Critical Review on the Effect of Different Metrics in Information Theoretic Privacy

Matin Mortaheb, *Member, IEEE,* and Cemil Vahapoglu, *Member, IEEE,*

Abstract

In the context of privacy, there is a trade-off between the utility and privacy of the published dataset. To obtain better privacy-utility tradeoff, two randomized-based approaches have been proposed in literature. (1) is context-free approaches that assume worst-case dataset statistics and adversaries. The most famous context-free approach is Differential Privacy (DP). (2) is context-aware approaches that explicitly model the dataset statistics and adversary's capabilities. Compared to the context-free privacy notions, context-aware privacy notions achieve a better privacy-utility tradeoff by incorporating the statistics of the dataset and placing reasonable restrictions on the capabilities of the adversary. Our goal for the project is to do a critical review on context-aware methods which is also known as information theoretic privacy. In the information theoretic method, utility and privacy constraints can be defined with various metrics such as mutual information, minimum mean square error, and maximal correlation. We tried to critically investigate the advantages and deficiencies of each metrics by comparing their lower-bound and privacy leakage separately for discrete and continuous case. Our review shows that mutual information sacrifices privacy to reach better utility. On the other hand, maximal correlation will behave better from the privacy point of view.

Index Terms

Information Theoretic Privacy, Mutual Information, Maximal correlation, MMSE.

I. Introduction

Consider two communicating agents Alice and Bob. To receive a payoff, Alice observes a random variable Y and wants to reveal it to Bob. On the other hand, nature chooses Y, dependent on X via a fixed channel $P_{Y|X}$. Alice wishes to disclose Y as accurately as possible, but in such a way that X is kept almost private from Bob. For instance, Y may represent the information that a social network (Alice) obtains from its users and X may represent political preferences of the users. Alice wants to disclose Y as accurately as possible to an advertising company and, simultaneously, wishes to protect the privacy of its users. Given a fixed joint distribution P_{XY}, Alice, hence, needs to choose a random mapping $P_{Z|Y}$, the so-called privacy filter, to release a new random variable Z, called the displayed data, such that X and Z satisfy a privacy constraint and Z maximizes a utility function (corresponding to the predictability of Y). More precisely, we seek to design a randomized mechanism M which maps Y to an auxiliary random variable Z such that the information leakage from X to Z is limited, and the "estimation efficiency" of Y given Z is maximal.

We assume throughout the project that X, Y, and Z form a Markov chain in that order, denoted by X–o–Y–o–Z. The system block diagram of this model is given in Fig. 1.

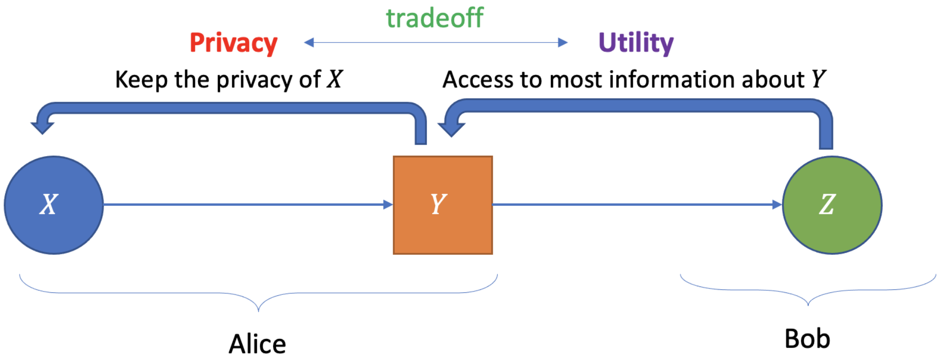

Figure 1. Information-theoretic Privacy.

Trade-off between the utility and privacy can be characterized as a following constrained optimization problem, denoted as *rate-privacy* function.

$$\sup_{P_{Z|Y}:\mathcal{L}(X\to Z)\le\epsilon} \mathcal{L}(Y\to Z) \tag{1}$$

M. Mortaheb and C. Vahapoglu are with the Department of Electrical and Computer Engineering, University of Maryland, College park, MD, 20742 USA
Manuscript received May 19, 2021; revised May 19, 2021.

where $\mathcal{L}(X \to Z)$ is an appropriate loss function measures the information leakage and $\mathcal{L}(Y \to Z)$ is a function which represents an estimation efficiency of Y given Z (utility). Mutual information (MI), maximal correlation, and minimum mean square error (MMSE) are three choices available for those quantities. In the information privacy literature, [1] investigates the case when $\mathcal{L}(X \to Z)$ and $\mathcal{L}(Y \to Z)$ are both measured by mutual information. [2] and [3] investigate the situation when the privacy constraints are given in terms of maximal correlation and mutual information respectively. They also use MMSE to see how resistant they are against privacy breaches.

Also, the above optimization problem can be viewed separately when $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ are both discrete or continuous cases. In our survey, we will separately compare the different privacy metrics.

The main contribution of present work are as follows:

- In the first part, we investigate the case where $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ are both discrete and the utility function is measured by mutual information. We will compare the case where privacy is measured by maximal correlation versus [1] which uses mutual information as a privacy metric. Our critique shows that mutual information will result tighter lower-bound in rate-privacy metric compared to the maximal correlation case. However, we will show that mutual information is in fact sacrificing the privacy to obtain the higher utility.
- In the second part, we investigate the case where $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ are both continuous and Gaussian random variables. MMSE is used to measure the resistance against the privacy breach. We will compare the case where privacy is measured by maximal correlation [2] versus the case which uses mutual information as a privacy metric [3]. Our critique shows that maximal correlation will result higher lower-bound. However, the bound we have obtained for end point of mutual information is not informative which is coherent with the argue that mutual information is not a good privacy metric.

II. Theoretical bound analysis for information extraction in Discrete Case

[1] analyzes the case when both utility and loss function defined in Eq. (1) are mutual information. In this case, the *rate-privacy* function can be defined as

$$g(p,\epsilon) := \sup_{P_{Z|Y} \in \mathcal{D}_\epsilon(P)} I(Y;Z) \tag{2}$$

where

$$\mathcal{D}_\epsilon(P) := \{P_{Z|Y} : X\text{–o–}Y\text{–o–}Z, I(X;Z) \le \epsilon\} \tag{3}$$

Remark. *Note that by the Markov condition X–o–Y–o–Z, we can always restrict $\epsilon \ge 0$ to only $0 \le \epsilon < I(X;Y)$, because by the data processing inequality we have $I(X;Z) \le I(X;Y)$ and hence for satisfying the privacy constraint ($I(X;Z) \le \epsilon$), it is just enough to satisfy $\epsilon \le I(X;Y)$. The equality is always satisfied by setting $Z = Y$, which yields $g(\epsilon) = H(Y)$.*

Lemma II.1. *The mapping $\epsilon \mapsto g(\epsilon)$ is concave for $\epsilon \ge 0$.*

Proof. Let's consider $P_{Z_1|Y} \in \mathcal{D}_{\epsilon_1}(P)$ and $P_{Z_3|Y} \in \mathcal{D}_{\epsilon_3}(P)$ be optimal with disjoint output alphabet $\mathcal{Z}_1$ and $\mathcal{Z}_3$. Therefore, we can say

$$g(\epsilon_1) = I(Y;Z_1) \text{ and } g(\epsilon_3) = I(Y;Z_3)$$

We introduce an auxiliary binary random variable $U \sim Bernoulli(\lambda)$ independent of (X,Y). Let's consider random privacy filter $P_{Z_\lambda|Y}$ where $\lambda := \frac{\epsilon_2 - \epsilon_1}{\epsilon_3 - \epsilon_1}$ where $\epsilon_2 := \lambda\epsilon_3 + (1-\lambda)\epsilon_1$. This channel is defined as

$$P_{Z_\lambda|Y} = \begin{cases} P_{Z_1|Y}, & U = 0. \\ P_{Z_3|Y}, & U = 1. \end{cases} \tag{4}$$

Note that (X,Y) –o– Z_λ –o– U. We need to show that $P_{Z_\lambda|Y} \in \mathcal{D}_{\epsilon_2}(P)$, where $0 \le \epsilon_1 \le \epsilon_2 \le \epsilon_3 \le I(X;Y)$. To show this:

$$I(X;Z_\lambda) = I(X;Z_\lambda,U) = I(X;Z_\lambda|U) = \lambda I(X;Z_3) + (1-\lambda)I(X;Z_1) \le \lambda\epsilon_3 + (1-\lambda)\epsilon_1 = \epsilon_2$$

Hence,

$$P_{Z_\lambda|Y} \in \mathcal{D}_{\epsilon_2}(P)$$

After satisfying the constraint of Eq. (2), i.e., $P_{Z_\lambda|Y} \in \mathcal{D}_{\epsilon_2}(P)$, the concavity of $g(\epsilon)$ would be resulted by:

$$g(\epsilon_2) \ge I(X;Z_\lambda) = \lambda I(X;Z_3) + (1-\lambda)I(X;Z_1) = (\frac{\epsilon_2 - \epsilon_1}{\epsilon_3 - \epsilon_1})g_{\epsilon_3}(X,Y) + (\frac{\epsilon_3 - \epsilon_2}{\epsilon_3 - \epsilon_1})g_{\epsilon_1}(X,Y)$$

which completes the concavity proof for $g(\epsilon)$. □

Lemma II.2. *The mapping $\epsilon \mapsto \frac{g(\epsilon)}{\epsilon}$ is non-increasing on $(0,\infty)$.*

Proof. We note that since $\epsilon \mapsto g(\epsilon)$ is concave, the chord slope $\frac{g(\epsilon)-g(0)}{\epsilon}$ is non-increasing in ϵ. The corollary then follows by noticing that $\frac{g(\epsilon)}{\epsilon} = \frac{g(\epsilon)-g(0)}{\epsilon} + \frac{g(0)}{\epsilon}$ □

Remark. *Since $g(\epsilon)$ is concave and strictly increasing on $[0, I(X;Y)]$ and hence the continuity of $I(Y;Z)$ and $I(X;Z)$ in $P_{Z|Y}$ implies that the feasible set $\mathcal{D}_{\epsilon_2}(P)$ can be replaced by $\mathcal{D}_{\epsilon}(P) := \{P_{Z|Y} : X\text{–}o\text{–}Y\text{–}o\text{–}Z, I(X;Z) = \epsilon\}$*

Theorem II.3. *$g(\epsilon)$ lies between two straight lines as follows:*

$$\epsilon \frac{H(Y)}{I(X;Y)} + g(0)(1 - \frac{\epsilon}{I(X;Y)}) \leq g(\epsilon) \leq H(Y|X) + \epsilon \tag{5}$$

Proof. For the upper-bound,

$$I(X;Z) = I(Y;Z) - I(Y;Z|X) \geq I(Y;Z) - H(Y|X)$$

The inequality is due to markov chain existing between X, Y, and Z (X–o–Y–o–Z). Therefore,

$$I(Y;Z) \leq I(X;Z) + H(Y|X) \leq \epsilon + H(Y|X)$$

For the lower-bound, since $\epsilon \mapsto g(\epsilon)$ is concave for $\epsilon \geq 0$, $g(\epsilon)$ must lie above the chord connecting $(I(X;Y), H(Y))$ and $(0, g(0))$ as it shown in Fig. 2.

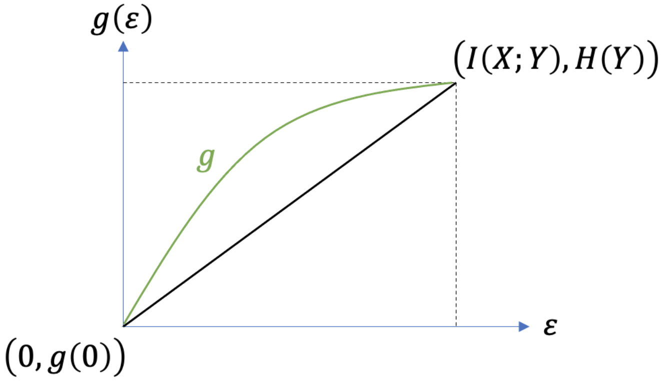

Figure 2. Proof for lower-bound of Theorem 1.

□

A. Original Critique on Discrete case rate-privacy functions

As we will mention in section (III), mutual information does not lead to an arguably operational privacy interpretation and thus cannot serve as an appropriate privacy leakage function. As a result, for defining rate-privacy function, different functions can be used for measuring privacy. In [2], the idea of using maximal correlation as measure of privacy ($\mathcal{L}(X \to Z)$) is introduced. However, the idea was just defined on continues case and also the authors could not find the bounds as mentioned in Theorem II.3. Here, we found the bound for this case same as analysis mentioned in [1]. Moreover, we will compare the bounds to see which bounds are tighter.

Definition 1 (Maximal Correlation)**.** *Given random variables U and V defined over general (discrete or continuous) alphabets $\mathcal{U}$ and $\mathcal{V}$, respectively, the maximal correlation $\rho_m(U,V)$ is defined as*

$$\rho_m(U,V) := \sup_{(f,g)\in\mathcal{S}} \mathbb{E}[f(U)g(V)]$$

where $\mathcal{S} := \{(f,g) : \mathbb{E}[f(U)] = \mathbb{E}[g(V)] = 0, var(f(U)) = var(g(V)) = 1\}$

Applying the Cauchy-Schwartz inequality, Renyi derived an equivalent ”one-function” characterization of the maximal correlation as follows:

$$\rho_m^2(U,V) = \sup_{(f)\in\mathcal{S}} \mathbb{E}[\mathbb{E}^2[f(U)|V]] \tag{6}$$

Definition 2 (Meaninmum Mean Square Error (MMSE))**.** *For a given pair of random variables (U,V), the MMSE of estimating Ugiven V is*

$$mmse(U|V) = \mathbb{E}[(U - (E)[U|V])^2] = \mathbb{E}[var(U|V)] \tag{7}$$

Definition 3 (ϵ-strong estimation privacy [2])**.** *Given a joint distribution P_{XY} and $\epsilon \geq 0$, Z is said to satisfy ϵ-strong estimation privacy, denoted as $Z \in \Gamma_\epsilon(P_{XY})$, if there exists a random mapping (channel) $P_{Z|Y}$ that induces a joint distribution $P_X \times P_{Z|X}$ on $\mathcal{X} \times \mathcal{Z}$, via the Markov condition, satisfying*

$$mmse(f(X)|Z) \geq (1 - \epsilon)var(f(X)) \tag{8}$$

for any non-degenerate Borel functions f on X. Similarly, Z is said to satisfy ϵ-weak estimation privacy, denoted as $Z \in \partial\Gamma_\epsilon(P_{XY})$, if Eq (8) is satisfied only for the identity function $f(x) = x$.

Theorem II.4 ([4])**.** *For a given P_{XY}, $Z \in \Gamma_\epsilon(P_{XY})$ if and only if there exists $P_{Z|Y}$ satisfying $\rho_m^2(X, Z) \le \epsilon$ for any $\epsilon \ge 0$.*

Proof. The correlation ratio of Y and Z, denoted by $\eta_Z(Y)$, is defined as

$$\eta_Z^2(Y) := \frac{var[\mathbb{E}[Y|Z]]}{var(Y)}$$

By using law of total variance, we can say

$$\frac{mmse(Y|Z)}{var(Y)} = 1 - \eta_Z^2(Y)$$

Thus, we have the following:

$$\inf_{f\in\mathcal{S}} \frac{mmse(f(X)|Z)}{var(f(X))} = 1 - \sup_{f\in\mathcal{S}} \eta_Z^2(f(X)) \tag{9}$$

At the same time, we know:

$$\eta_Z^2(Y) = \frac{\mathbb{E}[\mathbb{E}^2[f(X)|Z]]}{var(f(X))} \tag{10}$$

Also, in Eq. (6) we had:

$$\rho_m^2(U, V) = \sup_{(f)\in\mathcal{S}} \mathbb{E}[\mathbb{E}^2[f(U)|V]]$$

By substituting Eqs. (6) and (10) into eq. (9):

$$\inf_{f\in\mathcal{S}} \frac{mmse(f(X)|Z)}{var(f(X))} = 1 - \rho_m^2(X, Z) \tag{11}$$

Therefore, if $\rho_m^2(X, Z) \le \epsilon$, then:

$$mmse(f(X)|Z) \ge (1 - \epsilon)var(f(X))$$

which matches the ϵ-strong estimation privacy. Conversely, let P_{XZ} satisfy the ϵ-strong estimation privacy. Then for any f, Eq. (8) is satisfied. Also, in view of Eq. (11) for arbitrary $\delta > 0$, there exists $f \in \mathcal{S}$ such that

$$1 - \epsilon \le \frac{mmse(f(X)|Z)}{var(f(X))} \le 1 - \rho_m^2(X, Z) + \delta$$

and hence,

$$\rho_m^2(X, Z) \le \epsilon + \delta$$

□

Now, a new rate-privacy function can be defined by using maximal correlation metric as a notion of privacy instead of mutual information as follows:

$$\hat{g}(p, \epsilon) := \sup_{P_{Z|Y}\in\hat{\mathcal{D}}_\epsilon(P)} I(Y; Z) \tag{12}$$

where

$$\hat{\mathcal{D}}_\epsilon(P) := \{P_{Z|Y} : X \text{–o–} Y \text{–o–} Z, \rho_m^2(X; Z) \le \epsilon\} \tag{13}$$

Lemma II.5 (Idea from ref. [4])**.** *The mapping $\epsilon \mapsto \hat{g}(\epsilon)$ is concave for $\epsilon \ge 0$.*

Proof. Let's consider $P_{Z_1|Y} \in \hat{\mathcal{D}}_{\epsilon_1}(P)$ and $P_{Z_3|Y} \in \hat{\mathcal{D}}_{\epsilon_3}(P)$ be optimal with disjoint output alphabet $\mathcal{Z}_1$ and $\mathcal{Z}_3$. Therefore, we can say

$$\hat{g}(\epsilon_1) = I(Y; Z_1) \text{ and } \hat{g}(\epsilon_3) = I(Y; Z_3)$$

We introduce an auxiliary binary random variable $U \sim Bernoulli(\lambda)$ independent of (X, Y). Let's consider random privacy filter $P_{Z_\lambda|Y}$ where $\lambda := \frac{\epsilon_2 - \epsilon_1}{\epsilon_3 - \epsilon_1}$ where $\epsilon_2 := \lambda\epsilon_3 + (1 - \lambda)\epsilon_1$. This channel is defined as

$$P_{Z_\lambda|Y} = \begin{cases} P_{Z_1|Y}, & U = 0. \\ P_{Z_3|Y}, & U = 1. \end{cases} \tag{14}$$

Note that (X, Y) –o– Z_λ –o– U. We need to show that $P_{Z_\lambda|Y} \in \hat{\mathcal{D}}_{\epsilon_2}(P)$, where $0 \le \epsilon_1 \le \epsilon_2 \le \epsilon_3 \le \rho_m^2(X; Y)$. To show this, consider $f \in \mathcal{S}$, then:

$$\mathbb{E}[\mathbb{E}^2[f(X)|Z_\lambda]] = \mathbb{E}[\mathbb{E}[\mathbb{E}^2[f(X)|Z_\lambda]]|U] = \lambda\mathbb{E}[\mathbb{E}^2[f(X)|Z_3]] + (1 - \lambda)\mathbb{E}[\mathbb{E}^2[f(X)|Z_1]]$$

Therefore,

$$\rho_m^2(X;Z_\lambda) = \max_{f\in\mathcal{S}} \mathbb{E}[\mathbb{E}^2[f(X)|Z_\lambda]] = \max_{f\in\mathcal{S}} \lambda\mathbb{E}[\mathbb{E}^2[f(X)|Z_3]]+(1-\lambda)\mathbb{E}[\mathbb{E}^2[f(X)|Z_1]] \leq \lambda\rho_m^2(X;Z_3)+(1-\lambda)\rho_m^2(X;Z_1) \leq \lambda\epsilon_3+(1-\lambda)\epsilon_1$$

Hence,

$$P_{Z_\lambda|Y} \in \hat{\mathcal{D}}_{\epsilon_2}(P)$$

The remaining is just the same as what we did for **Lemma II.1**. □

Corollary II.5.1. *The mapping* $\epsilon \mapsto \frac{\hat{g}(\epsilon)}{\epsilon}$ *is non-increasing on* $(0,\infty)$.

Proof. We note that since $\epsilon \mapsto \hat{g}(\epsilon)$ is concave, the chord slope $\frac{\hat{g}(\epsilon)-\hat{g}(0)}{\epsilon}$ is non-increasing in ϵ. The corollary then follows by noticing that $\frac{\hat{g}(\epsilon)}{\epsilon} = \frac{\hat{g}(\epsilon)-\hat{g}(0)}{\epsilon} + \frac{\hat{g}(0)}{\epsilon}$. □

Corollary II.5.2. *For any* $\epsilon \in [0, \rho_m^2(X;Y)]$, *we have*

$$\hat{g}(\epsilon) \geq \epsilon\frac{H(Y)}{\rho_m^2(X;Y)} + \hat{g}(0)(1-\frac{\epsilon}{\rho_m^2(X;Y)}) \tag{15}$$

Proof. since $\epsilon \mapsto \hat{g}(\epsilon)$ is concave for $\epsilon \geq 0$, $\hat{g}(\epsilon)$ must lie above the chord connecting $(\rho_m^2(X;Y), H(Y))$ and $(0,\hat{g}(0))$ as it shown in Fig. 3.

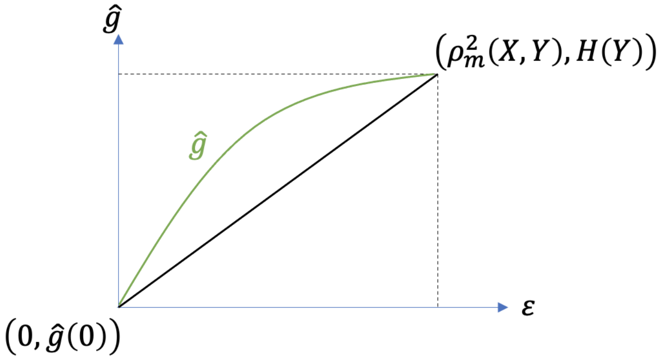

Figure 3. Proof for lower-bound of Corollary 2.

□

Therefore, by comparing Eqs. (5) and (15), we can see the difference is just $I(X;Y)$ and $\rho_m^2(X;Y)$. Hence, to compare the lower-bound in these two cases, it remains to find a relation between $I(X;Y)$ and $\rho_m^2(X;Y)$. The following theorem is one of our main contribution (original critique) for the survey on rate-privacy function which can show which privacy metric will result tighter bound.

Theorem II.6. *In a discrete case when* $|\mathcal{X}| = M$ *and* $|\mathcal{Y}| = N$, $I(X;Y) \leq (M-1)\rho_m^2(X;Y)$.

Proof.

$$I(X;Y) = \mathbb{E}[log\frac{P_{XY}(X,Y)}{P_X(X)P_Y(Y)}] \leq log\mathbb{E}[\frac{P_{XY}(X,Y)}{P_X(X)P_Y(Y)}] = \log(1+\chi^2(P_{XY}||P_X(X)P_Y(Y))) \leq \chi^2(P_{XY}||P_X(X)P_Y(Y))$$

where the first inequality follows from Jensen's inequality and the second inequality is based on the definition of χ^2-divergence as follows:

$$\chi^2(P||Q) := \int (\frac{dP}{dQ})^2 dQ - 1$$

Now, let's construct a matrix of size $M \times N$ with entries $\frac{P_{XY}(X,Y)}{\sqrt{P_X(X)P_Y(Y)}}$, and call it A. Let $\sigma_0 \leq \sigma_1 \leq ... \leq \sigma_k$ be the singular values of A where $k = \min\{M-1, N-1\}$. [5] has a series of beautiful proofs shows that matrix A has $\sigma_0 = 1$ and $\sigma_1 = \rho_m^2(X;Y)$. Therefore,

$$1+\sum_{i=0}^{k}\sigma_i^2 = Tr(AA^*) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\frac{P_{XY}^2(X,Y)}{P_X(X)P_Y(Y)} \to \sum_{i=0}^{k}\sigma_i^2 = \chi^2(P_{XY}||P_X(X)P_Y(Y)) \leq k\sigma_1^2 \leq (M-1)\rho_m^2(X;Y)$$

Also, at the beginning, we found:

$$I(X;Y) \leq \chi^2(P_{XY}||P_X(X)P_Y(Y))$$

Combining those above-mentioned formulas will result: $I(X;Y) \leq (M-1)\rho_m^2(X;Y)$ □

Therefore, the immediate result will appear for a binary case (M=2). In the binary case, $I(X;Y) \leq \rho_m^2(X;Y)$, and therefore, **lower-bound in mutual information case is tighter than lower-bound in maximal correlation case.**

This result is coherent with the deficiency of mutual information as a privacy metric which we will mention that in the next section. The heuristic behind mutual information's higher lower-bound is because mutual information is in fact sacrificing its privacy to obtain higher utility.

III. Deficiency of Mutual Information as Privacy Constraint Metric

In this section, it is shown that the value of classical mutual information does not ensure the safety from privacy breaches. In general, it is assumed that the larger $I(X;Z)$, the less privacy is preserved. Unfortunately, there are situations where privacy is obviously not preserved, but mutual information may not show any problem about it since the mutual information between the private data X and the displayed data Z can be written as the expectation of a Kullback-Leibler distance over Z, *i.e.*, $I(X;Z) = KL(p_{XZ} \| p_X p_Z) = \mathbb{E}_{z \sim Z} KL(p_{X|Z=z} \| p_X)$. In [6], a simple example is exhibited to see this case more clearly.

Example: Let us have one bit private data $X \in \{0,1\}$ where $P[X=0] = P[X=1] = \frac{1}{2}$. Now consider two randomizations, $Z_1 = R_1(X)$, and $Z_2 = R_2(X)$. In the first randomization R_1, we have

$$P[Z_1 = x | X = x] = 0.6$$
$$P[Z_1 = 1-x | X = x] = 0.4$$

Therefore, the probability of privacy breach is

$$\begin{aligned} P[X=1|Z_1=1] = P[X=0|Z_1=0] &= \\ &= \frac{P[Z_1=1|X=1]P[X=1]}{P[Z_1=1]} \\ &= \frac{0.6 \cdot 0.5}{0.6 \cdot 0.5 + 0.4 \cdot 0.5} = 0.6 \end{aligned}$$

In the second randomization R_2, let us have $R_2(X) = Z_2 \in \{e, 0, 1\}$ where e is "erased output". Consider the following R_2 randomization mechanism,

$$P[Z_2 = e | X = x] = 0.9999$$
$$P[Z_2 = x | X = x] = 99 \cdot 10^{-6}$$
$$P[Z_2 = 1-x | X = x] = 1 \cdot 10^{-6}$$

Then, we see that an observer can estimate private data X with high probability if the displayed data is not erased. Therefore, R_2 is a poor randomizer.

$$\begin{aligned} P[X=1|Z_2=1] = P[X=0|Z_2=0] &= \\ &= \frac{P[Z_2=1|X=1]P[X=1]}{P[Z_2=1]} \\ &= \frac{99 \cdot 10^{-6} \cdot 0.5}{99 \cdot 10^{-6} \cdot 0.5 + 10^{-6} \cdot 0.5} = 0.99 \end{aligned}$$

However, if we check $I(X;Z_1)$ and $I(X;Z_2)$, we obtain the following results.

$$\begin{aligned} \log \frac{P[X=z|Z_1=z]}{P[X=z]} &= \log \frac{0.6}{0.5} \approx 0.2630 \\ \log \frac{P[X=1-z|Z_1=z]}{P[X=1-z]} &= \log \frac{0.4}{0.5} \approx -0.3219 \\ KL(p_{X|Z_1=z} || p_X) &\approx 0.6 \cdot 0.2630 - 0.4 \cdot 0.3219 \approx 0.02905 \\ I(X;Z_1) &\approx 0.02905 \end{aligned}$$

$$\begin{aligned} \log \frac{P[X=z|Z_2=z]}{P[X=z]} &= \log \frac{0.99}{0.5} \approx 0.9855 \\ \log \frac{P[X=1-z|Z_2=z]}{P[X=1-z]} &= \log \frac{0.01}{0.5} \approx -5.6439 \\ KL(p_{X|Z_2=z} || p_X) &\approx 0.99 \cdot 0.9855 - 0.01 \cdot 5.6439 \approx 0.91921 \end{aligned}$$

For $Z_2 = e$, and $X \in \{0, 1\}$

$$\log \frac{P[X = x|Z_2 = e]}{P[X = x]} = \log \frac{0.5}{0.5} = 0$$
$$KL(p_{X|Z_2=e}||p_X) = 0$$

Then,

$$I(X; Z_2) = 0.9999 \cdot 0 + 0.0001 \cdot 0.91921 = 0.000091921$$

Unlike our intuition, mutual information says that R_2 is more privacy-preserving than R_1. Mutual information fails to detect privacy breaches because they are very infrequent: they occur only in 0.01% randomizations. In [6], authors suggest to use maximal operation instead of average operation over displayed data Z when measuring the privacy. In other words, the worst possible Kullback-Leibler distance rather than average of Kullback-Leibler distance is better at measuring privacy.

The use of maximal operation instead of average is also considered for different functions rather than Kullback-Leibler distance. As a result, this is one advantages of using maximal correlation operation instead of mutual information as $\mathcal{L}(X \to Z)$, since maximal operation does not have the above-mentioned issue.

IV. Theoretical bound analysis for information extraction in Continuous Case

In this section, we assume that the private data X and observable data Y are continuous random variables. We are interested in the case where the privacy filter $P_{Z|Y}$ is an Additive Gaussian channel. We approach the problem from the estimation theoretic point of view. In other words, we consider the problem as the estimation of observable data Y given the displayed data Z. We want the estimation error to be high for the sake of privacy. Then, after deriving the same channel forms, we do the original critique of MMSE lower bounds between [2] and [3].

Since we have fixed joint distribution $P_{XY} = p$, we use $g(\epsilon)$ instead of $g(p, \epsilon)$ given in equation (2). [3] proposes the following theorem.

Theorem IV.1. *Let (X_G, Y_G) be pair of Gaussian random variables with zero mean and correlation coefficient ρ. Then, for any $\epsilon \in [0, I(X; Y)]$ we have*

$$g(\epsilon) = \frac{1}{2} \log \left(\frac{\rho^2}{2^{-2\epsilon} + \rho^2 - 1} \right) \tag{16}$$

Proof. We can write $Y_G = aX_G + M_G$ where $a^2 = \rho^2 \frac{var(Y_G)}{var(X_G)}$ and $M_G \sim \mathcal{N}(0, \sigma^2)$ where $\sigma^2 = (1 - \rho^2)var(Y_G)$ and M_G is independent of X_G.

We also assume that $Z_\gamma = \sqrt{\gamma} Y_G + N_G$ (or $Z_\gamma = Y_G + \lambda N_G$) where $N_G \sim \mathcal{N}(0, 1)$ as we state at the beginning of the section. Therefore, we have $Z_\gamma = a\sqrt{\gamma} X_G + \sqrt{\gamma} M_G + N_G$ is also Gaussian random variable. Then, we have

$$I(X_G; Z_\gamma) = \frac{1}{2} \log \left(\frac{1 + \gamma Var(Y_G)}{1 + \gamma \sigma^2} \right)$$

For any $\epsilon \in [0, I(X; Y)]$, the unique solution satisfying $I(X; Z_\gamma) = \epsilon$ is

$$\gamma_\epsilon = \frac{1 - 2^{-2\epsilon}}{var(Y_G)(2^{-2\epsilon} + \rho^2 - 1)} \tag{17}$$

Then, by using $Z_\gamma = \sqrt{\gamma_\epsilon} Y_G + N_G$ and noting that $I(Y_G; Z_\gamma)$ is increasing function of γ, we obtain the given result. □

Before presenting our derivation, it would be helpful to see the preliminaries and the result obtained from [2].

Theorem IV.2. *Let (X,Y) be jointly Gaussian random variables, then we have*

$$\rho_m^2(X, Y) \leq 1 - 2^{-2I(X;Y)} \leq (2 \ln 2)\, I(X; Y) \tag{18}$$

Definition 4. ***(α-stable distribution)*** *A random variable X is said to have an α-stable distribution if the characteristic function of X is of the form*

$$\begin{aligned} \varphi(t) &:= \mathbb{E}[e^{itX}] \\ &= e^{itc - b|t|^\alpha (1 + i\kappa sgn(t) \omega_\alpha(t))} = S(\alpha, \kappa, b, c) \end{aligned}$$

where c is constant $-1 \leq \kappa \leq 1$ and

$$\omega_\alpha(t) = \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right) & \textit{if } \alpha \neq 1 \\ \frac{2}{\pi} \log|t| & \textit{if } \alpha = 1 \end{cases}$$

Note that Gaussian random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ has the characteristic function $\varphi(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2}$ which corresponds to $S(2, \kappa, \frac{\sigma}{\sqrt{2}}, \mu)$.

Theorem IV.3. *[7] Let (X,Y) be a given pair of random variables.*

1) *If N is a random variable with α-stable distribution and is independent of (X,Y), then $\lambda \to \rho_m(Y, X + \lambda N)$ is a non-increasing function for $\lambda \geq 0$.*
2) *If N and X are independent and have the same α-stable distribution for $0 < \alpha \leq 2$, then for any $\lambda \geq 0$,*

$$\rho_m(X, X + \lambda N) = \frac{1}{\sqrt{1 + \lambda^{\alpha}}} \tag{19}$$

This theorem shows that if the channel $X \to Y$ is an additive noise channel, *i.e.*, we have $Z = X + \lambda N$ where N and X have an α-stable distribution, then $\rho_m(X, Z)$ can be calculated analytically although it is hard to calculate it for general cases. Then, [2] proposes the following estimation theoretic result for α-stable distributions.

Theorem IV.4. *If the privacy filter $P_{Z|Y}$ is such that for random variables X–o–Y–o–Z, $P_{Z|X}$ can be modeled as $Z = X + \lambda N$, for N independent of (X, Y) and having same α-stable distribution as X for $\alpha \in (0, 2]$. Then,*

$$mmse_{\epsilon}(Y) \geq \left(1 - \varrho_{\epsilon}^2(X, Y)\right) var(Y) \tag{20}$$

where $\varrho_{\epsilon}(X, Y) = \sup_{\rho_m(X,Z) \leq \epsilon} \rho_m(Y, Z)$ and $\text{mmse}_{\epsilon}(Y) = \mathbb{E}[(Y - \mathbb{E}[Y|X + \lambda N])^2]$ when $\rho_m(X, Z) \leq \epsilon$.

A. Original Critique on Continuous case rate-privacy functions

As it is stated in the proof of Theorem IV.1, Z_{γ} can be written as $\sqrt{\gamma} Y_G + N_G$ where $N_G \sim \mathcal{N}(0, 1)$ when the privacy filter is Additive Gaussian channel. In addition to that, if we have Gaussian information and the channel $P_{Y|X}$ is an additive channel, we can obtain Z as $X + \lambda N$. So, we will derive it for [3] in this part so that we can compare the result in [3] with the lower bound in [2] which is given in Theorem IV.4 from the estimation theoretic point of view.

When we have Additive Gaussian channel $P_{Y|X}$ and Gaussian information X_G, we can write

$$Y_G = aX_G + M_G$$

here $a^2 = \rho^2 \frac{var(Y_G)}{var(X_G)}$ and $M_G \sim \mathcal{N}(0, \sigma^2)$ where $\sigma^2 = (1 - \rho^2) var(Y_G)$ and M_G is independent of X_G.

Let us have $\lambda = \frac{1}{\sqrt{\gamma}}$. Then, we can write

$$\begin{aligned} Z_{\gamma} &= \sqrt{\gamma} Y_G + N_G = a\sqrt{\gamma} X_G + \sqrt{\gamma} M_G + N_G \\ &= \frac{a}{\lambda} X_G + \frac{1}{\lambda} M_G + N_G \end{aligned}$$

Let $Z \triangleq \frac{\lambda}{a} Z_{\gamma}$. Then,

$$\begin{aligned} Z &= X_G + \underbrace{\frac{1}{a} M_G + \frac{\lambda}{a} N_G}_{\sim \mathcal{N}(0, \frac{\lambda^2 + \sigma^2}{a^2})} \\ &= X_G + \lambda^{'} N \end{aligned}$$

where $N \sim \mathcal{N}(0, 1)$ and $\lambda^{'} = \frac{\sqrt{\lambda^2 + \sigma^2}}{a}$. Therefore, we have

$$\lambda^{'}_{\epsilon} = \frac{\sqrt{\lambda_{\epsilon}^2 + \sigma^2}}{a}$$

for λ_{ϵ} satisfying $I(X_G; Z_{\gamma_{\epsilon}}) = \epsilon$. If we check $I(X_G; Z)$ for $Z = \frac{\lambda_{\epsilon}}{a} Z_{\gamma_{\epsilon}}$, we have

$$I(X_G; Z) = I(X_G; \frac{\lambda_{\epsilon}}{a} Z_{\gamma_{\epsilon}}) = H(X_G) - H(X_G | \frac{\lambda_{\epsilon}}{a} Z_{\gamma_{\epsilon}})$$

We know $\lambda_{\epsilon} = \frac{1}{\sqrt{\gamma_{\epsilon}}}$ by equation (17). It depends on the correlation ρ, the privacy constraint ϵ, and $var(Y_G)$. These parameters are fixed and known parameters of the system specifications. In addition to that, we know that $a^2 = \rho^2 \frac{var(Y_G)}{var(X_G)}$ which only depends on P_{XY}. Therefore, it is also a fixed and known parameter. Then, we can consider $\frac{\lambda_{\epsilon}}{a}$ as a fixed scaling factor. Scaling $Z_{\gamma_{\epsilon}}$ by a constant factor doesn't affect the change in the uncertainty. Then, we have

$$H(X_G) - H(X_G | \frac{\lambda_{\epsilon}}{a} Z_{\gamma_{\epsilon}}) = H(X_G) - H(X_G | Z_{\gamma_{\epsilon}}) = I(X_G; Z_{\gamma_{\epsilon}}) = \epsilon$$

Then, we eventually have,

$$Z = X_G + \lambda_{\epsilon}^{'} N \quad \text{where } I(X_G; Z) = \epsilon \tag{21}$$

For estimation theoretic comparison of results in [2] and [3], we need an MMSE bound. [3] proposes the following lemma.

Lemma IV.5. *For any given private data X and Gaussian observable data Y_G, we have for any $\epsilon \geq 0$*

$$\inf_{\substack{\gamma \geq 0 \\ I(X;Z_\gamma) \leq \epsilon}} \quad mmse(Y_G|Z_\gamma) \geq var(Y_G) 2^{-2g_\epsilon(X;Y_G)} \tag{22}$$

When we re-evaluate the constraints while substituting Z from (21) for Z_γ,

$$\lambda^{'} = \frac{\sqrt{\lambda^2 + \sigma^2}}{a} \Longleftrightarrow \lambda \geq 0 \Longleftrightarrow \gamma \geq 0$$

We also know that $I(X; Z_\gamma) \leq \epsilon$ takes value where $I(X; Z_\gamma) = \epsilon$, and we have already shown that

$$I(X; Z_\gamma) = \epsilon \Longleftrightarrow I(X; Z) = \epsilon$$

Therefore, we have the following result after replacing $\lambda^{'}$ with λ and X with X_G since X can have any distribution,

$$\inf_{\substack{\lambda \geq 0 \\ I(X;Z) \leq \epsilon \\ Z = X_G + \lambda N}} \quad mmse(Y_G|Z) \geq var(Y_G) 2^{-2g(\epsilon)} = var(Y_G) \frac{2^{-2\epsilon} + \rho^2 - 1}{\rho^2} \tag{23}$$

The equality follows from (16).

When we compare the lower bounds in equation (20) and in equation (23), the first thing that we can realize is that they are same for perfect privacy ($\epsilon = 0$). Moreover, when we check the point where $\epsilon = I(X;Y)$, the bound in equation (23) becomes negative. It follows directly from Theorem IV.2.

$$\rho^2(X,Y) \leq \rho_m^2(X,Y) \leq 1 - 2^{-2I(X;Y)}$$

Therefore,

$$2^{-2I(X;Y)} - 1 + \rho^2 \leq 0$$

It shows that mutual information privacy metric does not provide useful lower-bound when ϵ becomes $I(X;Y)$ while the bound in equation (20) is still non-negative. This observation is also consistent what we say about the deficiency of mutual information as a privacy metric in the previous section. Although, the mutual information doesn't show any sign of trouble, *i.e.*, $I(X;Z) \leq \epsilon$, it may not provide the privacy-preserving mechanism that it suggests.

We can also see it for general ϵ by rearranging the lower bound given in equation (23).

$$var(Y_G) \frac{2^{-2\epsilon} + \rho^2 - 1}{\rho^2} = \left(1 - \frac{1 - 2^{-2\epsilon}}{\rho^2}\right) var(Y_G)$$

$1 - 2^{-2\epsilon}$ is a simple concave function that converges to 1 for $\epsilon \in [0, \infty)$. Then, $\left(1 - \frac{1-2^{-2\epsilon}}{\rho^2}\right)$ term can be negative for some ϵ since $\rho^2 \in [0,1]$. On the other hand, $1 - \varrho_\epsilon^2(X,Y)$ is always non-negative since $\varrho_\epsilon^2(X,Y) \in [0,1]$.

V. Conclusion

In this project, the effect of using different metrics was investigated for measuring the privacy constraints on rate-privacy function separately when $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ are both discrete or continuous cases. According to our findings, mutual information could result tighter lower-bound for discrete case compared to maximal correlation metric. However, at the same time, mutual information could not ensure the safety from privacy beaches. Therefore, when mutual information was used as a privacy constraint in rate-privacy function, the higher lower-bound was resulted which means the higher utility is due to inefficiency of mutual information metric against privacy breaches. In the continuous domain, the rate-privacy function was analyzed by using MMSE estimator as a measure of resistivity against privacy breaches. According to our result, when X and Y are Gaussian random variables and the privacy filter is also Additive Gaussian channel, the maximal correlation would result higher value which indicates a better resistivity against privacy breaches.

REFERENCES

[1] S. Asoodeh, F. Alajaji, and T. Linder, “Notes on information-theoretic privacy,” in *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1272–1278, IEEE, 2014.

[2] S. Asoodeh, F. Alajaji, and T. Linder, “On maximal correlation, mutual information and data privacy,” in *2015 IEEE 14th Canadian workshop on information theory (CWIT)*, pp. 27–31, IEEE, 2015.

[3] S. Asoodeh, M. Diaz, F. Alajaji, and T. Linder, “Information extraction under privacy constraints,” *Information*, vol. 7, no. 1, p. 15, 2016.

[4] S. Asoodeh, F. Alajaji, and T. Linder, “Privacy-aware mmse estimation,” in *2016 IEEE International Symposium on Information Theory (ISIT)*, pp. 1989–1993, 2016.

[5] S. Feizi, A. Makhdoumi, K. Duffy, M. Kellis, and M. Medard, “Network maximal correlation,” in *Computer Science and Artificial Intelligence Laboratory Technical Report*, 2015.

[6] A. Evfimievski, J. Gehrke, and R. Srikant, “Limiting privacy breaches in privacy preserving data mining,” in *PODS ’03*, 2003.

[7] W. Bryc and A. Dembo, “On the maximum correlation coefficient,” *Theory of Probability and Its Applications*, vol. 49, pp. 132–138, 2005.