An Estimation-Theoretic View of Privacy

Hao Wang, Flavio P. Calmon [*]

Abstract

We study the central problem in data privacy: how to share data with an analyst while providing both privacy and utility guarantees to the user that owns the data. We present an estimation-theoretic analysis of the privacy-utility trade-off (PUT) in this setting. Here, an analyst is allowed to reconstruct (in a mean-squared error sense) certain functions of the data (utility), while other private functions should not be reconstructed with distortion below a certain threshold (privacy). We demonstrate how a χ^2-based information measure captures the fundamental PUT, and characterize several properties of this function. In particular, we give a sharp bound for the PUT. We then propose a convex program to compute privacy-assuring mappings when the functions to be disclosed and hidden are known *a priori*. Finally, we evaluate the robustness of our approach to finite samples.

[*] H. Wang and F. P. Calmon are with the John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA (emails: hao_wang@g.harvard.edu, flavio@seas.harvard.edu).

Contents

1 Introduction

Data sharing and publishing is increasingly common within scientific communities [1], businesses [2], government operations [3], medical fields [4] and beyond. Data is usually shared with an application in mind, from which the data provider receives some utility. For example, when a user shares her movie ratings with a streaming service, she receives utility in the form of suggestions of new, interesting movies that fit her taste. As a second example, when a medical research group shares patient data, their aim is to enable a wider community of researchers and statisticians to learn interesting patterns from that data. Utility is then gained through new scientific discoveries.

The disclosure of non-encrypted data incurs a privacy risk through unwanted inferences. In our previous examples, the streaming service may infer the user's political preference (potentially deemed private by the user) from her movie ratings, or an insurance company may determine the identity of a patient in the medical dataset. If privacy is a concern but the data has no immediate utility, then cryptographic methods suffice.

The dichotomy between privacy and utility has been widely studied by computer scientists, statisticians and information theorists alike. While specific metrics and models may vary among these communities, their desideratum is the same: to design mechanisms that perturb the data (or functions thereof) while achieving an acceptable privacy-utility trade-off (PUT). The feasibility of this goal depends on the chosen privacy and utility metric, as well as the topology and distribution of the data. The information-theoretic approach to privacy, and notably the results of Sankar *et al.* [5] [6], Issa *et al.* [7] [8], Asoodeh *et al.* [9] [10], Calmon *et al.* [11] [12], among others, seeks to quantify the best possible PUT for *any* privacy mechanism. Here, information-theoretic quantities, such as mutual information or maximal leakage, are used to characterize privacy, and, under assumptions of the distribution of the data, bounds on the fundamental PUT are derived. It is within this information-theoretic approach that the present work is inscribed.

Our aim is to characterize the fundamental performance limits of privacy-assuring mechanisms from an estimation-theoretic perspective, and to develop data-driven privacy-assuring mechanisms that provide estimation-theoretic guarantees. The specific privacy and utility metric used is the χ^2-information between probability distributions, and more generally, the *principal inertia components* (PICs) of the distribution of the private and disclosed data. The PICs reveal interesting facets of data disclosure under privacy constraints, and quantify the minimum mean-squared error (MMSE) achievable for reconstructing both private and useful information from the disclosed data. We do not seek to claim that the estimation-based approach subsumes other privacy metrics (e.g. differential privacy [13]). Rather, our goal is to show that the MMSE-view reveals an interesting facet of data disclosure which, in turn, can drive the design of privacy mechanisms used in practice.

The rest of the paper is organized as follows. In Section 2, we review the definition of PICs and present some properties of the PICs. In Section 3, we introduce the χ^2-privacy-utility function and give several properties based on this function. In Section 4, we discuss a fine-grained case and propose a PIC-based convex program to find privacy-assuring mappings. Finally, robustness of our approach to limited samples is discussed in Section 5.

1.1 Related work

Several papers, such as Sankar *et al.* [5], Calmon *et al.* [12], Asoodeh *et al.* [14], and Makhdoumi *et al.* [15], have studied information disclosure with privacy guarantees through an information-theoretic lens. For example, Sankar *et al.* [5] characterized PUTs in large databases using tools from rate-distortion theory. Calmon *et al.* [12] presented lower bounds for the minimum mean-squared-error of estimating private functions of a plaintext given knowledge of other, disclosed functions. Makhdoumi *et al.* [15] introduced the privacy funnel, where both privacy and utility are measured in terms of mutual information, and showed its connection with the information bottleneck [16]. The PUT was also explored in [17] and [18] using mutual information as a privacy metric. Currently, the most adpopted definition of privacy is differential privacy ([13], [19]), which enables queries to be computed over a database while simultaneously ensuring privacy of individual entries of the database. Fundamental bounds on composition of differentially private mechanisms were given by Kairouz *et al.* [20]. Other quantities from the information-theoretic literature have been used to quantify privacy and utility. For example, Asoodeh *et al.* [9] and Calmon *et al.* [11] used estimation-theoretic tools to characterize fundamental limits of privacy. Also of note, Liao *et al.* ([21], [22]) explored the PUT within a hypothesis testing framework.

We introduce the χ^2-privacy-utility function in this paper. Related privacy-utility functions and proof techniques that inspired our approach have appeared in [9], [10], and [23].

1.2 Notation

For a positive integer n, we define $[n] \triangleq \{1, ..., n\}$. Matrices are denoted in bold capital letters (e.g. $\mathbf{P}$) and vectors in bold lower-case letters (e.g. $\mathbf{p}$). For a vector $\mathbf{p}$, $\mathsf{diag}(\mathbf{p})$ is defined as the matrix with diagonal entries equal to $\mathbf{p}$ and all other entries equal to 0. Capital letters (e.g. X and Y) are used to denote random variables, and calligraphic letters (e.g. $\mathcal{X}$ and $\mathcal{Y}$) denote sets. The span of a set $\mathcal{S}$ of vectors is

$$\mathsf{span}(\mathcal{S}) \triangleq \left\{ \sum_{i=1}^{k} \lambda_i \mathbf{v}_i \middle| k \in \mathbb{N}, \mathbf{v}_i \in \mathcal{S}, \lambda_i \in \mathbb{R} \right\}. \tag{1}$$

We denote independence of X and Y by $X \perp\!\!\!\perp Y$ and write $X \sim Y$ to indicate that X and Y have the same distribution. When X, Y, Z form a Markov chain, we write $X \rightarrow Y \rightarrow Z$. For a discrete random variable X with probability mass function P_X, we denote $P_{Xmin} \triangleq \inf\{P_X(x)|x \in \mathcal{X}\}$. The MMSE of estimating X given Y is

$$\mathsf{mmse}(X|Y) \triangleq \min_{X\rightarrow Y\rightarrow \hat{X}} \mathbb{E}\left[(X-\hat{X})^2\right] = \mathbb{E}\left[(X-E(X|Y))^2\right]. \tag{2}$$

The χ^2-information between two random variables, X and Y, is defined as

$$\chi^2(X;Y) \triangleq \mathbb{E}\left[\left(\frac{P_{X,Y}(X,Y)}{P_X(X)P_Y(Y)}\right)\right] - 1. \tag{3}$$

Let P_X and Q_X be two probability mass functions with the same discrete support set $\mathcal{X}$. We denote $||P_X - Q_X||_1 \triangleq \sum_{x\in\mathcal{X}}|P_X(x) - Q_X(x)|$. For any real-valued random variable X, we denote the $\mathcal{L}_p$-norm of X as

$$||X||_p \triangleq (\mathbb{E}\left[|X|^p\right])^{1/p}.$$

The set of all functions that when composed with a random variable X with distribution P_X result in an $\mathcal{L}_2$-norm smaller than 1 is given by

$$\mathcal{L}_2(P_X) \triangleq \{f : \mathcal{X} \rightarrow \mathbb{R} \mid \|f(X)\|_2 \leq 1\}. \tag{4}$$

1.3 Privacy Setup

Throughout this paper, S denotes a variable to be hidden (e.g. political preference), and X is an observed variable that depends on S (e.g. movie ratings). Our goal is to disclose a realization of a random variable Y, produced from X through a randomized mapping $P_{Y|X}$, called the privacy-assuring mapping. Here, S, X and Y satisfy the Markov condition $S \rightarrow X \rightarrow Y$. We assume that an analyst will provide some utility based on an observation of Y (e.g. movie recommendations), while potentially trying to estimate S from Y. Privacy and utility will be quantified in terms of how well the analyst can reconstruct/estimate functions of S and X given Y, respectively. The support sets of S, X and Y are $\mathcal{S} = \{1, ..., |\mathcal{S}|\}$, $\mathcal{X} = \{1, ..., |\mathcal{X}|\}$ and $\mathcal{Y} = \{1, ..., |\mathcal{Y}|\}$, respectively. Furthermore, we assume $|\mathcal{S}|$, $|\mathcal{X}|$ and $|\mathcal{Y}|$ are bounded.

2 Principal Inertia Components

We present next the properties of the PICs that will be used in this paper. For a more detailed overview, we refer the reader to [11] and the references therein. We use the definition of PICs presented in [11], but note that the PICs predate [11] (e.g. [24], [25] and the references therein).

Definition 1 ([11], Definition 1)**.** Let X and Y be r.v.s with support sets $\mathcal{X}$ and $\mathcal{Y}$, respectively, and distribution $P_{X,Y}$. In addition, let $f_0 : \mathcal{X} \rightarrow \mathbb{R}$ and $g_0 : \mathcal{Y} \rightarrow \mathbb{R}$ be the constant functions $f_0(x) = 1$ and $g_0(y) = 1$. For $k \in \mathbb{Z}_+$, we (recursively) define

$$\begin{aligned}\lambda_k(X;Y) = \max\{&\mathbb{E}\left[f(X)g(Y)\right]^2 | f \in \mathcal{L}_2(P_X), g \in \mathcal{L}_2(P_Y), \mathbb{E}\left[f(X)f_j(X)\right] = 0,\\ &\mathbb{E}\left[g(Y)g_j(Y)\right] = 0, j \in \{0, \ldots, k-1\}\},\end{aligned} \tag{5}$$

where

$$(f_k, g_k) \triangleq \text{argmax}\left\{\mathbb{E}\left[f(X)g(Y)\right]^2 \middle| f \in \mathcal{L}_2(P_X), g \in \mathcal{L}_2(P_Y), \mathbb{E}\left[f(X)f_j(X)\right] = 0, \mathbb{E}\left[g(Y)g_j(Y)\right] = 0, j \in \{0, \dots, k-1\}\right\}. \tag{6}$$

The values $\lambda_k(X;Y)$ are called the *principal inertia components* (PICs) of $P_{X,Y}$. The functions f_k and g_k are called the *principal functions* of X and Y.

Observe that the PICs satisfy $\lambda_k(X;Y) \leq 1$, since $f_k \in \mathcal{L}_2(P_X)$ $g_k \in \mathcal{L}_2(P_Y)$ and

$$\mathbb{E}\left[f(X)g(Y)\right] \leq \|f(X)\|_2 \|g(Y)\|_2 \leq 1.$$

Thus, from Definition 1, $\lambda_{k+1}(X;Y) \leq \lambda_k(X;Y) \leq 1$.

Definition 2 ([11], Definition 14)**.** Let $d \triangleq \min\{|\mathcal{X}|, |\mathcal{Y}|\} - 1$, and $\lambda_d(X;Y)$ the smallest PIC of $P_{X,Y}$. We define

$$\delta(P_{X,Y}) \triangleq \begin{cases} \lambda_d(X;Y) & \text{if } |\mathcal{Y}| \leq |\mathcal{X}|, \\ 0 & \text{otherwise.} \end{cases}$$

We also denote $\lambda_d(X;Y)$ as $\lambda_{min}(X;Y)$ throughout this paper.

When both random variables X and Y have a finite support set, we have the following definition.

Definition 3. For $\mathcal{X} = [m]$ and $\mathcal{Y} = [n]$, let $\mathbf{P}_{X,Y} \in \mathbb{R}^{m \times n}$ be a matrix with entries $[\mathbf{P}_{X,Y}]_{i,j} = P_{X,Y}(i,j)$, and $\mathbf{D}_X \in \mathbb{R}^{m \times m}$ and $\mathbf{D}_Y \in \mathbb{R}^{n \times n}$ be diagonal matrices with diagonal entries $[\mathbf{D}_X]_{i,i} = P_X(i)$ and $[\mathbf{D}_Y]_{j,j} = P_Y(j)$, respectively, where $i \in [m]$ and $j \in [n]$. We define

$$\mathbf{Q}_{X,Y} \triangleq \mathbf{D}_X^{-1/2} \mathbf{P}_{X,Y} \mathbf{D}_Y^{-1/2}. \tag{7}$$

We denote the singular value decomposition of $\mathbf{Q}_{X,Y}$ by $\mathbf{Q}_{X,Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$.

The next theorem illustrates the different characterizations of the PICs used in this paper.

Theorem 1 ([11], Theorem 1)**.** *The following characterizations of the PICs are equivalent:*

1. *The characterization given in Definition 1 where, for f_k and g_k given in* (6), $g_k(Y) = \frac{\mathbb{E}[f_k(X)|Y]}{\|\mathbb{E}[f_k(X)|Y]\|_2}$ *and* $f_k(X) = \frac{\mathbb{E}[g_k(Y)|X]}{\|\mathbb{E}[g_k(Y)|X]\|_2}$.
2. *For any $k \in \mathbb{Z}_+$,*

$$1 - \lambda_k(X;Y) = \min\left\{\mathsf{mmse}(f(X)|Y) \middle| f \in \mathcal{L}_2(P_X), \|f(X)\|_2 = 1, \mathbb{E}\left[f(X)h_j(X)\right] = 0, j \in \{0, \dots, k-1\}\right\}, \tag{8}$$

where

$$h_k \triangleq \text{argmin}\left\{\mathsf{mmse}(f(X)|Y) \middle| f \in \mathcal{L}_2(P_X), \|f(X)\|_2 = 1, \mathbb{E}\left[f(X)h_j(X)\right] = 0, j \in \{0, \dots, k-1\}\right\}. \tag{9}$$

If $\lambda_k(X;Y)$ is unique, then $h_k = f_k$ given in (6).

Finally, if both $\mathcal{X}$ and $\mathcal{Y}$ are defined over finite supports, the following characterization is also equivalent.

3. *$\sqrt{\lambda_k(X;Y)}$ is the $(k+1)$-st largest singular value of $\mathbf{Q}_{X,Y}$. The principal functions f_k and g_k in* (6) *correspond to the columns of the matrices $\mathbf{D}_X^{-1/2}\mathbf{U}$ and $\mathbf{D}_Y^{-1/2}\mathbf{V}$, respectively, where $\mathbf{Q}_{X,Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$.*

The equivalent characterizations of the PICs in the above theorem have the following intuitive interpretation: the principal functions can be viewed as a basis that decompose the mean-squared error of estimating functions of a hidden variable X given an observation Y.

It has been shown (e.g. [11]) that $\chi^2(S;Y) = \sum_{i=1}^{d} \lambda_i(S;Y)$ where $d = \min\{|\mathcal{S}|, |\mathcal{Y}|\} - 1$. If $\chi^2(S;Y)$ is small, say $\chi^2(S;Y) \leq \epsilon < 1$, then, equivalently, the sum of all PICs is upper-bounded by ϵ. Since the PICs are non-negative, each PIC is also upper-bounded by ϵ. Consequently, from characterization 2 in Theorem 1, it follows that the MMSE of reconstructing *any* zero-mean, unit variance function of S given Y is lower bounded by $1-\epsilon$, i.e. all functions of S cannot be reconstructed with small MMSE given an observation of Y. In other words, functions of S cannot be reliably estimated from the disclosed variable Y in a mean-squared error sense, and, depending on the value of ϵ and the needs of the user, privacy is assured in this estimation-theoretic sense. If $1 \leq \epsilon \ll d$, then some functions of S may be (perfectly) revealed, but most PICs of $P_{S,Y}$ are still small and, thus, most functions of S are hard to estimate from Y. Analogously, if $\chi^2(X;Y)$ has a large lower bound, then certain functions of X can be, on average, reconstructed (i.e. estimated) with small MMSE from an observation of Y.

The previous discussion demonstrates how χ^2-information can be used to measure both privacy and utility from an estimation-theoretic view, where estimation is measured in terms of MMSE. We will adopt χ^2-information as a measure of both privacy and utility in the next section, and explore the fundamental PUT in terms of this metric.

3 Privacy-Utility Trade-off

The results introduced in this section use the following definition.

Definition 4. We define the χ^2-privacy-utility function as:

$$h_{S,X}(\epsilon) \triangleq \sup_{P_{Y|X} \in \mathcal{D}_{S,X}(\epsilon)} \chi^2(X;Y),$$

where (S,X) has fixed joint distribution $P_{S,X}$, $0 \leq \epsilon \leq \chi^2(S;X)$ and

$$\mathcal{D}_{S,X}(\epsilon) \triangleq \{P_{Y|X} | S \to X \to Y, \chi^2(S;Y) \leq \epsilon\}.$$

Remark 1. Note that $P_{Y|X} \to \chi^2(S;Y)$ is a continuous mapping for fixed $P_{S,X}$ and finite support set $\mathcal{S}$, $\mathcal{X}$, $\mathcal{Y}$. Therefore, $\mathcal{D}_{S,X}(\epsilon)$ is a compact set and the supremum in $h_{S,X}(\epsilon)$ is indeed a maximum.

Lemma 1. *$h_{S,X}(\epsilon)$ is a concave function. Furthermore, $\epsilon \to \frac{h_{S,X}(\epsilon)}{\epsilon}$ is a non-increasing mapping.*

Proof. The proof is given in the appendix. □

The χ^2-privacy-utility function has an upper-bound

$$h_{S,X}(\epsilon) \leq \epsilon + |\mathcal{X}| - 1 - \chi^2(S;X), \tag{10}$$

that follows immedieatly by the data-processing inequality:

$$\chi^2(S;X) + \chi^2(X;Y) \leq \chi^2(S;Y) + \chi^2(X;X). \tag{11}$$

We provide next a sharp bound for the χ^2-privacy-utility function that significantly improves (10) by using properties of the PICs. We will use the following definition and two lemmas to derive this result.

Lemma 2. *Suppose $S \to X \to Y$. Then*

$$\chi^2(X;Y) = \mathsf{tr}(\mathbf{A}) - 1, \tag{12}$$

$$\chi^2(S;Y) = \mathsf{tr}(\mathbf{BA}) - 1, \tag{13}$$

where

$$\mathbf{A} = \mathbf{Q}_{X,Y}\mathbf{Q}_{X,Y}^T, \mathbf{B} = \mathbf{Q}_{S,X}^T\mathbf{Q}_{S,X}.$$

Proof. The proof is given in the appendix. □

Definition 5. For $t_i \in [0,1]$ $(i \in [n])$, $\sum_{i\in[n]} t_i \geq \epsilon$ and $m \geq n$, $G_\epsilon^m(t_1,...,t_n)$ is defined as follows:

$$G_\epsilon^m(t_1,...,t_n) \triangleq \max\left\{\sum_{i=1}^m x_i \middle| (x_1,...,x_m) \in \mathcal{D}_\epsilon^m(t_1,...,t_n)\right\}, \tag{14}$$

where

$$\mathcal{D}_\epsilon^m(t_1,...,t_n) \triangleq \left\{(x_1,...,x_m) \middle| \sum_{j=1}^n t_j x_j \leq \epsilon, x_i \in [0,1], i \in [m]\right\}. \tag{15}$$

$G_\epsilon^m(t_1,...,t_n)$ is the solution of a linear program which, in turn, can be written in the closed form given in the next lemma.

Lemma 3. *Suppose $1 \geq t_1 \geq ... \geq t_{n-s} > t_{n-s+1} = ... = t_n = 0$ without loss of generality. If*

$$\sum_{j=n-s-i+1}^{n-s} t_j \leq \epsilon \leq \sum_{j=n-s-i}^{n-s} t_j$$

where $i = 0,...,n-1-s$, then

$$G_\epsilon^m(t_1,...,t_n) = s + (m-n) + i + \frac{\epsilon - \sum_{j=n-s-i+1}^{n-s} t_j}{t_{n-s-i}}. \tag{16}$$

And

$$x_j = 1, \text{ for } j = n-s-i+1,...,m,$$

$$x_{n-s-i} = \frac{\epsilon - \sum_{j=n-s-i+1}^{n-s} t_j}{t_{n-s-i}},$$

$$x_l = 0, \text{ for } l = 1,...,n-s-i-1$$

can achieve the maximum in the definition of $G_\epsilon^m(t_1,...,t_n)$.

We next give an upper bound and a lower bound for the χ^2-privacy-utility function. These bounds are illustrated in Fig. 1.

Theorem 2. *For the χ^2-privacy-utility function $h_{S,X}(\epsilon)$ defined in Definition 4,*

$$\frac{|\mathcal{X}|-1}{\chi^2(S;X)}\epsilon \le h_{S,X}(\epsilon) \le G_\epsilon^{|\mathcal{X}|-1}(\lambda_1(S;X),...,\lambda_d(S;X)),$$

where $d \triangleq \min\{|\mathcal{S}|,|\mathcal{X}|\}-1$ and $\lambda_1(S;X),...,\lambda_d(S;X)$ are the PICs of $P_{S,X}$.

Proof. The lower bound for $h_{S,X}(\epsilon)$ follows immediately from the concavity of $h_{S,X}(\epsilon)$.

Using Lemma 2, the χ^2-privacy-utility function can be simplified as follows:

$$h_{S,X}(\epsilon) = \max_{P_{Y|X}\in\mathcal{D}_{S,X}(\epsilon)} \mathsf{tr}(\mathbf{A}) - 1,$$

$$\mathcal{D}_{S,X}(\epsilon) = \{P_{Y|X}|S \to X \to Y, \mathsf{tr}(\mathbf{BA}) - 1 \le \epsilon\},$$

where

$$\mathbf{A} = \mathbf{Q}_{X,Y}\mathbf{Q}_{X,Y}^T, \mathbf{B} = \mathbf{Q}_{S,X}^T\mathbf{Q}_{S,X}.$$

We denote the singular value decomposition of $\mathbf{Q}_{S,X}$ and $\mathbf{Q}_{X,Y}$ by $\mathbf{Q}_{S,X} = \mathbf{W}\mathbf{\Sigma}_1\mathbf{U}^T$ and $\mathbf{Q}_{X,Y} = \mathbf{V}\mathbf{\Sigma}_2\mathbf{M}^T$, respectively. Then $\mathbf{B} = \mathbf{U}\mathbf{\Sigma}_1^T\mathbf{\Sigma}_1\mathbf{U}^T = \mathbf{U}\mathbf{\Sigma}_B\mathbf{U}^T$, $\mathbf{A} = \mathbf{V}\mathbf{\Sigma}_2\mathbf{\Sigma}_2^T\mathbf{V}^T = \mathbf{V}\mathbf{\Sigma}_A\mathbf{V}^T$ where $\mathbf{\Sigma}_B \triangleq \mathbf{\Sigma}_1^T\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_A \triangleq \mathbf{\Sigma}_2\mathbf{\Sigma}_2^T$.

Let $\mathbf{A}_1 = \mathbf{U}^T\mathbf{A}\mathbf{U} = \mathbf{L}\mathbf{\Sigma}_A\mathbf{L}^T$ where $\mathbf{L} \triangleq \mathbf{U}^T\mathbf{V}$. Suppose the diagonal elements of $\mathbf{A}_1$ are $a_1,...,a_{|\mathcal{X}|}$. Then

$$\mathsf{tr}(\mathbf{BA}) - 1 = a_1 - 1 + \sum_{i=2}^{d+1}\lambda_{i-1}(S;X)a_i. \tag{17}$$

Suppose the i-th row of $\mathbf{L}$ is $\mathbf{l}_i = (l_{i,1},...,l_{i,|\mathcal{X}|})$, the i-th column of $\mathbf{U}$ is $\mathbf{u}_i^T$ and $\mathbf{\Sigma}_A = \mathsf{diag}(\sigma_1,...,\sigma_{|\mathcal{X}|})$. By the definition of PICs, $\sigma_1 = 1$, $\sigma_{j+1} = \lambda_j(X;Y)$ for $j = 1,...,d$ and $\sigma_{j+1} = 0$ for $j = d+1,...,|\mathcal{X}|-1$. Then for $i \in \{1,2,...,|\mathcal{X}|\}$

$$0 \le a_i = \sum_{j=1}^{|\mathcal{X}|}\sigma_j l_{i,j}^2 \le \sum_{j=1}^{|\mathcal{X}|} l_{ij}^2 = 1.$$

Since the first column of $\mathbf{U}$ and $\mathbf{V}$ are both $(\sqrt{P_X(1)},...,\sqrt{P_X(|\mathcal{X}|)})^T$ following the properties of PICs, then $\mathbf{l}_1 = \mathbf{u}_1\mathbf{V} = (1,0,...,0)$. Therefore, $a_1 = \sigma_1 = 1$. If $P_{Y|X} \in \mathcal{D}_{S,X}(\epsilon)$, then by Equation (17)

$$\mathsf{tr}(\mathbf{BA}) - 1 = \sum_{i=2}^{d+1}\lambda_{i-1}(S;X)a_i \le \epsilon,$$

which implies

$$(a_2,...,a_{|\mathcal{X}|}) \in \mathcal{D}_\epsilon^{|\mathcal{X}|-1}(\lambda_1(S;X),...,\lambda_d(S;X)).$$

Thus,

$$h_{S,X}(\epsilon) \le \max\left\{\sum_{i=2}^{|\mathcal{X}|}a_i \middle| (a_2,...,a_{|\mathcal{X}|}) \in \mathcal{D}_\epsilon^{|\mathcal{X}|-1}(\lambda_1(S;X),...,\lambda_d(S;X))\right\} = G_\epsilon^{|\mathcal{X}|-1}(\lambda_1(S;X),...,\lambda_d(S;X)).$$

□

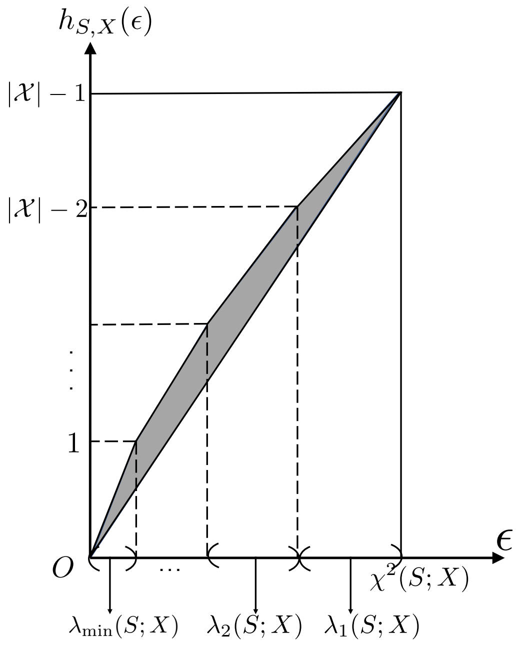

Figure 1: Upper bound and lower bound for χ^2-privacy-utility function when $\delta(P_{S,X}) > 0$.

Remark 2. The upper bound can also be proved by combining the trace inequality [26, Eq. (4)] with properties of the PICs.

If the value of $h_{S,X}(0)$ is known, a better lower bound can be obtained as follows:

$$\frac{|\mathcal{X}|-1-h_{S,X}(0)}{\chi^2(S;X)}\epsilon + h_{S,X}(0) \leq h_{S,X}(\epsilon). \tag{18}$$

When $S = X$, then $\chi^2(S;X) = |\mathcal{X}|-1$ and $h_{S,X}(\epsilon) = \epsilon$. Following Definition 5 and noticing that all PICs of $P_{S,X}$ are 1, then it can be shown that the upper bound and the lower bound for χ^2-privacy-utility function in Theorem 2 are both ϵ, which is equal to $h_{S,X}(\epsilon)$. Therefore, the upper bound and lower bound given in Theorem 2 is sharp.

Corollary 1. *$h_{S,X}(\epsilon)$ is a strictly increasing function for $\epsilon \in [0, \chi^2(S;X)]$.*

Proof. Firstly, $h_{S,X}(\epsilon)$ is non-decreasing. To see this, denote $\chi^2(S;X)$ by ϵ_0. Since $h_{S,X}(\epsilon)$ is a concave function, it is sufficient to show that, for every ϵ, $h_{S,X}(\epsilon) \leq h_{S,X}(\epsilon_0)$. For any Y, $\chi^2(X;Y) \leq |\mathcal{X}|-1$ which implies $h_{S,X}(\epsilon) \leq |\mathcal{X}|-1$. On the other hand, $h_{S,X}(\epsilon_0) = |\mathcal{X}|-1$. Thus, $h_{S,X}(\epsilon) \leq h_{S,X}(\epsilon_0)$.

Now suppose there exists $0 \leq \epsilon_1 < \epsilon_2 \leq \chi^2(S;X)$, such that $h_{S,X}(\epsilon_1) = h_{S,X}(\epsilon_2)$. Since $h_{S,X}(\epsilon)$ is a concave and non-decreasing function, then for any $\epsilon > \epsilon_1$, $h_{S,X}(\epsilon) = h_{S,X}(\epsilon_1)$. In particular, $h_{S,X}(\epsilon_1) = h_{S,X}(\epsilon_0) = |\mathcal{X}|-1$. This contradicts the upper bound of χ^2-privacy-utility function in Theorem 2 since the upper bound implies that $h_{S,X}(\epsilon) < |\mathcal{X}|-1$ when $\epsilon < \epsilon_0$. □

Definition 6. $\partial\mathcal{D}_{S,X}(\epsilon) \triangleq \{P_{Y|X} | S \rightarrow X \rightarrow Y, \chi^2(S;Y) = \epsilon\}$.

By Corollary 1, $h_{S,X}(\epsilon)$ is strictly increasing. Therefore,

$$h_{S,X}(\epsilon) = \max_{P_{Y|X} \in \mathcal{D}_{S,X}(\epsilon)} \chi^2(X;Y) = \max_{P_{Y|X} \in \partial\mathcal{D}_{S,X}(\epsilon)} \chi^2(X;Y). \tag{19}$$

By Corollary 7 in [11], when $\delta(P_{S,X}) = 0$, then $h_{S,X}(0) > 0$ (i.e. there exists a privacy-assuring mapping that allows the disclosure of a non-trivial amount of useful functions while guaranteeing perfect privacy). On the other hand, when $\delta(P_{S,X}) > 0$, then $h_{S,X}(0) = 0$. The following theorem shows that when $\delta(P_{S,X}) > 0$, the upper bound of $h_{S,X}(\epsilon)$ in Theorem 2 is achievable around zero implying that the upper bound is sharp around 0. This theorem also provides a specific way to construct the random variable Y which achieves the upper bound in the high privacy region.

Theorem 3. *Suppose $\delta(P_{S,X}) > 0$ and $P_{Xmin} > 0$. Then there exists a Y such that $S \to X \to Y$, $\chi^2(X;Y) = P_{Xmin}$ and $\chi^2(S;Y) = P_{Xmin}\lambda_{min}(S;X)$.*

Proof. From Theorem 1 in [11], there exists $f \in \mathcal{L}_2(P_X)$ such that $||f(X)||_2 = 1$, $\mathbb{E}[f(X)] = 0$ and $||\mathbb{E}[f(X)|S]||_2^2 = \lambda_{min}(S;X)$.
Fix $\mathcal{Y} = \{1, 2\}$ and the privacy-assuring mapping from X to Y is defined as follows:

$$P_{Y|X}(y|x) = \frac{1}{2} + (-1)^y \frac{\sqrt{P_{Xmin}} f(x)}{2}. \tag{20}$$

Since

$$1 = ||f(X)||_2^2 = \sum_{x=1}^{|\mathcal{X}|} f(x)^2 P_X(x) \geq f(x)^2 P_X(x),$$

for any x

$$|f(x)| \leq \frac{1}{\sqrt{P_X(x)}} \leq \frac{1}{\sqrt{P_{Xmin}}}.$$

Therefore, $|\frac{\sqrt{P_{Xmin}} f(x)}{2}| \leq \frac{1}{2}$, which implies that $P_{Y|X}(y|x)$ is feasible. Furthermore, $P_Y(y) = \frac{1}{2}$ because of $\mathbb{E}[f(X)] = 0$.

$$\begin{aligned}
\chi^2(X;Y) &= \sum_{x=1}^{|\mathcal{X}|} \sum_{y=1}^{|\mathcal{Y}|} \frac{P_{Y|X}(y|x)^2 P_X(x)}{P_Y(y)} - 1 \\
&= \sum_{x=1}^{|\mathcal{X}|} (P_X(x) + P_{Xmin} f(x)^2 P_X(x)) - 1 \\
&= P_{Xmin}.
\end{aligned}$$

Since

$$\begin{aligned}
P_{Y|S}(y|s) &= \sum_{x=1}^{|\mathcal{X}|} P_{Y|X}(y|x) P_{X|S}(x|s) \\
&= \sum_{x=1}^{|\mathcal{X}|} \left(\frac{1}{2} + (-1)^y \frac{\sqrt{P_{Xmin}} f(x)}{2} \right) P_{X|S}(x|s) \\
&= \frac{1}{2} + (-1)^y \frac{\sqrt{P_{Xmin}}}{2} \mathbb{E}[f(X)|S = s],
\end{aligned}$$

then

$$\begin{aligned}\chi^2(S;Y) &= \sum_{s=1}^{|\mathcal{S}|}\sum_{y=1}^{|\mathcal{Y}|}\frac{P_{Y|S}(y|s)^2P_S(s)}{P_Y(y)}-1\\ &= \sum_{s=1}^{|\mathcal{S}|}\Big(P_S(s)+P_{Xmin}\mathbb{E}\left[f(X)|S=s\right]^2P_S(s)\Big)-1\\ &= P_{Xmin}\lambda_{min}(S;X).\end{aligned}$$

□

Remark 3. When $\delta(P_{S,X})>0$ and $P_{Xmin}>0$, then $h_{S,X}(\hat{\epsilon})=P_{Xmin}$, where $\hat{\epsilon}=P_{Xmin}\lambda_{min}(S;X)$. Since $(\hat{\epsilon},P_{Xmin})$ is a point on the upper bound of χ^2-privacy-utility function given in Theorem 2, Theorem 3 shows that, in this case, the upper bound is achievable in the high privacy region.

4 A Convex Program for Computing Privacy-Assuring Mappings

In the previous section we studied χ^2-based metrics for both privacy and utility. This metric captures the overall error (in a mean-squared sense) of estimating functions of the private and the useful variables S and X, respectively. We used χ^2-information to measure both privacy and utility, and derived bounds for the PUT curve. The upper bound is shown to be achievable in the high privacy region in Theorem 3.

Next, we explore an alternative, finer-grained approach for measuring both privacy and utility based on the PICs (recall that χ^2-information is the sum of all PICs). This approach has a practical motivation, since often in there are well defined features (functions) of the data (realizations of a random variable) that should be hidden or disclosed. For example, a user might be comfortable revealing that his/her age is above a certain threshold, but not the age itself. Alternatively, a user may be willing to disclose that they prefer documentaries over action movies, but not exactly which documentary they like. More abstractly, we consider the case where certain known functions of a hidden variable should be revealed (utility), whereas others should be hidden (privacy). This is a finer-grained setting than the one used in the last section, since χ^2-information captures the aggregate reconstruction error across all zero-mean, unit variance functions.

We denote the set of functions to be hidden as

$$\mathcal{P}(S)\triangleq\{s_i(S)|s_i\in\mathcal{L}_2(P_S),\mathbb{E}\left[s_i(S)\right]=0,||s_i(S)||_2=1,i\in[m]\}.$$

We denote the set of functions to be disclosed as

$$\mathcal{U}(X)\triangleq\{u_i(X)|u_i\in\mathcal{L}_2(P_X),\mathbb{E}\left[u_i(X)\right]=0,||u_i(X)||_2=1,i\in[n]\}.$$

The goal is to find the privacy-assuring mapping, $P_{Y|X}$, such that $S\to X\to Y$ and Y satisfies the following privacy-utility constraints.

1. Utility constraints: $\max\{\mathsf{mmse}(u_i(X)|Y)\}_{i\in[n]}\le\Delta$ and $X\sim Y$.

2. Privacy constraints: $\mathsf{mmse}(s_i(S)|Y)\ge\theta_i, i\in[m]$.

In this section, we follow two steps, projection and optimization, to find this privacy-assuring mapping.

4.1 Projection

As a first step, we project all private functions to the observed variable and obtain a new set of functions:

$$\mathcal{P}(X) \triangleq \left\{ \hat{s}_i(X) \triangleq \frac{\mathbb{E}\left[s_i(S)|X\right]}{||\mathbb{E}\left[s_i(S)|X\right]||_2} \middle| i \in [m] \right\}.$$

The advantage of projection is twofold. First, it can significantly improve computational time of solving the optimization program, since after projection all functions are cast in terms of the observed random variable. Second, the hidden variable is not needed any more after the projection – the optimization solver only needs observed data. Therefore, the party that solves the optimization does not need access to the private data directly, further guaranteeing the safety of the sensitive information. Finally, the following theorem proves that privacy guarantees be cast in terms of the projected functions still hold for the original functions.

Theorem 4. *For $S \to X \to Y$, $f \in \mathcal{L}_2(P_S)$ and $\mathbb{E}\left[f(S)\right] = 0$, we have*

$$\mathbb{E}\left[\mathbb{E}\left[f(S)|X\right]\right] = 0, \tag{21}$$

and

$$\mathsf{mmse}\left(\frac{f(S)}{||f(S)||_2} \middle| Y \right) \geq \mathsf{mmse}\left(\frac{\mathbb{E}\left[f(S)|X\right]}{||\mathbb{E}\left[f(S)|X\right]||_2} \middle| Y \right). \tag{22}$$

Proof. Suppose $||f(S)||_2 = 1$ without loss of generality.
Observe that

$$\mathbb{E}\left[\mathbb{E}\left[f(S)|X\right]\right] = \mathbb{E}\left[f(S)\right] = 0.$$

Since $f(S) \to X \to Y$, then $\mathbb{E}\left[f(S)|X\right] = \mathbb{E}\left[f(S)|X, Y\right]$. Therefore,

$$\begin{aligned}
\mathsf{mmse}\left(\frac{\mathbb{E}\left[f(S)|X\right]}{||\mathbb{E}\left[f(S)|X\right]||_2} \middle| Y \right) &= \frac{\mathbb{E}\left[\mathbb{E}\left[f(S)|X\right]^2\right] - \mathbb{E}\left[\mathbb{E}\left[\mathbb{E}\left[f(S)|X\right]|Y\right]^2\right]}{||\mathbb{E}\left[f(S)|X\right]||_2^2} \\
&= \frac{\mathbb{E}\left[\mathbb{E}\left[f(S)|X\right]^2\right] - \mathbb{E}\left[\mathbb{E}\left[\mathbb{E}\left[f(S)|X, Y\right]|Y\right]^2\right]}{||\mathbb{E}\left[f(S)|X\right]||_2^2} \\
&= 1 - \frac{\mathbb{E}\left[\mathbb{E}\left[f(S)|Y\right]^2\right]}{||\mathbb{E}\left[f(S)|X\right]||_2^2} \\
&= \mathbb{E}\left[f(S)^2\right] - \frac{\mathbb{E}\left[\mathbb{E}\left[f(S)|Y\right]^2\right]}{||\mathbb{E}\left[f(S)|X\right]||_2^2} \\
&\leq \mathbb{E}\left[f(S)^2\right] - \mathbb{E}\left[\mathbb{E}\left[f(S)|Y\right]^2\right] \\
&= \mathsf{mmse}(f(S)|Y),
\end{aligned}$$

where the last inequality follows from Jensen's inequality:

$$1 = \mathbb{E}\left[f(S)^2\right] = \mathbb{E}\left[\mathbb{E}\left[f(S)^2|X\right]\right] \geq \mathbb{E}\left[\mathbb{E}\left[f(S)|X\right]^2\right] = ||\mathbb{E}\left[f(S)|X\right]||_2^2.$$

□

By Theorem 4, $\mathsf{mmse}(s_i(S)|Y) \geq \mathsf{mmse}(\hat{s}_i(X)|Y)$. Therefore, if the new set of functions satisfies the privacy constraints(i.e. $\mathsf{mmse}(\hat{s}_i(X)|Y) \geq \theta_i$), the original set of functions also satisfies the privacy constraints (i.e. $\mathsf{mmse}(s_i(S)|Y) \geq \theta_i$).

4.2 Optimization

After projection, both privacy and utility functions are based on the observed random variable. Next, we introduce a PIC-based convex optimization program to find the privacy-assuring mapping.

First, we construct $\mathbf{F}$ given by $(\mathbf{f}_0, \mathbf{f}_1, ..., \mathbf{f}_{|\mathcal{X}|-1})$ such that

$$\mathbf{F}^T \mathbf{D}_X \mathbf{F} = \mathbf{I}, \tag{23}$$

$$\mathsf{span}(\{\mathbf{f}_0, ..., \mathbf{f}_{n'}\}) = \mathsf{span}(\{\mathbf{f}_0, \mathbf{u}_1, ..., \mathbf{u}_n\}), \tag{24}$$

where

$$\mathbf{f}_i = (f_i(1), ..., f_i(|\mathcal{X}|))^T$$

and

$$\mathbf{u}_i = (u_i(1), ..., u_i(|\mathcal{X}|))^T.$$

From (23), $\{f_k(x)|k = 0, ..., |\mathcal{X}|-1\}$ is a basis of $\mathcal{L}_2(P_X)$, and the functions $\hat{s}_i(x)$ can be decomposed as

$$\hat{s}_i(x) = \sum_{k=0}^{|\mathcal{X}|-1} \alpha_{i,k} f_k(x), \tag{25}$$

If $\mathbf{P}_{X,Y} = \mathbf{D}_X \mathbf{F} \mathbf{\Sigma} \mathbf{F}^T \mathbf{D}_X$ is a feasible joint distribution matrix, then the following equations follow directly from Theorem 1:

$$\mathbb{E}\left[f_i(X) f_j(X)\right] = 0,$$

$$\mathbb{E}\left[g_i(Y) g_j(Y)\right] = \mathbb{E}\left[\frac{\mathbb{E}\left[f_i(X)|Y\right]}{\|\mathbb{E}\left[f_i(X)|Y\right]\|_2} \frac{\mathbb{E}\left[f_j(X)|Y\right]}{\|\mathbb{E}\left[f_j(X)|Y\right]\|_2}\right] = 0,$$

$$\mathsf{mmse}(f_i(X)|Y) = 1 - \lambda_i(X;Y),$$

where $i \neq j$. Moreover, it follows that

$$\mathsf{mmse}(\hat{s}_i(X)|Y) = \sum_{k=0}^{|\mathcal{X}|-1} \alpha_{i,k}^2 (1 - \lambda_k(X;Y)). \tag{26}$$

Therefore, the design of $P_{Y|X}$ is equivalent to solving the following convex program:

$$\begin{aligned}
&\max\ \sigma\\
\text{s.t. } &\sigma_0 = 1,\\
&\sigma_i \geq \sigma,\ (i = 1, ..., n')\\
&\sum_{k=0}^{|\mathcal{X}|-1} \alpha_{i,k}^2 \sigma_k^2 \leq 1 - \theta_i,\ (i = 1, ..., m)\\
&0 \leq \sigma_i \leq 1,\ (i = 1, ..., |\mathcal{X}|-1)\\
&\mathbf{\Sigma} = \mathsf{diag}(1, \sigma_1, ..., \sigma_{|\mathcal{X}|-1}),\\
&\mathbf{P}_{X,Y} = \mathbf{D}_X \mathbf{F} \mathbf{\Sigma} \mathbf{F}^T \mathbf{D}_X,\\
&\mathbf{P}_{X,Y} \text{ has non-negative entries.}
\end{aligned}$$

The previous convex program can be potentially solved by standard solvers. We also note that when all useful functions and private functions are based on the same random variable, we can use optimization without projection.

5 Robustness to Finite Sample Size

Fig. 2 illustrates one potential pipeline for designing privacy-assuring mappings in practice. We assume that samples of S and X are independently drawn from a source with distribution $Q_{S,X}$. During train time (step 1), a reference dataset with n samples is drawn from $Q_{S,X}$. The distribution of the source is estimated by computing the empirical distribution (type) $P_{\hat{S},\hat{X}}$ of the reference dataset. $P_{\hat{S},\hat{X}}$ and the privacy and utility constraints are then used as inputs to a convex program that returns the corresponding privacy-assuring mapping $P_{\hat{Y}|\hat{X}}$ (if feasible). We denote by $\hat{Y}$ the random variable produced by randomizing $\hat{X}$ according to $P_{\hat{Y}|\hat{X}}$, i.e. by applying the privacy-assuring mapping to a source with distribution $P_{\hat{S},\hat{X}}$. During test time, new i.i.d. samples from the source $Q_{S,X}$ are randomized using the privacy-assuring mapping $P_{\hat{Y}|\hat{X}}$ computed during train time, resulting in the disclosed variable Y.

The privacy and utility constraints used for computing the privacy-assuring mapping hold for a data source with distribution $P_{\hat{S},\hat{X}}$, since this is the distribution used as an input to the optimization program. However, during test time, $P_{\hat{Y}|\hat{X}}$ is applied to new samples from the source $Q_{S,X}$. *Do the privacy and utility guarantees still hold during test time?* Since as n increases $P_{\hat{S},\hat{X}}$ converges to $Q_{S,X}$, it is natural to expect that the privacy and utility guarantees during test time will not be far from the ones selected during train time.

In this section, we prove that for privacy and utility measured in terms of χ^2-information, with probability $1-\beta$ the privacy and utility guarantees at test time are at least $O\left(\sqrt{\frac{m}{n}\log\frac{n}{m\sqrt[m]{\beta}}}\right)$ close to the ones selected during train time where $m = |\mathcal{S}||\mathcal{X}||\mathcal{Y}|$. Similar analysis of robustness and generalization has appeared in [27] and [28].

We analyze next the robustness of the PUT optimization using χ^2-information, and characterize the gap between test and train time privacy and utility guarantees in terms of the number of samples in the reference dataset.

The main result makes use of the following technical lemmas.

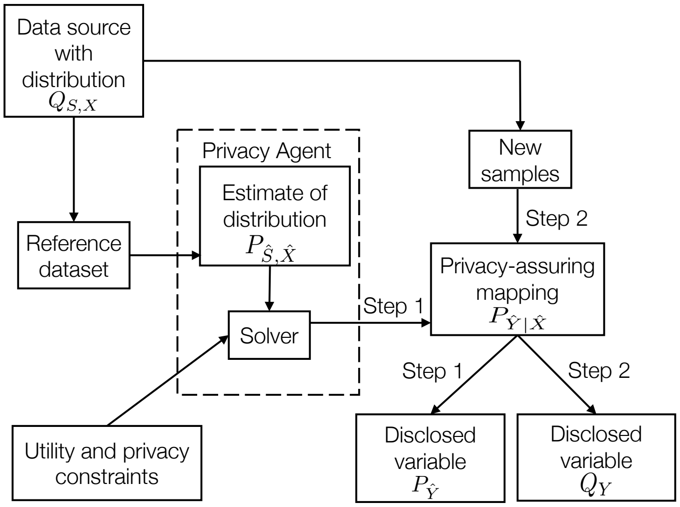
Data source with distribution $Q_{S,X}$
Privacy Agent
Estimate of distribution $P_{\hat{S},\hat{X}}$
Reference dataset
Solver
Utility and privacy constraints
New samples
Step 2
Privacy-assuring mapping $P_{\hat{Y}|\hat{X}}$
Step 1
Step 1
Step 2
Disclosed variable $P_{\hat{Y}}$
Disclosed variable Q_Y

Figure 2: Flowchart for train time (step 1) and test time (step 2).

Lemma 4. *Let $a, b, c, d \in [0, 1]$, then*

$$|ad - bc| \leq a|b - d| + b|a - c| \leq |b - d| + |a - c|.$$

Proof. The proof is given in the appendix. □

Lemma 5. *Let $P_{\hat{X}}(x)$ and $Q_X(x)$ be two probability mass functions with the same discrete and finite support set $\mathcal{X}$. Let $P_{\hat{Y}}(y)$ and $Q_Y(y)$ be two probability mass functions with the same discrete and finite support set $\mathcal{Y}$. Suppose $\min\{P_{\hat{Y}min}, Q_{Ymin}, P_{\hat{X}min}, Q_{Xmin}\} > 0$. Then*

$$|\chi^2(\hat{X};\hat{Y}) - \chi^2(X;Y)| \leq \frac{4||P_{\hat{X},\hat{Y}} - Q_{X,Y}||_1}{P_{\hat{X}min}P_{\hat{Y}min}Q_{Xmin}Q_{Ymin}} \leq \frac{4\sqrt{2D_{KL}(P_{\hat{X},\hat{Y}}||Q_{X,Y})}}{P_{\hat{X}min}P_{\hat{Y}min}Q_{Xmin}Q_{Ymin}}.$$

Proof.

$$|\chi^2(\hat{X};\hat{Y}) - \chi^2(X;Y)| \leq \sum_{x=1}^{|\mathcal{X}|}\sum_{y=1}^{|\mathcal{Y}|}\left|\frac{P_{\hat{X},\hat{Y}}(x,y)^2}{P_{\hat{X}}(x)P_{\hat{Y}}(y)} - \frac{Q_{X,Y}(x,y)^2}{Q_X(x)Q_Y(y)}\right|.$$

And

$$\left|\frac{P_{\hat{X},\hat{Y}}(x,y)^2}{P_{\hat{X}}(x)P_{\hat{Y}}(y)} - \frac{Q_{X,Y}(x,y)^2}{Q_X(x)Q_Y(y)}\right| \leq \frac{|P_{\hat{X},\hat{Y}}(x,y)^2 Q_X(x)Q_Y(y) - Q_{X,Y}(x,y)^2 P_{\hat{X}}(x)P_{\hat{Y}}(y)|}{P_{\hat{X}min}P_{\hat{Y}min}Q_{Xmin}Q_{Ymin}}.$$

Using Lemma 4, we can obtain:

$$\begin{aligned}
&|P_{\hat{X},\hat{Y}}(x,y)^2 Q_X(x)Q_Y(y) - Q_{X,Y}(x,y)^2 P_{\hat{X}}(x)P_{\hat{Y}}(y)| \\
\leq&|P_{\hat{X},\hat{Y}}(x,y)^2 - Q_{X,Y}(x,y)^2| + |P_{\hat{X}}(x)P_{\hat{Y}}(y) - Q_X(x)Q_Y(y)| \\
\leq&2|P_{\hat{X},\hat{Y}}(x,y) - Q_{X,Y}(x,y)| + P_{\hat{Y}}(y)|P_{\hat{X}}(x) - Q_X(x)| + Q_X(x)|P_{\hat{Y}}(y) - Q_Y(y)|.
\end{aligned}$$

Then

$$\sum_{x=1}^{|\mathcal{X}|}\sum_{y=1}^{|\mathcal{Y}|} P_{\hat{Y}}(y)|P_{\hat{X}}(x) - Q_X(x)| = ||P_{\hat{X}} - Q_X||_1 \leq ||P_{\hat{X},\hat{Y}} - Q_{X,Y}||_1.$$

The following inequality holds by the same reason.

$$\sum_{x=1}^{|\mathcal{X}|}\sum_{y=1}^{|\mathcal{Y}|} Q_X(x)|P_{\hat{Y}}(y) - Q_Y(y)| \leq ||P_{\hat{X},\hat{Y}} - Q_{X,Y}||_1.$$

Therefore,

$$|\chi^2(\hat{X};\hat{Y}) - \chi^2(X;Y)| \leq \frac{4||P_{\hat{X},\hat{Y}} - Q_{X,Y}||_1}{P_{\hat{X}min}P_{\hat{Y}min}Q_{Xmin}Q_{Ymin}} \leq \frac{4\sqrt{2D_{KL}(P_{\hat{X},\hat{Y}}||Q_{X,Y})}}{P_{\hat{X}min}P_{\hat{Y}min}Q_{Xmin}Q_{Ymin}},$$

where the second inequality follows from Pinsker's inequality. □

Remark 4. When $P_{\hat{Y}|\hat{X}} = Q_{Y|X}$, we can improve the result as follows:

$$|\chi^2(\hat{X};\hat{Y}) - \chi^2(X;Y)| \leq \frac{2||P_{\hat{X},\hat{Y}} - Q_{X,Y}||_1}{P_{\hat{Y}min}Q_{Ymin}} \leq \frac{2\sqrt{2D_{KL}(P_{\hat{X},\hat{Y}}||Q_{X,Y})}}{P_{\hat{Y}min}Q_{Ymin}}. \tag{27}$$

The following theorem is our main result in this section. It answers the question raised at the beginning of this section and provides the rate of the convergence in terms of n.

Theorem 5. *Let $P_{\hat{S},\hat{X}}$ be the empirical distribution obtained from n i.i.d. samples that is used to determine the mapping $P_{\hat{Y}|\hat{X}}$, and $Q_{S,X}$ be the true distribution of the data. In addition, denote by Q_Y the distribution after applying $P_{\hat{Y}|\hat{X}}$ to samples from $Q_{S,X}$. Suppose $\hat{S}$ and S, $\hat{X}$ and X, $\hat{Y}$ and Y have the same discrete and finite support set as $\mathcal{S}$, $\mathcal{X}$ and $\mathcal{Y}$, respectively. We define $m = |\mathcal{S}||\mathcal{X}||\mathcal{Y}|$. Assume* $\min\{P_{\hat{S}min}, Q_{Smin}, P_{\hat{Y}min}, Q_{Ymin}\} > 0$ *and $n \gg m$.*
(a) If $\chi^2(\hat{X};\hat{Y}) \geq \epsilon_1$, then with probability $1-\beta$,

$$\chi^2(X;Y) \geq \epsilon_1 - O\left(\sqrt{\frac{m}{n}\log\frac{n}{m\sqrt[m]{\beta}}}\right). \tag{28}$$

(b) If $\chi^2(\hat{S};\hat{Y}) \leq \epsilon_2$, then with probability $1-\beta$,

$$\chi^2(S;Y) \leq \epsilon_2 + O\left(\sqrt{\frac{m}{n}\log\frac{n}{m\sqrt[m]{\beta}}}\right). \tag{29}$$

Proof. The distribution $P_{\hat{S},\hat{X}}$ is the type (Chap.11 in [29]) of n observations of $Q_{S,X}$. Then from Corollary 2.1 in [30], for $\tau > 0$

$$P(D_{KL}(P_{\hat{S},\hat{X}}||Q_{S,X}) \geq \tau) \leq \binom{n+m-1}{m-1}e^{-n\tau} \leq \left(\frac{e(n+m)}{m}\right)^m e^{-n\tau}.$$

Also, since $P_{\hat{Y}|\hat{X}} = Q_{Y|X}$, then $D_{KL}(P_{\hat{S},\hat{X},\hat{Y}}||Q_{S,X,Y}) = D_{KL}(P_{\hat{S},\hat{X}}||Q_{S,X})$. By the chain rule,

$$D_{KL}(P_{\hat{S},\hat{Y}}||Q_{S,Y}) \leq D_{KL}(P_{\hat{S},\hat{X},\hat{Y}}||Q_{S,X,Y}) = D_{KL}(P_{\hat{S},\hat{X}}||Q_{S,X}),$$

$$D_{KL}(P_{\hat{X},\hat{Y}}||Q_{X,Y}) \leq D_{KL}(P_{\hat{S},\hat{X}}||Q_{S,X}).$$

Choosing

$$\tau = \frac{1}{n}\log\left(\frac{1}{\beta}\left(\frac{e(n+m)}{m}\right)^m\right).$$

Therefore, with probability $1-\beta$,

$$\begin{aligned} D_{KL}(P_{\hat{X},\hat{Y}}||Q_{X,Y}) \leq \tau, \\ D_{KL}(P_{\hat{S},\hat{Y}}||Q_{S,Y}) \leq \tau. \end{aligned}$$

(a)By Remark 4, with probability $1-\beta$,

$$\begin{aligned} |\chi^2(\hat{X};\hat{Y}) - \chi^2(X;Y)| \leq & \frac{2\sqrt{2D_{KL}(P_{\hat{X},\hat{Y}}||Q_{X,Y})}}{P_{\hat{Y}min}Q_{Ymin}} \\ \leq & \frac{2\sqrt{2\tau}}{P_{\hat{Y}min}Q_{Ymin}} \\ = & \frac{2\sqrt{2}}{P_{\hat{Y}min}Q_{Ymin}}\sqrt{\frac{1}{n}\log\left(\frac{1}{\beta}\left(\frac{e(n+m)}{m}\right)^m\right)} \\ = & O\left(\sqrt{\frac{m}{n}\log\frac{n}{m\sqrt[m]{\beta}}}\right). \end{aligned}$$

Thus, with probability $1-\beta$

$$\begin{aligned} \chi^2(X;Y) \geq & \chi^2(\hat{X};\hat{Y}) - |\chi^2(\hat{X};\hat{Y}) - \chi^2(X;Y)| \\ \geq & \epsilon_1 - O\left(\sqrt{\frac{m}{n}\log\frac{n}{m\sqrt[m]{\beta}}}\right). \end{aligned}$$

(b)By the same reason,

$$|\chi^2(\hat{S};\hat{Y}) - \chi^2(S;Y)| \leq O\left(\sqrt{\frac{m}{n}\log\frac{n}{m\sqrt[m]{\beta}}}\right).$$

Thus, with probability $1-\beta$

$$\begin{aligned} \chi^2(S;Y) \leq & \chi^2(\hat{S};\hat{Y}) + |\chi^2(\hat{S};\hat{Y}) - \chi^2(S;Y)| \\ \leq & \epsilon_2 + O\left(\sqrt{\frac{m}{n}\log\frac{n}{m\sqrt[m]{\beta}}}\right). \end{aligned}$$

□

6 Concluding remarks

In this paper, we studied the fundamental PUT in data disclosure, where χ^2-information was used to measure both privacy and utility. Several properties of the PUT were characterized through the χ^2-privacy-utility function. Moreover, a finer-grained, PIC-based convex optimization framework was proposed to design privacy-assuring mechanisms. Finally, we analyzed the robustness of our method to finite samples. Our analysis was based on the difference between the realizations of an observed random variable and its expected type. Analyzing the effectiveness of other distribution estimation methods is the subject of future work.

References

[1] C. Tenopir, S. Allard, K. Douglass, A. U. Aydinoglu, L. Wu, E. Read, M. Manoff, and M. Frame, "Data sharing by scientists: practices and perceptions," *PloS one*, vol. 6, no. 6, p. e21101, 2011.

[2] G. Stefansson, "Business-to-business data sharing: A source for integration of supply chains," *International journal of production economics*, vol. 75, no. 1, pp. 135–146, 2002.

[3] B. Otjacques, P. Hitzelberger, and F. Feltz, "Interoperability of e-government information systems: Issues of identification and data sharing," *Journal of Management Information Systems*, vol. 23, no. 4, pp. 29–51, 2007.

[4] L. Sweeney, "Guaranteeing anonymity when sharing medical data, the datafly system." in *Proceedings of the AMIA Annual Fall Symposium.* American Medical Informatics Association, 1997, p. 51.

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

[6] ——, "A theory of utility and privacy of data sources," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on.* IEEE, 2010, pp. 2642–2646.

[7] I. Issa, S. Kamath, and A. B. Wagner, "An operational measure of information leakage," in *Information Science and Systems (CISS), 2016 Annual Conference on.* IEEE, 2016, pp. 234–239.

[8] ——, "Maximal leakage minimization for the shannon cipher system," in *Information Theory (ISIT), 2016 IEEE International Symposium on.* IEEE, 2016, pp. 520–524.

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

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

[11] F. P. Calmon, A. Makhdoumi, M. Médard, M. Varia, M. Christiansen, and K. R. Duffy, "Principal inertia components and applications," *IEEE Transactions on Information Theory*, 2017.

[12] F. P. Calmon, M. Varia, and M. Médard, "On information-theoretic metrics for symmetric-key encryption and privacy," in *Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on.* IEEE, 2014, pp. 889–894.

[13] C. Dwork, "Differential privacy," in *Encyclopedia of Cryptography and Security.* Springer, 2011, pp. 338–340.

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

[15] A. Makhdoumi, S. Salamatian, N. Fawaz, and M. Médard, "From the information bottleneck to the privacy funnel," in *Information Theory Workshop (ITW), 2014 IEEE*. IEEE, 2014, pp. 501–505.

[16] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," *arXiv preprint physics/0004057*, 2000.

[17] A. Makhdoumi and N. Fawaz, "Privacy-utility tradeoff under statistical uncertainty," in *Communication, Control, and Computing (Allerton), 2013 51st Annual Allerton Conference on*. IEEE, 2013, pp. 1627–1634.

[18] D. Rebollo-Monedero, J. Forne, and J. Domingo-Ferrer, "From t-closeness-like privacy to postrandomization via information theory," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 11, pp. 1623–1636, 2010.

[19] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography Conference*. Springer, 2006, pp. 265–284.

[20] P. Kairouz, S. Oh, and P. Viswanath, "The composition theorem for differential privacy," *IEEE Transactions on Information Theory*, vol. 63, no. 6, pp. 4037–4049, 2017.

[21] J. Liao, L. Sankar, V. Y. Tan, and F. P. Calmon, "Hypothesis testing in the high privacy limit," *arXiv preprint arXiv:1607.00533*, 2016.

[22] J. Liao, L. Sankar, F. P. Calmon, and V. Y. Tan, "Hypothesis testing under maximal leakage privacy constraints," *arXiv preprint arXiv:1701.07099*, 2017.

[23] Y. Polyanskiy and Y. Wu, "Dissipation of information in channels with input constraints," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 35–55, 2016.

[24] M. Greenacre and T. Hastie, "The geometric interpretation of correspondence analysis," *Journal of the American statistical association*, vol. 82, no. 398, pp. 437–447, 1987.

[25] A. Rényi, "On measures of dependence," *Acta mathematica hungarica*, vol. 10, no. 3-4, pp. 441–451, 1959.

[26] J. B. Lasserre, "A trace inequality for matrix product," *IEEE Transactions on Automatic Control*, vol. 40, no. 8, pp. 1500–1501, 1995.

[27] O. Shamir, S. Sabato, and N. Tishby, "Learning and generalization with the information bottleneck," in *International Conference on Algorithmic Learning Theory*. Springer, 2008, pp. 92–107.

[28] F. P. Calmon, D. Wei, K. N. Ramamurthy, and K. R. Varshney, "Optimized data pre-processing for discrimination prevention," *arXiv preprint arXiv:1704.03354*, 2017.

[29] T. M. Cover and J. A. Thomas, "Elements of information theory 2nd edition (wiley series in telecommunications and signal processing)," 2006.

[30] I. Csiszár, P. C. Shields *et al.*, "Information theory and statistics: A tutorial," *Foundations and Trends® in Communications and Information Theory*, vol. 1, no. 4, pp. 417–528, 2004.

Appendix A Proof of Lemma 1

For $0 \le \epsilon_1 < \epsilon_2 < \epsilon_3 \le \chi^2(S;X)$, it suffices to show that

$$\frac{h_{S,X}(\epsilon_3) - h_{S,X}(\epsilon_1)}{\epsilon_3 - \epsilon_1} \le \frac{h_{S,X}(\epsilon_2) - h_{S,X}(\epsilon_1)}{\epsilon_2 - \epsilon_1},$$

which is equivalent to

$$\frac{\epsilon_2 - \epsilon_1}{\epsilon_3 - \epsilon_1} h_{S,X}(\epsilon_3) + \frac{\epsilon_3 - \epsilon_2}{\epsilon_3 - \epsilon_1} h_{S,X}(\epsilon_1) \le h_{S,X}(\epsilon_2). \tag{30}$$

Let $P_{Y_1|X}$ and $P_{Y_3|X}$ be two optimal solution in $\mathcal{D}_{S,X}(\epsilon_1)$ and $\mathcal{D}_{S,X}(\epsilon_3)$, respectively. Assume that Y_1 and Y_3 take values in $[m_1]$ and $[m_3]$, respectively. Furthermore, we denote $\lambda \triangleq \frac{\epsilon_2-\epsilon_1}{\epsilon_3-\epsilon_1}$. Next, we introduce a new privacy-assuring mapping defined as

$$P_{Y_\lambda|X}(y|x) \triangleq \begin{cases} \lambda P_{Y_3|X}(y|x) & \text{if } y \in [m_3], \\ (1-\lambda) P_{Y_1|X}(y - m_3|x) & \text{if } y - m_3 \in [m_1]. \end{cases} \tag{31}$$

Consequently, we have

$$P_{Y_\lambda}(y) = \begin{cases} \lambda P_{Y_3}(y) & \text{if } y \in [m_3], \\ (1-\lambda) P_{Y_1}(y - m_3) & \text{if } y - m_3 \in [m_1]. \end{cases} \tag{32}$$

Then

$$\begin{aligned}
\chi^2(X;Y_\lambda) =& \mathbb{E}\left[\frac{P_{X,Y_\lambda}(X,Y_\lambda)}{P_X(X)P_{Y_\lambda}(Y_\lambda)}\right] - 1 \\
=& \sum_{y\in[m_3]} \sum_{x=1}^{|\mathcal{X}|} \frac{P_{X,Y_\lambda}(x,y)^2}{P_X(x)P_{Y_\lambda}(y)} + \sum_{y-m_3\in[m_1]} \sum_{x=1}^{|\mathcal{X}|} \frac{P_{X,Y_\lambda}(x,y)^2}{P_X(x)P_{Y_\lambda}(y)} - 1 \\
=& \sum_{y\in[m_3]} \sum_{x=1}^{|\mathcal{X}|} \frac{P_{Y_\lambda|X}(y|x)^2 P_X(x)}{P_{Y_\lambda}(y)} + \sum_{y-m_3\in[m_1]} \sum_{x=1}^{|\mathcal{X}|} \frac{P_{Y_\lambda|X}(y|x)^2 P_X(x)}{P_{Y_\lambda}(y)} - 1 \\
=& \sum_{y\in[m_3]} \sum_{x=1}^{|\mathcal{X}|} \frac{\lambda^2 P_{Y_3|X}(y|x)^2 P_X(x)}{\lambda P_{Y_3}(y)} + \sum_{y\in[m_1]} \sum_{x=1}^{|\mathcal{X}|} \frac{(1-\lambda)^2 P_{Y_1|X}(y|x)^2 P_X(x)}{(1-\lambda) P_{Y_1}(y)} - 1 \\
=& \lambda \chi^2(X;Y_3) + (1-\lambda)\chi^2(X;Y_1).
\end{aligned}$$

Similarly, we have

$$\chi^2(S;Y_\lambda) = \lambda\chi^2(S;Y_3) + (1-\lambda)\chi^2(S;Y_1) \le \epsilon_2, \tag{33}$$

which implies that $P_{Y_\lambda|X} \in \mathcal{D}_{S,X}(\epsilon_2)$. Therefore,

$$\begin{aligned}
h_{S,X}(\epsilon_2) \ge& \chi^2(X;Y_\lambda) \\
&= \lambda\chi^2(X;Y_3) + (1-\lambda)\chi^2(X;Y_1) \\
&= \frac{\epsilon_2 - \epsilon_1}{\epsilon_3 - \epsilon_1} h_{S,X}(\epsilon_3) + \frac{\epsilon_3 - \epsilon_2}{\epsilon_3 - \epsilon_1} h_{S,X}(\epsilon_1),
\end{aligned} \tag{34}$$

which implies that (30) is true, so $h_{S,X}(\epsilon)$ is a concave function.

Then

$$\frac{h_{S,X}(\epsilon)}{\epsilon} = \frac{h_{S,X}(\epsilon) - h_{S,X}(0)}{\epsilon} + \frac{h_{S,X}(0)}{\epsilon}. \tag{35}$$

$\frac{h_{S,X}(0)}{\epsilon}$ is non-increasing because $h_{S,X}(0)$ is non-negative. Since $h_{S,X}(\epsilon)$ is concave, $\frac{h_{S,X}(\epsilon)-h_{S,X}(0)}{\epsilon}$ is also non-increasing. Therefore, $\frac{h_{S,X}(\epsilon)}{\epsilon}$ is non-increasing.

Appendix B Proof of Lemma 2

By the definition of χ^2-information,

$$\chi^2(X;Y)+1=\sum_{x=1}^{|\mathcal{X}|}\sum_{y=1}^{|\mathcal{Y}|}\frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}P_{X,Y}(x,y).$$

Note that $\mathbf{Q}_{X,Y}=\mathbf{D}_X^{-\frac{1}{2}}\mathbf{P}_{X,Y}\mathbf{D}_Y^{-\frac{1}{2}}$,

$$\begin{aligned}\mathsf{tr}(\mathbf{Q}_{X,Y}\mathbf{Q}_{X,Y}^T)&=\mathsf{tr}(\mathbf{D}_X^{-\frac{1}{2}}\mathbf{P}_{X,Y}\mathbf{D}_Y^{-1}\mathbf{P}_{X,Y}^T\mathbf{D}_X^{-\frac{1}{2}})\\&=\mathsf{tr}(\mathbf{D}_X^{-1}\mathbf{P}_{X,Y}\mathbf{D}_Y^{-1}\mathbf{P}_{X,Y}^T)\\&=\sum_{x=1}^{|\mathcal{X}|}\sum_{y=1}^{|\mathcal{Y}|}\frac{P_{X,Y}(x,y)}{P_X(x)}\frac{P_{X,Y}(x,y)}{P_Y(y)}.\end{aligned}$$

Therefore,

$$\chi^2(X;Y)=\mathsf{tr}(\mathbf{Q}_{X,Y}\mathbf{Q}_{X,Y}^T)-1=\mathsf{tr}(\mathbf{A})-1.$$

Since

$$\begin{aligned}\mathbf{Q}_{S,Y}&=\mathbf{D}_S^{-\frac{1}{2}}\mathbf{P}_{S,Y}\mathbf{D}_Y^{-\frac{1}{2}}\\&=\mathbf{D}_S^{-\frac{1}{2}}\mathbf{P}_{S,X}\mathbf{D}_X^{-\frac{1}{2}}\mathbf{D}_X^{-\frac{1}{2}}\mathbf{P}_{X,Y}\mathbf{D}_Y^{-\frac{1}{2}}\\&=\mathbf{Q}_{S,X}\mathbf{Q}_{X,Y},\end{aligned}$$

then

$$\begin{aligned}\chi^2(S;Y)&=\mathsf{tr}(\mathbf{Q}_{S,Y}\mathbf{Q}_{S,Y}^T)-1\\&=\mathsf{tr}(\mathbf{Q}_{S,X}\mathbf{Q}_{X,Y}\mathbf{Q}_{X,Y}^T\mathbf{Q}_{S,X}^T)-1\\&=\mathsf{tr}(\mathbf{BA})-1.\end{aligned}$$

Appendix C Proof of Lemma 4

Since $a,b,c,d\in[0,1]$, then

$$\begin{aligned}|ad-bc|&=|ad-ab+ab-bc|\\&\le a|b-d|+b|a-c|\\&\le|b-d|+|a-c|.\end{aligned}$$

Acknowledgement

The authors are indebted to Muriel Médard, Shahab Asoodeh, Salman Salamatian and Lalitha Sankar for discussions and feedback on the results presented here.