South African Statist. J. (1989) **23** 1-21

INVITED REVIEW PAPER

SOME RECENT DEVELOPMENTS IN THE ANALYSIS OF CHANGEPOINT DATA

F. Lombard: Department of Statistics, Rand Afrikaans University, P.O. Box 524, Johannesburg, 2000.

Keywords: Estimation; review; testing.

Summary: The purpose of this paper is to review some of the more recent developments in the statistical analysis of changepoint data. The paper is also partly expository. The review focuses on the simpler changepoint models and on aspects that should make the paper accessible to statisticians who are not specialists in the field.

1. INTRODUCTION

The purpose of this paper is to review some of the more recent developments in the statistical analysis of changepoint data. The paper is also partly expository. In view of the broad scope of the subject, both in terms of the variety of possible models and the levels of technical skill required, this review will be limited to aspects that, it is hoped, will be easily understood by statisticians who are not specialists in the field.

STMA 1989 SUBJECT CLASSIFICATION: 03:070(05:000)

A description of possibly the simplest type of changepoint problem will now be given. Consider a time-ordered sequence of independent observations $x_1, \cdots, x_n$ with distribution functions $F_1, \cdots, F_n$. If $F_1 = \cdots = F_n$ the data are homogeneous, that is, $x_1, \cdots, x_n$ constitute a random sample from a fixed distribution. Suppose, more generally, that for some integer $\tau (1 \leq \tau < n)$ $x_1, \cdots x_\tau$ have a common c.d.f. F while $x_{\tau+1}, \cdots, x_n$ have c.d.f. G. Then a change of distribution has taken place and τ is called the changepoint. The inference problem involved here is to test the null hypothesis $H_0 : F = G$, i.e. $F_1 = \cdots F_n$, against the alternative $H_a : F \neq G$ and to estimate τ in the event the H_0 is rejected.

If τ is *known*, this is just a two sample problem for which many parametric and nonparametric tests exist. The distinguishing feature of a changepoint problem is that τ is not known beforehend. This observation also suggests a simple technique for constructing tests of H_0 in a changepoint model on the basis of well-known two-sample tests. Let Z_τ be an appropriate test statistic for the two-sample problem involving $x_1, \cdots, x_\tau$ and $x_{\tau+1}, \cdots, x_n$. Then

$$max_{1 \leq \tau < n} Z_\tau \quad \text{or} \quad n^{-1} \sum_{1 \leq \tau < n} Z_\tau$$

are obvious choices of statistics for testing H_0 against an alternative involving an unknown changepoint. The main focus in the statistical literature has been on the derivation of the exact or approximate distribution of test statistics such as these.

In practice, more complicated situations than the one described above can arise. In the first place, there are often grounds for suspecting that more than one changepoint may be present. Secondly, the x_i may not be independent but instead be generated by, for instance, an autoregressive scheme such as

$$x_i - \mu = \rho(x_{i-1} - \mu) + e_i,$$

the problem being to detect changes in the structural parameters μ and ρ.

In many, if not most, applications the data are time-ordered, that is, x_{i+1} is observed after x_i. Nevertheless, the full data set $x_1, \cdots, x_T$ is available for analysis and the situation differs from that encountered in sequential data analysis. In the latter case the data are analyzed with the advent of every new observation, the objective being to detect a change as soon as possible after its onset, and then to terminate sampling. In the typical changepoint situation to be described in this paper the sample size is not determined by a data-dependent stopping rule and the analysis is of a retrospective nature.

The rest of the paper is structured as follows: In section 2 a number of examples are discussed briefly. Section 3 outlines some parametric test procedures, that is, where the type of underlying distribution (normal, exponential, etc.) is assumed known. In contrast, section 4 considers nonparametric tests which, because of the relatively mild distributional assumptions required, seem particularly well suited for use in practical applications. Section 5 is devoted to a discussion of the problems surrounding point and interval estimation of the changepoint. In section 6 some references are given to more difficult problems. Absent from our discussion is any reference to the Bayesian way of solving changepoint problems. This is not meant to suggest that the Bayesian view has little to offer this problem. On the contrary, such a substantial literature has developed over the years that a review of it is better left to someone well versed in the Bayesian philosophy.

2. SOME EXAMPLES

2.1 Flow volume of the Nile river

Cobb (1978) has considered the annual flow volume of the Nile at Aswan for the years 1871 through 1970. A time series plot of the data, presented in figure 1 below, suggests a possible decrease in flow volume around the turn of the century. A simple model for these data assumes that the observed flow volumes represent independent random variables $x_1, \cdots, x_{100}$ with constant variance and means

$$E(x_i) = \mu \text{ for } i \leq \tau; \; = \; \mu_* \text{ for } i > \tau,$$

with τ unknown ($1 \leq \tau < 100$). Under this model, the impression gained from figure 1 can be reinforced considerably by plotting the (retrospective) cusums

$$C_k = \sum_{i=1}^{k} (x_i - \bar{x}_n). \tag{2.1}$$

Observe that

$$E(x_i - \bar{x}_n) = \begin{cases} (\mu - \mu_*)(1 - \tau/n) & \text{for } i \leq \tau \\ -(\mu - \mu_*)\tau/n & \text{for } i > \tau. \end{cases}$$

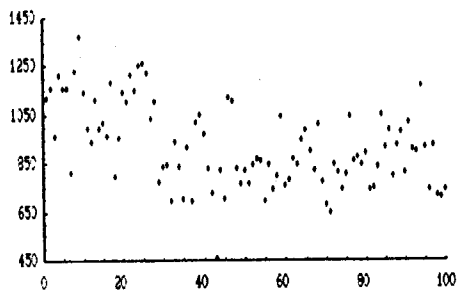

Figure 1. Annual flow volume (10^8 m^3) of the Nile at Aswan from 1871 to 1970

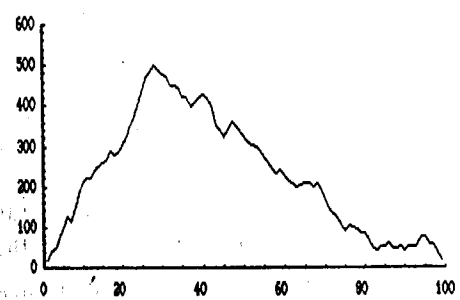

Figure 2. Cusum plot ($C_k \times 10^{-1}$) of Nile data

If the mean flow volume has not changed ($\mu = \mu_*$), then C_k can be expected to fluctuate around zero. In the contrary case, say $\mu > \mu_*$, an increasing trend, followed by a decrease after $k = \tau$ can be expected. The cusum plot, shown in figure 2, suggests a decrease in flow volume after 1898. Cusums such as C_k play an important role in many changepoint analyses. See Cobb (1978, section 4) for a fully parametric analysis of these data and Hinkley & Schechtman (1987) for a non-parametric approach.

2.2 Incidence of Talipes

Worsley (1983b, section 5) has considered the incidence of the birth defect Talipes (clubfoot) in a region of northern New Zealand for the years 1960 through 1976. His model for the data assumes that the number of Talipes births in year i is binomially distributed with parameters n_i (the known total births) and p_i (the unknown probability of a Talipes birth). In 1965 the herbicide 245-T was first used in the region, and it is of interest to see whether this coincides with an apparent increase in the incidence of Talipes births. Clearly, in view of the possible controversy surrounding such an issue, a rather careful analysis is required. As part of such an analysis, Worsley derives an exact test of

$$H_0 : p_1 = \cdots p_{17}$$

against an alternative involving a single change in incidence at some point in time.

2.3 Relapse rate in the treatment of leukemia

Matthews & Farewell (1982) have considered data on the rate of relapse after onset of remission of patients treated for leukemia. The usual experience is that the relapse rate of such patients is constant over time. If t denotes the time to relapse, they consider a model for the distribution of t specified by the hazard function

$$\lambda(t) = \begin{cases} \lambda & \text{if } t < \tau \\ \rho\lambda & \text{if } t \geq \tau. \end{cases}$$

The parameters λ and ρ, as well as the changepoint, τ, are unknown. Note that τ is now measured on a continuous time scale and that censoring can occur. The usual experience corresponds to $\rho = 1$ (equivalently, $\tau = \infty$). The alternative, $\rho < 1$, corresponds to an abrupt decrease in residual relapse rate for patients surviving past time τ after remission.

There are a number of other, very interesting, data sets available in the literature. Perusal of some of the following references should give the reader an appreciation for the range of possible applications: Hsu (1979), Pettitt (1979), Ali & Giaccotto (1982), Levin & Kline (1985), Commenges & Seal (1985), Lombard (1986, 1987), Srivastava & Worsley (1986), Worsley (1986) and Birks (1987).

3. PARAMETRIC TESTS

3.1 Normally distributed data

Let $x_1, \cdots, x_n$ be independent and normally distributed with means $\mu_1, \cdots, \mu_n$ and common variance $\sigma^2 > 0$. The simplest model for a change in mean after the τ-th observation $(1 \le \tau < n)$ is

$$\mu_i = \mu \text{ for } 1 \le i \le \tau; = \mu^* \text{ for } \tau + 1 \le i \le n. \tag{3.1}$$

The hypothesis of no change is $H_0 : \mu = \mu^*$, while the general changepoint alternative says that (3.1) holds for some (unspecified) τ. The likelihood ratio test of H_0 against (3.1) is, of course, based on the statistic $|T_\tau|$, where

$$T_\tau = (\bar{x}_\tau - \bar{x}^*_\tau)/\{\tau^{-1} + (n-\tau)^{-1}\}^{\frac{1}{2}}\tilde{\sigma}_{\tau,n}$$

is the well-known Student t-statistic, where

$$\bar{x}_\tau = \tau^{-1}\sum_{i=1}^{\tau} x_i \text{ and } \bar{x}^*_\tau = (n-\tau)^{-1}\sum_{i=\tau+1}^{n} x_i$$

are the means of the samples $x_1, \cdots, x_\tau$ and $x_{\tau+1}, \cdots, x_n$, and where

$$\tilde{\sigma}^2_{\tau,n} = (n-2)^{-1}\left\{\sum_{i=1}^{\tau}(x_i - \bar{x}_\tau)^2 + \sum_{i=\tau+1}^{n}(x_i - \bar{x}^*_\tau)^2\right\} \tag{3.2}$$

is the usual pooled estimate of σ^2. It follows that the likelihood ratio statistic for testing H_0 against the changepoint alternative ((3.1) with τ unspecified) is

$$U_n = max\{|T_\tau|; 1 \le \tau < n\}. \tag{3.3}$$

The T_τ all have t-distributions with $n-2$ degrees of freedom but, clearly, these are not independent. Thus some difficulty in finding the exact or large sample distribution of (3.3) can be anticipated.

Before proceeding with this, it is worthwhile to discuss some technical points. Let

$$\hat{\sigma}^2_n = (n-1)^{-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 \tag{3.4}$$

be the usual (unbiased under H_0) estimate of σ^2, and set

$$S_\tau = (\bar{x}_\tau - \bar{x}^*_\tau)/\{\tau^{-1} + (n-\tau)^{-1}\}^{\frac{1}{2}}\hat{\sigma}_n \tag{3.5}$$

and

$$V_n = max\{|S_\tau|; 1 \le \tau < n\}. \tag{3.6}$$

Simple algebraic manipulation shows that

$$S^2_\tau = (n-1)T^2_\tau/(n-2+T^2_\tau),$$

so that U_n and V_n are equivalent statistics. Furthermore, since U_n involves all $n-1$ quantities $\tilde{\sigma}_{\tau,n}$, it is clear that V_n requires less computational effort, especially when n is large. Also, if C_k denotes the cusum defined in (2.1), then

$$\bar{x}_\tau - \bar{x}^*_\tau = nC_\tau/\tau(n-\tau)$$

so that V_n in (3.6) can be expressed in terms of cusums as

$$V_n = max\{|C_\tau|/\{\tau(n-\tau)/n\}^{\frac{1}{2}}\hat{\sigma}_n; 1 \le \tau < n\}. \tag{3.7}$$

Finally, observe that

$$var\{C_\tau|H_0\} = n^{-1}\tau(n-\tau)\sigma^2 \tag{3.8}$$

so that V_n in (3.7) is just the maximum of the absolute cusums after standardization by their respective estimated standard errors.

In practice, a change occurring in some specified segment of the data may be more important than one occurring elsewhere. For instance, a change occurring very early (τ close to 1) or very late (τ close to n) may be relatively inconsequential. This leads to an *ad hoc* generalization of V_n, namely

$$max\{w(\tau)|S_\tau|; 1 \le \tau < n\}, \tag{3.9}$$

with w a nonnegative weight function that reflects the relative importance of detecting a changepoint at τ. In particular, the choice

$$w(\tau) = \{\tau(n-\tau)/n\}^{\frac{1}{2}}$$

gives most weight to changepoints in the central part of the data series. Expressed in terms of the cusums, this leads, via (3.9), to the statistic

$$W_n = max\{|C_\tau|/\hat{\sigma}_n; 1 \le \tau < n\}. \tag{3.10}$$

Worsley (1979), extending a technique first proposed by Hawkins (1977), has derived a numerical algorithm which, in principle, can yield percentage points of the exact distribution of V_n in (3.6). Worsley gives percentage points for $n \le 10$. The extent of the computations involved makes the technique impractical for larger sample sizes. However, highly accurate and easily computable approximations to the tail probabilities $pr\{W_n \ge b\}$ can be found by using the improved Bonferroni bounds developed by Worsley (1982); see, in particular, section 3.2 of his paper. The most complete results available to date for the statistics in (3.7) and (3.10) are summarized in James, James & Siegmund (1987). For instance, they give the approximation

$$pr\{W_n \ge b|H_0\} \approx 2(1-4\gamma^2)^{(n-3)/2}.\nu(4\gamma/(1-4\gamma^2)^{\frac{1}{2}}) \tag{3.11}$$

where $\gamma = b/n$ is assumed to lie between 0 and $\frac{1}{2}$, and

$$\nu(x) = 2x^{-2}exp\{-2\sum_{n=1}^{\infty} n^{-1}\Phi(-xn^{\frac{1}{2}}/2)\}, \qquad x > 0,$$

with Φ denoting the normal $(0,1)$ distribution function. They also give an approximation to $pr\{V_n \ge b|H_0\}$ which involves integration of the function $\nu(x)$; see their formula (21). In either case, numerical computations of a greater or lesser extent are required in order to evaluate significance probabilities.

The statistics considered thus far were derived by the maximum likelihood ratio principle. Chernoff & Zacks (1964), Gardner (1969) and MacNeill (1974) have considered tests derived from Bayesian likelihood ratio principles and much effort has been devoted to studying the frequentist properties of such tests. This approach leads to statistics such as

$$\hat{\sigma}_n^{-2}\sum_{k=1}^{n-1} C_k^2/(k(n-k)/n\}, \tag{3.12}$$

$$\hat{\sigma}_n^{-2}\sum_{k=1}^{n-1} C_k^2 \tag{3.13}$$

and

$$\hat{\sigma}_n^{-1}\sum_{k=1}^{n-1} C_k = \hat{\sigma}_n^{-1}\sum_{i=1}^{n} i(x_i - \bar{x}_n). \tag{3.14}$$

The exact distributions of (3.12) and (3.13) are not known. They do have well-known limiting forms, but it is not known for what values of n these constitute adequate approximations to the exact distributions. See, however, section 3.2 below.

Sen & Srivastava (1975b) have shown that the exact distribution of a suitably standardized version of the square of (3.14) is Beta $(1/2, n/2-1)$. Observe also that (3.14) is proportional to the correlation coefficient between x_i and i, $1 \leq i \leq n$. Thus, among all tests considered above, the one that rejects for large values of (3.14) seems to be the most obvious and useful one from an applied statistician's point of view. Furthermore, there is some evidence (Sen & Srivastava, 1975b, section 2, James, James & Siegmund, 1987, section 5) that (3.14) competes well with other tests in terms of its power to detect a change in mean, provided that the change does not occur near the beginning or the end of the sequence.

3.2 One-parameter exponential families

Suppose x_i belongs to a one-parameter exponential family with density

$$f(x_i) = exp\{x_i a(\mu_i) - b(\mu_i) + c(x_i)\}.$$

The likelihood ratio statistic for testing H_0 against the alternative (3.1), with τ unknown, is

$$\Lambda_n = max\{Z_\tau, 1 \leq \tau < n\}$$

where

$$\begin{aligned} Z_\tau =& 2\left[\tau\{\bar{x}_\tau a(\bar{x}_\tau) - b(\bar{x}_\tau)\} + (n-\tau)\{\bar{x}^*_\tau a(\bar{x}^*_\tau) - b(\bar{x}^*_\tau)\}\right. \\ & \left. -n\{\bar{x}_n a(\bar{x}_n) - b(\bar{x}_n)\}\right]. \end{aligned}$$

In the special case of a normal distribution (with known variance), Hawkins (1977) developed a recursion formula for $pr\{\Lambda_n \leq b|H_0\}$ from which he computed percentage points for $n \leq 50$. Worsley (1983b, 1986), by extending Hawkins' technique, has developed a recursion formula for the conditional probability $pr\{\Lambda_n \leq b|\bar{x}_n, H_0\}$. The motivation for considering conditional probabilities is that, under $H_0, \bar{x}_n$ is sufficient for the common parameter value μ. The conditional null distribution of Λ_n does therefore not involve the unknown μ. Worsley has successfully implemented his algorithm in the binomial and Poisson cases. For other

continuous distributions, such as the gamma and the inverse Gaussian, problems with numerical integration lead to seemingly insurmountable difficulties. In the negative exponential case he obtained the requisite conditional probabilities by a different method, exploiting a well-known characterization of this distribution. Also in this case, Hsu (1979) has investigated the behaviour of the statistic (3.14).

In the special case of Bernoulli variables (each x_i either zero or one) Pettitt (1979) proposes a test that rejects H_0 for large values of

$$K_n = n\, max\{|C_\tau|; 1 \leq \tau < n\},$$

where C_k is the cusum defined in (2.1). He shows that the exact null distribution of K_n, conditional on $\bar{x}_n$, is that of the well-known Kolmogorov-Smirnov two-sample statistic.

When the sample size, n, is large, considerable simplification can often be achieved by applying a variance stabilizing transformation to the data, that is, by replacing x_i by $y_i = h(x_i)$ where h satisfies the differential equation

$$\frac{d}{du}h(u) = \left[\frac{d^2}{du^2}b(u)\right]^{-\frac{1}{2}}$$

Then $y_1, \cdots, y_n$ are independent with a common variance ≈ 1 and means approximately equal to $h(\mu_1), \cdots, h(\mu_n)$. It is well known that such a transformation leads to a distribution that is more symmetric about its mean and, hence, improves the approximation to normality of partial sums of the observations. A particularly simple test of H_0 can then be based on the statistic

$$\sum_{k=1}^{n-1} C_k = \sum_{i=1}^{n} i(y_i - \bar{y}_n) \tag{3.15}$$

where C_k is the cusum defined at (2.1) with y_i in place of x_i. For large n, the null distribution of this is approximately normal with mean 0 and variance 1/12. Alternatively, tests can be based on the statistics

$$\sum_{k=1}^{n-1} C_k^2/\{k(n-k)/n\} \tag{3.16}$$

and

$$n^{-2}\sum_{k=1}^{n} C_k^2. \tag{3.17}$$

Simple approximations to the large sample tail probabilities of (3.16) and (3.17) are available. Namely,

$$pr\{(3.16) \geq b\} \approx (6/\pi b)^{\frac{1}{2}} e^{-b} \tag{3.18}$$

and

$$pr\{(3.17) \geq b\} \approx pr\{\chi_1^2 \geq \pi^2(b - 1/6) + 1\}, \tag{3.19}$$

where χ_1^2 denotes a random variable having a chi-square distribution with one degree of freedom. Some aspects of this transformation approach have been investigated by Van Wyk (1986).

3.3 Some general comments

1. The tests discussed thus far are all based on an assumed form of the underlying density. In real life, however, the statistician rarely has definite, *a priori*, knowledge of the form of the underlying distribution. In such cases, the fit of the data to the assumed distributional form would have to be assessed. But there is an obvious difficulty inherent in this if the data contain an unknown number of changepoints. From this perspective, procedures based on ranks, which are distribution-free under H_0, would seem to hold a definite advantage.

2. The presence of nuisance parameters also leads to considerable complications in testing for the presence of a changepoint. We saw that many procedures exist to test for a change in mean of a normal distribution. However, there seems to be no satisfactory procedure to test for a change in the variance of a normal distribution with fixed but unknown mean. By contrast, there exist simple rank tests for this problem.

3. It should also be noted that the tests discussed thus far have all been derived under a model involving at most one changepoint. The power of these tests against alternatives involving multiple changes is unknown.

4. There exists no single, comprehensive, study of the relative power properties of the tests discussed in section 3.1 and 3.2 or of those to be discussed in section 4 below. The available results are rather fragmentary, based mostly on Monte Carlo simulation (Sen & Srivastava, 1975a, section 3.4; 1975b, section 2.2; Wolfe & Schechtman, 1984, section 3), but also on theoretical calculation (Worsley 1986, section 4, James, James & Siegmund, 1987, sections 4 and 5). Our purpose here is not to review these results. Rather we remark that, probably due to limited availability of computer time, the available Monte Carlo power comparisons are based on relatively small sample sizes, $n \leq 50$ in all cases. Yet, with one or two exceptions (Worsley, 1983b, Pettitt, 1979), all published examples of changepoint data involve sample sizes substantially larger than 50. The relevance of these power studies to the applications that stimulated the research is therefore open to question.

4. RANK TESTS

Suppose $x_1, \cdots, x_n$ are independent with unspecified density functions $f_{\theta_1}, \cdots f_{\theta_n}$ and let r_i denote the rank of x_i, $1 \leq i \leq n$. It is well known that the distribution of the rank vector $(r_1, \cdots, r_n)$ does not depend on the underlying density when $H_0 : \theta_1 = \cdots = \theta_n$ is true. A simple way of constructing nonparametric tests is to replace x_i in your favourite parametric procedure by $\varphi\{r_i/(n+1)\}$, where φ is an appropriate "score function", defined and square integrable on the interval $(0, 1)$. The choice of score function is dictated by the type of parameter (location or scale) involved and, ideally, by the tail thickness of the underlying density. Commonly used choices are $\varphi(u) = 2u - 1$ (Wilcoxon score) or $\varphi(u) = \Phi^{-1}(u)$ (normal score) for location parameters and $\varphi(u) = (2u-1)^2$ (Mood score) or $\varphi(u) = -log\ u$ (Savage score) for scale parameters. Let

$$\bar{\varphi} = n^{-1}\sum_{i=1}^{n} \varphi\{i/(n+1)\},$$

$$\sigma_\varphi^2 = (n-1)^{-1}\sum_{i=1}^{n} \{\varphi(i/(n+1)) - \bar{\varphi}\}^2$$

and define the standardized rank score of x_i by

$$s(r_i) = [\varphi\{r_i/(n+1)\} - \bar{\varphi}]/\sigma_\varphi.$$

The rank analogues of the cusums C_k in (2.1) are

$$C_k = \sum_{i=1}^{k} s(r_i), \quad 1 \le k < n.$$

Any parametric test based on cusums can now be freed of distributional assumptions and the need to estimate nuisance location or scale parameters by the simple expedient of replacing the cusums by their rank analogues.

To illustrate, the rank analogue of the statistic (3.14) is

$$\sum_{i=1}^{n} is(r_i), \tag{4.1}$$

which is in the form of a simple linear rank statistic. The distribution theory of such statistics is well known; see Hájek and Šidák (1976, chapters 2 and 5). In particular, the null distribution does not depend on the variance of the underlying distribution. Thus, in contrast to (3.14), no estimate of variance is required in order to implement the test. Furthermore, it is known that such linear rank statistics approach their limiting normal distribution form quite rapidly. In particular, (4.1) is approximately normally distributed with mean zero and variance $\sigma_\varphi^2 n(n^2-1)/12$. Similar remarks also hold in respect of the rank analogues of the statistics (3.12) and (3.13). For samples of size $n > 30$ good approximations to the respective significance probabilities are given by the right-hand sides of (3.18) and (3.19). If a smaller sample size is involved, the required null distributions can be obtained easily by computer simulation.

As a further illustration, suppose that f_{θ_i} belongs to the scale parameter family generated by an unknown symmetric density with unknown median μ:

$$f_{\theta_i} = \theta_i g(\theta_i(x-\mu)), \qquad \theta_i > 0$$

with

$$g(\mu + x) = g(\mu - x); \qquad x \epsilon R^1.$$

If g is the normal (0,1) density this is the situation discussed in 2 of section 3.3 above, for which no satisfactory parametric procedure yet exists. However, if the Mood score $\varphi(u) = (2u-1)^2$ is used, the rank analogue of any of the statistics (3.12), (3.13) or (3.14) results in a valid test for a change in the scale parameter.

The rank analogue of W_n in (3.10) was first studied by Pettitt (1979) in the case of a Wilcoxon score. the general case was studied by Lombard (1983) and Sen (1984). A small Monte Carlo study by the author (unpublished) suggests that the approximation (James, James & Siegmund, 1987, formula (20))

$$P\{W_n \ge b|H_0\} \approx 2exp\{-2n^{-1}(b+0,583)^2\}$$

is quite accurate for $n \ge 30$ and a variety of score functions in common use. The situation regarding the rank analogue of the likelihood ratio statistic V_n in (3.7) is less satisfactory. No accurate large sample approximation to its distribution is known. Schechtman (1982), working with a Wilcoxon score, has found percentage points for a range of values $n \le 80$ by Monte Carlo simulation.

Wolfe & Schechtman (1984) have made some small-sample power comparisons between various statistics and have also reviewed the work in the area of rank tests for the changepoint problem prior to 1984.

Generalizations to the problem of multiple changes were considered by Lombard (1987) for data on the real line and by Lombard (1986) for circular data. The limiting forms of the test statistics appear as quadratic functionals of a Brownian bridge process and some general results regarding the distributions of such functionals are given by Erasmus & Lombard (1988).

5. ESTIMATION OF THE CHANGEPOINT

When the hypothesis of no change is rejected, one faces the problem of estimating the changepoint. An obvious estimate of τ is the point, $\hat{\tau}$, at which the absolute standardized cusum $|C_k|/\{k(n-k)/n\}^{\frac{1}{2}}$ attains its maximum. Here, C_k denotes either the cusum (2.1), or its rank analogue. In the case of a normal distribution with possible change in the mean, this is precisely the maximum likelihood estimate of τ under the model (3.1). Hinkley (1970) gave an approximation to the distribution of $\hat{\tau}-\tau$, valid in large samples, and noted that this distribution is independent of τ. He also noted that $\hat{\tau}$ is not a sufficient statistic for τ and that, consequently, inference about τ based on the marginal distribution of $\hat{\tau}$ would be inefficient. By the conditionality principle (Cox & Hinkley, 1974, section 2.3 (iii)) such inference should be made conditionally on an appropriate ancillary statistic.

Cobb (1978) constructed an approximately ancillary statistic and showed how a conditional confidence set for τ could be obtained. He points out, however, that a proper justification of his results requires that the distributions of the data before and after the changepoint be fully specified. Such a requirement is, of course, hardly likely to be met in any practical application. In an interesting subsequent development, Hinkley & Schechtman (1987) have considered the use of conditional bootstrap methods in this problem. These methods seem to hold a promise of successful application in more complicated situations.

It is interesting to note that consistent estimation of a changepoint, in the usual sense, is not possible. For if $x_1,\cdots,x_\tau$ have a common density f_1, while $x_{\tau+1},\cdots,x_n$ have a common density f_2, then the log-likelihood for τ is

$$\sum_{i=1}^{\tau} log f_1(x_i)+\sum_{i=\tau+1}^{n} log f_2(x_i). \tag{5.1}$$

Increasing the sample size n obviously has no effect on the shape of the log-likelihood in the immediate neighbourhood of τ, the true changepoint. A weak form of consistency does, however, hold. If it is assumed that τ varies with n in such a way that $\tau/n \to \lambda, 0<\lambda<1$, then $\hat{\tau}/n \to \lambda$ in probability when $\hat{\tau}$ is the value of τ at which (5.1) attains its maximum (Bhattacharya, 1987, Darkhovskh, 1976).

In completely nonparametric situations, very few results of practical use are available. Schechtman (1983) has derived a conservative confidence bound for the median of a distribution based on her (1982) rank analogue of the likelihood ratio statistic V_n. Lombard (1986) has used the bootstrap to obtain a confidence set in a nonparametric situation, but a proper justification for its use is lacking. Pettitt (1980) has studied the sampling distributions of some point estimates by Monte Carlo methods.

It is fair to say that in nonparametric cases the question of confidence set estimation of a changepoint still awaits a satisfactory solution.

6. OTHER PROBLEMS

The changepoint problems discussed thus far have all had a relatively simple structure. In this section some references to more complicated problems are given.

First, there is an obvious generalization of the situation considered in section 3.1 to multivariate normal distributions. Srivastava & Worsley (1986) have derived an upper bound for the type I error probability of the likelihood ratio test in this case. Exact results, or refined approximations, parallelling those in the univariate case are not available.

Next, suppose that $x_1,\cdots,x_n$ have a linear model structure

$$x_i = A_i\beta_i + e_i,$$

where A_i is a $1\times p$ row vector of known constants, β_i is a $p\times 1$ column vector of unknown regression constants and $e_1,\cdots,e_n$ are independent and identically distributed random variables. The simplest regression changepoint model assumes that

$$\beta_i=\beta \text{ for } i\le\tau; \beta_i=\beta_* \text{ for } i>\tau. \tag{6.1}$$

There exists a large body of work concerning tests of $H_0; \beta = \beta_*$ against the alternative that (6.1) holds for some τ. Shaban (1980) has published an annotated bibliography of work in this field. Even under the assumption that the errors are normally distributed, neither the exact nor the asymptotic distribution of the likelihood ratio statistic is known; see Feder (1975a, 1975b) and Hawkins (1980). Some useful approximations and bounds for the tail probabilities have been derived by Worsley (1983a). The reader can consult Sen (1984) for details concerning the use of rank tests in this problem. In a recent related development, Picard (1985) has considered a situation in which $x_1, \cdots, x_\tau$ and $x_{\tau+1}, \cdots, x_n$ form parts of two time series with spectral densities f_1 and f_2 respectively.

REFERENCES

ALI, M.M. & GIACCOTTO, C. (1982). The identical distribution hypothesis for stock market prices – Location- and scale alternatives. *J. Amer. Statist. Assoc.* **77**, 19-28.

BHATTACHARYA, P.K. (1987). Maximum likelihood estimation of a changepoint in the distribution of independent random variables: General multi-parameter case. *J. Mult. Anal.*, **23**, 183-208.

BIRKS, H.J.B. (1987). Multivariate analysis of stratigraphic data in geology: A review. *Chemometrics and Intelligent Laboratory Systems*, **2**, 109-126.

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

COBB, G.W. (1978). The problem of the Nile: Conditional solution to a changepoint problem. *Biometrika*, **65**, 243-251.

COMMENGES, D. & SEAL, J. (1985). The analysis of neuronal discharge sequences: Change-point estimation and comparison of variances. *Statist. in Med.*, **4**, 91-104.

COX, D.R. & HINKLEY, D.V. (1974). *Theoretical Statistics.* London: Chapman & Hall.

DARKHOVSKH, B.S. (1976). A nonparametric method for the a posteriori detection of the "disorder" time of a sequence of independent random variables. *Theor. Probab. Appl.*, **21**, 178-183.

ERASMUS, C.M. & LOMBARD, F. (1988). The asymptotic distribution of quadratic forms occurring in changepoint problems. *Canadian Journal of Statistics*, **16**, 259-268.

FEDER, P.I. (1975a). On asymptotic distribution theory in segmented regression problems – identified case. *Ann. Statist.*, **3**, 49-83.

FEDER, P.I. (1975b). The log likelihood ratio in segmented regression. *Ann. Statist.*, **3**. 84-97.

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

HÁJEK, J. & ŠIDÁK, Z. (1967). *Theory of rank tests.* New York: Academic Press.

HAWKINS, D.M. (1977). Testing a sequence of observations for a shift in location. *J. Amer. Statist. Assoc.*, **72**, 180-186.

HAWKINS, D.M. (1980). A note on continuous and discontinuous segmented regressions. *Technometrics*, **22**, 443-444.

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

HINKLEY, D. & SCHECHTMAN, E. (1987). Conditional bootstrap methods in the mean-shift model. *Biometrika*, **74**, 85-93.

HSU, D.A. (1979). Detecting shifts of parameter in gamma sequences with applications to stock price and air traffic flow analysis. *J. Amer. Statist. Assoc.*, **74**, 31-40.

JAMES, B., JAMES, K.L., & SIEGMUND, D. (1987). Tests or a changepoint. *Biometrika*, **74**, 71-83.

LEVIN, B. & KLINE, J. (1985). The cusum test of homogeneity with an application in spontaneous abortion epidemiology. *Statist. in Med.*, **4**, 469-488.

LOMBARD, F. (1983). Asymptotic distributions of rank statistics in the changepoint problem. *South African Statist. J.*, **17**, 83-105.

LOMBARD, F. (1986). The change-point problem for angular data: A non-parametric approach. *Technometrics*, **28**, 391-397.

LOMBARD, F. (1987). Rank tests for changepoint problems. *Biometrika*, **74**, 615-624.

MACNEILL, I.B. (1974). Tests for change of parameters at unknown times and distributions of some related functionals on Brownian motion. *Ann. Statist.*, **2**, 950-962.

MATTHEWS, D.E. & FAREWELL, V.T. (1982). On testing for a constant hazard against a change-point alternative. *Biometrics*, **38**, 463-468.

PETTITT, A.N. (1979). A non-parametric approach to the changepoint problem. *Appl. Statist.*, **28**, 126-135.

PETTITT, A.N. (1980). Some results on estimating a change-point using non-parametric type statistics. *J. Statist. Comput. Simul.* **11**, 261-272.

PICARD, D. (1985). Testing and estimating change-points in time series. *Adv. Appl. Prob.*, **17**, 841-867.

SCHECHTMAN, E. (1982). A nonparametric test for detecting changes in location. *Commun. Statist. - Theor. Meth.* **A11**(13), 1475-1482.

SCHECHTMAN, E. (1983). A conservative nonparametric distribution-free confidence bound for the shift in the changepoint problem. *Commun. Statist. - Theor. Meth.* **A12**(21), 2455-2464.

SEN, A. & SRIVASTAVA, M.S. (1975a). On tests for detecting change in mean. *Ann. Statist.*, **3**, 98-108.

SEN, A. & SRIVASTAVA, M.S. (1975b). Some one-sided tests for change in level. *Technometrics*, **17**, 61-64.

SEN, P.K. (1984). *Theory and applications of sequential nonparametrics.* Philadelphia: Society for Industrial and Applied Mathematics.

SHABAN, S.A. (1980). Change-point problem and two-phase regression: An annotated bibliography. *Int. Statist. Rev.*, **48**, 83-93.

SRIVASTAVA, M.S. & WORSLEY, K.J. (1986). Likelihood ratio tests for a change in the multivariate normal mean. *J. Amer. Statist. Assoc.*, **81**, 199-204.

VAN WYK, J.L. (1986). *Aspects of a changepoint problem and its applications.* M.Sc. dissertation. Department of Statistics, University of South Africa.

WOLFE, D.A. & SCHECHTMAN, E. (1984). Nonparametric statistical procedures for the changepoint problem. *J. Statist. Plann. Inference*, **9**, 389-396.

WORSLEY, K.J. (1979). On the likelihood ratio test for a shift in location of normal populations. *J. Amer. Statist. Assoc.*, **74**, 365-367.

WORSLEY, K.J. (1982). An improved Bonferroni inequality and applications. *Biometrika*, **69**, 297-302.

WORSLEY, K.J. (1983a). Testing for a two-phase multiple regression. *Technometrics*, **25**, 35-42.

WORSLEY, K.J. (1983b). The power of likelihood ratio and cumulative sum tests for a change in a binomial probability. Biometrika, **70**, 455-465.

WORSLEY, K.J. (1986). Confidence regions and tests for a changepoint in a sequence of exponential family random variables. *Biometrika*, **72**, 91-104.

Manuscript received, 1988.4.4, revised, 1988.10.4.