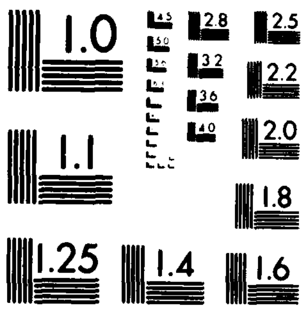
AD-A088 246 MARYLAND UNIV COLLEGE PARK DEPT OF MATHEMATICS F/G 12/1
SOME GRAPHICAL CONSIDERATIONS IN TIME SERIES ANALYSIS.(U)
JUN 80 B KEDEM AFOSR-80-0211
UNCLASSIFIED MD80-43-BK AFOSR-TR-80-0590 NL

END
DATE FILMED
9-80
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

LEVEL II

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR- 80-0590	2. GOVT ACCESSION NO.: AD-A088 246
	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): Some Graphical Considerations in Time Series Analysis	5. TYPE OF REPORT & PERIOD COVERED: Interim
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Benjamin Kedem	8. CONTRACT OR GRANT NUMBER(s): AFOSR-80-0211
9. PERFORMING ORGANIZATION NAME AND ADDRESS: University of Maryland, College Park, MD 20742	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 2304/A5, 61102F
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research, Bolling AFB, Washington, DC 20332	12. REPORT DATE: June 1980
	13. NUMBER OF PAGES: 23
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): MD 80-43-BK, TR 80-41	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

DTIC ELECTE AUG 26 1980 S D E

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Random Curves, Features, Crossings, Graphical

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The pictorial information in a stationary time series as depicted by crossings of levels and crossings of random levels and related quantities is studied. It is shown that such graphical features are directly connected with the covariance function and hence with the spectral density. Many of these features can be actually applied in estimation and in the study of extremes. In the Gaussian case, the finite dimensional distributions are completely determined by the axis crossings and by the crossings of a random curve (to be defined) if the process is essentially bounded.

DD FORM 1 JAN 73 1473

UNCLASSIFIED 400953

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD A088246

DDC FILE COPY

AFOSR-TR- 80-0590

SOME GRAPHICAL CONSIDERATIONS IN TIME SERIES ANALYSIS

Benjamin Kedem
Department of Mathematics
University of Maryland
College Park, Maryland
20742

TECHNICAL REPORT

TR80-41
MD80-43-BK

JUNE 1980

Research supported by AFOSR Grant-80-0211.

SOME GRAPHICAL CONSIDERATIONS IN TIME SERIES ANALYSIS

by Benjamin Kedem

ABSTRACT

The pictorial information in a stationary time series as depicted by crossings of levels and crossings of random levels and related quantities is studied. It is shown that such graphical features are directly connected with the covariance function and hence with the spectral density. Many of these features can be actually applied in estimation and in the study of extremes. In the Gaussian case, the finite dimensional distributions are completely determined by the axis crossings and by the crossings of a random curve (to be defined) if the process is essentially bounded.

SOME GRAPHICAL CONSIDERATIONS IN TIME SERIES ANALYSIS

by Benjamin Kedem

1. INTRODUCTION

From the review articles by Beniger and Robyn (1978) and Fienberg (1979) we learn that graphical methods for depiction of empirical data is an old and useful idea which may be traced hundreds of years back. However, no theory of statistical graphics exists although graphical depiction of data is a well accepted practice. Concerning this last remark, this is even so when the data at hand are graphical by their very nature, and the case we have in mind is time series analysis. True, theory as such does not exist but still we may advance graphical methods of analysis. The point to be made here is that graphical methods of analysis are almost always available given a long time series.

In this paper we shall demonstrate the usefulness of some graphical aspects in time series analysis by answering questions such as:

1. What does it mean that a time series appears to be bounded?
2. How can we interpret the number of crossings of fixed and random levels?
3. What can we learn from the succession of times spent by a series above and below a random level?
4. Are there useful graphical features to be reckoned with in time series data?
5. For very high levels, what should we observe in a time series order to determine the distribution of extremes?
6. Are the axis crossings by the successive differences of a time series useful and what happens if we difference indefinitely?

In fact we will show that the spectral density of a stationary process is a function of some graphical features which can be detected by a quick eye examination. This means that at least in principle spectral analysis amounts to a careful eye examination of time series graphs. This is independent of the Gaussian assumption. For a zero-mean stationary Gaussian process we can even say much more: all of the finite dimensional distributions can, at least in principle, be determined from graphical features, under some conditions often met in practice.

Our main concern in this paper is to express well-known quantities in terms of graphical features thus emphasizing the pictorial information contained in the plot of a stationary time series. For this purpose we shall sometimes create features by introducing useful random curves and by clipping the series at various levels. In what follows, axis crossings and crossings of "random curves" and other similar visual features play a dominant role.

2. RANDOM CLIPPING

For a 0-mean Gaussian process it is well-known that its correlation structure is a function of the "zeros" of its sample paths. Much the same can be achieved, at least in theory, for any bounded series.

Let $\{Z_t, t=0,\pm 1,\dots\}$ be a zero mean strictly stationary process. This is a strong assumption which is made here for simplicity. Assume that there exists an A such that

$$|Z_t| < A. \tag{2.1}$$

Then the process is also weakly stationary of order k for any k. In the range $(-A,A)$ we define a uniform process $\{U_t, t=0,\pm 1,\dots\}$ independent of Z_t and made of independent and identically distributed random variables, such that

$$f_{U_t}(u) = \begin{cases} \frac{1}{2A}, & -A<u<A \\ 0, & \text{otherwise.} \end{cases} \tag{2.2}$$

Thus, corresponding to a time series $Z_1,\dots,Z_N$ there is a "random curve" $U_1,\dots,U_N$ which helps us to define a clipped binary series $\{Y_t\}$ by

$$Y_t = \begin{cases} 1, & Z_t \ge U_t \\ 0, & Z_t < U_t \end{cases}, \quad t=1,\dots,N. \tag{2.3}$$

The time series $U_1,\dots,U_N$ will be referred to as the U-curve and throughout the paper it is assumed that N is large. Observe that $EY_t = \frac{1}{2}$ and for $t_i \neq t_j, i,j \in (1,\ldots,N)$

$$EY_{t_1} \ldots Y_{t_r} = P_r(Z_{t_1} \geq U_{t_1}, \ldots, Z_{t_r} \geq U_{t_r}) \tag{2.4}$$

$$= \frac{1}{(2A)^r} E(Z_{t_1}+A)\ldots(Z_{t_r}+A)$$

so that all the cross moments of $\{Z_t\}$ can be obtained from the moments of $\{Y_t\}$. In particular from (2.4) we obtain an important relation between the covariance functions

$$\gamma_Z(k) = 4A^2\gamma_Y(k) \ , \ k \neq 0, \ k=1, \ 2, \ \ldots \tag{2.5}$$

and so the covariances in the $\{Y_t\}$ process are reduced, which for binary data amounts to a lesser dependence.

Now consider the quantity

$$C_1 = 2\sum_1^N Y_t - 2\sum_2^N Y_t Y_{t-1} - (Y_1+Y_N). \tag{2.6}$$

This is readily seen to be the number of symbol changes in the binary series. But then it is also the number of "curve crossings" where the curve is the U-curve defined as the random series $U_1,\ldots,U_N$. (2.5) and (2.6) lead to

$$\gamma_Z(1) = A^2\left(1 - \frac{2E(C_1)}{N-1}\right). \tag{2.7}$$

We see that $\gamma_Z(1)$ is a linear function of the expected number of U-curve crossings by $Z_1,\ldots,Z_N$. An unbiased estimate of $\gamma_Z(1)$ is

$$\hat{\gamma}_Z(1) = A^2\left(1 - \frac{2C_1}{N-1}\right). \tag{2.8}$$

We can get hold of an approximation to the variance of $\hat{\gamma}_Z(1)$ as follows. Let $B > A$ and assume U_t is uniform in $(-B,B)$ while (2.1) still holds. Then in (2.4) B replaces A and as B increases

$$\Pr(Y_{t_1}=1,\ldots,Y_{t_r}=1) \to (\tfrac{1}{2})^r$$

in which case (Kedem (1980a)) we easily see that $C_1 \to b(N-1,\frac{1}{2})$ and

$$\mathrm{Var}(\hat{\gamma}_Z(1)) \simeq \frac{B^2}{N-1}\ . \tag{2.9}$$

The lesson to be learned here is two fold. First, we have made use of a conspicuous graphical feature and second, by employing an artificially large upper bound B on $|Z_t|$ in the definition of U_t, the clipped data are nearly independent. This is a desirable property of random clipping.

Next define

$$C_2 = 2\sum_1^N Y_t - 2\sum_3^N Y_tY_{t-2} - 2. \tag{2.10}$$

Then

$$\gamma_Z(2) = A^2\left(1 - \frac{2E(C_2)}{N-2}\right). \tag{2.11}$$

But C_2, apart from a negligible end effect, is twice the number of successive sojourns above and below the U-curve of at least two time periods. That is, twice the number of runs in the clipped binary series between the first and last 1's with at least two symbols plus 1. We can continue in this fashion. $\sum_1^N Y_t - \sum_r^N Y_t Y_{t-3}$ also defines the number of certain sojourns but it is simpler to switch now the eye to the binary series between the first and last 1's. Then

$$\sum_1^N Y_t - \sum_4^N Y_t Y_{t-3} = \text{\# of runs with at least 3 symbols} + \text{\# of } 0011 + \text{\# of } 0101 + 2.$$

For example in the binary series

00010100010101000011110101001111001

$$\Sigma Y_t - \Sigma Y_t Y_{t-3} = 16 - 5 = 11.$$

But the number of runs with 3 or more symbols (concentrating on the series between the first and last 1's) is 4, the number of 0011 is 2 and the number of 0101 is 3. The sum of these numbers plus 2 is 11 as it should. We can now define C_3 in terms of these features and then express $\gamma_Z(3)$ in terms of $E(C_3)$. In general, we define

$$C_k = 2\sum_1^N Y_t - 2\sum_{k+1}^N Y_t Y_{t-k} - k \ , \ k=2,3,\ldots \tag{2.12}$$

where C_k is obtained by counting the number of times $(Z_t \geq U_t)$ and the number of times $(Z_t \geq U_t, Z_{t-k} \geq U_{t-k})$. Then

$$\gamma_Z(k) = A^2\left(1 - \frac{2E(C_k)}{N-k}\right) . \qquad (2.13)$$

And by replacing $E(C_k)$ by C_k in (2.13) we obtain an unbiased estimate. If C_0 is defined so that

$$E(C_0) = \frac{N}{2}\left(1 - \frac{1}{A^2}\gamma_Z(0)\right),$$

(2.13) holds for $k=0,\pm 1,\ldots,$ and assuming that a spectral density $f(\lambda)$ exists we have for sufficiently large M and N

$$f(\lambda) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} e^{ik\lambda}A^2\left(1 - \frac{2E(C_k)}{N-k}\right)$$

$$\simeq A^2\left(\frac{\sin(M+\frac{1}{2})\lambda}{2\pi\sin\frac{1}{2}\lambda} - \frac{1}{\pi N}\sum_{k=-M}^{M} e^{-ik\lambda}E(C_k)\right). \qquad (2.14)$$

Whence for bounded stationary processes the spectral density is essentially the Fourier transform of the expected numbers of visual features such as U-curve crossings and lengths of sojourns above and below the U-curve.

3. THE GAUSSIAN CASE AND LOSS OF INFORMATION

How much information is lost due to clipping? This is a disturbing problem associated with hard limiting operations. In the Gaussian case however, at least in principle, the answer is that no information is lost if clipping and random clipping are combined. That is, our graphical features contain a great deal of information.

To show this choose an A so large relative to $\mathrm{Var}(Z_t)$ that (2.1) holds for all practical purposes. Next define

$$X_t = \begin{cases} 1, & Z_t \geq 0 \\ 0, & Z_t < 0 \end{cases}, \quad t=0,\pm 1,\ldots, \tag{3.15}$$

where $\{Z_t\}$ is a zero-mean stationary Gaussian process. Then it is well-known that

$$\gamma_Z(k) = \gamma_Z(0)\sin(2\pi\gamma_X(k)) \tag{3.16}$$

and by invoking (2.5)

$$\gamma_Z(0) = \frac{4A^2\gamma_Y(k)}{\sin(2\pi\gamma_X(k))}, \quad k=1,2,\ldots \tag{3.17}$$

If D_1 counts the number of axis-crossings by $Z_1,\ldots,Z_N$, then for $k=1$ we obtain from (2.7) and (2.17) after representing D_1 as in (2.6), but with Y's replaced by X's,

$$\gamma_Z(0) = \frac{A^2\left[1 - \frac{2E(C_1)}{N-1}\right]}{\mathrm{Cos}\,\frac{\pi E(D_1)}{N-1}} . \tag{3.18}$$

This means that the variance of a Gaussian process can be obtained by observing the U-curve-and axis-crossings by a time series from the process. Therefore from (2.13) and (3.18), for a bounded, practically speaking, zero-mean stationary Gaussian process, the finite dimensional distributions are completely determined graphically. To make this statement more precise we should attach a probabilistic statement to the bounds, require that long records be available and that the covariance function decays fast enough, in which case (3.18) and (2.13) provide consistent estimates when the expectations are replaced by observed values. This finding helps to explain the remarkable fact that a great deal of inference about a stationary Gaussian process can be made from clipped data (Kedem (1980a)).

4. HIGHER ORDER CROSSINGS

Can we get a formula equivalent to (3.18) in the general case for any bounded stationary process? The answer to this question is in the affirmative if we are willing to consider the $U^{(1)}$-curve, say, crossings by $\nabla Z_t = Z_t - Z_{t-1}$. Assume the same properties for $\{Z_t\}$ as in the previous section so that

$$|\nabla Z_t| \le 2A. \tag{4.1}$$

Let $\{U_t^{(1)}\}$ be the uniform independent process corresponding to $\{\nabla Z_t\}$ so that $U_t^{(1)}$ is uniform in $(-2A, 2A)$. Then from (2.7)

$$\gamma_{\nabla Z}(1) = 4A^2 \left[1 - \frac{2E(C^{(1)})}{N-1}\right] \tag{4.2}$$

where $C^{(1)}$ is the number of crossings by $\nabla Z_1, \dots, \nabla Z_N$ of the U-curve $U_1^{(1)}, \dots, U_N^{(1)}$. But

$$\gamma_{\nabla Z}(1) = 2\gamma_Z(1) - \gamma_Z(0) - \gamma_Z(2) \tag{4.3}$$

so that

$$\gamma_Z(0) = A^2 \left\{ 2\left[1 - \frac{2E(C_1)}{N-1}\right] - \left[1 - \frac{2E(C_2)}{N-2}\right] - 4\left[1 - \frac{2E(C^{(1)})}{N-1}\right]\right\}. \tag{4.4}$$

Thus the variance of a bounded stationary process is a linear function of the number of crossings by $Z_1, \dots, Z_N$ of $U_1, \dots, U_N$, the number of sojourns of at least two time periods above and below $U_1, \dots, U_N$, and the number of crossings by $\nabla Z_1, \dots, \nabla Z_N$ of $U_1^{(1)}, \dots, U_N^{(1)}$.

We have seen that the $U^{(1)}$-curve crossings by ∇Z_t are useful and a natural question arises as to how useful the $U^{(k)}$-curve, say, crossings by $\nabla^k Z_t$ are. To answer this question note that

$$|\nabla^k Z_t| \le 2^k A \ , \ k=0,1,2,... \tag{4.5}$$

and let $\{U_t^{(k)}\}$ be the corresponding independent process uniform in $(-2^k A,\ 2^k A)$ which consists of independent and identically distributed random variables. We call the time series $U_1^{(k)},...,U_N^{(k)}$ the $U^{(k)}$-curve which we introduce as above in order to create graphical features. Let $C^{(k)}$ be the number of crossings of the $U^{(k)}$-curve by $\nabla^k Z_1,...,\nabla^k Z_N$. Then from (2.7)

$$\gamma_{\nabla^k Z}(1) = 2^{2k}A^2\left[1 - \frac{2E(C^{(k)})}{N-1}\right],\ k=0,1,2,..., \tag{4.6}$$

where $C^{(0)}$ is taken as C_1.

Observe that

$$\gamma_{\nabla^k Z}(1) = -\binom{2k}{k-1}\gamma_Z(0) + \left[\binom{2k}{k} + \binom{2k}{k-2}\right]\gamma_Z(1) - + \ldots + (-1)^k\gamma_Z(k+1) \tag{4.7}$$

so that $\gamma_Z(k+1)$ can from (4.6), (4.7) be expressed as a function of $E(C^{(k)})$, $k=0,1,...,k$, and $E(C_2)$. We see that the $C^{(k)}$, the number of $U^{(k)}$-curve crossings by $\nabla^k Z_t$, $t=1,...,N$, together with C_2 completely determine the covariance function of a stationary bounded process. We call the $C^{(k)}$ the higher order $U^{(k)}$-curve crossings.

In analogy with the higher order $U^{(k)}$-curve crossings we can define the higher order axis crossings by $\nabla^k Z_t$. More precisely let $\{Z_t\}$ be a stationary process, not necessarily bounded, and define a binary process $\{X_t^{(k)}\}$ by

$$X_t^{(k)} = \begin{cases} 1, & \nabla^{k-1}Z_t \geq 0 \\ 0, & \nabla^{k-1} < 0 \end{cases}, \quad k=1,2,\ldots \tag{4.8}$$

Given a time series $Z_1,\ldots,Z_N$ we define

$$D_k = 2\sum_1^N X_t^{(k)} - 2\sum X_t^{(k)} X_{t-1}^{(k)} - (X_1^{(k)} + X_N^{(k)}). \tag{4.9}$$

This is the number of axis crossings by $\nabla^k Z_1,\ldots,\nabla^k Z_N$. But then D_1 is the number of axis crossings by the original series, and apart from end effects, D_2 is the number of local maxima and minima, D_3 is the number of inflection points, etc. Thus the first few D_k's correspond to conspicuous features in a time series. A natural question to ask is whether features which are not that conspicuous or not at all for that matter are still useful. Before considering this question we note that results similar to (4.6) can be obtained for the D_k's in the Gaussian case without reference to boundedness.

Now going back to our question, we note that $\{Z_t\}$ admits a spectral representation with respect to a process of orthogonal increments $\{\xi(\lambda), -\pi < \lambda \leq \pi\}$

$$Z_t = \int_{(-\pi,\pi]} e^{it\lambda} d\xi(\lambda). \tag{4.10}$$

It is convenient to give an answer to our question via (4.10). Observe that

$$(1-e^{-i\lambda}) = \frac{e^{-i\lambda/2}(e^{i\lambda/2}-e^{-i\lambda/2})2i}{2i}$$

$$= e^{-i(\frac{\lambda}{2} - \frac{\pi}{2})}(2(1-\cos\lambda))^{\frac{1}{2}}. \qquad (4.11)$$

Therefore by linearity

$$\nabla^k Z_t = \int_{(-\pi,\pi]} e^{it\lambda} e^{-ik(\frac{\lambda}{2} - \frac{\pi}{2})}(2(1-\cos\lambda))^{k/2} d\xi(\lambda) \qquad (4.12)$$

and we have

$$\frac{\nabla^k Z_t}{2^k} \xrightarrow{q.m.} \cos(t\pi)d\xi(\pi) \ , \ k \to \infty. \qquad (4.13)$$

This implies

$$E\frac{\nabla^k Z_t}{2^k}\,\frac{\nabla^k Z_{t-j}}{2^k} \to (-1)^j E|d\xi(\pi)|^2 = (-1)^j dF(\pi), \ k \to \infty$$

where F is the spectral distribution function. We shall assume $dF(\pi) > 0$. Then

$$\text{Corr}(\nabla^k Z_t, \nabla^k Z_{t-j}) \to (-1)^j \ , \ k \to \infty. \qquad (4.14)$$

This means that on each finite time (discrete) interval $X_t^{(k)}$ for large k tends to consist of binary strings in which a 0 s followed by a 1 and vice versa. In fact it was recently shown in Kedem and Slud (1980) that we actually have weak convergence of $\{X_t^{(k)}\}$

$$\{X_t^{(k)}\} \Rightarrow \{\ldots 010101 \ldots\}, \quad k \to \infty, \tag{4.15}$$

where the 0'th coordinate is either 0 or 1 with probability $\frac{1}{2}$. Whence in this sense the D_k's which count the number of symbol changes in finite records provide less and less information as k increases. In fact numerous simulations show that out of $N-1$ possible symbol changes about 80% are achieved already by D_{10}. Hence only D_k for rather low k are useful. (4.15) is called the Higher Order Crossings Theorem.

An important application of the higher order crossings D_k's is in the discrimination of time series. For this purpose we make use of another consequence of the Higher Order Crossings Theorem. It can be shown that for long and even moderate records lengths the D_k actually increase! This motivates the statistic

$$\psi_N^2 = \sum_{k=1}^{K} \frac{(\Delta_k - E\Delta_k)^2}{E\Delta_k}, \tag{4.16}$$

where

$$\Delta_k = \begin{cases} D_1, & k = 1 \\ D_k - D_{k-1}, & k = 2, \ldots, K-1 \\ (N-1) - D_{k-1}, & k = K \end{cases}$$

so that for sufficiently large N $\Delta_k \geq 0$ and $\sum_{k=1}^{K} \Delta_k = N-1$. Extensive simulations indicate that ψ_N^2 has an extremely robust distribution which varies only mildly from process to process. This statistic has proven useful in many cases and we shall report more on its applications in the near future elsewhere. It should be noted that for Gaussian processes the expected values $E\Delta_k$ can be computed exactly.

5. A REMARK ON THE CASE WHEN MOMENTS DO NOT EXIST

One of the great advantages of graphical features such as axis crossings of stationary sequences is that regardless of whether moments are finite or not or whether they exist at all, the number of axis crossings by a series of length N has moments of all orders. In such cases the axis crossings have a strong case for their use in inference. To bring a concrete example, consider the strictly stationary first order autoregressive Cauchy process

$$Z_t = \phi Z_{t-1} + u_t \ , \ t=0,\pm 1,\ldots \qquad (5.1)$$

where $|\phi|<1$, and the u_t are independent Cauchy random variables with characteristic function $e^{-(1-|\phi|)|s|}$. Then $\{Z_t\}$ is a strictly stationary process with Cauchy marginals having the standard characteristic function $e^{-|s|}$. In the usual situation when 2nd order moments of Z_t exist, ϕ is the correlation between Z_t and Z_{t-1} so that realizations appear either "smooth" or oscillatory depending on the sign of ϕ. In the Cauchy case ϕ is no longer a correlation but still the degree of oscillation depends on its sign! This means that ϕ is essentially a function of the degree of oscillation in the process or equivalently the number of axis crossings. This can also be seen from Table 1 which is the result of a simulation in which (5.1) was generated for different values of ϕ and the corresponding numbers of axis crossings in series of length $N=1000$ were recorded. It is interesting to observe that

1. AXIS CROSSINGS BY CAUCHY SERIES (5.1) OF LENGTH 1000

ϕ	D_1	$\hat{\phi}_\alpha$
.99	13	
.95	33	.95 , α = 2
.90	78	
.80	128	.82 , α = 0.5
.75	137	
.60	198	.60 , α = 0.5
.50	293	
.40	291	.41 , α = 0.25
.25	349	.20 , α = 0.25
.10	446	
.00	508	-.03 , α = 0.0
-.10	580	
-.25	656	-.28 , α = -0.1
-.40	705	-.41 , α = -0.1
-.50	747	
-.60	805	-.65 , α = -0.1
-.75	842	
-.80	891	-.81 , α = -0.1
-.90	921	
-.95	944	
-.99	985	

by comparison with previous results in the Gaussian case, the Cauchy series yields fewer axis crossings for positive ϕ but more crossings for negative ϕ. In addition, it was shown in Kedem (1980a) that in the Gaussian case the estimate $\cos(\pi D_1/(N-1))$, where D_1 stands for the number of axis crossings, is a remarkably good estimate. Taking these observations into account we suggest for the Cauchy case the estimate

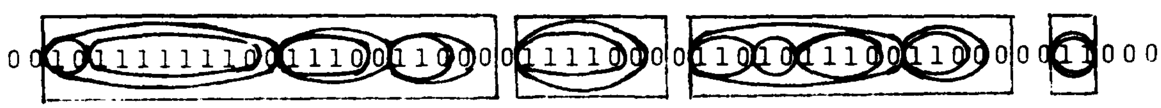
$$\hat{\phi}_\alpha = \cos\left(\frac{\pi D_1(1+\alpha)}{N-1}\right) \tag{5.2}$$

where α is a correction factor which has the same sign as ϕ. $\alpha = 0$ corresponds to the Gaussian case. Table 1 gives some $\hat{\phi}_\alpha$ for various α. Obviously α increases and decreases with ϕ. We intend to investigate this estimate in a future study.

6. FEATURES AND PROBABILITIES

We have seen above that the number of symbol changes in a clipped or randomly clipped binary series is a useful quantity which can and should be used in inference. Similarly, runs of various lengths and types are also important as they summarily portray useful graphical information about the original series. From these various types of runs or subsequences, we can actually compute various probabilities of interest. Specifically, we will briefly outline how to obtain the asymptotic distribution of the maximum in a stationary series provided certain conditions are assumed. But first we define a useful feature called a "unit."

An m'th order unit is a binary sequence which starts with a 1, ends with m separating 0's (if needed to separate it from other units) and in which each 0-run, if not an end run, consists of at most $m-1$ 0's. Note that the length of 1-runs is not restricted. For example in the binary series



There are 4 unit of order 3 and 7 units of order 2 and 10 units of first order. If 3 is the highest order under consideration, then there are 9 0's which do not belong to any unit and we call them the free 0's. Thus, if binary sequences are perceived in terms of units then the information in such sequences is neatly summarized.

Now, let Z_t be any stationary series of length N and clip the series at a certain level. This yields a binary series X_t. Let $n_{ij...k}$ be the frequency of $ij...k$ in the binary series, and let S be the number of 1's. Assume now that the highest order unit is m. Then the number of m'th order units is

$$s - n_{11} - n_{101} - n_{1001} - \cdots - n_{\underbrace{100...0}_{m-1}1} , \tag{6.1}$$

and the number of free 0's is

$$(N-s) - m(\text{\# of m'th order units } -1) - (m-1)n_{\underbrace{100...0}_{m-1}1}$$

$$- (m-2)n_{\underbrace{100...0}_{m-2}1} - \cdots - n_{101}. \tag{6.2}$$

Assume that as the level at which the Z_t series is clipped, the binary series displays an m-th order Markov dependence where m may be even very large. Observe that as the level increases the 1's become rare. It can be shown then under some conditions (Kedem (1980b) that for a high level and large N

$$\Pr(S=s) \sim \begin{pmatrix} \text{\# of permutations of the m'th} \\ \text{order units with the free 0's} \end{pmatrix} P^{s}_{\underbrace{100...0}_{m}} \; P^{n-s-m}_{\underbrace{00...0}_{m+1}} \tag{6.3}$$

where

$$P_{x_t x_{t-1} \cdots x_{t-m}} = \Pr(X_t = x_t | X_{t-1} = x_{t-1}, \ldots, X_{t-m} = x_{t-m}).$$

Now let $P_{100\ldots0} \to 0$ such that $NP_{100\ldots0} = \beta$ is fixed. Then it is not difficult to see that

$$\Pr(S=s) \to e^{-\beta} \beta^s/s! \ .$$

Therefore for large N

$$\Pr(\max_{1\le t\le N} Z_t < \text{a high level } u) \sim e^{-E(\#\text{ of exceedances above level } u)}.$$

If the 1's tend to cluster, a similar argument replaces the number of exceedances of level u by the number of upcrossings of level u.

7. SUMMARY

We have illustrated briefly the connection between visual quantities such as crossings of fixed and random levels and the covariance function of a stationary process, some parameters of interest and the distribution of the maximum in stationary series. In particular we have focused on some graphical features of time series which contain a great deal of information useful in inference. For this purpose we have created features by clipping at random levels and by changing the position of fixed levels in a controlled manner. Features such as peaks, troughs, inflection points, axis crossings by the k'th difference of a stationary process, etc., are useful up to a point. This is the subject of the Higher Order Crossings Theorem.

REFERENCES

Beniger, James. and Robyn, Dorothy L. (1978), "Quantitative Graphics in Statistics: A Brief History," *The American Statistician*, 32, 1-11.

Fienberg, Stephen E. (1979), "Graphical Methods in Statistics," *The American Statistician*, 33, 165-178.

Kedem, Benjamin (1980a), *Binary Time Series*, New York: Marcel Dekker.

_______________ (1980b), "On Nearest Neighbor Degeneracies of Indistinguishable Particles," Technical Report TR80-35, Department of Mathematics, University of Maryland.

_______________ and Slud, Eric (1980), "The Signature Problem for Stationary Time Series," submitted for publication.

END

DATE

FILMED

9-80

DTIC