The algorithm for calculating unimodal isotonic regression in `Iso`

Rolf Turner

For `Iso` version `0.0-18`

Abstract

The `Iso` package provides an algorithm for applying isotonic regression to data having an underlying unimodal structure. This algorithm consists essentially of "divide and conquer" approach to this class of isotonic regression problems. Repeated application of the algorithm permits the estimation of the location of the maximum of a data set assumed to have an underlying unimodal structure. This estimation procedure is "easily" (for some value of the word "easily") shown to be consistent. The performance of the resulting procedure for locating a maximum has been assessed through a simulation study described in one of the references. This document supplies some of the background on the algorithm used calculating unimodal isotonic regression and gives a theoretical justification of why this algorithm works.

Contents

1 Introduction

Algorithms for implementing isotonic regression under orderings other than the simple linear order are difficult to construct. The best known of such algorithms is the Maximum Lower Sets algorithm [1, p. 24]. This algorithm is complicated and hard to program. It is also reputed to run rather slowly, and indeed the number of operations required grows exponentially in certain cases.

The motivation for developing an improved algorithm for performing such regressions came in part from a data set being studied by members of the Faculty of Forestry at the University of New Brunswick. These data consisted of observations which had been made of the "vigour" of growth of five stands of black spruce. The stands each had different initial tree densities. It was expected that vigour would initially increase (as the trees increased in size) and then level off and start to decrease as the growing trees encroached upon each others' space and competed more strongly for resources such as moisture, nutrients, and light. It was further expected that the position of the mode of the vigour observations would depend upon the initial densities.

Plots of the data did not make it completely clear as to where the leveling-off point or mode occurred; the Forestry researchers requested a procedure for determining the location of this mode. A procedure which comes immediately to mind is to fit unimodal isotonic regressions with mode at each of the possible locations in turn. The location yielding minimal error sum of squares is then chosen as the location of the mode. It is thus desirable to be able to perform a large number of unimodal isotonic regressions quickly and efficiently.

Formally the unimodal isotonic regression problem may be stated as follows: Suppose that Y_{ij}, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$, are independent random variables such that $Y_{ij} = \mu_i + E_{ij}$ for all i and j, where the E_{ij} have mean 0 and variance σ^2. Further suppose that the μ_i have a *unimodal ordering*, i.e. that

$$\mu_1 \leq \mu_2 \leq \ldots \leq \mu_{k_0} \geq \mu_{k_0+1} \geq \ldots \geq \mu_p \tag{1}$$

for some k_0, $1 \leq k_0 \leq p$. Of course if $k_0 = p$ then we have the usual linear isotonic regression problem and if $k_0 = 1$ we the linear *decreasing* order isotonic regression problem.

The problem is to estimate the values of $\mu_1, \ldots, \mu_p$. The (weighted) least squares estimates of the μ_i are given by minimizing

$$SS = \sum_i \sum_j (Y_{ij} - \hat{\mu}_i)^2 w_i$$

subject to the constraint (1), where $w_1, \ldots, w_p$ are a (given) set of positive weights. This problem may initially be subdivided into three sub-problems involving only *linear* orderings: (a) estimating $\mu_1, \ldots, \mu_{k-1}$; (b) estimating μ_k; and (c) estimating $\mu_{k+1}, \ldots, \mu_n$. Sub-problem (b) is of course trivial as it stands, and sub-problems (a) and (c) can be solved by standard and well-known techniques. The question is how to combine the solutions of the three subproblems appropriately.

The answer is essentially to "interleaf" the estimates resulting from solving sub-problems (a) and (c) in *numerical* order, tack on $\hat{\mu}_k = \bar{Y}_{k\cdot}$ at the upper end, solve the corresponding isotonic regression with respect to the resulting linear ordering, and then put the estimates back in their original order.

In the next section we make this answer slightly more precise and demonstrate that it is indeed correct. The idea may be generalised to other partial orderings and to other "tree-like" structures as well as to unimodal ones but we will not elaborate on the details.

2 Notation and Terminology, and the Main Result

Let $k_0 \in S = \{1, \ldots, p\}$ be given (to avoid trivialitie assume $1 < k_0 < p$ and let $\prec$ be the partial order on given by $x \prec y$ if either $x \leq y \leq k_0$ or $x \geq y \geq k_0$. If $x < k_0$ and $y > k_0$ or vice versa then x and y are not comparable under $\prec$.
Recall that an isotonic function (with respect to the partial order $\prec$) is a (real-valued) function f such that $x \prec y$ implies $f(x) \leq f(y)$. If g is an arbitrary function on S, and w is a non-negative (weight) function on S, then the *isotonic regression* of g, with respect to $\prec$ and w, (denoted g_*) is that value of $\hat{g}$ which minimizes

$$\sum_{s \in S} [g(s) - \hat{g}(s)]^2 w(s)$$

over all *isotonic* functions $\hat{g}$.
Let S_1 and S_2 be two subsets of S. We say that S_2 *follows* S_1, (in symbols $S_1 \prec\prec S_2$) if $x \prec y$ for every x in S_1 and every y in S_2.
Let $S_1 = \{k \in S \mid k \neq k_0\}$ and $S_2 = \{k_0\}$. Let g_1 be the restriction of g to S_1, and let g_{1*} be the isotonic regression of g_1r. The weight function used to form g_{1*} is of course the restriction of the overall weight function w to S_1.
An elementary but important fact about isotonic regression is that g_* takes the form

$$g_*(s) = c_i \text{ on } L_i,\ i = 1, \ldots, r$$

where $L_1, \ldots, L_r$ form a disjoint and exhaustive collection of subsets of S, and $c_1 < c_2 < \ldots < c_r$. Moreover c_i is the weighted mean over L_i of the values of $g(s)$; i.e.

$$c_i = \frac{\sum_{s \in L_i} w(s) g(s)}{\sum_{s \in L_i} w(s)} \quad .$$

(See [1, p. 18, Theorem 1.3.5].) We call the sets L_i the *level* sets and the values c_i the *level* values of the isotonic regression.
Let the level sets and level values for g_{1*} be $L_1, \ldots, L_r$ and $c_1 < \ldots < c_r$, and let $L_{r+1} = \{k_0\}$ and let $c_{r+1} = g(k_0)$. Define a function f on $\{1, \ldots, r+1\}$, by $f(t) = c_t$ for $t = 1, \ldots r+1$, and a weight function u by

$$u(t) = \sum_{x \in L_t} w(x) \quad .$$

Theorem 1: Let f and u be as given above. Let f_* be the isotonic regression of f with respect to the usual order on $\{1, \ldots, r+1\}$ and the weight function u. Then the isotonic regression of g with respect to $\prec$ and w is given by

$$g_*(s) = f_*(t) \text{ for } s \in L_t \quad .$$

Remark: Note that S_1 consists of the two disjoint sets $\{1, \ldots, k-1\}$ and $\{k+1, \ldots, n\}$ which are unrelated with respect to $\prec$. It is easy to see (and well-known; see, e.g. [1, p. 57]) that an isotonic regression on their union is simply the amalgamation of separate isotonic regressions on each component. That is g_1* is obtained by doing an "ordinary" isotonic regression of the restriction of g to $\{1, \ldots, k-1\}$ and an isotonic regression of the restriction of g to $\{k+1, \ldots, p\}$ with respect to decreasing order on this set.
To prove Theorem 1 we require the following definitions and a couple of preliminary lemmas.
Definition: For any constant c we define

$$\mathcal{I}^c = \{g | g \text{ is isotonic and } g(s) \leq c \text{ for all } s \in S\}$$

and

$$\mathcal{I}_c = \{g | g \text{ is isotonic and } g(s) \geq c \text{ for all } s \in S\} \ .$$

Let $g_*(s)$ be the isotonic regression of g and define

$$g_{cu}(s) = \begin{cases} g_*(s) & \text{if } g_*(s) \leq c \\ c & \text{if } g_*(s) > c \ . \end{cases}$$

Lemma 1: The function g_{cu} uniquely minimizes

$$\sum_{s \in S} [g(s) - \hat{g}(s)]^2 w(s) \tag{2}$$

subject to $\hat{g} \in \mathcal{I}^c$.

Proof: For any $\hat{g}$ in $\mathcal{I}^c$,

$$\begin{aligned} \sum_{s \in S} [g(s) - g_{cu}(s)][g_{cu}(s) - \hat{g}(s)] w(s) &= \sum_{s \in S} [g(s) - g_*(s)][g_{cu}(s) - g_*(s)] w(s) \\ &\quad + \sum_{s \in S} [g_*(s) - g_{cu}(s)][g_{cu}(s) - \hat{g}(s)] w(s) \\ &\quad + \sum_{s \in S} [g(s) - g_*(s)][g_*(s) - \hat{g}(s)] w(s) \\ &= \Sigma_1 + \Sigma_2 + \Sigma_3 \end{aligned}$$

Now $\Sigma_1 = 0$ by [1, Theorem 1.3.6, p. 23] since $g_{cu}(s) - g_*(s)$ is a function of $g_*(s)$. It is also true that $\Sigma_3 \geq 0$ since g_* is the isotonic regression of g (applying [1, Theorem 1.3.1, p. 15]). Finally

$$\begin{aligned} \Sigma_2 &= \sum_{g_*(s) > c} [g_*(s) - g_{cu}(s)][g_{cu}(s) - \hat{g}(s)] w(s) \\ &= \sum_{g_*(s) > c} [g_*(s) - c][c - \hat{g}(s)] w(s) \geq 0 \ . \end{aligned}$$

Since $\mathcal{I}^c$ is a convex lattice we may apply the converse part of [1, Theorem 1.3.1, p. 15] and the result follows. ∎

Exactly analogous to Lemma 1 is

Lemma 2: The function

$$g_{cl}(s) = \begin{cases} g_*(s) & \text{if } g_*(s) \geq c \\ c & \text{if } g_*(s) < c \ . \end{cases}$$

uniquely minimizes (2) for $\hat{g} \in \mathcal{I}_c$.

Lemma 3, given below, is an immediate consequence of Lemma 1 and 2:

Lemma 3: Let $c_{k_1}, \ldots, c_{k_m}$ be a subset of the level values of g_*, and let

$$S' = S \setminus \bigcup_{l=1}^{m} \{s | g_*(s) = c_{k_l}\} \neq \phi$$

The isotonic regression of g restricted to S' is g_* restricted to S'.

We can now prove the main result:

Proof of Theorem 1: Since $x \prec k_0$ for all $x \in S_1$ it is easy to see that there is a constant c such that:

$$\begin{aligned} g_*(s) &< c \text{ implies } s \in S_1 \text{and} \\ g_*(s) &> c \text{ implies } s = k_0 \ . \end{aligned}$$

The set $\{s|g(s) = c\}$ may contain elements from S_1 and may contain k_0 as well.
For this c

$$g_*(s) = \begin{cases} g_{cu}(s) & \text{if } s \in S_1 \\ g_{cl}(s) & \text{if } s = k_0 \end{cases}$$

otherwise we would contradict the definition of g_*. Applying Lemmas 1 and 2, it follows that

$$g_{cu}(s) = \begin{cases} g_{1*}(s) & \text{if } g_{1*}(s) < c \\ c & \text{if } g_{1*}(s) \geq c \end{cases}$$

for $s \in S_1$ and

$$g_{lu}(s) = \begin{cases} c & \text{if } g_{2*}(s) \leq c \\ g_{2*}(s) & \text{if } g_{2*}(s) > c \end{cases}$$

for $s \in S_2$. Therefore $g_*(s)$ is a function of $g_{1*}(s)$ on S_1. In other words, $g_*(s)$ is constant on all of the level sets L_i of g_{1*}. (Since L_{r+1} consists of the single point k_0, $g_*(s)$ is trivially constant on L_{r+1}.) Let $g_*(s) = d_i$ on L_i for $i = 1, \ldots, r+1$. Now

$$\begin{aligned} \sum_S [g(s) - g_*(s)]^2 w(s) &= \sum_{S_1} [g(s) - g_{1*}(s) + g_{1*}(s) - g_*(s)]^2 w(s) \\ &\quad + \sum_{S_2} [g(s) - g_{2*}(s) + g_{2*}(s) - g_*(s)]^2 w(s) \\ &= \sum_{S_1} [g(s) - g_{1*}(s)]^2 w(s) + \sum_{S_2} [g(s) - g_{2*}(s)]^2 w(s) \\ &\quad + \sum_{S_1} [g_{1*}(s) - g_*(s)]^2 w(s) + \sum_{S_2} [g_{2*}(s) - g_*(s)]^2 w(s) \\ &\quad + 2 \sum_{S_1} [g(s) - g_{1*}(s)][g_{1*}(s) - g_*(s)] w(s) \\ &\quad + 2 \sum_{S_2} [g(s) - g_{2*}(s)][g_{2*}(s) - g_*(s)] w(s) \end{aligned}$$

The last two terms are zero by [1, Theorem 1.3.1, p. 15] since $g_{1*}(s) - g_*(s)$ is a function of $g_{1*}(s)$, and $g_{2*}(s) - g_*(s)$ is a function of $g_{2*}(s)$. The first two terms do not involve $g_*(s)$. Hence $g_*(s)$ minimizes

$$\sum_{S_1} [g_{1*}(s) - g_*(s)]^2 w(s) + \sum_{S_2} [g_{2*}(s) - g_*(s)]^2 w(s) \tag{3}$$

and hence is the isotonic regression of

$$h(s) = \begin{cases} g_{1*}(s) & \text{if } s \in S_1 \\ g(s) & \text{if } s = k_0 \end{cases}$$

It follows readily that the values of $g_*(s)$ on L_i, i.e. d_i, are in increasing order. Since $g_*(s)$ minimizes (3), equal to

$$\sum_{t=1}^{r} [c_t - d_t]^2 u(t)$$

under the assumption that g_* is isotonic, it follows that $d_1, d_2, \ldots, d_r$ minimize this expression under simple linear order on $1, 2, \ldots, r$, and hence $d_t = f_*(t)$ for all t. ∎

3 Estimating the Location of a Maximum

3.1 Consistency

Let Y_{ij} and w_i, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$, be as described in Section 1. Suppose that the value of k_0 is unknown and one wishes to estimate it in some rational manner. The (weighted) least squares estimate of k_0 may be determined by assuming that $k_0 = k$ for each $k = 1, \ldots, p$ and finding the (weighted) least squares estimates of the μ_i, say $\hat{\mu}_i(k)$ under this assumption. Let $SS(k)$ be the corresponding error sum of squares, i.e.

$$SS(k) = \sum_i \sum_j (Y_{ij} - \hat{\mu}_i(k))^2 w_i$$

The estimated value of k_0 is then that value of k which minimizes $SS(k)$.
If we assume that the mode is a strict one, i.e. that

$$\mu_1 \leq \mu_2 \leq \ldots \leq \mu_{k_0-1} < \mu_{k_0} > \mu_{k_0+1} \geq \ldots \geq \mu_p \,, \tag{4}$$

then it is not hard to demonstrate that this procedure yields a consistent estimate of k_0. We will not go into the details here.
There are other "obvious" ways of estimating the location of the maximum of a theoretical function underlying an observed data set. These include using the maximum of a fitted quadratic function or the single knot of a fitted "broken stick" (piecewise linear) model. The performance of unimodal isotonic regression is compared with these and other methods in [2].

3.2 Estimating a maximum in `Iso`

For a given data set, the `Iso` function `ufit` ("unimodal fit") calculates the best (least squares) unimodal fit with mode at a specified location given by the argument `lmode` ("location of mode"). If `lmode` is unspecified (i.e. left with its default value of `NULL`) then `ufit` searches over all possible modal locations and chooses that which yields the minimal error sum of squares.
The search is feasible since there are a finite and limited number of possibilities for the modal location. If the largely notional "predictor" vector is `x` then the possible modal locations are `x[i]`, with `i` running from `1` to `n` = `length(x)` and `(x[i] + x[i+1])/2` with `i` running from `1` to `n-1`. Note that all possible modal locations that are strictly between `x[i]` and `x[i+1]` are equivalent, so we restrict attention to the midpoints.
The possibilities are even more limited than that, however. Suppose that the optimal mode is at `x[i]` with `i` > 1. This says that the correponding isotoniation of `y`, `y*` say, is increasing on `x[1]` to `x[i]` and decreasing on `x[i]` to `x[n]`. Let the corresponding error sum of squares be SSE_i. Now consider the isotonisation of `y` with respect to the unimodal structure with mode at `(x[i-1]+x[i])/2`, say `y**` and let the corresponding error sum of squares be $\text{SSE}_{i-0.5}$. Note that `y*` satisfies the unimodal constraint that `y**` has to satisfy and hence $\text{SSE}_{i-0.5} \leq \text{SSE}_i$. But SSE_i is minimal over all possible modal locations, whence $\text{SSE}_i \leq \text{SSE}_{i-0.5}$ and so SSE_i is equal to $\text{SSE}_{i-0.5}$.
If the optimal mode is at `x[1]` then similar reasoning shows that SSE_1 is equal to $\text{SSE}_{1.5}$. Thus to find the optimal mode we need only search over the "half-points" `(x[i] + x[i+1])/2`, `i` running from `1` to `n-1`
If values of `y` are only meaningful at `x[1]`, ..., `x[n]`, e.g. if the values of `y` are some sort of annual amount or annual maximum, then the "half-points" only constitute a computational device and the optimal mode would be said to occur at the "whole-point" `x[i]` having the co-minimal value of SSE.

Note that if in searching over the "half-points" we find the minimal sum of squares to be at `(x[i] + x[i+1])/2`, then either `x[i]` or `x[i+1]` will give rise to a co-minimal value of SSE. Letting `y*` be the isotonisation of `y` corresponding to a mode at `(x[i] + x[i+1])/2`, we see that if `y*[i]` $\geq$ `y*[i+1]` then `y*` is also the isotonisation of `y` corresponding to a mode at `x[i]`. In this case `x[i]` will be an optimal modal location. Likewise if `y*[i]` $\leq$ `y*[i+1]` then `y*` is also the isotonisation of `y` corresponding to a mode at `x[i+1]`. In this case `x[i+1]` will be an optimal modal location.
If `y` consists of response values which can be observed over a continuum of `x` values but which *was* observed only at `x[1]`, ..., `x[n]`, then it is meaningful for the response in question to have a mode at a "half-point". In this case there is ambiguity — there are always (at least) two "optimal" modal locations.

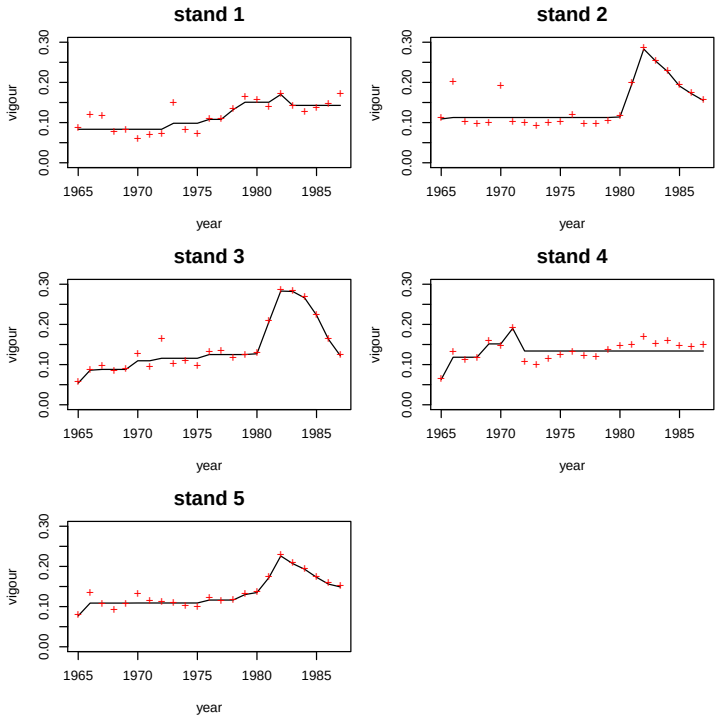

Figure 1: Unimodal isotonisation of growth vigour for each of five stands of spruce trees over the years 1965 to 1987. The black line represents the optimal unimodal isotonic fit. The red + symbols represent the raw data.

3.3 Examples

Consider the data set `vigour` which is included in the `Isopackage`. We can find the optimal location of maximum vigour over the years 1965 to 1987 for each stand. The code to fit the isotonic models and plot the graphs of the fits follows. The resulting plots are shown in Figure 1.

```
> par(mfrow=c(3,2),mar=c(4,4,3,1))
> for(i in 2:6) {
+    plot(ufit(vigour[,i],x=vigour[,1]),type="l",ylim=c(0,0.3),
+         xlab="year",ylab="vigour",main=paste("stand",i-1),cex.main=1.5)
+    points(vigour[,1],vigour[,i],pch="+",col="red")
+ }
```

Note that in this setting the "vigour" values are determined in terms of an annual growth cycle whence they make sense only for integrer values of "year". Hence "half=point" modes are not meaningful.
It may also be of interest to look for the optimal unimodal fit to the mean, over stands. A plot of the resulting fit is shown in Figure 2.

```
>     xm <- apply(vigour[,2:6],1,mean)
>     par(mar=c(4,4,3,1))
>     plot(ufit(xm,x=vigour[,1]),type="l",ylim=c(0,0.3),
+          xlab="year",ylab="vigour",main="Mean over stands",cex.main=1.5)
>     points(vigour[,1],xm,pch=22,col="red")
>     for(i in 2:6) points(vigour[,1],vigour[,i],pch="+",col="blue")
```

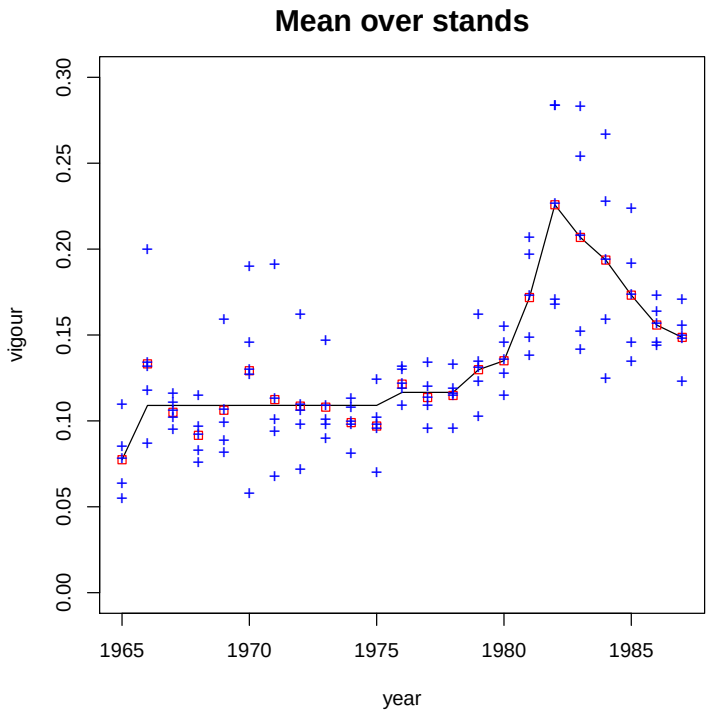

Figure 2: Unimodal isotonisation of the mean growth vigour over five stands of spruce trees for the years 1965 to 1987. The black line represents the optimal unimodal isotonic fit. The blue □ symbols represent the raw means. The red + symbols represent the data for all of the individual stands.

Acknowledgement: The author would like to thank Kirk Schmidt, a graduate student in the Department of Forest Engineering, U.N.B., and his advisor Professor Ted Needham, for drawing the problem on tree growth vigour to his attention.

References

[1] T. Robertson, F. T. Wright, and R. L. Dykstra. *Order Restricted Statistical Inference*. Wiley, New York, 1988.

[2] T. R. Turner and P. C. Wollan. Locating a maximum using isotonic regression. *Computational Statistics & Data Analysis*, 25(3):305–320, 1997.