The Annals of Mathematical Statistics
1971, Vol. 42, No. 2, 538–551

# SOME CHARACTERISTICS OF PAGE'S TWO-SIDED PROCEDURE FOR DETECTING A CHANGE IN A LOCATION PARAMETER

BY JACK NADLER[1] AND NAOMI B. ROBBINS
*Bell Telephone Laboratories*

We study a two-sided procedure proposed by E. S. Page for detecting a change in the location of the distribution of a sequence of independent observations which are ordered in time. We approximate the null distribution of Page's statistic and the power of his test for finite sequences. When the procedure is applied to an infinite sequence we approximate the average run length. In order to obtain these approximations we find the distribution function of the range of a Wiener process with drift and the Laplace transform of the time at which the range first exceeds some given value.

**1. Introduction.** In many applications of statistics, including areas as diverse as quality control and tracking an object following a ballistic trajectory, we are interested in detecting a change in the location of the distribution of a sequence of independent observations which are ordered in time. E. S. Page (1954) introduced a control chart procedure for quality control applications which is based on cumulative sums. Cumulative sum procedures have been applied widely (see Barnard (1959), Ewan (1963), Ewan and Kemp (1960), Freund (1962), Page (1961), Traux (1961), and others) and have many advantages including simplicity, ease of visual interpretation, and speed of detection. However, since not all of the properties of these procedures have been evaluated there has been little basis for the choice of control limits. Barnard (1959) suggested the use of cumulative sum charts, rejecting whenever the plotted points fall outside the edges of a V-mask, the parameters of which he chooses empirically by cut-and-try methods. Goldsmith and Whitfield (1961) choose the parameters of the V-mask on the basis of simulations. Johnson (1961) has provided simple though nonprecise approximations for the choice of the parameters. We study characteristics of the symmetric version of Page's original two-sided procedure and approximate its average run length (the expected number of articles sampled at a given quality level before action is taken).

Typically in control chart applications no upper limit is placed on the number of observations. We will call this the continuing case. We consider also the truncated case in which at most $n$ observations are sampled. In (1955) Page suggested basing a significance test for the truncated case on his process inspection scheme. However, he states that the properties of the test will be difficult to evaluate since it is a truncated form of a linear sequential test. These technical difficulties forced Page to modify his test. Chernoff and Zacks (1964) propose a Bayesian test for a change of a parameter and compare the power of their test to that of the modified Page test. We approximate the null distribution of Page's original two-sided statistic and the power of his test. First we show that Page's procedure is equivalent

Received January 22, 1970.

[1] Now at American Telephone and Telegraph Co.

to one based on the range of the sequence of partial sums of the observations. Then using a standard method of sequential analysis we convert to the Wiener process by replacing the sum $\sum_1^m Y_i$, $m = 1, 2, \cdots$ by the Wiener process $Y(t)$, $0 \leqq t < \infty$, where $Y(0) = 0$, $\{Y(t), t \geqq 0\}$ has independent and stationary increments, and $Y(t+s) - Y(s)$ is normally distributed with zero mean and variance $t$ for all $s > 0$. The probability distribution of the range of the Wiener process is found exactly, and is used as an approximation to the distribution of the range of the discrete process. A theorem of Donsker shows that in the null case this approximate distribution of the range is, in fact, its asymptotic distribution as the number of observations increases indefinitely. This technique of approximating the probability of an event when observations are taken discretely by the corresponding probability of a continuous stochastic process has been discussed by Anderson (1960), Darling and Siegert (1953), Doob (1949), and Feller (1951).

In computing the probabilities of interest we proved some new results in the theory of Wiener stochastic processes. Let $X(t)$ be a Wiener process with drift; that is, $X(t) = \sigma Y(t) + \mu t$ where $Y(t)$ is the Wiener process. Let $R_{\mu,\sigma^2}(0, T) = \max_{0 \leqq t \leqq T} X(t) - \min_{0 \leqq t \leqq T} X(t)$. In Theorem 1 we derive the distribution function of $R_{\mu,\sigma^2}(0, T)$. Define $\tau$ to be the random variable which represents the time at which $R_{\mu,\sigma^2}(0, T)$ first exceeds some value $r$ as $T \to \infty$. We find the distribution function, the Laplace transform, and therefore, any moment of $\tau$.

Assume that the variance $\sigma^2$ of the observations is finite. Then under the null hypothesis Donsker's theorem (1951) states that the asymptotic distribution of the test statistic does not depend on the particular distribution of the observations. Thus the test has a nonparametric character although the quality of the approximation for fixed $n$ will depend upon the underlying distribution.

**2. The procedure.** Consider a sequence of independent observations $x_1, x_2, \cdots$, ordered in time, where $x_i$ is distributed according to $F(x-\theta_i)$ with known finite variance $\sigma^2$. In the null state all $\theta_i = \theta_0$; alternatively, for some $m$, $\theta_i = \theta_0$ if $i = 1, \cdots, m$ and

$$\inf_{i>m} \theta_i = \theta_1 > \theta_0 \quad \text{or} \quad \sup_{i>m} \theta_i = \theta_2 < \theta_0. \tag{2.1}$$

The objective is to detect this possible shift in the value of the mean at the unknown time $m$. In the truncated case we test the null hypothesis that $x_1, x_2, \cdots, x_n$ come from $F(x-\theta_0)$ against the alternative that for some $m$ satisfying $0 \leqq m < n$, the observations $x_1, x_2, \cdots, x_m$ come from $F(x-\theta_0)$ while $x_{m+1}, \cdots, x_n$ are distributed according to $F(x-\theta_i)$, $\theta_i$ satisfying (2.1). (Since no loss of generality results, assume henceforth that $\theta_0$ is chosen so that $\mu = E(X) = 0$ in the null state.)

One situation with no limit on the number of observations sampled is a process inspection scheme designed to detect a change in the mean of a production process. E. S. Page (1954) proposed a sequential process inspection scheme (a scheme in which a decision is made after each observation, considering all previous observations) for this purpose based on cumulative sums. Page's procedure for detecting a one-sided change from the null state to $\mu = \mu' > 0$ is to form the cumulative sums $S_0 = 0$, $S_k = \sum_{j=1}^k x_j$, $k = 1, 2, \cdots$, and to take action after the $k$th observation if $S_k - \min_{0 \leqq i < k} S_i \geqq h$. This procedure is equivalent to forming the sums $S_0' = 0$, $S_k' = \max(S_{k-1}' + x_k, 0)$, $k \geqq 1$, and to taking action after the $k$th observation if $S_k' \geqq h$. To test for a change in either direction he suggests the simultaneous application of two one-sided schemes. His rule is to take action after the $k$th observation if either

(2.2) $$S_k - \min_{0 \leqq i < k} S_i \geqq h \quad \text{or} \quad \max_{0 \leqq i < k} S_i - S_k \geqq h'.$$

The difficulty of evaluating the characteristics of this procedure forced Page to consider an alternative scheme. However, since frequently one is equally interested in changes of $\mu$ in either direction, we consider his original procedure limited to the case $h = h'$. Thus, in the symmetric version of his procedure, the cumulative sums are plotted and action is taken if the sum rises a height $h$ from its previous minimum or falls $h$ from its previous maximum.

This procedure is equivalent to one based on $R_k$, the range of the sequence of partial sums after the $k$th observation. Let

$$V_k = \max_{0 \leqq i \leqq k} S_i \quad \text{and} \quad U_k = \min_{0 \leqq i \leqq k} S_i.$$

Then $R_k = V_k - U_k$. We now show that the rule "Take action after the $k$th observation if $R_k \geqq h$" is equivalent to the symmetric version of Page's procedure. Let $h = h'$ and assume that (2.2) holds for the first time at the $N$th observation. Clearly $R_N \geqq h$. Conversely, suppose that $N'$ is the index of the first observation such that $R_{N'} \geqq h$. Then either $S_{N'} = V_{N'}$ or $S_{N'} = U_{N'}$. Thus either $S_{N'} - \min_{0 \leqq i < N'} S_i \geqq h$ or $\max_{0 \leqq i < N'} S_i - S_{N'} \geqq h$. Hence $N = N'$ and the two procedures are equivalent.

**3. Distribution theory–truncated case.** In this section we consider the probability of rejecting the null hypothesis that $x_1, x_2, \cdots, x_n$ are independent observations from a common distribution with mean $\mu$ and variance $\sigma^2$. The probability of the event in (2.2) is needed (we will hereafter always assume that $h = h'$). However, because of the equivalence of the two procedures it suffices to find the probability that $R_n \geqq h$. Derivation of the exact distribution of $R_n$ is an extremely difficult problem (see Feller (1951)). Fortunately, Donsker's invariance principle applies; i.e., the limiting distribution of $R_n/n^{\frac{1}{2}}$ does not depend on the particular distribution of the $x_i$ (see Donsker (1951)). Let $Y(t)$ be a Wiener process and let $X(t) = \sigma Y(t) + \mu t$. We regard the sum $S_k$ as the value at time $t = k$ of the process $X(t)$ and calculate the exact distribution of the range of this process with continuous time parameter. We compute $P\{R_{\mu,\sigma^2}(0, T) \leqq r\}$. If we make the change of variable $q = r/T^{\frac{1}{2}}$ we then note that

$$P\{R_{\mu,\sigma^2}(0, T) \leqq qT^{\frac{1}{2}}\} = P\{R_{\mu,\sigma^2}(0, 1) \leqq q\}.$$

By Donsker's theorem, as $n \to \infty$

$$P\{R_n \leqq qn^{\frac{1}{2}}\} \to P\{R_{0,\sigma^2}(0, 1) \leqq q\},$$

so that our approximation to the distribution of $R_n$ provides the asymptotic distribution of $R_n/n^{\frac{1}{2}}$.

Define $V_T = \max_{0 \leqq t \leqq T} X(t)$ and $U_T = \min_{0 \leqq t \leqq T} X(t)$. To find the distribution function of the range we begin by finding the joint distribution function $F_{\mu,\sigma^2}(u, v; T)$ of $V_T$ and $U_T$. Then

$$(3.1) \qquad P\{R_{\mu,\sigma^2}(0, T) \leqq r\} = \int_0^r \int_{v-r}^0 \frac{\partial^2}{\partial v\, \partial u} F_{\mu,\sigma^2}(u, v;\, T)\, du\, dv$$

$$= \int_0^r \left[ \frac{\partial}{\partial v} F_{\mu,\sigma^2}(u, v;\, T) \Big|_{u=v-r}^{u=0} \right] dv.$$

Let $u < 0$ and $v > 0$. From Theorem (4.3) of Anderson (1960) it can be shown that

$$(3.2) \qquad P\{V_T \leqq v;\, U_T \geqq u\} = \Phi\left(\frac{v-\mu T}{\sigma T^{\frac{1}{2}}}\right) - \Phi\left(\frac{u-\mu T}{\sigma T^{\frac{1}{2}}}\right)$$

$$-\textstyle\sum_{k=1}^{\infty} \{g_k(u, v;\, \mu, \sigma T^{\frac{1}{2}}) - g_k(v, u;\, \mu, \sigma T^{\frac{1}{2}})\}$$

where

$$\Phi(x) = \textstyle\int_{-\infty}^{x} \phi(z)\, dx, \qquad \phi(z) = (2\pi)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}z^2) \qquad \text{and}$$

$$g_k(u, v; \mu, \sigma T^{\frac{1}{2}})$$

$$= \exp\left\{\frac{2\mu}{\sigma^2}[kv-(k-1)u]\right\}\left[\Phi\left(\frac{2(k-1)u-(2k-1)v-\mu T}{\sigma T^{\frac{1}{2}}}\right) - \Phi\left(\frac{(2k-1)u-2kv-\mu T}{\sigma T^{\frac{1}{2}}}\right)\right]$$

$$+\exp\left\{\frac{2\mu}{\sigma^2}[k(v-u)]\right\}\left[\Phi\left(\frac{(2k+1)u-2kv-\mu T}{\sigma T^{\frac{1}{2}}}\right) - \Phi\left(\frac{2ku-(2k-1)v-\mu T}{\sigma T^{\frac{1}{2}}}\right)\right].$$

The joint distribution of $V_T$ and $U_T$ therefore can be written

$$F_{\mu,\sigma^2}(u, v; T) = P\{V_T \leqq v\} - P\{V_T \leqq v;\, U_T > u\}$$

$$= \Phi\left(\frac{u-\mu T}{\sigma T^{\frac{1}{2}}}\right) - \exp\left(\frac{2\mu v}{\sigma^2}\right)\left[1 - \Phi\left(\frac{v+\mu T}{\sigma T^{\frac{1}{2}}}\right)\right]$$

$$+\sum_{k=1}^{\infty} \{g_k(u, v; \mu, \sigma T^{\frac{1}{2}}) - g_k(v, u; \mu, \sigma T^{\frac{1}{2}})\}.$$

Performing the calculations indicated in (3.1) we derive

THEOREM 1. *Let* $X(t) = \sigma Y(t) + \mu t$ *where* $Y(t)$ *is the Wiener process with* $EY(t) = 0$ *and* $EY^2(t) = t$. *Let* $R_{\mu,\sigma^2}(0, T) = \max_{0 \leqq t \leqq T} X(t) - \min_{0 \leqq t \leqq T} X(t)$. *Then*

$$
\begin{aligned}
P\{R_{\mu,\sigma^2}(0, T) \leqq r\} &\\
= \Phi\left(\frac{r+\mu T}{\sigma T^{\frac{1}{2}}}\right) + \Phi\left(\frac{r-\mu T}{\sigma T^{\frac{1}{2}}}\right) - 1 &\\
+ \sum_{k=1}^{\infty} \Bigg\{ (4k-1)\Bigg[ \exp\left(\frac{2k\mu r}{\sigma^2}\right) \Phi\left(\frac{(2k-1)r+\mu T}{\sigma T^{\frac{1}{2}}}\right) &\\
+ \exp\left(\frac{-2k\mu r}{\sigma^2}\right) \Phi\left(\frac{(2k-1)r-\mu T}{\sigma T^{\frac{1}{2}}}\right)\Bigg] &\\
- 8k\left[\exp\left(\frac{2k\mu r}{\sigma^2}\right)\Phi\left(\frac{2kr+\mu T}{\sigma T^{\frac{1}{2}}}\right) + \exp\left(\frac{-2k\mu r}{\sigma^2}\right)\Phi\left(\frac{2kr-\mu T}{\sigma T^{\frac{1}{2}}}\right)\right] &\\
+ (4k+1)\left[\exp\left(\frac{2k\mu r}{\sigma^2}\right)\Phi\left(\frac{(2k+1)r+\mu T}{\sigma T^{\frac{1}{2}}}\right) + \exp\left(\frac{2-k\mu r}{\sigma^2}\right)\Phi\left(\frac{(2k+1)r-\mu T}{\sigma T^{\frac{1}{2}}}\right)\right] &\\
+ \frac{2k\mu T^{\frac{1}{2}}}{\sigma}\exp\left(\frac{2k\mu r}{\sigma^2}\right)\left[\phi\left(\frac{(2k+1)r+\mu T}{\sigma T^{\frac{1}{2}}}\right) + \phi\left(\frac{(2k-1)r+\mu T}{\sigma T^{\frac{1}{2}}}\right)\right] & \qquad (3.3)\\
- \frac{2k\mu T^{\frac{1}{2}}}{\sigma}\exp\left(\frac{-2k\mu r}{\sigma^2}\right)\left[\phi\left(\frac{(2k+1)r-\mu T}{\sigma T^{\frac{1}{2}}}\right) + \phi\left(\frac{(2k-1)r-\mu T}{\sigma T^{\frac{1}{2}}}\right)\right] &\\
+ \frac{2k\mu}{\sigma^2}\exp\left(\frac{2k\mu r}{\sigma^2}\right)\Bigg[[(2k+1)r+\mu T]\Phi\left(\frac{(2k+1)r+\mu T}{\sigma T^{\frac{1}{2}}}\right) &\\
- 2[2kr+\mu T]\Phi\left(\frac{2kr+\mu T}{\sigma T^{\frac{1}{2}}}\right) &\\
+ [(2k-1)r+\mu T]\Phi\left(\frac{(2k-1)r+\mu T}{\sigma T^{\frac{1}{2}}}\right)\Bigg] &\\
- \frac{2k\mu}{\sigma^2}\exp\left(\frac{-2k\mu r}{\sigma^2}\right)\Bigg[[(2k+1)r-\mu T]\Phi\left(\frac{(2k+1)r-\mu T}{\sigma T^{\frac{1}{2}}}\right) &\\
- 2[2kr-\mu T]\Phi\left(\frac{2kr-\mu T}{\sigma T^{\frac{1}{2}}}\right) &\\
+ [(2k-1)r-\mu T]\Phi\left(\frac{(2k-1)r-\mu T}{\sigma T^{\frac{1}{2}}}\right)\Bigg]\Bigg\}. &
\end{aligned}
$$

If we evaluate (3.3) with $\mu = 0$ we find that

$$P\{R_{0,\sigma^2}(0, T) \leqq r\}$$

$$(3.4) \qquad = 2\Phi\left(\frac{r}{\sigma T^{\frac{1}{2}}}\right) - 1 + 2\sum_{k=1}^{\infty}\left\{(4k-1)\Phi\left(\frac{(2k-1)r}{\sigma T^{\frac{1}{2}}}\right) - 8k\Phi\left(\frac{2kr}{\sigma T^{\frac{1}{2}}}\right) + (4k+1)\Phi\left(\frac{(2k+1)r}{\sigma T^{\frac{1}{2}}}\right)\right\}.$$

This case was previously studied by Feller (1951), Anis (1954), and by Darling and Siegert (1953), whose results are equivalent to (3.4).

The probability of accepting the null hypothesis that $x_1, x_2, \cdots, x_n$ have a common distribution with mean zero is given approximately by (3.4). The probability that $R_{\mu,\sigma^2}(0, T) > r$, or one minus the probability in (3.3), represents the power of this test for the special class of alternatives in which the mean changed before the first observation. An example in which these alternatives might be appropriate is a production process in which an incorrect adjustment was made to some equipment before testing began.

Table 1 contains the appropriate limit $r$ for some values of Type I error $\alpha$ when $\sigma n^{\frac{1}{2}} = 1$. The expression in (3.4) is $1-\alpha$. For other values of $\sigma n^{\frac{1}{2}}$ multiply the given $r$ by $\sigma n^{\frac{1}{2}}$.

TABLE 1

| $\alpha$ | .1 | .05 | .025 | .01 | .005 | .001 |
|---|---|---|---|---|---|---|
| $r$ | 2.241 | 2.498 | 2.734 | 3.023 | 3.227 | 3.662 |

**4. Distribution theory–continuing case.** Let $x_1, x_2, \cdots$ be a sequence of independent observations, ordered in time, which, in the null state, are from a distribution with a known mean $\mu_0$ and known finite variance. (Without loss of generality we take $\mu_0 = 0$.) To detect a change in $\mu$ in either direction at unknown time $m$ we apply the procedure of Page stated in (2.2) with $h = h'$. But even if the mean remains constant at zero, (2.2) will hold for some $k$ with probability one, and action, though inappropriate, is taken. Thus one cannot choose $h$ to control the Type I and Type II errors; instead, one can use the criteria of controlling the average run length, which is the expected number of observations sampled before action is taken.

As in Section 3 we regard the sum $S_k$ as the value at time $t = k$ of a Wiener process $Y(t)$ with $EY(t) = 0$ and $EY^2(t) = 1$. Let $\tau$ be the random variable which represents the time at which $R_{0,1}(0, T)$ first exceeds $r$ as $T \to \infty$. Then the average run length for $\mu = 0$ is approximated by the expectation of $\tau$. We note that

$P\{\tau \leqq t\} = P\{R_{0,1}(0, t) > r\}$. Thus the distribution function of $\tau$ is one minus the expression in (3.4), or

$$F(t) = P\{\tau \leqq t\}$$

$$(4.1) \qquad = 2-2\Phi\left(\frac{r}{t^{\frac{1}{2}}}\right)-2\sum_{k=1}^{\infty}\left\{(4k-1)\Phi\left(\frac{(2k-1)r}{t^{\frac{1}{2}}}\right) - 8k\Phi\left(\frac{2kr}{t^{\frac{1}{2}}}\right)+(4k+1)\Phi\left(\frac{(2k+1)r}{t^{\frac{1}{2}}}\right)\right\}.$$

The series of derivatives is

$$(4.2) \qquad \frac{r}{t^{3/2}}\phi\left(\frac{r}{t^{\frac{1}{2}}}\right)+\frac{r}{t^{3/2}}\sum_{k=1}^{\infty}\left\{(4k-1)(2k-1)\phi\left(\frac{(2k-1)r}{t^{\frac{1}{2}}}\right) -16k^2\phi\left(\frac{2kr}{t^{\frac{1}{2}}}\right)+(4k+1)(2k+1)\phi\left(\frac{(2k+1)r}{t^{\frac{1}{2}}}\right)\right\},$$

which converges uniformly on any finite interval. Thus (4.2) represents the density function of $\tau$. To find the mean value of $\tau$ we take the Laplace transform

$$\psi(\lambda) = \int_0^{\infty} e^{-\lambda t}\, dF(t), \qquad \lambda > 0$$

and evaluate the $\lim_{\lambda\to 0} \psi'(\lambda)$ where $\psi'(\lambda) = d/(d\lambda)\psi(\lambda)$ (see Feller (1966); XIII, (2.6)). The transform is

$$\psi(\lambda) = \int_0^{\infty} e^{-\lambda t}\frac{r}{t^{3/2}}\phi\left(\frac{r}{t^{\frac{1}{2}}}\right)dt$$

$$+\int_0^{\infty}\sum_{k=1}^{\infty} e^{-\lambda t}\frac{(4k-1)(2k-1)r}{t^{3/2}}\phi\left(\frac{(2k-1)r}{t^{\frac{1}{2}}}\right)dt$$

$$-\int_0^{\infty}\sum_{k=1}^{\infty} e^{-\lambda t}\frac{16k^2r}{t^{3/2}}\phi\left(\frac{2kr}{t^{\frac{1}{2}}}\right)dt$$

$$+\int_0^{\infty}\sum_{k=1}^{\infty} e^{-\lambda t}\frac{(4k+1)(2k+1)}{t^{3/2}}\phi\left(\frac{(2k+1)r}{t^{\frac{1}{2}}}\right)dt.$$

By Fubini's theorem we may interchange the operations of summation and integration. Let $q = \exp(-r(2s)^{\frac{1}{2}})$. Then for $\lambda > 0$

$$\psi(\lambda) = q+\sum_{k=1}^{\infty}\{(4k-1)q^{2k-1}-8kq^{2k}+(4k+1)q^{2k+1}\}$$

$$= \frac{4q}{(1+q)^2} = \operatorname{sech}^2(r(\tfrac{1}{2}\lambda)^{\frac{1}{2}}).$$

Differentiating $\psi(\lambda)$ and taking the limit as $\lambda \to 0$ we find $E(\tau) = r^2/2$. All moments of $\tau$ can be found by successive differentiation of $\psi(\lambda)$.

The Laplace transform of the distribution of $\tau$ could also be found by applying the results of Darling and Siegert (1953). To illustrate this approach we will deduce the approximation to the average run length in the non-null case ($\mu \neq 0$) in this manner. Let $G_t(r) = P\{R_{\mu,\sigma^2}(0, t) \leqq r\}$, $g_t(r) = (\partial/\partial r)G_t(r)$, and recall that $F(t) = P\{\tau \leqq t\}$. We noted that $F(t) = 1 - G_t(r)$. Therefore

$$\frac{\partial}{\partial t}F(t) = -\frac{\partial}{\partial t}G_t(r).$$

As before we define $\psi(\lambda)$ as the Laplace transform of $\tau$, and let

$$\chi(\lambda) = \int_0^\infty e^{-\lambda t} g_t(r)\, dt.$$

Then

$$\psi(\lambda) = \int_0^\infty e^{-\lambda t}\left(-\frac{\partial}{\partial t}G_t(r)\right) dt. \tag{4.3}$$

Define $G_0(r) = 0$. Integrating (4.3) by parts we get

$$\psi(\lambda) = -\lambda \int_0^\infty e^{-\lambda t} G_t(r)\, dt$$

so that

$$\chi(\lambda) = -\frac{1}{\lambda}\frac{\partial}{\partial r}\psi(\lambda). \tag{4.4}$$

Applying Theorem (7.1) of Darling and Siegert (1953), as well as their relationship in (3.4) and the results of their Section 5($c$) we find that

$$\chi(\lambda) = -\frac{1}{\lambda}\frac{\partial^2}{\partial r^2}\int_{-r/2}^{r/2} \frac{\exp[\xi_2(x-v)]\sinh(\xi_1 r/2) - \exp[\xi_1(x-v)]\sinh(\xi_2 r/2)}{\sinh(\xi_1 r/2 - \xi_2 r/2)}\, dv$$

where

$$\xi_1 = \frac{-\mu + (\mu^2 + 2\sigma^2\lambda)^{\frac{1}{2}}}{\sigma^2} \quad \text{and} \quad \xi_2 = \frac{-\mu - (\mu^2 + 2\sigma^2\lambda)^{\frac{1}{2}}}{\sigma^2}.$$

After performing the required integration and substituting into (4.4) we find that

$$\psi(\lambda) = \frac{\partial}{\partial r}\left[\frac{-(\mu^2 + 2\sigma^2\lambda)^{\frac{1}{2}}}{\lambda}\frac{\cosh(\mu r/\sigma^2) - \cosh(r(\mu^2 + 2\sigma^2\lambda)^{\frac{1}{2}}/\sigma^2)}{\sinh(r(\mu^2 + 2\sigma^2\lambda)^{\frac{1}{2}}/\sigma^2)}\right]$$

$$= \frac{-(\mu^2 + 2\sigma^2\lambda)^{\frac{1}{2}}}{2\lambda\sinh^2 rq}\left\{\left(\frac{\mu}{\sigma^2} - q\right)\cosh\left(rq + \frac{\mu r}{\sigma^2}\right) - \left(\frac{\mu}{\sigma^2} + q\right)\cosh\left(rq - \frac{\mu r}{\sigma^2}\right) + 2q\right\} \tag{4.5}$$

where

$$q = \frac{(\mu^2 + 2\sigma^2\lambda)^{\frac{1}{2}}}{\sigma^2}.$$

Differentiating (4.5) and taking the limit as $\lambda \to 0$ we find that

$$E(\tau) = \frac{r}{\mu}\coth(\mu r/\sigma^2) - \frac{\sigma^2}{2\mu^2} - \frac{r^2}{2\sigma^2 \sinh^2(\mu r/\sigma^2)}$$

**5. Approximation to power function when mean changes at an unknown time.** It should be emphasized that the preceding work is approximate only in the sense of the usual approximation in sequential analysis. The probabilities that are calculated are the exact probabilities for the continuous process, and are used to approximate what would occur if observations were taken discretely. When the mean changes at an unknown time we cannot obtain results to the same degree of accuracy.

Let $Y(t)$ be the Wiener process and $X(t) = \sigma Y(t) + \mu_t(t-S)$, where $\mu_t = 0$ for $0 \leqq t \leqq S$ and $\mu_t = \mu$ (some fixed value) for $S < t \leqq T$. (If $\mu_t \geqq \mu$ for $S < t \leqq T$ then the resulting test would have even greater power.) Define $R_{\mu_t,\sigma^2}(0, T) = \max_{0 \leqq t \leqq T} X(t) - \min_{0 \leqq t \leqq T} X(t)$. Since in this section $\mu_t$ will only take on the values defined above, we will suppress the subscripts $\mu_t$ and $\sigma^2$ on $R(0, T)$. To find the probability of rejecting the null hypothesis when the mean changes at an unknown time we need to find $P\{R(0, T) \geqq r;\ R(0, S) < r)\}$. Let $X(S) = y$ and let $f(u_S, v_S, y)$ represent the joint density function of $U_S$, $V_S$, and $X(S)$. Then

$$P\{R(0, T) \geqq r;\ R(0, S) < r\}$$
$$= \int_0^r \int_{v_s-r}^0 \int_{u_s}^{v_s} P\{R(0, T) \geqq r \mid u_S, v_S, y\} f(u_S, v_S, y)\, dy\, du_S\, dv_S.$$

However, the probability in the integrand proves to be quite unwieldy, and this approach leads to expressions which are too cumbersome to employ. Thus we will have to content ourselves with an approximation to the power function which is approximate even for the continuous process.

We require the probability that $R(0, T)$ exceeds $r$. This can occur in any of four mutually exclusive ways: if for $S < t \leqq T$, $X(t)$ exceeds $u_S + r$ or is less than $v_S - r$, or if $X(t)$ achieves a new minimum $u < u_S$ and then increases beyond $u + r$ or achieves a new maximum $v > v_S$ and then decreases below $v - r$. Let us assume that the change in the mean is positive (a negative change would be handled similarly). Then the most probable path for which the range would exceed $r$ would be to achieve a new maximum greater than $u_S + r$. Since $P\{R(0, T) > r;\ R(0, S) < r\}$ is greater than $P\{X(t) > u_S + r;\ R(0, S) < r\}$ the latter probability will provide a lower bound for the power function of the continuous process.

Since we condition on the values $u_S$, $v_S$, and $y$, we need the joint density of $U_S$, $V_S$ and $X(S)$. The joint distribution function of $U_S$ and $V_S$ given that $X(S) = y$ is

$$F(u_S, v_S \mid X(S) = y) = P\{V_S \leqq v_S \mid X(S) = y\} - P\{V_S \leqq v_S;\ U_S > u_S \mid X(S) = y\}. \tag{5.1}$$

We find the second probability on the right from Theorem 4.2 of Anderson (1960) and the first by letting $U_S \to -\infty$ in the resulting expression. The joint density function of $U_S$, $V_S$, and $X(S)$ can be shown to be

$$
\begin{aligned}
f(u_S, v_S, y) &= f(u_S, v_S \mid y) \frac{1}{(2\pi\sigma^2 S)^{\frac{1}{2}}} \exp\left(-\frac{y^2}{2\sigma^2 S}\right), \\
&= \frac{2}{\sigma^3 (S^3)^{\frac{1}{2}}} \sum_{k=1}^{\infty} \left\{ 2k(k-1)\left[1 - \frac{(2(kv_S-(k-1)u_S)-y)^2}{\sigma^2 S}\right] \right. \\
&\qquad \cdot \phi\left(\frac{2(kv_S-(k-1)u_S)-y}{\sigma S^{\frac{1}{2}}}\right) \\
&\qquad -2k^2\left[1-\frac{(2k(v_S-u_S)-y)^2}{\sigma^2 S}\right]\phi\left(\frac{2k(v_S-u_S)-y}{\sigma S^{\frac{1}{2}}}\right) \\
&\qquad -2k^2\left[1-\frac{(2k(u_S-v_S)-y)^2}{\sigma^2 S}\right]\phi\left(\frac{2k(u_S-v_S)-y}{\sigma S^{\frac{1}{2}}}\right) \\
&\qquad +2k(k-1)\left[1-\frac{(2(ku_S-(k-1)v_S)-y)^2}{\sigma^2 S}\right] \\
&\qquad \left. \cdot \phi\left(\frac{2(ku_S-(k-1)v_S)-y}{\sigma S^{\frac{1}{2}}}\right)\right\}.
\end{aligned}
\tag{5.2}
$$

Our lower bound for $P\{R(0, T) \geqq r;\ R(0, S) < r\}$ is

$$
\begin{aligned}
&P\{X(t) > u_S + r,\ S < t \leqq T;\ R(0, S) < r\} \\
&= P\left\{Y(t) > \frac{u_S + r - y - \mu(t-S)}{\sigma},\ S < t \leqq T;\ R(0, S) < r\right\} \\
&= \int_0^r \int_{v_S - r}^0 \int_{u_S}^{v_S} P\left\{Y(t) > \frac{u_S + r - y - \mu(t-S)}{\sigma},\ S < t \leqq T \mid u_S, v_S, y\right\} \\
&\qquad \cdot f(u_S, v_S, y)\, dy\, du_S\, dv_S.
\end{aligned}
\tag{5.3}
$$

We now make a change of variable; let $r_S = v_S - u_S$ and $z = y - u_S$ where $0 < z < r_S$ and $-r_S < u_S < 0$. Then (5.3) can be expressed as

$$
\int_0^r \int_z^r \int_{-r_s}^0 P\left[Y(t) > \frac{r - z - \mu(t-S)}{\sigma},\ S < t \leqq T \mid u_S, r_S, z\right] \cdot f(u_S, r_S, z)\, du_S\, dr_S\, dz
\tag{5.4}
$$

where

$$f(u_S, r_S, z) = \frac{2}{\sigma^3(S^3)^{\frac{1}{2}}} \sum_{k=1}^{\infty} 2k(k-1)\left[1-\frac{(2kr_S+u_S-z)^2}{\sigma^2 S}\right]\phi\left(\frac{2kr_S+u_S-z}{\sigma S^{\frac{1}{2}}}\right)$$
$$-2k^2\left[1-\frac{(2kr_S-u_S-z)^2}{\sigma^2 S}\right]\phi\left(\frac{2kr_S-u_S-z}{\sigma S^{\frac{1}{2}}}\right)$$
$$-2k^2\left[1-\frac{(2kr_S+u_S+z)^2}{\sigma^2 S}\right]\phi\left(\frac{2kr_S+u_S+z}{\sigma S^{\frac{1}{2}}}\right)$$
$$+2k(k-1)\left[1-\frac{(2(k-1)r_S-u_S+z)^2}{\sigma^2 S}\right]\phi\left(\frac{2(k-1)r_S-u_S+z}{\sigma S^{\frac{1}{2}}}\right).$$

Performing the inner two integrations yields

$$f(z) = \int_z^r \int_{-r_S}^0 f(u_S, r_S, z)\, du_S dr_S$$
$$= \frac{2}{\sigma S^{\frac{1}{2}}}\left[\phi\left(\frac{z}{\sigma S^{\frac{1}{2}}}\right)+\sum_{k=1}^{\infty}\left\{-(2k-1)\phi\left(\frac{2kr-z}{\sigma S^{\frac{1}{2}}}\right)+(2k+1)\phi\left(\frac{2kr+z}{\sigma S^{\frac{1}{2}}}\right)\right.\right.$$
$$\left.\left.+2k\phi\left(\frac{(2k+1)r-z}{\sigma S^{\frac{1}{2}}}\right)-2k\phi\left(\frac{(2k-1)r+z}{\sigma S^{\frac{1}{2}}}\right)\right\}\right].$$

From equation (17.1) of Shepp (1966) we find

$$P\left\{Y(t) > \frac{r-z-\mu(t-S)}{\sigma},\ S < t \leqq T \mid z\right\}$$
$$= 1-\Phi\left(\frac{r-z-\mu(T-S)}{\sigma(T-S)^{\frac{1}{2}}}\right)+\exp\left(\frac{2\mu(r-z)}{\sigma^2}\right)\Phi\left(\frac{z-r-\mu(T-S)}{\sigma(T-S)^{\frac{1}{2}}}\right)$$

so that our approximation to the power function is

$$\int_0^r\left[1-\Phi\left(\frac{r-z-\mu(T-S)}{\sigma(T-S)^{\frac{1}{2}}}\right)+\exp\left(\frac{2\mu(r-z)}{\sigma^2}\right)\Phi\left(\frac{z-r-\mu(T-S)}{\sigma(T-S)^{\frac{1}{2}}}\right)\right]f(z)\,dz. \tag{5.5}$$

**6. Approximation to the average run length function when mean changes at an unknown time.** The average run length is approximated by the expected value of $\tau$ (the random variable representing the time at which $R(0, T)$ first exceeds $r$ as $T \to \infty$). We will consider the conditional average run length given that we do not reject before time $S$. Then we require $E\{\tau \mid R(0, S) < r\}$. Recall that $P\{\tau \leqq t \mid R(0, S) < r\} = P\{R(0, t) > r \mid R(0, S) < r\}$ which is given by (5.5) with $f(z)$ replaced by $f^*(z) = f(z)/P\{R(0, S) < r\}$.

$$E\{\tau \mid R(0, S) < r\} = \int_S^{\infty}\left\{1-\int_0^r\left[1-\Phi\left(\frac{r-z-\mu(t-S)}{\sigma(t-S)^{\frac{1}{2}}}\right)\right.\right. \tag{6.1}$$
$$\left.\left.+\exp\left(\frac{2\mu(r-z)}{\sigma^2}\right)\Phi\left(\frac{z-r-\mu(t-S)}{\sigma(t-S)^{\frac{1}{2}}}\right)\right]f^*(z)\,dz\right\}dt.$$

Letting $x = t - S$ and interchanging the order of integration we see that (6.1) equals

$$\int_0^r \int_0^\infty \left[ \Phi\left(\frac{r-z-\mu x}{\sigma x^{\frac{1}{2}}}\right) - \exp\left(\frac{2\mu(r-z)}{\sigma^2}\right) \Phi\left(\frac{z-r-\mu x}{\sigma x^{\frac{1}{2}}}\right)\right] dx f^*(z)\, dz,$$

which, after integration by parts, reduces to

$$\int_0^r \int_0^\infty \frac{(r-z)}{\sigma x^{\frac{1}{2}}}\, \phi\left(\frac{r-z-\mu x}{\sigma x^{\frac{1}{2}}}\right) dx f^*(z)\, dz.$$

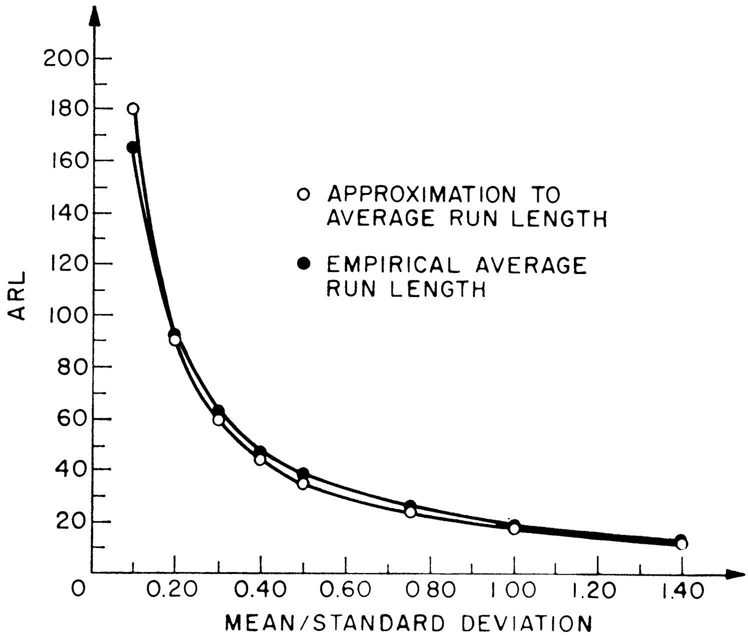


FIG. 1. Conditional average run length function for two-sided procedure when mean changed at the 76th observation.

Thus

$$(6.2)\quad E\{\tau \mid R(0, S) < r\} = \int_0^r \frac{r-z}{\mu} f^*(z)\, dz$$

$$= \frac{r}{\mu} - \frac{2\sigma S^{\frac{1}{2}}}{\mu P\{R(0, S) < r\}} \left[\frac{1}{(2\pi)^{\frac{1}{2}}} + 2 \sum_{k=1}^{\infty} (-1)^k \phi\left(\frac{kr}{\sigma S^{\frac{1}{2}}}\right)\right]$$

where $P\{R(0, S) < r\}$ is given by (3.4).

Selected points on the function (6.2) are plotted in Figure 1 and are compared to the results of sampling studies. In the empirical studies 500 sequences of pseudo-normal deviates were generated on the GE 635 computer for each mean and time of change combination. (Six times of change were studied; only one of which is reproduced here.) In Figure 2 we plot the approximation to $P\{R(0, T) \geqq r \mid R(0, S) < r\}$, the conditional power function given that we do not reject before time $S$. This function is given by (5.5) with $f^*(z)$ replacing $f(z)$. We compare these results to those of the sampling study.

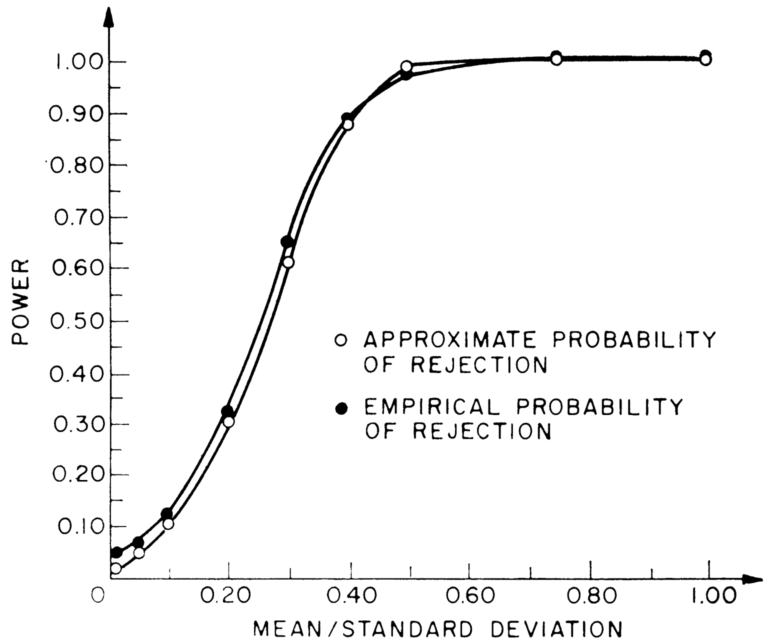


FIG. 2. Conditional power function for truncated two-sided procedure when mean changed at the 101-st of 200 observations.

**7. Unknown variance.** Throughout the preceding sections we have assumed that $\sigma^2$ is known. In industrial applications this is usually a reasonable assumption since large quantities of past data are usually available. We recall that $x_1, x_2, \cdots, x_n$ come from a distribution with variance $\sigma^2$, and we reject the null hypothesis when the range of the partial sums of the $x_i$ exceeds $r$. This is equivalent to rejecting the null hypothesis when the range of the partial sums of the $y_i$ exceeds $r/\sigma$ where $x_i = \sigma y_i$. We now consider the case when $\sigma^2$ is unknown. Suppose at the $j$th observation we estimate $\sigma^2$ by $s_j^2 = (j-1)^{-1} \sum_{i=1}^{j} (x_i - \bar{x}_j)^2$ where $\bar{x}_j = j^{-1} \sum_{i=1}^{j} x_i$. Then it is no longer true that the range for $n_1$ observations exceeding $r/s_{n_1}$ implies the range for $n_2$ observations exceeding $r/s_{n_2}$ where $n_2 > n_1$. However, $s_j^2$ converges to $\sigma^2$ almost surely. Thus as the number of observations becomes large this is not likely to be a serious problem. If $r$ is chosen so that the Type I error is small, then the probability will be small that the range will exceed $r/s_n$ for small $n$. However, should this event occur, it should be viewed with some suspicion.

## REFERENCES

[1] Anderson, T. W. (1960). A modification of the sequential probability ratio test to reduce the sample size. *Ann. Math. Statist.* **31** 165–197.

[2] Anis, A. A. (1954). On the distribution of the range of partial sums of independent random variables. *Proc. Math. Phys. Soc. U.A.R. (Egypt)* **5** 83–89.

[3] Barnard, G. A. (1959). Control charts and stochastic processes. *J. Roy. Statist. Soc. Ser. B.* **21** 239–271.

[4] Chernoff, H. and Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *Ann. Math. Statist.* **35** 999–1018.

[5] Darling, D. A. and Siegert, A. J. F. (1953). The first passage problem for a continuous Markov process. *Ann. Math. Statist.* **24** 624–639.

[6] Donsker, M. D. (1951). An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.* No. 6.

[7] Doob, J. L. (1949). Heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Statist.* **20** 393–403.

[8] Ewan, W. D. (1963). When and how to use Cu-sum charts. *Technometrics* **5** 1–22.

[9] Ewan, W. D. and Kemp, K. W. (1960). Sampling inspection of continuous processes with no autocorrelation between successive results. *Biometrika* **47** 363–380.

[10] Feller, W. (1951). The asymptotic distribution of the range of sums of independent random variables. *Ann. Math. Statist.* **22** 427–432.

[11] Feller, W. (1966). *An Introduction to Probability Theory and its Applications* **2**. Wiley, New York.

[12] Freedman, David A. (1967). Some invariance principles for functionals of a Markov chain. *Ann. Math. Statist.* **38** 1–7.

[13] Freund, Richard A. (1962). Graphical process control. *Industrial Quality Control* **18** 15–22.

[14] Goldsmith, R. L. and Whitfield, H. (1961). Average run lengths in cumulative chart quality control schemes. *Technometrics* **3** 11–20.

[15] Johnson, N. L. (1961). Simple theoretical approach to cumulative sum control charts. *J. Amer. Statist. Assoc.* **56** 835–840.

[16] Page, E. S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–114.

[17] Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika* **42** 523–527.

[18] Page, E. S. (1961). Cumulative sum charts. *Technometrics* **3** 1–9.

[19] Shepp, L. A. (1966). Radon–Nikodym derivatives of Gaussian measures. *Ann. Math. Statist.* **37** 321–354.

[20] Traux, H. M. (1961). Cumulative sum charts and their application to the chemical industry. *Industrial Quality Control* **18** 18–25.