PSYCHOMETRIKA—VOL. 42, NO. 1
MARCH, 1977
NOTES AND COMMENTS

CORRECTNESS OF KRUSKAL'S ALGORITHMS FOR MONOTONE REGRESSION WITH TIES

JAN DE LEEUW

UNIVERSITY OF LEIDEN

Kruskal has proposed two modifications of monotone regression that can be applied if there are ties in nonmetric scaling data. In this note we prove Kruskal's conjecture that his algorithms give the optimal least squares solution of these modified monotone regression problems. We also propose another (third) approach for dealing with ties.

Key words: monotone regression, optimal scaling.

Introduction

Most of the algorithms for nonmetric scaling are based on the concept of *monotone regression*. A very good discussion of the various aspects of monotone regression (under the name "isotone regression") can be found in Barlow, Bartholomew, Bremner, and Brunk [1972]. We only give a brief introduction. Suppose $x_1, x_2, \cdots, x_n$ is a given vector of real numbers, $w_1, w_2, \cdots, w_n$ is a given vector of positive real numbers called *weights*, and $\lesssim$ is a given partial order on the index set $\{1, 2, \cdots, n\}$. The monotone regression problem (for a given partial order, and a given vector of weights) is to find the vector $\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_n$ which minimizes

$$S(y) = \sum_{i=1}^{n} w_i(x_i - y_i)^2,$$

over all vectors $y_1, y_2, \cdots, y_n$ satisfying the *monotonicity condition*, i.e., $i \lesssim k \rightarrow y_i \leq y_k$.

If $\lesssim$ is a weak order, algorithms for finding the *monotone regression* $\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_n$ are well known. One of the more efficient ones is discussed by Kruskal [1964]. In nonmetric scaling the partial order $\lesssim$ is derived from the data. For convenience we may assume that the data is a vector $z_1, z_2, \cdots, z_n$ of real numbers, arranged in such a way that $z_1 \leq z_2 \leq \cdots \leq z_n$. If it is true that actually $z_1 < z_2 < \cdots < z_n$, (i.e. if there are no *ties*) we can define $\lesssim$ unambiguously by $i \lesssim k \leftrightarrow z_i \leq z_k$, and apply the algorithm for the weak order case. If the data contain blocks of ties, the situation is a bit more complicated. A first possibility is to impose no order restrictions on the y_i within tie blocks; a second is to require that the y_i within tie blocks be equal. Kruskal [1964] calls these the *primary* and *secondary approaches* to ties. He also discusses two simple modifications of the monotone regression algorithm for weak orders, and conjectures that these algorithms solve the monotone regression problem for the primary and secondary tie constraints. In this note

Comments by Dr. J. B. Kruskal have been most helpful.

Requests for reprints should be sent to Jan de Leeuw, Department of Data Theory, University of Leiden, Wassenaarseweg 80, Leiden, the Netherlands.

we give simple proofs of both conjectures. We also discuss a new, third approach to ties, in which we require the *means* of the y_i in the blocks to be in the appropriate order.

The Primary Approach

We first treat the primary approach. In the case of ties there is a partition of the index set $\{1, 2, \cdots, n\}$ into a number of tie blocks $I_1, I_2, \cdots, I_m$, with $1 \leq m \leq n$. The partial order $\lesssim_p$ is defined by $i \lesssim_p k$ if and only if there exist $1 \leq j < l \leq m$ such that $i \in I_j$ and $k \in I_l$. For convenience we also define an equivalence relation on $\{1, 2, \cdots, n\}$ by $i \simeq k$ if and only if there is a $1 \leq j \leq m$ such that $i \in I_j$ and $k \in I_j$. Suppose Y is the set of all feasible solutions to the monotone regression problem (i.e., the set of all vectors satisfying the constraints), and suppose $\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_n$ is the optimum solution (which is necessarily unique).

Theorem.

If $i \simeq k$ and $x_i \leq x_k$ then $\hat{x}_i \leq \hat{x}_k$.

Proof.

Suppose $\hat{x}_i > \hat{x}_k$. Then define $\bar{x}$ by $\bar{x}_i = \bar{x}_k = (w_i\hat{x}_i + w_k\hat{x}_k)/(w_i + w_k)$ and $\bar{x}_i = \hat{x}_i$ for all $v \neq i, k$. Then $\bar{x} \in Y$, and

$$S(\hat{x}) - S(\bar{x}) = -2\frac{w_iw_k}{w_i + w_k}(x_i - x_k)(\hat{x}_i - \hat{x}_k) + \frac{w_iw_k}{w_i + w_k}(\hat{x}_i - \hat{x}_k)^2.$$

Thus $S(\bar{x}) < S(\hat{x})$, contradicting the optimality of $\hat{x}$. Thus $\hat{x}_i \leq \hat{x}_k$. □

Now define the following extension $\lesssim_e$ of $\lesssim_p$: $i \lesssim_p k \rightarrow i \lesssim_e k$, $i \simeq k$ and $x_i \leq x_k \rightarrow i \lesssim_e k$. Suppose $\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_n$ is the solution of the monotone regression problem for the weak order $\lesssim_e$. By applying the theorem repeatedly we find the following result.

Corollary.

$\hat{x}_i = \tilde{x}_i$ for all $i = 1, 2, \cdots, n$.

Since the construction we have described is the one used by Kruskal, this proves the correctness of Kruskal's first algorithm. It must be emphasized that our theorem is a very special case of a general theorem given by Van Eeden [Note 1, p. 27, theorem 1.1; 4].

The Secondary Approach

The secondary approach is easier. We now define the partial order $\lesssim_s$ by $i \lesssim_s k$ if and only if either $i \lesssim_p k$ or $i \simeq k$. It follows that if $i \simeq k$ we require $y_i = y_k$. Suppose again that $\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_n$ is the solution to the monotone regression problem for the secondary approach. We now define a new, closely related, monotone regression problem without ties. Let, for each $j = 1, 2, \cdots, m$,

$$w_j^{\S} = \sum \{w_i \mid i \in I_j\},$$

and

$$x_j^{\S} = \sum \{w_ix_i \mid i \in I_j\}/w_j^{\S}.$$

Define the weak order $\lesssim_b$ on $\{1, 2, \cdots, m\}$ by $j \lesssim_b l$ if and only if $j \leq l$, and let

$$\hat{x}_1^{\S}, \hat{x}_2^{\S}, \cdots, \hat{x}_m^{\S}$$

be the vector that minimizes

$$S^{\S}(y^{\S}) = \sum_{j=1}^{m} w_j^{\S}(x_j^{\S} - y_j^{\S})^2,$$

over all $y_1^{\S}, y_2^{\S}, \cdots, y_m^{\S}$ satisfying $j \lesssim_b l \rightarrow y_j \leq y_l$.

Theorem.

If $i \in I_j$ then $\hat{x}_i = \hat{x}_j^{\S}$.

Proof.

Suppose $y_1, y_2, \cdots, y_n$ satisfies $y_i = y_k$ for all pairs i, k for which $i \simeq k$. Let $\bar{y}_j$ denote the common value of the y_i in the tie-block I_j. Then

$$S(y) = \sum_{j=1}^{m} \sum \{w_i(x_i - x_j^{\S})^2 \mid i \in I_j\} + \sum_{j=1}^{m} w_j(x_j^{\S} - \bar{y}_j)^2.$$

The first term is independent of $\bar{y}_1, \bar{y}_2, \cdots, \bar{y}_m$, and the second term is minimized by taking $\bar{y}_j = \hat{x}_j^{\S}$. □

The procedure implied by this theorem is to solve a monotone regression problem with a weak order defined over the block averages. This is precisely Kruskal's second algorithm, which is consequently also correct.

A Third Approach

It is clear that the conditions on the y_i defined by the primary approach are weaker than those of the secondary approach. We now define an even weaker approach, which is inspired by the work of Bradley, Katti, and Coons [1962], and Hayashi [1974]. For this third approach we require that the weighted means of the y_i in the tie-blocks are increasing, but we do *not* require the previous monotonicity condition. Thus if

$$y_j^{\S} = \sum \{w_i y_i \mid i \in I_j\}/w_j^{\S},$$

then we require

$$y_1^{\S} \leq y_2^{\S} \leq \cdots \leq y_m^{\S}.$$

Observe that the quadratic programming problem defined by these conditions is not strictly a monotone regression problem as defined in the first section. It is, however, closely related to monotone regression, and the following theorem shows that we can find the $\hat{x}_i$ for this approach to ties by using the algorithm for the secondary approach. We define $\hat{x}_j^{\S}$ as in the previous section.

Theorem.

If $i \in I_j$ then $\hat{x}_i = x_i + (\hat{x}_j^{\S} - x_j^{\S})$.

Proof.

We have the partition

$$S(y) = \sum_{j=1}^{m} w_j^{\S}(x_j^{\S} - y_j^{\S})^2 + \sum_{j=1}^{m} \sum (w_i\{(x_i - x_j^{\S}) - (y_i - y_j^{\S})\}^2 \mid i \in I_j).$$

The first term is minimized under the order restrictions by setting $y_j^{\S} = \hat{x}_j^{\S}$; the second term vanishes (and is consequently minimized) if we set $y_i = x_i + (y_j^{\S} - x_j^{\S})$ for all $i \in I_j$. Thus the overall minimum is attained if we use $\hat{x}_i$ as stated in the theorem. □

Computationally, this third approach is very easy to implement. We use the algorithm for the secondary approach to compute the optimal block means, and then adjust all y_i within the block by the same constant so that they have the appropriate mean.

REFERENCE NOTE

1. Van Eeden, C. *Testing and estimating ordered parameters of probability distributions.* Unpublished doctoral dissertation, University of Amsterdam, 1958.

REFERENCES

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., & Brunk, H. D. *Statistical inference under order restrictions.* New York: Wiley, 1972.

Bradley, R. A., Katti, S. K., & Coons, I. J. Optimal scaling for ordered categories. *Psychometrika,* 1962, *27,* 355–374.

Hayashi, C. Minimum dimension analysis MDA: One of the methods of multidimensional quantification. *Behaviormetrika,* 1974, *1,* 1–24.

Kruskal, J. B. Nonmetric multidimensional scaling: A numerical method. *Psychometrika,* 1964, *29,* 115–129.

Manuscript received 6/16/75
Final version received 4/28/76