

選択型SPデータによる嗜好の異質性の把握に関する基礎的分析*
A Preliminary Analysis on the Reproducibility of Individual Taste Heterogeneity
Using Multiple Stated Choice Data*

市岡秀之**・倉内慎也***・山本俊行****・森川高行*****

By Hideyuki ICHIOKA**・Shinya KURAUCHI***・Toshiyuki YAMAMOTO****・Takayuki MORIKAWA*****

1. はじめに

ライフスタイルや価値観の多様化, それに応じた様々な交通サービスならびに交通施策の提案を背景として, 個人の嗜好や施策に対する反応の異質性の把握することは, 交通行動分析や需要予測において益々重要度を増している. そのような認識のもと, 異質性を明示的に考慮した分析手法やモデルが数多く開発されているが¹⁾, 分析に用いるデータについては, 1個人に対して1つの行動結果があるクロスセクショナルデータが主流であり, そこから嗜好の異質性や交通施策に対する反応の差異を正確に計量することは困難であると思われる. また, 個人の嗜好に異質性が存在するにもかかわらず, サンプル全体に嗜好の同質性を仮定して推定を行った場合, パラメータ推定値に不偏性や一致がないことも指摘されている²⁾.

個人の嗜好・選好に関する情報をより詳細に抽出できる調査手法として, 仮想の状況下における選好意思表示を尋ねるSP (Stated Preference) 調査がある. SP調査は, 一種の実験データであるため極めて操作性が高く, また, 一度の調査において同一個人から複数のデータを収集できるため, これまで嗜好や選好に関する情報を効率的に取得するための調査設計法や, 回答の信頼性などについて様々な研究がなされている³⁾. しかし, それらは嗜好の同質性を仮定したロジットモデルのようなシンプルなモ

デルによる分析を前提としたものがほとんどであり, 嗜好の異質性が及ぼす影響や, 異質性を考慮したモデルの適用を前提とした場合の調査設計法については明らかにされていない.

そこで, 本研究では, 特に選択形式のSPデータに対して離散選択モデルを適用する場合を想定し, 嗜好や選好を個人レベルでいかに正確に計測できるかという点に着目して分析を行う. 分析においては, 異質な嗜好を持った仮想個人, および設問数等が異なる様々なSPデータをシミュレーションにより人工的に生成し, いくつかの分析手法やモデルの推定精度の比較を通じて, 最終的に, SPデータを用いて個人レベルでの嗜好や選好を把握するための調査設計およびモデル分析上の指針を提示することを目的とする.

2. 分析手法と着眼点

(1) シミュレーションによるデータ生成

本研究では, シミュレーションによって生成した人工データを用いる. これにより, 仮想個人やLOS について様々な設定ができ, SPデータの信頼性の問題⁴⁾のうち, 被験者に起因する「信憑性」の問題(政策操縦バイアスや正当化バイアス等)を排除し, 分析者が制御可能な調査設計に関連する「安定性」の問題(設問数や提示するLOS等)に焦点をあてることができる. また, 真の嗜好(パラメータの真値)が既知であるため, モデル推定精度の詳細な検証も可能となる.

本研究では, 選択肢数 2, 属性数 1 という極めてシンプルな選択状況を想定する. また, 被験者数 $n=1000$ 人とし, 被験者当たりの設問数 t 個を操作変数として分析を行う. 具体的なデータ生成プロセスは以下の通りである.

Step 1: 仮想個人 n について, 定数項 μ_n 及び属性

*キーワード: 交通行動分析, SP データ

**正員, 修(工), 株式会社片平エンジニアリング
(東京都文京区小石川 2-22-2, TEL: 03-5802-2041,
E-mail: ichioka@katahira.co.jp)

***正員, 修(工), 名古屋大学大学院工学研究科
(名古屋市千種区不老町, TEL: 052-789-3565,
E-mail: kurauchi@civil.nagoya-u.ac.jp)

****正員, 博(工), 名古屋大学大学院工学研究科
(E-mail: yamamoto@civil.nagoya-u.ac.jp)

*****正員, Ph.D., 名古屋大学大学院環境学研究科
(E-mail: morikawa@nagoya-u.jp)

に対する嗜好 α_n を乱数等を発生させて決定。

Step 2: 属性値について、2 肢選択であるため属性値の差 X を考え、 t 個の設問に対応する X を作成。

Step 3: 仮想個人 n の効用関数について、2 肢選択であるため以下のような効用差を考え、誤差項の差 $\Delta\varepsilon$ に標準正規乱数を発生させ効用差を計算。

$$\Delta U_{nt} = \alpha_n + \beta_n \Delta X_{nt} + \Delta\varepsilon_{nt} \quad (1)$$

なお、 ε_{nt} は個人および設問によらずランダムなホワイトノイズを表している。

Step 4: $U_{nt} > 0$ であれば代替案 1 を、そうでなければ代替案 2 を選択するとして、 $n \times t$ 個の選択結果のデータを生成。

(2) 分析の着眼点と分析手法

本研究では、以下の 3 点に着目して分析を行う。

(a) LOS の設定

本稿では LOS の設定に関して次の 2 点に着目して分析を行う。

1 点目は、LOS のばらつきの問題である。例えば、2 肢選択において、LOS は図 1 のように様々な設定が可能となる。case 1 のように、選好無差別な点付近を詳細に尋ねた場合、行動変化が生ずる点を詳細に把握できるが、両代替案の差異が微小であるため、回答が意思決定環境に付随するノイズに影響を受けやすい。逆に、case 3 のような場合、回答の安定性がある一方、行動変化に関する情報としては決して有益であると言えない。

2 点目は、LOS の偏りの問題である。繰り返し型の SP 調査では、設定する LOS の対称性もモデル推定結果に影響を及ぼすものと考えられる。例として

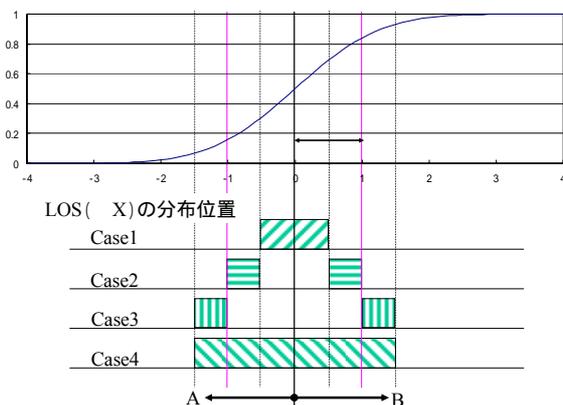


図 1 選択確率と設定する LOS の分布

ある個人に 4 つの設問を提示したとしよう。仮に、2 つの設問において代替案 1 を、残り 2 つの設問では代替案 2 を選んだとすれば、どのようなサービスレベルで行動変化が生ずるのかおおよそ把握することが可能となる。一方、4 つの設問全てにおいて代替案 1 を選択するようなサービスレベルを提示した場合、行動変化が生ずる点を把握することは不可能である。また、代替案 1 を 4 回中 3 回選択するようなケースでは、行動変化が生ずる点は把握できるが、代替案 1 を選んだ 3 つの設問の幾つかは冗長であり、調査として効率的とは言い難い。

以上の 2 点を鑑み、本研究では、LOS のばらつきについて、 $\Delta\varepsilon$ の標準偏差 $\sigma = 1.0$ を参考に以下の 4 つの Case を想定し、各々の範囲内で t 個の 1 様乱数を発生させて ΔX を決定した (図 1 参照)。

Case1: 選択確率 0.5 を中心に $-0.5\sigma \sim 0.5\sigma$ の間

Case2: $-\sigma \sim -0.5\sigma, 0.5\sigma \sim \sigma$

Case3: $-1.5\sigma \sim -\sigma, \sigma \sim 1.5\sigma$

Case4: $-1.5\sigma \sim 1.5\sigma$

その際、LOS の偏りの影響も把握するため、各ケースについて、図 1 の (A の領域からの抽出個数) : (B の領域からの抽出個数) = $(t : 0), (t-1 : 1)$ 、というように、抽出比率を変えて ΔX を設定した。

(b) 異質性存在下でのモデル推定精度

前述のように、SP 調査設計に関するガイドラインについては、嗜好の同質性を仮定したシンプルなモデルの適用を前提としたものがほとんどである。そこで、嗜好に異質性が存在する場合を想定し、嗜好の同質性を仮定したモデルと異質性を明示的に考慮したモデルで推定精度の比較を行い、モデルの違いによる嗜好の現況再現性を検証する。具体的には、データ生成プロセスの Step 1 において、 α_n および β_n を標準正規乱数により決定し、作成した選択データに対して、パラメータ値をサンプル全体で共通にした固定係数モデルとランダム係数モデルをそれぞれ適用し、パラメータ推定結果を比較する。

(c) 異質性を捉えるための設問数と分析手法

繰り返し選択型の SP データでは、1 個人当たりの設問数が多ければ、個人パラメータを推定することが可能である⁵⁾。しかし、設問数が多ければ回答の信憑性が低下し、アンケート回収率も下がるなどの問題が生ずる。そこで、まず設問数を変化させ

たデータを幾つか生成し、ランダム係数モデルを適用した場合の推定精度の差を検討する。次いで、個人パラメータを推定する方法として、個人ごとにモデルを推定する方法（方法1）、およびランダム係数モデルを適用し、推定されたパラメータを個人ごとの選択結果に応じて重み付けして個人パラメータを求める方法⁶⁾（方法2）を適用し、いずれがより個人レベルでの選好や嗜好を再現できるか検討する。

3. 分析結果

以下に分析結果を示す。なお ΔX は一様乱数に基づいて生成するため、推定結果に少なからず影響が及ぶものと推察される。そこで、乱数の影響を抑えるために、 ΔX を各設定に対して10ないしは5セット生成し、各々に対して選択データを作成しモデル推定を行った。

(1) 異質性が無い場合のLOSの設定の影響

まず、純粋にLOSの設定の影響を把握するために、嗜好に異質性が無い場合として、パラメータの真値を $\alpha = 1.0$, $\beta = 1.0$ として分析を行った。図2は2章(1)(a)で述べた設定状況ごとの推定結果を示している。 α , β の推定値の平均値はいずれも真値に近い値を示しており、LOSのばらつきや偏りによってバイアスは生じないことが確認された。また、標準偏差を見ると、0:8,8:0のように極端な偏りがある場合を除いて推定されたパラメータの値は安定している。調査設計時点においては、一般に、選好無差別であろう点を把握することは困難であるため、case 4のように、現実的な範囲内で属性値が一様に分布するようにLOSを設定することが賢明であると思われる。

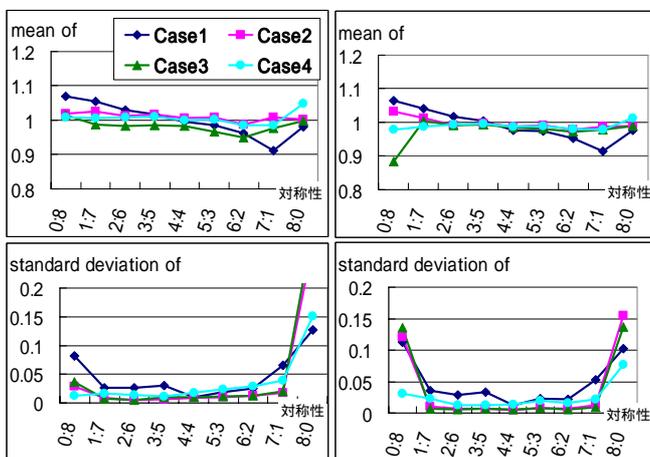


図2 及び の10回の推定値の平均値（上段）と標準偏差（下段）（ : 左列, : 右列）

(2) 異質性存在下でのモデル推定精度

異質性が存在する状況として、 α_n および β_n の真値として独立な標準正規乱数を発生させ、(1)と同様の分析を行った。図3に固定係数モデルの推定結果を、図4にランダム係数モデルの推定結果を示す。

固定係数モデルでは、 α , β 共に過小推定となっており、特に α についてはケースごとにシステムティックな差異が見受けられる。これは以下の理由によるものと考えられる。各個人に異質性が存在する場合、真の効用関数は以下ようになる。

$$U_{nt} = \alpha_n + \beta_n X_{nt} + \varepsilon_{nt} \\ = \bar{\alpha} + \bar{\beta} X_{nt} + (\varepsilon_{nt} + (\alpha_n - \bar{\alpha}) + (\beta_n - \bar{\beta}) X_{nt}) \quad (2)$$

この時、推定されるパラメータは α_n 及び β_n の平均値 $\bar{\alpha}$, $\bar{\beta}$ である。誤差項 $\Delta\varepsilon_{nt}$ は標準正規分布と仮定して推定を行っているが、上式のように、 α , β の異質性が誤差項と識別不能であるため、誤差項を過大評価してしまう。一方、推定されるパラメータは、式(3)で表されるようにスケール σ を含む β' である。

$$\Phi(\beta' \Delta X) = \Phi\left(\frac{\bar{\beta} \Delta X}{\sigma}\right) \quad (3)$$

つまり、 $\sigma=1$ と仮定して推定を行うが、実際は、パラメータ推定値は誤差項の標準偏差を含んだ形で推定されるため、誤差項が過大に評価される場合、推定されるパラメータ β' は過小推定となる。

α の推定値の Case ごとの違いについては、式(2)の $(\beta_n - \bar{\beta}) X_{nt}$ の部分が、定数項の推定に関わってくるためであると考えられる。定数項は、本来観測シェアと予測シェアを一致するように決定される。つまり、 ΔX の絶対値が変化する事により、 $(\beta_n - \bar{\beta}) X_{nt}$ の平均値が変動し、その平均値が定数項の推定値に反映されたものと考えられる。4:4における ΔX の絶対値の平均値を Case 別に計算すると、Case1 は 1.00, Case2 は 1.36, Case3 は 2.21, Case4 は 1.58 であり、絶対値が大きいほど α の推定値も大きくなっている事がわかる。

一方、ランダム係数モデルではバイアスが改善されている事がわかる。異質性が存在するため、必然的にパラメータ推定値の安定性は低下するものの、Case1を除いてLOSの設定に関わらず比較的安定し

た推定結果が得られた。

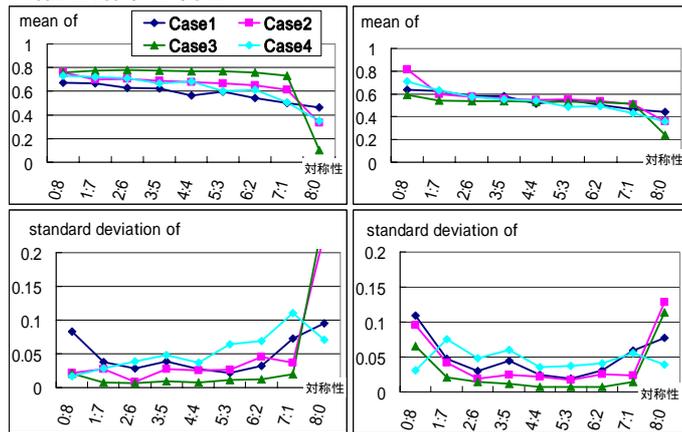


図3 固定係数モデルの推定結果

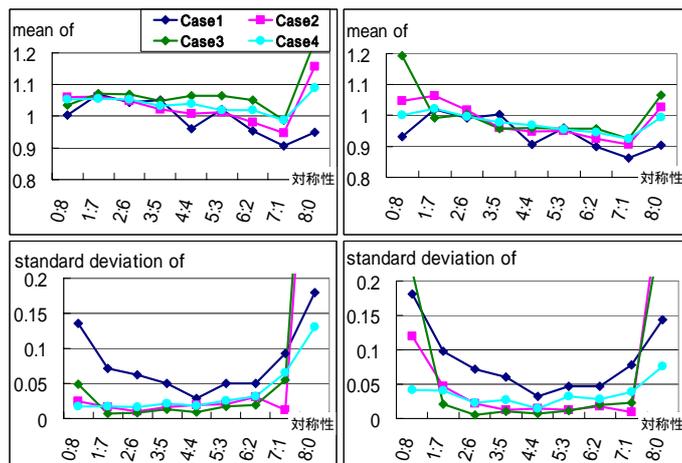


図4 ランダム係数モデルの推定結果

(3) 異質性を捉えるための設問数と分析手法

設問数が推定結果に及ぼす影響を把握するために、異質性存在下で設問数を2~32まで変化させ、ランダム係数モデルを推定した。なお、真の嗜好は α , β 共に標準正規乱数により決定し、LOSのばらつきはcase4, 偏りは4:4としてデータ作成した。

図5より、サンプルが1000人の場合、設問数が2個では不安定な解が得られるが、4以上であればある程度安定した推定値が得られる事が確認された。

次に、前章(2)(c)で述べた2つの個人パラメータの推計方法について検証を行う。個人パラメータの推定には1個人から多くのデータが必要となるため、ここでは1個人あたり最大64個のデータを作成した。本研究では、各パラメータの個人ごとの値が既知であるため、評価指標としては、推計された個人パラメータと真値との相関係数及びRMSEを用いた。

表1より、方法1については、1個人あたり64個のデータがない限り再現性は極めて低く、これは非現実的である。一方、方法2は方法1よりもかなり再現性は高いが、1個人あたりのデータ数が少なく

なるにつれ、再現性の低下率は大きくなっている。これは、方法2では選択結果に基づいて嗜好の重み付けを行うため、2肢選択の場合の重み付けパターンは最大で2²になるためである。それ故、方法2により個人パラメータを推計するためには、1個人あたり最低16回の繰返し回数が必要であり、調査設計においては信憑性の問題についてかなりの配慮が必要となる。

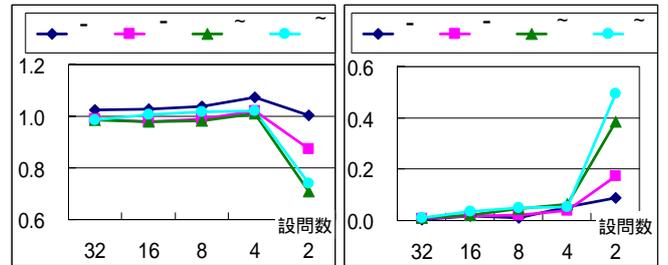


図5 設問数と推定値の平均値(左)・標準偏差(右)

表1 真値との相関係数及びRMSE

		方法1			方法2		
		設問数			設問数		
		64	32	16	64	32	16
相関係数	n	0.884	0.234	0.193	0.933	0.896	0.757
	n	0.849	0.298	0.216	0.933	0.896	0.757
RMSE	n	0.017	0.230	1.745	0.011	0.014	0.021
	n	0.026	0.233	2.441	0.012	0.014	0.021

4. おわりに

本研究では、嗜好の異質性を把握するために選択形式のSPデータ着目したシミュレーションスタディを行い、異質性を個人レベルで把握するための分析手法や調査設計について幾つかの知見を得た。しかし、状況設定は極めてシンプルなものであるため、今後、選択肢数や属性数が増えた場合に分析を拡張し、より一般的な知見を得る必要がある。

参考文献

- 1) レビューとして, McFadden, D.: Disaggregate behavioral travel demand's RUM side: a 30-year retrospective, In Hensher, D.A. (ed.) Travel Behaviour Research: The Leading Edge, Pergamon, pp.17-63, 2001.
- 2) Chamberlain, G.: Analysis of covariance with qualitative data, Review of Economic Studies, Vol.47, pp.225-238, 1980.
- 3) 例えば, 藤原章正, 杉恵頼寧: 選好意識調査設計の手引き, 交通工学, Vol.28, No.1, pp.63-71, 1993.
- 4) 森川高行: ステイテッド・プリファレンス・データの交通需要予測モデルへの適用に関する整理と展望, 土木学会論文集, No.413/IV-12, pp.9-18, 1990.
- 5) 佐々木邦明: 潜在的評価構造の差異を考慮した離散型選択モデル, 京都大学博士論文, 1997.
- 6) Train, K. E.: Discrete Choice Methods with Simulation, Cambridge University Press, 2003.