DBSJ Japanese Journal
Vol. 15-J, Article No. 11, March 2017

相関データにおけるベイジアン差分プライバシー

Bayesian Differential Privacy on Correlated Data

楊 斌♡ 佐藤 一誠♢ 中川 裕志♠

Bin YANG Issei SATO
Hiroshi NAKAGAWA

差分プライバシーはデータプライバシーを保護するための強力なプライバシー基準である。データに相関がある場合、差分プライバシーではデータのプライバシーが漏れるリスクがあることが知られている。この問題を解決するために、我々は、以下の三点について考察する：(1) データ相関性のプライバシーに及ぼす影響、(2) 攻撃者の事前知識がプライバシーに及ぼす影響、(3) 相関性のあるデータにおける任意の事前知識を持つ攻撃者に保護できる汎用なプライバシー摂動アルゴリズム。

上記三点に関する考察から、本稿では、Kifer らの提案した Pufferfish プライバシーに基づき、「ベイジアン差分プライバシー」という新しいプライバシー定義を提案する。この定義を用いることで、相関性の有無にかかわらず、任意のデータのプライバシーを保護でき、任意の事前知識を持つ攻撃者の攻撃からプライバシーを守ることができる。

更に、データの任意の構造に対してデータ相関性を表現できる「ガウシアン相関モデル」を提案する。このモデルに基づき、ベイジアン差分プライベートな摂動アルゴリズムの生成手法を提案する。これによって、効率的に摂動アルゴリズムを設計することができるようになる。

1 はじめに

ビッグデータと呼ばれる大量かつ多様なデータの活用がさまざまな業界で利用されるにともない、個人情報漏洩やプライバシー侵害は多数起こっている。プライバシー保護技術には、摂動法、匿名化法と暗号法 (秘密計算法) の三種類がある。本稿では、摂動法によるプライバシー保護方法に焦点を当てる。ある組織が個人のデータを収集し、そのデータに対するクエリ結果（クエリによって生成した結果）を計算し、データの利用者 (例えば、研究機関) と共有することで、ビッグデータは最大限に活用される。その一方、公開されたクエリ結果から、個人情報やプライバシーが漏れるリスクがある。個人情報やプライバシーを推定されないために、公開する前にノイズを加える方法がある。このような方法は摂動法と呼ばれる。

近年、摂動法の手法として、差分プライバシー [1] という概念が注目を集めている。差分プライバシーは、攻撃者の背景知識にかかわらず、強いプライバシー基準を提供する。直観的には、個体の変化による公開データに及ぼす影響が顕著ではないことを保証している。

♡ 非会員 楽天技術研究所
yangbin64@gmail.com
♢ 正会員 東京大学 大学院新領域創成科学研究科
sato@k.u-tokyo.ac.jp
♠ 非会員 東京大学 情報基盤センター
nakagawa@dl.itc.u-tokyo.ac.jp

表 1: 差分プライバシーの妥当性

データ＼攻撃者	強い	任意
独立	妥当	妥当
相関あり	妥当	妥当でない

1.1 相関データ

差分プライバシーが強力なプライバシー基準であることは広く認められているが、データに相関性がある場合、プライバシーを漏れる可能性があることは知られている [6]。10 人を含むソーシャルネットワークを考えてみよう。このネットワークにいる人々は、家族や友達や同僚など様々な関係がある。例えば、ある “攻撃者” はこのネットワークの構造 (人々の関係) を知っているが、人々の個人情報を知らない。“この 10 人に、インフルエンザである人は何人がいるか？” というクエリを聞くと、特定の人はインフルエンザであるかどうかの情報を推定できるかもしれない。個人の健康情報を保護するため、ラプラシアンノイズ $Lap(1/\epsilon)$ をクエリ結果に加えてから公開する必要がある。ラプラシアンノイズ $Lap(1/\epsilon)$ が $\epsilon-$ 差分プライバシーを満たすことは知られているが、(1) この 10 人全員が無相関な場合、特定の人がネットワークに含まれるかどうかにかかわらず、公開されたクエリ結果の確率分布はほとんど同様なので、その人の情報は推定しにくい、(2) この 10 人全員は極めて強い相関性を持ち（例えば、10 人の家族)、全員はほとんど同じ状態である場合、正しいクエリの値は 0 または 10 である確率が 1 に近い。ゆえに、一人の状態の変化はクエリ結果に 10 倍の変化を引き起こす。同じ水準のプライバシーを保証するために、ラプラシアンノイズ $Lap(1/10\epsilon)$ を加えることが必要である、(3) 中間的な例として、この 10 人は弱い相関性を持つ場合 (例えば、10 人の友達)、特定の人の情報が他人に影響するが、クエリの結果は固定な値ではない。この場合、加えるノイズの大きさが問題となる。

一方、実世界からデータの相関性を取得するのは比較的に容易である。例えば、ウェブクローラを用いることで、ソーシャルネットワークの構造を確かめることができる。“政治的傾向がリベラル派と保守派である人はそれぞれ何人いるか” というクエリの結果が与えられたら、特定な人の政治的傾向が推定されやすい。以上のようの状況から、データ相関性のプライバシーに及ぼす影響を慎重に考察するのが非常に重要であるといえる。

1.2 攻撃者の事前知識

差分プライバシーは、最も強い攻撃者の攻撃からプライバシーを保護することができる。潜在的な仮定としては、攻撃目標以外全ての情報を持つ、最も強い攻撃者は公開されたクエリ結果により、攻撃目標に関する最も多い情報を推定できる。データに相関性がない、又は攻撃者が最も強い場合 (攻撃目標以外全ての情報を知っている) には、差分プライバシーでプライバシーを保護することができる。しかし、データに相関性があり、攻撃者一部の事前知識しか持たない場合、プライバシーが侵害される恐れがある (表 1)。

2 関連研究

敵対的プライバシーというのは攻撃されることに関する事前確率と事後確率が明示的に表現されるプライバシー定義である。センシティブな変数 x とクエリ結果 r を確率変数とする。Miklau と Suciu[9] に提案された Perfect プライバシーとは、与えられた r に対し、$\Pr(x|r) \equiv \Pr(x)$ が成り立つことである。直観的に、この定義で、r が与えられた時の x の事後確率分布はその事前確率分布は全く同じである。残念ながら、一般的に、Perfect プライバシーを満たす出力はユーティリティ（データの有用性）がない。この定義を基づき、より緩和的な定義も考慮されている。Dwork ら [1] に提案された差分プライバシーが最も厳密なプライバシー基準であることは広く認識されている。

表 2: 記号

Symbol	Description
$\mathbf{x}$	データベースの事例 $\{x_1, x_2, \cdots, x_n\}$.
n	データベースのタプルの数 $\mathbf{x}$.
$[n]$	タプル集合 $\{1, 2, \cdots, n\}$.
i	攻撃目標となるタプル.
$\mathcal{K}, \mathcal{U}$	既知／未知タプル集合.
C	攻撃目的と既知タプル $\{i\} \cup \mathcal{K}$.
x_i, x_i'	タプル $i \in [n]$ の二つの事例.
$\mathbf{x}_{-i}$	データベース $\mathbf{x}$ から x_i を除いた集合.
$\mathbf{x}'$	データベース $\mathbf{x}$ に x_i を x_i' へ変換した集合.
$\mathbf{x}_\mathcal{K}, \mathbf{x}_\mathcal{U}$	既知／未知タプルの事例.
$\mathcal{A}(i, \mathcal{K})$	ある $\mathcal{K}$ を持つ x_i を攻撃したい攻撃者.
f	データベース $\mathbf{x}$ 上のクエリ関数.
q	正しいクエリ結果 $q = f(\mathbf{x})$.
$\mathcal{M}$	$\mathbf{x}$ 上のランダムメカニズム.
r	$\mathcal{M}$ に生成されたランダムリクエスト.

Kifer と Machanavajjhala[6] は上記の事例を考えたうえで、差分プライバシーにおける重要な問題を提起した。要するに、データが強い相関性を持つ場合、クエリ出力からセンシティブな情報を推定しやすいので、差分プライバシーでプライバシーを保証しないことがある。Kifer と Machanavajjhala[6] は隣接（相関あり）の概念を再定義することで、確定な制約を持つデータベースにも適用できるように、差分プライバシーの定義を修正した。Gehrke ら [3, 2] も相関データに注目し、zero-knowledge プライバシーという概念を提案した。提案された摂動法メカニズムで、相関データのプライバシーを保護できるが、集計関数の選択は難しい。ベイジアンプライバシーの定義とする Pufferfish[7] は本研究で提案された新しいプライバシー定義と類似のフレームワークである。Pufferfish に啓発された Blowfish[5] は特殊なプライバシー定義である。この定義を用いることで、必要な制限だけを特定し、出力結果のユーティリティを改善した。もうひとつのベイジアンプライバシーの定義 [12] では、タプルにある各属性はベイジアンネットワークと見なされる。Rastogi ら [10] は相関データを考慮し、差分プライバシーより弱い敵対的プライバシー定義を提案し、出力結果のユーティリティを改善した。[4] では、ソーシャルネットワーク向けの差分プライベートなメカニズムが提案された。

3 プライバシーと相関データ

n 個のタプルを含むデータベースを $[n]$ と表し、i 番目のタプルを単に i と表す。つまり、データベースを $[n] = \{1, 2, \cdots, n\}$ と表記する。データベース自体（タプルの変数の集合）である $\mathbf{x}$ を $\{x_1, x_2, \cdots, x_n\}$ とする。このデータベースを入力とするクエリ関数 f の結果 $q = f(\mathbf{x})$ を公開するとき、各タプルのプライバシーを保護するため、摂動されたクエリ結果 $r = \mathcal{M}(\mathbf{x})$ だけを公開する。ただし、$\mathcal{M}$ はあるランダム関数であり、確率分布 $\Pr(\mathcal{M}(\mathbf{x}) \in S)$ あるいは $\Pr(r \in S|\mathbf{x})$ で表せる ($S \subseteq Range(\mathcal{M})$)。$\mathcal{M}$ はクエリ q のランダム関数として扱える場合、$\Pr(r \in S|q)$ とも表せる。本稿でよく使われる記号を表 2 に載せる。

3.1 仮定

以下に、三つの仮定を記述する。

3.1.1 データの相関性

全てのタプルはランダム変数とみるので、その相関性はタプルの同時分布 $\Pr(x_1, x_2, \cdots, x_n)$ で表せる。マルコフ確率場を用い、この同時分布をグラフで表現できる。

仮定 1：タプル間は独立であるかまたは相関性があり、その相関性はだれでも知ることができる。

3.1.2 攻撃者

攻撃目標とするタプルと既知タプルセットと未知タプルセットを明示的に指定することで、任意の攻撃者を $\mathcal{A}(i, \mathcal{K})$ で定義する。ただし、タプル $i \in [n]$ は攻撃目標タプルであり、タプルセット $\mathcal{K} \subseteq [n] \setminus \{i\}$ を既知タプルセットとし、$\mathcal{U} = [n] \setminus \{i\} \setminus \mathcal{K}$ を未知タプルセットとする。特に、攻撃者 $\mathcal{A}(i, [n] \setminus \{i\})$ を強い攻撃者といい、攻撃者 $\mathcal{A}(i, \phi)$ を弱い攻撃者という。

仮定 2：任意の背景知識 $\mathcal{K} \subseteq [n] \setminus \{i\}$ を持つ攻撃者はタプル i を攻撃する。

3.1.3 メカニズム

多くの暗号アルゴリズムと同じよう、摂動メカニズム自体も公開されるとする。

仮定 3：条件確率分布 $\Pr(r \in S|\mathbf{x})$ と同値であるランダム関数 $\mathcal{M}(\mathbf{x})$ は公開されている。

情報の利用者は攻撃者となることもあるので、本稿で考えられた攻撃モデルは、データ相関性、メカニズム $\mathcal{M}$、出力 r と攻撃者の背景知識 $\mathcal{K}$ を用い、攻撃目標であるタプル i の情報を推定する問題になる。

3.2 差分プライバシーとラプラシアンメカニズム

上記の記号を用いて、差分プライバシーの定義を再掲する。

定義 3.1 (差分プライバシー) あるランダムメカニズム $\mathcal{M}$ が下記の条件を満たせば、差分プライバシーという。

$$DP(\mathcal{M}) := \sup_{i, \mathbf{x}_{-i}, x_i, x_i', S} \log \frac{\Pr(r \in S|x_i, \mathbf{x}_{-i})}{\Pr(r \in S|x_i', \mathbf{x}_{-i})} \leq \epsilon. \quad (1)$$

ラプラシアンメカニズム [1] では、クエリ結果にラプラス分布 $p(z) \sim Lap(\lambda)$ に従うノイズを加え、その結果を公開する。$\lambda = S(f)/\epsilon$ の場合、公開された結果は ϵ–差分プライバシーを満たす。ただし、$S(f) = \sup \|f(\mathbf{x}) - f(\mathbf{x}')\|_1$ は関数 f についてのセンシティブ（感度）と呼ばれる。$r = f(\mathbf{x}) + z$ だから、出力の密度はそれぞれ $p(r|\mathbf{x}) = \frac{1}{\lambda} e^{-\frac{1}{\lambda}|r - f(\mathbf{x})|}$ と $p(r|\mathbf{x}') = \frac{1}{\lambda} e^{-\frac{1}{\lambda}|r - f(\mathbf{x}')|}$ である。ϵ が小さいなら、ϵ–差分プライバシーでは $p(r|\mathbf{x}) \approx p(r|\mathbf{x}')$ であることを保証できる。

3.3 差分プライバシーの妥当性

差分プライバシーは $\mathcal{R} = \frac{p(r|x_i, \mathbf{x}_{-i})}{p(r|x_i', \mathbf{x}_{-i})}$ を狭い区域に抑える、すなわち、$e^{-\epsilon} \leq \mathcal{R} \leq e^{\epsilon}$。ここで、強い攻撃者であること ($\mathcal{K} = [n] \setminus \{i\}$) が仮定されている。任意の攻撃者の場合、事前知識 $\mathbf{x}_{-i}$ が $\mathbf{x}_\mathcal{K}$ になり、$\mathcal{R}$ が $\frac{p(r|x_i, \mathbf{x}_\mathcal{K})}{p(r|x_i', \mathbf{x}_\mathcal{K})}$ になる ($\mathcal{K} \subseteq [n] \setminus \{i\}$)。

表 1 に記載されるケースにおいて差分プライバシーの妥当性を考えてみよう。具体的に、相関データの場合に $\mathcal{R}$ が $[e^{-\epsilon}, e^{\epsilon}]$ に抑えられるかどうかを考察する。

3.3.1 独立データ

全てのタプルがお互いに独立であれば、強い攻撃者の場合でも、弱い攻撃者の場合でも、$\mathcal{R}$ は $[e^{-\epsilon}, e^{\epsilon}]$ に抑えられる。

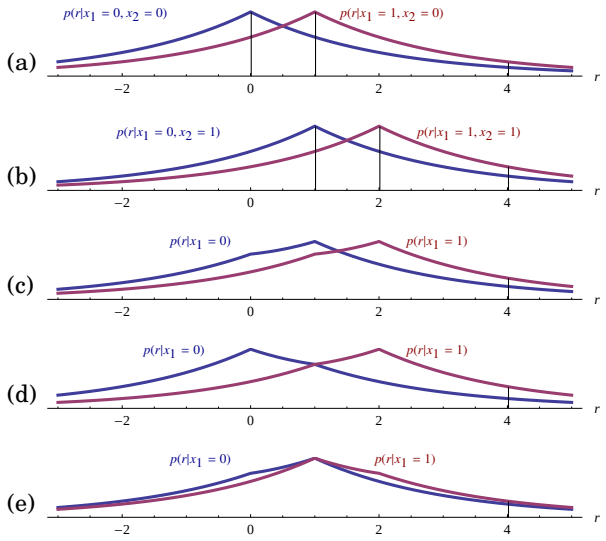


図 1: 相関データにおけるラプラシアンノイズ.

例えば、データベース $\mathbf{x} = \{x_1, x_2\}(x_1, x_2 \in \{0,1\})$ が確率分布表 3(a) に従う。クエリ関数 $f(x_1, x_2) = x_1 + x_2$ に、ラプラシアンノイズ $Lap(1/\epsilon)$ を加える。

x_2 の値を知っている強い攻撃者 $\mathcal{A}(1,\{2\})$ はタプル 1 の値を攻撃する。図 1 では、クエリ結果にノイズを加えた出力の分布を表す。メカニズム $\mathcal{M}$ は差分プライバシーを満たすので、図 1(a) と (b) に示されたよう、$\mathcal{R}$ 値は抑えられる。すなわち、$e^{-\epsilon} \leq \frac{p(r|x_1=0,x_2)}{p(r|x_1=1,x_2)} \leq e^{\epsilon}$。

一方、何も知らない弱い攻撃者 $\mathcal{A}(1,\phi)$ もタプル 1 の値を攻撃する。ベイズの定理により、$p(r|x_1) = \sum_{x_2} p(r|\mathbf{x}) \Pr(x_2|x_1)$。ゆえに、$p(r|x_1)$ が図 1(a) と (b) のようになる。x_1 と x_2 が独立であるから、$\Pr(x_2|x_1) \equiv \Pr(x_2)$。$\mathcal{M}$ は差分プライバシーを満たすので、$e^{-\epsilon} \leq \frac{p(r|x_1=0)}{p(r|x_1=1)} \leq e^{\epsilon}$。

3.3.2 相関データ

タプルに相関があれば、強い攻撃者の場合でも、$\mathcal{R}$ が $[e^{-\epsilon}, e^{\epsilon}]$ に抑えられるが、弱い攻撃者の場合に、$\mathcal{R}$ がその区域に抑えられないかもしれない。

例えば、データベース $\mathbf{x} = \{x_1, x_2\}(x_1, x_2 \in \{0,1\})$ が確率分布表 3(b) に従い、x_1 と x_2 は相関があるとする。

x_2 が分かる強い攻撃者の場合、図 1(a) と (b) のように、$\mathcal{R}$ が抑えられる。

一方、何も分からない弱い攻撃者 $\mathcal{A}(1,\phi)$ に対し、$p(r|x_1) = \sum_{x_2} p(r|\mathbf{x}) \Pr(x_2|x_1)$ である。極めて高い確率で x_1 と x_2 が等しいので、図 1(d) のように、曲線 $p(r|x_1 = 0)$ と $p(r|x_1 = 1)$ がお互いに離れる。結局、メカニズム $\mathcal{M}$ が差分プライバシーを満たすのに、$\mathcal{R}$ が抑えられない。

以上の例は、攻撃者の事前知識が少なければなるほどプライバシー情報を攻撃しやすいことを示唆する。実は、この結論も正しくない。以下に、その反例を挙げる。

例えば、データベース $\mathbf{x} = \{x_1, x_2\}(x_1, x_2 \in \{0,1\})$ が確率分布表 3(c) に従うとする。ここで、弱い攻撃者 $\mathcal{A}(1,\phi)$ に対し、$p(r|x_1) = \sum_{x_2} p(r|\mathbf{x}) \Pr(x_2|x_1)$ であるが、極めて高い確率で x_1 と x_2 が異なる。図 1(e) のように、二つの曲線はほとんど重なるので、$x_1 = 0$ と $x_1 = 1$ が非常に識別しにくい。この場合、$Lap(1/\epsilon)$ は極めて高いプライバシー水準を持つ。

以上の考察から、全攻撃者に対応できる、より厳密なプライバシー定義が必要となる。

表 3: 同時分布 $(\Pr(x_1, x_2))$

	$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	0.1	0.15
$x_2 = 1$	0.3	0.45

(a) 独立

	$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	0.49	0.01
$x_2 = 1$	0.01	0.49

(b) 正の相関

	$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	0.01	0.49
$x_2 = 1$	0.49	0.01

(c) 負の相関

4 ベイジアン差分プライバシー

以上の例でデータに相関性がある時に差分プライバシーの脆弱性を示した。情報漏洩のプライバシー水準を正しく評価するため、本節で、ベイズ的な新しいプライバシー定義を提案する。

4.1 定義

差分プライバシーの本質的な目標は攻撃目標であるタプルの追加、削除及び修正によらず、公開されたクエリの確率分布がほぼ変換されないようなメカニズムを設計することである。以下で提案する定義では、一部のタプルしか分からない弱い攻撃者にも適用される。主なアイデアは、未知のタプルを周辺化することで、既知のタプルとクエリ結果に関係を与えることである。

定義 4.1 *[ベイジアン差分プライバシー]* $\mathcal{A} = \mathcal{A}(i, \mathcal{K})$ を攻撃者とし、$\mathcal{M}(\mathbf{x}) = \Pr(r \in S|\mathbf{x})$ をデータベース $\mathbf{x}$ 上のランダム摂動メカニズムとすると、$\mathcal{A}$ における $\mathcal{M}$ のベイジアン差分プライバシー漏洩は

$$BDPL_{\mathcal{A}}(\mathcal{M}) := \sup_{x_i, x_i', \mathbf{x}_{\mathcal{K}}, S} \log \frac{\Pr(r \in S|x_i, \mathbf{x}_{\mathcal{K}})}{\Pr(r \in S|x_i', \mathbf{x}_{\mathcal{K}})} \tag{2}$$

である。ただし、$\mathbf{x}$ が連続変数であれば、周辺化して、

$$\Pr(r \in S|x_i, \mathbf{x}_{\mathcal{K}}) = \int \Pr(r \in S|\mathbf{x}) p(\mathbf{x}_{\mathcal{U}}|x_i, \mathbf{x}_{\mathcal{K}}) d\mathbf{x}_{\mathcal{U}} \tag{3}$$

であり、同様に、$\mathbf{x}$ が離散変数であれば、周辺化して、

$$\Pr(r \in S|x_i, \mathbf{x}_{\mathcal{K}}) = \sum_{\mathbf{x}_{\mathcal{U}}} \Pr(r \in S|\mathbf{x}) \Pr(\mathbf{x}_{\mathcal{U}}|x_i, \mathbf{x}_{\mathcal{K}}) \tag{4}$$

である。もし、以下の条件に満たせば、

$$\sup_{\mathcal{A}} BDPL_{\mathcal{A}}(\mathcal{M}) \leq \epsilon \tag{5}$$

$\mathcal{M}$ は ϵ–ベイジアン差分プライバシー *(ϵ–BDP)* を満たすという。

定義 4.1 の (2) にある S は任意の集合だから、r が連続的である場合、(2) は下記の式と同値であり、

$$BDPL_{\mathcal{A}}(\mathcal{M}) = \sup_{x_i, x_i', \mathbf{x}_{\mathcal{K}}, t} \log \frac{p(r = t|x_i, \mathbf{x}_{\mathcal{K}})}{p(r = t|x_i', \mathbf{x}_{\mathcal{K}})} \tag{6}$$

r が離散的である場合、(2) は下記の式と同値である。

$$BDPL_{\mathcal{A}}(\mathcal{M}) = \sup_{x_i, x_i', \mathbf{x}_{\mathcal{K}}, t} \log \frac{\Pr(r = t|x_i, \mathbf{x}_{\mathcal{K}})}{\Pr(r = t|x_i', \mathbf{x}_{\mathcal{K}})} \tag{7}$$

差分プライバシーとの差は、事前知識は $\mathbf{x}_{-i}$ ではなく、$\mathbf{x}_{\mathcal{K}}$ であることである。ベイズの定理を用いることで、(3) と (4) により、$\Pr(r \in S|x_i, \mathbf{x}_{\mathcal{K}})$ は二つの部分に分けられる。$\Pr(r \in S|\mathbf{x})$ は摂動メカニズムであり、$p(\mathbf{x}_{\mathcal{U}}|x_i, \mathbf{x}_{\mathcal{K}})$ はデータの相関性である。この定義で、$\Pr(r \in S|x_i, \mathbf{x}_{\mathcal{K}})$ と $\Pr(r \in S|x_i', \mathbf{x}_{\mathcal{K}})$ の比の最大値 $BDPL$ は $\exp(\epsilon)$ で抑えられれば、この摂動メカニズムは ϵ のプライバシーレベルを持つという。すなわち、ϵ–BDP を満たす。

4.2 BDP と DP の比較

BDP は差分プライバシーの拡張であり、より厳しいプライバシー基準である。(3) により、 $BDPL_{\mathcal{A}}(\mathcal{M}) =$

$$\sup_{x_i, x_i', \mathbf{x}_{\mathcal{K}}, S} \log \frac{\int \Pr(r \in S|\mathbf{x}) p(\mathbf{x}_{\mathcal{U}}|x_i, \mathbf{x}_{\mathcal{K}}) d\mathbf{x}_{\mathcal{U}}}{\int \Pr(r \in S|\mathbf{x}') p(\mathbf{x}_{\mathcal{U}}|x_i', \mathbf{x}_{\mathcal{K}}) d\mathbf{x}_{\mathcal{U}}}. \quad (8)$$

DP との区別は未知タプル $\mathbf{x}_{\mathcal{U}}$ が周辺化されたことである。

4.2.1 独立データ

全タプルが独立の場合、$p(\mathbf{x}_{\mathcal{U}}|x_i, \mathbf{x}_{\mathcal{K}}) \equiv p(\mathbf{x}_{\mathcal{U}})$ である (表 1 の上の行)。

定理 4.1 (独立データ) 全てのタプルがお互いに独立である場合、*ϵ–BDP* は *ϵ–DP* と同値である。□

4.2.2 相関データと強い攻撃者

タプルは相関があり、かつ強い攻撃者が限定されるケースを考えよう (表 1 の左下)。

定理 4.2 (相関データと強い攻撃者) 攻撃者は攻撃目標以外の全てのタプルが分かる場合、*ϵ–BDP* は *ϵ–DP* と同値である。□

4.2.3 相関データと任意の攻撃者

以上の分析は、全てのタプルが独立であるあるいは全ての攻撃者が強い場合に、BDP は DP と全く同じであることを示唆する。ただし、一般的に、タプルに相関性があるケース及び攻撃者は一部のタプルしか知らないケースにおいて、プライバシーを考える必要がある (表 1 の右下)。例えば、データベース $\mathbf{x} = \{x_1, x_2\}$ において、x_1 と x_2 の同時分布は表 3(b) に示され、クエリを $f(\mathbf{x}) = x_1 + x_2$ を考える。$\mathcal{A}(1, \{2\})$ が x_2 を知り、x_1 の値を攻撃したい強い攻撃者であり、$\mathcal{A}(1, \phi)$ が何も知らず、x_1 の値を攻撃したい弱い攻撃者である。ラプラシアンノイズ $Lap(1/\epsilon)(\epsilon = 0.1)$ をクエリ結果 r に加える。定理 4.2 により、攻撃者 $\mathcal{A}(1, \{2\})$ のプライバシー漏洩 $BDPL_{\mathcal{A}(1,\{2\})}(\mathcal{M})$ は ϵ であり、$\mathcal{A}(1, \phi)$ のプライバシー漏洩 $BDPL_{\mathcal{A}(1,\phi)}(\mathcal{M})$ は

$$\begin{aligned} & \sup_r \left| \log \frac{\sum_{x_2} \Pr(r|x_1 = 0, x_2) \Pr(x_2|x_1 = 0)}{\sum_{x_2} \Pr(r|x_1 = 1, x_2) \Pr(x_2|x_1 = 1)} \right| \\ = & \left| \log \frac{0.98 + 0.02e^{-\epsilon}}{0.02e^{-\epsilon} + 0.98e^{-2\epsilon}} \right| \approx 0.196 \approx 2\epsilon. \end{aligned}$$

である。$BDPL_{\mathcal{A}(1,\phi)}(\mathcal{M}) > \epsilon$ となるので、ϵ-BDP を満たさない。

$x_1 = 0$ であれば、高い確率で $x_2 = 0$ である。従って、高い確率で $f = 0$ である。同様に、$x_1 = 1$ であれば、高い確率で $f = 2$ である。つまり、x_1 の変化により f に大きい変化を引き起こす。故に、$\mathcal{M}$ の出力 r により元の x_1 の値を推定するのが容易となり、そのプライバシー漏洩が大きいといえる。

以上の例で、ラプラシアンメカニズム $\mathcal{M}$ で弱い攻撃者における $\mathcal{R}$ を制限できないので、$\mathcal{R}$ を $[e^{-\epsilon}, e^{\epsilon}]$ に制限できる $\epsilon'(\geq \epsilon)$ を探す必要がある。つまり、$Lap(1/\epsilon')$ は全ての攻撃者にプライバシーを保証できる。結果として、最も弱い攻撃者は最も大きいプライバシー漏洩を持ち、メカニズムのプライバシー水準を決める。しかし、この結論はいつも正しいだろうか？次の例を考えよう。

例えば、データベース $\mathbf{x}$ の各タプルの間に負の相関があり、同時分布が表 3(c) のように示される場合を考えてみる。この時、弱い攻撃者 $\mathcal{A}(1, \phi)$ のプライバシー漏洩は

$$\begin{aligned} & BDPL_{\mathcal{A}(1,\phi)}(\mathcal{M}) \\ = & \max_r \left| \log \frac{\sum_{d_2} \Pr(r|d_1 = 0, d_2) \Pr(d_2|d_1 = 0)}{\sum_{d_2} \Pr(r|d_1 = 1, d_2) \Pr(d_2|d_1 = 1)} \right| \\ = & \left| \log \frac{0.02 + 0.98e^{-\epsilon}}{0.98e^{-\epsilon} + 0.02e^{-2\epsilon}} \right| \approx 0.004 \ll \epsilon. \end{aligned}$$

となる。$BDPL_{\mathcal{A}}(\mathcal{M}) \leq \epsilon (\forall \mathcal{A})$ となるので、ϵ-BDP を満たす。

x_1 の値にかかわらず、f の値が 1 になりやすい。つまり、x_1 の変化は f に大きい変換を引き起こさない。結果として、$BDPL_{\mathcal{A}(1,\phi)}$ は ϵ より小さい。それにもかかわらず、強い攻撃者のプライバシー漏洩は ϵ であるので、任意の攻撃者におけるプライバシー漏洩は ϵ である。

要約すると、弱い攻撃者が大きいプライバシー漏洩を持つ場合もあり、強い攻撃者が大きいプライバシー漏洩を持つ場合もある。従って、BDP を計算する際に、全ての攻撃者を考える必要がある。さらに、ϵ–DP は $BDPL_{\mathcal{A}(i,[n]\setminus\{i\})}(\mathcal{M}) \leq \epsilon$ と同値であり、ϵ–BDP は $BDPL_{\mathcal{A}(i,\mathcal{K})}(\mathcal{M}) \leq \epsilon$ と同値であり、BDP は DP より強いことが知られている。

4.3 BDP と Pufferfish プライバシーの比較

Pufferfish プライバシー [7] はベイズ的なプライバシーフレームワークである。 潜在的秘密の集合 S（potential secrets）、差別的ペアの集合 S_{pairs}（discriminative pairs）及びデータ進化シナリオの集合 D（data evolution scenarios）が与えられて、任意の出力 ω と、任意のペア $s_i, s_j \in S_{pairs}$ と、任意の $\theta \in D$ に対して、メカニズム M は、下記の条件を満たせば、ϵ-Pufferfish プライバシーという。

$$e^{-\epsilon} \leq \frac{\Pr(M(Data) = \omega|s_i, \theta)}{\Pr(M(Data) = \omega|s_j, \theta)} \leq e^{\epsilon}. \quad (9)$$

ただし、S は保護するべき目標の集合であり、S_{pairs} は異なる S のペア（$S \times S$）であり、D は、データの生成方法や、データの相関性や、攻撃者が持つ知識についての仮定である。Pufferish プライバシーは、抽象的なプライバシー定義のフレームワークである一方、S、S_{pairs}、及び D を明示的に定義しないため、実際にプライバシーの評価はできない。提案した BDP は、S をタプルの定義域として、S_{pairs} を全ての潜在的な秘密のペアとして、D をベイズ的なデータ相関性として、明示的に定義することで、任意のメカニズムのプライバシーを定量的に評価できる。

4.4 議論

直観的に、強い攻撃者はより多い背景知識を持つから、プライバシーのリスクが高いのは当然であるが、なぜ、前の例で弱い攻撃者がプライバシーのリスクが高いのであろうか？

[7] の議論と同様に、ベイジアン差分プライバシー漏洩 $BDPL_{\mathcal{A}}(\mathcal{M})$ は事前オッズ比と事後オッズ比の差で表される。

$$\sup \left(\log \frac{\Pr(x_i|r, \mathbf{x}_{\mathcal{K}})}{\Pr(x_i'|r, \mathbf{x}_{\mathcal{K}})} - \log \frac{\Pr(x_i|\mathbf{x}_{\mathcal{K}})}{\Pr(x_i'|\mathbf{x}_{\mathcal{K}})} \right).$$

強い攻撃者 $\mathcal{A}(1, \{2\})$ に対し、事前分布 $\Pr(x_1|x_2 = 1)$ は $\{0.02, 0.98\}$ であり、出力 $r = 2$ を公開すると、その事後分布 $\Pr(x_1|r = 2, x_2 = 1)$ は $\{0.018, 0.982\}$ である。ここで、強い攻撃者の事後分布にも事前分布にも、x_1 に関して高い確信を与えるが、それぞれの差が小さいから、r の公開とともに、攻撃者は x_1 に関して少ない情報しか入手できないと言える。一方、弱い攻撃者 $\mathcal{A}(1, \phi)$ の事前と事後分布はそれぞれ $\{0.5, 0.5\}$ と $\{0.45, 0.55\}$ である。ここで、弱い攻撃者の事後分布にも事前分布にも、x_1 に関して低い確信を与えるが、それぞれの差が大きいから、r の公開とともに、攻撃者は x_1 に関して多い情報を入手できると言える。事前分布と事後分布の差の変化は、弱い攻撃者のほうが強い攻撃者より大きいから、BDPL はより大きい。

実は、差分プライバシーも Pufferfish プライバシーの一種類として考えられる。要するに、下記の式に書き直すことで、差分プ

ライバシーも事前分布と事後分布の差の変化値を評価することを言える。

$$\sup\left(\log\frac{\Pr(x_i|r,\mathbf{x}_{-i})}{\Pr(x_i'|r,\mathbf{x}_{-i})}-\log\frac{\Pr(x_i|\mathbf{x}_{-i})}{\Pr(x_i'|\mathbf{x}_{-i})}\right).$$

データ相関性と未知タプルはクエリ結果に大きい変化を引き起こし、大きい情報漏洩も引き起こす。ベイジアン差分プライバシーは未知タプルと明示的なデータ相関性を用いることで、この情報漏洩を捉えることができる。

5 ガウシアン相関モデル

合計クエリはデータ開示の領域で広く使われているので、以下では、合計クエリのみを考える。さらに、カウンティングクエリやヒストグラムなども合計クエリから導出される。

5.1 モデル

任意のマルコフ確率場はある多変量の同時分布である。ベイジアン差分プライバシーを計算するため、全ての $n2^{n-1}$ 個の攻撃者を考えるのが必要であるから、この計算は一般的に困難である。我々の目的は、複雑なデータ相関性を表現でき、そのベイジアン差分プライバシー漏洩を計算できるような相関モデルを作ることである。マルコフ確率場の一種類として、ガウシアンマルコフ確率場は下記の特徴を持つ。

第一に、タプルは相関あるいは独立であり、相関の強さも任意である。このような相関は重み付けのあるネットワークで表され、相関の強さは辺の重みで表現される。ガウシアンマルコフ確率場でこのネットワークから直接に定義できる (定義 5.1)。

第二に、BDPL を評価するため、一部のタプルが与えられ、ほかのタプルの条件確率分布を計算する必要がある。ガウシアンマルコフ確率場では、全ての条件分布がガウシアンであり、本稿の主な結論を容易に引き出せる (定理 5.2)。

第三に、ガウシアン分布は非常に計算しやすくて、効率的なアルゴリズムの設計が可能になる (Algorithm1)。

データベース $\mathbf{x}$ と相関を重み付けしたグラフ $G(\mathbf{x},W)$ で表わす。ただし、各頂点 $x_i\in\mathbf{x}$ が一つのタプル i に対応し、各重み $w_{ij}\in W(w_{ij}\geq 0)$ がタプル i と j の相関を表す。ここで、ネットワークが連結であることを仮定する。つまり、任意のタプル i と j に対し、i から j までの通路が少なくとも一通りがある。

定義 5.1 (ガウシアン相関モデル) $G(\mathbf{x},W)$ を重みつけた無向グラフとする。ただし、各頂点 $x_i\in\mathbf{x}$ はデータベースにあるタプルに対応し、各辺 $w_{ij}\in W(w_{ij}\geq 0)$ はタプル i と j の相関を表す。$\mathbf{W}=(w_{ij})$ を重みつけた隣接行列とし、$\mathbf{D}=diag(w_1,w_2,\cdots,w_n)$ を $G(\mathbf{x},W)$ の次数 *(degree)* 行列とし ($w_i=\sum_{j\neq i}w_{ij}$)、$\mathbf{L}$ をラプラシアン行列 $G(\mathbf{x},W)$ とする。

$$\mathbf{L}=\mathbf{D}-\mathbf{W}=\begin{pmatrix} w_1 & -w_{12} & \cdots & -w_{1,n}\\ -w_{12} & w_2 & \cdots & -w_{2,n}\\ \vdots & \vdots & \ddots & \vdots\\ -w_{1,n} & -w_{2,n} & \cdots & w_n\end{pmatrix}. \tag{10}$$

$\forall i\in[n]$ に対し、$\mathbf{x}_{-i}=\mathbf{x}_{[n]\setminus\{i\}}$ の条件確率は下記であれば、

$$p(\mathbf{x}_{-i}|x_i)\propto\exp\left(-\frac{\mathbf{x}^{\mathrm{T}}\mathbf{L}\mathbf{x}}{2}\right). \tag{11}$$

$(\mathbf{x},\mathbf{L})$ をガウシアン相関モデルといい、$G(\mathbf{x},\mathbf{L})$ と表記する。行列 $\mathbf{L}$ を $G(\mathbf{x},\mathbf{L})$ のガウシアン相関行列という。

任意の攻撃者 $\mathcal{A}(i,\mathcal{K})$ に対し、i、$\mathcal{K}$ と $\mathcal{U}$ の値にかかわらず、$p(\mathbf{x}_{\mathcal{U}}|x_i,\mathbf{x}_{\mathcal{K}})$ が以下の形式で表せる。

$$p(\mathbf{x}_{\mathcal{U}}|\mathbf{x}_C)\propto p(\mathbf{x}_{\mathcal{U}},\mathbf{x}_{\mathcal{K}}|x_i)\propto\exp\left(-\frac{\mathbf{x}^{\mathrm{T}}\mathbf{L}\mathbf{x}}{2}\right). \tag{12}$$

特に、$\mathcal{U}$ はただ一つのタプルである場合 ($\mathcal{U}=\{u\}$)、(12) は x_u の条件分布になる ($C=[n]\setminus\{u\}$)。すなわち、

$$p(x_u|\mathbf{x}_{-u})\propto\exp\left(-\frac{w_ux_u^2+2\mathbf{x}_{-u}^{\mathrm{T}}\tilde{\mathbf{L}}^{\mathrm{T}}x_u}{2}\right) \tag{13}$$

ただし、

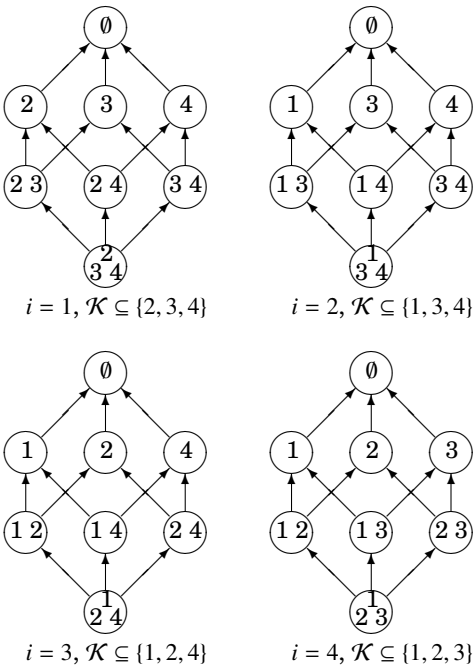
$$\mathbf{x}_{-u}=(x_1,\cdots,x_{u-1},x_{u+1},\cdots,x_n)^{\mathrm{T}}, \tag{14}$$

$$\tilde{\mathbf{L}}=(-w_{1u},\cdots,-w_{u-1,u},-w_{u,u+1},\cdots,-w_{un}). \tag{15}$$

結果として、(13) は下記のガウス分布になる。

$$p(x_u|\mathbf{x}_{-u})\propto\exp\left(-\frac{w_u}{2}(x_u-\sum_{j\neq u}\frac{w_{uj}}{w_u}x_j)^2\right). \tag{16}$$

この分布の平均 $\sum_{j\neq u}\frac{w_{uj}}{w_u}x_j$ は x_u の隣接タプルの重みつけた平均値である。大きい相関性 w_{uj} は x_j が x_u へ大きい影響があることを意味する。



$i=1,\mathcal{K}\subseteq\{2,3,4\}$　$i=2,\mathcal{K}\subseteq\{1,3,4\}$

$i=3,\mathcal{K}\subseteq\{1,2,4\}$　$i=4,\mathcal{K}\subseteq\{1,2,3\}$

図 2: $\mathcal{K}$ における階層的構造

5.2 相関性の分析

ガウシアン相関モデル上のタプルの集計クエリ $s=sum(\mathbf{x})$ に、ラプラシアンメカニズムでノイズを加え、$r=s+Lap(1/\epsilon)$ を出力する。ベイジアン差分プライバシー漏洩を計算するために、任意の $\mathcal{A}(i,\mathcal{K})$ に対し、$p(r|x_i,\mathbf{x}_{\mathcal{K}})$ を評価する必要がある。周辺化することで、以下が得られる。

$$p(r|x_i,\mathbf{x}_{\mathcal{K}})\propto\int p(r|s)p(s|x_i,\mathbf{x}_{\mathcal{K}})ds. \tag{17}$$

ここで、$p(s|x_i,\mathbf{x}_{\mathcal{K}})$ はタプルとクエリ結果の相関であり、$p(r|s)$ はラプラシアンメカニズムである。データ相関性 $p(s|x_i,\mathbf{x}_{\mathcal{K}})$ を本節で、ラプラシアンメカニズム $p(r|s)$ を次節で議論しよう。

2^{n-1} 個の異なる攻撃者 $\mathcal{A}(i,\mathcal{K})$ をそれぞれ評価するのは困難である。本稿で、高いプライバシー漏洩を持つ攻撃者の部分集合を求め、この部分集合にプライバシー漏洩が最も高い攻撃者を探す方法を提案する。まず、攻撃者のプライバシー漏洩の関係を反映できる階層的な構造を作る。

任意の攻撃者 $\mathcal{A}=\mathcal{A}(i,\mathcal{K})$ に対し、$\mathcal{K}'\subset\mathcal{K}$ であれば、$\mathcal{A}'=\mathcal{A}(i,\mathcal{K}')$ が $\mathcal{A}$ より弱いといい、$\mathcal{A}$ が $\mathcal{A}'$ より強いという。一方、

$|\mathcal{K}| - |\mathcal{K}'| = 1$ であれば、$\mathcal{A}'$ が $\mathcal{A}$ の先祖といい、$\mathcal{A}$ が $\mathcal{A}'$ の子孫という。図 2 で、四つのタプルを含むデータベースの階層的構造が示される。

一般的に、$\forall \mathcal{A}(i,\mathcal{K})$ に対し、

$$\begin{aligned} p(s|x_i, \mathbf{x}_\mathcal{K}) &= p(s_\mathcal{U} = s - s_\mathcal{K} - x_i | x_i, \mathbf{x}_\mathcal{K}) \\ &\propto p(s_\mathcal{U}, \mathbf{x}_\mathcal{K} | x_i) \end{aligned} \tag{18}$$

ただし、$s_\mathcal{U} = \sum_{i\in\mathcal{U}} x_i$。$\mathbf{x}_\mathcal{U}$ が未知だから、$s_\mathcal{U}$ が乱数と見なされる。$\forall \mathcal{A}(i,\mathcal{K})$ の先祖 $\mathcal{A}(i,\mathcal{K}')$ $(\mathcal{K}' = \mathcal{K} \setminus \{j\})$ に対し、$\mathcal{U}' = \mathcal{U} \cup \{j\}$ であり、$s_{\mathcal{U}'} = s_\mathcal{U} + x_j$ となる。そして、

$$p(s_{\mathcal{U}'}, \mathbf{x}_{\mathcal{K}'} | x_i) = p(s_\mathcal{U} + x_j, \mathbf{x}_{\mathcal{K}'} | x_i). \tag{19}$$

一方で、

$$p(s_\mathcal{U}, \mathbf{x}_\mathcal{K} | x_i) = p(s_\mathcal{U}, x_j, \mathbf{x}_{\mathcal{K}'} | x_i). \tag{20}$$

つまり、周辺化することで、$s_\mathcal{U}$ と x_j を一つの変数 $s_{\mathcal{U}'}$ に纏めると、(19) を (20) から導出できる。

命題 5.1 (組合せ) $G(\mathbf{x}, \mathbf{L})$ をあるガウシアン相関モデルとする。任意の攻撃者 $\mathcal{A} = \mathcal{A}(i,\mathcal{K})$ に対し、$C = \{i\} \cup \mathcal{K}$, $\mathcal{U} = [n] \setminus C$ とし、$m = |\mathcal{U}|$ を $\mathcal{U}$ のサイズとする。$x_0 = \frac{\sum_{u\in\mathcal{U}} x_u}{m}$ であり、$\mathbf{x}_\mathcal{A} = \{x_0, x_i\} \cup \mathbf{x}_\mathcal{K} = \{x_0\} \cup \mathbf{x}_C$ と仮定すると、$G(\mathbf{x}_\mathcal{A}, \mathbf{L}_\mathcal{A})$ もガウシアン相関モデルになる。更に、攻撃者 $\mathcal{A}' = \mathcal{A}(i,\mathcal{K}')$ を $\mathcal{A}$ の先祖とする。すなわち、$\mathcal{K}' = \mathcal{K} \setminus \{j\}$ $(\exists j \in \mathcal{K})$ である。$\mathcal{U}' = \mathcal{U} \cup \{j\}$ とし、$C' = \mathcal{K}' \cup \{i\}$ とする。もし、$\mathcal{A}$ のガウシアン相関行列が下記であれば、

$$\mathbf{L}_\mathcal{A} = \begin{pmatrix} w_0 & -\mathbf{w}_{0C}^\mathrm{T} \\ -\mathbf{w}_{0C} & \mathbf{W}_C \end{pmatrix} = \begin{pmatrix} w_0 & -w_{0j} & -\mathbf{w}_{0C'}^\mathrm{T} \\ -w_{0j} & w_j & -\mathbf{w}_{jC'}^\mathrm{T} \\ -\mathbf{w}_{0C'} & -\mathbf{w}_{jC'} & \mathbf{W}_{C'} \end{pmatrix}, \tag{21}$$

$\mathcal{A}'$ のガウシアン相関行列が下記のようになる。

$$\mathbf{L}_{\mathcal{A}'} = \begin{pmatrix} w'_0 & -\mathbf{w}'^\mathrm{T}_{0C'} \\ -\mathbf{w}'_{0C'} & \mathbf{W}'_{C'} \end{pmatrix}, \tag{22}$$

ただし、

$$\begin{aligned} w'_0 &= \frac{(m+1)^2(w_0 w_j - w_{0j}^2)}{m^2 w_j + 2m w_{0j} + w_0} \\ \mathbf{w}'_{0C'} &= (m+1)\frac{(w_{0j} + m w_j)\mathbf{w}_{0C'} + (w_0 + m w_{0j})\mathbf{w}_{jC'}}{m^2 w_j + 2m w_{0j} + w_0} \\ \mathbf{W}'_{C'} &= \mathbf{W}_{C'} - \frac{(\mathbf{w}_{0C'} - m\mathbf{w}_{jC'})(\mathbf{w}_{0C'}^\mathrm{T} - m\mathbf{w}_{jC'}^\mathrm{T})}{m^2 w_j + 2m w_{0j} + w_0}. \end{aligned}$$

□

5.3 出力摂動

本節で、(17) にある摂動メカニズム $p(r|s)$ をパラメータ λ と 0 を持つラプラシアンメカニズムに限定し、ベイジアン差分プライバシーを満たすように、λ を適当に選択する。すなわち、

$$z \sim Lap(\lambda) = \frac{1}{2\lambda} e^{-\frac{|z|}{\lambda}}. \tag{23}$$

このノイズをクエリ結果 s に加えて、その結果 $r = s + z$ を公開する。大きい λ は低いユーティリティを引き起こすので、以下で、ベイジアン差分プライバシーを満たすような最も小さい λ を見つける。

定理 5.1 で、摂動メカニズム $p(r|s)$ をデータ相関 $p(s|x_i, \mathbf{x}_\mathcal{K})$ に結びつけることで、ベイジアン差分プライバシー漏洩 BDPL の計算方法が与えられる。

定理 5.1 $G(\mathbf{x}, \mathbf{L})$ をガウシアン相関モデルとし、$s = \sum_{i\in[n]} x_i$ を $\mathbf{x}$ 上の集計クエリとする。ラプラシアンメカニズム $\mathcal{M}$ で、摂動出力 $r = \mathcal{M}(\mathbf{x}) = s + z$ を生成する $(z \sim Lap(\lambda))$。攻撃者 $\mathcal{A} = \mathcal{A}(i,\mathcal{K})$ に対し、$x_0 = \frac{\sum_{u\in\mathcal{U}} x_u}{|\mathcal{U}|}$ を未知タプルの平均値とし、$\mathbf{x}_\mathcal{A} = \begin{pmatrix} x_0 \\ \mathbf{x}_C \end{pmatrix} = \begin{pmatrix} x_0 \\ x_i \\ \mathbf{x}_\mathcal{K} \end{pmatrix}$ とし、対応するガウシアン相関行列は下記の式とする。

$$\mathbf{L}_\mathcal{A} = \begin{pmatrix} w_0 & -w_{0i} & -\mathbf{w}_{0\mathcal{K}}^\mathrm{T} \\ -w_{0i} & w_i & -\mathbf{w}_{i\mathcal{K}}^\mathrm{T} \\ -\mathbf{w}_{0\mathcal{K}} & -\mathbf{w}_{i\mathcal{K}'} & \mathbf{W}_\mathcal{K} \end{pmatrix}. \tag{24}$$

$M > 0$ と $m = |\mathcal{U}|$ を仮定すると、$\mathcal{A}$ についての $\mathcal{M}$ のベイジアン差分プライバシー漏洩は下記となる。

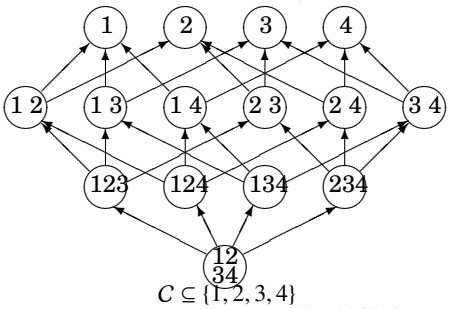
$$BDPL_\mathcal{A}(\mathcal{M}; M) = \frac{M}{\lambda}(1 + \ell_\mathcal{A}). \tag{25}$$

ただし、$\ell_\mathcal{A} = \frac{m w_{0i}}{w_0}$ を $\mathcal{A}$ に関する相関モデルの漏洩係数という。□

定義 4.1 により、$BDPL \le \epsilon$ が成立つと、ϵ-BDP を満たす。すなわち、

$$M(1 + \ell_\mathcal{A}) \le \epsilon\lambda. \tag{26}$$

データの相関性が与えられた時、左辺は定数である。右辺の ϵ はプライバシー指標であり、ϵ が小さければ、プライバシーは高い。λ はユーティリティ指標であり、λ が小さければ、ノイズの分散が小さいから、ユーティリティが高い。結局、式 (26) はプライバシー（ϵ）とユーティリティ（λ）のトレードオフ関係になる。与えられた ϵ に対して、$\ell_\mathcal{A}$ が小さければ、高いユーティリティ（小さい λ）のメカニズムを適用できる。



$C \subseteq \{1, 2, 3, 4\}$

図 3: C における階層的構造.

5.4 メカニズム

本節で、全ての攻撃者から、ベイジアン差分プライバシー漏洩が最も大きい攻撃者を見つける。ゆえに、全ての攻撃者に対し、ϵ-BDP を満たすようなラプラシアンメカニズムを作れる。

定理 5.2 $G(\mathbf{x}, \mathbf{L})$ はガウシアン相関モデルであり、$s = \sum_{i\in[n]} x_i$ は $\mathbf{x}$ 上の集計クエリである。ラプラシアンメカニズム $\mathcal{M}$ で、$r = \mathcal{M}(\mathbf{x}) = s + z$ $(z \sim Lap(\lambda))$ を生成する。任意の攻撃者 $\mathcal{A} = \mathcal{A}(i,\mathcal{K})$ とその先祖 $\mathcal{A}' = \mathcal{A}(i,\mathcal{K}')$ に対し、次の不等式が成り立つ。

$$\ell_\mathcal{A} \le \ell_{\mathcal{A}'}. \tag{27}$$

ゆえに、$\forall M > 0$ について、次の式は自明である。

$$BDPL_\mathcal{A}(\mathcal{M}; M) \le BDPL_{\mathcal{A}'}(\mathcal{M}; M). \tag{28}$$

□

定理 5.2 は、与えられた $Lap(\lambda)$ に対し、ある攻撃者の BDPL はその先祖の BDPL より小さいことを示唆する。ゆえに、最も弱い攻撃者は最も大きいプライバシー漏洩を持つ。逆に、全ての攻撃者にたいして ϵ-BDP を満たすために、ラプラシアンメカニズ

ムの分散が $\lambda \geq (1+\ell_{\mathcal{A}})\frac{M}{\epsilon}$ $(\forall \mathcal{A})$ を満たさなければならないので、λ は少なくても $\lambda_0 = \frac{M}{\epsilon}(1+\max_{\mathcal{A}} \ell_{\mathcal{A}})$ である。結局、最も大きい BDPL を持つ攻撃者を探せば良い。最も大きい BDPL を持つ攻撃者は必ず n 個の弱い攻撃者 $\mathcal{A}(i,\phi)$ のうちにいるから、命題 5.1 を用い、図 2 にあるパスを経て、各弱い攻撃者に対するガウシアン相関モデルを獲得できる。一方、命題 5.1 は $\mathcal{A}(i,\mathcal{K})$ に関するガウシアン相関モデルは $\mathcal{C} = \{i\} \cup \mathcal{K}$ にしか関係ないことを示唆する。例えば、$\mathcal{A}(i,\mathcal{K})$ と $\mathcal{A}(i', \mathcal{K} \cup \{i\} \setminus \{i'\})$ は異なる攻撃者であるが、それぞれの相関行列は完全に同じである。$\mathcal{C}$ の包含関係に基づき、図 3 のような階層的構造を作れる。結局、一番高い階層にある n 個の弱い攻撃者に対するプライバシー漏洩を評価し、そのうちから最も大きいのを選べば良い。Algorithm 1 で、各弱い攻撃者 $\mathcal{A}(i,\phi)$ $(\forall i \in [n])$ に対する漏洩係数を計算できる。

Algorithm 1 プライバシー漏洩係数を計算する

input: $\mathbf{L}$ - ガウシアン相関行列
input: m - $|\mathcal{U}|$ ($\mathbf{x}_{\mathcal{U}}$ にあるタプルの数)
output: ℓ - 漏洩係数

function CalcLeakageCoefficient($\mathbf{L}, m = 1$)
  $n \leftarrow$ SizeOf($\mathbf{L}$)
  if $n \leq 2$ **then**
    Return $\frac{w_0}{w_{01}}$
  end if
  $k \leftarrow \lceil \frac{n+1}{2} \rceil$
  $\mathbf{L}_1 \leftarrow \mathbf{L}$
  for $i = k \rightarrow 2$ **step** -1 **do**
    $\mathbf{L}_1 \leftarrow$ Combine($\mathbf{L}_1, i, m$) ▷ (命題 5.1)
    $m \leftarrow m + 1$
  end for
  $\ell_1 \leftarrow$ CalcLeakageCoefficient($\mathbf{L}_1, m$)
  $\mathbf{L}_2 \leftarrow \mathbf{L}$
  for $i = n \rightarrow k+1$ **step** -1 **do**
    $\mathbf{L}_2 \leftarrow$ Combine($\mathbf{L}_2, i, m$) ▷ (命題 5.1)
    $m \leftarrow m + 1$
  end for
  $\ell_2 \leftarrow$ CalcLeakageCoefficient($\mathbf{L}_2, m$)
  Return $\max(\ell_1, \ell_2)$
end function

定理 5.3 *Algorithm 1* の時間及び空間計算量はそれぞれ $O(n^3)$ と $O(n^2)$ である。□

5.5 事前分布を持つガウシアン相関モデル

弱い攻撃者 $\mathcal{A}(i,\phi)$ は大きいプライバシー漏洩を持つので、対応するガウシアン相関行列は 2×2 行列 $\begin{pmatrix} w_0 & -w_{0i} \\ -w_{0i} & w_0 \end{pmatrix}$ となる。ガウシアン相関行列の定義及び定理 5.1 により $(m = n-1)$、

$$BDPL_{\mathcal{A}(i,\phi)}(\mathcal{M}; M) \equiv \frac{nM}{\lambda}. \tag{29}$$

ゆえに、ベイジアン差分プライバシーを満たすために、n 倍の分散のラプラシアンノイズを加える必要がある。この結果は平凡であり、非常に低いユーティリティになってしまう。ガウシアン相関モデルに事前分布を追加し、全てのタプルの同時分布を定義することにより、攻撃者の仮定が弱くなる一方、ユーティリティを高めることができる。具体的に、ガウシアン事前分布 $\tau\mathbf{I}$ をガウシアン相関行列に $\mathbf{L}$ に加えると、τ が大きければ大きいほど、ユーティリティが高い。詳しくは、[11] を参考されたい。

5.6 正と負の相関

定義 5.1 にて、タプルは正の相関を持つから、あるタプルの値が大きくなると、ほかの未知タプルの値も大きくなる傾向があるので、クエリ結果はより大きく増加する。さらに、大量な未知タプルはクエリ結果の大きい増加を引き起こすので、弱い攻撃者が大きいセンシティブを持つ。ゆえに、大きいプライバシー漏洩を持つ。ただし、実際のデータがいつも正の相関性を持つとは限らない。例えば、「二つのタプルの足し算は 100 である」という相関性は「一つのタプルが大きくなると、別のタプルは小さくなる」を意味する。こんな場合、弱い攻撃者が大きいプライバシー漏洩を持つとは言えない。

6 実験

本節では、実験結果を示す。

6.1 人工データ

以下に、与えられたタプル数 n に対し、ランダムにネットワーク ($n\times n$ の隣接行列 $\mathbf{W}$) を生成する。まずは、$\mathbf{W}$ の各項目 w_{ij} を 0 に初期化する。次に、与えられたネットワーク度数 d に対し、$\mathbf{W}$ の上三角の部分の $\frac{n(n-1)}{2}$ 項目からランダムに $\frac{nd}{2}$ 個を選んでおく。さらに、$\mathbf{W}$ は必ず対称行列 ($w_{ij} \equiv w_{ji}$) であるので、選ばれた w_{ij} とその対称項目 w_{ji} を 1 に設定する。ゆえに、このネットワークの平均度数が d となる。異なるタプル数 n と度数 d に対し、このようなネットワークを生成し、それぞれのラプラシアン行列 $\mathbf{L} = \mathbf{D} - \mathbf{W}$ を計算し、最後に、事前分布を加える。具体的には、与えられた τ に対し、行列 $\mathbf{T} = \tau \cdot \mathbf{I}$ を $\mathbf{L}$ に加える。すなわち、$\hat{\mathbf{L}} = \mathbf{L} + \tau \cdot \mathbf{I}$ である。

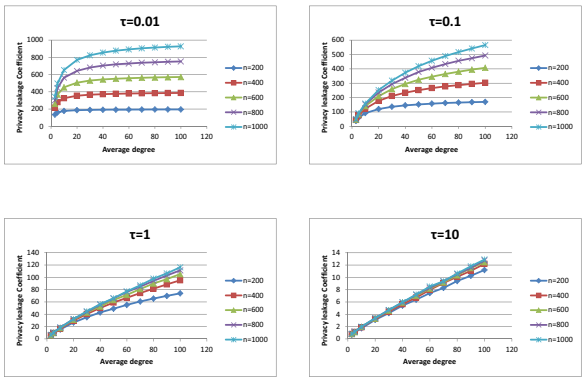


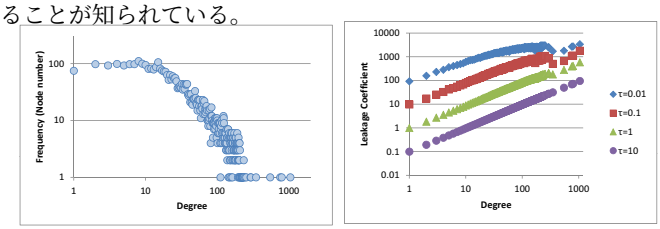
図 4: 実験結果（人工データ）.

ベイジアン差分プライバシー漏洩は漏洩係数より一意に決められるので、Algorithm 1 を用い、異なる係数 (n、d と τ) に対応するネットワークの漏洩係数のみを計算すればよい。漏洩係数は図 4 で示される。水平軸はネットワークの平均度数であり、垂直軸はネットワークの漏洩係数の最大値 $\max_{\mathcal{A}} \ell_{\mathcal{A}}$ である。n と τ はそれぞれタプル数と事前分布を表す。ネットワークの度数は大きければ大きいほど、そのプライバシー漏洩は大きい。なぜなら、大きい度数はタプル間の強い相関を意味するからである。図 4 により、τ が極めて小さければ、漏洩係数は n に近づき、ラプラシアンノイズの分散はほぼ差分プライバシーの n 倍になり、出力のユーティリティは非常に低い。一方で、τ が大きくなれば、漏洩係数が小さくなり、ユーティリティは大きくなる。

6.2 ソーシャルネットワークデータ

Stanford Large Network Dataset Collection [8] は、Facebook の参加者を見回し、4039 個の頂点と 88234 個の辺を集計し、作成したネットワークデータである。図 5(a) で、度数の分布を示す。度数の最大値は 1045 である。

タプル i と j が隣接していれば、隣接行列 $\mathbf{W}$ の項目 w_{ij} 及び w_{ji} を 1 に設定し、隣接していなければ、0 に設定する。さらに、事前分布をこの行列に加える。すなわち、$\hat{\mathbf{L}} = \mathbf{D} - \mathbf{W} + \tau \cdot \mathbf{I}$ である。図 5(b) では、異なるタプルの度数と事前分布に対し、評価された漏洩係数が示された。具体的には、各 τ に対し、タプルの度数と漏洩係数 ℓ を図 5(b) で比較する。この計算結果をみると、最大の度数 (1045) に対応するタプルは最大の漏洩係数を持つ。なぜなら、このタプルはほかのタプルと最大の相関性を持つからである。図 5(b) により、事前分布の増加とともに、漏洩係数が減少することが知られている。



(a). 度数分布. (b). 漏洩係数 vs. 度数.

図 5: 実験結果（ソーシャルネットワーク）.

7 おわりに

本稿では、データが相関性を持つ場合、任意の攻撃者に対する差分プライバシーを考えた。まず、データに相関があり、かつ攻撃者が一部の知識しか知らない場合、差分プライバシーを満たすメカニズムはプライベートではない可能性がある問題に対し、ベイジアン差分プライバシーという新しい Pufferfish プライバシー定義を提案した。さらに、ガウシアン相関モデルを提案することで、複雑な相関関係を持つデータを表現できる。最後に、ベイジアン差分プライバシーの計算は一般的に困難であるという問題に対し、ベイジアン差分プライベートなメカニズムを設計できる効率的なアルゴリズムを提案した。

[文献]

[1] C. Dwork, F. Mcsherry, K. Nissim, and A. Smith. “Calibrating noise to sensitivity in private data analysis.” In TCC, pages 265-284. Springer, 2006.

[2] J. Gehrke, M. Hay, E. Lui, and R. Pass. “Crowd-blending privacy.” In ECRYPTO, pages 479-496. 2012.

[3] J. Gehrke, E. Lui, and R. Pass. “Towards privacy for social networks: A zero-knowledge based definition of privacy.” In TCC, pages 432-449, 2011.

[4] M. Hay, C. Li, G. Miklau, and D. Jensen. “Accurate estimation of the degree distribution of private networks.” In ICDM, pages 169-178, 2009.

[5] X. He, A. Machanavajjhala, and B. Ding. “Blowfish privacy: Tuning privacy-utility trade-offs using policies.” In SIGMOD, pages 1447-1458, 2014.

[6] D. Kifer and A. Machanavajjhala. “No free lunch in data privacy.” In SIGMOD, pages 193-204, 2011.

[7] D. Kifer and A. Machanavajjhala. “Pufferfish: A framework for mathematical privacy definitions.” ACM Trans. Database Syst., 39(1):3:1-3:36, Jan. 2014.

[8] J. Leskovec and A. Krevl. “SNAP Datasets: Stanford large network dataset collection.” http://snap.stanford.edu/data.

[9] G. Miklau and D. Suciu. “A formal analysis of information disclosure in data exchange.” In SIGMOD, pages 575-586, 2004.

[10] V. Rastogi, M. Hay, G. Miklau, and D. Suciu. “Relationship privacy: output perturbation for queries with joins.” In PODS, pages 107-116, 2009.

[11] B. Yang, I, Sato, and H. Nakagawa. “Bayesian Differential Privacy on Correlated Data” In SIGMOD, pages 747-762, 2015.

[12] J. Zhang, G. Cormode, C. M. Procopiuc, D. Srivastava, and X. Xiao. “Privbayes: Private data release via bayesian networks.” In SIGMOD, pages 1423-1434, 2014.

楊 斌 Bin YANG

1998 年北京大学情報学科卒業．2010 年東京大学情報理工学系研究科修士課程終了．2013 年東京大学情報理工学系研究科博士課程終了．2013 年楽天技術研究所入社．機械学習、データサイエンス、プライバシーの研究に従事.

佐藤 一誠 Issei SATO

2011 年東京大学大学院情報理工学系研究科博士課程修了．2011 年より東京大学情報基盤センター助教．2013 年より科学技術振興機構 さきがけ研究員を兼務．2015 年 9 月より東京大学大学院新領域創成科学研究科講師．統計的機械学習およびデータマイニングの研究に従事.

中川 裕志 Hiroshi NAKAGAWA

1975 年東京大学工学部卒業。1980 年東京大学大学院工学系研究科博士課程修了。工学博士。1980 年から 1999 年、横浜国立大学。1999 年より現在まで東京大学情報基盤センター教授。自然言語処理、人工知能、機械学習、プライバシー保護の研究に従事。