UNIVERSITY OF GÖTEBORG

Department of Statistics

NONPARAMETRIC REGRESSION WITH SIMPLE

CURVE CHARACTERISTICS

by

Sture Holm and Marianne Frisén

Statistiska institutionen

Göteborgs Universitet

Viktoriagatan 13

S 411 25 Göteborg

Sweden

NONPARAMETRIC REGRESSION WITH SIMPLE CURVE CHARACTERISTICS

S. Holm

M. Frisén

1. Introduction

The character of nonparametric statistical methods
is that they are constructed for very general situations,
without the specific narrow assumptions, which appear in the
common parametric methods. Isotonic regression is a non-
parametric regression method, which has paid a well deserved
attention for some decades. In this case the only assump-
tion about the regression function is that it is non-decreas-
ing (or nonincreasing). The basic theory of isotonic
regression is contained in the book by Barlow, Bartholomew,
Brenner and Brunk (1972).

In many applications it is motivated to consider regression
functions which are not only monotonic but also have certain
convexity or concavity characteristics. For instance in
quantal response assays in biological applications, sigmoid
curves are used. These are increasing functions which are
first convex up to some point and then concave. There are
suggested a number of parametric sigmoid curves for ana-
lysis of such applications. See e.g. Finney (1978) section
17.

In economical applications involving demand, supply and
price, functions with prescribed monotonity and convexity or
concavity are common. For instance in Lipsey & Steiner
(1972) chapter 5 are found a number of convex decreasing
demand curves and convex increasing supply curves (in both
cases price as a function of quality). There is also given
an example of the quality demanded as a function of house-
hold income which might change character from increasing
concave to decreasing concave.

In all these applications the regression functions can be
assumed to satisfy some simple   nonparametric) curve
characteristics expressed in terms of increase or decrease
and convexity or concavity. Either the function is of one
type all the way, or it shifts character in a certain order
at some (unknown) points. The aim of the present paper
is to discuss the statistical problem of estimating such
nonparametric regression functions.

The papers by Dent (1973) and Holloway (1979) consider the problem of estimating convex (or concave) regression functions. In both papers the least squares estimates are obtained by linear programming methods. The case of unimodal regression is treated in Frisén (1985).

The regression functions with a single characteristic are of four types, increasing convex, increasing concave, decreasing convex and decreasing concave. It will be seen in the next section that the corresponding estimation problems are analogous, and we will give a procedure for determinating least square estimates.

For a regression function, shifting curve characteristic in a given point, we can obtain the least square estimate by a slight modification of the method for regression functions with a single curve characteristic. In the case when the regression function shifts curve characteristic in an unknown point we can find the least square estimate by calculating the sum of squares for the solutions for all possible shifting points. The general solution is then the one obtained for the shifting point giving the least sum of squares.

The estimation procedure will be illustrated by two simple examples in section 3. In section 4 we will give the utmost simplest consistency result for the estimates. Further statistical properties and further details on the estimation procedure will be given in forthcoming papers.

2. Estimation procedure

We will first discuss the estimation procedure for fitting a nondecreasing and convex function to a set of data by the least squares criterion. This means that to the observations (x_1, y_1), $(x_2, y_2), \ldots,$ (x_n, y_n) with $x_1 < x_2 < \ldots < x_n$ we will find a nondecreasing and convex function $f(x)$ such that

$$\sum_{k=1}^{n} (y_k - f(x_k))^2$$

is minimal. As mentioned in the introduction, the problems to fit a function with some other nonchang curve characteristic will be very similar to this one.

There is a reformulation of the problem which is good for theoretical as well as practical purposes. Let

$$f_0(x) = 1$$
$$f_1(x) = x - x_1$$

and

$$f_k(x) = \begin{cases} 0 & \text{for } x \geq x_k \\ x - x_k & \text{for } x > x_k \end{cases} \quad \text{for } k = 2, 3, \ldots, n-1$$

Then any function $f(x)$ which is nondecreasing and convex on the set $\{x_k : k = 1, 2, \ldots, n\}$ can be written:

$$f(x) = \sum_{k=0}^{n-1} \alpha_k f_k(x)$$

where

$$\alpha_k \geq 0 \quad \text{for } k \geq 1$$

There are no restrictions on the constant α_0.

Thus our problem can be formulated as the problem of finding

$$\alpha_k \geq 0 \quad k=1,2,\ldots,n-1$$

and α_0, which minimize

$$\sum_{l=1}^{n} \left(y_l - \sum_{k-0}^{n-1} \alpha_k f_k(x_l) \right)$$

For existance and uniqueness of solutions to the minimization problem, the following lemma is essential.

Lemma 1: The set of nondecreasing convex functions on the set

$$\left\{ x_k : k=1,2,\ldots,n \right\} \quad \text{is a closed convex cone.}$$

Proof: We can write the functions in the form

$$f(x) = \sum_{k=0}^{n-1} \alpha_k f_k(x)$$

with $\alpha_k \geq 0$ for $k=1,2,\ldots,n-1$.

If

$$g(x) = \sum_{k=0}^{n-1} \beta_k f_k(x)$$

is another function of this type, then so is

$$\lambda f(x) + (1-\lambda) g(x) = \sum_{k=0}^{n-1} (\lambda \alpha_k + (1-\lambda) \beta_k) f_k(x)$$

for all $\lambda, 0 \leq \lambda \leq 1$. Thus the set is convex. Further obviously $\gamma f(x)$ belongs to the set for all $\gamma > 0$ if $f(x)$ belongs to the set. Finally it is easily seen that if $f(x)$ is a limit of functions in this set, the function $f(x)$ also belongs to the set, since limits of sequences of nonnegative numbers are non-negative.

Q.E.D.

In order to be able to write things shorter we introduce the scalar product notation

$$(f,g) = \sum_{k=1}^{n} f(x_k)g(x_k)$$

and the norm notation

$$\|f\| = (f,f)^{-\frac{1}{2}}$$

for functions defined on

$$\left\{ x_k : \ k=1,2,\ldots,n \right\}.$$

In the minimization problem, where we should find $\alpha_k \geq 0$ $k=1,\ldots,n-1$ and α_0 to minimize

$$\left\| f - \sum_{k=0}^{n-1} \alpha_k f_k \right\|$$

we typically have some positive $\alpha_k = s$, while the others are equal to 0.

Denoting

$$I = \left\{ k \geq 1 : \alpha_k > 0 \right\} \cup \left\{ 0 \right\}$$

we can write the approximating function

$$\sum_{k \in I} \alpha_k f_k(x).$$

Then we can formulate the following lemma on a characterisation of the solution to the minimization problem.

Lemma 2. The function

$$\sum_{k \in I} \alpha_k f_k(x)$$

is the solution to the minimization problem if and only if

(i) $(f - \sum_{k \in I} \alpha_k f_k, f_j) = 0$ $\forall j \in I$

and

(ii) $(f - \sum_{k \in I} \alpha_k f_k, f_j) \leq 0$ $\forall j \notin I$

Proof   Denote

$$M = \| f - \sum_{k \in I} \alpha_k f_k \|$$

and suppose first that

$$\sum_{k \in I} \alpha_k f_k$$

is the solution. Then we can not have

$$(f - \sum_{k \in I} \alpha_k f_k, f_j) = k_j \neq 0 \text{ for } j \in I$$

since then

$$\| f - \sum_{k \in I} \alpha_k f_k - \varepsilon f_j \|^2 = M^2 - 2\varepsilon k_j + \varepsilon^2 \| f_j \|^2$$

is smaller than M^2 for some small positive or negative

ε. Further we can not have

$$(f - \sum_{k \in I} \alpha_k f_k - f_j) = k_j > 0 \text{ for } j \notin I$$

because then

$$\| f - \sum_{k \in I} \alpha_k f_k - \varepsilon f_j \|^2 = M^2 - 2\varepsilon K_j + \varepsilon^2 \| f_j \|^2$$

is smaller than M^2 for some small positive ε.

Thus the solution must satisfy (i) and (ii). On the

other hand if (i) and (ii) are satisfied, we have a local

minimum, and there is only one local minimum, the

global one.

                                        Q.E.D.

This characterisation lemma is closely related to the stepwise method to find the solution. The method consists of two parts, the exclusion part and the inclusion and substitution part.

The aim of the exclusion part is to find an index set I_0 such that

$$(f - \sum_{k \in I_0} \alpha_k f_k, f_j) = 0 \qquad \forall j \in I_0$$

and $\alpha_k \geq 0 \quad \forall j \in I_0 \setminus \{0\}$.

This is obtained by successive elimination.

Exclusion part of estimation procedure:

First write

$$f = \sum_{k \in \{0,\ldots,n-1\}} \alpha_k f_k$$

Let $I_0^{(1)}$ be the index set consisting of 0 and all $k \geq 1$ such that $\alpha_k \geq 0$. Next make a least square approximation of f by the sum $\sum_{k \in I_0^{(1)}} \alpha_k^{(1)} f_k$

Again exclude indices corresponding to $\alpha_k^{(1)} < 0$, let $I_0^{(2)}$ be the index set consisting of 0 and all $k \geq 1$ such that $\alpha_k^{(1)} \geq 0$, and make a least square approximation of f by the sum

$$\sum_{k \in I_0^{(2)}} \alpha_k^{(2)} f_k.$$

This is continued until we find an index set I_0 such that the least squares approximation

$$\sum_{k \in I_0} \alpha_k f_k$$

of f has koefficients $\alpha_k \geq 0$ for all $k \geq 1$ in I_0.

Note. It might happen that we end up with $I_0 = \{0\}$.
The exclusion part of the procedure is not necessary,
we could always skip it and start the inclusion and
substitution step with $I_0 = \{0\}$. But generelly the
exclusion part would give us a rough estimate, which is
a good starting point in the inclusion and substitution
part. In very simple cases it might also hit the solution
directly. For instance if the function f itself is
nondecreasing and convex, we would get $I_0 = \{0, 1, \ldots,$
$\ldots, n-1\}$.

Inclusion and substitution part of the estimation procedure.

A. The index set I_0 is such that

$$(f - \sum_{k \in I_0} \alpha_k f_k, f_j) = 0 \qquad \forall j \in I_0$$

and

$$\alpha_k \geq 0 \qquad \forall j \in I_0 \setminus \{0\}.$$ Calculate for each $j \notin I_0$ the
"projection"

$$\rho_j = (f - \sum_{k \in I_0} \alpha_k f_k, f_j) / \| f_j \|$$

Aa. If $\rho_j \leq 0 \quad \forall j \notin I_0$, the sum $\sum_{k \in I_0} \alpha_k f_k$
is the solution to the minimization problem.

Ab. If $\rho_j > 0$ for some $j \in I_0$ continue to B.

B. Let m be the index corresponding to the maximal

ρ_j (or one of the maximal ρ_j:s if there are several).

Let β_k for $k \in I_0 \cup \{m\}$ be the constants minimizing

$\| f - \sum_{k \in I_0 \cup \{m\}} \beta_k f_k \|$.

Ba. If $\beta_k \geq 0$ $\forall k \in I_0 \cup \{m\}$ start again from A with I_0

substituted by $I_0 \cup \{m\}$ and constants $\beta_k, k \in I_0 \cup \{m\}$.

Bb. If $\beta_k < 0$ for some $k \in I_0 \cup \{m\}$, calculate $\varepsilon_k = \alpha_k / (\alpha_k - \beta_k)$

and $\varepsilon^* = \min_{k \in I_0 \cup \{m\}} \varepsilon_k < 1$. Then in the sum

$\sum_{k \in I_0 \cup \{m\}} ((1-\varepsilon^*) \alpha_k + \varepsilon^* \beta_k) f_k$

at least one coefficient equals 0. Let I_1 be the index
set of the non-zero coefficients, and let

$\gamma_k = (1-\varepsilon^*) \alpha_k + \varepsilon^* \beta_k$, $k \in I_1$.

Further let ρ_k, $k \in I_1$ be the constants minimizing

$\| f - \sum_{k \in I_1} \rho_k f_k \|$

If $\rho_k \geq 0$ $\forall k \in I$ start again from A with I_0 substituted
by I_1.

If $\rho_k < 0$ for some $k \in I_1$ calculate new $\varepsilon_k = \gamma_k / (\gamma_k - \rho_k)$
$k \in I_1$ and $\varepsilon^* = \min_{k \in I_1} \varepsilon_k < 1$ and repeat Bb until only positive
coefficients are obtained. Then start from A again.

Note. In each "cycle" of the Bb part of the procedure,
the sum of squares of errors will strictly decrease.
In each "complete cycle" including A we will start anew
with a "presolution", with a smaller sum of squares of
errors than in the previous case, with positive co-
efficients in a least squares solution for a new set of
indeces. Since there is a finite number of possible
choices of indeces, the procedure will converge to the
solution in a finite number of steps. In A is contained

a check step where we stop when we have found the
solution as characterized in lemma 2. The inclusion
of the index m corresponding to the function with the
greatest "positive correlation" with the error in A
made for intuitive reasons. It ought to be good for
improving the solution as much as possible, and thus
ought to give fast convergence.

The procedure we have given here is easily modified
for other similar problems. For instance, if we want
to fit a nonincreasing convex function to data (x_1,y_1),
$(x_2,y_2),\ldots,$ (x_n,y_n) with $x_1<x_2<,\ldots,<x_n$ we use instead
the function system

$$f_0(x)=1$$

$$f_k(x)=\begin{cases} x_k-x & \text{for } x \leqslant x_k \\ 0 & \text{for } x>x_k \end{cases}$$

for $k=2,3,\ldots,n-1$

and

$$f_n(x)=x_n-x.$$

If there are several observations for some x_k we
just use a function system with functions corresponding
to all different values x_k. The observation in a point
is the mean of all y_k with the same x_k and in the scalar
product we use weights equal to the number of observations
in the different points.

For problems with shifting curve characteristics, there
are two cases, which are especially simple. If we want
to fit to data $(x_1,y_1),\ldots,(x_n,y_n)$ with $x_1<x_2<\ldots<x_n$
a function which is first nonincreasing convex and then
nondecreasing convex, we can use the function system of
the first problem in this section. But now there are
no sign restrictions of the coefficients of neither
$f_0(x)$ nor $f_1(x)$. The modification of the procedure for
this case is trivial.

A similar solution is obtained for the problem of
fitting a function which is first nondecreasing concave
and then nonincreasing concave.

The problem of fitting a sigmoid curve, which is first nondecreasing convex and then nondecreasing concave is not so simple.

3. Two simple examples

In order to illustrate how the estimation procedure works, we will show the steps in detail for two simple examples. Our first example is a very simple one, used by Holloway (1979).

Example 1. Fit a convex function (by least squares) to the data

x_k	2	4	6	9	10
y_k	10	2	6	4	8

When we make an approximation in form of a linear combination of f_0, f_1, f_2, f_3, f_4 there are no restrictions on the coefficients of f_0 and f_1 in this case.
After writing

$$f = 10 - 4f_1 + 6f_2 - \frac{8}{3} f_3 + \frac{14}{3} f_4$$

we exclude f_3, which has negative coefficient.

Fitting a linear combination of f_0, f_1, f_2, f_4 by least squares we get

$$f \sim 10 - 3,37f_1 + 3,69f_2 + 2,84f_4$$

The coefficients of f_2 and f_4 are positive, and the scalar product of f_3 and the error is negative. Thus we have the solution already at the end of the elimination part of the procedure.

This example was almost too simple. Also the next one is simple, but it is complicated enough to get also inclusion steps.

Example 2. Fit a nondecreasing convex function by least squares method to the data

x_k	1	3	5	9	10	11	14	15
y_k	3	5	4	5	7	10	11	14

After writing

$$f=3+f_1-1,5f_2+0,7f_3+1,75f_4+f_5-\frac{8}{3}f_6+\frac{8}{3}f_7$$

we exclude f_2 and f_6, which have negative coefficients. The least squares fit with a linear combination of f_0, f_1, f_3, f_4, f_5 and f_7 becomes

$$f\approx3,5+0,25f_1-0,125f_3+2,7981f_4-2,0769f_5+1,8462f_7.$$

Thus we next exclude f_3 and f_5. The least squares fit with a linear combination of f_0, f_1, f_4 and f_7,

$$f\approx3,3991+0,3149f_1+0,8404f_4+1,1508f_7$$

has positive coefficients for f_1, f_4 and f_7. This terminates the elimination part of the procedure. The error turns out to be negatively correlated with f_2, f_5 and f_6 but positively correlated with f_3. The least squares fit with f_3 included becomes

$$f\approx3,50+0,25f_1+0,1161f_3+0,7768f_4+1,1786f_7$$

which has positive coefficients for f_1, f_3, f_4 and f_7. It is not necessary to eliminate some other functions when f_3 is included. The error appears to have negative correlations with f_2, f_5 and f_6, which terminates the whole procedure. After a calculation including 3 least squares approximations we got the solution in the following table

x_k	1	3	5	9	10	11	14	15
$\hat{f}$	3,5	4,0	4,5	5,9643	7,1071	8,25	11,6786	14,0

In a procedure involving calculation for all possible subsets of variables f_1, f_2, f_3, f_4, f_5, f_6, f_7 would need 64 calculations of least squares estimates.

4. Consistency

In this paper we have no intention to treat the more intricate statistical properties of the estimates. We will only give a simple consistency property.

<u>Theorem.</u> Suppose that the mean of a random variable, Y is a strictly increasing and convex function $\mu(x)$ of x on the set $\left\{x_k; k=1,2,\ldots,n\right\}$, and that Y has a variance for all x_k, k=1,2,...,n. Suppose further that we make N_k observations of Y at x_k and that all

$$N = \sum_{k=1}^{n} N_k$$

Y:s are independent. Then the proposed estimator is uniformly consistant for estimating $\mu(x)$ for $x \in \left\{x_k; k=1,2,\ldots,n\right\}$ when

$$\min_{1 \leq k \leq n} N_k \to \infty$$

<u>Proof.</u> Because $\mu(x)$ is strictly increasing and convex there exists δ_o such that all functions $\mu^*(x)$ satisfying $|\mu^*(x)-\mu(x)| < \delta_o \forall x \in \left\{x_k : k=1,2,\ldots,n\right\}$ are also strictly increasing and convex. But by the Chebychev inequality and the Boole inequality there exists for each $\epsilon > 0$ and $\delta > 0$ a number $N(\epsilon,\delta)$ such that

$$P(|\overline{Y}_k - \mu(x_k)| < \delta \forall k=1,2,\ldots,n) \geq 1-\epsilon$$

$$\geq 1-\epsilon$$

when $\overline{Y}$ is the mean of at least $N(\epsilon,\delta)$ observations at x_k. If the mean function (taking value $\overline{Y}_k$ in x_k) is itself strictly increasing and convex the procedure will estimate $\mu(x_k)$ by $\overline{Y}_k$. Thus if

$$\min_{1 \leq k \leq n} N_k \geq N(\epsilon,\delta)$$

and $\delta \leq \delta_o$ the estimate $\hat{\mu}(x)$ will satisfy

$$P(|\hat{\mu}(x_k) - \mu(x_k)| < \delta \forall k=1,2,\ldots,n) \geq 1-\epsilon$$

Q.E.D.

References

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

Dent W. (1973). A Note on Least Squares Fitting of
    Functions Constrained to be Either Nonnegative,
    Nondecreasing or Convex. Management Science, 20,
    p 130-132.

Frisén M. (1985). Unimodal Regression. Research
    Report 1985:3 from Department of Statistics,
    University of Gothenburg.

Finney D.J. (1978). Statistical Methods in Biological
    Assay. Charles Griffin & Co, London.

Holloway C.A. (1979). On the Estimation of Convex
    Functions. Operations Research, 27, p 401-407.
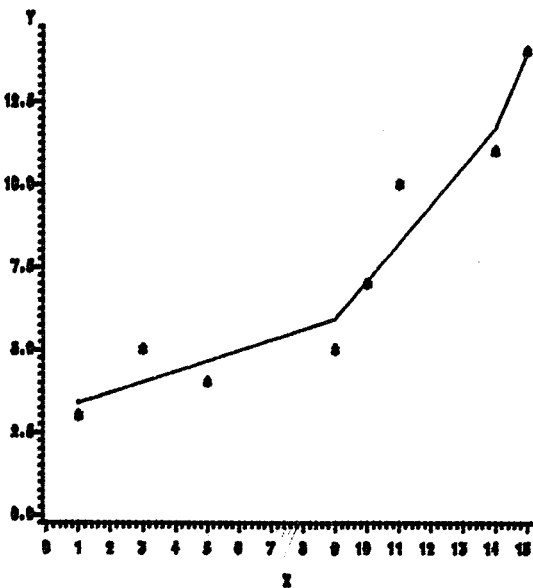
Illustration of example 2
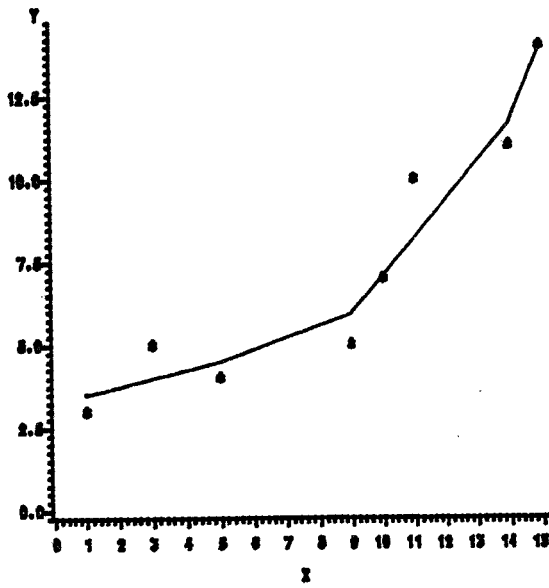The following figures show the succesive steps in the
estimation procedure
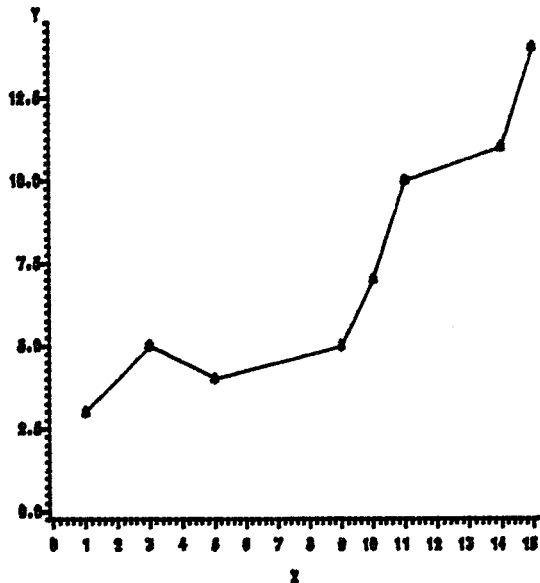
Starting approximation
All functions used
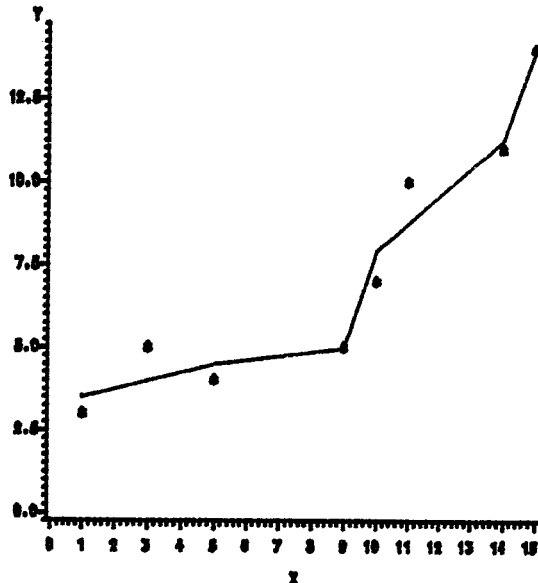
Second approximation
Functions f_2 and f_6
excluded

Third approximation
Also functions f_3 and f_5
excluded

Fourth and final approxi-
mation.
Function f_3 included

RESEARCH REPORT

1985:1   Frisén, M: The forgotten assumption in the
                          superpopulation model.

1985:2   Guilbaud, O: Stochastic order relations for one-sample
                          statistics of the Kolmogorov-Smirnov type.

1985:3   Frisén, M: Unimodal Regression.