mathrm{supp}(Y)} \\
&\quad \frac{\sup_{P_{\hat{U}|Y=y}} \mathbb{E}\left[g(P_{\hat{U}|Y}(U|y))\right]}{\sup_{P_{\hat{U}}} \mathbb{E}\left[g(P_{\hat{U}}(U))\right]} \\
&= \log \max_{y \in \mathrm{supp}(Y)} \\
&\quad \sup_{U:U-X-Y} \frac{\sup_{P_{\hat{U}|Y=y}} \mathbb{E}\left[g(P_{\hat{U}|Y}(U|y))\right]}{\sup_{P_{\hat{U}}} \mathbb{E}\left[g(P_{\hat{U}}(U))\right]} \qquad (224) \\
&= \log \max_{y:P_Y(y)>0} \max_{x:P_{X|Y}(x|y)>0} \frac{P_{X|Y}(x|y)}{P_X(x)} \qquad (225) \\
&= \log \max_{(x,y):P_{XY}(x,y)>0} \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)} \qquad (226) \\
&= D_{\infty}(P_{XY} \| P_X \times P_Y), \qquad (227)
\end{aligned}$$

where (225) follows from Theorem 6.

Appendix D
Variational Characterization for $D_{\infty}(P_X\|Q_X)$ With Gain Function $g(t) = \log t$

Here we show that (37) holds for the non-positive gain function $g(t) = \log t$ that does not satisfy the conditions in Theorem 6. The proof of the lower bound follows exactly along the same lines as that of Theorem 6 with the only difference that (182) holds for negative gain functions too noticing that $\frac{\sum_i a_i}{\sum_i b_i} \leq \max_i \frac{a_i}{b_i}$, for $b_i < 0$, $\forall i$.

For the upper bound, we first note that

$$\begin{aligned}
\sup_{P_{\hat{U}}} \mathbb{E}_{U\sim P_U}\left[\log P_{\hat{U}}(U)\right] &= -\inf_{P_{\hat{U}}} \left(H_P(U) + D(P_U\|P_{\hat{U}})\right) \\
&= -H_P(U). \qquad (228)
\end{aligned}$$

We lower bound the RHS in (37) with gain function $g(t) = \log t$ by using the "shattering" $P_{U|X}$ [13, Proof of Theorem 1], [15, Proof of Theorem 5]. Let $\mathcal{U} = \uplus_{x\in\mathcal{X}} \mathcal{U}_x$, $|\mathcal{U}_x| = m_x$. Define $P_{U|X}(u|x) = \frac{1}{m_x}$, $u \in \mathcal{U}_x$. So, we have

$$\begin{aligned}
&\sup_{P_{U|X}} \frac{\sup_{P_{\hat{U}}} \mathbb{E}_{U\sim P_U}\left[\log P_{\hat{U}}(U)\right]}{\sup_{P_{\hat{U}}} \mathbb{E}_{U\sim Q_U}\left[\log Q_{\hat{U}}(U)\right]} \geq \frac{-H_P(U)}{-H_Q(U)} \qquad (229) \\
&= \frac{\sum_u \left(\sum_x P_X(x)P_{U|X}(u|x)\right) \log \left(\sum_x P_X(x)P_{U|X}(u|x)\right)}{\sum_u \left(\sum_x Q_X(x)P_{U|X}(u|x)\right) \log \left(\sum_x Q_X(x)P_{U|X}(u|x)\right)} \qquad (230) \\
&= \frac{\sum_x P_X(x) \log \frac{P_X(x)}{m_x}}{\sum_x Q_X(x) \log \frac{Q_X(x)}{m_x}} \qquad (231) \\
&= \frac{-H_P(X)/\log m_{x^*} - P_X(x^*)}{-H_Q(X)/\log m_{x^*} - Q_X(x^*)} \qquad (232) \\
&= \frac{P_X(x^*)}{Q_X(x^*)} \qquad (233) \\
&= 2^{D_{\infty}(P_X\|Q_X)}, \qquad (234)
\end{aligned}$$

where (232) follows by fixing an $x^* \in \arg\max_x \frac{P_X(x)}{Q_X(x)}$ and choosing $m_x = 1$, for $x \neq x^*$, and (233) then follows by taking limit $m_{x^*} \to \infty$.

Appendix E
Opportunistic Maximal and Maximal Realizable $(\alpha = 1)$-Leakages

We first note that opportunistic maximal α-leakage in LHS of (48) can be written as (see [13, Equations (18)-(20)])

$$\begin{aligned}
\tilde{\mathcal{L}}_{\alpha}^{\max}(X \to Y) &= \sup_U \frac{\alpha}{\alpha-1} \log \sum_{y\in\mathrm{supp}(Y)} P_Y(y) \\
&\quad \frac{\left(\sum_u \left(\sum_x P_{U|XY}(u|x,y)P_{X|Y}(x|y)\right)^{\alpha}\right)^{\frac{1}{\alpha}}}{\left(\sum_u \left(\sum_x P_{U|XY}(u|x,y)P_X(x)\right)^{\alpha}\right)^{\frac{1}{\alpha}}}. \qquad (235)
\end{aligned}$$

Let us define opportunistic maximal and maximal realizable $(\alpha = 1)$-leakages as

$$\begin{aligned}
\tilde{\mathcal{L}}_{1}^{\max}(X \to Y) &= \sup_U \lim_{\alpha\to 1} \frac{\alpha}{\alpha-1} \log \sum_{y\in\mathrm{supp}(Y)} P_Y(y) \\
&\quad \frac{\left(\sum_u \left(\sum_x P_{U|XY}(u|x,y)P_{X|Y}(x|y)\right)^{\alpha}\right)^{\frac{1}{\alpha}}}{\left(\sum_u \left(\sum_x P_{U|XY}(u|x,y)P_X(x)\right)^{\alpha}\right)^{\frac{1}{\alpha}}}, \qquad (236) \\
\mathcal{L}_{1}^{\mathrm{r-max}}(X \to Y) &= \sup_{U:U-X-Y} \max_{y\in\mathrm{supp(Y)}} \\
&\quad \lim_{\alpha\to 1} \frac{\alpha}{\alpha-1} \log \frac{\left(\sum_u P_{U|Y}(u|y)^{\alpha}\right)^{\frac{1}{\alpha}}}{\left(\sum_u P_U(u)^{\alpha}\right)^{\frac{1}{\alpha}}}, \qquad (237)
\end{aligned}$$

by taking the limit first and the supremum next. When X and Y are independent, note that the expressions for both the leakages above are equal to zero. The following proposition plays a crucial role in proving that these leakages are equal to infinity.

Proposition 3: Let P_{XY} be a probability distribution over a finite alphabet $\mathcal{X} \times \mathcal{Y}$ such that X and Y are not independent. We have

$$\sup_{U:U-X-Y} \left(H(U) - H(U|Y=y)\right) = \infty. \qquad (238)$$

Remark 10: It is easy to see that $\sup_{U:U-X-Y} \left(H(U) - H(U|Y)\right) = \sup_{U:U-X-Y} I(U;Y) = I(X;Y)$, since we have $I(U;Y) \leq I(X;Y)$ for every U such that $U - X - Y$ by data processing inequality. However, if we replace the conditional entropy in the objective function with a conditional entropy where the conditioning is on a particular realization of Y (instead of the random variable itself), it is interesting that the supremum blows up to infinity.

Proof: For a fixed $Y = y$, we have

$$\begin{aligned}
&\sup_{U:U-X-Y} \left(H(U) - H(U|Y=y)\right) \\
&\geq \sup_{U:U-X-Y, H(X|U)=0} \left(H(U) - H(U|Y=y)\right). \qquad (239)
\end{aligned}$$

In the RHS of (239), we further assume that $\mathcal{U} = \bigcup_{x\in\mathcal{X}} \mathcal{U}_x$ be the alphabet of U such that

$$P_{U|X}(u|x) = \begin{cases} \frac{1}{n_x}, & u \in \mathcal{U}_x \\ 0, & \text{otherwise}, \end{cases} \qquad (240)$$

where $n_x = |\mathcal{U}_x|$ is to be fixed later. This together with the Markov chain $U - X - Y$ gives

$$P_U(u) = \sum_x P_X(x) P_{U|X}(u|x) = \frac{P_X(x)}{n_x}, \text{ for } u \in \mathcal{U}_x, \tag{241}$$

$$P_{U|Y}(u|y) = \sum_x P_{X|Y}(x|y) P_{U|X}(u|x) = \frac{P_{X|Y}(x|y)}{n_x}, \quad \text{for } u \in \mathcal{U}_x. \tag{242}$$

Continuing (239), we get

$$\sup_{U:U-X-Y} (H(U) - H(U|Y=y)) \tag{243}$$

$$\geq -\sum_{x\in\mathcal{X}} \sum_{u\in\mathcal{U}_x} P_U(u) \log P_U(u) + \sum_{x\in\mathcal{X}} \sum_{u\in\mathcal{U}_x} P_{U|Y}(u|y) \log P_{U|Y}(u|y) \tag{244}$$

$$= -\sum_{x\in\mathcal{X}} \sum_{u\in\mathcal{U}_x} \frac{P_X(x)}{n_x} \log\left(\frac{P_X(x)}{n_x}\right) + \sum_{x\in\mathcal{X}} \sum_{u\in\mathcal{U}_x} \frac{P_{X|Y}(x|y)}{n_x} \log\left(\frac{P_{X|Y}(x|y)}{n_x}\right) \tag{245}$$

$$= -\sum_{x\in\mathcal{X}} P_X(x) \log\left(\frac{P_X(x)}{n_x}\right) + \sum_{x\in\mathcal{X}} P_{X|Y}(x|y) \log\left(\frac{P_{X|Y}(x|y)}{n_x}\right) \tag{246}$$

$$= -\sum_x P_X(x) \log P_X(x) + \sum_x P_X(x) \log n_x + \sum_x P_{X|Y}(x|y) \log P_{X|Y}(x|y) - \sum_x P_{X|Y}(x|y) \log n_x \tag{247}$$

$$= H(X) - H(X|Y=y) + \sum_x \log n_x \left(P_X(x) - P_{X|Y}(x|y)\right) \tag{248}$$

where (245) follows from (241) and (241), and (246) follows because $|\mathcal{U}_x| = n_x$. Note that $1 \leq n_x \leq \infty$, for every $x \in \mathcal{X}$. Now choosing

$$n_x = \begin{cases} 1, & \text{for } x \text{ s.t. } P_X(x) \leq P_{X|Y}(x|y) \\ \infty, & \text{for } x \text{ s.t. } P_X(x) > P_{X|Y}(x|y), \end{cases} \tag{249}$$

implies that the summation term in (248) equals infinity. □

Then we have the following theorems.

Theorem 8: Let P_{XY} be a probability distribution over a finite alphabet $\mathcal{X} \times \mathcal{Y}$ such that X and Y are not independent. We have

$$\tilde{\mathcal{L}}_1^{\max}(X \to Y) = \infty. \tag{250}$$

Proof: Using L'Hospital's rule and the fact that,

$$\lim_{\alpha\to 1} \frac{d}{d\alpha} (a^\alpha + b^\alpha)^{\frac{1}{\alpha}} = a \log a + b \log b - (a+b) \log (a+b), \tag{251}$$

for $a, b \geq 0$, we can verify that, for $p_i, a_i, b_i, c_i, d_i \geq 0$, and $a_i + b_i = c_i + d_i = 1$, $i \in [1:t]$,

$$\lim_{\alpha\to 1} \frac{\alpha}{\alpha - 1} \log\left(\sum_{i=1}^t p_i \frac{(a_i^\alpha + b_i^\alpha)^{\frac{1}{\alpha}}}{(c_i^\alpha + d_i^\alpha)^{\frac{1}{\alpha}}}\right) = \sum_{i=1}^t p_i \left(a_i \log a_i + b_i \log b_i - c_i \log c_i - d_i \log d_i\right). \tag{252}$$

Now we have

$$\tilde{\mathcal{L}}_1^{\max}(X \to Y) = \sup_U \lim_{\alpha\to 1} \frac{\alpha}{\alpha - 1} \log \sum_{y\in\text{supp}(Y)} P_Y(y) \frac{\left(\sum_u \left(\sum_x P_{U|XY}(u|x,y) P_{X|Y}(x|y)\right)^\alpha\right)^{\frac{1}{\alpha}}}{\left(\sum_u \left(\sum_x P_{U|XY}(u|x,y) P_X(x)\right)^\alpha\right)^{\frac{1}{\alpha}}} \tag{253}$$

$$= \sup_U \sum_{y\in\text{supp}(Y)} \sum_u \Big(P_{U|Y}(u|y) \log(P_{U|Y}(u|y)) - \tilde{P}_{U|Y}(u|y) \log\left(\tilde{P}_{U|Y}(u|y)\right)\Big), \tag{254}$$

where (254) follows from (252) with

$$P_{U|Y}(u|y) = \sum_x P_{U|XY}(u|x,y) P_{X|Y}(x|y), \tag{255}$$

$$\tilde{P}_{U|Y}(u|y) = \sum_x P_{U|XY}(u|x,y) P_X(x). \tag{256}$$

Continuing (254), we get

$$\tilde{\mathcal{L}}_1^{\max}(X \to Y) = \sup_U \sum_{y\in\text{supp}(Y)} P_Y(y) \sum_u \Big(P_{U|Y}(u|y) \log\left(P_{U|Y}(u|y)\right) - \tilde{P}_{U|Y}(u|y) \log\left(\tilde{P}_{U|Y}(u|y)\right)\Big) \tag{257}$$

$$= \sup_{(U_y : y\in\mathcal{Y}) - X - Y} \sum_{y\in\text{supp}(Y)} P_Y(y) \sum_u \left(P_{U_y|Y}(u|y) \log P_{U_y|Y}(u|y)) - P_{U_y}(u) \log P_{U_y}(u)\right) \tag{258}$$

$$= \sum_{y\in\text{supp}(Y)} P_Y(y) \sup_{U_y : U_y - X - Y} \sum_u \left(P_{U_y|Y}(u|y) \log P_{U_y|Y}(u|y)) - P_{U_y}(u) \log P_{U_y}(u)\right) \tag{259}$$

$$= \sum_{y\in\text{supp}(Y)} P_Y(y) \sup_{U:U-X-Y} \sum_u \left(P_{U|Y}(u|y) \log P_{U|Y}(u|y)) - P_U(u) \log P_U(u)\right) \tag{260}$$

$$= \sum_{y\in\text{supp}(Y)} P_Y(y) \sup_{U:U-X-Y} (H(U) - H(U|Y=y)) \tag{261}$$

$$= \infty, \tag{262}$$

where (262) follows from Proposition 3. □

Theorem 9: Let P_{XY} be a probability distribution over a finite alphabet $\mathcal{X}\times\mathcal{Y}$ such that X and Y are not independent. We have

$$\mathcal{L}_1^{\mathrm{r-max}}(X\to Y)=\infty. \tag{263}$$

Proof: Consider

$$\mathcal{L}_1^{\mathrm{r-max}}(X\to Y)$$

$$:=\sup_{U:U-X-Y}\max_{y\in\mathrm{supp}(Y)}\lim_{\alpha\to1}\frac{\alpha}{\alpha-1}\log\frac{\left(\sum_u P_{U|Y}(u|y)^\alpha\right)^{\frac{1}{\alpha}}}{\left(\sum_u P_U(u)^\alpha\right)^{\frac{1}{\alpha}}} \tag{264}$$

$$=\sup_{U:U-X-Y}\max_{y\in\mathrm{supp}(Y)}\lim_{\alpha\to1}\left(\frac{\alpha}{\alpha-1}\log\left(\sum_u P_{U|Y}(u|y)^\alpha\right)^{\frac{1}{\alpha}}-\frac{\alpha}{\alpha-1}\log\left(\sum_u P_U(u)^\alpha\right)^{\frac{1}{\alpha}}\right) \tag{265}$$

$$=\sup_{U:U-X-Y}\max_{y\in\mathrm{supp}(Y)}\lim_{\alpha\to1}\left(H_\alpha(U)-H_\alpha(U|Y-y)\right) \tag{266}$$

$$=\sup_{U:U-X-Y}\max_{y\in\mathrm{supp}(Y)}\left(H(U)-H(U|Y=y)\right) \tag{267}$$

$$=\max_{y\in\mathrm{supp}(Y)}\sup_{U:U-X-Y}\left(H(U)-H(U|Y=y)\right) \tag{268}$$

$$=\infty, \tag{269}$$

where (269) follows from Proposition 3. □

We remark that Theorem 9 also follows from Theorem 8 by noticing that $\mathcal{L}^{r-\max}(X\to Y)\geq\tilde{\mathcal{L}}^{\max}(X\to Y)$.

Appendix F
Bregman Divergence

Let $F:\Omega\to\mathbb{R}$ be a continuously differentiable, strictly convex function on a closed, convex set Ω. The Bregman divergence associated with F [59] for points $p,q\in\Omega$ is the difference between the value of F at p and the value of the first-order Taylor expansion of F around q evaluated at point p, i.e.,

$$B_F(p,q)=F(p)-F(q)-\langle\nabla F(q),p-q\rangle. \tag{270}$$

Let $p=(p_1,\ldots,p_d)$ and $q=(q_1,\ldots,q_d)$ be two discrete probability distributions. Consider $F(p)=\frac{\alpha}{\alpha-1}\left(\left(\sum_i p_i^\alpha\right)^{\frac{1}{\alpha}}-1\right)$ which is a strictly convex function on the d-simplex. The associated Bregman divergence is given by

$$B_F(p,q)=F(p)-F(q)-\langle\nabla F(q),p-q\rangle \tag{271}$$

$$=\frac{\alpha}{\alpha-1}\left[\left(\sum_i p_i^\alpha\right)^{\frac{1}{\alpha}}-1-\left(\sum_i q_i^\alpha\right)^{\frac{1}{\alpha}}+1-\sum_i(p_i-q_i)\left(\sum_j q_j^\alpha\right)^{\frac{1-\alpha}{\alpha}}q_i^{\alpha-1}\right] \tag{272}$$

$$=\frac{\alpha}{\alpha-1}\left[\left(\sum_i p_i^\alpha\right)^{\frac{1}{\alpha}}-\left(\sum_j q_j^\alpha\right)^{\frac{1-\alpha}{\alpha}}\times\left(\sum_i q_i^\alpha+\sum_i p_i q_i^{\alpha-1}-\sum_i q_i^\alpha\right)\right] \tag{273}$$

$$=\frac{\alpha}{\alpha-1}\left[\left(\sum_i p_i^\alpha\right)^{\frac{1}{\alpha}}-\left(\sum_j q_j^\alpha\right)^{\frac{1-\alpha}{\alpha}}\sum_i p_i q_i^{\alpha-1}\right] \tag{274}$$

$$=\frac{\alpha}{\alpha-1}\left[\left(\sum_i p_i^\alpha\right)^{\frac{1}{\alpha}}-\left(\sum_i p_i^\alpha\right)^{\frac{1}{\alpha}}\times\sum_i\left(\frac{p_i^\alpha}{\sum_j p_j^\alpha}\right)^{\frac{1}{\alpha}}\left(\frac{q_i^\alpha}{\sum_j q_j^\alpha}\right)^{1-\frac{1}{\alpha}}\right] \tag{275}$$

$$=\frac{\alpha}{\alpha-1}\left[\left(\sum_i p_i^\alpha\right)^{\frac{1}{\alpha}}\left(1-\mathrm{e}^{\frac{1-\alpha}{\alpha}D_{\frac{1}{\alpha}}(p^{(\alpha)}||q^{(\alpha)})}\right)\right], \tag{276}$$

where $p^{(\alpha)}:=\left(\frac{p_i^\alpha}{\sum_j p_j^\alpha}:i\in[1:d]\right)$ and $q^{(\alpha)}:=\left(\frac{q_i^\alpha}{\sum_j q_j^\alpha}:i\in[1:d]\right)$ are tilted versions of the distributions p and q, respectively. Thus the associated Bregman divergence is related to the Rényi divergence of order $\frac{1}{\alpha}$ in the manner

$$B_F(p,q)=\frac{\alpha}{\alpha-1}\left[\left(\sum_i p_i^\alpha\right)^{\frac{1}{\alpha}}\left(1-\mathrm{e}^{\frac{1-\alpha}{\alpha}D_{\frac{1}{\alpha}}(p^{(\alpha)}||q^{(\alpha)})}\right)\right]. \tag{277}$$

Note that $\lim_{\alpha\to1}F(p)=\sum_{i=1}^d p_i\log p_i$ and the resulting Bregman divergence is equal to relative entropy.

Appendix G
Proofs of Lemma 2

Let $A=\{x:\tilde{P}_X(x)>0\}$. Then

$$-D(\tilde{P}_X||\tilde{Q}_X)=\sum_{x\in A}\tilde{P}_X(x)\log\left(\frac{\tilde{Q}_X(x)}{\tilde{P}_X(x)}\right) \tag{278}$$

$$\leq\log\left(\sum_{x\in A}\tilde{P}_X(X)\frac{\tilde{Q}_X(x)}{\tilde{P}_X(x)}\right) \tag{279}$$

$$=\log\left(\sum_{x\in A}\tilde{Q}_X(x)\right) \tag{280}$$

$$\leq\log(1) \tag{281}$$

$$=0, \tag{282}$$

where (279) follows by Jensen's inequality, (281) follows since $\sum_x\tilde{Q}_X(x)\leq1$.

Appendix H
Proof of Lemma 5

Let $\mathcal{X}=\{x_1,x_2,\ldots,x_n\}$ and $P_X(x_i)=p_i$, for $i\in[1:n]$. Consider $a_{[1:k]}$ such that $a_i=a_j$ for some $i\neq j$. There exists a $b_{[1:k]}$ such that for each $i\in[1:k]$, we have $a_i=b_j$ for some j and $b_r\neq a_j$ for some r and any j. Consider

$$\frac{\alpha}{\alpha-1}\left[\sum_{i=1}^n p_i\left(1-\mathrm{P}^*\left(\bigcup_{j=1}^k(\hat{X}_j=x_i)\right)^{\frac{\alpha-1}{\alpha}}\right)\right]. \tag{283}$$

Let $\mathcal{A}$ and $\mathcal{B}$ denote the sets of all multiset permutations of $a_{[1:k]}$ and $b_{[1:k]}$, respectively, when $a_{[1:k]}$ and $b_{[1:k]}$ are treated as multisets. Let $q_{a_1,a_2,\ldots,a_k}:=\sum_{r_{[1:k]}\in\mathcal{A}}P_{\hat{X}_{[1:k]}}(r_{[1:k]})$ and $q_{b_1,b_2,\ldots,b_k}:=\sum_{r_{[1:k]}\in\mathcal{B}}P_{\hat{X}_{[1:k]}}(r_{[1:k]})$. Each term out of the n terms in (283) will either contain both $q_{a_{[1:k]}}$ and $q_{b_{[1:k]}}$ (say, type 1), contain just $q_{b_{[1:k]}}$ alone (say, type 2), or does not contain both (say, type 3). We now construct a new strategy $P_{\hat{X}_{[1:k]}}$ by incorporating the value of $q_{a_{[1:k]}}$ into $q_{b_{[1:k]}}$ making the value of new $q_{a_{[1:k]}}$ equal to zero. Now

the values of the terms of type 2 strictly decrease as the α-loss function is strictly decreasing in its argument while retaining the values of the terms of types 1 and 3. This leads to a contradiction since $P^*_{X_{[1:k]}}$ is assumed to be an optimal strategy. So, $P_{\hat{X}_{[1:k]}}(a_{[1:k]}) = 0$. Repeating the same argument as above for all such $a_{[1:k]}$ s.t. $a_i = a_j$, for some $i \neq j$ completes the proof.

APPENDIX I
PROOF OF LEMMA 6

'Only if' Part: Suppose a vector $(t_1, t_2, \ldots, t_n)$ is admissible. Then there exists $P_{\hat{X}_{[1:k]}}$ satisfying (142). Using (141), since t_i is probability of a certain event, we have

$$0 \leq t_i \leq 1, \text{ for } i \in [1:n].$$

'If' Part: Suppose $0 \leq t_i \leq 1$, for $i \in [1:n]$. Summing up all the equations in (142) over $i \in [1:n]$ and using $\sum_{i=1}^n t_i = k$, we get

$$P_{\hat{X}_{[1:k]}}(a_{[1:k]}) = 0, \text{ for all } a_{[1:k]} \text{ s.t. } a_i = a_j, \text{ for some } i \neq j.$$

With this, (142) can be written in the form of system of linear equation only in terms of non-negative variables of the form

$$q_{i_1,i_2,\ldots,i_k} := \sum_{\sigma \in S_n} P_{\hat{X}_{[1:k]}}(x_{i_{\sigma(1)}}, x_{i_{\sigma(2)}}, \ldots, x_{i_{\sigma(n)}}), \quad (284)$$

where $i_1, i_2, \ldots, i_k$ are all distinct and belong to $[1:n]$. Here the sum is computed over all the permutations σ of the set $\{1, 2, \ldots, n\}$. The set of all such permutations is denoted by S_n. With this, the system of equations in (142) can be written in the form $AQ = b$, $Q \geq 0$. Here A is a $n \times \binom{n}{k}$-matrix, where the rows are indexed by $i \in [1:n]$ and columns are indexed by $(i_1, i_2, \ldots, i_k)$, where $i_1, i_2, \ldots, i_k$ are all distinct and belong to $[1:n]$. In particular, in the column indexed by $(i_1, i_2, \ldots, i_k)$, the entry of A corresponding to i_j^{th} row is 1, for $j \in [1:k]$. All the remaining entries of the matrix A are zeros. Q is $\binom{n}{k}$-length vector of variables of the form $q_{i_1,i_2,\ldots,i_k}$. b is an n-length vector with $b_i = t_i$. We are interested in the feasibility of the system $AQ = b$, $Q \geq 0$. We use the Farkas' lemma [60, Proposition 6.4.3] in linear programming for checking this. It states that the system $AQ = b$ has a non-negative solution if and only if every $y \in \mathbb{R}^n$ with $y^\top A \geq 0$ also implies $y^\top b \geq 0$. For our problem, $y^\top A \geq 0$ is equivalent to

$$\sum_{j=1}^k y_{i_j} \geq 0, \text{ for all distinct } i_1, i_2, \ldots, i_k \in [1:n]. \quad (285)$$

Without loss of generality, let us assume that $y_i \leq y_{i+1}$, $i \in [1:n-1]$. Then (285) is equivalent to

$$\sum_{i=1}^k y_i \geq 0. \quad (286)$$

Now consider

$$\sum_{i=1}^n y_i t_i$$

$$= \sum_{i=1}^k y_i t_i + y_{k+1} t_{k+1} + \sum_{i=k+2}^n y_i t_i \quad (287)$$

$$= \sum_{i=1}^k y_i + \sum_{i=1}^k y_i (t_i - 1) + y_{k+1} t_{k+1} + \sum_{i=k+2}^n y_i t_i \quad (288)$$

$$\geq \sum_{i=1}^k y_i + y_{k+1} \sum_{i=1}^k (t_i - 1) + y_{k+1} t_{k+1} + \sum_{i=k+2}^n y_i t_i \quad (289)$$

$$\geq \sum_{i=1}^k y_i + y_{k+1} \sum_{i=1}^k (t_i - 1) + y_{k+1} t_{k+1} + y_{k+1} \sum_{i=k+2}^n t_i \quad (290)$$

$$= \sum_{i=1}^k y_i + y_{k+1} \left(\sum_{i=1}^n t_i - k \right) \quad (291)$$

$$= \sum_{i=1}^k y_i \quad (292)$$

$$\geq 0, \quad (293)$$

where (289) follows because $y_i \leq y_{k+1}$ and $t_i - 1 \leq 0$, for $i \in [1:k]$, (290) follows because $y_i \geq y_{k+1}$, for $i \in [k+2:n]$, and (292) follows because $\sum_{i=1}^n t_i = k$, (293) follows from (286). Now using the Farkas' lemma, $AQ = b$, has a non-negative solution, i.e., the vector $(t_1, t_2, \ldots, t_n)$ is admissible.

ACKNOWLEDGMENT

The author Gowtham R. Kurri would like to thank Tyler Sypherd for helpful discussions on Bregman divergence and its connections to Rényi divergence in Lemma 1 motivated via α-loss.

REFERENCES

[1] C. Dwork, "Differential privacy," in *Proc. Int. Colloq. Automata, Lang., Program.*, M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, Eds. 2006, pp. 1–12.

[2] G. Smith, "On the foundations of quantitative information flow," in *Proc. Int. Conf. Found. Softw. Sci. Comput. Struct.*, 2009, pp. 288–302.

[3] C. Braun, K. Chatzikokolakis, and C. Palamidessi, "Quantitative notions of leakage for one-try attacks," *Electron. Notes Theor. Comput. Sci.*, vol. 249, pp. 75–91, Aug. 2009.

[4] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, "What can we learn privately?" *SIAM J. Comput.*, vol. 40, no. 3, pp. 793–826, Jan. 2011.

[5] M. S. Alvim, K. Chatzikokolakis, C. Palamidessi, and G. Smith, "Measuring information leakage using generalized gain functions," in *Proc. IEEE 25th Comput. Secur. Found. Symp.*, Jun. 2012, pp. 265–279.

[6] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy and statistical minimax rates," in *Proc. 51st Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2013, p. 1592.

[7] M. S. Alvim, K. Chatzikokolakis, A. Mciver, C. Morgan, C. Palamidessi, and G. Smith, "Additive and multiplicative notions of leakage, and their capacities," in *Proc. IEEE 27th Comput. Secur. Found. Symp.*, Jul. 2014, pp. 308–322.

[8] N. Merhav and E. Arikan, "The Shannon cipher system with a guessing wiretapper," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 1860–1866, 1999.

[9] F. D. P. Calmon and N. Fawaz, "Privacy against statistical inference," in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2012, pp. 1401–1408.

[10] C. Schieler and P. Cuff, "The henchman problem: Measuring secrecy by the minimum distortion in a list," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2014, pp. 596–600.

[11] I. Issa and A. B. Wagner, "Measuring secrecy by the probability of a successful guess," in *Proc. 53rd Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2015, pp. 980–987.

[12] S. Asoodeh, F. Alajaji, and T. Linder, "On maximal correlation, mutual information and data privacy," in *Proc. IEEE 14th Can. Workshop Inf. Theory (CWIT)*, Jul. 2015, pp. 27–31.

[13] I. Issa, A. B. Wagner, and S. Kamath, "An operational approach to information leakage," *IEEE Trans. Inf. Theory*, vol. 66, no. 3, pp. 1625–1657, Mar. 2020.

[14] B. Rassouli and D. Gündüz, "Optimal utility-privacy trade-off with total variation distance as a privacy measure," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 594–603, 2020.

[15] J. Liao, L. Sankar, O. Kosut, and F. P. Calmon, "Maximal α-leakage and its properties," in *Proc. IEEE Conf. Commun. Netw. Secur. (CNS)*, Jun. 2020, pp. 1–6.

[16] S. Saeidian, G. Cervia, T. J. Oechtering, and M. Skoglund, "Pointwise maximal leakage," *IEEE Trans. Inf. Theory*, vol. 69, no. 12, pp. 8054–8080, Dec. 2023.

[17] I. Issa, "An operational approach to information leakage," Ph.D. dissertation, Dept. Elect. Comput. Eng., Cornell Univ., Ithaca, NY, USA, 2017.

[18] C. Qian, R. Zhou, C. Tian, and T. Liu, "Improved weakly private information retrieval codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2022, pp. 2827–2832.

[19] J. Liao, L. Sankar, F. P. Calmon, and V. Y. F. Tan, "Hypothesis testing under maximal leakage privacy constraints," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 779–783.

[20] Y. Liu, L. Ong, S. Johnson, J. Kliewer, P. Sadeghi, and P. L. Yeoh, "Information leakage in zero-error source coding: A graph-theoretic perspective," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021, pp. 2590–2595.

[21] S. Saeidian, G. Cervia, T. J. Oechtering, and M. Skoglund, "Quantifying membership privacy via information leakage," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 3096–3108, 2021.

[22] N. Sathyavageeswaran, R. D. Yates, A. D. Sarwate, and N. Mandayam, "Privacy leakage in discrete-time updating systems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2022, pp. 2076–2081.

[23] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math. Statist. Probab.*, vol. 1. 1961, pp. 547–561.

[24] O. Shayevitz, "On Rényi measures and hypothesis testing," in *Proc. IEEE Int. Symp. Inf. Theory*, 2011, pp. 894–898.

[25] T. van Erven and P. Harremos, "Rényi divergence and Kullback–Leibler divergence," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3797–3820, Jul. 2014.

[26] I. Sason, "On the Rényi divergence, joint range of relative entropies, and a channel coding theorem," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 23–34, Jan. 2016.

[27] V. Anantharam, "A variational characterization of Rényi divergences," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 6979–6989, 2018.

[28] J. Birrell, P. Dupuis, M. A. Katsoulakis, L. Rey-Bellet, and J. Wang, "Variational representations and neural network estimation of Rényi divergences," *SIAM J. Math. Data Sci.*, vol. 3, no. 4, pp. 1093–1116, Jan. 2021.

[29] S. Arimoto, "Information-theoretical considerations on estimation problems," *Inf. Control*, vol. 19, no. 3, pp. 181–194, Oct. 1971.

[30] J. Liao, O. Kosut, L. Sankar, and F. D. P. Calmon, "Tunable measures for information leakage and applications to privacy-utility tradeoffs," *IEEE Trans. Inf. Theory*, vol. 65, no. 12, pp. 8043–8066, Dec. 2019.

[31] T. Sypherd, M. Diaz, L. Sankar, and P. Kairouz, "A tunable loss function for binary classification," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2019, pp. 2479–2483.

[32] F. P. Calmon, M. Varia, M. Médard, M. M. Christiansen, K. R. Duffy, and S. Tessaro, "Bounds on inference," in *Proc. 51st Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2013, pp. 567–574.

[33] C. T. Li and A. El Gamal, "Maximal correlation secrecy," *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3916–3926, May 2018.

[34] C. E. Shannon, "Communication theory of secrecy systems," *Bell Syst. Tech. J.*, vol. 28, no. 4, pp. 656–715, Oct. 1949.

[35] V. Prabhakaran and K. Ramchandran, "On secure distributed source coding," in *Proc. IEEE Inf. Theory Workshop*, Sep. 2007, pp. 442–447.

[36] H. Tyagi, P. Narayan, and P. Gupta, "When is a function securely computable?" *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6337–6350, Oct. 2011.

[37] L. Sankar, S. R. Rajagopalan, and H. V. Poor, "Utility-privacy tradeoffs in databases: An information-theoretic approach," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 6, pp. 838–852, Jun. 2013.

[38] W. Wang, L. Ying, and J. Zhang, "On the relation between identifiability, differential privacy, and mutual-information privacy," *IEEE Trans. Inf. Theory*, vol. 62, no. 9, pp. 5018–5029, Sep. 2016.

[39] S. Kamel, M. Sarkiss, M. Wigger, and G. Rekaya-Ben Othman, "Secrecy capacity-memory tradeoff of erasure broadcast channels," *IEEE Trans. Inf. Theory*, vol. 65, no. 8, pp. 5094–5124, Aug. 2019.

[40] S. Asoodeh, M. Diaz, F. Alajaji, and T. Linder, "Privacy-aware guessing efficiency," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 754–758.

[41] Z. Li, T. J. Oechtering, and D. Gündüz, "Privacy against a hypothesis testing adversary," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 6, pp. 1567–1581, Jun. 2019.

[42] B. Jiang, M. Seif, R. Tandon, and M. Li, "Context-aware local information privacy," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 3694–3708, 2021.

[43] P. Cuff and L. Yu, "Differential privacy as a mutual information constraint," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2016, pp. 43–54.

[44] F. McSherry, "Privacy integrated queries: An extensible platform for privacy-preserving data analysis," *Commun. ACM*, vol. 53, no. 9, pp. 89–97, Sep. 2010.

[45] J. Soria-Comas and J. Domingo-Ferrert, "Differential privacy via t-closeness in data publishing," in *Proc. 11th Annu. Conf. Privacy, Secur. Trust*, Jul. 2013, pp. 27–35.

[46] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *Proc. Annu. Int. Conf. Theory Appl. Cryptograph. Techn.*, 2006, pp. 486–503.

[47] I. Mironov, "Rényi differential privacy," in *Proc. IEEE Comput. Secur. Found. Symp. (CSF)*, 2017, pp. 263–275.

[48] I. Wagner and D. Eckhoff, "Technical privacy metrics: A systematic survey," *ACM Comput. Surveys*, vol. 51, no. 3, pp. 1–38, May 2019.

[49] M. Bloch et al., "An overview of information-theoretic security and privacy: Metrics, limits and applications," *IEEE J. Sel. Areas Inf. Theory*, vol. 2, no. 1, pp. 5–22, Mar. 2021.

[50] H. Hsu, N. Martinez, M. Bertran, G. Sapiro, and F. P. Calmon, "A survey on statistical, information, and estimation—Theoretic views on privacy," *IEEE BITS Inf. Theory Mag.*, vol. 1, no. 1, pp. 45–56, Sep. 2021.

[51] R. Sibson, "Information radius," *Zeitschrift Für Wahrscheinlichkeitstheorie Und Verwandte Gebiete*, vol. 14, no. 2, pp. 149–160, 1969.

[52] G. R. Kurri, O. Kosut, and L. Sankar, "Evaluating multiple guesses by an adversary via a tunable loss function," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021, pp. 2002–2007.

[53] G. R. Kurri, O. Kosut, and L. Sankar, "A variational formula for infinity-Rényi divergence with applications to information leakage," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2022, pp. 2493–2498.

[54] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997.

[55] N. Merhav and M. Feder, "Universal prediction," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2124–2147, Oct. 1998.

[56] X. Nguyen, M. J. Wainwright, and M. I. Jordan, "On surrogate loss functions and f-divergences," *Ann. Statist.*, vol. 37, no. 2, pp. 876–904, Apr. 2009.

[57] T. A. Courtade and R. D. Wesel, "Multiterminal source coding with an entropy-based distortion measure," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2011, pp. 2040–2044.

[58] P. L. Bartlett, M. I. Jordan, and J. D. McAuliffe, "Convexity, classification, and risk bounds," *J. Amer. Stat. Assoc.*, vol. 101, no. 473, pp. 138–156, Mar. 2006.

[59] L. M. Bregman, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming," *USSR Comput. Math. Math. Phys.*, vol. 7, no. 3, pp. 200–217, Jan. 1967.

[60] J. Matousek and B. Gartner, *Understanding and Using Linear Programming*. Berlin, Germany: Springer, 2007.

[61] W. Huleihel, S. Salamatian, and M. Médard, "Guessing with limited memory," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 2253–2257.

[62] R. Atar, K. Chowdhary, and P. Dupuis, "Robust bounds on risk-sensitive functionals via Rényi divergence," *SIAM/ASA J. Uncertainty Quantification*, vol. 3, no. 1, pp. 18–33, Jan. 2015.

[63] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004, pp. 41–52.

[64] S. Arimoto, "Information measures and capacity of order α for discrete memoryless channels," *Topics Inf. Theory*, 1977.

Gowtham R. Kurri (Member, IEEE) received the B.Tech. degree in electronics and communication engineering from the International Institute of Information Technology (IIIT), Hyderabad, India, in 2011, and the M.Sc. and Ph.D. degrees from the Tata Institute of Fundamental Research, Mumbai, India, in 2020. From 2011 to 2012, he was an Associate Engineer with Qualcomm India Pvt. Ltd., Hyderabad, India. From July 2019 to October 2019, he was a Research Intern with the Blockchain Technology Group, IBM Research, Bengaluru, India. From 2020 to 2023, he was a Post-Doctoral Researcher with the School of Electrical, Computer and Energy Engineering, Arizona State University. Since February 2023, he has been an Assistant Professor with the IIIT Hyderabad, where he is affiliated with the Signal Processing and Communications Research Centre. His research interests include information theory and statistical machine learning.

Lalitha Sankar (Senior Member, IEEE) received the B.Tech. degree from the Indian Institute of Technology Bombay, the M.S. degree from the University of Maryland, and the Ph.D. degree from Rutgers University. She is currently a Professor with the School of Electrical, Computer, and Energy Engineering, Arizona State University. Her research interests include applying information theory and data science to study reliable, responsible, and privacy-protected machine learning, cyber security, and resilience in critical infrastructure networks. She received the National Science Foundation CAREER Award in 2014, the IEEE Globecom 2011 Best Paper Award for her work on the privacy of side information in multi-user data systems, and the Academic Excellence Award from Rutgers in 2008.

Oliver Kosut (Senior Member, IEEE) received the B.S. degree in electrical engineering and mathematics from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2004, and the Ph.D. degree in electrical and computer engineering from Cornell University, Ithaca, NY, USA, in 2010.

From 2010 to 2012, he was a Post-Doctoral Research Associate with MIT. Since 2012, he has been a Faculty Member with the School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ, USA, where he is currently an Associate Professor. His research interests include information theory—particularly with applications to security and machine learning—and power systems.

Prof. Kosut received the NSF CAREER Award in 2015. He is an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY. He is an IEEE Information Theory Society Distinguished Lecturer from 2023 to 2024.