

ベイズ推定法を用いた離散選択モデルに関する基礎的研究*

A Preliminary Research on the Discrete Choice Model using Bayes Estimation Method*

大園渉**・稲田有香***・室町泰徳****

By Wataru OZONO**・Yuka INADA***・Yasunori MUROMACHI****

1. はじめに

近年、多くの分野でベイズ推定モデルの適用例が増えている¹⁾。従来のベイズ推定では、分析方法に多大な負担がかかったものの、マルコフ連鎖モンテカルロ(以下MCMC)法によりこの問題点を克服されており、近年、国内外で様々な研究されている。例えば、MCMC法によるベイズ推定法の適用には、我が国では目的地点の選択行動⁴⁾、地価モデルの推定⁵⁾などの例があり、海外では、外科医の勤務地選択行動⁶⁾、エネルギー供給選択行動⁷⁾などの例がある。また、交通計画分野では複数データの統合による交通需要モデルの推定方法³⁾などに適用されている。しかしながら、交通手段選択に対するMCMC法によるベイズ推定モデルの適用、およびそのモデルの比較・検証を行っている研究はほとんど無い。

一方、著しいモータリゼーションやライフスタイルの多様化により、消費者の行動範囲が広域化・多様化が生じている。人々の行動が多様化していく中で、この影響を考慮していく意義は大きい。より詳細な選択行動を予測するためには個人間の異質性による影響を考慮する必要がある。その際、MCMC法を用いたベイズ推定モデルによる交通行動選択モデルの有用性は小さくないと考えられる。

本研究では、港北ニュータウンにおける買物行動調査結果をもとに、住民の買物行動における交通手段選択モデルをベイズ推定法で推定し、最尤法による多項プロビット(以下、MNP)モデルとの比較結果を示す。

2. ベイズ推定法の概要

ベイズ推定法では、パラメータ自体が確率変数であり、パラメータの分布を知ることができる。また、いかなる関数の最大化も必要としないという利点がある。ProbitモデルやいくつかのMixed Logitモデル(特に対数正規分布を持ったもの)では、シミュレートされた尤度関数の最大化は、数値解析的に難しい場合が多い。以下にベイズの概念について示す²⁾。

パラメータ θ を持つモデルについて考えるとき、研究者はこれらのパラメータの値についての初期信念を持ち、この考えを改善するためにデータを集める、という状況を想定する。ベイズ分析の下では、パラメータについての研究者の考えは、そのパラメータが取りうる全ての値に対する確率分布によって表される。データを集める前段階では、研究者の考えは、論理、直感、過去の分析に基づいて決定される。これらの考えは、事前分布 $k(\theta)$ 、すなわち θ についての確率密度として表される。

研究者は θ の値についての考えを改善するためにデータを集める。 N 個の独立した意思決定主体のサンプルを、分析者が観察する場合を考える。 y_n は意思決定者 n に対して観察された選択結果を表し、全サンプルの観察された選択は集合的に $Y = \{y_1, \dots, y_n\}$ と表すことにする。これらのサンプルの情報に基づいて、研究者は θ についての考えを変更、更新する。その更新された考えは、 $K(\theta|Y)$ と表示され、事後分布と呼ばれる θ についての確率密度である。この事後分布は、観測されたサンプルに含まれた情報を取り込んでおり、 Y に従属している。

事前分布と事後分布の間には、ベイズのルールによって規定される関係がある。 $P(y_n|\theta)$ を意思決定 n が結果 y_n を生起する確率とする。説明変数の表記は省略するが、この確率は、説明変数とパラメータを行動結果に関連付ける行動モデルである。観察されたサンプルの結果 Y の生起する確率は

$$L(Y|q) = \prod_{n=1}^N P(y_n|q) \quad (1)$$

となる。これは観測された選択結果に対する尤度関数

*キーワード：MCMC、ベイズ推定法、離散選択分析

**学生員、工修、東京工業大学大学院総合理工学研究科
人間環境システム専攻
(神奈川県横浜市緑区長津田町4259、
TEL045-924-5606、FAX045-924-5574)

***工修、株式会社ザイマックス
(東京都新宿区四谷2-8、
TEL03-3561-2361、FAX03-5361-2406)

****正員、工修、東京工業大学大学院総合理工学研究科
人間環境システム専攻
(神奈川県横浜市緑区長津田町4259、
TEL045-924-5606、FAX045-924-5574)

(対数変換は行っていない)であり、 θ の関数である。
 ベイズのルールは、研究者が θ についての考えを改善する仕組みを与える。条件確率のルールより、

$$K(\theta | Y)L(Y) = L(Y | \theta) k(\theta)$$

また、 $L(Y)$ は、 θ の周りにおける Y の周辺確率であり、

$$L(Y) = \int L(Y | \theta) k(\theta) d(\theta)$$

以上の2式より、(1)式は以下のように表すことができる。

$$K(\theta | Y) = L(Y | \theta) k(\theta) / L(Y) \quad (2)$$

この式がベイズのルールを表しており、事前分布と事後分布に対して適用することができる。

(2)式は、 Y の周辺確率 $L(Y)$ は θ に関して一定であることから、以下のように表すことができる。

$$K(\theta | Y) = L(Y | \theta) k(\theta)$$

事後分布は事前分布と尤度関数を掛け合わせたものに比例していることをこの式は示す。 $k(\theta)$ に具体的な確率分布を仮定することにより、解析的、あるいはシミュレーションにより $K(\theta | Y)$ の確率分布が推定できる。

3. MCMC法の概要

MCMC法は多変量の確率分布からサンプルを得るための手法である。その特徴は、離散・連続変数を問わず様々な分布に使える、非常に多変量の場合でも適用できることである。ベイズ統計学においては事後分布が多次元であることが多く、1000次元以上ということもある。その周辺事後密度 $K(\theta | Y)$ を数値積分により求めたり、それに基づく推論を行うことは難しい。このため、MCMC法により、事後分布からの確率標本 θ^t をサンプリングし、得られた確率標本を用いることにより、事後分布に関する要約や、事後分布に基づく推論を行う。

マルコフ連鎖には、適当な初期値からはじめて十分な回数を繰り返していくと、確率標本の分布が正則条件の下で不変分布に収束していくという性質がある。この不変分布が事後分布になるようにマルコフ連鎖を構成することにより、マルコフ連鎖の確率標本 θ^t を事後分布からの確率標本としていく⁷⁾。

本研究ではMCMC法にギブスサンプリングを用いている。以下、ギブスサンプリングアルゴリズムについて示す⁷⁾。

- (1) 初期値 $\theta^0 = (\theta_1^0, \dots, \theta_p^0)$ を決め $t=1$ と置く。
- (2) $i=1, \dots, p$ について
 $\theta_i^t \sim \pi(\theta_i | \theta_{-i}^t, x)$,
 $\theta_{-i}^t = (\theta_1^0, \dots, \theta_{i-1}^t, \theta_{i+1}^{t-1}, \dots, \theta_p^{t-1})$,
 を発生させる。
- (3) t を $t+1$ として(2)に戻る。

(2)、(3)を繰り返し、十分繰り返したときの θ を確率分布の標本とする。このとき、 $\theta^t, t=1, 2, \dots$ は推移核が、

$$K(q^{(t-1)}, q^{(t)} | x) = \prod_{i=1}^p p(q_i^{(t)} | q_{-i}^{(t-1)}, x)$$

であるマルコフ連鎖であり、正則条件の下で $t \rightarrow \infty$ のときに θ^t の分布が事後分布に収束する。

ギブスサンプラーでは一様乱数や正規乱数などの乱数を用いて計算できることが前提となっている。よって、これができない場合においてはメトロポリスヘイスティングスアルゴリズムを用いる必要がある。

4. アンケート調査概要と集計結果

本調査は神奈川県横浜市都筑区の港北ニュータウンで行った。港北ニュータウン買物行動調査の概要を表1に、主な質問項目を表2に示す。アンケートの配布は74町丁目ごとに行い回答がどこの町丁目の住民によってなされたものかを把握できるようにした。

表1 調査概要

調査対象地域	港北ニュータウンの第一地区・第二地区
調査期間	2006年11月下旬～12月上旬
調査方法	ポスティング + 郵送回収
配布数	6920 (町帳目の世帯数で比例配分し、各町帳目に配布)
回収	1374
回収率	19.9%
有効サンプル数	1366

表2 アンケート調査の主な質問

買物場所	最寄りのコンビニ
	タウンセンター2箇所 (センター北駅周辺・センター南駅周辺)
	ニュータウン内やその近辺の商業地・商業施設35箇所のうち買物頻度が高い順に、最大3箇所を選択
	都心やその他の有名な商業地10箇所のうち買物頻度が高い順に、最大2箇所を選択
	で回答した各買物場所についての、買物頻度・交通手段・購入品目・買物以外の活動
	通販の利用頻度と購入品目
	買物場所決定に影響を与えそうな6項目(まとめ買いのしやすさ・価格の安さなど)についての、各影響度合い
	属性 (性別・年齢・職業・世帯人数・家族構成・車所有台数・免許の有無・世帯年収など)

次に、買物行動を把握するため、センター北とセンター南を目的地とする買物行動について集計を行った。各交通手段のシェアを表3、買物以外の目的の割合を図1に示す。分担率では自動車が半数以上を占めており、次いで徒歩となっている。センター北・南とも駅周辺には百貨店・大型ショッピングセンターがあり、駐車場も整備されている。また図1に示すように、買物のみの行動をする割合は低く、特に外食・サービスが両目的地で高

表3 買物交通手段分担率

	センター北	センター南
自動車	56.6%	55.5%
自転車	8.0%	10.5%
徒歩	19.4%	15.4%
バイク	0.6%	0.8%
バス	3.9%	5.8%
鉄道	11.4%	12.0%

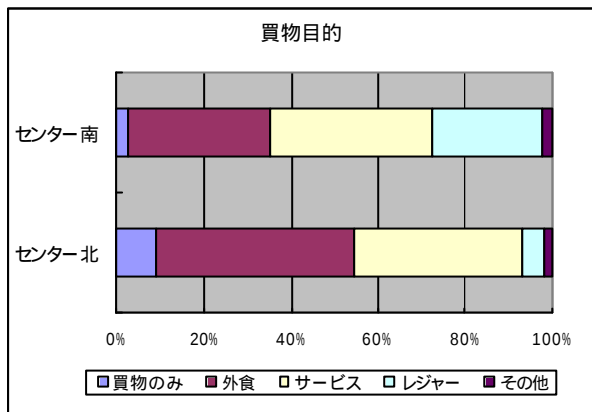


図1 買物の目的

いことが分かる。駅周辺にさまざまな施設が密集しているニュータウンの性質上、自動車や徒歩のような交通手段が、便利で融通の利く手段として選択されていると考えられる。

5. モデル推定結果と考察

アンケート調査結果をもとに、センター北およびセンター南を目的地とするサンプルを抽出し、自動車、自転車、徒歩、バイク、バス、鉄道の6選択肢による多項選択モデルを構築した。交通手段選択結果を被説明変数とし、各選択肢ダミー・所要時間・コスト(公共機関を使用するものと、そうでないもの)・性別・年齢・年収の11個を説明変数とした。また、第2章、第3章をもとに、ベイズ推定法を多項プロビットモデルに適用し、パラメータ推定を行った。MCMC法の収束判定であるGelman-Rubin統計量を図2に、モデル推定結果を表4に示す。

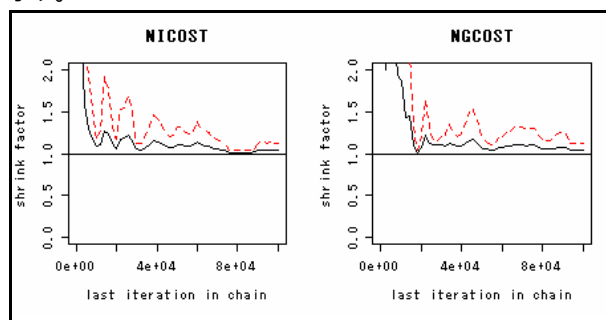


図2 Gelman-Rubin Statistic

まず、Gelman-Rubin 統計量とは、MCMC 法において発生させた各連鎖の推定値が収束しているかを判断するものであり、1.1 よりも小さければ有意となる。本研究におけるベイズモデルでも、図2に見るように全体的に1.1を下回っている。

次に、標本時系列プロットとは、繰り返し計算期間におけるパラメータの乱数を取る値を記録したものであり、収束していくと同じような範囲の乱数値を取るようになる。本研究では繰り返し計算期間を100000回行っており、図3においても状態空間をまんべんなくサンプリングしている様子が伺える。また、周辺事後密度とはパラメータ推定値の分布であり、推定値が取る値の程度を知ることができる。本研究の結果はほぼ正規分布となっており、一定の値に定まったことが分かる。以下、表4のモデル推定結果についての考察を行う。

- 最尤法によるMNPモデルとベイズモデルで、数値にかなりの相違が見られた。しかしながら、符合はほぼすべて同一となっている。
- 選択肢ダミーのパラメータに関しては、徒歩・自転車・自動車の値が相対的に高くなっている。これは、目的地への距離が近いこと、センターの大型店舗に付随する駐車場が多いため、これらを選択することの利便性が強く作用していると考えられる。
- 所要時間パラメータに関しては有意な負のパラメータとなっている。一方で、コストパラメータは予想に反して正の値となった。これは、目的地が近距離であり、各選択肢におけるコストがサンプル間でほぼ変わらないことが影響を及ぼしたと考えられる。
- $\bar{r}^2$ 値、的中率に関しては約50%という結果となった。ベイズモデルは的中率に偏りがあるが、最尤法によるMNPよりも全体的な中率は良い結果となった。

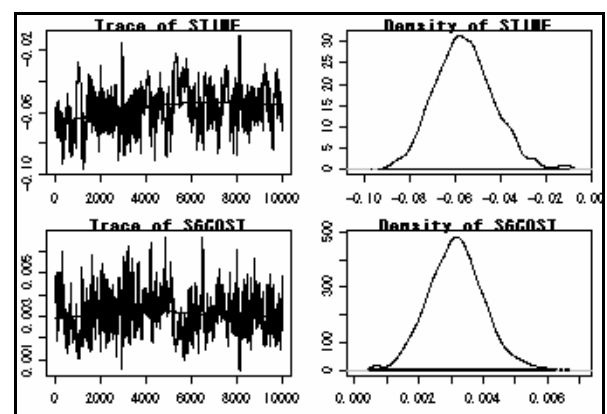


図3 標本時系列プロット、周辺事後密度

- ベイズモデルでのパラメータ分布を図3に示す。図からわかるように個人間の異質性が分かる。つまり、分布の最も高い部分の値を取るものもいれば、尾部の値を取るものもいるということである。この分布が広いほど個人にとって異質性が高いと考えられる。今回のモデルではすべてのパラメータにそれほど差は出なかったものの、定数項に関しては、大きいもので10程度の開きがあり、選択肢に対するの個人間異質性が表れていると考えられる。

6. おわりに

本稿では、買物行動調査結果をもとに、交通手段選択多項プロビットモデルを最尤法とベイズ推定法により推定し、各モデルのパラメータ推定値、モデルの有用性を検討した。ベイズ推定法によるパラメータ推定結果より個人間異質性が確認でき、これにより詳細な検討が可能となる。

なお、本研究では、買物場所をセンター北とセンター南の2つとして扱ったが、これら以外の目的地の影響が多々あると考えられる。またモデル構築は6選択肢で行ったが、選択肢の更なる細分化やNestedおよびMixed

ロジットモデルの検討を行い、ベイズ推定モデルの適用可能性、有用性と比較検討する必要があると考えられる。

参考文献

- 1) Rossi, P.E. and Allenby, G.M.: Bayesian statistics and marketing, Marketing Science, 22, pp.304-328, 2003
- 2) Train, K.: Discrete choice methods with simulation, Cambridge, 2003
- 3) 屋井鉄雄・森地茂・魚谷憲: 複数データの統合による交通需要モデルの推定方法, 土木学会論文集, 470/IV-20, pp.125-133, 1993
- 4) 北村菊池輝・藤井聡・北村隆一: 座標システムを導入した生活行動シミュレーションによる交通政策評価, 都市計画論文集, 36, pp.577-582, 2001
- 5) 古谷知之: ベイズ地理的加重回帰モデルの地価モデル推定への適用, 都市計画論文集, 39-3, pp.787-792, 2004
- 6) Bolduc, D., Fortin, B., and Gordon, S.: Multinomial probit estimation of spatially interdependent choices: An empirical comparison of two new techniques, International Regional Science Review, 20(1&2), pp.77-101, 1997
- 7) 伊庭幸人・種村正美・大森祐浩・和合肇・佐藤整尚・高橋明彦: 計算統計, 岩波書店, 2005
- 8) 稲田有香: ニュータウンにおける買物目的地選択に関する研究, 東京工業大学修士論文, 2007

表4 最尤法(MNP)とベイズ(Bayes)法による多項プロビットモデルの推定結果

モデル			MNP		Bayes	
			センター北	センター南	センター北	センター南
パラメータ	手段	所要時間(分)	-0.1495 (-1.175)	-0.1468 (-1.515)	-0.04187 [0.609]	-0.05155 [0.659]
	自動車	固有定数	4.369 (1.137)	1.879 (1.136)	1.112 [1.443]	0.08552 [0.717]
	自転車	固有定数	3.590 (1.046)	1.219 (0.730)	0.5626 [0.6521]	-0.3477 [0.651]
	徒歩	固有定数	6.486 (1.177)	4.011 (1.448)	1.818 [2.866]	0.7386 [3.76]
	バイク	固有定数	0.1678 (0.008)	-2.130 (-0.179)	-3.256 [0.429]	-5.63 [0.388]
	自動車 バイク	コスト(円)	0.002554 (0.193)	0.000579 (0.042)	0.01329 [0.0143]	0.01315 [0.0133]
	バス	固有定数	-1.577 (-0.517)	-2.100 (-1.087)	-0.519 [0.0098]	-0.9557 [0.0106]
	バス 鉄道	コスト(円)	0.01953 (1.185)	0.01115 (1.291)	0.0049 [0.0016]	0.00298 [0.00152]
	社会属性	性別	-0.5837 (-1.059)	0.2299 (0.905)	-0.1436 [0.1277]	0.1681 [0.131]
		年齢	-0.2057 (-1.116)	-0.02122 (-1.393)	-0.005791 [0.00444]	-0.004114 [0.00427]
		年収(万円)	0.000766 (1.148)	0.000635 (1.398)	0.0002379 [0.000121]	0.0002072 [0.000113]
p 2			0.4125	0.3597		
的中率(%)		計	51.0	47.1	62.7	57.4

():t 値、 []:std.dev