Testing a Sequence of Observations for a Shift in Location

DOUGLAS M. HAWKINS*

A possible alternative to the hypothesis that the sequence $X_1, X_2, \ldots, X_n$ are iid $N(\xi, \sigma^2)$ random variables is that at some unknown instant the expectation ξ shifts. The likelihood ratio test for the alternative of a location shift is studied and its distribution under the null hypothesis found. Tables of standard fractiles are given, along with asymptotic results.

KEY WORDS: Segmentation; Location shift; Testing for change-point.

1. INTRODUCTION

Suppose that $X_1, X_2, \ldots, X_n$ is a sequence of observations ordered in time. It may be reasonable to model the X_t by

$$X_t = \xi(t) + e_t \ , \quad t = 1, \ldots, n \ ,$$

where

$$\begin{aligned} \xi(t) &= \xi_1 \ , \quad t \le \theta_1 \ , \\ &= \xi_2 \ , \quad \theta_1 < t \le \theta_2 \ , \\ &\ \ \vdots \\ &= \xi_r \ , \quad \theta_{r-1} < t \ . \end{aligned} \tag{1.1}$$

The θ_i are parameters representing the epochs at which shifts in the location parameters ξ_i occur, and the e_t are independent errors. It is assumed that the θ_i and ξ_i are unknown and that r is possibly also unknown. There are two important inferential problems arising from this model: inference on the values of the θ_i and ξ_i, and testing hypotheses about r, the number of segments present in the data.

Most papers on these problems have assumed that the error terms e_t are normal with mean 0 and variance σ^2 (this will be abbreviated $e_t \sim N(0, \sigma^2)$). This model will also be assumed in this paper, and attention confined to literature on this model.

Chernoff and Zacks [2] study the problem of estimating the most recent mean ξ_r. The model is a Bayesian one in which a normal prior distribution is assigned to ξ_r, and it is shown that the Bayes estimator is, in general, nonlinear. A tentative estimator for θ_1 in the case $r = 2$ is also given. This Bayesian approach is taken further by Gardner [7].

A classical approach to the estimation of θ_1 for the case $r = 2$ is given in Hinkley [8]. The distribution of the maximum likelihood estimator of θ_1 is derived and used for setting up confidence intervals. This paper is closely related to Hinkley's, in that we shall be concerned with the testing problem which mirrors his estimation problem.

A test for whether a change has occurred is given in Page [10]. The test statistic is a cusum of $(X_i - \xi)$ where ξ is the supposed mean, and its is assumed that σ is known. This test is not dependent on the actual number of segments being specified, but is not efficient if it is known that the number of segments has one of two prespecified values. Chernoff and Zacks [2], under the Bayesian model mentioned earlier, show that an approximate test for the presence of two segments is based on $\sum_i (i-1)(X_i - \xi_1)$ if ξ_1 is known, and $\sum_i (i-1) \cdot (X_i - \bar{X})$ if ξ_1 is unknown, $\bar{X}$ being the mean of all n observations. Their assumption that the sign of $\xi_2 - \xi_1$ is known is relaxed by Gardner [7], who obtains a quadratic test statistic.

This paper is concerned with the classical test of the null hypothesis that a single segment is present against the alternative hypothesis that two segments are present. The likelihood ratio test will be given and its distribution under the null hypothesis derived.

Bayesian models will not be considered further in this paper, except to note that Sen and Srivastava [13, 14], in comparing the relative powers of the Bayesian procedure and the classical likelihood ratio test, showed the latter to have superior power when the two segments are of disparate lengths.

2. THE LIKELIHOOD RATIO TEST STATISTICS

Let

$$\bar{X}_k = \sum_{j=1}^{k} X_j/k \ , \quad \bar{X}_k{}' = \sum_{j=k+1}^{n} X_j/(n-k) \ , \quad 1 \le k \le n \ ,$$

$$S_k = \sum_{j=1}^{k} (X_j - \bar{X}_k)^2 + \sum_{j=k+1}^{n} (X_j - \bar{X}_k{}')^2 \ ,$$

and write $\bar{X}$ for $\bar{X}_n$, S for S_n.

We consider the hypotheses

$$\begin{aligned} H_0&: X_i \sim N(\xi_1, \sigma^2) \ , \quad i = 1, \ldots, n \ , \\ H_1&: X_i \sim N(\xi_1, \sigma^2) \ , \quad i = 1, \ldots, \theta_1 \ , \\ &\quad\ \sim N(\xi_2, \sigma^2) \ , \quad i = \theta_1 + 1, \ldots, n \ , \end{aligned}$$

with θ_1, ξ_1, and ξ_2 unknown ($\xi_1 \ne \xi_2$).

* Douglas M. Hawkins is Professor of Statistics, University of the Witwatersrand, Johannesburg, 2001 South Africa. The author is grateful to the editors and referee for suggesting a number of improvements.

March 1977, Volume 72, Number 357
Theory and Methods Section

Two cases may be distinguished: σ known and σ unknown. If σ is known, then the only unknown parameter under H_0 is ξ_1, and its maximum likelihood estimator (MLE) is $\bar{X}$. Under H_1 there are three unknown parameters. If θ_1 is fixed at k, then the MLE's are

$$\hat{\xi}_1 = \bar{X}_k \ , \qquad \hat{\xi}_2 = \bar{X}_k' \ .$$

The likelihood function is monotonically decreasing in S_k, and so the MLE of θ_1 is that k^* such that S_{k^*} is a minimum, over $k = 1, \ldots, n-1$.

Analysis of variance shows that $S = S_k + E_k$ where

$$E_k = k(\bar{X}_k - \bar{X})^2 + (n-k)(\bar{X}_k' - \bar{X})^2 \ , \qquad k = 1, \ldots, n-1 \ ,$$

so the criterion of minimum S_k is equivalent to maximum E_k.

It is then easy to show that

$$-2\sigma^2 \log \text{(likelihood ratio)} = S - S_{k^*} \ , \quad = E_{k^*} = U^2 \ , \quad \text{say} \ , \tag{2.1}$$

so that E_{k^*} defines the likelihood ratio test statistic.

If σ is unknown, then simple algebra shows that the likelihood ratio test is based on

$$S/S_{k^*} = 1 + E_{k^*}/S_{k^*} = 1 + W^2 \ , \quad \text{say} \ ,$$

and so is equivalent to

$$W^2 = E_{k^*}/S_{k^*} = E_{k^*}/(S - E_{k^*}) \ . \tag{2.2}$$

Under H_0, for arbitrary k, E_k/σ^2 follows a χ^2 distribution with one degree of freedom. The distribution of E_{k^*} will obviously be stochastically larger than $\sigma^2\chi_1^2$ because of the maximization over k.

3. THE DISTRIBUTION OF U

The first problem we will consider is the distribution of U. Since σ is assumed to be known we can set it equal to 1 without loss of generality.

An alternative expression for E_k is

$$E_k = n\{\sum_{i=1}^{k} (X_i - \bar{X})\}^2/\{k(n-k)\} \ .$$

Now k^* yields the maximum of E_k, or equivalently of $\sqrt{E_k}$. Thus

$$U = \sqrt{E_{k^*}} = \max_{1 \le k \le n-1} |T_k| \tag{3.1}$$

where $T_k = \{n/[k(n-k)]\}^{\frac{1}{2}} \sum_{i=1}^{k} (X_i - \bar{X})$. A simple calculation shows that for $k = 1, \ldots, n-1$, and $m < k$, the correlation between T_m and T_k is

$$\rho_{mk} = n/\sqrt{\{k(n-k)m(n-m)\}} \cdot E[\sum_{i=1}^{m} (X_i - \bar{X}) \sum_{j=1}^{k} (X_j - \bar{X})] \tag{3.2}$$

which simplifies [8] to $[m(n-k)/\{k(n-m)\}]^{\frac{1}{2}}$.

The distribution of U is that of the maximum absolute value attained by a Gaussian process in discrete time having zero mean, unit variance, and autocorrelation ρ_{mk}.

Lemma 1: The process $\{T_1, T_2, \ldots, T_{n-1}\}$ is Markovian.

Proof: It suffices to show that if $j < m < k$, then the partial covariance $\sigma_{jk.m}$ is zero. But this covariance is simply

$$\left\{\frac{j(n-k)}{k(n-j)}\right\}^{\frac{1}{2}} - \left\{\frac{j(n-m)m(n-k)}{m(n-j)k(n-m)}\right\}^{\frac{1}{2}} = 0 \ .$$

This property of the T_k process simplifies the distributional problems connected with U considerably, as is indicated in Theorem 1. Define

$$\phi(x, a, b) = 1/\{b(2\pi)\}^{\frac{1}{2}} \exp - \tfrac{1}{2}(x-a)^2/b^2 \ ,$$
$$g_1(x, s) = 1 \ , \quad x, s \ge 0 \ ,$$
$$g_k(x, s) = P[|T_i| < s, i = 1, \ldots, k-1 | T_k = x] \ , \quad x, s \ge 0 \ .$$

Theorem 1: The pdf of U is

$$f_U(x) = 2\phi(x, 0, 1)H(x) \ , \tag{3.3}$$

where $H(x) = \sum_{k=1}^{n-1} g_k(x, x) g_{n-k}(x, x)$.

Proof: Since $U = \max_k |T_k|$, it follows that

$$f_U(x)dx = \sum_{k=1}^{n-1} P[|T_k| \in (x, x+dx) \text{ and } |T_j| < |T_k| \text{ for all } j \ne k] \ .$$

To simplify the expressions that follow, we introduce the following notation for various events:

$$A = |T_k| \in (x, x+dx) \ ,$$
$$B = |T_j| < |T_k| \ , \quad j = 1, \ldots, k-1 \ ,$$
$$C = |T_j| < |T_k| \ , \quad j = k+1, \ldots, n \ .$$

Then the summand is $P[A \cap B \cap C]$. Now

$$P[A] = 2\phi(x, 0, 1)dx + O(dx)$$

since $T_k \sim N(0, 1)$, and

$$P[B|A] = P[|T_j| < x, \ j = 1, \ldots, k-1 \big| |T_k| = x] + o(dx)$$

since the events $|T_k| < |T_k|$ and $|T_j| < x$ are equivalent to order $O(dx)$ given A. Thus

$$P[B|A] = g_k(x, x) + O(dx) \ .$$

Next, the series T_k is Markovian, which implies that given T_k the sets $\{T_1, T_2, \ldots, T_{k-1}\}$ and $\{T_{k+1}, \ldots, T_n\}$ are independent.

$$P[C|A \cap B] = P[C|A] \ .$$

Finally, we note that the series obtained by taking the X_i in reverse order is a probabilistic replica of the given series $\{T_1, \ldots, T_{n-1}\}$. This implies that $P[C|A] = g_{n-k}(x, x) + O(dx)$. Combining these results yields the statement of the theorem.

Before Theorem 1 can be used, an expression for the $g_k(x, x)$ is required. This is given in the following theorem.

Theorem 2: The functions $g_k(x, s)$ satisfy the recursion

$$g_k(x, s) = \int_{y=0}^{s} g_{k-1}(y, s)[\phi(y, \rho x, \tau) + \phi(y, -\rho x, \tau)]dy \quad (3.4)$$

where $\rho = \rho_{k-1,k}$ and $\tau = \sqrt{(1 - \rho^2)}$.

Proof: By definition

$$g_k(x, s) = P[|T_j| < s, j = 1, \ldots, k-1 \,\big|\, |T_k| = x]$$
$$= \int_{y=-s}^{s} P[|T_j| < s, j = 1, \ldots, k-2 | (T_{k-1} = y) \cap (|T_k| = x)] \cdot dP[T_{k-1} < y \,\big|\, |T_k| = x] \; .$$

The Markovian nature of the T_k process implies that the first of these probabilities is $P[|T_j| < s,\; j = 1, \ldots, k-2 | T_{k-1} = y]$. Then since the distribution of each T_j is symmetric about zero, this simplifies to $P[|T_j| < s, j = 1, \ldots, k-2 \,\big|\, |T_{k-1}| = |y|]$ which is just $g_{k-1}(|y|, s)$.

Next we note that by this same symmetry it may be assumed that $T_k = x$, and hence that the conditional density of T_{k-1} is $N(\rho x, \tau)$. Thus,

$$P[T_{k-1} \in (y, y + dy) \,\big|\, |T_k| = x] = \phi(y, \rho x, \tau)dy + o(dy) \; .$$

Finally, gathering these results yields

$$g_k(x, s) = \int_{y=-s}^{s} g_{k-1}(|y|, s)\phi(y, \rho x, \tau)dy$$
$$= \int_{y=0}^{s} g_{k-1}(y, s)[\phi(y, \rho x, \tau) + \phi(y, -\rho x, \tau)]dy \; .$$

Tables of $g_k(x, s)$, $H(x)$, and the cumulative distribution function of T_{k^*} have been computed using Theorems 1 and 2 and a number of representative fractiles are listed in part a of Table 1.

1. Fractiles and Their Bonferroni Approximates[a]

n	α = .10	.05	.01
	a. Fractiles of U		
4	2.06 (2.13)	2.35 (2.39)	2.91 (2.94)
5	2.15 (2.24)	2.43 (2.50)	2.99 (3.02)
10	2.38 (2.54)	2.65 (2.77)	3.19 (3.26)
15	2.49 (2.69)	2.75 (2.91)	3.29 (3.38)
20	2.55 (2.79)	2.82 (3.01)	3.35 (3.47)
30	2.64 (2.92)	2.90 (3.13)	3.44 (3.58)
50	2.73 (3.09)	2.98 (3.28)	3.50 (3.71)
	b. Fractiles of W		
4	4.03 (5.34)	5.62 (7.65)	10.16 (17.28)
5	3.44 (4.18)	4.49 (5.39)	7.68 (9.56)
10	2.88 (3.28)	3.38 (3.76)	4.58 (4.96)
15	2.82 (3.19)	3.22 (3.55)	4.12 (4.40)
20	2.81 (3.20)	3.17 (3.49)	3.95 (4.21)
30	2.82 (3.20)	3.14 (3.47)	3.82 (4.07)
50	2.85 (3.26)	3.14 (3.50)	3.75 (4.02)

[a] The Bonferroni approximates are in parentheses.

Conservative tests of H_0 may also be made using Bonferroni's inequalities since

$$P[\max_{1 \le k \le n-1} |T_k| > c] \le (n-1)P[|T_1| > c] = 2(n-1)\Phi(-c) \; ,$$

where Φ is the standard normal distribution function. Thus a conservative level α test of H_0 may be based on the upper $\alpha/(2n-2)$ fractile of the standard normal distribution. These approximations are listed in Table 1 next to the exact values. As one would expect, the approximation is best when n and α are small, but even in the worst case the error is moderate.

The form of the extreme fractiles may be deduced from Theorems 1 and 2 by noting that given $T_k = x$, $T_m \sim N(\rho_{mk}x, 1 - \rho_{mk}^2)$ for $m < k$. Thus

$$P[T_m < x | T_k = x] = \Phi[x(1 - \rho_{mk})/\sqrt{(1 - \rho_{mk}^2)}]$$
$$= \Phi[x\{(1 - \rho_{mk})/(1 + \rho_{mk})\}^{\frac{1}{2}}]$$

and so $P[|T_m| < x | T_k = x] \to 1$ as $x \to \infty$. From this fact, it is a short step to show that as $x \to \infty$, $g_k(x, x) \to 1$, and so $H(x) \to (n-1)$.

This means that as $x \to \infty$,

$$P[U > x] \sim 2(n-1)\Phi(-x)$$

which is the bound given by the first Bonferroni inequality, and suggests that the significance of large U values may be assessed accurately from the Bonferroni inequality.

The distribution of k^* can also be obtained from Theorem 1.

Theorem 3: For $k = 1, \ldots, n-1$,

$$P[k^* = k] = \int_0^{\infty} g_k(x, x)g_{n-k}(x, x)\phi(x, 0, 1)dx \; . \quad (3.5)$$

Proof: From Theorem 1,

$$P[k^* = k] = \int_0^{\infty} dP[\{|T_k| < x\} \cap \{|T_j| < |T_k| j \ne k\}]$$
$$= \int_0^{\infty} g_k(x, x)g_{n-k}(x, x)\phi(x, 0, 1)dx \; .$$

The distribution of k^* for $n = 50$ is listed in Table 2, which also lists for comparison the distribution of the location of the maximum of a random walk with independently and identically distributed increments [6].

2. Distribution of k^* for $n = 50$

k^*	Probability of k^*: Equation 3.5	Random walk
1– 5	.213	.203
6–10	.100	.093
11–15	.073	.076
16–20	.062	.069
21–25	.058	.066

The latter random variable tends to an arc sine distribution, and it is of interest to note that k^* has a more extreme concentration near the endpoints of its range.

4. THE DISTRIBUTION OF W

Let us now turn to the situation in which σ is unknown. The likelihood ratio test statistic is E_{k^*}/S_{k^*}. Let us write

$$Z_k = T_k\{(n-2)/(S-T_k^2)\}^{\frac{1}{2}} \qquad (4.1)$$

which may be inverted to give

$$T_k = Z_k\{S/(n-2+Z_k^2)\}^{\frac{1}{2}} .$$

Now

$$E_{k^*} = (\max_k |T_k|)^2 , \quad \text{so} \quad E_{k^*}/S_{k^*} = (n-2)Z_{k^*}^2 .$$

Since Z_k is monotonic in T_k, the likelihood ratio statistic is equivalent to $W = \max |Z_k|$, whose distributions we now find.

Theorem 4: The pdf of W is

$$f_W(t) = 2f_\nu(t)\int_{s=0}^{\infty} h_\nu(s)H\{st/(\nu+t^2)^{\frac{1}{2}}\}ds , \qquad (4.2)$$

where $\nu = n-2$, $f_\nu(t)$ is a Student's-t density with ν degrees of freedom, $h_\nu(s)$ is the density of the square root of a χ^2 variate with $\nu+1$ degrees of freedom, and H is defined by (3.3).

Proof: Under H_0 the statistics $\bar{X}$ and S constitute complete sufficient statistics for ξ and σ^2. Now Z_k is a function of the studentized residuals $(X_i - \bar{X})/\sqrt{S}$ and so is distributed independently of $\bar{X}$ and S.

Suppose initially that S is fixed at s^2. Then

$$\begin{aligned} P[W \in (t, t+dt)|s] &= \sum_{k=1}^{n-1} P[|Z_k| \in (t, t+dt)|s]P[|Z_j| < |Z_k| , \\ &\qquad j \neq k|s, Z_k = t] + O(dt) . \end{aligned}$$

Considering the second probability we have given s and $Z_k = t$ that $T_k = ts/\sqrt{(n-2+t^2)}$. But by the monotonicity between Z and T, $|Z_j| < |Z_k|$ if and only if $|T_j| < |T_k|$ so that, say,

$$\begin{aligned} P[|Z_j| &< |Z_k| j \neq k|s, Z_k = t] \\ &= P[|T_j| < |T_k| j \neq k|s, T_k = ts/\sqrt{(n-2+t^2)}] \\ &= g_k\{ts(\nu+t^2)^{-\frac{1}{2}}, ts(\nu+t^2)^{-\frac{1}{2}}\} \\ &\qquad \cdot g_{n-k}\{ts(\nu+t^2)^{-\frac{1}{2}}, ts(\nu+t^2)^{-\frac{1}{2}}\} \\ &= G_k(s, t) . \end{aligned} \qquad (4.3)$$

Next, we note that Z_k is independent of S and follows a t distribution with ν degrees of freedom. Thus

$$P[|Z_k| \in (t, t+dt)|s] = 2f_\nu(t)dt + o(dt)$$

and so

$$P[W \in (t, t+dt)|s] = 2f_\nu(t)\sum_{k=1}^{n-1} G_k(s, t)dt .$$

Now mixing this distribution over that of S and using the definition of $H(x)$ leads to

$$\begin{aligned} P[W &\in (t, t+dt)] \\ &= 2f_\nu(t)\left[\int_{s=0}^{\infty} h_\nu(s)H\{st/(\nu+t^2)^{\frac{1}{2}}\}ds\right]dt + o(dt) . \end{aligned} \qquad (4.4)$$

A number of fractiles of this distribution have been computed, and are listed in part b of Table 1. As in the case when σ is known, a conservative test of H_0 may be found via Bonferroni's inequality applied to the t distribution with $n-2$ degrees of freedom. These approximates are also listed in Table 1b for comparison.

Table 1b shows that the Bonferroni approximation is adequate for small n and α, although a study of the Type 1 error probabilities shows that the approximation is considerably worse than it is for σ known. It seems that the fractiles for large values of n are approaching those of Table 1a.

When H_0 is rejected in favor of H_1, the problem of estimating θ_1, ξ_1, ξ_2, and σ^2 arises. The estimation of these parameters is discussed by Hinkley [8] who shows, *inter alia*, that $k^* = \theta_1 + O(1)$, and that the estimators $\bar{X}_{k^*}$ and $\bar{X}_{k^*}'$ have a bias of order n^{-1} as $n \to \infty$.

Analysis of variance considerations suggest the estimator $\hat{\sigma}^{2\prime} = S_{k^*}/(n-2)$ for σ^2. Since $S_{\theta_1}/(n-2)$ is an unbiased estimator, and $S_{k^*} \leq S_{\theta_1}$, this estimator is clearly biased downwards.

From the analysis of variance decompositions

$$S_{\theta_1} + E_{\theta_1} = S = S_{k^*} + E_{k^*} ,$$

we see that the bias in $\hat{\sigma}^2$ is $-\mathcal{E}[(E_{k^*} - E_{\theta_1})/(n-2)]$, where $\mathcal{E}$ denotes expectation. In this noncentral case, $\mathcal{E}(E_{k^*} - E_{\theta_1})$ cannot exceed the corresponding expectation for the central case, since for any k_1,

$$\mathcal{E}(E_k - E_{\theta_1}|H_0) = \mathcal{E}(E_k - E_{\theta_1}|H_1) ,$$

and by the consistency of k^*/n the range of k^* values is narrower in the noncentral case than the central one. Thus $\mathcal{E}(E_{k^*} - E_{\theta_1}|H_0)/(n-2)$ provides an upper bound for the bias in $\hat{\sigma}^2$. This expectation could be computed directly from the pdf of U. However, it is indicated next that U is plausibly $(2\log\log n)^{\frac{1}{2}} + O(1)$, and hence the bias in $\hat{\sigma}^2$ is asymptotically negligible.

5. ASYMPTOTIC RESULTS

In this section, the behavior of the distribution of U as $n \to \infty$ is discussed.

Since U^2 is defined as $-2\log$ (likelihood ratio) one might anticipate that asymptotically, U^2 would approach a χ_2^2 distribution. The standard asymptotic theory, however, assumes that the likelihood function is twice continuously differentiable in its parameters [9] while in this case, it is a discontinuous function of θ_1, and so the standard theory cannot be invoked. Quandt [12], studying a rather similar problem in piecewise regression, concluded from a simultation study that the explained variation, analogous to U^2, was stochastically larger than its imputed χ^2 distribution.

In an attempt to understand the asymptotic behavior of U, we consider a continuous process whose distributional and autocorrelation structure match those of the T_k process. The appropriate process[1] is a Gaussian process $x(t)$, $0 < t < 1$, having mean 0, variance 1, and autocorrelation $\rho_{st} = \sqrt{[s(1-t)/\{t(1-s)\}]}$ for $s < t$.

Following the general procedure of Doob [5], which has been validated by Donsker [4], we define a Brownian motion $\zeta(t) = t^{\frac{1}{2}}x\{t/(1+t)\}$, $0 < t < \infty$, and note that the event $|x(t)| < \lambda$ for all $t \in (0, 1)$ is equivalent to $|\zeta(t)| < \lambda\sqrt{t}$ for all $t \in (0, \infty)$.

Now Khintchine's law of the iterated logarithm (e.g., [15, p. 622]) states that with probability 1 for any $\epsilon < 0$ a Brownian motion exceeds $(1+\epsilon)\{2t \log|\log t|\}^{\frac{1}{2}}$ at least once in every neighborhood of 0 and ∞. Now $\log|\log t|$ is unbounded at 0 and ∞, and so with probability 1, for any λ, however large, $|x(t)|$ will exceed λ at least once in every neighborhood of 0 and 1.

Thus as $n \rightarrow \infty$, $U \rightarrow \infty$ with probability 1. In this sense, U does not have an asymptotic distribution, though it may have an extreme value distribution which depends on n.

The discussion relating $x(t)$ to a Brownian motion suggests (and this will be verified) that as $n \rightarrow \infty$, the location of the maximum k^*/n tends to either 0 or 1. This leads to the following approach for studying the dependence of the distributions of U and k^* on n.

Let the stochastic process $x(t)$ be observed at $M-1$ instants $t_i = i/M$ $(i = 1, 2, \ldots, M-1)$, where it is assumed that M is large. Augment this partial realization by $Y_{11}, Y_{12}, \ldots, Y_{1M}$ which are values of $x(t)$ at spacings halfway between the original t_i. Thus Y_{1j} corresponds to a time $(2j-1)/(2M)$. Then halve this mesh again, and let the $x(t)$ values at the new mesh points be $Y_{21}, Y_{22}, \ldots, Y_{2,2M}$. Repeat this halving process in total L times, finally observing $Y_{L1}, Y_{L2}, \ldots, Y_{LJ}$ where $J = 2^{L-1}M$.

From the construction of the Y_{ij} it can be seen that Y_{ij} is observed at time $t = 2^{-i}(2j-1)/M$ and so the aggregate of the $M - 1x(t)$ values and the $M(2^L - 1)Y_{ij}$ makes up a realization of $M2^L - 1$ points sampled from the $x(t)$ process at a spacing of $(2^LM)^{-1}$. In view of this equal spacing, this aggregate may be regarded as the set of T_k resulting from a series of length $n = 2^LM$. Thus the statistic U is distributed like the maximum of the $M - 1|x(t_i)|$ and $n - M|Y_{ij}|$, and k^*/n is distributed like the location t at which this maximum occurs.

Since Y_{ij} corresponds to a time $t = 2^{-i}(2j-1)/M$, it follows that Y_{ij} and Y_{kl} are $N(0, 1)$ with a correlation equal to the smaller of

$$\left[\frac{2^{-i}(2j-1)}{2^{-k}(2l-1)}, \frac{M - 2^{-k}(2l-1)}{M - 2^{-i}(2j-1)}\right]^{\frac{1}{2}}$$

and its inverse. This shows that pairs of diagonal subsequences which are located symmetrically about $t = \frac{1}{2}$, such as $Y_{11}, Y_{21}, \ldots, Y_{L1}$ and $Y_{1,M}, Y_{2,2M}, \ldots, Y_{L,J}$, apart from having the same joint distribution of their elements, are approximately independent provided M is large. This identity of distribution and approximate independence will be utilized later.

Define $Y_j = \max(|Y_{1j}|, |Y_{2j}|, \ldots, |Y_{Lj}|)$. Now the sequence over which this maximum is taken is of $N(0, 1)$ variates with the correlation between Y_{ij} and $Y_{i+N,j}$ being given by

$$\rho_{i,i+N}{}^2 = 2^{-N}\{[M - 2^{-i}(2j-1)]/[M - 2^{-i-N}(2j-1)]\} \; .$$

If M is large relative to j, this means that the sequence is approximately stationary with autocorrelation ρ_N at lag N given by $2^{-\frac{1}{2}N}$. Since $\rho_N \log N \rightarrow 0$ as $N \rightarrow \infty$, it follows from Berman [1] that Y_j is asymptotically distributed like the largest of L independent normal variables. That is, if $a_L = \sqrt{\{2 \log_e L\}}$, then

$$P[a_L{}^{-1}(Y_j - a_L) < y] \rightarrow \exp(-e^{-y}) = \Lambda(y) \; , \qquad (5.1)$$

a Type III extreme value distribution.

In particular, since U is greater than the larger of Y_1 and the maximum of its mirror image sequence, it follows that a stochastic lower bound for the asymptotic distribution of U is that of the largest of $2L$ independent $N(0, 1)$ variables. The centering constant for this distribution is

$$\begin{aligned} a_{2L} &= \sqrt{\{2 \log_e 2L\}} \\ &= \sqrt{[2 \log_e \{2(\log_2 n - \log_2 M)\}]} \; , \qquad (5.2) \end{aligned}$$

since $n = 2^LM$. Since $a_{2L} \rightarrow \infty$ as n increases, this lower bound implies that asymptotically the $M-1$ terms $x(t_i)$ do not contribute to the maximum, and so may be ignored.

Considering now the Y_{kl} with $l \neq 1$ which correspond to $t < \frac{1}{2}$, we see that all lie either near Y_{ij} for some i or near $x(t_j)$ for some j. In the former case, the correlation with Y_{i1} is at least $2^{-\frac{1}{4}} = .84$, and in the latter, at least .75. If the correlation between Y_{kl} and Y_{i1} is ρ, then it is easily shown that

$$\begin{aligned} &P[Y_{kl} \neq Y_{i1} + \delta \,|\, Y_{i1} = y] \\ &\quad = \Phi[-y\{1 - \rho/(1+\rho)\}^{\frac{1}{2}} - \delta/\sqrt{(1-\rho^2)}] \; . \end{aligned}$$

Since in the area of interest y is at least of order $\sqrt{\{2 \log_e (2L)\}}$, and so is large, it follows that this probability is small, and hence that the contribution of these Y_{kl} to the distribution of T_{k^*} is small, and possibly negligible.

As regards the distribution of k^*/n, since the Y_{i1} sequence is approximately stationary, all its terms have the same probability of yielding the maximum Y_1. Thus the median location of the maximum is at the $L/2$ term corresponding to $t = 2^{-\frac{1}{2}L}/M = (nM)^{-\frac{1}{2}}$. Taking into account the symmetry about $t = \frac{1}{2}$, we then conclude that the lower fractiles of k^*/n will vary as $n^{-\frac{1}{2}}$.

In an attempt to assess the adequacy of this model, a simulation was carried out. Sequences of independent $N(0, 1)$ variates of various lengths n were generated, and

[1] The referee has pointed out that this statement, while surely correct, does not seem to be proved in the literature. It should thus be regarded as an heuristic. The major result derived from it, that U tends to infinity with probability one, is verified directly in the following paragraphs.

for each sequence U and k^*/n were computed. The mean, standard deviation, skewness $\sqrt{b_1}$, and kurtosis b_2 of U are shown in Table 3 together with the lower quartile of k^*/n.

3. Statistics of U and k/n Obtained from Simulation*

Series length	Number simulated	Mean	Standard deviation	Skewness	Kurtosis	Quartile of k*/n
50	5000	1.989	.550	.677	3.62	.16
100	5000	2.094	.536	.677	3.69	.11
500	2000	2.274	.529	.778	4.20	.07
1000	3000	2.344	.521	.640	3.62	.05
5000	600	2.473	.547	.523	3.14	.03

The table shows that $\Lambda(\cdot)$, the Type III extreme value distribution, which has skewness 1.3 and kurtosis 5.4 [3] is a bad fit to the data for all the n values studied. This does not preclude $\Lambda(\cdot)$ as a limiting distribution of U. It is known that the approach of normal distribution extreme values to the limiting distribution is very slow. This fact was illustrated by a simulation of the largest of 320 independent $N(0, 1)$ variates which yielded a skewness of .69 and kurtosis 3.91.

Since $\Lambda(\cdot)$ is inadequate as a model for n as high as 5000, questions of its asymptotic goodness-of-fit are clearly academic, and the problem remains of adequate approximations to its distribution, and especially its fractiles, for n values in excess of 50. The moments of U shown in Table 3, and an inspection of stem and leaf plots of the actual values obtained in the simulations, suggest that to a good degree of approximation, the distribution may be taken as depending on n only through a centering constant.

In view of the observation that asymptotically U is centered at $\sqrt{[2 \log_e (2L)]}$ with $L = \log_2 n - \log_2 M$, a curve of the form $\sqrt{[2 \log_e \{a \log_2 n + b\}]}$ was fitted to the mean of U by least squares. This yielded $a = 1.966$, $b = -4.015$, and a correlation of .997. The value 2 for a suggested by the theory lay well within one standard error, as did the simpler value -4 for b. These values yield a centering constant of $\sqrt{[2 \log_e \{2 \log_2 (n/4)\}]}$.

Applying this centering correction and using the 90 percent fractile of U for $n = 50$ yielded the approximate fractiles listed in Table 4. Also shown is the proportion of the simulated U exceeding these values. All are seen to be close to the nominal .1, which suggests that the model is of acceptable accuracy for practical application to series of length up to 5000.

4. Simulation Check on Approximate Fractiles of U

n	Approximate fractile	Proportion of exceedances
50	2.73	.100
100	2.85	.090
500	3.03	.086
1000	3.09	.086
5000	3.20	.093

The lower quartile of k^*/n listed in Table 3 agrees fairly well with the model that it vary as $n^{-\frac{1}{2}}$.

A final question relates to the asymptotic behavior of W, the studentized equivalent of U. In this connection, note that S is the sum of $n - 1$ independently and identically distributed χ_1^2 variates. By the strong law of large numbers $(n - 1)^{-1}S \to 1$ with probability 1. On the other hand, due to the positive intercorrelations between the T_k, U is stochastically smaller than the largest of n independent normal variates, and this latter quantity, V_n say, satisfies

$$P[\limsup_{n\to\infty} a_n(V_n - a_n)/\log \log n = \tfrac{1}{2}] = 1$$

where $a_n = \sqrt{\{2 \log_e n\}}$ [11]. So U is of order $O(\sqrt{\log_e n})$ with probability 1. Hence, in the formula

$$W = U\{(n - 2)/(S - u^2)\}^{\frac{1}{2}} ,$$

as $n \to \infty$, the term in braces tends to 1 with probability 1. Thus the asymptotic distribution of W is the same as that of U.

6. EXAMPLE

A model of the Stock Exchange which is widely believed (and whose merits we will not argue here) is that a share price P_t at time t is

$$P_t = P_{t-1} + Y_t ,$$

where $Y_t \sim N(\xi_t, \sigma^2)$ and $Y_1, Y_2, \ldots$ are independent. A bull market is one in which $\xi_t > 0$, and a bear market one in which $\xi_t < 0$. We are concerned to know whether the market has changed posture during an interval, and set up the hypotheses

$$\begin{aligned} H_0&: \xi_t = \xi , \quad \text{for all } t , \\ H_1&: \xi = \xi_1 , \quad t \le \theta , \\ &\quad\;\; = \xi_2 , \quad t > \theta . \end{aligned}$$

A series of 50 week to week differences in the Eurosyndicat index (an index of the major stock markets on the European Continent) was computed for the weeks following August 3, 1971. This series yielded a Z_{k^*} of 4.88 which is highly significant. The basic statistics of the two segments are $\bar{X}_{k^*} = -1.77$, $\bar{X}_{k^*}' = 1.01$, $S_{k^*} = 162.59$, and $k^* = 15$.

The boundary between the segments occurs at November 9, 1971. This point marked the end of a long period of lack of confidence on the European stock markets, and was followed shortly by a devaluation of the dollar against the major European currencies.

The full series available started in January 1966 and contained 486 terms. An estimate of the variance was computed from the mean squared successive difference of the weekly changes and yielded the value 4.36. Under the model's assumptions of independent homoscedastic increments with possible occasional abrupt shifts in location, this estimate is preferable to the sample variance as it is hardly affected by a small number of changes in ξ.

[Received May 1974. Revised June 1976.]

REFERENCES

[1] Berman, Simeon M., "Limit Theorems for the Maximum Term in Stationary Sequences," *Annals of Mathematical Statistics*, 35 (1964), 502–16.

[2] Chernoff, H., and Zacks, S., "Estimating the Current Mean of a Normal Distribution Which Is Subjected to Changes in Time," *Annals of Mathematical Statistics*, 35 (1964), 999–1018.

[3] David, Herbert A., *Order Statistics*, New York: John Wiley & Sons, Inc., 1970.

[4] Donsker, Monroe D., "Justification and Extension of Doob's Heuristic Approach to the Kolmogorov-Smirnov Limit Theorems," *Annals of Mathematical Statistics*, 23 (1952), 277–81.

[5] Doob, J.L., "Heuristic Approach to the Kolmogorov-Smirnov Theorems," *Annals of Mathematical Statistics*, 20 (1949), 393–403.

[6] Feller, William, *An Introduction to Probability Theory and Its Applications, Vol. II*, New York: John Wiley & Sons, Inc., 1971.

[7] Gardner, L.A., Jr., "On Detecting Changes in the Mean of Normal Variates," *Annals of Mathematical Statistics*, 40 (1969), 116–26.

[8] Hinkley, David V., "Inference About the Change-Point in a Sequence of Random Variables," *Biometrika*, 57, 1 (1970), 1–17.

[9] Kendall, Maurice G., and Stuart, Alan, *The Advanced Theory of Statistics, Vol. II*, London: Charles W. Griffen & Co., Ltd., 1973.

[10] Page, E.S., "On Problems in Which a Change in Parameter Occurs at an Unknown Point," *Biometrika*, 44 (1957), 248–52.

[11] Pickands, James, III, "Sample Sequences of Maxima," *Annals of Mathematical Statistics*, 38 (1967), 1570–4.

[12] Quandt, Richard E., "Tests of the Hypothesis That a Linear Regression Line Obeys Two Separate Regimes," *Journal of the American Statistical Association*, 55 (June 1960), 324–30.

[13] Sen, Ashish, and Srivastava, Muni S., "On Tests for Detecting Change in Mean," *Annals of Statistics*, 3, 1 (1975), 98–108.

[14] ———, and Srivastava, Muni S., "Some One-Sided Tests for Change in Level," *Technometrics*, 17, 1 (1975), 61–4.

[15] Thomasian, Aram J., *The Structure of Probability Theory with Applications*, New York: McGraw-Hill Book Co., 1960.