Maximum Likelihood Estimation of Isotonic Normal Means with Unknown Variances*

Ning-Zhong Shi and Hua Jiang

Northeast Normal University, Changchun, China

Received November 5, 1996; revised June 23, 1997

To analyze the isotonic regression problem for normal means, it is usual to assume that all variances are known or unknown but equal. This paper then studies this problem in the case that there are no conditions imposed on the variances. Suppose that we have data drawn from $k$ independent normal populations with unknown means $\mu_i$'s and unknown variances $\sigma_i^2$'s, in which the means are restricted by a given partial ordering. This paper discusses some properties of the maximum likelihood estimates of $\mu_i$'s and $\sigma_i^2$'s under the restriction and proposes an algorithm for obtaining the estimates.

AMS subject classifications: 62H12, 62F10.

Key words and phrases: isotonic regression, partial order, restricted maximum likelihood estimation.

1. INTRODUCTION

For $k$ independent normal populations with unknown means $\mu_i$ and unknown variances $\sigma_i^2$, $i = 1, \ldots, k$, this paper studies the maximum likelihood estimation (MLE) of $\mu = (\mu_1, \ldots, \mu_k)$ and $\sigma^2 = (\sigma_1^2, \ldots, \sigma_k^2)$ subject to the condition that the means $\mu_i$'s are restricted by a given partial ordering. Many interesting partial orderings may be considered: (i) a simple order on the means $\mu_1 \leq \cdots \leq \mu_k$; (ii) in the study of dose-response relationships the means exhibit an unimodal trend $\mu_1 \leq \cdots \leq \mu_{h} \geq \cdots \geq \mu_k$, which is called the umbrella order and includes the simple order with $h = k$; (iii) a simple tree order on the means is of the form $\mu_1 \leq \mu_i$ for $i = 2, \ldots, k$, which implies that some experiments are designed such that several treatments are significantly more effective than a control.

* Project supported by the National Natural Science Foundation of China.

183
Let $x_{ij}$, $j = 1, ..., n_i$, be observations from the $i$th normal population. The log-likelihood function is given by

\[ l(\mu, \sigma^2) = \sum_{i=1}^{k} \left\{ -\frac{n_i}{2} \ln \sigma_i^2 - \frac{1}{2\sigma_i^2} \sum_{j=1}^{n_i} (x_{ij} - \mu)^2 \right\} + c, \]  

(1.1)

where $c$ is a constant which does not depend on the parameters.

By $\leq$ denote a partial order defined on a finite set $\Theta = (\theta_1, ..., \theta_k)$. A $k$-dimensional vector $\mu$ is said to be an isotonic function if $\theta_i, \theta_j \in \Theta, \theta_i \leq \theta_j$ implies $\mu_i \leq \mu_j$. By $D$ denote the set of all isotonic functions. The MLE of $(\mu, \sigma^2)$ is the maximum solution of (1.1) for $\mu \in D$ and $\sigma^2 \in R^+_k$. If all variances are known or unknown but equal, the MLE of $\mu$ subject to the order restriction is the maximum solution of (1.1) for $\mu \in D$ and equivalently is the solution of

\[ \min_{\mu \in D} \sum_{i=1}^{k} (x_{i} - \mu_i)^2 w_i \]  

(1.2)

for $\mu \in D$, where $x_i = \sum_j x_{ij}/n_i$ and $w_i = n_i/\sigma_i^2$ when variances are known; $w_i = n_i$ when all variances are unknown but equal. The solution now is called the isotonic regression of $(x, w)$ with $x = (x_1, ..., x_k)$ and $w = (w_1, ..., w_k)$. There are a number of elegant algorithms for obtaining the isotonic regression, see for example, Barlow et al. (1972) and Robertson et al. (1988).

In the study of the isotonic regression problem, the assumption about variances must be needed. In practice, sometimes we cannot obtain much information about the variances and this paper deals with the problem in the case that there are no conditions imposed on the variances. Shi (1994) considered a similar problem of estimating the MLE of $(\mu, \sigma^2)$, in which variances are also assumed to be restricted by a given partial order. The paper proposed an algorithm to compute the MLE and showed the convergence of the algorithm under the following

Condition A. For $i = 1, ..., k$, $\sigma_i^2 > (b - a)^2$, where $\sigma_i^2$ denotes the sample variance of the $i$th normal populations, and $a$ and $b$ denote the minimal and maximal sample means respectively.

Section 3 of this paper proposes an algorithm of obtaining the MLE of $(\mu, \sigma^2)$ for our problem and shows that the convergence of the algorithm does not need any imposed conditions. However, we do not know if the algorithm converges to the true MLE and hence a condition as the Condition A is also necessary to show that the algorithm converges to the MLE, which is discussed in Section 5. A numerical example using the algorithm is given in Section 4. It is known that the MLE is not unique for
our problem and it is interesting to study some properties of the MLE, which is given in the next section.

2. EXISTENCE OF THE MLE

In this section, for convenience, we assume that the normal means are restricted by the simple order, that is, $\mu_1 \leq \cdots \leq \mu_k$. Note that similar results given in this section may be obtained for any partial order restrictions.

Let $(\hat{\mu}, \hat{\sigma}^2)$ be the MLE of $(\mu, \sigma^2)$ subject to the order restriction. Then $\hat{\mu} \in D$ and $\hat{\sigma}^2 \in R^+_k$, which satisfy

$$l(\hat{\mu}, \hat{\sigma}^2) = \sup \{ l(\mu, \sigma^2); \mu \in D, \sigma^2 \in R^+_k \}, \quad (2.1)$$

where $l(\mu, \sigma^2)$ is given in (1.1). For any fixed $\sigma^2 \in R^+_k$, by the discussion in Section 1, the solution of $\sup \{ l(\mu, \sigma^2); \mu \in D \}$ must be the isotonic regression of $(\bar{x}, w)$. From Theorem 1.6 of Barlow et al. (1972) we have

$$a \leq \hat{\mu}_i \leq b \quad (2.2)$$

for $i = 1, \ldots, k$, where $a$ and $b$ are defined in Condition A.

On the other hand, it is easy to check that $\sup \{ l(\mu, \sigma^2); \mu \in D \}$ is the isotonic regression of $(\bar{x}, w)$. From Theorem 1.6 of Barlow et al. (1972) we have

$$L(\mu) = \sum_{i=1}^k -n_i \ln \left[ \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2 \right], \quad (2.4)$$

where the $\bar{x}_{ij}$'s are defined in Condition A. Thus $\hat{\mu}$ is the solution of (2.3) if and only if $\hat{\mu} \in D_0$ and satisfies

$$L(\hat{\mu}) = \sup \{ L(\mu); \mu \in D_0 \}. \quad (2.5)$$

Since $L(\mu)$ is a continuous function of $\mu$ and $D_0$ is a compact set, the solution of (2.5) and then of (2.1) exists. It means that the MLE of $(\mu, \sigma^2)$ under the order restriction exists.
As $D$ is a polyhedral convex cone, for any given $\mu \in D$, there uniquely exists a subscript set $\{i_1, \ldots, i_t\}$ with $1 \leq i_1 < \cdots < i_t < k$ such that $\mu$ may be written as

$$\mu_1 = \cdots = \mu_{i_1} < \mu_{i_1+1} = \cdots = \mu_{i_t} < \cdots < \mu_{i_t+1} = \cdots = \mu_k. \quad (2.6)$$

**Definition.** A vector $\mu \in D$ is said to be a favorable point if there is a subscript set $\{i_1, \ldots, i_t\}$ such that $\mu$ satisfies (2.6) and

$$\sum_{i=i_1}^{i_t} (\bar{x}_i - \mu_i) w_i(\mu) = 0 \quad (2.7)$$

for $s = 0, 1, \ldots, t$, where $w_i(\mu) = n_i/\sigma_i^2(\mu)$, $i_0 = 0$, and $i_{t+1} = k$.

**Theorem 2.1.** If $\hat{\mu}$ is the solution of (2.5), there is a subscript set $\{i_1, \ldots, i_t\}$ such that $\hat{\mu}$ satisfies (2.6) and (2.7), namely it is a favorable point.

The above theorem shows that the MLE of $\mu$ must be a favorable point and its proof is given in the Appendix. Because $L(\mu)$ is not a concave function, in general, the solution of (2.5) will not be unique. However, we have the following result.

**Theorem 2.2.** There are finitely many favorable points.

**Proof.** For a fixed subscript set, the components $\mu_i$ of a vector $\mu$ will be a constant for $i \in \{i_1 + 1, \ldots, i_t\}$, $s = 0, 1, \ldots, t$, if the vector satisfies (2.6). If $\theta_{i+1}$ denotes the constant, then (2.7) may be written as

$$\sum_{i=i_1}^{i_t} n_i (\bar{x}_i - \theta_{i+1}) g_i(\theta_{i+1}) = 0, \quad (2.8)$$

where $g_i(\theta_{i+1}) = 1/\sum (x_{i+s} - \theta_{i+1})^2$ and $s = 0, 1, \ldots, t$.

Since the left-hand side of (2.8) is a polynomial of $\theta_{i+1}$, the number of the solution of (2.8) is finite. As there are finitely many subscript sets, the theorem follows.

**Condition B.** For $i = 1, \ldots, k$, $\sigma_i^2 > \max\{(\bar{x}_i - a)^2, (\bar{x}_i - b)^2\}$, where $\sigma_i^2$, $a$, and $b$ are as defined in Condition A.

**Theorem 2.3.** If Condition B holds, the favorable point uniquely exists.

**Proof.** Condition B implies

$$\frac{\partial^2 L(\mu)}{\partial \mu_i^2} = -\sigma_i^2 + (\bar{x}_i - \mu_i)^2 < 0$$

186 SHI AND JIANG
for all \( \mu \) satisfying \( a \leq \mu_i \leq b \) and \( i = 1, \ldots, k \). Then \( L(\mu) \) is a concave function on \([a, b]\). By (2.2), the solution of (2.5) uniquely exists and, by Theorem 2.1, the favorable point uniquely exists.

Let \((\hat{\mu}, \hat{\sigma}^2)\) be a MLE subject to the given order restriction. By Theorem 2.1, \( \hat{\mu} \) must be a favorable point and, from (2.3), \( \hat{\sigma}^2 = \gamma_i(\hat{\mu}) \) for \( i = 1, \ldots, k \). Then Theorem 2.2 implies that the number of the MLE is finite. If Condition B holds, by Theorem 2.3 the MLE is unique.

3. THE PROPOSED ALGORITHM

For a given subscript set, one can obtain all favorable points according to the subscript set using (2.6) and (2.8), in which all real roots of some polynomials need to be found. Then the MLE of \((\mu, \sigma^2)\) may be obtained by comparing all favorable points for all subscript sets. However, this procedure is very hard to carry out. This section proposes an iteration algorithm.

As in the discussion in Section 1, if the variance vector \( \sigma^2 \) is given the MLE of \( \mu \) may be obtained as in the isotonic regression problem, and if the mean vector \( \mu \) is given the MLE of \( \sigma^2 \) is just \( \sigma^2(\mu) \) as defined in (2.3). The following algorithm is based on this consideration.

**Algorithm.**

Step(0, 0). Let \( \mu^{(0)} = \bar{\mu} \) and \( \sigma^{(0)} = \bar{\sigma}^2 \).

Step(n, 1). Find \( \mu^{(n)} \), the isotonic regression of \((\bar{\mu}, w^{(n-1)})\) for \( w^{(n-1)} = n_i \bar{\sigma}^{(n-1)} \).

Step(n, 2). Let \( \sigma^{(n)} = \gamma_i(\mu^{(n)}) \) and \( \sigma^{(n)} = (\sigma^{(n)}_1, \ldots, \sigma^{(n)}_k) \).

The above algorithm shows that

\[
L(\mu^{(n)}, \sigma^{(n-1)}) \leq L(\mu^{(n)}, \sigma^{(n)}) \leq L(\mu^{(n+1)}, \sigma^{(n)}) \tag{3.1}
\]

and \( L(\mu^{(n)}) \leq L(\mu^{(n+1)}) \) for \( n \geq 1 \). By the monotonicity, \( L(\mu^{(p)}) = L(\mu^{(p+q)}) \) for all integers \( q \geq 1 \) if \( L(\mu^{(p)}) = L(\mu^{(p+1)}) \). Thus we can give a termination criterion for the algorithm. For example, we stop the iteration at Step(n,2) if

\[
\max_{1 \leq r \leq k} |\mu^{(n)}_r - \mu^{(n)}_r| \leq 10^{-m}
\]

for some integers \( m \geq 1 \). The proof of the following theorem is given in the Appendix.
Theorem 3.1. The point sequence \( \{\mu^{(n)}\} \) given in the above algorithm converges to a favorable point as \( n \to \infty \).

Corollary 3.1 If Condition B holds and the limiting point of \( \{\mu^{(n)}\} \) is \( \mu^* \), then the MLE of \( (\mu, \sigma^2) \) is \( (\mu^*, \sigma^2(\mu^*)) \).

The above corollary may be shown by using Theorem 2.3. It must be noted that in the above corollary Condition B is not always needed. For example, if the sample means satisfy \( \bar{x}_1 \leq \cdots \leq \bar{x}_k \), the MLE will be \( (\mu^{(1)}, \sigma^{(1)}) \), the one step result of the algorithm. However, in general, the condition cannot be omitted. A detailed discussion is given in Section 5.

4. NUMERICAL EXAMPLE

For illustration, the proposed algorithm is used to treat the data shown in Shi (1994). There are five districts in Jilin Province of China: Liao yuan (Group 1), Qianfu (Group 2), Chang chu (Group 3), Tonghua (Group 4), and Jilin (Group 5). The data gave the scores of 100 students per district obtained in the National Matriculation Examination held in 1992. Past experience showed that the conditions of education of the district \( i + 1 \) was likely better than district \( i \) for \( i = 1, \ldots, 4 \). We proceed to estimate the examination scores of students of the five districts. By \( X_i \) we denote the examination score of district \( i \), then \( X_i \) follows a normal distribution with unknown mean \( \mu_i \) and unknown variance \( \sigma^2_i \) for \( i = 1, \ldots, 5 \). Prior information tells us that the means exhibit an increasing trend \( \mu_1 \leq \cdots \leq \mu_5 \). We use the proposed algorithm to estimate the MLE of \( \mu_i \)’s and \( \sigma^2_i \)’s subject to the simple order restriction.

It is easy to check that Condition B is satisfied, and by Corollary 3.1 we can obtain the MLE by the proposed algorithm. The computed results of estimating the means and the variances are listed in Table I, in which Cran’s (1980) program was used as a subroutine to compute isotonic

<table>
<thead>
<tr>
<th>TABLE I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computing Results of Examination Scores</td>
</tr>
<tr>
<td>( i = 1 )</td>
</tr>
<tr>
<td>( \mu(0) )</td>
</tr>
<tr>
<td>( \sigma(0) )</td>
</tr>
<tr>
<td>( \mu(1) )</td>
</tr>
<tr>
<td>( \sigma(1) )</td>
</tr>
<tr>
<td>( \mu(2) )</td>
</tr>
<tr>
<td>( \sigma(2) )</td>
</tr>
</tbody>
</table>
regression and the iteration is terminated when \( \max_i |\mu_i^{(n-1)} - \mu_i^{(n)}| \leq 10^{-3} \).

The computed results show that the iteration is terminated at \( n = 2 \). In the

table, for the terminated criterion \( 10^{-3} \), \( \mu^{(1)} = \mu^{(2)} \) means that the MLE of

\( \mu \) is the same as the isotonic regression of \( (x, \bar{w}) \), with \( \bar{w}_i = n_i/\hat{\sigma}_i^2 \). The MLE

of \( \sigma^2 \) is different from the sample variance, that is, \( \sigma^{(0)} \neq \sigma^{(2)} \).

5. DISCUSSION

Recall that our problem is to find the solution \( (\hat{\mu}, \hat{\sigma}^2) \) which

maximizes \( l(\mu, \sigma^2) \) subject to \( \mu \in D \) and \( \sigma^2 \in R^k_+ \),

(5.1)

where \( l(\mu, \sigma^2) \) is given in (1.1). For estimating \( \mu \), one may replace \( \sigma^2 \) by \( \hat{\sigma}^2 \)
in (1.2) to obtain the isotonic regression \( \mu' \), say, of \( (\bar{x}, \bar{w}) \) with \( \bar{w}_i = n_i/\hat{\sigma}_i^2 \) for

\( i = 1, \ldots, k \). The estimate of \( \sigma^2 \) will be \( \hat{\sigma}^2 \) or, by the likelihood principle,

\( \hat{\sigma}^2(\mu') \) as defined in (2.3). However, the estimates \( (\mu', \hat{\sigma}^2) \) and \( (\mu', \sigma^2(\mu')) \)

are the results of Step (1,1) and Step (1,2) in the proposed algorithm

respectively. Furthermore, the expression of \( L(\mu) \) in (2.4) may be written

as

\[
L(\mu) = \sum_{i=1}^k - n_i \ln \left[ 1 + \frac{(\bar{x}_i - \mu_i)^2}{\hat{\sigma}_i^2} \right] + c,
\]

where \( c \) is a constant which does not depend on the parameters. If

Condition B holds, by Taylor expansion,

\[
L(\mu) = \sum_{i=1}^k \sum_{j=1}^k \frac{(-1)^j}{j} \left[ (\bar{x}_i - \mu_i)^2 \bar{w}_i \right]^j.
\]

Then the estimate \( \mu' \), the isotonic regression of \( (\bar{x}, \bar{w}) \), is the first

approximation, \( j = 1 \), for the solution of (5.1).

On the other hand, by (2.3), a reasonable estimate of \( (\mu, \sigma^2) \) is of the

form \( (\hat{\mu}, \sigma^2(\hat{\mu})) \) and satisfies

\[
\sum_{i=1}^k E(\bar{x}_i - \mu_i)^2 w_i(\bar{x}) > \sum_{i=1}^k E(\bar{v}_i - \mu_i)^2 w_i(\bar{v})
\]

(5.2)

for any \( \mu \in D \), where \( w_i(\cdot) = n_i/\hat{\sigma}_i^2(\cdot) \) and \( i = 1, \ldots, k \); see Brunk (1965), Lee

(1981, 1988) and Hwang and Peddada (1994). The Eq. (5.2) implies that the mean square error of the estimate \( (\hat{\mu}, \sigma^2(\hat{\mu})) \) is strictly less than that of the

usual estimate. Let \( \mu^* \) be the limiting point of the proposed algorithm.

For every step \( n > 1 \) of the algorithm, by (3.1) we have
$$\sum_{i=1}^{k} E(\bar{x}_i - \mu_i)^2 w_i(\bar{x}) > \sum_{i=1}^{k} E(\mu_i^{(1)} - \mu_i)^2 w_i(\bar{x})$$

$$\geq \sum_{i=1}^{k} E(\mu_i^{(n)} - \mu_i)^2 w_i(\mu_i^{(n-1)})$$

and then the estimate ($\hat{\mu}^*, \sigma^2(\hat{\mu}^*)$) satisfies (5.2) even if it is not the MLE. Note that $\mu_i^{(1)} = \mu'$ in the second port of the above expression.

It will be very important to know the ratio at which Condition B is satisfied for some regular cases and to know how many favorable points, from the proposed algorithm, are the MLE even if Condition B is not satisfied. Therefore, some simulation studies are needed.

Some simulation results are listed in Table II, in which $k = 5$ and 7 respectively. The normal means are considered in two cases: (1) equal means $\mu_1 = \cdots = \mu_k$; (2) equal spacing means $\mu_{i+1} - \mu_i = A$, where

TABLE II

<table>
<thead>
<tr>
<th>Mean</th>
<th>Variance</th>
<th>NUNB</th>
<th>NUNC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$n=10$</td>
<td>$n=15$</td>
</tr>
<tr>
<td>$k = 5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>1.0 1.0 1.0 1.0 1.0</td>
<td>4290 1777</td>
<td>735</td>
</tr>
<tr>
<td>0.6 0.6 1.0 1.0 1.0</td>
<td>1.0 1.0 1.0 1.0 1.0</td>
<td>5747 3334</td>
<td>1674</td>
</tr>
<tr>
<td>0.6 1.0 1.0 0.1 0.6</td>
<td>1.0 1.0 1.0 0.6 0.6</td>
<td>3769 3221</td>
<td>1677</td>
</tr>
<tr>
<td>$-2 -2 0.0 0.1 0.2$</td>
<td>$1.0 1.0 1.0 1.0 1.0$</td>
<td>$5662 3291$</td>
<td>1806</td>
</tr>
<tr>
<td>$-2 -2 1.0 0.1 0.2$</td>
<td>$1.0 1.0 1.0 1.0 1.0$</td>
<td>$5272 3017$</td>
<td>1777</td>
</tr>
<tr>
<td>$0.6 0.6 0.0 0.1 0.2$</td>
<td>$1.0 1.0 1.0 0.6 0.6$</td>
<td>$6977 5410$</td>
<td>4367</td>
</tr>
<tr>
<td>$0.6 1.0 1.0 0.1 0.2$</td>
<td>$1.0 1.0 1.0 0.1 0.6$</td>
<td>$7280 5717$</td>
<td>4724</td>
</tr>
<tr>
<td>$1.0 1.0 1.0 0.6 0.6$</td>
<td>$1.0 1.0 1.0 0.6 0.6$</td>
<td>$7024 5514$</td>
<td>4335</td>
</tr>
<tr>
<td>$k = 7$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0 0.0 0.0 0.0 0.0 0.0</td>
<td>1.0 1.0 1.0 1.0 1.0</td>
<td>6270 3028</td>
<td>1272</td>
</tr>
<tr>
<td>0.8 0.8 0.8 0.8 1.0 1.0</td>
<td>1.0 1.0 1.0 1.0 1.0</td>
<td>6512 3346</td>
<td>1573</td>
</tr>
<tr>
<td>0.8 0.8 1.0 1.0 0.8 0.8</td>
<td>1.0 1.0 1.0 0.8 0.8</td>
<td>6543 3388</td>
<td>1582</td>
</tr>
<tr>
<td>1.0 1.0 0.8 0.8 0.8 0.8</td>
<td>1.0 1.0 0.8 0.8 0.8</td>
<td>6508 3378</td>
<td>1571</td>
</tr>
<tr>
<td>$-3 -2 -2 0.0 0.1 0.2 0.3$</td>
<td>$1.0 1.0 1.0 1.0 1.0$</td>
<td>$7916 5923$</td>
<td>4356</td>
</tr>
<tr>
<td>$0.8 0.8 0.8 0.8 1.0 1.0$</td>
<td>$1.0 1.0 1.0 1.0 1.0$</td>
<td>$8376 6912$</td>
<td>5704</td>
</tr>
<tr>
<td>$0.8 0.8 1.0 1.0 0.8 0.8$</td>
<td>$1.0 1.0 1.0 0.8 0.8$</td>
<td>$8527 7195$</td>
<td>6076</td>
</tr>
<tr>
<td>$1.0 1.0 0.8 0.8 0.8 0.8$</td>
<td>$1.0 1.0 0.8 0.8 0.8$</td>
<td>$8316 6920$</td>
<td>5661</td>
</tr>
</tbody>
</table>

Note: The simulations were run 10000 times. In the tables NUNB denotes the number of times that Condition B is not satisfied and NUNC denotes the number of times that the algorithm does not converge to true MLE.
i = 1, ..., k − 1 and d = 0.1. The variances are considered in four cases, equal, increasing, unimodal, and decreasing, as listed in the table. We take the simulations for equal sample sizes \( n_i = 10, \ n_i = 15, \) and \( n_i = 20 \) respectively. The simulations are run 10000 times for each case. In the table, NUNB denotes the number of times that Condition B is not satisfied and NUNC denotes the number of times that the favorable point obtained by the proposed algorithm is not the MLE.

The NUNB depends on the relationship of sizes of means and variances, and also depends on the sample size. NUNC = 0, almost everywhere, tells us that the favorable point obtained by the proposed algorithm is the MLE even if Condition B is not satisfied for the considered regular cases.

However, we can find some abnormal cases such that the NUNC does not equal zero, as listed in Table III. In the table, for \( k = 5 \), the means are 1.2, 0.9, 0.6, 0.3, and 0.0, and it is antitonic with respect to the simple order; the variances are 1.0, 0.8, 0.6, 0.4, and 0.2 respectively. The simulations are run 10000 times for equal sample size \( n_i = 20 \). It may be seen that Condition B is not satisfied for any of the cases. The favorable point from the proposed algorithm is not the MLE for 32 cases. For these cases, let \( \hat{\mu} \) denote the true MLE, \( \mu^* \) denote the favorable point obtained from the proposed algorithm, and \( \mu' \) denote the isotonic regression of

\[ \text{TABLE III} \]

The Simulation Results for Abnormal Cases: \( k = 5, \ n = 20, \) NUNB = 10000, NUNC = 32

\[
\begin{array}{ccccccccccc}
  L(\hat{\mu}) & L(u^*) & L(\hat{\mu}) - L(u^*) & L(\hat{\mu}) - L(u^*) & L(\hat{\mu}) - L(u^*) & L(\hat{\mu}) - L(u^*) & L(\hat{\mu}) - L(u^*) & L(\hat{\mu}) - L(u^*) \\
  80.272 & 78.415 & 77.802 & 1.856 & 2.470 & 70.736 & 70.436 & 70.436 & 0.299 & 0.300 \\
  81.901 & 81.846 & 81.729 & 0.054 & 0.171 & 89.989 & 89.971 & 89.971 & 0.018 & 0.092 \\
  74.776 & 73.073 & 72.314 & 1.703 & 2.462 & 66.467 & 66.121 & 66.120 & 0.345 & 0.347 \\
  75.906 & 75.893 & 75.786 & 0.013 & 0.121 & 88.378 & 84.953 & 83.559 & 3.425 & 4.819 \\
  83.899 & 82.157 & 80.929 & 1.741 & 2.969 & 78.913 & 78.522 & 78.509 & 0.391 & 0.404 \\
  67.869 & 67.727 & 67.727 & 0.142 & 0.142 & 71.949 & 68.645 & 66.986 & 3.304 & 4.963 \\
  64.990 & 64.432 & 64.346 & 0.558 & 0.644 & 80.794 & 78.160 & 78.147 & 2.634 & 2.647 \\
  86.406 & 86.372 & 85.312 & 0.034 & 1.094 & 91.185 & 88.650 & 88.489 & 2.535 & 2.697 \\
  66.055 & 65.925 & 65.603 & 0.160 & 0.483 & 93.171 & 92.671 & 91.568 & 0.500 & 1.603 \\
  79.911 & 79.665 & 79.649 & 0.246 & 0.262 & 65.661 & 65.529 & 64.526 & 0.132 & 1.136 \\
  87.382 & 87.362 & 87.199 & 0.020 & 0.183 & 64.521 & 60.697 & 60.204 & 3.824 & 4.317 \\
  83.770 & 83.019 & 82.313 & 0.751 & 1.457 & 86.497 & 85.284 & 85.277 & 1.213 & 1.220 \\
  67.143 & 66.080 & 66.080 & 1.064 & 1.064 & 77.468 & 77.107 & 77.107 & 0.361 & 0.361 \\
  83.109 & 83.014 & 83.002 & 0.094 & 0.106 & 69.079 & 69.065 & 68.048 & 0.013 & 1.031 \\
  49.508 & 48.187 & 47.437 & 1.321 & 2.070 & 91.336 & 90.629 & 90.628 & 0.708 & 0.709 \\
  80.666 & 77.191 & 73.371 & 3.475 & 7.295 & 96.127 & 96.115 & 96.097 & 0.011 & 0.030 \\
\end{array}
\]

\[\text{Note.} \quad \text{The simulations were run 10000 times. The values of the log-likelihood function are listed, in which } \hat{\mu} \text{ denotes the true MLE, } u^* \text{ denotes the estimate from the algorithm, and } u' \text{ denotes the isotonic regression using the sample variances.}\]
(x, w) with the estimate of variances $\sigma^2(\mu')$, the result of Step (1,2) in the algorithm. The values of $L(\hat{\mu}), L(\mu^*)$, and $L(\mu')$ are listed in Table III. The differences $L(\hat{\mu}) - L(\mu^*)$ and $L(\hat{\mu}) - L(\mu')$ are also listed in Table III. The differences, $L(\hat{\mu}) - L(\mu^*)$, are very small, corresponding to the values of $L(\hat{\mu})$ and $L(\mu^*)$ for all cases. The value of $L(\mu^*)$ is greater than that of $L(\mu')$.

Consequently, we recommend the iteration algorithm for practical use because its computation is simple and it has good properties of convergence.

APPENDIX

The Proof of Theorem 2.1. We need to find a vector $\hat{\mu}$ which belongs to $D$ and maximizes $L(\mu)$, given in (2.4), subject to $\mu \in D$. The Lagrangian function now is given by

$$\Phi(\mu, \lambda) = \frac{1}{2} L(\mu) + \sum_{i=1}^{k-1} \lambda_i(\mu_{i+1} - \mu_i),$$

where $\lambda = (\lambda_1, ..., \lambda_{k-1})$ and $\lambda_i$'s are the Lagrangian multipliers.

The Kuhn–Tucker conditions are usually used to deal with such problems. If $\hat{\mu}$ maximizes $L(\mu)$ subject to $\mu \in D$, then $\hat{\mu}$ satisfies the following conditions:

1. $\hat{\mu}_1 \leq \cdots \leq \hat{\mu}_k$;
2. $\nabla \Phi_i(\hat{\mu}) = 0$ for $i = 1, ..., k$;
3. $\lambda_i \geq 0$ for $i = 1, ..., k - 1$;
4. $\lambda_i(\hat{\mu}_{i+1} - \hat{\mu}_i) = 0$ for $i = 1, ..., k - 1$.

The second of these conditions corresponds to the equation

$$w_i(\hat{\mu}) + \hat{\mu}_{i-1} - \hat{\mu}_i = 0 \quad (A1)$$

for $i = 1, ..., k$, where $w_i(\hat{\mu})$ is given in (2.7) and $\hat{\lambda}_0 = \hat{\lambda}_k = 0$. Let $\{i_1, ..., i_t\}$ be the subscript set such that $\lambda_{i_j} = 0, j = 1, ..., t$; $\lambda_{j} > 0$, otherwise. Then the first, third and fourth conditions given in the above imply that $\hat{\mu}$ satisfies (2.6) and, from (A1), we have (2.7).

The Kuhn–Tucker conditions are necessarily satisfied for our problem. Furthermore, if $L(\mu)$ is a concave function, by the discussion in Mangasarian (1969, p. 94) these are also sufficient conditions and there exists uniquely a favorable point; see also Theorem 2.3 in this paper.

To prove Theorem 3.1, we need the following lemmas.
**Lemma A.1.** Let \( \{y_n\} \) be a uniformly bounded sequence in \( \mathbb{R}^k \). If \( y_n - y_{n-1} \to 0_k \), as \( n \to \infty \), and the sequence is not convergent, then there are infinitely many accumulation point of the sequence, where \( 0_k \) denotes the \( k \)-dimensional zero vector.

**Proof.** At first, we consider the case of \( k = 1 \). Let \( \alpha = \lim \inf \{ y_n \} \) and \( \beta = \lim \sup \{ y_n \} \). By the condition, \( \alpha < \beta \). For any \( z \in (\alpha, \beta) \), we will show that \( z \) is an accumulation point of \( \{y_n\} \), namely, there is a subsequence of \( \{y_n\} \) which converges to \( z \).

For any \( \varepsilon_2 > 0 \), by the assumption there is a positive integer \( N_1 \) such that \( |y_{n+1} - y_n| < \varepsilon_2 \) if \( n > N_1 \). On the other hand, \( \alpha < z < \beta \) implies that there is a positive integer \( n_1 > N_1 \) such that \( y_{n_1} < z \) and \( y_{n_1+1} > z \). Then we have

\[
|y_{n_1} - z| < y_{n_1+1} - y_{n_1} < \varepsilon_2.
\]

Similarly, for any \( \varepsilon_2 > 0 \) with \( \varepsilon_2 < \varepsilon_1 \), there is a term \( y_{n_2} \) with \( n_2 > n_1 \) such that \( |y_{n_2} - z| < \varepsilon_2 \). If one continues this procedure, a subsequence \( \{y_{n_j}\} \) may be obtained and it converges to \( z \).

For \( k > 1 \), denote \( y_n = (y_{1n}, \ldots, y_{kn}) \). Let \( \alpha_i = \lim \inf \{ y_{in} \} \) and \( \beta_i = \lim \sup \{ y_{in} \} \), \( i = 1, \ldots, k \). Without loss of generality, assume that \( \alpha_1 < \beta_1 \). For any \( z_i \in (\alpha_i, \beta_i) \), from the above discussion there is a subsequence \( \{y_{in}\} \) of \( \{y_{in}\} \) with \( y_{in} \to z_i \), as \( n \to \infty \). As the sequence is uniformly bounded, for \( i = 2 \), there is a subsequence of \( \{y_{in}\} \) which converges to \( z_2 \) for some \( z_2 \) in the interval \( (\alpha_2, \beta_2) \). So we can obtain a subsequence of \( \{y_n\} \) and a point \( z \), in \( \mathbb{R}^k \), with the first component \( z_1 \) such that the subsequence converges to \( z \). Because there are infinitely many points in \( (\alpha_1, \beta_1) \), the proof of this theorem is completed.

**Lemma A.2.** Let \( \{\mu^{(n)}\} \) be the sequence from the proposed algorithm and let \( \{\mu^{(n)}\} \) be a subsequence. If the subsequence is convergent, then \( \mu^{(n)} - \mu^{(n-1)} \to 0_k \), as \( n \to \infty \).

**Proof.** As \( \sigma^{(n)} \) is a continuous function of \( \mu^{(n)}, \sigma^{(n)} \) is also convergent. Recalling the expression (1.1), \( (\mu^{(n)}, \sigma^{(n)}) \) is convergent. From \( n_{j-1} \leq n_j - 1 \) and (3.1), we have

\[
0 \geq h(\mu^{(n_{j-1})}, \sigma^{(n_{j-1})}) - h(\mu^{(n_j)}, \sigma^{(n_j)})
\]

\[
\geq h(\mu^{(n_{j-1})}, \sigma^{(n_{j-1})}) - h(\mu^{(n_j)}, \sigma^{(n_j)})
\]

\[
\to 0
\]

(A2)
as \( n_j \to \infty \). For simplifying the notation, denote \( n_j \) by \( m \). Since the difference of \( 2l(\mu^{(m-1)}, \sigma^{(m-1)}) \) and \( 2l(\mu^{(m)}, \sigma^{(m-1)}) \) may be written as

\[
\sum_{i=1}^{k} \left[ (\bar{x}_i - \mu_i^{(m)})^2 - (\bar{x}_i - \mu_i^{(m-1)})^2 \right] w_i^{(m-1)},
\]

where \( w_i^{(m-1)} \) is defined in the Algorithm, (A2) implies that the difference converges to 0 when \( m \) tends to infinity. Because \( \mu^{(m)} \) is the isotonic regression of \( (\bar{x}, w^{(m-1)}) \), following the discussion in Shi (1994, p. 291), we have \( \mu^{(m)} - \mu^{(m-1)} \to 0 \), as \( m \to \infty \).

**Lemma A.3.** Let \( \{\mu^{(m)}\} \) be the sequence from the proposed algorithm and \( \{\mu^{(m)}\} \) be a subsequence. If the subsequence is convergent, then it converges to a favorable point.

**Proof.** Assume that \( \mu^{(m)} \to v \), as \( m \to \infty \). Since \( v \) belongs to \( D \), there is a subscript set \( \{i_1, ..., i_t\} \) such that \( v \) satisfies (2.6). Now we show that \( v \) satisfies (2.7). Let

\[
\alpha_{s+1}^{(m)} = \min\{\mu_{i_s}^{(m)}, i_s + 1 \leq i \leq i_{s+1} \}
\]

and

\[
\beta_{s+1}^{(m)} = \max\{\mu_{i_s}^{(m)}, i_s + 1 \leq i \leq i_{s+1} \},
\]

where \( s = 0, 1, ..., t \), \( i_0 = 0 \), and \( i_{t+1} = k \). Because \( \mu^{(m)} \) converges to \( v \), for any \( \delta > 0 \), there is an integer \( M' \) such that

\[
\max\{\beta_{s+1}^{(m)} - \alpha_{s+1}^{(m)}, s = 0, 1, ..., t\} < \delta,
\]

(A3)

for \( m > M' \). From the proposed algorithm, \( \mu^{(m)} \) is the isotonic regression of \( (\bar{x}, w(\mu^{(m-1)})) \). Therefore for any \( \varepsilon > 0 \), by the lemma in Barlow et al. (1972, p. 34) and (A3), there is an integer \( M \) with \( M > M' \) such that

\[
\left| \sum_{i=i_s}^{i_{s+1}} (\bar{x}_i - \mu_i^{(m)}) w_i(\mu^{(m-1)}) \right| < \varepsilon
\]

(A4)

for \( m > M \) and \( s = 0, 1, ..., t \). As \( w(\mu) \) is a continuous function of \( \mu \), we have

\[
\sum_{i=i_s}^{i_{s+1}} (\bar{x}_i - v_i) w_i(v) = \lim_{m \to \infty} \sum_{i=i_s}^{i_{s+1}} (\bar{x}_i - \mu_i^{(m)}) w_i(\mu^{(m)})
\]

\[
= \lim_{m \to \infty} \sum_{i=i_s}^{i_{s+1}} (\bar{x}_i - \mu_i^{(m)}) w_i(\mu^{(m-1)})
\]

\[
= 0
\]
for \( s = 0, 1, \ldots, t \). The second equation of the above follows Lemma A2 and the last follows (A4). Then \( v \) satisfies (2.7) and is a favorable point.

**The Proof of Theorem 3.1.** Let \( l_{2n-1} = l(\mu^{(n)} , \sigma^{(n-1)}) \) and \( l_{2n} = l(\mu^{(n)} , \sigma^{(n)}) \), where \( n > 1 \) and \( \mu^{(n)} \) and \( \sigma^{(n)} \) are as given in the algorithm. From \( l(\mu^{(n)} , \sigma^{(n)}) \leq l(x, \sigma^2) \) and (3.1), the real number sequence \( \{l_m; m = 1, 2, \ldots\} \) is monotone increasing and bounded. Then the sequence is convergent and \( l_{2n} - l_{2n-1} \to 0 \), as \( n \to \infty \). By using a method similar to that shown in the proof of Lemma A2, we can prove that \( \mu^{(n)} - \mu^{(n-1)} \to 0 \), as \( n \to \infty \).

Since \( \{\mu^{(n)}\} \) is uniformly bounded, if it is not convergent, by Lemma A1, there are infinitely many accumulation points and, by Lemma A3, they are all favorable points. This contradicts Theorem 2.2. Therefore \( \{\mu^{(n)}\} \) is convergent and, by Lemma A3 again, it converges to a favorable point.

**ACKNOWLEDGMENTS**

We thank the referee for helpful comments. We rewrote Section 5 taking his comments into account. We are also grateful to the editors for their help and support.

**REFERENCES**