

順序制約の下での情報量規準とペナルティー項について

安楽, 和夫
西南学院大学文学部 : 教授

<https://doi.org/10.15017/4363003>

出版情報 : 経済學研究. 66 (1), pp.37-50, 1999-06-30. 九州大学経済学会
バージョン :
権利関係 :

順序制約の下での情報量規準と ペナルティ項について

安 楽 和 夫

1. 序

k 個の母集団分布があり、それらのスカラー母数 $\theta_1, \theta_2, \dots, \theta_k$ の間に単順序: $\theta_1 \leq \theta_2 \leq \dots \leq \theta_k$ のような制約が仮定できるとする。真の母数の異なる値の組み合わせが

$$\theta_1 = \dots = \theta_{m_1} < \theta_{m_1+1} = \dots = \theta_{m_1+m_2} < \dots < \theta_{\sum_{j=1}^{k-1} m_j+1} = \dots = \theta_k$$

であるとき、 $\theta_j < \theta_{j+1}$ となるような変化点を見つけることがここでの問題である。

このような問題に対する統計的手法のひとつとして、多重比較検定法が挙げられよう (Williams (1971), Marcus (1976), 栗木 他 (1989), Hirotsu 他 (1992), Hoshino 他 (1996) 参照)。検定においては、通常有意水準を 0.05 などに設定するが、このような値は慣例的に用いられているに過ぎず、理論的根拠は持たないものである。帰無仮説 $\theta_1 = \dots = \theta_k$ を検定する検定問題において、帰無仮説を否定することに慎重でなければならない場合、このような有意水準を用いることは妥当と思われるが、そうでない場合に、通常の有意水準を用いることは必ずしも好ましくない。例えば、薬物の容量-反応実験で、投与量があるレベルに達すると、副作用や毒性が発現するようになるとする。このような変化点あるいは閾値を見つける場合に、有意水準を 0.05 のような小さな値に設定すると、一般に、検定による方法は変化点の検出に保守的となり、変化点を正しく見つけ出す確率は小さく押さえられてしまうことになる。有意水準をより大きな値に設定すると、このような誤りの確率はより小さくできるであろうが、実際に有意水準をいくらに設定するかを理論的に決めるのは難しいように思われる。このような意味で、上のような変化点の検出に対し、検定法からのアプローチには問題があるように思われる (永田・吉田 (1997), p. 28 参照)。

検定法に代わる統計的方法として、AIC (赤池情報量規準) (Akaike (1973, 1974)) のような情報量規準を用いたモデル選択法が考えられる。Kikuchi 他 (1993) は制約 $\theta_1 \leq \theta_2 \leq \dots \leq \theta_k$ の下での θ の最尤推定量に基づいて、AIC (赤池情報量規準) を計算する方法を提案した。これは多重比較検定などに比べて、変化点を検出する確率を大きく高める方法である (Yanagawa 他 (1994, 1997) 参照)。しかし、上のように順序制約のある問題に対しては、AIC の直接の適用は必ずしも適切ではないかも知れない。なぜなら、AIC におけるバイアス補正項は制約のない最尤推定量の漸近正規性から正当化されるが、

母数に順序制約があるような場合には、最尤推定量の漸近正規性は成り立たないからである。順序制約下での最尤推定量を用いた場合、このバイアス項は母数に依存し、一般に未知である。Anraku (1999)はこのバイアス項のとりうる値の範囲の中で、順序制約モデルにもっとも有利な値をペナルティー項として用いる方法を提案したが、このペナルティー項もバイアス補正項ではない。もし真のバイアス補正項を用いた情報量基準が使えるとすると、これとの比較に興味を持たれる。本稿では、分布を正規分布とし、あらかじめ設定したいくつかの場合について、真のバイアスをシミュレーションにより直接推定する。これをバイアス補正項として用いた情報量基準と Anraku (1999)の基準およびこれを修正したものとの比較を行う。

2. 順序制約下での情報量基準

2.1 準備

推定された確率分布 $\hat{g}(x)$ の、真の確率分布 $g(x)$ からの乖離度を測る尺度として Kullback-Leibler の情報量 (Kullback and Leibler (1951))

$$I(g, \hat{g}) = \int g(x) \log g(x) d\nu(x) - \int g(x) \log \hat{g}(x) d\nu(x)$$

を用いる。ただし、 g, \hat{g} は σ -有限な測度 ν に対する密度とする。上式の右辺第2項は期待対数尤度と呼ばれ、これを小さくする $\hat{g}(\cdot)$ が $g(\cdot)$ のより良い推定密度と考えられる。上式の右辺の2項はともに未知であるが、第2項は対数尤度で推定できるものである。 p 個の未知母数で表される確率モデルが真のモデルであるとするとき、母数の最尤推定量を用いた対数尤度は期待対数尤度の推定量と考えられる。そのときのバイアスは、真の分布が仮定した分布のクラスに属し、 p 個の未知母数への制約がある一定の条件を満たすものであれば、最尤推定量の漸近正規性より、漸近的には p で近似できる。AICはこのバイアスを補正した

$$AIC(\hat{\theta}) = l(\hat{\theta}) - p,$$

あるいは右辺を2倍したものとして定義される (Akaike (1973, 1974), Sakamoto 他 (1986), 竹内 (1976) 小西 (1998) 参照)。候補となるモデルの中で、この値が大きいほどより良いモデルと考えられる。しかし、 $\theta_1 \leq \theta_2 \leq \dots \leq \theta_k$ のような順序制約の下での最尤推定量については、漸近正規性は成り立たず、従って、バイアス項は AIC のものとは異なるものとなる。

一般に、モデルに多くの母数を用いるほどデータへの当てはまりが良くなるわけだが、AICのバイアス補正項はパラメータの濫用を抑制するようにも見られるためか、ペナルティー項とも呼ばれる。実際、バイアス補正としての意味を持たないペナルティー項を用いたモデル選択基準も提案されている。Schwarz (1978) の BIC などはベイズ流の考えから導かれたものであるが、対数尤度に、バイアス補正とは無関係なペナルティー項を加えたものとも見なせる (小西 (1998), Burnham and Anderson (1998) 参照)。Anraku (1999) の基準もバイアス補正項でないペナルティー項を用いたものである。

次の2.2節および2.3節で、正規分布を仮定したとき、順序制約のある最尤推定量を用いた場合の情

報量規準について, Anraku(1999)の結果を簡単にレビューし, バイアス項あるいはペナルティー項等について新たな考察を加える。

2.2 バイアス項の評価と情報量規準

今, 観測値 x_{ij} がそれぞれ正規分布 $N(\theta_i, \tau_i)$ (τ_i は既知) から得られたものであるとする ($i=1, \dots, k$; $j=1, \dots, n_i$)。また $\theta = (\theta_1, \dots, \theta_k)$ での対数尤度を $l(\theta)$ とし, 正規分布 $N(\theta_i, \tau_i)$ の密度関数を $f(\cdot, \theta_i, \tau_i)$ と表す。単順序 $\theta_1 \leq \theta_2 \leq \dots \leq \theta_k$ の仮定の下での θ_i の最尤推定量は次で与えられることが知られている (Robertson 他(1988), p. 24) :

$$\tilde{\theta}_i = \min_{t: t \geq \theta_i} \max_{s: s \leq \theta_i} \frac{\sum_{j=1}^{n_i} w_j \hat{\theta}_j}{\sum_{j=1}^{n_i} w_j}, \quad (i=1, \dots, k). \quad (1)$$

ただし, $w_j = n_j / \tau_j$ で, $\hat{\theta}_j$ は制約のない最尤推定量 $\hat{\theta}_j = n_j^{-1} \sum_{s=1}^{n_j} x_{js}$ である。また, $\tilde{\theta} = (\tilde{\theta}_1, \dots, \tilde{\theta}_k)$ での対数尤度は次のように与えられる:

$$l(\tilde{\theta}) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^k n_i \log \tau_i - \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^{n_i} \frac{(x_{ij} - \tilde{\theta}_i)^2}{\tau_i}.$$

ただし, $N = \sum_{i=1}^k n_i$ である。一方, $\tilde{\theta}$ を固定したとき期待対数尤度は

$$\sum_{i=1}^k n_i \int f(t, \theta_i, \tau_i) \log f(t, \tilde{\theta}_i, \tau_i) dt = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^k n_i \log \tau_i - \frac{N}{2} - \frac{1}{2} \sum_{i=1}^k \frac{n_i (\theta_i - \tilde{\theta}_i)^2}{\tau_i}$$

であり, これを $l(\tilde{\theta})$ で推定すると, 次のような意味でのバイアスを生じる:

$$\begin{aligned} B_l(\theta) &= E \left[l(\tilde{\theta}) - \sum_{i=1}^k n_i \int f(t, \theta_i, \tau_i) \log f(t, \tilde{\theta}_i, \tau_i) dt \mid \theta \right] \\ &= \frac{k}{2} + \frac{1}{2} E \left[\sum_{i=1}^k \frac{n_i (\tilde{\theta}_i - \theta_i)^2}{\tau_i} - \sum_{i=1}^k \frac{n_i (\hat{\theta}_i - \tilde{\theta}_i)^2}{\tau_i} \mid \theta \right] \end{aligned}$$

これは未知母数 θ に依存するものである。上の式で制約付きの最尤推定量 $\tilde{\theta}_i$ を制約無し最尤推定量 $\hat{\theta}_i$ に形式的に置き換えると, その値は k となり, AIC のバイアス項と等しくなる。

今, $H = \{y = (y_1, y_2, \dots, y_k) \mid y_1 = \dots = y_k\}$, $M = \{y = (y_1, \dots, y_k) \mid y_1 \leq \dots \leq y_k\}$ と置くと, 次の2つの不等式が成り立つ:

$$E \left[\sum_{i=1}^k n_i (\tilde{\theta}_i - \theta_i)^2 / \tau_i \mid \theta \in H \right] \leq E \left[\sum_{i=1}^k n_i (\tilde{\theta}_i - \theta_i)^2 / \tau_i \mid \theta \in M \right]$$

(Appendix の証明を参照);

$$E \left[\sum_{i=1}^k n_i (\hat{\theta}_i - \tilde{\theta}_i)^2 / \tau_i \mid \theta \in H \right] \geq E \left[\sum_{i=1}^k n_i (\hat{\theta}_i - \tilde{\theta}_i)^2 / \tau_i \mid \theta \in M \right]$$

(Perlman, 1969)。これより $\theta \in M$ のとき,

$$\sum P(i, k, w) i \leq B_l(\theta) < k \quad (2)$$

であり, 下限は $\theta \in H = \{y = (y_1, \dots, y_k) \mid y_1 = \dots = y_k\}$ の場合である (Anraku(1999))。ここで, $w = (w_1, \dots, w_k)$, $w_i = n_i / \tau_i$ であり, $P(i, k, w)$ は, $\theta \in H$ の下で, $\hat{\theta}_1, \dots, \hat{\theta}_k$ の中に異なるものが i 個ある確率であ

る ($1 \leq i \leq k$)。

このようにバイアス補正項 B_1 は θ に依存し、一般に未知である。これに対して Anraku (1999) は (2) 式の下限をペナルティー項として用い、次の規準を提案した。

$$ORIC(\tilde{\theta}) := l(\tilde{\theta}) - \inf_{\theta \in M} B_1(\theta) = l(\tilde{\theta}) - \sum_{i=1}^k P(i, k, w) i. \quad (3)$$

上のようなペナルティー項は $\theta \in M-H$ においては、本来のバイアス補正項より小さい。また $M \subset R^k$ より

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \hat{\theta}_i)^2 / \tau_i \leq \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \tilde{\theta}_i)^2 / \tau_i$$

となる (Robertson 他 (1988))。従って、常に $l(\hat{\theta}) - k \leq l(\tilde{\theta}) - k = (AIC(\hat{\theta}))$ となり、ペナルティー項として k を用いると、 $AIC(\hat{\theta})$ との比較は無意味となる。

モデル M は形式的にモデル H を含むものであるが、制約付きの最尤推定量を用いた ORIC と制約無し最尤推定量を用いた AIC の間には恒等的な大小関係はない。従って、比較可能である。例として、 $k=2$ の場合を考える。 $x_{ij} \sim N(\theta_i, 1)$ ($i=1, 2; j=1, \dots, n$) とし、3つのモデル $H: \theta_1 = \theta_2, M: \theta_1 \leq \theta_2, K: \theta_1 \neq \theta_2$ を考える。それぞれのモデルに対応する最尤推定量をそれぞれ $\hat{\theta}_H, \hat{\theta}_M, \hat{\theta}_K$ とすると、 H, K に対応する AIC と M に対応する ORIC はそれぞれ $AIC(\hat{\theta}_H) = l(\hat{\theta}_H) - 1, AIC(\hat{\theta}_K) = l(\hat{\theta}_K) - 2, ORIC(\hat{\theta}_M) = l(\hat{\theta}_M) - 3/2$ となり、これらの大小関係より3つのモデルの選択は次のように場合分けされる：

$$\begin{cases} M, & \text{if } \sqrt{n} \hat{\theta}_2 \geq \sqrt{n} \hat{\theta}_1 + \sqrt{2}, \\ H, & \text{if } \sqrt{n} \hat{\theta}_1 - 2 \leq \sqrt{n} \hat{\theta}_2 < \sqrt{n} \hat{\theta}_1 + \sqrt{2}, \\ K, & \text{if } \sqrt{n} \hat{\theta}_2 < \sqrt{n} \hat{\theta}_1 - 2. \end{cases}$$

$\sqrt{n} \hat{\theta}_1, \sqrt{n} \hat{\theta}_2$ をそれぞれ横軸と縦軸にとった場合のモデル選択の境界線を図1に示す。モデル M, H, K が選ばれるのは点 $(\sqrt{n} \hat{\theta}_1, \sqrt{n} \hat{\theta}_2)$ がそれぞれ、直線 (b) より上の領域、直線 (b) と (d) の間の領域、直線 (d) より下の領域にあるときである。 H, K についての AIC だけの比較であれば、直線 (a) と (d) の間の領域にあるときにモデル H が選ばれ、それ以外で K が選ばれる。直線 (a) と (b) に挟まれた領域に $(\sqrt{n} \hat{\theta}_1, \sqrt{n} \hat{\theta}_2)$ がある場合、AIC だけによる比較では、 H と判定され、ORIC と AIC では M と判定されることになる。標準正規分布関数を $\Phi(\cdot)$ と表すと、 $\theta_1 = \theta_2$ のとき、 $(\sqrt{n} \hat{\theta}_1, \sqrt{n} \hat{\theta}_2)$ がこの領域に入る確率は $\Phi(\sqrt{2}) - \Phi(1) = 0.08$ である。もし、問題が M か K のいずれかとなると、直線 (c) を境界とし、それより上方に $(\sqrt{n} \hat{\theta}_1, \sqrt{n} \hat{\theta}_2)$ があれば、モデル M を選択することになる。このような比較では、モデル M が H を含むことが反映されているように思われる。

ところで、(3) 式で与えた情報量規準の定義は、単順序 $\theta_1 \leq \theta_2 \leq \dots \leq \theta_k$ 以外の場合にも適用できるであろう。即ち M をある適当な条件を満たす制約領域とし、制約 $\theta \in M$ の下での θ の最尤推定量を $\tilde{\theta}_M = (\tilde{\theta}_1, \dots, \tilde{\theta}_k)'$ とする。また、 $l(\tilde{\theta}_M)$ で期待対数尤度を推定した場合のバイアスを $B_M(\theta)$ とするとき、上で与えた定義は次のように一般化されるであろう：

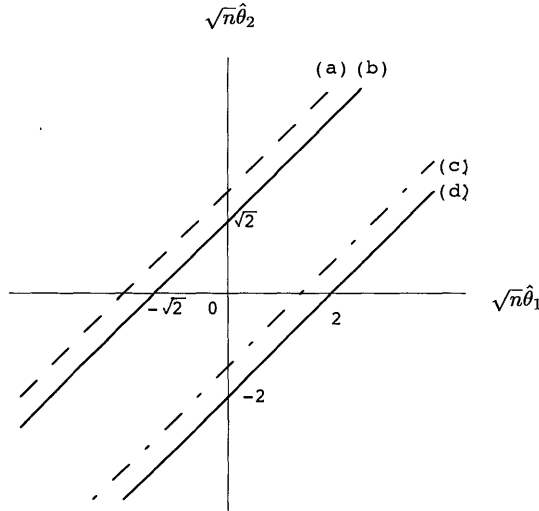


図1 モデル H, M, K の選択される境界線 (点 $(\sqrt{n}\hat{\theta}_1, \sqrt{n}\hat{\theta}_2)$ がそれぞれ、直線 (b) より上るとき M 、直線 (b) と (d) の間で H 、直線 (d) より小さいとき K が選択される。波線 (a) と (d) は、 $AIC(\hat{\theta}_H)$ と $AIC(\hat{\theta}_K)$ の選択の境界を表し、波線 (c) は $ORIC(\tilde{\theta}_M)$ と $AIC(\hat{\theta}_K)$ の2つを比較する場合の境界線である。)

$$ORIC_M(\tilde{\theta}_M) = l(\tilde{\theta}_M) - \inf_{\theta \in M} B_M(\theta).$$

このように定義すると、特に $M = \mathbf{R}^p$ の場合、上式は AIC に他ならない。すなわち、 $ORIC_H(\hat{\theta}_H) = AIC(\hat{\theta}_H)$ である。応用例として、単順序以外に、例えば次のような順序制約が挙げられる：

- (1) (simple tree order) $\theta_1 \leq \theta_i, (i=2, \dots, k)$,
- (2) (simple loop order) $\theta_1 \leq \theta_i \leq \theta_k, (i=2, \dots, k-1)$,
- (3) (umbrella order) $\theta_1 \leq \dots \leq \theta_p \geq \theta_{p+1} \geq \dots \geq \theta_k$.

ところで、2つのモデル $K: \theta \in \mathbf{R}^p, H: \theta = \theta(\beta) (\beta \in \mathbf{R}^q) (p > q)$ に対し、それぞれの最尤推定量を $\hat{\theta}_K, \hat{\theta}_H$ とすると、

$$2\{AIC(\hat{\theta}_K) - AIC(\hat{\theta}_H)\} = 2\{l(\hat{\theta}_K) - l(\hat{\theta}_H)\} - 2(p-q) \quad (4)$$

となる。上式の右辺の $2\{l(\hat{\theta}_K) - l(\hat{\theta}_H)\}$ は帰無仮説 $H: \theta \in H$ 、対立仮説 $K: \theta \in K-H$ に対する尤度比検定統計量であり、 $p-q$ は、帰無仮説 H の下での、尤度比検定統計量の期待値である(竹内, 1976)。 AIC と $ORIC$ の間にも同様に次が成り立つ(Anraku (1999))：

$$2\{ORIC(\tilde{\theta}_M) - AIC(\hat{\theta}_H)\} = 2\{l(\tilde{\theta}_M) - l(\hat{\theta}_H)\} - 2E(\bar{\chi}_{01}^2 | \theta \in H). \quad (5)$$

ここで $2\{l(\tilde{\theta}_M) - l(\hat{\theta}_H)\} = \bar{\chi}_{01}^2$ は、同等性の帰無仮説 (H) を単順序の対立仮説 (M) に対して検定する尤度比検定統計量である。

(4) 式および(5)式の右辺がそれぞれ正となれば、モデル K あるいは M がそれぞれ選ばれるわけだが、これは検定と類似している。(4) 式および(5)式の右辺がそれぞれ正となる確率の様子を図2に示

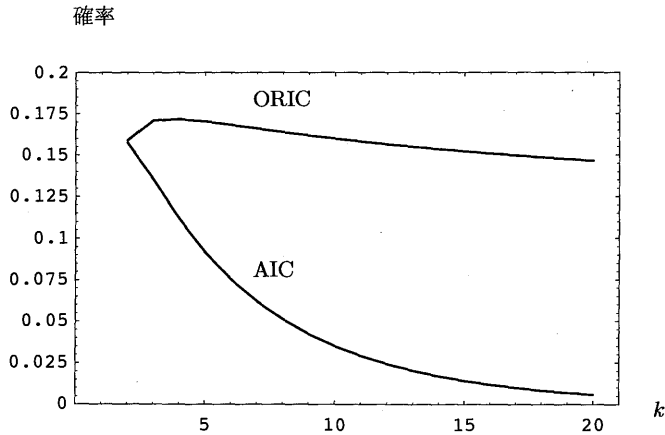


図2 確率 $P[\text{ORIC}(\hat{\theta}_M) > \text{AIC}(\hat{\theta}_H) | \theta \in H]$ と $P[\text{AIC}(\hat{\theta}_K) > \text{AIC}(\hat{\theta}_H) | \theta \in H]$ の様子

す。 k が大きくなるにつれ、AIC の方の確率は通常の有意水準よりも小さくなるのに対して、ORIC の方は、0.17前後から0.15へと減少は緩やかであり、AIC とはだいぶ異なっている。ORIC の場合の確率が急激に減少しないのは、ORIC のペナルティ項が H での値であり、 $\text{ORIC}(\hat{\theta}_M)$ と $\text{AIC}(\hat{\theta}_H)$ が近い値をとることによると考えられる。

2.3 分散比が既知の正規分布の場合

$x_{ij} \sim N(\theta_i, \tau_i \sigma^2)$ (τ_i は既知で、 σ^2 は未知) ($i=1, \dots, k, j=1, \dots, n_i$) とする。このとき、 $\theta_1 \leq \dots \leq \theta_k$ の下での θ_i の最尤推定量 $\hat{\theta}_i$ は(1)式と同じ形で与えられ、 σ^2 の最尤推定量は $\hat{\sigma}^2 = N^{-1} \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \hat{\theta}_i)^2$ で与えられる (Robertson 他(1988), p. 63)。このとき、期待対数尤度を対数尤度

$$l(\tilde{\theta}, \tilde{\sigma}^2) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^k n_i \log \tau_i - \frac{N}{2} \log \tilde{\sigma}^2 - \frac{N}{2}$$

で推定するとすると、次の意味でのバイアスを生じるであろう：

$$\begin{aligned} B_2(\theta, \sigma^2) &= E \left[l(\tilde{\theta}, \tilde{\sigma}^2) - \sum_{i=1}^k n_i \int f(t_i, \theta_i, \tau_i \sigma^2) \log f(t_i, \tilde{\theta}_i, \tau_i \tilde{\sigma}^2) dt_i \middle| \theta, \sigma^2 \right] \\ &= E \left[\frac{N}{2} \frac{\sigma^2}{\tilde{\sigma}^2} + \frac{1}{2} \sum_{i=1}^k n_i \frac{(\tilde{\theta}_i - \theta_i)^2}{\tilde{\sigma}^2} - \frac{N}{2} \middle| \theta, \sigma^2 \right]. \end{aligned}$$

ここで $B_2(\theta, \sigma^2)$ の最小値は $\theta \in H$ の時に得られ、 $\alpha = \sum_{i=1}^k P(i, k, w) i$, $\beta = \sum_{i=1}^k P(i, k, w) i^2$ とおくと、

$$E \left[\frac{\tilde{\sigma}^2}{\sigma^2} \middle| \theta \in H \right] = 1 - \frac{\alpha}{N}, \quad \text{Var} \left[\frac{\tilde{\sigma}^2}{\sigma^2} \middle| \theta \in H \right] = \frac{1}{N^2} (2N - 2\alpha + \beta - \alpha^2) \quad (6)$$

が成り立つ。Anraku(1999)は

$$\frac{N}{2} \left[E \left[\frac{\sigma^2}{\tilde{\sigma}^2} \middle| \theta \in H \right] - 1 \right] = 1 + \frac{\alpha}{2} + O(N^{-1}), \quad (7)$$

$$E\left[\sum_{i=1}^k \frac{n_i(\tilde{\theta}_i - \theta_i)^2}{\tau_i \tilde{\sigma}^2} \middle| \theta \in H\right] = \sum_{i=1}^k P(i, k, w) \frac{Ni}{N-i-2} = \alpha + O(N^{-1})$$

より, $B_2(\theta, \sigma^2) = \alpha + 1 + O(N^{-1})$ と評価して,

$$ORIC(\tilde{\theta}, \tilde{\sigma}^2) = l(\tilde{\theta}, \tilde{\sigma}^2) - \alpha - 1$$

を提案した。しかし, (6)式から(7)式の左辺ををそのまま評価すると,

$$\frac{N}{2} \left[E\left[\frac{\sigma^2}{\tilde{\sigma}^2} \middle| \theta \in H\right] - 1 \right] = \frac{(\alpha + 2)N^3 - (3\alpha^2 + 2\alpha - \beta)N^2 + \alpha^3 N}{2(N - \alpha)^3} + O(N^{-1})$$

となる。従って, $\theta \in H$ のとき

$$B_2(\theta, \sigma^2) = \frac{(\alpha + 2)N^3 - (3\alpha^2 + 2\alpha - \beta)N^2 + \alpha^3 N}{2(N - \alpha)^3} + \frac{1}{2} \sum_{i=1}^k P(i, k, w) \frac{Ni}{N-i-2} + O(N^{-1}) \quad (8)$$

である。(8)式をペナルティー項として用いた情報量規準を ORIC2 とする。

なお, $B_1(\theta)$ の場合と同様, $B_2(\theta, \sigma^2)$ は, $\theta \in M - H$ のとき, $\tilde{\theta} = \hat{\theta}$ とすると, これは AIC のペナルティー項と同じもので, $k+1$ と近似されるが, 正確には $N(k+1)/(N-k-2)$ である (Sugiura(1978))。

3. 変化点の検出について

真に異なる母数の組を見つけるために次の $2^{k-1} - 1$ 通りの場合について ORIC を計算する:

$$\theta_1 = \dots = \theta_{k-1} \leq \theta_k; \quad \theta_1 = \dots \leq \theta_{k-1} = \theta_k; \quad \dots; \quad \theta_1 \leq \dots \leq \theta_k.$$

また $\theta_1 = \dots = \theta_k$ の場合に対する AIC を計算し, これら 2^{k-1} 個の中で最大値を与えるものを見つける。これにより

$$\theta_1 = \dots = \theta_{m_1} < \theta_{m_1+1} = \dots = \theta_{m_1+m_2} < \dots < \theta_{\sum_{j=1}^{k-1} m_j+1} = \dots = \theta_k$$

とするモデルが選ばれたとすると, $\theta_{m_1}, \theta_{m_1+m_2}, \dots$, に変化点があるとみなされる。特に, $\theta_{m_1} < \theta_{m_1+1}$ となる θ_{m_1} を最初の変化点もしくは閾値とみなす。

4. シミュレーションによる比較評価について

ここでは, モンテカルロ・シミュレーションにより, 単順序が仮定できるとき, 真のモデルを同定するために, Kikuchi 他(1993)による AIC を使った方法と制約下での最尤推定量を用いた情報量規準の比較評価を行う。ここで, $k=4, \tau_1 = \dots = \tau_4 = 1$ とし, $x_{ij} \sim N(\theta_i, 1)$ ($i=1, \dots, 4; j=1, \dots, n_i$) となるような乱数を生成する。ただし, モデルにおいては分散は未知として扱う。また θ の組として次の 5 通りの場合を調べる:

$$\begin{aligned} C_0: (0, 0, 0, 0), & \quad C_1: (0, 0, 0, 0.1), & \quad C_2: (0, 0, 0.1, 0.1), \\ C_3: (0, 0.1, 0.1, 0.1), & \quad C_4: (0, 0.1, 0.2, 0.3). \end{aligned}$$

Kikuchi 他(1993)の方法では、はじめに $\theta_1 \leq \dots \leq \theta_k$ の下での最尤推定量を計算する。このとき $\theta_i (i=1, \dots, k)$ は m 個の異なる値をとるグループに分けられる ($1 \leq m \leq k$)。これに対して m 個のモデル

$$\theta(1) = \dots = \theta(j_0), \theta(j_0+1), \dots, \theta(m) \quad (j_0=1, \dots, m)$$

に対応する AIC を計算し、これらのうちで最大となるものがモデルとして選ばれることになる。

ORIC あるいは ORIC2においては、バイアス $B_2(\theta, \sigma^2)$ が未知であるために、とりうる値の範囲のうち、モデルに最も有利な値を用いるわけであるが、もしバイアスの真の値がわかるとすると、これを用いた方がバイアス補正という点から好ましいであろう。このような情報量規準を

$$ORIC3(\tilde{\theta}, \tilde{\sigma}^2) = I(\tilde{\theta}, \tilde{\sigma}^2) - B_2(\theta, \sigma^2)$$

とする。標本数を $n_1=n_2=n_3=n_4=10$ とし、上の C_0, C_1, \dots, C_4 の各場合について、 $B_2(\theta, \sigma^2)$ を 50,000 回のシミュレーションにより推定した結果が表 1 である。表中の H, M_1, M_2, \dots, M_7 はそれぞれ

$$\begin{aligned} H: \theta_1 = \theta_2 = \theta_3 = \theta_4, & \quad M_1: \theta_1 = \theta_2 = \theta_3 \leq \theta_4, & \quad M_2: \theta_1 = \theta_2 \leq \theta_3 = \theta_4, \\ M_3: \theta_1 \leq \theta_2 = \theta_3 = \theta_4, & \quad M_4: \theta_1 = \theta_2 \leq \theta_3 \leq \theta_4, & \quad M_5: \theta_1 \leq \theta_2 = \theta_3 \leq \theta_4, \\ M_6: \theta_1 \leq \theta_2 \leq \theta_3 = \theta_4, & \quad M_7: \theta_1 \leq \theta_2 \leq \theta_3 \leq \theta_4 \end{aligned}$$

を表し、表の値はそれらのモデルを仮定したときの $B_2(\theta, \sigma^2)$ の値の推定値である。

表 1 50,000回のシミュレーションによる $B_2(\theta, \sigma^2)$ の推定値
($n_1 = n_2 = n_3 = n_4 = 10$ の場合)

[仮定したモデル]

θ	H	M_1	M_2	M_3	M_4	M_5	M_6	M_7
C_0	2.140	2.730	2.722	2.721	3.091	3.197	3.087	3.432
C_1	2.136	2.857	2.793	2.759	3.249	3.347	3.175	3.603
C_2	2.136	2.812	2.868	2.803	3.273	3.346	3.269	3.649
C_3	2.136	2.769	2.793	2.848	3.179	3.346	3.244	3.603
C_4	2.121	2.963	2.992	2.952	3.504	3.632	3.499	3.982

ここで、真の θ に対して、これと矛盾する、異なったモデルを仮定した場合について考えてみる。真の $B_2(\theta, \sigma^2)$ と仮定したモデルが矛盾するような最も典型的な場合として、真の θ が M_7 のタイプであるのに、 H を仮定したとしよう。このとき $\hat{\theta} = N^{-1} \sum_{i=1}^k n_i \hat{\theta}_i = N^{-1} \sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij}$ とすると、

$$\frac{N\hat{\sigma}^2}{\sigma^2} = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \hat{\theta})^2 / \sigma^2$$

は自由度 $N-1$ 、非心率 $\lambda = \sum_{i=1}^k n_i (\theta_i - \bar{\theta})^2 / \sigma^2$ の非心 χ^2 分布に従う。ここで、 $\bar{\theta} = N^{-1} \sum_{i=1}^k n_i \theta_i$ である。また平均と分散はそれぞれ

$$E\left[\frac{\hat{\sigma}^2}{\sigma^2}\right] = \frac{N-1+\lambda}{N}, \quad Var\left[\frac{\hat{\sigma}^2}{\sigma^2}\right] = \frac{2(N-1+2\lambda)}{N^2}$$

で与えられる (Johnson 他(1995))。更に、 $1/x$ を $x = E(\hat{\sigma}^2 / \sigma^2)$ で Taylor 展開することにより

$$\frac{N}{2} \left[E\left[\frac{\sigma^2}{\hat{\sigma}^2}\right] - 1 \right] = \frac{N}{2} \left[\frac{1}{E(\hat{\sigma}^2 / \sigma^2)} + \frac{1}{\{E(\hat{\sigma}^2 / \sigma^2)\}^3} Var\left[\frac{\hat{\sigma}^2}{\sigma^2}\right] - 1 \right] + O(N^{-1})$$

$$= \frac{N}{2} \left[\frac{1-\lambda}{N-1+\lambda} + \frac{2N(N-1+2\lambda)}{(N-1+\lambda)^3} \right] + O(N^{-1})$$

を得る。一方、 N が大きいとき、 $\hat{\sigma}^2 \sim (N-1+\lambda)\sigma^2/N$ とすると、統計量

$$\frac{N}{2} \left(\frac{\sigma^2}{\hat{\sigma}^2} - 1 \right) + \frac{1}{2} \sum_{i=1}^k \frac{n_i(\hat{\theta}_i - \theta_i)^2}{\sigma^2} \quad (9)$$

の期待値はおよそ次のように近似できるであろう。

$$\begin{aligned} & \frac{N}{2} \left[\frac{1-\lambda}{N-1+\lambda} + \frac{2N(N-1+2\lambda)}{(N-1+\lambda)^3} + \frac{\sigma^2 + \lambda\sigma^2}{(N-1+\lambda)\sigma^2} \right] + O(N^{-1}) \\ &= \frac{N}{N-1+\lambda} + \frac{N^2(N-1+2\lambda)}{(N-1+\lambda)^3} + O(N^{-1}). \end{aligned}$$

ここで、 $g(\lambda) = N/(N-1+\lambda) + N^2(N-1+2\lambda)/(N-1+\lambda)^3$ とおくと、 $g(\lambda)$ は $\lambda \geq 0$ の範囲で λ の減少関数である。特に、 $g(0) = N(2N-1)/(N-1)^2$ である。このことは、真の θ に対して、これと矛盾する、異なったモデルを仮定した場合、バイアス項はより小さい値になってしまうことがあることを意味する。これは誤ったモデルに対して有利になることであり、真のバイアスあるいはその推定値を用いることがモデル選択においては必ずしも好ましくないことがあることを示唆している。

ORIC2 では(8)式からペナルティー項を計算する。ここでは、 $n_1=n_2=n_3=n_4$ より $w_1=w_2=w_3=w_4$ である。またモデル H, M_1, M_2, \dots, M_6 ではいくつかのグループをまとめることになる。例えば、 i から j ($i < j$) をまとめると、対応するウェイトは $\sum_{s=i}^j n_i$ となる。モデル H, M_1, \dots, M_7 に対応する w をそれぞれ $w(H), w(M_1), \dots, w(M_7)$ とすると、モデル H に対しては、 $P\{1, 1, w(H)\} = 1$ よりペナルティーの値 2.105 を得る。またモデル M_1, M_2, M_3 に対しては

$$P\{1, 2, w(M_i)\} = P\{2, 2, w(M_i)\} = \frac{1}{2} \quad (i=2, 3, 4)$$

であるから、2.688 を得る。さらに、モデル M_4, M_5, M_6 に対して $w(M_i) = (w_1^{(i)}, w_2^{(i)}, w_3^{(i)})$ とおくと

$$P\{1, 3, w(M_i)\} = \frac{1}{4} - \frac{1}{2\pi} \sin^{-1}(\rho^{(i)}), \quad P\{2, 3, w(M_i)\} = \frac{1}{2},$$

$$P\{3, 3, w(M_i)\} = \frac{1}{4} + \frac{1}{2\pi} \sin^{-1}(\rho^{(i)}).$$

ただし、 $\rho^{(i)} = -[w_1^{(i)}w_3^{(i)}/(w_1^{(i)}+w_2^{(i)})(w_2^{(i)}+w_3^{(i)})]1/2$ である (Robertson 他(1988)参照)。これより $w(M_4) = (20, 10, 10)$, $w(M_5) = (10, 20, 10)$, $w(M_6) = (10, 10, 20)$ より M_4, M_5, M_6 に対するペナルティー項はそれぞれ 3.079, 3.162, 3.079 となる。最後に、 $w_1=w_2=w_3=w_4$ のとき

$$P\{1, 4, w(M_7)\} = P\{3, 4, w(M_7)\} = \frac{1}{4}, \quad P\{2, 4, w(M_7)\} = \frac{11}{24}, \quad P\{4, 4, w(M_7)\} = \frac{1}{24}$$

となり (Robertson 他(1988))、ペナルティー項 3.392 を得る。ORIC における、 H, M_1, M_2, \dots, M_7 に対するペナルティー項がそれぞれ 2, 2.5, 2.5, 2.5, 2.804, 2.892, 2.804, 3.083 であり (Anraku(1999))、ORIC2 のペナルティー項はこれらよりそれぞれ大きな値となっている。また表 1 の C_0 に対応するペ

ナルティータ項の推定値により近い値となっている。

なお、一般の w に対して、 k が大きくなると ($k \geq 6$)、 $P(i, k, w)$ の正確な値を求めるのは一般に困難とされていた (Robertson 他 (1988), Sun (1988) 参照)。これに対し、Hayter and Liu (1996) は簡単な数値計算法によりこれを求める方法を提案している。この計算法により通常 k に対して実用的な値が得られると思われる (三輪 他 (1998) 参照)。

Kikuchi 他 (1993) による AIC を用いた方法と ORIC, ORIC2, ORIC3 のそれぞれについて、 H, M_1, \dots, M_7 の 8 つのモデルに対する規準量を計算し、その中で最大となったものを挙げるやり方で 10,000 回のシミュレーションを行った。このシミュレーションにおいて選ばれたモデルの頻度を表 2 に示す。真の θ が C_0 であるときは、 H が正しい選択となる。以下、 C_1, \dots, C_4 に対して、 M_1, M_2, M_3, M_7 がそれぞれ正しい選択となる。表からわかるように、AIC は、4 つの方法の中で、最も保守的である。即ち、 C_0 ではモデル H を選択する割合が高いが、変化点のあるモデルに対しては他の基準より選択率は低くなっており、特に、 M_2, M_3 ではその傾向が顕著である。ただし、モデル M_7 に対しては、他の方法より高い選択率を示している。この理由は、Kikuchi 他 (1993) の AIC による方法が、最初に最尤推定量を計算する段階で、特定のモデルに限定してしまうことによるかもしれない。一つの例として、4 つの母集団からの標本平均がそれぞれ 0.0, 0.5, 1.0, 1.1 のような場合、これらの値はそのまま $\theta_1, \dots, \theta_4$ の最尤推定量となり、AIC による方法ではこれから、(i) $\theta_1, \theta_2, \theta_3, \theta_4$, (ii) $\theta_1 = \theta_2, \theta_3, \theta_4$, (iii) $\theta_1 = \theta_2 = \theta_3, \theta_4$, (iv) $\theta_1 = \theta_2 = \theta_3 = \theta_4$ の 4 つのモデルについて AIC を計算し、この中で最大のものをモデルとして決めることになり、結果として、モデル (i) (M_7) が選ばれることになるかもしれない。これに対して、ORIC, ORIC2 および ORIC3 では 8 つのモデル全部について比較するので、この場合はモデル M_6 が選ばれる可能性が高い。このように、AIC を使った方法では、選択するモデルが制限されるため、このようなことが起きると考えられる。

最後に、全体的に M_7 の選択率は低いが、これは母数の差がさほど大きくなく、標本数も少ないためと考えられる。母数の差がより大きくなるほど、あるいは、標本数が大きくなるほど、正しい選択の比率は大きくなるのがシミュレーションからも確認できるが、ここでは割愛する。

表 2 10,000 回のシミュレーションにおける、AIC, ORIC, ORIC2, ORIC3 によるモデル選択の頻度分布 ($n_1 = n_2 = n_3 = n_4 = 10$ の場合)

AIC								
θ	H	M_1	M_2	M_3	M_4	M_5	M_6	M_7
C_0	8409	755	363	301	65	74	22	11
C_1	7878	1177	397	271	112	119	32	14
C_2	7785	950	636	303	150	109	47	20
C_3	7933	835	449	476	104	127	59	17
C_4	6489	1616	783	414	289	242	109	58

順序制約の下での情報量規準とペナルティー項について

ORIC								
θ	H	M_1	M_2	M_3	M_4	M_5	M_6	M_7
C_0	6661	1087	927	1087	70	107	57	4
C_1	5860	1625	1038	1093	128	169	79	8
C_2	5650	1273	1438	1246	121	157	107	8
C_3	5841	1115	1066	1612	78	180	99	9
C_4	4097	1652	1680	1700	238	389	212	32

ORIC2								
θ	H	M_1	M_2	M_3	M_4	M_5	M_6	M_7
C_0	6975	1007	848	1004	49	75	39	3
C_1	6209	1530	974	1024	83	127	47	6
C_2	5989	1204	1357	1170	88	112	74	6
C_3	6207	1026	1007	1502	57	134	59	8
C_4	4448	1610	1652	1649	166	309	141	25

ORIC3								
θ	H	M_1	M_2	M_3	M_4	M_5	M_6	M_7
C_0	6980	992	849	1005	53	74	45	2
C_1	6524	1306	934	1024	65	96	49	2
C_2	6455	1086	1148	1078	77	95	59	2
C_3	6551	1010	937	1275	65	104	56	2
C_4	5397	1439	1383	1423	96	165	93	4

表3 表2における最初の変化点選択の頻度分布

θ	AIR				ORIC			
	[1]	[2]	[3]	[4]	[1]	[2]	[3]	[4]
C_0	408	428	755	8409	1255	997	1087	6661
C_1	436	509	1177	7878	1349	1166	1625	5860
C_2	479	786	950	7785	1518	1559	1273	5650
C_3	679	553	835	7933	1900	1144	1115	5841
C_4	823	1072	1616	6489	2333	1918	1652	4097

θ	ORIC2				ORIC3			
	[1]	[2]	[3]	[4]	[1]	[2]	[3]	[4]
C_0	1121	897	1007	6975	1126	902	992	6980
C_1	1204	1057	1530	6209	1171	999	1306	6524
C_2	1362	1445	1204	5989	1234	1225	1086	6455
C_3	1703	1064	1026	6207	1437	1002	1010	6551
C_4	2124	1818	1610	4448	1685	1479	1439	5397

表2の結果を、最初の変化点の選択率としてまとめなおしたのが、表3である。 $\theta_1 = \dots = \theta_i < \theta_{i+1}$ となるときの i を $[i]$ と表している。また、 $\theta_1 = \dots = \theta_4$ の場合は $[4]$ とする。従って、 C_0, \dots, C_4 に対する正しい選択はそれぞれ $[4], [3], [2], [1], [1]$ である。この結果から、ORIC あるいは ORIC2 は C_0 を除いて、最初の変化点を高い率で選択していることがわかる。真の $B_2(\theta, \sigma^2)$ を推定した ORIC3 は C_0 の場合に、 H の選択率が ORIC や ORIC2 とほぼ同程度であるのに、他の場合の正しいモデルの選択率が ORIC や ORIC2 より劣っている。このように変化点をより敏感に検出する方法としては、ORIC あるいは ORIC2 による方法がバイアス補正項を用いた情報量基準 ORIC3 よりも有効であるように思われる。

5. バイアス項の推定における分散について

統計量(9)の分散について考える。 $\hat{\sigma}^2/\sigma^2$ の分散が $2(N-1+2\lambda)/N^2$ であるから、 $(N/2)(\sigma^2/\hat{\sigma}^2-1)$ の分散はおよそ $(N-1+2\lambda)/2$ となり、 N のオーダーである。また $\hat{\theta}$ と $\hat{\sigma}^2$ は独立であるから、

$$\text{Cov}\left[\frac{1}{\hat{\sigma}^2}, \frac{\sum_{i=1}^k n_i(\hat{\theta}-\theta_i)^2}{\hat{\sigma}^2}\right] = E\left[\sum_{i=1}^k n_i(\hat{\theta}-\theta_i)^2\right] \text{Var}\left[\frac{1}{\hat{\sigma}^2}\right] > 0.$$

つまり

$$\text{Cov}\left[\frac{N}{2}\left(\frac{\sigma^2}{\hat{\sigma}^2}-1\right), \frac{1}{2}\sum_{i=1}^k \frac{n_i(\hat{\theta}-\theta_i)^2}{\hat{\sigma}^2}\right] > 0$$

であるから、(9)の分散は $(N/2)(\sigma^2/\hat{\sigma}^2-1)$ の分散より大きく、 N のオーダーより小さくはないことがわかる。

このような点から、 $B_2(\theta, \sigma^2)$ をブートストラップ法等により直接推定しようとしても、分散が大きく、推定は不安定なものになるであろう。

このような問題に対して、Konishi and Kitagawa (1996) はブートストラップ推定で分散を低減する推定法 (variance reduction technique) を提案している。この方法が、母数に順序制約があるような問題についても有効に機能するか興味あるが、残念ながら本稿では取り上げられなかった。別の機会に報告できればと思う。

[謝辞] 確率 $P(i, k, w)$ の計算に関連して、Hayter 他(1996)、三輪 他(1998)の結果をご教示いただいた統計数理研究所の栗木哲助教授に感謝申し上げます。

参 考 文 献

- Akaike, H.(1973). "Information Theory and an Extension of the Mximum Likelihood Principle", In *2nd International Symposium on Information Theory*, Ed. B. N. Petrov and F. Csaki, pp.267-81. Budapest: Akademiai Kiado. (*Breakthroughs in Statistics 1*, (1992). Ed. S. Kotz and N. L. Johnson, pp.610-24. New York: Springer-Verlag. に再録)
- Akaike, H.(1974). "A New Look at the Statistical Model Identification", *IEEE Transactions on Automatic Control*, AC-19, 716-723.
- Anraku, K.(1999). "An Information Criterioin for Parameters under a Simple Order Restriction", *Biometrika* **86**, 141-152.

- Burnham, K.P. and Anderson, D.R.(1998). "Model Selection and Inference", New York: Springer-Verlag.
- Hater, A. and Liu, W.(1996). "A Note on the Calculation of $\Pr\{X_1 < X_2 < \dots < X_k\}$ ", *The American Statistician*, **50**, 365.
- Hirotsu, C., Kuriki, S. and Hayter, A.(1992). "Multiple Comparison Procedures based on the Maximal Component of the Cumulative Chi-Squared Statistic", *Biometrika* **79**, 381-92.
- Hoshino, N. and Seki, Y.(1996). "A Test Based on MDL Criterion for Comparing Increasing Dose Levels with a Zero Dose Control", *Communications in Statistics, Theory and Methods* **A25**, 1729-44.
- Johnson, N. L., Kotz, S. and Balakrishnan, N.(1995). "Continuous Univariate Distributions" Vol. 2, 2nd ed., New York: Wiley.
- Kikuchi, Y., Yanagawa, T. and Nishiyama, H.(1993). "Determining the No-Observed-Adverse-Effect Level in Continuous Response", In *Statistical Sciences and Data Analysis*, Ed. K. Matsusita, M. L. Puri and T. Hayakawa, pp.345-56, Netherlands: VSP International Science Publishers.
- Konishi, S. and Kitagawa, G.(1996). "Generalised Information Criteria in Model Selection", *Biometrika* **83**, 875-90.
- 小西貞則(1998). "情報量規準構成の理論と最近の発展", 応用統計学会第20回シンポジウム予稿集, 1-17.
- Kullback, S. and Leibler, R.A.(1951). "On Information and Sufficiency", *Annals of Mathematical Statistics* **22**, 79-86.
- 栗木哲, 広津千尋, Hayter, A. J.(1989). "累積カイ二乗の最大成分に基づく多重比較法-有意確率計算と用量水準比較への応用-", 応用統計学 **18**, 129-141.
- Marcus, R.(1976). "The Powers of Some Tests of the Equality of Normal Means against an Ordered Alternative", *Biometrika* **63**, 177-83.
- 三輪哲久, Hayter, A. J., Liu, W.(1998). "アンバランスモデルにおけるレベル確率計算と Bartholomew 検定", 第66回日本統計学会予稿集, 217-8.
- 永田靖, 吉田道弘(1997). 「統計的多重比較法の基礎」, サイエンス社.
- Perlman, M. D.(1969). "One-Sided Problems in Multivariate Analysis", *Annals of Mathematical Statistics* **40**, 549-67.
- Robertson, T., Wright, F. T. and Dykstra, R. L.(1988). "Order Restricted Statistical Inference", New York: Wiley.
- Sakamoto, Y., Ishiguro, M. and Kitagawa, G.(1986). "Akaike Information Criterion Statistics", Tokyo: D. Reidel.
- Schwarz, G.(1978). "Estimating the Dimension of a Model" *The Annals of Statistics* **6**, 461-4.
- Sugiura, N.(1978). "Further Analysis of the Data by Akaike's Information Criterion and the Finite Corrections", *Communication in Statistics, Theory and Methods* **A7**, 13-26.
- Sun, H. J.(1988). "A general reduction method for n-variate normal orthant probability", *Communications in Statistics, Theory and Methods* **A17**, 3913-21.
- 竹内啓(1976). "情報量統計量の分布とモデルの適切さの規準", 数理科学, No.153, pp.12-8.
- Williams, D. A.(1971). "A Test for Differences between Treatment Means when Several Dose Levels are Compared with a Zero Dose Control", *Biometrics* **27**, 103-17.
- Yanagawa, T., Kikuchi, Y. and Brown, K. G.(1994). "Statistical Issues on the No-Observed-Adverse-Effect Level in Categorical Response", *Environmental Health Perspectives Supplements* **Suppl. 1**, **102**, 95-101.
- Yanagawa, T., Kikuchi, Y. and Brown, K. G.(1997). "No-Observed-Adverse-Effect-Levels in Severity Data", *Journal of the American Statistical Association*, **92**, 449-454.

Appendix

$a, b \in \mathbf{R}^k$ に対し, 内積を $\langle a, b \rangle = a' \text{Diag}(w_1, \dots, w_k) b$, ノルムを $\|a\| = \sqrt{\langle a, a \rangle}$ で定義する. また, この内積のもとでの X の M への射影を $P(X|M)$ と表すと, $\tilde{\theta} = P(X|M)$ であり, 不等式は

$$E[\|P(X|M) - \theta\|^2 | \theta \in H] \leq E[\|P(X|M) - \theta\|^2 | \theta \in M]$$

と表せる. これを示すためには, 任意の $X, \theta_0 \in H, \lambda \in M, \theta_1 = \theta_0 + \lambda$ に対して,

$$\|P(X|M) - \theta_0\| \leq \|P(X + \lambda | M) - \theta_1\|$$

を示せばよい. 今, $P(X + \lambda | M) = \tilde{\theta}^* = (\tilde{\theta}_1^*, \dots, \tilde{\theta}_k^*)$ とおく. 一般に $\tilde{\theta}_1, \dots, \tilde{\theta}_k$ については

$$\tilde{\theta}_1 = \dots = \tilde{\theta}_{m_1} \leq \tilde{\theta}_{m_1+1} = \dots = \tilde{\theta}_{m_1+m_2} \leq \dots \leq \tilde{\theta}_{\sum_{j=1}^{h-1} m_j+1} = \dots = \tilde{\theta}_k$$

のように表せるが、 $\tilde{\theta} \leq \tilde{\theta}^*$ であるから、一般性を失うことなく、 $\tilde{\theta}_1 = \dots = \tilde{\theta}_k$ すなわち、 $P(X|M) = \tilde{\theta} \in H$ の場合に示せば十分である。今、線形空間 H の直行補空間を H^\perp とおくと、 $P(X|M) - X \in H^\perp$, $P(X + \lambda|M) - (X + \lambda) \in H^\perp$ だから

$$P(X + \lambda|M) - (X + \lambda) - \{P(X|M) - X\} = P(X + \lambda|M) - P(X|M) - \lambda \in H^\perp.$$

よって、 $\langle P(X + \lambda|M) - P(X|M) - \lambda, P(X|M) - \theta_0 \rangle = 0$ より、

$$\|P(X + \lambda|M) - \theta_1\|^2 = \|P(X + \lambda|M) - P(X|M) - \lambda\|^2 + \|P(X|M) + \lambda - \theta_1\|^2.$$

従って

$$\|P(X|M) - \theta_0\|^2 \leq \|P(X + \lambda|M) - \theta_1\|^2$$

を得る。

[西南学院大学文学部教授]