Monotonic Regression

JAN DE LEEUW

Volume 3, pp. 1260–1261

in

Encyclopedia of Statistics in Behavioral Science

ISBN-13: 978-0-470-86080-9
ISBN-10: 0-470-86080-4

Editors

Brian S. Everitt & David C. Howell

Monotonic Regression

In linear regression, we fit a linear function $y = \alpha + \beta x$ to a scatterplot of n points (x_i, y_i). We find the parameters α and β by minimizing

$$\sigma(\alpha, \beta) = \sum_{i=1}^{n} w_i (y_i - \alpha - \beta x_i)^2, \tag{1}$$

where the w_i are known positive weights (*see* **Multiple Linear Regression**).

In the more general **nonlinear regression** problem, we fit a nonlinear function $\phi_\theta(x)$ by minimizing

$$\sigma(\theta) = \sum_{i=1}^{n} w_i (y_i - \phi_\theta(x_i))^2 \tag{2}$$

over the parameters θ. In both cases, consequently, we select the minimizing function from a family of functions indexed by a small number of parameters.

In some statistical techniques, low-dimensional parametric models are too restrictive. In nonmetric **multidimensional scaling** [3], for example, we can only use the rank order of the x_i and not their actual numerical values. Parametric methods become useless, but we still can fit the best fitting monotone (increasing) function nonparametrically. Suppose there are no ties in x, and the x_i are ordered such that $x_1 < \cdots < x_n$. In monotone regression, we minimize

$$\sigma(z) = \sum_{i=1}^{n} w_i (y_i - z_i)^2 \tag{3}$$

over z, under the linear inequality restrictions that $z_1 \leq \cdots \leq z_n$. If the solution to this problem is $\hat{z}$, then the best fitting increasing function is the set of pairs $(x_i, \hat{z}_i)$. In monotone regression, the number of parameters is equal to the number of observations. The only reason we do not get a perfect solution all the time is because of the order restrictions on z.

Actual computation of the best fitting monotone function is based on the theorem that if $y_i > y_{i+1}$, then $\hat{z}_i = \hat{z}_{i+1}$. In words: if two consecutive values of y are in the wrong order, then the two corresponding consecutive values of the solution $\hat{z}$ will be equal. This basic theorem leads to a simple algorithm, because knowing that two values of $\hat{z}$ must be equal reduces the number of parameters by one. We thus have a monotone regression problem with $n - 1$ parameters. Either the elements are now in the correct order, or there is a violation, in which case we can reduce the problem to one with $n - 2$ parameters, and so on. This process always comes to an end, in the worst possible case when we only have a single parameter left, which is obviously monotone.

We can formalize this in more detail as the up-and-down-blocks algorithm of [4]. It is illustrated in Table 1, in which the first column is y. The first violation we find is $3 > 0$, or 3 is not up-satisfied. We merge the two elements to a block, which contains their weighted average 3/2 (in our example all weights are one). But now $2 > (3/2)$, and thus the new value $3/2$ is not down-satisfied. We merge all three values to a block of three and find $5/3$, which is both up-satisfied and down-satisfied. We then continue with the next violation. Clearly, the algorithm produces a decreasing number of blocks. The value of the block is computed using weighted averaging, where the weight of a block is the sum of the weights of the elements in the block. In our example, we wind up with only two blocks, and thus the best fitting monotone function $\hat{z}$ is a step function with a single step from $5/3$ to 4.

The result is plotted in Figure 1. The line through the points x and y is obviously the best possible fitting function. The best fitting monotone function, which we just computed, is the step function consisting of the two horizontal lines.

If x has ties, then this simple algorithm does not apply. There are two straightforward adaptions [2]. In the primary approach to ties, we start our monotone regression with blocks of y values corresponding to the ties in x. Thus, we require tied x values to correspond with tied z values. In the secondary approach, we pose no constraints on tied values, and it can be shown that in that case we merely have to order the y values such that they are increasing

Table 1 Computing monotone regression

y	$\rightarrow$	$\rightarrow$	$\rightarrow$	$\hat{z}$
2	2	$\frac{5}{3}$	$\frac{5}{3}$	$\frac{5}{3}$
3	$\frac{3}{2}$	$\frac{5}{3}$	$\frac{5}{3}$	$\frac{5}{3}$
0	$\frac{3}{2}$	$\frac{5}{3}$	$\frac{5}{3}$	$\frac{5}{3}$
6	6	6	6	4
6	6	6	3	4
0	0	0	3	4

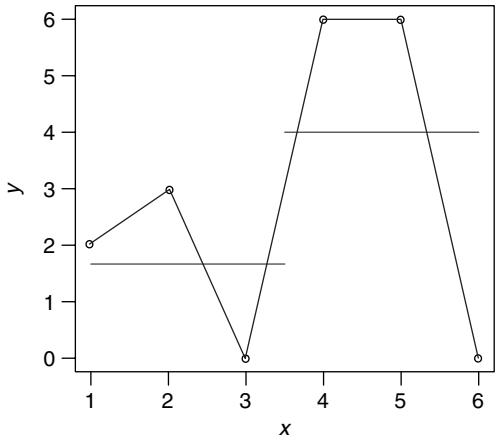

Figure 1 Plotting monotone regression

in blocks of tied x values. And then we perform an ordinary monotone regression.

Monotone regression can be generalized in several important directions. First, basically the same algorithm can be used to minimize any separable function of the form $\sum_{i=1}^{n} f(y_i - z_i)$, with f any convex function with a minimum at zero. For instance, f can be the absolute value function, in which case we merge blocks by computing medians instead of means. And second, we can generalize the algorithm from weak orders to partial orders in which some elements cannot be compared; for details, see [1]. Finally, it is sometimes necessary to compute the least squares monotone regression with a nondiagonal weight matrix. In this case, the simple block merging algorithms no longer apply, and more general quadratic programming methods must be used.

References

[1] Barlow, R.E., Bartholomew, R.J., Bremner, J.M. & Brunk, H.D. (1972). *Statistical Inference under Order Restrictions*, Wiley, New York.

[2] De Leeuw, J. (1977). Correctness of Kruskal's algorithms for monotone regression with ties, *Psychometrika* **42**, 141–144.

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

[4] Kruskal, J.B. (1964b). Nonmetric multidimensional scaling: a numerical method, *Psychometrika* **29**, 115–129.

JAN DE LEEUW