SHONAN INSTITUTE OF
TECHNOLOGY JOURNAL
Vol. 58, No. 1, 2024

情報の価値理論とその応用

鎌塚 明*

Theory of Value of Information and Its Application

Akira Kamatsuka

Abstract: Various attempts have been made to quantify the amount of information, mainly in information theory. On the other hand, various results on quantifying the value of information (VoI) to decision-making have been developed independently in decision theory. This paper outlines the author's previous results on a framework for treating these results in a unified manner. The value of information theory is also applied to the privacy-utility trade-off (PUT) problem in the information disclosure problem, considering privacy protection, which has been the subject of much interest in recent years.

Keywords: Value of Information, Statistical Decision Theory, Privacy-Utility Trade-off

要旨 情報を定量化するためのさまざまな試みは，主に，情報理論の分野で発展してきた．一方，これとは独立に，情報が意思決定に与える価値の定量化に関するさまざまな結果が，意思決定理論（統計的決定理論）の分野において「情報の価値理論」として得られている．本稿では，著者によるこれらの結果を統一的に扱うための理論に関するこれまでの結果について概説する．また，情報の価値理論の応用として，近年，盛んに研究がなされているプライバシ保護を考慮した情報公開問題におけるプライバシ・ユーティリティ・トレードオフ (PUT) 問題へ適用した場合の解釈を与える．

キーワード： 情報の価値理論，統計的決定理論，プライバシ・ユーティリティ・トレードオフ

1 はじめに

情報理論 (Information Theory) は，情報を定量的な量 (情報量) として数理的に扱う学問体系であり，Shannon [1] によって 1948 年に創始された．情報理論において基本となる情報量に，情報エントロピー (Entropy) $H(X)$ と相互情報量 (Mutual Information; MI) $I(X;Y)$ がある．前者はサンプル X のもつ平均的な情報量を意味し，後者はあるサンプル Y がサンプル X について持つ平均的な情報量を意味する．これらの基本的な量を基にして，情報理論は，情報通信，データ圧縮，情報セキュリティ，物理学等，さまざまな分野への応用がなされてきた．また近年では，機械学習や人工知能分野への応用も盛んに行われており，さまざまな意思決定 (decision making) に活用されつつある．

一般に，人間の意思決定に関する理論体系を意思決定理論 (Decision Theory) と呼ぶ．主な意思決定理論としては，統計学分野において Wald によって確立された統計的決定理論 (Statistical Decision Theory) [2] と，経済学分野において von Neumann と Morgenstern [3] によって体系付けられた期待効用理論 (Expected Utility Theory) および Savage [4] による主観的期待効用理論 (Subjective Expected Utility Theory) がある．

これら意思決定理論の枠組みに基づいて，情報がもつ価値 (Value of Information; VoI) を定量化しようするさまざまな研究が，1960 年代から 1970 年代にかけてなされてきた [5], [6],[7, 8],[9],[10]．Raiffa と Schlaifer は，サンプル Y の持つ価値を，それを意思決定に使わなかったときと比べたときの，最適な期待効用（もしくは期待損失）の差として定式化した．この量を Expected Value of Sample Information (EVSI) と呼ぶ [6, Chapter 4]．Stratonovich はあるサンプル Y が関心のあるサンプル X に関する情報を含むとき

*湘南工科大学 情報学部 情報学科 講師

の価値を，EVSI と Shannon の相互情報量を用いて定式化し，その価値の上界およびその達成可能条件を示した [7, 8][2].

情報理論分野においては, Shannon 相互情報量 $I(X;Y)$ が提案されて以来, Y が含む X に関する情報量の定量化として，さまざまなものが提案されてきた．代表的なものに，Sibson 相互情報量 [13]，Arimoto 相互情報量 [14], Csiszár 相互情報量 [15] および Lapidoth–Pfister 相互情報量 [16],[17] がある．一方，情報セキュリティ分野においては，Shannon 相互情報量 $I(X;Y)$ は基本的な情報漏えい尺度 (information leakage measure) として用いられるが，近年では，プライバシ保護情報公開問題の文脈において，さまざまな情報漏えい尺度が提案されている．

プライバシ保護情報公開問題においては，プライバシ情報を含むオリジナルデータ X の保有者 (Alice) とそのデータの正規の利用者 (Bob) およびプライバシ情報に対する情報を得ようとする攻撃者 (Eve) を想定する．Alice は Eve に対するプライバシ情報の漏洩を保護しつつ，Bob にとっての有用性 (Utility) を高めるために，オリジナルデータ X に変換[3]を施した加工情報情報 Y を公開しようとするが，このとき，どのような変換を用いるかが解くべき問題となる．より具体的には，プライバシ保護と有用性の間にはトレードオフ (Privacy-Utility Trade-off; PUT) があると考えられるため，それぞれの尺度を適当に定めた上で，PUT の理論解析を行った上で最適な変換方法を求めることが基本的な問題となる．

この情報公開問題におけるプライバシ保護の尺度として，近年，攻撃者 Eve がオリジナルデータ X あるいは X に相関する対象 U に関する推定を行うと仮定し[4], その推定能力に基づくプライバシ保護尺度がいくつか提案されている．例えば Calmon らは average cost gain を提案している [18]．Issa らは, maximal leakage を提案し [19, 20, 21, 22]，この量は後に，Liao らによって α-leakage や maximal α-leakage に拡張された [23, 24, 25, 26]．Liao らはさらに，α-leakage と Arimoto 相互情報量の等価性および maximal leakage と Arimoto–Sibson 通信路容量 [14] の等価性を示している．また Alvim らは, 効用関数 g に基づく g-leakage [27],[28], [29] を提案しており，Kurri らはそれをプライバシ情報保護における問題に拡張している [30]．これらのプライバシ保護尺度は，決定理論の観点から見ると，EVSI と本質的に等価な量あるいは EVSI を拡張した量として解釈することができる．

本論文は，これまでに提案されている相互情報量や情報漏えい尺度を統一的に扱うための情報の価値理論の構築に関する著者の一連の研究 [31, 32, 33, 34] の概説である．一般的な情報量に対する価値理論の構築にあたって本研究ではまず，1) 非負性，2) データ処理不等式 (Data Processing Inequality; DPI)，3) 独立性に着目し，これらの性質を公理として採用することで，一般化された情報漏えい尺度 (generalized information leakage measures) を導入する．また, サンプル Y を意思決定に用いることによる価値尺度として, EVSI の亜種として average ratio gain を導入する．また，EVSI や average ratio gain において，損失関数（もしくは効用関数）を適当に選ぶことで，既存のさまざまな相互情報量や情報漏えい尺度を EVSI もしくは average ratio gain によって表現できることを示す．その上で，Stratonovich の結果 [7, 8, 11] を拡張し，一般化された情報漏えい尺度に対する情報の価値理論を構築する．さらに，この結果をプライバシ情報公開問題に応用し，PUT 問題としての解釈を与える．

2　準備

X, Y および A をそれぞれ $\mathcal{X}, \mathcal{Y}$ および $\mathcal{A}$ 上に値をとる確率変数とし，$X - Y - A$ はこの順に Markov 連鎖をなす[5]とする．(X, Y) の同時分布 (joint distribution) を $p_{X,Y} = p_X p_{Y|X}$ とし，p_X を X に関する事前分布 (prior distribution) と呼び，$p_{Y|X}$ を通信路 (channel) と呼ぶ．さらに，p_Y を Y に関

[2] Stratonovich の結果は，統計的決定理論と情報理論を結びつける重要な結果であるとみなせるが，文献 [7, 8] がロシア語で書かれていることもあり，広くは知られていないようである．近年，文献 [8] の英語訳 [11] が出版されている．なお，Stratonovich と同様の問題を扱っている文献として，Kanaya と Nakagawa による研究がある [12].

[3] この変換を privacy mechanism と呼ぶが，情報理論における通信路 (channel) に相当する．

[4] このような攻撃者を guessing adversary と呼ぶ.

[5] $X - Y - A$ が Markov 連鎖をなす $\overset{\text{def.}}{\iff} X \perp\!\!\!\perp A \mid Y$.

する周辺分布 (marginal distribution) とする．また，$\mathcal{X} = \{1, 2, \dots, m\}$ とするとき，$\mathcal{X}$ 上の確率分布の全体を $(m-1)$ 次元確率単体と同一視し，$\Delta_{\mathcal{X}} := \left\{ p = (p_1, \dots, p_m) \in [0,1]^m \,\middle|\, \sum_{i=1}^m p_i = 1 \right\}$ と記す．確率変数 A を意思決定者 (decision maker; DM) の決定（行動）(action, decision) と呼ぶ．このとき，Y が与えられたもとでの A に関する確率分布を $q_{A|Y}$ と記し，確率的決定関数 (randomized decision rule) と呼ぶ．特に，A が Y の関数のとき，その関数を $\delta\colon \mathcal{Y} \to \mathcal{A}$ と記し，決定関数 (deterministic decision rule) と呼ぶ．ここで，$\mathcal{A}$ を決定（行動）空間 (action space, decision space) と呼ぶ．以上で述べた事項を，図 1 にシステムモデルとして表す．意思決定者の決定に対するペナルティやコストを表す関数 $\ell\colon \mathcal{X} \times \mathcal{A} \to \mathbb{R}$ を損失関数 (loss function, cost function)，利得を表す関数 $g\colon \mathcal{X} \times \mathcal{A} \to \mathbb{R}$ を効用関数 (utility function, gain function) とする．すなわち，$X = x$ のときに決定 $A = a$ をしたときの損失 (resp. 効用) を $\ell(x,a)$ (resp. $g(x,a)$) と記す．本稿では，簡単のため，特に断りのない限り，$\mathcal{X}$ および $\mathcal{Y}$ は有限集合とする．X の期待値を $\mathbb{E}_X[X] := \sum_x x p_X(x)$，$Y = y$ の条件のもとでの X に関する条件付き期待値を $\mathbb{E}_X[X \mid Y = y] := \sum_x x p_{X|Y}(x \mid y)$ と記す．また，本稿を通して，log の底は自然対数の底 $e = 2.718\cdots$ とする．

本節では，まず，統計的決定理論の基礎事項について述べる．次に，Shannon の相互情報量をはじめとするさまざまな相互情報量や情報漏えい尺度を統一的に扱うための generalized information leakage measure を導入する．

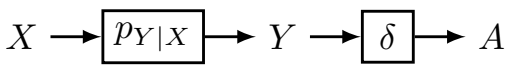

図 1: システムモデル

2.1 統計的決定理論

本節では，統計的決定理論における基本事項について述べる．

定義 2.1. $(X,Y) \sim p_{X,Y}$ とし，$\delta\colon \mathcal{Y} \to \mathcal{A}$ を決定関数とする．このとき，δ に関するベイズリスク（期待損失）$r(\delta)$ およびベイズ期待効用 $G(\delta)$ を以下で定める：

$$r(\delta) = \mathbb{E}_{X,Y}\left[\ell(X, \delta(Y))\right], \tag{1}$$

$$G(\delta) = \mathbb{E}_{X,Y}\left[g(X, \delta(Y))\right]. \tag{2}$$

このベイズリスク $r(\delta)$ を最小化する（あるいはベイズ期待効用 $G(\delta)$ を最大化する）決定関数 δ を求める問題が，統計的決定理論（ベイズ決定理論）における基本的な問題である．次の命題は，最適な決定関数に関するよく知られた結果である．

命題 2.2 ([35, Result 1], [36, Thm 2.7])**.** ベイズリスクの最小値は，以下で与えられる：

$$\min_{\delta} r(\delta) = r(\delta^*) \tag{3}$$

$$= \mathbb{E}_Y\left[\min_{a \in \mathcal{A}} \mathbb{E}_X\left[\ell(X,a) \mid Y\right]\right] \tag{4}$$

$$= \sum_y p_Y(y)\left[\min_{a \in \mathcal{A}} \sum_{x \in \mathcal{X}} p_{X|Y}(x \mid y)\ell(x,a)\right], \tag{5}$$

ここで，最小値を与える決定関数 $\delta^*\colon \mathcal{Y} \to \mathcal{A}$ は以下で定義される：

$$\delta^*(y) := \operatorname*{argmin}_{a \in \mathcal{A}} \mathbb{E}_X\left[\ell(X,a) \mid Y = y\right] \tag{6}$$

$$= \operatorname*{argmin}_{a \in \mathcal{A}} \sum_{x \in \mathcal{X}} p_{X|Y}(x \mid y)\ell(x,a), \tag{7}$$

ここで，$p_{X|Y}(x \mid y) := \frac{p_X(x) p_{Y|X}(y|x)}{\sum_x p_X(x) p_{Y|X}(y|x)}$ は $Y = y$ が与えられたもとでの X の事後確率 (posterior distribution) を表す．同様にして，ベイズ期待効用の最大値は以下のようになる：

$$\max_{\delta} G(\delta) = G(\delta^*) \tag{8}$$

$$= \mathbb{E}_Y\left[\max_{a \in \mathcal{A}} \mathbb{E}_X\left[g(X,a) \mid Y\right]\right], \tag{9}$$

$$\delta^*(y) := \operatorname*{argmax}_{a \in \mathcal{A}} \mathbb{E}_X\left[g(X,a) \mid Y = y\right]. \tag{10}$$

すなわち，ベイズリスクを最小（resp. ベイズ期待効用を最大）にする関数は事後期待損失最小化関数（resp. 事後期待効用最大化関数）である．

例 2.3. 意思決定者が X の推定 $A = \hat{X}$ を決定関数 $\delta\colon \mathcal{Y} \to \mathcal{X}$ を用いて，0-1 損失 $\ell_{0\text{-}1}(x, \hat{x}) = \mathbb{1}_{\{x=\hat{x}\}}$ のもとで行うとする[6]．このとき，ベイズリスクの最小値およびそれを与える決定関数は以下のようになる：

$$\min_{\delta} r(\delta) = 1 - \mathbb{E}_Y \left[\max_{x \in \mathcal{X}} p_{X|Y}(x \mid Y) \right], \tag{11}$$

$$\delta^*(y) = \operatorname*{argmax}_{x \in \mathcal{X}} p_{X|Y}(x \mid y). \qquad \text{(MAP 推定)} \tag{12}$$

注意 2.4. いま，決定空間 $\mathcal{A}$ が与えられたときに，$Y \in \mathcal{Y}$ に基づいて $\mathcal{A}$ 上の確率分布 $q \in \Delta_{\mathcal{A}}$ を決定する問題を新たに考える．このときの決定関数を $\tilde{\delta}\colon \mathcal{Y} \to \Delta_{\mathcal{A}}$ と記し，この決定問題における損失関数（resp. 効用関数）を $\ell(x, q)$ (resp. $g(x, q)$) と表す．また本稿では，$Y = y$ を観測したもとでの確率的決定関数 $q_{A|Y=y}$ に対する損失関数（resp. 効用関数）[35, Def 8] についても（記法の濫用を許容して）$\ell(x, q_{A|Y=y})$ (resp. $g(x, q_{A|Y=y})$) と記すことにする．このとき，$\tilde{\delta}$ に対するベイズリスク $r(\tilde{\delta})$ (resp. $G(\tilde{\delta})$) や確率的決定関数に対するベイズリスク $r(q_{A|Y})$ (resp. $G(q_{A|Y})$) が同様に定義され，両者の最小化問題（resp. 最大化問題）は等価な問題となる．このとき，命題 2.2 は，式 (4)(10) における $\min_{a \in \mathcal{A}}$ を $\inf_{q \in \Delta_{\mathcal{A}}}$，$\max_{a \in \mathcal{A}}$ を $\sup_{q \in \Delta_{\mathcal{A}}}$ に置き換えることで，$\tilde{\delta}$ や $q_{A|Y}$ を用いた決定問題についても同様に成り立つ．

例 2.5. 意思決定者が X の推定 $\hat{X}$ を確率的決定関数 $q_{\hat{X}|Y}$ を用いて，対数損失 $\ell_{\log}(x, q) = -\log q(x)$ のもとで行うとする[7]．このとき，ベイズリスクの最小値およびそれを与える確率的決定関数は以下のようになる：

$$\min_{q_{\hat{X}|Y}} r(q_{\hat{X}|Y}) = H(X \mid Y), \tag{13}$$

$$q^*_{\hat{X}|Y}(x \mid y) = p_{X|Y}(x \mid y), \tag{14}$$

ここで，

$$H(X \mid Y) := -\sum_y p_Y(y) \sum_x p_{X|Y}(x \mid y) \log p_{X|Y}(x \mid y) \tag{15}$$

は条件付きエントロピーを表す．

例 2.6 ([24, Lemma 1])**.** $\alpha > 0$ とする．例 2.5 と同様の設定のもとで，Liao らによって提案された次の α-loss $\ell_\alpha(x, q_{\hat{X}|Y=y})$ [24, Def 3][8] のもとでの推定問題を考える：

$$\ell_\alpha(x, q) = \begin{cases} -\log q(x), & \alpha = 1, \\ \frac{\alpha}{\alpha - 1} \left(1 - q(x)^{\frac{\alpha - 1}{\alpha}}\right), & \alpha \in (0, 1) \cup (1, \infty), \\ 1 - q(x), & \alpha = \infty. \end{cases} \tag{16}$$

このとき，ベイズリスクの最小値およびそれを与える決定関数は以下のようになる：

$$\min_{q_{\hat{X}|Y}} r(q_{\hat{X}|Y}) = \begin{cases} H(X \mid Y), & \alpha = 1, \\ \frac{\alpha}{\alpha - 1} \left(1 - \|p_X\|_\alpha\right), & \alpha \in (0, 1) \cup (1, \infty), \\ 1 - \sum_y p_Y(y) \max_x p_{X|Y}(x \mid y), & \alpha = \infty, \end{cases} \tag{17}$$

$$q^*_{\hat{X}|Y}(x \mid y) = \begin{cases} \frac{p_{X|Y}(x|y)^\alpha}{\sum_x p_{X|Y}(x|y)^\alpha}, & \alpha \in (0, \infty), \\ \begin{cases} 1/\left|\mathrm{MAX}(y)\right|, & x \in \mathrm{MAX}(y), \\ 0, & \text{otherwise}, \end{cases} & \alpha = \infty, \end{cases} \tag{18}$$

ここで，$\|p_X\|_\alpha := \left(\sum_x p_X(x)^\alpha\right)^{\frac{1}{\alpha}}$，
$\mathrm{MAX}(y) := \left\{ \tilde{x} \in \mathcal{X} \,\middle|\, p_{X|Y}(\tilde{x} \mid y) = \max_x p_{X|Y}(x \mid y) \right\}$
である．

例 2.7. $A = \hat{X}$（X の点推定問題）あるいは $A \in \Delta_{\mathcal{X}}$（確率分布の推定問題，$X$ の randomized decision rule を用いた推定問題）におけるその他の代表的な損失関数と効用関数および最適な決定関数を表 1 に示す．ここで，$\mathrm{MAX} := \{\tilde{x} \in \mathcal{X} \colon p_X(\tilde{x}) = \max_x p_X(x)\}$ であ

[6] $\mathbb{1}_{\{x=\hat{x}\}} := \begin{cases} 1, & x \neq \hat{x}, \\ 0, & \text{otherwise}. \end{cases}$

[7] これは決定関数 $\tilde{\delta}\colon \mathcal{Y} \to \Delta_{\mathcal{X}}$ を用いて，X の確率分布を決定する問題と等価である．

[8] $\alpha = 1$ のときが対数損失に相当する．

り，p_{X_α} および $p_{X_{\alpha|Y}}$ は以下で定義される α-tilted distribution [24] を表す：

$$p_{X_\alpha}(x) := \frac{p_X(x)^\alpha}{\sum_x p_X(x)^\alpha}, \tag{19}$$

$$p_{X_\alpha|Y}(x \mid y) := \frac{p_{X|Y}(x \mid y)^\alpha}{\sum_x p_{X|Y}(x \mid y)^\alpha}. \tag{20}$$

2.2 Generalized Information Leakage Measure

本節ではまず，これまでに提案されたさまざまな相互情報量や情報漏えい尺度を一般化し，サンプル Y が含むサンプル X に関する情報量を表す *generalized information leakage measure* $\mathcal{L}(X \to Y)$ を導入する．本研究では，既存の情報量が満たす 3 つの性質 1) 非負性，2) データ処理不等式 (Data Processing Inequality; DPI) および 3)（独立性 $\Rightarrow$ zero leakage）に着目し，これらを公理として採用する．

定義 2.8 (Generalized information leakage measure, [32, Def 3])．$(X, Y) \sim p_X p_{Y|X}$ とする．このとき，以下を満たす汎関数 $\mathcal{L}(p_X, p_{Y|X})$ を *generalized information leakage measure* と呼び，$\mathcal{L}(X \to Y)$ と記す．

1) 非負性：

$$\mathcal{L}(X \to Y) \geq 0. \tag{21}$$

2) DPI：任意の確率変数 Z に対して，$X - Y - Z$ が Markov 連鎖をなすならば

$$\mathcal{L}(X \to Z) \leq \mathcal{L}(X \to Y). \tag{22}$$

3) 独立性 $\Rightarrow$ zero leakage：

$$X \perp\!\!\!\perp Y \Longrightarrow \mathcal{L}(X \to Y) = 0. \tag{23}$$

注意 2.9. $\mathcal{L}(X \to Y)$ に対して，3) の逆は仮定しない．すなわち，

3′) zero leakage $\Longrightarrow$ 独立性

$$X \perp\!\!\!\perp Y \Longrightarrow \mathcal{L}(X \to Y) = 0. \tag{24}$$

は仮定しない．

例 2.10. 表 2 に情報理論分野における主要な相互情報量および情報漏えい尺度を示す．この表において，$\alpha \in (0,1) \cup (1,\infty)$ は調節可能なパラメータを表し，$H_\alpha(X) := \frac{\alpha}{1-\alpha} \log \left(\sum_x p_X(x)^\alpha\right)^{\frac{1}{\alpha}}$ は次数 α の Rényi エントロピー，$D_\alpha(p||q) := \frac{1}{\alpha-1} \log \left(\sum_z p^\alpha(z) q^{1-\alpha}(z)\right)$ は次数 α の Rényi divergence，U は X の任意の（randomized なものを含む）関数，$\hat{U}$ はその推定量，$D_f(p||q) := \sum_{z \in \mathcal{Z}} q(z) f\left(\frac{p(z)}{q(z)}\right)$ は f-divergence をそれぞれ表す．ここで，$f\colon [0,\infty) \to \mathbb{R}$ は $[0,\infty)$ 上の凸関数で $f(1) = 0$ かつ $t = 1$ において狭義凸であるとする．

注意 2.11. 表 2 で定義される多くの information leakage measure は，Shannon 相互情報量を特別な場合として含む．実際，次が成り立つ：

$$\begin{aligned} I(X;Y) &= \lim_{\alpha \to 1} I_\alpha^{\mathrm{A}}(X;Y) = \lim_{\alpha \to 1} I_\alpha^{\mathrm{S}}(X;Y) \\ &= \lim_{\alpha \to 1} I_\alpha^{\mathrm{C}}(X;Y) = \lim_{\alpha \to 1} I_\alpha^{\mathrm{LP}}(X;Y) \\ &= \lim_{\alpha \to 1} \mathcal{L}_\alpha^{\max}(X \to Y), \end{aligned} \tag{25}$$

$$I(X;Y) = I_f(X;Y) = \mathcal{L}_f(X \to Y), \tag{26}$$

ただし，$f(t) = t \log t$. また，$I_\infty^{\mathrm{S}}(X;Y) = \mathcal{L}_{\mathrm{maxL}}(X \to Y)$ であることが文献 [19, Thm 1] で示されており，$I_\alpha^{\mathrm{A}}(X;Y) = \mathcal{L}_\alpha(X \to Y)$ であることが文献 [24, Thm 1] で示されている．

注意 2.12. 多くの場合，1) 非負性は，$D_\alpha(p||q)$ や $D_f(p||q)$ の非負性から従う．α-leakage $\mathcal{L}_\alpha(X \to Y)$ の性質は，それと等価な Arimoto 相互情報量 $I_\alpha^{\mathrm{A}}(X;Y)$ の性質から従う．

3 情報の価値理論

本節では，2.2 節で導入した generalized information leakage measure $\mathcal{L}(X \to Y)$ 制約のもとでのサンプル Y に対する情報の価値を定義する．そのためにまず，無制約のもとでの Y の価値の定量化である EVSI および average ratio gain を導入する．

表 1. X の推定問題におけるさまざまな損失関数と効用関数

$\ell(x,a)$, $g(x,a)$	$\operatorname{argmin}_a \mathbb{E}_X[\ell(X,a)]$ $= \operatorname{argmax}_a \mathbb{E}_X[g(X,a)]$	$\min_a \mathbb{E}_X[\ell(X,a)]$, $\max_a \mathbb{E}_X[g(X,a)]$	$\operatorname{argmin}_\delta \mathbb{E}_{X,Y}[\ell(X,\delta(Y))]$ $= \operatorname{argmax}_\delta \mathbb{E}_{X,Y}[g(X,\delta(Y))]$	$\min_\delta \mathbb{E}_{X,Y}[\ell(X,\delta(Y))]$, $\max_\delta \mathbb{E}_{X,Y}[g(X,\delta(Y))]$
$\mathbb{1}_{\{x=\hat{x}\}}$ (0-1-loss), $1-\mathbb{1}_{\{x=\hat{x}\}}$	$\operatorname{argmax}_x p_X(x)$	$1-\max_x p_X(x)$, $\max_x p_X(x)$	$\operatorname{argmax}_x p_{X\mid Y}(x \mid y)$ (MAP estimation)	$1-\mathbb{E}_Y\left[\max_x p_{X\mid Y}(x \mid Y)\right]$, $\mathbb{E}_Y\left[\max_x p_{X\mid Y}(x \mid Y)\right]$
$(x-\hat{x})^2$ (squared-loss), $-(x-\hat{x})^2$	$\mathbb{E}_X[X]$	$\mathbb{V}(X)$, $-\mathbb{V}(X)$	$\mathbb{E}_X[X \mid Y=y]$	$\mathbb{E}_Y[\mathbb{V}(X \mid Y)]$, $-\mathbb{E}_Y[\mathbb{V}(X \mid Y)]$
$-\log q(x)$ (log-loss), $\log q(x)$ (log-score)	p_X	$H(X)$, $-H(X)$	$p_{X\mid Y=y}$	$H(X \mid Y)$, $-H(X \mid Y)$
$\frac{1}{\alpha-1}\left(1-\left(\frac{q(x)}{\|q\|_\alpha}\right)^{\alpha-1}\right)$, $\frac{1}{\alpha-1}\cdot\left(\frac{q(x)}{\|q\|_\alpha}\right)^{\alpha-1}$ (pseudo-spherical score)	p_X	$\frac{1}{\alpha-1}\left(1-\|p_X\|_\alpha\right)$ (Harvda–Tsallis entropy), $\frac{1}{\alpha-1}\cdot\|p_X\|_\alpha$	$p_{X\mid Y=y}$	$\frac{1}{\alpha-1}\left(1-\mathbb{E}_Y\left[\left\|p_{X\mid Y}(\cdot \mid Y)\right\|_\alpha\right]\right)$, $\frac{1}{\alpha-1}\cdot\mathbb{E}_Y\left[\left\|p_{X\mid Y}(\cdot \mid Y)\right\|_\alpha\right]$
$\frac{\alpha}{\alpha-1}\left(1-q(x)^{\alpha-1}\right)+\|q\|_\alpha^\alpha$, $\frac{\alpha}{\alpha-1}\cdot q(x)^{\alpha-1}-\|q\|_\alpha^\alpha$ (power score, Tsallis score)	p_X	$\frac{1}{\alpha-1}(1-\|p_X\|_\alpha^\alpha)$, $\frac{1}{\alpha-1}\cdot\|p_X\|_\alpha^\alpha$	$p_{X\mid Y=y}$	$\frac{1}{\alpha-1}\left(1-\mathbb{E}_Y\left[\left\|p_{X\mid Y}(\cdot \mid Y)\right\|_\alpha^\alpha\right]\right)$, $\frac{1}{\alpha-1}\cdot\mathbb{E}_Y\left[\left\|p_{X\mid Y}(\cdot \mid Y)\right\|_\alpha^\alpha\right]$
$\frac{\alpha}{\alpha-1}\left(1-q(x)^{\frac{\alpha-1}{\alpha}}\right)$ (α-loss), $\frac{\alpha}{\alpha-1}\cdot q(x)^{\frac{\alpha-1}{\alpha}}$	p_{X_α}	$\frac{\alpha}{\alpha-1}\left(1-\|p_X\|_\alpha\right)$, $\frac{\alpha}{\alpha-1}\cdot\|p_X\|_\alpha$	$p_{X_\alpha\mid Y=y}$	$\frac{\alpha}{\alpha-1}\left(1-\mathbb{E}_Y\left[\left\|p_{X\mid Y}(\cdot \mid Y)\right\|_\alpha\right]\right)$, $\frac{\alpha}{\alpha-1}\cdot\mathbb{E}_Y\left[\left\|p_{X\mid Y}(\cdot \mid Y)\right\|_\alpha\right]$
$1-q(x)$ (∞-loss, soft 0-1 loss), $q(x)$	$\begin{cases} 1/\lvert\mathrm{MAX}\rvert, & x\in\mathrm{MAX}, \\ 0, & \text{otherwise} \end{cases}$	$1-\max_x p_X(x)$, $\max_x p_X(x)$	$\begin{cases} 1/\lvert\mathrm{MAX}(y)\rvert, & x\in\mathrm{MAX}(y), \\ 0, & \text{otherwise} \end{cases}$	$1-\mathbb{E}_Y\left[\max_x p_{X\mid Y}(x \mid Y)\right]$, $\mathbb{E}_Y\left[\max_x p_{X\mid Y}(x \mid Y)\right]$

3.1 EVSI と average ratio gain

本節では，意思決定に用いるサンプル Y が持つ情報の価値の定量化として，EVSI および average ratio gain を定義する．これらはそれぞれ，サンプルを使わずに最適な意思決定を行うときと比べたときに，サンプル Y を用いて最適な意思決定を行うことがベイズリスク (resp. 期待効用) の意味でどのくらい有用であるのかをそれらの差 (difference) および比 (ratio) で定量化したものである．

定義 3.1 (EVSI, average gain, [6, Section 4.5.2])**.** $\ell(x,a)$ を損失関数，$g(x,a)$ を効用関数とする．このとき，以下で定義される量 $\mathsf{gain}^{(\cdot)}(X;Y)$ を *Expected value of sample information* (EVSI) あるいは *average gain* と呼ぶ：

$$\mathsf{gain}^{\ell}(X;Y) := \min_a \mathbb{E}\left[\ell(X,a)\right] - \min_\delta \mathbb{E}_{X,Y}\left[\ell(X,\delta(Y))\right] \tag{27}$$

$$= \min_a \mathbb{E}_X\left[\ell(X,a)\right] - \mathbb{E}_Y\left[\min_a \mathbb{E}_X\left[\ell(X,a) \mid Y\right]\right], \tag{28}$$

$$\mathsf{gain}^{g}(X;Y) := \max_\delta \mathbb{E}_{X,Y}\left[g(X,\delta(Y))\right] - \max_a \mathbb{E}\left[g(X,a)\right] \tag{29}$$

$$= \mathbb{E}_Y\left[\max_a \mathbb{E}_X\left[g(X,a) \mid Y\right]\right] - \max_a \mathbb{E}_X\left[g(X,a)\right]. \tag{30}$$

定義 3.2 (average ratio gain, [33, Def 8])**.** $\ell(x,a)$ を損失関数，$g(x,a)$ を効用関数とする．このとき，以下で定義される量 $\mathsf{Rgain}^{(\cdot)}(X;Y)$ を *average ratio gain* と呼ぶ：

$$\mathsf{Rgain}^{\ell}(X;Y) := \log \frac{\min_a \mathbb{E}\left[\ell(X,a)\right]}{\min_\delta \mathbb{E}_{X,Y}\left[\ell(X,\delta(Y))\right]} \tag{31}$$

$$= \log \frac{\min_a \mathbb{E}_X\left[\ell(X,a)\right]}{\mathbb{E}_Y\left[\min_a \mathbb{E}_X\left[\ell(X,a) \mid Y\right]\right]}, \tag{32}$$

$$\mathsf{Rgain}^{g}(X;Y) := \log \frac{\max_\delta \mathbb{E}_{X,Y}\left[g(X,\delta(Y))\right]}{\max_a \mathbb{E}\left[g(X,a)\right]} \tag{33}$$

$$= \log \frac{\mathbb{E}_Y\left[\max_a \mathbb{E}_X\left[g(X,a) \mid Y\right]\right]}{\max_a \mathbb{E}_X\left[g(X,a)\right]}. \tag{34}$$

注意 3.3. 近年，情報セキュリティ分野において，Alvim らによって additive leakage および g-leakage [27, 28, 29] が提案されているが，これらはそれぞれ $\mathsf{gain}^{g}(X;Y)$ および $\mathsf{Rgain}^{g}(X;Y)$ に相当する量である．

例 3.4. X の点推定問題（$A=\hat{X}$）を考える．表 1 に示した $\ell(x,a), g(x,a)$ の内，$\ell_{\log}(x,q) := -\log q(x)$,

表 2. 情報理論における代表的な相互情報量および情報漏えい尺度

Name	Definition	1)	2)	3′)
Shannon MI [37]	$I(X;Y) := H(X) - H(X \mid Y)$	✓	✓[38, Thm 2.8.1]	✓[38, Eq (2.90)]
Arimoto MI of order α [39]	$I_\alpha^{\mathrm{A}}(X;Y) := H_\alpha(X) - H_\alpha^{\mathrm{A}}(X \mid Y)$	✓[39, Thm 2]	✓[24, Footnote 4], [40, Cor 1]	✓[39, Thm 2]
Arimoto MI of order ∞ [24]	$I_\infty^{\mathrm{A}}(X;Y) := \log \frac{\sum_y \max_x p_{X,Y}(x,y)}{\max_x p_X(x)}$	✓	✓[40, Cor 1]	✗[41, Sec 6.6]
Sibson MI of order α [13]	$I_\alpha^{\mathrm{S}}(X;Y) := \min_{q_Y} D_\alpha(p_{X,Y} \| p_X q_Y)$	✓	✓[42, Eq (55)]	✓
Sibson MI of order ∞ [24]	$I_\infty^{\mathrm{S}}(X;Y) := \log \sum_y \max_x p_{Y\mid X}(y \mid x)$	✓	✓	✓
Augustin–Csiszár MI of order α [43], [15]	$I_\alpha^{\mathrm{C}}(X;Y) := \min_{q_Y} \mathbb{E}_X\left[D_\alpha(p_{Y\mid X}(\cdot \mid X) \| q_Y)\right]$	✓	✓ [15, Eq (22)]	✓
Lapidoth–Pfister MI of order α [16], [17]	$I_\alpha^{\mathrm{LP}}(X;Y) := \min_{q_X} \min_{q_Y} D_\alpha(p_X p_{Y\mid X} \| q_X q_Y)$	✓	✓ [16, Lemma 2]	✓ [16, Lemma 4]
f-information [44]	$I_f(X;Y) := D_f(p_{X,Y} \| p_X p_Y)$	✓	✓ [45, Thm 7.9]	✓ [46, Lem 4], [45, Thm 7.3]
f-leakage [24]	$\mathcal{L}_f(X \to Y) := \min_{q_Y} D_f(p_{X,Y} \| p_X q_Y)$	✓	✓	✓
maximal leakage [19]	$\mathcal{L}_{\max\mathrm{L}}(X \to Y) := \sup_{U-X-Y} \log \frac{\max_{q_{\hat{U}\mid Y}} \mathbb{E}_{U,Y}\left[q_{\hat{U}\mid Y}(U\mid Y)\right]}{\max_u q(u)}$	✓ [19, Lem 1]	✓ [19, Lem 1]	✓ [19, Cor 2]
α-leakage [24]	$\mathcal{L}_\alpha(X \to Y) := \frac{\alpha}{\alpha-1} \log \frac{\max_{q_{\hat{X}\mid Y}} \mathbb{E}_{X,Y}\left[q_{\hat{X}\mid Y}(X\mid Y)^{\frac{\alpha-1}{\alpha}}\right]}{\max_q \mathbb{E}_X\left[q(X)^{\frac{\alpha-1}{\alpha}}\right]}$	✓	✓	✓
maximal α-leakage [24]	$\mathcal{L}_\alpha^{\max}(X \to Y) := \sup_{U-X-Y} \mathcal{L}_\alpha(U \to Y)$	✓	✓ [24, Thm 3]	✓

$g_{\mathrm{PS}}(x,q) := \frac{1}{\alpha-1} \cdot \left(\frac{q(x)}{\|q\|_\alpha}\right)^{\alpha-1}$, $g_\alpha(x,q) := \frac{\alpha}{\alpha-1} \cdot q(x)^{\frac{\alpha-1}{\alpha}}$ については，その EVSI あるいは average ratio gain は，表 2 に示した information leakage と次のような対応がある:

$$\mathsf{gain}^{\ell_{\log}}(X;Y) = I(X;Y), \tag{35}$$

$$\mathsf{Rgain}^{g_{\mathrm{PS}}}(X;Y) = \mathsf{Rgain}^{g_\alpha}(X;Y) = \frac{\alpha-1}{\alpha} I_\alpha^{\mathrm{A}}(X;Y). \tag{36}$$

さらに，$g_{\mathrm{Power}}(x,q) := \frac{\alpha}{\alpha-1} \cdot q(x)^{\alpha-1} - \|q\|_\alpha^\alpha, \alpha > 1$ については，

$$\mathsf{Rgain}^{g_{\mathrm{Power}}}(X;Y) = (\alpha-1)\left(H_\alpha(X) - H_\alpha^{\mathrm{H}}(X \mid Y)\right) \tag{37}$$

と表すことができる．ここで，

$$H_\alpha^{\mathrm{H}}(X \mid Y) := \frac{1}{1-\alpha} \log \mathbb{E}_Y\left[\left\|p_{X\mid Y}(\cdot \mid Y)\right\|_\alpha^\alpha\right] \tag{38}$$

は，次数 α の Hayashi 条件付きエントロピー [47, Section II.A] を表す．

命題 3.5. 任意の $\ell(x,a), u(x,a)$ に対して，$\mathsf{gain}^{(\cdot)}(X;Y)$ および $\mathsf{Rgain}^{(\cdot)}(X;Y)$ は，定義 2.8 で導入した generalized information leakage measure の条件 1), 2), 3) を満たす．

Proof. 1) 非負性については定義から自明に成り立ち，2) DPI については，Américo らによる *core-concavity* に関する結果 [48, Thm 2, Sec V.F] から従う．3)（独立性 ⇒ zero leakage）については，$\mathcal{L}(X \to Y)$ の性質 3) から従う． □

一方，3′)（zero leakage ⇒ 独立性）が成り立つかどうかは，仮定する確率分布 $p_{X,Y}$ および $\ell(x,a)$ や $u(x,a)$ に依存する．実際，次が成り立つ．

命題 3.6 ([33, Prop 2])**.** $A = \hat{X}, \ell_{\mathrm{sq}}(x,\hat{x}) := (x-\hat{x})^2$ とする．このとき，$\mathsf{gain}^{\ell_{\mathrm{sq}}}(X;Y) = \mathbb{V}(X) - \mathbb{E}_Y\left[\mathbb{V}(X \mid Y)\right]$ および $\mathsf{Rgain}^{\ell_{\mathrm{sq}}}(X;Y) = \log \frac{\mathbb{V}(X)}{\mathbb{E}_Y[\mathbb{V}(X\mid Y)]}$ は一般には 3′)（zero leakage ⇒ 独立性）を満たさない．

Proof. $\mathsf{gain}^{\ell_{\mathrm{sq}}}(X;Y)$ についてのみ示す．

全分散の公式[9] より，

$$\mathsf{gain}^{\ell_{\mathrm{sq}}}(X;Y) = 0 \tag{39}$$

$$\iff \mathbb{V}\left(\mathbb{E}_Y\left[X \mid Y\right]\right) = 0 \tag{40}$$

$$\iff \mathbb{E}_Y\left[\left(\mathbb{E}_X\left[X \mid Y\right] - \mathbb{E}_X[X]\right)^2\right] = 0 \tag{41}$$

$$\iff \mathbb{E}_X\left[X \mid Y\right] = \mathbb{E}_X[X] \ \ a.s. \tag{42}$$

[9] $\mathbb{V}(X) = \mathbb{E}_Y\left[\mathbb{V}(X \mid Y)\right] + \mathbb{V}\left(\mathbb{E}_X\left[X \mid Y\right]\right)$.

最後の式は，*mean independence* と呼ばれ，一般には，独立性よりも弱い条件である．□

3.2 A Generalization of the Value of Information

本節では，2.2 節で導入した generalized information leakage measure と 3.1 節で導入した EVSI および average ratio gain に基づき，サンプル Y のもつ X に関する情報の価値を表す量として，Staratonovich の VoI を一般化した量を定義する．

定義 3.7 (generalized VoI, [32, Def 6])．$\ell(x,a)$ を損失関数，$g(x,a)$ を効用関数とし，$\mathcal{L}(X \to Y)$ を generalized information leakage measure とする．このとき，以下で定義される量を generalized value of information (generalized VoI) と呼ぶ：

$$\mathsf{V}^{\ell}_{\mathcal{L}}(R;\mathcal{Y}) := \sup_{\substack{p_{Y|X}:\\ \mathcal{L}(X\to Y)\le R}} \mathsf{gain}^{\ell}(X;Y) \tag{43}$$

$$= \min_{a} \mathbb{E}_X\left[\ell(X,a)\right] - \inf_{\substack{p_{Y|X}:\\ \mathcal{L}(X\to Y)\le R}} \mathbb{E}_Y\left[\min_{a} \mathbb{E}_X\left[\ell(X,a)\mid Y\right]\right], \tag{44}$$

$$\mathsf{V}^{\ell,\mathsf{R}}_{\mathcal{L}}(R;\mathcal{Y}) := \sup_{\substack{p_{Y|X}:\\ \mathcal{L}(X\to Y)\le R}} \mathsf{Rgain}^{\ell}(X;Y) \tag{45}$$

$$= \log \min_{a} \mathbb{E}_X\left[\ell(X,a)\right] - \log \inf_{\substack{p_{Y|X}:\\ \mathcal{L}(X\to Y)\le R}} \mathbb{E}_Y\left[\min_{a} \mathbb{E}_X\left[\ell(X,a)\mid Y\right]\right]. \tag{46}$$

$\mathsf{V}^{g}_{\mathcal{L}}(R;\mathcal{Y})$ および $\mathsf{V}^{g,\mathsf{R}}_{\mathcal{L}}(R;\mathcal{Y})$ についても，同様に定義する．

注意 3.8. Stratonovich は文献 [7, 8] において，$\mathcal{L}(X \to Y) = I(X;Y)$ の場合の VoI を定義している．

4 主結果と PUT 問題への応用

本節では，3.2 節で定義した generalized VoI $\mathsf{V}^{(\cdot)}_{\mathcal{L}}(R;\mathcal{Y})$ および $\mathsf{V}^{(\cdot),\mathsf{R}}_{\mathcal{L}}(R;\mathcal{Y})$ の達成可能な上界に関する結果を導出し，その PUT 問題としての解釈について述べる．

4.1 Generalized VoI の達成可能な上界

定義 4.1. $\ell(x,a)$ を損失関数とし，上界関数 $\mathsf{V}^{\ell}_{\mathcal{L}}(R)$ を以下で定める：

$$\mathsf{V}^{\ell}_{\mathcal{L}}(R) := \min_{a} \mathbb{E}_X\left[\ell(X,a)\right] - \inf_{\substack{p_{A|X}:\\ \mathcal{L}(X\to A)\le R}} \mathbb{E}_{X,A}\left[\ell(X,A)\right], . \tag{47}$$

$\mathsf{V}^{\ell,\mathsf{R}}_{\mathcal{L}}(R), \mathsf{V}^{g}_{\mathcal{L}}(R), \mathsf{V}^{g,\mathsf{R}}_{\mathcal{L}}(R)$ についても，同様に定義する．

定理 4.2 ([32, Thm 1])．任意のアルファベット $\mathcal{Y}$ と $R \ge 0$ に対して，

$$\mathsf{V}^{\ell}_{\mathcal{L}}(R;\mathcal{Y}) \le \mathsf{V}^{\ell}_{\mathcal{L}}(R). \tag{48}$$

さらに，$p^*_{A|X} = \operatorname{arginf}_{p_{A|X}:\, \mathcal{L}(X\to A)\le R} \mathbb{E}_{X,A}\left[\ell(X,A)\right]$ として，$\tilde{A} \in \mathcal{A}$ を $p^*_{A|X}$ から誘導される確率分布 p^*_A[10] に従う確率変数とする．また，$t(\tilde{A})$ を X に関する十分統計量とし，その値域を $t(\mathcal{A})$ とする．$\mathcal{Y} = t(\mathcal{A})$ のとき，式 (48) の等号は，以下で定義される $p^*_{Y|X}$ で達成される：

$$p^*_{Y|X}(y \mid x) := \sum_{a} p^*_{A|X}(a \mid x) \mathbb{1}_{\{y = t(a)\}}. \tag{49}$$

Proof. See Appendix A. □

注意 4.3. 定理 4.2 は，以下のように表現することができる：

$$\sup_{\mathcal{Y}} \mathsf{V}^{\ell}_{\mathcal{L}}(R;\mathcal{Y}) = \mathsf{V}^{\ell}_{\mathcal{L}}(R). \tag{50}$$

また，$\mathsf{V}^{\ell,\mathsf{R}}_{\mathcal{L}}(R;\mathcal{Y}), \mathsf{V}^{g}_{\mathcal{L}}(R;\mathcal{Y}), \mathsf{V}^{g,\mathsf{R}}_{\mathcal{L}}(R;\mathcal{Y})$ についても，同様の結果が成り立つ．

注意 4.4. $\mathcal{L}(X \to Y) = I(X;Y)$ のとき，Stratonovich はこの達成可能な上界 $\mathsf{V}^{\ell}_{I}(R)$ を *Value of Shannon's Information* と名付けた [11, Section 9.3]．

ここで，式 (48) の等号成立条件に関して，本研究で得られた結果が従来の条件の拡張になっていることを示す．いま $\mathcal{X} := \{1, 2, \ldots, m\}$ とおく．Stratonovich のオリジナルの結果では，$\mathcal{Y} = \Delta_{\mathcal{X}}$ として，$Y :=$

[10] $p^*_A(a) := \sum_x p_X(x) p^*_{A|X}(a \mid x)$.

$(p^*_{X|A}(1 \mid \tilde{A}), p^*_{X|A}(2 \mid \tilde{A}), \ldots, p^*_{X|A}(m \mid \tilde{A})) \in \Delta_{\mathcal{X}}$ を等号成立条件として示している．また，Raginsky は，$\mathcal{Y} = \mathcal{A}, Y = \tilde{A}$ の場合を等号成立条件として示している．これらの結果は，（適当な仮定のもとで）定理 4.2 における等号成立条件の特別な場合として表現できる．すなわち，次が成り立つ．

命題 4.5 ([32, Prop 4])．$t(\tilde{A}) := \tilde{A}$ は X に関する十分統計量である．さらに，$\{p_{A|X}(\cdot \mid x)\}_{x \in \mathcal{X}}$ が共通の台 (support) を持つとすると，$t(\tilde{A}) := (p_{X|A}(1|\tilde{A}), p_{X|A}(2|\tilde{A}), \ldots, p_{X|A}(m \mid \tilde{A}))$ は X に関する十分統計量である．

Proof. See [33, Appendix C]. □

4.2 達成可能な上界の性質

以下では，定理 4.2 における達成可能な上界の性質を示す．特に，$\mathsf{V}^{\ell}_{\mathcal{L}}(R)$ に限って示すが，同様の性質は他の上界についても成立する．

命題 4.6 ([32, Prop 5])．

1. $\mathsf{V}^{\ell}_{\mathcal{L}}(0) \geq 0$. 特に，$\mathcal{L}(X \to Y)$ が性質 3′)（zero leakage $\Rightarrow$ 独立性）をもつならば，$\mathsf{V}^{\ell}_{\mathcal{L}}(0) = 0$.
2. $\mathsf{V}^{\ell}_{\mathcal{L}}(R)$ は R について単調増加.
3. p_X を固定したときに $\mathcal{L}(X \to Y) = \mathcal{L}(p_X, p_{Y|X})$ が凸関数（resp. 準凸関数）ならば，$\mathsf{V}^{\ell}_{\mathcal{L}}(R)$ は凹関数（resp. 準凹関数）．
4. $\mathcal{L}_1(X \to Y), \mathcal{L}_2(X \to Y)$ を generalized information leakage measure とする．ある $c > 0$ が存在して，$\mathcal{L}_1(X \to Y) \leq c\mathcal{L}_2(X \to Y)$ ならば，

$$\mathsf{V}^{\ell}_{\mathcal{L}^{(2)}}(R) \leq \mathsf{V}^{\ell}_{\mathcal{L}^{(1)}}(cR), \tag{51}$$

$$\mathsf{V}^{\ell}_{\mathcal{L}^{(2)}}(R/c) \leq \mathsf{V}^{\ell}_{\mathcal{L}^{(1)}}(R). \tag{52}$$

特に，$c = 1$ のとき，すなわち，$\mathcal{L}_1(X \to Y) \leq \mathcal{L}_2(X \to Y)$ ならば

$$\mathsf{V}^{\ell}_{\mathcal{L}^{(2)}}(R) \leq \mathsf{V}^{\ell}_{\mathcal{L}^{(1)}}(R). \tag{53}$$

Proof. 1. のみ示す．まず，定理 4.2 の前半より，任意の $R \geq 0$ について $0 \leq \mathsf{V}^{\ell}_{\mathcal{L}}(R; \mathcal{Y}) \leq \mathsf{V}^{\ell}_{\mathcal{L}}(R)$ であることがわかる．次に，$\mathcal{L}(X \to Y)$ のもつ性質 3)（独立性 $\Rightarrow$ zero leakage）より，

$$\mathsf{V}^{\ell}_{\mathcal{L}}(0) = \min_a \mathbb{E}_X\left[\ell(X,a)\right] - \inf_{\substack{p_{A|X}\,:\\ \mathcal{L}(X \to A) \leq 0}} \mathbb{E}_{X,A}\left[\ell(X,A)\right] \tag{54}$$

$$= \min_a \mathbb{E}_X\left[\ell(X,a)\right] - \mathbb{E}^{p_X p_A}_{X,A}\left[\ell(X,A)\right] \tag{55}$$

$$\overset{(a)}{\leq} 0, \tag{56}$$

ここで (a) は (min) $\leq$ (average) の不等式から従う．ゆえに，$\mathsf{V}^{\ell}_{\mathcal{L}}(0) = 0$. 他の性質については，レート歪み理論におけるレート歪み関数の性質と同様に示せる (see, e.g.[49, Chapter 9], [33, Appendix D]) □

注意 4.7. 命題 4.6 の性質 1 は，性質 3′)（zero leakage $\Longrightarrow$ 独立性）を持つ $\mathcal{L}(X \to Y)$ については，$\mathcal{L}(X \to Y) = 0$ である Y を意思決定に用いることの gain は 0 になることを意味し，また，この性質 3′) を持たない $\mathcal{L}(X \to Y)$ を用いて leakage を測る場合，$\mathcal{L}(X \to Y) = 0$ であっても，この Y を利用して意思決定者にとっての gain が得られる可能性があることを意味する．命題 4.6 の性質 4 は，より強い information leakage 制約を課すほど，意思決定における gain は減少することを意味している．

例 4.8. Shannon 相互情報量 $I(X;Y) = I(p_X, p_{Y|X})$ は p_X を固定したときに $p_{Y|X}$ について凸 (see, e.g., [38, Thm 2.7.4]) なので，value of Shannon's information $\mathsf{V}^{\ell}_{I}(R)$ は R について凹関数である．図 2 に $\mathsf{V}^{\ell}_{I}(R)$ の概形を示す．

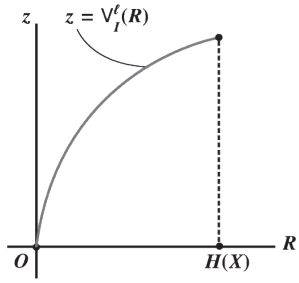

図 2: Value of Shannon's information

4.3　PUT 問題への応用

本節では，情報の価値理論を PUT 問題へ応用することを考える．具体的には，定理 4.2 で得られた結果の PUT 問題における解釈について説明する．

いま，データ $X = x$ の保有者 (Alice)，公開データ Y の正規の利用者 (Bob) および公開データ Y から個人情報に関する情報を得ようとする攻撃者 (Eve) を想定する．Alice は，privacy mechanism（通信路）$p_{Y|X}$ を用いて，Eve に対するプライバシ情報の漏えいを保護しつつ，Bob にとっての有用性を高めるような変換を X に施し，データ Y をパブリックに公開することを考える．このとき，プライバシ保護尺度は，Alice が任意に選ぶ generalized information leakage measure $\mathcal{L}(X \to Y)$ で測るものとする．Bob の公開データ Y に対する利用目的は，決定関数 $\delta\colon \mathcal{Y} \to \mathcal{A}, A := \delta(Y)$ および損失関数 $\ell(x,a)$（あるいは効用関数 $g(x,a)$）でモデル化されるものとし，Bob はベイズリスク（resp. ベイズ期待効用）の意味で最適な決定関数を用いた決定を行うと仮定する．さらに，Alice は Bob の公開データ Y の利用目的（すなわち，δ と $\ell(x,a)$）を事前に知っていると仮定する．

以上の想定のもとで，定理 4.2 は次のことを述べている：

Alice がプライバシ制約 $\mathcal{L}(X \to Y) \le R$ のもとで，Bob にとっての有用性 $\mathsf{gain}^{\ell}(X;Y)$ を最大化するためには，以下のステップで Y を公開すれば良い：

1. まず，次の最適化問題を解くことにより，$p^*_{A|X}$ を求める：

$$p^*_{A|X} = \operatorname*{arginf}_{p_{A|X}\colon \mathcal{L}(X\to A)\le R} \mathbb{E}_{X,A}\left[\ell(X,A)\right]. \tag{57}$$

2. $\tilde{A} \sim p^*_{A|X=x}$ を生成する
3. X の十分統計量 $Y := t(\tilde{A})$ を公開する

注意 4.9. 式 (57) の最適化問題は，レート歪み理論における歪みレート関数の計算問題に相当しているとみなせる．したがって，特に，$\mathcal{L}(X \to Y) = I(X;Y)$ のときには，Arimoto–Blahut アルゴリズム [14],[50] が適用できる [34].

5　おわりに

本論文では，一般化された情報漏えい尺度に対する情報の価値理論に関する，著者によるこれまでの研究を概説した．一般理論の構築にあたっては，これまでに提案されているすべての相互情報量に共通する性質（非負性，DPI，独立性 $\Rightarrow$ zero leakage）を持つ generalized information leakage measure を導入し，従来の EVSI に加えて，average ratio gain を導入し，既存の相互情報量との関係を示した．さらに，Stratonovich [8, 11] の解析手法を基にして，一般的な情報量に対する情報の価値に対する上界 $\mathsf{V}^{\ell}_{\mathcal{L}}(R)$ およびその達成可能条件を示した．さらに，この結果をプライバシ情報公開問題におけるプライバシ・ユーティリティ・トーレードオフ問題として考えた場合の解釈について述べた．

今後の研究としては，次の課題が挙げられる：

1. $\mathcal{L}(X \to Y) = I(X;Y)$ 以外の information leakage measure を考えた場合の $\mathsf{V}^{\ell}_{\mathcal{L}}(R)$ を達成する $p^*_{A|X}$ を求めるアルゴリズムの構築
2. EVSI および average ratio gain と既存の相互情報量との対応を利用した，通信路容量の計算アルゴリズムの提案
3. 情報の価値理論の他分野への適用

この内，1. については，$\mathcal{L}(X \to Y) = I^{\mathrm{A}}_{\alpha}(X;Y)$, $I^{\mathrm{S}}_{\alpha}(X;Y), I^{\mathrm{C}}_{\alpha}(X;Y)$ の場合に Arimoto–Blahut アルゴリズムの適用を試みた場合の考察が [34] で述べられている．2. については，例 3.4 で示した対応を利用した Arimoto Capacity $C^{\mathrm{A}}_{\alpha} := \max_{p_X} I^{\mathrm{A}}_{\alpha}(X;Y)$ および Sibson Capacity $C^{\mathrm{S}}_{\alpha} := \max_{p_X} I^{\mathrm{S}}_{\alpha}(X;Y)$ を計算する新たなアルゴリズムに関する結果がすでに得られている [51],[52].

参考文献

[1] C. E. Shannon, “A mathematical theory of communication,” vol. 27, no. 3, pp. 379–423, Jul. 1948.

[2] A. Wald, “Statistical Decision Functions,” *The Annals of Mathematical Statistics*, vol. 20, no. 2, pp. 165 – 205, 1949.

[3] J. von Neumann and O. Morgenstern, *Theory of games and economic behavior*. Princeton University Press, 1947.

[4] L. Savage, *The Foundations of Statistics*. New York: Wiley, 1954.

[5] R. A. Howard, "Information value theory," *IEEE Transactions on Systems Science and Cybernetics*, vol. 2, no. 1, pp. 22–26, 1966.

[6] H. Raiffa and R. Schlaifer, *Applied Statistical Decision Theory*, ser. Harvard Business School Publications. Division of Research, Graduate School of Business Adminitration, Harvard University, 1961.

[7] R. Stratonovich, "On value of information," *Izvestiya of USSR Academy of Sciences, Technical Cybernetics*, vol. 5, pp. 3–12, 1965.

[8] ——, **ТеориЯ Информации**. Москва Советское Радио, 1975.

[9] M. Gavurin, "On the value of information," *Yestnik Leningrad University Series, 4, 27-34; translation (1968), Selected Translations in Mathematical Statistics and Probability*, vol. 7, pp. 193–202, 1963.

[10] A. Perez, "Risk estimates in term of generalized f-entropies," in *Proc. Colloquium on Information Theory (Debrecen, 1967)*, vol. I, no. II. János Bolyai Math. Soc., Budapest, 1968, pp. 299–315.

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

[12] 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.

[13] R. Sibson, "Information radius," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 14, pp. 149–160, 1969.

[14] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 14–20, 1972.

[15] I. Csiszár, "Generalized cutoff rates and renyi's information measures," *IEEE Transactions on Information Theory*, vol. 41, no. 1, pp. 26–34, 1995.

[16] A. Lapidoth and C. Pfister, "Two measures of dependence," *Entropy*, vol. 21, no. 8, 2019. [Online]. Available: https://www.mdpi.com/1099-4300/21/8/778

[17] M. Tomamichel and M. Hayashi, "Operational interpretation of rényi information measures via composite hypothesis testing against product and markov distributions," *IEEE Transactions on Information Theory*, vol. 64, no. 2, pp. 1064–1082, 2018.

[18] 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.

[19] 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.

[20] 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.

[21] I. Issa, S. Kamath, and A. B. Wagner, “Maximal leakage minimization for the shannon cipher system,” in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 520–524.

[22] ——, “An operational measure of information leakage,” in *2016 Annual Conference on Information Science and Systems (CISS)*, March 2016, pp. 234–239.

[23] 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.

[24] 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.

[25] 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.

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

[27] M. S. Alvim, K. Chatzikokolakis, C. Palamidessi, and G. Smith, “Measuring information leakage using generalized gain functions,” in *2012 IEEE 25th Computer Security Foundations Symposium*, 2012, pp. 265–279.

[28] M. S. Alvim, K. Chatzikokolakis, A. Mciver, C. Morgan, C. Palamidessi, and G. Smith, “Additive and multiplicative notions of leakage, and their capacities,” in *2014 IEEE 27th Computer Security Foundations Symposium*, 2014, pp. 308–322.

[29] M. S. Alvim, K. Chatzikokolakis, A. McIver, C. Morgan, C. Palamidessi, and G. Smith, “Axioms for information leakage,” in *2016 IEEE 29th Computer Security Foundations Symposium (CSF)*, 2016, pp. 77–92.

[30] G. R. Kurri, L. Sankar, and O. Kosut, “An operational approach to information leakage via generalized gain functions,” 2022.

[31] 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.

[32] A. Kamatsuka, T. Yoshida, and T. Matsushima, “A generalization of the stratonovich’s value of information and application to privacy-utility trade-off,” in *2022 IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 1999–2004.

[33] ——, “A generalization of the stratonovich’s value of information and application to privacy-utility trade-off,” 2022. [Online]. Available: https://arxiv.org/abs/2201.11449

[34] A. Kamatsuka, T. Yoshida, K. Kazama, and T. Matsushima, “An algorithm for computing the stratonovich’s value of information,” in *2020 International Symposium on Information Theory and Its Applications (ISITA)*, 2022, pp. 98–102.

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

[36] J. K. Ghosh, *An introduction to Bayesian analysis : theory and methods / Jayanta K. Ghosh, Mohan Delampady, Tapas Samanta.*, ser. Springer texts in statistics. New York: Springer.

[37] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

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

[39] 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.

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

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

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

[43] U. Augustin, “Noisy channels,” Ph.D. dissertation, Habilitation thesis, Universitä Erlangen-Nürnberg, 1978.

[44] 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.

[45] Y. Polyanskiy. (2020) Information theory methods in statistics and computer science, lecture 1: f-divergences.

[46] I. Csiszár and P. C. Shields, “Information theory and statistics: A tutorial,” *Commun. Inf. Theory*, vol. 1, no. 4, pp. 417–528, Dec. 2004.

[47] M. Hayashi, “Exponential decreasing rate of leaked information in universal random privacy amplification,” *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3989–4001, 2011.

[48] 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.

[49] R. W. Yeung, *A First Course in Information Theory (Information Technology: Transmission, Processing and Storage)*. Berlin, Heidelberg: Springer-Verlag, 2006.

[50] R. Blahut, “Computation of channel capacity and rate-distortion functions,” *IEEE Transactions on Information Theory*, vol. 18, no. 4, pp. 460–473, 1972.

[51] 石川悠樹, 鎌塚明, and 風間皐希, “Arimoto-sibson capacity を計算する新たなアルゴリズム,” in **電子情報通信学会 技術研究報告**, ser. IT2023-22, vol. 123, no. 149, 神奈川, 8 月 2023, pp. 44–49, 2023 年 8 月 3 日 (木)-8 月 4 日 (金) 湘南工科大学 (IT).

[52] A. Kamatsuka, K. Kazama, and T. Yoshida, “Computing channel capacity: A statistical decision theoretic approach,” **第 *46* 回情報理論とその応用シンポジウム予稿集** *(SITA2023)*, pp. 123–128, Nov 2023.

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

付録

A　Proof of Theorem 4.2

Proof. 文献 [11, Chapter. 9.7] および [53] に倣って示す．いま $U^{\ell}_{\mathcal{L}}(R;\mathcal{Y})$ および $U^{\ell}_{\mathcal{L}}(R)$ を，それぞれ式 (44) および (47) の第二項目とする．すなわち，以下のように定義する：

$$U^{\ell}_{\mathcal{L}}(R;\mathcal{Y}) := \inf_{\substack{p_{Y|X}:\\ \mathcal{L}(X\to Y)\le R}} \mathbb{E}_Y\left[\inf_a \mathbb{E}_X\left[\ell(X,a)\mid Y\right]\right], \tag{58}$$

$$U^{\ell}_{\mathcal{L}}(R) := \inf_{\substack{p_{A|X}:\\ \mathcal{L}(X\to A)\le R}} \mathbb{E}_{X,A}\left[\ell(X,A)\right]. \tag{59}$$

(逆定理)：

$U^{\ell}_{\mathcal{L}}(R;\mathcal{Y}) \ge U^{\ell}_{\mathcal{L}}(R)$ を示せば十分である．

$p^*_{Y|X}$ およびベイズ決定関数 δ^{Bayes} を以下で定める：

$$p^*_{Y|X} := \operatorname*{arginf}_{\substack{p_{Y|X}:\\ \mathcal{L}(X\to Y)\le R}} U^{\ell}_{\mathcal{L}}(R;\mathcal{Y}), \tag{60}$$

$$\delta^{\mathrm{Bayes}}(y) := \operatorname*{argmin}_a \sum_{x\in\mathcal{X}} \ell(x,a)p^*_{X|Y}(x\mid y), \tag{61}$$

ここで，$p^*_{X|Y}(x\mid y) := \frac{p_X(x)p^*_{Y|X}(y|x)}{p_Y(y)}$．いま，$X-Y-A(:=\delta^{\mathrm{Bayes}}(Y))$ は任意の $p_{A|X}$ について Markov 連鎖をなすので，DPI の仮定 (23) から，

$$\mathcal{L}(X\to A)\le \mathcal{L}(X\to Y)\le R \tag{62}$$

ここで，$p^*_{A|X}$ を以下で定めると，

$$p^*_{A|X}(a\mid x) := \sum_{y\in\mathcal{Y}} p^*_{Y|X}(y\mid x)\mathbb{1}_{\{a=\delta^{\mathrm{Bayes}}(y)\}}. \tag{63}$$

不等式 (62) より，以下を得る：

$$U^{\ell}_{\mathcal{L}}(R) \le \sum_{x,a} p_X(x)p^*_{A|X}(a\mid x)\ell(x,a) \tag{64}$$

$$= \sum_{x,y} p_X(x)p^*_{Y|X}(y\mid x)\ell(x,\delta^{\mathrm{Bayes}}(y)) = U^{\ell}_{\mathcal{L}}(R;\mathcal{Y}). \tag{65}$$

(順定理)：$\mathcal{Y} := t(\mathcal{A})$ とおく．$U^{\ell}_{\mathcal{L}}(R;t(\mathcal{A})) \le U^{\ell}_{\mathcal{L}}(R)$ を示せば十分である．$p^*_{A|X}, p^*_A$ および $p^*_{X|A}$ を以下で定める：

$$p^*_{A|X} := \operatorname*{arginf}_{\substack{p_{A|X}:\\ \mathcal{L}(X\to A)\le R}} \mathbb{E}_{X,A}\left[\ell(X,A)\right], \tag{66}$$

$$p^*_A(a) := \sum_x p_X(x)p^*_{A|X}(a\mid x), \tag{67}$$

$$p^*_{X|A}(x\mid a) := \frac{p_X(x)p^*_{A|X}(a\mid x)}{p^*_A(a)}. \tag{68}$$

また，$\tilde{A}$ を p^*_A に従う確率変数とすると，$X-\tilde{A}-Y := t(\tilde{A})$ は Markov 連鎖をなすので，DPI の仮定 (23) から $\mathcal{L}(X\to Y)\le\mathcal{L}(X\to\tilde{A})\le R$ が成り立つ．よって，$p^*_{Y|X}(y\mid x) := \sum_a p^*_{A|X}(a\mid x)\mathbb{1}_{\{y=t(a)\}}$ とすると，以下が成り立つ：

$$U^{\ell}_{\mathcal{L}}(R;t(\mathcal{A})) := \inf_{\substack{p_{Y|X}:\\ \mathcal{L}(X\to Y)\le R}} \mathbb{E}_Y\left[\inf_a \mathbb{E}_X\left[\ell(X,a)\mid Y\right]\right] \tag{69}$$

$$\le \mathbb{E}_Y\left[\inf_a \mathbb{E}_X^{p^*_{X|Y}}\left[\ell(X,a)\mid Y\right]\right], \tag{70}$$

ここで，$\mathbb{E}_X^{p^*_{Y|X}}[\cdot]$ は $p^*_{X|Y}(x|y) = p_X(x)p^*_{Y|X}(y|x)/p^*_Y(y)$ により期待値を取ることを意味する．さらに，$\inf_a \mathbb{E}_X^{p^*_{X|Y}}\left[\ell(X,a)\mid Y=t(a')\right]$ を上から評価することを考えると，以下を得る：

$$\inf_a \mathbb{E}_X^{p^*_{X|Y}}\left[\ell(X,a)\,\middle|\, Y=t(a')\right]$$

$$\overset{(*)}{=} \inf_a \mathbb{E}_X^{p^*_{X|A}}\left[\ell(X,a)\,\middle|\, \tilde{A}=a'\right] \tag{71}$$

$$\le \mathbb{E}_X^{p^*_{X|A}}\left[\ell(X,a')\,\middle|\, \tilde{A}=a'\right], \tag{72}$$

ここで，$(*)$ は $t(\tilde{A})$ が十分統計量であることによる．したがって，

$$\mathbb{E}_Y\left[\inf_a \mathbb{E}_X^{p^*_{X|Y}}\left[\ell(X,a)\right]\mid Y\right] = \mathbb{E}_{\tilde{A}}\left[\inf_a \mathbb{E}_X^{p^*_{X|A}}\left[\ell(X,a)\right]\mid \tilde{A}\right] \tag{73}$$

$$\le \mathbb{E}_{X,\tilde{A}}^{p^*_{X|A}}\left[\ell(X,\tilde{A})\right] = \inf_{\substack{p_{A|X}:\\ \mathcal{L}(X\to A)\le R}} \mathbb{E}_{X,A}\left[\ell(X,A)\right] \tag{74}$$

$$= U^{\ell}_{\mathcal{L}}(R). \tag{75}$$

ゆえに，式 (70) とあわせると，$U^{\ell}_{\mathcal{L}}(R;t(\mathcal{A})) \le U^{\ell}_{\mathcal{L}}(R)$. □