isid/ms/2012/02
February 24, 2012
http://www.isid.ac.in/~statmath/eprints

Variance Estimation for Tree Order Restricted Models

Antar Bandyopadhyay
Sanjay Chaudhuri

Indian Statistical Institute, Delhi Centre
7, SJSS Marg, New Delhi–110 016, India

Variance Estimation for Tree Order Restricted Models

Antar Bandyopadhyay
Indian Statistical Institute,
Delhi and Kolkata.
antar@isid.ac.in

Sanjay Chaudhuri
Department of Statistics and Applied Probability,
National University of Singapore, Singapore.
sanjay@stat.nus.edu.sg

February 12, 2012

Abstract

In this article we discuss estimation of the common variance of several normal populations with tree order restricted means. We discuss the asymptotic properties of the maximum likelihood estimator of the variance as the number of populations tends to infinity. We consider several cases of various orders of the sample sizes and show that the maximum likelihood estimator of the variance may or may not be consistent or be asymptotically normal.

1 Introduction

1.1 Background and Motivation

Tree order restrictions arise naturally in many important applications. One classical scenario is the comparison of several, say s, treatments with a known control or a placebo treatment. It is then natural to model the effect of the i^{th} treatment, say μ_i to be at least as large as the effect of the control treatment denoted by μ_0, that is, $\mu_0 \leq \mu_i$, for all $i = 1, 2, \ldots, s$.

Under the tree order restriction the parameter space for $\mu := (\mu_0; \mu_1, \ \mu_2, \ldots, \mu_s)$ forms a symmetric polyhedral cone in $\mathbb{R}^{s+1}$ with its spine along the line $\mu_0 = \mu_1 = \cdots = \mu_s$. It can be shown that under the normality assumption the *constrained* maximum likelihood estimator (MLE) of the mean vector μ is biased in many situations. In fact in [7] it was shown that if μ_i's remain bounded and the sample sizes $n_i^{(s)}$s remain bounded then the bias for μ_0 diverges to $-\infty$ as $s \to \infty$. Because of this phenomenon the constrained MLEs have been criticised severely in literature. Hwang and Peddada [6] wrote that the MLE "fails disastrously" and Cohen and Sackrowitz [4] remarked that the MLE is "undesirable".

Under certain conditions however, μ_0 is unbiased. It was shown by Chaudhuri and Perlman [3, 2] that if either $\mu_{(1)} = \min_{i \geq 1} \ \mu_i$ or $n_0^{(s)}$ grow sufficiently fast, then the MLE of μ_0 can be bounded from below in probability and it may even be consistent.

In all earlier works it is generally assumed that the treatment groups are homoscedastic but none considered estimation of the variance σ^2. In this article we discuss maximum likelihood estimation of σ^2 under normality assumption with the tree order restriction on the population means. We consider the asymptotic properties of this estimator as $s \to \infty$, that is, in the limit the dimension of μ becomes large. We show that, depending on the growth of $n_i^{(s)}$ with s the MLE $\hat{\sigma}^2_{(s)}$ is consistent under mild conditions, even though in some of these cases $\hat{\mu}^{(s)}$ is not consistent. Under stricter assumptions we also prove asymptotically normality for the estimator $\hat{\sigma}^2_{(s)}$.

1.2 Constrained Maximum Likelihood Estimator of μ and σ^2

Suppose there are $s+1$ independent populations indexed by 0, 1, 2, ..., s with unknown means $(\mu_i)_{i\geq 0}$ and unknown common variance σ^2. Let X_{i1}, X_{i2}, ..., $X_{in_i^{(s)}}$, be an i.i.d. sample from the i^{th} population.

Under the tree order restriction and the assumptions made above, the maximum likelihood estimator of μ is given by (see [9, 10])

$$\hat{\mu}_0^{(s)} = \min_{I\subseteq\{1,2,\dots,s\}} \frac{n_0^{(s)}\bar{X}_{0\cdot}^{(s)} + \sum_{i\in I} n_i^{(s)}\bar{X}_{i\cdot}^{(s)}}{n_0^{(s)} + \sum_{i\in I} n_i^{(s)}} \quad \text{and} \tag{1}$$

$$\hat{\mu}_i^{(s)} = \max\left(\hat{\mu}_0^{(s)}, \bar{X}_{i\cdot}^{(s)}\right). \tag{2}$$

Note that, equations (1) and (2) also give the least squared estimators of the mean vector for any distribution.

Under the tree order restriction the constrained maximum likelihood estimator of σ^2 is given by

$$\hat{\sigma}^2_{(s)} = \frac{1}{N^{(s)}}\left\{\sum_{j=1}^{n_0^{(s)}}\left(X_{0j}-\hat{\mu}_0^{(s)}\right)^2 + \sum_{i=1}^{s}\sum_{j=1}^{n_i^{(s)}}\left(X_{ij}-\hat{\mu}_i^{(s)}\right)^2\right\}. \tag{3}$$

1.3 Main Results

We make the following assumptions throughout this article:

A1: The mean vector μ is tree order restriction, that is, $\mu_0 \leq \mu_i$ for all $i \geq 1$.

A2: There exists $B > 0$ such that $\mu_i \leq B$ for all $i \geq 0$.

A3: The populations are Normal.

For simplicity we further assume that

A4: $n_1^{(s)} = n_2^{(s)} = \cdots = n_s^{(s)} = n^{(s)}$ and both $n_0^{(s)}$ and $n^{(s)}$ are non-decreasing in s.

Our main interest is to study the asymptotic properties of the maximum likelihood estimator of σ^2 namely $\hat{\sigma}^2_{(s)}$ as the number of populations becomes large.

Let $N^{(s)} = \sum_{i=0}^{s} n_i^{(s)}$ be the total sample size and we write $\bar{X}_{i\cdot}^{(s)} := \frac{1}{n_i^{(s)}}\sum_{j=1}^{n_i^{(s)}} X_{ij}$ for the sample mean of the i^{th} population where $1 \leq i \leq s$.

We first consider an example with two populations, with tree-ordered means. The size of the sample drawn from the population with larger mean increases linearly with s, while size of the placebo sample remains constant.

Theorem 1. *Consider two populations with a common variance σ^2 and means $\mu_0 \leq \mu_1$. Let $n_0^{(s)} = m$ and $n_1^{(s)} = m's$, where m, $m' \geq 1$ are two fixed integers. Then as $s \to \infty$*

$$\hat{\sigma}^2_{(s)} \longrightarrow \sigma^2 \quad a.s. \tag{4}$$

Moreover,

$$\sqrt{N^{(s)}}\left(\hat{\sigma}^2_{(s)} - \sigma^2\right) \xrightarrow{\mathbf{d}} N\left(0, 2\sigma^4\right). \tag{5}$$

Note that as pointed out in Section 3 it is not difficult to show that $\hat{\mu}_0^{(s)}$ is biased in this case, yet $\hat{\sigma}^2_{(s)}$ is consistent and a CLT holds.

The assumptions of Theorem 1 can be interpreted in the following alternative way. Suppose we consider $s+1$ populations with an unknown common variance σ^2 and $\mu_1 = \mu_2 = \cdots = \mu_s$ with $\mu_0 \le \mu_1$. Both μ_0 and μ_1 are unknown. Let $n_0^{(s)} = m$ and $n^{(s)} = m'$ be the sample sizes from these distributions. Clearly the MLEs of μ_0, μ_1 and σ^2 are exactly same as in Theorem 1. So it follows that in the limit as $s \to \infty$, the MLE $\hat{\sigma}^2_{(s)}$ is consistent and admits a CLT. It is worth mentioning here that the assumption that $\mu_1 = \mu_2 = \cdots = \mu_s$ is very crucial for the consistency and also for the CLT. This is because as stated in Theorem 4 below the consistency may fail and the simulations presented in Section 2 shows that CLT may not hold either.

Our next theorem deals with the case when the total sample size $N^{(s)}$ grows at a faster rate than s.

Theorem 2. *Suppose $N^{(s)} \to \infty$ then under the assumptions* **A1** $-$ **A4**

$$\mathbf{P}\left(0 \le \limsup_{s\to\infty} \hat{\sigma}^2_{(s)} \le \sigma^2\right) = 1\,. \tag{6}$$

Further if we assume that $s/N^{(s)} \longrightarrow 0$ as $s \to \infty$ then

$$\hat{\sigma}^2_{(s)} \to \sigma^2 \ \ a.s.\,,$$

while if $s/\sqrt{N^{(s)}} \longrightarrow 0$ then

$$\sqrt{N^{(s)}}\left(\hat{\sigma}^2_{(s)} - \sigma^2\right) \ \xrightarrow{\mathbf{d}}\ N\left(0, 2\sigma^4\right)\,.$$

From the proof of Theorem 2 in Section 3 we see that the result holds for any constrain on the mean vector μ which need not be just the tree order restriction. So this result may be used in other constrained problems, for example, in the study of the *isotonic regression model*, where one assumes that $\mu_0 \le \mu_1 \le \mu_2 \le \cdots \le \mu_s$.

Following result is immediate from Theorem 2 which covers the case when $n^{(s)} \equiv m$ fixed but $n_0^{(s)}$ is increasing at an appropriate rate.

Theorem 3. *Let $n_0^{(s)}$ be such that $s/n_0^{(s)} \longrightarrow 0$ as $s \to \infty$. Then under the assumptions* **A1** $-$ **A4** *$\hat{\sigma}^2_{(s)}$ is strongly consistent. Moreover, if $s/\sqrt{n_0^{(s)}} \longrightarrow 0$ as $s \to \infty$ then,*

$$\sqrt{N^{(s)}}\left(\hat{\sigma}^2_{(s)} - \sigma^2\right) \ \xrightarrow{\mathbf{d}}\ N\left(0, 2\sigma^4\right).$$

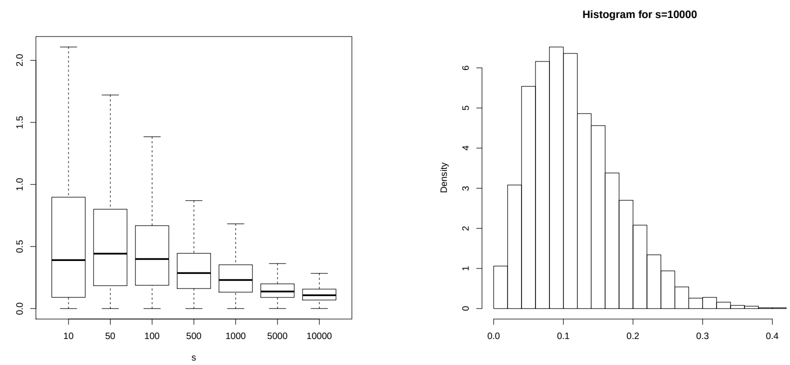
It is worth noting that under the conditions in Theorem 3 the MLE $\hat{\mu}_0^{(s)}$ is consistent [2].

From Theorem 2 the strong consistency holds if either $n_0^{(s)}/s \longrightarrow \infty$ or $n^{(s)} \longrightarrow \infty$ as $s \to \infty$. For the CLT to hold we need stronger condition, namely, $n_0^{(s)}/s^2 \longrightarrow \infty$ or $n^{(s)}/s \longrightarrow \infty$ as $s \to \infty$. In particular it covers the case when $n^{(s)}/\log s \longrightarrow \infty$ for which $\hat{\mu}_0^{(s)}$ is consistent if and only if $n_0^{(s)} \to \infty$ (see Proposition 1 in Section 4). In this case we have not been able to proof the CLT, which may hold (see Section 2 for simulated results).

Following theorem deals with the case when both $n_0^{(s)}$ and $n^{(s)}$ remain bounded.

Theorem 4. *Suppose $n_0^{(s)} = n^{(s)} = m$ for some fixed $m \ge 2$. Then under the assumptions* **A1** $-$ **A4**

$$\hat{\sigma}^2_{(s)} \longrightarrow \frac{m-1}{m}\sigma^2 \ \ a.s. \tag{7}$$



(a) (b)

Figure 1: Box plot (Figure 1(a)) and histogram for $s = 10000$ (Figure 1(b)) of $\xi_s = \sum_{i=0}^{s} n_i^{(s)} \left(\bar{X}_{i\cdot}^{(s)} - \hat{\mu}_i^{(s)}\right)^2 / \sqrt{N^{(s)}}$ when $n_0^{(s)} = n^{(s)} = m$.

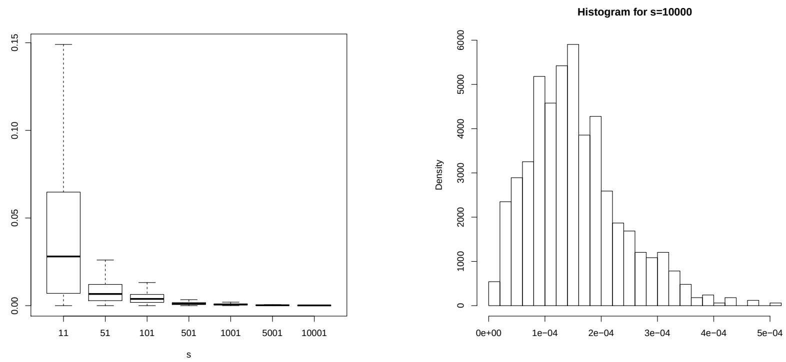
Like in Theorem 1 in this case also $N^{(s)} \sim ms$ but $\hat{\sigma}^2_{(s)}$ is not consistent. The difference is in the dimension of the mean parameter being estimated. In Theorem 1 it is exactly 2, however in Theorem 4 it grows unbounded with s. Once again in this case the CLT may not hold and we present some simulation results in Section 2.

The apparent ambiguity between Theorems 1 and 4 is reminiscent of the so called Neyman-Scott example [8]. They considered i.i.d. samples of equal finite size from several normal populations with unknown means and common variance. It was shown that in the limit if number of populations increases the MLE of the common variance is inconsistent. Here we observe the same phenomenon with tree order restriction on μ. For estimation of σ^2, μ is a nuisance parameter. In Theorem 1, $\mu_i = \mu_1$, for all $i \geq 1$. Thus even in the limit of $s \to \infty$, The number of nuisance parameters remain bounded. In contrast, in Theorem 4 this number increases unbounded. This explains the inconsistency of $\hat{\sigma}^2_{(s)}$ in the latter. Chaudhuri and Perlman [1] argue that in general, if the number of nuisance parameters is allowed to grow, the MLE of the parameter of interest may not be consistent. It even may not converge to any limit.

One important case which is not covered by the results described above is when $n_0^{(s)} = O(s)$ and $n^{(s)}$ remains bounded. Chaudhuri and Perlman [2] show that, in this case $\hat{\mu}_0^{(s)}$ remains bounded from below with high probability, but may not be consistent. In Section 2 we present some simulation results for this case.

1.4 Outline

The rest of the article is structured as follows. The next section gives the detailed simulation results of the cases mentioned above. In Section 3 we present the proofs of the main results. Section 4 contains some of the technical results and their proofs which we are used to prove the main results.



(a) (b)

Figure 2: Box plot (Figure 2(a)) and histogram for $s = 10000$ (Figure 2(b)) of ξ_s when $n_0^{(s)} = n^{(s)} = (\log s)^2$.

2 Simulation Studies for Some Unresolved Cases

In this section we perform a simulation study to explore the asymptotic behaviour of $\hat{\sigma}^2_{(s)}$ under some conditions which are not covered by the results in Section 1.3. In particular, we consider the following three situations:

1. We check if the CLT holds when $n_0^{(s)} = n^{(s)} = m$. Theorem 4 we know that $\hat{\sigma}^2_{(s)}$ is inconsistent for σ^2 in this case.

2. We check if CLT holds when $n_0^{(s)} = n^{(s)} = (\log s)^2$. Notice that in this case, $N^{(s)} = (s+1)(\log s)^2$ and from Theorem 2 it follows that $\hat{\sigma}^2_{(s)}$ is consistent for σ^2 in this case. However, $s/\sqrt{N^{(s)}} \not\to 0$, so the CLT cannot be derived from Theorem 2.

3. Asymptotic behaviour of $\hat{\sigma}^2_{(s)}$, when $n_0^{(s)} = O(s)$ and $n^{(s)}$ remains bounded is not covered by any result considered above. In this case even the consistency of $\hat{\mu}_0^{(s)}$ is not known, though it is bounded below with high probability [2].

In the simulation study for simplicity we assume all μ_i, $i \geq 0$ to be equal, which is equivalent to assuming $\mu = (0, 0, \ldots, 0)$. We assume $\sigma^2 = 1$. In order to study the asymptotic behaviour for large values of s, we consider s = 10, 50, 100, 500, 1000, 5000, 10000. The presented results are based on 2500 repetitions for each population size s.

Since, under all the above conditions the term $I_1 + I_3$ in (8) has a finite limit (in probability) and $\sqrt{N^{(s)}}(I_1 + I_3)$ converges in distribution, we concentrate on the random variable $\xi_s = \sqrt{N^{(s)}}(I_2 + I_4) = \sum_{i=0}^{s} n_i^{(s)} \left(\bar{X}_{i\cdot}^{(s)} - \hat{\mu}_i^{(s)}\right)^2 / \sqrt{N^{(s)}}$.

In the figures we present box and whisker plots of ξ_s for each case. We also present the histogram of ξ_s for $s = 10000$ in all cases. The top and the bottom of the boxes in the box whisker plots represent the first and third quartiles respectively. The line inside the box is the median. The whiskers were drawn to represent the most extreme data point still within the 1.5 times the inter quartile range of the first and the third quartiles. For better representation we have omitted more extreme points.

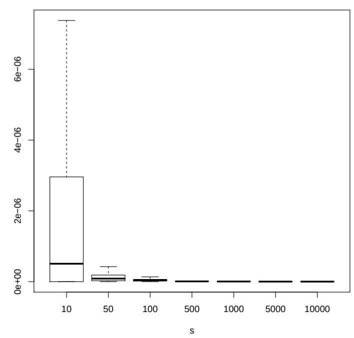
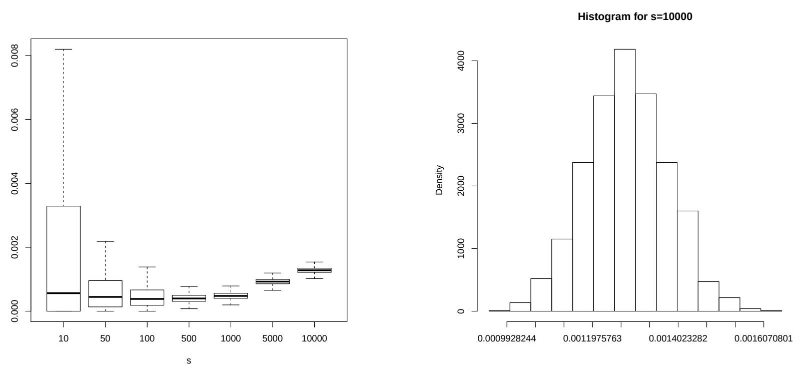


Figure 3: Box plot of $\xi_s/\sqrt{N^{(s)}} = I_2 + I_4$ when $n_0^{(s)} = O(s)$ and $n^{(s)} = m$.



(a) (b)

Figure 4: Box plot (Figure 4(a)) and histogram for $s = 10000$ (Figure 4(b)) of ξ_s when $n_0^{(s)} = O(s)$ and $n^{(s)} = m$

When $n_0^{(s)} = n^{(s)} = m$ from Figure 1(a) it is clear that the spread of the distribution reduces with s. Further, it is seen that the medians of the distributions have a decreasing trend with s. However, we cannot conclude that ξ_s converges to 0 in probability because the histogram in Figure 1(b) does not show any concentration near 0.

The case when $n_0^{(s)} = n^{(s)} = (\log s)^2$ is presented in Figures 2(a) and 2(b). From Figure 2(a) it seems that ξ_s converges to 0 in probability. The histogram (Figure 2(b)) also seems to be quite concentrated near 0. So a CLT may hold in this case.

The situation when $n_0^{(s)} = O(s)$ and $n^{(s)} = 100$ shows a different picture. As seen from the box plot in Figure 3, $\xi_s/\sqrt{N^{(s)}} = I_2 + I_4$ seems to be rapidly converging to 0 in probability, which indicates $\hat{\sigma}^2_{(s)}$ may be consistent. However, the box plot in Figure 4(a) and the histogram Figure 4(b) indicates that asymptotically ξ_s may not be converging to 0, thus the CLT may not hold. It is interesting to note that the histogram of ξ_s in Figure 4(b) is almost symmetric around its mean, for which we do not have an intuitive explanation.

3 Proofs of the Main Results

We start by observing that by Lemma 1 in Section 4

$$\hat{\sigma}^2_{(s)} = I_1 + I_2 + I_3 + I_4, \tag{8}$$

where

$$I_1 = \frac{1}{N^{(s)}} \sum_{j=1}^{n_0^{(s)}} \left(X_{0j} - \bar{X}_{0\cdot}^{(s)} \right)^2, \; I_2 = \frac{n_0^{(s)}}{N^{(s)}} \left(\bar{X}_{0\cdot}^{(s)} - \hat{\mu}_0^{(s)} \right)^2$$
$$I_3 = \frac{1}{N^{(s)}} \sum_{i=1}^{s} \sum_{j=1}^{n^{(s)}} \left(X_{ij} - \bar{X}_{i\cdot}^{(s)} \right)^2, \; I_4 = \frac{1}{N^{(s)}} \sum_{i=1}^{s} n^{(s)} \left(\bar{X}_{i\cdot}^{(s)} - \hat{\mu}_i^{(s)} \right)^2. \tag{9}$$

3.1 Proof of Theorem 1:

We have two populations and $N^{(s)} = m + m's$. Since $n_0^{(s)} = m$ and $n_1^{(s)} = m's$, $m/N^{(s)} \to 0$ and $m's/N^{(s)} \to 1$ as $s \to \infty$. Further $\bar{X}_{0\cdot}^{(s)}$ does not depend on s and X_{1j} are i.i.d random variables for $j = 1, 2, \ldots, m's$, with $E\left(X_{11}\right) = \mu_1$. So by the SLLN

$$\frac{m\bar{X}_{0\cdot}^{(s)} + \sum_{j=1}^{m's} X_{1j}}{N^{(s)}} \longrightarrow \mu_1 \qquad \text{a.s.}.$$

Now using the fact that $\bar{X}_{0\cdot}^{(s)}$ does not depend on s, it follows that:

$$\hat{\mu}_0^{(s)} \longrightarrow \min\left(\bar{X}_{0\cdot}^{(s)}, \mu_1\right) \;\; \text{a.s.} \tag{10}$$
$$\hat{\mu}_1^{(s)} \longrightarrow \max\left(\min\left(\bar{X}_{0\cdot}^{(s)}, \mu_1\right), \mu_1\right) = \mu_1 \;\; \text{a.s.} \tag{11}$$

Notice that in this case $I_1 = \frac{1}{N^{(s)}} \sum_{j=1}^{m} \left(X_{0j} - \bar{X}_{0\cdot}^{(s)} \right)^2$ and $\bar{X}_{0\cdot}^{(s)}$ does not depend on s so

$$\lim_{s\to\infty} I_1 = 0\,. \tag{12}$$

Also $I_2 = \frac{n_0^{(s)}}{N^{(s)}} \left(\bar{X}_{0\cdot}^{(s)} - \hat{\mu}_0^{(s)} \right)^2$, so by equation (10) we get

$$\lim_{s\to\infty} I_2 = 0\,. \tag{13}$$

Further $I_4 = \frac{m's}{N^{(s)}} \left(\bar{X}_{1\cdot}^{(s)} - \hat{\mu}_1^{(s)} \right)^2$, so using equation (11) we get

$$\lim_{s\to\infty} I_4 = 0\,. \tag{14}$$

Now observe that by standard SLLN

$$I_3 = \frac{1}{N^{(s)}} \sum_{j=1}^{m's} \left(X_{1j} - \bar{X}_{1\cdot}^{(s)} \right)^2 \longrightarrow \sigma^2 \;\; \text{a.s.} \tag{15}$$

So collecting the terms in (8) we get $\hat{\sigma}^2_{(s)} \longrightarrow \sigma^2$ a.s. proving the strong consistency.

To show the asymptotic normality we consider $\sqrt{N^{(s)}}\left(\hat{\sigma}^2_{(s)} - \sigma^2\right)$. Thus by similar argument as in equations (12) and (13) prove that

$$\lim_{s\to\infty} \sqrt{N^{(s)}} I_1 = 0 = \lim_{s\to\infty} \sqrt{N^{(s)}} I_2\,. \tag{16}$$

Further note that

$$
\begin{aligned}
\bar{X}_{1\cdot}^{(s)} - \hat{\mu}_1^{(s)} &= min\left(0, \bar{X}_{1\cdot}^{(s)} - \hat{\mu}_0^{(s)}\right) \\
&= min\left(0, \bar{X}_{1\cdot}^{(s)} - min\left(\bar{X}_{0\cdot}^{(s)}, \frac{m\bar{X}_{0\cdot}^{(s)} + m's\bar{X}_{1\cdot}^{(s)}}{m+m's}\right)\right) \\
&= min\left(0, max\left(\bar{X}_{1\cdot}^{(s)} - \bar{X}_{0\cdot}^{(s)}, \frac{m}{N^{(s)}}\left(\bar{X}_{1\cdot}^{(s)} - \bar{X}_{0\cdot}^{(s)}\right)\right)\right) \\
&= \frac{m}{N^{(s)}}\left(\bar{X}_{1\cdot}^{(s)} - \bar{X}_{0\cdot}^{(s)}\right)\mathbf{1}_{\left\{\bar{X}_{1\cdot}^{(s)} \leq \bar{X}_{0\cdot}^{(s)}\right\}}. \qquad (17)
\end{aligned}
$$

So it follows that

$$
\begin{aligned}
\sqrt{N^{(s)}} I_4 &= \frac{m's}{\sqrt{N^{(s)}}}\left(\bar{X}_{1\cdot}^{(s)} - \hat{\mu}_1^{(s)}\right)^2 \\
&= \frac{m^2 m'}{(N^{(s)})^{5/2}}\left(\bar{X}_{1\cdot}^{(s)} - \bar{X}_{0\cdot}^{(s)}\right)^2 \mathbf{1}_{\left\{\bar{X}_{1\cdot}^{(s)} \leq \bar{X}_{0\cdot}^{(s)}\right\}} \\
&\longrightarrow 0 \text{ a.s.} \qquad (18)
\end{aligned}
$$

Now using the standard CLT we get

$$
\sqrt{N^{(s)}}\left(I_3 - \sigma^2\right) = \sqrt{N^{(s)}}\left(\frac{1}{N^{(s)}}\sum_{j=1}^{m's}\left(X_{1j} - \bar{X}_{1\cdot}^{(s)}\right)^2 - \sigma^2\right) \xrightarrow{\mathbf{d}} \mathrm{N}\left(0, 2\sigma^4\right). \qquad (19)
$$

Finally using equations (16), (18) and (19) we conclude that

$$
\sqrt{N^{(s)}}\left(\hat{\sigma}_{(s)}^2 - \sigma^2\right) \xrightarrow{\mathbf{d}} \mathrm{N}\left(0, 2\sigma^4\right).
$$

□

3.2 Proof of Theorem 2

We start by noting that since $\hat{\sigma}_{(s)}^2$ is the least squared estimator of σ^2 so

$$
\hat{\sigma}_{(s)}^2 \leq \frac{1}{N^{(s)}}\sum_{i=0}^{s}\sum_{j=1}^{n_i^{(s)}}\left(X_{ij} - \mu_i\right)^2, \qquad (20)
$$

provided that μ satisfies the required tree order restriction. Here we note that specific constraint such as the tree order restriction is not needed to claim equation (20), it will hold for any general constraint under which least square estimator is obtained as long as μ satisfies it.

Now it follows that

$$
\begin{aligned}
\hat{\sigma}_{(s)}^2 \leq \sigma_{(s)}^2 &=: \frac{1}{N^{(s)}}\sum_{i=0}^{s}\sum_{j=1}^{n_i^{(s)}}\left(X_{ij} - \mu_i\right)^2 \qquad (21) \\
&= \frac{1}{N^{(s)}}\left\{\sum_{i=0}^{s} n_i^{(s)}\left(\bar{X}_{i\cdot}^{(s)} - \mu_i\right)^2 + \sum_{i=0}^{s}\sum_{j=1}^{n_i^{(s)}}\left(X_{ij} - \bar{X}_{i\cdot}^{(s)}\right)^2\right\}
\end{aligned}
$$

Since $X_{ij} - \mu_i$ are i.i.d. $\mathrm{N}\left(0, \sigma^2\right)$, so by SLLN the first assertion of the theorem follows. Furthermore, from the fundamental decomposition, (8) and (9) we get

$$
0 \leq \frac{1}{N^{(s)}}\left[I_2 + I_4\right] \leq \frac{1}{N^{(s)}}\sum_{i=0}^{s} n_i^{(s)}\left(\bar{X}_{i\cdot}^{(s)} - \mu_i\right)^2. \qquad (22)
$$

Note that $\sqrt{n_i^{(s)}}\left(\bar{X}_{i\cdot}^{(s)}-\mu_i\right)\sim \mathrm{N}\left(0,\sigma^2\right)$, which implies $n_i^{(s)}\left(\bar{X}_{i\cdot}^{(s)}-\mu_i\right)^2\sim\sigma^2\chi_1^2$. Thus from the SLLN it follows that $(s+1)^{-1}\sum_{i=0}^{s} n_i^{(s)}\left(\bar{X}_{i\cdot}^{(s)}-\mu_i\right)^2\to 2\sigma^2$ as $s\to\infty$.

By assumption $(s+1)/N^{(s)}\to 0$, thus

$$0\le \frac{1}{N^{(s)}}\left[I_2+I_4\right]\le \frac{s+1}{N^{(s)}}\frac{1}{s+1}\sum_{i=0}^{s} n_i^{(s)}\left(\bar{X}_{i\cdot}^{(s)}-\mu_i\right)^2\to 0 \text{ a.s.}$$

Moreover,

$$0\le \hat{\sigma}^2_{(s)}-\sigma^2_{(s)}=\frac{s+1}{N^{(s)}}\frac{1}{s+1}\sum_{i=0}^{s}\left(\bar{X}_{i\cdot}^{(s)}-\mu_i\right)^2-\frac{1}{N^{(s)}}\left[I_2+I_4\right]\to 0 \text{ a.s.}$$

Now using the SLLN as before $\sigma^2_{(s)}\to\sigma^2$ almost surely. So $\hat{\sigma}^2_{(s)}$ is strongly consistent.

Now we assume that $s/\sqrt{N^{(s)}}\to 0$. To prove the CLT we observe that

$$\sqrt{N^{(s)}}\left(\hat{\sigma}^2_{(s)}-\sigma^2\right)=\sqrt{N^{(s)}}\left(I_1+I_3-\sigma^2\right)+\sqrt{N^{(s)}}\left(I_2+I_4\right).$$

Now from (22)

$$\sqrt{N^{(s)}}\left(I_2+I_4\right)\le\frac{1}{\sqrt{N^{(s)}}}\sum_{i=0}^{s} n_i^{(s)}\left(\bar{X}_{i\cdot}^{(s)}-\mu_i\right)^2\le\frac{s+1}{\sqrt{N^{(s)}}}\frac{1}{s+1}\sum_{i=0}^{s} n_i^{(s)}\left(\bar{X}_{i\cdot}^{(s)}-\mu_i\right)^2.$$

Thus, using similar argument as above we get $s/\sqrt{N^{(s)}}\to 0\implies\sqrt{N^{(s)}}\left(I_2+I_4\right)\to 0$ a.s.

Let us now denote $Y_s=\sum_{i=0}^{s}\sum_{j=1}^{n_i^{(s)}}\left(X_{ij}-\bar{X}_{i\cdot}^{(s)}\right)^2$. Note that $\sum_{j=1}^{n_i^{(s)}}\left(X_{ij}-\bar{X}_{i\cdot}^{(s)}\right)^2$ are i.i.d. $\sigma^2\chi^2_{n_i^{(s)}-1}$ distributed random variables. Thus $Y_s\sim\sigma^2\chi^2_{N^{(s)}-s-1}$. Now

$$\begin{aligned}\sqrt{N^{(s)}}\left(I_1+I_3-\sigma^2\right)&=\sqrt{N^{(s)}}\left(\frac{Y_s}{N^{(s)}}-\sigma^2\right)=\frac{Y_s-E[Y_s]-(s+1)\sigma^2}{\sqrt{N^{(s)}}}\\&=\sqrt{\frac{Var[Y_s]}{N^{(s)}}}\left\{\frac{Y_s-E[Y_s]}{\sqrt{Var[Y_s]}}\right\}-\frac{s+1}{\sqrt{Var[Y_s]}}\sigma^2.\end{aligned}$$

Note that from properties of chi-square distribution

$$\frac{Y_s-E[Y_s]}{\sqrt{Var[Y_s]}}\xrightarrow{\mathbf{d}}\mathrm{N}\left(0,1\right).$$

Also under our assumption $Var[Y_s]/N^{(s)}=2(N^{(s)}-s-1)\sigma^4/N^{(s)}\to 2\sigma^4$. This completes the proof. □

3.3 Proof of Theorem 4

By assumption $n_0^{(s)}=n^{(s)}=m$ and $N^{(s)}=m(1+s)$ thus from (9) it follows that

$$I_1+I_3=\frac{m-1}{m(s+1)}\sum_{i=0}^{s}\left\{\sum_{j=1}^{m}\frac{1}{m-1}\left(X_{ij}-\bar{X}_{i\cdot}^{(s)}\right)^2\right\}=\frac{m-1}{m}\frac{1}{s+1}\sum_{i=0}^{s}W_i\,,$$

where $W_i=\frac{1}{m-1}\sum_{j=1}^{m}\left(X_{ij}-\bar{X}_{i\cdot}^{(s)}\right)^2$ for $i\ge 0$. Then $\{W_i\}_{i\ge 0}$ are i.i.d. random variables with mean σ^2 and variance $2\sigma^4$. So by standard SLLN we get

$$I_1+I_3\to\frac{m-1}{m}\sigma^2\qquad\text{as.}$$

Now consider

$$I_2 = \frac{m}{N^{(s)}} \left[\left(\bar{X}_{0\cdot}^{(s)} - \hat{\mu}_0^{(s)} \right)^2 \right] \le \frac{m}{N^{(s)}} \left(\frac{\rho^{(s)}}{1+\rho^{(s)}} \right)^2 \left[(\lambda(s))^2 \right] \le \frac{\log s}{1+s} \frac{\left[(\lambda(s))^2 \right]}{\log s},$$

where the first inequality follows from Lemma 2. Now

$$\lambda(s) = \min_{1 \le i \le s} \left(\bar{X}_{i\cdot}^{(s)} - \bar{X}_{0\cdot}^{(s)} \right) = \min_{1 \le i \le s} \bar{X}_{i\cdot}^{(s)} - \bar{X}_{0\cdot}^{(s)} = \min_{1 \le i \le s} \left(\mu_i + \sigma Z_i/\sqrt{m} \right) - \bar{X}_{0\cdot}^{(s)}, \tag{23}$$

where $Z_i = \left(\bar{X}_{i\cdot}^{(s)} - \mu_i \right) / (\sigma/\sqrt{m})$. Note that $\{Z_i\}_{i \ge 1}$ are i.i.d $\mathrm{N}(0,1)$ random variables. From assumption **A1** and **A2** we get

$$\mu_0 + \left(\frac{\sigma}{\sqrt{m}} \right) \min_{1 \le i \le s} Z_i - \bar{X}_{0\cdot}^{(s)} \le \lambda(s) \le B + \left(\frac{\sigma}{\sqrt{m}} \right) \min_{1 \le i \le s} Z_i - \bar{X}_{0\cdot}^{(s)} \tag{24}$$

Observe that $\bar{X}_{0\cdot}^{(s)}$ does not depend on s and from [5], it follows that $\lambda(s)/\sqrt{2 \log s} \xrightarrow{\mathbf{P}} \sigma/\sqrt{m}$. From (23) it now follows that $I_2 \xrightarrow{\mathbf{P}} 0$.

Finally we consider that

$$I_4 = \frac{1}{1+s} \sum_{i=1}^{s} \left(\bar{X}_{i\cdot}^{(s)} - \hat{\mu}_i^{(s)} \right)^2.$$

Let $V_{is} = \left(\bar{X}_{i\cdot}^{(s)} - \hat{\mu}_i^{(s)} \right)^2$. From the definition of $\hat{\mu}_i^{(s)}$, it follows that

$$\bar{X}_{i\cdot}^{(s)} - \hat{\mu}_i^{(s)} = \bar{X}_{i\cdot}^{(s)} - \max \left(\bar{X}_{i\cdot}^{(s)}, \hat{\mu}_0^{(s)} \right) = \left(\bar{X}_{i\cdot}^{(s)} - \hat{\mu}_0^{(s)} \right) \mathbf{1}_{\left\{ \hat{\mu}_0^{(s)} > \bar{X}_{i\cdot}^{(s)} \right\}}.$$

From Lemma 2, $\hat{\mu}_0^{(s)} \le \bar{X}_{0\cdot}^{(s)}$, so

$$\left(\bar{X}_{i\cdot}^{(s)} - \bar{X}_{0\cdot}^{(s)} \right) \mathbf{1}_{\left\{ \hat{\mu}_0^{(s)} > \bar{X}_{i\cdot}^{(s)} \right\}} \le \left(\bar{X}_{i\cdot}^{(s)} - \hat{\mu}_0^{(s)} \right) \mathbf{1}_{\left\{ \hat{\mu}_0^{(s)} > \bar{X}_{i\cdot}^{(s)} \right\}} \le 0. \tag{25}$$

Thus

$$0 \le V_{is} = \left(\bar{X}_{i\cdot}^{(s)} - \hat{\mu}_0^{(s)} \right)^2 \mathbf{1}_{\left\{ \hat{\mu}_0^{(s)} > \bar{X}_{i\cdot}^{(s)} \right\}} \le \left(\bar{X}_{i\cdot}^{(s)} - \bar{X}_{0\cdot}^{(s)} \right)^2 \mathbf{1}_{\left\{ \hat{\mu}_0^{(s)} > \bar{X}_{i\cdot}^{(s)} \right\}} \tag{26}$$

Let $\hat{\mu}_{0(-i)}^{(s)}$ be the estimate of μ_0 obtained after dropping the i^{th} population, that is, using only the data $(X_{0k})_{1 \le k \le m}$, $\{X_{jk} | 1 \le k \le m\}_{1 \le j \le s, j \ne i}$.

Recall that $n_0^{(s)} = n^{(s)} = m$ and notice that

$$\hat{\mu}_0^{(s)} \le \hat{\mu}_{0(-i)}^{(s)} \le \hat{\eta}_{0(-i)}^{(s)} = \min_{1 \le j \le s, i \ne j} \left(\frac{\bar{X}_{0\cdot}^{(s)} + \bar{X}_{j\cdot}^{(s)}}{2} \right) \qquad 1 \le k \le s.$$

So it follows that

$$\begin{aligned} V_{is} &= \left(\bar{X}_{i\cdot}^{(s)} - \bar{X}_{0\cdot}^{(s)} \right)^2 \mathbf{1}_{\left\{ \hat{\mu}_0^{(s)} > \bar{X}_{i\cdot}^{(s)} \right\}} & (27) \\ &\le \left(\bar{X}_{i\cdot}^{(s)} - \bar{X}_{0\cdot}^{(s)} \right)^2 \mathbf{1}_{\left\{ \hat{\eta}_{0(-i)}^{(s)} > \bar{X}_{i\cdot}^{(s)} \right\}} \\ &= \left(\bar{X}_{i\cdot}^{(s)} - \bar{X}_{0\cdot}^{(s)} \right)^2 \prod_{\substack{j=2 \\ j \ne i}}^{s} \mathbf{1}_{\left\{ \frac{1}{2} \left(\bar{X}_{0\cdot}^{(s)} + \bar{X}_{j\cdot}^{(s)} \right) > \bar{X}_{i\cdot}^{(s)} \right\}} \\ &= \left(\bar{X}_{i\cdot}^{(s)} - \bar{X}_{0\cdot}^{(s)} \right)^2 \prod_{\substack{j=2 \\ j \ne i}}^{s} \mathbf{1}_{\left\{ \bar{X}_{j\cdot}^{(s)} > 2\bar{X}_{i\cdot}^{(s)} - \bar{X}_{0\cdot}^{(s)} \right\}} & (28) \end{aligned}$$

Let us denote $Z_i^{(s)} = 2\bar{X}_{i\cdot}^{(s)} - \bar{X}_{0\cdot}^{(s)}$ which has $\mathrm{N}\left(2\mu_i - \mu_0, 5\sigma^2/m\right)$ distribution. By taking expectation on both sides of equation (28) we get

$$
\begin{aligned}
E\left[V_{is}\right] &\leq E\left[\left(\bar{X}_{i\cdot}^{(s)} - \bar{X}_{0\cdot}^{(s)}\right)^2 \prod_{\substack{j=2\\ j\neq i}}^{s} \left\{1 - \Phi\left(\frac{\sqrt{m}}{\sigma}\left(Z_i^{(s)} - \mu_j\right)\right)\right\}\right] \\
&\leq E\left[\left(\bar{X}_{i\cdot}^{(s)} - \bar{X}_{0\cdot}^{(s)}\right)^2 \left\{1 - \Phi\left(\frac{\sqrt{m}}{\sigma}\left(Z_i^{(s)} - B\right)\right)\right\}^{(s-1)}\right]. \qquad (29)
\end{aligned}
$$

The last inequality holds since $\mu_0 \leq \mu_j \leq B$, for all $1 \leq j \leq s$ by assumption A2.

Now applying the Cauchy-Schwartz inequality on (29) we get

$$
E\left[V_{is}\right] \leq \sqrt{E\left[\left(\bar{X}_{i\cdot}^{(s)} - \bar{X}_{0\cdot}^{(s)}\right)^4\right]}\sqrt{E\left[\left\{1 - \Phi\left(\frac{\sqrt{m}}{\sigma}\left(Z_i^{(s)} - B\right)\right)\right\}^{2(s-1)}\right]} \qquad (30)
$$

Now notice that $E\left[\left(\bar{X}_{i\cdot}^{(s)} - \bar{X}_{0\cdot}^{(s)}\right)^4\right]$ does not depend on s and $\sqrt{m}\left(Z_i^{(s)} - B\right)/\sigma \sim \mathrm{N}\left(\sqrt{m}(2\mu_i - \mu_0 - B)/\sigma, 5\right)$ distribution. Furthermore, since $2\mu_i - \mu_0 - B \geq \mu_0 - B$, it follows that

$$
E\left[V_{is}\right] \leq C\sqrt{E\left[\left\{1 - \Phi\left(W\right)\right\}^{2(s-1)}\right]}, \qquad (31)
$$

where $C = \sqrt{E\left[\left(\bar{X}_{i\cdot}^{(s)} - \bar{X}_{0\cdot}^{(s)}\right)^4\right]}$ and $W \sim \mathrm{N}\left(\sqrt{m}(\mu_0 - B)/\sigma, 5\right)$ distribution.

Notice that the right-hand side of (31) does not depend on i, so

$$
E\left[I_4\right] = \frac{1}{s+1}\sum_{i=1}^{s} E\left[V_{is}\right] \leq C\frac{s}{s+1}\sqrt{E\left[\left\{1 - \Phi\left(W\right)\right\}^{2(s-1)}\right]}
$$

Now using the DCT we conclude that $E\left[I_4\right] \to 0$ as $s \to \infty$. This completes the proof. $\square$

4 Technical Results on the MLEs of μ and σ^2

In this section we present some technical results on the constrained MLEs of μ and σ^2 which we have used to prove the theorems. Our first result gives an easy but very important decomposition of the MLE $\hat{\sigma}^2_{(s)}$ of σ^2 which we refer as fundamental decomposition. The proof of this lemma is obvious. So we omit it.

Lemma 1. *The constrained MLE $\hat{\sigma}^2_{(s)}$ of σ^2 admits the following decomposition*

$$
\begin{aligned}
\hat{\sigma}^2_{(s)} &= \frac{1}{N^{(s)}}\sum_{j=1}^{n_0^{(s)}}\left(X_{0j} - \bar{X}_{0\cdot}^{(s)}\right)^2 + \frac{n_0^{(s)}}{N^{(s)}}\left(\bar{X}_{0\cdot}^{(s)} - \hat{\mu}_0^{(s)}\right)^2 \\
&\quad + \frac{1}{N^{(s)}}\sum_{i=1}^{s}\sum_{j=1}^{n^{(s)}}\left(X_{ij} - \bar{X}_{i\cdot}^{(s)}\right)^2 + \frac{n^{(s)}}{N^{(s)}}\sum_{i=1}^{s}\left(\bar{X}_{i\cdot}^{(s)} - \hat{\mu}_i^{(s)}\right)^2. \qquad (32)
\end{aligned}
$$

Note that the first and the third term in (32) do not involve the order restricted MLEs of the means. Form our assumptions, $N^{(s)}$ increases strictly with s, so the asymptotic behaviours of these two terms can be determined from classical results such as the SLLN and CLT of i.i.d. random variables with finite second moment.

The next result gives two very useful upper and lower bounds on the MLE $\hat{\mu}_0^{(s)}$ of μ_0.

Lemma 2. *Let* $\rho^{(s)} := sn^{(s)}/n_0^{(s)}$ *and* $\lambda^{(s)} := \min_{1\le i\le s}\left(\bar{X}_{i\cdot}^{(s)} - \bar{X}_{0\cdot}^{(s)}\right)$. *Then*

$$\bar{X}_{0\cdot}^{(s)} + \frac{\rho^{(s)}}{1+\rho^{(s)}}\lambda^{(s)}\mathbf{1}_{\{\lambda^{(s)}<0\}} \le \hat{\mu}_0^{(s)} \le \bar{X}_{0\cdot}^{(s)}. \tag{33}$$

Proof: Let $S = \{1, 2, \ldots, s\}$. By definition

$$\hat{\mu}_0^{(s)} = \min_{I\subseteq S} \frac{n_0^{(s)}\bar{X}_{0\cdot}^{(s)} + n^{(s)}\sum_{i\in I}\bar{X}_{i\cdot}^{(s)}}{n_0^{(s)} + n^{(s)}|I|} = \bar{X}_{0\cdot}^{(s)} + \min_{I\subseteq S}\frac{n^{(s)}\sum_{i\in I}\left(\bar{X}_{i\cdot}^{(s)} - \bar{X}_{0\cdot}^{(s)}\right)}{n_0^{(s)} + n^{(s)}|I|}.$$

Suppose $\Lambda_i^{(s)} = \bar{X}_{i\cdot}^{(s)} - \bar{X}_{0\cdot}^{(s)}$ and $\lambda^{(s)} = \min_{1\le i\le s}\Lambda_i^{(s)}$. Fix $I \subseteq S$, $I \neq \emptyset$. Then

$$|I|\lambda^{(s)} \le \sum_{i\in I}\Lambda_i^{(s)} \Rightarrow \frac{n^{(s)}|I|\lambda^{(s)}}{n_0^{(s)} + |I|n^{(s)}} \le \frac{n^{(s)}\sum_{i\in I}\Lambda_i^{(s)}}{n_0^{(s)} + |I|n^{(s)}}.$$

Taking minimum on both sides we get:

$$\min_{I\subseteq S, I\neq\emptyset}\frac{n^{(s)}|I|\lambda^{(s)}}{n_0^{(s)} + |I|n^{(s)}} \le \min_{I\subseteq S, I\neq\emptyset}\frac{n^{(s)}\sum_{i\in I}\Lambda_i^{(s)}}{n_0^{(s)} + |I|n^{(s)}}. \tag{34}$$

Note that the function $f(x) = (n^{(s)}xc)/(n_0^{(s)} + xn^{(s)})$ is a non-decreasing function if $c > 0$ and non-increasing if $c < 0$. So in (34)

$$\min_{I\subseteq S, I\neq\emptyset}\frac{n^{(s)}|I|\lambda^{(s)}}{n_0^{(s)} + |I|n^{(s)}}\mathbf{1}_{\{\lambda^{(s)}<0\}} \ge \frac{sn^{(s)}\lambda^{(s)}}{n_0^{(s)} + sn^{(s)}}\mathbf{1}_{\{\lambda^{(s)}<0\}}.$$

Now using the observation that $\lambda^{(s)} > 0$, $\hat{\mu}_0^{(s)} = \bar{X}_{0\cdot}^{(s)}$ the inequality follows. □

We observe that from Lemma 2 it follows

$$\frac{\rho^{(s)}}{1+\rho^{(s)}}\lambda^{(s)}\mathbf{1}_{\{\lambda^{(s)}<0\}} \xrightarrow{\mathbf{P}} 0 \implies \bar{X}_{0\cdot}^{(s)} - \hat{\mu}_0^{(s)} \xrightarrow{\mathbf{P}} 0\,. \tag{35}$$

From the above lemma following result follows which we present as a stand alone fact. Note that in [2, Theorem 2.5] Chaudhuri and Perlman proved an weaker version using a different technique.

Proposition 1. *Suppose* $n^{(s)}/\log s \longrightarrow \infty$ *as* $s \to \infty$ *then* $\hat{\mu}_0^{(s)} \xrightarrow{\mathbf{P}} \mu_0$ *if and only if* $n_0^{(s)} \to \infty$.

Proof: Using similar argument which leads to equation (24) we conclude that if $n^{(s)}/logs \longrightarrow \infty$ then $\frac{\rho^{(s)}}{1+\rho^{(s)}}\lambda^{(s)}\mathbf{1}_{\{\lambda^{(s)}<0\}} \xrightarrow{\mathbf{P}} 0$ holds and hence from (35) we get

$$\bar{X}_{0\cdot}^{(s)} - \hat{\mu}_0^{(s)} \xrightarrow{\mathbf{P}} 0\,.$$

The result follows from the fact that $\bar{X}_{0\cdot}^{(s)} \xrightarrow{\mathbf{P}} \mu_0$ if and only if $n_0^{(s)} \to \infty$. □

Acknowledgement

The first author would like to thank the Department of Statistics and Applied Probability of National University of Singapore, for their kind hospitality.

References

[1] Sanjay Chaudhuri and Michael D. Perlman. On the bias and mean-square error of order-restricted maximum likelihood estimators. *J. Statist. Plann. Inference*, 130(1-2):229–250, 2005.

[2] Sanjay Chaudhuri and Michael D. Perlman. Consistent estimation of the minimum normal mean under the tree-order restriction. *J. Statist. Plann. Inference*, 137(11):3317–3335, 2007.

[3] Sanjay Chaudhuri and Michael D. Perlman. Biases of the maximum likelihood and Cohen-Sackrowitz estimators for the tree-order model. *Statist. Probab. Lett.*, 71(3):267–276, 2005.

[4] Arthur Cohen and H. B. Sackrowitz. Inference for the model of several treatments and a control. *J. Statist. Plann. Inference*, 107(1-2):89–101, 2002. ISSN 0378-3758. Statistical inference under inequality constraints.

[5] B. V. Gnedenko and A. Ya. Hinčin. *Èlementarnoe vvedenie v teoriyu veroyatnosteĭ*. Gosudarstv. Izdat. Tehn.-Teor. Lit., Moscow, Leningrad, 1952. 3d ed.

[6] J. T. Gene Hwang and Shyamal Das Peddada. Confidence interval estimation subject to order restrictions. *Ann. Statist.*, 22(1):67–93, 1994. ISSN 0090-5364.

[7] C.I.C. Lee. Quadratic loss of order restricted estimators for treatment means with a control. *The Annals of Statistics*, 16(2):751–758, 1998.

[8] J. Neyman and Elizabeth L. Scott. Consistent estimates based on partially consistent observations. *Econometrica*, 16:1–32, 1948. ISSN 0012-9682.

[9] Tim Robertson, F. T. Wright, and R. L. Dykstra. *Order restricted statistical inference*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Ltd., Chichester, 1988. ISBN 0-471-91787-7.

[10] Mervyn J. Silvapulle and Pranab K. Sen. *Constrained statistical inference*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2005. ISBN 0-471-20827-2. Inequality, order, and shape restrictions.