Model Selection Criteria for ANOVA Model with a Tree Order Restriction

Yu Inatsu[*1]

*1 *Department of Mathematics, Graduate School of Science, Hiroshima University*

ABSTRACT

In this paper, we consider Akaike information criterion (AIC) and C_p criterion for ANOVA model with a tree ordering (TO) $\theta_1 \leq \theta_j,\ (j = 2, \ldots, l)$ where $\theta_1, \ldots, \theta_l$ are population means. In general, under ANOVA model with the TO, the AIC and the C_p criterion have asymptotic biases which depend on unknown parameters. In order to solve these problems, we calculate (asymptotic) biases, and we derive its unbiased estimators. By using these estimators, we provide an asymptotically unbiased AIC and an "unbiased" C_p criterion for ANOVA model with the TO, called $\mathrm{AIC}_{\mathrm{TO}}$ and $\mathrm{TO}C_p$, respectively. Penalty terms of derived criteria are simply defined as a function of an indicator function and maximum likelihood estimators. Furthermore, we show that the $\mathrm{TO}C_p$ is the uniformly minimum-variance unbiased estimator (UMVUE).

Key Words: Order restriction, Tree ordering, AIC, C_p, UMVUE, ANOVA.

1. Introduction

In real data analysis, ANOVA model is often used for analyzing cluster data. Moreover, a model whose parameters $\mu_1, \ldots, \mu_l$ are restricted such as a Sinple Ordering (SO) given by $\mu_1 \leq \cdots \leq \mu_l$, is also important in the field of applied statistics (e.g., Robertson *et al.*, 1988). In addition, Brunk (1965), Lee (1981), Kelly (1989) and Hwang and Peddada (1994) showed that maximum likelihood estimators (MLEs) for mean parameters of ANOVA model with the SO are more efficient than those of ANOVA model without any restriction when the assumption of the SO is true.

On the other hand, in general, the classical asymptotic theory does not hold for the model with parameter restrictions. For example, Anraku (1999) showed that an ordinal Akaike information criterion (AIC, Akaike, 1973) for ANOVA model with the SO, whose penalty term is $2\times$ the number of parameters, is not an asymptotically unbiased estimator of a risk function. In order to solve this problem, Inatsu (2016) derived an asymptotically unbiased AIC for ANOVA model with the SO, called $\mathrm{AIC}_{\mathrm{SO}}$. Furthermore, a penalty term of the $\mathrm{AIC}_{\mathrm{SO}}$ can be simply defined as a function of MLEs of mean parameters. Nevertheless, there are other important restrictions in applied statistics.

In this paper, we consider ANOVA model with a Tree Ordering (TO) given by $\mu_1 \leq \mu_j$ $(j = 2, \ldots, l)$. For this model, we derive an asymptotically unbiased AIC, called $\mathrm{AIC}_{\mathrm{TO}}$. Similarly, we also derive an "unbiased" C_p criterion (Mallows, 1973) for this model.

The remainder of the present paper is organized as follows: In Section 2, we define the true model

[*1] Corresponding author

E-mail address: d144576@hiroshima-u.ac.jp

and candidate model. Moreover, we derive MLEs of parameters in the candidate model. In Section 3, we provide the AIC for ANOVA model with the TO, called $\mathrm{AIC_{TO}}$. In Section 4, we provide the C_p criterion for ANOVA model with the TO, called $\mathrm{TO}C_p$. In addition, we show that the $\mathrm{TO}C_p$ is the uniformly minimum-variance unbiased estimator (UMVUE). In Section 5, we confirm estimation accuracy of the $\mathrm{AIC_{TO}}$ and the $\mathrm{TO}C_p$ through numerical experiments. In Section 6, we conclude our discussion. Technical details are provided in Appendix.

2. ANOVA model with a tree order restriction

In this section, we define the true model, and candidate models with order restrictions. The MLE for the considered candidate model is given in Subsection 2.3.

2.1. True and candidate models

Let Y_{ij} be a observation variable on the jth individual in the ith cluster, where $1 \le i \le k^*$, $j = 1, \dots, N_i$ for each i, and $k^* \ge 2$. Here, we put $N = N_1 + \cdots + N_{k^*}$ and $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{iN_i})'$ for each i. Also we put $\boldsymbol{Y} = (\boldsymbol{Y}_1', \dots, \boldsymbol{Y}_{k^*}')'$ and $\boldsymbol{N} = (N_1, \dots, N_{k^*})'$.

Suppose that $Y_{11}, \dots, Y_{k^* N_{k^*}}$ are mutually independent, and Y_{ij} is distributed as

$$Y_{ij} \sim N(\mu_{i,*}, \sigma_*^2), \tag{2.1}$$

for any i and j. Here, $\mu_{i,*}$ and σ_*^2 are unknown true values satisfying $\mu_{i,*} \in \mathbb{R}$ and $\sigma_*^2 > 0$, respectively. In other words, the true model is given by (2.1).

Next, we define a candidate model. Let $Q_1, \dots, Q_k$ be non-empty disjoint sets satisfying $Q_1 \cup \cdots \cup Q_k = \{1, 2, \dots, k^*\}$, where $2 \le k \le k^*$. Then, we assume that $Y_{11}, \dots, Y_{k^* N_{k^*}}$ are mutually independent, and distributed as

$$Y_{ij} \sim N(\mu_i, \sigma^2), \tag{2.2}$$

where $\mu_1, \dots, \mu_{k^*}$ and $\sigma^2 (> 0)$ are unknown parameters. In addition, for the parameters $\mu_1, \dots, \mu_{k^*}$, we assume that

$$1 \le^{\forall} s \le k, \ {}^{\forall} u_1, u_2 \in Q_s, \quad \mu_{u_1} = \mu_{u_2}, \tag{2.3}$$

and

$$2 \le^{\forall} t \le k, \ {}^{\forall} \nu \in Q_t, \quad \mu_q \le \mu_\nu, \tag{2.4}$$

where $q \in Q_1$. Then, a candidate model $\mathcal{M}$ is defined as the model (2.2) with (2.3) and (2.4). In particular, the order restriction (2.4) is called a Tree Ordering (TO). For example, when $k^* = 7$, $k = 4$, $Q_1 = \{1, 3, 7\}$, $Q_2 = \{2\}$, $Q_3 = \{4, 5\}$ and $Q_4 = \{6\}$, the unknown parameters $\mu_1, \dots, \mu_7$ for the candidate model $\mathcal{M}$ are restricted as

$$\mu_1 = \mu_3 = \mu_7 \le \mu_2, \quad \mu_1 = \mu_3 = \mu_7 \le \mu_4 = \mu_5, \quad \mu_1 = \mu_3 = \mu_7 \le \mu_6.$$

2.2. Notation and lemma

In this subsection, we define several notations. After that, we provide the related lemma. Let l be an integer with $l \ge 2$. Then, define

$$\mathbb{N}_l = \{x \in \mathbb{N} \mid x \le l\} = \{1, \dots, l\}.$$

Moreover, let $x_1, \dots, x_l$ be real numbers, and let $N_1, \dots, N_l$ be positive numbers. We put $\boldsymbol{x} = (x_1, \dots, x_l)'$ and $\boldsymbol{N} = (N_1, \dots, N_l)'$. Furthermore, let $A = \{a_1, \dots, a_i\}$ be a non-empty subset of $\mathbb{N}_l$, where $a_1 < \cdots < a_i$ when $i \geq 2$.

Next, define

$$\boldsymbol{x}_A = (x_{a_1}, \dots, x_{a_i})', \quad \tilde{x}_A = \sum_{s \in A} x_s, \quad \bar{x}_A^{(\boldsymbol{N})} = \frac{\sum_{s \in A} N_s x_s}{\sum_{s \in A} N_s} = \frac{\sum_{s \in A} N_s x_s}{\tilde{N}_A}.$$

For example, when $l = 10$ and $A = \{2, 3, 5, 10\}$, $\boldsymbol{x}_A$, $\tilde{x}_A$ and $\bar{x}_A^{(\boldsymbol{N})}$ are given by

$$\begin{aligned} \boldsymbol{x}_A &= (x_2, x_3, x_5, x_{10})', \quad \tilde{x}_A = x_2 + x_3 + x_5 + x_{10}, \\ \bar{x}_A^{(\boldsymbol{N})} &= \frac{N_2 x_2 + N_3 x_3 + N_5 x_5 + N_{10} x_{10}}{N_2 + N_3 + N_5 + N_{10}}. \end{aligned}$$

In particular, when A has only one element a, i.e., $A = \{a\}$, it holds that $\boldsymbol{x}_A = (x_a)'$, $\tilde{x}_A = x_a$ and $\bar{x}_A^{(\boldsymbol{N})} = x_a$. On the other hand, when $A = \mathbb{N}_l$, it holds that $\boldsymbol{x}_A = \boldsymbol{x}$. For simplicity, we often represent $\bar{x}_A^{(\boldsymbol{N})}$ as $\bar{x}_A$. In addition, let $A^{(l)}$ be a set defined as

$$\begin{aligned} A^{(l)} &= \{(x_1, \dots, x_l)' \in \mathbb{R}^l \mid {}^{\forall} j \in \mathbb{N}_l \setminus \{1\},\ x_1 \leq x_j\} \\ &= \{(x_1, \dots, x_l)' \in \mathbb{R}^l \mid x_1 \leq x_2, \dots, x_1 \leq x_l\}. \end{aligned}$$

Furthermore, for any integer i with $1 \leq i \leq l$, we consider a family of sets $\mathcal{J}_i^{(l)}$ defined by

$$\mathcal{J}_i^{(l)} = \{J \subset \mathbb{N}_l \mid 1 \in J,\ \#J = i\},$$

where $\#J$ means the number of elements of the set J. For example, when $l = 4$, it holds that

$$\begin{aligned} &\mathcal{J}_1^{(4)} = \{\ \{1\}\ \},\ \mathcal{J}_2^{(4)} = \{\ \{1,2\}, \{1,3\}, \{1,4\}\ \},\ \mathcal{J}_3^{(4)} = \{\ \{1,2,3\}, \{1,2,4\}, \{1,3,4\}\ \}, \\ &\mathcal{J}_4^{(4)} = \{\ \{1,2,3,4\}\ \} = \{\ \mathbb{N}_4\ \}. \end{aligned}$$

Here, note that $\mathcal{J}_1^{(l)} = \{\ \{1\}\ \}$ and $\mathcal{J}_l^{(l)} = \{\ \mathbb{N}_l\ \}$ for any $l \geq 2$. Similarly, for any integer i with $1 \leq i \leq l$ and for any set J with $\mathcal{J}_i^{(l)}$, we consider the following set $A^{(l)}(J)$:

$$A^{(l)}(J) = \{(x_1, \dots, x_l)' \in \mathbb{R}^l \mid {}^{\forall} s \in J,\ x_1 = x_s, \quad {}^{\forall} t \in \mathbb{N}_l \setminus J,\ x_1 < x_t\}.$$

Note that when $J = \mathbb{N}_l$, it holds that $\mathbb{N}_l \setminus J = \emptyset$. In this case, the proposition

$${}^{\forall} t \in \emptyset,\ x_1 < x_t$$

is always true. For example, when $l = 4$, it holds that

$$\begin{aligned} A^{(4)}(\{1\}) &= \{\boldsymbol{x} = (x_1, \dots, x_4)' \in \mathbb{R}^4 \mid x_1 < x_2,\ x_1 < x_3,\ x_1 < x_4\}, \\ A^{(4)}(\{1,2\}) &= \{\boldsymbol{x} \in \mathbb{R}^4 \mid x_1 = x_2,\ x_1 < x_3,\ x_1 < x_4\}, \\ A^{(4)}(\{1,3\}) &= \{\boldsymbol{x} \in \mathbb{R}^4 \mid x_1 = x_3,\ x_1 < x_2,\ x_1 < x_4\}, \\ A^{(4)}(\{1,4\}) &= \{\boldsymbol{x} \in \mathbb{R}^4 \mid x_1 = x_4,\ x_1 < x_2,\ x_1 < x_3\}, \\ A^{(4)}(\{1,2,3\}) &= \{\boldsymbol{x} \in \mathbb{R}^4 \mid x_1 = x_2 = x_3,\ x_1 < x_4\}, \\ A^{(4)}(\{1,2,4\}) &= \{\boldsymbol{x} \in \mathbb{R}^4 \mid x_1 = x_2 = x_4,\ x_1 < x_3\}, \\ A^{(4)}(\{1,3,4\}) &= \{\boldsymbol{x} \in \mathbb{R}^4 \mid x_1 = x_3 = x_4,\ x_1 < x_2\}, \\ A^{(4)}(\{1,2,3,4\}) &= \{\boldsymbol{x} \in \mathbb{R}^4 \mid x_1 = x_2 = x_3 = x_4\}. \end{aligned}$$

It is clear that these eight sets are disjoint sets and

$$\bigcup_{i=1}^{4} \bigcup_{J \in \mathcal{J}_i^{(4)}} A^{(4)}(J) = \{\boldsymbol{x} \in \mathbb{R}^4 \mid x_1 \leq x_2,\ x_1 \leq x_3,\ x_1 \leq x_4\} = A^{(4)}.$$

Similarly, in the case of $l \geq 2$, it holds that

$$\bigcup_{i=1}^{l} \bigcup_{J \in \mathcal{J}_i^{(l)}} A^{(l)}(J) = \{\boldsymbol{x} \in \mathbb{R}^l \mid x_1 \leq x_2, \dots, x_1 \leq x_l\} = A^{(l)}, \tag{2.5}$$

and $A^{(l)}(J) \cap A^{(l)}(J^*) = \emptyset$ when $J \neq J^*$.

Next, given an integer s with $1 \leq s \leq l$ and a real number a. Then, for the vector $\boldsymbol{x} = (x_1, \dots, x_l)'$, let $\boldsymbol{x}[s; a]$ be an l-dimensional vector whose sth element is a and tth element ($t \in \mathbb{N}_l \setminus \{s\}$) is x_t. For example, if $\boldsymbol{x} = (1, 4, 4, 3)'$, then $\boldsymbol{x}[2; -1] = (1, -1, 4, 3)'$ and $\boldsymbol{x}[4; 5] = (1, 4, 4, 5)'$. Moreover, for any integer s with $1 \leq s \leq l$ and for any set $J = \{j_1, \dots, j_s\}$ of $\mathcal{J}_s^{(l)}$, we define a matrix $\boldsymbol{D}_J^{(\boldsymbol{N})}$ where $j_1 < \cdots < j_s$ when $s \geq 2$. First, in the case of $s = 1$, the family of sets $\mathcal{J}_1^{(l)}$ has only one set $J = \{1\}$, and we define $\boldsymbol{D}_J^{(\boldsymbol{N})} = 0$. On the other hand, in the case of $s \geq 2$, the matrix $\boldsymbol{D}_J^{(\boldsymbol{N})}$ is the $s - 1 \times s$ matrix whose ith row ($1 \leq i \leq s - 1$) is defined as

$$\frac{1}{\tilde{N}_{J \setminus \{j_{i+1}\}}} \boldsymbol{N}_J[i + 1; -\tilde{N}_{J \setminus \{j_{i+1}\}}]'.$$

For example, when $l = 4$, it holds that

$$\begin{aligned}
&\boldsymbol{D}_{\{1\}}^{(\boldsymbol{N})} = 0, \quad \boldsymbol{D}_{\{1,2\}}^{(\boldsymbol{N})} = \boldsymbol{D}_{\{1,3\}}^{(\boldsymbol{N})} = \boldsymbol{D}_{\{1,4\}}^{(\boldsymbol{N})} = (1 \quad -1), \\
&\boldsymbol{D}_{\{1,2,3\}}^{(\boldsymbol{N})} = \begin{pmatrix} \frac{N_1}{N_1+N_3} & -1 & \frac{N_3}{N_1+N_3} \\ \frac{N_1}{N_1+N_2} & \frac{N_2}{N_1+N_2} & -1 \end{pmatrix}, \ \boldsymbol{D}_{\{1,2,4\}}^{(\boldsymbol{N})} = \begin{pmatrix} \frac{N_1}{N_1+N_4} & -1 & \frac{N_4}{N_1+N_4} \\ \frac{N_1}{N_1+N_2} & \frac{N_2}{N_1+N_2} & -1 \end{pmatrix}, \\
&\boldsymbol{D}_{\{1,3,4\}}^{(\boldsymbol{N})} = \begin{pmatrix} \frac{N_1}{N_1+N_4} & -1 & \frac{N_4}{N_1+N_4} \\ \frac{N_1}{N_1+N_3} & \frac{N_3}{N_1+N_3} & -1 \end{pmatrix}, \\
&\boldsymbol{D}_{\{1,2,3,4\}}^{(\boldsymbol{N})} = \begin{pmatrix} \frac{N_1}{N_1+N_3+N_4} & -1 & \frac{N_3}{N_1+N_3+N_4} & \frac{N_4}{N_1+N_3+N_4} \\ \frac{N_1}{N_1+N_2+N_4} & \frac{N_2}{N_1+N_2+N_4} & -1 & \frac{N_4}{N_1+N_2+N_4} \\ \frac{N_1}{N_1+N_2+N_3} & \frac{N_2}{N_1+N_2+N_3} & \frac{N_3}{N_1+N_2+N_3} & -1 \end{pmatrix}.
\end{aligned}$$

For simplicity, we often represent $\boldsymbol{D}_J^{(\boldsymbol{N})}$ as $\boldsymbol{D}_J$.

Furthermore, we define a function $\boldsymbol{\eta}_l^{(\boldsymbol{N})}$ from $\mathbb{R}^l$ to $A^{(l)}$. For each vector $\boldsymbol{x} = (x_1, \dots, x_l)' \in \mathbb{R}^l$, $\boldsymbol{\eta}_l^{(\boldsymbol{N})}(\boldsymbol{x})$ is defined as

$$\boldsymbol{\eta}_l^{(\boldsymbol{N})}(\boldsymbol{x}) = \underset{\boldsymbol{y}=(y_1,\dots,y_l)' \in A^{(l)}}{\operatorname{argmin}} \sum_{i=1}^{l} N_i (x_i - y_i)^2. \tag{2.6}$$

In addition, let $\eta_l^{(\boldsymbol{N})}(\boldsymbol{x})[s]$ be the sth element ($1 \leq s \leq l$) of $\boldsymbol{\eta}_l^{(\boldsymbol{N})}(\boldsymbol{x})$. Note that well-definedness of $\boldsymbol{\eta}_l^{(\boldsymbol{N})}$ can be derived by using the Hilbert projection theorem (see, e.g., Rudin, 1986). For simplicity, we often represent $\boldsymbol{\eta}_l^{(\boldsymbol{N})}(\boldsymbol{x})$ as $\boldsymbol{\eta}_l(\boldsymbol{x})$.

Finally, we provide the following lemma:

Lemma 2.1. The following three propositions hold:

(1) It holds that

$$\mathbb{R}^l = \bigcup_{i=1}^{l} \bigcup_{J \in \mathcal{J}_i^{(l)}} \boldsymbol{\eta}_l^{-1}\left(A^{(l)}(J)\right),$$

$$\boldsymbol{\eta}_l^{-1}\left(A^{(l)}(J)\right) \cap \boldsymbol{\eta}_l^{-1}\left(A^{(l)}(J^*)\right) = \emptyset \quad (J \neq J^*).$$

(2) For any integer i with $1 \leq i \leq l$ and for any set J with $\mathcal{J}_i^{(l)}$, it holds that

$$\boldsymbol{\eta}_l^{-1}\left(A^{(l)}(J)\right) = \{\boldsymbol{x} = (x_1, \ldots, x_l)' \in \mathbb{R}^l \mid \boldsymbol{D}_J \boldsymbol{x}_J \geq \boldsymbol{0},\ {}^\forall t \in \mathbb{N}_l \setminus J,\ \bar{x}_J < x_t\}, \tag{2.7}$$

where the inequality $\boldsymbol{s} \geq \boldsymbol{0}$ means that all elements of the vector $\boldsymbol{s}$ are non-negative.

(3) Let i be an integer with $1 \leq i \leq l$, and let J be a set with $J \in \mathcal{J}_i^{(l)}$. Let $\boldsymbol{x} = (x_1, \ldots, x_l)'$ be an element of $\mathbb{R}^l$. Assume that $\boldsymbol{x}$ satisfies

$$\boldsymbol{x} \in \boldsymbol{\eta}_l^{-1}\left(A^{(l)}(J)\right).$$

Then, it holds that

$${}^\forall s \in J,\ \eta_l(\boldsymbol{x})[s] = \bar{x}_J, \quad {}^\forall t \in \mathbb{N}_l \setminus J,\ \eta_l(\boldsymbol{x})[t] = x_t.$$

In particular, for the case of $J = \mathbb{N}_l$, if $\boldsymbol{x}$ satisfies

$$\boldsymbol{x} \in \boldsymbol{\eta}_l^{-1}(A^{(l)}(J)) = \{\boldsymbol{x} \in \mathbb{R}^l \mid \boldsymbol{D}_J \boldsymbol{x}_J \geq \boldsymbol{0}\},$$

then, the following proposition holds:

$${}^\forall s \in J,\ \eta_l(\boldsymbol{x})[s] = \bar{x}_J.$$

The proof of Lemma 2.1 is given in Appendix 1.

2.3. Maximum likelihood estimators for unknown parameters

In this subsection, we derive MLEs for unknown parameters in the candidate model $\mathcal{M}$. First of all, we rewrite the candidate model. For any integer s with $1 \leq s \leq k$ and for all elements $q_1^{(s)}, \ldots, q_v^{(s)}$ of Q_s, let $\boldsymbol{X}_s = (\boldsymbol{Y}'_{q_1^{(s)}}, \ldots, \boldsymbol{Y}'_{q_v^{(s)}})'$, where v is the number of elements in Q_s. We put $\boldsymbol{X} = (\boldsymbol{X}'_1, \ldots, \boldsymbol{X}'_k)'$,

$$\mu_{q_1^{(s)}} = \cdots = \mu_{q_v^{(s)}} \equiv \theta_s,$$

and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)'$. In addition, define $n_s = N_{q_1^{(s)}} + \cdots + N_{q_v^{(s)}}$ and $\boldsymbol{n} = (n_1, \ldots, n_k)'$. Note that $n_1 + \cdots + n_k = N_1 + \cdots + N_{k^*} = N$. Then, the candidate model can be rewritten as

$$X_{st} \sim N(\theta_s, \sigma^2), \quad t = 1, \ldots, n_s,$$

with

$$\theta_1 \leq \theta_2, \ldots, \theta_1 \leq \theta_k.$$

Here, a parameter space Θ for the candidate model is defined as follows:

$$\Theta = \{(a_1, \ldots, a_k)' \in \mathbb{R}^k \mid {}^\forall u \in \mathbb{N}_k \setminus \{1\},\ a_1 \leq a_u\}.$$

Next, we consider a log-likelihood for the candidate model. Let

$$\bar{X}_s = \frac{1}{n_s}\sum_{v=1}^{n_s} X_{sv}, \quad s = 1, \ldots, k,$$

and let $\bar{\boldsymbol{X}} = (\bar{X}_1, \ldots, \bar{X}_k)'$. Then, since X_{st}'s are independently distributed as normal distribution, a log-likelihood function $l(\boldsymbol{\theta}, \sigma^2; \boldsymbol{X})$ is given by

$$\begin{aligned} l(\boldsymbol{\theta}, \sigma^2; \boldsymbol{X}) &= -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{s=1}^{k}\sum_{t=1}^{n_s}(X_{st} - \theta_s)^2 \\ &= -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{s=1}^{k}\sum_{t=1}^{n_s}(X_{st} - \bar{X}_s)^2 - \frac{1}{2\sigma^2}\sum_{s=1}^{k} n_s(\bar{X}_s - \theta_s)^2. \end{aligned}$$

Hence, for any $\sigma^2 > 0$, a maximizer of $l(\boldsymbol{\theta}, \sigma^2; \boldsymbol{X})$ on Θ is equivalent to a minimizer of

$$H(\boldsymbol{\theta}; \bar{\boldsymbol{X}}) = \sum_{s=1}^{k} n_s(\bar{X}_s - \theta_s)^2$$

on Θ. In other words, the MLE $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)'$ of $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmin}}\, H(\boldsymbol{\theta}; \bar{\boldsymbol{X}}). \tag{2.8}$$

We would like to note that the MLE $\hat{\boldsymbol{\theta}}$ can be written by using (2.6) as $\boldsymbol{\eta}_k^{(\boldsymbol{n})}(\bar{\boldsymbol{X}}) = \hat{\boldsymbol{\theta}}$. Here, we put $\bar{\boldsymbol{X}} = \boldsymbol{x} = (x_1, \ldots, x_k)'$. Then, from Lemma 2.1, there exists a unique integer α with $1 \leq \alpha \leq k$ and a unique set J with $J \in \mathcal{J}_\alpha^{(k)}$ such that

$$\boldsymbol{D}_J \boldsymbol{x}_J \geq \boldsymbol{0}, \quad {}^\forall\beta \in \mathbb{N}_k \setminus J,\ \bar{x}_J < x_\beta.$$

For this set J, it holds that

$$\begin{aligned} &{}^\forall w \in J, \quad \hat{\theta}_w = \bar{x}_J = \frac{\sum_{c\in J} n_c x_c}{\sum_{c\in J} n_c} = \frac{\sum_{c\in J} n_c \bar{X}_c}{\sum_{c\in J} n_c}, \\ &{}^\forall\beta \in \mathbb{N}_k \setminus J, \quad \hat{\theta}_\beta = x_\beta = \bar{X}_\beta. \end{aligned} \tag{2.9}$$

Therefore, the MLE $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \ldots, \hat{\mu}_{k^*})'$ of $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{k^*})'$ can be written as

$${}^\forall j \in Q_s, \quad \hat{\mu}_j = \hat{\theta}_s, \qquad (s = 1, \ldots, k). \tag{2.10}$$

On the other hand, the MLE $\hat{\sigma}^2$ of σ^2 can be written as

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{N}\sum_{s=1}^{k}\sum_{t=1}^{n_s}(X_{st} - \bar{X}_s)^2 + \frac{1}{N}\sum_{s=1}^{k} n_s(\bar{X}_s - \hat{\theta}_s)^2 \\ &= \frac{1}{N}\sum_{s=1}^{k}\sum_{t=1}^{n_s}(X_{st} - \hat{\theta}_s)^2 = \frac{1}{N}\sum_{i=1}^{k^*}\sum_{j=1}^{N_i}(Y_{ij} - \hat{\mu}_i)^2, \end{aligned} \tag{2.11}$$

because the function $l(\hat{\boldsymbol{\theta}}, \sigma^2; \boldsymbol{X})$ is a concave function with respect to (w.r.t.) σ^2.

3. Akaike information criterion for the candidate model

In this section, we derive an asymptotically unbiased AIC for the candidate model $\mathcal{M}$. Here, we assume the following two conditions:

(C1) The inequality $N - k - 6 > 0$ holds.

(C2) For the true parameters $\mu_{1,*}, \dots, \mu_{k^*,*}$, it holds that

$$1 \leq^{\forall} s \leq k, \ {}^{\forall}u_1, u_2 \in Q_s, \quad \mu_{u_1,*} = \mu_{u_2,*}$$

and

$$^{\forall}t \in \mathbb{N}_k \setminus \{1\}, \ {}^{\forall}\nu \in Q_t, \quad \mu_{q,*} \leq \mu_{\nu,*},$$

where $q \in Q_1$.

Hence, the condition (C2) means that the true model is included in the candidate model. In addition, for any integer s with $1 \leq s \leq k$ and for any integer u with $u \in Q_s$, we put $\mu_{u,*} = \theta_{s,*}$.

Next, we define a risk function. Let $\boldsymbol{Y}_* = (\boldsymbol{Y}'_{1,*}, \dots, \boldsymbol{Y}_{k^*,*})'$ be a random vector, and let $\boldsymbol{Y}_*$ be independent and identically distributed as $\boldsymbol{Y}$. Furthermore, for any integer s with $1 \leq s \leq k$ and for all elements $q_1^{(s)}, \dots, q_v^{(s)}$ of Q_s, we define $\boldsymbol{X}_{s,*} = (\boldsymbol{Y}'_{q_1^{(s)},*}, \dots, \boldsymbol{Y}'_{q_v^{(s)},*})'$. In addition, we put $\boldsymbol{X}_* = (\boldsymbol{X}'_{1,*}, \dots, \boldsymbol{X}'_{k,*})'$. Here, using the log-likelihood $l(\boldsymbol{\mu}, \sigma^2; \boldsymbol{Y}_*)$ of $\boldsymbol{Y}_*$, we define the following risk function R_1:

$$\begin{aligned} R_1 &= \mathrm{E}[\mathrm{E}_{\boldsymbol{Y}_*}[-2l(\hat{\boldsymbol{\mu}}, \hat{\sigma}^2; \boldsymbol{Y}_*)]] \\ &= \mathrm{E}\left[N \log(2\pi\hat{\sigma}^2) + \frac{N\sigma_*^2}{\hat{\sigma}^2} + \frac{\sum_{i=1}^{k^*} N_i(\mu_{i,*} - \hat{\mu}_i)^2}{\hat{\sigma}^2}\right]. \end{aligned} \tag{3.1}$$

Note that $-2\times$ the maximum log-likelihood is given by

$$-2l(\hat{\boldsymbol{\mu}}, \hat{\sigma}^2; \boldsymbol{Y}) = N \log(2\pi\hat{\sigma}^2) + N. \tag{3.2}$$

By using $-2l(\hat{\boldsymbol{\mu}}, \hat{\sigma}^2; \boldsymbol{Y})$, we estimate the risk function R_1. A bias B_1, which is the difference between the expected value of $-2l(\hat{\boldsymbol{\mu}}, \hat{\sigma}^2; \boldsymbol{Y})$ and R_1, can be expressed as

$$\begin{aligned} B_1 = \mathrm{E}[R_1 - \{-2l(\hat{\boldsymbol{\mu}}, \hat{\sigma}^2; \boldsymbol{Y})\}] &= \mathrm{E}\left[\frac{N\sigma_*^2}{\hat{\sigma}^2}\right] + \mathrm{E}\left[\frac{\sum_{i=1}^{k^*} N_i(\mu_{i,*} - \hat{\mu}_i)^2}{\hat{\sigma}^2}\right] - N \\ &= \mathrm{E}\left[\frac{N\sigma_*^2}{\hat{\sigma}^2}\right] + \mathrm{E}\left[\frac{\sum_{s=1}^{k} n_s(\theta_{s,*} - \hat{\theta}_s)^2}{\hat{\sigma}^2}\right] - N. \end{aligned}$$

Next, we evaluate B_1. Define

$$S = \frac{1}{\sigma_*^2} \sum_{s=1}^{k} \sum_{t=1}^{n_s} (X_{st} - \bar{X}_s)^2, \quad T = \frac{1}{\sigma_*^2} \sum_{s=1}^{k} n_s(\bar{X}_s - \hat{\theta}_s)^2.$$

Note that S and $\bar{\boldsymbol{X}}$ are independent, and S is distributed as the chi-squared distribution with $N - k$ degrees of freedom because X_{st}'s are independently distributed as normal distribution and the condition (C2) holds. Furthermore, from (2.9), since $\hat{\boldsymbol{\theta}}$ is a function of $\bar{\boldsymbol{X}}$, the statistic T is also a function of $\bar{\boldsymbol{X}}$. Hence, S and T are also independent. From (2.11), using S and T we can write

$N\hat{\sigma}^2/\sigma_*^2 = S + T$. Therefore, by using these results and the same technique given by Inatsu (2016), we obtain

$$B_1 = 2(k+1) - \frac{2N}{N-k-2}\mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{s=1}^{k} n_s(\bar{X}_s - \theta_{s,*})(\bar{X}_s - \hat{\theta}_s)\right] + O(N^{-1}). \tag{3.3}$$

Next, we calculate the expectation in (3.3). Here, the following theorem holds:

Theorem 3.1. Let l be an integer with $l \geq 2$. Let $n_1, \dots, n_l$ and τ^2 be positive numbers, and let $\xi_1, \dots, \xi_l$ be real numbers. Let $x_1, \dots, x_l$ be independent random variables, and let $x_s \sim N(\xi_s, \tau^2/n_s)$, $(s = 1, \dots, l)$. Put $\boldsymbol{n} = (n_1, \dots, n_l)'$, $\boldsymbol{\xi} = (\xi_1, \dots, \xi_l)'$ and $\boldsymbol{x} = (x_1, \dots, x_l)'$. Then, it holds that

$$\begin{aligned}
&\mathrm{E}\left[\frac{1}{\tau^2}\sum_{s=1}^{l} n_s(x_s - \xi_s)(x_s - \eta_l^{(\boldsymbol{n})}(\boldsymbol{x})[s])\right] \\
&= \sum_{i=2}^{l}(i-1)\mathrm{P}\left(\boldsymbol{\eta}_l(\boldsymbol{x}) \in \bigcup_{J \in \mathcal{J}_i^l} A^{(l)}(J)\right).
\end{aligned}$$

Details of the proof of Theorem 3.1 are given in Appendix 2 and 3.

Note that $\bar{X}_1, \dots, \bar{X}_k$ are mutually independent, and $\bar{X}_s \sim N(\theta_{s,*}, \sigma_*^2/n_s)$ for any integer s with $1 \leq s \leq k$. Also note that from (2.8) the MLE $\hat{\boldsymbol{\theta}}$ is given by $\hat{\boldsymbol{\theta}} = \boldsymbol{\eta}_k^{(\boldsymbol{n})}(\bar{\boldsymbol{X}})$. Therefore, from Theorem 3.1, the expectation in (3.3) can be expressed as

$$\begin{aligned}
\mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{s=1}^{k} n_s(\bar{X}_s - \theta_{s,*})(\bar{X}_s - \hat{\theta}_s)\right] &= \mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{s=1}^{k} n_s(\bar{X}_s - \theta_{s,*})(\bar{X}_s - \eta_k^{(\boldsymbol{n})}(\bar{\boldsymbol{X}})[s])\right] \\
&= \sum_{u=2}^{k}(u-1)\mathrm{P}\left(\hat{\boldsymbol{\theta}} \in \bigcup_{J \in \mathcal{J}_u^k} A^{(k)}(J)\right) = L, \ \text{(say)}.
\end{aligned}$$

Thus, since $L = O(1)$, we obtain

$$B_1 = 2(k+1) - \frac{2N}{N-k-2}L + O(N^{-1}) = 2(k+1) - 2L + O(N^{-1}).$$

Hence, in order to correct the bias, it is sufficient to add $2(k+1) - 2L$ to $-2l(\hat{\boldsymbol{\mu}}, \hat{\sigma}^2; \boldsymbol{Y})$. However, it is easily checked that L depends on the true parameters $\theta_{1,*}, \dots, \theta_{k,*}$ and σ_*^2. For this reason, we must estimate L. Here, we define the following random variable $\hat{m}$:

$$\hat{m} = 1 + \sum_{a=2}^{k} 1_{\{\hat{\theta}_1 < \hat{\theta}_a\}}. \tag{3.4}$$

It is clear that $\hat{m}$ is a discrete random variable and its possible values are 1 to k. Incidentally, from the definitions of $A^{(k)}(J)$, $\hat{m}$ and $\hat{\boldsymbol{\theta}}$, it holds that

$$\hat{\boldsymbol{\theta}} \in \bigcup_{J \in \mathcal{J}_u^k} A^{(k)}(J) \Longleftrightarrow \hat{m} = k + 1 - u \Longleftrightarrow k - \hat{m} = u - 1,$$

for any integer u with $1 \leq u \leq k$. Therefore, the random variable $k - \hat{m}$ satisfies

$$\mathrm{E}[k - \hat{m}] = \sum_{u=2}^{k}(u-1)\mathrm{P}\left(\hat{\boldsymbol{\theta}} \in \bigcup_{J \in \mathcal{J}_u^k} A^{(k)}(J)\right) = L.$$

Hence, in order to correct the bias, instead of $2(k+1)-2l$, we add

$$2(k+1)-2(k-\hat{m})=2(\hat{m}+1)$$

to $-2l(\hat{\boldsymbol{\mu}},\hat{\sigma}^2;\boldsymbol{Y})$. As a result, we obtain Akaike information criterion for the candidate model $\mathcal{M}$ with the TO, called $\mathrm{AIC}_{\mathrm{TO}}$.

Theorem 3.2. Let $l(\hat{\boldsymbol{\mu}},\hat{\sigma}^2;\boldsymbol{Y})$ be the maximum log-likelihood given by (3.2), and let $\hat{m}$ be a random variable given by (3.4). Then, Akaike information criterion for the candidate model $\mathcal{M}$ with the TO, called $\mathrm{AIC}_{\mathrm{TO}}$ is defined as

$$\mathrm{AIC}_{\mathrm{TO}}:=-2l(\hat{\boldsymbol{\mu}},\hat{\sigma}^2;\boldsymbol{Y})+2(\hat{m}+1).$$

Furthermore, for the risk function R_1 defined by (3.1), it holds that

$$\mathrm{E}[\mathrm{AIC}_{\mathrm{TO}}]=R_1+O(N^{-1}).$$

4. C_p criterion for the candidate model

In this section, we derive an unbiased C_p criterion for the candidate model $\mathcal{M}$. Here, we assume the following condition:

(C1*) The inequality $N-k^*-2>0$ holds.

Hence, we do not assume that the true model is included in the candidate model. First, we consider the risk function based on the prediction mean squared error (PMSE). The risk function R_2 based on the PMSE is given by

$$R_2=\mathrm{E}\left[\mathrm{E}_{\boldsymbol{Y}_*}\left[\frac{1}{\sigma_*^2}\sum_{i=1}^{k^*}\sum_{j=1}^{N_i}(Y_{ij,*}-\hat{\mu}_i)^2\right]\right]=N+\mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{i=1}^{k^*}N_i(\mu_{i,*}-\hat{\mu}_i)^2\right]. \tag{4.1}$$

Next, we define the following random variables:

$$\bar{Y}_i=\frac{1}{N_i}\sum_{j=1}^{N_i}Y_{ij}\quad(i=1,\dots,k^*),\quad \bar{\sigma}^2=\frac{1}{N}\sum_{i=1}^{k^*}\sum_{j=1}^{N_i}(Y_{ij}-\bar{Y}_i)^2. \tag{4.2}$$

Note that $\bar{Y}_1,\dots,\bar{Y}_{k^*}$ and $\bar{\sigma}^2$ are mutually independent, and $\bar{Y}_i\sim N(\mu_{i,*},\sigma_*^2/N_i)$ and $N\bar{\sigma}^2/\sigma_*^2\sim\chi^2_{N-k^*}$ because $Y_{11},\dots,Y_{kN_k}$ are independently distributed as normal distribution. Then, we estimate the risk function R_2 by using

$$(N-k^*-2)\frac{\hat{\sigma}^2}{\bar{\sigma}^2}. \tag{4.3}$$

Here, from (2.11) the MLE $\hat{\sigma}^2$ can be written as

$$\hat{\sigma}^2=\frac{1}{N}\sum_{i=1}^{k^*}\sum_{j=1}^{N_i}(Y_{ij}-\bar{Y}_i)^2+\frac{1}{N}\sum_{i=1}^{k^*}N_i(\bar{Y}_i-\hat{\mu}_i)^2=\bar{\sigma}^2+\frac{1}{N}\sum_{i=1}^{k^*}N_i(\bar{Y}_i-\hat{\mu}_i)^2. \tag{4.4}$$

Therefore, (4.3) can be expressed as

$$(N-k^*-2)\frac{\hat{\sigma}^2}{\bar{\sigma}^2}=N-k^*-2+\left(\frac{N-k^*-2}{N\bar{\sigma}^2/\sigma_*^2}\right)\frac{1}{\sigma_*^2}\sum_{i=1}^{k^*}N_i(\bar{Y}_i-\hat{\mu}_i)^2. \tag{4.5}$$

On the other hand, from (2.9) and (2.10), it can be seen that $\hat{\mu}_1, \ldots, \hat{\mu}_{k^*}$ are functions of $\bar{X}_1, \ldots, \bar{X}_k$. Moreover, for any integer s with $1 \le s \le k$, it holds that

$$\bar{X}_s = \frac{1}{n_s}\sum_{t=1}^{n_s} X_{st} = \frac{1}{\sum_{q\in Q_s} N_q}\sum_{q\in Q_s}\sum_{j=1}^{N_q} Y_{qj} = \frac{1}{\sum_{q\in Q_s} N_q}\sum_{q\in Q_s} N_q \bar{Y}_q. \tag{4.6}$$

Thus, $\bar{X}_1, \ldots, \bar{X}_k$ are functions of $\bar{Y}_1, \ldots, \bar{Y}_{k^*}$, and $\hat{\mu}_1, \ldots, \hat{\mu}_{k^*}$ are also functions of $\bar{Y}_1, \ldots, \bar{Y}_{k^*}$. Hence, noting that $\bar{Y}_1, \ldots, \bar{Y}_{k^*}$ and $\bar{\sigma}^2$ are independent, and $N\bar{\sigma}^2/\sigma_*^2 \sim \chi^2_{N-k^*}$ and $\mathrm{E}[(\chi^2_{N-k^*})^{-1}] = (N-k^*-2)^{-1}$, the expectation of (4.5) can be written as

$$\begin{aligned}
\mathrm{E}\left[(N-k^*-2)\frac{\hat{\sigma}^2}{\bar{\sigma}^2}\right] &= N-k^*-2+\mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{i=1}^{k^*} N_i\{(\bar{Y}_i-\mu_{i,*})+(\mu_{i,*}-\hat{\mu}_i)\}^2\right] \\
&= N-2+2\mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{i=1}^{k^*} N_i(\bar{Y}_i-\mu_{i,*})(\mu_{i,*}-\hat{\mu}_i)\right]+\mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{i=1}^{k^*} N_i(\mu_{i,*}-\hat{\mu}_i)^2\right] \\
&= N-2-2\mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{i=1}^{k^*} N_i(\bar{Y}_i-\mu_{i,*})\hat{\mu}_i\right]+\mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{i=1}^{k^*} N_i(\mu_{i,*}-\hat{\mu}_i)^2\right].
\end{aligned} \tag{4.7}$$

Therefore, by using (4.1) and (4.7), the bias B_2 which is the difference between the expected value of (4.3) and R_2, is given by

$$\begin{aligned}
B_2 = \mathrm{E}\left[R_2-(N-k^*-2)\frac{\hat{\sigma}^2}{\bar{\sigma}^2}\right] &= 2+2\mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{i=1}^{k^*} N_i(\bar{Y}_i-\mu_{i,*})\hat{\mu}_i\right] \\
&= 2+2\mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{s=1}^{k}\sum_{q\in Q_s} N_q(\bar{Y}_q-\mu_{q,*})\hat{\mu}_q\right].
\end{aligned} \tag{4.8}$$

Here, for any integer s with $1 \le s \le k$, we put

$$\frac{\sum_{q\in Q_s} N_q\mu_{q,*}}{\sum_{q\in Q_s} N_q} = \frac{\sum_{q\in Q_s} N_q\mu_{q,*}}{n_s} \equiv \alpha_{s,*}. \tag{4.9}$$

Then, combining (2.10), (4.6) and (4.9), (4.8) can be expressed as

$$\begin{aligned}
B_2 &= 2+2\mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{s=1}^{k} n_s(\bar{X}_s-\alpha_{s,*})\hat{\theta}_s\right] \\
&= 2-2\mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{s=1}^{k} n_s(\bar{X}_s-\alpha_{s,*})(\bar{X}_s-\hat{\theta}_s)\right]+2\mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{s=1}^{k} n_s(\bar{X}_s-\alpha_{s,*})\bar{X}_s\right].
\end{aligned}$$

Hence, noting that $\bar{X}_s \sim N(\alpha_{s,*}, \sigma^2/n_s)$, we have

$$B_2 = 2(k+1)-2\mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{s=1}^{k} n_s(\bar{X}_s-\alpha_{s,*})(\bar{X}_s-\hat{\theta}_s)\right].$$

Furthermore, by using the same argument as in Section 3, we get

$$\mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{s=1}^{k} n_s(\bar{X}_s-\alpha_{s,*})(\bar{X}_s-\hat{\theta}_s)\right] = \mathrm{E}[k-\hat{m}],$$

where $\hat{m}$ is given by (3.4). Thus, it is clear that

$$B_2 = 2(k+1)-2\mathrm{E}[k-\hat{m}] = \mathrm{E}[2(\hat{m}+1)].$$

This implies that in order to correct the bias, it is sufficient to add $2(\hat{m}+1)$ (instead of B_2) to (4.3). As a result, we obtain the C_p criterion for the candidate model $\mathcal{M}$ with the TO, called TOC_p.

Theorem 4.1. A C_p criterion for the candidate model $\mathcal{M}$ with the TO, called TOC_p is defined as

$$\mathrm{TO}C_p := (N - k^* - 2)\frac{\hat{\sigma}^2}{\bar{\sigma}^2} + 2(\hat{m} + 1),$$

where $\hat{\sigma}^2$, $\bar{\sigma}^2$ and $\hat{m}$ are given by (2.11), (4.2) and (3.4), respectively. Moreover, for the risk function R_2 given by (4.1), it holds that

$$\mathrm{E}[\mathrm{TO}C_p] = R_2.$$

Remark 4.1. The TOC_p is the unbiased estimator of R_2. Furthermore, unbiasedness of the TOC_p holds even if the true model is not included in the candidate model $\mathcal{M}$.

In addition, for unbiasedness of the TOC_p, the following theorem holds:

Theorem 4.2. The TOC_p is the uniformly minimum-variance unbiased estimator (UMVUE) of R_2.

Proof. As we mentioned before, the random variable $\hat{m}$ is a function of $\hat{\theta}_1, \ldots, \hat{\theta}_k$, and $\hat{\theta}_1, \ldots, \hat{\theta}_k$ are functions of $\bar{X}_1, \ldots, \bar{X}_k$. Furthermore, $\bar{X}_1, \ldots, \bar{X}_k$ are functions of $\bar{Y}_1, \ldots, \bar{Y}_{k^*}$. Thus, $\hat{m}$ is a function of $\bar{Y}_1, \ldots, \bar{Y}_{k^*}$. On the other hand, since $\hat{\mu}_1, \ldots, \hat{\mu}_{k^*}$ are functions of $\bar{Y}_1, \ldots, \bar{Y}_{k^*}$, from (4.4), we can see that both $\hat{\sigma}^2$ and $\bar{\sigma}^2$ are functions of $\bar{Y}_1, \ldots, \bar{Y}_{k^*}$. Therefore, from the definition of the TOC_p, the TOC_p is a function of $\bar{\sigma}^2$ and $\bar{Y}_1, \ldots, \bar{Y}_{k^*}$. Incidentally, noting that $Y_{11}, \ldots, Y_{k^* N_{k^*}}$ are mutually independent, and $Y_{ij} \sim N(\mu_{i,*}, \sigma_*^2)$ where $1 \le i \le k^*$ and $1 \le j \le N_i$, the joint distribution function $f(\boldsymbol{y}; \boldsymbol{\mu}_*, \sigma_*^2)$ can be written as

$$\begin{aligned} &f(\boldsymbol{y}; \boldsymbol{\mu}_*, \sigma_*^2) \\ &= \frac{1}{(2\pi\sigma_*^2)^{N/2}} \exp\left\{ -\frac{1}{2\sigma_*^2} \sum_{i=1}^{k^*} \left(N_i \bar{y}_i^2 + \sum_{j=1}^{N_i} (y_{ij} - \bar{y}_i)^2 \right) + \sum_{i=1}^{k^*} \frac{N_i \mu_{i,*}}{\sigma_*^2} \bar{y}_i - C \right\}, \end{aligned}$$

where $\bar{y}_i$ and C are given by

$$\bar{y}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} y_{ij}, \quad C = \frac{1}{2\sigma_*^2} \sum_{i=1}^{k^*} N_i \mu_{i,*}^2.$$

Here, define

$$T_0 = \sum_{i=1}^{k^*} \left(N_i \bar{Y}_i^2 + \sum_{j=1}^{N_i} (Y_{ij} - \bar{Y}_i)^2 \right), \quad T_i = \bar{Y}_i, \quad (i = 1, \ldots, k^*).$$

Then, $(T_0, T_1, \ldots, T_{k^*})'$ is a complete sufficient statistic (see, e.g., Lehmann and Casella, 1998). Moreover, since $\bar{\sigma}^2$ can be written by using $(T_0, T_1, \ldots, T_{k^*})'$ as

$$\bar{\sigma}^2 = \frac{1}{N} \left(T_0 - \sum_{i=1}^{k^*} N_i T_i^2 \right),$$

$\bar{\sigma}^2$ is a function of the complete sufficient statistic $(T_0, T_1, \ldots, T_{k^*})'$. Hence, the TO$C_p$ which is a function of $\bar{\sigma}^2$ and $\bar{Y}_1, \ldots, \bar{Y}_{k^*}$, is also a function of the complete sufficient statistic. Therefore, since the TOC_p is the unbiased estimator of R_2, from Lehmann-Scheffé theorem (see, e.g., Knight, 1999), the TOC_p is the UMVUE of R_2. □

5. Numerical experiments

In this section, we confirm estimation accuracies of the $\mathrm{AIC}_{\mathrm{TO}}$ and the $\mathrm{TO}C_p$ through numerical experiments. Let $X_{ij} \sim N(\theta_i, \sigma^2)$, where $i = 1, 2, 3, 4$ and $j = 1, \ldots, N_i$ for each i. We set $N_1 = N_2 = N_3 = N_4$. Furthermore, we put $N = N_1 + N_2 + N_3 + N_4$. In this setting, we consider the ANOVA model with the following restriction:

$$^{\forall}j \in \{2, 3, 4\}, \quad \theta_1 \leq \theta_j.$$

Hence, in this candidate model, the parameter space Θ is given by

$$\Theta \equiv \{\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)' \in \mathbb{R}^4 \mid {}^{\forall}j \in \{2, 3, 4\}, \quad \theta_1 \leq \theta_j\}.$$

Here, for comparison, we define the following two criteria:

$$\mathrm{fAIC} := -2l(\hat{\boldsymbol{\mu}}, \hat{\sigma}^2; \boldsymbol{Y}) + 2(4+1),$$
$$\mathrm{f}C_p := (N - k^* - 2)\frac{\hat{\sigma}^2}{\bar{\sigma}^2} + 2(4+1).$$

Thus, the penalty term of both the fAIC and the fC_p is $2\times$ the number of parameters. Note that since the parameters are restricted, the fAIC and the fC_p are not necessary (asymptotically) unbiased estimators of risk functions in general.

Next, in this numerical experiments, we consider the following true parameters:

$$\begin{aligned}
&\text{Case 1}: \ \theta_1 = 1, \ \theta_2 = 2, \ \theta_3 = 3, \ \theta_4 = 4, \ \sigma^2 = 1,\\
&\text{Case 2}: \ \theta_1 = 1, \ \theta_2 = 1.05, \ \theta_3 = 1.05, \ \theta_4 = 1.05, \ \sigma^2 = 1,\\
&\text{Case 3}: \ \theta_1 = 1, \ \theta_2 = 1, \ \theta_3 = 1, \ \theta_4 = 1, \ \sigma^2 = 1,\\
&\text{Case 4}: \ \theta_1 = 2, \ \theta_2 = 1.4, \ \theta_3 = 0.8, \ \theta_4 = 0.2, \ \sigma^2 = 1.
\end{aligned}$$

We would like to note that the vector of true parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_4)'$ is an interior point of Θ in Case 1. Similarly, in Case 2, $\boldsymbol{\theta}$ is an interior point of Θ, but $\boldsymbol{\theta}$ is very close to the boundary. On the other hand, $\boldsymbol{\theta}$ is a boundary point of Θ in Case 3. Moreover, in Case 4, $\boldsymbol{\theta}$ is not included in Θ. Therefore, the true model is included in the candidate model when Case 1–3. However, in Case 4, it is not included. From 1,000,000 Monte Carlo simulation runs, we confirm estimation accuracies of four criteria. Obtained results are given in Table 5.1 – 5.4.

Table 5.1 Estimation accuracy of four criteria in Case 1

N	R_1	$\mathrm{AIC}_{\mathrm{TO}}$	fAIC	R_2	$\mathrm{TO}C_p$	fC_p
12	47.08	37.47	37.70	15.82	15.82	16.05
36	108.75	106.75	106.78	39.97	39.97	40.00
100	289.30	288.64	288.64	104.00	104.00	104.00
200	572.81	572.53	572.53	204.00	204.00	204.00
1000	2842.93	2842.87	2842.87	1004.00	1004.00	1004.00
10000	28383.78	28383.90	28383.90	10004.00	10004.00	10004.00

Table 5.2 Estimation accuracy of four criteria in Case 2

N	R_1	$\mathrm{AIC}_{\mathrm{TO}}$	fAIC	R_2	$\mathrm{TO}C_p$	$\mathrm{f}C_p$
12	43.20	37.07	39.15	14.92	14.92	16.99
36	107.19	105.83	107.84	38.92	38.92	40.93
100	288.08	287.63	289.53	102.93	102.94	104.84
200	571.71	571.52	573.30	202.96	202.96	204.74
1000	2842.03	2842.01	2843.33	1003.11	1003.12	1004.44
10000	28383.61	28383.52	28383.73	10003.82	10003.83	10004.03

Table 5.3 Estimation accuracy of four criteria in Case 3

N	R_1	$\mathrm{AIC}_{\mathrm{TO}}$	fAIC	R_2	$\mathrm{TO}C_p$	$\mathrm{f}C_p$
12	43.02	37.12	39.29	14.91	14.91	17.09
36	107.13	105.83	108.01	38.91	38.91	41.09
100	288.04	287.60	289.78	102.91	102.91	105.09
200	571.65	571.47	573.64	202.91	202.91	205.09
1000	2841.82	2841.74	2843.91	1002.91	1002.91	1005.09
10000	28382.69	28382.59	28384.77	10002.91	10002.91	10005.09

Table 5.4 Estimation accuracy of four criteria in Case 4

N	R_1	$\mathrm{AIC}_{\mathrm{TO}}$	fAIC	R_2	$\mathrm{TO}C_p$	$\mathrm{f}C_p$
12	42.67	41.87	45.54	19.26	19.24	22.91
36	117.64	118.52	122.31	53.25	53.26	57.06
100	321.39	322.62	326.48	144.11	144.13	147.99
200	640.13	641.46	645.40	286.06	286.05	289.99
1000	3190.87	3192.31	3196.31	1422.00	1422.04	1426.04
10000	31887.66	31888.95	31892.95	14202.00	14201.98	14205.98

From Table 5.1, we can see that the $\mathrm{AIC}_{\mathrm{TO}}$ and the fAIC are asymptotically unbiased estimators of the risk R_1 in Case 1. Furthermore, the $\mathrm{TO}C_p$ and the $\mathrm{f}C_p$ are unbiased and asymptotically unbiased estimators of R_2, respectively. Similarly, from Table 5.2 we can see that the result of Case 2 is similar to that of Case 1. Nevertheless, estimation accuracies of the fAIC and the $\mathrm{f}C_p$ in Case 2 are not good even if the sample size N is less than 1000. On the other hand, in Case 3, from Table 5.3 we can see that the $\mathrm{AIC}_{\mathrm{TO}}$ is the asymptotically unbiased estimator of R_1 and the fAIC has the asymptotically bias. Similarly, the $\mathrm{TO}C_p$ is the unbiased estimator of R_2 and the $\mathrm{f}C_p$ has the asymptotic bias. Finally, from Table 5.4 we can see that the three criteria $\mathrm{AIC}_{\mathrm{TO}}$, fAIC and $\mathrm{f}C_p$ have asymptotic biases in Case 4. However, the $\mathrm{TO}C_p$ is the unbiased estimator of R_2.

6. Conclusion

Under ANOVA model with the tree ordering, we derived the asymptotically unbiased AIC and the unbiased C_p criterion, called $\mathrm{AIC}_{\mathrm{TO}}$ and $\mathrm{TO}C_p$, respectively. In particular, the $\mathrm{TO}C_p$ is the unbiased estimator even if the true model is not included in the set of candidate models. Moreover, we show that the $\mathrm{TO}C_p$ is the UMVUE. We confirmed these results through numerical experiments.

Appendix 1: Proof of Lemma 2.1

In this section, we prove Lemma 2.1. First, we provide the following lemma.

Lemma A. The following three propositions hold:

(1) Let A and B be non-empty subsets of $\mathbb{N}_l$, and let $A \cap B = \emptyset$. Then, it holds that

$$\bar{x}_A < \bar{x}_B \Rightarrow \bar{x}_A < \bar{x}_{A\cup B} < \bar{x}_B.$$

(2) Let A and $B_1, \ldots, B_i$ be non-empty subsets of $\mathbb{N}_l$, and let A and $B_1, \ldots, B_i$ be disjoint. Then, it holds that

$$^{\forall} j \in \{1, \ldots, i\},\ \bar{x}_A < \bar{x}_{B_j} \Rightarrow \bar{x}_A < \bar{x}_B, \tag{A.1}$$

where B is given by

$$B = \bigcup_{j=1}^{i} B_j.$$

Similarly, it also holds that

$$^{\forall} j \in \{1, \ldots, i\},\ \bar{x}_{B_j} \leq \bar{x}_A \Rightarrow \bar{x}_B \leq \bar{x}_A. \tag{A.2}$$

(3) Let A, B and C be non-empty subsets of $\mathbb{N}_l$, and let A, B and C be disjoint. Then, it holds that

$$\bar{x}_A < \bar{x}_C,\ \bar{x}_B \leq \bar{x}_C \Rightarrow \bar{x}_{A\cup B} < \bar{x}_C. \tag{A.3}$$

Proof. First, we prove (1). Let A and B be non-empty and disjoint subsets of $\mathbb{N}_l$, and let $\bar{x}_A < \bar{x}_B$. Then, multiplying both sides by $\tilde{N}_B = \sum_{b\in B} N_b$, we get

$$\tilde{N}_B \bar{x}_A < \tilde{N}_B \bar{x}_B = \sum_{b\in B} N_b x_b.$$

Furthermore, adding $\tilde{N}_A \bar{x}_A$ to both sides we have

$$(\tilde{N}_A + \tilde{N}_B)\bar{x}_A < \tilde{N}_A \bar{x}_A + \sum_{b\in B} N_b x_b = \sum_{a\in A} N_a x_a + \sum_{b\in B} N_b x_b.$$

In addition, dividing this inequality by $\tilde{N}_A + \tilde{N}_B = \sum_{a\in A} N_a + \sum_{b\in B} N_b$ we obtain

$$\bar{x}_A < \frac{\sum_{a\in A} N_a x_a + \sum_{b\in B} N_b x_b}{\sum_{a\in A} N_a + \sum_{b\in B} N_b}.$$

Here, recall that A and B are disjoint. Therefore, it holds that

$$\frac{\sum_{a\in A} N_a x_a + \sum_{b\in B} N_b x_b}{\sum_{a\in A} N_a + \sum_{b\in B} N_b} = \frac{\sum_{s\in A\cup B} N_s x_s}{\sum_{s\in A\cup B} N_s} = \bar{x}_{A\cup B}.$$

Hence, $\bar{x}_A < \bar{x}_{A\cup B}$ holds. By using the same argument, we can also prove that $\bar{x}_{A\cup B} < \bar{x}_B$ holds. Thus, the proposition (1) holds.

Next, we prove (2). Let A and $B_1, \dots, B_i$ be non-empty and disjoint subsets of $\mathbb{N}_l$. Assume that $\bar{x}_A < \bar{x}_{B_j}$ for any integer j with $1 \le j \le i$. Here, multiplying both sides of $\bar{x}_A < \bar{x}_{B_1}$ by $\tilde{N}_{B_1}$, we have

$$\tilde{N}_{B_1}\bar{x}_A < \tilde{N}_{B_1}\bar{x}_{B_1} = \sum_{s\in B_1} N_s x_s,$$

and multiplying both sides of $\bar{x}_A < \bar{x}_{B_2}$ by $\tilde{N}_{B_2}$, we get

$$\tilde{N}_{B_2}\bar{x}_A < \tilde{N}_{B_2}\bar{x}_{B_2} = \sum_{t\in B_2} N_t x_t.$$

Thus, using these two inequalities we obtain

$$\bar{x}_A < \frac{\sum_{s\in B_1} N_s x_s + \sum_{t\in B_2} N_t x_t}{\tilde{N}_{B_1} + \tilde{N}_{B_2}} = \frac{\sum_{s\in B_1} N_s x_s + \sum_{t\in B_2} N_t x_t}{\sum_{s\in B_1} N_s + \sum_{t\in B_2} N_t}.$$

Moreover, noting that B_1 and B_2 are disjoint, we get

$$\frac{\sum_{s\in B_1} N_s x_s + \sum_{t\in B_2} N_t x_t}{\sum_{s\in B_1} N_s + \sum_{t\in B_2} N_t} = \frac{\sum_{u\in B_1\cup B_2} N_u x_u}{\sum_{u\in B_1\cup B_2} N_u} = \bar{x}_{B_1\cup B_2}.$$

Hence, $\bar{x}_A < \bar{x}_{B_1\cup B_2}$ holds. Here, we put $B_1 \cup B_2 = C$. Then, it holds that $\bar{x}_A < \bar{x}_C$. From this inequality and $\bar{x}_A < \bar{x}_{B_3}$, using the same argument we obtain $\bar{x}_A < \bar{x}_{C\cup B_3} = \bar{x}_{B_1\cup B_2\cup B_3}$. By repeating this process, we get (A.1). Furthermore, (A.2) and (A.3) can be proved by using the same argument. Thus, the propositions (2) and (3) are proved. □

Next, we prove Lemma 2.1.

Proof. When $l = 2$, the statements of Lemma 2.1 are equivalent to Lemma C given by Inatsu (2016), and it is already proved. Therefore, we prove the case of $l \ge 3$.

First, we prove (1) of Lemma 2.1. From (2.5) it holds that

$$\bigcup_{i=1}^{l} \bigcup_{J\in\mathcal{J}_i^{(l)}} A^{(l)}(J) = \{\boldsymbol{x} \in \mathbb{R}^l \mid x_1 \le x_2, \dots, x_1 \le x_l\} = A^{(l)},$$

and $A^{(l)}(J) \ne A^{(l)}(J^*)$ where $J \ne J^*$. Therefore, from the definition of the inverse image, it is clear that (1) holds because $\boldsymbol{\eta}_l$ is the function from $\mathbb{R}^l$ to $A^{(l)}$.

Next, using mathematical induction we prove (2) and (3) of Lemma 2.1. Thus, assume that Lemma 2.1 is true when $l = 2, \dots, q-1$. In this assumption, we prove that Lemma 2.1 is also true when $l = q$. Here, in the case of $i = 1$, $\mathcal{J}_1^{(q)}$ has only one set $J = \{1\}$. First, for this set J, we show the inclusion relation $\supset$ of (2.7). Let $\boldsymbol{x} = (x_1, \dots, x_q)'$ be an element of $\mathbb{R}^q$ satisfying

$$\boldsymbol{D}_J\boldsymbol{x}_J \ge \boldsymbol{0}, \ {}^\forall t \in \mathbb{N}_q \setminus J, \ \bar{x}_J < x_t.$$

Here, note that $\bar{x}_J = x_1$. Hence, for any integer t with $2 \le t \le q$, the inequality $x_1 < x_t$ holds. This implies that $\boldsymbol{x} \in A^{(q)}(J) \subset A^{(q)}$. Meanwhile, let

$$H_q(\boldsymbol{\delta}; \boldsymbol{x}) = \sum_{u=1}^{q} N_u(x_u - \delta_u)^2.$$

Then, noting that $\boldsymbol{x} \in A^{(q)}$, we get

$$0 \leq \min_{\boldsymbol{\delta} \in A^{(q)}} H_q(\boldsymbol{\delta}; \boldsymbol{x}) \leq H_q(\boldsymbol{x}; \boldsymbol{x}) = 0.$$

Therefore, it holds that

$$\min_{\boldsymbol{\delta} \in A^{(q)}} H_q(\boldsymbol{\delta}; \boldsymbol{x}) = H_q(\boldsymbol{x}; \boldsymbol{x}) = 0.$$

This equality means that $\boldsymbol{\eta}_q(\boldsymbol{x}) = \boldsymbol{x} \in A^{(q)}(J)$. Thus, we obtain $\boldsymbol{\eta}_q(\boldsymbol{x}) \in A^{(q)}(J)$. Therefore, $\boldsymbol{x} \in \boldsymbol{\eta}_q^{-1}\left(A^{(q)}(J)\right)$ holds. Hence, the inclusion relation $\supset$ of (2.7) in the case of $J = \{1\}$ is proved. Next, we show $\subset$ of (2.7). Let $\boldsymbol{y} = (y_1, \ldots, y_q)'$ be an element of $\mathbb{R}^q$ satisfying $\boldsymbol{y} \in \boldsymbol{\eta}_q^{-1}\left(A^{(q)}(J)\right)$. In other words, we assume that

$$\boldsymbol{\eta}_q(\boldsymbol{y}) = \underset{\boldsymbol{\delta} \in A^{(q)}}{\operatorname{argmin}}\, H_q(\boldsymbol{\delta}; \boldsymbol{y}) \equiv \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)' \in A^{(q)}(J).$$

Here, noting that $A^{(q)}(J)$ is an open set, there exists an ε-neighborhood $\mathrm{U}(\boldsymbol{\alpha}; \varepsilon)$ of $\boldsymbol{\alpha}$ such that $\mathrm{U}(\boldsymbol{\alpha}; \varepsilon) \subset A^{(q)}(J)$. Thus, for any element $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)'$ of $\mathbb{R}^q$ satisfying $\boldsymbol{\gamma} \in \mathrm{U}(\boldsymbol{\alpha}; \varepsilon) \subset A^{(q)}$, it holds that

$$H_q(\boldsymbol{\alpha}; \boldsymbol{y}) \leq H_q(\boldsymbol{\gamma}; \boldsymbol{y}).$$

This implies that $\boldsymbol{\alpha}$ is a local minimizer of $H_q(\boldsymbol{\delta}; \boldsymbol{y})$. In addition, since $H_q(\boldsymbol{\delta}; \boldsymbol{y})$ is a strictly convex function on $\mathbb{R}^q$ with respect to (w.r.t.) $\boldsymbol{\delta}$, the local minimizer $\boldsymbol{\alpha}$ is a unique global minimizer. Moreover, it is clear that the global minimizer is $\boldsymbol{y}$ because $H_q(\boldsymbol{\delta}; \boldsymbol{y})$ is non-negative and $H_q(\boldsymbol{y}; \boldsymbol{y}) = 0$. Therefore, we get $\boldsymbol{\alpha} = \boldsymbol{y}$ and it holds that

$$\boldsymbol{\eta}_q(\boldsymbol{y}) = \boldsymbol{\alpha} = \boldsymbol{y} \in A^{(q)}(J).$$

Hence, for any s with $s \in \mathbb{N}_q \setminus J$, the inequality $y_1 < y_s$ holds. Consequently, the inclusion relation $\subset$ of (2.7) in the case of $J = \{1\}$ is proved.

Next, for any i with $2 \leq i \leq q-1$, we prove the inclusion relation $\supset$ of (2.7). Let i be an integer with $2 \leq i \leq q-1$, and let J be a set with $J \in \mathcal{J}_i^{(q)}$. Assume that $\boldsymbol{x} = (x_1, \ldots, x_q)'$ is an element of $\mathbb{R}^q$ satisfying $\boldsymbol{D}_J \boldsymbol{x}_J \geq \boldsymbol{0}$ and $\bar{x}_J < x_t$ for any $t \in \mathbb{N}_q \setminus J$. Here, the function $H_q(\boldsymbol{\alpha}; \boldsymbol{x})$ can be expressed as

$$\begin{aligned} H_q(\boldsymbol{\alpha}; \boldsymbol{x}) &= \sum_{d=1}^{q} N_d (x_d - \alpha_d)^2 = \sum_{s \in J} N_s (x_s - \alpha_s)^2 + \sum_{t \in \mathbb{N}_q \setminus J} N_t (x_t - \alpha_t)^2 \\ &= H_{\#J}(\boldsymbol{\alpha}_J; \boldsymbol{x}_J) + H_{\#\mathbb{N}_q \setminus J}(\boldsymbol{\alpha}_{\mathbb{N}_q \setminus J}; \boldsymbol{x}_{\mathbb{N}_q \setminus J}). \end{aligned}$$

Therefore, it is easily checked that

$$\min_{\boldsymbol{\alpha} \in A^{(q)}} H_q(\boldsymbol{\alpha}; \boldsymbol{x}) \geq \min_{\boldsymbol{\alpha}_J \in A^{(\#J)}} H_{\#J}(\boldsymbol{\alpha}_J; \boldsymbol{x}_J) + H_{\#\mathbb{N}_q \setminus J}(\boldsymbol{x}_{\mathbb{N}_q \setminus J}; \boldsymbol{x}_{\mathbb{N}_q \setminus J}). \tag{A.4}$$

In addition, we put $\boldsymbol{x}_J = (y_1, \ldots, y_{\#J})' = \boldsymbol{y}$, $\boldsymbol{\alpha}_J = (\beta_1, \ldots, \beta_{\#J})' = \boldsymbol{\beta}$, $\boldsymbol{N}_J = (n_1, \ldots, n_{\#J})' = \boldsymbol{n}$ and $J^* = \mathbb{N}_{\#J}$. By using these notations, we obtain

$$H_{\#J}(\boldsymbol{\alpha}_J; \boldsymbol{x}_J) = \sum_{s \in J} N_s (x_s - \alpha_s)^2 = \sum_{u=1}^{\#J} n_u (y_u - \beta_u)^2 = H_{\#J}(\boldsymbol{\beta}; \boldsymbol{y}),$$

and

$$\min_{\boldsymbol{\alpha}_J \in A^{(\#J)}} H_{\#J}(\boldsymbol{\alpha}_J; \boldsymbol{x}_J) = \min_{\boldsymbol{\beta} \in A^{(\#J)}} H_{\#J}(\boldsymbol{\beta}; \boldsymbol{y}).$$

Recall that Lemma 2.1 is true when $l = 2, \ldots, q-1$ from the assumption of mathematical induction. Moreover, it also holds that $\boldsymbol{D}_J^{(\boldsymbol{N})}\boldsymbol{x}_J \geq \boldsymbol{0}$. This inequality is equal to $\boldsymbol{D}_{J^*}^{(\boldsymbol{n})}\boldsymbol{y}_{J^*} \geq \boldsymbol{0}$. Furthermore, noting that $J^* = \mathbb{N}_{\#J}$ and $2 \leq \#J \leq q-1$, from (3) of Lemma 2.1 we get

$$\begin{aligned}\min_{\boldsymbol{\alpha}_J \in A^{(\#J)}} H_{\#J}(\boldsymbol{\alpha}_J; \boldsymbol{x}_J) &= \min_{\boldsymbol{\beta} \in A^{(\#J)}} H_{\#J}(\boldsymbol{\beta}; \boldsymbol{y}) \\ &= \sum_{u=1}^{\#J} n_u (y_u - \bar{y}_{J^*})^2 = \sum_{s \in J} N_s (x_s - \bar{x}_J)^2. \end{aligned} \tag{A.5}$$

Hence, from (A.4) and (A.5), it holds that

$$\min_{\boldsymbol{\alpha} \in A^{(q)}} H_q(\boldsymbol{\alpha}; \boldsymbol{x}) \geq \sum_{s \in J} N_s (x_s - \bar{x}_J)^2 + \sum_{t \in \mathbb{N}_q \setminus J} N_t (x_t - x_t)^2. \tag{A.6}$$

Here, let $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)'$ be a q-dimensional vector whose sth element ($s \in J$) is $\bar{x}_J$ and tth element ($t \in \mathbb{N}_q \setminus J$) is x_t. Then, from the assumption, for any $t \in \mathbb{N}_q \setminus J$ it holds that $\bar{x}_J < x_t$. Thus, from the definition of $\boldsymbol{\gamma}$, we obtain $\boldsymbol{\gamma} \in A^{(q)}$. Hence, the following inequality holds:

$$\min_{\boldsymbol{\alpha} \in A^{(q)}} H_q(\boldsymbol{\alpha}; \boldsymbol{x}) \leq H_q(\boldsymbol{\gamma}; \boldsymbol{x}) = \sum_{s \in J} N_s (x_s - \bar{x}_J)^2 + \sum_{t \in \mathbb{N}_q \setminus J} N_t (x_t - x_t)^2. \tag{A.7}$$

Therefore, from (A.6) and (A.7) we get

$$\min_{\boldsymbol{\alpha} \in A^{(q)}} H_q(\boldsymbol{\alpha}; \boldsymbol{x}) = H_q(\boldsymbol{\gamma}; \boldsymbol{x}).$$

This implies that

$$\boldsymbol{\eta}_q(\boldsymbol{x}) = \operatorname*{argmin}_{\boldsymbol{\alpha} \in A^{(q)}} H_q(\boldsymbol{\alpha}; \boldsymbol{x}) = \boldsymbol{\gamma}.$$

Noting that from the definition of $\boldsymbol{\gamma}$, we have $\boldsymbol{\gamma} \in A^{(q)}(J)$, i.e., $\boldsymbol{x} \in \boldsymbol{\eta}_q^{-1}\left(A^{(q)}(J)\right)$. Consequently, for any i with $2 \leq i \leq q-1$, the inclusion relation $\supset$ of (2.7) is proved.

Next, we prove the inclusion relation $\subset$ of (2.7). Let i be an integer with $2 \leq i \leq q-1$, and let J be a set with $J \in \mathcal{J}_i^{(q)}$. Also let $\boldsymbol{x} = (x_1, \ldots, x_q)'$ be an element of $\mathbb{R}^q$ satisfying $\boldsymbol{x} \in \boldsymbol{\eta}_q^{-1}\left(A^{(q)}(J)\right)$. In other words, we assume that

$$\boldsymbol{\eta}_q(\boldsymbol{x}) = (\alpha_1, \ldots, \alpha_q)' = \boldsymbol{\alpha} \in A^{(q)}(J).$$

Here, from the definition of $A^{(q)}(J)$, for any $s \in J$ and for any $t \in \mathbb{N}_q \setminus J$, it holds that $\alpha_1 = \alpha_s$ and $\alpha_1 < \alpha_t$. Incidentally, from the definition of $\boldsymbol{\eta}_q$, we get

$$\begin{aligned}\min_{\boldsymbol{\delta} \in A^{(q)}} \sum_{i=1}^{q} N_i (x_i - \delta_i)^2 &= \sum_{s \in J} N_s (x_s - \alpha_s)^2 + \sum_{t \in \mathbb{N}_q \setminus J} N_t (x_t - \alpha_t)^2 \\ &= \sum_{s \in J} N_s (x_s - \alpha_1)^2 + \sum_{t \in \mathbb{N}_q \setminus J} N_t (x_t - \alpha_t)^2.\end{aligned}$$

In addition, for the subvector $\boldsymbol{\gamma}^* = (\gamma_1, \boldsymbol{\gamma}'_{\mathbb{N}_q \setminus J})'$, we consider the following function:

$$H(\boldsymbol{\gamma}^*; \boldsymbol{x}) = \sum_{s \in J} N_s (x_s - \gamma_1)^2 + \sum_{t \in \mathbb{N}_q \setminus J} N_t (x_t - \gamma_t)^2.$$

Noting that $\boldsymbol{\alpha}^* = (\alpha_1, \boldsymbol{\alpha}'_{\mathbb{N}_q \setminus J})' \in A^{(q-\#J+1)}(\{1\})$ and $A^{(q-\#J+1)}(\{1\})$ is an open set, there exists an ε-neighborhood $\mathrm{U}(\boldsymbol{\alpha}^*; \varepsilon)$ of $\boldsymbol{\alpha}^*$ such that $\mathrm{U}(\boldsymbol{\alpha}^*; \varepsilon) \subset A^{(q-\#J+1)}(\{1\})$. Let $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_q)'$, and let $\boldsymbol{\zeta}^* = (\zeta_1, \boldsymbol{\zeta}'_{\mathbb{N}_q \setminus J})' \in \mathrm{U}(\boldsymbol{\alpha}^*; \varepsilon)$. Moreover, let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_q)'$ be a q-dimensional vector whose sth element $(s \in J)$ is $\xi_s = \zeta_1$, and tth element $(t \in \mathbb{N}_q \setminus J)$ is $\xi_t = \zeta_t$. Then, noting that $\boldsymbol{\xi} \in A^{(q)}$ we obtain

$$\begin{aligned} H(\boldsymbol{\zeta}^*; \boldsymbol{x}) &= \sum_{s \in J} N_s(x_s - \zeta_1)^2 + \sum_{t \in \mathbb{N}_q \setminus J} N_t(x_t - \zeta_t)^2 = \sum_{s \in J} N_s(x_s - \xi_s)^2 + \sum_{t \in \mathbb{N}_q \setminus J} N_t(x_t - \xi_t)^2 \\ &\geq \min_{\boldsymbol{\delta} \in A^{(q)}} \sum_{i=1}^{q} N_i(x_i - \delta_i)^2 = \sum_{s \in J} N_s(x_s - \alpha_1)^2 + \sum_{t \in \mathbb{N}_q \setminus J} N_t(x_t - \alpha_t)^2 = H(\boldsymbol{\alpha}^*; \boldsymbol{x}). \end{aligned}$$

Thus, $\boldsymbol{\alpha}^*$ is a local minimizer of $H(\boldsymbol{\gamma}^*; \boldsymbol{x})$. In addition, since $H(\boldsymbol{\gamma}^*; \boldsymbol{x})$ is a strictly convex function on $\mathbb{R}^{q - \#J + 1}$ w.r.t. $\boldsymbol{\gamma}^*$, the local minimizer $\boldsymbol{\alpha}^*$ is a unique global minimizer of $H(\boldsymbol{\gamma}^*; \boldsymbol{x})$. Moreover, the global minimizer can be obtained by differentiating $H(\boldsymbol{\gamma}^*; \boldsymbol{x})$ w.r.t. $\boldsymbol{\gamma}^*$ as

$$\alpha_1 = \bar{x}_J, \quad \alpha_t = x_t \ \ (t \in \mathbb{N}_q \setminus J).$$

Therefore, noting that $\alpha_1 < \alpha_t$, we have $\bar{x}_J < x_t$.

Next, we prove $\boldsymbol{D}_J^{(\boldsymbol{N})} \boldsymbol{x}_J \geq \boldsymbol{0}$. We replace $\boldsymbol{x}_J$ and $\boldsymbol{N}_J$ with $\boldsymbol{y} = (y_1, \ldots, y_i)'$ and $\boldsymbol{n} = (n_1, \ldots, n_i)'$, respectively. In addition, we put $J^* = \mathbb{N}_i$. Note that $\boldsymbol{x}_J = \boldsymbol{y} = \boldsymbol{y}_{J^*}$. Also note that $\boldsymbol{y}$ is an i-dimensional vector and $2 \leq i \leq q-1$. Recall that from (1) of Lemma 2.1, it holds that

$$\mathbb{R}^i = \bigcup_{s=1}^{i} \bigcup_{J \in \mathcal{J}_s^{(i)}} \boldsymbol{\eta}_i^{-1}\left(A^{(i)}(J)\right),$$

$$\boldsymbol{\eta}_i^{-1}\left(A^{(i)}(J)\right) \cap \boldsymbol{\eta}_i^{-1}\left(A^{(i)}(J^*)\right) = \emptyset \quad (J \neq J^*).$$

In order to prove $\boldsymbol{D}_J^{(\boldsymbol{N})} \boldsymbol{x}_J \geq \boldsymbol{0}$, we show $\boldsymbol{y} \in \boldsymbol{\eta}_i^{-1}\left(A^{(i)}(\mathbb{N}_i)\right)$ using proof by contradiction. Thus, we assume that there exists an integer s with $1 \leq s \leq i-1$ and a set J^{**} of $\mathcal{J}_s^{(i)}$ such that $\boldsymbol{y} \in \boldsymbol{\eta}_i^{-1}\left(A^{(i)}(J^{**})\right)$. Recall that from the assumption of mathematical induction, Lemma 2.1 is true when $l = 2, \ldots, q-1$. Furthermore, since $i \leq q-1$, from (2) of Lemma 2.1, $\boldsymbol{y} \in \boldsymbol{\eta}_i^{-1}\left(A^{(i)}(J^{**})\right)$ is equivalent to

$$\boldsymbol{D}_{J^{**}}^{(\boldsymbol{n})} \boldsymbol{y}_{J^{**}} \geq \boldsymbol{0}, \quad \bar{y}_{J^{**}} < y_t \ \ (t \in \mathbb{N}_i \setminus J^{**}).$$

Here, by using (2) of Lemma A, we get $\bar{y}_{J^{**}} < \bar{y}_{\mathbb{N}_i \setminus J^{**}}$. Moreover, using (1) of Lemma A we have $\bar{y}_{J^{**}} < \bar{y}_{\mathbb{N}_i} = \bar{x}_J$. Therefore, combining $\bar{x}_J < x_t$ $(t \in \mathbb{N}_q \setminus J)$, we get

$$\bar{y}_{J^{**}} < x_r \quad (r \in \mathbb{N}_q \setminus J). \tag{A.8}$$

Note that there exists a set J^{***} with $J^{***} \subsetneqq J$ satisfies $\bar{y}_{J^{**}} = \bar{x}_{J^{***}}$ and

$$\boldsymbol{D}_{J^{**}}^{(\boldsymbol{n})} \boldsymbol{y}_{J^{**}} = \boldsymbol{D}_{J^{***}}^{(\boldsymbol{N})} \boldsymbol{x}_{J^{***}} \geq \boldsymbol{0}, \quad \bar{x}_{J^{***}} < x_v \quad (v \in J \setminus J^{***}). \tag{A.9}$$

Hence, for the set J^{***}, from (A.8) and (A.9) it holds that

$$\boldsymbol{D}_{J^{***}}^{(\boldsymbol{N})} \boldsymbol{x}_{J^{***}} \geq \boldsymbol{0}, \quad \bar{x}_{J^{***}} < x_u \quad (u \in \mathbb{N}_q \setminus J^{***}).$$

As we proved before, this implies that $\boldsymbol{x} \in \boldsymbol{\eta}_q^{-1}\left(A^{(q)}(J^{***})\right)$. However, this result is a contradiction because $J \neq J^{***}$, $\boldsymbol{x} \in \boldsymbol{\eta}_q^{-1}\left(A^{(q)}(J)\right)$ and $\boldsymbol{\eta}_q^{-1}\left(A^{(q)}(J)\right) \cap \boldsymbol{\eta}_q^{-1}\left(A^{(q)}(J^{***})\right) = \emptyset$. Therefore, we obtain $\boldsymbol{y} \in \boldsymbol{\eta}_i^{-1}\left(A^{(i)}(\mathbb{N}_i)\right)$. From (2) of Lemma 2.1, this result is equivalent to $\boldsymbol{D}_{\mathbb{N}_i}^{(\boldsymbol{n})} \boldsymbol{y} \geq \boldsymbol{0}$. This inequality can be written by using $\boldsymbol{N}$, J and $\boldsymbol{x}_J$ as $\boldsymbol{D}_J^{(\boldsymbol{N})}\boldsymbol{x}_J \geq \boldsymbol{0}$. Thus, for any i with $2 \leq i \leq q-1$, the inclusion relation $\subset$ of (2.7) is proved.

Finally, in the case of $i = q$, i.e., $J = \mathbb{N}_q \in \mathcal{J}_q^{(q)}$, we prove (2.7). First, we prove the inclusion relation $\supset$ of (2.7). Let $\boldsymbol{x} = (x_1, \ldots, x_q)' \in \mathbb{R}^q$, and let $\boldsymbol{D}_J\boldsymbol{x}_J \geq \boldsymbol{0}$. Recall that the following relation holds:

$$\mathbb{R}^q = \bigcup_{s=1}^{q} \bigcup_{J \in \mathcal{J}_s^{(q)}} \boldsymbol{\eta}_q^{-1}\left(A^{(q)}(J)\right),$$

$$\boldsymbol{\eta}_q^{-1}\left(A^{(q)}(J)\right) \cap \boldsymbol{\eta}_q^{-1}\left(A^{(q)}(J^*)\right) = \emptyset \quad (J \neq J^*).$$

Again, we consider proof by contradiction. Hence, we assume that there exists an integer s with $1 \leq s \leq q-1$ and a set J^* of $\mathcal{J}_s^{(q)}$ satisfying $\boldsymbol{x} \in \boldsymbol{\eta}_q^{-1}\left(A^{(q)}(J^*)\right)$. Thus, as we mentioned before, it holds that

$$\boldsymbol{D}_{J^*}\boldsymbol{x}_{J^*} \geq \boldsymbol{0}, \quad \bar{x}_{J^*} < x_t \quad (t \in \mathbb{N}_q \setminus J^*).$$

We would like to recall that $1 \in J^*$ and the number of elements in J^* is s. Here, if $s = q-1$, then $\mathbb{N}_q \setminus J^*$ has only one element a satisfying $a > 1$. Therefore, it holds that

$$\bar{x}_{\mathbb{N}_q \setminus \{a\}} < x_a.$$

However, this inequality is a contradiction because $\boldsymbol{D}_J\boldsymbol{x}_J \geq \boldsymbol{0}$. Hence, s satisfies $1 \leq s \leq q-2$. Incidentally, note that there exists a element t^* of $\mathbb{N}_q \setminus J^*$ which satisfies

$$^{\forall}t \in N_q \setminus (J^* \cup \{t^*\}),\ x_t \leq x_{t^*}$$

Therefore, form (2) of Lemma A we get

$$\bar{x}_{N_q \setminus (J^* \cup \{t^*\})} \leq x_{t^*}$$

In addition, since $\bar{x}_J < x_{t^*}$, from (3) of Lemma A we obtain

$$\bar{x}_{\mathbb{N}_q \setminus \{t^*\}} < x_{t^*}$$

However, this inequality is also contradiction because $\boldsymbol{D}_J\boldsymbol{x}_J \geq \boldsymbol{0}$. Thus, we get $s = q$. This implies that $J^* = \mathbb{N}_q \in \mathcal{J}_q^{(q)}$ and $\boldsymbol{x} \in \boldsymbol{\eta}_q^{-1}\left(A^{(q)}(\mathbb{N}_q)\right)$. Therefore, the inclusion relation $\supset$ of (2.7) in the case of $i = q$ is proved. Next, we prove $\subset$. Assume that $\boldsymbol{x} \in \boldsymbol{\eta}_q^{-1}\left(A^{(q)}(\mathbb{N}_q)\right)$. In other words, it holds that

$$\boldsymbol{\eta}_q(\boldsymbol{x}) \equiv \boldsymbol{\alpha} \in A^{(q)}(\mathbb{N}_q).$$

From the definition of $A^{(q)}(\mathbb{N}_q)$, we get $\boldsymbol{\alpha} = \boldsymbol{1}_q\alpha$, where $\boldsymbol{1}_q$ is a q-dimensional vector and every element of $\boldsymbol{1}_q$ is equal to one. Here, again we consider proof by contradiction. Therefore, we assume that there exists an integer s with $2 \leq s \leq q$ which satisfies

$$\bar{x}_{\mathbb{N}_q \setminus \{s\}} < x_s. \tag{A.10}$$

Meanwhile, for the function $H_q(\boldsymbol{\delta}; \boldsymbol{x})$ given by

$$H_q(\boldsymbol{\delta}; \boldsymbol{x}) = \sum_{a=1}^{q} N_a(x_a - \delta_a)^2,$$

it is easily checked that

$$\min_{\boldsymbol{\delta}\in A^{(q)}} H_q(\boldsymbol{\delta};\boldsymbol{x}) = H_q(\boldsymbol{\alpha};\boldsymbol{x}) = \sum_{a=1}^{q} N_a(x_a-\alpha)^2, \tag{A.11}$$

because $\boldsymbol{x} \in \boldsymbol{\eta}_q^{-1}\left(A^{(q)}(\mathbb{N}_q)\right)$ is true. Here, it is clear that the following inequality holds:

$$\sum_{a=1}^{q} N_a(x_a-\alpha)^2 \geq \min_{\beta\in\mathbb{R}} \sum_{a=1,\ a\neq s}^{q} N_a(x_a-\beta)^2 = \sum_{a=1,\ a\neq s}^{q} N_a(x_a-\bar{x}_{\mathbb{N}_q\setminus\{s\}})^2. \tag{A.12}$$

Hence, combining (A.11) and (A.12) we get

$$\min_{\boldsymbol{\delta}\in A^{(q)}} H_q(\boldsymbol{\delta};\boldsymbol{x}) \geq \sum_{a=1,\ a\neq s}^{q} N_a(x_a-\bar{x}_{\mathbb{N}_q\setminus\{s\}})^2. \tag{A.13}$$

Let $\boldsymbol{\beta}$ be a q-dimensional vector whose sth and tth ($t \in \mathbb{N}_q \setminus \{s\}$) elements are x_s and $\bar{x}_{\mathbb{N}_q\setminus\{s\}}$, respectively. Then, the inequality (A.13) can be written by using $\boldsymbol{\beta}$ as

$$\min_{\boldsymbol{\delta}\in A^{(q)}} H_q(\boldsymbol{\delta};\boldsymbol{x}) \geq H_q(\boldsymbol{\beta};\boldsymbol{x}).$$

On the other hand, from the assumption (A.10), we obtain

$$\min_{\boldsymbol{\delta}\in A^{(q)}} H_q(\boldsymbol{\delta};\boldsymbol{x}) \leq H_q(\boldsymbol{\beta};\boldsymbol{x}),$$

because $\boldsymbol{\beta} \in A^{(q)}$. Thus, we have

$$\min_{\boldsymbol{\delta}\in A^{(q)}} H_q(\boldsymbol{\delta};\boldsymbol{x}) = H_q(\boldsymbol{\beta};\boldsymbol{x}),$$

and this means that $\boldsymbol{\eta}_q(\boldsymbol{x}) = \boldsymbol{\beta}$. However, this result is a contradiction because $\boldsymbol{\eta}_q(\boldsymbol{x}) = \boldsymbol{\alpha}$ and $\boldsymbol{\alpha} \neq \boldsymbol{\beta}$. Hence, for any integer s with $2 \leq s \leq q$, it holds that $\bar{x}_{\mathbb{N}_q\setminus\{s\}} \geq x_s$. This inequality is equivalent to $\boldsymbol{D}_{\mathbb{N}_q}\boldsymbol{x}_{\mathbb{N}_q} \geq \boldsymbol{0}$. Therefore, the inclusion relation $\subset$ of (2.7) in the case of $i = q$ is proved. Consequently, (2) of Lemma 2.1 is proved.

Finally, we prove (3) of Lemma 2.1. When $J \neq \mathbb{N}_q$, we have already proved in the proof of (2) of Lemma 2.1. Thus, we prove the case of $J = \mathbb{N}_q$. Let $\boldsymbol{x} \in \boldsymbol{\eta}_q^{-1}\left(A^{(q)}(\mathbb{N}_q)\right)$. Then, it holds that $\boldsymbol{\eta}_q(\boldsymbol{x}) \equiv \boldsymbol{\alpha} \in A^{(q)}(\mathbb{N}_q)$ and $\boldsymbol{\alpha}$ can be written as $\boldsymbol{\alpha} = \alpha\boldsymbol{1}_q$. Here, for the function $H_q(\boldsymbol{\delta};\boldsymbol{x})$ defined by

$$H_q(\boldsymbol{\delta};\boldsymbol{x}) = \sum_{a=1}^{q} N_a(x_a-\delta_a)^2,$$

we obtain

$$\begin{aligned}\min_{\boldsymbol{\delta}\in A^{(q)}} H_q(\boldsymbol{\delta};\boldsymbol{x}) &= H_q(\boldsymbol{\alpha};\boldsymbol{x}) = \sum_{a=1}^{q} N_a(x_a-\alpha)^2 \\ &\geq \min_{\beta\in\mathbb{R}} \sum_{a=1}^{q} N_a(x_a-\beta)^2 = \sum_{a=1}^{q} N_a(x_a-\bar{x}_{\mathbb{N}_q})^2 = H_q(\bar{x}_{\mathbb{N}_q}\boldsymbol{1}_q;\boldsymbol{x}),\end{aligned} \tag{A.14}$$

because $\boldsymbol{x} \in \boldsymbol{\eta}_q^{-1}\left(A^{(q)}(\mathbb{N}_q)\right)$ holds. On the other hand, since $\bar{x}_{\mathbb{N}_q}\boldsymbol{1}_q \in A^{(q)}$, we get

$$\min_{\boldsymbol{\delta}\in A^{(q)}} H_q(\boldsymbol{\delta};\boldsymbol{x}) \leq H_q(\bar{x}_{\mathbb{N}_q}\boldsymbol{1}_q;\boldsymbol{x}).$$

By combining this inequality and (A.14), we have

$$\min_{\boldsymbol{\delta}\in A^{(q)}} H_q(\boldsymbol{\delta};\boldsymbol{x}) = H_q(\bar{x}_{\mathbb{N}_q}\boldsymbol{1}_q;\boldsymbol{x}).$$

This implies $\boldsymbol{\eta}_q(\boldsymbol{x}) = \boldsymbol{\alpha} = \bar{x}_{\mathbb{N}_q}\boldsymbol{1}_q$. Therefore, (3) of Lemma 2.1 is proved. □

Appendix 2: Technical lemma

In this section, we provide two technical lemmas. Using Lemma 2.1 and provided two lemmas, we prove Theorem 3.1 in Appendix 3.

Lemma B. Let $v_1, \ldots, v_l$ be independent random variables, and let $v_s \sim N(\xi_s, \tau^2/N_s)$ where $1 \le s \le l$, $\tau^2 > 0$, $\xi_1, \ldots, \xi_l \in \mathbb{R}$ and $N_1, \ldots, N_l \in \mathbb{R}_{>0}$. Let $\boldsymbol{N} = (N_1, \ldots, N_l)'$, $\boldsymbol{v} = (v_1, \ldots, v_l)'$ and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_l)'$. In addition, for any integer i with $1 \le i \le l$ and for any set J with $J \in \mathcal{J}_i^{(l)}$, define

$$S(J) = \sum_{s \in J} N_s (v_s - \xi_s)(v_s - \bar{v}_J).$$

Then, the following two propositions hold:

(1) If $J \ne \mathbb{N}_l$, then $\boldsymbol{v}_{\mathbb{N}_l \setminus J}$, $((\boldsymbol{D}_J \boldsymbol{v}_J)', S(J))'$ and $\bar{v}_J$ are mutually independent.

(2) if $J = \mathbb{N}_l$, then $((\boldsymbol{D}_J \boldsymbol{v}_J)', S(J))'$ and $\bar{v}_J$ are mutually independent.

Proof. First, we prove (1). From the assumption, $\boldsymbol{v}$ is distributed as the multivariate normal distribution with a diagonal covariance matrix. Therefore, noting that the two sets J and $\mathbb{N}_l \setminus J$ are disjoint sets, it can be shown that the two subvectors $\boldsymbol{v}_J$ and $\boldsymbol{v}_{\mathbb{N}_l \setminus J}$ are also distributed as (multivariate) normal distributions and these are mutually independent.

Next, we prove that $((\boldsymbol{D}_J \boldsymbol{v}_J)', S(J))'$ and $\bar{v}_J$ are functions of $\boldsymbol{v}_J$, and these are mutually independent. Here, the case of $J = \{1\}$ is clear because $((\boldsymbol{D}_J \boldsymbol{v}_J)', S(J))' = (0, 0)'$. Thus, we consider the case of $J \ne \{1\}$. Since

$$\sum_{s \in J} N_s \bar{v}_J (v_s - \bar{v}_J) = 0,$$

it holds that

$$\begin{aligned} S(J) = \sum_{s \in J} N_s (v_s - \xi_s)(v_s - \bar{v}_J) &= \sum_{s \in J} N_s (v_s - \bar{v}_J - \xi_s)(v_s - \bar{v}_J) \\ &= \sum_{s \in J} N_s (v_s - \bar{v}_J)^2 - \sum_{s \in J} N_s \xi_s (v_s - \bar{v}_J). \end{aligned}$$

Here, let

$$\boldsymbol{A} = (\mathrm{diag}(\boldsymbol{N}_J))^{1/2} \left\{ \boldsymbol{I}_{\#J} - \frac{\boldsymbol{1}_{\#J}}{\tilde{N}_J} \boldsymbol{N}_J' \right\}, \tag{B.1}$$

where $\mathrm{diag}(\boldsymbol{N}_J)$ means the diagonal matrix whose (a, a) element is the ath element of the vector $\boldsymbol{N}_J$. Then, $S(J)$ can be expressed as

$$S(J) = (\boldsymbol{A}\boldsymbol{v}_J)'(\boldsymbol{A}\boldsymbol{v}_J) - (\boldsymbol{\xi}_J' (\mathrm{diag}(\boldsymbol{N}_J))^{1/2}) \boldsymbol{A}\boldsymbol{v}_J.$$

Hence, $((\boldsymbol{D}_J \boldsymbol{v}_J)', S(J))'$ is the function of $((\boldsymbol{D}_J \boldsymbol{v}_J)', (\boldsymbol{A}\boldsymbol{v}_J)')'$. Therefore, it is sufficient to prove that $((\boldsymbol{D}_J \boldsymbol{v}_J)', (\boldsymbol{A}\boldsymbol{v}_J)')'$ and $\bar{v}_J$ are independent. Note that the vector $((\boldsymbol{D}_J \boldsymbol{v}_J)', (\boldsymbol{A}\boldsymbol{v}_J)', \bar{v}_J)'$ can written by

$$\begin{pmatrix} \boldsymbol{D}_J \boldsymbol{v}_J \\ \boldsymbol{A}\boldsymbol{v}_J \\ \bar{v}_J \end{pmatrix} = \begin{pmatrix} \boldsymbol{D}_J \\ \boldsymbol{A} \\ \boldsymbol{N}_J' / \tilde{N}_J \end{pmatrix} \boldsymbol{v}_J,$$

and $\boldsymbol{v}_J$ are distributed as multivariate normal distribution. Thus, $((\boldsymbol{D}_J\boldsymbol{v}_J)', (\boldsymbol{A}\boldsymbol{v}_J)')'$ and $\bar{v}_J$ are distributed as (multivariate) normal distributions. Hence, in order to prove its independence, it is sufficient to prove that the covariance of $((\boldsymbol{D}_J\boldsymbol{v}_J)', (\boldsymbol{A}\boldsymbol{v}_J)')'$ and $\bar{v}_J$ is the zero vector. Here, the covariance of $\boldsymbol{D}_J\boldsymbol{v}_J$ and $\bar{v}_J$ can be expressed as

$$\mathrm{Cov}[\boldsymbol{D}_J\boldsymbol{v}_J, \bar{v}_J] = \boldsymbol{D}_J\mathrm{Var}[\boldsymbol{v}_J]\boldsymbol{N}_J/\tilde{N}_J. \tag{B.2}$$

Furthermore, noting that $\mathrm{Var}[\boldsymbol{v}_J] = \tau^2(\mathrm{diag}(\boldsymbol{N}_J))^{-1}$, (B.2) can be written as

$$\mathrm{Cov}[\boldsymbol{D}_J\boldsymbol{v}_J, \bar{v}_J] = (\tau^2/\tilde{N}_J)\boldsymbol{D}_J(\mathrm{diag}(\boldsymbol{N}_J))^{-1}\boldsymbol{N}_J = (\tau^2/\tilde{N}_J)\boldsymbol{D}_J\boldsymbol{1}_{\#J}.$$

In addition, from the definition of the matrix $\boldsymbol{D}_J$, it holds that $\boldsymbol{D}_J\boldsymbol{1}_{\#J} = \boldsymbol{0}$. Therefore, we get $\mathrm{Cov}[\boldsymbol{D}_J\boldsymbol{v}_J, \bar{v}_J] = \boldsymbol{0}$. Similarly, the covariance of $\boldsymbol{A}\boldsymbol{v}_J$ and $\bar{v}_J$ is given by

$$\mathrm{Cov}[\boldsymbol{A}\boldsymbol{v}_J, \bar{v}_J] = (\tau^2/\tilde{N}_J)\boldsymbol{A}\boldsymbol{1}_{\#J},$$

and it holds that $\boldsymbol{A}\boldsymbol{1}_{\#J} = \boldsymbol{0}$ from (B.1). Thus, we have $\mathrm{Cov}[\boldsymbol{A}\boldsymbol{v}_J, \bar{v}_J] = \boldsymbol{0}$. Therefore, $((\boldsymbol{D}_J\boldsymbol{v}_J)', (\boldsymbol{A}\boldsymbol{v}_J)')'$ and $\bar{v}_J$ are independent. This implies that $((\boldsymbol{D}_J\boldsymbol{v}_J)', S(J))'$ and $\bar{v}_J$ are independent. Hence, (1) is proved. On the other hand, by using the same argument, we can also prove (2).

□

Lemma C. Let $v_1, \ldots, v_l$ be independent random variables defined as in Lemma B, and let

$$A^{(l)}(\{1\}) = \{(x_1, \ldots, x_l)' \in \mathbb{R}^l \mid x_1 < x_2, \ldots, x_1 < x_l\}.$$

Then, it holds that

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

Proof. From the definition of an indicator function, it is clear that the fourth equality holds. On the other hand, for the first and third equalities, we must prove

$$\boldsymbol{v} \in \boldsymbol{\eta}_l^{-1}(A^{(l)}(\{1\})) \Leftrightarrow \boldsymbol{v} \in A^{(l)}(\{1\}).$$

However, we have already proved this relation in (2.7). Therefore, we prove the second equality. For any integer s with $1 \leq s \leq l$, we define

$$\frac{\sqrt{N_s}(v_s - \xi_s)}{\tau} = z_s, \quad b_s = \frac{\xi_s\sqrt{N_s}}{\tau}.$$

Note that $z_1, \ldots, z_l$ are independent and identically distributed as $N(0, 1)$. Furthermore, it holds that

$$\frac{1}{\tau^2}\sum_{s=1}^{l} N_s v_s(v_s - \xi_s) = \sum_{s=1}^{l} z_s(z_s + b_s). \tag{C.2}$$

In addition, for any integer t with $2 \le t \le l$, putting

$$\frac{\sqrt{N_t}}{\sqrt{N_1}} = a_t,$$

the following relation holds:

$$\boldsymbol{v} \in A^{(l)}(\{1\}) \Leftrightarrow 2 \le t \le l,\ v_1 < v_t \Leftrightarrow 2 \le t \le l,\ a_t(z_1 + b_1) - b_t < z_t.$$

Here, define

$$E_l = \{(c_1, \ldots, c_l) \in \mathbb{R}^l \mid 2 \le t \le l,\ a_t(c_1 + b_1) - b_t < c_t\}.$$

Then, for the vector $\boldsymbol{z} = (z_1, \ldots, z_l)'$, it holds that $\boldsymbol{v} \in A^{(l)}(\{1\}) \Leftrightarrow \boldsymbol{z} \in E_l$. Using this result and (C.2), we obtain

$$\mathrm{E}\left[1_{\{\boldsymbol{v} \in A^{(l)}(\{1\})\}} \times \frac{1}{\tau^2} \sum_{s=1}^{l} N_s v_s (v_s - \xi_s)\right] = \mathrm{E}\left[1_{\{\boldsymbol{z} \in E_l\}} \times \sum_{s=1}^{l} z_s (z_s + b_s)\right]$$
$$= \int \cdots \int_{E_l} \left\{\sum_{s=1}^{l} z_s (z_s + b_s)\right\} \prod_{s=1}^{l} \phi(z_s) dz_1 \cdots dz_l, \tag{C.3}$$

where $\phi(x)$ is the probability density function of standard normal distribution. Here, when $l = 2$, Inatsu (2016) proved that (C.3) is equal to $l\mathrm{E}[1_{\{\boldsymbol{v} \in A^{(l)}(\{1\})\}}]$. Hence, we prove the case of $l \ge 3$.

First, for any integer s with $2 \le s \le l$ we define

$$F_s(x) = \int_{a_s(x+b_1)-b_s}^{\infty} \phi(y) dy.$$

In addition, let

$$G_1 = \int_{-\infty}^{\infty} z_1 (z_1 + b_1) \left(\prod_{s=2}^{l} F_s(z_1)\right) \phi(z_1) dz_1,$$

and let

$$G_s = \int_{-\infty}^{\infty} \left(\int_{a_s(z_1+b_1)-b_s}^{\infty} z_s (z_s + b_s) \phi(z_s) dz_s\right) \left(\prod_{2 \le t \le l,\ t \ne s} F_t(z_1)\right) \phi(z_1) dz_1, \tag{C.4}$$

where $s = 2, \ldots, l$. Then, (C.3) can be written as

$$\int \cdots \int_{E_l} \left\{\sum_{s=1}^{l} z_s (z_s + b_s)\right\} \prod_{s=1}^{l} \phi(z_s) dz_1 \cdots dz_l = \sum_{s=1}^{l} G_s. \tag{C.5}$$

Next, we calculate G_1 and G_s. Using the integration by parts, G_1 can be expressed as

$$G_1 = \left[-\phi(z_1)(z_1 + b_1) \left(\prod_{s=2}^{l} F_s(z_1)\right)\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \phi(z_1) \left(\prod_{s=2}^{l} F_s(z_1)\right) dz_1$$
$$+ \int_{-\infty}^{\infty} \phi(z_1)(z_1 + b_1) \frac{d}{dz_1} \left(\prod_{s=2}^{l} F_s(z_1)\right) dz_1. \tag{C.6}$$

Here, noting that

$$\frac{d}{dz_1} F_s(z_1) = -a_s \phi(a_s(z_1 + b_1) - b_s)$$

and the first term of the right hand side of (C.6) is zero, (C.6) can be written as

$$G_1 = \int_{-\infty}^{\infty} \phi(z_1) \left(\prod_{s=2}^{l} F_s(z_1) \right) dz_1$$
$$+ \int_{-\infty}^{\infty} \phi(z_1)(z_1 + b_1) \left\{ \sum_{s=2}^{l} \{-a_s \phi(a_s(z_1 + b_1) - b_s)\} \left(\prod_{2 \le t \le l,\ t \ne s} F_t(z_1) \right) \right\} dz_1. \quad \text{(C.7)}$$

Next, we calculate G_s. Here, note that

$$\int_{a_s(z_1+b_1)-b_s}^{\infty} z_s(z_s + b_s)\phi(z_s)dz_s = [-\phi(z_s)(z_s + b_s)]_{a_s(z_1+b_1)-b_s}^{\infty} + \int_{a_s(z_1+b_1)-b_s}^{\infty} \phi(z_s)dz_s$$
$$= a_s(z_1 + b_1)\phi\{a_s(z_1 + b_1) - b_s\} + F_s(z_1). \quad \text{(C.8)}$$

Hence, substituting (C.8) into (C.4) yields

$$G_s = \int_{-\infty}^{\infty} \phi(z_1) \left(\prod_{s=2}^{l} F_s(z_1) \right) dz_1$$
$$+ \int_{-\infty}^{\infty} \phi(z_1)(z_1 + b_1)\{a_s \phi(a_s(z_1 + b_1) - b_s)\} \left(\prod_{2 \le t \le l,\ t \ne s} F_t(z_1) \right) dz_1. \quad \text{(C.9)}$$

Therefore, using (C.7) and (C.9) we get

$$\sum_{s=1}^{l} G_s = l \int_{-\infty}^{\infty} \phi(z_1) \left(\prod_{s=2}^{l} F_s(z_1) \right) dz_1 = l \int \cdots \int_{E_l} \prod_{s=1}^{l} \phi(z_s) dz_1 \cdots dz_l$$
$$= l\mathrm{E}[1_{\{\boldsymbol{z} \in E_l\}}] = l\mathrm{E}[1_{\{\boldsymbol{v} \in A^{(l)}(\{1\})\}}]. \quad \text{(C.10)}$$

Thus, by substituting (C.10) into (C.5), we obtain (C.1). □

Appendix 3: Proof of Theorem 3.1

In this section, we prove Theorem 3.1. First, we provide the following lemma.

Lemma D. Let n_1, n_2 and τ^2 be positive numbers, and let ξ_1, and ξ_2 be real numbers. Put $\boldsymbol{n} = (n_1, n_2)'$. Let x_1 and x_2 be independent random variables distributed as $x_s \sim N(\xi_s, \tau^2/n_s)$, $(s = 1, 2)$, and let $\boldsymbol{x} = (x_1, x_2)'$. Then, the following two propositions hold:

(P1) For any integer i with $1 \le i \le 2$, and for any set J with $J \in \mathcal{J}_i^{(2)}$, it holds that

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

(P2) The following equality holds:

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

Proof. First, we prove (D.1). When $i = 1$, i.e., $J = \{1\}$, noting that $\bar{x}_J = x_1$, the equality (D.1) is clear. On the other hand, when $i = 2$, i.e., $J = \mathbb{N}_2$, the equality (D.1) is equivalent to (P1) of Lemma F given by Inatsu (2016), and it is already proved. Similarly, the proof of (D.2) is equivalent to the proof of (P2) of Lemma F given by Inatsu (2016). Therefore, lemma D is proved. □

Next, we consider the following lemma:

Lemma E. Let l be an integer with $l \geq 2$. Assume that the following proposition (P) is true:

(P) Let $N_1, \ldots, N_l$ and ς^2 be positive numbers, and let $\zeta_1, \ldots, \zeta_l$ be real numbers. Let $y_1, \ldots, y_l$ be independent random variables, and let $y_s \sim N(\zeta_s, \varsigma^2/N_s)$ where $s = 1, \ldots, l$. Put $\boldsymbol{N} = (N_1, \ldots, N_l)'$, $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_l)'$ and $\boldsymbol{y} = (y_1, \ldots, y_l)'$. Then, for any integer i with $1 \leq i \leq l$ and for any set J with $J \in \mathcal{J}_i^{(l)}$, it holds that

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

Under the assumption (P), the following proposition (P*) holds:

(P*) Let $n_1, \ldots, n_{l+1}$ and τ^2 be positive numbers, and let $\xi_1, \ldots, \xi_{l+1}$ be real numbers. Let $x_1, \ldots, x_{l+1}$ be independent random variables, and let $x_s \sim N(\xi_s, \tau^2/n_s)$ where $s = 1, \ldots, l+1$. Put $\boldsymbol{n} = (n_1, \ldots, n_{l+1})'$, $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_{l+1})'$ and $\boldsymbol{x} = (x_1, \ldots, x_{l+1})'$. Then, for any integer i with $1 \leq i \leq l+1$ and for any set J with $J \in \mathcal{J}_i^{(l+1)}$, it holds that

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

Moreover, the following equality holds:

$$\begin{aligned} &\mathrm{E}\left[\frac{1}{\tau^2} \sum_{s=1}^{l+1} n_s(x_s - \xi_s)(x_s - \eta_{l+1}^{(\boldsymbol{n})}(\boldsymbol{x})[s])\right] \\ &= \sum_{i=2}^{l+1}(i-1)\mathrm{P}\left(\boldsymbol{\eta}_{l+1}(\boldsymbol{x}) \in \bigcup_{J\in\mathcal{J}_i^{l+1}} A^{(l+1)}(J)\right). \end{aligned} \quad \text{(E.3)}$$

Note that Lemma D and Lemma E yield Theorem 3.1. Hence, we prove Lemma E.

Proof. First, we prove (E.2). Suppose that i is an integer satisfying $1 \leq i \leq l$ and suppose also that J is a set satisfying $J \in \mathcal{J}_i^{(l+1)}$. In this case, we replace $\boldsymbol{n}_J$, $\boldsymbol{x}_J$ and $\boldsymbol{\xi}_J$ with $\boldsymbol{N} = (N_1, \ldots, N_i)'$, $\boldsymbol{y} = (y_1, \ldots, y_i)'$ and $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_i)'$, respectively. We put $J^* = \mathbb{N}_i$. Then, from the assumption

(E.1), the left hand side of (E.2) can be expressed as

$$\begin{aligned}
&\mathrm{E}\left[1_{\{\boldsymbol{D}_J^{(\boldsymbol{n})}\boldsymbol{x}_J\geq\boldsymbol{0}\}}\frac{1}{\tau^2}\sum_{s\in J}n_s(x_s-\xi_s)(x_s-\bar{x}_J^{(\boldsymbol{n})})\right]\\
&= \mathrm{E}\left[1_{\{\boldsymbol{D}_{J^*}^{(\boldsymbol{N})}\boldsymbol{y}_{J^*}\geq\boldsymbol{0}\}}\frac{1}{\tau^2}\sum_{t\in J^*}N_t(y_t-\zeta_t)(y_t-\bar{y}_{J^*}^{(\boldsymbol{N})})\right]\\
&= (i-1)\mathrm{P}(\boldsymbol{D}_{J^*}^{(\boldsymbol{N})}\boldsymbol{y}_{J^*}\geq\boldsymbol{0})=(i-1)\mathrm{P}(\boldsymbol{D}_J^{(\boldsymbol{n})}\boldsymbol{x}_J\geq\boldsymbol{0}). \qquad \text{(E.4)}
\end{aligned}$$

Hence, we get (E.2). Therefore, it is sufficient to prove the case of $i=l+1$, i.e., $J=\mathbb{N}_{l+1}\in\mathcal{J}_i^{(l+1)}$. Here, the left hand side of (E.2) can be rewritten as

$$\mathrm{E}\left[1_{\{\boldsymbol{D}_J^{(\boldsymbol{n})}\boldsymbol{x}_J\geq\boldsymbol{0}\}}\frac{1}{\tau^2}\sum_{s\in J}n_s(x_s-\xi_s)(x_s-\bar{x}_J^{(\boldsymbol{n})})\right]=X-Y, \qquad \text{(E.5)}$$

where X and Y are given by

$$\begin{aligned}
X&=\mathrm{E}\left[1_{\{\boldsymbol{D}_J^{(\boldsymbol{n})}\boldsymbol{x}_J\geq\boldsymbol{0}\}}\frac{1}{\tau^2}\sum_{s=1}^{l+1}n_s(x_s-\xi_s)x_s\right],\\
Y&=\mathrm{E}\left[1_{\{\boldsymbol{D}_J^{(\boldsymbol{n})}\boldsymbol{x}_J\geq\boldsymbol{0}\}}\frac{1}{\tau^2}\sum_{s=1}^{l+1}n_s(x_s-\xi_s)\bar{x}_J^{(\boldsymbol{n})}\right].
\end{aligned}$$

First, we calculate Y. Noting that

$$\frac{1}{\tau^2}\sum_{s=1}^{l+1}n_s(x_s-\xi_s)\bar{x}_J^{(\boldsymbol{n})}=\frac{\tilde{n}_J}{\tau^2}(\bar{x}_J^{(\boldsymbol{n})}-\bar{\xi}_J^{(\boldsymbol{n})})\bar{x}_J^{(\boldsymbol{n})}$$

and $\bar{x}_J^{(\boldsymbol{n})}\sim N(\bar{\xi}_J^{(\boldsymbol{n})},\tau^2/\tilde{n}_J)$, from (2) of Lemma B we obtain

$$\begin{aligned}
Y&=\mathrm{E}\left[1_{\{\boldsymbol{D}_J^{(\boldsymbol{n})}\boldsymbol{x}_J\geq\boldsymbol{0}\}}\frac{1}{\tau^2}\sum_{s=1}^{l+1}n_s(x_s-\xi_s)\bar{x}_J^{(\boldsymbol{n})}\right]\\
&=\mathrm{E}\left[1_{\{\boldsymbol{D}_J^{(\boldsymbol{n})}\boldsymbol{x}_J\geq\boldsymbol{0}\}}\right]\mathrm{E}\left[\frac{\tilde{n}_J}{\tau^2}(\bar{x}_J^{(\boldsymbol{n})}-\bar{\xi}_J^{(\boldsymbol{n})})\bar{x}_J^{(\boldsymbol{n})}\right]\\
&=\mathrm{E}\left[1_{\{\boldsymbol{D}_J^{(\boldsymbol{n})}\boldsymbol{x}_J\geq\boldsymbol{0}\}}\right]\times 1=\mathrm{P}(\boldsymbol{D}_J^{(\boldsymbol{n})}\boldsymbol{x}_J\geq\boldsymbol{0}). \qquad \text{(E.6)}
\end{aligned}$$

Next, we calculate X. From (1) of Lemma 2.1, it is easily checked that the following equality holds:

$$1_{\{\boldsymbol{D}_J^{(\boldsymbol{n})}\boldsymbol{x}_J\geq\boldsymbol{0}\}}=1-\sum_{u=1}^{l}\sum_{J^*\in\mathcal{J}_u^{(l+1)}}1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(A^{(l+1)}(J^*))\}}. \qquad \text{(E.7)}$$

Therefore, X can be expressed by using (E.7) as

$$\begin{aligned}
X&=\mathrm{E}\left[\frac{1}{\tau^2}\sum_{s=1}^{l+1}n_s(x_s-\xi_s)x_s\right]\\
&\quad-\sum_{u=1}^{l}\sum_{J^*\in\mathcal{J}_u^{l+1}}\mathrm{E}\left[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(A^{(l+1)}(J^*))\}}\frac{1}{\tau^2}\sum_{s=1}^{l+1}n_s(x_s-\xi_s)x_s\right]\\
&=(l+1)-\sum_{u=1}^{l}\sum_{J^*\in\mathcal{J}_u^{l+1}}\mathrm{E}\left[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(A^{(l+1)}(J^*))\}}\frac{1}{\tau^2}\sum_{s=1}^{l+1}n_s(x_s-\xi_s)x_s\right], \qquad \text{(E.8)}
\end{aligned}$$

where the first term of the last equality in (E.8) is derived by $x_s \sim N(\xi_s, \tau^2/n_s)$. Next, for any integer u with $1 \leq u \leq l$ and for any set J^* with $J^* \in \mathcal{J}_u^{l+1}$, we calculate

$$\mathrm{E}\left[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(A^{(l+1)}(J^*))\}}\frac{1}{\tau^2}\sum_{s=1}^{l+1} n_s(x_s-\xi_s)x_s\right]. \tag{E.9}$$

Here, recall that from (2) of Lemma 2.1, the following relation holds:

$$\boldsymbol{x} \in \boldsymbol{\eta}_{l+1}^{-1}(A^{(l+1)}(J^*)) \Leftrightarrow \boldsymbol{D}_{J^*}\boldsymbol{x}_{J^*} \geq \boldsymbol{0},\ {}^{\forall}t \in \mathbb{N}_{l+1} \setminus J^*,\ \bar{x}_{J^*} < x_t. \tag{E.10}$$

Thus, noting that

$$\begin{aligned}
&\frac{1}{\tau^2}\sum_{s=1}^{l+1} n_s(x_s-\xi_s)x_s \\
&= \frac{1}{\tau^2}\sum_{s\in J^*} n_s(x_s-\xi_s)x_s + \frac{1}{\tau^2}\sum_{t\in\mathbb{N}_{l+1}\setminus J^*} n_t(x_t-\xi_t)x_t \\
&= \frac{1}{\tau^2}\sum_{s\in J^*} n_s(x_s-\xi_s)(x_s-\bar{x}_{J^*}+\bar{x}_{J^*}) + \frac{1}{\tau^2}\sum_{t\in\mathbb{N}_{l+1}\setminus J^*} n_t(x_t-\xi_t)x_t \\
&= \frac{1}{\tau^2}\sum_{s\in J^*} n_s(x_s-\xi_s)(x_s-\bar{x}_{J^*}) + \frac{\tilde{n}_{J^*}}{\tau^2}(\bar{x}_{J^*}-\bar{\xi}_{J^*})\bar{x}_{J^*} + \frac{1}{\tau^2}\sum_{t\in\mathbb{N}_{l+1}\setminus J^*} n_t(x_t-\xi_t)x_t,
\end{aligned}$$

the expectation (E.9) can be rewritten as

$$\mathrm{E}\left[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(A^{(l+1)}(J^*))\}}\frac{1}{\tau^2}\sum_{s=1}^{l+1} n_s(x_s-\xi_s)x_s\right] = G + H, \tag{E.11}$$

where G and H are given by

$$\begin{aligned}
G &= \mathrm{E}\left[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(A^{(l+1)}(J^*))\}}\frac{1}{\tau^2}\sum_{s\in J^*} n_s(x_s-\xi_s)(x_s-\bar{x}_{J^*})\right], \\
H &= \mathrm{E}\left[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(A^{(l+1)}(J^*))\}}\left(\frac{\tilde{n}_{J^*}}{\tau^2}(\bar{x}_{J^*}-\bar{\xi}_{J^*})\bar{x}_{J^*} + \frac{1}{\tau^2}\sum_{t\in\mathbb{N}_{l+1}\setminus J^*} n_t(x_t-\xi_t)x_t\right)\right].
\end{aligned}$$

By using (E.10), Lemma B and (E.4), G can be expressed as

$$\begin{aligned}
G &= \mathrm{E}[1_{\{{}^{\forall}t\in\mathbb{N}_{l+1}\setminus J^*,\ \bar{x}_{J^*}<x_t\}}] \times \mathrm{E}\left[1_{\{\boldsymbol{D}_{J^*}\boldsymbol{x}_{J^*}\geq\boldsymbol{0}\}}\frac{1}{\tau^2}\sum_{s\in J^*} n_s(x_s-\xi_s)(x_s-\bar{x}_{J^*})\right] \\
&= \mathrm{E}[1_{\{{}^{\forall}t\in\mathbb{N}_{l+1}\setminus J^*,\ \bar{x}_{J^*}<x_t\}}] \times (u-1)\mathrm{E}[1_{\{\boldsymbol{D}_{J^*}\boldsymbol{x}_{J^*}\geq\boldsymbol{0}\}}] \\
&= (u-1)\times\mathrm{E}[1_{\{\boldsymbol{D}_{J^*}\boldsymbol{x}_{J^*}\geq\boldsymbol{0},\ {}^{\forall}t\in\mathbb{N}_{l+1}\setminus J^*,\ \bar{x}_{J^*}<x_t\}}] = (u-1)\times\mathrm{E}[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(A^{(l+1)}(J^*))\}}].
\end{aligned}$$

On the other hand, using (E.10), Lemma B and Lemma C, H can be written as

$$\begin{aligned}
H &= \mathrm{E}[1_{\{\boldsymbol{D}_{J^*}\boldsymbol{x}_{J^*}\geq\boldsymbol{0}\}}] \\
&\quad \times \mathrm{E}\left[1_{\{{}^{\forall}t\in\mathbb{N}_{l+1}\setminus J^*,\ \bar{x}_{J^*}<x_t\}}\left(\frac{\tilde{n}_{J^*}}{\tau^2}(\bar{x}_{J^*}-\bar{\xi}_{J^*})\bar{x}_{J^*} + \frac{1}{\tau^2}\sum_{t\in\mathbb{N}_{l+1}\setminus J^*} n_t(x_t-\xi_t)x_t\right)\right] \\
&= \mathrm{E}[1_{\{\boldsymbol{D}_{J^*}\boldsymbol{x}_{J^*}\geq\boldsymbol{0}\}}] \times (l+1-u+1)\mathrm{E}\left[1_{\{{}^{\forall}t\in\mathbb{N}_{l+1}\setminus J^*,\ \bar{x}_{J^*}<x_t\}}\right] \\
&= (l+1-u+1)\times\mathrm{E}[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(A^{(l+1)}(J^*))\}}].
\end{aligned}$$

Hence, substituting G and H into (E.11) yields

$$\mathrm{E}\left[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(A^{(l+1)}(J^*))\}}\frac{1}{\tau^2}\sum_{s=1}^{l+1}n_s(x_s-\xi_s)x_s\right]=(l+1)\times\mathrm{E}[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(A^{(l+1)}(J^*))\}}]. \tag{E.12}$$

Furthermore, combining (E.12) and (E.8) we get

$$\begin{aligned}X&=(l+1)-\sum_{u=1}^{l}\sum_{J^*\in\mathcal{J}_u^{l+1}}(l+1)\times\mathrm{E}[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(A^{(l+1)}(J^*))\}}]\\&=(l+1)\mathrm{E}\left[1-\sum_{u=1}^{l}\sum_{J^*\in\mathcal{J}_u^{l+1}}1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(A^{(l+1)}(J^*))\}}\right]=(l+1)\mathrm{E}[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(A^{(l+1)}(J))\}}]\\&=(l+1)\mathrm{E}[1_{\{\boldsymbol{D}_J\boldsymbol{x}_J\geq\boldsymbol{0}\}}]=(l+1)\mathrm{P}(\boldsymbol{D}_J\boldsymbol{x}_J\geq\boldsymbol{0}).\end{aligned} \tag{E.13}$$

Thus, substituting (E.6) and (E.13) into (E.5) yields

$$\mathrm{E}\left[1_{\{\boldsymbol{D}_J^{(\boldsymbol{n})}\boldsymbol{x}_J\geq\boldsymbol{0}\}}\frac{1}{\tau^2}\sum_{s\in J}n_s(x_s-\xi_s)(x_s-\bar{x}_J^{(\boldsymbol{n})})\right]=l\mathrm{P}(\boldsymbol{D}_J\boldsymbol{x}_J\geq\boldsymbol{0}).$$

Hence, the expectation (E.2) for the case of $i=l+1$ (i.e., $J=\mathbb{N}_{l+1}$), is proved.

Finally, we prove (E.3). By using (1) and (3) of Lemma 2.1, the left hand side of (E.3) can be expressed as

$$\begin{aligned}&\mathrm{E}\left[\frac{1}{\tau^2}\sum_{s=1}^{l+1}n_s(x_s-\xi_s)(x_s-\eta_{l+1}^{(\boldsymbol{n})}(\boldsymbol{x})[s])\right]\\&=\mathrm{E}\left[\sum_{i=1}^{l+1}\sum_{J\in\mathcal{J}_i^{(l+1)}}\left(1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(A^{(l+1)}(J))\}}\frac{1}{\tau^2}\sum_{s=1}^{l+1}n_s(x_s-\xi_s)(x_s-\eta_{l+1}^{(\boldsymbol{n})}(\boldsymbol{x})[s])\right)\right]\\&=\sum_{i=2}^{l+1}\sum_{J\in\mathcal{J}_i^{(l+1)}}\mathrm{E}\left[\left(1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(A^{(l+1)}(J))\}}\frac{1}{\tau^2}\sum_{r\in J}n_r(x_r-\xi_r)(x_r-\bar{x}_J)\right)\right].\end{aligned} \tag{E.14}$$

Here, using (E.2), Lemma B and

(2) of Lemma 2.1, we obtain

$$\begin{aligned}&\mathrm{E}\left[\left(1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(A^{(l+1)}(J))\}}\frac{1}{\tau^2}\sum_{r\in J}n_r(x_r-\xi_r)(x_r-\bar{x}_J)\right)\right]\\&=\mathrm{E}[1_{\{\forall u\in\mathbb{N}_{l+1}\setminus J,\ \bar{x}_J<x_u\}}]\times\mathrm{E}\left[1_{\{\boldsymbol{D}_J\boldsymbol{x}_J\geq\boldsymbol{0}\}}\frac{1}{\tau^2}\sum_{r\in J}n_r(x_r-\xi_r)(x_r-\bar{x}_J)\right]\\&=\mathrm{E}[1_{\{\forall u\in\mathbb{N}_{l+1}\setminus J,\ \bar{x}_J<x_u\}}]\times(i-1)\mathrm{E}[1_{\{\boldsymbol{D}_J\boldsymbol{x}_J\geq\boldsymbol{0}\}}]=(i-1)\mathrm{P}(\boldsymbol{\eta}_{l+1}(\boldsymbol{x})\in A^{(l+1)}(J)).\end{aligned} \tag{E.15}$$

Thus, substituting (E.15) into (E.14) yields

$$\begin{aligned}&\mathrm{E}\left[\frac{1}{\tau^2}\sum_{s=1}^{l+1}n_s(x_s-\xi_s)(x_s-\eta_{l+1}^{(\boldsymbol{n})}(\boldsymbol{x})[s])\right]\\&=\sum_{i=2}^{l+1}(i-1)\sum_{J\in\mathcal{J}_i^{(l+1)}}\mathrm{P}(\boldsymbol{\eta}_{l+1}(\boldsymbol{x})\in A^{(l+1)}(J))\\&=\sum_{i=2}^{l+1}(i-1)\mathrm{P}\left(\boldsymbol{\eta}_{l+1}(\boldsymbol{x})\in\bigcup_{J\in\mathcal{J}_i^{l+1}}A^{(l+1)}(J)\right),\end{aligned}$$

because $A^{(l+1)}(J) \cap A^{(l+1)}(J^*) = \emptyset$ when $J \neq J^*$. Therefore, (E.3) is proved. □

Acknowledgements

The author would like to thank Professor Hirofumi Wakaki and Associate Professor Hirokazu Yanagihara of Hiroshima University for their helpful comments and suggestions.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory* (eds. B. N. Petrov & F. Csáki), 267–281, Akadémiai Kiadó, Budapest.

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.

Inatsu, Y. (2016). Akaike information criterion for ANOVA model with a simple order restriction. *TR 16-13, Statistical Research Group, Hiroshima University*, Hiroshima.

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

Knight, K. (1999). *Course in Mathematical Statistics.* Chapman & Hall.

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

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd edition. Springer.

Mallows, C. L. (1973). Some comments on C_p. *Technometrics*, **15**, 661–675.

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

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