AD 696596

APPLIED RESEARCH LABORATORIES

THE UNIVERSITY OF TEXAS AT AUSTIN

ARL-TR-69-22

Copy No. 67

26 June 1969

ACOUSTIC MODELING, SIMULATION, AND ANALYSIS OF COMPLEX UNDERWATER TARGETS

II. STATISTICAL EVALUATION OF EXPERIMENTAL DATA

David Middleton
Consultant

NAVAL SHIP SYSTEMS COMMAND
Contract N00024-69-C-1129
Project Serial No. SF 1010316, Task 8212

66

ARL-TR-69-22

26 June 1969

ACOUSTIC MODELING, SIMULATION, AND ANALYSIS OF COMPLEX UNDERWATER TARGETS
II. STATISTICAL EVALUATION OF EXPERIMENTAL DATA

David Middleton
Consultant

NAVAL SHIP SYSTEMS COMMAND
Contract N00024-69-C-1129
Project Serial No. SF 1010316, Task 8212

APPLIED RESEARCH LABORATORIES
THE UNIVERSITY OF TEXAS AT AUSTIN
AUSTIN, TEXAS 78712

ABSTRACT

Some methods of analyzing experimental data for complex targets and reverberation are outlined, including a heuristic discussion of what constitutes an ensemble and how to obtain one experimentally. The problem of then determining whether or not the actual data form a valid ensemble is considered. Tests for the stability, or stationarity, of the underlying random mechanism are briefly described (test of independence and homogeneity); tests of whether or not a particular data set belongs to some postulated probability distribution are provided (goodness-of-fit), including a powerful test for establishing the normality or non-normality of the sample. Among the tests considered are the χ^2, the Kolmogorov-Smirnov, the runs test, and the W-test for normality. These tests are carried out for some hypothetical reverberation data to illustrate the individual tests, at particular ranges, which can include the domain of a target in the reverberation. Some second-order properties of various classes of second moments of these data are also discussed, and an approach to relating simulated data to those from a real environment is briefly sketched. This memorandum is intended as a preliminary guide to the statistical treatment of data that are obtained in target and background measurements. As a subsequent task, additional tests and techniques remain to be chosen for this class of problems, including the explicit analysis of data already obtained.

TABLE OF CONTENTS

I. Introduction*:

The purpose of this memorandum, the second in a continuing series, is to provide a number of statistical measures of experimental data, here obtained in the study of the simulated, modeled, and actual versions of complex targets, observed in a reverberatory or ambient noise environment. The same types of statistical measures are to be similarly applied to the background noise and reverberation themselves. Since target data and reverberation are nonstationary phenomena, as the background (ambient) noise may be on occasion, meaningful results must be obtained from a suitably defined set of observations: measurements on a single member (representation) of the set are usually quite inadequate. Furthermore, for obvious practical reasons of stability and economy it is not possible to generate the theoretically desirable infinite set: we are necessarily limited to the generation of sets of finite size and representations of finite duration.

Before we can proceed in a technically clear fashion, we must define a number of fundamental terms, whose meanings are not necessarily identical with those of more conventional statistical usage, but which are quiet common to statistical communication theory (Ref. 1). Here we shall refer to the set of data, or observations, as an <u>ensemble</u>, and, furthermore, we shall say that an

ensemble ≡ a set of objects, data, observations, etc., which possess statistical regularity, <u>i.e.</u>, to which some probability measure can be meaningfully assigned. The probability measure in question is usually a probability distribution of one (or more) of the defining attributes of the set, for example, magnitude.

---

*Portions of the material in this memorandum are based on some results of the author's current work under Contract Nonr-4883(00) with the Office of Naval Research.

Our term "ensemble" here, accordingly implies, reproduceable, or invariant statistical properties. Similarly, a random process is an ensemble of functions (or representations) of time (and often, space, as well) where, of course, the underlying mechanism producing these functions is random in some sense. A theoretical ensemble, when descriptive of a random process like reverberation here, consists of an infinite number of representations, or elements. An experimentally obtained ensemble, of course, can contain only a finite number of representations. In conventional statistical usage the theoretical ensemble is equivalent to the (infinite) "population," while the finite, or "experimental" ensemble is usually called the "sample." In what follows, the nature of this terminology will be quite clear. For further details, see Ref. 1.

Key questions accordingly are:

A. Do we have a valid ensemble, or sample, to begin with; i.e., does the set of representations we have constructed, or observed, truly represent an ensemble, and hence can we legitimately deduce from it the various desired statistical properties of these data? In technical language, is the set (or sample) "homogeneous" and its members "independently generated?"

B. Is the ensemble stationary or nonstationary?

C. What probability distribution best "fits," or "describes" within preset bounds, the statistical process in question?

Other questions are:

D. What is the effect of sample duration, (the duration of each member or representation in the sample) and ensemble size (the number of representations, or members, in the sample) on the accuracy with which various statistics, e.g., means, intensity, covariances, etc., can be measured?

E. How closely (in time or space) can one sample the data (of each representation) and still retain statistical independence of the sampled values?, and other questions.

To answer these we turn to the appropriate statistical methods, and in particular to those that put the least premium on prior knowledge. Many sources for this are available, and are described, for example, in the books by Cramér (Ref. 2), Hoel (Ref. 3), Mood (Ref. 4), Wilks (Ref. 5), Fraser (Ref. 6), and others (Ref. 7). A particularly useful text, describing the applied properties of some of the more important statistical techniques, is that of Siegel (Ref. 8), from which we shall take some of the needed tables (Appendix, end of this memorandum). Our general task here is thus twofold: (1), to select appropriate statistical measures and tests for answering questions like the above about our data; and (2), to show in detail how these measures and tests are to be specifically applied.

Let us state the basic questions more fully, and arrange them in a hierarchy of descending order of generality (although not necessarily of practical importance), by means of which we attempt to interpret our data statistically. Our ordering becomes:

A. <u>Tests for independence of the sampled data:</u> essentially, the question here is "are the individual members of the "ensemble" independent, so that the sampled data, from member to member, can be treated as representative of independent random variables, with the same distribution?" (See pp. 195-198, Wilks (Ref. 5).) This is really the joint question of (1) "homogeneity" of the ensemble--<u>i.e.</u>, of whether or not the members are <u>generated, by the same underlying mechanism</u>, and (2), <u>that the members are independently generated</u>.

Thus, we have the equivalent:

<u>Tests for the existence of the ensemble:</u> fundamentally we must establish whether or not our set of observations (of size M, j = 1,..., M)

constitutes a valid ensemble. Tests of this property are known as tests of homogeneity (of the ensemble), and independence (of the sample members) where we wish to show that our data can be considered as generated independently by the same statistical mechanism. Thus, one has here a test of the "stability" of the underlying mechanisms as the members of the ensemble are interchanged, for all times in the interval (t_1, t_1+T) during which each member is defined; (see Sec. 30.6, Ref. 2, for example).

B. Tests for "goodness-of-fit:" in this case we desire to determine whether or not our set (of M random members) belongs to a specific distribution, *e.g.*, gauss (for high-density reverberation), Poisson, or others (such as those theoretically predicted for envelope and phase in the case of narrow band illumination (Ref. 9), for example). Various tests for this are discussed in Ref. 2, Sec. (30.1)--Secs. (30.4),(30.8); see also the nonparametric tests of Chap. 14, Ref. 5, esp. 14.2(b); Sec. (14.3), order statistics, etc.), and p. 460, and, in particular, Siegel (Ref. 8). Some others are considered here below. We can also test for the *a priori* probability of an event's occurrence: *e.g.*, signal in noise vs. noise alone. with tests of this type.

C. First- and Second-Order Sample Statistics: here we wish to determine the dependence of the mean and variance of the sample mean and variance, covariance, etc., on sample duration and ensemble (*i.e.*, "sample") size, and on the (infinite population) parameters of the corresponding theoretical ensemble. In addition, we wish to examine the dependence and independence of samples on sampling intervals. (See Secs. 8.1, 8.2, of Wilks (Ref. 5).)

We distinguish various types of tests, accordingly as the samples are large or small, or the tests themselves are distribution-free, or "nonparametric" (Refs. 6 and 8). In these latter, we avoid the often unsupportable assumption of a particular or specific distribution, with, of course, an attendant loss of efficiency in deciding an hypothesis for a given sample-size vis-à-vis the case of known distributions. Tests of homogeneity (cf., A, above) may also be applied to situations involving a known form of the distribution, but unknown distribution parameters--usually the mean and/or variance. Thus, one can ask for tests of homogeneity of the means, or the variances, etc. Other pertinent questions in the analysis and interpretation of data are: (1), confidence limits and regions; (2), the distributions of the sample statistics (e.g., means, variances, etc., cf.Wilks (Ref. 5), Sec. (8.3) et seq.); (3), stationarity and non-stationarity: if the process appears to be nonstationary, is this non-stationarity removable by a model based on the concept of stationary increments? (See Yaglom (Ref. 10), Sec. (18); also Sec. (14).)

In the present treatment we are not concerned with the derivation of statistical tests, nor with such important questions as what is a "best" test. We intend here merely to summarize and illustrate a few useful results, and quote any pertinent properties they may possess, referring the reader to the statistical literature for the analytical details (Refs. 2, 5, and 6). Nor is any special originality claimed in this respect here, with the possible exception of actual application to our present class of problems. Other tests and measures will be introduced as the work progresses. What we shall do, however, is to provide enough "mechanism" for the reader to make his own applications to specific problems.

We emphasize that in the kinds of problems we are dealing with here--the measurement of random phenomena--targets and background noise--it is essential to establish the reliability of these data: pertinent results are ensemble measures, not individual measures, so that it is vital to determine, within acceptable bounds, that we have a valid ensemble and that useful measures can be obtained from it. The fact that we have to deal with finite and even small samples makes this a definite problem.

Accordingly, we shall begin in Sec. II with a brief discussion of statistical tests, both conditional and unconditional; Sec. III is devoted to the question of the existence of an ensemble, given certain data; Sec. IV is concerned with a number of procedures for "goodness-of-fit:" do these data belong to a specified distribution? Section V discusses various properties of the sample statistics. At present we consider both small- and large-sample statistics. Section VI suggests how statistical measures may be used to connect simulation with reality, and in Sec. VII we conclude with a short summary of the principal results and their interpretation.

II. Some Brief Remarks on Statistical Tests:

The problems outlined above are all examples of the statistical test of one hypothesis against another. Let us review very briefly the salient features of such binary ("two-alternative") tests. Generally, we have

H_o: the "null" or "true" hypothesis vs H_1: the alternative to H_o.

H_o states that the data sample in question belongs to the assumed population; or "true" hypothesis state.

(2.1)

H_1: the alternative hypothesis to H_o, which states that H_o is not "true," viz. that the data sample belongs to some other population, or, equivalently, represents some other alternative hypothesis.

On the basis of a data sample, $\underset{\sim}{X} = (X_1,\ldots,X_n)$, where the X_j, $j = 1,\ldots, n$, are n observations (e.g., measurements) of a quantity X, we are asked to decide whether this data sample belongs to H_o, or alternatively, to H_1--or, in other words, on the basis of the observation X, whether or not H_o is "true." For this purpose a test statistic, $Z = Z(\underset{\sim}{X})$, is constructed and compared with a prechosen threshold, Z_α. If $Z_{(sample)} > Z_\alpha$, we reject H_o, and accordingly accept the alternative H_1, and conversely, if $Z_{sample} \leq Z_\alpha$, we say that H_o is "true," rejecting H_1.

Of course, neither of these decisions is always correct, since $\underset{\sim}{X}$ (and...$Z(\underset{\sim}{X})$) are random variables. Two types of error can occur in any one decision "decide H_o," or "decide H_1." These are known as Type I and II errors, where H_o is rejected (and H_1 accepted) when H_o is the true state, and where H_o is decided when H_1 is the true state, respectively. Let us now determine these error probabilities for the statistical test of H_o vs. H_1 described above. For this we need

$$\left.\begin{aligned} W_1(Z|H_o) &= \text{p.d. of the test statistic } Z(\underset{\sim}{X}), \text{ under the hypothesis } H_o \\ W_1(Z|H_1) &= \text{p.d. of the test statistic } Z(\underset{\sim}{X}), \text{ under the hypothesis } H_1. \end{aligned}\right\} \tag{2.2}$$

Our test of H_o against the alternative H_1 is for any $Z = Z(\underset{\sim}{X})$,

$$\underline{\text{decide } H_o}, \text{ if } Z \leq Z_\alpha; \text{ or, } \underline{\text{decide } H_1}, \text{ if } Z > Z_\alpha. \tag{2.3}$$

The situation is sketched in Fig. (2.1) for the ensemble of possible test statistics, Z, corresponding to the ensemble of possible data vectors $\underset{\sim}{X}$. If $Z > Z_\alpha$ we decide H_1, but there will be a Type I error probability α that this decision is incorrect, and that these data are really representative of the H_o state. Similarly, if $Z \leq Z_\alpha$ we decide H_o, but there will be a Type II error probability β that this decision is also incorrect, and that these data really belong to the H_1 state This situation is shown in Fig. (2.1). With the help of Eq. (2.2) we can accordingly say that

$$\alpha = \int_{Z_\alpha}^{\infty} W_1(Z|H_o)dZ = P_o(Z>Z_\alpha) \quad , \tag{2.4}$$

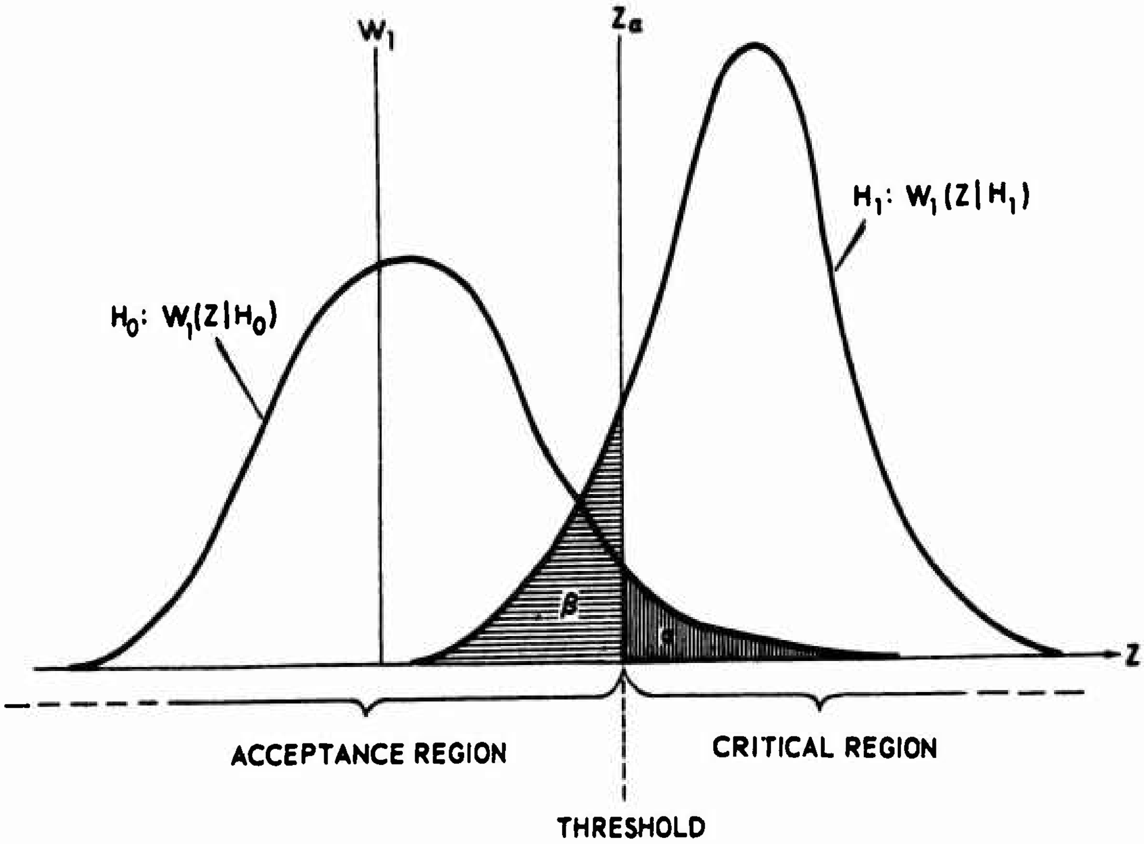

FIGURE 2.1
THE PROBABILITY DENSITIES OF THE TEST STATISTIC Z
WITH TYPE I AND II ERROR PROBABILITIES (α,β)
Z_α = THRESHOLD

ARL - UT
AS-69-750
DM - JEW
7 - 18 - 69

and that

$$\beta = \int_{-\infty}^{Z_\alpha} W_1(Z|H_1)dZ = P_1(Z \leq Z_\alpha) \quad , \tag{2.5}$$

where P_o, P_1 are probabilities under H_o, H_1 respectively. The quantity α is called the significance level of the test H_o vs. H_1; the interval $Z \leq Z_\alpha$ is the acceptance region, while $Z > Z_\alpha$ is the rejection or critical region. The quantity

$$1 - \beta = \int_{Z_\alpha}^{\infty} W_1(Z|H_1)dZ = P_1(Z > Z_\alpha) \quad , \tag{2.6}$$

is termed the power of the test. Thus, in the test (2.3) we say that if $Z_{sample} > Z_\alpha$, H_o is false (and H_1 is true) at significance level α, *i.e.*, the test $Z_{sample} > Z_\alpha$ is significant (H_o false) at level α. On the other hand, if $Z_{sample} \leq Z_\alpha$, we say the test of H_o(vs. H_1) is consistent, at level α, with the hypothesis H_o. Ideally, we would like to choose a test statistic, Z, to make both α and β indefinitely small, but this is not possible for finite sample sizes ($n<\infty$). A good test of H_o(vs. H_1) is one in which α, say, is fixed, and the power of the test, cf. Eq. (2.6) is large, or equivalently, the Type II error probability β is small. A "best" test here is one where β is made as small as possible, *i.e.*, the power of the test is maximized (for given sample size), with the Type I error probability fixed. In general, increasing the sample-size increases the power, $1-\beta$, of the test.

In actual practice our data $\mathbf{X}$ have a certain *a priori* probability, q, of belonging to the null or H_o hypothesis class, and conversely, an *a priori* probability, $p = 1-q$, of belonging to the state H_1. (If $q = 1$ we would know *for sure*, *a priori*, that $\mathbf{X}$ belonged to the process described under H_o and no test of H_o vs. H_1, would be necessary. Similarly, for $p = 1$, $\mathbf{X}$ is known surely *a priori* to belong to H_1 and no test would again be needed.) The error probabilities α, β, and the power of the test, $1 - \beta$, are all *conditional probabilities*--conditional on the "true" state of affairs (i.e., on H_o or H_1).

Thus, we have more explicitly

$$\left.\begin{aligned}
&\alpha = \alpha(H_1|H_o) = \text{(conditional) Type I error probability of deciding} \\
&\quad H_1,\ \text{when } H_o \text{ is really true,} \\
&\beta = \beta(H_o|H_1) = \text{(conditional) Type II error probability of deciding} \\
&\quad H_o,\ \text{when } H_1 \text{ is true,}
\end{aligned}\right\} \tag{2.7}$$

and

$$\therefore \left\{\begin{aligned}
&q\alpha = \text{unconditional probability of deciding } H_1 \text{ when } H_o \text{ is true;} \\
&p\beta = \text{unconditional probability of deciding } H_o \text{ when } H_1 \text{ is true.}
\end{aligned}\right\} \tag{2.7a}$$

It can be shown (Ref. 11) for the "best" test in the Neyman-Pearson sense of fixing α and minimizing β, *i.e.*,

$$\min_{\delta}(p\beta + \lambda q\alpha) \equiv R^*_{NP} \quad , \tag{2.8}$$

(where δ is the rule, analogous to Eq. (2.3), for making decisions, and λ is a Lagrange multiplier), that the optimum test statistic Z^* here is the likelihood ratio

$$Z^* \equiv \frac{W_1(\underset{\sim}{X}|H_1)}{W_1(\underset{\sim}{X}|H_o)} \quad , \tag{2.9}$$

and the test itself is

$$\underline{\text{decide } H_o}, \text{ if } Z^* \leq Z_\alpha(= \frac{\lambda}{\mu}); \text{ or, } \underline{\text{decide } H_1}, \text{ if } Z^* > Z_\alpha(= \frac{\lambda}{\mu}); \ \mu \equiv p/q. \tag{2.10}$$

The threshold Z_α is now seen to be a ratio of λ and the a priori probability ratio μ; λ itself, in the more general sense of statistical decision theory (Ref. 1, Chap. 18), is a cost ratio. Thus, our threshold Z_α embodies both the subjective elements of cost assignment (value judgements) and priori probabilities, and takes account of the fact that the occurrance of the particular data $\underset{\sim}{X}$ at hand is a priori weighted vis-à-vis H_o and H_1. Similar remarks apply for sub-optimum tests, where $Z \neq Z^*$ now: the threshold Z_α still embodies λ and μ.

Finally, the validity of a test, i.e., that it yield meaningful results on the basis of the experimental data and the significance level chosen--depends on randomization and replication (Ref. 12). By randomization is meant effectively the generation of a true ensemble in the process of data production, cf. the remarks A., Sec. 1, above. Replication is the generation of sufficient amounts of data, i.e., the establishment of an ensemble ("sample") of sufficient size--to provide statistically accurate decisions, or in other words, to yield decisions with acceptably small probabilities (α,β) of error. What is "acceptable," of course, will depend on the experimental environment and the accuracies desired.

To summarize then, we proceed as follows to construct a statistical test:

A. Select the hypothesis (H_o) whose validity at significance level α is to be tested (vs some alternative H_1, i.e., all other possibilities, for example), appropriate to the experiment undertaken.

B. Select an appropriate test statistic, $Z(\underset{\sim}{X})$. This generally means selecting an appropriate type of test.

C. Choose α, and therefore (in principle, at least) obtain the threshold Z_α from Eq. (2.4). This usually means that $W_1(Z|H_o)$ is known, or approximable.

D. With an experimental sample $Z_{sample} = Z(\underset{\sim}{X}_{sample})$, apply the decision rules:

Accept H_o (at significance level α) if $Z_{sample} \leq Z_\alpha$; or

Reject H_o (at significance level α) if $Z_{sample} > Z_\alpha$ (and $\therefore$ accept H_1).

Particular care, moreover, must be taken to insure (1), that these data, $\underset{\sim}{X}_{sample}$, belong to a valid ensemble, and (2), that sufficient replication is made, i.e., the sample is large enough, under proper ensemble conditions, to yield acceptable (i.e., "reasonable") probabilities of correct decisions. Usually, only $W_1(Z|H_o)$ is available, so the only control on the test of H_o vs. H_1 is given by the Type I error probability α. If at all possible, the attempt to determine the Type II error probability β should be made; this can be done when the mechanism governing the process X under the hypothesis H_1 is known, which, unfortunately, is not a frequent experimental situation, however.

With the above in mind, let us proceed to the program sketched in Sec. 1 earlier.

III. Tests for Independence of the Data and Validity of the Ensemble (Homogeneity and Independence):*

As we have mentioned earlier (Sec. 1) a collection of data samples does not necessarily constitute an ensemble (of finite size) which may be regarded as a subset of the infinite-member ensemble that represents the usual theoretical limit. If the underlying mechanism is changing in the course of the observation we cannot expect the statistical properties of the experimental subset to approach those of the infinite population, under stable conditions, so that the subset itself is not in any way representative of other subsets, obtained subsequently, and so on. The defining statistical character of an ensemble is that it have reproducible measures or properties--in the case of finite samples, these should converge in some sense to the corresponding measures or properties characteristic of the infinite population.

Let us elaborate further and discuss the construction of a valid ensemble. A key question here is what is meant by an ensemble (Ref. 13). As noted above in Sec. 1,

"An ensemble is a set of similarly prepared functions (here of time, space, or both, etc.) to which can be assigned a probability measure."

The critical point here is that the ensemble is an <u>unordered</u> set, in that the same properties (<u>e.g.</u>, statistics, distribution, etc.) of the set remain invariant under an arbitrary ordering. Changing the order of an <u>ordered</u> set, <u>i.e.</u>, disordering it, however, <u>implies changing the mechanism of the representations</u>, and that, of course, destroys the original set properties. Thus, if $\{X^{(j)}\}$ is the ensemble $(j = 1,\ldots,M)$, with various statistical properties, such as the mean $\overline{X}$, mean-square $\overline{X^2}$, p.d., $W_1(X)$, etc., these remain unchanged (except for a subset of measure zero) as the ordering (or "indexing") j is arbitrarily changed, <u>e.g.</u>, $j \to j'$, randomly or otherwise. For example, if

*Adapted from ONR notes of D. Middleton.

$$\overline{X} \equiv \frac{1}{M}\sum_{j=1}^{M} X(j) = \text{(ensemble) mean of } X \quad , \tag{3.1}$$

then, if the $\{X^{(j)}\}$ form a true ensemble, we must have

$$\overline{X} = \frac{1}{M}\sum_{j'=1}^{M} X^{(j')} = \frac{1}{M}\sum_{j=1}^{M} X^{(j)}; \ (j \neq j' \text{ arbitrarily}), \tag{3.2}$$

and so on for the other statistical properties of the set, which are accordingly invariant with respect to the index (j).

Ideally, we can think of generating an ensemble by producing the set of random functions simultaneously, under identical conditions and identical statistical mechanisms. For example, an ensemble of M independent "pings" from a sonar producing reverberation in a medium and/or returns from a target is theoretically obtained by introducing an identical, single ping into each of M independent media, each of which itself possesses the same basic scattering properties. The identical pings are all released at the same instant from the M distinct sources. Each medium also has a receiver at some point in it, and each receiver is identical with all the others vis-à-vis location and processing properties. The observed reverberation then consists of M representations $X^{(1)}(t), X^{(2)}(t), \ldots, X^{(M)}(t)$, $t_1 < t < t_1 + T$, on the same interval $(t_1, t_1 + T)$. By definition of the manner in which each representation is generated, each is independent of the other, in the sense that given $X^{(j)}(t)$ we can in no way construct $X^{(k)}(t)$ $(k \neq j)$ from it: the random mechanisms producing $X^{(j)}$ and $X^{(k)}$ are separate and unrelated. From this follows the required invariance of ensemble properties with respect to ordering. If, however, the M scatter mechanisms of the idealized experiment above are not independent or are not the same, ordering will depend on coupling and mechanism, and we no longer have an ensemble, in the desired sense, whose properties are invariant of index.

Theoretically, we can always construct (on paper) M identical random mechanisms and operate them simultaneously. Practically, this is nearly always impossible. Instead, we must take our data sequentially in time, generating one long record which under certain conditions we can split up into independent representations which then can constitute an ensemble. In place of our example above we now have a single source of pings and a single scatter medium. The medium is "pinged" at different times t_{oj}, $j=1,\ldots,M$, none of which is commensurable with another. Sufficient intervals $(t_{oj+1} - t_{oj})$ between successive pings are allowed, so that each reverberatory return from the medium dies down at the receiver before the next appears. Then, we take the single record of M consecutive ping returns and split it up into M separate returns, or representations, one for each individual ping. Each representation is then potentially a member of an ensemble (of size M). If each representation is independent and the underlying scatter mechanism of the medium does not change during the overall time $(t_{oM} - t_{o1})$, we then have an ensemble in the statistical sense defined above. In more recondite language, we require that the mechanism be "ergodic,"* or more simply, "stable," with independent excitations. Thus, we see that "independence" of the representations is closely related to the existence of the ensemble, so that a test for independence of the representations is essential to a test for the existence of the ensemble, as is a test of "homogeneity," implying invariance with respect to ordering of the members, etc.: there should be no change in ensemble properties (cf. (3.1), (3.2)) if we have a true ensemble, as the ordering (j) is changed.

Similar remarks apply to a target enbedded in reverberation, or any background noise. Ideally, we would prefer that position and aspect of the target be unchanging in the course of time, or at worst, changing in a completely predictable way. Physically, however, this is also never strictly the case: there are always aspect and positional variations that are both deterministic and entirely random, and which are not predictable to the observer. We can still obtain a valid ensemble, for target plus

*See Sec. 1.6 of Ref. 1 for a more precise definition.

background, provided again that the mechanism of random target variation remains basically unchanged in the observation period, and that the deterministic (i.e., systematic) effects are compensated for. For instance, with the target at a given aspect (and fixed transmitting and receiving platforms) there will be (small) random variations in the target returns, due to random rotational and translational target movement. If the aspect is alowly shifted from one bearing to another in the course of the overall observation, the record at the initial bearing will obviously be different from that at a later period of the experiment. Consequently, whereas these data at the earlier time can be split up into representations that may form an ensemble, i.e., may be homogeneous and independent (if the rate of change of aspect is slow compared to ping-rate, of course), and similarly for these data taken at later aspect periods, the various sets of representations will not in any combination constitute a valid ensemble, since the underlying mechanism is no longer stable, or "stationary," but changed. Tests of homogeneity also provide a measure of the size of the ensemble (and interval) which can be obtained under such slowly varying, "unstable" conditions.

With the above in mind, we must now cite some specific tests for the existence of a valid ensemble or, equivalently, for the homogeneity and independence of a data set (sample ensemble). Let us consider, again as our example, a set of reverberation data, containing a target at some aspect, as shown in Fig. (3.1). This set is obtained, as above, by repeated, independent pinging of the target and medium. Our basic question here is, "Do we have a valid ensemble?". We proceed first by splitting the set into two subsets and then testing for the homogeneity of the two subsets. By successively interchanging the members of each subset and repeating the test, we can then establish the homogeneity of the total set. (Other variations of this procedure are briefly indicated below.)

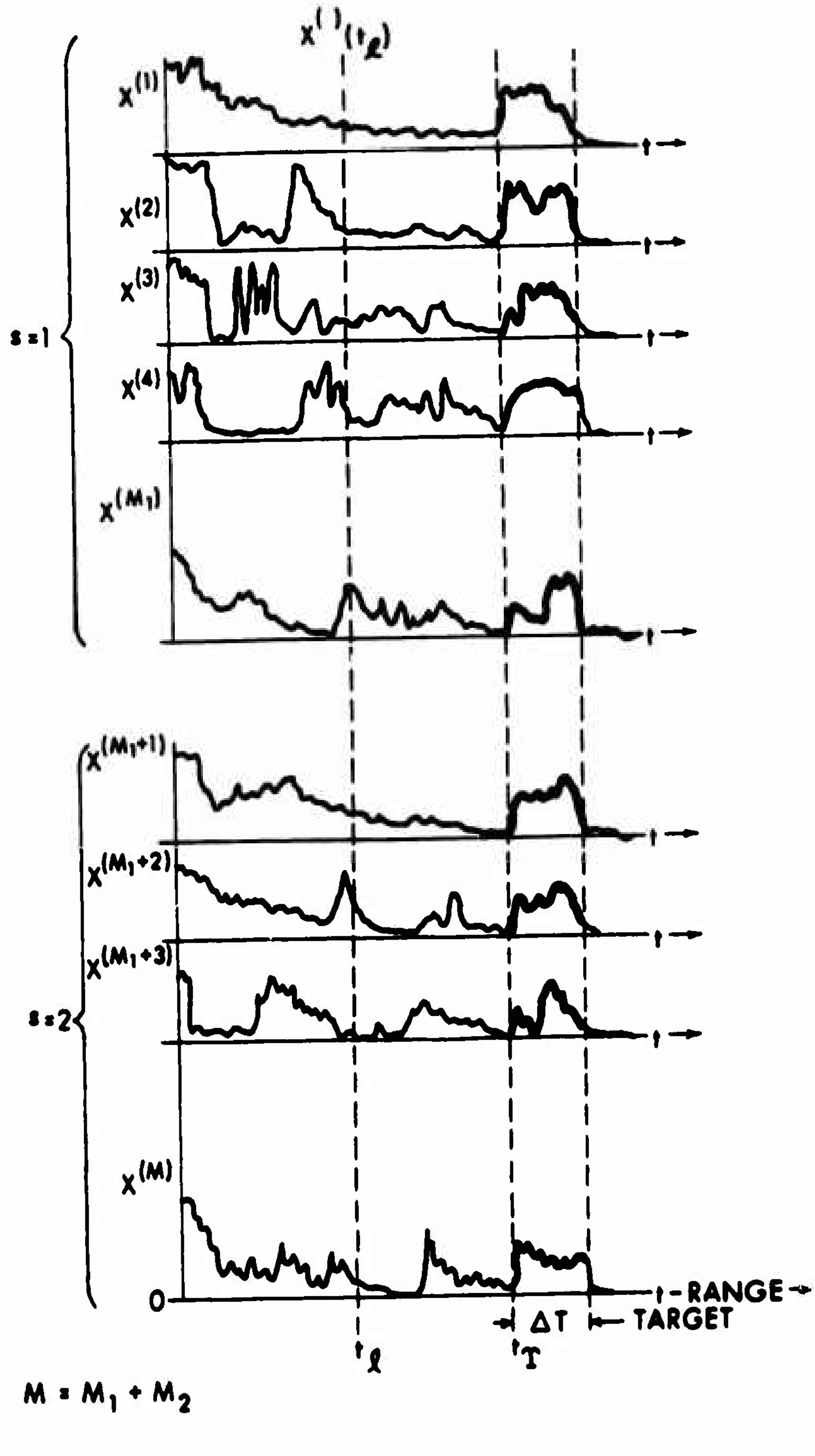

FIGURE 3.1
TWO INDEPENDENT SETS
(s=1,2) OF M_1, M_2 SAMPLES OF REVERBERATION TRACES
WITH A TARGET (SKETCHES OF ENVELOPE DATA)

ARL - UT
AS-69-751
DM - JEW
7 - 18 - 69

Let $M_1 + M_2 = M$ = total number of reverberation (and target) records generated in an experiment, and let M_1, M_2 represent the number in a pair of sequences s_1, (sample size M_1), s_2(sample size M_2). Successive pings are produced sufficiently far apart in time that there is no overlap of returns from ping to ping, and the total duration of the experiment may be not so long that the underlying propagation mechansims change. Our initial question is: "Is the set s_1 drawn from the same population as the set s_2 (and vice versa)?". Now we form a grouping into r_1, r_2 categories, as follows: Let

$$\left\{\begin{array}{l} \nu_i = \text{no. of samples (in } s_1 \text{ set) where } X(t_\ell) \text{ falls in range } X_i(t_\ell), \\ \quad X_i(t_\ell) + \Delta X; \qquad \sum_{i=1}^{r_1} \nu_i = M_1; \\ \\ \mu_i = \text{no. of samples (in } s_2 \text{ set) where } X(t_\ell) \text{ falls in range } X_i(t_\ell), \\ \quad X_i(t_\ell) + \Delta X; \qquad \sum_{i=1}^{r_2} \mu_i = M_2. \end{array}\right. \tag{3.3}$$

Each group, therefore, has a set of probabilities ($p_1^{(1)}, \ldots, p_i^{(1)}, \ldots, p_{r_1}^{(1)}$; $p_1^{(2)}, \ldots, p_i^{(2)}, \ldots, p_{r_2}^{(2)}$) associated with the $i = 1, \ldots, r_1$; $i = 1, \ldots, r_2$ groups of amplitude levels. Here it is convenient to set the number of levels the same in each group, so that $r_1 = r_2 = r$. The probabilities $(\sum_{i=1}^{r} p_i = 1)$ are not known here, but the test of homogeneity of the sequences is equivalent now to <u>establishing the hypothesis</u> (H_o) <u>that these probabilities are constants, which remain the same</u> (for the same interval levels) <u>from sequence</u> (s_1) <u>to sequence</u> (s_2). In other words, the underlying statistical description, as given by the various p.d.'s of the (infinite) ensemble, remain the same for the set s_1 as for the set s_2.

We are now ready to describe a number of tests for homogeneity of the (sample) ensemble: the first is distribution-dependent, while the second and third are distribution-free*:

A. A χ^2-test for Homogeneity of the Sample.

It is shown by Cramér (Ref. 14) that the test statistic for testing the homogeneity of the two sequences (subsets), that comprise the experimental ensemble $s_{1,2}$ described above is

$$Z_{sample} = \chi^2 = M_1 M_2 \sum_{i=1}^{r} \frac{1}{\nu_i + \mu_i} \left(\frac{\nu_i}{M_1} - \frac{\mu_i}{M_2}\right)^2 (> 0) \quad , \tag{3.4}$$

where χ^2 is distributed with a χ^2 - p.d. of r - 1 degrees of freedom ($r \geq 2$). Specifically, this p.d. is

$$W_1(Z|H_o) = \left[W_1(\chi^2|H_o) =\right] \left\{2^{\frac{r-1}{2}} \Gamma\left(\frac{r-1}{2}\right)\right\}^{-1} Z^{\frac{r-3}{2}} e^{-Z/2}, \quad Z > 0, \tag{3.5}$$

(cf. Ref. 2, Eq. 30.1.2). In particular, $r \geq 2$ and we must have $M_1, M_2 >> r$ for the sampling distribution, i.e., $W_1(Z_{sample})$, to converge in probability to the infinite sample p.d. $W_1(Z|H_o)$ above (cf. Secs. 30.1, 30.6, Ref. 2). If $M_1 = M_2 = M/2$, which is often a convenient "sizing" for comparison, the test statistic (3.4) becomes

$$Z_{sample} = \sum_{i=1}^{r} \frac{(\nu_i - \mu_i)^2}{\nu_i + \mu_i} \quad , \tag{3.6}$$

with $r \geq 2$, and $M/2 >> r$ for us to be able to use Eq. (3.5) as the p.d. of sample Z. We next select our significance level, α, and determine the threshold Z_α according to

*Here by distribution-dependent we mean that the test statistic, Z, obeys a specified form of probability distribution, whereas in the "distribution-free" case, Z does not depend on a specific form of p.d.

$$\alpha = P(Z = \chi^2 > Z_\alpha | r) = \int_{Z_\alpha = \chi_\alpha^2}^{\infty} W_1(Z|H_o)dZ; \text{ with } W_1(Z|H_o) = \text{Eq. (3.5)}, \tag{3.7}$$

for r degrees of freedom from the Table of the χ^2 distribution (Table 1, Appendix, or cf. p. 559, Cramér, Ref. 2, for example). Then we determine Z_{sample} from Eqs. (3.4), or (3.6), and apply the test (2.3):

$$\begin{cases} \underline{\text{decide } H_o}: & \text{if } Z_{sample} < Z_\alpha: \text{ the two sequences came from the same statistical population, i.e., belong to the same distribution, or} \\ \underline{\text{decide } H_1}: & \text{if } Z_{sample} \geq Z_\alpha: \text{ the two sequences do not come from the same population: sequence } s_1 \text{ is not homogeneous with sequence } s_2. \end{cases} \tag{3.8}$$

The decision H_o then has a (conditional) probability $1 - \alpha$ of being correct. Note: This test is a first-order test, i.e., only the first-order p.d.'s of $\{X^{(j)}\}$ are involved in determining the levels X_i, $X_i + \Delta X$, etc.

<u>Example:</u>

As an example, let us consider the hypothetical case of a set of reverberation pings, presumably independently generated, of the experiment described above and illustrated in Fig. (3.1). Let M = 200, i.e., 200 records with 1 ping each is the total sample ensemble size, and let us consider the (envelopes) $X^{(j)}$ of the returns at a time t_ℓ, ($t_\ell \in T$). We divide the range of amplitudes of $X^{(j)}$ into 10 intervals, all equal except for the tenth, which ranges from $9\Delta X$ to ∞. Then $M_1 = M_2 = M/2 = 100$ and $r_1 = r_2 = r = 10$, if we divide the set into two equal parts for the test of homogeneity. From our data let us say that we find that

$$\left\{\begin{array}{l} s_1: \ \nu_1 = 3;\ \nu_2 = 1;\ \nu_3 = 17;\ \nu_4 = 16;\ \nu_5 = 28;\ \nu_6 = 14;\ \nu_7 = 11; \\ \qquad \nu_8 = 5;\ \nu_9 = 3;\ \nu_{10} = 2, \\ \\ s_2: \ \mu_1 = 1;\ \mu_2 = 6;\ \mu_3 = 16;\ \mu_4 = 21;\ \mu_5 = 17;\ \mu_6 = 13;\ \mu_7 = 10; \\ \qquad \mu_8 = 8;\ \mu_9 = 7;\ \mu_{10} = 1. \end{array}\right. \tag{3.9}$$

From Eq. (3.6) our test statistic is computed to be

$$Z_{\text{sample}} = 10.67, \text{ at } t = t_\ell \quad , \tag{3.10}$$

and from the Table 1 (or Ref. 2, p. 559), for $r - 1 = 9$ deg of freedom we see that $Z_\alpha = 16.92$ at the 5% ($\alpha = 0.05$) significance level. Evidently, $Z_{\text{sample}} < Z_{\alpha=0.05} = 16.92$ here, and H_o is therefore accepted: the two sequences are homogeneous, with a 0.95 probability of this being so, and with 0.05 probability that this hypothesis (H_o) is falsely rejected. This test is then repeated for the different ranges ($t_\ell \epsilon$ T). If homogeneity prevails (at this α-significance level) we can conclude that these data everywhere in the range interval considered effectively belong to a valid ensemble. If there are ranges (t_ℓ') where H_o is rejected, we conclude tentatively that different random mechanisms are at work in the two subsets, at least at these ranges.

Of course, we could raise our threshold Z_α by decreasing the false rejection probability α, and tighten the test with respect to H_o. However, at the same time we would increase β, the (conditional) probability of falsely deciding H_o when H_1: "inhomogeneity" of $s_{1,2}$ is true. Since we do not here know the limiting distribution (let alone the sample distribution) of the test statistic Z under the alternative H_1, we cannot use Eq. (2.5) to determine β. Clearly, one reason for looking for "best" tests, in the sense at least of minimizing β (or maximizing the power, $1-\beta$), is that for a given control or significance level, α, we can hope to keep β acceptably small. In any case, choosing α too small penalizes us by a correspondingly large β: falsely deciding H_o when it really is not true. As noted in Sec. 2, the threshold Z_α really is the ratio of two other subjective factors: λ/μ - ratio of cost or value assignments to the possible decisions, to the ratio of a priori probabilities that the data at hand do or do not belong to H_o, H_1, cf. (2.10). Selecting a threshold, accordingly, involves elements outside the physical data themselves and includes the general significance and "value" of the experiment and its consequences. (Statistical Decision Theory (SDT) includes such elements, in a broader formulation of the decision process in the face of uncertainty (Ref. 15, Chaps. 1, 2, and 5) and (Ref. 1, Chaps. 18 and 19).)

So far, besides independence of the representations, we have assumed that each data group s_1, s_2 is itself a valid subensemble and what we are testing for is whether or not the combined data belong to a single statistical population (at each of the ranges $t_\ell \epsilon T$). This is really a test of the "stability" of the ensemble, here at times t_ℓ. In our example above we have shown that there is no apparent change of statistical mechanism between the time when the first representation was obtained, through the time of the last, so that if subsets 1 and 2 are themselves valid subensembles, the complete data set is likewise. However, to establish that each subset is a valid subensemble we must be able to interchange representations without damaging ensemble properties--i.e., we must have an unordered set, and this in turn implies, basically, independence of the mechanisms generating each representation.

There is no unique way of establishing this fact. One possibility is to divide up the complete data into now four equal parts and test these against each other pair by pair (or) jointly, by an extension of Cramér (Ref. 2, Sec. 30.6, for S = 4, etc.). If homogeneous, continue, splitting each quarter in half, etc. The trouble with carrying this too far (e.g., when the number of degrees of freedom (r) becomes comparable to or larger than the number of representations in the subgroup) is that we begin to get significant results too often: there are not enough data to provide adequate sample sizes.

Another possibility is to exchange every odd-indexed representation in s_1 with its counterpart in s_2 and test for homogeneity using the above procedure. Then, replace these elements and exchange all even-numbered representations and repeat the test. If H_o is rejected in one or both tests we strongly suspect that the subensembles are not valid, i.e., do not consist of independent (i.e., unordered) representations. (More refined (and complicated) tests of independence exist; these involve calculation of "contingency," cf. Ref. 2, Sec. 30.5, but will not be discussed here further.) Usually, the above may suffice, at least as an initial approach. Of course, if there are cyclical mechanisms with several cycles at least in a subgroup, these will not be revealed, and a valid ensemble may be decided, when such is not the case. (See C. below, on the "runs test.") Again, we must refer as much as we can to our knowledge of the basic physical mechanisms that govern these data, to assist us in choosing our tests and interpreting the results.

B. A Nonparametric Test for Homogeneity of the Data; The Kolmogorov-Smirnov Test:

This is the Kolmogorov-Smirnov test (Refs. 16 and 17, and in particular, pp. 47-52; 127-136 of Ref. 8), which has the considerable advantage over the χ^2-test above in that it is distribution-free, i.e., it does not depend on the (infinite) population statistics, but only on those of the sample. Furthermore, it is usually more powerful (i.e., given larger values of $1-\beta$) than does the χ^2. (Ref. 8, p. 127, et seq.)

For the basic problem here of testing the homogeneity of the data set, we again divide the sample (of size M) into two sets of representation, of size M_1, M_2, with $M_1 + M_2 = M$. The test statistic now is (Ref. 8, p. 127, *et seq.*):

$$Z_{\text{sample}} = \max_{-\infty < X(t_\ell) < \infty} \left| S_{M_1}(X_\ell)_1 - S_{M_2}(X_\ell)_2 \right| \quad , \tag{3.11}$$

where $S_{M_1}(X_\ell)_1$, $S_{M_2}(X_\ell)_2$ are the cumulative experimental distributions of X, taken over the sets $j = 1,\ldots,M_1$, $j = M_1 + 1,\ldots, M_1 + M_2 = M$, at time t_ℓ, when the X_ℓ are arranged in order of (ascending) magnitude in each subset M_1, M_2. The test based on (3.11) is a two-tailed test. It is sensitive to any kind of difference in the distributions from which the two subsets are drawn--such as location, dispersion, etc., as well as the form of the distributions themselves.

The test of whether the subensembles (s_1, s_2) are from the same population is then, as before,

$$\left\{\begin{array}{ll} \underline{\text{decide } H_0\text{:}} & \text{the two sets are homogeneous, } \underline{\text{i.e.}}\text{, belong to the same} \\ & \text{distribution, if } Z_{\text{sample}} \leq Z_\alpha, \\ & \\ \underline{\text{decide } H_1\text{:}} & \text{the two sets are not homogeneous, } \underline{\text{i.e.}}\text{, belong to different} \\ & \text{distributions, if } Z_{\text{sample}} > Z_\alpha, \end{array}\right. \tag{3.12}$$

where now, however, the threshold Z_α is given by

$$Z_\alpha = \lambda_\alpha \sqrt{\frac{M_1+M_2}{M_1 M_2}} \quad ; \tag{3.13a}$$

and λ_α is determined by

$$\alpha = 1 - \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2\lambda^2_\alpha} , \tag{3.13b}$$

which relates the significance level, α, to the threshold parameter λ_α. When $M_1 = M_2 = M/2$, (3.13a) simplifies to $Z_\alpha = 2\lambda_\alpha/\sqrt{M}$. To carry out an actual test, with chosen α, we must first obtain λ_α. (These values are provided for in Table 3, Appendix; λ_α is the coefficient of $(M_1+M_2)(M_1, M_2)^{\frac{1}{2}}$ in Table 3. Table 2 is used in the small sample case when $M_1 = M_2 \leq 40$, and the test statistic is $K = M_1 Z_{sample}$, i.e., the maximum numerator (3.11), and the threshold Z_α is replaced by K_α.)

One-tailed tests can also be constructed. These use the test statistic

$$Z_{sample} = \max_{-\infty < X(t_\ell) < \infty} \left\{S_{M_1}(X_\ell)_1 - S_{M_2}(X_\ell)_2\right\} , \tag{3.14a}$$

and are employed to decide whether or not the values (magnitudes) of the population (i.e., the theoretical ensemble) from which one of the subsets was drawn are statistically larger than the magnitudes of the population from which the other subset is taken. For small samples ($M_1 = M_2 \leq 40$) critical thresholds K_α are given by Table 2 (Appendix), where now the test statistic is modified to $K_{sample} = M_1 Z_{sample} = M_1 \cdot$ Eq. (3.14a). For large samples ($M_1, M_2 \geq 40$), the test statistic is shown to be (Ref. 8, p. 131)

$$Z'_{sample} = 4Z^2_{sample} \frac{M_1M_2}{M_1+M_2} , \tag{3.14b}$$

where Z'_{sample} obeys the χ^2-distribution of (3.5) with 2 degrees of freedom; the significance level α is determined by (3.7); and the test itself is given by (3.8). (See Table 1, Appendix.)

From this point on we test for homogeneity and validity of the sample ensembles precisely as in the preceding Case A above, repeating for each time (range) t_ℓ in the records. In practice, we would choose only a few such ranges usually, unless we had reason to suspect possible variations in the statistical mechanisms for some local range interval. Over the target our selection of t_ℓ's would be denser, to examine the target return for any level features. Whereas the amount of computation for the Kolmogorov-Smirnov test is usually larger than the χ^2 test, the effectiveness of the former is less sensitive to the true distribution involved and is more powerful. (Again, unless we know the distribution of the test statistic under H_1, we cannot determine fully how accurate our decisions are.)

<u>Example:</u>

Let us again return to the data of our reverberation example, discussed in A, above. The $X^{(j)}$ are the magnitudes of the reverberation envelopes (cf., Fig. (3.1)), which for the unordered data set are

$$\left\{\begin{array}{l} X^{(1)}(t_\ell) = 5;\ X^{(2)}(t_\ell) = 2;\ldots;\ X^{(j)}(t_\ell) = 7;\ldots;\ X^{(M_1)}(t_\ell) = 4 \\ \qquad (s_1 \text{ sequence of } M_1 \text{ representations}) \\ \text{and} \\ X^{(j=M_1+1)}(t_\ell) = 8;\ X^{(M_1+2)}(t_\ell) = 3;\ldots;\ X^{(M_1+M_2)}(t_\ell) = 6. \\ \qquad (s_2 \text{ sequence of } M_2 \text{ representations}). \end{array}\right. \tag{3.15a}$$

These become, on being ordered,

$$\left\{\begin{array}{l} (s_1):\ X^{(2)}_{i=1} = 2;\ X^{(M_1)}_2 = 4;\ X^{(1)}_3 = 5;\ldots;\ X^{(j')}_{\nu_1} = 16; \\ (s_2):\ X^{(M_1+2)}_{i=1} = 3;\ X_2\ldots;\ X^{(M_1+M_2)}_i = 6;\ldots;\ X^{(M_1+1)}_{\nu_1} = 8;\ldots; \end{array}\right. \tag{3.15b}$$

Next, we determine the cumulative distributions $S_{M_1}(X_\ell)_1$, $S_{M_2}(X_\ell)_2$, noting that there may be more than one sample member having the same value. This is done by adding increments ν_i/M_1, μ_i/M_2 to the preceding value of S_{M_1}, S_{M_2}, at S_{i-1}, respectively for the two data subsets. (Again, ν_i, μ_i are the number of ordered X_i having the same values, cf., (3.3).) From these distributions we take the largest difference, according to (3.11) or (3.14), and this is the desired test statistic Z_{sample}. This is shown in Fig. (3.2) and in Table (3.1), (including first and second sample moments) for our reverberation example.

Here $M_1 = M_2 = 100$ and again we choose the significance level $\alpha = 0.05$. From Table 3 (Appendix) applied to (3.12), we get

$$Z_{0.05} = 2\lambda_{0.05}/\sqrt{200} = \sqrt{2} \cdot 10^{-1} \lambda_{0.05} = 0.192; \; \lambda_{0.05} = 1.36 \quad , \qquad (3.16a)$$

and from Table 3.1 and Fig. 3.2 we see that the test statistic (3.11) here becomes (for the present two-tailed test)

$$Z_{sample} = 0.07 \quad , \qquad (3.16b)$$

so that $Z_{sample} < Z_{0.05}$, and therefore we conclude from (3.12) that (at the 0.05 significance level) the two subensembles (M_1, M_2) are homogeneous, i.e., belong to the same distribution.

We can also determine whether the values of X belonging to subset 1 are statistically comparable to those belonging to subset 2 by using the one-tailed test statistic (3.14), which for the present large samples is modified to (3.14a), viz

$$Z'_{sample} = 4(-0.07)^2 \frac{10^2 \cdot 10^2}{200} = 0.98 \doteq \chi^2 \text{ (2 deg. of freedom)} \quad . \qquad (3.17)$$

TABLE (3.1)

i	ΔX_i range	No. of magnitudes of X_i in interval ΔX_i		$S_{M_1}(\ )_1$	$S_{M_2}(\ )_2$	$S_{M_1} - S_{M_2}$
1	0 - 1	$\nu_i = 3$	$\mu_i = 1$	0.03	0.01	0.02
2	1 - 2	1	6	0.04	0.07	-0.03
3	2 - 3	17	16	0.21	0.23	-0.02
4	3 - 4	16	21	0.37	0.44	-0.07
5	4 - 5	28	17	0.65	0.61	0.04
6	5 - 6	14	13	0.79	0.74	0.05
7	6 - 7	11	10	0.90	0.84	0.06
8	7 - 8	5	8	0.95	0.92	0.03
9	8 - 9	3	7	0.98	0.99	-0.01
10	9 - ∞	2	1	1.00	1.00	0.00
		$M_1 = 100$	$M_2 = 100$			

$$\langle X_1 \rangle_{exp} = 4.08 \ ; \quad \langle X_{2_1} \rangle_{exp} = 4.15$$

$$\langle X_1^2 \rangle_{exp} = 20.06 \ ; \quad \langle X_2^2 \rangle_{exp} = 21.37$$

$$\sigma^2_{X_1} = 20.3 \ ; \quad \sigma^2_{X_2} = 21.6$$

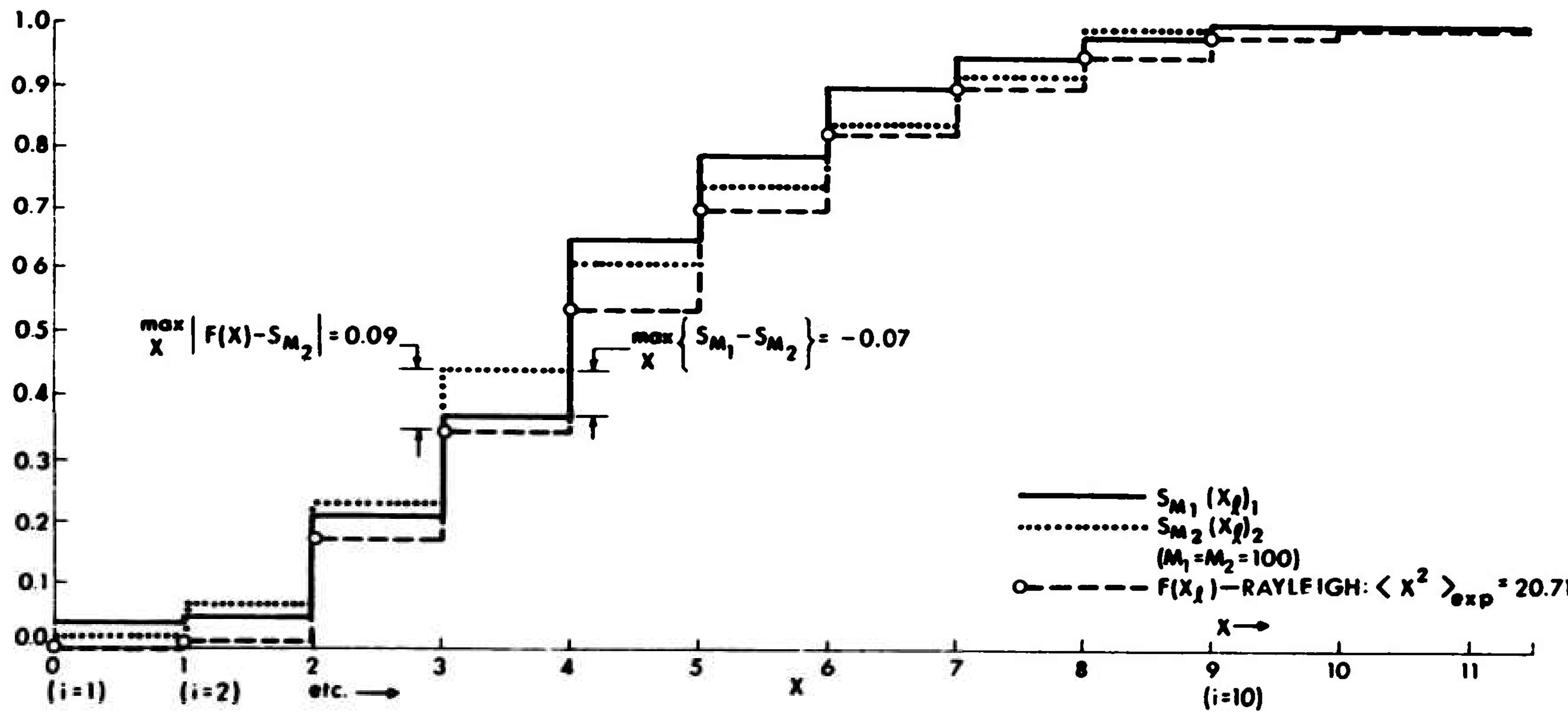

FIGURE 3.2
CUMULATIVE EXPERIMENTAL DISTRIBUTIONS S_{M_1}, S_{M_2}

ARL - UT
AS-69-752
DM - JEW
7 - 18 - 69

From Table 1 we see that $\chi^2_{0.05}|_{2df} = 5.99$, while $P(Z > \chi^2_{0.05-2df} = 0.98)$ · 0.62 ($> \alpha = 0.05$). Thus, we conclude that H_o: "no statistical difference in magnitudes"--applies, at $\alpha = 0.05$, which is, as expected, consistent with the result of the two-tailed test, that the two subensembles are homogeneous.

C. A Test of Sample Randomness and Validity of the Ensemble: One-Sample Runs Test

A key condition that the data sample constitutes a valid (experimental) ensemble is that the member representation be statistically independent,* as well as having the same underlying statistical mechanism. (See the remarks at the beginning of Sec. 3, above.) A useful test for establishing this fact, viz. the "randomness" of the sample,* in addition to what is known about the physical mechanism producing the random phenomenon--e.g., ping-rate in our specific example, is the so-called "runs test" (cf. Ref. 8, pp. 52-58), which uses the order or sequence in which the original data are obtained.

This test is based on the number of runs the data sample exhibits, where a run is defined as a succession of identical symbols (numbers, + or - signs, etc.) which are preceded and followed by different symbols (or none at all). The idea behind a test using runs is as follows: the total number of runs in a sample ensemble of any specified size gives an indication of whether or not the sample is random. For very few runs, some causal mechanism producing a time-trend or data bunching indicating a lack of element independence is suggested, whereas if an excessive number of runs occurs, some systematic cyclic, short-period mechanisms appears to be operating. In either instance, independence of the representations does not seem to hold and the sample is said not to be random (at some level of significance, α, of course, which is selected for the actual tests). The importance of a test based on runs is the fact that the order in which the events occur

*In conventional statistical language, the sample values are said to be "random."

is specifically used, unlike A and B above, where such information is neglected in the testing. Thus, it is quite possible that a χ^2 or Kolmogorov-Smirnov test might indicate acceptance of H_o, while a runs test would indicate an inhomogeneity in the statistical mechanism, i.e., rejection of H_o.

To set up a runs test we select two attributes of the sample, which appear in sequence. For example, one attribute might be all values of an observed magnitude that are positive, the other, all negative values. In a sample with M representations, let M_1 be the number having attribute No. 1 (positive values, say), and M_2 the number having attribute No. 2 (negative values), e.g., $M = M_1 + M_2$. Next, observe the order in which these attributes occur in the original data and count the number of runs, R. Thus, the test statistic is $Z_{sample} = R$ here. We distinguish two cases:

Small Samples $[M_1, M_2 < 20]$:

decide H_o, if $R_{I-\alpha} < R < R_{II-\alpha}$; i.e., decide that the sample is random - the elements are independent and unordered.

decide H_1, if $R < R_{I-\alpha}$, or $R > R_{II-\alpha}$, i.e., decide that the elements are not independent, but are ordered in some fashion.

If $R < R_{I-\alpha}$, there are two few runs, while for $R > R_{II-\alpha}$, there are too many. Here $R_{I-\alpha}$ is given in the body of Table 4 and $R_{II-\alpha}$ in Table 5, for the $\alpha = 0.05$ significance level.

Large Samples $[M_1, M_2 \geq 20]$: Here the sampling p.d. of R is essentially normal, with mean and variance

$$\bar{R} = \mu_R = \frac{2M_1M_2}{M_1+M_2} + 1; \quad \sigma^2_R = \frac{2M_1M_2(2M_1M_2-M_1-M_2)}{(M_1+M_2)^2\,(M_1+M_2-1)} \quad . \tag{3.18}$$

The test statistic is taken to be the standardized normal random variable

$$Z_{sample} = \frac{R-\mu_R}{\sigma_R} \quad , \tag{3.19}$$

and the test itself is two-tailed, specifically:

$$\begin{cases} \underline{\text{decide } H_o}: & \text{if } |Z_{sample}| < |Z_\alpha|, \text{ we conclude that the sample is randomly (\underline{i.e.}, independently) generated, with an invariant statistical mechanism,} \\ \text{or} & \\ \underline{\text{decide } H_1}: & \text{if } |Z_{sample}| > |Z_\alpha|, \text{ where we conclude that the sample is not independently generated and has a varying statistical mechanism.} \end{cases} \tag{3.20}$$

In short, the runs test provides an alternative and nonparametric approach to determining whether or not the sample ensemble is a valid one, <u>i.e.</u>, whether it can be regarded as belonging to a single random process, with specifyable measures. Here we have

$$P(|Z| > |Z_\alpha|) = \alpha \quad , \tag{3.21}$$

and P/2 is given in Table 6 (Appendix) for this large-sample case; (the probability P is twice that shown in Table 6 because we are using a two-tailed test).

<u>Example</u>:

Let us now apply the runs test to the reverberation example of the preceding portions of this report, keeping as before a significance level $\alpha = 0.05$. First, we note that our data [cf., Fig. (3.1) and Table (3.1)] refer to envelope values, which are always positive. To set up a meaningful runs test let us use as our sample attributes the positive and negative <u>differences</u> between successive envelope magnitudes, taken in independent pairs $\Delta\xi_i = X_{i+1} - X_i$, $\Delta_{i+1} = X_{i+3} - X_{i+2}$, etc., and in particular, the pluses and minuses so generated in the sequence of $M = 100$ pairs of representations (taken at time $t = t_\ell$, as before), a typical section of which, let us say, looks like

---- ++ - ++ ------ +++++ -- + ---- +++++++ ---

FIGURE (3.3)

A portion of the sequence of +'s and -'s (signs of the envelope differences $\Delta\xi$ of the reverberation sample ensemble of Fig. (3.1) at t_ℓ. (Each run is underlined.)

Let us also say that in the sample of $M = 100$ representation (and $\therefore$ (+ or -) values of $\Delta\xi$; no zeros) we find that there are $M_1 = 54$ (+'s) and $M_2 = 46$ (-'s), and $R = 58$ runs. Applying (3.18) through (3.20) we obtain

$$\mu_R = \bar{R} = \frac{2 \cdot 54 \cdot 46}{100} + 1 \doteq 50.6; \ \sigma_R \doteq 4.94;$$

$$|Z_{sample}| = \frac{|58 - 50.6|}{4.94} = 1.49 \quad , \tag{3.22}$$

so that from Table 6 we see that $P(|Z| > |Z_{sample}|) = 2 \cdot (0.0681) = 0.1362$, which exceeds $\alpha = 0.05$. Accordingly, we retain the hypothesis H_0 and conclude that our sample ensemble is a random sample, from a single population.

In the preceding three cases (A through C) we have shown how a variety of useful statistical tests may be used to establish the validity of the sample ensemble. The Kolmogorov-Smirnov test, which is applicable in all cases where the population distribution is continuous (as is certainly the case in our physical applications), is usually the most powerful. (For the runs test no meaningful statement as to its power can be made, apart from the possible context of a specific problem.)

IV. Tests of "Goodness-of-Fit:"

Here we wish to test the hypothesis (H_0) that our sample ensemble belongs to some specific distribution; (the alternative (H_1) is any other distribution). In other words, we are testing for a "fit" of our data to the postulated distribution. Again, we may distinguish both parametric and nonparametric tests, like A vs. B, C in Sec. 3 above. We begin with

A. A χ^2-test of "Goodness-of-Fit:"

In this case we have M independent* data samples $\{X^{(j)}(t_\ell)\}$, $j = 1, \ldots M$, say, of reverberation, at range t_ℓ. We wish to determine whether or not the X's here belong to some *a priori specified* distribution, *e.g.*, Rayleigh, Hoyt, etc., (Ref. 9), where the X's are the envelopes of narrow-band reverberation. We precede as follows; cf. Sec. 30.1, Ref. 2:

1. Divide the range of amplitudes into r (equal) intervals (except for the r^{th}):

$$X_1;\ X_1 + \Delta X = X_2;\ X_3 = X_2 + \Delta X, \text{ etc., where } X_{i+1} = X_i + \Delta X;$$

2. Let ν_i' = number of values of the set of M sample values having X's in the range X_i, $X_i + \Delta X$ (as in (3.3) above)

$$\therefore \sum_i^r \nu_i' = M \quad ; \qquad (4.1)$$

3. Let p_i = probability that X_i falls in the interval X_i, $X_i + \Delta X$, determined from the postulated p.d. against which we are testing

*We have established that our data is a valid sample ensemble by the methods of Sec. 3.

the sample for "fit." These p_i are given by*

$$p_i = W_1(X_i)\ \Delta X \quad , \tag{4.2}$$

where $W_1(X_i)$ is the postulated p.d., and the amplitude interval ΔX is chosen according to

$$r\Delta x = X_{(b)} - X_{(a)} \quad , \tag{4.3}$$

where $X_{(b)}$, $X_{(a)}$ are the upper and lower limits on the range of values of X in our sample set.

Now, it can be shown that (cf., Sec. 30.1, Ref. 2) the test statistic here is

$$Z_{sample} = \chi^2_{sample} = \sum_{i=1} (\nu_i' - Mp_i)^2 / Mp_i \quad , \tag{4.4}$$

and that as $M\to\infty$, $W_1(Z_{sample}|H_o) \to W_1(Z|H_o) = \chi^2$-probability density with r-1 degrees of freedom. Again, α is determined by (3.7), for this $Z_{sample} \approx Z$, cf., (4.4), provided $M >> r$, $(r \geq 2)$. The test is, accordingly,

or $\left\{ \begin{array}{ll} \underline{\text{decide } H_o}: & \text{the sample ensemble belongs to a population with p.d. } W_1(X), \text{ if } Z_{sample} \leq Z_\alpha, \\ \underline{\text{decide } H_1}: & \text{the sample ensemble belongs to a population with a p.d.} \neq W_1(X), \text{ if } Z_{sample} > Z_\alpha. \end{array} \right.$ (4.5)

---

*The X's here can be standardized variables, e.g., $X_i = (Y_i - \overline{Y}_{exp})/\sigma_{Y_{exp}}$, where $\overline{Y}_{exp}$, σ^2_{exp} are the experimental means and variances of the original, unstandardized data $\{Y_i\}$. Or, equivalently, the X_i's are nonstandardized, and we use the experimental means and variances in the a priori p.d. $W_1(X_1)$, as above.

This is, of course, for range t_ℓ. We must repeat for various ranges $t_\ell \epsilon T$ in order to establish with more reasonable certainty that the reverberation has the same first-order p.d. as that postulated. With a target present (at ranges $t_\ell \epsilon t_T + \Delta T$, say), cf. Fig. (3.1), we would expect a non-fit to the postulated d.d. of reverberation alone, even when $W_1(X)$ for this is not correctly chosen. Again, from the structure of the physical model of reverberation (Ref. 9), and possible models of target, we should often be able to make a good guess at the form of $W_1(X)$.

Example:

Let us use the data of our reverberation example (cf., Table (3.1) earlier)to test whether the observed envelopes obey a Rayleigh distribution, e.g.,

$$W_1(X|H_o) = \frac{2Xe^{-X^2/\overline{X^2}}}{\overline{X^2}} \quad , X > 0 \quad . \tag{4.6}$$

Here $M = 200$; and $\nu_i = \nu_i + \mu_i$; $r = 10$, (Table (3.1)). The p_i are determined from (4.6) and (4.2). However, we do not know the true ensemble value of $\overline{X^2}$, so we are forced to use an estimate, namely the sample mean-square

$$\overline{X^2} \approx \langle X_1^2 \rangle_{exp} = \frac{\langle X_1^2 \rangle_{exp} + \langle X_2^2 \rangle_{exp}}{2} = 20.71, \quad (\text{cf., Table (3.1)}). \tag{4.7}$$

Consequently, our estimated p_i's are obtained from

$$p_i^* = \frac{2X_i}{20.71} e^{-X_i^2/20.71} \; ; (X_1 = 0; X_2 = 1, X_3 = 2, \text{etc.}), \tag{4.8}$$

and are specifically, with the accompanying values of ν_i':

$$\begin{array}{ll} p_1^* = 0.00; \; \nu_1' = 4 & p_6^* = 0.144; \; \nu_6' = 27 \\ p_2^* = 0.095; \; \nu_2' = 7 & p_7^* = 0.102; \; \nu_7' = 21 \\ p_3^* = 0.161; \; \nu_3' = 33 & p_8^* = 0.064; \; \nu_8' = 13 \\ p_4^* = 0.188; \; \nu_4' = 37 & p_9^* = 0.033; \; \nu_9' = 10 \\ p_5^* = 0.178; \; \nu_5' = 45 & p_{10}^* = 0.016; \; \nu_{10}' = 3 \end{array} \tag{4.8a}$$

Applying (4.8a) to Eq. (4.4) gives us (on discarding the values at $i = 1$, $X_1 = 0$)

$$Z_{sample} = 11.17 = \chi^2 \quad , \tag{4.9}$$

which with $9 = (r-1)$ degrees of freedom, gives $P(Z > Z_{sample}) \doteq 0.27$ from Table 1 (Appendix), which is certainly larger than our chosen significance level $\alpha = 0.05$. Consequently, we accept H_o: the observed data do fit a Rayliegh distribution (with experimental mean-square 20.71).

B. A Nonparmetric Test for "Goodness-of-Fit" (Kolmogorov-Smirnov Test):

This, again, is a Kolmogorov-Smirnov two-tailed test (Ref. 8, pp. 47-51 and Refs. 16 and 17), constructed as follows: the test statistic is now

$$Z_{sample} = \max_{-\infty < X_i(t_\ell) < \infty} |F(X(t_\ell)) - S_M(X(t_\ell))| \quad , \tag{4.10}$$

where F is the theoretical (or postulated) cumulative (first-order) distribution of the ordered $X_i(t_\ell)$, cf., Eq. (3.15), against which we are testing the experimental cumulative distribution $S_M(X_\ell)$. Thus,

$$F(X_i) = \int_{-\infty}^{X_i} W_1(X)\, dX \quad , \tag{4.11}$$

and we can use standardized variables for X, or apply the same experimental means and variances to the explicit structuring of F_1 through $W_1(X)$, (4.11), (cf., the footnote preceding Eq. (4.2)). The threshold at level α is now given by

$$Z_\alpha = \lambda_\alpha/\sqrt{M} \quad , \tag{4.12}$$

where λ_α is determined by (3.13b) and Table 3. Thus, the test takes the largest of the differences $|F-S_M|$ and compares it to the threshold Z_α in the usual way, cf., (3.12). (Figure (3.2) applies here with s_1 replaced by F and M by M_2.) The test is repeated for the different ranges, t_ℓ, of interest, including those about the target.

Example:

For the reverberation example treated in A above we find from (4.6), (4.7) in (4.11) that

$$F(X_i) = 1 - e^{-X_i^2/20.71}, \quad (X_1 = 0;\ X_2 = 1;\ X_i = i-1), \tag{4.13}$$

where again we have replaced the unknown value of $\overline{X^2}$ by the sample estimate $\langle x^2\rangle_{exp}$, (4.7). [This procedure is an approximation, of course, vis-à-vis the case for which Table 7 is calculated under the assumption of F(X) completely known, including its statistical parameters. However, there is evidence that when the K-S test is applied, with estimated parameters of the distribution used in place of the (unknown) true values, the employment of Table 7 will lead to a conservative test--i.e., when $Z_{sample} > Z_\alpha$, the hypothesis H_o may be even more safely rejected than in the case of F completely specified (and test at level α).] Equation (4.13) for our example is shown in Fig. (3.2), from which we obtain for (4.10)

$$Z_{sample} = 0.09, \quad M_2 = 100 \quad , \tag{4.14}$$

and $M = M_2$ for purposes of numerical illustration. [Of course, a better test would be to use all data $M = M_1+M_2 = 200$, and obtain Z_{sample} for $M = 200$. The present case is chosen simply for illustration.] From Table 7 we see that at the $\alpha = 0.05$ level of significance,

$$Z_{0.05} = \frac{1.36}{\sqrt{M_2}} = 0.136 > Z_{sample} = 0.09 \quad , \tag{4.15}$$

so that we retain H_o and accordingly say (for this significance level) that our data fit the Rayliegh distribution (4.13). This is consistant with our conclusion under A above, based on the χ^2-test.

C. A New Nonparametric Test for the Normal Distribution (Ref. 18):

In a great many applications it is reasonable, for physical reasons, to expect normal, or gaussian statistics--for instance, in the case of ambient sea noise (Ref. 17), and in our example discussed above, where the (instantaneous) reverberation amplitudes (as distinct from the envelope magnitudes) will generally obey a normal distribution if the number of independent illuminated scatterers is sufficiently large (Ref. 9, Sec. 10.1, 2). Accordingly, it is practically very important to be able to handle the gaussian cases in as powerful fashion as possible. Now, while both the χ^2 and Kolmogorov-Smirnov (K-S) tests are applicable to the problem of testing whether or not the sample ensemble belongs to a normal process, at least with respect to the first-order p.d. $W_1(W;\ t_\ell)$, a recent test, called the W-test for normality (Ref. 18), has been derived that is particularly powerful (more so than either the χ^2 and K-S test) and is quite simple to calculate. The test statistic here is

$$Z_{sample} = W_{sample} = b^2/S^2 \quad , \tag{4.16}$$

where, if we have M sample values $X^{(j)}(t_\ell)$, $j = 1, \ldots, M$, as before, we order the $\{X^{(j)}\} \rightarrow \{X_i\}$ into an ordered sample set, and compute

$$S^2 = \sum_{j=1}^{M} (X^{(j)} - \langle X \rangle)^2 = \sum_{i=1}^{M} (X_i - \langle X \rangle)^2 \quad , \tag{4.17}$$

where $\langle X \rangle$ = sample mean, and we obtain b from

$$\text{(M even):} \quad b = \sum_{i=1}^{M/2} a_{M-i+1}(X_{M-i+1} - X_i) \tag{4.18a}$$

$$\text{(M odd):} \quad b = a_M(X_M - X_1) + \cdots + a_{\frac{M+3}{2}}\left(X_{\frac{M+3}{2}} - X_{\frac{M-1}{2}}\right), \tag{4.18b}$$

where the a_n are obtained from Table 8 (Table 5 of Ref. 18). Next, we compute

W_{sample} according to (4.16) through (4.18b) and use Table 9 (Table 6 of Ref. 18) to find Z_α for various values of α. Table 9 gives values of Z_α for various values of

$$(1-\alpha)\ 100\% = 100 \int_{W_o}^{W_\alpha = Z_\alpha} dP(W|H_o) = 100 \int_{W_o}^{W_\alpha(>W_o)} W_1(W|H_o)\ dW \quad , \tag{4.19}$$

and sample-size M, from which one at once obtains

$$\alpha = \int_{W_\alpha}^{1} W_1(W|H_o)\ dW \quad . \tag{4.20}$$

The upper limit on $Z = W$ is unity, and there is a positive nonzero lower limit, W_o. Specifically, here, the test statistic W is scale-($\sim\sigma_X$) and origin-($\sim\bar{X}$) invariant, and has a p.d. which depends only on sample size M, for samples from a normal population. Furthermore, W is statistically independent of S^2 and of $\langle X \rangle$, the mean of the ordered sample values. (See p. 593, Ref.18.) Note here that <u>small</u> values of $Z_{sample} = W$ are significant--<u>i.e.</u>, indicate non-normality. However, by (4.19) and (4.20), the test remains,

$$\begin{cases} \underline{\text{decide } H_o\text{: normality of sample,}} \text{ if } Z_{sample} = W_{sample} < W_\alpha(=Z_\alpha) \\ \underline{\text{decide } H_1\text{: non-normality of sample,}} \text{ if } Z_{sample} > W_\alpha. \end{cases} \tag{4.21}$$

Again, we proceed in similar fashion for each range t_ℓ, including those covering the target.

V. <u>First- and Second-Order Sample Statistics:</u>

Frequently, the lower-order moments of the observed data provide useful information about the process and possible indication of stability, stationarity, etc. In particular, if the corresponding infinite population moments are known we can determine how finite sample-size influences their measurement.

If $\{X(t)\}$ is the observed reverberation (and target) process, consisting of M representations, statistics of interest to us are:

1. The sample mean:

$$\langle X(t)\rangle_{exp} \equiv \frac{1}{M}\sum_{j=1}^{M} X^{(j)}(t) \equiv m_\epsilon (\equiv \overline{X} \text{ of Wilks (Ref. 5), Sec. (8.2)}); \tag{5.1}$$

2. The sample intensity:

$$\langle X(t)^2\rangle_{exp} \equiv \frac{1}{M}\sum_{j=1}^{M} X^{(j)}(t)^2 \quad ; \tag{5.2}$$

3. The sample covariance:

$$\langle x(t_1)\ x(t_2)\rangle_{exp} \equiv K_x(t_1,t_2)_{exp} = \frac{1}{M}\sum_{j=1}^{M} x^{(j)}(t_1)\ x^{(j)}(t_2) \quad , \tag{5.3a}$$

$$x^{(j)}(t) \equiv X^{(j)}(t) - \langle X(t)\rangle_{exp} \quad . \tag{5.3b}$$

We also have the following relations:

$$\langle x(t)^2\rangle_{exp} \equiv \frac{1}{M}\sum_{j=1}^{M} x^{(j)}(t)^2 = \frac{1}{M}\sum_{j=1}^{M} (X^{(j)}(t)-m_\epsilon)^2 = \frac{M-1}{M}\, s_X{}^2 \quad , \tag{5.4a}$$

with

$$s_X{}^2 \equiv \frac{1}{M-1}\sum_{j=1}^{M} (X^{(j)}-m_\epsilon)^2 \quad , \tag{5.4b}$$

the usual sample variance (Ref. 5, Eq. 8.2.6). We have also

$$s_X^2 = \frac{M\langle X(t)^2\rangle_{exp} - Mm_\epsilon^2}{M-1} \text{ or } \langle X^2\rangle_{exp} = \left(\frac{M-1}{M}\right)s_X^2 + m_\epsilon^2 \quad . \tag{5.4c}$$

We assume in the above that in the set of representations $\{X^{(j)}\}$, the $X^{(j)}$ represent independent random variables, with the same p.d.'s. (This assumption, of course, must be justified, by procedures like those discussed in Sec. 3 above.)

Let

$$\left\{\begin{array}{l} \mu_X = \overline{X} = (\infty \text{ - population) ensemble* mean;} \\ \sigma_X^2 = \overline{X^2} - \overline{X}^2 = (\infty \text{ - population) ensemble* variance;} \\ \\ K_X(t_1,t_2) = \overline{x(t_1)x(t_2)} = \overline{(X_1-\overline{X}_1)(X_2-\overline{X}_2)} = (\infty \text{ - population) ensemble covariance.} \end{array}\right. \tag{5.5}$$

We want now:

1. The (ensemble)* mean of the sample mean: this shows whether there is any departure from, or "bias" between, the two measures, i.e., between the sample mean vs. ensemble mean;

2. The (ensemble)* variance of the sample mean: this gives a measure of spread in the experimental mean as a function of ensemble size and the true variance; and

3. The (ensemble)* covariance of the sample mean: this provides a measure of correlation between means at different times (ranges) and some idea of how close samples (in time) can be taken before such correlation becomes significant.

*"Ensemble" here refers to the theoretical, infinite population ensemble, not to the subensemble of the experimental data set.

From previous work* we find, finally, that (with $x \equiv X-\overline{X}$)

1. <u>Variance of the experimental variance</u> $[x_1=x_2=x_3=x_4;(x_i=x(t_i))]$:

$$\mathrm{var}\,\langle x_1^2\rangle_{exp} \equiv \langle x_1^2\rangle^2_{exp} - \langle x_1^2\rangle^2_{exp} = \frac{\overline{x_1^4} - \overline{x_1}^2}{M} \quad . \tag{5.6a}$$

(Also: $\overline{s_X^2} = \sigma_X^2;\ \overline{(s_X^2)^2} = \frac{1}{M}\left\{\mu_4 - \left(\frac{M-3}{M-1}\right)\sigma_X^4\right\};\ \mu_4 = \overline{(X-\overline{X})^4}$, (5.6b)

[cf., Wilks (Ref. 5), Eqs. 8.2.7, 8.2.9.])

2. <u>Variance of the experimental covariance</u> $[x_1=x_3;\ x_2=x_4]$:

$$\mathrm{var}\,\langle x_1x_2\rangle_{exp} = \sigma^2_{K_x\text{-}exp} \equiv \overline{K_x(t_1,t_2)^2_{exp}} - \overline{K_x(t_1,t_2)_{exp}}^2$$

$$= \frac{\overline{x_1^2x_2^2} - K^2_{x\text{-}12}}{M} \quad , \tag{5.7a}$$

with

$$K_{x\text{-}12} = \overline{x(t_1)x(t_2)} \quad . \tag{5.7b}$$

3. <u>Covariance of the experimental variance</u> $[x_1=x_2;\ x_3=x_4]$:

$$(\langle x_1^2\rangle_{exp} - \langle x_1\rangle^2_{exp})(\langle x_2^2\rangle_{exp} - \langle x_2\rangle^2_{exp}) \equiv K_{K_X}(t_1,t_1;\ t_3,t_3)$$

$$= \frac{\overline{x_1^2x_3^2} - \overline{x_1^2}\cdot\overline{x_3^2}}{M} \tag{5.8}$$

*D. Middleton, current ONR Research Notes.

4. <u>Covariance of the experimental covariance:</u>

$$K_{K_{x\text{-exp}}}(t_1,\ldots,t_4) = \overline{K_x(t_1,t_2)_{exp}\cdot K_x(t_3,t_4)_{exp}} - \overline{K_x(t_1,t_2)_{exp}}\cdot\overline{K_x(t_3,t_4)_{exp}} \qquad (5.9a)$$

$$= \frac{\overline{x_1x_2x_3x_4} - K_{x-12}\cdot K_{x-34}}{M} \quad . \qquad (5.9b)$$

By specializing (5.9) we get the other cases (5.6) through (5.8) directly. When the process x is gaussian we may use Eq. 7.29a of Ref. 1 to obtain directly

1. $\text{var}\left\langle x_1^{\,2}\right\rangle_{exp} = 2\,\overline{x_1^{\,2}}^{\,2}/M \quad ,$ (5.10a)

2. $\text{var}\left\langle x_1\,x_2\right\rangle_{exp} = \dfrac{K_{x-12}^{\,2} + \overline{x_1^{\,2}}\cdot\overline{x_2^{\,2}}}{M} \quad ,$ (5.10b)

3. $(K_{K_x})_{12} = 2\,\overline{x_1x_3}^{\,2}/M \quad ,$ (5.10c)

4. $(K_{K_x})_{1234} = (K_{x-13}\cdot K_{x-24} + K_{x-13}\cdot K_{x-14})/M.$ (5.10d)

A practical difficulty with using these moment relations is the fact that very often we do not know the ensemble average in question, <u>e.g.</u>, $\overline{x^2}$, K_{x-12}, etc. A number of partial resolutions of this difficulty is:

1. Use the sample statistics involved, e.g., for $\overline{x_1^2}$ use

$$\sum_j^M x^{(j)}(t_1)^2/M = \langle x(t_1)^2\rangle_{exp} \quad ,$$

etc., with the largest sample sizes available;

2. Use the alternative approach of making repeated observations on subensembles of the data, and estimate the variance (and mean) between samples (of these subensembles) by (5.6a), now given by

$$S_z^2 = \frac{1}{M-1}\sum_{k=1}^{M} (z_k - \langle z\rangle_{exp})^2 = \frac{1}{M-1}\sum_{k=1}^{M} z_k^2 - \frac{M}{M-1}\langle z\rangle_{exp}^2 \quad , \tag{5.11}$$

with

$$z_k \equiv \sum_{j=1}^{J} G_k(x^{(j)})/J \quad , \tag{5.12}$$

where G_k is the statistical form in question, e.g., $G_k = x^{(j)^2}$, with $j = j', \ldots, j' + J_k)$. Of course, S_z^2 is itself a random variable, whose p.d.'s depend on still other unknown parameters. (For some simple applications and a discussion of the implementing circuitry, see Korn (Ref. 19).)

VI. Simulation and Reality:

Simulation--in tank, by computer, in a lake--provides a powerful tool for obtaining quantitative results under varying degrees of control. The crucial problem in applying the results of simulation is to determine the extent to which the results are truly representative of the corresponding real-life situations. Statistical methods appear to be the needed link between simulation and the real-data environment. We very briefly sketch the approach:

1. First, verify that the ensemble data from simulation constitute a valid ensemble (testing for independence and homogeneity of the member representations, (cf. Sec. 3).

2. Next, establish the valid ensemble size for the "real" data to be studied.

3. Simulation, of course, must be carried out under as close similarity to the actual, known conditions under which the real-data ensemble was established. In practise, one may conveniently be able to treat a range of conditions in the simulation, and thus "straddle" the real environment's constraints.

4. Next, "calibrate" the simulation in terms of the real data set: specifically, for reverberation without target, we use the statistical measures of the real data to establish the form and relative scale of the distributions and their (lower-order) moments ("goodness-of-fit" tests and estimation, etc.). It is understood that the conditions of simulation are matched as closely as possible to those under which the real data were taken. Then, we seek isomorphisms between the results from the real data and the hoped-for equivalent simulation. The absolute scale is not important, but isomorphisms and departures from them on the part of the simulated data then provide quantitative and qualitative clues as to possible differences (and relative equivalences) between the real and simulated models.

5. Once an acceptable calibration of the simulation has been achieved, including known (relative) departures from the results of the real data, we can proceed to exploit the simulated model for other properties, checking back with the real data when feasible. This is, of course, a kind of boot-strap operation, but it is typical of the kind of joint statistical,

experimental interaction aided by physical insights appropriate to the problem at hand, that must be devised if one is to relate simulation and reality in a convincing fashion. The details of what statistical properties (usually first-order probability densities and first- and second-order moments to begin with) are particularly useful, will depend on the type of measurements undertaken. In any case, statistical methods provide the needed connection between reality and simulation, so that the techniques and results of the latter may be relied upon with confidence (determined by the significance levels, α, chosen).

VII. Concluding Remarks

In the preceding sections we have briefly described some of the principal features of statistical tests in general and specific tests in particular, for determining whether or not the observed data set represents a valid ensemble, whether two data sequences belong to the same underlying process, i.e., have the same distributions, and whether a data set has a particular (first-order) distribution. Examples of both parametric (i.e., distribution-dependent) and nonparametric (distribution-free) tests are provided. Generally, the nonparametric tests (Kolmogorov-Smirnov; Runs, W-test, etc.) are the most reliable, since they require few, if any, unavailable *a priori* statistical data. The Kolmogorov-Smirnov tests are also among the more powerful (larger $1-\beta$ values) even for small samples, with the χ^2-tests breaking down when sample-size is too small.

All of the above is cast in the particular framework of measuring the characteristics of reverberation data, which can contain a target at certain ranges, and is intended to provide a working preliminary for the analysis of data in such situations, with attention to target properties. Included, also, is a short "program" for relating simulated and "real" data, which is a critical problem in ultimately practical applications. The present discussion is far short of offering a comprehensive account of "best" rests for all such environments: one next step here is to extend, from working experience, this small collection of procedures, to still others that may prove particularly useful in future experiments.

Finally, we emphasize that statistical measures of data by themselves can be worse than useless: they can be seriously misleading*. We must always embed our statistical analysis as deeply as we can in the physical background of the experiment, and use this background as fully and carefully as possible in our choice of data selection processing, choice of tests, and interpretation. In this important sense practical statistical analysis remains an art, and is not purely a technical device to be used unthinkingly. In a later memorandum, we hope to extend the above, stressing the results for actual experiments.

*See, for example, the remarks of W. Feller, "Are Life Scientists Overawed by Statistics?" p. 24 of "Scientific Research," Vol. 4, No. 3, February 3, 1969 (published by McGraw-Hill).

APPENDIX

TABLE 1. TABLE OF CRITICAL VALUES OF CHI SQUARE*

df	Probability α under H_0 that $Z \geq Z_\alpha = \chi^2_\alpha$; α = .99	.98	.95	.90	.80	.70	.50	.30	.20	.10	.05	.02	.01	.001
1	Z_α = .00016	.00063	.0039	.016	.064	.15	.46	1.07	1.64	2.71	3.84	5.41	6.64	10.83
2	.02	.04	.10	.21	.45	.71	1.39	2.41	3.22	4.60	5.99	7.82	9.21	13.82
3	.12	.18	.35	.58	1.00	1.42	2.37	3.66	4.64	6.25	7.82	9.84	11.34	16.27
4	.30	.43	.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	11.67	13.28	18.46
5	.55	.75	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	13.39	15.09	20.52
6	.87	1.13	1.64	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	15.03	16.81	22.46
7	1.24	1.56	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	16.62	18.48	24.32
8	1.65	2.03	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	18.17	20.09	26.12
9	2.09	2.53	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	19.68	21.67	27.88
10	2.56	3.06	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	21.16	23.21	29.59
11	3.05	3.61	4.58	5.58	6.99	8.15	10.34	12.90	14.63	17.28	19.68	22.62	24.72	31.26
12	3.57	4.18	5.23	6.30	7.81	9.03	11.34	14.01	15.81	18.55	21.03	24.05	26.22	32.91
13	4.11	4.76	5.89	7.04	8.63	9.93	12.34	15.12	16.98	19.81	22.36	25.47	27.69	34.53
14	4.66	5.37	6.57	7.79	9.47	10.82	13.34	16.22	18.15	21.06	23.68	26.87	29.14	36.12
15	5.23	5.98	7.26	8.55	10.31	11.72	14.34	17.32	19.31	22.31	25.00	28.26	30.58	37.70
16	5.81	6.61	7.96	9.31	11.15	12.62	15.34	18.42	20.46	23.54	26.30	29.63	32.00	39.29
17	6.41	7.26	8.67	10.08	12.00	13.53	16.34	19.51	21.62	24.77	27.59	31.00	33.41	40.75
18	7.02	7.91	9.39	10.86	12.86	14.44	17.34	20.60	22.76	25.99	28.87	32.35	34.80	42.31
19	7.63	8.57	10.12	11.65	13.72	15.35	18.34	21.69	23.90	27.20	30.14	33.69	36.19	43.82
20	8.26	9.24	10.85	12.44	14.58	16.27	19.34	22.78	25.04	28.41	31.41	35.02	37.57	45.32
21	8.90	9.92	11.59	13.24	15.44	17.18	20.34	23.86	26.17	29.62	32.67	36.34	38.93	46.80
22	9.54	10.60	12.34	14.04	16.31	18.10	21.34	24.94	27.30	30.81	33.92	37.66	40.29	48.27
23	10.20	11.29	13.09	14.85	17.19	19.02	22.34	26.02	28.43	32.01	35.17	38.97	41.64	49.73
24	10.86	11.99	13.85	15.66	18.06	19.94	23.34	27.10	29.55	33.20	36.42	40.27	42.98	51.18
25	11.52	12.70	14.61	16.47	18.94	20.87	24.34	28.17	30.68	34.38	37.65	41.57	44.31	52.62
26	12.20	13.41	15.38	17.29	19.82	21.79	25.34	29.25	31.80	35.56	38.88	42.86	45.64	54.05
27	12.88	14.12	16.15	18.11	20.70	22.72	26.34	30.32	32.91	36.74	40.11	44.14	46.96	55.48
28	13.56	14.85	16.93	18.94	21.59	23.65	27.34	31.39	34.03	37.92	41.34	45.42	48.28	56.89
29	14.26	15.57	17.71	19.77	22.48	24.58	28.34	32.46	35.14	39.09	42.56	46.69	49.59	58.30
30	14.95	16.31	18.49	20.60	23.36	25.51	29.34	33.53	36.25	40.26	43.77	47.96	50.89	59.70

* Table 1 is abridged from Table IV of Fisher and Yates: *Statistical tables for biological, agricultural, and medical research*, published by Oliver and Boyd Ltd., Edinburgh, by permission of the authors and publishers.

APPENDIX

TABLE 2. TABLE OF CRITICAL VALUES K_{α} of $(M/2)Z_{sample}$ IN THE KOLMOGOROV-SMIRNOV TWO-SAMPLE TEST

(Small samples)**

$M_1 = M_2 = M/2$	One-tailed test*		Two-tailed test†	
	$\alpha = .05$	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$
3	3	—	—	—
4	4	—	4	—
5	4	5	5	5
6	5	6	5	6
7	5	6	6	6
8	5	6	6	7
9	6	7	6	7
10	6	7	7	8
11	6	8	7	8
12	6	8	7	8
13	7	8	7	9
14	7	8	8	9
15	7	9	8	9
16	7	9	8	10
17	8	9	8	10
18	8	10	9	10
19	8	10	9	10
20	8	10	9	11
21	8	10	9	11
22	9	11	9	11
23	9	11	10	11
24	9	11	10	12
25	9	11	10	12
26	9	11	10	12
27	9	12	10	12
28	10	12	11	13
29	10	12	11	13
30	10	12	11	13
35	11	13	12	
40	11	14	13	

**As modified from Seigel (Ref. 8, pp. 278)

APPENDIX

TABLE 3. TABLE OF CRITICAL VALUES Z_α OF Z_{sample} IN THE KOLMOGOROV-SMIRNOV TWO-SAMPLE TEST

(Large samples: two-tailed test)* (N_1, $N_2 \geq 40$)

Level of significance α	Value of $\hat{Z}_{sample}$ so large as to call for rejection of H_c at the indicated level of significance, where Z_{sample} = maximum $\lvert S_N(X)_1 - S_N(X)_2 \rvert$
.10	$Z_\alpha = 1.22\sqrt{\frac{N_1+N_2}{N_1 N_2}}$
.05	$1.36\sqrt{\frac{N_1+N_2}{N_1 N_2}}$
.025	$1.48\sqrt{\frac{N_1+N_2}{N_1 N_2}}$
.01	$1.63\sqrt{\frac{N_1+N_2}{N_1 N_2}}$
.005	$1.73\sqrt{\frac{N_1+N_2}{N_1 N_2}}$
.001	$1.95\sqrt{\frac{N_1+N_2}{N_1 N_2}}$

*Modified from Seigel (Ref. 8, pp. 279)

APPENDIX

TABLE 4. TABLE OF CRITICAL VALUES OF R IN THE RUNS TEST*

Given in the bodies of Table 4 and Table 5 are various critical values of R for various values of M_1 and M_2. For the one-sample runs test, any value of R which is equal to or smaller than that shown in Table 4 or equal to or larger than that shown in Table 5 is significant at the .05 level. (For the Wald-Wolfowitz two-sample runs test, any value of R which is equal to or smaller than that shown in Table 4 is significant at the .05 level.)

Table 4

M_1 \ M_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2											2	2	2	2	2	2	2	2	2
3					2	2	2	2	2	2	2	2	2	3	3	3	3	3	3
4				2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4
5			2	2	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5
6		2	2	3	3	3	3	4	4	4	4	5	5	5	5	5	5	6	6
7		2	2	3	3	3	4	4	5	5	5	5	5	6	6	6	6	6	6
8		2	3	3	3	4	4	5	5	5	6	6	6	6	6	7	7	7	7
9		2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8
10		2	3	3	4	5	5	5	6	6	7	7	7	7	8	8	8	8	9
11		2	3	4	4	5	5	6	6	7	7	7	8	8	8	9	9	9	9
12	2	2	3	4	4	5	6	6	7	7	7	8	8	8	9	9	9	10	10
13	2	2	3	4	5	5	6	6	7	7	8	8	9	9	9	10	10	10	10
14	2	2	3	4	5	5	6	7	7	8	8	9	9	9	10	10	10	11	11
15	2	3	3	4	5	6	6	7	7	8	8	9	9	10	10	11	11	11	12
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11	11	11	12	12
17	2	3	4	4	5	6	7	[illegible]	8	9	9	10	10	11	11	11	12	12	13
18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12	13	13
19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	13
20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	13	14

*Adapted from Seigel (Ref. 8, pp. 252-253)

APPENDIX

TABLE 5. TABLE OF CRITICAL VALUES OF R IN THE RUNS TEST

TABLE 5

N_1 \ N_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2																			
3																			
4				9	9														
5			9	10	10	11	11												
6			9	10	11	12	12	13	13	13	13								
7				11	12	13	13	14	14	14	14	15	15	15					
8				11	12	13	14	14	15	15	16	16	16	16	17	17	17	17	17
9					13	14	14	15	16	16	16	17	17	18	18	18	18	18	18
10					13	14	15	16	16	17	17	18	18	18	19	19	19	20	20
11					13	14	15	16	17	17	18	19	19	19	20	20	20	21	21
12					13	14	16	16	17	18	19	19	20	20	21	21	21	22	22
13						15	16	17	18	19	19	20	20	21	21	22	22	23	23
14						15	16	17	18	19	20	20	21	22	22	23	23	23	24
15						15	16	18	18	19	20	21	22	22	23	23	24	24	25
16							17	18	19	20	21	21	22	23	23	24	25	25	25
17							17	18	19	20	21	22	23	23	24	25	25	26	26
18							17	18	19	20	21	22	23	24	25	25	26	26	27
19							17	18	20	21	22	23	23	24	25	26	26	27	27
20							17	18	20	21	22	23	24	25	25	26	27	27	28

*Adapted from Seigel (Ref. 8, pp. 252-253)

APPENDIX

TABLE 6. TABLE OF PROBABILITIES ASSOCIATED WITH VALUES AS EXTREME AS OBSERVED VALUES OF z IN THE NORMAL DISTRIBUTION*

The body of the table gives one-tailed probabilities under H_0 of z. The left-hand marginal column gives various values of z to one decimal place. The top row gives various values to the second decimal place. Thus, for example, the one-tailed p of $z \geq .11$ or $z \leq -.11$ is $p = .4562$.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
	$p/2 = \alpha/2$									
.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
3.2	.0007									
3.3	.0005									
3.4	.0003									
3.5	.00023									
3.6	.00016									
3.7	.00011									
3.8	.00007									
3.9	.00005									
4.0	.00003									

*Taken from Siegel (Ref. 8, pp. 247)

APPENDIX

TABLE 7. TABLE OF CRITICAL VALUES OF Z IN THE KOLMOGOROV-SMIRNOV ONE-SAMPLE TEST*

Sample size M	Level of significance for Z_{sample} = maximum $\lvert F(X) - S_M(X) \rvert$				
	α = .20	.15	.10	.05	.01
1	Z_α = .900	.925	.950	.975	.995
2	.684	.726	.776	.842	.929
3	.565	.597	.642	.708	.828
4	.494	.525	.564	.624	.733
5	.446	.474	.510	.565	.669
6	.410	.436	.470	.521	.618
7	.381	.405	.438	.486	.577
8	.358	.381	.411	.457	.543
9	.339	.360	.388	.432	.514
10	.322	.342	.368	.410	.490
11	.307	.326	.352	.391	.468
12	.295	.313	.338	.375	.450
13	.284	.302	.325	.361	.433
14	.274	.292	.314	.349	.418
15	.266	.283	.304	.338	.404
16	.258	.274	.295	.328	.392
17	.250	.266	.286	.318	.381
18	.244	.259	.278	.309	.371
19	.237	.252	.272	.301	.363
20	.231	.246	.264	.294	.356
25	.21	.22	.24	.27	.32
30	.19	.20	.22	.24	.29
35	.18	.19	.21	.23	.27
Over 35	$\frac{1.07}{\sqrt{N}}$	$\frac{1.14}{\sqrt{N}}$	$\frac{1.22}{\sqrt{N}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.63}{\sqrt{N}}$

*Adapted from Seigel (Ref. 8, pp. 251)

APPENDIX

TABLE 8. COEFFICIENTS a_{M-i+1} FOR THE W TEST FOR NORMALITY,*
FOR M = 2(1)50.

i \ M	2	3	4	5	6	7	8	9	10
1	0·7071	0·7071	0·6872	0·6646	0·6431	0·6233	0·6052	0·5888	0·5739
2	—	·0000	·1677	·2413	·2806	·3031	·3164	·3244	·3291
3	—	—	—	·0000	·0875	·1401	·1743	·1976	·2141
4	—	—	—	—	—	·0000	·0561	·0947	·1224
5	—	—	—	—	—	—	—	·0000	·0399

i \ M	11	12	13	14	15	16	17	18	19	20
1	0·5601	0·5475	0·5359	0·5251	0·5150	0·5056	0·4968	0·4886	0·4808	0·4734
2	·3315	·3325	·3325	·3318	·3306	·3290	·3273	·3253	·3232	·3211
3	·2260	·2347	·2412	·2460	·2495	·2521	·2540	·2553	·2561	·2565
4	·1429	·1586	·1707	·1802	·1878	·1939	·1988	·2027	·2059	·2085
5	·0695	·0922	·1099	·1240	·1353	·1447	·1524	·1587	·1641	·1686
6	0·0000	0·0303	0·0539	0·0727	0·0880	0·1005	0·1109	0·1197	0·1271	0·1334
7	—	—	·0000	·0240	·0433	·0593	·0725	·0837	·0932	·1013
8	—	—	—	—	·0000	·0196	·0359	·0496	·0612	·0711
9	—	—	—	—	—	—	·0000	·0163	·0303	·0422
10	—	—	—	—	—	—	—	—	·0000	·0140

i \ M	21	22	23	24	25	26	27	28	29	30
1	0·4643	0·4590	0·4542	0·4493	0·4450	0·4407	0·4366	0·4328	0·4291	0·4254
2	·3185	·3156	·3126	·3098	·3069	·3043	·3018	·2992	·2968	·2944
3	·2578	·2571	·2563	·2554	·2543	·2533	·2522	·2510	·2499	·2487
4	·2119	·2131	·2139	·2145	·2148	·2151	·2152	·2151	·2150	·2148
5	·1736	·1764	·1787	·1807	·1822	·1836	·1848	·1857	·1864	·1870
6	0·1399	0·1443	0·1480	0·1512	0·1539	0·1563	0·1584	0·1601	0·1616	0·1630
7	·1092	·1150	·1201	·1245	·1283	·1316	·1346	·1372	·1395	·1415
8	·0804	·0878	·0941	·0997	·1046	·1089	·1128	·1162	·1192	·1219
9	·0530	·0618	·0696	·0764	·0823	·0876	·0923	·0965	·1002	·1036
10	·0263	·0368	·0459	·0539	·0610	·0672	·0728	·0778	·0822	·0862
11	0·0000	0·0122	0·0228	0·0321	0·0403	0·0476	0·0540	0·0598	0·0650	0·0697
12	—	—	·0000	·0107	·0200	·0284	·0358	·0424	·0483	·0537
13	—	—	—	—	·0000	·0094	·0178	·0253	·0320	·0381
14	—	—	—	—	—	—	·0000	·0084	·0159	·0227
15	—	—	—	—	—	—	—	—	·0000	·0076

*Adapted from Shapiro and Wilk (Ref. 18, pp. 603-604)

TABLE 8. COEFFICIENTS a_{M-i+1} FOR THE W TEST FOR NORMALITY, FOR M = 2(1)50 (cont.)

i \ M	31	32	33	34	35	36	37	38	39	40
1	0·4220	0·4188	0·4156	0·4127	0·4096	0·4068	0·4040	0·4015	0·3989	0·3964
2	·2921	·2898	·2876	·2854	·2834	·2813	·2794	·2774	·2755	·2737
3	·2475	·2463	·2451	·2439	·2427	·2415	·2403	·2391	·2380	·2368
4	·2145	·2141	·2137	·2132	·2127	·2121	·2116	·2110	·2104	·2098
5	·1874	·1878	·1880	·1882	·1883	·1883	·1883	·1881	·1880	·1878
6	0·1641	0·1651	0·1660	0·1667	0·1673	0·1678	0·1683	0·1686	0·1689	0·1691
7	·1433	·1449	·1463	·1475	·1487	·1496	·1505	·1513	·1520	·1526
8	·1243	·1265	·1284	·1301	·1317	·1331	·1344	·1356	·1366	·1376
9	·1066	·1093	·1118	·1140	·1160	·1179	·1196	·1211	·1225	·1237
10	·0899	·0931	·0961	·0988	·1013	·1036	·1056	·1075	·1092	·1108
11	0·0739	0·0777	0·0812	0·0844	0·0873	0·0900	0·0924	0·0947	0·0967	0·0986
12	·0585	·0629	·0669	·0706	·0739	·0770	·0798	·0824	·0848	·0870
13	·0435	·0485	·0530	·0572	·0610	·0645	·0677	·0706	·0733	·0759
14	·0289	·0344	·0395	·0441	·0484	·0523	·0559	·0592	·0622	·0651
15	·0144	·0206	·0262	·0314	·0361	·0404	·0444	·0481	·0515	·0546
16	0·0000	0·0068	0·0131	0·0187	0·0239	0·0287	0·0331	0·0372	0·0409	0·0444
17	—	—	·0000	·0062	·0119	·0172	·0220	·0264	·0305	·0343
18	—	—	—	—	·0000	·0057	·0110	·0158	·0203	·0244
19	—	—	—	—	—	—	·0000	·0053	·0101	·0146
20	—	—	—	—	—	—	—	—	·0000	·0049

i \ M	41	42	43	44	45	46	47	48	49	50
1	0·3940	0·3917	0·3894	0·3872	0·3850	0·3830	0·3808	0·3789	0·3770	0·3751
2	·2719	·2701	·2684	·2667	·2651	·2635	·2620	·2604	·2589	·2574
3	·2357	·2345	·2334	·2323	·2313	·2302	·2291	·2281	·2271	·2260
4	·2091	·2085	·2078	·2072	·2065	·2058	·2052	·2045	·2038	·2032
5	·1876	·1874	·1871	·1868	·1865	·1862	·1859	·1855	·1851	·1847
6	0·1693	0·1694	0·1695	0·1695	0·1695	0·1695	0·1695	0·1693	0·1692	0·1691
7	·1531	·1535	·1539	·1542	·1545	·1548	·1550	·1551	·1553	·1554
8	·1384	·1392	·1398	·1405	·1410	·1415	·1420	·1423	·1427	·1430
9	·1249	·1259	·1269	·1278	·1286	·1293	·1300	·1306	·1312	·1317
10	·1123	·1136	·1149	·1160	·1170	·1180	·1189	·1197	·1205	·1212
11	0·1004	0·1020	0·1035	0·1049	0·1062	0·1073	0·1085	0·1095	0·1105	0·1113
12	0891	·0909	·0927	·0943	·0959	·0972	·0986	·0998	·1010	·1020
13	·0782	·0804	·0824	·0842	·0860	·0876	·0892	·0906	·0919	·0932
14	·0677	·0701	·0724	·0745	·0765	·0783	·0801	·0817	·0832	·0846
15	·0575	·0602	·0628	·0651	·0673	·0694	·0713	·0731	·0748	·0764
16	0·0476	0·0506	0·0534	0·0560	0·0584	0·0607	0·0628	0·0648	0·0667	0·0685
17	·0379	·0411	·0442	·0471	·0497	·0522	·0546	·0568	·0588	·0608
18	·0283	·0318	·0352	·0383	·0412	·0439	·0465	·0489	·0511	·0532
19	·0188	·0227	·0263	·0296	·0328	·0357	·0385	·0411	·0436	·0459
20	·0094	·0136	·0175	·0211	·0245	·0277	·0307	·0335	·0361	·0386
21	0·0000	0·0045	0·0087	0·0126	0·0163	0·0197	0·0229	0·0259	0·0288	0·0314
22		—	·0000	·0042	·0081	·0118	·0153	·0185	·0215	·0244
23	—	—	—	—	·0000	·0039	·0076	·0111	·0143	·0174
24	—	—	—	—	—	—	·0000	·0037	·0071	·0104
25	—	—	—	—	—	—	—	—	·0000	·0035

TABLE 9. PERCENTAGE POINTS OF THE W TEST* FOR M = 3(1)50

M	α = 0.99	0.98	0.95	0.90	0.50	0.40	0.05	0.02	0.01
	Level								
	λ = 0.01	0.02	0.05	0.10	0.50	0.90	0.95	0.98	0.99
$W_\alpha = Z_\alpha$ → 3	0.753	0.756	0.767	0.789	0.759	0.998	0.999	1.000	1.000
4	·687	·707	·748	·792	·935	·987	·002	·906	·907
5	·686	·715	·762	·806	·927	·979	·986	·991	·993
6	0·713	0·743	0·788	0·826	0·027	0·974	0·981	0·986	0·989
7	·730	·760	·803	·838	·028	·972	·979	·985	·988
8	·749	·778	·818	·851	·932	·972	·978	·984	·987
9	·764	·791	·829	·859	·935	·972	·978	·984	·986
10	·781	·806	·842	·869	·938	·972	·978	·983	·986
11	0·792	0·817	0·850	0·876	0·940	0·973	0·979	0·984	0·986
12	·805	·828	·859	·883	·943	·973	·979	·984	·986
13	·814	·837	·866	·889	·945	·974	·979	·984	·986
14	·825	·846	·874	·895	·947	·975	·980	·984	·986
15	·835	·855	·881	·901	·950	·975	·980	·984	·987
16	0·844	0·863	0·887	0·906	0·952	0·976	0·981	0·985	0·987
17	·851	·869	·892	·910	·954	·977	·981	·985	·987
18	·858	·874	·897	·914	·956	·978	·982	·986	·988
19	·863	·879	·901	·917	·957	·978	·982	·986	·988
20	·868	·884	·905	·920	·959	·979	·983	·986	·988
21	0·873	0·888	0·908	0·923	0·960	0·980	0·983	0·987	0·989
22	·878	·892	·911	·926	·961	·980	·984	·987	·989
23	·881	·895	·914	·928	·962	·981	·984	·987	·989
24	·884	·898	·916	·930	·963	·981	·984	·987	·989
25	·888	·901	·918	·931	·964	·981	·985	·988	·989
26	0·891	0·904	0·920	0·933	0·965	0·982	0·985	0·988	0·989
27	·894	·906	·923	·935	·965	·982	·985	·988	·990
28	·896	·908	·924	·936	·966	·982	·985	·988	·990
29	·898	·910	·926	·937	·966	·982	·985	·988	·900
30	·900	·912	·927	·939	·967	·983	·985	·988	·900
31	0·902	0·914	0·929	0·940	0·967	0·983	0·986	0·988	0·990
32	·904	·915	·930	·941	·968	·983	·986	·988	·990
33	·906	·917	·931	·942	·968	·983	·986	·989	·990
34	·908	·919	·933	·943	·969	·983	·986	·989	·990
35	·910	·920	·934	·944	·969	·984	·986	·989	·990
36	0·912	0·922	0·935	0·945	0·970	0·984	0·986	0·989	0·990
37	·914	·924	·936	·946	·970	·984	·987	·989	·990
38	·916	·925	·938	·947	·971	·984	·987	·989	·990
39	·917	·927	·939	·948	·971	·984	·987	·989	·991
40	·919	·928	·940	·949	·972	·985	·987	·989	·991
41	0·920	0·929	0·941	0·950	0·972	0·985	0·987	0·989	0·991
42	·922	·930	·942	·951	·972	·985	·987	·989	·991
43	·923	·932	·943	·951	·973	·985	·987	·990	·991
44	·924	·933	·944	·952	·973	·985	·987	·990	·991
45	·926	·934	·945	·953	·973	·985	·988	·990	·991
46	0·927	0·935	0·943	0·953	0·974	0·985	0·988	0·990	0·991
47	·928	·936	·946	·954	·974	·985	·988	·990	·991
48	·929	·937	·947	·954	·974	·985	·988	·990	·991
49	·929	·937	·947	·955	·974	·985	·988	·990	·991
50	·930	·938	·947	·955	·974	·985	·988	·990	·991

*Adapted from Shapiro and Wilk (Ref. 18, pp. 605)

REFERENCES

1. Middleton, D., An Introduction to Statistical Communication Theory, McGraw-Hill Book Company, Inc., New York, N. Y., 1960, Secs. 1.3 - 1.6.

2. Cramér, H., Mathematical Methods of Statistics, Princeton University Press, Princeton, New Jersey, 1946; cf. Chapter 30.

3. Hoel, P. G., Introduction to Mathematical Statistics, John Wiley & Sons, Inc., New York, N. Y. 1947; cf. Chapter 11.

4. Mood, A. F., Introduction to the Theory of Statistics, McGraw-Hill Book Company, Inc., New York, N. Y., 1950.

5. Wilks, S. S., Mathematical Statistics, John Wiley & Sons, Inc., New York, N. Y., 1962; cf. Chapter 14.

6. Fraser, D. A. S., Nonparametric Methods in Statistics, John Wiley & Sons, Inc., New York, N. Y., 1957.

7. Bendat, J. S. and A. G. Piersol, Measurement and Analysis of Random Data, John Wiley & Sons, Inc., New York, N. Y., 1966.

8. Siegel, S., Nonparametric Statistics for the Behaviorial Sciences, McGraw-Hill Book Company, Inc., New York, N. Y., 1956.

9. Middleton, D., "A Statistical Theory of Reverberation and Similar First-Order Scattered Fields," Trans. IEEE on Information Theory, IT-13, No. 3, pp. 372-414, July 1967, Sec. 10.

10. Yaglom, A. M., Introduction to the Theory of Stationary Random Processes, Prentice-Hall, Inc., New York, N. Y., 1962.

11. Middleton, D., An Introduction to Statistical Communication Theory, McGraw-Hill Book Company, Inc., New York, N. Y., 1960, Sec. 19.2-1.

12. Hoel, P. G., Introduction to Mathematical Statistics, John Wiley & Sons, Inc., New York, N. Y. 1947; cf. Chapter 11, pp. 215-216.

13. Middleton, D., An Introduction to Statistical Communication Theory, McGraw-Hill Book Company, Inc., New York, N. Y., 1960, Secs. 1.3, 1.4.

14. Cramér, H., Mathematical Methods of Statistics, Princeton University Press, Princeton, New Jersey, 1946; cf. Chapter 30, pp. 447, 448.

15. Middleton, D., Topics in Communication Theory, McGraw-Hill Book Company, Inc., New York, N. Y., 1965.

16. Wilks, S. S., Mathematical Statistics, John Wiley & Sons, Inc., New York, N. Y., 1962; cf. Chapter 14, Sec. 14.3 6(e), pp. 454 457.

17. Arase, T., and E. M. Arase, "Deep-Sea Ambient-Noise Statistics," J. Acoust. Soc. Am., 44, No. 6, 1679-1684, December 1968, for some pertinent applications.

18. Shapiro, S. S., and M. B. Wilk, "An Analysis of Variance Test for Normality," Biometrika, 52, 3 and 4, pp. 591-611, 1965.

19. Korn, G. A., Random Process Simulation and Measurements, McGraw-Hill Book Company, Inc., New York, N. Y., 1966, Secs. 5-7, 11, 14, Chapter 6.

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) Applied Research Laboratories The University of Texas at Austin Austin, Texas 78712	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP ---
3. REPORT TITLE Acoustic Modeling, Simulation, and Analysis of Complex Underwater Targets II. Statistical Evaluation of Experimental Data	
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Technical Report No. 69-22	
5. AUTHOR(S) (First name, middle initial, last name) David Middleton	
6. REPORT DATE 26 June 1969	7a. TOTAL NO. OF PAGES 75 7b. NO. OF REFS 19
8a. CONTRACT OR GRANT NO N00024-69-C-1129 b. PROJECT NO. SF 11121100, Task 08212 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) ARL-TR-69-22 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) ----
10. DISTRIBUTION STATEMENT This document has been approved for public release and sale; its distribution is unlimited.	
11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Naval Ship Systems Command Department of the Navy Washington, D. C. 20360

13. ABSTRACT

Some methods of analyzing experimental data for complex targets and reverberation are outlined, including a heuristic discussion of what constitutes an ensemble and how to obtain one experimentally. The problem of then determining whether or not the actual data form a valid ensemble is considered. Tests for the stability, or stationarity, of the underlying random mechanism are briefly described (test of independence and homogeneity); tests of whether or not a particular data set belongs to some postulated probability destribution are provided (goodness-of-fit), including a powerful test for establishing the normality or non-normality of the sample. Among the tests considered are the χ^2, the Kolmogorov-Smirnov, the runs test, and the W-test for normality. These tests are carried out for some hypothetical reverberation data to illustrate the individual tests, at particular ranges, which can include the domain of a target reverberation. Some second-order properties of various classes of second moments of these data are also discussed, and an approach to relating simulated data to those from a real environment is briefly sketched. This memorandum is intended as a preliminary guide to the statistical treatment of data that are obtained in target and background measurements. As a subsequent task, additional tests and techniques remain to be chosen for this class of problems, including the explicit analysis of data already obtained.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6801

UNCLASSIFIED

Security Classification

UNCLASSIFIED
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Statistical Evaluation of Reverberation Data						
Reverberation						
Tests for Ensemble						
Tests for Goodness-of-Fit						
Target Data						
Simulation Data						
Sample Statistics						
Simulation and Reality						

DD FORM 1 NOV 65 1473 (BACK)
(PAGE 2)

UNCLASSIFIED
Security Classification