複雑な内生抽出法に基づく標本への離散選択モデルの適用

北村隆一[1]・酒井弘[2]・山本俊行[3]

[1]正会員　Ph. D.　京都大学教授　工学研究科土木システム工学専攻（〒606-8501 京都府京都市左京区吉田本町）
[2]正会員　社団法人システム科学研究所室長　調査研究部（〒600-8223 京都府京都市中京区新町通四条上ル小結棚町428　新町アイエスビル）
[3]正会員　博(工)　京都大学助手　工学研究科土木システム工学専攻（〒606-8501 京都府京都市左京区吉田本町）

本論文では内生抽出標本の重み付け理論を拡張し，加重層別標本抽出法や多次元選択肢別標本抽出法による標本への適用を論じる．これらの複雑な内生標本抽出法により得られた標本に適用可能な一意的な重みが存在することを示し，従来一般的になされてきた重み付け法の問題点を指摘するとともに，その改良を提案する．さらに，重み付きと重み無しで離散選択モデルを推定した場合の係数値と t-値を比較し，共分散行列を適切に推定することの重要性を示す．

Key Words : *weighting, endogenous sampling, roadside surveys, WESML, covariance estimator*

1. はじめに

交通行動の調査において複雑な抽出法が望まれる場合が多々ある．例えば本論文で例として用いた調査が対象としている観光行動，あるいは低密度地域での公共交通機関利用トリップのように，低頻度で発生する事象を対象とする場合，通常の無作為抽出による調査は極めて効率性の低いものとなる．このような場合，観光地や鉄道駅のように対象とする事象を直接観測できる場所で標本の抽出を行うことによりより効率的な調査が可能となると考えられる[1]．

調査において，研究の対象とする事象 (あるいはそれにより決定される事象) に基づいて標本が抽出される場合，これを内生標本抽出 (endogenous sampling) と呼ぶ．典型的な例は選択肢別抽出 (choice-based sampling) で，路側で抽出された自動車利用者と，駅や車内で抽出された公共交通機関利用者を組み合わせることにより交通手段選択解析のための標本を準備する場合などがこれにあたる．また，交通行動調査でしばしば用いられる加重層別標本 (enriched sample) は，無作為抽出による標本を選択肢別抽出による標本で補完したもので，これも内生標本抽出によるものである．

無作為抽出や，外生的に定義された層を用いた層別抽出の場合と異なり，内生抽出法により得られた標本の解析にあたっては，標本平均や標本分布などを用いた単純集計による母数の推定の場合のみならず，モデル推定の際にも重み付けによるひずみ修正が必要となる[1,2]．選択肢別抽出の場合，重み付けおよび離散選択モデルの推定の方法論はこれまでに開発，適用されてきた[1,3,4]．本稿では重み付け方法論をより一般化し，より複雑な内生標本抽出法を用いた標本への適用を試みる．また，路側調査や乗客調査などの内生標本抽出が複数箇所で（あるいは複数時点で）行われる場合，各調査地点での抽出率に基づいて単純に重みを算出するという従来の方法の問題点を指摘し，本研究で得られた重み付け法の適用を提案する．

次章ではこれまでの重み付け方法論を多次元選択肢別標本抽出へと一般化する．この結果を用い，第3章では外生的層別抽出による標本と内生的抽出から得られた標本から成る加重層別標本抽出について，一元的な重み付け法を提案する．第4章では複雑な多次元選択肢別標本抽出を用いた観光行動調査の標本を例にとり，重みの算定を具体的に示す．この重みを用いた離散選択モデルの推定例を第5章に示し，モデルの係数の推定値と，共分散行列の推定値から得られる t-値について，各々ひずみを検討する．結論を第6章に示す．

2. 複雑な選択肢別抽出法に基づく標本の重みづけ

本章では従来の研究[3,4,5,6,7]に基づき，選択肢別抽出により得られた標本の重みづけについて論じ，それを多

次元選択肢別抽出の場合へと拡張する．互いに排反な M ($<\infty$) 個の離散選択肢からなる選択肢集合を C，説明変数の標本空間を X と表し，対象とする母集団が空間 $C\times X$ に含まれるとする．また，$j\in C$，$x\in X$ とし，この母集団から得られる標本を (j,x) で表し，それが抽出される同時確率密度を

$$g(j,x\,|\,\theta)=q(x)\Pr[j\,|\,x,\theta] \tag{1}$$

と表す．ここに

$g(j,x\,|\,\theta)$ ＝ 標本 (j,x) の同時確率密度関数，
$q(x)$ ＝ 説明変数 x の同時周辺密度関数
＝ $\sum_{j\in C} g(j,x|\theta)$，
θ ＝ 選択 j と説明変数 x の関係を規定する母集団パラメータのベクトル

である．この定式化を用いると，純粋な無作為抽出が用いられた場合の標本 (j,x) の確率密度，$L_r(j,x)$，は

$$L_r(j,x)=g(j,x\,|\,\theta) \tag{2}$$

と表される．

選択肢別抽出を行うにあたり，選択肢集合，C，が B 個の，必ずしも排反ではない部分集合，$C_b, b=1,2,\ldots,B$, に分割され，各々の部分集合を層とみなし標本が抽出されたとする．各層毎の母集団を $A_b=C_b\times X$ と示す．また $C_1\cup C_2\cup\cdots\cup C_B=C$ である．この場合，C_bの母集団での比率，$Q(b|\theta)$，は

$$Q(b|\theta)=\sum_{j\in C_b}\int_{x\in X} g(j,x\,|\,\theta)dx=\sum_{j\in C_b}\int_{x\in X} q(x)\Pr[j\,|\,x,\theta]dx$$
$$=\sum_{j\in C_b} q(j\,|\,\theta) \tag{3}$$

と与えられる．ここに

$q(j\,|\,\theta)$ ＝ 選択肢 j が選ばれる周辺確率

である．$Q(b|\theta)$ は C_b に属する選択肢の市場でのシェアに対応することに注目されたい．

始めに，j が 1 つの層のみに含まれる場合を考える．選択肢別抽出の場合に層 b が選出され標本 (j,x) が抽出される同時確率密度，$L_c(j,x,b)$，は

$$L_c(j,x,b)=H(b)\Pr[(j,x)\in A_b\,|\,b,\theta]=\frac{H(b)g(j,x\,|\,\theta)}{Q(b\,|\,\theta)} \tag{4}$$

と与えられる．ここに

$H(b)$ ＝ 選択肢別抽出において層 b から標本が抽出される確率

である．ここで

$$L_r(j,x)=L_c(j,x,b)\frac{Q(b|\theta)}{H(b)} \tag{5}$$

であることに着目すると，$[H(b)/Q(b|\theta)]^{-1}$ を選択肢別抽出法により得られた標本の抽出確率密度に適用することにより，その抽出確率密度は無作為抽出の場合と同一のものとなることが分かる．すなわち，$[H(b)/Q(b|\theta)]^{-1}$ は選択肢別抽出法により得られた標本の重みとして適用可能である．

次にこの結果を j が複数の層に含まれる場合に一般化する．後述のように，特定の個体が複数の層から標本として選出される確率は微小であり，近似的に無視することができるとすれば，

$$L_c(j,x)=\sum_{b\in z_j} L_c(j,x,b)=\sum_{b\in z_j}\frac{H(b)g(j,x\,|\,\theta)}{Q(b\,|\,\theta)}$$
$$=g(j,x\,|\,\theta)\sum_{b\in z_j}\frac{H(b)}{Q(b\,|\,\theta)} \tag{6}$$

が成立する．ここで Z_j は選択肢 j を含む層の集合である．式 (6) と式 (2) の比較から，選択肢 j を含む標本の重みは

$$\omega(j)=\left[\sum_{b\in z_j}\frac{H(b)}{Q(b\,|\,\theta)}\right]^{-1} \tag{7}$$

と得られる．

本研究では，上記の重み付け理論を拡張し，多次元選択肢別抽出法が用いられた場合の標本の重み付け法を提案する．観測される行動が D 次元の選択行動であるとし，標本を $(j_1,j_2,\ldots,j_D;x)=(\boldsymbol{j},x)$ と表そう．ここに $j_d\in C_d$, $d=1,2,\ldots,D$, C_d は d-次元目の選択肢集合，x はこれまでのように説明変数のベクトルである．さらに各々の選択肢集合が B_d に分割されており，$C_d=C_1^d\cup C_2^d\cup\cdots\cup C_{B_d}^d$ が成立するとする．また b_{j_d} により j_d が属する層を表し，$\boldsymbol{b}=(b_{j_1},b_{j_2},\ldots,b_{j_D})$ とする．各次元での各々の層において選択肢別標本が抽出されるとする．

ここで注意されたいのは，$\boldsymbol{j}$ に含まれている選択肢は

必ずしも調査の結果得られる標本を用いて推定される離散選択モデルの選択肢と一致するわけではないという点である（例えば推定されるモデルは (j_1, j_2) のみを対象とするものであるかもしれない）．しかしながら，多次元選択肢別標本抽出が内生的標本抽出と見なされる場合，適切な重みを用いずにモデルが推定されたときに歪みが生じることに変わりはない．

さて，現実の標本抽出においては母集団が標本数に比べ格段に大きく，抽出率は極めて小さいことが通常である．したがって特定の個体が標本として複数次元において抽出される確率は極めて小さいと考えられる．そこで，j を持つ個体が抽出される確率，$p(j)$，は，その個体が層 b_{j_d} で抽出されるという事象を B_{j_d} と表すと

$$
\begin{aligned}
p(j) &= \Pr[B_{j_1} \cup B_{j_2} \cup \cdots \cup B_{j_D}] \\
&= \Pr\left[\sum_{d=1}^{D} B_{j_d} - \sum_{j_e<j_f} \left(B_{j_e} \cap B_{j_f}\right) + \cdots \right. \\
&\quad \left. + (-1)^{D-1} \left(B_{j_1} \cap B_{j_2} \cap \cdots \cap B_{j_D}\right)\right] \\
&\cong \Pr[B_{j_1}] + \Pr[B_{j_2}] + \cdots + \Pr[B_{j_D}]
\end{aligned} \tag{8}
$$

により十分に近似されると考えられる．簡単な例として，4 抽出次元が存在し，ある個体の抽出確率は

$$
\Pr[B_{j_1}] = \Pr[B_{j_2}] = \cdots = \Pr[B_{j_4}] = 0.02
$$

また次元毎に抽出は独立で，

$$
\begin{aligned}
&\Pr[B_{j_k} \cap B_{j_{k'}}] = \Pr[B_{j_k}]\Pr[B_{j_{k'}}], \\
&\Pr[B_{j_k} \cap B_{j_{k'}} \cap B_{j_{k''}}] = \Pr[B_{j_k}]\Pr[B_{j_{k'}}]\Pr[B_{j_{k''}}], \\
&\Pr[B_{j_1} \cap B_{j_2} \cap B_{j_3} \cap B_{j_4}] = \Pr[B_{j_1}]\Pr[B_{j_2}]\cdots\Pr[B_{j_4}] \\
&\qquad\qquad k \neq k',\ k \neq k'',\ k' \neq k''
\end{aligned}
$$

が成立するとしよう．すると

$$
\begin{aligned}
p(j) &= 4 \times 0.02 - 6 \times (0.02)^2 + 4 \times (0.02)^3 - (0.02)^4 \\
&= 0.0776 \cong 0.08
\end{aligned}
$$

と，式 (8) の近似が成立している．

この近似を用い式 (6) と同様，

$$
L_c(j, x) = \sum_{d} \sum_{b \in z_{j_d}} L_c(j, x, b) = g(j, x \mid \theta) \sum_{d} \sum_{b \in z_{j_d}} \frac{H^d(b)}{Q^d(b \mid \theta)} \tag{9}
$$

が得られる．ここに $H^d(b)$ と $Q^d(b|\theta)$ は第 d-次元について定義された $H(b)$ と $Q(b|\theta)$ である．選択肢ベクトル j を含む標本の重みは

$$
\omega(j) = \left[\sum_{d} \sum_{b \in z_{j_d}} \frac{H^d(b)}{Q^d(b \mid \theta)}\right]^{-1} \tag{10}
$$

と表される．

式 (10) の適用には $H^d(b)$ および $Q^d(b|\theta)$ の値が必要となる．前者が標本内での層の構成比により，また後者がマーケットにおける層の構成比により表されると考え，各々の推定値を

$$
\hat{H}^d(b) = \frac{\sum_{j_d \in C_b^d} n_{j_d}}{n}, \quad \hat{Q}^d(b \mid \theta) = \frac{\sum_{j_d \in C_b^d} \hat{N}_{j_d}}{\hat{N}} \tag{11}
$$

と表す．ここに

n_{j_d} = 標本内で d-次元目の選択肢別抽出で得られた標本のうち，選択肢 j_d を持つ標本の頻度，
$\hat{N}_{j_d}$ = 母集団内で次元 d の選択肢 j_d を持つ個体の頻度の推定値，
n = 総標本数，
$\hat{N}$ = 母集団内の総個体数の推定値

である．すると選択肢ベクトル j を持つ標本の重みは

$$
\hat{\omega}(j) = \left[\frac{\hat{N}}{n} \sum_{d} \sum_{b \in z_{j_d}} \frac{\sum_{j_d \in C_b^d} n_{j_d}}{\sum_{j_d \in C_b^d} \hat{N}_{j_d}}\right]^{-1} = K\left[\sum_{d} \sum_{b \in z_{j_d}} \frac{\sum_{j_d \in C_b^d} n_{j_d}}{\sum_{j_d \in C_b^d} \hat{N}_{j_d}}\right]^{-1} \tag{12}
$$

と表される．ここに K は定数で，その値を調整することにより標本の拡大率を定めることができる．式 (12) では，拡大率は抽出された次元に依存せず，各々の標本に対して一意的に拡大率が決定されることに着目されたい．

3. 路側調査と家庭訪問調査に基づく標本への適用

上記の重み付け方法の適用例として，新規高速道路が都市圏住民の一日の交通パターンに与えるインパクトを把握することを目的とした，圏域における一日の総トリップを母集団とした調査を考えよう．この調査の抽出単

位は個人で，標本は新規高速道路と競合する路線での路側調査で車種別に抽出された個人（運転者）と，住民登録台帳に基づき地域別に層化抽出された個人から成るとする．調査の目的が一日の交通パターンの推定であり，道路網上での各トリップの通過点は個人の交通パターンにより規定されるから，路側調査での標本抽出は内生的であると考えるのが妥当である．結果として得られる標本は，路側調査による内生標本抽出と，住民台帳に基づく層別無作為抽出が組み合わされたもので，加重層別標本と見なすことができる．なお，住民台帳に基づき抽出された個人が，路側調査で再抽出される，あるいは同一の個人が路側調査で複数回抽出される可能性はあるものとし，個人の抽出確率は式 (8) のように近似されるものとする．

路側調査が M 個所で行われたとし，ある標本個人が i 番目の調査地点を通過した回数に等しい値を採る変数 $m_i (= 0, 1, 2, ...)$ を定義し，その個人が一日に路側調査地点の集合を通過した頻度を $(m_1, m_2, \ldots, m_M)$ と表す．変数 m_i の値は個人が調査地点 i を通過した頻度のみにより定義され，地点 i でその個人が抽出されたか否かによらないことに注意されたい．また，車種が総計 V あるとし，標本個人の利用車輌の車種を v で，居住地域が L あるとし，標本個人の居住地域を l で表す．標本個人の路側調査地点通過頻度と居住地域を離散選択の結果と見なし，$j = (m_1, m_2, \ldots, m_M, v, l)$ とすると

$$\hat{\omega}(j) = K\left[\sum_{i=1}^{M} m_i \frac{n_i^v}{N_i^v} + \frac{w_l}{W_l}\right]^{-1} \tag{13}$$

が得られる．ここに

n_i^v = 標本内で地点 i での路側調査で抽出された車種 v の車輌台数,
N_i^v = 一日に地点 i を通過する車種 v の車輌の観測台数,
w_l = 標本内で住民台帳により抽出された地域 l に居住する個人の総数,
W_l = 地域 l の住民台帳に登録された個人総数

である．ここで定義された重みは，路側調査で抽出されたか住民台帳で抽出されたかにかかわらず，各々の標本個人に一意的に適用される．

以上の解析は住民台帳で抽出された個人を路側調査で抽出されることを妨げないという前提に立っている．住民台帳で抽出された個人は路側調査では抽出しないよう調査がなされた場合，特定の j を持つ個人が抽出される確率は，その個人が住民台帳から抽出される確率と，住民台帳からは抽出されず路側調査で抽出される確率の和である．j が与えられたとして，後者は路側調査の抽出率で近似される．すなわち，住民台帳で抽出された標本については m_i をすべて 0 とした上で式 (13) を適用し，路側調査で抽出された標本には式 (13) をそのまま適用することにより重み付けがなされる．

この章での解析は住民台帳から抽出された個人と路側調査で抽出された個人の双方に一意的な重みを適用することにより適切な重み付けが可能となることを示している．すなわち，式 (13) を適用し

① 路側調査で得られた標本も，住民台帳から得られた標本と同様，住民台帳からの抽出確率に基づき重み付けをおこなうこと，また
② 全路側調査地点についての通過情報に基づき得られる路側調査での抽出確率を，路側調査および住民台帳からの標本に適用すること

により適切な重み付けが可能となる．住民登録台帳から抽出された個人は路側調査で抽出しない場合でも，②に示されるように全路側調査地点についての通過情報に基づき得られる路側調査での抽出確率を算定し，路側調査から得られた標本の重み付けを行うことが必要となることに着目されたい．

選択肢別標本抽出法や加重層別標本抽出法は交通機関選択のように単独のトリップを対象とした場合に適用される場合が多い．このような場合でも，路側調査や乗客調査などの選択肢別抽出が複数地点で（あるいは複数交通機関について）行われる場合，適切な重み付けのためには全調査地点の通過情報（あるいは全調査対象交通機関の利用に関する情報）が各標本について必要となるのが一般である．したがって調査票の設計にあたり，これらの重み付けに必要となる情報に関する設問を設けることが肝要である．

現存の調査結果の重み付けに当たって問題となるのはこのような情報が必ずしも存在しないということであろう．例えば，路側調査から得られた標本については他の調査地点の通過についての情報は一切得られていないのが一般である．したがって，これらの標本について $(m_1, m_2, \ldots, m_M)$ を推定することは不可能である．一つの対応策として，住民台帳等から抽出された標本から j の分布が推定されるとすれば，それに基づき路側調査地点 i で抽出された標本について，その属性とそれが地点 i を通過したことを所与として，j の未知要素を impute し，$\hat{\omega}(j)$ を算定するという方法が考えられる．このような方法は重みの決定に当たり複雑な推定を必要とするものであるが，既存の調査結果を用い上記の①と②に対応しよ

表—1　調査票配布地別配布数，回収数及び回収率

配布地区分	配布個所数	配布数	回収数	回収率
鉄道駅	9	5,500	867	15.7%
高速インター	2	5,000	809	16.2%
一日乗車券売場	3	1,000	218	21.8%
観光地	23	11,640	2,983	25.6%
宿泊施設	66	3,548	815	23.0%
合計	103	26,688	5,692	21.3%

表—2　鉄道駅での標本抽出状況

鉄道駅	乗降客数推計値	回収調査票数	推定抽出率
A	222,990	424	0.190%
B	6,100	76	1.246%
C	32,760	78	0.238%
D	61,250	132	0.216%
E	13,560	157	1.158%
合計	336,660	867	0.258%

表—3　高速道路出路での標本抽出状況

インター	祝祭日平均全車種台数	推定乗用車台数	回収調査票数	推定抽出率
京都東	32,120	24,830	321	1.293%
京都南	54,950	42,360	488	1.152%
合計	87,070	67,190	809	1.204%

表—4　宿泊施設での標本抽出状況

宿泊施設	部屋数	推定宿泊者数	回収調査票数	推定抽出率
ホテル	6,576	13,152	507	3.85%
公的宿泊施設	808	1,616	144	8.91%
旅館	1,067	2,134	113	5.30%
ペンション	240	480	50	10.42%
合計†	8,691	17,382	814	4.68%

† 総回収調査票数は、宿泊施設に配布していることは確認されているが、宿泊施設が不明になっているデータが1件あるため、814件となっている.

うとする場合，重み付けの過程が複雑化することは避けえないといえよう.

4. 京都市観光調査への適用

京都市への観光客の交通行動を調査する目的で，秋の観光シーズンのピークにあたる1997年11月3日に，アンケート調査，歩行者・自動車交通量調査等を含む大規模調査が京都市により実施された．アンケート調査は一日の観光交通行動に関する情報を収集することを目的とし，地図を媒体として用い，全ての目的地および滞在時間の記入を回答者に促すものである．また基本的な個人属性に加え，当日の来洛体験についての態度に関する質問等も含まれている．総計26,688部の調査票を配布，5,692が郵送回収されている．返答率は21.3%と高いものとなったが，これは調査票が色刷りで見やすいものであった，抽選で回答者500名に京舞妓の写真入りの「オリジナル・テレフォンカード」を贈呈した，また調査が対象とする観光行動が回顧するに楽しいものである，等の理由によると考えられる.

アンケート調査は複雑な内生抽出法を用いている．調査票は

- 入洛地点（鉄道駅，高速道路出口）
- 観光集客地（社寺仏閣，その他観光地）
- 宿泊施設（ホテル，旅館，公的宿泊施設等）

で手渡し配布された[1]. 表—1に調査票配布個所別に配布数と回収票数を示す．この表から明らかなように訪洛者の抽出確率は，用いられた入洛交通手段，来訪地，市内での宿泊の有無により大きく変化する．標本の適切な重み付けが必要なことは明らかであろう．以下に重み算定の手順を京都市の調査について簡単に記す.

入洛地点: 調査において調査票が配布されたのは鉄道駅，名神高速道路インターチェンジ出路，市バス・地下鉄一日乗車券売場の計14個所である．鉄道乗降客に対して5駅，9個所で調査票が配布された．これらについて結果を表—2に示す．休日の乗降客数に関する情報はいくつかの駅について入手できなかったため，類似の駅での休日－平日比を用いるなどして推定したものを含む.

調査では名神高速道路京都東および京都南インターから入洛する車輌のうち乗用車のみを調査票配布対象としている．乗用車の台数は，道路公団大阪管理局管内交通量統計平成6年分[8]から得られた各インター出路の祝祭日平均利用台数（全車種）に，平成6年度道路交通センサスに基づく京都東，京都南インターそれぞれの両側区間での平均乗用車比率を乗じることにより推定している．推計車輌台数，回収調査票数，推定抽出率を表—3に示す.

バス・地下鉄一日乗車券発売所への一日の来訪人数については情報が全く無く，一年間の一日乗車券発売数を基に調査日の来訪数を大雑把に推定せざるをえなかった.

観光地: 調査票が配布された23個所の観光集客地のうち4個所について事前調査から入り込み客数が推定されている．京都市産業観光局の京都市観光調査年報[9]に示されている各観光地の来訪比率を基に，今回の休日調査箇所と一致しているところはそのままの来訪比率，一致していない箇所は一致している箇所の最低値（ここでは「東寺」の3.3%）を採用し，各々の観光地への入り込み数を推定した.

宿泊施設: 66宿泊施設での調査票回収結果を表—4に示す．推定宿泊者数は一部屋当たり二名を想定してい

表—5　重み算出の数値例に用いた標本

標本	入洛地点	来訪観光集客地	宿泊施設
1	A駅	清水寺，金閣寺，銀閣寺	ホテル
2	京都南ｲﾝﾀｰ	嵐山観月橋	(日帰り)

表—6　各次元での推定抽出率

標本	入洛地点	来訪観光集客地	宿泊施設
1	0.0019	0.0111，0.0248，0.0181	0.0385
2	0.0115	0.0160	0.0

る. 非常に限定された場所で調査票が配布されたためか，抽出率はきわめて高いものとなっている.

以上，重み設定の過程を簡単に記したが，$\hat{N}_{j_d}$ の推定には数々の困難が伴い，一日乗車券売り場の例に示されるように，ごく大雑把な推定値しか得られなかった場合もある. 今後，複雑な内生標本抽出法を用いた調査の設計にあたっては，重み設定に必要となる情報の入手可能性を事前に検討することが重要であろう. なお，森地・屋井[1]は離散選択モデルの定数項を除く係数のひずみのない推定値を得るにあたり，$Q^d(b|\theta)$ には高い精度が要求されないこと，また定数項については $Q^d(b|\theta)$ の真値が判明した時点で事後的な修正が可能であることを報告していることを附記する.

数値例:　以上に示した重み付け手法適用の数値例として表—5 に示す 2 標本の重みを算定する. 標本 1 は鉄道を用い A 駅より入洛, 京都に宿泊し,調査日には清水寺を始めとする 3 集客地を訪れている. 標本 2 は車を用いた日帰り旅行で，嵐山地域のみを訪れている. これら標本の，各次元での推定抽出率を表—6 に示す. 本研究で提案する重み付け法を適用した場合，標本 i の重み $\hat{\omega}(j_i)$，$i=1,2$，は以下のように算定される.

$$\hat{\omega}(j_1)=K[0.0019+0.0111+0.0248+0.0181+0.0385]^{-1}$$
$$=10.6K$$
$$\hat{\omega}(j_2)=K[0.0115+0.0160+0.0]^{-1}=36.4K$$

従来用いられてきた単純な重み付け法では，各ケースの重みはそのケースが抽出された次元の推定抽出率の逆数により決定されてきた. 例えば，標本 1 が A 駅で抽出されたとすれば重みは $K/0.0019=526.3K$，清水寺で抽出されたとすれば $K/0.0111=90.1K$，また宿泊地で抽出されたとすれば $K/0.0385=26.0K$ という重みが算定され，同じ j を持つ標本であっても，抽出された次元によって異なる重みが与えられるという非合理が生じる. 本論文で展開した重み付け方法論によれば，抽出された次元に拘わらず，標本抽出の全次元にわたる標本の属性，j，に基づき重みが $10.59K$ と一意的に算定される.

表—7 に京都までの利用交通手段と京都市内での移動に用いられた代表交通手段を，重み無し，従来用いられてきた方法による重み付き，本研究で提案する方法による重み付きでクロス集計したものを示す. 複数の手段を利用している場合には，鉄道，タクシー，自家用車（レンタカー含む），バスの優先順位で代表交通手段を決定した. ここでは重み付け後に全体の頻度が 5,692 に保たれるよう定数 K を調整している. 入洛地点での標本抽出は京都までの利用交通手段に基づく選択肢別抽出であるため, 当然の事ながら利用交通手段の分布は重みの適用, 不適用, 重みの種類によって著しく異なったものとなる. 京都までの交通手段と京都市内での代表交通手段の双方について，いずれの重み付けによっても自家用車の頻度が低下しているものの，従来法の方が低下が著しい. ま

表—7　入洛交通手段と市内交通手段の分布

重み無し

京都までの交通手段	京都市内での代表交通手段				
	鉄道	バス	タクシー	自家用車	合計
鉄道	2,564 (45.0)	327 (5.7)	940 (16.5)	131 (2.3)	3,962 (69.6)
バス(その他手段を含む)	60 (1.1)	329 (5.8)	2 (-)	10 (0.2)	401 (7.0)
タクシー	2 (-)	7 (0.1)	29 (0.5)	2 (-)	40 (0.7)
自家用車	33 (0.6)	118 (2.1)	24 (0.4)	1,114 (19.6)	1,289 (22.6)
合計	2,659 (46.7)	781 (13.7)	995 (17.5)	1,257 (22.1)	5,692 (100.0)

従来法

京都までの交通手段	京都市内での代表交通手段				
	鉄道	バス	タクシー	自家用車	合計
鉄道	2,761 (48.5)	408 (7.2)	1,060 (18.6)	125 (2.2)	4,354 (76.5)
バス(その他手段を含む)	38 (0.7)	380 (6.7)	1 (-)	5 (0.1)	424 (7.4)
タクシー	1 (-)	2 (-)	32 (0.6)	1 (-)	36 (0.6)
自家用車	23 (0.4)	91 (1.6)	16 (0.3)	750 (13.2)	879 (15.4)
合計	2,822 (49.6)	881 (15.5)	1,109 (19.5)	880 (15.5)	5,692 (100.0)

提案法

京都までの交通手段	京都市内での代表交通手段				
	鉄道	バス	タクシー	自家用車	合計
鉄道	2,140 (37.6)	801 (14.1)	562 (9.9)	95 (1.7)	3,599 (63.2)
バス(その他手段を含む)	37 (0.6)	808 (14.2)	1 (-)	5 (0.1)	851 (14.9)
タクシー	1 (-)	4 (0.1)	39 (0.7)	- (-)	45 (0.8)
自家用車	30 (0.5)	195 (3.4)	10 (0.2)	964 (16.9)	1198 (21.0)
合計	2,208 (38.8)	1,808 (30.8)	612 (10.8)	1,064 (18.7)	5,692 (100.0)

( ): 総合計のパーセント. 四捨五入のため合計は一致しない.
-: 0.05 以下

た，従来法による重み付けにより鉄道の頻度が上昇している一方，本研究で提案する重み付けでは鉄道の頻度が低下，さらに，バス（その他手段を含む）の頻度が増加している．京都市内での代表交通手段については，従来法では自家用車の頻度が低下している一方，本研究で提案する方法ではタクシーの頻度が低下していることが見て取れる．本研究で提案する重み付けにより，鉄道で京都に到着しバス等で市内を移動した人の頻度が327から801に，またバス等で到着しバス等で市内を移動した人の頻度が329から808へと大幅に増加している．これは入洛地点での選択肢別抽出で，京都市内移動の主要交通手段であるバスの利用者が抽出されていないことが主要因と考えられる．しかしながら，従来の重み付けではバス等で市内を移動した人の頻度は重み無しの場合からほとんど変化していない．従来法ではバス利用者に対して適切な重み付けが行われていないと考えられ，本研究で提案する方法の優位性を示すものと言えよう．

5. 内生的重みを用いた離散選択モデルとパラメータ共分散行列の推定

本節では本研究で提案する重み付け法の離散選択モデル推定への適応例として，上記の京都市でのデータと重みを用い来訪交通手段選択モデルを WESML (weighted exogenous sample maximum likelihood)[2), 10)] 法により推定する．ここで解析の目的は重み付けがモデルパラメータの推定値にどう影響するかを検討することに加え，パラメータの共分散行列の推定法が，得られる t-値にどう影響を及ぼすかを定めることにある．したがってモデル自体は入洛時の自動車の利用確率を表す簡単な二項ロジットモデルである．選択肢別抽出標本を用いた推定法の比較検討については森地ら[11)] を参照されたい[12]．

WESML 法の対数尤度関数を

$$L=\sum_{i=1}^{N}\hat{\omega}(j_i)\ln \Pr[J_i|x_i,\theta] \tag{14}$$

とする．ここで i は標本を指し，N は総標本数，J_i は標本 i について観測された選択肢である．内生抽出標本を用い WESML 法を適用して離散選択モデルを推定した場合，共分散行列は

$$\Sigma=\Omega^{-1}\Lambda\Omega^{-1} \tag{15}$$

と与えられる[4)]．ここに

$$\Omega=E\left\{\left[\frac{\partial \ln \Pr(J_i|x_i,\theta)}{\partial\theta}\right]\left[\frac{\partial \ln \Pr(J_i|x_i,\theta)}{\partial\theta'}\right]\right\} \tag{16}$$

$$\Lambda=E\left\{\hat{\omega}(j_i)\left[\frac{\partial \ln \Pr(J_i|x_i,\theta)}{\partial\theta}\right]\left[\frac{\partial \ln \Pr(J_i|x_i,\theta)}{\partial\theta'}\right]\right\} \tag{17}$$

である．推定にあたっては，

$$\hat{\Omega}=\frac{1}{N}\sum_{i=1}^{N}\left\{\left[\frac{\partial \ln \Pr(J_i\mid x_i,\theta)}{\partial\theta}\right]\left[\frac{\partial \ln \Pr(J_i\mid x_i,\theta)}{\partial\theta'}\right]\right\} \tag{16'}$$

$$\hat{\Lambda}=\frac{1}{N}\sum_{i=1}^{N}\left\{\hat{\omega}(j_i)\left[\frac{\partial \ln \Pr(J_i\mid x_i,\theta)}{\partial\theta}\right]\left[\frac{\partial \ln \Pr(J_i\mid x_i,\theta)}{\partial\theta'}\right]\right\} \tag{17'}$$

を適用する．この推定量の過去の適用例には山本ら[19)] がある．

周知のようにモデル係数と t-値の推定値は重みを適用する場合としない場合で大きく異なったものとなる．表-8 に示されるように，重み無しの推定結果と本研究で提案する重み付けによる推定結果を比べると，交通手段別費用の係数の差異が特に顕著で，自動車費用の係数が重み無しの推定値の -0.42 に対して重み付き推定値はその約 4.1 倍の -1.71 である．鉄道費用の係数の推定値も重み無しの 0.49 に対し重み付きが 1.08 と 2 倍を超える値をとっている．同様の差異が t-値の推定値についても見られる．従来法による重み付けでは，鉄道費用の係数の推定値は重み無しに対して約 1.6 倍となり，本研究で提案する方法と同様の結果が得られているものの，自動車費用の係数については重み無しの場合とほとんど変化せず，本研究で提案する方法による結果とは大きく異なる．

表-8 の t^* 及び t^{**} の各列は，式 (15) に基づいて推定された共分散行列から得られた t-値と，重みが外生的であるとして $\hat{\Sigma}=\hat{\Lambda}^{-1}$ に基づいて推定された共分散行列から得られた t-値を比較したものである．式 (15) に基づく値を真値と考えた場合，従来法，本研究で提案する方法のいずれもほとんどの変数について t-値の誤差は ±20% 以内であるが，本研究で提案する重み付けによる推定結果の「初訪洛 (D)」の t-値のように約 40% の誤差を見せるものもある．ここでの事例では生じていないものの，$\hat{\Sigma}=\hat{\Lambda}^{-1}$ を用いた場合に有意性が誤って判断されモデルの同定に支障をきたすことは十分に考えられる．

6. 結論

本論文では内生標本抽出の重み付け理論を拡張し，加重層別標本抽出法への適用に触れ，多次元選択肢別標本

表—8 WESMLによる自動車利用離散選択モデルの推定†

係数およびt-値の比較 説明変数	重み無し β	t	従来法 β	t^*	t^{**}	提案法 β	t^*	t^{**}
男性 (D)	0.64	7.59	0.60	6.23	6.25	0.56	6.32	6.38
50歳以上 (D)	-0.61	-7.47	-0.90	-9.10	-9.64	-0.84	-9.15	-10.06
就業者 (D)	0.21	2.17	0.51	4.01	4.52	0.43	3.78	4.41
毎日車を運転 (D)	0.92	11.76	0.97	11.23	11.11	0.97	11.99	12.41
家族と訪洛 (D)	0.88	11.06	0.97	10.41	10.58	0.97	10.75	12.21
初訪洛 (D)	-0.41	-3.16	-0.55	-4.21	-3.43	-0.55	-5.30	-3.29
自動車所要時間/10000 (分)	-1.11	-5.34	-1.08	-4.94	-4.57	-0.65	-3.25	-2.73
自動車費用/10000 (円)	-0.42	-0.90	-0.36	-0.73	-0.69	-1.71	-3.87	-3.10
鉄道所要時間/10000 (分)	0.96	6.11	0.84	5.32	4.58	0.91	6.61	5.02
鉄道費用/10000 (円)	0.49	2.32	0.78	3.96	3.04	1.08	6.05	4.23
定数項	-1.88	-16.95	-2.70	-18.67	-20.93	-2.40	-17.80	-21.40
標本数	4,987							
L(0)	-3456.7		-3456.7			-3456.7		
L(C)	-2708.6		-2213.4			-2585.18		
L(β)	-2285.9		-1875.6			-2193.73		
χ^2	2341.7		3162.3			2526.00		

†: 正の係数はその変数の値の増加につれ自動車利用の確率が増加することを示す.
t^*: (15) 式に示される方法で推定された共分散行列に基づく.
t^{**}: 共分散行列 $\hat{\Sigma}=\hat{\Lambda}^{-1}$ に基づく.
(D): 0-1 ダミー変数を表す.

抽出法による標本への適用を数値例をもって示した．路側調査や乗客調査などの内生標本抽出が複数箇所で行われる場合，各調査地点での抽出率に基づいて抽出地点別に標本の重みを算出するという単純な方法が用いられるのが一般的であったが，この方法では母集団内で同一の抽出確率を持つ個体が，抽出された場所により異なった重みを付与されるという非合理が生じる．さらに加重層別標本抽出法の場合，選択肢別抽出法により得られた標本も無作為抽出により抽出される可能性を持っていたため，無作為抽出により抽出される確率を用い重み付けを行なわない場合ひずみが生じる．本稿ではこれらの問題点を指摘するとともに，多次元選択肢別標本や加重層別標本に一元的な重みが存在することを示し，重み付け方法の改良を提案した．さらに，重み付きと重み無しで離散選択モデルを推定した場合の係数値と t-値を比較し，共分散行列を適切に推定することの重要性を示した．

本研究の解析では複雑な内生標本抽出法に基づいた標本に，汎用的に適用可能な重みの一般形を導出している．しかしながら，本稿で紹介した事例にも示されるように，この重みを算定するに際して必要な情報は，これまでの調査では収集されないことが一般であったといえよう．本論文の解析は，標本抽出法が与えられたとして，どのような情報が重み付けに必要となるかを示すもので，付帯調査等の設計に有用であると考えられる．また，欠落している情報を推定することが可能ならば，現存のデータの重みを再算定することによってより適切な推定，解析が可能となり，既存情報の有効利用につながろう．このために必要となる imputation などの手法の開発が今後の課題として残る．

謝辞：本研究で用いたデータは京都市休日交通体系調査の一環として行われた調査に基づくものである．調査の実施に際してご協力いただいた，京都市，京都市政策企画室(当時)の中村嘉次氏，社団法人システム科学研究所，ならびに，調査に関連して設置された勉強会において有益なご助言をいただいた，山梨大学の西井和夫助教授(現教授)，京都大学の藤井聡助手(現助教授)に対して深甚な謝意を表したい．

注

[1] これらに加え都心繁華街でも調査票が配布されたが，得られたデータは本研究の解析には用いていない．

[2] 離散選択モデルの重み付き推定としては，パネルデータの解析における消耗バイアスの修正という目的で，いくつかの研究で WESML を適用した分析が行われている[12)-16)]．また，Kitamura et al. の一連の研究では，消耗バイアスと選択肢別抽出を共に考慮した重みの算出を行っている[7),17),18)]．

参考文献

1) 森地茂，屋井鉄夫：非日常的交通への非集計行動モデルと選択肢別標本抽出法の適用性，土木学会論文報告集，343号，pp. 161-170，1984.

2) 土木学会土木計画学研究委員会 (編) 非集計行動モデルの理論と実際．東京: 土木学会，1995.

3) Cosslett, S.R.: Maximum likelihood estimator for choice-based samples. *Econometrica*, Vol. 49, No. 5, pp. 1289-1316, 1981.

4) Manski, C.F. and McFadden, CD.: Alternative estimators and sample designs for discrete choice analysis. In C.F. Manski and D. McFadden (eds.) *Structural Analysis of Discrete Data*,

Cambridge: MIT Press, pp. 2-50, 1981.
5) Lancaster, T. and Imbens, G.: Choice-based sampling of dynamic populatins. In J. Hartog, G. Ridder and J. Theeuwes (eds.) *Panel Data and Labor Market Studies*, Amsterdam: North-Holland, pp. 21-43, 1990.
6) Thill, J.C. and Horowitz, J.L.: Estimating a destination-choice model from a choice-based sample with limited information. *Geographical Analysis*, Vol. 23, No. 4, pp. 298-315, 1991.
7) Kitamura, R., Pendyala, R.M. and Goulias, K.G.: Weighting methods for choice-based panels with correlated attrition and initial choice. In C.F. Daganzo (ed.) *Transportation and Traffic Theory*, Amsterdam, Elsevier Science Publishers, pp. 275-294, 1993.
8) 日本道路公団大阪管理局：大阪管理局管内　交通量統計　平成６年分, pp. 40, 1995.
9) 京都市産業観光局：京都市観光調査年報　平成７年, pp. 18, 1995.
10) Manski, C. and Lerman, S.: The estimation of choice probabilities from choice-based samples. *Econometrica*, Vol. 45, pp. 1977-88, 1977.
11) 森地茂, 屋井鉄夫, 石田東生：非日常的交通行動への非集計モデルの適用－チョイスベイストサンプルに対する推定問題の検討－, 土木計画学研究・講演集, No. 5, pp. 442-449, 1983.
12) 西井和夫, 近藤勝直, 古屋秀樹, 鈴木隆：パネルアトリッションを考慮した買物場所選択モデル：甲府買物パネルデータを用いて, 土木計画学研究・論文集, No. 12, pp. 389-396, 1995
13) 佐々木邦明, 森川高行, 杉山幸司：パネルサンプルの初期摩耗を考慮した動的な買物目的地選択モデル, 土木計画学研究・論文集, No. 13, pp. 595-602, 1996
14) 倉内文孝, 飯田恭敬, 塚口博司, 宇野伸宏：駐車場案内システム導入によるドライバーの駐車行動変化の実証分析, 都市計画論文集, No. 31, pp. 457-462, 1996.
15) 杉恵頼寧, 藤原章正, 小笹俊成：選好意識パネルデータを用いた交通機関選択モデルの予測精度, 土木学会論文集, No. 576/IV-37, pp. 11-22, 1997.
16) 西井和夫, 近藤勝直, 古屋秀樹, 栃木秀典：多時点パネルのアトリションバイアスとその修正法に関する研究：甲府買物パネルデータを用いて, 土木計画学研究・論文集, No. 14, pp. 653-662, 1997.
17) Pendyala, R. M., Goulias, K. G. and Kitamura, R.: Development of weights for a choice-based panel survey sample with attrition, *Transportation Research A*, Vol. 27A, No. 6, pp. 477-492, 1993.
18) Pendyala, R. M. and Kitamura, R.: Weighting methods for attrition in choice-based panels, In T. F. Golob, R. Kitamura and L. Long (eds.) *Panels for Transportation Planning*, Massachusetts, Kluwer Academic Publishers, pp. 233-257, 1997.
19) 山本俊行, 松田忠士, 北村隆一：保有予定期間との比較に基づく世帯における自動車保有期間の分析. 土木計画学研究・論文集, No. 14, pp. 799-808, 1997.

(2000. 3. 2 受付)

WEIGHTED ESTIMATION OF DISCRETE CHOICE MODELS WITH DATA FROM COMPLEX ENDOGENOUS SAMPLING

Ryuichi KITAMURA, Hiromu SAKAI and Toshiyuki YAMAMOTO

The methodology for weighting endogenous samples is extended in this study and applied to a multi-dimensional choice-based sample. It is shown that a unified weight exists for samples obtained from complex endogenous sampling schemes. Problems with conventional weighting methods for samples from endogenous sampling, such as roadside surveys, are pointed out and an improved weighting scheme is proposed. Coefficient estimates of discrete choice models with and without weights are compared, and the importance of applying an appropriate covariance estimator is pointed out.