プライバシー保護技術の新展開

南　和宏
（統計数理研究所データ科学研究系教授）

データ駆動社会において、データ利活用は各方面で急速に進み、そのための分析技術はますます高度かつ緻密化している。そのようなデータ利活用は我々の社会に有益な知見をもたらす一方、個人情報や企業秘密の情報漏えいのリスクを高める諸刃の剣である。

この問題には公的統計も以前から取り組んでおり、公表される統計表の秘匿手法やミクロデータが安全かつ有用に利用されるため提供形態や関連する技術が研究されてきた（伊藤(2016)）。英国国家統計庁がバーチャルミクロデータラボ（オンサイト施設）を開設するにあたり2003年に考案されたデータアクセスの５つの安全モデルFive Safes Model（表）は、ヨーロッパ等における行政記録情報の活用やデジタル経済に在り方にとっての参照モデルになるなど、公的統計の取り組みはデータ駆動社会に対して影響を与えている（Ritchie (2017), 伊藤 (2020)）。

近年、様々な情報の組み合わせを大量かつ高速に高度に扱えるようになってきたことにより生じうる情報漏えいリスクが米国で指摘されるようになり、米国センサス局のAbowd（2019）は、プライバシー保護を単なる情報科学、統計学の技術的な問題ではなく経済学の問題と位置づけ、データの利活用とプライバシー保護の相反する要件のトレードオフの決定問題と定式化している。そして、公的統計のプライバシー保護の状況は新たな転換点を迎えつつある。

本特集では、プライバシー保護において大きな潮流となりつつある攪乱的な秘匿処理技術とその安全性基準である「差分プライバシー」を主要な話題とし、公的統計の分野におけるプライバシー保護に関する最新の動向を海外の事例を含めて解説する。さらに差分プライバシーがもたらす公的ミクロデータの２次的利用に関する今後の課題を制度面、技術面から幅広く議論し、様々な利用が期待される合成データへの適用の可能性を紹介する。

以下、各論文（順不同）の概要を紹介する。

槙田論文は、公的統計におけるプライバシー保護に関する法制度を紹介し、データの作成、利用に関わるステークホルダーを整理したうえで、国連のタスクチームによるプライバシー保護技術の位置づけを明らかにする。特に、2020年の米国センサスで実施された差分プライバシーの「識別不可能性」が重要な概念であること

安全なプロジェクト	このプロジェクトのためにデータを使用することは適用か	運営上の管理
安全な利用者	利用者は適切に利用するという信用に足るか	
安全な施設	データアクセス施設は認められない利用に制限をかけるようになっているか	
安全なデータ	データ自身に開示リスクはないか	統計上の管理
安全な成果物	統計的成果物が開示になってしまわないか	

を指摘し、日本における適用可能性について、その方向性と課題を論じている。

南論文は、公的統計の分野においてこれまで主流であった非攪乱的な秘匿処理の安全性に関する問題点をk-匿名化の事例で紹介し、攪乱的な秘匿処理による差分プライバシーの概念とその優位性を解説する。また公的統計の二次的利用においては、プライバシー予算の管理、安全性審査が課題であることを指摘し、多目的の分析に資する合成データへの適用が有望であることを論じている

伊藤論文は、ヨーロッパ諸国における匿名化措置の動向を紹介し、特にイギリス国家統計局の事例を詳しく説明している。近年は攪乱的な手法も取り入れられている状況に触れ、オーストラリア統計局のオンデマンド集計のシステムTableBuilderのノイズ付加手法や現在欧州統計局においても差分プライバシーの導入が検討されている状況を紹介する。また匿名データの拡張可能性に関して、法解釈や匿名化処理の基準の変更の可能性を議論している。

寺田論文は、2020年米国センサスへの差分プライバシー適用の事例を詳しく解説する。その背景にあったモザイク効果に対する再構築攻撃によるプライバシーリスクを説明し、地理空間上の大規模な集計表間で整合性を保つための差分プライバシーの適用方式を紹介する。またデータの有用性とプライバシー保護のトレードオフを決めるプライバシー損失予算の値についての妥当性を技術面、社会的な面から議論する。

千田論文では、ビジネス展開が進むプライバシー保護型の合成データについて、技術と法規制の両面をふまえた実用動向を多くの事例とともに紹介する。合成データの生成手法として、ベイジアンネットワークなどの機械学習モデルを作成して生成する手法、深層学習モデルによる手法を紹介し、安全性検証については、差分プライバシーによる理論的な保証と合成データに対する実証的にリスク評価の２つに整理する。今後の課題として、差分プライバシーの予算パラメータの基準値の設定、リスクアセスメント手法のコンセンサスの形成、法規制との関係整理等を挙げている。

このように本特集では、各界の専門家がプライバシー保護の最新動向について、それぞれの立場の知見を述べた。その中で「差分プライバシーの展開」が共通のメッセージとして現れ、その概念、応用事例等が多面的に解説された。本特集が、公的統計の分野におけるプライバシー保護の現状を整理し、今後の方向性を考える一助となれば幸いである。

＜参考文献＞

Felix Ritchie（2017）, The ‘Five Safes’: a framework for planning, designing and evaluating data access solutions, Data for Policy 2017, London, 6-7 September 2017 https://doi.org/10.5281/zenodo.897821

John M. Abowd（2019）, The U.S. Census Bureau Tries to Be a Good Data Steward in the 21st Century, 9th Annual FDIC Consumer Research Symposium, October, 2019.

伊藤伸介（2016）「諸外国における政府統計ミクロデータの提供の現状とわが国の課題」, 中央大学経済研究所年報第48号

伊藤伸介（2020）「諸外国における公的統計と行政記録データの二次利用に関する展開方向」『経済学論纂（中央大学）』第61巻第２号