

市区町村の統計表を考慮した都道府県単位の個票データの合成

原田 拓 弥*・村 田 忠 彦**

Synthetic Individual Households by Prefecture Considering Statistics of Local Governments

Takuya HARADA* and Tadahiko MURATA**

In this paper, we employ statistics of local governments (city, town, or village) in synthesizing individual households of a prefecture to reduce the error between synthesized statistics and actual statistics. Our previous method used estimated statistics of local governments with fewer than 200,000 people since those smaller local governments do not release their detailed statistics. To reduce the errors with prefectural statistics, we simultaneously employ statistics of smaller local governments in synthesizing population in a prefecture. Our experimental results show that we can make a 1/7 to 1/140 reduction in the error between synthesized statistics and actual statistics compared to the previous method.

Key Words: synthetic population, 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⁾。社会シミュレーションは少数のパラメータからなるシンプルなモデルを記述し、社会現象の原理的理解を目的とする Abstract Model, Abstract Model より詳細にモデルを記述する Middle Range Model, 社会現象を可能な限り忠実に再現する Facsimile Model に分類される^{2),3)}。特定の地域や社会現象を詳細に分析するためには Facsimile Model の構築が求められる場合がある。このようなモデルを作成するためには、環境のデータとそこで活動する人々のデータが必要となる。環境のデータは地理情報や地域メッシュといった地理情報システムの利用が可能である。モデルの粒度を細かくするほど、エージェントの意思決定においても、可能な限り現実の個人の振る舞いを模倣する必要がある。個人の意思決定モデルを記述する際には、年齢、性別、職業などの市民個人の属性が必要になることが少なくない。

個人を対象にしたエージェントの意思決定の実現における課題の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、実データによ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、仮想的な個人がもつ属性データ（以下、個票データ）を合成し、その個票データを前提とした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なわれるようになっている^{4)~6)}。

統計情報に基づく個票データの合成^(注1)に関する研究の歴史は古く、Synthetic Reconstruction Method (SR法)⁷⁾として知られている。その後、数多くの個票データ合成法が提案されているが、基本的にSR法に基づく、実個票データのサンプルを用いたアルゴリズムとなっている。Barthelemyら⁸⁾は、SR法の弱点として、個人の統計表と世帯の統計表のどちらかに適合する合成ができたとしても、両方に適合する合成が困難であることを指摘している。この課題を解決するため、Gargiuloら⁹⁾やBarthelemyら⁸⁾は、サンプルを用いない合成手法を提案している。LenormandとDeffuan¹⁰⁾は、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。これらの海外の研究ではそれぞれの国において利用可能な統計表と特徴に基づいた手法が開発されており、ほかの地域へ適用する際には留意が必要と指摘されている⁹⁾。

日本においても、日本の利活用可能な情報と特徴に基づい

* 青山学院大学理工学部 相模原市中央区淵野辺 5-10-1

** 関西大学総合情報学部 高槻市霊仙寺町 2-1-1

* College of Science and Engineering, Aoyama Gakuin University, 5-10-1 Fuchinobe, Chuo-ku, Sagami-hara

** Faculty of Informatics, Kansai University, 2-1-1 Reisenji-cho, Takatsuki

(Received April 2, 2021)

(Revised July 11, 2021)

(注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。

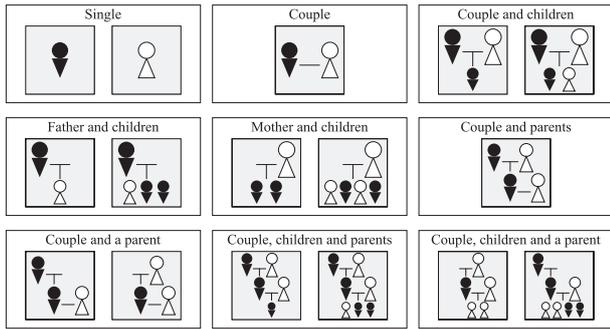


Fig. 1 Nine family types of household

た個票データが合成されている。日本における合成手法として、国勢調査のサンプルを用いた花岡の手法¹¹⁾とサンプルを用いない池田ら¹²⁾およびその派生手法^{13)~17)}がある。花岡¹¹⁾の手法により合成された人口は使用データの規約により一定期間のみ利用可能かつ第三者提供は不可能である。利用期間終了後はサンプルデータの返却とサンプルデータの個々の情報を判別可能な中間データを削除する必要がある。一方、池田ら¹²⁾およびその派生手法^{13)~17)}では、公開されている情報のみを用いているため、このような制限は存在しない。第三者提供が可能な手法で合成された個票データが公開されることで、シミュレーションを実施する研究者が独自に個票データを合成する必要がなくなる。

これらの国内外の研究を踏まえ、本論文では第三者提供可能な日本の個票データを合成するために、池田らの手法¹²⁾をもとにした手法^{16),17)}を改良する。池田ら¹²⁾は Fig. 1 に示す 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を提案した。家族類型とは、一般世帯^(注 2)を世帯員と世帯主の続柄により区分した分類である。柁井ら¹⁴⁾は、池田らが提案した合成手法をもとに目的関数や合成手法の改良を行なった。

しかし、池田ら¹²⁾や柁井ら¹⁴⁾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。柁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、家族類型ごと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、合成した家族類型ごとの世帯数と統計情報に相違が発生した。

著者らは、これらの問題を解決し、対象地域と同規模の個票データを合成する手法を提案した^{16),17)}。まず、初期世帯の合成方法と目的関数計算に用いる統計表を変更し、SA 法による最適化の過程で個人の年齢を変更する際に家族類型ごとの構成員の人口分布を考慮した合成手法を提案した(年齢変更法)¹⁶⁾。つぎに、年齢変更法では、年齢別の人口分布が変化する可能性があるため、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)¹⁷⁾。

これらの手法^{16),17)}では都道府県単位を対象として個票データの合成手法が開発されていた。一方で個票データへ居住地を割り当てる手法¹⁸⁾や所得属性を追加する手法¹⁹⁾では市区町村単位の個票データが必要となる。従来手法^{16),17)}が用いていた統計表のうち、父子年齢差と母子年齢差の統計表は都道府県のみ、夫婦年齢差と家族類型別、男女別 1 歳階級の人口分布は人口 20 万人以上の市区および都道府県を対象に公開されている。したがって、従来手法^{16),17)}を用いる場合、人口 20 万人以上の市区の合成には父子年齢差と母子年齢差を、人口 20 万人未満の市町村の合成には上記すべての統計表を推計する必要がある。年齢差の統計表の推計では、都道府県を対象に推計された統計表を市区町村の人口や世帯数の規模に応じて縮小^(注 3)していたため、高齢の夫と若年の妻のような珍しい属性をもつ少数の世帯が縮小された統計表には数値として表われず、考慮することができなかった。人口分布の推計では、5 歳階級の人口分布を 1 歳階級へ推計して合算した後、都道府県の実統計表と比較すると相違が発生していた。これにより、所得属性の追加時¹⁹⁾に個票データ内のある年齢の人口が労働者数を下回ることがあり、年齢別労働者数の統計表を個票データに合わせて変更する必要があった。

本論文では、ある都道府県に属するすべての市区町村の個票データを同時に合成する手法を提案する。提案手法では、都道府県単位で公開されている統計表の縮小を可能な限り避け、珍しい属性をもつ少数の世帯を合成できるようにする。なお、本論文では統計表と適合する個票データを合成するために、年齢交換法¹⁷⁾をもとにした手法を提案する。

本論文の構成はつぎのとおりである。まず、2 章で従来手法について説明する。つぎに 3 章で本論文が提案する手法を説明する。4 章で合成結果を示しその考察を述べる。5 章は本論文のまとめである。

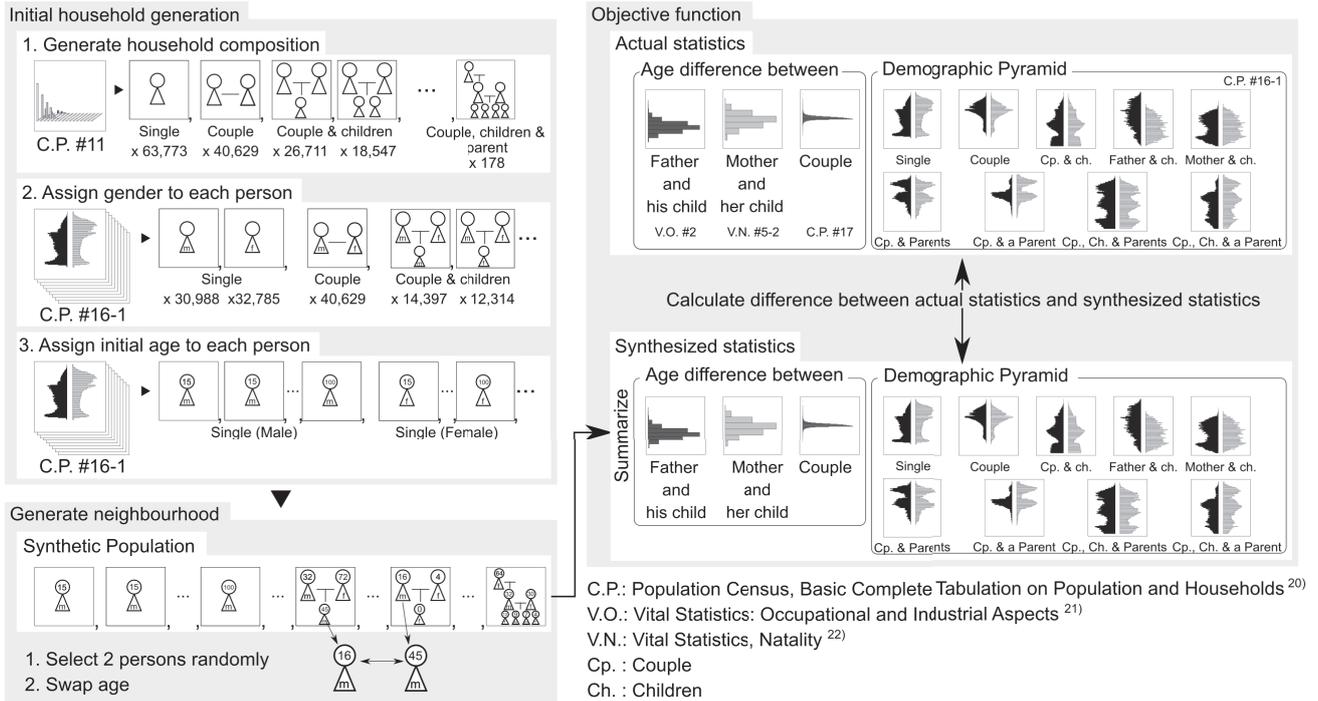
2. 従来手法

従来手法^{16),17)}は、統計情報をもとに作成した仮想の世帯構成を、SA 法を用いて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などの公開されている統計表とコンピュータ上で合成した世帯構成のデータ集合(個票データ)から集計する統計表(合成した個票データから作成する統計表を合成統計表と呼ぶ)を作成し、その誤差最小化を目的とし解を探索している。

個票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。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、世帯内の役割、親族関係の 5 つの属性をもっている。世帯内の役割は、夫や妻、子など所属する世帯内の役割を表わしている。

(注 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え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。

(注 3) 本論文では、対象とする都道府県や市区町村の人口や世帯数と一致するように統計表に集計されている統計量を割合に応じて減少される操作を統計表の縮小と表わす。

Fig. 2 Previous method^{16), 17)}

著者らの手法は、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²⁰⁾に掲載されている Fig. 1 の 9 種類を用いている。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% を占めている。

2.1 都道府県および人口 20 万人以上の市区の合成方法

従来手法^{16), 17)}を用いた都道府県および人口 20 万人以上の市区の合成方法を Fig. 2 に示す。従来手法^{16), 17)}は初期世帯合成、近傍解生成、目的関数の 3 つの要素から構成されている。

初期世帯合成では、対象地域と同じ規模の世帯構成を合成するために、世帯数や男女別人口を統計情報どおりに生成している。まず、家族類型、世帯人員別世帯数が集計されている国勢調査 人口等基本集計²⁰⁾ 表番号 11 を用いて、統計表と同数の世帯を生成する。つぎに、個人の性別と初期の年齢を設定する。これらは家族類型別、男女別、1 歳階級の人口分布が集計されている国勢調査 人口等基本集計²⁰⁾ 表番号 16-1 を用いて、男女の数と人口を統計表どおりに設定している。詳細は文献^{16), 17)}を参照されたい。

近傍解生成は、年齢交換法¹⁷⁾を用いた以下の手続きで行なう。

Step 1 合成データ内の個人をランダムに 1 人選択

Step 2 Step 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 Step 1 と Step 2 で選択した個人の年齢を交換
従来手法¹⁷⁾では以下の目的関数の最小化を目的とし最適化をしている。

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

ここで、 A は合成された個票データ、 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} は合成統計表 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} は実統計表 s の項目 j の割合、 m_{sj} は合成統計表 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である。すなわち、上記の目的関数では、実統計表の統計量と個票データから作成する合成統計表の統計量の絶対誤差を算出している。上記の目的関数を統計表 $s \in S$ に対して計算し、その総和の最小化を目的に SA 法を用いて探索している。

従来手法¹⁷⁾では目的関数の計算につぎの統計表を用いている。

- 父子の年齢差^(注 4)
- 母子の年齢差^(注 5)
- 夫婦の年齢差^(注 6)
- 家族類型別、男女別、人口分布^(注 7)

すべての統計表は年齢不詳の項目を除いている。父子の年齢差の形式を Table 1 に、夫婦の年齢差の形式を Table 2 に、家族類型別、男女別、人口分布を Table 3 に示す。なお Table 1~Table 3 は平成 27 年に調査された鳥取県の統計

(注 4) 人口動態職業・産業別統計 出生²¹⁾ 表

(注 5) 人口動態調査 確定数 保管統計表²²⁾ 出生 5-2

(注 6) 国勢調査 人口等基本集計²⁰⁾ 表番号 17

(注 7) 国勢調査 人口等基本集計²⁰⁾ 表番号 16-1

Table 1 Age difference between father and his child (statistics P-1)

X	Y	Rate (%)
Father-Child	Age difference -19	0.491
Father-Child	Age difference 20-24	6.58
...
Father-Child	Age difference 65-69	0.000
Father-Child	Age difference 70-	0.000

Table 2 Age difference between couple (statistics P-3)

X	Y	Rate (%)
Married Couple	Age difference -70 or less	0.000
Married Couple	Age difference -69	0.000
...
Married Couple	Age difference -1	7.177
Married Couple	Age difference ±0	13.763
Married Couple	Age difference +1	13.073
...
Married Couple	Age difference +69	0.000
Married Couple	Age difference +70 or more	0.000

Table 3 Male demographic pyramid in single households (statistics P-4)

X	Y	Rate (%)
Single household	Age 0	0.000
Single household	Age 1	0.000
...
Single household	Age 99	0.013
Single household	Age 100	0.010

Table 4 Number of G_s and age group

Statistics	G_s	age group
Age difference between father and his child	13	Five-Year child
Age difference between mother and her child	9	Five-Year child
Age difference between couples, LC _x -1	141	Single-Year
Demographic pyramid by nine family types of household	1,818	Single-Year

表を正規化した値である。父子の年齢差と母子年齢差は都道府県単位で集計されている。人口 20 万人以上の市区の個票データを合成する場合には、これらの統計表を対象の市区の規模に応じて縮小し、使用している。

従来手法で用いている統計表 s の項目数 G_s を **Table 4** に示す。家族類型別、男女別、人口分布の項目数 G_s は 9 種類の家族類型について男女別に 0 歳から 100 歳以上までの 1 歳階級であるため、合計で 1,818 となる。

2.2 人口 20 万人未満の市町村の合成方法

本節では、従来手法¹⁷⁾を用いて、人口 20 万人未満の市町村を合成する方法を説明する。従来手法¹⁷⁾では、人口 20 万人以上の市区のみ公開されている国勢調査 人口等基本集計²⁰⁾表番号 16-1 や表番号 17 を用いているため、人口 20 万人未満の市町村の個票データを合成する際には、推計した統計表を用いて初期世帯合成と目的関数を計算している。

初期世帯合成では、人口 20 万人未満の市町村を対象として

集計されている 5 歳階級の人口分布の表番号 16-2 を用いて家族類型別に 1 歳階級へ推計する。推計には次式を用いる。

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

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、国勢調査 人口等基本集計²⁰⁾表番号 3-2、 $pr_{t,g,a5}$ は表番号 16-2 から取得する。(2) 式により、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^{16),17)}と同様に初期世帯を合成する。

人口 20 万人未満の市町村を対象に個票データを合成する場合、目的関数は (1) 式を使用している。一方で、目的関数の計算に使用する統計表はつぎのように変更している。父子年齢差と母子年齢差は人口 20 万人以上の市区を合成する場合と同様に、都道府県単位で集計されている統計表を縮小し、使用している。夫婦年齢差は合成対象が市の場合は市部を集計対象とする統計表を、合成対象が町・村の場合は郡部を集計対象とする統計表を縮小し、使用している。家族類型別、男女別、1 歳階級の人口分布は (2) 式を用いて推計した統計表を用いている。

3. 提案手法

本論文では、ある都道府県に属するすべての市区町村の個票データを同時に合成することで、個票データ内の家族類型別、男女別、1 歳階級の人口が実統計表と一致し、父子年齢差、母子年齢差、夫婦年齢差の統計表の縮小を可能な限り避けることで、珍しい属性をもつ少数の世帯を含めて合成することが可能な手法を提案する。従来手法では、市区町村単位の個票データの合成を行なう際に、都道府県単位でしか得られない統計表を縮小して対応していたため、珍しい属性をもつ少数の世帯が合成できないという課題があった。本節では、提案手法を初期世帯合成、近傍解生成、目的関数に分けて説明する。

3.1 初期世帯合成

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同時に合成する。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^{16),17)}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、都道府県単位で集計された家族類型別、男女別、1 歳階級の人口分布と一致するように、家族類型別、男女別、1 歳階級の人口分布を人口 20 万人未満の市町村ごとに推計する。従来手法における (2) 式を用いた縮小化は、市町村ごとに独立に縮小が行なわれてい

るため、都道府県単位の統計表との整合性は保証されなかった。なお、目的関数の計算する際には、この推計した統計表を使用せず、表番号16-2で集計されている家族類型別、男女別、5歳階級の人口分布を用いる。

人口20万人未満の市町村の家族類型別、男女別、1歳階級の人口分布を推計するために、次式を満たす統計量 $\hat{ps}_{t,g,a}^Q$ を求める。

$$ps_{t,g,a}^Q = \sum_q \hat{ps}_{t,g,a}^q \quad (3)$$

$$ps_{t,g,a5}^q = \sum_a \hat{ps}_{t,g,a}^q \quad (4)$$

ここで、 $ps_{t,g,a}^Q$ は都道府県 Q において、人口20万人未満の市町村に居住する家族類型 t 、性別 g 、年齢 a の人口、 $\hat{ps}_{t,g,a}^q$ は求める人口20万人未満の市町村 q の家族類型 t 、性別 g 、年齢 a の人口、 $ps_{t,g,a5}^q$ は国勢調査人口等基本集計²⁰⁾ 表16-2に集計されている人口20万人未満の市町村 q における家族類型 t 、性別 g 、5歳階級 $a5$ の人口である。 $\hat{ps}_{t,g,a}^q$ は人口20万人未満の市町村 q における家族類型 t 、性別 g 、年齢 a の係数 $\alpha_{t,g,a}^q$ を使用し次式で算出している。

$$\hat{ps}_{t,g,a}^q = \text{Round}(ps_{t,g,a5}^q \times \alpha_{t,g,a}^q) \quad (5)$$

5歳階級の人口である $ps_{t,g,a5}^q$ を階級 $a5$ の1歳階級の係数である $\alpha_{t,g,a}^q$ を掛け合わせることで1歳階級の人口 $\hat{ps}_{t,g,a}^q$ を求めている。ここで Round は四捨五入する関数である^(注8)。(5)式に用いる係数 $\alpha_{t,g,a}^q$ は次式で求めている。

$$\alpha_{t,g,a}^q = \hat{ps}_{t,g,a}^q / \sum_{a \in a5} \hat{ps}_{t,g,a}^q \quad (6)$$

ここで、 $\hat{ps}_{t,g,a}^Q$ は都道府県 Q に属する人口20万人未満の市町村 q 、家族類型 t 、性別 g 、年齢 a の人口 $ps_{t,g,a}^Q$ を初期値にもつ変数である。この $\hat{ps}_{t,g,a}^Q$ を人口20万人未満の市町村 q の初期世帯を合成するたびに次式を用いて値を更新している。この操作により、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。

$$\hat{ps}_{t,g,a}^Q \leftarrow \hat{ps}_{t,g,a}^Q - \hat{ps}_{t,g,a}^q \quad (7)$$

なお、(7)式の \leftarrow は代入である。したがって、都道府県 Q の初期世帯合成後は次式となる。

$$\sum_a \hat{ps}_{t,g,a}^Q = 0 \quad (8)$$

3.2 近傍解生成

本論文では以下の手続きにより近傍解を生成する。

- Step 1** 合成データ内の個人をランダムに1人選択
Step 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 Step 1とStep 2で選択した個人の年齢を交換
 Step 2では、Step 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 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 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 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 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
- Step 1で選択された個人と異なる人口20万人未満の市町村に居住する個人のうち、Step 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、0~4歳、5~9歳など、5歳階級の年齢が等しいという意味である。

3.3 目的関数

提案手法においても目的関数は(1)式を使用する。目的関数計算に用いる統計表およびその集計方法をFig. 3に示す。Fig. 3の左は合成中の個票データを、中央は個票データから作成する合成統計表を、右は実統計表を示している。個票データと合成統計表の間の矢印は合成統計表の集計先を示している。たとえば、人口20万人以上の市である鳥取市の個票データは、鳥取県を集計対象とする統計表と鳥取市を集計対象とする統計表に計上される。また、人口20万人未満の市である米子市の個票データは、鳥取県を集計対象とする統計表、人口20万人以上の市を除く市部を集計対象とする統計表、米子市を対象とする統計表に計上される。なお、Fig. 3は鳥取県を例示している。また、Fig. 3は紙面の都合上、倉吉市、境港市、岩見町を対象とする統計表と集計先への矢印は省略している。

父子年齢差・母子年齢差は実統計表が都道府県単位で集計されていることから、合成対象のすべての市区町村を対象に集計し目的関数値を計算する。人口20万人以上の市区では、従来手法が用いていた夫婦年齢差、家族類型別、男女別、1歳階級の人口分布が公開されている。提案手法においてもこれらの統計表を用いて、Fig. 3の鳥取市のように、人口20万人以上の市区ごとに集計し目的関数値を計算する。

一方で、人口20万人未満の市町村では夫婦年齢差、家族類型別、男女別、1歳階級の人口分布が公開されていない。夫婦年齢差は市部と郡部を集計し、公開されている。したがって、市部を対象とする夫婦年齢差は個票データのうち、市に該当するものを集計、郡部は町・村に該当するものを集計し、目的関数値を計算する。家族類型別、男女別、1歳階級の人口分布の代わりに以下の統計表を新たに採用し、市町村別に目的関数値を計算する。

- 家族類型別、男女別、人口分布 (5歳階級)^(注9)
 - 男女別人口分布 (1歳階級)^(注10)
- 男女別人口分布 (1歳階級)を採用することで、初期世帯合

(注8) なお、この Round 関数は $ps_{t,g,a5}^q = \sum_a \hat{ps}_{t,g,a}^q$ を満たすように、四捨五入した要素を ± 1 調整している。

(注9) 国勢調査人口等基本集計²⁰⁾ 表番号16-2

(注10) 国勢調査人口等基本集計²⁰⁾ 表番号3-2

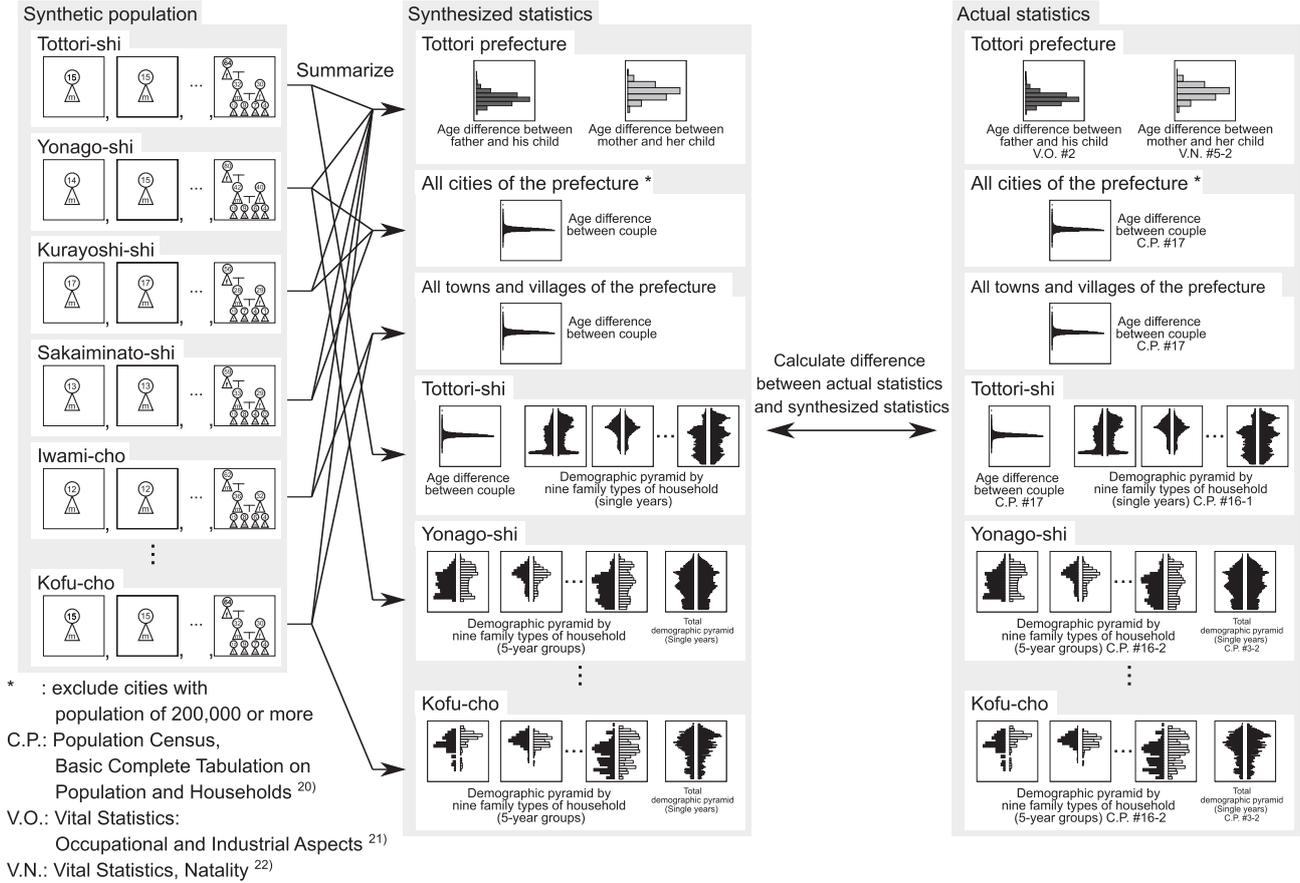


Fig. 3 Objective function

成時に考慮できていない人口 20 万人未満の市町村の 1 歳ごとの傾向を個票データへ反映することができる。しかし、男女別人口分布（1 歳階級）はある市区町村のすべての人口を対象に集計されているため、9 種類の家族類型を対象とする本論文ではこの統計表を直接使用し目的関数値を計算することには適さない。そこで、Fig. 4 に例示する方法により、人口 20 万人未満の各市町村の男女別人口分布（1 歳階級）を同市町村の 9 種類の家族類型の人口と都道府県の 1 歳階級の人口と一致するように 9 種類の家族類型に限定した男女別人口分布を推計する。

まず、都道府県の 9 種類の家族類型の 1 歳階級の人口分布から人口 20 万人以上の 9 種類の家族類型の 1 歳階級の人口分布を差し引き、人口 20 万人未満の市町村合算の 9 種類の家族類型の 1 歳階級の人口分布を求める。つぎに、この人口分布を 5 歳階級ごとに分割する。最後に、人口 20 万人未満の市町村別の 9 種類の家族類型の 5 歳階級の人口と分割した 1 歳階級の人口分布と一致するように人口 20 万人未満の市町村別、1 歳階級の人口分布を求める。この作業を分割した 5 歳階級すべてに実施する。

4. 実験結果

本論文では、鳥取県を対象に世帯構成の合成を行なう。鳥取県の合成対象は 207,143 世帯、その人口は 498,403 人であ

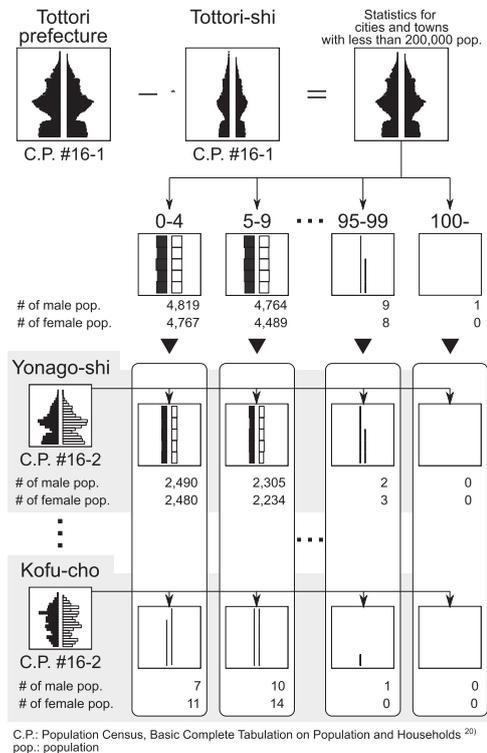


Fig. 4 Statistics calculation for the sum of cities, towns, and villages with less than 200,000 populations

Table 5 Total error (evaluation by statistics in Tottori prefecture)

Statistics	# of iterations per individual	Previous method ¹⁷⁾		Proposed method
		Prefecture	Local government	
Age difference between father and his child	1	164,149.6	164,419.8	181,358.6
	10	72,046.2	73,515.0	105,591.0
	100	4,204.2	5,173.0	13,546.8
	1,000	301.8	718.4	915.4
	10,000	32.0	179.4	161.6
	100,000	2.6	60.4	14.6
Age difference between mother and her child	1	197,704.4	197,924.2	226,547.0
	10	79,870.4	81,270.4	124,302.6
	100	5,887.4	6,727.0	16,325.0
	1,000	698.2	1,244.0	1,520.2
	10,000	96.0	296.8	459.2
	100,000	3.2	84.0	27.0
Age difference between couples	1	149,391.4	149,637.8	167,809.8
	10	78,973.4	79,514.2	118,841.0
	100	4,683.6	6,109.2	18,228.8
	1,000	207.0	873.6	996.0
	10,000	20.0	296.2	265.0
	100,000	2.4	195.4	120.0
Demographic pyramid by nine family types of household (single years)	1	0.0	22,334.0	0.0
	10	0.0	22,334.0	0.0
	100	0.0	22,334.0	0.0
	1,000	0.0	22,334.0	0.0
	10,000	0.0	22,334.0	0.0
	100,000	0.0	22,334.0	0.0
Total	1	511,245.4	534,315.8	575,715.4
	10	230,890.0	256,633.6	348,734.6
	100	14,775.2	40,343.2	48,100.6
	1,000	1,207.0	25,170.0	3,431.6
	10,000	148.0	23,106.4	885.8
	100,000	8.2	22,673.8	161.6
St.dev.	1	453.8	542.9	572.1
	10	679.7	622.7	537.1
	100	208.6	199.1	339.2
	1,000	61.0	60.2	83.7
	10,000	19.2	23.4	28.6
	100,000	2.6	18.4	10.8

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。本論文が用いた計算機の CPU は、AMD Ryzen Threadripper 1950X (16 コア, 3.4 GHz) で、メモリは DDR4-1866 16 GB×8, OS は Microsoft Windows 10 Pro 64 bit である。

本論文では以下の手法の比較を行なう。

都道府県単位 従来手法¹⁷⁾を用い、都道府県単位で合成する手法^(注 11)

市区町村単位 従来手法¹⁷⁾を用い、都道府県下のすべての市区町村を個別に合成し、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。本論文では、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、鳥取県を集計対象とする父子年齢差、母子年齢差、夫婦年齢差、家族類型別、男女別 1 歳階級の人口分布を用いて評価を行なった。

鳥取県を調査対象とする統計表と合成データから作成する合成統計表の絶対誤差の総和 ((1) 式の値) 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¹⁷⁾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 1×10^n ($n = 0, 1, \dots, 5$) 回とした。総探索回数は 445,578~44,557,800,000 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 つの手法すべてにおいて共通の設定である。

Table 5 の誤差の合計と標準偏差から、従来手法¹⁷⁾を用いて都道府県単位で合成する場合 (1 人あたり探索回数 100,000 回の場合で、誤差 8.6, 標準偏差 2.6) 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¹⁷⁾を用いて市区町村単位で合成する手法である。

従来手法¹⁷⁾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都道府県単位で集計された統計表である家族類型、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

(注 11) 合成された各世帯は市区町村属性をもたない。

万人未満である。個票データを統合し、その人口分布を鳥取県の実人口分布との差を(1)式により求めると22,334.0の誤差が生じている。これは、個票データの合成をする際に推計した統計表を用いたことが原因である。この誤差は各市町村の最適化時に計上されないため、従来手法¹⁷⁾では削減不可能な誤差である。

提案手法は市区町村単位で合成する手法と比べ、人口分布の誤差は削減できているが、年齢差の統計表との誤差は増加している。父子・母子年齢差の統計表については、年齢を交換する2人目の個人を選択(3.2節のStep 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、ほか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歳階級の父子年齢差、母子年齢差に影響を与える個人の選択確率が低下する。

従来手法¹⁷⁾では、都道府県単位の夫婦年齢差の統計表をそれぞれの市の規模に応じて縮小し、使用している。そのため、夫婦の年齢が大きく離れている世帯など、実統計表にごく少数存在する珍しい世帯は合成統計表には反映されず、そのような世帯が合成されたとしても、誤差として計上される。一方で、提案手法では、実統計表の集計地域と同じく、人口20万人以上の市区を除くすべての市から市部の夫婦年齢差を、すべての町・村から郡部の夫婦年齢差の統計表を集計するため、珍しい世帯が合成されていない場合に誤差として計上される。したがって、提案手法によって、統計表上にごく少数存在する珍しい世帯の合成が可能となる。しかし、提案手法では従来手法¹⁷⁾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¹⁷⁾と比べ困難である。

Table 5の探索回数1人あたり100,000回の誤差を人口で割った1人あたりの誤差は、都道府県単位で合成する手法は0.000016、市区町村単位で合成する手法は0.045493、提案手法で合成する手法は0.000324である。これはそれぞれ60,780人に1人、22人に1人、3,084人に1人、年齢に誤りがあることを意味する。シミュレーションモデルによって求められる精度が異なるものの、探索回数を増加させても誤差の削減が困難な市区町村単位で合成した個票データより都道府県単位で合成した個票データか提案手法で合成した個票データを使用すべきである。

5. おわりに

本論文では、ある都道府県下の市区町村の個票データを都道府県単位で同時に合成する手法を提案した。従来手法¹⁷⁾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は、推計した統計表を用いていた。そのため、所得属性追加時に労働者数が人口を上回るなど、属性追加やシミュレーションで活用するには留意が必要な個票データが合成されていた。提案手法では、ある都道府県に属するすべての市区町村の個票データを同時に合成することで、個票データ内の家族類型別、男女別、1歳階級の人口が実統計表と一致し、父子年齢

差、母子年齢差、夫婦年齢差の統計表の縮小を可能な限り避けることで、珍しい属性をもつ少数の世帯を合成することができた。

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、提案手法では、探索回数1人あたり1,000回では約1/7に、10,000回では約1/26に、100,000回では約1/140に誤差を削減できた。一方、従来手法¹⁷⁾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1人あたり1,000回では2.8倍、1人あたり10,000回では6.0倍、1人あたり100,000回では19.7倍誤差が増加していた。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、現時点においては、市区町村や位置情報、所得の属性を必要としない都道府県単位の社会シミュレーションを行なう場合は従来手法¹⁷⁾を用いた個票データを、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なう場合は提案手法を用いた個票データを使用すべきである。提案手法の誤差は、近傍解を生成する際に、実統計表との誤差が発生している世帯の構成員を集中的に選択することで改善が可能と考えられる。これは今後の課題とする。

謝辞 本論文の一部は、JSPS 科研費19K23229、JST 未来社会創造事業JPMJMI20B3、青山学院大学総合研究所、関西大学研究拠点支援経費研究課題の助成を受け、学際大規模情報基盤共同利用・共同研究拠点および革新的ハイパフォーマンス・コンピューティング・インフラ(jh190056, jh200022, jh210040)、HPCI 共用ストレージ(hp190215, hp200262)の支援を受けたものである。

参考文献

- 1) J.M. Epstein and R. Axtell: *Growing Artificial Societies: Social Science from the Bottom Up*, 1st edition, The MIT Press (1996)
- 2) N. Gilbert: *Agent-Based Models*, **153**, SAGE Publications (2008)
- 3) 高橋●●: 社会システムの研究動向3—評価・分析手法(1)—モデルの解像度と妥当性評価, 計測と制御, **52-7**, 582/587 (2013)
- 4) 市川●●: 医療分野におけるリスクマネジメント 地理情報分析と社会シミュレーション技術を用いた検討, 計測と制御, **57-6**, 407/412 (2018)
- 5) Y. Goto: Stylized Fact Analysis of Cash-for-Work Programs in the Disaster Reconstruction Process, *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 1144/1149 (2018)
- 6) 村田, 原田, 杉浦: リアルスケール社会シミュレーションの実現に向けて, シミュレーション, **36**, 58/62 (2017)
- 7) A.G. Wilson and C.E. Pownall: A New Representation of the Urban System for Modelling and for the Study of Micro-Level Interdependence, *Area*, **8-4**, 246/254 (1976)
- 8) J. Barthelemy and P.L. Toint: Synthetic Population Generation without a Sample, *Transportation Science*, **47-2**, 266/279 (2013)
- 9) F. Gargiulo, S. Ternes, S. Huet and G. Deffuant: An Iterative Approach for Generating Statistically Realistic Pop-

- ulations of Households, *PLOS ONE*, **5-1**, 1/9 (2010)
- 10) M. Lenormand and 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**, 112 (2013)
 - 11) 花岡●●: 全国版の小地域マイクロデータの構築と災害分析への活用, 地域安全学会論文集, **29**, 247/255 (2016)
 - 12) 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, 第 43 回システム工学部研究会, 11/14 (2010)
 - 13) 福田, 喜多: エージェントベースの人口推計モデルにおける属性決定手法の評価, システム制御情報学会論文誌, **27-7**, 279/289 (2014)
 - 14) 栢井, 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)
 - 15) 原田, 村田: 並列計算を用いた SA 法による都道府県レベルの大規模世帯の復元, 計測自動制御学会論文集, **54-4**, 421/429 (2018)
 - 16) 原田, 村田, 栢井: 家族類型と世帯内の役割を考慮した SA 法による大規模世帯の合成, 計測自動制御学会論文集, **54-9**, 705/717 (2018)
 - 17) T. Murata, T. Harada and 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)
 - 18) T. Harada and T. Murata: Projecting Households of Synthetic Population on Buildings Using Fundamental Geospatial Data, *SICE Journal of Control, Measurement, and System Integration*, **10-6**, 505/512 (2017)
 - 19) 杉浦, 村田, 原田: 賃金構造基本統計調査に基づく合成人口の労働者への就業属性別の所得の割当て, システム制御情報学会論文誌, **32-2**, 70/79 (2019)
 - 20) 総務省統計局: e-Stat 平成 27 年度国勢調査 人口等基本集計, <https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00200521&tstat=000001080615&tclass1=000001089055> (2016)
 - 21) 総務省統計局: e-Stat 人口動態職業・産業別統計 2015 年度, <https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00450011&tstat=000001028897&year=20151&tclass1=000001053122> (2018)
 - 22) 総務省統計局: e-Stat 人口動態統計 出生 年次 2015 年, https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00450011&bunya_1=02&tstat=000001028897&cycle=7&year=20150&month=0&tclass1=000001053058&tclass2=000001053061&tclass3=000001053064 (2016)

村田 忠彦 (正会員)



1997 年大阪府立大学大学院工学研究科博士後期課程修了。博士 (工学)。現在, 関西大学総合情報学部教授, シカゴ大学計算研究所客員研究員, IEEE SMCS 日本支部長, IEEE SMCS Soft Computing 技術委員長, IEEE SMCS Awareness Computing 技術委員長, IEEE CIS Neural Networks 技術委員, IEEE SMCS Associate Vice President, 日本進化計算学会会長などを歴任。システム制御情報学会奨励賞, IEEE SMCS 最優秀技術委員会賞を受賞。現在, 社会シミュレーションの研究に従事。システム制御情報学会, 知能情報フuzzy学会, 情報処理学会, IEEE などの会員。

[著者紹介]

原田 拓弥 (正会員)



2018 年関西大学大学院総合情報学研究科博士課程後期課程修了。博士 (情報学)。青山学院大学理工学部経営システム工学科助教を経て, 現在, 芝浦工業大学システム理工学部電子情報システム学科助教。大規模な社会シミュレーションの研究に従事。2019 年度計測自動制御学会論文賞受賞。経営情報学会の会員。