The Remedian: A Robust Averaging Method for Large Data Sets

PETER J. ROUSSEEUW and GILBERT W. BASSETT, JR.*

It is often assumed that to compute a robust estimator on n data values one needs at least n storage elements (contrary to the sample average, that may be calculated with an updating mechanism). This is one of the main reasons why robust estimators are seldom used for large data sets and why they are not included in most statistical packages. We introduce a new estimator that takes up little storage space, investigate its statistical properties, and provide an example on real-time curve "averaging" in a medical context. The remedian with base b proceeds by computing medians of groups of b observations, and then medians of these medians, until only a single estimate remains. This method merely needs k arrays of size b (where $n = b^k$), so the total storage is $O(\log n)$ for fixed b or, alternatively, $O(n^{1/k})$ for fixed k. Its storage economy makes it useful for robust estimation in large data bases, for real-time engineering applications in which the data themselves are not stored, and for resistant "averaging" of curves or images. The method is equivariant for monotone transformations. Optimal choices of b with respect to storage and finite-sample breakdown are derived. The remedian is shown to be a consistent estimator of the population median, and it converges at a nonstandard rate to a median-stable distribution.

KEY WORDS: Curve averaging; Median-stable laws; Recursive estimation; Single-pass algorithms; Storage space.

1. INTRODUCTION

In spite of a growing awareness of the dangers posed by outliers, many statistical packages still restrict themselves to the sample average as a summary value of n data points, rather than include the sample median as well. In many scientific disciplines, alternatives to the average are not even considered. One of the main reasons is that the sample average can easily be computed with an updating mechanism, so only a single pass through the data is necessary. For instance, the following FORTRAN lines may be used:

```
      DO 10 I = 1, N
10    SUM = SUM + ENTER(I)
      AVERA = SUM/N
```

where ENTER is a function that reads, records, generates, or otherwise accesses the ith observation (for instance, by looking it up in a large data base residing on an external device). Therefore, it is never necessary to store the data in central memory, so software vendors can rightly claim that their packages can cope with 100,000 observations. It is commonly thought that all robust estimators would need to store at least the data, so the resulting storage would be essentially $O(n)$ or more. In many applications $O(n)$ storage is infeasible, especially when a lot of estimations has to be carried out simultaneously, as in the case of averaging sequences of curves or images with many pixels, or in real-time engineering applications where the data are not stored.

To remedy this problem, we propose a new robust estimator that can also be computed by means of a single-pass updating mechanism, without having to store the observations. Let us assume that $n = b^k$, where b and k are integers (the case where n is not of this form will be treated in Sec. 7). The *remedian with base b* proceeds by computing medians of groups of b observations, yielding b^{k-1} estimates on which this procedure is iterated, and so on, until only a single estimate remains. When implemented properly, this method merely needs k arrays of size b that are continuously reused. Figure 1 illustrates the remedian with base 11 and exponent 4. The data enter at the top, and array 1 is filled with the first 11 observations. Then the median of these 11 observations is stored in the first element of array 2, and array 1 is used again for the second group of 11 observations, the median of which will be put in the second position of array 2. After some time array 2 is full too, and its median is stored in the first position of array 3, and so on. When $11^4 = 14{,}641$ data values have passed by, array 4 is complete and its median becomes the final estimate. This method uses only 44 storage positions, and its speed is of the same order of magnitude as that of the ordinary average.

In general, the remedian with base b and exponent k merely needs bk storage spaces for sample size $n = b^k$. (We call b the *base* by analogy to positional number systems. We could take $b = 10$, but we prefer odd b because then the medians are easier to handle.) The basic idea of the remedian is quite natural, since both of us arrived at it independently.

The remedian could easily be incorporated in software packages: By means of just 15 arrays of 11 real numbers each (or a 15×11 matrix) one would be able to process as many as 11^{15} numbers, which ought to be sufficient for all applications. Figure 2 shows a FORTRAN implementation of the remedian corresponding to Figure 1.

The remedian is affine equivariant, because it transforms well when all observations x_i are replaced by $cx_i + d$, where c and d are arbitrary constants. Like the sample median, it is even equivariant with respect to any *monotone* transformation of the x_i, such as a power function or a logarithm. (The data need not even be numbers at all!)

* Peter J. Rousseeuw is Professor, Vrije Universiteit Brussel, Vesaliuslaan 24, B-2520 Edegem, Belgium. Gilbert W. Bassett, Jr., is Professor, Department of Economics, University of Illinois at Chicago, IL 60680. The authors are grateful to Bill Eddy and John Tukey for providing information on related methods. Thanks also go to the editor, associate editor, and referees for useful suggestions.

Journal of the American Statistical Association
March 1990, Vol. 85, No. 409, Theory and Methods

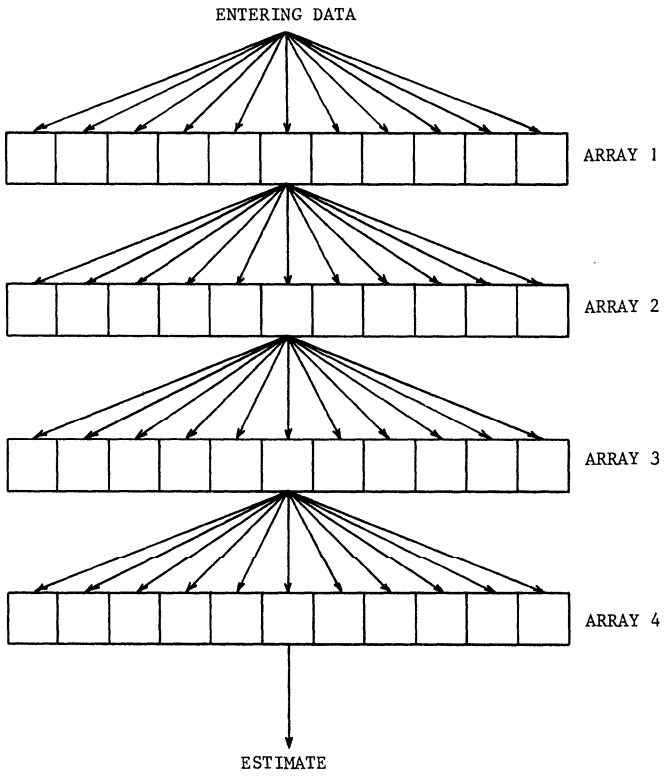

Figure 1. Mechanism of the Remedian With Base 11 and Exponent 4, Using 44 Storage Spaces for a Data Set of Size $n = 11^4 = 14{,}641$.

On the other hand, the remedian is not invariant with respect to permutations of the observations, which appears to be the price for its low storage. It can be seen as an order statistic with random rank, for which the distribution of the ranks does not depend on the original distribution of the x_i (see Sec. 4). Looking at the rank distribution, it can be verified that the remedian is *nearly* permutation invariant.

Sections 3 and 4 discuss some possible choices of b and k from the perspectives of storage economy and robustness. In Sections 5 and 6 we show that the remedian is a consistent estimator of the underlying population median, and that it converges at a nonstandard rate to a limiting distribution that is close to a Gaussian law. In Section 7 we look at some extensions and alternative approaches.

2. APPLICATIONS TO AVERAGING

Suppose we want to obtain a certain curve corresponding to a physical phenomenon. A curve can be registered by means of a list of its function values $x(t)$ at equally spaced arguments t (usually t represents time). Because the observed values of $x(t)$ are subject to noise one repeats the experiment several times, yielding n curves in all, so the data are of the form

$$\{x_i(t): t = 1, \ldots, T\} \quad \text{for } i = 1, \ldots, n. \qquad (2.1)$$

One wants to combine the n curves to estimate the true underlying shape. The classical approach is *averaging*, yielding the curve

$$\bar{x}(t) = \frac{1}{n}\sum_{i=1}^{n} x_i(t), \qquad t = 1, \ldots, T. \qquad (2.2)$$

This, however, assumes Gaussian noise and no outliers! The averaging technique is very common in engineering and medicine. For instance, averaging is built into many special-purpose instruments used in hospitals [e.g., the microprocessor-based average recorders in Pauwels, Vogeleer, Clement, Rousseeuw, and Kaufman (1982) and Trau et al. (1983)].

Usually T and n are quite large, so one cannot store all of the observed curves in central memory. This precludes calculation of the "median curve"

$$\underset{i=1,\ldots,n}{\text{median}}\, x_i(t), \qquad t = 1, \ldots, T, \qquad (2.3)$$

as well as many other robust summaries. We propose to compute the remedian instead, because it is a robust single-pass method.

The program of Figure 2 can be easily adapted to produce the remedian curve, by replacing the arrays A1, A2, A3, and A4 of length 11 by matrices with 11 rows and T columns. In this way the total storage becomes $44T$, whereas the plain median would have needed $14{,}641T$ positions.

Let us consider a medical example. The *electroretinagram* (ERG) is used in ophthalmology to examine disorders of the visual system. When the eye is exposed to a white flash of light, it develops a small electric potential. The ERG curve shows the evolution of this evoked potential (in microvolts) as a function of time (in milliseconds). The bottom curve in Figure 3b is a standard ERG of a healthy patient (from Trau et al. 1983). The important features are the four peaks (denoted by a, b, OP_1, and OP_2) and in particular their t coordinates, which are used for medical diagnosis.

When the ERG curve is recorded only once, the noise typically dominates the signal so that no peak can be found. The current solution is to record many curves by repeating the stimulus flash of light, and then to average them. The average curve is often deformed and difficult to interpret, however, because of a high amount of contamination caused by electrical interference, involuntary eye movements, and other artifacts.

It is quite feasible to replace the averaging routine in

```
CC     A PROGRAM FOR THE REMEDIAN
CC     -------------------------
       DIMENSION A1(11),A2(11),A3(11),A4(11)
       DO 40 M=1,11
       DO 30 L=1,11
       DO 20 K=1,11
       DO 10 J=1,11
       I=I+1
   10  A1(J)=ENTER(I)
   20  A2(K)=FMED(A1)
   30  A3(L)=FMED(A2)
   40  A4(M)=FMED(A3)
       REMED=FMED(A4)
       WRITE(*,*)REMED
       STOP
       END
```

Figure 2. Small Program Yielding the Remedian With Base 11 and Exponent 4. Here, ENTER is a function that reads or otherwise accesses the ith observation, and FMED returns the median of an array of 11 numbers.

the recording instrument by the remedian, because the latter is equally fast and does not need too much storage. To verify if this replacement is worthwhile, computer simulations were performed in which both the average and the remedian were calculated for a bundle of curves, some of which were contaminated. The basic curve was the standard ERG of Figure 3b, measured at $T = 320$ time units. Figure 3a contains $n = 81$ curves (in ophthalmology more curves are used, but this would make the display overcrowded). The curves were generated as follows: With probability .7, curve i is the basic ERG plus some Gaussian noise with modest scale. With probability .1, the $x(t)$ values are multiplied by a random factor greater than 1. With probability .2, the curve models a response at half the standard speed, again with magnified $x(t)$ values.

The upper curve in Figure 3b is the average of the ERG curves in Figure 3a. It has been greatly affected by the contamination, which caused a substantial upward shift. What is worse, the average has one peak too many, rendering medical diagnosis difficult. Averaging often produces results like this in actual clinical practice. On the other hand, the 3^4 remedian lies near the original ERG and is virtually undamaged by the contamination.

Many other applications of robust averaging are possible—for instance, in spectroscopy. Median-type procedures can also be used to estimate horizontal shifts between spectrograms (Rousseeuw 1987).

Averaging also occurs in image analysis. An image may be described as a rectangular grid of pixels, each with a corresponding number $x(r, c)$ indicating its gray intensity. When n images are read one after another, the data are

$$\{x_i(r, c): r = 1, \ldots, R; c = 1, \ldots, C\} \quad \text{for } i = 1, \ldots, n, \quad (2.4)$$

where R is the number of rows and C is the number of columns. In one application, a physicist recorded images of a crystallographic lattice by means of an electron microscope, with $R = 512$, $C = 512$, and $n \approx 10{,}000$. Usually such images are averaged to obtain a sharp result, but in this case averaging did not work well because in many images a part of the lattice was contaminated or even

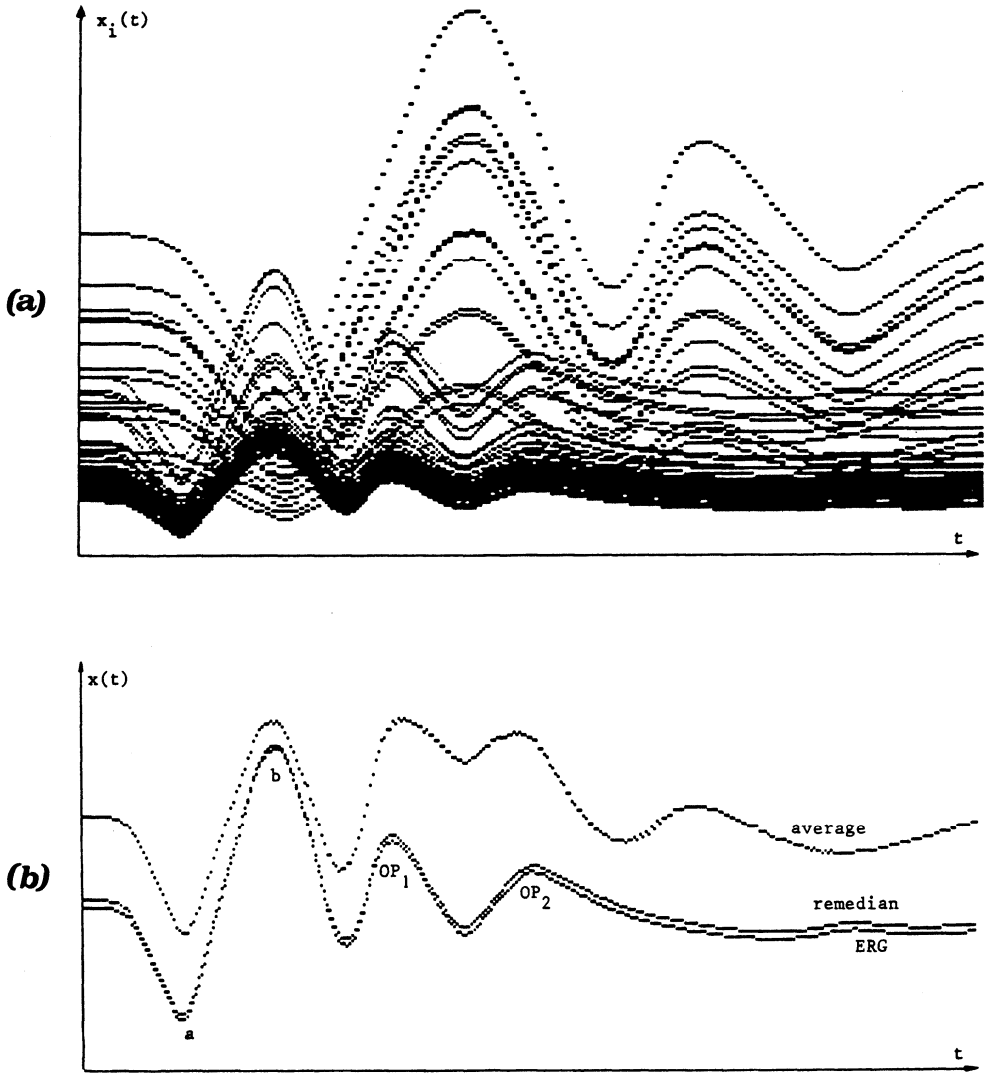

Figure 3. (a) Bundle of Simulated ERG Curves, Some With Pure Gaussian Noise and Others With Various Kinds of Contamination. (b) Plot With the Standard ERG (bottom curve), the Average of the Simulated ERG's (upper curve), and Their Remedian (middle curve).

destroyed by the radiation of the microscope itself. Computing plain medians was not feasible because there were $nRC \approx 2{,}621{,}400{,}000$ data values in all, which could not be stored in central memory. One can, however, compute the remedian image given by

$$\underset{i=1,\ldots,n}{\text{remedian}}\ x_i(r, c), \qquad r = 1, \ldots, R, \quad c = 1, \ldots, C. \tag{2.5}$$

The computation of remedian curves and images may be speeded if one has access to parallel computing facilities, because one could let each processor work on a different element. For instance, Hillis (1987, p. 87) described how his 65,536-processor Connection Machine deals with images of 256×256 pixels, by identifying each processor with a single pixel.

3. STORAGE AND COMPUTATION TIME

For given n, which base b minimizes the remedian's storage? We can write the exponent k as a function of n and b, yielding $k = \log_b(n)$. Therefore, we have to minimize

$$bk = b \log_b(n) = b \frac{\ln(n)}{\ln(b)} \tag{3.1}$$

over all b, where ln denotes the natural logarithm with respect to $e = 2.718\ldots$. Note that $\ln(n)$ is a constant, so we have to minimize $h(b) = b/\ln(b)$, which does not depend on n. Only integer values of b are allowed, and we find $h(2) = 2.885$, $h(3) = 2.731$, $h(4) = 2.885$, and $h(5) = 3.107$, after which h is monotonically increasing, although very slowly. Thus $b = 3$ gives the smallest total storage, but even if we take another fixed value of b the total storage is not necessarily much larger (for instance, for $b = 9$ the total storage is increased by a constant factor of 1.5, as seen in Table 1). We will show that the larger the base, the more robust the estimator is. Therefore, the relatively small gain in storage by taking $b = 3$ rather than, say, $b = 9$ is not worthwhile, except if we want to apply the method by hand.

We have seen that for a fixed value of b the storage is $O(\log n)$. On the other hand, we can also keep k fixed; then $b = n^{1/k}$, so the storage is $bk = kn^{1/k} = O(n^{1/k})$. Table 1 contains the remedian with $k = 2$, which needs the most storage but still looks economical compared with the plain median (given by $k = 1$) in the last column.

Let us also consider the total computation time. It is well known that the median of n numbers can be computed in $O(n)$ time (see Knuth 1973, p. 216). For the b^k remedian, we have to compute b^{k-1} medians of b observations at the first level, b^{k-2} such medians at the second level, and so on. The total time is thus proportional to

$$b^k + b^{k-1} + \cdots + b = n\left(1 + \frac{1}{b} + \cdots + \frac{1}{b^{k-1}}\right). \tag{3.2}$$

If k is fixed then b increases and (3.2) becomes asymptotically equivalent to n, so the remedian has the same speed as the usual median, with proportionality factor 1. On the other hand, when b is fixed the exponent k will grow, so

$$1 + \frac{1}{b} + \cdots + \frac{1}{b^{k-1}} \to \frac{b}{b-1},$$

which implies that the computation time is again $O(n)$, but now the proportionality factor with respect to the median is $b/(b-1)$. The larger the base, the faster the estimator will be.

The remedian could be calculated much faster by means of parallel computing. The b^{k-1} medians at the first level could be computed simultaneously by different processors, then followed by the b^{k-2} medians of the second level, and so on, yielding a total time of bk. But then all of the data would need to be stored, and the total space would again be n. Note that parallel computing interchanges the requirements of space and time in this case.

4. BREAKDOWN POINT

The finite-sample breakdown point of an estimator is defined as the smallest fraction of the observations that have to be replaced to carry the estimator over all bounds. In this definition, both the configuration and the magnitude of the outliers can be chosen in the least favorable way. The median has the best possible breakdown, because at least $\lceil n/2 \rceil$ observations have to be replaced to carry it outside the range of the original data. (The "ceiling" $\lceil q \rceil$ is the nearest integer $\geq q$.) Therefore, the break-

Table 1. Total Storage Space and Finite-Sample Breakdown Point for Different Versions of the Remedian

	Remedian ($b = 3$)		Remedian ($b = 9$)		Remedian ($k = 2$)		Median ($k = 1$)	
n	Storage	ε_n^*	Storage	ε_n^*	Storage	ε_n^*	Storage	ε_n^*
9	6	44%	9	56%	6	44%	9	56%
81	12	20%	18	31%	18	31%	81	51%
729	18	9%	27	17%	54	27%	729	50%
6,561	24	4%	36	10%	162	26%	6,561	50%
59,049	30	2%	45	5%	486	25%	59,049	50%
531,441	36	1%	54	3%	1,458	25%	531,441	50%
4,782,969	42	0%	63	2%	4,374	25%	4,782,969	50%

down point of the sample median equals $\lceil n/2 \rceil / n$, which is the upper bound for all affine equivariant estimators (see Rousseeuw and Leroy 1987, p. 185).

To break down the b^k remedian we need at least $\lceil b/2 \rceil$ new values in the last round (which is a b-median). In each of the $\lceil b/2 \rceil$ corresponding cells at least $\lceil b/2 \rceil$ values must be new, and so on. In all, we need to replace at least $\lceil b/2 \rceil^k$ observations. It is also easy to see that this number of outliers is sufficient, provided they are put at the worst possible positions. Therefore, we have proven that the breakdown point of the remedian is

$$\varepsilon_n^* = \lceil b/2 \rceil^k / n = (\lceil b/2 \rceil / b)^k. \tag{4.1}$$

This also entails an *exact fit* result: If at least $n - \lceil b/2 \rceil^k + 1$ observations are identical, then the remedian will equal that value, no matter what the other observations may be.

Section 3 shows that the smallest storage is obtained for fixed b, but then the breakdown point will tend to 0 because $\lceil b/2 \rceil / b < 1$ and k goes to infinity with n. This is clearly visible in Table 1: For $b = 3$ (optimal storage) the breakdown point goes to 0 as $(\frac{2}{3})^k$, whereas for $b = 9$ it goes to 0 a little more slowly. The worst case is $b = 2$ (sample average) with breakdown point $(\frac{1}{2})^k = 1/n$, so even a single outlier may spoil the estimate.

On the other hand, if we keep k fixed then $\lceil b/2 \rceil / b \to \frac{1}{2}$ because now b tends to infinity, so the breakdown point tends to $(\frac{1}{2})^k$. In Table 1 we see that the breakdown point goes to 25% for the remedian with $k = 2$. This is the best possible value because $k \geq 2$ for any remedian. There is a trade-off between robustness and storage, as the higher breakdown points require more storage space. But even if we take the most robust version ($k = 2$) the storage is still merely $O(\sqrt{n})$, which is far less than that of the plain median ($k = 1$).

The remedian provides an interesting paradox. For fixed $b \geq 3$, its finite-sample breakdown point tends to 0% for increasing n. On the other hand, Section 5 shows that the remedian is a consistent estimator of the population median, which is a functional with a breakdown point of 50%! The finite-sample breakdown point does not converge to the asymptotic version because the finite-sample remedian is not permutation invariant, and the least-favorable outlier patterns are very peculiar. To make a remedian break down with just $\lceil b/2 \rceil^k$ outliers, the outliers must be placed in some particular subset of the indexes $\{1, \ldots, n\}$. Assuming that all sets of $\lceil b/2 \rceil^k$ indexes are equally likely, the probability of a breakdown subset goes to 0 very quickly when n increases.

When b is odd, the remedian coincides with one of the original observations. Unlike the plain median, this observation does not always have rank $\lceil n/2 \rceil$, but it may have one of several ranks. The breakdown argument shows that the smallest possible rank is exactly $\lceil b/2 \rceil^k$, whereas the largest possible rank is $n - \lceil b/2 \rceil^k + 1$. Moreover, if the observations are iid with respect to a continuous distribution function, all $n!$ orderings of the data are equally likely. Then we can even compute the *probability* that a particular rank will come out, by dividing the total number of orderings yielding that remedian by $n!$. For instance, for the 3^2 remedian the ranks 1, 2, 3, 7, 8, and 9 have probability 0, because $\lceil 3/2 \rceil^2 = 4$. We can also verify that there are 77,760 orderings for which the remedian becomes the fourth order statistic, so the probability of rank 4 becomes $77{,}760/9! = 3/14 \approx .2143$. This is also the probability of rank 6. For rank 5, we similarly find $207{,}360/9! = 4/7 \approx .5714$. Therefore, the remedian is restricted to the middle ranks, with the largest probability at the median itself.

5. CONSISTENCY

Let $x_1, \ldots, x_n$ be independent observations with common distribution F. (Note that we do *not* need any symmetry!) Assume that the number of observations is $n = b^k$, where b is fixed and $k = 1, 2, \ldots$. The base $b \geq 3$ is assumed odd.

The sample median of b observations is denoted by T_b. The remedian with base b and exponent k is denoted by T_{b^k}. This estimate can be expressed recursively in terms of the previous remedian estimates, which were based on b^{k-1} observations:

$$T_{b^k} = T_b(T_{b^{k-1}}, \ldots, T_{b^{k-1}}). \tag{5.1}$$

Theorem 1. Assume that F has a continuous density f that is strictly positive at $\mathrm{Med}(F)$, and let the base $b \geq 3$ be odd. Then the remedian T_{b^k} is consistent for $\mathrm{Med}(F)$ when $k \to \infty$.

Proof. Denote the distribution function of the sample median by $G_b(x)$. It is equal to the probability that at least $m + 1$ of the $b = 2m + 1$ observations are less than x, so

$$G_b(x) = \sum_{j=m+1}^{b} \binom{b}{j} (F(x))^j (1 - F(x))^{b-j}. \tag{5.2}$$

Therefore, we can write

$$G_b(x) = R_b(F(x)), \tag{5.3}$$

where $R_b(u) = \sum_{j=m+1}^{b} \binom{b}{j} u^j (1-u)^{b-j}$ is a monotone function from [0, 1] onto [0, 1] for which $R_b(0) = 0$, $R_b(\frac{1}{2}) = \frac{1}{2}$, and $R_b(1) = 1$. When differentiating $R_b(u)$ most terms vanish, yielding

$$R_b'(u) = \frac{b!}{m!\, m!} u^m (1-u)^m,$$

which is strictly increasing for $u \leq \frac{1}{2}$ and strictly decreasing for $u \geq \frac{1}{2}$ because $m \geq 1$.

Let us denote the distribution function of the remedian by $G_{b^k}(x)$. Since the remedian with $n = b^k$ is just the plain median of b previous remedians by Equation (5.1), it follows that

$$G_{b^k}(x) = R_b(G_{b^{k-1}}(x)). \tag{5.4}$$

The distribution $G_{b^{k-1}}$ can similarly be expressed in terms of $G_{b^{k-2}}$ and so on, all the way back to G_b, which depends

on F; hence

$$G_{b^k}(x) = R_b(R_b(\cdots R_b(F(x)))) = R_b^{(k)}(F(x)), \quad (5.5)$$

where $R_b^{(k)}$ denotes the k-fold composition of the function R_b.

Figure 4 illustrates how the recursion formula works. The horizontal axis initially contains a value of $F(x)$. The curve is the plot of the function R_b, and it shows how $F(x)$ is mapped to $G_b(x)$, which is read on the vertical axis. To find $G_{b^2}(x)$ one can put the $G_b(x)$ value on the horizontal axis and again read up to the curve. An easier way is to read from $G_b(x)$ over to the 45° line and then up to the curve, as indicated in the figure. Repeating this process generates all of the G_{b^k} values on the 45° line.

The consistency of the remedian follows directly from (5.5) and the fact that $R_b(u)$ is strictly convex for $u \leq \frac{1}{2}$ and strictly concave for $u \geq \frac{1}{2}$. In Figure 4 we see that as $k \to \infty$ the values of $G_{b^k}(x)$ gravitate toward 1 for any initial x value such that $F(x) > \frac{1}{2}$. On the other side, the values of $G_{b^k}(x)$ drift to 0 for any x such that $F(x) < \frac{1}{2}$. The probability that the remedian is outside a neighborhood of the population median therefore tends to 0; hence the remedian is consistent.

6. ASYMPTOTIC DISTRIBUTION

The following heuristic argument gives the remedian's convergence rate, which differs from $n^{1/2}$ and depends on the choice of the base b.

Assume that $\text{Med}(F) = 0$. Differentiating (5.4) with respect to x yields the density of the remedian at 0:

$$g_{b^k}(0) = R_b'(\tfrac{1}{2})g_{b^{k-1}}(0) = (R_b'(\tfrac{1}{2}))^k f(0) = \beta_b^k f(0),$$

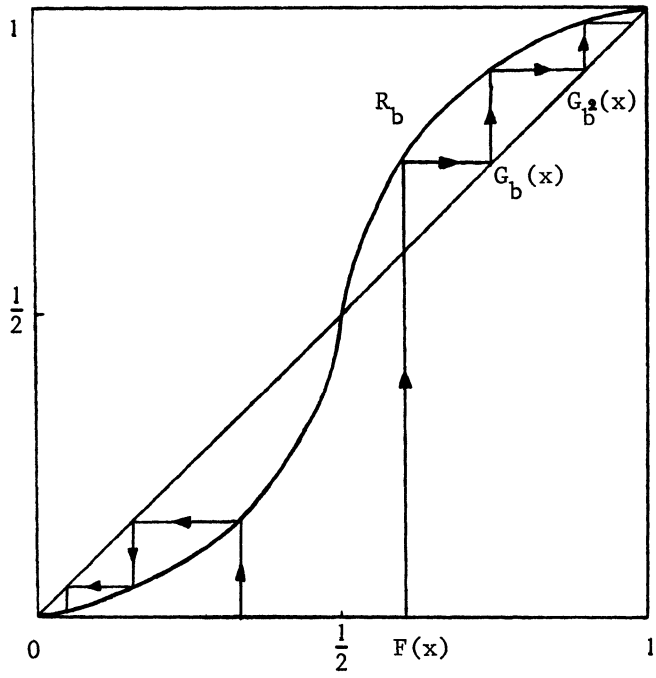

Figure 4. Plot of the Function R_b, Which Recursively Determines the Distribution of the Remedian.

in which the positive constant β_b is given by

$$\beta_b = R_b'\left(\frac{1}{2}\right) = \frac{b!}{m!\,m!}\left(\frac{1}{2}\right)^{2m}. \quad (6.1)$$

Therefore, $g_{b^k}(0)$ is increasing like β_b^k; hence to stabilize the density it needs to be stretched horizontally by the same scaling factor β_b^k. This gives the correct convergence rate, as will be confirmed by Theorem 2.

We can write the convergence rate in the usual form n^δ by putting $\beta_b^k = n^\delta$ and keeping in mind that $n = b^k$. Thus

$$\delta = \ln(\beta_b)/\ln(b). \quad (6.2)$$

For $b = 3$ this yields $\delta = .369$, whereas $b = 11$ gives $\delta = .415$. When b increases, the exponent δ approaches $\frac{1}{2}$; indeed, the remedian then approaches the sample median and its $n^{1/2}$ convergence rate.

Theorem 2. Under the conditions of Theorem 1, $\beta_b^k(T_{b^k} - \text{Med}(F))$ converges in law to an H_b distribution with location 0 and scale parameter $\sigma(F) = 1/\{f(\text{Med}(F))\sqrt{2\pi}\}$.

Proof. See the Appendix.

Theorem 2 shows that the limiting distribution only depends on the base b and on the density f at $\text{Med}(F)$. The general H_b distribution is denoted by $H_b(x; \mu, \sigma)$, where μ and σ are location and scale parameters. Here, μ is defined as the median of H_b. The parameter σ is most conveniently defined in terms of the height of the density at the median. For any distribution H with a density $h(\mu)$ at its median, define $\sigma(H)$ as $1/(h(\mu)\sqrt{2\pi})$. With this convention, the standard H_b distribution $H_b(x; 0, 1)$ and the standard Gaussian distribution $\Phi(x)$ have densities that are equal at 0.

For any distribution F with $\text{Med}(F) = 0$, Theorem 2 yields

$$H_b(x) = \lim_{k\to\infty} R_b^{(k)}(F(\beta_b^{-k}x)), \quad (6.3)$$

which can be used to evaluate $H_b(x)$ numerically. It turns out that $H_b(x; 0, 1)$ is very close to $\Phi(x)$, and already for $b = 3$ they agree to three decimal places.

The H_b distribution also satisfies the functional equation

$$H(x) = R_b(H(x/\beta_b)), \quad (6.4)$$

which defines the so-called *median-stable* laws. These laws were previously considered by Österreicher (1984). They arise when considering the following question: For what type of distribution will the median's sampling distribution be identical (except for scale) to the population distribution? The answer will be a law that under the remedian is its own limit; this follows from (5.5). By Theorem 2, H_b is also the remedian limit law when sampling from non-median-stable distributions. The same property holds for the usual stable laws with regard to the sample average.

It is easy to see that H_b is not a Gaussian distribution itself, by noting that Gaussian distributions do not satisfy (6.4).

7. EXTENSIONS AND RELATED APPROACHES

How should we proceed when the sample size n is less than b^k? The remedian algorithm then ends up with n_1 numbers in the first array, n_2 numbers in the second array, and n_k numbers in the last array, such that $n = n_1 + n_2 b + \cdots + n_k b^{k-1}$. For our final estimate we then compute a *weighted median* in which the n_1 numbers in the first array have weight 1, the n_2 numbers in the second array have weight b, and the n_k numbers in the last array have weight b^{k-1}. This final computation does not need much storage because there are fewer than bk numbers and they only have to be ranked in increasing order, after which their weights must be added until the sum is at least $n/2$.

To estimate the remedian's sampling distribution (and to obtain the associated confidence intervals) we can extrapolate the asymptotic results of Section 6. Alternatively, we could use the spread of the b values of the last step (or the b^2 values of the previous step) to compute a nonparametric confidence interval.

Instead of computing medians, one could also insert other estimators in the b^k scheme. In this way the same storage bk is needed. For an M estimator with maximal breakdown point, the recursive version also has breakdown point $([b/2]/b)^k$. Recursive estimators are usually different from their base generators, except for the sample average and the sample extremes.

Essentially three kinds of low-storage robust estimators have appeared in the literature. The first kind is based on stochastic approximation (Englund, Holst, and Ruppert 1988; Holst 1985; Martin and Masreliez 1975; Tierney 1983). Tukey (1978) and Weide (1978) computed medians of subsamples, followed by classical averaging. The third approach (Pearl 1981) is based on trees, in which minima and maxima are alternated at each level.

Like the remedian, none of these existing methods is permutation invariant. The main advantages of the remedian are its monotone equivariance and high breakdown point. On the other hand, most of its competitors have a lower variance, particularly those based on stochastic approximation. This is somewhat counterbalanced by the fact that the stochastic approximation methods need reliable starting values. A promising approach (proposed by an associate editor) would be a hybrid estimator that starts with the remedian and then switches to a stochastic approximation algorithm. If the switch were made after half the data were processed, then the hybrid estimator would have half the remedian's breakdown point and at most twice the asymptotic variance of the stochastic approximation estimator. Other choices of the change point are possible, involving a compromise between good breakdown properties and accuracy.

APPENDIX: PROOF OF THEOREM 2

In this proof we assume that $\mathrm{Med}(F) = 0$ without loss of generality.

Denote the distribution of $\beta_b^k T_{b^k}$ by $L_k(x) = R_b^{(k)}(F(\beta_b^{-k}x))$. As F is continuously differentiable near 0 we can expand it in a Taylor series yielding $F(\beta_b^{-k}x) = \frac{1}{2} + f(0)\beta_b^{-k}x(1 + o_k(1))$, so

$$L_k(x) = R_b^{(k)}(\tfrac{1}{2} + f(0)\beta_b^{-k}x) + o_k(1) \qquad (A.1)$$

because for all $z > 0$ and $\alpha > 0$ such that $\frac{1}{2} + z(1 + \alpha) < 1$ we have

$$R_b^{(k)}\left(\frac{1}{2} + z\right) \le R_b^{(k)}\left(\frac{1}{2} + z(1 + \alpha)\right) \le R_b^{(k)}\left(\frac{1}{2} + z\right) + \frac{\alpha}{2}.$$

The first inequality follows directly from R_b being increasing. For the second we note that

$$\frac{1}{2} + z = \frac{1}{1 + \alpha}\left(\frac{1}{2} + z(1 + \alpha)\right) + \frac{\alpha}{1 + \alpha}\left(\frac{1}{2}\right).$$

Combining concavity of R_b with $R_b(\frac{1}{2}) = \frac{1}{2}$ yields

$$R_b\left(\frac{1}{2} + z\right) \ge \frac{1}{1 + \alpha} R_b\left(\frac{1}{2} + z(1 + \alpha)\right) + \frac{\alpha}{1 + \alpha}\left(\frac{1}{2}\right);$$

hence $R_b(\frac{1}{2} + z(1 + \alpha)) \le \frac{1}{2} + (R_b(\frac{1}{2} + z) - \frac{1}{2})(1 + \alpha)$. Repeated application of this result gives

$$R_b^{(k)}(\tfrac{1}{2} + z(1 + \alpha)) \le \tfrac{1}{2} + (R_b^{(k)}(\tfrac{1}{2} + z) - \tfrac{1}{2})(1 + \alpha),$$

yielding the desired result because $R_b^{(k)}(\frac{1}{2} + z) < 1$ due to $\frac{1}{2} + z < 1$. Similar inequalities can be proved for $\alpha < 0$ as well as for $z < 0$. By putting $z = \beta_b^{-k} f(0)x$ and $\alpha = o_k(1)$ Equation (A.1) follows.

By (A.1) we may now restrict attention to the limit of

$$J_k(x) = R_b^{(k)}(\tfrac{1}{2} + \beta_b^{-k} f(0)x) \quad \text{as } k \to \infty.$$

For each x we know that $J_k(x)$ is in $[0, 1]$ for sufficiently large k. Indeed, for large k it holds that $|\beta_b^{-k} f(0)x| < \frac{1}{2}$, so $\frac{1}{2} + \beta_b^{-k} f(0)x$ belongs to $[0, 1]$. Applying $R_b^{(k)}$ then keeps $J_k(x)$ in $[0, 1]$.

We show that $J_k(x)$ converges for each x. Let $x > 0$ and write

$$\begin{aligned} J_{k+1}(x) &= R_b^{(k+1)}(\tfrac{1}{2} + \beta_b^{-k-1} f(0)x) \\ &= R_b^{(k)}(R_b(\tfrac{1}{2} + \beta_b^{-k-1} f(0)x)). \end{aligned}$$

Expanding R_b around $\frac{1}{2}$ gives

$$R_b(\tfrac{1}{2} + \beta_b^{-k-1} f(0)x) = \tfrac{1}{2} + \beta_b(1 - 4\theta^2)^m \beta_b^{-k-1} f(0)x,$$

where $0 < \theta < \beta_b^{-k-1} f(0)x$ and

$$\begin{aligned} R_b'\left(\frac{1}{2} + \theta\right) &= \frac{b!}{m!\,m!}\left(\frac{1}{2} + \theta\right)^m \left(1 - \left(\frac{1}{2} + \theta\right)\right)^m \\ &= 2^{2m}\beta_b \left(\frac{1}{2} + \theta\right)^m \left(\frac{1}{2} - \theta\right)^m \\ &= \beta_b(1 - 4\theta^2)^m. \end{aligned}$$

We restrict attention to large enough k for which $\beta_b^{-k-1} f(0)x < \frac{1}{2}$; hence $0 < 1 - 4\theta^2 < 1$. Then

$$\begin{aligned} J_{k+1}(x) &= R_b^{(k)}(\tfrac{1}{2} + (1 - 4\theta^2)^m \beta_b^{-k} f(0)x) \\ &\le R_b^{(k)}(\tfrac{1}{2} + \beta_b^{-k} f(0)x) = J_k(x). \end{aligned}$$

The sequence $J_k(x)$ is bounded and monotonically decreasing for large k; hence it converges to a value in $[0, 1]$.

The location μ of H_b is 0 because $H_b(0) = \lim_{k\to\infty} J_k(0) = \lim_{k\to\infty} R_b^{(k)}(\frac{1}{2}) = \frac{1}{2}$. For the scale parameter σ we compute $H_b'(0) = \lim_{k\to\infty} J_k'(0) = \lim_{k\to\infty}(R_b'(\frac{1}{2}))^k \beta_b^{-k} f(0) = f(0)$; hence H_b has the same density at 0 as the original population F.

[*Received August 1988. Revised April 1989.*]

REFERENCES

Englund, J., Holst, U., and Ruppert, D. (1988), "Recursive M-Estimators of Location and Scale for Dependent Sequences," *Scandinavian Journal of Statistics,* 15, 147–159.

Hillis, W. D. (1987), "The Connection Machine," *Scientific American,*

257, No. 6 (June), 86–93.
Holst, U. (1985), "Recursive Estimation of Quantiles," in *Contributions to Probability and Statistics in Honour of G. Blom,* eds. J. Lanke and G. Lindgren, Sweden: Lund University, pp. 179–188.
Knuth, D. E. (1973), *The Art of Computer Programming* (Vol. 3), Reading, MA: Addison-Wesley.
Martin, R. D., and Masreliez, C. J. (1975), "Robust Estimation via Stochastic Approximation," *IEEE Transactions on Information Theory,* 21, 263–271.
Österreicher, F. (1984), "On Median-Stable Laws," unpublished paper presented at the Oberwolfach meeting on robust statistics, Mathematisches Forschungsinstitut Oberwolfach (West Germany), September.
Pauwels, H. P., Vogeleer, M., Clement, P. A. R., Rousseeuw, P. J., and Kaufman, L. (1982), "Brainstem Electric Response Audiometry in Newborns," *International Journal of Pediatric Otorhinolaryngology,* 4, 317–323.
Pearl, J. (1981), "A Space-Efficient On-Line Method of Computing Quantile Estimates," *Journal of Algorithms,* 2, 164–177.
Rousseeuw, P. J. (1987), "An Application of L_1 to Astronomy," in *Statistical Data Analysis Based on the L_1 Norm and Related Methods,* ed. Y. Dodge, Amsterdam: North-Holland, pp. 437–445.
Rousseeuw, P. J., and Leroy, A. M. (1987), *Robust Regression and Outlier Detection,* New York: Wiley-Interscience.
Tierney, L. (1983), "A Space-Efficient Recursive Procedure for Estimating a Quantile of an Unknown Distribution," *SIAM Journal on Scientific and Statistical Computing,* 4, 706–711.
Trau, R., Salu, P., Wisnia, K., Kaufman, L., Rousseeuw, P. J., and Pierreux, A. (1983), "Simultaneous ERG–VER Recording," *Bulletin of the Belgian Ophthalmological Society,* 206, 61–67.
Tukey, J. W. (1978), "The Ninther: A Technique for Low-Effort Robust (Resistant) Location in Large Samples," in *Contributions to Survey Sampling and Applied Statistics in Honor of H. O. Hartley,* ed. H. A. David, New York: Academic Press, pp. 251–257.
Weide, B. (1978), "Space-Efficient On-Line Selection Algorithms," in *Proceedings of Computer Science and Statistics: Eleventh Annual Symposium on the Interface,* Raleigh: North Carolina State University, pp. 308–311.