数理解析研究所講究録
第 1999 巻 2016 年 47-71

順序制約つき ANOVA モデルの AIC 規準

広島大学・理学研究科数学専攻　稲津 佑
Yu Inatsu
Department of Mathematics, Graduate School of Science, Hiroshima University

1. Introduction

本稿では順序制約が課された ANOVA モデルにおける AIC 規準について考察する．なお，本稿の内容は現在執筆中の論文 “ Akaike information criterion for ANOVA model with a simple order restriction ” (Hiroshima Statistical Research Group Technical Report にて公開予定) の要約である．詳細な議論や証明はこちらを参考されたい．

実データ解析において，解析者は様々な統計モデルの構築を行う．このとき，解析者はモデル (e.g., パラメータ，分布，構造，etc.) に対して何らかの正則条件を仮定している場合が多い．例えば，あるパラメータの推定を尤度最大化に基づき行う場合，すなわち，最尤推定量 (MLE) に基づいた推測を行うときは MLE が (対数) 尤度の極大化点である (i.e., MLE が尤度方程式の解である) という正則条件を仮定することが多い．この正則条件を仮定することで，MLE の一致性や漸近正規性，漸近有効性が導かれ，また，MLE に基づく統計量の漸近性質，例えば，AIC の罰則項が 2 倍のパラメータ数となることや尤度比統計量の帰無分布がカイ二乗分布に分布収束すること等が導かれる．すなわち，正則条件が成立しているときは”精度”と”扱いやすさ”の 2 つの観点から見て良い結果が導かれる．

一方，正則条件が成立しない場合の代表例として，パラメータに対する順序制約が挙げられる．特に，パラメータ $\theta_1,\ldots,\theta_k$ に対する simple order restriction (SO) と呼ばれる制約，$\theta_1 \le \theta_2 \le \cdots \le \theta_k$ は応用の上で非常に重要である．順序制約を課すことの意義の 1 つとして，事前に得られた情報を，あるいは自然に考えられるべきであろう仮定をモデルに反映することで，推定量の精度の改善が期待できる点が挙げられる．実際，互いに独立に正規分布 $N(\mu_i,\sigma^2/n_i)$ に従う確率変数 X_i, $(i=1,\ldots,k,\ n_i>0)$, におけるパラメータ μ_i の推定に関しては，SO の仮定が正しいとき，SO の下での MLE は通常の MLE に比べ精度が改善されている．具体的には，順序制約が無い下での μ_i の MLE, $\hat{\mu}_i$ は $\hat{\mu}_i = X_i$ で与えられ，SO の下での MLE, $\hat{\mu}_{i,\mathrm{SO}}$ は Robertson *et* *al.* (1988) より

$$\hat{\mu}_{i,\mathrm{SO}} = \min_{v;i\le v}\max_{u;u\le i}\frac{\sum_{j=u}^{v} n_j X_j}{\sum_{j=u}^{v} n_j}, \quad (i=1,\ldots,k),$$

で与えられ, この $\hat{\mu}_i$, $\hat{\mu}_{i,\mathrm{SO}}$ に対して Brunk (1965), Lee (1981) および Kelly (1989) がそれぞれ

(a) $\sum_{i=1}^{k} n_i \mathrm{E}[(\hat{\mu}_i-\mu_i)^2] > \sum_{i=1}^{k} n_i \mathrm{E}[(\hat{\mu}_{i,\mathrm{SO}}-\mu_i)^2]$,

(b) $\mathrm{E}[(\hat{\mu}_i-\mu_i)^2] > \mathrm{E}[(\hat{\mu}_{i,\mathrm{SO}}-\mu_i)^2], \quad (i=1,\ldots,k)$,

(c) $\mathrm{P}(|\hat{\mu}_{i,\mathrm{SO}}-\mu_i|\le t) > \mathrm{P}(|\hat{\mu}_i-\mu_i|\le t), \quad (t>0,\ i=1,\ldots,k)$,

を示している. 加えて, Hwang and Peddada (1994) は (c) の結果を楕円分布の場合に対してまで拡張している. 従って, 順序制約を課すことで"精度"の意味で正則条件を仮定した場合より良い結果が得られることが期待できる.

しかしながら, 順序制約を課した場合, "扱いやすさ"の点に関しては問題がある. Anraku (1999) は k クラスターの ANOVA における母平均 $\mu_1,\ldots,\mu_k$ に SO を仮定した下での AIC について考察しており, 通常の AIC はリスクに対する漸近不偏な推定量ではないことを示している. 加えて, そのバイアスは未知パラメータの真値に依存してしまうことを明らかにしている. また, Yokoyama (1995) は Yokoyama and Fujikoshi (1993) で扱われたパラレルプロファイルモデルにおいて, 基準化変換後に現れる分散パラメータ σ^2, τ^2 に対して SO を仮定した下での尤度比検定について考察しており, 帰無分布が必ずしもカイ二乗分布に分布収束するとは限らないことを示している. 加えて, 帰無分布の収束先は未知パラメータ τ^2 の真値に依存してしまうことも明らかにしている. この 2 つの例からもわかるように, 比較的単純な制約である SO の場合であっても, 導かれる結果は扱いやすいとは言い難いものとなってしまう.

以上を踏まえ, 本論文では特に SO の下での ANOVA における AIC 規準に着目する. 未知パラメータの真値に依存してしまう漸近バイアスの不偏な推定量を導出し, 漸近バイアスのない AIC を導出することで, "扱いやすさ"の問題の改善を図る. なお, 一般的な状況下での通常の AIC の期待値はリスクに対して N^{-1} のオーダーの誤差を持つため, 本論文で導出する AIC も誤差のオーダーが $O(N^{-1})$ となるまでバイアス補正を行う.

ここで, 本論文における記号の使い方についていくつか約束しておく. ベクトルの右肩のプライム, $'$ をベクトルの転置を意味するものとする. また, $\mathbf{1}_p$ を 1 を p 個並べた p 次元列ベクトル, $\mathbf{0}_p$ を 0 を p 個並べた p 次元列ベクトルとする. 後の都合上, $\mathbf{0}_0 = 0$ と定義しておく. 更に, p 次元ベクトル $\boldsymbol{x} = (x_1,\ldots,x_p)', \boldsymbol{y} = (y_1,\ldots,y_p)' \in \mathbb{R}^p$, お

よび $\boldsymbol{a} = (a_1, \ldots, a_p)' \in \mathbb{R}^p_{>0}$ に対して, 内積 $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\boldsymbol{a}}$ およびノルム $\|\boldsymbol{x}\|_{\boldsymbol{a}}$ をそれぞれ

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\boldsymbol{a}} = \sum_{i=1}^{p} a_i x_i y_i, \quad \|\boldsymbol{x}\|_{\boldsymbol{a}} = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle_{\boldsymbol{a}}} = \sqrt{\sum_{i=1}^{p} a_i x_i^2},$$

と定義する. $\|\boldsymbol{x}\|_{\boldsymbol{a}}$ は完備なノルムであることに注意されたい. また, $\boldsymbol{x} \geq \boldsymbol{0}_p$ および $\boldsymbol{x} \geq \boldsymbol{y}$ をそれぞれ

$$\boldsymbol{x} \geq \boldsymbol{0}_p \Leftrightarrow x_i \geq 0 \ (1 \leq i \leq p),$$
$$\boldsymbol{x} \geq \boldsymbol{y} \Leftrightarrow \boldsymbol{x} - \boldsymbol{y} \geq \boldsymbol{0}_p,$$

で定義する. また, ある命題 P に対して, 定義関数 $1_{\{P\}}$ を

$$1_{\{P\}} = \begin{cases} 1 & \text{if} \quad P \text{ is true} \\ 0 & \text{if} \quad P \text{ is not true} \end{cases},$$

で定義する. 次に, 確率変数 (ベクトル) $\boldsymbol{X}$ と $\boldsymbol{Y}$ に対して, $\boldsymbol{X}$ と $\boldsymbol{Y}$ が互いに独立であることを $\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y}$ で表すことにする. 最後に, 有限集合 A に対して, A の要素数を $\#A$ で表すことにする.

2. 母平均に順序制約が課せられた ANOVA モデル

X_{ij} を第 i クラスターにおける第 j 番目の個体から得られた目的変数とする. ただし, $i = 1, \ldots, k$, $j = 1, \ldots, N_i$ である. また, $k \geq 2$, $N = N_1 + \cdots, N_k$ とし, $N - k - 6 > 0$ であるとする. $X_{11}, \ldots, X_{kN_k}$ は互いに独立な確率変数であるとし, 以下のモデル

$$X_{ij} \sim N(\theta_i, \sigma^2), \tag{2.1}$$

を仮定する. ただし, $\theta_1, \ldots, \theta_k$ は未知の平均パラメータであり, $\sigma^2 > 0$ は未知の分散パラメータである. さらに, $\theta_1, \ldots, \theta_k$ に対して, simple order restriction (SO)

$$\theta_1 \leq \theta_2 \leq \cdots \leq \theta_k, \tag{2.2}$$

を仮定する. 集合 Θ を $\Theta = \{(\theta_1, \ldots, \theta_k)' \in \mathbb{R}^k \mid \theta_1 \leq \theta_2 \leq \cdots \leq \theta_k\}$ とおくとき, 制約 (2.2) の下でのモデル (2.1) は, 平均パラメータを Θ 上に制約した下での ANOVA モデルである. ここで, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)'$ とおく. また, $\boldsymbol{\theta}$ と σ^2 の真値をそれぞれ $\boldsymbol{\theta}_* = (\theta_{1,*}, \ldots, \theta_{k,*})'$, σ_*^2 とおく. 真値 $\boldsymbol{\theta}_*$ および σ_*^2 に対して, $\boldsymbol{\theta}_* \in \Theta$ および $\sigma_*^2 > 0$ を仮定する.

2.1. 順序制約下での最尤推定量

本小節では, SO の下でのモデル (2.1) における未知パラメータの最尤推定量 (MLE) の導出を行う. $\boldsymbol{N} = (N_1, \ldots, N_k)'$ とおく. $\boldsymbol{X}$ をすべての目的変数を並べたベクトルとする. すなわち, $\boldsymbol{X} = (X_{11}, \ldots, X_{ij}, \ldots, X_{kN_k})'$ である. また, $1 \le i \le k$ なる各 i に対して, 確率変数 $\bar{X}_i$, $\bar{\sigma}^2$ を

$$\bar{X}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} X_{ij}, \quad \bar{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{N_i} (X_{ij} - \bar{X}_i)^2, \tag{2.3}$$

と定義する. すなわち, $\bar{X}_i$, $\bar{\sigma}^2$ は標本平均, 標本分散である. また, $\bar{\boldsymbol{X}} = (\bar{X}_1, \ldots, \bar{X}_k)'$ とおく. SO の制約が無い通常の ANOVA モデルにおいては, $\boldsymbol{\theta}$, σ^2 の MLE はそれぞれ $\bar{\boldsymbol{X}}$, $\bar{\sigma}^2$ であることに注意されたい. さて, X_{ij} の正規性と独立性から, 対数尤度関数 $l(\boldsymbol{\theta}, \sigma^2; \boldsymbol{X})$ は

$$\begin{aligned} l(\boldsymbol{\theta}, \sigma^2; \boldsymbol{X}) &= -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{k} \sum_{j=1}^{N_i} (X_{ij} - \theta_i)^2 \\ &= -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{k} \sum_{j=1}^{N_i} (X_{ij} - \bar{X}_i)^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{k} N_i (\bar{X}_i - \theta_i)^2, \end{aligned} \tag{2.4}$$

で与えられる. 故に, 任意の $\sigma^2 > 0$ に対して, $l(\boldsymbol{\theta}, \sigma^2; \boldsymbol{X})$ の Θ 上での最大化は

$$H(\boldsymbol{\theta}) = \sum_{i=1}^{k} N_i (\bar{X}_i - \theta_i)^2 = \|\bar{\boldsymbol{X}} - \boldsymbol{\theta}\|_{\boldsymbol{N}}^2,$$

の Θ 上での最小化を考えればよい. あるいは, $H^*(\boldsymbol{\theta}) = \sqrt{H(\boldsymbol{\theta})} = \|\bar{\boldsymbol{X}} - \boldsymbol{\theta}\|_{\boldsymbol{N}}$ の Θ 上での最小化を考えてもよい. ここで, ノルム $\|\cdot\|_{\boldsymbol{N}}$ は完備なノルムであり, 集合 Θ は空でない閉凸集合であるから, 任意の $\bar{\boldsymbol{X}} \in \mathbb{R}^k$ に対して, Θ 上での $H^*(\boldsymbol{\theta})$ の最小化点がただ一つ存在する (see, e.g., Rudin, 1986). すなわち, $\boldsymbol{\theta}$ の SO の下での MLE, $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)'$ の一意存在性は保証されており, $1 \le i \le k$ なる i に対し, $\hat{\theta}_i$ は

$$\hat{\theta}_i = \min_{v; i \le v} \max_{u; u \le i} \frac{\sum_{j=u}^{v} N_j \bar{X}_j}{\sum_{j=u}^{v} N_j}, \tag{2.5}$$

で与えられる (see, e.g., Robertson *et al.*, 1988). 一方, σ^2 の MLE, $\hat{\sigma}^2$ は $l(\hat{\boldsymbol{\theta}}, \sigma^2; \boldsymbol{X})$ が σ^2 に関して凹関数であることに注意すれば

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{N_i}(X_{ij} - \bar{X}_i)^2 + \frac{1}{N}\sum_{i=1}^{k} N_i(\bar{X}_i - \hat{\theta}_i)^2,$$

となる.

2.2. リスク関数とバイアス

本小節では, K-L ダイバージェンスに基づくリスク関数の定義, および, それを対数尤度で推定した際のバイアスについて考える. $\boldsymbol{X}^\star$ を $\boldsymbol{X}$ と互いに独立に同一の分布に従う確率変数ベクトルとする. このとき, K-L ダイバージェンスに基づくリスク, R を

$$\begin{aligned} R &= \mathrm{E}[\mathrm{E}_\star[-2l(\hat{\boldsymbol{\theta}}, \hat{\sigma}^2; \boldsymbol{X}^\star)]] \\ &= \mathrm{E}\left[N\log(2\pi\hat{\sigma}^2) + \frac{N\sigma_*^2}{\hat{\sigma}^2} + \frac{\sum_{i=1}^{k} N_i(\theta_{i,*} - \hat{\theta}_i)^2}{\hat{\sigma}^2}\right], \end{aligned} \tag{2.6}$$

で定義する. なお, 通常の ANOVA モデルにおいては, リスク $\bar{R}$ は $\bar{R} = \mathrm{E}[\mathrm{E}_\star[-2l(\bar{\boldsymbol{X}}, \bar{\sigma}^2; \boldsymbol{X}^\star)]]$ である. 一方, 最大対数尤度 $l(\hat{\boldsymbol{\theta}}, \hat{\sigma}^2; \boldsymbol{X})$ は

$$l(\hat{\boldsymbol{\theta}}, \hat{\sigma}^2; \boldsymbol{X}) = -\frac{N}{2}\log(2\pi\hat{\sigma}^2) - \frac{N}{2}, \tag{2.7}$$

であるから, リスク, R を $-2l(\hat{\boldsymbol{\theta}}, \hat{\sigma}^2; \boldsymbol{X})$ で推定した際のバイアス, B は

$$B = \mathrm{E}[R - \{-2l(\hat{\boldsymbol{\theta}}, \hat{\sigma}^2; \boldsymbol{X})\}] = \mathrm{E}\left[\frac{N\sigma_*^2}{\hat{\sigma}^2}\right] + \mathrm{E}\left[\frac{\sum_{i=1}^{k} N_i(\theta_{i,*} - \hat{\theta}_i)^2}{\hat{\sigma}^2}\right] - N, \tag{2.8}$$

で与えられる.

次に, B を詳しく評価していく. 確率変数 S および T をそれぞれ

$$S = \frac{1}{\sigma_*^2}\sum_{i=1}^{k}\sum_{j=1}^{N_i}(X_{ij} - \bar{X}_i)^2, \quad T = \frac{1}{\sigma_*^2}\sum_{i=1}^{k} N_i(\bar{X}_i - \hat{\theta}_i)^2,$$

と定義する. X_{ij} の正規性と独立性から, S は自由度 $N-k$ のカイ二乗分布に従い, $S \perp \bar{\boldsymbol{X}}$ であることに注意されたい. また, (2.5) より, $\hat{\boldsymbol{\theta}}$ は確率変数ベクトル $\bar{\boldsymbol{X}}$ の関数であるから, T も $\bar{\boldsymbol{X}}$ の関数であり, $S \perp T$ であることにも注意されたい. この S, T を用いると, $N\hat{\sigma}^2/\sigma_*^2 = S + T$ と書けることを利用して B を評価していく.

$$\frac{N\sigma_*^2}{\hat{\sigma}^2} = \frac{N^2}{N\hat{\sigma}^2/\sigma_*^2} = \frac{N^2}{S+T} = \frac{N^2}{S}\frac{1}{1+T/S}, \tag{2.9}$$

であり, $x \geq 0$ なる x に対して, $(1+x)^{-1} = 1 - x + c^* x^2$, ただし, $0 \leq c^* \leq 1$ とかけることを利用すれば, (2.9) は

$$\frac{N\sigma_*^2}{\hat{\sigma}^2} = \frac{N^2}{S} - \frac{N^2 T}{S^2} + C^* \frac{N^2 T^2}{S^3},$$

とかける. ただし, C^* は $0 \leq C^* \leq 1$ なる確率変数である. よって, $S \sim \chi^2_{N-k}$, $S \perp T$ であることを利用すれば

$$\begin{aligned} \mathrm{E}\left[\frac{N\sigma_*^2}{\hat{\sigma}^2}\right] &= \frac{N^2}{N-k-2} - \frac{N^2 \mathrm{E}[T]}{(N-k-2)(N-k-4)} + \mathrm{E}\left[C^* \frac{N^2T^2}{S^3}\right] \\ &= N + k + 2 + O(N^{-1}) - \mathrm{E}[T] + O(N^{-1})\mathrm{E}[T] + \mathrm{E}\left[C^* \frac{N^2T^2}{S^3}\right], \end{aligned} \tag{2.10}$$

を得る. 一方, $y \geq 0$ なる y に対して, $(1+y)^{-1} = 1 - c^\star y$, ただし, $0 \leq c^\star \leq 1$ とかけることを利用すれば

$$\begin{aligned} &\frac{\sum_{i=1}^k N_i(\theta_{i,*} - \hat{\theta}_i)^2}{\hat{\sigma}^2} \\ &= \frac{N}{\sigma_*^2} \frac{\sum_{i=1}^k N_i(\theta_{i,*} - \hat{\theta}_i)^2}{S+T} = \frac{N}{\sigma_*^2} \frac{\sum_{i=1}^k N_i(\theta_{i,*} - \hat{\theta}_i)^2}{S} \frac{1}{1+T/S} \\ &= \frac{N}{\sigma_*^2} \frac{\sum_{i=1}^k N_i(\theta_{i,*} - \hat{\theta}_i)^2}{S} - C^\star \frac{NT}{\sigma_*^2} \frac{\sum_{i=1}^k N_i(\theta_{i,*} - \hat{\theta}_i)^2}{S^2} \\ &= \frac{N}{\sigma_*^2} \frac{\sum_{i=1}^k N_i(\theta_{i,*} - \bar{X}_i + \bar{X}_i - \hat{\theta}_i)^2}{S} - C^\star \frac{NT}{\sigma_*^2} \frac{\sum_{i=1}^k N_i(\theta_{i,*} - \hat{\theta}_i)^2}{S^2} \\ &= \frac{N}{\sigma_*^2} \frac{\sum_{i=1}^k N_i(\theta_{i,*} - \bar{X}_i)^2}{S} - \frac{2N}{\sigma_*^2} \frac{\sum_{i=1}^k N_i(\bar{X}_i - \theta_{i,*})(\bar{X}_i - \hat{\theta}_i)}{S} \\ &\quad + \frac{NT}{S} - C^\star \frac{NT}{\sigma_*^2} \frac{\sum_{i=1}^k N_i(\theta_{i,*} - \hat{\theta}_i)^2}{S^2}, \end{aligned}$$

を得る. ただし, $C^\star$ は $0 \leq C^\star \leq 1$ なる確率変数である. ここで, $1 \leq i \leq k$ なる i に対して $S \perp \bar{X}_i$, $S \perp \hat{\theta}_i$, $S \perp T$ であること, および $\bar{X}_i \sim N(\theta_{i,*}, \sigma_*^2/N_i)$ であることに

注意すれば

$$
\begin{aligned}
&\mathrm{E}\left[\frac{\sum_{i=1}^{k} N_i(\theta_{i,*}-\hat{\theta}_i)^2}{\hat{\sigma}^2}\right] \\
&= \frac{Nk}{N-k-2} - \frac{2N}{N-k-2}\mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{i=1}^{k} N_i(\bar{X}_i-\theta_{i,*})(\bar{X}_i-\hat{\theta}_i)\right] \\
&\quad + \frac{N\mathrm{E}[T]}{N-k-2} - \mathrm{E}\left[C^{*}\frac{NT}{\sigma_*^2}\frac{\sum_{i=1}^{k} N_i(\theta_{i,*}-\hat{\theta}_i)^2}{S^2}\right] \\
&= k + O(N^{-1}) - \frac{2N}{N-k-2}\mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{i=1}^{k} N_i(\bar{X}_i-\theta_{i,*})(\bar{X}_i-\hat{\theta}_i)\right] \\
&\quad + \mathrm{E}[T] + O(N^{-1})\mathrm{E}[T] - \mathrm{E}\left[C^{*}\frac{NT}{\sigma_*^2}\frac{\sum_{i=1}^{k} N_i(\theta_{i,*}-\hat{\theta}_i)^2}{S^2}\right], \qquad (2.11)
\end{aligned}
$$

を得る. よって, (2.10), (2.11) より

$$
\begin{aligned}
&\mathrm{E}\left[\frac{N\sigma_*^2}{\hat{\sigma}^2}\right] + \mathrm{E}\left[\frac{\sum_{i=1}^{k} N_i(\theta_{i,*}-\hat{\theta}_i)^2}{\hat{\sigma}^2}\right] \\
&= N + 2(k+1) - \frac{2N}{N-k-2}\mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{i=1}^{k} N_i(\bar{X}_i-\theta_{i,*})(\bar{X}_i-\hat{\theta}_i)\right] + J, \quad (2.12)
\end{aligned}
$$

が成り立つ. ただし, J は

$$
J = O(N^{-1}) + O(N^{-1})\mathrm{E}[T] + \mathrm{E}\left[C^{*}\frac{N^2T^2}{S^3}\right] - \mathrm{E}\left[C^{*}\frac{NT}{\sigma_*^2}\frac{\sum_{i=1}^{k} N_i(\theta_{i,*}-\hat{\theta}_i)^2}{S^2}\right],
$$

である. ところで, $\hat{\boldsymbol{\theta}}$ は $\|\bar{\boldsymbol{X}}-\boldsymbol{\theta}\|_{\boldsymbol{N}}$ を Θ 上で最小にする点であり, また, 仮定より $\boldsymbol{\theta}_* \in \Theta$ であったから, $\|\bar{\boldsymbol{X}}-\hat{\boldsymbol{\theta}}\|_{\boldsymbol{N}} \le \|\bar{\boldsymbol{X}}-\boldsymbol{\theta}_*\|_{\boldsymbol{N}}$ が成り立つ. 故に,

$$
\begin{aligned}
T = \frac{1}{\sigma_*^2}\sum_{i=1}^{k} N_i(\bar{X}_i-\hat{\theta}_i)^2 &= \frac{1}{\sigma_*^2}(\|\bar{\boldsymbol{X}}-\hat{\boldsymbol{\theta}}\|_{\boldsymbol{N}})^2 \le \frac{1}{\sigma_*^2}(\|\bar{\boldsymbol{X}}-\boldsymbol{\theta}_*\|_{\boldsymbol{N}})^2 \\
&= \frac{1}{\sigma_*^2}\sum_{i=1}^{k} N_i(\bar{X}_i-\theta_{i,*})^2 \equiv K, \ (\text{say}),
\end{aligned}
$$

であり, $K \sim \chi_k^2$ である. よって, $0 \le \mathrm{E}[T] \le \mathrm{E}[K] = k$ から, $\mathrm{E}[T] = O(1)$ を得る. また, $0 \le C^* \le 1$ に注意すれば,

$$
\begin{aligned}
\left|\mathrm{E}\left[C^{*}\frac{N^2T^2}{S^3}\right]\right| \le \mathrm{E}\left[\frac{N^2T^2}{S^3}\right] &= \frac{N^2\mathrm{E}[T^2]}{(N-k-2)(N-k-4)(N-k-6)} \\
&\le O(N^{-1})\mathrm{E}[K^2] = O(N^{-1})(2k+k^2) = O(N^{-1}),
\end{aligned}
$$

より,

$$\mathrm{E}\left[C^* \frac{N^2T^2}{S^3}\right] = O(N^{-1}),$$

を得る．最後に，三角不等式 $\|\boldsymbol{\theta}_* - \hat{\boldsymbol{\theta}}\|_{\boldsymbol{N}} \le \|\boldsymbol{\theta}_* - \bar{\boldsymbol{X}}\|_{\boldsymbol{N}} + \|\bar{\boldsymbol{X}} - \hat{\boldsymbol{\theta}}\|_{\boldsymbol{N}}$ と先に述べた $\|\bar{\boldsymbol{X}} - \hat{\boldsymbol{\theta}}\|_{\boldsymbol{N}} \le \|\bar{\boldsymbol{X}} - \boldsymbol{\theta}_*\|_{\boldsymbol{N}}$ から $\|\boldsymbol{\theta}_* - \hat{\boldsymbol{\theta}}\|_{\boldsymbol{N}} \le 2\|\boldsymbol{\theta}_* - \bar{\boldsymbol{X}}\|_{\boldsymbol{N}}$ が得られること，および，$0 \le C^* \le 1$ であることから，$T \le K$ に注意すれば

$$\begin{aligned}
&\left|\mathrm{E}\left[C^* \frac{NT}{\sigma_*^2} \frac{\sum_{i=1}^k N_i(\theta_{i,*} - \hat{\theta}_i)^2}{S^2}\right]\right| \\
&\le \mathrm{E}\left[\frac{NT}{\sigma_*^2} \frac{\sum_{i=1}^k N_i(\theta_{i,*} - \hat{\theta}_i)^2}{S^2}\right] \\
&\le \frac{N}{(N-k-2)(N-k-4)} \mathrm{E}\left[\frac{T}{\sigma_*^2}(\|\boldsymbol{\theta}_* - \hat{\boldsymbol{\theta}}\|_{\boldsymbol{N}})^2\right] \le O(N^{-1})\mathrm{E}[4K^2] = O(N^{-1}),
\end{aligned}$$

より

$$\mathrm{E}\left[C^* \frac{NT}{\sigma_*^2} \frac{\sum_{i=1}^k N_i(\theta_{i,*} - \hat{\theta}_i)^2}{S^2}\right] = O(N^{-1}),$$

を得る．従って，$J = O(N^{-1})$ であり，この結果と (2.12) より，(2.8) は

$$B = 2(k+1) - \frac{2N}{N-k-2}\mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{i=1}^k N_i(\bar{X}_i - \theta_{i,*})(\bar{X}_i - \hat{\theta}_i)\right] + O(N^{-1}), \quad (2.13)$$

と表される．従って，$O(N^{-1})$ のオーダーのバイアス補正を行うためには，(2.13) 中の期待値を求めさえすればよい．次の節では，いくつかの新しい記号の定義と期待値を求めるための lemma を与える．

3. 新しい記号と補題

本節ではまず，いくつかの新しい記号の定義を行う．その後，期待値を求めるための lemma を与える．

3.1. 新しい記号

いくつかの新しい記号を定義する．l を 2 以上の自然数とし，$n_1, \ldots, n_l$ を正数とする．また，$\boldsymbol{n} = (n_1, \ldots, n_l)'$ とする．l 次元ベクトル $\boldsymbol{x} = (x_1, \ldots, x_l)' \in \mathbb{R}^l$ と $1 \le i \le j \le l$ なる i, j に対し，$\boldsymbol{x}_{[i,j]}$ を $\boldsymbol{x}$ の第 i 成分から第 j 成分までを並べた $j-i+1$

次元ベクトルとする. すなわち, $\boldsymbol{x}_{[i,j]} = (x_i, \ldots, x_j)'$ であり, $\boldsymbol{x}_{[i,i]} = x_i$, $\boldsymbol{x}_{[1,l]} = \boldsymbol{x}$ である. 更に, $\tilde{x}_{[i,j]}$ および $\bar{x}^{(\boldsymbol{n})}_{[i,j]}$ をそれぞれ

$$\tilde{x}_{[i,j]} = \sum_{s=i}^{j} x_s, \quad \bar{x}^{(\boldsymbol{n})}_{[i,j]} = \frac{\sum_{s=i}^{j} n_s x_s}{\sum_{s=i}^{j} n_s} = \frac{\sum_{s=i}^{j} n_s x_s}{\tilde{n}_{[i,j]}} = \frac{\boldsymbol{n}'_{[i,j]}\boldsymbol{x}_{[i,j]}}{\tilde{n}_{[i,j]}},$$

で定義する. 簡略化のため, 以後特に断りがない限り, $\bar{x}^{(\boldsymbol{n})}_{[i,j]}$ を $\bar{x}_{[i,j]}$ とかくことにする. また, $\bar{x}_{[i,i]} = x_i$ であることに注意されたい.

次に, 集合 $\mathcal{A}^l$ を

$$\begin{aligned}\mathcal{A}^l &= \{(a_1, \ldots, a_l)' \in \mathbb{R}^l \mid a_1 \leq a_2 \leq \cdots \leq a_l\} \\ &= \{(a_1, \ldots, a_l)' \in \mathbb{R}^l \mid 1 \leq t \leq l-1,\ a_t \leq a_{t+1}\},\end{aligned}$$

で定義する. また, 集合 $\mathcal{A}^l_1$, $\mathcal{A}^l_l$ を

$$\mathcal{A}^l_1 = \{(x_1, \ldots, x_l)' \in \mathbb{R}^l \mid x_1 = x_2 = \cdots = x_l\},$$

および

$$\begin{aligned}\mathcal{A}^l_l &= \{(x_1, \ldots, x_l)' \in \mathbb{R}^l \mid x_1 < x_2 < \cdots < x_l\} \\ &= \{(x_1, \ldots, x_l)' \in \mathbb{R}^l \mid 1 \leq t \leq l-1,\ x_t < x_{t+1}\},\end{aligned}$$

で定義する. 後のために, $\mathcal{A}^1 = \mathbb{R}^1$ と定義する. 更に, $1 \leq i \leq l$ なる自然数 i に対し, 集合 $\mathcal{W}^l_i$ を

$$\mathcal{W}^l_i = \{(w_1, \ldots, w_i)' \in \mathbb{N}^i \mid 1 \leq t \leq i,\ w_{t-1} < w_t,\ w_0 = 0,\ w_i = l\},$$

と定義する. よって, 例えば $l = 2$ の場合

$$\mathcal{W}^2_1 = \{(2)'\}, \quad \mathcal{W}^2_2 = \{(1,2)'\},$$

であり, $l = 3$ の場合

$$\mathcal{W}^3_1 = \{(3)'\}, \quad \mathcal{W}^3_2 = \{(1,3)',\ (2,3)'\}, \quad \mathcal{W}^3_3 = \{(1,2,3)'\},$$

であり, $l = 4$ の場合

$$\begin{aligned}&\mathcal{W}^4_1 = \{(4)'\}, \quad \mathcal{W}^4_2 = \{(1,4)',\ (2,4)',\ (3,4)'\}, \\ &\mathcal{W}^4_3 = \{(1,2,4)',\ (1,3,4)',\ (2,3,4)'\}, \quad \mathcal{W}^4_4 = \{(1,2,3,4)'\},\end{aligned}$$

となる. ここで, 集合 $\mathcal{W}^l_i$ の要素数は ${}_{l-1}C_{i-1}$ であることに注意されたい. また, 集合 $\mathcal{W}^l_i$ の元, $\boldsymbol{w} = (w_1, \ldots, w_i)'$ は i 次元ベクトルであり, その第 i 成分 w_i は常に $w_i = l$

であることにも合わせて注意されたい. 定義より, $\mathcal{W}_1^l$ の元 $\boldsymbol{w}$ は $\boldsymbol{w} = (l)'$ ただひとつであり, $\mathcal{W}_l^l$ の元 $\boldsymbol{w}$ も $\boldsymbol{w} = (1, \ldots, l)'$ ただひとつである.

次に, 任意の i, $(i = 1, \ldots, l)$ と, 任意の $\boldsymbol{w} \in \mathcal{W}_i^l$ に対し, 集合 $\mathcal{A}_i^l(\boldsymbol{w})$ を以下のように定める. まず, $i = 1$ の場合, $\mathcal{W}_1^l$ の元 $\boldsymbol{w}$ は $\boldsymbol{w} = (l)'$ のみであり, $\mathcal{A}_1^l(\boldsymbol{w})$ を

$$\mathcal{A}_1^l(\boldsymbol{w}) = \{(x_1, \ldots, x_l)' \in \mathbb{R}^l \mid x_1 = x_2 = \cdots = x_l\} = \mathcal{A}_1^l,$$

と定める. $2 \leq i \leq l$ の場合, $\mathcal{W}_i^l$ の任意の元 $\boldsymbol{w} = (w_1, \ldots, w_i)'$ に対して,

$$\begin{aligned}
\mathcal{A}_i^l(\boldsymbol{w}) = \{(a_1, \ldots, a_l)' \in \mathcal{A}^l \mid & \ 1 \leq t \leq i-1, \ a_{w_t} < a_{w_{t+1}}, \\
& 0 \leq s \leq i-1, \ w_0 = 0, \ a_{1+w_s} = a_{w_{s+1}}\}, \\
= \{(x_1, \ldots, x_l)' \in \mathbb{R}^l \mid & \ 1 \leq t \leq i-1, \ x_{w_t} < x_{w_{t+1}}, \\
& 0 \leq s \leq i-1, \ w_0 = 0, \ x_{1+w_s} = \cdots = x_{w_{s+1}}\},
\end{aligned} \tag{3.1}$$

とする. すなわち, (3.1) において, $\mathcal{A}_i^l(\boldsymbol{w})$ の元 $\boldsymbol{x} = (x_1, \ldots, x_l)'$ は

$$x_1 = \cdots = x_{w_1} < x_{1+w_1} = \cdots = x_{w_2} < \cdots < x_{1+w_{i-1}} = \cdots = x_l, \tag{3.2}$$

を満足し, $\mathcal{A}_i^l(\boldsymbol{w})$ は (3.2) を満たす $\mathbb{R}^l$ の元すべてをあつめた集合であることに注意されたい. ここで, 特に $i = l$ の場合, $\mathcal{W}_l^l$ の元 $\boldsymbol{w}$ は $\boldsymbol{w} = (w_1, \ldots, w_l)' = (1, \ldots, l)'$ ただひとつであったから,

$$\mathcal{A}_l^l(\boldsymbol{w}) = \{(x_1, \ldots, x_l)' \in \mathbb{R}^l \mid x_1 < x_2 < \cdots < x_l\} = \mathcal{A}_l^l,$$

である. ここで, いくつかの具体例を紹介する. $l = 2$ の場合

$$\mathcal{A}_1^2(\boldsymbol{w}) = \mathcal{A}_1^2 = \{\boldsymbol{x} \in \mathbb{R}^2 \mid x_1 = x_2\}, \quad \mathcal{A}_2^2(\boldsymbol{w}) = \mathcal{A}_2^2 = \{\boldsymbol{x} \in \mathbb{R}^2 \mid x_1 < x_2\},$$

であり, $l = 3$ の場合

$$\begin{aligned}
&\mathcal{A}_1^3(\boldsymbol{w}) = \mathcal{A}_1^3 = \{\boldsymbol{x} \in \mathbb{R}^3 \mid x_1 = x_2 = x_3\}, \\
&\mathcal{A}_2^3(\boldsymbol{w}) = \{\boldsymbol{x} \in \mathbb{R}^3 \mid x_1 < x_2 = x_3\}, \quad (\text{if } \boldsymbol{w} = (1,3)' \in \mathcal{W}_2^3) \\
&\mathcal{A}_2^3(\boldsymbol{w}) = \{\boldsymbol{x} \in \mathbb{R}^3 \mid x_1 = x_2 < x_3\}, \quad (\text{if } \boldsymbol{w} = (2,3)' \in \mathcal{W}_2^3) \\
&\mathcal{A}_3^3(\boldsymbol{w}) = \mathcal{A}_3^3 = \{\boldsymbol{x} \in \mathbb{R}^3 \mid x_1 < x_2 < x_3\},
\end{aligned}$$

となる. ここで, $\mathcal{A}^l$, $\mathcal{A}_i^l(\boldsymbol{w})$ に対して,

$$\mathcal{A}^l = \bigcup_{i=1}^{l} \bigcup_{\boldsymbol{w}; \boldsymbol{w} \in \mathcal{W}_i^l} \mathcal{A}_i^l(\boldsymbol{w}),$$

および

$$(i, \boldsymbol{w}) \neq (i^*, \boldsymbol{w}^*) \Rightarrow \mathcal{A}_i^l(\boldsymbol{w}) \cap \mathcal{A}_{i^*}^l(\boldsymbol{w}^*) = \emptyset,$$

であることに注意されたい.

次に, $1 \leq i \leq j \leq l$ なる i, j に対し, 行列 $\boldsymbol{D}_{i,j}^{(\boldsymbol{n})}$ を定義する. $i = j$ の場合, $\boldsymbol{D}_{i,j}^{(\boldsymbol{n})}$ は 1 行 1 列の行列であり, $\boldsymbol{D}_{i,j}^{(\boldsymbol{n})} = 0$ と定義する. また, $i < j$ の場合, $\boldsymbol{D}_{i,j}^{(\boldsymbol{n})}$ は $j - i$ 行 $j - i + 1$ 列の行列であり, $\boldsymbol{D}_{i,j}^{(\boldsymbol{n})}$ の第 s 行 $(1 \leq s \leq j - i)$ を

$$\left(\frac{1}{\tilde{n}_{[i,i+s-1]}} \boldsymbol{n}'_{[i,i+s-1]}, \frac{-1}{\tilde{n}_{[i+s,j]}} \boldsymbol{n}'_{[i+s,j]} \right),$$

と定義する. よって, 例えば $l = 4$ の場合

$$\begin{aligned}
&\boldsymbol{D}_{1,1}^{(\boldsymbol{n})} = \boldsymbol{D}_{2,2}^{(\boldsymbol{n})} = \boldsymbol{D}_{3,3}^{(\boldsymbol{n})} = \boldsymbol{D}_{4,4}^{(\boldsymbol{n})} = 0, \\
&\boldsymbol{D}_{1,2}^{(\boldsymbol{n})} = \boldsymbol{D}_{2,3}^{(\boldsymbol{n})} = \boldsymbol{D}_{3,4}^{(\boldsymbol{n})} = (1 \ \ -1), \\
&\boldsymbol{D}_{1,3}^{(\boldsymbol{n})} = \begin{pmatrix} 1 & \frac{-n_2}{n_2+n_3} & \frac{-n_3}{n_2+n_3} \\ \frac{n_1}{n_1+n_2} & \frac{n_2}{n_1+n_2} & -1 \end{pmatrix}, \quad \boldsymbol{D}_{2,4}^{(\boldsymbol{n})} = \begin{pmatrix} 1 & \frac{-n_3}{n_3+n_4} & \frac{-n_4}{n_3+n_4} \\ \frac{n_2}{n_2+n_3} & \frac{n_3}{n_2+n_3} & -1 \end{pmatrix}, \\
&\boldsymbol{D}_{1,4}^{(\boldsymbol{n})} = \begin{pmatrix} 1 & \frac{-n_2}{n_2+n_3+n_4} & \frac{-n_3}{n_2+n_3+n_4} & \frac{-n_4}{n_2+n_3+n_4} \\ \frac{n_1}{n_1+n_2} & \frac{n_2}{n_1+n_2} & \frac{-n_3}{n_3+n_4} & \frac{-n_4}{n_3+n_4} \\ \frac{n_1}{n_1+n_2+n_3} & \frac{n_2}{n_1+n_2+n_3} & \frac{n_3}{n_1+n_2+n_3} & -1 \end{pmatrix},
\end{aligned}$$

となる. 簡略化のため, 以後特に断りがない限り $\boldsymbol{D}_{i,j}^{(\boldsymbol{n})}$ を $\boldsymbol{D}_{i,j}$ と書くことにする.

最後に, 写像をひとつ定義しておく. $\boldsymbol{\eta}_l^{(\boldsymbol{n})}$ を $\mathbb{R}^l$ から $\mathcal{A}^l$ への写像とし, $\boldsymbol{x} = (x_1, \ldots, x_l)' \in \mathbb{R}^l$ に対して

$$\boldsymbol{\eta}_l^{(\boldsymbol{n})}(\boldsymbol{x}) = \underset{\boldsymbol{y} \in \mathcal{A}^l}{\operatorname{argmin}} \|\boldsymbol{x} - \boldsymbol{y}\|_{\boldsymbol{n}}^2,$$

で定義する. 簡略化のため, 以後特に断りがない限り $\boldsymbol{\eta}_l^{(\boldsymbol{n})}$ を $\boldsymbol{\eta}_l$ でかくことにする. ここで, $(\mathbb{R}^l, \| \ \|_{\boldsymbol{n}})$ はヒルベルト空間であり, $\mathcal{A}^l$ は空でない閉凸集合であるから, 任意の $\boldsymbol{x} \in \mathbb{R}^l$ に対して $\boldsymbol{\eta}_l(\boldsymbol{x})$ の一意存在性は保証されていることに注意されたい (see, e.g., Rudin, 1986). $\boldsymbol{\eta}_l(\boldsymbol{x})$ は l 次元ベクトルであり, $\boldsymbol{\eta}_l(\boldsymbol{x})$ の第 s 成分 $(1 \leq s \leq l)$ を $\eta_l(\boldsymbol{x})[s]$ と表すことにする. このとき, Robertson *et al.* (1988) より, $\eta_l(\boldsymbol{x})[s]$ は

$$\eta_l(\boldsymbol{x})[s] = \min_{v; v \geq s} \max_{u; u \leq s} \frac{\sum_{j=u}^{v} n_j x_j}{\sum_{j=u}^{v} n_j} = \min_{v; v \geq s} \max_{u; u \leq s} \bar{x}_{[u,v]},$$

と表されることに注意されたい. また, 写像 $\boldsymbol{\eta}_1(\boldsymbol{x}) = \boldsymbol{x}$ と定義しておく.

3.2. Main lemma

次の補題が成立する.

Lemma 3.1. k を 2 以上の任意の自然数とする. $n_1, \ldots, n_k$ を任意の正数とし, $\boldsymbol{n} = (n_1, \ldots, n_k)'$ とする. $\xi_1, \ldots, \xi_k$ を任意の実数, τ^2 を任意の正の数とし, $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_k)'$ とする. $x_1, \ldots, x_k$ は互いに独立な確率変数とし, $\boldsymbol{x} = (x_1, \ldots, x_k)'$, $x_i \sim N(\xi_i, \tau^2/n_i)$, $(i = 1, \ldots, k)$ とする. このとき, 以下が成立する.

$$\mathrm{E}\left[\frac{1}{\tau^2}\sum_{i=1}^{k} n_i(x_i - \xi_i)(x_i - \eta_k^{(\boldsymbol{n})}(\boldsymbol{x})[i])\right]$$
$$= \sum_{i=1}^{k-1}(k-i)\mathrm{P}\left(\boldsymbol{\eta}_k^{(\boldsymbol{n})}(\boldsymbol{x}) \in \bigcup_{\boldsymbol{w};\boldsymbol{w}\in\mathcal{W}_i^k} \mathcal{A}_i^k(\boldsymbol{w})\right).$$

4. 順序制約下での ANOVA モデルに対する AIC 規準とその性質

本節では AIC 規準の導出および関連する定理の紹介を行う.

4.1. AIC 規準の導出

本小節では, SO の下でのモデル (2.1) における AIC 規準を導出する. まず, (2.13) 中の期待値を求める. (2.3) より, $\bar{X}_1, \ldots, \bar{X}_k$ は互いに独立な確率変数であり, 各 i, $1 \le i \le k$ に対して $\bar{X}_i \sim N(\theta_{i,*}, \sigma_*^2/N_i)$ である. また, (2.5) より, MLE $\hat{\boldsymbol{\theta}}$ は $\hat{\boldsymbol{\theta}} = \boldsymbol{\eta}_k^{(\boldsymbol{N})}(\bar{\boldsymbol{X}})$ と表すことができる. 故に, Lemma 3.1 より, (2.13) 中の期待値は

$$\mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{i=1}^{k} N_i(\bar{X}_i - \theta_{i,*})(\bar{X}_i - \hat{\theta}_i)\right] = \mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{i=1}^{k} N_i(\bar{X}_i - \theta_{i,*})(\bar{X}_i - \eta_k^{(\boldsymbol{N})}(\bar{\boldsymbol{X}})[i])\right]$$
$$= \sum_{i=1}^{k-1}(k-i)\mathrm{P}\left(\hat{\boldsymbol{\theta}} \in \bigcup_{\boldsymbol{w};\boldsymbol{w}\in\mathcal{W}_i^k} \mathcal{A}_i^k(\boldsymbol{w})\right) \equiv Q, \text{ (say)}$$

で与えられる. 明らかに $Q = O(1)$ であることに注意すれば, Q を (2.13) に代入することにより

$$B = 2(k+1) - \frac{2N}{N-k-2}Q + O(N^{-1}) = 2(k+1) - 2Q + O(N^{-1}), \quad (4.1)$$

を得る. 従って, バイアス補正を行うには $2(k+1)-2Q$ を $-2l(\hat{\boldsymbol{\theta}},\hat{\sigma}^2;\boldsymbol{X})$ に足せばよい. しかしながら, 容易に確かめられることであるが Q は真値 $\theta_{1,*},\ldots,\theta_{k,*}$ および σ_*^2 に依存している. 故に, Q を推定する必要がある. ここで, 集合 $\hat{\mathcal{M}}$ を

$$\hat{\mathcal{M}}=\bigcup_{i=1}^{k}\{\hat{\theta}_i\},$$

と定義する. また, 確率変数 $\hat{m}$ を

$$\hat{m}=\#\hat{\mathcal{M}}, \tag{4.2}$$

で定義する. 定義より, $\hat{m}$ は 1 から k までの自然数を値にとる離散型確率変数であることに注意されたい. 例えば, $\hat{\theta}_1=\cdots=\hat{\theta}_k$ ならば $\hat{m}=1$ であり, $\hat{\theta}_1<\hat{\theta}_2<\cdots<\hat{\theta}_k$ ならば $\hat{m}=k$ であり, $\hat{\theta}_1<\hat{\theta}_2=\cdots=\hat{\theta}_k$ ならば $\hat{m}=2$ である. ここで, $\hat{m}$ の定義と $\mathcal{A}_i^k(\boldsymbol{w})$ の定義より, 明らかに

$$\hat{\boldsymbol{\theta}}\in\bigcup_{\boldsymbol{w};\boldsymbol{w}\in\mathcal{W}_i^k}\mathcal{A}_i^k(\boldsymbol{w})\Leftrightarrow\hat{m}=i,$$

が成立する. 従って, 確率変数 $k-\hat{m}$ に対して,

$$\mathrm{E}[k-\hat{m}]=\sum_{i=1}^{k}(k-i)\mathrm{P}(\hat{m}=i)=\sum_{i=1}^{k-1}(k-i)\mathrm{P}\left(\hat{\boldsymbol{\theta}}\in\bigcup_{\boldsymbol{w};\boldsymbol{w}\in\mathcal{W}_i^k}\mathcal{A}_i^k(\boldsymbol{w})\right)=Q,$$

であるから, $k-\hat{m}$ は Q の不偏推定量である. 従って, (4.1) より

$$\mathrm{E}[2(\hat{m}+1)]=\mathrm{E}[2(k+1)-2(k-\hat{m})]=2(k+1)-2Q=B+O(N^{-1}),$$

であるから, N^{-1} のオーダーのバイアス補正を行うためには, $2(k+1)-2Q$ のかわりに $2(\hat{m}+1)$ を $-2l(\hat{\boldsymbol{\theta}},\hat{\sigma}^2;\boldsymbol{X})$ に足せばよい. これにより, SO の下での ANOVA モデルにおける AIC 規準, $\mathrm{AIC_{SO}}$ を得る.

Theorem 4.1. $l(\hat{\boldsymbol{\theta}},\hat{\sigma}^2;\boldsymbol{X})$ を (2.7) で与えられる最大対数尤度とし, $\hat{m}$ は (4.2) で与えられるものとする. このとき, SO の下での ANOVA モデルにおける AIC 規準, $\mathrm{AIC_{SO}}$ は

$$\mathrm{AIC_{SO}}:=-2l(\hat{\boldsymbol{\theta}},\hat{\sigma}^2;\boldsymbol{X})+2(\hat{m}+1),$$

で与えられる. また, (2.6) で与えられるリスク R に対して,

$$\mathrm{E}[\mathrm{AIC_{SO}}]=R+O(N^{-1}),$$

が成立する.

Remark 4.1. $\mathrm{AIC}_{\mathrm{SO}}$ は順序制約 (2.2) の下で与えられたが, (2.2) 中の “ $\leq$ ”のいくつかを “ $=$ ”に置き換えても $\mathrm{AIC}_{\mathrm{SO}}$ が導出できる. 実際, 例えば $k=4$ のときモデル (2.1) に

$$\theta_1 = \theta_2 \leq \theta_3 = \theta_4, \tag{4.3}$$

というような制約を課すこともできる. この場合, $N_1^* = N_1 + N_2$, $N_2^* = N_3 + N_4$, $\theta_1 = \theta_2 = \mu_1$, $\theta_3 = \theta_4 = \mu_2$ とおき,

$$X_{11}, \ldots, X_{1N_1}, X_{21}, \ldots, X_{2N_2} \to Y_{11}, \ldots, Y_{1N_1^*},$$
$$X_{31}, \ldots, X_{3N_3}, X_{41}, \ldots, X_{4N_4} \to Y_{21}, \ldots, Y_{2N_2^*},$$

と記号を置き換えれば, (4.3) の下でのモデル (2.1) は, $\mu_1 \leq \mu_2$ の下での ANOVA モデル

$$Y_{ij} \sim N(\mu_i, \sigma^2), \quad (i = 1, 2,\ j = 1, \ldots, N_i^*)$$

を考えることと等しい. あとは, このモデルに対して同様の手順で $\mathrm{AIC}_{\mathrm{SO}}$ を導けばよい.

Remark 4.2. $\mathrm{AIC}_{\mathrm{SO}}$ はリスクに対する漸近不偏な推定量であり, 誤差のオーダーは N^{-1} である. 一方, 制約なしの ANOVA において, 通常の AIC もまたリスク $\bar{R}$ に対する漸近不偏な推定量であり, その誤差のオーダーは N^{-1} である. 故に, リスクに対する推定の精度という観点から見たとき, $\mathrm{AIC}_{\mathrm{SO}}$ は AIC と同等に優れている. また, $\mathrm{AIC}_{\mathrm{SO}}$ の罰則項は $2(\hat{m}+1)$ であり, $\hat{m}$ は MLE の実現値のうちで異なっているものの数をカウントするだけでよかった. $\mathrm{AIC}_{\mathrm{SO}}$ を計算する段階ではすでに MLE の計算は終わっているのだから, そこから $\hat{m}$ を計算するのに要するコストは無いに等しい. 故に, $\mathrm{AIC}_{\mathrm{SO}}$ は扱いやすさの面からみても AIC と同等に優れている.

次の小節では, 特別な場合における $\mathrm{AIC}_{\mathrm{SO}}$ をひとつ紹介する.

4.2. 真の分散 σ_*^2 が既知の場合

ANOVA モデル (2.1) において, 分散パラメータ σ^2 の真値 σ_*^2 が既知であるとする. このとき, SO の下での $\theta_1, \ldots, \theta_k$ の MLE, $\hat{\theta}_1, \ldots, \hat{\theta}_k$ は (2.5) で与えられる. また,

K-L ダイバージェンスに基づくリスク, R_1 は (2.6) 中の $\hat{\sigma}^2$ を σ_*^2 に置き換えれば

$$\begin{aligned} R_1 &= \mathrm{E}[\mathrm{E}_\star[-2l(\hat{\boldsymbol{\theta}}, \sigma_*^2; \boldsymbol{X}^\star)]] \\ &= \mathrm{E}\left[N\log(2\pi\sigma_*^2) + N + \frac{\sum_{i=1}^k N_i(\theta_{i,*} - \hat{\theta}_i)^2}{\sigma_*^2}\right] \\ &= N\log(2\pi\sigma_*^2) + N + \mathrm{E}\left[\frac{\sum_{i=1}^k N_i(\theta_{i,*} - \bar{X}_i + \bar{X}_i - \hat{\theta}_i)^2}{\sigma_*^2}\right] \\ &= N\log(2\pi\sigma_*^2) + N + k - 2Q + \mathrm{E}\left[\frac{\sum_{i=1}^k N_i(\bar{X}_i - \hat{\theta}_i)^2}{\sigma_*^2}\right], \end{aligned} \tag{4.4}$$

となる. SO の制約が無い ANOVA においては, σ_*^2 が既知の時, リスク $\bar{R}_1$ は $\bar{R}_1 = \mathrm{E}[\mathrm{E}_\star[-2l(\bar{\boldsymbol{X}}, \sigma_*^2; \boldsymbol{X}^\star)]]$ である. 一方, 最大対数尤度 $l(\hat{\boldsymbol{\theta}}, \sigma_*^2; \boldsymbol{X})$ は (2.4) より

$$\begin{aligned} & l(\hat{\boldsymbol{\theta}}, \sigma_*^2; \boldsymbol{X}) \\ &= -\frac{N}{2}\log(2\pi\sigma_*^2) - \frac{1}{2\sigma_*^2}\sum_{i=1}^k\sum_{j=1}^{N_i}(X_{ij} - \bar{X}_i)^2 - \frac{1}{2\sigma_*^2}\sum_{i=1}^k N_i(\bar{X}_i - \hat{\theta}_i)^2, \end{aligned} \tag{4.5}$$

であるから, リスク, R_1 を $-2l(\hat{\boldsymbol{\theta}}, \sigma_*^2; \boldsymbol{X})$ で推定した際のバイアス, B_1 は

$$\begin{aligned} B_1 = \mathrm{E}[R_1 - \{-2l(\hat{\boldsymbol{\theta}}, \sigma_*^2; \boldsymbol{X})\}] &= N + k - 2Q - \mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{i=1}^k\sum_{j=1}^{N_i}(X_{ij} - \bar{X}_i)^2\right] \\ &= N + k - 2Q - (N - k) = 2k - 2Q, \end{aligned}$$

で与えられる. (4.2) で定義した $\hat{m}$ に関して, $\mathrm{E}[k - \hat{m}] = Q$ に注意すれば次の corollary を得る.

Corollary 4.1. 順序制約 SO を課した ANOVA モデル (2.1) における σ^2 の真値, σ_*^2 が既知とする. $l(\hat{\boldsymbol{\theta}}, \sigma_*^2; \boldsymbol{X})$ を (4.5) で与えられる最大対数尤度とし, $\hat{m}$ は (4.2) で与えられるものとする. このとき, $\mathrm{AIC}_{\mathrm{SO}}$ は

$$\mathrm{AIC}_{\mathrm{SO}} = -2l(\hat{\boldsymbol{\theta}}, \sigma_*^2; \boldsymbol{X}) + 2\hat{m},$$

と表される. また, (4.4) で与えられるリスク R_1 に対して,

$$\mathrm{E}[\mathrm{AIC}_{\mathrm{SO}}] = R_1,$$

が成立する.

Remark 4.3. 分散 σ_*^2 が既知の場合, $\mathrm{AIC_{SO}}$ はリスク R_1 に対する不偏推定量である. また, 制約が無い ANOVA モデルにおいて, 分散が既知の場合, 通常の AIC もリスク $\bar{R}_1$ の不偏推定量である.

モデル (2.1) において, 分散の真値 σ_*^2 が未知の場合と既知の場合における $\mathrm{AIC_{SO}}$ と AIC の関係を Table 4.1 にまとめる.

Table 4.1 $\mathrm{AIC_{SO}}$ と AIC の関係

		$\mathrm{AIC_{SO}}$	AIC
分散未知	制約	SO	Non
	リスク	$\mathrm{E}[\mathrm{E}_\star[-2l(\hat{\boldsymbol{\theta}},\hat{\sigma}^2;\boldsymbol{X}^\star)]]$	$\mathrm{E}[\mathrm{E}_\star[-2l(\bar{\boldsymbol{X}},\bar{\sigma}^2;\boldsymbol{X}^\star)]]$
	罰則項	$2(\hat{m}+1)$	$2(k+1)$
	リスクに対する不偏性	漸近不偏	漸近不偏
	バイアスのオーダー	$O(N^{-1})$	$O(N^{-1})$
分散既知	制約	SO	Non
	リスク	$\mathrm{E}[\mathrm{E}_\star[-2l(\hat{\boldsymbol{\theta}},\sigma_*^2;\boldsymbol{X}^\star)]]$	$\mathrm{E}[\mathrm{E}_\star[-2l(\bar{\boldsymbol{X}},\sigma_*^2;\boldsymbol{X}^\star)]]$
	罰則項	$2\hat{m}$	$2k$
	リスクに対する不偏性	不偏	不偏
	バイアスのオーダー	0	0

Note: $\hat{m}$ は (4.2) で与えられる.

4.3. モデル選択規準としての $\mathbf{AIC_{SO}}$ と AIC の振る舞い

本小節では, SO が課された特定のモデルを比較する場合における, $\mathrm{AIC_{SO}}$ と AIC のモデル選択規準としての振る舞いに関する定理をひとつ与える. 特定のモデルについて説明する前に, まず例として 3 クラスター ANOVA モデル, $X_{ij} \sim N(\theta_i,\sigma^2)$, $i=1,2,3,\ j=1,\ldots,N_i$ を考える. このモデルにおいて, 母平均 θ_1, θ_2, θ_3 への SO の入れ方 (課し方) には次の 4 通りが考えられる:

$$\text{CASE 1}:\ \theta_1=\theta_2=\theta_3,\quad \text{CASE 2}:\ \theta_1=\theta_2\le\theta_3,$$
$$\text{CASE 3}:\ \theta_1\le\theta_2=\theta_3,\quad \text{CASE 4}:\ \theta_1\le\theta_2\le\theta_3.$$

言い換えれば, 異なる SO の入れ方をした 4 個のモデルを考えることができる. 無論, $k\ (\ge 2)$ クラスター ANOVA モデルにおいては, 母平均への SO の入れ方は 2^{k-1} 通

りあり, 2^{k-1} 個のモデルを考えることができる. このようなモデルを $\mathrm{AIC_{SO}}$ および AIC を用いて選択する際, 以下の定理が成立する.

Theorem 4.2. $k\ (\geq 2)$ クラスター ANOVA モデルにおいて, 母平均に異なる SO の入れ方をした 2^{k-1} 個のモデルを $\mathrm{AIC_{SO}}$ および AIC を用いて選択した場合, $\mathrm{AIC_{SO}}$ 最小化によって選ばれるモデルと AIC 最小化によって選ばれるモデルは常に一致する.

Theorem 4.2 は, SO の入れ方に関するフルサーチを行う場合, リスクに対して漸近不偏ではない AIC を用いたとしても漸近不偏性を持った $\mathrm{AIC_{SO}}$ を用いた場合とかわらないことを示している. これは AIC を通常通り用いて良いということを示唆している.

Remark 4.4. Theorem 4.2 は AIC を用いても良いと述べているだけであり, もちろん, $\mathrm{AIC_{SO}}$ を用いても良い. また, Theorem 4.2 の結果はあくまで SO の入れ方に関してフルサーチした場合に成立するものであり, 例えば, ネストされた SO の入れ方に関してモデル比較を行う場合, $\mathrm{AIC_{SO}}$ 最小化によって選ばれるモデルと AIC 最小化によって選ばれるモデルは異なることもある.

5. 数値実験

本節では, 4 クラスター ANOVA モデル $X_{ij} \sim N(\theta_i, \sigma^2/N_i)$, (ただし, $1 \leq i \leq 4$, $1 \leq j \leq N_i$, $N_1 = \cdots = N_4$) において, 次の 4 つのモデル,

$$\begin{aligned}
&\text{Model 1}: \ \text{ANOVA モデル} \ \ \text{with} \ \ \theta_1 = \theta_2 = \theta_3 = \theta_4,\\
&\text{Model 2}: \ \text{ANOVA モデル} \ \ \text{with} \ \ \theta_1 \leq \theta_2 = \theta_3 = \theta_4,\\
&\text{Model 3}: \ \text{ANOVA モデル} \ \ \text{with} \ \ \theta_1 \leq \theta_2 \leq \theta_3 = \theta_4,\\
&\text{Model 4}: \ \text{ANOVA モデル} \ \ \text{with} \ \ \theta_1 \leq \theta_2 \leq \theta_3 \leq \theta_4,
\end{aligned}$$

に対してモデル選択を行った際の $\mathrm{AIC_{SO}}$ と AIC の性能を 100000 回のモンテカルロシミュレーションにより比較する. 第 q $(1 \leq q \leq 100000)$ 回目のシミュレーションに

おいて, 次の記号を定義する:

$$\hat{\theta}^{(q)}_{1,\mathrm{AIC_{SO}}},\dots,\hat{\theta}^{(q)}_{4,\mathrm{AIC_{SO}}},\ \hat{\sigma^2}^{(q)}_{\mathrm{AIC_{SO}}}:\ \mathrm{AIC_{SO}}\ \text{最小化によって選ばれたモデルにおける}\ \theta_1,\dots,\theta_4,\ \sigma^2\ \text{の MLE},$$

$$\hat{\theta}^{(q)}_{1,\mathrm{AIC}},\dots,\hat{\theta}^{(q)}_{4,\mathrm{AIC}},\ \hat{\sigma^2}^{(q)}_{\mathrm{AIC}}:\ \mathrm{AIC}\ \text{最小化によって選ばれたモデルにおける}\ \theta_1,\dots,\theta_4,\ \sigma^2\ \text{の MLE}.$$

さて, SO を課した ANOVA モデルにおけるリスクは (2.6) で与えられていたのだから,

$$R(\hat{\theta}_1,\dots,\hat{\theta}_4,\hat{\sigma}^2)=N\log(2\pi\hat{\sigma}^2)+\frac{N\sigma_*^2}{\hat{\sigma}^2}+\frac{\sum_{i=1}^4 N_i(\theta_{i,*}-\hat{\theta}_i)^2}{\hat{\sigma}^2},$$

はリスクに対する不偏推定量である. これを踏まえ, $\mathrm{AIC_{SO}}$ および AIC の性能を

$$\mathrm{PE}_{\mathrm{AIC_{SO}}}=\frac{1}{100000}\sum_{q=1}^{100000}R(\hat{\theta}^{(q)}_{1,\mathrm{AIC_{SO}}},\dots,\hat{\theta}^{(q)}_{4,\mathrm{AIC_{SO}}},\hat{\sigma^2}^{(q)}_{\mathrm{AIC_{SO}}}),$$

$$\mathrm{PE}_{\mathrm{AIC}}=\frac{1}{100000}\sum_{q=1}^{100000}R(\hat{\theta}^{(q)}_{1,\mathrm{AIC}},\dots,\hat{\theta}^{(q)}_{4,\mathrm{AIC}},\hat{\sigma^2}^{(q)}_{\mathrm{AIC}}),$$

を用いて測ることとする. すなわち, $\mathrm{PE}_{\mathrm{AIC_{SO}}}$ は $\mathrm{AIC_{SO}}$ 最小化によって選ばれたモデルのリスク (の推定値) であり, $\mathrm{PE}_{\mathrm{AIC}}$ は AIC 最小化によって選ばれたモデルのリスク (の推定値) である. 無論, より小さいほうが望ましい.

次に, 本シミュレーションでは真のモデルとして次の 3 通りの場合を考える:

$$\text{Case 1}:\ \theta_1=\theta_2=2,\ \theta_3=\theta_4=2.8,\ \sigma^2=2,$$
$$\text{Case 2}:\ \theta_1=1.5,\ \theta_2=1.8,\ \theta_3=2.1,\ \theta_4=2.4,\ \sigma^2=2,$$
$$\text{Case 3}:\ \theta_1=\theta_2=\theta_3=\theta_4=2.5,\ \sigma^2=2.$$

Case 1 においては, 真のモデルを含んでいるものは Model 3, 4 であり, Case 2 においては, 真のモデルを含んでいるものは Model 4 であり, Case 3 においては, 真のモデルを含んでいるものは Model 1, 2, 3, 4 である. また, 各 Case において, 真のモデルを含む最小モデルはそれぞれ Model 3, Model 4, Model 1 である. これらの設定の下, $N=N_1+\cdots+N_4$ が 40 の場合と 200 の場合で数値実験を行った. その結果を Table 5.1 – 5.3 にまとめた.

Table 5.1　Case 1 における AIC_{SO} と AIC の性能

N		Model 1	Model 2	Model 3	Model 4	$PE_{AIC_{SO}}$	PE_{AIC}
40	リスク	146.51	146.49	145.30	145.84	**146.43**	146.73
	AIC_{SO}	146.37	146.10	144.47	144.81		
	AIC	146.37	146.40	145.64	147.13		
200	リスク	723.54	719.53	709.82	710.34	**710.31**	710.42
	AIC_{SO}	723.69	719.66	709.69	710.18		
	AIC	723.69	719.69	710.69	712.18		

Table 5.2　Case 2 における AIC_{SO} と AIC の性能

N		Model 1	Model 2	Model 3	Model 4	$PE_{AIC_{SO}}$	PE_{AIC}
40	リスク	145.61	145.37	145.39	145.76	**146.34**	146.42
	AIC_{SO}	145.49	144.92	144.60	144.63		
	AIC	145.49	145.16	145.69	146.76		
200	リスク	719.14	713.68	711.20	710.75	**711.85**	712.04
	AIC_{SO}	719.18	713.62	710.97	710.42		
	AIC	719.18	713.63	711.33	711.30		

Table 5.3　Case 3 における AIC_{SO} と AIC の性能

N		Model 1	Model 2	Model 3	Model 4	$PE_{AIC_{SO}}$	PE_{AIC}
40	リスク	143.55	144.20	144.62	144.99	144.40	**144.16**
	AIC_{SO}	143.26	143.73	144.01	144.27		
	AIC	143.26	144.72	146.39	148.09		
200	リスク	708.26	708.76	709.08	709.37	708.86	**708.67**
	AIC_{SO}	708.26	708.75	709.05	709.33		
	AIC	708.26	709.74	711.44	713.15		

Table 5.1–5.3 より, $\mathrm{AIC_{SO}}$ はリスクに対する漸近不偏な推定量であることが確認できる. また, $\mathrm{AIC_{SO}}$ の定義と AIC の定義より, 常に $\mathrm{AIC_{SO}} \leq \mathrm{AIC}$ が成り立つが, このことが Table 5.1–5.3 からも確認できる. すなわち, AIC はリスクを大きく推定してしまい, 漸近不偏な推定量でないことがわかる.

また, $\mathrm{PE_{AIC_{SO}}}$ および $\mathrm{PE_{AIC}}$ に関しては, Case 1, 2 の場合, Table 5.1, 5.2 より $\mathrm{PE_{AIC_{SO}}}$ のほうが小さくなっている. すなわち, この場合においては $\mathrm{AIC_{SO}}$ を用いたほうが良いということになる. 一方で, Case 3 の場合, Table 5.3 より $\mathrm{PE_{AIC}}$ のほうが小さくなっている. すなわち, この場合においては AIC を用いたほうが良いということになる. このような結果となった理由のひとつは AIC の推定値から説明がつく. AIC はリスクを大きく推定するため, リスクが小さいモデルが選ばれにくくなることがある. 実際, Case 1 の $N = 40$ の場合, リスクは小さい順に Model 3, 4, 2, 1 であるが, AIC の値は小さい順に Model 3, 1, 2, 4 であり, 本来 2 番目にリスクが小さい Model 4 が最も AIC の値を大きくしてしまっている. 故に, リスクが大きい Model 1 や 2 をより多く選んでしまい, 結果 $\mathrm{PE_{AIC}}$ の値を大きくしてしまっている. すなわち, リスクを大きく評価していることが悪い方へ働いた結果といえる. しかしながら, Case 3 の場合, $\mathrm{PE_{AIC}}$ の方が値が小さくなっている. この理由は, Table 5.3 より, リスク, $\mathrm{AIC_{SO}}$, AIC の値はいずれも小さい順に Model 1, 2, 3, 4 であり, AIC はリスクを大きく評価するため, リスク最小モデルである Model 1 をより選びやすくなっているからである. すなわち, リスクを大きく評価したことが功を奏した結果といえる.

6. まとめと今後の課題

順序制約, SO が課せられた ANOVA モデルにおける AIC 規準, $\mathrm{AIC_{SO}}$ を導出した. $\mathrm{AIC_{SO}}$ はリスクに対する漸近不偏な推定量であり, また, 扱いやすいという点から通常の AIC と同等に良い推定量である. 加えて, Table 4.1 より, $\mathrm{AIC_{SO}}$ と AIC の間には類似の対応が見られた. 正則条件が成立している下で導かれた AIC と正則条件が成立していない複雑な設定下で導かれた $\mathrm{AIC_{SO}}$ がこのような対応を見せることは興味深い. また, モデル選択を行う場合において, SO の課し方に関するフルサーチを行う場合はリスクに対する漸近不偏な推定量ではない AIC を用いても良いということを理論的に示すことができた. 多くの分野で用いられ, また, 解析者にとっても馴染みがある AIC を使ってもよいという結果を導けたことは有意義である.

しかしながら, まだいくつかの残された課題がある. まずはじめに, 今回は母平均に対して SO を課したが, 他の制約, 例えば, 傘型制約 ($\theta_1 \leq \cdots \leq \theta_{k_1} \geq \cdots \geq \theta_{k_2}$, $2 \leq k_1 < k_2$) 等も考えるべきであろう. また, 今回扱ったモデルは正規性を仮定して

いたが, より一般的な楕円分布に従うといったような拡張も重要であると考えられる. 同様に, 平均構造に関しても, ANOVA モデルをより一般の重回帰モデルや GLM へと拡張することも重要である. 最後に, モデル選択規準としての $\mathrm{AIC_{SO}}$ と AIC の比較に関してはまだ十分な検証ができているとは言えないことである. 実際, 数値実験の結果は $\mathrm{AIC_{SO}}$ を用いたほうが良い場合もあれば, AIC を用いたほうが良い場合もあるということを明らかにしたにすぎない. どのような状況下でどちらが優れているのか, あるいは, 現実のデータに対して両者を適用した際はどちらが優れているのか, 等を明らかにしていく必要がある. これらの残された問題は今後の課題としておく.

Appendix

本節では, Lemma 3.1 の証明に必要となる 7 つの lemma, Lemma A–G を紹介する. Lemma 3.1 は Lemma F と Lemma G を示すことで得られる.

Lemma A. l を 2 以上の自然数とし, $n_1, \ldots, n_l \in \mathbb{R}_{>0}$, $\boldsymbol{n} = (n_1, \ldots, n_l)'$ とする. $\boldsymbol{x} = (x_1, \ldots, x_l)'$ を $\mathbb{R}^l$ の任意の元とする. このとき, 以下が成り立つ.

(i) $1 \leq a \leq b < c \leq l$ なる任意の自然数 a, b, c に対し,

$$\bar{x}_{[a,b]} \geq \bar{x}_{[a,c]} \Leftrightarrow \bar{x}_{[a,b]} \geq \bar{x}_{[b+1,c]} \Leftrightarrow \bar{x}_{[a,c]} \geq \bar{x}_{[b+1,c]},$$

および

$$\bar{x}_{[a,b]} < \bar{x}_{[a,c]} \Leftrightarrow \bar{x}_{[a,b]} < \bar{x}_{[b+1,c]} \Leftrightarrow \bar{x}_{[a,c]} < \bar{x}_{[b+1,c]}.$$

(ii) i を $2 \leq i \leq l$ なる任意の自然数とする. $w_1, \ldots, w_i$ を $w_1 < w_2 < \cdots < w_i$ を満たす任意の自然数とする. ただし, w_i は常に $w_i = l$ とし, また, w_0 を $w_0 = 0$ と定める. このとき,

$$\bar{x}_{[1+w_0,w_1]} < \bar{x}_{[1+w_1,w_2]} < \cdots < \bar{x}_{[1+w_{i-1},w_i]}$$

が成り立っているのならば, $1 \leq s < t \leq i$ なる任意の自然数 s, t に対して以下が成立する.

$$\bar{x}_{[1+w_{s-1},w_s]} < \bar{x}_{[1+w_{s-1},w_t]}.$$

(iii) i, j を $1 \leq i < j \leq l$ を満たす任意の自然数とする. このとき, 以下が成り立つ.

$$\bar{x}_{[i,b]} \geq \bar{x}_{[b+1,j]},\ ({}^{\forall}b \in \mathbb{N} \text{ with } i \leq b < j) \Leftrightarrow \boldsymbol{D}_{[i,j]}\boldsymbol{x}_{[i,j]} \geq \boldsymbol{0}_{j-i}.$$

Lemma B. l を 2 以上の自然数とし, $n_1, \ldots, n_l \in \mathbb{R}_{>0}$, $\boldsymbol{n} = (n_1, \ldots, n_l)'$ とする. $\boldsymbol{x} = (x_1, \ldots, x_l)'$ を $\mathbb{R}^l$ の任意の元とする. i を $2 \le i \le l$ なる任意の自然数とする. $w_1, \ldots, w_i$ を $w_1 < w_2 < \cdots < w_i$ を満たす任意の自然数とする. ただし, w_i は常に $w_i = l$ とし, また, w_0 を $w_0 = 0$ と定める. また,

$$\begin{aligned} \eta_l(\boldsymbol{x})[1] = \cdots &= \eta_l(\boldsymbol{x})[w_1], \\ \eta_l(\boldsymbol{x})[w_1 + 1] = \cdots &= \eta_l(\boldsymbol{x})[w_2], \\ &\vdots \\ \eta_l(\boldsymbol{x})[w_{i-1} + 1] = \cdots &= \eta_l(\boldsymbol{x})[w_i], \end{aligned}$$

および

$$\eta_l(\boldsymbol{x})[w_j] = \bar{x}_{[1+w_{j-1}, w_j]}, \quad (1 \le j \le i),$$

更に

$$\bar{x}_{[1,w_1]} < \bar{x}_{[1+w_1,w_2]} < \cdots < \bar{x}_{[1+w_{i-1},w_i]},$$

が成り立っているとする. このとき, 以下が成り立つ.

(i) s を $1 < s \le i$ なる任意の自然数とする. このとき, $s \le t \le i$ なる任意の自然数 t に対して,

$$\boldsymbol{D}_{1+w_{t-1},w_t} \boldsymbol{x}_{[1+w_{t-1},w_t]} \ge \boldsymbol{0}_{w_t - w_{t-1} - 1},$$

ならば,

$$\boldsymbol{D}_{1+w_{s-2},w_{s-1}} \boldsymbol{x}_{[1+w_{s-2},w_{s-1}]} \ge \boldsymbol{0}_{w_{s-1} - w_{s-2} - 1}.$$

(ii) $1 \le t \le i$ なる任意の自然数 t に対して,

$$\boldsymbol{D}_{1+w_{t-1},w_t} \boldsymbol{x}_{[1+w_{t-1},w_t]} \ge \boldsymbol{0}_{w_t - w_{t-1} - 1},$$

が成立する.

Lemma C. l を 2 以上の自然数とし, $n_1, \ldots, n_l \in \mathbb{R}_{>0}$, $\boldsymbol{n} = (n_1, \ldots, n_l)'$ とする. $\xi_1, \ldots, \xi_l \in \mathbb{R}$, $\tau^2 > 0$ とする. $x_1, \ldots, x_l$ を互いに独立な確率変数とし, $1 \le s \le l$ に対し, $x_s \sim N(\xi_s, \tau^2/n_s)$ とする. このとき, $\boldsymbol{x} = (x_1, \ldots, x_l)'$ に対し, 以下が成り立つ.

(i)

$$\mathbb{R}^l = \bigcup_{i=1}^{l} \bigcup_{\boldsymbol{w};\boldsymbol{w} \in \mathcal{W}_i^l} \boldsymbol{\eta}_l^{-1}(\mathcal{A}_i^l(\boldsymbol{w})),$$

$$\boldsymbol{\eta}_l^{-1}(\mathcal{A}_i^l(\boldsymbol{w})) \cap \boldsymbol{\eta}_l^{-1}(\mathcal{A}_{i^*}^l(\boldsymbol{w}^*)) = \emptyset, \quad ((i, \boldsymbol{w}) \ne (i^*, \boldsymbol{w}^*)).$$

(ii) $\mathcal{A}_1^l(\boldsymbol{w}) = \mathcal{A}_1^l$ に対して,

$$\boldsymbol{x} \in \boldsymbol{\eta}_l^{-1}(\mathcal{A}_1^l(\boldsymbol{w})) \Leftrightarrow \boldsymbol{D}_{1,l}\boldsymbol{x}_{[1,l]} \geq \boldsymbol{0}_{l-1},$$

が成り立つ. また, $2 \leq i \leq l$, $\boldsymbol{w} = (w_1, \ldots, w_i)' \in \mathcal{W}_i^l$ なる i, $\boldsymbol{w}$ に対し,

$$\begin{aligned}\boldsymbol{x} \in \boldsymbol{\eta}_l^{-1}(\mathcal{A}_i^l(\boldsymbol{w})) \Leftrightarrow 0 \leq t \leq i-1,\ &\boldsymbol{D}_{1+w_t, w_{t+1}}\boldsymbol{x}_{[1+w_t, w_{t+1}]} \geq \boldsymbol{0}_{\rho_{t,\boldsymbol{w}}},\\ 0 \leq s \leq i-2,\ &\bar{x}_{[1+w_s, w_{s+1}]} < \bar{x}_{[1+w_{s+1}, w_{s+2}]},\end{aligned}$$

が成り立つ. ただし, $w_0 = 0$ であり, また, $\rho_{t,\boldsymbol{w}} = w_{t+1} - w_t - 1$ である.

(iii) $1 \leq i \leq l$, $\boldsymbol{w} = (w_1, \ldots, w_i)' \in \mathcal{W}_i^l$ なる i, $\boldsymbol{w}$ に対し,

$$\begin{aligned}\boldsymbol{x} \in \boldsymbol{\eta}_l^{-1}(\mathcal{A}_i^l(\boldsymbol{w})) \Rightarrow 0 \leq t \leq i-1,\ \eta_l(\boldsymbol{x})[1+w_t] = \cdots &= \eta_l(\boldsymbol{x})[w_{t+1}]\\ &= \bar{x}_{[1+w_t, w_{t+1}]},\end{aligned}$$

ただし, $w_0 = 0$.

(iv) $1 \leq i \leq l$ なる i に対し,

$$\sum_{\boldsymbol{w};\boldsymbol{w} \in \mathcal{W}_i^l} \mathrm{P}\left(\boldsymbol{x} \in \boldsymbol{\eta}_l^{-1}(\mathcal{A}_i^l(\boldsymbol{w}))\right) = \mathrm{P}\left(\boldsymbol{\eta}_l(\boldsymbol{x}) \in \bigcup_{\boldsymbol{w};\boldsymbol{w} \in \mathcal{W}_i^l} \mathcal{A}_i^l(\boldsymbol{w})\right).$$

Lemma D. $v_1, \ldots, v_l$ を互いに独立な確率変数とし, $1 \leq s \leq l$ に対し, $v_s \sim N(\xi_s, \tau^2/n_s)$ とする. $\tau^2 > 0$, $\xi_1, \ldots, \xi_l \in \mathbb{R}$, $n_1, \ldots, n_l \in \mathbb{R}_{>0}$, $\boldsymbol{n} = (n_1, \ldots, n_l)'$, $\boldsymbol{v} = (v_1, \ldots, v_l)'$ とする. このとき, $1 \leq i \leq j \leq l$ に対して,

$$\boldsymbol{D}_{i,j}\boldsymbol{v}_{[i,j]} \perp\!\!\!\perp \bar{v}_{[i,j]},$$

および

$$\bar{v}_{[i,j]} \perp\!\!\!\perp \sum_{s=i}^{j} n_s(v_s - \xi_s)(v_s - \bar{v}_{[i,j]}),$$

が成立する.

Lemma E. 確率変数 $v_1, \ldots, v_l$ を Lemma D で定義されたものとする. このとき, 集合

$$\begin{aligned}\mathcal{A}_l^l &= \{(x_1, \ldots, x_l)' \in \mathbb{R}^l \mid x_1 < x_2 < \cdots < x_l\}\\ &= \{(x_1, \ldots, x_l)' \in \mathbb{R}^l \mid 1 \leq t \leq l-1,\ x_t < x_{t+1}\},\end{aligned}$$

に対して, 以下が成立する.

$$\begin{aligned}
&\mathrm{E}\left[1_{\{\boldsymbol{v}\in\boldsymbol{\eta}_l^{-1}(\mathcal{A}_l^l)\}}\times\frac{1}{\tau^2}\sum_{s=1}^{l}n_s v_s(v_s-\xi_s)\right]\\
&=\mathrm{E}\left[1_{\{\boldsymbol{v}\in\mathcal{A}_l^l\}}\times\frac{1}{\tau^2}\sum_{s=1}^{l}n_s v_s(v_s-\xi_s)\right]\\
&=l\mathrm{E}[1_{\{\boldsymbol{v}\in\mathcal{A}_l^l\}}]=l\mathrm{E}[1_{\{\boldsymbol{v}\in\boldsymbol{\eta}_l^{-1}(\mathcal{A}_l^l)\}}]=l\mathrm{P}(\boldsymbol{v}\in\boldsymbol{\eta}_l^{-1}(\mathcal{A}_l^l)).
\end{aligned}$$

Lemma F. n_1, n_2 を任意の正数とし, $\boldsymbol{n}=(n_1,n_2)'$ とする. ξ_1, ξ_2 を任意の実数, τ^2 を任意の正の数とする. x_1, x_2 は互いに独立な確率変数とし, $\boldsymbol{x}=(x_1,x_2)'$, $x_s\sim N(\xi_s,\tau^2/n_s)$, $(s=1,2)$ とする. このとき, 以下の命題が成立する.

(P1) $1\le i\le j\le 2$ なる任意の自然数 i, j に対し,

$$\begin{aligned}
&\mathrm{E}\left[1_{\{\boldsymbol{D}_{i,j}^{(\boldsymbol{n})}\boldsymbol{x}_{[i,j]}\ge\boldsymbol{0}_{j-i}\}}\frac{1}{\tau^2}\sum_{s=i}^{j}n_s(x_s-\xi_s)(x_s-\bar{x}_{[i,j]}^{(\boldsymbol{n})})\right]\\
&=(j-i)\mathrm{P}(\boldsymbol{D}_{i,j}^{(\boldsymbol{n})}\boldsymbol{x}_{[i,j]}\ge\boldsymbol{0}_{j-i}).
\end{aligned}$$

(P2) $\boldsymbol{w}=(2)'\in\mathcal{W}_1^2$ に対して,

$$\mathrm{E}\left[\frac{1}{\tau^2}\sum_{s=1}^{2}n_s(x_s-\xi_s)(x_s-\eta_2^{(\boldsymbol{n})}(\boldsymbol{x})[s])\right]=\mathrm{P}\left(\boldsymbol{\eta}_2^{(\boldsymbol{n})}(\boldsymbol{x})\in\mathcal{A}_1^2(\boldsymbol{w})\right).$$

Lemma G. l を 2 以上の任意の自然数とする. 次の命題 (P) が真であると仮定する.

(P) $N_1,\dots,N_l$ を任意の正数とし, $\boldsymbol{N}=(N_1,\dots,N_l)'$ とする. $\zeta_1,\dots,\zeta_l$ を任意の実数, ς^2 を任意の正の数とし, $\boldsymbol{\zeta}=(\zeta_1,\dots,\zeta_l)'$ とする. $y_1,\dots,y_l$ は互いに独立な確率変数とし, $\boldsymbol{y}=(y_1,\dots,y_l)'$, $y_s\sim N(\zeta_s,\varsigma^2/N_s)$, $(s=1,\dots,l)$ とする. このとき, $1\le i\le j\le l$ なる任意の自然数 i, j に対し, 以下が成立する.

$$\begin{aligned}
&\mathrm{E}\left[1_{\{\boldsymbol{D}_{i,j}^{(\boldsymbol{N})}\boldsymbol{y}_{[i,j]}\ge\boldsymbol{0}_{j-i}\}}\frac{1}{\varsigma^2}\sum_{s=i}^{j}N_s(y_s-\zeta_s)(y_s-\bar{y}_{[i,j]}^{(\boldsymbol{N})})\right]\\
&=(j-i)\mathrm{P}(\boldsymbol{D}_{i,j}^{(\boldsymbol{N})}\boldsymbol{y}_{[i,j]}\ge\boldsymbol{0}_{j-i}).
\end{aligned}$$

このとき, 以下の命題 (P*) が成立する.

(P*) $n_1,\dots,n_{l+1}$ を任意の正数とし, $\boldsymbol{n}=(n_1,\dots,n_{l+1})'$ とする. $\xi_1,\dots,\xi_{l+1}$ を任意の実数, τ^2 を任意の正の数とし, $\boldsymbol{\xi}=(\xi_1,\dots,\xi_{l+1})'$ とする. $x_1,\dots,x_{l+1}$

は互いに独立な確率変数とし, $\boldsymbol{x}=(x_1,\ldots,x_{l+1})'$, $x_s \sim N(\xi_s,\tau^2/n_s)$, $(s=1,\ldots,l+1)$ とする. このとき, $1\le i\le j\le l+1$ なる任意の自然数 i, j に対し, 以下が成立する.

$$\mathrm{E}\left[1_{\{\boldsymbol{D}_{i,j}^{(n)}\boldsymbol{x}_{[i,j]}\ge \boldsymbol{0}_{j-i}\}}\frac{1}{\tau^2}\sum_{s=i}^{j} n_s(x_s-\xi_s)(x_s-\bar{x}_{[i,j]}^{(n)})\right]$$
$$=(j-i)\mathrm{P}(\boldsymbol{D}_{i,j}^{(n)}\boldsymbol{x}_{[i,j]}\ge \boldsymbol{0}_{j-i}).$$

更に, 次の式が成り立つ.

$$\mathrm{E}\left[\frac{1}{\tau^2}\sum_{s=1}^{l+1} n_s(x_s-\xi_s)(x_s-\eta_{l+1}^{(n)}(\boldsymbol{x})[s])\right]$$
$$=\sum_{i=1}^{l}(l+1-i)\mathrm{P}\left(\eta_{l+1}(\boldsymbol{x})\in \bigcup_{\boldsymbol{w};\boldsymbol{w}\in\mathcal{W}_i^{l+1}} \mathcal{A}_i^{l+1}(\boldsymbol{w})\right).$$

References

Anraku, K. (1999). An information criterion for parameters under a simple order restriction. *Biometrika*, **86**, 141–152.

Brunk, H. D. (1965). Conditional expectation given a σ-lattice and application. *Ann. Math. Statist.*, **36**, 1339–1350.

Hwang, J. T. and Peddada, S. D. (1994). Confidence interval estimation subject to order restrictions. *Ann. Statist.*, **22**, 67–93.

Kelly R. (1989). Stochastic reduction of loss in estimating normal means by isotonic regression. *Ann. Statist.*, **17**, 937–940.

Lee, C. C. (1981). The quadratic loss of isotonic regression under normality. *Ann. Statist.*, **9**, 686–688.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inference.* Wiley.

Rudin, W. (1986). *Real and Complex Analysis.* McGraw-Hill.

Yokoyama, T. (1995). LR test for random-effects covariance structure in a parallel profile model. *Ann. Inst. Statist. Math.*, **47**, 309–320.

Yokoyama, T. and Fujikoshi, Y. (1993). A parallel profile model with random-effects covariance structure. *J. Japan Statist. Soc.*, **23**, 83–89.