数理解析研究所講究録
第 272 巻 1976 年 61-71



# Incomplete Quantal Response Data にもとづく最尤解とその存在定理について

川崎医大　　仮谷太一
中村　忠

## §1．要約

標本に含まれる個体の tolerance ( or threshold ) value (閾値)が、個体ごとに定まる特定の区間に属するという情報――以下区間データとよぶ――にもとづいて，tolerance distribution の母数を推定する問題を考察する。いくつかの具体例を示した後　問題の定式化を行ない，母集団分布が正規分布である場合について　最尤解の存在条件を提示する。

なお　このような考え方は、quantal な反応の解析だけでなく、一般に正確な観測値が得がたい場合，微妙なまたは複雑な内容を答え易い尺度で押えていく調査，および脱落あり、新規加入あり、1回とばしありといった検診の時系列に似たデータを有効に利用する場合など　広い適用範囲をもっている。

## §2．序

生体に刺激を与えた場合、その刺激の応答として ある特定の反応が起こったか起こらなかったかだけが観測されることがある。このような応答は *quantal* であるといわれる。生物学・医学などの分野では しばしばこのような場合に遭遇する。

いま問題としている応答が *quantal* であるとき、特定の反応が起こるか起こらないかは 刺激の強さに依存するであろう。ところで ある一定の条件の下では、各個体には、刺激の強さに対して それぞれ固有の水準があり、その水準以上だと反応を起こすが、未満だと反応を起こさないようなある値が存在すると考えられる。この値のことを *tolerance* ( *or threshold* ) *value* (閾値)といい、*quantal* な事象の議論ではこの閾値の分布が問題となる。

*quantal* な事象の例を2・3あげてみよう。

(i) 特定の臓器に有害物質が蓄積され、それがある水準に達すると発病するが、その閾値には個体差が見られる。適当な母集団を考えれば これらの閾値は ある確率分布に従うであろう。

(ii) 乳児がある年令に達すると 乳歯の萌出(反応)が始まるが、その年令(刺激)は、適当な母集団をとれば 一定の確率分布に従う。

このほか、女児の初潮開始年令、ある種の心臓疾患の治癒に必要なジギタリスの投与量、分娩時死亡に対する出生時の体重など、類似の事象はきわめて多い。

さて、これらの場合に閾値の従う確率分布の母数を推定しようとすれば 2つの大きな問題につき当る。すなわち1つは標本の大きさと無作為性の問題であり、他の1つは個体から得られるデータが、普通の意味での観測値ではなく、個体ごとに定まる区間データであることである。例をあげてこれらの問題を具体的に述べることにしよう。

「ある地域の生後 a か月から b か月までの幼児」について何か月かの間隔で S 回乳歯の集団検診を実施し、特定の乳歯の萌出年令分布の母数推定を行なおうとする場合を考察することにしよう。

このとき S 回の検診にすべて参加した人だけを標本とみなすならば極端に偏ったものになり、しかも標本の大きさは非常に小さくなってしまうであろう。1回しか来なかった人、途中何回か欠診した人、2回来てあと来なかった人達などすべて標本に含めたい。こうすることによって始めて標本の大きさと無作為性の確保を可能にすることができる。従来このようなデータの処理にあたって全然工夫がなされていなかったわけではない。人年法（Pearson-year method）

，人日法（Pearson-day method）などと呼ばれているものがそれであるが，きわめて経験的であり人為的な便法に過ぎないように思われる。

一方萌出データの観点からは，標本に含まれる個体を3つのグループに分類することができる。すなわち

(a) 2回以上受診し，それら受診日の間に乳歯が萌出した人

(b) 最初の受診日にすでに萌出していた人

(c) 最後の受診日まで萌出のなかった人

である。この調査から得られる，信頼のおける確実な萌出データは，(a)グループに属する人の場合は，萌出年月日をはさむ直前直後の受診日における満年令を両端とする区間データであり，(b)グループに属する人の場合は分布の下限値と最初の受診日における満年令を両端とする区間データであり，(c)グループに属する人の場合は最後の受診日における満年令と分布の上限値とを両端とする区間データである。

S回の集団検診からなる全調査を終った時点で，全受診者（$n$人とする）について上述した区間データの端点をなす満年令をそれぞれ計算しいっしょにして大きさの順に並べ替えたものをあらためて

(1) $x_{(0)} < x_{(1)} < \cdots < x_{(j)} < \cdots < x_{(i)} < \cdots < x_{(m)} < x_{(m+1)}$

とする，ここに $x_{(0)}$，$x_{(m+1)}$ はそれぞれ分布の下限値，上限

値である。次にこれら $m+2$ 個の値の すべての組合わせを作り $(x_{(j)}, x_{(i)})$ $(j<i)$ が 上述した区間データであるときに限り $C_{ij}=1$ $(j<i)$，その他のとき $C_{ij}=0$ と定義する。区間データ $(x_{(0)}, x_{(m+1)})$ をもつような受診者はいないから $C_{m+1\,0}=0$ である。

ところで生年月日を連続変量と考えれば 受診者の受診時点における満年令は連続変量となり 同じ区間データに属する受診者が複数になる確率は 0 となる。このとき

$$
(2)\quad \begin{cases} C_{ij}=0 \text{ または } 1 \ (i>j), \quad C_{m+1\,0}=0 \\ \displaystyle\sum_{i=1}^{m+1}\sum_{j=0}^{i-1} C_{ij}=n \quad (n \text{ は全受診者数}) \\ \displaystyle\sum_{j=0}^{k-1} C_{kj}+\sum_{i=k+1}^{m+1} C_{ik}=1 \quad (k \neq 0,\ m+1) \end{cases}
$$

が成り立つ。

また、実際的なデータ処理としては 同一の生年月日の人がいる場合 同一の区間データをもつ人が 2人以上になる可能性も出てくるので，$C_{ij}$ については 次の関係が成立する。

$$
(2)' \quad \begin{cases} C_{ij}=0 \text{ または正整数} \ (i>j), \quad C_{m+1\,0}=0, \\ \displaystyle\sum_{i=1}^{m+1}\sum_{j=0}^{i-1} C_{ij}=n \quad (n \text{ は全受診者数}), \\ C_{k0}, C_{k1}, \cdots, C_{k\,k-1};\ C_{k+1\,k}, \cdots, C_{m+1\,k} \text{ のうち 1 個} \\ \text{を除き他はすべて 0 である} \quad (k \neq 0,\ m+1). \end{cases}
$$

母集団分布を $F(x, \theta)$ とすれば 対数尤度関数は

(3) $$M(\theta) = \sum_{i=1}^{m+1} \sum_{j=0}^{i-1} C_{ij}\, \ln \{ F(x_{(i)}, \theta) - F(x_{(j)}, \theta) \}$$

となるから，母数 $\theta$ の最尤推定の問題は 次のように定式化することができる。

## §3. 定式化と最尤解の存在定理

ここでは $F(x, \theta)$ が正規分布関数である場合について考察する。$\Theta$ は上半面を表わすものとし、$x_1, \cdots, x_m$ は $-\infty = x_0 < x_1 < \cdots < x_m < x_{m+1} = \infty$ をみたす実数とする。標準正規分布の分布関数，密度関数をそれぞれ $\Phi$，$\phi$ で表わし，表示を簡単にするために 次のような $\Theta$ または $\Theta \times R^*$ ($R^*$ は拡張された実数系）上で定義された関数を導入する．

$$t(x, \theta) \equiv (x - \mu)/\sigma \quad (x \in R), \quad t(\pm\infty, \theta) = \pm\infty$$

$$\phi_i(\theta) \equiv \phi(t(x_i, \theta)), \quad \Phi_i(\theta) \equiv \Phi(t(x_i, \theta)) \quad (0 \leqq i \leqq m+1).$$

$$\Phi_{ij}(\theta) \equiv \ln \{ \Phi_i(\theta) - \Phi_j(\theta) \} \quad (0 \leqq j < i \leqq m+1).$$

$$M(\theta) \equiv \sum_{i=1}^{m+1} \sum_{j=0}^{i-1} C_{ij}\, \Phi_{ij}(\theta),$$

ここに $\{C_{ij}\}$ $0 \leqq j < i \leqq m+1$ は，$C_{m+1,0}$ を除き，すべては $0$ でないような非負の実数とする。$\Phi_{m+1,0}(\theta) \equiv 0$ だから $M(\theta)$ は $C_{m+1,0}$ の値には無関係である。

さて上述した $\theta$ の最尤解の存在に関する問題は，次のように一般化して定式化することができる，

| |
|---|
| 問題 1.　$M(\theta^*) = \sup\{ M(\theta) : \theta \in \Theta \}$<br>となるような $\theta^* \in \Theta$ を求めよ。 |

以後上のような $\theta^*$ を最適解ということにし、これが存在するとき，問題 1 は最適解をもつということにする。$\{C_{ij}\}$ に対して，単に，これらのことごとくが 0 ではない非負の実数であるとの条件の下で問題 1 を定式化したのであるが，この条件は もちろん (2) または (2)′ を特殊の場合として含んでいる。$\{C_{ij}\}$ が (2) または (2)′ をみたすときは問題 1 の最適解 $\theta^*$ は 最尤解となる。

$M(\theta)$ の定義から，$M(\theta)$ は $\theta$ に関して連続であり，かつ $M(\theta) < 0$ $(\theta \in \Theta)$ であることは容易に分る。

$R^m$ の部分集合 $S$ に対し，$\partial S = \overline{S} \cap S^c$ とおく。また $F(\theta) = (\Phi_1(\theta), \cdots, \Phi_m(\theta))$ とし，$\Theta$ の $F$ による像を $F(\Theta)$ とおけば，$\partial F(\Theta)$ は次のように分解される。

定理 1. $\partial F(\Theta) = \bigcup_{k=0}^{m} L_k$，ここに $L_k = \{ z = (z_1, \cdots, z_m) \in R^m ;\ z_j = 0\ (j < k),\ 0 \leqq z_k \leqq 1,\ z_j = 1\ (j > k) \}$ $(1 \leqq k \leqq m)$ であり，$L_0 = \{ z = (z_1, \cdots, z_m) \in R^m ;\ z_1 = \cdots = z_m,\ 0 \leqq z_1 \leqq 1 \}$.

$L_k$ からその両端点を除いた集合を $\mathring{L}_k$ で表わす。このとき定理 1 より次のことが分る。

補助定理 1. $\lim_{n \to \infty} F(\theta_n) \in \mathring{L}_k$ $(1 \leqq k \leqq m)$ となる $\Theta$ 内の列 $\{\theta_n\}$ に対し，$\{M(\theta_n)\}$ が有界列であることと，$\{C_{ij}\}$ が条件

$(C_k)$：$j<i<k$ または $k<j<i$ なるすべての組 $(i,j)$ に対し $C_{ij}=0$；をみたすこととは同値である。

<u>補助定理2.</u> $\lim_{n\to\infty} F(\theta_n) \in \mathring{L}_0$ となる Ⓗ 内の列 $\{\theta_n\}$ に対し、$M(\theta_n)$ が有界列であることと，$\{C_{ij}\}$ が条件

$(C_0)$：$1 \leqq j<i \leqq m$ となるすべての組 $(i,j)$ に対し $C_{ij}=0$； をみたすこととは 同値である。

次に $F(Ⓗ)$ の境界付近での $M(\theta)$の値を計算するために，$\lambda \in (0,1)$ 上で関数 $\widetilde{M}_k(\lambda)$, $\widetilde{M}_0(\lambda)$ を次のように定義する。

$$\widetilde{M}_k(\lambda) \equiv \left(\sum_{j=0}^{k-1} C_{kj}\right) \ln \lambda + \left(\sum_{i=k+1}^{m+1} C_{ik}\right) \ln (1-\lambda), \quad 1 \leqq k \leqq m$$

$$\widetilde{M}_0(\lambda) \equiv \left(\sum_{i=1}^{m} C_{i0}\right) \ln \lambda + \left(\sum_{j=1}^{m} C_{m+1,j}\right) \ln (1-\lambda).$$

このとき

<u>補助定理3.</u> $\{C_{ij}\}$ は条件 $(C_k)$ をみたし，列 $\{\theta_n\}$ は $\lim_{n\to\infty} F(\theta_n) = (z_1, z_2, \cdots, z_m) \in \mathring{L}_k$ をみたすとする。このとき (i) $k \geqq 1 \Rightarrow \lim_{n\to\infty} M(\theta_n) = \widetilde{M}_k(z_k)$, (ii) $k=0 \Rightarrow \lim_{n\to\infty} M(\theta_n) = \widetilde{M}_0(z_1)$.

<u>補題1.</u> $f(\lambda) = a \ln \lambda + b \ln (1-\lambda)$, $\lambda \in (0,1)$ とする。$a>0$ かつ $b>0$ ならば $f(\lambda^*) = \sup\{f(\lambda); \lambda \in (0,1)\}$ となる $\lambda^*$ がただ一つ存在し，しかも $\lambda^* = a/(a+b)$ である。

次に Ⓗ 内の半直線上での $M(\theta)$の増減を調べるために 方向微分を導入しよう。任意の実数 $x$，および $\lambda \in (0,1)$ に対し

$D_\lambda(x) \equiv \{\theta = (\mu, \sigma) \in \Theta ;\ x - \mu = \sigma \Phi^{-1}(\lambda)\}$ （$\Phi^{-1}$ は $\Phi$ の逆関数）とおく。$D_\lambda(x)$ は $\Theta$ 内の半直線である。$\theta = (\mu, \sigma) \in D_\lambda(x)$ に対し，$\theta$ での $D_\lambda(x)$ に沿った $M(\theta)$ の方向微分 $M'_{\lambda,x}(\theta)$ は次式で定義される。$M'_{\lambda,x}(\theta) \equiv \lim_{\substack{\sigma' \to \sigma \\ \theta' = (\mu', \sigma') \in D_\lambda(x)}} (M(\theta') - M(\theta)) / (\sigma' - \sigma)$。このとき容易に次式が得られる。

(4) $\sigma^2 M'_{\lambda,x}(\theta) = \sum_{i=2}^{m} \sum_{j=1}^{i-1} C_{ij} [(x - x_i)\phi_i(\theta) - (x - x_j)\phi_j(\theta)] \cdot [\Phi_i(\theta) - \Phi_j(\theta)]^{-1} + \sum_{i=1}^{m} C_{i0}(x - x_i)\phi_i(\theta)\Phi_i(\theta)^{-1} - \sum_{j=1}^{m} C_{m+1,j}(x - x_j) \cdot \phi_j(\theta)[1 - \Phi_j(\theta)]^{-1}$.

この式と補題１および補助定理３より次の補題２を得る。

<u>補題２</u>．$\{C_{ij}\}$ は条件（$C_k$）（$1 \leqq k \leqq m$）のうち少くとも一つはみたすものとし，みたしている条件（$C_k$）の $k$ の集合を $I$ とおく。このとき各 $\theta \in \Theta$ と，各 $k \in I$ に対して

(5) $M(\theta) < \sup\{\tilde{M}_k(\lambda) ;\ \lambda \in (0, 1)\}$ が成立する。

補題２は問題１が最適解をもたない場合についての主張であるが，$\{C_{ij}\}$ が条件（$C_0$）をみたしている場合については次が成り立つ。いま $A(x) = \sum_{i=1}^{m} C_{i0}(x - x_i)$，$B(x) = \sum_{j=1}^{m} C_{m+1,j}(x - x_j)$，$g(\lambda) = [\lambda(1 - \lambda)]^{-1} \phi(\Phi^{-1}(\lambda))$.

(6) $G(\lambda, x) = g(\lambda)[A(x) - (A(x) + B(x))\lambda]$ とおく。

<u>補助定理４</u>．$\{C_{ij}\}$ が条件（$C_0$）をみたしているならば

$$\lim_{\sigma \to \infty} \sigma^2 M'_{\lambda,x}(\theta) = G(\lambda, x)$$

以上の準備のもとで目指す次の２定理が得られる。

<u>定理 2.</u> $\{C_{ij}\}$ は条件 $(C_0)$ をみたさないとする。このとき問題 1 が最適解をもつための必要十分条件は，$\{C_{ij}\}$ が条件 $(C_k)$ $(1\leqq k\leqq m)$ のいずれをもみたさないことである。

<u>定理 3.</u> $\{C_{ij}\}$ は条件 $(C_0)$ をみたし，条件 $(C_k)$ $(1\leqq k\leqq m)$ のいずれをもみたさないとき，$a=\sum_{i=1}^{m} C_{i0}>0$，$b=\sum_{j=1}^{m} C_{m+1,j}>0$ でかつ不等式 $\left[\sum_{i=1}^{m} x_i\,(C_{i0}+C_{m+1,i})\right]/(a+b)<\left[\sum_{i=1}^{m} x_i C_{i0}\right]/a$ が成立するならば 問題 1 は最適解をもつ。

最尤解の存在については、定理 2，3 から容易に次の定理 4，5 が導かれる。

<u>定理 4.</u> $\{C_{ij}\}$ が (2) または (2)′ の条件をみたし，かつ条件 $(C_0)$ をみたさないとする。このとき問題 1 が最尤解をもつための必要十分条件は，$\{C_{ij}\}$ が条件 $(C_k)$ $(1\leqq k\leqq m)$ のいずれをもみたさないことである。

<u>定理 5.</u> $\{C_{ij}\}$ が (2) または (2)′，および条件 $(C_0)$ をみたし，条件 $(C_k)$ $(1\leqq k\leqq m)$ のいずれをもみたさないとき，$a=\sum_{i=1}^{m} C_{i0}>0$，$b=\sum_{j=1}^{m} C_{m+1,j}>0$ で、かつ不等式

$$\frac{\sum_{i=1}^{m} x_i\,(C_{i0}+C_{m+1,i})}{a+b}<\frac{\sum_{i=1}^{m} x_i C_{i0}}{a}$$

が成り立つならば，問題 1 は最尤解をもつ。

〔参考文献〕

〔1〕 Kariya, T.：Incomplete Quantal Response Data Analysis. Kawasaki Medical Journal Vol 1. No. 2, pp 85～93, 1975.

〔2〕 仮谷太一：初潮年令分布の最尤推定．川崎医学会誌，第1巻第1号，pp 75～83, 1975.

〔3〕 Nakamura, T. & T. Kariya：On the Existence Theorem of the Optimal Solution of the Weighted Least-Square Problem. 川崎医学会誌一般教養篇 第1巻 pp 1～11, 1975.