TESTS FOR RANDOMNESS, STATIONARITY, NORMALITY AND COMPARISON OF SPECTRA

J. S. BENDAT, L. D. ENOCHSON, and A. G. PIERSOL

MEASUREMENT ANALYSIS CORPORATION

FOREWORD

This report was prepared by the Measurement Analysis Corporation, Los Angeles, California, for the Aero-Acoustics Branch, Vehicle Dynamics Division, AF Flight Dynamics Laboratory, Wright-Patterson Air Force Base Ohio, under Contract AF33(615)-1314. This research is part of a continuing effort to obtain significant information on sound environment simulation and dynamic response to acoustic excitation under the Research and Technology Division, Air Force Systems Command's exploratory development program. The Project No. is 4437, "High Intensity Sound Environment Simulation," and Task No. 443706, "Advanced Instrumentation Study for Sonic Fatigue Experimental Work." Mr. W. K. Shilling III was the project engineer.

The contractor's report number is MAC 402-03A. The manuscript was released by the authors April 1965 for publication as an Air Force Flight Dynamics Laboratory Technical Report.

This report covers work conducted from February 1964 to April 1965.

This technical report has been reviewed and is approved.

Walter J. Mykytow

WALTER J. MYKYTOW
Asst. for Research & Technology
Vehicle Dynamics Division
AF Flight Dynamics Laboratory

ABSTRACT

This report contains material for carrying out certain practical tests for randomness, stationarity, and normality of physical data, as well as details on how to test for equivalence of two power spectral density functions. The tests for randomness are qualitative inspections of measured properties from sample records, such as power spectra, probability density, auto-correlation functions, which may be part of a regular data reduction procedure. The tests for stationarity are nonparametric statistical procedures, based upon Run and Trend Distributions, which require simple quantitative calculations for acceptance or rejection. The test for normality is a statistical hypothesis test based upon sample measurements of the amplitude probability density function at seven equispaced amplitude levels from zero to plus and minus three sigma. To test for equivalence of two spectra, a straight-forward procedure is recommended which requires knowledge only of the number of degrees-of-freedom associated with the spectral estimates, and the full bandwidth occupied by the data. A special case of this procedure is to compare an unknown spectrum to a "white" spectrum.

Contrails

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

1. INTRODUCTION

Before measured data can be properly interpreted for any physical problem, it is necessary to know or assume certain basic properties about the data, and if possible to verify these properties. The three most important properties are randomness, stationarity, and normality. By randomness is meant that the data is not deterministic and in particular is not periodic, since different analysis procedures are required for random data as opposed to periodic data. In order to predict future properties of random data from analysis of its past properties, one needs to investigate the stationarity (time invariance) of the data. Here again, different analysis procedures are required for dealing with nonstationary data as opposed to stationary data. Finally, the establishment or not of normality is desirable so as to estimate the probability of exceeding various amplitude levels, as well as to provide confidence limits for other quantities which are calculated from the data, such as power spectra.

A separate question is the problem of measuring two different power spectra from different data and determining whether or not these two power spectra are statistically equivalent. Related to this question is the problem of shaping a power spectrum to a desired form, such as bandwidth limited "white" noise. This second question is clearly a special case of the first question where one desires to compare a measured power spectrum to a given "white" power spectrum.

This report discusses practical techniques for testing measured data for randomness, stationarity, normality, and for comparing spectra. Tests for randomness are covered in Section 2, followed by tests for stationarity in Section 3 and tests for normality in Section 4. The final Section 5 develops ideas for testing two power spectra for equivalence.

2. TESTS FOR RANDOMNESS

Random vibration is that type of time-varying motion which consists of randomly varying amplitudes and frequencies such that its behavior can be described only in statistical terms. No analytical representation for the vibratory motion is possible. The vibration does not repeat itself after a finite time period. For all practical purposes, a continuous vibration may be considered random unless motion with a periodic or "almost-periodic" form is present. Almost-periodic vibrations are those which are aperiodic but have a frequency spectrum consisting of discrete components (for example, the sum of two sine waves whose frequency ratio is an irrational number is almost-periodic). Hence, almost-periodic data may be visualized as a series of sinusoids, just as for the case of periodic data.

If vibration data is periodic, or almost-periodic, the fact is usually obvious by simple observation of an amplitude time history plot. However, when vibration data is a mixture of both random and periodic or almost-periodic portions, this fact is often not obvious. There are three different techniques which may be used to detect the presence of sinusoidal components produced by periodic or almost-periodic portions in an otherwise random vibration response. These techniques consist of qualitative inspections of the vibration properties which might be measured from sample records as a customary part of the random data reduction.

The presence of sinusoidal components in an otherwise random vibration response may often be detected by visual inspection of a power spectral density function, an amplitude probability density function, and/or an autocorrelation function measured from stationary sampled data.

To illustrate how a power spectrum can reveal the presence of a sinusoidal component in an otherwise random signal, refer to Figure 1. In this example, the output of a random noise generator was mixed with a sinusoidal signal. The sinusoidal signal was given an rms amplitude equal

to 1/20th that of the random signal. Plot A which was made using a relatively wide filter gives little or no indication of the presence of the sinusoid. Plot B which was made using a medium filter indicates a possible sinusoid quite clearly. Plot C which was made using a narrow filter gives a strong indication.

Figure 1 illustrates how a highly resolved power spectrum will reveal sinusoidal components as sharp peaks,even when the sinusoids are of relatively small intensity. However, a sharp peak in the power spectrum for a vibration response may also represent the narrow band random response of a lightly damped structural resonance. These two cases can sometimes be distinguished from one another by repeating the power spectral measurement with the narrowest available PSD analyzer filter bandwidth. If the measured spectral peak represents a sine wave, the indicated bandwidth of the peak will always be equal to the bandwidth of the PSD analyzer filter, no matter how narrow the filter is. This method of detection will clearly not work unless the bandwidth for the narrowest PSD analyzer filter is smaller than the bandwidth for a possible narrow band random response peak. For the case of spectral peaks with relatively low center frequencies, say less than 50 cps, a structural resonance may have a bandwidth of less than 1 cps making it very difficult to distinguish from a sine wave.

The presence of sinusoidal components in an otherwise random vibration response may also be revealed by an amplitude probability density function for the response. The probability density plots for a sine wave and a random signal are markedly different. A random signal will usually have a probability density function which at least resembles the familiar bell-shaped Gaussian characteristic, while a sine wave has a dish-shaped probability density function. A mixture of the two takes on prominent characteristics of both. This is clearly illustrated by the actual measurements in Plots A, B, and C of Figure 2.

Perhaps the most powerful method of detecting sinusoidal components in an otherwise random vibration response is presented by an autocorrelation

plot. For any purely random signal, the autocorrelation function will always approach zero (assuming the signal has no DC component) as the time displacement becomes large. On the other hand, the autocorrelation function for a periodic or almost-periodic signal is also periodic or almost-periodic, and will continue to oscillate no matter how large the time displacement becomes. Thus, the autocorrelation plot for a signal representing a mixed random and periodic or almost-periodic vibration response will decay to a perpetual oscillation as the time displacement becomes large. These matters are illustrated by Plots A, B, and C of Figure 3.

POWER SPECTRA FOR A SINE WAVE MIXED WITH A RANDOM SIGNAL OBTAINED USING THREE DIFFERENT PSD ANALYZER FILTER BANDWIDTHS B.

RMS AMPLITUDE OF SINE WAVE = 0.05 VOLTS

RMS AMPLITUDE OF RANDOM SIGNAL = 1.0 VOLTS

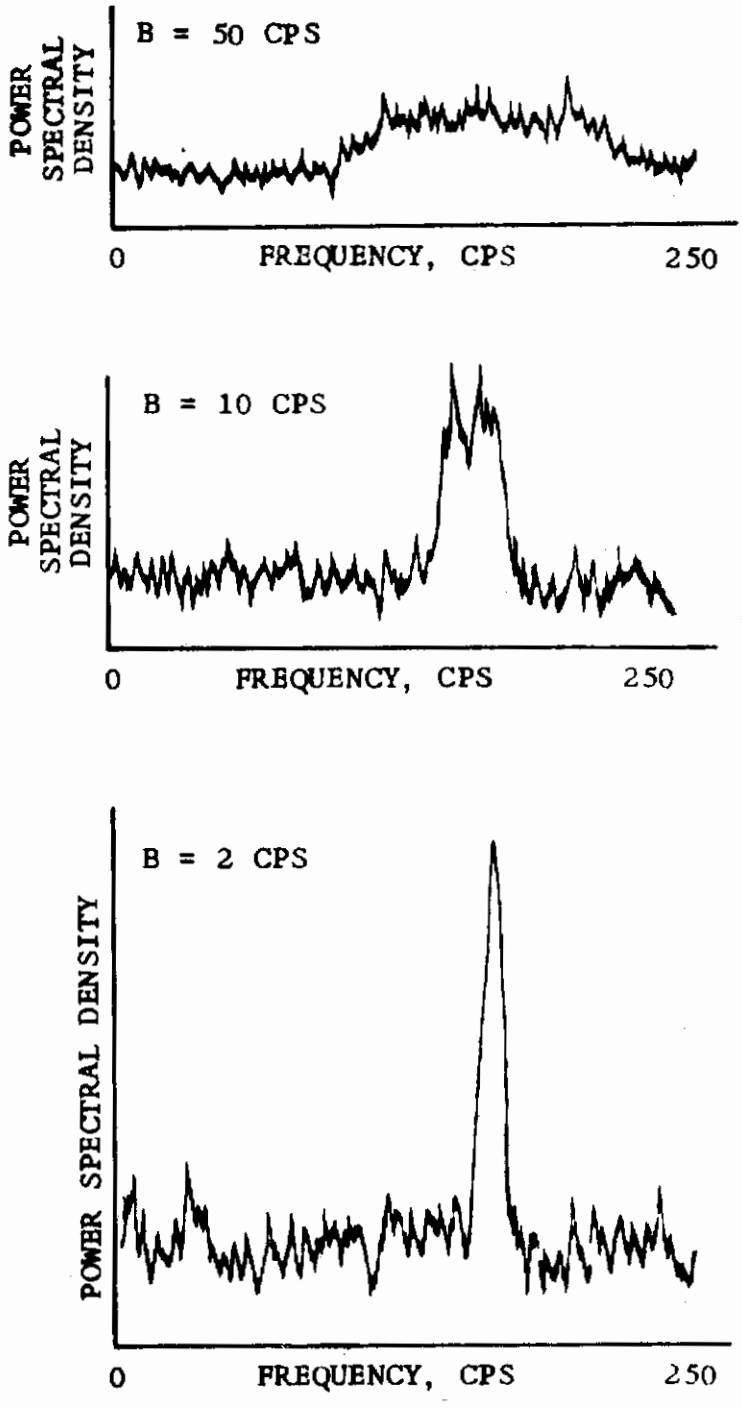
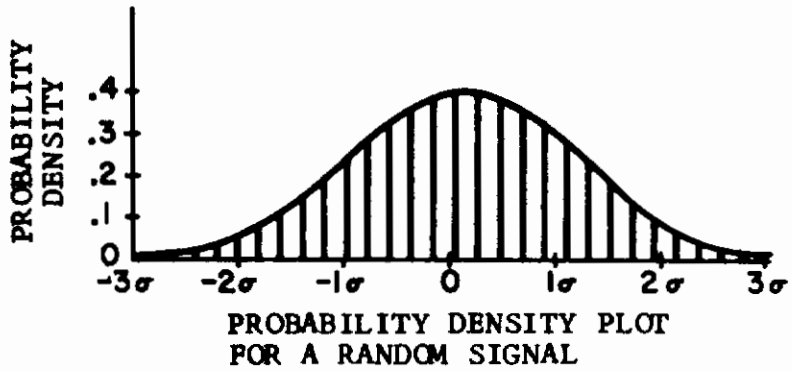


FIGURE 1. ACTUAL POWER SPECTRA PLOTS

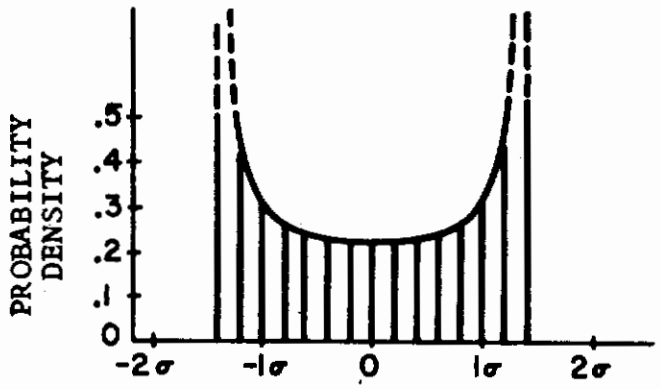
RMS AMPLITUDE OF A SINE WAVE = 1.0 VOLTS

RMS AMPLITUDE OF A RANDOM SIGNAL = 0.3 VOLTS

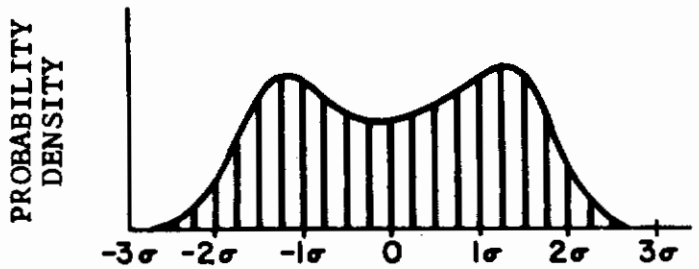
RMS AMPLITUDE OF TOTAL SIGNAL = σ = 1.04 VOLTS



PROBABILITY DENSITY PLOT FOR A RANDOM SIGNAL

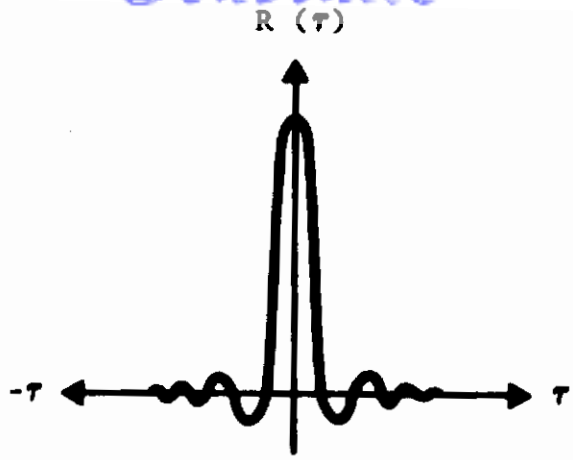


PROBABILITY DENSITY PLOT FOR A SINE WAVE

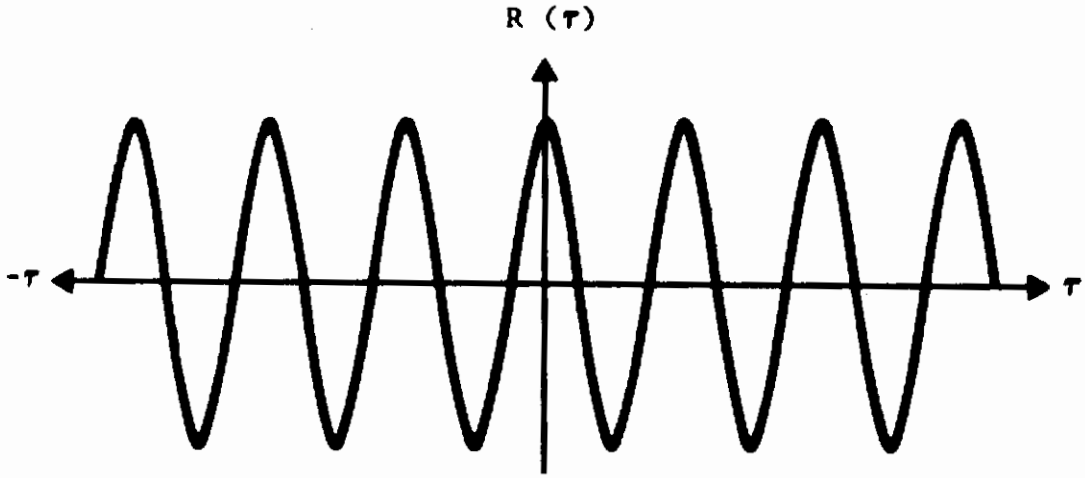


PROBABILITY DENSITY PLOT FOR A SINE WAVE MIXED WITH A RANDOM SIGNAL

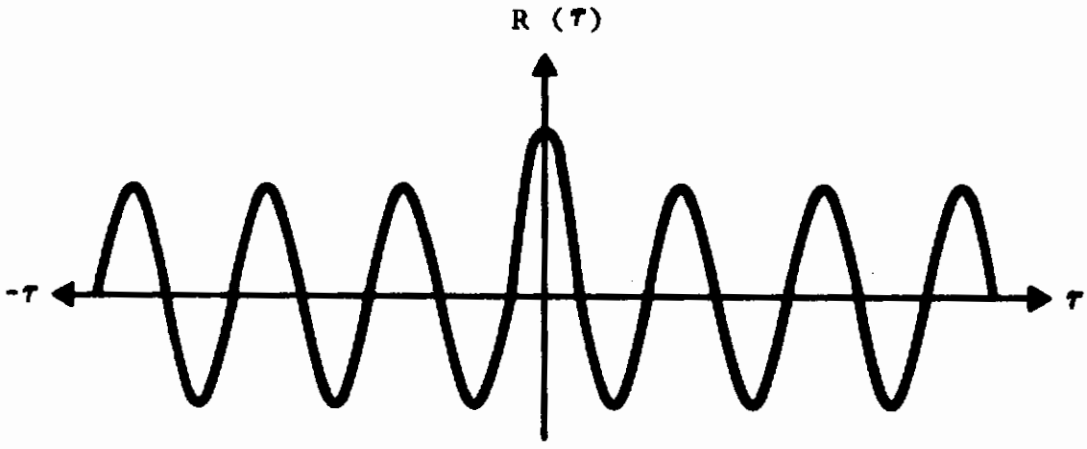
FIGURE 2. ACTUAL PROBABILITY DENSITY PLOTS



AUTOCORRELATION PLOT FOR A RANDOM SIGNAL



AUTOCORRELATION PLOT FOR A SINE WAVE



AUTOCORRELATION PLOT FOR A SINE WAVE MIXED WITH A RANDOM SIGNAL

FIGURE 3. TYPICAL AUTOCORRELATION PLOTS

3. TESTS FOR STATIONARITY

Previous work in Reference 1, Section 6. 1. 8 and Reference 2, Section 16, developed quantitative tests for stationarity which assumed that the random variable being sampled had an underlying normal distribution. Uncertainty expressions which were used required information about the bandwidth occupied about the data, and knowledge about the exact averaging times used to measure properties of the data. Also, in applying these tests, it was necessary to remove all periodic components from the data so as to deal only with completely random data.

In actual practice, some of these requirements are not satisfied, and it is difficult to obtain all of the needed information. These problems can often be avoided by using statistical procedures which do not assume any distribution for the random variable in question. Such procedures are called distribution free or nonparametric procedures. Two such practical procedures will be discussed here, known as the Run Test and the Trend Test. These will then be applied to yield practical nonparametric tests for stationarity.

3. 1 RUN TEST

Consider a sequence of N observations of a random variable x where each observation is classified into one of two mutually exclusive categories, which may be identified simply by plus (+) or minus (-). The simplest example would be a sequence of flips of a coin where each observation is either a heads (+) or a tails (-). A second example might be a sequence of measured values x_i ($i = 1, 2, 3, \ldots$) with a mean value $\bar{x}$, where each observation is $x_i \geq x$, (+) or $x_i < x$, (-). A third example might be a simultaneous sequence of two sets of measured values x_i and y_i ($i = 1, 2, 3, \ldots$), where each observation is $x_i \geq y_i$(+) or $x_i < y_i$(-). In any case, the sequence of plus and minus observations might be as follows.

+	+	-	+	+	-	+	+	+	-	+	-	-	+	-	-	+	-	-	-
1		2	3		4	5			6	7	8		9	10		11	12		

A run is defined as a sequence of identical observations that are followed or preceded by a different observation or no observation at all. In the above example, there are $r = 12$ runs in the sequence of $N = 20$ observations. The distribution for r is the basis for the Run Test.

The number of runs which occur in a sequence of observations gives an indication as to whether or not results are independent random observations of the same random variable. Specifically, if a sequence of N observations are independent observations of the same random variable, that is, the probability of a (+) or (-) result does not change from one observation to the next, then the sampling distribution for the number of runs in the sequence is a random variable r with a mean value and variance as follows.

$$\mu_r = \frac{2N_1N_2}{N} + 1 \qquad ; \qquad N = N_1 + N_2 \tag{1}$$

$$\sigma_r^2 = \frac{2N_1N_2(2N_1N_2 - N)}{N^2(N - 1)} \tag{2}$$

Here, N_1 is the number of (+) observations and N_2 is the number of (-) observations.

For the special case where $N_1 = N_2 = N/2$, Eqs. (1) and (2) reduce to

$$\mu_r = \frac{N}{2} + 1 \tag{3}$$

$$\sigma_r^2 = \frac{N(N - 2)}{4(N - 1)} \tag{4}$$

For $N \geq 10$, the distribution for the random variable r is approximately normal with the above mean value and variance. A limited tabulation of the distribution function for r at the 5% level of significance is presented in Table 1, using $\mu_r \mp 1.96\,\sigma_r$ for the endpoints of the 95% region, and then rounding off to conservative acceptance values. For example, values of (8.1, 17.9) become (9, 17), to give the smallest acceptance region.

TABLE 1

Run Test Distribution at 5% Level of Significance

Number of Measurements $N_1 = N_2 = N/2$	Expected Number Runs μ_r	95% Acceptance Region for Independence Hypothesis	
		0.975	0.025
6	7	4	10
10	11	7	15
12	13	9	17
14	15	10	20
16	17	12	22
18	19	14	24
20	21	15	27
30	31	24	38
40	41	33	49
50	51	42	60
60	61	51	71
70	71	60	82
80	81	69	93
90	91	78	104
100	101	88	114

Perhaps the most direct application of runs to data evaluation problems involves the testing of a single sequence of observations for independence. Assume there is reason to believe that there is an underlying trend in a sequence of observations. That is, there is reason to believe that the probability of a (+) or (-) is changing from one observation to the next. The existence of a trend can be tested for as follows. Hypothesize that the

sequence of N observations are independent observations of the random variable. Then, the number of runs in the sequence has a sampling distribution as given in Table 1. The hypothesis can be tested at any desired level of significance α by comparing the observed runs to the interval between $r_{n;(1-\alpha/2)}$ and $r_{n;\alpha/2}$ where $n = N/2$. If the observed runs fall outside the interval, the hypothesis is rejected at the α level of significance. Otherwise, the hypothesis is accepted.

The quantity $r_{n;p}$ as used here means that the random variable r satisfies the one-sided probability condition

$$\text{Prob}\left[r \geq r_{n;p}\right] = p \tag{5}$$

Hence, when $p = 1 - (\alpha/2)$ and when $p = (\alpha/2)$,

$$\text{Prob}\left[r \geq r_{n;(1-\alpha/2)}\right] = 1 - (\alpha/2) \tag{6}$$

$$\text{Prob}\left[r \geq r_{n;\alpha/2}\right] = \alpha/2 \tag{7}$$

The two-sided region for r is now governed by the probability condition

$$\text{Prob}\left[r_{n;(1-\alpha/2)} \leq r \leq r_{n;\alpha/2}\right] = 1 - \alpha \tag{8}$$

For example, if $\alpha = 5\% = 0.05$, then

$$\text{Prob}\left[r_{n;0.975} \leq r \leq r_{n;0.025}\right] = 0.95 \tag{9}$$

In words, there is 95% probability that the random variable r will fall in the interval bounded by $r_{n;0.975}$ and $r_{n;0.025}$. There is 5% probability that the random variable r will fall outside this interval, when in fact it should fall inside the interval. This 5% value for α is called the level of significance for the test.

Example 1:

Assume a sequence of N = 20 observations of a random variable produces results as noted in the table below.

(1)	5.5	(6)	7.1	(11)	5.8	(16)	5.8
(2)	4.4	(7)	5.0	(12)	7.2	(17)	5.0
(3)	5.7	(8)	6.1	(13)	6.7	(18)	4.6
(4)	5.2	(9)	5.5	(14)	5.4	(19)	7.6
(5)	4.8	(10)	8.0	(15)	6.4	(20)	5.4

Determine if the observations are independent by testing the runs which occur in the variation of the observations about their median value. Perform the test at the $\alpha = 0.05$ level of significance.

By visual inspection of the data, it is seen that $x = 5.6$ is the median value of the 20 observations. Let all observations with a value greater than 5.6 be identified by (+) and all with a value less than 5.6 be identified by (-). The result is

-	-	+	-	-	+	-	+	-	+	+	+	+	-	+	+	-	-	+	-
1		2	3		4	5	6	7		8			9	10		11		12	13

Hence, there are 13 runs represented by the sequence of 20 observations. Assume the observations are independent. The acceptance region for this hypothesis is

$$r_{10;(1-\alpha/2)} \leq r \leq r_{10;(\alpha/2)}$$

From Table 1, for $\alpha = 0.05$, $r_{10;(1-\alpha/2)} = r_{10;0.975} = 7$ and $r_{10;\alpha/2} = r_{10;0.025} = 14$. Hence, the hypothesis is accepted since $r = 13$ falls within the range between 7 and 14. That is, there is no reason to believe that the observations are not independent.

3.2 TREND TEST

Consider a sequence of N observations of a random variable x. Let the observations be denoted by $x_1, x_2, x_3 \ldots, x_N$. Now, count the number of times that $x_i > x_j$ for $i < j$. Each such inequality is called a reverse arrangement. The total number of reverse arrangements is a random variable which will be denoted by A. The distribution for A is the basis for the trend test.

To clarify just what a reverse arrangement is and how to count them in a given sequence of observations, consider the following sequence:

$$x_1 = 5,\ x_2 = 3,\ x_3 = 8,\ x_4 = 9,\ x_5 = 4,\ x_6 = 1,\ x_7 = 7,\ x_8 = 5$$

In the above sequence $x_1 > x_2$, $x_1 > x_5$, and $x_1 > 6$ which gives a number of reverse arrangements for x_1 of $A_1 = 3$. Now, choosing x_2 and comparing it against subsequent observation (i.e., for $i = 2$ and $i < j = 3, 4, \ldots, 8$) one notes $x_2 > x_6$ only so that the number of reverse arrangements for x_2 is $A_2 = 1$. For x_3 one finds that $A_3 = 4$. Continuing on, one sees that $A_4 = 4$, $A_5 = 1$, $A_6 = 0$, and $A_7 = 1$. The total number of reverse arrangements is therefore

$$A = A_1 + A_2 + \ldots + A_7 = 3 + 1 + 4 + 1 + 0 + 1 = 14$$

A general definition for the random variable A is as follows. From the set of observations $\{x_i\}$ j $i = 1, 2, \ldots, N$, define

$$h_{ij} = \begin{cases} 1 & \text{if } x_i > x_j \\ 0 & \text{otherwise} \end{cases} \qquad (10)$$

Then

$$A = \sum_{i=1}^{N-1} A_i \tag{11}$$

where

$$A_i = \sum_{j=i+1}^{N} h_{ij} \tag{12}$$

For example,

$$A_1 = \sum_{j=2}^{N} h_{1j} \quad ; \quad A_2 = \sum_{j=3}^{N} h_{2j} \quad ; \quad A_3 = \sum_{j=4}^{N} h_{3j}$$

If the sequence of N number of observations are independent observations of the same random variable, then the sampling distribution for the total number of reverse arrangements A in the sequence will have a mean value and variance as follows.

$$\mu_A = \frac{N(N-1)}{4} \tag{13}$$

$$\sigma_A^2 = \frac{2N^3 + 3N^2 - 5N}{72} = \frac{N(2N+5)(N-1)}{72} \tag{14}$$

For $N \geq 10$, the distribution for the random variable A will be approximately normal with the above mean value and variance. Limited tabulation of the distribution function for A at the 5% level of significance is presented in Table 2, using $\mu_r \mp 1.96\, \sigma_A$ for the endpoints of the 95% region, and then rounding off to conservative acceptance values as in Table 1.

The trend test is applied in basically the same way that the run test is applied, and generally speaking, is more powerful than the run test for detecting monotonic trends in a sequence of observations. The trend test is not powerful, however, for detecting fluctuating trends.

TABLE 2

Trend Test Distribution at 5% Level of Significance

Number of Measurements N	Expected Number Trends μ_A	95% Acceptance Region for Independence Hypothesis 0.975	0.025
10	22.5	12	33
12	33.0	19	47
14	45.5	28	63
16	60.0	39	81
18	76.5	51	102
20	95.0	65	125
30	217.5	163	272
40	390.0	306	474
50	612.5	496	729
60	885.0	732	1038
70	1207.5	1015	1400
80	1580.0	1345	1815
90	2002.5	1722	2283
100	2475.0	2146	2804

Example 2:

Test the sequence of N = 20 observations in Example 1 for a trend at the $\alpha = 0.05$ level of significance. The number of reverse arrangements in the observations are as follows.

$A_1 = 8$ $A_6 = 11$ $A_{11} = 4$ $A_{16} = 3$

$A_2 = 0$ $A_7 = 1$ $A_{12} = 7$ $A_{17} = 1$

$A_3 = 8$ $A_8 = 7$ $A_{13} = 6$ $A_{18} = 0$

$A_4 = 4$ $A_9 = 4$ $A_{14} = 2$ $A_{19} = 1$

$A_5 = 1$ $A_{10} = 10$ $A_{15} = 4$

The total number of reverse arangements is A = 82.

Assume the observations are independent observations of a random variable x where there is no trend. The acceptance region for the hypothesis is

$$A_{20;(1-\alpha/2)} \leq A \leq A_{20;(\alpha/2)}$$

From Table 2, for $\alpha = 0.05$, $A_{20;(1-\alpha/2)} = A_{20;0.975} = 64$ and $A_{20;\alpha/2} = A_{20;0.025} = 126$. Hence, the hypothesis is accepted since A = 82 falls within the range between 64 and 126.

3.3 RECOMMENDED PROCEDURE

The general concept of stationarity is presented in rigorous terms in Reference 1, Section 4.4. From that material, proof of stationarity would theoretically involve verification that all statistical properties for the random process of interest are invariant with time translations. Such verifications are clearly not feasible in practical terms since there are an infinite number of possible statistics, and a complete description of the random process by an ensemble would be required to compute them. However, by noting certain important assumptions which are generally valid for the vast majority of random data in nature, practical tests for stationarity can be developed.

The first important assumption is: If the data of interest is nonstationary, then the statistical properties computed by time averaging over each of a sequence of short time intervals from a single sample record vary significantly from one interval to the next. Furthermore, if the data of interest is stationary, then the statistical properties computed for each of the sequence of short time intervals do not vary significantly from one time interval to the next. Here the word "significantly" means that variations are greater than would be expected because of statistical sampling variations. If the assumption is accepted, random data are tested for stationarity by investigating the behavior of individual sample records rather than an ensemble of sample records. In

simpler terms, the above assumption means that proof of self-stationarity for individual sample records, as defined in Section 1. 3. 3, can be accepted as proof of stationarity for the random process from which the records were obtained. Although one can contrive a hypothetical nonergodic stationary random process where this assumption is not valid, it is generally acceptable for random data representing actual random phenomena.

The second important assumption is: Verification of weak stationarity (time invariance of the mean value and autocorrelation function) is acceptable for most desired analyses and applications. If this assumption is accepted, then verification of stationarity can be restricted to investigations of only the mean value and autocorrelation functions for the data. This assumption is acceptable in most cases for two reasons, 1) important stationary-data-analysis procedures such as power-spectra and autocorrelation analysis are valid for weak stationarity, and 2) random data representing actual physical phenomena is generally strongly stationary if the data is weakly stationary. Note that for the case of data with a Gaussian probability density function, weakly stationary data is automatically strongly stationary since all higher order statistical properties of Gaussian data are determined by the mean value and autocorrelation function.

A third important assumption is: The sample record of the data to be investigated is very long compared to the random fluctuations of the data. This condition is necessary so that short time averages truly reflect average properties of the data and not just the random fluctuations inherent in the data. In other words, the sample record must be long enough to permit long term time trends to be differentiated from the time history fluctuations. For example, assume one is interested in determining if there are time trends (nonstationarity) in equipment operation. It is clear that long term time trends could not be detected from data collected over a period of only one hour. Note that the validity of this assumption is a function of the specific circumstances for each problem. If the assumption is not considered valid, the stationary

tests discussed here are not applicable, and the data must be studied in terms of ensemble averages.

A fourth important assumption which is not necessary, but which will simplify practical testing procedures is: If the mean square value (or variance) of the data of interest is stationary, then the autocorrelation function (or autocovariance function) for the data is also stationary. This assumption is certainly not as dependable as the first two assumptions. Nevertheless, it is usually valid for the following simple reason. It is highly unlikely for nonstationary data to have a time varying autocorrelation for any time displacement τ without the value at $\tau = 0$ varying. Since the mean square value is equal to the autocorrelation function at $\tau = 0$ (the variance equals the autocovariance at $\tau = 0$), attention can usually be restricted to the mean square value (or variance) rather than to the entire autocorrelation or autocovariance) function.

With the above assumptions in mind, the stationarity of random data can be tested by investigating a single sample record x(t) as follows:

1. Divide the sample record into N equal intervals where the data in each interval may be considered independent. For the case of relatively broadband data, these intervals are contiguous. If the data is narrow band or of low frequency relative to the individual interval length, it is necessary to allow a space between each interval to assure the data in each interval is effectively independent of adjacent intervals.

2. Compute a mean value and mean square value (or variance) for each interval and align these sample values in time sequence, as follows.

$$\bar{x}_1 , \bar{x}_2 , \bar{x}_3 , \ldots , x_N$$
$$\overline{x_1^2} , \overline{x_1^2} , \overline{x_3^2} , \ldots , \overline{x_N^2}$$

3. Test the sequence of mean and mean square (or variance) values for the presence of underlying trends or variations other than those due to normal sampling variations.

The final test of the sample values for nonstationary trends may be accomplished in many ways. If the sampling distribution for the values is known, various parametric statistical tests discussed in Reference 1, Section 6.1.8 and Reference 2, Section 16 can be applied. However, the sampling distribution for these tests requires a detailed knowledge of the frequency composition of the data. Such knowledge is generally unavailable at the time one wishes to establish whether or not data is stationary. Hence, a nonparametric approach which does not require a knowledge of sampling distributions for the data is desirable. Two such nonparametric tests which are applicable to this problem are the run test and the trend test of Sections 2 and 3. Either test may be applied directly as follows.

Assume that the sequence of mean values $(x_1, x_2, x_3, \ldots, x_N)$ and the sequence of mean square values $(x_1^2, x_2^2, x_3^2, \ldots, x_N^2)$ are each independent sample values of a random variable with a true mean value μ_x and mean square value ψ_x^2, respectively. If this assumption is true, the variations in the sequence of sample values are random and display no trends. Hence, the number of runs in the sequence relative to any given value (for simplicity, take the median value) will be as expected for a sequence of independent random observations of the same variable, as presented in Table 2. If the number of runs or reverse arrangements is significantly different from the expected number given in Tables 1 and 2, the hypothesis of stationarity is rejected. Otherwise, the hypothesis is accepted. The details of applying either test with numerical illustrations are presented in Examples 1 and 2. Note that the sample mean values and sample mean square values must be tested separately.

Several important features of these nonparametric tests for stationarity are:

1. A knowledge of the frequency bandwidth of the data under investigation is not required.

2. A knowledge of the exact averaging time used to measure the mean and mean square values is not required.

3. It is unnecessary for the data under investigation to be completely random. Valid conclusions are obtained even when sinusoidal components are present in the data, as long as the fundamental period is short compared to the averaging time used for each mean and mean square measurement.

4. TESTS FOR NORMALITY

The instantaneous amplitude values of a random signal are said to follow a normal or Gaussian distribution if their first-order probability density function is defined by the equation

$$p(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(y-\bar{y})^2}{2\sigma^2}\right] \tag{15}$$

where y denotes instantaneous amplitude values assumed by the random signals, $\bar{y}$ is the mean value of the random signal, and σ is the standard deviation of the random signal. In terms of p(y),

$$\bar{y} = \int_{-\infty}^{\infty} y\, p(y)\, dY \tag{16}$$

$$\sigma = \left[\int_{-\infty}^{\infty} (y-\bar{y})^2\, p(y)\, dy\right]^{1/2} \tag{17}$$

Also, p(y) satisfies the condition

$$\int_{-\infty}^{\infty} p(y)\, dy = 1 \tag{18}$$

A plot of p(y) is shown in Figure 4, and its values are tabulated in many standard references. Note that p(y) is symmetric about the mean value $y = \bar{y}$.

An assumption that the amplitude values of a random signal follow a normal distribution is made in many physical problems with good justification when the random signal is due to a combination of a large number of small effects. This is a consequence of the Central Limit Theorem and requires that no effect be dominant with respect to the other effects. It is true also that if a normally distributed random signal passes through a linear system,

$$\text{Prob}\left[y_1 < y \leq y_2\right] = \int_{y_1}^{y_2} p(y)\, dy = P_{1,2}$$

$$\text{Prob}\left[-\infty < y < \infty\right] = \int_{-\infty}^{\infty} p(y)\, dy = 1$$

$$p(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(y-\bar{y})^2}{2\sigma^2}\right]$$

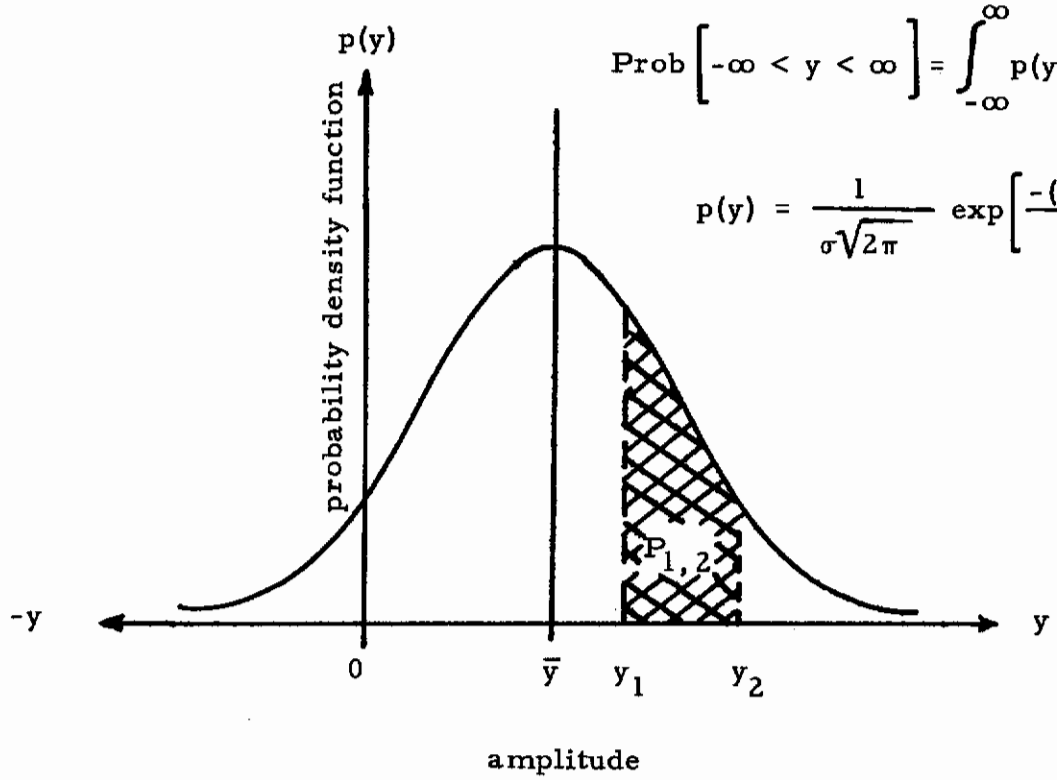


Figure 4. Gaussian Probability Density Plot

then the system output is normally distributed. Thus linear operations preserve normality. In particular, linear narrow band filtering operations tend to make non-normal random data more normal by virtue of the Central Limit Theorem, and keep normal random data normal by virtue of linearity.

When nonlinear operations are involved at any point in a system, the subsequent system output is no longer normal. The deviation from normality is a strong indicator of the nonlinearity. Peak signal distributions and extreme signal distributions are two important classes of distributions which by their fundamental nature are highly skewed and hence non-normal. Nonstationary trends in data because of transient conditions or underlying factors represent another class of situations which cause deviations from normality. Thus, non-normal data may be quite meaningful and would require special detailed analysis to determine precisely what factors are responsible for the calculated results.

Various error formulas for the statistical accuracy to be associated with correlation function measurements and spectral density function measurements assume that the random signal follows a normal distribution. These error formulas are approximations only for non-normal data and their practical limitations for non-normal data are unknown. When data is shown to be normal, many quantitative statements can be made about the range of expected variations of the random variable by using available tables. For non-normal data, where no mathematical formula is available, it is necessary to obtain experimental results to predict the range of expected variations. Clearly, the establishment of normality or lack of normality is an essential step for complete data analysis.

A practical statistical test for normality will now be outlined, called the Zero to Three Sigma Test. This test extends material in Section 17 of Reference 2, and is based upon measurements of the probability at seven equispaced amplitude levels from zero to three sigma. Since this test is an application of Hypothesis Tests, a discussion will be given first of these fundamental statistical ideas.

4.1 HYPOTHESIS TESTS

Consider the case where some estimator $\hat{\theta}$ is computed from a sample of N independent observations of a random variable x. Assume that there is reason to believe that the true parameter θ being estimated has a specific value θ_0. Now, even if $\theta = \theta_0$, the sample value $\hat{\theta}$ will probably not come out exactly equal to θ_0 because of the sampling variability associated with $\hat{\theta}$. Hence, the following question arises. If it is hypothesized that $\theta = \theta_0$, how much difference between $\hat{\theta}$ and θ_0 must occur before the hypothesis should be rejected as invalid? This question can be answered in statistical terms by considering the probability of any noted difference between $\hat{\theta}$ and θ_0 based upon the sampling distribution for $\hat{\theta}$. If the probability of a given difference is small, the difference is considered significant and the hypothesis that $\theta = \theta_0$ is accepted as normal statistical variability and the hypothesis that $\theta = \theta_0$ is accepted.

The above discussions outline the simplest form of a statistical procedure called hypothesis testing. To clarify the general technique, assume a sample value $\hat{\theta}$ has a probability density function of $p(\hat{\theta})$. Now, if a hypothesis that $\theta = \theta_0$ is true, then $p(\hat{\theta})$ has a mean value of θ_0 as illustrated in Figure 5.

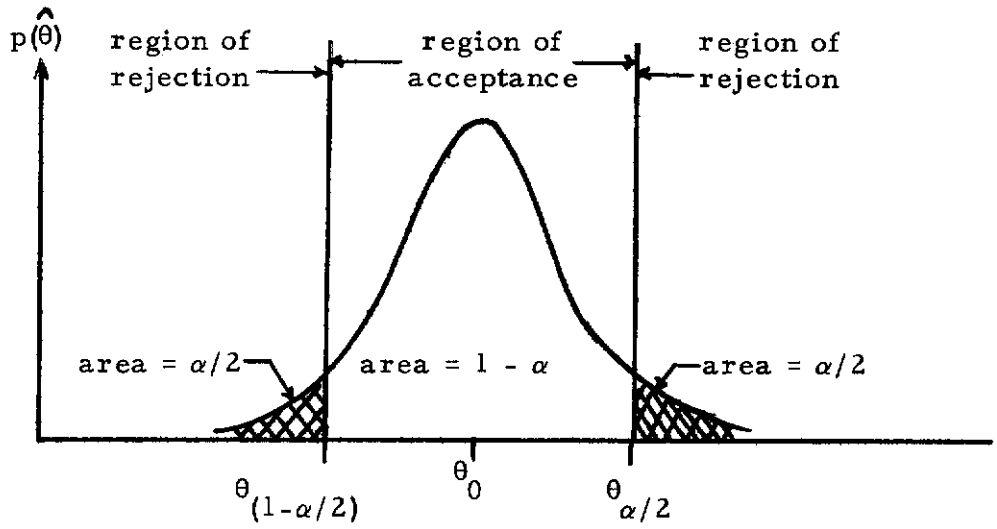


Figure 5. Acceptance and Rejection Regions for Hypothesis Tests

The probability that $\hat{\theta}$ falls below the lower value $\theta_{(1-\alpha/2)}$ is

$$P\left[\hat{\theta} \leq \theta_{(1-\alpha/2)}\right] = \int_{-\infty}^{\theta_l} p(\hat{\theta})\, d\hat{\theta} = P(\hat{\theta}_{1-\alpha/2}) = \alpha/2 \tag{19}$$

The probability that $\hat{\theta}$ falls above the upper value $\theta_{\alpha/2}$ is

$$P\left[\hat{\theta} \geq \theta_{\alpha/2}\right] = \int_{\theta_{\alpha/2}}^{\infty} p(\hat{\theta})\, d\hat{\theta} = 1 - P(\theta_{\alpha/2}) = \alpha/2 \tag{20}$$

Hence, the probability, that $\hat{\theta}$ is outside the range between $\theta_{(1-\alpha/2)}$ and $\theta_{\alpha/2}$, is α. Now, let α be small so that it is very unlikely that θ would fall outside the range between $\theta_{(1-\alpha/2)}$ and $\theta_{\alpha/2}$. If a sample is collected and a value of $\hat{\theta}$ is computed which in fact falls outside the range between $\theta_{(1-\alpha/2)}$ and $\theta_{\alpha/2}$, there is strong reason to question the original hypothese that $\theta = \theta_0$ because such a value for θ is very unlikely. Hence, the hypothesis that $\theta = \theta_0$ would be rejected. On the other hand, if the value θ falls within the range between $\theta_{(1-\alpha/2)}$ and $\theta_{\alpha/2}$, there is no strong reason to question the original hypothesis. Hence, the hypothesis that $\theta = \theta_0$ would be accepted.

The small probability α used for the hypothesis test is called the level of significance for the test. The range of values for $\hat{\theta}$ where the hypothesis is not accepted is called the region of rejection or critical region. The range of values of $\hat{\theta}$ where the hypothesis is accepted is called the region of acceptance. The simple hypothesis test outlined above is called a two-sided test because, if the hypothesis is not true, the value of θ could be either greater or less than θ_0. Hence, it is necessary to test for significant differences between θ and θ_0 in both directions. In other cases, a one-sided test might be sufficient.

For example, assume that $\theta \geq \theta_0$. For this case, the hypothesis would be false only if θ is less than θ_0. Hence, the test is performed using the lower side of the probability density function for $p(\hat{\theta})$.

There are two possible errors which can occur when a hypothesis test is performed. First, the hypothesis might be rejected when in fact it is true. This possible error is called a Type I Error. Second, the hypothesis might be accepted when in fact it is false. This possible error is called a Type II Error. From Figure 5, a Type I Error occurs if the hypothesis were true and θ falls in the region of rejection. Hence, the probability of a Type I Error is simply equal to α, the level of significance for the test.

In order to establish a probability of a Type II Error, it is necessary to specify some deviation of the true parameter θ from the hypothesized parameter θ_0 which one desires to detect. For example, assume that the true parameter actually has a value of $\theta = \theta_0 \pm d$, as illustrated in Figure 6.

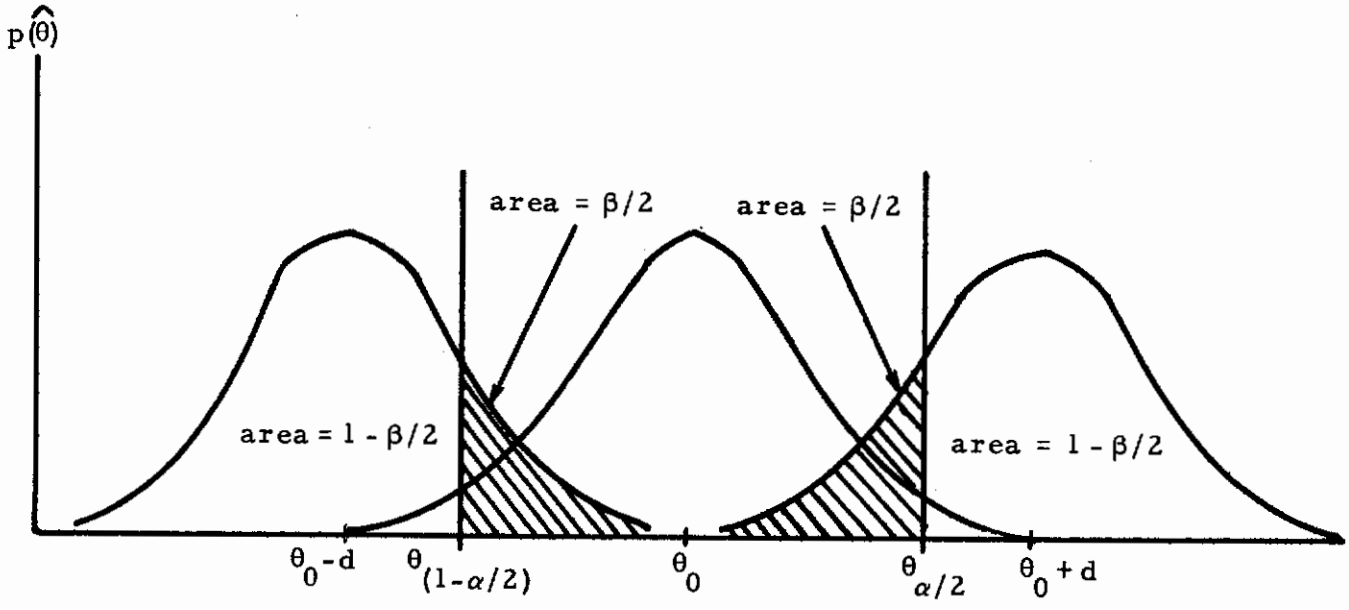


Figure 6. Type II Error Regions for Hypothesis Tests

If it is hypothesized that $\theta = \theta_0$ when in fact $\theta = \theta_0 \pm d$, the probability that θ falls inside the acceptance region between $\theta_{(1-\alpha/2)}$ and $\theta_{\alpha/2}$ is β. Hence, the probability of a Type II Error is β for detecting a difference of $\pm d$ from the hypothesized value θ_0.

The probability $1-\beta$ is called the power of the test. Clearly, for any given sample size N, the probability of a Type I Error can be reduced by reducing the level of significance α. However, this will increase the probability β of a Type II Error (reduce the power of the test). The only way to reduce both α and β is to increase the sample size N for the estimate θ. These ideas form the basis for selecting the necessary sample sizes for statistical experiments.

4.2 ZERO TO THREE SIGMA TEST

Assume a signal y(t) is stationary and has a true normal probability density function p(y) with a mean value of zero and a standard deviation of σ, namely,

$$p(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-y^2}{2\sigma^2}\right) \tag{21}$$

Assume that the probability density function of this signal is symmetric about $y = 0$ and is measured at seven different positive amplitude values starting at $y = 0$ which are separated 0.5σ apart. These seven values

$$y = 0,\ 0.5\sigma,\ 1.0\sigma,\ 1.5\sigma,\ 2.0\sigma,\ 2.5\sigma,\ 3.0\sigma$$

vary from zero to three sigma and account for the name of the test. It is assumed that the seven measurements are statistically independent.

If the probability density function at each amplitude is tested at the α level of significance, then the probability of correctly accepting a normality hypothesis at each amplitude is $(1-\alpha)$. The probability of accepting the normality hypothesis at all seven amplitudes is $(1-\alpha)^7$. It follows that the probability of not accepting at least one of the seven amplitudes (making at least one Type I Error) is given by

$$\alpha' = 1 - (1-\alpha)^7 \tag{22}$$

Thus, if a Type I Error probability of α' is desired for the over-all test, then each of the seven amplitudes must have an individual Type I Error probability of α where

$$\alpha = 1 - (1-\alpha')^{1/7} \tag{23}$$

Corresponding values of α and α' are shown in Table 3.

TABLE 3

Over-all and Individual Type I Errors

α'	0.007	0.048	0.100	0.152	0.204
α	0.001	0.007	0.015	0.023	0.032

Consider now a sample record of length T which is to be tested for normality. The first requirement is to test the record for stationarity. This can be done by nonparametric techniques which do not require bandwidth information. Two such practical tests are described in Sections 3.1 and 3.2,

(1) Run Test

(2) Trend Text

The next step is to test the record for randomness and the presence of periodicities. This can be done qualitatively by visual inspection of the data and certain of its basic amplitude, time, and frequency properties, as discussed in Section 2. A quantitative test for randomness is derived and experimentally verified in Chapter 15 of Reference 2. This test requires assumptions or explicit knowledge about the bandwidth B occupied by the sample signal. A crude estimate of the bandwidth B should therefore be included as part of the data analysis procedure by using appropriate filtering techniques. Any obvious periodicities in the data should be filtered out from the data before conducting further detailed analysis.

From experimental tests carried out in Chapter 14 of Reference 2, the standard error (normalized standard deviation) associated with a particular instrument used in measuring amplitude probability density functions was found to be

$$\epsilon_y = \frac{k}{\sqrt{BTW\hat{p}(y)}} \tag{24}$$

where

B = signal bandwidth

T = record length

W = amplitude window

$\hat{p}(y)$ = estimated probability density function at amplitude y

k = experimental coefficient ≈ 0.20

The coefficient $k \approx 0.2$ is associated with the particular instrument and is deemed to be typical of coefficients which are obtained with other probability density measuring instruments. See Figure 7 for a plot of Eq. (24) when $k = 0.20$ as a function of BTW for various $\hat{p}(y)$.

It is shown also in Reference 2, Chapter 14, that probability density measurements may be considered normally distributed about the true probability density at any specified amplitude. Hence, an estimate $\hat{p}_v \equiv \hat{p}(v)$ for a specific amplitude $y = v$ has a normal sampling distribution about the true value $p_v = p(v)$ such that

$$\text{Prob}\left[-(\epsilon_v p_v)\, z_{\alpha/2} \leq (\hat{p}_v - p_v) \leq (\epsilon_v p_v)\, z_{\alpha/2}\right] = 1 - \alpha \qquad (25)$$

where α lies in the range $0 < \alpha < 1$, where $z_{\alpha/2}$ is the normal deviate corresponding to α. For example, at $\alpha = 0.05$, (5%), the quantity $z_{\alpha/2} = 1.96$.

Assume that the sample record follows a normal distribution $p_{0y} = p_0(y)$ with a mean value of zero. Let the hypothesis H_0 at any amplitude level $y = v$ b

$$H_0 : \hat{p}_v = p_{0v} \qquad (26)$$

where $\hat{p}_v$ is the measured value, and p_{0v} is the corresponding normal value. The first step in testing H_0 at $y = v$ is to compute ϵ_v in Eq. (24) using p_{0v} in place of $\hat{p}(v)$. Then H_0 is tested at the α level of significance by accepting H_0 if the difference $\left|\hat{p}_v - p_{0v}\right|$ falls inside the range

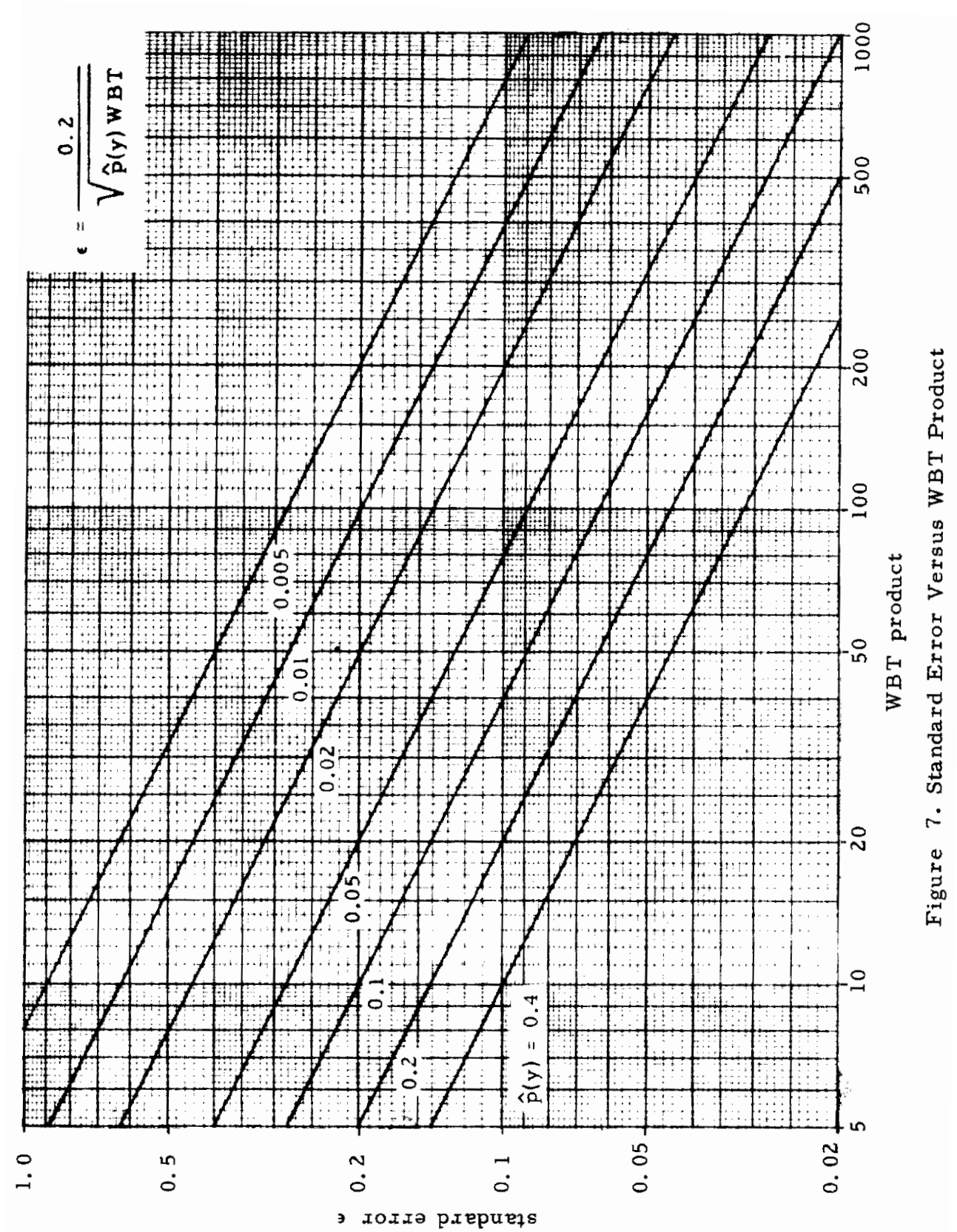


Figure 7. Standard Error Versus WBT Product

$$\left| \hat{p}_v - p_{0v} \right| \leq (\epsilon_v p_{0v})\, z_{\alpha/2} \tag{27}$$

where

$$\epsilon_v = \frac{k}{\sqrt{BTW\, p_{0v}}} \tag{28}$$

This hypothesis is tested for as many different amplitude levels as desired, and the Type I Error for each level tested is α.

By hypothesis, since p_{0v} has a zero mean value

$$p_{0v} = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-v^2/2\sigma^2} \tag{29}$$

Substitution of Eq. (15) into Eq. (14) and solving for ϵ_v^2 yields

$$\epsilon_v^2 = \frac{k^2}{BTW\, p_{0v}} = \frac{k^2 \sigma \sqrt{2\pi}\, e^{v^2/2\sigma^2}}{BTW} \tag{30}$$

Thus, for given B, T, W, and σ, it is clear that ϵ_v increases as v increases. In the zero to three sigma test for normality, ϵ_v will be calculated only at the seven levels $v = 0,\ 0.5\sigma,\ 1.0\sigma,\ 1.5\sigma,\ 2.0\sigma,\ 2.5\sigma$, and 3.0σ. If it is desired to keep ϵ_v constant as v increases, this can be accomplished by increasing T so as to satisfy the formula

$$T = \frac{k^2}{BW\, p_{0v}\, \epsilon_v^2} = \frac{k^2 \sigma \sqrt{2\pi}\, e^{v^2/2\sigma^2}}{BW\, \epsilon_v^2} \tag{31}$$

From a table of normal ordinates, assuming the data is normalized so that σ = unity, the seven chosen amplitude values have the probability density values p_{0v} and reciprocal values $(1/p_{0v})$ listed in Table 2.

TABLE 4

Normal Probability Density Values

v	0	0.5	1.0	1.5	2.0	2.5	3.0
p_{0v}	0.399	0.352	0.242	0.130	0.0540	0.0175	0.0044
$1/p_{0v}$	2.51	2.84	4.13	7.72	18.52	57.14	227.3

The probability of a Type II Error, β, is a function of the level of significance α, the standard deviation $\epsilon_v p_{0v}$, and the non-Gaussian probability density p_v which one wishes to detect with a probability of β. Unfortunately, for cases where one assumes a normal sampling distribution, the risk of a Type II Error for detecting a difference $\left| p_v - p_{0v} \right| = (\epsilon_v p_{0v})\, z_{\alpha/2}$ is always $\beta = 0.50$. This point is illustrated in Figure 5. It is clear that $\beta = 0.50$ since exactly half of the normal distribution is covered at either extreme.

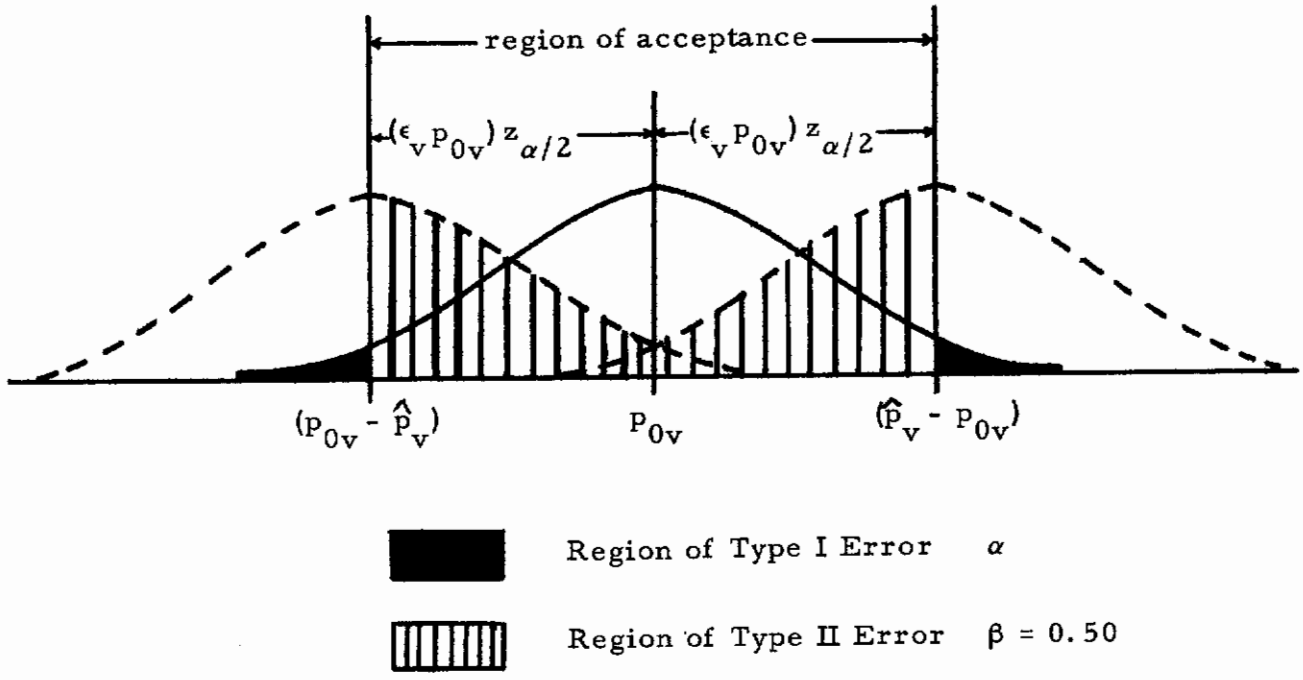


Figure 8. Probable Type I and Type II Errors

It is important to note that the α and β risks are determined by $(\epsilon_v p_{0v})$ which in turn is determined by the record length T and the amplitude $y = v$ at which the test is to be applied. Thus, if a predetermined α and β risk are desired, these matters must be considered to arrive at the necessary record length T to be obtained.

4.3 RECOMMENDED PROCEDURE

A recommended procedure for applying the zero to three sigma test for normality is as follows:

1. Decide upon the over-all desired Type I Error α' and the corresponding individual Type I Error α; use Eq. (22) or Table 3.

2. Use a table of normal deviates and determine the value for the normal deviate $z_{\alpha/2}$.

3. Decide upon the deviation from normality $\left|\hat{p}_v - p_{0v}\right|$ which is to be detected at each amplitude with a Type II Error $\beta = 0.50$. (A practical value to choose is $\pm$ 25% deviation from normality at any of the seven amplitudes to be tested, namely, $\left|\hat{p}_v - p_{0v}\right| = 0.25 p_{0v}$.)

4. Determine the required value for the standard error ϵ_v from the equation

$$\epsilon_v = \frac{\left|\hat{p}_v - p_{0v}\right|}{p_{0v} z_{\alpha/2}} \tag{32}$$

Note: for $\hat{p}_v - p_{0v} = 0.25\, p_{0v}$, this equation is independent of p_{0v} and is simply

$$\epsilon_v = \frac{0.25}{z_{\alpha/2}} \tag{33}$$

5. Determine the necessary value of T from Eq. (31) as a function of the probability density p_{0v} ; use ϵ_v calculated in Step 4. This requires knowledge of the parameters B, W, and k. If the available record length $T_r \geq T$, then use only a subrecord of length T. However, if the available record length $T_r < T$ specified from Eq. (31), then it is necessary to compute the appropriate error ϵ_v from Eq. (30). Equation (32) then gives $\epsilon_v z_{\alpha/2}$ as the percentage deviation from normality which is being detected with a Type II Error of $\beta = 0.50$.

6. Measure the probability density estimate $\hat{p}_v$ at the desired seven values. The acceptance region for the hypothesis H_0 at each of the seven values is given by Eq. (27).

7. If a rejection is obtained at any one of the seven values, the sample record was probably not obtained from a signal with a normal probability density function.

4.4 NUMERICAL EXAMPLE

Assume a signal representing a stationary record occupies a bandwidth B = 1000 cps. Assume the amplitude window width W of the probability density analyzer is equal to W = 0.10 for an input signal level of σ = unity, and that k = 0.20 is an acceptable value for k. Assume the probability density function is symmetric about $\bar{y} = 0$ and that seven positive amplitude levels are tested spaced 0.50 units apart from 0 to 3.0. The Type I Error probability for the entire test is $\alpha' = 0.10$, and a 25% deviation from normality at any of the seven amplitudes is detected with a Type II Error of $\beta = 0.50$.

From Table 3, the required level of significance for the test at each level is found to be $\alpha = 0.015$. From a table of normal deviates, the value $z_{\alpha/2} = 2.43$ when $\alpha = 0.015$. The determination of α and $z_{\alpha/2}$ are

Steps 1 and 2 in the recommended procedure. Step 3 was taken care of in the problem statement. The associated value of ϵ_v (Step 4) is here found to be

$$\epsilon_v = \frac{0.25}{2.43} \approx 0.10$$

The required record lengths can now be calculated as a function of p_{0v} from Eq. (31), namely,

$$T = \frac{k^2}{BW\, p_{0v} \epsilon_v^2} = \frac{(0.20)^2}{(1000)(0.10)(0.10)^2 p_{0v}} = \frac{0.04}{p_{0v}}$$

At the maximum amplitude value $v = 3.0$, the required T from Table 4 is

$$T = (0.04)(227) \approx 9.0 \text{ seconds}$$

Thus, a sample record of at least $T_r = 9$ seconds is required to perform this test at the level $v = 3.0$ if the Type I Error is to be $\alpha = 0.10$, and the percentage deviation from normality which is detected is $\pm 25\%$ with a Type II Error of $\beta = 0.50$. At the lower amplitude values $v = 0, 0.50, 1.0, 1.5, 2.0$, and 2.5, the required averaging times are, respectively,

$$T = 0.10,\ 0.12,\ 0.16,\ 0.31,\ 0.75,\ 2.3 \text{ seconds}$$

Assuming that $T_r \geq 9.0$ seconds, then the probability density estimate $\hat{p}_v$ must be measured at each of the seven amplitude values using the appropriate averaging times in each case. The acceptance region at each amplitude v is

$$\left| \hat{p}_v - p_{0v} \right| \leq (\epsilon_v z_{\alpha/2})\, p_{0v} \approx 0.25\, p_{0v}$$

If $T_r < 9.0$ seconds, say $T_r = 4.0$ seconds, then all of the amplitude values except the maximum value $v = 3.0$ can be tested by the above procedure. For the value $v = 3.0$, a different standard error will apply as determined by Eq. (30),

$$\epsilon_{3.0} = \frac{k}{\sqrt{BTW\, p_{0v}}} = \frac{0.20}{\sqrt{(1000)(4.0)(0.10)(0.0044)}} \approx 0.15$$

It follows that $\epsilon_{3.0}\, z_{\alpha/2} = (0.15)(2.43) \approx 0.37$, and the acceptance region at the value $v = 3.0$ becomes

$$\left| \hat{p}_3 - p_{03} \right| \leq 0.37\, p_{03}$$

For this point, the Type I Error will still be $\alpha = 0.10$. However, the percentage deviation from normality which is being detected will now be $\pm 37\%$ with a Type II Error of $\beta = 0.50$.

5. COMPARISON OF SPECTRA

Assume that two signals with distinct power spectra $G_1(f)$ and $G_2(f)$ are obtained under different conditions. Assume that these spectra have the same average value over a given frequency interval (f_a, f_b) of bandwidth $B = f_b - f_a$. Then

$$\sigma_1^2(f_a, f_b) = \int_{f_a}^{f_b} G_1(f)\, df = \int_{f_a}^{f_b} G_2(f)\, df = \sigma_2^2(f_a, f_b) \tag{34}$$

where $\sigma_{1,2}^2(f_a, f_b)$ denotes the mean square value of the two signals over the frequency interval (f_1, f_2). Equation (34) represents, in a sense, a normalizing condition for the two spectra.

Confidence intervals for the single spectra $G_1(f)$ or $G_2(f)$ alone can be described in terms of their sample mean values and sample variances. For a large number of degrees-of-freedom, these sampling distributions will be approximately normal. Comparison of the two spectra can also be carried out to determine statistically whether or not they differ by more than a reasonable amount. In particular, one can determine whether a given spectrum should be considered equivalent to a "white" spectrum over (f_a, f_b). By "white" spectrum is meant that $G(f) = c$, a constant over (f_a, f_b). For this special case,

$$\sigma^2(f_a, f_b) = \int_{f_a}^{f_b} c\, df = c(f_b - f_a) = cB \tag{35}$$

Thus,

$$G(f) = \frac{\sigma^2(f_a, f_b)}{B} \tag{36}$$

indicates how G(f) may be determined from $\sigma^2(f_a, f_b)$ and B when G(f) is "white." Various methods for comparing spectra will now be discussed.

5.1 MEAN SQUARE ERROR CRITERIA

A mean square error criteria is first derived to determine when one spectrum is more white than another. Define the mean square error ϵ^2 for arbitrary G(f) by

$$\epsilon^2 = \int_{f_a}^{f_b} \left[G(f) - c\right]^2 df \tag{37}$$

Clearly, $\epsilon^2 \geq 0$ for all G(f) and $\epsilon^2 = 0$, its minimum value, if G(f) = c inside (f_a, f_b). Also, it is clear that the closer G(f) approximates c inside (f_a, f_b), the closer ϵ^2 approaches zero. In fact, the minimum value of $\epsilon^2 = 0$ occurs if and only if G(f) is white noise inside (f_a, f_b).

A reasonable method for comparing the "whiteness" of $G_1(f)$ to $G_2(f)$ is now obvious. One has merely to compute:

$$\epsilon_1^2 = \int_{f_a}^{f_b} \left[G_1(f) - c\right]^2 df$$

$$\epsilon_2^2 = \int_{f_a}^{f_b} \left[G_2(f) - c\right]^2 df \tag{38}$$

and see which gives the smaller value. The smaller value corresponds to the spectrum which is the more "white" of the two.

This procedure is not completely rigorous because of statistical sampling fluctuations that occur in measurements of $G_1(f)$ and $G_2(f)$. One can obtain different results for $G_1(f)$ and $G_2(f)$ even though they are from the

same population, and this in turn gives different results for ϵ_1^2 and ϵ_2^2. To decide whether or not $G_1(f)$ and $G_2(f)$ are equivalent, the following test for equivalence of two spectra should be used. A special case of this procedure is to let $G_1(f)$ represent "white" noise and test whether or not $G_2(f)$ is equivalent to this $G_1(f)$.

5.2 CONFIDENCE INTERVAL FOR SPECTRAL ESTIMATES

From Reference 1, it is known that estimates $\left\{\hat{G}(f)\right\}$ of a true power spectral density function $G(f)$ have a mean value approximately equal to $G(f)$ and a variance

$$\mathrm{Var}\left[\hat{G}(f)\right] \approx \frac{2}{n} G^2(f) \tag{39}$$

where n is the number of degrees-of-freedom associated with the estimate. For analog processing of data, $n = 2B_e T$ where B_e is the equivalent bandwidth of the narrow resolution filter and T is the record length. It follows that for n sufficiently large, a $(1 - \alpha)$ confidence interval for $G(f)$ may be obtained from

$$\left[\hat{G}(f)\left(1 - z_{\alpha/2}\sqrt{2/n}\right) \leq G(f) \leq \hat{G}(f)\left(1 + z_{\alpha/2}\sqrt{2/n}\right)\right] \tag{40}$$

where $z_{\alpha/2}$ is the $100(\alpha/2)$ percentage point of the normal distribution.

The logarithmic transformation of $\hat{G}(f)$ yields a random variable $\log \hat{G}(f)$ with mean value approximately equal to $\log G(f)$ and variance

$$\mathrm{Var}\left[\log \hat{G}(f)\right] \approx \frac{2}{n} \tag{41}$$

Thus, the sampling properties of $\left\{\log \hat{G}(f)\right\}$ are independent of frequency, and tend to be more nearly normal than $\left\{\hat{G}(f)\right\}$. A heuristic proof for Eq. (41) is obtained by noting that the logarithm of the terms inside Eq. (40) yields

$$\left[\log\ \hat{G}(f) + \log\left(1 - z_{\alpha/2}\sqrt{2/n}\right) < \log G(f) \leq \log\ \hat{G}(f) + \log\left(1 + z_{\alpha/2}\sqrt{2/n}\right)\right] \quad (42)$$

Also, for $z_{\alpha/2}\sqrt{2/n} \ll 1$,

$$\log\left(1 \pm z_{\alpha/2}\sqrt{2/n}\right) \approx \pm z_{\alpha/2}\sqrt{2/n} \quad (43)$$

Hence, a $(1 - \alpha)$ confidence interval for log G(f) is estimated by

$$\left[\log\ \hat{G}(f) - z_{\alpha/2}\sqrt{2/n} \leq \log G(f) \leq \log \hat{G}(f) + z_{\alpha/2}\sqrt{2/n}\right] \quad (44)$$

which for a normal distribution implies Eq. (41). An equivalent expression to Eq. (44) shows that G(f) has the $(1 - \alpha)$ confidence interval

$$\left[\exp\left(\log \hat{G}(f) - z_{\alpha/2}\sqrt{2/n}\right) \leq G(f) \leq \exp\left(\log \hat{G}(f) + z_{\alpha/2}\sqrt{2/n}\right)\right] \quad (45)$$

5.3 TEST FOR EQUIVALENCE OF TWO SPECTRA

Assume now that two spectra $G_1(f)$ and $G_2(f)$ have been obtained under different conditions, filtered, and normalized so as to satisfy Eq. (1). Assume also that both spectra are analyzed using the same number of degrees-of-freedom, and that the frequency interval (f_a, f_b) of bandwidth B is divided into ℓ parts where

$$\ell = \frac{B}{B_e} \tag{46}$$

Assume that ℓ statistically independent spectral estimates of $G_1(f)$ and $G_2(f)$ are obtained from these ℓ parts, and that the number of degrees-of-freedom is reasonably large to justify a normal sampling distribution.

The test for equivalence is carried out by computing

$$D = \frac{\sqrt{n}}{2\sqrt{\ell}} \sum_{i=1}^{\ell} \log \frac{\hat{G}_1(f_i)}{\hat{G}_2(f_i)} \tag{47}$$

and comparing its value with the normal distribution of zero mean and unit standard deviation. To be specific, if $G_1(f)$ and $G_2(f)$ are equivalent, then there is $(1 - \alpha)$ probability that D will fall inside the range

$$\left[-z_{\alpha/2} \leq D \leq z_{\alpha/2}\right] \tag{48}$$

For example, if the coefficient for a level of significance (Type I Error) is $\alpha = 0.01$, there is 99% probability that D will fall inside the range

$$\left[-2.58 \leq D \leq 2.58\right] \tag{49}$$

Hence, if D falls outside the range (-2.58, 2.58), it is reasonable to reject the hypothesis that $G_1(f)$ is equivalent to $G_2(f)$. There is at most a 1% risk that one is rejecting a true hypothesis. If D falls inside the range (-2.58, 2.58), the hypothesis is accepted.

The derivation of Eq. (47) is based upon the following considerations. For each frequency f_i,

$$\begin{aligned} \log \hat{G}_1(f_i) &\sim N\left[\log G_1(f_i) \ , \quad 2/n\right] \\ \log \hat{G}_2(f_i) &\sim N\left[\log G_2(f_i) \ , \quad 2/n\right] \end{aligned} \tag{50}$$

where $N\left[\mu, \sigma^2\right]$ represents a normal distribution with mean value μ and variance σ^2, while $\sim$ denotes "is distributed as." Now,

$$\log \frac{\hat{G}_1(f_i)}{\hat{G}_2(f_i)} \sim N\left[0, 4/n\right] \tag{51}$$

since it is the difference of the two quantities in Eq. (50). Next, as the average sum of ℓ identically distributed quantities,

$$\frac{1}{\ell} \sum_{i=1}^{\ell} \log \frac{\hat{G}_1(f_i)}{\hat{G}_2(f_i)} \sim N\left[0, 4/n\ell\right] \tag{52}$$

Dividing by $2/\sqrt{n\ell}$ then gives the stated result

$$D \sim N(0, 1) \tag{53}$$

5.4 TEST FOR "WHITE" SPECTRUM

Assume that $G_1(f)$ is a white spectrum inside the interval (f_a, f_b) where $B = f_b - f_a$. Then

$$G_1(f) = c = \frac{\sigma^2(f_1, f_2)}{B} \tag{54}$$

For this special case, Eq. (47) becomes

$$D = \frac{\sqrt{n}}{2\sqrt{\ell}} \sum_{i=1}^{\ell} \log \frac{c}{G_2(f)}$$

$$= \frac{\sqrt{n\ell}\,\log c}{2} - \frac{\sqrt{n}}{2\sqrt{\ell}} \sum_{i=1}^{\ell} \log \hat{G}_2(f_i) \tag{55}$$

As before, D should be computed and the hypothesis that $G_2(f)$ is a white spectrum accepted at the 1% level of significance if D falls inside the range (-2.58, 2.58). Otherwise, the hypothesis should be rejected.

REFERENCES

1. Bendat, J. S., Enochson, L. D., Klein, G. H., and A. G. Piersol, "The Application of Statistics to the Flight Vehicle Vibration Problem," ASD TR 61-123, Aeronautical Systems Division, AFSC, USAF, Wright-Patterson AFB, Ohio. December 1961. (AD 271 913)

2. Bendat, J. S., Enochson, L. D., Klein, G. H., and A. G. Piersol, "Advanced Concepts of Stochastic Processes and Statistics for Flight Vehicle Vibration Estimation and Measurement," ASD TR 62-973, Aeronautical Systems Division, AFSC, USAF, Wright-Patterson AFB, Ohio. December 1962. (AD 297 031)

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Measurement Analysis Corporation
10962 Santa Monica Blvd.
Los Angeles, California 90025

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP

3. REPORT TITLE
Tests for Randomness, Stationarity, Normality, and Comparison of Spectra

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Final

5. AUTHOR(S) (Last name, first name, initial)
Bendat, Julius S.
Enochson, Loren D.
Piersol, Allan G.

6. REPORT DATE
August 1965

7a. TOTAL NO. OF PAGES
45

7b. NO. OF REFS
2

8a. CONTRACT OR GRANT NO.
AF 33(615)-1314

b. PROJECT NO.
4437

c.
Task No. 443706

d.

9a. ORIGINATOR'S REPORT NUMBER(S)
AFFDL-TR-65-74

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
MAC 402-03A

10. AVAILABILITY/LIMITATION NOTICES
Qualified requesters may obtain copies of this report from DDC. This report has been furnished to CFSTI for sale to the public

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
Air Force Flight Dynamics Laboratory
Wright-Patterson AFB, Ohio 45433

13. ABSTRACT

This report contains material for carrying out certain practical tests for randomness, stationarity, and normality of physical data, as well as details on how to test for equivalence of two power spectral density functions. The tests for randomness are qualitative inspections of measured properties from sample records, such as power spectra, probability density, auto-correlation functions, which may be part of a regular data reduction procedure. The tests for stationarity are nonparametric statistical procedures, based upon Run and Trend Distributions, which require simple quantitative calculations for acceptance or rejection. The test for normality is a statistical hypothesis test based upon sample measurements of the amplitude probability density function at seven equispaced amplitude levels from zero to plus and minus three sigma. To test for equivalence of two spectra, a straight-forward procedure is recommended which requires knowledge only of the number of degrees-of-freedom associated with the spectral estimates, and the full bandwidth occupied by the data. A special case of this procedure is to compare an unknown spectrum to a "white" spectrum.

DD FORM 1 JAN 64 1473

UNCLASSIFIED
Security Classification

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Mathematics Statistics Spectral Analysis Random Processes						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through _____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS), (S), (C), or (U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.

UNCLASSIFIED

Security Classification