LEAST ABSOLUTE DEVIATION ESTIMATES IN STABLE AUTOREGRESSION

by: S. Gross
W. Steiger

D.C.S. TR-#51

January 1977

Dr. S. Gross
Dept. of Mathematics
Montclair State College

Dr. William Steiger
Rutgers University
Dept.of Computer Science

Department of Computer Science
Rutgers, The State University of New Jersey
New Brunswick, New Jersey 08903

Technical Report #51

Classification: Probability/Statistics

Title: Least Absolute Deviation Estimates in Stable Autoregression

Authors:

S. Gross
Department of Mathematics
Montclair State College

W. L. Steiger
Department of Computer Science
Rutgers University

Key Words: Least Absolute Deviation, Autoregression, Infinite Variance

Running Title: Least Absolute Deviation Estimates

AMS (MOS) Classification: Primary 62J05, 60G10
Secondary 62P20, 93E10

Abstract:

We consider an L_1 analogue of the least squares estimator for the parameters of a stationary, finite order autoregressive scheme based on stable random variables. This estimator, the least absolute deviation (LAD), is shown to be strongly consistent via a result that may have independent interest. Finally, the sampling properties are compared to those of least squares. Together with a known convergence rate result for least squares, this provides evidence for a conjecture concerning the rate of convergence of LAD estimators.

1. Introduction and Summary

There has been a good deal of interest in stochastic processes based on random variables with infinite variance (e.g. [2] - [5], [7] - [13]. No doubt this is due to the inherent challenge ([9], e.g.) and theoretical interest provided by the non-normal stable laws as well as the possibility that processes constructed from these laws may be appropriate models for many, diverse phenomena ([10], [11], [12]).

Kanter and Steiger [7] studied estimation problems for linear regressions and for stationary, finite order autoregressions where the underlying random variables are in the domain of attraction of a symmetric stable law. This study was motivated by the fact that the components of a symmetric, stable vector $(X,Z) \in L_1$ have a linear regression

$$E(X|Z) = \lambda Z \quad \text{a.s.} \tag{1}$$

in which, if $E(X^2) = \infty$, the least squares estimator of λ usually fails to converge to λ. Other, consistent estimators of λ were discussed for this context and it was pointed out, via a conditional expectation analogue of the Yule-Walker equations, that the estimation problem for autoregressions may be solved using any consistent estimator of λ in the regression problem. Accordingly, the main result of [7] was somewhat surprising: the least squares estimator is consistent in finite order, stationary autoregressions with infinite variance.

The present paper considers least absolute deviation estimators (LAD), an L_1 analogue of least squares, and shows that the behavior mimics that of least squares. First, it is easy to see that in (1), the LAD estimator of λ, L_n, defined as a (not necessarily unique) location of the minimum of the distance function

$$f_n(a) = N^{-1} \sum_{i=1}^{N} | X_i - a Z_i |, \tag{2}$$

converges a.s. as $N \to \infty$ to a constant not necessarily equal to λ; here $\{(X_i, Z_i)\}$ is a stationary, ergodic sequence in which each $(X_i, Z_i) = (X, Z)$ in law. Thus, like least squares, LAD fails to be consistent in stable regression. We shall not consider that part of the analogy with least squares in this paper. Rather, we study autoregressions and show that, again like least squares, the LAD estimator of the parameters of stationary, finite order autoregressions based on symmetric stable random variables is consistent: this is accomplished in the next section. The final section presents some Monte-Carlo results upon which we base a conjecture concerning the convergence rate of LAD estimators.

It may be useful to close this introduction by recalling some basic facts about stable laws: further details may be found in Feller [6]. A random variable X is stable of index $\alpha \in (0,2]$ in case for each n there is a constanct c_n that makes

$$(3) \qquad L(X) = L((X_1 + \dots + X_n)/n^{1/\alpha} - c_n),$$

the X_i being i.i.d. variables with $L(X_i) = L(X)$ and $L(\cdot)$ referring to the law of the variable inside parentheses. A random variable variable Z is in the domain of attraction of a stable law of index $\alpha < 2$ if $t^\alpha P\{ |Z| > t \} \to k \neq 0$, a constant as $t \to \infty$, in which case a sequence $\{ c_i \}$ may be found that makes $L((Z_1 + \dots + Z_n)/n^{1/\alpha} - c_n) \to L(X)$, the Z_i being i.i.d. copies of Z. Finally if $X \neq 0$ is symmetric and stable of index α, $E(e^{itX}) = e^{-c|t|^\alpha}$ for some $c > 0$ and if X, Y, Z are i.i.d. symmetric stable variables, then for any a, b,

$$(4) \qquad L(aY + bZ) = L((|a|^\alpha + |b|^\alpha)^{1/\alpha} X)$$

2. Strong Consistency of LAD

Let $\{\dots, U_{-1}, U_0, U_1, \dots\}$ denote a sequence of i.i.d. random variables defined on a probability space (Ω, A, p). Fix an integer $k \geq 1$ and constants $a_1, \dots, a_k$ and consider the k^{th} order autoregression

$$(5) \qquad X_n = a_1 X_{n-1} + \ldots + a_k X_{n-k} + U_n$$

It was shown in [7] that if $a_1 z + \ldots, + a_k z^k \neq 1$ for any complex z, $|z| \leq 1$, and if the $\{ U_i \}$ are in the domain of attraction of a stable law of index $\alpha \in (0, 2]$ then (4) may be written as

$$(6) \qquad X_n = \sum_{j=0}^{\infty} b_j U_{n-j},$$

for appropriate constants b_j; the sum on the right converges almost surely. In other words, there is a stationary process $\{X_i\}$ that satisfies (5). For this stationary, k^{th} order autoregression, we study the problem of estimating $\underline{a} = (a_1, \ldots, a_k)$ based on N observations.

Consider the distance function

$$(7) \qquad f_N(\underline{c}) = N^{-1} \sum_{j=1}^{N} |R_j(\underline{c})|$$

where $R_j(\underline{c}) = X_{j+k} - (c_1 X_{j+k-1} + \ldots + c_k X_j)$ is the j^{th} residual determined by the vector $\underline{c}$. Clearly f_N is almost surely a continuous, convex function $R^k \to R$ because each term of (7) is. The least absolute deviation (LAD) estimator for $\underline{a}$ based on a sample of size N, $\underline{L}_N$, is the location of the minimum of f_N, unambiguously defined almost surely. However, to avoid unnecessary technicalities we define $\underline{L}_N$ as any vector that satisfies

(8) $$f_N(\underline{L}_N) \leq f_N(\underline{u}), \text{ all } \underline{u} \in R^k.$$

The salient property of $\underline{L}_N$ is that if the U_i are symmetric stable of index $\alpha \in (1, 2]$ then it converges to $\underline{a}$ almost surely as $N \to \infty$. This is shown by the following facts.

Let $C = \{\underline{c} \in R^k : 1 \neq c_1 z + \ldots + c_k z^k, |z| \leq 1\}$ the set of possible coefficients for stationary, k^{th} order autoregressions that may be written as (6). Clearly C is a bounded, open set. It is then easy to establish the following simple result.

Lemma 1: Let $\{\ldots, U_{-1}, U_0, U_1, \ldots\}$ be an i.i.d sequence, stable of index $\alpha \in (1,2]$ and $\{X_i\}$ a stationary process satisfying (5). Then for each $\underline{c} \in C$, f_N almost surely converges to the limit function

(9) $$f(\underline{c}) = E\,|X_{k+1} - (c_1 X_k + \ldots + c_k X_1)|$$

Proof: Fix $\underline{c} \in C$. There exists an i.i.d. sequence $\{\ldots, V_{-1}, V_0, V_1, \ldots\}$ of symmetric stable random variables of index α and a stationary process $\{Y_i\}$ that satisfies

(10) $$Y_n = c_1 Y_{n-1} + \ldots + c_k Y_{n-k} + V_n$$

Each term $R_j(\underline{c})$ in (7) has the same law as V_0 and $\{R_j(\underline{c})\}$ is stationary and ergodic because $\{X_i\}$ is (see Doob [1]). By the ergodic theorem $f_N(\underline{c}) \to E \mid R_0(\underline{c}) \mid = f(\underline{c})$ a.s. ▨

It is more complicated to establish the next fact, interesting in its own right.

Lemma 2: Let $\{\ldots, U_{-1}, U_0, U_1, \ldots\}$ be an i.i.d. sequence, symmetric and stable of index $\alpha \in (1, 2]$ and $\{X_i\}$ a stationary process satisfying (5). Then the limit function f is convex on C and has a unique minimum attained at $\underline{a}$; i.e.,

$$(11) \qquad f(\underline{a}) < f(\underline{c}),\ \underline{c} \neq \underline{a}$$

Proof: Because $| X_{k+1} - (c_1X_k + \ldots + c_kX_1)|$ is a convex function of $\underline{c}$ on the set C, $E(| X_{k+1} - (c_1X_k + \ldots + c_kX_1)| = f(\underline{c})$ is convex.

To prove (11) write $\underline{c} = \underline{a} + \underline{d}$ and note that $f(\underline{c}) = f(\underline{a} + \underline{d}) =$ $E(| X_{k+1} - (a_1X_k + \ldots + a_kX_1) - (d_1X_k + \ldots + d_kX_1)|) =$ $E(|U_{k+1} - (d_1X_k + \ldots + d_kX_1)|$, the last equality a consequence of (5). Using (6) for each of the X_i, $i = 1, \ldots, k$, we obtain

$$(12) \qquad f(\underline{c}) = E \mid U_{k+1} - \sum_{j=0}^{\infty} e_j U_{k-j} \mid$$

where $e_j = d_1 b_j + \dots + d_k b_{j-k}$, and $b_\ell = 0$ for $\ell < 0$. Since the $b_j \to 0$ geometrically fast as is well-known, so do the e_j, so $\sum_{j=0}^{N} e_j U_{k-j}$ converges a.s. as $N \to \infty$ (see [7 ; p. 781]). Its limit is a symmetric random variable, Z, independent of U_{k+1}, whose law satisfies

$L(Z) = L((\sum_{j=0}^{\infty} |e_j|^\alpha)^{1/\alpha} U_0)$ by an extension of (4), so that

$$f(\underline{c}) = E \mid U_{k+1} - (\sum_{j=0}^{\infty} |e_j|^\alpha)^{1/\alpha} U_o \mid \tag{13}$$

Since $E \mid A + t B \mid$ has a unique minimum when $t = 0$ if A and B are non-trivial, independent, symmetric random variables in L_1, f attains its unique minimum when $\sum_{j=0}^{\infty} |e_j|^\alpha = 0$. This occurs only when $d_1 = \dots = d_k = 0$ or at $\underline{c} = \underline{a}$. ▨

The last fact is of independent interest and seems to be new.

Lemma 3: Let $\{h_i\}$ be an equicontinuous sequence of convex functions $R^k \to R$ that converge pointwise to a continuous convex function h on an open set $U \subset R^k$. Suppose h_n has a minimum at $\underline{x}_n$ and h, a unique minimum at $\underline{x} \in U$; i.e.,

$$\begin{aligned} &h_n(\underline{x}_n) \le h_n(\underline{u}), \text{ all } \underline{u} \\ &h(\underline{x}) < h(\underline{u}), \quad \underline{u} \ne \underline{x} \end{aligned} \tag{14}$$

Then

(15) $\underline{x}_n \to \underline{x}$ as $n \to \infty$.

Proof: If $\underline{x}_n \not\to \underline{x}$ there is $\varepsilon > 0$ and a subsequence n_j for which $||\underline{x}_{n_j} - \underline{x}|| \geq \varepsilon$. Choose $t \leq \varepsilon$ so that $h_n(\underline{u}) \to h(\underline{u})$ whenever $||\underline{u}-\underline{x}|| \leq t$. Write $S(t)$ for the surface of the sphere $\{\underline{u}: ||\underline{u}-\underline{x}|| = t\}$ and note that h_{n_j} attains a minimum (not necessarily uniquely) at, say, $\underline{u}_{n_j} \in S_t$. There is a limit point, $\underline{u}_o$ and a subsequence m_j of n_j such that $\underline{u}_{m_j} \to \underline{u}_o$. Equicontinuity implies $h_{m_j}(\underline{u}_{m_j}) \to h(\underline{u}_o)$. Pointwise convergence combines to imply that for large enough m_j, $h_{m_j}(\underline{u}_{m_j}) > h(\underline{u}_o) - \sigma/4$ and $h_{m_j}(\underline{x}) < h(\underline{x}) + \sigma/4$ where $\sigma = h(\underline{u}_o) - h(\underline{x}) > 0$; therefore $h_{m_j}(\underline{u}_{m_j}) - h_{m_j}(\underline{x}) > \sigma/2$. This shows that for m_j large enough, each h_{m_j} attains a minimum in $B(t) = \{\underline{u}: ||\underline{u}-\underline{x}|| < t\}$ which, together with convexity, contradicts $\underline{x}_n \not\to \underline{x}$. ▨

Theorem 1: Let $\{\ldots, U_{-1}, U_o, U_1, \ldots\}$ be an i.i.d. sequence of symmetric stable random variables of index $\alpha \in (1,2]$ and $\{X_i\}$ a stationary sequence that satisfies (5) for some $\underline{a} \in C \subset R^k$. The least absolute deviation estimators $\underline{L}_N$, satisfy

(16) $\underline{L}_N \to \underline{a}$ a.s. as $N \to \infty$

(19) $$g_N(\underline{c}) = N^{-1} \sum_{j=1}^{N} |X_{k+j} - (c_o + c_1 X_{k+j-1} + \ldots + c_k X_j)|$$

and

(20) $$g(\underline{c}) = E|X_n - (c_o + c_1 X_{n-1} + \ldots + c_k X_{n-k})|$$

$g_N \to g$ a.s. on $R \times C$. Furthermore Lemma 2 adapts to show that f has a unique minimum at $(a_o, a_1, \ldots, a_k) = (a_o, \underline{a})$ from which, using Lemma 3, $\underline{L}_N \to (a_o, \underline{a})$ a.s. This gives the parameters of (5).

3. Further Remarks

In this section we present Monte-Carlo results that compare sampling distributions of $\underline{L}_n$ with those of the least squares estimators. These lend support to conjectures about the rate of convergence of $\underline{L}_n$ to $\underline{a}$ and possible generalizations of Theorem 1.

In the experiments we generated stable random numbers using the fact that for $\alpha < 1$ the random variable $S_\alpha = [f(U)/V]^{1/\alpha-1}$ is positive stable of index α, where U and V are independent with $P\{U \le t\} = t$, $0 \le t \le 1$, and $P\{V \le t\} = 1-e^{-t}$, $0 \le t < \infty$, and $f(x) = (\sin \pi\alpha x/\sin \pi x)^{1/(1-\alpha)}(\sin \pi x(1-\alpha)/\sin \pi\alpha x)$, $0 \le x \le 1$ (see [8]). Moreover if X is symmetric stable of index $\alpha \in (0,2]$ and Y is positive stable of index $\beta < 1$ and independent of X, then

$$R = X(Y)^{1/\alpha}$$

is symmetric stable of index $\alpha\beta$ (see [6 ; p172]) from which it follows that

$$(21) \qquad N(S_{\alpha/2})^{1/2}$$

is symmetric stable of index α, where N is normal, $E(N)=0$, $E(N^2)=2$, and $S_{\alpha/2}$ positive stable of index $\alpha/2$ and independent of N; in fact, the variable in (21) is "standard", with characteristic function $e^{-|t|^\alpha}$.

We simulated the stationary auto regression

$$(22) \qquad X_n = a_1 X_{n-1} + a_2 X_{n-2} + n\ldots + a_k X_{n-k} + U_n$$

with U_i symmetric stable of index α, generated as in (21). This was done via the moving average representation of the process,

$$X_n = \sum_{j=1}^{\infty} b_j U_{n-j}$$

and approximating the solution to (22) by the (non stationary) process

(23) $$X_n = \sum_{j=0}^{200} b_j U_{n-j} .$$

For $N > 0$, a sample of size N is $X_1, \ldots, X_N$, computed from $U_{-199}, \ldots, U_N$ according to (23).

Twenty samples of size N were generated and for each, a value of $\underline{L}_N$ was computed using (8). For comparison the least squares estimator $\underline{C}_N$ was also computed as

$$\underline{C}_N = R_N^{-1} \underline{\rho}_N$$

where $r_N(i) = \left(\sum_{n=1}^{N-k} X_n X_{i+n}\right) / \sum_{n=1}^{N-k} X_n^2$, $\rho_N^T = (r_N(1), \ldots, r_N(k))$ and

$R_N = (r_N(|i-j|))$, $1 \leq i, j \leq k$. The following tables depict the sampling distributions of $\underline{L}_N$ and $\underline{C}_N$ for $k = 2$, $a_1 = 1.4$, $a_2 = -.7$, $N = 32, 102, 502$, and $\alpha = 1.6, .6$.

The entries are the residuals $\underline{L}_N - \underline{a}$ and $\underline{C}_N - \underline{a}$ for each estimator. MAX and MIN refer respectively to the largest and smallest of the 20 residuals for each estimator and Q refers to quartiles. Thus, $-.0429 \leq C_{32}(1) - 1.4 \leq .0966$ was satisfied for 10 of the 20 least squares estimators of $a_1 = 1.4$ based on samples of size 32, from Table 1.

References

[0] Bloomfield, P. and Steiger, W. L. (1977) Least absolute deviation curve-fitting. (in preparation).

[1] Doob, J. L. (1953) Stochastic Processes, John Wiley, New York.

[2] du Mouchel, W. H. (1973) Stable distributions in statistical inference: 1. Symmetric stable distributions compared to other symmetric long-tailed distributions. J. Amer. Statist. Assoc. 68, 469-477.

[3] du Mouchel, W. H. (1973) On the asymptotic normality of the maximum likelihood estimates when sampling from a stable distribution. Ann. Statistics 1, 948-957.

[4] Fama, E. and Roll, R. (1968) Some properties of symmetric stable distributions. J. Amer. Statist. Assoc. 63, 817-836.

[5] Fama, E. and Roll, R. (1971) Parameter estimation for symmetric stable distributions. J. Amer. Statist. Assoc. 66, 331-338.

[6] Feller, W. (1966) An Introduction to Probability Theory and its Applications, Vol. 2, John Wiley, New York.

[7] Kanter, Marek and Steiger, W. L. (1974) Regression and autoregression with infinite variance, Adv. Appl. Probability 6, 768-783.

[8] Kanter, Marek (1975) Stable densities under change of scale and total variation inequalities, Ann. Probability 3, 697-707.

[9] Kanter, Marek (1976) On the unimodality of stable densities, Ann. Probability 4, 1006-1008.

[10] Mandelbrot, Benoit (1963) The variation of certain speculative prices, J. Business U. Chicago 26, 294-419.

[11] Mandelbrot, Benoit (1967) The variation of some other speculative prices, J. Business U. Chicago 40, 393-413.

[12] Mandelbrot, Benoit (1974) A population birth-and-mutation process I: explicit distributions for the number of mutants in an old culture of bacteria. J. Appl. Probability 11, 437-444.

[13] Mandelbrot, Benoit and Taylor, H. (1967) On the distribution of stock price differences, Operations Research 15, 1057-1062.