

居住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 including Housing 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の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事例および実際の都市への適用事例の蓄積が進められている¹⁾。

住宅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が、シミュレーションを実行するためには、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年次データを用意する必要がある。しかし、住民基本台帳などから個人や個別世帯に関するデータを入手することは一般的に困難であり、またプライバシー保護の観点からも望ましくない。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、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、世帯タイプを設定した上で、IPF法によりタイプ別世帯数を推定する方法^{2) 3) 4)}が多く用いられているが、多数の世帯属性が取り扱われる場合、ゼロセル問題による信頼性の低下等の問題がある。また、これらとは異なる手法として、モンテカルロサンプリング等を用いて個別世帯に対し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）を作成する手法⁵⁾が提案されている。

本研究グループでは、シミュレーション基準年におけるマイクロデータの作成に関して、体系的な推定手法の開発に取り組んでいる^{6) 7) 8)}。本稿は参考論文⁹⁾に英文で

*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(情報理工学)、東京都市大学環境情報学部

*****正員、博士(工学)、Kasetsart University

発表した内容を基本とし、世帯構成員の年齢、性別、世帯主との関係を属性とする場合に関して提案した先行研究⁹⁾の評価方法を発展させ、居住ゾーンと住居タイプを属性として加えた初期マイクロ世帯データの推定手法に拡張するものである。具体的には、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に加えて、既存統計による各ゾーンの住宅タイプ別住居数をコントロールトータルとして利用した、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。その上で、道央都市圏パーソントリップ調査データより抽出した10,000世帯のマイクロデータを母集団とした適用を行う。母集団よりランダム抽出された1,000サンプルを用いて住宅タイプおよび居住ゾーンの配分確率を定義して初期分布を推定し、構築されたシステムの有効性を検討する。

2. 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

本研究の推定手法は、Moeckelら⁷⁾によって構築されたエージェントベースの手法を拡張したアプローチを発展させたものであり、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、参考文献^{3) 4)}において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した作成システムを構築している。これらのシステムでは、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た。

本稿ではこれに住居タイプと居住ゾーンを属性として加えたマイクロデータの推計に拡張するものであり、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。

- ・対象地域においては、周辺制約データとしてタイプ別住宅数が国勢調査より利用可能であるものとする。
- ・対象地域はゾーンに分割されており、各ゾーン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、居住ゾーン、住宅タイプの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。

これらの前提条件のもとで、対象地域におけるすべての世帯の世帯構成員の年齢を推定生成することが目的となる。以上の設定は非常に限定的なものであるが、世帯の相違を取り扱うという点で既存手法を拡張するものであると考えられる。

(2) マイクロデータの定義

先行研究⁹⁾において、マイクロデータは年齢、性別、世帯主との続柄を属性の集合として定義している。 m 人からなる世帯 s のデータ h_{ms} を式(1)で表わす。

$$h_{ms} = \{c_{ms}, x_{ms}\} \quad (1)$$

ここで、 c_{ms} は m 人からなる世帯 s の世帯構成、 x_{ms} は年齢構成である。世帯データは、世帯構成員カテゴリの一般形に対する年齢ベクトルとして表現される。世帯構成員カテゴリは、世帯主との続柄および性別によって定義される。マイクロデータセット全体の一般的な表現は式(2)で表わされる。

$$A = \{a_i = (a_{i1}, a_{i2}, \dots, a_{iR}) \mid 1 \leq i \leq N\} \quad (2)$$

ここで、 R は構成員タイプ数、 N は対象地域の世帯数、 a_{ik} は世帯構成員 k の年齢である。世帯 i に世帯構成員 k が存在しない場合、 a_{ik} には不在を表すダミー変数(999)を割り当てる。

本稿では、式(1)を住宅タイプおよび居住ゾーンを属性として含む場合に拡張し、個々の世帯マイクロデータを式(3)のように定義する。

$$(hhl)_{ms} = \{c_{ms}, x_{ms}, j, z\} \quad (3)$$

ここで、 j は住宅タイプ、 z は居住ゾーンである。

(3) 推定手法

分析の基本的な考え方は次のとおりである。

- ・世帯サンプルから得られる確率に基づいて初期年次の人口マイクロデータの推定を行う。
- ・各段階の推定、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。
- ・サンプルから得られる属性間の相関性(各世帯構成員の年齢)を考慮してマイクロデータの属性を決定する。
- ・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、住宅タイプ、居住ゾーンに関して調整を行う。

以上の前提条件及び定義に基づいた初期マイクロデータの推定フローを図1に示す。フロー図の左半分は世帯構成員の属性を推定するものであり、右半分は住宅タイプ及び居住ゾーンを推定するものである。以下に、フロ

ーの各ステップについて詳細を記述する。

- ・(1)–(3)：世帯構成員数 1 からサンプルデータセットにおける世帯構成員数の最大値 M まで推定を実行する。
- ・(4)–(6)：世帯 1 から S_m まで推定を実行する。 S_m は世帯構成員 m 人の世帯数であり、対象地域におけるコントロールトータルとして、国勢調査により外生的に与えられるものとする。
- ・(7)–(9)：世帯 s に対して乱数 ran_s ($0 \leq ran_s \leq 1$) を発生させ、世帯構成 c_{ms} を世帯サンプル j_{ms} より決定する。
- ・(10)–(11)：出現頻度の低い世帯タイプ (C_{rare}) については、年齢構成 x_{ms} は世帯サンプル j_{ms} と同様に決定する。
- ・(10), (12)–(18)：十分なサンプルが得られる世帯タイプに対しては先行研究における手法⁹⁾を用いて属性(各世帯構成員の年齢)間の相関性を考慮して年齢を決定する。
- ・(19)–(20)：初期データセットによる性別年齢階層別人口 t_{gy} は周辺制約である年齢階層別人口 T_{gy} を満たさない。
- ・(21)–(24)：モンテカルロ法により世帯(m, s)をランダム抽出し、 x_{ms} が所属する性別年齢階層(g, y)において $t_{gy} > T_{gy}$ である場合は x_{ms} の再推定を実行し、そうでない場合は世帯(m, s)の再サンプリングを行う。
- ・(7)–(17), (25)–(27)：推定データと周辺制約間の誤差が減少する場合は、再推定された世帯構成 c_{ms} および各世帯構成員の性別と年齢によって x_{ms} を置換する。
- ・(28)：推定データによる性別年齢階層別人口が周辺制約を満たすまで、調整計算を繰り返し試行する。
- ・(29)–(34)：ステップ(1)–(6)と同様。
- ・(35)–(37)：世帯 s に対して乱数 ran_s ($0 \leq ran_s \leq 1$) を発生させ、サンプルデータより構築されたロジットモデルにより住宅タイプ j を決定する。
- ・(38)–(41)：世帯 s に対して乱数 ran_s ($0 \leq ran_s \leq 1$) を発生させ、サンプルデータより構築されたロジットモデルにより居住ゾーン z を決定する。
- ・(42)–(43)：初期データセットによるタイプ別ゾーン別住宅数 t_{jk} は周辺制約であるタイプ別ゾーン別住宅数 T_{jk} を満たさない。
- ・(44)–(47)：モンテカルロ法により世帯(m, s)をランダム抽出し、 $(hhl)_{ms}$ が所属する(j, z)において $t_{jk} > T_{jk}$ である場合は住宅タイプおよび居住ゾーンの再推定を実行し、そうでない場合は世帯(m, s)の再サンプリングを行う。
- ・(35)–(40), (48)–(50)：推定データと周辺制約間の誤差が減少する場合は、再推定された住宅タイプおよび居住ゾーンによって $(hhl)_{ms}$ の(j, z)を置換する。
- ・(51)：推定データによるタイプ別ゾーン別住宅数が周辺制約を満たすまで、調整計算を繰り返し試行する。

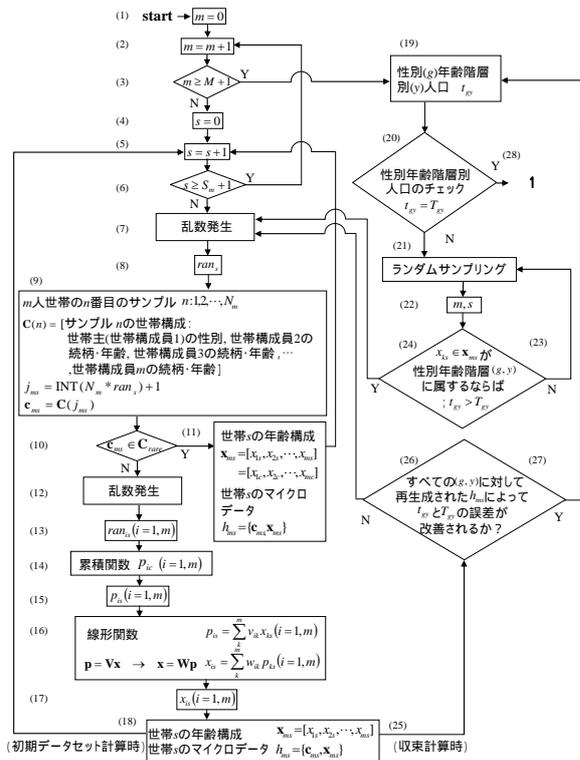


図 - 1 初期マイクロデータ推定フロー

4. 初期マイクロ世帯推定システムの適用

(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において取得された世帯構成員、住宅タイプ、居住ゾーンの詳細情報を含む19,394世帯マイクロデータを用いて、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。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Aとした。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。住宅タイプは5タイプ（戸建持家・戸建賃貸・集合持家・集合賃貸・その他）設定し、ゾーンの設定は対象地域を8ゾーンに区分した。また、世帯属性に関する集計により、世帯人員別世帯数、性別年齢階層別人口、タイプ別ゾーン別住宅数に関する周辺制約データを作成した。

(2) サンプルデータおよびパラメータ推定

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データセットBを設定した。これらのサンプルより20の世帯構成員タイプ、および16の世帯タイプを設定した。自由度10以上のサンプル数が得られる10の世帯タイプについては、先行研究⁹⁾において推定された世帯構成に関するパラメータを適用し、推定フローにおけるステップ(10)–(11)の方法で年齢構成を決定する。

住宅タイプ選択およびゾーン選択に関するロジットモデルパラメータについても同様のサンプルを用いて推定を行い、それぞれ表 - 1 および表 - 2 に示すような結果が得られた。

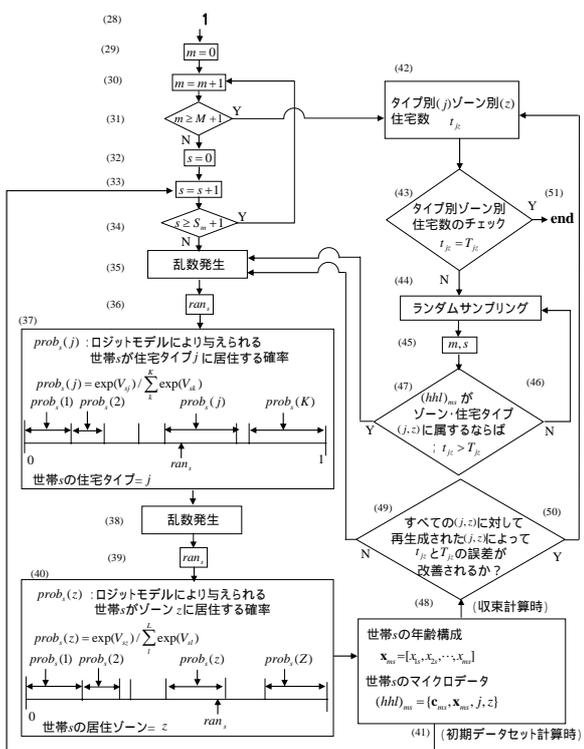


表 - 1 住宅タイプ選択パラメータ推定結果

説明変数	パラメータ値 (t 値)			
	戸建持家	戸建賃貸	集合持家	集合賃貸
世帯人員	0.667(3.32)	0.645(2.72)	0.405(1.95)	0.503(2.47)
世帯主年齢	0.166(4.28)	-0.116(-2.12)	0.106(2.64)	0.128(3.24)
サンプル数: 1,000 対数尤度: -1,118.6 尤度比: 0.304				

表 - 2 ゾーン選択パラメータ推定結果

説明変数	パラメータ値 (t 値)		
	zone 1	zone 2 - 5	zone 6 - 8
都心距離	-	-0.010 (-0.76)	-
世帯人員	-	0.442 (6.21)	0.672 (7.23)
世帯主年齢	-	-	-0.055 (-2.45)
戸建持家ダミー	-	1.553 (6.21)	0.902 (2.39)
集合持家ダミー	-	-	-2.045 (-4.77)
集合賃貸ダミー	-	1.079 (5.14)	-0.639 (-1.78)
サンプル数: 1,000 対数尤度: -757.8 尤度比: 0.307			

(3) シミュレーション結果

図 - 1 の初期データ作成フローに従って、ケーススタディに対する各世帯の構成員データの推定を行った。また、比較のため、簡易手法として、推定フローにおけるステップ(12)–(18)を実行せず、従来型の固定的な年齢決定による推定を同時に実施した。

提案手法(E1)および簡易手法(E2)による推定結果についてそれぞれ8回ずつ推定を行い、観測データからの誤差の平均値および標準偏差を算出した。戸建持家住宅について世帯構成ごとの結果を表 - 3 に示す。一部結果が逆転している世帯構成も見られるが、特に各世帯構成員の年齢間の相関性を考慮して年齢が決定されている世帯タイプにおいて、提案手法による推定結果は簡易手法よりも観測データとの誤差が少なく良好な結果が得られる傾向がある。ゾーン別世帯数の推定結果について、夫婦

+子供男+子供女の世帯構成の場合を例として図 - 2 に示すが、提案手法と簡易手法の間に大きな差は見られなかった。両手法は世帯構成員の年齢決定を除き同一であり、ゾーン決定に関しては年齢構成以外の要素が支配的であるためと考えられる。また、別途開発したマイクロデータの適合度評価手法¹⁰⁾を提案手法(E11~18)および簡易手法(E21~28)の推定結果に適用し、適合度の評価を行った結果を表 - 4 に示す。一部の推定結果では提案手法の方が大きな値となっているが、平均的には簡易手法よりも提案手法の方が平均値、標準偏差ともに小さな値となっており、良好かつ安定的な結果が得られているものと考えられる。

5. おわりに

本研究では、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に関して、先行研究を住宅タイプおよび居住ゾーンの属性を含む場合に拡張し、推定システムを構築した。また、実データを対象としたケーススタディを実行し、別途構築中の適合度評価手法を用いて検証を行った結果、その有効性がある程度確認された。

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。ここに記して感謝の意を表したい。

参考文献

- 1) 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.
- 2) 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.
- 3) 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.
- 4) 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.
- 5) Moeckel, R., Spiekermann, K., and Wegener, M.: Creating a Synthetic Population, Proceedings of CUPUM '03, Sendai, CD-ROM, 2003.
- 6) 杉木直, 宮本和明, Varameth VICHIANSEN: 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, 第39回土木計画学研究発表会論文集, CD-ROM, 2009.

表 - 3 提案手法および簡易手法による推定結果による観測データからの誤差（戸建持家住宅）

世帯構成	観測サンプル数	提案手法 (E1)		簡易手法 (E2)	
		平均値	標準偏差	平均値	標準偏差
単身・男	300	207.8	24.9	210.1	23.3
単身・女	591	88.0	17.9	93.1	16.7
夫婦	176	-36.4	5.7	-39.0	12.3
本人女+子供女	1535	-131.4	24.0	-128.1	22.5
2人世帯その他	120	-23.9	10.0	-22.1	6.7
夫婦+子供男	150	1.4	12.4	2.1	12.9
夫婦+子供女	456	-81.0	15.8	-96.4	14.5
夫婦+母親	507	12.3	21.5	34.1	14.6
3人世帯その他	83	25.8	11.3	20.8	7.9
夫婦+子供男1	132	-30.4	5.9	-37.0	13.4
+子供男2					
夫婦+子供男+子供女	193	28.9	10.2	27.4	4.9
夫婦+子供女1	394	11.3	23.3	3.4	23.3
+子供女2					
4人世帯その他	192	-31.1	9.4	-26.0	12.9
5人世帯	322	-30.4	8.4	-30.8	8.6
6人世帯	70	-11.5	2.3	-11.0	4.9
7人世帯	12	0.8	1.8	-0.6	2.4

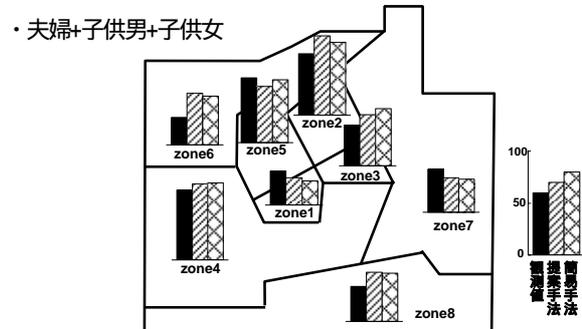


図 - 2 居住ゾーン別世帯構成別世帯数の例

表 - 4 適合度評価結果

	提案手法		簡易手法		
	平均	標準偏差	平均	標準偏差	
E11	0.17115	0.00065	E21	0.18456	0.00099
E12	0.17721	0.00198	E22	0.18506	0.00203
E13	0.17738	0.00067	E23	0.18522	0.00149
E14	0.18378	0.00329	E24	0.19013	0.00169
E15	0.19247	0.00202	E25	0.19262	0.00277
E16	0.19405	0.00206	E26	0.20019	0.00347
E17	0.19985	0.00102	E27	0.20103	0.00406
E18	0.20570	0.00136	E28	0.20117	0.00405
平均	0.18770	0.00163	平均	0.19250	0.00257

- 7) 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.
- 8) 杉木直, 宮本和明, 大谷紀子, Varameth VICHIANSEN: 質的属性を含む初期マイクロ世帯データの推定手法, 第40回土木計画学研究発表会論文集, CD-ROM, 2009.
- 9) Miyamoto, K., Sugiki, N., Otani, N. and Vichiansan, V.: Agent-based estimation method of household microdata for base year in land use microsimulation, Compendium of the 89th Transportation Research Board Annual Meeting, DVD, 2010.
- 10) 大谷紀子, 杉木直, 宮本和明: 居住情報を含む世帯マイクロデータの推定集合の適合度評価, 第41回土木計画学研究発表会論文集, CD-ROM, 2010.