世帯及び個人属性分布を考慮したPT調査データの拡大係数算出手法の適用可能性*

Applicability of the Methodology to Match Distributions of Household and Person Attributes in the Generation of Synthetic Populations Based on Person-trip Survey Data*

萩尾龍彦**・倉内慎也***・石村龍則****

By Tatsuhiko HAGIO**・Shinya KURAUCHI***・Tatsunori ISHIMURA****

1．はじめに

我が国は，人口減少社会という大きな転換期に入り，加えて増加が著しい高齢者の交通事故対策やモビリティの確保が喫緊の課題となっている．それらの問題に対して，コンパクトシティや TOD 等の長期的かつ土地利用と一体となった政策が必要不可欠である．近年では，その実現性を検討するために，個人の行動に着目したミクロシミュレーションが適用されるようになり，より精緻な交通需要予測が可能となった．

しかし，ミクロシミュレーションを行う際には個人あるいは世帯レベルでの非集計情報が必要となる．我が国では国勢調査を含む全数調査が実施されているもののこれらの情報は開示されていない．したがって，需要予測には PT 調査で得られるサンプリングの非集計情報に，国勢調査等から得られる全数の集計情報から求められる拡大係数を寄与するというアプローチが頻繁に行われてきた．従来，世帯属性を考慮して拡大係数を算出する IPF 法が用いられていた．しかし，世帯属性である世帯構成人数や自動車保有台数は一致するものの，世帯内の個人を考慮していないため，個人属性である性別や年齢層が一致する保証はない．そこで，世帯属性，個人属性を両方考慮する IPU 法が開発されている．しかし，必ず厳密に各属性の制約に収束するとは限らないため，値によって信頼性にばらつきが生ずるという問題が残っている．

本研究では，以上の認識のもと，IPU 法をベースとし，PT データからより母集団の再現性が高い拡大係数を算定する新たな拡大係数算出手法を構築することを目的とする．本稿では，まず従来の手法である IPF 法，IPU 法の構造を示し，特に IPU 法の問題点を明確に把握することで，それに対する改善点を示していく．

*キーワーズ：拡大係数，非集計モデル，PT調査

**学生員，愛媛大学大学院理工学研究科

***正会員，博(工)，愛媛大学大学院理工学研究科

（愛媛県松山市文京町3,

E-mail: kurauchi@dpc.ehime-u.ac.jp）

****学生員，愛媛大学大学院理工学研究科

また，IPU 法で PT データを拡大し，項目ごとの変化の推移を分析し，誤差の原因を探り，世帯及び個人属性分布を考慮した拡大係数算出手法の適用可能性を検討する．

2．拡大係数算出手法の概要

本研究で構築するシステムの拡大係数算定手法のベースとなる IPU 法は，IPF 法を世帯レベルと個人レベルの両方の特性に合わせることができるように発展させたものであり，基本的な考え方は IPF 法と同様である．本章では，IPF 法と IPU 法の特徴を説明し，また 2 つの手法の関係性を示す．

（１）IPF（Iterative Proportional Fitting）法

個人属性１

個人属性２	1	2	3	合計
1	8	2	10	20
2	15	10	5	30
合計	23	12	15	50

標本調査の周辺分布

	属性	振り分け	合計
個人属性1	1	300	1000
	2	500	
	3	200	
個人属性2	1	700	1000
	2	300	

図１　分割表と周辺分布の例

IPF 法とは，標本抽出データから得られる非集計データでの複数の属性間のオッズ比を保持したままで，国勢調査等の全数調査から得られる各属性の周辺分布に一致するように，各セグメントの拡大係数を算出する手法である．ここで，複数の属性間のオッズ比とは，属性間の相関関係を表わすものであり，図１で示すように分割表が考慮する周辺分布が 2 次でそれぞれの属性の分類が 3，2 カテゴリーである場合，オッズ比は以下の式で表わされる．

$$\phi = \frac{p_{1,1}p_{2,2}}{p_{1,2}p_{2,1}} = \frac{0.16\times 0.2}{0.3\times 0.04} \quad \cdots\cdots(1)$$

ただし，$p_{i,j}$ は属性 1 のカテゴリーが i，属性 2 のカテゴリーが j の分類に属するセグメントのサンプルの割合を表す．周辺分布が m 次で次元 i のカテゴリー数が n_i の場合，属性 j のカテゴリー i_j，$i_j + c_1$，属性 k のカテゴリー i_k，$i_k + c_2$ のオッズ比は以下の式で

表わされる．

$$\phi = \frac{(p_{i_1\ldots,i_{j+c1},\ldots,i_{k+c2},\ldots,i_m})(p_{i_1\ldots,i_{j+c1},\ldots,i_{k+c2},\ldots,i_m})}{(p_{i_1\ldots,i_{j+c1},\ldots,i_k,\ldots,i_m})(p_{i_1\ldots,i_j,\ldots,i_{k+c2},\ldots,i_m})} \cdots\cdots\cdots\cdots\cdots(2)$$

ただし，c_1 は $i_j + c_1 \leq n_j$ を満たす正の整数を表わし，c_2 は $i_k + c_2 \leq n_k$ を持たす正の整数を表わす．

（２）IPU（Iterative Proportional Updating）法

セグメント番号	重み	個人属性1			個人属性2		重み1	重み2	重み3	重み4	重み5	重み（最終）
		1	2	3	1	2						
1	1	8	0	0	8	0	13.04	13.04	13.04	28.44	28.44	29.06
2	1	15	0	0	0	15	13.04	13.04	13.04	13.04	5.76	4.50
3	1	0	2	0	2	0	1.00	41.67	41.67	90.86	90.86	140.93
4	1	0	10	0	0	10	1.00	41.67	41.67	41.67	18.41	21.81
5	1	0	0	10	10	0	1.00	1.00	13.33	29.07	29.07	18.56
6	1	0	0	5	0	5	1.00	1.00	13.33	13.33	5.89	2.87
合計		24.00	14.00	18.00	21.00	32.00						
周辺分布		300.00	500.00	200.00	700.00	300.00						
δ_b		0.9233	0.9760	0.9250	0.9714	0.9000						
重み合計1		**300.00**	12.00	15.00	116.35	210.65						
重み合計2		300.00	**500.00**	15.00	197.68	617.32						
重み合計3		300.00	500.00	**200.00**	321.01	678.99						
重み合計4		423.19	598.38	357.41	**700.00**	678.99						
重み合計5		313.99	365.81	320.20	700.00	**300.00**						
δ_a		0.0466	0.2684	0.6010	0.0000	0.0000						
重み合計（最終）		300.00	500.00	200.00	700.00	300.00						

図２　IPU 法における IPF 法データの持たせ方

IPU 法とは，IPF 法で用いたデータの持たせ方に工夫を行い，拡大係数を算出する手法である．**図２**に示すように，各セグメントは個人属性に属するもの以外の要素は 0 として格納している．IPF 法では，分割表を作成し，複数の属性間のオッズ比を保持したまま，周辺分布に一致させていく．一方，IPU 法では，セグメントごとではなく世帯ごとにデータを持たせて周辺分布に一致させている相違点がある．しかし，IPU 法も同様に拡大係数算出後のオッズ比が保持されるため，以下の IPU 法のアルゴリズムと IPF 法は等価であると言える．

a）IPU 法による計算アルゴリズム

図２に示した簡単な例を用いて，IPU 法のアルゴリズムの手順を説明する．IPU 法では，厳密な解が保証されていないため，アルゴリズムの性能や適合度の特性を事前に検証しておく必要がある．IPU 法は初期値としてサンプルに含まれる全ての世帯の重みは等しいと仮定し，計算を開始する．このアルゴリズムは，全ての個人属性が一致するまで重みを調整しながら反復していく．まず初めに個人属性 1 のカテゴリー1 の制約に合わせるために，周辺分布を重み合計で割る．300/24=12.5 となり，制約を満たすためにその重みをカテゴリー1 の全てにかける．これで，カテゴリー1 の全てが 13.04 となった．同様に，カテゴリー2，3 は，500/12=41.67，200/15＝13.33 となる．重み 3 を更新し，この時点で個人属性 1 の重みは完全に一致した．次に，個人属性 2 の重みについても同様に行う．最初の個人属性２のカテゴリー1 は，個人属性 1 の調整が一致した後，制約に合わせる比率を個人属性 1 の重みに掛けて計算する．それは，700/321.01=2.18 となり，個人属性 2 のカテゴリー1 に値が含まれる全てのセグメントについて更新する．つまり，セグメント 2，4，6 については 0 であるため更新されず，重み 3 の値を用いる．同様に，重み 5 も更新する．以上で，1 回の計算過程が終了し，世帯，個人属性を調整した重みの合計と制約条件との誤差が縮小したことがわかる．この相対的な差の絶対値は，適合度として次式で表わされる．

$$\delta_j = \frac{|d_{i,j} w_i - c_j|}{c_j} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(3)$$

各々の繰り返し後，全体的な適合度の評価として用いられる．補正前のδの平均値は 0.9392，補正後は 0.1832 である．繰り返し間の増幅率は，次式で表わされ，0.7560 となる．

$$\Delta = |\delta_a - \delta_b| \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(4)$$

重みは Δ の値が 1×10^{-10} 等小さい値になるまで繰り返し調整する．約 20 回の繰り返し後，$\Delta = 8.15\times10^{-11}$ となり，1×10^{-10} 以下となる．

また，IPU 法独自の特徴として，図３に示すように世帯属性と個人属性を考慮した上で拡大係数を算出することも可能である．計算アルゴリズムは上記で示した例と同様である．

世帯番号	重み	世帯属性1	世帯属性2	個人属性1	個人属性2	重み1	重み2	重み3	重み4	重み（最終）
1	1	1	0	1	1	17.50	17.50	16.63	13.84	32.50
2	1	1	0	1	2	17.50	17.50	16.63	13.84	32.50
3	1	0	1	2	1	1.00	16.25	15.44	12.85	13.75
4	1	0	1	1	2	1.00	16.25	15.44	12.85	13.75
5	1	0	1	0	2	1.00	16.25	16.25	13.53	53.75
6	1	0	1	1	1	1.00	16.25	15.44	12.85	13.75
合計		2.00	4.00	6.00	9.00					
周辺分布		35.00	65.00	95.00	120.00					
δ_b		0.9429	0.9385	0.9368	0.9250					
重み合計1		***35.00***	4.00	39.00	58.50					
重み合計2		35.00	***65.00***	100.00	150.00					
重み合計3		33.25	62.56	***95.00***	144.13					
重み合計4		27.68	52.09	79.10	***120.00***					
δ_a		0.2090	0.1986	0.1674	0.0000					
重み合計（最終）		160.00	65.00	95.00	120.00					

図３　IPU 法アルゴリズムの計算例

３．IPU 法の適用と計算結果の考察

（１）データ概要

今回の IPU 法で用いるデータについて説明を行う．同時分布の世帯属性及び個人属性は，2007 年に松山都市圏で行われた松山パーソントリップ調査から得られたものから，5033 人のサンプルを用いる．概要について**表２**に示す．本研究で用いるゾーンは，都市計画で一般的に用いられている B ゾーンレベルである．まず，同時分布は世帯属性として世帯数，個人属性として性年齢階層を用いる．世帯数は，松山市を B ゾーンで分割した 25 カテゴリーである．性別は 2 カテゴリー，年齢は 0～4 歳の情報は PT 調査では得られ

ていないため，5〜9 歳から 5 歳区切りの 19 カテゴリーである．併せて，975 カテゴリーの情報がある．制約条件として利用する母集団の周辺分布は，2005 年の国勢調査データから得られたものを用いる．このデータ精度も同様に B ゾーンレベルである．しかし，自動車保有台数のみ B ゾーンレベルでの情報が得られていないため，市町村レベルで得られている松山市のデータを用いる．以上のデータから，拡大係数を IPU 法より算出する．

表１　松山パーソントリップ調査概要

調査名	松山パーソントリップ調査
調査主体	松山市，伊予市，東温市，松前町，砥部町
配布日時	2007年10月～12月
対象者	松山都市圏在住者からランダム抽出
回収サンプル数	29130名
総トリップ数	79389人･トリップ
主な調査項目	個人票，世帯票，居住に関する調査

（２）計算結果

データ概要で説明したように，周辺分布で用いるデータは自動車保有台数のみゾーンレベルが異なっている．そこで，拡大係数を算出する際に自動車保有台数を含むデータ，含まないデータの 2 パターンに分ける．2 パターンの PT データを IPU 法で拡大し，項目ごとの変化の推移を分析し，誤差の原因を探り，世帯及び個人属性分布を考慮した拡大係数算出手法の適用可能性を検討する．

a）各項目における誤差変化の推移

拡大後のゾーン別の世帯数，**性年齢階層の**推計精度結果を図４，５に示す．これを見ると，自動車保有台数を制約条件に入れると，自動車保有台数がない場合と比べて精度が低下することがわかる．それに比べ，自動車保有台数がない場合では，制約への精度が非常に高いといえる．

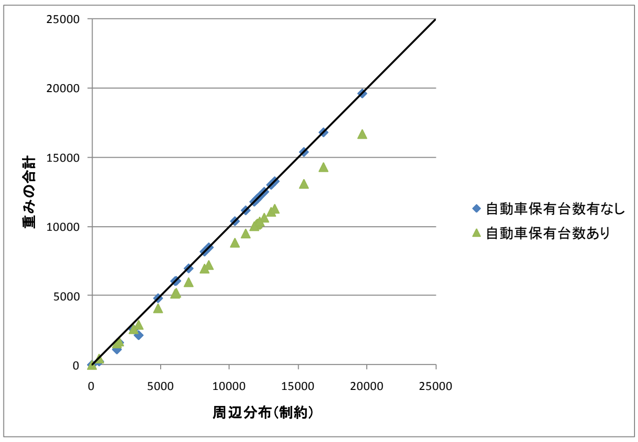

図４　拡大後のゾーン別世帯数の推計精度

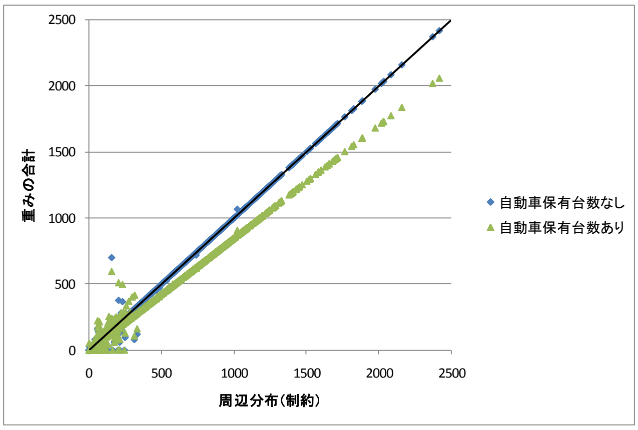

図５　拡大後のゾーン別性年齢階層の推計精度

（３）誤差原因の考察

IPU 法では，先ほどの計算結果で示したように，値によって必ず厳密に各属性の制約に収束するとは限らないという問題点がある．その原因のひとつとして，端点解に陥っていることが考えられる．

例えば，世帯属性が同じ 2 つの世帯があると仮定する．世帯 1 はどの個人にも特別な条件を持たせず，世帯 2 には個人属性カテゴリーに属する 1 人が存在する．世帯属性，個人属性の制約が 4，3 である場合，それぞれの制約を満たすための重みは，次式の同時一時方程式の解決により得ることができ，**図６**のような形で表わされる．

$$\left.\begin{array}{l} w_1 + w_2 = 4 \\ w_2 = 3 \end{array}\right\} \cdots\cdots\cdots\cdots\cdots (5)$$

w_1， w_2 ：世帯 1，2 の重み

交点 I は，方程式（4）の中の連立方程式及び点座標の解を示しており，与えられた世帯属性，個人属性の制約を満たす重みである．IPU 法アルゴリズムは，w_1， w_2 によって定義された第一象限のどの点でもスタートでき，アルゴリズムが反復した結果，交点 I に到達するメカニズムである．しかし，個人は実際多くの制約に対応するため，解が完全に点 I に一致しないことがある．**図７**に示すように，第 2 制約が $w_2 = 5$ になった場合は解が第４象限となり，アルゴリズムは解である交点 I に近づきはするが，完全に点 I に到達することはできない．アルゴリズムは，2 つの制約方程式が水平軸と交差する点 I_1， I_2 間を重みの座標が交互に移動し，値が定まらない．そのため，値により各属性の制約に収束しないことがあると考えられる．

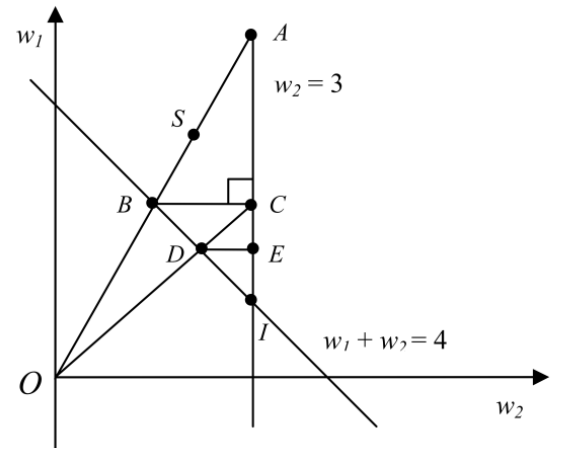

図６　実行可能な解の場合

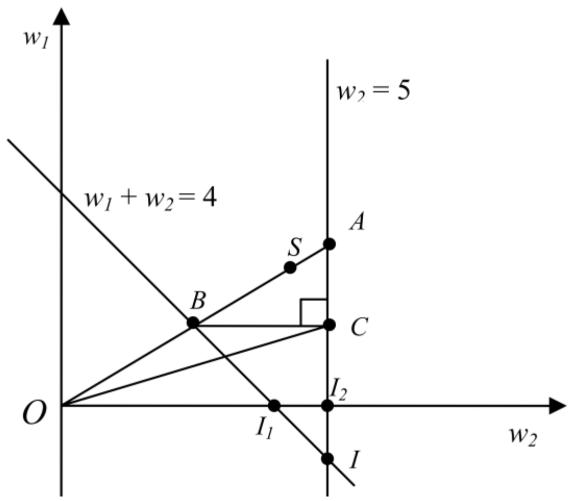

図７　実行不可能な解の場合

もうひとつの原因として，制約条件のゾーンが異なることが考えられる．今回制約条件で用いた自動車保有台数は，松山市等の市町村レベルの周辺分布しか得ることができない．そのため，IPU 法で，自動車保有台数の制約に重みを調整すると，B ゾーンレベルに比べ市町村レベルは自由度が高いため，自動車保有台数の制約は一致するものの年齢，性別，世帯数の制約が大きく離れてしまうと考えられる．

４．おわりに

本研究では，IPU 法をベースとし，PT データからより母集団の再現性が高い新たな拡大係数算出手法を構築することを目的とし，その適用可能性を分析した．本稿では，従来の手法である IPF 法，IPU 法の構造を示し，特に IPU 法の問題点を明確に把握した．その中で，IPU 法の問題として，制約条件として用いる周辺分布のゾーンレベルが異なると，精度が低下することがわかった．今後は，この問題の改善点を把握し，ゾーンレベルに関係なく精度の高い拡大係数算出手法を開発したいと考えている．そうすることで，個人の行動に着目したミクロシミュレーションが適用されるようになり，より精緻な交通需要予測が可能となるといえよう．

参考文献

1) 西田悟史，山本俊行，藤井聡，北村隆一：非集計交通需要分析のための将来世帯属性生成システムの構築，土木計画学研究・論文集，No. 17，2000.

2) Xin Ye, Karthik Konduri, Ram M.Pendyala : A methodology to match distributions of both household and person attributes in the generation of synthetic populations,