2つの匿名化情報の組み合わせによる k-匿名度の定式化に関する考察

秋山 寛子[1]　中山 雅哉[2]　加藤 朗[1]　砂原 秀樹[1]

概要：　近年，IC カード等による公共交通機関の利用記録やクレジットカード等による購買記録，スマートフォン等による移動記録などの様々なライフログデータや，小型気象センサ等による環境情報など複合したビッグデータを利用した社会的サービスに関心が高まっているが，その利活用のためにはパーソナル情報を安全に取り扱うことが不可欠である．プライバシ保護技術の一つとして，個人を特定・識別できないような形にパーソナル情報を加工する匿名化技術がある．しかし，匿名化した情報であっても，他の情報と突き合わせることで個人を識別できてしまう再識別可能性のリスクがあることが問題となっている．このリスクは，誰がどの情報を持っているかのコンテキストに依存するため，完全な安全性を保証する匿名化情報の作成は不可能となる．このため，現在は，パーソナル情報の利用は匿名化技術とともに，データの使用や保管に関するルールなどを契約として併せて利用している．しかし，契約は文言による表記であるため，その意味の解釈に個人差が生じるため曖昧性が生じ，契約の遵守を客観的に示す方法が難しい．このため本研究では，パーソナル情報の匿名性を客観的に表す指標を定式化することで，この問題の解決を試みる．

A Study on the Formulation of k-anonymity with a Combination of 2 Anonymized Personal Information

Hiroko Akiyama[1]　Masaya Nakayama[2]　Akira Kato[1]　Hideki Sunahara[1]

1. はじめに

位置情報や購買履歴等の，人の活動履歴は，大きな価値を生み出す可能性を持っている．しかし，これらの情報は，パーソナル情報であるため，その取り扱いには注意が必要である．

その活用に際しては，プライバシを保護する必要があり，情報利用の倫理や勧告等と，技術とを組み合わせて実現する，そこで，技術が実現できることと，その程度を明確にする必要がある．プライバシ保護の技術的な方法の一つに，匿名化技術がある．匿名化とは，特定の個人を識別できないような状態に，パーソナル情報を加工することである．安全な匿名化情報の生成のために，個人を識別できないように適切に匿名化処理をほどこす．そのために，パーソナル情報そのものの値を加工するという処理と，匿名化処理を施された情報が，個人を特定できない状態であるということを確認する作業が必要である．本稿では，後者の問題にフォーカスし，匿名化情報の突き合わせによって，個人が識別できる条件を定式化することを目的とし，匿名化情報を k-匿名度を用いて表し，匿名化情報の安全性確認のための指標とする方法について述べる．

匿名化情報の安全性を検証するために，匿名度指標の 1 つである k-匿名度を用いる．k-匿名性とは，あるレコードに対して，少なくとも他の $k-1$ レコードと区別できないことを保証するというものである．複数の匿名化情報が組み合わされた場合，その匿名度は，個々の匿名化情報の k 匿名度に対して，どのような値を取るのかを，匿名化情報とその元となるデータの分布等を把握した上で，算出する方法を定義する．

第 2 章では，匿名化情報の安全性指標の設定についての考察について述べる．第 3 章では，複数の匿名化情報を手にした人が個人を識別できる可能性の度合いについて考察

1　慶應義塾大学大学院メディアデザイン研究科
2　東京大学情報基盤センター

し，定式化を行う．第４章では，匿名化情報を開示する人が安全性のチェックとして匿名度を使用することについて考察し，定式化する方法を提案する．第５章では，まとめと今後の課題について述べる．

2. 匿名化情報の安全性指標についての考察

2.1 プライバシ保護とパーソナル情報

OECD ガイドライン [1] によると，「『個人データ』とは，識別されたまたは識別されうる個人（データ主体）に関するすべての情報」とされている．パーソナル情報とは，個人を表す情報と，位置情報や購買履歴等の，活動履歴とがある．さらに，個人を表す情報は，「識別子」と「準識別子」とがある [2]．識別子とは，個人を特定可能な識別情報であり，名前や電話番号等がある．準識別子とは，性別や年齢などの属性情報である．これらの情報は，数値であったり，文字列であったり，さまざまなデータフォーマットを持つ．数値であれば精度，文字列情報であれば，その文字列の表すものが何であるかにより，匿名化の処理内容は異なってくる．また，識別子以外の情報は，それ自体では，個人を特定することは出来ないが，複数の情報を組み合わせることで，個人を特定できる可能性を持っている [3]．そこで，匿名化情報の安全性を確認するという問題に対して，どのような匿名化情報を組み合わせると，個人の識別が可能になるのかという問題を解決する方法について考察する．

また，保護すべきプライバシの定義としては，「実質的個人識別性」という考え方があり，パーソナルデータの利用・流通に関する研究会報告書 [4] によると，「取得などの際に特定の個人が識別されなかったとしても，他のパーソナルデータとあわせて分析されること等により，特定の個人が識別される可能性があることについて，十分に配慮する必要あり」とされている．したがって，個人を特定できないような情報であっても，データの突き合わせによる識別リスクを明確にする必要がある．

2.2 プライバシ保護における懸念事項

現状、プライバシの保護は，契約と技術との組み合わせで行っている．契約とは，使用についてのルールであり，情報の使用用途を明確にする，他への譲渡を禁じる，指定した方法以外の利用を禁じる，等の内容で構成される．一方，技術の面では，通信や情報処理全般の実装を担っており，情報へのアクセス制御や，認証システム，パーソナル情報の匿名化処理などが挙げられる．本研究では，技術的手法により実現できるプライバシ保護について明確にすることを目的とし，匿名化情報の安全性を科学的に示す方法についての検討を行う．

契約と技術によるプライバシ保護について，それぞれについて懸念すべき点が考えられる．まず，契約についてであるが，

- 「言葉」での表現のため，表現の捉え方に差異が生じる
- 情報の使用の同意の取り方が明確でない，統一されていない．
- プライバシ懸念についての事項が，個々人によって異なるため，それぞれに対応することが難しいため，情報提供者の満足・納得のいく形でのプライバシ保護とならない

などが挙げられる．一方，技術の中でも匿名化情報についてであるが，

- 他の情報との突き合わせによる再識別可能性のリスク
- 安全性の客観的な提示指標がない
- 匿名度と情報エントロピーとのトレードオフ

などが挙げられる．

上記のような懸念項目の解決のためには，匿名化情報の状態を表す客観的な数値が有用であると考える．なぜならば，契約における言葉の曖昧性をなくすことができ，また，技術的に求める匿名化情報の生成のために利用することが可能となるからである．

2.3 匿名化処理と匿名度指標

主な匿名化処理と匿名度指標を以下に示す．

2.3.1 匿名化処理

匿名化処理とは，元のデータが個人を特定できるようなものであったり，個人の性質を表すものであるとき，個人を特定できないように値を加工する処理のことである [5]．

- 記号化
  データそのものを，暗号化や文字列置換などにより置き換える操作
- あいまい化
  データの値を，グルーピングや値域の変更により曖昧にする操作
- 削除・切り落とし
  情報の一部あるいは全部を消去する操作

2.3.2 匿名度指標

他の個人とどの程度区別がつかないかの指標について，有名なものを挙げる [6]．

- k-anonymity（k-匿名度）
  あるデータの集合において，どの要素を取り出しても，他の k 個と区別がつかない状態
- l-diversity（l-多様性）
  ある条件を満たすレコードの任意の属性において，バリエーションが l 個以上ある状態
- t-closeness（t-近似性）
  既知の分布との距離が t 以下である状態

本稿の実験で行う匿名化処理はあいまい化，匿名度指標は k-匿名度とする．

2.4 特定と識別

特定とは，ある個人とある情報とが一意に結びつくことである．対して識別とは，情報の持ち主が誰であるかはわからなくても，他の情報と区別できることである．技術検討ワーキンググループ報告書 [7]（2013 年 12 月）によると，匿名化技術により加工・作成される情報のカテゴリーを以下としている．

(1) 識別特定情報
　個人が識別されかつ特定される状態の情報（すなわち「個人情報」）（それが誰か一人の情報であることがわかり，さらにその一人が誰であるかわかる情報）

(2) 識別非特定情報
　一人ひとりは識別されるが個人が特定されない状態の情報（それが誰か一人の情報であることがわかるが，その一人が誰であるかまではわからない情報）

(3) 非識別非特定情報
　一人ひとりが識別されない（かつ個人が特定されない）状態の情報（それが誰の情報であるかがわからず，さらに，それが誰か一人の情報であることがわからない情報）

さらに，非識別非特定情報は，個人の識別化の困難性という指標において，程度には大きな差がある [7]．そこで，本研究では，個人識別かの困難性を，情報の突き合わせによる再識別性を突き合わせたときの k-匿名度を用いて表す方法により，その指標を提示する方法について提案する．

2.5 k-匿名度指標の定式化

本稿では，匿名化情報の匿名性を数値として表す手法を提案する．この手法の利点は，匿名性を客観的に評価のできる数値として表すことで，様々な場面で利用が可能な点である．以下に活用例を挙げる．

- 文言による契約内容の曖昧性をなくすことができ，実施内容の確認や評価を正しく行うことができる．
- データのフォーマットや種別に依らず，匿名化された状態や，情報加工方法によって，匿名度を表すことができる．
- プライバシ保護における安全についての対策を，技術で守る側および契約の履行で守る側の双方が活用できる．

3. 2つの匿名化情報の突き合わせによる k-値の期待値の定式化

本章では，元データの値は知らずに匿名化情報を複数取得し突き合わせたときの，匿名度を求める方法について考察する．また，データ間の関係を用いた匿名度テーブルの作成について検討した．

3.1 匿名化情報の突き合わせによる再識別可能性定量的表現方法

パーソナル情報の母集団が同じ場合，匿名化された状態であってもデータ間に相関があり，その関係が既知である場合，

- ある値に対応する値を予測されてしまう
- 突き合わせたときの匿名度が低い場合，個人を識別できる危険性がある

というリスクが考えられる．以下に例を示す．

例：あるクラスの身長と体重の情報を別々に入手した．
　身長 $x$cm 台で体重 $y$kg 台の人は何人か？

身長	人数
140cm 台	4
150cm 台	20
160cm 台	20
170cm 台	6

体重	人数
30kg 台	5
40kg 台	25
50kg 台	15
60kg 台	5

一方の値を入手できれば他方の値を “ある程度” の度合いで推定することは可能である．その推定できる度合いを把握することができれば，匿名化情報を突き合わせたときの安全性（匿名度）の指標とすることが可能となる．

3.2 定式化

匿名化処理はあいまい化とし，その処理は，集合の要素に対して，あるルールについて部分集合をつくることとする．例を図 1 に示す．関係 R のある 2 つの属性データを

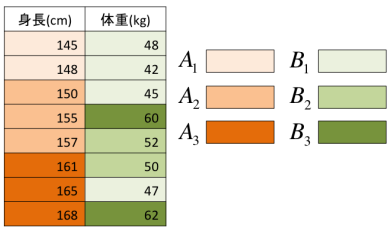

図 1　匿名化処理と集合の分割

A, B とする．属性の要素を集合と考え，A, B の要素数を n とすると，

$$A = \{a_1, a_2, \cdots, a_n\}$$
$$B = \{b_1, b_2, \cdots, b_n\}$$

と表す．これら 2 つの集合 A, B をそれぞれ s, t 個に分割し，部分集合を作る．匿名化情報 A', B' はそれぞれの部分集合を用いて，

$$A' = \{A_1, A_2, \cdots, A_s\}$$

$$B' = \{B_1, B_2, \cdots, B_t\}$$

と表す．

2 つの匿名化情報を突き合わせた場合の匿名度とは，A の要素が A_i に該当し，かつ，B の要素が B_j に該当する要素数である．これを，同時確率 $P(A_i, B_j)$ で表す．これは，周辺確率 $P(A_i), P(B_j)$ と，条件付き確率 $P(A_i|B_j), P(B_j|A_i)$ を用いて表すと，

$$P(A_i, B_j) = P(A_i|B_j) \times P(B_j) = P(B_j|A_i) \times P(A_i)$$

となる．A_i かつ B_j である人数は，同時確率に全体数をかけたものとなる．

$$n(A_i \cap B_j) = P(A_i, B_j) \times n$$

したがって，匿名化情報 A', B' の匿名度は，A, B のすべての部分集合の組み合わせの同時確率の最小値となり，以下のように表せる．

$$min(n(A_i \cap B_j))$$

3.3 k-値の期待値テーブルの作成

3.3.1 データ間の関係を用いた k-値の定式化

前節にて，属性 A の要素が条件 A_i に該当し，かつ，属性 B の要素が条件 B_j に該当する確率を，同時確率を用いて表した．求めたい確率（同時確率）を計算するのに必要な情報は，周辺確率と条件付き確率である．周辺確率は，集合の分割の要素数，つまり，匿名化処理の状態（あいまい化処理の結果）により計算が可能である．それに対して，条件付き確率の部分は，取得できる情報からはわからない．そこで，条件付き確率の部分に組み合わせるデータ間の関係を代入する方法について考える．

3.3.2 期待値テーブルの作成とリスク度の評価

2 つの属性情報 A, B について，各々の匿名化と，その情報間の関係の特性を計算することを考える．

まず，属性情報をそれぞれルールによって部分集合を作りグルーピングする．ルールとは，例えば，年齢の情報であれば 10 歳毎に区切り，「20 代」のようにする等である．次に，2 つの情報間の関係について，統計的な性質を表す値を計算する．相関係数や回帰直線を求める．これらの値により，突き合わせたときの匿名度の期待値を表すテーブル $T(x, y)$ を作成する．行方向に A の部分集合 $A_1, A_2, \cdots, A_s$ を並べ，同様に列方向に B の部分集合 $B_1, B_2, \cdots, B_t$ を並べる．その行と列の交差する部分に突き合わせたときの匿名度の期待値 $n(A_i \cap B_j)$ を記入する．

テーブルの安全性の評価は，l-diversity を用いて行う．ある 1 つの行（あるいは列）に注目し，すべての列（あるいは行）の値を比較する．すべての列に均等かつ同値の期待値が存在する場合，リスクは低いと言える．反対に，期待値が一つの列にしか存在せず，かつその期待値が小さな値である場合，l 値は 1 であり，また，該当する人数が少ないため，リスクは高いと言える．

このテーブルの値をもとにして，テーブル全体の危険度を判断し，データの開示を検討することが可能となる．組み合わせるデータの種別に依って，数値の表す危険度は異なってくるため，検討の上での客観的な指標として有用であると考える．

3.3.3 期待値テーブル作成と評価の例

アヤメの花弁の幅と長さのデータを用いて例を示す．図 2 は，x 方向へ花弁の幅，y 方向へ花弁の長さの値をプロットしたものである．相関係数や回帰直線の値から，k の期待値を表すテーブル $T(x, y)$ を作成する（図 3）．リスクの

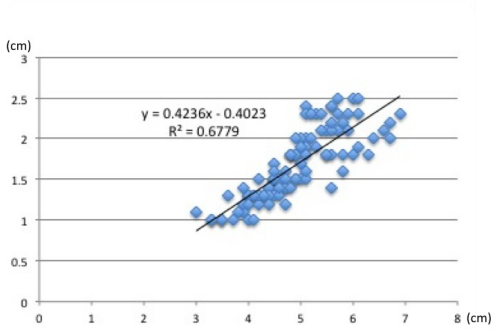

図 2　アヤメの花弁の幅と長さ

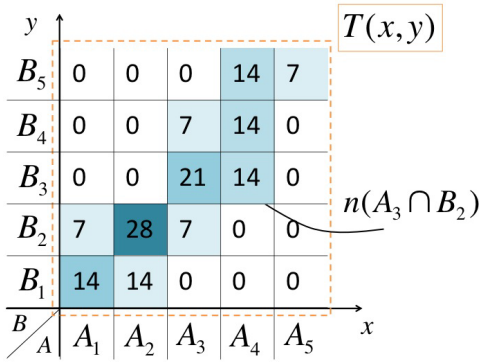

図 3　匿名度の期待値テーブル

評価については，l-diversity 値をみることで行う．

- $A = a_2$ の場合

  $B = b_1 \cdots 20\%$

  $B = b_2 \cdots 60\%$

  $B = b_3 \cdots 20\%$

- $A = a_3$ の場合

  $B = b_2 \cdots 33\%$

$B = b_3 \cdots 33\%$

$B = b_4 \cdots 33\%$

- $A = a_4$ の場合

  $B = b_4 \cdots 100\%$

- リスクが大きい順に並べると以下になる．

  $A = a_4 > a_2 > a_3$

ただし，これは列毎の評価であり，テーブル全体の評価は突き合わせるデータの種別という要素を考慮して行う必要がある．

4. 求める k-値を満たす匿名化情報生成についての考察

本章では，複数の匿名化情報を突き合わせたときの匿名度が求める値となるための匿名化処理を検討する．各属性情報に対する匿名化処理の変化により，それぞれを突き合わせたときの匿名度の変化について着目し，実験と考察を行う．

4.1 匿名化処理による匿名度と情報量の考察

匿名化とは各値を均整にすることと言い換えることが出来る．しかし，値を均整にするということは，データとしては特徴がない状態であるので，傾向を見たいデータであれば，特徴を押さえ込むようなあいまい化であると，データとしての価値がなくなってしまう．データの種別や傾向を把握した上で，データの値として意味があり有益かつ適当に匿名度を上げるような匿名化処理を行う必要がある．

また，匿名度 k_a である属性データ A と，匿名度 k_b である属性データ B があるとき，それらを突き合わせたときの匿名度 k_{ab} が極端に小さくなると，再識別可能性が上がってしまう．

以下に例を示す．

- 年齢と住所の情報あいまい化した匿名化情報をつくる
- $k = 100$ にする

処理方法	年齢	住所	匿名度	情報の特性
α	10 歳区切り	県単位	$k = 10,000$	曖昧度が高い 情報の価値が低い
β	3 歳区切り	町単位	$k = 100$	年齢の区切りが悪い 使いにくい

表 1　匿名化処理による匿名度と情報量の変化

ある匿名化処理 α,β を行った結果，表 1 となったとする．いずれの処理も，上に挙げたように，求める匿名度に近い値の匿名度を持ち，かつ，活用において価値のある匿名化情報の形である状態は作れないため，適当な情報であるとは言えない．そこで，適用する匿名化処理によって，組み合わせたときの匿名度がどのような値になるのかを，式で表すことを考える．

4.2 あいまい化処理による匿名度の変化の一例

実際のデータを用いて，匿名化と匿名度の実験を行う．東京大学空間情報科学センターの空間データ利用を伴う共同研究を行っており，「人の流れプロジェクト」[8] のデータを使用する．内容は，約 60 万人の属性情報 (職業，年齢など) と位置情報である．ここで使用するデータは，平成 20 年 10 月 1 日午前 8:00 のデータである．

年齢と職業の関係について，あいまい化を行う．何らかの職に従事している人を対象とし，人数は約 30 万人である．年齢は，5 歳から 80 歳までは 5 歳区切り，80 歳以上は同じとなっている．職業は，10 種の職種に分類されている．年齢をあいまい化し，職種の加工は行わない．ここで求める匿名化情報は，突き合わせたときの匿名度は $k = 100$ 以上であり，さらに，あいまい化処理において各グループは区切りのよい活用しやすい代表値をとるものとする．

- 元データ

  この場合，1 人に認識されるデータが出現しており，識別可能な状態であった (表 4).

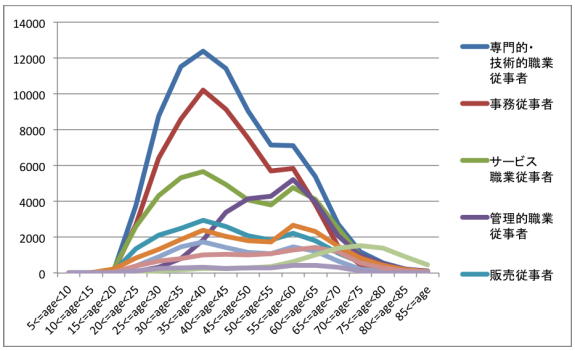

図 4　元データ

- 匿名化 1

  年齢を 10 歳区切りのデータにした (表 5)．すると，年齢については 10〜80 代の 8 種，職種 10 種の 80 パターンの組み合わせができた．匿名度が 1 であるデータが 1 件あった．また，総数が約 30 万人であるのに対し，匿名度が 1 桁であるデータの組み合わせが 3 件，6 件であった．

- 匿名化 2

  20 代以下を一つにまとめ，30〜69 歳までは 10 歳区切りにし，70 歳以上は 1 つにまとめた (表 6)．すると，1 桁および 2 桁の匿名度のデータの組み合わせはなくなった．3 桁の組み合わせは 12 通りあり，突き合わせの最小の匿名度は 165 となった．しかし，この 165 となったデータの職種は全体数が 3193 人であり，約 30 万人のうちの 1%程度の量であり，匿名度が 3 桁となるデータのうち 6 つがこの職種に該当している．

実際のデータにおいて，個々の属性情報の匿名度は大きく変えず，突き合わせたときの匿名度を増やすことが可能と

	10代	20代	30代	40代	50代	60代	70代	80代
専門的・技術的職業従事者	106	12,516	23,904	20,439	14,250	8,120	1,747	278
事務従事者	108	9,114	18,807	16,667	11,549	5,354	660	82
サービス職業従事者	208	6,907	10,994	9,026	8,576	6,639	1,360	290
管理的職業従事者	1	421	2,623	7,553	9,510	6,048	1,481	329
販売従事者	90	3,498	5,447	4,688	4,030	2,899	891	232
生産工程・労務作業者	231	2,156	4,273	3,883	4,400	3,835	1,260	270
運輸・通信従事者	30	1,307	3,207	2,582	2,531	1,863	296	25
その他職業	41	1,060	1,842	2,082	2,345	2,591	854	222
農林水産業従事者	15	183	448	527	991	2,409	2,917	1,360
保安職業従事者	9	375	616	531	718	779	156	9

表 **2**　匿名化 1

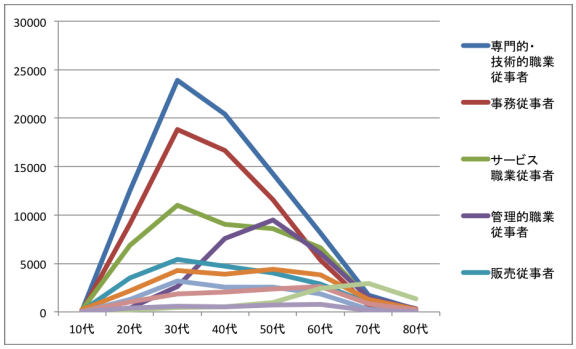

図 **5**　匿名化 1

	20代以下	30代	40代	50代	60代	70代以上
専門的・技術的職業従事者	12,622	23,904	20,439	14,250	8,120	2,025
事務従事者	9,222	18,807	16,667	11,549	5,354	742
サービス職業従事者	7,115	10,994	9,026	8,576	6,639	1,650
管理的職業従事者	422	2,623	7,553	9,510	6,048	1,810
販売従事者	3,588	5,447	4,688	4,030	2,899	1,123
生産工程・労務作業者	2,387	4,273	3,883	4,400	3,835	1,530
運輸・通信従事者	1,337	3,207	2,582	2,531	1,863	321
その他職業	1,101	1,842	2,082	2,345	2,591	1,076
農林水産業従事者	198	448	527	991	2,409	4,277
保安職業従事者	384	616	531	718	779	165

表 **3**　匿名化 2

なり，また，データ自体の価値も大きく損なうことなく匿名化を行うことが出来た．

4.3　定式化

関係 R のある 2 つの属性データを A, B とすると，各属性情報を列と考え，A, B の要素数を n とすると，

$$A = \{a_1, a_2, \cdots, a_n\}$$
$$B = \{b_1, b_2, \cdots, b_n\}$$

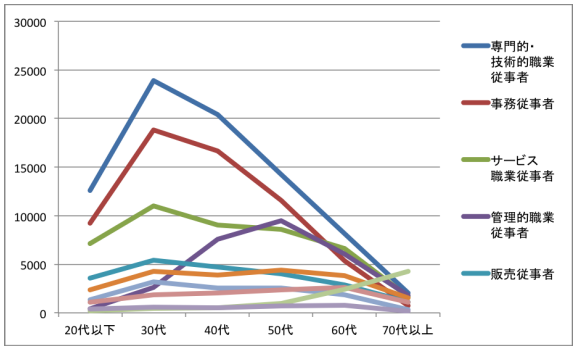

図 **6**　匿名化 2

$$AB = \{(a_1, b_1), (a_2, b_2), \cdots, (a_n, b_n)\}$$

と表す．A, B のそれぞれの要素に匿名化処理を行ったものを A', B' とする．A, B の要素の値のとりうる範囲について，s, t 個に分類する関数を f, g とする．

$$A' = f(A), B' = g(B)$$
$$A' = \{A'_1, A'_2, \cdots, A'_s\}$$
$$B' = \{B'_1, B'_2, \cdots, B'_t\}$$

となる．AB を A', B' を突き合わせたものを AB' とし，A, B の要素のうち A'_i, B'_j へ分類されるものを AB'_{ij} とし，

$$AB' = \{\{AB'_{11}, AB'_{12}, \cdots, AB_{1s}\},$$
$$\{AB'_{21}, AB'_{22}, \cdots, AB_{2s}\},$$
$$\cdots, \{AB'_{t1}, AB'_{t2}, \cdots, AB_{ts}\}\}$$

とする．このとき，AB' の各集合の要素の数が突き合わせたときの匿名度である．

A, B が数値データのように，ある値に対応している場合，3 次元空間を用いて匿名度 AB' の値を表す方法について考える．A, B の情報をそれぞれ，x, y 軸上へプロットし，その点を $z = 1$ とする．$n(AB'_{ij})$ の値は，A, B をそれぞれ分類した区間の積分（総和）になる．

前章で例に挙げたアヤメの花弁についてのデータを用いて，例を示す．集合 A をアヤメの花弁長，集合 B をアヤメの花弁幅の値とし，これらの値を x, y 方向へプロットする．プロットされた点は，$z = 1$ とする．ここで，2 種類の匿名化処理の結果について考える (図 7)．

(1) 匿名化 1

A, B のとる範囲に対して，それぞれ 20 等分する処理を，匿名化関数 f_1, g_1 とする．各区間に入る (x, y) の対の個数を z 軸に表す (破線)．この場合，z 軸の値が 1 になる点が存在するので，再識別可能と言える．

(2) 匿名化 2

A, B のとる範囲に対して，それぞれ 10 等分する処理を，匿名化関数 f_2, g_2 とする．各区間に入る (x, y) の

対の個数を z 軸に表す (実線)．この場合，z 軸の値は 0 または 1 より大きいので，ただ 1 人に識別できることはないと言える．

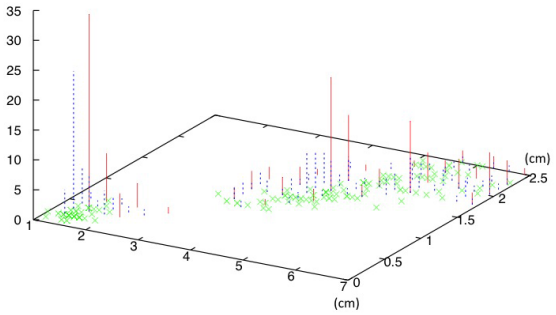

図 7　匿名化処理の違いによる匿名度の変化

5. まとめと今後の課題

匿名化情報のリスクについて検討し，再識別可能性という問題に焦点を当て，安全性を示す客観的な指標を作成することを目的とし，2 つの考察を行った．

1 つは，情報を取得する側について，2 つの匿名化情報の突き合わせについて，情報を取得した人がある条件を持つ人を再識別できるリスクについて検討し，同時確率を用いて定式化した，そして，それらの情報間の相関関係を組み込んだ定式化方法について検討した，2 つ目は，情報を開示する側に必要なこととして，複数の属性のパーソナル情報の開示による安全性の確認のため，行う匿名化処理に応じて，組み合わせたときの匿名度がどうなるかを定式化し，例示を行った．

今後の課題として，組み合わせる匿名化情報同士の相関関係を反映した匿名度の定式化として，様々な統計的特性を表す情報をパラメータとして，パラメータの設定ごとの実験結果をまとめる．また，本稿では 2 つの匿名化情報の突き合わせによる匿名度を定式化したが，3 つ以上の突き合わせについての拡張可能かの確認と，その定式化について検討を行っていく．

参考文献

[1] OECD プライバシーガイドライン，入手先 ⟨http://www.jipdec.or.jp/publications/oecd/⟩ 2013.12, (参照 2014.5.16)

[2] 村本俊祐，上土井陽子，若林真一，“k-匿名性を利用したデータ一般化によるプライバシー保護，” DEWS2007, 2007

[3] 高橋克巳，“匿名化技術の現状について，” 入手先 ⟨http://www.kantei.go.jp/jp/singi/it2/pd/wg/dai1/siryou2_3.pdf⟩, 2013.9, (参照 2014.5.16)

[4] “パーソナルデータの利用・流通に関する研究会報告書（案）～パーソナルデータの適正な利用・流通の促進に向けた方策～”, 入手先 ⟨http://www.soumu.go.jp/main_content/000225513.pdf⟩, 2013.5, (参照 2014.5.16)

[5] 松崎和賢，廣田啓一，高橋克巳，白井康之，“パーソントリップデータにおける匿名性に関する考察,” 土木計画額研究・公演集, 2010

[6] 白井康之，“データ匿名化に関する検討，” 入手先 ⟨http://hdl.handle.net/2115/48479⟩，2011.6, (参照 2014.5.16)

[7] 技術検討ワーキンググループ報告書，入手先 ⟨http://www.kantei.go.jp/jp/singi/it2/pd/dai5/siryou2-1.pdf⟩ 2013.12.10, (参照 2014.5.16)

[8] 東京大学人の流れプロジェクト，入手先 ⟨http://pflow.csis.u-tokyo.ac.jp/index-j.html⟩ (参照 2014.5.16)