

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

建設省建築研究所 正員 ○ 堀 越 信
(佛長大橋設計センター 正員 塚 本 直 幸

1. はじめに

都市交通計画の分野において、交通需要の予測モデルあるいは交通政策の評価モデルとしての非集計モデルの理論的および実証的研究が近年盛んに進められている。この非集計モデルは、個々の意思決定単位における選択行動をゾーンごとに集計せず、意思決定単位レベルのデータをそのままモデル構築の際のデータとするもので、意思決定単位の実選択行動メカニズムを明示的にモデル化できるところにその特徴がある。非集計モデルは、交通機関選択だけでなく住宅地選択や職業選択などのいわゆる選択行動問題に広く適用されている^{1,2)}。非集計モデルの使用目的は種々考えられるが、このモデルを交通需要の予測や交通政策の評価のためのモデルとして用いる場合には、パラメータ推定値や選択比率の推定値の安定性の問題が重要な問題となる。そこで、本研究では、非集計モデルのうち最も適用例の多いロジットモデルをとりあげ、パラメータ推定値や選択比率の推定値の安定性の問題について、効用値の分布特性とデータサイズとの関連において検討を行う。すなわち、効用値のある分布特性において安定したパラメータ推定値や選択比率の推定値を得るためにはどの程度のデータサイズが必要なのか、またその必要なデータサイズは効用値の分布特性に応じてどのように変化するのかなどについて、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仮想観測値データを作成することによって検討を行い、今後の非集計ロジットモデルの構築に資することを目的とする。

本研究では、図-1に示す手順に従って検討を行う。はじめに、効用値の分布特性とデータサイズを設定することにより、効用値の分布特性に対応した仮想観測値データを作成し、このデータをもとに、効用値の分布特性とデータサイズがパラメータ推定値の安定性にどのような影響を与えるかを調べ、つぎに効用値の分布特性とデータサイズ、さらに集計化の方法が選択比率の推定値の安定性にどのような影響を与えるのかを調べる。

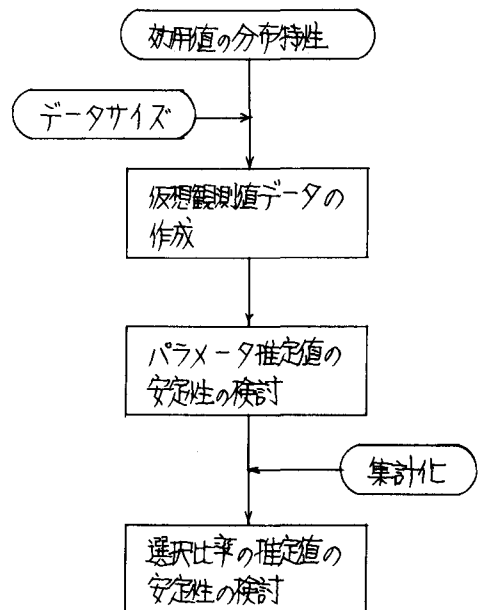


図-1 研究のフロー

2. 仮想観測値データの作成

2.1 効用値の分布特性の表現方法

いま、2つの選択可能な代替案 l, m がある場合に、個人 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) で表わされるものとする。式(1), (2)において、 U_{il}, U_{im} は代替案の特性や個人 i の社会経済的屬性

$$U_{il} = U_{il} + \varepsilon_{il} = \beta_{0l} + \sum_{k=1}^p \beta_k X_{ilk} + \varepsilon_{il} \quad (1)$$

$$U_{im} = U_{im} + \varepsilon_{im} = \beta_{0m} + \sum_{k=1}^p \beta_k X_{imk} + \varepsilon_{im} \quad (2)$$

などから構成される部分で、それぞれ変数 X_{ilk}, X_{imk} の一次結合で表わされるものとする。また、 $\varepsilon_{il}, \varepsilon_{im}$ は誤差項として確率変動する部分で、それぞれ互いに独立に式(3)で表現される確率密度関数 $f(\varepsilon)$ をもつガンベ

1/分布に従うものとする。

$$f(\varepsilon) = \text{EXP}(-\varepsilon) \cdot \text{EXP}\{-\text{EXP}(-\varepsilon)\} \quad (3)$$

式(1)から式(2)を引いて、 $U_i = U_{ie} - U_{im}$, $U_i = U_{ie} - U_{im}$, $\beta_0 = \beta_{0e} - \beta_{0m}$, $\chi_{ik} = \chi_{iek} - \chi_{imk}$ とおく
と式(4)が得られる。

$$U_i = U_i + \varepsilon_{ie} - \varepsilon_{im} = \beta_0 + \sum_{k=1}^p \beta_k \chi_{ik} + \varepsilon_{ie} - \varepsilon_{im} \quad (4)$$

この場合に、個人*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)で表現される。

$$\begin{aligned} P_{il} &= \text{Prob}[U_{ie} > U_{im}] \\ &= \text{Prob}[U_i > 0] = \frac{1}{1 + e^{-U_i}} \end{aligned} \quad (5)$$

ここで、 U_i は連続的に変化するものとして効用値 U_i を式(6)で表わし、その確率密度関数を $g(U_i)$ 、各説明変数 χ_{ik} の期待値と分散をそれぞれ $E[\chi_{ik}]$, $V[\chi_{ik}]$ とし、各説明変数間の相関係数を $r_{kk'}$ とすると、効用値 U_i の期待値 $E[U_i]$, U_i の期待値 $E[U_i^2]$, 対象とする集団全体の代替案*l*の選択比率 $E[p(w)]$ は、それぞれ式(7), (8), (9)で表現される。

$$U_i = \beta_0 + \sum_{k=1}^p \beta_k \chi_{ik} \quad (6)$$

$$E[U_i] = \beta_0 + \sum_{k=1}^p \beta_k E[\chi_{ik}] \quad (7)$$

$$\begin{aligned} E[U_i^2] &= \beta_0^2 + 2\beta_0 \sum_{k=1}^p \beta_k E[\chi_{ik}] + \sum_{k=1}^p \sum_{k'=1}^p \beta_k \beta_{k'} \{ \text{Cov}[\chi_{ik}, \chi_{ik'}] + E[\chi_{ik}] E[\chi_{ik'}] \} \\ &= \beta_0^2 + 2\beta_0 \sum_{k=1}^p \beta_k E[\chi_{ik}] + \sum_{k=1}^p \sum_{k'=1}^p \beta_k \beta_{k'} \{ r_{kk'} \sqrt{V[\chi_{ik}] V[\chi_{ik'}]} + E[\chi_{ik}] E[\chi_{ik'}] \} \end{aligned} \quad (8)$$

$$E[p(w)] = \int_{-\infty}^{\infty} p(w) g(w) dw = \int_{-\infty}^{\infty} \frac{g(w)}{1 + e^{-w}} dw \quad (9)$$

ここで、 $E[p(w)]$ に代えて、効用値 U_i の期待値 $E[U_i]$ をもとに、対象とする集団全体の代替案*l*の選択比率 $P(E[U_i])$ を求めると式(10)のようになる。

$$P(E[U_i]) = \frac{1}{1 + e^{-E[U_i]}} \quad (10)$$

$E[\chi_{ik}]$, $V[\chi_{ik}]$ と $U_i = \beta_0 + \sum_{k=1}^p \beta_k \chi_{ik}$ によって規定される分布特性を効用値 U_i が持つ場合に、式(4)における U_i と U_i が同符号である確率を S とすると、 S は式(11)で表現される。式(11)において、 $\varphi(\varepsilon^*)$ は2つの誤差

$$\begin{aligned} S &= \text{Prob}(U_i > 0 \cap U_i > 0) + \text{Prob}(U_i \leq 0 \cap U_i \leq 0) \\ &= \int_0^{\infty} \left[\int_{-\infty}^{\infty} \varphi(\varepsilon^*) d\varepsilon^* \right] g(w) dw + \int_{-\infty}^0 \left[\int_{-\infty}^{-w} \varphi(\varepsilon^*) d\varepsilon^* \right] g(w) dw \\ &= \int_{-\infty}^0 \frac{g(w) + g(-w)}{1 + e^w} dw \end{aligned} \quad (11)$$

項 ε_{ie} , ε_{im} の差 $\varepsilon^* = \varepsilon_{ie} - \varepsilon_{im}$ の確率密度関数を表わしている。

以上の $E[U_i]$, $E[U_i^2]$, $E[p(w)]$, $P(E[U_i])$, S は設定した効用値の分布特性を表現するものと考えることができ。

2.2 仮想観測値データの作成方法³⁾

仮想観測値データの作成は図-2の手順に従って行う⁴⁾。はじめに、式(6)の U_i を構成するパラメータ β_0 , β_k と説明変数 χ_{ik} の期待値 $E[\chi_{ik}]$, 分散 $V[\chi_{ik}]$ および説明変数間の相関係数 $r_{kk'}$ を設定する。つぎに、 $E[\chi_{ik}]$, $V[\chi_{ik}]$, $r_{kk'}$ の設定値を母数としてもつような説明変数群をデータサイズを指定してシミュレーションにより作成する。さらに、2つの代替案*l*, m の効用 U_{ie} , U_{im} のうち誤差項として確率変動する部分 ε_{ie} , ε_{im} を確率密度関数 $f(\varepsilon)$ が式(3)で表わされるガンベル乱数を発生させることによって作成する。これら χ_{iek} , ε_{ie} , ε_{im} を式(4)に代入して U_i を求め、 U_i が正($U_{ie} > U_{im}$)なら仮想観測値 P_{il}^* を1(P_{im}^* を0)とおき、 U_i が正で

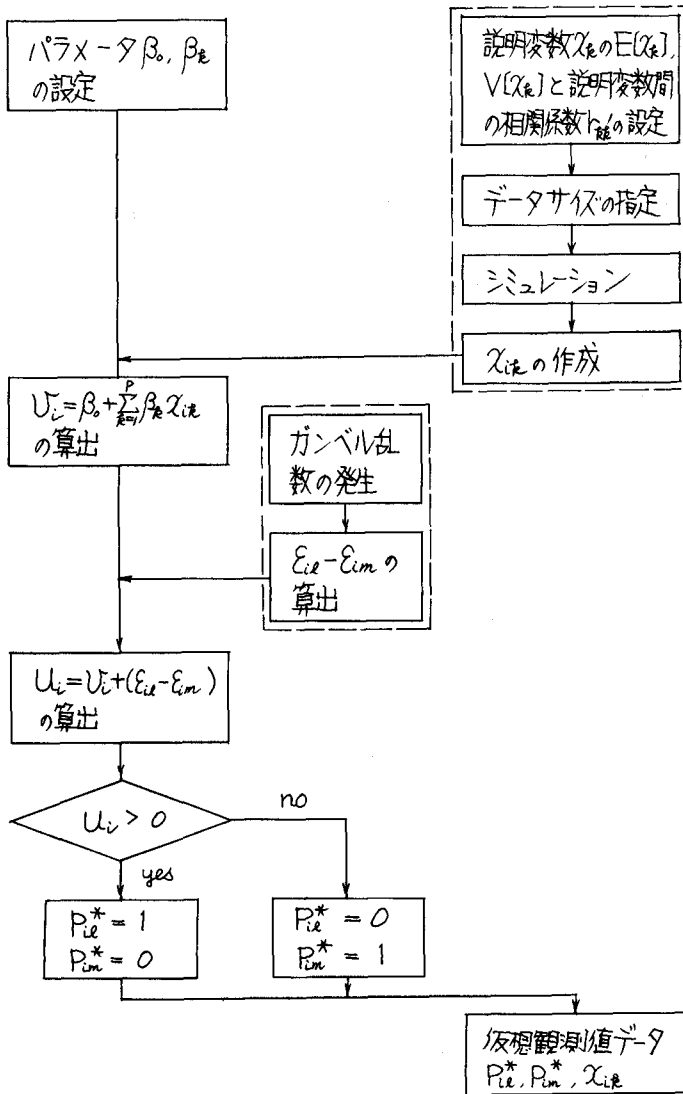


図-2 仮観測値データの作成方法

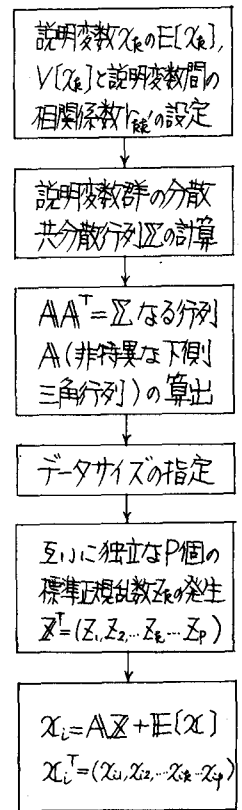


図-3 説明変数群の作成方法

なり ($U_{ie} \leq U_{im}$) なり仮観測値 P_{ie}^* を 0 (P_{im}^* を 1) とおいて、仮観測値データ $P_{ie}^*, P_{im}^*, x_{ik}$ を作成する。仮観測値データを作成するにあたって必要となる説明変数群の作成方法とガンベル乱数の発生方法を以下に説明する。

$E[x_{ik}]$, $V[x_{ik}]$, $r_{kk'}$ によって規定される構造をもつ説明変数群の作成は図-3に示す手順に従って行う⁵⁾。ただし、ここでは説明変数群は多重正規分布に従うものとする。はじめに、 $E[x_{ik}]$, $V[x_{ik}]$, $r_{kk'}$ より説明変数群の分散共分散行列 Σ の計算を行い、この分散共分散行列 Σ を用いて式(12)を満たす非特異な下側三角行列 A を求める。

$$AA^T = \Sigma \quad (12)$$

つぎに、互いに独立な P 個の標準正規乱数 z_k ($k=1, 2, \dots, P$) を発生させることにより、式(13)によって説明

変数群を算出する。

$$X_i = A_i Z + E(X_i) \quad (13)$$

ここで、

$$X_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ik} \\ \vdots \\ x_{ip} \end{pmatrix}, \quad Z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_k \\ \vdots \\ z_p \end{pmatrix}, \quad E(X_i) = \begin{pmatrix} E(x_1) \\ E(x_2) \\ \vdots \\ E(x_k) \\ \vdots \\ E(x_p) \end{pmatrix}$$

確率変動する誤差項を作成するためのガンベル乱数は、ガンベル分布の確率密度関数 $f(\varepsilon)$ が式(3)で表われるので、その分布関数 $F(\varepsilon)$ を式(14)によって求め、この式(14)を式(15)のように変形し、その変形された式(15)の $F(\varepsilon)$ に $[0, 1]$ の一様乱数を代入することによって発生させる。

$$F(\varepsilon) = \int_0^\varepsilon f(\varepsilon) d\varepsilon = \exp\{-\exp(-\varepsilon)\} \quad (14)$$

$$\varepsilon = -\ln\{-\ln[F(\varepsilon)]\} \quad (15)$$

2.3 作成した仮観測値データ

ここでは、モデルに取り入れられた説明変数を 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-1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、それぞれ正規分布 $N(E(x_k), V(x_k))$ に従うものとしている。設定した各説明変数 x_k の期待値 $E(x_k)$ 、分散 $V(x_k)$ と各説明変数間の相関係数の値を表-2に示す。表-2における設定値をもとに、式(7)、(8)、(9)、(10)、(11)に従って算出した $E(u)$ 、 $E(u^2)$ 、 $E(pu)$ 、 $P(E(u))$ 、 S の値を表-3に示す。表-2において、CASE I、CASE II、CASE IIIでは各説明変数 x_k は互いに独立とし、 $r_{kk'} = 0.0$ とされている。CASE IVでは説明変数間に強い相関関係がある場合のパラメータ推定値や選別比率の推定値の安定性がどのように変化するのかを把握することをねらいとして $r_{kk'} = 0.8$ としている。また、CASE V、CASE VIでは、 $E(x_k)$ の変化がパラメータ推定値や選別比率の推定値の安定性にどのように影響するかを把握することをねらいとしてCASE IIIをもとに $E(x_k)$ を変化させている。説明変数の分散 $V(x_k)$ が大きくなるに従って、また説明変数間の相関係数 $r_{kk'}$ が大きくなるに従って、式(8)の $E(u^2)$ が大きくなるので、表-3のCASE I ~ CASE IVについては、式(11)の S の値がづぎのようになっている。

ねらいとして $r_{kk'} = 0.8$ としている。また、

CASE V、CASE VIでは、 $E(x_k)$ の変化がパラメータ推定値や選別比率の推定値の安定性にどのように影響するかを把握することをねらいとしてCASE IIIをもとに $E(x_k)$ を変化させている。説明変数の分散 $V(x_k)$ が大きくなるに従って、また説明変数間の相関係数 $r_{kk'}$ が大きくなるに従って、式(8)の $E(u^2)$ が大きくなるので、表-3のCASE I ~ CASE IVについては、式(11)の S の値がづぎのようになっている。

$$S_I < S_{II} < S_{III} < S_{IV} \quad (16)$$

また、式(7)の $E(u)$ が0から離れることも、 $E(u^2)$ を大きくするので、表-3のCASE III、CASE V、CASE VIについては S の値がづぎのよう

表-1 パラメータの設定値

パラメータ	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

表-2 $E(x_k)$ 、 $V(x_k)$ 、 $r_{kk'}$ の設定値

	CASE I	CASE II	CASE III	CASE IV	CASE V	CASE VI
$E(x_1)$	0	0	0	0	1	1
$V(x_1)$	0.5	1	5	5	5	5
$E(x_2)$	0	0	0	0	1	1
$V(x_2)$	0.5	1	5	5	5	5
$E(x_3)$	0	0	0	0	1	1
$V(x_3)$	0.5	1	10	10	10	10
$E(x_4)$	0	0	0	0	1	1
$V(x_4)$	0.5	1	15	15	15	15
$E(x_5)$	0	0	0	0	1	2
$V(x_5)$	0.5	1	20	20	20	20
$r_{kk'}$	0.0	0.0	0.0	0.8	0.0	0.0

なっている。

$$S_{II} < S_V < S_{VI} \quad (17)$$

仮想観測値データは、設定したCASEごとに、データサイズを100から1000まで100ずつ増加させ、各々のデータサイズについて5組作成している。従って、ひとつのCASEについて50組の仮想観測値データを作成していることになる。

表-3 効用値の分布特性

	CASE I	CASE II	CASE III	CASE IV	CASE V	CASE VI
$E(u)$	0.0	0.0	0.0	0.0	1.5	2.0
$E(u^2)$	0.275	0.550	8.550	25.079	10.800	12.550
$E(p(u))$	0.500	0.500	0.500	0.500	0.670	0.721
$P(E(u))$	0.500	0.500	0.500	0.500	0.818	0.881
S	0.600	0.637	0.833	0.895	0.849	0.861

3. パラメータ推定値の安定性

前章で作成した仮想観測値データに対するパラメータの推定結果について以下に説明する。パラメータの推定は最尤法によって行っている。パラメータの推定結果の一部を図-4～図-13に示す。図-4～図-8はCASE IIIにおける推定結果 $\hat{\beta}_k$ をパラメータごとに示したものであり、図-9、図-10、図-11、図-12、図-13はそれぞれCASE I, CASE II, CASE IV, CASE V, CASE VIのパラメータの推定結果 $\hat{\beta}_k$ を示したものである。

図-4～図-13によると、一般にデータサイズが小さい場合には、得られた推定値と表-1で設定したパラメータの真値との乖離が大きく、推定結果のバラツキも大きい。データサイズが大きくなるに従って、推定値が設定したパラメータの真値に近くなり、バラツキも小さくなる。図-4、図-9～図-13によりパラメータ β_1 の推定結果 $\hat{\beta}_1$ について比較すると、各説明変数間に相関関係がない場合には、表-3のSの値の高いCASEほど小さなサンプルサイズで安定することゝわかる。ただ、CASE III, CASE V, CASE VI, 2はSの値にあまり差がなかったため同様な安定性になっている。また、CASE III (図-4)とCASE IV (図-11)とを比較すると、CASE IVのほうがSの値は高いが、パラメータ推定値の安定性は多重共線性により著しく低下している。

ここで、式(18)で定義される相対誤差 $S(\hat{\beta}_k)$ の計算結果の一部(パラメータ β_4 の推定における $S(\hat{\beta}_4)$)を図-14、図-15に示す。

$$S(\hat{\beta}_k) = \frac{\sqrt{\frac{5}{\sum_{l=1}^5 (\hat{\beta}_{k,l} - \beta_k)^2} / 5}}{\beta_k} \quad (18)$$

ここで、 l は各データサイズにおけるパラメータの推定番号、 $l=1, 2, \dots, 5$

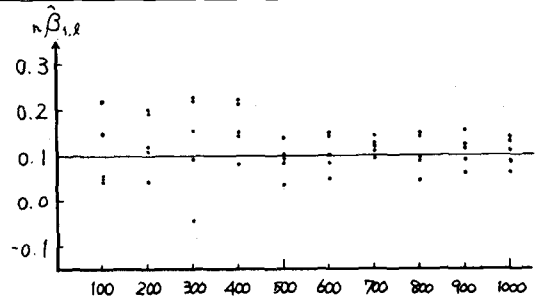


図-4 CASE IIIの $\hat{\beta}_1$

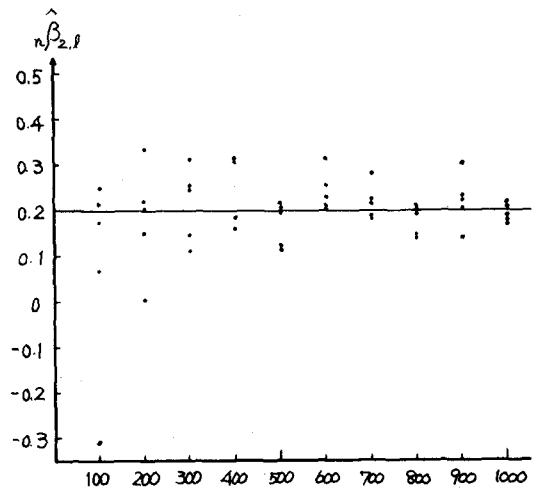


図-5 CASE IIIの $\hat{\beta}_2$

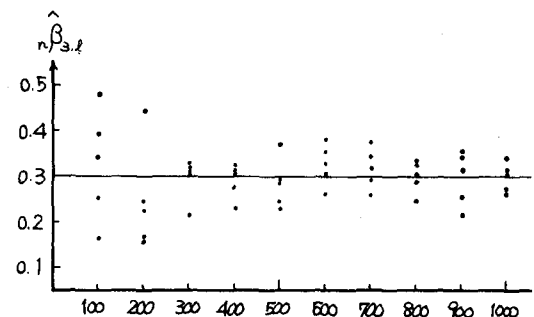
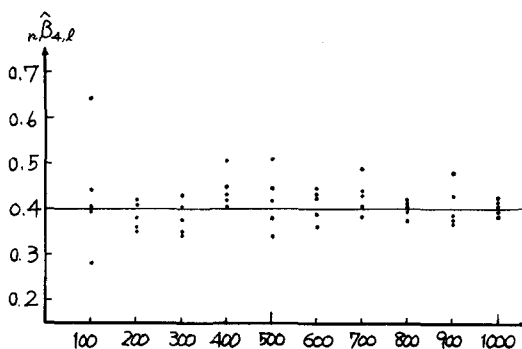
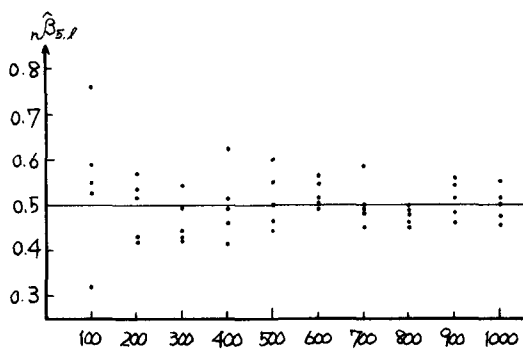


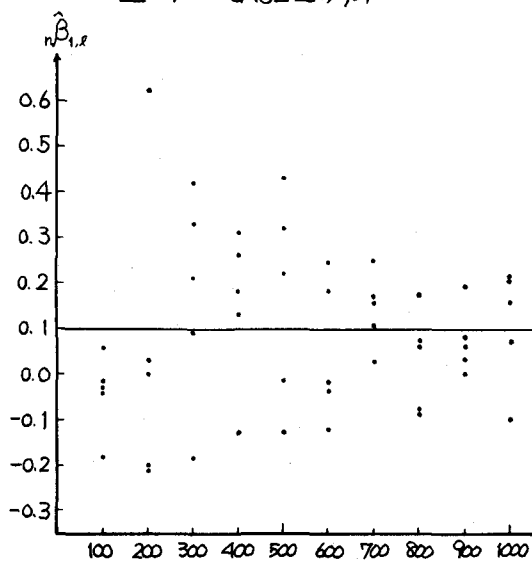
図-6 CASE IIIの $\hat{\beta}_3$



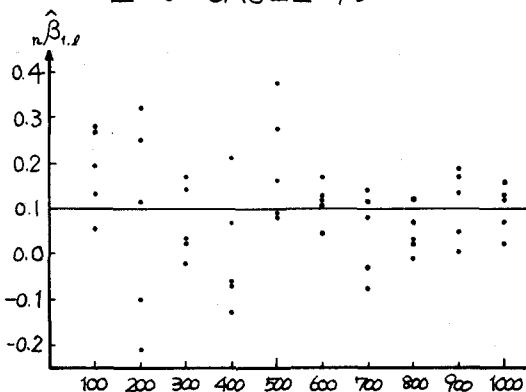
☒-7 CASE III の $\hat{\beta}_4$



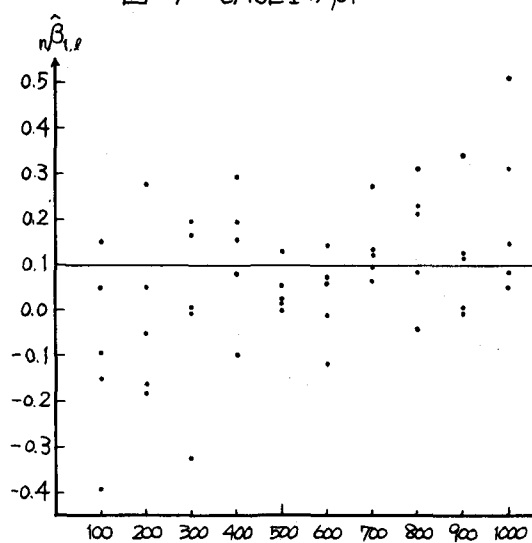
☒-8 CASE III の $\hat{\beta}_5$



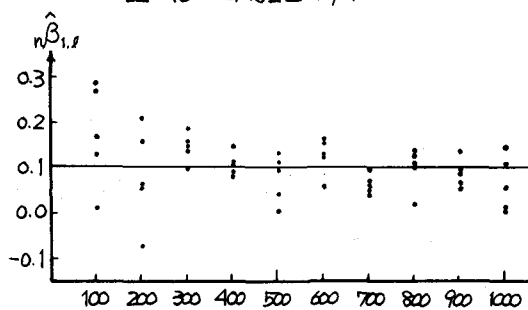
☒-9 CASE I の $\hat{\beta}_1$



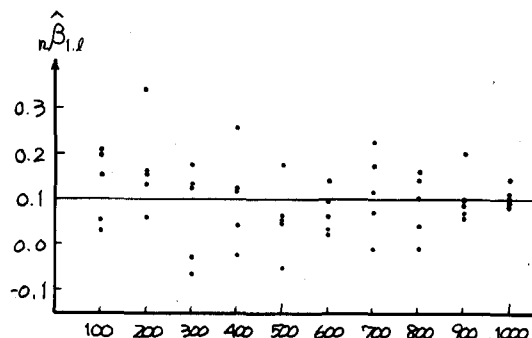
☒-10 CASE II の $\hat{\beta}_1$



☒-11 CASE IV の $\hat{\beta}_1$



☒-12 CASE V の $\hat{\beta}_1$



☒-13 CASE VI の $\hat{\beta}_1$

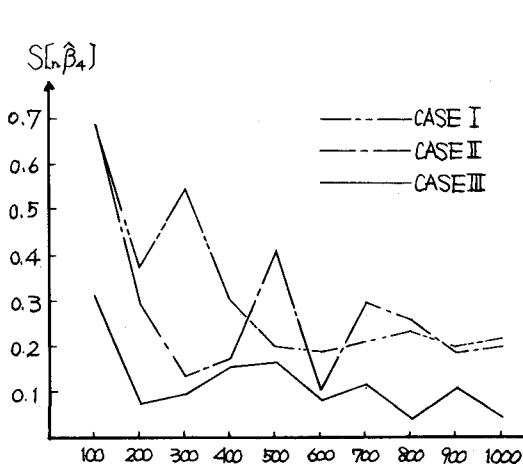


図-14 データサイズと $S[n, \hat{\beta}_4]$ -(1)-

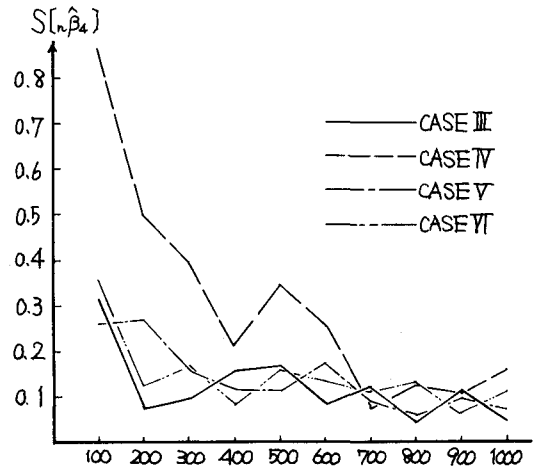


図-15 データサイズと $S[n, \hat{\beta}_4]$ -(2)-

n はデータサイズ, $n=100, 200, \dots,$

1000

$\hat{\beta}_{k, n}$ はデータサイズ n , パラメータの推定番号 k のときのパラメータ β_k の推定値

図-14, 図-15によると, どのCASEでもデータサイズが大きくなるに従って, 相対誤差 $S[n, \hat{\beta}_4]$ は小さくなる傾向にあるが, CASE I, CASE II, CASE IVではデータサイズが小さいときにはかなり大きな値をとることがわかる。CASE V, CASE VIはデータサイズの大きさに応じてCASE IIIと同じような変化をするが, CASE IVでは強い多重共線性のためにほとんどのデータサイズでCASE IIIよりも大きな相対誤差となっている。

ここで, 各パラメータごとに, あるデータサイズ以上の場合には相対誤差 $S[n, \hat{\beta}_k]$ がある水準以下になるデータサイズに注目して整理すると表-4~表-8のようになる。 $\hat{\beta}_5$ を例として, 表-4~表-8について説明すると, $\hat{\beta}_5$ を相対誤差 $(S[n, \hat{\beta}_5])0.1$ 以下で推定するためには, CASE IIで1000以上, CASE IIIで600以上, CASE Vで500以上, CASE VIで800以上のデータサイズが必要となることがわかる。CASE I, CASE IVではデータサイズを1000程度としても相対誤差0.1以下で推定することは困難である。相対誤差0.2以下とすると, CASE IIで800以上, CASE IIIで200以上,

表-4 $S[n, \hat{\beta}_k] \leq 0.1$ の場合

	CASE I	CASE II	CASE III	CASE IV	CASE V	CASE VI
$\hat{\beta}_1$	—	—	—	—	—	—
$\hat{\beta}_2$	—	—	—	—	—	—
$\hat{\beta}_3$	1000	1000	1000	—	—	1000
$\hat{\beta}_4$	—	—	1000	—	700	—
$\hat{\beta}_5$	—	1000	600	—	500	800

表-5 $S[n, \hat{\beta}_k] \leq 0.2$ の場合

	CASE I	CASE II	CASE III	CASE IV	CASE V	CASE VI
$\hat{\beta}_1$	—	—	—	—	—	—
$\hat{\beta}_2$	—	—	1000	—	—	—
$\hat{\beta}_3$	1000	1000	500	—	700	700
$\hat{\beta}_4$	—	900	200	700	300	200
$\hat{\beta}_5$	—	800	200	700	300	300

表-6 $S[n, \hat{\beta}_k] \leq 0.3$ の場合

	CASE I	CASE II	CASE III	CASE IV	CASE V	CASE VI
$\hat{\beta}_1$	—	—	1000	—	—	1000
$\hat{\beta}_2$	—	—	1000	—	700	700
$\hat{\beta}_3$	1000	1000	300	1000	300	100
$\hat{\beta}_4$	500	600	200	600	100	200
$\hat{\beta}_5$	600	100	100	200	100	100

CASE IVで700以上, CASE Vで300以上, CASE VIで300以上となり, CASE Iを除くどのCASEでも必要データサイズは相対誤差0.1以下のときよりも小さくなっている。相対誤差0.3以下とするとCASE Iで600以上, CASE II, CASE IIIで100程度, CASE IVで200以上, CASE V, CASE VIで100程度となり, CASE II~CASE VIではかなり小さなデータサイズで推定が可能となる。相対誤差0.4以下としてもこの傾向は変わらないが, 相対誤差0.5以下とするとCASE Iでは600以上から100程度へと必要とするデータサイズはかなり減少する。

説明変数間に相関関係がないCASE IV以外のCASEでは, 一般にSの値の大きいCASEほど小さなデータサイズで安定的なパラメータ推定値を得ることが出来る。ただし, 100, 200程度のデータサイズで安定的なパラメータ推定が可能なのは, 許容される相対誤差の水準が高く, モデル

の適合度がよい場合に限られることがわることがわかる。CASE IIIとCASE IVとを比較すると, 説明変数間に相関関係のあるCASE IVのほうがより大きなデータサイズを必要とすることもわかる。

4. 選択比率の推定値の安定性

各CASEごとの選択比率の推定値を図-16~図-27に示す。図-16~図-21は効用関数のパラメータを推定したのちに, 効用値の平均値 $\bar{u}$ を求め, その平均値 $\bar{u}$ における推定確率 $P(\bar{u})$ によって選択比率を求めたものである。一方, 図-22~図-27は効用値 $\bar{u}$ ごとの推定確率を算出し, これらの平均値 $\bar{P}(\bar{u})$ を求めることにより, 選択確率としたものである。各CASEごとに $P(\bar{u})$ と $\bar{P}(\bar{u})$ について比較すると, CASE I, CASE IIでは同じような安定性を示すが, CASE III, CASE IVでは $P(\bar{u})$ の値が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 V, CASE VIでは $P(\bar{u})$ が大きなバイアスをもっていることがわかる。

ここで, これらの推定値をもとに, 式(19), (20)の値を図示すると図-28~図-31のようになる。式(19), (20)

$$S[nP(\bar{u})] = \sqrt{\frac{\sum_{k=1}^5 \{nP_k(\bar{u}) - E[p(u)]\}^2}{5}} \quad (19)$$

$$S[n\bar{P}(\bar{u})] = \sqrt{\frac{\sum_{k=1}^5 \{\bar{n}P_k(\bar{u}) - E[p(u)]\}^2}{5}} \quad (20)$$

表-7 $S[n\hat{\beta}_k] \leq 0.4$ の場合

	CASE I	CASE II	CASE III	CASE IV	CASE V	CASE VI
$\hat{\beta}_1$	—	—	900	—	—	1000
$\hat{\beta}_2$	—	800	500	—	500	700
$\hat{\beta}_3$	1000	500	100	800	300	100
$\hat{\beta}_4$	400	600	100	300	100	100
$\hat{\beta}_5$	600	100	100	100	100	100

表-8 $S[n\hat{\beta}_k] \leq 0.5$ の場合

	CASE I	CASE II	CASE III	CASE IV	CASE V	CASE VI
$\hat{\beta}_1$	—	1000	500	—	—	900
$\hat{\beta}_2$	—	800	500	—	400	300
$\hat{\beta}_3$	600	300	100	800	100	100
$\hat{\beta}_4$	400	200	100	200	100	100
$\hat{\beta}_5$	100	100	100	100	100	100

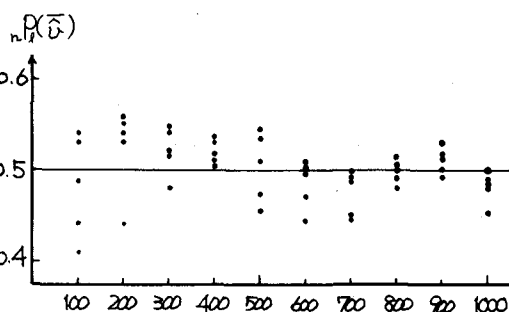


図-16 CASE Iの $P(\bar{u})$

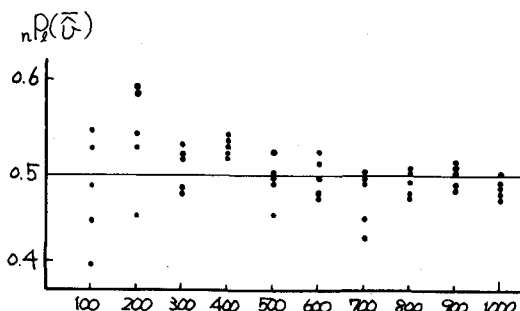


図-17 CASE IIの $P(\bar{u})$

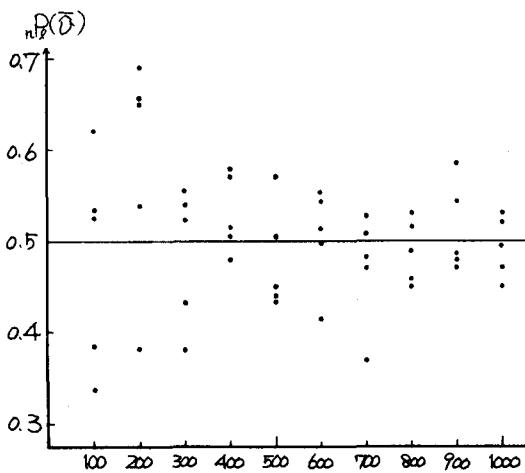


図-18 CASE III の $P(\hat{p})$

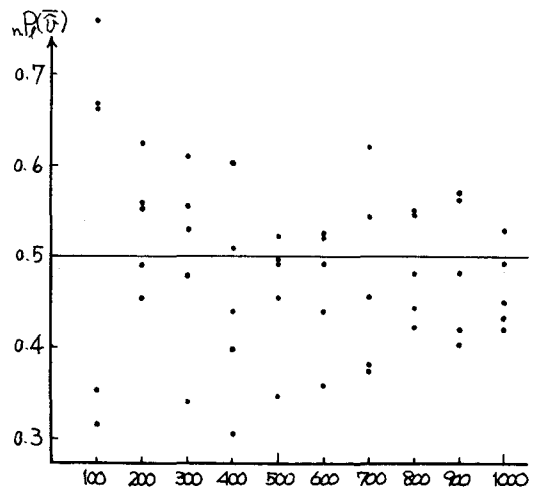


図-19 CASE IV の $P(\hat{p})$

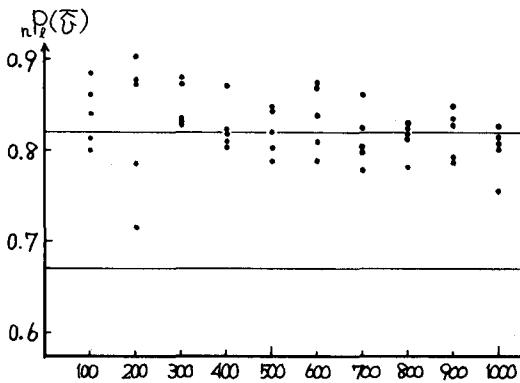


図-20 CASE V の $P(\hat{p})$

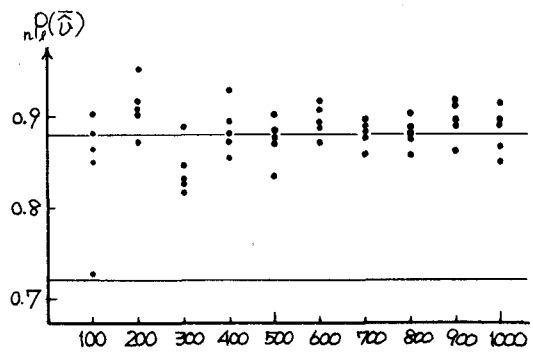


図-21 CASE VI の $P(\hat{p})$

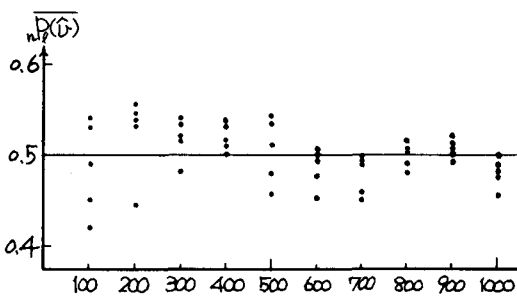


図-22 CASE I の $P(\hat{p})$

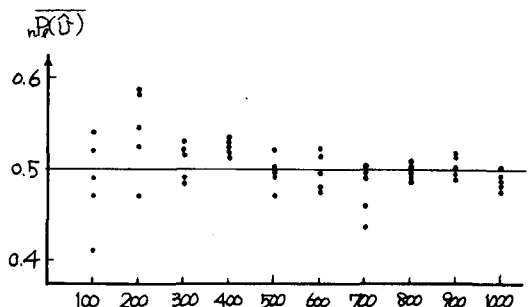


図-23 CASE II の $P(\hat{p})$

において、 ℓ は各データサイズにおける選択比率の推定番号 ($\ell = 1, 2, \dots, 5$) であり、 n はデータサイズ ($n = 100, 200, \dots, 1000$) である。また、 $nP(\hat{p})$ 、 $n\overline{P}(\hat{p})$ はそれぞれデータサイズ n 、推定番号 ℓ のときの $P(E|u)$ 、 $E(p|u)$ の推定値である。CASE I、CASE II では $S[\overline{P}(\hat{p})]$ 、 $S[\overline{P}(\hat{p})]$ が同じような値を示すが、CASE III、CASE IV では前述の如く $P(\hat{p})$ がデータサイズが小さい場合には不安定なため、 $S[\overline{P}(\hat{p})]$ は $S[\overline{P}(\hat{p})]$ に比較するとデータサイズが小さい場合に著しく大きな値となっている。従って、 $E(p|u)$ と $P(E|u)$ とに差がない CASE でも $V(u)$ が大きい場合には、データサイズが小さければ大きな誤差をもつことになる。CASE V、CASE VI では $E(p|u)$ と $P(E|u)$

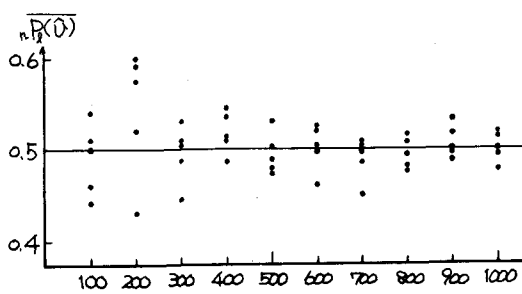


図-24 CASE III の $\overline{P(\hat{p})}$

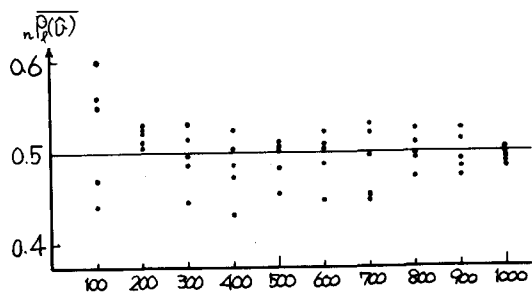


図-25 CASE IV の $\overline{P(\hat{p})}$

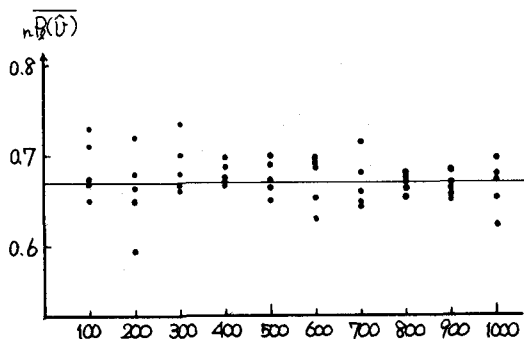


図-26 CASE V の $\overline{P(\hat{p})}$

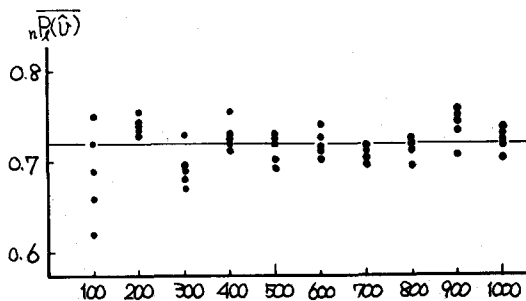


図-27 CASE VI の $\overline{P(\hat{p})}$

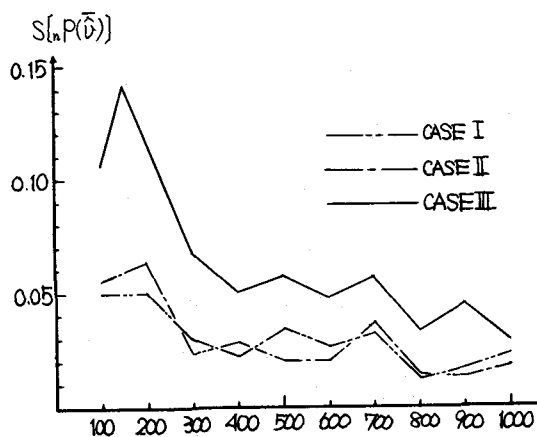


図-28 データサイズと $S[nP(\hat{p})]$ - (1)-

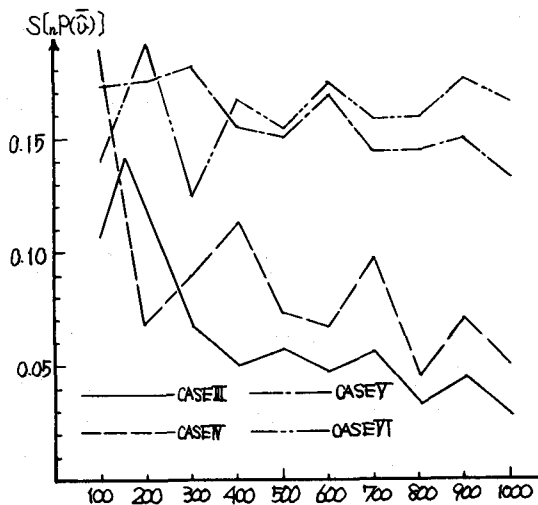


図-29 データサイズと $S[nP(\hat{p})]$ - (2)-

とに差があり、 $P(E(u))$ がバイアスをもっているため $S[nP(\hat{p})]$ はデータサイズを大きくしても小さくならない。ここで、あるデータサイズ以上の場合には $S[nP(\hat{p})]$ 、 $S[nP(\hat{p})]$ がある水準以下になるデータサイズに注目して整理すると、表-9、表-10のようになる。表-9によると、 $S[nP(\hat{p})]$ を0.025以下とするにはCASE I、CASE IIで800以上のデータサイズが必要で、0.05以下とすると300以上のデータサイズでよいことがわかる。CASE V、CASE VIではデータサイズが1000であっても $S[nP(\hat{p})]$ 0.1以下で $P(\hat{p})$ を求めることは困難である。表-10によると、選定比率を $\overline{P(\hat{p})}$ を求める場合には、どのCASEでも同程度のデータサイズでよいことがわかる。CA

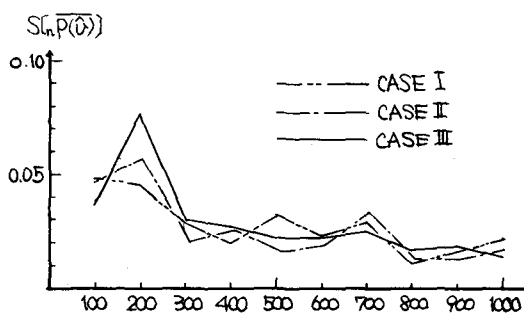


図-29 データサイズと $S_n(P(\hat{u}))$

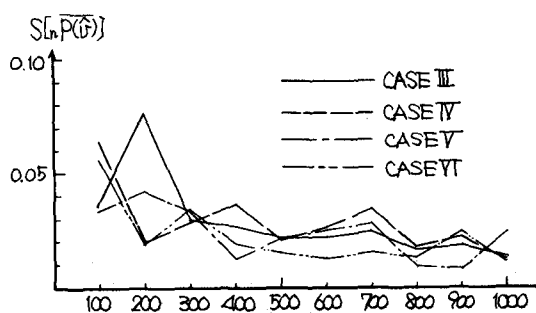


図-30 データサイズと $S_n(P(\hat{u}))$

SEⅢを例にとると、 $S_n(P(\hat{u}))$ 0.025以下の場合には、800以上、0.050以下の場合には300以上、0.075以下の場合には、0.050以下の場合と同じように300以上、0.100以下の場合には100程度のデータサイズを必要とする。

以上のことより、推定された効用値 $\hat{u}$ の平均値 $\bar{P}(\hat{u})$ をもとに安定的な選択比率を求めるためには、 $V[u]$ が大きいときには、大きなデータサイズを必要とすることがわかる。また、推定された効用値 $\hat{u}$ ごとの確率 $P(\hat{u})$ の平均値 $\bar{P}(\hat{u})$ によって選択比率を求める場合には、効用値の分布特性が異なっても、許容される誤差に対応してほぼ同程度のデータサイズが必要となる。このことから、選択比率を求める場合に必要とされるデータサイズについては、集計化の方法との関連で考える必要があることがわかる。

5. まとめ

本研究では、現在都市交通計画の分野で注目されている非集計ロジットモデルをとりあげ、パラメータ推定値や選択比率の推定値の安定性の問題について、効用値の分布特性とデータサイズとの関連において検討を行った。具体的には、説明変数が5変数のバイナリー・チョイス型の非集計ロジットモデルを考え、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仮観測値データを作成することによって検討を行った。パラメータを推定する場合と選択比率を推定する場合とにわけて、その結果を要約すると以下のようになる。

- (1) モデルの適合度が高い場合でも、モデルに取り入れられた説明変数間に高い相関関係が存在する場合には、小さなデータサイズで安定的なパラメータ推定値を得ることは困難である。100, 200のデータサイズで安定的なパラメータ推定値が求められるのは、説明変数間に高い相関関係が存在せず、モデルの適合度がよく、かつ許容される相対誤差の水準が高い場合などに限られる。
- (2) 推定された効用値 $\hat{u}$ ごとの確率 $P(\hat{u})$ の平均値 $\bar{P}(\hat{u})$ を選択比率とする場合には、効用値の分布特性が異なっても、許容される誤差の水準に対応して、ほぼ同程度のデータサイズが必要となる。推定された効用値 $\hat{u}$ の平均値 $\bar{P}(\hat{u})$ をもとにして選択比率を求める場合には、 $V[u]$ が大きいときにはとくに大きなデータサイズを必要とする。安定性のある選択比率の推定値を求めるためのデータサイズの検討にあたっては、集計化の

表-9 許容誤差とデータサイズ ($P(\hat{u})$ の場合)

	0.025	0.050	0.075	0.100
CASE I	800	300	100	100
CASE II	800	300	100	100
CASE III	—	800	300	300
CASE IV	—	—	600	500
CASE V	—	—	—	—
CASE VI	—	—	—	—

表-10 許容誤差とデータサイズ ($P(\hat{u})$ の場合)

	0.025	0.050	0.075	0.100
CASE I	800	100	100	100
CASE II	800	300	100	100
CASE III	800	300	300	100
CASE IV	800	200	100	100
CASE V	800	100	100	100
CASE VI	400	200	100	100

方法と效用値の分布特性とを合わせて考える必要がある。

なお、本研究では、效用値の分布特性に対応した仮想観測値データの作成方法を提案し、それによって作成された仮想観測値データをもとに検討を進めたが、この方法は非集計ロジットモデルの構築および適用上における他のいくつかの課題の検討にも有用な方法と考えられる。

参考文献

- 1) Amemiya, T: "Qualitative Response Models", Journal of Economic Literature, Vol. XIX, pp. 1483 ~ 1536, December 1981
- 2) 森嶋公夫: "ミクロ・モデルの計量分析", 経済セミナー, No. 326, pp. 58 ~ 65, 1982
- 3) 栢越 信, 塚本直幸: "非集計ロジットモデルの誤差分析", 交通工学, 第17巻第6号, pp. 13 ~ 22, 1982.
- 4) 栢越 信, 塚本直幸: "シミュレーションによる非集計ロジットモデルの誤差分析", 第6回交通工学研究発表会論文集, pp. 81 ~ 83, 1982.
- 5) 佐和隆光: "回帰分析", 朝倉書店, pp. 45 ~ 49, 1979.