

小地域人口マイクロデータの推計手法に関する検討

—Iterative Proportional Updating と焼きなまし法の比較—

花岡 和 聖

I. 研究背景・目的

社会科学分野において、個人や世帯を識別できるマイクロデータ（個票データ）に対する関心は、近年、急速に高まっている。仮に、それが町丁・字等の小地域別に整理されていれば、マイクロデータの詳細な人口属性に基づき高齢者福祉や消費購買行動、交通需要などの地域差の把握が可能となり、学術的にも、実務的にも有益なデータとなる。しかし、国や自治体による公的な統計調査においては、個人情報保護のこともあり、そうした小地域が表章されたマイクロデータ（以下、小地域人口マイクロデータ）は、一般に提供されないことが多い。一方で、集計表の形式では、国勢調査小地域集計が整備されており、年齢や職業、家族構成等が表章された町丁・字等別の集計表を利用できる。ただし、オーダーメイド集計を利用しない限り、一般に公開される集計表の次元数や種類は限定されることから、世帯・人口分布を断片的にしか捉えられない。

こうした公的調査の利用上の課題に対して、複数の集計表を組み合わせたり、再抽出された部分的な世帯・個人サンプルを援用したりすることで、原データの代わりとなる小地域人口マイクロデータを推計する試みが進められてきた。1980年代の Birkin and Clarke¹⁾ による研究などに始まり、1990年代以降、地理学や交通工学のマイクロシミュレーション研究の中で、推計手法に関する研究が深化していく。これまでに様々な手法が提案されてきたが、それらは Iterative Proportional Fitting (IPF) 系と組合せ最適化アルゴリズム (Combinatorial Optimization: CO) 系に大別できる²⁾。近年、前者に関して、従来の手法の課題点を改善する形で、Iterative Proportional Updating (IPU)³⁾ や階層的 IPF⁴⁾ が提案されており、新たな方法論的な展開が認められる。

様々な手法が提案される中で、それぞれの手法には、小地域人口マイクロデータを推計する上での長所と短所が存在する。そこで、本研究では、第4回（平成12年）京阪神都市圏パーソントリップ調査の世帯サンプルを利用し、IPU 及び CO の一手法である焼きなまし法を適用することで、小地域人口マイクロデータの生成を行い、その推計結果を、国勢調査小地域集計との適合度や使用する世帯サンプル、地域特性などの観点から比較検討することを目的とする。なお、本研究では、小地域とは、国勢調査小地域集計の町丁・字等を意味し、世帯サンプルとは、実際の調査で収集された世帯と世帯内の個人（世帯員）を識別できるマイクロデータを意味する。

Ⅱ. 小地域人口マイクロデータの推計手法

1. IPF 系

IPF は、複数の集計表の周辺度数を制約条件として、多次元クロス表を推計する手法である。国勢調査小地域集計を制約条件（以下、制約表）として利用することで、小地域区分が表章された多次元クロス表を推計できる。その具体的な手続きは、制約表とする国勢調査小地域集計と多次元クロス表の周辺度数の比を用いて、クロス表のセル度数を繰り返し更新していくことで、その周辺度数を制約表と一致させる。国内では、浅見・木戸⁵⁾ や石坂他⁶⁾ による IPF の適用事例があり、特に、石坂他は、マルチスケールな IPF を提案しており、基本単位区から市区町村までの集計表と整合的な多次元クロス表を作成した。

IPF の課題として、世帯と世帯員レベルの集計表を同時に制約条件として考慮できないため、世帯数もしくは世帯員数のいずれかを選択しなければならない。また、世帯サンプルを用いないため、個々の世帯について、そこに属する個人の属性まで把握できない。それ故、小地域人口マイクロデータを生成するためには、多次元クロス表から求めた条件付き確率とモンテカルロ・シミュレーションを用いて、マイクロデータ化する作業が必要となる。この作業には、様々な条件を考慮して世帯構成を再現しなければならない、その計算プログラムは複雑なものになる上、非現実的な属性の組合せが生じる可能性もある⁷⁾。

こうした IPF の課題に対処する方法として、近年、IPU が提案されている。この手法は、第 1 図に示されるように、公的調査等の世帯サンプルをベースに、IPF と同様に、国勢調査小地域集計と世帯サンプルの周辺度数との比を用いて、繰り返し、拡大係数を更新していくことで、両者を一致させる手法である。IPU は、IPF と類似した推計手法を採用するが、世帯サンプルを利用する点で IPF とは異なる。

2. CO 系

組合せ最適化アルゴリズム (CO) による方法では、公的調査等の世帯サンプルを、それが国勢調査小地域集計と一致するようになるまで、繰り返しリサンプリングすることで、小地域人口マイクロデータを生成できる。これまでに、山登り法や焼きなまし法、遺伝的アルゴリズムを用いた手法が提案されてきた。その中でも、焼きなまし法 (Simulated Annealing: SA) は、他の手法と比較して良好な結果が得られるとされている⁸⁾。

SA は、高温に熱せられた金属を冷却する工程を模した組合せ最適化アルゴリズムである。この手法では、まず、世帯サンプルの全体から推計対象地区の世帯数分だけ無作為にその一部を抽出する。次に、対象地区内と全体の世帯サンプルからそれぞれ無作為にサンプルを一つ選び交換する。それによって、制約条件とする国勢調査小地域集計との適合度が改善されれば、そのサンプルの交換を正式に採用する。そうでなければ、温度パラメータに応じて、サンプルの交換の採用を判断する。この作業を数十万～数百万回繰り返すことで、国勢調査小地域集計と、集計表レベルで整合的な世帯サンプルの新たな組合せが求められる。

SA を用いた小地域人口マイクロデータの研究事例として、高齢者介護の需要予測⁹⁾ に始まり、地域労働市場の予測¹⁰⁾、消費者特性分析¹¹⁾、災害の社会的脆弱性分析¹²⁾ などが挙げられる。ここでは、「匿名データ」(イギリスの場合、Samples of Anonymised Records) などの公的調査のマイクロデー

| ステップ1 | | | | | | ステップ4 | | | | | |
|--------|-------------|-----|------|------|-----------------|-----------------|-------------|------|-----|------|-------------|
| 世帯番号 | 世帯人員数 | | 世帯数 | | 拡大係数 | 世帯番号 | 年齢1 | 年齢2 | 世帯A | 世帯B | 拡大係数 |
| | 年齢1 | 年齢2 | 世帯A | 世帯B | | | | | | | |
| 1 | 2 | 0 | 0 | 1 | 1 × 0.67 | 1 | 2 | 0 | 0 | 1 | 0.67 |
| 2 | 1 | 1 | 0 | 1 | 1 × 0.67 | 2 | 1 | 1 | 0 | 1 | 1.08 |
| 3 | 1 | 0 | 1 | 0 | 1 × 0.67 | 3 | 1 | 0 | 1 | 0 | 2.00 |
| 4 | 0 | 3 | 0 | 1 | 1 | 4 | 0 | 3 | 0 | 1 | 1.61 |
| 5 | 2 | 1 | 0 | 1 | 1 × 0.67 | 5 | 2 | 1 | 0 | 1 | 1.08 |
| 総数(推定) | 6 | 5 | 1 | 4 | | 総数(推定) | 6.58 | 7 | 2 | 4.44 | |
| 総数(国調) | 4 | 7 | 2 | 3 | | 総数(国調) | 4 | 7 | 2 | 3 | |
| 調整率 | 0.67 | | | | | 調整率 | 0.68 | | | | |
| ステップ2 | | | | | | 暫定的な推定結果 | | | | | |
| 世帯番号 | 年齢1 | 年齢2 | 世帯A | 世帯B | 拡大係数 | 世帯番号 | 年齢1 | 年齢2 | 世帯A | 世帯B | 拡大係数 |
| 1 | 2 | 0 | 0 | 1 | 0.67 | 1 | 2 | 0 | 0 | 1 | 0.46 |
| 2 | 1 | 1 | 0 | 1 | 0.67 | 2 | 1 | 1 | 0 | 1 | 0.73 |
| 3 | 1 | 0 | 1 | 0 | 0.67 | 3 | 1 | 0 | 1 | 0 | 2.00 |
| 4 | 0 | 3 | 0 | 1 | 1.00 | 4 | 0 | 3 | 0 | 1 | 1.09 |
| 5 | 2 | 1 | 0 | 1 | 0.67 | 5 | 2 | 1 | 0 | 1 | 0.73 |
| 総数(推定) | 4 | 4.3 | 0.67 | 3 | | 総数(推定) | 5.11 | 4.75 | 2 | 3 | |
| 総数(国調) | 4 | 7 | 2 | 3 | | 総数(国調) | 4 | 7 | 2 | 3 | |
| 調整率 | 1.61 | | | | | | | | | | |
| ステップ3 | | | | | | ※ステップ1に戻り、繰り返す。 | | | | | |
| 世帯番号 | 年齢1 | 年齢2 | 世帯A | 世帯B | 拡大係数 | | | | | | |
| 1 | 2 | 0 | 0 | 1 | 0.67 | | | | | | |
| 2 | 1 | 1 | 0 | 1 | 1.08 | | | | | | |
| 3 | 1 | 0 | 1 | 0 | 0.67 | | | | | | |
| 4 | 0 | 3 | 0 | 1 | 1.61 | | | | | | |
| 5 | 2 | 1 | 0 | 1 | 1.08 | | | | | | |
| 総数(推定) | 5.25 | 7 | 0.67 | 4.44 | | | | | | | |
| 総数(国調) | 4 | 7 | 2 | 3 | | | | | | | |
| 調整率 | 2.99 | | | | | | | | | | |

第1図 IPUによる推計手法の概要

図中の太字の数字は、各ステップの調整率とそれが適用されることを意味する。

国調：国勢調査小地域集計

タが使用される場合が多いが、消費者アンケート調査のマイクロデータを利用した事例も挙げられる¹³⁾。

3. 両手法の比較研究

IPFとCOの適合度を比較した研究成果は、Huang and Williamson¹⁴⁾やWilliamson¹⁵⁾、Ryan et al.¹⁶⁾で報告されてきた。そこでの結論は、総じてIPFよりもCOの方が、制約表との適合度の点で優れていることが示される。ただし、COは、学生寮や病院、老人ホームなどが立地するなど、地区の人口構成が画一的であると、同じ属性をもつ世帯サンプルだけを抽出することが難しく、適合度が低下する傾向にある。一方で、新たに提案されたIPUにおいても、世帯サンプルを利用するため、推計結果の適合度がIPFと同様の結果になるとは限らない。これまでにIPF系の異なる手法同士を比較した研究は行われてきたが¹⁷⁾、COとIPUの比較は、管見の限りにおいて実施されていない。それ故、IPF系の新たな手法であるIPUとCOの一つであるSAの推計結果を比較することは、小地域人口マイクロデータの生成手法を選定する上で、重要な判断材料となる。

Ⅲ. 研究方法

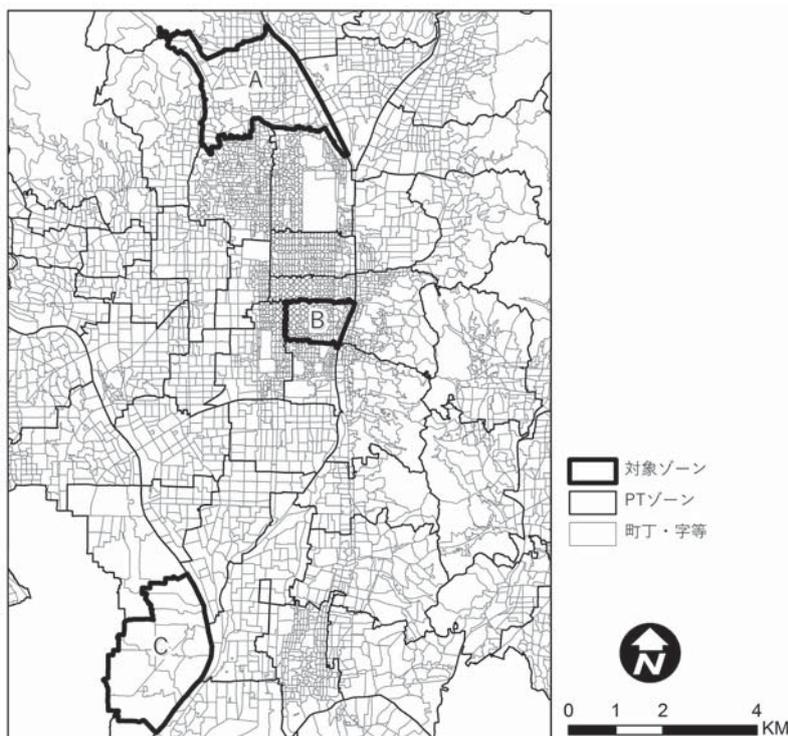
1. 使用データ及び検証地域

本研究では、第4回（平成12年）京阪神都市圏パーソントリップ調査のマイクロデータを使用する。その理由として、パーソントリップ調査は、地理学や交通工学等で共通して関心が高く、これまでにマイクロシミュレーション研究などで活用されてきた点が挙げられる。推計に使用する世帯サンプルは、以下に示す検証地域内で回収された1032世帯分である。制約表として、平成12年国勢調査小地域集計の男女別年齢階級（31項目）、職業（10項目）、世帯規模（7項目）に関する3つの集計表を用意した。

パーソントリップ調査には、PTゾーンと呼ばれる地域区分が表章される。京都市は約60のPTゾーンから構成され、1ゾーンあたり80程度の町丁・字等が含まれる。このうち、検証地域として、地域特性が異なる3つのPTゾーンを選定する（第2図）。それらは、①北区の都市的な郊外住宅地に位置するPTゾーンA、②下京区の都心部に位置し、業務地区と住宅地が混在するPTゾーンB、③伏見区にあり、郊外住宅地化が進んだ地域に位置するPTゾーンCである。無人や秘匿の地域を除くと、各PTゾーンには、国勢調査の町丁・字等が、それぞれ156地区（A）、139地区（B）、15地区（C）含まれ、合計で310地区となる。推計に利用する世帯サンプルの内訳は、それぞれ637世帯（A）、144世帯（B）、251世帯（C）である。

2. 適合度指標と推計手法

小地域人口マイクロデータの適合度指標として、制約表*i*について、表章項目*j*の国勢調査小地域



第2図 本研究で選定したPTゾーンの位置（京都市）
町丁・字等の境界は、平成12年国勢調査に基づく。

集計の観測度数 O_{ij} と、小地域人口マイクロデータに基づく推定度数 E_{ij} の差の絶対和 (Total Absolute Error: TAE) を合計した Overall TAE を用いる (式 (1)・(2))。Overall TAE のスコアは、地区の世帯数と比例して増加することが想定されるため、Overall TAE を地区の世帯数で除した世帯あたり Overall TAE も併せて使用する。

$$TAE_i = \sum_j |O_{ij} - E_{ij}| \quad (1)$$

$$Overall\ TAE = \sum_i TAE_i \quad (2)$$

小地域人口マイクロデータの推計には IPU と SA を用いる。両手法の各種パラメータは次の通りに設定した。IPU の収束計算は、地区あたり 1000 回とする。また制約表中に度数がゼロの項目がある場合、それを十分に小さな値 (0.0001) に置き換えて計算を行った。SA に関しては、Williamson¹⁸⁾ を参考に、世帯サンプルの交換の最大回数は 500 万回、初期温度は 48、温度減少率が適用されるまでの世帯サンプルの交換回数は 480 回、温度減少率は 5% に設定した。世帯サンプルの交換を採用するか否かは Overall TAE のスコアに基づく。

本研究では、次の 3 つの観点から両手法による推計結果を評価する。それらは、① Overall TAE と世帯あたり Overall TAE に基づく適合度と拡大係数の基本統計量、② 推計に使用する世帯サンプル数、③ 制約表と世帯サンプルの類似度と度数ゼロの表章項目割合である。

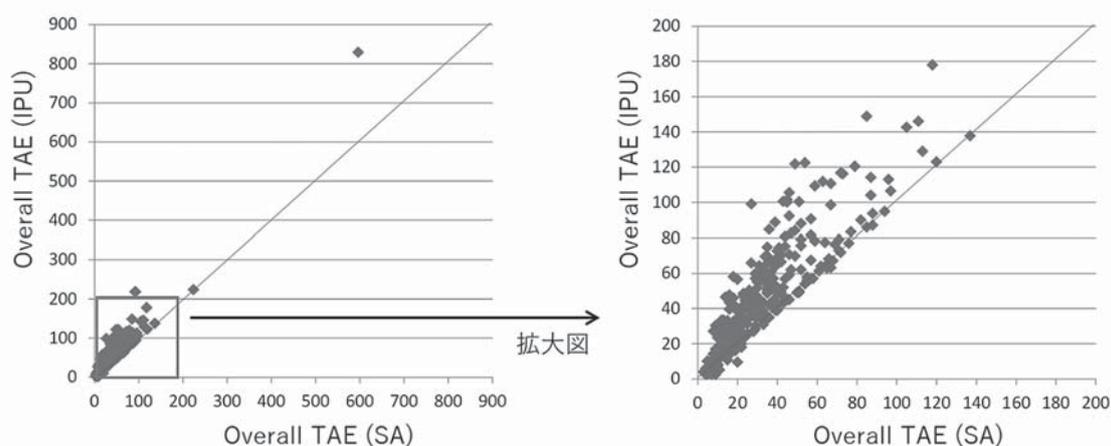
IV. IPU と SA の比較

1. 適合度及び拡大係数

検証地域内の 310 地区について、IPU と SA で得られた小地域人口マイクロデータの Overall TAE を散布図に示す (第 3 図)。両手法の適合度が同水準であることを示す対角線よりも上に、大半の地区の点が分布しており、IPU の方が、SA と比較して Overall TAE が大きいことがわかる。正確には 86% の地区で、SA の方が、国勢調査小地域集計の制約表と集計レベルでの整合性が高い。また、SA の適合度の方が高い地区のみを対象に集計すると、SA と IPU の Overall TAE の差は、平均で 17.3 であった。逆のケースでは、その差は平均で 2.1 であった。この結果からは、IPU の方が高い適合度を示す地区において、SA は、ほぼ同水準の適合度を示すことがわかる。

第 1 表に、地域特性が異なる 3 つの PT ゾーン別に、Overall TAE 及び世帯あたり Overall TAE の平均値を整理した。その結果をみると、まず、全てのゾーンにおいて、SA の方で、Overall TAE と世帯あたり Overall TAE の平均値が小さく、より高い適合度を達成したと言える。特に、業務・住宅混在地区において、SA の方が世帯あたりの Overall TAE が大幅に小さく、適合度の目安となる 1 を下回る結果がみられた¹⁹⁾。業務・住宅混在地区のように、世帯数が少ないと、特定の表章項目への度数分布の偏りや度数ゼロの表章項目が生じやすいと考えられる。この点に関しては第 3 節で言及する。

次に、拡大係数に関する基本統計量 (第 2 表) からは、各世帯サンプルが、小地域人口マイクロデータで最終的に何回使用されたかを把握できる。特定のサンプルが、繰り返し利用されることは、



第3図 IPU と SA の Overall TAE の分布

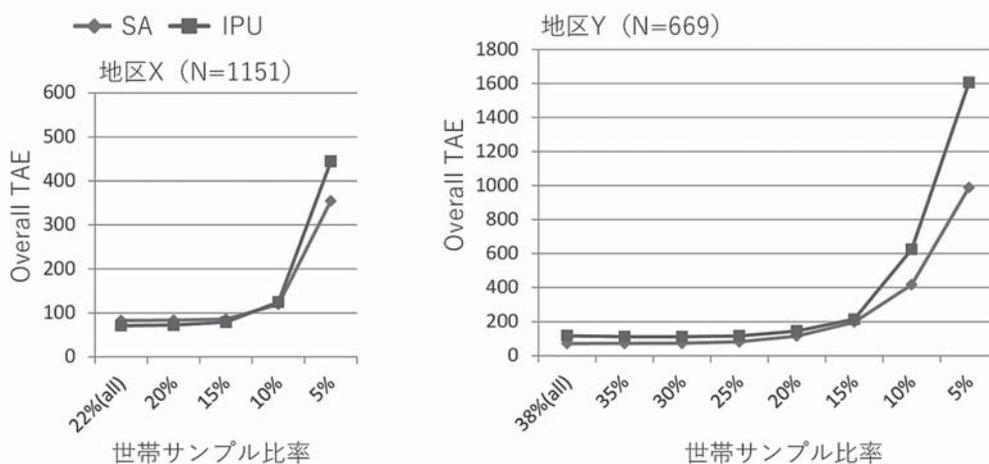
地区別の結果をプロットした。

第1表 PT ゾーンの地域特性別にみた推計結果

| PT ゾーン | 地区数 (町丁目) | 世帯数 平均 | 世帯サン プル数 | Overall TAE・平均 | | 世帯あたり Overall TAE・平均 | |
|---------------|--------------|-----------|-------------|----------------|-------|-------------------------|-------|
| | | | | IPU | SA | IPU | SA |
| A (近郊住宅地区) | 156 | 163.5 | 637 | 38.4 | 32.5 | 0.298 | 0.249 |
| B (業務・住宅混在地区) | 139 | 49.2 | 144 | 52.5 | 30.5 | 1.586 | 0.954 |
| C (郊外住宅地区) | 15 | 607.8 | 251 | 151.5 | 113.1 | 0.266 | 0.206 |
| 3ゾーン全体 | 310 | 133.7 | 1032 | 50.2 | 35.5 | 0.874 | 0.563 |

第2表 拡大係数の基本統計量 (3ゾーン全体)

| 手法 | 最小値 | 最大値 | 平均 | 標準偏差 | 推計結果に含まれる 世帯サンプル数 |
|-----|--------|------|------|------|----------------------|
| IPU | 0.0001 | 51.7 | 0.67 | 1.77 | 1032 |
| SA | 1.0 | 53.0 | 1.75 | 1.70 | 964 |



第4図 世帯サンプル比率別にみた IPU と SA の推計結果 (100セット・平均)

地区 X・Y は、PT ゾーン C から選定した。

世帯サンプル比率 = 推計に利用する世帯サンプル数 / 地区の総世帯数

制約表との適合度が高くとも、制約表以外の項目を利用する場合に問題となる。たとえば、パーソントリップ調査であれば目的別のトリップ数を集計した際に、特定の目的に分布が集中することになる。第2表に示した拡大係数の基本統計量をみると、IPUは、拡大係数が小数值をとることができるため、その平均値は1を下回る。一方で、SAの結果をみると、最大値や平均値が大きく、同じサンプルがより高い頻度で繰り返し採用されている。加えて、IPUの場合は、1032世帯すべてのサンプルが利用されているが、SAでは、本来含まれるべき68世帯のサンプルの拡大係数が0であり、一度も利用されていない。この点は、拡大係数が整数値に限定されるSAのデメリットである。

2. 推計に使用する世帯サンプル数

既往研究²⁰⁾では、推計に使用する世帯サンプル数が少ない地区ほど、Overall TAEが高くなる傾向が示された。そこで、本節では、同一の地区を対象に、パーソントリップ調査の世帯サンプル数を減少させた場合、適合度がどのように変化するかを確認する。検証地区は、PTゾーンCから、世帯数が多く、比較的良好な適合度が得られた2つの地区を選定した。1つ目の地区Xは、世帯数が最も多い町丁目(1151世帯)で、全世帯サンプルを用いた場合はIPUの方で適合度が良い。2つ目の地区Yは、世帯数がゾーン内の平均値に近い世帯数を有する町丁目(669世帯)で、全世帯サンプルを用いた場合はSAの方で適合度が良い。分析方法としては、地区の総世帯数に対する世帯サンプル数の比率を5%ずつ減らした場合におけるOverall TAEのスコアをそれぞれ求めた。推計から除外するサンプルは無作為に選択することとし、比率ごとに100セットを用意して、推計を行った。

第4図に、2カ所の検証地区のOverall TAEスコア(100セットの平均値)のグラフを示す。両地区のグラフをみると、地区Xでは、総世帯数に対する世帯サンプルの比率が5%で、地区Yでは、その比率が10%以下になると、Overall TAEが大きく増加することがわかる。特に、その増加分はSAよりもIPUで顕著であった。このような傾向から、推計に活用できる世帯サンプル数が少ない状況下では、SAの方が適合度の点で良好な結果が得られることがわかる。興味深い点として、各地区でOverall TAEが大きく変化する比率は、世帯サンプル数に換算すると60~70世帯程度であった。この結果は、今回、3つの国勢調査小地域集計を制約表(表章項目数は48項目)として考慮したが、これらに整合する小地域人口マイクロデータを生成するためには、地区の総世帯数に比例して世帯サンプル数を増やす必要があるというよりも、一定数の世帯サンプルが必要なことを示唆している²¹⁾。

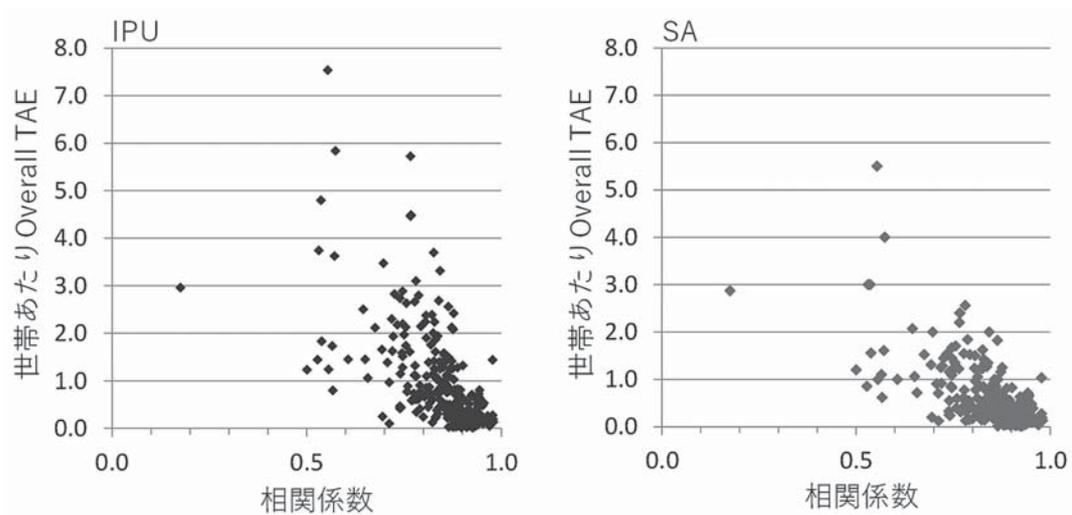
3. 世帯サンプルの類似度及び度数ゼロの項目割合

SAにおいて、地区内に大学や病院、刑務所などの大規模施設が立地し、世帯・人口構成が画一的な場合、適合度が低下する傾向にあることが既往研究で指摘される²²⁾。この点に関して、本節では、次の2点から推計結果の比較検討を行う。

第一に、世帯サンプルが、地区の制約表の世帯・人口分布と大きく異なる場合、IPUとSAともに適合度が低下すると考えられる。そこで、制約表と世帯サンプルの度数分布の類似度を求め、それと適合度との関連性を検討する。類似度として、地区ごとに、3つの制約表とそれに対応する世帯サンプルの度数分布の相関係数を求めた。相関係数が1.0に近いほど、両者の世帯・人口構成の類似度が高いと判断できる。その分析結果として、各地区の相関係数と世帯あたりOverall TAEの散布図を第5図に示す。両手法ともに、相関係数が0.9以上であれば、大半の地区で世帯あたりOverall

TAE のスコアが1 よりも小さくなる。しかし、いずれの手法において、相関係数が低下すると、点のばらつきが大きくなり、適合度が低い地区が増加することがわかる。ただし、両手法を比較すると、相関係数が0.9 未満になると、IPU の方が点のばらつきがより大きくなる。したがって、SA の方が、世帯サンプルと制約表の度数分布が多少異なっていたとしても、それに対処する形で、小地域人口マイクロデータを生成できるものと考えられる。

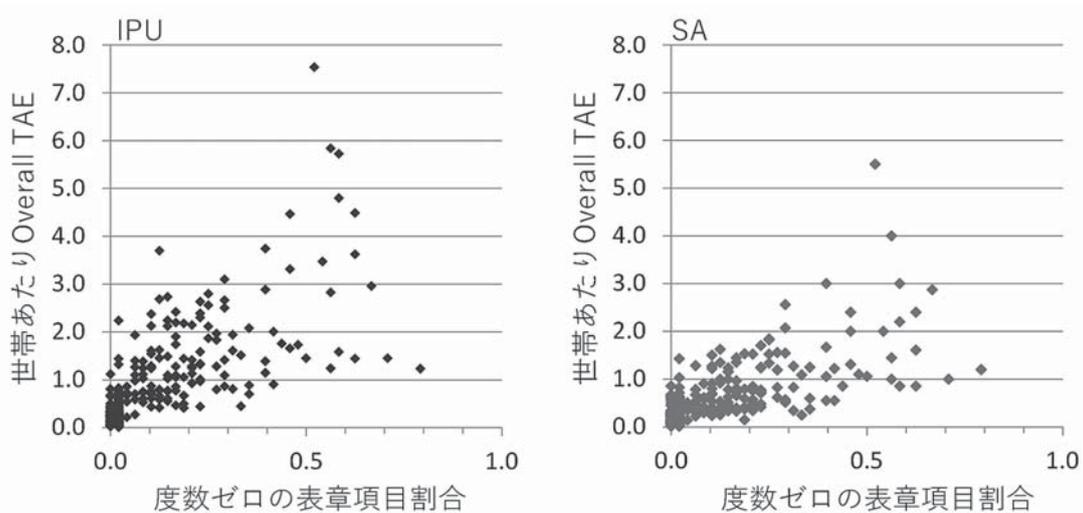
第二に、地区の世帯・人口構成が画一的な場合、制約表中の表章項目が度数ゼロになる可能性が高くなると想定される。そこで、制約表中にある度数ゼロの表章項目数の割合を求め、それと適合度の関連性を検討した。第6図に示す散布図からは、度数ゼロの表章項目割合が高いほど、世帯あたり Overall TAE のスコアが増加する傾向にあることがわかる。つまり、世帯・人口構成に大きな偏りがある地区では、両手法とも高い適合度を達成することが難しい。しかし、両手法を比較する



第5図 制約表と世帯サンプルの相関係数別に見た IPU と SA の推計結果

相関係数は、国勢調査小地域集計と世帯サンプルの集計表の度数分布に基づく。

地区別の結果をプロットした。



第6図 度数ゼロの表章項目割合別に見た IPU と SA の推計結果

地区別の結果をプロットした。

と、IPU に対して、SA の方が、度数ゼロの表章項目割合が大きくとも、世帯あたり Overall TAE のスコアが低い水準でばらつきが小さく、比較的良好な結果が導かれることがわかった。ただし、度数ゼロの表章項目割合は、地区の世帯数が少ないと高くなるため、その点に留意する必要がある。

V. 考察・結論

本研究では、京都市内の3つのPTゾーンを検証地域として選定し、パーソントリップ調査の世帯サンプルを用いて、IPU と SA による小地域人口マイクロデータの推計結果を比較検討した。本研究成果の知見と今後の課題は、次のように整理できる。

第一に、小地域人口マイクロデータの適合度指標 (Overall TAE) に関して、86%の町丁目において、IPU よりも SA の方で、より良好な結果が得られた。このような SA の優位性は、地域特性が異なる3つのPTゾーン別にみても同様であった。ただし、IPU の拡大係数は小数值であるため全世帯サンプルが推計結果に反映されるが、SA の場合は整数値に限定されるため、結果的に約7%の世帯サンプルが、小地域マイクロデータで使用されていない。対象地域に本来あるべき一部の世帯サンプルが除外されたことは、適合度の点で問題がないとしても、母集団の代表性を考えた場合、望ましい結果ではないと言える。このような課題に対しては、1世帯サンプルを0.5世帯や0.1世帯とカウントすることで、より多くの世帯サンプルが使用されるようになり、結果的に、全世帯サンプルが利用される確率が高くなる。

第二に、推計に使用できる世帯サンプル数が極端に少ない場合、両手法とも適合度が大きく低下することがわかった。特に、そのような傾向は、IPU の方でより顕著であった。興味深い点として、3つの国勢調査小地域集計を制約表に考慮した今回のケースでは、世帯サンプル数が60～70程度を境に、適合度が大きく変化することが示された。このことは、いずれの手法ともに、地区の総世帯数に応じて世帯サンプル数を増やす必要があるというよりも、一定数の世帯サンプルを用意しなければならないことを示唆している。今後の研究では、制約表の種類や表章項目数に応じて、最低限必要な世帯サンプル数を明らかにしていく必要がある。

第三に、地区の世帯・人口構成の偏りについて、制約表と世帯サンプルの類似度及び度数ゼロの表章項目割合をもとに評価した結果、両手法の適合度は、類似度が低いほど、度数ゼロの割合が高いほど、低下することがわかった。ただし、いずれのケースにおいても、IPU と比較して SA の方が、推計が難しい条件下でも、世帯あたり Overall TAE のスコアが低い水準でばらつきが小さく、相対的に良好な結果が得られた。このような結果は、世帯サンプルの交換による方法や一定程度の改悪解を受け入れる工夫など、世帯・人口構成の偏りに対して柔軟に対処できる SA 特有の仕組みによるものであろう。この点に関しても、どの程度の世帯・人口構成の偏りであれば IPU や SA で対処できるのか、今後、より具体的に評価していく必要がある。

以上、本研究の結論としては、小地域人口マイクロデータの生成において、IPU よりも SA の方が総じて高い適合度を達成できると指摘できる。特に、世帯サンプル数が少ない地区や、従来、推計が難しい世帯・人口構成に偏りがある地区において、SA の方で、より高い適合度が示された点は重要であろう。今後、上記の研究課題を解決できる工夫を提案し SA や IPU の推計に組み込むことで、小地域人口マイクロデータの適合度をさらに改善できるのではないだろうか。そうなれば、冒

頭で述べたように、社会科学分野で急速に拡大するマイクロデータへの高いニーズに応えられる小地域人口マイクロデータを、今後、提供できるようになるものと期待できる。

付記

本研究は JSPS 科研費 15K16883 の助成を受けたものである。なお、本研究は、京都府交通需要マネジメント施策検証業務基礎調査の一環である。

注

- 1) Birkin, M. and Clarke, M.: SYNTHESIS-A Synthetic Spatial Information System for Urban and Regional Analysis: Methods and Examples, *Environment and Planning A* 20, 1988, pp.1645-1671.
- 2) Tanton, R.: A Review of Spatial Microsimulation methods, *International Journal of Microsimulation* 7-1, 2014, pp.4-25.
- 3) ①Ye, X., Konduri, K., Waddell, P., Sana, B. and Pendyala, R.M.: A Methodology to Match Distributions of Both Household and Person Attributes in the Generation of Synthetic Populations, Paper presented at the 88th Annual Meeting of the Transportation Research Board, 2009., ② Konduri, K. You, D., Garikapati, V. and Pendyala, R.: Application of an Enhanced Population Synthesis Model That Accommodates Controls at Multiple Geographic Resolutions, *Transportation Research Record: Journal of the Transportation Research Board* 2563, 2016, pp.40-50, doi: 10.3141/2563-08.
- 4) Müller, K. and Kay, W. A.: Hierarchical IPF: Generating a Synthetic Population for Switzerland, Paper presented at the 4th European Regional Science Association Conference 2011, 2011.
- 5) 浅見泰司・木戸浩司「国勢調査住宅関連統計の IPF 法による度数分布表推計の精度—東京大都市圏を例として—」、日本建築学会計画系論文報告集 514、1998、185-189 頁。
- 6) 石坂公一・内海康也・大橋佳子「小地域における居住状況分析用データの推計方法」、日本建築学会住宅系研究報告会論文集 6-6、2011、169-176 頁。
- 7) Huang, Z. and Williamson, P.: *A Comparison of Synthetic Reconstruction and Combinatorial Optimization Approaches to the Creation of Small-area Microdata*, Department of Geography Working Paper 2001/2, University of Liverpool, 2001.
- 8) Williamson, P., Birkin, M. and Rees, P. H.: The Estimation of Population Microdata by Using Data from Small Area Statistics and Samples of Anonymised Records, *Environment and Planning A* 30, 1998, pp. 785-816, doi:10.1068/a300785.
- 9) Williamson, P.: Community Care Policies for the Elderly, 1981 and 1991: A Microsimulation Approach. *Microsimulation for Urban and Regional Policy Analysis*, in Clarke, G. ed.: *Microsimulation for Urban and Regional Policy Analysis*, Pion, 1996, pp.64-78.
- 10) Ballas, D. and Clarke, G. P.: GIS and Microsimulation Approach to Local Labour Market Policy Analysis, *Computers, Environment and Urban Systems* 24, 2000, pp.305-330.
- 11) 花岡和聖「公的統計『匿名データ』を用いた小地域単位での地理空間分析の可能性—空間的マイクロシミュレーションによる地理的な合成マイクロデータの生成—」、人文地理 64-3、2012、195-211 頁。
- 12) 花岡和聖「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—国勢調査・匿名データの活用—」地域安全学会論文集 29、2016、247-255 頁。
- 13) Hanaoka, K. and Clarke, G. P.: Spatial Microsimulation Modelling for Retail Market Analysis at the Small-area Level, *Computers, Environment and Urban Systems* 31, 2007, pp.162-187.
- 14) 前掲 7)。
- 15) Williamson, P.: An Evaluation of Two Synthetic Small-area Microdata Simulation Methodologies: Synthetic Reconstruction and Combinatorial Optimisation, in Tanton, R. and Edwards, K. eds.: *Spatial Microsimulation: A Reference Guide for Users*, Springer, 2013, pp.19-47.
- 16) Ryan, J., Maoh, H., Kanaroglou, P.: Population Synthesis: Comparing the Major Techniques Using a

Small, Complete Population of Firms, *Geographical Analysis* 41, 2009, pp.181-203.

- 17) ①前掲 4)、② Lenormand, M. and Deffuant, G.: Generating a Synthetic Population of Individuals in Households: Sample-Free Vs Sample-Based Methods, *Journal of Artificial Societies and Social Simulation* 16-4, 2013, doi: 10.18564/jasss.2319.
- 18) Williamson, P.: *CO Instruction Manual*, Department of Geography Working Paper 2007/1, University of Liverpool, 2007.
- 19) 前掲 15)。
- 20) 花岡和聖「焼きなまし法を用いたパーソントリップ調査データの拡大補正法に関する研究—平成 12 年度京阪神都市圏パーソントリップ調査データを用いて—」、都市計画論文集 41-3、2006、91-96 頁。
- 21) 同様の結果は、京都府全域を対象とした研究でも認められる。前掲 20)。
- 22) 前掲 20)。

(本学文学部准教授)

Estimating a Geographically Disaggregated Synthetic Microdata Set:
A Comparative Study of Iterative Proportional Updating and Simulated Annealing

by
Kazumasa Hanaoka

The purpose of this paper is to compare goodness-of-fit scores of synthetic microdata sets generated by two different approaches: iterative proportional updating (IPU) and combinatorial optimization algorithm using simulated annealing (SA). Three zones in Kyoto City, Japan, with different regional characteristics wherein the Person Trip Survey was conducted were selected as the study areas. Household samples from the survey were used as initial datasets for both the approaches, and synthetic microdata sets were generated such that they agreed with the three selected constraining tables from the small area statistics of the Japanese population census. The results of this study are as follows:

(1) In comparison with the IPU, a synthetic microdata set generated by SA showed a better goodness-of-fit score for the 86% of the small areas. The same results were found for the three zones. Further, all initial household samples were included in a synthetic microdata set generated by IPU but not SA. Such missing samples are problematic in terms of fully representing an area population although SA achieved better results than IPU in the goodness-of-fit score.

(2) By reducing the number of household samples available for estimation, the effects on goodness-of-fit scores were examined for two selected small areas. The results indicated that the scores were suddenly changed when only a small proportion (5% or 10%) of the household samples were available for estimation. Interestingly, such a proportion was equivalent to approximately 60-70 household samples for both small areas, suggesting that the minimum number of samples is necessary for achieving good results.

(3) In terms of the differences in the distributions between household samples and constraining census tables, as per the similarity index and proportions of zero frequency cells, SA attained better fit than IPU in both cases. In particular, SA results showed that the range of scores was not as large as IPU even if a relatively larger deviation and more zero frequencies existed.

Therefore, combinatorial optimization algorithm using SA achieved better goodness-of-fit scores than IPU. Although a deeper analysis is necessary, this study provides information on determining an approach for generating a synthetic microdata set.