A Generalization of the Stratonovich’s Value of Information and Application to Privacy-Utility Trade-off

Akira Kamatsuka  
Shonan Institute of Technology  
Email: kamatsuka@info.shonan-it.ac.jp

Takahiro Yoshida  
Nihon University  
Email: yoshida.takahiro@nihon-u.ac.jp

Toshiyasu Matsushima  
Waseda University  
Email: toshimat@waseda.jp

Abstract—The Stratonovich’s value of information (VoI) is a quantity that measures how much inferential gain is obtained from a perturbed sample under information leakage constraint. In this paper, we introduce a generalized VoI for a general loss function and general information leakage. Then we derive an upper bound of the generalized VoI. Moreover, for a classical loss function, we provide a achievable condition of the upper bound which is weaker than that of in previous studies. Since VoI can be viewed as a formulation of a privacy-utility trade-off (PUT) problem, we provide an interpretation of the achievable condition in the PUT context.

I. INTRODUCTION

Research on decision-making under a constraint of information leakage has been studied in 1960s in the academy of sciences of the Soviet Union (USSR Academy). In particular, Stratonovich’s work [1] is pioneering, however, it does not appear to be widely known. In [1] and [2], he introduced Value of Information (VoI) to quantify how much inferential gain is obtained from a perturbed sample $Y$ which contains some information about original sample $X$. His formulation of the VoI was based on the Shannon’s mutual information (MI) $I(X;Y)$ in the information theory [3] and a loss (cost) function $\ell(x,a)$, where $a$ is some action (e.g. point estimation on $X$, hypothesis testing on $p_X$, prediction), in the statistical decision theory (see, e.g., [4]).

Since Shannon’s proposal of MI, various information leakage measures have been proposed. Some examples are Arimoto’s MI [5], Sibson’s MI [6], and Csiszar’s MI [7]. Recently, new information leakage measures have been proposed in the privacy-utility trade-off (PUT) problem, such as f-information [8] and f-leakage [9], as privacy measures. In addition to these measures, by assuming a “guessing” adversary, information leakage measures that have operational meanings have been proposed. For example, Asodeh et al. introduced probability of correctly guessing in [10], [11]. In [12], Issa et al. introduced maximal leakage which quantifies the maximal logarithmic gain of correctly guessing any arbitrary function of the original sample. Extending the maximal leakage, Liao et al. introduced $\alpha$-leakage and $\alpha$-maximal leakage in [13], [16]–[18]. Liao et al. also showed the relationships between the (maximal) $\alpha$-leakage and both Arimoto’s MI and Sibson’s MI. It is worth noting that Liu et al. introduced an $\alpha$-loss to define the $\alpha$-leakage.

In this study, we first introduce an information leakage measure in a general manner by extracting common properties from these specific information leakage measures. Then we define a generalized VoI for the information leakage measure and a general loss function containing the $\alpha$-loss. For the generalized VoI, we derive an upper bound next. Moreover, for a classical loss function $\ell(x,a)$, we also provide an achievable condition of the upper bound which is weaker than that of in previous studies [1], [2] and [19]. We also show basic properties of the achievable upper bound and some extended results. Finally, since VoI can be viewed as a formulation of a PUT problem in a certain situation, based on our prior work [20], we provide an interpretation of the achievable condition in the PUT context.

II. PRELIMINARY

\[ X \xrightarrow{Py|X} Y \xrightarrow{\delta^*} \delta^* \xrightarrow{A} \]

Fig. 1. System Model

In this section, we first review the statistical decision theory and the concept of information leakage in information theory on the system model in Figure 1. For simplicity, unless otherwise stated, we will assume that all alphabets are finite.

A. Notations

Let $X, Y$ and $A$ be random variables on alphabets $\mathcal{X}, \mathcal{Y}$ and $\mathcal{A}$. Let $p_{X,Y} = p_X \times p_{Y|X}$ and $p_Y$ be a given joint distribution of $(X,Y)$ and a marginal distribution of $Y$, respectively. Let $\delta^*: \mathcal{Y} \times \mathcal{A} \to [0,1]$ and $\delta: \mathcal{Y} \to \mathcal{A}$ be a randomized decision rule and a deterministic decision rule, respectively. Since $\delta^*(y,a)$ is equivalent to a conditional probability $p_{A|Y}(a|y)$, we will use these notations interchangeably. The classical notation for a loss function in the statistical decision theory is $\ell(x,a)$, which represents a loss for making an action $A = a$.\[\]
when the true state is $X = x$. In this study, however, we extend the concept of the loss function to a loss for making an action $A = a$ from a sample $Y = y$ using the (randomized) decision rule $\delta^*$ when the true state is $X = x$, denoted as $\ell(x, y, a, \delta^*)$. Finally, we use log to represent the natural logarithm.

### B. Statistical decision theory

We review the basic concepts and results in the statistical decision theory next.

**Definition 1.** The loss function for a randomized decision rule $\delta^* : \mathcal{Y} \times \mathcal{A} \to [0, 1]$ is defined as

$$L(x, \delta^*(y, \cdot)) := \mathbb{E}_A [\ell(x, y, A, \delta^*) \mid Y = y]$$

where $\ell(x, y, a, \delta^*) = \sum_a p_{A|Y}(a \mid y)\ell(x, y, a, \delta^*)$. Finally, we use log to represent the natural logarithm.

**Definition 2.** The risk function and the Bayes risk function for a randomized decision rule $\delta^*$ is defined as

$$R(x, \delta^*, p_{Y|X}) := \mathbb{E}_Y[L(x, \delta^*(Y, \cdot)) \mid X = x] = \sum_y p_{Y|X}(y \mid x) L(x, \delta^*(y, \cdot)), \quad (4)$$

$$r(\delta^*, p_{Y|X}) := \mathbb{E}_X[R(X, \delta^*, p_{Y|X})] = \sum_x p_X(x) R(x, \delta^*, p_{Y|X}). \quad (5)$$

**Proposition 1** (20 Prop 1). The minimal Bayes risk is given by

$$\inf_{\delta^*} r(\delta^*, p_{Y|X}) = r(\delta^*_{\text{Bayes}}, p_{Y|X}) \quad (7)$$

with the optimal randomized decision rule $\delta^*_{\text{Bayes}}$ given by

$$\delta^*_{\text{Bayes}}(y, \cdot) := \arg\inf_{\delta(y, \cdot)} \mathbb{E}_X[L(X, \delta^*(Y, \cdot)) \mid Y = y], \quad (9)$$

where infimum is over all randomized decision rule $\delta^*(y, \cdot) = p_{A|Y}(\cdot \mid y)$ for fixed $y$. In particular, when a channel is $p_{Y|X} = p_Y$ (i.e., $X$ and $Y$ are independent, denoted by $X \perp Y$),

$$\inf_{\delta^*} r(\delta^*, p_Y) = r(\delta^*_{\text{Bayes}}, p_Y)$$

$$= \mathbb{E}_Y \left[ \inf_{\delta(y, \cdot)} \mathbb{E}_X[L(X, \delta^*(Y, \cdot))] \right]. \quad (11)$$

**Remark 1.** The corresponding result of the Proposition 1 for a deterministic decision rule $\delta$ and a classical loss function $\ell(x, a)$ is given by

$$\delta^\text{Bayes}(y) := \arg\inf_a \mathbb{E}_X[\ell(X, a) \mid Y = y], \quad (12)$$

$$\inf_{\delta} r(\delta, p_{Y|X}) = r(\delta^\text{Bayes}, p_{Y|X}) = \mathbb{E}_Y \left[ \inf_a \mathbb{E}_X[\ell(X, a)] \mid Y \right], \quad (13)$$

$$\inf_{\delta} r(\delta, p_Y) = r(\delta^\text{Bayes}, p_Y) = \inf_a \mathbb{E}_X[\ell(X, a)]. \quad (14)$$

### C. Information leakage

In this study, we introduce information leakage measure, denoted as $L(X \to Y)$, to quantify how much information $Y$ leak about $X$. To this end, we extract some properties in common to well-known information leakage measures in information theory.

**Definition 3.** The information leakage $L(X \to Y) = L(p_X, p_{Y|X})$ is defined as a functional of $p_X$ and $p_{Y|X}$ that satisfies following properties:

1. **Non-negativity:**

$$L(X \to Y) \geq 0. \quad (15)$$

2. **Data Processing Inequality (DPI):**

If $X - Y - Z$ forms a Markov chain, then

$$L(X \to Z) \leq L(X \to Y). \quad (16)$$

3. **Independence:**

$$L(X \to Y) = 0 \iff X \perp Y. \quad (17)$$

1. **Examples of the information leakage:** Table I shows the typical information leakage measures in information theory that have these properties and their references, where

- $\alpha \in (0, 1) \cup (1, \infty)$. Note that the value of the information leakage measures in the table are extended by continuity to $\alpha = 1$ and $\alpha = \infty$.
- $H_\alpha(X) := \frac{1}{1-\alpha} \log \left( \sum_x p_X(x)^\alpha \right)^{\frac{1}{\alpha}}$ is the Rényi entropy of order $\alpha$.
- $H^\alpha_{\text{Bayes}}(X|Y) := \frac{1}{1-\alpha} \log \left( \sum_y p_{X,Y}(x,y)^\alpha \right)^{\frac{1}{\alpha}}$ is Arimoto’s conditional entropy of $X$ given $Y$ of order $\alpha$.
- $D_\alpha(p||q) := \frac{1}{1-\alpha} \log \left( \sum_z p^\alpha(z) q^{1-\alpha}(z) \right)$ is the Rényi divergence of order $\alpha$.
- $U$ represents an arbitrary (potentially random) function of $X$ and $U$ represents its estimator.
- $D_f(p||q) := \sum_{z \in Z} q(z) f \left( \frac{p(z)}{q(z)} \right)$ is the $f$-divergence, where $f : \mathbb{R}_+ \to \mathbb{R}$ is a convex function such that $f(1) = 0$ and strictly convex at $t = 1$, where $\mathbb{R}_+ := [0, \infty)$.

Note that relationships between these information leakage measures are given as follows:

- $I(X; Y) = I_1^\alpha(X; Y) = I_1(X; Y) = I_1^\alpha(X; Y) = I_1(X; Y)$, where $f(t) = t \log t$.
- $I_\alpha^\text{Bayes}(X; Y) = L_\alpha(X \to Y)$ (see [10] Thm 1).
- $L_\alpha(X \to Y)$

$$= \left\{ \begin{array}{ll}
\sup_{p_X} I_\alpha^X(\tilde{X}; Y) = \sup_{p_X} I_\alpha^X(\tilde{X}; Y), & \alpha > 1, \\
L_{\text{Max}}(X \to Y), & \alpha = \infty, \\
I(X; Y), & \alpha = 1,
\end{array} \right.$$

$\tilde{X}$ is a probability distribution over support of $p_X$. See [10] Thm 2 for detail.

Most of the non-negativity properties 1) in the Table I follow from the non-negativity of $D_{\alpha}(p||q)$ and $D_f(p||q)$. Note that properties of $\alpha$-leakage $L_\alpha(X \to Y)$ follows

$^2$Note that these properties are part of requirements for reasonable information leakage measures proposed by Issa et al.
from that of Arimoto’s MI $I^\alpha_a(X;Y)$ because of their identity mentioned above. Independence property 3) of maximal $\alpha$-leakage $L^\alpha_{\text{leak}}(X \to Y)$ follows from the property in the $\alpha$-leakage, while the property of Sibson’s MI follows can be derived in a similar manner of [6, Thm 2].

Csiszár’s MI $I^\alpha_c(X;Y)$ and $f$-leakage $L_f(X \to Y)$ also have the independence property 3). In fact, for Csiszár’s MI, it follows from the non-negativity of the $\alpha$-divergence that $I^\alpha_c(X;Y) = E_X [D_\alpha(p_Y|X) || q_Y^\alpha] = 0 \iff D_\alpha(p_Y|X) || q_Y^\alpha = 0$ a.s. $\iff p_Y|X(y|x) = q_Y^\alpha(y) = p_Y(y)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, $X \perp \perp Y$, where $q_Y^\alpha := \text{argmin}_{q_Y} E_X [D_\alpha(q_Y|X) || q_Y]$, $\text{sup} (p_Y)$, $y \in \mathcal{Y}$.

For $f$-leakage, it can be shown in a similar way. Finally, DFI property of $f$-leakage follows from [26, Lem 4], [25, Thm 7.2], and a discussion in [24, Sec V].

2) $\text{mmse}$-leakage: In addition to the typical information leakage measures, we can define a new information leakage measure, minimum mean squared error-leakage $L_{\text{mmse}}(X \to Y)$, which has the properties 1), 2) but does not satisfy 3) in general. Note that we assume that alphabets are continuous here, i.e., $\mathcal{X} = \mathcal{Y} = \mathbb{R}$.

**Definition 4** (Minimum mean squared error-leakage). The minimum mean squared error-leakage $L_{\text{mmse}}(X \to Y)$ is defined as

$$L_{\text{mmse}}(X \to Y) := \mathbb{V}(X) - \inf_{f: \mathcal{Y} \to \mathcal{X}} E_X E_Y [(X - f(Y))^2]$$

(18)

$$= \mathbb{V}(X) - E_Y [\mathbb{V}(X | Y)],$$

(19)

where infimums is over all (measurable) function $f: \mathcal{Y} \to \mathcal{X}$ and $\mathbb{V}(X) := E_X [(X - E_X[X])^2], \mathbb{V}(X | Y) := E_X [(X - E_X[X | Y])^2]$ are variance of $X$ and conditional variance of $X$ given $Y = y$, respectively.

**Proposition 2.** $L_{\text{mmse}}(X \to Y)$ has the properties 1), 2) but does not satisfy 3) in general.

**Proof.** Property 1) is trivial from the definition of the $\alpha$-leakage. Property 2) can be proved as follows: If $X \perp \perp Z$ forms a Markov chain, then $L_{\text{mmse}}(X \to Y) - L_{\text{mmse}}(X \to Z) = E_Y E_Z [(E_X[X | Y] - E_X[X | Z])^2] \geq 0$, where we used the *orthogonal principle* in the first equality.

Finally, it follows from the *law of total variance* $\mathbb{V}(X) = E_Y E_X [X | Y] + E_X [E_Y [X | Y]]$ that $L_{\text{mmse}}(X \to Y) = 0 \iff E_Y E_X [X | Y] = E_X [E_Y [X | Y]] = 0 \iff E_X [X | Y] = E_X [X] a.s.$ The equality condition $E_X [X | Y] = E_X [X]$ a.s. is often called a *mean independence*, which is known as a weaker condition than independence 3), i.e., $X \perp \perp Y \implies E_X [X | Y] = E_X [X] a.s.$

**Remark 2.** As with the $\text{mmse}$-leakage $L_{\text{mmse}}(X \to Y)$, Arimoto’s MI of order $\alpha = \infty$, i.e., $I^\infty_a(X;Y)$ does not have the independence property 3) (see [23, Sec 6.6]).

### III. A Generalization of the Value of Information

In this section, we introduce the *Stratonovich’s Value of Information* (Vol) in a general manner to formulate the leakage-
utility trade-off problem. We also show that the generalized VoI can be viewed as an analogue of the distortion-rate function and the information bottleneck.

A. Average gain

We first introduce average gain to quantify the utility of using $Y$ for a decision-making as largest reduction of the minimal Bayes risk compared to independent case.

**Definition 5** (Average gain). The average gain of using $Y$ on $X$ for making an action $A$ when a loss function is $\ell(x, y, a, \delta^*)$ is defined as

$$
gain^\ell(X; Y) := \inf_{\delta^*} \left[ E_Y \left[ L(X, \delta^*(Y, \cdot)) \right] \right]$$

where $\ell(x, y, a, \delta^*)$ is extended to $\ell(x, y, a, \delta^*)$.

$$
= E_Y \left[ \inf_{\delta^*(y, \cdot)} E_X \left[ L(X, \delta^*(Y, \cdot)) | Y \right] \right] - E_Y \left[ \inf_{\delta^*(y, \cdot)} E_X \left[ \ell(X, a) | Y \right] \right].
$$

**Remark 3.** Note that the average gain is a statistical decision-theoretic counterpart of the average cost gain $\Delta C$ defined in [28].

Using the similar argument as in [29], see V.F, it follows that the average gain satisfies the DPI.

**Proposition 3** ([29], see V.F). For any loss function $\ell(x, y, a, \delta^*)$, the average gain $gain^\ell(X; Y)$ satisfies DPI.

**Example 1.** When a decision maker’s action is to estimate $X$ deterministically under a squared-loss, i.e., $A = \hat{X} = \delta(Y)$, $\ell_{sq}(x, \hat{x}) := (x - \hat{x})^2$, $gain^{\ell_{sq}}(X; Y) = \mathcal{L}_{mse}(X \rightarrow Y)$.

**Example 2.** When a decision maker’s action is to estimate $X$ deterministically under an $\alpha$-loss proposed by Liao et al. in [10] Def 3, i.e., $A = \hat{X}, \ell_{\alpha}(x, y, \hat{x}, \delta^*) := \frac{\alpha}{\alpha^2 - 1} \left( 1 - \delta^*(y, \hat{x}) \right)^{\alpha - 2} \mathbb{I}(\hat{x} = x)$.

$$
gain^{\ell_{\alpha}}(X; Y)
= \left\{ \begin{array}{ll}
\frac{\alpha}{\alpha^2 - 1} \left( e^{\frac{1-n}{n}} H_n(Y|X) - e^{\frac{1-n}{n}} H_n(X) \right), & \alpha > 1 \\
H(X) - H(X | Y) = I(X; Y), & \alpha = 1.
\end{array} \right.
$$

where (23) follows from [10] Lem 1.

Intuitively, the optimal decision rule [9] seems not to depend on $y$ when the independent channel $p_Y$ is used, however, it is not the case in general loss function $\ell(x, y, a, \delta^*)$. Thus we restrict the loss function to the following standard loss class.

**Definition 6** (Standard loss). The loss function $\ell(x, y, a, \delta^*)$ is said to be a standard loss if there exists a function $\ell: \mathcal{X} \times \mathcal{A} \times [0, 1] \rightarrow \mathbb{R}^+_s, (x, a, p) \mapsto \ell(x, a, p)$ such that for all $x, y, a$ and $\delta^*$,

$$
\ell(x, y, a, \delta^*) = \tilde{\ell}(x, a, \delta^*(y, a)).
$$

**Example 3.** The classical loss function $\ell(x, a)$ and the $\alpha$-loss $\ell_{\alpha}(x, y, \hat{x}, \delta^*)$ in the Example 2 are typical examples of the standard loss.

**Proposition 4.** For a standard loss $\ell(x, y, a, \delta^*)$, the optimal decision rule [9] does not depend on $y$ when a channel is independent.

**Proof.** Since

$$
\inf_{\delta^*(y, \cdot)} E_X \left[ L(X, \delta^*(y, \cdot)) \right]
= \inf_{\delta^*(y, \cdot)} \sum_x p_X(x) \delta^*(y, a) \tilde{\ell}(x, a, \delta^*(y, a))
$$

is constant regardless of the value of $y$, the optimal decision rule [9] does not depend on $y$.

B. A Generalization of the Value of Information

We define VoI for information leakage to formulate the leakage-utility trade-off problem. In the following, we assume that the information leakage $\mathcal{L}(X \rightarrow Y)$ is bounded, i.e., there exists an upper bound $K(X)$ that can depend on $p_X$ such that for all $p_Y|X$, $\mathcal{L}(X \rightarrow Y) \leq K(X)$.

**Definition 7.** Let the loss function $\ell(x, y, a, \delta^*)$ be a standard loss. For $0 \leq R \leq K(X)$, the generalized value of information for information leakage $\mathcal{V}_L^\ell(R; Y)$ is defined as

$$
\mathcal{V}_L^\ell(R; Y) := \sup_{p_Y|X: \mathcal{L}(X \rightarrow Y) \leq R} \text{gain}^\ell(X; Y)
= \inf_{\delta^*(y, \cdot)} \mathbb{E}_Y \left[ \sup_{p_Y|X: \mathcal{L}(X \rightarrow Y) \leq R} \text{gain}^\ell(X; Y) \right].
$$

In particular, for a deterministic decision rule and a classical loss function $\ell(x, a)$, it is given as

$$
\mathcal{V}_L^\ell(R; Y) = \inf_{\alpha \geq 1} \mathbb{E}_Y \left[ \inf_{p_Y|X: \mathcal{L}(X \rightarrow Y) \leq R} \text{gain}^\ell(X; Y) \right].
$$

**Remark 4.** Stratonovich’s original formulation of VoI is when $\mathcal{L}(X \rightarrow Y) = I(X; Y)$ and classical loss $\ell(x, a)$. Note that the second term of the generalized VoI $\mathcal{U}_L^V(R; Y) := \inf_{p_Y|X: \mathcal{L}(X \rightarrow Y) \leq R} \mathbb{E}_Y \left[ \inf_{\delta^*(y, \cdot)} \mathbb{E}_X \left[ L(X, \delta^*(Y, \cdot)) | Y \right] \right]$ will be the distortion-rate function $D(R; Y)$ under a non-standard loss function $\ell(x, y, a, \delta^*) = d(x, y)$, where $d(x, y)$ is a distortion.
function, which is not appropriate loss for a decision-making context since it only measures the distortion between $x$ and $y$.

Example 4. From the Example 1 and Example 2 it follows immediately that

$$V_{\alpha=1}^\ell(R;\mathcal{Y}) = R, \quad 0 \leq R \leq V(X),$$

$$V_{\alpha}^\ell(R;\mathcal{Y}) = R, \quad 0 \leq R \leq H(X),$$

for all alphabet $\mathcal{Y}$.

Example 5. When an action is to estimate $U$ correlated only with $X$, i.e., $A = U$ under $\alpha = 1$-loss $\ell_{\alpha=1}^U(u, y, \hat{u}, \delta^*) := \frac{\alpha}{\alpha=1}(1 - \delta^*(y, \hat{u})) = \mathbb{I}(\hat{u} = u)$ and the information leakage constraint $\mathcal{L}(X \rightarrow Y) = I(X; Y) \leq R$, the generalized VoI is given as

$$V_{\alpha=1}^U(R;\mathcal{Y}) := \sup_{P_Y|X: I(X;Y) \leq R} \text{gain}_{\alpha=1}^U(U;Y)\quad(31)$$

$$V_{\alpha}^U(R;\mathcal{Y}) = \sup_{P_Y|X: I(Y;X,Y) \leq R} I(U;Y).\quad(32)$$

Note that this quantity is the well-known information bottleneck [30].

IV. MAIN RESULTS

The main results of this paper are an upper bound of the VoI for a standard loss and a fundamental limit of the VoI for a classical loss.

A. Upper bound and Fundamental Limit

For a standard loss $\ell(x, y, a, \delta^*)$, following upper bound holds.

Proposition 5. For a standard loss $\ell(x, y, a, \delta^*)$, define a function as follows:

$$V_{\ell}^R(R;\mathcal{Y}) := \inf_{\delta^*(y, \cdot)} \mathbb{E}_X [L(X, \delta^*(y, \cdot))] - \inf_{P_{Y|X}, \delta^*} \mathbb{E}_{X,Y} [L(X, \delta^*(Y, \cdot))].\quad(33)$$

Then $V_{\ell}^R(0) = 0$ and for $0 \leq R \leq K(X)$ and arbitrary alphabet $\mathcal{Y}$,

$$V_{\ell}^R(R;\mathcal{Y}) \leq \bar{V}_{\ell}^R(R;\mathcal{Y}).\quad(34)$$

Proof. See Appendix A.

Note that the upper bound (33) still depends on the alphabet $\mathcal{Y}$. Interestingly, when it comes to the classical loss function $\ell(x, a)$, corresponding upper bound is independent on the alphabet $\mathcal{Y}$ and it is even achievable.

Theorem 1. For a classical loss $\ell(x, a)$, define a function as follows:

$$V_{\ell}^R(R) := \inf_a \mathbb{E}_X [\ell(X, a)] - \inf_{P_{A|X}, L(X \rightarrow A) \leq R} \mathbb{E}_{X,A} [\ell(X, A)].\quad(35)$$

Then $V_{\ell}^R(0) = 0$ and for $0 \leq R \leq K(X)$ and arbitrary alphabet $\mathcal{Y}$,

$$V_{\ell}^R(R;\mathcal{Y}) \leq \bar{V}_{\ell}^R(R;\mathcal{Y}).\quad(36)$$

Moreover, let $t(A)$ be a sufficient statistic of $A$ for $X$ and $t(A)$ be set of all values of the statistic. Then the equality in the inequality (36) holds when $\mathcal{Y} = t(A)$ and the optimal mechanism is given by

$$p_{Y|X}(y | x) := \sum_a p_{A|x}(a | x) \mathbb{I}(y = t(a)),\quad(37)$$

where $p_{A|x} = \arg \inf_{\mathcal{P}(A | X)} \ell(X \rightarrow A) \leq R \mathbb{E}_{X,A} [\ell(X, A)]$.

The statement above can be summarized as follows:

$$\sup_{\mathcal{Y}} V_{\ell}^R(R;\mathcal{Y}) = \bar{V}_{\ell}^R(R).\quad(38)$$

Proof. See Appendix B.

Remark 5. Stratonovich call $\bar{V}_{\ell}^R(R)$ as Value of Shannon’s Information in [2, Chapter. 9.3]. Thus we can call $V_{\alpha}^R(R)$ (resp. $V_{\alpha}^R(R)$, $V_{\alpha}^R(R)$) and $V_{\ell}^R(R)$ (resp. $V_{\ell}^R(R)$, $V_{\ell}^R(R)$) as Value of Arimoto’s (resp. Sibson’s, Csiszár’s, $f$-) Information and Value of $\alpha$- (resp. maximal $\alpha$, $f$-) leakage.

Let the alphabet $\mathcal{X}$ be $\mathcal{X} := \{1, 2, \ldots, m\}$ and $\mathbb{P}(X)$ be a probability simplex in $\mathbb{R}^m$. In Storatonovich’s original proof of the achievability, he showed the equality condition as $\mathcal{Y} = \mathbb{P}(X)$ and $Y = (p_X|A(1 | A), p_X|A(2 | A), \ldots, p_X|A(m | A)) \in \mathbb{P}(X)$. In [19], Raginsky gave much shorter proof with $\mathcal{Y} = A$ and $Y = A$. Note that both equality conditions are special cases of the Theorem 1 i.e., following holds.

Proposition 6. $t(A) = A$ is a sufficient statistic of $A$ for $X$. Moreover, if a family of distributions $\{p_{A|X}(\cdot | x) \}_{x \in X}$ have the same support, then $t(A) = (p_{X|A}(1 | A), p_{X|A}(2 | A), \ldots, p_{X|A}(m | A))$ is also sufficient for $X$.

Proof. See Appendix C.

Remark 6. Even though mmse-leakage $\mathcal{L}_{\text{mmse}}(X \rightarrow Y)$ and Arimoto’s MI of order $\alpha = \infty$, i.e., $I_{\alpha}^\infty(X; Y)$ does not have the independence property 3, almost the same result holds for $V_{\ell}^R(R)$ and $V_{\ell}^R(R)$ since the only part that we use the independence property is to prove $V_{\ell}^R(0) = 0$. Note that $V_{\ell}^R(0) \geq 0$ and $V_{\ell}^R(0) \geq 0$ in general.

B. Basic properties of the Fundamental Limit

The following basic properties hold for the fundamental limit $V_{\ell}^R(R)$.

Proposition 7.

1) $V_{\ell}^R(R)$ is non-decreasing in $R$.

2) $V_{\ell}^R(R)$ is concave (resp. quasi-concave) if $\mathcal{L}(X \rightarrow A)$ is convex (resp. quasi-convex) in $p_{A|X}$. 

Remark 7. Even though mmse-leakage $\mathcal{L}_{\text{mmse}}(X \rightarrow Y)$ and Arimoto’s MI of order $\alpha = \infty$, i.e., $I_{\alpha}^\infty(X; Y)$ does not have the independence property 3, almost the same result holds for $V_{\ell}^R(R)$ and $V_{\ell}^R(R)$ since the only part that we use the independence property is to prove $V_{\ell}^R(0) = 0$. Note that $V_{\ell}^R(0) \geq 0$ and $V_{\ell}^R(0) \geq 0$ in general.

Proposition 8.
3) Let \( L_1(X \rightarrow Y), L_2(X \rightarrow Y) \) be information leakage measures. If there exists a constant \( c > 0 \) such that \( L_1(X \rightarrow Y) \leq c L_2(X \rightarrow Y) \), then
\[
V^{L_1}_{\log}(R) \leq V^{L_2}_{\log}(cR), \quad (39)
\]
\[
V^{\ell}_{\log}(R/c) \leq V^{\ell}_{L_1}(1R). \quad (40)
\]

**Proof.** See Appendix D. \( \blacksquare \)

**Corollary 1.** From the property 2) above, following holds.

- \( V_f(R) \) is concave since \( I(X; A) \) is convex in \( p_{A|X} \) for fixed \( p_X \) (see, e.g., [21 Thm 2.7.4])
- \( V^\ell_{\log}(R) \) is quasi-concave since \( L_\alpha(X \rightarrow A) = I^\alpha(X; A) \) is quasi-convex in \( p_{A|X} \) for fixed \( p_X \) (see [10 Thm 10])
- For \( 0 < \alpha < 1 \), \( V^\ell_{\log}(R) \) is concave since \( I^\alpha(X; A) \) is convex in \( p_{A|X} \) for fixed \( p_X \) (see [31 Thm 10])
- For \( 0 < \alpha < 1 \), \( V_f(R) \) is convex since \( I_f(X; A) \) and \( L_f(X \rightarrow A) \) are convex in \( p_{A|X} \) for fixed \( p_X \) (see [32 Thm 9 (c)])
- \( V_f(R) \) and \( V^\ell_{\log}(R) \) are concave since \( I_f(X; A) \) and \( L_f(X \rightarrow A) \) are both convex in \( p_{A|X} \) for fixed \( p_X \)
- For \( 0 < \alpha < 1 \), \( V^\ell_{\log}(R) \) is quasi-convex since \( L^\alpha_{\max}(X \rightarrow A) \) is quasi-convex in \( p_{A|X} \) for fixed \( p_X \) (see [10 Thm 3]). For \( 0 < \alpha < 1 \), \( V^\ell_{\log}(R) \) is concave since \( L^\alpha_{\max}(X \rightarrow A) \) is convex in \( p_{A|X} \) for fixed \( p_X \)

Figure 2 shows a graph of the value of Shannon’s information.

![Figure 2. Value of Shannon’s information](image)

**C. Extension: logarithmic value of information**

Instead of the average gain in Definition 5, we can consider the logarithmic gain to capture utility.

7From the convexity of \( f \)-divergence [26 Lem 4.1], one can derive the convexity of \( I_f(X; A) \) and \( L_f(X \rightarrow A) \) in \( p_{A|X} \).

8Convexity of \( L^\alpha_{\max}(X \rightarrow A) \) in \( p_{A|X} \) follows from [10 Thm 2] and [31 Thm 10].

**Definition 8.** The logarithmic gain of using \( Y \) on \( X \) for making an action \( A \) when a loss function is \( l(x, a) \) and the logarithmic value of information are defined as follows:
\[
\text{Lgain}^f(X; Y) := \log \inf_{\delta^*: \rho_Y = \rho_X} r(\delta^*, p_Y | X) \quad (41)
\]
\[
= \log \inf_{\delta^*: \rho_Y = \rho_X} E_Y \{ l(X, a) | Y \}, \quad (42)
\]
\[
L^\ell(Y; X) := \sup_{p_Y/X; \mathcal{L}(X \rightarrow Y) \leq R} \text{Lgain}^\ell(X; Y). \quad (43)
\]

**Example 6.** Let \( A = \hat{X} \) and \( \ell_{sq}(x, \hat{x}) = (x - \hat{x})^2 \). Then
\[
\text{Lgain}^{\ell_{sq}}(X; Y) = \log \frac{\mathbb{E}[V(X)]}{\mathbb{E}_{Y}[\mathbb{V}(X | Y)]} = \mathcal{L}_{\max}(X \rightarrow Y). \quad (44)
\]

From Proposition 2 it follows that \( \text{Lgain}^{\ell_{sq}}(X; Y) = \mathcal{L}_{\max}(X \rightarrow Y) \) (properties 1), 2) and does not have the independence property 3).

**Remark 7.** It is worth noting that Issa *et al.* introduce maximal versions of the logarithmic gain in [15]. For example, they introduced the variance leakage \( \mathcal{L}''(X \rightarrow Y) \) as follows:
\[
\mathcal{L}''(X \rightarrow Y) := \sup_{U \rightarrow X \rightarrow Y} \mathcal{L}_{\max}(U \rightarrow Y) \quad (45)
\]
\[
= -\log(1 - \rho_m(X; Y)), \quad (46)
\]
(see [15 Def 10 and Lem 16]) where
\[
\rho_m(X; Y) := \sup_{f, g} \frac{\mathbb{E}[f(X)g(Y)]}{\mathbb{E}[f(X)]=\mathbb{E}[g(X)]=0, \mathbb{E}[f(X)^2]=\mathbb{E}[g(X)^2]=1} \quad (47)
\]
is the maximal correlation. Note that the variance leakage \( \mathcal{L}''(X \rightarrow Y) \) have all properties 1), 2) and 3) in Definition 3 (see [23 Prop 5.2]). They also introduced a maximal version of all the logarithmic gain, called maximal cost leakage \( \mathcal{L}^c(X \rightarrow Y) \), as follows:
\[
\mathcal{L}^c(X \rightarrow Y) := \sup_{U \rightarrow X \rightarrow Y} \text{Lgain}^c(U; Y) \quad (48)
\]
\[
= -\log \sum_{y \in \text{supp}(p_Y)} \min_{x \in \text{supp}(p_X)} p_Y(X | Y, x) \quad (49)
\]
(see [15 Def 11 and Thm 15]). Note also that the maximal cost gain \( \mathcal{L}^c(X \rightarrow Y) \) have all properties 1), 2) and 3) in Definition 3 (see [15 Cor 5]). In addition to these loss (cost) based information leakage measures, they also introduced several utility based information leakage measures and showed relationships to the maximal information leakage \( \mathcal{L}_{\max}(X \rightarrow Y) \). See [15] for detail.

For the logarithmic gain, a similar result as in Theorem 1 holds as follows.

9Here we used the term ‘utility’ in a statistical decision-theoretic sense. Note that Issa *et al.* call ‘utility based information leakage’ as ‘gain based information leakage’. 
Corollary 2. For a classical loss \( \ell(x, a) \), define a function as follows:

\[
\mathbf{LV}_\ell^\ell(R) := \log \inf_a \mathbb{E}_X \left[ \ell(X, a) \right] - \inf_{p_{A|X}: \mathcal{L}(X \rightarrow A) \leq R} \log \mathbb{E}_{X, A} \left[ \ell(X, A) \right].
\]

(50)

Then, following holds.

\[
\sup_\mathcal{Y} \mathbf{LV}_\ell^\ell(R; \mathcal{Y}) = \mathbf{LV}_\ell^\ell(R).
\]

(51)

V. APPLICATION TO PRIVACY-UTILITY TRADE-OFF

In this section, we provide an interpretation of the achievability condition in Theorem 1 in the PUT context. We assume three parties: data curator (Alice), a legitimate user (Bob), and an adversary (Eve). Alice has the original data \( X \) and disclose perturbed data \( Y \) through a privacy mechanism \( p_{Y|X} \) to prevent information leakage to Eve. A privacy constraint is represented as \( \mathcal{L}(X \rightarrow Y) \leq R \), where the information leakage measure \( \mathcal{L}(X \rightarrow Y) \) is chosen arbitrarily by Alice. While Bob’s purpose of using the published data \( Y \) is represented as an action, a deterministic decision rule and a loss function, i.e., \( A = \delta(Y) \) and \( \ell(x, a) \), respectively. Suppose that Alice knows the Bob’s purpose of using the published data \( Y \) before disclosure. We also assume that Bob makes his action with the optimal decision rule \( \delta^\text{Bayes} \) under the loss functions \( \ell(x, a) \).

In the situation above, Theorem 1 states that in order to maximize utility measured by \( \text{gain}^\ell(X; Y) \) under the privacy constraint \( \mathcal{L}(X \rightarrow Y) \leq R \), Alice should take the following steps:

1) Find the channel \( p_{A|X}^* \) such that

\[
p_{A|X}^* = \arg \inf_{p_{A|X}: \mathcal{L}(X \rightarrow A) \leq R} \mathbb{E}_{X, A} \left[ \ell(X, A) \right].
\]

(52)

2) Generate a random variable \( \tilde{A} \) drawn to \( p_{A|X}^* \).

3) Finally, disclose \( Y = t(\tilde{A}) \), a sufficient statistic of \( \tilde{A} \) for \( X \), to public.

Remark 8. When Alice assumes Eve’s purpose of using \( Y \), say \( \delta^\text{eve} \) and \( \ell^\text{eve}(x, y, a, \delta^\text{eve}) \), she can choose a privacy constraint as an average gain for Eve, i.e., \( \mathcal{L}(X \rightarrow Y) := \text{gain}^{\ell^\text{eave}}(X; Y) \). Note that she can even adopt the privacy constraint as the maximal gain \( \text{Mgain}^{\ell^\text{eave}}(X; Y) \) defined as follows, which is the inferential gain for using \( Y \) in the most favorable situation for Eve.

Definition 9 (Maximal gain). For a standard loss \( \ell(x, y, a, \delta^*) \), the maximal gain of using \( Y \) on \( X \) for making an action \( A \) is defined as

\[
\text{Mgain}^\ell(X; Y) := \mathbb{E}_Y \left[ \inf_{\delta^*(y, \cdot)} \mathbb{E}_X \left[ L(X, \delta^*(Y, \cdot)) \right] \right]
\]

\[
- \min_y \mathbb{E}_X \left[ L(X, \delta^*(Y, \cdot)) \mid Y = y \right].
\]

(53)

Note that it follows immediately from [33, Prop 23] that the maximal gain satisfies DPI.

VI. CONCLUSION

In this study, we generalized the Stratonovich’s Vol to formulate a problem of decision-making under a general information leakage constraint and a general loss function. We derived upper bound for the Vol and showed weaker achievability condition than ever for a classical loss function. We presented an interpretation of these results in the PUT context and some extended results. Future work includes deriving calculation algorithms for the upper bound.

APPENDIX A

PROOF OF Proposition 5

Proof. Define \( \tilde{U}_\ell^\ell(R; \mathcal{Y}) \) and \( \tilde{U}_\ell^\ell(R) \) as the second terms of the RHS in (28) and (35), respectively, i.e.,

\[
\tilde{U}_\ell^\ell(R; \mathcal{Y}) := \inf_{p_{Y|X} \in \mathcal{L}(X \rightarrow Y) \leq R} \mathbb{E}_{\mathcal{Y}} \left[ \inf_{\delta^*(y, \cdot)} \mathbb{E}_X \left[ L(X, \delta^*(Y, \cdot)) \mid Y \right] \right],
\]

(54)

\[
\tilde{U}_\ell^\ell(R; \mathcal{Y}) := \inf_{p_{Y|X} \in \mathcal{L}(X \rightarrow Y) \leq R} \mathbb{E}_{X, Y} \left[ L(X, \delta^*(Y, \cdot)) \right].
\]

(55)

It suffices to show that \( \tilde{U}_\ell^\ell(R; \mathcal{Y}) \geq \tilde{U}_\ell^\ell(R) \) for arbitrary alphabet \( \mathcal{Y} \). Define the privacy mechanism \( \tilde{p}_{Y|X} \) and the optimal randomized decision rule \( \delta^*_{\text{Bayes}} = \tilde{p}_{A|Y} \) as

\[
\tilde{p}_{Y|X} := \arg \inf \mathbb{E}_{X, A} \left[ L(X, \delta^*(Y, \cdot)) \right],
\]

(56)

\[
\delta^*_{\text{Bayes}}(y, a) = \tilde{p}_{A|Y} (a \mid y)
\]

(57)

\[
= \arg \inf \sum_{x, a} \mathbb{E}_X \left[ L(X, \delta^*(Y, \cdot)) \right] p_{X|Y}(x \mid y),
\]

(58)

where \( \tilde{p}_{X|Y}(x \mid y) := \frac{p_X(x) \tilde{p}_{Y|X}(y \mid x)}{p_Y(y)} \). Since \( X - Y - A \) forms a Markov chain for the distributions \( \tilde{p}_{Y|X} \) and \( \delta^*_{\text{Bayes}} = \tilde{p}_{A|Y} \),

\[
\mathcal{L}(X \rightarrow A) \leq \mathcal{L}(X \rightarrow Y) \leq R
\]

(59)

holds from DPI (16) and (56). Then from (59),

\[
\tilde{U}_\ell^\ell(R; \mathcal{Y}) := \inf_{p_{Y|X} \in \mathcal{L}(X \rightarrow Y) \leq R} \mathbb{E}_{X, Y} \left[ L(X, \delta^*(Y, \cdot)) \right]
\]

(60)

\[
\leq \sum_{x, y, a} p_X(x) \tilde{p}_{Y|X}(y \mid x) \delta^*_{\text{Bayes}}(y, a) \ell(x, y, a, \delta^*_{\text{Bayes}})
\]

(61)

\[
= \tilde{U}_\ell^\ell(R; \mathcal{Y}).
\]

(62)

APPENDIX B

PROOF OF Theorem 1

Based on [2, Chapter. 9.7] and a refined proof in [19], we prove Theorem 1 as follows.
Proof. Define \( U^L_x(R; Y) \) and \( U^L_x(R) \) as the second terms of RHS in (28) and (35), respectively, i.e.,

\[
U^L_x(R; Y) := \inf_{p_{A|X} : \mathcal{L}(X \rightarrow Y) \leq R} \mathbb{E}_Y \left[ \inf_a \mathbb{E}_X [\ell(X, a) \mid Y] \right],
\]

\[
U^L_x(R) := \inf_{p_{A|X} : \mathcal{L}(X \rightarrow A) \leq R} \mathbb{E}_{X,A} [\ell(X, A)].
\]  

(Converse part): It suffices to show that \( U^L_x(R; Y) \geq U^L_x(R) \) to prove \( V^L_x(R; Y) \leq V^L_x(R) \) for arbitrary \( Y \). This can be proved in a similar way to that in the proof of Proposition [5] (see [29] Appendix D).

(Achievable part): Let \( Y := t(A) \). It suffices to show that \( U^L_x(R; t(A)) \leq U^L_x(R) \). Define \( p^*_A | X, p^*_A \) and \( p^*_X | A \) as follows:

\[
p^*_A | X := \arg \min_{p_{A|X} : \mathcal{L}(X \rightarrow A) \leq R} \mathbb{E}_{X,A} [\ell(X, A)],
\]

\[
p^*_A(a) := \sum_x p_X(x) p^*_A | X (a \mid x),
\]

\[
p^*_X | A(a | x) := \frac{p_X(x)p^*_A | X (a \mid x)}{p^*_A(a)}.
\]

Let \( \hat{A} \) be a random variable drawn to \( p^*_A \). Since \( X - \hat{A} - Y := t(\hat{A}) \) forms a Markov chain,

\[
\mathcal{L}(X \rightarrow Y) \leq \mathcal{L}(X \rightarrow \hat{A}) \leq R
\]  

holds from DPI (16) and (65). Now, define a privacy mechanism \( p^*_Y | X \) as

\[
p^*_Y | X(y \mid x) := \sum_a p^*_A | X(a \mid x) \mathbb{I}_{y = t(a)}.
\]

Then

\[
U^L_x(R; t(A)) := \inf_{p_{Y|X} : \mathcal{L}(X \rightarrow Y) \leq R} \mathbb{E}_Y \left[ \inf_a \mathbb{E}_X [\ell(X, a) \mid Y] \right]
\]

\[
\leq \mathbb{E}_Y \left[ \inf_a \mathbb{E}_{X \mid Y} [\ell(X, a) \mid Y] \right],
\]

where the expectation \( \mathbb{E}_{X \mid Y} [\cdot] \) is taken over the distribution \( p^*_Y | X(x \mid y) = p_X(x)p^*_Y | X(y \mid x)p^*_X(y) \). Now, we will evaluate \( \inf_a \mathbb{E}_{X \mid Y} [\ell(X, a) \mid Y = t(a')] \) from above.

\[
\inf_a \mathbb{E}_{X \mid Y} [\ell(X, a) \mid Y = t(a')] \overset{(*)}{=} \inf_a \mathbb{E}_X [\ell(X, a) \mid \hat{A} = a']
\]

\[
\leq \mathbb{E}_X [\ell(X, a') \mid \hat{A} = a'],
\]

where the equality \((*)\) follows from the sufficiency of \( t(\hat{A}) \) \( \overset{10}{10} \). Thus we have

\[
\mathbb{E}_Y \left[ \inf_a \mathbb{E}_X [\ell(X, a) \mid Y] \right] = \mathbb{E}_{\hat{A}} \left[ \inf_a \mathbb{E}_X [\ell(X, a) \mid \hat{A}] \right]
\]

\[
\leq \mathbb{E}_{\hat{A},X} [\ell(X, \hat{A})] = \inf_{p_{A|X} : \mathcal{L}(X \rightarrow A) \leq R} \mathbb{E}_{X,A} [\ell(X, A)]
\]

\[
= U^L_x(R).
\]  

By combining with (71), \( U^L_x(R; t(A)) \leq U^L_x(R) \).

\[\text{(End of proof)}\]

\section*{APPENDIX C}
\textbf{Proof of Proposition 6}

The sufficiency of \( t(A) = A \) is trivial. To prove the sufficiency of \( t(A) = (p_{X \mid A}(1 \mid A), p_{X \mid A}(2 \mid A), \ldots, p_{X \mid A}(m \mid A)) \), we first introduce the following lemmas.

\textbf{Lemma 1} ([34] Thm 6.12). Assume that a family of distributions \( \{p_{A \mid X}(\cdot \mid x)\}_{x \in X} \) have the same support. Then

\[
s(A) := \left( p_{A \mid X}(A \mid 1), \ldots, p_{A \mid X}(A \mid m), p_{A \mid X}(1 \mid A), \ldots, p_{A \mid X}(m \mid A) \right)
\]  

is a minimal sufficient statistic of \( A \) for \( X \).

\textbf{Lemma 2}. Let \( T_1 = t_1(A) \) be a sufficient statistic of \( A \) for \( X \). If there exists a (measurable) function \( f \) such that \( T_1 = f(t_2(A)) \), then \( T_2 = t_2(A) \) is also sufficient for \( X \).

\textbf{Proof}. The statement follows immediately from the Fisher’s factorization theorem (see, e.g., [34] Thm 6.5) or DPI for Shanon’s mutual information (see, e.g., [21] Eq (2.124)).

\textbf{Lemma 3}.

\[
e(A) = \left( \frac{p_{X \mid A}(2 \mid A)}{p_{X \mid A}(1 \mid A)}, \ldots, \frac{p_{X \mid A}(m \mid A)}{p_{X \mid A}(1 \mid A)} \right)
\]

is a (minimal) sufficient statistic of \( A \) for \( X \).

\textbf{Proof}. Since \( s(A) := \left( \frac{p_{X \mid A}(2 \mid A)}{p_{X \mid A}(1 \mid A)}, \ldots, \frac{p_{X \mid A}(m \mid A)}{p_{X \mid A}(1 \mid A)} \right) = \left( \frac{p_X(1)}{p_X(2)}, \frac{p_X(2)}{p_X(1)} \cdot \frac{p_X(1)}{p_X(4)}, \ldots, \frac{p_X(m)}{p_X(1)} \cdot \frac{p_X(1)}{p_X(4)} \right) \) is a function of \( e(A) \), it follows from Lemma 2 that \( e(A) \) is also sufficient. The minimality follows immediately as follows: For arbitrary \( a, b \in \mathcal{A} \), it holds that \( s(a) = s(b) \iff e(a) = e(b) \).

Making use of these results, we prove Proposition 6 as follows.

\textbf{Proof}. Since \( e(A) = \left( \frac{p_{X \mid A}(2 \mid A)}{p_{X \mid A}(1 \mid A)}, \ldots, \frac{p_{X \mid A}(m \mid A)}{p_{X \mid A}(1 \mid A)} \right) \) is a function of \( t(A) = (p_{X \mid A}(1 \mid A), p_{X \mid A}(2 \mid A), \ldots, p_{X \mid A}(m \mid A)) \), from Lemma 2 \( e(A) \) is also sufficient for \( X \).}

\[\text{(End of proof)}\]
APPENDIX D
PROOF OF PROPOSITION

Proof. The property 1) is trivial. To prove the property 2), it suffices to show that $U(R) := \inf_{P_{A|X}} \mathbb{E}_{X,A}[\ell(X, A)]$ is convex (resp. quasi-convex) when $L(X \rightarrow A) = \mathcal{L}(p_X, p_{A|X})$ is convex (resp. quasi-convex). We will only prove the convexity. For arbitrary $0 \leq \lambda \leq 1$ and $0 \leq R_1, R_2 \leq K(X)$, define

$$p_{A|X}^{*,1} := \arg\inf_{P_{A|X}} \mathbb{E}_{X,A}[\ell(X, A)], \quad (79)$$

$$p_{A|X}^{*,2} := \arg\inf_{P_{A|X}} \mathbb{E}_{X,A}[\ell(X, A)], \quad (80)$$

$$p_{A|X}^{*,\lambda} := \lambda p_{A|X}^{*,1} + (1 - \lambda)p_{A|X}^{*,2}. \quad (81)$$

Then let denote $\mathcal{L}^{*,1}(X \rightarrow A), \mathcal{L}^{*,2}(X \rightarrow A)$ and $\mathcal{L}^{*,\lambda}(X \rightarrow A)$ as the $\alpha$-leakages defined by $p_{A|X}^{*,1}, p_{A|X}^{*,2}$ and $p_{A|X}^{*,\lambda}$, respectively. Then

$$\mathcal{L}^{*,\lambda}(X \rightarrow A) \leq \lambda \mathcal{L}^{*,1}(X \rightarrow A) + (1 - \lambda)\mathcal{L}^{*,2}(X \rightarrow A) \quad (82)$$

$$\leq \lambda R_1 + (1 - \lambda)R_2. \quad (83)$$

Therefore,

$$U(\lambda R_1 + (1 - \lambda)R_2) \leq \mathbb{E}_{X,A}[\ell(X, A)] \quad (84)$$

$$= \sum_{x,a} p_X(x)p_{A|X}^{*,\lambda}(a \mid x)\ell(x, a) \quad (85)$$

$$= \lambda U(R_1) + (1 - \lambda)U(R_2). \quad (86)$$

The quasi-convexity can be proved in a similar way.

To prove the property 3), it suffices to show that

$$V_{\mathcal{L}_2}^\ell(R; \mathcal{Y}) \leq V_{\mathcal{L}_1}^\ell(cR; \mathcal{Y}) \quad (87)$$

for arbitrary alphabet $\mathcal{Y}$. To this end, define

$$p_{Y|X}^{*,2} := \arg\sup_{P_{Y|X}} \text{gain}^\ell(X; Y) \quad (88)$$

for arbitrary alphabet $\mathcal{Y}$. Since

$$\mathcal{L}_1(p_X, p_{Y|X}^{*,2}) \leq c \mathcal{L}_2(p_X, p_{Y|X}^{*,2}) \leq cR, \quad (89)$$

it holds that

$$V_{\mathcal{L}_1}^\ell(cR; \mathcal{Y}) := \sup_{P_{Y|X}: \mathcal{L}_1(X \rightarrow Y) \leq cR} \text{gain}^\ell(X; Y) \quad (90)$$

$$\leq \text{gain}^\ell(p_X, p_{Y|X}^{*,2}) = V_{\mathcal{L}_2}^\ell(R; \mathcal{Y}), \quad (91)$$

where $\text{gain}^\ell(p_X, p_{Y|X}^{*,2}) := r(\delta^*\text{Bayes}, p_Y^{*,2}) - r(\delta^*\text{Bayes}, p_{Y|X}^{*,2})$ and $p_Y^{*,2}(y) := \sum_x p_X(x)p_{Y|X}^{*,2}(y \mid x)$.
REFERENCES

[1] R. Stratonovich, “On value of information,” Izvestiya of USSR Academy of Sciences, Technical Cybernetics, vol. 5, pp. 3–12, 1965.

[2] R. Stratonovich, R. Belavkin, P. Pardalos, and J. Principe, Theory of Information and its Value. Springer International Publishing, 2020.

[3] F. Kanaya and K. Nakagawa, “On the practical implication of mutual information for statistical decisionmaking,” IEEE Transactions on Information Theory, vol. 37, no. 4, pp. 1151–1156, July 1991.

[4] C. E. Shannon, “A mathematical theory of communication,” The Bell System Technical Journal, vol. 27, pp. 379–423, 1948. [Online]. Available: http://plan9.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf

[5] J. Berger, Statistical decision theory and Bayesian analysis, 2nd ed., ser. Springer series in statistics. New York, NY: Springer, 1985.

[6] S. Arimoto, “Information measures and capacity of order α for discrete memoryless channels,” in 2nd Colloquium, Keszthely, Hungary, 1975, I. Csiszar and P. Elias, Eds., vol. 16. Amsterdam, Netherlands: North Holland: Colloquia Mathematica Societatis Jano’s Bolyai, 1977, pp. 41–52.

[7] R. Sibson, “Information radius,” Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, vol. 14, pp. 149–160, 1969.

[8] I. Csiszar, “Generalized cutoff rates and renyi’s information measures,” IEEE Transactions on Information Theory, vol. 41, no. 1, pp. 26–34, 1995.

[9] H. Wang, M. Diaz, F. P. Calmon, and L. Sankar, “The utility cost of robust privacy guarantees,” in 2018 IEEE International Symposium on Information Theory (ISIT), June 2018, pp. 706–710.

[10] J. Liao, O. Kosut, L. Sankar, and F. du Pin Calmon, “Tunable measures for information leakage and applications to privacy-utility tradeoffs,” IEEE Transactions on Information Theory, vol. 65, no. 12, pp. 8043–8066, 2019.

[11] S. Asoodeh, M. Diaz, F. Alajaji, and T. Linder, “Privacy-aware guessing efficiency,” in 2017 IEEE International Symposium on Information Theory (ISIT), 2017, pp. 754–758.

[12] ——, “Estimation efficiency under privacy constraints,” IEEE Transactions on Information Theory, vol. 65, no. 3, pp. 1512–1534, 2019.

[13] I. Issa and A. B. Wagner, “Operational definitions for some common information leakage metrics,” in 2017 IEEE International Symposium on Information Theory (ISIT), June 2017, pp. 769–773.

[14] I. Issa, S. Kamath, and A. B. Wagner, “An operational measure of information leakage,” in 2015 Annual Conference on Information Science and Systems (CISS), March 2016, pp. 234–239.

[15] I. Issa, A. B. Wagner, and S. Kamath, “An operational approach to information leakage,” IEEE Transactions on Information Theory, vol. 66, no. 3, pp. 1625–1657, 2020.

[16] J. Liao, O. Kosut, L. Sankar, and F. P. Calmon, “A tunable measure for information leakage,” in 2018 IEEE International Symposium on Information Theory (ISIT), 2018, pp. 701–705.

[17] J. Liao, L. Sankar, O. Kosut, and F. P. Calmon, “Maximal α-leakage and its properties,” in 2020 IEEE Conference on Communications and Network Security (CNS), 2020, pp. 1–6.

[18] ——, “Robustness of maximal α-leakage to side information,” in 2019 IEEE International Symposium on Information Theory (ISIT), 2019, pp. 642–646.

[19] M. Roginsky, “Value of information, bayes risks, and rate-distortion theory,” The Information Structuralist (Blog), 2010.

[20] A. Kamatsuka, T. Yoshida, and T. Matsushima, “Privacy-utility trade-off with the stratonovich’s value of information,” in 2021 IEEE Information Theory Workshop (ITW), 2021, pp. 1–6.

[21] T. M. Cover and J. A. Thomas, Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing). Wiley-Interscience, 2006.

[22] S. Fehr and S. Berens, “On the conditional renyi entropy,” IEEE Transactions on Information Theory, vol. 60, no. 11, pp. 6801–6810, 2014.

[23] S. Asoodeh, “Information and estimation theoretic approaches to data privacy,” Ph.D. dissertation, Queen’s University at Kingston, 2017.

[24] Y. Polyanskiy and S. Verdú, “Arimoto channel coding converse and renyi divergence,” in 2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2010, pp. 1327–1333.

[25] Y. Polyanskiy. (2020) Information theory methods in statistics and computer science, lecture 1: f-divergences. [Online]. Available: http://people.lids.mit.edu/yp/homepage/data/LN_fdiv.pdf

[26] I. Csiszár and P. C. Shields, “Information theory and statistics: A tutorial,” Commun. Inf. Theory, vol. 1, no. 4, pp. 417–528, Dec. 2004. [Online]. Available: http://dx.doi.org/10.1361/01000000004

[27] Y. Wu and S. Verdú, “Functional properties of minimum mean-square error and mutual information,” IEEE Transactions on Information Theory, vol. 58, no. 3, pp. 1289–1301, 2012.

[28] F. du Pin Calmon and N. Fawaz, “Privacy against statistical inference,” in 2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Oct 2012, pp. 1401–1408.

[29] A. Américo, M. Khouzani, and P. Malacaria, “Conditional entropy and data processing: An axiomatic approach based on core-concavity,” IEEE Transactions on Information Theory, vol. 66, no. 9, pp. 5537–5547, 2020.

[30] N. Tishby, F. C. Pereira, and W. Bialek, “The information bottleneck method,” 1999, pp. 368–377.

[31] S. Ho and S. Verdú, “Convexity/concavity of renyi entropy and α-mutual information,” in 2015 IEEE International Symposium on Information Theory (ISIT), 2015, pp. 745–749.

[32] S. Verdú, “Error exponents and α-mutual information,” Entropy, vol. 23, no. 2, 2021. [Online]. Available: https://www.mdpi.com/1099-4300/23/2/199

[33] M. Alvim, K. Chatzikokolakis, A. McIver, C. Morgan, C. Palamidessi, and G. Smith, “An axiomatization of information flow measures,” Theoretical Computer Science, vol. 777, 10 2018.

[34] E. L. Lehmann and G. Casella, Theory of Point Estimation (Springer Texts in Statistics), 2nd ed. Springer, Aug. 1998.

[35] Y. Yannis, “Causal inference for multiple treatments via sufficiency and ratios of generalized propensity scores,” Jan 2012.