On Linear Refinement of Differential Privacy-Preserving Query Answering

Xiaowei Ying, Xintao Wu, and Yue Wang

University of North Carolina at Charlotte
{xying,xwu,ywang91}@uncc.edu

Abstract. Recent work showed the necessity of incorporating a user's background knowledge to improve the accuracy of estimates from noisy responses of histogram queries. Various types of constraints (e.g., linear constraints, ordering constraints, and range constraints) may hold on the true (non-randomized) answers of histogram queries. So the idea was to apply the constraints over the noisy responses and find a new set of answers (called refinements) that are closest to the noisy responses and also satisfy known constraints. As a result, the refinements expect to boost the accuracy of final histogram query results. However, there is one key question: is the ratio of the distributions of the results after refinements from any two neighbor databases still bounded? In this paper, we introduce a new definition, ρ-differential privacy on refinement, to quantify the change of distributions of refinements. We focus on one representative refinement, the linear refinement with linear constraints and study the relationship between the classic ϵ-differential privacy (on responses) and our ρ-differential privacy on refinement. We demonstrate the conditions when the ρ-differential privacy on refinement achieves the same ϵ-differential privacy. We argue privacy breaches could incur when the conditions do not meet.

Keywords: differential privacy, linear constraint, refinement, background knowledge.

1 Introduction

Research on differential privacy [1, 2] has shown that it is possible to carry out data analysis on sensitive data while ensuring strong privacy guarantees. Differential privacy is a paradigm of post-processing the output of queries. Differential privacy is defined as a property of a query answering mechanism, and a query answering mechanism satisfying differential privacy must meet the requirement that the distribution of its noisy query responses change very little with the addition or deletion of any record, so that the analyst can not infer the presence or absence of some record from the responses. Formally, differential privacy uses a user-specified privacy threshold ϵ to bound the ratio of the probabilities of the noisy responses from any two neighbor databases (differing one record).

Recent work [3–5] showed the necessity of incorporating a user's background knowledge to improve the accuracy of estimates from noisy responses of histogram queries. Various types of constraints (e.g., linear constraints, ordering

J. Pei et al. (Eds.): PAKDD 2013, Part II, LNAI 7819, pp. 353–364, 2013.

constraints, and range constraints) may hold on the true (non-randomized) answers of histogram queries. So the idea was to apply the constraints over the noisy responses and find a new set of answers (called refinements) that are closest to the noisy responses and also satisfy known constraints. As a result, the refinements expect to boost the accuracy of final histogram query results.

However, there is one key question: is the ratio of the distributions of the results after refinements from any two neighbor databases still bounded? In this paper, we introduce a new definition, ρ-differential privacy on refinement, to quantify the change of distributions of refinements. We focus on one representative refinement, the linear refinement with linear constraints and study the relationship between the classic ϵ-differential privacy (on responses) and our ρ-differential privacy on refinement. We demonstrate the conditions when the ρ-differential privacy on refinement achieves the same ϵ-differential privacy. We argue privacy breaches could incur when the conditions do not meet.

2 Differential Privacy Revisited

We revisit the formal definition and the mechanism of differential privacy. We denote the original database as $\mathcal{D}$, and its neighboring database as $\mathcal{D}'$. We will concentrate on pairs of databases (D, D') differing only in one row, meaning one is a subset of the other and the larger database contains just one additional row.

Definition 1. *(ϵ-differential privacy) [1]. A mechanism $\mathcal{K}$ is ϵ-differentially private if for all databases D and D' differing on at most one element, and any subsets of outputs $S \subseteq Range(\mathcal{K})$,*

$$\Pr[\mathcal{K}(D) \in S] \leq e^{\epsilon} \times \Pr[\mathcal{K}(D') \in S] \tag{1}$$

Theorem 1. *[1] For $f : D \to \mathbf{R}^d$, the mechanism $\mathcal{K}_f$ that adds independently generated noise with distribution $Lap(\Delta f/\epsilon)$ to each of the d output terms satisfies ϵ-differential privacy, where the sensitivity, Δf, is $\Delta f = max_{D,D'} \|f(D) - f(D')\|_1$ for all D, D' differing in at most one element.*

The mechanism for achieving differential privacy computes the sum of the true answer and random noise generated from a Laplace distribution. The magnitude of the noise distribution is determined by the sensitivity of the computation and the privacy parameter specified by the data owner. Differential privacy maintains composability, i.e., differential privacy guarantees can be provided even when multiple differentially-private releases are available to an adversary, and can extend to group privacy, i.e., changing a group of k records in the data set induces a change of at most a multiplicative $e^{k\epsilon}$ in the corresponding output distribution [6].

3 ρ-Differential Privacy on Refinement

In this section, we first describe the notations and then formally define refinement based on background knowledge. We present definitions of unbiased refinement

and constrained refinement. We finally introduce our key concept, ρ-**differential privacy on refinement**, and use an illustrating example to show the difference between the proposed ρ-differential privacy on refinement and the classic ϵ-differential privacy.In a differentially private query answering mechanism, the analyst submits queries, the mechanism generates true values for the query, and perturbs them with calibrated noise to derive the responses, then returns the responses to the analyst. Usually, the analyst may possess some background knowledge about the database. With background knowledge, the analyst can refine the responses given by the mechanism, and may obtain more accurate values for his queries.

3.1 Definition

We denote the original database as $\mathcal{D}$, and its neighboring database as $\mathcal{D}'$ which differs from the original database by a single record. The vector-valued query is denoted as $\boldsymbol{Q}$, $\boldsymbol{Q} = (q_1, q_2, \cdots, q_n)^T$. We denote the true value from database D for the query as $\boldsymbol{\mu}$, $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_n)^T$, and the response from database $\mathcal{D}$ for the query as X, $X = (X_1, X_2, \cdots, X_n)^T$. And we denote the true value from database $\mathcal{D}'$ for the query as $\boldsymbol{\mu}'$, $\boldsymbol{\mu}' = (\mu'_1, \mu'_2, \cdots, \mu'_n)^T$, and the response from database $\mathcal{D}'$ for the query as X', $X' = (X'_1, X'_2, \cdots, X'_n)^T$. The randomization mechanism satisfies ϵ-differential privacy, i.e., for an arbitrary set of integers $S = \{i, j, \ldots, k\} \subseteq \{1, \ldots, n\}$,

$$e^{-\epsilon} \leq \frac{\Pr[X_S]}{\Pr[X'_S]} \leq e^{\epsilon} \tag{2}$$

Assume that the user knows some background knowledge about $\mathcal{D}$ and $\mathcal{D}'$ denoted by $\mathcal{B}$ and $\mathcal{B}'$ respectively. For database $\mathcal{D}$, we denote the estimated value as $\widehat{X}$ derived by the analyst from the response using background knowledge, $\widehat{X} = (\widehat{X}_1, \widehat{X}_2, \cdots, \widehat{X}_n)^T$, for database $\mathcal{D}'$, we denote it as $\widehat{X}'$, $\widehat{X}' = (\widehat{X}'_1, \widehat{X}'_2, \cdots, \widehat{X}'_n)^T$.

Definition 2. *(**Refinement**) Given the background knowledge $\mathcal{B}$ on database $\mathcal{D}$, the refinement $\widehat{X} = (\widehat{X}_1, \ldots, \widehat{X}_n)^T$ is the user's estimation on the true value of query $\boldsymbol{Q}(\mathcal{D})$ based on the response X: $\widehat{X} = \mathrm{rf}(X|\mathcal{B}, \mathcal{D})$.*

Similarly, given response X' from $\mathcal{D}'$, the refinement to estimate $\boldsymbol{Q}(\mathcal{D}')$ is $\widehat{X}' = \mathrm{rf}(X'|\mathcal{B}', \mathcal{D}')$.

Definition 3. *(**Unbiased Refinement**) The refinement $\widehat{X}$ is unbiased if $\mathbf{E}(\widehat{X}) = \boldsymbol{\mu}$ stands for any $\boldsymbol{\mu}$.*

Definition 4. *(**Constrained Refinement**) The refinement $\widehat{X}$ is a constrained refinement if $\widehat{X}$ always satisfies the background knowledge $\mathcal{B}$ for any response X.*

The two refinements, $\widehat{X} = (\widehat{X}_1, \ldots, \widehat{X}_n)$ from $\mathcal{D}$ and $\widehat{X}' = (\widehat{X}'_1, \ldots, \widehat{X}'_n)$ from $\mathcal{D}'$, may be mapped to two disjoint spaces by the refinement function $\mathrm{rf}()$. In this

case, either the numerator or the denominator of ratio $\frac{\Pr(\widehat{X}=\boldsymbol{x})}{\Pr(\widehat{X}'=\boldsymbol{x})}$ is 0. However, this difference is due to the refinement strategy and does not disclose any privacy information.

Definition 5. *(ρ-**differential privacy on refinement**) Given the refinements $\widehat{X}$ and $\widehat{X}'$ and an arbitrary set of integers $S = \{i, j, \ldots, k\} \subseteq \{1, \ldots, n\}$, define*

$$\widehat{X}_S = (\widehat{X}_i, \widehat{X}_j, \ldots, \widehat{X}_k) \text{ and } \widehat{X}'_S = (\widehat{X}'_i, \widehat{X}'_j, \ldots, \widehat{X}'_k).$$

Let $\mathcal{R}_S$ and $\mathcal{R}'_S$ be the sets of all possible values of $\widehat{X}_S$ and $\widehat{X}'_S$ respectively. The refinement satisfies differential privacy, if $\mathcal{R}_S \cap \mathcal{R}'_S \neq \varnothing$ and for any subset $\Omega \subseteq \mathcal{R} \cap \mathcal{R}'$ the following inequality stands

$$e^{-\rho} \leq \frac{\Pr[\widehat{X}_S \in \Omega]}{\Pr[\widehat{X}'_S \in \Omega]} \leq e^{\rho}. \tag{3}$$

3.2 An Illustrating Example

Example 1. The analyst submits a vector-valued query $\boldsymbol{Q}$, $\boldsymbol{Q} = (q_1, q_2)^T$, $\max|\boldsymbol{\mu} - \boldsymbol{\mu}'| = 1$, and $\sigma = \frac{1}{\epsilon}$. The analyst has the background knowledge that $\mu_1 + \mu_2 = c$ and $\mu'_1 + \mu'_2 = c'$. One method to refine the response is shown in (4):

$$\widehat{X}_1 = \frac{1}{2}(X_1 + c - X_2),\ \widehat{X}_2 = \frac{1}{2}(X_2 + c - X_1). \tag{4}$$

Equivalently expressed in matrix:

$$\begin{pmatrix} \widehat{X}_1 \\ \widehat{X}_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} c \tag{5}$$

The refinement in (5) belongs to constrained refinement. So we can calculate the ratio $\Pr(\widehat{X}_1 = x_1)/\Pr(\widehat{X}'_1 = x_1)$ to obtain the bound. First, from formula (5), we derive the probability density function of $\widehat{X}_1$ and $\widehat{X}'_1$, shown in (6) and (7).

$$f_{\widehat{X}_1}(x_1) = \frac{1}{2\sigma} \int_{\mathbb{R}} \exp\left\{ -\frac{|2x_1 + x_2 - c - \mu_1| + |x_2 - \mu_2|}{\sigma} \right\} \mathrm{d}x_2 \tag{6}$$

$$f_{\widehat{X}'_1}(x_1) = \frac{1}{2\sigma} \int_{\mathbb{R}} \exp\left\{ -\frac{|2x_1 + x_2 - c' - \mu'_1| + |x_2 - \mu'_2|}{\sigma} \right\} \mathrm{d}x_2 \tag{7}$$

Without loss of generality, we assume that $\mu_1 - \mu'_1 = 1$. When x_1 is sufficiently large, we can simplify formulas (6) and (7) to formulas (8) and (9) respectively.

$$f_{\widehat{X}_1}(x_1) = \frac{1}{2\sigma}(\sigma + 2x_1 - 2\mu_1) \exp\left\{ -\frac{2(x_1 - \mu_1)}{\sigma} \right\} \tag{8}$$

$$f_{\widehat{X}'_1}(x_1) = \frac{1}{2\sigma}(\sigma + 2x_1 - 2\mu'_1) \exp\left\{ -\frac{2(x_1 - \mu'_1)}{\sigma} \right\} \tag{9}$$

The ratio of the two PDFs can then be calculated as shown in (10), which tends to be $e^{2\epsilon}$ for large value of response X_1.

$$\begin{aligned}\frac{f_{\widehat{X}_1}(x_1)}{f_{\widehat{X}_1'}(x_1)} &= \frac{(\sigma + 2x_1 - 2\mu_1)}{(\sigma + 2x_1 - 2\mu_1')} \exp\left\{\frac{2(\mu_1 - \mu_1')}{\sigma}\right\} \\ &= \frac{(\sigma + 2x_1 - 2\mu_1)}{(\sigma + 2x_1 - 2\mu_1')} e^{2\epsilon} \to e^{2\epsilon} \text{ (as } x_1 \to \infty) \end{aligned} \tag{10}$$

So we can conclude that the ratio between the distributions of refinements for databases $\mathcal{D}$ and $\mathcal{D}'$ could be different from the ratio between the distributions of responses. In this example, the classic ϵ-differential privacy incurs 2ϵ-differential privacy on refinement. □

4 Background Knowledge and Refinement Analysis

In this section, we formally the linear constraint based background knowledge and conduct theoretical analysis on how refinement strategies affect differential privacy on refinement. We will use the following scenario as a running example throughout this section. Consider that a data publisher (such as a school) has collected grade information about a group of students and would like to allow the third party to query the data while preserving the privacy of the individuals involved. Assume the analyst submits a simple vector query: $\boldsymbol{Q} = (q_A, q_B, q_C, q_D, q_F, q_p, q_t)$. q_A, q_B, q_C, q_D, and q_F represent the numbers of students receiving grades A, B, C, D, and F respectively; q_p represents the number of passing students (grade D or higher) and q_t represents the query for the number of all the students.

$$\begin{cases} \mu_A + \mu_B + \mu_C + \mu_D - \mu_p = 0 \\ \mu_F + \mu_p - \mu_t = 0 \\ \mu_A + \mu_B = 80 \end{cases} \tag{11}$$

The analyst may have the background knowledge in terms of the linear constraints shown in (11). The first two constraints are by the definition and independent of the underlying database whereas the third constraint holds specifically on the current database.

The analyst may have the background knowledge in terms of the ordering constraint, e.g., $\mu_A \leq \mu_p$. Ordering constraint can also be enforced when the user submits the vector query. For example, the analyst may submit a simple ascending ordering query that shows the number of students in each category. In other words, the analyst knows for sure that $\mu_1 \leq \mu_2 \leq ... \leq \mu_n$ although the responses may not hold the order constraints due to calibrated noises. Similarly the range constraint denotes the true answer of a particular query is within some finite range, e.g., $\mu_A \in [1, 10]$. Range constraints are often implicitly used in the post-process of noisy output. For example, users apply non-negative constraints when dealing with responses with negative values for attributes like age or salary. We will study these types of background knowledge in our future work.

4.1 Refinement with Linear Constraint

Assume that the user knows m linear combinations of the true answers:

$$b_{1i}\mu_1 + b_{2i}\mu_2 + \cdots + b_{ni}\mu_n = c_i, \quad i = 1, \ldots, m.$$

Equivalently, $\boldsymbol{b}_i^T \boldsymbol{\mu} = c_i$, where b_{ji} is the j-th entry of $\boldsymbol{b}_i$. Let $\boldsymbol{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m]$ and $\boldsymbol{c} = (c_1, \ldots, c_m)^T$.

Definition 6. *(**Linear Constraint**) The background knowledge with linear constraint can be expressed as*

$$\boldsymbol{B}^T \boldsymbol{\mu} = \boldsymbol{c},$$

where $\boldsymbol{B}$ is an $n \times m$ matrix and $\boldsymbol{c}$ is an m-dimensional constant vectors.

Under the linear constraint based background knowledge, a constrained refinement $\widehat{X}$ must satisfy $\boldsymbol{B}^T \widehat{X} = \boldsymbol{c}$ for any response X.

Definition 7. *(**Refinement with Linear Constraint**) The refinement $\widehat{X}$ is linear if it can be expressed as*

$$\widehat{X} = \boldsymbol{A}X + \boldsymbol{D}\boldsymbol{c} + \boldsymbol{h}, \tag{12}$$

where $\boldsymbol{A}$ and $\boldsymbol{D}$ are $n \times n$ and $n \times m$ matrices respectively, and $\boldsymbol{h}$ is an n-dimensional constant vector.

4.2 A General Result

Theorem 2. *Suppose that the user possesses the linear background knowledge $\boldsymbol{B}^T\boldsymbol{\mu} = \boldsymbol{c}$ and $\boldsymbol{B}^T\boldsymbol{\mu}' = \boldsymbol{c}'$ for database $\mathcal{D}$ and $\mathcal{D}'$ respectively, and he implements some constrained linear refinement as shown in (12) to estimate $\boldsymbol{\mu}$. Assume* $\mathrm{rank}(\boldsymbol{B}) = m$ *and* $\mathrm{rank}(\boldsymbol{A}) = r = n - m$. *Let $\boldsymbol{U} = (\boldsymbol{U}_1, \boldsymbol{U}_2), \boldsymbol{V} = (\boldsymbol{V}_1, \boldsymbol{V}_2), \Sigma = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$, be the SVD of $\boldsymbol{A}$: $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$, and $\boldsymbol{A}^* = \boldsymbol{V}\begin{pmatrix} \boldsymbol{\Sigma}_1^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_m \end{pmatrix}\boldsymbol{U}^T$. Adding noise from distribution $Lap(\sigma)$, $\sigma = \Delta\boldsymbol{Q}/\epsilon$, would result in*

$$\rho = \frac{\epsilon \|\boldsymbol{A}^*\boldsymbol{D}(\boldsymbol{c} - \boldsymbol{c}') - (\boldsymbol{\mu} - \boldsymbol{\mu}')\|_1}{\|\boldsymbol{\mu} - \boldsymbol{\mu}'\|_1},$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\mu}'$ are from the two databases that achieves $\|\boldsymbol{\mu} - \boldsymbol{\mu}'\|_1 = \Delta\boldsymbol{Q}$.

Proof. Let $\Omega = \{\omega_1, \ldots, \omega_k\} \subseteq \{1, 2, \ldots, n\}$, and P_Ω be the $n \times k$ matrix with $P(\omega_i, i) = 1$ and 0 elsewhere. Similarly, $\bar{\Omega} = \{1, \ldots, n\} - \Omega$. with $n \times (n-k)$ matrix $P_{\bar{\Omega}}$ defined likewise. We can rewrite the refinement function to

$$\widehat{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T X + \boldsymbol{D}\boldsymbol{c} + \boldsymbol{h}.$$

Let $Z = \begin{pmatrix} \boldsymbol{Z}_1 \\ \boldsymbol{Z}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix} \boldsymbol{V}^T X$. With $\boldsymbol{V}^T = \boldsymbol{V}^{-1}$ and $|\boldsymbol{V}| = 1$, we can have that the PDF of Z is

$$f_Z(\boldsymbol{z}) = \frac{1}{|\boldsymbol{\Sigma}_1|} f_X(\boldsymbol{V}\boldsymbol{\Sigma}^*\boldsymbol{z}), \text{ where } \boldsymbol{\Sigma}^* = \begin{pmatrix} \boldsymbol{\Sigma}_1^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}.$$

Let $W = UZ$, $S = \widehat{X} - \boldsymbol{D}\boldsymbol{c} - \boldsymbol{h} = \boldsymbol{U}_1 Z_1$, and $T = \boldsymbol{U}_2 Z_2$. Then, $W = S + T$, $S \in \mathcal{U}_\infty$, $S \in \mathcal{U}_\in$, and the PDF of W is given by

$$f_W(\boldsymbol{w}) = \frac{1}{|\boldsymbol{\Sigma}_1|} f_X(\boldsymbol{V}\boldsymbol{\Sigma}^*\boldsymbol{U}^T\boldsymbol{w}) = \frac{1}{|\boldsymbol{\Sigma}_1|} f_X(\boldsymbol{A}^*\boldsymbol{w}).$$

Notice that S and T are actually the projection of W onto the space spanned by $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ respectively, and the two spaces are orthogonal. For any $\boldsymbol{s} \in \mathcal{U}_\infty$ and $\boldsymbol{t} \in \mathcal{U}_\in$, $W = \boldsymbol{s} + \boldsymbol{t}$ would always give $S = \boldsymbol{s}$. Therefore, if $\boldsymbol{s} \in \mathcal{U}_\infty$, the PDF of S is given by

$$\begin{aligned} f_S(\boldsymbol{s}) &= \frac{1}{|\boldsymbol{\Sigma}_1|} \int_{\mathcal{U}_\in} f_X[\boldsymbol{A}^*(\boldsymbol{s}+\boldsymbol{t})]\mathrm{d}\boldsymbol{t} = \frac{1}{|\boldsymbol{\Sigma}_1|} \int f_X\left[\boldsymbol{A}^*\left(\boldsymbol{s} + U\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{z}_2\end{pmatrix}\right)\right] \mathrm{d}U\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{z}_2\end{pmatrix} \\ &= \frac{1}{|\boldsymbol{\Sigma}_1|} \int f_X\left(\boldsymbol{A}^*\boldsymbol{s} + \boldsymbol{V}_2\boldsymbol{z}_2\right) \mathrm{d}\boldsymbol{z}_2. \end{aligned}$$

Hence, the PDF of $\widehat{X}$ can be given by

$$f_{\widehat{X}}(\boldsymbol{x}) = \frac{1}{|\boldsymbol{\Sigma}_1|} \int f_X\left[\boldsymbol{A}^*(\boldsymbol{x} - \boldsymbol{D}\boldsymbol{c} - \boldsymbol{h}) + \boldsymbol{V}_2\boldsymbol{z}_2\right] \mathrm{d}\boldsymbol{z}_2,$$

if $\boldsymbol{U}_2^T(\boldsymbol{x} - \boldsymbol{D}\boldsymbol{c} - \boldsymbol{h}) = 0$, and $f_{\widehat{X}}(\boldsymbol{x}) = 0$ otherwise.

Similarly, the PDF of $\widehat{X}'$ is given by

$$f_{\widehat{X}'}(\boldsymbol{x}) = \frac{1}{|\boldsymbol{\Sigma}_1|} \int f_{X'}\left[\boldsymbol{A}^*(\boldsymbol{x} - \boldsymbol{D}\boldsymbol{c}' - \boldsymbol{h}) + \boldsymbol{V}_2\boldsymbol{z}_2\right] \mathrm{d}\boldsymbol{z}_2,$$

if $\boldsymbol{U}_2^T(\boldsymbol{x} - \boldsymbol{D}\boldsymbol{c}' - \boldsymbol{h}) = 0$, and $f_{\widehat{X}}(\boldsymbol{x}) = 0$ otherwise.

Notice that $\widehat{X}_\Omega = P_\Omega^T\widehat{X}$, $\widehat{X}_{\bar{\Omega}} = P_{\bar{\Omega}}^T\widehat{X}$ and $\widehat{X} = P_\Omega\widehat{X}_\Omega + P_{\bar{\Omega}}\widehat{X}_{\bar{\Omega}}$. The PDF of $\widehat{X}_\Omega$ can be expressed as

$$f_{\widehat{X}_\Omega}(\boldsymbol{x}_\Omega) = \frac{1}{|\boldsymbol{\Sigma}_1|} \iint_{\mathcal{D}(\boldsymbol{x}_\Omega)} \mathrm{d}\boldsymbol{z}_2 \mathrm{d}\boldsymbol{x}_{\bar{\Omega}} f_X\left[\boldsymbol{A}^*\left(P_\Omega\boldsymbol{x}_\Omega + P_{\bar{\Omega}}\boldsymbol{x}_{\bar{\Omega}} - \boldsymbol{D}\boldsymbol{c} - \boldsymbol{h}\right) + \boldsymbol{V}_2\boldsymbol{z}_2\right], \tag{13}$$

where $\mathcal{D}(\boldsymbol{x}_\Omega) = \{\boldsymbol{x}_{\bar{\Omega}} : \boldsymbol{U}_2^T(P_\Omega\boldsymbol{x}_\Omega + P_{\bar{\Omega}}\boldsymbol{x}_{\bar{\Omega}} - \boldsymbol{D}\boldsymbol{c} - \boldsymbol{h}) = 0\}$.

The PDF of $\widehat{X}'_\Omega$ can be derived in a similar manner. When the ratio of the integral kernels in (13) is bounded, i.e.,

$$e^{-\epsilon} \leq \frac{f_X\left[\boldsymbol{A}^*(\boldsymbol{x} - \boldsymbol{D}\boldsymbol{c} - \boldsymbol{h}) + \boldsymbol{V}_2\boldsymbol{z}_2\right]}{f_{X'}\left[\boldsymbol{A}^*(\boldsymbol{x} - \boldsymbol{D}\boldsymbol{c}' - \boldsymbol{h}) + \boldsymbol{V}_2\boldsymbol{z}_2\right]} \leq e^{\epsilon}, \tag{14}$$

the ratio of the integrals, $f_{\widehat{X}_\Omega}(\boldsymbol{x}_\Omega)/f_{\widehat{X}'_\Omega}(\boldsymbol{x}_\Omega)$, is also bounded by $[e^{-\epsilon}, e^{\epsilon}]$. Note that f_X and f_X are the Laplace distribution p.d.f., and hence

$$\begin{aligned}\frac{f_{\widehat{X}_\Omega}(\boldsymbol{x}_\Omega)}{f_{\widehat{X}'_\Omega}(\boldsymbol{x}_\Omega)} \leq & \frac{f_X\left[\boldsymbol{A}^*(\boldsymbol{x}-\boldsymbol{D}\boldsymbol{c}-\boldsymbol{h})+\boldsymbol{V}_2\boldsymbol{z}_2\right]}{f_{X'}\left[\boldsymbol{A}^*(\boldsymbol{x}-\boldsymbol{D}\boldsymbol{c}'-\boldsymbol{h})+\boldsymbol{V}_2\boldsymbol{z}_2\right]} \\ \leq & \frac{\exp\{\frac{1}{\sigma}\|\boldsymbol{A}^*(\boldsymbol{x}-\boldsymbol{D}\boldsymbol{c}-\boldsymbol{h})+\boldsymbol{V}_2\boldsymbol{z}_2-\boldsymbol{\mu}\|_1\}}{\exp\{\frac{1}{\sigma}\|\boldsymbol{A}^*(\boldsymbol{x}-\boldsymbol{D}\boldsymbol{c}'-\boldsymbol{h})+\boldsymbol{V}_2\boldsymbol{z}_2-\boldsymbol{\mu}'\|_1\}} \\ \leq & \exp\left\{\pm\frac{1}{\sigma}\|\boldsymbol{A}^*\boldsymbol{D}(\boldsymbol{c}-\boldsymbol{c}')-(\boldsymbol{\mu}-\boldsymbol{\mu}')\|_1\right\}.\end{aligned}$$

Therefore, when the noise is added according to the classical schema , i.e., take $\sigma = \frac{\Delta \boldsymbol{Q}}{\epsilon} = \frac{1}{\epsilon}\|\boldsymbol{\mu}-\boldsymbol{\mu}'\|_1$, we can have $f_{\widehat{X}_\Omega}(\boldsymbol{x}_\Omega)/f_{\widehat{X}'_\Omega}(\boldsymbol{x}_\Omega)$ is bounded by $e^{\pm\rho}$ where

$$\rho = \frac{\epsilon\|\boldsymbol{A}^*\boldsymbol{D}(\boldsymbol{c}-\boldsymbol{c}')-(\boldsymbol{\mu}-\boldsymbol{\mu}')\|_1}{\|\boldsymbol{\mu}-\boldsymbol{\mu}'\|_1}. \qquad \square$$

A special case of the above result is that $\rho = \epsilon$ when $\boldsymbol{c} = \boldsymbol{c}'$ (no difference on constants of linear background constraints over two neighbor databases). However, in practice, $\boldsymbol{c}$ could be different from $\boldsymbol{c}'$ (refer to the example shown in Appendix), the ρ-differential privacy on refinement is generally different from the ϵ-differential privacy. A direct result from the above theorem is that, in order to guarantee $e^{-\epsilon} \leq f_{\widehat{X}_\Omega}(\boldsymbol{x}_\Omega)/f_{\widehat{X}'_\Omega}(\boldsymbol{x}_\Omega) \leq e^{\epsilon}$, we can choose $\sigma = \frac{1}{\epsilon}\max_{\mathcal{D},\mathcal{D}'}$ $\|\boldsymbol{A}^*\boldsymbol{D}(\boldsymbol{c}-\boldsymbol{c}')-(\boldsymbol{\mu}-\boldsymbol{\mu}')\|_1$.

4.3 The Best Linear Refinement

Consider the following least square refinement based on the linear background knowledge:

$$\min \|\widehat{X}-X\|_2 \quad \text{s.t. } \boldsymbol{B}^T\widehat{X} = \boldsymbol{c}. \tag{15}$$

Theorem 3. *The least square refinement from the optimization problem in* (15) *is given by*

$$\widehat{X} = \left[\boldsymbol{I}-\boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\right]X + \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{c}. \tag{16}$$

The refinement shown in (16) *is a constrained unbiased refinement. It has the minimum variance of* $\widehat{X}_i$, $i = 1, \ldots, n$, *among all linear unbiased refinements.*

Proof. The Lagrange function of (16) is

$$\mathcal{L} = (\widehat{X}-X)^T(\widehat{X}-X) - 2\boldsymbol{\Lambda}(\boldsymbol{B}^T\widehat{X}-\boldsymbol{c}).$$

Taking $\frac{\partial \mathcal{L}}{\partial \widehat{X}} = \boldsymbol{0}$, we can have $\widehat{X} = X + \boldsymbol{B}^T\boldsymbol{\Lambda}$, and hence

$$\begin{aligned}\boldsymbol{B}^T\widehat{X} &= \boldsymbol{B}^T(X+\boldsymbol{B}\boldsymbol{\Lambda}) = \boldsymbol{c} \\ \boldsymbol{\Lambda} &= (\boldsymbol{B}^T\boldsymbol{B})^{-1}(\boldsymbol{c}-\boldsymbol{B}^TX) \\ \widehat{X} &= X + \boldsymbol{B}[(\boldsymbol{B}^T\boldsymbol{B})^{-1}(\boldsymbol{c}-\boldsymbol{B}^TX)]\end{aligned}$$

Equivalently, $\widehat{X}$ can be expressed as follows:

$$\widehat{X} = [\boldsymbol{I} - \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T]X + \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{c}.$$

Next, we show that $\widehat{X}$ is a unbiased constrained refinement:

$$\begin{aligned}\boldsymbol{B}^T\widehat{X} &= \boldsymbol{B}^T X - \boldsymbol{B}^T\boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T X + \boldsymbol{B}^T\boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{c} = \boldsymbol{c}.\\ \mathbf{E}(\widehat{X}) &= [\boldsymbol{I} - \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T]\,\mathbf{E}(X) + \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{c}\\ &= \boldsymbol{\mu} - \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{\mu} + \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{c} = \boldsymbol{\mu}.\end{aligned}$$

Next, we prove the minimal variance property. We use $\boldsymbol{M}$ to denote the matrix $\boldsymbol{I} - \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T$. Then we have $\boldsymbol{M} = \boldsymbol{M}^T$, $\boldsymbol{M}\boldsymbol{M}^T = \boldsymbol{M}$. We can further show that $(\boldsymbol{A} - \boldsymbol{I})\boldsymbol{M} = \boldsymbol{0}$. Notice that the following equalities stand for any $\boldsymbol{\mu}$,

$$\mathbf{E}(\widehat{X}) = \boldsymbol{A}\,\mathbf{E}(X) + \boldsymbol{D}\boldsymbol{c} + \boldsymbol{h} \Rightarrow \boldsymbol{\mu} = \boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{D}\boldsymbol{B}^T\boldsymbol{\mu} + \boldsymbol{h}.$$

We can thus have $\boldsymbol{I} - \boldsymbol{A} = \boldsymbol{D}\boldsymbol{B}^T$ and $\boldsymbol{h} = \boldsymbol{0}$. Therefore,

$$(\boldsymbol{A} - \boldsymbol{I})\boldsymbol{M} = -\boldsymbol{D}\boldsymbol{B}^T[\boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T - \boldsymbol{I}] = \boldsymbol{0}.$$

Since $\mathrm{Cov}(\widehat{X}) = \boldsymbol{A}\,\mathrm{Cov}(X)\boldsymbol{A}' = 2\sigma^2\boldsymbol{A}\boldsymbol{A}^T$, $\mathbf{V}(\widehat{X}_i)/2\sigma^2$ is the i-th diagonal entry of matrix $\boldsymbol{A}\boldsymbol{A}^T$. With $\boldsymbol{M}\boldsymbol{M}^T = \boldsymbol{M}$ and $(\boldsymbol{A} - \boldsymbol{I})\boldsymbol{M} = \boldsymbol{0}$, we can have

$$\boldsymbol{A}\boldsymbol{A}^T = [(\boldsymbol{A} - \boldsymbol{M}) + \boldsymbol{M}][(\boldsymbol{A} - \boldsymbol{M})^T + \boldsymbol{M}^T] = (\boldsymbol{A} - \boldsymbol{M})(\boldsymbol{A} - \boldsymbol{M})^T + \boldsymbol{M}.$$

Since $(\boldsymbol{A}-\boldsymbol{M})(\boldsymbol{A}-\boldsymbol{M})^T$ is the semi-positive definite matrix, and the the diagonal entries are non-negative, and hence $(\boldsymbol{A}\boldsymbol{A}^T)_{ii} \geq \boldsymbol{M}_{ii}$ with $\boldsymbol{A} = \boldsymbol{M}$ minimizes $(\boldsymbol{A}\boldsymbol{A}^T)_{ii}$, $i = 1, \ldots, n$.

5 Conclusion and Further Discussion

In this paper we have introduced a new definition, ρ-differential privacy on refinement, to quantify the change of distributions of results after refinements. We focus on one representative refinement, the linear refinement with background knowledge as linear constraints and investigate the relationship between the classic ϵ-differential privacy (on responses) and our ρ-differential privacy on refinement.

Three techniques were proposed to use constraints to boost accuracy of answering range queries over histograms [3–5]. The refinement approach (also called constrained inference) [3] focused on using consistency constraints, which should hold over the noisy output, to improve accuracy for a variety of correlated histogram queries. The idea was to find a new set of answers $\bar{q}$ that is the closet set to the set of noisy answers $\tilde{q}$ and that also satisfies the consistency constraints. The proposed approach, *the minimum least squares solution*, was a special case of our linear refinement with linear constraints presented in this paper. Hay et al. in [3] also showed that the inferred $\bar{q}$ based on the minimum L_2 solution satisfies

ϵ-differential privacy. In our work, we introduced the general linear refinement and showed the conditions on when the refinement based on the general linear constraints achieves the same ϵ-differential privacy as defined over distributions of responses. The authors extended to refine degree distribution of networks under the context of publishing private network data [7]. Xiao et al. in [4] proposed an approach based on the Haar wavelet. In [5], the authors unified the two approaches [3, 4] in one general framework based on the matrix mechanism that can answer a workload of predicate counting queries.

One key question is whether background knowledge can be exploited by adversaries to breach privacy. It is well known that for the pre-processing based privacy preserving data mining models, several works [8, 9] showed the risks of privacy disclosure by incorporating a user's background knowledge in the reasoning process. In contrast, in the context of differential privacy, the authors in [1, 2] stated that differential privacy provides formal privacy guarantees that do not depend on an adversary's background knowledge (including access to other databases) or computational power. In [10], the authors gave an explicit formulation of *resistance to background knowledge*. The formulation follows the implicit statement: *Regardless of external knowledge, an adversary with access to the sanitized database draws the same conclusions whether or not my data is included in the original data.* They presented a mathematical formulation of background knowledge and belief. The belief is modeled by the posteriori distribution: given a response, the adversary draws his belief about the database using Baye's rule to obtain a posterior refinement. In [3–5], the authors also stated that the refinement has *no impact* on the differential privacy guarantee. This is because the analyst performs the refinement without access to the private data, using only the constraints and the perturbed responses. The perturbed responses are simply the output of a differentially private mechanism and post-processing of responses cannot diminish the rigorous privacy guarantee.

In [11], the authors examined the assumptions of differential privacy from the data generation perspective and proposed a participation-based guideline - *does deleting an individual's tuple erase all evidence of the individual's participation in the data-generation process?* - for determining the applicability of differential privacy. They showed that the privacy guarantee from differential privacy can degrade when applied to social networks or when deterministic statistics (of a contingency table) have been previously released. The deterministic statistics can be modeled as linear constraints with fixed $\boldsymbol{c}$ values. In this case, $\boldsymbol{c}$ could be different from $\boldsymbol{c}'$. Based on our Theorem 2, the ρ-differential privacy on refinement is different from the ϵ-differential privacy and we have to add larger noise to prevent privacy breaches. In practice, the adversary may possess any kind of background knowledge, which may even include the a-priori knowledge of the exact values of all other $n-1$ individuals. We refer readers to the example shown in Appendix where the adversary can exploit the background knowledge of the other $n-1$ individuals in the database to infer the value of a specific individual. We argue that the privacy breach is caused by the combination of the randomization mechanism and the background knowledge. In our future work, we would

explore whether refinements with some particular background knowledge (e.g., ordering or range constraints) can incur privacy breaches, i.e., enabling the adversary to draw *significantly* different beliefs about the databases.

Acknowledgments. This work was supported in part by U.S. National Science Foundation IIS-0546027 and CCF-0915059.

References

1. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating Noise to Sensitivity in Private Data Analysis. In: Halevi, S., Rabin, T. (eds.) TCC 2006. LNCS, vol. 3876, pp. 265–284. Springer, Heidelberg (2006)
2. Dwork, C.: A Firm Foundation for Private Data Analysis. Communications of the ACM (January 2011)
3. Hay, M., Rastogi, V., Miklau, G., Suciu, D.: Boosting the Accuracy of Differentially Private Histograms Through Consistency. Proceedings of the VLDB Endowment 3(1) (2010)
4. Xiao, X., Wang, G., Gehrke, J.: Differential Privacy via Wavelet Transforms. In: Proceedings of the 26th IEEE International Conference on Data Enginering, pp. 225–236. IEEE (2010)
5. Li, C., Hay, M., Rastogi, V., Miklau, G., McGregor, A.: Optimizing Linear Counting Queries Under Differential Privacy. In: Proceedings of the 29th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems of Data, pp. 123–134. ACM (2010)
6. Dwork, C., Lei, J.: Differential Privacy and Robust Statistics. In: Proceedings of the 41st Annual ACM Symposium on Theory of Computing, pp. 371–380. ACM (2009)
7. Hay, M., Li, C., Miklau, G., Jensen, D.: Accurate Estimation of the Degree Distribution of Private Networks. In: Proceedings of the 9th IEEE International Conference on Data Mining, pp. 169–178. IEEE (2009)
8. Martin, D., Kifer, D., Machanavajjhala, A., Gehrke, J., Halpern, J.: Worst-Case Background Knowledge for Privacy-Preserving Data Publishing. In: Proceedings of the 26th IEEE International Conference on Data Enginering. IEEE (2007)
9. Du, W., Teng, Z., Zhu, Z.: Privacy-MaxEnt: Integrating Background Knowledge in Privacy Quantification. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, ACM (2008)
10. Ganta, S., Kasiviswanathan, S., Smith, A.: Composition Attacks and Auxiliary Information in Data Privacy. In: Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 265–273. ACM (2008)
11. Kifer, D., Machanavajjhala, A.: No Free Lunch in Data Privacy. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 193–204. ACM (2011)

A Appendix

A.1 Example When $c \neq c'$

Database $\mathcal{D}$ with n records is obtained by adding one record to database $\mathcal{D}_0$. Every record in $\mathcal{D}$ belongs to one of two categories. The attacker knows that

in $\mathcal{D}_0$, $k\mu_1 = \mu_2$, where μ_i denotes the count of category i in $\mathcal{D}_0$, $i = 1, 2$. The added record belongs to either of the two categories, denoted by $\mathcal{D}'$ and $\mathcal{D}''$ respectively. Let $\boldsymbol{\mu}' = \binom{\mu_1'}{\mu_2'}$ and $\boldsymbol{\mu}'' = \binom{\mu_1''}{\mu_2''}$ be the counts of $\mathcal{D}'$ and $\mathcal{D}''$ respectively. The background knowledge can be expressed as:

$$\begin{aligned} &\text{if } \mathcal{D}' \text{ is true: } k\mu_1' - \mu_2' = \boldsymbol{B}^T\boldsymbol{\mu}' = \boldsymbol{c}' = k, \\ &\text{if } \mathcal{D}'' \text{ is true: } k\mu_1'' - \mu_2'' = \boldsymbol{B}^T\boldsymbol{\mu}'' = \boldsymbol{c}'' = -1, \end{aligned}$$

where $\boldsymbol{B} = \binom{k}{-1}$.

Response $X = (X_1, X_2)$ is obtained by adding noise $Lap(\frac{2}{\epsilon})$. Next, we show that, for $\mathcal{D}'$ and $\mathcal{D}''$, the refinements $\widehat{X}'$ and $\widehat{X}''$ do not satisfy differential privacy. Consider the following refinement:

$$\text{For } \mathcal{D}' : \widehat{X}_1 = \frac{X_1 + X_2 + k}{k+1}, \text{ and } \widehat{X}_2 = X_2; \tag{17}$$

$$\text{For } \mathcal{D}'' : \widehat{X}_1 = \frac{X_1 + X_2 - 1}{k+1}, \text{ and } \widehat{X}_2 = X_2. \tag{18}$$

Comparing (17) and (18) with the general linear refinement formula in (12), we can have

$$\boldsymbol{A} = \begin{pmatrix} \frac{1}{k+1} & \frac{1}{k+1} \\ 0 & 1 \end{pmatrix}, \ \boldsymbol{D} = \begin{pmatrix} \frac{1}{k+1} \\ 0 \end{pmatrix}, \text{ and } \boldsymbol{h} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

When X_1 satisfies $x_1 \geq \frac{\mu_1' + \mu_2' + k}{k+1}$, we can have

$$\begin{aligned} f_{\widehat{X}_1'}(x_1) &= \int_{\mathbb{R}} f_{X_1}(z) f_{X_2}[(k+1)x_1 - k - z]\mathrm{d}z \\ &\propto \int_{\mathbb{R}} \exp\left\{ -\frac{|z - \mu_1'| + |z - (k+1)x_1 + k + \mu_2'|}{\sigma} \right\} \mathrm{d}z \\ &= \exp\left\{ \frac{n + k - (k+1)x_1}{\sigma} \right\} [(k+1)x_1 - k - n] \\ &\quad (\text{note } n = \mu_1' + \mu_2'). \end{aligned}$$

For $\mathcal{D}''$, we can similarly have that when $x_1 \geq \frac{\mu_1'' + \mu_2'' - 1}{k+1}$,

$$f_{\widehat{X}_1''}(x_1) \propto \exp\left\{ \frac{n - 1 - (k+1)x_1}{\sigma} \right\} [(k+1)x_1 + 1 - n].$$

With $\sigma = \frac{2}{\epsilon}$ (satisfying ϵ-differential privacy), we can have:

$$\lim_{x_1 \to \infty} \frac{f_{\widehat{X}_1'}(x_1)}{f_{\widehat{X}_1''}(x_1)} = \exp\left[\frac{(k+1)\epsilon}{2} \right].$$

Therefore, the ratio $f_{\widehat{X}_1'} / f_{\widehat{X}_1''}$ reaches $e^{\frac{(k+1)\epsilon}{2}}$ for sufficiently large $X_1 + X_2$, which indicates the adversary can tell which database of $\mathcal{D}'$ and $\mathcal{D}''$ the response is from. In other words, the adversary can derive the value of the added record by refinement.