The Annals of Statistics
1983, Vol. 11, No. 2, 467–477

THE MIN-MAX ALGORITHM AND ISOTONIC REGRESSION[1]

BY CHU-IN CHARLES LEE

Memorial University of Newfoundland and University of Iowa

There is growing interest in statistical inference under order restrictions. A major demand in this subject is to have a fast, direct method to solve the least squares problem of partially ordered isotonic regression. The Min-Max algorithm is such a method in which the user searches for the global minimum and the local maximum successively. A comparison of algorithms for partially ordered isotonic regression is included. As an application, using this efficient algorithm, it is feasible to approximate critical values of isotonic tests by simulation.

1. Introduction. Isotonic regression is a least squares problem under order restrictions. It may appear as (2.1) below, or as its generalizations in [4] and [15]. Pioneering work is given by Ayer et al. [3] and a summary can be found in Barlow and coauthors [4]. Applications of simply ordered isotonic regression include maximum likelihood estimation of ordered binomial parameters [3], Poisson extreme problems [4] and [16], ordered regression lines [1], and multidimensional scalings [11], among others. The prototypical partially ordered isotonic regression problem involves estimating a real valued function on the plane, required to be monotonic with respect to the ordinary partial ordering. There are of course many generalizations of this. Applications of partially ordered isotonic regression include negative estimates of variance components [17], representation of similarity matrices by trees [10], juice composition of mandarin oranges [13], legal interruptions of pregnancy in Hungary [9], and additive structure in qualitative data [12], among others. A quasi-ordered isotonic regression can be reduced to a partially ordered isotonic regression.

The first algorithm for the simply ordered isotonic regression was the efficient Pool-Adjacent-Violators of Ayer et al. [3], but it does not apply in general to partially ordered isotonic regression. The first algorithm for the partially ordered isotonic regression was the Minimum Lower Sets of Brunk [6], but this method requires computation of many weighted averages. Van Eeden [8] introduced an algorithm which is suitable for small problems. An improved algorithm was given by Gebhardt [9] in 1970. These methods are not efficient for problems of large size, as shown in Table 7.1. A fast, direct method for partially ordered isotonic regression, the Min-Max algorithm, is presented in Section 3. Because of the efficiency of this algorithm, it is feasible for simulations in applications to the test of homogeneity among means against ordered alternatives and to the test of goodness-of-fit of isotonicity.

It is well known that isotonic regression with respect to any partial order is constant on level sets and the value on each of these level sets is given by the weighted average of the observations in the level set. The problem of computing the isotonic regression therefore reduces to that of determining these level sets. A level set B cannot be a constant set of the isotonic regression unless $\mathrm{Av}(U \cap B) \leq \mathrm{Av}(B)$ for all upper sets U intersecting B. Such a set is termed a block. Which level sets are blocks depends on the values of the observations. Thus those level sets which are not blocks can be ignored. The Min-Max algorithm therefore searches for the level sets of the isotonic regression among block classes.

Received September 1980; revised October 1982.

[1] Research supported by Natural Sciences and Engineering Research Council of Canada, Grant No. A-4541.

AMS 1980 *subject classification.* Primary 62F10; secondary 62F03.

Key words and phrases. Min-Max algorithm, isotonic regression, likelihood ratio tests.

2. Some terminology. Let X be a finite set. A binary relation $\preccurlyeq$ on X is a partial order if it is reflexive, transitive and antisymmetric. An element x is an immediate predecessor of an element y if $x \preccurlyeq y$ but there is no z in X distinct from x and y such that $x \preccurlyeq z \preccurlyeq y$. An element x is minimal in X if there is no z in X distinct from x such that $z \preccurlyeq x$. Two elements x and y are comparable if $x \preccurlyeq y$ or $y \preccurlyeq x$. The partial order $\preccurlyeq$ on X is a simple order if every pair of elements are comparable. A binary relation $\preccurlyeq$ on X is a quasi-order if it is reflexive and transitive but not necessarily antisymmetric. In this article, we shall let X be a partially ordered finite set.

A subset L of X is called a lower set if $y \in L$, $x \in X$ and $x \preccurlyeq y$ imply $x \in L$. A subset U of X is called an upper set if $x \in U$, $y \in X$ and $x \preccurlyeq y$ imply $y \in U$. The complement of a lower set is an upper set and the complement of an upper set is a lower set. The intersection of a lower set and an upper set is called a level set. A subset S of X is a level set if and only if $x, y \in S$, $z \in X$ and $x \preccurlyeq z \preccurlyeq y$ imply $z \in S$. A real valued function f on X is isotonic with respect to the given partial order $\preccurlyeq$ if $x, y \in X$, and $x \preccurlyeq y$ imply $f(x) \leq f(y)$. The indicator function of an upper set is isotonic.

Let g be a given function on X and let w be a given positive weight function on X. An isotonic function g^* on X is called an isotonic regression of g with weights w if in the space of isotonic functions f on X it minimizes the sum

$$\sum_{x \in X} \{g(x) - f(x)\}^2 w(x). \tag{2.1}$$

The existence and the uniqueness of the isotonic regression have been shown by Brunk [7].

Let S be a nonempty subset of X. The weighted average of g over S is denoted by $\mathrm{Av}(S)$, that is,

$$\mathrm{Av}(S) = \sum_{x \in S} w(x)g(x) / \sum_{x \in S} w(x). \tag{2.2}$$

Let $[g^* = \alpha]$ be defined as $\{x : g^*(x) = \alpha\}$. It is well known that if α is a value such that $[g^* = \alpha]$ is nonempty then

$$g^*(x) = \mathrm{Av}([g^* = \alpha]) \tag{2.3}$$

for each $x \in [g^* = \alpha]$. By Theorem 2.4 and Theorem 1.7 of Barlow et al. [4],

$$\mathrm{Av}(L \cap [g^* = \alpha]) \geq \alpha \tag{2.4}$$

for each lower set L which meets $[g^* = \alpha]$, and

$$\mathrm{Av}(U \cap [g^* = \alpha]) \leq \alpha \tag{2.5}$$

for each upper set U which meets $[g^* = \alpha]$.

A subset B of X is called a block (with respect to given g and w) if it is a nonempty level set and if it satisfies

$$\mathrm{Av}(U \cap B) \leq \mathrm{Av}(B) \tag{2.6}$$

for each upper set U of X which meets B. A block B is called a solution block if $g^*(x) = \mathrm{Av}(B)$ for each $x \in B$. By (2.3) and (2.5), each nonempty set $[g^* = \alpha]$ is a solution block. A collection Δ of mutually disjoint blocks of X is called a block class if (i) it can be partially ordered by an induced binary relation $<$ so that $A < B$ for $A, B \in \Delta$ if there are $x \in A$ and $y \in B$ such that $x \preccurlyeq y$, and if (ii) the union of all blocks in Δ is a level set. The collection of all singleton subsets of X is a block class and the collection of all solution blocks $[g^* = \alpha]$ is also a block class. A block $A \in \Delta$ is an immediate predecessor of block $B \in \Delta$ if $A < B$ but there is no block $C \in \Delta$ distinct from A and B such that $A < C < B$. If A is an immediate predecessor of B then there are $x \in A$ and $y \in B$ such that x is an immediate predecessor of y. A block $B \in \Delta$ is a lower set of the union of all blocks in Δ if and only if B is minimal in Δ.

3. The Min-Max algorithm. The first block class, Δ^1, consists of the individual elements of the design space X. The algorithm then iterates in the following manner:

PART I. The block V^I with the lowest possible weighted average in Δ^I, the Ith block class, is selected.

1. If that block is a lower set of X^I, V^I is a solution block. The problem is then reduced to that of finding the isotonic regression on

$$X^{I+1} = X^I - V^I. \tag{3.1}$$

The new block class is

$$\Delta^{I+1} = \Delta^I - \{V^I\}. \tag{3.2}$$

2. If V^I is not a lower set of X^I, it will be pooled with an immediate predecessor. Only certain predecessors will be considered (see Part II). The eligible immediate predecessor with the highest possible weighted average (the one most able to bring V^I into the appropriate order) will be joined to V^I. The operation space for the next stage remains the same, that is,

$$X^{I+1} = X^I. \tag{3.3}$$

However, this union $P^I \cup V^I$, which shall be denoted by G^I, may not result in a block:

(i) If it does result in a block, the new block class is

$$\Delta^{I+1} = \Delta^I - \{V^I, P^I\} + \{G^I\}. \tag{3.4}$$

(ii) If it does not result in a block, the upper set U^I of G^I with the highest possible weighted average will be a block. If there are several upper sets having the same highest average take U^I as their union. The remaining fragments of individual design points are also blocks. The new block class is therefore

$$\Delta^{I+1} = \Delta^I - \{V^I, P^I\} + \{U^I\} + \{\{x_j\} : x_j \in F^I\} \tag{3.5}$$

where $F^I = G^I - U^I$.

PART II. Which immediate predecessors are considered depends on the value of the block V^I. If $\mathrm{Av}(V^I) < \mathrm{Av}(V^J)$ for some $J < I$, V^I must not have been a block in Δ^J. Let J be the last such index. Therefore V^I must have been created from the fragments when G^J was split. In this case, only the remaining fragments are searched for immediate predecessors. If, however, $\mathrm{Av}(V^I) \geq \mathrm{Av}(V^J)$ for all $J < I$, then there is no restriction on immediate predecessors of V^I.

We start with every member of the design space X being an identification element and then combining members and preserving the identification element of the immediate predecessor member. The purpose of identification element and that of Part II of the Min-Max algorithm are to simplify the task of finding the upper set U^I with the highest possible value. As the algorithm iterates, we have a sequence $\{\mathrm{Av}(V^I)\}$. So long as G^I are blocks, the sequence will be monotone nondecreasing. The sequence ceases to be monotone nondecreasing at J when G^J is not a block. The converse, however, is not true. When G^J is not a block and $\mathrm{Av}(V^{J+1}) < \mathrm{Av}(V^J)$, the algorithm is executed within the fragments F^J for the stages $J+1, J+2, \cdots$, until stage I such that $\mathrm{Av}(V^I) \geq \mathrm{Av}(V^J)$.

The isotonic regression $g^*(x)$ is $\mathrm{Av}(V^I)$ for each $x \in V^I$ when V^I is a lower set of X^I. The justification of the Min-Max algorithm is given in Section 6 while some methods for finding the maximum upper set U^I are presented in the next section. We shall illustrate the Min-Max algorithm by a numerical example in Section 5. The speed of the Min-Max algorithm depends on the number of stages required to execute a problem and the total

number of times that weighted averages needed to be computed. Consider the problem given in Section 5 with data g_{ij} taken as random normal digits. A study of 1000 cases showed that 888 of them were executed by the minimum number of 16 stages. The largest number of stages in the study was 20, which occurred only once. We shall discuss the speed of the Min-Max algorithm in Section 7.

4. Methods for searching the maximum upper set. A major task in the Min-Max algorithm is to find the maximum upper set U^I of $G^I = P^I \cup V^I$. The maximum upper set U^I is the largest upper set of G^I such that

$$\text{Av}(U^I) = \max\{\text{Av}(U): U \text{ upper set of } G^I\}. \tag{4.1}$$

If there are several upper sets having the same highest average, take U^I as their union. But upper sets of G^I may be numerous when G^I contains a large number of elements. Five methods are presented in this section with their comparison being given in Section 7.

Every block B at each stage of the Min-Max algorithm contains a special element. This special element is called an identification element and is denoted by $e(B)$. An identification element is a minimal element of the block to which it belongs. If B is a singleton block $\{x\}$ then $e(B) = x$. If B is not a singleton block, then B must be a maximum upper set at some stage. The identification element $e(U^I)$ of the maximum upper set U^I is defined as the identification element $e(P^I)$ of P^I. We shall show in Theorem 4.2 below that $e(P^I)$ belongs to U^I.

LEMMA 4.1. *Let I be such that $\text{Av}(V^I) < \text{Av}(V^H)$ for some $H < I$. Then there exists a J such that* (a) $\text{Av}(V^J) > \text{Av}(V^I)$, (b) $\text{Av}(V^K) \le \text{Av}(V^I)$ *for* $K = J+1, \cdots, I$, (c) V^K, P^K *are covered by* F^J *for* $K = J+1, \cdots, I$, (d) V^K *is not a lower set of* F^J *for* $K = J+1, \cdots, I$.

PROOF. Let $J < I$ be the last index such that $\text{Av}(V^I) < \text{Av}(V^J)$. Then (a) and (b) hold. Clearly, $\text{Av}(V^J) > \text{Av}(V^{J+1})$ and U^J is a proper subset of G^J. Every block in Δ^J has a weighted average no less than $\text{Av}(V^J)$. So is U^J. The blocks that can have weighted averages less than $\text{Av}(V^J)$ are the ones covered by F^J. By Part II of the Min-Max algorithm, (c) holds. Let L be a lower set of F^J. Then $\text{Av}(L)$ is a weighted average of $\text{Av}(L \cap P^J)$ and $\text{Av}(L \cap V^J)$. By (2.6), the last two values are no less than $\text{Av}(V^J)$. Therefore $\text{Av}(L) \ge \text{Av}(V^J)$ and (d) holds. □

THEOREM 4.2. *Let U be an upper set of G^I. If U does not contain $e(P^I)$ then* $\text{Av}(U) \le \text{Av}(V^I)$. *Consequently, $e(P^I)$ belongs to U^I and U^I is the largest upper set of G^I such that*

$$\text{Av}(U^I) = \max\{\text{Av}(U): U \text{ upper set of } G^I,\ e(P^I) \in U\}. \tag{4.2}$$

PROOF. Assume that each nonsingleton block B at the beginning of the Ith stage can be expressed by

$$B = \{e(P^{i_1})\} \cup (V^{i_1} \cap \cap_{h=1}^{r} U^{i_h}) \cup \cdots \cup (V^{i_r} \cap U^{i_r}) \tag{4.3}$$

with $\text{Av}(V^{i_j}) \le \text{Av}(V^{i_r})$ for $j = 1, \cdots, r$ where $e(P^{i_1}) = e(B)$ is the unique element in P^{i_1}. The only new nonsingleton block, if any, created at the Ith stage is U^I. By mathematical induction, it suffices to show that (4.3) holds for U^I so that (4.3) holds for every nonsingleton block.

Suppose that P^I is nonsingleton. Then P^I can be expressed by (4.3). By Lemma 4.1 and Part II of the Min-Max algorithm, we have that $\text{Av}(V^{i_j}) \le \text{Av}(V^I)$ for $j = 1, \cdots, r$. Let U be an upper set of G^I. By (2.6), each nonempty level set $V^{i_j} \cap (\cap_{h=j}^{r} U^{i_h} \cap U)$ has a weighted average no more than $\text{Av}(V^{i_j})$ for each j. It follows that $\text{Av}(U) \le \text{Av}(V^I)$ if U

does not contain $e(P^I) = e(P^{I_1})$. Consequently, $e(P^I)$ belongs to U^I and U^I can be expressed by (4.3). It is trivial for the case that P^I is singleton and the proof is complete. □

Non-singleton blocks are listed in the last column of Table 5.2 with their expressions (4.3) while the leading elements are the identification elements for the numerical example given in the next section. The number of upper sets in (4.1) is reduced significantly to that in (4.2). Further reductions can be achieved. Let $e = e(P^I)$ and let $U_e = \{x : x \in G^I, e \preccurlyeq x\}$. An element y in $G^I - U_e$ is said to be excessive if $g(y) \geq \mathrm{Av}(G^I)$. An element z in $G^I - U_e$ is said to be deficient if $g(z) < \mathrm{Av}(G^I)$. A sequence of elements $x_1, \cdots, x_n$ are said to be non-comparable if every pair of elements are not comparable. Let $y_1, \cdots, y_n$ be n non-comparable excessive elements. The smallest upper set of $G^I - U_e$ which contains $y_1, \cdots, y_n$ is called an excessive upper set; we denote it by $U[y_1, \cdots, y_n]$. Let $z_1, \cdots, z_n$ be n non-comparable deficient elements. The smallest lower set of $G^I - U_e$ which contains $z_1, \cdots, z_n$ is called a deficient lower set; we denote it by $L[z_1, \cdots, z_n]$.

METHOD 1. *Excessive upper sets.* It is obvious that (4.2) can be reduced to

$$\mathrm{Av}(U^I) = \max\{\mathrm{Av}(U_e \cup U) : U \text{ excessive upper set } U[y_1, \cdots, y_n]\} \tag{4.4}$$

where $\mathrm{Av}(U_e \cup U[y_1, \cdots, y_n])$ can be computed as the weighted average of the two values $\mathrm{Av}(U_e)$ and $\mathrm{Av}(U[y_1, \cdots, y_n])$.

METHOD 2. *Connected excessive upper sets.* Let B be a level set of X. It is said to be connected if for any of its partitions of B_1 and B_2, there are a pair of elements, one in B_1 and one in B_2, which are comparable. It follows from Theorem 4.2 that every block in the Min-Max algorithm is connected. Therefore, (4.4) can be reduced to

$$\mathrm{Av}(U^I) = \max\{\mathrm{Av}(U_e \cup U[y_1, \cdots, y_n]) : U_e \cup U[y_1, \cdots, y_n] \text{ connected}\}. \tag{4.5}$$

The amount of reduction from (4.4) to (4.5) depends on the partial order $\preccurlyeq$ over the set $G^I - U_e$.

METHOD 3. *Deficient lower sets.* In the Min-Max algorithm, it is likely that the number of excessive elements is larger than that of deficient elements. It is economical to consider lower sets instead of upper sets in (4.4) and (4.5). In this method, we find the largest deficient lower set L^* which satisfies

$$\mathrm{Av}(L[z_1, \cdots, z_n]) < \mathrm{Av}(G^I). \tag{4.6}$$

The maximum upper set U^I is $G^I - L^*$. It is trivial that lower sets $L[z_1, \cdots, z_n]$ covered by P^I do not satisfy (4.6) and they shall not be considered in searching for L^*.

METHOD 4. *Deficient lower sets as complements of excessive upper sets.* This method is the same as Method 3 but we exclude those lower sets $L[z_1, \cdots, z_n]$ such that $G^I - U_e - L[z_1, \cdots, z_n]$ contains a minimal element which is deficient. In other words, the lower sets $L[z_1, \cdots, z_n]$ we shall consider are complements of excessive upper sets with respect to $G^I - U_e$.

METHOD 5. *Deficient lower sets as complements of connected excessive upper sets.* This method is the same as Method 4 but we consider only those lower sets $L[z_1, \cdots, z_n]$ which are complements of connected excessive upper sets.

When $G^I - U_e$ contains no more than ten elements, Method 3 performs well. When $G^I - U_e$ contains a large number of elements, Method 4 and Method 5 are better. The degree of improvement of the last two methods over Method 3 can be illustrated in Table 7.1.

5. A numerical example. The following hypothetical example has been created to illustrate the operation of the Min-Max algorithm. The index set $X = \{(i, j) : 1 \leq i \leq 4, 1 \leq j \leq 4\}$ is partially ordered so that $(i, j) \preccurlyeq (i', j')$ if $i \leq i'$ and $j \leq j'$. Element (1, 1) has no immediate predecessor, elements (1, 2), (1, 3), (1, 4), (2, 1), (3, 1) and (4, 1) have one immediate predecessor each, while the remaining nine elements have two immediate predecessors each. Let $g_{ij} = g[(i, j)]$ be given in Table 5.1 and $w_{ij} = w[(i, j)] \equiv 1$.

The stagewise procedure is shown in Table 5.2. At stage 1, block $V^1 = \{(2, 2)\}$ has two eligible immediate predecessors $\{(1, 2)\}$ and $\{(2, 1)\}$. The latter is selected as P^1 because it has a larger average. At stage 8, $G^8 = \{(1, 2), (2, 1), (2, 2), (3, 1), (4, 1)\}$ and it contains ten upper sets among which four of them contain the identification element $e = (1, 2)$. By Method 1, we need to consider three upper sets. They are U_e, $U_e \cup U[(2, 1)]$ and $U_e \cup U[(3, 1)]$ where $U_e = \{(1, 2), (2, 2)\}$, $U[(2, 1)] = \{(2, 1), (3, 1), (4, 1)\}$ and $U[(3, 1)] = \{(3, 1), (4, 1)\}$. The upper set $U_e \cup U[(3, 1)]$ is not connected and it is excluded in Method 2. There is only one deficient lower set $L[(4, 1)] = \{(4, 1), (3, 1), (2, 1)\}$ to be considered by Method 3, Method 4 and Method 5. Since $L[(4, 1)]$ fails to satisfy (4.6), the maximum upper set U^8 is G^8. At stage 12, $U^{12} = U_e = \{(1, 4), (2, 4), (3, 4)\}$ is a proper subset of G^{12} and elements (3, 2) and (3, 3) are now placed in singleton blocks $\{(3, 2)\}$ and $\{(3, 3)\}$. At the next stage, $\{(3, 2)\}$ is the only eligible immediate predecessor of $V^{13} = \{(3, 3)\}$. At stage 14, there is one excessive upper set $U[(4, 3)] = \{(4, 3)\}$ but there are no deficient lower sets. Therefore, it must be that $U^{14} = G^{14}$. There are four levels of solution blocks V^4, V^{10}, V^{17} and V^{18}. The problem is solved by computing weighted averages 25, 24, 17, 17 and 17 times respectively for the five methods in the preceding section.

TABLE 5.1
Given values g_{ij}

4	48	16	14	6
3	37	12	9	26
2	19	2	25	17
1	8	27	21	4
j \ i	**1**	**2**	**3**	**4**

TABLE 5.2
Stagewise Procedure of the Min-Max Algorithm

Stage I	Av(V^I)	$e(V^I)$	Av(P^I)	$e(P^I)$	Av(U^I)	U^I
1	2.00	(2,2)	27.00	(2,1)	14.50	{(2,1)} ∪ {(2,2)}
2	4.00	(4,1)	21.00	(3,1)	12.50	{(3,1)} ∪ {(4,1)}
3	6.00	(4,4)	26.00	(4,3)	16.00	{(4,3)} ∪ {(4,4)}
4	8.00	(1,1)				
5	9.00	(3,3)	25.00	(3,2)	17.00	{(3,2)} ∪ {(3,3)}
6	12.00	(2,3)	37.00	(1,3)	24.50	{(1,3)} ∪ {(2,3)}
7	12.50	(3,1)	14.50	(2,1)	13.50	{(2,1)} ∪ {(2,2)} ∪ {(3,1), (4,1)}
8	13.50	(2,1)	19.00	(1,2)	14.60	{(1,2)} ∪ {(2,1), (2,2), (3,1), (4,1)}
9	14.00	(3,4)	17.00	(3,2)	16.00	{(3,2)} ∪ {(3,3)} ∪ {(3,4)}
10	14.60	(1,2)				
11	16.00	(2,4)	48.00	(1,4)	32.00	{(1,4)} ∪ {(2,4)}
12	16.00	(3,2)	32.00	(1,4)	26.00	{(1,4)} ∪ {(2,4)} ∪ {(3,4)}
13	9.00	(3,3)	25.00	(3,2)	17.00	{(3,2)} ∪ {(3,3)}
14	16.00	(4,3)	26.00	(1,4)	22.00	{(1,4)} ∪ {(2,4)} ∪ {(3,4)} ∪ {(4,3), (4,4)}
15	17.00	(3,2)	24.50	(1,3)	20.75	{(1.3)} ∪ {(2,3)} ∪ {(3,2), (3,3)}
16	17.00	(4,2)	20.75	(1,3)	20.00	{(1,3)} ∪ {(2,3)} ∪ {(3,2), (3,3)} ∪ {(4,2)}
17	20.00	(1,3)				
18	22.00	(1.4)				

6. Justification of Min-Max algorithm. Let X be a partially ordered finite set with a partial order $\preccurlyeq$. Let $\Delta^1, \cdots, \Delta^I$ be a sequence of block classes with a transition from Δ^I to Δ^{I+1} as defined by (3, 2), (3, 4) or (3.5) of the Min-Max algorithm, where Δ^1 is the initial class of all singleton subsets of X. The union of all blocks in Δ^I is the space X^I of all elements in operation and $X - X^I$ is the set for which the isotonic regression has been solved at the beginning of stage I. If V^I is a lower set of X^I then the isotonic regression $g^*(x) = \mathrm{Av}(V^I)$ for each $x \in V^I$. We shall justify the Min-Max algorithm by the following four theorems.

THEOREM 6.1. *If V^I is not a lower set of X^I, then it has at least one eligible immediate predecessor.*

PROOF. Suppose that $\mathrm{Av}(V^I) \geq \mathrm{Av}(V^J)$ for $J = 1, \cdots, I$. Since V^I is not a lower set of X^I, it is not minimal in the partially ordered block class Δ^I. Therefore, V^I has at least one eligible immediate predecessor.

Suppose that $\mathrm{Av}(V^I) < \mathrm{Av}(V^J)$ for some $J < I$. Let J be the last such index. By Lemma 4.1, $V^I \subset F^J$ and V^I is not a lower set of F^J. Therefore, V^I has an immediate predecessor which is covered by F^J. □

THEOREM 6.2. *The algorithm produces a block class at each stage.*

PROOF. Let Δ^I be the block class at stage I of the Min-Max algorithm. We shall show that Δ^{I+1} defined by (3.2), (3.4), or (3.5) is also a block class. Suppose that V^I is a lower set of X^I. Then V^I is a minimal element in the partially ordered block class Δ^I. By (3.2), the collection Δ^{I+1} of blocks is a block class with a partial order inherited from Δ^I.

Suppose that V^I is not a lower set of X^I. The blocks in Δ^I can be ordered as a sequence $B_1, \cdots, B_m$ so that A precedes B for each pair of blocks $A < B$ in Δ^I and there is a suitable i such that $B_i = P^I$ and $B_{i+1} = V^I$. Since U^I is the maximum upper set of G^I, it is a block which satisfies (2.6). Suppose that $F^I = G^I - U^I$ is nonempty. The elements of F^I can be ordered as a sequence $x_1, \cdots, x_r$ so that x precedes y for each pair of elements $x \preccurlyeq y$ in F^I. Let A_h be the hth block of the sequence

$$B_1, \cdots, B_{i-1}, \{x_1\}, \cdots, \{x_r\}, U^I, B_{i+2}, \cdots, B_m. \tag{6.1}$$

Members of Δ^{I+1} as defined by (3.5) are the blocks listed in (6.1). Let a binary relation $<'$ be defined on Δ^{I+1} so that $A_i <' A_j$ if there are $x \in A_i$ and $y \in A_j$ such that $x \preccurlyeq y$. Apply the reflexivity and the transitivity to $<'$. The binary relation $<'$ is a quasi-order on Δ^{I+1}. Since it is not possible to have $A_j <' A_i$ for $j > i$, $<'$ is a partial order on Δ^{I+1}. For the case that U^I coincides with G^I, the collection Δ^{I+1} of blocks defined by (3.4) can be shown similarly to be a block class by deleting $\{x_1\}, \cdots, \{x_r\}$ in (6.1). This completes the proof. □

THEOREM 6.3. *The sequence of block classes $\Delta^1, \Delta^2, \cdots$ of the Min-Max algorithm will terminate after a finite number of stages.*

PROOF. Since X is finite, the number of distinct block classes of X is finite. Therefore the theorem will be proved once it has been established that the block classes selected by the algorithm are distinct. Clearly Δ^{I+1} differs from Δ^I. Furthermore, unless $U^I \neq G^I$, Δ^{I+1} contains one fewer block than Δ^I (either by elimination of a block or by amalgamation of two). Suppose that $\Delta^I = \Delta^{I+C}$ for some $C > 1$. Then there exists a loop (if each step in the algorithm is uniquely defined). Furthermore, there exists a J, $I \leq J < I + C$ such that U^J is a proper subset of G^J and $\mathrm{Av}(V^J) \geq \mathrm{Av}(V^K)$ for $K = I, I+1, \cdots, I+C$. If U^J is a proper subset of P^J then from (2.6) $\mathrm{Av}(V^J) < \mathrm{Av}(U^J) \leq \mathrm{Av}(P^J)$. If U^J intersects both P^J and V^J then $\mathrm{Av}(U^J)$ is a weighted average of $\mathrm{Av}(U^J \cap P^J)$ and $\mathrm{Av}(U^J \cap V^J)$. By (2.6) the last two values are no more than $\mathrm{Av}(P^J)$ and hence $\mathrm{Av}(V^J) < \mathrm{Av}(U^J) < \mathrm{Av}(P^J)$. By Theorem

4.2, U^J contains $e(P^J)$. Clearly U^J cannot appear as V^K for $K = J + 1, \cdots, J + C$. If $U^J = P^K$ for some K, $J + 1 \leq K < J + C$ then by Part II of the Min-Max algorithm either $\text{Av}(P^K) \geq \text{Av}(U^K) > \text{Av}(V^K) = \text{Av}(V^J)$ with U^K a proper subset of P^K or $\text{Av}(P^K) > \text{Av}(U^K) > \text{Av}(V^K) = \text{Av}(V^J)$. The block U^K is the one in Δ^K which contains $e(P^K) = e(P^J)$. Therefore, if $B \in \Delta^{J+C}$ is a block which contains $e(P^J)$ then either B is a proper subset of P^J or $\text{Av}(B) < \text{Av}(P^J)$. In either case, we have that $P^J \notin \Delta^{J+C}$. Since it is not possible to have $\Delta^{J+C} = \Delta^J$, the block classes selected by the algorithm are distinct. □

THEOREM 6.4. *The isotonic regression at a point in a terminal block is equal to the average over that block, and hence the isotonic regression is the average over the blocks which the algorithm selects.*

PROOF. Let V^I be a lower set of X^I, let $x \in V^I$ and let $\alpha = g^*(x)$. We shall show that

$$g^*(x) = \text{Av}(V^I). \tag{6.2}$$

By (3.1), X^I is an upper set of X. It follows from theorem 2.4 of [4] that

$$\text{Av}(X^I \cap [g^* \leq \alpha]) \leq \alpha. \tag{6.3}$$

From (2.6), we have that

$$\text{Av}([g^* \leq \alpha] \cap B) \geq \text{Av}(B) \geq \text{Av}(V^I) \tag{6.4}$$

for each block $B \in \Delta^I$ which intersects the lower set $[g^* \leq \alpha]$. Since the left hand side of (6.3) is a weighted average of those left hand sides of (6.4), we have that $\text{Av}(V^I) \leq \alpha$. It remains to show that $\text{Av}(V^I) \geq \alpha$.

Let $L^I = V^I \cup (X - X^I)$. Then L^I is a lower set of X. This lower set is the union of all blocks $B_1, \cdots, B_t$ which have been eliminated by the Min-Max algorithm at the end of stage I. It follows from Theorem 2.4 of [4] that

$$\text{Av}(L^I \cap [g^* \geq \alpha]) \geq \alpha. \tag{6.5}$$

Since V^I is a lower set of X^I, by Lemma 4.1 we have that $\text{Av}(V^I) \geq \text{Av}(V^J)$ for $J = 1, \cdots, I$. From (2.6), we have that

$$\text{Av}([g^* \geq \alpha] \cap B_j) \leq \text{Av}(B_j) \leq \text{Av}(V^I) \tag{6.6}$$

for each block $B_j, j = 1, \cdots, t$, which intersects the upper set $[g^* \geq \alpha]$. Since the left hand side of (6.5) is a weighted average of those left hand sides of (6.6), we have that $\text{Av}(V^I) \geq \alpha$. The proof is complete. □

7. A comparison of algorithms for partially ordered isotonic regression. The purpose of this section is to compare algorithms for partially ordered isotonic regression, to examine the five methods in Section 4, and to discuss the speed of the Min-Max algorithm. We shall proceed to investigate these by considering the following five problems:

PROBLEM 1. Let $X = \{(i, j) : i = 1, \cdots, 4, j = 1, \cdots, 4\}$ be partially ordered so that $(i, j) \preccurlyeq (i', j')$ if $i \leq i'$ and $j \leq j'$. Let $w_{ij} \equiv 1$ and let g_{ij} be random normal digits. Four sets of 16 values of g_{ij} are executed independently. The data are listed in the order (1, 1), (2, 1), $\cdots$, (4, 4).

PROBLEM 2. Let $X = \{(i, j) : i = 1, \cdots, 8, j = 1, \cdots, 8\}$ be partially ordered so that $(i, j) \preccurlyeq (i', j')$ if $i \leq i'$ and $j \leq j'$. Let $w_{ij} \equiv 1$ and let g_{ij} be random normal digits. The data are listed in the order (1, 1), (2, 1), $\cdots$, (8, 8).

PROBLEM 3. Let $X = \{(i, j, k) : i = 1, \cdots, 4, j = 1, \cdots, 4, k = 1, \cdots, 4\}$ be partially ordered so that $(i, j, k) \preccurlyeq (i', j', k')$ if $i \leq i', j \leq j'$ and $k \leq k'$. Let $w_{ijk} \equiv 1$ and let g_{ijk} be random normal digits. The data are listed in the order (1, 1, 1), (2, 1, 1), $\cdots$, (4, 4, 4).

PROBLEM 4. The data are the legal interruptions of pregnancy in Hungary appearing in Gebhardt [9]. There are 93 observations in a partially ordered set X of 76 elements among which 61 of them have one observation, 13 of them have two observations and 2 of them have three observations so that w_{ij} is the number of observations and g_{ij} is the arithmetic average of observations at each (i, j). The data are listed in the order (43, 1), (44, 1), $\cdots$, (65, 15).

PROBLEM 5. This problem is the same as Problem 4 but the data are listed in the order (38, 8), (43, 1), $\cdots$, (71, 4).

For convenience, the 64 given values of g for the first three problems are identically the first 64 random normal digits in [5], reading along columns. We solved these five problems by the Minimum Lower Sets algorithm of Brunk [6], Van Eeden's [8] algorithm, Gebhardt's [9] algorithm, and the Min-Max algorithm with five methods. The efficiency of Van Eeden's algorithm and that of Gebhardt's algorithm depend on the order in which the data is listed. We considered the last two problems of the same data but two different ways in which the data are ordered. The efficiency of the Minimum Lower Sets algorithm does not depend on the order of the data while that of the Min-Max algorithm is affected by the order of the data when there are ties among averages.

The total number of times that weighted averages were computed for the four algorithms to solve each of the five problems are given in Table 7.1. In the last four problems, those total numbers were too costly to find by Van Eeden's algorithm so we inserted their lower bounds instead. It is clear that Van Eeden's algorithm and the Minimum Lower Sets algorithm are not efficient for problems with 16 or more values. It is also clear that the Min-Max algorithm is the most efficient one. For larger problems we have difficulty in finding the total numbers of computations needed by the Minimum Lower Sets algorithm. For problems of smaller size, the Min-Max algorithm can be shown again as the most efficient one.

The speed of the Min-Max algorithm depends on the total number of stages and the total number of times that weighted averages needed to be computed. Those stages such that V^I are lower sets are easier to execute. The number of those stages are the number of solution blocks. Those numbers were 22, 7, 7, 21 and 24 respectively for the five problems. If we exclude those stages, then the number of stages needed for the five problems were 43, 57, 64, 58 and 56. Therefore, the total number of stages needed by the Min-Max algorithm is about the size of the problem.

The total number of times that weighted averages needed to be computed depends on the problem as well as the method. There were 43, 57, 64, 58 and 56 computations of weighted average of $\mathrm{Av}(P^I)$ and $\mathrm{Av}(V^I)$ for the five problems. The remainder of the computations were due to searching for the maximum upper sets U^I. They were computations of $\mathrm{Av}(U_e \cup U[y_1, \cdots, y_n])$ or $\mathrm{Av}(L[z_1, \cdots, z_n])$ as defined in Section 4. It is clear that Method 3 is superior to Method 1 and Method 2. By this method, the computations of $\mathrm{Av}(L[z_1, \cdots, z_n])$ were 6, 61, 325, 7 and 4 times respectively for the five problems. The last number of 4 for the problem of size 76 is smaller than the first number of 6 for the problem of size 16 in four tries. This is because the data in Problem 5 showed a strong isotonicity while the data in Problem 1 did not. Consequently, a very large problem can be solved easily if the given function g shows a strong isotonicity. Only in Problem 3 is there an appreciable improvement from Method 3 to Method 4 and Method 5. Even though the last two methods are not so easy to use as Method 3, they are superior for the case that the order restricted regression has more than two independent variables and the number of data points is large.

8. Applications to hypothesis testing. Let y_{ij}, $i = 1, \cdots, k$, $j = 1, \cdots, n_i$, be independent normal variates with unknown means $\mu(x_i)$ and with given variances σ_i^2. Suppose that the regression curve μ is known to satisfy some order restrictions. The set $X = \{x_1, \cdots, x_k\}$ is partially ordered according to those restrictions so that μ is isotonic on X.

TABLE 7.1[a]
Total Numbers of Computations for Isotonic Regression Algorithms

Algorithm	Problem 1	Problem 2	Problem 3	Problem 4	Problem 5
Minimum Lower Sets	692	25702	681687	227172	227172
Van Eeden's	211205	(1.4×10^{17})[b]	(1.4×10^{17})	(8.8×10^{12})	(8.8×10^{12})
Gebhardt's	118	4086	22138	133	557
Min-Max: Method 1	74	407	869	82	71
Min-Max: Method 2	74	346	617	78	70
Min-Max: Method 3	49	118	389	65	60
Min-Max: Method 4	49	101	190	65	60
Min-Max: Method 5	49	95	157	65	60

[a] Gebhardt's subalgorithm in Section 4 of [9] is defective, and a suitable replacement has been used in the execution. It was intended to include the Minimax Order algorithm of [2] in the study but the algorithm is defective in the sense that it fails to maintain the excesses of adjacent independent subblocks in nondecreasing order.
[b] These are the lower bounds.

Barlow et al. [4] considered a likelihood ratio test of homogeneity against ordered alternatives. The null hypothesis is

$$H_0 : \mu(x_1) = \cdots = \mu(x_k)$$

and the alternative hypothesis is $H_1 - H_0$ where

$$H_1 : \mu(\cdot) \text{ is isotonic.}$$

The test statistic is

$$\bar{\chi}_k^2 = \sum_{i=1}^{k} w_i(\hat{\mu}_i^* - \hat{\mu})^2, \tag{8.1}$$

where $\hat{\mu}_i^*$ are the isotonic regression of $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i$ with weights $w_i = n_i\sigma_i^{-2}$ and $\hat{\mu} = \sum_{i=1}^{k} w_i\bar{y}_i / \sum_{i=1}^{k} w_i$. The P-value of an observed test statistic $\bar{\chi}_k^2 > 0$ is

$$\bar{P} = \sum_{\ell=2}^{k} P(\ell, k)\Pr(\chi_{\ell-1}^2 \geq \bar{\chi}_k^2) \tag{8.2}$$

where $P(\ell, k)$ is the probability that the isotonic regression $\hat{\mu}^*$ takes exactly ℓ distinct values when H_0 is true, and the notation χ_ν^2 is used to denote a random variable having the χ^2 distribution with ν degrees of freedom.

It has been shown in Barlow et al. [4] that if X is simply ordered and $w_1 = \cdots = w_k$, then

$$P(\ell, k) = |S_k^\ell|/k!, \quad \ell = 1, \cdots, k,$$

where $|S_k^\ell|$ is the coefficient of z^ℓ in the expansion of $z(z+1)\cdots(z+k-1)$. However, there is no closed form expression for these probabilities $P(\ell, k)$ in the general situation and direct computation is not feasible for moderate or large k. But the P-value $\bar{P}$ can be approximated by simulation with the aid of the efficient Min-Max algorithm. Consider the example in Section 5. Suppose that 16 values of g_{ij} are observations of independent normal variates with $n_{ij} \equiv 1$ and $\sigma_{ij}^2 \equiv 100$. The observed test statistic (8.1) was $\bar{\chi}_k^2 = 2.572$. The approximated P-values of $\bar{\chi}_k^2$ in a simulation study were 0.4889, 0.5263, 0.5306, 0.5258, $\cdots$, 0.5324 respectively after the 50th, 100th, 150th, 200th, $\cdots$, 1000th cases. The approximation may be terminated at some point when the convergence of the P-values is satisfactory. The rate of convergence is governed by the multinomial distribution of $P(\ell, k)$.

This approximation can also be applied to the test of goodness-of-fit by Robertson and Wegman [14]. The isotonicity of $\mu(\cdot)$ is tested against all possible alternatives. The likelihood ratio principle leads to a test statistic

$$\tilde{\chi}_k^2 = \sum_{i=1}^{k} w_i(\bar{y}_i - \hat{\mu}_i^*)^2. \tag{8.3}$$

The P-value of an observed test statistic $\tilde{\chi}_k^2 > 0$ is

$$\tilde{P} = \sum_{\ell=1}^{k-1} P(\ell, k)\Pr(\chi_{k-\ell}^2 \geq \tilde{\chi}_k^2). \tag{8.4}$$

The observed test statistic (8.3) for the same example was $\tilde{\chi}_k^2 = 20.412$. The approximated P-values of $\tilde{\chi}_k^2$ in the same simulation study were 0.0629, 0.0568, 0.0555, 0.0561, $\cdots$, 0.0552 respectively after the 50th, 100th, 150th, 200th, $\cdots$, 1000th cases.

In the situation with variances $\sigma_i^2 = \alpha_i\sigma^2$, where α_i are known constants and σ^2 is unknown, see [4] and [14] for likelihood ratio statistics. The P-values in this situation have expressions similar to (8.2) and (8.4).

Acknowledgment. The author is very grateful to Professor H. D. Brunk for suggesting this investigation. The author wishes to thank the referees for many helpful comments and suggestions.

REFERENCES

[1] ADICHIE, J. N. (1976). Testing parallelism of regression lines against ordered alternatives. *Commun. Statist. A* **5** 985–997.

[2] ALEXANDER, M. J. (1970). An algorithm for obtaining maximum-likelihood estimates of a set of partially ordered parameters. Rocketdyne Technical Paper.

[3] AYER, M., BRUNK, H. D., EWING, G. M., REID, W. T. and SILVERMAN, E. (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist.* **26** 641–647.

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

[5] BEYER, W. H. (1968). *Handbook of Tables for Probability and Statistics*, 2nd edition. Chemical Rubber Co., Cleveland, Ohio.

[6] BRUNK, H. D. (1955). Maximum likelihood estimates of monotone parameters. *Ann. Math. Statist.* **26** 607–616.

[7] BRUNK, H. D. (1965). Conditional expectation given a σ-lattice and applications. *Ann. Math. Statist.* **36** 1339–1350.

[8] EEDEN, C. VAN (1957). Maximum likelihood estimation of paritally or completely ordered parameters. I. *Indag. Math.* **19** 128–136.

[9] GEBHARDT, F. (1970). An algorithm for monotone regression with one or more independent variables. *Biometrika* **57** 263–271.

[10] HARTIGAN, J. A. (1967). Representation of similarity matrices by trees. *J. Amer. Statist. Assoc.* **62** 1140–1158.

[11] KRUSKAL, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29** 1–27.

[12] LEEUW, J. DE, YOUNG, F. W. and TAKANE, Y. (1976). Additive structure in qualitative data—alternating least-squares method with optimal scaling features. *Psychometrika* **41** 471–503.

[13] LOMBARD, P. B. and BRUNK, H. D. (1963). Evaluating the relation of juice composition of mandarin oranges to percent acceptance of a taste panel. *Food Technol.* **17** 113–115.

[14] ROBERTSON, T. and WEGMAN, E. J. (1978). Likelihood ratio tests for ordered restrictions in exponential families. *Ann. Statist.* **6** 485–505.

[15] ROBERTSON, T. and WRIGHT, F. T. (1980). Algorithms in order restricted statistical inference and the Cauchy mean value property. *Ann. Statist.* **8** 645–651.

[16] SHAKED, M. (1979). Estimation of starshaped sequences of Poisson and normal means. *Ann. Statist.* **7** 729–741.

[17] THOMPSON, W. A., JR. (1962). The problem of negative estimates of variance components. *Ann. Math. Statist.* **33** 273–289.

DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE
THE UNIVERSITY OF IOWA
IOWA CITY, IOWA 52242