Proc. Natl. Acad. Sci. USA
Vol. 74, No. 2, pp. 434–436, February 1977
Statistics

Robust nonlinear data smoothers: Definitions and recommendations

(Monte-Carlo trials/time-series data/data noise reduction/nonlinear filters/exploratory data analysis)

PAUL F. VELLEMAN*

Department of Statistics, Princeton University, Princeton, New Jersey 08540

Communicated by John W. Tukey, November 22, 1976

ABSTRACT Nonlinear data smoothers provide a practical method of finding smooth traces for data confounded with possibly long-tailed or occasionally "spikey" noise. While they are natural tools for analyzing time-series data, they can be applied to any data set for which a sequencing order can be established. Their resistance to the effects of unsupported extreme observations and their ability to respond rapidly to well-supported patterns make them valuable as tools for finding patterns not constrained to specific parametric form and as versatile data-cleaning algorithms. This paper defines some robust nonlinear smoothers that have performed well in Monte-Carlo trials and makes brief recommendations based upon that study.

Much of data analysis, and practically all of exploratory data analysis, consists of looking for patterns in data. Frequently the underlying pattern is obscured by measurement error and other "noise." Extracting the pattern can be difficult, especially when this noise has a long-tailed distribution. Data patterns are frequently expected to be smooth; i.e., each data point well supported by the points in its vicinity. This can happen in several ways. A point can be at about the same level as its neighbors, or it can be consistent with changes at a steady or steadily changing rate. If there is a jump, the points on either side of it can be supported on one side. Peaks can be supported by a consistent trend up to and down from the top. However, a small number of extreme data points not otherwise supported will often be best considered not to be part of the pattern.

These deliberately vague definitions implicitly define noise as the extent of any excursion insufficiently supported in its vicinity, and suggest seeking data smoothers resistant to influence by noise with occasional "spikes" or a long-tailed distribution. Such smoothers must be nonlinear in the sense that, for arbitrary sequences $\{x_t\}$ and $\{y_t\}$, and smoother Sm, it need not be true that $Sm\{x_t + y_t\} = Sm\{x_t\} + Sm\{y_t\}$ (otherwise the low frequency component of noise spikes cannot be removed). Consequently, they cannot be analyzed precisely within the classical transfer function framework of linear filter theory. We do not yet have a deep mathematical understanding of the smoothers presented in this paper, but we have empirical evidence of their good performance from both computer simulation studies and practical data analysis experience. In some situations—especially when finding smooth traces for data confounded with long-tailed or spikey noise—these smoothers out-perform linear methods. In general application, they are effective tools for finding patterns not constrained to specific forms and can be used as versatile data-cleaning algorithms.

Invariances and symmetries

All the smoothers considered here commute with the simplest modifications of data sequences. For input sequence $\{y_t\}$ and constants a or c,

$$Sm\{y_t + a\} = Sm\{y_t\} + a$$

$$Sm\{cy_t\} = cSm\{y_t\}, \text{ all } c \text{ including } c < 0 \text{ and } c = 0.$$

The smoothers described here are invariant to choice of t-axis origin and direction.

Definitions and notation

For a data sequence, $\{y_t\}$, the t dimension can be any sequencing variable, not necessarily time. For simplicity, t is treated as if equispaced, but the smoothers seem to perform well when spacing is somewhat uneven. Any data smoother produces two sequences, the Smooth, $\{z_t\} = Sm\{y_t\}$ and the Rough, $\{r_t\} = \{y_t - z_t\}$. Compound smoothers can be constructed by *resmoothing* the Smooth, $\{z_t\} = Sm_2\{Sm_1\{y_t\}\}$, or by its dual operation, *reroughing*, $\{z_t\} = Sm_1\{y_t\} + Sm_2\{y_t - Sm_1\{y_t\}\}$, which could be written $Sm_1\{y_t\} + Sm_2\{r_t\}$. Resmoothing is denoted by juxtaposing the smoother names: Sm_1Sm_2.

Reroughing attempts to recover pattern from the residual of the first smoothing by applying a second smoother and adding the smoothed Rough to the first Smooth sequence. Reroughing of smoother Sm_1 with smoother Sm_2 is denoted $Sm_1 + Sm_2$. If the second smoother is identical to the first, then we speak of "Sm_1, twice." Reroughing is associative but not commutative.

Elementary data-smoothing units

The simplest nonlinear smoothing procedure is the running median of length v. It is defined as follows:

For odd length, $v = 2u + 1$

$$z_t = \text{med}\,\{y_{t-u}, \ldots, y_t, \ldots, y_{t+u}\}.$$

For even length, $v = 2u$, the median falls naturally at a point halfway between two values of t. For equispaced t this can be remedied by resmoothing with a running mean of length 2 to produce a smoothing unit which is properly centered in t, and can be thought of as:

$$z_t = \tfrac{1}{2}\,\text{med}\,\{y_{t-u}, \ldots, y_t, \ldots, y_{t+u-1}\} + \tfrac{1}{2}\,\text{med}\,\{y_{t-u+1}, \ldots, y_t, \ldots, y_{t+u}\}.$$

Note that this smoother in fact employs $2u + 1$ data values, but gives only half weight to the outer ones.

At the ends of a finite sequence, the length of the median, v, must decrease by 2 for each step toward the end. The endpoints themselves can be copied or estimated with a special algorithm.

Odd-length running medians do not modify monotone sequences. A running median of length v will annihilate unsupported excursions of length $[(v + 1)/2] - 1$, where $[x]$ is the greatest integer not exceeding x. Running medians are named

* Present address: Department of Economic and Social Statistics, New York State School of Industrial and Labor Relations, Cornell University, Ithaca, N.Y. 14850.

with their lengths and denoted by the numeral of their length. Once an application of running medians has removed outliers, it is safe to apply linear smoothers. These are usually weighted running means. They are defined, for $v = 2u + 1$, by

$$z_t = \frac{1}{W}\sum_{j=1}^{v} w_j y_{t-u+j-1},\ W = \sum_{j=1}^{v} w_j.$$

For even length, $v = 2u$, the smoother uses $v + 1$ data points, but the outer weights are reduced (typically by a factor of ½ for symmetric weights).

Some special running means are used often enough to deserve names. One of these is the running mean of length 3 with weights ¼, ½, ¼, called hanning (after J. von Hann, who liked to use it) and denoted H.

A more sophisticated weighted running mean, the cosine bell of length v, has weights

$$w_{[v/2]+1-i} = w_{[(y+1)/2]+i} = 1 + \cos\left(\frac{\pi i}{(v/2)+1}\right),\quad i = 0, 1, 2, \ldots, [v/2].$$

Cosine bell running means can be used with linear smoothers to reduce the size of the side lobes in the transfer function. They have a similar effect when incorporated in nonlinear smoothers, and are often used to "taper" the robust smoothers, defined below.

Robust theory

The robust M-estimate of location, discussed in (ref. 1) and (ref. 2), suggests a general form that encompasses both the above smoothers and a class of adaptive robust nonlinear smoothers. These behave like linear smoothers in the presence of short-tailed noise, but like robust nonlinear smoothers in the presence of long-tailed noise.

The M-smoothed, T-tapered sequence $\{z_t\}$ would be the sequence of solutions at each t of

$$\sum_{i\in V_t} T_i \psi(y_{t+i} - z_t) = 0 \qquad [1]$$

where $i \in V_t$ is read "i is a number such that $(t + i)$ is in the vicinity of t" or briefly, "i in the vicinity of t." The $\{T_i\}$ taper the smoother in the t dimension, and $\Sigma_{i\in V_t} T_i = 1$. The function $\psi(.)$ is an influence function (refs. 1, 3, and 4). (If [1] were to be solved, it would have to be by iteration—Newton-Raphson has worked for similar location estimates, but would be expensive. A computationally practical modification appears below.) Note that for $\psi(u) = u$, [1] becomes a standard linear smoother with weights $\{T_i\}$. For $\psi(u) = \mathrm{sgn}(u)$ and T_i constant, we obtain medians of length determined by the definition of V_t.

We usually use a standardized argument, $u_i = (y_{t+i} - Z_t)/cS_t$, in $\psi(.)$, where Z_t is a first approximate Smooth (usually produced by an elementary nonlinear smoother), S_t is a measure of scale in some vicinity of t (usually one including V_t as a subset), and c is chosen according to the nature of S_t.

Practical robust computation

A computationally practical approximation to [1] is the w-smoother, derived by substituting

$$w(u) = \begin{cases} \frac{\psi(u)}{u}, & u \neq 0 \\ \psi'(0), & u = 0 \end{cases}$$

in [1] to obtain

$$z_t = \frac{\sum_{i\in V_t} T_i w\left(\frac{y_{t+i} - Z_t}{cS_t}\right) y_{t+i}}{\sum_{i\in V_t} T_i w\left(\frac{y_{t+i} - Z_t}{cS_t}\right)} \qquad [2]$$

Specific values for Z_t, S_t, $w(.)$, T_i, and c in [2] are discussed below.

Testing and calibration

Smoothers were tested by applying them to zero-phase sinusoids, $\sin(2\pi jt/N)$, $j = 1, 2, \ldots, N/2$; $t = 0, 1, \ldots,$ to which white noise of differing distribution shapes and scales had been added. The regression of the smoother output on a noise-free version of the signal sinusoid to obtain the regression coefficient B_j provided, in B_j^2, a measure of the power present at the signal frequency. This served as one measure of smoother performance. The power present in the smoother output at other frequencies was also monitored to trap excessive transport of signal power to other frequencies—modulation which can occur with some nonlinear smoothers. Plots of $\log(B_j^2)$ against j serve many of the purposes of linear-filter transfer functions, (refs. 5 and 6), although affected by the amount and type of noise introduced.

Compound smoothers

Few of the smoothing units discussed here perform well by themselves, so they are frequently combined. Two compound smoothers due to Tukey have appeared in print (refs. 2, 7, and 8) and perform well in many circumstances. They are 53H and (53H), twice. Two smoothers better in many ways, but more difficult to compute by hand, are 4253H and (4253H), twice, and these are recommended at this time for simple smoothing tasks. (4253H), twice was affected only slightly by long-tailed noise and negligibly by Gaussian white noise. It appears to have good low-pass transfer behavior with side lobes down 30 dB, and modulation no worse than −20dB at any frequency.

For more sensitive smoothing tasks, the w-smoother using

Z_t = Smooth when 4253H is applied to the original data

S_t = $Sm(\mathrm{med}_{i\in V_t}\{|y_{t+i} - Z_t|\})$ where Sm is 53H, (the smoothed median absolute deviation, or SMAD)

T_i = scaled cosine-bell weights for length determined by V_t

c = 6

$$w(u) = \begin{cases}(1-u^2)^2, & |u| \le 1 \\ 0, & \text{else}\end{cases} \text{ (the "biweight")}$$

performed as well as the best simple nonlinear compound smoothers in any single situation but exhibited more consistent performance for signals contaminated with Gaussian noise and better robustness against long-tailed noise.

I wish to thank John Tukey and Peter Bloomfield for their help and support during this research. I thank Alan Gross and Douglas Dunn for their comments on earlier drafts of this manuscript, and the referees for valuable suggestions to clarify the exposition. This work was supported in part by AEC (now Energy Research and Development Administration). Contract no. AT(11-1)2310 awarded to the Department of Statistics, Princeton University.

1. Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H. & Tukey, J. W. (1972) *Robust Estimates of Location: Survey and Advances* (Princeton University Press, Princeton, N.J.).

2. Beaton, A. E. & Tukey, J. W. (1974) "The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data," *Technometrics* **16**, 147–186.
3. Hampel, F. R. (1974) "The influence curve and its role in robust estimation," *J. Am. Stat. Assoc.*, **69**, 383–393.
4. Huber, P. J. (1972) "Robust statistics: A review," *Ann. Math. Stat.* **43**, 1041–1067.
5. Bingham, C., Godfrey, M. D. & Tukey, J. W. (1967) "Modern techniques of power spectrum estimation," AU-15, *Trans. IEEE in Audio and Electroacoustics*, pp. 56–66.
6. Blackman, R. B. & Tukey, J. W. (1958) *The Measurement of Power Spectra from the Point of View of Communications Engineering* (Dover Press, New York).
7. Tukey, J. W. (1971) *Exploratory Data Analysis* (Addison-Wesley, Reading, Mass.) 3 Vols. (limited preliminary edition).
8. Tukey, J. W. (1977) *Exploratory Data Analysis* (Addison-Wesley, Reading, Mass.).