Vol.2022-CSEC-97 No.5
Vol.2022-IOT-57 No.5
2022/5/19

プライバシー保護枠組みの全体像と差分プライバシーの有用性と限界

菅和聖 [1]

概要：現代社会では、かつてない細かい粒度の個人情報が継続的かつ自動的に収集できるようになった。個人情報の産業的な価値は高い一方で、データを利用する際にはプライバシー保護も両立することが求められる。もっとも、データの有用性とプライバシー保護の強さにはトレード・オフが存在するため、両者の適度なバランスをとる必要がある。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。2020 年には、米国センサス局が国勢調査に差分プライバシーを採用した。学術的にも、差分プライバシーは、プライバシー保護技術を評価する標準的な尺度となっている。ただし、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からの要請に応えるためには、差分プライバシーを含む数理的手法のみでは対処できず、法律、システムを総動員した対策が求められる。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づけを明確化する。次に、差分プライバシーの理論やその応用研究を解説する。さらに、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。

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

Overview of the Privacy Preserving Framework and Utility and Limitations of Differential Privacy

KAZUTOSHI KAN[†1]

Abstract: In modern society, unprecedentedly granular personal information can be collected persistently and automatically. While personal information is industrially valuable, privacy must be preserved amid the use of the data. In general, there is a trade-off between the usefulness of personal data and the privacy, thus a reasonable balance between the two should be struck. Differential privacy enables to achieve moderate privacy through quantifying the degree of privacy protection. It provides the academic standard for evaluating the privacy-preserving technologies. The U.S. Census Bureau adopted the differential privacy for the 2020 Census. However, mathematical methodologies including differential privacy cannot suffice social demands for privacy, such as the control over personal information, thus a comprehensive approach which also includes laws and IT systems are desirable. This paper gives an overview of the privacy preserving framework and reveals the location of differential privacy in the framework. It also explains the theory of differential privacy and its application studies, and discusses the desirable privacy preserving framework in the society considering the utility and limitations of the differential privacy.

Keywords: Differential Privacy, Privacy Protection, Control over Personal Information, Anonymization, ELSI

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（個人に関する情報）を継続的かつ自動的に収集できる。このため、プライバシー保護の重要性は高まっている。本稿では、プライバシー保護の枠組みの全体像を整理する。さらに、情報理論的な安全性基準を提供する差分プライバシー（differential privacy、Dwork [2006]）の原理を解説し、その応用研究の事例を紹介する。差分プライバシーは、プライバシー保護技術を評価する学術的な標準となっている。

個人情報の利用は、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等による社会システムの効率性向上に寄与する。社会的便益をもたらす一方で、個人情報の利用は､プライバシー侵害の脅威も増大させる。例えば、個人情報は、機械学習（人工知能、AI）と組合せることにより、精緻な個人のプロファイリングを可能とする。ターゲティング広告への応用では悪影響は必ずしも大きくないが、久木田［2020］が指摘するように、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。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; ELSI）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。個人情報のビッグデータを利用する際には、個人の権利保護と社会的利益のバランスを考慮することが求められる。数理的手法は、両者のトレード・オフがあるもとで最善のバランスを実現するうえで有用である。

プライバシー問題を巡る状況は、個人データベースの整備、判断を自動化する AI の普及、匿名加工データの第 3 者

1 日本銀行金融研究所

Vol.2022-CSEC-97 No.5
Vol.2022-IOT-57 No.5
2022/5/19

流通、個票データの利用拡大などによって変化している。第 1 に、個人情報の用途が広範かつ複雑になったため、プライバシー侵害の影響を予見または認識することが難しくなった。これに伴い、プライバシーの概念自体として、「機微な情報の漏洩防止」を拡大した「自己情報のコントロール権」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や要求が高度化したため、これらの達成は、情報セキュリティ分野の技術的課題を超えて、法規制、数理的手法、情報システムの設計なども総動員して対処すべきものとなった。第 2 に、攻撃者がさまざまな情報や計算資源を利用できるため、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。とくに、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 。このため、任意の背景知識をもつ攻撃者に対して有効な情報理論的アプローチの重要性が高まった。

後者の理由により、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） やデータ・スワッピング（data swapping） 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データベース再構築攻撃（database reconstruction attack）の脅威が、理論上の脅威にとどまらなくなったことを挙げた。また、いずれの手法も、特定の攻撃手法のもとでしか安全性を証明できず、攻撃手法によらないプライバシーの保護を達成しない。このほか、Google や Uber 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。

差分プライバシーは、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）の基準であり、プライバシー保護技術を評価する学術的な標準である。また、差分プライバシーの理論研究は成熟しつつある。他方、差分プライバシーには以下のような難点もあり、その実用面での普及は道半ばである。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、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。差分プライバシーの理論が要請する個人データベースに関する前提を、実務データが必ずしも満たさない。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、あらゆる状況で有望な選択肢とまでは言えない。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認識しておくことが求められる。

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。さらに、差分プライバシーを含む数理的手法を概観する。3 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。4 節では、応用研究を紹介する。5 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枠組み

プライバシー保護の枠組みは、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である。本節では、これらを便宜的に、「原理」、「ルール」、「手段」の３つのカテゴリに分類する（図１）。「原理」は基本的な考え方であり、プライバシーの概念やその保護のあり方を規定するものである。「ルール」は国内法や国際的な取り決めを表す。「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 IT システム上で処理される現代では、情報技術と数理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。

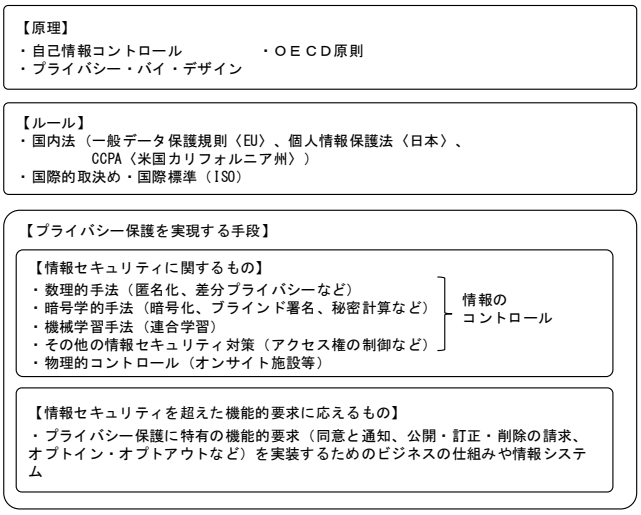

図 1　プライバシー保護枠組みの全体像

2.1 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）」、または「追跡拒否権（do not track）」が含まれうるものとなった。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）という考え方も提唱されている。

1980 年に制定された 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。この理念は、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・デザインはで 7 原則から成り、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。

2.2 ルール

ルールは、国内法、国際的な取り決め、国際標準、業界団体による自主規制などから成る。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a。「プライバシー権」は憲法 13 条に定める幸福追求権に基礎づけられる人格権の 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、プライバシー等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等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排除するのではなく、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。

2.3 プライバシー保護を実現するための手段

(1) 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 。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、機械学習手法に主として分けられる。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。

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 。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

a 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。

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）、秘密計算（secure computation） 、秘密分散（secret sharing）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 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 や Encrypted Server Name Indication）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。

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。機械学習手法として、秘密計算と似た状況のもとで、機械学習モデルを訓練する連合学習（federated learning）も提案されている。連合学習は、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する場合に利用される。各企業は、互いに個人データベースの情報を開示しあうことはない。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。

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。

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「オンサイト施設」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、カメラ等による監視のもと、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。

(2) 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 IT 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、IT 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、実装されなければならない。

2.4 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、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。端的な利用の拡大例は、個人データを自動処理する機械学習システムによる判断の自動化、名寄せによる個人データベースの整備、企業や国境を跨ぐ第 3 者への個人データの提供・共有である。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、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。

2.5 プライバシー保護のための数理的手法

数理的手法は、統計学、データベース、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。表１のとおり、これらの数理的手法によるプライバシー保護の安全性の原理は、それぞれ異なる。

	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	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	利用者	仮名から人物を特定できない（決定論的）
k-匿名化	利用者	同一IDを持つk人を区別できない
合成データ	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	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	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	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。

表 1　数理的手法と安全性の原理

3. 差分プライバシーの理論

3.1 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。こうし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。個人データベースに含まれない外部情報（背景知識）の脅威は増している。これに加えて、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。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。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も普及している。個人データベースの集計表は、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 。

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。複数の統計表を組合せると、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。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。

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手法は、任意の背景知識をもつ攻撃者に対して理論的なプライバシー保証を与える。その利点は、米国センサス局が懸念する再構築攻撃に加え、想定外の攻撃に対しても安全性を保証できる堅牢性の高さである。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言えず、どんな状況でも利用すべき万能薬ではない。

3.2 差分プライバシーの定義

差分プライバシーの定義は、以下のとおりである。まず、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。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$m_q\colon \mathcal{D}^n \to \mathcal{R}$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。

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。

差分プライバシーは、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、任意の背景知識をもつ攻撃者に対して有効であるため、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（事後処理定理、post-processing theorem）というものである。

また、差分プライバシーの定義により、相異なるレコード数がu個の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（group privacy）が保証される。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、保証されるプライバシーの水準が弱まることが分かる。差分プライバシーの定義では、各データ・レコードは独立であることが暗に仮定されているため、同一人物の寄与が複数のレコードにわたり、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。

3.3 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、ϵ =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言えるかについては、現在のと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）と考えられている。

この情報量の単位は、データベースの隣接性の定義により決まる。すなわち、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、保証されるプライバシーの意味合いが変化する。Dwork [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。したがって、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した。もっとも、メンバーシップが暴露されても、データの値のみが漏洩しなければよい場合もある。この場合には、メンバーシップの保護は過剰となる。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合せは、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。

直列合成定理　メカニズム$m_1, m_2, \ldots, m_k$が、それぞれ$\epsilon_1, \epsilon_2, \ldots, \epsilon_k$の差分プライバシーを満たすとき、これらの出力データの組合せは、$(\sum_{1\le i\le k}\epsilon_i)$-差分プライバシーを満たす。

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を意味する。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。

3.4 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。

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、ランダム化応答（randomized response）に基づくものがある。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。

4. 差分プライバシーの応用研究

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、公的統計や、企業によるユーザ統計などへの差分プライバシーの応用研究を紹介する。

4.1 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 。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データ・レコードの数の質問）への応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。

(1)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、粗い地理単位に集約していくものである。この方法では、集計値の非負制約を逸脱する 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、次節で紹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。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で誤差が大きくなることを指す。

(2)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） 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: privlet）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$O(V/\epsilon^2)$であるのに対して、プライブレット法では$O((\log_2(V))^3/\epsilon^2)$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の非負制約の逸脱や、データの疎性を喪失する 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 。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。

4.2 局所差分プライバシーの応用

RAPPOR （ randomized aggregatable privacy-preserving

ordinal response、Erlingsson, Pihur, and Korolova [2014]）は、2 種類のランダム化応答と Bloom filter を組合せることで、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、主としてデータ圧縮の役割を果たす。この手法の特長は、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に変換する。次に、得られたデータをビットごとに以下の 2 段階のランダム化応答で無作為化する。

恒久的なランダム化応答により、真値BをB'に置き換えて、これを恒久的に利用する。これにより、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。攻撃者は RAPPOR の出力からBを確定的に割り出すことができない。一時的なランダム化応答は、クエリ毎に再実行される。これにより、攻撃者にとって、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。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 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。

4.3 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が進められている。
CHORUS（Johnson et al. [2020]）は、SQL データベースへのクエリを改変することにより、差分プライバシーを満たす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ほか、数え上げ以外の幅広いクエリにも対応できるため、スケーラビリティがある。

4.4 機械学習への応用

差分プライバシーは、機械学習などのより複雑な分析手法に応用される。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恐れがある。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）や、訓練データの逆算攻撃（model inversion attack）が知られている。

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の知識を有する強力な攻撃者に対しても有効である。例えば、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。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 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 と CIFAR-10 の公開データセットを用いたベンチマークの画像分類タスクでは、$(8, 10^{-5})$-差分プライバシーを達成しつつ、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。少ないプライバシー予算（ϵ =0.5）の下でも、91%～96%の高い精度を示した。

5. 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。

5.1 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は、必ずしも差分プライバシーの理論が置く仮定を満たすとは限らない。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言えない。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

5.2 総合的なプライバシー保護措置の必要性

数理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、社会的便益と個人の利益保護のトレード・オフを改善するものであり、個人情報を活用する上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、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である。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得ないため、技術以外の対策は必要である。

5.3 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（oil）または通貨（currency）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・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言えないが、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る。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、社会的な合意の形成を経ながら、模索し続ける取組みが不可欠である。

謝辞 本稿の作成に当たっては、菊池浩明氏（明治大学）および山本慶子氏（日本銀行）から有益なコメントを頂いた。謹んで感謝の意を表する。また、ありうべき誤りはすべて筆者個人に属する。

参考文献

[1] 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
[2] 久木田水生、「人工知能の倫理とその教育」、電子情報通信学会 SITE シンポジウム「データサイエンスの ELSI と教育」、2021 年
[3] 曽我部真裕・林秀弥・栗田昌裕、『情報法概説第 2 版』、弘文堂、2016 年
[4] Abadi, Marti'n, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, Li Zhang, "Deep learning with differential privacy," Proceedings of the 23rd ACM SIGSAC Conference on Computer and Communications Security, 2016, pp. 308-318.
[5] 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
[6] 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, Volume 7(7), 2020
[7] Cavoukian, Ann, “Privacy by design the 7 foundational principles,” Technical report, Information and Privacy Commissioner of Ontario (January 2011, revised version).
[8] Duchi C. Duchi, 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.
[9] Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith, “Calibrating Noise to Sensitivity in Private Data Analysis,” Proceedings of Theory of Cryptography Conference 2006, pp. 265-284, Springer, 2006.
[10] 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, pp. 1054-1067, 2014.
[11] 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, pp. 535-551, 2020.
[12] OECD, “The OECD privacy framework,” 2013
[13] United States Census Bureau, "A history of census privacy protections," 2019
[14] Xiao, Xiaokui, Guozhang Wang, and Johannes Gehrke, "Differential privacy via wavelet transforms," Proceedings in 26th International Conference on Data Engineering, pp. 225-236, 2010.