https://ntrs.nasa.gov/search.jsp?R=19660005535 2020-03-16T21:52:05+00:00Z

General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

AD614271

SEQUENTIAL RANKING PROCEDURES

BY

ELIAS ALPHONSE PARENT, JR.

TECHNICAL REPORT NO. 80

April 23, 1965

DDC
APR 29 1965
TISIA B

SUPPORTED BY THE ARMY, NAVY, AIR FORCE AND NASA UNDER
CONTRACT Nonr-225(53) (NR-042-002)
WITH THE OFFICE OF NAVAL RESEARCH

COPY 2 OF 3
HARD COPY $. 3.00
MICROFICHE $. 0.75
77p

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

ARCHIVE COPY

SEQUENTIAL RANKING PROCEDURES

by

Elias Alphonse Parent, Jr.

TECHNICAL REPORT NO. 80

April 23, 1965

Supported by the Army, Navy, Air Force and NASA under

Contract Nonr-225(53) (NR-042-002)

with the Office of Naval Research

Gerald J. Lieberman, Project Director

Reproduction in Whole or in Part is Permitted for

any Purpose of the United States Government

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

BLANK PAGE

SEQUENTIAL RANKING PROCEDURES

By

Elias Alphonse Parent, Jr.

1. Introduction. Many statistical procedures used and studied today are sequential in nature. By this we mean that the time when a statistical decision is reached is random. In contrast to such procedures are the fixed sample size procedures. Best known perhaps is sequential analysis and the sequential probability ratio test as formulated by Wald [6]. There are other sequential procedures, for example in process inspection schemes, where, based on a sequence of observations a decision is made to stop the process and take some adjusting action, the time at which the process is stopped being a random variable. There are many other sequential-like procedures.

In the theory of hypothesis testing for the case of a simple hypothesis against a simple alternative it is known that a most powerful test can be determined by the Neyman-Pearson lemma, which is of the form:

$$\text{reject } f = f_0 \text{ if } \Lambda_n = \frac{f_1(X_1, X_2, \cdots, X_n)}{f_0(X_1, X_2, \cdots, X_n)} > K$$

where the hypotheses to be tested are $f = f_0$ against $f = f_1$, f_0 and f_1 are the joint densities of the observations $X_1, X_2, \cdots, X_n$, corresponding to each hypothesis. This is an example of a nonsequential procedure. To extend such a procedure to the sequential idea we need only modify the test as follows:

take a sample of size of size m and

reject f_0 if $\Lambda_m \geq K_1$

accept f_0 if $\Lambda_m \leq K_2$

draw another sample of size $n-m$ if $K_2 < \Lambda_m < K_1$

if the second sample is required compute Λ_n and

$$\text{reject } f_0 \text{ if } \Lambda_n > K$$
$$\text{accept } f_0 \text{ if } \Lambda_n \leq K .$$

Such a simple modification gives us a two stage procedure with a new feature in that the total sample size is random, being either m or n, depending upon the outcome of the first stage. This basic idea of a sequential test was proposed by Dodge and Romig in [8], and has been extended to multiple stage sampling plans.

Sequential hypothesis testing as proposed by Wald requires that a computation of Λ_n and a decision be made as each observation is taken. Briefly, to test $f = f_0$ against $f = f_1$ select constants $B < A$ and compute Λ_n as each observation is taken, and proceed according to the rule

$$\text{if } \Lambda_n \geq A \text{ reject } f = f_0$$
$$\text{if } \Lambda_n \leq B \text{ reject } f = f_1$$
$$\text{if } B < \Lambda_n < A \text{ take another observation and compute } \Lambda_{n+1}$$

Since the sequential probability ratio test is formulated in terms of the ratio which leads to most powerful tests according to the Neyman-Pearson theory we would expect it to have good properties. This indeed is the case in that of all tests with the same power the sequential probability ratio test requires on the average fewest observations. This optimal property was conjectured by Wald and finally proved by Wald and Wolfowitz in [9].

In order to carry out these sequential tests of hypotheses we note that an assumption as to the specific form of f_0 and f_1 must be made. It often happens that the form of the underlying distribution is not assumed known and in this case nonparametric statistical methods are used. In nonparametric statistics many tests of statistical hypotheses are based on the set of ranks $\{T_1, T_2, \dots, T_n\}$ determined from a random sample $\{X_1, X_2, \dots, X_n\}$, or the signs of the observations (± 1 according as X_i in positive or negative) or on a combination of both of these sets of statistics derived from the basic observations. The sign test, signed rank test, Wilcoxon-Mann-Whitney test, Fisher-Yates test and many others are examples of such fixed sample size nonparametric tests.

Contrary to the case in parametric statistics (as opposed to nonparametric statistics) there are very few sequential procedures in nonparametric statistics, particularly sequential procedures based on signs, ranks, or both. One reason for this is that for most specified alternatives to the null hypothesis it is difficult to compute probabilities for statistics based on signs and ranks which in turn makes it difficult to properly evaluate the properties and operating characteristics of the procedures. This difficulty can be circumvented by restricting attention to special classes of alternatives such as those proposed by Lehmann in [1], where to the null hypothesis $F(x)$ he proposed alternatives of the form $F^a(x)$, $a > 0$. This of course does not solve the basic problem of alternatives as the question of whether or not the Lehmann alternative is appropriate for the problem being considered arises. However it is a first step inasmuch as it does allow us to develop some sequential procedures where exact distribution theory calculations are possible. In the fixed sample size problem it simplifies considerations of power of rank tests.

An example of a nonparametric sequential test is the following adaptation of Wald's sequential probability ratio test for binomial observations. Consider a sequence of independent identically distributed random variables $X_1, X_2, \ldots$ with cumulative distribution function $F(t) = P(X_1 \leq t)$. We wish to test $F(t_0) = p_0$ against $F(t_0) = p_1$ for some fixed value t_0. The number of observations less than or equal to t_0, say N, after taking n observations, is a binomial random variable with parameters $F(t_0)$ and n. The probability ratio reduces to

$$(1.1) \qquad \Lambda_n = \frac{P(N|\ F(t_0) = p_1)}{P(N|\ F(t_0) = p_0)} = \left(\frac{p_1}{p_0}\,\frac{1-p_0}{1-p_1}\right)^N \left(\frac{1-p_1}{1-p_0}\right)^n$$

and the sequential test based on this ratio is discussed in Wald [6]. For the special case where $t_0 = 0$, N is equivalent to the number of negative observations after n trials and this would be a sequential test based on the signs of the observations.

An example of a nonparametric sequential procedure based on ranks of observations is the grouped rank test developed by Wilcoxon, Rhodes and Bradley [4]. Actually two sequential procedures are developed in [4], the Configural Rank Test and the Rank Sum Test. Basically, observations are taken in groups of m X's and n Y's and the observations are ranked within each group. For each group a statistic is computed based on the ranks and Wald's sequential probability ratio test is applied to the sequence of statistics so generated. Each group of m X's and n Y's becomes the basic unit used in the probability ratio. Suppose the X- population has distribution $F(x)$ and the Y- population has distribution $G(y)$, and observations are taken as follows

$$(X_{11}, X_{12}, \ldots, X_{1m}, Y_{11}, Y_{12}, \ldots Y_{1n}) \text{ - group 1}$$

$$(X_{21}, X_{22}, \ldots, X_{2m}, Y_{21}, Y_{22}, \ldots Y_{2n}) \text{ - group 2}$$

$$\vdots$$

$$(X_{\gamma 1}, X_{\gamma 2}, \ldots, X_{\gamma m}, Y_{\gamma 1}, Y_{\gamma 2}, \ldots, Y_{\gamma n}) \text{ - group } \gamma$$

Let $R_\gamma = (R_{\gamma 1}, R_{\gamma 2}, \ldots, R_{\gamma m}, S_{\gamma 1}, S_{\gamma 2}, \ldots, S_{\gamma n})$ be the rank vector associated with group γ where $R_{\gamma i}$ is the rank of $X_{\gamma i}$ and $S_{\gamma i}$ is the rank of $Y_{\gamma i}$, the ranks taken from the combined ranking of the X's and Y's. Taking a function of R_γ, say $T_\gamma = T(R_\gamma)$, we generate a new sequence of random variables $T_1, T_2, \ldots$ and the Wald sequential probability ratio test may now be applied to the T_i. For independent group to group sampling we have

$$\Lambda_n = \prod_{\gamma=1}^{n} \frac{P(T_\gamma \mid Y \sim G(y))}{P(T_\gamma \mid Y \sim F(y))} \tag{1.2}$$

as the probability ratio to test the hypothesis that the Y- population has distribution $F(y)$ against $G(y)$. In [4] the authors consider Lehmann alternatives $G(y) = F^k(y)$, $k > 0$ and the function T in one case is the actual configuration of X's and Y's, which is equivalent to the vector $(S_{\gamma 1}, S_{\gamma 2}, \ldots, S_{\gamma n})$, and in the second case T is taken to be the sum of the Y ranks.

Wilcoxon, Rhodes and Bradley observe that the test could be improved by taking observations in pairs and reranking from the beginning each time a new observation pair is taken. One reason for the reduced efficiency of the group ranking method is that the observations in one group are not compared with observations from any other group. The reranking suggestion would take into account all comparisons. However, this is very cumbersome, and moreover reranking introduces non-independence of successive probability ratios making an analysis of the properties of such a procedure difficult.

Thus in order to attack the problem of nonparametric sequential tests of hypotheses based on ranks we should consider procedures such that the distribution theory is tractable and such that ranks are assigned in a truly sequential manner, avoiding as much as possible the complexities introduced by reranking. To this end two new sequential ranking methods will be defined in this dissertation.

In order to be led somewhat naturally to these new ranking methods we now consider the reranking procedure in more detail. Let T_{ij} be the rank of X_j at the i^{th} stage in the reranking process. We observe $X_1, X_2, \ldots, X_n, \ldots$ and each time a new observation is taken the entire set of observations is reranked. We have

Observation vectors	Rank vectors
(X_1)	(T_{11})
(X_1, X_2)	(T_{21}, T_{22})
(X_1, X_2, X_3)	(T_{31}, T_{32}, T_{33})
$\vdots$	$\vdots$
$(X_1, X_2, \dots, X_n)$	$(T_{n1}, T_{n2}, \dots T_{nn})$

Notice that the vector $(T_{11}, T_{22}, \dots, T_{nn})$ completely determines the n rank vectors listed above in the sense that each vector could be reconstructed given only T_{ii} $i = 1, 2, \dots, n$. T_{ii} is the rank of X_i relative to the set $\{X_1, X_2, \dots, X_i\}$. Thus we can rank an observation as it is observed, relative to the preceeding observations without reranking the previous observations and still retain the information contained in the n rank vectors which would come from reranking. This method of ranking observations is one way of assigning ranks which fits in naturally with the idea of sequential procedures and lends itself to developing sequential procedures in non-parametric problems. This ranking procedure also takes into account all comparisons among the observations.

Analogous to the fixed sample size signed rank test we will define a second sequential ranking procedure based upon the absolute values of the observations and taking into account the signs of the observations. This signed sequential ranking procedure will be applied to a problem in process control. By process control we mean a procedure where the aim is to determine when a given sequence of random variables changes from being distributed according to a distribution $F(x)$ to a different distribution $G(x)$. The term process control enjoys a broader definition today including those cases where the process is adjusted according to some statistic based upon the sequence of observations. Such procedures are referred to as adaptive control methods.

The early methods used to control a process were based on control charts (Shewhart charts) and modifications of these control charts. To control the mean value of some dimension of a process at a particular value μ_o, samples of size n are taken at frequent intervals of time and the sample mean $\bar{X}$ is compared with $\mu_o \pm k\sigma/\sqrt{n}$. If $\bar{X}$ falls outside these lines the process is stopped and adjustments to the process are carried out, and for $\mu_o - k\sigma/\sqrt{n} \leq \bar{X} \leq \mu_o + k\sigma/\sqrt{n}$ the process is allowed to continue without adjustment. Modifications to the basic control chart method came in the form of "warning lines" inside the action lines $\mu_o \pm k\sigma/\sqrt{n}$. Further modifications were introduced which changed the action rule to rules of the type "If K consecutive points on the chart fall outside control lines, take action." These early procedures failed to take advantage of all the information contained in the sequence $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_n$. At best the modified action rules used only the information contained in a fixed number of sample values in the immediate past.

In order to take advantage of this unused information the stopping rule should incorporate the entire sample. A step in this direction was taken by Page in [7] with the introduction of cumulative sum schemes. If the mean of a process is to be controlled the cumulative sums $S_n = \sum_{i=1}^{n}(X_i - k)$ are plotted on a chart against n. The entire history of the process is presented and changes in the process mean are visible through changes in direction of the mean path. To detect one-sided deviations in the mean, say increases, the stopping rule used is to stop the process when the current point of the path (n, S_n) rises a given amount $h > 0$ above the previous lowest point of the path. Two-sided deviations are treated by applying two one-sided schemes simultaneously. For normal observations the cumulative sum schemes have been found to be more sensitive than the Shewhart control chart.

When no assumption is made as to the form of the underlying distributions we might look to non parametric methods for a control procedure. For example, the sequential rank of X_i is equally likely to be $1, 2, \ldots, i$ as long as no change takes place in the distribution of $X_1, X_2, \ldots, X_i$. But when a location change takes place, say an increase in the process mean, larger ranks would be more probable. We will consider the sequential rank of $|X_i|$ relative to $|X_1|, |X_2|, \ldots |X_i|$, multiplied by the sign of X_i ($+1$ if $X_i \geq 0$ and -1 if $X_i < 0$) in a process control problem. This method of sequentially assigning ranks, as noted before, will be called signed sequential ranking.

This dissertation defines two methods of assigning ranks in a sequential manner to observations $X_1, X_2 \ldots$. Basic properties of the sequential ranks are studied and distribution theory is determined. Section 2 contains some preliminary results including some relating to order statistics of observations taken from non identical distributions. These results are used in the later sections. In Section 3 the method of sequential ranking is defined and it is shown that for a fixed sample size, ordinary ranks and sequential ranks are equivalent for the purpose of hypothesis testing. Section 4 is an application to sequential hypothesis testing for the two sample problem where the alternative is of the form proposed by Lehmann in [1]. The signed sequential ranking scheme is defined in Section 5 and a condition on the distribution of the sequence of observations is given which implies that the signed sequential ranks are independent. Distribution theory is given for the signed sequential ranks. Section 6 contains an application of signed sequential ranking to a process control problem.

2. Preliminary results. Let $X_1, X_2, \ldots, X_n$ be any random variables with continuous comulative distribution functions $F_1, F_2, \ldots, F_n$. Define X_{nk} to be the k^{th} smallest in the set $\{X_1, X_2, \ldots, X_n\}$. We can obtain a general expression for the distribution of X_{nk} as follows:

$$(2.1) \qquad F_{nk}(x) = P(X_{nk} \leq x)$$

$$= \sum_{i=k}^{n} P(i \text{ X's are } \leq x \text{ and } n-i \text{ X's are } > x)$$

Letting E_i denote the event [i X's are $\leq x$ and n-i X's are $> x$] there are $\binom{n}{i}$ ways to select the X's which are less than or equal to x, and a typical way in which E_i could occur is

$$E_{ij} = [X_{j_1} \leq x, \ldots, X_{j_i} \leq x, x < X_{j_{i+1}}, \ldots, x < X_{j_n}]$$

where $j = 1, 2, \ldots, \binom{n}{i}$ to take into account all possible cases.

For $j \neq j'$ the events E_{ij} and $E_{ij'}$ are disjoint and $E_i = \bigcup_j E_{ij}$. Thus we have

$$F_{nk}(x) = \sum_{i=k}^{n} P(E_i) = \sum_{i=k}^{n} \sum_{j=1}^{\binom{n}{i}} P(E_{ij})$$

and further, when the X_i are assumed to be independent we obtain

$$P(E_{ij}) = \prod_{m=1}^{i} P(X_{j_m} \leq x) \prod_{m=i+1}^{n} (1 - P(X_{j_m} \leq x)) .$$

As a special case of (2.1), to be used later, we have the following result when the X's are distributed according to only two different distributions.

Lemma 2.1. Let $X_1, X_2, \ldots , X_N$ be independent random variables where $\{X_i, 1 \leq i \leq m\}$ are distributed according to $F(x)$ and $\{X_i, m + 1 \leq i \leq N\}$ are distributed according to $G(x)$. Then

$$(2.2) \qquad F_{Nk}(x) = \sum_{i=k}^{N} \sum_{j=o}^{i} \binom{m}{j}\binom{N-m}{i-j} F^j(x) (1-F(x))^{m-j} G^{i-j}(x)(1-G(x))^{N-m-i+j}$$

Proof: Each of the basic events E_i (defined above) can be written as a union of disjoint events E_{ij}, $j = 0, 1, 2, \ldots , i$ where E_{ij} consists of j X's (with distribution $F(x)) \leq x$ and $i - j$ X's (with distribution $G(x)) \leq x$, the remaining X's are $> x$. There are $\binom{m}{j} \binom{N-m}{i-j}$ ways to select such an event, each having probability $F^j(x) (1-F(x))^{m-j} G^{i-j}(x) (1-G(x))^{N-m-i+j}$. We use the convention that $\binom{a}{b} = 0$ if $a < b$.

Remark: When $G \equiv F$ we can use the fact that $\sum_{j=0}^{i} \binom{m}{j}\binom{N-m}{i-j} = \binom{N}{i}$ to get the known result

$$(2.3) \qquad F_{Nk}(x) = \sum_{i=k}^{N} \binom{N}{i} F^i(x)(1-F(x))^{N-i} .$$

In order to derive the distribution theory associated with the sequential ranking procedures proposed in this paper the next lemma will be useful. We consider a random variable X with a continuous distribution function $F(x)$ and define the sign of X to be 1 if $X \geq 0$ and -1 if $X < 0$. Letting E = sign of X, we can compute the joint distribution function for E and $|X|$ as

$$(2.4) \qquad F(x,y) = \begin{cases} 0 & -\infty < y < 0,\ -\infty < x < \infty \\ 0 & -\infty < y < \infty,\ -\infty < x < -1 \\ F(0) - F(-y) & 0 \leq y < \infty,\quad -1 \leq x < 1 \\ F(y) - F(-y) & 0 \leq y < \infty,\quad 1 \leq x < \infty \end{cases}$$

where $F(x,y) = P(E \leq x,\ |X| \leq y)$,

since for $-\infty < y < 0$, $-\infty < x < \infty$, $|X| \geq 0$ with probability 1 implies $F(x,y) = 0$, for $-\infty < y < \infty$, $-\infty < x < -1$, $E = \pm 1$ with probability 1 implies $F(x,y) = 0$, for $0 \leq y < \infty$, $-1 \leq x < 1$, $F(x,y)$ $= P(-y \leq X \leq 0) = F(0) - F(-y)$ and for $0 \leq y < \infty$, $1 \leq x < \infty$, $F(x,y) = P(-y \leq X \leq y) = F(y) - F(-y)$.

In developing the properties of the signed sequential rank an important role will be played by the dependency of the sign of X and $|X|$ and thus we establish a condition whereby E and $|X|$ are independent random variables in

Lemma 2.2 $|X|$ and sign of X ($= E$) are independent if and only if $F(-x) = F(0)\,[1 - F(x) + F(-x)]$ for all $x \geq 0$.

Proof: The marginal distribution for E and $|X|$ are

$$P(E \leq x) = \begin{cases} 0 & x < -1 \\ F(0) & -1 \leq x < 1 \\ 1 & 1 \leq x \end{cases} \quad \text{and} \quad P(|X| \leq y) = \begin{cases} 0 & y < 0 \\ F(y) - F(-y) & 0 \leq y \end{cases}$$

and the product of the marginals is

$$P(E \leq x)\,P(|X| \leq y) = \begin{cases} 0 & -\infty < y < 0, \quad -\infty < x < \infty \\ 0 & -\infty < y < \infty, \quad -\infty < x < -1 \\ F(0)[F(y) - F(-y)] & 0 \leq y < \infty, \quad -1 \leq x < 1 \\ F(y) - F(-y) & 0 \leq y < \infty, \quad 1 \leq x < \infty \end{cases}$$

Thus the joint distribution function of E and $|X|$ will factor into the product of the marginal distributions if and only if $F(0) - F(-y) = F(0)\,[F(y) - F(-y)]$ for all $0 \leq y$ which is equivalent to the condition in the lemma.

Remark: Throughout, we will assume that the basic random variables, usually denoted by X or Y, are defined on the same probability space and have continuous cumulative distribution functions. Thus the ranking procedures to be defined will always be determined uniquely except possibly for sets of measure zero.

3. The Sequential Rank. In the introduction we mentioned the possibility of ranking observations as they are taken without reranking the previous observations. We make this idea formal by

Definition 3.1 The sequential rank of X_n relative to X_1, $X_2, \dots, X_n$ is k if $X_{nk} = X_n$, $k = 1, 2, \dots, n$ where X_{nk} is the k^{th} smallest in the set $\{X_1, X_2, \dots, X_n\}$.

Thus the sequential rank of X_1 is always 1, the sequential rank of X_2 is 1 or 2 according as $X_2 < X_1$ or $X_1 < X_2$, the sequential rank of X_3 is 1, 2 or 3 according as X_3 is the smallest, next largest or largest of the set $\{X_1, X_2, X_3\}$, etc. We use the notation Z_i for the sequential rank of X_i.

Lemma 3.1 There is a one to one correspondence between the set of $n!$ possible orderings $X_{i_1} < X_{i_2} < \dots < X_{i_n}$ and the $n!$ possible sequential rank vectors $(Z_1, Z_2, \dots, Z_n)$.

Proof: We can consider $(X_1, X_2, \dots, X_n) = (x_1, x_2, \dots, x_n)$ where the x_i are n distinct real numbers and the set $\{(x_{i_1}, x_{i_2}, \dots, x_{i_n})\}$ consisting of the $n!$ vectors obtained by permuting the coordinates of $(x_1, x_2, \dots, x_n)$. The corresponding set $\{(X_{i_1}, X_{i_2}, \dots, X_{i_n})\}$ gives the $n!$ possible orderings. Now define the mapping φ from the set $\{(x_{i_1}, x_{i_2}, \dots, x_{i_n})\}$ into the set $\{(r_1, r_2, \dots, r_n)$: $r_1 = 1, r_2 = 1, 2, \dots, r_n = 1, 2, \dots, n\}$ by setting the j^{th} coordinate of $\varphi(x_{i_1}, x_{i_2}, \dots, x_{i_n})$ equal to the rank of x_{i_j} in the set $x_{i_1}, x_{i_2}, \dots x_{i_j}$ i.e. the j^{th} coordinate is r if x_{i_j} is the r^{th} smallest among $x_{i_1}, x_{i_2}, \dots, x_{i_j}$. The mapping φ is one-to-one and onto. (This is almost identical to part of the proof of Theorem 1.1 in [2] page 993.).

By this lemma we mean that if we consider each ordering, say $X_{i_1} < X_{i_2} < \dots < X_{i_n}$ of a set of observations $(X_1, X_2, \dots, X_n)$ and use definition 3.1 to obtain the associated sequential rank vector $(Z_1, Z_2, \dots, Z_n)$, the sequential rank vector is uniquely determined and moreover the sequential rank vector uniquely determines the ordering.

Since a particular ordering of $X_1, X_2, \dots, X_n$ also determines the ordinary rank vector $(T_1, T_2, \dots, T_n)$ in a one-to-one manner, there exists a one-to-one mapping between the set of sequential rank vectors and the set of ordinary rank vectors.

In order to obtain the probability distribution for sequential rank vectors notice that since a particular ordering $X_{i_1} < X_{i_2} < \dots < X_{i_n}$ determines in a one-to-one manner an ordinary rank vector and a sequential rank vector, it is enough to determine a mapping from the ordinary rank vector determined by the ordering, to the sequential rank vector determined by the same ordering. The distribution of $(Z_1, Z_2, \dots, Z_n)$ is then available for a wide class of distributions of the basic variables $X_1, X_2, \dots, X_n$ since Hoeffding has given the distribution of $(T_1, T_2, \dots, T_n)$ in [3].

Consider the indicator function

$$\chi(x,y) = \begin{cases} 1 & \text{if } x \le y \\ 0 & \text{if } x > y \end{cases}$$

and for $X_1, X_2, \dots, X_n$ define the mapping

$$\varphi(X_1, X_2, \ldots, X_n) = \Big(1, \sum_{j=1}^{2} \chi(X_j, X_2), \ldots, \sum_{j=1}^{i} \chi(X_j, X_i),$$

$$\ldots, \sum_{j=1}^{n} \chi(X_j, X_n)\Big)$$

The i^{th} coordinate $\sum_{j=1}^{i} \chi(X_j, X_i)$ is equal to the number of X's in $(X_1, X_2, \ldots, X_i)$ which are less than or equal to X_i, that is, the sequential rank of X_i. But since $X_i < X_j$ iff $T_i < T_j$ $(i \neq j)$ we have

$$\chi(X_i, X_j) = \chi(T_i, T_j),$$

and this holds for all i and j. Hence we have

(3.1) $\varphi(X_1, X_2, \ldots, X_n) = \varphi(T_1, T_2, \ldots, T_n) = (Z_1, Z_2, \ldots, Z_n)$,

and φ is a mapping from the ordinary rank vectors to the sequential rank vectors corresponding to a particular ordering of the basic variables.

Let f_i $i = 1, 2, \ldots, n$ be continuous, non-decreasing functions defined on the unit interval such that $f_i(0) = 1 - f_i(1) = 0$ for each i. Denote by $\mathcal{F}(f_1, f_2, \ldots, f_n)$ the family of all $(F_1, F_2, \ldots, F_n)$ such that $F_i = f_i(F)$ where F runs through all continuous distributions. Now if $X_1, X_2, \ldots, X_n$ are independent and distributed according to $F_1, F_2, \ldots, F_n$, Lehmann has shown in [1] that

(a) the distribution of the ordinary ranks $T_1, T_2, \ldots, T_n$ obtained from $X_1, X_2, \ldots, X_n$ is constant within each family

$\mathcal{F}(f_1, f_2, \ldots, f_n)$, this is lemma 3.2, and

(b) the power of any rank test depends only on $f_1, f_2, \ldots, f_n$, and that uniformly most powerful tests exist, this is Theorem 3.1.

Because of the one-to one correspondence between rank vectors and sequential rank vectors properties (a) and (b) are preserved for sequential ranks. The reason for this is that in computing sequential rank vectors we are merely identifying different points in n - dimensional space with each possible ordering $X_{i_1} < X_{i_2} < \ldots < X_{i_n}$ than when ordinary rank vectors are computed. Thus the probability associated with any subset of ordinary rank vectors can also be associated with a unique subset of sequential rank vectors and we have, analogously as in [1],

<u>Theorem 3.1</u>. Given n functions $f_1^o, f_2^o, \ldots, f_n^o$ and any sequential rank test of the hypothesis H: $(F_1, F_2, \ldots, F_n) \in \mathcal{F}(f_1^o, f_2^o, \ldots, f_n^o)$ (i.e. a test based on the sequential ranks), the power of this test depends only on $f_1^o, f_2^o, \ldots, f_n^o$. That is, if $(F_1, F_2, \ldots, F_n)$ and $(F_1', F_2', \ldots, F_n')$ belong to the same class $\mathcal{F}(f_1, f_2, \ldots, f_n)$ the test has the same power against these two alternatives. Furthermore given any class of alternatives K: $(F_1, F_2, \ldots, F_n) \in \mathcal{F}(f_1', f_2', \ldots, f_n')$ there exists a uniformly most powerful test based on the sequential ranks for testing H against K.

When $X_1, X_2, \ldots, X_n$ are independent and identically distributed the sequential ranks are independent with distribution

$$P(Z_i = k) = 1/i \qquad k = 1, 2, \ldots, i \qquad i = 1, 2, \ldots, n .$$

A proof of this is given in [2]. We see that the mapping defined in (3.1) takes the vector of dependent ranks $(T_1, T_2, \cdots, T_n)$ into the vector of independent sequential ranks $(Z_1, Z_2, \cdots, Z_n)$. Thus according to Theorem 3.1 and the discussion leading to it we lose nothing in the matter of hypothesis testing by considering sequential ranks instead of ordinary ranks, and in fact when we are dealing with independent and identically distributed random variables we find that the sequential ranks are independent.

Since there is a one-to-one correspondence between the ordered observations and the sequential rank vector, the distribution theory for sequential rank vectors is also completely specified by

$$P(X_{i_1} \leq X_{i_2} \leq \cdots \leq X_{i_n}) = \int\cdots\int_{-\infty < x_{i_1} \leq x_{i_2} \leq \cdots \leq x_{i_n} < \infty} \prod_{j=1}^{n} dF_{i_j}(x_{i_j})$$

$$(3.2) \qquad = \int\cdots\int_{-\infty < x_{i_1} \leq \cdots \leq x_{i_n} < \infty} \prod_{j=1}^{n} df_{i_j}(F(x_{i_j}))$$

$$= \int\cdots\int_{0 \leq y_{i_1} \leq y_{i_2} \leq \cdots \leq y_{i_n} \leq 1} \prod_{j=1}^{n} df_{i_j}(y_{i_j})$$

where $y_{i_j} = F(x_{i_j})$ and the X_i are assumed to be independent in this calculation. Let $f = (f_1, f_2, \cdots, f_n)$ and write $P(X_1 \leq X_2 \leq \cdots \leq X_n) = P(f)$. The distribution function for the $n!$ vectors $(Z_1, Z_2, \cdots, Z_n)$ is obtained by computing $P(f)$ for all possible permutations of the components of f. In order to determine the marginal distribution for Z_i we notice that $Z_i = k$ if only if

X_i is the k^{th} smallest among the first i observations, and we get

$$(3.3) \qquad P(Z_i = k) = \sum P(f) \qquad f = (f_{j_1}, f_{j_2}, \ldots, f_{j_i})$$

where f_i is the k^{th} coordinate of f and the summation is taken over the $(i-1)!$ permutations of the coordinates leaving f_i fixed at the k^{th} coordinate.

For the special case where the X_i are taken to be identically distributed, we can take $f_i(x) = x$ without loss of generality, and it is easy to compute (3.2) and (3.3) to get

$$(3.4) \qquad P(f) = 1/n!$$

$$\text{and} \quad P(Z_i = k) = 1/i \qquad k = 1, 2, \ldots, i, \; i = 1, 2, \ldots, n$$

yielding the independence of $Z_1, Z_2, \ldots, Z_n$ as noted above.

Another special case, to be used later, is when the f_i are taken to be the Lehmann alternatives, introduced in [1]. We let $F_i(x) = F^{a_i}(x)$ $a_i > 0$, and in this case a straight forward computation gives

$$(3.5) \qquad P(X_1 \leq X_2 \leq \cdots \leq X_n) = \frac{\prod_{i=1}^{n} a_i}{\prod_{i=1}^{n} \left(\sum_{j=1}^{i} a_j \right)}$$

By relabeling the X's, the probability of any order of the X's can be found using (3.5), giving all the values needed in (3.2) to specify the distribution of the sequential rank vectors.

4. <u>An Application of Sequential Ranking to Hypothesis Testing</u>. In the nonparametric, fixed sample size, two sample problem, it is assumed that there are available two sets of observations $\{X_1, X_2, \ldots, X_m\}$ and $\{Y_1, Y_2, \ldots, Y_n\}$ each set from some probability distribution. The problem is to test the hypothesis that the distributions are the same, against the alternative that they are different. Usually the alternative is more restrictive as when only a shift in location is considered. In this section we consider the nonparametric two sample problem as a sequential problem rather than fixed sample size.

Let X_i $i = 1, 2, \ldots$ and Y_j $j = 1, 2, \ldots$ be independent random variables and assume we wish to test

$$H: \; G = F \quad \text{against} \quad K: \; G = f(F)$$

where F is the continuous cumulative distribution of the X's and G the continuous cumulative distribution of the Y's. We propose to use the sequential probability ratio statistic based on the sequential ranks and we can assume the observations to be taken alternatively as

$$X_1, Y_1, X_2, Y_2, \ldots, X_n, Y_n, \ldots .$$

Let $Z^N = (Z_1, Z_2, \ldots, Z_N)$ be the sequential rank vector based on the first N observations and write $P_1(Z^N)/P_o(Z^N)$ as the sequential probability ratio, P_1 referring to the alternative to the hypothesis, P_o to the hypothesis.

Under the hypothesis $P(Z^N = z) = 1/N!$ and $P_o(Z^N) = 1/N!$ Under the alternative we can compute $P(Z^N = z)$ by noting that each outcome vector z corresponds, in a one-to-one manner, to a particular order of the X's and Y's. For example

$$z^3 = (1, 1, 1) \leftrightarrow X_2 < Y_1 < X_1, \quad z^3 = (1, 2, 1) \leftrightarrow X_2 < X_1 < Y_1 .$$

Each Z^N in turn corresponds to a vector ((F, G, F) or (F, F, G) as in our example) of F's and G's meaning that the observation appearing in the i^{th} smallest position in the ordering of X's and Y's has the distribution F or G according as F or G appears as the i^{th} coordinate of the F, G vector. Thus to compute $P(Z^N = z)$ for all possible values of z we need only compute

$$P(U_1 \leq U_2 \leq \cdots \leq U_N)$$

where U_i is an X or a Y according to the outcome. In particular when f is a continuous increasing function on the unit interval with $f(0) = 1 - f(1) = 0$, the probability distribution is constant for all continuous distributions F and depends only on f. In fact we have

$$P(U_1 \leq U_2 \leq \cdots \leq U_N) = \int_{-\infty < t_1 \leq \cdots \leq t_N < \infty} \cdots \int \prod_{i=1}^{N} df_i(F(t_i))$$

$$= \int_{0 \leq y_1 \leq \cdots \leq y_N \leq 1} \cdots \int \prod_{i=1}^{N} df_i(y_i)$$

by letting $y_i = F(t_i)$ where $f_i(F(t_i)) = F(t_i)$ when $U_i = X_i$ and $f_i(F(t_i)) = f(F(t_i))$ when $U_i = Y_i$.

In the special case of Lehmann alternatives $f(x) = x^a$, $a > 0$ and by (3.5) we get, for N even,

$$P_1(Z^N) = \frac{a^{N/2}}{\prod_{i=1}^{N} \left(\sum_{j=1}^{i} A_j \right)} \quad \text{where} \quad A_i = \begin{cases} 1 & \text{if } U_i = X_i \\ a & \text{if } U_i = Y_i \end{cases}$$

and the probability ratio reduces to

$$\frac{P_1(Z^N)}{P_o(Z^N)} = \frac{N!\, a^{N/2}}{\prod_{i=1}^{N} \left(\sum_{j=1}^{i} A_j \right)} .$$

A similar result holds for N odd. The vector $\bar{A}_N = (A_1, A_2, \ldots, A_N)$, corresponding to the vector of F's and G's determines $P(Z^N = z)$ for $[(N/2)!]^2$ outcomes z out of the $N!$ possible. We can compute the probability ratios at each stage using the following relations:

$$(4.1) \qquad S_N = \frac{P_1}{P_o} = \begin{cases} \dfrac{N!\, a^{\frac{N-1}{2}}}{\prod_{i=1}^{N} \left(\sum_{j=1}^{i} A_j \right)} & N \text{ odd} \\[2ex] \dfrac{N!\, a^{N/2}}{\prod_{i=1}^{N} \left(\sum_{j=1}^{i} A_j \right)} & N \text{ even} \end{cases}$$

$$
(4.2) \qquad S_{N+1} = \frac{P_1}{P_o} = \begin{cases} \dfrac{(N+1)!\, a^{\frac{N+1}{2}}}{\prod\limits_{i=1}^{Z-1}\left(\sum\limits_{j=1}^{i} A_j\right) \prod\limits_{i=Z-1}^{N}\left(a + \sum\limits_{j=1}^{i} A_j\right)} & N \text{ odd} \\[2em] \dfrac{(N+1)!\, a^{N/2}}{\prod\limits_{i=1}^{Z-1}\left(\sum\limits_{j=1}^{i} A_j\right) \prod\limits_{i=Z-1}^{N}\left(1 + \sum\limits_{j=1}^{i} A_j\right)} & N \text{ even} \end{cases}
$$

where Z = sequential rank of $Y_{\frac{N+1}{2}}$, N odd, and Z = sequential rank of $X_{\frac{N+2}{2}}$, N even. At the $N + 1^{st}$ observation Z is determined and if $Z = k$, the $N + 1^{st}$ observation came between the $k - 1^{st}$ and the k^{th} smallest observations of the preceding N observations. Thus $\bar{A}_{N+1} = (A_1, A_2, \ldots, A_{k-1}, A^*, A_k, \ldots, A_N)$ where $A^* = 1$ if the $N + 1^{st}$ observation is an X and $A^* = a$ if the observation is a Y. Using (4.1) and (4.2) and Z we can pass from S_N to S_{N+1} as the observations are taken. For example $S_1 = 1$

$$
S_2 = \begin{cases} 2a/1+a & \text{if } X_1 < Y_1 \leftrightarrow \bar{A}_2 = (1, a) \\[1em] 2/1+a & \text{if } Y_1 < X_1 \leftrightarrow \bar{A}_2 = (a, 1) \end{cases}
$$

$$S_3 = \begin{cases} \frac{6a}{(1+a)(2+a)} & \text{if} \begin{array}{l} X_1 < Y_1 < X_2 \\ X_2 < Y_1 < X_1 \end{array} \leftrightarrow \bar{A}_3 = (1, a, 1) \\ \frac{6}{(1+a)(2+a)} & \text{if} \begin{array}{l} Y_1 < X_1 < X_2 \\ Y_1 < X_2 < X_1 \end{array} \leftrightarrow \bar{A}_3 = (a, 1, 1) \\ \frac{3a}{(2+a)} & \text{if} \begin{array}{l} X_1 < X_2 < Y_1 \\ X_2 < X_1 < Y_1 \end{array} \leftrightarrow \bar{A}_3 = (1, 1, a) \end{cases}$$

We noted before that under the hypothesis $a = 1$ the sequential ranks $Z_1, Z_2, \ldots, Z_N$ are independent. However, when $a \neq 1$ we do not have this independence property. Consider the case $N = 3$ where we observe X_1, Y_1, X_2 in that order. The possible outcomes are

Ordered observations	Sequential ranks	Probability
$X_1 < X_2 < Y_1$	(1, 2, 2)	a/2(2+a)
$X_2 < X_1 < Y_1$	(1, 2, 1)	a/2(2+a)
$X_1 < Y_1 < X_2$	(1, 2, 3)	a/(1+a)(2+a)
$X_2 < Y_1 < X_1$	(1, 1, 1)	a/(1+a)(2+a)
$Y_1 < X_1 < X_2$	(1, 1, 3)	1/(1+a)(2+a)
$Y_1 < X_2 < X_1$	(1, 1, 2)	1/(1+a)(2+a)

and the marginal distributions are easily computed as

$P(Z_1 = 1) = 1$ $\qquad$ $P(Z_3 = 1) = a(3+a)/2(1+a)(2+a)$

$P(Z_2 = 1) = 1/1+a$ $\qquad$ $P(Z_3 = 2) = (2+a+a^2)/2(1+a)(2+a)$

$P(Z_2 = 2) = a/1+a$ $\qquad$ $P(Z_3 = 3) = 1/2+a$

Now $P((Z_1, Z_2, Z_3) = (1, 1, 1)) = a/(1+a)(2+a)$ and $P(Z_1 = 1)$ $P(Z_2 = 1)$ $P(Z_3 = 1) = \frac{a}{(1+a)(2+a)} \frac{3+a}{2(1+a)}$ and it follows that Z_1, Z_2, Z_3 cannot be independent unless $a = 1$ since independence of Z_1, Z_2, Z_3 implies $(3+a)/2(1+a) = 1$ which in turn implies $a = 1$. Thus we have

Theorem 4.1 Let $X_1, Y_1, X_2, \ldots, X_N, Y_N$ be independent random variables with X_i distributed according to F and Y_i distributed according to F^a, $a > 0$. The sequential ranks based on such a sequence are independent if and only if $a = 1$.

As an illustration of the sequential probability ratio test based on the sequential ranks consider the data given below.

$X_1 = 3.926$	$X_{10} = 4.08$	$Y_1 = 4.70$	$Y_{10} = 1.56$
$X_2 = 3.45$	$X_{11} = 3.67$	$Y_2 = 4.15$	$Y_{11} = 4.29$
$X_3 = 2.00$	$X_{12} = 2.94$	$Y_3 = 4.55$	$Y_{12} = 1.74$
$X_4 = 2.28$	$X_{13} = 5.90$	$Y_4 = 3.31$	$Y_{13} = 2.17$
$X_5 = 3.494$	$X_{14} = 2.18$	$Y_5 = 2.13$	$Y_{14} = 1.97$
$X_6 = 4.25$	$X_{15} = 5.39$	$Y_6 = 4.686$	$Y_{15} = 4.689$
$X_7 = 2.382$	$X_{16} = 2.74$	$Y_7 = 2.68$	$Y_{16} = 2.87$
$X_8 = 3.02$	$X_{17} = 3.492$	$Y_8 = 2.36$	$Y_{17} = 3.17$
$X_9 = 3.26$	$X_{18} = 2.70$	$Y_9 = 3.95$	

The data is taken from Table 600A page 600 of "Statistics, A New Approach," W. A. Wallis and H. V. Roberts, The Free Press, Glencoe, Illinois. If we assume X has some continuous distribution F and Y has F^a as a distribution then

$$P(X < Y) = \int_{-\infty}^{\infty} F(y)\, dF^a(y) = \frac{a}{1+a} .$$

Suppose we consider $a = 4$, $P(X < Y) = .8$ as the alternative to the hypothesis $a = 1$. We take as boundaries for the sequential probability ratio test

$$A = \frac{1 - \beta}{\alpha} = \frac{1 - .05}{.05} = 19$$

$$B = \frac{\beta}{1 - \alpha} = \frac{.05}{1 - .05} = .0526$$

and if $S_N \leq B$ we accept H: $a = 1$, if $S_N \geq A$ we accept K: $a = 4$ and if $B < S_N < A$ we take another observation and compute S_{N+1}, repeating the test. Using the computational formulas (4.1) and (4.2) we get

$S_1 = 1$	$S_{11} = .454$
$S_2 = 1.6$	$S_{12} = .725$
$S_3 = 2.0$	$S_{13} = .764$
$S_4 = 3.2$	$S_{14} = .586$
$S_5 = 4.15$	$S_{15} = .467$
$S_6 = 6.65$	$S_{16} = .234$
$S_7 = 8.75$	$S_{17} = .168$
$S_8 = 4.0$	$S_{18} = .242$
$S_9 = 3.31$	$S_{19} = .138$
$S_{10} = .809$	$S_{20} = .0288$

and since $S_{20} \leq .0526$ we accept H at the 20^{th} observation.

Notice that even though the probability ratio S_N is written as a function of the sequential ranks, in (4.1) and (4.2), it can also be computed as a function of the order configuration. By this we mean, for example, the order configuration 1 a 1 stands for $X_1 < Y_1 < X_2$ or $X_2 < Y_1 < X_1$ and a 1 1 stands for $Y_1 < X_1 < X_2$ or $Y_1 < X_2 < X_1$. Each order configuration determines a value of S_N as a function of a. It can happen that for some value of $a \neq 1$ and two different configurations, S_N takes on the same value. As an example consider $N = 6$ and the configurations a 1 1 1 a a and 1 a a a 1 1. The denominators in S_6 for these configurations are

$$g_1(a) = a(a + 1)(a + 2)(a + 3)(2a + 3)(3a + 3)$$

$$g_2(a) = 1(1 + a)(1 + 2a)(1 + 3a)(2 + 3a)(3 + 3a)$$

respectively. For $a = 1/2$ and $a = 2$ we get

$$g_1(1/2) = g_2(1/2) = 945/8 \quad \text{and} \quad g_1(2) = g_2(2) = 7560.$$

Let $c(t)$ be the number of different configurations such that $S_N = t$. We have

$$(4.3) \qquad P(S_N = t \mid a) = \frac{[\frac{N}{2}]! \left(N - [\frac{N}{2}]\right)! \; c(t) \; a^{[\frac{N}{2}]}}{\prod_{i=1}^{N}\left(\sum_{j=1}^{i} a_j\right)}$$

where the a_j's correspond to any particular configuration making $S_N = t$ ($a_j = 1$ or a according as X or Y is in the j^{th} place).

(4.3) follows because any two configurations which make $S_N = t$ have the same probability under the alternative to the hypothesis. Under the hypothesis

$$(4.4) \qquad P(S_N = t \mid a = 1) = \frac{[\frac{N}{2}]!\left(N - [\frac{N}{2}]\right)!\, c(t)}{N!} .$$

In (4.3) and (4.4) $[x]$ is the greatest integer function.

In Wald's sequential probability ratio test the approximations $A \leq \frac{1-\beta}{\alpha}$ and $B \geq \frac{\beta}{1-\alpha}$ are valid when the probability of termination of the test is 1. These inequalities were derived under the assumption that the basic sequence of probability ratios was determined from an independent sequence of observations and that the sequential probability ratio at the n^{th} observation is formed as a product of independent and identically distributed random variables. Under the alternative hypothesis we have found that the sequential ranks are not independent. Thus we must now show that the test terminates with probability 1 in order to interpret α and β as error probabilities.

It is enough to show that the test terminates with probability 1 considering only N even. For N even we can write

$$(4.5) \qquad S_N^{-1} = \prod_{i=1}^{N} a^{-1/2} \frac{1}{i} \sum_{j=1}^{i} A_j$$

and define $\bar{A}_1^N = \frac{1}{i} \sum_{j=1}^{i} A_j$ with $Y_1^N = a^{-1/2} \bar{A}_1^N$ and $Z_1^N = \log Y_1^N$. We consider first the case where the null hypothesis is true. A_1, $A_2, \ldots, A_N$ are dependent random variables with

(4.6) $$P(A_j = 1) = P(A_j = a) = 1/2$$

giving $E(A_j) = E(\bar{A}_1^N) = \frac{1+a}{2}$. For $i \neq j$ we have

$$P(A_i = 1, A_j = 1) = P(A_i = a, A_j = a) = \frac{1}{4}\frac{N-2}{N-1}$$

(4.7)

$$P(A_i = 1, A_j = a) = P(A_i = a, A_j = 1) = \frac{1}{4}\frac{N}{N-1}$$

and a simple computation gives $Var(A_j) = (\frac{1-a}{2})^2$ and $Cov(A_i, A_j)$ $= \frac{-1}{N-1}(\frac{1-a}{2})^2$. Also $E(Y_1^N) = a^{-1/2} E(\bar{A}_1^N) = \frac{a^{-1/2} + a^{1/2}}{2}$ and

$$Var(Y_1^N) = \frac{a^{-1}}{i^2} Var(\bar{A}_1^N) = \frac{a^{-1}}{i^2}\left\{\sum_{j=1}^{i} Var(A_j) + 2\sum_{j=1}^{i-1}\sum_{k=j+1}^{i} Cov(A_j, A_k)\right\}$$

$$= \frac{a^{-1}}{i^2}\left\{i\left(\frac{1-a}{2}\right)^2 + (i^2 - i)\frac{-1}{N-1}\left(\frac{1-a}{2}\right)^2\right\}$$

$$= \frac{1}{a}\left(\frac{1-a}{2}\right)^2 \frac{1}{N-1}\frac{N-i}{i}$$

and notice that $Var(Y_1^N)$ is decreasing in i as $i = 1, 2, \ldots, N$. If $1 < a$ then $1 \leq \bar{A}_1^N \leq a$ and $a^{-1/2} \leq Y_1^N \leq a^{1/2}$. If $a < 1$ then $a^{1/2} \leq Y_1^N \leq a^{-1/2}$.

In order to show that the test terminates with probability one it is enough to show that $S_N^{-1} \to \infty$ in probability. Thus for arbitrary positive B we show that

$$\lim_{N\to\infty} P(S_N^{-1} \le 1/B) = \lim_{N\to\infty} P(\log S_N^{-1} \le \log 1/B) = 0 .$$

Let $K = \log 1/B$ and use Chebyshev's inequality to get

$$P\left(\sum_{i=1}^{N} Z_i^N \le K\right) = P\left(\sum_{i=1}^{N} Z_i^N - \sum_{i=1}^{N} E(Z_i^N) \le K - \sum_{i=1}^{N} E(Z_i^N)\right)$$

$$\le P\left(\left| \sum_{i=1}^{N} Z_i^N - \sum_{i=1}^{N} E(Z_i^N) \right| \ge - K + \sum_{i=1}^{N} E(Z_i^N)\right)$$

by taking N large enough to make $K - \sum_{i=1}^{N} E(Z_i^N) < 0$. This can be done since Y_i^N is bounded, and bounded away from 0, we have

$$Z_i^N = \log Y_i^N = \log \lambda_a + \frac{Y_i^N - \lambda_a}{\lambda_a} - (Y_i^N - \lambda_a)^2 \left(\frac{1}{\xi_i^N}\right)^2$$

where $\lambda_a = E(Y_i^N) = \dfrac{a^{-1/2} + a^{1/2}}{2} > 1$ and ξ_i^N is bounded away from 0, and further

$$E(Z_i^N) \ge \log \lambda_a - c\,\frac{N-i}{i(N-1)} \qquad c > 0$$

$$\sum_{i=1}^{N} E(Z_i^N) = N \log \lambda_a - O(\log N) = O(N) > 0 .$$

Thus

$$P\left(\sum_{i=1}^{N} Z_i^N \le K\right) \le \frac{\operatorname{Var}\left(\sum_{i=1}^{N} Z_i^N\right)}{O(N^2)}$$

and

$$\mathrm{Var}\left(\sum_{i=1}^{N} Z_i^N\right) = \sum_{i=1}^{N} \mathrm{Var}(Z_i^N) + 2 \sum_{i<j} \mathrm{Cov}(Z_i^N, Z_j^N)$$

$$\leq \sum_{i=1}^{N} \mathrm{Var}(Z_i^N) + 2 \sum_{i<j} (\mathrm{Var}(Z_i^N) \cdot \mathrm{Var}(Z_j^N))^{1/2} .$$

Now, expanding $\log Y_i^N$ in only two terms

$$\mathrm{Var}(Z_i^N) = E(\log Y_i^N - E(\log Y_i^N))^2 \leq E(\log Y_i^N - \log \lambda_a)^2$$

$$= E\left(\frac{Y_i^N - \lambda_a}{\eta_i^N}\right)^2 \leq c' \ \mathrm{Var}(Y_i^N) = c'' \frac{N-i}{i(N-1)}$$

and $\frac{N-i}{i(N-1)}$ is decreasing in i. Now we can write

$$\mathrm{Var}\left(\sum_{i=1}^{N} Z_i^N\right) \leq O(\log N) + 2\, c \sum_{i=1}^{N} (N-i)\ \mathrm{Var}(Z_i^N)$$

$$= O(N \log N)$$

and finally

$$P\left(\sum_{i=1}^{N} Z_i^N \leq K\right) \leq \frac{O(N \log N)}{O(N^2)} \rightarrow 0 \text{ as } N \rightarrow \infty .$$

Since $\log S^{-1} = \sum_{i=1}^{N} Z_i^N$ we have $S_N^{-1} \rightarrow \infty$ in probability, and when the null hypothesis is true, the test terminates with probability 1.

More generally now consider

(4.8) $P_N = P(A^{-1} \leq S_N^{-1} \leq B^{-1}) = P(K_1 - \mu_N \leq \log S_N^{-1} - \mu_N \leq K_2 - \mu_N)$

where $K_1 = \log A^{-1}$, $K_2 = \log B^{-1}$ and $\mu_N = E(\log S_N^{-1})$. For large enough values of N

(4.9)
$$P_N \leq \begin{cases} P(|\log S_N^{-1} - \mu_N| \geq \mu_N - K_2) \leq \dfrac{\mathrm{Var}(\log S_N^{-1})}{(\mu_N - K_2)^2} & \text{if} \quad \mu_N \nearrow \infty \\ \\ P(|\log S_N^{-1} - \mu_N| \geq K_1 - \mu_N) \leq \dfrac{\mathrm{Var}(\log S_N^{-1})}{(K_1 - \mu_N)^2} & \text{if} \quad \mu_N \searrow -\infty \end{cases}$$

The test will terminate with probability 1 as long as $P_N \to 0$, and this is independent of the true distribution of the Y population since the inequalities in (4.9) were obtained without reference to the distribution of the Y's. In particular we found that when the X and Y populations are identically distributed, $\mu_N = O(N)$ and $\mathrm{Var}(\log S_N^{-1}) = O(N \log N)$.

The method just given to show that the probability of termination of the test is one is not satisfactory for all alternatives since the verification of condition (4.9) is difficult. We now consider a better approach. As before take N even and write the probability ratio as

(4.10) $$T_N = S_N^{-1} = \prod_{i=1}^{N} \left\{ a^{-1/2} \frac{1}{i} \sum_{j=1}^{i} A_j \right\} .$$

In order to show that the probability of termination is one it is enough to show that $N^{-1} \log T_N$ converges to some non-zero constant since for fixed boundaries A, B, the equivalent formulation

$$\frac{1}{N} \log A^{-1} < \frac{1}{N} \log T_N < \frac{1}{N} \log B^{-1}$$

will terminate with probability one, provided $N^{-1} \log T_N$ converges in probability to some non-zero constant.

Let $2n = N$ and let $Z_1, Z_2, \ldots Z_N$ be the order statistics for the combined sample. Define the empirical cumulative distribution functions for the X's and Y's as

$$F_n(t) = \frac{(\text{number of X's} \leq t)}{n}$$

$$G_n(t) = \frac{(\text{number of Y's} \leq t)}{n}$$

Since $\sum_{j=1}^{i} A_j$ = (number of X's in $Z_1, Z_2, \ldots, Z_i$) + a(number of Y's in $Z_1, Z_2, \ldots, Z_i$) we can write

$$\frac{1}{i} \sum_{j=1}^{i} A_j = \frac{n}{i} F_n(Z_i) + \frac{n}{i} a\, G_n(Z_i)$$

and

$$\frac{1}{N} \log T_N = -\frac{1}{2} \log a - \log 2 + \log N - \frac{1}{N} \log N!$$
$$+ \frac{1}{N} \sum_{i=1}^{N} \log (F_n(Z_i) + a\, G_n(Z_i)).$$

Since $\lim_{N \to \infty} (\log N - \frac{1}{N} \log N!) = 1$, we have

$$\lim_{N \to \infty} N^{-1} \log T_N = \log e/2\sqrt{a} + \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \log (F_n(Z_i) + a\, G_n(Z_i))$$

$$= \log e/2\sqrt{a} + \frac{1}{2} \int_{-\infty}^{\infty} \log (F(x) + a\, G(x))(dF(x) + dG(x)),$$

the latter limit following from a result of I. R. Savage and J. Sethuraman communicated to the author by Sethuraman as

Theorem (Savage-Sethuraman) Let $X_1, X_2 \ldots X_n, Y_1, Y_2, \ldots Y_n$ be independent random variables where the X_i are distributed according to the continuous distribution F and the Y_i according to the continuous distribution G. Let $Z_1, Z_2, \ldots, Z_N$ $(N = 2n)$ be the order statistics of the combined sample and let F_n and G_n be the empirical cumulative distribution functions of the X's and Y's respectively. Then

$$N^{-1} \sum_{i=1}^{N} \log (F_n(Z_i) + a\, G_n(Z_i)) \to \frac{1}{2} \int_{-\infty}^{\infty} \log(F(x) + a\, G(x))(dF(x) + dG(x))$$

in probability. (see [10])

In our case $G = F$ or F^a depending upon which hypothesis holds. However we will consider the entire class of alternatives F^b, $b > 0$ which could hold. Let $N^{-1} \log T_N \to L_a(b)$. Then

$$L_a(b) = \log e/2\sqrt{a} + \frac{1}{2}\int_{-\infty}^{\infty} \log(F + aF^b)\, d\,\frac{(F + aF^b)}{a}$$

$$+ \frac{1}{2}\int_{-\infty}^{\infty} \log(F + aF^b)\, d((1 - 1/a)F)$$

(4.12)

$$= \log e/2\sqrt{a} + \frac{1}{2a}\int_0^{1+a} \log t\, dt + \frac{a-1}{2a}\int_0^1 \log(t + at^b)dt$$

$$= -\log 2 - \frac{1}{2}\log a + \frac{1+a}{2a}\log(1+a) + \frac{a-1}{2a}\int_0^1 \log(1+at^{b-1})dt$$

The function $\int_0^1 \log(1+at^{b-1})dt$ decreases as b increases, and thus $L_a(b)$ is monotone in b, decreasing when $1 < a$, and increasing when $a < 1$.

Under the null hypothesis $b = 1$ and

$$L_a(1) = \log \frac{a^{-1/2} + a^{1/2}}{2} > 0 \quad \text{for } a \neq 1 .$$

Under the alternative hypothesis $b = a$ and

$$L_a(a) = \log \frac{a^{-1/2} + a^{1/2}}{2} - \frac{(a-1)^2}{2}\int_0^1 \frac{1}{a + t^{1-a}}\, dt.$$

In order to show that the test terminates with probability 1 we must have $L_a(a) \neq 0$ for $a \neq 1$. In fact we will show that $L_a(a) < 0$ for $a \neq 1$. Notice first that it is enough to consider $0 < a < 1$ since

$$L_{\frac{1}{a}}(1/a) = \log \frac{a^{1/2} + a^{-1/2}}{2} - \frac{(a^{-1} - 1)^2}{2} \int_0^1 \frac{1}{a^{-1} + t^{1-1/a}}\, dt$$

$$= \log \frac{a^{-1/2} + a^{1/2}}{2} - \frac{(a-1)^2}{2a^2} \int_0^1 \frac{a}{1+at^{1-1/a}}\, dt$$

$$= \log \frac{a^{-1/2} + a^{1/2}}{2} - \frac{(a-1)^2}{2a^2} \int_0^1 \frac{a^2\, s^{a-1}}{1+a\, s^{a-1}}\, ds \qquad (t = s^a)$$

$$= \log \frac{a^{-1/2} + a^{1/2}}{2} - \frac{(a-1)^2}{2} \int_0^1 \frac{1}{a+s^{1-a}}\, ds = L_a(a) \ .$$

We can write

$$2L_a(a) = \log\left(1 + \frac{(a-1)^2}{4a}\right) - (a-1)^2 \int_0^1 \frac{1}{a+t^{1-a}}\, dt$$

$$= \int_0^1 \frac{(a-1)^2}{4a + (a-1)^2 t}\, dt - \int_0^1 \frac{(a-1)^2}{a+t^{1-a}}\, dt$$

$$= (a-1)^2 \int_0^1 \left(\frac{1}{4a + (a-1)^2\, t} - \frac{1}{a+t^{1-a}}\right) dt$$

and we wish to show that $a+t^{1-a} < 4a + (a-1)^2\, t$ for $0 \leq t \leq 1$ and $0 < a < 1$. Define

$$h(a,\ t) = 3a + (a-1)^2 t - t^{1-a}$$

and notice that

$$\frac{\partial h}{\partial t} = (a-1)^2 - (1-a)\, t^{-a} < (a-1)^2 - (1-a) = (a-1)\, a < 0$$

$$\frac{\partial^2 h}{\partial t^2} = a(1-a)\, t^{-(a+1)} > 0 \ .$$

Since $h(a, 0) = 3a$, $h(a, 1) = a(a + 1)$ we may conclude that $h(a, t) > 0$, which makes the integrand in $2L_a(a)$ negative as was to be proved.

We have shown that the sequential test terminates with probability one under the null and alternative hypothesis and moreover the test will terminate with probability one when the Y's are distributed according to F^b for $b > 0$ except possibly for only one value of b. This follows from the monotonicity of $L_a(b)$.

We also remark here that for a fixed sample size test of

$$\mathcal{H}_0:\ X \sim F,\ \ Y \sim F$$

$$\text{against } \mathcal{H}_1:\ X \sim F,\ \ Y \sim F^a \quad a \neq 1,\ a > 0$$

using ranks of observations, the Neyman-Pearson theory would give a most powerful test of the form

$$\text{accept } \mathcal{H}_0 \text{ for } S_N^{-1} > K \ .$$

An equivalent test would be to accept $\mathcal{H}_0$ if $\frac{1}{N} \log S_N^{-1} > \frac{1}{N} \log K$. Assume $a > 1$ and let $L_a(b_0) = 0$. Then

$$\lim_{N \to \infty} P\left(\frac{1}{N} \log S_N^{-1} > \frac{1}{N} \log K\right) = 1 \quad \text{if} \quad Y \sim F^b \quad b < b_0$$

$$\lim_{N \to \infty} P\left(\frac{1}{N} \log S_N^{-1} < \frac{1}{N} \log K\right) = 1 \quad \text{if} \quad Y \sim F^b \quad b > b_0$$

and thus for a test of the composite hypotheses

$$\mathcal{H}_0: \quad X \sim F, \; Y \sim F^b \quad 0 < b < b_0(a)$$

$$\text{against} \quad \mathcal{H}_1: \quad X \sim F, \; Y \sim F^b \quad b_0(a) < b$$

the test is consistent (in probability).

5. The Signed Sequential Rank. We now extend the ranking procedure defined in Section 3 to include the sign of the observation. This corresponds to the signed rank statistic used in fixed sample size problems.

Definition 5.1 The signed sequential rank of X_n relative to $X_1, X_2, \dots, X_n$ is the product of the sequential rank of $|X_n|$ relative to $|X_1|, |X_2|, \dots, |X_n|$ and sign (X_n), where sign $(X_n) = 1$ if $X_n \geq 0$ and sign $(X_n) = -1$ if $X_n < 0$.

In the case of sequential rank vectors there are $N!$ points in the sample space corresponding to a sample $X_1, X_2, \dots, X_N$ and in the case of signed sequential rank vectors there are $2^N N!$ points corresponding to the same sample. Of course if the basic variables (the X_i) are positive random variables (or negative) the signed sequential ranks are equivalent to the sequential ranks.

We found in Section 3 that when the basic random variables are independent and identically distributed the sequential ranks are independent. This result does not hold in general for signed sequential ranks and so we now determine a sufficient condition for this result to hold in this case.

Let $X_1, X_2, \dots, X_N$ be independent and identically distributed random variables and let Z_i = sequential rank of $|X_i|$ relative to $|X_1|, |X_2|, \dots, |X_i|$, E_i = sign (X_i) with $Y_i = E_i Z_i$, $i = 1, 2, \dots, N$. If $F(x) = P(X_1 \leq x)$ satisfies the condition in lemma 2.2. $E_i, |X_1|, |X_2|, \dots, |X_i|$ are independent and it follows that E_i and Z_i are independent. Thus we get

$$P(Y_i = j) = P(E_i = 1, Z_i = j) = P(E_i = 1)\, P(Z_i = j)$$

(5.1) $$= (1-F(0))\, 1/i$$

$$P(Y_i = -j) = P(E_i = -1, Z_i = j) = P(E_i = -1)\, P(Z_i = j)$$

$$= F(0)\, 1/i$$

for $j = 1, 2, \ldots, i$, $i = 1, 2, \ldots, N$. $P(Z_i = j) = 1/i$ follows from (3.4)

We will now show that the condition given in lemma 2.2 is a sufficient condition to guarantee the independence of the signed sequential ranks.

Theorem 5.1. If $X_1, X_2, \ldots, X_N$ are independent and identically distributed according to $F(x)$ where $F(-x) = F(0)[1-F(x) + F(-x)]$ for all $x \geq 0$ then the signed sequential ranks $Y_1, Y_2, \ldots, Y_N$ are independent random variables.

Proof: Let $(i_1, i_2, \ldots, i_N)$ be an arbitrary outcome vector for $(Y_1, Y_2, \ldots, Y_N)$ and let k be the number of positive integers in $(i_1, i_2, \ldots, i_N)$. We have

$$\prod_{m=1}^{N} P(Y_m = i_m) = \frac{[1-F(0)]^k\, [F(0)]^{N-k}}{N!}$$

from (5.1). Each outcome vector corresponds to a particular ordering of the X's, with N-k of the X's negative. The absolute values of these N-k X's have a particular ordering among the positive X's. So each outcome vector is equivalent to an event like $[0 \leq \epsilon_1 X_{j_1} < \epsilon_2 X_{j_2} < \cdots < \epsilon_N X_{j_N}]$ where k of the ϵ_j are 1 and N-k are -1. The distribution function for -X is $1-F(-x)$ and using $F(-x) = F(0)[1-F(x) + F(-x)]$ we have

$$d(1-F(-x)) = -dF(-x) = \frac{F(0)}{1-F(0)}\, dF(x) \ .$$

Hence

$$P(Y_1 = i_1, Y_2 = i_2, \ldots , Y_N = i_N)$$

$$= P(0 \leq \epsilon_1 X_{j_1} < \ldots < \epsilon_N X_{j_N}) = \int_{0 \leq y_1 < \ldots < y_N < \infty} \cdots \int \prod_{i=1}^{N} dF_{X_{j_i}}(y_i)$$

$$= \left[\frac{F(0)}{1-F(0)}\right]^{N-k} \int_{0 \leq y_1 < \ldots < y_N < \infty} \cdots \int \prod_{i=1}^{N} dF_{X_{j_i}}(y_i)$$

$$= \left[\frac{F(0)}{1-F(0)}\right]^{N-k} P(0 \leq X_{j_1} < \ldots < X_{j_N})$$

$$= \left[\frac{F(0)}{1-F(0)}\right]^{N-k} \frac{P(0 \leq X_i, \text{ for all } i\ )}{N!} = \frac{[F(0)]^{N-k}\,[1-F(0)]^{k}}{N!}$$

Thus $P(Y_1 = i_1, Y_2 = i_2, \ldots , Y_N = i_N) = \prod_{j=1}^{N} P(Y_j = i_j)$ establishing the independence.

Remark: In the proof of the theorem we have assumed that $F(0) \neq 1$. If $F(0) = 1$, the X_i are negative random variables and the signed sequential ranks reduce to $\{-(\text{sequential rank of } |X_i|)\}$, which are independent.

The condition $F(-x) = F(0)[1-F(x) + F(-x)]$ for all $x > 0$ is satisfied by distributions of positive, negative and symmetric (about 0) random variables. A larger class of distributions satisfies the condition. If we consider all measurable sets $A \subset [0, \infty)$ and define $-A = \{x: \ -x \in A\}$, then the condition

$$\Pr\{X \in A\} = k \Pr\{X \in -A\} \qquad k \geq 0, \text{ all } A$$

is enough to insure that $F(-x) = F(0)[1-F(x) + F(-x)]$ for all $x \geq 0$, since taking $A = [0, \infty)$ we get $k = \frac{1-F(0)}{F(0)}$ and taking $A = [0, x]$ we get $F(-x) = F(0)[1-F(x) + F(-x)]$ for all $x \geq 0$. On the other hand, starting with $F(-x) = F(0)[1-F(x) + F(-x)]$ for all $x \geq 0$ we get

$$dF(x) = \frac{1-F(0)}{F(0)} d(-F(-x))$$

and

$$\Pr(X \in A) = \int_A dF(x) = \frac{1-F(0)}{F(0)} \int_A d(-F(-x)) = \frac{1-F(0)}{F(0)} \Pr(X \in -A) .$$

We now consider the asymtotic distribution of sums of signed sequential ranks based on observations from a distribution satisfying the condition in Theorem 5.1. Let $X_1, X_2, \ldots, X_n$ be independent identically distributed random variables with common distribution function $F(x)$ such that for all $x \geq 0$ $F(-x) = F(0)[1-F(x) + F(-x)]$ holds. Define Y_n = signed sequential rank of X_n. Using (5.1) we get easily

$$E(Y_n) = (1-2F(0)) \frac{n+1}{2}$$

(5.2)

$$\mathrm{Var}(Y_n) = \left(\frac{1}{3} - \frac{(1-2F(0))^2}{4}\right) n^2 + \left(\frac{1}{2} - \frac{(1-2F(0))^2}{2}\right) n + \left(\frac{1}{6} - \frac{(1-2F(0))^2}{4}\right)$$

When $F(x)$ satisfies the condition of Theorem 5.1 the signed sequential ranks are independent, but not identically distributed, and forming the partial sums, $S_n = \sum_{i=1}^{n} Y_i$ we have

$$E(S_n) = \left(\frac{1-2F(0)}{4}\right) n^2 + \left(\frac{3(1-2F(0))}{4}\right) n$$

(5.3)

$$\mathrm{Var}(S_n) = \left(\frac{4-3(1-2F(0))^2}{72}\right) n(n+1)(2n+1) + \left(\frac{1-(1-2F(0))^2}{4}\right) n(n+1) + \left(\frac{4-6(1-2F(0))^2}{24}\right) n \,.$$

Now for $\epsilon > 0$, $k = 1, 2, \ldots, n$, $\sigma_n^2 = \mathrm{Var}(S_n)$ it follows that for large enough values of n

$$\int_{|y-E(Y_k)| > \epsilon \sigma_n} (y - E(Y_k))^2 \, dF_{Y_k}(y) = 0$$

because the range of integration becomes a set with zero probability since Y_k is bounded according to $|Y_k| \leq k$ and $\sigma_n \approx c \cdot n^{3/2}$. Thus as $n \to \infty$ the integral is zero for all $k = 1, 2, \ldots, n$ and by the Lindegerg-Feller Theorem it follows that S_n is asymtotically normal.

If we normalize the signed sequential ranks and then consider partial sums we get

$$S_n^* = \sum_{i=1}^{n} \frac{Y_i - E(Y_i)}{[\mathrm{Var}(Y_i)]^{1/2}}$$

and

$$0 \leq \left| \frac{Y_i - E(Y_i)}{[\mathrm{Var}(Y_i)]^{1/2}} \right| \leq \frac{2i}{(\alpha_1 i^2 + \alpha_2 i + \alpha_3)^{1/2}} = \frac{2}{(\alpha_1 + \alpha_2/i + \alpha_3/i^2)^{1/2}} \to \frac{2}{\sqrt{\alpha_1}}$$

as $i \to \infty$ where $\alpha_1 = \frac{1}{3} - \frac{(1-2F(0))^2}{4}$. Hence the normalized signed sequential ranks are uniformly bounded and by the bounded Lyapounov Theorem $S_n^*/\sqrt{n}$ is asymtotically distributed as a unit normal random variable.

As was noted in the introduction, some statistical problems are concerned with detecting a change in the distribution of a sequence of observations obtained from some process. We now consider the case where in the basic set of independent random variables $(X_1, X_2, \ldots, X_N)$ the first m are distributed according to $F(x)$ and the remaining $N-m$ are distributed according to $G(x)$. As before let Y_i denote the signed sequential rank of X_i. Since each possible outcome vector for $(Y_1, Y_2, \ldots, Y_N)$ corresponds to an event of the form

$$[0 \leq \epsilon_1 X_{i_1} < \epsilon_2 X_{i_2} < \ldots < \epsilon_N X_{i_N}]$$

where $\epsilon_i = \pm 1$ and $(i_1, i_2, \ldots, i_N)$ is a permutation of $(1, 2, \ldots, N)$, the joint distribution of the signed sequential ranks is obtainable, in principle, from

$$(5.4) \qquad P(0 \leq \epsilon_1 X_{i_1} < \epsilon_2 X_{i_2} < \ldots < \epsilon_N X_{i_N}) = P(F, G, \epsilon)$$

where $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_N)$. In general (when $F \neq G$) the Y_i are not independent. For example if we are sampling from an unknown distribution $F(x)$ and we wish to detect a change in distribution to $F^a(x)$, $a > 1$ (a stochastically larger distribution) where $F(0) = 0$, we lose the property of independence. In this simple case signed sequential ranks and sequential ranks are equivalent and taking $N = 3$ with $m = 1$ we have

$$P(Y_1 = 1) = 1, \quad P(Y_2 = 1) = \frac{1}{1+a}, \quad P(Y_3 = 1) = \frac{1+3a}{2(1+a)(1+2a)}$$

and $P(Y_1 = 1, Y_2 = 1, Y_3 = 1) = \frac{1}{2(1+2a)}$. In general, for $a > 1$

$$\frac{1}{2(1+2a)} \neq \frac{1}{1+a} \frac{1+3a}{2(1+a)(1+2a)} = \frac{1}{2(1+2a)} \frac{1+3a}{(1+a)^2}$$

Since there are cases when the signed sequential ranks are independent we now determine the marginal distributions for signed sequential ranks in the case where a change in distribution from $F(x)$ to $G(x)$ occurs for arbitrary continuous distributions $F(x)$ and $G(x)$. Let $X_1, X_2, \ldots, X_N$ be independent random variables with X_i $1 \leq i \leq m$ distributed as $F(x)$ and X_i $m + 1 \leq i \leq N$ distributed as $G(x)$. Take $N = m + n$, and let Y_i be the signed sequential rank of X_i and $H_k(t)$ be the distribution function of the k^{th} order statistic from the set $\{|X_1|, |X_2|, \ldots, X_{N-1}\}$. It is enough to determine the distribution of Y_N. Using lemma 2.1 and $P(|X_1| \leq x) = F(x) - F(-x)$ for $x \geq 0$ we get

(5.5)

$$H_k(t) = \sum_{i=k}^{N-1} \sum_{j=0}^{i} \binom{m}{j}\binom{n-1}{i-j}(F(t) - F(-t))^j (1 - F(t) + F(-t))^{m-j}$$
$$\cdot (G(t) - G(-t))^{i-j} (1-G(t) + G(-t))^{n-i+j-1} \qquad t \geq 0$$

Now let Z_k be the k^{th} order statistic from $\{|X_1|, |X_2|, \ldots, |X_{N-1}|\}$. Then

$$P(Y_N = 1) = P(0 \leq X_N < Z_1) = E(G(Z_1)) - G(0)$$

$$= \int_0^\infty G(t)\, dH_1(t) - G(0)$$

$$= 1 - G(0) - \int_0^\infty H_1(t)\, dG(t) \quad .$$

Also for $2 \leq k \leq N - 1$

$$P(Y_N = k) = P(Z_{k-1} < X_N < Z_k) = E(G(Z_k)) - E(G(Z_{k-1})) \ .$$

Now $E(G(Z_k)) = \int_0^\infty G(t)\, dH_k(t) = 1 - \int_0^\infty H_k(t)\, dG(t)$ and

$$P(Y_N = k) = \int_0^\infty \{H_{k-1}(t) - H_k(t)\}\, dG(t)$$

$$= \sum_{j=0}^{k-1} \binom{m}{j}\binom{n-1}{k-1-j} \int_0^\infty (F(t) - F(-t))^j\, (1-F(t) + F(-t))^{m-j}$$

$$(G(t) - G(-t))^{k-1-j} \cdot (1-G(t) + G(-t))^{n-k+j}\, dG(t)$$

For $k = N$ we get $P(Y_N = N) = P(Z_{N-1} < X_N) = 1 - E(G(Z_{N-1}))$
$= \int_0^\infty H_{N-1}(t)\, dG(t)$. For negative values of Y_N we can calculate
$P(Y_N = -k) = P(Z_{k-1} < - X_N < Z_k)$ in a similar manner to obtain finally
for $2 \leq k \leq N - 1$

$$P(Y_N = 1) = 1 - G(0) - \int_0^\infty H_1(t)\, dG(t)$$

$$P(Y_N = -1) = G(0) + \int_0^\infty H_1(t)\, dG(-t)$$

$$P(Y_N = k) = \sum_{j=0}^{k-1} \binom{m}{j}\binom{n-1}{k-1-j} \int_0^\infty (F(t) - F(-t))^j\, (1-F(t) + F(-t))^{m-j}$$

(5.6)

$$\cdot\, (G(t) - G(-t))^{k-1-j}\, (1-G(t) + G(-t))^{n-k+j}\, dG(t)$$

$$P(Y_N = -k) = -\sum_{j=0}^{k-1} \binom{m}{j}\binom{n-1}{k-1-j} \int_0^\infty (F(t) - F(-t))^j\, (1-F(t) + F(-t))^{m-j}$$

$$\cdot\, (G(t) - G(-t))^{k-1-j}\, (1-G(t) + G(-t))^{n-k+j}\, dG(-t)$$

$$P(Y_N = N) = \int_0^\infty H_{N-1}(t)\, dG(t)$$

$$P(Y_N = -N) = -\int_0^\infty H_{N-1}(t)\, dG(-t)$$

The equations given in (5.6) can be written in one formula as

$$P(Y_N = \epsilon k) = \epsilon \sum_{j=0}^{k-1} \binom{m}{j}\binom{n-1}{k-1-j} \int_0^\infty (F(t) - F(-t))^j\, (1-F(t) + F(-t))^{m-j}$$

(5.7)

$$\cdot\, (G(t) - G(-t))^{k-1-j}\, (1-G(t) + G(-t))^{n-k+j}\, dG(\epsilon t)$$

where $\epsilon = \pm 1$ and $k = 1, 2, \ldots, N$. Verification that (5.7) reduces to (5.6) in the case $k = 1$ and $\epsilon = \pm 1$ can be accomplished through the following result

Lemma 5.1. $\sum_{i=0}^{N} \sum_{j=0}^{i} \binom{m}{j}\binom{n}{i-j} p^j(1-p)^{m-j} q^{i-j}(1-q)^{n-i+j} \equiv 1 \quad (N = m+n).$

Proof: Let $a_{ij} = \binom{m}{j}\binom{n}{i-j} p^j(1-p)^{m-j} q^{i-j} (1-q)^{n-i+j}$ and recall the convention of $\binom{a}{b} = 0$ if $b > a$. Instead of summing as indicated we sum along diagonals and get

$$\sum_{i=0}^{N} \sum_{j=0}^{i} a_{ij} = \sum_{\ell=0}^{N} \sum_{j=0}^{N-\ell} a_{\ell+j\ j}$$

$$= \sum_{\ell=0}^{N} \sum_{j=0}^{N-\ell} \binom{m}{j}\binom{n}{\ell} p^j(1-p)^{m-j} q^{\ell}(1-q)^{n-\ell}$$

$$= \sum_{\ell=0}^{N} \binom{n}{\ell} q^{\ell}(1-q)^{n-\ell} \sum_{j=0}^{N-\ell} \binom{m}{j} p^j(1-p)^{m-j}$$

$$= \sum_{\ell=0}^{n} \binom{n}{\ell} q^{\ell}(1-q)^{n-\ell} \sum_{j=0}^{N-\ell} \binom{m}{j} p^j(1-p)^{m-j} \quad \text{since} \quad \binom{n}{\ell} = 0 \quad \ell > n$$

Since $0 \leq \ell \leq n$ the upper limit in the second sum is $N \geq N - \ell \geq N - n = m$ implying $m \leq N - \ell$ and making the second sum always equal to 1. Using the binomial theorem a second time gives the result.

Letting $p = F(t) - F(-t)$, $q = G(t) - G(-t)$ we can write $H_1(t) = 1 - [1-p]^m [1-q]^{n-1}$ to complete the verification.

Using Lemma 5.1 and (5.7) we can compute the characteristic function for Y_N as

(5.8) $$\varphi(u) = E(e^{iuY_N}) =$$

$$\int_0^\infty e^{iu}[1-q(t) + q(t)\, e^{iu}]^{n-1}\, [1-p(t) + p(t)\, e^{iu}]^m\, dG(t)$$

$$- \int_0^\infty e^{-iu}[1-q(t) + q(t)\, e^{-iu}]^{n-1}\, [1-p(t) + p(t)\, e^{-iu}]^m\, dG(-t)$$

where $p(t) = F(t) - F(-t)$ and $q(t) = G(t) - G(-t)$.

Differentiating (5.8) and setting $u = 0$ we get

(5.9) $$E(Y_N) = \int_0^\infty \{1 + (n-1)\, q(t) + mp(t)\}\, d\{G(t) + G(-t)\}$$

(5.10)
$$E(Y_N^2) = 1 + (n-1)\,\frac{(2n+5)}{6} + \int_0^\infty \{3mp(t) + 2m(n-1)\, q(t)\, p(t) + m(m-1)\, p^2(t)\}\, dq(t)$$

The marginal distribution of Y_N, equation (5.7), holds for arbitrary continuous distribution functions F and G and thus (5.8), (5.9) and (5.10) are the general expressions for the characteristic function, mean and second moment of the Y_N. Thus to generalize (5.2) to arbitrary continuous distributions F we let $F = G$ in (5.9) and (5.10) and we get

(5.11) $$E(Y_N) = (N-1)\left(\frac{1}{2} - G^2(0) - \int_0^\infty G(-t)\, dG(t)\right) + 1 - 2G(0)$$

(5.12) $$E(Y_N^2) = \frac{N^2}{3} + \frac{N}{2} + \frac{1}{6}$$

6. <u>An Application of Signed Sequential Ranking to Process Control</u>. As stated in the introduction, in the process control problem we wish to determine a procedure which will determine when a given sequence of random variables changes from being distributed according to $F(x)$ to a different distribution $G(x)$. In particular we will consider the case where $F(x)$ satisfies the condition of Theorem 5.1 and changes to $G(x)$ which also satisfies the condition. Inasmuch as the distribution of the signed sequential ranks depends on the parameter $F(0)$ we will of course require $G(0) \neq F(0)$. The procedure described in this section is still applicable to cases where G does not satisfy the condition in Theorem 5.1 but we do not have exact results in such instances. However empirical results are presented at the end of this section bearing on the effectiveness of the procedure for special cases.

Let $X_1, X_2, \ldots$ be a sequence of independent random variables (observations on a process) with common distribution function $F(x)$ where for all $x \geq 0$ the condition $F(-x) = F(0)[1-F(x) + F(-x)]$ holds, and let $Y_1, Y_2, \ldots$ be the corresponding signed sequential ranks. We define the cumulative sums $S_n = Z_1 + Z_2 + \ldots + Z_n$ where $Z_i = Y_i/i$. Since the condition in Theorem 5.1 is satisfied the Z_i are independent and

$$(6.1) \quad P(Z_n = t) = \begin{cases} \frac{1-F(0)}{n} & t = \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1 \\ \\ \frac{F(0)}{n} & t = -\frac{1}{n}, -\frac{2}{n}, \ldots, -\frac{n-1}{n}, -1 . \end{cases}$$

Some easy computations yield

(6.2)

$$E(Z_n) = \frac{1-2F(0)}{2} \cdot \left(1 + \frac{1}{n}\right)$$

$$\mathrm{Var}(Z_n) = \left\{\frac{1}{3} - \left(\frac{1-2F(0)}{2}\right)^2\right\} \left(1 + \frac{1}{n}\right) + \left\{\frac{1}{6} - \left(\frac{1-2F(0)}{2}\right)^2\right\} \left(\frac{1}{n} + \frac{1}{n^2}\right)$$

$$E(Z_n^2) = \frac{1}{3} + \frac{1}{2n} + \frac{1}{6n^2}$$

$$E(S_n) = \frac{1-2F(0)}{2} \left(n + \sum_{i=1}^{n} 1/i\right)$$

$$\mathrm{Var}(S_r) = \left\{\frac{1}{3} - \left(\frac{1-2F(0)}{2}\right)^2\right\} \left(n + \sum_{i=1}^{n} \frac{1}{i}\right) + \left\{\frac{1}{6} - \left(\frac{1-2F(0)}{2}\right)^2\right\} \sum_{i=1}^{n} \frac{1}{i} + \frac{1}{i^2}$$

Although tedious, the distribution $P(S_n = t)$ can be computed exactly. For example

$$P(S_1 = t) = P(Z_1 = t) = \begin{cases} 1-F(0) & t = 1 \\ F(0) & t = -1 \end{cases}$$

$$P(S_2 = t) = P(S_1 + Z_2 = t) = \begin{cases} \frac{(1-F(0))^2}{2} & t = \frac{3}{2}, 2 \\ \frac{(1-F(0))\, F(0)}{2} & t = \frac{1}{2} \\ (1-F(0))\, F(0) & t = 0 \\ \frac{(1-F(0))\, F(0)}{2} & t = -\frac{1}{2} \\ \frac{(F(0))^2}{2} & t = -\frac{3}{2}, -2 \end{cases}$$

and in general

(6.3) $$P(S_n = t) = P(S_{n-1} = t - Z_n) = \sum_x P(S_{n-1} = t-x)\, P(Z_n = x)$$

where x ranges over $-1, -\frac{n-1}{n}, \ldots, -\frac{1}{n}, \frac{1}{n}, \ldots, \frac{n-1}{n}, 1$.

The procedure we will propose will stop the process whenever S_n does not lie in some fixed open interval (b, a) where

$-\infty < b < 0 < a < \infty$. In order to determine the operating characteristics of such a procedure such as the average number of observations until the process is stopped we must compute

$$(6.4) \quad P(N = n) = P(b < S_i < a,\ i = 1, 2, \ldots, n-1,\ S_n \notin (b,a))$$

N being the smallest integral value for which S_N does not lie in the open interval (b, a). Then $E(N) = \sum_{n=1}^{\infty} n\, P(N = n)$ gives the average number of observations as a function of a, b and $F(0)$. In order to compute the probability of reaching the boundaries b and a for the first time at time n the following procedure may be used. We define $F_1(x) = P(S_1 \leq x)$, $F_2(x) = P(S_2 \leq x,\ b < S_1 < a)$ and in general

$$(6.5) \qquad F_n(x) = P(S_n \leq x,\ b < S_i < a \quad i = 1, 2, \ldots, n-1)$$

It follows that $F_2(x) = P(Z_2 \leq x-S_1,\ b < S_1 < a) = \int_b^a F_{Z_2}(x-y)\, dF_1(y)$ and in general

$$(6.6) \qquad F_n(x) = \int_b^a F_{Z_n}(x-y)\, dF_{n-1}(y)\ .$$

The probability of reaching boundary a for the first time at n is $F_n(\infty) - F_n(a)$ and the probability of reaching boundary b for the first time at n is $F_n(b) - F_n(-\infty)$. Using these probabilities we can also calculate $E(N)$.

Computations of the probability functions in (6.6) could be carried out and the computational burden lessened somewhat by noting that for large values of n, the Z_n tend to become identically distributed. We now consider some approximations to $E(N)$ using some results from sequential analysis.

Using (5.1) the characteristic function of Z_n is given by

$$(6.7) \qquad \varphi_n(u) = \frac{1-F(0)}{n} \frac{e^{i\,u/n} - e^{iu(1+1/n)}}{1-e^{i\,u/n}} + \frac{F(0)}{n} \frac{e^{-i\,u/n} - e^{-iu(1+1/n)}}{1-e^{-i\,u/n}}$$

and using limited expansions of exponentials we have

$$(6.8) \qquad \varphi(u) = \lim_{n \to \infty} \varphi_n(u) = (1-F(0)) \frac{e^{iu} - 1}{iu} + F(0) \frac{1-e^{-iu}}{iu}$$

which is the characteristic function of a random variable with density

$$(6.9) \qquad f(x) = \begin{cases} F(0) & -1 < x < 0 \\ 1-F(0) & 0 \leq x < 1 \end{cases}$$

For large values of n, Z_n has approximately the density of (6.9). The moment generating function associated with (6.9) is

$$(6.10) \qquad M(t) = F(0) \frac{1-e^{-t}}{t} + (1-F(0)) \frac{e^t - 1}{t}$$

which exists for all real values of t. As an approximation we will use $E(Z_n) = \frac{1-2F(0)}{2}$. In the cumulative sums $S_n = Z_1 + Z_2 + \ldots + Z_n$ the Z_i are independent and as noted, not identically distributed. However if we disregard the first few signed sequential ranks and start later in the sequence the approximation to identically distributed random variables improves. As before, we take N to be the smallest integral value for which S_N does not lie in (b, a). We use the results of Wald [5] in the sequel.

Consider first the case where $F(0) = 1/2$ (F is symmetric about 0)
Here $E(Z_n) = 0$ and using (3.8) of [5]

$$E(N) = \frac{E(S_N^2)}{E(Z_n^2)} = \frac{-ab}{E(Z_n^2)} = -3ab$$

When $F(0) \neq 1/2$, $E(Z_n) \neq 0$ and we can use $E(S_N) = E(Z_n) \cdot E(N)$ and the approximation $E(S_N) = aP(S_N \geq a) + b(1-P(S_N \geq a))$ to get

$$(6.11) \qquad E(N) = \begin{cases} -3ab & F(0) = 1/2 \\ \\ \dfrac{2b + 2(a-b)P(S_N \geq a)}{1-2F(0)} & F(0) \neq 1/2 \end{cases}$$

Let h be the non zero root of $M(t) = 1$. A further approximation gives

$$(6.12) \qquad P(S_N \geq a) = \frac{1 - e^{bh}}{e^{ah} - e^{bh}}$$

where of course h depends on the value of $F(0)$. Setting $M(t) = 1$ we get $F(0) = \dfrac{1 + t - e^t}{2 - e^t - e^{-t}}$ which must be solved for t. Each solution corresponding to a fixed value of $F(0)$ is a value for h in (6.12) yielding, in turn, a solution to (6.11).

Let $g(t) = \dfrac{1 + t - e^t}{2 - e^t - e^{-t}}$. Then $g'(t) = \dfrac{4(1-\cosh t) + 2t \sinh t}{4(1-\cosh t)^2}$

and considering the numerator $a(t) = 4(1-\cosh t) + 2t \sinh t$ we find $a'(t) = \sinh t + 2t \cosh t$ and moreover

BLANK PAGE

$$a'(t) < 0 \qquad t < 0$$

$$a'(0) = 0$$

$$a'(t) > 0 \qquad t > 0$$

Thus $a(t) \geq 0$, making $g'(t) \geq 0$ and $g(t)$ is monotone increasing in t. As $F(0)$ increases from 0 to 1 the solution to $F(0) = \frac{1 + t - e^t}{2 - e^t - e^{-t}}$ say $h(F(0))$ increases from $-\infty$ to ∞. Notice that

$$\lim_{t \to \infty} g(t) = \lim_{t \to \infty} \frac{e^{-t} + te^{-t} - 1}{2e^{-t} - 1 - e^{-2t}} = 1$$

$$\lim_{t \to -\infty} g(t) = \lim_{t \to -\infty} \frac{e^t + te^t - e^{2t}}{2e^t - e^{2t} - 1} = 0 \quad .$$

Now for $h = h(F(0))$ increasing, $P(S_N \geq a) \doteq \frac{1 - e^{bh}}{e^{ah} - e^{bh}}$ is decreasing in h since for

$$g(h) = \frac{1 - e^{bh}}{e^{ah} - e^{bh}}$$

we have

$$g'(h) = \frac{(a-b)\, e^{(a+b)h} - [ae^{ah} - be^{bh}]}{(e^{ah} - e^{bh})^2}$$

and considering the numerator after factoring out $e^{(a+b)h}$ we have to show

$$a-b - ae^{-bh} + be^{-ah} \leq 0 \qquad \text{for all } h.$$

Writing $\alpha = a/a-b$, $\beta = -b/a-b$ we have $\alpha + \beta = 1$, $\alpha, \beta > 0$ and we must show that $1 \leq \alpha e^{\beta(a-b)h} + \beta e^{-\alpha(a-b)h}$. Let $f(h) = \alpha e^{\beta(a-b)h} + \beta e^{-\alpha(a-b)h}$ and notice that $f(0) = 1$, $f'(0) = 0$ with $f''(h) > 0$ since

$$f'(h) = \alpha\beta(a-b)\, e^{\beta(a-b)h} - \alpha\beta(a-b)\, e^{-\alpha(a-b)h}$$

$$f''(h) = \alpha\, \beta^2(a-b)^2\, e^{\beta(a-b)h} + \alpha^2\beta(a-b)^2\, e^{-\alpha(a-b)h}$$

Thus $f(h)$ attains its minimum value at $h = 0$. For increasing values of $F(0)$ the corresponding values of $h = h(F(0))$ increase and $P(S_N \geq a)$ decreases. For $F(0) \neq 1/2$ we have

$$(6.13) \qquad E(N) = \frac{2b + 2(a-b)\left(\dfrac{1 - e^{bh}}{e^{ah} - e^{bh}}\right)}{1-2F(0)}$$

In particular taking $b = -a$, $h \neq 0$ we have

$$(6.14) \qquad E(N) = \frac{2a(1-e^{-ah} - \sinh(ah))(1-\cosh(h))}{\sinh(ah)(\sinh(h) - h)}$$

For $h = 0$, $E(N) = 3a^2$ and (6.14) is plotted in Figure 1 for selected values of a. $g(t) = \dfrac{1 + t - e^t}{2 - e^t - e^{-t}}$ is shown in Figure 2. $E(N)$ is plotted against $F(0)$ in Figure 3.

Suppose now that a process is observed according to some measurable characteristic and we have a sequence $X_1, X_2, \ldots$, distributed according to $F(x)$ where $F(x)$ satisfies the condition of Theorem 5.1 and moreover we assume $F(0) = 1/2$. If we set boundaries $(-a, a)$ $a > 0$

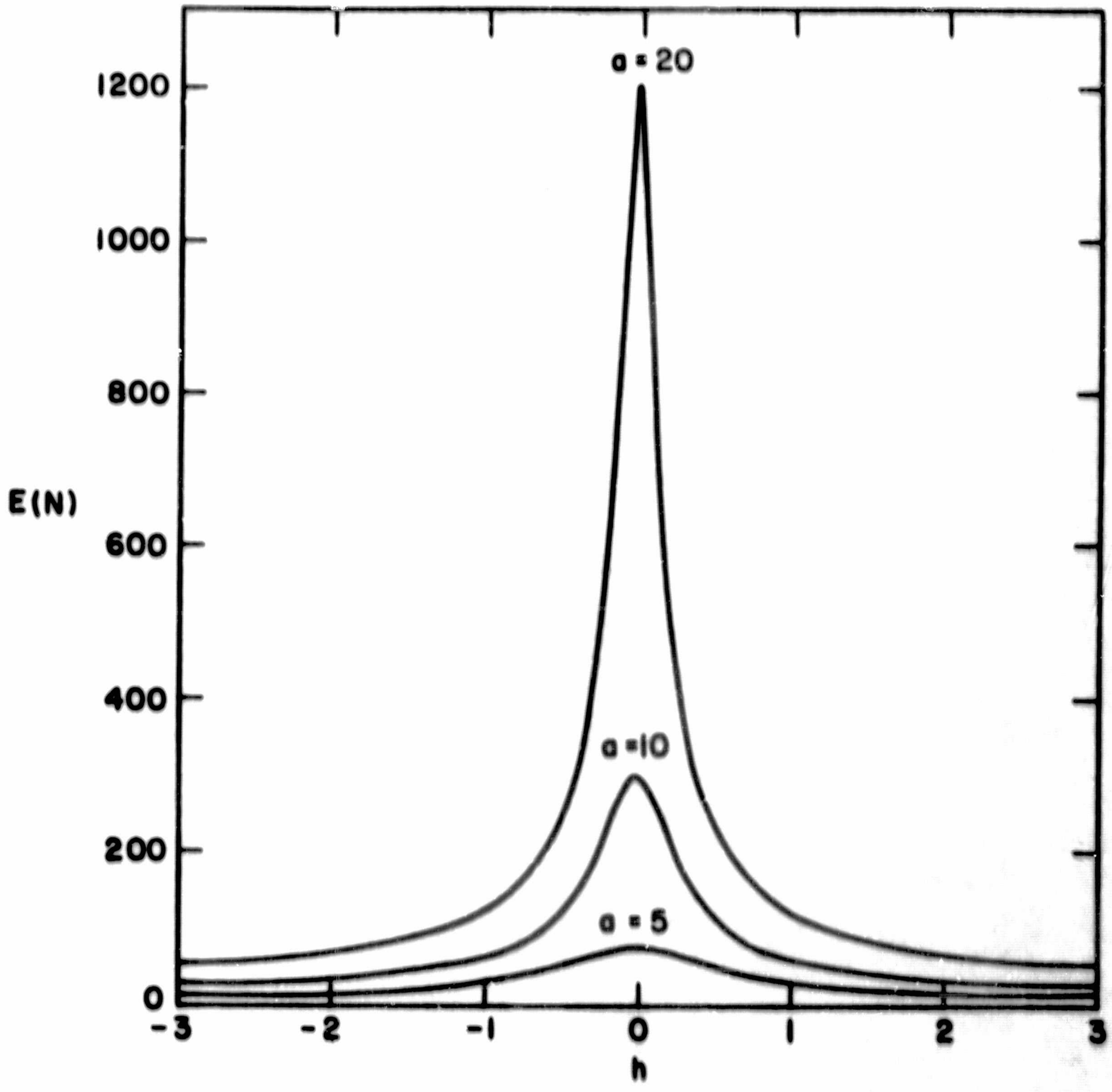

Figure 1

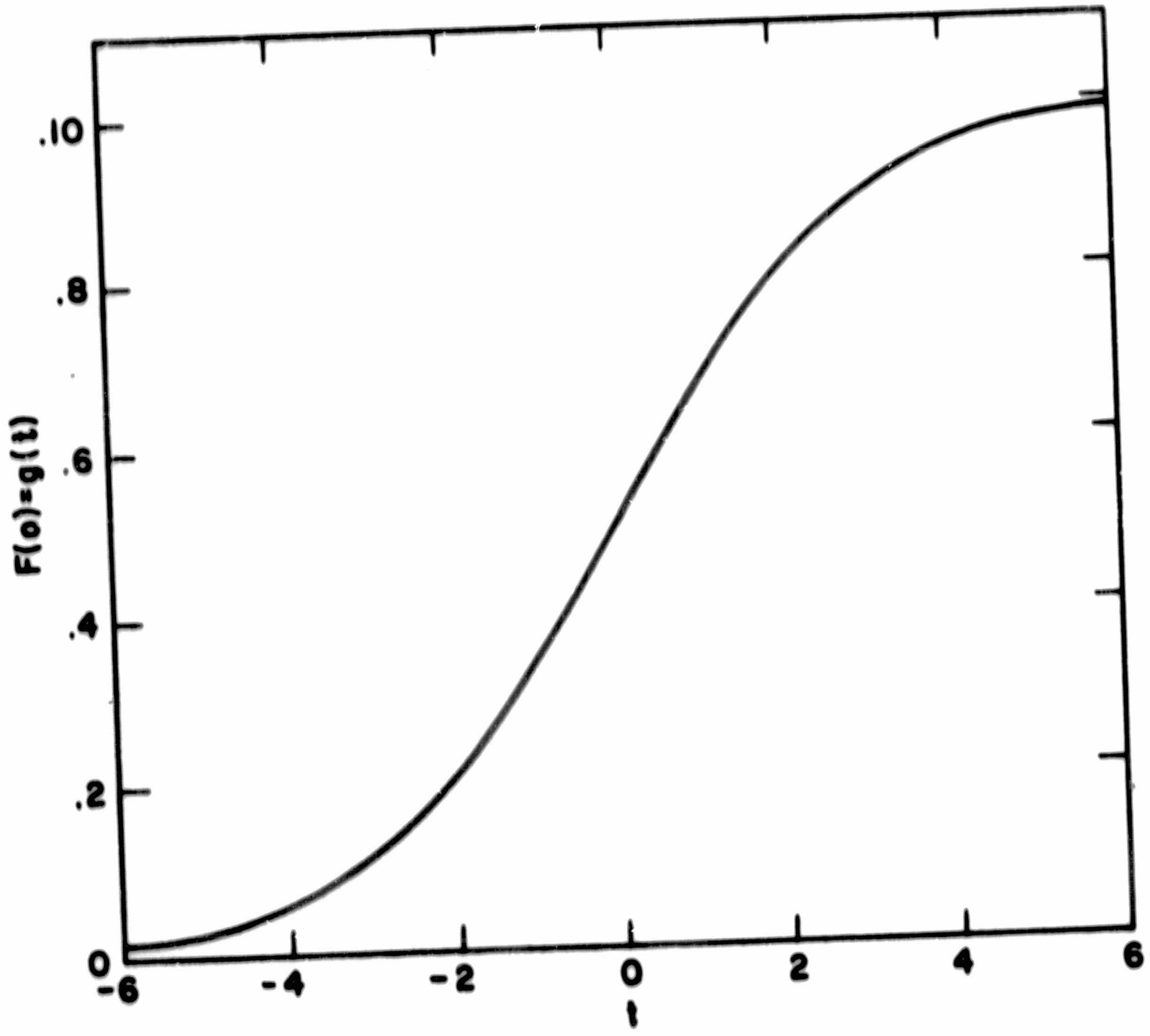

Figure 2

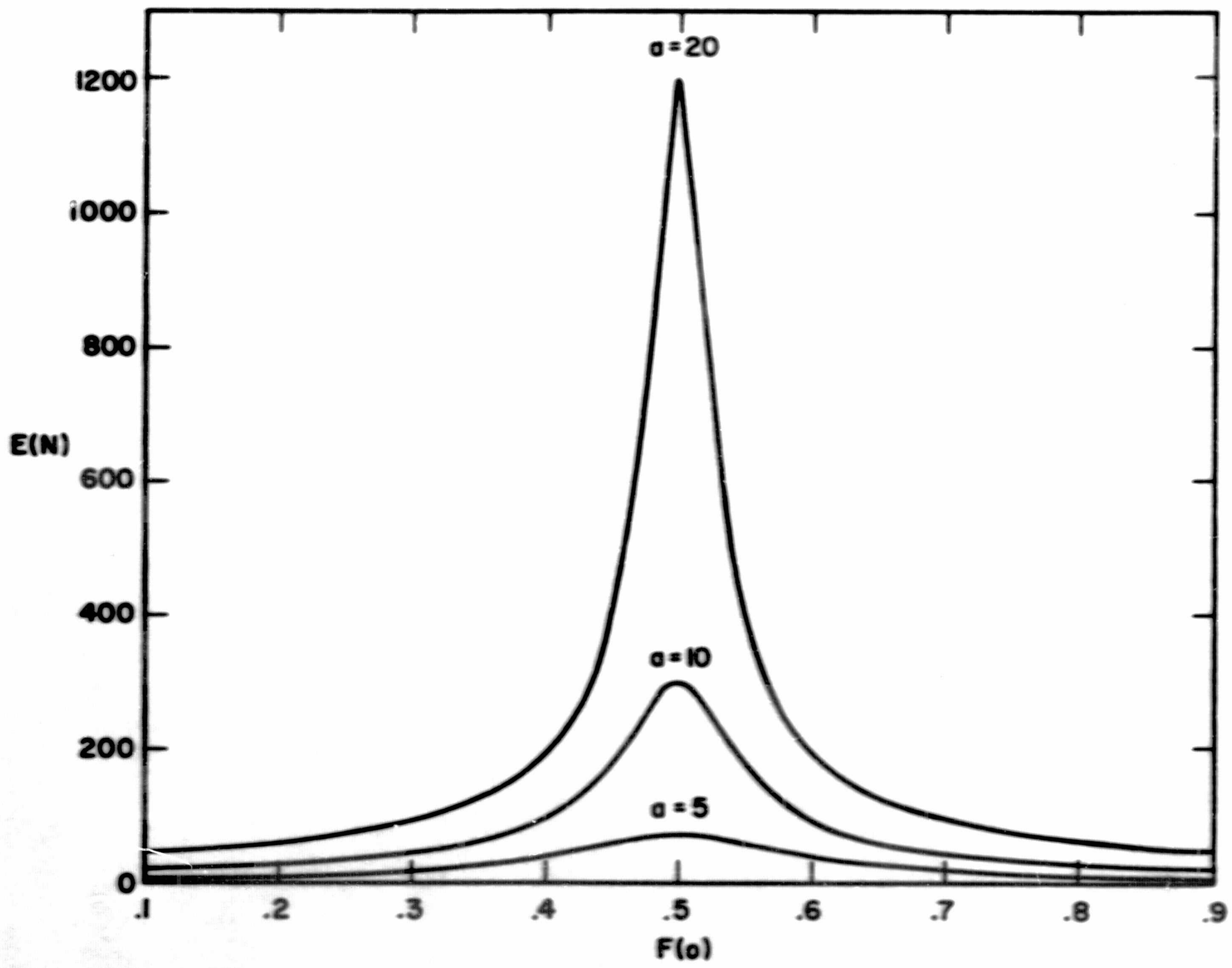

Figure 3

and use the rule which requires us to stop the process when $S_n \notin (-a, a)$ for the first time we can expect to continue for $3a^2$ observations before stopping. However if the process is such that $F(0) \neq 1/2$ we will stop the process in the reduced average time as given in Figure 1. Similar computations can be made for arbitrary intevals (b, a) using (6.13). However, in the process control problem we wish to detect when a change takes place in the distribution of the basic random variables. We have seen that when a change takes place from $F(x)$ to $G(x)$ at some point in the sequence, the signed sequential ranks are no longer independent in general. Suppose the change is to a distribution $G(x)$ such that the condition of Theorem 5.1 is still satisfied and the change takes place at time m. Intuitively, one might feel that for large values of n the distribution of the $m + n^{th}$ signed sequential rank would depend very little on F and m. This being so we could assume the sequence $\{Z_i\}$ to be independent for the purpose of determing the expected number of observations until the process is stopped. For example suppose we take (b, a) as the continuation interval and denote (6.13) by $E(a, b, F(0))$. Given that $S_m = x$, $b < x < a$ the expected number of additional observations under $G(x)$ is

(6.15)
$$E(N|S_m = x) = E(a-x, b-x, G(0))$$

$$= \frac{2(b-x)}{1-2G(0)} + \frac{2(a-b)}{1-2\ G(0)} \cdot \frac{e^{xh} - e^{bh}}{e^{ah} - e^{bh}}$$

The conditional distribution for S_m is

$$P(S_m \leq x \mid b < S_m < a,\ b < x < a) = \frac{P(b < S_m \leq x)}{P(b < S_m < a)}$$

$$(6.16) \qquad = \begin{cases} 1 & x \geq a \\ \dfrac{F_{S_m}(x) - F_{S_m}(b)}{F_{S_m}(a) - F_{S_m}(b)} & b < x < a \\ 0 & x \leq b \end{cases}$$

and the unconditional total expected number of observations is given by

$$(6.17) \qquad E(N, m, G(0)) = m + \frac{1}{F_{S_m}(a) - F_{S_m}(b)} \int_b^a E(N \mid S_m = x)\, dF_{S_m}(x)$$

To lend some support to the statement that for large values of n, the distribution of Z_{m+n} does not depend too much on m and the distribution of $X_1, X_2, \ldots, X_m$ (and thus could be taken as $G(x)$ to justify (6.15)) we examine its characteristic function as $n \to \infty$. We have

$$\lim_{n \to \infty} \varphi_n(u) = \lim_{n \to \infty} E\left(e^{iuZ_{n+n}}\right)$$

$$= \lim_{n \to \infty} \int_0^\infty e^{i\frac{u}{m+n}} \left(1-q(t) + q(t)\, e^{i\frac{u}{m+n}}\right)^{n-1}$$

$$\left(1-p(t) + p(t)\, e^{i\frac{u}{m+n}}\right)^{m} dG(t)$$

$$- \lim_{n \to \infty} \int_0^\infty e^{-i \frac{u}{m+n}} \left(1-q(t) + q(t)\, e^{-i \frac{u}{m+n}}\right)^{n-1}$$

$$\left(1-p(t) + p(t)\, e^{-i \frac{u}{m+n}}\right)^{m} dG(-t)$$

$$= \int_0^\infty \lim_{n \to \infty} \left(1-q(t) + q(t)\, e^{i \frac{u}{m+n}}\right)^{n-1} dG(t)$$

$$- \int_0^\infty \lim_{n \to \infty} \left(1-q(t) + q(t)\, e^{-i \frac{u}{m+n}}\right)^{n-1} dG(-t)$$

$$= \int_0^\infty e^{iq(t)u}\, dG(t) - \int_0^\infty e^{-iq(t)u}\, dG(-t)$$

Also, since $q(t) = \frac{G(t)}{1-G(0)} - \frac{G(0)}{1-G(0)}$ and $-q(t) = \frac{G(-t)}{G(0)} - 1$

we have

$$\lim_{n \to \infty} \varphi_n(u) = G(0)\, \frac{1-e^{-iu}}{iu} + (1-G(0))\, \frac{e^{iu} - 1}{iu}$$

corresponding to (6.10).

We now consider a case where we have a change from a distribution satisfying the condition in Theorem 5.1 to another such distribution. Imagine a production process where some dimension is measured on the items being produced. Let these measurements be $X_1, X_2, \ldots$ assumed to be independent and identically distributed as $F(x)$. Each item is subject to inspection and if $X_i \leq 0$ the item is removed from the production line with probability p. The result is a new sequence say $C_1, C_2, \ldots$ and we call this random censoring and $\{C_i\}$ the censored sequence. The distribution of C_i can be found by

$$P(C_i \leq t) = \sum_{j=1}^{\infty} P(C_i \leq t \mid C_i = X_{i+j-1})\, P(X_{i+j-1} = C_i)$$

$$= \sum_{j=1}^{\infty} P(X_{i+j-1} \leq t \mid C_i = X_{i+j-1})\, P(X_{i+j-1} = C_i)$$

$$= P(X_1 \leq t \mid X_1 = C_1) \sum_{j=1}^{\infty} P(C_i = X_{i+j-1})$$

$$= P(X_1 \leq t \mid X_1 = C_1)$$

For $t \leq 0$

$$P(C_i \leq t) = P(X_1 \leq t \mid X_1 = C_1)$$

$$= \frac{P(X_1 \leq t,\ X_1 \leq 0,\ X_1 \text{ not censored})}{P(C_1 = X_1)}$$

$$= \frac{P(X_1 \leq t,\ X_1 \text{ not censored})}{1-pF(0)}$$

$$= (1-p)\, F(t)/1-pF(0) \quad .$$

For $t > 0$

$$P(C_i \leq t) = \frac{P(X_1 \leq t,\ X_1 \leq 0,\ X_1 \text{ not censored}) + P(X_1 \leq t,\ 0 < X_1)}{1-pF(0)}$$

$$= \frac{F(t) - pF(0)}{1-pF(0)} \quad .$$

For random censoring when $X_1 \leq 0$ we have

$$(6.9) \qquad P(C_1 \leq t) = \begin{cases} \dfrac{1-p}{1-pF(0)} F(t) & t \leq 0 \\ \\ \dfrac{F(t) - pF(0)}{1-pF(0)} & t > 0 \end{cases}$$

In a similar way if we censor with probability p when $X_1 > 0$ we have

$$(6.10) \qquad P(C_1 \leq t) = \begin{cases} \dfrac{F(t)}{1-p + pF(0)} & t \leq 0 \\ \\ \dfrac{(1-p)\ F(t) + pF(0)}{1-p + pF(0)} & t > 0 \end{cases}$$

Suppose now that the symmetry condition $F(-t) = F(0)[1-F(t) + F(-t)]$ holds for all $t \geq 0$. In the case of random censoring for $X_1 \leq 0$ we have for $t \geq 0$

$$G(t) = P(C_1 \leq t)$$

$$F(t) = [1-pF(0)]\ G(t) + pF(0)$$

$$F(-t) = \frac{1-pF(0)}{1-p} G(-t)$$

$$F(0) = \frac{1-pF(0)}{1-p} G(0)$$

Using these relations it follows that

$$G(0)[1-G(t) + G(-t)] = \frac{(1-p)\ F(0)}{1-p\ F(0)}\ \frac{1-F(t) + (1-p)\ F(-t)}{1-p\ F(0)}$$

and from the symmetry condition on F we get

$$G(0)[1-G(t) + G(-t)] = G(-t) \quad \text{for all} \quad t \geq 0 ,$$

with a change from $F(0)$ to $G(0) = \frac{(1-p)\,F(0)}{1-p\,F(0)}$. In particular for $F(0) = 1/2$, $G(0) = (1-p)/2-p$. A similar calculation for censoring when $X_1 > 0$ shows that the symmetry condition holds for $G(t)$ and $G(0) = F(0)/1-p + pF(0)$. For $F(0) = 1/2$, $G(0) = 1/2 + p$.

We shall now compare the expected number of observations needed to stop a process subject to random sampling using a Shewhart type control chart with the expected number needed using the procedure described above. Consider a sequence of independent observations $X_1, X_2 \ldots$ with common continuous symmetric distribution $F(x)$. Subjecting the X_i to random censoring when $X_i \leq 0$ we get from (6.9) the distribution of the censored observations $C_1, C_2, \ldots$ as

$$(6.11) \qquad P(C_1 \leq t) = \begin{cases} \frac{1-p}{2-p}\, 2F(t) & t \leq 0 \\ \frac{2F(t) - p}{2-p} & t > 0 \end{cases}$$

We assume here that when $p = 0$ the process is in control and that when the process starts some fixed value of p, $0 \leq p \leq 1$ is in effect. If $p > 0$ we want to stop the process as quickly as possible. We consider three procedures:

procedure 1 - when $C_1 > b > 0$ for the first time, stop the process

procedure 2 - when $|C_1| > b > 0$ for the first time, stop the process

procedure 3 - when $|S_n| > a > 0$ for the first time, stop the process

Procedures 1 and 2 are Shewhart type procedures and b is usually taken so that the probability of stopping at a particular stage is small when $p = 0$. Procedure 3 is the signed sequential rank procedure previously described in this section. Define $p_1 = P(C_1 > b)$ and $p_2 = P(|C_1| > b)$ assuming $p = 0$. For each procedure the probability of falling outside the control limit for the first time at the n^{th} observation is

$$p_i(1-p_i)^{n-1} \qquad i = 1, 2$$

and $E_1(N) = 1/p_1$, $E_2(N) = 1/p_2$ are the expected number of observations taken before stopping. $E_3(N) = 3a^2$ and setting $E_1(N) = E_2(N) = E_3(N)$ we get

$$p_1 = 1-F(b) = F(-b) = 1/3a^2$$

$$p_2 = 1-F(b) - F(-b) = 2F(-G) = 1/3a^2 \ .$$

For $p > 0$ $p_1' = P(C_1 > b) = 1-P(C_1 \leq b) = \frac{2F(-b)}{2-p} = \frac{2}{(2-p)\ 3a^2}$ and

$$p_2' = P(C_1 > b) + P(C_1 < -b) = \frac{2F(-b)}{2-p} + \frac{1-p}{2-p}\, 2\ F(-b) = 2F(-b) = 1/3a^2$$

Thus for $p > 0$

$$E_1(N) = 1/p_1' = \frac{(2-p)\ 3a^2}{2}$$

$$E_2(N) = 1/p_2' = 3a^2$$

$$E_3(N) = \frac{-2a + 4a\left(\frac{e^{ah} - 1}{e^{2ah} - 1}\right)}{p/2 - p}$$

and notice that since $P(C_1 \leq 0) = \frac{1-p}{2-p} < 1/2$, it follows that $h < 0$.

$E_1(N)$ and $E_2(N)$ increase quadratically with a and E_3 is essentially linear in a. For example

	a = 10				a = 20			
p	1	3/4	1/2	1/4	1	3/4	1/2	1/4
$E_1(N)$	150	187.5	225	262.5	600	750	900	1050
$E_3(N)$	20	33.3	60	140	40	66.6	120	280

and procedure 3 is insensitive to values of $p \geq 0$. The values of h corresponding to $p = 1, 3/4, 1/2, 1/4$ are $-\infty$, -2.2, -.9, -.5 respectively.

The following tabulated results were obtained empirically to determine the effect of translation of the mean of the observations. We considered normal observations with mean μ and variance 1 and stopped sampling when $|S_N| > a$ for the first time where

$$S_n = \sum_{i=1}^{n} Z_i = \sum_{i=1}^{n} \frac{Y_i}{i}$$

and Y_i is the signed sequential rank of X_i, $X_i \sim \mathcal{N}(\mu,1)$. For each parameter pair (a,μ) twenty trials were performed except for $\mu = .1$, .2, .3 where fifty trials were used. Sample averages, sample variances and sample standard deviations for termination time N are given.

	a = 10			a = 20		
μ	$\bar{N}$	s^2	s	$\bar{N}$	s^2	s
.1	180.78	13449.27	115.97	364.56	31279.43	176.85
.2	101.78	3306.46	57.50	179.04	3345.18	57.83
.3	69.95	710.12	26.64	130.40	1437.18	37.91
.4	52.55	324.99	18.02	108.65	1160.87	34.07
.5	42.25	139.14	11.79	77.25	367.14	19.16
.6	39.60	121.41	11.01	72.70	171.69	13.10
.7	36.55	128.05	11.31	67.45	130.26	11.41
.8	28.70	31.48	5.61	62.05	115.31	10.73
.9	28.00	38.31	6.18	53.55	67.31	8.20
1.0	28.80	29.64	5.44	52.25	39.77	6.30
1.5	23.40	7.93	2.81	44.00	26.94	5.19
2.0	22.40	4.98	2.23	42.45	12.05	3.47
2.5	20.90	5.25	2.29	41.55	9.20	3.03
3.0	21.65	7.60	2.75	40.95	13.83	3.72

7. Summary and Conclusions. We remarked in the introduction on the paucity of nonparametric sequential procedures, particularly those based on ranks of observations. The author feels that the absence of a natural way of assigning ranks to observations, as the observations are taken, without reranking, was a significant cause for the lack of such procedures. The sequential ranking schemes defined and studied in this dissertation provide us with methods whereby ranks may be assigned in just such a manner.

In order to use the methods of sequential parametric hypothesis testing (Wald's sequential probability ratio test) in our nonparametric setting, we must replace the sequence of observations $X_1, X_2, \ldots$ by a sequence of ranks $R_1, R_2, \ldots$ and base the test on the probability ratio of the ranks. This can be done by the sequential ranking scheme defined in Section 3. One basic nonparametric problem is the two-sample problem where we must decide whether or not an X- population and a Y- population have the same probability distribution. This problem was treated in Section 4 in the special case where the alternatives are of the form proposed by Lehmann [1]. However the method proposed in Section 4 is general in the sense that in order to carry out the test one must only be able to compute $P(U_1 \leq U_2 \leq \cdots \leq U_N)$ where the U's are X's and Y's. In general this computation is difficult, but for special alternatives where the computation is feasible, the method in Section 4 applies directly.

Notice that in the finite sample size problem nothing is sacrificed by ranking sequentially (Theorem 3.1) instead of using ordinary ranks. In fact a little is gained inasmuch as the sequential ranks may be viewed as a transformation of the <u>dependent</u> ordinary ranks into the <u>independent</u> sequential ranks.

Merely ranking observations tells us nothing of their location, except relative to each other. In order to take into account the location of each observation relative to the origin as well as its size (absolute value) and relative location, the method of signed sequential ranking was devised. Contrary to sequential ranks, signed sequential ranks obtained from independent identically distributed observations are not independent in general. A sufficient condition on the distribution of the observations is given in Theorem 5.1 to insure that the signed sequential ranks will be independent. In the process control problem we used signed sequential ranks of observations whose distributions satisfied this condition. This simplified the calculations since sums of independent random variables were involved in the analysis.

The methods of sequential ranking and signed sequential ranking proposed in this dissertation are new, as far as the author can determine, and provide a natural way of assigning ranks to observations which fits into the theory of sequential analysis (hypothesis testing) and sequential procedures (process control). All the attendant distribution theory results are new and the condition of Theorem 5.1 which insures the independence of signed sequential ranks is the only one known to the author.

There are many areas for further investigation suggested by this research. In the sequential probability ratio test of Section 4 we did not use the sequential ranks explicitly (except for Z in equation (4.2)) in the definition of the probability ratio S_N. S_N can be written in terms of $(Z_1, Z_2, \ldots, Z_N)$, the sequential ranks, but the expression is quite complicated and it is much more convenient to use (4.1) and (4.2) which incorporate the most recent sequential rank only. Thus the behavior of S_N was obtained by reference to $A_1, A_2, \ldots A_N$. More general results are needed as to the probability of termination of $P_1(Z^N)/P_0(Z^N)$ for alternatives other than Lehmann alternatives. This is necessary because under the alternative hypothesis the sequential ranks are not independent generally and the conservative approximations $A = 1-\beta/1-\alpha$ remain valid for successive dependent observations when the probability is one that the procedure will ultimately terminate.

A second area for further study is the evaluation of the rule given in Section 6 for process control problems when changes from F to G are not of the form presented (e.g. $G(x) = F(x \pm \Delta)\ \Delta > 0$). Also there are other ad hoc rules which could be proposed using signed sequential ranks (or sequential ranks) in process control problems.

REFERENCES

[1] E. L. Lehmann, "The Power of Rank Tests," Ann. Math. Stat. Vol. 24, (1953), pp. 23-43.

[2] Ole Barndorff-Nielsen, "On the Limit Behavior of Extreme Order-Statistics," Ann. Math. Stat., Vol. 34, (1963), pp. 992-1002.

[3] W. Hoeffding, "Optimum Non Parametric Tests," Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, (1951) pp. 83-92.

[4] F. Wilcoxon, L. J. Rhodes and R. A. Bradley," Two Sequential Two-Sample Grouped Rank Tests with Applications," Biometrics, Vol. 19, No. 1 (1963), pp. 58-84.

[5] A. Wald, "Differentiation Under the Expectation Sign in the Fundamental Identity of Sequential Analysis." Ann. Math. Stat. Vol. 17, (1946), pp. 493-497.

[6] A. Wald, "Sequential Analysis," John Wiley and Son, New York, 1947.

[7] E. S. Page, "Continuous Inspection Schemes" Biometrika, Vol. 41, (1954), pp. 100-115.

[8] H. F. Dodge and H. G. Romig, "A Method of Sampling Inspection," Bell System Tech. Jour. Vol. 8, (1929), pp. 613-631.

REFERENCES

[9] A. Wald and J. Wolfowitz, "Optimum Character of the Sequential Probability Ratio Test," Ann. Math. Stat., Vol. 19, (1948), pp. 326-339.

[10] Personal Communication from J. Sethuraman in March 1965.