古典量子差分プライバシーの数学的側面

名古屋大学大学院 多元数理科学研究科 多元数理科学専攻
吉田 裕哉 (Yuuya YOSHIDA)

概要

情報公開におけるプライバシー保護の指標として有名なものに, 差分プライバシーがある. 通常, 差分プライバシーは古典確率論を用いて研究されるため, その量子拡張はほとんど研究されていない. また, 応用面に焦点を当てた研究が多く, 数学的側面に焦点を当てた研究は少ない. 本稿では, 量子拡張の 1 つである古典量子差分プライバシーの数学的側面について, 古典と量子の比較を中心に述べる.

1 導入

データ解析において, 個人情報 X (たとえば, 離婚の有無) を保護しつつ X に関する統計 (たとえば, 離婚率) を知りたいことがある. 個人情報を保護する方法の 1 つに, **ランダム応答**というものがある [10]. ランダム応答において, データ提供者は X の代わりに別のデータ Y を答えるが, この際, Y は条件付き確率分布 $\mathbb{P}_{Y|X}$ に従う. ランダム応答を用いると, ランダム化された Y から元の X を復元できないため, X は保護される. 一方, Y は X の値に依存した確率分布に従うため, 集まった Y から X に関する統計をある程度推定できる.

ランダム応答においてプライバシー保護をより強固にするため, 条件付き確率分布 $\mathbb{P}_{Y|X}$ に対して次の条件を課す:

$$\forall x, x',\ \mathbb{P}_{Y|X}(\cdot|x') \leq e^{\varepsilon}\mathbb{P}_{Y|X}(\cdot|x). \tag{1}$$

ただし, $\varepsilon > 0$ は定数で, 不等式は要素ごとに成り立つことを意味する. この条件は ε **差分プライバシー (ε-DP)** と呼ばれる [2–4]. パラメータ ε が小さいと保護が強く, ε が大きいと保護が弱い. 通常, 差分プライバシーは古典確率論を用いて研究されるため, その量子拡張はほとんど研究されていない. 本稿では, 量子拡張の 1 つである古典量子差分プライバシーを定義し, その数学的側面について, 古典と量子の比較を中心に述べる. より詳しい内容については, full ver. [1] を参照してほしい.

2　古典量子差分プライバシー

以下, データ Y の代わりに量子状態 (数学的には密度行列, つまり, トレース 1 の半正定値行列) を用いる. つまり, データ提供者は $X = x$ の代わりに量子状態 ρ_x を答える (**古典量子の設定**). データ解析者は, 集まった ρ_x から X に関する統計を推定する. さて, 量子状態に関する情報を得るためには, 量子状態を測定しなければならない. 測定すると, (量子状態と測定から定まる) 確率分布に従い, 測定値が得られる. プライバシー保護をより強固にするため, この測定値の確率分布に ε-DP を課す:

$$\forall (M_y)_{y=1}^m \text{ POVM, the c.p.d. } \mathbb{P}(y|x) = \operatorname{Tr} \rho_x M_y \text{ satisfies } \varepsilon\text{-DP}, \tag{2}$$

ただし, "POVM" は "positive-operator-valued measure" の略で測定を表し, "c.p.d." は "conditional probability distribution" の略である. ここでは POVM の定義は重要でないため, 省略する. 条件 (2) は**古典量子 ε 差分プライバシー (CQ ε-DP)** と呼ばれ [11], 次の条件と同値である:

$$\forall x, x',\ \rho_{x'} \le e^{\varepsilon} \rho_x.$$

ただし, 不等式は両辺の差が半正定値であることを意味する. ここで, 古典的な ε-DP の定義 (1) を思い出す. 確率分布 $\mathbb{P}_{Y|X}(\cdot|x)$ を確率ベクトル p_x に置き換えると, (1) は次のように書き換えられる:

$$\forall x, x',\ p_{x'} \le e^{\varepsilon} p_x.$$

そのため, CQ ε-DP は古典的な ε-DP の単純な拡張であることがわかる.

これまでに出てきた定義をまとめる.

Definition 1 (Classical ε-DP [2] and CQ ε-DP [11])**.** $\varepsilon > 0$ を実数, $n \ge 2$ を整数とする. 任意の $i, j = 1, \dots, n$ に対して $p_i \le e^{\varepsilon} p_j$ が成り立つとき, 確率ベクトルの組 $(p_i)_{i=1}^n$ は **ε-DP** であるという. 任意の $i, j = 1, \dots, n$ に対して $\rho_i \le e^{\varepsilon} \rho_j$ が成り立つとき, 密度行列の組 $(\rho_i)_{i=1}^n$ は **CQ ε-DP** であるという. また, 集合 $\mathrm{C}_n^{(d)}(\varepsilon)$, $\mathrm{CQ}_n^{(d)}(\varepsilon)$, $\mathrm{C}_n(\varepsilon)$, $\mathrm{CQ}_n(\varepsilon)$ を次のように定める:

$$\mathrm{C}_n^{(d)}(\varepsilon) = \{\varepsilon\text{-DP } (p_i)_{i=1}^n : \text{all } p_i \text{ are probability vectors in } \mathbb{R}^d\} \quad (d \ge 2),$$
$$\mathrm{CQ}_n^{(d)}(\varepsilon) = \{\text{CQ } \varepsilon\text{-DP } (\rho_i)_{i=1}^n : \text{all } \rho_i \text{ are density matrices on } \mathbb{C}^d\} \quad (d \ge 2),$$
$$\mathrm{C}_n(\varepsilon) = \bigcup_{d \ge 2} \mathrm{C}_n^{(d)}(\varepsilon), \quad \mathrm{CQ}_n(\varepsilon) = \bigcup_{d \ge 2} \mathrm{CQ}_n^{(d)}(\varepsilon).$$

古典状態 (確率ベクトル) $p \in \mathbb{R}^d$ を量子状態 (密度行列) として扱いたいときは, p の代わりに対角成分 $p(1), \dots, p(d)$ を持つ対角行列 $\operatorname{diag}(p)$ を用いる. そのため, $\mathrm{CQ}_n(\varepsilon)$ の部分集合

$$\operatorname{diag}(\mathrm{C}_n(\varepsilon)) \coloneqq \{(\operatorname{diag}(p_i))_{i=1}^n : (p_i)_{i=1}^n \in \mathrm{C}_n(\varepsilon)\}$$

は $\mathrm{C}_n(\varepsilon)$ に対応する集合である.

3　本質的に古典的な元

実は, $\mathrm{diag}(\mathrm{C}_n(\varepsilon))$ より大きい “本質的に古典的な” 集合を定義できる. 説明のため, 次の最適化問題を考える:

$$S_n^{\mathrm{C}}(\varepsilon;\Phi) = \underbrace{\sup_{(p_i)_{i=1}^n \in \mathrm{C}_n(\varepsilon)}}_{\text{Privacy protection}} \underbrace{\Phi(\mathrm{diag}(p_1),\dots,\mathrm{diag}(p_n))}_{\text{Utility}}.$$

これは, “X を保護しつつ X に関する統計の推定精度を最大にしたい” という自然な動機から生じる問題であり, 情報理論的な DP の研究でしばしば扱われる [5–9, 11]. 量子の場合は

$$S_n^{\mathrm{CQ}}(\varepsilon;\Phi) = \underbrace{\sup_{(\rho_i)_{i=1}^n \in \mathrm{CQ}_n(\varepsilon)}}_{\text{Privacy protection}} \underbrace{\Phi(\rho_1,\dots,\rho_n)}_{\text{Utility}}$$

である. 上の Φ は密度行列 n 個の実数値関数であり, 個人情報の有用性を表す. 一方, 条件 $(p_i)_{i=1}^n \in \mathrm{C}_n(\varepsilon)$ や $(\rho_i)_{i=1}^n \in \mathrm{CQ}_n(\varepsilon)$ はプライバシー保護条件である.

次に, 目的関数 Φ に対して以下を仮定する (通常は成り立つ). その際, CPTP 写像という用語が出てくるが, ここでは “密度行列を密度行列に写す線型写像” という認識で十分である.

Definition 2 (Monotonicity for CPTP maps)**.** 任意の密度行列 $\rho_1,\dots,\rho_n$ と CPTP 写像 Λ に対して,

$$\Phi(\Lambda(\rho_1),\dots,\Lambda(\rho_n)) \le \Phi(\rho_1,\dots,\rho_n)$$

が成り立つとき, Φ は CPTP 写像に対して単調であるという. この不等式を**データ処理不等式**という.

CPTP 写像に関する単調性から, 不等式

$$\sup_{\substack{(p_i)_{i=1}^n \in \mathrm{C}_n(\varepsilon) \\ \Lambda \text{ CPTP map}}} \Phi\big(\Lambda(\mathrm{diag}(p_1)),\dots,\Lambda(\mathrm{diag}(p_n))\big) \le S_n^{\mathrm{C}}(\varepsilon;\Phi)$$

が従う. 恒等写像は CPTP 写像なので, 反対向きの不等式も成り立つ. こうして, 次の定義に到達する.

Definition 3 (Essentially classical element)**.** $\varepsilon > 0$ を実数, $n \ge 2$ を整数とする. 任意の $i = 1,\dots,n$ に対して $\Lambda(\mathrm{diag}(p_i)) = \rho_i$ となるような $(p_i)_{i=1}^n \in \mathrm{C}_n(\varepsilon)$ と CPTP 写像 Λ が存在するとき, $(\rho_i)_{i=1}^n \in \mathrm{CQ}_n(\varepsilon)$ は**本質的に古典的**であるという. 集合 $\mathrm{CQ}_n(\varepsilon)$ の本質的に古典的な元全体の集合を $\mathrm{EC}_n(\varepsilon)$ で表す.

さて, $S_n^{\mathrm{EC}}(\varepsilon;\Phi)$ を $S_n^{\mathrm{CQ}}(\varepsilon;\Phi)$ と同様に定める. 集合 $\mathrm{EC}_n(\varepsilon)$ は $\mathrm{diag}(\mathrm{C}_n(\varepsilon))$ より大きいが,

$$S_n^{\mathrm{EC}}(\varepsilon;\Phi) = S_n^{\mathrm{C}}(\varepsilon;\Phi)$$

が成り立つ. 我々は $S_n^{\mathrm{C}}(\varepsilon;\Phi)$ と $S_n^{\mathrm{CQ}}(\varepsilon;\Phi)$ の比較に興味があるが, それは $S_n^{\mathrm{EC}}(\varepsilon;\Phi)$ と $S_n^{\mathrm{CQ}}(\varepsilon;\Phi)$ の比較と同じである. では, そもそも $\mathrm{EC}_n(\varepsilon)$ と $\mathrm{CQ}_n(\varepsilon)$ は一致するのだろうか? これに関して, 次の命題が知られている [11, Theorem 1].

Proposition 1. 任意の $\varepsilon > 0$ に対して, $\mathrm{EC}_2(\varepsilon) = \mathrm{CQ}_2(\varepsilon)$.

一方, 後で述べるように, 任意の $n \geq 3$ に対して $\mathrm{EC}_n(\varepsilon) \neq \mathrm{CQ}_n(\varepsilon)$ である (Corollary 2). そのため, 次の定義を用いて $\mathrm{EC}_n(\varepsilon)$ と $\mathrm{CQ}_n(\varepsilon)$ の違いを評価する.

Definition 4. 実数 $\varepsilon > 0$ と整数 $n \geq 2$ に対して, 集合 $\mathcal{E}_n(\varepsilon)$ を

$$\mathcal{E}_n(\varepsilon) = \{\varepsilon' > 0 : \mathrm{CQ}_n(\varepsilon) \subset \mathrm{EC}_n(\varepsilon')\}$$

と定める.

実は, $\mathcal{E}_n(\varepsilon)$ は空でない (Theorem 1). しかも, 集合 $\mathrm{EC}_n(\varepsilon)$ は $\varepsilon > 0$ に関して単調増加なので, $\mathcal{E}_n(\varepsilon)$ は上に有界でない区間, つまり, $[\varepsilon_{\inf}, \infty)$ または $(\varepsilon_{\inf}, \infty)$ の形である. 簡単にわかるように, $\varepsilon_{\inf} \geq \varepsilon$ である. また, $\varepsilon_{\inf}/\varepsilon$ が大きければ $\mathrm{EC}_n(\varepsilon)$ と $\mathrm{CQ}_n(\varepsilon)$ の差も大きく, $\varepsilon_{\inf}/\varepsilon$ が小さければ $\mathrm{EC}_n(\varepsilon)$ と $\mathrm{CQ}_n(\varepsilon)$ の差も小さい. このように, $\mathrm{EC}_n(\varepsilon)$ と $\mathrm{CQ}_n(\varepsilon)$ の違いは $\mathcal{E}_n(\varepsilon)$ の下限で表せるため, これを評価する.

4 主結果

Theorem 1. 任意の $\varepsilon > 0$ と $n \geq 2$ に対して, $e^{\varepsilon'} - 1 = (n-1)(e^{\varepsilon} - 1)$ となるような $\varepsilon' \in \mathcal{E}_n(\varepsilon)$ が存在する.

Theorem 2. 任意の $n \geq 2$, $\varepsilon > 0$, $\varepsilon' \in \mathcal{E}_n(\varepsilon)$ に対して,

$$\frac{e^{\varepsilon'} - 1}{e^{\varepsilon} - 1} \geq F_n(e^{\varepsilon} - 1)$$

である. ただし, F_n は次のように定義される:

$$g_n(t) = \begin{cases} \dfrac{2}{t-1}(\sqrt{(n-1)(n-t)} + n - t) & 1 < t \le n, \\ \infty & t = 1, \end{cases}$$

$$a_{n,t}(x) = t\Big(\frac{n-t}{n-1}(x+2)^2 - x^2\Big) \quad (1 \le t \le n,\ x \ge 0),$$

$$f_{n,k}(x) = \frac{(n+2k)x + \sqrt{(n+2k)^2x^2 + 8na_{n,k}(x)}}{2a_{n,k}(x)} \quad (1 \le k \le n/2,\ 0 \le x < g_n(k)),$$

$$F_n(x) = \min\{f_{n,k}(x) : 1 \le k \le n/2 \text{ with } x < g_n(k)\} \quad (x \ge 0).$$

関数 g_n, $a_{n,t}$, $f_{n,k}$, F_n の性質に関しては, full ver. [1] の Lemmas 4.1 and 4.2 を参照.

Theorem 1 で $n = 2$ とすると, Proposition 1 が従う. しかも, Theorems 1 and 2 により, 下限 $\varepsilon_{\inf} = \varepsilon_{\inf}(n, \varepsilon) = \inf \mathcal{E}_n(\varepsilon)$ は不等式

$$F_n(e^{\varepsilon} - 1) \le \frac{e^{\varepsilon_{\inf}} - 1}{e^{\varepsilon} - 1} \le n - 1 \tag{3}$$

を満たす. 関数 F_n は次の性質を持つ [1, Lemma 4.2]:

- 任意の $x \ge 0$ に対して, $F_2(x) = 1$;
- 任意の $n \ge 3$ に対して, F_n は狭義単調増加;
- 任意の $n \ge 2$ に対して,

$$F_n(0) = \sqrt{\frac{n(n-1)}{2\lfloor n/2 \rfloor \lceil n/2 \rceil}} \quad \text{and} \quad \lim_{x \to \infty} F_n(x) = \frac{n+2}{4}.$$

よって, $\varepsilon_{\inf}(2, \varepsilon) = \varepsilon$ である. 次に, 大きい $n \ge 2$ を固定する. $\varepsilon > 0$ が十分小さいなら $F_n(e^{\varepsilon} - 1)$ は $\sqrt{2} = \sup_{m \ge 2} F_m(0)$ 未満であるため, 不等式 (3) の左辺と右辺にはかなり差がある. しかし, $\varepsilon > 0$ が十分大きいなら $F_n(e^{\varepsilon} - 1)$ は $(n+2)/4$ に近いため, 不等式 (3) の左辺と右辺は (n に関する増加レートの意味で) 同程度である.

Theorem 2 から次の系を得る.

Corollary 1. 任意の $n \ge 3$, $\varepsilon > 0$, $\varepsilon' \in \mathcal{E}_n(\varepsilon)$ に対して,

$$\frac{e^{\varepsilon'} - 1}{e^{\varepsilon} - 1} > \sqrt{\frac{n(n-1)}{2\lfloor n/2 \rfloor \lceil n/2 \rceil}}. \tag{4}$$

不等式 (4) の右辺は 1 より大きいので, 次の系を得る.

Corollary 2. 任意の $\varepsilon > 0$ と $n \ge 3$ に対して, $\mathrm{EC}_n(\varepsilon) \ne \mathrm{CQ}_n(\varepsilon)$.

実は, full ver. [1, Theorem 2.2] で, 任意の $n \geq 3$ に対して $S_n^{\mathrm{C}}(\varepsilon;\Phi) = S_n^{\mathrm{EC}}(\varepsilon;\Phi) < S_n^{\mathrm{CQ}}(\varepsilon;\Phi)$ となるような目的関数 Φ を構成している. そのため, このことからも Corollary 2 が従う.

参考文献

[1] Y. Yoshida. Mathematical comparison of classical and quantum mechanisms in optimization under local differential privacy. arXiv:2011.09960.

[2] J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science—FOCS 2013*, pages 429–438. IEEE Computer Soc., Los Alamitos, CA, 2013.

[3] C. Dwork. Differential privacy. In *Automata, languages and programming. Part II*, volume 4052 of *Lecture Notes in Comput. Sci.*, pages 1–12. Springer, Berlin, 2006.

[4] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography*, volume 3876 of *Lecture Notes in Comput. Sci.*, pages 265–284. Springer, Berlin, 2006.

[5] Q. Geng, P. Kairouz, S. Oh, and P. Viswanath. The staircase mechanism in differential privacy. *IEEE J. Sel. Topics Signal Process.*, 9(7):1176–1184, 2015.

[6] Q. Geng and P. Viswanath. The optimal noise-adding mechanism in differential privacy. *IEEE Trans. Inform. Theory*, 62(2):925–951, 2016.

[7] Q. Geng and P. Viswanath. Optimal noise adding mechanisms for approximate differential privacy. *IEEE Trans. Inform. Theory*, 62(2):952–969, 2016.

[8] N. Holohan, D. J. Leith, and O. Mason. Optimal differentially private mechanisms for randomised response. *IEEE Trans. Inf. Forensics Secur.*, 12(11):2726–2735, 2017.

[9] P. Kairouz, S. Oh, and P. Viswanath. Extremal mechanisms for local differential privacy. *J. Mach. Learn. Res.*, 17:Paper No. 17, 51, 2016.

[10] S. L. Warner. Randomized response: a survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc.*, 60(309):63–69, 1965.

[11] Y. Yoshida and M. Hayashi. Classical mechanism is optimal in classical-quantum differentially private mechanisms. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 1973–1977. 2020.