The Annals of Statistics
1981, Vol. 9 No. 3, 544–554

A NONPARAMETRIC CONTROL CHART FOR DETECTING SMALL DISORDERS[1]

BY P. K. BHATTACHARYA AND DARGAN FRIERSON, JR.

University of Arizona and University of North Carolina at Wilmington

We consider sequential observation of independent random variables $X_1, \cdots, X_N$ whose distribution changes from F to G after the first $[N\theta]$ variables. The object is to detect the unknown change-point quickly without too many false alarms. A nonparametric control chart based on partial weighted sums of sequential ranks is proposed. It is shown that if the change from F to G is small, then as $N \to \infty$, the appropriately scaled and linearly interpolated graph of partial rank sums converges to a Brownian motion on which a drift sets in at time θ. Using this, the asymptotic performance of the one-sided control chart is compared with one based on partial sums of the X's. Location change, scale change and contamination are considered. It is found that for distributions with heavy tails, the control chart based on ranks stops more frequently and faster than the one based on the X's. Performance of the two procedures are also tested on simulated data and the outcomes are compatible with the theoretical results.

1. Introduction. An important problem in industrial quality control is to detect changes in the distribution of sequentially observed independent random variables (rv's) X_i, $i = 1, 2, \cdots$. The existing literature mainly deals with a change in the mean of normal rv's or a change in the probability of Bernoulli trials. If the mean changes from a known μ_0 to μ_1 after some unknown point, then the simplest procedure to detect such a change is to stop at the smallest n for which $S_n = \sum_{i=1}^{n} (X_i - \mu_0)$ falls outside some specified limits. For the one-sided version of this method Page (1954) proposed a variation which stops as $S_n - \min_{1 \le k \le n} S_k$ gets large. Bather (1963) and Shiryaev (1963) obtained optimum stopping rules in the Bayesian formulation of the problem.

The present paper develops a nonparametric procedure for detecting a change in the distribution without any knowledge of the initial distribution. In the area of nonsequential inference about change-points, Chernoff and Zacks (1964), Gardner (1969), Hinkley (1970), Sen and Srivastava (1975) and P. K. Bhattacharya and Brockwell (1976) have studied the change in mean of a univariate normal distribution, Kander and Zacks (1966) have considered a change in a single-parameter exponential distribution and P. K. Bhattacharya (1978) has considered the problem for an arbitrary multi-parameter family of distributions satisfying some regularity conditions. Nonparametric methods based on ranks for detecting a location change in the nonsequential case have been proposed by G. K. Bhattacharya and Johnson (1968), Sen and Srivastava (1975) and Darkhovskv (1976). Likewise, sequential ranks will play a key role in a nonparametric sequential approach. The type of change to be considered here will be quite general. Location change, scale change and contamination are special cases. However, it will be assumed that a small change in distribution (in a sense to be defined later) takes place after a large number of observations. The main concern of this paper is to study the asymptotic behavior of cumulative sums of sequential ranks under these assumptions. In the nonsequential parametric case, P. K. Bhattacharya

Received July, 1978; revised March, 1980.

[1] This research was supported by National Science Foundation Grant MCS 77-01663.

AMS 1970 *subject classification.* Primary 62N10; secondary 62E20, 62G99, 62L10.

Key words and phrases. Nonparametric control chart, sequential rank, weak convergence, Brownian motion.

and Brockwell (1976) and P. K. Bhattacharya (1978) have investigated the behavior of the maximum likelihood estimator of a change-point from such an asymptotic point of view.

For each positive integer N, let $\{X_j^{(N)}, 1 \le j \le N\}$ be a sequence of independent rv's of which the first $[N\theta]$ have continuous cumulative distribution function (cdf) F and the remaining $N - [N\theta]$ have continuous cdf G_N, where $\theta \in (0, 1]$, F and G_N are unknown. The rv's $X_j^{(N)}, j = 1, 2, \cdots$ are observed sequentially and the object is to stop soon after $X_{[N\theta]+1}$, keeping the probability of stopping before $X_{[N\theta]}$ small. We assume that N is large and the change from F to G_N is small in the sense that

$$\begin{aligned} \xi_N &= \int F(x)\,dG_N(x) = \tfrac{1}{2} + \delta N^{-1/2} + o(N^{-1/2}), \qquad \delta > 0, \\ \alpha_N &= 2\int F(x)G_N(x)\,dG_N(x) = \tfrac{2}{3} + O(N^{-1/2}), \\ \beta_N &= \int F^2(x)\,dG_N(x) = \tfrac{1}{3} + O(N^{-1/2}). \end{aligned} \tag{1}$$

For simplicity, we shall write $X_j^{(N)} = X_j$, $G_N = G$, $\xi_N = \xi$, $\alpha_N = \alpha$, $\beta_N = \beta$ and $[N\theta] = N\theta$ when there is no danger of confusion.

This model arises in the following context. A machine produces items from which random samples of size 1 are taken at frequent intervals and X_j is observed on the jth sample. X_j has cdf F until and unless a small disorder in the machine changes the distribution to G which is close to F as in (1). If the sampling cost is also small, then the sampling interval can be made small and a large number of observations are available before any disorder occurs. The condition $\delta > 0$ in (1) means that the observations after change will tend to be larger than those before change and for that reason, a one-sided stopping rule based on cumulative sums of sequential ranks will be considered. For $\delta \neq 0$, a corresponding two-sided rule will be appropriate. The reason for considering a finite sequence $\{X_j, 1 \le j \le N\}$ is that the machine is assumed to be routinely adjusted at certain intervals even if no disorder has been detected, and $\theta = 1$ means no disorder takes place in the entire interval.

REMARK 1. Large N can be related to a small change in distribution in the following manner. Suppose we want to detect a change for which $\int F dG - \tfrac{1}{2} = \varepsilon$. We then want $N = O(\epsilon^{-2})$. For example if $F(x) = \Phi(x)$, $G(x) = \Phi(x - \mu)$, where Φ is the standard normal cdf, then for small μ, $\int F dG - \tfrac{1}{2} \simeq \mu/(2\sqrt{\pi})$. To detect $\mu = 0.1$, if we take $N = 1000$, then (1) is satisfied with $\delta = 0.892$.

In sequential analysis the usual ranks have been used by Wilcoxon, Rhodes and Bradley (1963), Savage and Sethuraman (1966) and later by Sethuraman (1970) for the two-sample problem. For a nonparametric control chart we shall use sequential ranks, which were first considered by Parent (1965).

DEFINITION. The sequential rank of X_i among $X_1, \cdots, X_i$ is defined by

$$R_i = 1 + \textstyle\sum_{k=1}^{i-1} u(X_k - X_i), \tag{2}$$

where u is the indicator function of $(-\infty, 0]$.

Parent proved that sequential ranks from i.i.d observations are independent, the ith being uniformly distributed on $\{1, \cdots, i\}$. Thus if

$$Z_i = i^{-1}[R_i - (i+1)/2], \qquad S_k = \textstyle\sum_{i=1}^{k} Z_i, \tag{3}$$

then for $i \leq N\theta$, the Z_i are independently distributed with mean 0 and $\mathrm{Var}(Z_i) = (1 - i^{-2})/12$ and though the behavior of Z_i for $i > N\theta$ is quite complicated, the mean of Z_i is clearly positive for $i > N\theta$. Hence it is reasonable to use S_k, $1 \leq k \leq N$, for detecting disorders in the following manner.

Consider the normalized and linearly interpolated continuous time version of $\{S_k, 1 \leq k \leq N\}$ given by

$$S_N(t) = (12/N)^{1/2}\{S_{[Nt]} + (Nt - [Nt])Z_{[Nt]+1}\}, \qquad 0 \leq t \leq 1. \tag{4}$$

NONPARAMETRIC CONTROL CHART. On the cartesian plane let the horizontal axis denote time, $0 \leq t \leq 1$, and draw a horizontal line called the control limit at a distance $c > 0$ from this axis. The graph of $S_N(t)$, plotted sequentially in relation to the control limit as each new X_i is observed, is a nonparametric control chart. Observation is terminated and a disorder is announced as soon as a point on this control chart falls above the control limit.

In order to determine the constant c for the control chart and to study its properties when an actual change in distribution occurs, we shall derive the weak limit of the stochastic process $\{S_N(t), 0 \leq t \leq 1\}$. This limit, as stated below, is the main result of this paper.

THEOREM 1. *Let Z_i, S_k and $S_N(t)$ be as in (3) and (4). Then, with respect to uniform convergence in $C[0, 1]$, $\{S_N(t), 0 \leq t \leq 1\}$ converges weakly to $\{X(t), 0 \leq t \leq 1\}$ given by*

$$X(t) = B(t) + \sqrt{12}\,\delta\theta \log(t/\theta) I_{[\theta,1]}(t), \qquad 0 \leq t \leq 1, \tag{5}$$

where $B(t)$ is standard Brownian motion and $\delta > 0$ is determined from (1).

The proof of this theorem will be accomplished in Section 3 and the Appendix after obtaining some basic properties of sequential ranks in Section 2. The asymptotic behavior of the nonparametric control chart is then understood by analyzing the properties of the following stopping rule for the process $X(t)$ described in Theorem 1, which we shall refer to as the Logarithmic Stopping Rule because of the logarithmic drift in $X(t)$.

LOGARITHMIC STOPPING RULE: *Stop at the smallest t for which $X(t) \geq c$.*

On the other hand, suppose the original observations have finite variances both before and after change. Let M_F and σ_F^2 denote the mean and the variance of F and define M_G and σ_G^2 likewise. Suppose the change from F to G_N is such that (1) holds and $\sqrt{N}(M_{G_N} - M_F)/\sigma_F \to \Delta > 0$ and $\sigma_{G_N} \to \sigma_F$ as $N \to \infty$. In such a case if M_F and σ_F^2 of the initial distribution are known, then a control chart based on the cumulative sums $N^{-1/2} \sum_{i=1}^{k} Z_i^*$ with a control limit c can be used, where $Z_i^* = (X_i - M_F)/\sigma_F$. Let

$$S_N^*(t) = N^{-1/2}\{\textstyle\sum_{i=1}^{[Nt]} Z_i^* + (Nt - [Nt])Z_{[Nt]+1}^*\}, \qquad 0 \leq t \leq 1. \tag{6}$$

Then it is easy to see that $\{S_N^*(t), 0 \leq t \leq 1\}$ converges weakly to the process

$$X^*(t) = B(t) + \Delta(t - \theta) I_{[\theta,1]}(t), \qquad 0 \leq t \leq 1, \tag{7}$$

where $B(t)$ is standard Brownian motion and Δ is as given above. The asymptotic behavior of the control chart based on the normalized continuous time version $S_N^*(t)$, $0 \leq t \leq 1$ of $N^{-1/2} \sum_{i=1}^{k} Z_i^*$, $1 \leq k \leq N$, is then understood by analyzing the properties of the following stopping rule based on $X^*(t)$ which we shall refer to as the Linear Stopping Rule because of the linear drift in $X^*(t)$.

LINEAR STOPPING RULE: *Stop at the smallest t for which $X^*(t) \geq c$.*

In Section 4, the performances of the above Linear and Logarithmic Stopping Rules are compared both theoretically and empirically. There it is seen that for distributions with heavy tails, the nonparametric control chart works better than the chart based on $S_N^*(t)$ in the sense that it stops more frequently and more quickly.

REMARK 2. In practice, random samples are often taken in batches of $m > 1$. In such a case sequential ranks of a batch of m observations would depend on their arbitrary time ordering. However, for large N, such arbitrary time orderings will have a negligible effect on the nonparametric control chart. To see this, consider the ith batch of m observations. Between any two time orderings of these observations, the maximum discrepancy in $\sum_{j=1}^m Z_{m(i-1)+j}$ is $\sum_{j=2}^m (j-1)\{m(i-1)+j\}^{-1}$. Hence the maximum discrepancy in $S_N(t)$ due to arbitrary time orderings within batches is

$$(12/N)^{1/2} \sum_{i=1}^{Nt/m} \sum_{j=2}^m (j-1)\{m(i-1)+j\}^{-1} \leq (3/N)^{1/2}(m-1)[m+\log(Nt/m)],$$

which tends to 0 as $N \to \infty$.

2. Properties of sequential ranks. The following lemma strengthens Parent's (1965) result that for i.i.d observations the sequential ranks are independent.

LEMMA 2.1. *The random vectors $(R_1, \cdots, R_i)$ and $(R_{i+1}, \cdots, R_N)$ of sequential ranks are independent whenever $i \leq [N\theta]$.*

PROOF. Fix $i \leq [N\theta]$. Observe that the vector of sequential ranks $(R_1, \cdots, R_i)$ is in one-one correspondence with the vector of usual ranks $(\hat{R}_{i1}, \cdots, \hat{R}_{ii})$. It is, therefore, enough to show that $(\hat{R}_{i1}, \cdots, \hat{R}_{ii})$ and $(R_{i+1}, \cdots, R_N)$ are independent. But this follows because $X_1, \cdots, X_i$ are i.i.d and independent of $(X_{i+1}, \cdots, X_N)$, so that the joint distribution of $X_1, \cdots, X_N$ is invariant under any transformation of the sample space which permutes the first i coordinates and leaves the remaining $N-i$ coordinates unchanged.

An immediate consequence of Lemma 2.1 is that $R_1, \cdots, R_{[N\theta]}$ are mutually independent and for $i \leq [N\theta]$, R_i is uniformly distributed on $\{1, \cdots, i\}$.

The relevant results on means, variances and covariances of sequential ranks are given below,

LEMMA 2.2. (i) *For $i \leq [N\theta]$, $E(R_i) = (i+1)/2$ and $\mathrm{Var}(R_i) = (i^2-1)/12$.*

(ii) *For $i > [N\theta]$, $E(R_i) = (i+1)/2 + [N\theta](\xi - ½)$ and $\mathrm{Var}(R_i) = (i^2-1)/12 + [N\theta]^2(\xi - \xi^2 + \beta - \alpha - 1/12) + [N\theta](2\xi - \beta - \alpha) + [N\theta]\, i(-\xi + \alpha - 1/6)$.*

(iii) *For $i < j$ and $i \leq [N\theta]$, $\mathrm{Cov}(R_i, R_j) = 0$, and for $[N\theta] < i < j$, $\mathrm{Cov}(R_i, R_j) = [N\theta](-2\xi^2 + 5\xi - 3\alpha)/2$.*

PROOF. (i) and the first part of (iii) are immediate from Lemma 2.1 and the remark immediately following it. To prove the remaining parts, express R_i, R_j by (2) and find $E(R_i)$ and $\mathrm{Var}(R_i)$ from $E(R_i \mid X_i)$ and $\mathrm{Var}(R_i \mid X_i)$, and $\mathrm{Cov}(R_i, R_j)$ from $E(R_i \mid X_i, X_j)$, $E(R_j \mid X_i, X_j)$ and $\mathrm{Cov}(R_i, R_j \mid X_i, X_j)$, observing that for $[N\theta] < i < j$, X_i and X_j both have cdf G. □

For Z_i defined by (3), we now use (1) to obtain:

$$(8)\quad \begin{cases} E(Z_i) = 0 \quad \text{and} \quad \mathrm{Var}(Z_i) = (1-i^{-2})/12 \quad \text{for} \quad i \leq [N\theta] \\ E(Z_i) = N^{1/2}\delta\theta i^{-1}\{1+o(1)\} \text{ and } \mathrm{Var}(Z_i) = (1-i^{-2})/12 + 0(N^{-1/2}) \text{ for } i > [N\theta], \\ \mathrm{Cov}(Z_i, Z_j) = 0 \quad \text{for} \quad \min(i,j) \leq [N\theta]. \\ \mathrm{Cov}(Z_i, Z_j) = 0(N^{-3/2}) \quad \text{for} \quad \min(i,j) > [N\theta]. \end{cases}$$

Finally, we give some results about conditional expectations of sequential ranks given the original observations.

LEMMA 2.3.

$$E(R_i \mid X_j) = \begin{cases} (i+1)/2 & \textit{if} \quad i \le [N\theta] \textit{ and } j > i \\ (i+1)/2 + [N\theta](\xi - 1/2) & \textit{if} \quad i > [N\theta] \textit{ and } j > i \\ 1 + (j-1)F(X_j) & \textit{if} \quad i = j \le [N\theta] \\ 1 + [N\theta]F(X_j) + (j - [N\theta] - 1)G(X_j) & \textit{if} \quad i = j > [N\theta] \\ (i+1)/2 + 1/2 - F(X_j) & \textit{if} \quad j < i \le [N\theta] \\ ([N\theta]-1)\,\xi + (i - [N\theta] - 1)/2 + 2 - G(X_j) & \textit{if} \quad j \le [N\theta] < i \\ [N\theta]\xi + (i - 1 - [N\theta])/2 + 3/2 - G(X_j) & \textit{if} \quad [N\theta] < j < i. \end{cases}$$

PROOF. For $j > i$, R_i is independent of X_j, so $E(R_i \mid X_j) = E(R_i)$ and the first two cases follow from Lemma 2.2. For $i = j$, $E(R_i \mid X_j) = 1 + \sum_{k=1}^{j-1} E(u(X_k - X_j) \mid X_j]$ and the next two cases follow because $E[u(X_k - X_j) \mid X_j]$ is $F(X_j)$ or $G(X_j)$ according as $k \le [N\theta]$ or $k > [N\theta]$. For the remaining cases, $j < i$ and so

$$E(R_i \mid X_j) = 1 + \textstyle\sum_{k=1}^{j-1} E[u(X_k - X_i)] + E[u(X_j - X_i) \mid X_j] + \sum_{k=j+1}^{i-1} E[u(X_k - X_i)]$$

from which the lemma follows. □

3. Convergence of finite-dimensional distributions (fdd) of $\{S_N\}$. In this section we prove that the fdd of $\{S_N(t)\}$ given by (4) converge to those of $\{X(t)\}$ given by (5). This convergence, stated as Theorem 1A below, constitutes half of Theorem 1 stated in the Introduction. The other half, viz. the tightness of $\{S_N(t)\}$, will be established in Theorem 1B.

THEOREM 1A. *For arbitrary positive integers d, d′ and for arbitrary* $0 < t_1' < \cdots < t_{d'}' = \theta < t_1 < \cdots < t_d \le 1$,

$$(S_N(t_1') - \mu(t_1'), \cdots, S_N(t_d) - \mu(t_d)) \to_{\mathscr{L}} (B(t_1'), \cdots, B(t_d)),$$

where

$$\mu(t) = \sqrt{12}\,\delta\theta \log(t/\theta) I_{[\theta,1]}(t), \qquad 0 \le t \le 1. \tag{9}$$

Since $|Z_i| \le 1/2$, the linear interpolation term of $S_N(t)$ may be disregarded in proving Theorem 1A. Replacing $[Nt]$ by Nt for notational simplicity, we then have $S_N(t) = (12/N)^{1/2} \sum_1^{Nt} Z_i$.

The following elementary facts are needed in this section. For $0 < s < t \le 1$, as $N \to \infty$,

$$\begin{aligned} &\textstyle\sum_{i=Ns+1}^{Nt} i^{-1} = \log((t/s) + o(1) \\ &N^{-1} \textstyle\sum_{j=1}^{Nt} (1 - \sum_{i=j}^{Nt} i^{-1})^2 = t + o(1) \\ &N^{-1} \textstyle\sum_{j=Ns+1}^{Nt} (1 - \sum_{i=j}^{Nt} i^{-1})^2 = t - s - s(\sum_{i=Ns+1}^{Nt} i^{-1})^2 + o(1). \end{aligned} \tag{10}$$

These are obtained by approximation of integrals by Riemann sums and rearrangements of terms by interchanging order of summation. From (8) and (10) we obtain

$$ES_N(t) = \mu(t) + o(1) \qquad \text{and} \qquad \operatorname{Var} S_N(t) = t + o(1), \tag{11}$$

where $\mu(t)$ is given by (9).

Since the Z_i are dependent for $i > N\theta$, we shall approximate $S_N(t)$ by

$$A_N(t) = (12/N)^{1/2} A_{Nt} \qquad \textit{where} \quad A_k = \textstyle\sum_{j=1}^{k} E(S_k \mid X_j) - (k-1)E(S_k) \tag{12}$$

on $\theta < t \le 1$. A process analogous to (12) has been used by Reynolds (1975) in a similar situation. For each t, $A_N(t)$ is a sum of independent rv's, but for $t_1 < t_2$, $A_N(t_2)$ is obtained in a more complicated manner than by addition of new independent summands to $A_N(t_1)$. By Lemma 4.1 of Hájek (1968),

$$E[S_N(t) - A_N(t)]^2 = 12N^{-1}[\text{Var } S_{Nt} - \text{Var } A_{Nt}],$$

which tends to 0 as $N \to \infty$ by Lemma 2.3, (10) and (11). Hence

LEMMA 3.1. *$S_N(t) - A_N(t) \to 0$ in probability.*

Finally, we analyze the increments of $\{A_N(t), 0 \le t \le 1\}$ in the following lemma, which is obtained by using Lemma 2.3 and rearranging terms.

LEMMA 3.2. *Let $\theta = t_o < t_1 < \cdots < t_d \le 1$. Then for arbitrary $l_1, \cdots, l_d$,*

$$\begin{aligned}\sum_{r=1}^{d} l_r\{(A_N(t_r) - EA_N(t_r)) - (A_N(t_{r-1}) - EA_N(t_{r-1}))\} &= (12/N)^{1/2}[-(\sum_{r=1}^{d} l_r\lambda_{Nr}) \sum_{j=1}^{Nt_o} H(X_j)\\ &\quad - \sum_{s=1}^{d-1} \sum_{j=Nt_{s-1}+1}^{Nt_s} \{\sum_{r=s+1}^{d} l_r\lambda_{Nr} - l_s(1 - \sum_{i=j}^{Nt_s} i^{-1})\}H(X_j)\\ &\quad + l_d \sum_{j=Nt_{d-1}+1}^{Nt_d} (1 - \sum_{i=j}^{Nt_d} i^{-1})H(X_j)] + o_p(1),\end{aligned}$$

where $\lambda_{Nr} = \sum_{i=Nt_{r-1}+1}^{Nt_r} i^{-1}$ for $1 \le r \le d$, $H(X_j) = F(X_j) - \frac{1}{2}$ for $1 \le j \le N\theta$ and $H(X_j) = G(X_j) - \frac{1}{2}$ for $N\theta + 1 \le j \le N$.

PROOF OF THEOREM 1A. For each N, the d' increments before θ and the vector of d increments after θ are based on

$$\sum_{i=1}^{Nt'_1} Z_i, \sum_{i=Nt'_1+1}^{Nt'_2} Z_i, \cdots, \sum_{i=Nt'_{d'-1}}^{Nt'_{d'}} Z_i, \sum_{Nt_o+1}^{Nt_d} Z_i,$$

which are mutually independent by Lemma 2.1. It is, therefore, enough to show

$$S_N(t'_r) - S_N(t'_{r-1}) \to_{\mathscr{L}} N(0, t'_r - t'_{r-1}) \quad \text{for} \quad 1 \le r \le d', \tag{13}$$

and

$$\sum_{r=1}^{d} l_r\{(A_N(t_r) - \mu(t_r)) - (A_N(t_{r-1}) - \mu(t_{r-1}))\} \to_{\mathscr{L}} N(0, \sum_{r=1}^{d} l_r^2(t_r - t_{r-1})) \tag{14}$$

for all $(l_1, \cdots, l_d) \ne (0, \cdots, 0)$, because then by Lemma 3.1 and the Cramér-Wold Theorem,

$$\begin{aligned}\{S_N(t'_r) - S_N(t'_{r-1}), 1 \le r \le d';\quad (S_N(t_r) - \mu(t_r)) - (S_N(t_{r-1}) - \mu(t_{r-1})), 1 \le r \le d\}\\ \to_{\mathscr{L}} \{\textstyle\prod_{r=1}^{d'} N(0, t'_r - t'_{r-1})\} \times \{\prod_{r=1}^{d} N(0, t_r - t_{r-1})\},\end{aligned}$$

which should be the case for standard Brownian motion. Since $S_N(t'_r) - S_N(t'_{r-1})$ is the normalized sum of independent Z_i and $|Z_i| \le \frac{1}{2}$, (13) is immediate. To prove (14), use $EA_N(t) = \mu(t) + o(1)$ to write the left-hand side of (14) in the form

$$\sum_{j=1}^{Nt_d} b_{Nj}H(X_j) + o_p(1),$$

where b_{Nj} and $H(X_j)$ are given in Lemma 3.2. But the $H(X_j)$ are independent with mean 0, variance $\frac{1}{12}$ and $|H(X_j)| \le 1$, so that

$$\sum_{j=1}^{Nt_d} b_{Nj}H(X_j)/\lim_{N\to\infty} (12^{-1} \sum_{j=1}^{Nt_d} b_{Nj}^2)^{1/2} \to_{\mathscr{L}} N(0, 1),$$

and by Lemma 3.2 and (10), $\lim_{N\to\infty} 12^{-1} \sum_{j=1}^{Nt_d} b_{Nj}^2 = \sum_{r=1}^{d} l_r^2(t_r - t_{r-1})$. This completes the proof.

The other half of Theorem 1 is stated below, but its proof will be outlined in the Appendix.

THEOREM 1B. *The sequence $\{S_N(t), 0 \le t \le 1\}$ is tight.*

4. Properties of the nonparametric control chart in some special cases. In this section F and G_N have finite variances σ_F^2 and $\sigma_{G_N}^2$ satisfying $\lim_{N\to\infty} \sigma_{G_N} = \sigma_F$ and their means M_F and M_{G_N} satisfy $\lim_{n\to\infty} \sqrt{N}(M_{G_N} - M_F)/\sigma_F = \Delta > 0$. We shall use Theorem 1 to compare the performance of the nonparametric control chart with that of the one based on $S_N^*(t)$ given by (6) when N is large. This is done by comparing the constant boundary crossing behaviors of the process $X(t)$ of Theorem 1 and the process $X^*(t)$ given by (7). The control limit c is chosen so that the probability of false alarm is held at a preassigned level α. For asymptotic purposes, we need

$$P\{\max_{0\le t\le 1} B(t) \ge c\} = 2P\{B(1) \ge c\} = \alpha,$$

so that $c = \Phi^{-1}\{(1-\alpha)/2\}$, where Φ is the standard normal cdf. Define

$$\psi(t) = c - \sqrt{12}\,\delta\theta\log(t/\theta)I_{[\theta,\,1]}(t), \qquad \psi^*(t) = c - \Delta(t-\theta)I_{[\theta,\,1]}(t),$$

where δ is given in (1) and Δ is as described above. Then the stopping rules stated in Section 1 are equivalent to:

LOGARITHMIC STOPPING RULE: *Stop at the smallest t for which $B(t) \ge \psi(t)$.*

and

LINEAR STOPPING RULE: *Stop at the smallest t for which $B(t) \ge \psi^*(t)$.*

The two rules can be compared through the probability of stopping before time 1 or by the conditional expectation of the stopping time given that a stoppage has occurred. For this, we compare $\psi(t)$ and $\psi^*(t)$on $\theta < t \le 1$, and we observe that three cases may arise. In the following, $\psi'(\theta)$ and $\psi^{*\prime}(\theta)$ are right-hand derivatives.

Case 1. $\psi(t) > \psi^*(t)$ for all $\theta < t \le 1$. This holds when $\psi'(\theta) \ge \psi^{*\prime}(\theta)$, i.e.,

(15a) $$\sqrt{12}\,\delta\Delta^{-1} \le 1.$$

In this case, the linear stopping rule performs better in both ways, so that the nonparametric control chart is asymptotically less efficient.

Case 2. There is a point $a \in (\theta, 1)$ such that $\psi(t) < \psi^*(t)$ for all $\theta < t < a$ and $\psi(t) > \psi^*(t)$ for all $a < t \le 1$. This holds when both $\psi'(\theta) < \psi^{*\prime}(\theta)$ and $\psi(1) > \psi^*(1)$ hold, i.e.,

(15b) $$1 < \sqrt{12}\,\delta\Delta^{-1} < (1-\theta)/\{\theta\log(1/\theta)\}.$$

This case will subdivide into further cases where one or the other stopping rule may stop before time 1 with a higher probability, but the logarithmic stopping rule has the advantage that it performs better in terms of early stopping because of the advantage that $\psi(t)$ has over $\psi^*(t)$ immediately after θ. If the constant a is close to 1, then the nonparametric control chart certainly looks more attractive.

Case 3. $\psi(t) \le \psi^*(t)$ for all $\theta < t \le 1$, the only possibility of an equality being at $t = 1$. This holds when $\psi(1) \le \psi^*(1)$, i.e.,

(15c) $$\sqrt{12}\,\delta\Delta^{-1} \ge (1-\theta)/\{\theta\log(1/\theta)\}.$$

In this case the logarithmic stopping rule performs better in both ways, so that the nonparametric control chart is more efficient.

REMARK 3. It is interesting to note that classification into the above three cases is determined solely on the basis of the criterion $\psi'(\theta)/\psi^{*\prime}(\theta) = \sqrt{12}\;\delta\Delta^{-1}$, which is the ratio of the Pitman efficacy of the Wilcoxon test to that of the t-test for comparing the observations before and after change nonsequentially (see Lehmann (1975), Theorem 11, page 372). The relevance of the t-test in comparing $\sum_1^{N\theta} X_i$ and $\sum_{N\theta+1}^{N(\theta+\epsilon)} X_i$, $\epsilon > 0$, is obvious. On the other hand, for small ϵ, $\sum_{N\theta+1}^{N(\theta+\epsilon)} i^{-1}R_i$ is approximately a linear function of the Wilcoxon statistic computed from $X_1, \cdots, X_{N\theta}$ and $X_{N\theta+1}, \cdots, X_{N(\theta+\epsilon)}$.

We shall now examine three types of changes in the context of some well-known distributions and will classify each situation in one of the above cases by means of $\sqrt{12}\ \delta\Delta^{-1}$.

4.1 *Location Change.* Suppose F has density f for which $\int f^2(x)\,dx < \infty$. Let $G_N(x) = F(x - \mu N^{-1/2})$, $\mu > 0$. Then $\Delta = \mu\sigma_F^{-1}$. Moreover,

$$\xi_N = \int F\,dG_N = \frac{1}{2} + \mu N^{-1/2}\int h_N\,dF, \qquad \text{where} \quad h_N = H_{\mu N^{-1/2}},$$

$$H_\theta(x) = \theta^{-1}\{F(x+\theta) - F(x)\} = \theta^{-1}\int_x^{x+\theta} f(y)\,dy.$$

By the Cauchy-Schwarz inequality $H_\theta^2(x) \le \theta^{-1}\int_x^{x+\theta} f^2(y)\,dy$, so that

$$\int H_\theta^2(x)\,dx \le \theta^{-1}\int dx\int_x^{x+\theta} f^2(y)\,dy = \theta^{-1}\int f^2(y)\,dy\int_{y-\theta}^{y} dx = \int f^2(y)\,dy$$

for all θ. Hence $\int h_N^2(x)\,dx \le \int f^2(x)\,dx$ for all N. Using the Cauchy-Schwarz inequality again, we have $\int h_N(x)\,dF(x) \le \int f^2(x)\,dx < \infty$ for all N. Theorem 4.2, page 64 of Hájek and Šidák (1967) can now be used to show that $\lim_{N\to\infty}\int h_N(x)\,dF(x) = \int f^2(x)\,dx$. Thus (1) holds with $\delta = \mu\int f^2(x)\,dx$ (the other parts of (1) being verified in the same way), so that the criterion is given by $\sqrt{12}\ \delta\Delta^{-1} = \sqrt{12}\ \sigma_F\int f^2(x)\,dx$. We now look at two special cases.

Example 1. If F is $N(0, 1)$ in the location model, then $\sqrt{12}\sigma_F\int f^2(x)\,dx = (3/\pi)^{1/2} < 1$. Thus for all $\theta \in (0, 1)$, (15a) is satisfied and Case 1 holds. However, since $(3/\pi)^{1/2}$ is close to 1, the nonparametric control chart is not much worse than the one based on $S_N^*(t)$ unless θ is very small.

Example 2. If F is a t-distribution with n df in the location model, then the criterion of classification is

$$A(n) = \sqrt{12}\ \sigma_F\int f^2(x)\,dx = \sqrt{\frac{12}{\pi(n-2)}}\left(\frac{\Gamma\left(\frac{n+1}{2}\right)^2}{\Gamma\left(\frac{n}{2}\right)}\right)\Gamma\left(n+\frac{1}{2}\right),$$

which is to be compared with $g(\theta) = (1-\theta)/\{\theta\log(1/\theta)\}$. For selected values of n and θ, $A(n)$ and $g(\theta)$ are tabulated below in tables 1 and 2.

Comparing $A(n)$ and $g(\theta)$ we now see that

(i) For $3 \le n \le 10$, either (15b) or (15c) is satisfied, so Case 2 or Case 3 holds. For each of these values of n, Case 2 holds for smaller values of θ and Case 3 for larger values of θ. Thus for small n, when the distributions have heavy tails, the nonparametric control chart

TABLE 1.
$A(n)$ for Selected Values of n.

n	3	4	5	6	7	8	9	10	15	20
$A(n)$	1.38	1.18	1.11	1.08	1.06	1.04	1.03	1.03	1.00	0.99

TABLE 2.
$g(\theta) = (1-\theta)/\{\theta \log(1/\theta)\}$ *for Selected Values of* θ.

θ	.125	.25	.33	.50	.55	.60	.66	.75	.875	.90
$g(\theta)$	3.37	2.16	1.82	1.44	1.37	1.31	1.23	1.31	1.07	1.05

is clearly better for larger values of θ and has the advantage of early stopping even for smaller values of θ.

(ii) For $n \geq 15$, (15a) is satisfied, so Case 1 holds. In these cases, the control chart based on $N^{-1/2} \sum_{i=1}^{k} Z_i^*$ is asymptotically better.

The results of simulation also exhibit this pattern.

4.2. *Scale Change.* Suppose F has density f for which $\int |x| f^2(x)\, dx < \infty$, and M_F and $\int xf^2(x)\, dx$ are both positive. Let $G_N(x) = F(c_N x)$ where $c_N = 1 - \mu N^{-1/2}$, $\mu > 0$. Then $\Delta = \mu M_F \sigma_F^{-1}$, and by arguments similar to those used in the case of location change, we get $\delta = \mu \int xf^2(x)\, dx$. Thus the criterion of classification is $\sqrt{12}\ \delta\Delta^{-1} = \sqrt{12}\ \sigma_F M_F^{-1} \int xf^2(x)\, dx$.

Example 3. If in the scale model, F is a gamma distribution with density

$$f(x) = \{\lambda^{\nu}/\Gamma(\nu)\} x^{\nu-1} \exp(-\lambda x), \qquad x > 0, \quad \nu > 0, \lambda > 0,$$

then $\sqrt{12}\ \sigma_F M_F^{-1} \int xf^2(x)\, dx = (12/\nu)^{1/2} \Gamma(2\nu)/\{(\Gamma(\nu))^2 2^{2\nu}\} = k(\nu)$, say. Since $k(\nu)$ increases with ν and $\lim_{\nu\to\infty} k(\nu) = (3/\pi)^{1/2} < 1$, (15a) is always satisfied, so Case 1 holds. However, $k(\nu)$ is close to 1 for most values of ν, so the comment made in the context of a location change in a normal distribution also applies here.

Example 4. If in the scale model, F is a log-normal distribution with parameters m and s^2, i.e., F is the cdf of exp (X) where X is $N(m, s^2)$ then $\int xf^2(x)\, dx = (2s\sqrt{\pi})^{-1}$, so that $\sqrt{12}\ \sigma_F M_F^{-1} \int xf^2(x)\, dx = \sqrt{3}(\pi s^2)^{-1/2}(\exp(s^2)-1)^{1/2}$. For large s^2, this criterion exceeds $g(\theta)$ unless θ is too small, so that (15c) is satisfied and Case 3 holds. For example, for $s^2 = 3$ and $\theta = 1/3$, the criterion takes the value 2.4648, while $g(\theta) = 1.8205$, which makes the nonparametric control chart more efficient. The results of simulation for these values of parameter are discussed in Section 5.

4.3 *Contamination.* F and H are continuous cdf's with finite variances. Suppose $M_H > M_F$, $\int FdH > 1/2$ and let $G_N(x) = (1 - \mu N^{-1/2})F(x) + \mu N^{-1/2} H(x)$, $0 < \mu < 1$, be obtained by contamination of F by the contaminating distribution H. Then $\Delta = \mu(M_H - M_F)\sigma_F^{-1}$ and (1) holds with $\delta = \mu(\int FdH - 1/2)$, so that the classification criterion is $\sqrt{12}\ \delta\Delta^{-1} = \sqrt{12}\ \sigma_F(\int FdH - 1/2)(M_H - M_F)^{-1}$.

Example. If F is $N(0, 1)$ and H is $N(1, 1)$ in the contamination model, then $\int FdH - 1/2 = .23958$ by numerical integration, and the criterion takes the value .82993. Thus (15a) is satisfied and Case 1 holds. Since the criterion is close to 1, the comments made in connection with location change of a normal distribution applies here.

5. Simulation Results. In a small-scale simulation study, 200 runs of $N = 900$ observations were generated in each of several different situations and were subjected to both the nonparametric control chart using S_k given by (3) and the one using $S_k^* = \sum_{i=1}^{k} (X_i - M_F)/\sigma_F$. Let T and T^* denote respectively the smallest k needed for $(12/N)^{1/2} S_k$ and $N^{-1/2} S_k^*$ to exceed c, which was taken to be 1.645 to set the false alarm probability at the asymptotic level $\alpha = .10$. In case of no crossing, define T and T^* to be $N + 1$. We discuss the results of our simulation in terms of

$$\hat{p} = \text{proportion of runs with } T \leq 900, \text{ and}$$

$$\hat{E}(T) = \text{mean } T \text{ for all runs with } T \leq 900$$

for the nonparametric control chart and $\hat{p}^*$ and $\hat{E}(T^*)$, defined analogously for the other control chart.

For a location change by an amount $\mu N^{-1/2} = \frac{1}{6}$ from a t-distribution with 3 df, the nonparametric control chart performed better when the change took place at $\theta = .62$ with $\hat{p} = .390$ and $\hat{E}(T) = 703.7$ against $\hat{p}^* = .360$ and $\hat{E}(T^*) = 739.0$ for the other chart. For the same amount of location change from t_3 at an earlier point $\theta = .33$, the nonparametric control chart maintained its superiority in early stopping by $\hat{E}(T) = 625.4$ against $\hat{E}(T^*) = 671.7$, but did worse in terms of frequency of stoppage by $\hat{p} = .525$ against $\hat{p}^* = .715$. For the same amount of location change from $N(0, 1)$, the nonparametric chart was clearly worse. For $\theta = .33$, it scored $\hat{p} = .705$ and $\hat{E}(T) = 633.7$ against $\hat{p}^* = .965$ and $\hat{E}(T^*) = 579.6$. The nonparametric chart was also found superior for detecting a scale change by $c_N = 1 - \mu N^{-1/2} = \frac{5}{6}$ from a log-normal distribution with $s^2 = 3$. Even for an early change at $\theta = .33$, it scored $\hat{p} = .480$ and $\hat{E}(T) = 623.9$ against $\hat{p}^* = .240$ and $\hat{E}(T^*) = 637.9$. The above results, though based on rather limited simulations, illustrate the three cases discussed in the previous section.

6. Acknowledgment. Thanks are due to Lee Applebaum for programming the simulations and other numerical work in Section 4. Helpful comments by a referee are also gratefully acknowledged.

APPENDIX

Outline of the Proof of Theorem 1B. Since $S_N(0) \equiv 0$, by Theorem 12.3 of Billingsley (1968), it is enough to show that the moment condition

$$E(|S_N(t_2) - S_N(t_1)|^4) \leq c(t_2 - t_1)^2 \tag{16}$$

holds for all N, $0 \leq t_1 < t_2 \leq 1$ and for some $c > 0$. Use $|Z_i| \leq \frac{1}{2}$ throughout the proof. For $Nt_2 - Nt_1 \leq 1$, (16) holds trivially with $c = 9$, and for $Nt_2 - Nt_1 > 1$, (16) is implied by

$$E(\textstyle\sum_{k=[Nt_1]+1}^{[Nt_2]} Z_k)^4 \leq c^2 N^2 (t_2 - t_1)^2. \tag{17}$$

Expanding $(\sum_{k=[Nt_1]+1}^{[Nt_2]} Z_k)^4$, it is seen that the nontrivial terms to consider are $E(Z_k^2 Z_l Z_m)$ are $E(Z_k Z_l Z_m Z_n)$ for $k \neq l \neq m \neq n$, because the number of the other terms are $O(N^2)$. Using Lemma 2.1 and (8), we now see that the main thing is to prove that for some $c > 0$,

$$\textstyle\sum_{[Nt_1]+1 \leq k \neq l \neq m \leq [Nt_2]} E(Z_k^2 Z_l Z_m) \leq cN^2(t_2 - t_1)^2$$

and

$$\textstyle\sum_{[Nt_1]+1 \leq k \neq l \neq m \neq n \leq [Nt_2]} E(Z_k Z_l Z_m Z_n) \leq cN^2(t_2 - t_1)^2,$$

for which it is enough to show that

$$E[\{R_{s_1} - (s_1 + 1)/2\}^2 \textstyle\prod_{i=2}^3 \{R_{s_i} - (s_i + 1)/2\}] = O(N^3) \tag{18}$$

and

$$E[\textstyle\prod_{i=1}^4 \{R_{s_i} - (s_i + 1)/2\}] = O(N^2) \tag{19}$$

hold for all distinct integers s_1, s_2, s_3, s_4 between $[N\theta] + 1$ and N. Finally, by using $E(R_i) - (i + 1)/2 = [N\theta](\xi - \frac{1}{2}) = O(N^{1/2})$, $\mathrm{Var}(R_i) = O(N^2)$ and $\mathrm{Cov}(R_i, R_j) = O(N^{1/2})$ for

$i, j > [N\theta]$, it is seen that (18) and (19) are implied by

$$E[\prod_{i=1}^{3} (R_{s_i} - ER_{s_i})] = O(N),$$

$$E[\prod_{i=1}^{4} (R_{s_i} - ER_{s_i})] = O(N^2),$$

$$E[(R_{s_1} - ER_{s_1})^2(R_{s_2} - ER_{s_2})] = O(N^2), \quad \text{and}$$

$$E[(R_{s_1} - ER_{s_1})^2 \prod_{i=2}^{3} (R_{s_i} - ER_{s_i})] = O(N^3)$$

for all distinct integers s_1, s_2, s_3, s_4 between $[N\theta] + 1$ and N. Writing R_i by (2), these relations are established by repeated smoothing taking appropriate conditional expectations. The details have been given by Frierson (1977).

REFERENCES

BATHER, J. A. (1963). Control charts and the minimization of costs. *J. Roy. Statist. Soc. Ser. B*, **25** 49–80.

BHATTACHARYA, P. K. (1978). Estimation of change-point in the distribution of independent random variables. Unpublished manuscript.

BHATTACHARYA, P. K. and BROCKWELL, P. J. (1976). The minimum of an additive process with applications to signal estimation and storage theory. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete* **37** 51–75.

BHATTACHARYA, G. K. and JOHNSON, R. A. (1968). Nonparametric tests for shift at unknown time point. *Ann. Math. Statist.* **39** 1731–1743.

BILLINGSLEY, PATRICK (1968). *Convergence of Probability Measures.* Wiley, New York.

CHERNOFF, H. and ZACKS, S. (1964). Estimating the current mean of a normal distribution which is subjected to change in time. *Ann. Math. Statist.* **35** 999–1018.

DARKHOVSKV, B. S. (1976). A nonparametric method for the a posteriori detection of the 'disorder' time of a sequence of independent random variables. *Theor. Probability Appl.* **21** 178–183.

FRIERSON, DARGAN, JR. (1977). A nonparametric approach to sequential detection of small changes in distribution. Doctoral Dissertation. Univ. Arizona.

GARDNER, L. A. (1969). On detecting changes in the mean of normal variables. *Ann. Math. Statist.* **40** 116–126.

HAJÉK, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. *Ann. Math. Statist.* **39** 325–346.

HAJÉK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests.* Academic, New York.

HINKLEY, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika* **57** 1–17.

KANDER, Z. and ZACKS. S. (1966). Test procedures for possible changes in parameters of statistical distributions occurring at unknown time points. *Ann. Math. Statist.* **37** 1196–1210.

LEHMANN, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks.* Holden-Day, San Francisco.

PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–114.

PARENT, E. A., JR. (1965). Sequential ranking procedures. Doctoral Dissertation, Stanford Univ.

REYNOLDS, M. R., JR. (1975). A sequential signed-rank test for symmetry. *Ann. Statist.* **3** 382–400.

SAVAGE, I. R. and SETHURAMAN, J. (1966). Stopping time of a rank order sequential probability ratio test on Lehmann alternatives. *Ann. Math. Statist.* **37** 1154–1160.

SEN, A. and SRIVASTAVA, M. S. (1975). On tests for detecting change in mean. *Ann. Statist.* **3** 90–108.

SETHURAMAN, J. (1970). Stopping time of a rank order sequential probability ratio test on Lehmann alternatives II. *Ann. Math. Statist.* **41** 1322–1333.

SHIRYAEV, A. N. (1963). On optimum methods in quickest detection problems. *Theor. Probability Appl.* **8** 22–46.

WILCOXON, FRANK, RHODES, L. J. and BRADLEY, R. A. (1963). Two sequential two-sample grouped rank tests with applications to screening experiments. *Biometrics* **19** 58–84.

DIVISION OF STATISTICS
UNIVERSITY OF CALIFORNIA
DAVIS, CALIFORNIA 95616

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF NORTH CAROLINA AT WILMINGTON
P. O. BOX 3725
WILMINGTON, NORTH CAROLINA 28401