Smooth Regression Analysis

Author(s): Geoffrey S. Watson

Source: *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, Dec., 1964, Vol. 26, No. 4 (Dec., 1964), pp. 359-372

Published by: Indian Statistical Institute

Stable URL: https://www.jstor.org/stable/25049340

SMOOTH REGRESSION ANALYSIS

By GEOFFREY S. WATSON*

The Johns Hopkins University

SUMMARY. Few would deny that the most powerful statistical tool is graph paper. When however there are many observations (and/or many variables) graphical procedures become tedious. It seems to the author that the most characteristic problem for statisticians at the moment is the development of methods for analyzing the data poured out by electronic observing systems. The present paper gives a simple computer method for obtaining a "graph" from a large number of observations.

1. Introduction

Large sample methods have been devised recently for handling time series on the assumption that they are stationary and have respectable spectral distributions. The methods involve rather arbitrary numerical smoothing and permit the use of judgement. The practice of following rather castiron inferential rules seems to be breaking down. Smoothing methods have been suggested, in the same spirit by the author and others for the estimation of probability densities and conditional failure rates and the next task in the programme is naturally the estimation of regression functions. (See Rosenblatt (1956), Whittle (1957, 1958), Tukey (1960), Watson and Leadbetter (1963, 1964, 1964a), Parzen (1962b), Bartlett (1964)).

Mahalanobis (1961) has suggested a "distribution-free" regression analysis and christened it "Fractile Graphical Analysis." Though it has properties of consistency, similar to the method to be proposed here, the aim of the method is rather to produce distribution-free tests than to track arbitrary curves with large bodies of data. Because it uses order statistics it is not in fact well adapted for this purpose. The limiting theory is given by Parthasarathy and Bhattacharya (1961) and other results occur in the same issue of *Sankhyā.*

The plan of this paper is as follows : Section 2 gives a review of previous work on the estimation of probability densities and then extends the methods for regression functions. The treatment is heuristic—a more rigorous treatment will be given elsewhere by Leadbetter (1964). Section 3 contains several artificial numerical examples and Section 4 shows a biological application. Some generalizations to regression functions of several variables and to time series are discussed in Section 5.

2. Theory

Let $(X_1, Y_1) \dots (X_n, Y_n)$ be n pairs of independent random variables with joint density function $f(x, y)$. Let $f_1(x)$ be the marginal density of X. Then the regression function of Y on X is

$$m(x) = E(Y \mid X = x)$$

i.e.,

$$m(x) = \frac{\int y f(x, y) dy}{\int f(x, y) dy} \qquad \dots \quad (2.1)$$

*This research was supported in part by the Office of Naval Research, Contract Nonr 4010(09).

The problem is to obtain an estimate of $\hat{m}(x)$, using the sample of n, that tends to $m(x)$ as $n \to \infty$ regardless (almost) of the nature of $f(x, y)$ and in particular of $m(x)$. It is necessary first to sketch, non-rigorously, earlier work on the estimation of the probability density.

Let $\{\delta_n(z)\}$ be a sequence of non-negative functions of z, of total area unity. As $n \to \infty$, $\delta_n(z)$ is to "tend" to the Dirac delta function. (There is no mathematical reason why the functions should be non-negative.) Then

$$\hat{f}_1(x) = \frac{1}{n} \sum_{i=1}^{n} \delta_n(x - X_i) \qquad \ldots \quad (2.2)$$

is a smoothed estimate of the probability density $f_1(x)$. For

$$E(\hat{f}_1(x)) = \int_{-\infty}^{\infty} \delta_n(x-z) f_1(z) dz, \qquad \ldots \quad (2.3)$$

$$\text{var}\,(\hat{f}_1(x)) = \frac{1}{n}\left\{ \int_{-\infty}^{\infty} \delta_n^2(x-z) f_1(z) dz - \left[\int_{-\infty}^{\infty} \delta_n(x-z) f_1(z) dz \right]^2 \right\}. \qquad \ldots \quad (2.4)$$

Since $\delta_n(z) \to \delta(z)$, $E(\hat{f}_1(x)) \to f_1(x)$ provided $f_1(x)$ is continuous at x. To see that also var $(\hat{f}_1(x))$ may tend to zero, define $\alpha_n = \int_{-\infty}^{\infty} \delta_n^2(z) dz$. Then $\alpha_n \to \infty$ and $\{\delta_n^2(z)/\alpha_n\}$ is another sequence of functions tending to $\delta(z)$.

Now (2.4) may be rewritten as

$$\frac{n}{\alpha_n} \text{var}\,(\hat{f}_1(x)) = \int_{-\infty}^{\infty} \frac{\delta_n^2(x-z)}{\alpha_n} f_1(z) dz - \frac{1}{\alpha_n}\left[\int_{-\infty}^{\infty} \delta_n(x-z) f_1(z) dz \right]^2 \qquad \ldots \quad (2.5)$$

so that we have, as $n \to \infty$,

$$\text{var}\ (\hat{f}_1(x)) \sim \frac{\alpha_n}{n} f_1(x) \qquad \ldots \quad (2.6)$$

which tends to zero if α_n/n tends to zero. Thus estimate (2.2) converges, at continuity points, to $f_1(x)$ as $n \to \infty$ provided $\delta_n(z) \to \delta(z)$ at the right speed. The reader will see that if $\delta_n(z)$ is the "square wave" function ($\delta_n(z) = \delta$ for $|z| \leqslant \frac{1}{2\delta}$ and zero elsewhere) the estimator (2.2) is almost the histogram where one knows that too fine a subdivision gives little bias but much variance, while too coarse a subdivision gives the reverse. For spectral densities (Parzen, 1958) and probability densities (Watson and Leadbetter, 1963), a formal way out of this dilemma has been suggested: choose $\delta_n(z)$ so that

$$E \int (\hat{f}_1(z) - f_1(z))^2\, dz \quad \text{minimum}. \qquad \ldots \quad (2.7)$$

For probability densities, this means that the fourier transform of $\delta_n(z)$, $\varphi_{\delta_n}(t)$, should be related to the fourier transform of f_1, $\varphi_{f_1}(t)$, by

$$\varphi_{\delta_n}(t) = \frac{|\varphi_{f_1}|^2}{\frac{1}{n} + \frac{n-1}{n} |\varphi_{f_1}|^2} \qquad \ldots \quad (2.8)$$

Thus the "best" $\delta_n(z)$ depends on the function $f_1(z)$ being estimated. While it may be shown that the minimum is "flat", this approach is not very useful in practice. Thus it seems best to use a family of smoothing functions e.g. a fixed function $k(z) > 0$, $\int k(z)dz = 1$ and a δ_n so that the family is $\left\{\frac{1}{\delta_n} k\left(\frac{z}{\delta_n}\right)\right\}$ with $\delta_n \to 0$ as $n \to \infty$. By experimenting numerically one chooses what appears to be the "best" member of the family.

To apply these ideas to the regression problem suppose $\delta_n(x, y)$ is a two dimensional smoothing function or "window." Then it will be required that

$$\delta_n(x, y) \geqslant 0, \ \int\int \delta_n(x, y)dx\, dy = 1. \qquad \dots \quad (2.9)$$

To estimate $m(x)$ in (2.1), it is natural first to estimate $f(x, y)$ by the analogue of (2.2), namely,

$$\hat{f}_n(x, y) = \frac{1}{n} \sum_{i=1}^{n} \delta_n(x - X_i, y - Y_i). \qquad \dots \quad (2.10)$$

With this estimate, one can form

$$\hat{m}_n(x) = \int_{-\infty}^{\infty} y\hat{f}_n(x, y)dy / \int_{-\infty}^{\infty} \hat{f}_n(x, y)dy. \qquad \dots \quad (2.11)$$

If we further require that

$$\int_{-\infty}^{\infty} y\delta_n(x, y)dy = 0, \qquad \dots \quad (2.12)$$

and define
$$\delta_n(x) = \int_{-\infty}^{\infty} \delta_n(x, y)dy, \qquad \dots \quad (2.13)$$

then (2.11) becomes
$$\hat{m}_n(x) = \frac{\sum_{i=1}^{n} Y_i\, \delta_n(x - X_i)}{\sum_{i=1}^{n} \delta_n(x - X_i)}. \qquad \dots \quad (2.14)$$

This estimator is seen to be the ratio of the estimator

$$\frac{1}{n} \sum_{1}^{n} Y_i\, \delta_n(x - X_i) \quad \text{of} \quad \int y f(x, y)dy$$

and
$$\frac{1}{n} \Sigma\, \delta_n(x - X_i) \quad \text{of} \quad \int f(x, y)dy = f_1(x).$$

When (2.14) is written as

$$\hat{m}_n(x) = \sum_{i=1}^{n} Y_i \frac{\delta_n(x - X_i)}{\sum_{j=1}^{n} \delta_n(x - X_j)} \qquad \dots \quad (2.15)$$

it is clearly a weighted average of the Y_i where the weights fall off as one goes away from x and vary with the sample. The use of fixed weights when Y is observed for equally spaced and known x is well known—(2.15) constitutes a generalization.

The mathematical study of (2.14) is awkward because it is a ratio. If $\delta_n(z)$ is taken, as in the numerical examples below, as the triangular window ($\delta_n(z) = \delta(1-\delta|z|)$ for $|z| < \delta$, zero elsewhere), the denominator may be zero with non-zero probability so that the estimator may have no moments. To get a criterion for choosing $\delta_n(z)$ we could either demand that it minimizes

$$J = E \int \left(\sum_{i=1}^{n} \delta_n(x-X_i)m(x) - \sum_{i=1}^{n} Y_i \, \delta_n(x-X_i) \right)^2 dx \qquad \dots \quad (2.16)$$

or something similar to (2.7) when the denominator of (2.14) is replaced by $n f_1(x)$. In both cases, the integrals will not in general converge because of the arbitrary nature of $m(x)$. This reflects the difficulty inherent in the problem, that while in estimating $f_1(x)$ we are dealing, roughly, with a function that must be essentially zero except for a finite part of the x-axis, the estimation of $m(x)$ deals with functions that are non-zero and with values of interest everywhere. Assuming away this convergence difficulty, the minimization of (2.16) requires a $\delta_n(z)$ with a fourier transform given by

$$\varphi_{\delta_n}(t) = \frac{|\varphi_{f_1 m}|^2}{\frac{E(Y^2)}{n} + \frac{n-1}{n} |\varphi_{f_1 m}|^2} \qquad \dots \quad (2.17)$$

where

$$\varphi_{f_1 m}(t) = \int e^{ixt} f_1(x) m(x) dx. \qquad \dots \quad (2.18)$$

While a study could be made of this optimum choice, as was done before, experimental studies seem more profitable.

To show directly that, for suitable $\{\delta_n(z)\}$, $\hat{m}_n(x) \to m(x)$ it will be sufficient for practical applications to observe that

$$\hat{m}_n(x) = \frac{\Sigma \; Y_i \delta_n(x-X_i)}{n \, f_1(x)} \; \frac{n f_1(x)}{\Sigma \; \delta_n(x-X_i)}. \qquad \dots \quad (2.19)$$

From earlier discussions it follows that the last factor tends, in probability, to unity provided $\alpha_n \to \infty$, $\frac{\alpha_n}{n} \to 0$. Denoting the first term by $m_n^*(x)$ we have

$$E(m_n^*(x)) = \frac{1}{f_1(x)} \int_{-\infty}^{\infty} m(z) \delta_n(x-z) f_1(z) dz$$

$$\to m(x) \quad \text{as } n \to \infty$$

provided $\delta_n(z) \to \delta(z)$, and $m(z) \, f_1(z)$ continuous at x.

Now

$$\operatorname{var}(m_n^*(x)) = \frac{1}{n f_1^2(x)} \operatorname{var}(Y_i \delta_n(x-X_i)),$$

$$= \frac{1}{n f_1^2(x)} \left\{ \int E(Y^2 | X = z) \delta_n^2(x-z) f_1(z) dz - f_1^2(x) E^2(m_n^*) \right\}.$$

But

$$E(Y^2 \mid X = z) = \sigma^2(z) + m^2(z)$$

so that

$$\operatorname{var}(m_n^*(x)) = \frac{\alpha_n}{n f_1^2(x)} \left\{ \int \frac{\delta_n^2(x-z)}{\alpha_n} (\sigma^2(z) + m^2(z)) f_1(z) dz - \frac{f_1^2(x) E^2(m_n^*(x))}{\alpha_n} \right\}.$$

But if $\alpha_n \to \infty$, $\delta_n^2(z)/\alpha_n \to \delta(z)$ and if $(\sigma^2(z) + m^2(z))\ f_1(z)$ is continuous at x, we have

$$\operatorname{var}(m_n^*(x)) \sim \frac{\alpha_n}{n f_1(x)} (\sigma^2(x) + m^2(x)) \to 0, \text{ as } n \to \infty. \qquad \ldots \quad (2.20)$$

Our heuristic result is that the $\hat{m}_n(x)$, given by (2.14) is a consistent estimator of $m(x)$ and, in large samples, its variance is given by (2.20).

Quite different types of estimators may be suggested, starting out with the ideas used by Hodges and Fix (1952) on non-parametric discrimination. For example, one might consider

$$\tilde{m}(x) = \frac{1}{k(n)} \sum_{S(x)} Y_i \qquad \ldots \quad (2.21)$$

where $\qquad S(x) = \{i \mid X_i \text{ is one of the } k(n) X_i \text{ is nearest to } x\}.$

Estimators of this type, though perhaps convenient mathematically (no ratio problem is raised), are not well adapted to large scale computers. They bear the same relation to the Mahalanobis estimates as the density estimator (2.2) does to the histogram.

3. Artificial examples

Two samples of 100 were drawn from a bivariate normal distribution with zero means, unit variances and a correlation of $\rho = 0.8$ so $E(Y \mid X = x) = m(x) = .8x$. The estimator (2.14) was used with the triangular function

$$\delta_n(z) = \begin{cases} \delta(1 - \delta|z|) & \text{for } |z| \leqslant \frac{1}{\delta}, \\ 0 & \text{for } |z| > \frac{1}{\delta}. \end{cases} \qquad \ldots \quad (3.1)$$

A series of values of δ were used and $\hat{m}_n(x)$ and $\hat{f}_1(x)$ were printed out for a range of values of x—these choices must be made with an eye to the data but it is natural to use a print-out interval of $1/\delta$. Because X is $N(0, 1)$ (standard normal), and $N = 100$, $\delta = 2, 5, 10$ and print out points -2 to 2 at intervals of 0.5 are reasonable choices. The results are shown in Fig. 1, in which computed points are simply joined by straight

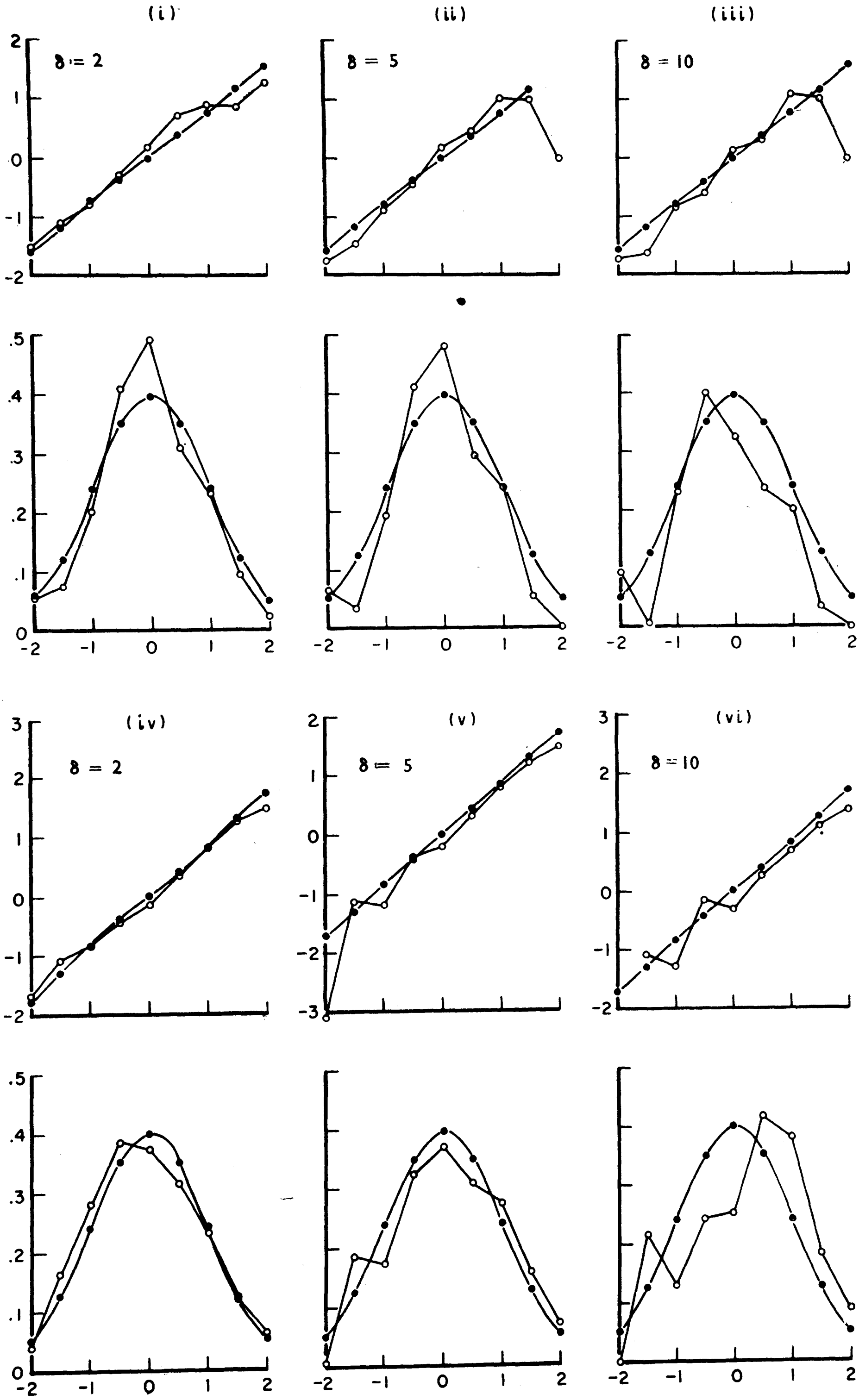


Fig. 1. Estimates from 2 samples of 100 of the regression $y = 0.8x$ and the density $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ are shown for $\delta = 2, 5, 10$.

lines. The programme also found (i) the least squares line through the origin which is indistinguishable from the true line, (ii) the estimates of the probability density of X, which are also shown.

In the next experiment, X was $N(0, 1)$ and Y, given $X = x$, was $N(3x^2, 1)$. Thus $m(x) = 3x^2$. Again (3.1) was used with $\delta = 2$, 5, 10 and the same print-out points. The results are shown in Fig. 2 for one sample of 100. The least squares quadratic through the origin was not perceptibly different from the true line.

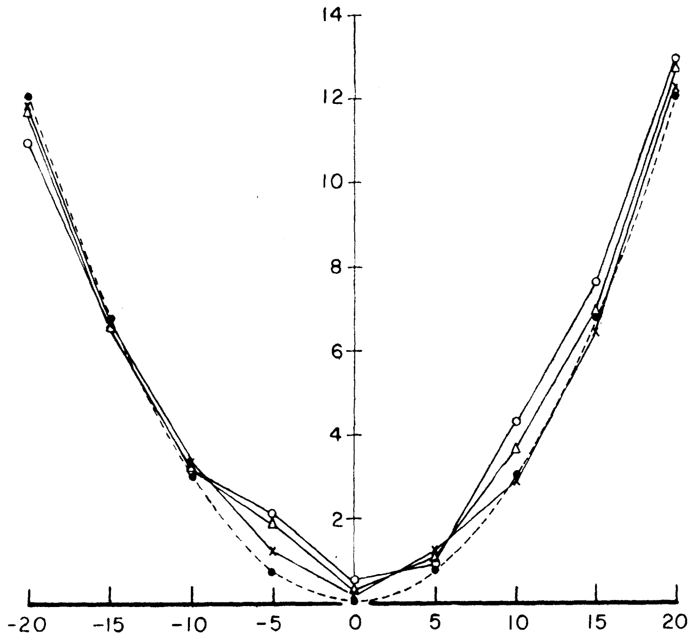


Fig. 2. The regression estimate of the true regression $y = 3x^2$ is shown for $\delta = 2(\cdot)$ $\delta = 5(\Delta)$ and $\delta = 10(\cdot)$.

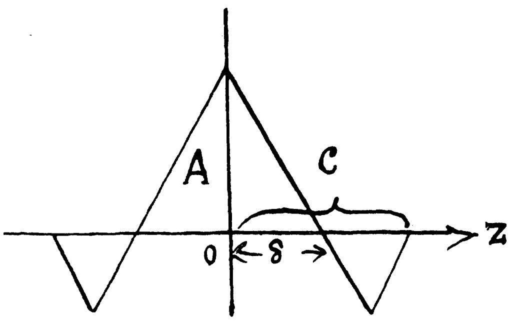
The results seem to be very encouraging and speak for themselves. As might be expected (i) the estimators go wild at the ends of the range (ii) estimators smooth out the minimum of x^2 slightly. Both these difficulties could be reduced by using an estimator that involves more smoothing in the tails of the X distribution and less smoothing for X near zero. Estimators of this type have been experimented with extensively in our other investigations on the hazard function and probabilities densities. They are easier to construct and use numerically than to study mathematically. Typical of these estimators would be one of the general form (2.14) but where $\delta(x-X_i)$ is replaced by a triangle of area one, top vertex above X_i, bottom vertices m_l observations to left of X_i and m_r observations to the right of X_i. Many of these have been tried without dramatic improvements being noticed. Fig. 3(i) and (ii) shows the results of using two such triangular smooting "functions." Estimators of a different type follow if one chooses $\delta(z)$ to pass a quadratic unchanged. All $\delta(z)$ pass a constant in the sense that

$$\int_{-\infty}^{\infty} \delta(z) \cdot c \, dz = c.$$

Further

$$\int_{-\infty}^{\infty} \delta(z) \, z \, dz = 0$$

if $\delta(z)$ is symmetrical about 0. To pass a quadratic R. H. Jones suggested a "triple triangle"



... (3.2)

where $$\delta A = \frac{(2-4\sqrt{2})^2}{2-\sqrt{2}}, \qquad c = \frac{\delta}{2-\sqrt{2}}.$$

The use of (3.2) on the parabola with $\delta = 0.25, 0.5, 1.0$ is shown in Fig. 4. The error, particularly at $x = 0$, does seem slightly less. That a variety of windows lead to much the same results with samples of size 100 was to be expected—the general amount of smoothing involved rather than the detailed method of smoothing is what matters. In the above examples, this means choice of δ.

4. Biological example

As a real example with some interesting features, we were able to use data from a survey, kindly put at our disposal by Dr. Warren Winkelstein, Jr. It was being analyzed by Mr. Roger Priore then of the department of Biostatistics, The Johns Hopkins University. The data was in the form of 586 IBM cards, the first 65 columns all conveying information. The relation of blood pressure and age was one object of the study. Ignoring the many subclassifications, there is much scatter and nothing to suggest a simple polynomial regression relationship. Applying the method of Section 2, to one of the blood pressure measurements, the results illustrated in Fig. 5 were obtained. The value of δ was varied until the "right" amount of smoothing was obtained—the judgement being based mainly on whether the associated estimate of

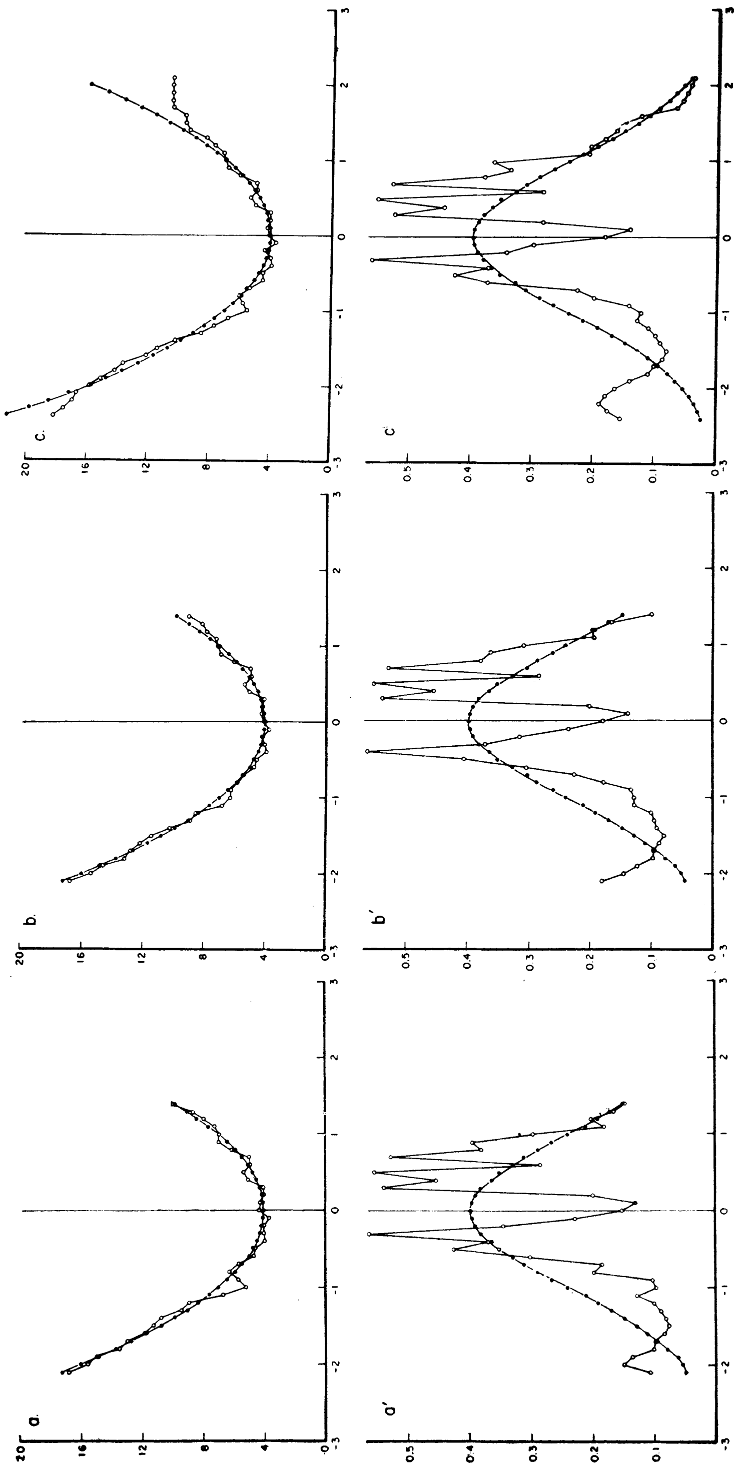


Fig. 3. Estimates of the regression $4+3x^3$ and the standard normal density from a sample of 100, using a triangular window with a vertex above x and a base chosen to be least subject to (a) 4 sample points on each side of the vertex, base greater than 0.2 (b) 5 sample points on each side of the vertex, base greater than 0.2 (c) 10 points in base of triangle whose length is greater than 0.2.

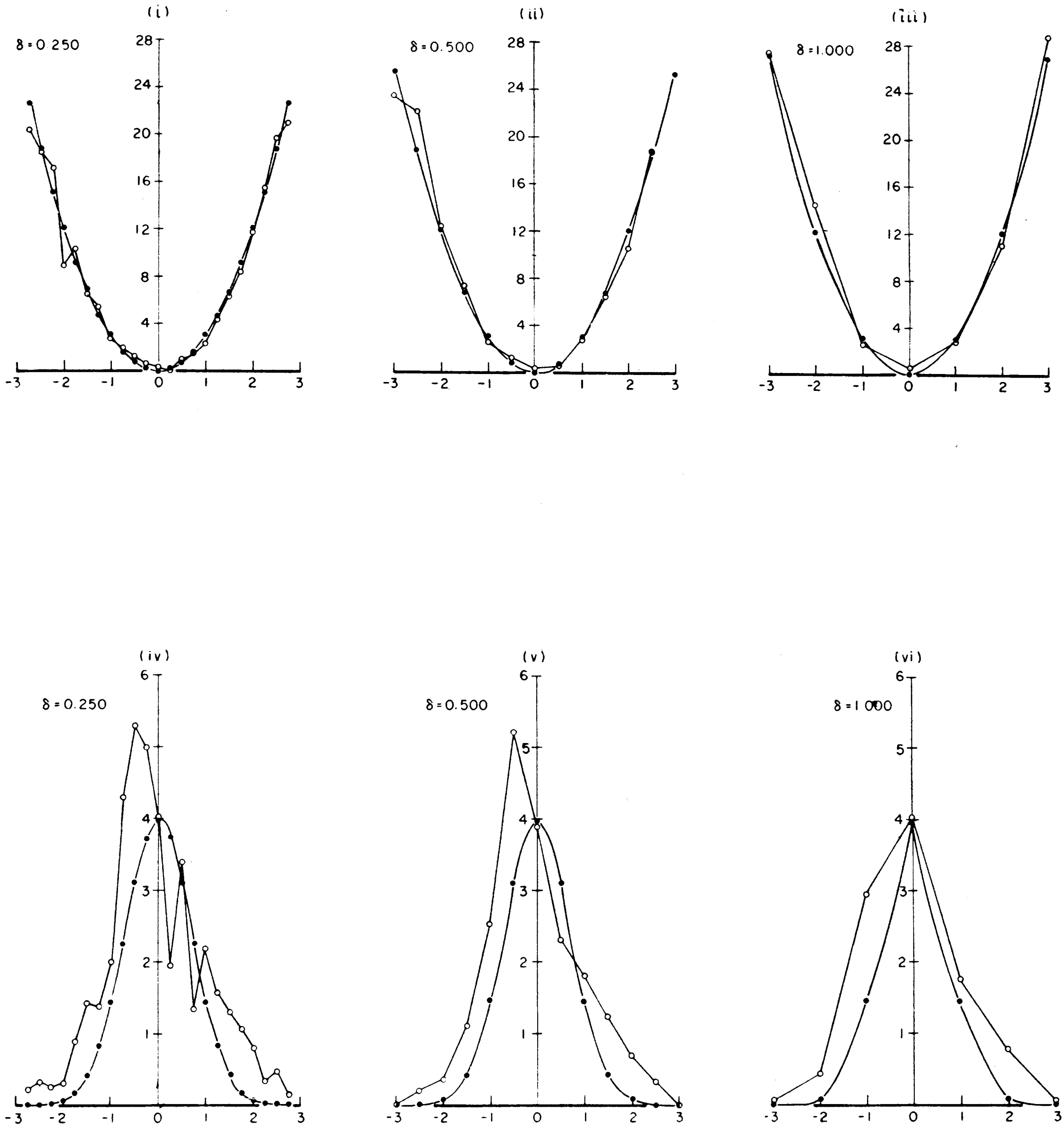


Fig. 4. The estimates of the regression $y = 3x^2$ and the density $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ from a sample of 100 using Jones' window (3.2). The smoothing increases with δ.

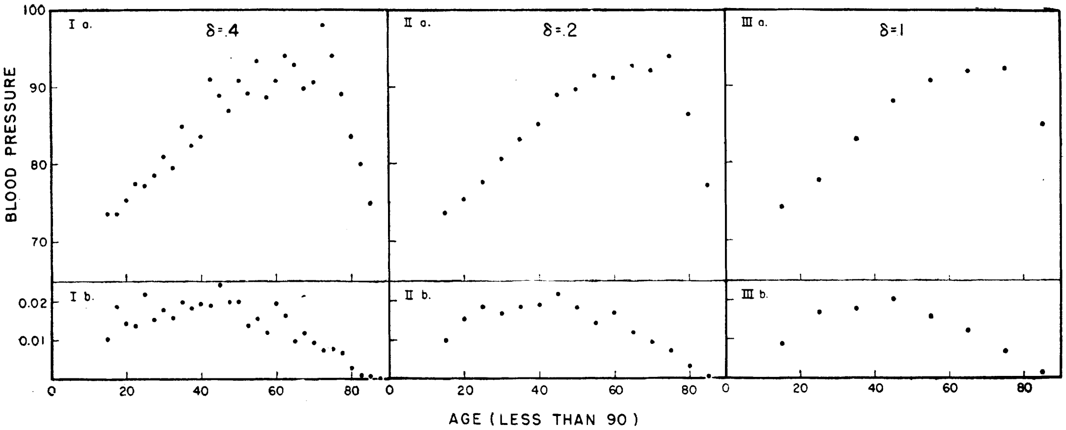


Fig. 5. The regression (I, II, III(a)) of a blood pressure measurement on age for a large body of data is shown, the smoothing being adjusted until the age distributions (I, II, III(b)) looked reasonable. The triangular window was used.

the age density seemed sensible. The estimates were computed at intervals of $1/\delta$. In this way an objective estimated regression curve (Fig. 5 (iii)) was obtained. While this might be considered for point prediction, the scatter about the line is still large (the individual points are not shown because of the graphical difficulties). This prompted Mr. Priore to perform a rather different "non-parametric regression" analysis—the subjects were grouped by age and the percentiles of these distributions plotted against age. This suggests yet another variant of these smoothing methods—grouping the ages by the percentiles of the X sample and estimating the conditional distribution of Y and hence its percentiles, or directly its percentiles. The method of deviation of (2.14) contains yet another alternative method since it is based on an estimate of $f(x, y)$ and from this we can clearly get plots of the percentiles against x.

5. Generalizations

It is common to assume linear models for time series—for example the (linear) autoregressive process

$$x_t = \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p} + \varepsilon_t, \qquad \dots \quad (5.1)$$

where the ε_t are independent identically distributed variables. For the particular case $p = 1$ one is tempted to check the linearity by using the theory of Section 2 on the pairs (x_{t-1}, x_t). This makes intuitive sense and, no doubt the theory could be extended to cover this case. Thus we are checking the model $x_t = \rho x_{t-1} + \varepsilon_t$ against the alternative of $x_t = m(x_{t-1}) + \varepsilon_t$ for some general function m.

Jones (1963) analyzed 20 years of records of the daily maximum temperatures in Central Park, New York City. After subtracting the daily means he found that the best linear predictor used the last three days; in fact he gave the model

$$\begin{cases} x_t = .605 x_{t-1} + .102 x_{t-2} - .040 x_{t-3} + \varepsilon_t \\ \hat{\mathrm{var}}(\varepsilon_t) = 49.46, \quad \hat{\mathrm{var}}(x_t) = 73.97. \end{cases} \qquad \dots \quad (5.2)$$

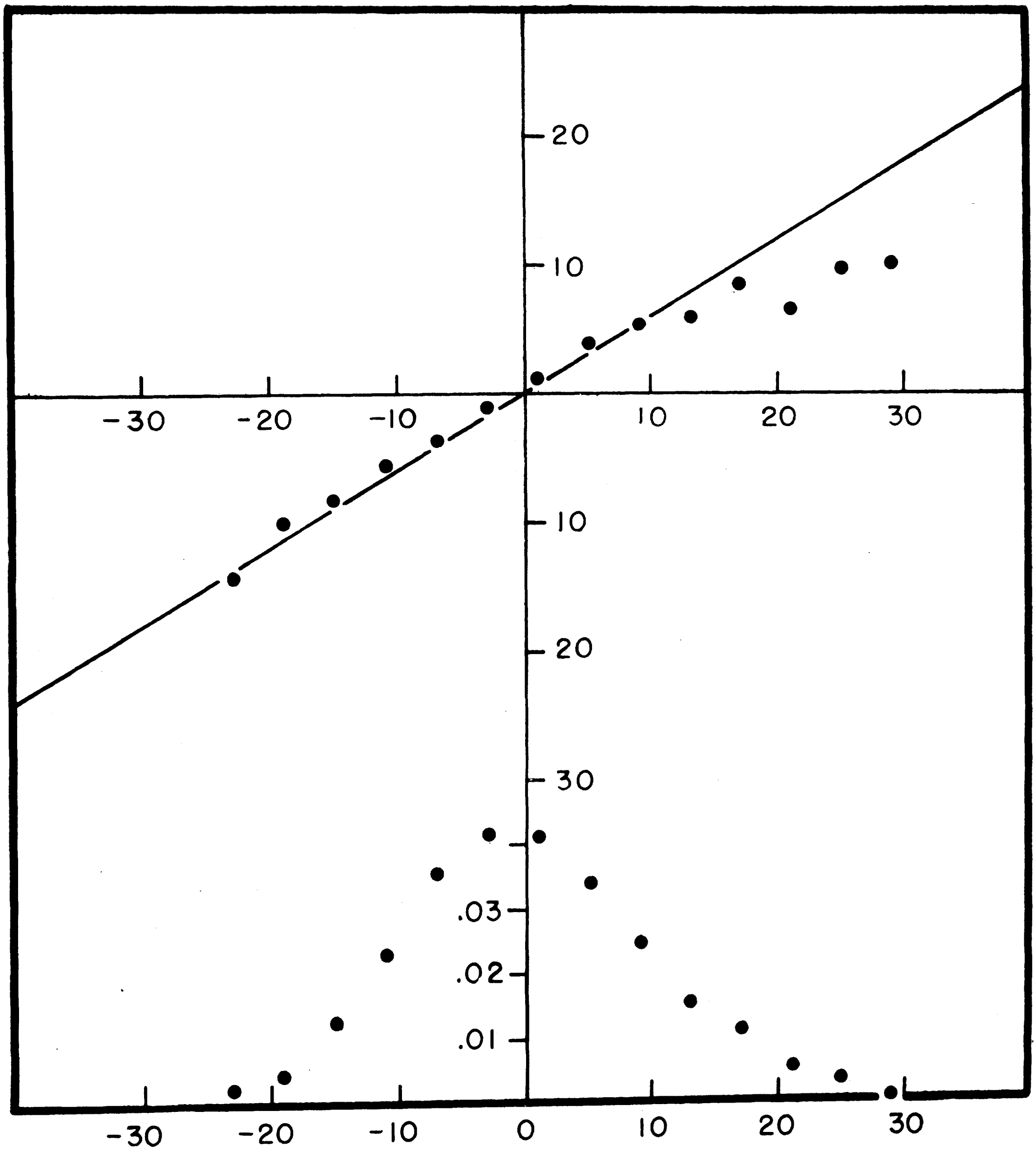


Fig. 6. The regression of a daily temperature on the previous day and the density of temperature.

Hence the pairs (x_{t-1}, x_t) should, if the model is truely linear, fall near the line $x_t = .6x_{t-1}$. The result of computing (2.14) (using the data : 1951-1960, the 73 days each year from March 1 onwards) is shown in Fig. 6, along with the estimate (2.2) and the line $x_t = .6x_{t-1}$. The temperature deviations covered a range (−30°F, +30°F) roughly so that (2.14) and (2.2) were printed out at unit intervals from −20 to 20. A triangular window with $\delta = 0.25$ was used which seems reasonable in view of the temperature distribution suggested by (2.2). The linearity for smaller deviations is obvious. For large deviations the method (2.14) may become unreliable because of a paucity of observations. Thus, in summary, Fig. 6 shows that the dependence of the temperature *is* "usually linear."

Regression problems often involve conditioning on more than one variable. The time series example above suggests the need for checking the linearity of the process with more than one time lag.

Theoretically this generalization is no problem. Presume as given a sample of n values of (X_{1i}, X_{2i}, Y_i), $i = 1, \ldots, n.$, and a two dimensional approximation $\delta_n(z_1, z_2)$ to a two dimensional Dirac delta function with

$$\int \int \delta_n (z_1, z_2)\, dz_1\, dz_2 = 1.$$

The estimator of

$$m(x_1, x_2) = E(Y \mid X_1 = x_1,\ X_2 = x_2) \quad \ldots \quad (5.3)$$

that would naturally be suggested is

$$\hat{m}(x_1, x_2) = \frac{\sum\limits_{i=1}^{n} Y_i\, \delta_n(x_1 - X_{1i},\ X_2 - X_{2i})}{\sum\limits_{i=1}^{n} \delta_n(x_1 - X_{1i},\ x_2 - X_{2i})}. \quad \ldots \quad (5.4)$$

In practice one could try the "pyramid" function—the rational generalization of the triangle function — or any bivariate density function. Some experimenting is suggested to get the relative scales on the two axes correct.

This "choice of scale" problem suggests the desirability of the scales of all fixed variables being transformed to ones where the random variables are uniformly distributed on (0, 1)—or approximately so. This corresponds to "prewhitening" in spectral density estimation, and leads to simpler estimates of densities and regressions since the denominators in (2.14) and (2.14) and (5.4) may be replaced by n. Monte Carlo studies did *not* show any superior virtues of the method in practice.

6. Acknowledgements

The writer is grateful to Eugene Sobel for doing the computations and Dr. Richard H. Jones for some helpful conversations and for providing the time series data. The writer also wishes to thank Dr. Warren Winkelstein, Jr., Chief of Department of Epidemiology, Public Health Research Institute for Chronic Diseases of the State of New York for the use of his data.

References

Bartlett, M. S. (1964): Statistical estimation of density functions. *Sankhyā*, Series A, **25**, 245–254.

Leadbetter, M. R. (1964): *Technical Report*, Research Triangle Institute, North Carolina.

Mahalanobis, P. C. (1961): A method of fractile graphical analysis. *Sankhyā*, Series A, **23**, 41–64.

Parthasarathy, K. R. and Bhattacharya, P. K. (1961): Some limit theorems in regression theory, *Sankhyā*, Series A, **23**, 91–102.

PARZEN, E. (1958) : On asymptotically efficient estimates of the spectral density of a stationary time series. *J. Roy. Stat. Soc.*, Series B, **20**, 303–322.

——— (1962) : On estimation of a probability density and mode. *Ann. Math. Stat.*, **33**, 1065–1076.

ROSENBLATT, M. (1956) : Remarks on some non-parametric estimates of a density function. *Ann. Math. Stat.*, **27**, 832–837.

TUKEY, J. W. (1961) : Curve as parameters and touch estimation. *Proc. 4th Berkeley Symp.*, **1**, 681–694, University of California Press, Berkeley.

WHITTLE, P. (1957) : Curve and periodogram smoothing. *J. Roy. Stat. Soc.*, Series B, **19**, 38–46.

——— (1958) : On the smoothing of probability density functions. *J. Roy. Stat. Soc.*, Series B, **20**, 334–343.

WATSON, G. S. and LEADBETTER, M. R. (1963) : On the estimation of the probability density I. *Ann. Math. Stat.*, **34**, 480–491.

——— (1964a) : Hazard analysis I. *Biometrika*, **51**, 175–184.

——— (1964b) : Hazard analysis II. *Sankhyā*, Series A, **26**, 101–116.

Paper received : November, 1964.