IMES DISCUSSION PAPER SERIES

望ましいプライバシー保護のあり方を巡って：差分プライバシーの有用性と限界

菅 和聖（かん かずとし）

Discussion Paper No. 2022-J-5

IMES

INSTITUTE FOR MONETARY AND ECONOMIC STUDIES

BANK OF JAPAN

日本銀行金融研究所

〒103-8660 東京都中央区日本橋本石町 2-1-1

IMES Discussion Paper Series 2022-J-5
2022 年 5 月

望ましいプライバシー保護のあり方を巡って：
差分プライバシーの有用性と限界

菅 和聖（かん かずとし）*

要　　旨

現代社会では、個人情報の産業的な価値が高い一方で、個人情報の利活用にあたってはプライバシー保護との両立が求められる。差分プライバシーは、プライバシー保護の強さを定量的に評価する安全性基準であり、適度なプライバシー保護の実現に有用な概念である。本稿では、プライバシーを保護する枠組みの全体像を整理し、その中での差分プライバシーの位置付けを明確化する。さらに、差分プライバシーの理論やその応用研究を解説したあと、差分プライバシーの有用性と限界を踏まえながら、望ましいプライバシー保護のあり方を巡る課題を考察する。すなわち、自己情報のコントロールといったプライバシー保護に対する社会的要請に応えるためには、差分プライバシーのような数理的基準のみでは対処できない。社会的課題としてのプライバシー保護を実現するには、数理的技術や情報セキュリティに加えて、法制度、情報システムやビジネスの仕組みなどを総動員した対策が求められる。

キーワード：差分プライバシー、プライバシー保護、自己情報コントロール、匿名化、ELSI
JEL classification: Z00、K22

* 日本銀行金融研究所企画役（E-mail: kazutoshi.kan@boj.or.jp）

本稿の作成に当たっては、菊池浩明氏（明治大学）、金融研究所スタッフから有益なコメントを頂いた。また、山本慶子氏（金融研究所）には、法律に係る記述を校正して頂いた。ここに記して感謝したい。ただし、本稿に示されている意見は、筆者個人に属し、日本銀行の公式見解を示すものではない。また、ありうべき誤りはすべて筆者個人に属する。

目　次

１．はじめに

現代社会では、普及したスマートフォンやデジタル化したサービスを通じて、かつてない細かい粒度の個人情報[1]が継続的かつ自動的に収集されている。個人情報の産業的価値は高く、個人情報の利用が拡大している。例えば、個人データベースの整備、与信や入会など個人に利益／不利益をもたらしうるサービスに関する判断の機械学習（人工知能、AI）による自動化、企業や国境を跨ぐ第 3 者への個人データの提供・共有、個票データの利用拡大などが進んでいる。こうした状況下においては、プライバシー侵害の脅威も増しているため、プライバシー保護の重要性が高まっている。

個人情報を活用する際には、個人の権利保護と社会的利益の適度なバランスをとることが求められる。社会的利益としては、新産業の創出、防災や防犯による国民の安全確保、マイナンバーカードなどによる社会システムの効率性向上が挙げられる。他方で、個人情報の利用拡大は、プライバシー侵害の脅威も増大させる。例えば、単体では悪用が難しい個人情報であっても、名寄せによりデータベース化されることで脅威となりうる[2]。個人データベースは、AI と組み合せることにより、精緻な個人のプロファイリングを可能とする。久木田［2020］が指摘するように、プロファイリングが保険、人事、警察、裁判などに応用されると、差別的な偏見の助長など重大な問題を生じる[3]。こうした脅威は、AI が引き起こす倫理的・法的・社会的課題（ethical, legal, and social issues、以降本稿では ELSI と呼ぶ）の 1 つであり、AI 判断の公平性やプライバシーなどの観点から AI 倫理の領域で議論されている。

差分プライバシー（differential privacy、Dwork *et al.* [2006]）は、個人情報の有用性とプライバシー保護の強さにトレード・オフがあるもとで、両者の適度なバランスを実現するうえで有用である。差分プライバシーは、プライバシー保護技術の安全性を定量的に評価する基準であり、学術的な標準となっている。その基準は、特定の攻撃モデルに依存せず無条件で成立する安全性（unconditional security、または情報理論的安全性）に基づいている。この強力な性質は、以下の

---

[1] 本稿では、個人に関する情報という意味で用いる。特段の断りがない限り、個人情報保護法における「個人情報」とは定義が異なる。また、本稿では、個人情報の保護は、後述する自己情報コントロールを実現する意味でのプライバシー保護に含まれるものと位置付ける。

[2] 米国では、民間の信用情報機関が個人情報を収集し信用スコアの算出を行っている。代表的な例に Fair Isaac Corporation による FICO スコアが挙げられる。米国では、信用スコアが融資・保険などの金融サービスを通じて、消費者の生活に大きな影響を与えている。米国の信用スコアとそれに対する法規制を参照しつつ、日本における信用スコアを巡る法的問題を論じたものとして、例えば林［2022］を参照。

[3] AI がその設計者や社会の持っている差別的な偏見を学習してしまう問題については、Benjamin [2019]、O’Neil [2016]などを参照。

理由から重要性が増している。第 1 に、プライバシーの暴露を試みる攻撃者が利用できる情報や計算資源が増大しているため、攻撃者に特定の背景知識（個人データベースに含まれない攻撃対象者に関する情報）や攻撃手法を仮定する攻撃モデルに基づく安全性解析では、プライバシーを保証することが一般に困難になった[4]。これに伴い、攻撃モデルによらず安全性を証明できる情報理論的アプローチの重要性が高まった。第 2 に、プライバシー侵害の脅威が増すなかで、より予防的な措置が望まれるようになった。そうした需要に対し、差分プライバシーが保証する情報理論的安全性は、将来発見される惧れのある未知の攻撃手法に対しても有効であるという利点を提供している。

米国のセンサス局（United States Census Bureau [2019]）は、2020 年の国勢調査から差分プライバシーを採用し、セル秘匿（suppression）[5]やデータ・スワッピング（data swapping）[6]と呼ばれる従来の「場当たり的（ad-hoc）」な手法と決別した。その理由として、個人データベースから得られる統計値の組合せから、元データの一部または全部を逆算する再構築攻撃（reconstruction attack）の脅威が、理論上の脅威にとどまらなくなったことなどを挙げた（Garfinkel, Abowd, and Martindale [2019]、Census Scientific Advisory Committee [2021]）。このほか、Google 社や Uber 社などの大手企業も、個人データを収集する過程に差分プライバシーを導入している。

もっとも、差分プライバシーは万能の処方箋ではない（詳しくは 4 節を参照）。すなわち、プライバシー保護の強さを表すパラメータ（ϵ、プライバシー予算）を定める方法には、理論的にも実務的にも明確な合意がない。また、差分プライバシーを満たす代表的な手法であるラプラス・メカニズムからは、実用に足る十分な精度や性質を備えた出力データが得られない場合がある。この問題に対処するためのメカニズムの設計は数理技術的に容易でない。差分プライバシーを理論的に保証するために個人データベースに置いた前提を、実務データが必ずしも満たすとは限らない。実務に応用するには、こうした課題を個別の応用例ごとに克服する必要がある。さらに、差分プライバシーは、個人データベースが任意の統計クエリを受け付けるという利用状況を想定しており、あらゆる状況で

---

[4] 背景知識によってプライバシーが暴露される問題として、Sweeney[2002]は、マサッチューセッツ州の医療保険の個人データと、同州ケンブリッジの選挙人名簿の個人データを、氏名の情報を用いずに、性別・郵便番号・生年月日を手掛かりに突合できることを指摘した。この他のプライバシー暴露の事例については、例えば佐久間［2016］を参照されたい。

[5] セル秘匿は、一定の基準に基づいて統計表またはその一部を秘匿する手法である。例えば、頻度が一定値以下のセルを秘匿するといった基準を設定する。

[6] データ・スワッピング（Dalenius and Reiss [1982]、 Willenborg and Waal[2001]）は、個票のレコード・データ同士で属性値を入れ替える手法である。2010 年の米国国勢調査で利用された。

有望な選択肢とまではいえない。したがって、差分プライバシーを安全性基準として採用する際には、差分プライバシーの有用性と限界の両方を考慮することが求められる。

さらに、プライバシー保護は社会的課題であり、差分プライバシーのような数理的な基準のみでは対処できない。膨大な個人情報を企業や国家が活用しうることへの懸念から、プライバシーの概念として、「機微な情報[7]の漏洩防止」を拡大した「自己情報のコントロール権[8]」を採用する動きが国際的に拡がった。プライバシー保護の目標を定めることは技術的課題の範疇を超えた社会的課題である。こうした社会的要求に応えるには、数理的技術や情報セキュリティといった技術的要素に加えて、法制度、情報システムやビジネスの仕組みなども総動員して対処することが望ましい。

本稿の構成は以下のとおりである。2節では、プライバシー保護の枠組みの全体像の整理を試みる。3節では、差分プライバシーを含む数理的手法を概観する。4節では、差分プライバシーの理論とメカニズムを解説する。5節では、応用研究を紹介する。6節では、差分プライバシーの有用性と限界を踏まえて、その普及に向けた課題と望ましいプライバシー保護のあり方を考察する。

2．プライバシーの概念とその保護の枠組みの全体像

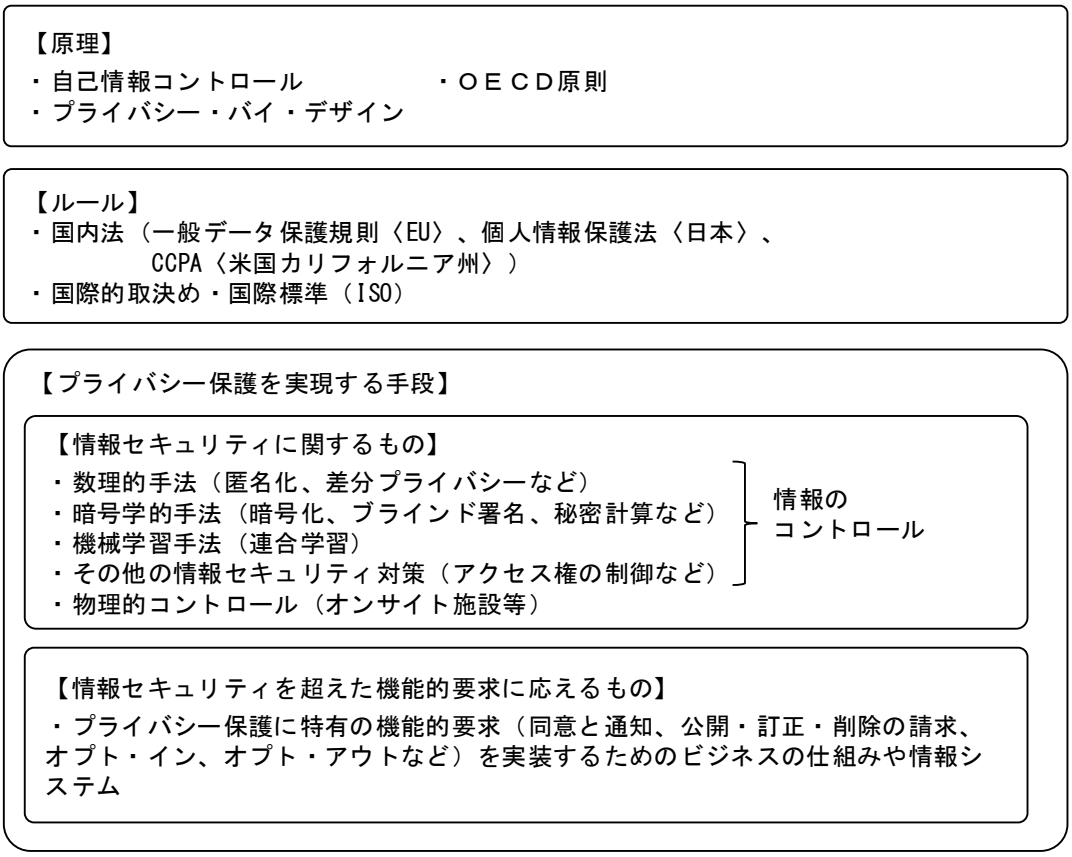
社会受容性に制約があるもとで、個人データの有用性を最大限に引き出すための対応策の総体を、プライバシー保護の枠組みと呼ぶ。本節では、これらを「原理」、「ルール」、「手段」の3つのカテゴリに分類する（図1）。「原理」はプライバシーの概念やその保護のあり方の理念である。「ルール」は国内法や国際的な取決めを表す。「手段」は、プライバシー保護を実現するために実装される技術や仕組みを表す。さらに、手段は、その目的が情報セキュリティの範疇におさまるものと、情報セキュリティの範疇を超えたプライバシー保護に特有の機能的要求に応えるものとに分けられる。これらのカテゴリの各要素が整合的かつ総合的に作用しながら、総体としてプライバシー保護が達成される。とくに、個人情報がデータベース化されて情報システム上で処理される現代では、情報技術と数理的技術が重要な役割を果たす。以下では、各カテゴリとプライバシー保護への社会的要請について説明する。

---

[7] 本稿では、他人に知られたくない個人に関する情報という意味で用いる。金融分野における個人情報保護に関するガイドラインにおける「機微情報」とは定義が異なる。

[8] 自己情報コントロール権については、例えば中川［2016］、曽我部・林・栗田［2019］を参照。

図１　プライバシー保護の枠組みの全体像



（１）プライバシーの概念とその保護の原理

プライバシーの概念は、「他人に知られたくない個人情報の漏洩防止」から、「自己情報のコントロール権」に拡大している。1980 年代以前は、プライバシーは、都市などの物理的な空間の中で「一人にしてもらう権利」であると考えられた。インターネットが普及した現代では、プライバシーの定義が、「忘れられる権利（the right to be forgotten）」[9]、または「追跡拒否権（do not track）」[10]が含まれうるものとなった[11]。1960 年代以降の情報化の進展により、国家や私企業が個人情報を大量に保有することが問題視されるようになったことを背景として、自身に関する個人データについて開示、訂正、削除を要求できる「自己情報コントロール」をプライバシーとする考え方が提唱されている。さらに、個人データがどのような目的に使われているかを知り、かつその目的に応じて自己情報の

[9] 忘れられる権利とは、検索エンジンに対して、自分に関する情報が書かれた Web ページへのアクセスを遮断させる権利である。

[10] 追跡拒否権は、検索エンジンに対して、自分に関する情報が記載されたページを現時点以降収集させない権利である。

[11] EU や米国における法制度との関係については、例えば石井［2014］を参照。

利用を許可あるいは拒否できることをプライバシーと定義するパーソナル・データ・エコシステム（personal data ecosystem、Cavoukian [2012, 2013]）という考え方も提唱されている。

1980 年に制定された経済協力開発機構（Organisation for Economic Co-operation and Development: OECD）のプライバシー・ガイドラインは、自己情報のコントロールの概念を本質的に内包しており、日本を始め、OECD 加盟各国のプライバシー保護法制の基本となっている。また、インターネットの普及や、個人データの国境を跨いだ流通（越境）が盛んになったことを受け、2013 年に改正された（OECD [2013]）。

ガイドラインでは、データ主体への通知と同意（第 7 項）、利用目的の明確化（第 9 項）、利用目的の範囲内で（第 8 項）収集したデータに安全管理措置を講じること（第 11 項）、データの開示・訂正・削除の請求権の保障（第 13 項）などの普遍的な原理が謳われている。利用目的の範囲内で、可能な限り収集するデータを減らして第 3 者とも共有しない原則を、データ最小化（data minimization）と呼ぶ。2013 年の改正では、プライバシー執行機関の設置、国際間でのプライバシー保護の執行協力、一貫したプライバシー法整備を求めている。

プライバシー・バイ・デザイン（privacy by design、Cavoukian [2011]）は、プライバシー保護技術（privacy enhancing technologies: PETs）や法規制の遵守のみでは、プライバシー保護が達成できないとの課題意識のもとで提唱された理念である。とりわけ、巨大な個人データベースの所有者に比べてデータを提供する個々人の立場は弱い。この力の非対称性により、適切なプライバシー保護措置がとられない状況が放置された場合には、個人データを提供するインセンティブが失われ、個人情報活用による社会的な恩恵を享受できなくなる惧れがある。現在では、プライバシー・バイ・デザインの理念は、EU や米国のプライバシー保護制度に取り入れられている。

プライバシー・バイ・デザインは 7 原則からなり、事前の予防措置をとること（原則 1）、プライバシー保護をデフォルトとし（原則 2）、システムの設計に組み込むこと（原則 3）、プライバシー保護は事業者と利用者の双方に利益のあるポジティブ・サムであること（原則 4）、ライフサイクルにわたってデータを保護すること（原則 5）、プライバシー保護の仕組みを可視化・透明化し（原則 6）、利用者中心のものとすること（原則 7）が提唱されている。

（２）ルール

ルールは、国内法、国際的な取決め、国際標準、業界団体による自主規制などからなる。ここでは、主要な国内法のみを紹介する。

欧州では、OECD のガイドラインに基づいて、一般データ保護規則（General

Data Protection Regulation: GDPR）が法制化された。欧州においては、プライバシー権は基本的人権の 1 つと考えられている。米国では、カリフォルニア州消費者プライバシー法（California Consumer Privacy Act: CCPA）が制定されている。基本的な考え方は、憲法上に記された消費者のプライバシー権を推進することである。

日本において、プライバシー保護に関連する法律には、民法、個人情報保護法などがある[12]。「プライバシー権」は憲法 13 条に定める幸福追求権に基礎付けられる人格権の 1 つとして保護されており、自己情報コントロール権を指すとの考えが通説である（曽我部・林・栗田［2019］）。個人情報保護法は、OECD のガイドラインに沿うかたちで制定され、「個人の権利利益を保護することを目的としている」（個人情報保護法第 1 条）。この権利利益には、プライバシーなどの人格的な権利利益と財産的な権利利益が含まれる。ただし、個人情報保護法には、プライバシー権や自己情報コントロールとの関係が、共通の理解が確立していないなどの理由で明記されていない（曽我部・林・栗田［2019］）。2017 年の改正では「匿名加工情報」のカテゴリが新設され、情報の第 3 者流通の要件が緩和された。匿名加工情報は、特定の個人を識別することができないように個人情報を加工し、当該個人情報を復元できないようにした情報と定義される。「特定の個人を識別することができない」との文言の解釈は、高度な技術による特定を適用対象としておらず、一般的な企業が通常の方法により特定できなければよいこととなっている[13]。

（3）プライバシー保護を実現するための手段

イ．情報セキュリティに関するもの

プライバシーを機微な情報の改竄や漏洩の防止といった狭い意味で捉える場合は、プライバシー保護は情報セキュリティの範疇におさまる[14]。情報セキュリティは、情報のコントロールと物理的コントロールに分けられる（図１）。情報のコントロールは、一部要素の削除や追加といったデータの劣化を伴う数理的手法と、データの劣化を伴わない暗号学的手法、および機械学習手法に主として

12 このほか、刑法は、名誉棄損、侮辱罪、名誉感情の侵害に対する罰則を規定している。

13 高度な技術による個人識別の可能性は排除されないため、一般論としては、匿名加工情報が社会全体に漏洩すれば、プライバシーが侵害される可能性は否定できない。このほか、個人情報を訓練データに用いた機械学習モデルのパラメータからは、5 節(4)で後述するように、訓練データの情報が逆算できる可能性があるが、訓練済みの機械学習モデルは個人情報保護法の対象外となっている。

14 情報セキュリティの 3 原則は、情報の完全性（integrity）、機密性（confidentiality）、可用性（availability）を確保することである。機微な情報の漏洩防止は機密性の確保に相当する。機微な情報に無謬性や耐改ざん性を持たせることは完全性の確保に相当する。

分けられる。数理的手法には、主に統計やデータ・マイニングの分野で利用される、差分プライバシーや統計的開示制御などが含まれる。数理的手法は、３節で詳述する。

暗号学的手法は、暗号化によって匿名性や情報の機密性を達成するものである[15]。暗号鍵の情報があれば情報を復元できるため、情報は劣化しない。通常の通信における暗号化手法や、金融取引などでの匿名性を達成するブラインド署名およびゼロ知識証明の１つである zk-SNARKs（zero-knowledge succinct non-interactive argument of knowledge、Gennaro *et al.* [2013]）、秘密計算（secure computation）[16]、秘密分散（secret sharing）[17]が含まれる。近年、プライバシー保護を、追加的なセキュリティ目標として考慮する研究（Ramacher, Slamanig, and Weninger [2021]、Lian *et al.* [2021]）が進展している。インターネットで広く利用されている暗号化プロトコル TLS でも、プライバシー保護の観点から、通信相手の特定につながる通信内容の一部（Client Hello やサーバ名）を秘匿する修正（Encrypted Client Hello[18]や Encrypted Server Name Indication[19]）が提案されている。ここでのプライバシー保護では、情報の機密性だけではなく、追跡可能性（traceability）、リンク可能性（linkability）などの安全性レベルの異なる複数のプライバシー要件が考慮されている[20]。

機械学習手法は、プライバシー保護を考慮した機械学習モデルの訓練手法である。その代表例である連合学習（federated learning、McMahan [2017]）は、複数の企業がそれぞれの個人データベースを使って１つの機械学習モデルを構築

15 ただし、個人情報保護法上の「個人情報」を暗号化したものは、情報の機密性が保たれていたとしても引き続き「個人情報」に該当する点には変わりがない。

16 秘密計算は、暗号化したままデータを処理する技術である。複数の企業が機密データを持ち寄り、互いに秘密を開示せずにデータを共同利用しながら分析を行う場合に有用である。具体的には、準同型暗号を利用することで、データを暗号化したまま演算処理できる。準同型暗号とは、平文に対する演算処理の結果と、対応する暗号文に対する演算処理の結果を復号したものが一致する性質をもつ暗号である。秘密計算によるプライバシー保護は、限られた利用者間でのみ提供される。

17 秘密分散は、１つの個人データベースを複数のサーバに分散して暗号化して保有する。データを復号するには、2 つ以上のサーバのデータを組み合せる必要があるため、単一のデータ保有主体ではデータを読み出せない。秘密分散は、セキュリティ問題における「データ保有主体を信頼する」という前提を緩和することができる。

18 TLS Encrypted Client Hello (draft, https://www.ietf.org/archive/id/draft-ietf-tls-esni-08.html)

19 Encrypted Server Name Indication for TLS 1.3 (draft, https://www.ietf.org/archive/id/draft-ietf-tls-esni-06.txt)

20 個人データの秘匿は、データ内容の秘匿（数値などがわからないこと）と特定人物との対応関係の秘匿（数値はわかるが誰のデータかわからないこと）に分解できる。後者の紐付けが可能である性質をリンク可能性と呼ぶ。追跡可能性は、匿名化された個人データについて、ある（匿名の）人物に関する複数のデータを関連付けられる性質を表す。リンク可能性が満たされていれば追跡可能性を持つ。

する手法である。機械学習モデルの訓練時に、各企業は互いに個人データベースの情報を開示しあう必要はない。このように個人データベースを 1 つに集約することなく、分散的に訓練を行う点に特徴がある。ただし、各個人データベースから送信される機械学習モデルの更新情報から個人データに関する情報が漏洩するリスクがある。このリスクへの対策のため、モデルの更新情報を第 3 者に読み取られないようにする手法（secure aggregation）が Bonawitz *et al.* [2017]により提案されている。

その他の情報セキュリティ対策は、脆弱性のあるプログラムの修正や個人データベースへのアクセス権の設定といった安全管理措置である。情報セキュリティの基本的な対策は、一般的に情報システムを利用する際に必ず求められる。プライバシー保護の観点では、社内利用など利用者が限定された環境で個人データベースが運用される場合に、アクセス権を利用者のみに与えるよう設定するといった対応が想定される。

物理的なコントロールは、個人情報を扱う状況を物理的に制限するものである。例えば、日本の公的統計の「オンサイト利用施設」[21]は、学術研究など公益性のある利用目的に限り、物理的に隔離された空間内で、カメラなどによる監視のもと、個票データへのアクセスが許可される。公益性のある用途に限られるものの、公的統計のデータについても 2 次利用の制度が整備されてきている [22]。

ロ．情報セキュリティを超えた機能的要求に応えるもの

プライバシーを自己情報のコントロール権と広い意味で捉える場合には、情報セキュリティの範疇を超えたプライバシー保護に特有の機能的要求に応えるため、制度や情報システムのデザインなどに関する措置も必要となる。個人データベースに対する情報の公開・削除・訂正の請求権、オプト・アウト、オプト・インといった仕組みは、暗号学ではカバーされないプライバシー保護目標であ

---

21 行政機関および大学などの学術研究機関に設けられている。施設一覧は以下を参照。
https://www.e-stat.go.jp/microdata/data-use/on-site-facilities

22 公的統計の分野では、統計調査で得た情報の利用を促進するため、学術研究など公益性の高い目的などを想定した 2 次利用の制度を整備している。2 次利用の方法には、(1)データ利用者からのニーズに応じて新しい統計を作成するオーダーメイド集計、(2)個人や団体が識別・推定できないよう加工された匿名データの提供、(3)調査票データの提供がある。

とくに、(3)について、統計法の改正（「統計法及び独立行政法人統計センター法の一部を改正する法律」〈平成 30 年法律第 34 号〉、令和元年 5 月 1 日施行）では、複数の統計の（氏名などの識別子を削除して仮名化した）個票データを接続して探索的に高度な統計分析を行う「高度利用」への道が開かれている。この利用方法の特徴は、個票データ分析の自由度や利便性を高くする一方、利用状況の監視や分析結果の持ち出しを厳格に管理するというものである。分析は、「オンサイト施設」と呼ばれる物理的に隔離された場所で、カメラなどによる監視のもとで行う必要がある。

詳しくは、https://soumu.go.jp/toukeitoukatsu/index/seido/2jiriyou.htm を参照。

る。これらは、プライバシー・バイ・デザインの考え方に則り、プライバシー保護を目的とした機能として、事業者によって、情報システムやビジネスの仕組みの中に計画的に織り込まれ、実装されることが望ましい。

（４）社会からの要請としてのプライバシー保護

プライバシー保護の枠組みが整備されるとともに、1 節で述べたように、個人データの活用が一層拡大してきており、脅威に晒される個人情報が量的に増大している。個人情報の利用が拡大すると、プライバシーの概念や安全性に関する社会からの要求が高まる。とりわけ、プライバシー侵害は原状回復が困難であり、悪用による潜在的な脅威も知覚しにくいため、プライバシー侵害の脅威が増大すると、より予防的な措置が望まれる。

社会からの要請としてのプライバシー保護は、純粋に技術のみで解決できる課題ではない。プライバシーの本質は侵害されて初めて認識できるものであり、その被害は個人差のある主観的な要素を含む。また、同一個人への影響についても状況依存性[23]を帯びる。

プライバシーの概念の定義やその保護の目標の定め方を巡っては、個人の主観や主義の多様性に配慮しながら、社会的な合意を形成する過程を経る必要がある。図 1 のプライバシー保護の枠組みは、個人情報の利用の進展とこれを受けた社会的合意の形成を反映しながら、変化していくものと考えられる。このような社会課題としてのプライバシー保護は、６節（3）で述べるように、新技術が社会にもたらす ELSI への対応策の文脈に位置付けることもできる。

３．プライバシー保護のための数理的手法

プライバシー保護は、統計学、データベース、暗号学、情報理論などの複数の分野にまたがる学際分野として発展している。関連する技術群は、プライバシー保護データ・マイニング（privacy preserving data mining: PPDM）と総称される。訓練済みの機械学習モデルなどの高度な加工情報からプライバシーを保護する手法（Agrawal and Srikant [2000]）も研究されている。本節では、これらのうち、情報の劣化を伴うような確率的または統計的な手法を説明する。

（１）数理的手法の分類

数理的手法は、①適用される段階、②有効範囲、③安全性の原理、という軸に従って便宜的に分類することができる（図２、表１）。図２において、(a)の個人

[23] 例えば、ある個人が賭博目的で非合法の金融から資金を借り入れたとの情報は、その人物が就職活動中であれば、とくに重大な被害をもたらす惧れがある。

データに変更を加えるアプローチでは、個人データベースの管理者が信頼できる者であるとの前提が不要である。(b)の個人データベース自体を個人の特定ができないように加工するアプローチでは、データを第 3 者に流通させる場合にも有用である。このほか、(c)の出力データを加工するアプローチがある。

図２　個人情報分析の流れ

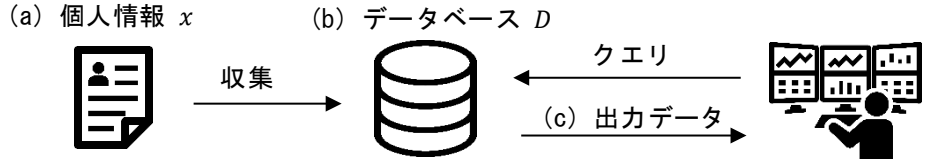


表１　数理的手法と安全性の原理

	適用段階	有効範囲 (注)	安全性の原理
局所差分プライバシーを満たす手法	(a)	中央管理者、利用者	データを確率的にしか推定できない（確率的）
仮名化／匿名化	(b)	利用者	仮名から人物を特定できない（決定論的）
k-匿名化	(b)	利用者	同一IDを持つk人を区別できない
合成データ	(b)	利用者	確率分布から生成した合成データから、確率分布の推定に利用した元データを復元することが困難
統計的開示制御	(c)	利用者	セルの秘匿（決定論的）
差分プライバシーを満たす手法	(c)	利用者	データを確率的にしか推定できない（確率的）
ランダム・サンプリング	(c)	利用者	出力を得るために、どのデータが利用されたのかが断定できない（確率的）

（注）データを秘匿できる対象者の範囲。「中央管理者」は個人データベースの管理者、「利用者」は個人データベースから作成された統計情報の利用者を表す。

（２）各数理的手法の紹介

本節では、概ね表 1 の順に沿って各数理手法を紹介する。

イ．局所差分プライバシーを満たす手法

局所差分プライバシー（local differential privacy）を満たす手法は、入力データをランダム化することで、個人データの推定を困難にする。差分プライバシーと同様に、任意の背景知識をもつ攻撃者に対して安全性を保証できることに加え、個人データベースの管理者を信頼する前提が不要である。差分プライバシーについては、本節（2）ホ.、および４節で詳しく述べる。局所差分プライバシーについては、４節（5）で整理する。

ロ．匿名化・仮名化

匿名化（anonymization）は、個人データの中から氏名などの個人を直接的に識別できる情報（個人 ID、疑似 ID）を削除する操作である。単体で個人を識別できる情報を「個人 ID」と呼ぶ。複数の情報の組合せで個人を識別できる情報を「疑似 ID（quasi-indentifier）」と呼ぶ[24]。

仮名化（pseudonymization）は、個人 ID をランダムに発生させた番号（仮名、pseudonym）に置き換える操作を指す。仮名の生成には、乱数を発生させて個人 ID との対応表を保存しておく方法や、個人 ID のハッシュ値を用いる方法などがある。個人データベースの中に 1 人の個人データが複数レコード含まれている場合には、複数の仮名を割り当てて匿名性を高められる（多重仮名化）。

ハ．*k*-匿名化

k-匿名化（*k*-anonymization、Sweeney [2002]）は、疑似 ID を変形する[25]ことにより、任意の個人を同一の疑似 ID を持つ *k*-1 人の中に紛れこませる手法である。同一の疑似 ID を *k* 人以上が持つ状態を *k*-匿名性（*k*-anonymity）という。ただし、*k*-匿名化には属性暴露のリスクがある。例えば、医療データベースにおいて、ある疑似 ID を持つ個人が *k* 人以上おり、その全員が消化器系の疾患を抱えていることが判明したとする。この場合には、その疑似 ID を持つ個人を *k* 人の中から特定することはできないが、消化器系の疾患を持つという属性情報は漏洩してしまう。

この問題点を克服するために考案された *l*-多様性（*l*-diversity、Machanavajjhala *et al.* [2007]）は、同一の疑似 ID を持つ *k* 人の属性の種類がある値（*l*）よりも多くなることを保証することにより、属性暴露を防ぐ。上記例を援用すると、*k* 人の疾患タイプが *l* 種類以上あるように疑似 ID が変形されるため、変形後の疑似 ID から疾患タイプを絞り込めないことに相当する。なお、*l*-多様性があっても、属性の値の分布に偏りがある場合には、プライバシー侵害のリスクが残る[26]。こ

---

24 例えば、住民基本台帳に記されている基本 4 情報（氏名、性別、生年月日、住所）の組合せは、ほぼ確実に個人を識別できる。

25 疑似 ID を変形する方法には、一般化とレコード削除の 2 つがある。一般化とは、疑似 ID の情報量を削減することである。例えば、生年月日の情報（1990 年 3 月 14 日）から年（1990 年）だけを取り出すことが一般化に相当する。レコード削除とは、ある個人のデータ（レコード）自体を削除することである。レコード削除は、一般化によって他のレコードと同一の疑似 ID を作成できない場合に利用される。

26 例えば、*k* 人からなるグループの年収の範囲が｛100 万円～110 万円、110 万円～120 万円｝のいずれかの場合には、年収の範囲が近接しているため、年収のバンドが 100 万円～120 万円に絞り込まれてしまう。仮に、年収の範囲が｛100 万円～110 万円、600 万円～610 万円｝のいずれかの場合には、上記のような 20 万円の狭いバンドに年収を絞り込むことができない。

のような属性値の近接性を考慮して*k*-匿名性を保つ手法に、*t*-近似性（t-closeness、Li, Li, and Venkatasubramanian [2007]）がある。

k-匿名化は、実装が容易である反面、消去する属性情報が多いと失われる情報が多くなる。また、安全性の解析は攻撃モデルに依存する。さらに、変形される疑似 ID の数が最も少ないという意味で最適な *k*-匿名化を行う問題は、NP 困難である（効率的な求解が困難であると信じられている計算問題のクラスに属する）ことが知られている（Meyerson and Williams [2004]）。

二．統計的開示制御

統計的開示制御（statistical disclosure control、Hundepool *et al*. [2010, 2012]）は、匿名化データや集計表を加工する技術の総称である。プライバシー保護を巡る研究の歴史において、統計的開示制御は 1980 年代ごろより、統計分野とデータベース分野において研究されてきた。

統計的開示制御は、国内外の公的統計の分野で、集計表を公表する際に、それらを利用してプライバシーの暴露が生じないよう、情報量を制限するものである。具体的には、機微にふれる集計値を非公開とするセル秘匿基準を設定する。集計値のうち上位n個体が総和のk%以上を占めるものを秘匿する$n-k$%占有ルール（*n*-*k* dominance rule）などが用いられる。ただし、*k*-匿名性と同様にプライバシー侵害のリスクを評価する際には、特定の攻撃モデルの前提を置く必要がある。

ホ．差分プライバシーを満たす手法

差分プライバシーは、プライバシー保護技術を情報理論的安全性に基づいて定量的に評価する基準である。この概念は、2006 年に暗号学者のドウォーク（Dwork *et al*. [2006]）により、暗号学的手法に基づいて定式化された。従前の研究との関係では、情報理論とともに発展した暗号学の観点から、プライバシー保護技術の 1 つとして研究されていた推論制御（inference control）を発展させた概念が差分プライバシーである、と位置付けられる（暗号学との関係は４節（2）を参照）。推論制御は、任意のクエリを許容する場合のプライバシー保護方法の模索するものとして 1980 年頃より存在している（岩村・西島 [1991]、Denning, Denning, and Schwarts [1979]、 Denning [1980]、Beck [1980]、Denning [1982]）[27]。

出力データが確率変数となる確率的手法が差分プライバシーを満たすと、任意の背景知識を持つ攻撃者に対してプライバシーを保証できる。差分プライバシーについては、４節で詳述する。

---

[27] 差分プライバシーのプライバシー保護データ・マイニングの中での位置付けについては、例えば、五十嵐・高橋［2012］を参照。

へ．ランダム・サンプリング

ランダム・サンプリング（例えば、Adam and Wortmann [1989]を参照）は、個人データベースのなかから一部をランダムにサンプリングすることで、統計量などを算出する手法である。この手法の安全性の原理は、個人データベースから出力データを作成する際に、ある個人のデータが使われているか否かを攻撃者は完全には判別できない、というものである。

ト．合成データ

合成データは、個票データの統計的性質を保つ新しい個票データを生成する技術である。元の個票データを公表せずに、合成データを公表することにより、プライバシーが保護される。ユーザには、施されたプライバシー保護技術に影響されずに、元の個票データに対するデータ分析手法を踏襲できるメリットがある。一般的には、元の個票データから統計的な特徴を捉えるパラメータを推定し、これを用いて合成データを生成する。合成データから元データが推定されるリスクは低いとされる。もっとも、極端なデータ値は情報が漏洩する可能性があることから、差分プライバシーを満たす合成データの生成方法も研究されている。また、個票データが持つ統計的性質は多様であるため、合成データの広範な用途での活用は容易ではない。

チ．複合的な手法

これまでに紹介した手法を組み合せることもある。Li, Qardaji, and Su [2012]は、差分プライバシーの定義を拡張したうえで、ランダム・サンプリングと組み合せた *k*-匿名化が拡張された差分プライバシーを満たすことを示した。

また、匿名化と連合学習を組み合せて、各個人にグループ ID を割り振る手法もある。Google 社の研究者らは、連合学習（２節（3）イ．を参照）を利用して、属性の似た個人をグループにまとめ、グループに ID を付与する「フロック（federated learning of cohorts: FLoC）」いう手法を提案した[28]。フロックは、Web ブラウザの閲覧履歴を収集し、ターゲティング広告に利用されるサード・パーティ・クッキーの代替技術の候補である。近年、ターゲティング広告の過度な追跡がプライバシーの侵害にあたるとして、サード・パーティ・クッキーの利用を制限する動きが拡がっている。

４．差分プライバシーの理論

本節では、差分プライバシーの理論的基礎を簡単に紹介する。詳細については、例えば、Dwork and Roth [2014]や佐久間［2016］を参照されたい。

[28] https://github.com/WICG/floc

（１）情報理論的安全性の重要性の高まり

差分プライバシーが提供する情報理論的安全性の重要性は高まっている。その背景として、以下の点が挙げられる。

第１に、攻撃者が利用できる個人情報が増している。流通する個人情報は個々には利用価値が小さいものであっても、名寄せによりデータベース化されると、個人の特性を包括的に暴露するものとなりうる。名寄せされたデータベースを攻撃者が悪用できる場合には、それ自体が脅威であるほか、匿名化などのプライバシー保護措置が施された個人データベースでも個人を特定しうる[29]。個人データベースに含まれない外部情報（背景知識）や、個人データベースから公表される加工情報の量も増している。また、リレーショナル・データベースに対するSQLによる任意のクエリの処理が普及し、統計ユーザの要求に応じてテイラー・メイドで個人データベースを集計する高度な利用が拡がっている。集計表や各種のクエリ結果のほかデータベースには含まれない外部情報（背景知識）は、個別には安全であっても、これらの組合せにより個人データの一部が暴露する惧れがある。公表情報の組合せは膨大であり、すべての組合せについてリスクを検証することは不可能である。

第２に、攻撃者が利用できる計算資源の性能が向上している。集計表から元データを逆算するデータベース再構築攻撃（database reconstruction attack）[30]は、計算能力の制約から現実的には実行が難しいとされてきた。しかし、計算機の能力向上に伴い、再構築攻撃の脅威が理論上のリスクから対策が求められる課題に変化しており、米国の国勢調査が差分プライバシーに基づくプライバシー保護に移行する動機となった[31]。

第３に、事前にあらゆる攻撃を想定することは困難である。公的統計では、個人データベースの集計結果を統計表などのかたちで公表している。前述のように、複数の統計表を組み合せると、個人の情報が漏洩するリスクがある。このリスクは、伝統的には統計的開示制御によって、事前に想定できる限りのプライバシー暴露攻撃を防げるように専門家によって注意深く抑制されてきた（Hundepool *et al*. [2010]）。もっとも、このアプローチでは、統計発表時に未知である攻撃には備えられない。

---

29 例えば、個人の位置情報と時刻を蓄積した匿名化済みの個人データベースを攻撃者が利用できる場合、「ある時刻にある場所で特定の人物を見かけた」という情報が新たに入手できると、個人データベースから当該人物のレコードを探し当てうる。

30 複数の統計表を制約条件として、データベースに含まれる個人データを逆算する攻撃。問題の種類としては、いわゆる虫食い算や数独パズルのようなもの。

31 米国センサス局が 2019 年 10 月に公開した「A History of Census Privacy Protections」と題したパンフレットでは、前回（2010 年）の国勢調査について、「その場しのぎのプライバシー保護（ad-hoc privacy protections）を使う最後の国勢調査」としている。

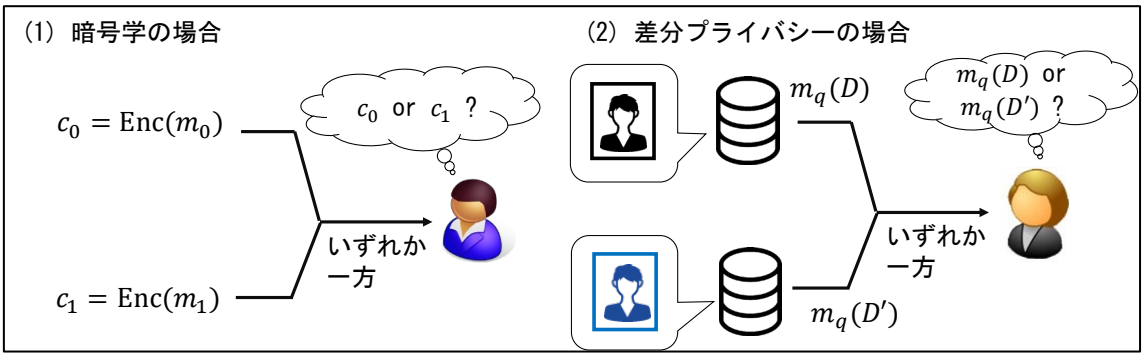
一般に、背景知識をすべて考慮することは難しいため、仮名化や匿名化といった手法のみでは、プライバシーを保証することは技術的には不可能に近い。実際、k-匿名化は差分プライバシーを満たさない。これらの手法の安全性評価は、それぞれ特定の攻撃者のモデルに依存している。

これに対して、情報理論的な安全性を達成する差分プライバシーに基づく手法は、任意の背景知識をもつ攻撃者に対して有効である。その利点は、堅牢性の高さであり、米国センサス局が懸念する再構築攻撃に加え、将来発見される惧れのある想定外の攻撃に対しても安全性を保証できる点にある。この強力な性質により、差分プライバシーに基づく安全性評価は、予防的なプライバシー保護措置として望ましい。ただし、差分プライバシーは、データベースに任意クエリを投げて統計処理を行う状況を主として想定しているので、定型的な統計表しか公表しない場合にも必要とまではいえず、どんな状況でも利用すべき手法ではない。

（２）差分プライバシーの定義

差分プライバシーは、暗号学の識別不可能性の概念をプライバシーに転用したものである。現在では、これがプライバシーの学術的に正式（formal）な定義とされている。暗号学では、2 つの平文m_0とm_1が与えられたとき、これらに対応する暗号文$c_0 = \mathrm{Enc}(m_0)$、$c_1 = \mathrm{Enc}(m_1)$を識別することが（計算量的に）困難であること（indistinguishability）をもって、暗号方式の安全性の証とする[32]。

図３ 暗号学の識別不可能性とプライバシーの識別不可能性



差分プライバシーは、2 つの確率分布の識別が困難であることに基づいて定義される（図３）。ある個人xのデータのみが異なる 2 つの個人データベースDとD'を用意する。両者から得られる出力値の確率分布$m_q(D)$と$m_q(D')$の識別が困難

[32] 厳密には、IND-CCA（indistinguishability under chosen ciphertext attack）と呼ばれる対話ゲームとして数学的に定式化される。

であれば、出力値から得られるxに関する情報は制限される。端的にいえば、2 つの確率分布の差分から、xの情報を抽出することは困難である。これが、差分プライバシーの考え方である。差分プライバシーの「差分（differential）」は、ある個人のデータが含まれるデータベースと、そうでないデータベースの差分に着目することに由来する。厳密な定義は、以下のとおりである。以下では、個人データベースDがあり、その各要素が個人 1 人分のデータであると想定する。まず、個人データベースの隣接性を定義する。

定義 2 つの個人データベースD、D'が「隣接する（adjacent）」とは、これらが高々 1 人分のデータしか異ならないことを指す。

統計値を求める統計クエリをq、統計値に乱数を付加するランダム化メカニズム（以下、単にメカニズム）をm_qとする。$m_q(D)$は、データベースDに対して統計クエリqを投げ、算出された厳密な統計値にメカニズムm_qによって確率的なゆらぎを付加した統計値を意味する。

定義 あるクエリqが与えられたもとで、メカニズムm_q: $\mathcal{D}^n \to \mathcal{R}$が「$\epsilon$-差分プライバシーを満たす」とは、ある正の定数$\epsilon$が存在して、任意の隣接するデータベースの組$D \sim D'$および、任意の統計値の集合$S \subseteq \mathcal{R}$に対して

$$\Pr[m_q(D) \in S] \leq \exp(\epsilon) \times \Pr[m_q(D') \in S]$$

を満たすことである。

上記の定義は、$m_q(D)$と$m_q(D')$の確率分布が識別できない（ϵ-識別不可能である）ことを厳密に定式化したものである。ϵに応じた不等式評価の解釈は以下のとおりである：

- $\epsilon = 0$の場合、右辺$\exp(\epsilon) = 1$となり、$m_q(D)$と$m_q(D')$の確率分布が完全に一致する。統計値がデータベースDに依存しないため、プライバシーは完全に保護（完全秘匿）されるが、統計値としての価値は失われる。
- $\epsilon = +\infty$の場合、右辺$\exp(\epsilon) = +\infty$となり、ある 1 人のデータが変動しただけで、統計量の確率分布が大きく変化することを許容する。統計値の有用性は最大となるものの、その個人のデータの部分情報を統計値から確定的に推定しうるため、データの秘匿性は全く保証されない。
- $0 < \epsilon < +\infty$の場合、適度なプライバシー保護を達成する。ϵの値が小さいほど、より高いレベルのプライバシーを保護することが可能となる反面、統計値の有用性は低下する。

上記の差分プライバシーを緩和した、(ϵ, δ)-差分プライバシーに基づく定義（Dwork *et al.* [2006]）

$$\Pr\left[m_q(D) \in S\right] \leq \exp(\epsilon) \times \Pr\left[m_q(D') \in S\right] + \delta$$

も利用される。この定義は、確率δでϵ-差分プライバシーが破れることを許容するものと解釈できる。

差分プライバシーは、任意の 1 レコード分のデータによる情報の流出量の最悪ケースでの評価に基づいて定義されており、かつ、特定の攻撃モデルを仮定しないため、保守的かつ強力な安全性基準である。その重要な性質の 1 つは、差分プライバシーを満たすメカニズムの出力データからどのような関数値を計算しても、元のデータベースに関する追加的な情報なしに、安全性を引き下げることはできない、という以下の定理の主張である：

事後処理定理（post-processing theorem、Dwork and Roth [2014]） メカニズムm_q: $\mathcal{D}^n \to \mathcal{R}$が$(\epsilon, \delta)$-差分プライバシーを満たすとき、任意の写像$f: \mathcal{R} \to \mathcal{R}'$に対して、$f \circ m_q$も$(\epsilon, \delta)$-差分プライバシーを満たす。

また、差分プライバシーの定義により、相異なるレコードがu個存在するデータベースについては、$u\epsilon$-差分プライバシーが保証されることが簡単に導かれる。すなわち、サイズuのデータの集合に対しても差分プライバシーが保証される。この場合、group privacy と呼ばれる。これによると、レコードのデータ同士の相関が強い場合には、そうでない場合に比べて、保証されるプライバシーの水準が弱まることがわかる。差分プライバシーの要件の充足を理論的に評価する際には、各レコード・データが独立であることを暗に仮定することが多い。もっとも、データ・レコードが従う真の確率分布が想定したものと異なり、レコード・データの独立性の仮定が正しくなかった場合には、実際に保証される安全性は理論的な評価よりも低くなってしまう。同一人物の影響が複数のレコードにわたり、複数のレコード・データが互いに強く相関しうるような実用的なデータベースでは、実態的にどの程度の差分プライバシーが保証されうるかが問題となる[33]。

（３）ϵの解釈

定数ϵはプライバシー保護の強さの基準と解釈できる。ϵが小さいほど、出力データから得られる個人の情報量が少ない。ϵの値は、統計の提供者が外生的に決める必要がある。実務的には、$\epsilon = 0.1$から 1 桁程度に設定することが多い。もっとも、どの程度ϵが小さければ実際に安全といえるかについては、現在のと

[33] 例えば、同一人物の検索履歴のデータがタブレット端末と PC から別々に収集された場合には、2 つのデータは強く相関する可能性が高い。

ころ合意は存在せず、政策的に判断されるべきもの（policy maker's choice）[34]と考えられている。

差分プライバシーが表す安全性の基準は、データベースの隣接性の定義に依存する。平たくいえば、差分プライバシーは安全性の強さを測る定規であり、定規の目盛はデータベースの隣接性により決まる。ϵは目盛で測った値であり、目盛幅が異なれば同じ値でも異なる安全性を指すことになる。より厳密には、あるレコードx、x'のみが異なる隣接データベース$D \sim D'$において、x、x'の異なり方に依存して、（ϵの値が同一であっても）保証されるプライバシーの意味合いが変化する。Dwork *et al.* [2006]は隣接性を、ある個人のデータの存在の有無の違い、すなわち、メンバーシップ（membership）であると解釈している。この場合、D、D'は、ある個人のデータを含むデータベースと、含まないデータベースと定義される。もっとも、個人データベースの性質や用途によっては、メンバーシップ（ある個人のデータが含まれるか否か）が暴露されても、データの値のみが漏洩しなければ問題ない場合もある。この場合には、メンバーシップに関する情報の保護は過剰なプライバシー保護の目標となる。隣接性は、応用事例に即して定義されるものであり、この定義に応じてϵの意味が変化する。

差分プライバシーを満たすメカニズムの組は、差分プライバシーを満たすことが、直列合成定理（composition theorem）により保証される[35]。

直列合成定理：メカニズム$m_1, m_2, \dots, m_k$が、それぞれ$\epsilon_1, \epsilon_2, \dots, \epsilon_k$の差分プライバシーを満たすとき、これらのメカニズムの組は、$(\sum_{1 \le i \le k} \epsilon_i)$-差分プライバシーを満たす。

直列合成定理は、(ϵ, δ)-差分プライバシーでも成立する。この定理は、クエリの種類と回数が増すとその都度ϵが加算されてプライバシー保護の度合いが下がることも示唆している。こうした弱点があるため、実務上は、ϵの値に応じてクエリの許容回数に上限を設ける必要がある。データベース全体で保持すべきプライバシー保護の強さ ϵ を予算に見立て、それぞれのクエリに割り当てるため、ϵはプライバシー予算（privacy budget）とも呼ばれる。

---

[34] ただし、経済学的な最適化の観点から、望ましい ϵ の水準にアプローチする研究も見られ始めている。詳細は、John M. Abowd, and Ian M. Schmutte, "An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choice", *American Economic Review*, 109(1), 2019, pp.171–202, を参照。

[35] 例えば、ある個人データベースに対して、それぞれに差分プライバシーを満たす 2 つのメカニズムから統計値を出力する場合、2 つのメカニズムのペアを、統計値のペアを出力する 1 つのメカニズムとみなせる。このとき、このメカニズムも差分プライバシーを満たす。

（４）メカニズムの設計

差分プライバシーは、プライバシー保護の基準に過ぎない。そのため、プライバシー予算ϵと統計クエリqが与えられたとき、ランダム化メカニズムm_qをどのように設計するかという問題が残る。

良いランダム化メカニズムの条件は、「高い安全性基準の差分プライバシーを満たしつつ、有用性の高い統計量を得られる」ことである。メカニズムの有用性は、ランダム化した出力データ$m_q(D)$と、真の値$q(D)$との近さ

$$\Pr\left[\left\|q(D) - m_q(D)\right\| > g(n)\right] \leq \beta$$

で評価できる。$g(n)$はサンプルのサイズnに対してどの程度速く有用性が高まるかを表す関数である。両者の近さが有用性の程度の決定要因となるため、その近さを効用（utility）と呼ぶ。ここでの有用性は統計クエリqに依存する概念である。一般に、有用性とプライバシー保護の強さにはトレード・オフがある。より良いランダム化メカニズムを設計しようとすると、各統計クエリqの個別的な性質を考慮することになるが、このような設計方針には汎用性がない。

そこで、単純な統計解析のためのメカニズムの組として、高度な統計解析を可能にするメカニズムを構成するモジュラー・アプローチ（modular approach）が広く利用される。このアプローチの正当性は、差分プライバシーを満たすメカニズムを組み合せたメカニズムもまた、差分プライバシーを満たすという直列合成定理の主張を根拠にしている。これにより、複数のクエリに対応できるメカニズムの設計は、個々のクエリに対応するメカニズムの設計に帰着できる。以下では、個々のメカニズムについて述べる。

イ．大域敏感度によるメカニズム

（イ）大域敏感度の定義

任意の統計クエリに対して、差分プライバシーを満たすノイズを付与する方法を設計する汎用的な方法として、「大域敏感度（global sensitivity）」に基づくものがある。大域敏感度とは、データベースの中で１つのレコードに変化があったときの統計量$q(D)$の変化幅の最大値（最悪のケース）で定義される。

定義（大域敏感度）クエリ$q: \mathcal{D}^n \to \mathbb{R}^d$の大域敏感度は、

$$\mathrm{GS}_q = \max_{D,D':\ D \sim D'} ||q(D) - q(D')||.$$

ここで、dはクエリの出力データの次元を表す。ノルムは、（後述する）ラプラス・メカニズムの場合にはℓ_1ノルム、ガウシアン・メカニズムの場合にはℓ_2ノルムを表す。以降では、ノルムの定義は省略する。

定義により、大域敏感度はデータベース値には依存しない。直感的には、プライバシー保護の観点から、大域敏感度の大きなクエリへの応答に対して、大きなノイズを付加する必要がある。例えば、値域が[0,1]に基準化されたデータベースに対して、平均値を返すクエリの大域敏感度は、レコード数Nを用いて、$1/N$となる。最大値を返すクエリの大域敏感度は1である。したがって、最大値をとるクエリには、より大きなノイズを付加する必要がある。これは、最大値を返すレコードの値をそのまま公開するクエリは、プライバシーが暴露しやすいという直感と符号する。大域敏感度は、あるクエリについて、どの程度公開に慎重になるべきか、を表す尺度であるとも解釈できる。

（ロ）大域敏感度によるメカニズムの設計

大域敏感度を用いて設計されたランダム化メカニズムの代表例は、統計量に特定の確率分布に従うノイズを付加する方法である。

ラプラス・メカニズム（Dwork and Roth [2014]）は、統計量$q(D)$にラプラス分布 [36]に従うノイズを付加し、$m_q(D) = q(D) + \text{Laplace}(\text{GS}_q/\epsilon)$を出力する手法である。ノイズの大きさ（標準偏差）は、大域敏感度に比例する。ラプラス・メカニズムは、差分プライバシーを満たすことが知られている。

定理：クエリqが与えられたとき、$\text{Laplace}(\text{GS}_\text{q}/\epsilon)$に従うノイズを付加するラプラス・メカニズムは、$\epsilon$差分プライバシーを満たす。

ラプラス・メカニズムは、実装が容易であるため、現在でも広く使われている。ラプラス・メカニズムのほかに、ノイズがガウス分布に従うガウシアン・メカニズムも、差分プライバシーを満たす。すなわち、$m_q(D) = q(D) + N(0, \text{GS}_q^2 \cdot \sigma^2)$とすると、一定のパラメータ条件のもとで、メカニズムm_qは、(ϵ, δ)-差分プライバシーを満たす（Dwork and Roth[2014]）。このほか、指数メカニズム（exponential mechanism、McSherry and Talwar [2007]）は、効用関数（utility score）$q: \mathcal{D}^n \times \mathcal{A} \rightarrow \mathbb{R}$を最大化する要素を確率的に返すものである [37]。ある要素$a \in \mathcal{A}$が選ばれる確率を、qの大域敏感度に基づくものとすることで、要素を返すメカニズムが差分プライバシーを満たす。具体的には、$\Pr[a] \sim \exp(\epsilon \cdot q(D, a)/2 \cdot \text{GS}_\text{q})$とする。こ

---

[36] ラプラス分布に従う確率変数$x \sim \text{Laplace(b)}$の確率密度関数は$f(x; b) = \frac{1}{2b}\exp\left(-\frac{|x|}{b}\right)$で表される。確率分布の形状は中央で上向きに尖っている。分散は、$2b^2$である。

[37] 指数メカニズムは、データ値にノイズを加えると効用が大きく変化する場合に利用される。例えば、需要曲線を決定するビッド価格の集合｛1 万円、2 万円、5 万円｝が与えられたもとで、ビッド価格というプライバシー情報を保護しつつ、利益（＝効用）を最大化する最適価格の決定することを考える。価格そのものにノイズを加えるとき、価格を 5 万円に設定する場合には売上は 5 万円になるが、5.1 万円に設定する場合には売上は 0 円になり、利益も大きく減少する。指数メカニズムを利用して、ビッド価格とは独立な集合から価格を選択することで、プライバシー保護と利益の最大化を両立できる。

こで、~は比例関係にあることを表す。大域敏感度によるメカニズムの設計は、１レコードが変化した場合に生じる統計値の変化幅の最悪ケースに基づいてノイズを付加するため、ノイズが過大になるという問題がある。

ロ．局所敏感度の定義

大域敏感度による方法よりも、小さいノイズを付加する方法として、局所敏感度（local sensitivity、Nissim, Raskhodnikova, and Smith [2007]、Dwork and Lei [2009]）に基づくアプローチがある。

定義（局所敏感度）データベースDが与えられたもとで、クエリ$q:\mathcal{D}^n \to \mathbb{R}^d$の$D$における局所敏感度は、

$$\mathrm{LS}_q(D) = \max_{D':\ D\sim D'} ||q(D) - q(D')||.$$

定義により、局所敏感度は大域敏感度よりも小さく、任意の$D \in \mathcal{D}^n$について$\mathrm{LS}_q(D) \leq \mathrm{GS}_q$が成立する。局所敏感度は、データベース$D$の値に依存するため、$D$の値の情報が統計量から漏出するリスクがある。一般に、局所敏感度に比例するサイズのノイズを付加しても、差分プライバシーは満たされない。したがって、局所敏感度を用いたメカニズムを設計する場合には、データベース値の情報が漏出しないよう工夫する必要がある。

Johnson, Near, and Song [2018]は、実用的なSQLクエリに対して、局所敏感度の上界となる変数（elastic sensitivity）を定義し、これに比例するノイズを付加するFLEXメカニズムを提案した。このメカニズムは、複雑な構文のSQLクエリに対するelastic sensitivityの計算を、単純な構文のSQLクエリに対するelastic sensitivityの計算に帰着することにより、幅広いSQLクエリへの対応を可能にしている。ただし、このアプローチは、クエリ自体に変更を加えずに、データベースから出力される返り値のみを変更（ノイズ付加）するものであるため、異常値を除外して平均をとるといった刈込み（clipping）つきのクエリには十分に対応できない。この制約を解消するため、FLEXのアプローチを発展させたものが、５節（3）で紹介するCHORUS（Johnson *et al.* [2020]）である。

Nissim, Raskhodnikova, and Smith [2007]は、局所敏感度の上界を近似する変数（smooth sensitivity）を定義し、これに比例するノイズを加える手法を提案した。中央値を求めるクエリなど、幾らかのクエリに対して高速にsmooth sensitivityを計算できることを示した。

局所敏感度に基づく手法では、差分プライバシーを満たしつつ、高速に動作し、かつ幅広い統計解析に対応できる可用性の高いメカニズムの設計が容易ではない。

（５）局所差分プライバシー

差分プライバシーは、統計量を公開する段階でのプライバシー保護に焦点を当てたものであり、個人データベースを保有する統計の提供者は無条件に信頼する前提を置く。これに対して、局所差分プライバシー（local differential privacy:LDP、Duchi, Jordan, and Wainwright [2013]）は、データを収集する段階でノイズを乗せるアプローチであり、個人データベースの保有者を信頼する前提を要しない。定義は以下のとおりである。

定義 メカニズムm_q: $\mathcal{D} \to \mathcal{R}$ が「ϵ-局所差分プライバシーを満たす」とは、任意のユーザのデータの組$v,\ v' \in \mathcal{D}$、および任意の$S \subseteq R$に対して、

$$\Pr[m_q(v) \in S] \leq \exp(\epsilon) \times \Pr[m_q(v') \in S]$$

を満たすことである。

LDP を満たすメカニズムには、データ値にノイズを付加するものと、データ値を他の値に入れ替えるランダム化応答（randomized response）に基づくものがある（Yang *et al.* [2020]）。前者は、データが連続値をとる場合に適用される。ノイズは、ラプラス分布やガウス分布に従う。後者は、データが離散値をとる場合に適用される。

ランダム化応答は、「あなたは過去に罪を犯したことがありますか？」といった類の機微にふれる質問に対する回答のバイアスを取り除くためのサーベイ手法として、Warner [1965]によって提案された。この手法は、一定の確率で真の値と偽の値を入れ替えるものであり、回答者は自身の回答に対して尤もらしく抗弁する余地（plausible deniability）が残る。

一例として次のメカニズムを考える。回答者は、秘密裡にコインを投げる。表が出れば正直に回答し、裏が出れば、もう一度コインを投げる。その際、表が出れば必ず「はい」と答え、裏が出れば「いいえ」と回答する。この場合には、簡単な計算により$\epsilon = \ln(3)$の局所差分プライバシーを満たすことが示される [38]。また、メカニズムの構成から自明であるが、ある個人が「はい」と回答したとの情報が漏洩した場合でも、当該個人はその回答が真実に基づくものではないと抗弁する余地が残る。他方、調査主体は、回答者が上記のプロトコルに忠実に従って回答するとの仮定のもとで、ランダム化された回答から真の肯定回答者（表が出て「はい」と回答したもの）の割合を統計的に推定できる。

Holohan, Leith, and Mason [2017]は、上述のように回答値が 2 値の場合の最適な（最尤推定量の誤差を最小化する）ランダム化応答を与えた。回答値が 3 通り

---

[38] 条件付き確率$\Pr[\text{回答値}|\text{真の値}]$を考える。$\Pr[\text{Yes}|\text{Yes}] = 1/2 + 1/2 \times 1/2 = 3/4$、$\Pr[\text{Yes}|\text{No}] = 1/2 * 1/2 = 1/4$である。以上より、$\Pr[\text{Yes}|\text{Yes}] / \Pr[\text{Yes}|\text{No}] = 3$となる。同様に、$\Pr[\text{No}|\text{No}] / \Pr[\text{No}|Yes] = 3$である。以上より$\ln(3)$-局所差分プライバシーを満たす。

以上の値をとりうる場合には、一般ランダム化応答（Kairouz, Bonawitz, and Ramage [2016]）が適用できる。

５．差分プライバシーの応用研究と適用事例

差分プライバシーや局所差分プライバシーを満たすメカニズムは数多く提案されている。本節では、これらの手法と、公的統計や、企業によるユーザ統計などへの応用事例をあわせて紹介する。

（１）集計表への応用

国勢調査の人口データやスマートフォンの位置情報に基づく人口・人流データは、階層的な地理単位で集計される[39]。米国の国勢調査では、それぞれの地理単位において人種、性別、民族などの属性別の内訳も示される。数え上げクエリ（counting query、特定の条件を満たすレコード・データ数の算出）へのデータベースの回答をランダム化することにより、これらの集計表に差分プライバシーを適用できる。

イ．トップダウン法

階層的な地理単位で集計するための単純なアプローチは、最も細かい単位の集計データにラプラス・メカニズムを適用してノイズを加えながら、粗い地理単位に集約していくものである。ただし、この方法では、人口のような負にならない集計値が負値をとってしまう可能性があるほか、ノイズの重畳により部分和精度が劣化する。また、米国の国勢調査では、国や州単位では正確な人口が公表される。集計表は、公表済みの公知の事実との整合性を満たすことが望ましい。

こうした問題への解決策として、米国センサス局はトップダウン法（Abowd *et al.* [2019]）を採用した。この方法は、上位階層から下位階層に向かって集計表を再帰的に細分化していく。各階層において、ラプラス・メカニズムなどで集計表をランダム化したあと、制約付きの整数計画問題を解いて集計値を補正する。この制約条件では、（州単位の正確な人口などの）公知の計数との一致、非負制約、部分和と合計の関係などを勘案する。h層ある地理単位の階層のそれぞれに、ϵ/hのプライバシー予算を充てることで、集計表全体ではϵ-差分プライバシーを満たす。この手法の利点は、国家全体の人口など厳密な数値が求められる部分の正確性とプライバシー保護を両立している点である。この反面、５節（1）ロ．で紹

[39] 米国の国勢調査では、国家（nation）、州（state）、郡（country）から最も細かいセンサス・ブロック（Census block）までの階層的な地理単位ごとに集計される。

介するプライブレット法と比較して、トップダウン法は部分和精度に優れない[40]。

ロ．プライブレット法

Xiao, Wang, and Gehrke [2010] は、離散ウェーブレット変換（discrete wavelet transformation）[41]を利用して部分和精度を改善するプライブレット（privacy preserving wavelet: privelet）法を提案した。この手法は、集計表をウェーブレット変換して得られたウェーブレット係数にラプラス・メカニズムによりノイズを付加し、逆変換で戻す。元データを直接的にランダム化する手法と比べて、より小さなノイズで差分プライバシーを満たすため、出力データの有用性が高い。付加されるノイズの分散は、集計値 V に対して、直接的な方法では$\mathrm{O}(V/\epsilon^2)$であるのに対して、プライブレット法では$\mathrm{O}((\log_2(V))^3/\epsilon^2)$である。

プライブレット法は、差分プライバシーを満たし、部分和精度に優れる利点がある。他方、数え上げクエリへの応答が負値を採りうる欠点や、データの疎性を喪失する[42]欠点もある。ランダム化された返り値は負となりうるため、用途によっては、本来的に取りえない負値が集計表に混入することでデータの有用性が下がる惧れがある[43]。ランダム化された集計表の至るところで非ゼロ値が大量に出現し、データの密度が増加すると、データサイズが増大する。例えば人流をリアルタイムで集計処理する場合に遅延が発生する惧れが高まる。

寺田ほか［2015］は、プライブレット法を改良し、非負制約を満たし、データの疎性を維持する手法を提案した。改良法では、Haar ウェーブレット逆変換を施す際に、非負制約を満たすような処理（トップダウン精緻化）を組み込んだ。さらに、本郷ほか［2020］は、トップダウン精緻化の一部の処理を省略（2 分木に沿った処理について、ある分岐点以降を省略する「枝刈り」）することで、計算効率を改善した。

---

40 一般的に、部分和精度の劣化は、狭い区画での集計値を合計して、より広い区画の集計値を算出していくと、ノイズが重畳されることによってもたらされる。

41 主に画像処理に応用される線型変換。フーリエ変換と概念的に似た手法であり、連続関数を、局所的な波であるウェーブレット（wavelet）の和で表現する。Xiao, Wang, and Gehrke [2010]では、Haar 基底を利用する Haar ウェーブレット変換（Stollnitz, Derose, and Salesin [1996]）を採用した。

42 データの疎性の喪失は、元データの多くが疎である（ゼロ値を採る）ため、集計値においても多くの区画でゼロとなるようなデータセットにおいて、ノイズの負荷により集計値が非ゼロ値となることを指す。

43 なお、単純に負値をゼロに修正することは、集計表に正のバイアスが生じるため望ましくない。

（２）局所差分プライバシーの応用

イ．RAPPOR

RAPPOR（randomized aggregatable privacy-preserving ordinal response、Erlingsson, Pihur, and Korolova [2014]）は、2 段階のランダム化応答と Bloom filter（Bloom [1970]）[44]を組み合せることで、局所差分プライバシーを満たしながらユーザから個人データを収集する手法である。Bloom filter は、文字列から数値へのデータ形式の変換と、データ圧縮の役割を果たす。この手法の特長は、数値だけでなく任意の文字列データにも適用できること、および複数回のクエリに対しても頑健にプライバシーを保護できることである。

RAPPOR は、まず、元のデータvを Bloom filter に通し、長さ k ビットの固定長のデータBに変換する。次に、得られたデータをビットごとに 2 段階のランダム化応答で無作為化する。第 1 段階は、各データに対して 1 回のみ実行される恒久的な（permanent）ランダム化応答により、BをB'に変換する。第 2 段階は、クエリごとに再実行される一時的な（instantaneous）ランダム化応答を実行し、B'から長さ k のビット列 S を生成する。ユーザの端末から送られるデータとして、S が中央サーバに送信され、集計に利用される。

恒久的なランダム化応答により、同一クエリの反復により真値 B を割り出す攻撃（averaging attack）からプライバシーを保護できる。データ B に替えて B' が恒久的に利用されるため、攻撃者は RAPPOR の出力から B を確定的に割り出せないことが保証できる。一時的なランダム化応答は、クエリごとに再実行されるため、攻撃者は B'を手がかりに同一ユーザを追跡することが困難になる。すなわち、恒久的、一時的なランダム化応答は、それぞれ長期的、短期的なリスクからプライバシーを保護している。

具体的には、恒久的なランダム化応答では、Bの各ビットi $(0 \le i \le k)$ごとに

$$B_i' = \begin{cases} 1, \text{ with probability } \frac{1}{2}f \\ 0, \text{ with probability } \frac{1}{2}f \\ B_i, \text{with probability } 1-f \end{cases}$$

とする。ここで、B_iは、Bのi番目のビット値を表す。fは回答値を入れ替える確率を表す。一時的なランダム化応答では、すべての桁をゼロで初期化したあと、各ビットi $(1 \le i < k)$ごとに

---

[44] Bloom filter は、固定長の確率的データ構造であり、ある要素が集合の中に含まれるか否かを確率的かつ高速に判定できる。利点としては、メモリの消費空間が少なく、判定の計算量が O(1)と高速である。欠点としては、集合に存在しないデータを存在すると誤判定する可能性がある。

$$\Pr[S_i = 1] = \begin{cases} q, \text{ if } B_i' = 1 \\ p, \text{ if } B_i' = 0 \end{cases}$$

とする。

実証実験では、特定の単語の出現をカウントする数え上げクエリに対する有用性を示された。RAPPOR はオープンソース・プロジェクトである Chromium において実装[45]されている。これを介して、Google 社が提供するブラウザ Chrome に組み込まれ、ユーザの検索エンジンの利用などに関する統計情報の取得に活用されている。

ロ．感染症の接触確認アプリ

Apple and Google [2020]は、COVID-19 による感染症対策の一環として、プライバシーに配慮しながら追跡調査による接触確認を行う仕組みとして Exposure Notification System（ENS）を提案した。この方式に基づく接触確認アプリでは、ローカル端末には濃厚接触の可能性を示す通知が表示できる一方で、アプリの中央管理者（Apple 社や Google 社）からは通知が表示された端末の特定や位置情報の取得などができない。公衆衛生機関のみが、端末の位置情報などは取得できないものの、通知が表示された端末の特定が可能である。

Google and Apple [2021]は、局所差分プライバシー、秘密分散、ゼロ知識証明といったプライバシー保護技術を組み合わせて、通知件数の統計を作成する手法として Exposure Notification Privacy-preserving Analytics（ENPA）を提案した。ENPA では、システムを構成するサーバのうち、統計作成を担うサーバは総検知数などの統計情報以外の一切の個人データを知ることができない。また、これ以外の 2 台のサーバは一切の個人データを知ることができない。

これを実現する仕組みの概要は以下のとおりである。まず、ローカル端末が生成する濃厚接触の可能性に関する複数の指標データ[46]を、離散的なデータに変換する。典型的には、連続値を、ヒストグラムの区間に対応する番号に置き換える操作が想定される。離散的なデータは、2 値ベクトル（0 または 1 の列）の形で表現されている。次に、このデータを、局所差分プライバシーを満たすようにランダム化応答などによりランダム化する。最後に、秘密分散とゼロ知識証明を組み合わせたプロトコル（secret-shared non-interactive proofs、Corrigan-Gibbs and Boneh [2017]）を通じて、ランダム化したデータの集合から統計が作成される。このステップでは、互いに秘密を開示し合わない前提の 2 台のサーバが、それぞれ秘密分散により断片化されたデータを処理する。個々の断片化されたデー

[45] The Chromium Projects design documents, “RAPPOR (randomized aggregatable privacy-preserving ordinal response)” (https://www.chromium.org/developers/design-documents/rappor）

[46] 論文では具体例の明記はないが、例えば、濃厚接触の時間、位置情報などが考えられる。

タ値が真正のものであること（2 台のサーバにそれぞれ送られた断片化データのペアを合計すると 2 値ベクトルに戻ること）を、2 台のサーバが協力してゼロ知識証明する。両者の協力により、このうち 1 台の統計作成を担うサーバは信頼性のある統計を作成できる。このとき、いずれのサーバも、個々の端末のデータそのものを知ることはできない。

（３）SQL データベースと親和性の高い汎用的なフレームワーク

統計量を算出するアルゴリズムそのものを差分プライバシーを満たすように改変するアプローチでは、データ分析を行うために、差分プライバシーに関する高い専門性が要求されるほか、さまざまな統計クエリに対応することが困難であることから汎用性に欠ける。そこで、分析者にプライバシー保護技術を意識させずに分析を実行可能にするための汎用フレームワークの研究 Kotsogiannis *et al.*[2019]、Bater *et al.* [2020]、Wilson [2020]）が進められている。CHORUS（Johnson *et al.* [2020]）は、SQL データベースから出力された回答値を改変（ノイズ付加）する事後処理に加えて、クエリを事前に改変する前処理を許容することにより、後述する刈込み付きのクエリへの対応を可能とする汎用性の高いフレームワークである。このアプローチでは、既存のデータベース・システムに一切の変更を加えずに、差分プライバシーを満たすことができるため、スケーラビリティがある。全体の処理は、クエリの書き換え（query rewriting）、敏感度分析（sensitivity analysis）、ノイズ付加（post-processing）の 3 段階である（図４）。例えば、distance という系列データについて特定の範囲に収まる値のみを対象に平均値を算出するといった刈込み付きのクエリを動作させたい場合には、まず、次のようにクエリを書き換える：

書換え前：SELECT SUM (distance) FROM database

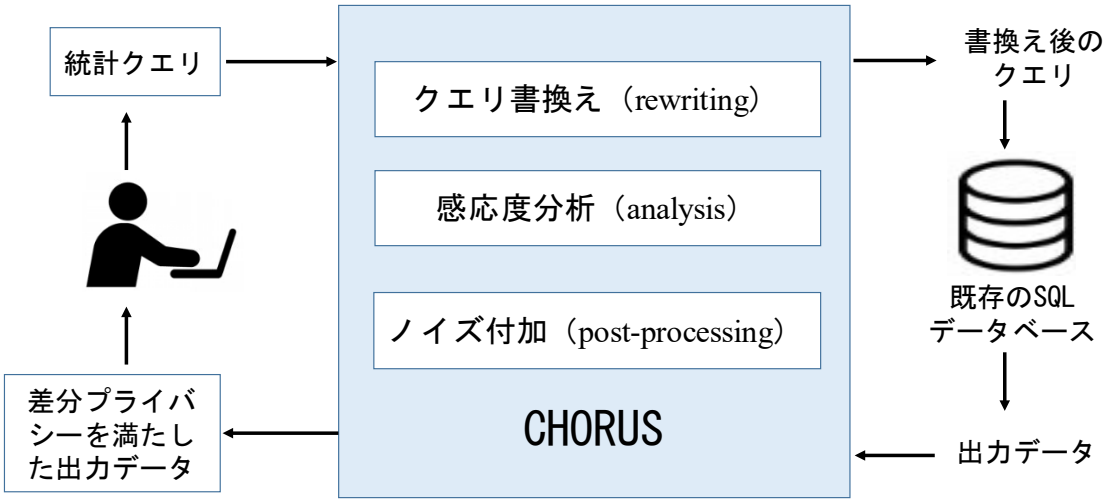
書換え後：SELECT SUM (max(0, min(100, distance)) AS SUM FROM database

次に、感応度分析で算出した大域敏感度$\mathrm{GS} = (u - l) * s$に従って、クエリの返り値にノイズ$\mathrm{Laplace}(\mathrm{GS}/\epsilon)$を付加する。ここで、$u$、$l$はそれぞれ、刈込み条件の上限（100）と下限（0）を表す。sは、刈込みにより除外されるレコード数の上限であり、安定性（stability）と呼ばれる。

CHORUS はオープンソースで開発されており[47]、Uber 社は GDPR に準拠する目的で内部のリサーチに活用している。

[47] https://github.com/uvm-plaid/chorus

図４ CHORUS の概要



（４）機械学習への応用

機械学習モデルの訓練に用いたプライバシー情報が、機械学習モデルの出力を通じて漏出する惧れがある。典型的な攻撃手法として、メンバーシップ推定攻撃（membership inference attack、Shokri *et al.* [2017]）や、訓練データの逆算攻撃（model inversion attack、Fredrikson, Jha, and Ristenpart [2015]）が知られている。こうした脅威に対する防御策として、差分プライバシーを機械学習に組み込む研究も行われている。

Abadi *et al.* [2016]は、深層学習に差分プライバシーを適用する手法を提案した。この手法は、モデル・パラメータと学習アルゴリズムに関する情報を有する強力な攻撃者に対しても有効である。例えば、既知の学習アルゴリズムがモバイル上のアプリケーションに組み込まれ、攻撃者にモデルの内部情報が知られている場合にも適用できる。ϵが 1 桁程度の適度な（modest）プライバシー予算のもとで、高い計算効率と精度（正解率、accuracy）を達成したと評価している。実証実験では、TensorFlow[48]上で数万から数百万個のパラメータの深層学習モデルを訓練した。MNIST[49]と CIFAR-10[50]の公開データセットを用いたベンチマークの

[48] オープンソースで開発されている機械学習のソフトウェア・ライブラリ。2015 年に Google 社によって公開された。

[49] The MNIST database（Modified National Institute of Standards and Technology database）は、手書きの数字の画像と、その数字を表すラベル・データからなる大規模データベース。画像認識の AI 手法の学習用データセットやパフォーマンス計測ベンチマークとして広く利用されている。The CIFAR-10 datase も同様。

[50] The CIFAR-10 dataset は、32×32 ピクセルの物体のカラー画像と、その物体の分類を表す 10 種類のラベルデータからなるデータベース。

画像分類タスクでは、$(8, 10^{-5})$-差分プライバシー（定義は４節（2）を参照）を達成しつつ、それぞれ 97%、73%の正解率となった。

Arachchige *et al.* [2020]は、深層学習に局所差分プライバシーを適用する手法を提案した。この手法では、各ユーザから機械学習モデルにデータを送信する前に、データに乱数を付加するランダム化層（randomization layer）を通す。上記と同様に、MNIST と CIFAR-10 のデータセットを用いた実証実験では、少ないプライバシー予算（$\epsilon = 0.5$）のもとでも、91%～96%の高い正解率を示した。

６．考察

差分プライバシーの企業での応用は拡がっている。本節では、差分プライバシーの制約と限界について述べたあと、プライバシー保護の望ましいあり方について考察する。

（１）差分プライバシーの課題

差分プライバシーの理論研究は概ね成熟しており、今後の課題は社会インフラへの普及である。差分プライバシーは、攻撃モデルに依存しない無条件で成立する安全性を達成する。社会に流通する個人情報が増加していくに従い、攻撃者の背景知識を考慮することが困難になるため、将来的には差分プライバシーを一層活用していくことがより望ましくなっていく。差分プライバシーを満たすメカニズムを機動的に開発することは容易でないため、データ分析者がプライバシー保護技術を意識せずに分析を進められる汎用的なフレームワークの活用は有用な選択肢である。

もっとも、差分プライバシーは万能の処方箋ではない。実用的な個人データベースの各レコード・データが従う真の確率分布は必ずしも明らかではない。このため、この確率分布に一定の仮定を置いたもとで評価した差分プライバシーの安全性基準が、実際には満たされない惧れがある。例えば、レコード・データ同士が強い相関を持つ場合、強いプライバシーは保証できない惧れがある。また、データベースの運用環境に応じて、最善のプライバシー保護方法は変わりうる。差分プライバシーを満たすメカニズムは、さまざまなクエリを受け付ける利用方法を想定することが多い。このため、適切なアクセス制御が行われるもとでの社内利用などであれば、差分プライバシーが最善の選択肢とまではいえない。さらに、どの程度のプライバシー予算があれば安全といえるか、については執筆時点ではコンセンサスは存在しない。

（２）総合的なプライバシー保護措置の必要性

数理的技術と情報技術のみでは、個人データの開示・訂正・削除の請求の仕組

みを提供できない。こうした要求に対処するには、プライバシー・バイ・デザインの概念に基づき、法規制やITシステムの設計段階からプライバシー保護措置をデフォルトで組み込むことが必要となる。セキュリティ・ルール・システムを総動員したプライバシー保護措置により、自己情報のコントロールを達成することが望ましい。

総合的なプライバシー保護措置は、社会的利益と個人の権利保護のトレード・オフを改善するものであり、個人情報を活用するうえでの単なる制約ではない。適切なプライバシー保護措置を導入することで始めて、社会から受容される個人情報の活用法もありうる。とりわけ、デジタル・プラットフォーム提供者や金融機関などの大量の個人データを収集する企業にとって、自社ビジネスが社会から受容されることは必須条件であり、それゆえプライバシー保護は社会の中での企業の社会的意義やこれらに対する規制のあり方に関わる重大な問題となる。また、個人データの活用が一層進めば、事前に想定していなかった新しいプライバシー侵害の脅威が現れる。このような脅威に対して、技術は常に後追いにならざるをえないため、技術以外の対策は必要である。

（3）テクノロジーとの共生と望ましいプライバシー保護のあり方

プライバシー保護の枠組みは重要であるが、個人情報の使われ方にはより大きな注意を払うべきである。個人情報はインターネット時代の石油または通貨などと呼ばれる。個人情報から最大限に価値を引きだすべく、近年では、融資、保険、人事、司法などの分野で、AIと組み合せて個人情報が利用されはじめている。もっとも、便益の最大化、リスクや損失の最小化といった合理的な目標の追及が、必ずしも人類社会に幸福をもたらすとは限らない。AIと個人情報の活用は、公平性やプライバシー保護にまつわる新しいELSIをもたらしている。久木田［2021］は、テクノロジーは使い方次第で良い結果も悪い結果ももたらす、といったテクノロジー中立論は、悪用が容易なAIにはあてはまらない、と指摘した。例えば、個人に関するあらゆる情報が管理・統制される監視社会は、必ずしも望ましいとはいえないが、犯罪や汚職への対策といった合理的かつ善良な目標を追及した結果、こうした社会に意図せず到達する惧れがある。

この一方で、テクノロジーが発達することで、実現できる社会のあり方の選択肢は増えていく。あらゆる個人情報を把握・管理・統制することが潜在的に可能であっても、あえて行わない領域を確保することができる。プライバシー保護は、こうしたELSIへの対処の一環と位置付けることもできる。また、一般に、AIを含めた新しいテクノロジーは、価値観や社会の規範を変容させうる。社会規範の変化に応じて、望ましいプライバシー保護のあり方も変わっていくと考えられるため、社会的な合意の形成を経ながら、望ましいプライバシー保護のあり方

を模索し続ける取組みが不可欠である。

以　上

【参考文献】

五十嵐　大・高橋克巳、「注目のプライバシー Differential Privacy」、『コンピュータソフトウェア』第 29 巻第 4 号、2012 年、40～49 頁

石井夏生利、「アメリカのプライバシー保護に関する動向」、『情報処理』第 55 巻第 12 号、2014 年、1346～1352 頁

岩村　充・西島裕子、「統計データの個票公開とプライバシーの保護―推論制御の理論、その紹介と応用」、『金融研究』第 10 巻第 4 号、日本銀行金融研究所、1991 年、67～93 頁

久木田水生、「人工知能と人間のよりよい共生のために」、『RAD-IT21 WEB マガジン』、2020 年（https://rad-it21.com/ai/kukita-minao_20200317/、2022 年 4 月 7 日）

――――、「人工知能の倫理とその教育」、『信学技報』第 121 巻 119 号、2021 年、50～55 頁

佐久間　淳、『データ解析におけるプライバシー保護』、講談社、2016 年

曽我部真裕・林　秀弥・栗田昌裕、『情報法概説第 2 版』、弘文堂、2019 年

寺田雅之・鈴木亮平・山口高康・本郷節之、「大規模集計データへの差分プライバシの適用」、『情報処理学会論文誌』第 56 巻第 9 号、2015 年、1801～1816 頁

中川裕志、『プライバシー保護入門―法制度と数理的基礎』、勁草書房、2016 年

林　眞子、「信用スコアに関する情報の取扱いとその規律のあり方について」、金融研究所ディスカッション・ペーパーNo.2022-J-4、日本銀行金融研究所、2022 年

本郷節之・寺田雅之・鈴木昭弘・稲垣　潤、「非負精緻化をともなう privelet 法における演算効率化法の性能向上」、『情報処理学会論文誌』第 61 巻 9 号、2020 年、1458～1471 頁

Abadi, Marti'n, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang, “Deep Learning with Differential Privacy,” Proceedings of the 23rd ACM SIGSAC Conference on Computer and Communications Security, 2016, pp. 308–318.

Abowd, John, Daniel Kifer, Brett Moran, Robert Ashmead, Philip Leclerc, William Sexton, Simson Garfinkel, and Ashwin Machanavajjhala, “Census topdown:

Differentially private data, incremental schemas, and consistency with public knowledge," 2019 (available at: https://systems.cs.columbia.edu/private-systems-class/papers/Abowd2019Census.pdf、2022 年 4 月 7 日).

Adam, Nabil R., and John C. Wortmann, "Security-Control Methods for Statistical Databases: A Comparative Study," *ACM Computing Surveys*, 21(4), 1989, pp. 515–556.

Agrawal, Rakesh, and Ramakrishnan Srikant, "Privary-Preserving Data Mining," *ACM SIGMOD Record*, 29(2), 2000, pp. 439–450.

Apple, and Google, "Exposure Notifications: Using Technology to Help Public Health Authorities Fight COVID-19," 2020 (available at https://www.google.com/covid19/exposurenotifications/、2022 年 5 月 11 日).

Apple, and Google, "Exposure Notification Privacy-preserving Analytics (ENPA)," White Paper, 2021 (available at https://covid19-static.cdn-apple.com/applications/covid19/current/static/contact-tracing/pdf/ENPA_White_Paper.pdf、2022 年 5 月 10 日).

Arachchige, Pathum Chamikara Mahawaga, Peter Bertok, Ibrahim Khalil, Dongxi Liu, Seyit Camtepe, and Mohammed Atiquzzaman, "Local Differential Privacy for Deep Learning," *IEEE Internet of Things Journal*, 7(7), 2020.

Bater, Johes, Yongjoo Park, Xi He, Xiao Wang, and Jennie Rogers, "SAQE: Practical Privacy-Preserving Approximate Query Processing for Data Federations," Proceedings of the VLDB Endowment, 13(12), 2020, pp. 2691–2705.

Benjamin, Ruha, *Race after technology: Abolitionist tools for the New Jim Code*, Polity Press, 2019.

Bloom, H. Burton, "Space/Time Trade-Offs in Hash Coding with Allowable Errors," Communications of the ACM, 13(7), 1970, pp. 422–426.

Bonawitz, Keith, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brentdan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth, "Practical Secure Aggregation for Privacy-Preserving Machine Learning," Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017, pp. 1175–1191.

Cavoukian, Ann, "Privacy by Design the 7 Foundational Principles," Technical Report, Information and Privacy Commissioner of Ontario (January 2011, revised version), 2011 (available at https://www.ipc.on.ca/wp-

content/uploads/resources/7foundationalprinciples.pdf、2022 年 4 月 7 日).

————, “Privacy by Design and the Emerging Personal Data Ecosystem,” 2012 (available at https://www.ipc.on.ca/wp-content/uploads/Resources/pbd-pde.pdf、2022 年 4 月 7 日).

————, and Drummond Reed, “Big Privacy: Bridging Big Data and the Personal Data Ecosystem through Privacy by Design,” 2013.

Census Scientific Advisory Committee, “Determining the Privacy-Loss Budget: Research into Alternatives to Differential Privacy,” 2021 (available at https://www2.census.gov/about/partners/cac/sac/meetings/2021-05/presentation-research-on-alternatives-to-differential-privacy.pdf、2021 年 10 月 18 日).

Corrigan-Gibbs, Henry, and Dan Boneh, “Prio: Private, Robust, and Scalable Computation of Aggregate Statistics,” Proceedings of 14th USENIX Symposium on Networked Systems Design and Implementation, 2017.

Dalenius, Tore, and Steven P. Reiss, “Data-Swapping: A Technique for Disclosure Control,” *Journal of Statistical Planning and Inference*, 6(1), 1982, pp.73–85.

Denning, E. Dorothy, “Secure Statistical Databases with Random Sample Queries,” *ACM Transaction on Database Systems*, 5(3), 1980, pp 291–315.

————, *Cryptography and Data Security*, Addison-Wesley Publishing Company, 1982.

————, Peter J. Denning, and Mayer D. Schwarts, “The Tracker: A Threat to Statistical Database Security,” *ACM Transaction on Database System*, 4(1), 1979.

Duchi John C., Michael I. Jordan, and Martin J. Wainwright, “Local Privacy and Statistical Minimax Rates,” Proceedings of 2013 IEEE 54th Annual Symposium on Foundations of Computer Science, 2013, pp. 429–438.

Dwork, Cynthia, “Differential Privacy: A Survey of Results,” Theory and Applications of Models of Computation, Lecture Notes in Computer Science, 4978 Springer Verlag, 2008.

————, and Jing Lei, “Differential Privacy and Robust Statistics,” Proceedings of the 41st annual ACM Symposium on Theory of Computing, 2009, pp. 371–380.

————, Frank McSherry, Kobbi Nissim, and Adam Smith, “Calibrating Noise to

Sensitivity in Private Data Analysis," Proceedings of Theory of Cryptography Conference 2006, Lecture Notes in Computer Science, 3876, Springer, 2006, pp. 265–284,

————, and Aaron Roth, "The Algorithmic Foundations of Differential Privacy", *Foundations and Trends in Theoretical Computer Science*, 9(3–4), 2014, pp. 211–407.

Erlingsson, Úlfar, Vasyl Pihur, and Aleksandra Korolova, "RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response," Proceedings of the 2014 ACM SIGSAC Conference on Computer Communications Security, ACM, 2014, pp. 1054–1067.

Fredrikson, Matt, Somesh Jha, and Thomas Ristenpart, "Model Iinversion Attacks that Exploit Confidence Information and Basic Countermeasures," Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, 2015, pp. 1322–1333.

Garfinkel, Simson, John M. Abowd, and Christian Martindale, "Understanding Database Reconstruction Attacks on Public Data," *Communications of the ACM*, 62(3), 2019, pp. 46–53.

Gennaro, Rosario, Craig Gentry, Bryan Parno, and Mariana Raykova, "Quadratic Span Programs and Succinct NIZKs without PCPs," Proceedings of EUROCRYPT 2013, Lecture Notes in Computer Science, 7881, Springer, 2013, pp. 626–645.

Holohan, Naoise, Douglas J. Leith, and Oliver Mason, "Optimal Differentially Private Mechanisms for Randomised Response," *IEEE Transactions on Information Forensics and Security*, 12(11), 2017, pp. 2726–2735.

Hundepool, Anco, Josep Domingo-Ferrer, Luisa Franconi, Sarah Giessing, Rainer Lenz, Jane Naylor, Eric Schulte Nordholt, Giovanni Seri, and Peter-Paul De Wolf, *Handbook on Statistical Disclosure Control*, Statistics Netherlands, 2010.

————, ————, ————, ————, Eric Schulte Nordholt, Keith Spicer, and Peter-Paul de Wolf, *Statistical Disclosure Control*, John Wiley & Sons, 2012.

Johnson, Noah , Joseph P. Near, Dawn Song, "Towards Practical Differential Privacy for SQL Queries," Proceedings of the VLDB Endowment, 11(5), 2018, pp.526–

539.

Kairouz, Peter, Keith Bonawitz, and Daniel Ramage, “Discrete Distribution Estimation under Local Privacy,” Proceedings of the 33rd International Conference on Machine Learning, 48, 2016, pp. 2436–2444.

Kotsogiannis, Ios, Yuchao Tao, Xi He, Maryam Fanaeepour, Ashwin Machanavajjhala, Michael Hay, and Gerome Miklau,” PrivateSQL: A Differentially Private SQL Query Engine,” Proceedings of the VLDB Endowment. 12(11), 2019, pp. 1371–1384.

Johnson, Noah, Joseph P. Near, Joseph M. Hellerstein, and Dawn Song, “CHORUS: A Programming Framework for Building Scalable Differential Privacy Mechanisms,” Proceedings of IEEE Europen Symposium on Security and Privacy, 2020, pp. 535–551.

Li, Ninghui, Tiancheng Li, and Suresh Venkatasubramanian, “t-Closeness: Privacy Beyond k-Anonymity and l-Diversity,” 2007 IEEE 23rd International Conference on Data Engineering, IEEE, 2007, pp. 106–115.

————, Wahbeh Qardaji, and Dong Su, “On Sampling, Anonymization, and Differential Privacy or, *k*-Anonymization Meets Differential Privacy,” Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security, 2012, pp. 32–33.

Lian, Huanhuan, Tianyu Pan, Huige Wang, and Yunlei Zhao, “Identity-Based Identity-Concealed Authenticated Key Exchange,” Proceedings in European Symposium on Research in Computer Security, Lecture Notes in Computer Science, 12973, 2021, pp. 651–675.

Machanavajjhala, Ashwin, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam, “*l*-Diversity: Privacy beyond *k*-Anonimity,” *ACM Transaction on Knowledge Discovery from Data*, 1(1), 2007.

McMahan, Brendan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas, “Communication-Efficient Learning of Deep Networks from Decentralized Data,” Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, 54, 2017, pp. 1273–1282.

McSherry, Frank, and Kunal Talwar, “Mechanism Design via Differential Privacy,” Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science, 2007, pp. 94–103.

Meyerson, Adam, and Ryan Williams, “On the Complexity of Optimal *k*-Anonymity,” Proceedings of 23rd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, pp. 223–228, 2004.

Nissim, Kobbi, Sofya Raskhodnikova, and Adam Smith, “Smooth Sensitivity and Sampling in Private Data Analysis,” Proceedings of the 39th Annual ACM Symposium on Theory of Computing, 2007, pp. 75–84.

O’Neil, Cathy, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown Press, 2016.

Organisation for Economic Co-operation and Development (OECD), “The OECD Privacy Framework,” OECD, 2013 (available at https://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf、2022 年 4 月 7 日).

Ramacher, Sebastian, Daniel Slamanig, and Andreas Weninger, “Privacy-Preserving Authenticated Key Exchange: Stronger Privacy and Generic Constructions,” Proceedings of European Symposium on Research in Computer Security, 2021, pp. 676–696.

Shokri, Reza, Marco Stronati, Congzheng Song, and Vitaly Shmatikov, “Membership Inference Attacks against Machine Learning Models,” *2017 IEEE Symposium on Security and Privacy*, 2017, pp. 3–18.

Stollnitz, Eric J., Anthony D. Derose, and David H. Salesin, *Wavelets for computer graphics: Theory and applications*, Morgan Kaufmann Publishers Inc., 1996.

Sweeney, Latanya, “k-Anonymity: A Model for Protecting Privacy,” *International Journal on Uncertainty, Fuzziness, and Knowledge-Based Systems*, 10(5), 2002, pp. 557–570.

United States Census Bureau, “A History of Census Privacy Protections,” United States Census Bureau, 2019 (available at https://www2.census.gov/library/visualizations/2019/communications/history-privacy-protection.pdf、2021 年 10 月 4 日).

Warner, Stanley L., “Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias,” *Journal of the American Statistical Association*, 60(309), 1965, pp. 63–69.

Willenborg, Leon, and Ton de Waal, *Elements of statistical disclosure control*, Springer, 2001.

Wilson, Royce J., Celia Yuxin Zhang, William Lam, Damien Desfontaines, Daniel

Simmons-Marengo, and Bryant Gipson, “Differentially Private SQL with Bounded User Contribution,” Proceedings on Privacy Enhancing Technologies Symposium, 2020(2), 2020, pp. 230–250.

Xiao, Xiaokui, Guozhang Wang, and Johannes Gehrke, “Differential Privacy via Wavelet Tansforms,” Proceedings of the 26th International Conference on Data Engineering, 2010, pp. 225–236.

Yang, Mengmeng, Lingjuan Lyu, Jun Zhao, Tianqing Zhu, and Kwok-Yan Lam, “Local Differential Privacy and Its Applications: A Comprehensive Survey,” arXiv:2008.03686, 2020.