



Research Report
Department of Statistics
Göteborg University
Sweden

**Least squares estimates of regression
functions with certain monotonicity
and concavity/convexity restrictions**

Ulla Dahlbom

**Research Report 1998:6
ISSN 0349-8034**

Mailing address:	Fax	Phone	Home Page:
Department of Statistics	Nat: 031-773 12 74	Nat: 031-773 10 00	http://www.stat.gu.se
Göteborg University	Int: +46 31 773 12 74	Int: +46 31 773 10 00	
Box 660			
SE 405 30 Göteborg			
Sweden			

Least squares estimates of regression functions with certain monotonicity and concavity/convexity restrictions.

Summary: In all regression problems the choice of model and estimation method is due to a priori information about the regression function. In some situations it is motivated to consider regression functions with specific non-parametric characteristics, for instance monotonicity and/or concavity/convexity.

We propose a new least squares estimation method for curves that fulfil monotonicity and concavity/convexity restrictions. The least squares estimate of such a regression function is a piecewise linear continuous function with bending points contained in the set of the observed values of the independent variable.

The set of bending points, which makes the function a least squares solution can be determined by an iterative algorithm within a finite number of steps.

Keywords: Concave/convex, non-parametric regression, piecewise linear.

1. Introduction.

When you deal with regression problems the choice of model and estimation method is due to the a priori information about the regression function. In some applications the shape of the regression function is known and the estimation problem is reduced to an estimation of some parameters. But often this is not the case. Instead the relationship between the dependent and independent variables is determined by some constraints on the variables. A simple characteristic is monotonicity. In this case the only assumption about the regression function is, that it has a non-decreasing phase or/and a non-increasing phase. If the regression function consists of one non-decreasing (or non-increasing) phase only

a suitable estimation method is isotonic regression. This is a non-parametric regression method that is very often referred to. The basic theory of isotonic regression is described in Barlow et al (1972) and Robertson et al (1989). In case of unimodality the same basic technique is used for two pieces of data.

Another simple shape characteristic is concavity (or convexity). In many applications it is motivated to consider regression functions with both monotonicity and concavity/convexity restrictions. The concave regression problem was first formulated by Hildreth (1954) for the estimation of marginal productivity curves. He adopted the LS estimation method and formulated it as a quadratic programming problem. Dent (1973) and Holloway (1979) continued the work and the concave regression function was obtained by a more general framework of quadratic programming with linear inequality constraints. The consistency of these LS concave regression estimators has been proved by Hanson and Pledger (1976). Wu (1982) proposed two algorithms for concave regression. One involves some quadratic programming. Unfortunately this algorithm has no assured convergence. The other is just an approximation, but very easy to implement. An iterative method for regression with convexity and monotonicity restrictions was proposed by Holm and Frisén (1985). This method gives the LS estimate in a finite number of steps. Fraser and Massam (1987) formulated an algorithm using cylinder projections and dual spaces. Theoretically they obtained the LS solution but the estimation method involves a matrix inversion which in practice seems to imply numerical instability in the obtained results.

In this report we propose a new algorithm for all kinds of regression problems with different concavity/convexity restrictions, such as increasing and concave, unimodal and sigmoid regression. It is an iterative procedure for determining least

squares estimates. The proposed estimation method is closely related to the estimation method proposed by Holm and Frisén (1985). The iteration with an inclusion and an exclusion procedure is identical for both estimation methods. In Holm and Frisén the estimation problem is solved by using a function system including a matrix inversion. In our proposed method we instead use restrictions on the differences between two successive regression coefficients. Using these differences it is easier to extend the problem from increasing and concave regression to unimodal and sigmoid regression. These two later problems are mentioned but not studied by Holm and Frisén.

In this report only non-parametric concave regression is discussed. Unimodal and sigmoid regression together with further properties are examined in Dahlbom (1994).

2. The estimation procedure.

The purpose of this paper is to find a regression function with simple curve characteristics like isotonic restrictions and concavity. Thus for the observations $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ with $x_1 < x_2 < \dots < x_N$ and with non-negative weights w_1, w_2, \dots, w_N we will search for a regression function $f(x_i)$ that minimizes

$$\sum_{i=1}^N (y_i - f(x_i))^2$$

and satisfies the restrictions. First we check if the observations themselves fulfil the non-decreasing and concave restrictions. If this is the case, $f(x_i) = y_i, i=1, \dots, N$, is the solution to the problem.

But usually this is not the case and instead we have to use some estimation procedure. When a regression function is monotonic and concave one condition is fulfilled, namely

$$\frac{\hat{y}_{i+1} - \hat{y}_i}{\hat{x}_{i+1} - \hat{x}_i} \geq \frac{\hat{y}_{i+2} - \hat{y}_{i+1}}{\hat{x}_{i+2} - \hat{x}_{i+1}} \geq 0 \quad \text{for } i=1, \dots, (N-2)$$

Equality in the left inequality means that two or more estimated observations will lie on a straight line. This means that the regression function consists of continuous linear pieces where the estimate of the regression coefficient for each linear part will change monotonically. The continuous regression function will bend in some x-values and for this reason they will be called 'bending points'.

A major problem in the estimation method is to find the bending points that minimizes the sum of squares. We start describing the proposed estimation procedure by introducing some notations:

Suppose that the observations are divided into M ordered groups of some sizes, where $x_1 \leq \dots \leq x_{i_1}$ belong to the first group, $x_{i_1+1} \leq \dots \leq x_{i_2}$ to the second and so on.

The smallest x-observation in each group will have a certain rule in the construction and it is called the bending point for the corresponding x-interval.

Since bending points will appear a few times in the description of the estimation method they will have a special notation, $x_0^{(k)}$, where "k" is the interval in which the bending point lies, ie $x_0^{(1)} = x_1$, $x_0^{(2)} = x_{i_1+1}$, $x_0^{(3)} = x_{i_2+1}$ and so on.

We need also a notation for a predicted y-value corresponding to the bending point. Therefore we introduce a notation according to the following:

Let $x_0^{(k)}$ be the bending point of the k:th x-interval. Predict a y-value, $y_0^{(k)}$, at $x_0^{(k)}$ using the weighted least squares estimate of the regression line in the k:th x-interval. The estimate of the regression line is obtained from the observed y-values in this x-interval together with a predicted y-value, corresponding to the bending point $x_0^{(k-1)}$, with a weight $w_0^{(k-1)}$, obtained from the regression line in the (k-1):th x-interval, which will be described later.

DEF: A predicted point, $(x_0^{(k)}, y_0^{(k)})$, where $x_0^{(k)}$ is the bending point of the k:th x-interval and $y_0^{(k)}$ is the corresponding y-value, predicted according to the above description, is called the fiction point from the k:th x-interval.

We first demonstrate how the procedure works on some fixed bending points. This case may result in a solution that does not fulfil the isotonic and concave restrictions.

2.1. Fixed bending points.

We study first the problem of estimating (by least squares) a piecewise linear continuous regression function with fixed bending points $x_0^{(k+1)} = x_{i_k+1}$, $k = 1, \dots, (M-1)$.

Consider the first interval with observation points x_1, x_2, \dots, x_i . A straight-forward calculation shows that the minimal sum of squares for these observation points for a line through a point $(x_0^{(2)}, y) = (x_{i+1}, y^*)$ is a quadratic function of y^* . This function

has a minimum for $y^* = y_0^{(2)}$, the value of the ordinary regression line for the observation points, x_1, x_2, \dots, x_i , taken in the point $x_0^{(2)}$. The coefficient of the square in the function of y^* equals

$$w_0^{(2)} = \frac{\sum_{i=1}^{i_1} w_i \sum_{i=1}^{i_1} [x_i - \bar{x}_1]^2 w_i}{\sum_{i=1}^{i_1} [x_i - x_0^{(2)}]^2 w_i}$$

where \bar{x}_1 denotes the weighted mean, x_1, x_2, \dots, x_i . This means that if the rest of the regression function (for $x \geq x_0^{(2)}$) were determined the sum of squares from the first interval would be equivalent to a constant plus the influence of an artificial observation in the point $(x_0^{(2)}, y_0^{(2)})$ with weight $w_0^{(2)}$.

Next we can analogously study the sum of squares from the second interval $x_{i_1+1}, \dots, x_{i_2}$ for a regression line through a point $(x_0^{(3)}, y^*)$ when the artificial point from the previous interval is included. Again it is a quadratic function with minimum for y^* equal to the value of the regression line. The coefficient of the square in the function of y^* equals

$$w_0^{(3)} = \frac{\sum_{i=i_1+1}^{i_2} w_i \sum_{i=i_1+1}^{i_2} [x_i - \bar{x}_2]^2 w_i}{\sum_{i=i_1+1}^{i_2} [x_i - x_0^{(3)}]^2 w_i}$$

where \bar{x}_2 denotes the weighted mean of the ordinary x -observations in the second interval including the bending point $x_0^{(2)}$. Thus the third x -interval now contains an

extra observation, which includes all information about the sum of squares in the two first x-intervals. This procedure will continue in the same way up to the last x-interval, where we beyond the observed points will have a fiction point with a calculated weight which contains all information of the sum of squares from all former x-intervals.

Up to this point we have estimated M straight lines. The only purpose of this is to calculate the weights, $w_0^{(k)}$, and the fiction points $(x_0^{(k)}, y_0^{(k)})$, $k=2, \dots, M$, one in each x-interval. To determine the least squares estimate of the continuous piecewise linear regression we must continue the procedure.

Let \bar{y}_k and \bar{x}_k denote the weighted averages of the y- and x-observations in the k:th x-interval and denote the ordinary weighted least squares estimate of β of a straight line through a predetermined point by $\hat{\beta}_M$. The notation $f_k(x)$ means the LSE of a regression line through a predetermined point.

In the piecewise procedure we start at the last x-interval. In this interval we estimate a straight line for the observed values and the fiction point $(x_0^{(M)}, y_0^{(M)})$.

This means that in the M:th x-interval we obtain $\hat{\beta}_M$ and the estimate of the regression line $f_M(x) = \bar{y}_M + \hat{\beta}_M \cdot (x - \bar{x}_k)$. Calculate the value of $f_M(x)$ for the bending point, $x_0^{(M)}$, in the last interval. Next determine the least squares estimate of the regression line in the second last x-interval including the fiction observation,

$(x_0^{(M-1)}, y_0^{(M-1)})$, in a way that will make the regression line go through the predicted point, $(x_0^{(M)}, f_M(x_0^{(M)}))$, in the last x-interval. Since the weight, $w_0^{(M-1)}$, of the fiction point $(x_0^{(M-1)}, y_0^{(M-1)})$ contains all information from the sum of squares of the first (M-2) x-intervals, this method will give us the minimum sum of squares in the (M-1):th x-interval taking the first (M-2) x-intervals into account.

The least squares estimate of β will be $\hat{\beta}_{M-1} = K_1/K_2$ where

$$K_1 = \sum_{i=1}^{m_{(M-1)}} [y_i - y_0^{(M)}] [x_i - x_0^{(M)}] w_i + [f_{M-2}(x_0^{(M-1)}) - y_0^{(M)}] [x_0^{(M-1)} - x_0^{(M)}] w_0^{(M-1)}$$

and

$$K_2 = \sum_{i=1}^{m_{(M-1)}} [x_i - x_0^{(M)}]^2 w_i + [x_0^{(M-1)} - x_0^{(M)}]^2 w_0^{(M-1)}$$

The notation $m_{(M-1)}$ means the number of x-observations in the second last x-interval excluding the fiction point, $(x_0^{(M-1)}, y_0^{(M-1)})$, and the weighted sums symbolize the summary of these observations. We have now connected two straight lines, one in each of the two last x-intervals, and the connection point is $[x_0^{(M)}, f_M(x_0^{(M)})] = [(x_0^{(M)}, f_{M-1}(x_0^{(M)}))]$.

Next calculate the value of the regression line $f_{M-1}(x)$ in the bending point, $x_0^{(M-1)}$, of the second last x-interval. Estimate the weighted least squares regression line through this point $[x_0^{(M-1)}, f_{M-1}(x_0^{(M-1)})]$ for the third last x-interval including the fiction

point containing information from the first to the (M-3):th x-interval. Now the straight lines, one in each of the 2:nd last and the 3:rd last x-intervals, are also connected and the connection point is $[x_0^{(M-1)}, f_{M-1}(x_0^{(M-1)})] = [x_0^{(M-1)}, f_{M-2}(x_0^{(M-1)})]$.

This procedure is repeated until all x-intervals are connected with weighted least squares estimates of regression lines through successively calculated points, $[x_0^{(i)}, f_i(x_0^{(i)})]$ and we have a continuous curve. Since the weights $w_0(i)$, $i=2, \dots, M$, contain all information from the sum of squares of the preceding (i-1) x-intervals the procedure will give us the least squares estimate of the whole curve under the concave restrictions. We now have to check if the restrictions 'monotonic' and 'concave' are fulfilled. This is done by checking if $\hat{\beta}_{k-1} > \hat{\beta}_k$ for $k=2, \dots, M$. As mentioned before this does not have to be the case when the bending points are fixed.

We have described the estimation method where the procedure starts by estimating the continuous piecewise linear regression function from the last x-interval to the first. This is in fact one case out of several alternatives, which give the same unique solution. In the addition to the above solution it is also possible to start the whole estimation procedure using continuous piecewise linear regression from the first up to the last x-interval. We can then check if the restrictions, ie monotonic and concave, are fulfilled.

The procedure can just as well start with estimated bending points in each end of an inner interval. We continue by determining a least squares regression function in this x-interval and successively use continuous piecewise linear regression in

each direction, both to the first and the last x-intervals. It is obvious that especially the last method gives some understanding of how the estimation method works. As stated before, in each of these three cases we get the same unique solution.

2.2. Changing bending points.

The whole procedure starts by letting every observed value serve as a bending point. That means that we check if the observations themselves fulfil the isotonic and concave restrictions. If this is not the case we start an estimation procedure that uses an inclusion-exclusion method, which is described in Holm and Frisén (1985) according to the following:

Determine the ordinary weighted least squares estimate of the regression line for all the observations. Thus we have only one bending point, namely x_1 . To examine which other points that could be possible bending points we reformulate the problem. Let

$$f_1(x) = 1$$

$$f_k(x) = \begin{cases} x - x_k & \text{for } x < x_k \\ 0 & \text{for } x \geq x_k \end{cases} \quad k=2, 3, \dots, N$$

We can also write any concave regression function $f(x)$ as $f(x) = \sum_{k=1}^N a_k f_k(x)$, where

a_1 has no restrictions and $a_k \geq 0$ for $k \geq 2$. Thus our weighted least squares

estimate is obtained by minimizing

$$\sum_{j=1}^N \left[y_j - \sum_{k=1}^N a_k f_k(x) \right]^2 w_j$$

1) Inclusion-procedure:

Let $\{I_0\}$ be the set of index-values corresponding to the x-observations that are bending points. When we start the procedure $\{I_0\}$ only contains one value namely $i=1$.

A. Calculate the scalar product between $f_j(x)$ and $f_k(x)$ for every $k \notin \{I_0\}$. Divide the product by the norm of $f_j(x)$:

$$\rho_j = \frac{\left[y - \sum_{k \notin \{I_0\}} a_k f_k(x), f_j(x) \right]}{\left[f_j(x), f_j(x) \right]^{1/2}}$$

A1. If $\rho_j \leq 0$ for every j then we have found the least squares solution.

A2. If $\rho_j > 0$ for some $j \notin \{I_0\}$ denote the largest positive ρ -value by ρ_{j_i} and let the corresponding index, j_i , be included in the set $\{I_0\}$. This index set now contains one further index value and we have also got one further bending point. Use the procedure described for fixed bending points, which corresponds to index values contained in $\{I_0\}$. Estimate $f_k(x) = \alpha_k + \beta_k x$ where $k \in \{I_0\}$. Compare $\hat{\beta}_{k-1}$ to $\hat{\beta}_k$.

i) If the monotonic and concave restrictions are fulfilled, ie if $\hat{\beta}_{k-1} \geq \hat{\beta}_k$ for $k-1, k \in \{I_0\}$ then start again from point A.

ii) If $\hat{\beta}_{k-1} < \hat{\beta}_k$ for some $k-1, k \in \{I_0\}$ then start the exclusion-part.

2) Exclusion procedure:

Go back to the former solution, which was the last solution, where the monotonic and concave restrictions were fulfilled. Denote the estimated β_k -values in this

solution by $\hat{\beta}_k^+$. Let

$$\gamma_k = \hat{\beta}_{k-1}^+ - \hat{\beta}_k^+$$

$$\lambda_k = \hat{\beta}_{k-1} - \hat{\beta}_k.$$

Calculate

$$\varepsilon_k = \frac{\gamma_k}{\gamma_k - \lambda_k}$$

for $k \in \{I_0$ - the index for the last obtained bending point}.

B. If there exists a minimal ε -value < 1 then let $\varepsilon_{jk} = \min \varepsilon_k < 1$, where minimum is obtained for $k \in \{I_0\}$. The index value jk is excluded from the index set $\{I_0\}$ and x_{jk} will no longer be a bending point. Thus we have one bending point less compared to the last time we used the procedure for fixed bending points.

The minimum value ε_{jk} is used to obtain a new estimate of β_k , $k=1, \dots, M$, where the monotonic and concave restrictions are fulfilled. This is obtained by calculating a

new weighted estimate of β_k , denoted by $\hat{\beta}_k^w$, according to $\hat{\beta}_k^w = \varepsilon_{jk}^* \hat{\beta}_k + (1 - \varepsilon_{jk})^* \hat{\beta}_k^+$

where $k \in \{I_0\}$.

Next we start the procedure for fixed bending points again using the bending

points x_k , $k \in \{I_0\}$. Estimate $f_k(x) = \alpha_k + \beta_k x$ where $k \in \{I_0\}$. Compare these new

estimates $\hat{\beta}_{k-1}$ to $\hat{\beta}_k$.

B1. If the monotonic and concave restrictions are fulfilled ie if $\hat{\beta}_{k-1} \geq \hat{\beta}_k$ for

$k-1$, $k \in \{I_0\}$ then start the inclusion procedure again from point A.

B2. If $\hat{\beta}_{k-1} < \hat{\beta}_k$ for some $k-1$, $k \in \{I_0\}$ then continue with the exclusion part. Go

back to the former solution, which was the last solution, $\hat{\beta}_k^w$, where the monotonic and concave restrictions were fulfilled.

Let

$$v_k = \hat{\beta}_{k-1}^w - \hat{\beta}_k^w$$

$$\delta_k = \hat{\beta}_{k-1} - \hat{\beta}_k.$$

Then calculate a new set of ε_k :

$$\varepsilon_k = \frac{v_k}{v_k - \delta_k}$$

After that the exclusion procedure continues from point B.

This inclusion-exclusion part of the estimation method will go on until $\rho_k \leq 0$ for every $k \notin \{l_0\}$ at point A1 in the inclusion procedure. We have then received the least squares estimate of the regression function.

The estimation method is of course not limited to just monotonic and concave regression. We can just as well use it for isotonic and concave regression, isotonic and convex regression, pure convex and suitable combinations of regression functions of the shapes concave-up, concave-down, convex-up and convex-down. But for each of these four cases we must make some adjustments in the continuous piecewise linear regression, use different function systems in the inclusion procedure and also change the restrictions for ρ_k and the comparisons between $\hat{\beta}_{k-1}$ and $\hat{\beta}_k$ in the inclusion-exclusion procedure. These adjustments are discussed in Dahlbom (1994) together with different properties of the estimation procedure.

3. Some properties of the L.S. estimation method.

3.1. Convergence of the algorithm.

The algorithm consists of an inclusion part and an exclusion part. The inclusion part works in a way that gives new bending points, one by one according to a specific criterion. After each obtained bending point the algorithm checks if the form restrictions are fulfilled before it picks another bending point. This procedure gives us an ever smaller sum of squares of errors. The inclusion part continues until the restrictions about the form are not fulfilled.

Then the exclusion part starts. Thus one bending point is excluded according to a special criterion and a new solution is determined. The algorithm checks if the new solution fulfils the form restrictions. If this is not the case, another bending point is excluded. This continues until we obtain a solution which makes the restrictions fulfilled again.

We now start with a new presolution with a smaller sum of squares of errors than the previous one. Since we only have a finite number of possible sets of bending points and since each cycle gives us a smaller sum of squares of errors the procedure will converge to the LS solution in a finite number of steps.

3.2. Consistency.

Consistency in concave regression in general is a very intricate problem. We will present only one consistency property given by Hansen & Pledger (1976).

Let I be an interval on the real line. For each x in I let F_x be a cumulative distribution function with the mean $\mu(x)$. Suppose that $\mu(x)$ is continuous and concave on I . Also

suppose that $Y_j, j=1, \dots, N$ is a sequence of independent random variables which gives Y_k the distribution F_{x_k} . For each positive integer N we can define an estimator $\mu_N(x) = \mu_N(x; Y_j, j=1, \dots, N)$ of $\mu(x)$ which minimizes a sum of squares. In this situation with probability one, $\mu_N(x)$ will converge to $\mu(x)$ uniformly on any subinterval of I and in addition $\mu_N(x)$ satisfies a condition which means that $\mu_N(x)$ will not get too large at the ends of the interval I .

THEOREM 3.1: Suppose that $I = [0,1]$ and that $0 < a < b < 1$. Then

$$P\{ \limsup_{N \rightarrow \infty} \max_{x \in I} [\mu_N(x) - \mu(x)] \leq 0 ;$$

$$\liminf_{N \rightarrow \infty} \max_{a \leq x \leq b} [\mu_N(x) - \mu(x)] \geq 0 \} = 1.$$

The proof of this theorem is given in Hanson & Pledger (1976).

Summarizing the estimation method we can conclude that the estimate of concave/convex regression is a piecewise linear function. However, the formulas for continuous piecewise regression function estimates are not directly applicable to this case. But the sum of the mean variance and the mean squared bias for piecewise linear regression is probably a rough estimate of the mean variance of a concave/convex estimate with the same mean interval length.

One common property for continuous piecewise linear regression is that the least squares regression function estimates are linear functions of the observations. This is proved in the following lemma.

LEMMA 3.1: If the bending points are fixed then the LS regression functions estimates are linear functions of the observations.

PROOF: For any of the x_i -intervals where the function is linear, the estimates are determined by using the estimation method simultaneously from the first x-interval to the i :th and from the last x-interval to the i :th. Then information from the observations in the first interval to the $(i-1)$:th is contained in a fiction point in the i :th bending point and information from the $(i+1)$:th to the last x-interval is contained in a fiction point in the $(i+1)$:th bending point. The regression function estimate in the i :th x-interval is the ordinary linear regression function estimate with the two fiction points included and for each observation in the interval the estimate is a linear function of all observations. But the fiction points are also a linear function of the observations in the other intervals respectively to the left of the $(i-1)$:th and to the right of the $(i+1)$:th intervals including other fiction points and so on. Therefore we can conclude that each regression function estimate is a linear function of the observations.

Q.E.D.

If we use the assumptions of equidistant x-observations and equally sized x-intervals it is easy to evaluate some theory about $w_0(i)$. A useful property describes the influence of the observations in the $(i-k)$:th or $(i+k)$:th x-interval on the weighted steepness in the i :th x-interval:

Suppose that (x_k, y_k) , $k=1, \dots, N$, are observations with the corresponding weights w_k , and that the values of x are equidistant. The estimate of the regression function is a continuous function consisting of piecewise straight lines. In order to get rough estimates of the influence of the bounds we make a calculation for equidistant bending points. This means that in each x -interval, corresponding to an estimated straight line, we have m ordinary observations x_j , $j=i, \dots, i+m-1$, with weights $w_j = 1$ together with an extra estimated point $(x_0^{(i)}, y_0^{(i)})$ with weight $w_0^{(i)}$. This estimated observation, with corresponding weight, contains the influence of the observations in the neighbouring x -interval, $i-1$. This means that when we use the weighted LSE of the regression line the influence of $(x_0^{(i)}, y_0^{(i)})$ on the slope is $w_0^{(i)}/(m+w_0^{(i)})$. It is easily shown that $w_0^{(i)}$ is limited by 0.1 and 0.3 and

$$\frac{w_0^{(i)}}{m+w_0^{(i)}} < \frac{0.3}{1+0.1} < 0.275.$$

Thus the influence of the observations in the $(i-k)$:th or $(i+k)$:th x -interval on the weighted steepness in the i :th x -interval is generally $(0.275)^k$. If $k > 4$ this influence will be very small.

Suppose we have a regression problem according to the described situation with observations (x_k, y_k) , $k=1, \dots, N$ with weights w_k . Also suppose that the x -observations are equidistant. Divide the observations into intervals, each containing m x -observations, $m \geq 2$. The first x -observation in each interval will be an ordinary observation with weight w_k and also coincide with the bending point with weight $w_0^{(i-1)}$

for the previous interval. The rest of the $(m-1)$ x -observations in each interval will be ordinary points. In this situation we can give a general expression for the weight $w_0^{(i)}$

$$w_0^{(i)} = \frac{\sum_{k=1}^m w_k \sum_{k=1}^m (x_k - \bar{x})^2 w_k + \left[\sum_{k=1}^m (x_k - \bar{x})^2 w_k + \frac{(\sum_{k=1}^m (x - x_0^{(i-1)}) w_k)^2}{\sum_{k=1}^m w_k} \right] w_0^{(i-1)}}{\sum_{k=1}^m (x_k - x_0^{(i)})^2 w_k + (x_0 - x_0^{(i)})^2 w_0^{(i-1)}}$$

It is obvious that if the series $w_0^{(i)}$, successively obtained, will converge to the limit value, w_0 , then this value is the solution of the general expression when $w_0^{(i)} = w_0^{(i-1)}$.

The general expression of the weight, $w_0^{(i)}$, can be written in a simplified way as

$$w_0^{(i)} = \frac{A + Bw_0^{(i-1)}}{C + Dw_0^{(i-1)}} = f(w_0^{(i-1)})$$

This equation behaves very nicely. It is obvious that the size of the derivative f' will influence the distance between $w_0^{(i)}$ and w_0 and also the convergence of $\lim_{i \rightarrow \infty} w_0^{(i)}$.

Generally, if $f' < 1$ then $\lim_{i \rightarrow \infty} w_0^{(i)}$ will converge to a limit value. If on the other hand

$f' > 1$ then this is not the case. We can also conclude that if the equation

$(C + Dw_0^{(i-1)})w_0^{(i)} = (A + Bw_0^{(i-1)})$ has complex roots then the series $w_0^{(i)}$ will not

converge. In other cases we can obtain two solutions.

In simple situations the limit value, w_0 , is very easy to calculate. Let Δ be the notation of the distance between two successive x-observations. In the special case when $\Delta=1$ and $w_k=1$, $k=1, \dots, N$, the solution of the general expression gives us the limit value

$$w_0 = -\frac{1}{2} + \sqrt{\frac{m^2 + 2}{12}}.$$

In this situation we can prove the convergence.

4. Example

Consider the concave regression example by Hildreth (1954), Wu (1982) and Fraser and Massam (1987):

Let y_i be the average corn yield, x_i be the amount of nitrogen fertilizer and m_i be the number of y-observations for each x_i . The data are

x_i	0	20	40	60	80	120	160	180
m_i	27	9	8	10	9	19	10	8
y_i	22.94	41.58	65.46	58.81	81.74	82.15	96.59	94.01

First we estimate an ordinary least squares regression line from the

observations $\hat{y}(x) = 34.43 + 0.3310x$. Now, for each x_i we calculate a ρ -value. If any ρ -value > 0 this indicates that we have not found the least squares estimate.

$$\rho(x) = (0.00, 59.69, 56.91, 45.96, 38.05, 19.87, 5.80, 0.00)$$

The largest ρ -value is $\rho(x_2) = 59.69$. This indicates that we should include $x_2 = 20$ as a bending point. Now we have to estimate two straight lines connected in x_2 .

From the linear regression through a predetermined point we obtain

$$\hat{y}(x) = 22.94 + 1.5490x \quad 0 \leq x \leq 20$$

$$\hat{y}(x) = 48.92 + 0.2505x \quad 20 \leq x \leq 180.$$

When we compare the regression coefficients we can see that the concavity restriction is fulfilled. Therefore we continue with the inclusion part to examine if any other point should be included in the regression procedure as a bending point. For each x_i , we calculate a ρ -value.

$$\rho(x) = (0.00, 0.00, 10.27, 10.04, 11.90, 7.11, 2.60, 0.00)$$

The largest ρ -value is $\rho(x_5) = 11.90$, ie $x_5 = 80$ is included as a bending point.

Now we have to estimate three straight lines connected in x_2 and in x_5 . From the linear regression procedure we obtain

$$\hat{y}(x) = 22.94 + 1.0810x \quad 0 \leq x \leq 20$$

$$\hat{y}(x) = 33.53 + 0.5513x \quad 20 \leq x \leq 80$$

$$\hat{y}(x) = 64.52 + 0.1638x \quad 80 \leq x \leq 180.$$

When we compare the regression coefficients we can see that the concavity restriction is fulfilled. Therefore we continue with the inclusion part to examine if a third point should be included in the regression procedure as a bending point.

For each x_i , we calculate a ρ -value.

$$\rho(x) = (0.00, 0.00, 2.47, -1.51, 0.00, -0.60, 0.58, 0.00)$$

The largest ρ -value is $\rho(x_3) = 2.47$, ie $x_3 = 40$ is the new bending point. From the linear regression procedure with bending points in x_2 , x_3 and x_5 we obtain

$$\hat{y}(x) = 22.94 + 0.9320x \quad 0 \leq x \leq 20$$

$$\hat{y}(x) = 23.46 + 0.9059x \quad 20 \leq x \leq 40$$

$$\hat{y}(x) = 43.03 + 0.4167x \quad 40 \leq x \leq 80$$

$$\hat{y}(x) = 62.25 + 0.1765x \quad 80 \leq x \leq 180.$$

From this result we can see that the regression coefficients fulfil the concavity restriction. We continue the inclusion part. For each x_i , we calculate a ρ -value.

$$\rho(x) = (0.00, 0.00, 0.00, -2.72, 0.00, -0.12, 0.72, 0.00)$$

The only ρ -value > 0 is $\rho(x_7) = 0.72$ and $x_7 = 160$ is included as a bending point.

From the linear regression procedure we obtain

$$\hat{y}(x) = 22.94 + 0.9320x \quad 0 \leq x \leq 20$$

$$\hat{y}(x) = 23.83 + 0.9274x \quad 20 \leq x \leq 40$$

$$\hat{y}(x) = 45.70 + 0.3606x \quad 40 \leq x \leq 80$$

$$\hat{y}(x) = 54.72 + 0.2479x \quad 80 \leq x \leq 160$$

$$\hat{y}(x) = 97.40 - 0.01882x \quad 160 \leq x \leq 180.$$

If we continue the inclusion part and calculate a p -value for each x_i then we can see that no obtained p -value > 0 and all b -values fulfil the concavity restriction. Thus we have found the least squares estimate which gives us the estimate of the regression function

$$\hat{y}(x) = (22.94, 41.58, 60.13, 67.13, 74.55, 84.47, 94.39, 94.01).$$

The least squares estimate gives us the sum of squares 1570.812.

Acknowledgements

I would like to thank professor Sture Holm for reading of this paper carefully and suggesting improvements in the text. I am also very grateful to professor Marianne Frisén for her good advice.

REFERENCES.

- BARLOW, R.E., BARTHOLOMEW, D.J., BREMNER, J.M. and BRUNK, H.D.; (1972): Statistical Inference Under Order Restrictions, Wiley, New York.
- DAHLBOM, U.; (1994): Estimation of regression functions with monotonicity and concavity/convexity restrictions. Almqvist & Wiksell International, Stockholm.
- DENT, W.; (1973): A Note on Least Squares Fitting of Functions Constrained to be Either Nonnegative, Nondecreasing or Convex, Management Science 20, pg 130-132.
- FRASER, D. A. S. and MASSAM, H.; (1987): An Algorithm for Concave Regression, Foundations of Statistical Inference, D. Reidel Company.
- HANSON, D.L. and PLEDGER, G.; (1976): Consistency in Regression, Annals of Statistics 4, pg 1038-1050.
- HILDRETH, C.; (1954): Point Estimates of Ordinates of Concave Functions, Journal of American Statistical Association 49, pg 598-619.
- HOLLOWAY, C.A.; (1979): On the Estimation of Convex Functions, Operations Research 27, pg 401-407.
- HOLM, S. and FRISÉN, M.; (1985): Nonparametric Regression with simple curve characteristics, Research report 1985:4, University of Göteborg, Dept. of Statistics.
- ROBERTSON, T., WRIGHT, F.T. and DYKSTRA, R.L.; (1989): Order Restricted Statistical Inference, Wiley, New York.
- WU, C-F.; (1982): Some Algorithms for Concave and Isotonic Regression, Studies in the Management Sciences 19: pg 105-116.

Research Report

- | | | |
|--------|-----------------------------|---|
| 1996:1 | Ekman, A | Sequential analysis of simple hypotheses when using play-the-winner allocation. |
| 1996:2 | Wessman, P | Some principles for surveillance adopted for multivariate processes with a common change point. |
| 1996:3 | Frisén, M. & Wessman, P | Evaluations of likelihood ratio methods for surveillance. |
| 1996:4 | Wessman, P. | Evaluation of univariate surveillance procedures for some multivariate problems. |
| 1996:5 | Särkkä, A. | Outlying observations and their influence on maximum pseudo-likelihood estimates of Gibbs point processes. |
| 1997:1 | Ekman, A. | Sequential probability ratio tests when using randomized play-the-winner allocation. |
| 1997:2 | Pettersson, M. | Monitoring a Freshwater Fishpopulation: Statistical Surveillance of Biodiversity. |
| 1997:3 | Jonsson, R. | Screening-related prevalence and incidence for non-recurrent diseases. |
| 1997:4 | Johnsson, T. | A stepwise regression procedure applied to an international production study - a multiple inference solution. |
| 1998:1 | Särkkä, A. & Högmänder, H.: | Multiple spatial point patterns with hierarchical interactions. |
| 1998:2 | Frisén, M. & Wessman, P.: | Evaluations of likelihood ratio methods for surveillance.
Differences and robustness. |
| 1998:3 | Järpe, E.: | Surveillance of spatial patterns.
Change of the interaction in the Ising model. |
| 1998:4 | Pettersson, M.: | Evaluation of some methods for statistical surveillance of an autoregressive process. |
| 1998:5 | Pettersson, M.: | On monitoring of environmental and other autoregressive processes. |