

非集計ロジットモデルの集計バイアスに関する研究*

A Study on Aggregation Bias of Disaggregate Logit Model

桐 越 信**
浅 野 光 行***
塚 本 直 幸****

1. はじめに

非集計モデルは、個々の選択行動単位における選択行動をゾーンごとに集計せずに、選択行動単位レベルのデータをそのままモデル構築のためのデータとするところにその特徴がある。この非集計モデルを交通需要の予測や交通政策の評価のために用いる場合には、対象とする集団全体における代替案の選択比率の値やその変化値によって需要予測や交通政策の効果分析を行うので、選択行動単位における代替案の選択確率をもとにした対象とする集団全体における代替案の選択比率を求める必要がある。選択行動単位における代替案の選択確率をもとに対象とする集団全体における代替案の選択比率を求めることが「集計化」といわれるものであり、現在までにこの集計化に関してすでにいくつかの方法が提案されている。本研究では、そのうち比較的適用が簡単とされている平均値法とモーメント法の2つの方法を取りあげ、集計化の解析的な意味について非集計ロジットモデルを例にあらためて考察し、その結果をもとに集計バイアスの大きさについて理論的検討を行う。

はじめに、平均値法、モーメント法による集計化は、選択確率関数であるロジット関数を他の関数で置き換えたりえて期待値演算を行うことに等しく、いわゆる集計バイアスはこの関数の代用により生ずるものであることを示す。つぎに、平均値法、モーメント法の集計バイアスについて効用値の分布特性との関連で理論的検討を行い、各方法の集計バイアスの大きさを集計バイアス曲線として表わす。さらに、効用値の分布特性をあらかじめ設定してシミュレーションにより作成した仮想観測値データをもとに平均値法、モーメント法による選択比率の推定精度とデータサイズとの関係について把握する。

2. 集計化の理論的検討

(1) 非集計ロジットモデル

いま、2つの選択可能な代替案 l, m がある場合に、選択行動単位 i が代替案 l, m を選択したときの効用 u_{il}, u_{im} はそれぞれ式(1)、(2)で表されるものとする。式(1)、(2)において、 v_{il}, v_{im} は代替案の特性や選

$$u_{il} = v_{il} + \epsilon_{il} = \beta_{0l} + \sum_{k=1}^P \beta_k x_{ilk} + \epsilon_{il} \quad (1)$$

$$u_{im} = v_{im} + \epsilon_{im} = \beta_{0m} + \sum_{k=1}^P \beta_k x_{imk} + \epsilon_{im} \quad (2)$$

択行動単位 i の社会経済的属性などから構成される部分で、それぞれ変数 x_{ilk}, x_{imk} の一次結合で表されるものとする。また、 $\epsilon_{il}, \epsilon_{im}$ は誤差項として確率的な変動をする部分で、それぞれ互いに独立に式(3)で表される確率密度関数 $f(\epsilon)$ をもつガンベル分布に従うものとする。

$$f(\epsilon) = e^{-\epsilon} e^{-e^{-\epsilon}} \quad (3)$$

* 交通需要予測、非集計ロジットモデル、集計バイアス

** Shin KIRIKOSHI, 正会員 工修 建設省都市局都市交通調査室係長

*** Mitsuyuki ASANO, 正会員 工博 建設省建築研究所都市施設研究室長

**** Naoyuki TSUKAMOTO, 正会員 工修 幹長大橋設計センター交通計画部

式(1)から式(2)を引いて、 $u_i = u_{il} - u_{im}$, $v_i = v_{il} - v_{im}$, $x_{ik} = x_{ilk} - x_{imk}$, $\beta_0 = \beta_{0l} - \beta_{0m}$ とおくと式(4)が得られる。

$$u_i = v_i + \epsilon_{il} - \epsilon_{im} = \beta_0 + \sum_{k=1}^P \beta_k x_{ik} + \epsilon_{il} - \epsilon_{im} \quad (4)$$

この場合に、選択行動単位 i によって代替案 l が選択される確率を p_{il} とすると、 p_{il} は式(5)で表される。

$$\begin{aligned} p_{il} &= \text{Prob}[u_{il} > u_{im}] \\ &= \text{Prob}[u_i > 0] \\ &= \frac{1}{1 + e^{-v_i}} \end{aligned} \quad (5)$$

ここで、式(4)の x_{ik} が連続的に変化するものとする、 v_i , u_i も同様に連続的に変化する、連続的に変化する x_{ik} を x_k で表し、 v_i , u_i をそれぞれ式(6)、(7)のように v , u と表す。各説明変数 x_k の期待値と分散をそれぞれ $E[x_k]$, $V[x_k]$ とし、各説明変数間の相関係数を $\gamma_{kk'}$ とすると、式(6)の v の期待値 $E[v]$ 、分散 $V[v]$ はそれぞれ式(8)、(9)のようになる。

$$v = \beta_0 + \sum_{k=1}^P \beta_k x_k \quad (6)$$

$$u = v + \epsilon_l - \epsilon_m \quad (7)$$

$$E[v] = \beta_0 + \sum_{k=1}^P \beta_k E[x_k] \quad (8)$$

$$\begin{aligned} V[v] &= \sum_{k=1}^P \sum_{k'=1}^P \beta_k \beta_{k'} \text{Cov}[x_k, x_{k'}] \\ &= \sum_{k=1}^P \sum_{k'=1}^P \beta_k \beta_{k'} \gamma_{kk'} \sqrt{V[x_k] V[x_{k'}]} \end{aligned} \quad (9)$$

式(4)の v_i を式(6)のように v と表したので、それに応じて式(5)の p_{il} を式(10)のように $p(v)$ で表現し、このロジット関数 $p(v)$ を選択確率関数とよぶことにする。

$$p(v) = \frac{1}{1 + e^{-v}} \quad (10)$$

式(6)の v を誤解のない範囲で効用値とよび、その確率密度関数を $g(v)$ で表すと、対象とする集団全体における代替案 l の選択比率 P_0 は式(11)のようになる。この式(11)は、効用値 v の確率分布に対する選択確率 $p(v)$

$$\begin{aligned} P_0 &= E[p(v)] \\ &= \int_{-\infty}^{\infty} p(v) g(v) dv \end{aligned} \quad (11)$$

の期待値演算を行っていると考えられる。ここで、選択確率関数として式(10)のロジット関数 $p(v)$ をそのまま用いるのではなく、式(11)の期待値演算の結果が、式(6)の効用値 v の簡単な統計量（例えば、期待値 $E[v]$ や分散 $V[v]$ など）の関数となるようにロジット関数 $p(v)$ を他の関数で置き換えることが考えられるが、このことが平均値法やモーメント法による集計化にはかならない。

(2) 平均値法

平均値法とは、本来式(11)で表される対象とする集団全体での代替案 l の選択比率、すなわち個々の選択行動単位での選択確率の期待値の代わりに、式(12)に示すように効用値 v の期待値 $E[v]$ における式(10)の値 $p(E[v])$ をもって、対象とする集団全体における代替案 l の選択比率とするものである。すなわち、平均値

$$\begin{aligned} P_1 &= p(E[v]) \\ &= \frac{1}{1 + e^{-E[v]}} \end{aligned} \quad (12)$$

法とは、集団全体における選択比率をその集団における「平均的な選択行動単位」での選択確率で代用しようとするものである。このことは、選択確率関数として式(10)のロジット関数の代わりに式(13)で表される $p_1(v)$ ($= \text{const}$)を用いて、式(11)の期待値演算を行ったものと考えられる(式(14))。

$$p_1(v) = p(E[v])$$

$$= \frac{1}{1 + e^{-E[v]}} \quad (13)$$

$$\begin{aligned} E[p_1(v)] &= \int_{-\infty}^{\infty} \frac{1}{1 + e^{-E[v]}} g(v) dv \\ &= \frac{1}{1 + e^{-E[v]}} = P_1 \end{aligned} \quad (14)$$

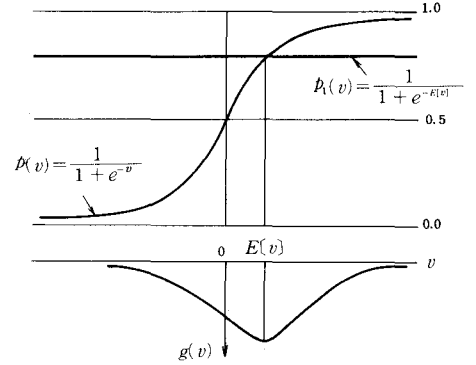


図-1 $p_1(v)$ による近似 (平均値法)

従って、平均値法では、代替案 i の選択確率が対象とする集団のすべての選択行動単位において同じで、その選択確率は効用値 v の期待値 $E[v]$ におけるロジット関数 $p(v)$ の値 $p(E[v])$ に等しいと仮定していることになる。ここで、 $p(v)$ 、 $p_1(v)$ 、 $g(v)$ の関係を図-1に示す。以上のことから、平均値法では、効用値 v の分布特性を表す指標のうち、期待値 $E[v]$ のみが考慮されていることがわかる。

(3) モーメント法

モーメント法は、式(10)の $p(v)$ を式(15)のように $E[v]$ のまわりでTaylor展開して、2次項までで近似し、その近似されたものの期待値をもって式(11)の代わりとするものである。従って、モーメント法とは、式(10)の値をもって対象とする集団全体における代替案 i の選択比率とするものである。^{1), 2)}

$$p(v) \doteq p_2(v)$$

$$= p(E[v]) + (v - E[v]) \left. \frac{dp(v)}{dv} \right|_{v=E[v]} + \frac{1}{2} (v - E[v])^2 \left. \frac{d^2p(v)}{dv^2} \right|_{v=E[v]} \quad (15)$$

$$\begin{aligned} P_2 &= E[p_2(v)] = \int_{-\infty}^{\infty} p_2(v) g(v) dv \\ &= p(E[v]) - V[v] \cdot p(E[v]) \{1 - p(E[v])\} \{2p(E[v]) - 1\} / 2 \end{aligned} \quad (16)$$

式(15)の $p_2(v)$ は効用値 v の2次多項式となっているので、モーメント法は式(10)にかえて、式(17)を選択確率関数として期待値演算を行ったものと考えることができる。式(17)で a 、 b 、 c はそれぞれ式(18)、(19)、(20)に示す値をとる。

$$p_2(v) = av^2 + bv + c \quad (17)$$

$$a = -p(E[v]) \{1 - p(E[v])\} \{2p(E[v]) - 1\} / 2 \quad (18)$$

$$b = p(E[v]) \{1 - p(E[v])\} \{1 + 2p(E[v]) \cdot E[v] - E[v]\} \quad (19)$$

$$\begin{aligned} c &= p(E[v]) - p(E[v]) \{1 - p(E[v])\} E[v] \\ &\quad - p(E[v]) \{1 - p(E[v])\} \{2p(E[v]) - 1\} (E[v])^2 / 2 \end{aligned} \quad (20)$$

また、式(20)より $p(v)$ と $p_2(v)$ は $v = E[v]$ において接しているので、モーメント法では、各選択行動単位

$$\left. \frac{dp(v)}{dv} \right|_{v=E[v]} = \left. \frac{dp_2(v)}{dv} \right|_{v=E[v]} \quad (21)$$

での選択確率は効用値 v の期待値 $E[v]$ の点で $p(v)$ と接するような2次多項式 $p_2(v)$ で表すことができ

ると仮定していることになる。その結果、対象とする集団全体における代替 l の選択比率は式⑭に示すように効用値 v の期待値 $E[v]$ と分散 $V[v]$ の関数として表されることになる。ここで、 $p(v)$ 、 $p_2(v)$ 、 $g(v)$ の関係を図-2に示す。モーメント法は、前述のように、選択確率関数 $p(v)$ を $v = E[v]$ のまわりでTaylor展開し、2次の項までで近似したものであるから、別名Taylor series methodともよばれるが、その意味からすると平均値法は式⑬、⑭からもわかるように選択確率関数 $p(v)$ を0次の項(=const.)で近似したものと考えることができる。

平均値法とモーメント法における選択確率関数と選択比率についてまとめると表-1のようになる。非集計ロ

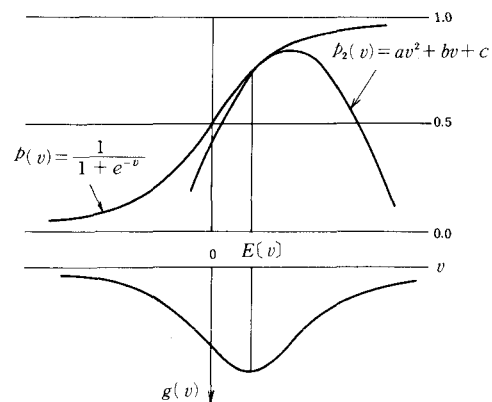


図-2 $p_2(v)$ による近似(モーメント法)

表-1 平均値法とモーメント法における選択確率関数と選択比率

集計方法	選 択 確 率 関 数	選 択 比 率
平均値法	$p_1(v) = \frac{1}{1 + e^{-E[v]}} \text{ (const.)}$	$P_1 = E[p_1(v)]$ $= \frac{1}{1 + e^{-E[v]}}$
モーメント法	$p_2(v) = av^2 + bv + c \text{ (2次多項式)}$ $a = -P_1(1 - P_1)(2P_1 - 1)/2$ $b = P_1(1 - P_1)(1 + 2P_1E[v] - E[v])$ $c = P_1 - P_1(1 - P_1)E[v]$ $- P_1(1 - P_1)(2P_1 - 1)(E[v])^2/2$	$P_2 = E[p_2(v)]$ $= P_1 - V[v]P_1(1 - P_1)$ $\times (2P_1 - 1)/2$

ジットモデルの平均値法とモーメント法による集計バイアスは、以上のことから、式⑭のロジット関数で表すべき選択確率関数を他の関数で置き換えたことにより生じた効用値 v の各点での選択確率の乖離が効用値 v の確率分布($g(v)$)によって重み付けられた結果であるといえることができる。従って、効用値 v の分布特性によっては、モーメント法による選択比率が平均値法による選択比率よりもつねに精度が高くなるわけではなく、とくに式⑭が成立する近似可能な範囲を越えて適用された場合には、著しく精度が低下することもある。このことより、各方法の集計バイアスの大きさについては効用値 v の分布特性との関連でとらえることが極めて重要となる。

3. 効用値の分布特性と集計バイアス

ここでは、平均値法とモーメント法による選択比率の集計バイアスについて、効用値 v の分布特性との関連でその大きさについて具体的に検討を行う。すなわち、平均値法やモーメント法による選択比率の集計バイアスが、効用値のある分布特性においてどのような値をとるのか、また、その集計バイアスは効用値の分布特性の変化に応じてどのように変化するのかなどについて検討を行う。対象とする集団全体における真の選択比率 P_0 、平均値法による選択比率 P_1 、モーメント法に

よる選択比率 P_2 はそれぞれ前述の式(11)、(12)、(16)によって算出される。効用値 v の期待値 $E(v)$ 、分散 $V(v)$ がある値をとるときのこれら P_0 、 P_1 、 P_2 の関係を集計バイアス曲線として整理したものを図-3～図-7に示す。図-3～図-7では効用値 v は正規分布 $N(E(v), V(v))$ に従うものとしている。図-3～図-7に示す集計バイアス曲線より効用値 v の分布特性との関連で平均値法、モーメント法による選択比率の集計バイアスについてつぎのことがいえる。

- a) 集団全体での真の選択比率が0、0.5、1.0の場合には、いずれの方法による選択比率にも理論的には集計バイアスは生じない。
- b) 平均値法による選択比率は、真の選択比率が0.5以上の場合にはより高い値を示し、真の選択比率が0.5以下の場合にはより低い値を示す。

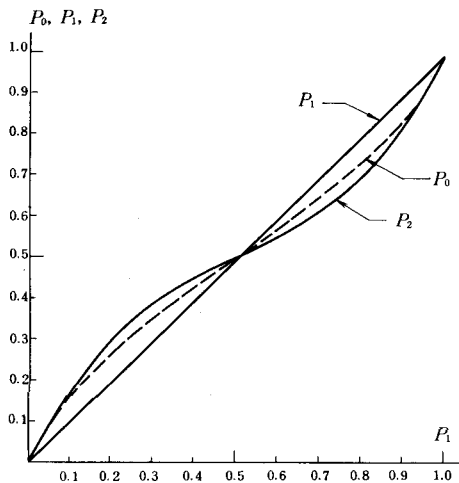


図-3 集計バイアス曲線 ($V(v)=2.0$)

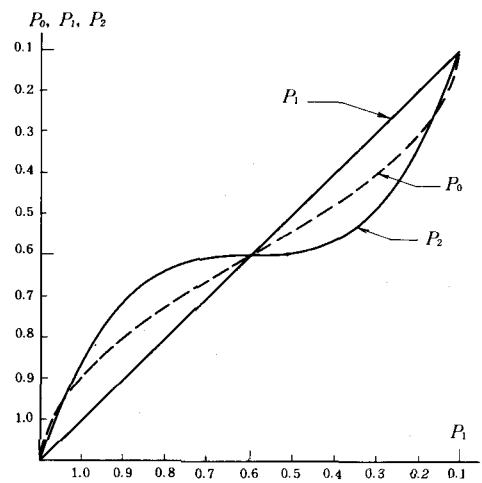


図-4 集計バイアス曲線 ($V(v)=4.0$)

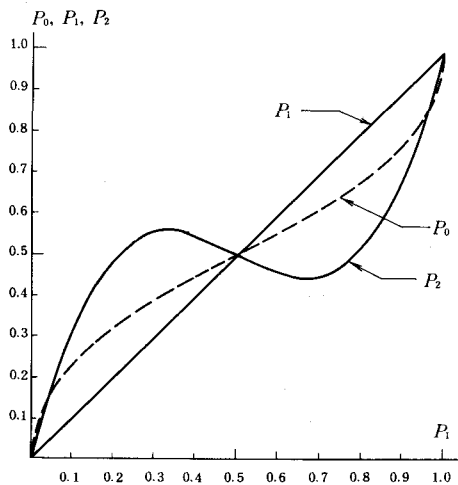


図-5 集計バイアス曲線 ($V(v)=6.0$)

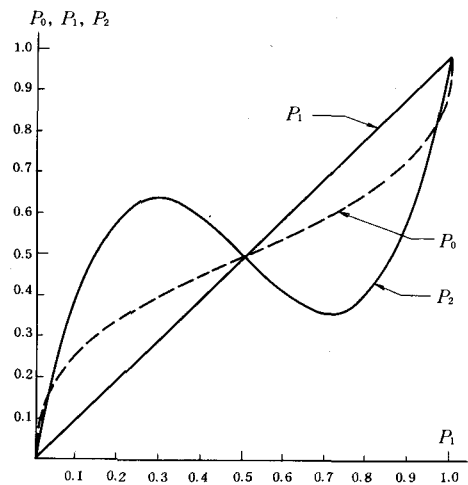


図-6 集計バイアス曲線 ($V(v)=8.0$)

c) 効用値 v の期待値 $E[v]$ が同じ場合でも分散 $V[v]$ が大きくなるに従って集計バイアスは大きくなり、また分散 $V[v]$ が同じでも期待値 $E[v]$ の大きさによって集計バイアスは異なる。さらに、平均値法、モーメント法ともに集計バイアスが最大となる $E[v]$ は分散 $V[v]$ の大きさによって異なる。

d) 平均値法、モーメント法による選択比率はともに効用値 v の分散 $V[v]$ が大きい場合には著しく大きな集計バイアスをもつ。たとえば、 $V[v]=10.0$ の場合、 P_0 が 0.65 のときには、 $P_1=0.80$, $P_2=0.32$ であり、また P_0 が 0.59 のときには、 $P_1=0.70$, $P_2=0.28$ である。とくに、モーメント法による選択比率は、効用値の分散 $V[v]$ が大きい場合には、式 (15) が成立する近似可能な範囲を越えて式 (16) の期待値演算を行うことになるために著しく大きな集計バイアスをもつことになる。従って、モーメント法の適用は式 (15) が成立するような範囲に限定されるべきであろう。

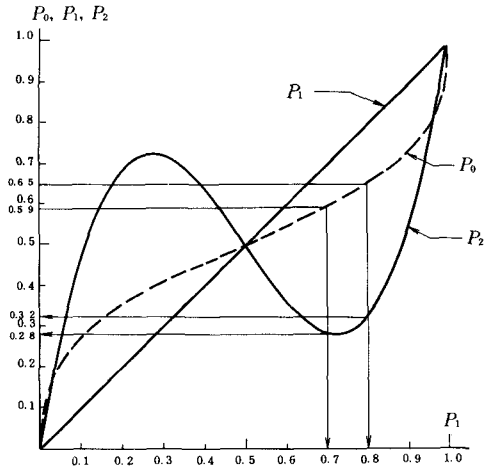


図-7 集計バイアス曲線 ($V[v]=10.0$)

4. 選択比率の推定精度とデータサイズ

前章までにおいて、対象とする集団全体における選択比率を表す 3 つの方法 P_0, P_1, P_2 について説明するとともに、 P_1, P_2 による選択比率の集計バイアスについて効用値の分布特性との関連で検討を行ったが、実際には非集計ロジットモデルのパラメータを推定したのちにそれぞれ以下の推定式に従って具体的な選択比率を推定することになる。

a) $P_0=E[p(v)]$ の推定:

$$\bar{P}_0 = \overline{p(\bar{v})} = \frac{1}{n} \sum_{i=1}^n p(\bar{v}_i) = \frac{1}{n} \sum_{i=1}^n \frac{1}{1 + e^{-\bar{v}_i}} \quad (22)$$

b) $P_1=E[p_1(v)]$ の推定:

$$\bar{P}_1 = p(\bar{v}) = p\left(\frac{1}{n} \sum_{i=1}^n \bar{v}_i\right) = \frac{1}{1 + e^{-\frac{1}{n} \sum_{i=1}^n \bar{v}_i}} \quad (23)$$

c) $P_2=E[p_2(v)]$ の推定:

$$\bar{P}_2 = \bar{P}_1 - \frac{1}{2} \bar{V}[v] \bar{P}_1 (1 - \bar{P}_1) (2 \bar{P}_1 - 1) \quad (24)$$

式 (22)、(23) において、 $\bar{v}_i$ は効用値 v_i の推定値で式 (25) によって表される。式 (22) は集計方法のうちの実サンプル法と

$$\bar{v}_i = \hat{\beta}_0 + \sum_{k=1}^p \hat{\beta}_k x_{ik} \quad (25)$$

よばれるものに等しいが、式 (22) ~ (24) によって求められる選択比率の推定値が $\bar{v}_i$ あるいはその平均値によって規定される以上、これらの式より求められる選択比率の推定精度はデータサイズによって影響を受ける。そこで、選択比率の推定精度とデータサイズとの関係について把握することが必要となるが、本研究で

表-3 $E[x_k], V[x_k]$ の設定値

	CASE A		CASE B	
	$E[x_k]$	$V[x_k]$	$E[x_k]$	$V[x_k]$
x_1	0	5	1	5
x_2	0	5	1	5
x_3	0	10	1	10
x_4	0	15	1	15
x_5	0	20	1	20

表-2
パラメータ
の設定値

β_0	0.0
β_1	0.1
β_2	0.2
β_3	0.3
β_4	0.4
β_5	0.5

$E[v]$	0.0	1.5
$V[v]$	8.550	8.550
P_0	0.500	0.670
P_1	0.500	0.818
P_2	0.500	0.413

は、効用値の分布特性をあらかじめ設定し、シミュレーションにより分布特性に対応した仮想観測値データの作成を行い、その作成された仮想観測値データをもとに検討を行う。なお、シミュレーションによる仮想観測値データの作成方法は文献3) によっている。ここでは、モデルに取り入れられた説明変数は x_1, x_2, x_3, x_4, x_5 の5変数とし、パラメータについては、表-2に示す値をパラメータの真値として設定した。モデルに取り入れられた各説明変数 x_k ($k=1, 2, \dots, 5$) はそれぞれ独立に正規分布 $N(E[x_k], V[x_k])$ に従うものとしている。各説明変数 x_k の設定した期待値 $E[x_k]$ 、分散 $V[x_k]$ を表-3に示す。表-3には、設定値により算出される $E[v]$ 、 $V[v]$ P_0, P_1, P_2 の値も合わせて示している。

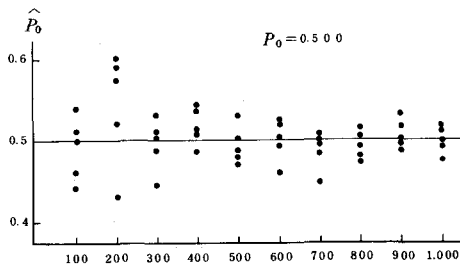


図-8 選択比率の推定結果 (CASE A の $\hat{P}_0$)

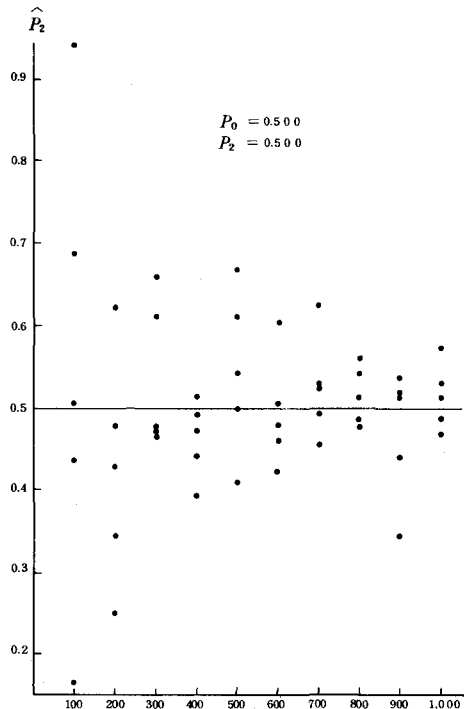


図-10 選択比率の推定結果 (CASE A の $\hat{P}_2$)

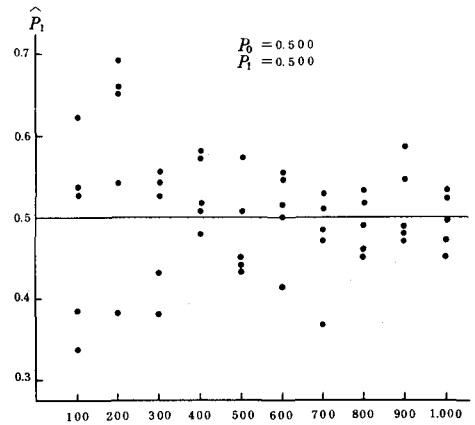


図-9 選択比率の推定結果 (CASE A の $\hat{P}_1$)

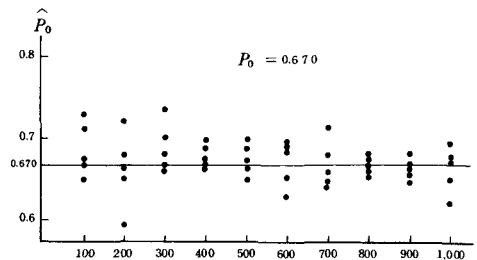


図-11 選択比率の推定結果 (CASE B の $\hat{P}_0$)

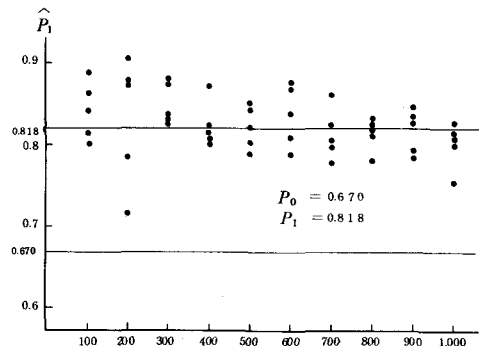


図-12 選択比率の推定結果 (CASE B の $\hat{P}_1$)

選択比率の3つの推定方法 $\hat{P}_0, \hat{P}_1, \hat{P}_2$ による推定結果をもとに、CASEごと推定精度について比較すると、CASE A (図-8、図-9、図-10) では、いずれの方法においてもデータサイズが大きくなるに従って、選択比率の推定精度は高くなるが、 $\hat{P}_1, \hat{P}_2$ ではデータサイズが小さい場合には著しく安定性に欠けていることがわかる。CASE B (図-11、図-12、図-13) では、 P_0 による推定値の精度はデータサイズが大きくなるに従って幾分高くなるが、 $\hat{P}_1, \hat{P}_2$ による推定値の精度は集計バイアスのためデータサイズを大きくしても必ずしもよくならない。

5. まとめ

本研究では、非集計ロジットモデルを交通需要の予測や交通政策の評価のために用いる場合に必要となる集計化の方法について比較的適用が簡単とされている平均値法とモーメント法の2つの方法を取りあげ、解析的な意味について考察するとともに集計バイアスの大きさについて検討を行った。本研究における結論を要約すると以下のようになる。

- 平均値法、モーメント法による集計化は、選択確率関数であるロジット関数を他の関数で置き換えたうえで期待値演算を行うことに等しく、代用される関数は、平均値法では定数、モーメント法では2次多項式である。いわゆる集計バイアスはこの関数の代用により生じた効用値の各点での選択確率の乖離が効用値の確率分布によって重み付けられた結果である。
- 集団全体での真の選択比率が0、0.5、1.0の場合には、平均値法、モーメント法のいずれの方法による選択比率にも理論的には集計バイアスは生じない。平均値法による選択比率は、真の選択比率が0.5以上の場合にはより高い値を示し、真の選択比率が0.5以下の場合にはより低い値を示す。いわゆる集計バイアスは、効用値の分布特性によって異なり、平均値法、モーメント法ともに効用値の分散が大きくなるに従って集計バイアスは大きくなるが、この傾向はモーメント法においてとくに著しい。これは、モーメント法が式(4)が成立するような近似可能な範囲を越えて式(5)の期待値演算を行っていることによる。
- 効用値の分布特性から集計バイアスがとくに問題とならない場合には、データサイズの増加に従って選択比率の推定精度は高くなるが、データサイズが小さい場合のモーメント法による推定精度は著しく低い。理論的に集計バイアスが存在する場合には、データサイズを大きくしても集計バイアスは解消されない。

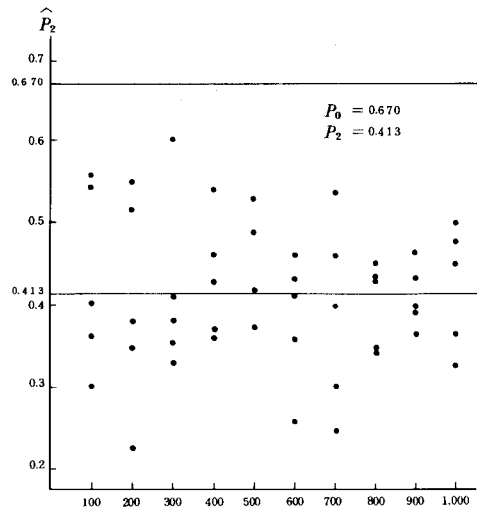


図-13 選択比率の推定結果
(CASE Bの $\hat{P}_2$)

参考文献

- Talvitie: "Aggregate travel demand analysis with disaggregate or aggregate travel demand models", Transportation Research Forum Proceeding, pp. 583~601, 1973
- 財) 運輸経済研究センター: "大都市圏における交通機関選択分析調査報告書", pp. 194~202, 1978
- 桐越 信、浅野光行: "都市交通計画における交通需要予測モデルの評価方法", 建築研究報告、第103号, pp. 50~57, 建設省建築研究所、1983
- 桐越 信、塚本直幸: "シミュレーションデータによる非集計ロジットモデルの誤差解析", 第5回土木計画学研究発表会講演集, pp. 450~461, 1983