

# 総合的属性からなる世帯マイクロデータの推計方法\*

An Estimation Method of Household Micro-Data Consists of Comprehensive Attributes\*

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<sup>1)</sup>。

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。

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<sup>2)3)4)</sup>。本稿は、世帯構成員の年齢、性別、世帯主との関係に居住ゾーンと住居タイプを属性として考慮した加えた場合に関して提案した先行研究<sup>5)6)</sup>をベースとして、自動車保有台数及び世帯所得を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。先行研究では、周辺制約が存在する属性についてのみ取り扱ってきたが、本稿では周辺制約が存在せず世帯に付加される属性として、離散変数である自動車保有台数と、連続変数である世帯所得を新たに加え

\*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

た点に特徴がある。初期マイクロデータの属性としてはこれらの他にも就業・就学状況、通勤地など様々なものが考えられるが、本稿で新たに追加したものを含む一連の推計手法のいずれかにより作成可能であり、総合的な属性からなる初期マイクロ世帯データの推定手法が構築されるものとする。具体的には、自動車保有台数についてはロジット型の離散選択確率にモンテカルロ法を適用し、世帯所得については所得回帰式の標準誤差にボックス・ミュラー法による正規乱数を考慮することにより、各属性を決定する手法を構築する。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サンプルを用いて自動車保有台数の配分確率および所得回帰式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<sup>7)</sup>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<sup>2)3)</sup>においては属性を連続変数に限定した場合、参考文献<sup>4)</sup>においてはこれに質的変数を含んだ場合、参考文献<sup>5)6)</sup>においてはさらに居住属性として居住ゾーンと住宅タイプを加えた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、周辺制約に対する調整手法を提案した。

本稿ではこれらに加え、周辺制約のない属性として離散変数である自動車保有台数と、連続変数である世帯所得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、以下のようなマイクロ世帯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。
- ・対象地域はゾーンに分割されており、周辺制約データとしてタイプ別住宅数、および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、自動車保有台数、世帯所得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・性別、居住ゾーン、住宅タイプ、自動車保有台数、世帯所得を推定生成することが目的となる。

## (2) マイクロデータの定義

先行研究<sup>9)</sup>において、マイクロデータは年齢、性別、世帯主との続柄、住宅タイプおよび居住ゾーンを属性の集合として定義している。居住ゾーン  $z$  で住宅タイプ  $j$ 、に居住する  $m$  人からなる世帯  $s$  のデータ  $h_{ms}$  を式(1)で表わす。

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

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)を自動車保有台数および世帯所得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。

$$h_{ms} = \{c_{ms}, x_{ms}, j, z, nc, inc\} \quad (3)$$

ここで、 $nc$  は自動車保有台数、 $inc$  は世帯所得である。

## (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。

- ・調整後のマイクロ世帯に、世帯主年齢、居住ゾーン、住宅タイプ等の付加された属性を含む世帯属性を説明変量として、ロジット型の選択モデルにより自動車保有台数を決定する。

- ・同様の世帯属性を説明変量として所得回帰式に、所得回帰式の標準誤差にボックス・ミュラー法による正規乱数を考慮して、確率的に世帯所得を決定する。

ボックス・ミュラー法は、2つの一様乱数から正規正規乱数を発生させる方法である。0以上1未満の2つの一様乱数  $ran1$ 、 $ran2$  を次式で変換すると、2つの独立した平均  $\mu$ 、標準偏差  $\sigma$  の正規乱数  $RAN1$ 、 $RAN2$  が生成される。

$$RAN1 = \sigma \cdot (-2 \log_e ran1)^{1/2} \cdot \cos 2\pi \cdot ran2 + \mu \quad (4)$$

$$RAN2 = \sigma \cdot (-2 \log_e ran1)^{1/2} \cdot \sin 2\pi \cdot ran2 + \mu \quad (5)$$

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。図左は世帯構成員の属性を推定するものであり、図中は住宅タイプ及び居住ゾーンを推定するものである。また、図右は自動車保有台数および世帯所得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<sup>9)</sup>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。

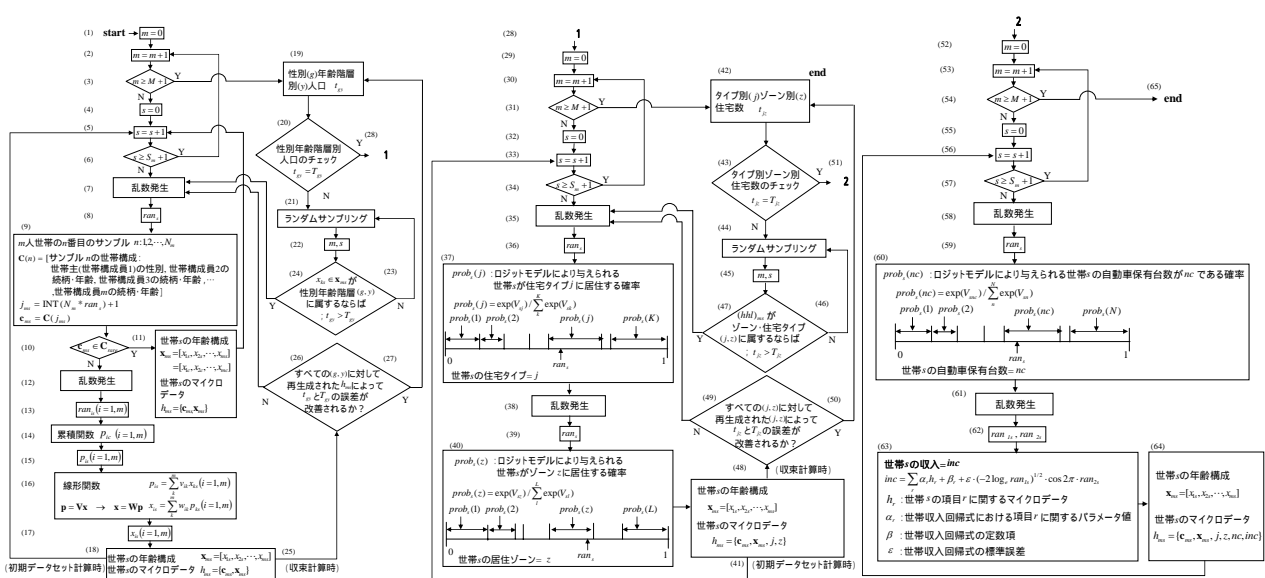


図 - 1 初期マイクロデータ推定フロー

- (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  $t_{jz}$  は周辺制約であるタイプ別ゾーン別住宅数  $T_{jz}$  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、 $h_{ms}$  が所属する  $(j, z)$  において  $t_{jz} > T_{jz}$  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  $h_{ms}$  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。
- (52)–(57) : ステップ(1)–(6)と同様。
- (58)–(60) : 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自動車保有台数  $nc$  を決定する。
- (61)–(62) : 世帯  $s$  に対して 2 つの一樣乱数  $ran_{1s}, ran_{2s}$  ( $0 \leq ran_{is} \leq 1 (i=1,2)$ ) を発生させる。
- (63)–(64) : サンプルデータより構築された収入回帰式により推定される世帯収入に、2 つの一樣乱数より生成した正規乱数を考慮した回帰式の標準誤差を加え確率的に世帯収入  $inc$  を決定する。
- (51) : すべての世帯に関するを実行し、推定計算を終了する。

#### 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。推定対象とする属性のうち、世帯収入のみ対応データがないため、家計調査の世帯人員および世帯主年齢別年間収入十分位階級別世帯数、世帯人員別・世帯主年齢別世帯数に関する集計データを利用して、仮想の収入データを作成し、母集団データセットAに付加した。

##### ( 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<sup>4)</sup>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て年齢構成を決定し、住宅タイプ選択およびゾーン選択については先行研究<sup>5)</sup>において推定されたロジットモデルパラメータを適用して決定する。

同様のサンプルを用いて自動車保有台数の選択に関するロジットパラメータ、世帯収入に関する重回帰式のパラメータ推定を行い、それぞれ表 - 1 および表 - 2 に示すような結果が得られた。

### (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属性データの推定を行った。

自動車保有台数の推定結果を表 - 3 に示す。全体では4割以上の的中率となったが、台数が多いほどの中率が低くなっており、これらについては推定式等の見直しを行ってゆく予定である。また、図 - 2 に世帯所得推定結果として、世帯収入の観測値および推計値に対する累積世帯数を示す。推計結果は対象マイクロデータの所得分布状況を概ね表現していると考えられる。その他の世帯属性等、より詳細な分析結果については講演時に報告する予定である。

## 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、先行研究を自動車保有台数及び世帯所得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、実データを用いたケーススタディを実行し、その有効性を検証した。これらにより、総合的属性からなる初期マイクロ世帯データの推定手法が構築された。

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) 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, 第39回土木計画学研究発表会論文集, CD-ROM, 2009.
- 3) 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.
- 4) 杉木直, 宮本和明, 大谷紀子, Varameth VICHENSAN: 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.
- 5) Miyamoto, K., Sugiki, N., Otani, N. and Vichiensan, V.:

表 - 1 自動車保有台数選択パラメータ推定結果

説明変数	パラメータ値 (t 値)			
	0台	1台	2台	3台以上
世帯人数	- (-)	0.18(1.78)	0.51(5.28)	0.53(4.35)
単身世帯ダミー	- (-)	-1.67(-7.41)	-3.86(-5.27)	-1.91(-3.08)
世帯主				
20-29 歳	- (-)	- (-)	-2.27(-3.75)	- (-)
30-39 歳	- (-)	1.53(5.81)	- (-)	- (-)
40-49 歳	- (-)	- (-)	-2.00(-5.22)	-2.48(-3.69)
50-59 歳	- (-)	- (-)	-2.10(-5.94)	-1.04(-2.4)
60-69 歳	- (-)	- (-)	-2.64(-7.19)	-4.09(-3.88)
70 歳以上	- (-)	-1.54(-7.17)	-4.65(-10.66)	-4.11(-6.12)
住宅タイプ				
戸建持家	- (-)	1.41(4.65)	2.31(7.32)	- (-)
戸建賃貸	- (-)	1.07(1.83)	1.37(2.05)	- (-)
集合持家	- (-)	1.42(4.23)	1.26(3.06)	- (-)
集合賃貸	- (-)	0.73(2.41)	- (-)	-3.89(-3.71)
サンプル数: 1,000 対数尤度: -880.026 尤度比: 0.360				

表 - 2 所得回帰式推定結果

説明変数	パラメータ値 (t 値)
切片	375.92(15.47)
世帯人数	69.57(9.05)
単身ダミー	-57.64(-2.83)
世帯主ダミー	
40-49 歳	76.84(4.02)
50-59 歳	124.55(7.46)
70 歳以上	-43.44(-2.54)
Zone1 ダミー	44.91(2.19)
サンプル数: 1,000 標準誤差: 197.69 補正済み決定係数: 0.313	

表 - 3 自動車保有台数推定結果

観測値	推定値					的中率
	0台	1台	2台	3台以上	計	
0台	1,099	1,281	213	91	2,684	40.9%
1台	1,284	2,700	988	217	5,189	24.7%
2台	259	922	537	106	1,824	14.2%
3台以上	41	139	106	17	303	13.5%
計	2,683	5,042	1,844	431	10,000	43.5%

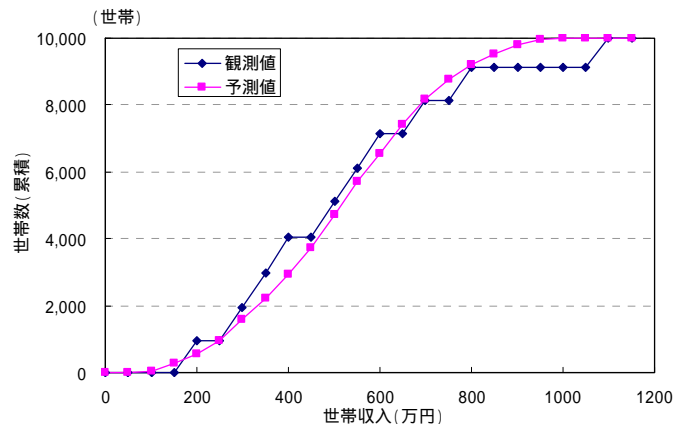


図 - 2 世帯所得推定結果

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.

- 6) 杉木直, 宮本和明, 大谷紀子, Varameth VICHENSAN: 居住属性を含む初期マイクロ世帯データの推定手法, 第41回土木計画学研究発表会論文集, CD-ROM, 2010.
- 7) Moeckel, R., Spiekermann, K., and Wegener, M.: Creating a Synthetic Population, Proceedings of CUPUM '03, Sendai, CD-ROM, 2003.