【土木計画学研究・論文集 No.3 1986年1月】

共役性概念に基づくロジットモデルのパラメータ推定法とその統計的検定について*

Parameter Estimation and Statistical Tests of Logit Models with the Conjugate Theory

宮城俊彦** 小川俊幸***
by Toshihiko MIYAGI and Toshiyuki OGAWA

This paper aims at providing disaggregate tests theoreticaly consistent with a parameter estimation method. The parameter estimation method presented here is based on a straightforward application of the conjugacy theory on the statisfaction function of disaggregate logit model. Because the conjugacy theory indicates that the optimum function of conjugate problem induces an entropy of choice behavior, we can use the entropy as a measure of uncertainty involved in estimating of parameters. Then, it can be show that Shannon's information can be derived from the entropy as the conjugate function of satisfaction function and that Shannon's information is useful for comprehensive tests such as "significance", "usefulness" and "accuracy" of models.

1. はじめに

非集計行動モデルは、個人の選択行動を選択確率（$0<P<1$）として表現する。分析者は最良のモデルを得るためモデルの精度を計量的に評価する尺度を必要とするが、観測される選択行動が０－１のランダム変数であるので、モデルの精度を統計的に検定する手法が必要となる。

従来、非集計モデルを評価する尺度として、的中率、尤度比χ^2有意性検定としての尤度比指数ρ^2などがよく用いられる。的中率は直感的には分かりやすい尺度であるが、何等の統計的背景を持たない。尤度比は帰無モデルが検定すべきモデルのパラメータの部分集合で定義されるときχ^2分布する。したがって、χ^2検定を用いてモデルの仮説検定が行なえる。しかし、χ^2検定はモデルの予測精度が評価できず、また、異なるモデルの比較の場合は、一方のモデルの説明変数が他方のそれの部分集合であるという制約が必要となる[1]。モデルの予測能力の最も適切な尺度はρ^2であるが、ρ^2は一回の観測データ（単純サンプリングと呼ぶことにする）に基づく尺度であり、データを反復して収集したとき（反復サンプリングと呼ぶことにする）にはそれぞれに対応したρ^2が出現するので、ρ^2値だけでモデルの予測能力を判断するのは問題が残ろう。これらのことは、モデルの良し悪しの判断は１個の尺度だけでは不十分であり、複数の尺度が必要で、できればそれらを統一する尺度が望まれること、また、反復して選択が行なわれるという事実をも考慮した尺度が必要なことを示している。

*キーワーズ：ロジットモデル、パラメータ推定、仮説検定

**正会員　工博　岐阜大学助教授　工学部建設工学科（〒501-11　岐阜市柳戸１－１）
***学生会員　岐阜大学工学研究科大学院　土木工学専攻（同上）

上述の目的を達成する尺度は、Hauserによって提案されている[2]。Hauserは、Kullback情報量基準と類似の情報量を用いて、モデルの"有意性""有効性"そして"精度"を検定する方法を提案しているが、"有意性"は従来のχ^2検定、"有効性"はρ^2に一致する。

本研究の目的は、非集計モデルのパラメータ推定の新しい手法と概念を提案し、そしてその推定手法と密接に関係したモデルの統計的検定法を提案することにある。パラメータ推定手法は満足度関数の共役性概念から直接的に導かれ、また、推定プロセスそのものが選択の不確実性を表すことを利用して、不確実性の減少の度合を示す尺度、すなわち情報量が定義される。このように導かれる情報量はHauserの提案する情報量に結果的には一致するが、Shannonの情報量を基準におく点でHauserの導出とは異なる。また、共役性概念を用いて定式化することにより、パラメータ推定とその検定が一貫した理論によって行なえることを明らかにしている点でも、従来の研究およびHauserのアプローチとは異なる。

2．ロジットモデルのパラメータ推定法

（1）ランダム効用モデルと表記法

まず初めに、選択可能選択肢集合を$A=\{a_j\}$あるいは$J=\{j\}$で表現し、各々の次元をJとする。このとき、個人iが選択肢jを選択するときの効用を次の加法的ランダム効用として定義する。

$$\tilde{U}_j^i = V_j^i + \tilde{\varepsilon}_j^i \qquad (1)$$

ここで各変数は個人iに関し

$\tilde{U}_j^i$ ：選択肢jの全効用

V_j^i ：選択肢jの測定可能効用

$\tilde{\varepsilon}_j^i$ ：選択肢jの測定しえない効用（誤差項）

である。なお記号「～」はその変数が確率変数であることを表している。選択肢jの測定可能効用は次式に示すように個人iおよび選択肢jに関する説明変数ベクトルとパラメータベクトルの関数として表現できる。

$$V_j^i = V_j^i(\theta, \mathbb{X}_j^i) \qquad (2)$$

ここで

$\mathbb{X}_j^i$ ：個人iについての選択肢jの説明変数ベクトル。k番目の説明変数をX_{jk}^iで表現するなら

$\mathbb{X}_j^i = (X_{j1}^i, X_{j2}^i, \cdots X_{jk}^i, \cdots X_{jK}^i)$

$|\mathbb{X}_j^i| = K$

θ ：パラメータベクトル

一般には次式に示すような線形関数が用いられる。

$$V_j^i = \theta_{0j} + \sum_{k=1}^{K} \theta_{jk} X_{jk}^i \qquad (3)$$

個人iの測定可能効用ベクトル$\mathbb{V}^i = \{V_1^i, V_2^i, \cdots V_j^i\}$は、あらゆる個人の経験する測定可能効用ベクトルの実現可能集合W上での1つの実現直である。同様に、説明変数ベクトル$\mathbb{X}^i = \{\mathbb{X}_1^i, \mathbb{X}_2^i, \cdots \mathbb{X}_j^i\}$も説明変数ベクトルの実現可能集合Xからの1つの実現値である。実現可能集合W, Xは各々有限の可算個集合であると仮定する。

（2）満足度関数

式（1）のランダム効用の下で、選択肢選択に伴う個人iの最大効用の期待値を満足度関数と呼ぶことにしS^iで表現する。式（1）で定義された効用の誤差項が、互いに独立で同一のGumbel分布に従うと仮定するならば、個人iの満足度関数は次式で与えられる[3]。

$$S^i(\mathbb{V}^i) = \frac{1}{\alpha} \ln \sum_{j \in J} \exp(\alpha V_j^i) \qquad (4)$$

S^iはある特定の$\mathbb{V}^i$が実現したときの満足度関数であり、したがって、集合Wにわたっての期待満足度関数は次式で定義できる。

$$\overline{S}(\mathbb{V}) = \frac{1}{\alpha} \sum_{\mathbb{V}^i \in \mathbb{W}} Pr(\mathbb{V}^i) \ln \sum_{j \in J} \exp(\alpha V_j^i) \qquad (5)$$

ここに、$Pr(\mathbb{V}^i)$はWにおいて$\mathbb{V}^i$が出現する確率である。$Pr(\mathbb{V}^i)$は特に事情がなければ個人に対し等確率であると仮定することができよう。また、式（3）の線型効用関数のもとでは、ある固定された推定パラメータベクトルに対し、$\mathbb{V}^i$が出現する確率は、$\mathbb{X}^i$が出現する確率に等しいと仮定してよいであろう。このとき式（5）は次のように定義できる。

$$\overline{S}(\mathbb{V}) = \frac{1}{\alpha} \sum_{\mathbb{X}^i \in X} Pr(\mathbb{X}^i) \ln \sum_{j \in J} \exp(\alpha V_j^i) \qquad (6)$$

今、N人が経験する効用レベルの全体が、個々人が経験しうる実現可能効用空間Wを形成すると仮定す

る。すなわち、$\mathbb{W}=\{\mathbb{V}^1,\mathbb{V}^2,\cdots\mathbb{V}^N\}$。このとき、個々人の説明変数ベクトル$\mathbb{X}^i$がすべて異なると仮定すると、$\mathbb{X}^i$の出現確率は$Pr(\mathbb{X}^i)=1/N$とおくことができる。したがって、式(5)または式(6)は次のようになる。

$$\overline{S}(\mathbb{V})=\frac{1}{\alpha}\sum_{i=1}^{N}(1/N)\ln\sum_{j\in J}\exp(\alpha V_j^i) \qquad (7)$$

以下の議論においては、主に式(7)の満足度関数を対象に分析する。満足度関数は種々の興味ある性質をもつが、その一つは満足度関数の接線勾配が選択確率を表すという性質である。すなわち、式(4)の満足度関数に対し、

$$\frac{\partial S^i(\mathbb{V}^i)}{\partial V_j^i}=\frac{\exp(\alpha V_j^i)}{\sum_{j\in J}\exp(\alpha V_j^i)}=P_j^i \qquad (8)$$

が成立する。式(8)は個人iという条件のもとでの選択肢jを選ぶ条件付確率を表し $\Sigma_j P_j^i=1$ が成立する。同様に式(6)に対しては

$$\frac{\partial \overline{S}(\mathbb{V})}{\partial V_j^i}=\sum_{i=1}^{N}Pr(\mathbb{X}^i)P_j^i=\sum_{i=1}^{N}P_{ji}=q_j \qquad (9)$$

が成立する。P_{ji}はN人の集団のうち個人iが選択肢jを選ぶ同時確率を表している。また、q_jは周辺確率であり、市場占有率を表す。式(7)の満足度関数の下では式(9)の周辺確率は

$$(1/N)\sum_{i=1}^{N}P_j^i=q_j \qquad (10)$$

となることに注意する。

（3）共役性理論

共役性とは同一の関数が二通りの方法で表現できることを表す概念であり、共役関数は関数のもつ共役性から自然に誘導される関数である。一般的に関数は点の集合の軌跡として表現されるが、別の方法では関数のエピグラフを支える包絡支持平面の共通集合で表現できる。共役関数は後者の表現方法そのものであり、関数の接線勾配を共役変数としてその関数の異なる表現を行なうものである。満足度関数の共役関数はその接線勾配すなわち選択確率を共役変数として表現することになる。式(7)の満足度関数$\overline{S}(\mathbb{V})$の共役関数$\overline{S}^*(\mathbb{P})$は次の最適化問題で与えられる。

[P 0]

$$-\overline{S}^*(\mathbb{P})=\min_{\mathbb{V}}.\{\overline{S}(\mathbb{V})-\frac{1}{N}\sum_{i=1}^{N}\sum_{j\in J}P_j^iV_j^i\}$$

$$=\min_{\mathbb{V}}.\{\frac{1}{\alpha N}\sum_{i=1}^{N}\ln\sum_{j\in J}\exp(\alpha V_j^i)-\frac{1}{N}\sum_{i=1}^{N}\sum_{j\in J}P_j^iV_j^i\} \qquad (11)$$

このとき、P_j^iは与えられていることが前提であり[P 0]は$\mathbb{V}$に関する最適化を扱っていることに注意する。このときP_j^iはV_j^iに対する共役変数であり最適解$\hat{\mathbb{V}}$は次式を満足しなければならない。

$$P_j^i=\frac{\exp(\alpha\hat{V}_j^i)}{\sum_{j\in J}\exp(\alpha\hat{V}_j^i)} \qquad (12)$$

このとき式(11)の最適値関数は次式に定義するエントロピーに等しい[4]。

$$-\overline{S}^*(\mathbb{P})=\frac{1}{\alpha}(-\frac{1}{N}\sum_{i=1}^{N}\sum_{j\in J}P_j^i\ln P_j^i)$$

$$=\frac{1}{\alpha}H(\mathbb{P}) \qquad (13)$$

（4）パラメータ推定法

ところで[P 0]で与えられる共役関数を求めるプロセスは次のことを意味している。まず、満足度関数Sの確率的表現は次のようなものである。

$$S=Pr[(U_1=\bar{U},U_2<\bar{U},\cdots U_J<\bar{U})\cup(U_1<\bar{U},U_2=\bar{U},\cdots U_J<\bar{U})\cup\cdots\cup(U_1<\bar{U},U_2<\bar{U},\cdots U_J=\bar{U})] \qquad (14)$$

すなわち、選択肢のうちどれかが最大効用を示す確率を表しており、したがって、式(11)の第1項は個人iが実際には選択しない選択肢おも含めた選択肢各々の効用の期待値を表すものである。一方、第2項は明らかに選択された選択肢についての測定可能効用期待値である。したがって式(11)は、測定不可能な効用をも含めた期待効用と測定できる効用に基づく期待効用の差をなるべく小さくするように測定可能効用を決定する問題である。ところで、測定可能効用は、式(2)で表現したようにパラメータベクトルと説明変数ベクトルで定義される。したがって、最適化問題[P 0]は、説明変数が与えられるという条件の下での、測定不可能な効用を含む期待効用関数と測定可能効用で定義される期待効用関数の差を

小さくするようにパラメータを決定するという、いわばパラメータ推定問題に他ならない。このように、式(11)は従来のパラメータ推定問題と全く異なる概念を与える。

今、個人iの選択パターンに関する反復サンプリングの結果、調査に基づく個人の選択確率分布 $\mathbb{P}^i=\{\mathring{P}^i_1, \mathring{P}^i_2, \cdots \mathring{P}^i_J\}$が与えられるものと仮定する。このとき、$\mathring{\mathbb{P}}^i$に矛盾しない測定可能効用 $\mathbb{V}^i$を満足度関数と測定可能効用の期待値から求める問題、すなわち、与えられた選択確率に矛盾しないようにパラメータベクトルを決定する問題が[P 0]の意味するパラメータ推定問題に他ならない。したがって、測定可能効用が、式(3)のような線形効用関数で与えられる場合のパラメータ推定問題は、

$\beta_{oj}=\alpha\theta_{oj}$, $\beta_k=\alpha\theta_k$ とおくならば次の最適化問題となる。

[P 1]

$$\min_{\boldsymbol{\beta}}. L_1(\boldsymbol{\beta})=\frac{1}{\alpha}\sum_{1=1}^{N}\ln\sum_{j\in J}\exp(\beta_{oj}+\sum_{k=1}^{K}\beta_k X^i_{jk})$$

$$-\frac{1}{N}\sum_{i=1}^{N}\sum_{j\in J}\mathring{P}^i_j(\beta_{oj}+\sum_{k=1}^{K}\beta_k X^i_{jk}) \qquad (15)$$

式(15)の最適解が満足すべき条件を求めると、

$$\frac{\partial L_1(\hat{\boldsymbol{\beta}})}{\partial \beta_{oj}}=\frac{1}{N}\sum_{i=1}^{N}P^i_j(\hat{\boldsymbol{\beta}})-\frac{1}{N}\sum_{i=1}^{N}\mathring{P}^i_j=0 \qquad (16)$$
$$j=1\cdots J$$

$$\frac{\partial L_1(\hat{\boldsymbol{\beta}})}{\partial \beta_k}=\frac{1}{N}\sum_{i=1}^{N}P^i_j(\hat{\boldsymbol{\beta}})X^i_{jk}-\frac{1}{N}\sum_{i=1}^{N}\mathring{P}^i_j X^i_{jk}=0$$
$$k=1\cdots K \qquad (17)$$

で与えられる。したがって、式(16),(17)の非線形連立方程式をNewton－Raphson法などにより解けばよい。ところで、真値に近い選択確率$\mathring{P}^i_j$を求めるにはデータを反復収集しなければならず、現実問題としては不可能に近い。一回の調査のみによって得られる選択行動は、個人iが選択肢jを選択したかしなかったかを示す単純データのみである。このような状況のもとでのパラメータ推定問題は次のようになる。

[P 2]

$$\min_{\boldsymbol{\beta}}. L_2(\boldsymbol{\beta})=\frac{1}{\alpha}\sum_{i=1}^{N}\ln\sum_{j\in J}\exp(\beta_{oj}+\sum_{k=1}^{K}\beta_k X^i_{jk})$$

$$-\frac{1}{N}\sum_{i=1}^{N}\sum_{j\in J}\triangle^i_j(\beta_{oj}+\sum_{k=1}^{K}\beta_k X^i_{jk}) \qquad (18)$$

ここで、

$$\triangle^i_j=\begin{cases}1: & \text{個人iが選択肢jを選択したとき}\\ 0: & \text{それ以外のとき}\end{cases}$$

式(18)の最適解が満足すべき条件は次式で与えられる。

$$\frac{\partial L_2(\hat{\boldsymbol{\beta}})}{\partial \beta_{oj}}=\frac{1}{N}\sum_{i=1}^{N}P^i_j(\hat{\boldsymbol{\beta}})-\frac{1}{N}\sum_{i=1}^{N}\triangle^i_j=0 \qquad (19)$$
$$j=1\cdots J$$

$$\frac{\partial L_2(\hat{\boldsymbol{\beta}})}{\partial \beta_k}=\frac{1}{N}\sum_{i=1}^{N}P^i_j(\hat{\boldsymbol{\beta}})X^i_{jk}-\frac{1}{N}\sum_{i=1}^{N}\triangle^i_j X^i_{jk}=0$$
$$k=1\cdots K \qquad (20)$$

この場合も[P 1]と同様にNewton－Raphson法で計算できる。式(19)および式(20)のそれぞれにおいて、第2項は各々サンプルより計算される市場占有率およびk番目説明変数の平均値を表していることに留意する。すなわち、

$$\frac{1}{N}\sum_{i=1}^{N}\triangle^i_j=\frac{n_j}{N}=r_j \qquad (21)$$

$$\frac{1}{N}\sum_{i=1}^{N}\triangle^i_j X^i_{jk}=\overline{X}_k \qquad (22)$$

ここで、 r_j ：選択肢jの市場占有率
n_j ：選択肢jが選ばれた回数
$\overline{X}_k$ ：k番目説明変数の平均値

3．情報量に基づくモデルの仮説検定

(1) エントロピーと情報量

パラメータ推定問題[P 1],[P 2]によって得られる選択確率は、説明変数$\mathbb{X}^i$が与えられたという件の下で選択肢a_jを選択する条件付確率であるのでこれを$P^i_j=Pr(a_j \mid \mathbb{X}^i)$と表そう。このとき、[P 1]によって推定されたパラメータを持つ選択確率による選択パターン推定に伴う不確実性は、式(13)より、次の条件付エントロピーで表せる。

$$H(A \mid X)=\sum_{x^i\in X}Pr(\mathbb{X}^i)H(A \mid \mathbb{X}^i)$$

$$=-\sum_{i=1}^{N}\frac{1}{N}\sum_{a_j\in A}Pr(a_j \mid \mathbb{X}^i)\ln Pr(a_j \mid \mathbb{X}^i) \qquad (23)$$

同様に[P 2]については式(13)より次式を得る。

$$H'(A \mid X) = -\sum_{i=1}^{N} \frac{1}{N} \sum_{a_j \in A} \triangle_j^i \ln Pr(a_j \mid \mathbb{X}^i) \tag{24}$$

式(23),(24)は、説明変数$\mathbb{X}^i$に関する情報が得られた後の事後エントロピーを表している。$\mathbb{X}^i$が得られないとき、我々は選択確率として$Pr(a_j)$を仮定するであろう。このときの事前エントロピーは次式で与えられる。

$$H(A) = -\sum_{a_j \in A} Pr(a_j) \ln Pr(a_j) \tag{25}$$

したがって、$\mathbb{X}^i$の情報を得てのエントロピーの減少分は、すなわち、獲得情報量は次式で定義できる。

$$I(A, X) = H(A) - H(A \mid X) \tag{26}$$

式(26)はShannonの情報量である。事前確率が特に次式で与えられる場合を考えてみよう。

$$Pr(a_j) = \frac{1}{J} \tag{27}$$

あるいは

$$Pr(a_j) = \sum_{i=1}^{N} Pr(\mathbb{X}^i) Pr(a_j \mid \mathbb{X}^i)$$

$$= \frac{1}{N} \sum_{i=1}^{N} Pr(a_j \mid \mathbb{X}^i) = q_j \tag{28}$$

このとき、式(26)は次式のように表現できる。

$$I(A, X) = \frac{1}{N} \sum_{i=1}^{N} \sum_{a_j \in A} Pr(a_j \mid \mathbb{X}^i) \ln \left\{ \frac{Pr(a_j \mid \mathbb{X}^i)}{Pr(a_j)} \right\} \tag{29}$$

式(29)のもつ意味を図-1を用いて説明する。今、ある確率パターン$\mathbb{Y}^i = \{Y_j^i\}$に対応したエントロピー関数を次式で定義する。

$$F(\mathbb{Y}^i) = -\Sigma_j Y_j^i \ln Y_j^i \tag{30}$$

また、$Pr(\mathbb{X}^i) = P^i$, $Pr(a_j \mid \mathbb{X}^i) = P_j^i$ とおく。このとき

$$H(A) = F\left(\sum_{i=1}^{N} P^i P_j^i\right) \tag{31}$$

であり、また

$$H(A \mid X) = \sum_{i=1}^{N} P^i F(\mathbb{P}^i) \tag{32}$$

Fは凹関数であるので、次式が成立する。

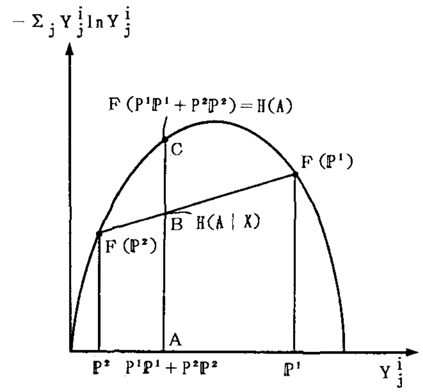

図-1　情報量の概念図

$$F\left(\sum_{i=1}^{N} P^i P_j^i\right) \geq \sum_{i=1}^{N} P^i F(\mathbb{P}^i) \tag{33}$$

したがって、N=2については図-1のような関係が成立しており、I(A, X)は線分BCに他ならない。事前情報による不確実性ACは、情報を得ての不確実性線分ABに減少し、したがって線分BCが獲得情報量I(A, X)を表す。

次に1回のサンプルに伴う情報量を定義しよう。事前確率を$Pr(a_j) = (1/N)\Sigma_i \triangle_j^i$とおくならば、式(25)は$H(A) = -(1/N)\Sigma_i \Sigma_j \ln Pr(a_j)$で与えられる。このときの条件付エントロピーが式(24)で与えられることにより、単純サンプリングに伴う経験情報量が次のように定義できる。

$$I'(A, X) = \frac{1}{N} \sum_{i=1}^{N} \sum_{a_j \in A} \triangle_j^i \ln \left\{ \frac{Pr(a_j \mid \mathbb{X}^i)}{Pr(a_j)} \right\} \tag{34}$$

単純サンプルデータにより推定される選択確率が正しいものと仮定すると、このとき、その選択確率から導かれる個人の0-1選択パターンは無数にあり、したがって、そのときの情報量I'(A, X)も無数に出現するランダム変量となる。

I'(A, X)は次に示す興味ある性質を持つ[2]。

① I'(A, X)は期待値がI(A, X)で分散が次式で与えられる正規分布に従う。

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} \left\{ \sum_{a_j \in A} Pr(a_j \mid \mathbb{X}^i) [\ln(Pr(a_j \mid \mathbb{X}^i) / Pr(a_j))]^2 - \left[\sum_{a_j \in A} Pr(a_j \mid \mathbb{X}^i) \ln(Pr(a_j \mid \mathbb{X}^i) / Pr(a_j))\right]^2 \right\} \tag{35}$$

②$2NI'(A,X)$はχ^2分布する。

次に$I(A,X)$，$I'(A,X)$を用いて次の指標を定義し、これらを不確実性係数と呼ぶことにする。

$$EU^2=\frac{I(A,X)}{H(A)} \tag{36}$$

$$U^2=\frac{I'(A,X)}{H(A)} \tag{37}$$

EU^2,U^2は図－1に示した例によって明らかなように、$\mathbb{X}^i$を得ての選択確率推定が、事前情報に基づく不確実性をどの程度減少させたのかの割合を示している。EU^2は不確実性の減少割合の期待値を示す一方、U^2は単純サンプリングに基づくそれを表している。また、U^2は一般に用いられている尤度比指数ρ^2に等価であることを示すことができる。

（2）情報量に基づく仮説検定

今、帰無仮説として「選択確率は個人iに無関係に定まる」を考えてみよう。このことは、選択確率として、式(27)あるいは(28)を用いることを意味する。Hauserは、この仮説検定のためには、①有意性検定、②有効性検定、そして③モデルの精度 の検定を行なうことを提案している。これらの各種検定法は、本研究で導出した指標を用いるならば、次のように説明できる。

①有意性検定

$2NI'(A,X)$がχ^2分布することを利用して、帰無仮説$Pr(a_j)$から推定確率$Pr(a_j \mid \mathbb{X}^i)$がいかに乖離するかを検定できる。$2NI'(A,X)$が大きな値をとれば、帰無仮説$P(a_j)$は棄却される。したがって、推定確率$P(a_j \mid \mathbb{X}^i)$が有意であるためには、$2NI'(A,X)$が想定された有意水準における$\chi^2$値よりも大でなければならない。ただし、この検定力は非常に弱く、多くの場合(データ数が適度の大きさを持っていれば)仮説は棄却される。

②有効性検定

不確実性係数EU^2,U^2は推定モデルがどの程度不確実性を説明しているのかを表している点で寄与率R^2に近い指標である。これらの指標は、推定確率が真の確率に一致しているのならば、あるいはP^i_jがサンプル値と全く同じ値を与えるならば1.0となる。また、逆にすべての選択肢に等確率を与えるようなモデルの場合には0.0となる。たとえば、今、(0，1)という選択データが得られているとし、そのときの推定確率が$\mathbb{P}^1=(0.4,0.6)$と $\mathbb{P}^2=(0.2,0.8)$の2つを得たとしよう。このとき、$\mathbb{P}^1$よりは$\mathbb{P}^2$の方が2番目の選択肢が選ばれると判断するのに適切である。事実、$\mathbb{P}^1$によるエントロピーは$\mathbb{P}^2$の場合のそれより大きく、したがって、U^2値は後者の方が大きな値をとる。なお、異なる説明変数集合を持つ複数個のモデルを比較する場合には、各々のEU^2やU^2を計算することによってモデルの有効性の比較が可能である。

③モデルの精度の比較

$I'(A,X)$は期待値が$I(A,X)$で分散がσ^2の正規分布に従う。このことは、もし推定された選択確率が正しいものとし、それによって発生される個人の0－1選択パターンが観測されるとき、その0－1選択パターンより計算される情報量は$I(A,X)$で与えられる値が最も頻度高く生じ、$I(A,X)$より離れるに従って、$I'(A,X)$を結果として与える0－1選択パターンの頻度が少なくなることを意味している。したがって、単純サンプルより計算される$I'(A,X)$の値が$I(A,X)$に近いほど推定された確率は精度が高いと判断できる。言い替えれば、$I'(A,X)$と$I(A,X)$が同じとなるような推定確率を得たならば、それによって予測される0－1選択パターンは、他のどの選択確率で予測される0－1選択パターンよりも頻度が高い。逆に、$I'(A,X)$が$I(A,X)$と大きく離れているならば、それはたまたま生じた0－1選択パターンを観測したか誤った確率を得たかのどちらかであり、再度データを収集するか、あるいは説明変数を追加して別の選択確率を得る努力をすべきであることを示唆している。なお、$I'(A,X)$が$I(A,X)$とどれ位離れているかを統計的に検定するには、$I'(A,X)=I(A,X)$という仮説を正規確率分布を用いて検定すればよい。

4．計算例

（1）少数サンプル2肢選択の例

情報量を基準にしたモデルの検定を直感的に理解するため、2選択肢でデータ数5の場合の例を考えてみよう。個人の0－1選択パターンを表－1に示し、異なる2つのモデルによる推定確率を表－2、表－3に示めす。市場占有率は$r_1=0.6, r_2=0.4$である。的中率はモデル1が100%、モデル2が80%であるが、個人1，2についてはモデル2のほうが実際の状況をよく説明していると思われ、どちらのモデルを採用するかは判断に迷うところである。この問題に対して、情報量基準によるモデルの検定を実行してみる。なお、事前確率は市場占有率を用い

る。このときの各種検定指標を表－4に示す。

まず、検定の第1段階はχ^2有意性検定であるが、この例ではあまりにもデータ数が少ないので、両者とも「事前確率分布と差はない」という帰無仮説を棄却できない。本来ならば、この時点でモデルを再検討するが、この場合は説明のための単なる例にすぎないので、χ^2検定の結果は問題にせず次に進むことにする。

第2段階ではU²値によるモデルの有効性を検討する。モデル1は事前確率の不確実性を41％説明し、モデル2は48％説明し、したがって、両者とも市場占有率を推定確率に使用するよりも調査データを説明するのには有効であると判断できる。また、モデル2の方がモデル1よりも有効である。

さて、有効性検定は1回の調査データに基づくモデルの有効性を議論するのみである。モデルの良さの別の尺度は、得られた推定確率が最も高い尺度で生じるU²値を保証するものであるかという点であり、このため、I'(A,X)とI(A,X)がどれ位近いかを判断する必要がある。この検定がモデルの精度の検定である。有意水準20％(信頼区間80％)でモデル1はI'(A,X)＝I(A,X)という仮説が棄却されるが、モデル2は棄却できない。

以上の検定結果より、的中率は低いがモデル2を採択すべきであると結論できる。

(2)多数サンプル3肢選択の例

Daganzoによって与えられた3径路選択問題の50人のデータを用いたパラメータ推定とその検定の例を示す。このデータでは、個人ごとの各径路走行時間t_j^iと実際の選択径路がデータとして与えられている[5]。ここで次のような2つのモデルを想定する。

(モデルA)定数項をもつ線形効用モデル

各径路の測定可能効用を$V_j^i=\theta_{0j}-t_j^i$とすると

$$\alpha V_j^i=\alpha(\theta_{0j}-t_j^i)=\beta_{0j}+\beta_t t_j^i \qquad (38)$$

$$j=1,2,3$$

このとき、β_{0j}はそれぞれの相対的な差が選択確率に影響を与え絶対量は直接には関係しない。したがって$\beta_{03}=0$と固定する。

(モデルB)定数項なしの線形効用モデル

各径路の測定可能効用を走行時間のみで表現する。すなわち、

$$\alpha V_j^i=\beta t_j^i \qquad j=1,2,3 \qquad (39)$$

これらのパラメータ推定の結果を表－5に示す。さ

表-1　選択結果

i＼j	1	2
1	1	0
2	1	0
3	0	1
4	0	1
5	1	0

表-2
モデル1の推定確率

i＼j	1	2
1	0.8	0.2
2	0.8	0.2
3	0.4	0.6
4	0.4	0.6
5	0.6	0.4

表-3
モデル2の推定確率

i＼j	1	2
1	0.9	0.1
2	0.9	0.1
3	0.4	0.6
4	0.4	0.6
5	0.4	0.6

表-4　検定指標 その1

指標＼モデル	1	2
Hit.R	1.00	0.80
I'(A,X)	0.277	0.324
I(A,X)	0.069	0.139
2NI'(A,X)	2.77	3.24
U²	0.412	0.481
EU²	0.103	0.207
$\nu=\dfrac{I'-I}{\sigma/\sqrt{N}}$	1.318	0.902

表-5　パラメータ推定結果

β_{01}	β_{02}	β_t	β
-0.314	0.101	-0.349	-0.357

表-6　検定指標 その2

P(a_j)	等確率		市場占有率	
指標＼モデル	A	B	A	B
Hit.R	0.70	0.72	0.70	0.72
I'(A,X)	0.443	0.432	0.292	0.288
I(A,X)	0.443	0.432	0.292	0.307
2NI'(A,X)	44.3	43.2	29.2	28.8
U²	0.403	0.393	0.308	0.296
EU²	0.403	0.393	0.308	0.316
$\nu=\dfrac{I'-I}{\sigma/\sqrt{N}}$	0.0	0.0	0.0	-0.069

て、これらの結果を用いて各モデルの検討を行なう。このときの各種指標は表－6となる。なおここでは、事前確率を等確率の場合および市場占有率の場合の両方についての各種指標の計算を行なった。これらより次のことが言えよう。

(i) χ^2値は事前確率のとり方に関係なくA，Bの両モデルとも大きく、事前確率と差はないという仮説は成立しない。

(ii)事前確率を等確率とおいた場合：モデルAの方が不確実性係数U^2が大きく、より多くの不確実性(40.3%)を説明している。$I'(A,X)$は$I(A,X)$に一致しており、この点でも十分信頼のおけるモデルであると言える。ただし、モデルBのU^2も39.3%であり、両者の差はほとんどない。

(iii)事前確率を市場占有率とおいた場合：モデルAは$I'(A,X)$と$I(A,X)$が一致しており、データ収集を反復した場合にも十分適用できるであろう。ただし、U^2は30.8%であり、事前確率を等確率とおいた場合よりモデルにより説明される不確実性が約10%減少している。一方、モデルBはU^2の値はモデルAと同程度の29.6%であるが$I'(A,X)$と$I(A,X)$が一致しておらず、繰り返し生ずる0－1選択パターンの予測への適用に対し、モデルAほど強く推薦できるものではない。しかし、有意水準20%とした場合には両者とも仮説を棄却できず、その意味では両モデルの差はほとんどない。

これらの結果から、この例については選択肢固有ダミーを使用したモデルAを採用するのが望ましいと結論することができる。

5．まとめ

本研究では、まず、共役性概念に基づく新しいパラメータ推定方法を提案した。この方法によれば、選択に伴う不確実性(エントロピー)をパラメータ推定プロセスの中で直接計測することが可能になる。また、推定された個人レベルでの選択確率の良さを計量的に評価するために、パラメータ推定に伴うエントロピーを利用する方法を提案した。すなわち、選択確率を説明変数に関連づけることによる選択パターンの不確実性の減少の度合を情報量として定義し、この情報量を尺度としてモデルの良し悪しを判断しようとする方法である。結果として得られた情報量はHauserの提案した指標と一致するが、Hauserの誘導した経験情報量はその成立根拠が不明であり、また、期待情報量についても本研究のようにパラメータ推定プロセスと密接に関連した形で誘導されたものではなかった。従来、モデルの評価は、尤度比χ^2検定や尤度比指数で行なわれてきたが、情報量を用いた検定法はこれらの検定法を包括するものであり、従来のχ^2検定を"有意性検定"またρ^2値は"有効性検定"と位置づけることに加えて、さらに"精度の検定"も1つの尺度として行えることが可能になった。精度の検定は全く新しい考えかたで、従来用いられてきた的中率の概念をより発展させ一般化した概念であると考えてよかろう。すなわち、選択は反復して行なわれるという事実を考慮して、最も頻度が高く生じる状況を説明する選択確率が得られたかどうかを判断する尺度が、ここでいう"精度"の概念である。本研究では、線形効用関数モデルを前提として議論を展開したが、より一般的な非線形効用関数モデルの場合にも適用可能である。また、精度の検定において有意水準を20%としたが、有意水準を決定することは非常に複雑な問題であり、今後の分析結果に負うところが大きい。

【参考文献】

(1) Stopher,P.R. and A.H.Meyburg (1979) Survey sampling and multivariate analysis for social scientists and engineers, Lexington Books, Toronto.

(2) Hauser,J.R. (1978) Testing the accuracy, usefulness, and significance of probabilistic choice models : an information-theoretic approach, Operations Res., Vol.26 No.3, pp.406－421.

(3) Williams,HCWL (1977) On the formation of travel demand models and economic evaluation measures of user benefit, Environment and Planning A, Vol.9, pp.285－344.

(4) 宮城俊彦、加藤晃 (1984) ランダム効用理論を基礎とした交通統合モデル、土木計画学研究・論文集1、pp.99－106.

(5) Daganzo,C. (1979) Multinomial probit : The theory and its application to demand forecasting, Academic Press, New York.