

ロジスティックモデルにおける信頼領域

山 添 史 郎

Confidence Regions of Logistic Model

Shiro YAMAZOE

Abstract

Quantal response analysis is concerned with the relationship between dependent and response variables. Various algebraic forms have been proposed for the response relationship. In this paper we present the "exact" confidence regions for the logistic model with two parameters. The confidence region is constructed by the method which is applied by Sterne to the problem of determining confidence intervals for the proportion parameter in a Binomial Distribution.

1. はじめに

本論文においては量・効果関係モデルにおける信頼領域を作る問題を考える。量 d を決めて一つの反応を観測するとき、反応がおこる確率を $p(d)$ とする。 m レベル d_1, d_2, \dots, d_m で反応を観測し、その結果から反応曲線 $p(d)$ または $p(d)$ に関するパラメータについての推測が行なわれる。 d を薬剤の投与量、 $p(d)$ をその投与量での治験効果を生ずる確率、または副作用を生ずる確率とするとこの反応曲線 $p(d)$ は一般には単調増加関数とみなされ、ふつう S 字曲線がモデルとして使われる。母数を持つモデルとしてはロジスティック、プロビット、ホッケースティックといったものがよく使われる。反応確率 $p(d)$ をパラメトリックな数式モデルとして推測するのは LD_{50} 、 ED_{50} 、すなわち、 $p(d) = 0.5$ となる d 、または実質的安全量、すなわち、 ε を 10^{-6} などといったごく小さい確率として $p(d) = \varepsilon$ をみたす d を推測したいためである。このような量を推測するとき単に点

推定としてではなく信頼領域を求めることが必要となる。

量・効果モデルにおける信頼領域は近似定理を使うことによって行なわれている¹⁾。上記に示した例などでは観測個数はそれほど大きくはない。また設定されるモデルによって信頼領域が大きく変化することが予測される。特に実質的安全量の推測は外挿問題であるのでモデルに大きく依存する。従って近似理論によってではなく、正確に計られた信頼領域が必要となる。

二項分布の信頼区間を構成する方法として Sterne によるものがある¹⁾。その方法は信頼水準を正確に計算して作っており、かつ信頼区間の長さの総和を最小にするというよい性質をもっている。しかしこの方法でもとめられた信頼区間は単一の区間とはならない。その点を修正したのが Crow²⁾、Yamazoe³⁾ である。Sterne¹⁾ による方法は確率を正確に計算すること、計算機による解が得やすいことの特徴をもつ。Sterne の方法は各パラメータをとめて、その分布の値を大きい方から集めてその和が信頼水準 δ まで集めたものから信頼領域を構成

していく。この方法を以下で最大確率法とよぶことにする。

2. ロジスティックモデル

量 d に応じて反応する確率 $p(d)$ が

$$p(d) = 1 / \exp(\alpha - \beta d)$$

で与えられるとき、その量・効果モデルをロジスティックモデルという。量 d は正のみの値をとる場合が多いが、ここではその対数値をとったと考え、全実数値をとるものとする。また量・効果モデルは d に関し、単調増加と仮定することが多く、ここで $\beta > 0$ とする。本論文の目的は α 、 β の同時信頼領域を求めることである。

反応をみる d のレベルの個数を m とし、各レベルを $d_1 < d_2 < \dots < d_m$ とする。各 d_i において n_i 個の互いに独立な実験を行ない、それぞれ k_i 個の反応を観測したとする。この事象の確率は

$$L = \prod_{i=1}^m \binom{n_i}{k_i} p(d_i)^{k_i} (1-p(d_i))^{n_i-k_i}$$

で与えられる。尤度 $L = L(\alpha, \beta)$ を $k = (k_1, k_2, \dots, k_m)$ の関数とみて、その値が大きい方から順に加え、その和が信頼水準 δ に達するまでの k の集合を $A(\alpha, \beta)$ とする。この $A(\alpha, \beta)$ を仮説検定の用語を流用して許容域という。 $k^0 = (k_1^0, \dots, k_m^0)$ を観測したときの信頼水準 δ の信頼領域は

$$B(k^0) = \{(\alpha, \beta); k^0 \in A(\alpha, \beta)\}$$

で与えられる。この最大確率法によって得られた信頼領域の良さの一つは許容域 $A(\alpha, \beta)$ が一様に $k = (k_1, \dots, k_m)$ の最小個数から成りたっていることである。従って $B(k)$ が有界であるならば $\sum_k \{B(k) \text{ の長さ} \}$ が最小である信頼領域を最大確率法は与えている。このことは必ずしも $B(k)$ が一様に最小の信頼領域を与えていることにはならないが、最大確率法よりも一様に小さい信頼領域の存在しないことを意味している。

最大確率法のよい点は離散型分布の場合には、

計算機の助けをかりてではあるが、正確に信頼水準 δ の信頼領域を与えることができることである。一方この方法の欠点は求められた信頼領域が必ずしも単一の閉領域を与えるわけではないことである。このことは二項分布の場合にも指適され、Sterne の方法は修正されて、同じ最小性をもち、かつ、単一の区間からなる信頼区間を持つような方法が提出されている^{2,3)}。

ロジスティックモデルにおける α 、 β の最尤推定量は対数尤度を微分した

$$\sum_i (k_i - n_i p_i) = 0$$

$$\sum_i (k_i - n_i p_i) d_i = 0$$

をみたす α 、 β を求めればよい。ここで

$$p_i = p(d_i) = 1 / (1 + \exp(\alpha - \beta d_i))$$

である。また情報行列を J とするとその各成分は

$$J_{11} = \sum n_i p_i (1-p_i)$$

$$J_{12} = J_{21} = - \sum n_i d_i p_i (1-p_i)$$

$$J_{22} = \sum n_i d_i^2 p_i (1-p_i)$$

で与えられる。 $n = \sum n_i$ が大きいときは

$$\sqrt{n} \{(\hat{\alpha}, \hat{\beta})' - (\alpha, \beta)'\}$$

が近似的に 2 次元正規分布 $N((0, 0)', J^{-1})$ に分布する⁴⁾。よって

$$\chi^2 = ((\hat{\alpha}, \hat{\beta}) - (\alpha, \beta)) J \left(\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} - \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \right)$$

は自由度 2 の χ^2 分布に近似的に従う。

3. 例

ロジスティックモデルに対し最大値法で信頼領域を求めるにはパラメータ α 、 β を細分化し、かく α 、 β に対して与えられた $k = (k_1, \dots, k_m)$ が許容域 $A(\alpha, \beta)$ に入るかどうかを決めればよい。まず観測レベルは $m = 4$ とし、各レベルを $d_1 = -1.5$ 、 $d_2 = -0.5$ 、 $d_3 = 0.5$ 、 $d_4 = 1.5$ とし、観測個数は一定の $n_i = 4$ とした。反応

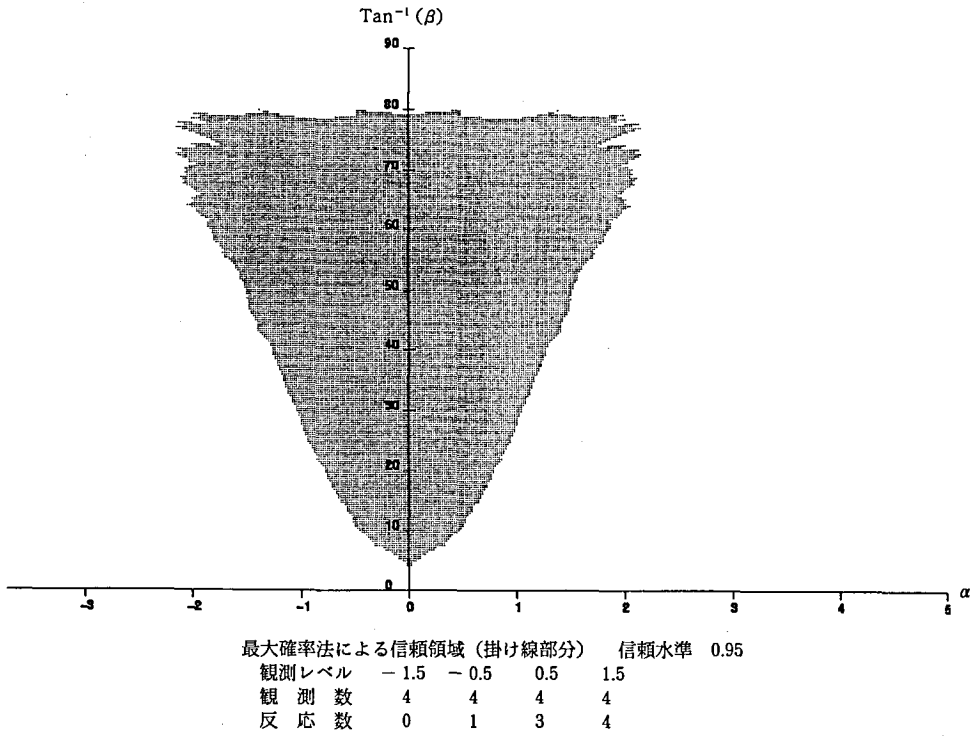


図1 logistic モデルの信頼領域

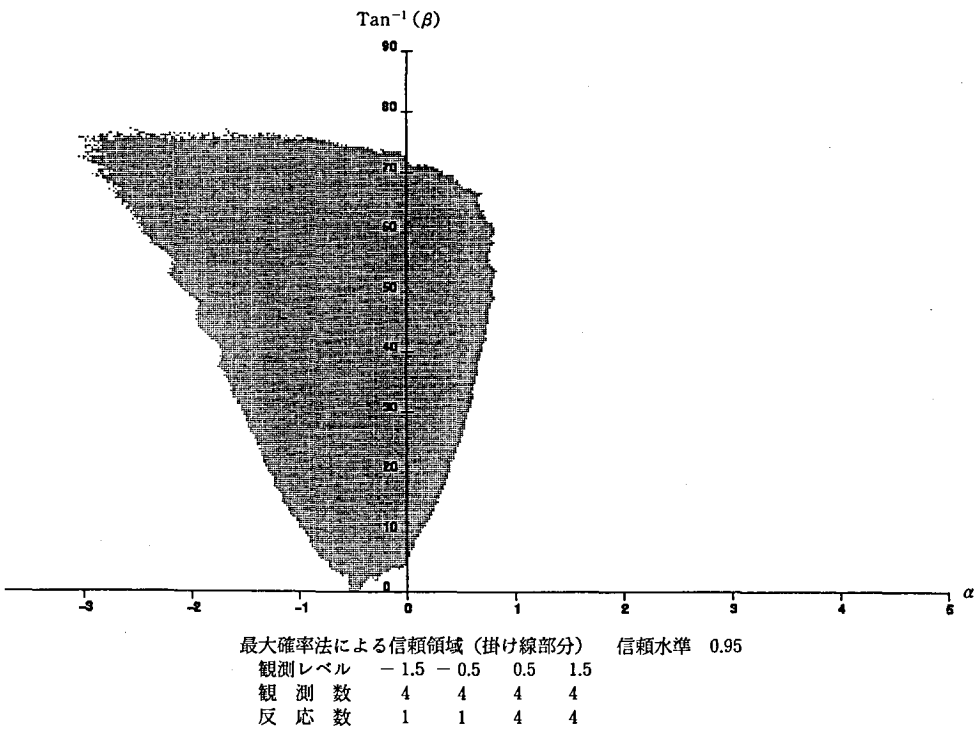


図2 logistic モデルの信頼領域

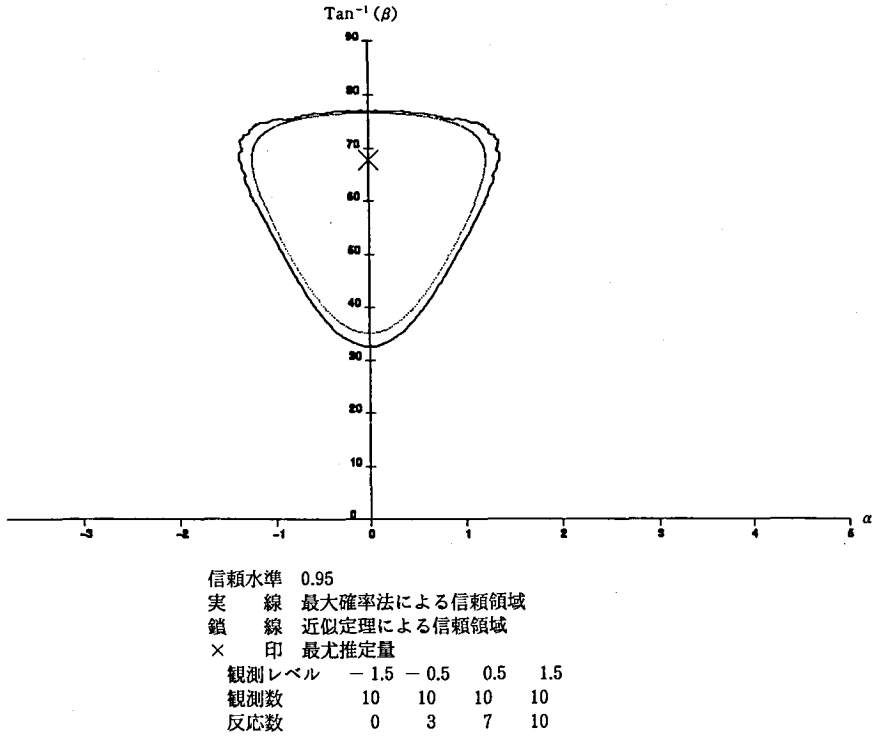


図3 logistic モデルの信頼領域

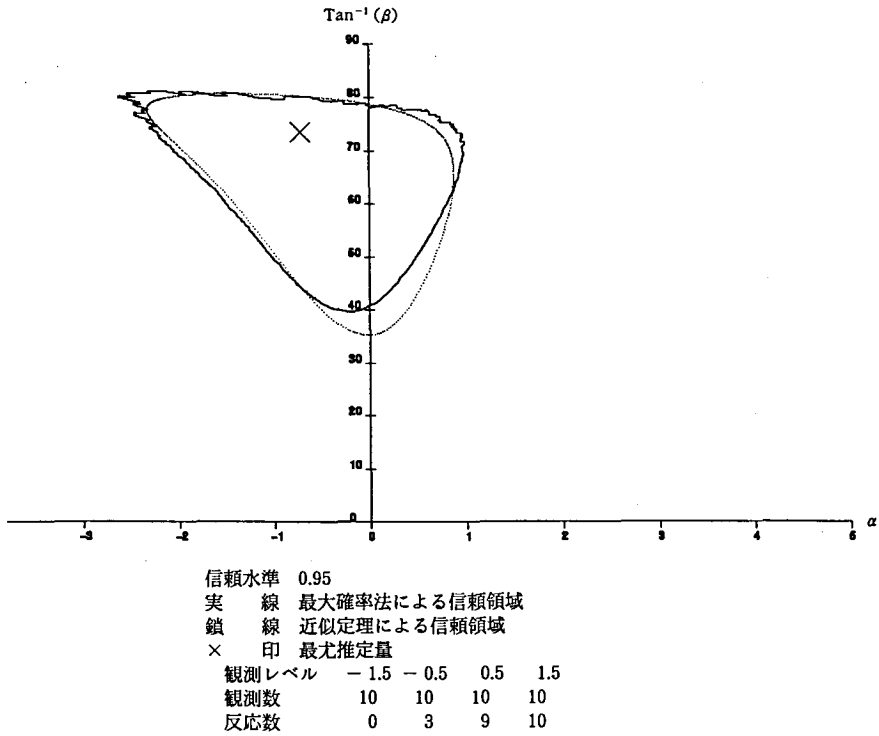


図4 logistic モデルの信頼領域

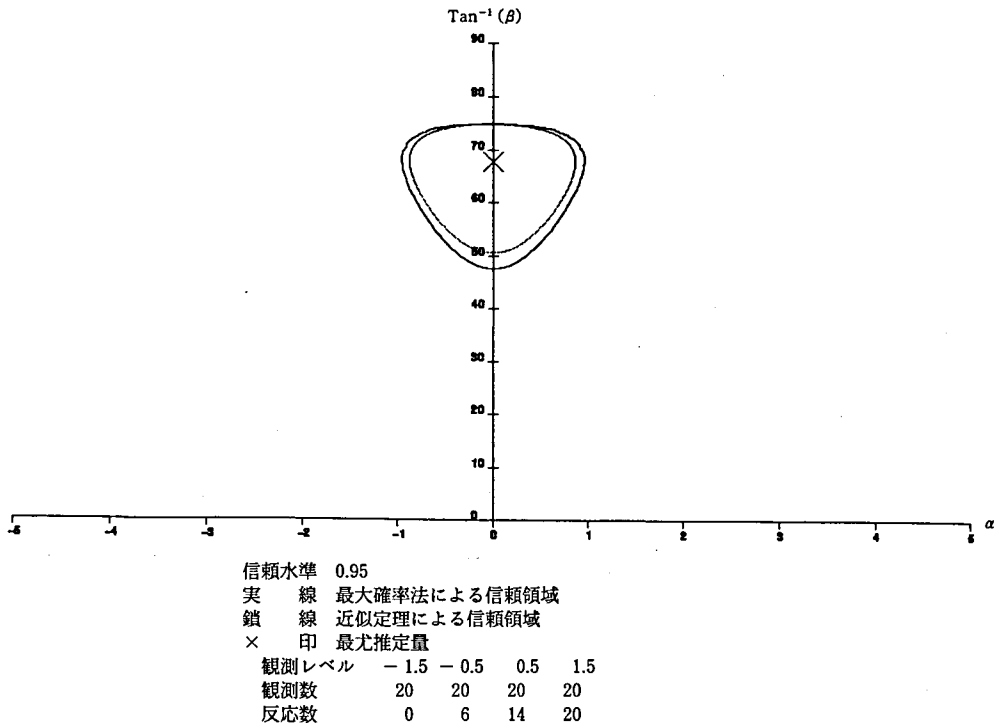


図5 logistic モデルの信頼領域

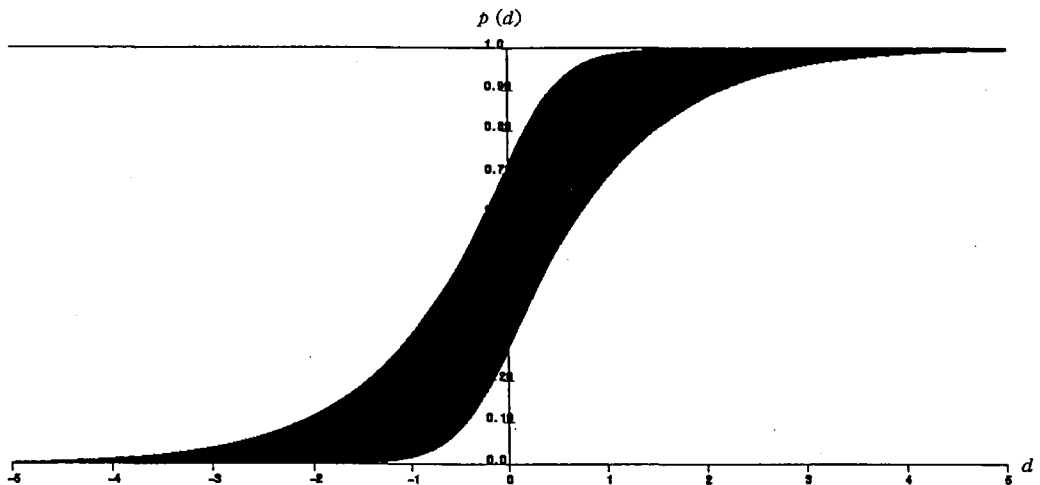
図5の最大確率法から作った $p(d)$ の信頼区間

図6 反応確率の信頼領域

個数が $k_1 = 0$ 、 $k_2 = 1$ 、 $k_3 = 3$ 、 $k_4 = 4$ の場合の信頼水準 $\delta = 0.95$ の信頼領域を図1に示す。図1で黒い点をうってある領域が信頼領域である。母数 β は $p(d) = \frac{1}{2}$ の傾きを示すため図では $\text{Tan}^{-1}(\beta)$ で示した。単位は角度である。

観測レベルが $d = 0$ に対称、かつ反応値も対称であるから信頼領域も $\alpha = 0$ に関し対称になっている。またこの例の場合も信頼領域は単一の閉領域にはなっていない。二項分布の場合もそうであったが、得られた信頼領域は圧倒的大部

分を占める一つの閉な領域と少しはなれたごく小さな点の集合となっている。最大確率法で信頼領域を求める場合最大の部分のみを求めることでほぼ十分と考えられる。この部分のみを求めるためのプログラム実行時間はもとの信頼領域を求めるものよりもはるかに短縮できる。以下信頼領域を求める場合最大部分のみを求め表示する。同じ条件の下で反応個数が $k=(1, 1, 4, 4)$ の信頼領域を図2に示す。この場合は非対称である。

図3は最大確率法と情報行列を用いて求めた信頼領域を共に示した。鎖線で示した領域が情報行列を用いたものである。この領域は α, β の楕円となるのであるが、 β の方を $\tan^{-1}(\beta)$ で示したため、上につまった形になっている。この場合観測個数が各レベルで10個ずつと図1の場合より大きく、信頼領域は小さくなっている。なお正規近似によって求めた信頼領域は最大値法によって求めたものより小さくでがちである。図4は観測個数が同じく10個ずつの場合に反応個数が $k=(0, 3, 9, 10)$ と非対称の場合の信頼領域である。

分布を決める母数 α, β 以外も分布の信頼領域から求めることができる。例えば各 $p(d)$ の信頼区間を図示したのが図6である。図5はもとの α, β の最大値法による信頼領域を示した。この信頼区間に“正確な”ものでなく信頼水準は $\delta=0.95$ より大きくなっている。

4. 結 論

環境中の化学物質や医薬品の安全性を考えるとき、何らかのモデルを設定し得られた観測値

からパラメータを推測し、外挿して実質的な安全量を推定することが行なわれる。このような場合、推定は信頼区間によって行なわれることが望ましい。この場合いくつかのパラメータを持つ分布を仮定し、データの個数が大きいとき、パラメータの信頼領域を中心極限定理を利用して求めることが多い。本論文では分布を正確に計算し、その信頼領域を与える新しい方法を試みた。この方法では得られた信頼領域が信頼水準を正確に計算し、与えた信頼領域も小さいものであることが示される。ここで与えたのは2つの母数をもつロジスティックモデルに限ったが、他のモデルに対しても適用可能なものである。2つの母数を含むため信頼領域を求めるプログラムを実行するにはレベル数が4、5個でそれぞれのレベルでの観測個数が10から20あたりでは10数分で得られるものである。その意味でも最大確率法は実用可能な手法といえ、また他の近似方法に対する規準を与えるといえよう。

参 考 文 献

- 1) Sterne, T. E. (1954). Some remarks on confidence or fiducial limits. *Biometrika* 41, 272-278.
- 2) Crow, E. L. (1956). Confidence intervals for proportion. *Biometrika* 43, 423-435.
- 3) Yamazoe, S. (1993). Confidence belts for the binomial parameter. *J. Japan Statist. Soc.* 23, 161-169.
- 4) Brand, R. J., Pinnock, D. E. and Jackson, L. (1973). Large Sample Confidence Bands for the logistic response curve and its inverse. *American Statistician*. 27, 157-160.