

市区町村の統計表を考慮した都道府県単位の仮想個票の合成

○原田拓弥 村田忠彦（関西大学）

Synthetic Individual Households by Prefecture Considering Statistics of Local Governments

*T. Harada and T. Murata (Kansai University)

概要— 本研究では、より統計表と適合する仮想の世帯構成（以下、仮想個票）を合成するために、都道府県下の市区町村を一度に合成する手法を提案する。従来手法において、人口20万人未満の市町村の仮想個票を合成する場合、必要な統計表が公開されていないため、統計表の推計と対象とする市町村より広範囲で調査された統計表を縮小するなど、加工した統計表を多用していた。本研究の提案手法により、統計表の加工を避けるために、都道府県下の市区町村を一度に合成することで、これらの統計表の加工を最小限に抑えることが可能となる。実験結果より、市区町村単位で合成する従来手法の誤差を0.58倍から0.16倍に削減する仮想個票の合成が可能となった。

キーワード: Synthetic Reconstruct Method, Simulated Annealing, Statistics, Real-Scale Social Simulation

1 はじめに

国家的また国際的な災害対策や経済政策において、より精度が高く、きめの細かい対応が求められている。これらの分析と将来の可能性を可視化する社会シミュレーションへの関心が高まっている¹⁾。これまでの多くの社会シミュレーションではモデルを単純化せよというKeep It Simple, Stupid (以下、KISS原理)²⁾に基づいてモデル化されていた。しかし、KISS原理では現実社会の複雑な現象のモデル化は不可能であると指摘されている³⁾。そのため、可能な限り忠実に現実社会を模倣するモデルを用いた社会シミュレーションが期待されている。このようなモデルにおいて、モデルの粒度を現実社会に近づけるほど、エージェントの意思決定においても、可能な限り現実社会を模倣する必要がある。現実社会を可能な限り模倣するエージェントの意思決定の実現には様々な課題がある。その課題の1つがエージェントが保持する属性の設定である。エージェントの属性の設定に政府や行政が収集している戸籍や納税のデータを用いることができれば、現実社会と整合するエージェントの属性の設定が可能である。しかし、これらの市民のデータは個人情報保護やプライバシーの観点から利活用が困難である。このような状況から、政府統計をはじめとする利用可能な統計情報から、仮想的な属性を持つ個人で構成される人工社会を生成し、その人工社会の中でどのような事象が発生するかを観察する社会シミュレーションが行われるようになってきている。

統計情報に基づく個票データの合成¹⁾に関する研究の歴史は古く、Synthetic Reconstruction Method (SR法)⁴⁾として知られている。LenormandとDeffuanは⁵⁾、サンプルを用いて合成するSR法と、サンプルを用いない合成手法とを比較し、後者が個人と世帯をよりよく合成できていることを示した。

本研究が用いる合成手法もサンプルを用いない手法である。池田らは9種類の家族類型を対象に、Simulated Annealing (SA)法を用いて9つの統計表への適合を目的とした世帯集団を合成する手法⁶⁾を提案した。家

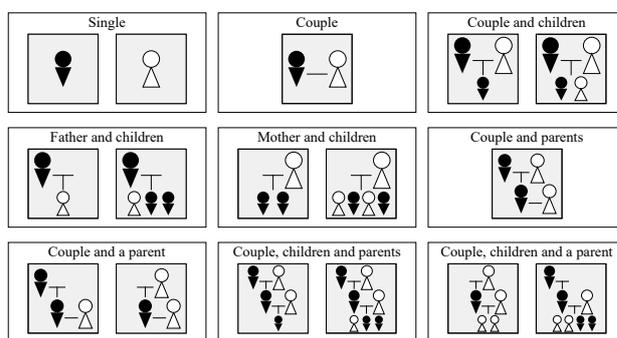


Fig. 1: 合成対象の家族類型

族類型とは、一般世帯²⁾を世帯員と世帯主の続柄により区分した分類である。柘井ら⁷⁾は、池田らが提案した合成手法を基に目的関数や合成手法の改良を行った。

しかし、池田らや柘井らの合成手法は日本の統計情報を500世帯や1000世帯に縮小した合成手法であった。柘井らの手法を用いて現実と同じ規模の世帯集団を合成する際には、家族類型毎の世帯数を統計情報に基づき、確率的に合成していたため、合成した家族類型毎の世帯数と統計情報に相違が発生した。また、Fig. 1に示す合成対象の9種類の家族類型のうち、「単独」世帯と「夫婦のみ」世帯のみ人口分布を用いて最適化を試みている。他の7種類の家族類型は、全ての家族類型を含んだ男女別の人口分布を用いて最適化を行っている。そのため、結婚できない年齢の夫婦が発生する可能性があった。

著者らは、これらの問題を解決し、対象地域と同じ規模の仮想個票を合成するために、対象地域と同規模で合成する手法を提案した^{8),9)}。まず、初期世帯の合成方法と最適化に用いる統計表の数を変更し、SA法による最適化の過程で個人の年齢を変更する際に用いる確率分布を変更した合成手法を提案した（年齢変更法⁸⁾。次に、初期世帯の人口分布を維持した上でSA法による最適化をする手法を提案した（年齢交換法⁹⁾。年齢変更法は設定した探索回数が少ない場合に年齢交換法より誤差を多く削減可能である。一方、年齢交換法

¹⁾ここで、個票データの「復元」ではなく「合成」という用語を用いている。復元の場合、実際の人口構成と同一の個票の復元が期待されるが、合成される個票はあくまでも統計的特徴が類似した仮想的な個票である。

²⁾一般世帯とは、住居と生計を共にしている人の集まりや一戸を構成して住んでいる単身者、間借り、下宿、独身寮に住んでいる単身者の世帯である。

は設定した探索回数が多い場合に年齢変更法より多くの誤差を削減可能である。

しかし、これらの手法^{8),9)}は都道府県単位で個票を合成するために提案されていた。一方、合成した仮想個票へ位置情報属性を追加する手法¹⁰⁾や所得属性を追加する手法¹¹⁾では市区町村単位で合成された仮想個票が必要となる。人口20万人未満の市町村では、個人が特定されないように、人口20万人以上の市区町村と比べ粗い統計情報のみ公開されている。仮想個票を合成する手法^{8),9)}では都道府県や人口20万人以上の市区町村を対象とした統計表を用いる。人口20万人未満の市町村を合成するためには、公開されていない統計表を推計するか、都道府県や市部・郡部の統計表を縮小する必要があった。そのため、市区町村単位で合成した仮想個票を統合し、都道府県単位の仮想個票を作成する場合、例えば市区町村単位の仮想個票が使用した統計表と完全に一致している場合においても、統合した仮想個票を都道府県単位の統計表と比較すると差異が発生する恐れがあった。

本研究では、各仮想個票が市区町村属性を保持しつつ、都道府県や市部・郡部単位の統計表にも適合する仮想個票の合成手法を提案する。具体的には、ある都道府県を構成する市区町村の仮想個票を1度に合成する。その際に、市区町村・市部・郡部・都道府県単位の統計表を用いる。市部とは、都道府県下の全ての市と区である。郡部とは、都道府県下の全ての町と村である。用いる統計表を仮想個票へ適合させる際には、統計表それぞれの調査範囲と仮想個票の市区町村とを対応させる。すなわち、郡部の統計表の場合は、ある都道府県下の全ての町村を対応させる。統計表の調査範囲と仮想個票の市区町村を対応させることにより、統計表の推計や過度な縮小作業が不要となる。なお、本研究では、従来手法より統計表に適合した仮想個票を合成するために、年齢交換法⁹⁾を基にした手法を提案する。

2 従来手法を用いた世帯構成合成手法

2.1 都道府県及び人口20万人以上の市区の合成方法

著者らが提案した世帯合成手法⁹⁾は、統計情報を基に作成した仮想の世帯構成を、複数の統計に適合させる手法である。個人の年齢や親子の年齢差の統計情報に対する、コンピュータ上で合成した世帯構成のデータ集合(仮想個票, 合成データ)の誤差を計算し、SA法を用いて誤差を最小化する。

合成データは複数の世帯とその構成員である個人によって構成している。合成データのモデルをFig. 2に示す。著者らの手法は、山形県を対象に、山形県の統計情報に基づく世帯構成の合成を試みており、家族類型は実統計¹²⁾に掲載されているFig. 1の9種類の家族類型を用いている。なお、これらの9種類の世帯数で日本全体の全世帯数の約95%、山形県については約90%を占めていることが知られている。その他の世帯数の家族類型は、他の親族を含む世帯や兄弟姉妹のみからなる世帯、非親族を含む世帯、他に分類されない世帯である。これらの家族類型は、公開されている複数の統計情報に整合させることが困難である。

世帯の中には世帯を構成する構成員が存在している。それぞれの個人は年齢、性別、所属する家族類型、世

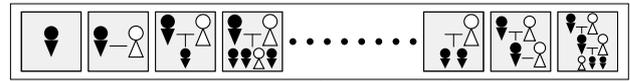


Fig. 2: 合成データのモデル

帯の役割、親族関係の5つの属性を持っている。世帯の役割は、所属する世帯の中での役割を表している。

著者らは、対象地域と同じ規模の世帯構成を合成するために、対象地域を集計対象とする統計表を用いて世帯数や人口などの統計情報通りに初期世帯を生成する。都道府県単位で合成する際には都道府県の、市区単位で合成する際には市区の統計表を用いる。著者らの初期世帯合成法⁹⁾では、家族類型、世帯人員別世帯数の統計表である国勢調査人口等基本集計¹²⁾表番号11を用いている。表番号11を考慮することにより、合成した世帯数と国勢調査と相違なく初期世帯を合成可能である。詳細は文献⁹⁾を参照されたい。また、個人の性別と初期の年齢を設定する際には、国勢調査人口等基本集計¹²⁾表番号16-1を用いている。個人の性別は統計表の男女の数に相違が発生しないように設定する。詳細は文献^{8),9)}を参照されたい。

著者らはSA法を用いて合成データと複数の統計表の差を最小化している。著者らが用いた21種類の統計表を以下に示す。

- (1) 父子の年齢差
(人口動態職業・産業別統計 出生¹³⁾表2)
- (2) 母子の年齢差
(人口動態調査 確定数 保管統計表¹⁴⁾出生5-2)
- (3) 夫婦の年齢差
(国勢調査 人口等基本集計¹²⁾表番号17)
- (4) 単独世帯の男性の人口分布(1歳階級)
- (5) 単独世帯の女性の人口分布(同上)
- (6) 夫婦のみ世帯の男性の人口分布(同上)
- (7) 夫婦のみ世帯の女性の人口分布(同上)
- (8) 夫婦と子供世帯の男性の人口分布(同上)
- (9) 夫婦と子供世帯の女性の人口分布(同上)
- (10) 男親と子供世帯の男性の人口分布(同上)
- (11) 男親と子供世帯の女性の人口分布(同上)
- (12) 女親と子供世帯の男性の人口分布(同上)
- (13) 女親と子供世帯の女性の人口分布(同上)
- (14) 夫婦と両親世帯の男性の人口分布(同上)
- (15) 夫婦と両親世帯の女性の人口分布(同上)
- (16) 夫婦とひとり親世帯の男性の人口分布(同上)
- (17) 夫婦とひとり親世帯の女性の人口分布(同上)
- (18) 夫婦、子供と両親世帯の男性の人口分布(同上)
- (19) 夫婦、子供と両親世帯の女性の人口分布(同上)
- (20) 夫婦、子供とひとり親世帯の男性の人口分布(同上)
- (21) 夫婦、子供とひとり親世帯の女性の人口分布(同上)

全ての統計表は年齢不詳の項目を除いて調整した。なお、統計表(4)~(21)は国勢調査人口等基本集計¹²⁾表番号16-1を用いている。統計表(1)、統計表(2)の形式をTable 1に、統計表(3)の形式をTable 2に、統計表(4)~統計表(21)の形式をTable 3に示す。なおTable 1~Table 3は平成22年度に調査された鳥取県の統計表を正規化した値である。

著者らが用いたSA法では、以下の目的関数を用いて最適化をしている。

Table 1: 統計表 (1) 父子の年齢差

条件 X	条件 Y	割合 (%)
父子関係	年齢差 ~19	0.495
父子関係	年齢差 20~24	8.645
...
父子関係	年齢差 65~69	0.022
父子関係	年齢差 70~	0.000

Table 2: 統計表 (3) 夫婦の年齢差

条件 X	条件 Y	割合 (%)
夫婦関係	年齢差 ~-70	0.000
夫婦関係	年齢差 -69	0.000
...
夫婦関係	年齢差 -1	6.960
夫婦関係	年齢差 -0	12.701
夫婦関係	年齢差 +1	13.087
...
夫婦関係	年齢差 +69	0.000
夫婦関係	年齢差 +70~	0.000

$$f(A) = \sum_{s=1}^S \sum_{j=1}^{G_s} |c_{sj}(A) - \text{Round}(r_{sj} \times m_{sj}(A))| \quad (1)$$

ここで、 A は合成データ、 S は統計表の数、 s は統計表の種類、 G_s は統計表 s の項目数、 c_{sj} は統計表 s の条件 X_{sj} と条件 Y_{sj} を満たす合成データの個人の数、 r_{sj} は統計表 s の項目 j の割合、 m_{sj} は統計表 s の条件 X_{sj} を満たす合成データの個人の数、 Round 関数は小数点以下を四捨五入する関数を表している。上記の目的関数を統計表 $s = 1, 2, \dots, S$ に対して行い、その総和の最小化を SA 法を用いて行う。著者らが用いた統計表 s の項目数 G_s を Table 4 に示す。統計表 (4)~統計表 (21) の項目数 G_s は全て 0 歳から 100 歳までの 1 歳区分の 101 である。

著者らが用いた SA 法⁹⁾ では以下の手続きにより最適化を行う。

Step 1 合成データを初期生成

Step 2 探索回数が規定数に達するか、 $f(A) = 0$ になるまで探索を終了

Step 3-1 合成データ内の個人をランダムに 1 人選択

Step 3-2 Step 3-1 で選択した個人と同じ家族類型かつ性別の個人をランダムに 1 人選択

Step 3-3 Step 3-1 と Step 3-2 で選択した個人の年齢を交換

Step 4 解の遷移判定

Step 5 探索回数を更新して SA の温度を冷却

Step 6 Step 2 の処理に戻る

2.2 人口 20 万人未満の市町村の合成方法

本節では、2.1 節で説明した従来手法⁹⁾ を用いて、人口 20 万人未満の市町村を対象とし、仮想個票を合成する方法を説明する。従来手法⁹⁾ では、仮想個票を合成するために、国勢調査 人口等基本集計¹²⁾ 表番号 16-1 や表番号 17 を用いている。これらの統計は人口 20 万人以上の市区のみ公開されている。そのため、人口 20 万人未満の市町村の仮想個票を合成する際には、統計

Table 3: 統計表 (4) 単独世帯における男性の人口分布

条件 X	条件 Y	割合 (%)
単独世帯の男性	年齢 0	0.000
単独世帯の男性	年齢 1	0.000
...
単独世帯の男性	年齢 99	0.016
単独世帯の男性	年齢 100	0.028

Table 4: 統計表 s の項目数 G_s と年齢階級

統計表	G_s	年齢階級
統計表 (1)	13	5 歳
統計表 (2)	9	5 歳
統計表 (3)	141	1 歳
統計表 (4)	101	1 歳
...
統計表 (21)	101	1 歳

表の推計か対象地域を含んだより広域を対象とした統計表を縮小し使用する必要がある。本節では、従来手法^{8),9)} を用いて人口 20 万人未満の市町村を合成する手法を説明する。

まず、初期世帯を合成する際に用いる統計表の推計方法について説明する。従来手法^{8),9)} における初期世帯生成法では、国勢調査 人口等基本集計¹²⁾ 表番号 11 と表番号 16-1 を用いる。表番号 11 は日本全国全ての市区町村において公開されているが、表番号 16-1 は人口 20 万人未満の市町村では公開されていない。表番号 16-1 は家族類型別、婚姻状況別、男女別の 1 歳階級の人口分布が記されている。一方、家族類型別、婚姻状況別、男女別の 5 歳階級の人口分布が記されている。国勢調査 人口等基本集計¹²⁾ 表番号 16-2 は日本全国全ての市区町村において公開されている。従来手法^{8),9)} の初期世帯合成手法を用いて人口 20 万人未満の市町村を合成するために、表番号 16-2 を 1 歳階級に推計している。これは、家族類型別、男女別の 1 歳階級の人口分布を得るためである。

国勢調査 人口等基本集計¹²⁾ 表番号 16-2 を家族類型別に 1 歳階級へ推計する際には、表番号 16-2 の他に国勢調査 人口等基本集計¹²⁾ 表番号 3-2 を用いる。表番号 3-2 には男女別の 1 歳階級の人口分布が記載されている。表番号 16-2 を家族類型別、男女別に 1 歳階級へ推計する際には以下の式を用いる。なお、表番号 16-2 には婚姻状況別の人口分布も記載されているが、従来手法⁹⁾ の初期世帯合成法では婚姻状況別の人口分布を用いていないため、婚姻状況別の人口分布の推計は行わない。

$$p_{t,g,a} = \text{Round} \left(\frac{p_{g,a} \times pr_{t,g,a5}}{\sum_{i \in T} pr_{i,g,a5}} \right) \quad (2)$$

ここで、 $p_{t,g,a}$ は家族類型 t 、性別 g 、年齢 a の人口、 $p_{g,a}$ は性別 g 、年齢 a の総人口、 $pr_{t,g,a5}$ は家族類型 t 、性別 g 、年齢 $a \sim a+4$ の人口、 T は家族類型の最大数である。 $p_{g,a}$ は表番号 3-2、 $pr_{t,g,a5}$ は表番号 16-2 から取得する。式 (2) により、各市町村の 1 歳階級別人口 $p_{g,a}$ を同じ市町村を対象にした 5 歳階級別の家族類型 t の割合により推計している。1 歳階級の家族類型別、男女別の人口分布を推計後は従来手法と^{8),9)} 同様に初期世帯を合成する。

SA法を用いて適合させる統計表(1)~(21)について説明する。統計表(1),(2)は都道府県単位で調査された統計表である。市区町村を対象とし仮想個票を合成する際には都道府県の統計表を用いる。統計表(3)は人口20万人以上の市区と市部、郡部を対象に調査された統計表である。人口20万人未満の市を対象とする場合は市部の統計表を、人口20万人未満の町村を対象とする場合は郡部の統計表を用いる。統計表(4)~統計表(21)は式(2)を用いて推計した家族類型、男女別の1歳階級の人口分布を用いる。

3 提案手法

2.2節では、従来手法⁹⁾を用いて人口20万人未満の市町村の仮想個票を合成する方法を説明した。従来手法⁹⁾を用いる際に、人口20万人未満の市町村では公開されていない統計が必要な場合、統計表の推計や、より広い範囲を対象とした統計表を縮小し使用していた。そのため、市区町村単位で合成した仮想個票を統合し、都道府県の仮想個票を作成すると、都道府県を対象とする統計表に適合しない恐れがある。本章では、市区町村属性を保持しつつ、都道府県下の市区町村を一度に合成する提案手法を説明する。

3.1 初期世帯合成法

提案手法の初期世帯合成法では、合成対象の都道府県下の市区町村を一度に合成する。その際に、人口20万人以上の市区と人口20万人未満の市町村では利用可能な統計表が異なるため、それぞれ異なる手法で初期世帯を合成する。人口20万人以上の市区では、従来手法⁹⁾の初期世帯合成法で用いる全ての統計表が利用可能であるため、従来手法⁹⁾と同様に初期世帯を合成する。

一方、人口20万人未満の市町村では、初期世帯合成に用いる統計表の一部の利用が不可能である。従来手法⁹⁾の初期世帯合成手法では、国勢調査人口等基本集計¹²⁾表番号11と表番号16-1を用いている。表番号16-1は人口20万人未満の市町村では公開されていない。そのため、本研究では、表番号16-1の代替となる統計表を次式で求める。

$$p_{t,g,a}^q = \text{Round} \left(\frac{\hat{p}_{t,g,a}^Q \times pr_{t,g,a5}^q}{\sum_{i \in T} pr_{i,g,a5}^q} \right) \quad (3)$$

ここで、 $p_{t,g,a}^q$ は都道府県 Q に属する市区町村 q 、家族類型 t 、性別 g 、年齢 a の人口、 $\hat{p}_{t,g,a}^Q$ は家族類型 t 、性別 g 、年齢 a の都道府県 Q の人口、 $pr_{t,g,a5}^q$ は市区町村 q 、家族類型 t 、性別 g 、年齢 $a \sim a+4$ の人口である。なお、 $\hat{p}_{t,g,a}^Q$ は人口20万人以上の市区を含む、各市区町村の初期世帯を合成後、次式を用いて値を更新する。

$$\hat{p}_{t,g,a}^Q = \hat{p}_{t,g,a}^Q - p_{t,g,a}^q \quad (4)$$

$\hat{p}_{t,g,a}^Q$ の初期値は都道府県 Q の国勢調査人口等基本集計¹²⁾表番号16-1の家族類型 t 、性別 g 、年齢 a の統計値である。式(4)では、各市区町村の初期世帯合成後に都道府県 Q の統計値 $\hat{p}_{t,g,a}^Q$ から式(3)で求めた統計値 $p_{t,g,a}^q$ を差し引いている。この操作により、

$$p_{t,g,a}^Q = \sum_{q \in Q} p_{t,g,a}^q \quad (5)$$

となるような各市区町村の家族類型別、男女別、1歳階級の人口分布が計算される。本研究では、式(4)を用いて、都道府県 Q の人口 $\hat{p}_{t,g,a}^Q$ を差し引きながら、都道府県 Q に属する各市区町村の初期世帯を順に合成する。式(4)を用いて $\hat{p}_{t,g,a}^Q$ を更新することにより、都道府県 Q の人口分布と相違ない初期世帯が合成可能となる。例えば、都道府県 Q に7の市区町村がある場合、最初の市区町村 $q=1$ は、 $\hat{p}_{t,g,a}^Q$ の初期値である都道府県全体の人口に対して5歳階級の割合を掛け合わせて1歳階級の統計表を作成する。その後、 $q=1$ の市区町村の人口を差し引いて得られる更新後の $\hat{p}_{t,g,a}^Q$ に対して、 $q=2$ の都市の1歳階級の統計表を作成する。これらの操作を $q=7$ の市区町村の1歳階級の統計表を得るまで繰り返す。なお、都道府県 Q に属する市区町村の初期世帯を合成する順序は、総務省が設定した全国地方公共団体コードの昇順に行う。全国地方公共団体コードは国勢調査の地域コードと対応している。

3.2 適合させる統計表

本節ではSA法を用いて適合させる統計表とその適用範囲について説明する。人口20万人以上の市区では2.1節で示した21種類の統計表を用いる。人口20万人未満の市町村では2.1節で示した21種類の統計表に加えて以下の統計表を用いる。

- (22) 男性の人口分布(1歳階級)
(国勢調査人口等基本集計¹²⁾表番号3-2)
- (23) 女性の人口分布(1歳階級)
(国勢調査人口等基本集計¹²⁾表番号3-2)

なお、統計表(4)~統計表(21)の家族類型、男女別人口分布には国勢調査人口等基本集計¹²⁾表番号16-2に記載されている5歳階級の人口分布を用い、式(2)や式(3)で推計した統計表を用いない。推計した統計表に適合した仮想個票を合成すると、都道府県や日本など、より広域を調査範囲とする統計表に適合しない恐れがある。統計表(22)と統計表(23)を新たに追加した理由は、統計表(4)~統計表(21)が5歳階級の人口分布であるため、1歳毎の人口分布への適合させるためである。そのため、本研究では、SA法で適合させる統計表への調整を最小限にとどめる³⁾。

従来手法^{8),9)}と提案手法において使用する統計表の例をFig.3に示す。Fig.3は、人口20万人以上のA市、B市、C市と人口20万人未満のD市、E市、F町、G町から成り立つ仮想の都道府県の例である。従来手法において市区町村属性を保持し合成する場合は、市区町村単位で合成する必要がある。そのため、統計表(1)、統計表(2)、人口20万人未満の統計表(3)は同じ統計表を各市区町村の人口に合わせて縮小している(Fig.3(a)の青)。

提案手法では、統計表の調査範囲を可能な限り再現する。すなわち、都道府県単位で調査された統計表(1)と統計表(2)は、全ての市区町村の人口を母数にする。また、市部と郡部で調査された統計表(3)についても同様に、市部の統計表(3)は人口20万人未満の市を、郡部の統計表(3)は人口20万人未満の町村の人口を母

³⁾本研究で行う統計表の調整について、統計表(1)~(3)は全人口を対象とし集計されているため、9種類の家族類型の人口に合わせて調整している。統計表(4)~(21)は統計表から年齢不詳の人口を取り除いている。

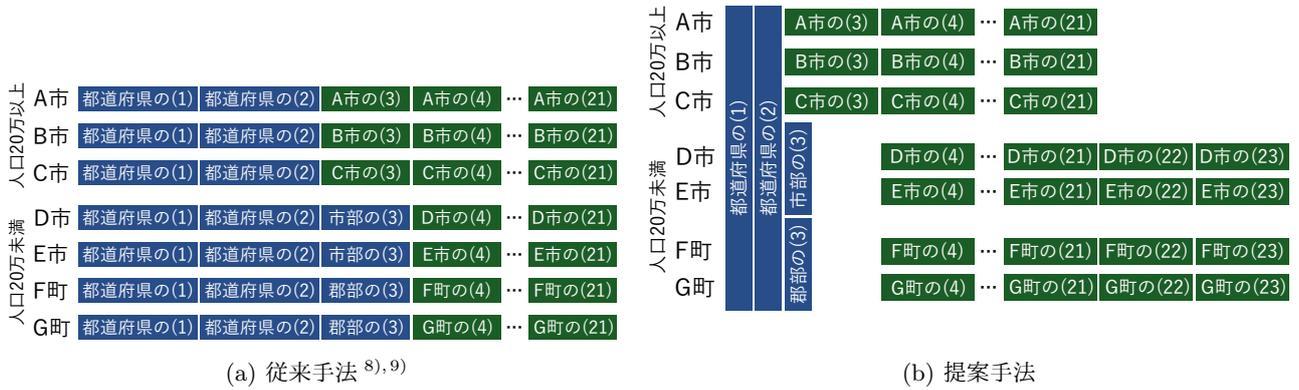


Fig. 3: 従来手法^{8),9)}と提案手法における統計表の適用範囲
(図中の(1)~(23)は統計表(1)~(23)である.)

数にする。提案手法では、都道府県下の市区町村を一度に合成することにより、都道府県や市部、郡部といった市区町村単位より広範囲で調査された統計表をSA法を用いて適合させる際に、統計表の調査範囲と同一の範囲で合成可能である。(Fig. 3 (b)の青)。

3.3 SA法の手続き

本研究が用いるSA法では以下の手続きにより行う。

- Step 1 3.1節の手法で合成データを初期生成
- Step 2 探索回数が規定数に達するか、 $f(A) = 0$ になるまで探索を終了
- Step 3-1 合成データ内の個人をランダムに1人選択
- Step 3-2 後述する条件に該当する個人をランダムに1人選択
- Step 3-3 Step 3-1とStep 3-2で選択した個人の年齢を交換
- Step 4 解の遷移判定
- Step 5 探索回数を更新してSAの温度を冷却
- Step 6 Step 2の処理に戻る

Step 3-2では、Step 3-1で選択した個人によって選択する個人の条件を変更する。Step 3-1で選択された個人が人口20万人以上の市区に属している場合、Step 3-2では、選択されている個人と同じ市区かつ家族類型かつ性別の個人をランダムに選択する。一方、Step 3-1で選択された個人が人口20万人未満の市町村に属している場合、以下の候補からランダムに選択する。

- Step 3-1で選択された個人と同じ市町村かつ家族類型かつ性別の個人
- 人口20万人未満の市町村の個人のうち、Step 3-1で選択された個人と同じ家族類型かつ性別かつ年代の個人

ここで、同じ年代とは、0歳~4歳、5歳~9歳など、5歳階級の年齢が等しいという意味である。他の市町村の同じ年代の個人を選択候補に加えることにより、統計表(22)と統計表(23)の1歳階級の人口分布との最適化が可能である。これは、提案手法の初期世帯合成法では、都道府県の人口分布を基にしており、初期世帯の合成時点では市町村単位の1歳階級の人口分布へ適合していないためである。

4 実験結果

本研究では、鳥取県を対象に世帯構成の合成を行う。鳥取県の合成対数は207,143世帯、その人口は498,403人である。鳥取県は19の市町村からなっており、鳥取市を除く18の市町村は人口20万人未満である。本研究が用いた計算機のCPUは、Intel Core-i7 3930K (3.2 GHz, 6コア)で、メモリはDDR3-1600 8GB×8、OSはMicrosoft Windows 10 Pro 64 bitである。

本研究では以下の手法の比較を行う。

- 都道府県単位 従来手法⁹⁾を用い、都道府県単位で合成する手法⁴⁾
- 市区町村単位 従来手法⁹⁾を用い、都道府県下の全ての市区町村を合成し、統合する手法
- 提案手法 3章で説明した提案手法

これらの3つの手法では、それぞれ用いる統計表やその数が異なる。本研究では、これらの手法を比較するために、それぞれの手法においてSA法の探索を終えた後に、鳥取県を調査対象とする統計表(1)~(21)を用いて評価を行った。鳥取県を調査対象とする統計表(1)~(21)と合成データの誤差の総和(式(1)の値)をTable 5に示す。SA法の設定として、初期温度を1.0、収束温度を0.1と設定し、冷却関数には指数冷却を用いた。これらの設定は従来手法⁹⁾と同様の設定である。探索回数は単独世帯を除く1人あたり1,000回、5,000回、10,000回とした。総探索回数はそれぞれ、445,578,000回、2,227,890,000回、4,455,780,000回である。なお、試行回数は10回であり、Table 5は平均値を示している。これらは3つの手法全てにおいて共通の設定である。また、Table 5の(4)~(21)の誤差は、統計表(4)~統計表(21)の誤差の総和である。

Table 5の誤差の合計と標準偏差から、従来手法⁹⁾を用いて都道府県単位で合成する場合が最も統計表との誤差が削減できている。次いで、提案手法、従来手法⁹⁾を用いて市区町村単位で合成する手法である。

従来手法⁹⁾を用いて市区町村単位で合成する手法が最も誤差が大きい理由は統計表(4)~(21)の家族類型、男女別、1歳階級の人口分布に適合できていない点である。鳥取県の19の市町村のうち、18の市町村が人口20万人未満である。仮想個票を合成する際に、推計

⁴⁾合成された各世帯は市区町村属性をもたない

Table 5: 統計表との誤差

統計表	探索回数 (回/人)	従来手法 ⁹⁾		提案手法
		都道府県	市区町村	
(1)	1,000	3,180.0	3,507.6	7,453.4
	5,000	505.4	817.2	2,588.2
	10,000	233.8	492.2	1,398.2
(2)	1,000	3,969.0	4,478.2	9,309.6
	5,000	597.8	1,041.2	3,146.4
	10,000	251.6	601.6	1,677.2
(3)	1,000	552.8	1,289.6	2,185.6
	5,000	53.8	895.6	1,021.6
	10,000	12.0	865.0	892.8
(4)~(21)	1,000	0.0	23,292.0	0.0
	5,000	0.0	23,292.0	0.0
	10,000	0.0	23,292.0	0.0
合計	1,000	7,701.8	32,567.4	18,948.6
	5,000	1,157.0	26,046.0	6,756.2
	10,000	497.4	25,250.8	3,968.2
標準偏差	1,000	121.02	146.46	224.90
	5,000	43.16	99.03	123.87
	10,000	27.55	91.25	84.78

した統計表 (4)~(21) を用いるため、統合後に鳥取県の人口分布と比較すると大きな誤差が発生していた。

提案手法は市区町村単位で合成する手法と比べ、統計表 (4)~(21) との誤差は削減できているが、統計表 (1)~(3) との誤差は増加している。父子・母子年齢差の統計表である統計表 (1) と (2) については、年齢を交換する 2 人目の個人を選択 (3.3 節の Step 3-2) する際に、人口 20 万人未満の場合は、1 人目の市町村の個人に加えて、他の市町村の同年代の個人も選択候補に加えている。そのため、5 歳階級の年齢差の統計表である統計表 (1)、統計表 (2) に影響する個人を選択する確率が低下する。夫婦年齢差の統計表である統計表 (3) については、従来手法⁹⁾ のような過度な縮小は行わないため、統計表に夫婦の年齢が大きく離れている珍しい世帯が含まれている。また、提案手法では従来手法⁹⁾ と比べ、同じ年代の個人を選択する確率が高い。そのため、夫婦の年齢差が大きく離れた世帯の合成が従来手法⁹⁾ と比べ困難である。

5 おわりに

本研究では、市区町村属性を保持しつつ、都道府県単位で世帯構成を合成する手法を提案した。従来手法⁹⁾ では、人口 20 万人未満の市町村の世帯構成を合成する際に、統計表の推計とより広範囲で調査された統計表を縮小し用いていた。そのため、ある都道府県下の市区町村の世帯構成を全て合成した後に統合し、都道府県の統計表と比較すると誤差が発生する恐れがあった。提案手法では、世帯構成を合成する際に、ある都道府県下の市区町村を同時に合成することで、統計表の調査範囲と相違なく仮想個票へ適合できた。

実験結果から、従来手法を用いて市区町村単位で合成し統合する手法と比べ、提案手法の誤差を、探索回数が 1 人あたり 1,000 回では 0.58 倍、5,000 回では 0.26 倍、10,000 回では 0.16 倍に削減する仮想個票を合成できた。一方、従来手法⁹⁾ を用いて都道府県単位で合成する手法と比較すると、探索回数が 1 人あたり 1,000 回では 2.46 倍、1 人あたり 5,000 回では 5.83 倍、1 人あたり 10,000 回では 7.98 倍、悪化していた。これは、

提案手法では年齢を交換する個人を選択する自由度が低いことと、人口 20 万人未満の市町村では全人口を対象とする 1 歳階級の人口分布を適合させる統計表に加えたことが原因である。

これらの結果から、市区町村や位置情報、所得の属性を必要としない社会シミュレーションを行う場合は従来手法⁹⁾ を用いて合成した仮想個票を、市区町村や位置情報、所得の属性を用いる社会シミュレーションを行う場合は提案手法を用いて合成した仮想個票を用いるべきである。

謝辞

本研究の一部は、科学技術融合振興財団、JSPS 科研費 17K03669 の助成を受けたものです。

参考文献

- 1) J. M. Epstein, R. L. Axtell: Growing Artificial Societies –Social Science from the Bottom Up–, MIT Press (1996)
- 2) R. Axelrod: The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration, Princeton University Press (1997)
- 3) 寺野：エージェントベースモデリング：KISS 原理を超えて、人工知能学会誌, **18-6**, 710/715 (2003)
- 4) A. G. Wilson, C. E. Pownall: A new representation of the urban system for modeling and for the study of micro-level interdependence, *Area*, **8-4**, 246/254 (1976)
- 5) M. Lenormand, G. Deffuant: Generating a synthetic population of individuals in households: Sample free vs sample-based methods, *Journal of Artificial Societies and Social Simulation*, **16-4**, 1/9 (2013)
- 6) 池田, 喜多, 薄田: 地域人口動態シミュレーションのためのエージェント推計手法, 計測自動車制御学会第 43 回システム工学部会研究会, 11/14 (2010)
- 7) 柘井, 村田: 統計データからの市民の属性復元のための進化計算と SA による 2 段階最適化, システム制御情報学会論文誌, **30-6**, 216/227 (2017)
- 8) T. Murata, T. Harada, D. Masui: Modified SA-based Household Reconstruction from Statistics for Agent-Based Social Simulations, *Proc. of 2016 IEEE International Conference on Systems, Man, and Cybernetics*, 3600/3605 (2016)
- 9) T. Murata, T. Harada, D. Masui: Comparing Transition Procedures in Modified Simulated-Annealing-Based Synthetic Reconstruction Method Without Samples, *SICE Journal of Control, Measurement, and System Integration*, **10-6**, 513/519 (2017)
- 10) T. Harada, T. Murata: Projecting household of synthetic population on buildings using fundamental geospatial data, *SICE Journal of Control, Measurement, and System Integration*, **10-6**, 505/512 (2017)
- 11) T. Murata, S. Sugiura, T. Harada: Income allocation to each worker in synthetic populations using basic survey on wage structure, *Proc. of 2017 IEEE Symposium Series on Computational Intelligence*, 471/476 (2017)
- 12) 厚生労働省:平成 22 年度 国勢調査 人口等基本集計 全国結果, http://www.e-stat.go.jp/SG1/estat/GL08020103.do?_toGL08020103_&tclassID=000001034991 (2011)
- 13) 厚生労働省:人口動態職業・産業別統計 出生 2010 年度, http://www.e-stat.go.jp/SG1/estat/GL08020103.do?_toGL08020103_&listID=000001109777 (2013)
- 14) 厚生労働省:人口動態統計 出生 2010 年, http://www.e-stat.go.jp/SG1/estat/GL08020103.do?_toGL08020103_&listID=000001101841 (2011)