The Annals of Statistics
1994, Vol. 22, No. 1, 67–93

CONFIDENCE INTERVAL ESTIMATION SUBJECT TO ORDER RESTRICTIONS

BY J. T. GENE HWANG[1] AND SHYAMAL DAS PEDDADA[2]

Cornell University and University of Virginia

This article deals with the construction of confidence intervals when the components of the location parameter μ of the random variable $\mathbf{X}$, which is elliptically symmetrically distributed, are subject to order restrictions. Several domination results are proved by studying the derivative of the coverage probability of the confidence intervals centered at the improved point estimators. Consequently, we strengthen the previously known results regarding the simple ordering and obtain several new results for other general forms of order restrictions, including the simple tree ordering, the umbrella ordering, the simple and the double loop ordering and some combination of these. These domination results are obtained under the assumption that Σ is a diagonal matrix. When Σ is nondiagonal, some new intervals are introduced which dominate the standard intervals centered at the unrestricted maximum likelihood estimator for various types of order restrictions. Similar results are obtained for scale parameters as well. Contrary to the location problems, in case of the scale parameters satisfying the simple ordering we find that the restricted maximum likelihood estimator of the largest parameter fails to universally dominate the unrestricted maximum likelihood estimator. A similar negative result is noted for simple tree order restriction.

1. Introduction. When estimating the components of the parameter $\mu = (\mu_1, \mu_2, \dots, \mu_k)'$, often some additional information regarding the order of the parameters μ_i is available to the researcher. For instance, if μ_i is the average height of US children of age i, then it is reasonable to assume the *simple ordering*, that is,

$$\mu_1 \le \mu_2 \le \cdots \le \mu_k. \tag{1.1}$$

Like the simple ordering, the *simple tree ordering*, defined as

$$\mu_1 \le \mu_i, \quad \text{for all } i, \tag{1.2}$$

arises very naturally in many problems of practical interest. For example, suppose μ_1 is the average yield of a crop with no fertilizer added and μ_i is the

Received December 1990; revised April 1993.
[1]Formerly Jiunn T. Hwang. Partially supported by NSF Grant DMS-88-09016.
[2] Supported in part by ONR Grant N-00014-92-J-1009. This author is thankful to the University of Virginia for providing support for a leave to visit the Department of Mathematics, Cornell University, whose hospitality is also appreciated.
AMS 1991 *subject classifications.* 62F10, 62F25, 62F30.
Key words and phrases. Coverage probability, isotonic regression, location parameter, order restrictions, restricted maximum likelihood estimator, scale parameter, simple ordering, simple tree ordering, unrestricted maximum likelihood estimator.

average yield of the crop when the ith brand of fertilizer is added. Then it is reasonable to expect $\mu_i \geq \mu_1$, for all i, although one may have no information regarding the relative performance of the various brands. For other types of order restrictions one may refer to the book by Robertson, Wright and Dykstra (1988).

There exists a considerable amount of literature on the estimation of parameters under order restrictions using isotonic regression [cf. Robertson, Wright and Dykstra (1988)]. Suppose $\mathbf{X} = (X_1, X_2, \ldots, X_k)'$ is an observation vector whose mean vector is μ. Then $\hat{\mu}$ is said to be the isotonic regression estimator of μ if it minimizes $(\mathbf{X} - \mu)'\Omega(\mathbf{X} - \mu)$, subject to the order restriction on μ. Here Ω is a positive definite matrix. If $\mathbf{X}$ has an elliptically symmetric unimodal distribution with mean μ and covariance matrix proportional to Σ and if Ω is chosen to be Σ^{-1}, then the isotonic regression estimator is a restricted maximum likelihood estimator of μ.

When the components of μ are estimated simultaneously, the restricted maximum likelihood estimator is known to perform better than the unrestricted maximum likelihood estimator. However, it depends upon the type of order restriction when one is interested in estimating the individual components of μ. For instance, when $X_1, X_2, \ldots, X_k$ are independently normally distributed, then under simple order restriction the restricted maximum likelihood estimator dominates the unrestricted one [cf. Kelly (1989) and Lee (1981)]. However, in the case of simple tree ordering, as observed by Lee (1988), the restricted maximum likelihood estimator fails disastrously, especially when k is large. Recently, Kelly (1990) found some isotonic regression estimators that stochastically dominate the unrestricted maximum likelihood estimator under simple tree ordering when $k \to \infty$.

As seen in the book of Robertson, Wright and Dykstra (1988), much of the effort during the past two decades has been expended on the problem of testing hypotheses under ordered alternatives. It seems very little is known regarding an important companion problem, namely, the estimation of confidence intervals. For instance, there are only two pages of discussion on this topic in Robertson, Wright and Dykstra (1988), and the authors remark that "this is primarily due to the general intractability of these types of problems."

In this article, as a first step, we focus on the one-dimensional constant length confidence intervals centered at some improved point estimators, typically the isotonic regression estimators. When Σ is a diagonal matrix, we obtain in Section 2 some derivative formulas of the coverage probability of the confidence intervals centered at the improved estimators. These formulas are used to study the domination of the proposed confidence intervals over the standard confidence intervals. Here and below, standard confidence intervals refer to the interval centered at the unrestricted maximum likelihood estimator. In the process we strengthen and unify the previously known results of Lee (1981, 1988) and Kelly (1989) for the simple order and simple tree order. A new scheme is introduced to construct better confidence intervals for general order restriction problems. The proposed scheme, unlike the intervals centered at the restricted maximum likelihood estimator, is simple to implement and can be used for any general order restriction. It is proved, under some sufficient conditions, that some of the confidence intervals centered at the proposed estimator are superior to the standard ones. Extensive simulation studies suggest that the proposed confidence intervals are significantly better than the standard ones. Section 3 deals with the case when Σ is nondiagonal. Some of the domination results are carried over to Section 3 by considering a simple alternative to the restricted maximum likelihood estimator. Using the general scheme introduced in Section 2, we develop better confidence intervals for general order restriction problems. Here again, the analytic results and the simulation studies establish that the proposed intervals are substantially better than the standard ones.

In Section 4 we discuss the problem of estimating the scale parameters under various types of order restrictions and some partial domination results are obtained. It is shown that the restricted maximum likelihood estimator of the smallest parameter in a simple order stochastically dominates the standard unrestricted maximum likelihood estimator. Surprisingly, however, the restricted maximum likelihood estimator of the largest parameter fails to dominate the standard one. We also note that the restricted maximum likelihood estimator of μ_i in a simple tree order fails to universally dominate X_i.

Conclusion of this article along with some open research problems in this area are given in Section 5.

2. The diagonal case. For a given order restriction on μ, the isotonic regression estimator minimizes $(\mathbf{X}-\mu)'\Omega(\mathbf{X}-\mu)$ under the order restriction. In this section we assume Ω to be a diagonal matrix with diagonal elements $W_1, W_2, \dots, W_k$ and $\mathbf{W} = (W_1, W_2, \dots, W_k)'$. Then the isotonic regression estimator can be expressed in a nice form using the minimax formula given on page 23 of Robertson, Wright and Dykstra (1988). To describe the isotonic regression estimator, we let "$\preceq$" denote the pair relation between i and j, both in $\mathcal{I} = \{1, 2, \dots, k\}$, such that $i \preceq j$ if and only if it is known that $\mu_i \le \mu_j$. The minimax formula for estimating μ_i is

$$\hat{\mu}_i = \min_{\mathcal{L}:\, i\in\mathcal{L}} \max_{\mathcal{U}:\, i\in\mathcal{U}} A_{\mathbf{X}}(\mathcal{L}\cap\mathcal{U}) = \max_{\mathcal{U}:\, i\in\mathcal{U}} \min_{\mathcal{L}:\, i\in\mathcal{L}} A_{\mathbf{X}}(\mathcal{L}\cap\mathcal{U}), \tag{2.1}$$

where $\mathcal{U}$ and $\mathcal{L}$ are, respectively, the upper and lower sets of $\mathcal{I}$. An upper set $\mathcal{U}$ is a set such that if $i \in \mathcal{U}$ and $i \preceq j$, then $j \in \mathcal{U}$. A lower set $\mathcal{L}$ is defined similarly with $i \preceq j$ replaced by $j \preceq i$. Furthermore, for an arbitrary set $\mathcal{S}$,

$$A_{\mathbf{X}}(\mathcal{S}) = \frac{\sum_{i\in\mathcal{S}} W_i X_i}{\sum_{i\in\mathcal{S}} W_i}. \tag{2.2}$$

The two well-studied order restrictions in the context of isotonic regression are (i) the simple order and (ii) the simple tree order. For (i), expression (2.1) reduces to

$$\hat{\mu}_i^{SO} = \min_{i\le t} \max_{s\le i} A_{\mathbf{X}}(s:t) = \max_{s\le i} \min_{i\le t} A_{\mathbf{X}}(s:t). \tag{2.3}$$

The notation $(s:t)$ denotes the set of integers $\{s, s+1, \dots, t\}$. In case of (ii), expression (2.1) reduces to

$$\hat{\mu}_1^{ST} = \min_{1 \in \mathcal{S}} A_{\mathbf{X}}(\mathcal{S}), \tag{2.4}$$

where the minimization is taken over all $\mathcal{S}$ containing 1 and

$$\hat{\mu}_i^{ST} = \max(\hat{\mu}_1^{ST}, X_i), \quad \text{for all } i \geq 2. \tag{2.5}$$

The results contained in this article are basically driven by the lemmas of the following type. Although we establish the lemmas for $i \preceq k$, and $1 \preceq i$, they can be proved for any pair of symbols $i, j \in \mathcal{I}$ with $i \preceq j$. In the next two lemmas we assume that the pdf of $\mathbf{X}$ is $f(\mathbf{X} - \mu)$.

LEMMA 2.1. *Let $i \preceq k$, $i \in \mathcal{I}$, with $i \neq k$, and let $c_\pm$ be two real numbers, with $c_- < c_+$. Then*

$$\frac{\partial}{\partial \mu_k} P(c_- < \hat{\mu}_i - \mu_i < c_+)$$
$$= \int_{\mathcal{A}} \cdots \int \left\{ f(\mathbf{Z} - \mu + (\mu_i + c_-)\mathbf{1}) - f(\mathbf{Z} - \mu + (\mu_i + c_+)\mathbf{1}) \right\} \prod_{l \neq k} dX_l^*,$$

where $\mathbf{Z} = (X_1^*, X_2^*, \dots, X_{k-1}^*, y)'$,

$$\mathcal{A} = \left\{ (X_1^*, X_2^*, \dots, X_{k-1}^*): \min_{i \in \mathcal{L}, k \notin \mathcal{L}} \max_{i \in \mathcal{U}} A_{\mathbf{X}^*}(\mathcal{L} \cap \mathcal{U}) > 0 \right\}$$

and

$$y = -\min_{i,k \in \mathcal{L}} \max_{i \in \mathcal{U}} \sum_{j \in \mathcal{L} \cap \mathcal{U}, j \neq k} W_j X_j^* / W_k.$$

PROOF.

$$P(c_- < \hat{\mu}_i - \mu_i < c_+) = P(\hat{\mu}_i > c_- + \mu_i) - P(\hat{\mu}_i > c_+ + \mu_i)$$
$$= Q(c_- + \mu_i) - Q(c_+ + \mu_i),$$

where $Q(c) = P(\hat{\mu}_i > c)$. Performing the change of variables $X_i^* = X_i - c$, one can write $Q(c) = P(\min_{i \in \mathcal{L}} \max_{i \in \mathcal{U}} A_{\mathbf{X}^*}(\mathcal{L} \cap \mathcal{U}) > 0)$. Note that the event $\min_{i \in \mathcal{L}} \max_{i \in \mathcal{U}} A_{\mathbf{X}^*}(\mathcal{L} \cap \mathcal{U}) > 0$ is equivalent to

$$\min\left\{ \min_{i \in \mathcal{L}, k \notin \mathcal{L}} \max_{i \in \mathcal{U}} A_{\mathbf{X}^*}(\mathcal{L} \cap \mathcal{U}), \min_{i,k \in \mathcal{L}} \max_{i \in \mathcal{U}} A_{\mathbf{X}^*}(\mathcal{L} \cap \mathcal{U}) \right\} > 0,$$

which is equivalent to

$$\min_{i \in \mathcal{L}, k \notin \mathcal{L}} \max_{i \in \mathcal{U}} A_{\mathbf{X}^*}(\mathcal{L} \cap \mathcal{U}) > 0 \quad \text{and} \quad \min_{i,k \in \mathcal{L}} \max_{i \in \mathcal{U}} A_{\mathbf{X}^*}(\mathcal{L} \cap \mathcal{U}) > 0.$$

Observe that the event

$$\min_{i,k\in\mathcal{L}}\ \max_{i\in\mathcal{U}}\ A_{\mathbf{X}^*}(\mathcal{L}\cap\mathcal{U})>0$$

is equivalent to

$$\max_{i\in\mathcal{U}} A_{\mathbf{X}^*}(\mathcal{L}\cap\mathcal{U})>0,\quad \text{for every } \mathcal{L} \text{ such that } i,k\in\mathcal{L}.$$

Note that $i\in\mathcal{U}$ implies that $k\in\mathcal{U}$. The last displayed inequality is equivalent to $X_k^*\geq y$. Making a transformation of variables $t_k=X_k^*-\mu_k$, we write

$$Q(c)=\int_{\mathcal{A}}\cdots\int_{t_k\geq y-\mu_k} f\left(X_1^*-\mu_1+c,\ldots,X_{k-1}^*-\mu_{k-1}+c,t_k+c\right)dt_k\prod_{l\neq k}dX_l^*.$$

Since $\mathcal{A}$ does not involve μ_k, we therefore have

$$\frac{\partial}{\partial\mu_k}Q(c)=\int_{\mathcal{A}}\cdots\int f(\mathbf{Z}-\boldsymbol{\mu}+c\mathbf{1})\prod_{l\neq k}dX_l^*.$$

This, together with

$$P(c_-<\hat{\mu}_i-\mu_i<c_+)=Q(c_-+\mu_i)-Q(c_++\mu_i),$$

proves the lemma. □

Similarly we have the following lemma.

LEMMA 2.2. *If* $1\preceq i,i\in\mathcal{I}$, *with* $i\neq 1$, *and if* $c_\pm$ *are two real numbers as before, then*

$$\begin{aligned}\frac{\partial}{\partial\mu_1}&P(c_-<\hat{\mu}_i-\mu_i<c_+)\\&=\int_{\mathcal{A}}\cdots\int\left\{f(\mathbf{Z}-\boldsymbol{\mu}+(\mu_i+c_-)\mathbf{1})-f\left(\mathbf{Z}-\boldsymbol{\mu}+(\mu_i+c_+)\mathbf{1}\right)\right\}\prod_{l\neq 1}dX_l^*,\end{aligned}$$

where $\mathbf{Z}=(y,X_2^*,X_3^*,\ldots,X_k^*)'$,

$$\mathcal{A}=\left\{X^*:\max_{i\in\mathcal{U},1\notin\mathcal{U}}\ \min_{i\in\mathcal{L}} A_{\mathbf{X}^*}(\mathcal{L}\cap\mathcal{U})<0\right\}$$

and

$$y=-\max_{1,i\in\mathcal{U}}\ \min_{i\in\mathcal{L}}\ \sum_{j\in\mathcal{L}\cap\mathcal{U},j\neq 1}\frac{W_jX_j^*}{W_1}.$$

PROOF. Use the formula

$$\hat{\mu}_i=\max_{\mathcal{U}:\ i\in\mathcal{U}}\ \min_{\mathcal{L}:i\in\mathcal{L}} A_{\mathbf{X}}(\mathcal{L}\cap\mathcal{U}),\tag{2.6}$$

and proceed as in the proof of Lemma 2.1. □

2.1. *Estimation of the nodes of a graph.* To motivate the development of some alternate estimators to the restricted maximum likelihood estimators, we first consider the estimation of the smallest parameter, μ_1, in a simple tree ordering. Under the simple tree order restriction, Lee (1988) proved that the restricted maximum likelihood estimator (2.4) for μ_1 fails to dominate X_1 in the mean squared error sense as $k \to \infty$. In the following we note a similar phenomenon in terms of the confidence interval estimation when k becomes very large.

THEOREM 2.3. *Suppose for $i = 1, 2, \ldots, k, X_i$ are independently distributed with mean μ_i and variance σ_i^2 and with a support $(-\infty, \infty)$. Suppose $\mu_i - \mu_1$ is bounded above by B and the σ_i^2's are bounded below and above by positive constants as $k \to \infty$. Then, for every c_- and c_+,*

$$\lim_{k\to\infty} P(c_- < \hat{\mu}_1^{ST} - \mu_1 < c_+) = 0,$$

where the weights $W_i = 1/\sigma_i^2$.

PROOF. Let $\varepsilon_i = (X_i - \mu_i)/\sigma_i$, and let $\varepsilon_{(2)} = \min_{i\geq 2} \varepsilon_i$ and $W_{(2)}$ be the weight corresponding to $\varepsilon_{(2)}$. Direct calculation shows that

$$\begin{aligned}\hat{\mu}_1^{ST} - \mu_1 &\leq \min_{i\geq 2} \frac{\sqrt{W_1}\varepsilon_1 + \sqrt{W_i}\varepsilon_i}{W_1 + W_i} + B \\ &\leq \frac{\sqrt{W_1}\varepsilon_1 + \sqrt{W_{(2)}}\varepsilon_{(2)}}{W_1 + W_{(2)}} + B \\ &\leq \frac{|\varepsilon_1|}{\sqrt{W_1}} + B + \frac{\sqrt{W_{(2)}}\varepsilon_{(2)}}{W_1 + W_{(2)}}.\end{aligned}$$

Since the W_i's are bounded below and above by positive real numbers, the coefficient $\sqrt{W_{(2)}}/(W_1 + W_{(2)})$ is bounded below by a constant $K > 0$. Therefore, the last term of the last upper bound is bounded above by $K\varepsilon_{(2)}$ if $\varepsilon_{(2)} < 0$, which holds almost surely as $k \to \infty$. These, together with the fact that $\varepsilon_{(2)} \to -\infty$ as $k \to \infty$, imply that the last displayed expression approaches $-\infty$ and hence the theorem. □

Thus we see that, apart from being complicated to derive, it is not necessary that the restricted maximum likelihood estimator will improve upon the unrestricted one. Hence, one would like to know if there exists a simple alternative estimator which dominates the unrestricted maximum likelihood estimator. The answer to this question is contained in Theorem 2.4, proved below.

Before proceeding any further, we now develop some useful notation. A pictorial representation of the population parameters denoted by solid circles, which are joined together by line segments, will be called a *graph*. For instance, in Figure 1 each of the pictures is a graph, and each solid circle represents a parameter $\mu_i, i = 1, 2, \ldots, k$. As examples, graph (a) corresponds to the

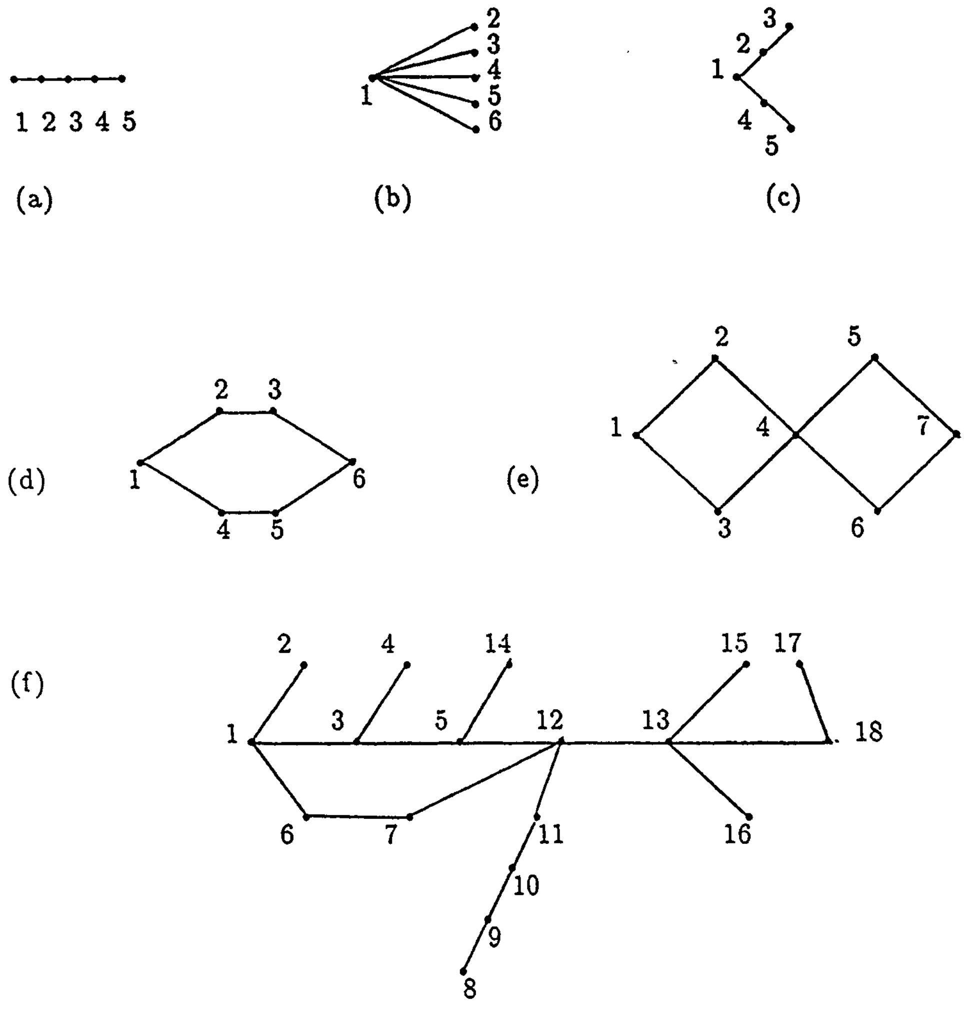

FIG. 1.

simple ordering, (b) corresponds to the simple tree ordering, (c) is umbrella ordering, (e) is double loop ordering and (f) is a combination of (a)–(d).

We omit writing μ on the graphs but only write the subscripts. A line segment joining two circles indicates that the circle to the right is known to correspond to the larger μ than the one to the left. Furthermore, μ_i is said to be a node if, for any, s, it is known that either

$$\mu_s \leq \mu_i \quad \text{or} \quad \mu_s \geq \mu_i.$$

We shall number the means μ_i's so that if it is known that $\mu_i \leq \mu_j$, then $i \leq j$.

Therefore if μ_i is a node, then

(2.7) $$\mu_s \le \mu_i \text{ and } \mu_i \le \mu_t, \quad \text{for all } s, t \text{ such that } s \le i \le t.$$

We now note the following important remark [Proposition 4.1.1 in Tong (1990)], which plays a crucial role in the subsequent arguments. We shall say that a probability density function (pdf) is elliptically symmetric unimodal if it is of the form

(2.8) $$f(\mathbf{x}-\mu) = g((\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)) \quad \text{with } g(u) \text{ nonincreasing in } u.$$

REMARK 2.1. If $\mathbf{X}$ has an elliptically symmetric unimodal pdf then so does any subvector of $\mathbf{X}$.

The main theorem of this section is as follows.

THEOREM 2.4. *Suppose the pdf of* $\mathbf{X}$ *satisfies* (2.8) *with diagonal* Σ *and suppose* $\hat{\mu}_i^{SO}$ *is as in* (2.3), *where* $\mathbf{W}$ *consists of the diagonal elements of* Σ^{-1}. *Then* $\hat{\mu}_i^{SO}$ *stochastically dominates* X_i *that is, for all* $c > 0$,

(2.9) $$P(|\hat{\mu}_i^{SO} - \mu_i| < c) \ge P(|X_i - \mu_i| < c),$$

as long as μ_i *is a node. Moreover, the inequality in* (2.9) *is a strict one if* $g(u) > 0$ *for all* $u > 0$.

PROOF. Assume without loss of generality that $\mu_i = 0$. We now consider two cases: (i) $i \ne k$ and $\sum W_j\mu_j \ge 0$; (ii) $i \ne 1$ and $\sum W_j\mu_j \le 0$. These two cases exhaust all possibilities. The assertion is obviously true for $1 < i < k$. Furthermore, for $i = 1$, case (i) applies since by (2.7) $\mu_j \ge \mu_i$, for all j, which implies that $\sum W_j\mu_j \ge 0$. A similar argument shows that, for $i = k$, case (ii) applies.

Now consider case (i). By Lemma 2.1,

$$\frac{\partial}{\partial \mu_k} P(-c < \hat{\mu}_i^{SO} - \mu_i < c) \le 0,$$

if $f(\mathbf{Z} - \mu - c\mathbf{1}) \le f(\mathbf{Z} - \mu + c\mathbf{1})$, which is true if

(2.10) $$\begin{aligned} &\sum_{j=1}^{k-1} W_j(X_j^* - \mu_j - c)^2 + W_k(y - \mu_k - c)^2 \\ &\quad \ge \sum_{j=1}^{k-1} W_j(X_j^* - \mu_j + c)^2 + W_k(y - \mu_k + c)^2. \end{aligned}$$

Note here that

$$y = -\max_{s \le i} \sum_{j=s}^{k-1} \frac{W_j X_j^*}{W_k}.$$

Now (2.10) is equivalent to requiring

$$\sum_{j=1}^{k-1} W_j(X_j^* - \mu_j) + W_k(y - \mu_k) \leq 0,$$

which is equivalent to

$$\sum_{j=1}^{k-1} W_j X_j^* - \max_{s \leq i} \sum_{j=s}^{k-1} W_j X_j^* - \sum_{j=1}^{k} W_j \mu_j \leq 0.$$

This inequality obviously holds for case (i). Letting $\mu_k \to \infty$ and using (2.3), we obtain a lower bound for the left-hand side of (2.9), which is itself in the absence of X_k.

Next, consider case (ii). Lemma 2.2 provides a formula for

$$\frac{\partial}{\partial \mu_1} P\left(-c < \hat{\mu}_i^{SO} - \mu_i < c\right).$$

Then, following an argument similar to the preceding paragraph, we conclude that this derivative is positive. Hence by letting $\mu_1 \to -\infty$ we obtain a lower bound for $P(|\hat{\mu}_i^{SO} - \mu_i| < c)$, which is itself in the absence of X_1.

Applying these arguments inductively $(k-1)$ times, we get the lower bound for $P(|\hat{\mu}_i^{SO} - \mu_i| < c)$, which is exactly the right-hand side of (2.9) with nonstrict inequality. Now assume that $g(u) > 0$ for all $u > 0$. To argue for the strict inequality, suppose to the contrary that at every stage of the above induction process the partial derivatives are zero. Then for $k = 2$ one can verify that $g(u)$ is constant for sufficiently large u. But since g is a pdf this implies that the constant is zero, contradicting the assumption that $g(u) > 0$ for all $u > 0$. Hence the theorem follows. □

When the parameters are subject to simple order restriction, through a different argument, assuming normality, Kelly (1989) proved that $\hat{\mu}_i$ universally dominates X_i. Kushary and Cohen (1989) proved a similar result when $k = 2$. Kelly's result implies Lee's (1981) result which shows that $\hat{\mu}_i$ has a smaller mean squared error (MSE). Theorem 2.4 not only extends both these results to spherically symmetric distributions, but provides a method for constructing better estimators for other forms of order restrictions (see Corollary 2.6 and Section 2.2).

Note that the stochastic domination in (2.9) is equivalent to the universal domination criterion (with respect to Euclidean error). That is, with respect to the loss function $L(|\delta - \mu_i|)$, where L is an arbitrary nondecreasing and nonconstant function, the risk of $\hat{\mu}_i^{SO}$ is always smaller than that of X_i for every μ_i. See Hwang (1985, 1986) for further discussion on universal domination. In the following discussion we shall use stochastic domination and universal domination interchangeably.

The following corollaries for the simple ordering and simple tree ordering follow from Theorem 2.4 under the same distributional assumptions of **X** stated in the theorem.

COROLLARY 2.5 (Simple ordering). *$\hat{\mu}_i^{SO}$ universally dominates X_i, for estimating μ_i, when* (1.1) *holds.*

COROLLARY 2.6 (Simple tree ordering). *$\hat{\mu}_1 = \hat{\mu}_1^{SO}$ universally dominates X_1, for estimating μ_1, when* (1.2) *holds.*

The significance of the above results is that the simple order MLE performs well for other orderings including simple tree. However, strikingly, under simple tree, the corresponding MLE ($\hat{\mu}_1^{ST}$) fails in the sense of Theorem 2.3.

For estimating $\mu_i, i > 1$, in a simple tree order restriction we propose the following estimator:

$$\hat{\mu}_i = \max(X_i, \hat{\mu}_1).$$

Qualitatively, the estimator makes sense in that, together with $\hat{\mu}_1$, it will satisfy the simple tree constraint and it also mimics (2.5), except that the restricted maximum likelihood estimator $\hat{\mu}_1^{ST}$ has been replaced by a more appropriate estimator $\hat{\mu}_i$. The following theorem supplies evidence that $\hat{\mu}_i$ performs well.

THEOREM 2.7. *Let $i \neq 1$ be fixed. Assume that the pdf of* **X** *satisfies* (2.8), *and suppose*

$$\mu_i > \frac{\sum_{j \neq i} \mu_j W_j}{\sum_{j \neq i} W_j}. \tag{2.11}$$

Then, for every $c > 0$, the coverage probability of the interval $\hat{\mu}_i \pm c$ is no less than the coverage probability of $X_i \pm c$. It is strictly larger if $g(u) > 0$ for all $u > 0$.

PROOF. We shall show that

$$P(|\max(X_i, \hat{\mu}_1) - \mu_i| < c) \tag{2.12}$$

increases as a function of μ_1. This will complete the proof since as $\mu_1 \to -\infty$, $\hat{\mu}_1 \to -\infty$ and hence $\hat{\mu}_i$ reduces to X_i.

Write (2.12) as $Q(\mu_i + c) - Q(\mu_i - c)$, where for any constant $c_*, Q(c_*) = P(\hat{\mu}_i < c_*)$. Then $Q(c_*) = P(X_i < c_*) - P(X_i < c_*, \hat{\mu}_1 \geq c_*)$. Hence,

$$\frac{\partial}{\partial \mu_1} Q(c_*) = -\frac{\partial}{\partial \mu_1} P(X_i < c_*, \hat{\mu}_1 \geq c_*).$$

Although Lemmas 2.1 and 2.2 do not provide a formula for the differentiation of the right-hand side of this expression, similar calculations yield the following formula:

$$\frac{\partial}{\partial \mu_1} P(X_i < c_*, \hat{\mu}_1 \geq c_*) = \int_{X_i^* < 0} \cdots \int f(\mathbf{Z} - \mu + c_* \mathbf{1}) \, dX_2^* \cdots dX_k^*,$$

where

$$\mathbf{Z} = \left(-\min_{s \geq 1} \frac{\sum_{j=2}^{s} W_j X_j^*}{W_1}, X_2^*, \ldots, X_k^* \right)',$$

and when $s = 1$ for any a_j, $\sum_{j=2}^{s} a_j$ is defined to be zero. Putting all this together, the derivative of (2.12) with respect to μ_1 is

$$\int_{X_i^* < 0} \cdots \int \{ f(\mathbf{Z} - \mu + (\mu_i - c)\mathbf{1}) - f(\mathbf{Z} - \mu + (\mu_i + c)\mathbf{1}) \} \, dX_2^* \, dX_3^* \cdots dX_k^*.$$

We can now establish the theorem by showing that this integrand is nonnegative, which holds if

$$\begin{aligned} &(\mathbf{Z} - \mu + (\mu_i - c)\mathbf{1})' \Sigma^{-1} (\mathbf{Z} - \mu + (\mu_i - c)\mathbf{1}) \\ &\quad \leq (\mathbf{Z} - \mu + (\mu_i + c)\mathbf{1})' \Sigma^{-1} (\mathbf{Z} - \mu + (\mu_i + c)\mathbf{1}). \end{aligned}$$

This can be simplified to

$$-\min_{s \geq 1} \sum_{j=2}^{s} X_j^* W_j + \sum_{j=2}^{k} X_j^* W_j + \sum_{j=1}^{k} (\mu_i - \mu_j) W_j \geq 0,$$

which follows from the assumption of the theorem. The strict inequalities can be argued as in Theorem 2.4. □

In a related result, Lee [(1988), Theorem 4.1] proved that $\hat{\mu}_i^{ST}, i \geq 2$, has a smaller mean squared error than X_i under a condition which is equivalent to (2.11). In fact, under the same condition his result can be strengthened to stochastic domination by directly applying our Lemma 2.1. Unfortunately its companion estimator $\hat{\mu}_1^{ST}$ may be inappropriate, as shown in Theorem 2.3.

When applying $\hat{\mu}_1$ to a given problem involving simple tree ordering, the best way is to label the population, according to some prior information or a guess, so that the μ_i's are increasing in i. Corollary 2.6, however, assures us that $\hat{\mu}_1$ will stochastically dominate X_1 even if the labeling is wrong. The numerical studies are consistent with this fact. In Table 1, cases (i) and (iii) correspond to a correct labeling while (ii) corresponds to an incorrect one.

Both $\hat{\mu}_1$ and $\hat{\mu}_2$ dominate X_1 and X_2 substantially even though in cases (i) and (iii) (2.11) is not satisfied. Furthermore it has larger coverage probability than the interval centered at the restricted maximum likelihood estimator, namely, $\hat{\mu}_i^{ST} \pm c$.

TABLE 1

Coverage probabilities: Performance of the proposed estimators $\hat{\mu}_i$ and the restricted maximum likelihood estimators $\hat{\mu}_i^{ST}$ in the simple tree ordering problem [*graph* (b)]. *Results are based on* 2500 *simulation runs generated from* 15 *normal populations, with the variance of the ith population being i: case* (i) $\mu_1 = 0, \mu_2 = 0$ *and, for* $3 \le j \le 15, \mu_j = 0.1(1+j)$; *case* (ii) $\mu_1 = 0, \mu_2 = 1$ *and, for* $3 \le j \le 15, \mu_j = 0.1(1+j)$; *case* (iii) $\mu_1 = 0, \mu_2 = 1$ *and, for* $3 \le j \le 15, \mu_j = 0.1(10+j)$

Case	Nominal level	$\hat{\mu}_1$	$\hat{\mu}_1^{ST}$	$\hat{\mu}_2$	$\hat{\mu}_2^{ST}$	Case	Nominal level	$\hat{\mu}_1$	$\hat{\mu}_1^{ST}$	$\hat{\mu}_2$	$\hat{\mu}_2^{ST}$
(i)	0.95	0.97	0.72	0.97	0.97	(iii)	0.68	0.74	0.46	0.76	0.72
(ii)	0.68	0.77	0.31	0.75	0.70		0.90	0.94	0.74	0.94	0.93
	0.90	0.94	0.60	0.94	0.92		0.95	0.97	0.84	0.97	0.97
	0.95	0.97	0.72	0.97	0.97						

In case of the simple tree ordering, Lee [(1988), Theorem 3.1] studied a modified weight isotonic regression estimator for the control mean μ_1. He proved in Theorem 3.1 of his paper that if the weight W_1 corresponding to μ_1 is chosen to be sufficiently large, then, under the assumption of normality, the mean squared error of the isotonic regression estimator is strictly smaller than the mean squared error of X_1. Kelly (1990) also studied a similar kind of estimator for μ_1 using stochastic domination criterion. No modified estimators were proposed for other means and neither of their estimators were extended to other forms of order restrictions.

2.2. *Other order restrictions.* In general it is not easy to compute the restricted maximum likelihood estimators for any arbitrary graph. Further, as seen in Theorem 2.3, the restricted maximum likelihood estimator can be disastrous for some types of order restrictions such as the simple tree order. Thus there is a need to develop a simple scheme to construct improved estimators for parameters subject to different types of order restrictions. In this section we shall develop a general scheme, based upon simple order estimators, which is motivated and justified to some extent by Theorem 2.4. A very fast way of calculating simple order estimators is PAVA as depicted in Robertson, Wright and Dykstra (1988).

CASE A (Graphs with at least one node).

Step 1. *Estimate the nodal means of the graph.* While maintaining the known inequalities between the means of a graph and guessing the inequalities between the rest of the means, we first estimate the nodes of a graph using the simple order estimator $\hat{\mu}_i^{SO}$. Thus, for instance, in graph (e), the double loop order restriction, we shall first estimate the means μ_1,μ_4 and μ_7 while guessing the inequality between μ_2 and μ_3 and between μ_5 and μ_6. Theorem 2.4 assures that the simple order estimators for the nodes so constructed will stochastically dominate the unrestricted maximum likelihood estimators of the nodes. The improvements will be more substantial if we guess the correct orders. However, there is no harm even if our guess is completely wrong.

TABLE 2

Coverage probabilities: Performance of the proposed estimators of μ_1 and μ_2 under the umbrella ordering [similar to graph (c)]. Results are based on 2500 simulation runs generated from 15 normal populations, each population having the same variance, 1: case (i) $\mu_1 = -1, \mu_2 = 0, \mu_3 = 0.5$ and, for $4 \le j \le 15, \mu_j = 0.1(10+j)$; case (ii) $\mu_1 = -1, \mu_2 = 0, \mu_3 = 1$ and, for $4 \le j \le 15, \mu_j = 0.1(10+j)$; case (iii) $\mu_1 = -1, \mu_2 = 0, \mu_3 = 2$ and, for $4 \le j \le 15, \mu_j = 0.1(10+j)$

Case	Nominal level	$\hat{\mu}_1$	$\hat{\mu}_2$	Case	Nominal level	$\hat{\mu}_1$	$\hat{\mu}_2$
(i)	0.68	0.74	0.81	(iii)	0.68	0.72	0.75
	0.90	0.94	0.97		0.90	0.94	0.94
	0.95	0.97	0.99		0.95	0.97	0.98
(ii)	0.68	0.73	0.79				
	0.90	0.94	0.96				
	0.95	0.97	0.99				

For instance, in Table 2 we study the performance of the proposed estimators of the parameters μ_1 and μ_2 under the umbrella ordering [similar to graph (c)]. The estimator for μ_2 will be depicted in Step 2. There are two branches to the umbrella, one of them contains μ_1, μ_2, μ_3, and the second branch contains $\mu_1, \mu_4, \mu_5, \ldots, \mu_{15}$. In the construction of $\hat{\mu}_1$ we guessed in each of the three cases (i), (ii) and (iii) that $\mu_1 \le \mu_2 \le \cdots \le \mu_{15}$. Thus in cases (i) and (ii) we made correct guesses but made a wrong guess in case (iii). In all the cases the proposed estimators perform better than the unrestricted maximum likelihood estimators.

Step 2. Estimate the nonnodal means of the graph. To estimate the nonnodal mean, μ_i, we remove the smallest number of circles from the graph so that μ_i becomes a node in the resulting subgraph $\mathcal{G}_{\mu_i}$. Based on the circles in $\mathcal{G}_{\mu_i}$ we then construct the simple order estimator for μ_i using the procedure described above. It should be noted, however, that some of the circles in $\mathcal{G}_{\mu_i}$ might have been estimated earlier. For those circles, while forming the simple order estimator for μ_i we shall use the estimators as the observations and give a weight of B to them, where $B \to \infty$. Consider, for example, μ_2 in graph (e). If we delete μ_3, then μ_2 is a node in the resulting subgraph $\mathcal{G}_{\mu_2}$. So we shall construct the simple order estimator for μ_2, while guessing the inequality between μ_5 and μ_6. Let us suppose that our prior information indicates $\mu_5 \le \mu_6$. Note that $\mu_i, i = 1, 4$ and 7, have been estimated as $\hat{\mu}_i$ in Step 1. Therefore we shall construct the simple order estimator $\hat{\mu}_i^{SO}(B)$ for μ_i based on the "observations" $\{\hat{\mu}_1, X_2, \hat{\mu}_4, X_5, X_6, \hat{\mu}_7\}$ with respect to weights $\{B, W_2, B, W_5, W_6, B\}$. The recommended estimator $\hat{\mu}_2$ for μ_2 is then $\lim_{B\to\infty} \hat{\mu}_2^{SO}(B)$. Note that this estimator is the projection of X_2 on to the interval $[\hat{\mu}_1, \hat{\mu}_4]$. Thus we have estimated μ_1, μ_2, μ_4 and μ_7 by $\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_4$ and $\hat{\mu}_7$, respectively. By eliminating the second population, μ_3 becomes a node. Thus, similarly, μ_3 is estimated using the "observations" $\hat{\mu}_1, \hat{\mu}_4, X_5, X_6$ and $\hat{\mu}_7$. Estimators for μ_5 and μ_6 are similarly obtained one by one.

It can be argued in general that the estimators constructed this way satisfy the original constraint. For instance, in the preceding paragraph, since $\hat{\mu}_i^{SO}(B) \le \hat{\mu}_2^{SO}(B) \le \hat{\mu}_4^{SO}(B)$ and $\lim_{B\to\infty} \hat{\mu}_i^{SO}(B) \to \hat{\mu}_i$, for $i = 1, 4$, it follows that $\hat{\mu}_1 \le \hat{\mu}_2 \le \hat{\mu}_4$. In practice, there is no difficulty in evaluating the limits. One simply chooses a number B which is, say, 10^5 times larger than the maximum weight and uses the corresponding $\hat{\mu}^{SO}(B)$, which should be fairly close to its limit.

We are thankful to a referee who pointed out correctly that there is an alternative way to calculate directly our estimators for the nonnodal means without letting $B \to \infty$. Since our problem is indeed equivalent to the bounded isotonic regression problem where the bounds are formed by the estimators of the previously estimated parameters, we may consequently use the modified PAVA algorithm on page 57 of Barlow, Bartholomew, Bremner and Brunk (1972). This method has been demonstrated to work when the covariance matrix is a diagonal matrix and is recommended for such a case. However, we are not aware of its validity when the covariance matrix is nondiagonal and hence a general approach is described as above.

When we apply the proposed method to graph (a), we obtain the same estimator as $\hat{\mu}_i^{SO}$ of (2.3). Furthermore, its application to graph (b) yields the estimators recommended in Corollary 2.6 and Theorem 2.7.

CASE B (Graphs with no nodes). Suppose there are no nodes in a graph, such as graph (f). Then as in Step 2 of Case A, corresponding to each circle μ_i we shall remove the smallest number of circles so that μ_i is a node in the resulting subgraph $\mathcal{G}_{\mu_i}$. Then among all the subgraphs $\mathcal{G}_{\mu_1}, \mathcal{G}_{\mu_2}, \ldots, \mathcal{G}_{\mu_k}$ select the one which has largest number of circles. Estimate all the nodes in that subgraph using the simple order estimator $\hat{\mu}^{SO}$. Due to Theorem 2.4 these estimators will dominate the unrestricted maximum likelihood estimators. Then go on to the subgraph which has next largest number of circles and obtain the simple order estimator of all the nodes in the subgraph and proceed this way till all the parameters are estimated. In any subgraph, if a circle appears that has already been estimated previously, then a weight of B should be given to that estimator as done in Step 2 of Case A, where $B \to \infty$. As an example, consider the graph (f), where there are no nodes. Note that $\mathcal{G}_{\mu_{12}} = \mathcal{G}_{\mu_{13}} = \{1, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 18\}$ is the largest subgraph with 14 circles in it. We construct the simple order estimators $\hat{\mu}_{12}^{SO}$ and $\hat{\mu}_{13}^{SO}$ using the data on the 14 populations. The next largest subgraph is $\mathcal{G}_{\mu_1} = \{1, 2, 3, 4, 5, 6, 7, 12, 13, 14, 15, 16, 18\}$. The node in this subgraph is μ_1. Since μ_{12} and μ_{13} have already been estimated in the last step, we shall replace the data X_{12} and X_{13} by their estimates $\hat{\mu}_{12}$ and $\hat{\mu}_{13}$. Thus the "data" used for estimating μ_1 is $\{X_1, X_2, X_3, X_4, X_5, X_6, X_7, \hat{\mu}_{12}, \hat{\mu}_{13}, X_{14}, X_{15}, X_{16}, X_{18}\}$ with weights $\{W_1, W_2, W_3, W_4, W_5, W_6, W_7, B, B, W_{14}, W_{15}, W_{16}, W_{18}\}$, where $B \to \infty$. We then construct the simple order estimator $\hat{\mu}_1^{SO}(B)$ for the node μ_1. Then the recommended estimator for μ_1 is $\hat{\mu}_1 = \lim_{B\to\infty} \hat{\mu}_1^{SO}(B)$. Continue this process till all the required parameters are estimated.

As described in Step 2 of Case A, the proposed estimators would preserve the known order restrictions on the parameters.

3. Nondiagonal case. There are two problems with the restricted maximum likelihood estimators when Σ is nondiagonal. First, they are computationally quite intensive. One needs to determine which of the 2^k partitions the data $\mathbf{X}$ belongs to, and in each determination a calculation of the inverse of a suitable covariance matrix is required. Second and more important, the restricted maximum likelihood estimators do not always behave properly even for the simple ordering. The following theorem exhibits such cases.

THEOREM 3.1. *Suppose the pdf of* $\mathbf{X}$ *satisfies* (2.8) *with* Σ *known and the components of* μ *satisfy the simple ordering, and suppose that* $f(u)$ *is strictly decreasing in* u. *Then there exists a* k *and a* Σ *such that the restricted maximum likelihood estimator* $\hat{\mu}_1^M$ *of* μ_1 *fails to dominate* X_1 *both stochastically and in mean squared error. In fact, the confidence interval centered at* $\hat{\mu}_1^M$ *can have coverage probability as small as one wishes by choosing* k.

PROOF. Without loss of generality, assume $\mathbf{X}$ is normally distributed. Choose a matrix $\mathbf{L}$ such that $\mathbf{X}^* = \mathbf{L}\mathbf{X}$ is normally distributed with mean and covariance matrix given by

$$\boldsymbol{\mu}^* = \mathbf{L}\boldsymbol{\mu} = (\mu_1, \mu_2 - \mu_1, \dots, \mu_k - \mu_{k-1})' \quad \text{and} \quad \Sigma^* = \mathbf{L}\Sigma\mathbf{L}',$$

respectively. Under this linear transformation, the simple order restriction in the parameters $\mu_1, \mu_2, \dots, \mu_k$ is equivalent to the following:

(P) Estimate $\boldsymbol{\mu}^*$ subject to the constraint $\mu_j^* \geq 0$ for all $j > 1$.

Now consider a simple tree problem based on $\mathbf{Y} \sim N(\boldsymbol{\mu}, \mathbf{D})$, where k is large enough and $\mathbf{D}$ is diagonal, and that the restricted maximum likelihood estimator fails for μ_1. Apply a linear transformation $\mathbf{M}$ such that

$$\boldsymbol{\mu}^{**} = \mathbf{M}\boldsymbol{\mu} = (\mu_1, \mu_2 - \mu_1, \dots, \mu_k - \mu_1)'.$$

This will transform the problem to P if Σ satisfies

$$\Sigma = \mathbf{L}^{-1}\mathbf{M}\mathbf{D}\mathbf{M}'\mathbf{L}'^{-1},$$

since then $\mathbf{MY}$ and $\mathbf{LX}$ are identically distributed. Since the first components of $\boldsymbol{\mu}, \boldsymbol{\mu}^*$ and $\boldsymbol{\mu}^{**}$ are the same, the restricted maximum likelihood estimator of μ_1 is the same for all of them. Hence the restricted maximum likelihood estimator fails for the simple ordering problem too. □

To describe our proposed procedure, let

$$A_{\mathbf{X}}(\mathcal{S}) = \mathbf{W}_{\mathcal{S}}'\mathbf{X}_{\mathcal{S}}/\mathbf{1}'\mathbf{W}_{\mathcal{S}},$$

where $\mathbf{W}_{\mathcal{S}} = (\Sigma_{\mathcal{S}})^{-1}\mathbf{1}$ and $\Sigma_{\mathcal{S}}$, a submatrix of Σ, is proportional to the covariance matrix of the subvector $\mathbf{X}_{\mathcal{S}}$ when (2.8) is assumed. The vector $\mathbf{X}_{\mathcal{S}}$ denotes the subvector of $\mathbf{X}$ consisting of elements X_i, $i \in \mathcal{S}$. Below, corresponding to a dummy vector $\mathbf{X}^*$ defined in Lemma 3.2, we shall similarly define a subvector $\mathbf{X}^*_{\mathcal{S}}$. Also below, for a set $\mathcal{S}$ that contains 1 (or k), we let $\mathbf{X}^*_{\mathcal{S}_{(1)}}$ and $\mathbf{W}_{\mathcal{S}_{(1)}}$ (or $\mathbf{X}^*_{\mathcal{S}_{(k)}}$ and $\mathbf{W}_{\mathcal{S}_{(k)}}$), respectively, denote $\mathbf{X}^*_{\mathcal{S}}$ and $\mathbf{W}_{\mathcal{S}}$ with the first (or the last) element deleted. Further, $\mathbf{W}_{\mathcal{S}_1}$ or $\mathbf{W}_{\mathcal{S}_k}$ denotes the first (or the last) element of $\mathbf{W}_{\mathcal{S}}$. We will use $\mathbf{W}$ to denote $\mathbf{W}_{\mathcal{S}}$ when $\mathcal{S} = \mathcal{I}$. With this generalization, let $\hat{\mu}_i^{\Sigma}$ denote $\hat{\mu}_i$ given in (2.1). The estimator we propose is

$$\hat{\mu}_i^{SO} = \min_{i \le t} \max_{s \le t} A_{\mathbf{X}}(s\colon t). \tag{3.1}$$

When Σ is diagonal, the estimators proposed in this section coincide with their counterparts given in Section 2.

Similar to Lemma 2.1, we can derive the following lemma, which provides the derivative formula to study the coverage probability of some confidence intervals.

LEMMA 3.2. *Let $i \preceq k, i \ne k, i \in \mathcal{I}$, and let $c_{\pm}$ be two real numbers, with $c_- < c_+$. Assume that $\mathbf{W}_{\mathcal{S}_k} > 0$ for every $\mathcal{S}$ that contains k. Then*

$$\frac{\partial}{\partial \mu_k} P(c_- < \hat{\mu}_i^{\Sigma} - \mu_i < c_+)$$
$$= \int_{\mathcal{A}} \cdots \int f(\mathbf{Z} - \mu + (\mu_i + c_-)\mathbf{1}) - f(\mathbf{Z} - \mu + (\mu_i + c_+)\mathbf{1}) \prod_{l \ne k} dX_l^*,$$

where $\mathbf{Z} = (X_1^*, X_2^*, \dots, X_{k-1}^*, y)'$,

$$\mathcal{A} = \left\{ (X_1^*, X_2^*, \dots, X_{k-1}^*)\colon \min_{i \in \mathcal{L}, k \notin \mathcal{L}} \max_{i \in \mathcal{U}} A_{\mathbf{X}^*}(\mathcal{L} \cap \mathcal{U}) > 0 \right\}$$

and

$$y = -\min_{i,k \in \mathcal{L}} \max_{i \in \mathcal{U}} \frac{\mathbf{W}'_{(\mathcal{L}\cap\mathcal{U})_{(k)}} \mathbf{X}^*_{(\mathcal{L}\cap\mathcal{U})_{(k)}}}{\mathbf{W}_{(\mathcal{L}\cap\mathcal{U})_k}}.$$

It is unfortunate that we cannot derive a result similar to Lemma 2.2 for $\hat{\mu}_i^{\Sigma}$ for every Σ. This is because, unlike in the diagonal case, the estimator given in (3.1) does not remain the same if the min and max operators are interchanged. We thank one of the referees for pointing out this fact to us through a counterexample. In certain cases, such as Examples 3.1 and 3.2, the min and max operators are interchangeable. We derive the following lemma similar to Lemma 2.2 for such cases.

LEMMA 3.3. *Assume that $\mathbf{W}_{\mathcal{S}_1} > 0$ for every $\mathcal{S}$ that contains 1. If $1 \preceq i$, $i \ne 1, i \in \mathcal{I}$, if $c_{\pm}$ are two real numbers as before and if Σ is such that*

$$\min_{\mathcal{L}: i \in \mathcal{L}} \max_{\mathcal{U}: i \in \mathcal{U}} A_{\mathbf{X}}(\mathcal{L} \cap \mathcal{U}) = \max_{\mathcal{U}: i \in \mathcal{U}} \min_{\mathcal{L}: i \in \mathcal{L}} A_{\mathbf{X}}(\mathcal{L} \cap \mathcal{U}), \tag{3.2}$$

then

$$\frac{\partial}{\partial \mu_1} P(c_- < \hat{\mu}_i^{\Sigma} - \mu_i < c_+)$$
$$= \int_{\mathcal{A}} \cdots \int f(\mathbf{Z} - \boldsymbol{\mu} + (\mu_i + c_-)\mathbf{1}) - f(\mathbf{Z} - \boldsymbol{\mu} + (\mu_i + c_+)\mathbf{1}) \prod_{l \neq 1} dX_l^*,$$

where $\mathbf{Z} = (y, X_2^*, X_3^*, \ldots, X_k^*)'$,

$$\mathcal{A} = \left\{ (X_2^*, X_3^*, \ldots, X_k^*) \colon \max_{i \in \mathcal{U}, 1 \notin \mathcal{U}} \min_{i \in \mathcal{L}} A_{\mathbf{X}^*}(\mathcal{L} \cap \mathcal{U}) < 0 \right\}$$

and

$$y = -\max_{1, i \in \mathcal{U}} \min_{i \in \mathcal{L}} \frac{\mathbf{W}'_{(\mathcal{L}\cap\mathcal{U})_{(1)}} \mathbf{X}^*_{(\mathcal{L}\cap\mathcal{U})_{(1)}}}{\mathbf{W}_{(\mathcal{L}\cap\mathcal{U})_1}}.$$

The following are two examples where (3.2) is satisfied.

EXAMPLE 3.1 (Intraclass correlation matrix). Suppose that $\Sigma = \sigma^2\{(1 - \rho)\mathbf{I} + \rho\mathbf{J}\}$, where $\mathbf{I}$ is the identity matrix, $\mathbf{J}$ is the matrix of unities and ρ is such that the matrix is a positive definite matrix, that is, $-1/(k-1) < \rho < 1$. Then observe that $\mathbf{W}_{\mathcal{S}} = (\Sigma_{\mathcal{S}})^{-1}\mathbf{1}$ is proportional to the vector $\mathbf{1}$, with a positive constant of proportionality. Hence condition (3.2) follows from the fact that it is satisfied for the diagonal case or specifically for the case where the weights W_i are all 1.

EXAMPLE 3.2 (Estimation of the smallest or the largest mean for a general Σ). If μ_i is known to be either the smallest or the largest mean, one of the operators can be eliminated and hence (3.2) obviously holds for any matrix Σ.

Next we show how to apply the preceding lemmas to establish domination results. In the rest of the section we assume that $\mathbf{X}$ has an elliptically symmetric unimodal density function (2.8) and $\mathbf{W} = \Sigma^{-1}\mathbf{1}$. We shall also assume that the equation (3.2) holds (with respect to the simple ordering). We first deal with the case when the μ's satisfy the simple ordering.

3.1. *Simple ordering.* In the following, a vector $\mathbf{V} = (V_1, V_2, \ldots, V_n)'$ is said to satisfy the nonnegative backward average property if, for all $i, \Sigma_{j=i}^n V_j \geq 0$ and is said to satisfy the nonnegative forward average property if, for all $i, \Sigma_{j=1}^i V_j \geq 0$.

REMARK 3.1. Note that $\mathbf{V}$ has a nonnegative backward average property if and only if $\mathbf{V}'\boldsymbol{\mu} \geq \mu_1 \mathbf{V}'\mathbf{1}$ for every $\boldsymbol{\mu}$ satisfying the simple ordering. Similarly, it has a nonnegative forward average property if and only if $\mathbf{V}'\boldsymbol{\mu} \leq \mu_k \mathbf{V}'\mathbf{1}$ for every $\boldsymbol{\mu}$ satisfying the simple ordering.

THEOREM 3.4. *For every $s, t, s \le i \le t$, suppose that the first and the last components of $\mathbf{W}_{(s:t)}$ are positive. Also suppose, for every $t > i, \mathbf{W}_{(i:t)}$ satisfies the nonnegative backward average property and, for every $s < i, \mathbf{W}_{(s:i)}$ satisfies the nonnegative forward average property. Then $\hat{\mu}_i^{SO}$, given in* (3.1), *universally dominates the unrestricted maximum likelihood estimator X_i. That is, the coverage probability of the interval $\hat{\mu}_i^{SO} \pm c$ is as large as that of $X_i \pm c$, for every $c > 0$. It is strictly larger if $g(u) > 0$ for all $u > 0$.*

PROOF. Without loss of generality assume that $\mu_i = 0$. As in the proof of Theorem 2.4, we shall consider two cases: (i) $i \neq k$ and $\mathbf{W}'\boldsymbol{\mu} \ge 0$; (ii) $i \neq 1$ and $\mathbf{W}'\boldsymbol{\mu} \le 0$. For $1 < i < k$, these two cases obviously make up all the situations. The same can be said for $i = 1$ or k, as can be established using Remark 3.1 and $\mu_i = 0$.

Using Lemma 3.2, we have

$$\frac{\partial}{\partial \mu_k} P(|\hat{\mu}_i^{SO} - \mu_i| < c) \le 0$$

if

$$f\big((\mathbf{Z} - \boldsymbol{\mu} - c\mathbf{1})'\Sigma^{-1}(\mathbf{Z} - \boldsymbol{\mu} - c\mathbf{1})\big) \le f\big((\mathbf{Z} - \boldsymbol{\mu} + c\mathbf{1})'\Sigma^{-1}(\mathbf{Z} - \boldsymbol{\mu} + c\mathbf{1})\big)$$

or, equivalently, if

$$(\mathbf{Z} - \boldsymbol{\mu} - c\mathbf{1})'\Sigma^{-1}(\mathbf{Z} - \boldsymbol{\mu} - c\mathbf{1}) \ge (\mathbf{Z} - \boldsymbol{\mu} + c\mathbf{1})'\Sigma^{-1}(\mathbf{Z} - \boldsymbol{\mu} + c\mathbf{1}), \tag{3.3}$$

where $\mathbf{Z}$ is as defined in the lemma with

$$y = -\max_{s \le i} \frac{\mathbf{W}'_{(s:k)_{(k)}} \mathbf{X}^*_{(s:k)_{(k)}}}{\mathbf{W}_{(s:k)_k}}.$$

The inequality (3.3) is equivalent to

$$-c\mathbf{1}'\Sigma^{-1}(\mathbf{Z} - \boldsymbol{\mu}) \ge 0$$

or $\mathbf{W}'(\mathbf{Z} - \boldsymbol{\mu}) = \mathbf{W}'\mathbf{Z} - \mathbf{W}'\boldsymbol{\mu} \le 0$, which obviously holds under case (i) since, by the definition of $\mathbf{Z}, \mathbf{W}'\mathbf{Z} \le 0$. By letting $\mu_k \to \infty$, we obtain a lower bound for $P(|\hat{\mu}_i^{SO} - \mu_i| < c)$, which is itself in the absence of X_k.

Next, consider case (ii). One can then pursue the same argument using Lemma 3.3 and conclude that $P(|\hat{\mu}_i^{SO} - \mu_i| < c)$ is nondecreasing in μ_1. Letting $\mu_1 \to -\infty$, we obtain the lower bound for $P(|\hat{\mu}_i^{SO} - \mu_i| < c)$, which is itself in the absence of X_1.

Arguing inductively and employing Remark 2.1, we establish the theorem for nonstrict inequalities. The strict inequalities can be argued as in Theorem 2.4. □

3.2. *Other orderings.* With the extra assumption that

$$\mathbf{W}_S \text{ has positive elements } \forall S, \tag{3.4}$$

one can establish analytically that the proposed estimators dominate the corresponding unrestricted maximum likelihood estimators for estimating the nodes of a graph.

THEOREM 3.5. *Under* (3.2) *and* (3.4), *the coverage probability of* $\hat{\mu}_i^{SO} \pm c$ *is as large as that of* $X_i \pm c$, *for every* $c > 0$, *as long as* μ_i *is a node. It is strictly larger if* $g(u) > 0$ *for all* $u > 0$.

PROOF. The proof follows along the lines of the proof of Theorem 3.4 except noting that $\mathbf{W}'_S\boldsymbol{\mu} \geq \mu_1\mathbf{W}'_S\mathbf{1}$, for every $\boldsymbol{\mu}$ satisfying the simple tree ordering, iff (3.4) holds. □

For other order restrictions such as the graphs (a) through (f), we can construct estimators as done in Section 2, except using $\mathbf{W}_S = \boldsymbol{\Sigma}_S^{-1}\mathbf{1}$ as the weight vector. The resultant estimators have the same domination property as long as the assumptions of Theorem 3.5 are satisfied. We provide two concrete cases that satisfy Theorem 3.5.

EXAMPLE 3.3 (Continuation of Example 3.1). When $\boldsymbol{\Sigma}$ is an intraclass correlation matrix, both assumptions of Theorem 3.5 are satisfied.

EXAMPLE 3.4 (Continuation of Example 3.2). For estimating the smallest or the largest mean, Theorem 3.5 applies as long as $\mathbf{W}_S$ has positive elements. In addition to the intraclass correlation matrix, this assumption is also satisfied by the *M-matrix.* Note that if $\boldsymbol{\Sigma}$ is an M-matrix, then it follows from Berman and Plemmons [(1979), Theorem 2.4, page 140] that the inverse of every principal submatrix has positive elements.

3.3. *Numerical studies.* Since the intraclass correlation matrix has a special significance in statistics, we therefore study this matrix in greater detail. One can construct better confidence intervals for the parameters satisfying any graph as decribed earlier, by following the scheme suggested in Section 2.2. Extensive simulation studies were performed for simple tree restriction and the umbrella restriction. All our studies indicate that the proposed estimators are at least as good as the unrestricted maximum likelihood estimators and significantly better in other cases. For illustration, in Table 3 we provide results corresponding to the umbrella order restriction. In our study $\boldsymbol{\Sigma} = (1 - \rho)\mathbf{I} + \rho\mathbf{J}$. Performance of the proposed estimators of μ_1 and μ_2 are studied. As in Table 2, there are two branches to the umbrella, one of them contains μ_1, μ_2, μ_3, and the second branch contains $\mu_1, \mu_4, \mu_5, \dots, \mu_{15}$. In the construction of $\hat{\mu}_1$, we guessed in each of the four cases (i)–(iv) that $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_{15}$. Thus in cases (i) and (ii) the guesses are correct, but they

TABLE 3

Coverage probabilities: Performance of the proposed estimators of μ_1 and μ_2 under the umbrella ordering [graph (c)]. Results are based on 2500 simulation runs generated from 15 normal populations, each population having the same variance, 1: case (i) $\mu_1 = -1, \mu_2 = 0, \mu_3 = 0.5$ and, for $4 \le j \le 15, \mu_j = 0.1(10+j)$ and $\rho = -0.05$; case (ii) $\mu_1 = -1, \mu_2 = 0, \mu_3 = 0.5$ and, for $4 \le j \le 15, \mu_j = 0.1(10+j)$ and $\rho = 0.5$; case (iii) $\mu_1 = -1, \mu_2 = 3, \mu_3 = 4$ and, for $4 \le j \le 15, \mu_j = 0.1(10+j)$ and $\rho = -0.05$; case (iv) $\mu_1 = -1, \mu_2 = 3, \mu_3 = 4$ and, for $4 \le j \le 15, \mu_j = 0.1(10+j)$ and $\rho = 0.5$

Case	Nominal level	$\hat{\mu}_1$	$\hat{\mu}_2$	Case	Nominal level	$\hat{\mu}_1$	$\hat{\mu}_2$
(i)	0.68	0.75	0.82	(iii)	0.68	0.68	0.74
	0.90	0.94	0.97		0.90	0.90	0.93
	0.95	0.97	0.99		0.95	0.95	0.97
(ii)	0.68	0.70	0.72	(iv)	0.68	0.68	0.70
	0.90	0.90	0.93		0.90	0.90	0.91
	0.95	0.95	0.97		0.95	0.95	0.96

are wrong in cases (iii) and (iv). Although the performance of the proposed intervals was impressive when ρ was positive, the gains were much more for $\rho \le 0$.

4. Estimation of scale parameters. Suppose, for $i = 1, 2, \dots, k$, X_i is a nonnegative continuous random variable with a positive scale parameter μ_i. We assume in this section that the joint pdf of X_i/μ_i, for $i = 1, 2, \dots, k$, is denoted by $f(\cdot)$. The problem of interest is to estimate the unknown parameter μ_i for some i, when the parameters are subject to order restrictions. There are only a few results available in the literature on this subject. Kushary and Cohen (1989) and Kaur and Singh (1991) studied this problem when $k = 2$ for some special models. In this article we shall obtain some improved confidence intervals for the smallest and the largest parameters when the parameters are subject to order restrictions. We also obtain some universal domination results.

When $n_i X_i/\mu_i, 1 \le i \le k$, are independently distributed according to chi-square distribution with n_i degrees of freedom (df), the isotonic regression estimator, $\hat{\mu}_i^{SO}$ in (2.3), with weights $W_i = n_i$ is the restricted maximum likelihood estimator for μ_i under simple order constraint [Robertson, Wright and Dykstra (1988), page 36].

For a scale parameter μ_i, the standard confidence interval is given by $(X_i/c_+, X_i/c_-)$, where the two positive constants $c_\pm$ are such that $c_- < c_+$. It may seem that one can apply a logarithm transformation and reduce the problem to a location problem and then apply the results of Section 2. This approach, however, cannot be carried out since, after the transformation, the estimator will be quite different. Therefore we take a more direct approach in what follows.

LEMMA 4.1. *Suppose $W_j > 0$, for all j. Then*

$$\frac{\partial}{\partial \mu_k} P\left(c_- < \frac{\hat{\mu}_1^{SO}}{\mu_1} < c_+\right) = \int_{\mathcal{A}} \cdots \int \frac{\mu_1^k}{\mu_k^2} y\left(c_-^k f(c_-\mu_1 \mathbf{Z}) - c_+^k f(c_+\mu_1 \mathbf{Z})\right) \prod_{j=1}^{k-1} \frac{dX_j^*}{\mu_j}, \tag{4.1}$$

where $\mathbf{Z} = (X_1^*/\mu_1, X_2^*/\mu_2, \ldots, X_{k-1}^*/\mu_{k-1}, y/\mu_k)'$,

$$\mathcal{A} = \left\{(X_1^*, X_2^*, \ldots, X_{k-1}^*)\colon \min_{1 \le t \le k-1} A_{\mathbf{X}^*}(1:t) > 1\right\}$$

and

$$y = \left\{\mathbf{W}'_{(1:k)}\mathbf{1} - \mathbf{W}'_{(1:k-1)}\mathbf{X}^*_{(1:k-1)}\right\}_+ / W_k.$$

Here, for any real number a, $\{a\}_+ = \max(a, 0)$.

PROOF. Write

$$P(c_- < \hat{\mu}_1^{SO}/\mu_1 < c_+) = Q(\mu_1 c_-) - Q(\mu_1 c_+),$$

where $Q(c) = P(\hat{\mu}_1^{SO} > c)$. Then note that $Q(c)$ can be written as

$$Q(c) = \int_{\min_{t \ge 1} A_{\mathbf{X}}(1:t) > c} \cdots \int f\left(\frac{X_1}{\mu_1}, \ldots, \frac{X_k}{\mu_k}\right) \prod_{j=1}^{k} \frac{dX_j}{\mu_j}.$$

Performing the change of variables $X_i^* = X_i/c$, we deduce

$$Q(c) = \int_{\min_{t \ge 1} A_{\mathbf{X}^*}(1:t) > 1} \cdots \int c^k f\left(\frac{cX_1^*}{\mu_1}, \ldots, \frac{cX_k^*}{\mu_k}\right) \prod_{j=1}^{k} \frac{dX_j^*}{\mu_j}.$$

Performing another transformation, $t_k = X_k^*/\mu_k$, and observing that $\min_{t \ge 1} A_{\mathbf{X}^*}(1:t) > 1$ is equivalent to $X_k^* > y$ intersecting $\mathcal{A}$, we obtain

$$Q(c) = \int_{\mathcal{A}} \cdots \int_{y/\mu_k}^{\infty} c^k f\left(\frac{cX_1^*}{\mu_1}, \ldots, ct_k\right) dt_k \prod_{j=1}^{k-1} \frac{dX_j^*}{\mu_j}.$$

By differentiating we obtain

$$\frac{\partial}{\partial \mu_k} Q(c) = \int_{\mathcal{A}} \cdots \int \frac{c^k y}{\mu_k^2} f\left(\frac{cX_1^*}{\mu_1}, \ldots, \frac{cy}{\mu_k}\right) \prod_{j=1}^{k-1} \frac{dX_j^*}{\mu_j},$$

which implies (4.1). □

Similarly, we prove the following lemma.

LEMMA 4.2. *Suppose* $W_j > 0$, *for all* j. *Then*

$$\frac{\partial}{\partial \mu_1} P\left(c_- < \frac{\hat{\mu}_k^{SO}}{\mu_k} < c_+\right) = \int_{\mathcal{A}} \cdots \int \frac{\mu_k^k}{\mu_1^2} y\left(c_-^k f(c_- \mu_k \mathbf{Z}) - c_+^k f(c_+ \mu_k \mathbf{Z})\right) \prod_{j=2}^{k} \frac{dX_j^*}{\mu_j},$$

where $\mathbf{Z} = (y/\mu_1, X_2^*/\mu_2, \ldots, X_k^*/\mu_k)'$,

$$\mathcal{A} = \left\{(X_2^*, \ldots, X_k^*) \colon \max_{2 \le s \le k} A_{\mathbf{X}^*}(s:k) < 1\right\}$$

and

$$y = \left\{\mathbf{W}'_{(1:k)} 1 - \mathbf{W}'_{(2:k)} \mathbf{X}^*_{(2:k)}\right\}_+ / W_1.$$

Using these lemmas, one can establish some domination results in the sense of coverage probability. Note that in case of the scale parameters, the size of a confidence interval is usually measured in terms of the ratio of the two endpoints of the interval. An advantage is that this measure is invariant under scale transformations. According to the measure, the following two confidence intervals have the same size.

THEOREM 4.3. (i) *For all* μ *such that* $\mu_1 \le \mu_i$, *for all* i, *the confidence interval* $(\hat{\mu}_1^{SO}/c_+, \hat{\mu}_1^{SO}/c_-)$ *has a coverage probability at least as large as that of* $(X_1/c_+, X_1/c_-)$ *if, for all* $s, 1 < s \le k$, *and* $y > 0$,

$$c_-^s f_s(c_- \mu_1 \mathbf{Z}) \le c_+^s f_s(c_+ \mu_1 \mathbf{Z}), \tag{4.2}$$

where $f_s(\cdot)$ *is the pdf of* $(X_1/\mu_1, \ldots, X_s/\mu_s)$ *and* $\mathbf{Z}$ *and* y *are as defined in Lemma* 4.1 *except* k *is replaced by* s.

(ii) *For all* μ *such that* $\mu_i \le \mu_k$, *for all* i, *the confidence interval* $(\hat{\mu}_k^{SO}/c_+, \hat{\mu}_k^{SO}/c_-)$ *has a coverage probability at least as large as that of* $(X_k/c_+, X_k/c_-)$ *if, for all* $s, 1 \le s < k$ *and* $y > 0$,

$$c_-^s f_s(c_- \mu_k \mathbf{Z}) \le c_+^s f_s(c_+ \mu_k \mathbf{Z}),$$

where $f_s(\cdot)$ *is the pdf of* $(X_s/\mu_s, \ldots, X_n/\mu_n)$ *and* $\mathbf{Z}$ *and* y *are as defined in Lemma* 4.2 *with* 1 *and* 2 *being replaced by* s *and* $s + 1$.

PROOF. We will prove only (i), since the proof of (ii) is similar. Note that (4.2), together with Lemma 4.1, implies that

$$\frac{\partial}{\partial \mu_k} P\left(c_- < \frac{\hat{\mu}_1^{SO}}{\mu_1} < c_+\right) \le 0.$$

Therefore by letting $\mu_k \to \infty$, we obtain a lower bound for $P(c_- < \hat{\mu}_1^{SO}/\mu_1 < c_+)$, which is itself in the absence of X_k. Arguing inductively and dropping all

$X_j, j > 1$, in this manner, we arrive at a lower bound for $P(c_- < \hat{\mu}_1^{SO}/\mu_1 < c_+)$, which is exactly the coverage probability for $(X_1/c_+, X_1/c_-)$. □

As an example for Theorem 4.3, suppose that $u_i = n_i X_i/\mu_i, 1 \le i \le k$, are independently distributed as chi-square random variables with n_i degrees of freedom. Then the pdf of $\mathbf{X}$ with $\mu_i = 1$ is $f(\mathbf{X}) = \Pi_{i=1}^k f_i(X_i)$, where $f_i(X_i)$ is proportional to

$$X_i^{n_i/2-1} \exp(-n_i X_i/2), \qquad X_i > 0. \tag{4.3}$$

THEOREM 4.4. (i) *If the constants c_- and c_+ are chosen to satisfy*

$$(c_-/c_+) \exp(c_+ - c_-) \le 1, \tag{4.4}$$

then the coverage probability of the confidence interval $(\hat{\mu}_1^{SO}/c_+, \hat{\mu}_1^{SO}/c_-)$ is larger than that of $(X_1/c_+, X_1/c_-)$, for all $\mu_i \ge \mu_1, 1 \le i \le k$.

(ii) *If c_- and c_+ satisfy*

$$(c_-/c_+) \exp(c_+ - c_-) \ge 1, \tag{4.5}$$

then the coverage probability of the confidence interval $(\hat{\mu}_k^{SO}/c_+, \hat{\mu}_k^{SO}/c_-)$ is larger than that of $(X_k/c_+, X_k/c_-)$, for all $\mu_k \ge \mu_i, 1 \le i \le k$.

REMARK 4.1. If nX/μ is distributed as a chi-square random variable with n df, then the equal-tailed confidence interval $(X/c_+, X/c_-)$ uses the constants $c_\pm$ which satisfy (4.4). This can be proved for $n = 2$ and when n is sufficiently large. Exact numerical computations show that for $n = 5, 10, 15, 20$ and for $c_\pm$ such that the coverage probabilities of $(X/c_+, X/c_-)$ are 0.55, 0.9 and 0.95, the inequality (4.4) holds. On the other hand it can be analytically shown, for all degrees of freedom, that the shortest-width confidence interval $(X/c_+, X/c_-)$ uses constants which satisfy (4.5).

PROOF OF THEOREM 4.4. We will prove (i), and (ii) will follow similarly. Using (4.3), we note that $f_s(c\mu_1 \mathbf{Z})$ in (4.2) is proportional to

$$\left(\frac{c\mu_1 y}{\mu_s}\right)^{n_s/2-1} \exp\left(-\frac{cS(\mathbf{X}^*)}{2}\right) \prod_{i=1}^{s-1} \left(\frac{c\mu_1 X_i^*}{\mu_i}\right)^{n_i/2-1}$$

where

$$S(\mathbf{X}^*) = \sum_{i=1}^{s-1} \frac{n_i X_i^* \mu_1}{\mu_i} + \frac{n_s y \mu_1}{\mu_s}.$$

Hence (4.2) is equivalent to

$$\left(\frac{c_-}{c_+}\right)^{\Sigma_{i=1}^s n_i/2} \exp\left((c_+ - c_-)\frac{S(\mathbf{X}^*)}{2}\right) \le 1. \tag{4.6}$$

Furthermore, since $\mu_1 \leq \mu_i$ and $y = (\Sigma_{i=1}^{s} n_i - \Sigma_{i=1}^{s-1} n_i X_i^*)_+ / n_s > 0$, one can establish that $S(\mathbf{X}^*) \leq \Sigma_{i=1}^{s} n_i$. Hence (4.6) is true if

$$\left(\frac{c_-}{c_+}\right)^{\Sigma_{i=1}^{s} n_i/2} \exp\left((c_+ - c_-)\sum_{i=1}^{s} \frac{n_i}{2}\right) \leq 1,$$

which follows from (4.4). We have therefore established that the coverage probability of $(\hat{\mu}_1^{SO}/c_+, \hat{\mu}_1^{SO}/c_-)$ is at least as large as that of $(X_1/c_+, X_1/c_-)$. The strict inequalities can easily be established. □

Although a direct application of the preceding theorems will not yield the result, all numerical studies (over 100 different combinations of μ_i's, n_i's, $c_\pm$ and k) indicate that the equal-tailed confidence interval $(X_k/c_+, X_k/c_-)$ has a smaller coverage probability than the interval $(\hat{\mu}_k^{SO}/c_+, \hat{\mu}_k^{SO}/c_-)$.

If one considers reverse simple tree problems, that is, $\mu_j \leq \mu_k$ for all j, then the restricted maximum likelihood estimator does not perform well according to the following theorem, which can be established along the lines of Theorem 2.3. The question then is: what will be an improved confidence interval for μ_k? Theorem 4.4(ii) shows that in fact the confidence intervals therein will be better than the standard intervals even for the reverse simple tree situation. The same domination is true with any other restriction as long as μ_1 is the smallest and μ_k the largest parameter. Because of this, we can construct better estimators for the graphs (a) through (f) as in Section 2.

THEOREM 4.5. *If, for all i, we suppose $\mu_i \leq \mu_k$, then as $k \to \infty$, any fixed-length confidence interval, centered at the restricted maximum likelihood estimator of μ_k, will have a coverage probability converging to* 0.

Theorem 4.4 does not yield the stochastic domination results. However, as demonstrated below, a simple application of Lemma 4.1 to a different interval will yield the result.

THEOREM 4.6. *Under* (4.3), *the following hold*:

(i) *$\hat{\mu}_1^{SO}$ universally dominates X_1, that is,*

$$P(|\hat{\mu}_1^{SO} - \mu_1| < c) > P(|X_1 - \mu_1| < c), \quad \text{for all } c > 0, \tag{4.7}$$

whenever $\mu_i \geq \mu_1$;

(ii) *$\hat{\mu}_k^{SO}$ fails to universally dominate X_k; in fact for every μ_i there exists a positive constant c such that*

$$P(|\hat{\mu}_k^{SO} - \mu_k| < c) < P(|X_k - \mu_k| < c).$$

PROOF. (i) Consider $P(|\delta - \mu_1| < c)$, where δ represents either $\hat{\mu}_1^{SO}$ or X_1. Then we can rewrite this probability as $P(c_- < \delta/\mu_1 < c_+)$, where $c_\pm = 1 \pm c/\mu_1$.

There are two possibilities: (a) $c_- < 0$ and (b) $c_- \geq 0$. When (a) is true the problem reduces to showing

$$P(0 \leq \hat{\mu}_1^{SO}/\mu_1 < c_+) > P(0 < X_1/\mu_1 < c_+), \quad \text{for all } c.$$

However, this inequality is obvious since $\hat{\mu}_1^{SO} \leq X_1$, where strict inequality holds with positive probability.

On the other hand (b) implies that $d \leq 1$, where $d = c/\mu_1$. Hence $0 \leq d \leq 1$. We are done if we are able to verify the condition (4.4) stated in Theorem 4.4. Note that condition (4.4) is equivalent to showing

$$\frac{(1-d)\exp(2d)}{1+d} \leq 1. \tag{4.8}$$

When $d = 0$, the left-hand side of (4.8) is 1. Thus to establish (4.8) it will be enough to show that the left-hand side is a decreasing function of d. The derivative of the left-hand side with respect to d yields

$$\frac{-2d^2\exp(2d)}{(1+d)^2},$$

which is nonpositive. The result follows now by appealing to Theorem 4.4.

(ii) This part of the theorem is not restricted to gamma distributed random variables but it is true for general scale families. To prove (ii) it suffices to find a constant c such that $P(|\hat{\mu}_k^{SO} - \mu_k| < c) < P(|X_k - \mu_k| < c)$. For every μ_i, there exists a constant c such that $c_- = 1 - c/\mu_k < 0$; then, since $X_k \leq \hat{\mu}_k^{SO}$, where strict inequality holds with a positive probability,

$$P(0 < \hat{\mu}_k^{SO}/\mu_k < c_+) < P(0 < X_k/\mu_k < c_+), \quad \text{where } c_+ = 1 + c/\mu_k.$$

Hence the theorem follows. □

Following arguments similar to those in the above theorem, we prove the following theorem regarding the simple tree order restriction.

THEOREM 4.7. *If $\mu_1 \leq \mu_i$ for all i, then the isotonic regression estimator $\hat{\mu}_i^{ST}$ of μ_i fails to universally dominate X_i.*

REMARK 4.2. This theorem is true for all scale families.

5. Generalizations and open problems. In this paper, derivative formulas for the coverage probability of confidence intervals centered at the isotonic regression type estimators are given. Using these formulas, we strengthen the results of the previous research and establish many new theorems. In particular we are able to develop new improved confidence intervals for different types of constraints on (a) location parameters, both when the underlying random variables are uncorrelated and when they are correlated, and (b) scale parameters.

In the case of the location problems, we focused on the case when the covariance matrix is known. If the covariance matrix is known up to a constant multiplier σ^2 and if there is an independent estimator $\hat{\sigma}^2$, the standard estimator which is usually of the form $X_i \pm c\hat{\sigma}$ can be improved by $\hat{\mu}_i \pm c\hat{\sigma}$ as long as $\hat{\mu}_i$ stochastically dominates X_i. This is based on a simple conditioning argument.

Our framework is also applicable to a linear model with restrictions involving linear combinations such as $\mathbf{a}_i'\mu + b_i$ as long as we are interested in estimating one of the linear combinations. Such restrictions are often reasonable since the mean responses are of these forms. By renaming $\mathbf{a}_i'\mu + b_i$ as μ_i^* and, working with the least squares estimator X_i^*, we can transform the problem back to our framework with covariance matrix of $\mathbf{X}^*$ not necessarily diagonal, a setting studied in Section 3.

5.1. *Open problems.* Although partial results are available in Section 3, we are unable to deal with an arbitrary covariance matrix. A technically equivalent question is how to deal with estimation of μ under an arbitrary linear restriction $\mathbf{A}\mu \leq \mathbf{b}$, where $\mathbf{A}$ is a known rectangular matrix and $\mathbf{b}$ is a known column vector, with "$\leq$" being a componentwise inequality. It will also be important to study the situation when the covariance matrix is completely unknown or has a variance component structure with the variance components unknown.

Concerning the estimation of the scale parameters, there are several questions that remain unanswered at this point. The performance of the isotonic regression estimators (and their modifications) is not completely understood. For example, when the parameters are subject to simple order restriction we are unable to study the performance of the isotonic regression estimator of the i-th parameter $\mu_i, 1 < i < k$.

The intervals studied here have constant length. It would be nice to construct shorter-length confidence intervals with the coverage probabilities above some nominal level.

It will also be interesting to generalize the results of this paper to the discrete distributions.

In solving these problems, we feel that the use of the derivative formulas, such as the ones developed in this paper, are essential.

Acknowledgments. We thank Professors R. Kelly, C. I. C. Lee, F. T. Wright and H. Mukerjee for their valuable discussions. We also thank the two referees for their careful reading of the original manuscript and for detecting an error in Section 3, which along with other suggestions have led to a substantial improvement.

REFERENCES

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference under Order Restrictions: The Theory and Applications of Isotonic Regression*. Wiley, New York.

BERMAN, A. and PLEMMONS, R. J. (1979). *Non-Negative Matrices in Mathematical Sciences*. Academic, New York.

HWANG, J. T. (1985). Universal domination and stochastic domination. *Ann. Statist.* **13** 295–314.

HWANG, J. T. (1986). Universal domination and stochastic domination. In *Encyclopedia of Statistical Sciences* **8** 781–784.

KAUR, A. and SINGH, H. (1991). On the estimation of ordered mean of two exponential populations. *Ann. Inst. Statist. Math.* **47** 347–356.

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

KELLY, R. (1990). Distributional properties of isotonic regression estimators under simple tree order restriction. Technical Report 93, Dept. Statistics, Pennsylvania State Univ.

KUSHARY, D. and COHEN, A. (1989). Estimating ordered location and scale parameters. *Statist. Decisions* **7** 201–213.

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

LEE, C. I. C. (1983). The min-max algorithm and isotonic regression. *Ann. Statist.* **11** 467–477.

LEE, C. I. C. (1988). The quadratic loss of order restricted estimators for several treatment means and a control mean. *Ann. Statist.* **16** 751–758.

ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, New York.

TONG, Y. L. (1990). *The Multivariate Normal Distribution*. Springer, New York.

DEPARTMENT OF MATHEMATICS
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853

DIVISION OF STATISTICS
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF VIRGINIA
CHARLOTTESVILLE, VIRGINIA 22903