Ⅳ-143　離散データの回帰分析に関する一考察

近畿大学工学部　正　員　難　波　義　郎
〃　　〃　保　野　健治郎
〃　　〃　大　森　豊　裕

1．　はじめに　　土木計画においては，離散量のデータ，とくに計数データ（自然数）を扱うことが少なくない．そして，このようなデータに対して回帰分析を行なうことも多い．一般に回帰式を求める方法としては多くの場合，最小２乗法が用いられてきた．この最小２乗法は，データの誤差分布を正規分布と仮定したときの最尤法の１つとして体系づけられる．[1] 正規分布は，理論的に＋∞から－∞の連続変数の分布である．したがって，取り扱うデータが連続量でその範囲が＋∞から－∞の範囲の値を取り得る場合に適用されるものである．しかし，件数などの自然数は，けっして負の値をとらない離散量である．従来はこのようなデータにも最小２乗法が用いられてきたが，厳密には明らかに不適切な用い方をしてきたようである．とくに，母平均が０に近い標本においては，正規分布には負の部分に分布が存在するので，非負であるデータに正規分布を当てはめると矛盾が生じるので，そのまま適用することはできないことが指摘できる．

2．　理論的考察[2),3)]　n組のデータ（x_1, y_1），…，（x_n, y_n）が与えられたものとし，回帰式は

$$y = g(x \mid a_1 \cdots a_m) \tag{1}$$

とする．データyの確率密度関数を正規分布とすると，尤度は

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\sigma^2}} \exp\{-\frac{1}{2\sigma^2}[y_i - g(x_i \mid a_1 \cdots a_m)]^2\} \tag{2}$$

となる．与えられたデータ（x_1, y_1），…，（x_n, y_n）に対して，この尤度を最大にするa_1，…，a_mおよびσ^2が回帰係数および残差分散の最尤推定量を与える．この尤度（または対数尤度）を最大にするa_1，…，a_mを求めるためには，残差２乗和

$$S = \sum_{i=1}^{n} \{y_i - g(x_i \mid a_1 \cdots a_m)\}^2 \tag{3}$$

を最小にする$a_1 \cdots a_m$を求めればよいことは周知の通りである．

一方，離散データに対しては，非負の分布を考慮して確率関数をポアソン分布とすれば，尤度は

$$L = \prod_{i=1}^{n} \frac{\{g(x_i \mid a_1 \cdots a_m)\}}{y_i\,!} \exp\{-g(x_i \mid a_1 \cdots a_m)\} \tag{4}$$

となる．

3．　回帰式の指標　　回帰式を評価する際に参考となる指標には以下のようなものが考えられる。

a．重相関係数：R

b．残差２乗和：S　連続量のデータの場合は，この残差が最小のとき，尤度が最大となる．

c．尤度（対数尤度）：l

d．赤池情報量規準：ＡＩＣ

$$\mathrm{AIC} = -2 \times (\text{最大対数尤度}) + 2 \times (\text{自由パラメータの数}) \tag{5}$$

「エントロピーとの観点からＡＩＣの値の差が，１～２以上あれば，有意と考えられる[2)]」とされる．

e．推定値の総和：Σy_i

f．データと回帰式の最大偏差：δ_+, δ_-

$$\delta_+(\text{回帰式より上側の最大偏差}) = \max\{\hat{y}_i - y_i\} \tag{6}$$

$$\delta_-(\text{回帰式より下側の最大偏差}) = \max\{y_i - \hat{y}_i\} \tag{7}$$

4．　分析結果と検討　　本稿では回帰モデルとしていわゆる数量化Ｉ類

$$Y=\sum_{j=1}^{2}\sum_{k=1}^{K_j} x_{jk}\delta_{jk}+c \qquad (8)$$

ここに，Ｙは計数データ（自然数），x_{jk}はｶﾃｺﾞﾘｳｪｲﾄ，ｊはｱｲﾃﾑの通し番号，δ_{jk}はﾃﾞｰﾀがｊ番目のｱｲﾃﾑのｋ番目のｶﾃｺﾞﾘに反応したときに1,その他のとき0となる変数，ｋはｶﾃｺﾞﾘの通し番号（ｋ$=1,2,\cdots K_j$．ただし$K_1=2$,　$K_2=3$)，ｃは定数項である．

を考え，データはポアソン乱数を使って人為的に作成したもの（表１参照）を最小２乗法（従来の解法：式（３）を最小にする方法）と式（４）を最大にする方法（ポアソン回帰と呼ぶ）との計算結果（表２参照）を比較検討してみることとする．

表２のように，重相関係数は全く同じ値となっているが，ポアソン回帰法の場合の方が対数尤度の値は大きく，ＡＩＣは小さい。このことはデータの与え方からも当然の結果であるが，このようなデータの場合はポアソン回帰法によって分析すべきであることがわかる．また，カテゴリウエイトで注意の要する点はx_{22}の部分で，ポアソン回帰法では負の値であるのに最小２乗法では正の値となっていることである．

もっとも，両者の式からＹの推定値を求めた結果はそれほど違わないので，この例では致命的な結果とはなってはいない．手法の選択は，分析の立場によって異なると思われるが，簡便性や経験的に従来の方法で十分と言い切れるかどうかは疑問である。それぞれのケースでその都度両者の比較をしてみなければわからないと思われる．

以上，今回は離散データの回帰分析における手法と１つの例題を示したにすぎないが，さらに種々のデータを分析していく必要があると思われる．また，残差の絶対値を最小にする方法や残差の最大値を最小にする方法[4),5)]等についても今後の課題である．

表１　データ

Y	i	j	Y	i	j
1	1	1	5	2	1
0	1	1	6	2	1
0	1	1	4	2	1
1	1	1	0	2	1
0	1	1	13	2	1
5	1	2	0	2	2
2	1	2	11	2	2
1	1	2	13	2	2
4	1	2	4	2	2
0	1	2	6	2	2
6	1	3	2	2	3
8	1	3	6	2	3
6	1	3	10	2	3
2	1	3	8	2	3
1	1	3	1	2	3

表２　分析結果

項目　　方法	最小２乗法	ﾎﾟｱｿﾝ回帰法
c	5.233	1.917
x_{11}	−2.767	−0.440
x_{12}	2.767	4.939
x_{21}	−2.233	−1.088
x_{22}	−0.333	0.836
x_{23}	2.567	3.454
重相関係数	0.711	0.711
残差２乗和	338.87	339.54
対数尤度	−78.93	−73.21
ＡＩＣ値	171.87	158.42
推定値の総和	157.00	157.00
最大偏差（δ+）	9.43	9.69
（δ-）	−5.77	−5.77

参考文献

1)　S. ブラント：データ解析の方法，みすず書房，1976
2)　坂元慶行ほか：情報量統計学，共立出版，1983
3)　粟屋隆：データ解析，学会出版センター，1983
4)　S.H.ZANAKIS et al : OPTIMIZATIIN IN STATISTICS, NORTH-HOLLAND PUB., 1982
5)　ALBERT TARANTOLA : INVERSE PROBLEM THEORY, ELSEVIER,1987