Akaike Information Criterion for ANOVA Model with a Simple 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) for ANOVA model with a simple ordering (SO) $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_l$ where $\theta_1, \ldots, \theta_l$ are population means. Under an ordinary ANOVA model without any order restriction, it is well known that an ordinal AIC, whose penalty term is 2 $\times$ the number of parameters, is an asymptotically unbiased estimator of a risk function based on the expected K-L divergence. However, in general, under ANOVA model with the SO, the ordinal AIC has an asymptotic bias which depends on unknown population means. In order to solve this problem, we calculate the asymptotic bias, and we derive its unbiased estimator. By using this estimator we provide an asymptotically unbiased AIC for ANOVA model with the SO, called $\mathrm{AIC}_{\mathrm{SO}}$. A penalty term of the $\mathrm{AIC}_{\mathrm{SO}}$ is simply defined as a function of maximum likelihood estimators of population means.

Key Words: Order restriction, Simple ordering, AIC, ANOVA.

1. Introduction

In real data analysis, analysts can consider many statistical models. Nevertheless, in many cases, we assume that considered models satisfy some regularity conditions. For example, in the case of deriving a maximum likelihood estimator (MLE), we often assume that the MLE is a solution of a likelihood equation. If this assumption holds, in general, the MLE has good properties such as consistency and asymptotic normality. Furthermore, if additional mild conditions hold, statistics based on the MLE have also good properties, e.g., Akaike information criterion (AIC) becomes an asymptotically unbiased estimator of a risk function based on the expected K-L divergence, and a penalty term of AIC can be simply expressed as $2\times$ the number of parameters. In addition, it can be shown that the null distribution of a likelihood ratio statistic converges to chi-squared distribution. Thus, when certain regularity conditions hold, we can get good models (or

*1 Corresponding author
E-mail address: d144576@hiroshima-u.ac.jp

statistics) from the viewpoint of usefulness and estimation accuracy.

On the other hand, when some parameters of the considered model are restricted, regularity conditions are not satisfied. In particular, when the parameters $\theta_1, \ldots, \theta_k$ are restricted as $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$, this restriction is called a simple ordering (SO), and the SO is very important in applied statistics. The main advantage of considering order restrictions is that some information can be reflected in the model. As a result, we can expect that estimation accuracy can be improved. For example, let $X_1, \ldots, X_k$ be independent random variables, and let $X_i \sim N(\mu_i, \sigma^2/n_i)$ where $n_i > 0$. Then, when the assumption of the SO is true, the MLE of μ_i for the model with the SO is more efficient compared with the MLE for the model without any restriction. Specifically, the MLE $\hat{\mu}_i$ of μ_i for the non-restricted model is given by $\hat{\mu}_i = X_i$. On the other hand, under the assumption of the SO, from Robertson *et al.* (1988) the MLE $\hat{\mu}_{i,\mathrm{SO}}$ can be obtained as

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

For these MLEs, Brunk (1965), Lee (1981) and Kelly (1989) showed that

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

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

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

respectively. Furthermore, from Hwang and Peddada (1994), the result of (c) was extended to the case of elliptical distributions. Thus, considering order restrictions yields good estimators from the viewpoint of estimation accuracy.

However, models with order restrictions are not easy to use. Anraku (1999) considered AIC for k-clusters ANOVA model with the SO, and showed that a general AIC, whose penalty term is $2\times$ the number of parameters, is not an asymptotically unbiased estimator of a risk function. Furthermore, its asymptotic bias depends on unknown parameters. Moreover, Yokoyama (1995) considered a parallel profile model with a random-effects covariance structure proposed by Yokoyama and Fujikoshi (1993). Variance parameters of the random-effects covariance structure are restricted as the SO, and Yokoyama (1995) investigated the likelihood ratio test for testing the adequacy of this structure. In this test, they showed that the null distribution of the likelihood ratio test statistic does not necessarily converge to chi-squared distribution. In addition, they also showed that the limiting distribution of the test statistic depends on unknown variance parameters. As can be seen from these two examples, derived results from the model with order restrictions are not easy to use even if the assumed restriction is very simple such as the SO. Based on these, in this paper we focus on AIC for ANOVA model with the SO. Deriving an unbiased estimator of the asymptotic bias which depends on unknown parameters, we propose AIC for ANOVA model with the SO, called $\mathrm{AIC}_{\mathrm{SO}}$.

Finally, we would like to recall that AIC is defined as

$$\mathrm{AIC} = -2l(\hat{x}) + 2p, \tag{1.1}$$

where $\hat{x}$ is the MLE of a parameter x, $l(\cdot)$ is a log-likelihood function and p is the number of independent parameters. Hereafter, in order to avoid confusion, if $\hat{x}$ is derived based on the model without any order restriction, we refer to the AIC given by (1.1) as an ordinal AIC. Similarly, if $\hat{x}$ is derived based on the model with a order restriction, we refer to the AIC given by (1.1) as a pseudo AIC (pAIC).

The remainder of the present paper is organized as follows: In Section 2, we derive MLEs of parameters and a risk function for ANOVA model with the SO. In Section 3, we define several notations, and we provide one important lemma for calculating the asymptotic bias. In Section 4, we provide AIC for ANOVA model with the SO, called $\mathrm{AIC}_{\mathrm{SO}}$. In Section 5, we introduce different $\mathrm{AIC}_{\mathrm{SO}}$s for several special cases. In Section 6, we confirm that performance of the $\mathrm{AIC}_{\mathrm{SO}}$ through numerical experiments. In Section 7, we conclude our discussion. Technical details are provided in Appendix.

2. ANOVA model with a simple order restriction

Let X_{ij} be a observation variable on the jth individual in the ith cluster, where $i = 1, \ldots, k$ and $j = 1, \ldots, N_i$. Here, let $k \geq 2$ and $N = N_1 + \cdots, N_k$, and let $N - k - 6 > 0$. Moreover, assume that $X_{11}, \ldots, X_{kN_k}$ are mutually independent random variables. In this setting, we consider the model

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

where $\theta_1, \ldots, \theta_k$ and $\sigma^2 > 0$ are unknown parameters. Furthermore, we assume that the parameters $\theta_1, \ldots, \theta_k$ are restricted as

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

Thus, the restriction (2.2) is the SO. Let Θ be a set defined as $\Theta = \{(\theta_1, \ldots, \theta_k)' \in \mathbb{R}^k \mid \theta_1 \leq \theta_2 \leq \cdots \leq \theta_k\}$. Then, the model (2.1) with the restriction (2.2) is equal to ANOVA model whose mean parameters are restricted on Θ. Here, we put $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)'$. In addition, let $\boldsymbol{\theta}_* = (\theta_{1,*}, \ldots, \theta_{k,*})'$ and σ_*^2 denote the true parameters of $\boldsymbol{\theta}$ and σ^2, respectively. Finally, for the true parameters $\boldsymbol{\theta}_*$ and σ_*^2, we assume that $\boldsymbol{\theta}_* \in \Theta$ and $\sigma_*^2 > 0$.

2.1. Maximum likelihood estimator

In this subsection, we derive MLEs of unknown parameters for the model (2.1) with the SO. Let $\boldsymbol{N} = (N_1, \ldots, N_k)'$. Suppose that $\boldsymbol{X}$ is an N-dimensional vector which has all X_{ij}, $(i = 1, \ldots, k,\ j = 1, \ldots, N_i)$. In other words, $\boldsymbol{X}$ can be written as $\boldsymbol{X} = (X_{11}, \ldots, X_{ij}, \ldots, X_{kN_k})'$. Furthermore, for any i with $1 \leq i \leq k$, define

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

Hence, $\bar{X}_i$ and $\bar{\sigma}^2$ are the sample mean and variance, respectively. We put $\bar{\boldsymbol{X}} = (\bar{X}_1, \ldots, \bar{X}_k)'$. Note that under the ordinal ANOVA model without any order restriction, the MLEs of $\boldsymbol{\theta}$ and σ^2 are $\bar{\boldsymbol{X}}$ and $\bar{\sigma}^2$, respectively. Here, since X_{ij}'s are mutually independent, from normality of X_{ij}, a log-likelihood function $l(\boldsymbol{\theta}, \sigma^2; \boldsymbol{X})$ can be expressed as

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

Here, for any $\boldsymbol{a} = (a_1, \ldots, a_p)' \in \mathbb{R}^p$ and $\boldsymbol{b} = (b_1, \ldots, b_p)' \in \mathbb{R}^p_{>0}$, we define

$$\|\boldsymbol{a}\|_{\boldsymbol{b}} = \sqrt{\sum_{i=1}^{p} b_i a_i^2}. \tag{2.5}$$

Note that (2.5) is a complete norm. Then, for any $\sigma^2 > 0$, the maximization problem of $l(\boldsymbol{\theta}, \sigma^2; \boldsymbol{X})$ on Θ is equal to the minimization problem of

$$H(\boldsymbol{\theta}) = \sum_{i=1}^{k} N_i(\bar{X}_i - \theta_i)^2 = \|\bar{\boldsymbol{X}} - \boldsymbol{\theta}\|^2_{\boldsymbol{N}}, \tag{2.6}$$

on Θ. Needless to say, this minimization problem is equal to the minimization of $H^*(\boldsymbol{\theta}) = \sqrt{H(\boldsymbol{\theta})} = \|\bar{\boldsymbol{X}} - \boldsymbol{\theta}\|_{\boldsymbol{N}}$ on Θ. Recall that the norm $\|\cdot\|_{\boldsymbol{N}}$ is the complete norm, and the set Θ is the non-empty closed convex set. Therefore, for any $\bar{\boldsymbol{X}} \in \mathbb{R}^k$, there exists a unique point $\hat{\boldsymbol{\theta}}$ in Θ such that $\hat{\boldsymbol{\theta}}$ minimizes $H^*(\boldsymbol{\theta})$ on Θ, (see, e.g., Rudin, 1986). This implies that existence and uniqueness for the MLE $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)'$ of $\boldsymbol{\theta}$ hold. Moreover, from Robertson *et al.* (1988), for any i with $1 \leq i \leq k$, $\hat{\theta}_i$ is given by

$$\hat{\theta}_i = \min_{v; i \leq v}\max_{u; u \leq i}\frac{\sum_{j=u}^{v} N_j \bar{X}_j}{\sum_{j=u}^{v} N_j}. \tag{2.7}$$

On the other hand, it is easily checked that the MLE $\hat{\sigma}^2$ of σ^2 can be obtained by differentiating the function $l(\hat{\boldsymbol{\theta}}, \sigma^2; \boldsymbol{X})$ with respect to (w.r.t.) σ^2 as

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

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

2.2. Risk function and bias

Let $\boldsymbol{X}^\star$ be a random variable, and let $\boldsymbol{X}^\star \sim$ i.i.d. $\boldsymbol{X}$. Then, a risk function based on the expected Kullback-Leibler divergence can be defined as

$$\begin{aligned}
&\mathrm{E}[\mathrm{E}_\star[-2l(\hat{\boldsymbol{\theta}}, \hat{\sigma}^2; \boldsymbol{X}^\star)]] \\
&= \mathrm{E}\left[N\log(2\pi\hat{\sigma}^2) + \frac{N\sigma_*^2}{\hat{\sigma}^2} + \frac{\sum_{i=1}^k N_i(\theta_{i,*} - \hat{\theta}_i)^2}{\hat{\sigma}^2}\right] \equiv R \quad \text{(say)}.
\end{aligned} \tag{2.8}$$

Note that in the case of the ordinal ANOVA model, a risk function $\bar{R}$ is given by $\bar{R} = \mathrm{E}[\mathrm{E}_\star[-2l(\bar{\boldsymbol{X}}, \bar{\sigma}^2; \boldsymbol{X}^\star)]]$. Since the maximum log-likelihood $l(\hat{\boldsymbol{\theta}}, \hat{\sigma}^2; \boldsymbol{X})$ can be written as

$$l(\hat{\boldsymbol{\theta}}, \hat{\sigma}^2; \boldsymbol{X}) = -\frac{N}{2}\log(2\pi\hat{\sigma}^2) - \frac{N}{2}, \tag{2.9}$$

if we estimate the risk function R by $-2l(\hat{\boldsymbol{\theta}}, \hat{\sigma}^2; \boldsymbol{X})$, then the bias B, which is the difference between the expected value of $-2l(\hat{\boldsymbol{\theta}}, \hat{\sigma}^2; \boldsymbol{X})$ and R, can be expressed as

$$B = \mathrm{E}[R - \{-2l(\hat{\boldsymbol{\theta}}, \hat{\sigma}^2; \boldsymbol{X})\}] = \mathrm{E}\left[\frac{N\sigma_*^2}{\hat{\sigma}^2}\right] + \mathrm{E}\left[\frac{\sum_{i=1}^k N_i(\theta_{i,*} - \hat{\theta}_i)^2}{\hat{\sigma}^2}\right] - N. \tag{2.10}$$

Next, we evaluate the bias B. Let S and T be random variables defined by

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

Note that S is distributed as χ^2_{N-k} where χ^2_{N-k} is the chi-squared distribution with $N-k$ degrees of freedom. Furthermore, since $X_{11}, \ldots, X_{kN_k}$ are independently distributed as normal distributions, we obtain $S \perp\!\!\!\perp \bar{\boldsymbol{X}}$ where the notation $* \perp\!\!\!\perp \star$ means that $*$ and $\star$ are mutually independent. In addition, from (2.7), $\hat{\boldsymbol{\theta}}$ is a function of the random vector $\bar{\boldsymbol{X}}$. Thus, T is also a function of $\bar{\boldsymbol{X}}$ and it holds that $S \perp\!\!\!\perp T$. Using S and T, it holds that $N\hat{\sigma}^2/\sigma_*^2 = S + T$ and we obtain

$$\frac{N\sigma_*^2}{\hat{\sigma}^2} = \frac{N^2}{N\hat{\sigma}^2/\sigma_*^2} = \frac{N^2}{S+T} = \frac{N^2}{S}\frac{1}{1+T/S}. \tag{2.11}$$

In addition, noting that $(1+x)^{-1} = 1 - x + c^* x^2$ where $x \geq 0$ and $0 \leq c^* \leq 1$, (2.11) can be expanded as

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

where C^* is a random variable with $0 \leq C^* \leq 1$. Hence, from $S \sim \chi^2_{N-k}$ and $S \perp\!\!\!\perp T$, we have

$$\begin{aligned} \mathrm{E}\left[\frac{N\sigma_*^2}{\hat{\sigma}^2}\right] &= \frac{N^2}{N-k-2} - \frac{N^2 \mathrm{E}[T]}{(N-k-2)(N-k-4)} + \mathrm{E}\left[C^* \frac{N^2 T^2}{S^3}\right] \\ &= N + k + 2 + O(N^{-1}) - \mathrm{E}[T] + O(N^{-1})\mathrm{E}[T] + \mathrm{E}\left[C^* \frac{N^2 T^2}{S^3}\right]. \end{aligned} \tag{2.12}$$

Similarly, using $(1+y)^{-1} = 1 - c^\star y$ where $y \geq 0$ and $0 \leq c^\star \leq 1$, we get

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

where $C^\star$ is a random variable with $0 \leq C^\star \leq 1$. Here, for any i with $1 \leq i \leq k$, it holds that $S \perp\!\!\!\perp \bar{X}_i$, $S \perp\!\!\!\perp \hat{\theta}_i$, $S \perp\!\!\!\perp T$ and $\bar{X}_i \sim \mathrm{N}(\theta_{i,*}, \sigma_*^2/N_i)$. Therefore, we obtain

$$\begin{aligned} &\mathrm{E}\left[\frac{\sum_{i=1}^k N_i(\theta_{i,*} - \hat{\theta}_i)^2}{\hat{\sigma}^2}\right] \\ &= \frac{Nk}{N-k-2} - \frac{2N}{N-k-2} \mathrm{E}\left[\frac{1}{\sigma_*^2} \sum_{i=1}^k N_i(\bar{X}_i - \theta_{i,*})(\bar{X}_i - \hat{\theta}_i)\right] \\ &\quad + \frac{N\mathrm{E}[T]}{N-k-2} - \mathrm{E}\left[C^\star \frac{NT}{\sigma_*^2} \frac{\sum_{i=1}^k N_i(\theta_{i,*} - \hat{\theta}_i)^2}{S^2}\right] \\ &= k + O(N^{-1}) - \frac{2N}{N-k-2} \mathrm{E}\left[\frac{1}{\sigma_*^2} \sum_{i=1}^k N_i(\bar{X}_i - \theta_{i,*})(\bar{X}_i - \hat{\theta}_i)\right] \\ &\quad + \mathrm{E}[T] + O(N^{-1})\mathrm{E}[T] - \mathrm{E}\left[C^\star \frac{NT}{\sigma_*^2} \frac{\sum_{i=1}^k N_i(\theta_{i,*} - \hat{\theta}_i)^2}{S^2}\right]. \end{aligned} \tag{2.13}$$

Thus, from (2.12) and (2.13), it holds that

$$\mathrm{E}\left[\frac{N\sigma_*^2}{\hat{\sigma}^2}\right] + \mathrm{E}\left[\frac{\sum_{i=1}^k N_i(\theta_{i,*} - \hat{\theta}_i)^2}{\hat{\sigma}^2}\right]$$
$$= N + 2(k+1) - \frac{2N}{N-k-2}\mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{i=1}^k N_i(\bar{X}_i - \theta_{i,*})(\bar{X}_i - \hat{\theta}_i)\right] + J, \qquad (2.14)$$

where J is given by

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

Here, from the definition of $\hat{\boldsymbol{\theta}}$, it holds that $\|\bar{\boldsymbol{X}} - \hat{\boldsymbol{\theta}}\|_{\boldsymbol{N}} \leq \|\bar{\boldsymbol{X}} - \boldsymbol{\theta}\|_{\boldsymbol{N}}$ for any $\boldsymbol{\theta} \in \Theta$. Moreover, from the assumption, the true value $\boldsymbol{\theta}_*$ satisfies $\boldsymbol{\theta}_* \in \Theta$. Thus, it holds that $\|\bar{\boldsymbol{X}} - \hat{\boldsymbol{\theta}}\|_{\boldsymbol{N}} \leq \|\bar{\boldsymbol{X}} - \boldsymbol{\theta}_*\|_{\boldsymbol{N}}$ and

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

where $K \sim \chi_k^2$. Therefore, by using the above inequality, we get $0 \leq \mathrm{E}[T] \leq \mathrm{E}[K] = k$ and $\mathrm{E}[T] = O(1)$. In addition, noting that $0 \leq C^* \leq 1$, we have

$$\left|\mathrm{E}\left[C^* \frac{N^2T^2}{S^3}\right]\right| \leq \mathrm{E}\left[\frac{N^2T^2}{S^3}\right] = \frac{N^2\mathrm{E}[T^2]}{(N-k-2)(N-k-4)(N-k-6)}$$
$$\leq O(N^{-1})\mathrm{E}[K^2] = O(N^{-1})(2k + k^2) = O(N^{-1}).$$

This implies

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

Noting that the triangle inequality $\|\boldsymbol{\theta}_* - \hat{\boldsymbol{\theta}}\|_{\boldsymbol{N}} \leq \|\boldsymbol{\theta}_* - \bar{\boldsymbol{X}}\|_{\boldsymbol{N}} + \|\bar{\boldsymbol{X}} - \hat{\boldsymbol{\theta}}\|_{\boldsymbol{N}}$ and $\|\bar{\boldsymbol{X}} - \hat{\boldsymbol{\theta}}\|_{\boldsymbol{N}} \leq \|\bar{\boldsymbol{X}} - \boldsymbol{\theta}_*\|_{\boldsymbol{N}}$, we obtain $\|\boldsymbol{\theta}_* - \hat{\boldsymbol{\theta}}\|_{\boldsymbol{N}} \leq 2\|\boldsymbol{\theta}_* - \bar{\boldsymbol{X}}\|_{\boldsymbol{N}}$. Hence, since $0 \leq C^\star \leq 1$ and $T \leq K$, the following inequality holds:

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

This implies

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

Thus, from the definition of J, we obtain $J = O(N^{-1})$. From (2.10) and (2.14), the bias B can be expressed as

$$B = 2(k+1) - \frac{2N}{N-k-2}\mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{i=1}^k N_i(\bar{X}_i-\theta_{i,*})(\bar{X}_i-\hat{\theta}_i)\right] + O(N^{-1}). \tag{2.15}$$

Hence, in order to correct the bias up to the order of N^{-1}, we must calculate the expected value in (2.15).

3. Notation and main lemma

In this section, we provide the lemma to calculate the expected value in (2.15). First, we define several notations.

3.1. Notation

Let l be an integer with $l \geq 2$ and let $n_1, \ldots, n_l$ be positive numbers. We put $\boldsymbol{n} = (n_1, \ldots, n_l)'$. For any l-dimensional vector $\boldsymbol{x} = (x_1, \ldots, x_l)' \in \mathbb{R}^l$, and for any i, j with $1 \leq i \leq j \leq l$, we write $\boldsymbol{x}_{[i,j]} = (x_i, \ldots, x_j)'$. Note that $\boldsymbol{x}_{[i,j]}$ is a $(j-i+1)$-dimensional vector whose the sth element is x_{i+s-1} where $1 \leq s \leq j-i+1$. In particular, $\boldsymbol{x}_{[i,i]} = x_i$ and $\boldsymbol{x}_{[1,l]} = \boldsymbol{x}$. Let

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

For simplicity, we often represent $\bar{x}^{(\boldsymbol{n})}_{[i,j]}$ as $\bar{x}_{[i,j]}$. Note that $\bar{x}_{[i,i]} = x_i$.

Next, let $\mathcal{A}^l$ be a set defined by

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

and let $\mathcal{A}^l_1$ and $\mathcal{A}^l_l$ be sets defined by

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

and

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

We define $\mathcal{A}^1 = \mathbb{R}^1$. Moreover, for any integer i with $1 \le i \le l$, we write

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

Hence, for example, in the case of $l = 2$, the sets $\mathcal{W}_1^2$ and $\mathcal{W}_2^2$ are given by

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

and in the case of $l = 3$, the sets $\mathcal{W}_1^3$, $\mathcal{W}_2^3$ and $\mathcal{W}_3^3$ are given by

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

Furthermore, in the case of $l = 4$, the sets $\mathcal{W}_1^4$, $\mathcal{W}_2^4$, $\mathcal{W}_3^4$ and $\mathcal{W}_4^4$ are given by

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

Note that the number of elements of $\mathcal{W}_i^l$ is ${}_{l-1}C_{i-1}$. Also note that, for any element $\boldsymbol{w} = (w_1, \dots, w_i)'$ in $\mathcal{W}_i^l$, $\boldsymbol{w}$ is an i-dimensional vector and $w_i = l$. From the definitions of $\mathcal{W}_1^l$ and $\mathcal{W}_l^l$, $\mathcal{W}_1^l$ has the unique element $\boldsymbol{w} = (l)'$ and $\mathcal{W}_l^l$ has the unique element $\boldsymbol{w} = (1, \dots, l)'$. Furthermore, for any i $(i = 1, \dots, l)$ and for any $\boldsymbol{w} \in \mathcal{W}_i^l$, we define a set $\mathcal{A}_i^l(\boldsymbol{w})$ as follows. First, in the case of $i = 1$, $\mathcal{W}_1^l$ has the unique element $\boldsymbol{w} = (l)'$, and we define

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

On the other hand, in the case of $2 \le i \le l$, for any element $\boldsymbol{w} = (w_1, \dots, w_i)'$ in $\mathcal{W}_i^l$, we define

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

Thus, from (3.1), the element $\boldsymbol{x} = (x_1, \dots, x_l)'$ in $\mathcal{A}_i^l(\boldsymbol{w})$ satisfies

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

In particular, when $i = l$, $\mathcal{W}_l^l$ has the unique element $\boldsymbol{w} = (w_1, \dots, w_l)' = (1, \dots, l)'$, and it holds that

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

Here, we provide several examples. When $l = 2$, $\mathcal{A}_1^2(\boldsymbol{w})$ and $\mathcal{A}_2^2(\boldsymbol{w})$ can be expressed as

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

In addition, when $l = 3$, for each s $(s = 1, 2, 3)$, $\mathcal{A}_s^3(\boldsymbol{w})$ can be expressed as

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

Therefore, in general, it holds that

$$
\mathcal{A}^l = \bigcup_{i=1}^{l} \bigcup_{\boldsymbol{w};\boldsymbol{w}\in\mathcal{W}_i^l} \mathcal{A}_i^l(\boldsymbol{w}), \tag{3.2}
$$

and

$$
(i, \boldsymbol{w}) \neq (i^*, \boldsymbol{w}^*) \Rightarrow \mathcal{A}_i^l(\boldsymbol{w}) \cap \mathcal{A}_{i^*}^l(\boldsymbol{w}^*) = \emptyset. \tag{3.3}
$$

Next, for any i and j with $1 \leq i \leq j \leq l$, we define a matrix $\boldsymbol{D}_{i,j}^{(\boldsymbol{n})}$. First, when $i = j$, let $\boldsymbol{D}_{i,j}^{(\boldsymbol{n})}$ be a 1×1 matrix and let $\boldsymbol{D}_{i,j}^{(\boldsymbol{n})} = 0$. On the other hand, when $i < j$, let $\boldsymbol{D}_{i,j}^{(\boldsymbol{n})}$ be a $(j-i) \times (j-i+1)$ matrix and let the sth row of $\boldsymbol{D}_{i,j}^{(\boldsymbol{n})}$ $(1 \leq s \leq j-i)$ be defined by

$$
\left(\frac{1}{\tilde{n}_{[i,i+s-1]}} \boldsymbol{n}'_{[i,i+s-1]}, \frac{-1}{\tilde{n}_{[i+s,j]}} \boldsymbol{n}'_{[i+s,j]} \right). \tag{3.4}
$$

Hence, for example, when $l = 4$, $\boldsymbol{D}_{i,j}^{(\boldsymbol{n})}$'s $(1 \leq i \leq j \leq 4)$ are given by

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

For simplicity, we often represent $\boldsymbol{D}_{i,j}^{(\boldsymbol{n})}$ as $\boldsymbol{D}_{i,j}$.

Finally, we define a function $\boldsymbol{\eta}_l^{(\boldsymbol{n})}$. Let $\boldsymbol{\eta}_l^{(\boldsymbol{n})}$ be a function from $\mathbb{R}^l$ to $\mathcal{A}^l$, and let $\boldsymbol{\eta}_l^{(\boldsymbol{n})}(\boldsymbol{x})$ be defined by

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

for any $\boldsymbol{x} = (x_1, \ldots, x_l)' \in \mathbb{R}^l$. For simplicity, we often represent $\boldsymbol{\eta}_l^{(\boldsymbol{n})}$ as $\boldsymbol{\eta}_l$. Note that $\boldsymbol{\eta}_l(\boldsymbol{x})$ is well-defined because $(\mathbb{R}^l, \| \|_{\boldsymbol{n}})$ is a Hilbert space and $\mathcal{A}^l$ is the non-empty closed convex set (see, e.g., Rudin, 1986). Also note that $\boldsymbol{\eta}_l(\boldsymbol{x})$ is an l-dimensional vector. Let

$\eta_l(\boldsymbol{x})[s]$ be a sth element of $\boldsymbol{\eta}_l(\boldsymbol{x})$ $(1 \leq s \leq l)$. Then, from Robertson *et al.* (1988), $\eta_l(\boldsymbol{x})[s]$ can be expressed as

$$\eta_l(\boldsymbol{x})[s] = \min_{v;v\geq s} \max_{u;u\leq s} \frac{\sum_{j=u}^{v} n_j x_j}{\sum_{j=u}^{v} n_j} = \min_{v;v\geq s} \max_{u;u\leq s} \bar{x}_{[u,v]}. \tag{3.5}$$

In addition, we define $\boldsymbol{\eta}_1(\boldsymbol{x}) = \boldsymbol{x}$.

3.2. Main lemma

The following lemma holds.

Lemma 3.1. Let k be an integer with $k \geq 2$, and let $n_1, \ldots, n_k$ be positive numbers. Let $\xi_1, \ldots, \xi_k$ be real numbers, and let τ^2 be a positive number. Suppose that $x_1, \ldots, x_k$ are independent random variables, and $x_i \sim N(\xi_i, \tau^2/n_i)$, $(i = 1, \ldots, k)$. We put $\boldsymbol{n} = (n_1, \ldots, n_k)'$, $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_k)'$ and $\boldsymbol{x} = (x_1, \ldots, x_k)'$. Then, it holds that

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

Proof. See Appendix. □

4. AIC for ANOVA model with the simple ordering

In this section, we derive AIC for ANOVA model (2.1) with the SO. First, we calculate the expected value in (2.15). From (2.3), $\bar{X}_1, \ldots, \bar{X}_k$ are mutually independent, and for any i, with $1 \leq i \leq k$, it holds that $\bar{X}_i \sim N(\theta_{i,*}, \sigma_*^2/N_i)$. Furthermore, from (2.7) the MLE $\hat{\boldsymbol{\theta}}$ can be expressed as $\hat{\boldsymbol{\theta}} = \boldsymbol{\eta}_k^{(\boldsymbol{N})}(\bar{\boldsymbol{X}})$. Hence, from Lemma 3.1, the expected value in (2.15) can be written as

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

Thus, noting that $Q = O(1)$, substituting Q into (2.15) yields

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

Therefore, in order to correct the bias up to the order of N^{-1}, we only have to add $2(k+1)-2Q$ to $-2l(\hat{\boldsymbol{\theta}}, \hat{\sigma}^2; \boldsymbol{X})$. However, it is easily checked that Q depends on the true values $\theta_{1,*}, \ldots, \theta_{k,*}$ and σ_*^2. Thus, we must estimate Q. Here, let

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

and let

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

where the notation $\#\hat{\mathcal{M}}$ means that the number of elements of $\hat{\mathcal{M}}$. From the definition of $\hat{m}$, $\hat{m}$ is a discrete random variable, and its possible values are 1 to k. For example, if $\hat{\theta}_1 = \cdots = \hat{\theta}_k$, then $\hat{m} = 1$, and if $\hat{\theta}_1 < \hat{\theta}_2 < \cdots < \hat{\theta}_k$, then $\hat{m} = k$. Similarly, if $\hat{\theta}_1 < \hat{\theta}_2 = \cdots = \hat{\theta}_k$, then $\hat{m} = 2$. Here, from the definitions of $\hat{m}$ and $\mathcal{A}_i^k(\boldsymbol{w})$, we have

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

This implies

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

Thus, $k-\hat{m}$ is an unbiased estimator of Q. Therefore, from (4.1) we obtain

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

Hence, adding $2(\hat{m}+1)$ (instead of $2(k+1)-2Q$) to $-2l(\hat{\boldsymbol{\theta}}, \hat{\sigma}^2; \boldsymbol{X})$, we obtain AIC for ANOVA model with the SO, called $\mathrm{AIC}_{\mathrm{SO}}$.

Theorem 4.1. Let $l(\hat{\boldsymbol{\theta}}, \hat{\sigma}^2; \boldsymbol{X})$ be the maximum log-likelihood given by (2.9), and let $\hat{m}$ be the random variable given by (4.2). Then, the $\mathrm{AIC}_{\mathrm{SO}}$ is defined by

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

In addition, for the risk function R given by (2.8), it holds that

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

Remark 4.1. The $\mathrm{AIC}_{\mathrm{SO}}$ is derived under the order restriction (2.2). However, we can also derive the $\mathrm{AIC}_{\mathrm{SO}}$ even if we change a part of inequalities in (2.2) to " = ". For example, when $k = 4$ we can derive the $\mathrm{AIC}_{\mathrm{SO}}$ for the model (2.1) with

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

In this case, putting $N_1^* = N_1 + N_2$, $N_2^* = N_3 + N_4$, $\theta_1 = \theta_2 = \mu_1$ and $\theta_3 = \theta_4 = \mu_2$, and replacing

$$\begin{aligned} X_{11}, \dots, X_{1N_1}, X_{21}, \dots, X_{2N_2} &\to Y_{11}, \dots, Y_{1N_1^*}, \\ X_{31}, \dots, X_{3N_3}, X_{41}, \dots, X_{4N_4} &\to Y_{21}, \dots, Y_{2N_2^*}, \end{aligned}$$

the model (2.1) under (4.3) is equal to the model

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

under the restriction $\mu_1 \leq \mu_2$. Hence, by using the same argument, we can derive the $\mathrm{AIC}_{\mathrm{SO}}$.

Remark 4.2. The $\mathrm{AIC}_{\mathrm{SO}}$ is an asymptotically unbiased estimator of the risk function R and the order of the bias is N^{-1}. Similarly, for the ordinal ANOVA model without any order restriction, the ordinal AIC is also an asymptotically unbiased estimator of the risk function $\bar{R}$, and the order of the bias is N^{-1}. Thus, the $\mathrm{AIC}_{\mathrm{SO}}$ is as good as the AIC from the viewpoint of estimation accuracy of risk functions. In addition, the penalty term of $\mathrm{AIC}_{\mathrm{SO}}$ is $2(\hat{m} + 1)$, and from (4.2), $\hat{m}$ is simply defined as the function of the MLE. Therefore, also from the viewpoint of usefulness, the $\mathrm{AIC}_{\mathrm{SO}}$ is as good as AIC.

5. AIC$_{\mathrm{SO}}$ for several special cases

In this section, we provide the $\mathrm{AIC}_{\mathrm{SO}}$ for several special cases.

5.1. AIC$_{\mathrm{SO}}$ when the true variance σ_*^2 is known

In this subsection, we assume that the true variance σ_*^2 is known in ANOVA model (2.1). Then, under this assumption and the SO, the MLEs $\hat{\theta}_1, \dots, \hat{\theta}_k$ of $\theta_1, \dots, \theta_k$ are given by (2.7) because (2.6) does not depend on the variance parameter. Furthermore, in this case, the risk function based on the K-L divergence, R_1 can be written by replacing $\hat{\sigma}^2$ with σ_*^2 in (2.8) as

$$\begin{aligned} R_1 &= \mathrm{E}[\mathrm{E}_\star[-2l(\hat{\boldsymbol{\theta}}, \sigma_*^2; \boldsymbol{X}^\star)]] \\ &= \mathrm{E}\left[N \log(2\pi\sigma_*^2) + N + \frac{\sum_{i=1}^k N_i(\theta_{i,*} - \hat{\theta}_i)^2}{\sigma_*^2}\right] \\ &= N \log(2\pi\sigma_*^2) + N + \mathrm{E}\left[\frac{\sum_{i=1}^k N_i(\theta_{i,*} - \bar{X}_i + \bar{X}_i - \hat{\theta}_i)^2}{\sigma_*^2}\right] \\ &= N \log(2\pi\sigma_*^2) + N + k - 2Q + \mathrm{E}\left[\frac{\sum_{i=1}^k N_i(\bar{X}_i - \hat{\theta}_i)^2}{\sigma_*^2}\right]. \end{aligned} \tag{5.1}$$

Note that under the ordinal ANOVA model without the SO, when σ_*^2 is known the risk $\bar{R}_1$ is given by $\bar{R}_1 = \mathrm{E}[\mathrm{E}_\star[-2l(\bar{\boldsymbol{X}}, \sigma_*^2; \boldsymbol{X}^\star)]]$. Here, from (2.4), the maximum log-likelihood $l(\hat{\boldsymbol{\theta}}, \sigma_*^2; \boldsymbol{X})$ can be expressed as

$$\begin{aligned} & l(\hat{\boldsymbol{\theta}}, \sigma_*^2; \boldsymbol{X}) \\ &= -\frac{N}{2}\log(2\pi\sigma_*^2) - \frac{1}{2\sigma_*^2}\sum_{i=1}^{k}\sum_{j=1}^{N_i}(X_{ij} - \bar{X}_i)^2 - \frac{1}{2\sigma_*^2}\sum_{i=1}^{k} N_i(\bar{X}_i - \hat{\theta}_i)^2. \end{aligned} \tag{5.2}$$

Hence, the bias B_1 which is the difference between the expected value of $-2l(\hat{\boldsymbol{\theta}}, \sigma_*^2; \boldsymbol{X})$ and R_1, can be expressed as

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

Recall that the random variable $\hat{m}$ given by (4.2) satisfies $\mathrm{E}[k - \hat{m}] = Q$. Therefore, we obtain the following corollary.

Corollary 5.1. Let $l(\hat{\boldsymbol{\theta}}, \sigma_*^2; \boldsymbol{X})$ be the maximum log-likelihood given by (5.2), and let $\hat{m}$ be the random variable given by (4.2). Then, under ANOVA model (2.1) with the SO and known variance σ_*^2, the $\mathrm{AIC_{SO}}$ is given by

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

Moreover, it holds that

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

where R_1 is the risk function given by (5.1).

Remark 5.1. When the true variance σ_*^2 is known, the $\mathrm{AIC_{SO}}$ is an "unbiased" estimator of the risk function R_1. In addition, under the ordinal ANOVA model without the SO, when σ_*^2 is known the ordinal AIC is an "unbiased" estimator of the risk function $\bar{R}_1$.

5.2. $\mathrm{AIC_{SO}}$ with known variance weights

In this subsection, we consider the following model:

$$X_{ij} \sim N(\theta_i, \iota_i\sigma^2), \quad (i = 1, \ldots, k,\ j = 1, \ldots, N_i), \tag{5.3}$$

where $\theta_1, \ldots, \theta_k$ and σ^2 are unknown parameters, and $\iota_1, \ldots, \iota_k$ are known positive weights. Furthermore, also in this model, we assume the SO given by (2.2) for the parameters $\theta_1, \ldots, \theta_k$. Here, let

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

Note that under the ordinal ANOVA model (5.3) without the SO, the MLEs of θ_i and σ^2 are given by $\bar{\bar{X}}_i$ and $\bar{\bar{\sigma}}^2$, respectively. In this setting, we put $\bar{\bar{\boldsymbol{X}}} = (\bar{\bar{X}}_1, \ldots, \bar{\bar{X}}_k)'$. Next, define $n_i = N_i/\iota_i$, for any i $(1 \le i \le k)$. Then, $\bar{\bar{X}}_i$ is distributed as $N(\theta_i, \sigma^2/n_i)$. Therefore, putting $\boldsymbol{n} = (n_1, \ldots, n_k)'$ and $\boldsymbol{\iota} = (\iota_1, \ldots, \iota_k)'$, the log-likelihood function $l(\boldsymbol{\theta}, \sigma^2; \boldsymbol{X}, \boldsymbol{\iota})$ can be written by

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

Thus, by using the same argument as in Subsection 2.1, the MLEs of θ_i and σ^2, $\hat{\hat{\theta}}_i$ and $\hat{\hat{\sigma}}^2$, respectively, are give by

$$
\begin{aligned}
\hat{\hat{\theta}}_i &= \min_{v; i \le v} \max_{u; u \le i} \frac{\sum_{j=u}^{v} n_j \bar{\bar{X}}_j}{\sum_{j=u}^{v} n_j}, \quad (i = 1, \ldots, k), \\
\hat{\hat{\sigma}}^2 &= \frac{1}{N}\sum_{i=1}^{k}\frac{1}{\iota_i}\sum_{j=1}^{N_i}(X_{ij} - \bar{\bar{X}}_i)^2 + \frac{1}{N}\sum_{i=1}^{k} n_i(\bar{\bar{X}}_i - \hat{\hat{\theta}}_i)^2.
\end{aligned}
$$

Next, we put $\hat{\hat{\boldsymbol{\theta}}} = (\hat{\hat{\theta}}_1, \ldots, \hat{\hat{\theta}}_k)'$ and $R_2 = \mathrm{E}[\mathrm{E}_\star[-2l(\hat{\hat{\boldsymbol{\theta}}}, \hat{\hat{\sigma}}^2; \boldsymbol{X}^\star, \boldsymbol{\iota})]]$ where R_2 is the risk function. Furthermore, the maximum log-likelihood $l(\hat{\hat{\boldsymbol{\theta}}}, \hat{\hat{\sigma}}^2; \boldsymbol{X}, \boldsymbol{\iota})$ is given by

$$
l(\hat{\hat{\boldsymbol{\theta}}}, \hat{\hat{\sigma}}^2; \boldsymbol{X}, \boldsymbol{\iota}) = -\frac{1}{2}\sum_{i=1}^{k} N_i \log \iota_i - \frac{N}{2}\log(2\pi\hat{\hat{\sigma}}^2) - \frac{N}{2}. \tag{5.4}
$$

Moreover, by using the same argument as in Subsection 2.2, the bias $B_2 = \mathrm{E}[R_2 - \{-2l(\hat{\hat{\boldsymbol{\theta}}}, \hat{\hat{\sigma}}^2; \boldsymbol{X}, \boldsymbol{\iota})\}]$ can be expressed as

$$
B_2 = 2(k+1) - \frac{2N}{N-k-2}\mathrm{E}\left[\frac{1}{\sigma_*^2}\sum_{i=1}^{k} n_i(\bar{\bar{X}}_i - \theta_{i,*})(\bar{\bar{X}}_i - \hat{\hat{\theta}}_i)\right] + O(N^{-1}).
$$

Here, define

$$
\mathcal{M}^* = \bigcup_{i=1}^{k}\{\hat{\hat{\theta}}_i\}, \quad m^* = \#\mathcal{M}^*. \tag{5.5}
$$

Then, we obtain the following corollary.

Corollary 5.2. Let $l(\hat{\hat{\boldsymbol{\theta}}}, \hat{\hat{\sigma}}^2; \boldsymbol{X}, \boldsymbol{\iota})$ be the log-likelihood given by (5.4), and let m^* be the random variable given by (5.5). Then, under ANOVA model (5.3) with the SO, the

$\mathrm{AIC}_{\mathrm{SO}}$ is given by

$$\mathrm{AIC}_{\mathrm{SO}} = -2l(\hat{\hat{\boldsymbol{\theta}}}, \hat{\hat{\sigma}}^2; \boldsymbol{X}, \boldsymbol{\iota}) + 2(m^* + 1).$$

Furthermore, it holds that $\mathrm{E}[\mathrm{AIC}_{\mathrm{SO}}] = R_2 + O(N^{-1})$ where R_2 is the risk function defined by $R_2 = \mathrm{E}[\mathrm{E}_\star[-2l(\hat{\hat{\boldsymbol{\theta}}}, \hat{\hat{\sigma}}^2; \boldsymbol{X}^\star, \boldsymbol{\iota})]]$.

Remark 5.2. Under ANOVA model (5.3) with the SO, when σ_*^2 is known, the $\mathrm{AIC}_{\mathrm{SO}}$ can be derived as $\mathrm{AIC}_{\mathrm{SO}} = -2l(\hat{\hat{\boldsymbol{\theta}}}, \sigma_*^2; \boldsymbol{X}, \boldsymbol{\iota}) + 2m^*$. Furthermore, for the risk function $R_3 = \mathrm{E}[\mathrm{E}_\star[-2l(\hat{\hat{\boldsymbol{\theta}}}, \sigma_*^2; \boldsymbol{X}^\star, \boldsymbol{\iota})]]$, it holds that $\mathrm{E}[\mathrm{AIC}_{\mathrm{SO}}] = R_3$.

5.3. Multivariate ANOVA model with the SO

Let $\boldsymbol{V}_j^{(i)} = (V_{j1}^{(i)}, \ldots, V_{jp}^{(i)})'$ be a p-dimensional random vector on the jth individual in the ith cluster, where $i = 1, \ldots, k$ and $j = 1, \ldots, N_i$. Here, let $k \geq 2,\ p \geq 2$ and $N = N_1 + \cdots + N_k$. In this setting, we assume $N - k - 6 > 0$. Moreover, we assume that $\boldsymbol{V}_1^{(1)}, \ldots, \boldsymbol{V}_{N_k}^{(k)}$ are mutually independent. Then, we consider the following model

$$\boldsymbol{V}_j^{(i)} \sim N_p(\boldsymbol{\omega} + \delta_i \boldsymbol{a}, \tau^2 \boldsymbol{I}_p + \rho \boldsymbol{a}\boldsymbol{a}'), \quad (\tau^2 > 0,\ \tau^2 + \rho \boldsymbol{a}'\boldsymbol{a} > 0), \tag{5.6}$$

where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_p)'$, and $\delta_1, \ldots, \delta_k$, τ^2 and ρ are unknown parameters. In addition, $\boldsymbol{I}_p$ is a $p \times p$ unit matrix, and $\boldsymbol{a} = (a_1, \ldots, a_p)'$ is a known non-zero vector. Here, without loss of generality, we may assume that $\delta_1 = 0$. Moreover, the parameters $\delta_1, \ldots, \delta_k$ are restricted as

$$\delta_1 \leq \delta_2 \leq \cdots \leq \delta_k. \tag{5.7}$$

In other words, we consider the SO restriction for the parameters $\delta_1, \ldots, \delta_k$. For example, under the model (5.6), when $\boldsymbol{a} = \boldsymbol{1}_p$ this model is a parallel profile model considered by Yokoyama and Fujikoshi (1993), where $\boldsymbol{1}_p$ is a p-dimensional vector of ones.

Next, we decompose the model (5.6). Let $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_p$ be p-dimensional vectors with $\boldsymbol{h}_u' \boldsymbol{h}_{u^*} = 0,\ (u \neq u^*)$, $\boldsymbol{h}_u' \boldsymbol{h}_u = 1$ and $\boldsymbol{h}_1 = (\boldsymbol{a}'\boldsymbol{a})^{-1/2} \boldsymbol{a}$. Define $\boldsymbol{H}_2 = (\boldsymbol{h}_2, \ldots, \boldsymbol{h}_p)$ and $\boldsymbol{H} = (\boldsymbol{h}_1, \boldsymbol{H}_2)$. Then, considering $\boldsymbol{H}'\boldsymbol{V}_j^{(i)}$ we get

$$\boldsymbol{h}_1' \boldsymbol{V}_j^{(i)} \sim N(\boldsymbol{h}_1' \boldsymbol{\omega} + (\boldsymbol{a}'\boldsymbol{a})^{1/2} \delta_i, \tau^2 + \rho \boldsymbol{a}'\boldsymbol{a}), \quad (1 \leq i \leq k,\ 1 \leq j \leq N_i), \tag{5.8}$$

and

$$\boldsymbol{H}_2' \boldsymbol{V}_j^{(i)} \sim N_{p-1}(\boldsymbol{H}_2' \boldsymbol{\omega}, \tau^2 \boldsymbol{I}_{p-1}), \quad (1 \leq i \leq k,\ 1 \leq j \leq N_i). \tag{5.9}$$

Here, we replace $\boldsymbol{h}_1' \boldsymbol{V}_j^{(i)}$ with Y_{ij}. In addition, we put $\boldsymbol{h}_1' \boldsymbol{\omega} + (\boldsymbol{a}'\boldsymbol{a})^{1/2} \delta_i = \vartheta_i$ and $\tau^2 + \rho \boldsymbol{a}'\boldsymbol{a} = \varsigma^2$. Then, the model (5.8) is equal to

$$Y_{ij} \sim N(\vartheta_i, \varsigma^2), \quad (\varsigma^2 > 0,\ 1 \leq i \leq k,\ 1 \leq j \leq N_i), \tag{5.10}$$

and the parameters $\vartheta_1, \dots, \vartheta_k$ are restricted as

$$\vartheta_1 \le \vartheta_2 \le \dots \le \vartheta_k.$$

Furthermore, since $\boldsymbol{H}_2'\boldsymbol{V}_1^{(1)} \dots, \boldsymbol{H}_2'\boldsymbol{V}_{N_k}^{(k)}$ are independent and identically distributed, putting $\boldsymbol{H}_2'\boldsymbol{\omega} = (\mu_1, \dots, \mu_{p-1})' = \boldsymbol{\mu}$ the model (5.9) can be expressed as

$$Z_{st} \sim N(\mu_s, \tau^2), \quad (\tau^2 > 0,\ 1 \le s \le p-1,\ 1 \le t \le N). \tag{5.11}$$

Note that $Y_{11}, \dots, Y_{kN_k}, Z_{11}, \dots, Z_{(p-1)N}$ are mutually independent. Also note that the parameters $\mu_1, \dots, \mu_{p-1}$ are not restricted. Here, let

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

$$\bar{Z}_s = \frac{1}{N}\sum_{t=1}^{N} Z_{st},\ \bar{\tau}^2 = \frac{1}{N(p-1)}\sum_{s=1}^{p-1}\sum_{t=1}^{N}(Z_{st} - \bar{Z}_s)^2, \quad (s = 1, \dots, p-1),$$

and let $\bar{\boldsymbol{Y}} = (\bar{Y}_1, \dots, \bar{Y}_k)'$. Since μ_s and τ^2 are not restricted, it is easily checked that the MLEs of μ_s and τ^2 are $\bar{Z}_s$ and $\bar{\tau}^2$, respectively.

Next, we put $\boldsymbol{Y} = (Y_{11}, \dots, Y_{kN_k})'$, $\boldsymbol{Z} = (Z_{11}, \dots, Z_{(p-1)N})'$ and $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_k)'$. Then, the log-likelihood function $l(\boldsymbol{\vartheta}, \varsigma^2; \boldsymbol{Y})$ of $\boldsymbol{Y}$, is given by

$$\begin{aligned} l(\boldsymbol{\vartheta}, \varsigma^2; \boldsymbol{Y}) &= -\frac{N}{2}\log(2\pi\varsigma^2) - \frac{1}{2\varsigma^2}\sum_{i=1}^{k}\sum_{j=1}^{N_i}(Y_{ij} - \vartheta_i)^2 \\ &= -\frac{N}{2}\log(2\pi\varsigma^2) - \frac{1}{2\varsigma^2}\sum_{i=1}^{k}\sum_{j=1}^{N_i}(Y_{ij} - \bar{Y}_i)^2 - \frac{1}{2\varsigma^2}\sum_{i=1}^{k} N_i(\bar{Y}_i - \vartheta_i)^2. \end{aligned}$$

Similarly, the log-likelihood function $l(\boldsymbol{\mu}, \tau^2; \boldsymbol{Z})$ of $\boldsymbol{Z}$ is given by

$$l(\boldsymbol{\mu}, \tau^2; \boldsymbol{Z}) = -\frac{N(p-1)}{2}\log(2\pi\tau^2) - \frac{1}{2\tau^2}\sum_{s=1}^{p-1}\sum_{t=1}^{N}(Z_{st} - \mu_s)^2.$$

By using the same argument as in Subsection 2.1, the MLEs of ϑ_i and ς^2, $\hat{\vartheta}_i$ and $\hat{\varsigma}^2$ can be expressed as

$$\begin{aligned} \hat{\vartheta}_i &= \min_{v;i\le v}\max_{u;u\le i}\frac{\sum_{j=u}^{v} N_j\bar{Y}_j}{\sum_{j=u}^{v} N_j}, \\ \hat{\varsigma}^2 &= \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{N_i}(Y_{ij} - \bar{Y}_i)^2 + \frac{1}{N}\sum_{i=1}^{k} N_i(\bar{Y}_i - \hat{\vartheta}_i)^2, \end{aligned}$$

respectively. Note that the joint log-likelihood of $\boldsymbol{Y}$ and $\boldsymbol{Z}$, $l(\boldsymbol{\vartheta}, \varsigma^2, \boldsymbol{\mu}, \tau^2; \boldsymbol{Y}, \boldsymbol{Z})$ satisfies that $l(\boldsymbol{\vartheta}, \varsigma^2, \boldsymbol{\mu}, \tau^2; \boldsymbol{Y}, \boldsymbol{Z}) = l(\boldsymbol{\vartheta}, \varsigma^2; \boldsymbol{Y}) + l(\boldsymbol{\mu}, \tau^2; \boldsymbol{Z})$ because $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are independent. Here, let $\boldsymbol{Y}^\star$ and $\boldsymbol{Z}^\star$ be random vectors satisfying $(\boldsymbol{Y}^\star, \boldsymbol{Z}^\star) \sim$ i.d.d. $(\boldsymbol{Y}, \boldsymbol{Z})$. Then, the risk function R_4 can be written as $R_4 = \mathrm{E}[\mathrm{E}_\star[-2l(\hat{\boldsymbol{\vartheta}}, \hat{\varsigma}^2, \bar{\boldsymbol{Z}}, \bar{\tau}^2; \boldsymbol{Y}^\star, \boldsymbol{Z}^\star)]] = \mathrm{E}[\mathrm{E}_\star[-2l(\hat{\boldsymbol{\vartheta}}, \hat{\varsigma}^2; \boldsymbol{Y}^\star)]] + \mathrm{E}[\mathrm{E}_\star[-2l(\bar{\boldsymbol{Z}}, \bar{\tau}^2; \boldsymbol{Z}^\star)]]$, where , $\hat{\boldsymbol{\vartheta}} = (\hat{\vartheta}_1, \ldots, \hat{\vartheta}_k)'$ and $\bar{\boldsymbol{Z}} = (\bar{Z}_1, \ldots, \bar{Z}_{p-1})'$. In order to calculate the bias which is the difference between the expected value of $-2l(\hat{\boldsymbol{\vartheta}}, \hat{\varsigma}^2, \bar{\boldsymbol{Z}}, \bar{\tau}^2; \boldsymbol{Y}, \boldsymbol{Z})$ and R_4, it is sufficient to calculate

$$\mathrm{E}[\mathrm{E}_\star[-2l(\hat{\boldsymbol{\vartheta}}, \hat{\varsigma}^2; \boldsymbol{Y}^\star)] + 2l(\hat{\boldsymbol{\vartheta}}, \hat{\varsigma}^2; \boldsymbol{Y})],$$

and

$$\mathrm{E}[\mathrm{E}_\star[-2l(\bar{\boldsymbol{Z}}, \bar{\tau}^2; \boldsymbol{Z}^\star)] + 2l(\bar{\boldsymbol{Z}}, \bar{\tau}^2; \boldsymbol{Z})].$$

Here, it is easily checked that

$$\mathrm{E}[\mathrm{E}_\star[-2l(\bar{\boldsymbol{Z}}, \bar{\tau}^2; \boldsymbol{Z}^\star)] + 2l(\bar{\boldsymbol{Z}}, \bar{\tau}^2; \boldsymbol{Z})] = 2p + O(N^{-1}).$$

On the other hand, define

$$\mathcal{M}^\dagger = \bigcup_{i=1}^{k} \{\hat{\vartheta}_i\}, \quad m^\dagger = \#\mathcal{M}^\dagger. \tag{5.12}$$

Then, using the same argument as in Section 4, we have

$$\mathrm{E}[\mathrm{E}_\star[-2l(\hat{\boldsymbol{\vartheta}}, \hat{\varsigma}^2; \boldsymbol{Y}^\star)] + 2l(\hat{\boldsymbol{\vartheta}}, \hat{\varsigma}^2; \boldsymbol{Y})] = 2(m^\dagger + 1) + O(N^{-1}). \tag{5.13}$$

Therefore, we obtain the following corollary.

Corollary 5.3. Under the model (5.6) with the order restriction (5.7), the $\mathrm{AIC}_{\mathrm{SO}}$ is given by

$$\mathrm{AIC}_{\mathrm{SO}} = -2l(\hat{\boldsymbol{\vartheta}}, \hat{\varsigma}^2, \bar{\boldsymbol{Z}}, \bar{\tau}^2; \boldsymbol{Y}, \boldsymbol{Z}) + 2(m^\dagger + 1 + p).$$

Furthermore, it holds that $\mathrm{E}[\mathrm{AIC}_{\mathrm{SO}}] = R_4 + O(N^{-1})$.

Remark 5.3. Under the model (5.6) with the order restriction (5.7), when both τ_*^2 and ρ_* are known (i.e., both ς_*^2 and τ_*^2 are known), the $\mathrm{AIC}_{\mathrm{SO}}$ can be derived as $\mathrm{AIC}_{\mathrm{SO}} = -2l(\hat{\boldsymbol{\vartheta}}, \varsigma_*^2, \bar{\boldsymbol{Z}}, \tau_*^2; \boldsymbol{Y}, \boldsymbol{Z}) + 2(m^\dagger + p - 1)$. Moreover, for the risk function $R_5 = \mathrm{E}[\mathrm{E}_\star[-2l(\hat{\boldsymbol{\vartheta}}, \varsigma_*^2, \bar{\boldsymbol{Z}}, \tau_*^2; \boldsymbol{Y}^\star, \boldsymbol{Z}^\star)]]$, it holds that $\mathrm{E}[\mathrm{AIC}_{\mathrm{SO}}] = R_5$.

We introduced six models thus far. In other words, the model (2.1) when σ_*^2 is unknown (Case A), and known (Case B). Moreover, the model (5.3) when σ_*^2 is unknown (Case C), and known (Case D). Finally, the model (5.10) and (5.11) when both ς_*^2 and τ_*^2 are unknown (Case E), and known (Case F). The properties of the $\mathrm{AIC}_{\mathrm{SO}}$ and the ordinal AIC for these six models are summarized in Table 5.1.

5.4. Comparison of the $\mathrm{AIC}_{\mathrm{SO}}$ and the pseudo AIC (pAIC) under certain candidate models

Let k be an integer with $k \geq 2$, and let $\mathcal{W}_i^k$ be the set defined as in Subsection 3.1 where i is an integer with $1 \leq i \leq k$. Moreover, for any i ($i = 1, \dots, k$) and for any $\boldsymbol{w} \in \mathcal{W}_i^k$, we define a set $\mathcal{C}_i^k(\boldsymbol{w})$ as follows. First, in the case of $i = 1$, $\mathcal{W}_1^k$ has the unique element $\boldsymbol{w} = (k)'$, and we define

$$\mathcal{C}_1^k(\boldsymbol{w}) = \{(x_1, \dots, x_k)' \in \mathbb{R}^k \mid x_1 = x_2 = \cdots = x_k\} = \mathcal{C}_1^k.$$

On the other hand, in the case of $2 \leq i \leq k$, for any element $\boldsymbol{w} = (w_1, \dots, w_i)'$ in $\mathcal{W}_i^k$, we define

$$\mathcal{C}_i^k(\boldsymbol{w}) = \{(x_1, \dots, x_k)' \in \mathbb{R}^k \mid 1 \leq t \leq i-1,\ x_{w_t} \leq x_{w_{t+1}}, \quad 0 \leq s \leq i-1,\ w_0 = 0,\ x_{1+w_s} = \cdots = x_{w_{s+1}}\}. \tag{5.14}$$

Thus, from (5.14), the element $\boldsymbol{x} = (x_1, \dots, x_k)'$ in $\mathcal{C}_i^k(\boldsymbol{w})$ satisfies

$$x_1 = \cdots = x_{w_1} \leq x_{1+w_1} = \cdots = x_{w_2} \leq \cdots \leq x_{1+w_{i-1}} = \cdots = x_k.$$

In particular, when $i = k$, $\mathcal{W}_k^k$ has the unique element $\boldsymbol{w} = (w_1, \dots, w_k)' = (1, \dots, k)'$, and it holds that

$$\mathcal{C}_k^k(\boldsymbol{w}) = \{(x_1, \dots, x_k)' \in \mathbb{R}^k \mid x_1 \leq x_2 \leq \cdots \leq x_k\} = \mathcal{C}_k^k.$$

Here, let X_{st} be independent random variables where $s = 1, \dots, k$ and $t = 1, \dots, N_s$. Then, for any i with $1 \leq i \leq k$ and for any $\boldsymbol{w} \in \mathcal{W}_i^k$, we consider ANOVA model

$$X_{st} \sim N(\theta_s, \sigma^2), \quad (s = 1, \dots, k, \quad t = 1, \dots, N_s),$$

with $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)' \in \mathcal{C}_i^k(\boldsymbol{w})$. For example, when $k = 5$, $i = 3$ and $\boldsymbol{w} = (w_1, w_2, w_3)' = (1, 3, 5)' \in \mathcal{C}_3^5$, above model is equal to ANOVA model with $\theta_1 \leq \theta_2 = \theta_3 \leq \theta_4 = \theta_5$. Recall that the number of elements of $\mathcal{W}_i^k$ is ${}_{k-1}C_{i-1}$. Hence, it holds that

$$\sum_{i=1}^{k} \#\mathcal{W}_i^k = 2^{k-1}.$$

This implies that we can consider 2^{k-1} models. In this subsection, these models are candidate models.

Next, we consider the $\mathrm{AIC}_{\mathrm{SO}}$ and the pseudo AIC (pAIC) for these candidate models. Recall that the pAIC is defined as

$$\mathrm{pAIC} = -2l(\hat{\theta}) + 2p,$$

where $l(\cdot)$ is a maximum log-likelihood, p is the number of independent parameters in the candidate model, and $\hat{\theta}$ is the MLE which is derived under the order restricted model. In this setting, we define the minimum $\mathrm{AIC}_{\mathrm{SO}}$ model and the minimum pAIC model. Let $\mathcal{M}_1, \ldots, \mathcal{M}_{2^{k-1}}$ be candidate models, and let $\mathrm{AIC}_{\mathrm{SO}}(\mathcal{M}_q)$ and $\mathrm{pAIC}(\mathcal{M}_q)$ be values of the $\mathrm{AIC}_{\mathrm{SO}}$ and the pAIC in the candidate model $\mathcal{M}_q$, respectively. Then, we define that the candidate model $\mathcal{M}_q$ is the minimum $\mathrm{AIC}_{\mathrm{SO}}$ model if $\mathcal{M}_q$ satisfies the following two conditions:

(m1) For any candidate model $\mathcal{M}_{q^*}$, it holds that $\mathrm{AIC}_{\mathrm{SO}}(\mathcal{M}_q) \leq \mathrm{AIC}_{\mathrm{SO}}(\mathcal{M}_{q^*})$.

(m2) For any candidate model $\mathcal{M}_{q^\star}$ with $\mathrm{AIC}_{\mathrm{SO}}(\mathcal{M}_q) = \mathrm{AIC}_{\mathrm{SO}}(\mathcal{M}_{q^\star})$, it holds that $\#(\mathcal{M}_q) \leq \#(\mathcal{M}_{q^\star})$ where $\#(\mathcal{M}_j)$ is the number of independent parameters in the candidate model $\mathcal{M}_j$, $(j = 1, \ldots, 2^{k-1})$.

Similarly, by replacing $\mathrm{AIC}_{\mathrm{SO}}$ with pAIC in the conditions (m1) and (m2), we also define the minimum pAIC model. Then, the following theorem holds.

Theorem 5.1. Let k (≥ 2) be an integer, and let $\mathcal{M}_1, \ldots, \mathcal{M}_{2^{k-1}}$ be candidate models defined as in Subsection 5.4. Then, the minimum $\mathrm{AIC}_{\mathrm{SO}}$ model is equal to the minimum pAIC model.

Proof. Let $\mathcal{M}_q$ be the minimum $\mathrm{AIC}_{\mathrm{SO}}$ model, and let $\hat{\theta}_1^{(q)}, \ldots, \hat{\theta}_k^{(q)}$ be the MLEs of $\theta_1, \ldots, \theta_k$ in the model $\mathcal{M}_q$, respectively. First, we consider the case of not $\hat{\theta}_1^{(q)} = \cdots = \hat{\theta}_k^{(q)}$. Hence, there exists a number i $(2 \leq i \leq k)$ and natural numbers $w_1, \ldots, w_i$ with $w_1 < \cdots < w_i$ where $w_i = k$ such that

$$\hat{\theta}_{w_{j-1}+1}^{(q)} = \cdots = \hat{\theta}_{w_j}^{(q)} = \frac{\sum_{s=w_{j-1}+1}^{w_j} N_s \bar{X}_s}{\sum_{s=w_{j-1}+1}^{w_j} N_s}, \quad (j = 1, \ldots, i), \tag{5.15}$$

and $\hat{\theta}_{w_1}^{(q)} < \cdots < \hat{\theta}_{w_i}^{(q)}$. Note that $w_0 = 0$. Here, let $l^{(q)}$ be a $-2\times$ maximum log-likelihood in the model $\mathcal{M}_q$. Furthermore, from (5.15) it holds that $\hat{m} = i$. Therefore, $\mathrm{AIC}_{\mathrm{SO}}(\mathcal{M}_q)$ can be written as

$$\mathrm{AIC}_{\mathrm{SO}}(\mathcal{M}_q) = l^{(q)} + 2(1 + i).$$

Moreover, from the definition of the minimum $\mathrm{AIC}_{\mathrm{SO}}$ model, the model $\mathcal{M}_q$ is ANOVA model with

$$\theta_{w_0+1} = \cdots = \theta_{w_1} \leq \theta_{w_1+1} = \cdots = \theta_{w_2} \leq \cdots \leq \theta_{w_{i-1}+1} = \cdots = \theta_{w_i}.$$

In this model, the number of independent parameters is $i + 1$. Thus, $\mathrm{pAIC}(\mathcal{M}_q)$ is also $l^{(q)} + 2(1 + i)$. Hence, we get $\mathrm{AIC}_{\mathrm{SO}}(\mathcal{M}_q) = \mathrm{pAIC}(\mathcal{M}_q)$. On the other hand, from the definition of $\mathcal{M}_q$ it holds that

$$\begin{aligned}\mathrm{pAIC}(\mathcal{M}_q) = \mathrm{AIC}_{\mathrm{SO}}(\mathcal{M}_q) &= \min_{1 \leq u \leq 2^{k-1}} \mathrm{AIC}_{\mathrm{SO}}(\mathcal{M}_u) \\ &\leq \min_{1 \leq u \leq 2^{k-1},\ u \neq q} \mathrm{AIC}_{\mathrm{SO}}(\mathcal{M}_u). \end{aligned} \tag{5.16}$$

Furthermore, from the definitions of the $\mathrm{AIC}_{\mathrm{SO}}$ and the pAIC, it is clear that $\mathrm{AIC}_{\mathrm{SO}}(\mathcal{M}_u) \leq \mathrm{pAIC}(\mathcal{M}_u)$. Therefore, combining this inequality and (5.16), we obtain

$$\mathrm{pAIC}(\mathcal{M}_q) \leq \min_{1 \leq u \leq 2^{k-1},\ u \neq q} \mathrm{AIC}_{\mathrm{SO}}(\mathcal{M}_u) \leq \min_{1 \leq u \leq 2^{k-1},\ u \neq q} \mathrm{pAIC}(\mathcal{M}_u).$$

Hence, for any candidate model $\mathcal{M}_u$, it holds that $\mathrm{pAIC}(\mathcal{M}_q) \leq \mathrm{pAIC}(\mathcal{M}_u)$. In addition, for any candidate model $\mathcal{M}_{u^\star}$ with $\mathrm{pAIC}(\mathcal{M}_q) = \mathrm{pAIC}(\mathcal{M}_{u^\star})$, it holds that $\#(\mathcal{M}_q) \leq \#(\mathcal{M}_{u^\star})$. In fact, if $\mathrm{pAIC}(\mathcal{M}_q) = \mathrm{pAIC}(\mathcal{M}_{u^\star})$ and $i^\star = \#(\mathcal{M}_{u^\star}) < \#(\mathcal{M}_q) = i$, it holds that

$$\mathrm{AIC}_{\mathrm{SO}}(\mathcal{M}_q) = \mathrm{pAIC}(\mathcal{M}_q) = \mathrm{pAIC}(\mathcal{M}_{u^\star}) \geq \mathrm{AIC}_{\mathrm{SO}}(\mathcal{M}_{u^\star}).$$

However, since $i^\star = \#(\mathcal{M}_{u^\star}) < \#(\mathcal{M}_q) = i$, this implies that $\mathcal{M}_q$ is not the minimum $\mathrm{AIC}_{\mathrm{SO}}$ model. This is a contradiction. Hence, for any candidate model $\mathcal{M}_{u^\star}$ with $\mathrm{pAIC}(\mathcal{M}_q) = \mathrm{pAIC}(\mathcal{M}_{u^\star})$, it holds that $\#(\mathcal{M}_q) \leq \#(\mathcal{M}_{u^\star})$. Therefore, the minimum pAIC model is $\mathcal{M}_q$. Similarly, by using the same argument, we can also prove the case of $\hat{\theta}_1^{(q)} = \cdots = \hat{\theta}_k^{(q)}$. □

Recall that the $\mathrm{AIC}_{\mathrm{SO}}$ is the asymptotically "unbiased" estimator of the risk function. Furthermore, in general, the pAIC is the asymptotically "biased" estimator of the risk function. However, Theorem 5.1 means that the minimum $\mathrm{AIC}_{\mathrm{SO}}$ model based on the $\mathrm{AIC}_{\mathrm{SO}}$ is equal to the minimum pAIC model based on the pAIC although the $\mathrm{AIC}_{\mathrm{SO}}$ and pAIC are asymptotically unbiased and biased estimators, respectively. In other words, when we consider the model selection problem for these candidate models using the $\mathrm{AIC}_{\mathrm{SO}}$ or the pAIC, we may use the pAIC.

Remark 5.4. Needless to say, for these candidate models, we can also use the $\mathrm{AIC}_{\mathrm{SO}}$. Here, we would like to note that, in general, the result of Theorem 5.1 does not hold when the number of candidate models is smaller than 2^{k-1}. For example, when we consider the nested candidate models, in general, the minimum $\mathrm{AIC}_{\mathrm{SO}}$ model is not equal to the minimum pAIC model.

6. Numerical experiments

Let X_{ij} be a random variable distributed as $N(\theta_i, \sigma^2/N_i)$ where $1 \leq i \leq 4$, $1 \leq j \leq N_i$ and $N_1 = \cdots = N_4$. Moreover, let $N = N_1 + N_2 + N_3 + N_4$. In this section, we consider the following four candidate models:

$$\begin{aligned}&\text{Model 1}: \text{ ANOVA model with } \theta_1 = \theta_2 = \theta_3 = \theta_4,\\&\text{Model 2}: \text{ ANOVA model with } \theta_1 \le \theta_2 = \theta_3 = \theta_4,\\&\text{Model 3}: \text{ ANOVA model with } \theta_1 \le \theta_2 \le \theta_3 = \theta_4,\\&\text{Model 4}: \text{ ANOVA model with } \theta_1 \le \theta_2 \le \theta_3 \le \theta_4.\end{aligned}$$

Thus, these four models are nested. From 100,000 monte calro simulation runs, we compare performance of the $\mathrm{AIC_{SO}}$ and the pAIC. In the qth simulation, where $(1 \le q \le 100000)$, let $\hat{\theta}^{(q)}_{1,\mathrm{AIC_{SO}}}, \ldots, \hat{\theta}^{(q)}_{4,\mathrm{AIC_{SO}}}$ and $\hat{\sigma^2}^{(q)}_{\mathrm{AIC_{SO}}}$ be MLEs of the parameters $\theta_1, \ldots, \theta_4$ and σ^2 for the minimum $\mathrm{AIC_{SO}}$ model, respectively. Similarly, let $\hat{\theta}^{(q)}_{1,\mathrm{pAIC}}, \ldots, \hat{\theta}^{(q)}_{4,\mathrm{pAIC}}$ and $\hat{\sigma^2}^{(q)}_{\mathrm{pAIC}}$ be MLEs of the parameters $\theta_1, \ldots, \theta_4$ and σ^2 for the minimum pAIC model, respectively. Here, since the risk function of ANOVA model with the SO is given by (2.8), the estimator

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

is an unbiased estimator of the risk function. Based on this, we evaluate performance of the $\mathrm{AIC_{SO}}$ and the pAIC as

$$\begin{aligned}\mathrm{PE}_{\mathrm{AIC_{SO}}} &= \frac{1}{100000}\sum_{q=1}^{100000} R(\hat{\theta}^{(q)}_{1,\mathrm{AIC_{SO}}}, \ldots, \hat{\theta}^{(q)}_{4,\mathrm{AIC_{SO}}}, \hat{\sigma^2}^{(q)}_{\mathrm{AIC_{SO}}}),\\ \mathrm{PE}_{\mathrm{pAIC}} &= \frac{1}{100000}\sum_{q=1}^{100000} R(\hat{\theta}^{(q)}_{1,\mathrm{pAIC}}, \ldots, \hat{\theta}^{(q)}_{4,\mathrm{pAIC}}, \hat{\sigma^2}^{(q)}_{\mathrm{pAIC}}).\end{aligned}$$

Thus, the $\mathrm{PE}_{\mathrm{AIC_{SO}}}$ and the $\mathrm{PE}_{\mathrm{pAIC}}$ are estimated values of risk functions for the minimum $\mathrm{AIC_{SO}}$ model (the model selected by using the $\mathrm{AIC_{SO}}$) and the minimum pAIC model (the model selected by using the pAIC), respectively.

Next, in this simulation, we consider the following true models:

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

In Case 1, Model 3 and 4 include the true model, and in Case 2, Model 4 includes the true model. Moreover, in Case 3, Model 1, 2, 3 and 4 include the true model. For these cases, we set $N = 40$ and $N = 200$. The values of the $\mathrm{PE}_{\mathrm{AIC_{SO}}}$ and the $\mathrm{PE}_{\mathrm{pAIC}}$ in the cases 1–3 are given in Table 6.1–6.3, respectively.

From Table 6.1–6.3, we can see that the $\mathrm{AIC_{SO}}$ is an asymptotically unbiased estimator of the risk function. Recall that from the definitions of the $\mathrm{AIC_{SO}}$ and the pAIC, the value of the $\mathrm{AIC_{SO}}$ is equal to or smaller than that of the pAIC. We can confirm that this inequality holds for all cases. Moreover, for the values of the $\mathrm{PE}_{\mathrm{AIC}_{\mathrm{SO}}}$ and the $\mathrm{PE}_{\mathrm{pAIC}}$, from Table 6.1 and 6.2, we can see that the $\mathrm{PE}_{\mathrm{AIC}_{\mathrm{SO}}}$ is smaller than the $\mathrm{PE}_{\mathrm{pAIC}}$ in Case 1 and 2. Thus, compared with the pAIC, model selections using the $\mathrm{AIC}_{\mathrm{SO}}$ are better from the viewpoint of the risk for the selected model in Case 1 and 2. On the other hand, in case 3, the model selection using the pAIC is better.

7. Conclusion

In this paper, we derived the $\mathrm{AIC}_{\mathrm{SO}}$ for ANOVA model with the simple order restriction. We showed that the $\mathrm{AIC}_{\mathrm{SO}}$ is the asymptotically unbiased estimator of the risk function. Furthermore, we also showed that if the true variance is known, the $\mathrm{AIC}_{\mathrm{SO}}$ is the unbiased estimator. We would like to note that since the penalty term of the $\mathrm{AIC}_{\mathrm{SO}}$ is simply defined as the function of the MLEs, the $\mathrm{AIC}_{\mathrm{SO}}$ is very useful for analysts. Thus, from the viewpoint of usefulness and estimation accuracy of the risk function, the $\mathrm{AIC}_{\mathrm{SO}}$ is as good as the ordinal AIC. Furthermore, Theorem 5.1 shows that under certain candidate models, the selected models based on minimizing the $\mathrm{AIC}_{\mathrm{SO}}$ and the pAIC are the same model. In addition, from numerical experiments we could confirm that the $\mathrm{AIC}_{\mathrm{SO}}$ is an unbiased estimator of the risk function.

Appendix

In this section, we define several notations. Next, we show seven lemmas, Lemma A–G, and using Lemma F and Lemma G we prove Lemma 3.1.

First, we define the inequality of vectors. Let $\boldsymbol{x} = (x_1, \ldots, x_p)'$ and $\boldsymbol{y} = (y_1, \ldots, y_p)'$ be p-dimensional vectors, and let $\mathbf{0}_p$ be a p-dimensional vector of zeros. Then, define

$$\boldsymbol{x} \geq \mathbf{0}_p \Leftrightarrow {}^\forall i \in \{1, \ldots, p\},\ x_i \geq 0,$$
$$\boldsymbol{x} \geq \boldsymbol{y} \Leftrightarrow \boldsymbol{x} - \boldsymbol{y} \geq \mathbf{0}_p.$$

Furthermore, for some proposition P, we define an indicator function $1_{\{P\}}$ as

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

Appendix A: Lemma A and its proof

Lemma A. Let l be an integer with $l \geq 2$, and let $n_1, \ldots, n_l$ be elements of $\mathbb{R}_{>0}$. Let $\boldsymbol{n} = (n_1, \ldots, n_l)'$, and let $\boldsymbol{x} = (x_1, \ldots, x_l)'$ be a vector of $\mathbb{R}^l$. Then, the following (i), (ii) and (iii) hold:

(i) For all integers a, b and c with $1 \leq a \leq b < c \leq l$, it holds that

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

and

$$\bar{x}_{[a,b]} < \bar{x}_{[a,c]} \Leftrightarrow \bar{x}_{[a,b]} < \bar{x}_{[b+1,c]} \Leftrightarrow \bar{x}_{[a,c]} < \bar{x}_{[b+1,c]}. \tag{A.2}$$

(ii) Let i be an integer with $2 \leq i \leq l$, and let $w_1, \ldots, w_i$ be integers with $w_1 < w_2 < \cdots < w_i$ and $w_i = l$. Put $w_0 = 0$. Then, if

$$\bar{x}_{[1+w_0,w_1]} < \bar{x}_{[1+w_1,w_2]} < \cdots < \bar{x}_{[1+w_{i-1},w_i]} \tag{A.3}$$

is true, for all integers s and t with $1 \leq s < t \leq i$, it holds that

$$\bar{x}_{[1+w_{s-1},w_s]} < \bar{x}_{[1+w_{s-1},w_t]}. \tag{A.4}$$

(iii) Let i and j be integers with $1 \leq i < j \leq l$. Then, it holds that

$$\bar{x}_{[i,b]} \geq \bar{x}_{[b+1,j]}, \ ({}^{\forall} b \in \mathbb{N} \text{ with } i \leq b < j) \Leftrightarrow \boldsymbol{D}_{[i,j]}\boldsymbol{x}_{[i,j]} \geq \boldsymbol{0}_{j-i}. \tag{A.5}$$

Proof. First, we prove (i). Let a, b and c be integers with $1 \leq a \leq b < c \leq l$. In this setting, we show $\bar{x}_{[a,b]} < \bar{x}_{[a,c]} \Leftrightarrow \bar{x}_{[a,b]} < \bar{x}_{[b+1,c]}$. Let $\bar{x}_{[a,b]} < \bar{x}_{[a,c]}$, i.e., $\bar{x}_{[a,b]} - \bar{x}_{[a,c]} < 0$. Then, we get

$$\begin{aligned}
\bar{x}_{[a,b]} - \bar{x}_{[a,c]} &= \frac{\sum_{j=a}^{b} n_j x_j}{\tilde{n}_{[a,b]}} - \frac{\sum_{j=a}^{c} n_j x_j}{\tilde{n}_{[a,c]}} \\
&= \frac{\sum_{j=a}^{b} n_j x_j}{\tilde{n}_{[a,b]}} - \frac{\sum_{j=a}^{b} n_j x_j + \sum_{j=b+1}^{c} n_j x_j}{\tilde{n}_{[a,c]}} \\
&= \sum_{j=a}^{b} n_j x_j \left(\frac{1}{\tilde{n}_{[a,b]}} - \frac{1}{\tilde{n}_{[a,c]}} \right) - \frac{\sum_{j=b+1}^{c} n_j x_j}{\tilde{n}_{[a,c]}} \\
&= \sum_{j=a}^{b} n_j x_j \left(\frac{\tilde{n}_{[a,c]} - \tilde{n}_{[a,b]}}{\tilde{n}_{[a,b]}\tilde{n}_{[a,c]}} \right) - \frac{\sum_{j=b+1}^{c} n_j x_j}{\tilde{n}_{[a,c]}} \\
&= \sum_{j=a}^{b} n_j x_j \left(\frac{\tilde{n}_{[b+1,c]}}{\tilde{n}_{[a,b]}\tilde{n}_{[a,c]}} \right) - \frac{\sum_{j=b+1}^{c} n_j x_j}{\tilde{n}_{[a,c]}} \\
&= \frac{\tilde{n}_{[b+1,c]}}{\tilde{n}_{[a,c]}} \left(\frac{\sum_{j=a}^{b} n_j x_j}{\tilde{n}_{[a,b]}} - \frac{\sum_{j=b+1}^{c} n_j x_j}{\tilde{n}_{[b+1,c]}} \right) = \frac{\tilde{n}_{[b+1,c]}}{\tilde{n}_{[a,c]}} (\bar{x}_{[a,b]} - \bar{x}_{[b+1,c]}).
\end{aligned}$$

Hence, noting that $\tilde{n}_{[b+1,c]}/\tilde{n}_{[a,c]}$ is positive, we have

$$\begin{aligned}
\bar{x}_{[a,b]} - \bar{x}_{[a,c]} < 0 &\Leftrightarrow \frac{\tilde{n}_{[b+1,c]}}{\tilde{n}_{[a,c]}} (\bar{x}_{[a,b]} - \bar{x}_{[b+1,c]}) < 0 \\
&\Leftrightarrow \bar{x}_{[a,b]} - \bar{x}_{[b+1,c]} < 0 \Leftrightarrow \bar{x}_{[a,b]} < \bar{x}_{[b+1,c]}.
\end{aligned}$$

Therefore, it holds that $\bar{x}_{[a,b]} < \bar{x}_{[a,c]} \Leftrightarrow \bar{x}_{[a,b]} < \bar{x}_{[b+1,c]}$. Moreover, by considering its contraposition, $\bar{x}_{[a,b]} \geq \bar{x}_{[a,c]} \Leftrightarrow \bar{x}_{[a,b]} \geq \bar{x}_{[b+1,c]}$ also holds. In addition, noting that

$$\begin{aligned}
&\bar{x}_{[a,b]} \geq \bar{x}_{[b+1,c]} \\
\Leftrightarrow\ &\tilde{n}_{[a,b]}\tilde{n}_{[b+1,c]}\bar{x}_{[a,b]} \geq \tilde{n}_{[a,b]}\tilde{n}_{[b+1,c]}\bar{x}_{[b+1,c]} \\
\Leftrightarrow\ &\tilde{n}_{[b+1,c]}\sum_{j=a}^{b} n_j x_j \geq \tilde{n}_{[a,b]}\sum_{j=b+1}^{c} n_j x_j \\
\Leftrightarrow\ &\tilde{n}_{[b+1,c]}\sum_{j=a}^{b} n_j x_j + \tilde{n}_{[b+1,c]}\sum_{j=b+1}^{c} n_j x_j \geq \tilde{n}_{[a,b]}\sum_{j=b+1}^{c} n_j x_j + \tilde{n}_{[b+1,c]}\sum_{j=b+1}^{c} n_j x_j \\
\Leftrightarrow\ &\tilde{n}_{[b+1,c]}\sum_{j=a}^{c} n_j x_j \geq \tilde{n}_{[a,c]}\sum_{j=b+1}^{c} n_j x_j \\
\Leftrightarrow\ &\frac{\sum_{j=a}^{c} n_j x_j}{\tilde{n}_{[a,c]}} \geq \frac{\sum_{j=b+1}^{c} n_j x_j}{\tilde{n}_{[b+1,c]}} \Leftrightarrow \bar{x}_{[a,c]} \geq \bar{x}_{[b+1,c]},
\end{aligned}$$

we get $\bar{x}_{[a,b]} \geq \bar{x}_{[b+1,c]} \Leftrightarrow \bar{x}_{[a,c]} \geq \bar{x}_{[b+1,c]}$. Finally, by considering its contraposition, we also get $\bar{x}_{[a,b]} < \bar{x}_{[b+1,c]} \Leftrightarrow \bar{x}_{[a,c]} < \bar{x}_{[b+1,c]}$. Thus, it holds that (A.1) and (A.2).

Next, we prove (ii). Assume that (A.3) is true. Let s and t be integers with $1 \leq s < t \leq i$. When $t = 2$ and $s = 1$, from (A.3) it holds that $\bar{x}_{[1+w_{1-1},w_1]} < \bar{x}_{[1+w_1,w_2]}$. Moreover, from (A.2) $\bar{x}_{[1+w_{1-1},w_1]} < \bar{x}_{[1+w_1,w_2]}$ yields $\bar{x}_{[1+w_{1-1},w_1]} < \bar{x}_{[1+w_{1-1},w_2]}$. Thus, we get $\bar{x}_{[1+w_{s-1},w_s]} < \bar{x}_{[1+w_{s-1},w_t]}$. Hence, if $t = 2$, (ii) is proved. Therefore, we consider the case of $t \geq 3$. Since (A.3) is true, we obtain

$$\bar{x}_{[1,w_1]} < \cdots < \bar{x}_{[1+w_{t-3},w_{t-2}]} < \bar{x}_{[1+w_{t-2},w_{t-1}]} < \bar{x}_{[1+w_{t-1},w_t]}. \tag{A.6}$$

Here, using (A.2) and the last inequality of (A.6), $\bar{x}_{[1+w_{t-2},w_{t-1}]} < \bar{x}_{[1+w_{t-1},w_t]}$, we get

$$\bar{x}_{[1+w_{t-2},w_{t-1}]} < \bar{x}_{[1+w_{t-2},w_t]}. \tag{A.7}$$

Thus, if $s = t - 1$, (ii) is proved.

Finally, we consider the case of $s < t - 1$, i.e., there exists q with $q \geq 2$, such that $s = t - q$. Here, we put $v = t - 1$. Note that from (A.7) the inequality $\bar{x}_{[1+w_{v-1},w_v]} < \bar{x}_{[1+w_{v-1},w_t]}$ holds. In this setting, (ii) is proved as follows:

1. Combining $\bar{x}_{[1+w_{v-1},w_v]} < \bar{x}_{[1+w_{v-1},w_t]}$ and the inequality $\bar{x}_{[1+w_{v-2},w_{v-1}]} < \bar{x}_{[1+w_{v-1},w_v]}$ in (A.6), we obtain $\bar{x}_{[1+w_{v-2},w_{v-1}]} < \bar{x}_{[1+w_{v-1},w_t]}$.
2. Again, by using (A.2), we get $\bar{x}_{[1+w_{v-2},w_{v-1}]} < \bar{x}_{[1+w_{v-2},w_t]}$.
3. Here, if $v - 1 = s$, (ii) is proved. On the other hand, if $s < v - 1$, replacing $v - 1$ with v, and we go back to the step 1.

Therefore, using above method we obtain (ii).

Finally, we prove (iii). Let i and j be integers with $1 \leq i < j \leq l$, and let b be an integer with $i \leq b < j$, i.e., $i \leq b \leq j-1$. We put $s = b - i + 1$. Note that $1 \leq s \leq j - i$. Recall that from (3.4), the sth row of the matrix $\boldsymbol{D}_{i,j}$ is given

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

Therefore, the sth element of the vector $\boldsymbol{D}_{i,j}\boldsymbol{x}_{[i,j]}$ can be expressed as

$$\begin{aligned}
&\left(\frac{1}{\tilde{n}_{[i,i+s-1]}} \boldsymbol{n}'_{[i,i+s-1]}, \frac{-1}{\tilde{n}_{[i+s,j]}} \boldsymbol{n}'_{[i+s,j]} \right) \boldsymbol{x}_{[i,j]} \\
&= \left(\frac{1}{\tilde{n}_{[i,i+s-1]}} \boldsymbol{n}'_{[i,i+s-1]}, \frac{-1}{\tilde{n}_{[i+s,j]}} \boldsymbol{n}'_{[i+s,j]} \right) (\boldsymbol{x}'_{[i,i+s-1]}, \boldsymbol{x}'_{[i+s,j]})' \\
&= \frac{\boldsymbol{n}'_{[i,i+s-1]}\boldsymbol{x}_{[i,i+s-1]}}{\tilde{n}_{[i,i+s-1]}} - \frac{\boldsymbol{n}'_{[i+s,j]}\boldsymbol{x}_{[i+s,j]}}{\tilde{n}_{[i+s,j]}} = \bar{x}_{[i,i+s-1]} - \bar{x}_{[i+s,j]} = \bar{x}_{[i,b]} - \bar{x}_{[b+1,j]}.
\end{aligned}$$

Hence, if $\boldsymbol{D}_{i,j}\boldsymbol{x}_{[i,j]} \geq \boldsymbol{0}_{j-i}$, then we obtain $\bar{x}_{[i,b]} - \bar{x}_{[b+1,j]} \geq 0$, i.e., $\bar{x}_{[i,b]} \geq \bar{x}_{[b+1,j]}$. On the other hand, if $\bar{x}_{[i,b]} \geq \bar{x}_{[b+1,j]}$, i.e., $\bar{x}_{[i,b]} - \bar{x}_{[b+1,j]} \geq 0$ for any integer b with $i \leq b < j$, then, we get $\boldsymbol{D}_{i,j}\boldsymbol{x}_{[i,j]} \geq \boldsymbol{0}_{j-i}$. Thus, (A.5) holds. □

Appendix B: Lemma B and its proof

Lemma B. Let l be an integer with $l \geq 2$, and let $n_1, \ldots, n_l \in \mathbb{R}_{>0}$ and $\boldsymbol{n} = (n_1, \ldots, n_l)'$. Let $\boldsymbol{x} = (x_1, \ldots, x_l)' \in \mathbb{R}^l$, and let i be an integer with $2 \leq i \leq l$. In addition, let $w_1, \ldots, w_i \in \mathbb{N}$, and let $w_1 < w_2 < \cdots < w_i$ where $w_i = l$. Put $w_0 = 0$. Assume that

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

and

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

Moreover, also assume that

$$\bar{x}_{[1,w_1]} < \bar{x}_{[1+w_1,w_2]} < \cdots < \bar{x}_{[1+w_{i-1},w_i]}. \tag{B.2}$$

Then, the following two propositions hold:

(i) Let s be an integer with $1 < s \leq i$. If the inequality

$$\boldsymbol{D}_{1+w_{t-1},w_t}\boldsymbol{x}_{[1+w_{t-1},w_t]} \geq \boldsymbol{0}_{w_t - w_{t-1} - 1} \tag{B.3}$$

holds for any integer t with $s \leq t \leq i$, then, the following inequality also holds:

$$\boldsymbol{D}_{1+w_{s-2},w_{s-1}} \boldsymbol{x}_{[1+w_{s-2},w_{s-1}]} \geq \mathbf{0}_{w_{s-1}-w_{s-2}-1}, \tag{B.4}$$

where we define $\mathbf{0}_0 = 0$.

(ii) For any integer t with $1 \leq t \leq i$, it holds that

$$\boldsymbol{D}_{1+w_{t-1},w_t} \boldsymbol{x}_{[1+w_{t-1},w_t]} \geq \mathbf{0}_{w_t-w_{t-1}-1}. \tag{B.5}$$

Proof. First, we prove (i). We would like to recall that, from (3.5) $\eta_l(\boldsymbol{x})[w_{s-1}]$ is given by

$$\eta_l(\boldsymbol{x})[w_{s-1}] = \min_{v;v \geq w_{s-1}} \max_{u;u \leq w_{s-1}} \bar{x}_{[u,v]}. \tag{B.6}$$

Here, assume that

$$^{\exists}v^* > w_{s-1} \quad \text{s.t.} \quad v^* = \operatorname*{argmin}_{v;v \geq w_{s-1}} \left(\max_{u;u \leq w_{s-1}} \bar{x}_{[u,v]} \right). \tag{B.7}$$

Note that the assumption (B.7) is equal to

$$\min_{v;v \geq w_{s-1}} \max_{u;u \leq w_{s-1}} \bar{x}_{[u,v]} = \max_{u;u \leq w_{s-1}} \bar{x}_{[u,v^*]}. \tag{B.8}$$

Then, from (B.6) and (B.8) we have

$$\eta_l(\boldsymbol{x})[w_{s-1}] = \max_{u;u \leq w_{s-1}} \bar{x}_{[u,v^*]}.$$

Furthermore, noting that

$$\max_{u;u \leq w_{s-1}} \bar{x}_{[u,v^*]} \geq \bar{x}_{[1+w_{s-2},v^*]}$$

we also get

$$\eta_l(\boldsymbol{x})[w_{s-1}] \geq \bar{x}_{[1+w_{s-2},v^*]}. \tag{B.9}$$

Incidentally, since v^* satisfies the inequality $v^* > w_{s-1}$, there exists a number t such that $s \leq t \leq i$ and $1 + w_{t-1} \leq v^* \leq w_t$. Based on this, we consider the following two cases:

Case 1 : $\bar{x}_{[1+w_{s-2},w_{t-1}]} < \bar{x}_{[1+w_{t-1},v^*]}$,
Case 2 : $\bar{x}_{[1+w_{s-2},w_{t-1}]} \geq \bar{x}_{[1+w_{t-1},v^*]}$.

It is clear that Case 1 is the negation of Case 2. Next, we show that both Case 1 and Case 2 are false. In fact, if Case 1 is true, i.e., the inequality $\bar{x}_{[1+w_{s-2},w_{t-1}]} < \bar{x}_{[1+w_{t-1},v^*]}$ is true, from (A.2) we obtain $\bar{x}_{[1+w_{s-2},w_{t-1}]} < \bar{x}_{[1+w_{s-2},v^*]}$. Thus, using this inequality and (B.9) we get

$$\eta_l(\boldsymbol{x})[w_{s-1}] > \bar{x}_{[1+w_{s-2},w_{t-1}]}. \tag{B.10}$$

Recall that we assume the inequality (B.2). Hence, from (A.4) it holds that

$$\bar{x}_{[1+w_{s-2},w_{s-1}]} < \bar{x}_{[1+w_{s-2},w_{t-1}]}. \tag{B.11}$$

Therefore, combining (B.10) and (B.11), we obtain

$$\eta_l(\boldsymbol{x})[w_{s-1}] > \bar{x}_{[1+w_{s-2},w_{s-1}]}. \tag{B.12}$$

However, from the assumption (B.1), it holds that $\eta_l(\boldsymbol{x})[w_{s-1}] = \bar{x}_{[1+w_{s-2},w_{s-1}]}$. This result and (B.12) contradict. Hence, Case 1 is false. Next, we consider Case 2. Suppose that $\bar{x}_{[1+w_{s-2},w_{t-1}]} \geq \bar{x}_{[1+w_{t-1},v^*]}$ is true. Then, from (A.1) we have $\bar{x}_{[1+w_{s-2},v^*]} \geq \bar{x}_{[1+w_{t-1},v^*]}$. Combining this inequality and (B.9), we get

$$\eta_l(\boldsymbol{x})[w_{s-1}] \geq \bar{x}_{[1+w_{t-1},v^*]}. \tag{B.13}$$

Here, when $v^* = w_t$, from (B.13) it holds that

$$\eta_l(\boldsymbol{x})[w_{s-1}] \geq \bar{x}_{[1+w_{t-1},w_t]}.$$

On the other hand, when $1 + w_{t-1} \leq v^* < w_t$, from the assumption (B.3), it holds that $\boldsymbol{D}_{1+w_{t-1},w_t}\boldsymbol{x}_{[1+w_{t-1},w_t]} \geq \mathbf{0}_{w_t-w_{t-1}-1}$. Using this inequality and (A.5) we obtain

$$\bar{x}_{[1+w_{t-1},v^*]} \geq \bar{x}_{[1+v^*,w_t]}.$$

Again, from (A.1) it holds that $\bar{x}_{[1+w_{t-1},v^*]} \geq \bar{x}_{[1+w_{t-1},w_t]}$. Substituting this inequality into (B.13) yields $\eta_l(\boldsymbol{x})[w_{s-1}] \geq \bar{x}_{[1+w_{t-1},w_t]}$. Thus, in both cases it holds that

$$\eta_l(\boldsymbol{x})[w_{s-1}] \geq \bar{x}_{[1+w_{t-1},w_t]}. \tag{B.14}$$

Here, recall that from the assumption (B.1), the equality $\eta_l(\boldsymbol{x})[w_{s-1}] = \bar{x}_{[1+w_{s-2},w_{s-1}]}$ holds. Therefore, combining this equality and (B.14) it holds that

$$\bar{x}_{[1+w_{s-2},w_{s-1}]} \geq \bar{x}_{[1+w_{t-1},w_t]}. \tag{B.15}$$

Note that from the definitions of s and t, the inequality $s-1 \leq t-1$ holds. Thus, (B.15) and (B.2) contradict. Therefore, Case 2 is false. Hence, both Case 1 and Case 2 are false. This implies that the assumption (B.7) is not true. Thus, we obtain

$$\underset{v;v\geq w_{s-1}}{\operatorname{argmin}} \left(\max_{u;u\leq w_{s-1}} \bar{x}_{[u,v]} \right) = w_{s-1},$$

in other words, it holds that

$$\eta_l(\boldsymbol{x})[w_{s-1}] = \min_{v;v\geq w_{s-1}} \max_{u;u\leq w_{s-1}} \bar{x}_{[u,v]} = \max_{u;u\leq w_{s-1}} \bar{x}_{[u,w_{s-1}]}. \tag{B.16}$$

Therefore, from (B.16) it holds that $\eta_l(\boldsymbol{x})[w_{s-1}] \geq \bar{x}_{[r,w_{s-1}]}$ for any integer r with $1 + w_{s-2} < r \leq w_{s-1}$. Again, using $\eta_l(\boldsymbol{x})[w_{s-1}] = \bar{x}_{[1+w_{s-2},w_{s-1}]}$ we get $\bar{x}_{[1+w_{s-2},w_{s-1}]} \geq \bar{x}_{[r,w_{s-1}]}$. Moreover, from (A.1) we have $\bar{x}_{[1+w_{s-2},r-1]} \geq \bar{x}_{[r,w_{s-1}]}$. Here, we replace $r-1$ with b. Then, b is the integer satisfying $1 + w_{s-2} \leq b < w_{s-1}$ and it holds that

$\bar{x}_{[1+w_{s-2},b]} \geq \bar{x}_{[b+1,w_{s-1}]}$. Thus, from (A.5), this implies $\boldsymbol{D}_{1+w_{s-2},w_{s-1}}\boldsymbol{x}_{[1+w_{s-2},w_{s-1}]} \geq \mathbf{0}_{w_{s-1}-w_{s-2}-1}$. Consequently, (B.4) is proved, This implies that (i) holds.

Next, we prove (ii). Since we have already proved the first proposition (i), in order to prove (ii), it is sufficient to prove that

$$\boldsymbol{D}_{1+w_{i-1},w_i}\boldsymbol{x}_{[1+w_{i-1},w_i]} \geq \mathbf{0}_{w_i-w_{i-1}-1}.$$

Here, we consider $\eta_l(\boldsymbol{x})[w_i]$. Recall that from (3.5) $\eta_l(\boldsymbol{x})[w_i]$ is given by

$$\eta_l(\boldsymbol{x})[w_i] = \min_{v;v\geq w_i}\max_{u;u\leq w_i}\bar{x}_{[u,v]}.$$

Noting that $w_i = l$, we get

$$\eta_l(\boldsymbol{x})[w_i] = \min_{v;v\geq w_i}\max_{u;u\leq w_i}\bar{x}_{[u,v]} = \max_{u;u\leq w_i}\bar{x}_{[u,w_i]}. \tag{B.17}$$

Also note that, from (B.1) the equality $\eta_l(\boldsymbol{x})[w_i] = \bar{x}_{[1+w_{i-1},w_i]}$ holds. Therefore, using this equality and (B.17) we obtain $\bar{x}_{[1+w_{i-1},w_i]} \geq \bar{x}_{[r,w_i]}$ for any integer r with $1 + w_{i-1} < r \leq w_i$. Again, by using the same argument as in the proof of (i), we have $\boldsymbol{D}_{1+w_{i-1},w_i}\boldsymbol{x}_{[1+w_{i-1},w_i]} \geq \mathbf{0}_{w_i-w_{i-1}-1}$. Thus, combining this result and (i), (ii) is proved.

□

Appendix C: Lemma C and its proof

Lemma C. Let l be an integer with $l \geq 2$, and let $n_1, \ldots, n_l \in \mathbb{R}_{>0}$, $\boldsymbol{n} = (n_1, \ldots, n_l)'$, $\xi_1, \ldots, \xi_l \in \mathbb{R}$ and $\tau^2 > 0$. Let $x_1, \ldots, x_l$ be independent random variables, and for any integer s with $1 \leq s \leq l$, let $x_s \sim N(\xi_s, \tau^2/n_s)$. Put $\boldsymbol{x} = (x_1, \ldots, x_l)'$. Then, the following four propositions hold:

(i)

$$\mathbb{R}^l = \bigcup_{i=1}^{l}\bigcup_{\boldsymbol{w};\boldsymbol{w}\in\mathcal{W}_i^l}\boldsymbol{\eta}_l^{-1}(\mathcal{A}_i^l(\boldsymbol{w})),$$
$$\boldsymbol{\eta}_l^{-1}(\mathcal{A}_i^l(\boldsymbol{w})) \cap \boldsymbol{\eta}_l^{-1}(\mathcal{A}_{i^*}^l(\boldsymbol{w}^*)) = \emptyset, \quad ((i,\boldsymbol{w}) \neq (i^*,\boldsymbol{w}^*)).$$

(ii) For the set $\mathcal{A}_1^l(\boldsymbol{w}) = \mathcal{A}_1^l$, it holds that

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

Moreover, for any integer i with $2 \leq i \leq l$ and for any element $\boldsymbol{w} = (w_1, \ldots, w_i)' \in \mathcal{W}_i^l$, it holds that

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

where $w_0 = 0$ and, $\rho_{t,\boldsymbol{w}} = w_{t+1} - w_t - 1$.

(iii) For any integer i with $1 \le i \le l$ and for any element $\boldsymbol{w} = (w_1, \ldots, w_i)' \in \mathcal{W}_i^l$, it holds that

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

where $w_0 = 0$.

(iv) For any integer i with $1 \le i \le l$, it holds that

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

Proof. First, we prove (i). From the definition of the function $\boldsymbol{\eta}_l(\cdot)$, we get $\boldsymbol{\eta}_l(\mathbb{R}^l) \subset \mathcal{A}^l$. Hence, from (3.2) and (3.3) it is clear that (i) holds. Second, we prove (iv). Here, note that from (i) it holds that $\boldsymbol{\eta}_l^{-1}(\mathcal{A}_i^l(\boldsymbol{w})) \cap \boldsymbol{\eta}_l^{-1}(\mathcal{A}_{i^*}^l(\boldsymbol{w}^*)) = \emptyset$. Thus, the events $\boldsymbol{x} \in \boldsymbol{\eta}_l^{-1}(\mathcal{A}_i^l(\boldsymbol{w}))$ and $\boldsymbol{x} \in \boldsymbol{\eta}_l^{-1}(\mathcal{A}_{i^*}^l(\boldsymbol{w}^*))$ are disjoint. Therefore, from the definition of the probability we obtain

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

Furthermore, from the inverse image, we also get

$$\boldsymbol{x} \in \bigcup_{\boldsymbol{w};\boldsymbol{w}\in\mathcal{W}_i^l} \boldsymbol{\eta}_l^{-1}(\mathcal{A}_i^l(\boldsymbol{w})) \Leftrightarrow \boldsymbol{\eta}_l(\boldsymbol{x}) \in \bigcup_{\boldsymbol{w};\boldsymbol{w}\in\mathcal{W}_i^l} \mathcal{A}_i^l(\boldsymbol{w}).$$

Hence, (iv) is proved

Next, we prove (ii). First, we prove the right-arrow $\Rightarrow$ in (C.1). $\boldsymbol{x} \in \boldsymbol{\eta}_l^{-1}(\mathcal{A}_1^l(\boldsymbol{w}))$, i.e., $\boldsymbol{\eta}_l(\boldsymbol{x}) \in \mathcal{A}_1^l(\boldsymbol{w})$. Then, from the definition of $\mathcal{A}_1^l(\boldsymbol{w})$, we get

$$\eta_l(\boldsymbol{x})[1] = \eta_l(\boldsymbol{x})[2] = \cdots = \eta_l(\boldsymbol{x})[l] \equiv \hat{\alpha} \quad \text{(say)}.$$

This implies that $\boldsymbol{\eta}_l(\boldsymbol{x}) = \hat{\alpha}\mathbf{1}_l$. In addition, from the definition of the function $\boldsymbol{\eta}_l$ it holds that

$$\min_{\boldsymbol{y}\in\mathcal{A}^l} \|\boldsymbol{x} - \boldsymbol{y}\|_{\boldsymbol{n}}^2 = \|\boldsymbol{x} - \boldsymbol{\eta}_l(\boldsymbol{x})\|_{\boldsymbol{n}}^2 = \|\boldsymbol{x} - \hat{\alpha}\mathbf{1}_l\|_{\boldsymbol{n}}^2. \tag{C.3}$$

Moreover, noting that $\mathcal{A}_1^l(\boldsymbol{w}) \subset \mathcal{A}^l$ we have

$$\min_{\boldsymbol{y}\in\mathcal{A}^l} \|\boldsymbol{x} - \boldsymbol{y}\|_{\boldsymbol{n}}^2 \le \min_{\boldsymbol{y}^*\in\mathcal{A}_1^l(\boldsymbol{w})} \|\boldsymbol{x} - \boldsymbol{y}^*\|_{\boldsymbol{n}}^2 = \min_{\alpha\in\mathbb{R}^1} \|\boldsymbol{x} - \alpha\mathbf{1}_l\|_{\boldsymbol{n}}^2.$$

Here, note that the norm $\|\boldsymbol{x} - \alpha\mathbf{1}_l\|_{\boldsymbol{n}}^2$ is a convex function with respect to (w.r.t.) α on $\mathbb{R}^1$. Thus, there exists a unique point $\alpha_{\min}$ which maximizes $\|\boldsymbol{x} - \alpha\mathbf{1}_l\|_{\boldsymbol{n}}^2$ w.r.t. α. Therefore, we obtain

$$\min_{\boldsymbol{y}\in\mathcal{A}^l} \|\boldsymbol{x} - \boldsymbol{y}\|_{\boldsymbol{n}}^2 \le \|\boldsymbol{x} - \alpha_{\min}\mathbf{1}_l\|_{\boldsymbol{n}}^2. \tag{C.4}$$

Hence, combining (C.3) and (C.4), the inequality $\|\boldsymbol{x} - \hat{\alpha}\mathbf{1}_l\|_{\boldsymbol{n}}^2 \leq \|\boldsymbol{x} - \alpha_{\min}\mathbf{1}_l\|_{\boldsymbol{n}}^2$ holds. Therefore, from uniqueness of $\alpha_{\min}$ we obtain $\hat{\alpha} = \alpha_{\min}$. On the other hand, $\alpha_{\min}$ can be obtained by differentiating the function $\|\boldsymbol{x} - \alpha\mathbf{1}_l\|_{\boldsymbol{n}}^2$ w.r.t. α as $\alpha_{\min} = \bar{x}_{[1,l]}$ because the function $\|\boldsymbol{x} - \alpha\mathbf{1}_l\|_{\boldsymbol{n}}^2$ is the convex function. Thus, it holds that

$$\eta_l(\boldsymbol{x})[1] = \eta_l(\boldsymbol{x})[2] = \cdots = \eta_l(\boldsymbol{x})[l] = \bar{x}_{[1,l]}. \tag{C.5}$$

Here, recall that $\eta_l(\boldsymbol{x})[s]$ is given by (3.5). Hence, $\eta_l(\boldsymbol{x})[1]$ can be written as

$$\eta_l(\boldsymbol{x})[1] = \min_{v \geq 1} \bar{x}_{[1,v]}. \tag{C.6}$$

Moreover, from (C.5) we get $\eta_l(\boldsymbol{x})[1] = \bar{x}_{[1,l]}$. Therefore, combining this equality and (C.6), it holds that $\bar{x}_{[1,v]} \geq \bar{x}_{[1,l]}$ for any integer v with $1 \leq v < l$. Thus, from (A.1) we obtain $\bar{x}_{[1,v]} \geq \bar{x}_{[v+1,l]}$. Hence, from (A.5) this implies that $\boldsymbol{D}_{1,l}\boldsymbol{x}_{[1,l]} \geq \mathbf{0}_{l-1}$. Consequently, the right-arrow $\Rightarrow$ in (C.1) is proved.

Next, we prove the left-arrow $\Leftarrow$ in (C.1). Let $\boldsymbol{D}_{1,l}\boldsymbol{x}_{[1,l]} \geq \mathbf{0}_{l-1}$. Then, from (A.5) it holds that $\bar{x}_{[1,v]} \geq \bar{x}_{[v+1,l]}$ for any integer v with $1 \leq v < l$. Again, from (A.1) we have $\bar{x}_{[1,v]} \geq \bar{x}_{[1,l]}$. Hence, combining this result and (C.6) we get $\eta_l(\boldsymbol{x})[1] = \bar{x}_{[1,l]}$. On the other hand, from (3.5), $\eta_l(\boldsymbol{x})[l]$ can be expressed as

$$\eta_l(\boldsymbol{x})[l] = \max_{u \leq l} \bar{x}_{[u,l]}. \tag{C.7}$$

Here, since the inequality $\bar{x}_{[1,v]} \geq \bar{x}_{[v+1,l]}$ holds, from (A.1) we obtain $\bar{x}_{[1,l]} \geq \bar{x}_{[v+1,l]}$. This result and (C.7) yield $\eta_l(\boldsymbol{x})[l] = \bar{x}_{[1,l]}$. Thus, it holds that $\eta_l(\boldsymbol{x})[1] = \eta_l(\boldsymbol{x})[l]$. In addition, from the definition of $\boldsymbol{\eta}_l$ we have $\eta_l(\boldsymbol{x})[1] \leq \cdots \leq \eta_l(\boldsymbol{x})[l]$. Therefore, combining this inequality and the equality $\eta_l(\boldsymbol{x})[1] = \eta_l(\boldsymbol{x})[l]$, we get $\eta_l(\boldsymbol{x})[1] = \cdots = \eta_l(\boldsymbol{x})[l]$. This implies that $\boldsymbol{\eta}_l(\boldsymbol{x}) \in \mathcal{A}_1^l(\boldsymbol{w})$, i.e., $\boldsymbol{x} \in \boldsymbol{\eta}_l^{-1}(\mathcal{A}_1^l(\boldsymbol{w}))$. Hence, the left-arrow $\Leftarrow$ in (C.1) is proved. Therefore, (C.1) is proved.

Next, we prove (C.2). First, we show the right-arrow $\Rightarrow$ in (C.2). Let i be an integer with $2 \leq i \leq l$, and let $\boldsymbol{w} = (w_1, \ldots, w_i)'$ be an element with $\boldsymbol{w} \in \mathcal{W}_i^l$. Here, we put $w_0 = 0$. Furthermore, assume that $\boldsymbol{x} \in \boldsymbol{\eta}_l^{-1}(\mathcal{A}_i^l(\boldsymbol{w}))$. Note that $w_i = l$. Then, since $\boldsymbol{\eta}_l(\boldsymbol{x}) \in \mathcal{A}_i^l(\boldsymbol{w})$ the following equalities hold:

$$\begin{aligned}
\eta_l(\boldsymbol{x})[1] = \cdots = \eta_l(\boldsymbol{x})[w_1] &\equiv \hat{\delta}_1, \\
\eta_l(\boldsymbol{x})[w_1+1] = \cdots = \eta_l(\boldsymbol{x})[w_2] &\equiv \hat{\delta}_2, \\
&\vdots \\
\eta_l(\boldsymbol{x})[w_{i-1}+1] = \cdots = \eta_l(\boldsymbol{x})[w_i] &\equiv \hat{\delta}_i, \quad \text{(say)}.
\end{aligned}$$

Moreover, the inequality $\hat{\delta}_1 < \cdots < \hat{\delta}_i$ also holds. We put $\hat{\boldsymbol{\delta}} = (\hat{\delta}_1, \ldots, \hat{\delta}_i)'$. Then, $\boldsymbol{\eta}_l(\boldsymbol{x})$ can be written by using $\hat{\boldsymbol{\delta}}$ as $\boldsymbol{\eta}_l(\boldsymbol{x}) = (\hat{\delta}_1\mathbf{1}'_{w_1-w_0}, \ldots, \hat{\delta}_i\mathbf{1}'_{w_i-w_{i-1}})'$. From the definitions of $\boldsymbol{\eta}_l$ and $\|\ \|_{\boldsymbol{n}}$, using $\boldsymbol{\eta}_l(\boldsymbol{x}) = (\hat{\delta}_1 \mathbf{1}'_{w_1-w_0}, \dots, \hat{\delta}_i \mathbf{1}'_{w_i-w_{i-1}})'$ we get

$$\min_{\boldsymbol{y}\in\mathcal{A}^l} \|\boldsymbol{x}-\boldsymbol{y}\|^2_{\boldsymbol{n}} = \|\boldsymbol{x}-\boldsymbol{\eta}_l(\boldsymbol{x})\|^2_{\boldsymbol{n}} = \sum_{s=0}^{i-1} \sum_{u=1+w_s}^{w_{s+1}} n_u (x_u - \hat{\delta}_{s+1})^2. \tag{C.8}$$

Define for each $\boldsymbol{\delta} = (\delta_1, \dots, \delta_i)' \in \mathbb{R}^i$ a function

$$f(\boldsymbol{\delta}) = \sum_{s=0}^{i-1} \sum_{u=1+w_s}^{w_{s+1}} n_u (x_u - \delta_{s+1})^2.$$

Therefore, the right hand side in (C.8) can be written as $f(\hat{\boldsymbol{\delta}})$. Incidentally, since $\mathcal{A}^l_i(\boldsymbol{w}) \subset \mathcal{A}^l$ the following inequality holds:

$$\min_{\boldsymbol{y}\in\mathcal{A}^l} \|\boldsymbol{x}-\boldsymbol{y}\|^2_{\boldsymbol{n}} \le \min_{\boldsymbol{y}^*\in\mathcal{A}^l_i(\boldsymbol{w})} \|\boldsymbol{x}-\boldsymbol{y}^*\|^2_{\boldsymbol{n}} = \min_{\boldsymbol{\delta}\in\mathcal{A}^i_i} f(\boldsymbol{\delta}). \tag{C.9}$$

Here, there exists a positive number ε such that the ε-neighborhood of $\hat{\boldsymbol{\delta}}$, $\mathrm{U}(\hat{\boldsymbol{\delta}};\varepsilon)$ satisfies $\mathrm{U}(\hat{\boldsymbol{\delta}};\varepsilon) \subset \mathcal{A}^i_i$ because $\hat{\boldsymbol{\delta}} \in \mathcal{A}^i_i$ and $\mathcal{A}^i_i$ is an open set. By combining this result and (C.9) we have

$$\min_{\boldsymbol{y}\in\mathcal{A}^l} \|\boldsymbol{x}-\boldsymbol{y}\|^2_{\boldsymbol{n}} \le \min_{\boldsymbol{\delta}\in\mathcal{A}^i} f(\boldsymbol{\delta}) \le \min_{\boldsymbol{\delta}^*\in\mathrm{U}(\hat{\boldsymbol{\delta}};\varepsilon)} f(\boldsymbol{\delta}^*). \tag{C.10}$$

Hence, from (C.8) and (C.10) it holds that $f(\hat{\boldsymbol{\delta}}) \le f(\boldsymbol{\delta}^*)$ for any $\boldsymbol{\delta}^* \in \mathrm{U}(\hat{\boldsymbol{\delta}};\varepsilon)$. Thus, the point $\hat{\boldsymbol{\delta}}$ minimizes the function $f(\boldsymbol{\delta})$. On the other hand, since $f(\boldsymbol{\delta})$ is a convex function w.r.t. $\boldsymbol{\delta}$ on $\mathbb{R}^i$, there exists a unique point $\boldsymbol{\delta}_{\min} = (\delta_{1,\min}, \dots, \delta_{i,\min})'$ which minimizes $f(\boldsymbol{\delta})$. Therefore, noting that $f(\boldsymbol{\delta})$ is convex, we get $\hat{\boldsymbol{\delta}} = \boldsymbol{\delta}_{\min}$. Furthermore, the point $\boldsymbol{\delta}_{\min}$ can be obtained by differentiating the function $f(\boldsymbol{\delta})$ w.r.t. $\boldsymbol{\delta}$ as

$$\eta_l(\boldsymbol{x})[1+w_t] = \cdots = \eta_l(\boldsymbol{x})[w_{t+1}] = \hat{\delta}_{t+1} = \delta_{t+1,\min} = \bar{x}_{[1+w_t,w_{t+1}]}, \tag{C.11}$$

for any integer t with $0 \le t \le i-1$. Here, since $\hat{\delta}_1 < \cdots < \hat{\delta}_i$, for any integer s with $0 \le s \le i-2$ it holds that

$$\bar{x}_{[1+w_s,w_{s+1}]} < \bar{x}_{[1+w_{s+1},w_{s+2}]}. \tag{C.12}$$

Therefore, (C.11) and (C.12) imply that the all assumptions in Lemma B are satisfied. Thus, from (B.5), for any integer t with $0 \le t \le i-1$ it holds that $\boldsymbol{D}_{1+w_t,w_{t+1}} \boldsymbol{x}_{[1+w_t,w_{t+1}]} \ge \mathbf{0}_{w_{t+1}-w_t-1}$. Hence, by considering this result and (C.12), the right-arrow $\Rightarrow$ in (C.2) is proved.

Next, we prove the left-arrow $\Leftarrow$ in (C.2). Let i be an integer with $2 \le i \le l$, and let $\boldsymbol{w} = (w_1, \dots, w_i)'$ be an element with $\boldsymbol{w} \in \mathcal{W}^l_i$. We put $w_0 = 0$. Assume that

$$\boldsymbol{D}_{1+w_t,w_{t+1}} \boldsymbol{x}_{[1+w_t,w_{t+1}]} \ge \mathbf{0}_{w_{t+1}-w_t-1} \quad (0 \le t \le i-1), \tag{C.13}$$

and

$$\bar{x}_{[1+w_s,w_{s+1}]} < \bar{x}_{[1+w_{s+1},w_{s+2}]} \quad (0 \le s \le i-2). \tag{C.14}$$

Then, from the definitions of ノ $\| \, \|_{\boldsymbol{n}}$ and $\boldsymbol{\eta}_l(\boldsymbol{x})$, it holds that

$$\begin{aligned} \min_{\boldsymbol{\delta} \in \mathcal{A}^l} \|\boldsymbol{x} - \boldsymbol{\delta}|_{\boldsymbol{n}}^2 &= \|\boldsymbol{x} - \boldsymbol{\eta}_l(\boldsymbol{x})|_{\boldsymbol{n}}^2 \\ &= \sum_{t=0}^{i-1} \|\boldsymbol{x}_{[1+w_t,w_{t+1}]} - \boldsymbol{\eta}_l(\boldsymbol{x})_{[1+w_t,w_{t+1}]}\|_{\boldsymbol{n}_{[1+w_t,w_{t+1}]}}^2. \end{aligned} \tag{C.15}$$

In addition, since $\boldsymbol{\eta}_l(\boldsymbol{x})_{[1+w_t,w_{t+1}]} \in \mathcal{A}^{w_{t+1}-w_t}$, for any integer t with $0 \le t \le i-1$, the inequality holds:

$$\begin{aligned} &\|\boldsymbol{x}_{[1+w_t,w_{t+1}]} - \boldsymbol{\eta}_l(\boldsymbol{x})_{[1+w_t,w_{t+1}]}\|_{\boldsymbol{n}_{[1+w_t,w_{t+1}]}}^2 \\ \ge &\min_{\boldsymbol{\delta}_{[1+w_t,w_{t+1}]} \in \mathcal{A}^{w_{t+1}-w_t}} \|\boldsymbol{x}_{[1+w_t,w_{t+1}]} - \boldsymbol{\delta}_{[1+w_t,w_{t+1}]}\|_{\boldsymbol{n}_{[1+w_t,w_{t+1}]}}^2. \end{aligned} \tag{C.16}$$

Next, we evaluate the right hand side in (C.16). Here, we replace $w_{t+1} - w_t$, $\boldsymbol{x}_{[1+w_t,w_{t+1}]}$ and $\boldsymbol{n}_{[1+w_t,w_{t+1}]}$ with g, $\boldsymbol{y}_{[1,g]} = \boldsymbol{y} = (y_1, \ldots, y_g)'$ and $\boldsymbol{N}_{[1,g]} = \boldsymbol{N} = (N_1, \ldots, N_g)'$, respectively. Then, the right hand side in (C.16) can be rewritten as

$$\begin{aligned} &\min_{\boldsymbol{\delta}_{[1+w_t,w_{t+1}]} \in \mathcal{A}^{w_{t+1}-w_t}} \|\boldsymbol{x}_{[1+w_t,w_{t+1}]} - \boldsymbol{\delta}_{[1+w_t,w_{t+1}]}\|_{\boldsymbol{n}_{[1+w_t,w_{t+1}]}}^2 \\ = &\min_{\boldsymbol{\delta}_{[1+w_t,w_{t+1}]} \in \mathcal{A}^{g}} \|\boldsymbol{y} - \boldsymbol{\delta}_{[1+w_t,w_{t+1}]}\|_{\boldsymbol{N}}^2 = \|\boldsymbol{y} - \boldsymbol{\eta}_g^{(\boldsymbol{N})}(\boldsymbol{y})\|_{\boldsymbol{N}}^2. \end{aligned} \tag{C.17}$$

in the case of $g = 1$, i.e., $w_{t+1} = w_t + 1$, since $\boldsymbol{\eta}_1^{(\boldsymbol{N})}(\boldsymbol{y}) = \boldsymbol{y}$, it is clear that $\boldsymbol{\eta}_g^{(\boldsymbol{N})}(\boldsymbol{y}) = \boldsymbol{y} = \boldsymbol{x}_{[1+w_t,w_{t+1}]} = x_{t+1} = \bar{x}_{[1+w_t,w_{t+1}]}$. On the other hand, in the case of $g \ge 2$, from (C.13) and the definition of the matrix $\boldsymbol{D}_{i,j} = \boldsymbol{D}_{i,j}^{(\boldsymbol{n})}$, we get

$$\begin{aligned} \boldsymbol{D}_{1+w_t,w_{t+1}} \boldsymbol{x}_{[1+w_t,w_{t+1}]} \ge \boldsymbol{0}_{w_{t+1}-w_t-1} &\Leftrightarrow \boldsymbol{D}_{1+w_t,w_{t+1}}^{(\boldsymbol{n})} \boldsymbol{y}_{[1,g]} \ge \boldsymbol{0}_{g-1} \\ &\Leftrightarrow \boldsymbol{D}_{1,g}^{(\boldsymbol{N})} \boldsymbol{y}_{[1,g]} \ge \boldsymbol{0}_{g-1}. \end{aligned} \tag{C.18}$$

Moreover, we obtain

$$\boldsymbol{D}_{1,g}^{(\boldsymbol{N})} \boldsymbol{y}_{[1,g]} \ge \boldsymbol{0}_{g-1} \Rightarrow \boldsymbol{y} \in (\boldsymbol{\eta}_g^{(\boldsymbol{N})})^{-1}(\mathcal{A}_1^g), \tag{C.19}$$

because we have already proved (C.1). Furthermore, from (C.5) we get

$$\boldsymbol{y} \in (\boldsymbol{\eta}_g^{(\boldsymbol{N})})^{-1}(\mathcal{A}_1^g) \Rightarrow \eta_g^{(\boldsymbol{N})}(\boldsymbol{y})[1] = \cdots = \eta_g^{(\boldsymbol{N})}(\boldsymbol{y})[g] = \bar{y}_{[1,g]}^{(\boldsymbol{N})} = \bar{x}_{[1+w_t,w_{t+1}]}. \tag{C.20}$$

Therefore, combining (C.18), (C.19) and (C.20) we obtain

$$\begin{aligned} &\boldsymbol{D}_{1+w_t,w_{t+1}} \boldsymbol{x}_{[1+w_t,w_{t+1}]} \ge \boldsymbol{0}_{w_{t+1}-w_t-1} \\ \Rightarrow\ &\eta_g^{(\boldsymbol{N})}(\boldsymbol{y})[1] = \cdots = \eta_g^{(\boldsymbol{N})}(\boldsymbol{y})[g] = \bar{x}_{[1+w_t,w_{t+1}]}. \end{aligned}$$

Thus, by using this result, (C.15), (C.16) and (C.17) imply that

$$\|\boldsymbol{x}-\boldsymbol{\eta}_l(\boldsymbol{x})\|_{\boldsymbol{n}}^2 \geq \sum_{t=0}^{i-1} \|\boldsymbol{x}_{[1+w_t,w_{t+1}]} - \bar{x}_{[1+w_t,w_{t+1}]}\boldsymbol{1}_{w_{t+1}-w_t}\|_{\boldsymbol{n}_{[1+w_t,w_{t+1}]}}^2. \tag{C.21}$$

Here, we put $\boldsymbol{h} = (\bar{x}_{[1,w_1]}\boldsymbol{1}'_{w_1}, \bar{x}_{[1+w_1,w_2]}\boldsymbol{1}'_{w_2-w_1}, \ldots, \bar{x}_{[1+w_{i-1},w_i]}\boldsymbol{1}'_{w_i-w_{i-1}})'$. From (C.14), since $\boldsymbol{h} \in \mathcal{A}^l$ we get

$$\begin{aligned}\|\boldsymbol{x}-\boldsymbol{\eta}_l(\boldsymbol{x})\|_{\boldsymbol{n}}^2 &= \min_{\boldsymbol{\delta}\in\mathcal{A}^l} \|\boldsymbol{x}-\boldsymbol{\delta}\|_{\boldsymbol{n}}^2 \\ &\leq \|\boldsymbol{x}-\boldsymbol{h}\|_{\boldsymbol{n}}^2 \\ &= \sum_{t=0}^{i-1} \|\boldsymbol{x}_{[1+w_t,w_{t+1}]} - \bar{x}_{[1+w_t,w_{t+1}]}\boldsymbol{1}_{w_{t+1}-w_t}\|_{\boldsymbol{n}_{[1+w_t,w_{t+1}]}}^2. \end{aligned} \tag{C.22}$$

Hence, (C.21) and (C.22) imply that $\boldsymbol{\eta}_l(\boldsymbol{x}) = \boldsymbol{h}$. In addition, noting that $\boldsymbol{h} \in \mathcal{A}_i^l(\boldsymbol{w})$, it also holds that $\boldsymbol{x} \in \boldsymbol{\eta}_l^{-1}(\mathcal{A}_i^l(\boldsymbol{w}))$. Thus, the left-arrow $\Leftarrow$ in (C.2) is proved. Consequently, (ii) is proved.

Finally, in the proof of (ii), we have already proved (C.5) and (C.11). Thus, (iii) is proved. Therefore, Lemma C is proved. □

Appendix D: Lemma D and its proof

Lemma D. Let $v_1, \ldots, v_l$ be independent random variables, and let $v_s \sim N(\xi_s, \tau^2/n_s)$ where $1 \leq s \leq l$. Let $\tau^2 > 0$, $\xi_1, \ldots, \xi_l \in \mathbb{R}$, $n_1, \ldots, n_l \in \mathbb{R}_{>0}$, $\boldsymbol{n} = (n_1, \ldots, n_l)'$ and $\boldsymbol{v} = (v_1, \ldots, v_l)'$. Then, for any i and j with $1 \leq i \leq j \leq l$, it holds that

$$\boldsymbol{D}_{i,j}\boldsymbol{v}_{[i,j]} \perp\!\!\!\perp \bar{v}_{[i,j]}, \tag{D.1}$$

and

$$\bar{v}_{[i,j]} \perp\!\!\!\perp \sum_{s=i}^{j} n_s(v_s-\xi_s)(v_s-\bar{v}_{[i,j]}). \tag{D.2}$$

Proof. First, we prove (D.1). when $i = j$, since $\boldsymbol{D}_{i,j} = 0$ it is clear that $\boldsymbol{D}_{i,j}\boldsymbol{v}_{[i,j]} \perp\!\!\!\perp \bar{v}_{[i,j]}$. Hence, we prove the case of $i < j$. Noting that $\bar{v}_{[i,j]}$ can be written as

$$\bar{v}_{[i,j]} = \frac{\boldsymbol{n}'_{[i,j]}}{\tilde{n}_{[i,j]}}\boldsymbol{v}_{[i,j]},$$

we get

$$\begin{pmatrix} \boldsymbol{D}_{i,j}\boldsymbol{v}_{[i,j]} \\ \bar{v}_{[i,j]} \end{pmatrix} = \begin{pmatrix} \boldsymbol{D}_{i,j} \\ \frac{\boldsymbol{n}'_{[i,j]}}{\tilde{n}_{[i,j]}} \end{pmatrix} \boldsymbol{v}_{[i,j]} \quad \sim N_{j-i+1}(*, \star).$$

Thus, it is sufficient to prove $\text{Cov}[\boldsymbol{D}_{i,j}\boldsymbol{v}_{[i,j]}, \bar{v}_{[i,j]}] = \boldsymbol{0}_{j-i}$. Here, for any s with $(1 \leq s \leq j-i)$, the sth row of $\boldsymbol{D}_{i,j}$ is given by (3.4), it holds that

$$\left(\frac{1}{\tilde{n}_{[i,i+s-1]}}\boldsymbol{n}'_{[i,i+s-1]}, \frac{-1}{\tilde{n}_{[i+s,j]}}\boldsymbol{n}'_{[i+s,j]}\right)\boldsymbol{1}_{j-i+1} = 0.$$

Therefore, we obtain

$$\begin{aligned}\text{Cov}[\boldsymbol{D}_{i,j}\boldsymbol{v}_{[i,j]}, \bar{v}_{[i,j]}] &= \boldsymbol{D}_{i,j}\tau^2\text{diag}(n_i^{-1}, \ldots, n_j^{-1})\frac{1}{\tilde{n}_{[i,j]}}\boldsymbol{n}_{[i,j]} \\ &= \frac{\tau^2}{\tilde{n}_{[i,j]}}\boldsymbol{D}_{i,j}\boldsymbol{1}_{j-i+1} = \boldsymbol{0}_{j-i},\end{aligned}$$

where $\text{diag}(a_1, \ldots, a_p)$ is a $p \times p$ diagonal matrix whose (s,s)th element is a_s. This implies $\boldsymbol{D}_{i,j}\boldsymbol{v}_{[i,j]} \perp\!\!\!\perp \bar{v}_{[i,j]}$.

Next, we prove (D.2). when $i = j$, since $\bar{v}_{[i,j]} = v_i$ we get

$$\sum_{s=i}^{j} n_s(v_s - \xi_s)(v_s - \bar{v}_{[i,j]}) = 0.$$

Hence, it is clear that (D.2) holds. Thus, we prove the case of $i < j$. From the definition of $\bar{v}_{[i,j]}$, it is easily checked that

$$\sum_{s=i}^{j} n_s\bar{v}_{[i,j]}(v_s - \bar{v}_{[i,j]}) = 0.$$

By using this result, we have

$$\begin{aligned}\sum_{s=i}^{j} n_s(v_s - \xi_s)(v_s - \bar{v}_{[i,j]}) &= \sum_{s=i}^{j} n_s(\{v_s - \xi_s - \bar{v}_{[i,j]}\} + \bar{v}_{[i,j]})(v_s - \bar{v}_{[i,j]}) \\ &= \sum_{s=i}^{j} n_s(v_s - \bar{v}_{[i,j]} - \xi_s)(v_s - \bar{v}_{[i,j]}) \\ &= \sum_{s=i}^{j} n_s(v_s - \bar{v}_{[i,j]})^2 - \sum_{s=i}^{j} n_s\xi_s(v_s - \bar{v}_{[i,j]}).\end{aligned}$$

Here, putting

$$\boldsymbol{A} = \text{diag}(n_i^{1/2}, \ldots, n_j^{1/2})\left\{\boldsymbol{I}_{j-i+1} - \frac{\boldsymbol{1}_{j-i+1}}{\tilde{n}_{[i,j]}}\boldsymbol{n}'_{[i,j]}\right\},$$

we get

$$\sum_{s=i}^{j} n_s(v_s - \xi_s)(v_s - \bar{v}_{[i,j]}) = (\boldsymbol{A}\boldsymbol{v}_{[i,j]})'(\boldsymbol{A}\boldsymbol{v}_{[i,j]}) - (\sqrt{n_i}\xi_i, \ldots, \sqrt{n_j}\xi_j)\boldsymbol{A}\boldsymbol{v}_{[i,j]}.$$

Therefore, it is sufficient to prove $\boldsymbol{A}\boldsymbol{v}_{[i,j]} \perp\!\!\!\perp \bar{v}_{[i,j]}$. by using the same argument, it is easily checked that $((\boldsymbol{A}\boldsymbol{v}_{[i,j]})', \bar{v}_{[i,j]})' \sim N_{j-i+2}(*, \star)$. Thus, we prove $\mathrm{Cov}[\boldsymbol{A}\boldsymbol{v}_{[i,j]}, \bar{v}_{[i,j]}] = \boldsymbol{0}_{j-i+1}$. From the definitions of $\boldsymbol{A}\boldsymbol{v}_{[i,j]}$ and $\bar{v}_{[i,j]}$, we obtain

$$\begin{aligned}\mathrm{Cov}[\boldsymbol{A}\boldsymbol{v}_{[i,j]}, \bar{v}_{[i,j]}] &= \frac{\tau^2}{\tilde{n}_{[i,j]}}\boldsymbol{A}\ \mathrm{diag}(n_i^{-1}, \ldots, n_j^{-1})\ \boldsymbol{n}_{[i,j]} = \frac{\tau^2}{\tilde{n}_{[i,j]}}\boldsymbol{A}\boldsymbol{1}_{j-i+1} \\ &= \frac{\tau^2}{\tilde{n}_{[i,j]}}\mathrm{diag}(n_i^{1/2}, \ldots, n_j^{1/2})\left\{\boldsymbol{1}_{j-i+1} - \frac{\boldsymbol{1}_{j-i+1}}{\tilde{n}_{[i,j]}}\boldsymbol{n}'_{[i,j]}\boldsymbol{1}_{j-i+1}\right\} = \boldsymbol{0}_{j-i+1}.\end{aligned}$$

This implies $\boldsymbol{A}\boldsymbol{v}_{[i,j]} \perp\!\!\!\perp \bar{v}_{[i,j]}$. Therefore, (D.2) holds. □

Appendix E: Lemma E and its proof

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

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

Then, it holds that

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

Proof. From the definition of the indicator function, it is clear that the fourth equality holds. Therefore, first, we show the first and third equalities. In other words, we show

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

If $\boldsymbol{v} \in \mathcal{A}_l^l$, it holds that

$$\min_{\boldsymbol{y} \in \mathcal{A}^l} \|\boldsymbol{v} - \boldsymbol{y}\|_{\boldsymbol{n}}^2 = 0,$$

because $\mathcal{A}_l^l \subset \mathcal{A}^l$. Hence, noting that $\boldsymbol{\eta}_l(\boldsymbol{v}) = \boldsymbol{v} \in \mathcal{A}_l^l$, we get $\boldsymbol{v} \in \boldsymbol{\eta}_l^{-1}(\mathcal{A}_l^l)$. On the other hand, recall that for the element $\boldsymbol{w} = (w_1, \ldots, w_l)' = (1, \ldots, l)' \in \mathcal{W}_l^l$, the set $\mathcal{A}_l^l$ is equal to the set $\mathcal{A}_l^l(\boldsymbol{w})$. Thus, if $\boldsymbol{v} \in \boldsymbol{\eta}_l^{-1}(\mathcal{A}_l^l) = \boldsymbol{\eta}_l^{-1}(\mathcal{A}_l^l(\boldsymbol{w}))$, from (C.2) of Lemma C we obtain

$$\bar{v}_{[1+0,1]} < \bar{v}_{[1+1,2]} < \cdots < \bar{v}_{[1+(l-1),l]}.$$

Hence, noting that $\bar{v}_{[s,s]} = v_s$, we get $v_1 < v_2 < \cdots < v_l$. This implies $\boldsymbol{v} \in \mathcal{A}_l^l$.

Next we show the second equality. For any s with $1 \le s \le l$, we put

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

These $z_1, \ldots, z_l$ are independently distributed as $N(0,1)$. Moreover, using z_s and b_s we have

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

Furthermore, for any t with $2 \le t \le l$, putting

$$\frac{\sqrt{n_t}}{\sqrt{n_{t-1}}} = a_{t-1},$$

it holds that

$$\boldsymbol{v} \in \mathcal{A}_l^l \Leftrightarrow 2 \le t \le l,\ v_{t-1} < v_t \Leftrightarrow 2 \le t \le l,\ a_{t-1}(z_{t-1} + b_{t-1}) - b_t < z_t.$$

Here, let

$$E_l = \{(c_1, \ldots, c_l) \in \mathbb{R}^l \mid 2 \le t \le l,\ a_{t-1}(c_{t-1} + b_{t-1}) - b_t < c_t\}.$$

Then, for the element $\boldsymbol{z} = (z_1, \ldots, z_l)'$, it holds that $\boldsymbol{v} \in \mathcal{A}_l^l \Leftrightarrow \boldsymbol{z} \in E_l$. By using this result and (E.1), we get

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

where $\phi(x)$ is the probability density function of standard normal distribution. We prove (E.2) in the order of $l = 2$, $l = 3$ and $l \ge 4$.

First, when $l = 2$, (E.2) can be written as

$$\begin{aligned} &\int_{-\infty}^{\infty} \int_{a_1(z_1+b_1)-b_2}^{\infty} \{z_1(z_1 + b_1) + z_2(z_2 + b_2)\} \phi(z_1)\phi(z_2) dz_1 dz_2 \\ &= \int_{-\infty}^{\infty} z_1(z_1 + b_1)\phi(z_1) \left\{\int_{a_1(z_1+b_1)-b_2}^{\infty} \phi(z_2) dz_2\right\} dz_1 \\ &\quad + \int_{-\infty}^{\infty} \phi(z_1) \left\{\int_{a_1(z_1+b_1)-b_2}^{\infty} z_2(z_2 + b_2)\phi(z_2) dz_2\right\} dz_1. \end{aligned} \tag{E.3}$$

Using the integration by parts formula, the first part of the right hand side in (E.3) can be expressed as

$$\begin{aligned}
&\int_{-\infty}^{\infty} z_1(z_1+b_1)\phi(z_1)\left\{\int_{a_1(z_1+b_1)-b_2}^{\infty}\phi(z_2)dz_2\right\}dz_1 \\
&= \left[-\phi(z_1)(z_1+b_1)\left\{\int_{a_1(z_1+b_1)-b_2}^{\infty}\phi(z_2)dz_2\right\}\right]_{-\infty}^{\infty} \\
&\quad + \int_{-\infty}^{\infty}\int_{a_1(z_1+b_1)-b_2}^{\infty}\phi(z_1)\phi(z_2)dz_1dz_2 \\
&\quad + \int_{-\infty}^{\infty}\phi(z_1)(z_1+b_1)\{-a_1\phi(a_1(z_1+b_1)-b_2)\}dz_1 \\
&= \int_{-\infty}^{\infty}\int_{a_1(z_1+b_1)-b_2}^{\infty}\phi(z_1)\phi(z_2)dz_1dz_2 \\
&\quad - \int_{-\infty}^{\infty} a_1(z_1+b_1)\{\phi(a_1(z_1+b_1)-b_2)\}\phi(z_1)dz_1.
\end{aligned}$$

On the other hand, noting that

$$\begin{aligned}
\int_{a_1(z_1+b_1)-b_2}^{\infty} z_2(z_2+b_2)\phi(z_2)dz_2 &= [-\phi(z_2)(z_2+b_2)]_{a_1(z_1+b_1)-b_2}^{\infty} \\
&\quad + \int_{a_1(z_1+b_1)-b_2}^{\infty}\phi(z_2)dz_2 \\
&= a_1(z_1+b_1)\phi(a_1(z_1+b_1)-b_2) \\
&\quad + \int_{a_1(z_1+b_1)-b_2}^{\infty}\phi(z_2)dz_2,
\end{aligned}$$

the second part of the right hand side in (E.3) can be written as

$$\begin{aligned}
&\int_{-\infty}^{\infty}\phi(z_1)\left\{\int_{a_1(z_1+b_1)-b_2}^{\infty} z_2(z_2+b_2)\phi(z_2)dz_2\right\}dz_1 \\
&= \int_{-\infty}^{\infty} a_1(z_1+b_1)\{\phi(a_1(z_1+b_1)-b_2)\}\phi(z_1)dz_1 \\
&\quad + \int_{-\infty}^{\infty}\int_{a_1(z_1+b_1)-b_2}^{\infty}\phi(z_1)\phi(z_2)dz_1dz_2.
\end{aligned}$$

Therefore, the right hand side in (E.3) is equal to

$$2\int_{-\infty}^{\infty}\int_{a_1(z_1+b_1)-b_2}^{\infty}\phi(z_1)\phi(z_2)dz_1dz_2 = 2\mathrm{E}[1_{\{\boldsymbol{z}\in E_2\}}] = 2\mathrm{E}[1_{\{\boldsymbol{v}\in\mathcal{A}_2^2\}}].$$

Therefore, when $l=2$, Lemma E is proved.

Next, we consider the case of $l = 3$. In this case, (E.2) can be written as

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{a_1(z_1+b_1)-b_2}^{\infty}\int_{a_2(z_2+b_2)-b_3}^{\infty}\left\{\sum_{s=1}^{3} z_s(z_s+b_s)\right\}\prod_{s=1}^{3}\phi(z_s)dz_1dz_2dz_3\\
&=\int_{-\infty}^{\infty} z_1(z_1+b_1)\phi(z_1)\left\{\int_{a_1(z_1+b_1)-b_2}^{\infty}\phi(z_2)\left(\int_{a_2(z_2+b_2)-b_3}^{\infty}\phi(z_3)dz_3\right)dz_2\right\}dz_1\\
&\quad+\int_{-\infty}^{\infty}\phi(z_1)\left\{\int_{a_1(z_1+b_1)-b_2}^{\infty} z_2(z_2+b_2)\phi(z_2)\left(\int_{a_2(z_2+b_2)-b_3}^{\infty}\phi(z_3)dz_3\right)dz_2\right\}dz_1\\
&\quad+\int_{-\infty}^{\infty}\int_{a_1(z_1+b_1)-b_2}^{\infty}\phi(z_1)\phi(z_2)\left(\int_{a_2(z_2+b_2)-b_3}^{\infty} z_3(z_3+b_3)\phi(z_3)dz_3\right)dz_1dz_2. \qquad \text{(E.4)}
\end{aligned}$$

Again, using the integration by parts formula, the first part of the right hand side in (E.4) can be expressed as

$$\begin{aligned}
&\left[-\phi(z_1)(z_1+b_1)\left\{\int_{a_1(z_1+b_1)-b_2}^{\infty}\phi(z_2)\left(\int_{a_2(z_2+b_2)-b_3}^{\infty}\phi(z_3)dz_3\right)dz_2\right\}\right]_{-\infty}^{\infty}\\
&\quad+\int_{-\infty}^{\infty}\int_{a_1(z_1+b_1)-b_2}^{\infty}\int_{a_2(z_2+b_2)-b_3}^{\infty}\phi(z_1)\phi(z_2)\phi(z_3)dz_1dz_2dz_3\\
&\quad+\int_{-\infty}^{\infty}\phi(z_1)(z_1+b_1)\{-a_1\phi(a_1(z_1+b_1)-b_2)\}\int_{a_1a_2(z_1+b_1)-b_3}^{\infty}\phi(z_3)dz_3dz_1\\
&=\int_{-\infty}^{\infty}\int_{a_1(z_1+b_1)-b_2}^{\infty}\int_{a_2(z_2+b_2)-b_3}^{\infty}\phi(z_1)\phi(z_2)\phi(z_3)dz_1dz_2dz_3\\
&\quad-\int_{-\infty}^{\infty}\int_{a_1a_2(z_1+b_1)-b_3}^{\infty}a_1(z_1+b_1)\phi(z_1)\phi\{a_1(z_1+b_1)-b_2\}\phi(z_3)dz_1dz_3.
\end{aligned}$$

Moreover, noting that

$$\begin{aligned}
&\left\{\int_{a_1(z_1+b_1)-b_2}^{\infty} z_2(z_2+b_2)\phi(z_2)\left(\int_{a_2(z_2+b_2)-b_3}^{\infty}\phi(z_3)dz_3\right)dz_2\right\}\\
&=\left[-\phi(z_2)(z_2+b_2)\left(\int_{a_2(z_2+b_2)-b_3}^{\infty}\phi(z_3)dz_3\right)\right]_{a_1(z_1+b_1)-b_2}^{\infty}\\
&\quad+\int_{a_1(z_1+b_1)-b_2}^{\infty}\int_{a_2(z_2+b_2)-b_3}^{\infty}\phi(z_2)\phi(z_3)dz_2dz_3\\
&\quad-\int_{a_1(z_1+b_1)-b_2}^{\infty}a_2(z_2+b_2)\phi(z_2)\phi\{a_2(z_2+b_2)-b_3\}dz_2\\
&=a_1(z_1+b_1)\phi\{a_1(z_1+b_1)-b_2\}\int_{a_1a_2(z_1+b_1)-b_3}^{\infty}\phi(z_3)dz_3\\
&\quad+\int_{a_1(z_1+b_1)-b_2}^{\infty}\int_{a_2(z_2+b_2)-b_3}^{\infty}\phi(z_2)\phi(z_3)dz_2dz_3\\
&\quad-\int_{a_1(z_1+b_1)-b_2}^{\infty}a_2(z_2+b_2)\phi(z_2)\phi\{a_2(z_2+b_2)-b_3\}dz_2,
\end{aligned}$$

the second term of the right hand side in (E.4) can be written as

$$\int_{-\infty}^{\infty}\int_{a_1a_2(z_1+b_1)-b_3}^{\infty} a_1(z_1+b_1)\phi(z_1)\phi\{a_1(z_1+b_1)-b_2\}\phi(z_3)dz_1dz_3$$
$$+\int_{-\infty}^{\infty}\int_{a_1(z_1+b_1)-b_2}^{\infty}\int_{a_2(z_2+b_2)-b_3}^{\infty}\phi(z_1)\phi(z_2)\phi(z_3)dz_1dz_2dz_3$$
$$-\int_{-\infty}^{\infty}\int_{a_1(z_1+b_1)-b_2}^{\infty}\phi(z_1)a_2(z_2+b_2)\phi(z_2)\phi\{a_2(z_2+b_2)-b_3\}dz_1dz_2.$$

Similarly, noting that

$$\left(\int_{a_2(z_2+b_2)-b_3}^{\infty} z_3(z_3+b_3)\phi(z_3)dz_3\right)$$
$$=[-\phi(z_3)(z_3+b_3)]_{a_2(z_2+b_2)-b_3}^{\infty}+\int_{a_2(z_2+b_2)-b_3}^{\infty}\phi(z_3)dz_3$$
$$=a_2(z_2+b_2)\phi\{a_2(z_2+b_2)-b_3\}+\int_{a_2(z_2+b_2)-b_3}^{\infty}\phi(z_3)dz_3,$$

the third term of the right hand side in (E.4) can be expressed as

$$\int_{-\infty}^{\infty}\int_{a_1(z_1+b_1)-b_2}^{\infty}\phi(z_1)a_2(z_2+b_2)\phi(z_2)\phi\{a_2(z_2+b_2)-b_3\}dz_1dz_2$$
$$+\int_{-\infty}^{\infty}\int_{a_1(z_1+b_1)-b_2}^{\infty}\int_{a_2(z_2+b_2)-b_3}^{\infty}\phi(z_1)\phi(z_2)\phi(z_3)dz_1dz_2dz_3.$$

Therefore, using these results the right hand side in (E.4) is given by

$$3\int_{-\infty}^{\infty}\int_{a_1(z_1+b_1)-b_2}^{\infty}\int_{a_2(z_2+b_2)-b_3}^{\infty}\phi(z_1)\phi(z_2)\phi(z_3)dz_1dz_2dz_3=3\mathrm{E}[1_{\{\boldsymbol{z}\in E_3\}}]$$
$$=3\mathrm{E}[1_{\{\boldsymbol{v}\in\mathcal{A}_3^3\}}].$$

Thus, when $l=3$, Lemma E is proved.

Finally, we prove the case of $l\geq 4$. In this case also we use the same argument as in the proof of $l=2$ and $l=3$. For any s with $1\leq s\leq l-1$, let

$$F_s(x)=\int_{a_s(x+b_s)-b_{s+1}}^{\infty}F_{s+1}(z_{s+1})\phi(z_{s+1})dz_{s+1},$$

and let $F_l(x) = 1$. Then, it holds that

$$\begin{aligned}
&\int \cdots \int_{E_l} \prod_{s=1}^{l} \phi(z_s) dz_1 \cdots dz_l \\
&= \int_{-\infty}^{\infty} F_1(z_1)\phi(z_1)dz_1 \\
&= \int_{-\infty}^{\infty} \left\{ \int_{a_1(z_1+b_1)-b_2}^{\infty} F_2(z_2)\phi(z_2)dz_2 \right\} \phi(z_1)dz_1 \\
&= \int_{-\infty}^{\infty} \int_{a_1(z_1+b_1)-b_2}^{\infty} \left\{ \int_{a_2(z_2+b_2)-b_3}^{\infty} F_3(z_3)\phi(z_3)dz_3 \right\} \phi(z_1)\phi(z_2)dz_1dz_2 \\
&= \int_{-\infty}^{\infty} \int_{a_1(z_1+b_1)-b_2}^{\infty} \cdots \int_{a_{i-2}(z_{i-2}+b_{i-2})-b_{i-1}}^{\infty} \left\{ \int_{a_{i-1}(z_{i-1}+b_{i-1})-b_i}^{\infty} F_i(z_i)\phi(z_i)dz_i \right\} \\
&\qquad \prod_{j=1}^{i-1} \phi(z_j) dz_1 \cdots d_{i-1}. \qquad \text{(E.5)}
\end{aligned}$$

Furthermore, for any i with $1 \le i \le l-1$, it holds that

$$\frac{d}{dz_i} F_i(z_i) = -a_i F_{i+1}\{a_i(z_i + b_i) - b_{i+1}\}\phi\{a_i(z_i + b_i) - b_{i+1}\}. \qquad \text{(E.6)}$$

Using these results, (E.2) can be expressed as

$$\int \cdots \int_{E_l} \left\{ \sum_{s=1}^{l} z_s(z_s + b_s) \right\} \prod_{s=1}^{l} \phi(z_s) dz_1 \cdots dz_l = G_1 + G_2 + G_3, \qquad \text{(E.7)}$$

where

$$\begin{aligned}
G_1 &= \int_{-\infty}^{\infty} z_1(z_1 + b_1)F_1(z_1)\phi(z_1)dz_1, \\
G_2 &= \sum_{i=2}^{l-1} \left[\int_{-\infty}^{\infty} \int_{a_1(z_1+b_1)-b_2}^{\infty} \cdots \int_{a_{i-2}(z_{i-2}+b_{i-2})-b_{i-1}}^{\infty} \right. \\
&\qquad \left. \left\{ \int_{a_{i-1}(z_{i-1}+b_{i-1})-b_i}^{\infty} z_i(z_i + b_i)F_i(z_i)\phi(z_i)dz_i \right\} \prod_{j=1}^{i-1} \phi(z_j) dz_1 \cdots d_{i-1} \right], \\
G_3 &= \int_{-\infty}^{\infty} \int_{a_1(z_1+b_1)-b_2}^{\infty} \cdots \int_{a_{l-2}(z_{l-2}+b_{l-2})-b_{l-1}}^{\infty} \\
&\qquad \left\{ \int_{a_{l-1}(z_{l-1}+b_{l-1})-b_l}^{\infty} z_l(z_l + b_l)\phi(z_l)dz_l \right\} \prod_{j=1}^{l-1} \phi(z_j) dz_1 \cdots d_{l-1}.
\end{aligned}$$

Next, we evaluate G_2. From (E.6), the brace { } of G_2 can be expanded as

$$\left\{\int_{a_{i-1}(z_{i-1}+b_{i-1})-b_i}^{\infty} z_i(z_i+b_i)F_i(z_i)\phi(z_i)dz_i\right\}$$
$$= [-\phi(z_i)(z_i+b_i)F_i(z_i)]_{a_{i-1}(z_{i-1}+b_{i-1})-b_i}^{\infty} + \int_{a_{i-1}(z_{i-1}+b_{i-1})-b_i}^{\infty} \phi(z_i)F_i(z_i)dz_i$$
$$- \int_{a_{i-1}(z_{i-1}+b_{i-1})-b_i}^{\infty} \phi(z_i)(z_i+b_i)a_iF_{i+1}\{a_i(z_i+b_i)-b_{i+1}\}\phi\{a_i(z_i+b_i)-b_{i+1}\}dz_i$$
$$= a_{i-1}(z_{i-1}+b_{i-1})\phi\{a_{i-1}(z_{i-1}+b_{i-1})-b_i\}F_i\{a_{i-1}(z_{i-1}+b_{i-1})-b_i\}$$
$$+ \int_{a_{i-1}(z_{i-1}+b_{i-1})-b_i}^{\infty} \phi(z_i)F_i(z_i)dz_i$$
$$- \int_{a_{i-1}(z_{i-1}+b_{i-1})-b_i}^{\infty} a_i(z_i+b_i)F_{i+1}\{a_i(z_i+b_i)-b_{i+1}\}\phi\{a_i(z_i+b_i)-b_{i+1}\}\phi(z_i)dz_i.$$

Hence, using this expansion and (E.5), the bracket [ ] of G_2 can be expressed as

$$\Bigg[\int_{-\infty}^{\infty}\int_{a_1(z_1+b_1)-b_2}^{\infty}\cdots\int_{a_{i-2}(z_{i-2}+b_{i-2})-b_{i-1}}^{\infty}$$
$$\left\{\int_{a_{i-1}(z_{i-1}+b_{i-1})-b_i}^{\infty} z_i(z_i+b_i)F_i(z_i)\phi(z_i)dz_i\right\}\prod_{j=1}^{i-1}\phi(z_j)dz_1\cdots d_{i-1}\Bigg]$$
$$= \int_{-\infty}^{\infty}\int_{a_1(z_1+b_1)-b_2}^{\infty}\cdots\int_{a_{i-2}(z_{i-2}+b_{i-2})-b_{i-1}}^{\infty}$$
$$a_{i-1}(z_{i-1}+b_{i-1})\phi\{a_{i-1}(z_{i-1}+b_{i-1})-b_i\}F_i\{a_{i-1}(z_{i-1}+b_{i-1})-b_i\}$$
$$\prod_{j=1}^{i-1}\phi(z_j)dz_1\cdots dz_{i-1}$$
$$+ \int\cdots\int_{E_l}\prod_{s=1}^{l}\phi(z_s)dz_1\cdots dz_l$$
$$- \int_{-\infty}^{\infty}\int_{a_1(z_1+b_1)-b_2}^{\infty}\cdots\int_{a_{i-1}(z_{i-1}+b_{i-1})-b_i}^{\infty}$$
$$a_i(z_i+b_i)\phi\{a_i(z_i+b_i)-b_{i+1}\}F_{i+1}\{a_i(z_i+b_i)-b_{i+1}\}$$
$$\prod_{j=1}^{i}\phi(z_j)dz_1\cdots dz_i. \tag{E.8}$$

Here, when $i=2$, from (E.5) we define

$$\int_{-\infty}^{\infty}\int_{a_1(z_1+b_1)-b_2}^{\infty}\cdots\int_{a_{i-2}(z_{i-2}+b_{i-2})-b_{i-1}}^{\infty} \equiv \int_{-\infty}^{\infty}.$$

Therefore, from (E.8) we obtain

$$G_2 = \int_{-\infty}^{\infty} a_1(z_1+b_1)\phi\{a_1(z_1+b_1)-b_2\}F_2\{a_1(z_1+b_1)-b_2\}\phi(z_1)dz_1$$
$$+(l-2)\int\cdots\int_{E_l}\prod_{s=1}^{l}\phi(z_s)dz_1\cdots dz_l$$
$$-\int_{-\infty}^{\infty}\int_{a_1(z_1+b_1)-b_2}^{\infty}\cdots\int_{a_{l-2}(z_{l-2}+b_{l-2})-b_{l-1}}^{\infty}$$
$$a_{l-1}(z_{l-1}+b_{l-1})\phi\{a_{l-1}(z_{l-1}+b_{l-1})-b_l\}\prod_{j=1}^{l-1}\phi(z_j)dz_1\cdots dz_{l-1}. \quad \text{(E.9)}$$

Next, we evaluate G_1 and G_3. From (E.5) and (E.6) we get

$$G_1 = [-\phi(z_1)(z_1+b_1)F_1(z_1)]_{-\infty}^{\infty} + \int_{-\infty}^{\infty}\phi(z_1)F_1(z_1)dz_1$$
$$-\int_{-\infty}^{\infty} a_1(z_1+b_1)\phi\{a_1(z_1+b_1)-b_2\}F_2\{a_1(z_1+b_1)-b_2\}\phi(z_1)dz_1$$
$$= \int\cdots\int_{E_l}\prod_{s=1}^{l}\phi(z_s)dz_1\cdots dz_l$$
$$-\int_{-\infty}^{\infty} a_1(z_1+b_1)\phi\{a_1(z_1+b_1)-b_2\}F_2\{a_1(z_1+b_1)-b_2\}\phi(z_1)dz_1. \quad \text{(E.10)}$$

Similarly, noting that

$$\left\{\int_{a_{l-1}(z_{l-1}+b_{l-1})-b_l}^{\infty} z_l(z_l+b_l)\phi(z_l)dz_l\right\} = [-\phi(z_l)(z_l+b_l)]_{a_{l-1}(z_{l-1}+b_{l-1})-b_l}^{\infty}$$
$$+\int_{a_{l-1}(z_{l-1}+b_{l-1})-b_l}^{\infty}\phi(z_l)dz_l$$
$$= a_{l-1}(z_{l-1}+b_{l-1})\phi\{a_{l-1}(z_{l-1}+b_{l-1})-b_l\}$$
$$+\int_{a_{l-1}(z_{l-1}+b_{l-1})-b_l}^{\infty}\phi(z_l)dz_l,$$

G_3 can be written as

$$G_3 = \int_{-\infty}^{\infty}\int_{a_1(z_1+b_1)-b_2}^{\infty}\cdots\int_{a_{l-2}(z_{l-2}+b_{l-2})-b_{l-1}}^{\infty}$$
$$a_{l-1}(z_{l-1}+b_{l-1})\phi\{a_{l-1}(z_{l-1}+b_{l-1})-b_l\}\prod_{j=1}^{l-1}\phi(z_j)dz_1\cdots dz_{l-1}$$
$$+\int\cdots\int_{E_l}\prod_{s=1}^{l}\phi(z_s)dz_1\cdots dz_l. \quad \text{(E.11)}$$

Hence, substituting (E.9), (E.10) and (E.11) into (E.7) yields

$$\int \cdots \int_{E_l} \left\{ \sum_{s=1}^{l} z_s(z_s + b_s) \right\} \prod_{s=1}^{l} \phi(z_s) dz_1 \cdots dz_l$$
$$= l \int \cdots \int_{E_l} \prod_{s=1}^{l} \phi(z_s) dz_1 \cdots dz_l = l\mathrm{E}[1_{\{\boldsymbol{z} \in E_l\}}] = l\mathrm{E}[1_{\{\boldsymbol{v} \in \mathcal{A}_l^l\}}].$$

Thus, when $l \geq 4$, Lemma E is proved. □

Appendix F: Lemma F and its proof

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

(P1) Suppose that i and j are integers with $1 \leq i \leq j \leq 2$. Then, it holds that

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

(P2) For the element $\boldsymbol{w} = (2)' \in \mathcal{W}_1^2$, it holds that

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

Proof. First, we prove (P1). Let i and j be integers with $1 \leq i \leq j \leq 2$. Here, when $i = j$ it holds that

$$\frac{1}{\tau^2} \sum_{s=i}^{j} n_s(x_s - \xi_s)(x_s - \bar{x}_{[i,j]}^{(\boldsymbol{n})}) = 0,$$

because $\bar{x}_{[i,j]}^{(\boldsymbol{n})} = x_i$. Thus, it is clear that (F.1) holds. Hence, it is sufficient to consider the case of $i < j$, (i.e., $i = 1$ and $j = 2$). In this case, the following equality holds:

$$\mathrm{E}\left[1_{\{\boldsymbol{D}_{1,2}^{(\boldsymbol{n})} \boldsymbol{x}_{[1,2]} \geq \boldsymbol{0}_1\}} \frac{1}{\tau^2} \sum_{s=1}^{2} n_s(x_s - \xi_s)(x_s - \bar{x}_{[1,2]}^{(\boldsymbol{n})}) \right] = X - Y, \quad \text{(F.3)}$$

where X and Y are given by

$$X = \mathrm{E}\left[1_{\{\boldsymbol{D}_{1,2}^{(\boldsymbol{n})} \boldsymbol{x}_{[1,2]} \geq \boldsymbol{0}_1\}} \frac{1}{\tau^2} \sum_{s=1}^{2} n_s(x_s - \xi_s)x_s \right],$$
$$Y = \mathrm{E}\left[1_{\{\boldsymbol{D}_{1,2}^{(\boldsymbol{n})} \boldsymbol{x}_{[1,2]} \geq \boldsymbol{0}_1\}} \frac{1}{\tau^2} \sum_{s=1}^{2} n_s(x_s - \xi_s)\bar{x}_{[1,2]}^{(\boldsymbol{n})} \right].$$

Here, we would like to note that

$$\begin{aligned}
\frac{1}{\tau^2}\sum_{s=1}^{2} n_s(x_s-\xi_s)\bar{x}_{[1,2]}^{(\boldsymbol{n})} &= \frac{\bar{x}_{[1,2]}^{(\boldsymbol{n})}}{\tau^2}\{n_1x_1+n_2x_2-(n_1\xi_1+n_2\xi_2)\} \\
&= \frac{\bar{x}_{[1,2]}^{(\boldsymbol{n})}}{\tau^2}\frac{n_1x_1+n_2x_2-(n_1\xi_1+n_2\xi_2)}{n_1+n_2}(n_1+n_2) \\
&= \frac{\bar{x}_{[1,2]}^{(\boldsymbol{n})}}{\tau^2}\left(\bar{x}_{[1,2]}^{(\boldsymbol{n})}-\frac{n_1\xi_1+n_2\xi_2}{n_1+n_2}\right)(n_1+n_2).
\end{aligned}$$

Thus, from Lemma D, noting that $\boldsymbol{D}_{1,2}^{(\boldsymbol{n})}\boldsymbol{x}_{[1,2]} \perp\!\!\!\perp \bar{x}_{[1,2]}^{(\boldsymbol{n})}$ we get

$$\begin{aligned}
Y &= \mathrm{E}\left[1_{\{\boldsymbol{D}_{1,2}^{(\boldsymbol{n})}\boldsymbol{x}_{[1,2]}\geq \boldsymbol{0}_1\}}\frac{1}{\tau^2}\sum_{s=1}^{2} n_s(x_s-\xi_s)\bar{x}_{[1,2]}^{(\boldsymbol{n})}\right] \\
&= \mathrm{E}\left[1_{\{\boldsymbol{D}_{1,2}^{(\boldsymbol{n})}\boldsymbol{x}_{[1,2]}\geq \boldsymbol{0}_1\}}\frac{\bar{x}_{[1,2]}^{(\boldsymbol{n})}}{\tau^2}\left(\bar{x}_{[1,2]}^{(\boldsymbol{n})}-\frac{n_1\xi_1+n_2\xi_2}{n_1+n_2}\right)(n_1+n_2)\right] \\
&= \mathrm{E}\left[1_{\{\boldsymbol{D}_{1,2}^{(\boldsymbol{n})}\boldsymbol{x}_{[1,2]}\geq \boldsymbol{0}_1\}}\right]\times \mathrm{E}\left[\frac{\bar{x}_{[1,2]}^{(\boldsymbol{n})}}{\tau^2}\left(\bar{x}_{[1,2]}^{(\boldsymbol{n})}-\frac{n_1\xi_1+n_2\xi_2}{n_1+n_2}\right)(n_1+n_2)\right].
\end{aligned}$$

In addition, since

$$\bar{x}_{[1,2]}^{(\boldsymbol{n})} \sim N\left(\frac{n_1\xi_1+n_2\xi_2}{n_1+n_2}, \frac{\tau^2}{n_1+n_2}\right),$$

it is clear that the second expectation of the last row is one. Hence, we have

$$Y = \mathrm{E}\left[1_{\{\boldsymbol{D}_{1,2}^{(\boldsymbol{n})}\boldsymbol{x}_{[1,2]}\geq \boldsymbol{0}_1\}}\right] = \mathrm{P}(\boldsymbol{D}_{1,2}^{(\boldsymbol{n})}\boldsymbol{x}_{[1,2]} \geq \boldsymbol{0}_1). \tag{F.4}$$

Next, we consider X. Recall that, from (C.1), the fact $\boldsymbol{x} \in \boldsymbol{\eta}_2^{-1}(\mathcal{A}_1^2(\boldsymbol{w})) \Leftrightarrow \boldsymbol{D}_{1,2}^{(\boldsymbol{n})}\boldsymbol{x}_{[1,2]} \geq \boldsymbol{0}_1$ holds. Moreover, from (i) of Lemma C, it holds that $\mathbb{R}^2 = \boldsymbol{\eta}_2^{-1}(\mathcal{A}_1^2(\boldsymbol{w}))\cup\boldsymbol{\eta}_2^{-1}(\mathcal{A}_2^2(\boldsymbol{w}^*))$ and $\boldsymbol{\eta}_2^{-1}(\mathcal{A}_1^2(\boldsymbol{w}))\cap\boldsymbol{\eta}_2^{-1}(\mathcal{A}_2^2(\boldsymbol{w}^*)) = \emptyset$. These imply that

$$1_{\{\boldsymbol{D}_{1,2}^{(\boldsymbol{n})}\boldsymbol{x}_{[1,2]}\geq \boldsymbol{0}_1\}} = 1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_2^{-1}(\mathcal{A}_1^2(\boldsymbol{w}))\}} = 1 - 1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_2^{-1}(\mathcal{A}_2^2(\boldsymbol{w}^*))\}} = 1 - 1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_2^{-1}(\mathcal{A}_2^2)\}}.$$

Therefore, we obtain

$$X = \mathrm{E}\left[\frac{1}{\tau^2}\sum_{s=1}^{2} n_s(x_s-\xi_s)x_s\right] - \mathrm{E}\left[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_2^{-1}(\mathcal{A}_2^2)\}}\frac{1}{\tau^2}\sum_{s=1}^{2} n_s(x_s-\xi_s)x_s\right]. \tag{F.5}$$

Here, it is easily checked that the first expectation of (F.5) is two because $x_s \sim N(\xi_s, \tau^2/n_s)$. On the other hand, from Lemma E the second expectation can be written as $2\mathrm{E}[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_2^{-1}(\mathcal{A}_2^2)\}}]$. Thus, using these results we get

$$\begin{aligned}
X = 2 - 2\mathrm{E}[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_2^{-1}(\mathcal{A}_2^2)\}}] = 2\mathrm{E}[1 - 1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_2^{-1}(\mathcal{A}_2^2)\}}] &= 2\mathrm{E}[1_{\{\boldsymbol{D}_{1,2}^{(\boldsymbol{n})}\boldsymbol{x}_{[1,2]}\geq \boldsymbol{0}_1\}}] \\
&= 2\mathrm{P}(\boldsymbol{D}_{1,2}^{(\boldsymbol{n})}\boldsymbol{x}_{[1,2]} \geq \boldsymbol{0}_1). \quad \text{(F.6)}
\end{aligned}$$

Therefore, substituting (F.4) and (F.6) into (F.3) yields (F.1). Hence, (P1) is proved.

Next, we prove (P2). Recall that from (i) of Lemma C we obtain $\mathbb{R}^2 = \boldsymbol{\eta}_2^{-1}(\mathcal{A}_1^2) \cup \boldsymbol{\eta}_2^{-1}(\mathcal{A}_2^2)$ and $\boldsymbol{\eta}_2^{-1}(\mathcal{A}_1^2) \cap \boldsymbol{\eta}_2^{-1}(\mathcal{A}_2^2) = \emptyset$. In addition, from (iii) of Lemma C it holds that

$$\boldsymbol{x} \in \boldsymbol{\eta}_2^{-1}(\mathcal{A}_1^2) \Rightarrow \eta_2(\boldsymbol{x})[1] = \eta_2(\boldsymbol{x})[2] = \bar{x}_{[1,2]},$$

and

$$\boldsymbol{x} \in \boldsymbol{\eta}_2^{-1}(\mathcal{A}_2^2) \Rightarrow \eta_2(\boldsymbol{x})[1] = x_1,\ \eta_2(\boldsymbol{x})[2] = x_2.$$

Hence, using these results and $\boldsymbol{x} \in \boldsymbol{\eta}_2^{-1}(\mathcal{A}_1^2) \Leftrightarrow \boldsymbol{D}_{1,2}^{(\boldsymbol{n})}\boldsymbol{x}_{[1,2]} \geq \boldsymbol{0}_1$, from (P1) of Lemma F we get

$$\begin{aligned}
&\mathrm{E}\left[\frac{1}{\tau^2}\sum_{s=1}^{2} n_s(x_s - \xi_s)(x_s - \eta_2^{(\boldsymbol{n})}(\boldsymbol{x})[s])\right] \\
&= \mathrm{E}\left[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_2^{-1}(\mathcal{A}_1^2)\}}\frac{1}{\tau^2}\sum_{s=1}^{2} n_s(x_s - \xi_s)(x_s - \bar{x}_{[1,2]})\right] \\
&= \mathrm{E}\left[1_{\{\boldsymbol{D}_{1,2}^{(\boldsymbol{n})}\boldsymbol{x}_{[1,2]}\geq\boldsymbol{0}_1\}}\frac{1}{\tau^2}\sum_{s=1}^{2} n_s(x_s - \xi_s)(x_s - \bar{x}_{[1,2]})\right] \\
&= \mathrm{P}(\boldsymbol{D}_{1,2}^{(\boldsymbol{n})}\boldsymbol{x}_{[1,2]} \geq \boldsymbol{0}_1) = \mathrm{P}(\boldsymbol{x} \in \boldsymbol{\eta}_2^{-1}(\mathcal{A}_1^2)).
\end{aligned}$$

Finally, from (iv) of Lemma C, we have $\mathrm{P}(\boldsymbol{x} \in \boldsymbol{\eta}_2^{-1}(\mathcal{A}_1^2)) = \mathrm{P}(\boldsymbol{\eta}_2^{(\boldsymbol{n})}(\boldsymbol{x}) \in \mathcal{A}_1^2)$. Therefore, (F.2) holds because $\mathcal{A}_1^2 = \mathcal{A}_1^2(\boldsymbol{w})$ for the element $\boldsymbol{w} = (2)' \in \mathcal{W}_1^2$. Consequently, Lemma F is proved. □

Appendix G: Lemma G and proofs of both Lemma G and Lemma 3.1

Lemma G. 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 $y_s \sim N(\zeta_s, \varsigma^2/N_s)$, $(s = 1, \ldots, l)$. 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 all integers i and j with $1 \leq i \leq j \leq l$, it holds that

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

Then, the following proposition (P*) is also true:

(P*) Let $n_1, \dots, n_{l+1}$ and τ^2 be positive numbers, and let $\xi_1, \dots, \xi_{l+1}$ be real numbers. Let $x_1, \dots, x_{l+1}$ be independent random variables, and let $x_s \sim N(\xi_s, \tau^2/n_s)$, $(s = 1, \dots, l+1)$. Put $\boldsymbol{n} = (n_1, \dots, n_{l+1})'$, $\boldsymbol{\xi} = (\xi_1, \dots, \xi_{l+1})'$ and $\boldsymbol{x} = (x_1, \dots, x_{l+1})'$. Then, for all integers i and j with $1 \le i \le j \le l+1$, it holds that

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

and

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

Proof. First, we prove (G.2). Let i and j be integers with $1 \le i \le j \le l+1$. Here, when $0 \le j-i \le l-1$, without loss of generality we may replace $n_i, \dots, n_j$, $\xi_i, \dots, \xi_j$, $x_i, \dots, x_j$ and τ^2 with $N_1, \dots, N_g$, $\zeta_1, \dots, \zeta_g$, $y_1, \dots, y_g$ and ς^2, respectively. Note that $g = j-i+1$ and $1 \le g \le l$. We put $\boldsymbol{N} = (N_1, \dots, N_g)'$ and $\boldsymbol{y} = (y_1, \dots, y_g)'$. Since $x_{i-1+t} = y_t \sim N(\zeta_t, \varsigma^2/N_t)$ $(1 \le t \le g)$, from the definitions of the matrix $\boldsymbol{D}_{i,j}^{(\boldsymbol{n})}$ and $\bar{x}_{[i,j]}^{(\boldsymbol{n})}$, using (G.1) we get

$$\begin{aligned} &\mathrm{E}\left[1_{\{\boldsymbol{D}_{i,j}^{(\boldsymbol{n})}\boldsymbol{x}_{[i,j]} \ge \boldsymbol{0}_{j-i}\}} \frac{1}{\tau^2} \sum_{s=i}^{j} n_s (x_s - \xi_s)(x_s - \bar{x}_{[i,j]}^{(\boldsymbol{n})})\right] \\ &= \mathrm{E}\left[1_{\{\boldsymbol{D}_{1,g}^{(\boldsymbol{N})}\boldsymbol{y}_{[1,g]} \ge \boldsymbol{0}_{g-1}\}} \frac{1}{\varsigma^2} \sum_{t=1}^{g} N_t (y_t - \zeta_t)(y_t - \bar{y}_{[1,g]}^{(\boldsymbol{N})})\right] \\ &= (g-1)\mathrm{P}(\boldsymbol{D}_{1,g}^{(\boldsymbol{N})}\boldsymbol{y}_{[1,g]} \ge \boldsymbol{0}_{g-1}) = (j-i)\mathrm{P}(\boldsymbol{D}_{i,j}^{(\boldsymbol{n})}\boldsymbol{x}_{[i,j]} \ge \boldsymbol{0}_{j-i}). \end{aligned} \tag{G.4}$$

Hence, when $0 \le j-i \le l-1$, (G.2) is proved. Therefore, it is sufficient to prove the case of $j-i = l$, i.e., $i = 1$ and $j = l+1$. In this case, the following equality holds:

$$\mathrm{E}\left[1_{\{\boldsymbol{D}_{1,l+1}^{(\boldsymbol{n})}\boldsymbol{x}_{[1,l+1]} \ge \boldsymbol{0}_l\}} \frac{1}{\tau^2} \sum_{s=1}^{l+1} n_s (x_s - \xi_s)(x_s - \bar{x}_{[1,l+1]}^{(\boldsymbol{n})})\right] = X - Y, \tag{G.5}$$

where

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

Here, noting that

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

and $\bar{x}^{(\boldsymbol{n})}_{[1,l+1]} \sim N(\bar{\xi}^{(\boldsymbol{n})}_{[1,l+1]}, \tau^2/\tilde{n}_{[1,l+1]})$, from (D.1), Y can be expressed as

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

On the other hand, from (i) of Lemma C and (C.1) we obtain

$$1_{\{\boldsymbol{D}^{(\boldsymbol{n})}_{1,l+1}\boldsymbol{x}_{[1,l+1]}\geq \boldsymbol{0}_l\}} = 1 - \sum_{u=2}^{l+1}\sum_{\boldsymbol{w};\boldsymbol{w}\in\mathcal{W}_u^{l+1}} 1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(\mathcal{A}_u^{l+1}(\boldsymbol{w}))\}}. \tag{G.7}$$

Therefore, X can be expressed as

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

where the first term of the last row in (G.8) can be derived by using $x_s \sim N(\xi_s, \tau^2/n_s)$. Next, for any integer u with $2 \leq u \leq l+1$ and for any element $\boldsymbol{w} = (w_1, \ldots, w_u)'$ with $\boldsymbol{w} \in \mathcal{W}_u^{l+1}$, we calculate

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

From (ii) of Lemma C, it holds that

$$\begin{aligned} \boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(\mathcal{A}_u^{l+1}(\boldsymbol{w})) \Leftrightarrow\ & 0 \leq t \leq u-1,\ \boldsymbol{D}_{1+w_t,w_{t+1}}\boldsymbol{x}_{[1+w_t,w_{t+1}]} \geq \boldsymbol{0}_{w_{t+1}-w_t-1}, \\ & 0 \leq s \leq u-2,\ \bar{x}_{[1+w_s,w_{s+1}]} < \bar{x}_{[1+w_{s+1},w_{s+2}]}, \end{aligned} \tag{G.10}$$

where $w_0 = 0$ and $w_u = l+1$. Here, noting that

$$\begin{aligned}
&\frac{1}{\tau^2}\sum_{s=1}^{l+1} n_s(x_s-\xi_s)x_s \\
&= \frac{1}{\tau^2}\sum_{q=0}^{u-1}\sum_{s=1+w_q}^{w_{q+1}} n_s(x_s-\xi_s)x_s \\
&= \frac{1}{\tau^2}\sum_{q=0}^{u-1}\sum_{s=1+w_q}^{w_{q+1}} n_s(x_s-\xi_s)\{(x_s-\bar{x}_{[1+w_q,w_{q+1}]})+\bar{x}_{[1+w_q,w_{q+1}]}\} \\
&= \frac{1}{\tau^2}\sum_{q=0}^{u-1}\sum_{s=1+w_q}^{w_{q+1}} n_s(x_s-\xi_s)(x_s-\bar{x}_{[1+w_q,w_{q+1}]}) \\
&\quad +\frac{1}{\tau^2}\sum_{q=0}^{u-1}\sum_{s=1+w_q}^{w_{q+1}} n_s(x_s-\xi_s)\bar{x}_{[1+w_q,w_{q+1}]} \\
&= \frac{1}{\tau^2}\sum_{q=0}^{u-1}\sum_{s=1+w_q}^{w_{q+1}} n_s(x_s-\xi_s)(x_s-\bar{x}_{[1+w_q,w_{q+1}]}) \\
&\quad +\frac{1}{\tau^2}\sum_{q=0}^{u-1}\tilde{n}_{[1+w_q,w_{q+1}]}\bar{x}_{[1+w_q,w_{q+1}]}(\bar{x}_{[1+w_q,w_{q+1}]}-\bar{\xi}_{[1+w_q,w_{q+1}]}),
\end{aligned}$$

(G.9) can be rewritten as

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

where

$$\begin{aligned}
G &= \mathrm{E}\left[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(\mathcal{A}_u^{l+1}(\boldsymbol{w}))\}}\frac{1}{\tau^2}\sum_{q=0}^{u-1}\sum_{s=1+w_q}^{w_{q+1}} n_s(x_s-\xi_s)(x_s-\bar{x}_{[1+w_q,w_{q+1}]})\right], \\
H &= \mathrm{E}\left[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(\mathcal{A}_u^{l+1}(\boldsymbol{w}))\}}\frac{1}{\tau^2}\sum_{q=0}^{u-1}\tilde{n}_{[1+w_q,w_{q+1}]}\bar{x}_{[1+w_q,w_{q+1}]}(\bar{x}_{[1+w_q,w_{q+1}]}-\bar{\xi}_{[1+w_q,w_{q+1}]})\right].
\end{aligned}$$

It is clear that

$$(\bar{x}_{[1+w_0,w_1]},\boldsymbol{D}_{1+w_0,w_1}\boldsymbol{x}_{[1+w_0,w_1]})' \perp\!\!\!\perp \cdots \perp\!\!\!\perp (\bar{x}_{[1+w_{u-1},w_u]},\boldsymbol{D}_{1+w_{u-1},w_u}\boldsymbol{x}_{[1+w_{u-1},w_u]})', \tag{G.12}$$

and from (D.1) it holds that $\bar{x}_{[1+w_q,w_{q+1}]} \perp\!\!\!\perp \boldsymbol{D}_{1+w_q,w_{q+1}}\boldsymbol{x}_{[1+w_q,w_{q+1}]}$. Thus, using these and (G.10) we obtain

$$\begin{aligned}
H &= \mathrm{E}\left[1_{\{0\le t\le u-1,\ \boldsymbol{D}_{1+w_t,w_{t+1}}\boldsymbol{x}_{[1+w_t,w_{t+1}]}\ge \boldsymbol{0}_{w_{t+1}-w_t-1}\}}\right] \\
&\quad\times\mathrm{E}\Bigg[1_{\{\bar{x}_{[1+w_0,w_1]}<\cdots<\bar{x}_{[1+w_{u-1},w_u]}\}} \\
&\qquad \frac{1}{\tau^2}\sum_{q=0}^{u-1}\tilde{n}_{[1+w_q,w_{q+1}]}\bar{x}_{[1+w_q,w_{q+1}]}(\bar{x}_{[1+w_q,w_{q+1}]}-\bar{\xi}_{[1+w_q,w_{q+1}]})\Bigg],
\end{aligned} \tag{G.13}$$

Here, note that $\bar{x}_{[1+w_0,w_1]} < \cdots < \bar{x}_{[1+w_{u-1},w_u]}$ is equivalent to

$$(\bar{x}_{[1+w_0,w_1]}, \ldots, \bar{x}_{[1+w_{u-1},w_u]})' \in \mathcal{A}_u^u.$$

Furthermore, $\bar{x}_{[1+w_0,w_1]}, \ldots, \bar{x}_{[1+w_{u-1},w_u]}$ are independent random variable, and it holds that $\bar{x}_{[1+w_q,w_{q+1}]} \sim N(\bar{\xi}_{[1+w_q,w_{q+1}]}, \tau^2/\tilde{n}_{[1+w_q,w_{q+1}]})$ for any q with $0 \le q \le u-1$. Hence, from Lemma E we get

$$\begin{aligned} &\mathrm{E}\Bigg[1_{\{\bar{x}_{[1+w_0,w_1]} < \cdots < \bar{x}_{[1+w_{u-1},w_u]}\}} \\ &\qquad \frac{1}{\tau^2}\sum_{q=0}^{u-1} \tilde{n}_{[1+w_q,w_{q+1}]}\bar{x}_{[1+w_q,w_{q+1}]}(\bar{x}_{[1+w_q,w_{q+1}]} - \bar{\xi}_{[1+w_q,w_{q+1}]})\Bigg] \\ &= u\mathrm{E}\left[1_{\{\bar{x}_{[1+w_0,w_1]} < \cdots < \bar{x}_{[1+w_{u-1},w_u]}\}}\right]. \end{aligned} \tag{G.14}$$

From (G.10), substituting (G.14) into (G.13) yields

$$\begin{aligned} H &= \mathrm{E}\left[1_{\{0\le t\le u-1,\ \boldsymbol{D}_{1+w_t,w_{t+1}}\boldsymbol{x}_{[1+w_t,w_{t+1}]} \ge \mathbf{0}_{w_{t+1}-w_t-1}\}}\right] \\ &\qquad \times u\mathrm{E}\left[1_{\{\bar{x}_{[1+w_0,w_1]} < \cdots < \bar{x}_{[1+w_{u-1},w_u]}\}}\right] \\ &= u\mathrm{E}\left[1_{\{0\le t\le u-1,\ \boldsymbol{D}_{1+w_t,w_{t+1}}\boldsymbol{x}_{[1+w_t,w_{t+1}]} \ge \mathbf{0}_{w_{t+1}-w_t-1}\}} \times 1_{\{\bar{x}_{[1+w_0,w_1]} < \cdots < \bar{x}_{[1+w_{u-1},w_u]}\}}\right] \\ &= u\mathrm{E}[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(\mathcal{A}_u^{l+1}(\boldsymbol{w}))\}}]. \end{aligned} \tag{G.15}$$

On the other hand, using (G.10), (G.12) and both $(D.1)$ and $(D.2)$ of Lemma D, we obtain

$$\begin{aligned} G = {} &\mathrm{E}\left[1_{\{\bar{x}_{[1+w_0,w_1]} < \cdots < \bar{x}_{[1+w_{u-1},w_u]}\}}\right] \\ &\times \mathrm{E}\Bigg[1_{\{0\le t\le u-1,\ \boldsymbol{D}_{1+w_t,w_{t+1}}\boldsymbol{x}_{[1+w_t,w_{t+1}]} \ge \mathbf{0}_{w_{t+1}-w_t-1}\}} \\ &\qquad \frac{1}{\tau^2}\sum_{q=0}^{u-1}\sum_{s=1+w_q}^{w_{q+1}} n_s(x_s - \xi_s)(x_s - \bar{x}_{[1+w_q,w_{q+1}]})\Bigg]. \end{aligned} \tag{G.16}$$

Note that $\boldsymbol{D}_{1+w_0,w_1}\boldsymbol{x}_{[1+w_0,w_1]} \perp\!\!\!\perp \cdots \perp\!\!\!\perp \boldsymbol{D}_{1+w_{u-1},w_u}\boldsymbol{x}_{[1+w_{u-1},w_u]}$. Moreover, for any q and q^* with $q \ne q^*$, the random vector (or variable) $\boldsymbol{D}_{1+w_{q^*-1},w_{q^*}}\boldsymbol{x}_{[1+w_{q^*-1},w_{q^*}]}$ and

$$\sum_{s=1+w_q}^{w_{q+1}} n_s(x_s - \xi_s)(x_s - \bar{x}_{[1+w_q,w_{q+1}]})$$

are also independent. Therefore, (G.16) can be written as

$$
\begin{aligned}
G = & \mathrm{E}\left[1_{\{\bar{x}_{[1+w_0,w_1]}<\cdots<\bar{x}_{[1+w_{u-1},w_u]}\}}\right] \\
& \times \mathrm{E}\left[\sum_{q=0}^{u-1}\left\{1_{\{0\le t\le u-1,\ \boldsymbol{D}_{1+w_t,w_{t+1}}\boldsymbol{x}_{[1+w_t,w_{t+1}]}\ge \boldsymbol{0}_{w_{t+1}-w_t-1}\}}\right.\right. \\
& \qquad \left.\left.\frac{1}{\tau^2}\sum_{s=1+w_q}^{w_{q+1}} n_s(x_s-\xi_s)(x_s-\bar{x}_{[1+w_q,w_{q+1}]})\right\}\right] \\
= & \mathrm{E}\left[1_{\{\bar{x}_{[1+w_0,w_1]}<\cdots<\bar{x}_{[1+w_{u-1},w_u]}\}}\right] \\
& \times \mathrm{E}\left[\sum_{q=0}^{u-1}\left\{1_{\{0\le t\le u-1,\ t\neq q,\ \boldsymbol{D}_{1+w_t,w_{t+1}}\boldsymbol{x}_{[1+w_t,w_{t+1}]}\ge \boldsymbol{0}_{w_{t+1}-w_t-1}\}}\right\}\right. \\
& \qquad \left\{1_{\{\boldsymbol{D}_{1+w_q,w_{q+1}}\boldsymbol{x}_{[1+w_q,w_{q+1}]}\ge \boldsymbol{0}_{w_{q+1}-w_q-1}\}}\right. \\
& \qquad \left.\left.\frac{1}{\tau^2}\sum_{s=1+w_q}^{w_{q+1}} n_s(x_s-\xi_s)(x_s-\bar{x}_{[1+w_q,w_{q+1}]})\right\}\right] \\
= & \mathrm{E}\left[1_{\{\bar{x}_{[1+w_0,w_1]}<\cdots<\bar{x}_{[1+w_{u-1},w_u]}\}}\right] \\
& \times \sum_{q=0}^{u-1}\mathrm{E}\left[1_{\{0\le t\le u-1,\ t\neq q,\ \boldsymbol{D}_{1+w_t,w_{t+1}}\boldsymbol{x}_{[1+w_t,w_{t+1}]}\ge \boldsymbol{0}_{w_{t+1}-w_t-1}\}}\right] \\
& \qquad \mathrm{E}\left[1_{\{\boldsymbol{D}_{1+w_q,w_{q+1}}\boldsymbol{x}_{[1+w_q,w_{q+1}]}\ge \boldsymbol{0}_{w_{q+1}-w_q-1}\}}\right. \\
& \qquad \left.\frac{1}{\tau^2}\sum_{s=1+w_q}^{w_{q+1}} n_s(x_s-\xi_s)(x_s-\bar{x}_{[1+w_q,w_{q+1}]})\right]. \qquad (G.17)
\end{aligned}
$$

In addition, since $0 \le w_{q+1} - w_q - 1 \le l - 1$, from (G.4) we have

$$
\begin{aligned}
& \mathrm{E}\left[1_{\{\boldsymbol{D}_{1+w_q,w_{q+1}}\boldsymbol{x}_{[1+w_q,w_{q+1}]}\ge \boldsymbol{0}_{w_{q+1}-w_q-1}\}}\right. \\
& \qquad \left.\frac{1}{\tau^2}\sum_{s=1+w_q}^{w_{q+1}} n_s(x_s-\xi_s)(x_s-\bar{x}_{[1+w_q,w_{q+1}]})\right] \\
& = (w_{q+1}-w_q-1)\mathrm{E}\left[1_{\{\boldsymbol{D}_{1+w_q,w_{q+1}}\boldsymbol{x}_{[1+w_q,w_{q+1}]}\ge \boldsymbol{0}_{w_{q+1}-w_q-1}\}}\right]. \qquad (G.18)
\end{aligned}
$$

Thus, substituting (G.18) into (G.17) yields

$$
\begin{aligned}
G = & \mathrm{E}\left[1_{\{\bar{x}_{[1+w_0,w_1]}<\cdots<\bar{x}_{[1+w_{u-1},w_u]}\}}\right] \\
& \times \sum_{q=0}^{u-1}(w_{q+1}-w_q-1)\mathrm{E}\left[1_{\{0\le t\le u-1,\ \boldsymbol{D}_{1+w_t,w_{t+1}}\boldsymbol{x}_{[1+w_t,w_{t+1}]}\ge \boldsymbol{0}_{w_{t+1}-w_t-1}\}}\right] \\
= & \mathrm{E}\left[1_{\{\bar{x}_{[1+w_0,w_1]}<\cdots<\bar{x}_{[1+w_{u-1},w_u]},\ 0\le t\le u-1,\ \boldsymbol{D}_{1+w_t,w_{t+1}}\boldsymbol{x}_{[1+w_t,w_{t+1}]}\ge \boldsymbol{0}_{w_{t+1}-w_t-1}\}}\right] \\
& \times (w_u - w_0 - u) \\
= & (l+1-u)\mathrm{E}[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(\mathcal{A}_u^{l+1}(\boldsymbol{w}))\}}]. \qquad (G.19)
\end{aligned}
$$

Hence, substituting (G.15) and (G.19) into (G.11), we obtain

$$\mathrm{E}\left[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(\mathcal{A}_u^{l+1}(\boldsymbol{w}))\}}\frac{1}{\tau^2}\sum_{s=1}^{l+1}n_s(x_s-\xi_s)x_s\right]=(l+1)\mathrm{E}[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(\mathcal{A}_u^{l+1}(\boldsymbol{w})\}}]. \qquad \text{(G.20)}$$

Consequently, substituting (G.20) into (G.8) yields

$$\begin{aligned}X&=(l+1)\left\{1-\sum_{u=2}^{l+1}\sum_{\boldsymbol{w};\boldsymbol{w}\in\mathcal{W}_u^{l+1}}\mathrm{E}[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(\mathcal{A}_u^{l+1}(\boldsymbol{w}))\}}]\right\}\\&=(l+1)\mathrm{E}\left[1-\sum_{u=2}^{l+1}\sum_{\boldsymbol{w};\boldsymbol{w}\in\mathcal{W}_u^{l+1}}1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(\mathcal{A}_u^{l+1}(\boldsymbol{w}))\}}\right]\\&=(l+1)\mathrm{E}[1_{\{\boldsymbol{D}_{1,l+1}^{(\boldsymbol{n})}\boldsymbol{x}_{[1,l+1]}\geq\mathbf{0}_l\}}]=(l+1)\mathrm{P}(\boldsymbol{D}_{1,l+1}^{(\boldsymbol{n})}\boldsymbol{x}_{[1,l+1]}\geq\mathbf{0}_l), \qquad \text{(G.21)}\end{aligned}$$

where the third equality in (G.21) is derived by using (G.7). Finally, substituting (G.6) and (G.21) into (G.5), we obtain (G.2).

Next, we prove (G.3). From (i), (ii) and (iii) of Lemma C, we get

$$\begin{aligned}&\mathrm{E}\left[\frac{1}{\tau^2}\sum_{s=1}^{l+1}n_s(x_s-\xi_s)(x_s-\eta_{l+1}^{(\boldsymbol{n})}(\boldsymbol{x})[s])\right]\\&=\mathrm{E}\left[\sum_{u=1}^{l+1}\sum_{\boldsymbol{w};\boldsymbol{w}\in\mathcal{W}_u^{l+1}}\left\{1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(\mathcal{A}_u^{l+1}(\boldsymbol{w}))\}}\frac{1}{\tau^2}\sum_{s=1}^{l+1}n_s(x_s-\xi_s)(x_s-\eta_{l+1}^{(\boldsymbol{n})}(\boldsymbol{x})[s])\right\}\right]\\&=\mathrm{E}\left[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(\mathcal{A}_1^{l+1})\}}\frac{1}{\tau^2}\sum_{s=1}^{l+1}n_s(x_s-\xi_s)(x_s-\eta_{l+1}^{(\boldsymbol{n})}(\boldsymbol{x})[s])\right]\\&\quad+\mathrm{E}\left[\sum_{u=2}^{l+1}\sum_{\boldsymbol{w};\boldsymbol{w}\in\mathcal{W}_u^{l+1}}\left\{1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(\mathcal{A}_u^{l+1}(\boldsymbol{w}))\}}\frac{1}{\tau^2}\sum_{s=1}^{l+1}n_s(x_s-\xi_s)(x_s-\eta_{l+1}^{(\boldsymbol{n})}(\boldsymbol{x})[s])\right\}\right]\\&=\mathrm{E}\left[1_{\{\boldsymbol{D}_{1,l+1}^{(\boldsymbol{n})}\boldsymbol{x}_{[1,l+1]}\geq\mathbf{0}_l\}}\frac{1}{\tau^2}\sum_{s=1}^{l+1}n_s(x_s-\xi_s)(x_s-\bar{x}_{[1,l+1]})\right]\\&\quad+\mathrm{E}\left[\sum_{u=2}^{l+1}\sum_{\boldsymbol{w};\boldsymbol{w}\in\mathcal{W}_u^{l+1}}\left\{1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(\mathcal{A}_u^{l+1}(\boldsymbol{w}))\}}\frac{1}{\tau^2}\sum_{q=0}^{u-1}\sum_{s=1+w_q}^{w_{q+1}}n_s(x_s-\xi_s)(x_s-\bar{x}_{[1+w_q,w_{q+1}]})\right\}\right]\\&=\mathrm{E}\left[1_{\{\boldsymbol{D}_{1,l+1}^{(\boldsymbol{n})}\boldsymbol{x}_{[1,l+1]}\geq\mathbf{0}_l\}}\frac{1}{\tau^2}\sum_{s=1}^{l+1}n_s(x_s-\xi_s)(x_s-\bar{x}_{[1,l+1]})\right]+\sum_{u=2}^{l+1}\sum_{\boldsymbol{w};\boldsymbol{w}\in\mathcal{W}_u^{l+1}}G. \qquad \text{(G.22)}\end{aligned}$$

Therefore, from (G.2) and (G.19), (G.22) can be expressed as

$$\begin{aligned}
&\mathrm{E}\left[\frac{1}{\tau^2}\sum_{s=1}^{l+1} n_s(x_s-\xi_s)(x_s-\eta_{l+1}^{(\boldsymbol{n})}(\boldsymbol{x})[s])\right] \\
&= l\mathrm{P}(\boldsymbol{D}_{1,l+1}^{(\boldsymbol{n})}\boldsymbol{x}_{[1,l+1]} \geq \boldsymbol{0}_l) + \sum_{u=2}^{l+1}\sum_{\boldsymbol{w};\boldsymbol{w}\in\mathcal{W}_u^{l+1}}(l+1-u)\mathrm{E}[1_{\{\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(\mathcal{A}_u^{l+1}(\boldsymbol{w}))\}}] \\
&= l\mathrm{P}(\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(\mathcal{A}_1^{l+1})) + \sum_{u=2}^{l}(l+1-u)\sum_{\boldsymbol{w};\boldsymbol{w}\in\mathcal{W}_u^{l+1}}\mathrm{P}(\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(\mathcal{A}_u^{l+1}(\boldsymbol{w}))).
\end{aligned}$$

Here, note that $\mathcal{A}_1^{l+1}=\mathcal{A}_1^{l+1}(\boldsymbol{w})$ for the element $\boldsymbol{w}\in\mathcal{W}_1^{l+1}$. Thus, from (iv) of Lemma C, we have

$$\begin{aligned}
&l\mathrm{P}(\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(\mathcal{A}_1^{l+1})) + \sum_{u=2}^{l}(l+1-u)\sum_{\boldsymbol{w};\boldsymbol{w}\in\mathcal{W}_u^{l+1}}\mathrm{P}(\boldsymbol{x}\in\boldsymbol{\eta}_{l+1}^{-1}(\mathcal{A}_u^{l+1}(\boldsymbol{w}))) \\
&= l\mathrm{P}(\boldsymbol{\eta}_{l+1}(\boldsymbol{x})\in\mathcal{A}_1^{l+1}) + \sum_{u=2}^{l}(l+1-u)\mathrm{P}\left(\boldsymbol{\eta}_{l+1}(\boldsymbol{x})\in\bigcup_{\boldsymbol{w};\boldsymbol{w}\in\mathcal{W}_u^{l+1}}\mathcal{A}_u^{l+1}(\boldsymbol{w})\right) \\
&= \sum_{i=1}^{l}(l+1-i)\mathrm{P}\left(\boldsymbol{\eta}_{l+1}(\boldsymbol{x})\in\bigcup_{\boldsymbol{w};\boldsymbol{w}\in\mathcal{W}_i^{l+1}}\mathcal{A}_i^{l+1}(\boldsymbol{w})\right).
\end{aligned}$$

This implies that (G.3) holds. Hence, Lemma G is proved. □

Consequently, combining Lemma F and Lemma G we obtain Lemma 3.1.

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

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

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

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

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

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

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

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

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

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

Table 5.1. Some properties of the $\mathrm{AIC}_{\mathrm{SO}}$ and the ordinal AIC in Case A–F

		$\mathrm{AIC}_{\mathrm{SO}}$	Ordinal AIC
Case A	Restriction	SO	Non
	Risk	$\mathrm{E}[\mathrm{E}_\star[-2l(\hat{\boldsymbol{\theta}}, \hat{\sigma}^2; \boldsymbol{X}^\star)]]$	$\mathrm{E}[\mathrm{E}_\star[-2l(\bar{\boldsymbol{X}}, \bar{\sigma}^2; \boldsymbol{X}^\star)]]$
	Penalty term	$2(m+1)$	$2(k+1)$
	Bias to the risk	Asymptotically unbiased	Asymptotically unbiased
	Order of the bias	$O(N^{-1})$	$O(N^{-1})$
Case B	Restriction	SO	Non
	Risk	$\mathrm{E}[\mathrm{E}_\star[-2l(\hat{\boldsymbol{\theta}}, \sigma_*^2; \boldsymbol{X}^\star)]]$	$\mathrm{E}[\mathrm{E}_\star[-2l(\bar{\boldsymbol{X}}, \sigma_*^2; \boldsymbol{X}^\star)]]$
	Penalty term	$2m$	$2k$
	Bias to the risk	Unbiased	Unbiased
	Order of the bias	0	0
Case C	Restriction	SO	Non
	Risk	$\mathrm{E}[\mathrm{E}_\star[-2l(\hat{\hat{\boldsymbol{\theta}}}, \hat{\hat{\sigma}}^2; \boldsymbol{X}^\star, \boldsymbol{\iota})]]$	$\mathrm{E}[\mathrm{E}_\star[-2l(\bar{\bar{\boldsymbol{X}}}, \bar{\bar{\sigma}}^2; \boldsymbol{X}^\star, \boldsymbol{\iota})]]$
	Penalty term	$2(m^*+1)$	$2(k+1)$
	Bias to the risk	Asymptotically unbiased	Asymptotically unbiased
	Order of the bias	$O(N^{-1})$	$O(N^{-1})$
Case D	Restriction	SO	Non
	Risk	$\mathrm{E}[\mathrm{E}_\star[-2l(\hat{\hat{\boldsymbol{\theta}}}, \sigma_*^2; \boldsymbol{X}^\star, \boldsymbol{\iota})]]$	$\mathrm{E}[\mathrm{E}_\star[-2l(\bar{\bar{\boldsymbol{X}}}, \sigma_*^2; \boldsymbol{X}^\star, \boldsymbol{\iota})]]$
	Penalty term	$2m^*$	$2k$
	Bias to the risk	Unbiased	Unbiased
	Order of the bias	0	0
Case E	Restriction	SO	Non
	Risk	$\mathrm{E}[\mathrm{E}_\star[-2l(\hat{\boldsymbol{\vartheta}}, \bar{\varsigma}^2, \bar{\boldsymbol{Z}}, \bar{\tau}^2; \boldsymbol{Y}^\star, \boldsymbol{Z}^\star)]]$	$\mathrm{E}[\mathrm{E}_\star[-2l(\bar{\boldsymbol{Y}}, \bar{\varsigma}^2, \bar{\boldsymbol{Z}}, \bar{\tau}^2; \boldsymbol{Y}^\star, \boldsymbol{Z}^\star)]]$
	Penalty term	$2(m^\dagger+1+p)$	$2(k+1+p)$
	Bias to the risk	Asymptotically unbiased	Asymptotically unbiased
	Order of the bias	$O(N^{-1})$	$O(N^{-1})$
Case F	Restriction	SO	Non
	Risk	$\mathrm{E}[\mathrm{E}_\star[-2l(\hat{\boldsymbol{\vartheta}}, \varsigma_*^2, \bar{\boldsymbol{Z}}, \tau_*^2; \boldsymbol{Y}^\star, \boldsymbol{Z}^\star)]]$	$\mathrm{E}[\mathrm{E}_\star[-2l(\bar{\boldsymbol{Y}}, \varsigma_*^2, \bar{\boldsymbol{Z}}, \tau_*^2; \boldsymbol{Y}^\star, \boldsymbol{Z}^\star)]]$
	Penalty term	$2(m^\dagger+p-1)$	$2(k+p-1)$
	Bias to the risk	Unbiased	Unbiased
	Order of the bias	0	0

Note: m, m^* and $m^\dagger$ are given by , (4.2) , (5.5) and (5.12), respectively.

Table 6.1 Some properties of the $\mathrm{AIC_{SO}}$ and the pAIC in Case 1

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

Table 6.2 Some properties of the $\mathrm{AIC_{SO}}$ and the pAIC in Case 2

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

Table 6.3 Some properties of the $\mathrm{AIC_{SO}}$ and the pAIC in Case 3

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