Isabelle/HOL を用いた差分プライバシーの形式的検証

佐藤 哲也
東京工業大学
e-mail : tsato@c.titech.ac.jp

1　差分プライバシー

差分プライバシー [1, 2] は Dwork らによって提案されたデータベースのプライバシーの基準である。差分プライバシーの基本的な考え方は、計算の過程でノイズを加えることで内部の個人情報を秘匿するというものである。差分プライバシーの身近な実用例として、iOS の入力候補の推定 [3] や、Google Map の混雑状況の算出 [4] が挙げられる (正確には局所差分プライバシー [5])。これらのサービスでは、入力文字列の断片や位置情報といった個人情報を大規模に集計し、結果を公開していて、無策だと個人情報漏洩のリスクが高い。しかし、データを集計する段階でノイズを加えることで、サービスに影響を及ぼすことなく、個人情報漏洩のリスクを許容可能なレベルに抑えている。

差分プライバシーの標準的な定義 (教科書 [6]) に移ろう。

定義 1. *(ノイズ加算処理のある)* 確率的アルゴリズム $M\colon \mathbb{N}^{|X|} \to \mathrm{Prob}(Y)$ が (ε, δ)-差分プライベート (DP) であるとは、隣接するデータセット $D, D' \in \mathbb{N}^{|X|}$ について以下を満たすことである。

$$\forall S \subseteq Y.\ \Pr[M(D) \in S] \leq \exp(\varepsilon) \Pr[M(D') \in S] + \delta. \tag{1}$$

ここで、データセット $D, D' \in \mathbb{N}^{|X|}$ が 1 か所だけ異なるとき、隣接すると呼ぶ。$0 \leq \varepsilon, \delta$ はそれぞれ、プライバシー損失 (確率の比) の上限と誤差で小さいほど確率分布 $M(D)$ と $M(D')$ は近くなる。

差分プライバシーは有用な性質を数多く持つが、紙幅のため本稿では触れない ([6] を参照)。

2　定理証明支援系 Isabelle/HOL を用いた差分プライバシーの形式的検証

本研究の動機は、精密化した評価を行う場合やプログラムが複雑な場合に、差分プライバシーを保証する何らかのツールがほしいというものである。差分プライバシーのプロジェクトである OpenDP [7] では、ソースコードに証明が与えられているが、そのような証明が正しいことを厳密に保証したい。このような目的に対して、定理証明支援系によるアプローチが有効だと考えた。

定理証明支援系は、数学の定理証明を支援するソフトウェアで、数学の定義や公理や命題を (プログラミング言語を用いて) 形式的に表現し、形式的証明によって、各ステップで誤りがない厳密な定理証明を可能とする。

以下、定理証明支援系 Isabelle/HOL [8] による差分プライバシーの形式的検証の概略を紹介する。本研究の設定として、確率的アルゴリズムを $M\colon \mathbb{N}^{|X|} \to \mathrm{Prob}(Y)$ を可測関数として定式化、Isabelle/HOL の数学的表現を用いて記述し、その差分プライバシーについて、Isabelle/HOL 上で形式的証明を与えることを目指す。

差分プライバシーの定義の不等式 (1) を以下のように Isabelle/HOL で記述し、差分プライバシーを示す述語 `differential_privacy` を構成する (`adj` は対称的とは限らない抽象的な隣接関係)。

```
definition DP_inequality::"'a measure ⇒ 'a measure ⇒ real ⇒ real ⇒ bool" where
"DP_inequality M N ε δ ≡ (∀ A ∈ sets M.  (measure M A ≤ (exp ε) * measure N A + δ))"

definition differential_privacy :: "('a rel) ⇒ real ⇒ real ⇒ bool " where
```



```
"differential_privacy adj ε δ ≡  (∀ (d1,d2) ∈ adj.
  DP_inequality (M d1) (M d2) ε δ ∧ DP_inequality (M d2) (M d1) ε δ) "
```

差分プライバシーの形式的証明の例として、以下の Report Noisy Max メカニズムを考える。

```
definition RMN_counting :: "real ⇒ nat list ⇒ nat measure" where
"RMN_counting ε x = do {y ← Lap_dist_list (1 / ε) (counting_query x);
                        return (count_space UNIV) (argmax_list y)}"
```

このアルゴリズムは、データセット `x` の要素を数え上げるクエリのリスト `counting_query` の中で最大のクエリ番号を返すものである。ただし、計算の途中で、各数え上げクエリの返却値に、尺度 $1/\varepsilon$、平均 0 のラプラス分布で生成したノイズを加算する (処理は `Lap_dist_list (1 / ε)`)。

差分プライバシーの基本的性質を用いると、数え上げクエリのリストのサイズ m に対し、Report Noisy Max メカニズムは $(m\varepsilon, 0)$-DP であることが容易に得られる。ところが実は、出力の分布を解析することで m によらず $(\varepsilon, 0)$-DP であることが示される (`adjacency_RNM_counting` は隣接関係)。

```
theorem differential_privacy_LapMech_RNM:
assumes pose: "(ε::real) > 0"
shows "differential_privacy (RMN_counting ε) adjacency_RNM_counting ε 0"
```

本講演では、差分プライバシーについてと、この定理に至るまでの形式化を、主に紹介したい。

参考文献

[1] Cynthia Dwork. Differential privacy. In Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener, editors, *Automata, Languages and Programming*, volume 4052 of *Lecture Notes in Computer Science*, pages 1–12. Springer Berlin Heidelberg, 2006.

[2] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In Shai Halevi and Tal Rabin, editors, *Theory of Cryptography*, volume 3876 of *LNCS*, pages 265–284. Springer Berlin Heidelberg, 2006.

[3] Apple Inc. Differential privacy overview. `https://www.apple.com/privacy/docs/Differential\_Privacy\_Overview.pdf`. Accessed: March 5. 2024.

[4] Miguel Guevara. Enabling developers and organizations to use differential privacy. `https://developers.googleblog.com/2019/09/enabling-developers-and-organizations.html`. Accessed: March 5. 2024.

[5] Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 2011.

[6] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3-4):211–407, 2013.

[7] OpenDP developers. Opendp proof initiation. `https://docs.opendp.org/en/stable/contributing/proof-initiation.html`. Accessed: August 1. 2024.

[8] Tobias Nipkow, Markus Wenzel, and Lawrence C. Paulson. *Isabelle/HOL: A Proof Assistant for Higher-Order Logic*. Springer-Verlag, Berlin, Heidelberg, 2002.

