AN ANALYSIS OF AMPLITUDE PROBABILITY MEASUREMENTS

G. P. THRALL

FOREWORD

This report was prepared by Measurement Analysis Corporation, Los Angeles, California, for the Aerospace Dynamics Branch, Vehicle Dynamics Division, AF Flight Dynamics Laboratory, Wright-Patterson Air Force Base, Ohio, under Contract AF33(615)-1418. The research performed is part of a continuing effort to provide advanced techniques in the application of random process theory and statistics to vibration problems which is part of the Research and Technology Division, Air Force Systems Command's exploratory development program. The contract was initiated under Project No. 1370, "Dynamic Problems in Flight Vehicles", Task No. 137005, "Prediction and Control of Structural Vibration". Mr. R. G. Merkle of the Aerospace Dynamics Branch, Vehicle Dynamics Division, AF Flight Dynamics Laboratory, was the project engineer.

This report covers work conducted from March 1964 to December 1964. Contractors report number is MAC 403-01.

The manuscript for this report was released by the authors in January 1965 for publication as an FDL Technical Report.

This report has been reviewed and is approved.

Walter J. Mykytow

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

ABSTRACT

Techniques for measuring amplitude probabilities and probability densities are summarized and previous results on the statistical uncertainty of such measurements are reviewed. A rigorous mathematical model of probability measurements is derived. It is shown that unknown correlations in the parameters of the model make it impossible to develop explicit expression for the mean square estimation error. Results are presented of a computer simulation of amplitude probability estimates and comparisons are made between experimental and computed mean square errors.

CONTENTS

AN ANALYSIS OF AMPLITUDE PROBABILITY MEASUREMENTS

1. INTRODUCTION

In ASD-TDR-62-973, Ref. [1], two methods are discussed for determining the mean square error in the estimation of amplitude probability densities from a record of finite duration. In addition, this report presents results of experiments performed in an attempt to verify the two analytic expressions. These experiments were performed quite carefully. However, to quote from ASD-TDR-62-973, "Neither of the two theoretical uncertainty expressions considered appears to be completely valid for all the conditions studied." The purpose of this report is to clarify several points concerning probability density estimates, and to present a more rigorous evaluation of amplitude probability measurements.

Section 2 presents a brief summary of the definition of amplitude probabilities and associated measurement techniques. In Section 3, the analytical work on amplitude probability estimates which was presented in ASD-TDR-62-973 is reviewed and the limitations discussed. A more rigorous approach to the evaluation of amplitude probability measurements is given in Section 4. It is shown that the samples are not independent but may be correlated in an unknown way; thus, making it impossible to develop an explicit expression for the mean square estimation error. In Section 5, the analysis is extended to probability density estimates and a comparison is made between the new analysis and the previous results. Section 6 presents the results of a computer simulation of amplitude probability estimates and comparisons are made between experimental and computed mean square errors. Section 7 briefly reviews the conclusions reached.

2. AMPLITUDE PROBABILITY MEASUREMENTS

Consider a stationary random process $X(t)$. The probability that $X(t)$ lies in an amplitude interval (a, b) at any time t is defined as the fraction of the total time that $a \leq X(t) \leq b$. Symbolically, this may be expressed by

$$p = P\left[a \leq X(t) \leq b\right] = \lim_{T \to \infty} \frac{1}{T} \sum_i \tau_i \qquad (1)$$

where τ_i is the length of the ith interval in which $X(t)$ is between a and b. The notation $P\left[\cdot\right]$ will always mean the probability of the event described between the square brackets. Since any physical measurement cannot extend over an infinite time, the probability p is estimated experimentally be averaging over a finite time. Thus, if $\hat{p}$ denotes the estimate of p, it follows that

$$\hat{p} = \frac{1}{T} \sum_i \tau_i \qquad (2)$$

The averaging operation may be accomplished with analog techniques by summing each of the τ_i with a clock, or through digital methods by sampling $X(t)$ with very narrow pulses and counting the number which fall within (a, b). Since each type of device performs essentially the same operation, i.e., summing time intervals, the particular measuring device will not be of concern in this report.

To evaluate how close $\hat{p}$ is to p for a given record length, it is desirable to know the statistical properties of the time intervals, τ_i. Unfortunately, the time statistics of a random process are very difficult, if not impossible, to obtain so that an alternate analytic approach must be employed to determine the errors involved in the measurement over a finite time interval.

A quantity of great practical importance is the probability density function which is given by

$$f(x) = \lim_{(b-a)\to 0} \frac{P\left[a \le X(t) \le b\right]}{b - a} \tag{3}$$

where x lies in the interval (a, b). In Eq. (3), the procedure of taking the limits as (b - a) approaches zero is beyond the capability of physical instruments. However, if (b - a) is sufficiently small, the probability density may be approximated by

$$\hat{f}(x) = \frac{\hat{p}}{b - a} = \frac{1}{(b - a)T} \sum_i \tau_i \tag{4}$$

Thus, an error analysis of the estimation of p by $\hat{p}$ is equivalent to the estimation of $f(x)$ by $\hat{f}(x)$, the essential difference being division by the scale factor (b - a). There is one small difference between the errors associated with the estimation of p and $f(x)$; $\hat{f}(x)$ is slightly biased whereas $\hat{p}$ is unbiased. This effect will be taken up in Section 5 which deals with probability density estimates.

3. REVIEW OF PREVIOUS ERROR ESTIMATION TECHNIQUES

In Ref. [1], the mean square error in estimating the probability density at an amplitude x is defined by

$$\epsilon^2 = \frac{\sigma_p^2}{N} \tag{5}$$

where σ_p^2 is the population variance and N is the equivalent number of events upon which the estimate is based. The underlying assumption in Eq.(5) is that the equivalent number of events are statistically independent. As will be indicated later, this is generally not the case.

The numerical value of N was estimated in two ways; each of which will now be discussed.

3.1 LEVEL CROSSINGS

One expression for N was obtained in terms of the number of crossings of the amplitude interval. To quote from Ref. [1, p. 14-3] (with slight notational change): "For analyzing a sample record of length T with an amplitude window of width (b - a), the total number of times that data is observed is equal to $\bar{\nu}_{(b-a)}T$ where $\bar{\nu}_{(b-a)}$ is the number of crossings per second of the amplitude interval (a, b). The number of events may be thought of as the number of crossings of the interval (a, b) multiplied by the width (b-a) of the interval. If the interval width (b-a) is small, the number of crossings of the interval (a, b) is approximately equal to the number of crossings of the level a denoted by $\bar{\nu}_a T$. Thus, $N = (b-a)\bar{\nu}_a T$."

Although this approach appeared to give reasonable estimates of the mean square error in the experiments which were performed, it has several major faults. In the first place, the level crossings will not generally be independent events so that the basic assumption of Eq. (5) is violated. Secondly, for a Gaussian distribution of amplitudes, the average number of level crossings may be found without difficulty. However, for a non-Gaussian random process, the level crossing calculation is quite difficult since it depends upon the joint distribution of the process and the derivative of the process. If the process is known to be Gaussian, probability density measurements are obviously not needed and all pertinent information can be obtained from estimates of the mean value and covariance function. Thus, principal interest is in the cases of unknown and/or non-Gaussian distributions for which the level crossing computation is not possible. Because of the above facts, it is felt that the "level crossing method" offers little possibility for further extension and will not be considered further.

3.2 SAMPLING COEFFICIENTS

The second method considered in Ref. [1] is based upon a statistical study of the coefficients defined by the sampling theorem for bandwidth limited random processes. Again, to quote from Ref. [1, p. 14-3]; "The number of events represented by a continuous random signal is given by $N = 2BT$ where B is an equivalent ideal bandwidth in cps and T is the available record length in seconds. To be more exact, T represents the total time the signal is actually observed and analyzed. For the problem at hand, T is only that time spent by the signal within the amplitude window $(b-a)$ since the signal is not actually observed and analyzed when the amplitudes are outside the window $(b-a)$. This actual analysis time is given by $\sum_i \tau_i$ in Eq. (5). Thus, for analyzing a sample record of length T with an amplitude window of width $(b-a)$, the equivalent number of events becomes $N = 2B\sum \tau_i$. From Eq. (4), $\sum \tau_i = (b-a)\hat{f}(x)T$. Substituting this into the expression for N gives $N = 2(b-a)\hat{f}(x)BT$."

As before, the assumption is made that the sampling coefficients are independent random variables. It will be shown in the next section that this is not usually true and consideration must be given to the correlations between the samples. Also, use of the equivalent ideal bandwidth is not correct as will be seen.

4. ANALYTICAL MODEL OF PROBABILITY MEASUREMENTS

The approach to evaluating probability measurements based upon the sampling coefficients is valid. However, as mentioned before, a more rigorous development is required. To begin, a brief review of the necessary random process theory is presented below.

4.1 DESCRIPTION OF RANDOM PROCESSES

The random process $X(t)$ is assumed to be stationary with a power spectral density function, $G(f)$, which is limited to a bandwidth B and is zero elsewhere, but is otherwise arbitrary. It may be assumed that the frequency interval of interest starts at $f = 0$, i.e.,

$$\begin{aligned} G(f) &\geq 0 \quad , \quad 0 \leq f \leq B \\ &= 0 \quad , \quad f > B \end{aligned} \tag{6}$$

This results in a simplification of certain equations which follow, but does not change any of the results. The covariance function of $X(t)$ is given by

$$R(\tau) = \int_0^B G(f) \cos 2\pi f\tau \, df \tag{7}$$

Since $X(t)$ is stationary and band-limited, it may be represented by

$$X(t) = \underset{M \to \infty}{\mathrm{l.i.m.}} \sum_{m=-M}^{M} X\left(\frac{m}{2B}\right) \frac{\sin (2\pi Bt - m\pi}{2\pi Bt - m\pi} \tag{8}$$

for all t, Ref. [2].

Equation (8) expresses the content of the sampling theorem for random functions. The notation l. i. m. stands for limit-in-the-mean and states that the right side of Eq. (8) is the best linear estimate, in the mean square sense, of X(t) in terms of the values at the sample points. It is clear that all the statistical properties of X(t) are contained in the coefficients of the expansion.

The function, $S_m(t) = \frac{\sin(2\pi Bt - m\pi)}{2\pi Bt - m\pi}$, has the property that

$$S_m\left(\frac{n}{2B}\right) = \begin{cases} 1 & \text{if} \quad n = m \\ 0 & \text{if} \quad n \neq m \end{cases}$$

where n is an integer. For values of t such that $2\pi BT - m\pi \geq \frac{\pi}{2}$, $|S_m(t)|$ varies as $(2\pi BT - m\pi)^{-1}$ which implies that the value of X(t) is described primarily by the sample points which lie nearest to t.

For a time interval of length T, where T is chosen such that

$$2BT >> 1 \tag{9}$$

X(t) is closely approximated by a finite sum of terms derived from the sampling points lying within T. Thus,

$$X(t) \simeq \sum_{n=1}^{N} X\left(\frac{n}{2B}\right) \frac{\sin(2\pi Bt - n\pi)}{2\pi Bt - n\pi}, \qquad \begin{array}{l} t \text{ contained in } T \\ N = 2BT \end{array} \tag{10}$$

The major source of error is near the endpoints of T, but will be neglected in the analysis which follows. From Eq. (10), X(t), defined over the interval T, may be approximated by a finite sum of terms with random coefficients

$\left\{ X\left(\frac{n}{2B}\right)\right\}$ and each of the coefficients has the same statistical properties as $X(t)$ for any t.

4.2 ANALYTICAL MODEL

Let a new set of random variables $\left\{ Y_n \right\}$ be defined as follows:

$$Y_n = Y_n(a, b) = 1 \qquad \text{if} \qquad a \leq X\left(\frac{n}{2B}\right) \leq b \tag{11}$$

$$= 0 \qquad\qquad \text{otherwise}$$

Since Y_n can only be zero or one, the kth moment of Y_n is given by

$$E\left[Y_n^k\right] = (1)^k P\left[Y_n = 1\right] = p \quad , \qquad k = 1, 2, \ldots \tag{12}$$

and the variance of Y_n is

$$V\left[Y_n\right] = E\left[Y_n^2\right] - E^2\left[Y_n\right] = p(1 - p) \tag{13}$$

Let

$$Z_N = \frac{1}{N} \sum_{n=1}^{N} Y_n \tag{14}$$

Then Z_N is an unbiased estimate of p since

$$E\left[Z_N\right] = p \tag{15}$$

Thus, an analysis of Z_N is equivalent to an analysis of amplitude probability measurements which use a sum of time intervals to estimate p.

An extremely important property of an estimate of some quantity is that it be consistent. Mathematically, this means that the estimate must converge in probability to the desired quantity. Thus, if Z_N is to be a consistent estimate of p, it is required that, for any $\epsilon > 0$,

$$\lim_{N\to\infty} P\left[|Z_N - p| > \epsilon\right] = 0 \tag{16}$$

In general, an error analysis based on Eq. (16) would be difficult since the computation of the required probabilities for any value of N would not be an easy task.

The mean square error in the estimation of p is defined by

$$E\left[(Z_N - p)^2\right] = E\left[Z_N^2\right] - p^2 = V\left[Z_N\right]$$

$$= \frac{1}{N^2} \sum_{n=1}^{N} V\left[Y_N\right] + \frac{1}{N^2} \sum_{\substack{m\neq n \\ m,n=1}}^{N} \left(E\left[Y_m Y_n\right] - p^2\right) \tag{17}$$

By the Tchebycheff inequality, Ref. [3, p. 225],

$$P\left[|Z_N - p| \geq \epsilon\right] \leq \frac{V\left[Z_N\right]}{\epsilon^2} \tag{18}$$

so that mean square convergence of Z_N to $p\left(\lim_{N\to\infty} V\left[Z_N\right] = 0\right)$ implies convergence in probability. Equation (18) is also useful in that it gives an upper bound on the probability that Z_N differs from p by more than a fixed amount. The above statements are true independently of the amplitude probability distribution of X(t).

Conditions under which Z_N converges in mean square to p will now be given. The following theorem which is proved in Ref. [3, p. 419] serves as a basis. It is quoted here in the context of the random variables under discussion.

Theorem: $\lim_{N\to\infty} V\left[Z_N\right] = 0$ if and only if $\lim_{N\to\infty} C\left[Y_N, Z_N\right] = 0$

where $C\left[Y_N, Z_N\right]$ is the covariance of Y_N and Z_N.

From the definition of the random variables $\left\{Y_N\right\}$, $C\left[Y_N, Z_N\right]$ may be calculated as follows:

$$\begin{aligned} C\left[Y_N, Z_N\right] &= E\left[(Y_N - p)(Z_N - p)\right] \\ &= E\left[Y_N Z_N\right] - p^2 \\ &= \frac{1}{N} \sum_{n=1}^{N} E\left[Y_N Y_n\right] - p^2 \end{aligned} \tag{19}$$

The expectations occurring in the above summation may be expressed as

$$\begin{aligned} E\left[Y_n Y_N\right] &= P\left[Y_n = 1, Y_N = 1\right] \\ &= P\left[a \le X\left(\frac{n}{2B}\right) \le b,\ a \le X\left(\frac{N}{2B}\right) \le b\right] \end{aligned} \tag{20}$$

where the right side of Eq. (20) is the joint probability that the two sample points both lie in the interval (a, b). Therefore, Z_N is a consistent estimate of p if and only if

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} P\left[a \le X\left(\frac{n}{2B}\right) \le b,\ a \le X\left(\frac{N}{2B}\right) \le b\right] = p^2 \tag{21}$$

One interpretation of Eq. (21), although not the only one, is that the random variables, $X\left(\frac{n}{2B}\right)$, defined on the sample points are required to become independent, and remain so, after some finite time separation of sample points. This means that all but a finite number of the joint probabilities are equal to the product of the individual probabilities, and thus equal to p^2.

Having established necessary and sufficient conditions for the mean square convergence of Z_N to p, the mean square error resulting from a finite sample size will now be investigated for several situations of interest.

4.3 INDEPENDENT SAMPLES

Suppose that $X(t)$ is such that

$$C\left[Y_m, Y_n\right] = 0 \quad , \qquad m \neq n \tag{22}$$

for all m and n. From the definition of the sequence of random variables $\left\{Y_n\right\}$, Eq. (11), it follows that

$$\begin{aligned} C\left[Y_m, Y_n\right] &= P\left[Y_m = 1,\ Y_n = 1\right] - p^2 \\ &= P\left[Y_m = 1\right]\ P\left[Y_n = 1 \,\middle|\, Y_m = 1\right] - p^2 \end{aligned} \tag{23}$$

where $P\left[Y_n = 1 \,\middle|\, Y_m = 1\right]$ is the conditional probability that $Y_n = 1$ when it is known that $Y_m = 1$. Thus, the fact that all of the covariances are zero implies

$$P\left[Y_n = 1 \,\middle|\, Y_m = 1\right] = p = P\left[Y_n = 1\right] \tag{24}$$

which is the condition for statistical independence of the random variables $\left\{Y_n\right\}$. Since

$$P\left[Y_n = 1 \,\middle|\, Y_m = 1\right] = P\left[a \leq X\left(\frac{n}{2B}\right) \leq b \,\middle|\, a \leq X\left(\frac{m}{2B}\right) \leq b\right] \tag{25}$$

the independence of $\left\{Y_n\right\}$ implies the independence of $\left\{X\left(\frac{n}{2B}\right)\right\}$. The converse statement is also true.

If it can be shown that the samples obtained from $X(t)$ are independent, the mean square error in the estimation of p becomes

$$V\left[Z_N\right] = \frac{1}{N^2} \sum_{n=1}^{N} V\left[Y_n\right] = \frac{p(1-p)}{2BT} \tag{26}$$

where N has been replaced by its value in terms of the time-bandwidth product, namely $N = 2BT$. When the true numerical value of p is unknown, it is desirable to replace $p(1-p)$ by its maximum value which is $(1/4)$. Thus, for independent samples

$$V\left[Z_N\right] \leq \frac{1}{8BT} \tag{27}$$

Let Q be the required probability that Z_N lies within $\pm\,\epsilon$ of p after $2BT$ samples. Then, from Eq. (18),

$$Q = 1 - P\left[\left|Z_N - p\right| > \epsilon\right] \geq 1 - \frac{1}{8BT\epsilon^2} \tag{28}$$

Since Z_N is the number of "successes" in N independent trials, the distribution of Z_N is binomial. For large N, the binomial law may be closely approximated by a normal distribution, and in this case, Eq. (28) may be replaced by

$$Q > \Phi\left(\frac{\epsilon}{(8BT)^{-1}}\right) - \Phi\left(\frac{-\epsilon}{(8BT)^{-1}}\right) = 2\Phi(8BT\epsilon) - 1 \tag{29}$$

where

$$\Phi(x_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_0} e^{-x^2/2}\, dx$$

which is a tabulated function.

As an example of independent coefficients in the expansion of Eq. (10), assume X(t) is a stationary, normal random process with zero mean and a uniform spectral density limited to a bandwidth B. The covariance function of X(t) is given by

$$R(\tau) = R(0) \int_0^B \cos 2\pi f\tau \, df = R(0) \frac{\sin 2\pi B\tau}{2\pi B\tau} \tag{30}$$

X(t) may be represented by its values at the sample points which are spaced at time intervals of 1/2B. Thus, letting $\tau = (m - n)/2B$, Eq. (30) gives

$$\begin{aligned} R\left(\frac{m-n}{2B}\right) = C\left[X\left(\frac{m}{2B}\right),\; X\left(\frac{n}{2B}\right)\right] &= R(0)\,\frac{\sin \pi(m-n)}{\pi(m-n)} \\ &= \begin{cases} 0 & , \; m \neq n \\ R(0) & , \; m = n \end{cases} \end{aligned} \tag{31}$$

In the case of normal random variables, zero covariance implies statistical independence and the mean square error in estimating p is given by Eq. (27).

4.4 CORRELATED SAMPLES

In the general case for which the random variables Y_n are correlated in some unspecified way, it does not appear possible to establish an upper bound on $V\left[Z_N\right]$ which converges to zero with increasing N. To see this, let

$$C_{n-m} = C\left[Y_m, Y_n\right] = E\left[Y_n Y_m\right] - p^2 \tag{32}$$

and upon substituting this into Eq. (17), it is seen that

$$V\left[Z_N\right] = \frac{p(1-p)}{N} + \frac{2}{N^2} \sum_{n=1}^{N-1} (N-n)\, C_n \tag{33}$$

The maximum value of C_m is $p(1-p)$, thus

$$0 \le V\left[Z_N\right] \le p(1-p)\left[\frac{1}{N} + \frac{2}{N^2} \sum_{n=1}^{N-1} (N-n)\right] = p(1-p) \tag{34}$$

which is independent of N.

Without further knowledge or assumptions concerning the covariance of the sequence $\left\{Y_n\right\}$, no useful estimate of the mean square error is available. In fact, knowledge of the covariances of $\left\{Y_n\right\}$ is equivalent to a knowledge of the joint distribution of $X(t)$ and $X(t+\tau)$. This joint distribution is a higher order statistic of the random process $X(t)$ than the first order amplitude distribution and thus could not reasonably be expected to be known. In fact, the first order amplitude distribution is directly obtainable from the joint distribution.

Suppose it is known, or can be established that the values of $\left\{X\left(\frac{n}{2B}\right)\right\}$ are independent when the sample points are separated by an interval greater than $\frac{k}{2B}$, where k is an integer. This means that $C_n = 0$ for $n > k$. Under these conditions it follows from Eq. (33) that

$$V\left[Z_N\right] \leq \frac{2BT(2k+1) - k(k+1)}{4(2BT)^2} \tag{35}$$

where the value of 1/4 has been substituted for $p(1 - p)$ to make the bound independent of p. Thus, if adjacent sample points are statistically dependent, $k = 1$, an upper bound on the mean square error is $3/(8BT)$.

The result presented above, Eq. (35), indicates that useful bounds can be obtained if independence only over a finite interval is assumed. However, unless the relation between the dependence interval and the total length of the record, T, is known, there is insufficient information to determine the appropriate value of k.

5. AMPLITUDE PROBABILITY DENSITY ESTIMATES

5.1 GENERAL ANALYSIS

In the analysis that follows, it will be assumed that the random variables $\{Y_n\}$ are independent. The previous analysis has been concerned with the estimation of probabilities rather than probability densities, and it has been shown that Z_N is an unbiased consistent estimate. When Z_N is used to estimate a probability density function, however, a bias is introduced which increases the mean square error.

Let $f(x)$ be the probability density function of the amplitude of $X(t)$, and $F(x)$ the corresponding probability distribution function $\left[F'(x) = f(x)\right]$. From the prior definitions, the quantity

$$f_N(x) = \frac{Z_N(a, b)}{b - a} \quad , \quad x = \frac{a + b}{2} \quad , \quad a < b \tag{36}$$

is an estimate of $f(x)$. The decision as to how large $(b - a)$ should be will be discussed later. The estimate Z_N may be expressed in terms of an empirical distribution function $F_N(x)$ so that

$$f_N(x) = \frac{F_N(b) - F_N(a)}{b - a} \tag{37}$$

where, for example,

$$F_N(b) = \frac{\text{number of } X\left(\frac{n}{2B}\right) \leq b}{N} \tag{38}$$

Thus,

$$E\left[f_N(x)\right] = \frac{F(b) - F(a)}{b - a} \tag{39}$$

which is not usually equal to $f(x)$, and a bias error is introduced into the estimation of the probability density. To relate the above notation to the previous analysis, note that

$$(b - a) E\left[f_N(x)\right] = E\left[Z_N\right] = p$$

The mean square error in estimating $f(x)$ by $f_N(x)$ is given by

$$E\left[(f_N(x) - f(x))^2\right] = V\left[f_N(x)\right] + E^2\left[f_N(x) - f(x)\right] \tag{40}$$

From Eqs. (32) and (37), and the definition of variance,

$$V\left[f_N(x)\right] = \frac{V\left[F_N(b)\right] + V\left[F_N(a)\right] - 2C\left[F_N(b), F_N(a)\right]}{(b - a)^2} \tag{41}$$

where

$$C\left[F_N(b), F_N(a)\right] = E\left[F_N(b) F_N(a)\right] - F(b) F(a) \tag{42}$$

and

$$V\left[F_N(b)\right] = F(b)\left[1 - F(b)\right]$$

Since the elements of the sequence $\left\{Y_N\right\}$ are independent, and setting $Y_N(-\infty, b) = Y_n(b)$,

$$E\left[F_N(b) F_N(a)\right] = \frac{1}{N^2} \sum_{n=1}^{N} E\left[Y_n(a) Y_n(b)\right] + \frac{1}{N^2} \sum_{\substack{m \neq n \\ m,\, n=1}}^{N} E\left[Y_m(a) Y_n(b)\right]$$

$$= \frac{F\left[\min(a, b)\right]}{N} + \frac{N - 1}{N} F(a) F(b)$$

$$= \frac{F(a)}{N} + \frac{N - 1}{N} F(a) F(b) \tag{43}$$

Therefore, substituting Eq. (43) into Eq. (42),

$$C\left[F_N(b), F_N(a)\right] = \frac{1}{N}\left[F(a) - F(a) F(b)\right] \tag{44}$$

and Eq. (41) becomes

$$V\left[f_N(x)\right] = \frac{F(b)\left[1 - F(b)\right] + F(a)\left[1 - F(a)\right] - 2F(a) + 2F(a)F(b)}{N(b - a)^2}$$

$$= \frac{F(b) - F(a) - \left[F(b) - F(a)\right]^2}{N(b - a)^2} \tag{45}$$

Thus, the mean square error becomes

$$E\left[(f_N(x) - f(x))^2\right] = \frac{F(b) - F(a) - \left[F(b) - F(a)\right]^2}{N(b - a)^2} + \left[\frac{F(b) - F(a)}{b - a} - f(x)\right]^2 \tag{46}$$

It is of interest to express the mean square error strictly in terms of the density function rather than the associated distribution function. To this end, let F be expanded in a Taylor series about the point x, then letting $\Delta = (b - a)$,

$$F(b) = F(x) + \left(\frac{\Delta}{2}\right)f(x) + \frac{1}{2}\left(\frac{\Delta}{2}\right)^2 f'(x) + \frac{1}{6}\left(\frac{\Delta}{2}\right)^3 f''(x) + \ldots$$

$$F(a) = F(x) - \left(\frac{\Delta}{2}\right)f(x) + \frac{1}{2}\left(\frac{\Delta}{2}\right)^2 f'(x) - \frac{1}{6}\left(\frac{\Delta}{2}\right)^3 f''(x) + \ldots$$

and

$$F(b) - F(a) \cong \Delta f(x) + \frac{\Delta^3 f''(x)}{24} \tag{47}$$

The next term in the series is $\dfrac{\Delta^5 f^{IV}(x)}{1920}$ which can reasonably be expected to be negligible if Δ is small and $f(x)$ has no sharp peaks. Assuming that the quantity $\left[F(b) - F(a)\right]$ can be approximated closely enough by the two term expansion of Eq. (47), the final expression for the mean square error is

$$E\left[\left|f_N(x) - f(x)\right|^2\right] = \frac{f(x)}{N\Delta}\left[1 - \Delta f(x)\right] + \frac{\Delta f''(x)}{24N} + \frac{\Delta^2 f(x) f''(x)}{12N} + \left(\frac{N-1}{N}\right)\frac{\Delta^4 (f''(x))^2}{576} \quad (48)$$

and the bias error becomes

$$E\left[f_N(x) - f(x)\right] = \frac{\Delta^2 f''(x)}{24} \quad (49)$$

5.2 COMPARISON WITH PREVIOUS RESULTS

A comparison of previous approaches to the estimation of probability densities (see Ref. [1]) indicates close agreement with the results derived here. Following Ref. [1], let ϵ^2 be the normalized mean square error defined by

$$\epsilon^2 = \frac{E\left[\left(f_N(x) - f(x)\right)^2\right]}{f^2(x)} \quad (50)$$

If all terms involving $f''(x)$ are neglected in Eq. (48), then setting $N = 2BT$ yields

$$\epsilon^2 = \frac{1 - f(x)\Delta}{2BT\, f(x)\Delta} \quad (51)$$

In Ref. [1], the expression given for ϵ^2 is

$$\epsilon^2 = \frac{1}{2BT\, f(x)\Delta} \quad (52)$$

However, since Δ is usually small, the value of ϵ^2 as computed by Eq. (52) should be a good approximation to the mean square error if $f(x)$ is reasonably smooth and the samples are independent.

Experimental results presented in Ref. [1] give a calculated ϵ^2 which is smaller than the ϵ^2 of Eq. (52) by a factor of about 10. These experiments were performed quite carefully; therefore, the difference between the theory leading to Eq. (52) and the experimental results cannot be attributed to calibration errors, etc. A reasonable explanation of the differences is that the amplitude values of the random processes used were statistically dependent over significant time intervals. If the majority of the correlations were negative, this would serve to explain the apparent discrepancy.

To illustrate the effect of statistical dependence upon the mean square error, suppose that only adjacent samples are dependent and that the bias error can be neglected. In this case the correct expression for the normalized mean square becomes, using Eq. (33),

$$\epsilon^2 = \frac{1}{p^2}\left[\frac{p(1-p)}{N} + \frac{2(N-1)C_1}{N^2}\right] \tag{53}$$

where the relation $E\left[Z_N\right] = p$ has been used.

If the measured value of the normalized mean square error is

$$\epsilon_m^2 = 0.1\,\frac{(1-p)}{Np} \tag{54}$$

the value of C_1 may be found by equating the right sides of Eqs. (53) and (54). Thus, it follows that

$$C_1 = \frac{-0.9(1-p)\,p}{2} = -.45\,(1-p)\,p \tag{55}$$

when N is large. If p is about 0.04 (such as would be the case in measuring the peak value of a unit variance normal density function with $\Delta = 0.1$), then the value of C_1 is -0.017. This result indicates that only a slight correlation can produce significant changes in the mean square error.

6. EXPERIMENTAL PROGRAM

In order to further evaluate the effect of correlation, a digital simulation of amplitude probability measurements was carried out. The basic approach was to generate white Gaussian noise which was then passed through a digital filter to shape the output spectrum. Amplitude probabilities were estimated by counting the number of times the filtered process fell within the amplitude window. Mean square errors in the probability estimation procedure were estimated by repeating the simulation one hundred times to obtain a good statistical sample. A total of thirteen cases were run using two different digital filters. The first filter approximated either a lowpass or bandpass filter. The output spectrum when this filter was used is given by

$$G_1(f) = \frac{1}{1 + \left(\frac{f - f_c}{f_0}\right)^8} \tag{56}$$

where f_c is the center frequency and f_0 is the half power frequency. A single tuned filter was used for the second filter so that the corresponding output spectrum was

$$G_2(f) = \frac{1}{(f^2 - f_n^2)^2 + \frac{f_n^2 f^2}{Q^2}} \tag{57}$$

where f_n is the natural frequency and Q is a measure of the narrowness of the filter bandwidth.

The results for the thirteen cases are presented in Table 1. Before attempting to interpret the results, the methods used in computing the various quantities shown will be indicated.

Case	$\overline{P}$	T (sec)	Spectrum	B_n	σ_e^2	σ_c^2	σ_e^2/σ_c^2
1	.666	3.25	$G_1(f)$, $f_c = 0$, $f_0 = 10$	10.3	1.73×10^{-3}	3.33×10^{-3}	.520
2	.634	1.30	$G_1(f)$, $f_c = 0$, $f_0 = 25$	25,7	1.86×10^{-3}	3.46×10^{-3}	.538
3	.864	0.65	$G_1(f)$, $f_c = 0$, $f_0 = 77$	79.1	8.86×10^{-4}	1.15×10^{-3}	.770
4	.495	0.65	$G_1(f)$, $f_c = 0$, $f_0 = 100$	102.8	1.04×10^{-3}	1.87×10^{-3}	.555
5	.360	0.65	$G_1(f)$, $f_c = 100$, $f_0 = 100$	205.6	6.24×10^{-4}	8.65×10^{-4}	.721
6	.640	0.65	$G_2(f)$, $f_n = 30$, $Q = 20$	2.6	4.36×10^{-3}	6.77×10^{-2}	.064
7	.677	0.65	$G_2(f)$, $f_n = 30$, $Q = 10$	5.2	3.80×10^{-3}	4.18×10^{-2}	.091
8	.419	0.65	$G_2(f)$, $f_n = 30$, $Q = 5$	10.4	2.98×10^{-3}	2.34×10^{-2}	.127
9	.548	0.65	$G_2(f)$, $f_n = 30$, $Q = 2$	24.5	2.77×10^{-3}	7.80×10^{-3}	.355
10	.477	0.65	$G_2(f)$, $f_n = 100$, $Q = 20$	7.9	8.60×10^{-3}	2.45×10^{-2}	.351
11	.623	0.65	$G_2(f)$, $f_n = 100$, $Q = 10$	15.7	5.58×10^{-3}	1.15×10^{-2}	.485
12	.375	0.65	$G_2(f)$, $f_n = 100$, $Q = 5$	31.1	1.76×10^{-3}	5.00×10^{-3}	.352
13	.514	0.65	$G_2(f)$, $f_n = 100$, $Q = 2$	73.5	1.26×10^{-3}	2.62×10^{-3}	.480

Table 1. Simulation Results

The true value of the probability being estimated on each run was taken to be the average over the 100 runs. Thus,

$$\overline{P} = \frac{1}{100} \sum_{i=1}^{100} P_i \tag{58}$$

where P_i is the estimate of P in the ith run.

Measurement time T was governed by

$$T = \frac{M}{2B_F} \tag{59}$$

where

M = number of data points in filter output

B_F = 385 cps = folding frequency associated with the sampling interval

The experimental mean square error was computed from

$$\sigma_e^2 = \frac{1}{100} \sum_{i=1}^{100} (P_i - \overline{P})^2 \tag{60}$$

In order to use Eq. (26) for the computed mean square error, σ_c^2, it is necessary to specify the bandwidth of the random process being analyzed. However, for the spectra given by Eq. (56) and Eq. (57), it is not possible to define a unique bandwidth and an equivalent one must be employed. To this end the noise equivalent bandwidth, B_n, has been used in the calculation of σ_c^2. The noise equivalent bandwidth is defined by

$$B_n = \frac{\int_0^\infty G(f)\, df}{G_{max}} \tag{61}$$

where G_{max} is the maximum value of the spectral density function. For the filters employed in the simulation program, it may be shown that

$$\begin{aligned} B_{n1} &= 1.028\, f_0 \quad , \quad \text{lowpass filter} \\ &= 2.056\, f_0 \quad , \quad \text{bandpass filter} \\ B_{n2} &= \frac{\pi f_n}{2Q}\left(1 - \frac{1}{2Q^2}\right) \end{aligned} \tag{62}$$

Using the above relations, the computed mean square error was found from

$$\sigma_c^2 = \frac{\overline{P}(1 - \overline{P})}{2 B_n T} \tag{63}$$

The last column in the table gives the ratio of σ_e^2 to σ_c^2. It is of interest to note that this ratio is consistently less than one. This indicates that the computed mean square error is a conservative estimate of the true mean square error, at least for the cases run. Of course, it is not possible to generalize this conclusion to other random processes.

Since each of the σ_e^2 was determined from 100 runs, there is a sampling variability associated with the values. Thus, to test whether σ_e^2 is in fact the same as σ_c^2, a two-sided χ^2 test with a 5% level of significance and 99 degrees-of-freedom was applied to the results. At this level, the ratio σ_e^2/σ_c^2 must fall between .741 and 1.30 before the hypothesis that $\sigma_e^2 = \sigma_c^2$ is accepted. Referring to the table, it is seen that the hypothesis is accepted only for Case 3. However, a comparison of Case 3 with Cases 4 and 5 indicates that if Case 3 passed the test, so should have Cases 4 and 5. This follows from the fact that the bandwidths of the latter cases were larger, and thus the roll-off effects should have been smaller. The fact that Cases 4 and 5 failed to test makes plausible the conclusion that sampling variability caused Case 3 to pass the test when it should have failed.

7. CONCLUSIONS

The previous work points up a central fact common to the analysis of a sequence of random variables: a mean square error analysis of properties of the sequence produces conclusive results only when the random variables are independent or the associated covariances are known to some degree. In a survey of recent investigations into the estimation of amplitude probability densities, Refs. [1, 4, 5, 6, 7, 8], the assumption of independent samples was either explicit or implicit in all cases. Thus there appears to be no analytical formulation upon which to base an exact expression for the mean square error which would be valid in all cases.

The analysis which led to the mean square error expression of Eq. (35) has verified that the error varies as $1/BT$, where B is the total bandwidth of the process and T is the total measurement. This result is in essential agreement with the expression derived in Ref. [1] with the exception that the equivalent noise bandwidth was used there. It has not been possible, however, to determine the explicit form of the mean square error for all cases of interest.

The simulation program which was described in Section 6 clearly showed that the mean square error does not follow the simple expression of Eq. (26). However, it appears reasonable to conclude that Eq. (26) will provide a useful guide in the selection of record lengths but not in determining the associated mean square error.

REFERENCES

1. 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-TDR-62-973, Aeronautical Systems Division, AFSC, USAF, WPAFB, Ohio. December 1962.
2. Balakrishnan, A.V., "A Note on the Sampling Principle for Continuous Signals," <u>IRE Trans. Info. Theory</u>, vol. IT-3, June 1957.
3. Parzen, E., <u>Modern Probability Theory and Its Applications</u>, John Wiley and Sons, New York. 1960.
4. Rosenblatt, M., "Remarks on Some Nonparametric Estimates of a Density Function," <u>Annals Math. Stat</u>. vol. 27, pp. 832-837. 1956.
5. Parzen, E., "On the Estimation of a Probability Density Function and Mode," <u>Annals Math. Stat.</u> vol. 33, pp. 1065-1076. 1962.
6. Leadbetter, M. and G. Watson, "On the Estimation of the Probability Density," ASTIA AD 264 811.
7. Park, J., "Statistical Estimation of Normalized Linear Signal Parameters," ASTIA AD 230 887.
8. Leadbetter, M., "On the Non-Parametric Estimation of Probability Densities," ASTIA AD 415 764.

APPENDIX I

(Supplement to Section 6)

DETAILS OF EXPERIMENTAL PROGRAM

The program developed to evaluate the effect of correlation, MAC004C, was written mainly in the FORTRAN IV language for the Univac 1107 digital computer.

For each case processed, the program did the following:

i. Read in a control card describing the filter to be used.

ii. Generated filter weights based on the filter parameters

iii. Generated one-hundred sets of random numbers, each set consisting of 500 to 550 individual numbers. The sequences were developed so that their probability density functions were Gaussian and their power spectral densities were flat (white noise).

iv. Filtered each sequence in turn using the filter weights developed in step ii.

v. Computed the sample probability $P_i\left[a \leq y < b\right]$ for each sequence. The parameters a and b were preselected so that $P\left[a \leq y < b\right] \approx 1/2$.

vi. The mean and variance of the set $\left\{P_i\right\}$ was then computed.

vii. As a control, probability density and power spectral density functions were computed of both the original filtered data for the last case.

The numerical filters employed were of two types. The first was a lowpass filter whose transfer function had the general form

$$G_1(f) = \frac{1}{1 + \left\{\dfrac{\sin\left(\dfrac{f - f_0}{2}\right)}{\sin\left(f_0/2\right)}\right\}^N} \qquad \text{(I-1)}$$

The filter weights were obtained from Eq. (I-1) by evaluation of the Fourier transform of $G_1(f)$.

$$g_1(t) = \Delta f\left[2 \sum_{i=1}^{m-1} G_1(i\Delta f)\cos(ti\Delta f) + G_1(0) + \cos(tm\Delta f)\, G(m\Delta f)\right]$$

where

$$\Delta f = \frac{1}{(2m\Delta t)}$$

The data was filtered through the use of

$$y(i\Delta t) = \Delta t \sum_{j=-N}^{N} x\Big((i+j)\,\Delta t\Big)\, g(j\Delta t) \tag{I-2}$$

The second filtering process had the transfer function

$$\left|G_2(f)\right|^2 \approx \frac{1}{\left(f^2 - f_n^{\,2}\right)^2 + \dfrac{f_n^{\,2} f^2}{Q^2}}$$

More precisely,

$$G_2(f) = \frac{1}{1 - 2e^{-\Delta t(j\omega + \omega_n \xi)} \cos\left[\omega_n\left(1 - \xi^2\right)^{1/2} \Delta t\right] + e^{-2\Delta t(j\omega + \omega_n \xi)}}$$

which is the transfer function of the numerical filter

$$y_i = x_i + 2e^{-\omega_n \xi \Delta t} \cos\left[\omega_n \left(1 - \xi^2\right)^{1/2} \Delta t\right] y_{i-1} - e^{-2\omega_n \xi \Delta t} y_{i-2}$$

This may be shown to have the same transfer function characteristics as the differential equation

$$\ddot{y} + 2\xi\, \omega_n \dot{y} + \omega_n^2 y = x$$

provided that f and f_n are less than $1/2\Delta t$.

The Gaussian random numbers were generated in the standard manner; the sequence $\{x_j\}$ was derived from a sequence $\{\xi_i\}$ through use of the expression

$$x_j = \sum_{i=k}^{k+11} \xi_i \qquad k = 12j - 11$$

where the $\{\xi_i\}$ are independent random variables uniformly distributed in the interval (-1/2, 1/2). As $E\left[\xi_i\right] = 0$ and $E\left[(\xi_i - E\,\xi)^2\right] = \frac{1}{12}$, then $E\left[x_j\right] = 0$, and $E\left[(x_j - Ex_j)^2\right] = 1$. The central limit theorem states that such processes as x_j become Gaussian in character for a large enough summation of ξ_i terms. Experience has shown that the addition of twelve of the uniformly distributed and independent random variables does indeed appear to be Gaussian. The uniformly distributed numbers $\{\xi_i\}$ were generated using certain numerical properties of the Univac 1107.

Although only 500 filtered values were used for each run, more than that number were generated of the x_i's because of end point and transient problems with the numerical filters.

The sample probability $P_i = P\left[a \leq y_j < b\right]$ for each run was obtained using procedures such as those discussed in MAC 402-07, "Probability Calculations on a Digital Computer,"

The sample mean and variance of $\left\{P_i\right\}$ were computed using the usual formulas:

$$\overline{P} = \frac{1}{100} \sum_{i=1}^{100} P_i \quad , \quad \sigma_P^2 = \frac{1}{99} \sum_{i=1}^{100} (P_i - P)^2$$

The final step in the program was to compute sample probability density functions and power spectral densities of $\left\{x_i\right\}$ and $\left\{y_i\right\}$ for the last run as a quality check of the processing.

One example of these outputs, Figure I-1, I-2, and I-3, is included. These were made from data generated from the last run of case 7, as listed on page 22.

The power spectral densities were computed using too many lags, resulting in a very low figure (10) for the number of degrees-of-freedom, so that the confidence bands on the PSD are quite wide. This is reflected in the scattered effect of the plot of the white noise spectra.

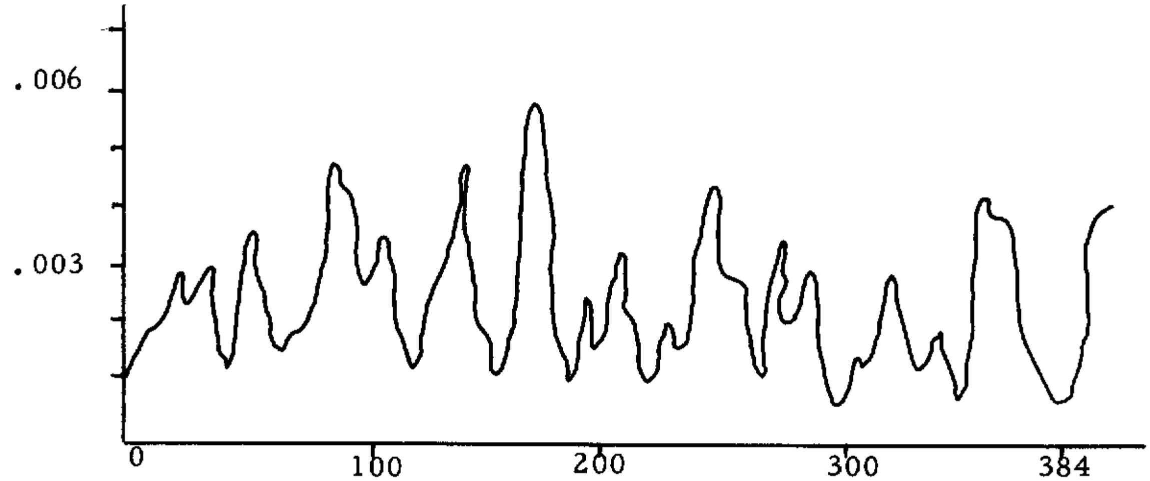

Figure I-1. Spectra of Uncorrelated Noise

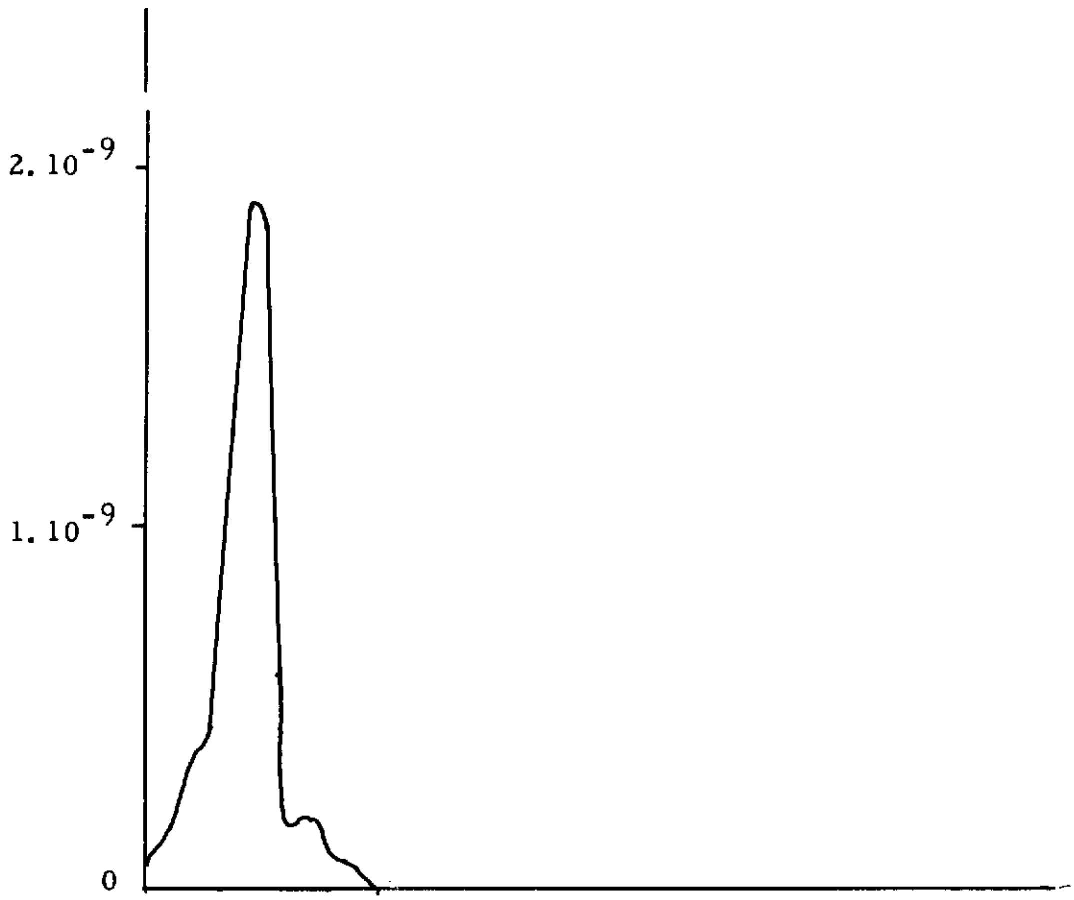

Figure I-2. Spectra of Filtered Noise

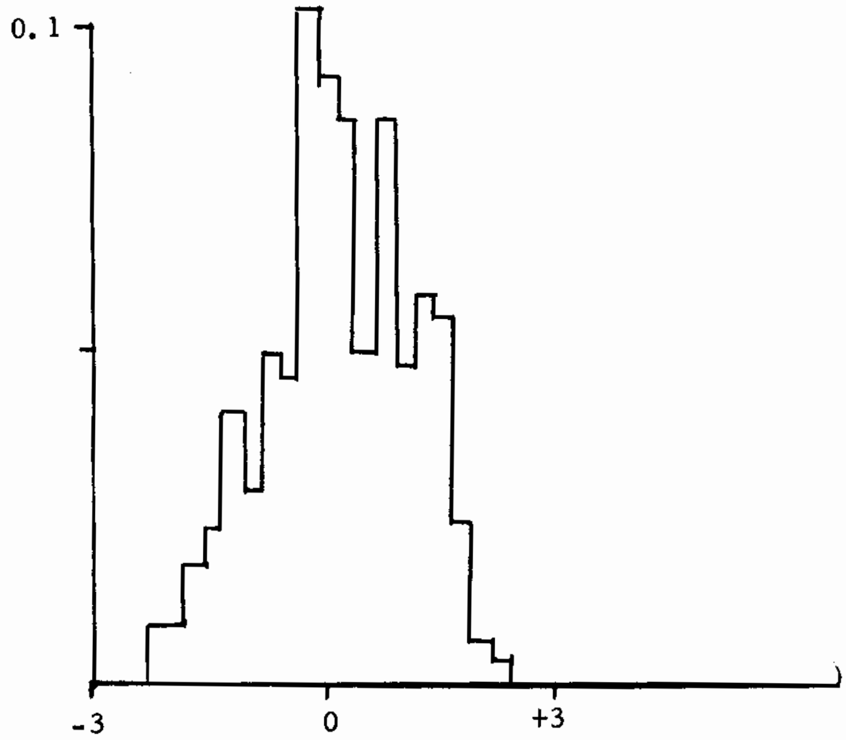

Figure I-3. Sample Probability Density Function (Histogram) of Unfiltered Data

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 Boulevard Los Angeles, California 90025	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP Not Applicable

3. REPORT TITLE

Statistical Uncertainty in Amplitude Probability Measurements

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Final Report (one of several under Contract AF33(615)-1418)

5. AUTHOR(S) *(Last name, first name, initial)*

George P. Thrall

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
January 1965	32	8

8a. CONTRACT OR GRANT NO. AF33(615)-1418	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO. 1370	FDL-TDR-64-116
c. Task No. 137005	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d.	MAC 403-01

10. AVAILABILITY/LIMITATION NOTICES

Qualified requesters may obtain copies of this report from DDC. Released to Office of Technical Services, Department of Commerce, for sale to public.

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

13. ABSTRACT

Techniques for measuring amplitude probabilities and probability densities are summarized and previous results on the statistical uncertainty of such measurements are reviewed. A rigorous mathematical model of probability measurements is derived. It is shown that unknown correlations in the parameters of the model make it impossible to develop explicit expression for the mean square estimation error. Results are presented of a computer simulation of amplitude probability estimates and comparisons are made between experimental and computed mean square errors.

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
Probability measurements						
Random Processes						
Measurement errors						
Probability density						

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.

2a. 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.

2b. 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.

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

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

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

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

9a. 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.

9b. 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