

# 中華人民共和国の国勢調査を用いた街路単位の人口合成

張錦旭 村田忠彦 (関西大学)

## Street-Level Synthetic Populations Using Population Census of People's Republic of China

\* J. Zhang and T. Murata (Kansai University)

**概要**— 本研究では、中華人民共和国の世帯単位の人口データを公開されている統計データに基づいて合成する。世帯単位の合成人口データには、世帯の家族構成とともに世帯構成員の年齢、性別の属性が含まれる。中国全国のどの地域でも入手可能な同じ形式の統計データのみを使用し、中国全国の街道単位（居住者数が数万人規模、日本の町単位相当）の世帯単位の人口データを合成する手法を提案し、提案手法の有効性を検証する。

**キーワード:** 合成人口データ, SA 法, 中華人民共和国

### 1 はじめに

本研究では、中華人民共和国（以下、中国）の世帯単位の人口データを、公開されている統計データに基づいて合成する。近年、コンピューターを用いた多くのシミュレーションモデルが構築され、社会シミュレーションの適用分野が、交通計画<sup>1)</sup>、健康<sup>2)</sup>、社会経済<sup>3)</sup>、政策効果<sup>4)</sup>など大きく拡がりを見せている。

社会シミュレーションの中でも、モデルの自由度の高さからエージェント・ベース・モデル (Agent-Based model, ABM) を用いたアプローチが注目されている。マクロモデルによる社会シミュレーションでは、集計データに基づく数理モデルが構築されるが、ABMでは、独立した意思決定の機能をもつ多数のエージェントからシステムが構成される。各エージェントは、モデル化された行動をそれぞれの状況に応じて試行し、他のエージェントや環境との相互に影響を与え合う。

近年、現実社会の個々の要素を反映したリアルスケール社会シミュレーション (Real-Scale Social Simulation, RSSS) が注目を集めている<sup>5-6)</sup>。ABMでも、社会的に実規模のモデルを組むためには、各エージェントの年齢や性別、所属する家族構成などの個票データが必要である。しかし、人口データとして、公開されている統計データまたは利用制限が課せられている一部の個票情報しか入手できず、すべての個人や世帯の情報を入手することは困難である。人口統計データの収集時に、利用範囲の説明をしたとしても、統計作成以外の利用目的に全員が同意するわけではない。一方、個人情報を利用することができたとしても、プライバシー侵害などの問題が起きる恐れがある。RSSSに取り組む際は、事前に対象とする地域の人口データを合成する必要があり、研究者の参入障壁となっている<sup>5-6)</sup>。これらのことから、プライバシー情報を用いず、公開されている統計情報から、統計的特徴を同一にする個人情報や世帯情報をもつ合成人口データがあれば、実規模のモデル構築の助けとなる。

本研究では、中国全国のどの地域でも入手可能な同じ形式の統計データのみを使用し、中国全国の街道単位（居住者数が数万人規模、日本の町単位相当）の世帯単位の人口データを合成する手法を提案し、提案手

法の有効性を検証する。

### 2 従来手法

人口合成手法の中でよく使われている手法は、反復比例フィッティング法 (Iterative Proportional Fitting Procedure, IPF法) と組合せ最適化法 (Combinatorial Optimization, CO法) である。

#### 2.1 IPF法

IPF法は、DemingとStephan<sup>7)</sup>によって初めて提案され、最も広く使用されている人口合成手法である。エージェントに基づく合成人口は、Beckman<sup>8)</sup>に実施されたトラフィック分析シミュレーションシステム (TRANSIMS) の研究で初登場した。そのシステムには合成人口、社会活動、経路計画などの部分が含まれている。Arentzeら<sup>9)</sup>は、IPF法を用いて個人レベルの人口を合成し、家庭に割り当てる人口合成手法 (ALBATROSS) を提案し、ヨーロッパ地域の人口を合成している。Wheatonら<sup>10)</sup>は、TRANSIMシステムの合成人口 (Population Synthesizer)、米国国勢調査局のデータ (US Census Bureau TIGER Data)、および 5% のサンプリングされた家庭データ (Public Use Microdata Sample, PUMS) に基づき、IPF法を用いて個人レベルの合成データを世帯に割り当て、ヨーロッパの人口を合成している。

Rich<sup>11)</sup>は、利用可能なデータの量、種類および質は国ごとに異なるため、それぞれで人口の合成手法を考える必要があることを示唆している。詳細なデータが利用できる場合、どのような手法を利用するのが重要である。IPF法は多くの研究者に推奨されているが、いくつかの制限がある。例えば、合成される人口データの精度はサンプルデータの代表性に大きく依存している。サンプルデータに含まれていない個人の属性値は、合成される人口においても、含まれることはない。

#### 2.2 CO法

CO法は、調査のマイクロデータから個人と世帯の最適な組合せを見つけ、合計されたデータを照合する手法である。CO法の利点は、IPF法よりもデータ要求が少ないことである。オーストラリア国立社会経済モデ

リングセンター (National Centre for Social and Economic Modeling, NATSEM) は、CO法を用いてオーストラリアの人口を合成している。しかし、Huangら<sup>12)</sup>は、CO法がIPF法より計算量が多いと指摘している。制約条件間の不一致もモデリングに問題を引き起こす可能性があるため、IPF法は収束できない恐れがある。CO法の場合、各制約条件に完全に一致する組み合わせが完璧に存在しない可能性があることを示している。

焼きなまし法 (Simulated Annealing, SA)はKirkpatrickら<sup>13)</sup>に考案された最適化アルゴリズムでCO法の一つとして用いることができる。Harlandら<sup>14)</sup>は、SA法を用いて多様な規模の集団が合成できることを示している。また、橋井ら<sup>15)</sup>はサンプルを用いないSA法を用い、政府統計総合窓口 (e-Stat) で公開されているデータに基づいて日本の世帯構成員の情報を含む人口を合成している。合成された各世帯を国土交通省国土地理院が公開する基盤地図情報を用いて建物座標に割り当てている<sup>16)</sup>。また、国勢調査と賃金構造基本統計調査を用いて、各世帯の就業者属性の割り当てを行っている<sup>17)</sup>。さらに、従業地属性を加えることにより、夜間人口 (居住地) と昼間人口 (従業地) の接続を行っている<sup>18)</sup>。

### 2.3 中国の人口合成手法

Chenら<sup>19)</sup>は、地方統計局、公安局、保健局などのデータに基づき、R言語のIPF法を用い、中国の直轄市である重慶市の人口を合成している。2015年の重慶市の1%統計調査を用い、家族の人数と家族の世代数の統計データを用いた人口合成を行っている。誤差の平均差は0.014%、0.010%である。なお、誤差の測定は家族の人数と家族の世代数によっておこなわれている。人数は1~5人以上、世代は1~4代以上である。したがって、世帯構成員の年齢などの属性は考慮されていない。

Yeら<sup>20)</sup>は、重みを調整して再分配することによって人口を合成する反復比例更新法 (Iterative Proportional Update, IPU) を提案している。世帯レベルと個人レベルの両方の特性が一致する重みを調整して再割当てを行う手順を用いている。Yeら<sup>21)</sup>は、国家統計局 (National Bureau of Statistics: NBS) のデータに基づき、IPU法を用い、中国の省単位 (各省の人口は、5千万人規模) の個人レベルの人口を合成している。NBSのデータの最小範囲は省であるため、Yeらは省単位の合成人口を完成させている。合成された世帯情報は、省の単位でしか割り当てられていないため、それより詳細の位置情報をえることはできない。本研究では、中国国勢調査で公開されている年齢構成、世帯構成に関する統計情報を用いて、中国における統計の最小単位である街路単位で世帯情報の合成を行う。

## 3 人口データ

本節では、中国の人口データの合成において、従来研究で使用されたデータと、本研究で使用するデータについて説明する。

### 3.1 従来研究

Yeら<sup>21)</sup>は、「中国国勢調査2000」<sup>22)</sup> (「第5回国勢調査」とも呼ばれる)の統計データを使用している。

中国国勢調査では、2000年11月1日0時現在の統計データが集計されている。統計データは、全人口を対象とした総合統計 (短表) と10%の人口を対象としたサンプル統計 (長表) の2種類に分かれている。Yeら<sup>19)</sup>は、短表と長表の2種類の統計データを同時に使用して人口合成を行なっている。長表は10%の人口を対象にしているため、合成された人口は必ずしも、全ての属性値を含んでいるわけではない。本研究は、全人口の属性値を反映させるため、短表のみを使用した人口合成を行う。

本研究では、街路単位の国勢調査の結果を用いた人口合成手法を提案する。2000年の国勢調査では、街路単位の統計は公開されていなかった。街路単位の統計は2020年から全ての自治体で義務付けられたため、街路単位の統計を用いた人口合成手法の開発が待たれている。

### 3.2 本研究

本研究で使用する統計データは、中国統計出版社 (China Statistics Press) が発行した「中国国勢調査年鑑2020 (China Population Census Yearbook 2020)」<sup>23)</sup> (「第7回国勢調査」とも呼ばれる) と中国各都市の市統計局が発行した「第7回全国都市別人口調査」<sup>24)</sup>である。

「中国国勢調査年鑑2020」は、2020年の中国の国単位の居住人口について、2020年11月1日0時現在の統計をエクセル形式で公開している。調査期間は2020年11月-12月である。国勢調査の未調査率は0.05%である。2020年11月の中国の人口は14億1178万人であり、2010年 (「中国国勢調査2010 (Tabulation on The 2010 Population Census of The People's Republic of China)」<sup>25)</sup>「第6回国勢調査」とも呼ばれる)と比較すると7206万人、5.38%増加している。全国の世帯数は5億2268万世帯である。そのうち、単身または家族単位の世帯 (家庭世帯) の数は4億9416万である。家庭世帯に属する人口 (家庭世帯人口) は12億9281万人である。一方、老人ホームなどの施設で暮らしている世帯 (集体世帯) は2853万世帯である。集体世帯に属する人口 (集体世帯人口) は1億1897万人である。家庭世帯は全世帯の94.5%を占めている。家庭世帯人口は全人口の91.7%を占めている。

また、国勢調査をもとに、都市単位で編纂されている「第7回全国都市別人口調査」では、居住人口について、2020年11月1日0時現在の統計を冊子形式で公開している。この集計結果により、各都市に含まれる街路単位 (万人規模) の個人情報と世帯情報を把握することができる。各都市は、規定された種類およびフォーマットの統計データを公表することが法律により義務づけられている。

「中国国勢調査年鑑2020」と「第7回全国都市別人口調査」は同じ調査データを使った統計である。統計の範囲と区分に違いがある。

本研究で使用する11種類の統計表を以下に示す。

- (1) A0105 省単位の年齢別 (5歳区分) ・性別人口
- (2) A0108 省単位の世帯タイプ別 (10種類) 世帯数
- (3) A0301 全国の年齢別 (1歳区分) ・性別人口
- (4) A0501 全国の世帯タイプ別 (50種類) 世帯数
- (5) A0502 省単位の年齢別 (12区分) ・性別単身世帯数

Table 1: 統計表の単位と年齢区分

統計表	単位	年齢区分
(1)	省	0, 1-4, 5-9, 10-14, ..., 95-99, 100以上
(2)	省	
(3)	全国	0, 1, 2, ..., 99, 100以上
(4)	全国	
(5)	省	0-14, 15-19, 20-24, ..., 60-64, 65以上
(6)	省	
(7)	省	
(8)	市, 区	0-14, 15-59, 60-64, 65以上
(9)	街道	
(10)	街道	0-14, 15-59, 60-64, 65以上
(11)	街道	

Table 2: 世帯と世帯構成員の属性値とその個数

属性	属性値	属性値の個数
世帯属性		
街道コード	110101001, ...	38734
世帯id	1, 2, 3, ...	522689264
世帯人数	1, 2, 3, ..., 10	10
世代数	1, 2, 3, 4, 5	5
世帯構成員属性		
街道コード	110101001, ...	38734
人口id	1, 2, 3, ...	1409778724
性別	男性, 女性	2
年齢	0, 1, 2, ..., 100	101
世帯タイプ	家庭世帯, 集体世帯	2
世帯id	1, 2, 3, ...	522689264

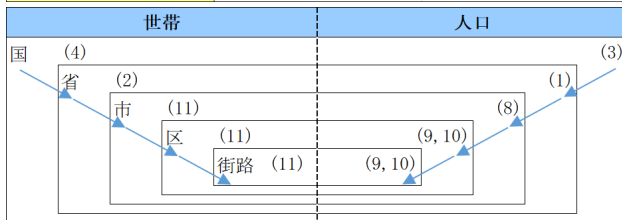


Fig.1: 初期世帯と初期世帯構成員の属性割当手続き

Table 3: 下位層での推計方法

国 統計表(3) 男			省 統計表(1) 男		省 推計 男	
年齢	万人	%	年齢	万人	年齢	万人
5歳	876.6	18.3	5-9歳	876.4	5歳	160.4
6歳	988.2	20.6			6歳	180.5
7歳	954.9	19.9			7歳	174.4
8歳	1031.8	21.5			8歳	188.4
9歳	950.4	19.8			9歳	173.5

Table 4: A0501 全国の世帯タイプ別の家庭世帯数

	家庭世帯数	一世代	二世代	三世代	四世代	五世代以上
合計	494157423	244615023	181471659	65528182	2540015	2544
一人世帯	125490007	125490007	0	0	0	0
二人世帯	146690059	110685779	36004280	0	0	0
三人世帯	103700982	5134646	93163291	5403045	0	0
四人世帯	65100986	1834602	42919633	20215417	131334	0
五人世帯	30513352	701437	7315318	21906684	589759	154
六人世帯	15125667	293766	1453830	12552955	824670	446
七人世帯	4589308	139356	386946	3439349	622936	721
八人世帯	1557638	89363	130530	1102904	234231	610
九人世帯	655060	53700	48137	476038	76840	345
十人以上世帯	734364	192367	49694	431790	60245	268

- (6) A0504 省単位の60歳以上の人を含む世帯数
  - (7) A0505 省単位の65歳以上の人を含む世帯数
  - (8) 県(市, 区)単位の年齢別(4区分)人口
  - (9) 街道単位の性別(2区分)人口
  - (10) 街道単位の年齢別(4区分)人口
  - (11) 街道単位の家庭世帯と集体世帯の世帯数, 人口
- (1)から(7)の統計表は「中国国勢調査年鑑2020」からのものである。「中国国勢調査年鑑2020」では, 短表がAテーブル, 長表がBテーブルと表示されている。(8)から(11)の統計表は「第7回全国都市別人口調査」

からのものである。「第7回全国都市別人口調査」では, 全ての統計が短表である。

(1)から(11)の統計表の集計単位と年齢区分をTable 1に示す。中国の行政区分は, 基本的には省(千万人規模), 市(百万人規模), 県または区(十万人規模), 街路(万人規模)という行政区で構成されている。なお, 街路は郷や鎮とも呼ばれ, 住民自治の基層組織居委会や村委会などの基層自治組織が設けられている。

統計表(1), (3), (5), (8), (10)において, 「100歳以上」を「100歳」に変更している。つまり, 最終的な合成人口では, 100歳として表される人口は, 実際に100歳以上である可能性がある。100歳以上の人口は全人口の0.008%を占めている。

また, 統計表(2), (4), (5), (11)の「10人以上の世帯」を「10人世帯」としている。「第7回国勢調査計画」<sup>26)</sup>では, 1世帯あたり20人までの国勢調査票を記入することができる。20人を超える大規模な世帯であれば, 複数の世帯に分けて記入することとなっている。そのため, 20人以上の世帯については, 統計上2つの世帯に分けられている。10人以上の世帯は全世帯の0.15%を占めている。

さらに, 統計表(2), (4), (5), (11)の「5世代以上」を「5世代」としている。NetEase<sup>27)</sup>によると, 中国には6世代の家庭世帯が1つ存在する。5世代以上は全世代の0.0005%を占めている。

本研究では, Yeら<sup>21)</sup>が使用したデータ「中国国勢調査2000」から「中国国勢調査年鑑2020」に更新する。さらに「第7回全国都市別人口調査」を追加することにより, 街路単位の人口を合成することが可能になる。

## 4 提案手法

2で, IPF法とCO法のメリットとデメリットを紹介した。SA法は, サンプルを必要としないことや, 他のアルゴリズムに比べて精度が高いというメリットがあるため, 本研究では中国の人口合成にSA法を利用する。

村田ら<sup>28)</sup>が公開している日本の合成人口データには, 年齢, 性別, 世帯タイプ, 世帯内の役割, 住所, 従業地などの情報が含まれている。Rich<sup>11)</sup>が指摘しているように, 各国で公開されている統計データは異なるため, 国ごとに人口合成手法を開発する必要がある。中国と日本の統計局が提供する統計表はかなり異なっているため, 中国で公開されている統計データに応じ, 柘井ら<sup>15)</sup>が提案したSA法を修正する必要がある。統計表の違いにより, 初期世帯と初期世帯構成員の生成や解の遷移判定に影響を与える。

本研究のSA法では以下の手続きを用いる。

- Step 1 初期世帯と初期世帯構成員を生成
- Step 2 探索回数が規定数に達すれば探索を終了
- Step 3 人口をランダムに2人選択, 交換
- Step 4 解の遷移判定
- Step 5 探索回数を更新
- Step 6 Step 2に戻る

### 4.1 Step 1: 初期世帯と初期世帯構成員の合成

街路単位で最終的に合成する世帯と世帯構成員の属

性, 属性値, 属性値の個数を Table 2 に示す. Fig. 1 の順に街路単位の初期世帯と初期世帯構成員の属性を計算する. 基本的には, 上位層の統計的特徴を下位層に適用する. 例えば, Table 3 のように国単位の統計表(3)に示される 1 歳区分の割合を反映させて, 省単位の統計表(1)の 12 の年齢区分内の人口を分割する. 同様の手続きにより, 街路単位で性別, 年齢 (1 歳区分) を得るため, 統計表(3), (1), (8), (9), (10)を順に用いて, 各年齢区分の人口を求める.

世帯と世帯構成員の属性に対しても, 年齢別の人口と同様の手続きを行い, 街路ごとの世帯タイプ別世帯数と世帯構成員の 1 歳区分の人口を得る. 世帯タイプとして, 中国には Table 4 のような 50 種類の世帯タイプの世帯数が集計されている. これらの 50 の世帯タイプを用いて, 街路単位の 50 種類の初期世帯数を計算する. 具体的には, 統計表(4), (2), (11)を順に用いて各世帯タイプの世帯数を計算する.

このようにして得られた街路単位の初期世帯と初期世帯構成員の属性の計算の結果は, 統計表(4), (2), (11)と統計表(3), (1), (8), (9), (10)に適合している. 初期世帯構成員に, 計算した街路単位の統計の項目に合うように年齢と数を割り当てる.

なお, 集体世帯と集体世帯人口については, 統計表がないため, 最初に家庭世帯の世帯ごとの人口を合成し, 残りの合成人口を集体世帯の初期世帯人口とする. 次に, 集体世帯人口は, 集体世帯にランダムに割り当てられ, 初期集体世帯を合成する.

#### 4.2 Step 3: 人口をランダムに 2 人選択, 交換

全人口の中からランダムに 2 人を選び, 所属する世帯を入れ替える. 選択された 2 人が同時に集体世帯となっている場合は, 後続の計算を試行せず, 再度ランダムに 2 人を選択する.

#### 4.3 Step 4: 解の遷移判定

本研究では, 統計表(5), (6), (7) の誤差の最小化を目的とする. 式(1)の各統計表の項目の誤差の総和  $\varepsilon_j$  を用いて, 合成人口から得られた統計における誤差を評価する.

$$\varepsilon_j = \sum_{i=1}^{n_j} | \text{合成統計値}_i - \text{実統計値}_i | \quad (1)$$

ここで,  $j$  は統計表番号 ((5),(6),(7)のいずれか),  $n_j$  は統計表  $j$  の項目数である. 式(1)の「合成統計値」は合成人口の属性のデータを表し, 「実統計値」は国勢調査に含まれる該当の属性のデータを表す. 解の遷移判定では, 式(1)の値を合計した以下の合計誤差の最小化を行う.

$$\varepsilon = \sum_{j=5}^7 \varepsilon_j \quad (2)$$

解の遷移判定において, 最小誤差値の合成人口を一時保存する.

Table 5: 統計表との平均誤差と平均標準偏差

統計表	探索回数 (人/回)	平均誤差	平均標準偏差
(5)	0	17.8	18.7
	1000	15.6	13.7
	5000	14.3	11.5
	10000	13.9	10.5
(6)	0	327.6	248.1
	1000	317.8	226.5
	5000	281.2	218.6
	10000	272.1	227.9
(7)	0	243.5	193.9
	1000	227.8	179.2
	5000	182.2	148.8
	10000	174.2	144.8

Table 6: 統計表との誤差の削減率

統計表	探索回数	平均誤差削減	削減率 (/1000回)
(5)	0~1000	2.2	2.20
	1000~5000	1.3	0.33
	5000~10000	0.4	0.08
(6)	0~1000	9.8	9.80
	1000~5000	36.6	9.15
	5000~10000	9.1	1.82
(7)	0~1000	15.7	15.70
	1000~5000	45.6	11.40
	5000~10000	8.0	1.60

## 5 実験結果

本研究では, 「甘肅省蘭州市永登県」の街路である「柳鎮」を対象に人口を合成する. 柳鎮には, 5,470 世帯, 13,981 人がいる. その中で 4,839 家庭世帯(88.5%), 13,503 家庭世帯人口(96.6%)がいる. 柳鎮の人口は, 全国の街道ごとの人口の中央値に近く, 典型的な街路であるとみなすことができる. さらに, 96% 以上の人口が家庭世帯に属している.

なお, 最適化に用いる統計(5), (6), (7)は省単位の統計であるため, 街道の合成人口の最適化においては, 省における割合を街道単位に適用して, 街道での値を推定したものを式(1)の「実統計値」とする.

合成人口と統計表(5), (6), (7)の誤差  $\varepsilon$  の平均と標準偏差の平均を Table 5 に示す. 人口を合成する回数は 3 回である. 「平均誤差」は統計表の各項目に対する平均誤差の平均値である. 例えば, 統計表(5)には 12 の項目(年齢区分)があり, 各項目に対する推定値と合成人口データの差の絶対値の平均値を計算する. 次に, 各試行で得られた平均値の平均を計算することで「平均誤差」を計算している. 「平均標準偏差」は項目間の誤差の標準偏差である. 各試行で得られた標準偏差の平均値を「平均標準偏差」とする. 探索回数を 1 人あたり 1000 回, 5000 回, 10000 回とする. 探索回数が増えるにつれて, 誤差と標準偏差が徐々に小さくなるのがわかる.

Table 6 に示すように, 探索回数が進むほど, 誤差の削減率も低くなる. 「平均誤差削減」は 0 回, 1000 回, 5000 回, 10000 回の式(1)の値の差の 1000 回の探索あたりの平均値である. 例えば, 統計表(5)に対して, 0 回と 1000 回の式(1)の値の差を計算し, 誤差削減とする. 1000 回から 5000 回までの間は 4000 回の探索に対する誤差削減になるため, 1000 回あたりとして誤差の

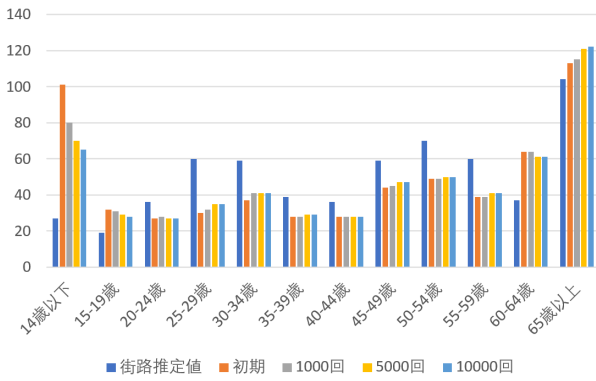


Fig.2: 統計表(5)に対する合成人口 (男性)

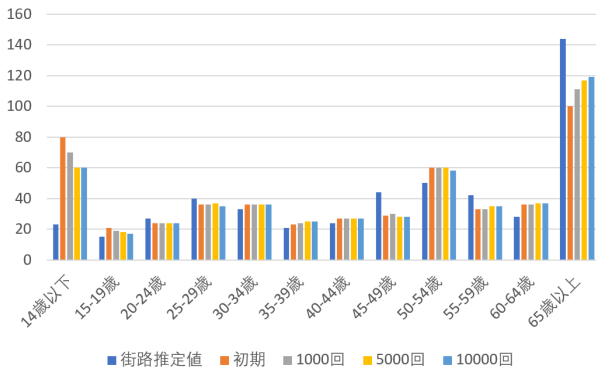


Fig.3: 統計表(5)に対する合成人口 (女性)

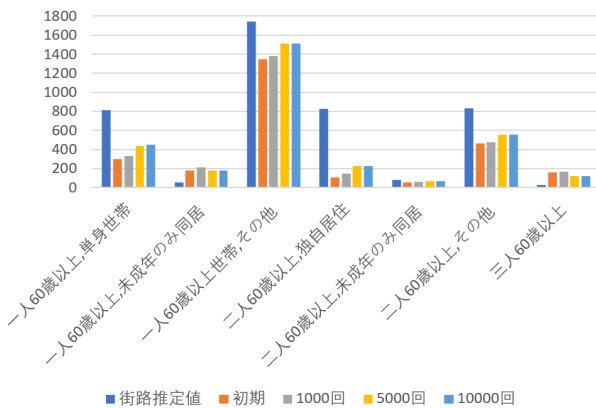


Fig.4: 統計表(6)に対する合成人口 (60歳以上)

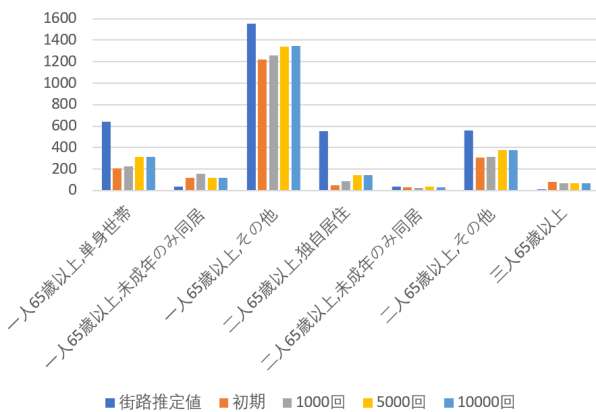


Fig.5: 統計表(7)に対する合成人口 (65歳以上)

差を4で割る. これらの値の試行数に対する平均を計算し, 「平均誤差削減」とする. したがって, 「平均誤差削減率」は1000回あたりの「平均誤差削減」である. 統計表(5), (6), (7)のいずれにおいても, 探索回数5000~10000では, 誤差の削減率が小さくなっており, 探索が収束しつつあることがわかる.

男性と女性に対する統計表(5)の街道に対する推定値と合成人口データから集計された合成統計値の値をFig.2とFig.3に示す. これらの図から, 10代と60代以上については, 合成人口データにおける人口が多く, 20代から50代については, 合成人口データにおける人口が街路の推定値に比べて少ないことがわかる. これは合成人口データにおいて, 年齢に関係なく単身世帯に割り当てられてしまったためである. つまり, 本来単身では生活していないと考えられる10代以下と60代以上の人口が合成人口では多く単身世帯として合成されてしまっていることを意味している.

統計表(6)に対する合成人口(60歳以上の人を含む世帯数)をFig.4に示す. 統計表(7)に対する合成人口(65歳以上の人を含む世帯数)をFig.5に示す. これらの図から, 一人60歳以上, 未成年のみ同居と三人60歳以上については, 合成人口データにおける人口が多く, それ以外については, 合成人口データにおける人口が少ないことがわかる. 特に, 一人60歳以上, 単身世帯と二人60歳以上, 独自居住の誤差が大きいことがわかる.

Fig.2~Fig.5に示した結果から, Table 5に示したように探索回数を増やすごとに, 街路推定値に近づく結果が得られていることがわかる. 探索回数が増えるにつれ, 合成人口は推定値に近くなっていく. 今後, さらに誤差を減らすため, 計算回数を増やし, 計算を効率に改善する.

本研究で用いた計算機のCPUは, Intel(R) Core(TM) i7-9700F (3.00GHz, 8コア)で, メモリーは8GB×2, OSはMicrosoft Windows 10 Pro 64 bitである. 探索回数1億3975万であり, 一人当たり10000回の人口合成にかかる計算時間の平均値は5369秒(1時間23分29秒)であり, 標準偏差は80.0である.

## 6 おわりに

本研究では, 中国国内における全ての街道単位で収集され, 公開される統計データの短表のみを使用し, 中国全国の街道単位の世界構成員を含む人口合成手法を提案した. 省単位で行われていた従来研究と比較し, 合成世帯の単位を省から街道にすることができた. また, 2020年に実施された最新の国勢調査の結果を利用し, 現在の実社会に近い人口を合成した. 特に短表のみを使用することにより, サンプル調査に起因する相違点が発生しないようにした. 悉皆調査を元にして, SA法による個票の誤差削減を行い, IPF法のデメ

リットを回避できた。提案手法を用いて、中国全国の人口を合成することが可能である。

## 謝辞

本研究の一部は、JSPS KAKENHI Grant Number JP20K10362 および JST 未来社会創造事業 (JPMJMI20B3) の支援を受けました。

## 参考文献

- 1) R. J. Beckman, K. A. Baggerly, M. D. McKay, Creating synthetic baseline populations. *Transportation Res Part A Policy Practice*, 30 (6): 415/429 (1996)
- 2) M. L. C. d. Atti, S. Meler, C. Rizzo, M. Ajelli, M. Massari, P. Manfredi, C. Furlanello, G. S. Tomba, M. Lannelli, Mitigation measures for pandemic influenza in Italy: an individual based model considering different scenarios. *PLoS One*, 3 (3): e1790 (2008)
- 3) A. J. Heppenstall, O. E. McFarland, A. J. Evans, Application of Multi-agent Systems and Social Network Theory to Petrol Pricing on UK Motorways, *International Central and Eastern European Conference on Multi-Agent Systems, CEEMAS 2005: Multi-Agent Systems and Applications IV*, 551/554 (2005)
- 4) D. Ballas, G.P. Clarke, E. Wiemers, Building a dynamic spatial microsimulation model for Ireland. *Population Space Place*, 11 (3): 157/172 (2005)
- 5) 原田, 杉浦, 村田, リアルスケール社会シミュレーションのための日本全国・複数年度の仮想個票の合成と活用可能性, 第 15 回社会システム部会研究会, 第 59 回システム工学部会研究会, 252 (2018)
- 6) T. Murata, T. Harada, Synthetic and Distribution Method of Japanese Synthesized Population for Real-Scale Social Simulations, *The 33rd Annual Conference of the Japanese Society for Artificial Intelligence* (2019)
- 7) W. E. Deming, F. F. Stephan, on a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known. *Annals of Mathematical Stats*, 11 (4): 427/444 (1940)
- 8) R. Beckman, K. Baggerly, McKay M. Creating Synthetic Baseline Populations. *Transportation Research Part A*, 30: 415/429 (1996)
- 9) T. Arentze, H. Timmermans, F. Hofman, Creating synthetic household populations: Problems and approach. *Transportation Research Record*, 2014 (1): 85/91 (2007)
- 10) W. D. Wheaton, J. C. Cajka, B. M. Chasteen, D. K. Wagener, P. C. Cooley, L. Ganapathi, D. J. Roberts, J. L. Allpress, Synthesized population databases: A US geospatial database for agent-based models. *Methods Rep RTI Press*. 2009, 2009 (10): 905 (2009)
- 11) J. Rich, Large-scale spatial population synthesis for Denmark. *European Transport Research Review*, 10 (2): 1/18 (2018)
- 12) Z. Huang, P. Williamson, A comparison of synthetic reconstruction and combinatorial optimization approaches to the creation of small-area microdata. Department of Geography, University of Liverpool, 1/86 (2001)
- 13) S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, "Optimization by Simulated Annealing". *Science* 220 (4598): 671/680 (1983)
- 14) K. Harland, A. Heppenstall, D. Smith, M. Birkin: Creating realistic synthetic populations at varying spatial scales, A comparative critique of population synthesis techniques, *Journal of Artificial Societies and Social Simulation*, 15-1, 1/24 (2012)
- 15) 栴井, 村田, 原田, 市民属性の合成手法における年齢交換による誤差最小化, 計測自動制御学会第 12 回社会システム部会研究会資料, 31/39 (2017)
- 16) T. Harada, T. Murata, Projecting Households of Synthetic Population on Buildings Using Fundamental Geospatial Data, *JCMSI*, Vol. 10, No 6, 505/512 (2017)
- 17) 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, Vol.32, No.2, 69/78 (2019)
- 18) T. Murata, D. Iwase, T. Harada, Workplace Assignment to Workers in Synthetic Populations in Japan, *IEEE*, 1/10 (2022)
- 19) X. Chen, X. Wang, Z. He, Population fitting research and system development based on urban micro-scale, *GEOSPATIAL INFORMATION*, Vol.18, No.3, 61/65 (2020)
- 20) X. Ye, K. Konduri, R. M. Pendyala, B. Sana, and P. Waddell, "A methodology to match distributions of both household and person attributes in the generation of synthetic populations," in *Proc. 88th Annu. Meeting Transp. Res. Board*, 1/24 (2009)
- 21) P. Ye, B. Tian, Y. Lv, Q. Li, F. Wang, On Iterative Proportional Updating: Limitations and Improvements for General Population Synthesis, *IEEE TRANSACTIONS ON CYBERNETICS*, VOL. 52, NO. 3, 1726/1735 (2022)
- 22) <http://www.stats.gov.cn/tjsj/pcsj/rkpc/5rp/index.htm>
- 23) <http://www.stats.gov.cn/tjsj/pcsj/rkpc/7rp/zk/indexch.htm>
- 24) [https://tjj.lanzhou.gov.cn/art/2021/5/28/art\\_4860\\_1006824.html](https://tjj.lanzhou.gov.cn/art/2021/5/28/art_4860_1006824.html)
- 25) <http://www.stats.gov.cn/tjsj/pcsj/rkpc/6rp/indexch.htm>
- 26) <http://www.chaozhou.gov.cn/attachment/0/490/490880/3685645.pdf>
- 27) <https://www.163.com/dy/article/HH3RAUCE0553SE9C.html>
- 28) <http://www.res.kutc.kansai-u.ac.jp/~murata/rsss-distribution/>