Ⅳ-68　交通需要予測の誤差分析

東京大学都市工学教室　正員　太田勝敏

この研究は，数学的モデルによる予測値に含まれる誤差の原因を考察して，それらの原因の特性から，その予測値に含まれる誤差の大いさを予測時点で事前に推定する方法（予測誤差分析法）を検討し，交通需要予測モデルのひとつであるトリップ発生モデルの予測誤差についての分析応用例を示したものである。

1．数学的モデルによる予測に伴なう誤差の原因

一般に，ある事象をある因果関係を仮定した数学的モデルにより定量的に推定する場合，その予測値に伴なう誤差の原因としては次のようなものがある。

(1)　モデルの理論的構造の誤り--------真の関係　$Y_t = F_t(X_t ; A_t)$，但し Y_tは予測対象，$F_t(X_t : A_t)$ は X_tを独立変数，A_tをパラメータとする構造式，を $Y_a = F_a(X_a : A_a)$ と仮定する誤り．

(2)　予測モデル式の誤差-----------データの入手可能性，取扱い上の難易等により，$Y_a = F_a(X_a : A_a)$ の近似式 $Y_b = F_b(X_b : A_b)$ を用いることによる誤差。（例．影響の小さい独立変数の省略，代理変数の利用，線型モデル化　）

(3)　パラメータ推定値の誤差--------パラメータ A_bの推定値 $\hat{A}_b$ を用いることによる誤差．

(4)　独立変数推定値の誤差---------独立変数 X_b に対し，その推定値 $\hat{X}_b$ を用いることによる誤差．

予測に伴なう最大の誤差原因は，明らかに(1)であるが，将来の予測時点でそのモデルの理論的構造が真であるかを予め推定することは困難である。　(2)，(3)，(4)　の誤差は，基になったデータの精度，パラメータ推定方法，現状との適合性などからある程度その大いさを推定できる場合が多い。

2．予測誤差分析の諸手法

$Y = F(X, A) + D$，但し Dは上記(2)のモデル式の誤差を示すかく乱項で平均を0とする確率変数，という予測モデルによるYの予測値 $\hat{Y}$ に含まれる誤差の内，モデルの理論的構造の誤り以外の3原因によるものについては，X，A，Dの誤差特性から$\hat{Y}$の確率分布を推定することによって求めることができる。この予測誤差分析の本質は，X，A，Dに含まれる誤差が予測モデルを通してYにどのように"伝播"してゆくかを分析することであり，理論的には，X，A，D　という確率変数が予測モデルによりYという新しい確率変数にどう変換されるかという確率論における"変数変換"の問題である。　このことから，予測誤差分析の代表的ケースと手法には下表のようなものがある。[1),2)]

ケース	推定内容	（インプット）	（アウトプット）	分析手法：厳密法	分析手法：近似法
ケースⅠ	確率分布形	$f(\hat{x}), f(\hat{a}), f(\hat{d})$	$\longrightarrow f(\hat{y})$	変数変換法	モンテカルロ・シミュレーション Quantile Arithmetic
ケースⅡ	モーメント （平均，分散）	$E(\hat{x}), E(\hat{a}), E(\hat{d}), Cov(a_i, a_j)$ $V(\hat{x}), V(\hat{a}), V(\hat{d}), Cov(x_i, x_j)$	$\longrightarrow E(\hat{y}), V(\hat{y})$	モーメント計算法	テイラー展開法
ケースⅢ	範囲（信頼限界）	$[\underline{x}, \overline{x}], [\underline{a}, \overline{a}], [\underline{d}, \overline{d}]$	$\longrightarrow [\underline{y}, \overline{y}]$	————	信頼限界代数 Interval Analysis

注．1　予測モデル：　$\hat{y} = F(\hat{x} : \hat{a}) + \hat{d}$，　但し $\overline{d}(= E(\hat{d})) = 0$．

2　記号：　$f(x)$は，確率変数Xの確率密度分布．$E(x)$，$V(x)$，$Cov(x, y)$は，それぞれ期待値，分散，共分散を示す。$[\underline{x}, \overline{x}] (= \underline{x} \leq x \leq \overline{x}) = CI(x : \alpha\%)$は，$x$の信頼限界($\alpha\%$)である。

3. 線型回帰モデルによる予測に伴なう誤差

下に示す①式のような線型回帰モデルによる予測誤差（分散）は，テイラー展開により②式で近似的に求めることができる。　この場合，かく乱項はパラメータおよび独立変数と相関関係はなく，また，パラメータと独立変数の間にも相関関係がないものと仮定している。

線型回帰モデル　$\hat{y} = \sum_i \hat{a}_i \hat{x}_i + \hat{\alpha}$,　$\bar{\hat{\alpha}} = 0$　①

推定値の分散（予測値）　$V(\hat{y}) \fallingdotseq \sum_i\sum_j \bar{\hat{a}}_i \bar{\hat{a}}_j Cov(\hat{x}_i, \hat{x}_j) + \sum_i\sum_j \bar{\hat{x}}_i \bar{\hat{x}}_j Cov(\hat{a}_i, \hat{a}_j) + V(\hat{\alpha})$　②

$\left[or \fallingdotseq \sum_i \{\bar{\hat{a}}_i^2 V(\hat{x}_i) + \bar{\hat{x}}_i^2 V(\hat{a}_i)\} + 2\sum_i\sum_j^{i-1}\{\bar{\hat{a}}_i \bar{\hat{a}}_j Cov(\hat{x}_i, \hat{x}_j) + \bar{\hat{x}}_i \bar{\hat{x}}_j Cov(\hat{a}_i, \hat{a}_j)\} + V(\hat{\alpha}) \right]$

但し，$\bar{\hat{a}}$，$\bar{\hat{x}}$ は $\hat{a}$，$\hat{x}$ の平均値を示す。尚，通常予測にはこの平均値を用いる。

②式の第1項は，独立変数の推定値，第2項はパラメータ推定値，第3項は回帰モデル式，のそれぞれの誤差の影響を表わしている。　これらの内，$V(\hat{\alpha})$，$V(\hat{a}_i)$，$Cov(\hat{a}_i, \hat{a}_j)$ は回帰分析の際に求めることができる。[3]　$V(\hat{x}_i)$，$Cov(\hat{x}_i, \hat{x}_j)$ は，$\hat{x}$ の推定根拠にもとづいて推定することになる。

4. トリップ発生モデルへの応用例

Deutschman は，①式のような多重回帰式によるゾーン別世帯当り平均トリップ発生モデルについて，独立変数に一定比率の誤差を与えた場合，予測値の誤差がどのようになるかについて，シュミレーションにより推定している。[4]　Deutchman の方法は，ゾーン j の独立変数 i の推定値 $\hat{x}_{ij}$ に対し，± w %（符号はランダムに選ぶ）の誤差を仮定して，ゾーン j の世帯当り平均トリップ発生量 $\hat{y}_j$ を求め，その予測誤差 $\hat{y}_j - y_j$ の全ゾーン平均値（標準誤差 S.E.）を全ゾーン平均世帯当りトリップ発生量 $\bar{y}$ との比率（%S.E.）で示している。　このシュミレーションは，②式のテイラー展開により③式のように近似できる。

$$V(\hat{y}_j) \fallingdotseq \sum_i \bar{\hat{a}}_i^2 V(\hat{x}_{ij}) + V(\hat{\alpha}) = \sum_i \bar{\hat{a}}_i^2 \left(\frac{w}{100} \times \bar{\hat{x}}_{ij}\right)^2 + s^2 \qquad ③$$

但し，s は回帰式の標準誤差。

この場合，パラメータ推定値の誤差は不明のためその影響を省略し，独立変数の推定値相互間には相関がないものと仮定した。また，全ゾーンについての平均的誤差は，$\bar{x}_{ij}$ の代りに全ゾーン平均値 $\bar{x}_i$ により近似した。　左表は，この分析結果の一部であるが，テイラー展開による近似式がシュミレーションの結果とよく一致していることがわかる。　この応用例から予測誤差分析法は，ある予測精度を得るために必要な独立変数の推定精度の決定や，モデルに入れる独立変数とモデルの構造の選定などに有益な情報を与えてくれるものであることが示唆される。

No.	回帰モデルと x の誤差	R	S	%S.E.	
				Deutsch.	テイラー
1	$x_1 = -2.157 + .0583x_6 + 2.06\times10^{-4}x_4$	.94	.84	16	—
2	$w_6 = 10\%$　$w_4 = 10\%$			17	16
3	10%　20%			17	17
4	10%　25%			17	17
5	$x_1 = 3.292 + .06597x_6$	.93	.90	17	—
6	$w_6 = 10\%$			18	17
7	$x_1 = 1.578 + 4.784x_2$	.94	.83	15	—
8	$w_2 = 15\%$			20	19
9	$x_1 = 2.195 + .0290x_6 + 2.854x_2$	.95	.75	14	—
10	$w_6 = 10\%$　$w_2 = 15\%$			16	15
11	$x_1 = 1.156 + 4.501x_2 + 9.7\times10^{-5}x_4$	.94	.82	15	—
12	$w_2 = 15\%$　$w_4 = 10\%$			19	18
13	15%　20%			19	18
14	15%　25%			20	18

注．x_1 ＝ 世帯当りパーソントリップ（ゾーン平均）　$\bar{x}_1 = 5.37$
x_2 ＝ 〃　乗用車台数　$\bar{x}_2 = .79$
x_4 ＝ 平均世帯所得　$\bar{x}_4 = 6670$
x_6 ＝ 一戸建住宅の比率（%）　$\bar{x}_6 = 31.4$

参考文献

1) K. Ohta, 'Prediction Analysis in Planning' (1971)
2) W. Alonso, "Predicting Best with Imperfect Data", AIP Journal, July 1968.
3) ジョンストン「計量経済学の方法」(訳 1964)
4) H. D. Deutchman, 'Establishing a Statistical Criterion for Selecting Trip Generation Procedures", HRR No.191