Ⅳ-136　交通需要推定モデルにおける入力データ精度の影響について

北海道大学　工学部　　正員　　山形耕一

1．はじめに

交通需要推定には多くの数学的モデルが用いられるが，モデルによる推定値には，その操作の諸段階において種々の原因により生じる誤差が含まれている。これらの誤差を発生段階によって分けると，①モデル設定段階での誤差……現象解析の誤り，説明変数の省略，代理変数，線型化等の構造の簡略化等，②パラメータ設定段階での誤差……パラメータの推定法，推定のための入力データの誤差等，③予測段階での誤差……説明変数の予測値の誤差等，がある。交通需要推定においては，交通発生から交通配分に至る各段階で用いられるモデルのそれぞれが，これらの誤差を生じると共に，各段階の推定値は次の段階の入力となるため，誤差が累積される。したがって，個々のモデルで生じる誤差を検討すると共に，データの収集から推定の諸段階を通した全プロセスとしての推定の精度を検討し，精度的斉合性を図ることが必要である。本研究は，パラメータ推定段階において，入力データの精度がパラメータの推定値の誤差に及ぼす影響を考察し，モデルで必要とする精度に応じた調査やデータ作成を行うこと，あるいは，利用可能なデータの精度に適したモデル選択を行うための基礎研究である。

2．パラメータ推定の誤差

パラメータの推定においては，先ず，現象解析にもとづいて，モデル型

$$Y = F(X_1, \cdots\cdots, X_l : Q_1, \cdots\cdots, Q_m) \qquad [1]$$

ここに，Y；従属変数，　$X_1, \cdots\cdots, X_l$；説明変数，　$Q_1, \cdots\cdots, Q_m$；パラメータ

を設定し，次に，観測された現象における従属変数と独立変数の観測値に対して，モデルによる推定値$\hat{Y}_\nu$が，観測値Y_νに最もよく適合するようにパラメータが定められる。ここで，Y_νと$\hat{Y}_\nu$の間の残差の原因を考えると，第1にはモデル設定段階での誤差のため，モデルが現象を記述し切れないことやν観測値における特殊性，第2には現象の観測に際しての測定誤差が挙げられる。交通需要推定モデルでは，従属変数は通常発生交通量や分布交通量等交通量データであり，その多くはパーソントリップ調査や自動車OD調査等の標本調査によって得られる。したがって，調査誤差として測定誤差が生じるが，本研究ではこの誤差のパラメータ推定値に及ぼす影響を考える。上記2つの原因によって生じる残差を，それぞれ確率変数U，Vで表すと，観測値と推定値との間の関係は

$$Y_\nu = F(X_{1\nu}, \cdots\cdots, X_{l\nu} : Q_1, \cdots\cdots, Q_m) + U_\nu + V_\nu \qquad [2]$$

で表される。

〔2〕式の構造を前提としてパラメータの推定を行う場合，先ず，観測値Y_νと推定値$\hat{Y}_\nu = F(X_{1\nu}, \cdots X_{l\nu} : Q_1 \cdots Q_m)$の間の適合度にいかなる基準を設けて最小化するかが1つの問題となる。とくに，対数等の変換を行って線型化されるモデルでは，例えば，$Y_\nu - \hat{Y}_\nu$の2乗和最小の基準と，対数変換後の$\log Y_\nu - \log \hat{Y}_\nu$の2乗和最小の基準とでは，異なったパラメータが得られる。

モデルが線型である場合は，〔2〕式の構造は

$$Y_\nu = \beta_0 + \beta_1 X_{1\nu} + \cdots\cdots + \beta_l X_{l\nu} + U_\nu + V_\nu \qquad [3]$$

である。残差2乗和最小の基準を用いて，回帰分析を行うと，パラメータβ_iと推定量$\hat{\beta}_i$との関係は

$$\hat{\beta}_i = \beta_i + \sum_{j=1}^{l} S_{ij}^{-1} s_{uj} + \sum_{j=1}^{l} S_{ij}^{-1} s_{vj}$$

ここに，S_{ij}^{-1}；$S_{ij} = \sum_{\nu=1}^{n} (X_{i\nu} - \bar{X}_i)(X_{j\nu} - \bar{X}_j)$なる平方和・積和行列の逆行列の$ij$要素

$s_{uj} = \sum_\nu (X_{j\nu} - \bar{X}_j)(U_\nu - \bar{U})$，　$s_{vj} = \sum_\nu (X_{j\nu} - \bar{X}_j)(V_\nu - \bar{V})$，　n；データ数

となる。各説明変数と確率変数UとV，およびUV相互間は独立であると仮定してよい。したがって，U，Vの分散をそれぞれσ_u^2，σ_v^2とすると，$\hat{\beta}_i$の分散は，

$$Var(\hat{\beta}_i) = S_{ii}^{-1}\ (\sigma_U^2 + \sigma_V^2) \qquad [5]$$

である。すなわち、パラメータ推定における従属変数データの分散の影響は，攪乱項Uの影響と共に現れる。そして，その影響の大きさは，σ_U^2とσ_V^2の相対的な大きさが関係している。確率変数UおよびVの実現値は，観測値からは知り得ないが，UとVとの和である確率変数の分散として，

$$\sigma_U^2 + \sigma_V^2 = \sum_{\nu=1}^{n} (Y_\nu - \hat{Y}_\nu)^2 / (k - l - 1) = S_{yy}(1 - R^2) / (n - l - 1) \qquad [6]$$

ここに，k；データの数，l；説明変数の数，R；重相関係数

を推定することができる。したがって，調査法の側から，データの分散σ_V^2を知ることができれば，σ_U^2も算出されるので，データ精度によるパラメータ推定の誤差を知ることができる。

3．シュミレーションによる分析

モデルが非線型である場合には，前節の方法は適用できず，シュミレーションによる方法が有効であると考えられる。本研究では，グラビティモデルのパラメータを，対数変換せずに残差2乗和最小の基準から直接推定する場合を取り上げ，シュミレーションによる分析を行った。モデル構造は

$$T_{ij} = k(G_i A_j)^{\alpha} D_{ij}^{\beta} + U_{ij} + V_{ij} \qquad [7]$$

を用いた。対象データには，道央都市圏パーソントリップ調査全目的OD表を用いた。このデータは，調査達成率5.2％で，トリップ数5000規模のゾーンペア交通量で約6％の標準誤差をもつ。シュミレーションでは，このOD表を正しいものと仮定し，そのパラメータを真のパラメータα，β，kとした。そして，調査誤差項V_{ij}に人為的に誤差を与えたとき，推定されたパラメータ$\hat{\alpha}$，$\hat{\beta}$，$\hat{k}$の真の値周辺における変動を調べた。

標本調査における標本誤差は，調査単位が単純ランダム抽出されているとき，交通量の如く特定の属性をもつ調査単位の数を推定する場合には，

$$Var(T_{ij}) = \sqrt{\frac{T_{ij}(N - T_{ij})}{n} \frac{N-n}{N}} \qquad [8]$$

ここでN，n；母集団および標本の大きさ(トリップ)で表される。そこで，各V_{ij}の変動係数が一定となるようにそれぞれのV_{ij}の標準誤差を定め，これに確率的な誤差の実現値を表す標準正規乱数を乗じてV_{ij}の値とした。そして，変動係数の大きさに対応するパラメータ推定値の誤差の大きさを調べている。なお，パラメータ推定法としては，テーラー展開を用いた線型化法を用いている。結果を表1に示す。

表1　パラメータ推定誤差のシュミレーション

		α	β
パラメータ		0.959	0.537
標準誤差の水準	0.05	0.0012 (0.1)%	0.0028 (0.5)%
	0.10	0.0029 (0.3)	0.0073 (1.4)
	0.15	0.0046 (0.5)	0.0110 (2.1)
	0.20	0.0110 (1.1)	0.0190 (3.6)

試行回数　10

表1にみられるように，グラビティモデルでは，入力データの精度の低下にもとずくパラメータ推定誤差の増加は極めて小さい。すなわち，入力データの精度に対して感度の低いモデルであると言えよう。これは，グラビティモデルの適合度が十分でないため，攪乱項Uの分散σ_U^2が大きくなり，入力データの分散σ_V^2の影響がパラメータ推定に現れてこないためと推測される。

4．おわりに

本研究では，標本調査等のように，観測誤差を含むデータを用いてモデルのパラメータ推定を行う場合，データの精度が推定値にもたらす影響について考察した。また，グラビティモデルにおける影響をシュミレーションにより検討し，グラビティモデルはデータ精度の影響が小さい，いわば，データの精度に感度の低いモデルであることが判った。このことは，グラビティモデルを用いることを前提とするならば，交通調査は比較的低い精度で設計してもよいこと，また逆に，低い精度のデータであっても，グラビティモデルは適用可能であることを意味する。今後更に，交通需要推定の諸段階のモデルについてデータ精度の影響の研究を行い，調査・データ作成から需要推定の各段階を通した全プロセスの推定精度の問題を検討していくことが必要と考える。