Ⅳ－187　ディリクレ分布を用いた新たなﾍﾞｲｽﾞ更新法

東京工業大学　学生員　魚谷　憲
東京工業大学　正　員　屋井　鉄雄
東京工業大学　正　員　岩倉　成志
東京工業大学　学生員　高田　和幸

１．はじめに

交通需要ﾃﾞｰﾀは、定期的、継続的に収集されるｹｰｽが多く、これらは長年に渡り各種交通計画策定の重要な基礎統計として、また予測技術発展の原動力として大いに貢献してきた。しかし、近年の交通技術進歩に伴って、様々な環境変化が生じつつある。また、従来の調査方法においては様々な問題も指摘され始めている。そのため、将来を見越して、ﾃﾞｰﾀ収集技術を見直す必要性は高い。また、交通需要予測の目的や要求も時代と共に変化し、わが国の国内においては必ずしも大規模調査に基づく技術論ばかりが要求されるとは考えにくい。本研究ではﾍﾞｲｽﾞ更新法において追加的に統合利用する集計量の誤差の考え方を変え、誤差分布を新たな分布に置き直した更新過程の提案を行う。

２．従来のパラメータ更新法について

Ben-Akivaら(1988)のｸﾞﾙｰﾌﾟでは、集計ﾃﾞｰﾀの信頼性(分散など)が不明なことから、ﾎﾟｱｿﾝ分布を当てはめた方法論を開発しているが、個々の観測値が平均と標準偏差の等しい独立な分布に従う仮定は現実的とはいえない。また、森地ら(1987)は、離散選択ﾓﾃﾞﾙのﾊﾟﾗﾒｰﾀ推定時や移転時に追加的に得られる集計ﾃﾞｰﾀを用いて、信頼性や再現性のより高いﾊﾟﾗﾒｰﾀ値を得る方法論を提案した。集計ﾃﾞｰﾀの誤差分布を正規分布とすることで、集計ﾃﾞｰﾀの持つ信頼性を更新手続きに取り込むことに成功している。以下にこのﾍﾞｲｽﾞ更新法の概要を示す。

離散選択ﾓﾃﾞﾙ(例えば、ﾛｼﾞｯﾄ型ﾓﾃﾞﾙ)のﾊﾟﾗﾒｰﾀθdが、個人ﾃﾞｰﾀだけを用いて最尤推定された場合を考える。このﾊﾟﾗﾒｰﾀは漸近的に正規分布に従うため、以下のような密度関数を持つ確率変数と見なすことができる。

$$f(\theta)\propto \exp\{-\frac{1}{2}(\theta-\theta d)^t\Sigma_d^{-1}(\theta-\theta d)\} \tag{1}$$

これは平均θdで分散共分散ﾏﾄﾘｸｽがΣ$_d$の多項正規分布である。ここで、集計ﾃﾞｰﾀ行列がQで与えられ、その値の離散選択ﾓﾃﾞﾙを用いた推計値をQ(θ)で表し、Qを正規分布に従う誤差εを用いて、

$$Q=Q(\theta)+\varepsilon \tag{2}$$

と表せば、ﾊﾟﾗﾒｰﾀθの事後確率密度は、

$$f(\theta\mid Q)=f(\theta)\cdot L(Q\mid\theta) \tag{3}$$

によって表される。ここで、尤度関数L(Q｜θ)は、

$$L(Q\mid\theta)\propto \exp\{-\frac{1}{2}(Q-Q(\theta))^t\Sigma_q^{-1}(Q-Q(\theta))\} \tag{4}$$

であるから、結局事後密度関数は、

$$f(\theta\mid Q)\propto \exp\{-\frac{1}{2}(\theta-\theta d)^t\Sigma_d^{-1}(\theta-\theta d) -\frac{1}{2}(Q-Q(\theta))^t\Sigma_q^{-1}(Q-Q(\theta))\} \tag{5}$$

であり、この式の最大化によって、新たなﾊﾟﾗﾒｰﾀﾍﾞｸﾄﾙθを決められる。実際の計算において、Qに何を用いるかは、それが離散選択ﾓﾃﾞﾙを用いて推計できる値でありさえすれば他に際だった制約条件はいらない。したがって、大変応用範囲の広い更新方法である。

３．多項分布を用いたパラメータの更新方法

以上の更新方法では、追加的な集計ﾃﾞｰﾀの信頼性を表す分散が必要であったが、必ずしも明らかでない。また、誤差の設定における今一つの仮定である「選択肢間の独立性」については、更新するﾓﾃﾞﾙがｼｪｱﾓﾃﾞﾙであることを考えると必ずしも適当であったとは言えない。そこで、本研究では集計ﾃﾞｰﾀの分布に多項分布を用いることによって、交通機関などの選択肢間の競合関係を考慮した分布形を設定し、ﾃﾞｰﾀの誤差分散について明示的な設定行わない方法をまず提案した。ここでは、段階的な更新方法を述べる。したがって、通常の最尤法で離散選択ﾓﾃﾞﾙのﾊﾟﾗﾒｰﾀを推定した後に、集計ﾃﾞｰﾀの分布に次の分布を当てはめて、ﾍﾞｲｽﾞ更新による新しいﾊﾟﾗﾒｰﾀを得る計算手順となる。

交通量x_1、x_2、x_3、・・・、x_kが

$$\frac{X!}{x_1!\cdot x_2!\cdots x_k!}p_1^{x_1}\cdots p_{k-1}^{x_{k-1}}(1-p_1-\cdots-p_{k-1})^{x_k} \tag{6}$$

なる多項分布に従うと考える。ここで、Ｘは総交通量、ｐにはﾛｼﾞｯﾄﾓﾃﾞﾙなど離散選択ﾓﾃﾞﾙより得られるｼｪｱを当てはめることによって、従来と同様な考え方に基づくﾓﾃﾞﾙﾊﾟﾗﾒｰﾀの更新方法を作り出せる。上式(6)が追加ﾃﾞｰﾀによる尤度関数に相当するため、複数のＯＤﾍﾟｱで複数の交通手段の交通量x_jが観測された場合を想定すると、尤度$La(=L(x \mid \theta))$は、

$$La=\prod_{j=1}^{J}\frac{X!}{x_{j1}!\cdots x_{jk}!}p_{j1}^{X_{j1}}\cdots p_{jk-1}^{X_{jk-1}}(1-p_{j1}-\cdots-p_{jk-1})^{X_{jk}} \tag{7}$$

によって表される。ここでＪはＯＤﾍﾟｱの個数を表し、ｊがそのうちの各ＯＤに対応する。したがって、x_{ji}、p_{ji}は各々、ＯＤﾍﾟｱｊの交通機関ｉの観測交通量とﾓﾃﾞﾙによる推計ｼｪｱを表す。この場合、事前分布が非集計ｻﾝﾌﾟﾙから作成した離散選択ﾓﾃﾞﾙのﾊﾟﾗﾒｰﾀ分布（(1)式）である。したがって、事後確率密度は(3)式と同様でありこの分布を対数変換すれば

$$\begin{aligned}\ln f(\theta \mid q)&=\ln f(\theta)+\ln L(q \mid \theta)\\&=-\frac{1}{2}(\theta-\theta d)^t\Sigma_d^{-1}(\theta-\theta d)\\&\quad+\sum_{j=1}^{J}\sum_{k=1}^{K+1}(x_{jk})\ln p_{jk}+C\end{aligned} \tag{8}$$

$$C=\ln\prod_{j=1}^{J}\frac{X!}{x_{j1}!\cdots x_{jk}!} \tag{9}$$

である。(8)式の最大値を求めれば新しいﾊﾟﾗﾒｰﾀθを得られる。(8)式の第１項は最小2乗法と同等な線形式であるが、第2項には選択確率p_{jk}が含まれθに関する非線形項になっているため、この計算には非線形最適化計算が必要である。

４．ディリクレ分布を用いたパラメータの更新方法

交通量x_1、x_2、x_3…、x_kがﾊﾟﾗﾒｰﾀ$(\alpha、q_1)$、$(\alpha、q_2)$、$(\alpha、q_3)$、…$(\alpha、q_k)$の独立なｶﾞﾝﾏ分布に従うとすれば、その比z_i

$$z_i=x_i/(x_1+\cdots+x_k)、i=1、\cdots、k \tag{10}$$

の同時確率密度関数は、

$$\frac{\Gamma(q_1+\cdots+q_k)}{\Gamma(q_1)\cdots\Gamma(q_k)}z_1^{q_1-1}\cdots z_{k-1}^{q_{k-1}-1}(1-z_1-\cdots-z_{k-1})^{q_k-1} \tag{11}$$

なるﾃﾞｨﾘｸﾚ分布に従う。ここでq_iは以下の式で表されると仮定する。

$$q_i=T\cdot p_i(\theta) \tag{12}$$

Ｔは総交通量、$p_i(\theta)$は離散選択ﾓﾃﾞﾙより得られるｼｪｱである。この時、q_iが交通量であるため整数で近似しｶﾞﾝﾏ関数にｽﾀｰﾘﾝｸﾞ公式を当てはめることによりより

$$\frac{\Gamma(q_1+\cdots+q_k)}{\Gamma(q_1)\cdots\Gamma(q_k)}=\frac{T!\cdot q_1\cdots q_k}{q_1!\cdots q_k!\cdot T} \tag{13}$$

を得る。この関係を用いれば、尤度関数$L(q \mid \theta)$は、

$$\begin{aligned}\ln(q \mid \theta)&=\ln(T!)-\Sigma\ln(q_i!)\\&\quad+\Sigma\ln(q_i)+\ln T+\Sigma(q_i-1)\ln z_i\end{aligned} \tag{14}$$

$$=\Sigma(q_i-1)\ln\frac{z_i}{q_i}+C \tag{15}$$

である。ここで３．で述べた多項分布の場合と同様にﾍﾞｲｽﾞ更新の手続きに乗せると、

$$\begin{aligned}\ln f(\theta \mid q)&=-\frac{1}{2}(\theta-\theta d)^t\Sigma_d^{-1}(\theta-\theta d)\\&\quad+\sum_{j=1}^{J}\sum_{k=1}^{K}(q_{jk}-1)\ln\frac{Z_{jk}}{q_{jk}}+C'\end{aligned} \tag{16}$$

である。この(16)式の最大化をすることによって新しいﾊﾟﾗﾒｰﾀθを得られる。ＣとＣ’はﾊﾟﾗﾒｰﾀθには無関係な定数項である。

表－１にﾊﾟﾗﾒｰﾀ更新結果を示す。ﾊﾟﾗﾒｰﾀ更新を行ったﾓﾃﾞﾙは、平成２年度に実施した家庭訪問調査より作成した非集計ﾛｼﾞｯﾄﾓﾃﾞﾙ型、高速道選択ﾓﾃﾞﾙである。調査は京浜地域を対象にし、目的地は伊豆・箱根・三浦・房総に限定した。回収票は1048票、ﾓﾃﾞﾙ作成に用いたｻﾝﾌﾟﾙは777ｻﾝﾌﾟﾙである。このﾓﾃﾞﾙをもとに、同年実施の道路交通量調査より得た２９ｿﾞｰﾝ間の経路別集計ﾃﾞｰﾀを用いて更新した。

表１．ﾊﾟﾗﾒｰﾀ更新結果

	非集計ﾓﾃﾞﾙのﾊﾟﾗﾒｰﾀ	多項分布による更新結果	ﾃﾞｨﾘｸﾚ分布による更新結果
総時間（時間）	-0.00845	-0.0173	-0.0156
総費用（円）	-0.00201	-0.000181	-0.000121
海沿い距離（km）	0.0171	0.0133	0.0141
定数項	-0.0889	-0.905	-0.922

５．おわりに

本研究は、従来より提案されていたﾍﾞｲｽﾞ更新方法を集計ﾃﾞｰﾀに対する異なる仮定に基づいて定式化したものである。構造化された多項分布を用いた方法と、ﾃﾞｨﾘｸﾚ分布を用いた方法の２つの更新法の提案を行った。

［参考文献］

森地　茂、屋井　鉄雄、平井節生：
個人データと集計データとの統合利用によるモデル構築方法
土木計画学研究論文集，Ｎｏ５，pp51-58，1987