The Annals of Statistics
1980, Vol. 8, No. 3, 645–651

ALGORITHMS IN ORDER RESTRICTED STATISTICAL INFERENCE AND THE CAUCHY MEAN VALUE PROPERTY[1]

BY TIM ROBERTSON AND F. T. WRIGHT

University of Iowa, University of Missouri-Rolla

Most algorithms in order restricted statistical inference express the estimates in terms of certain summary statistics computed from pooled samples. These algorithms may or may not yield optimal estimates depending on whether or not the Cauchy mean value property holds strictly for the summary statistics. In this paper a minimum lower sets algorithm, which holds generally, is described and used to prove the optimality of estimates described by a max-min formula.

1. Introduction. A number of estimates studied in order restricted statistical inference can be expressed using "max-min" formulas. These formulas are, perhaps, the most succinct way of describing these estimates and are used extensively in consistency proofs. However, these formulas are not very handy for the actual computation of the estimates. Moreover, optimality is generally argued by induction and other algorithms have proven to be more useful here. If the ordering is not linear a minimum lower sets algorithm (cf. Barlow et al. (1972)) has often been used for proving optimality. Such an algorithm was incorrectly stated in Theorem 2.4 of Robertson and Wright (1973) for a least absolute deviations problem. The proof given there implicitly uses the incorrect assumption that the median is a strict Cauchy mean value function. In this note we prove the optimality of estimates given by a max-min formula by developing an alternative (and much more complicated) minimum lower sets algorithm which is valid in a general setting, which includes the least absolute deviations problem and several other problems discussed in the literature.

2. Algorithms. Assume that we have samples from k distributions; that these distributions are indexed by what we shall refer to as observation points, $s_1, s_2, \cdots, s_k$, and that we wish to estimate a real valued parameter from each distribution, say $\theta(s_i)$, $i = 1, 2 \cdots, k$. We also assume that it is known a priori that each of the parameters belongs to a nondegenerate interval I and these parameters satisfy a certain order restriction such as, $\theta(s_1) \leqslant \theta(s_2) \leqslant \cdots \leqslant \theta(s_k)$. Thus, we are interested in estimates which satisfy these conditions.

We consider order restrictions that can be specified in terms of a partial order. In particular, with $\ll$ a partial order on $S = \{s_1, s_2, \cdots, s_k\}$, we consider the order

Received February 1976; revised December 1978.

[1]This research was supported by the National Science Foundation under Grants MCS75-23576 and MCS78-01514.

AMS 1970 *subject classifications.* Primary 62G05; secondary 62F10.

Key words and phrases. Optimality, L_p problems, Cauchy mean value function, computation algorithms, isotonic regression.

restrictions $\theta(s_i) \leqslant \theta(s_j)$ if $s_i \ll s_j$. (Such a function, $\theta(\cdot)$ on S, is said to be isotone with respect to $\ll$.)

For $i = 1, 2, \cdots, k$, we denote the sample items at s_i by x_{ij} with $j = 1, 2, \cdots, n_i$. Thus, for each nonempty subset A of S, let $M(A)$ be a statistic based on the collection of observations x_{ij} with $j = 1, 2, \cdots, n_i$ and i such that $s_i \in A$. For instance, $M(A)$ could be the mean, midrange, median or some other percentile of this collection of observations. If for each A, $M(A)$ is the median of the appropriate collection we refer to M as the median function. The description of these estimates also involves the collection of upper layers of S. A subset L of S is called an upper layer if and only if $s_j \in L$ whenever $S_i \in L$ and $s_i \ll s_j$. We use the symbol L with or without subscripts, etc., to denote an upper layer and $\mathfrak{L}$ to denote the collection of upper layers.

It is easy to see that

$$\bar{\theta}(s_i) = \max_{\{L:\, s_i \in L\}} \min_{\{L':\, s_i \notin L'\}} M(L - L') \tag{2.1}$$

is an isotone function on S and hence could be used as an estimate of θ. In many situations, if the estimates $M(A)$ are chosen to minimize certain objective functions then $\bar{\theta}$ minimizes a closely related objective function subject to the isotonic constraint. Typically, this is established using algorithms other than (2.1). A minimum lower sets algorithm (MLSA) is discussed in Barlow, et al., (1972) and it is not difficult to see that it provides the same estimates as $\bar{\theta}$ provided M is a strict Cauchy mean value function, which we now define. (A generalized version of the MLSA will be described later.)

DEFINITION. A function M defined on the nonempty subsets of S is said to be a Cauchy mean value function provided $M(A + B)$ is between $M(A)$ and $M(B)$ (i.e., $M(A) \wedge M(B) \leqslant M(A + B) \leqslant M(A) \vee M(B)$) whenever A and B are nonempty, disjoint subsets of S. If, in addition, $M(A + B)$ is strictly between $M(A)$ and $M(B)$ for such A and B with $M(A) \neq M(B)$ then we say M is a strict Cauchy mean value function.

In Theorem 2.4 of Robertson and Wright (1973) it was claimed that the MLSA provides the same estimates as the max-min formula if M is the median function. It is not difficult to construct examples to show that this is not correct and, in fact, one can construct examples for which the algorithm can not be implemented. This is due to the fact that the median is not a strict Cauchy mean value function. They have noted that if the median is defined by averaging the two middle ordered observations for even sample sizes then it is a Cauchy mean value function.

We now describe the modifications needed for this algorithm in the case of general Cauchy mean value functions. Since a lower layer is the complement of an upper layer the algorithm could be described in terms of either. We have found upper layers more convenient to work with, but will still refer to the algorithm as the MLSA and, in fact, we will refer to the modified version as the minimum lower sets algorithm with the understanding that the simplified version (the one discussed in Barlow, et al., (1972)) should be used if $M(\cdot)$ is a strict Cauchy mean value function.

Minimum lower set algorithm. Let $L_1 = S$ and let $L_2^{(1)}, L_2^{(2)}, \cdots, L_2^{(b(2))}$ be the upper layers which are proper subsets of L_1 and which minimize $M(L_1 - L)$. It will be shown later that for at least one i, $L_2^{(i)}$ satisfies

$$M(L_2^{(j)} - L_2^{(i)}) \leqslant M(L_1 - L_2^{(j)}) \text{ for all } j \text{ with } L_2^{(i)} \subsetneqq L_2^{(j)}.$$

Set L_2 equal to the smallest (intersection) of the $L_2^{(i)}$ which satisfy this inequality. If $L_2 \neq \phi$ let $L_3^{(1)}, L_3^{(2)}, \cdots, L_3^{(b(3))}$ be the upper layers which are proper subsets of L_2 and minimize $M(L_2 - L)$. Again there is at least one i for which $L_3^{(i)}$ satisfies $M(L_3^{(j)} - L_3^{(i)}) \leqslant M(L_2 - L_3^{(j)})$ for all j with $L_3^{(i)} \subsetneqq L_3^{(j)}$ and we set L_3 equal to the smallest (intersection) of the $L_3^{(i)}$ which satisfies this inequality. Continuing this process, we obtain $S = L_1 \supset L_2 \supset \cdots \supset L_{H+1} = \phi$ with $L_{\alpha+1}$ the smallest of the upper layers $L_{\alpha+1}^{(i)}$, which are proper subsets of L_α, which minimize $M(L_\alpha - L)$ and satisfy

(2.2)

$$M(L_{\alpha+1}^{(j)} - L_{\alpha+1}^{(i)}) \leqslant M(L_\alpha - L_{\alpha+1}^{(j)}) \text{ for all } 1 \leqslant j \leqslant b(\alpha+1) \text{ with } L_{\alpha+1}^{(i)} \subsetneqq L_{\alpha+1}^{(j)}$$

where $L_{\alpha+1}^{(1)}, \cdots, L_{\alpha+1}^{(b(\alpha+1))}$ are the upper layers that are proper subsets of L_α and minimize $M(L_\alpha - L)$.

THEOREM 2.1. *At each step in the* MLSA *there exists an upper layer $L_{\alpha+1}$ which is a proper subset of L_α, which minimizes $M(L_\alpha - L)$ and satisfies* (2.2). *If $s_j \in L_\alpha - L_{\alpha+1}$, then*

$$\bar{\theta}(s_j) = M(L_\alpha - L_{\alpha+1}).$$

The L_α also satisfy

(2.3a) $$M((L_\alpha - L_{\alpha+1}) \cap L) \leqslant M((L_\alpha - L_{\alpha+1}) - L) \textit{ if } (L_\alpha - L_{\alpha+1}) \cap L \neq \phi \textit{ and } (L_\alpha - L_{\alpha+1}) - L \neq \phi$$

and

(2.3b) $$M(L_\alpha - L_{\alpha+1}) < M(L_{\alpha+1} - L) \textit{ if } L_{\alpha+1} - L \neq \phi.$$

The latter implies that $M(L_1 - L_2) < M(L_2 - L_3) < \cdots < M(L_H - L_{H+1})$ and that the $L_\alpha - L_{\alpha+1}$ are the level sets.

It should be noted that if $M(\cdot)$ is a strict Cauchy mean value function then (2.2) holds for all $L_{\alpha+1}^{(i)}$. For if not there is a $L_{\alpha+1}^{(i)} \subsetneqq L_{\alpha+1}^{(j)}$ with $M(L_{\alpha+1}^{(j)} - L_{\alpha+1}^{(i)}) > M(L_\alpha - L_{\alpha+1}^{(j)})$, which implies that $M(L_\alpha - L_{\alpha+1}^{(j)}) < M(L_\alpha - L_{\alpha+1}^{(i)}) < M(L_{\alpha+1}^{(j)} - L_{\alpha+1}^{(i)})$. This contradicts the assumption that $L_{\alpha+1}^{(i)}$ and $L_{\alpha+1}^{(j)}$ both minimize $M(L_\alpha - L)$. Thus this MLSA reduces to the one given in Barlow, et al., (1972) if $M(\cdot)$ is a strict Cauchy mean value function.

If the order is linear, say $\theta(s_1) \leqslant \theta(s_2) \leqslant \cdots \leqslant \theta(s_k)$ then these estimates are most easily computed using a pool adjacent violators algorithm (PAVA) (c.f. Barlow, et al., (1972)). The fact that the PAVA gives the same estimates as (2.1) is proved in Robertson and Waltman (1968) assuming M is the median function. It is easily seen that the proof only requires that M is a Cauchy mean value function; however, it contains a slight error and needs to be modified. In the second full paragraph on page 1033, it is claimed that $M(R, i) \geqslant M(R, i + 1)$ implies that $M(R, i + 1) \geqslant M_{i+1}$ ($M(R,S)$ is the median of the sample obtained by pooling samples $R, R + 1, \cdots, S$ and M_i is the median of the ith sample). This is true if $M(\cdot)$ is a strict Cauchy mean value function, but the median is not such a function. We now give the modifications needed in that paragraph. The $\bar{\theta}_j$ should be defined as there and then note that if $R < i$ and if

$$\max_{R \leqslant S \leqslant H+1} M(R, S) \neq \max_{R \leqslant S \leqslant H+1;\, S \neq i} M(R, S)$$

then $M(R, i) > M(R, i + 1)$. It follows from the Cauchy mean value property that $M(R, i) > M(R, i + 1) \geqslant M_{i+1} > M_i$. Using the fact that $M(R, i) > M_i$ and the Cauchy mean value property again, we conclude that $M(R, i - 1) \geqslant M(R, i)$. This contradiction establishes the next claim in the proof given for Theorem 3.1 and the remainder of the proof is correct as it stands.

Theorem 2.7 of Robertson and Wright (1973) states that $\bar{\theta}$ minimizes $\sum_{i=1}^{k} \sum_{j=1}^{n_i} |x_{ij} - \phi(s_i)|$ subject to the restriction that ϕ is isotone on S provided M is the median function. While the result is correct, the proof needs to be modified since it is based on the version of the MLSA which does not apply to medians. We now state a general result which includes this l_1 problem as a special case.

Suppose that for each A, a nonempty subset of S, there is an objective function $D(\phi|A)$ defined for ϕ: $S \to R$. For each real number x, let ϕ_x be constant on S with $\phi_x(s_i) = x$ for $i = 1, 2, \cdots, k$ and let $d(x|A) = D(\phi_x|A)$. We also suppose that $M(A)$ minimizes $d(x|A)$. In the l_1 problem

$$D(\phi|A) = \sum_{\{i:\, s_i \in A\}} \sum_{j=1}^{n_i} |x_{ij} - \phi(s_i)|,$$

$$d(x|A) = \sum_{\{i:\, s_i \in A\}} \sum_{j=1}^{n_i} |x_{ij} - x|,$$

and M is the median function. We make the following assumptions:

A-1. If $A \subset S$, then $M(A) \in I$ and $d(\cdot|A)$ is nonincreasing on $I \cap (-\infty, M(A)]$ and nondecreasing on $I \cap [M(A), \infty)$.

A-2. If ϕ and σ are functions on S such that $\phi(s_i) = \sigma(s_i)$ for each $s_i \in A$, then $D(\phi|A) = D(\sigma|A)$.

A-3. If $A \subset B \subset S$, if $\phi(s_i) = \sigma(s_i)$ for $s_i \in B - A$, and if $D(\phi|A) \leqslant D(\sigma|A)$, then $D(\phi|B) \leqslant D(\sigma|B)$.

For each $L \in \mathfrak{L}$ let $\mathcal{V}(L)$ denote the collection of all upper layers for which $L - L'$ is nonempty.

THEOREM 2.2. *Suppose $D(\cdot|A)$, $d(\cdot|A)$ and $M(\cdot)$ satisfy conditions* A-1, A-2 *and* A-3 *and that $M(\cdot)$ is a Cauchy mean value function. Then for $i = 1, 2, \cdots, k$*

$$\bar{\theta}(s_i) = \max_{\{L:\, s_i \in L\}} \min_{L' \in \mathcal{V}(L)} M(L - L')$$
$$= \min_{\{L':\, s_i \notin L'\}} \max_{\{L:\, s_i \in L\}} M(L - L')$$

and $\bar{\theta}$ minimizes $D(\phi) = D(\phi|S)$ subject to the restriction that ϕ is isotone.

Barlow, et al., (1972) discuss the l_2 problem and the ordered proportions problem giving the original references. Barlow and Ubhaya (1971) and Ubhaya (1974 a, b) consider the l_p problems $1 < p \leqslant \infty$. Robertson and Wright (1975) treat the l_1 problem as well as some other ordered estimation problems.

3. Proofs (outlines). We first prove Theorem 2.1 by showing that the MLSA may be used to obtain the representation given in Equation 2.1. Then Theorem 2.2 is proved by showing that the function obtained from the minimum lower sets algorithm is a minimizing function and that these three representations for $\bar{\theta}(\cdot)$ are equivalent.

PROOF OF THEOREM 2.1. We suppose that the first α upper layers, $L_1, L_2, \cdots, L_\alpha$ have been determined and show how $L_{\alpha+1}$ is obtained. Since S is finite, $L_{\alpha+1}^{(i)}$ satisfies (2.2) for some i and if $L_{\alpha+1}^{(i)}$ satisfies (2.2) it can be shown by contradiction that it also satisfies the following strengthened version of (2.2):
(3.1)

$$M(L_\alpha - (L \cup L_{\alpha+1}^{(i)})) \geqslant M(L - L_{\alpha+1}^{(i)}) \text{ for } L \subset L_\alpha \text{ with } L_\alpha - (L \cup L_{\alpha+1}^{(i)}) \neq \phi \text{ and } L - L_{\alpha+1}^{(i)} \neq \phi.$$

Because the algorithm chooses $L_{\alpha+1}$ to satisfy (2.2), it will also satisfy (3.1) and replacing L by $L \cap L_\alpha$, (3.1) yields (2.3a).

The first part of this proof is completed by showing that if $L_{\alpha+1}^{(r)}$ and $L_{\alpha+1}^{(s)}$ satisfy (2.2) then $L_{\alpha+1}^{(r)} \cap L_{\alpha+1}^{(s)}$ is one of the minimizing upper layers and satisfies (2.2). Clearly $L_\alpha - (L_{\alpha+1}^{(r)} \cap L_{\alpha+1}^{(s)}) = (L_\alpha - L_{\alpha+1}^{(s)}) + (L_{\alpha+1}^{(s)} - L_{\alpha+1}^{(r)})$, where $+$ denotes a disjoint union. Either $L_{\alpha+1}^{(s)} - L_{\alpha+1}^{(r)} = \phi$, $L_\alpha - (L_{\alpha+1}^{(r)} \cup L_{\alpha+1}^{(s)}) = \phi$ or both are nonempty. In all three cases $M(L_\alpha - (L_{\alpha+1}^{(r)} \cap L_{\alpha+1}^{(s)})) = M(L_\alpha - L_{\alpha+1}^{(s)})$. The first two cases are obvious and for the third case use the Cauchy mean value property and (3.1) with $L = L_{\alpha+1}^{(s)}$. Next it is shown that $L_{\alpha+1}^{(r)} \cap L_{\alpha+1}^{(s)}$ satisfies (2.2). For $L_{\alpha+1}^{(j)} \supsetneq L_{\alpha+1}^{(r)} \cap L_{\alpha+1}^{(s)}$, we need to show that $M(L_{\alpha+1}^{(j)} - (L_{\alpha+1}^{(r)} \cap L_{\alpha+1}^{(s)})) \leqslant M(L_\alpha - L_{\alpha+1}^{(j)})$. Since

$$L_{\alpha+1}^{(j)} - (L_{\alpha+1}^{(r)} \cap L_{\alpha+1}^{(s)}) = ((L_{\alpha+1}^{(s)} - L_{\alpha+1}^{(r)}) \cap L_{\alpha+1}^{(j)}) + (L_{\alpha+1}^{(j)} - L_{\alpha+1}^{(s)}), \tag{3.2}$$

we show that each term on the right-hand side of (3.2) is empty or has M value less than or equal to $M(L_\alpha - L_{\alpha+1}^{(j)})$. For $L_{\alpha+1}^{(j)} - L_{\alpha+1}^{(s)} \neq \phi$ consider separately the two possibilities, $L_\alpha - (L_{\alpha+1}^{(s)} \cup L_{\alpha+1}^{(j)})$ is empty or nonempty. In both cases, $M(L_{\alpha+1}^{(j)} - L_{\alpha+1}^{(s)}) \leqslant M(L_\alpha - L_{\alpha+1}^{(j)})$. This is obvious in the former case and in the latter case, first apply (3.1) to $L_{\alpha+1}^{(s)}$ with $L = L_{\alpha+1}^{(j)}$ and then apply the averaging property. The other term in (3.2) is treated similarly. So the L_α can be chosen as specified in the algorithm and they satisfy (2.3a).

We next show that they satisfy (2.3b). If not, there is an integer $\alpha \in [1, H]$ and an upper layer L, a proper subset of $L_{\alpha+1}$, with

$$M(L_\alpha - L_{\alpha+1}) \geqslant M(L_{\alpha+1} - L). \tag{3.3}$$

It follows from the Cauchy mean value property that $L = L_{\alpha+1}^{(j_0)}$ for some j_0 and $L_{\alpha+1}^{(j_0)}$ does not satisfy (2.2). This means there is a $L_{\alpha+1}^{(j)} \supsetneqq L_{\alpha+1}^{(j_0)}$ with

$$M(L_{\alpha+1}^{(j)} - L_{\alpha+1}^{(j_0)}) > M(L_\alpha - L_{\alpha+1}^{(j)}). \tag{3.4}$$

Considering separately the two cases $L_{\alpha+1}^{(j)} - L_{\alpha+1}$ is empty or nonempty, we see that $L^* = L_{\alpha+1} \cap L_{\alpha+1}^{(j)}$ is one of the $L_{\alpha+1}^{(i)}$, say $L_{\alpha+1}^{(j_1)}$. In the latter case, use $L_\alpha - L_{\alpha+1} = (L_\alpha - (L_{\alpha+1}^{(j)} \cup L_{\alpha+1})) + (L_{\alpha+1}^{(j)} - L_{\alpha+1})$ and for $L_\alpha - (L_{\alpha+1}^{(j)} \cup L_{\alpha+1}) \neq \phi$ apply (3.1) with $L = L_{\alpha+1}^{(j)}$ and $L_{\alpha+1}^{(i)} = L_{\alpha+1}$ and the averaging property. For future reference note that it has been shown that

$$M(L_{\alpha+1}^{(j)} - L_{\alpha+1}) \leqslant M(L_\alpha - L_{\alpha+1}). \tag{3.5}$$

If $L_{\alpha+1}^{(j_1)} - L_{\alpha+1}^{(j_0)} = \phi$, then $L_{\alpha+1}^{(j)} - L_{\alpha+1}^{(j_0)} = L_{\alpha+1}^{(j)} - L_{\alpha+1}$ and so $M(L_{\alpha+1}^{(j)} - L_{\alpha+1}) = M(L_{\alpha+1}^{(j)} - L_{\alpha+1}^{(j_0)}) > M(L_\alpha - L_{\alpha+1}^{(j)})$. This contradicts (3.5), and hence $L_{\alpha+1}^{(j_0)}$ is a proper subset of $L_{\alpha+1}^{(j_1)}$.

Next we show that (3.4) holds with $L_{\alpha+1}^{(j)}$ replaced by $L_{\alpha+1}^{(j_1)}$. This can be seen by considering separately the two cases $L_{\alpha+1}^{(j)} - L_{\alpha+1}$ is empty or nonempty. In the latter case, use $M(L_{\alpha+1}^{(j)} - L_{\alpha+1}^{(j_0)}) \leqslant \max(M(L_{\alpha+1}^{(j_1)} - L_{\alpha+1}^{(j_0)}), M(L_{\alpha+1}^{(j)} - L_{\alpha+1}))$, (3.4) and (3.5).

So $L_{\alpha+1}^{(j_1)} \supsetneqq L_{\alpha+1}^{(j_0)}$ and either $L_{\alpha+1}^{(j_1)} = L_{\alpha+1}$ or it does not satisfy (2.2). Continuing this process, it must stop after a finite number of iterations with $L_{\alpha+1}^{(j_0)} \subsetneqq L_{\alpha+1}^{(j_1)} \subsetneqq \cdots \subsetneqq L_{\alpha+1}^{(j_\beta)} = L_{\alpha+1}$ and $M(L_{\alpha+1}^{(j_{i+1})} - L_{\alpha+1}^{(j_i)}) > M(L_\alpha - L_{\alpha+1})$ for $i = 0, 1, \cdots, \beta - 1$. Using the average property we have

$$M(L_{\alpha+1} - L_{\alpha+1}^{(j_0)}) > M(L_\alpha - L_{\alpha+1}),$$

which contradicts (3.3). So the L_α satisfy (2.3b).

The last part of the proof of Theorem 2.1 is to show that this algorithm provides the same estimate as Equation (2.1). Fix $s_j \in L_\alpha - L_{\alpha+1}$ and the proof is the same as the one given for Theorem 2.4 in Robertson and Wright (1973). Inequality (2.3a) is used to establish (2.4) given there (those h's for which $(L \cap L_h) - L_{h+1} = \phi$ may be ignored). Also (2.3b) is used to establish (2.8).

PROOF OF THEOREM 2.2. To show that $\bar{\theta}$ is optimum one only needs to modify the proof of Theorem 2.7 of Robertson and Wright (1973). We do not use their Lemma 2.6 and we employ the general MLSA presented here. The first part of their proof shows that if $H = 1$ and g is any isotone function on S then there is a constant function on S, say, $\hat{g}$, with $D(\hat{g}) \leqslant D(g)$. In that proof it is claimed that

$M(L_1) \leqslant M(L_1 - L_2')$ and the Cauchy mean value property imply that $M(L_2') \leqslant M(L_1) \leqslant M(L_1 - L_2')$. This need not be true if $M(L_1) = M(L_1 - L_2')$ and M is not a strict Cauchy mean value function, however if $M(L_1) = M(L_1 - L_2)$ then $L_2' = L_2^{(j)}$ for some j and applying (2.2) with $L_{\alpha+1}^{(i)} = \phi$, we have $M(L_2') \leqslant M(L_1 - L_2')$. So the desired inequalities follow.

Next an induction on H is performed. As in their proof, the observation points in $L_1 - L_2$ and those in L_2 are considered separately. The proof that $\bar{\theta}$ is optimum is completed by applying the case $H = 1$, the induction hypothesis, and the definition of the MLSA.

The proof of Theorem 2.2 is completed by showing that the three representations given there are equal. Clearly for $s_i \in L_\alpha - L_{\alpha+1}$,

$$\min_{\{L':\, s_i \notin L'\}} \max_{\{L:\, s_i \in L\}} M(L - L') \geqslant \max_{\{L:\, s_i \in L\}} \min_{\{L':\, s_i \notin L'\}} M(L - L')$$

$$\geqslant \max_{\{L:\, s_i \in L\}} \min_{L' \in \mathcal{V}(L)} M(L - L') = M(L_\alpha - L_{\alpha+1}).$$

So we need only show that the first expression above is less than or equal to $M(L_\alpha - L_{\alpha+1})$, but this is established by the argument beginning at the last line of page 425 of Robertson and Wright (1973).

REFERENCES

[1] BARLOW, R. E., BARTHOLOMEW, D. J., BREMMER, J. M. and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions.* Wiley, New York.

[2] BARLOW, R. E. and UBHAYA, V. A. (1971). Isotonic approximation. In *Proceedings of the Symposium on Optimizing Methods in Statistics.* Academic Press, New York.

[3] ROBERTSON, TIM and WALTMAN, PAUL (1968). On estimating monotone parameters. *Ann. Math. Statist.* **39** 1030-1039.

[4] ROBERTSON, TIM and WRIGHT, F. T. (1973). Multiple isotonic median regression. *Ann. Statist.* **1** 422-432.

[5] ROBERTSON, TIM and WRIGHT, F. T. (1975). Consistency in generalized isotonic regression. *Ann Statist.* **3** 350-362.

[6] UBHAYA, VASANT A. (1974a). Isotone Optimization I. *J. Approximation Theory* **12** 146-159.

[7] UBHAYA, VASANT A. (1974b). Isotone Optimization II. *J. Approximation Theory* **12** 315-331.

DEPARTMENT OF STATISTICS
UNIVERSITY OF IOWA
IOWA CITY, IOWA 52242

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MISSOURI-ROLLA
ROLLA, MISSOURI 65401