

シミュレーション法によるプロビットモデルの推定可能性*

Estimatability of Multinomial Probit Model with Structured Covariance

屋井鉄雄**・中川隆広***・石塚順一****

By Tetsuo YAI**・Takahiro NAKAGAWA***・Junichi ISHITSUKA****

1. はじめに

鉄道ネットワークの高密度化が進む大都市域において鉄道整備を効果的に行うためには、高精度の需要推計技法が求められる。この大都市域を対象に鉄道経路選択を取り扱うためには、各経路間での重複区間増加に伴う経路間の類似性を無視できない。これに対し従来の非集計ロジットモデルでは I I A 特性等の制約条件を緩和するために様々な改良がなされてきたが未だに十分であるとは言えない。一方、筆者らはプロビットモデルの共分散項の構造化を行うことで、同モデルの欠点である複雑さの軽減を可能にすると同時に、選択肢間の任意の類似性を表現可能にし、これを構造化プロビットモデルと呼んだ。しかし、パラメータ推定や便益計測法については不明瞭な部分が残されていた。そこで、本研究では構造化プロビットモデルの特性を捉え、近年発展しつつあるシミュレーション法を採り入れるなど同モデルの完成及び操作性の向上を目的として分析を進めた。

2. 構造化プロビットモデルの定式化

意志決定者が各選択肢から得る効用 U を、観測可能な部分(確定項) V と不可能な部分(誤差項) ε からなる以下の式で表現する。

$$U = V + \varepsilon \quad (1)$$

プロビットモデルは選択肢ごとの誤差項に正規分布を仮定することにより導出され、この誤差項間に相関を持たせることが可能である。選択肢 R 個の中から選択肢 r を選択する確率 Pr は、以下の様に表される。

$$Pr = \int_{-\infty}^{x_r} \dots \int_{-\infty}^{x_r} \dots \int_{-\infty}^{x_r} \Phi(\varepsilon) d\varepsilon_1 \dots d\varepsilon_R \quad (2)$$

ここで、密度関数は、

$$\Phi(\varepsilon) = \frac{1}{(2\pi)^{\frac{R}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \varepsilon \Sigma^{-1} \varepsilon^T\right) \quad (3)$$

であり、分散共分散行列は次のように表される。

*キーワード：経路選択、構造化プロビットモデル、シミュレーション、便益計測

**正員、工博、東京工業大学工学部土木工学科
(東京都目黒区大岡山 2 丁目 12 番地 1 号、
TEL03-5734-2693, FAX03-3726-2201)

***正員、工修、JR 東海

****学生員、東京工業大学理工学研究科土木工学専攻

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1R} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2R} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1R} & \sigma_{2R} & \dots & \sigma_R^2 \end{pmatrix} \quad (4)$$

この時、(4)の各要素を効用関数のパラメータと同時に推定するのが従来のプロビットモデルの基本的な考え方である。しかし、利用者個人毎に出発地・目的地が異なるため、経路集合の組み合わせも異なり、個人毎に経路間の類似性が任意に変わる。そのため分散共分散項をパラメータとして母集団で共通に推定する従来のプロビットモデルでは鉄道経路選択には適用できなかった。そこで、経路間の任意の類似性を表現可能とするため共分散項の構造化を行う。今、経路 r に関する確率効用のランダム項を相互独立な 2 つの項(長さに依存する誤差 ε^1 ・経路固有の誤差 ε^0)に分割して考える。このとき誤差及び分散共分散行列は以下の式で表現できる。

$$\varepsilon_r = \varepsilon_r^1 + \varepsilon_r^0 \quad r = 1, \dots, R \quad (5)$$

$$\Sigma = \Sigma^1 + \Sigma^0 \quad (6)$$

ここで、 ε^1 は単位長さ毎に独立に分布すると仮定する。この仮定より誤差は経路の長さに依存すると考えられ、どの単位からも誤差が発生し、さらにその誤差が互いに独立であると考えられる。先の仮定より分散は

$$\text{Var}(\varepsilon_r^1) = \sigma^2(x) + \dots + \sigma^2(x) = n_r(x) \sigma^2(x) \quad (7)$$

$\sigma^2(x)$: 単位長さ毎の分散

$n_r(x)$: 単位数

いま、 x は空間距離とし、単位数が距離に比例して算出されるとすれば、経路 r の路線長 L_r を用いて(7)式は

$$\text{Var}(\varepsilon_r^1) = L_r \sigma^2 \quad (8)$$

と表すことができる。次に重複区間を含む 2 経路 r, q 間の共分散について考えてみる。先の仮定より

$$\begin{aligned} \text{Cov}(\varepsilon_r^1, \varepsilon_q^1) &= E(\varepsilon_r^1 \varepsilon_q^1) \\ &= E(\varepsilon_r^{\text{ov}} + \varepsilon_r^{\text{nov}})(\varepsilon_q^{\text{ov}} + \varepsilon_q^{\text{nov}}) \\ &= E(\varepsilon_r^{\text{ov}} \varepsilon_q^{\text{ov}}) + E(\varepsilon_r^{\text{ov}} \varepsilon_q^{\text{nov}}) + E(\varepsilon_r^{\text{nov}} \varepsilon_q^{\text{ov}}) + E(\varepsilon_r^{\text{nov}} \varepsilon_q^{\text{nov}}) \\ &= E(\varepsilon_r^{\text{ov}})^2 \end{aligned} \quad (9)$$

ov : 重複区間

nov : 非重複区間

経路 r, q 間の重複区間での単位数 $n_{rq}(x)$ を用いて

$$\text{Cov}(\varepsilon_r^1, \varepsilon_q^1) = n_{rq}(x) \sigma^2(x) \tag{10}$$

であり、経路 r, q 間の重複区間長 L_{rq} とを用いれば、

$$\text{Cov}(\varepsilon_r^1, \varepsilon_q^1) = L_{rq} \sigma^2 \tag{11}$$

と表すことができる。一方、経路固有の誤差は各経路毎に発生するため共分散は 0 となる。

$$\begin{aligned} \text{Cov}(\varepsilon_r^0, \varepsilon_q^0) &= \sigma_0^2, \quad q = r \\ &= 0, \quad q \neq r \end{aligned} \tag{12}$$

よって分散共分散行列は次のように表すことができる。

$$\Sigma = \sigma^2 \begin{pmatrix} L_1 & L_{12} & \cdots & L_{1R} \\ L_{12} & L_2 & \cdots & L_{2R} \\ \vdots & \vdots & \ddots & \vdots \\ L_{1R} & L_{2R} & \cdots & L_R \end{pmatrix} + \sigma_0^2 I \tag{13}$$

L_r : 経路 r の路線長

L_{rq} : 経路 r, q 間の重複区間長

分散比で表現すれば、

$$\Sigma = \sigma_0^2 \begin{pmatrix} \eta L_1 + 1 & \eta L_{12} & \cdots & \eta L_{1R} \\ \eta L_{12} & \eta L_2 + 1 & \cdots & \eta L_{2R} \\ \vdots & \vdots & \ddots & \vdots \\ \eta L_{1R} & \eta L_{2R} & \cdots & \eta L_R + 1 \end{pmatrix} \tag{14}$$

$$\eta = \frac{\sigma^2}{\sigma_0^2} \tag{15}$$

と表すことができ、分散共分散行列のパラメータは、 η の 1 つだけとなり、効用関数に含まれる説明変数ごとのパラメータとの同時推定が可能である。

3. パラメータ推定

3-1 最尤推定法

プロビットモデルのパラメータ推定方法では最尤推定法が一般的である。最尤推定法は選択確率の算出に際し、(12)式より多重積分を行う必要があり、尤度関数を最大化するためには、この計算を繰り返さなければならず、かなりの計算時間を必要とする。そこで、簡便な推定方法の開発が求められ、代替的な推定方法が開発されてはいるが、満足なレベルには達していなかった。しかし近年、シミュレーション法が発展しておりその実用性が高まっている。

3-2 シミュレーション法

シミュレーション法とは、誤差項に多数の乱数を基に確率的に生成された値をあて各意志決定者の選択をシミュレートし、選択確率を近似する方法である。シミュレーション法は選択確率算出過程とパラメータ推定過程に分別できる(表 1)。シミュレーション法の特徴は、選択確率算出の際に多重積分を必要としない点であり、最尤推定法の欠点を克服できる。本研究では、最も正確であるとされる GHK 法と計算時間の少なく

すむ SML を組み合わせてパラメータの推定を行う。

ここで、GHK 法について説明する。4 肢選択のケースにおいて選択肢 1 が選ばれるとする。その際効用関数は、以下の関係が成立している。

$$\begin{pmatrix} U_1^* \\ U_2^* \\ U_3^* \end{pmatrix} = \begin{pmatrix} U_2 - U_1 \\ U_3 - U_1 \\ U_4 - U_1 \end{pmatrix} = X\theta + \begin{pmatrix} w_{11} & 0 & 0 \\ w_{12} & w_{22} & 0 \\ w_{13} & w_{23} & w_{33} \end{pmatrix} \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{pmatrix} \leq 0 \tag{16}$$

$$WW^t = M_1 \Sigma \Sigma^t M_1^t \tag{17}$$

$$M_1 = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} \tag{18}$$

ζ は、0 ~ 1 の一様乱数を処理し得られる。

ある乱数 r に対する選択肢 1 の選択確率は、以下の様に条件付確率の積で求められる。

$$\begin{aligned} P_{1r} &= P(U_1^* \leq 0, U_2^* \leq 0, \text{and } U_3^* \leq 0) \\ &= P(U_1^* \leq 0) * P(U_2^* \leq 0 | U_1^* \leq 0) \\ &\quad * P(U_3^* \leq 0 | U_2^* \leq 0, U_1^* \leq 0) \end{aligned} \tag{19}$$

以上の過程を乱数発生回数 (R 回) だけ繰り返し選択確率の近似を行う。

$$E(P_1) = \frac{1}{R} \sum_{r=1}^R P_{1r} \tag{20}$$

この確率を対数尤度関数に代入して(SML)、パラメータ推定を行う。

3-3 鉄道経路選択モデルへの適用

本研究で利用するデータはすべて 3 肢選択であるが、大都市交通センサス(平成 2 年)から実際の選択経路が確認できるサンプルに対し、新たに 2 つの代替経路を加えている。サンプル数は 1074 である。乱数発生回数は 25, 50, 100 回で推定を行い、分散共分散行列(14)式は、基準化のため $\sigma_0^2=1$ とし、 $\sigma^2(x)$ を効用関数のパラメータと同時に推定する。モデル構築結果を表 2 に示す。表 2 におけるパラメータ平均誤差、最大誤差とは最尤法で得られたパラメータ値に対する数値である。

表 1 各シミュレーション法

選択確率算出法	パラメータ推定法
-Frequency Simulator (不連続点が存在する)	-Simulated Maximum Likelihood (SML)
-Kernel-Smoothed Frequency Simulator (不連続点の解消)	-Method of Simulated Moments (MSM)
-The GHK Recursive Simulator (効用差のマトリクスに変形) (条件付確率の積)	(少ない乱数利用回数でバイアスが消える) (頑強である)

混雑率は平均混雑率を利用している。

表2より尤度比は5つのモデルにおいて同値であり、平均誤差は最大で約3%であることからシミュレーション法を利用しても精度を確保できることが見てとれる。

シミュレーション法では、同じ乱数発生回数でも利用乱数次第で結果が異なる。その傾向は乱数発生回数が多くなる程、平均誤差、最大誤差及びそれらの変動は小さくなり信頼性は増す。他方、推定計算に要する時間は伸び、パラメータの精度を確保できる乱数発生回数の見極めがシミュレーション法には求められる。今回の推定に要した計算時間については、3肢選択のため多重積分の負荷が少なく、乱数発生回数25より最尤法の方が速い。しかし、選択肢が増えればシミュレーション法の利点は大きい。

4. 便益計測方法

4-1 プロビットモデルによる

支払意思額の算出法

効用関数を(1)式の様に考えれば、最大効用 U^* の確率分布は式(20)で表される。

$$F(U^*) = \text{Prob}(U_1 \leq U^*, \dots, U_R \leq U^*) \\ = \text{Prob}(U_1 \leq U^*, \dots, U_R \leq U^*) \quad (21)$$

$$= \int_{\varepsilon_1 = -\infty}^{U^* - V_1} \dots \int_{\varepsilon_R = -\infty}^{U^* - V_R} \Phi(\varepsilon) d\varepsilon_R \dots d\varepsilon_1 \quad (22)$$

最大効用の密度関数は、この分布関数を U^* で微分し得られる。密度関数の特性を考えるにあたり、ここでは2肢選択を例にとる。選択肢間が無相関の場合、密度関数 S は(22)式で表される。

$$S(U^*) = \frac{1}{2\pi} \exp\left(-\frac{\varepsilon_1^2}{2}\right) \int_{-\infty}^{\varepsilon_1 + V_1 - V_2} \exp\left(-\frac{t^2}{2}\right) dt \\ + \frac{1}{2\pi} \exp\left(-\frac{(\varepsilon_1 + V_1 - V_2)^2}{2}\right) \int_{-\infty}^{\varepsilon_1} \exp\left(-\frac{t^2}{2}\right) dt$$

$$U^* = \varepsilon_1 + V_1 \quad (23)$$

図1は確定項値を変えて密度関数を比較したものである。 $V_1 = V_2 = 0$ では、最頻値が期待値より左に位置しており、確定項値の差が広がるにつれ正規分布に近づく。また、相関が高くなるにつれ正規分布に近づくことも確認した。最大効用の密度関数は正規分布ではないが、極めて似た分布を示す。したがってプロビットモデルでは最大効用の期待値ではなく最頻値を利用する。最大効用の最頻値 M は(22)式を U^* で2回微分することで得られる。

$$M = \frac{\partial^2 F(U^*)}{\partial U^{*2}} = 0 \quad (24)$$

消費者余剰は、整備前の最頻値 M^0 と整備後の最頻値 M^1 の差分により求められる。最大効用最頻値の一定増分に対する不変性(式(25))を考慮すれば、1人あたりの支払意思額 ΔC は、ラインホール費用パラメータ β を用いて式(26)で表される。

$$M(V+a) = M(V) + a \quad (25)$$

$$\Delta C = \frac{M^1 - M^0}{\beta} \quad (26)$$

このように支払意思額は整備前後の最頻値差を用いて簡便に算出できる。消費者便益は支払意思額と利用者数から算出できるため、整備による便益を計測できる。

4-2 最頻値算出

ここでは、(21)式の多重積分をGHK法により近似させた分布関数を考える。3肢選択の分布関数を近似するには2つの乱数が必要である。パラメータ推定ではサンプル毎に乱数が異なる様に設定したため、(サンプル数)×(乱数発生回数)の乱数が必要であったが、最頻値はODごとに1つであるため乱数発生回数は、パラ

表2 モデルの構築結果 (t値) サンプル数1074

		Logit	Probit(最尤法)	(乱数25)	(乱数50)	(乱数100)
ラインホール費用 β	(円)	-6.34e-3(6.33)	-5.83e-3(5.22)	-5.92e-3(5.20)	-5.93e-3(5.20)	-5.91e-3(5.11)
アクセス時間	(分)	-1.42e-1(6.11)	-1.27e-1(5.23)	-1.29e-1(5.23)	-1.29e-1(5.18)	-1.29e-1(5.09)
イグレス時間	(分)	-1.66e-1(5.99)	-1.51e-1(5.45)	-1.53e-1(5.48)	-1.53e-1(5.43)	-1.53e-1(5.39)
ラインホール時間	(分)	-7.51e-2(7.87)	-6.95e-2(5.80)	-7.10e-2(5.79)	-7.05e-2(5.73)	-7.06e-2(5.58)
乗換時間(上り階段)	(分)	-3.91e-1(2.64)	-3.35e-1(2.44)	-3.34e-1(2.43)	-3.39e-1(2.47)	-3.38e-1(2.42)
乗換時間(下り+水平)	(分)	-1.31e-1(5.63)	-1.16e-1(5.21)	-1.18e-1(5.19)	-1.17e-1(5.17)	-1.18e-1(5.16)
待ち時間	(分)	-1.27e-1(4.41)	-1.18e-1(3.90)	-1.20e-1(3.92)	-1.19e-1(3.90)	-1.19e-1(3.81)
乗換回数	(回)	-4.19e-1(4.04)	-3.82e-1(3.92)	-3.95e-1(3.97)	-3.92e-1(3.96)	-3.91e-1(3.95)
混雑率*ラインホール時間(分%)		-1.38e-7(1.20)	-9.08e-8(0.82)	-8.73e-8(0.80)	-8.82e-8(0.75)	-9.10e-8(0.80)
分散 $\sigma^2(x)$			0.302(1.45)	0.343(1.51)	0.332(1.47)	0.333(1.41)
尤度比		0.182	0.182	0.182	0.182	0.182
パラメータ平均誤差	(%)			3.09	2.45	2.26
パラメータ最大誤差	(%)			13.4	9.93	10.4

メータ推定値の1サンプル分程度に少なくできる。
ここでは1ODに対し4組と25組の乱数を発生させて計算を行った。

4-3 便益計測結果

図2は様々なパターン(Pと記す)での支払意思額算出結果の1例である。構造化(14)式の利点はP1,3の様にネットワーク構成が異なる場合のみならず、P3,4の様にネットワーク構成が同一でも経路長の違いによる類似性の相違を表現可能な点である。算出結果の比較を行う際、パラメータ推定と支払意思額算出法の組み合わせとしてロジットモデル、プロビットモデル(両方で多重積分)に加え、以下の5つの場合(シミュレーション法)を考えた。

- case1: 多重積分 - 乱数25組
- case2: 乱数100組 - 多重積分
- case3: 乱数100組 - 乱数25組
- case4: 乱数25組 - 乱数25組
- case5: 乱数25組 - 乱数4組

ロジットモデルではネットワーク構成を反映しないため支払意思額は一律97円である。構造化によりcase毎に支払意思額に差異があり、case1ではP1とP4で14.5円の差が生じている。シミュレーション法では、すべてのパターンにおいて乱数利用回数の多いcase3がprobit(積分法)に近い支払意思額を得ている。しかし、case5においても支払意思額の減少の割合はcase1と同じ傾向を示し、case1との差は最大2.2円であり精度が保たれていることが分かる。またP1,2では最頻値計算の際、乱数の影響はないため支払意思額の誤差は

推定パラメータの誤差に依存している。以上の簡便な試算より、シミュレーション法は最尤推定法に対して、若干誤差を生じさせているが、ロジットモデルによる過大評価よりはるかに小さいことを確認できた。

5. まとめ

本研究の成果は、1) 構造化プロビットモデルの定式化を行い、個人差の表現を可能としたこと、2) シミュレーション法によるモデル推定の有効性を示したこと、3) 最大効用の特性を踏まえ、簡便な支払意思額の算出法を提案したこと、の3つである。なお、シミュレーション法によるパラメータ推定と便益計算の算定法については更に検討を進める予定である。

参考文献

- 1) Daniel McFadden: A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration, *Econometrica*, Vol.57, No.5, 995-1026, 1989.
- 2) John Geweke, Michael Keane, David Runkle: Alternative Computational Approaches to Inference in the Multinomial Probit Model, *The Review of Economics and Statistics*, 609-632, 1994.
- 3) 屋井鉄雄、岩倉成志、伊東誠: 鉄道ネットワークの需要と余剰の推計法について、*土木計画学研究論文集* 11, 81-88, 1993.
- 4) Tetsuo Yai, Seiji Iwakura, Shigeru Morichi: Multinomial Probit Model with Structured Covariance Matrix for Route Choice Behavior, *Transportation Research B*, forthcoming.
- 5) 屋井鉄雄、中川隆広: 構造化プロビットモデルの発展性、*土木計画学研究論文集* 13, 1996.
- 6) 屋井鉄雄、中川隆広、石塚順一: 鉄道サービスの改善効果の計測法、*j-rail 論文集*, 1996.

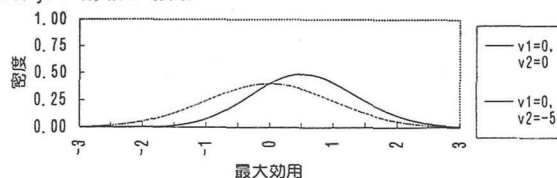


図1 最大効用の密度関数

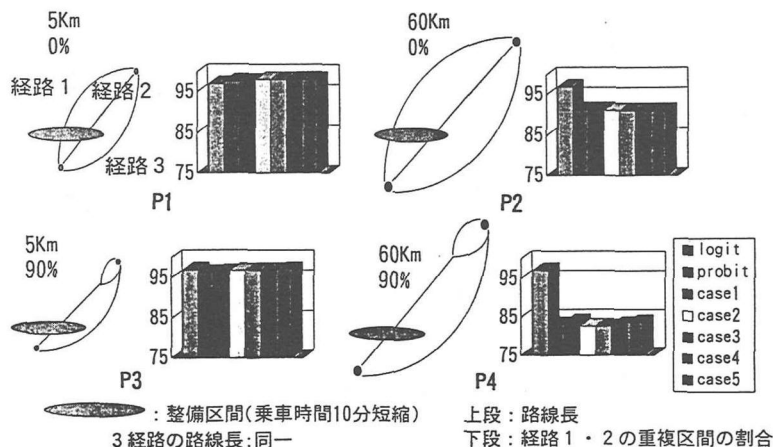


図2 ネットワークパターンと支払意思額