差分プライベートな国勢調査データの有用性に関する定量的な評価研究

伊藤 伸介†
寺田 雅之††
加藤 駿典†
松井 秀俊††

Empirical Assessment of Potential Applications for Differential Private Data from the Japanese Population Census

ITO Shinsuke
TERADA Masayuki
KATO Shunsuke
MATSUI Hidetoshi

アメリカセンサス局は、公表された人口センサスに対する「データベース再構築攻撃」に伴う個体情報の特定化のリスクへの対策として、2020 年人口センサスにおいて、Top down アルゴリズムによる差分プライバシーの方法論を適用した。こうした状況を踏まえ、わが国の国勢調査に対する差分プライバシーの方法論の適用可能性を追究するために、本稿では、令和 2 年国勢調査の個票データを用いて作成した集計表をもとに、各種の差分プライバシーの実現手法を適用した場合の有用性を定量的に評価した。本稿においては、特定の地域区分において対象となるすべての変数でクロス表を作成し、差分プライバシーに基づきノイズを付与した場合と、より上位の区分において按分した上で、差分プライバシーに基づきノイズを付与した場合の結果を比較している。本研究の成果によれば、上位の地域区分において調査項目の分布特性に基づいて按分を行った場合、差分プライバシーの実現方式によっては、作成された集計表に関する平均絶対誤差 (MAE)は、対象となるすべての変数を用いて作成された同様の集計表と比較して小さくなる場合もあることが実証的に確認された。

キーワード: 差分プライバシー、アメリカ人口センサス、国勢調査、生態学的誤謬、有用性

The U.S. Census Bureau has used differential privacy based on a top-down algorithm in the 2020 Population Census as a countermeasure against the risk of identifying individual information due to a "database reconstruction attack". In order to investigate the applicability of differential privacy methodology to the Japanese Population Census, this paper quantitatively evaluates the usefulness of various differential privacy implementation methods when applied to the tabulation tables created using individual data from the 2020 Population Census. Specifically, we compare the results of noise addition to cross tables created with all target variables in a specific regional category and after proportionally dividing the data in higher-level categories. The results of this study empirically confirm that when prorating based on the distributional characteristics of the survey items in the higher-level regional categories, the mean absolute error (MAE) for the generated tabulations may be smaller than for similar tabulations generated using all the target variables, depending on how differential privacy is applied.

Keywords: differential privacy, U.S. Census, Japanese Population Census, ecological fallacy, usefulness

† 中央大学経済学部 Email: ssitoh@tamacc.chuo-u.ac.jp
†† (株)NTT ドコモ/京都橘大学 Email: teradam@nttdocomo.com
† 総務省統計研究研修所 Email: skato@nstac.go.jp
†† 滋賀大学データサイエンス学部 E-mail: hmatsui@biwako.shiga-u.ac.jp

1. はじめに

公的統計におけるプライバシー保護のための秘匿技法として、海外の統計作成部局は、攪乱的手法(perturbative methods)を積極的に採用してきた(伊藤・寺田(2023))。ヨーロッパでは、2021 年人口センサスの作成・公表にあたって、攪乱的手法としての cell key method[1]とターゲット・スワッピング(targeted data swapping)の実用化が追究されてきた。例えば、イギリス国家統計局(Office for National Statistics=ONS)が開発した「オンデマンド型公表システム(Flexible Dissemination System)」は、2021 年人口センサスの作成・公表において cell key method の実用化を追究したシステムである(Office for National Statistics(2017))。そして、ONS は、2021 年人口センサスに関して、「Create a custom dataset」という web サイトで、オンデマンドによる多次元集計表の提供サービスを開始した。このサービスでは、集計項目の数やそれに含まれるカテゴリーの数に関する利用者の要求に応じて、多次元集計表の作成・提供が可能になっている。そのために、オンデマンド型公表システムでは、ターゲット・スワッピングが施された個票データを元データとした上で、cell key method についてはデータの有用性を重視する形での攪乱(“a light touch cell key perturbation”)が実施されてきた。

一方、アメリカセンサス局(以下「センサス局」と呼称)は、人口センサス(以下「センサス」と略称)について個体情報のプライバシーの暴露に対する懸念を有していた。そのため、アメリカ国民およびアメリカ経済に関するデータを効率的かつ合理的に収集・利用することを保証することを任務とする、センサス局の DSEP(=Data Stewardship Executive Policy Committee)が、2020 年センサスに関して、2010 年センサスで適用されたスワッピングとは異なる新たな方法論の適用を模索していた。それは、センサスデータに対して差分プライバシー(differential privacy) (Dwork(2006))の実現方式の適用可能性(Jamin(2021))を追究することであった。

差分プライバシーは、データベース再構築攻撃(database reconstruction attack)(Abowd(2018))への対策としてセンサス局で検討されてきた。データベース再構築攻撃(あるいは再構築攻撃)とは、あるデータベースから生成された (一見して安全に見える) データを重ね合わせることによって制約充足問題を構築し、その問題を解いて元のデータベースを復元することにより、データに含まれる個人のプライバシーを暴露する攻撃である。

センサス局は、2020 年センサスの公表統計表に差分プライバシーを適用するにあたって、2010 年のセンサスデータを用いて差分プライバシーの実用性に関する検証を行った(伊藤・寺田(2020), 伊藤他(2022))。具体的には、センサス局が採用した TopDown アルゴリズムに基づき、統計表の公表によって消費されるプライバシー損失予算(privacy loss budget)ε を設定し、地域のレベルにおけるパラメータ ε の適切な割り当てに関する検証を進めてきた(Garfinkel et al.(2019))[2]。これに関しては、統計数値の秘匿性の観点だけでなく、データの利用者や利害関係者が要求する統計数値の精度も考慮した上で数度の修正がなされたが、最終的には、DSEP が、2020 年センサスの統計表を公表する上で求められるパラメータ ε の数値の決定を行った

---

[1] 攪乱的手法としての cell key method の特徴については、伊藤・寺田(2023)を参照。

[2] センサス局が実施した TopDown アルゴリズムでは、以下のような手順で統計表を作成し、差分プライバシーの方法論が適用される。最初に全国レベルで集計を行い、数理的に最適化されたプライバシー損失予算 ε に基づいてノイズを付与した上で、差分プライベートな統計表が作成される。つぎに、州のレベルにおいて、プライバシーの強度とデータの有用性の両面を考慮した上で、ノイズを付加された差分プライベートな統計表が作成される。以下、同じような形で、郡レベル、センサストラクトレベル、センサスブロックレベルの順に、階層的な地域区分に沿った形で差分プライベートな統計表を作成する(伊藤・寺田(2020))。

(伊藤他(2022))。具体的には、2021 年 6 月に公開された最終版のプライバシー保護済ミクロデータファイル(Privacy-Protected Microdata Files=PPMFs)の作成のために、全体のプライバシー損失予算のパラメータ ε は 19.61 に設定された。このパラメータに基づき、2020 年センサスを対象に、差分プライバシーの実現方式が適用された統計表(区画改定データ(PL94-171))が 2021 年 8 月に公表された。さらに、一般公開型ミクロデータ(public use microdata)である差分プライベートな PPMFs が 2024 年に公開されている。

伊藤他(2024)は、わが国の公的統計における差分プライバシーの適用可能性を追究するために、平成 27 年国勢調査の個票データを用いて、地理的区分が異なる統計表に各種の差分プライバシーの実現方式を適用した場合の有用性に関する定量的な評価を行った。これについては、評価指標として平均絶対誤差 (mean absolute error=MAE)を用いた場合の有用性の比較・検証を行った。本研究からは、国勢調査に差分プライバシーを適用するにあたって、最上位の地域区分でノイズを生成し、トップダウンで調整を図りながら統計表の各セルにノイズを割り当てる場合では、最小地域区分ごとの集計表のセルにノイズを付与し、畳み上げて集計した場合よりも、相対的に精度の高い数値が得られることが明らかになった。しかしながら、本研究は、地域区分の粒度のみを比較の対象にしていたことから、より高次元で差分プライベートな統計表の有用性についてもさらなる検証が必要だと考える。

本稿は、令和 2 年国勢調査の個票データをもとに、差分プライバシーの実現手法を適用した集計表を対象に、その有用性に関する定量的な評価方法について議論する。さらに、本研究では、差分プライバシーの方法論が適用された集計表を作成するにあたって、調査項目を追加したり、調査項目の分布特性に基づいて按分を行ったりする場合に、それが集計表の有用性に及ぼす影響についての評価を行う。

2. 秘匿処理を施したデータにおける有用性の評価方法に関する論点

本節では、差分プライバシー等の秘匿処理を施したデータに対する有用性の評価方法に関する論点について述べることにしたい。

公的統計の分野で、個票データに、リコーディング、トップ(ボトム)・コーディング等の非攪乱的手法(non-perturbative methods)やノイズ付与やスワッピングといった攪乱的手法(perturbative methods)を含む各種の秘匿処理の方法を適用した場合、匿名化されたミクロデータに対する有用性の定量的な評価方法については、主として①記述統計量やクロス表等を用いた元データとの分布特性の差異の把握、および②情報量損失(information loss)に関する指標を用いた評価を指摘することができる(伊藤(2019))。また、センサスデータを例に、スワッピングと差分プライバシーの実現方式を比較・検証した Christ et al. (2022)の研究では、有用性の指標として、平均平方誤差(mean square error)および平滑化されたカルバック・ライブラー情報量(μ-smoothed Kullback-Leibler divergence)が用いられている。これらの指標はそれぞれ、上記の①と②に該当していると言える。

差分プライベートなデータの有用性については、差分プライバシーが適用される前の元データからの分布特性の差異を定量的に把握するために、センサス局が 2010 年センサスの PPDMs の作成・公開にあたって有用性を測るための指標として用いられた平均絶対誤差(MAE)、さらには 2 乗平均平方根誤差(RMSE)等、セル単位での平均誤差を用いた評価を行ってきた。これらの指標は、匿名化ミクロデータの有用性に関する評価方法の中の、①記述統計量やクロス表等を用いた元データとの分布特性の差異の把握に対応している。このように元データと差分プライベートなデータの間で分布特性を比較することは、有用性を定量的に評価する有効な方法になっている。そこでの論点は、分布特性の比較に用いる指標をどう

いった観点から有用性評価のために用いるかということである。具体的には、差分プライバシーに伴う攪乱によって統計数値の安全性が保証された上で、様々な地理的規模の下でノイズが有用性に与える影響という観点から、ノイズ付与済みの集計データと元になる個票データから作成された集計データとの分布特性の差異を定量的に明らかにすることである。

3 生態学的誤謬から見た差分プライベートな統計表の有用性評価について

異なる粒度の地域区分において、個票データに基づく変数間の相関性といった有用性の指標と集計データから算出された場合のそれとを比較・検討することが考えられる。こうした集計データと元の個票データにおける分布特性との関係性は、Robinson によって「生態学的誤謬(ecological fallacy)」(Robinson(1950))という形で概念化された(伊藤(2002))。生態学的誤謬とは、個別主体群の社会経済的特性に関する「個別的相関(individual correlation)」ではなく、地域レベルの集団的特性について算出された「生態学的相関(ecological correlation)」によって個別主体の社会経済的属性間の関連性を把握することを表している(Robinson(1950), 伊藤(2002), 伊藤(2011))。この場合、個別的相関と生態学的相関の 2 つの数値が異なるのにもかかわらず、生態学的相関を誤って個別的関係に当てはめることによって、生態学的誤謬が生じる。具体的には、複数の変数をすべて用いたクロス表の作成において、変数のより少ない複数の表から按分して擬似的に作成する場合に、按分によって生じる誤差が生態学的誤謬を発生させる要因となっている。

それに対して、集計データを用いて個々人の社会経済的属性と社会的行為との関係を推測することは、生態学的推論(ecological inference)と呼ばれており(Openshaw(1984), Holt et al.(1996), 伊藤(2011))、「生態学的誤謬の慎重な適用」(Allardt(1969, p.42))という観点から、個々人の社会的行為を対象にした場合の集計データによる実証的な社会研究の可能性が追究されてきた。生態学的推論の一手法が、生態学的変数間における回帰分析を行うことによって、個別主体の社会的な行為事象に関する傾向的な把握を試みる生態学的回帰(ecological regression)である(Goodman(1953), 伊藤(2011, p.40))。生態学的回帰を用いた有用性の評価については、例えば、公表されたセンサスデータを用いて、生態学的回帰による検証結果に基づき、差分プライバシーの実現方式の有効性を明らかにした Cohen et al. (2022)の研究がある(伊藤他(2024))。

差分プライバシーの実現方式が適用される集計表を対象に、地域の粒度の観点から有用性の評価を行う上では、差分プライバシーの適用によって付与されるノイズつき集計データの元データに基づく集計データとの誤差(差分プライバシーに起因する誤差)、および元の個票データにおける分布特性の地域区分と地域の粒度ごとに集計されたデータの分布特性との差異によって、生態学的誤謬をもたらす誤差(生態学的誤謬に起因する誤差)の両方を勘案することが求められる。この差分プライバシーに起因する誤差と生態学的誤謬に起因する誤差の総計値が最も小さくなるように、集計表の作成に用いた変数群と地域区分の粒度の組み合わせが選択された場合、それは最も有用性の高い分布特性を有するデータとみなすことができる。このとき、選択されたデータが最も細かい地域区分を有するミクロデータとは限らないことに留意する必要がある。粒度が細かいほど、集計表の分布特性は、個票データにおけるそれと類似する傾向にあるため、生態学的誤謬に起因する誤差は小さくなるものの、集計表に含まれるセルの度数は小さくなることから、差分プライベートなノイズが各セルに付与された場合、ノイズが度数に与える影響が相対的に大きくなることが考えられるからである。

差分プライバシーに起因する誤差と生態学的誤謬に起因する誤差との総計値に関しては、定式化を行うことが求められる。これらの 2 つの誤差の総計を定量的に評価するための

有用性指標の検討も必要になるだろう。これについては、各種の差分プライバシーの実現方式と異なる地域区分の粒度のもとで、様々な有用性指標を用いた定量的な評価を行う必要がある。それによって、各種の差分プライバシーの実現方式が適用されたデータに対する有用性の評価に関する体系化を図ることも可能になる。

次節では、令和2年国勢調査の個票データを対象に、特定の地域区分において対象となるすべての変数を用いて作成する方法(以下「フルクロス集計法」)に基づくクロス集計表(以下「フルクロス表」と呼称)とより上位の地域区分からの按分を適用して作成する方法(以下「按分集計法」)に基づく擬似的なクロス集計表(以下「按分クロス表」と呼称)をそれぞれ差分プライバシーに基づき作成し、比較・検証を行う。具体的には、フルクロス表の作成にあたって、伊藤他(2024)で用いた手法の中の3種類の差分プライバシーの実現方式(単純なLaplace メカニズム(+負値の切り上げ)、ボトムアップ構成法、トップダウン構成法)を適用した[3]。つぎに、按分クロス表の作成においては、より上位の地域区分における属性別のクロス表と当該の地域区分における属性なしの総計値のそれぞれに差分プライバシーに基づくノイズを付与した。これらのフルクロス表と按分クロス表の結果を比較・検証することによって、地域区分の粒度の相違と差分プライバシーに伴う誤差との関連性を追究したい。

4. 2020年の国勢調査データに対する差分プライバシーの適用に関する実証実験について

本節では、わが国の国勢調査における個票データを用いて、差分プライバシーの適用に関する実証実験の手順とその特徴について述べる。

本実験においては、令和2年国勢調査における個票データ (調査票情報)を使用する。本実験では、性別、年齢と住居の種類の3つの変数を対象に、各種のクロス集計表を作成するだけでなく、地域区分の粒度が異なる小地域別の集計表も作成した上で、各種の差分プライバシーの実現方式を適用した。さらに、調査項目の分布特性に基づいて按分を行った場合、それが集計表にどのような影響を及ぼすかについて定量的な評価を行った。

本研究は、地域区分の粒度とクロス表で用いる変数群の関係、およびそれぞれのクロス表において差分プライバシーに基づくノイズ付与が有用性に及ぼす影響を把握することを指向している。具体的には、性別(2区分)、年齢(18区分)と住居の種類(3区分)の3つの変数を対象に、都道府県、市区町村、町・字および基本単位区のそれぞれについて、差分プライベートなフルクロス表と按分クロス表を作成し、それらの比較・検証を行った。按分クロス表に関しては、①日本全国のクロス表から、都道府県別の人口数に基づいて按分した都道府県別のクロス表の推計値、②都道府県別のクロス表から、市区町村別の人口数に基づいて按分したあらゆるパターンの市区町村別のクロス表の推計値、③市区町村別のクロス表から、各町・字別の人口数に基づいて按分した町・字別のクロス表の推計値、④町・字別のクロス表から、基本単位区別の差分プライベートな人口数に基づいて按分した基本単位区別のクロス表の推計値のそれぞれに対して、差分プライバシーに基づくノイズを適用した。

より上位の地域区分から按分された性別、年齢と住居の種類のクロス表におけるセルの推計値と、同じ変数のすべてを用いて作成した表におけるセルの数値との差異は、生態学的誤謬に起因する誤差に相当する。本実験では双方のクロス表に差分プライベートなノイズを付与した按分クロス表とフルクロス表を比較することによって、生態学的誤謬に起因する誤差

[3] 伊藤他(2024)では、PRAM(=Post RAndomization Method)による検証も行ったが、他の差分プライバシーの実現方式と比較して、有用性が顕著に低いことが定量的に確認されたことから、本研究では、PRAMによる実験を行っていない。

を考慮する形で、差分プライベートなノイズがセルの数値に及ぼす影響について定量的に把握することを目指している。

本実験では、負値の切り上げを含む Laplace メカニズム、ボトムアップ構成法、トップダウン構成法の 3 種類の方法を用いた。前述の通り、伊藤他 (2024) で検証した手法のうち PRAM については、これらの方法と比較して有用性が顕著に低いことが確認されたことから、本実験では対象外としている。ここで、ボトムアップ構成法とは、(Laplace メカニズムと同様に) それぞれの集計表における最小集計区分のセル値に対して Laplace ノイズを付加した後に、多次元ベクトル空間における正規単体 (canonical simplex) への射影問題を解くことにより、セル値の総数 (総数制約) を保持しつつ、Laplace ノイズの付与により生じる負のセル値を除去する (非負制約の充足) 方法である (寺田他(2017a),寺田他(2017b))。その一方、トップダウン構成法とは、まず全国の人口総数を総数制約とした都道府県単位の人口に対して上記を適用することによりプライバシー保護済みの都道府県単位の人口を算出し、次にその (プライバシー保護済みの) 都道府県単位の人口を総数制約としてプライバシー保護済みの市区町村単位の人口を算出するようなやり方で、トップダウンの方向で再帰的に総数制約を保持しつつ非負制約を充足したプライバシー保護済みの人口を得る構成法である。なお、トップダウン構成法においては、それぞれの地域区分ごとにプライバシー損失予算を配分する必要があるが、本実験ではこれを均等に配分している。

プライバシー損失予算 (ε) は、伊藤他(2024)と同様に、0.1、0.2、0.7、1.0、1.1、5、10、20 の 8 種類を設定した。なお、按分クロス表の作成の際に用いられる 2 種類の表(上位の地域区分における属性別の表、当該の地域区分での属性を含まない総計値のみの表)の作成にあたっては、プライバシー損失予算をそれぞれの表に 2 分の 1 ずつ均等に配分している[4]。さらに、本実験では、フルクロス表と按分クロス表の比較・検証のために、地域区分ごとに平均絶対誤差 (MAE)と二乗平均平方根誤差(RMSE)を誤差指標として用いている。ただし本稿では、紙面の制約上、MAE による比較の結果のみに基づいて議論する。

本研究の特徴としては、前節での議論を踏まえ、差分プライバシーに起因する誤差と生態学的誤謬に起因する誤差の総計値を元データからの誤差という形で MAE や RMSE を用いて計測しようとしていることが指摘できる。さらに、対象となるすべての変数でクロス表を作成し、差分プライバシーを適用した場合の結果と、それより上位の区分でクロス表を作成した後、差分プライバシー実現方式を適用し、下位の区分で按分した場合の結果を比較することで、差分プライバシーに起因する誤差と生態学的誤謬に起因する誤差の関係性を定量的に把握しようとしている点が注目される。

5. 実験の結果と考察

表 1〜8 は、性別、年齢と住居の種類の 3 変数のクロス表を例に、基本単位区、町・字、市区町村と都道府県のそれぞれを対象にしてフルクロス表と按分クロス表を作成した上で、MAE を指標とした本実験の評価結果を表している。各表において、(a) Laplace、(b) BottomUp、(c) TopDown はそれぞれ、Laplace メカニズム (+ 負値の切り上げ)、ボトムアップ構成法、およびトップダウン構成法を指す。なお、表中における太字は、同一条件において最も MAE が小さい (誤差が小さい) 差分プライバシー適用手法を示す。また、背景が網掛けのセルは、

[4] 地域区分ないし属性区分ごとに異なるプライバシー損失予算を配分するよう構成することも可能であるが、本実験では、予算の配分比率は一定にした。

表 1　性別×年齢×住居の種類に関するフルクロス表―基本単位区を対象

ε	手法	MAE(全国)	MAE(都道府県)	MAE(市区町村)	MAE(町字)	MAE(基本単位区)
0.1	(a)Laplace	9368343.77	199326.46	4938.51	91.30	5.16
	(b)BottomUp	11415.90	2706.27	113.71	**4.90**	**0.66**
	(c)TopDown	**52.50**	**36.58**	**19.80**	5.79	0.77
0.2	(a)Laplace	4580701.41	97461.73	2414.72	44.77	2.63
	(b)BottomUp	6553.29	1865.22	79.71	**3.41**	**0.51**
	(c)TopDown	**24.88**	**18.30**	**11.55**	3.80	0.72
0.7	(a)Laplace	1239727.50	26377.18	653.57	12.27	0.79
	(b)BottomUp	2043.03	657.09	29.25	**1.38**	**0.23**
	(c)TopDown	**7.42**	**5.51**	**4.09**	1.56	0.54
1	(a)Laplace	855047.25	18192.49	450.80	8.50	0.56
	(b)BottomUp	1448.49	479.82	21.44	**1.05**	**0.18**
	(c)TopDown	**4.07**	**3.79**	**2.99**	1.19	0.47
1.1	(a)Laplace	774578.91	16480.40	408.38	7.71	0.51
	(b)BottomUp	1260.85	437.03	19.67	**0.97**	**0.16**
	(c)TopDown	**3.93**	**3.41**	**2.74**	1.10	0.45
5	(a)Laplace	165950.78	3530.87	87.51	1.67	0.11
	(b)BottomUp	305.43	109.69	4.99	**0.27**	**0.04**
	(c)TopDown	**1.05**	**0.79**	**0.67**	0.33	0.17
10	(a)Laplace	82949.81	1764.89	43.74	0.84	0.06
	(b)BottomUp	150.59	55.30	2.51	**0.13**	**0.02**
	(c)TopDown	**0.48**	**0.39**	**0.34**	0.18	0.10
20	(a)Laplace	41467.23	882.28	21.87	0.42	0.03
	(b)BottomUp	68.13	27.51	1.25	**0.07**	**0.01**
	(c)TopDown	**0.23**	**0.20**	**0.17**	0.09	0.05

表 2　性別×年齢×住居の種類に関する按分クロス表―基本単位区を対象

ε	手法	MAE(全国)	MAE(都道府県)	MAE(市区町村)	MAE(町字)	MAE(基本単位区)
0.1	(a)Laplace	910658.64	19375.72	484.49	11.14	0.78
	(b)BottomUp	5915.24	1127.86	51.94	**3.50**	**0.36**
	(c)TopDown	**64.12**	**59.14**	**28.07**	7.39	0.58
0.2	(a)Laplace	437610.58	9311.27	234.33	5.75	0.51
	(b)BottomUp	2942.42	583.23	28.65	**2.10**	**0.30**
	(c)TopDown	**34.87**	**29.77**	**16.70**	5.10	0.46
0.7	(a)Laplace	115895.64	2466.44	63.18	1.73	0.32
	(b)BottomUp	999.59	208.31	10.71	**0.81**	**0.26**
	(c)TopDown	**10.76**	**8.53**	**6.09**	2.20	0.31
1	(a)Laplace	79628.75	1694.99	43.66	1.23	0.30
	(b)BottomUp	558.41	157.58	8.17	**0.61**	**0.26**
	(c)TopDown	**6.93**	**6.07**	**4.48**	1.69	0.29
1.1	(a)Laplace	71933.97	1531.35	39.51	1.12	0.29
	(b)BottomUp	635.96	145.43	7.59	**0.57**	**0.26**
	(c)TopDown	**6.23**	**5.43**	**4.17**	1.58	0.29
5	(a)Laplace	14897.50	317.46	8.32	0.25	0.26
	(b)BottomUp	133.98	38.98	2.09	**0.15**	**0.25**
	(c)TopDown	**1.50**	**1.25**	**1.05**	0.48	0.26
10	(a)Laplace	7415.91	158.04	4.15	0.13	0.25
	(b)BottomUp	59.51	19.66	1.07	**0.08**	**0.25**
	(c)TopDown	**0.78**	**0.62**	**0.54**	0.27	0.25
20	(a)Laplace	3705.32	78.97	2.07	0.06	0.25
	(b)BottomUp	28.72	9.93	0.54	**0.04**	**0.25**
	(c)TopDown	**0.35**	**0.32**	**0.28**	0.15	0.25

表 3　性別×年齢×住居の種類に関するフルクロス表―町・字を対象

ε	手法	MAE(全国)	MAE(都道府県)	MAE(市区町村)	MAE(町字)	MAE(基本単位区)
0.1	(a)Laplace	437662.06	9312.20	234.38	5.75	
	(b)BottomUp	3076.17	582.22	28.64	**2.10**	
	(c)TopDown	**33.18**	**29.93**	**16.76**	5.10	
0.2	(a)Laplace	209658.04	4461.18	113.33	2.96	
	(b)BottomUp	1481.22	319.83	16.36	**1.24**	
	(c)TopDown	**18.99**	**15.08**	**9.66**	3.28	
0.7	(a)Laplace	55762.55	1187.01	30.72	0.89	
	(b)BottomUp	401.13	119.29	6.27	**0.47**	
	(c)TopDown	**5.66**	**4.25**	**3.36**	1.31	
1	(a)Laplace	38382.05	817.60	21.26	0.63	
	(b)BottomUp	299.08	89.03	4.71	**0.35**	
	(c)TopDown	**3.46**	**3.07**	**2.46**	1.00	
1.1	(a)Laplace	34825.83	741.59	19.30	0.57	
	(b)BottomUp	305.32	82.27	4.36	**0.32**	
	(c)TopDown	**3.79**	**2.80**	**2.25**	0.93	
5	(a)Laplace	7417.11	158.11	4.15	0.13	
	(b)BottomUp	61.11	19.75	1.07	**0.08**	
	(c)TopDown	**0.84**	**0.64**	**0.54**	0.27	
10	(a)Laplace	3704.45	78.95	2.07	0.06	
	(b)BottomUp	31.63	9.93	0.54	**0.04**	
	(c)TopDown	**0.40**	**0.32**	**0.28**	0.15	
20	(a)Laplace	1851.78	39.47	1.04	0.03	
	(b)BottomUp	15.56	4.95	0.27	**0.02**	
	(c)TopDown	**0.21**	**0.16**	**0.14**	0.07	

表 4　性別×年齢×住居の種類に関する按分クロス表―町・字を対象

ε	手法	MAE(全国)	MAE(都道府県)	MAE(市区町村)	MAE(町字)	MAE(基本単位区)
0.1	(a)Laplace	11687.38	277.69	13.99	1.93	
	(b)BottomUp	871.96	95.32	**10.06**	**1.84**	
	(c)TopDown	**50.52**	**44.53**	22.64	2.05	
0.2	(a)Laplace	5253.09	128.23	7.25	1.84	
	(b)BottomUp	392.00	50.57	**5.59**	**1.80**	
	(c)TopDown	**26.14**	**21.73**	13.36	1.89	
0.7	(a)Laplace	1221.97	32.74	2.22	1.79	
	(b)BottomUp	121.06	15.04	**1.84**	**1.78**	
	(c)TopDown	**8.23**	**6.54**	4.79	1.80	
1	(a)Laplace	807.71	22.29	1.58	1.79	
	(b)BottomUp	77.84	10.67	**1.32**	**1.78**	
	(c)TopDown	**4.90**	**4.57**	3.50	1.79	
1.1	(a)Laplace	723.72	20.13	1.45	1.79	
	(b)BottomUp	76.51	9.62	**1.21**	**1.78**	
	(c)TopDown	**4.78**	**4.09**	3.24	1.79	
5	(a)Laplace	139.53	4.31	0.33	1.78	
	(b)BottomUp	18.38	2.16	**0.28**	**1.78**	
	(c)TopDown	**1.30**	**0.94**	0.80	1.78	
10	(a)Laplace	67.76	2.14	0.16	1.78	
	(b)BottomUp	8.24	1.09	**0.14**	**1.78**	
	(c)TopDown	**0.65**	**0.49**	0.41	1.78	
20	(a)Laplace	34.89	1.07	0.08	1.78	
	(b)BottomUp	4.43	0.55	**0.07**	**1.78**	
	(c)TopDown	**0.32**	**0.23**	0.21	1.78	

表 5　性別×年齢×住居の種類に関するフルクロス表―市区町村を対象

ε	手法	MAE(全国)	MAE(都道府県)	MAE(市区町村)	MAE(町字)	MAE(基本単位区)
0.1	(a)Laplace	5284.88	128.82	7.28		
	(b)BottomUp	477.99	50.82	**5.63**		
	(c)TopDown	**27.65**	**22.14**	13.32		
0.2	(a)Laplace	2374.19	60.06	3.78		
	(b)BottomUp	221.24	26.21	**3.05**		
	(c)TopDown	**14.96**	**11.47**	7.67		
0.7	(a)Laplace	562.16	15.76	1.14		
	(b)BottomUp	64.08	7.60	**0.96**		
	(c)TopDown	**3.79**	**3.25**	2.60		
1	(a)Laplace	374.76	11.14	0.81		
	(b)BottomUp	43.11	5.36	**0.68**		
	(c)TopDown	**2.86**	**2.28**	1.88		
1.1	(a)Laplace	336.34	9.82	0.74		
	(b)BottomUp	39.98	4.82	**0.62**		
	(c)TopDown	**2.89**	**2.07**	1.73		
5	(a)Laplace	68.16	2.14	0.16		
	(b)BottomUp	8.52	1.07	**0.14**		
	(c)TopDown	**0.63**	**0.47**	0.41		
10	(a)Laplace	35.52	1.09	0.08		
	(b)BottomUp	4.73	0.55	**0.07**		
	(c)TopDown	**0.27**	**0.24**	0.21		
20	(a)Laplace	17.21	0.53	0.04		
	(b)BottomUp	2.38	0.28	**0.04**		
	(c)TopDown	**0.15**	**0.12**	0.10		

表 6　性別×年齢×住居の種類に関する按分クロス表―市区町村を対象

ε	手法	MAE(全国)	MAE(都道府県)	MAE(市区町村)	MAE(町字)	MAE(基本単位区)
0.1	(a)Laplace	222.79	17.16	51.87		
	(b)BottomUp	139.13	**15.75**	**51.83**		
	(c)TopDown	**34.36**	30.01	51.91		
0.2	(a)Laplace	120.64	8.85	51.81		
	(b)BottomUp	70.15	**7.70**	**51.79**		
	(c)TopDown	**17.03**	15.13	51.81		
0.7	(a)Laplace	34.27	2.51	51.78		
	(b)BottomUp	20.94	**2.23**	**51.77**		
	(c)TopDown	**5.61**	4.30	51.77		
1	(a)Laplace	22.98	1.76	51.78		
	(b)BottomUp	14.54	**1.59**	**51.77**		
	(c)TopDown	**3.37**	3.04	51.77		
1.1	(a)Laplace	21.86	1.58	51.78		
	(b)BottomUp	12.24	**1.43**	**51.77**		
	(c)TopDown	**3.80**	2.77	51.77		
5	(a)Laplace	4.56	0.35	51.77		
	(b)BottomUp	2.91	**0.32**	51.77		
	(c)TopDown	**0.87**	0.63	**51.77**		
10	(a)Laplace	2.18	0.18	51.77		
	(b)BottomUp	1.65	**0.17**	**51.77**		
	(c)TopDown	**0.39**	0.32	51.77		
20	(a)Laplace	1.11	0.09	51.77		
	(b)BottomUp	0.83	**0.08**	51.77		
	(c)TopDown	**0.18**	0.16	**51.77**		

表７　性別×年齢×住居の種類に関するフルクロス表―都道府県を対象

ε	手法	MAE(全国)	MAE(都道府県)	MAE(市区町村)	MAE(町字)	MAE(基本単位区)
0.1	(a)Laplace	121.58	8.68			
	(b)BottomUp	62.62	**7.63**			
	(c)TopDown	**18.92**	14.71			
0.2	(a)Laplace	57.61	4.35			
	(b)BottomUp	36.29	**3.88**			
	(c)TopDown	**8.86**	7.63			
0.7	(a)Laplace	14.75	1.25			
	(b)BottomUp	10.05	**1.12**			
	(c)TopDown	**2.84**	2.18			
1	(a)Laplace	11.03	0.89			
	(b)BottomUp	6.98	**0.79**			
	(c)TopDown	**2.01**	1.55			
1.1	(a)Laplace	10.11	0.79			
	(b)BottomUp	6.27	**0.71**			
	(c)TopDown	**1.89**	1.41			
5	(a)Laplace	2.25	0.17			
	(b)BottomUp	1.51	**0.16**			
	(c)TopDown	**0.43**	0.32			
10	(a)Laplace	1.11	0.09			
	(b)BottomUp	0.77	**0.08**			
	(c)TopDown	**0.15**	0.16			
20	(a)Laplace	0.57	0.04			
	(b)BottomUp	0.36	**0.04**			
	(c)TopDown	**0.09**	0.08			

表 8　性別×年齢×住居の種類に関する按分クロス表―都道府県を対象

ε	手法	MAE(全国)	MAE(都道府県)	MAE(市区町村)	MAE(町字)	MAE(基本単位区)
0.1	(a)Laplace	**18.40**	**1534.59**			
	(b)BottomUp	19.85	1534.62			
	(c)TopDown	19.28	1534.64			
0.2	(a)Laplace	10.23	1534.60			
	(b)BottomUp	**8.62**	**1534.57**			
	(c)TopDown	9.38	1534.60			
0.7	(a)Laplace	**2.66**	**1534.57**			
	(b)BottomUp	3.04	1534.57			
	(c)TopDown	2.99	1534.57			
1	(a)Laplace	1.76	1534.57			
	(b)BottomUp	2.17	**1534.57**			
	(c)TopDown	**1.74**	1534.57			
1.1	(a)Laplace	1.92	1534.57			
	(b)BottomUp	1.96	**1534.57**			
	(c)TopDown	**1.92**	1534.57			
5	(a)Laplace	**0.38**	1534.57			
	(b)BottomUp	0.40	1534.57			
	(c)TopDown	0.41	**1534.56**			
10	(a)Laplace	**0.18**	**1534.56**			
	(b)BottomUp	0.20	1534.56			
	(c)TopDown	0.20	1534.56			
20	(a)Laplace	0.10	1534.56			
	(b)BottomUp	0.11	**1534.56**			
	(c)TopDown	**0.09**	1534.56			

フルクロス表と按分クロス表を比較したときに、MAE がもう一方より小さいことを示す。

5.1 フルクロス集計法の結果の全体的な傾向について

まず、フルクロス集計法の結果 (表 1、表 3、表 5、表 7) についての全体的な傾向を議論する。

いずれの表においても、単純な Laplace メカニズムとボトムアップ構成法を比較すると、集計条件やプライバシー損失予算 (ε の値) が等しいもとでは、すべての条件においてボトムアップ構成法の誤差が単純な Laplace メカニズムの誤差より小さいか同等である。つまり、誤差の観点ではボトムアップ構成法が単純な Laplace メカニズムより純粋に優れていると言える。

次に、ボトムアップ構成法とトップダウン構成法を比較すると、たとえば表 1 において基本単位区などの細かい集計区分においてはボトムアップ構成法の誤差がトップダウン構成法の誤差より小さいが、ボトムアップ構成法の誤差は地域区分を粗くするにつれて大きく増大するのに対し、トップダウン構成法は地域区分を粗くしても誤差の増大はわずかに抑えられている。この傾向は他の表においても同様に見られる。

これらの結果は、伊藤他 (2024) における実験結果を裏付けるものであり、集計条件の変化にかかわらず、上記の傾向については変化がないことが示された。

5.2 按分集計法の結果の全体的な傾向について

同様に、按分集計法の結果 (表 2、表 4、表 6、表 8) についての全体的な傾向を議論する。

これらの実験結果も、手法間の関係については前節のフルクロス集計法の実験結果と類似の傾向を示した。すなわち、ボトムアップ構成法は単純な Laplace メカニズムに対していずれの集計条件においても同等ないし優れている。また、集計区分の細かさによりボトムアップ構成法とトップダウン構成法の優劣が逆転する点についてもほぼ同様である。

ただし、フルクロス集計法の実験結果においては、いずれの手法においても集計区分を粗くするにつれて (程度の大小は異なりつつも) 誤差が増大している一方で、按分集計法ではプライバシー損失予算が大きい、すなわち安全性が低い条件下において、(按分により作成された) 最小の地域区分と、その上位の地域区分との間で誤差が逆転している。たとえば、表 2 において、基本単位区の誤差と町字の誤差が逆転しており、具体的には、単純な Laplace メカニズムとボトムアップ構成法では $\varepsilon \geqq 5$ の条件下、トップダウン構成法では $\varepsilon = 20$ の条件下で逆転が見られる。この逆転が発生する ε の値は最小の地域区分の粒度によって異なり、その粒度が粗くなるにつれて逆転が発生する ε の値は小さくなる。

この現象は、伊藤他(2024)には見られず、按分により生じた影響と考えられる。この点については後にあらためて考察する。

5.3 フルクロス集計法と按分集計法の比較

次に、これらの 2 つの集計法の結果について比較する。

まず、(最小の地域区分を基本単位区とした) 表 1 と表 2 の結果に着目すると、単純な Laplace メカニズムとボトムアップ構成法において、町字以上での地域区分では、いずれのプライバシー予算でも按分集計法におけるセルの背景が網掛けで示されている。つまり、按分集計法のほうがフルクロス集計法より誤差が小さい。その一方、基本単位区では、プライバシー損失予算が小さい条件では按分集計法のほうがフルクロス集計法より誤差が小さい (按分集計法のセルの背景が網掛けである) が、プライバシー損失予算が大きい条件ではそれが逆転

する (フルクロス法のセルの背景が網掛けである)。具体的には、単純な Laplace メカニズムでは $\varepsilon \geqq 5$、ボトムアップ構成法では $\varepsilon \geqq 0.7$ の条件下において、この逆転が発生している。この逆転が発生する ε の値は最小地域区分の粒度が大きくなるにつれて小さくなり、たとえば最小地域区分を町字とした場合 (表 3 と表 4) では、逆転が発生する ε の値はそれぞれ $\varepsilon \geqq 0.7$ と $\varepsilon \geqq 0.2$ となる。最小地域区分が市区町村以上のときは、ε の値にかかわらず逆転が発生した。

その一方、トップダウン構成法は傾向が異なる。たとえば表 1 と表 2 の比較において、他の 2 つとは逆に、町字以上の地域区分ではいずれのプライバシー予算でもフルクロス集計法のほうが按分集計法より誤差が小さい。基本単位区では、単純な Laplace メカニズムやボトムアップ構成法と同様に、プライバシー損失予算が小さい条件では按分集計法のほうがフルクロス集計法より誤差が小さく、プライバシー損失予算が大きい条件、具体的には $\varepsilon \geqq 5$ の条件下でそれが逆転する。最小地域区分の粒度が大きくなるにつれて逆転が発生する ε の値が小さくなる点についても同様であり、町字を最小地域区分とした場合 (表 3 と表 4) では $\varepsilon \geqq 0.7$ において逆転が発生し、市区町村以上が最小地域区分のときは発生しない。

5.4 比較結果に関する考察

前節で示した通り、たとえば表 1 と表 2 の比較において、単純な Laplace メカニズムとボトムアップ構成法は、町字以上での地域区分ではいずれのプライバシー予算でも按分集計法のほうがフルクロス集計法より誤差が小さいという結果が得られた。

これは、それぞれの手法の性質を考えると自然な結果である。すなわち、これらの手法においては、最小地域区分以外の値は、最小地域区分の値から畳みあげて作成されることから、その誤差も最小地域区分が細かいほど多く累積され、その影響も著しく大きい。したがって、フルクロス集計法よりも最小地域区分が「粗い」按分集計法のほうが、誤差の累積が小さく抑えられることから、このような結果が得られたと考えられる。

このことは、トップダウン構成法において、町字以上での集計地域で逆の結果、つまりフルクロス集計法のほうが按分集計法より誤差が小さいという結果が一貫して得られたことにも符合する。

つまり、トップダウン構成法においては、地域区分の畳みあげによる誤差の蓄積は理論的にほとんど存在しない (なお、地域区分が小さいほうが誤差の値が小さいのは、主に非負精緻化の改善効果[5]によるものである) ことから、按分による誤差の発生がほぼそのまま按分集計法における誤差の悪化として現れたものと推測される。

その一方で、表 1 と表 2 における最小地域区分である基本単位区での結果は、いずれの手法においても同じ傾向が得られ、プライバシー損失予算が小さい条件では按分集計法のほうが、プライバシー損失予算が大きい条件ではフルクロス集計法が、より誤差が小さいという傾向を示した。

まず、この傾向が示された点については自然な結果と言える。つまり、差分プライバシーを保証するために必要となるノイズ強度はプライバシー損失予算が小さいほど (安全性が高いほど) 大きくなることから、プライバシー損失予算が小さいときは、このノイズによる影響が大きくなるが、その一方で、上位の地域区分からの按分による誤差は、プライバシー損失予算とは関係がなく一定である。したがって、プライバシー損失予算が小さいほど相対的

[5] トップダウン構成法では、集計結果が疎 (sparse) であるほど非負精緻化による誤差の改善効果が高いため、地域区分が細かいほどその効果により誤差が改善される。

にノイズによる影響が強くなることから、按分集計法のほうが有利な状況にあると言える。

ただし、具体的にプライバシー損失予算の値がいくつのときに優劣が逆転するかについては手法により結果が分かれ、具体的には単純な Laplace メカニズムでは $\varepsilon \geqq 5$、ボトムアップ構成法では $\varepsilon \geqq 0.7$、トップダウン構成法では $\varepsilon \geqq 5$ のとき、それぞれフルクロス集計法のほうが按分集計法より誤差が小さくなった。

これは、按分集計法による誤差と、差分プライバシーを保証するために必要となるノイズ強度との兼ね合いにより発生するものである。たとえば極端にプライバシー損失予算が大きい条件、たとえば $\varepsilon=20$ の条件下においては、表 1 においていずれの手法においても基本単位区の誤差は 0.01〜0.05 であり、ノイズによる誤差の影響は極めて小さいと言える。その一方で、表 2 ではいずれも 0.25 という値を示しており、これはほとんど按分により発生した誤差であるとみなすことができる。実際、表 2 においてプライバシー損失予算がより小さい値のときの基本単位区の誤差は 0.25 以上であり、この値が本実験での按分集計法における基本単位区の誤差の下限となっている。そこで、あらためて表 1 における基本単位区の誤差の値と、両集計法の優劣との関係について確認してみると、誤差の値が 0.25 を上回るか否かによってその優劣が定まっている。

これは、最小地域区分を町字以上とした結果 (表 3〜8) によっても裏付けられる。たとえば最小地域区分を町字とした結果 (表 3 と表 4) の比較において、按分による誤差は 1.78 程度と推定され、表 3 において町字別の誤差の値が 1.78 を上回るか否かによって優劣が定まっている。また、市区町村や都道府県を最小地域区分とした場合、按分による誤差が極めて大きいものとなり (市区町村別で 51.77、都道府県別で 1534.56)、常にフルクロス集計法が按分集計法より誤差が小さいという結果が得られた。

実際には、按分により発生する誤差は対象とするデータによって異なり、また、按分集計法においてもノイズによる影響を受けるため、この値が絶対的な基準にはならないが、ノイズによる誤差が按分誤差を大きく上回るほど細かい集計区分でフルクロス集計表を作っても、より大きな集計区分からの按分により細かい集計区分の表を作ったほうが良い、つまりあまり細かすぎる集計区分で表を作成しても意味がないこと、およびその粒度はプライバシー損失予算が小さいほど粗いものとなることを示唆している。

6. まとめ

本稿では、差分プライバシーの実現方式が適用された国勢調査データに対する有用性の評価方法を追究するために、令和 2 年国勢調査の個票データを用いて作成した集計表をもとに、各種の差分プライバシーの実現手法を適用した場合の有用性を定量的に評価した。本研究では、フルクロス表と按分クロス表を作成した上で、より粗い集計表に基づく按分クロス表が、相対的に細かな集計表であるフルクロス表と比べて、元データに対してより近似的であるかを検証するために、MAE に基づく有用性の比較・検証を行った。第 3 章で述べたように、クロス表の粒度が細かくなるにつれて、生態学的誤謬に起因する誤差は小さくなるが、差分プライバシーに起因する誤差が大きくなる。そこで、本稿では、これらの関係性を想定した上で実験を行った。本実験の結果から、上位の地域区分において調査項目の分布特性に基づいて按分を行った場合に、地域区分の粒度と差分プライバシーの実現方式によっては、按分によって作成された集計表に関する MAE は、対象となるすべての変数を用いて作成された同様の集計表と比較して小さくなるという興味深い結果が明らかになった。按分クロス表では按分に起因した誤差が定量的に評価されるため、それがフルクロス表において計測された

誤差よりも小さければ、対象となる変数の一部から作成された複数のクロス表の按分によって生成されたデータは、誤差評価の観点から相対的に高い有用性を有するとみなすことができる。このことは、差分プライバシーが適用されたノイズ付与のデータに関しては、差分プライバシーに起因する誤差と生態学的誤謬に起因する誤差との関係性によっては、対象となる変数のすべてがミクロデータに含まれず、按分によって変数を生成したとしても、相対的により高い有用性が期待できることを示唆している。

他方、本研究では、差分プライバシーに起因する誤差と生態学的誤謬に起因する誤差を評価するための指標として MAE や RMSE を用いた。伊藤他(2024)は、地域区分の粒度に着目し、MAE というセル単位での平均誤差を用いた場合の有用性の評価に関する実証実験を行った。本実験の結果から、異なる粒度において基本単位区といった粒度が細かな地域区分の場合、元データに含まれる統計数値と差分プライバシーの実現方式が適用された数値とのずれをセル単位での平均誤差だけでは適切な評価が困難なことを指摘している。このことから、国勢調査への差分プライバシーの適用可能性をさらに追究するにあたって、セル単位の MAE や RMSE だけでなく、それとは異なる統計量を用いた有用性の検証方法についても検討することが求められる。しかしながら、本稿では、伊藤他(2024)で指摘した、粒度が細かな地域区分におけるセル単位での平均誤差による適切な評価の困難さに対する解決策については提示されていない。

したがって、差分プライベートなデータにおける有用性の定量的な評価方法については、さらに実証研究を進めていきたいと考えている。また、差分プライバシーに起因する誤差と生態学的誤謬に起因する誤差の総計値に関する定式化については、国勢調査といった公的統計を用いた差分プライバシーの実現方式の適用可能性に関するさらなる研究成果を踏まえ、将来的な研究課題として引き続き検討を行うことが求められよう。

参考文献

[1] 伊藤伸介(2002)「アメリカにおけるミクロ社会モデルの体系化の試み―オーカットの社会人口モデルと所得移転モデル―」,『統計学』第83号, pp.11-31.
[2] 伊藤伸介(2011)「わが国におけるミクロデータの新たな展開可能性について―イギリスにおける地域分析用ミクロデータを例に―」, 明海大学『経済学論集』Vol.23, No.3, pp.36-54.
[3] 伊藤伸介(2019)「公的統計データにおける秘匿性と有用性の評価のあり方に関する一考察―スワッピングを中心に―」, 坂田幸繁編『公的統計情報―その利活用と展望』中央大学出版部, pp.39-62.
[4] 伊藤伸介・寺田雅之 (2020)「詳細な地域データにおける秘匿処理の適用可能性について」『日本統計学会誌』第 50 巻第 1 号, pp.139-166.
[5] 伊藤伸介・寺田雅之・赤塚裕人・北井宏昌(2022)「海外における公的統計に対する攪乱的手法の新たな取り組み―アメリカセンサス局による差分プライバシーの適用を中心に―」『統計研究彙報』第 79 号, pp.131-150.
[6] 伊藤伸介・寺田雅之(2023)「海外における公的統計に関するプライバシー保護の現状―アメリカとイギリスの事例をもとに―」『統計研究彙報』第 80 号, pp.117-136.
[7] 伊藤伸介・寺田雅之・加藤駿典(2024) 「公的統計に対する差分プライバシーの適用と有効性の評価に関する検討―国勢調査を例に―」, 『統計研究彙報』第 81 号, pp.69-88.
[8] 寺田雅之・山口高康・本郷節之(2017a)「匿名個票開示への差分プライバシの適用」『情報処理学会論文誌』第 58 巻第 9 号, pp.1483-1500.

[9] 寺田雅之・山口高康・本郷節之(2017b)「高次元大規模データへの差分プライバシー適用のための最適精緻化法」『SCIS2017 予稿集』3B3-5, pp.1-8.
[10] Abowd, J. M. (2018). Staring-down the database reconstruction theorem, Joint Statistical Meetings, Vancouver, BC, Canada. https://www.census.gov/content/dam/Census/newsroom/press-kits/2018/jsm/jsmpresentation-database-reconstruction.pdf
[11] Allardt, E. (1969) "Aggregate Analysis: The Problem of Its Informative Value", Dogan, M. and Rokkan, S.(eds.) *Quantitative Ecological Analysis in the Social Sciences,* the M.I.T Press, London, pp.41-51.
[12] Christ, M., Radway, S., Bellovin, S. M. (2022) "Differential Privacy and Swapping: Examining De-Identification's Impact on Minority Representation and Privacy Preservation in the U.S. Census", Paper Presented at Conference: 2022 IEEE Symposium on Security and Privacy, pp.457-472.
[13] Cohen, A., Duchin, M., Matthews, J.N., and Suwal, B. (2022) "Private Numbers in Public Policy: Census, Differential Privacy, and Redistricting." *Harvard Data Science Review*, Special Issue 2, MIT Press.
[14] Dwork, C. (2006). Differential privacy, ICALP.
[15] Garfinkel, S. Abowd, J. M., and Martindale, C. (2019) "Understanding Database Reconstruction Attack in Public Data", Communications of the ACM, Vol. 62 No. 3, ACM, pp. 46-53.
[16] Goodman, L.A. (1953) "Ecological Regression and Behavior of Individuals", *American Sociological Review*, vol.18, pp.663-664.
[17] Holt, D., Steel, D. G., Tranmer, M., Wrigley, N.(1996) "Aggregation and Ecological Effects in Geographically Based Data", *Geographical Analysis,* Vol.28, No.3, pp.244-261.
[18] Jamin, R. (2021) "Disclosure Avoidance for the 2020 Census: An Introduction", U.S. Census Bureau.
[19] Office for National Statistics (2017) "Development of flexible dissemination for 2021 Census".
[20] Openshaw, S. (1984) "Ecological Fallacies and the Analysis of Area Census Data", *Environment and Planning A*, Vol.16, pp.17-31.
[21] Robinson, W. S. (1950) "Ecological Correlations and the Behavior of Individuals", *American Sociological Review*, vol.15, pp.351-357.