DEPARTMENT OF STATISTICS

The University of Wisconsin
Madison, Wisconsin

Technical Report No. 120

August, 1967

NONPARAMETRIC TESTS FOR SHIFT AT AN UNKNOWN TIME POINT

by

G. K. Bhattacharyya
and
R. A. Johnson

This research was supported in part by Wisconsin Alumni Research Foundation and the National Science Foundation under contract GK-1055.

1. Introduction and summary. This work is an investigation of a nonparametric approach to the problem of testing for a shift in the level of a process occurring at an unknown time point when a fixed number of observations are drawn consecutively in time. This problem frequently arises in quality control. Chernoff and Zacks [2] mention that it may also be considered as being a simplification of a tracking problem. We observe successively the independent random variables $X_1, X_2, \ldots, X_N$ which are distributed according to the continuous cdf F_i, $i = 1, 2, \ldots, N$. An upward shift in the level shall be interpreted to mean that the random variables after the change are stochastically larger than those before. Two versions of the testing problem are studied. The first deals with the case when the initial process level is known and the second when it is unknown. In the first case, we make the simplifying assumption that the distributions F_i are symmetric before the shift and introduce the known initial level by saying that the point of symmetry γ_o is known. Without loss of generality, we set $\gamma_o = 0$. Defining a class of cdf's $\mathfrak{F}_o = \{F\text{: } F \text{ continuous}, F(x) = 1 - F(-x), -\infty < x < \infty\}$, the problem of detecting an upward shift becomes that of testing the null hypothesis

$$H_o\text{: } F_0(x) = F_1(x) = \ldots = F_N(x), \text{ some } F_0 \in \mathfrak{F}_o$$

against the alternative

$$H_1\text{: } F_0(x) = F_1(x) = \ldots = F_m(x) > F_{m+1}(x) = \ldots = F_N(x), \text{ some } F_0 \in \mathfrak{F}_o$$

$$F_{m+1} \neq F_m,$$

where m $(0 \leq m \leq N - 1)$ is unknown.

Sometimes it may be of interest to investigate whether a process level is stable without having precise knowledge about its initial value. In this Situation, the problem of detecting an upward shift in level becomes that of testing the null hypothesis

$$H_o^*: \; F_1(x) = F_2(x) = \ldots = F_N(x)$$

against the alternatives

$$H_1^*: \; F_1(x) = \ldots = F_m(x) \geq F_{m+1}(x) = \ldots = F_N(x), \; F_{m+1} \neq F_m,$$

where m $(1 \leq m \leq N - 1)$ is unknown. The distributions are not assumed to be symmetric.

The testing problem in the case of known initial level has been considered by Page [11], Chernoff and Zacks [2] and Kander and Zacks [7]. Assuming that the observations are initially from a symmetric distribution with known mean γ_o, Page proposes a test based on the variables sgn $(X_i - \gamma_o)$. Chernoff and Zacks assume that the F_i are normal cdf's with constant known variance and they derive a test for shift in the mean through a Bayesian argument. Their approach is extended to the one parameter exponential family of distributions by Kander and Zacks. Except for the test based on signs, all the previous work lies within the framework of parametric statistics. The second formulation of the testing problem, the case of unknown initial level, has not been treated in detail. The only test proposed thus far is the one derived by Chernoff and Zacks for normal distributions with constant known variance. In both problems, our approach generally is to find optimal invariant tests for certain local shift alternatives and then to examine their properties. Our optimality criterion is the maximization of local average power where the average is over the space of the nuisance parameter m with respect to an arbitrary weighting $\{q_i, \; i = 1, 2, \ldots, N: \; q_i \geq 0, \; \sum_{i=1}^{N} q_i = 1\}$. From the Bayesian viewpoint, q_i may be interpreted as the prior probability that X_i is the first shifted variate. Although the average power concept has a Bayesian interpretation, our derivation of locally optimal tests follows quite closely the developments given in Lehmann [8] for the one sample location and the two sample shift problems. Invariant tests with maximum local average power are derived for the case of known initial level in Section 2 and for the case of unknown initial level in Section 3. In both cases the tests are shown to be unbiased for general classes of shift alternatives and for all possible points of shift. The test statistics are distribution-free under the null hypothesis and their large sample distributions are shown to be normal. They all depend on the weight function $\{q_i\}$ and this allows for flexibility in the choice of test with regard to whatever information is available on the possible point of shift. This could vary from complete information where one would use a degenerate weight function to the case of complete ignorance where a choice of uniform weights would seem appropriate. With uniform weights, certain tests in Section 3 reduce to the standard tests for trend while a degenerate weight function leads to the usual two sample tests. In Section 4, we obtain the asymptotic distributions of the test statistics under local translation alternatives and investigate their Pitman efficiencies. Some small sample powers for normal alternatives are given in Section 5.

2. Locally best invariant test (initial process level known). For testing H_o versus H_1, we use invariance considerations to reduce the data and then develop distribution-free tests which maximize local average power against specific translation alternatives. The problem remains invariant under the group of all transformations $x_i' = h(x_i)$, $i = 1, 2, \ldots N$ where h is continuous, odd and strictly increasing. A maximal invariant under the group is $(\underset{\sim}{R}, \underset{\sim}{A})$ where $\underset{\sim}{R} = (R_1, R_2, \ldots, R_N)$ is the vector of ranks of $|X_1|, \ldots, |X_N|$ and $\underset{\sim}{A} = (A_1, A_2, \ldots, A_N)$ with $A_i = 0$ (1) if $X_i < 0$ (>0). If $\alpha = k/2^N N!$, any invariant test of size α will reject H_o for exactly k realizations of $(\underset{\sim}{R}, \underset{\sim}{A})$.

Let F(x) denote the common cdf under H_o. For the subfamily of translation alternatives, $F_{m+1}(x) = F(x - \Delta)$, $\Delta > 0$, the power $\beta(\Delta \mid m)$ depends not only on F and the amount of translation Δ, but also on the nuisance parameter m. In order to remove the parameter m, we turn our attention to the average power $\overline{\beta}(\Delta) = \sum_{i=1}^{N} q_i \, \beta(\Delta \mid i-1)$ where the weights satisfy $q_i \geq 0$ and $\sum_{i=1}^{N} q_i = 1$. The structure of the invariant test which maximizes $\overline{\beta}(\Delta)$ is exhibited in the following theorem.

Theorem 2.1. Let the cdf $F \in \mathcal{F}_0$ possess a density $f(x)$ having the following properties:

(A) $f(x) > 0$ a. e. (Lebesgue) and f is absolutely continuous.

(B) For a sufficiently small $\varepsilon > 0$, there exists a function $H(x)$ with $\int_{-\infty}^{\infty} H(x)\,dx < \infty$ and for almost all x

$$\sup_{|\delta| \leq \varepsilon} \left|\frac{f(x+\delta) - f(x)}{\delta}\right| \leq H(x).$$

Then the invariant test which maximizes the derivative of the average power at $\Delta = 0$ has a rejection region of the form

$$(2.1) \qquad T = \sum_{i=1}^{N} Q_i \quad \operatorname{sgn}(X_i)\, E\left[-\frac{f'(V^{(R_i)})}{f(V^{(R_i)})}\right] > C$$

where the $Q_i = \sum_{m=1}^{i} q_m$ are the cumulative weights and $V^{(1)} < V^{(2)} \ldots < V^{(N)}$ is an ordered sample from a population having density $2f(x)$, $x > 0$. (2.1) also maximizes the average power itself for all sufficiently small $\Delta > 0$.

Proof. Let $n = N - m$ where X_{m+1} is the first shifted variate. With amount of translation Δ, the probability of any specific realization $\underset{\sim}{A} = \underset{\sim}{a}$ is given by

$$(2.2) \qquad P_\Delta(\underset{\sim}{a} \mid m) = 2^{-m}\,[F(-\Delta)]^{n-a_o}[F(\Delta)]^{a_o}$$

where $a_o = \sum_{i=m+1}^{N} a_i$. Due to the symmetry of F, if X has cdf $F(x-\Delta)$, the conditional density of $|X|$ given $X > 0$ is $f(x-\Delta)/F(\Delta)$ and given $X < 0$, it is $f(x+\Delta)/F(-\Delta)$. Using this together with condition (A), we follow Lehmann [8], p. 254, and express the conditional probability of any specific realization $\underset{\sim}{r}$ of $\underset{\sim}{R}$ given $\underset{\sim}{A} = \underset{\sim}{a}$ as

$$(2.3) \qquad P_\Delta(\underset{\sim}{r} \mid \underset{\sim}{a}, m) = \{N!\, 2^{N-m}\,[F(-\Delta)]^{n-a_o}[F(\Delta)]^{a_o}\}^{-1}$$

$$E\left[\prod_{i=m+1}^{N} \frac{f(V^{(r_i)} + (1-2a_i)\,\Delta)}{f(V^{(r_i)})}\right]$$

From (2.2) and (2.3), we obtain

$$(2.4) \qquad P_\Delta(\underset{\sim}{r}, \underset{\sim}{a} \mid m) = (N!\ 2^N)^{-1}\ E[\prod_{i=1}^{N} \frac{f(V^{(r_i)} + b_{mi}\Delta)}{f(V^{(r_i)})}]$$

where $b_{mi} = 0$ for $i \leq m$ and $b_{mi} = (1 - 2a_i)$ for $i > m$. If $P_\Delta(\underset{\sim}{r}, \underset{\sim}{a}) = \sum_{m=1}^{N} q_m P_\Delta(\underset{\sim}{r}, \underset{\sim}{a} \mid m - 1)$, the average power of an invariant test for shift Δ is obtained by summing $P_\Delta(\underset{\sim}{r}, \underset{\sim}{a})$ over all $(\underset{\sim}{r}, \underset{\sim}{a})$ belonging to the critical region.

By Neyman-Pearson's Lemma, it follows that the derivative of the average power function at $\Delta = 0$ is maximized by the rejection region $\frac{\partial}{\partial\Delta} P_\Delta(\underset{\sim}{r}, \underset{\sim}{a}) \mid_{\Delta = 0} > C$. Under condition (B), the dominated convergence theorem allows differentiation of (2.4) under the expectation sign. To see this write (2.4) as $P_\Delta(\underset{\sim}{r}, \underset{\sim}{a} \mid m) = \int_S [\prod_{i=1}^{N} f(v^{(r_i)} + b_{mi}\ \Delta)]\ dv^{(1)} \ldots dv^{(N)}$ where $S = \{v^{(1)}, v^{(2)}, \ldots, v^{(N)}: v^{(1)} < v^{(2)} < \ldots < v^{(N)}\}$. For all $|\Delta| \leq \varepsilon \leq 1$, $G(x) = H(x) + f(x)$ dominates both $f(x + b_{mi}\ \Delta)$ and $|\Delta^{-1}[f(x + b_{mi}\ \Delta) - f(x)]|$ almost everywhere. Using the identity

$$\prod_{i=1}^{N} \alpha_i - \prod_{i=1}^{N} \beta_i = \sum_{i=1}^{N} (\alpha_i - \beta_i) \prod_{j=0}^{i-1} \alpha_j \prod_{k=i+1}^{N+1} \beta_k, \text{ with } \alpha_o = \beta_{N+1} = 1,$$

we then have

$$(2.5) \qquad \Delta^{-1} \left| \prod_{i=1}^{N} f(v^{(r_i)} + b_{mi}\ \Delta) - \prod_{i=1}^{N} f(v^{(r_i)}) \right| \leq N \prod_{i=1}^{N} G(v^{(r_i)})$$

and the right hand side is Lebesgue integrable over Euclidean N space and hence over the subspace S. Differentiating under the expectation in (2.4), it follows after straightforward manipulation that the rejection region is of the form (2.1).

The type of argument used by Lehmann [8], p. 287, shows that the average power is also maximized for all sufficiently small Δ. This completes the proof of the theorem.

For a few specific choices of the distribution F, the test statistics T of (2.1) are given in Table 1. Large values of the test statistic are critical in each case.

Table 1. Tests with locally best average power against translation alternatives in a process with known initial level.

F	Test Statistic		
	General weights	Uniform weights $q_i = 1/N$	Degenerate weights $q_{m+1}=1,\ q_i=0,\ i\neq m+1$
Double exponential	$T_{(1)} = \sum_{i=1}^{N} Q_i\ sgn(X_i)$	$\frac{1}{N}\sum_{i=1}^{N} i\ sgn(X_i)$	$\sum_{i=m+1}^{N} sgn(X_i)$
Logistic	$T_{(2)} = \sum_{i=1}^{N} Q_i\ sgn(X_i)\ R_i$	$\frac{1}{N}\sum_{i=1}^{N} i\ sgn(X_i)\ R_i$	$\sum_{i=m+1}^{N} sgn(X_i)\ R_i$
Normal	$T_{(3)} = \sum_{i=1}^{N} Q_i\ sgn(X_i)E(W^{(R_i)})$*	$\frac{1}{N}\sum_{i=1}^{N} i\ sgn(X_i)E(W^{(R_i)})$	$\sum_{i=m+1}^{N} sgn(X_i)E(W^{(R_i)})$

* $W^{(1)} < W^{(2)} \ldots < W^{(N)}$ is an ordered sample from a χ_1 distribution.

The uniform weighting used in the third column allows for the possibility that a shift might occur before the observations are taken. Chernoff and Zacks [2] and also Kander and Zacks [7] have assumed that the known process level corresponds to the distribution of X_1 and this has led them to the uniform prior $q_i = (N - 1)^{-1}$, $i = 2, 3, \ldots, N$. Apart from this minor difference, our optimal invariant test $T_{(1)}$ with uniform weighting coincides with Kander and Zack's test which is based on the posterior likelihood ratio for a binomial sample. Some power comparisons between this and Page's test have been made in [2]. Hájek [5] and Adichie [1] have studied the large sample properties of test statistics of the form (2.1) which arise in connection with a linear regression model having the Q_i's as values of the independent variable.

When the point of possible shift is known, the weight function becomes $q_{m+1}=1$, $q_i=0$ for $i\neq m+1$. The three test statistics for this case are given in the fourth column of Table 1. $T_{(1)}$ reduces to the sign test for location based on the observations $X_{m+1}, \ldots, X_N$. The forms of $T_{(2)}$ and $T_{(3)}$ in this case are structurally similar to the Wilcoxon signed rank and the one sample normal score tests based on the above $N - m$ observations. The intrinsic difference is that for $T_{(2)}$ and $T_{(3)}$, the ranking is considered over all N ovservations. It is interesting to note that in a two sample shift problem where one sample is known to be from a distribution symmetric about 0, the locally optimal invariant tests for logistic and normal distributions are $T_{(2)}$ and $T_{(3)}$ and not the Wilcoxon and the normal score tests. The reason is that a smaller invariance group is appropriate here.

We now investigate the unbiasedness of the class of tests (2.1) and more generally, of any tests of the form

$$(2.6) \qquad T(\underset{\sim}{X}) = \sum_{i=1}^{N} Q_i \operatorname{sgn}(X_i)\, U(R_i)$$

where $0 \leq Q_1 \leq Q_2 \cdots \leq Q_N \leq 1$ is a given set of constants and $U(\cdot)$ is a function of the ranks of the $|X_i|$, $i = 1, 2 \ldots N$.

<u>Theorem</u> 2.2. <u>If</u> $U(\cdot)$ <u>is a nondecreasing function, any test which rejects</u> H_o <u>for large values of</u> $T(\underset{\sim}{x})$ <u>is unbiased for testing</u> H_o <u>against</u> H_1.

<u>Proof</u>. Let m $(0 \leq m \leq N - 1)$ be arbitrary but fixed. Define a class of mappings $\mathcal{C}$: $(x_1, x_2, \ldots, x_N) \to (x_1', x_2', \ldots, x_N')$ by $x_i' = x_i$ for $i \leq m$ and $x_i' = h(x_i)$ for $i > m$ where h is continuous, nondecreasing and $h(x) \geq x$ for all x. For any cdf $\prod_{i=1}^{N} F_i$ under H_1, there exists an $F_o \in \mathcal{F}_o$ and an h such that if $(X_1, X_2, \ldots, X_N)$ is distributed as $\prod_{i=1}^{N} F_o$, $(X_1', X_2', \ldots, X_N')$ will be distributed as $\prod_{i=1}^{N} F_i$. It is then sufficient to show that for each map of $\mathcal{C}$, $T(\underset{\sim}{X}') \geq T(\underset{\sim}{X})$ a.e. (Lebesgue), ([8], p. 246).

Consider first a point x where the map is sign preserving in addition to having the above properties. Let $\underset{\sim}{r}$ and $\underset{\sim}{r}'$ denote the vectors of ranks of the absolute values for $\underset{\sim}{x}$ and $\underset{\sim}{x}^{+}$ respectively. Introduce the index sets

$$I_1 = \{i: x_i > 0, i \le m\} \qquad I = I_1 \cup I_2$$

the stepwise procedure of first mapping the points $\{x_i: i \in J_o\}$ to the left of zero but closer to zero than any other $x_i^{'}$, then moving them across zero and finally to $\{x_i^{'}: i \in J_o\}$, we see that at each stage the value of T cannot decrease. This completes the proof of the theorem.

The Q_i represent cumulative weights and hence are nondecreasing. Therefore any test of the form (2.1) is unbiased for every weight function $\{q_i\}$ provided that $E[-f^{'}(V^{(i)})/f(V^{(i)})]$ is nondecreasing in i. In particular, the tests in Table 1 are all unbiased.

Except for $T_{(1)}$, any statistic of the form (2.1) will generally have a sample space consisting of $2^N N!$ points so that the setting of the exact critical region might be very difficult even for moderate sample sizes. To obtain a large sample approximation to the null distribution, consider the sequence of test statistics $T_N = \sum_{i=1}^{N} Q_{Ni} \text{ sgn } (X_{Ni}) E[-g^{'}(V^{(R_{Ni})})/g(V^{(R_{Ni})})]$ where g is a known density having cdf $G \in \mathcal{F}_o$ and satisfying the conditions of Theorem 2.1. Let $Z_{N1} < Z_{N2} < \ldots < Z_{NN}$ be an ordered sample from a uniform distribution on (0, 1) and define a function $\psi(u)$ on $0 < u < 1$ by

$$(2.9) \qquad \psi(u) = -g^{'}(G^{-1}(\frac{u+1}{2}))/g(G^{-1}(\frac{u+1}{2})).$$

In terms of ψ the test statistic T_N can be written as $T_N = \sum_{i=1}^{N} Q_{Ni} \text{ sgn } (X_{Ni}) E\psi(Z_{Ni})$. Under H_o, the distribution of T_N depends only on the choice of g and the weight function $\{q_{Ni}\}$ and not on the particular population cdf. We may therefore assume that the common cdf of X_i under H_o is G. Define a class of cdf's by

$$(2.10) \qquad \mathcal{F} = \{F: \int_{-\infty}^{\infty} (\frac{f^{'}(x)}{f(x)})^2 f(x)\, dx < \infty, \text{ (A) and (B) of Theorem 2.1 hold}\}.$$

The following theorem is a direct consequence of Theorem 7.1 of Hájek [5].

Theorem 2.3. *If* $G \in \mathcal{F}_o$ *is symmetric and if the sequence of weights* $\{q_{Ni}\}$ *satisfies* $\lim_{N\to\infty} \sum_{i=1}^{N} Q^2_{Ni}/N = b^2$, $0 < b^2 < \infty$, *then under* H_o,

where $Q_i = \sum_{m=1}^{i} q_m$ and $V^{(1)} < V^{(2)} < \ldots < V^{(N)}$ is an ordered sample from F.

The proof is similar to that of Theorem 2.1 and hence is omitted.

The simplified forms of the test statistic (3.1) for logistic, normal and double exponential distributions are $T^{(1)} = \sum_{i=1}^{N} Q_i S_i$, $T^{(2)} = \sum_{i=1}^{N} Q_i \, E_\Phi (V^{(S_i)})$ and $T^{(3)} = \sum_{i=1}^{N} Q_i \, E[\operatorname{sgn} W^{(S_i)}]$ respectively, where $E_\Phi (V^{(S_i)})$ are the normal scores and $W^{(1)} < W^{(2)} < \ldots < W^{(N)}$ is an ordered sample from the double exponential distribution.

Chernoff and Zacks [2] obtained the test $\sum_{i=1}^{N} (i - 1)(X_i - \bar{X}) > C$ from the posterior likelihood ratio for normal observations with known variance. For the special case of uniform weights the test statistics (3.1) have the same structure except that functions of ranks are involved instead of the actual observations. Note also that $T^{(1)}$ becomes $\sum_{i=1}^{N} (i - 1) S_i$ and the test is equivalent to Spearman's rank correlation test for trend. Because of this correspondence it is expected that our tests would perform well even when more than one jump occurs in the same direction. With the weight function $q_{m+1}=1$, $q_i=0$, $i \neq m+1$. $T^{(1)}$ and $T^{(2)}$ reduce to the standard two sample Wilcoxon and normal score tests respectively.

Theorem 3.2. If $U(\cdot)$ is a nondecreasing function, any test which rejects H_o^* for large values of $M(\underset{\sim}{X}) = \sum_{i=1}^{N} Q_i \, U(S_i)$ is unbiased for testing H_o^* vs. H_1^*.

Proof. Let m $(1 \leq m < N)$ be arbitrary but fixed. Consider the same class of transformations $\mathcal{T}$ introduced in the proof of Theorem 2.2. It is sufficient again to show that $M(\underset{\sim}{X}') \geq M(\underset{\sim}{X})$ a. e. Let $\underset{\sim}{s}' = (s_1', s_2', \ldots, s_N')$ be the vector of ranks of the x_i'. Clearly $i > m$ $(\leq m) \Rightarrow s_i' \geq (\leq)\, s_i \Rightarrow U(s_i') \geq (\leq)\; U(s_i)$. Hence

(3.2) $$T(\underset{\sim}{x}') - T(\underset{\sim}{x}) = \sum_{i=1}^{m} Q_i \, [U(s_i') - U(s_i)] + \sum_{i=m+1}^{N} Q_i \, [U(s_i') - U(s_i)]$$

$$\geq Q_m \sum_{i=1}^{m} [U(s_i') - U(s_i)] + Q_{m+1} \sum_{i=m+1}^{N} [U(s_i') - U(s_i)]$$

$$= (Q_{m+1} - Q_m) \sum_{i=m+1}^{N} [U(s_i') - U(s_i)] \geq 0. \quad \text{QED.}$$

We now consider the asymptotic distribution of the class of statistics

$$(3.3) \qquad T_N = \sum_{i=1}^{N} Q_{Ni}\, E[-g'(V^{(S_{Ni})})/g(V^{(S_{Ni})})].$$

Defining a function ψ by

$$(3.4) \qquad \psi(u) = -g'(G^{-1}(u))/g(G^{-1}(u)), \qquad 0 < u < 1$$

and letting $Z_{N1} < \dots Z_{NN}$ be an ordered sample from the uniform distribution on $(0, 1)$, the test statistic T_N can be expressed as $T_N = \sum_{i=1}^{N} Q_{Ni}\, E[\psi(Z_{Ni})]$. The next theorem follows directly from Hájek [4], Section 6.

<u>Theorem</u> 3.3. Let $\bar{Q}_N = \sum_{i=1}^{N} Q_{Ni}/N$. <u>If</u> $G \in \mathcal{F}$ <u>where</u> $\mathcal{F}$ <u>is defined by</u> (2.10) <u>and if the sequence of weights</u> $\{q_{Ni}\}$ <u>satisfies</u>

$$(A_1) \qquad \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} (Q_{Ni} - \bar{Q}_N)^2 = c^2, \qquad 0 < c^2 < \infty$$

$$(A_2) \qquad \lim_{N\to\infty} \frac{\max_{1 \le i \le N} (Q_{Ni} - \bar{Q}_N)^2}{\sum_{i=1}^{N} (Q_{Ni} - \bar{Q}_N)^2} = 0$$

<u>then under</u> H_0^*,

$$\frac{T_N - E(T_N)}{N^{1/2}\, c\, [\int_0^1 \psi^2(u)\, du]^{1/2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

4. <u>Asymptotic distribution under local alternatives and Pitman ARE</u>. Although desirable, an exact power comparison of our tests with those of [11] and [2] for various parent distributions would involve tremendous computational difficulties even for moderate sample sizes. Consequently, we devote this section to the derivation of the Pitman asymptotic relative efficiency (ARE). The usefulness of this measure in our time series situation is somewhat questionable because the assumption of a single shift in the process level may make little sense when the sample size can be increased only by taking observations over an extended period of time. However, this objection could be ruled out in many cases where it is

possible to increase the size by sampling more frequently in a fixed time period. We will treat in detail the class of tests (3.3). The development of the corresponding results for the tests of Section 2 is similar.

Define a sequence of local translation alternatives $\{K_N\}$ by

$$K_N: \quad F_i(x) = F(x), \ i = 1, 2, \ldots, m \qquad F \in \mathcal{F}$$

$$= F(x - \theta N^{-1/2}), \ i = m+1, \ldots, N$$

$$\lim_{N\to\infty} (m/N) = \lambda, \quad 0 < \lambda < 1.$$

where $\mathcal{F}$ is defined by (2.10). Let $\psi_N\left(i/(N+1)\right) = E[-g'(V^{(Ni)})/g(V^{(Ni)})]$, where $V^{(N1)} < V^{(N2)} < \ldots < V^{(NN)}$ is an ordered sample from $G \in \mathcal{F}$. Noting that $\sum_{i=1}^{N} \psi_N\left(S_{Ni}/(N+1)\right)$ is constant for every N, we express the test statistic (3.3) as

(4.2) $$S_N^o = N^{-1/2} \sum_{i=1}^{N} (Q_{Ni} - \overline{Q}_N)\, \psi_N\left(S_{Ni}/(N+1)\right).$$

Set $d_\psi^2 = \int_0^1 \psi^2(u)\,du$ and $d_\phi^2 = \int_0^1 \phi^2(u)\,du$ where $\psi(u)$ is defined by (3.4) and $\phi(u)$ is the same function with g replaced by f and G by F.

<u>Theorem</u> 4.1. <u>Let</u> $G \in \mathcal{F}$ <u>where</u> $\mathcal{F}$ <u>is defined by</u> (2.10). <u>If the sequence of weights</u> $\{q_{Ni}\}$ <u>satisfies</u>

(A_3) $$\lim \sum_{i=m+1}^{N} (Q_{Ni} - \overline{Q}_N)/N = a < \infty$$

<u>in addition to</u> (A_1) <u>and</u> (A_2) <u>of Theorem</u> 3.3, <u>then</u> $\lim \mathcal{L}(S_N^o \mid K_N) = \mathcal{N}(\mu, c^2 d_\psi^2)$, where

(4.3) $$\mu = \theta\, a \int_0^1 \phi(u)\, \psi(u)\, du.$$

<u>Proof.</u> The proof uses the principle of contiguity and is methodically based on Hájek [5]. The important difference is that the coefficients $(Q_{Ni} - \overline{Q}_N)$ occurring in the test statistic (4.2) do not appear in the alternatives $\{K_N\}$ while in [5]

they do. We will sketch the main steps leaving out the details. Introduce $s(x) = f^{1/2}(x)$, $U_N = -N^{-1/2} \sum_{i=m+1}^{N} [f'(X_i)/f(X_i)]$, $W_N = 2 \sum_{i=1}^{N} [r_{Ni}^{1/2}(X) - 1]$ and $L_N = \sum_{i=1}^{N} \log r_{Ni}$, where

$$(4.4) \qquad r_{Ni}(x) = \begin{cases} 1, & i = 1, 2, \ldots, m \\ \dfrac{f(x - \theta N^{-1/2})}{f(x)}, & i = m+1, \ldots, N. \end{cases}$$

We have

$$(4.5) \qquad E(U_N \mid H_o^*) = 0, \ \ell im \ \mathrm{Var}\,(U_N \mid H_o^*)/(1 - \lambda)\, d_\phi^2 = 1$$

and the application of Lemma 4.3 of [5] yields

$$(4.6) \qquad \ell im \ 4E(W_N \mid H_o^*)/(1 - \lambda)\, \theta^2 d_\phi^2 = 1.$$

Further

$$(4.7) \qquad E[(W_N - E(W_N) - \theta U_N)^2 \mid H_o^*]$$
$$\leq 4 \sum_{i=m+1}^{N} E\Big[\Big(\frac{s(X_i - \theta N^{-1/2})}{s(X_i)} - 1 - \frac{1}{2}\frac{\theta}{N^{1/2}}\frac{f'(X_i)}{f(X_i)}\Big)^2 \mid H_o^*\Big]$$
$$= \frac{4\theta^2}{N} \sum_{i=m+1}^{N} \int_{-\infty}^{\infty} \Big[\frac{s(x - \theta N^{-1/2}) - s(x)}{\theta N^{-1/2}} - s'(x)\Big]^2 dx.$$

Using (4.5) and (4.7) together with Lemma 4.3 of [5], we obtain

$$(4.8) \qquad W_N + \tfrac{1}{4}(1 - \lambda)\, \theta^2 d_\phi^2 + \theta U_N \xrightarrow{P_{H_o^*}} 0,$$

and then by application of the central limit theorem to the sequence $\{U_N\}$, it follows that

$$(4.9) \qquad \ell im \mathcal{L}(W_N \mid H_o^*) = \mathcal{N}(-\tfrac{1}{4}(1 - \lambda)\, \theta^2 d_\phi^2, (1 - \lambda)\, \theta^2 d_\phi^2).$$

The conditions of Lemma 4.2 of [5] are satisfied. Therefore

$$(4.10) \qquad W_N - L_N \xrightarrow{P_{H_o^*}} \tfrac{1}{4}(1 - \lambda)\, \theta^2\, d_\phi^2,$$

$$\lim \mathcal{L}(L_N \mid H_o^*) = \mathcal{N}(-\tfrac{1}{2}(1 - \lambda)\, \theta^2\, d_\phi^2,\ (1 - \lambda)\, \theta^2\, d_\phi^2)$$

and the probability measures are contiguous. The proof of the theorem may be completed by showing that the limiting joint distribution of S_N^o and L_N under H_o^* is bivariate normal with correlation coefficient

$$(4.11) \qquad \rho = \frac{a}{c(1 - \lambda)^{1/2}} \int_0^1 \frac{\phi(u)\ \psi(u)\ du}{d_\phi\ d_\psi}\ .$$

Under H_o^*, S_N^o can be approximated in the mean square (c.f. [4]) by $S_N^* = N^{-1/2} \sum_{i=1}^{N} (Q_{Ni} - \overline{Q}_N)\ \psi(F(X_i))$. By (4.9) and (4.10), we have

$$L_N + \tfrac{1}{2}(1 - \lambda)\ \theta^2\ d_\phi^2 + \theta\ U_N \xrightarrow{P_{H_o^*}} 0$$

so that

$$\lim \mathcal{L}(S_N,\ L_N \mid H_o^*) = \lim \mathcal{L}(S_N^*,\ -\theta\, U_N - \tfrac{1}{2}(1 - \lambda)\ \theta^2\ d_\phi^2 \mid H_o^*).$$

Applying the bivariate central limit theorem (Cramér [3], p. 114) to (S_N^*, U_N) and using the conditions (A_2) and (A_3), we complete the proof.

In the special case of uniform weights, we have $\overline{Q}_N = 1/2$ and it is easy to see that the conditions of Theorem 4.1 are satisfied with $a = \lambda(1 - \lambda)/2$ and $c^2 = 1/12$. Under $\{K_N\}$ the limiting distribution of S_N^o is therefore $\mathcal{N}\left(\frac{\lambda(1 - \lambda)}{2} \int_0^1 \phi(u)\ \psi(u)\ du,\ d_\psi^2/12\right)$. In order to arrive at the usual expressions for the ARE, we shall assume that the conditions of Lemma 3 of [6] are also satisfied. Under these additional conditions, the application of Theorem 4.1 to $T_N^{(1)} = (N - 1)^{-1} \sum_{i=1}^{N} (i - 1)\ S_i$ and $T_N^{(2)} = (N - 1)^{-1} \sum_{i=1}^{N} (i - 1)\ E_\phi(V^{(S_i)})$ yields

$$(4.12)\qquad \ell\mathrm{im}\,\mathcal{L}\left(\frac{2}{N^{1/2}}\left[\frac{T_N^{(1)}}{N+1} - \frac{N}{4}\right] \mid K_N\right) = \mathcal{N}\left(\theta\,\lambda(1-\lambda)\int_{-\infty}^{\infty} f^2(x)\,dx,\ \frac{1}{36}\right)$$

$$\ell\mathrm{im}\,\mathcal{L}\left(\frac{2\,T_N^{(2)}}{N^{1/2}} \mid K_N\right) = \mathcal{N}\left(\theta\,\lambda(1-\lambda)\int_{-\infty}^{\infty} \frac{f^2(x)\,dx}{\phi[\Phi^{-1}(F(x))]},\ \frac{1}{3}\right).$$

When the initial process level is unknown Chernoff-Zack's test statistic has the form $Z_N = \sum_{i=1}^{N} (i-1)(X_i - \bar{X})$. Application of this test to normal populations requires the knowledge of the standard deviation σ. With σ unknown, a Studentized form $Z_N^* = (N-2)^{1/2} Z_N / (D_N S_e)$, with $D_N^2 = N(N^2-1)/12$ and $S_e^2 = \sum_{i=1}^{N} (X_i - \bar{X})^2 - Z_N^2/D_N^2$, may be used. Under normality, the null distribution of Z_N^* is student's t with $(N-2)$ d.f. The asymptotic distribution of Z_N and Z_N^* under the sequence $\{K_N\}$ is given in the following theorem

Theorem 4.2. *If for some* $\delta > 0$, *F has* $(2+\delta)$*-th absolute moment then*

$$(4.13)\qquad \ell\mathrm{im}\,\mathcal{L}\left(\frac{Z_N}{D_N\sigma} \mid K_N\right) = \ell\mathrm{im}\,\mathcal{L}(Z_N^* \mid K_N) = \mathcal{N}\left(\frac{\theta\sqrt{3}\,\lambda(1-\lambda)}{\sigma},\ 1\right).$$

Proof. Apply Liapounov's central limit theorem to the sequence of random variables $Y_{Ni} = (i - \frac{N+1}{2})(X_i - \nu)$, $i = 1, 2, \ldots, m$; $Y_{Ni} = (i - \frac{N+1}{2})(X_i - \nu - \theta N^{-1/2})$, $i = m+1, \ldots, N$ where $\nu = \int_{-\infty}^{\infty} x\,dF(x)$, and note that $S_e^2/(N-2) \xrightarrow{P} \sigma^2$ under K_N.

It follows that for uniform weights the ARE of the test T_N of (3.3) relative to Chernoff-Zack's test is given by

$$(4.14)\qquad e_{T:Z} = \sigma^2 d_\psi^{-2} \left(\int_0^1 \phi(u)\,\psi(u)\,du\right)^2.$$

For the particular tests $T_N^{(1)}$ and $T_N^{(2)}$ this reduces to

$$e_{T^{(1)}:Z} = 12\sigma^2 \left(\int_{-\infty}^{\infty} f^2(x)\,dx\right)^2, \quad e_{T^{(2)}:Z} = \sigma^2 \left\{\int_{-\infty}^{\infty} \frac{f^2(x)\,dx}{\phi[\Phi^{-1}(F(x))]}\right\},$$

and these are precisely the ARE of the two sample Wilcoxon and the normal score tests relative to the t-test.

The selection of a test T_N of the form (3.3) or equivalently of an S_N^o involves the choice of a function ψ defined through a density g as well as a weight function $\{q_i\}$. If two such tests, T_N and T_N^* defined through ψ and ψ^*, are based on identical or asymptotically equivalent weight functions(i.e. a and c of (A_2) and (A_3) are equal), Theorem 4.1 shows that their ARE is given by

$$(4.15) \qquad e_{T:\,T^*} = \left[\frac{d^*_\psi}{d_\psi} \frac{\int_0^1 \phi(u)\ \psi(u)\ du}{\int_0^1 \phi(u)\ \psi^*(u)\ du}\right]^2$$

which is independent of the particular weights used. Since this expression holds for a degenerate weight function, the ARE is equal to that of the standard two sample rank order tests for shift. Thus the investigation of the ARE of our tests poses no new problem if the weight functions are the same or asymptotically equivalent.

It is also of interest to study the sensitivity of the ARE in relation to the choice of the weight function. Suppose T and T' are two tests defined through the same ψ-function but involve two different weight functions $\{q_i\}$ and $\{q_i'\}$ which satisfy the conditions (A_1), (A_2) and (A_3) with the limits (a, c^2) and (a', c'^2) respectively. From Theorem 4.1, we obtain $e_{T:\,T'} = (ac'/a'c)^2$ which is independent of ψ. Suppose that T' has the degenerate weight $q'_{m+1}=1$, $q'_i=0$, $i\neq m+1$. which one uses when the possible point of shift m is known. Under the assumption $m/N \to \lambda$ as $N \to \lambda$, we then have $a' = c'^2 = \lambda(1 - \lambda)$. If we neglect the fact that m is known and use the test T with uniform weights, we will have $a = \lambda(1 - \lambda)/2$, $c^2 = 1/12$. The ARE is $e_{T:\,T'} = 3\lambda(1 - \lambda) \leq 3/4$. This indicates that the loss of efficiency incurred in using a uniform weight instead of the correct degenerate weight is at least 25% and could be much higher if the point of shift is near the beginning or the end of the observation period. Some small sample power comparisons for different choices of weight function are given in the next section.

For the sake of completeness, we state the asymptotic distribution of the test statistic $T_N = \sum_{i=1}^{N} Q_{Ni}\ \mathrm{sgn}\ (X_{Ni})\ E\ \psi(Z_{Ni})$ of Section 2 under the sequence of alternatives $\{K_N\}$ with the additional assumption that F is symmetric. In this case ψ is defined by (2.9).

Theorem 4.3. Let F and G be symmetric and members of $\mathcal{J}$ where $\mathcal{J}$ is defined by (2.10). If the sequence of weights $\{q_{Ni}\}$ satisfies

$$\ell im \sum_{i=1}^{N} Q_{Ni}^2/N = b^2 < \infty \text{ and } \ell im \sum_{i=m+1}^{N} Q_{Ni}/N = \xi < \infty$$

then

$$(4.16) \qquad \ell im \mathcal{L}(T_N | K_N) = \mathcal{N}(\theta\ \xi \int_0^1 \phi(u)\ \psi(u)\ du,\ b^2\ d_\psi^2).$$

The proof is similar to that of Theorem 4.1.

5. Small sample power. Determination of the exact power of our rank order tests being extremely difficult even for moderately large sample sizes, we present here some power computations for very small sample sizes. The power of the test $T^{(1)} = \sum_{i=1}^{N} Q_i\ S_i$ for testing H_o^* vs. H_1^* is calculated in the special case of translations in the distribution $F_1(x) = \Phi(x)$ where Φ is the standard normal cdf. Let us begin by examining the relationship between the power and the amount of translation Δ when the weights $\{q_i\}$ are uniform.

For test size α and sample size N, the critical region contains the rank orders $\underset{\sim}{S} = (S_1, S_2, \ldots, S_N)$ having the $\alpha \cdot N!$ largest values of $\sum_{i=1}^{N} i\ S_i$. Randomization is necessary when $\alpha \cdot N!$ is not an integer or when ties occur in the value of the test statistic at the boundary of the critical region. Fixing the point of shift m+1, we proceed by coding the critical rank vectors into the vectors $\underset{\sim}{Z} = (Z_1, Z_2, \ldots, Z_N)$ according to the rule $Z_i = 0$ if $i \in \{S_1, S_2, \ldots, S_m\}$ and $Z_i = 1$ otherwise. The probabilities of the Z vectors under various normal translation alternatives are tabulated by Milton [9] to nine decimal places. Power is computed by using Table A of [9] and the fact that m! (N - m)! different rank orders yield the same Z. Table 2 gives the power under different amounts of translation Δ and points of shift m+1. Only the values of m which are $\leq N/2$ are tabulated. The powers for other values of m follow by symmetry.

Table 2. Power of the test $\sum_{i=1}^{N} i\, S_i$ for normal translation alternatives.

$\alpha = .05$

N	m	Δ			
		0.2	0.8	1.5	3.0
4	1	.060	.095	.135	.181
4	2	.064	.116	.182	.268
5	1	.060	.094	.132	.174
5	2	.067	.136	.232	.365
6	1	.059	.092	.127	.166
6	2	.068	.141	.244	.384
6	3	.072	.170	.327	.572

Since $T^{(1)}$ is designed for the situation where the process level is unknown, a comparison of the power with Page's test [11] would not be relevant. A comparison of $\sum_{i=1}^{N} i\, S_i$ with the Studentized form Z_N^* of Chernoff and Zack's test would be interesting, but this has been deferred because of the difficulty in obtaining the distribution of Z_N^* in a convenient form under the above alternatives.

For the situation where the initial process level is known, some power comparisons between the test $T_{(1)} = \sum_{i=1}^{N} i \operatorname{sgn}(X_i)$ and Page's test were made by Chernoff-Zacks [2] for normal alternatives. $T_{(1)}$ was found to have slightly more power unless the point of shift is near either end in which case Page's test performs better. It would be interesting to compare these tests with the tests $T_{(2)}$ and $T_{(3)}$ of Table 1. However the powers of the latter tests are difficult to compute due to the absence of tables of rank order probabilities for the absolute values of observations from a normal population. Perhaps power would have to be studied by Monte Carlo techniques.

To illustrate the effect of the selection of weights $\{q_i\}$ on the power, we again consider the test $T^{(1)} = \sum_{i=1}^{N} Q_i S_i$ of Section 3. With sample size N = 5, four systems of weights are chosen for comparison. These range from the degenerate (0, 0, 1, 0, 0) which corresponds to the known point of shift m+1 = 3 to the uniform (0, 1/4, 1/4, 1/4, 1/4) which corresponds to the complete ignorance of the point of shift. Powers of each test for normal translation alternatives are calculated as above and are presented in Table 3. As one would expect, the power is maximum for the choice of the correct degenerate weighting and it falls off with the approach towards the uniform weighting. In the same manner, the entries of Table 2 may be compared to the powers of the corresponding Wilcoxon tests which are available in Milton [10].

Table 3. Effect of the weight function on power of $\sum_{i=1}^{N} Q_i S_i$.

N = 5, m = 2, α = .10

Weight function	Δ			
	0.2	0.8	1.5	3.0
(0, 0, 1, 0, 0)	.137	.296	.540	.921
$(0, \frac{2}{11}, \frac{6}{11}, \frac{2}{11}, \frac{1}{11})$	.135	.283	.498	.813
$(0, \frac{2}{9}, \frac{4}{9}, \frac{2}{9}, \frac{1}{9})$	.135	.278	.484	.777
$(0, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$	.131	.251	.407	.602

The study of the small sample power of the tests derived in this paper is being continued and the results will be communicated later.

REFERENCES

[1] ADICHIE, J. N. (1967). Asymptotic efficiency of a class of nonparametric tests for regression parameters. Ann. Math. Statist. 38 884-893.

[2] 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.

[3] CRAMÉR, H. (1936). Random Variables and Probability Distributions (2nd edition). Cambridge Univ. Press.

[4] HÁJEK, J. (1961). Some extensions of the Wald-Wolfowitz-Noether theorem. Ann. Math. Statist. 32 506-523.

[5] HÁJEK, J. (1962). Asymptotically most powerful rank order tests. Ann. Math. Statist. 33 1124-1147.

[6] HODGES, J.L. Jr. and LEHMANN, E.L. (1961). Comparison of the normal scores and Wilcoxon tests. Proc. Fourth Berk. Symp. Math. Statist. Prob. 1 307-317. Univ. of California Press.

[7] 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.

[8] LEHMANN, E.L. (1959). Testing Statistical Hypotheses. Wiley, New York.

[9] MILTON, R. C. (1965). Rank order probabilities: Two-sample normal shift alternatives. Technical Report No. 53. Dept. of Stat., Univ. of Minnesota.

[10] MILTON, R. C. (1966). Rank order probabilities: Two-sample normal shift alternatives. Technical Report No. 53a. Dept. of Stat., Univ. of Minnesota.

[11] PAGE, E. S. (1955). A test for a change in a parameter occurring at an unknown point. Biometrika 42 523-526.