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 ──

東京工業大学　正　森地　茂
東京工業大学　学○屋井　鉄雄
筑波大学　　正　石田　東生

1 はじめに

本研究の目的は大別して次の2つである。　第1は観光交通への非集計モデルの適用性についての検討であり、第2は選択肢別抽出サンプル（choice-based sample）に対するモデル推定上の問題点の検討である。

非集計モデル作成に要するデータは、多くの場合家庭訪問調査より得られるが、交通目的によってはそれが効率的でないことがある。　都市間の観光トリップや買廻り品の買物トリップなど発生頻度の少ない非日常的な交通行動のサンプルを家庭訪問調査から必要量得るには、大量の家庭を抽出するか、過去数ヶ月間以上に行なった交通行動を聞き出す必要がある。　前者の場合は莫大な費用を要し、後者の場合はデータの信頼性に問題がある。

これに対し実際にトリップしている人間を直接サンプリングできれば、コストまた調査の容易さからも非常に効率的である。　例えば交通機関選択分析のデータを、鉄道利用者については駅や列車内で得自動車利用者については駐車場や有料道路料金所で得る場合や、買物トリップの目的地選択分析のデータをショッピングセンターごとに得る方法がこれに相当する。　この方法は選択肢別標本抽出（choice-based sampling）と呼ばれ、一種の層別サンプリングとみなせる。

層別サンプリングは一般に行動を決定する要因ごとに層を分割することを意味するが、選択肢別標本抽出は行動の際の選択肢ごとにサンプリングする方法である。　本研究における層別サンプリングとは広義の層別サンプリング（general stratified sampling）を意味し、それは狭義の層別サンプリング（行動決定要因をもとに層に分ける方法で、本研究では特性値別標本抽出；exogenous sampling と呼ぶ）と選択肢別標本抽出の両者の層により分けられるものと定義する。

選択肢別標本抽出では効率的にデータを収集でき、そこで得たサンプルを家庭訪問調査などのサンプルに加え、シェアの小さい選択肢のサンプルを増してモデルを推定する方法が有効である。　このような方法はenriched sampling と呼ばれる。　例えばアクセス交通手段を分析するとき、家庭訪問調査では自転車利用のサンプルを多数集めることは困難である。　このとき選択肢別標本抽出により自転車利用者を駅で直接つかまえ、それらを家庭訪問調査のサンプルに加えモデルを推定する。　このように選択肢別標本抽出は、それ自体でモデルを推定するばかりではなく、enriched sampling として家庭訪問調査を補完する使い道がある。

選択肢別標本抽出は調査方法の特殊さに依存する幾つかの問題を有するが、その利用価値の高さを考慮すると、それらの検討が必要であると考える。

2 選択肢別抽出サンプルに対するモデルの推定

選択肢別抽出サンプルの尤度関数は無作為抽出の場合と異なる。　それを示すため、まず層別サンプルの尤度を考える。　層別サンプリングでは、①母集団をG個の層に分割し、②総サンプル数Nと各層の構成比 $H(g)$ $(g=1,\cdots,G;\ \sum_{g=1}^{G} H(g)=1)$ を決定する。　そして③各層ごとに $N_{(g)}$ 個 $(=N\cdot H(g))$ のサンプルを無作為に抽出する。　すなわち層別サンプリングは一組の $[G, H(g)(g=1,\cdots,G)]$ とNによって決定される。

非集計モデルに対し、層別は選択肢（実際に選択した代替案）と特性（例えば年収、住所といった社会経済特性）の2つに関してなされる。　ここでCを選択肢iの全体集合、Xを特性xの可測集合とし、$i\in C$, $x\in X$ なる (i,x) の組の全体集合を $C\times X$ で表わす。

$C \times X$ をG個の層に分ける場合を考えると、各層における(i,z)の集合は$(C\times X)_g$ $(g=1,\cdots,G)$で表わされる。 $(C\times X)_g$は層gに含まれるすべての(i,z)からなり、相互に排他的な集合である。

また母集団における(i,z)の確率密度$f(i,z)$は、

$$f(i,z) = P(i|z,\beta^*)\cdot P(z) \quad , (i,z)\in C\times X \qquad (2-1)$$

で与えられる。 $P(z)$は母集団における特性zの周辺分布、$P(i|z,\beta^*)$は未知なパラメータを有し選択を表わす関数、β^*はパラメータの真値である。
$P(i|z,\beta^*)$の関数形にはLogitモデルやProbitモデルがある。

各種の抽出方法に応じて複数個の(i,z)が観測され、β^*を推定する。 層別サンプリング[G, H(g)(g=1~G)]のとき、(i,z)はそれを含む層gが確率$H(g)$で選択され、続けてその層から無作為に抽出されると考える。

このとき観測値(i,z)を抽出する尤度は、

$$\ell(i,z) = \frac{f(i,z)}{\sum_{(j,y)\in(C\times X)_g} f(j,y)}\cdot H(g) \qquad (2-2)$$

で表わされる。 右辺の$f(i,z)/\sum f(j,y)$は層gが選択された場合に(i,z)が観測される割合を表わす条件付きの尤度である。

総サンプルN $(=\sum_{g=1}^{G} N_{sg})$の尤度関数L^*は、

$$L^* = \prod_{g=1}^{G}\prod_{t=1}^{N_{sg}} \frac{f(i_t,z_t)}{\sum_{(j,y)\in(C\times X)_g} f(j,y)}\cdot H(g) \qquad (2-3)$$

となる。 tは各サンプルを表わし、(i_t,z_t)はtの選択肢と特性を表わす。

以上が層別サンプリングの場合であるが、単純無作為抽出であれば、$\sum_{(j,y)\in(C\times X)_g} f(j,y) = H(g)$ $(g=1,\cdots,G)$となり、尤度ℓ_Rは、

$$\ell_R(i,z) = f(i,z) \qquad (2-4)$$

で表わされ、尤度関数L_R^*は、

$$L_R^* = \prod_{t=1}^{N} f(i_t,z_t) = \prod_{t=1}^{N} P(i_t|z_t,\beta^*)\cdot P(z_t) \qquad (2-5)$$

対数尤度関数L_Rは、

$$L_R = \sum_{t=1}^{N} \ln P(i_t|z_t,\beta^*) + \sum_{t=1}^{N} \ln P(z_t) \qquad (2-6)$$

右辺第2項はβ^*を含んでおらず、L_Rの最大化において省略できる。

また特性値別標本抽出のとき、層別は特性に関してのみなされ、特性の全体集合Xがg個の部分集合に分割される。 これを$(C\times X)_g = C\times X_g$で表わす。

$$\sum_{(j,y)\in C\times X_g} f(j,y) = \sum_{y\in X_g}\sum_{j\in C} P(j|y,\beta^*)\cdot P(y) = \sum_{y\in X_g} P(y) \qquad (2-7)$$

となるため尤度ℓ_Eは、

$$\ell_E(i,z) = \frac{f(i,z)}{\sum_{y\in X_g} P(y)}\cdot H(g) \qquad (2-8)$$

と表わされ、尤度関数L_E^*は、

$$L_E^* = \prod_{g=1}^{G}\prod_{t=1}^{N_{sg}} \frac{f(i_t,z_t)}{\sum_{y\in X_g} P(y)}\cdot H(g) = \prod_{g=1}^{G}\prod_{t=1}^{N_{sg}} P(i_t|z_t,\beta^*)\cdot \frac{P(z_t)\cdot H(g)}{\sum_{y\in X_g} P(y)} \qquad (2-9)$$

したがって対数尤度関数L_Eは、

$$L_E = \sum_{g=1}^{G}\sum_{t=1}^{N_{sg}} \ln P(i_t|z_t,\beta^*) + \sum_{g=1}^{G}\sum_{t=1}^{N_{sg}} \ln \frac{P(z_t)\cdot H(g)}{\sum_{y\in X_g} P(y)} \qquad (2-10)$$

で表わされる。
上式右辺第2項もβ^*を含んでおらず、無作為抽出の場合同様、第1項だけでβ^*の推定ができる。 このときの推定量をESML(Exogenous Sampling Maximum Likelihood)推定量と言う。

選択肢別標本抽出では、選択肢についてのみ層別され、選択肢の全体集合Cがg個の層に分割されるならば、各層は部分集合$(C\times X)_g = C_g\times X$で表わされる。

$$\sum_{(j,y)\in C_g\times X} f(j,y) = \sum_{j\in C_g}\sum_{y\in X} P(j|y,\beta^*)\cdot P(y) = \sum_{j\in C_g} Q(j) \qquad (2-11)$$

上式で$Q(j)=\sum_{y\in X} P(j|y,\beta^*)$は母集団での選択肢$j$のシェアを表わす。 ここで各層が唯一つの選択肢で構成されている場合を考える。(各層が複数の選択肢を有する場合はCosslett (1981)に詳しい。)[2] そのとき$C_g=\{i\}$と表わせ、(2-11)より、$\sum_{j\in C_g} Q(j) = Q(i)$となる。 したがって尤度$\ell_c$は、

$$\ell_c(i,z) = \frac{f(i,z)}{Q(i)}\cdot H(i) \qquad (2-12)$$

また尤度関数L_c^*は、

$$L_c^* = \prod_{i\in C}\prod_{t=1}^{N_{si}} \frac{f(i_t,z_t)}{Q(i_t)}\cdot H(i_t) = \prod_{i\in C}\prod_{t=1}^{N_{si}} \frac{P(i_t|z_t,\beta^*)P(z_t)}{Q(i_t)}\cdot H(i_t) \qquad (2-13)$$

対数尤度関数L_cは、

$$L_c = \sum_{i\in C}\sum_{t=1}^{N_s} \ln P(i_t|z_t,\beta^*) + \sum_{i\in C}\sum_{t=1}^{N_{si}} \ln \frac{P(z_t)\cdot H(i_t)}{Q(i_t)} \qquad (2-14)$$

(2-14)式の右辺第2項のQ(i)は(2-11)で示したようにθ*の関数となっており、尤度関数の最大化にあたり、無作為抽出や特性値別標本抽出でしたような省略はできない。この点が大きく異なり最大化を困難とする。

一般に分析者側で有する情報はたかだかQ(i)程度で、P(z)が既知であることはまず無い。Q(i)が既知のとき、制約式$Q(i)=\sum_{y\in X} P(i|y,\theta^*)\cdot P(y)$のもとで、θ*, P(z)の同時推定を行なえるが、P(z)の設定等に問題がある。[1] これに対し幾つかの代替的な推定量が提案されている。それらはP(z)を含まず推定上扱いやすい。代替推定量には、①WESML推定量や②MM推定量などがある。

①WESML(Weighted-ESML)推定量は、漸近正規性、一致性を有するもので(2-15)式で求まる。

$$WESML = \max_{\theta} \sum_{i\in C}\sum_{t=1}^{Nsi} \frac{Q(i_t)}{H(i_t)} \cdot \ln P(i_t|z_t,\theta) \qquad (2-16)$$

分散共分散行列や一致性の証明はManski & Lerman(1977)に詳しい。[4]

②MM推定量(Manski & McFadden)では、条件付き尤度を対象に尤度関数を構成する。条件付き尤度は、(2-12)より、

$$l_c(i|z) = \frac{l_c(i,z)}{\sum_{j\in C} l_c(j,z)} = \frac{P(i|z,\theta^*)\cdot\frac{H(i)}{Q(i)}}{\sum_{j\in C} P(j|z,\theta^*)\cdot\frac{H(j)}{Q(j)}} \qquad (2-17)$$

上式は、分布P(z)を含んでおらず、(2-18)でθが推定できる。

$$MM = \max_{\theta} \sum_{i\in C}\sum_{t=1}^{Nsi} \ln \frac{P(i_t|z_t,\theta)\cdot\frac{H(i_t)}{Q(i_t)}}{\sum_{j\in C} P(j|z_t,\theta)\cdot\frac{H(j)}{Q(j)}} \qquad (2-18)$$

MM推定量も漸近正規性および一致性を有する。さらにP(i|z,θ*)が代替案マイナス1個の定数項を含んだLogitモデルであれば、漸近有効となる。[2]

また特別なケースとして、モデル式にLogitモデルを採用し、定数項を代替案マイナス1個導入すれば、ESML推定により、一致推定量を得ることができる。このとき定数項は、

$$C_i' = C_i + \ln\frac{Q(i)}{H(i)}\cdot\frac{H(M)}{Q(M)}, \quad i=1,\cdots,M-1 \qquad (2-19)$$

によって修正を要する。C_iは推定された定数項、Mは代替案数である。この推定量が一致推定量となることの証明は、Lerman & Manski(1977)を参照されたい。[4]

3 調査データと観光交通への非集計モデルの適用結果

本研究の分析に用いるサンプルは、表1に示す鉄道と車の選択肢別標本抽出から得られた。分析には全サンプルより再抽出した鉄道利用者644、車利用者632を用いる。長距離の観光トリップは発生頻度が少なく、所要時間等のLOS(Level of Service)値の認識は正確でない。それは代替交通機関について顕著であり、アンケート調査からそれらの情報を集めても信頼性は低く問題がある。

ここでは1つの対応策として、個人ごと代替機関(利用機関も含む)ごとにLOS値を設定した。時刻表、地図等を用い、ある一定の判定基準のもとで、所要時間最小となる鉄道・車のルートを探索し、そのときの値を各代替案のLOS情報とした。

本章のモデル推定に用いるサンプルは、実数調査の結果より推計した鉄道対車のシェア(0.3:0.7)に合うよう1276より再抽出した900である。選択肢別抽出サンプルにおける推定上の問題を避けるためにランダムサンプルとしており、これが母集団を偏りなく表わすと考える。モデル式にはLogitモデルを採用した。モデル変数組の決定は、非集計モデルが効用理論に基づく個人行動モデルであることに留意し行なった。各パラメータ値の意味が明確となるよう高い相関を持つ変数の同時導入を避け、符号条件・七値等をチェックしつつ、尤度比・的中率が改善されるように

表1 調査概要とサンプル

対象交通機関	鉄道	自動車
調査時期	昭和55年 5月20日(水)～23日(土)	
調査対象	・中央本線下り特急 急行乗客	・中央自動車道下り利用自動車のうち乗用車
調査方法	・中央本線特急 急行列車内で新宿発後に調査票配布 八王子付近で回収	・中央自動車道 大月 勝沼 河口湖 各ICで調査票配布 郵送にて回収
旅客数 通過台数※	21085人	24167台
回収数(回収率)	4976(23.6%)	2320(9.6%)
分析に用いるサンプル数	644票	632票
サンプルの抽出条件	・発地が1都3県 ・観光目的(帰宅トリップは除く)	・鉄道利用時には新宿から中央本線に乗車すると回答した者

※同日行なった実数調査より

変数を増加した。

推定結果の例を、表2に示す。モデルはいずれも説明力の高いもので、観光交通への適用性の高さが確かめられた。トリップの特性を示すダミー変数の有意性は高くなっている。また係数の符号から、目的地が多く、日帰りで同伴者が1～5人、また車を保有していることが、鉄道代替案の効用を下げることがわかる。これは常識的な傾向である。

次にLOS変数の係数を評価する。所要時間の1分の減少は鉄道利用時のイグレスコストの約13円の増加と同等で、これは4つのモデルでほとんど変化しない。またモデル4より、鉄道利用時アクセスコストについては、それが14.8円となり、若干アクセスの評価が低いことになる。またモデル1で所要時間と自動車イグレス時間の係数を比較すると前者は後者の2.5倍である。これは旅行者の時間に対する全体的な評価が、イグレスという一部の評価より高いことを示し、観光交通のようにトリップの長い場合には納得できる結果である。

各モデル間で的中率、尤度比に有意な差はないが、本研究ではモデル2を基本的なモデルに位置付け、以後の分析に用いることとする。

表2　観光交通への非集計モデルの適用結果

MODEL NO	1	2	3	4
鉄道アクセスコスト(円)　鉄				−0.001099 (1.57)
鉄道イグレスコスト(円)　鉄	−0.001245 (3.80)	−0.001241 (3.79)	−0.001142 (3.56)	−0.001231 (3.76)
自動車イグレス時間(分)　車	−0.006330 (4.30)			
高速からのイグレス距離(Km)車		−0.01122 (4.42)		−0.01098 (4.30)
自動車総走行距離(Km)　車			−0.009134 (3.56)	
総所要時間(分)　共通	−0.01595 (4.55)	−0.01545 (4.37)	−0.01558 (4.14)	−0.01631 (4.52)
目的地数2以上ダミー　鉄	−1.161 (4.70)	−1.157 (4.68)	−1.123 (4.60)	−1.147 (4.63)
日帰りダミー　鉄	−0.9518 (3.08)	−0.9347 (3.01)	−0.9772 (3.14)	−0.9227 (2.97)
同伴者1～5人ダミー　鉄	−2.479 (10.18)	−2.480 (10.16)	−2.492 (10.28)	−2.489 (10.13)
車保有ダミー　鉄	−3.760 (11.70)	−3.778 (11.72)	−3.754 (11.74)	−3.771 (11.73)
定数項　鉄	4.558 (9.80)	4.524 (9.70)	3.628 (5.75)	4.779 (9.54)
カイ2乗値	571.6	572.8	565.7	575.4
ユウ度比	0.516	0.517	0.510	0.519
機関別的中率(%)　鉄道	74.8	74.8	75.6	75.6
機関別的中率(%)　自動車	94.4	94.4	94.1	94.1
全体の的中率	88.6	88.6	88.6	88.6

サンプル数　900（鉄：270　車：630）

4選択肢別抽出サンプルに対する推定問題の検討

選択肢別抽出サンプルを用いモデルを推定するとき、サンプル数はどれほどあれば十分であるか、そのとき代替案ごとのサンプルの割合をどう設定するのが望ましいか。また推定量にはどれを採用すれば良いか。さらに推定時に必要な母集団でのシエアの値にはどれほどの精度が要求されるか等、モデル作成にあたり多くの検討課題がある。本章ではこれらの問題点を検討した。分析は全サンプル(1276)を対象とし、モデル変数組は3章で求めた基本モデル（表2、モデル2）と同一とした。

1)モデル推定結果

2章で示した各種推定量によるモデルを表3に記す。モデル構造にはLogitモデルとProbitモデルの2つを、推定量にはESML、WESML、MMの3つを取り扱っている。表中LogitモデルでMM推定量がない理由は、この場合MM推定量が漸近有効であり、ESMLと一致するからである。またProbitモデルでESMLがないのは、ESMLによる推定が一致性を持つのがLogitモデルの場合だけであることによる。

WESML、MM推定量のカイ2乗値、尤度比（自由度調整済）は、個々の関数ごとESMLと同様に算出している。初期尤度は個々の確率に母集団でのシ

表3　選択肢別抽出サンプルに対するモデル推定結果

MODEL TYPE		LOGIT MODEL		PROBIT MODEL	
ESTIMATOR		ESML	WESML	WESML	MM
θ1	総所要時間(分)　共通	−0.01569 (5.71)	−0.01530 (4.91)	−0.008003 (4.90)	−0.008068 (5.71)
θ2	鉄道イグレスコスト(円)　鉄	−0.001375 (5.31)	−0.001301 (5.13)	−0.0007227 (5.45)	−0.0007659 (5.57)
θ3	高速からのイグレス距離(Km)車	−0.01165 (5.72)	−0.01074 (4.98)	−0.006315 (5.55)	−0.006735 (6.21)
θ4	目的地数2以上ダミー　鉄	−1.211 (6.27)	−1.255 (6.12)	−0.6534 (6.11)	−0.6180 (6.12)
θ5	日帰りダミー　鉄	−1.237 (4.89)	−1.223 (4.94)	−0.6559 (4.93)	−0.6535 (4.94)
θ6	同伴者1～5人ダミー　鉄	−2.323 (11.86)	−2.340 (11.42)	−1.283 (11.53)	−1.263 (11.98)
θ7	車保有ダミー　鉄	−3.805 (14.15)	−3.827 (13.71)	−2.125 (14.22)	−2.104 (14.87)
θ8	定数項　鉄	4.514 (11.75)	4.584 (10.45)	2.463 (10.53)	2.405 (12.27)
カイ2乗値		1060.62	816.74	813.60	919.78
ユウ度比		0.527	0.521	0.519	0.517
機関別的中率(%)	鉄道	75.6	75.6	75.5	75.0
機関別的中率(%)	自動車	94.8	94.6	94.8	94.8
全体での的中率(%)		89.0	88.9	89.0	88.9

サンプル数　1276（鉄：644　車：632）

エアQ(i)を代入することにより定義している。 すなわち、ESML, WESML, MMに対し、$N\cdot\sum_{i\in C}H(i)\cdot\ln Q(i)$, $N\cdot\sum_{i\in C}Q(i)\ln Q(i)$, $N\cdot\sum_{i\in C}H(i)\ln H(i)$ となる。 ランダムサンプルであれば3者とも同一の値となる。 全体での的中率とは、$\sum_{i\in C}Q(i)\cdot PC(i)$ (PC(i), i代替案の的中率) により算出した。 これは母集団に対する説明力指標である。 本データでは、Q(1)(母集団での鉄道利用シエア)は0.3, またH(1)(サンプルでの鉄道利用シエア)は約0.5である。

4種の推定結果で説明力が他に比べ著しく劣るものはない。 MM推定量の鉄道的中率が若干低いが有意な差ではない。 パラメータ値もモデルごとにみれば大きな差とは言えない。 LogitとProbitでパラメータ値が大きく異なるのはProbitで誤差の分布をN(0,1)としているためである。 全サンプルを用いた本節の分析ではモデル・推定量間でほとんど差のない結果が得られた。 次節以降では、それをより細かく分析する。

2)分担率の推計誤差のモデルへの影響

選択肢別抽出サンプルを用いた推定は一般にQ(i)(母集団での選択肢iのシエア:分担率)の情報を必要とする。 ここでQ(i)の推計誤差がモデル推定結果に与える影響を把握しておく必要がある。

ESML推定量ではQ(i)の差は定数項の値だけを変化させ、その結果的中率が変化する。 しかしWESML, MM推定量ではすべてのパラメータが歪んでしまう可能性がある。 本節では、Logit, Probit両モデルでWESML推定量を用いた場合の検討結果を示す。 図の1と2で、縦軸は、Q(1)が0.3のときの係数($\theta_{0.3}$)とその分散($\sigma^2_{0.3}$)を基準とし、設定したQ(1)(=0.1, 0.2, 0.4, 0.5, 0.7, 0.9; これらを誤って推計された値とする。)ごとのθのひずみを表わす。 $TV=|\theta-\theta_{0.3}|/\sqrt{\sigma^2_{0.3}}$ である。
図より、設定値が0.3から遠ざかるにしたがい、TV値が大きくなることが読み取れる。 パラメータ間の相関があり、個々のパラメータごとに差を評価することは望ましくないが、設定値が0.2, 0.4(推計誤差0.1)程度のとき、全般にTV値が小さくかなり安全側で評価しても有意な差とは言えない。 しかし定数項については、0.1の誤差でもかなり大きな差がある。 そしてこのような傾向はLogit, Probitでさほど変わらない。 Q(i)に関する情報がまったく無い場合を除くと、誤差を0.1以下に抑えることは十分可能である。 したがって定数項以外のパラメータの歪みのない値を得るためには、Q(i)に高い精度が要求されないことがわかる。

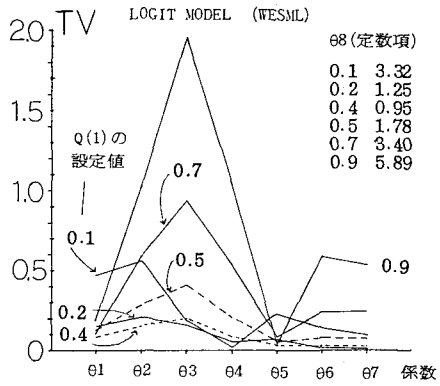


図1 Q(1)の推計誤差と係数の変動(Logit Model)

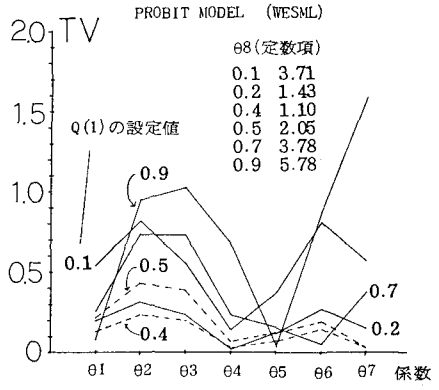


図2 Q(2)の推計誤差と係数の変動(Probit Model)

3)サンプルデザインと推定量についての分析

本研究で対象とする推定量の統計的性質は漸近理論に基づく。 したがって実際にモデル推定する有限個のサンプルに対しては実証的な検討が必要となる。
それによって各推定量の安定性を知り、推定に必要な代替案ごとのサンプル数を把握できる。

本節の分析では、全サンプル(1276)を母サンプルとし、そこから任意のサンプル数(N)とサンプルにおける代替案のシエア(H(i): 2章で示した層ごとの構成

比）を満すよう再サンプリングしたものを対象とする。　しかし約1300というサンプルサイズが十分大きく、かつ母集団を偏りなく表わしているとは言いきれない。　そのことが本節の分析上の問題であることを付記しておく。

Logitモデルに対する分析のフローを、図3に示す。　各ケース（N, H(1)）ごとに、サンプル抽出、モデル推定を複数回繰り返し、パラメータの平均値を求める。　そしてそれに最も近いモデルを平均的モデルと称し、そのケースを代表させる。　異なるN, H(1)間で、複数回の抽出によるパラメータのばらつきを直接比較するためには、十分な数の母サンプルを必要とする。　なぜならば、サンプル数が十分でないと、そこからの抽出率が、ばらつきに影響するからである。　しかし、その場合にもパラメータの平均ベクトルは変わらないと考え、平均的モデルの有意性によって、ケース間の比較を行なうことにした。　また同一ケースで推定量間の比較を行なう場合には、以下に示すMRMS値などの、ばらつきを直接表わす値を用いた。

$$MRMS = \frac{1}{Np-1}\sum_{\substack{i=1 \\ (i\neq M)}}^{Np} RMS_i \qquad (4-1)$$

$$RMS_i = \sqrt{\frac{1}{k}\sum_{k=1}^{k}(\theta_{ik}-\theta_{Mk})^2} \qquad (4-2)$$

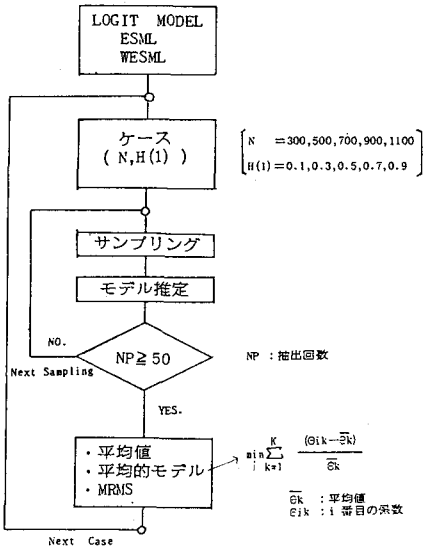


図3　分析のフロー

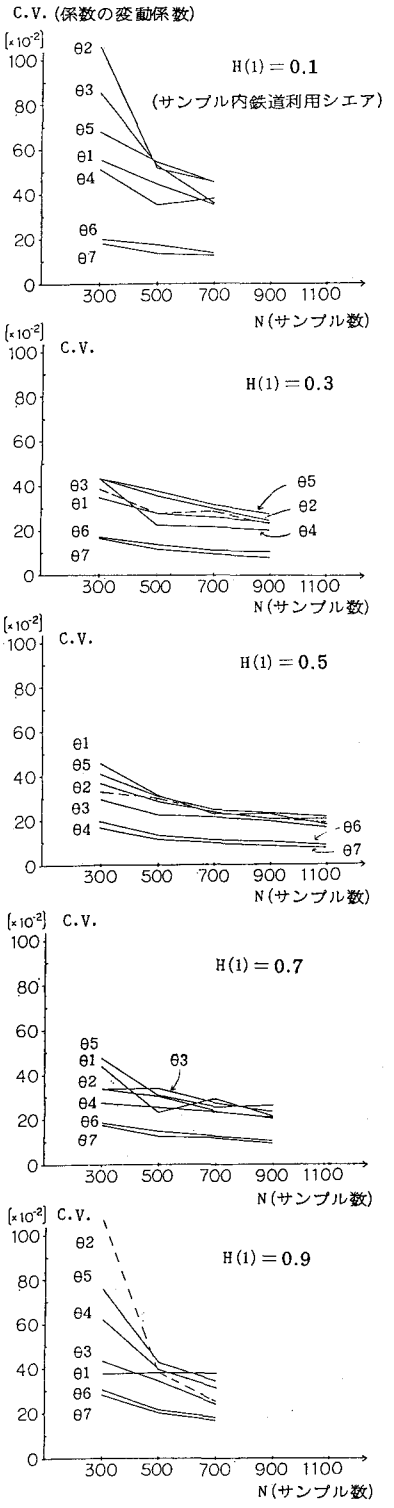


図4　各ケースの平均的モデルの変動係数

(4-1), (4-2)で、Kはパラメータ数、Npはサンプルの抽出回数、θ_{ik}はi番目の抽出サンプルによるk番目のパラメータ、θ_{Mk}は平均的モデルのパラメータをそれぞれ表わす。(4-1), (4-2)を各ケースごとに算出した。

なお、各ケースごとに算出した的中率や尤度比の平均値にはほとんど差がなく、平均的モデルの全サンプルにおける的中率も変わらない。

a)サンプルデザイン

図4は平均的モデルの各パラメータの変動係数を各ケースごとに示したものである。(C.V. = σ_k/θ_{ok})これらはLogit-ESMLの推定量の場合である。H(1)が0.1や0.9のとき、C.V.は他のH(1)と比べ全般に大きく、C.V.が0.5以下であれば統計的に5%有意であることを考慮すると、300サンプルでは不十分と言える。またH(1)が0.3, 0.5, 0.7のとき、各パラメータともC.V.は小さく有意であり、サンプル数の増加に伴うC.V.の減少はあまりない。この場合には300サンプルでも有意な推定がなされている。

したがって本分析に用いたデータ、モデル変数組について言えば、各代替案ごと150サンプルづつ抽出しても十分であり、有意な推定結果の得られる可能性の高いこと伏、事後的に判明した。

b)推定量間でのパラメータの安定性

図5はLogitモデルにおける推定量間でのMRMSの差を図示したものである。全般にWESMLのパラメータのばらつきが大きいこと伏読み取れるが、それはサンプル数が少なく、かつH(1)が極端に大きい場合に顕著である。逆にサンプル数が700でH(1)が0.7以下では、両者で差はほとんど認められない。

次にProbitモデルに対して、WESML, MM推定量間で若干の検討を行なった。ここでは計算時間等の制約から、検討するケースを限定し、(N, H(1))として、(300, 0.5)と(700, 0.5)の2ケースを対象とした。H(1)=0.5は良好なシエアと考えられ、N=300は必要最小限のサンプル数、700は十分なサンプル数と考えられる。(a)より)

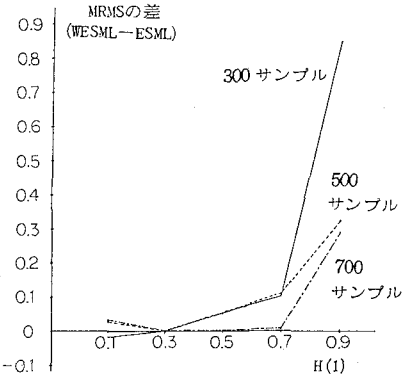
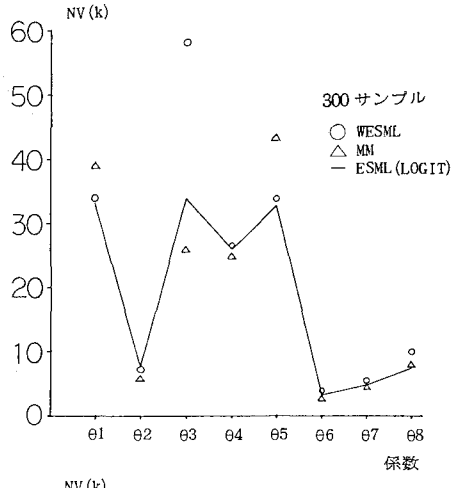


図5 推定量と係数のばらつき(Logit Model)



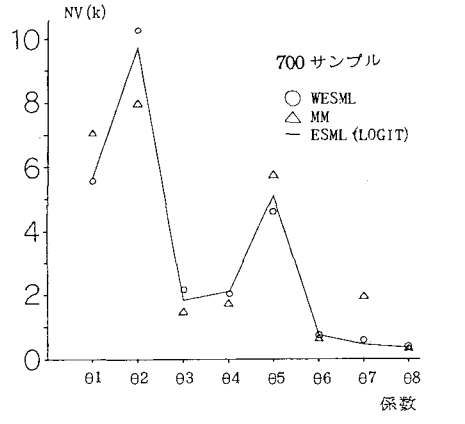


図6 推定量と係数のばらつき(Probit Model)

表4 推定量ごとのNV値

推定量＼サンプル	300	700
WESML	177.6	26.00
MM	153.4	26.96
ESML	148.9	26.32

各ケースで10回づつサンプリングを行ない、推定量の変動を比較した。比較指標として、

$$NV(k) = \sum_{j=1}^{Np} \sum_{i=1}^{Np} (\theta_{ik} - \theta_{jk})^2 / \theta_{jk}^2 \qquad (4-3)$$

$$NV = \sum_{k=1}^{K} NV(k) \qquad (4-4)$$

を用いた。図6はNV(k)をパラメータごとに示したものである。直線は同一ケースでLogit-ESMLの結果である。700サンプルでは各推定量とも同様な傾向を示し大きな差は認められない。表4に示したNVの値も大差ない。しかし300サンプルでは、θ3, θ4で差があり、θ3ではWESMLが、またθ4ではMMのばらつきが大きい。NVの値もそれを受け、WESMLではかなり大きい。

抽出回数が少なく結論付けることは危険であるが、ここではサンプル数が少ない場合にWESMLのばらつきのより大きいことがわかった。そしてこれはLogitモデルの結果と一致する。

5 まとめと今後の課題

本研究では大きく分けて2つの問題を扱った。すなわち、①観光交通行動への非集計モデルの適用性と②選択肢別抽出サンプルに対する推定問題の2つである。

①では、行動を記述する優れた要因の抽出がなされ、観光交通における機関選択の特性の把握ができた。したがってそこで得たモデルは説明力の十分高いものであった。

②では、小数のサンプルに対する推定量の特性を検討し、幾つかの有用な知見が得られた。

すなわち、

i) Logit-ESML, Logit-WESML, Probit-WESML, Probit-MMという4種の推定結果は同一サンプルに対しては同程度の説明力を有すること。

ii) 母集団における シェア(Q(i))の推計については、さほどの精度を必要としないこと。但し、定数項については大きく偏る可能性があり、注意を要すること。

iii) 推定に用いるサンプル数は、代替案ごと同数とすることが望ましく、そのときサンプル総数は、300～500であれば十分であること。

iv) パラメータの安定性は、Logitモデルについては全般にWESMLよりもESMLの方が良く、サンプル数が少なく、サンプルにおける代替案のシェア(H(i))がQ(i)と大きく異なる場合に、それが顕著であること。

しかし、これらは本研究で用いたデータに関する結論であり、観光交通における二肢選択を対象とし、実データより得た最良なモデル変数組のうちの1つによる検討結果である。多肢選択では、H(i)の設定における自由度が増し、またモデル変数の導入形式も複雑になる等の問題があり、検討すべき点は多い。また推定量の安定性の検討などは実データによらずとも機械的に行なえる。これらの検討は今後の課題としたい。

参考文献

1 Manski, C.F. & McFadden, D. (1981)
Alternative Estimators and Sample Designs for Discrete Choice Analysis
Structural Analysis of Discrete Data with Econometric Applications

2 Cosslett, S.R. (1981)
Efficient Estimation of Discrete-Choice Models
Structural Analysis of Discrete Data with Econometric Applications

3 Lerman, S.R. (1980, Unpublished paper)
Theory of Sampling and Sample Design

4 Manski, C.F. & Lerman, S.R. (1977)
The Estimation of Choice Probabilities from Choice based Samples
Econometrica Vol 45, No8

5 Lerman, S.R. & Manski, C.F. (1975)
Alternative Sampling Procedures for Calibrating Disaggregate Choice Models
TRR592

6 Lerman, S.R. & Manski, C.F. (1979)
Sample Design for Discrete Choice Analysis of Travel Behavior: The state of the art
Transpn. Res. Vol 13A

7 屋井鉄雄, 森地茂, 石田東生 (1982)
チョイスベイストサンプルを用いた非集計交通機関選択モデル
第37回年次学術講演会講演概要集IV P379～380

8 三宅光一, 森地茂, 屋井鉄雄 (1982)
観光交通のモーダルスプリット分析
第37回年次学術講演会講演概要集IV P377～378