「プライバシー保護データマイニング」特集号 解　説

差分プライバシーとは何か

寺田　雅之*

1. はじめに

ビッグデータや AI 技術への注目の高まりが示すように，さまざまなデータを活用することにより社会や産業，そして私たちの生活を豊かにする知見を得られるようになることが期待されている．その一方，それらのデータは往々にして個人のプライバシーを含み，その取り扱いには法的・社会的に厳しい制約や責任が求められる．

そこで，プライバシーを保護しながら安全にデータを活用するための技術，すなわちプライバシー保護技術が重要となる．プライバシー保護技術は，データに含まれる個人に関する情報を開示することなく（安全性の保証），さまざまな知見を得るために必要となるデータの性質や特徴を得ること（有用性の確保）を目的とする. [5]

一般に，上記の安全性と有用性はトレードオフの関係にあり，どの程度の安全性を与えた場合にどの程度の有用性を得られるかはプライバシー保護技術ごとにそれぞれ異なる．つまり，十分な安全性を適切に保持したうえで，より高い有用性を備えたデータを出力できる方式が優れたプライバシー保護技術であるといえる．

さて，それでは安全性が十分に保持されているかどうかはどのように測ればよいのであろうか．残念なことに適切な安全性指標がどのようなものかは自明ではない．たとえば特定の攻撃のみに対する安全性を想定して構築された安全性指標は，「想定外」の攻撃により深刻なプライバシー開示を起こしうる．

差分プライバシー (differential privacy)[9,6] はこのような課題を背景として提唱された，プライバシー保護の安全性を網羅的に定義するための指標である．2016 年 6 月に Apple WWDC において iOS 10 への採用が表明されたほか，米国の国勢調査局 (Bureau of the Census) がその適用に向けた検討を進める [4] など，いまでは実用に向けた取り組みも広く進められつつある．

本稿では，プライバシー保護に関する安全性を評価することの難しさについて具体的な事例を交えながら議論したうえで，差分プライバシーの定義，および定義が意味するところを説明する．

* 株式会社 NTT ドコモ 先進技術研究所

Key Words: differential privacy, privacy measures, mosaic effects, composition theorems.

2. データ活用とプライバシー

個人に関する情報を含むデータベース（D とする）を加工し，なんらかの有益な情報を取り出すことを考える．データベース D は一つ以上のレコードから構成され，それぞれのレコードは各個人に対応する．

D から情報を取り出すための加工処理を問合せとよび，Q と表記する．これには D から有益な情報を取り出すためのあらゆる処理が考えられる．たとえば平均や分散などの統計量の取得や，クロス集計表やデータキューブなどの集計データの作成，D を学習データとした機械学習の予測モデル生成など，さまざまな「データ活用」におけるプロセスが問合せに相当する．

ここで，問合せの出力 $Q(D)$ は有益な情報だけでなく，D の個々のレコードに関する情報をなんらかの形で含む．そのため，$Q(D)$ から個人のプライバシーが開示 (disclosure)[1] されうることに注意が必要となる．

$Q(D)$ から個人のプライバシーを開示しようとする者を攻撃者とよぶ．攻撃者はさまざまな攻撃能力（攻撃に用いるアルゴリズムや計算資源など）や背景知識 (D の部分情報や，それ以外の外部情報) をもちうる．

以下，$Q(D)$ からのプライバシー開示がどのように発生しうるかについて簡単な例を用いて説明する．

2.1 プライバシー開示の例

男女 20 人のクラスで試験が実施された．その試験結果 D から受験者の性別と合否を抽出し（Q に相当する），第 1 表の形で公開したとする（$Q(D)$ に相当する）．

さて，このクラスの一人である A 君（男性）は，自分の試験結果を他人に知られたくないと考えている．このデータを得た攻撃者から A 君のプライバシーは守られるであろうか？

第 1 表　男女別の試験結果

性別	合否	人数 ($n=20$)
女性	合格	10 人
女性	落第	5 人
男性	落第	5 人

[1] D 中に含まれる個人に関する情報が推定できること．“disclosure” は「露見」や「暴露」とも訳されるが，本稿では日英統計用語集 [24] に従い「開示」とする．

答えは「守られない」．第 1 表に含まれる全人数を足すとクラスの人数になる $(10+5+5=20)$ ことから，男性の合格者は 0 人であることがわかる．したがって，A 君が落第したこともわかってしまう．

なお，このように「0」値の存在から特定個人のプライバシーが開示されることは「非構造的ゼロ (non-structural zero) による開示」とよばれる [14]．この攻撃にあたって，攻撃者にとって特別な背景知識は必要なく（「A 君は男性でこのクラスの一員である」以上の情報は必要ない），必要となる攻撃能力も「足し算をして n と比較する」ということだけである．

ここで，個人に対応するレコードが特定されていないにもかかわらず，特定個人のプライバシー開示が発生していることに注意したい．すなわち，第 1 表 には 5 個の同じ「男性，落第」というレコードが存在し，そのうちどれが A 君に対応するかは特定できないが，A 君という個人が試験に落ちたというプライバシーが決定的に開示される．つまり，データベース内のレコードが特定個人に対応づけられるかどうかと，個人のプライバシーが守られるかどうかは別の話であることがわかる．

2.2 プライバシー保護技術と安全性指標

そこで，問合せによるプライバシーの開示を防ぎつつ有益な情報を得るために，$Q(D)$ をそのまま出力するのではなく，$Q(D)$ と似ているがプライバシー保護の措置が施された $Q'(D)$ を出力することを考える．つまり（危険な）問合せ Q を（安全な）Q' へ置き換える．これを実現する技術は，PPDM (Privacy-Preserving Data Mining)[1] や PPDP (— Publishing)[11], または SDC (Statistical Disclosure Control)[14] などとよばれる．

本稿では，これらのプライバシー保護技術の適用により Q を書き換えて構築された Q' をプライバシー保護方式とよぶ．また，Q' がどのくらいの強度でプライバシーの開示を防ぐか，つまりプライバシーを保護するかを示す指標を，（プライバシーに関する）安全性指標とよぶ．どのような攻撃者の背景知識や攻撃能力を想定するかは，安全性指標の種類ごとにさまざまに異なる．

さきほど A 君のプライバシーを開示することが明らかになった第 1 表を，プライバシーに関する安全性指標の一つである k-匿名性 [18] に基づいて加工することを考える．k-匿名性とは，簡単にいえば同じ属性（準識別子）をもつ人がデータベース中に k 人以上存在する，という条件[1]により安全性を定義した指標である．

第 2 表は，$k=10$ の強度で k-匿名性を満たすように第 1 表を加工した例である．試験に落第した人の数が男女のいずれも 10 人に満たないため，この二つを一般化して性別を「※」（非開示）とした．今度は A 君のプライバシーは守られるであろうか．

[1] この例では試験の合否が攻撃者にとって知りたい情報（プライバシー）なので，性別を準識別子と考える．

第 2 表　男女別の試験結果 $(k=10)$

性別	合否	人数 $(n=20)$
女性	合格	10 人
※	落第	10 人

やはり守られず，A 君は試験に落ちたことがわかる．これは，この例におけるプライバシー開示は k の値が不足しているから発生したわけではなく，k-匿名性が「想定していない」攻撃により発生したためである．k-匿名性が保証する安全性はリンケージ攻撃 (linkage attack) とよばれる攻撃に対するものに限られ，非構造的ゼロからの開示については何も保証しない．

2.3 安全性指標の「改善」

k-匿名性だけを満たしても非構造的ゼロの問題によりプライバシーが開示されることがわかった．そこで「非構造的ゼロを含まない」ことを要件に追加して改善することを考える．これを仮に「k-匿名性（改）」としよう．

第 3 表はこの k-匿名性（改）を満たしている．先ほどまでのような非構造的ゼロからの開示は発生しない．

第 3 表　試験結果（非構造的ゼロの対策ずみ）

性別	合否	人数 $(n=20)$
※	合格	10 人
※	落第	10 人

ただし，性別の情報も完全に失われてしまったので，第 3 表からは合否それぞれの人数しかわからない．

そこで，さらに別の観点として「所属する部活動」別にデータをまとめることを考える．このクラスでは全員が必ず一つ以上の部活動に所属し，掛け持ちも許される．なお，A 君は運動部と文化部を掛け持ちしている．

A 君のクラスで運動部に所属している人の試験結果は第 4 表の通りであった．これも非構造的ゼロからの開示は発生しない．

第 4 表　運動部に所属する人の試験結果

合否	人数
合格	5 人
落第	7 人

また，文化部に所属している人の試験結果は第 5 表の通りであった．これも大丈夫そうに見える．

第 5 表　文化部に所属する人の試験結果

合否	人数
合格	5 人
落第	6 人

さて，それぞれの試験結果から非構造的ゼロによる開示は発生しないことが確認できた．A 君のプライバシーは今度こそ守られるだろうか？

2.4 モザイク効果問題

残念ながら，やはり A 君が試験に落ちたことはわかってしまう．第 3 表〜第 5 表のそれぞれ単体からは A 君の合否はわからないが，それらを重ね合わせると，

- 運動部と文化部を掛け持ちしている人は 3 人で，
- そのいずれも試験に落第した．

ということが明らかになる（そうでないと矛盾が生じる）ため，掛け持ちをしている A 君の落第が確定する．

このように，複数の（それぞれ安全そうに見える）問合せ結果を「重ね合わせ」ることによりプライバシー開示が発生する問題はモザイク効果 (mosaic effect)[17] 問題とよばれる[1]．モザイク効果によるプライバシー開示のリスクは，扱う属性が多い大規模なデータに対してより大きく表れ，たとえば公的機関によるオープンデータの充実に向けた課題として議論されている [3]．上記の k-匿名性（改）は，単体での非構造的ゼロの問題は考慮しているが，モザイク効果は想定していなかったため，これを用いたプライバシー開示に対して脆弱となった．

したがって，k-匿名性（改）もまだ安全性指標として不十分である．さらに改善が必要となるが，どこまで「改善」すれば十分といえるだろうか．

3. 差分プライバシー

前章の議論の通り，ある特定の背景知識や攻撃能力をもつ攻撃者によるプライバシー開示のみを考慮して安全性指標を構築しても，その指標が想定していない攻撃に対しては脆弱になる．また，それをさらに別の攻撃に対応できるよう改善を繰り返しても「想定外」をなくすことは難しく，際限なき「いたちごっこ」になりかねない．

そこで，データベースに含まれる個人のプライバシーが適切に守られているかどうかを判断するためには，

- 任意の背景知識をもつ攻撃者による，
- どのようなアルゴリズムによる攻撃に対しても，

プライバシーの開示を一定以下に抑えることを保証できる安全性指標と，これによるプライバシー保護方式 Q' やその出力 $Q'(D)$ の評価が重要となる．

このような性質をもつプライバシーに関する安全性指標，すなわち特定の攻撃に対してのみ “ad hoc” にしか安全性を保証しない指標ではなく，任意の攻撃に対して “ad omnia” に安全性を保証できることを目的として構築された安全性指標が差分プライバシーである．[7]

3.1 差分プライバシーの定義と意味

差分プライバシー [6,8] は，識別不能性に基づくプライバシー保護指標の一種である．差分プライバシーにおいて，あるプライバシー保護方式 Q' の安全性はパラメータ ϵ を用いて以下のように定義される．

【定義 1】 任意の隣接したデータベース D_1 および D_2 $(D_1, D_2 \in \mathcal{D})$ に対し，ランダム化関数 (randomized function) $Q' : \mathcal{D} \to \mathcal{R}$ が下式を満たすとき，Q' は ϵ-差分プライバシー (ϵ-DP) を満たす．ただし，ここで S は Q' の出力空間 $\mathcal{R}$ の任意の部分空間である $(S \subseteq \mathcal{R})$.

$$\Pr[Q'(D_1) \in S] \leq e^{\epsilon} \cdot \Pr[Q'(D_2) \in S] \tag{1}$$

さて，**定義 1** は何を意味するだろうか．一見だけではプライバシー保護と何の関係があるかすらわかりにくい．そこで，以下に（なるべく）直感的な説明を試みる．

基本的な考え方としては，**定義 1** は以下を意味する．

- あるデータベース D_1 と，D_1 からある人（A さんとする）の情報を抜いた（もしくは他人の情報と入れ替えた）データベース D_2 があるとする．
- D_2 は A さんの情報を含まないので，そこから得られた問合せ結果 $Q'(D_2)$ は，A さんのプライバシーを開示しようがない（安全である）．
- （A さんの情報を含む）データベース D_1 から得られた問合せ結果 $Q'(D_1)$ が，$Q'(D_2)$ とほとんど見分けがつかないなら，$Q'(D_1)$ も A さんにとって安全といえる．どのくらい見分けがつかなければ安全とするかは ϵ の値により定める．
- 任意の D_1 と D_2 の組合せに対して上記が成立するなら，A さん以外も含めたすべての人にとって Q' は安全といえる．
- このとき，Q' は ϵ-差分プライバシーを満たす．

以降において，**定義 1** がなぜ上記を意味するかについて順番に説明する．

まず，D_1 と D_2 は「隣接」したデータベースである．隣接の定義には 2 種類ある[2]が，ここでは D_1 と D_2 は置換の関係で隣接する，つまり D_1 に含まれているある一人（A さんとする）に関するレコードが，D_2 では別のレコードに置換されているとする．

Q' はデータベースへの問合せを表す．ここでは「データベース中に含まれる，ある条件を満たすレコードの数」を数える計数問合せを考える．ただし，プライバシー保護のため Q' は実際の計数結果に対して整数ノイズを加える（ランダム化する）ものとする．

計数問合せの結果はレコード数（自然数）なので，そこに整数ノイズを加えた出力の値域 $\mathcal{R}$ は（負値を含む）整数の集合となる[3]．なお，このように $\mathcal{R}$ が離散空間となるとき，(1) 式が成立することは任意の $s \in \mathcal{R}$ に対して下式が成立することと等価である[4]．

[1] 差分による開示 (disclosure by differencing)[14] の一種ともみなせる．

[2] D_1 と D_2 が，1 レコードの追加／削除の関係にあるとする定義と，置換の関係にあるとする定義 [16]．ここでの問合せ（単一の計数問合せ）ではどちらの定義を用いても議論は変わらない．

[3] ノイズの加算により，負の値をとりうることに注意．

[4] (2) 式における Pr を確率密度と読み替えれば，$\mathcal{R}$ が連続空間であっても特殊な場合を除いて成立する．

$$\Pr[Q'(D_1)=s] \leq e^{\epsilon}\cdot\Pr[Q'(D_2)=s] \quad (2)$$

Q' はランダム化により確率的にふるまうため，$Q'(D_1)$ と $Q'(D_2)$ はいずれも確率変数となる．(2)式 は，$Q'(D_1)$ と $Q'(D_2)$ の分布が「ほとんど同じ」であること，すなわち任意の s に対して $\Pr[Q'(D_1)=s]$ と $\Pr[Q'(D_2)=s]$ の比が最大でも $e^{\epsilon}:1$ を超えないことを求める．なお，このような性質を満たすノイズとしては，両側幾何分布に従う乱数が挙げられる[1]．

したがって，ϵ が十分に小さければ，問合せ Q' により s を得た攻撃者は，s がどのような値であっても，それが（A さんのデータを含む）D_1 に対してなされたものか，（別のデータを含む）D_2 に対してなされたものかを識別できない．つまり，s を得てもデータベースの中に A さんのデータがどのように含まれていたかわからないため，s は A さんのプライバシーを開示しない．

より厳密には，問合せ Q' により s を得たときに，

帰無仮説 s は D_1 から得られた $(s\sim Q'(D_1))$，

対立仮説 s は D_2 から得られた $(s\sim Q'(D_2))$，

とする仮説検定をしたとき，第 1 種の過誤が起こる確率 $\Pr_{\mathrm{FP}}$ と第 2 種の過誤が起こる確率 $\Pr_{\mathrm{FN}}$ の関係について，以下が成立する．[19,15]

$$\Pr_{\mathrm{FP}}+e^{\epsilon}\Pr_{\mathrm{FN}} \geq 1 \cap e^{\epsilon}\Pr_{\mathrm{FP}}+\Pr_{\mathrm{FN}} \geq 1 \quad (3)$$

D_1 と D_2 は（$\mathcal{D}$ 上の）「任意の」隣接したデータベースであるため，これはデータベースの中身がどのようなものであっても（たとえば D_1 には A さん 1 人のレコードしか含まれていなかったとしても）成立し，A さんではなくほかのレコードに対応する人，つまり B さんでも C さんでも任意の人に対して同様に成立する．

つまり Q' が ϵ-DP を満たすとき，ϵ が十分に小さければ，Q' は（$\mathcal{D}$ 上の）任意のデータベースに含まれるすべての人に対して，その人の情報に関する識別困難性，すなわちプライバシー保護を提供する．

3.2 ϵ はいくつなら安全か？

前節の説明において，「ϵ が十分に小さければ」ϵ-差分プライバシーを満たす問合せ Q' はプライバシーを保護することを示した．では，ϵ がどのくらいであれば「十分に小さい」といえるのだろうか？

ϵ が 0.1 とか 10 とかいわれても，それぞれがどの程度の安全性をもたらすか，**定義 1** から直感的には理解しにくい．満たすべき ϵ の具体的な基準は扱うデータの性質や活用の目的によって変わりうるとしても，なんの目安もないと「このデータをこういう目的で使うから ϵ はこの値に設定しよう」という合意形成も困難となる．

そこで，ϵ がいくつならどの程度の安全性が与えられるかをイメージしやすくするために，単一のレコードに対する二値問合せをランダム化し，「でたらめ」な回答を混ぜることにより得られる安全性について検討する．

ある人が Yes/No の二択で答えるアンケートを受けたとする．質問の内容はプライバシーに関わるためそのまま答えたくない．そこで，ある確率 p で「正直に」本当の答えを回答し，それ以外は「でたらめに」Yes/No を半々の確率で回答してよいとした．

このアンケートは二値のランダム化回答 (randomized response) に相当し，差分プライバシーを満たす．具体的には，p と ϵ との間に以下の関係が成立する．

$$p=(e^{\epsilon}-1)/(e^{\epsilon}+1) \quad (4)$$

ϵ にいくつか具体的な値を与えたとき，正直に答える確率 p がどのように変化するかを第 6 表に示す．

第 6 表　ϵ (および e^{ϵ}) と p の関係

ϵ	e^{ϵ}	p
0	1	0%
0.1	1.11	5%
0.2	1.22	10%
ln2 ($\simeq$0.7)	2	33%
ln3 ($\simeq$1.1)	3	50%
3	20.1	90%
5	148	99%
∞	∞	100%

たとえば $\epsilon=0.1$ とは，質問に対して 5% しか本当のことを答えず，95% の確率で「でたらめ」を答える場合の安全性に相当する．質問の内容によるかもしれないが，これは安全性が高いとして受け入れやすい数値と考えられる．その一方，$\epsilon\geq 5$ のとき，99% 以上の確率で本当のことを答えることに相当する．1% 未満の確率で回答がでたらめである可能性は残されている（決定的な開示が発生するわけではない）にしても，これで「安全である」と主張することはさすがに難しそうである．

なお，このシナリオは攻撃者にとって最も有利な状況である．$Q'(D)$ は攻撃対象以外の情報を含まず，攻撃者は（$Q'(D)$ が示す回答が本当か嘘かの）二択問題に正解すればよい．一般には，$Q'(D)$ には攻撃対象以外に関する情報も含まれるため攻撃者はなんらかの手段でその影響を排除しなければならず，攻撃目標も二択問題にまで絞りこまれているとは限らない．

そのため，一般的な問合せでは，第 6 表に示した p が与える以上の安全性を実質的に備える（安全性マージンをもつ）こともある．しかし，その有無や程度について差分プライバシーは何も保証しない．また，高次元のデータに対する問合せ結果からは情報の絞り込みがしやすい（攻撃対象以外の情報を排除しやすい）ため，大きなマージンは見込みにくい．したがって，この安全性マージンに対して過度に期待することは望ましくない．

[1] 幾何メカニズム [12] とよばれる．Laplace 分布に従う乱数もこの性質を満たすが，$\mathcal{R}$ は実数の集合となる．

3.3 合成定理

差分プライバシーが備える性質はさまざまなものが知られているが，そのうち重要な一つである合成定理 (composition theorem)[10] について簡単に説明する．

合成定理とは，複数の（差分プライバシーを満たす）問合せを合成したときに得られる安全性を保証する定理であり，具体的には以下で示される．

【定理 1】 それぞれ ϵ_i-DP を満たす n 個の問合せ Q'_i があり，Q' はそれらの Q'_i の列挙，つまり $Q'(D)=(Q'_1(D),Q'_2(D),\ldots,Q'_n(D))$ とする．このとき，Q' は $\sum_{i=1}^{n}\epsilon_i$-DP を満たす．

これは，たとえばあるデータベース D に ϵ_1-DP を満たす問合せ Q'_1 を適用して $Q'_1(D)$ を得たのちに，同じ D に対して ϵ_2-DP を満たす問合せ Q'_2 を適用して $Q'_2(D)$ を得たとき，つまり問合せ結果の「重ね合わせ」が発生するときでも，$(\epsilon_1+\epsilon_2)$-DP の安全性が保証されることを示す．つまり，第 **2** 章の例で示したようなモザイク効果によるプライバシーの決定的な開示が生じることはない．このことは，差分プライバシーの適用がモザイク効果問題の解決もしくは緩和の手段となりうることを示している．

なお，合成定理の特殊な場合として，上記の Q'_i が適用されるレコードの集合が互いに素であるとき，**定理 1** より強い安全性が与えられる．具体的には，このとき Q' は $(\max_i \epsilon_i)$-DP を満たし，これを並列合成定理とよぶ．

並列合成定理は，たとえば分割表（集計単位が互いに素な集計表）を作成する際に，表の次元数（セル数）が増えても各セルに加えるノイズの強度は一定でよい（安全性は変わらない）ことを保証する．この性質は，大規模な集計データのプライバシー保護にとくに有用である．

4. 課題と今後の展望

本稿では，プライバシー保護に関する適切な安全性指標の必要性とその難しさについて議論し，「安全に」安全性を測るための指標として注目を集めている差分プライバシーについて，その定義や安全性パラメータである ϵ がもつ意味について説明した．

本稿では差分プライバシーの実現方式（メカニズム）に関する議論は割愛したが，差分プライバシーを満たすだけならその実現はきわめて簡単である．たとえばクロス集計表などの分割表は，普通に分割表を作成した後に，表中のそれぞれの値（セル値）に対し，ϵ の値により定まる強度の Laplace ノイズを加えるだけで ϵ-DP が与えられる [9]．これを Laplace メカニズムとよぶ．また，いわゆる匿名化個票は，**3.2** 節のランダム化応答を，合成定理に基づいて個票データ上のすべての値に適用することにより得られる．これは SDC の分野で広く知られている，維持置換攪乱とよばれる手法に基づく PRAM[13] と等価である [1]．

ただし，Laplace メカニズムや PRAM などの素朴な適用により得られたデータは，（安全性は ϵ により保証されているとしても）有用であるとは限らない．たとえば Laplace メカニズムにより加工された分割表は，セル値として（本来ありえない）負の値を含んだり，小計などの部分和がノイズを多く含む（部分和の精度が低い）など，実用上の問題を含む．また，PRAM により十分な安全性をもつ匿名化個票を得ようとすると，維持確率（元の個票での値を「正直に」反映させる確率）が悲観的なほど小さい値となる．これらの問題の詳細や，その改善に向けた議論については文献 [21–23] やその参考文献などを参照されたい．また，分割表や匿名化個票以外にも，さまざまな応用において有用性を向上させるためのメカニズムの改善や提案が進められている [8,20]．

差分プライバシーは厳しすぎる（有用性が不足する）と指摘されることもある．公開しようとするデータが細かすぎる場合（たとえば履歴データを公開しようとする場合など），適切な ϵ を満たしたうえで高い有用性をもつデータを作成することは困難である．しかし，このようなデータはそもそもプライバシーを保護すること自体が困難であり，差分プライバシーは「適切に」そのリスクの高さを反映しているだけとも考えられる．

また，不適切な方法で差分プライバシーを満たそうとすると，必要以上に有用性を損なうことにも注意が必要である．上記で示した通り，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おり，（安全性は同じでも）有用性はそれぞれ異なる．応用にしたがって最適なメカニズムを選択することが重要であるとともに，さらなる有用性改善に向けたメカニズムの発展が期待される．

なお，**3.2** 節で触れたように，差分プライバシーが保証する安全性と，実際に得られる安全性との間にはマージンが存在しうる．何の保証もなしにこの安全性マージンに依存することは危険であるが，たとえば合成定理により得られた ϵ が保証する安全性と実際の合成結果が備える安全性との間には，ϵ だけでは表現しきれない差分が存在することが明らかにされており，それに基づく安全性定義（**定義 1**）の緩和が議論されている [15,2]．これらの検討の進展により上記の安全性マージンを「安全に」活用できるようになれば，差分プライバシーに基づくデータのさらなる有用性向上に繋がると考えられる．

ビッグデータは流行期から実用期に入って久しく，AI 技術の流行やオープンデータの進展により活用されるデータの規模も複雑性も増している．これは，モザイク効果などにより誰も意図しない形で深刻なプライバシー開示が発生するリスクを高めることにも繋がる．適切な安全性に基づいた健全なデータ活用の普及に向け，今後の関連技術の進化や議論の進展に期待したい．

[1] 維持確率 p の値は (4) 式とは異なる．

謝　辞

本稿執筆にあたり数々の貴重なコメントを頂きました，中央大学経済学部 伊藤伸介教授，北海道科学大学工学部 本郷節之教授に感謝いたします．

(2018 年 9 月 3 日受付)

参 考 文 献

[1] R. Agrawal and R. Srikant: Privacy-preserving data mining; *ACM SIGMOD Record*, Vol. 29, No. 2, pp. 439–450 (2000)

[2] M. Bun and T. Steinke: Concentrated differential privacy: Simplifications, extensions, and lower bounds; *Theory of Cryptography*, pp. 635–658 (2016)

[3] J. Czajka, C. Schneider, A. Sukasih and K. Collins: Minimizing disclosure risk in HHS open data initiatives; *Technical Report 09/29/2014, U.S. Department of Health and Human Services* (2014)

[4] A. N. Dajani, A. D. Lauger, P. E. Singer, D. Kifer, J. P. Reiter, A. Machanavajjhala, S. L. Garfinkel, S. A. Dahl, M. Graham, V. Karwa, H. Kim, P. Leclerc, I. M. Schmutte, W. N. Sexton, L. Vilhuber and J. M. Abowd: The modernization of statistical disclosure limitation at the U.S. Census Bureau; *Proc. Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality 2017* (2017)

[5] J. D. Ferrer: A survey of inference control methods for privacy-preserving data mining; *Privacy-Preserving Data Mining: Models and Algorithms*, pp. 53–80, Springer (2008)

[6] C. Dwork: Differential privacy; *Proc. 33rd Intl. Conf. Automata, Languages and Programming - Volume Part II*, Vol. 4052 of *LNCS*, pp. 1–12 (2006)

[7] C. Dwork: An ad omnia approach to defining and achieving private data analysis; *Proc. 1st Intl. Conf. Privacy, Security, and Trust in KDD*, pp. 1–13 (2007)

[8] C. Dwork: Differential privacy; A survey of results; *Proc. 5th Intl. Conf. Theory and Applications of Models of Computation*, pp. 1–19, Springer (2008)

[9] C. Dwork, F. McSherry, K. Nissim and A. Smith: Calibrating noise to sensitivity in private data analysis; *Proc. 3rd Conf. Theory of Cryptography*, Vol. 3876 of *LNCS*, pp. 265–284 (2006)

[10] C. Dwork and A. Roth: The algorithmic foundations of differential privacy; *Foundations and Trends in Theoretical Computer Science*, Vol. 9, Nos 3–4, pp. 211–407 (2014)

[11] B. C. M. Fung, K. Wang, R. Chen and P. S. Yu: Privacy-preserving data publishing; *ACM Computing Surveys*, Vol. 42, No. 4, pp. 1–53 (2010)

[12] A. Ghosh, T. Roughgarden and M. Sundararajan: Universally utility-maximizing privacy mechanisms; *SIAM J. Computing*, Vol. 41, No. 6, pp. 1673–1693 (2012)

[13] J. Gouweleeuw, P. Kooiman, L. Willenborg and P. P. de Wolf: The post randomisation method for protecting microdata; *Quaderns d'Estadística i Investigació Operativa (QÜESTIIÓ)*, Vol. 22, No. 1, pp. 145–156 (1998)

[14] A. Hundepool, J. Domingo-Ferrer, L. Franconi, S. Giessing, E. S. Nordholt, K. Spicer and P.-P. de Wolf: *Statistical Disclosure Control*, John Wiley & Sons (2012)

[15] P. Kairouz, S. Oh and P. Viswanath: The composition theorem for differential privacy; *Proc. 32nd Intl. Conf. Machine Learning (ICML'15)*, Vol. 37, pp. 1376–1385 (2015)

[16] D. Kifer and A. Machanavajjhala: No free lunch in data privacy; *Proc. 2011 Intl. Conf. Management of Data (SIGMOD '11)*, ACM, pp. 193–204 (2011)

[17] H. J. Smith, S. J. Milberg and S. J. Burke: Information privacy: Measuring individuals' concerns about organizational practices; *MIS Quarterly*, Vol. 20, No. 2, pp. 167–196 (1996)

[18] L. Sweeney: K-anonimity: A model for protecting privacy; *Intl. J. Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 10, No. 05, pp. 557–570 (2002)

[19] L. Wasserman and S. Zhou: A statistical framework for differential privacy; *J. American Statistical Association*, Vol. 105, No. 489, pp. 375–389 (2010)

[20] T. Zhu, G. Li, W. Zhou and P. S. Yu: Differentially private data publishing and analysis: A survey; *IEEE Trans. Knowledge and Data Engineering*, Vol. 29, No. 8, pp. 1619–1638 (2017)

[21] 寺田, 山口, 本郷: 大規模高次元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; 2017 年暗号と情報セキュリティシンポジウム (SCIS2017) 予稿集, 電子情報通信学会, pp. 1–8 (2017)

[22] 寺田, 山口, 本郷: 匿名化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, Vol. 58, No. 9, pp. 1483–1500 (2017)

[23] 寺田, 鈴木, 山口, 本郷: 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, Vol. 56, No. 9, 1801–1816 (2015)

[24] 総務省 統計局: 日英統計用語集; https://www.e-stat.go.jp/classifications/terms/90

著 者 略 歴

寺田（てらだ）　雅之（まさゆき）

1995 年 神戸大学大学院工学研究科 修士課程修了．同年 日本電信電話（株）入社，2003 年 （株）NTT ドコモへ転籍，2008 年 電気通信大学大学院 博士後期課程修了，2009 年より現職．プライバシー保護技術とその応用，携帯電話ネットワークからの人口推計技術，人口からの交通予測技術の研究に従事．博士（工学）．2015 年度情報処理学会論文賞，山下記念研究賞受賞．情報処理学会，電子情報通信学会会員．