63-4-1

407 122

RADC-TDR-62-499

September, 1962

ASTIA Document No:

Development of Non-Parametric Techniques to the Reliability Testing of Air Force Ground Electronic Equipment

Thomas J. Burke

Daniel Goss

Contract AF-30 (602)-2282

Utica College of Syracuse University

Rome Air Development Center

Air Force Systems Command

United States Air Force

Griffiss Air Force Base, New York

AD 407122

407122

Errata

Recently, Technical Report RADC-TDR-62-499 dated September 1962, and entitled "Development of Non-Parametric Techniques to the Reliability Testing of Air Force Ground Electronic Equipment", and written by T. Burke and D. Goss, and prepared for the Rome Air Development Center, USAF, Griffiss Air Force Base, New York was sent to you.

Attach this errata sheet containing the three notices listed below to the inside of the front cover of the above report. (in accordance with the requirements of RADC 3002C, section 3.5.2)

1. PATENT NOTICE: When government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

2. DDC NOTICE: Qualified requestors may obtain copies of this report from the DDC Document Service Center, Cameron Station, Alexandria, Virginia. DDC Services for the Department of Defense contractors are available through the "Field of Interest Register" on a "need-to-know" certified by the cognizant military agency of their project or contract.

3. OTS NOTICE: This report has been released to the Office of Technical Services U. S. Department of Commerce, Washington 25, D.C., for sale to the general public.

RADC-TDR-62-499

September, 1962

ASTIA Document No:

Development of Non-Parametric Techniques to the Reliability Testing of Air Force Ground Electronic Equipment

Thomas J. Burke

Daniel Goss

Contract AF-30 (602)-2282

Utica College of Syracuse University

Rome Air Development Center

Air Force Systems Command

United States Air Force

Griffiss Air Force Base, New York

FORWARD

This effort had as its purpose the study of:

(a) Randomization Tests of Statistical Inference. These tests may be utilized to minimize the risk of biased results resulting from non-random sets of failure data. Since, generally, reliability estimations must be based upon the assumption that the occurring failures are random in nature, such tests of randomness may be considered as a necessity.

(b) Statistical Decision Functions. These functions may be used to determine the advisability of terminating reliability tests in the presence of relatively sparse data. From decision functions, risk functions can be developed which may be capable of mathematically equating potential costs to each possible alternate decision (accept, reject, or continue test).

In the area of randomization tests the contractor has investigated and modified (for reliability test purposes) two testing procedures, the runs test and the serial correlation test.

Of the two, the runs test is the easiest to apply (can be implemented using a simple table). The runs test can also be used to determine whether or not a modification or redesign of an equipment has resulted in a significant improvement in reliability.

The serial correlation test, although more involved than the runs test, can readily be applied if one is willing to set up and solve a series of elementary statistical relationships.

In the area of statistical decision functions the contractor has

performed a literature search and analysis of all available data on the subject. The conclusion was reached that it may well be possible to develop a risk function for reliability test purposes. However, before this is accomplished more knowledge concerning appropriate decision functions and their practical ramifications must be acquired.

ABSTRACT

This is a study of several randomization tests of statistical inference. For testing the randomness of a sequence of observations, the Runs Test and the Serial Correlation Test are discussed.

Some aspects of decision theory are discussed. Among the topics considered are definitions, basic principles, the pay-off matrix, utility theory, and principles of choice. The report concludes with critical comments on utility theory and principles of choice.

TABLE OF CONTENTS

page

Introduction 1

Discussion 2

1. Randomization Tests

a) Random Sampling. 2
b) Runs Test. 3
c) Serial Correlation Test. 9

2. Statistical Decision Theory

a) Definitions and Basic Principles 14
b) Utility. 16
c) The Pay-Off Matrix 18
d) Principles of Choice 19
e) Bayes Estimate and Sequential Decision Problems 21
f) Critical Comments on Utility Theory and Principles of Choice 23

Conclusions. 27

Bibliography 29

Development of Non-Parametric Techniques to the Reliability Testing of Air Force Ground Electronic Equipment

INTRODUCTION

Randomization tests are necessary to avoid biased results when the experiment (as is almost always the case) is based on conditions of random sampling. If there is any slight suspicion that the observations do not form a random set when taken over some time interval, it is important to test for randomness before applying any statistical technique. The Runs Test and the Serial Correlation Test, which can be used to determine a lack of randomness in sequences of observations if such a lack exists, are discussed. These tests are nonparametric since they can be used when the underlying frequency is unknown. Examples are given which illustrate how the Runs Test can be used to test for the identity of two distributions.

The works of von Neumann and Morgenstern (1944, 1947) and Wald (1950) concerning the theory of games and the theory of statistical decisions stimulated much of the basic research in decision theory during the last decade. In the second part of this report it is our purpose to discuss some of the basic concepts of this theory. Decision functions can often be used to determine a rule for carrying out the reliability experimentation and for making a terminal decision. In this report, we will point out the difficulties encountered in bridging the gap between theory and practice.

DISCUSSION

1. Randomization Tests

a) Random Sampling

A set of elements which have a common measurable or observable characteristic is usually called the population or the universe. A subset of these elements, chosen in any manner, is called a sample of the population.

One type of population consists of elements or observations which actually exist. Examples of this type of universe might be: (a) all the registered voters of Philadelphia and (b) the seven members of the Board of Education of a certain city. In studying some characteristic of the voters of Philadelphia, it would be impractical to contact every voter whereas it would be reasonable to contact only a subset or sample of this population. Usually it is possible to obtain adequate information for most purposes from relatively small samples. In the case of the Board of Education members, because of the small number, it probably would be easy to measure every element (individual) of the population. Both of the above were illustrations of finite populations. A second type of population, usually involved in experiments, is obtained if we consider all the hypothetical measurements of the weight of an object. Also, the population consisting of all the hypothetical tosses of a die is a infinite population since "all possible tosses" can never be made.

The purpose of a statistical test is to make some generalization [the meantime between failures is X hours] about a population from a subset or sample of the population. The way we choose this sample plays

an important part in the degree of confidence which we can put on the results of the experiment. If some individual elements of our population are more likely to be chosen than others, then the sample is certainly biased. Whether the population is finite or infinite, we would like every element of the population to have an equal chance of being included in the sample. A sample satisfying this condition is called a random sample. This definition implies that some device must be used so that selection of elements of the sample be left to chance. However, in reliability tests aimed at verifying a mean time to failure, after each failure, the time to failure is recorded, the equipment is repaired and then put back on test. The assumption is that failure i and failure i+1 are independent. It is important to test this assumption since most techniques for testing statistical hypotheses assume a random set from some population. In particular, in the median and other tests which involve sequences of observations, it is assumed that these sequences are random. Therefore, before applying these statistical methods, some methods must be used to test the randomness of the sequence. The following two sections discuss such techniques.

b) The Run Test

One of the most useful and easiest to apply tests of the randomness of a sequence is the runs test. In order to illustrate the use of this test consider the following set of times to failure of a certain electronic equipment gathered over a period of several months:

(I) 31, 27, 37, 41, 32, 36, 28, 23, 41, 30, 37, 24, 19, 26, 39, 35, 25, 28, 40, 23.

We then found the median. In this case, since we have 10 elements above and including 31 and 10 elements below and including 30, the median is 30.5. We now replace each mean time between failures by + if above the

median and by - if below the median of 30.5 This yields the following sequence:

(II) +-++++-+-+---++-+-

A run is defined as a sequence of identical letters or symbols which is followed and preceded by different letters or no symbols at all. In the above sequence of + and - signs, we find that we first have a run of 1, then another run of 1, and then a run of 4, next a run of 2 etc. for a total number of 12 runs. This total number of runs which we will call υ is often a good indication of a possible lack of randomness. For example, υ would equal 3 if the twenty elements of sequence (II) were arranged as follows:

(III) ++++++----------+++

This could mean that we have too few runs, a total number much smaller than that expected under the randomness hypothesis. In another example, υ = 18 if the twenty elements of sequence (II) were arranged in some alternating or almost alternating pattern such as:

(IV) +-+-+-+-++-+-+-+-+--

This probably means that we have <u>too many runs</u>, a number much larger than we could expect by chance. In either of the last two illustrations we probably would reject the hypothesis of randomness.

In order to determine whether υ, the observed total number of runs, is too few or too many, let us consider an arrangement of n_1 letters or symbols of one kind and n_2 letters or symbols of the second kind. If we assume that this sequence is a random sample from a given population, it is possible to obtain the sampling distribution of the variable υ for repeated random arrangements by the laws of probability. These probabilities have been used to construct tables which enable us to

test whether a sample value of χ is unusually small or large. The following are portions of tables complied for several levels of significance which enable us to make exact tests when n_1 and n_2 are small (less than 20):

TABLE I—Critical values of χ in the Runs Test for $\chi_{.025}$ and $\chi_{.975}$. (Ref. 6)

n_1 \ n_2	5	8	10	12	14	16	18	20
3		2 7	2 7	2 7	2 7	3 7	3 7	3 7
5	2 9	3 10	3 11	4 11	4 11	4 11	5 11	5 11
8		4 13	5 14	6 15	6 15	6 16	7 16	7 16
10			6 15	7 16	7 17	8 18	8 18	9 19
12				7 18	8 19	9 20	9 20	10 21
14					9 20	10 21	10 22	11 23
16						11 22	11 24	12 24
18							12 25	13 26
20								14 27

The values listed are such that a number less than or equal to the $\chi_{.025}$ value (in the upper left hand corner of each rectangle) will occur not more than 2½% of the time; and a number greater than or equal to $\chi_{.975}$ value (in the lower right hand corner of rectangle) will occur not more than 2.5% of the time.

Example (1):

Suppose we have the following set of times between failure of a certain electronic equipment: 31, 27, 37, 41, 32, 36, 28, 23, 41, 30, 37, 24, 19, 26, 39, 35, 25, 28, 40, 23. We wish to test the randomness of this sequence of observations at the 5% level of significance.

Solution:

As shown on page 3, the median of this sequence is 30.5. Designating a mean time between failures above the median by + and mean time between failures below the median by -, the above sequence becomes:

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

The total number of runs in this sequence is $u = 12$. The hypothesis H_o that we wish to test is that +'s and -'s occur in random order. The alternate hypothesis H_1 is that the order of +'s and -'s (or the total number of runs) is not random. We use a two tail test since we cannot predict the direction of the deviation from randomness. In this case, $n_1=n_2=10$ and the $u = 12$. From Table I, we see that the number 12 lies between $u_{.025}=6$ and $u_{.975}=16$ and therefore we do not reject the hypothesis of randomness at the 5% level of significance.

Table I can be used when n_1 and n_2 are equal to or less than 20. More extensive tables are available but it has been shown that if n_1 and n_2 are greater than 10, then the sampling distribution of u is approximately normal with the mean and standard deviation given by the following:

(A) $$\mu_u = \frac{2n_1n_2}{n_1+n_2} + 1$$

(B) $$\sigma_u^2 = \frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$$

Example (2)

We wish to test the randomness of a sequence of 50 observations if $n_1=25$ and $n_2=25$ and the number of runs is 17, at the 5% level of significance.

Solution

Substituting in formulas (A) and (B) give:

$$\mu_v = \frac{(2)(25)(25)}{50} + 1 = 26$$

$$\sigma_v^2 = \frac{(1250)(1250-50)}{(2500)(49)} = 12.2 = (3.5)^2$$

(C) The formula

$$\bar{Z} = \frac{v - \mu_v}{\sigma_v}$$

tells us the number of standard deviations a particular Z is from the mean. Substituting in this formula (C) we obtain

$$\bar{Z} = \frac{17-26}{3.5} = \frac{-9}{3.5} = -2.6$$

Since, in a normal distribution, 95% of the cases lie between -1.96σ and $+1.96\sigma$ from the mean, we reject the hypothesis of randomness. In particular $26 \pm (1.96)(3.5)$ or 26 ± 7 is the region of acceptance of the hypothesis of randomness at the 5% level.

The runs test can also be used to determine whether or not two random samples are from populations having the same frequency distributions. This is useful in testing whether or not a modification or a redesign of equipment has resulted in a significant improvement. The following examples will illustrate this technique.

Example (3)

Suppose 8 observations of the times to failure of an equipment have been recorded (sample A) and after a redesign, 10 new measurements

are made (sample B):

(A) 23, 32, 51, 43, 45, 33, 58, 37

(B) 38, 27, 47, 38, 40, 55, 61, 49, 43, 52

Is this apparent improvement [a mean of 40.3 in A to a mean of 45 in B] a significant one at 2½% level?

Solution

We first order the 18 observations of samples A and B into a single sequence according to size. We order by a random device those observations which occur more than once. [An alternative method is to break the ties in all possible ways and note the resulting value of u. If all these values of u produce significant results or all values of u produce results which are not significant, then the ties present no difficulty.] We underline the observations from Sample A to preserve their identity when combined with the observations of sample B:

23, 27, 32, 33, 37, 38, 38, 40, 43, 43, 45, 47, 49, 51, 52, 55, 58, 61

This arrangement has 12 runs. The hypothesis being tested is that the two samples A and B have the same distribution (that is, the redesign has resulted in no improvement). A significant improvement would result in very few runs. This is because in the single ordered sequence, a definite improvement would mean that the majority of the measurements in sample B would fall in the right portion of the sequence and thus reduce the number of runs. This indicates the use of a one-tailed test. We can use Table I at the .025 level. We see that, for $n_1=8$ and $n_2=10$ and $u=12$, from this table that $12>5$ and therefore we do not reject the hypothesis. The hypothesis would be rejected (indicating a definite improvement) only if u were less than

or equal to 5.

Example (4)

Same as example (3) with samples A and B as follows:

(A) 12, 25, 43, 19, 24, 46, 18, 44

(B) 29, 11, 49, 56, 61, 36, 58, 53, 38, 72, 53, 64

Solution

The combined sequence ordered according to size is: 11, 12, 18, 19, 24, 25, 29, 36, 38, 43, 44, 46, 49, 53, 55, 56, 58, 61, 64, 72

Here $n_1=8$ and $n_2=12$, and $\sigma=5$. From Table I again, we note that $5<6$ and therefore we reject the hypothesis of no improvement.

c) Serial Correlation Test

A second test which is useful for testing the randomness of a sequence is the serial correlation test. If we are studying a sequence of observations which are truly random, we would not expect a relationship between two consecutive elements. The probability that the smallest and largest elements will be consecutive is the same as the probability that the two largest or two smallest—or, in fact, any two elements will be consecutive. Therefore, if we pair each element with its successor and ordinary techniques of correlation are used, we would expect that perfect randomness would result in a correlation of zero.

If our sequence consists of the n elements $X_1, X_2, X_3, \ldots X_i, \ldots X_n$, then the successor of X_1 is X_2, of X_2 is X_3, and X_i is X_{i+1} for $i=1,2,3,\ldots, n-1$ and we also define that the successor of X_n is X_1. The pairings of the two variables X and Y which we are to study are as follows:

$$X: X_1 \quad X_2 \quad \ldots \quad X_i \quad \ldots \quad X_n$$

$$Y: X_2 \quad X_3 \quad \ldots \quad X_{i+1} \quad \ldots \quad X_1$$

If the correlation coefficient is close to +1 or to -1 we probably would reject the hypothesis of randomness (since +1 represents perfect positive correlation and -1 represents perfect negative correlation) and if the correlation is zero or near zero, we probably would accept the hypothesis that the sequence is random since this would indicate no or almost no relation between X_i and X_{i+1}. It is possible to obtain a sampling distribution of the serial correlation coefficient by the laws of probability. Even though the formula for this correlation coefficient is:

$$\rho = \sum_{i=1}^{n} \frac{X_i X_{i+1} - n\bar{x}\bar{y}}{n S_x S_y}$$

it is necessary only to consider:

(D) $R = \sum_{i=1}^{n} X_i X_{i+1}$ since $\bar{X}$, $\bar{Y}$, S_x and S_y are unchanged under the various permutations. It has been shown that an approximation of the distribution of R is normal with a mean and standard deviation as follows:

(E) $\mu_R = \frac{S_1^2 - S_2}{n-1}$ and

(F) $\sigma_R = \sqrt{\frac{S_2^2 - S_4}{n-1} + \frac{S_1^4 - 4S_1^2 S_2 + 4S_1 S_3 + S_2^2 - 2S_4}{(n-1)(n-2)} - \mu_R^2}$

(G) Where $S_K = X_1^K + X_2^K + \ldots X_n^K$

(H) The formula $Z = \frac{R - \mu_R}{\sigma_R}$ gives us the number of standard deviations a particular serial correlation coefficient is from the mean and therefore will enable us to reject or accept the hypothesis at a given level of confidence. As n becomes larger the computations involved become laborious. However, modifications, as will be illustrated in

example (6) below, often will simplify these computations.

Example (5)

Using the serial correlation coefficient method, test the following sequence of 12 members for randomness:

5, 17, 18, 8, 19, 13, 9, 1, 3, 16, 11, 2 at the 5% level of significance.

Solution

The given sequence produces the following pairings:

x: 5 17 18 8 19 13 9 1 3 16 11 2

y: 17 18 8 19 13 9 1 3 16 11 2 5

We first find $R = \sum_{i=1}^{n} x_i \cdot x_{i+1} = (5)(17) + (17)(18) + (18)(8) + \ldots + (2)(5) = 1319$

Then, using formula (G) we find S_1, S_2, S_3, and S_4:

$S_1 = 5+17+18+\ldots+2=122$

$S_2 = (5)^2+(17)^2+(18)^2+\ldots+(2)^2=1704$

$S_3 = (5)^3+(17)^3+(18)^3+\ldots+(2)^3=26,630$

$S_4 = (5)^4+(17)^4+(18)^4+\ldots+(2)^4=438,936$

Now (E) gives $\mu_R = \frac{(122)^2-1704}{11}=1198$ and (F) produces:

$$\sigma_R = \sqrt{\frac{(2)(464)(680)}{11} + \frac{(135)(105)(296)}{(11)(10)} - (1198)^2}$$

$$\sigma_R = \sqrt{224,062 + 1,228,230 - 1,435,204}$$

$$\sigma_R = \sqrt{17,088} = 131$$

and finally:

$$Z = \frac{R-\mu_R}{\sigma_R} = \frac{1319-1198}{131} = .93$$

At the 5% level of significance, a $Z \geq 1.65$ is required to reject

the hypothesis and therefore the hypothesis of randomness is accepted.

Example (6)

Again, using the serial correlation technique, we wish to test, at the 5% level of significance, the following sequence for randomness: 6.74, 6.72, 6.77, 6.79, 6.75, 6.80, 6.82, 6.88, 6.92, 6.89

Solution

Since the result of the test is not influenced by adding the same constant to each term of the sequence or by multiplying each term of the sequence by the same constant, we add -6.80 and then multiply each term of the sequence by 100 to produce to following simpler sequence: -6, -8, -3, -1, -5, 0, +2, +8, +12, +9. The solution now follows as in the preceding example

X: -6 -8 -3 -1 -5 +0 +2 +8 +12 +9

Y: +9 -6 -8 -3 -1 -5 0 +2 +8 +12

$$R = (-6)(+9)+(-8)(-6)+(-3)(-8)+\ldots.+(+9)(12) = 246$$

$$S_1 = -6 -8 -3 -1+\ldots..+9 = 8$$

$$S_2 = (-6)^2 + (-8)^2 + (-3)^2 + \ldots + (9)^2 = 428$$

$$S_3 = (-6)^3 + (-8)^3 + (-3)^3 + \ldots + (9)^3 = 2096$$

$$S_4 = (-6)^4 + (-8)^4 + (-4)^4 + \ldots + (9)^4 = 37,508$$

$$\mu_R = \frac{64-428}{9} = -40.4$$

$$\sigma_R = \sqrt{\frac{145,675}{9} + \frac{69,768}{72} - (40.4)^2} =$$

$$\sigma_R = \sqrt{16186 + 969 - 1,632} =$$

$$\sigma_R = \sqrt{15,523} = 125$$

again, finally

$$Z = \frac{246+40.4}{125} = \frac{286.4}{125} = 2.3$$

Therefore, since the probability of obtaining an R of this size or greater is approximately .011, we reject the hypothesis of randomness.

When we correlate each member with its successor as in the above problems, this is known as serial correlation with lag 1. The test is applied in the usual manner if any other lag k is used.

Example (7)

Solve problem (5) using the runs technique

Solution

In this problem the median is 10. Using a + sign to indicate an element above the median and - sign to indicate an element below, the sequence becomes:

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

and U, the number of runs, is 7. Using Table I with $n_1 = n_2 = 6$, we see that $3 < 7 < 10$ and therefore we accept the hypothesis of randomness at 5% level.

Example (8)

Solve problem (6) using the runs technique.

Solution

It is obvious that here the first 5 terms are above the median and the last 5 terms are below the median. Thus the sequence becomes:

-----+++++ and $U = 2$

From Table I, we see that $U = 2$ does not fall between 2 and 9. Therefore we reject the hypothesis of randomness.

It is clear from the last two examples that the computations involved in the serial correlation test are much more involved than those for the total runs test. For large n, the condition of normality is reasonable, and then the serial correlation test may be more reliable.

2. Some Statistical Decision Theory

a) Definitions and Basic Principles.

A statistician, a technician, or any experimenter who makes observations and measurements is finally faced with the problem of making a decision on the basis of data available (usually incomplete). He may have to decide between the acceptance or rejection of a given hypothesis. In sequential testing, there are three possible decisions he might make, namely: accept, reject, or continue testing. He may wish to estimate the mean life to failure of a piece of equipment on the basis of the results of an experiment. Or in other words, he must make a decision as to which element X_o of a class X should be adopted as the estimate of the true (but unknown) value of the parameter. These three familiar statistical problems can be considered as special cases of the general decision problem.

Principles must be applied to aid in making the decision. In any particular problem, the principle which will guide the decision making will depend on the nature of the problem and often will involve numerical constants which can only be estimated. A. Wald (Ref. 14) developed the minimax principle, that is the principle of selecting a rule which minimizes the maximum risk which could occur. This means that the investigator anticipates the worst and acts accordingly. This is frequently a wise action but has been criticized as being too pessimistic, and several useful modifications of the minimax theory have been presented. A second method originated by Bayes and used by LaPlace is to assume that the a priori probabilities are all equal, unless evidence to the contrary is available. According to the Bayes principle, we choose a decision function (defined later) so as to maximize the average gain. A further discussion of the various principles

of choice will be found in section d.

Statistical decision theory and the theory of games are closely related. Many statistical problems can be regarded as a two person game in which the statistician plays against Nature. As an aid in making a decision, the statistician performs an experiment whose outcome is $S_i (i=1,2,\ldots n)$ The set of all such possible outcomes will be called S, the sample space. The set of all possible actions (or strategies) we call A. A may consist of only two elements a_1, and a_2 as in the case where the decision is to accept or reject a hypothesis. A may contain three elements a_1, a_2, and a_3 as in sequential testing when the decision is to accept, reject or continue testing. Also A may contain an infinite number of elements as in the case of point or interval estimation. A statistical decision function, $d(s)=a$ is a rule associating a possible action, a, with each possible outcome S. For example, we wish to estimate the unknown mean life μ of a population using a random sample of size n from the population. We can consider that a possible action is a statement which says that the mean of the population is the mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} Xi$ where one outcome is $S = (X_1\ X_2\ X_3 \ldots X_n)$. Here $d(S) = \bar{X}$. A second example would be the test of a hypothesis $H_o: \mu = 40$ which requires the rejection of the hypothesis if $\bar{X} \leq 45$ and the acceptance of the hypothesis if $X > 45$. Here we have two elements in A namely, a_1, the rejection of the hypothesis and a_2 the acceptance of the hypothesis. (This is an arbitrary rule whichcould be modified readily to include a 3rd action, continue testing.) Here $d(s) = a_1$ if $s = (X_1, X_2, \ldots, X_n)$ and $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \leq 45$, and $d(s)=a_2$ if $S = (X_1, X_2, \ldots, X_n)$ and $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i > 45$

Considering the above two examples as a game, statistician versus nature, the

decisions arrived at are often called pure decisions since the strategies or actions taken are clear-cut, that is for a given outcome a definite action is prescribed. In other games, the decision may depend on some randomising device to select an action (decision) from an existing probability distribution over a set of possible decisions. Such decisions are called mixed decisions. In the game of Parchesi, the decisions are of the mixed variety since the decision as to which marker to move depends on the roll of a pair of dice.

b) Utility

The major problem of the statistician is how to choose between alternate decision functions. Associated with each end result facing the decision maker is a numerically measured value called utility and in general he should make those decisions for which his end results would have as large a utility as possible. This criterion may not necessarily always be measured in dollars but other valuable considerations such as time, reputation and even life itself may be applied.

Let us suppose the existence of a parameter space Ω consisting of elements ω_i : $i=1,2,3,\ldots n$ For each ω_i there is a corresponding probability distribution on S, the sample space. For a fixed ω, a certain action a $[a=d(s)]$ will yield an end result $e=e(a;\omega)$. Now let E denote the set of all possible end results e. Then ω and d together determine a probability distribution on E. If the experimenter can determine which of two probability distributions he prefers, then a utility can be assigned to each of these. Therefore each ω and d will have as utility, the utility of the probability distribution which they determine. A second way of assigning a utility to a pair, (d,ω) is possible if a precise numerical loss L resulting from a given action

can be determined. This expected loss is often called the risk, regret or inutility.

The preceding utility theory is obviously very general in character and it is often extremely difficult to determine a utility in these ways.

A simple illustration of how numbers may be assigned to basic alternatives is given by Savage. He proposes the following problem. On a given occasion a person faces the decision of whether or not to carry an umbrella. For simplicity, only two possible states of nature, future rain and future shine are considered. The possible consequences are given by this table:

State / Act	Rain	Shine
Carry	inconvenience and wet feet	inconvenience and slight embarrassment
Don't Carry	Miserable drenching	bliss unalloyed

In other situations, the possible consequences of each act, in each state, could be measured in dollars and in the preceding rain-shine example, the following "income matrix" might be reasonable:

State / Act	Rain	Shine
Carry	4	5
Don't Carry	-10	10

c) The Pay-off Matrix

The two person game mentioned in the preceding section can be considered as the product of space X, all possible strategies of player I, and of space Y, all strategies of player II. We define R(x,y) as a function which gives the outcome r for each pair (x,y), x belonging to X and y belonging to Y. In the theory of utility it has been shown that with each r of our outcome space R there may be associated an M (x,y) called the pay-off function which may be regarded as the monetary gain or "utility" to player II (the decision maker). In the theory of games there is a special class of games called two-person zero sum games. This is a game in which the total payoff to the two players is Zero so that if the pay-off to player II is M(x,y) then the payoff to player I is -M(x,y).

Our game can be considered as a matrix (Mij) in which the columns are the strategies of the first player and the rows the strategies of the second player:

$$
\begin{array}{cc}
 & \text{I} \\
\text{II} & \begin{array}{c|cccc}
 & X_1 & X_2 & X_3 & \cdots X_n \\
y_1 & M_{11} & M_{21} & M_{31} & \cdots M_{m1} \\
y_2 & M_{12} & M_{22} & M_{32} & \cdots M_{m2} \\
y_3 & M_{13} & M_{23} & M_{33} & \cdots M_{m3} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
y_n & M_{1n} & M_{2n} & M_{3n} & M_{mn}
\end{array}
\end{array}
$$

Independently, the first player selects the ith column and at the same time the second player selects the jth row. The pay-off to the second player is the entry M_{ij} of the matrix. The final choice of a decision function must be based on some reasonable principle of choice. Therefore, before illustrating the pay-off matrix, principles of choice will be discussed somewhat further in the next section.

d) Principles of Choice.

In section (a) reference was made to several principles which could aid the statistician in making a decision. Suppose for a given problem we have been successful in assigning the utility $U(d;\omega)$ to each decision function d for each ω in Ω. Then, as previously mentioned, the most conservative principle of choice is the minimax criterion. Here we choose the decision function so that the min $U(d;\omega)$ is the greatest. We will illustrate this principle by devising a pay-off matrix as explained in the preceding section.

	X_1	X_2	X_3	X_4
y_1	5	0	4	5
y_2	0	3	6	3
y_3	2	3	3	2
y_4	4	1	4	3

The minimax principle states that we should choose the row for which the least possible gain is a maximum. In the four rows in the pay-off matrix above, the minimum gains are 0,0,2,1. Therefore we choose strategy y_3 since regardless of what happens our minimum gain will be 2 units.

The second method mentioned in section (a) was the Bayes principle or as it is often called "the principle of insufficient reason." This criterion asserts that if one is "completely ignorant" of the status of nature, we assume that the probabilities of all are equal and we choose a strategy which maximises the average gain. In the pay-off matrix above, the average gain (utility) for the four rows are $\frac{14}{4}$, $\frac{12}{4}$, $\frac{10}{4}$ and $\frac{12}{4}$. Therefore we choose y_1.

The minimax principle of Wald as previously noted is unduly pessimistic since concentration is made on the worst possible consequence of each act. A number of modifications have been devised to remove this objection and Hurwiz in 1951 suggested the selection of a number α $(0 \leq \alpha \leq 1)$ which he called the pessimism-optimism index. For each act (or row) of our matrix let M_i be the maximum utility number and m_i the smallest utility number. We now choose the row for which $\alpha M_i + (1-\alpha) m_i$ is a maximum. When $\alpha = 0$, this is equivalent to Wald's minimax principle. In the previous example the four rows give 5α, 6α, $3\alpha+2(1-\alpha) = \alpha+2$, $4\alpha+1(1-\alpha) = 3\alpha+1$. For $\alpha = \frac{1}{4}$ the 4 rows would give the values $\frac{5}{4}$, $\frac{6}{4}$, $\frac{9}{4}$, $\frac{7}{4}$ and y_3 would be the strategy chosen. For $\alpha = \frac{1}{2}$, the results become $\frac{5}{2}$, $\frac{6}{2}$, $\frac{5}{2}$ and $\frac{5}{2}$ so that y_2 is preferred. And finally, for $\alpha = 3/4$, we have $\frac{15}{4}$, $\frac{18}{4}$, $\frac{11}{4}$ and $\frac{13}{4}$, so that y_2 is again preferred. It seems clear that for any $\alpha > \frac{2}{5}$ y_2 is preferred.

Several other modifications involve the use of inutility, risk or regret. Savage suggested that we minimise the maximum regret where the regret is defined as the amount that must be added to the utility to equal the maximum utility pay-off. Or it can be also described as the difference between the pay-off for his actual strategy and the pay-off he

would receive if he knew in advance the state of nature (Player I strategy). Symbolically $r_{ij} = \max_j M_{ij} - M_{ij}$. Again, using the previous matrix as an example, the regret matrix associated with it is:

	X_1	X_2	X_3	X_4
y_1	0	3	2	0
y_2	5	0	0	2
y_3	3	0	3	3
y_4	1	2	2	2

Here we have maximum regrets for the four rows 3, 5, 3, and 2. Therefore we choose the act which minimizes the maximum risk (regret) which is y_4.

The above example illustrates that the use of four different principles of choice lead to four different decisions regarding the choice of strategy. The minimax principle indicates the use of strategy y_3; the Bayes principle leads to the choice of y_1; the modification developed by Hurwicz suggests the choice of y_2 as the preferred strategy; and finally the regret principle indicates the use of strategy y_4.

e) Bayes Estimates, and Sequential Decision Problems

In order to estimate a parameter ω, the statistician performs an experiment with outcomes S_1 where S is, as previously stated, the set of all possible outcomes under consideration. An event S_a a subset of S will have a probability depending on ω:

$$P_\omega(S_a) = \sum_{s \in S_a} P\omega(s)$$

Then the decision function d(S)=a enables us to select an action a from the set A of all possible actions. If we now assume that there exists a numerical loss function L(d, w), then the expected value of this loss will be defined by:

$$M(d,\omega) = \sum_{S} L(d,\omega)\, P\omega(S)$$

where d=d(S) and the statistician obviously tries to minimize this expected loss. An estimate which does minimize the expected loss is called a Bayes estimate.

In the case of the binomial distribution where:

θ = the probability of success based on the observation that the n trials, and

$P(x) = {}_nC_x \cdot \theta^x(1-\theta)^{n-x}$

$\lambda(\theta)\, d\theta$ = probability that θ lies in the interval θ to $\theta+d\theta$

$f_n(x)$ = our estimate of the parameter θ

$KE\left\{f_n(x)-\theta\right\}^2 = K\sum_{x=0}^{n}\left\{f_n(x)-\theta\right\}^2 P_n(x)$ [K a constant]

L = expectation of loss

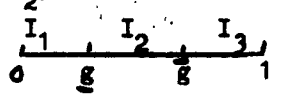
it is shown (Ref 1) that:

M) $$L = \int_0^1 KE\left\{f_n(x)-\theta\right\}^2 \lambda(\theta)\, d\theta .$$

The θ that minimises L in equation M is the Bayes estimate.

A well known testing procedure is Wald's sequential probability ratio test. After each observation the statistician must make a decision whether to continue to make observations or to take final action (usually accepting or rejecting a hypothesis). Let us suppose that we have to decide between two hypotheses H_1 and H_2 of which the a priori probabilities are g and 1-g respectively. We assume that there is no loss in accepting the true hypothesis, and a loss ℓ_{12} if we accept H_2 when H_1 is true, and a loss of ℓ_{21} if we accept H_1 when H_2 is true. If g is small we of course

accept H_2 since the risk $R_1 = \ell_{12} g$ is small. Also if g is large, 1-g is small, and we similarly accept H_1 since the risk $R_2 = \ell_{21}(1-g)$ is small. In order to determine the size of g for acceptance of H_2, the whole interval $0 \leq g \leq 1$ is divided into three subintervals



where $I_1 (0 \leq g \leq \underline{g})$, $I_2 (\underline{g} < g < \bar{g})$ and $I_3 (\bar{g} \leq g \leq 1)$. Then if g belongs to I_2 we continue to make observations, but if g belongs to I_1 or I_3 we stop and take action: accept H_2 if in I_1 and accept H_1 if in I_3. If we take n observations, the a posteriori probability of H_1 is given by Bayes Theorem (Ref. 5):

$$g_n = \frac{g P_{1n}}{g P_{1n} + (1-g) P_{2n}}$$

where P_{1n} is the probability of having n observations under H_1 and P_{2n} the probability of n observations under H_2. On pp. 226-230 in Vol. 17, Econometrica, general formulas are developed for $\bar{g}$ and $\underline{g}$ and also for a number of special frequency distributions. In all cases, however, it is very difficult to compute the operating characteristics involved in the equations for $\bar{g}$ and $\underline{g}$. In particular, the losses ℓ_{12} and ℓ_{21} can only be guessed at. In the case where the hypotheses H_1 and H_2 do not differ much from each other the approximations developed by Wald give fairly close results.

f) Critical Comments on Utility Theory and Principles of Choice.

Utility theory (as stated before) is at the present still very general in character. The constructing of the utility function $U(d;\omega)$ is extremely difficult and highly subjective. However, even though some statisticians assert that the theory is doomed from a practical standpoint, our feeling is that it has given many new insights to statistical decision theory and offers possibilities for great usefulness in the future. The

first obvious difficulty with utility theory is that numerical values can very seldom be assigned to the consequences of decisions. This usually rests upon an infinity of paired comparisons. Some statisticians have tried to avoid this difficulty by making a finite or a relatively few paired comparisons. Verification then is based on these as well as on the assumption that the utility function exists and is linear.

A number of authors have developed axiom systems in order to have both a plausible set of consistent requirements based on idealized human preferences as well as a procedure for proving the validity of the utility assignments. A second difficulty arises as a result of these axiom systems since the preferences or paired comparisons almost never satisfy the stated axioms.

The principles of choice discussed earlier were Walds minimax principle, Bayes principle, and modifications by Hurwicz and Savage. Serious criticisms, objections, and drawbacks have been cited by many statisticians to all of these. In criticism of Walds minimax principle we quote from Savage (Ref. 12): "There is a general principle for finding minimax actions when the number of states and mixed actions is finite, but it leads in general to very extensive computations and is not applicable at all when either of these numbers is infinite. Devices for solving special classes of minimax problems are therefore much sought after, and even now many of the most commonplace situations of statistics lead to difficult minimax problems. Few, if any, new minimax solutions of immediate practical importance have yet been found." A second serious objection to Walds minimax principle (as stated before) is based on extreme pessimism, that is that we assume nature to be in the worst

possible state.

The Savage modification, minimising the maximum regret, has been criticized by Chernoff (Ref. 5) as follows:

"Unfortunately, the minimax regret (risk) criterion has several drawbacks. First, it has never been clearly demonstrated that differences in utility do in fact measure what one may call regret (risk). In other words, it is not clear that the "regret" of going from a state of utility 5 to a state of utility 3 is equivalent in some sense to that of going from a state of utility 11 to one of utility 9. Secondly, one may construct examples where an arbitrarily small advantage in one state of nature outweighs a considerable advantage in another state."

Objections to the principle of insufficient reason have been many. The first criticism is that the principle is extremely vague and may lead to contradictory results. A second objection is that the principle is not strictly applicable for a person who has had any experience with the problem at hand since utility theory depends on a series of preferences. A third criticism, much like the first, is that this criterion is highly subjective. For example, if we are faced with a real problem in decision making, we must first list the mutually exclusive states of nature. The objection here is that many such lists are possible, and therefore will in general give different results.

Finally, the fourth and last principle of choice discussed was the Hurwicz modification of Wald minimax principle in an effort to make it less pessimistic. The major objection to this modification is that it involves the selection of an α (considered as measuring the optimism of the statistician [player II] in a game against nature [player I] which

is highly subjective and vague. A criticism which applies not only to the Hurwicz α modification but also to the other principles of choice discussed is that they depend on some notion of complete ignorance. In reality, however, the statistician usually is not in complete ignorance of the true state of nature but has some idea of the various possibilities.

CONCLUSION

Since most methods used in statistics were derived upon the basis of random sampling, it is essential that this condition be satisfied in order that the results obtained by valid. In particular, when the tests involve sequences of observations ordered with respect to time, some test for randomness should be applied before methods based on randomness are used. The Runs Test, discussed and illustrated in section 1b, is one of the easiest to apply. The computations are simple, the test is nonparametric (knowledge of underlying frequency function not necessary), and can be used for both small (see Table I) and large samples. The Serial Correlation Test, 1c, involves laborious computations for large n but for smaller n the arithmetic is not too time consuming. It should be pointed out that in the case of nonparametric tests there is no established theory to determine what constitutes a "best" test.

The construction of utility functions and applications of decision theory is difficult and subjective. However, we conclude that:

(1) A subjective risk function or decision function may well be an optimum one if the individual designing it has knowledge of what is important and what is not important to the purpose of the test. In the particular instance of a risk function applied to a reliability test, the Air Force (consumer) should have the prerogative of choosing which parameters of a risk function are most important and of designing the function to suit these.

(2) While decision functions in general are difficult to manipulate due to the lack of a common denominator for different consequences, in the particular case of a decision function for reliability

test purposes it may be found to be less difficult. This is due to the fact that the majority of actions involved may be broken down to but two prime factors, time and money, and the definition of the relationship and relative importance of those should be the responsibility of the consumer.

(3) It is conceivable that a decision function tailored to reliability tests will be an eventuality. However, before this eventuality may be realized, more knowledge of decision functions is necessary. At this time little is known about possible applications of decision functions, or their practicality of use. This area first came into prominence about 15 years ago, and significant work was accomplished. However, since that time little more has been accomplished in this area than the study of its more obvious fundamentals.

BIBLIOGRAPHY

1. Arrow, K., Blackwell, D., and Girshik, M.A., "Bayes and Minimax Solutions of Sequential Decision Problems," Econometrica, 17 (1949), pp. 213-244.

2. Blackwell, D., and Girshik, M.A., Theory of Games and Statistical Decisions, New York, John Wiley & Sons, Inc., 1954.

3. Brownlee, K.A., Statistical Theory and Methodology in Science and Engineering, New York, John Wiley & Sons, Inc., 1960.

4. Brunk, H.D., An Introduction to Mathematical Statistics, New York, Ginn & Co., 1960.

5. Chernoff, H., and Moses L.E., Elementary Decision Theory, New York, John Wiley & Sons, Inc., 1959.

6. Dixon, W.J., and Massey, F.J., Introduction to Statistical Analysis, New York, McGraw-Hill, 1951.

7. Hoel, P.G., Introduction to Mathematical Statistics, New York, John Wiley & Sons, Inc., 1954.

8. Keeping, E.S., "Statistical Decisions," American Mathematics Monthly, 63 (1956), pp. 147-159.

9. Luce, R.D., and Raifa, H., Games and Decisions, New York, John Wiley & Sons, Inc., 1958.

10. Parzen, E., Modern Probability Theory and Its Applications, New York, John Wiley & Sons, Inc., 1960.

11. Savage, L.J., The Foundations of Statistics, New York, John Wiley & Sons, Inc., 1954.

12. Savage, L.J., "The Theory of Statistical Decision," Journal of American Statistical Association, 46 (1951), pp. 55-67.

13. Siegal, S., Nonparametric Statistics for the Behavioral Sciences, New York, McGraw-Hill, 1956.

14. Wald, A., Statistical Decision Theory, New York, John Wiley & Sons, Inc., 1950.