

質的属性を含む初期マイクロ世帯データの推定手法*

An Estimation Method of Household Micro-Data containing Discrete Attributes for the Base Year*

杉木直**・宮本和明***・大谷紀子****・Varameth VICHENSAN*****

By Nao SUGIKI**・Kazuaki MIYAMOTO***・Noriko OTANI****・Varameth VICHENSAN*****

1. はじめに

都市モデル分野においては、土地利用と交通の詳細な変化を記述するマイクロシミュレーションへの関心が高まっており、欧米諸国を中心として、複数の研究グループよっての研究開発および実際の都市への適用事例の蓄積が精力的に進められている¹⁾。近年の発表件数ベースで見るとメソスケールのモデルに比べて明らかに卓越していると言える²⁾。

居住立地モデルのような世帯を対象としたマイクロシミュレーションモデルの場合、各世帯には世帯収入、世帯人数、各世帯構成員の年齢、自動車保有、居住地、住宅タイプ等の多くの属性が定義される。このようなシミュレーションを実行するためには、基準年において全ての世帯に対してこれらの属性を定義したシミュレーション初期データを用意する必要がある。しかし、国勢調査や住民基本台帳等から個人や個別世帯に関するデータを入手することは、わが国はもとより一般の諸外国においても禁止されている。従って、マイクロシミュレーションモデルでは、国勢調査などの入手可能な集計データと、個別世帯の属性情報を追加的に提供するサンプル調査を組み合わせて、人口データ推計を実施する必要がある。

人口データの作成手法としては、まず世帯をいくつかのタイプに分類した上で、IPF法により各タイプに属する世帯数を推定する方法が多く用いられている。しかし、モデルにおいて多数の世帯属性が取り扱われる場合、IPF法では推定が困難になるという問題がある。また、これらとは異なる手法としては、個別世帯（一般にはエージェント）に対してモンテカルロサンプリング等を用いて複数の属性の組み合わせを設定したデータ（以降、マイクロデータと呼ぶ）を作成する手法が提案されている。

本研究は、シミュレーション基準年におけるマイクロ

*キーワード：マイクロシミュレーション、マイクロデータ、初期データ推定、世帯属性
**正員 修士（情報科学）（株）ドーコン総合計画部
（〒004-8585札幌市厚別区厚別中央1条5丁目4-1、
TEL011-801-1555、FAX011-801-1556）

***フェロー 工博 東京都市大学環境情報学部

****正員 博士（工学）東京都市大学環境情報学部

*****正員 博士（工学） Kasetsart University

データを作成するための体系的な手法を構築することを目的に実施している一連の研究に含まれるものであり、マイクロデータの属性が年齢等の連続変量の場合に関しては既に報告している³⁾⁴⁾。本稿ではその成果をもとに、マイクロデータに性別や世帯主との関係等の質的属性を含む場合に拡張することを目的としている。

本稿では、まず人口データ推計に関する既存研究のレビューを行い、既存推計手法の課題を整理する。その上で、サンプルマイクロデータによって得られる各属性間の相関性に関する情報、および既存統計による各ゾーンのコントロールトータル情報を利用して、シミュレーション基準時点のマイクロデータを推定するシステムを構築する。このシステムは、各世帯のマイクロ属性データを決定するための複数のアプローチによって構成するものである。その過程において、属性間の相関をマイクロデータに取り組み工夫を加えている。そして、構築されたシステムを、道央都市圏パーソントリップ調査で取得したデータに対して適用し、その有効性を検討している。

2. 人口・世帯データ推計に関する既存研究

(1) IPF法およびその拡張

Demingら⁵⁾によって提案され、Beckmanら⁶⁾によって人口データ推計問題⁷⁾への適用がなされたIPF法は、人口データ作成における一般的な手法である。同様の手法は、都市圏における世帯データ作成を対象として宮本ら⁸⁾によって提案されており、周辺分布を制約とした同時確率最大化問題として定式化されているが、その解法に用いる式の展開から基本的にはIPF法と同一となる。

IPF法では、多次元のテーブルにおける各セルの世帯数が生成される。次元は、限定的な数の世帯属性によって定義される。次元や属性間の相関性を考慮するためにサンプルデータセットが用いられる。低い次元への分配には国勢調査等によって与えられる周辺分布に一致することが条件となる。IPF法によって推定されたデータはマイクロシミュレーションにおいても十分活用可能であるが、基本的には個々の世帯別のマイクロデータではなく、あくまでもタイプ別の世帯数である。

Guoら⁷⁾はゼロセル値問題、および世帯および個人レベルの属性の両方に関する統計的分布のコントロールの

無能力さに関して緩和し、IPF法を改良している。また、Pritchardら⁹⁾も、モンテカルロシミュレーションを用いて各エージェントにより多くの属性を考慮する機能を追加し、IPF法を改良している。これらは非常に有用な拡張であるが、IPF法を用いる限りセルベースのアプローチであることには変わらない。

限定数のタイプのみを扱う場合は問題がないが、より複雑なマイクロデータを作成する場合、IPF法には限界がある。行列のセルに対する初期データの信頼性の観点において非常に高い基準を必要とする。IPF法では、0のセルを0.1もしくは0.01に設定するため、これらの設定は確率に影響を与えるが、理論的な根拠はない¹⁰⁾。

また、Moeckelら¹⁰⁾によって指摘されているこのようなセル要素の問題に加え、改良されたIPF法を含むこれらのセルベースのアプローチには次のような問題がある。

- ・属性およびそれらのカテゴリーを任意に設定することができない。
- ・年齢や収入などは本来連続変数であるが、カテゴリーを離散的に設定する必要があり、これらの設定はゾーニングにおけるMAUP (Modifiable Area Unit Problem) と同様のMTUP (Modifiable Type Unit Problem) と呼ぶべき問題を引き起こす。
- ・属性数が増えるほどゼロセル問題が顕著になり、信頼性が低下する。

IPF法は、限られた数の属性を設定することによって、コンピュータの計算負荷量を削減することを目的として考案されたものであるという側面もあるが、属性カテゴリーの組み合わせが膨大になってしまう傾向があり、その一方で大部分のセルが0となってしまう。

(2) モンテカルロサンプリング

Moeckelら¹⁰⁾はエージェントベースの手法を採用している。モンテカルロサンプリング手法が用いられており、マイクロシミュレーションモデルの実行に必要な多くの属性を、世帯の構成員ごとに考慮することが可能である。この場合、考慮可能な属性数は選択された属性間に有意義な関係が存在する場合のみに制限される。

本研究の推定手法は、これらのアプローチを発展させるものであり、まず、参考文献^{3) 4)}においては属性を連続変数に限定した場合の作成システムを構築している。このシステムでは、後述するように、相関する属性変数をモンテカルロシミュレーションで決定する際に、無相関変数である主成分を介して行う手法を新たに提案している。本稿ではこれに質的変数を含んだより一般的なマイクロデータの推計に拡張するものである。

3. 初期マイクロ世帯データ推定手法の構築

(1) 前提条件の設定

本研究では、既往研究^{3) 4)}においてエージェントベ

ースの手法を拡張し連続変数の属性を対象に構築した手法を、性別および世帯主との関係等の質的属性を含む場合に拡張したアプローチにより、マイクロデータ推計手法を構築する。構築にあたっては、以下のような人口データ推定問題を前提条件として設定している。

- ・対象エージェントは世帯およびその構成員とする。
- ・対象とする世帯属性は世帯人数および世帯構成（世帯内の世帯主との続柄の組み合わせによって定義）であり、世帯内の各世帯構成員は年齢、性別、世帯主との続柄を属性として持つ。
- ・対象地域においては、周辺制約データとして性別5歳年齢階層別の人口データおよび世帯人数別の世帯数が国勢調査より利用可能であるものとする。
- ・すべての世帯の世帯構成、各世帯構成員の年齢、性別、続柄情報を含む限定的な数の世帯サンプルが入手可能であるものとする。

(2) 推定手法

分析の基本的な考え方は次のとおりである。

- ・世帯人数別の世帯数に対して、世帯サンプルより各世帯の構成員の質的属性である性別、続柄を決定する。
- ・十分なサンプルが得られる世帯タイプに対しては、属性（各世帯構成員の年齢）間の相関性を考慮して年齢を決定し、出現頻度の低い世帯タイプ（ C_{rare} ）については世帯サンプルと同様の年齢を与える。
- ・人口データ推計においては、周辺分布（性別5歳年齢階層別人口）に一致するように世帯構成および各世帯構成員の年齢に関して調整を行う。
- ・推定、データ生成、調整はすべてモンテカルロ法等を用いて確率的に行う。

本推定手法の最も特徴的な点は、連続変数である属性間の相関性の処理方法である。まずサンプルにおける m 人世帯データの属性変数 ($X_{is} = (x_{1s}, \dots, x_{ms})$) を、主成分分析を用いて無相関変数 ($P_{is} = (p_{1s}, \dots, p_{ms})$) に変換する。

$$P = aX \quad (1)$$

世帯サンプルの無相関変数 p_{is} の値に基づいて、図 - 1 に示すように、 $i = 1 \sim m$ について累積曲線を作成する。式(1)より次式が導出される。

$$X = A^{-1}P = BP \quad (2)$$

世帯を生成する際には属性 i に対して乱数 ran_i を発生させ、図 - 1 の累積曲線より $i = 1 \sim m$ について生成世帯の各構成員に対する p_{is} を求める。生成世帯の x_i は式(2)より $i = 1 \sim m$ について求められる。同様の処理を対象地域内の全ての世帯人別世帯数に対して実行することで、世帯推定の初期データセットが作成される。この初期データセットは、周辺分布である性別5歳階級人口を満たさないため、モンテカルロ法を用いて世帯をランダム抽出し、世帯構成員の年齢が更新された新たなデータによ

って置き換える。このような調整処理を、周辺分布を満たすまで繰り返し実行する。本研究の初期マイクロデータ推定手法のフローを図 - 2 に示す。

4. 初期マイクロ世帯データ推定手法の適用

(1) 利用データ

本研究では、第4回道央都市圏パーソントリップ調査データを用いて、構築された初期マイクロデータ推定手法に関するケーススタディを行う。第4回道央都市圏パーソントリップ調査では、19,394世帯について世帯構成員の詳細情報を含むマイクロデータが取得されている。このうち、10,000世帯をランダム抽出したデータを母集団データセットとしてケーススタディを行う。母集団データは単身世帯から7人世帯までの様々な世帯構成からなり、人口は24,115人分のデータである。また、各世帯の構成員を性別年齢5歳階級別に集計し、人口に関する周辺制約データを作成した。

母集団に対してサンプリング調査が実施されたことを想定し、母集団データセットより1,000世帯を抽出し世帯サンプルとして設定した。これらのサンプル世帯データにおいて、自由度10以上のサンプル数が得られる10の世帯タイプ(単身男、単身女、夫婦、世帯主女+子供女、夫婦+子供男、夫婦+子供女、夫婦+母親、夫婦+子供男2人、夫婦+子供男+子供女、夫婦+子供女2人)について、主成分分析によるパラメータ行列 A の逆行列よりパラメータ B を導出した。推定結果の一部として、夫婦と2人の男の子供からなる4人男世帯に対する結果を

表 - 1 に示す。乱数 ran_i によって与えられる非相関変数 p_{is} を、パラメータ B に作用することによって、世帯構成員の属性変数 x_{ms} が生成される。

(2) シミュレーション結果

ケーススタディとして、図 - 2 の初期マイクロ世帯データ作成フローに従って、10,000世帯に対する各世帯の構成員データの推定を行った。部分的な検証ではあるが、世帯人数別に推定世帯データセットと母集団データセット(観測データ)に関して集計し、性別年齢階層別人口の再現状況を検証した。分析結果の一部として、単身世帯と3人世帯に関する結果を図 - 3 に示す。本手法による世帯マイクロデータの推計は母集団の構造をよく再現しており推定手法の妥当性が示されているが、3人世帯男性の31~40歳など部分的に乖離が大きい部分も見られる。シミュレーション時間等の制約により、現在の推定結果では、周辺人口制約に対して性別5歳年齢階級の人口誤差の総和が、総人口の5%以下となる条件で収束させ

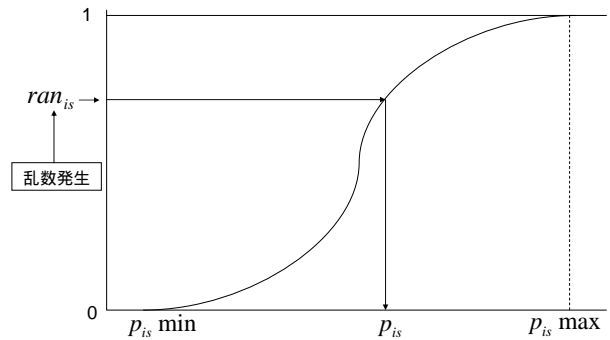


図 - 1 無相関変数を用いた相関の統合

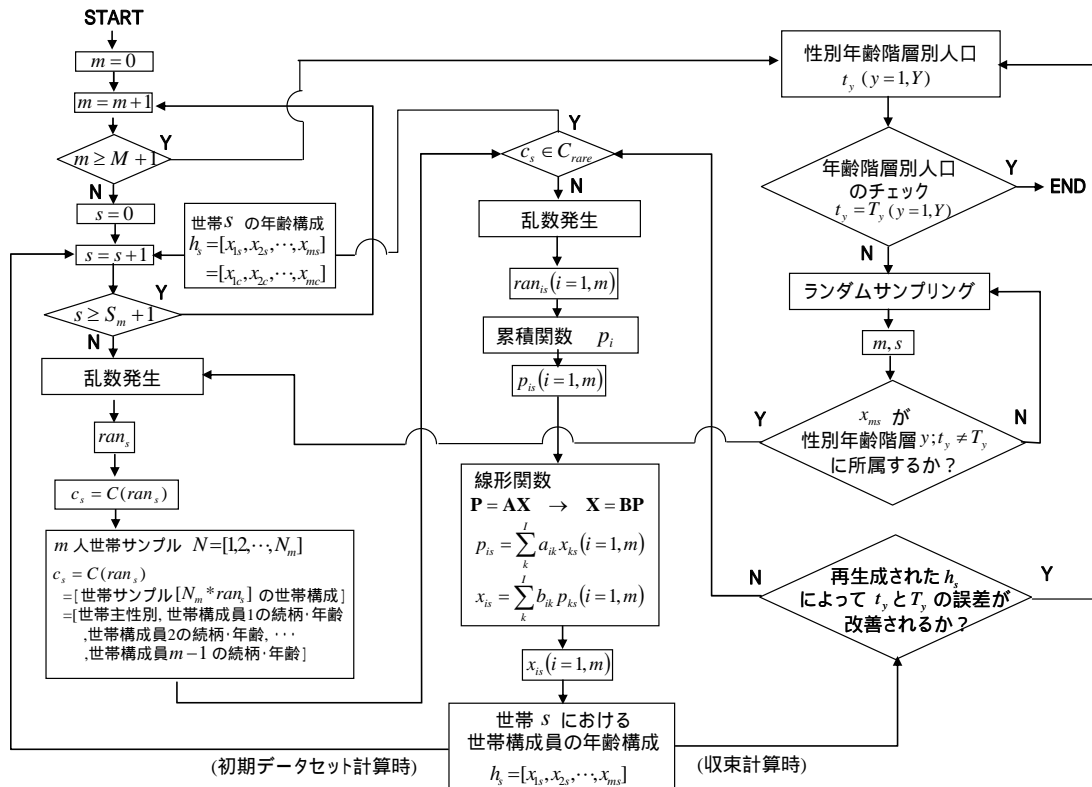


図 - 2 初期マイクロデータ推定フロー

ており、アルゴリズム等を改良し、より精度の高い推定の開発を行ってゆく予定である。

5. おわりに

本研究では、新たなエージェントベースアプローチによる世帯マイクロデータ推定のためのシステムに関して、これまで連続変量の属性に限定していたものを質的変数まで拡張した。この結果、実際のシミュレーションに必要な最低限のマイクロデータの作成は可能になったと考えられる。また、構築された部分の妥当性はケーススタディによってある程度確認された。

今後は、マイクロデータを要素とする2つの集合間の適合度を評価するために新たに作成した指標⁽¹¹⁾⁽¹²⁾を活用することにより、現システムの改良と適用範囲の拡張を行う予定である。

表 - 1 パラメータ推定結果
(4人世帯：夫婦+子供2人世帯)

b_{ik}	p_1	p_2	p_3	p_4	C
x_1	0.531	0.486	0.471	0.509	9.643
x_2	-0.237	0.384	0.593	-0.668	3.833
x_3	-0.725	-0.160	0.411	0.529	3.214
x_4	-0.369	0.769	-0.508	0.121	9.429

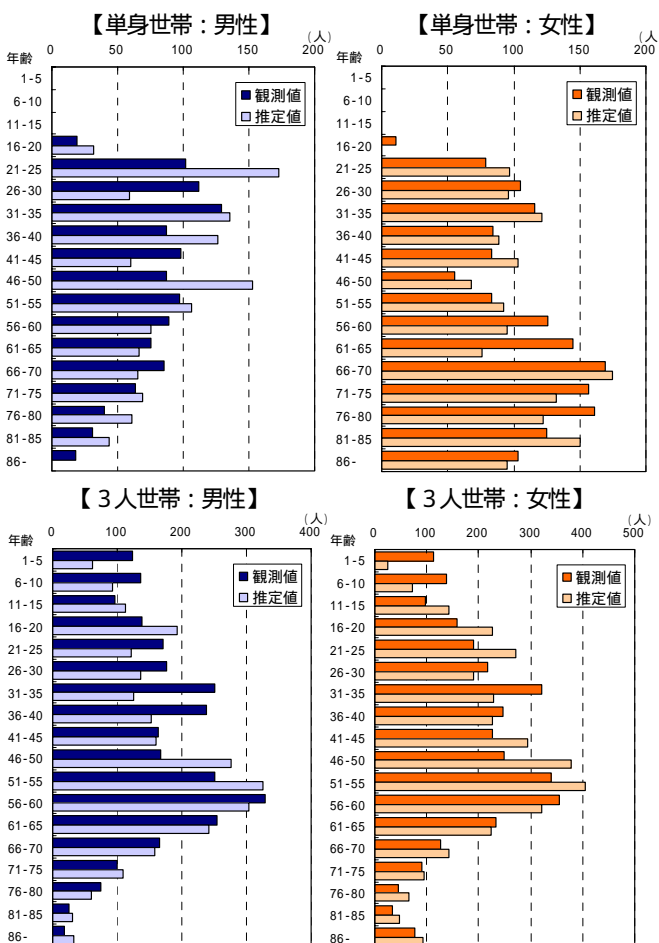


図 - 3 マイクロ世帯データの推計結果

なお、本論文は、平成20~21年度科学研究費補助金(基盤研究(B)), 課題番号: 20360232, 研究課題名: 詳細属性情報を含む世帯の空間分布予測のためのマイクロシミュレーションシステム)の研究成果の一部を取りまとめたものである。ここに記して謝意を表したい。

参考文献

- 1) Wegener, M.: Overview of Land-Use Transport Models, Proceedings of CUPUM '03, Sendai, CD-ROM, 2003.
- 2) 宮本和明, 北詰恵一, 鈴木温: 世界における実用都市モデルの実態調査とその理論・機能と適用対象の体系化, 平成18年度~19年度科学研究費補助金(基盤研究(C)), 課題番号: 18560524) 研究成果報告書, 2008.
- 3) 杉木直, 宮本和明, Varameth VICHENSAN: 土地利用マイクロシミュレーションにおける初期マイクロ世帯データの推定手法, 土木計画学研究・講演集, 39, 2009
- 4) Miyamoto, K., and Sugiki, N.: An Estimation Method of Household Micro-Data for the Base Year in Land-Use Micro Simulation, Proceedings of CUPUM '09, Hong Kong, CD-Rom, 2009.
- 5) Deming, W.E. and Stephan, F.F.: On a Least Squares Adjustment of a Sampled Frequency Table when the Expected Marginal Totals are known, Annals of Mathematical Statistics, Vol.11, pp.427-444, 1940.
- 6) Beckman, R. J., Baggerly, K. A. and McKay, M. D.: Creating Synthetic Baseline Populations, Transportation Research A, Vol.30, No.6, pp.415-435, 1996.
- 7) Guo, J. Y. and Bhat, C. R.: Population Synthesis for Microsimulating Travel Behavior, Transportation Research Record: Journal of the Transportation Research Board, No.2014, pp.92-101, 2007.
- 8) 宮本和明, 安藤淳, 清水英範: 非集計行動分析に基づく都市圏住宅需要モデル, 土木学会論文集, No.365/ -4, pp.79-88, 1986.
- 9) Pritchard, D. R. and Miller, E.J.: Advances in Agent Population Synthesis and Application in an Integrated Land Use / Transportation Model, 88th Annual Meeting Compendium of Papers, Transportation Research Board, DVD, 2009.
- 10) Moeckel, R., Spiekermann, K., and Wegener, M.: Creating a Synthetic Population, Proceedings of CUPUM '03, Sendai, CD-ROM, 2003.
- 11) 大谷紀子, 杉木直, 宮本和明: 土地利用マイクロシミュレーションにおける質的属性を含むマイクロデータの適合度評価, 40, 2009
- 12) Otani, N., Miyamoto, K., and Sugiki, N.: Goodness-of-Fit Evaluation Method between Observed and Estimated Sets of Micro-Data in Land-Use Micro-Simulation, Proceedings of CUPUM '09, Hong Kong, CD-Rom, 2009.