THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 53, No. 7, September 1974
Printed in U.S.A.

A Statistical Analysis of Telephone Noise

By B. W. STUCK and B. KLEINER

(Manuscript received January 28, 1974)

The results of a statistical analysis of telephone noise are presented. The analysis consists of two stages: an exploratory data analysis stage, where the data are characterized through various nonparametric statistics and a model-building stage, where the data are matched to models.

The exploratory data analysis stage involved examination of noise waveforms, power spectra, and covariance estimates. The results show that telephone noise consists of a deterministic component (sinusoids at various frequencies) and a stochastic component. It is assumed that these components add. The data are filtered to remove the deterministic component. Next, central moment estimates are presented, as well as first-order amplitude statistics (histograms and empirical cumulative distributions) for these filtered data. The results indicate that the filtered data appear wide-sense stationary over short periods of time (typically 1 second) and, although close to gaussian, are distinctly nongaussian.

The model-building stage involved fitting the filtered data to two classes of models. The first class of models is based on symmetric stable distributions that arise from the central limit theorem. The second class of models assumes two different physical processes that contribute to the random component of telephone noise: The low-variance process is assumed to be gaussian, while the high-variance component is assumed to be a filtered Poisson process. Both classes of models agree intuitively with the physical processes generating telephone noise and are mathematically tractable. Based largely on graphical tests, both models appear to fit the filtered data reasonably well.

I. INTRODUCTION

Noise on telephone lines has puzzled and plagued people since the invention of the telephone. While it is common knowledge that telephone channel noise is nongaussian, nowhere in the literature is there a clear account of an adequate statistical characterization of telephone

noise. In part, this is due to the fact that only recently have statistical tools been developed that are equal to the task.

This paper attempts to adequately characterize some statistical properties of telephone channel noise by means of various nonparametric statistics and by mathematical models. It is encouraging to note that the results presented here do not contradict those found in earlier works. However, since only a small number of telephone line noise sample functions were examined, the results must be regarded as tentative, awaiting independent checks by other investigators. It is hoped the results presented here will stimulate communication theorists to investigate new methods for optimally processing signals corrupted by the nongaussian noise models presented here. Work along these lines might lead to optimum and practical suboptimum receiver structures for combating telephone noise. This in turn might permit greater insight into how noise limits telephone channel performance with regard to voice communication or data transmission.

The authors have tried to keep the notation and nomenclature consistent with that used in statistics and probability theory. The reader is reminded, for example, that "empirical cumulative distribution function" refers to an estimate of the true "cumulative distribution function" based on observations of "empirical" data. The words "sample" and "empirical" are often used interchangeably, as in "sample mean" and "empirical mean," as compared with the ensemble mean.

1.1 Summary of past work

Broadly speaking, previous investigators characterized telephone channel noise in two different ways, based on different ways of measuring the data and with different problems in mind. First, direct measurements of sample functions of telephone channel noise have been carried out[1–3] and mathematical models for the noise statistics have been constructed. Second, digital signals have been transmitted over telephone lines and the difference between the transmitted and received signals has been analyzed, providing an indirect measurement of telephone channel noise.[4–16] It is extremely difficult to correlate these two types of measurements. This paper is solely concerned with direct measurements of telephone channel noise sample functions.

Both types of measurements indicated the nongaussian nature of the noise. The analog measurements suggested that telephone noise could be considered a mixture of a nongaussian random process with sinusoids at various frequencies.[1–3] The first-order amplitude statistics for the random process appeared to be adequately modeled by a Pareto

distribution.[2] Analysis of errors in digital signals transmitted over telephone lines revealed that errors cluster in time, an indirect measure that the noise cannot be adequately modeled as white and gaussian.[4,13]

Some causes of telephone noise are thermal noise in amplifiers,[1,17] switches sparking on opening or closing contact,[1–3,17–19] lightning,[3,17] electromagnetic crosstalk,[17] fading on microwave links and switching to guard bands,[3,17] echo suppressor turnaround,[17] disturbances because of maintenance,[17] power line harmonics as well as harmonics of all sinusoids used in the telephone switching system,[17] and noise generated at switch interfaces.[17] The main causes of telephone noise are felt to be due to thermal noise, switch arcing, and pickup of unwanted sinusoidal harmonics.[17] The main cause of error clustering in digital signals is felt to be due to the impulsive component of the noise, generated by switch arcing.[17]

1.2 Problem statement

The problem is twofold:

(*i*) To provide an adequate statistical characterization of telephone channel noise by means of various nonparametric statistics.

(*ii*) To allow the data plus knowledge of the physical processes generating telephone noise to lead to a mathematically tractable class of models.

1.3 Outline of the paper

The data from five telephone lines and the processing necessary to convert the data into a form suitable for further analysis are described first. The analysis is broken down into two steps, an exploratory data analysis stage where the data are characterized through various nonparametric statistics and a model-building stage where the data are matched to models.

The exploratory data analysis stage involved examination of noise waveforms, power spectra, and covariance estimates. The results show that the data consisted of a deterministic component (sinusoids at various frequencies) plus a stochastic component, which were assumed to be independent. The data were filtered to remove the sinusoids that were significantly larger than the stochastic component. Histograms and empirical cumulative distribution functions for the filtered data were examined, as well as central moment estimates. The filtered data appeared to be wide-sense stationary over short periods of time, typically 1 second. Based largely on quantile-quantile plots, it was concluded that, although close to gaussian, the filtered data for three out

of the five lines were distinctly nongaussian; the filtered data for the remaining two lines appeared to be gaussian.

The model-building stage involved fitting the filtered data to two classes of models. The first class of models is highly unstructured; it was based on stable distributions, infinitely divisible distributions that arise from the central limit theorem. Based on a series of parameter estimation procedures including robust estimation, maximum likelihood estimation, and quantile-quantile plots and backed up by a likelihood ratio test on the goodness-of-fit, the three nongaussian lines could be adequately modeled by a stable distribution with characteristic index ≈ 1.95 (a gaussian has characteristic 2.0).

The second class of models is much more structured than the first. Two different physical processes were assumed to contribute to the filtered data: the low-variance component was a stationary gaussian process, while the high-variance component was a filtered Poisson process. Parameters for the gaussian component were estimated using trimmed means and trimmed variances. The parameters specifying the filtered Poisson process were much more complicated to estimate. The instants of time at which noise bursts occurred and the intervals between bursts were first examined; based on power spectra as well as covariance estimates, the intervals appeared to come from a renewal process. Histograms and empirical cumulative distribution functions indicated that the time intervals came from a Poisson process; empirical survivor and hazard function plots showed that a Poisson process with constant rate parameter was not an adequate model, however. Because of the small number of bursts observed, it was quite difficult to fit the time intervals to a Poisson process with varying rate parameter, and for expediency a constant Poisson rate parameter was chosen to model noise burst times of occurrence. The amplitudes of the noise bursts were adequately modeled by a log normal and power Rayleigh, or generalized gamma. The durations of actual noise bursts were used to estimate parameters in the noise burst shaping filter. A simple indication is presented of how well the filtered data fit the gaussian-plus-filtered-Poisson-process model. A number of other models and extensions of these models are discussed.

II. EXPLORATORY DATA ANALYSIS

2.1 Description of the data

The data, supplied by J. Fennick, consist of analog tape recordings of noise on five telephone lines. Figure 1 is a diagram of the measure-

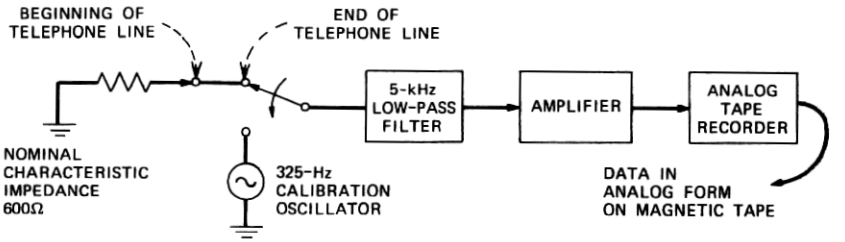


Fig. 1—Telephone noise measurement system.

ment system. Table I describes the origin and destination of each line, as well as the date and time of the recording.

The origin of the telephone line was terminated with the nominal characteristic impedance of the line, 600 ohms. The output of the line was low-pass filtered to remove spurious out-of-band signals, amplified, and recorded on an analog tape at 18.75 cm per second. The system was calibrated before each recording with a 15-second burst of a 325-Hz sinusoid at a predetermined amplitude. The dynamic range of the recording system was approximately 100.[20] No attempt was made to eliminate dc offset and drift. Each recording was approximately 15 minutes long.

Figure 2 is a block diagram of the analog-to-digital tape conversion system. The analog tapes were played back on an analog tape recorder (of a different model than that on which they were recorded) at 18.75 cm per second. The calibration signal set the gain on the playback amplifier so that the calibration signal amplitude was roughly equal to its value at the recording site. There was no attempt to remove wow and flutter in the tape recording.[21] The signal was low-pass filtered (to lessen the chance for spectral aliasing), amplified, sampled 10,000 times a second, passed through an analog-to-digital converter, and subsequently put into digital format on a tape suitable for further processing

Table I — Description of data

Line	Origin	Destination	Approximate Length (km)	Date	Time
1	Monroe, Mich.	Detroit, Mich.	55	7/8/64	2:06 p.m.
2	Saginaw, Mich.	Detroit, Mich.	160	7/8/64	2:49 p.m.
3	St. Louis, Mo.	Ft. Smith, Ark.	690	8/4/64	10:30 a.m.
4	Little Rock, Ark.	Ft. Smith, Ark.	250	8/4/64	1:55 p.m.
5	Newark, N. J.	Binghamton, N. Y.	400	12/12/63	12:52 p.m.

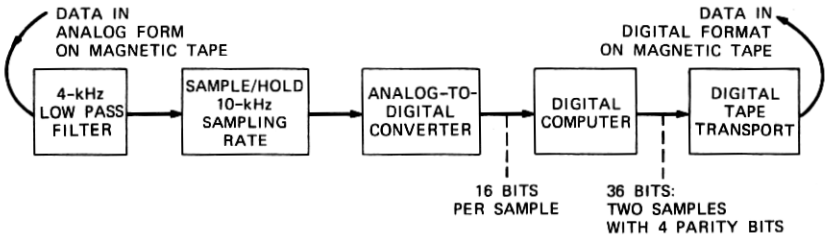


Fig. 2—Analog-to-digital conversion of telephone noise data.

by a Honeywell 6070 digital computer. The levels of the analog-to-digital converter will be the units specifying the amplitudes of the telephone noise data. No measures of the degradation in data on digital tapes resulting from jitter during the sampling process are available; it is assumed to be negligible compared to other sources of measurement error. No bounds are available on the loss of information entailed by examining a continuous-time random process at only discrete instants of time.[22]

Two critical remarks concerning the method of recording the data should be made: first, there are no quantitative measures available on the amount of noise introduced into the data by the measurement system alone. Presumably, any measurement system noise was insignificant compared to the telephone channel noise. Second, the dynamic range of the recording system is probably insufficient to faithfully record telephone noise; a much more satisfactory dynamic range would be 1000 to 10,000. Both issues have been dealt with elsewhere (in a different context); a possible solution would be to convert the data into digital format directly at the recording site.[23] Consideration of these problems is left to future research; the data analysis proceeded with these caveats in mind.

Figure 3 shows a typical telephone channel noise waveform from line 1 after conversion to digital format.

How typical are these data compared with that observed on other telephone lines? A search of the literature as well as private communications from engineers shows that the data discussed here appear to be typical of telephone channel noise. Throughout this investigation, nothing was uncovered that contradicted earlier work; rather, this work tends to clarify and place in perspective that of earlier investigators. Note too that the telephone lines examined here were typically several hundred miles long, presumably passing through a variety of equipment, and hence quite representative of telephone noise. Finally,

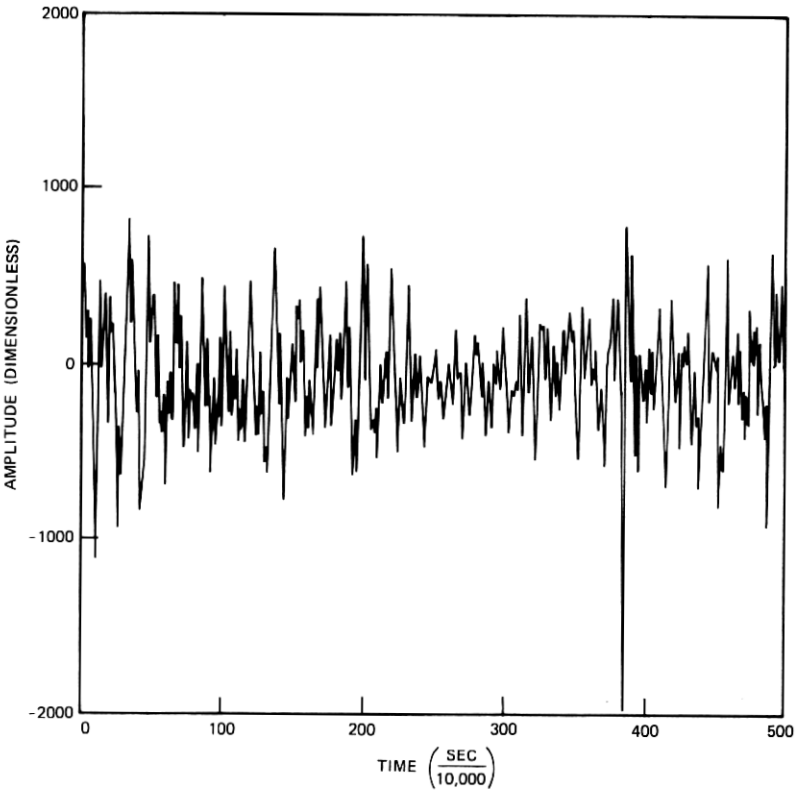


Fig. 3—Typical line 1 telephone noise waveform (unfiltered).

the analog tape recordings were played into loudspeakers, and the authors felt the noise sounded typical.

A much more serious objection to the present analysis is that not enough of the data at hand was examined. If all five 15-minute noise recordings were sampled 10,000 times a second and then put on to tape in digital format, more than 45 million noise data must be analyzed; in particular, 9 million data must be processed and statistically characterized for just one telephone line sample function. In practice, only 10-second segments from the beginning and middle of a recording were examined in detail and compared with each other. No unusual statistical differences were observed between these segments for any telephone line. The main reason for examining so little data was the great cost of analyzing these data statistically and, in particular, of digitally filtering the data to remove sinusoids. It is difficult to predict

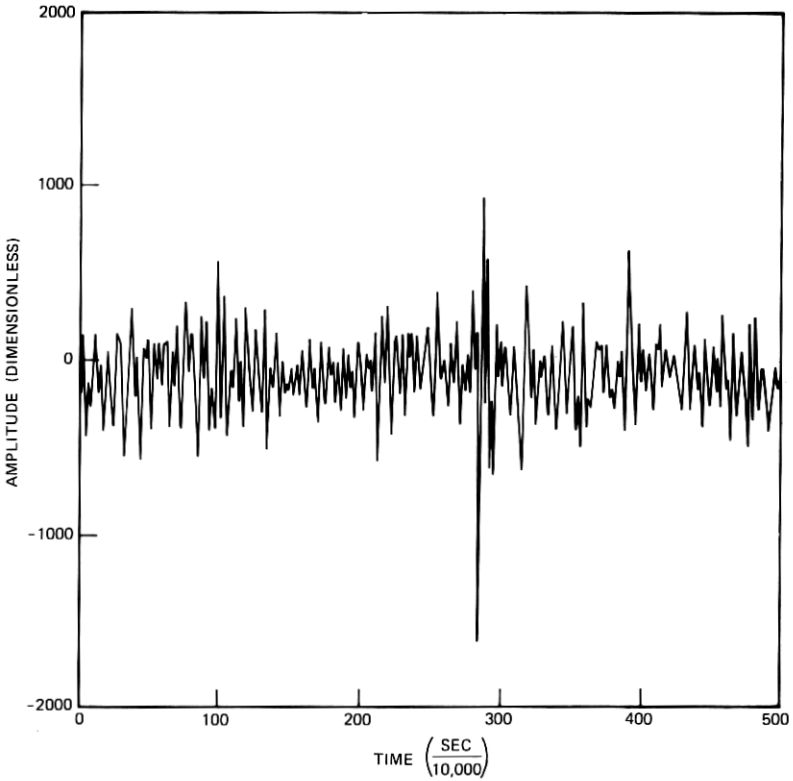


Fig. 4—Typical line 1 telephone noise waveform (filtered).

in advance exactly which sinusoids are present on a particular telephone line; it is easier to filter these out digitally after the measurement without distortion than to accomplish this with analog filters.

2.2 Data modifications and estimation of power density spectra

As mentioned above, the data went through several stages before they were available in digital form. Some further processing is necessary to remove the effects of this pre-processing, as well as to remove unwanted sinusoids. Since the frequency response of each line was unknown, nothing was done to compensate for it.

The first step is to compute estimates of the power spectrum. The data were segmented into blocks (typically of length 1000, corresponding to $\frac{1}{10}$ of a second of noise). Each block was tapered and enlarged to 1024 values by adding zeros, then transformed into the frequency

domain by the fast Fourier transform (FFT);[24] the power spectrum was estimated from the transformed data. The method used is comparable in bias and variance to other spectral estimation procedures, but requires considerably less computer time than other non-FFT-based methods.[25–27] Furthermore, it is possible to calculate cross-spectrum estimates easily, as well as to check for nonstationarity by computing the spectra of successive segments of data. The discrete Fourier transform of N successive data at frequency $w_l = l/N\,\Delta t$ is $\hat{f}(w_l)$,

$$\hat{f}(w_l) = \sum_{K=1}^{N} n_K e^{-i2\pi Kl/N} \qquad l = 0, 1, \cdots, \left[\frac{N}{2}\right],$$

where

n_K = sample of noise waveform at time $K\Delta t$

and

Δt = time interval between samples.

The estimate of the power spectrum density at frequency w_l is $\hat{S}(w_l)$,

$$\hat{S}(w_l) = \frac{\Delta t}{N} \sum_{j=-M}^{M} |\hat{f}(w_{l+j})|^2 g_j,$$

where

$$\sum_{j=-M}^{M} g_j = 1,$$

and the weights $\{g_j\}$ are introduced to smooth the estimate. Unequal weights can be used to lower the bias of the estimate, but increase its variance. The value $\hat{S}(w_l)$ represents the average noise power density in a frequency band centered at w_l. All power spectrum density estimates shown here were computed with g_j equal to $(M - |j|)/M^2$, where $M = 5$. Figure 5 shows the power spectrum for the waveform in Fig. 3 with two sharp peaks probably reflecting sinusoids at 650 and 4300 Hz. Figure 6 shows a succession of 24 power spectrum estimates for line 1 for $\frac{1}{10}$-second segments of filtered data. The first 13 are from successive segments recorded at the beginning, while the final 11 are from successive segments recorded 5 minutes later. These results indicate that line 1 data can be regarded as wide-sense stationary over at least 1-second time intervals. The variance of these spectral estimates is unknown; if the process were gaussian, the distribution of $S(\hat{w}_l)$ can be approximated by a χ^2 distribution with $[2/\sum_{j=-M}^{+M} g_j^2]$ degrees of freedom.[28] For $g_j = (M - |j|)/M^2$ with $M = 5$, this results in approximately 14 degrees of freedom. From the data shown here for line

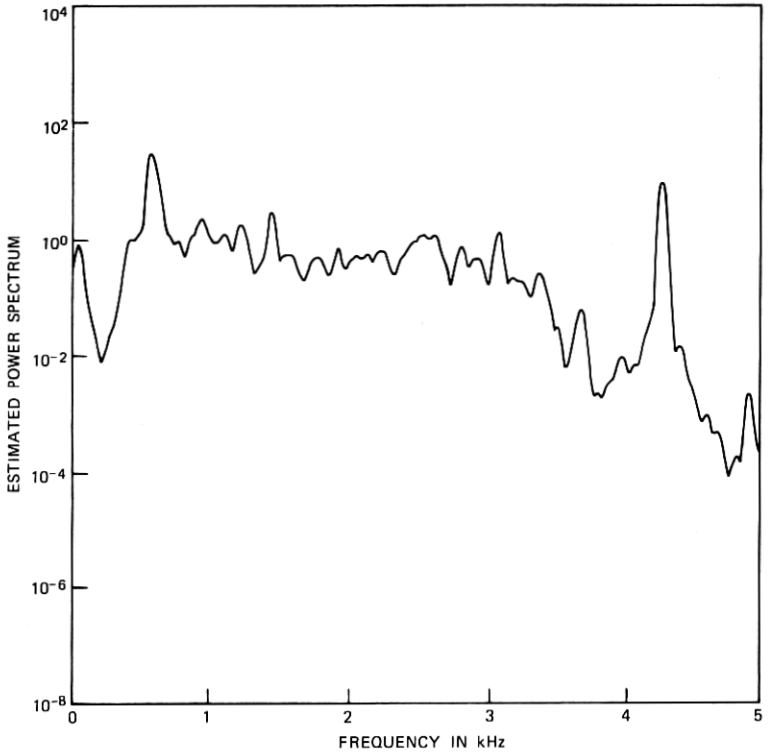


Fig. 5—Typical power spectrum for unfiltered line 1 data ($N = 1000$).

1, as well as data from the other four lines, telephone noise power density spectra appear to have the same shape, but different scales.

Table II summarizes the estimates of frequencies of signals which were quite probably sinusoids, and whose estimated power spectrum density was at least a factor of 10 above the estimated wideband power

Table II — Estimated frequencies of sinusoids that were subsequently filtered out

Line	Estimated Frequency of Sinusoid (Hz)
1	650, 4300
2	650, 4300
3	3900
4	2000, 3600
5	60

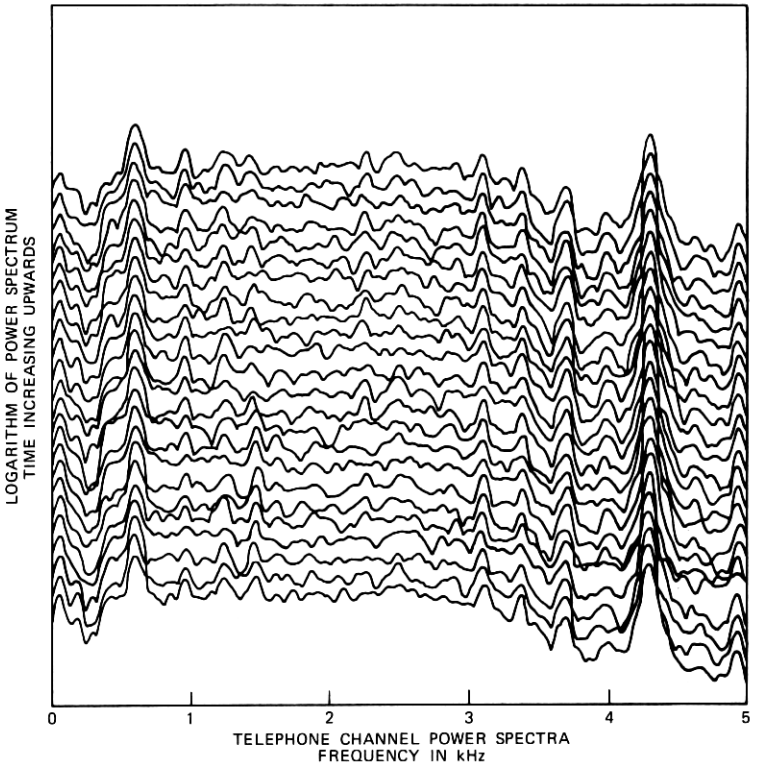


Fig. 6—Line 1 power spectra: bottom 13 from beginning of line 1 data, top 11 from middle of line 1 data ($N = 1000$).

spectrum density. While this criterion is arbitrary, independent experimental evidence to be discussed shortly indicates it is adequate from a statistical point of view.

Since many statistical tests require uncorrelated samples, it is necessary to filter out these sinusoids, as well as to compensate for distortions in the data from the measurement system. This implicitly assumes that telephone noise can be modeled as the sum of a deterministic process, sinusoids at various frequencies, and a purely stochastic process, which will be characterized in greater detail. This was accomplished using low-pass, band-stop, and high-pass linear-phase digital filters designed by computer programs developed by L. Rabiner; the filtering was carried out in the discrete time domain by convolution. Figure 7 shows the power spectrum of a filtered segment of the data shown in Fig. 4.

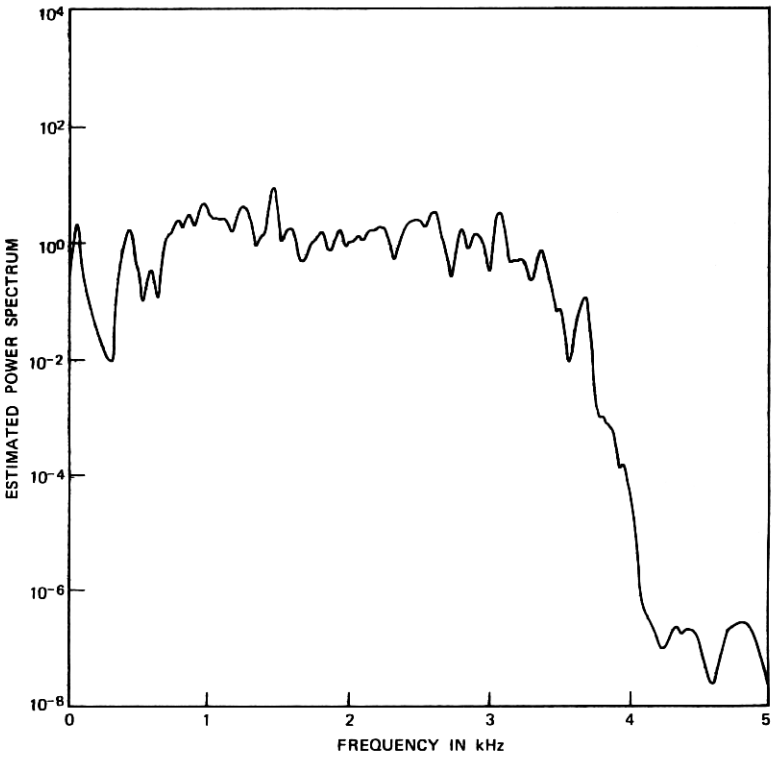


Fig. 7—Typical power spectrum for filtered line 1 data.

Because of the difficulty in finding telephone lines completely free from sinusoidal interference, the question arises as to how much harmonic content can be tolerated in performing various statistical tests. Work carried out elsewhere in a different context has examined this issue from an experimental viewpoint;[29] the principal findings were that the amplitude statistics are apparently not significantly degraded by the linear filtering, if the sinusoid is the same size or smaller than the observed noise levels. This topic can be a subject for future research.

2.3 Covariance estimation

It is assumed in many statistical computations that the data are statistically independent. In practice, the data usually depend to some extent on each other, and it is often quite difficult to quantitatively estimate the effects of this lack of independence. One indication of

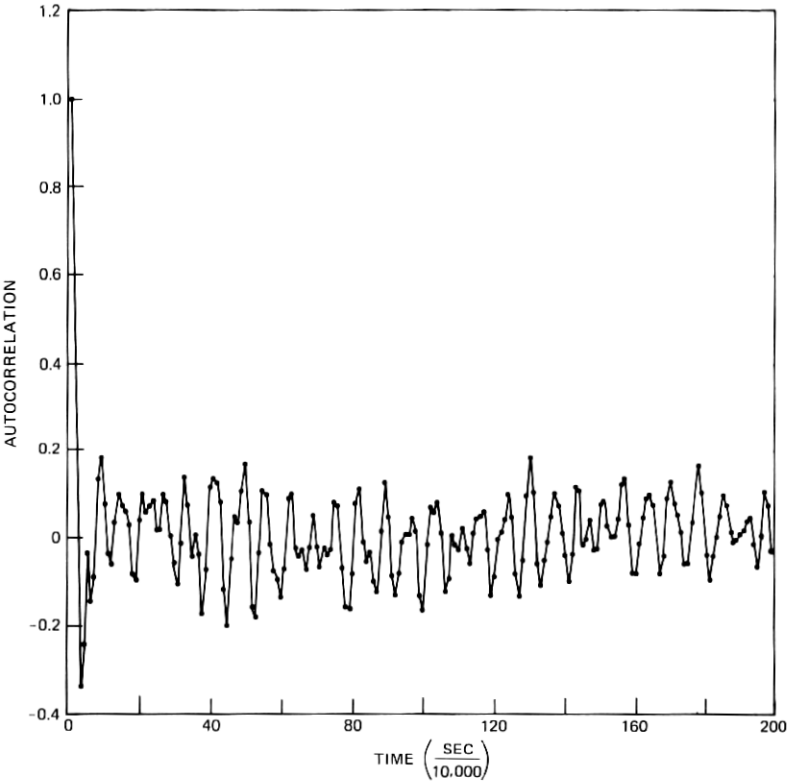


Fig. 8—Typical line 1 autocorrelation function ($N = 1000$).

independence is the estimated autocorrelation function,

$$\hat{R}_{nn}(l\Delta t) = \frac{1}{\bar{R}_{nn}(0)} \cdot \frac{1}{N} \sum_{K=1}^{N-l} n'(K\Delta t)n'(K\Delta t + l\Delta t) \quad l = 1, \cdots, N-1,$$

where

$$n'(K\Delta t) = n(K\Delta t) - \frac{1}{N} \sum_{l=1}^{N} n(l\Delta t)$$

and

$$\bar{R}_{nn}(0) = \frac{1}{N} \sum_{K=1}^{N} n'^2(K\Delta t).$$

A typical autocorrelation of filtered data is plotted in Fig. 8. A sinusoid that was not filtered out is quite evident at approximately 1400 Hz (see also Fig. 7); ignoring this sinusoid,[29] the autocorrelation appears to be approximately zero for $l \geqq 3$.

If the data are wide-sense stationary and ergodic, then the autocorrelation and the power density spectrum are a Fourier transform pair.[30]

Examination of the filtered waveform in Fig. 4 indicates that the samples appear uncorrelated, i.e., they are scattered in a random fashion about a location parameter.

The sample normalized cross covariance for two different segments of data, $\{x(\Delta t), x(2\Delta t), \cdots, x(N\Delta t)\}$ and $\{y(\Delta t), y(2\Delta t), \cdots, y(N\Delta t)\}$, is defined as

$$\hat{R}_{xy}(l\Delta t) = \frac{\frac{1}{N}\sum_{K=1}^{N-l} x'(K\Delta t)y'(K\Delta t + l\Delta t)}{\tilde{R}_{xx}^{\frac{1}{2}}(0)\tilde{R}_{yy}^{\frac{1}{2}}(0)} \quad l = 0, 1, \cdots, N-1,$$

where

$$x'(K\Delta t) = x(K\Delta t) - \frac{1}{N}\sum_{l=1}^{N} x(l\Delta t),$$

$$y'(K\Delta t) = y(K\Delta t) - \frac{1}{N}\sum_{l=1}^{N} y(l\Delta t),$$

$$\tilde{R}_{xx}(0) = \frac{1}{N}\sum_{K=1}^{N} x'^2(K\Delta t),$$

and

$$\tilde{R}_{yy}(0) = \frac{1}{N}\sum_{K=1}^{N} y'^2(K\Delta t),$$

and is shown in Fig. 9 for two typical segments of filtered data. From this as well as other data, the filtered telephone noise data examined appear to be uncorrelated over short time intervals.

Since the data, after filtering, appear approximately uncorrelated, they will now be characterized in greater detail.

2.4 First-order filtered data amplitude statistics

A nonparametric statistical description of first-order noise amplitude statistics provides a great deal of useful information. For example, if the data are independent identically distributed random variables, they can be completely characterized by their empirical cumulative distribution function.[31] The work in this section relies heavily on graphical methods for data analysis, to give more physical insight into the nature of the data.[32]

The empirical or sample cumulative distribution function is defined as

$$\hat{F}(X) \triangleq \frac{\text{number of observations less than or equal to } X}{\text{total number of observations}},$$

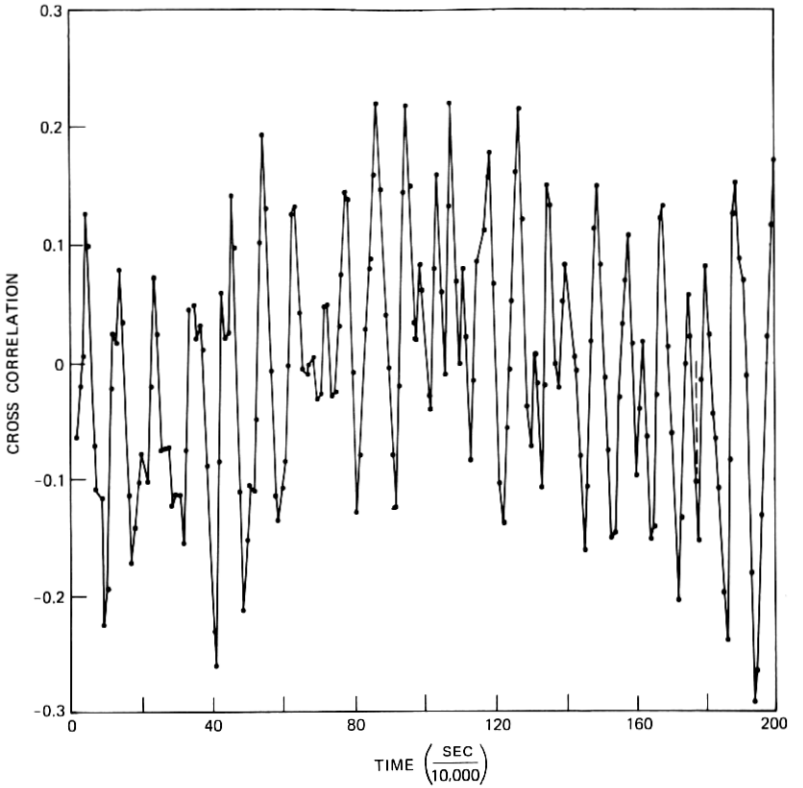


Fig. 9—Typical line 1 crosscorrelation function ($N = 1000$).

which is a function of $\{x_K\}$, the set of observations. The sample histogram is defined as

$$P(X, X + \Delta) \triangleq \text{number of sample values in } [X, X + \Delta],$$

where Δ is the bin width. Figure 10 is a plot of a typical empirical cumulative distribution function, and Fig. 11 shows a typical sample histogram. These two figures imply that the first-order probability density for the data is roughly bell-shaped and symmetric. A simple graphical symmetry check on the empirical cumulative distribution is shown in Fig. 12; x_K is plotted against x_{N-K+1}, where $K = 1, 2, \cdots, [N/2]$, $N = 1000$ is the total number of observations, and $\{x_K\}$ is the set of ordered observations. If the empirical cumulative distribution

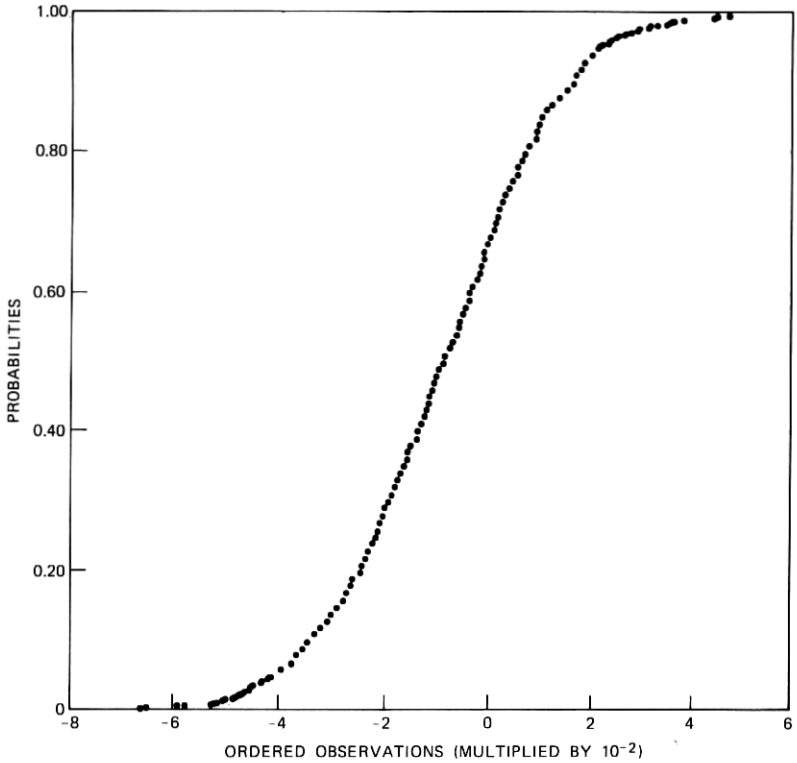


Fig. 10—Typical line 1 empirical cumulative distribution function ($N = 1000$).

is symmetric, these points lie on a straight line with negative unit slope; this is apparently the case.

The next quantities of interest are central moment estimates, which are defined as follows:[33]

$$\bar{x} = \text{sample mean} = \frac{1}{N}\sum_{j=1}^{N} x_j,$$

$$s^2 = \text{sample variance} = \frac{1}{N}\sum_{j=1}^{N} (x_j - \bar{x})^2,$$

$$\hat{a}_3 = \text{sample skewness} = \frac{1}{N}\sum_{j=1}^{N} (x_j - \bar{x})^3/(s^2)^{\frac{3}{2}},$$

and

$$\hat{a}_4 = \text{sample kurtosis} = \frac{1}{N}\sum_{j=1}^{N} (x_j - \bar{x})^4/(s^2)^2.$$

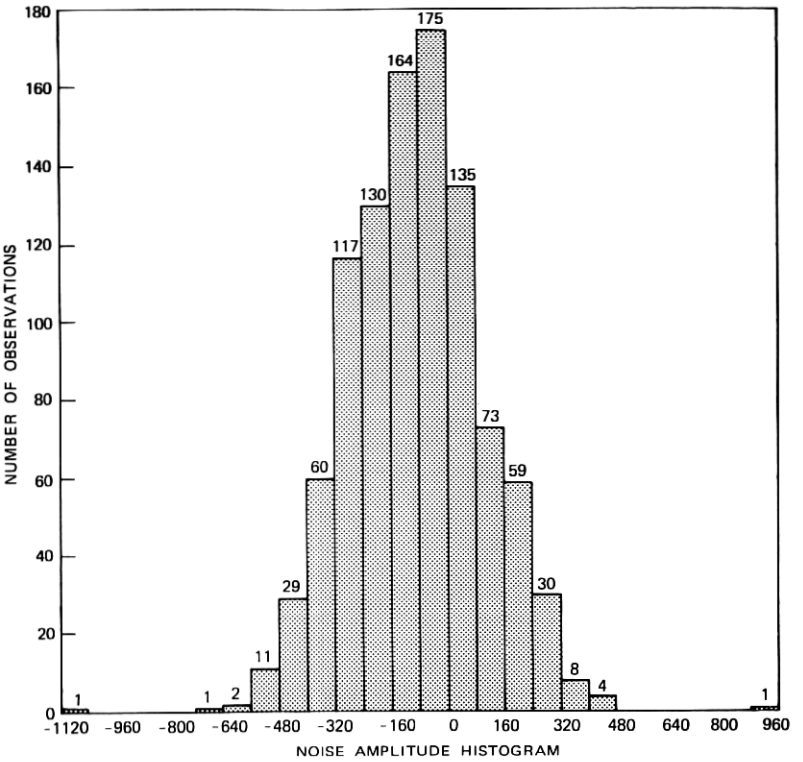


Fig. 11—Typical line 1 histogram ($N = 1000$).

These parameters were estimated for ten segments of 1000 data for each of the five lines. Table III shows these estimates for the segment of each line whose fourth central moment was the median of all ten fourth-central-moment estimates of this line. The 5-percent significance level for 1000 independent identically distributed gaussian random variables with known mean and variance are[33]

$$-0.127 < \hat{a}_3 < 0.127$$
$$2.76 < \hat{a}_4 < 3.26.$$

Figure 13 shows a scatter plot of $\hat{a}_3$ vs. $\hat{a}_4$ for successive segments of 1000 data for each of the five lines. Based on this evidence, it can be conjectured that lines 1, 2, and 4 are nongaussian, while the gaussian hypothesis cannot be rejected for lines 3 and 5. Since quite a large body of literature exists on gaussian random processes and these random processes are well understood, the gaussian hypothesis is not

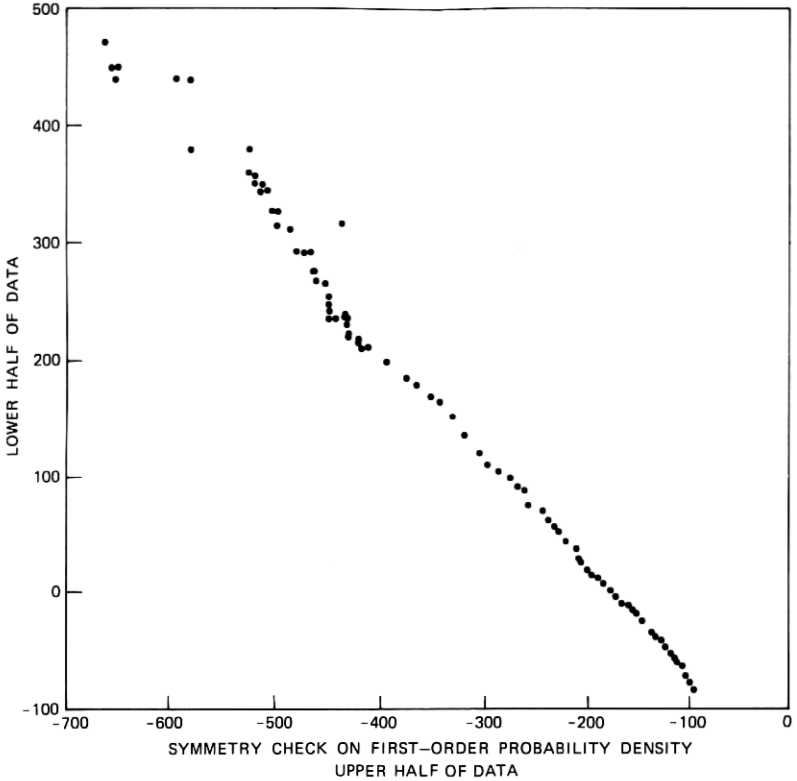


Fig. 12—Typical symmetry check on line 1 empirical cumulative distribution function ($N = 1000$).

lightly discarded: the evidence that the data are nongaussian should be much more convincing than that presented so far.

A very convenient graphical method to check how well data fit a theoretical distribution function is the quantile-quantile, or Q-Q, plot.[32] The qth quantile of a cumulative distribution function $F(x)$ is defined here as the value x for which $F(x) = q$, $0 \leqq q \leqq 1$. A Q-Q plot plots

Table III — Estimated telephone noise central moments

Line	$\bar{x}$	s^2	$\hat{a}_3$	$\hat{a}_4$
1	−87.9	38,600	0.05	3.4
2	−80.1	18,200	0.08	3.5
3	−80.1	44,200	−0.05	3.1
4	−82.3	87,200	0.02	3.3
5	− 1.06	1,990	−0.08	2.9

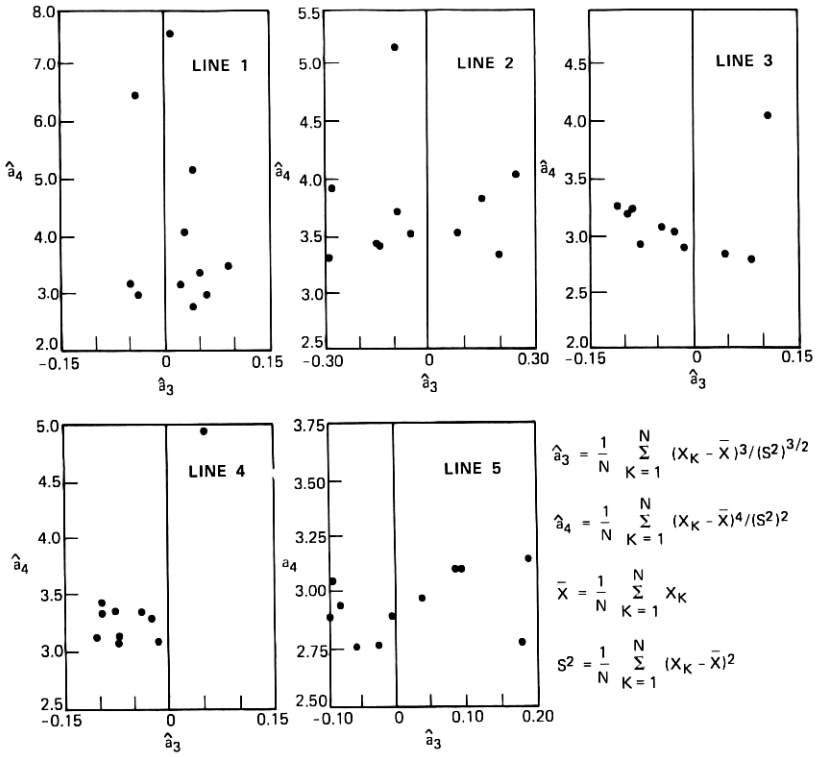


Fig. 13—Scatter plot of estimated third central moment $\hat{a}_3$ vs estimated fourth central moment $\hat{a}_4$ for at least ten successive segments of 1000 data ($N = 1000$) for all five lines.

quantiles of the empirical cumulative distribution function against quantiles of the theoretical distribution. If the empirical and theoretical distribution functions are the same, the plot is a straight line with slope +1 passing through the origin. If the empirical and theoretical distribution functions are the same to within a linear transformation (i.e., to within a scale and location parameter) the plot is still a straight line. A typical quantile-quantile plot for line 1 filtered data against a gaussian distribution is shown in Fig. 14; the sample size was 13,000. The first 100 and last 100 quantiles, as well as every hundredth quantile in the middle, have been plotted, giving the illusion of discontinuity during the transition from the middle to the tail quantiles.[32] Ten observations in each tail are widely scattered. Figure 15 shows the central portion of the quantile-quantile plot with these observations excluded. The tails curve toward horizontal lines,

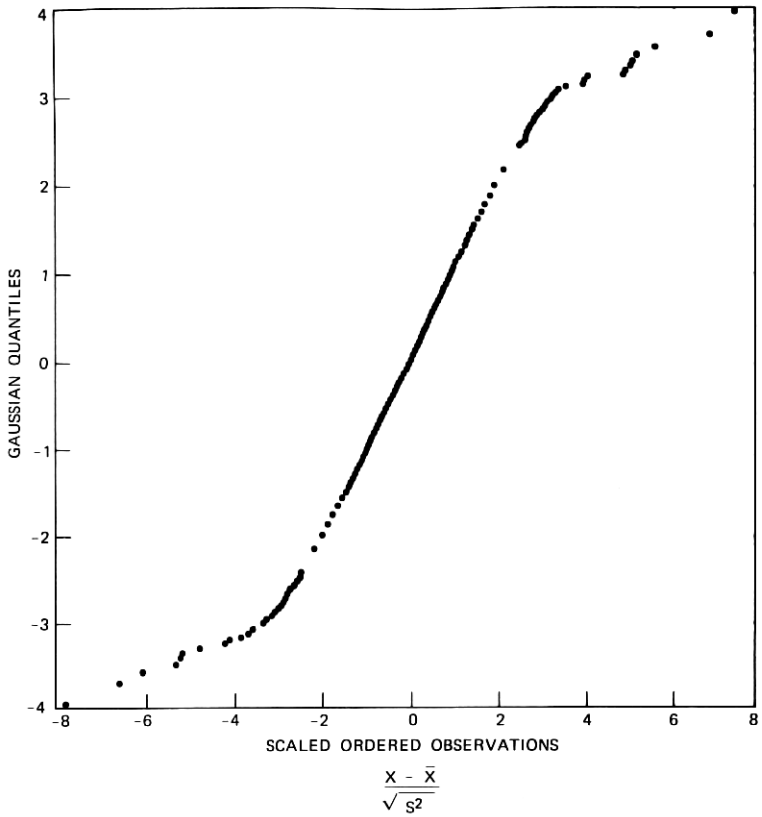


Fig. 14—Q-Q plot for 13,000 line 1 data against gaussian distribution ($\bar{X}$ = sample mean, S^2 = sample variance).

another indication of the long-tailed nongaussian nature of the data. The 10 points on each tail were found to be highly correlated: These very large excursions occurred in clumps of two, three, and five at a time, violating the assumption that the data are independent. For comparison, Fig. 16 shows a quantile-quantile plot for line 5 filtered data against a gaussian distribution; the sample size was 11,000. The straight line is a good indication that these data are gaussian.

III. MODELS

3.1 Central limit theorem

Since noise on telephone lines is presumably due to a large number of independent causes, it is worthwhile to digress and review the central

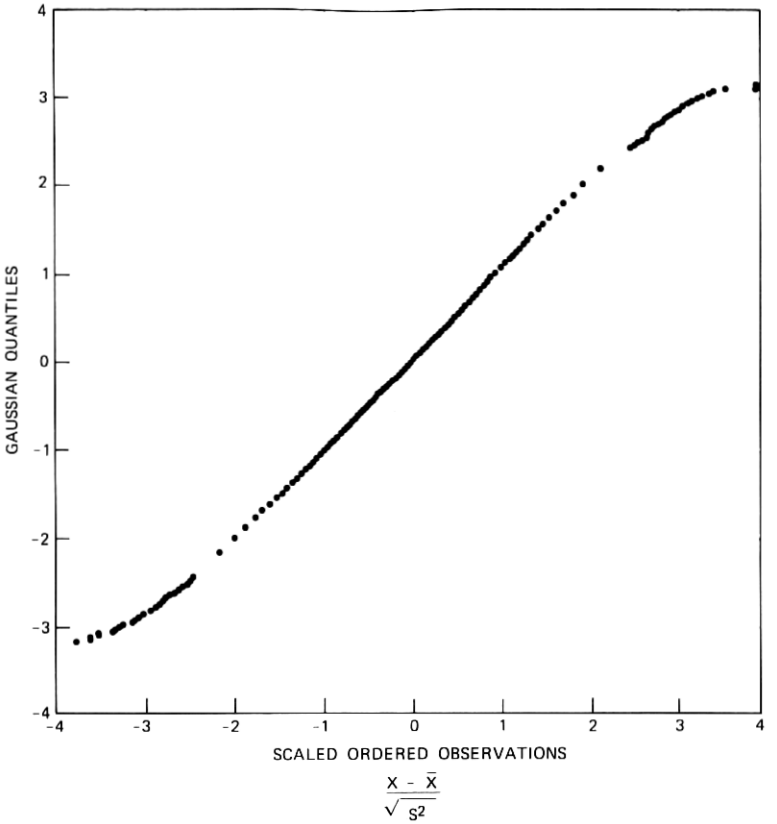


Fig. 15—Center portion of line 1 Q-Q plot for gaussian model (N = 12,980) ($\bar{X}$ = sample mean, S^2 = sample variance).

limit theorem. The material presented here is largely tutorial, following standard references.[34,35] The close association between the central limit theorem and the gaussian distribution is remarkable because of its algebraic closure property: If two independent random variables are both gaussian, their sum is also. It is much less widely known that the gaussian distribution is a special case of a larger family of distributions, which arise from the central limit theorem and exhibit the same closure properties as the gaussian, the stable distributions.

The reason for the importance of the gaussian rather than the entire stable distribution family is that only the gaussian distribution has a finite variance, and infinite variance is felt to be physically inappropriate in virtually any context. However, this is naive in that

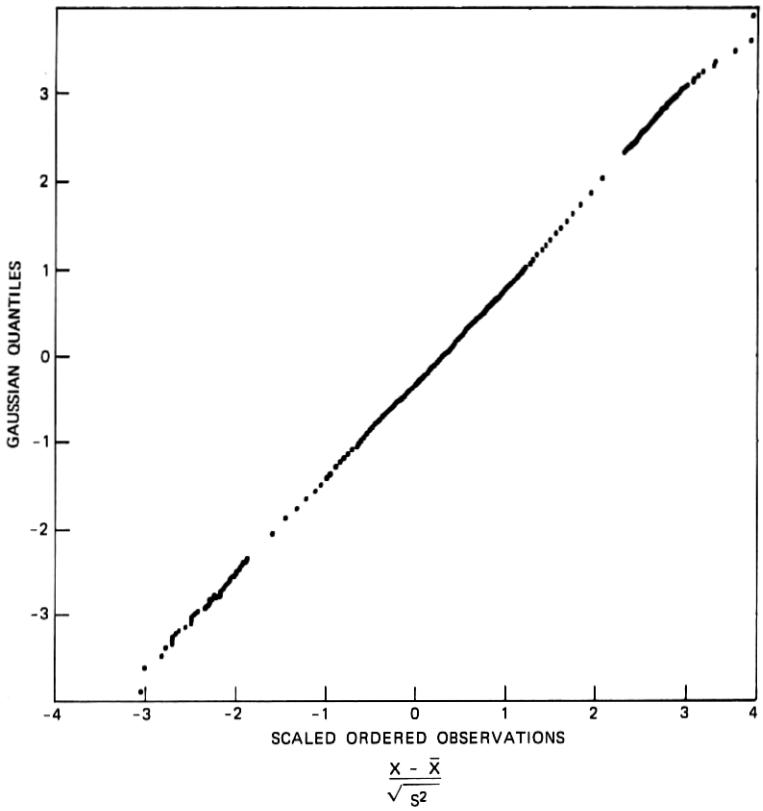


Fig. 16—Q-Q plot for 11,000 line 5 data against gaussian distribution ($\bar{X}$ = sample mean, S^2 = sample variance).

the gaussian distribution is unbounded and unbounded quantities are also felt to be physically inappropriate. The gaussian distribution may describe a particular situation adequately over a certain range; an infinite-variance distribution may model an actual situation over a greater range of a parameter. Both distributions may be physically inappropriate, but the infinite variance may, in this sense, account for the observations better than the gaussian.

Mandelbrot has pioneered in developing and popularizing this notion,[36–38] for example, in connection with economic phenomena[39–42] and in error statistics of digital signals transmitted over telephone lines.[9] Consider, for example, the distribution of changes in stock market prices. Mandelbrot[39] and Fama[43] have shown that, although

the change in stock market prices is bounded, the probability of very large deviations is so great that many statistical techniques that assume an underlying distribution with finite variance are not applicable. Stock market prices may be modeled as a sum of a large number of random variables; similarly, at any instant of time, telephone noise is presumably the sum of a large number of random variables. The sum of a large number of infinite-variance variables is often dominated by one or a few of the summands[44]—a theoretical property of infinite-variance distributions. The key feature common to these models is that the limiting distribution remains the same if an arbitrary but finite number of terms are dropped from the sum. This intuitive notion can be made precise and, subject to a mild restriction on the distribution from which the summands are drawn, leads naturally to the central limit theorem.[45]

Among infinite-variance distributions, the stable distributions play an important role, because only stable distributions can be limiting distributions of suitably normalized sums of independent identically distributed random variables, as well as because stable distributions are closed under convolution. Some pioneering work on the statistical analysis of data from a stable distribution has been carried out already; the analysis described here is a straightforward application of this work.[46–48] Before detailing that work, a summary is presented of some properties of stable distributions.

A distribution function $P(x)$ is called *stable* if, for all $a_1 > 0$, b_1, $a_2 > 0$, b_2, there exist constants $a > 0$, b such that

$$P(a_1x + b_1)*P(a_2x + b_2) = P(ax + b)$$

holds.[49] Every stable distribution has a continuous density; the stable distributions discussed in this work have unimodal densities that are analytic throughout their support.[50] The random variable x is stable if and only if the logarithm of its characteristic function is[51]

$$\ln [E(e^{ixw})] \triangleq \ln [\varphi_x(w)]$$
$$= \begin{cases} -|cw|^{\alpha}[1 - i\beta \cdot \text{sign } (w) \tan (\pi\alpha/2)] + i\delta w & \alpha \neq 1 \\ -|cw|[1 - i\beta 2/\pi \text{ sign } (w) \ln |cw|] + i\delta w & \alpha = 1 \end{cases}$$
$$-1 \leqq \beta \leqq 1 \qquad 0 < \alpha \leqq 2.$$

Thus, every stable law is described by four parameters α, β, c (or $\gamma = c^{\alpha}$), δ, where α is the characteristic index, β is associated with the skewness of the distribution, c is a scale parameter, and δ is a location parameter. If $\beta = 0$, the distribution is symmetric about $x = \delta$. If

$\beta > 0$ and $0 < \alpha < 2$, the distribution is skewed to the right, and the degree of skewness increases as β increases; conversely, if $\beta < 0$, the distribution is skewed to the left and the degree of skewness increases as β decreases. For $\alpha = 2$, β is irrelevant.[52]

If s_n is the suitably normalized sum of n independent identically distributed random variables $x_1, x_2, \cdots, x_n$,

$$s_n = \frac{1}{B_n}(x_1 + x_2 + \cdots + x_n) - A_n$$

where B_n and A_n are normalizing constants, then the distribution of x is said to belong to the domain of attraction of a stable distribution with characteristic index α if the distribution of s_n converges to this stable law as n goes to infinity;[53] this distribution belongs to the domain of partial attraction of a stable distribution if the distribution of s_n converges only for some subsequence.[54] A stable distribution with index α has absolute moments of all orders strictly less than α, i.e., $E[|x|^p] < \infty$ for $0 \leqq p < \alpha$.[55]

Stable distributions and densities can be expressed as a power series.[56] In several cases, this series can be considerably simplified to yield analytic closed-form expressions; these cases are $\alpha = 2$ (gaussian), $\alpha = 1$ and $\beta = 0$ (Cauchy), and $\alpha = \frac{1}{2}$ with $\beta = \pm 1$. Figure 17 depicts several stable density functions.

If $x_1, x_2, \cdots, x_n, \cdots$, are independent random variables drawn from r distributions, each within the domain of attraction of stable laws with indices drawn from the finite set $(\alpha_1, \cdots, \alpha_r)$, then under certain conditions on the number of representatives of each distribution, the suitably normalized sum of these variables converges to a distribution that is the convolution of r stable distributions.[57,58]

The question of rate of approach to the limiting distribution of a sum of suitably normalized, independent, identically distributed random variables is well understood if the limiting distribution is gaussian ($\alpha = 2$).[59–61] If the limiting distribution is in the domain of attraction of a stable distribution, a variety of results are available.[30,35,43,47] The most useful result[62] available at present, from a data analysis point of view (see Ref. 63), loosely states that the difference between the actual distribution of the sum of N suitably normalized random variables and the limiting stable distribution ($0 < \alpha < 2$) is bounded by a linear combination of terms of the order of $N^{-1/\alpha}$ and $N^{-(2-\alpha)/\alpha}$. As an example, consider the case $\alpha = 1.9$: one term is $N^{-1/\alpha} = N^{-0.53}$ while the other term is $N^{-(2-\alpha)/\alpha} = N^{-0.053}$; N must be

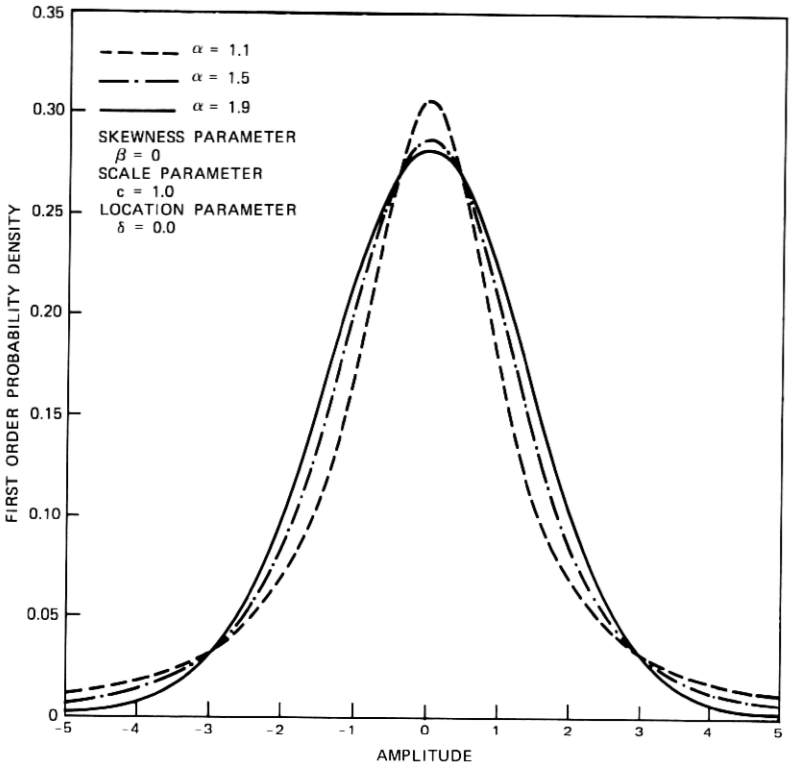


Fig. 17—Some symmetric stable first-order probability density function.

astronomically big to reduce this second term to a value smaller than 0.1, which indicates how slow this rate of convergence to a limiting stable distribution may be.[63] Thus, in many practical situations, caution must be shown in going to the limiting distribution.[64–69] Noise on telephone lines is possibly a case in point.

Section 3.2 discusses how filtered data from the three nongaussian telephone lines are fit to stable distributions. Since these distributions have no second moments, the modifications necessary to properly interpret power spectra and covariance estimates, as well as auto- and crosscorrelation estimates, for these three lines are not clear. This whole area must be subject to further research.[38]

As a final aside, the question of ergodicity, of relating time average statistics to ensemble average statistics, will not be addressed here: The filtered data are assumed to be an ergodic random process.

3.2 Symmetric stable distribution model

In this section, various statistical tests are described for determining if telephone noise on the three lines that appear nongaussian can be adequately modeled by a symmetric stable distribution ($0 < \alpha \leqq 2$, $\beta = 0$).

A series of estimators for symmetric stable distribution ($1 < \alpha < 2$) parameters have recently been developed.[47,48] These estimators are based on statistics easily derived from the empirical distribution function; they have been compared with maximum likelihood estimates and found to offer reasonable agreement when suitable precautions, such as a large sample size for α near 2, are taken.[46] These parameter estimates are

$$\hat{\delta} = \frac{1}{2Np} \sum_{K=(0.5-p)}^{(0.5+p)} \tilde{x}_K \qquad p = 0.125, 0.250, 0.375$$

$$\hat{c} = \frac{1}{1.654} (\tilde{x}_{0.72} - \tilde{x}_{0.28}),$$

where $\tilde{x}_r$ is the value of the rth empirical quantile, $\hat{\delta}$ is a trimmed mean, and $\hat{c}$ measures the spread of the distribution. To estimate the characteristic index α, an auxiliary variable z_q is first computed

$$z_q = \frac{\tilde{x}_q - \tilde{x}_{1-q}}{2\hat{c}} = 0.827 \frac{\tilde{x}_q - \tilde{x}_{1-q}}{\tilde{x}_{0.72} - \tilde{x}_{0.28}}$$

$$q = 0.9995, 0.995, 0.99, 0.985, 0.98, 0.975, 0.97, 0.96, 0.95, 0.94, 0.92,$$

and then $\hat{\alpha}$ is obtained as a function of z_q from tables in Ref. 48. $\hat{\alpha}$ is a measure of how rapidly the distribution approaches its asymptotic values. For line 1, with a sample size of 13,000, it was found that

$$\hat{\delta} = -87.4,$$
$$\hat{c} = 132.0,$$
$$\hat{\alpha}(z_{0.99}) = \hat{\alpha}(z_{0.995}) = 1.95.$$

In addition, this was carried out for a sample size of 1000 thirteen times, a sample size of 2000 six times, a sample size of 3000 four times, a sample size of 4000 three times, and a sample size of 5000 two times. The results are tabulated in Table IV for the $q = 0.98$ fractile. Different choices of q resulted in practically the same estimates.

Larger and larger samples were used because, if the data really come from a stable distribution, then the parameter estimates would presumably converge to their true values with increasing N.

Table IV — Line 1 symmetric stable distribution parameter estimates

	$N = 1000$	$N = 2000$	$N = 3000$	$N = 4000$	$N = 5000$
$\hat{\delta}$	−88.6, −88.8, −86.9, −87.2 −88.2, −88.0, −87.6, −87.6 −88.3, −88.7, −86.6, −87.5 −87.2	−87.0, −88.1 −87.6, −88.5 −87.7, −88.7	−88.1, −87.8 −87.8, −88.0	−87.9, −87.8 −88.1	−87.9, −88.0
$\hat{c}$	132.1, 127.4, 122.9, 124.7 145.2, 134.5, 128.6, 119.8 143.4, 134.5, 130.5, 124.2 137.6	123.7, 141.1 124.0, 137.6 135.1, 128.7	125.8, 135.7 129.1, 135.1	125.5, 132.3 135.9	130.6, 132.1
α	1.92, 1.94, 1.96, 1.90 1.90, 1.89, 2.00 1.90, 1.99, 2.00, 2.00 1.91, 1.98	1.93, 1.94, 1.92 1.92, 2.00, 1.93	1.94, 1.92, 1.93 1.97	1.93, 1.91 1.97	1.94, 1.96

Figure 18 shows a Q-Q plot of 13,000 line 1 data against a symmetric stable distribution with $\alpha = 1.94$, while Fig. 21 shows the same plot with 10 points on either end excluded. These points were excluded because they were possibly atypical observations, and because they were highly correlated to one another. Again, as in the gaussian Q-Q plots, only the first and last 100 empirical quantiles, as well as every one hundredth between have been plotted, giving the false illusion of discontinuity in the observations. The eye is quite sensitive to deviations from a straight line for quantile-quantile plots; in particular, $\alpha = 1.94$, 1.95, 1.96 could easily be distinguished from one another (Figs. 18 to 23). The data appear to be slightly skewed, so a non-symmetric ($|\beta| \ll 1$, $\beta \neq 0$) stable distribution might indeed provide a better fit to the data. As α increases from 1.94 to 1.96, the stable distribution has shorter and shorter tails, and the points in the tails bending towards the vertical for $\alpha = 1.94$ align with the rest of the data for increasing α.

As a check on these estimates, W. DuMouchel has supplied the authors with a computer program that numerically calculates maximum likelihood estimates of parameters of stable distributions, as well as of their covariances.[46] DuMouchel has shown the maximum likelihood parameter estimates are asymptotically normal, so that some statistical techniques developed for data analysis of gaussian samples can be brought to bear.[70] The method used is numerical, nonetheless, so two possible pitfalls must be kept in mind.

(*i*) For ease in numerical calculations, the data were aggregated into bins, thereby losing information.

(*ii*) A Newton-Raphson-type of algorithm was used that approximates the first and second derivatives of the likelihood function with differences.

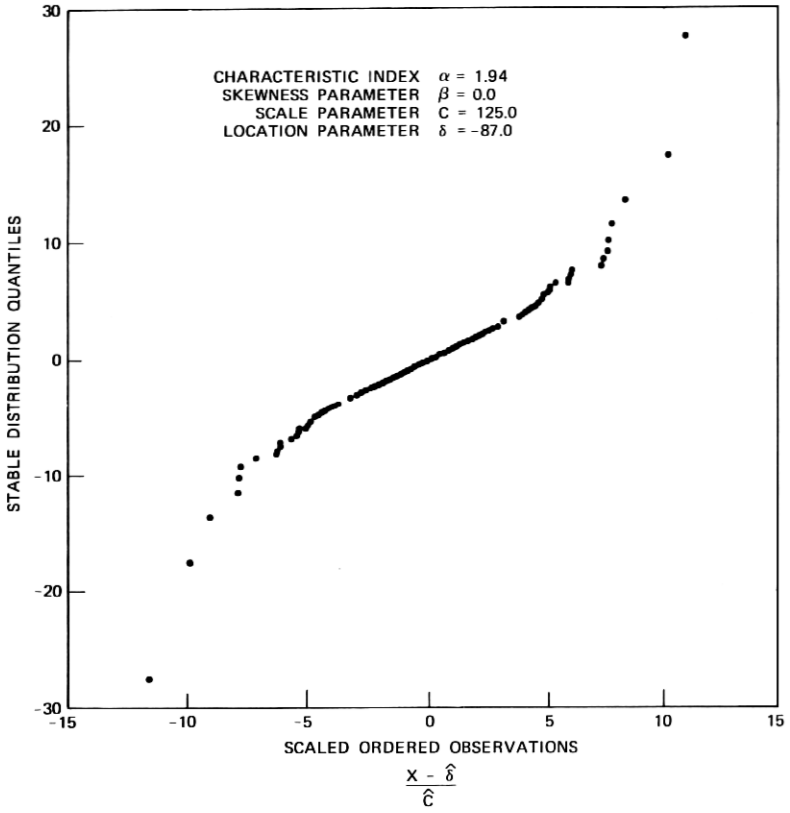


Fig. 18—Line 1 Q-Q plot against a stable distribution, $\alpha = 1.94$ ($N = 13{,}000$).

DuMouchel[71] has observed that the first approximation is the more critical of the two. The second approximation was investigated using a simplex algorithm rather than Newton-Raphson which did not compute discrete approximations to derivatives, with results consistent to those now described.

For line 1 data, with a sample size of 13,000, the numerical maximum likelihood stable distribution parameter estimates were

$$\hat{\alpha} = 1.95$$
$$\hat{\beta} = -0.006$$
$$\hat{c} = 132.7$$
$$\hat{\delta} = -88.8.$$

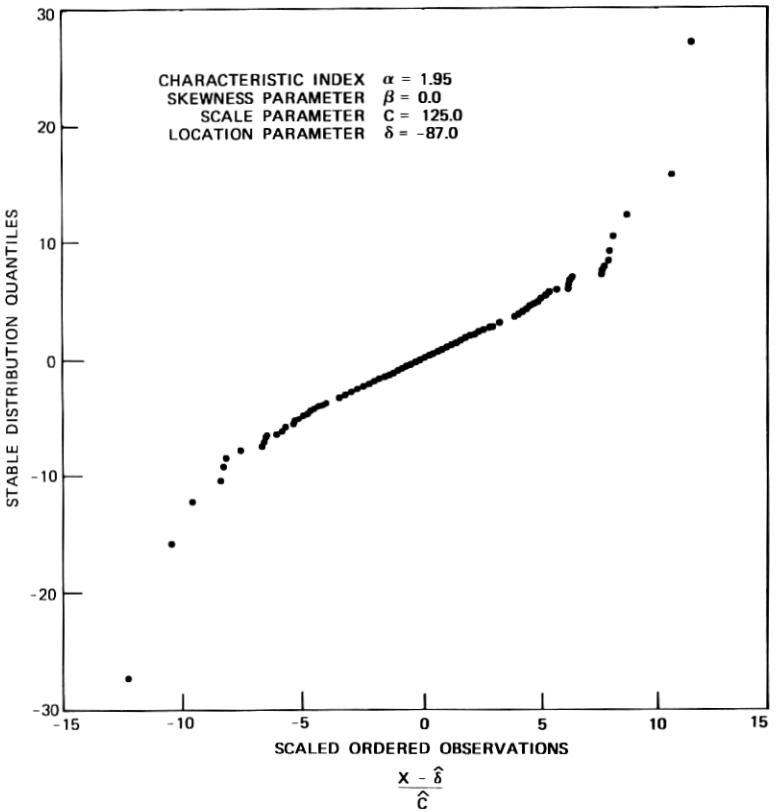


Fig. 19—Line 1 Q-Q plot against a stable distribution, $\alpha = 1.95$ ($N = 13{,}000$).

The numerical approximation to the estimated parameter covariances are shown in Table V.

The large variance of $\hat{\beta}$ compared to the other estimates has been observed by DuMouchel;[46] the cause is unknown. Although Q-Q plots

Table V — Parameter estimate covariances ($\times 10^6$)

	$\hat{\alpha}$	$\hat{c}$	$\hat{\beta}$	$\hat{\delta}$
$\hat{\alpha}$	410	11	−57	−1
$\hat{c}$	11	63	−41	−2
$\hat{\beta}$	−57	−41	13,551	430
$\hat{\delta}$	−1	−2	430	248

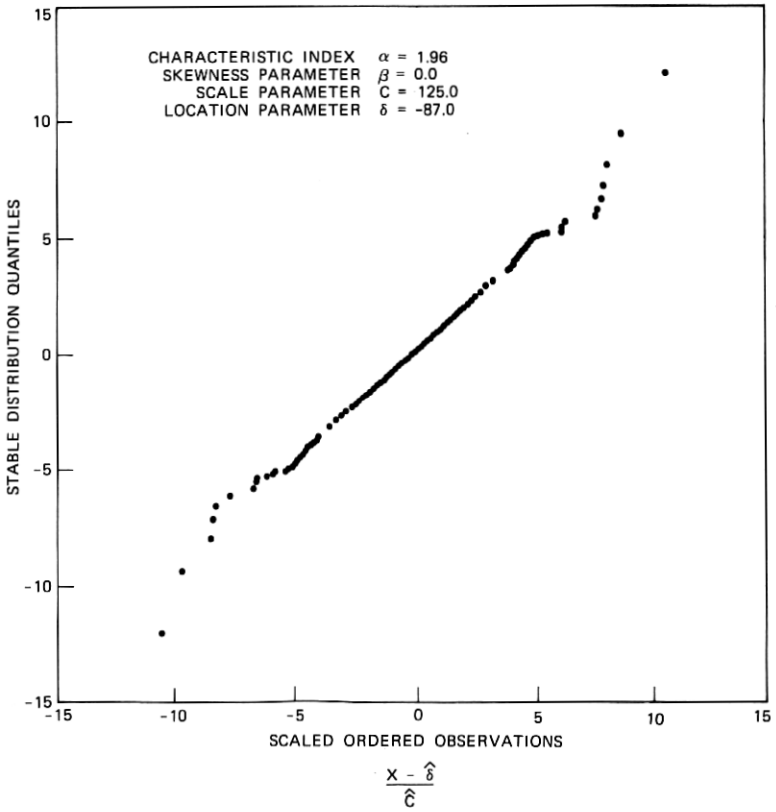


Fig. 20—Line 1 Q-Q plot against a stable distribution, $\alpha = 1.96$ ($N = 13{,}000$).

indicated a slight skewness, i.e., $\beta < 0$ and $|\beta| \ll 1$, the interpretation of the maximum likelihood estimate for β was obscured by this large variance.

As a check on these results, maximum likelihood parameters of stable distributions were estimated for 78,750 filtered data from lines 1 and 2, corresponding to approximately 10 seconds of telephone noise. A Newton-Raphson-type algorithm was used; the parameter estimate covariances were comparable to those just discussed. The results were:

	$\hat{\alpha}$	$\hat{\beta}$	$\hat{c}$	$\hat{\delta}$
Line 1	1.96	−0.0084	218.0	−81.35
Line 2	1.94	−0.0014	93.5	−80.30

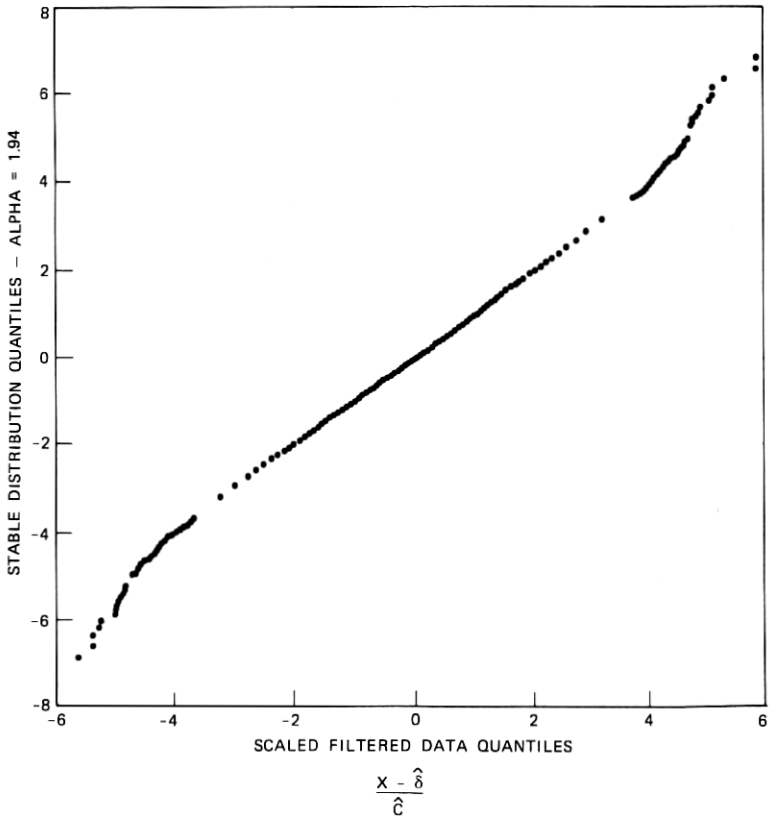


Fig. 21—Center portion of line 1 Q-Q plot against a stable distribution, $\alpha = 1.94$ ($N = 12{,}980$).

Two other pieces of evidence that telephone noise can be fitted by a stable distribution are now presented: the studentized range test and the likelihood ratio test. These have been discussed elsewhere;[47,71] for large amounts of data, caution is necessary to interpret the results of these tests properly. On the other hand, since the data here are apparently close to gaussian, large amounts of data must be examined to make clear the nongaussian nature of the noise. Thus, the results of these tests must be very carefully interpreted, and are included for the sake of completeness.

For line 1 data, testing the gaussian hypothesis at a 0.5-percent significance level via the studentized range test for sample sizes of 1000 led to mixed results: some segments of data fell within these

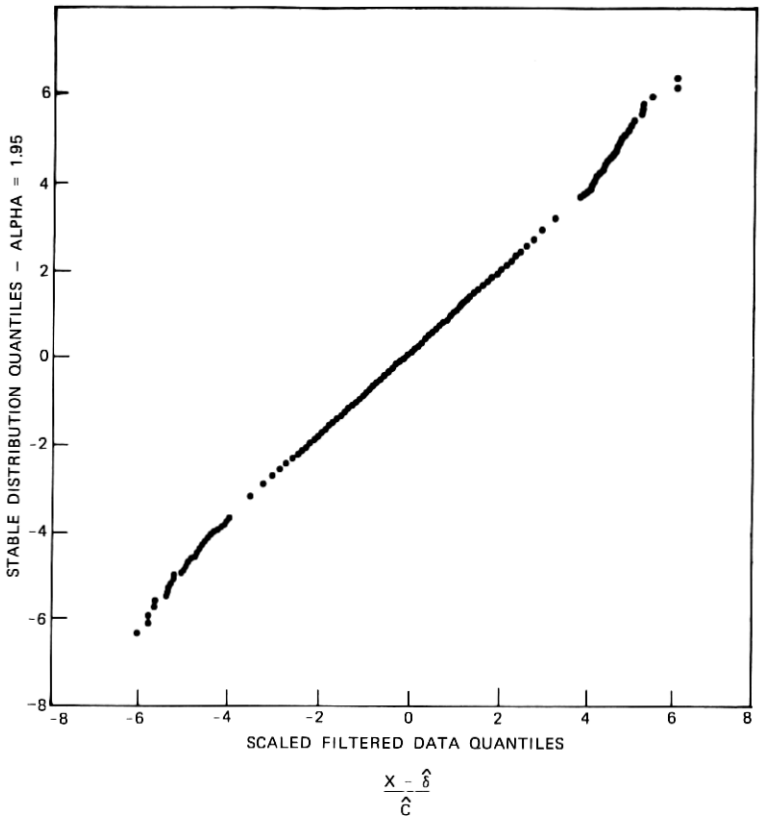


Fig. 22—Center portion of line 1 Q-Q plot against a stable distribution, $\alpha = 1.95$ ($N = 12{,}980$).

limits, others fell outside. However, for a sample size of 10,000 the result of the studentized range test clearly fell outside the confidence intervals.

A likelihood ratio test was used to test 10,000 line 1 data at 1-percent significance levels against five hypotheses. The stable distribution hypothesis was rejected for $\alpha = 2.00, \alpha = 1.98, \alpha = 1.90$, and $\alpha = 1.85$, but could not be rejected for $\alpha = 1.95$.

3.3 Other central-limit-theorem-based models

What other models arise that might adequately account for the data, while having the same central-limit-theorem-based appeal as the stable distributions? First, it is possible the data examined lie in a

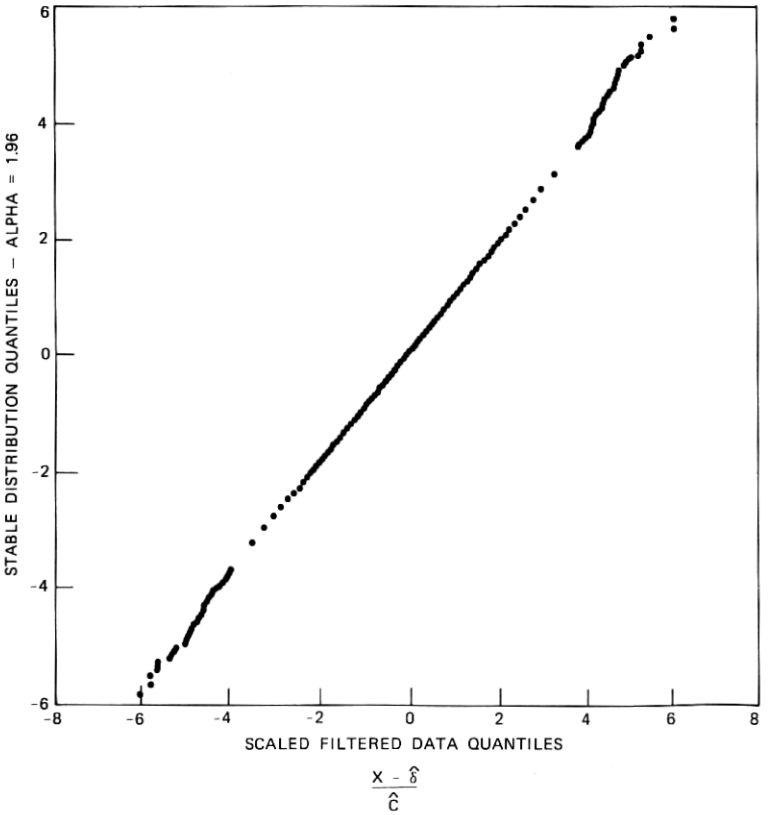


Fig. 23—Center portion of line 1 Q-Q plot against a stable distribution, $\alpha = 1.96$ ($N = 12{,}980$).

domain of partial attraction of a stable distribution (which is wider than the domain of attraction[72]). If enough data were examined, it might be possible that α would approach 2. A second possibility is to model the data as a convolution of r stable distributions, each with its own domain of attraction;[73] presumably, each distribution could be attributed to a separate physical process. A third possibility is that the data are drawn randomly from m gaussian distributions, each with different mean and variance; for example, the data could be drawn from a low-variance gaussian a fraction P of the time, and from a high-variance gaussian a fraction $(1 - P)$ of the time. A fourth possibility is to model the data as a nonstationary gaussian random process, which is a special case of a nonstationary stable random

process, or by a doubly stochastic gaussian random process (where the mean and the variance are themselves random processes), which is a special case of a doubly stochastic stable random process (where all four parameters are themselves random processes). While these nonstationary and doubly stochastic models do not appear to be necessary to adequately model the data discussed here, over longer time intervals, such as days, weeks, months, or years, the simple models might be inadequate while these more complicated models might be more appropriate. Presumably, other classes of models exist.

It is difficult to refute these alternative models offhand. Recall that the original goal was to find a mathematically tractable model for telephone noise; the model discussed here is simple and agrees intuitively with the physics of the noise. The other models are more complicated. To be of practical use, however, they must be so oversimplified that the intuitive agreement with the physics of the noise is lost. It is hoped that the class of models based on stable distributions will lead to more insight into how telephone noise limits voice communication and data transmission and, more important, into new ways for combating this noise.

3.4 Gaussian-plus-filtered-Poisson-process model

A model involving more parameters than the previous one is now investigated. This model assumes that telephone noise is due to a sum of two independent random processes. The low-variance part is assumed to be white and gaussian, while the high-variance process is assumed to be a filtered Poisson process. This type of model was popularized by Snyder,[74] and has been used in optical communication[75,76] and ELF communication[77] to assess theoretically optimum and suboptimum receiver structures. It has intuitive physical appeal: for instance, the low-variance component can be attributed to thermal noise and electromagnetic crosstalk, while the high-variance component can be attributed to switch arcing and thunderstorms. It is convenient to view the filtered Poisson process as the output of a linear dynamical system, whose input is an impulse train; the area of the impulse is

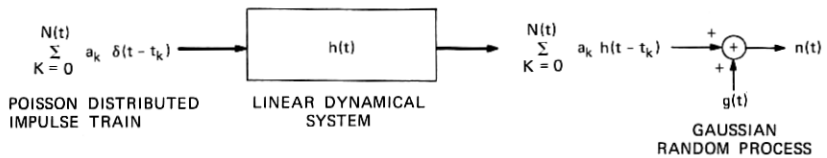


Fig. 24—Block diagram for generating a gaussian-plus-filtered-Poisson process.

assumed to be a random variable, while the instants of time at which the impulses occur follow a Poisson distribution with rate or intensity parameter λ. Figure 24 and the following equation summarize this discussion.

$$n(t) = \begin{cases} g(t) + \sum_{K=1}^{N(t)} a_K h(t - t_K) & N(t) > 0 \\ g(t) & N(t) = 0, \end{cases}$$

where

$n(t)$ = gaussian-plus-filtered-Poisson process,
$g(t)$ = stationary gaussian random process,
a_K = area of Kth impulse,
$h(t)$ = impulse response of linear dynamical system,
t_K = time at which the Kth impulse occurs,
$N(t)$ = number of impulses that occur in $[0, t)$.

To completely describe this model, the following parameters must be estimated

(*i*) The mean and variance of the gaussian random process.
(*ii*) The probability density function for the impulse areas.
(*iii*) The Poisson process rate parameter λ.
(*iv*) The linear system structure.

We recall that the original motivation for this work was to stimulate interest of communication theorists in receiver structures that detect or estimate signals corrupted by nongaussian noise. One advantage of this type model is that parameters can be related to receiver performance limitations as well as to physical causes of noise. This helps in determining how much effort should go into improving the receiver as opposed to reducing the noise (e.g., by designing switches to operate at lower voltages). One disadvantage of this type of model is its great analytical complexity; it may be quite difficult to find analytic performance limitations, and to determine how sensitive these limitations are to model parameters.[75]

If the impulse areas $\{a_K\}$ are assumed to be independent identically distributed random variables that are independent of the times the impulses occur, the characteristic function for the first-order probability density function can be shown to be

$$E[e^{i\omega n(t)}] = \exp\left\{im\omega - \tfrac{1}{2}\sigma^2\omega^2 + \lambda \int_0^t [E_a(e^{i\omega a h(\tau)}) - 1]d\tau\right\},$$

where

ω = frequency,
m = mean of gaussian random process,
σ^2 = variance of gaussian random process,
$E_a(\ )$ = expectation of ( ) with respect to the random variable a, the impulse area.

It is quite difficult to analytically invert $E(e^{i\omega n(t)})$ to find the probability density for $n(t)$. This in turn means maximum likelihood parameter estimates, and Cramér-Rao lower bounds on parameter estimate covariances are difficult to calculate analytically. For this reason, numerical approximations must often be used. To avoid these problems, a suboptimum parameter estimation method was developed: Each parameter of the model was estimated by itself. There is no guarantee that these estimates, when put together, will be close to the true parameter values. The sole reason for doing this was to make the problem tractable. Evidence presented later indicates this method provides an excellent (but perhaps suboptimum) fit to the data.

Although the dynamics of the linear system can be quite complicated, only three simple cases are considered here.

$$(i)\ h(t) = Ae^{-At}u_{-1}(t)$$

$$(ii)\ h(t) = \left(\frac{A^2+\omega^2}{A}\right)e^{-At}\cos\omega t\, u_{-1}(t) \quad u_{-1}(t) = \begin{cases} 1 & t > 0 \\ 0 & t < 0 \end{cases}$$

$$(iii)\ h(t) = \left(\frac{A^2+\omega^2}{\omega}\right)e^{-At}\sin\omega t\, u_{-1}(t),$$

which are perhaps the cases of greatest practical interest.[77]

Assuming the amplitude burst statistics to be independent of the instants of time at which bursts occur, and assuming the gaussian process to be independent of the filtered Poisson process, the mean and variance can be calculated (Table VI) for the steady state noise. $E(a)$ is assumed to be zero in all models presented here. This completes a general discussion of the gaussian-plus-filtered-Poisson-process model; the methods used to estimate the model parameters are now described in detail.

3.5 Gaussian random process parameter estimation

If $E(a) = 0$, then $E[n(t)] = m$, and the sample mean is an unbiased estimate of the true mean of the gaussian process. If the data are

Table VI — Mean and variance for $n(t)$

$h(t)$	$E[n(t)]$	Variance $[n(t)]$
$Ae^{-At}u_{-1}(t)$	$m + \lambda E(a)$	$\sigma^2 + \frac{\lambda A}{2} E(a^2)$
$\frac{A^2+\omega^2}{A} e^{-At} \cos \omega t \, u_{-1}(t)$	$m + \lambda E(a)$	$\sigma^2 + \frac{\lambda}{4}\left(\frac{A^2+\omega^2}{A}\right)^2 \left[\frac{1}{A} + \frac{A}{A^2+\omega^2}\right] E(a^2)$
$\frac{A^2+\omega^2}{\omega} e^{-At} \sin \omega t \, u_{-1}(t)$	$m + \lambda E(a)$	$\sigma^2 + \frac{\lambda}{4}\left(\frac{A^2+\omega^2}{\omega}\right)^2 \left[\frac{1}{A} - \frac{A}{A^2+\omega^2}\right] E(a^2)$

trimmed to exclude a fraction (e.g., 25 percent) of the data with largest absolute deviation from the sample mean, then presumably most values of $n(t)$ that contain large contributions from the filtered Poisson process will be excluded.

The estimates for the mean and variance of the gaussian process were consistent with estimates to be presented later for λ, A, ω, and $E(a^2)$. No bounds are available on the bias or variance of these parameter estimates. The results are summarized in Table VII. The sample variance has been rescaled, based on the assumption that the data were drawn from a truncated gaussian distribution.

3.6 Poisson process parameter estimation

The Poisson process intensity is closely related to the times at which bursts of high-amplitude telephone noise occur. Many definitions of a noise burst are possible. The definition chosen here, although arbitrary, was found to be qualitatively insensitive to the parameters defining a burst. The absolute value of a zero mean waveform is shown

Table VII — Gaussian random process trimmed mean and variance (Total data = 10,000, with a fraction p trimmed from either side)

Line	Truncated Sample Mean			Rescaled Truncated Sample Variance		
	$p = 0.125$	0.250	0.375	$p = 0.125$	0.250	0.375
1	−88.0	−87.6	−87.3	35,414	35,813	35,208
2	−79.5	−78.8	−78.7	17,151	17,058	16,815
3	−77.6	−75.2	−73.6	49,306	48,840	46,554
4	−81.1	−80.9	−81.0	83,076	88,430	90,864
5	− 1.0	− 1.0	− 1.1	1,664	1,672	1,683

in Fig. 25. The duration of the burst is the time difference between the moment that the absolute value of the waveform climbs above an upper threshold, T_{upper}, and the time that the absolute value of the waveform drops below a lower threshold, T_{lower}, provided the waveform stays below the lower threshold for at least a predetermined period of time, called the guard band G_B, which separates one burst from the next. The instant of time a burst occurs, $t_{\max}$, is the first instant at which the absolute value of the burst attains its maximum value, $P_{\max}$.

A large number of statistics can characterize a point process. A number of nonparametric statistics were used to characterize the points in time at which bursts occur, and then, based on this evidence, the burst data were examined in greater detail to see if they could be adequately modeled by a renewal process in general, or a Poisson process in particular (e.g., see Ref. 78).

The two statistics that were first examined were

(*i*) The sample mean time between bursts as a function of T_{upper}, T_{lower}, and G_B.
(*ii*) The empirical cumulative distribution function and the histogram for time intervals between events as a function of T_{upper}, T_{lower}, and G_B.

The effect on these statistics of variations of T_{upper}, T_{lower}, and G_B is now discussed. For line 1 data, for example, T_{lower} was fixed at 600 (roughly three standard deviations from the sample mean), T_{upper} was set at 600, and G_B was varied from 0.1 to 0.9 millisecond, in steps of 0.2 millisecond. T_{upper} was then set at 800, and G_B was varied in an identical manner. Finally, T_{upper} was set at 1000, and G_B was again varied in the same fashion. The number of events observed was found to be insensitive to the choice of guard band as well as to T_{upper}. The guard band was therefore set at 0.5 millisecond, and T_{upper} was set equal to T_{lower} (which also avoids ambiguity in the meaning of threshold).

A typical empirical cumulative distribution function and a histogram for the time intervals between bursts are shown in Figs. 26 and 27, for $T_{\text{upper}} = T_{\text{lower}} = 800$. Typical histograms and empirical cumulative distribution functions for time intervals between bursts for $T_{\text{upper}} = T_{\text{lower}} = 600$ and $T_{\text{upper}} = T_{\text{lower}} = 1000$ had the same shapes as those in Figs. 26 and 27. If the bursts were Poisson-distributed, the distribution function would be completely specified by this information.

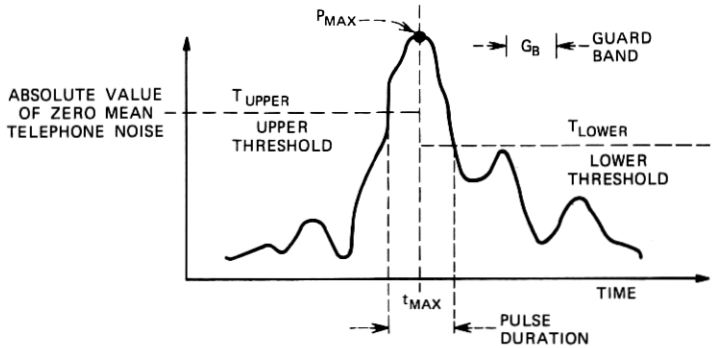


Fig. 25—Definition of burst parameters.

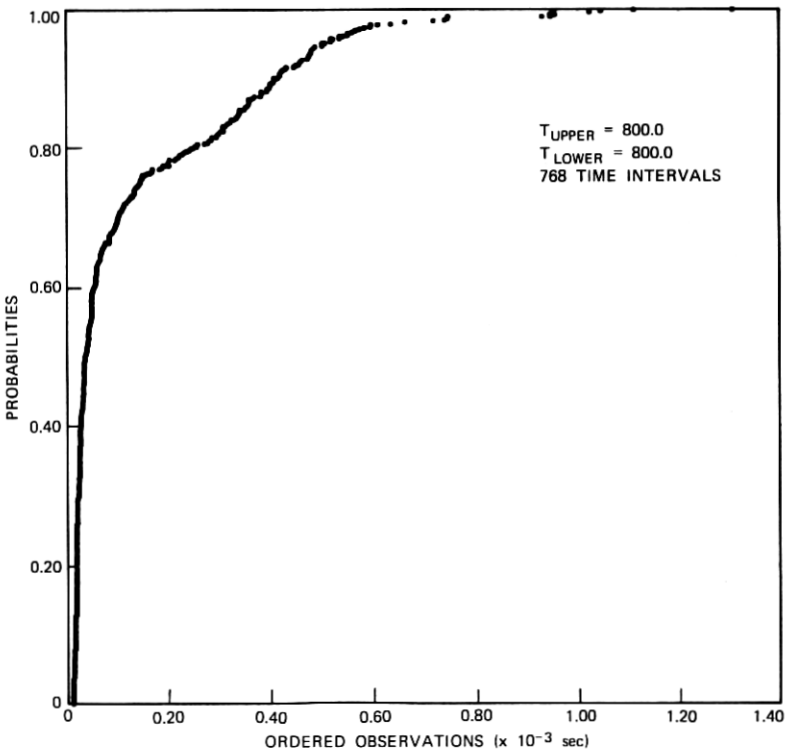


Fig. 26—Empirical cumulative distribution function for line 1 time intervals between bursts (N = 768).

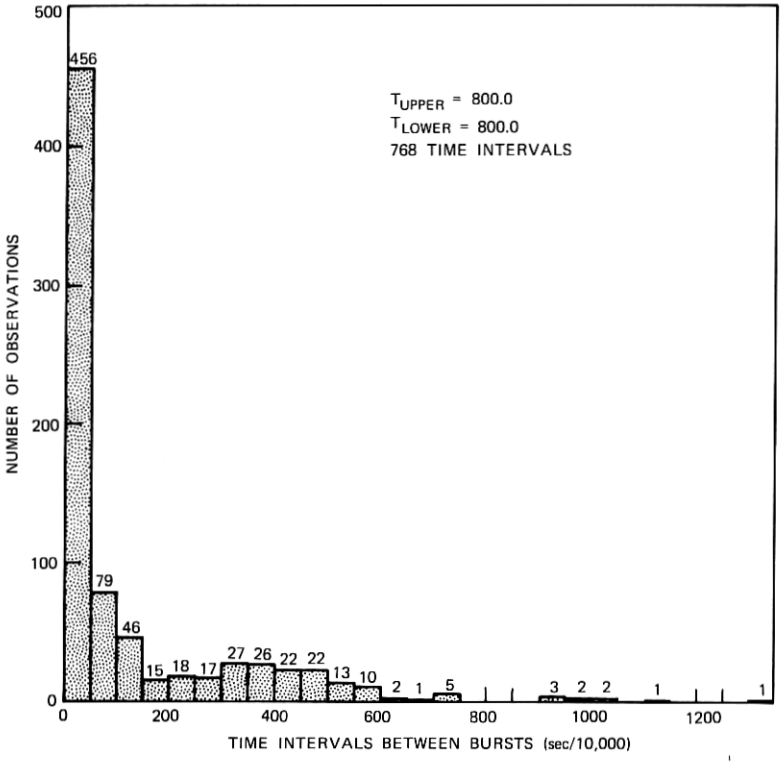


Fig. 27—Histogram for line 1 time intervals between bursts ($N = 768$).

Investigation of the burst statistics of the other two nongaussian lines yielded findings similar to those for line 1.

Next, the second-order statistics of the time intervals were investigated. Figure 28 shows a scatter plot of the $(K + 1)$th interval against the Kth interval. This plot shows that long intervals followed by long intervals are unlikely compared with long intervals followed by short intervals, or short intervals followed by long or short intervals. Note that it is still possible that the intervals are generated by a renewal process with a nonexponential distribution. Note also that approximately half the points plotted fall in the lower left corner square, $0 < \text{time}_K \leqq 100$ and $0 < \text{time}_{K+1} \leqq 100$.

Another set of second-order statistics of interest is

(*i*) The estimated autocorrelation of the time intervals between bursts

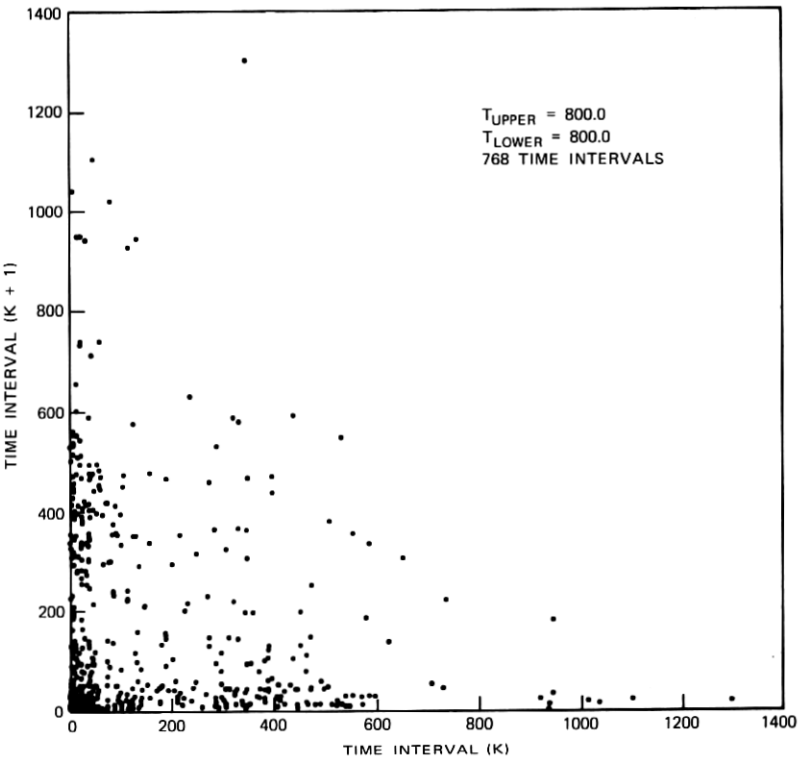


Fig. 28—Scatter plot of time intervals between bursts for line 1 data ($N = 768$).

$$\hat{R}_{tt}(K) = \frac{\frac{1}{N}\sum_{l=1}^{N-K}(t_l - \bar{t})(t_{l+K} - \bar{t})}{\tilde{R}_{tt}(0)}, \quad \bar{t} = \frac{1}{N}\sum_{j=1}^{N} t_j,$$

where

$$\tilde{R}_{tt}(0) = \frac{1}{N}\sum_{K=1}^{N}(t_K - \bar{t})^2$$

and

$$t_j = \text{length of } j\text{th time interval.}$$

(*ii*) The estimated power spectrum of the time intervals between bursts, which is the Fourier transform of $R_{tt}(K)$ if the process is wide-sense stationary and ergodic.[79] The same issues that were discussed earlier when the noise amplitude power spectrum was estimated are relevant here.

(*iii*) Contingency tables,[80] which correspond to estimates of the second-order joint density of intervals that are a specified number of intervals apart from each other.

(*iv*) The estimate of the conditional expectation of the length of an interval given the length of the interval j intervals earlier

$$E[t_{K+j}|t_K] = \int t_{K+j} dp(t_{K+j}|t_K).$$

The data analysis presented here focuses on the first two of these second-order statistics. Since only a small number of events was observed typically (e.g., 147 for $T_{\text{upper}} = T_{\text{lower}} = 1000$, 768 for $T_{\text{upper}} = T_{\text{lower}} = 800$, for line 1 bursts), statistical fluctuations would have obscured the interpretation of the last two statistics. Figure 29 shows a typical sample autocorrelation function for 1000 intervals,

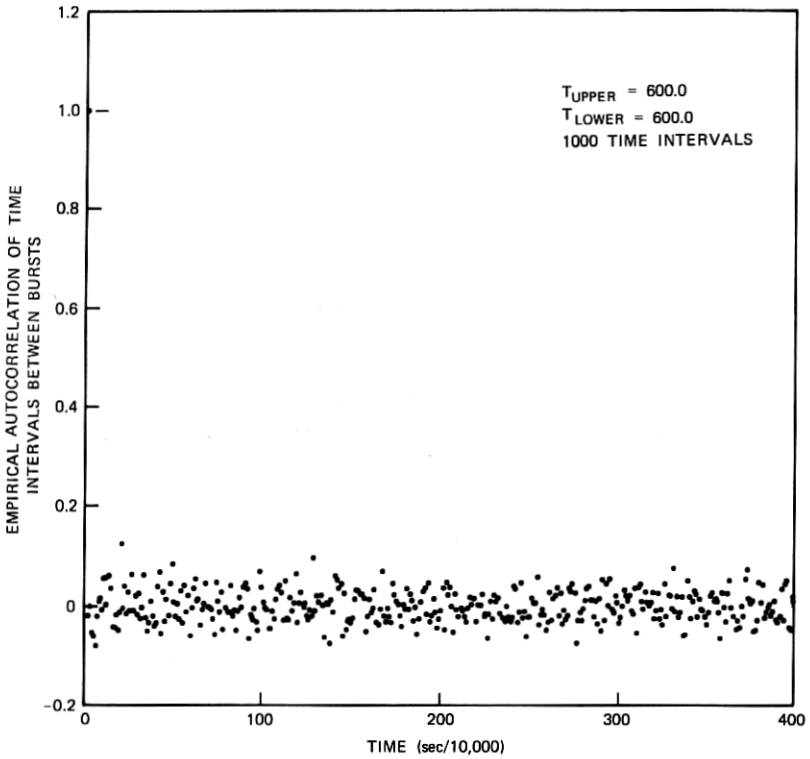


Fig. 29—Autocorrelation function for line 1 time intervals between bursts ($N = 1000$).

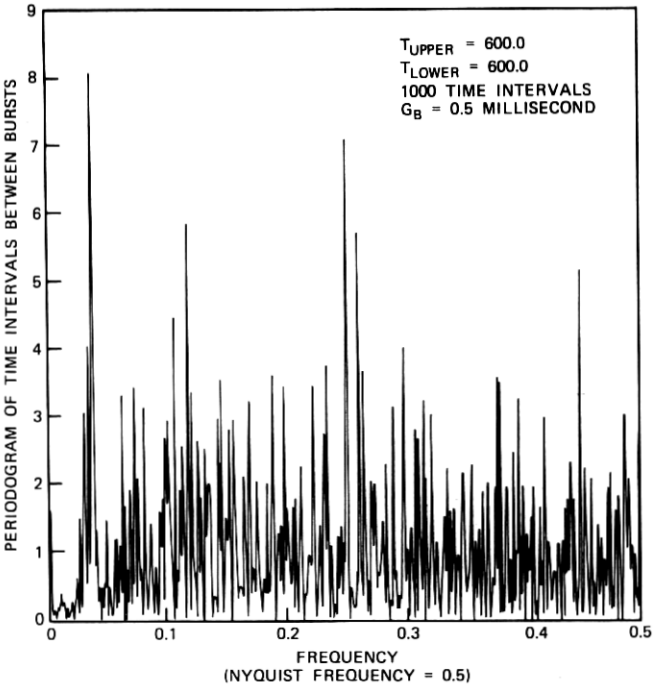


Fig. 30—Periodogram of line 1 time intervals between bursts ($N = 1000$).

while Fig. 30 shows the corresponding periodogram. Based on this evidence, it seems possible the bursts arise from a renewal process. To test this model, the so-called summed empirical periodogram $\tilde{S}(lf_0)$, defined as

$$\tilde{S}(lf_0) = \sum_{K=1}^{l} |\hat{f}(Kf_0)|^2,$$

where $|\hat{f}(Kf_0)|^2$ = periodogram at frequency Kf_0 is plotted in Fig. 31, along with 5 percent significance limits to test the renewal hypothesis, according to which $\tilde{S}$ should be straight. From this and other evidence not presented here, it appears that the bursts analyzed can be reasonably modeled as a renewal process.

Next, it is useful to characterize the renewal process model in greater detail. Such a process can be statistically described by a variety of measures.[81] The two considered here are

(*i*) The survivor function $S(t) \triangleq$ fraction of intervals greater than or equal to t.

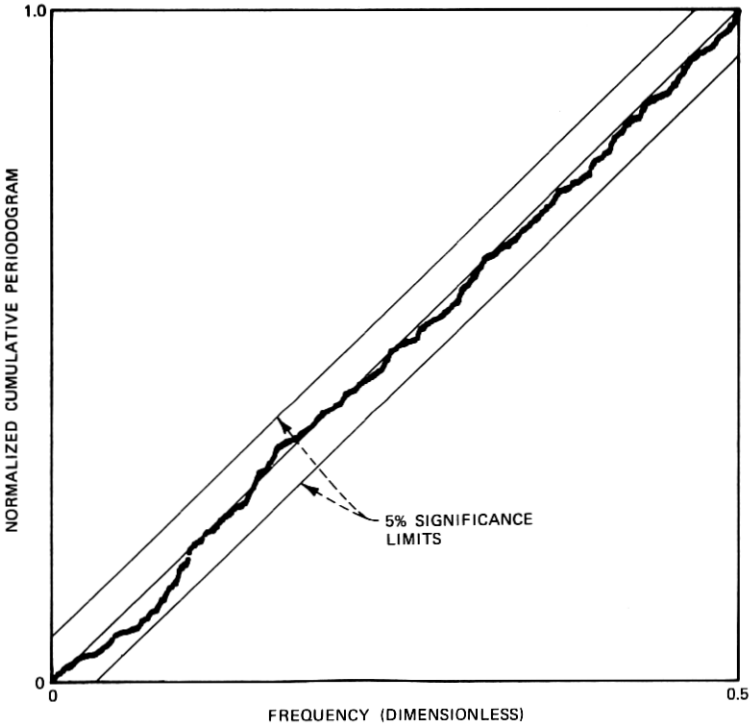


Fig. 31—Normalized cumulative periodogram for line 1 time intervals between bursts, $T_{\text{upper}} = T_{\text{lower}} = 800$ ($N = 768$).

(*ii*) The hazard function $H(t) \triangleq P(t)/S(t)$, where $P(t)$ is the probability density function of the renewal process and $S(t)$ is the survivor function.

Note that $H(t) \cdot \Delta t$ is the probability of an event in an interval of length Δt seconds centered at t, which can be interpreted as the fraction of intervals in the range $(t - \Delta/2, t + \Delta/2)$, given that the last event was t time units ago.

The survivor and hazard functions are related by[82]

$$S(t) = \exp\left[- \int_0^t H(x)dx \right] = \int_t^\infty P(x)dx.$$

In a Poisson process with constant intensity λ, these simplify to

$$\left.\begin{aligned} P(t) &= \lambda e^{-\lambda t} \\ S(t) &= e^{-\lambda t} \\ H(t) &= \lambda \end{aligned}\right\} t > 0.$$

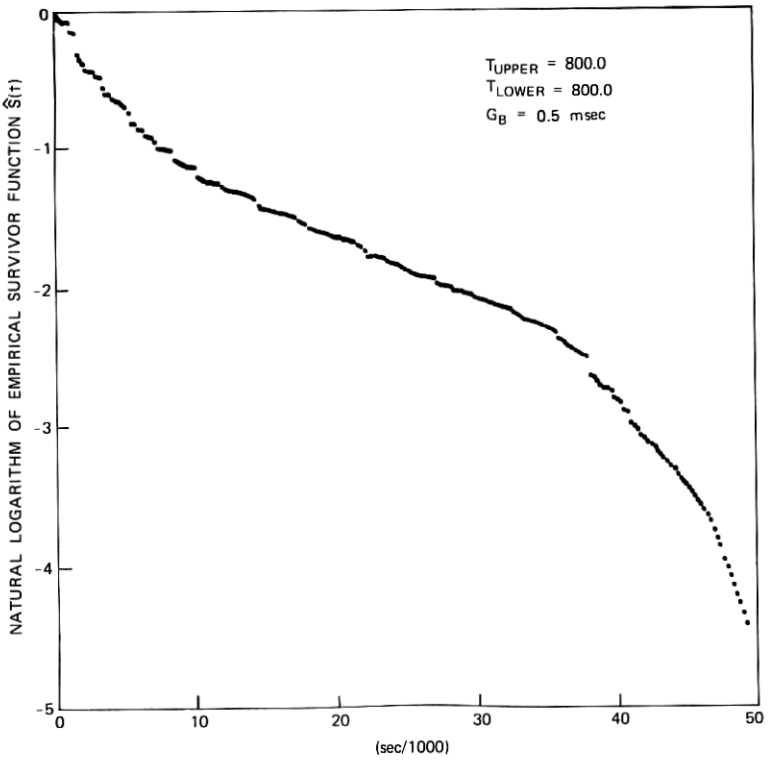


Fig. 32—Natural logarithm of empirical survivor function for line 1 time intervals between bursts ($N = 768$).

In practice, only estimates of the survivor and hazard function are obtained. The empirical survivor function $\hat{S}(t)$ is the fraction of observed intervals greater than or equal to t. The empirical hazard function $\hat{H}(t)$ equals $\hat{P}(t)/\hat{S}(t)$, where $\hat{P}(t)$ is the fraction of observed intervals in the range $[t - \Delta/2,\ t + \Delta/2]$. Figures 32 and 33 show representative empirical survivor and hazard functions for bursts observed on line 1. Statistical fluctuations are quite apparent. For long time intervals, only one or two events fall in any particular bin, giving the appearance of a trend in $\hat{H}(t)$; the survivor function is more stable for long intervals. The log survivor function roughly follows a series of straight-line segments with different slopes, indicating the process can be modeled as a Poisson process whose rate parameter is equal to the absolute value of the slope of the straight line approximations.

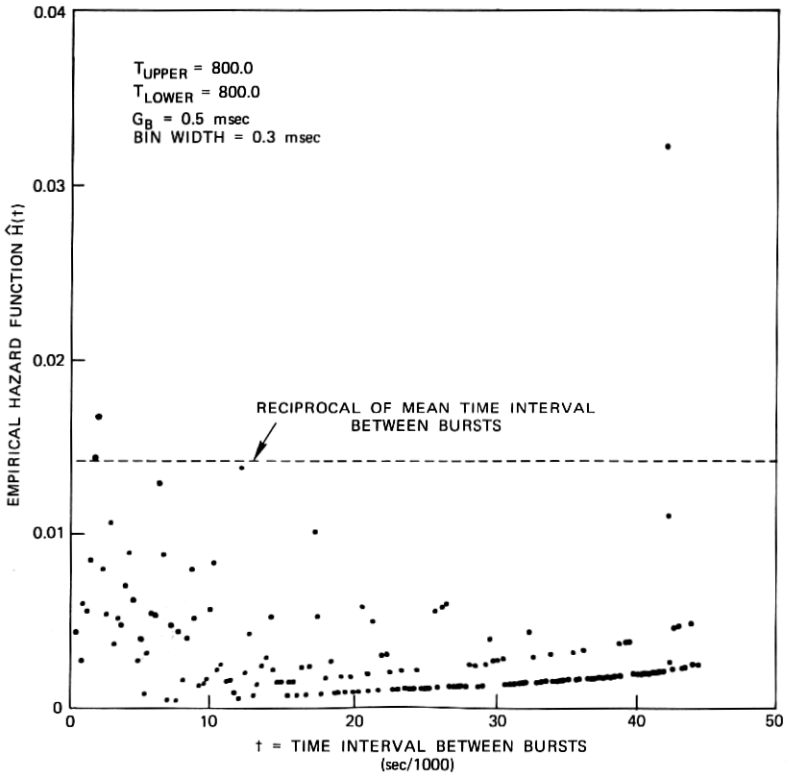


Fig. 33—Empirical hazard function for line 1 time intervals between bursts (N = 768).

In order to adequately model the time intervals between bursts, two models more complicated than a simple Poisson process were investigated. The first model was a pth order autoregressive process, while the second was a doubly stochastic Poisson process, where the Poisson intensity λ was a random variable specified by a two-state Markov process (see Ref. 8). The autoregressive model (for p = 5 and 15) did not adequately account for long time intervals between events. The doubly stochastic Poisson process model did not adequately account for short time intervals followed by long time intervals. Therefore, both these models were dropped in favor of a Poisson process model with constant intensity, even though the log survivor function could not be approximated by a single straight line. Since this is only one of at least six model parameters, it was hoped the overall goodness

of fit of the model would not be seriously degraded; the evidence presented later indicates this might be true.

3.7 *Pulse amplitude statistics*

It was assumed the burst amplitude and the instant of time at which the burst occurred were independent random variables. Figures 34 and 35 show representative empirical cumulative distribution functions and histograms for line 1 burst amplitude data. Based on these curves, a number of distributions can be fitted to the data; only two will be discussed here, a two-sided log normal and a two-sided power Rayleigh[77] (also known as generalized gamma[83]). A two-sided log normal random variable l equals Rq, where R and q are independent random variables, with q equally likely to be $+1$ or -1 and R defined by a log normal

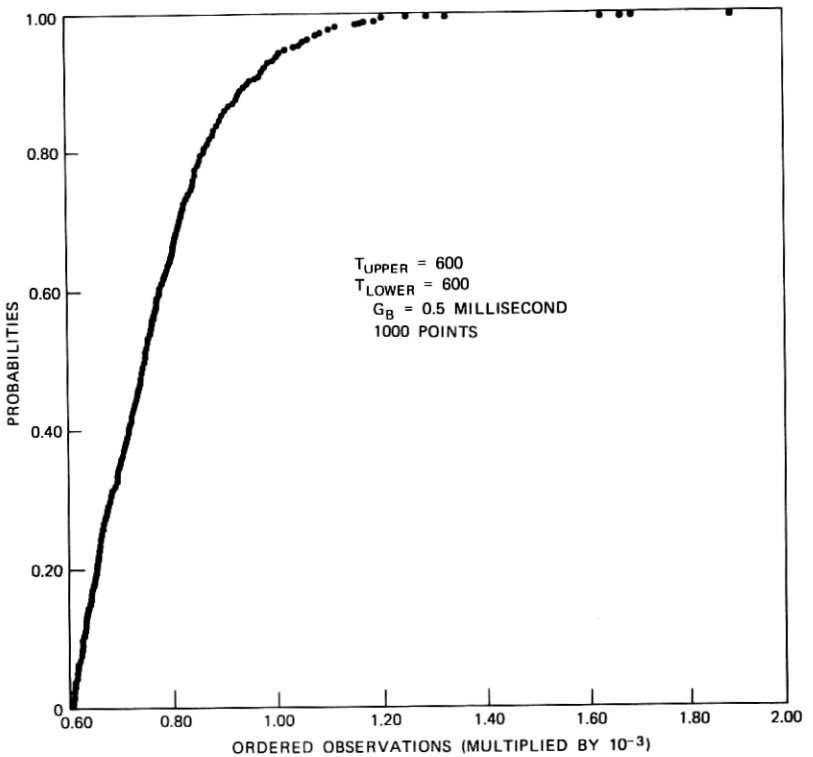


Fig. 34—Empirical cumulative distribution function of maximum burst amplitude (N = 1000).

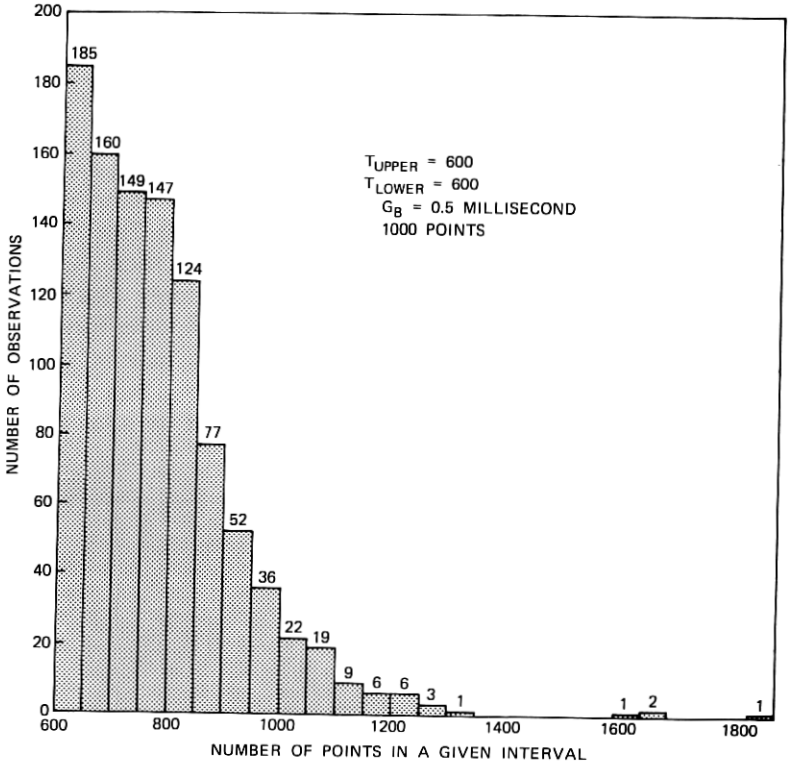


Fig. 35—Histogram of maximum burst amplitudes ($N = 1000$).

density

$$p_R(u) = \frac{1}{\sqrt{2\pi}\sigma u} \exp\left[\frac{-(ln|u| - m)^2}{2\sigma^2}\right] \qquad 0 < u < \infty.$$

Similarly, a two-sided power Rayleigh random variable p equals Rq, where R and q are independent random variables, with q equally likely to be $+1$ or -1, while R is defined by a power Rayleigh density

$$p_R(u) = \left(\frac{K}{R_0}\right)\left(\frac{|u|}{R_0}\right)^{K-1} e^{-(|u|/R_0)^{K-1}} \qquad 0 < K \leqq 2, \quad 0 \leqq |u| < \infty.$$

Each density has zero mean; the log normal variance is $e^{\frac{1}{2}\sigma^2+m}$, while the power Rayleigh variance is $R_0^2\Gamma(1 + 2/K)$.

The parameters of each distribution could be fit to empirical cumulative distribution functions such as that shown in Fig. 34. The range

of parameters for line 1 data, for example, was:

log normal:	mean $5 \leqq m \leqq 6$
	variance $0.01 \leqq \sigma^2 \leqq 0.10$
power Rayleigh:	scale factor $500 \leqq R_0 \leqq 2000$
	exponent $0.50 \leqq K \leqq 0.70$.

The investigation of a much larger amount of data probably would have narrowed the range of these estimates substantially.

3.8 Linear dynamical system parameters

The approach chosen here for estimating linear dynamical system parameters was trial and error. The sample mean burst duration was set equal to the damping constant A^{-1} in both the first- and second-order systems. The second-order system oscillates at frequency ω, which was arbitrarily chosen as $(\frac{1}{2})$ (damping constant)$^{-1}$, to obtain qualitative agreement with actually observed noise bursts (e.g., Fig. 3). For line 1 bursts, for example, $A^{-1} = 0.1$ millisecond for $T_{\text{upper}} = T_{\text{lower}} = 600$, $G_B = 0.5$ millisecond.

3.9 Goodness-of-fit to data of gaussian-plus-filtered-Poisson-process model

Only one test was carried out to provide some heuristic measure of goodness-of-fit of this model to the data. The test was analogous to a quantile-quantile plot. Using typical parameter estimates such as those just described, a computer generated a sample function of a gaussian-plus-filtered-Poisson process. These artificial data were sorted and plotted against actual (sorted) telephone noise from line 1 as shown in Figs. 36 and 37.

The reason for performing just one test is the great difficulty in expressing analytically the distribution function for the gaussian-plus-filtered-Poisson-process model. Hence, it is very difficult to perform quantile-quantile plots of the actual data versus model quantiles, as well as to find maximum likelihood parameter estimates and Cramér-Rao lower bounds on parameter estimate covariances.

3.10 Criticism of the gaussian-plus-filtered-Poisson-process model

Many criticisms of this statistical analysis are possible. First, the question of optimally choosing parameters was never addressed and is still open. Since a large number of parameters must be estimated, a series of presumably suboptimum but easy-to-calculate estimates appeared to be the only feasible course.

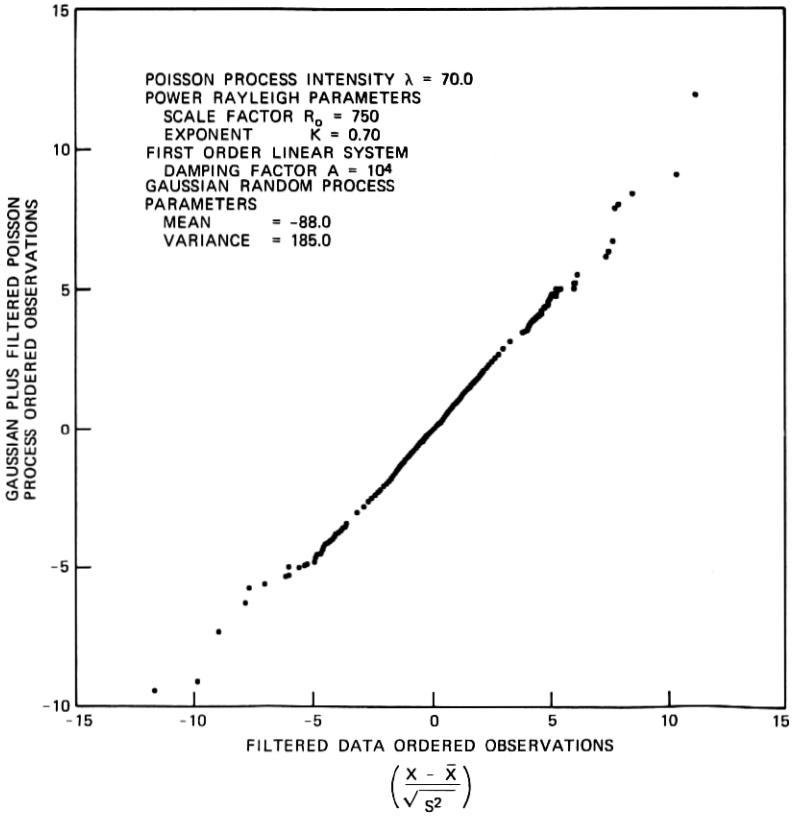


Fig. 36—Line 1 data vs gaussian-plus-filtered-Poisson-process data (N = 13,000) ($\bar{X}$ = sample mean, S^2 = sample variance).

Second, it is not clear how to relate the linear dynamical system parameters in the model to actual telephone system parameters. Where does the filtering occur in reality? Why should it be linear and stationary? Other evidence[19] suggests that the linear dynamical system parameters are not as well defined for other telephone lines as for the data examined in the present work.

Third, the time intervals between bursts are not adequately modeled over the entire observation by a Poisson process. Two other more complicated models were investigated in order to account for this. Many other models can still be investigated.

Fourth, a more general class of models was never investigated that includes the gaussian-plus-filtered-Poisson process as a special case. This model, mentioned briefly earlier, is a mixture of a low-variance gaussian distribution and a high-variance gaussian distribution; during

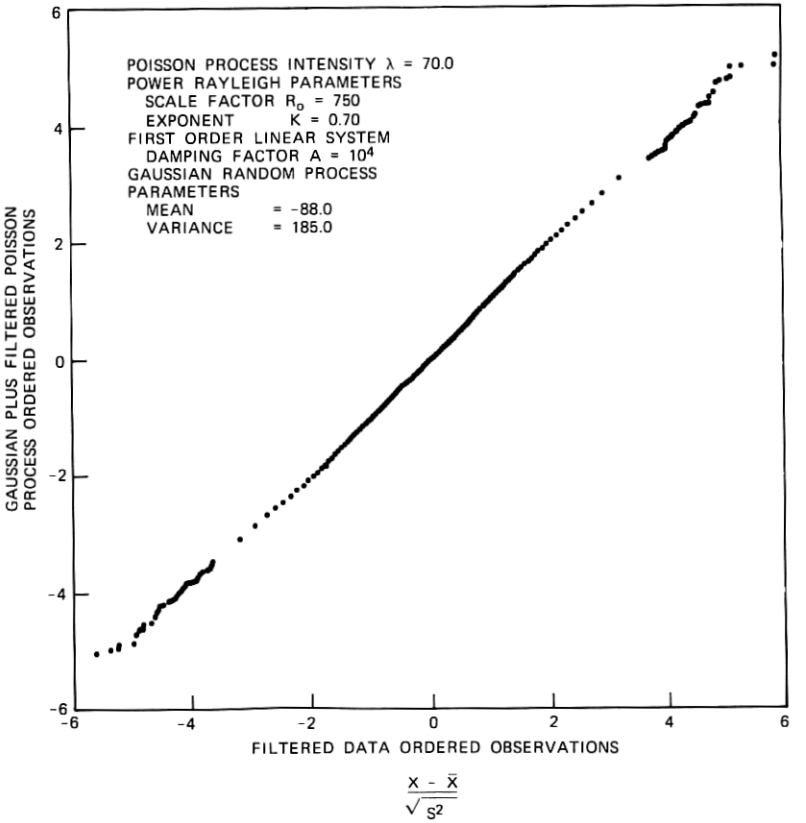


Fig. 37—Center portion of line 1 data vs gaussian-plus-filtered-Poisson-process data (N = 12,980) ($\bar{X}$ = sample mean, S^2 = sample variance).

a fraction P of the time, the low-variance gaussian distribution is chosen to model the data, while during the other $(1 - P)$ fraction of time the high-variance gaussian distribution is chosen. The reasons for not investigating this class of models were that the gaussian-plus-filtered-Poisson-process model comes closer to describing intuitively the physical process of telephone noise generation, and that it has been used by communication theorists in other applications[74–76] more than a mixture of gaussians.

IV. CONCLUSIONS

This study has presented evidence that noise on some lines consists of a deterministic component (e.g., sinusoids at various frequencies) and a purely stochastic component. It was assumed that these com-

ponents add. The data analyzed here suggest that the stochastic component is stationary over short periods of time (typically 1 second) and distinctly nongaussian. Two simple models have been proposed for the nongaussian noise, one based on stable distributions, the other on a mixture of a gaussian process and a nongaussian-filtered Poisson process. Based on the data analyzed here, both models agree intuitively with the physical processes generating telephone noise and appear to fit the data reasonably well.

It is hoped this work will stimulate further research in this area; since only a small number of telephone lines were examined, the models presented here await confirmation by independent investigators. Other models than those discussed here may indeed more adequately account for noise on telephone lines. It is hoped this work will lead to greater insight into how telephone noise limits voice communication and data

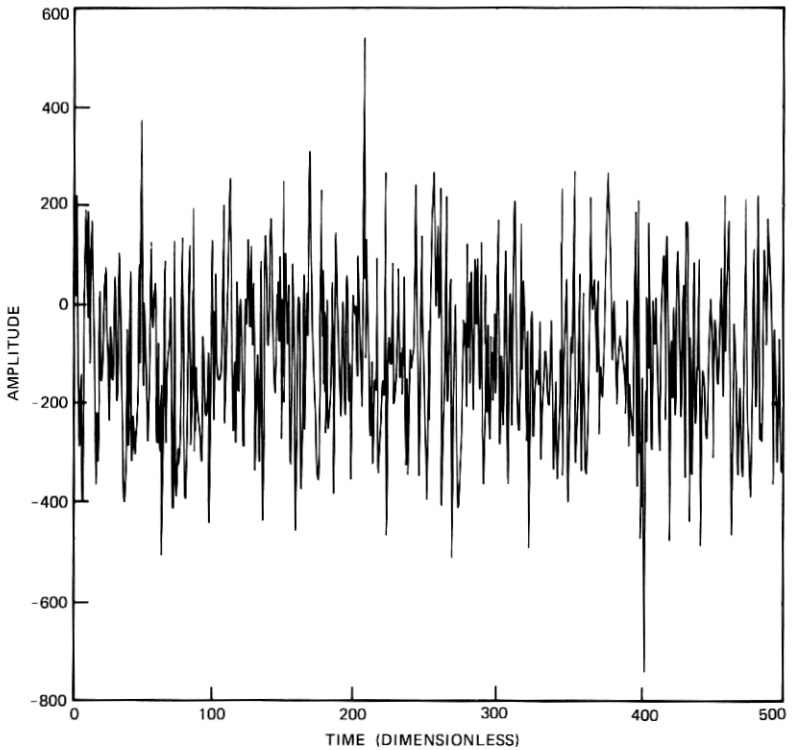


Fig. 38—Computer-generated gaussian-plus-filtered-Poisson-process sample function (same process parameters as in Figs. 36 and 37; $\bar{X}$ = sample mean, S^2 = sample variance).

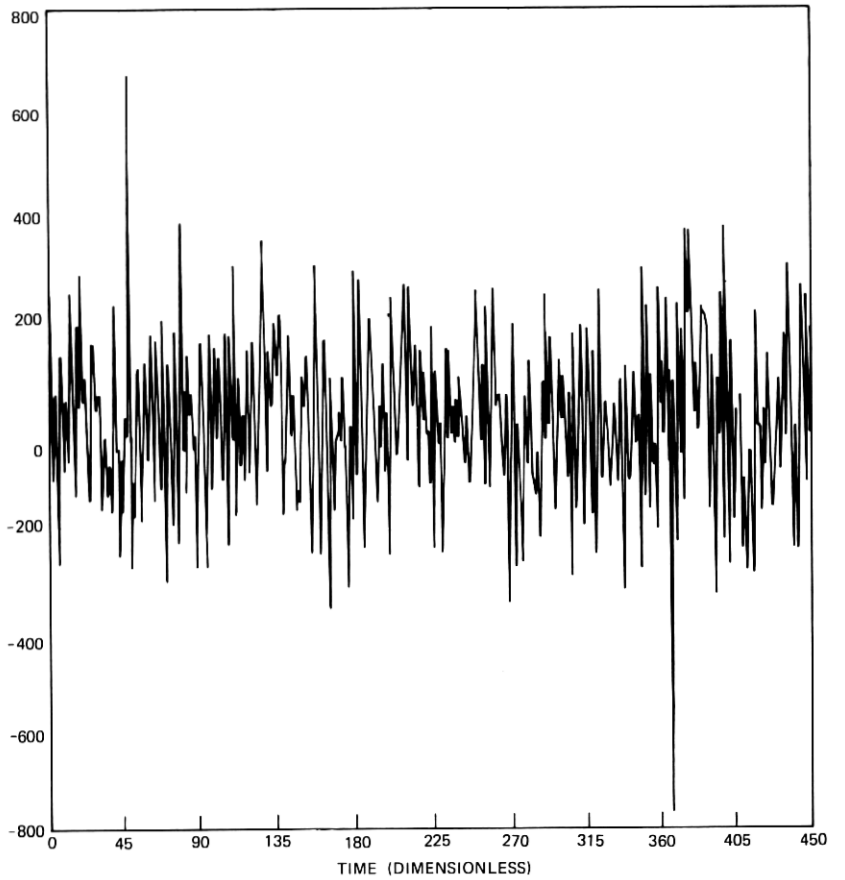
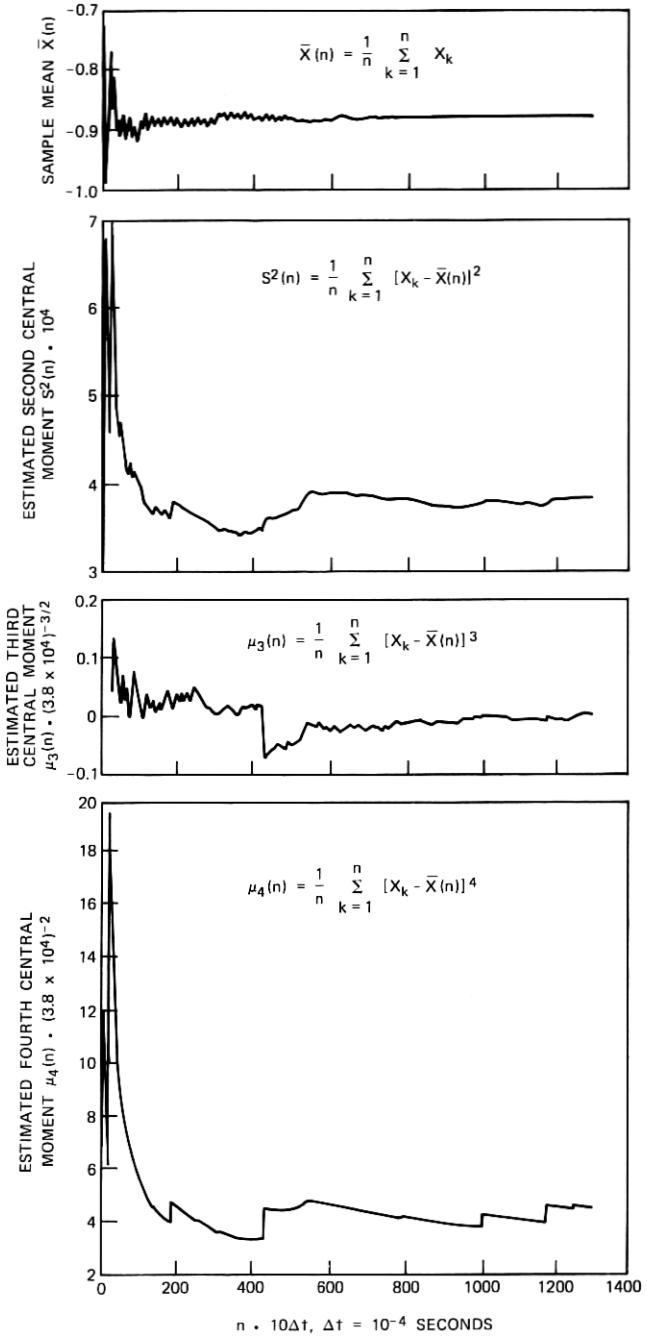


Fig. 39—Computer-generated symmetric stable random process sample function $\alpha = 1.9$ ($c = 100$, $\delta = 0$).

transmission and, more importantly, will lead to new methods for combating this noise.

V. ACKNOWLEDGMENTS

The authors thank J. Fennick for supplying the data, D. Bock for help in the analog-to-digital conversion, L. Rabiner for helpful advice on various aspects of digital filtering, W. DuMouchel for supplying a maximum likelihood stable distribution parameter estimation computer program, M.-J. Cross for accurate calculations of stable distributions used in quantile-quantile plots and elsewhere, J. Chambers for help in calculating and interpreting numerical approximations to


SAMPLE MEAN $\bar{X}(n)$
$\bar{X}(n) = \frac{1}{n}\sum_{k=1}^{n} X_k$
-0.7
-0.8
-0.9
-1.0
ESTIMATED SECOND CENTRAL MOMENT $S^2(n) \cdot 10^4$
$S^2(n) = \frac{1}{n}\sum_{k=1}^{n} [X_k - \bar{X}(n)]^2$
7
6
5
4
3
ESTIMATED THIRD CENTRAL MOMENT $\mu_3(n) \cdot (3.8 \times 10^4)^{-3/2}$
$\mu_3(n) = \frac{1}{n}\sum_{k=1}^{n} [X_k - \bar{X}(n)]^3$
0.2
0.1
0
-0.1
ESTIMATED FOURTH CENTRAL MOMENT $\mu_4(n) \cdot (3.8 \times 10^4)^{-2}$
$\mu_4(n) = \frac{1}{n}\sum_{k=1}^{n} [X_k - \bar{X}(n)]^4$
20
18
16
14
12
10
8
6
4
2
0
200
400
600
800
1000
1200
1400
$n \cdot 10\Delta t$, $\Delta t = 10^{-4}$ SECONDS

maximum likelihood parameter estimates, and J. McKenna and W. M. Boyce for constant encouragement throughout the course of this work. C. Mallows, W. M. Boyce, and D. Slepian had many helpful suggestions for improving this paper. Finally, Miss J. Langoski has done an excellent job in typing and retyping this manuscript.

APPENDIX A

This appendix is included to give the reader some feeling for the two models discussed here. Using typical parameter estimates and a first-order system, a computer generated and plotted a discrete time sample function of a gaussian-plus-filtered-Poisson process (Fig. 38). For comparison, Fig. 39 shows a computer-generated discrete time series, where each point was drawn independently from a symmetric stable distribution (characteristic index $\alpha = 1.9$).

APPENDIX B

A graphic indication that the data can be better modeled by a nongaussian rather than gaussian distribution is now presented. The motivation for this work is found in Mandelbrot.[84]

Estimates for the sample mean, as well as second, third, and fourth central moments can be calculated recursively for larger and larger amounts of data. Figure 40 plots these estimates for 13,000 filtered data (only every tenth estimate is plotted). Note the tendency for the second, third, and fourth central moment estimates to wander rather than stabilize as more and more data are included, as evidenced by the jumps in the estimates. The sample mean, however, does stabilize; note the small ripple in this estimate, which presumably is due to a sinusoid at approximately 600 Hz that was not filtered from the data (see Table II).

These results are qualitatively consistent with results for central moment estimates of computer-generated stable random variables (10,000 independent identically distributed samples, $\alpha = 1.96$, $\beta = 0$). Central moment estimates of computer-generated gaussian random variables (10,000 independent identically distributed samples) did stabilize at the correct values, while exhibiting no apparent jumps such as in Fig. 40. These results are also consistent with the gaussian-plus-filtered-Poisson-process model, with the large jumps in the estimates presumably a result of the filtered Poisson process.

Fig. 40—Sample mean as well as second, third, and fourth central moment estimates for line 1 ($N = 13{,}000$).

REFERENCES

1. J. H. Fennick, "Amplitude Distributions of Telephone Channel Noise and a Model for Impulse Noise," B.S.T.J., *48*, No. 10 (December 1969), pp. 3243–3263.
2. P. Mertz, "Impulse Noise and Error Performance in Data Transmission," Rand Memo RM-4526-PR, DDC AD 614 416, April 1965.
3. H. L. Yudkin, "Some Results in the Measurement of Impulse Noise on Several Telephone Circuits," Proc. Nat. Elec. Conf., Chicago, Ill., *16*, pp. 222–231, 10–12 October 1960.
4. A. Fontaine and R. G. Gallagher, "Error Statistics and Coding for Binary Transmission Over Telephone Circuits," Proc. IRE, *49*, 1961, pp. 1059–1065.
5. E. J. Hofmann, "Error Statistics Utilizing the Code Translation Data-System Over Various Media," M.I.T. Lincoln Laboratory Report 25G-0026, 28 March 1961.
6. E. J. Hofmann, "Error Statistics Utilizing the Milgo Data System Over Various Media," M.I.T. Lincoln Laboratory Report 25G-0025, 23 January 1961.
7. E. J. Hofmann, "Measured Error Distributions on the Bell A1 Facility Over Various Media," Proc. Nat. Elec. Conf., Chicago, Ill., 1960, pp. 37–44.
8. E. N. Gilbert, "Capacity of a Burst-Noise Channel," B.S.T.J., *39*, No. 5 (September 1960), pp. 1253–1265.
9. J. M. Berger and B. Mandelbrot, "A New Model for Error Clustering in Telephone Circuits," IBM J. Research and Dev., *7*, 1963, pp. 224–236.
10. E. O. Elliott, "A Model of the Switched Telephone [Data] Network for Data Communication," B.S.T.J., *44*, No. 1 (January 1965), pp. 89–109.
11. B. D. Frichtman, "A Binary Channel Characterization Using Partitioned Markov Chains," IEEE Trans. Info. Thy., *18*, 1967, pp. 221–227.
12. S. M. Sussman, "Analysis of the Bursts Model for Error Statistics in Telephone Circuits," IEEE Trans. Comm. Sys., *11*, 1963, pp. 213–221.
13. P. A. W. Lewis and D. R. Cox, "A Statistical Analysis of Telephone Circuit Error Data," IEEE Trans. Comm. Tech., *14*, 1966, pp. 382–389.
14. A. A. Alexander, R. M. Gryb, and D. W. Nast, "Capabilities of the Telephone Network for Data Transmission," B.S.T.J., *39*, No. 3 (May 1960), pp. 431–476.
15. H. C. Fleming and R. M. Hutchinson, Jr., "Low-Speed Data Transmission Performance on the Switched Telecommunications Network," B.S.T.J., *50*, No. 4 (April 1971), pp. 1385–1405.
16. M. D. Balkovic, H. W. Klancer, S. W. Klare, and W. G. McGruther, "High-Speed Voiceband Data Transmission Performance on the Switched Telecommunications Network," B.S.T.J., *50*, No. 4 (April 1971), pp. 1349–1384.
17. H. O. Burton and D. D. Sullivan, "Errors and Error Control," Proc. IEEE, *60*, 1972, pp. 1293–1301.
18. A. M. Curtis, "Contact Phenomena in Telephone Switching Circuits," B.S.T.J., *19*, No. 1 (January 1940), pp. 40–62.
19. M. Kurland and D. A. Molony, "Observations on the Effects of Pulse Noise In Digital Data Transmission Systems," IEEE Trans. Comm. Tech., *15*, 1967, pp. 552–556.
20. J. Fennick, private communication.
21. J. E. Evans, "Preliminary Analysis of ELF Noise," M.I.T. Lincoln Laboratory Technical Note 1969-18, DDC AD 691 814, 26 March 1969.
22. A. Ephremides and L. Brandenburg, "On the Reconstruction Error of Sampled Data Estimates," IEEE Trans. on Info. Thy., *IT-19*, 1973, pp. 365–367.
23. J. E. Evans, D. K. Willim, and J. R. Brown, "Description of the Lincoln Laboratory Wideband ELF Noise Recording Systems," M.I.T. Lincoln Laboratory, TN 1972-19, DDC AD 743 006.
24. J. Cooley, P. Lewis, and P. Welch, "The Fast Fourier Transform Algorithm and Its Applications," IBM Research Paper RC-1743, 9 February 1967.
25. J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures*, New York: Wiley-Interscience, 1971.
26. J. W. Cooley, P. A. W. Lewis, and P. D. Welch, "The Application of the Fast Fourier Transform Algorithm to the Estimation of Spectra and Cross-Spectra," J. Sound Vib., *12*, 1970, pp. 339–352.

27. G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, San Francisco: Holden Day, 1970.
28. R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra*, New York: Dover, 1958.
29. Ref. 21, pp. 6–7.
30. J. L. Doob, *Stochastic Processes*, New York: Wiley, 1953, p. 519.
31. H. Cramér, *Mathematical Methods of Statistics*, Princeton: Princeton University Press, 1946.
32. M. B. Wilk and R. Gnanadesikan, "Probability Plotting Methods for the Analysis of Data," Biometrika, *55*, 1968, pp. 1–17.
33. E. S. Pearson and H. O. Hartley, *Biometrika Tables for Statisticians*, Cambridge: Cambridge University Press, 1966, pp. 207–208.
34. B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*, rev. ed., Reading, Mass.: Addison-Wesley, 1968, Chs. 3, 4, 6, 7.
35. W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. II, New York: Wiley, 1966, Chs. 6, 9, 17.
36. B. Mandelbrot, "How Long is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension," Science, *156*, 1967, pp. 636–638.
37. B. Mandelbrot, "Long-Run Linearity, Locally Gaussian Processes, H-Spectra, and Infinite Variances," Int. Econ. Rev., *10*, 1969, pp. 82–111.
38. B. Mandelbrot, "Sporadic Random Functions and Conditional Spectral Analysis: Self Similar Examples and Limits," *Fifth Berkeley Symp. Math. Stat. Prob.*, Vol. III, Berkeley and Los Angeles: University of California Press, 1967, pp. 155–179.
39. B. Mandelbrot, "The Variation of Certain Speculative Prices," J. Business, *36*, 1963, pp. 394–419.
40. B. Mandelbrot, "The Pareto-Levy Law and the Distribution of Income," Int. Econ. Rev., *1*, 1960, pp. 79–106.
41. B. Mandelbrot, "The Stable Paretian Income Distribution When the Apparent Exponent is Near Two," Int. Econ. Rev., *4*, 1963, pp. 111–114.
42. B. Mandelbrot, "The Variation of Some Other Speculative Prices," J. Bus. Univ. Chicago, *40*, 1967, pp. 393–413.
43. E. Fama, "Behavior of Stock Market Prices," J. Bus. U. Chicago, *38*, 1965, pp. 34–105.
44. D. A. Darling, "The Influence of the Maximum Term in the Addition of Independent Random Variables," Trans. Am. Math. Soc., *73*, 1952, pp. 95–107.
45. M. Loève, *Probability Theory*, New York: Van Nostrand Reinhold, 1960, pp. 288–318.
46. W. H. DuMouchel, "Stable Distributions in Statistical Inference," Yale Ph.D. dissertation, 1971.
47. E. F. Fama and R. Roll, "Properties of Symmetric Stable Distributions," J.A.S.A., *63*, 1968, pp. 817–836.
48. E. F. Fama and R. Roll, "Parameter Estimates for Symmetric Stable Distributions," J.A.S.A., *66*, 1971, pp. 331–338.
49. Ref. 34, p. 162.
50. Ref. 34, p. 197.
51. Ref. 34, p. 164.
52. M.-J. Cross, "Tables of Finite Mean Nonsymmetric Stable Distributions as Computed from their Convergent and Asymptotic Series," to appear in J. Stat. Comp. Simulation.
53. Ref. 34, p. 171.
54. Ref. 34, p. 183.
55. Ref. 35, p. 169.
56. Ref. 35, p. 459.
57. J. D. Mason, "Convolution of Stable Laws as Limit Distributions of Partial Sums," Ann. Math. Stat., *41*, 1970, pp. 101–114.
58. V. M. Zolotarev and V. S. Korolyak, "On a Hypothesis Proposed by B. V. Gnedenko," Thy. Prob. Appl., *6*, 1961, pp. 431–435.
59. P. Billingsley, "Limit Theorems for Randomly Selected Partial Sums," Ann. Math. Stat., *33*, 1962, pp. 85–92.

60. A. E. Berry, "The Accuracy of the Gaussian Approximation to the Sum of Independent Variates," Trans. Am. Math. Soc., *49*, 1941, pp. 122–136.
61. R. G. Esseen, "Fourier Analysis of Distribution Functions. A Mathematical Study of the Laplace-Gaussian Laws," Acta. Math., *77*, 1945, pp. 1–125.
62. H. Cramér, "On Asymptotic Expansions for Sums of Independent Random Variables with a Limiting Stable Distribution," Sankhyā, Series A, 1963, pp. 13–24.
63. W. DuMouchel, "Stable Distributions in Statistical Inference: 1. Symmetric Stable Distributions Compared to Other Symmetric Long-Tailed Distributions," J.A.S.A., *68*, 1973, pp. 469–477.
64. R. M. Dudley, "Speeds of Metric Probability Convergence," Z. Wahrschein., *22*, 1972, pp. 323–332.
65. C. C. Heyde, "On Large Deviation Probabilities in the Case of Attraction to a Non-Normal Stable Law," Sankhyā, Series A, *30*, 1968, pp. 253–258.
66. V. Boonyasombut and J. M. Shapiro, "The Accuracy of Infinitely Divisible Approximations to Sums of Random Variables With Application to Stable Laws," Ann. Math. Stat., *41*, 1970, pp. 237–250.
67. T. A. Hern, "Error Estimates for Weak Convergence to Certain Infinitely Divisible Laws," Ann. Math. Stat., *43*, 1972, pp. 1592–1602.
68. M. Lipschultz, "On Strong Bounds for Sums of Independent Random Variables Which Tend to a Stable Distribution," Trans. Am. Math. Soc., *81*, 1956, pp. 135–154.
69. W. Doeblin, "Sur l'Ensemble de Puissances d'une Loi de Probabilité," Studia Math., *9*, 1940, pp. 71–96.
70. W. H. DuMouchel, "On the Asymptotic Normality of the Maximum Likelihood Estimate when Sampling from a Stable Distribution," Ann. Stat., *1*, 1973, pp. 948–957.
71. Ref. 46, p. 87.
72. Ref. 34, p. 189.
73. H. G. Tucker, "Convolutions of Distributions Attracted to Stable Laws," Ann. Math. Stat., *39*, 1968, pp. 1381–1390.
74. D. L. Snyder, "Optimal Detection of Known Signals in a Non-Gaussian Noise Resembling VLF Atmospheric Noise," 1968 Wescon Convention Record, Part 4.
75. E. V. Hoversten, R. O. Harger, and S. J. Halme, "Communication Theory for the Turbulent Atmosphere," Proc. IEEE, *58*, 1970, pp. 1626–1650.
76. S. D. Personick, "Statistics of a General Class of Avalanche Detectors with Applications to Optical Communication," B.S.T.J., *50*, No. 10 (December 1971), pp. 3075–3095.
77. J. W. Modestino, "A Model for ELF Noise," M.I.T. Lincoln Laboratory Tech. Rpt. 493, DDC AD 737 368, 16 December 1971.
78. D. R. Cox and P. A. W. Lewis, *The Statistical Analysis of Series of Events*, London: Methuen, 1966.
79. Ref. 78, p. 71.
80. Ref. 78, p. 176.
81. Ref. 78, pp. 134–171.
82. Ref. 78, p. 135.
83. N. L. Johnson and S. Kotz, *Continuous Univariate Distribution I*, Boston: Houghton Mifflin, 1970, p. 197.
84. Ref. 39, p. 396.