*Biometrika* (1989), 76, 3, *pp.* 409–23
*Printed in Great Britain*

# The likelihood ratio test for a change-point in simple linear regression

BY HYUNE-JU KIM

*Department of Mathematics, Syracuse University, Syracuse, New York 13244, U.S.A.*

AND DAVID SIEGMUND

*Department of Statistics, Stanford University, Stanford, California 94305, U.S.A.*

SUMMARY

We consider likelihood ratio tests to detect a change-point in simple linear regression (a) when the alternative specifies that only the intercept changes and (b) when the alternative permits the intercept and the slope to change. Approximations for the significance level are obtained under reasonably general assumptions about the empirical distribution of the independent variable. The approximations are compared with simulations in order to assess their accuracy. For the model in which only the intercept is allowed to change, a confidence region for the change-point and an approximate joint confidence region for the change-point, the difference in intercepts, and the slope are obtained by inversion of the appropriate likelihood ratio tests.

*Some key words*: Change-point; Linear regression; Maxima of random fields.

## 1. INTRODUCTION

Given $x_1, \ldots, x_m$, suppose that $y_i$ $(i=1, \ldots, m)$ are independent and normally distributed with common variance $\sigma^2$. For some $j$, the change-point, the expectation of $y_i$ equals $\alpha_0+\beta_0 x_i$ if $i \leqslant j$ and equals $\alpha_1+\beta_1 x_i$ if $i>j$. This paper is primarily concerned with the likelihood ratio tests of the hypothesis of no change, $H_0$: $\beta_0=\beta_1$ and $\alpha_0=\alpha_1$, against one of the alternatives, $H_1$: $\beta_0=\beta_1=\beta$ and there exists a $j$ $(1 \leqslant j<m)$ such that $\alpha_0 \neq \alpha_1$ or $H_2$; there exists a $j$ $(1 \leqslant j<m)$ such that $\beta_0 \neq \beta_1$ or $\alpha_0 \neq \alpha_1$. A secondary consideration is when $H_1$ holds, to obtain confidence regions for $j$ and jointly for $j$, $\alpha_0-\alpha_1$, and $\beta$. Since problems similar to these have been widely discussed in a variety of formulations with a variety of applications in mind, we begin with a brief review and motivation for our formulation.

The test of $H_0$ against a more general alternative, where the variance is also allowed to change, was discussed by Quandt (1958, 1960) in reference to econometric models which may change over time. Quandt also noted on the basis of a Monte Carlo experiment that a proposed chi-squared approximation to the significance level of the likelihood ratio test is very poor. Research in this general direction was stimulated by the seminal paper of Brown, Durbin & Evans (1975), who discussed multiple regression models, noted the mathematical difficulties associated with the sampling distribution of Quandt's likelihood ratio statistic, and introduced recursive residuals to circumvent this difficulty in testing the constancy of regression models over time. Brown et al. (1975) considered their test to be a data analytic tool and steadfastly refused, in spite of prompting by discussants of their paper, to specify alternative hypotheses or indicate in detail what one should do if one found $H_0$ to be untenable. Worsley (1983a) gave approximate upper bounds for the significance level of the likelihood ratio test. Numerical results indicate that his bounds are reasonably good in small samples.

The change-point problem discussed here should be contrasted with problems of 'broken line' regression, for example Hinkley (1969, 1971), where one constrains the regression function to be continuous at the change-point. A motivating example for broken line regression is dose response curves, where there is a more or less sudden change in the rate at which response varies with dose. The quite different mathematical theory associated with the sampling distribution of the likelihood ratio statistic in this case has been discussed by Davies (1987), Johnstone & Siegmund (1989), and in an unpublished paper by M. Knowles and D. Siegmund.

Recently, motivated by problems involving changes in the rates of industrial accidents, of congenital defects in new born babies, or of spontaneously aborted foetuses, a number of authors have studied simpler change-point models without covariates. See, for example, Worsley (1983b, 1986), Raferty & Akman (1986) and Levin & Kline (1985). In these problems the exact form of the alternative hypothesis plays an important role. If the simple model of no change is discarded, one wants to estimate the times of changes in order to formulate hypotheses about possible causes.

Our motivation comes from this epidemiological perspective. Thus, for us the test of $H_0$ against $H_1$, where only the intercept of the regression line is allowed to change, is an extension of the simple change-point problem discussed by Worsley (1983b) and others to allow the inclusion of a covariate. See Maronna & Yohai (1978) for another possible application. The importance of such an extension can be motivated by consideration of the British coal mining data (Jarrett, 1979), which gives the time intervals between accidents involving more than ten deaths in British coal mines, from 1851 to 1962. These data have been treated by regression methods (Cox & Lewis, 1966, p. 42) and by change-point methods (Worsley, 1986; Raferty & Akman, 1986). Worsley (1986) estimates the time of change to lie in the period 1887-1895 and offers a tentative explanation in terms of changes in safety practices implemented at about that time. In contrast, regression of the accident rate on time might lead one to attribute long run changes to gradual changes in the coal mining industry, in particular increased mechanization and/or decreased use of coal, leading to a decrease in the number of man-hours worked. A change-point model without covariates may suggest an abrupt change where only a gradual trend exists, exaggerate the magnitude of a change, or introduce unwanted bias into an estimate of its location.

Our goal in this paper is to indicate the potential for using recently developed methods for approximating boundary crossing probabilities to study analytically likelihood ratio tests of constancy of a regression model over time. A difficulty associated with this problem is that under the null hypothesis the distribution, even the asymptotic distribution, of the test statistic depends on the values of the independent variable. We consider two quite different cases: (i) the values of the $x_i$ are effectively random, i.e. for all $j$ the empirical distribution of the $x_i$ for $i \leq j$ is about the same as that for $i > j$; (ii) the values of the $x_i$ are a function of $i$, say $x_i = f(i/m)$. In case (i) the limiting distribution under $H_0$ of the likelihood ratio statistic is the same as when there is no covariate; the second case is more difficult and interesting. Under some assumptions on the empirical distribution of the $x$'s, which include equally spaced $x$'s as a special case, we obtain approximations to the significance level of the likelihood ratio tests of $H_0$ against either $H_1$ or $H_2$. The approximation involves an asymptotic expression for the tail of the distribution of the maximum of some novel random fields. Either of two methods can be adapted for our purposes: (i) that developed in a series of papers by Pickands (1969), Qualls & Watanabe (1973) independently of and parallel to Bickel & Rosenblatt (1973) and Hogan & Siegmund (1986), or (ii) the method developed by Woodroofe (1976, 1982) in one-dimensional time and recently extended to higher dimensions by Siegmund (1988b). See also Aldous (1989).

In view of the importance we attach to problems of estimation, for the model specified by $H_1$ we also give a confidence region for the change-point $j$ and an approximate joint confidence region for $j$, the difference $\alpha_0-\alpha_1$, and $\beta$. We believe that in some cases the latter confidence region can shed light on a problem posed by Cox (1961): to choose between a change-point model without covariates and a regression model without change-points.

The paper is organized as follows. Section 2 contains a description of our results and some numerical examples illustrating the accuracy of the approximations. Section 3 contains a sketch of a proof of one of these results. Confidence regions are discussed in § 4. Some general remarks and open problems form the contents of § 5.

## 2. Approximations and numerical examples

The following notation will be used throughout this paper:

$$\bar{y}_i = i^{-1}\sum_{k=1}^{i} y_k, \quad \bar{y}_i^* = (m-i)^{-1}\sum_{k=i+1}^{m} y_k,$$

$$Q_{yyi} = \sum_{k=1}^{i}(y_k-\bar{y}_i)^2, \quad Q_{yyi}^* = \sum_{k=i+1}^{m}(y_k-\bar{y}_i^*)^2, \quad Q_{xyi} = \sum_{k=1}^{i}(x_k-\bar{x}_i)(y_k-\bar{y}_i), \ldots;$$

$$\hat{\beta} = Q_{xym}/Q_{xxm}, \quad \hat{\alpha}_i = \bar{y}_i - \hat{\beta}\bar{x}_i, \quad \hat{\alpha}_i^* = \bar{y}_i^* - \hat{\beta}\bar{x}_i^*, \quad \hat{\sigma}^2 = m^{-1}(Q_{yym} - Q_{xym}^2/Q_{xxm}).$$

A tedious calculation shows that the likelihood ratio test of $H_0$ against $H_1$, generalized slightly as suggested by James, James & Siegmund (1987), rejects $H_0$ for large values of

$$\max_{m_0 \le i \le m_1} |U_m(i)|/\hat{\sigma}, \tag{2·1}$$

where

$$\begin{aligned} U_m(i) &= \left(\frac{i}{1-i/m}\right)^{\frac{1}{2}}\left(\frac{\bar{y}_i - \bar{y}_m - \hat{\beta}(\bar{x}_i-\bar{x}_m)}{[1-i(\bar{x}_i-\bar{x}_m)^2/\{Q_{xxm}(1-i/m)\}]^{\frac{1}{2}}}\right) \\ &= (\hat{\alpha}_i - \hat{\alpha}_i^*)\left[\frac{i(1-i/m)}{1-i(\bar{x}_i-\bar{x}_m)^2/\{Q_{xxm}(1-i/m)\}}\right]^{\frac{1}{2}}. \end{aligned} \tag{2·2}$$

If it were known that the only possible value of the change-point $j$ is $j=i$, the appropriate two-sample statistic for testing $\alpha_1=\alpha_2$ would be $|U_m(i)|/\hat{\sigma}$. The maximization over $i$ in (2·1) searches for the unknown value of $j$.

*Remark.* Brown et al. (1975), among others, mention the possibility of using a functional of the process of cumulative sums of residuals,

$$\hat{\sigma}^{-1}\sum_{k=1}^{i}[y_k - \{\bar{y}_m - \hat{\beta}(x_k-\bar{x}_m)\}] \quad (i=1,\ldots,m) \tag{2·3}$$

to test $H_0$ against unspecified, presumably quite general alternatives. The suggestion seems not to have been pursued, evidently because of difficulties in determining the associated sampling distributions. The statistics (2·1) is a particular functional of the process (2·3). The methods discussed below are also applicable to other suitably normalized maxima of (2·3).

Under $H_0$, $U_m(i)$ $(i=1,\dots,m-1)$ is a zero mean Gaussian process with covariance function given for $i \leq k$ by

$$\sigma^{-2}\operatorname{cov}\{U_m(i), U_m(k)\} = \left\{\frac{i(m-k)}{k(m-i)}\right\}^{\frac{1}{2}} \frac{D_m(i,k)}{\{D_m(i,i)D_m(k,k)\}^{\frac{1}{2}}}, \quad (2\cdot4)$$

where

$$D_m(i,k) = 1-(\bar{x}_i-\bar{x}_m)(\bar{x}_k-\bar{x}_m)k/\{Q_{xxm}(1-k/m)\} \quad (i \leq k). \quad (2\cdot5)$$

The first factor in (2·4), namely $[i(m-k)/\{k(m-i)\}]^{\frac{1}{2}}$, is the covariance function of

$$W_0(i/m)/\{(i/m)(1-i/m)\}^{\frac{1}{2}} \quad (i=1,\dots,m-1), \quad (2\cdot6)$$

where $W_0$ is a Brownian Bridge process. This process is central to the developments of James et al. (1987).

In order to proceed it is helpful to make some assumptions about the empirical distribution of the $x$'s.

A particularly simple case occurs if $x_1,\dots,x_m$ form the initial segment of an infinite sequence $x_1, x_2, \dots$ such that $\bar{x}_m \to \gamma$ and $m^{-1}Q_{xxm} \to \tau^2$ for some $\gamma$ and $\tau^2>0$ as $m\to\infty$. This would happen if the $x$'s are observed values of a stationary process with finite second moment. Then $D_m([ms],[mt]) \to 1$, and, for any $0<t_0<t_1<1$, $\sigma^{-1}U_m([mt])$ converges weakly on $[t_0, t_1]$ to $W_0(t)/\{t(1-t)\}^{\frac{1}{2}}$. In this case it seems reasonable to use as an approximation to the distribution of (2·1) the distribution of the maximum over $m_0 \leq i \leq m_1$ of the absolute value of (2·6), for which a good approximation to tail probabilities is known (Siegmund, 1985, p. 237).

At the opposite extreme from the case of 'random' $x$'s is the case where $x_i$ denotes the time at which the $i$th observation is made, so $x_1<\dots<x_m$. Particularly interesting is the special case of equally spaced observations, $x_i=i/m$ $(i=1,\dots,m)$. More generally assume that $x_i=x_i^{(m)}$ is a doubly indexed array of the form $x_i=f(i/m)$ for some bounded, continuous, nonconstant function $f$ defined on $[0,1]$; let

$$g(t)=\left\{\int_0^1 f(u)\,du - t^{-1}\int_0^t f(u)\,du\right\}\Big/\left((1-t)\left[\int_0^1 f^2(u)\,du-\left\{\int_0^1 f(u)\,du\right\}^2\right]^{\frac{1}{2}}\right), \quad (2\cdot7)$$

$$D(s,t)=1-t(1-s)g(s)g(t) \quad (s\leq t). \quad (2\cdot8)$$

For the special case $f(t)\equiv t$, $g(t)\equiv\sqrt{3}$. It follows easily from (2·5) that

$$D_m([ms],[mt]) \to D(s,t) \quad (m\to\infty). \quad (2\cdot9)$$

An approximation for

$$p_1 = \mathrm{pr}\{\max_{m_0\leq i\leq m_1} \hat{\sigma}^{-1}|U_m(i)| \geq b\} \quad (2\cdot10)$$

is given in Theorem 1. The approximation involves the special function

$$\nu(x)=2x^{-2}\exp\left\{-2\sum_{n=1}^{\infty} n^{-1}\Phi(-\tfrac{1}{2}x\sqrt{n})\right\} \quad (x>0), \quad (2\cdot11)$$

where $\Phi$ denotes the standard normal distribution function. For numerical purposes it often suffices to use the small $x$ approximation (Siegmund, 1985, Ch. 10)

$$\nu(x) \doteq \exp(-\rho x), \tag{2·12}$$

where $\rho$ is a numerical constant approximately equal to 0·583.

THEOREM 1. *Assume $H_0$ is true, so*

$$y_i = \alpha + \beta x_i + \varepsilon_i \quad (i = 1, \ldots, m),$$

*where the $\varepsilon_i$ are independent $N(0, \sigma^2)$. Suppose $x_i = f(i/m)$, let g be defined by* (2·7) *and let*

$$\mu(t) = 1/[2t(1-t)\{1 - g^2(t)t(1-t)\}].$$

*Assume b, $m \to \infty$ such that, for some $0 < c < 1$ and $0 \leq t_0 < t_1 \leq 1$,*

$$b/\sqrt{m} \to c, \quad m_i \sim mt_i \quad (i = 0, 1).$$

*Then for $U_m$ defined in* (2·2), *the probability $p_1$ in* (2·10) *satisfies*

$$p_1 \sim (2/\pi)^{\frac{1}{2}} b(1 - b^2/m)^{\frac{1}{2}(m-5)} \int_{t_0}^{t_1} \mu(t)\nu[\{2c^2\mu(t)/(1-c^2)\}^{\frac{1}{2}}]\, dt, \tag{2·13}$$

*where $\nu$ is defined in* (2·11) *and given approximately in* (2·12).

*Remark.* There is also a version of Theorem 1 when $\sigma^2$ is known. An appropriate test statistic is

$$\max_{m_0 \leq i \leq m_1} \sigma^{-1}|U_m(i)|.$$

If $b \to \infty$, $m \to \infty$ with $b/\sqrt{m} \to c > 0$, then for any $0 \leq t_0 < t_1 \leq 1$

$$\operatorname{pr}\{\max_{t_0 < i/m < t_1} \sigma^{-1}|U_m(i)| \geq b\} \sim 2b\phi(b) \int_{t_0}^{t_1} \mu(t)\nu[\{2c^2\mu(t)\}^{\frac{1}{2}}]\, dt. \tag{2·14}$$

This relation might also serve as a large-sample approximation even if the $\varepsilon$'s are not normally distributed. For $0 < t_0 < t_1 < 1$, as $m \to \infty$

$$\hat{\sigma}^{-1} U_m([mt]) \to \sigma^{-1} U(t)$$

weakly on $[t_0, t_1]$. Here $U(t)$ is the zero mean Gaussian process with covariance function

$$\sigma^{-2} \operatorname{cov}\{U(s), U(t)\} = \left\{\frac{s(1-t)}{t(1-s)}\right\}^{\frac{1}{2}} \frac{D(s,t)}{\{D(s,s)D(t,t)\}^{\frac{1}{2}}} \quad (s \leq t), \tag{2·15}$$

where $D(s, t)$ is defined in (2·8). For this limiting process the tail behaviour of

$$\operatorname{pr}\{\max_{t_0 < i/m < t_1} \sigma^{-1}|U(i/m)| \geq b\}$$

is also described by (2·14).

Table 1 shows the accuracy of (2·13) and (2·14) when $x_i = i/m$ $(i = 1, \ldots, m)$, $m_0 = 0{\cdot}1m$, and $m_1 = 0{\cdot}9m$. For each $m$, for the probabilities 0·10, 0·05 and 0·01 the tail percentiles of the distributions of max $\hat{\sigma}^{-1}|U_m(i)|$ and max $\sigma^{-1}|U_m(i)|$ were estimated by a 10 000 repetition Monte Carlo experiment and are recorded in the third and fourth columns of Table 1. The right-hand side of (2·13) evaluated at the estimated percentiles of max $\hat{\sigma}^{-1}|U_m(i)|$ and the right-hand side of (2·14) evaluated at the estimated percentiles of max $\sigma^{-1}|U_m(i)|$ are given in the last two columns. The approximations are quite good, especially for larger values of $m$ and smaller probabilities.

A somewhat simpler large-sample approximation is obtained by considering the maximum of the limiting process $|U(t)|$ over all real $t$ in $[m_0/m, m_1/m]$. However, this

Table 1. *Accuracy of approximations*

| | | Estimated tail percentiles | | Approximations | |
|---|---|---|---|---|---|
| $m$ | Probability | $\max \hat{\sigma}^{-1} U_m(i)$ | $\max \sigma^{-1} U_m(i)$ | (2·13) | (2·14) |
| 10 | 0·10 | 2·46 | 2·45 | 0·098 | 0·108 |
| | 0·05 | 2·62 | 2·72 | 0·042 | 0·052 |
| | 0·01 | 2·84 | 3·27 | 0·007 | 0·009 |
| 20 | 0·10 | 2·66 | 2·62 | 0·097 | 0·107 |
| | 0·05 | 2·84 | 2·87 | 0·048 | 0·054 |
| | 0·01 | 3·19 | 3·39 | 0·009 | 0·010 |
| 40 | 0·10 | 2·76 | 2·75 | 0·107 | 0·108 |
| | 0·05 | 2·98 | 3·01 | 0·052 | 0·051 |
| | 0·01 | 3·43 | 3·51 | 0·009 | 0·010 |
| 70 | 0·10 | 2·83 | 2·83 | 0·110 | 0·108 |
| | 0·05 | 3·07 | 3·10 | 0·053 | 0·050 |
| | 0·01 | 3·52 | 3·58 | 0·010 | 0·010 |

$m_0 = 0{\cdot}1m$, $m_1 = 0{\cdot}9m$, $x_i = i/m$.

approximation usually overestimates the true probability by about forty to one hundred percent. It is also interesting to consider the approximation given by Siegmund (1985, p. 237) for the much simpler problem of testing for a change-point in the mean of a sequence of normally distributed random variables with common, known variance. It was mentioned above that this would be an appropriate large-sample approximation in the present problem if the $x$'s are 'randomly' distributed. However, for equally spaced $x$'s this approximation often substantially underestimates the true probability. For example, for $b = 2{\cdot}62$ and $m = 20$ it gives 0·075; for $b = 2{\cdot}83$ and $m = 70$ it gives 0·061. The discrepancies between these values and the corresponding entries in Table 1 illustrate the effect that the distribution of the $x$'s can have on tail probabilities of the likelihood ratio statistic.

The proof of Theorem 1 is omitted. It is similar in principle but much simpler in detail than the proof of Theorem 2 sketched below.

We consider next the test of $H_0$ against $H_2$, which allows both the slope and the intercept to change. The likelihood ratio test rejects $H_0$ for large values of

$$\hat{\sigma}^{-2} \max_{m_0 \le i \le m_1} \left\{ \frac{mi(\bar{y}_i - \bar{y}_m)^2}{m-i} + \frac{Q_{xyi}^2}{Q_{xxi}} + \frac{Q_{xyi}^{*2}}{Q_{xxi}^*} - \frac{Q_{xym}^2}{Q_{xxm}} \right\}. \tag{2·16}$$

A useful alternative expression for (2·16) is as follows: let

$$A_1' = (1, -1, 0), \quad A_2' = (0, 1, 0, -1),$$

$$X_{1i} = \begin{bmatrix} 1 & 0 & x_1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & x_i \\ 0 & 1 & x_{i+1} \\ \vdots & \vdots & \vdots \\ 0 & 1 & x_m \end{bmatrix}, \quad X_{2i} = \begin{bmatrix} 1 & x_1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_i & 0 & 0 \\ 0 & 0 & 1 & x_{i+1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & x_m \end{bmatrix},$$

and $Y' = (y_1, \ldots, y_m)$. Then (2·16) is of the form

$$\hat{\sigma}^{-2} \max_{m_0 \le i \le m_1} \{U_{1,m}^2(i) + U_{2,m}^2(i)\}, \tag{2·17}$$

where for $\mu = 1$ or 2

$$U_{\mu,m}(i) = A'_\mu(X'_{\mu i}X_{\mu i})^{-1}X'_{\mu i}Y/\{A'_\mu(X'_{\mu i}X_{\mu i})^{-1}A_\mu\}^{\frac{1}{2}}. \tag{2·18}$$

It is easy to see that $U_{1,m}(i)$ for $i = 1, \ldots, m$ is precisely the process $U_m(i)$ defined in (2·2). Let

$$C_{\lambda\mu}(i, k) = \sigma^{-2}\operatorname{cov}\{U_{\lambda,m}(i), U_{\mu,m}(k)\}.$$

Then for $i \leqslant k$, $C_{11}(i, k)$ is given in (2·4),

$$C_{22}(i, k) = \left(\frac{Q_{xxi}Q^*_{xxk}}{Q^*_{xxi}Q_{xxk}}\right)^{\frac{1}{2}} \frac{D_m(i, k)}{\{D_m(i, i)D_m(k, k)\}^{\frac{1}{2}}}, \tag{2·19}$$

$$C_{12}(i, k) = \left\{\frac{miQ^*_{xxk}}{(m-i)Q_{xxm}Q_{xxk}}\right\}^{\frac{1}{2}} \frac{\bar{x}_i - \bar{x}_k}{\{D_m(i, i)D_m(k, k)\}^{\frac{1}{2}}}, \tag{2·20}$$

$$C_{21}(i, k) = \left\{\frac{m(m-k)Q_{xxi}}{kQ_{xxm}Q^*_{xxi}}\right\}^{\frac{1}{2}} \frac{\bar{x}^*_k - \bar{x}^*_i}{\{D_m(i, i)D_m(k, k)\}^{\frac{1}{2}}}. \tag{2·21}$$

Note that, for each $i$, $U_{1,m}(i)$ and $U_{2,m}(i)$ are independent random variables, although the processes

$$\{U_{1,m}(i),\ i = 1, \ldots, m-1\}, \quad \{U_{2,m}(i),\ i = 1, \ldots, m-1\}$$

are by no means independent.

Again there is the simple case that $\bar{x}_m \to \mu$ and $m^{-1}Q_{xxm} \to \tau^2 > 0$ as $m \to \infty$. Under $H_0$, for each $0 < t_0 < t_1 < 1$, $\hat{\sigma}^{-1}\{U_{1,m}([mt]), U_{2,m}([mt])\}$ converges weakly on $[t_0, t_1]$ to $\{t(1-t)\}^{-\frac{1}{2}}\{W_{01}(t), W_{02}(t)\}$, where $W_{0\lambda}(t)$ $(\lambda = 1, 2)$ are independent Brownian bridges; and a minor extension of the one-dimensional results of Siegmund (1985, p. 237) provides an approximation to the tail probability of (2·16).

Now assume $x_i = i/m$. Results under the more general condition $x_i = f(i/m)$ are given in H.-J. Kim's unpublished dissertation; but to avoid some cumbersome expressions, we discuss here only the special case.

THEOREM 2. *Assume $H_0$ is true, so*

$$y_i = \alpha + \beta x_i + \varepsilon_i \quad (i = 1, 2, \ldots, m),$$

*where the $\varepsilon_i$ are independent $N(0, \sigma^2)$. For $0 < t < 1$, $0 \leqslant \theta < 2\pi$ let*

$$u(t, \theta) = \frac{\frac{1}{2} + \{1 - 6t(1-t)\}\sin^2\theta - \surd 3(2t-1)\cos\theta\sin\theta}{t(1-t)D(t, t)}. \tag{2·22}$$

*Assume $b, m \to \infty$ such that for some $0 < c < 1$ and $0 < t_0 < t_1 < 1$,*

$$b/\surd m \to c, \quad m_i \sim mt_i \quad (i = 0, 1).$$

*Then the probability, $p_2$, that the random variable in (2·17) exceeds $b^2$ satisfies*

$$p_2 \sim (2\pi)^{-1}b^2\left(1 - \frac{b^2}{m}\right)^{\frac{1}{2}(m-6)} \int_{t_0}^{t_1}\int_0^{2\pi} \mu(t, \theta)\nu\left[\left\{\frac{2c^2\mu(t, \theta)}{1 - c^2}\right\}^{\frac{1}{2}}\right] d\theta\, dt, \tag{2·23}$$

*where $\nu$ is defined in (2·11) and given approximately in (2·12).*

*Remark.* Again there is a similar result if $\sigma^2$ is known and used in place of $\hat{\sigma}^2$ in (2·16); and this result might be considered a large-sample approximation even if the $\varepsilon$'s are not normally distributed. If $(U_1(t), U_2(t))$ $(0<t<1)$ denotes the Gaussian process which is the weak limit of $(U_{1,m}([mt]), U_{2,m}([mt]))$ as $m\to\infty$, then

$$\mathrm{pr}\left[\max_{t_0<i/m<t_1} \sigma^{-2}\left\{U_1^2\left(\frac{i}{m}\right)+U_2^2\left(\frac{i}{m}\right)\right\}\geqslant b^2\right]$$

$$\sim(2\pi)^{-1}b^2\exp\left(-\tfrac{1}{2}b^2\right)\int_{t_0}^{t_1}\int_0^{2\pi}\mu(t,\theta)\nu[\{2c^2\mu(t,\theta)\}^{\frac{1}{2}}]\,d\theta\,dt. \qquad (2\cdot24)$$

This time, for $0<t_0<t_1<1$, if one takes the maximum over all real $t\in(t_0, t_1)$, the result would be to replace $\nu$ by 1 in (2·24). The double integral could then be evaluated and the right-hand side of (2·24) would simplify to

$$2^{-1}b^2\exp\left(-\tfrac{1}{2}b^2\right)\log\{t_1(1-t_0)/t_0(1-t_1)\}.$$

Remarkably, this is exactly the corresponding approximation for the Brownian Bridge case, although the limiting process is not a two-dimensional Brownian Bridge.

Table 2 gives the results of a 10 000 repetition Monte Carlo experiment to check the accuracy of (2·23). As in Table 1, selected percentiles of the tail of the distribution of (2·16) were estimated by means of the Monte Carlo experiment; and the right-hand side of (2·23) was evaluated at the estimated percentiles. The agreement is very good.

Table 2. *Accuracy of the approximation* (2·23)

| $m$ | Probability | Estimated tail percentile | Approximation (2·23) |
|---|---|---|---|
| 20 | 0·10 | 2·96 | 0·108 |
| | 0·05 | 3·14 | 0·051 |
| | 0·01 | 3·44 | 0·010 |
| 40 | 0·10 | 3·12 | 0·108 |
| | 0·05 | 3·33 | 0·051 |
| | 0·01 | 3·73 | 0·009 |

$m_0=0\cdot1m$, $m_1=0\cdot9m$, $x_i=i/m$.

## 3. Proof of Theorem 2

To discuss the essential ideas of the proof, free insofar as possible from unpleasant computational technicalities, we consider (2·24) instead of (2·23). Let $Z(t,\theta)=U_1(t)\cos\theta+U_2(t)\sin\theta$ and observe that

$$\{U_1^2(t)+U_2^2(t)\}^{\frac{1}{2}}=\sup_{0\leqslant\theta<2\pi}Z(t,\theta).$$

Hence the left-hand side of (2·24) equals

$$\mathrm{pr}\{\sup_{t_0<i/m<t_1}\ \sup_{0\leqslant\theta<2\pi}Z(t,\theta)\geqslant b\}. \qquad (3\cdot1)$$

This probability can be analysed using recently developed methods for approximating the tail of the distribution of the maxima of random fields. We are able to obtain an explicit, computable formula because locally for small $h$ and $\delta$ the conditional random field $Z(t+h,\theta+\delta)-Z(t,\theta)$ given $Z(t,\theta)$ behaves like a sum of two independent one-dimensional processes. The advantage of dealing with (2·24) instead of (2·23) is that this local behaviour of the Gaussian field $Z(t, \theta)$ can be inferred from the behaviour of the covariance function rather than requiring a substantially more elaborate calculation of joint densities.

After taking limits in (2·4), (2·19), (2·20) and (2·21) to evaluate the covariances of the process $(U_1(t), U_2(t))$ and a subsequent messy calculation one can show that

$$E\{Z(t+h, \theta+\delta) \mid Z(t, \theta)=\xi\} = \xi\{1-\tfrac{1}{2}\delta^2 - |h|\mu(t, \theta) + o(\delta^2) + o(h)\} \quad (\delta, h \to 0), \tag{3·2}$$

where $\mu(t, \theta)$ is defined in (2·22) and

$$\begin{aligned} &\operatorname{cov}\{Z(t+h_1, \theta+\delta_1), Z(t+h_2, \theta+\delta_2) \mid Z(t, \theta)=\xi\} \\ &\qquad = 2\mu(t, \theta) \min(|h_1|, |h_2|) + \delta_1\delta_2 + o(\delta_1\delta_2) + o\{\max(|h_1|, |h_2|)\}. \end{aligned} \tag{3·3}$$

In (3·3) the $h_i$ can be positive or negative, but must have the same sign.

The following lemma is an easy consequence of (3·2) and (3·3).

LEMMA 1. *Let $\Delta>0$. Assume $\xi \to \infty$, $b \to \infty$, $m \to \infty$ with $\xi \sim b$ and $b \sim c\surd m$. For any $0<t<1$, $0<\theta<2\pi$, given $Z(t, \theta)=\xi$, the conditional finite dimensional joint distributions of*

$$b\{Z(t+i/m, \theta+k\Delta/\surd m)-\xi\} \quad (i=0, \pm 1, \pm 2, \ldots, k=0, 1, \ldots)$$

*converge to those of*

$$S_i + V_k \quad (i=0, \pm 1, \ldots, k=0, 1, \ldots),$$

*where $S_i$ is a sum of $|i|$ independent $N\{-c^2\mu(t, \theta), 2c^2\mu(t, \theta)\}$ random variables, $V_k = k\Delta c V - \tfrac{1}{2}k^2\Delta^2 c^2$, and $V$ is $N(0, 1)$ and independent of the $S_i$.*

We now sketch a proof of Theorem 2, in which we follow Siegmund's (1988b) extension to random fields of Woodroofe's (1976, 1982) method. The novelty of the present case arises from the fact, as expressed in Lemma 1, that conditionally $Z(t, \theta)$ behaves locally like Brownian motion in the $t$ coordinate and is a smooth Gaussian process in $\theta$. Let

$$J = \{(i_0, k_0)\colon t_0 \leqslant i_0/m \leqslant t_1, 0 \leqslant k_0\Delta/\surd m < 2\pi\},$$

and for each $(i_0, k_0) \in J$ let

$$J(i_0, k_0) = \{(i, k) \in J\colon k > k_0 \text{ or } k = k_0 \text{ and } i > i_0\}.$$

Now

$$\begin{aligned} &\operatorname{pr}\left\{\max_{(i,k)\in J} Z\left(\frac{i}{m}, \frac{k\Delta}{\surd m}\right) \geqslant b\right\} \\ &= \textstyle\sum_{(i_0, k_0)\in J} \operatorname{pr}\left\{Z\left(\frac{i_0}{m}, \frac{k_0\Delta}{\surd m}\right) \geqslant b, \max\limits_{(i,k)\in J(i_0, k_0)} Z\left(\frac{i}{m}, \frac{k\Delta}{\surd m}\right) < b\right\} \\ &= \textstyle\sum_{(i_0, k_0)\in J} \displaystyle\int_0^\infty \operatorname{pr}\left\{Z\left(\frac{i_0}{m}, \frac{k_0\Delta}{\surd m}\right) \in b + \frac{dx}{b}\right\} \\ &\qquad \times \operatorname{pr}\left\{\max_{(i,k)\in J(i_0, k_0)} Z\left(\frac{i}{m}, \frac{k\Delta}{\surd m}\right) < b \,\Big|\, Z\left(\frac{i_0}{m}, \frac{k_0\Delta}{\surd m}\right) = b + \frac{x}{b}\right\}. \end{aligned} \tag{3·4}$$

Since $Z(t, \theta)$ is $N(0, 1)$ for all $(t, \theta)$,

$$\operatorname{pr}\left\{Z\left(\frac{i_0}{m}, \frac{k_0\Delta}{\surd m}\right) \in b + \frac{dx}{b}\right\} \sim \phi(b) \exp(-x) \frac{dx}{b}. \tag{3·5}$$

The conditional probability on the right-hand side of (3·4) equals

$$\mathrm{pr}\left(\max_{(i_0+i,\,k_0+k)\in J(i_0,\,k_0)} b\left[Z\left\{\frac{(i_0+i)}{m},\frac{(k_0+k)\Delta}{\sqrt{m}}\right\}-\left(b+\frac{x}{b}\right)\right]<-x \,\Big|\, Z\left(\frac{i_0}{m},\frac{k_0\Delta}{\sqrt{m}}\right)=b+\frac{x}{b}\right),$$

which by Lemma 1 and the reasoning in the proof of Lemma 4 of Siegmund (1988b), except for an inconsequential number of $(i_0, k_0)$ near the edges of $J$, converges as $b\to\infty$ to

$$\mathrm{pr}\,(\max_{i\geq 1} S_i \leq -x)\ \mathrm{pr}\,(\max_{i\leq 0} S_i+\max_{k\geq 1} V_k\leq -x).$$

Substitution of this limit together with (3·5) into (3·4) yields the preliminary result

$$\mathrm{pr}\left\{\max_{(i,k)\in J} Z\left(\frac{i}{m},\frac{k\Delta}{\sqrt{m}}\right)\geq b\right\}\sim b^{-1}\varphi(b)$$
$$\times\sum_{(i_0,\,k_0)\in J}\int_0^\infty e^{-x}\ \mathrm{pr}\,(\max_{i\geq 1} S_i\leq -x)\ \mathrm{pr}\,(\max_{i\leq 0} S_i+\max_{k\geq 1} V_k\leq -x)\,dx. \quad (3\cdot 6)$$

Additional, technical effort shows that (3·6) holds uniformly in $\Delta$ close to 0, and hence an asymptotic approximation for

$$\mathrm{pr}\left\{\max_{\substack{t_0\leq i/m\leq t_1\\ 0\leq\theta<2\pi}} Z(i/m,\theta)\geq b\right\}$$

is obtained by evaluating the right-hand side of (3·6) as $\Delta\to 0$. Lemma 2 below is concerned with the individual terms in (3·6). Together with an approximation of the resulting sum by a Riemann integral it completes this outline of a proof of Theorem 2.

The statement of Lemma 2 is slightly more general than the preceding argument requires. The notation is locally defined and is not necessarily consistent with that used elsewhere in the paper.

LEMMA 2. *Let $\mu>0$, $\Delta>0$. Assume $X_1, X_2, \ldots$ are independent $N(-\mu,\sigma^2)$ random variables, and let*

$$S_i = X_1+\ldots+X_i \quad (i=1,2,\ldots,S_0=0).$$

*Let $V$ be $N(0,1)$, independent of the $X$'s and let $V_k=\Delta kV-\tfrac{1}{2}\Delta^2k^2$. As $\Delta\to 0$*

$$\Delta^{-1}\int_0^\infty \exp\left(-2\frac{\mu x}{\sigma^2}\right)\mathrm{pr}\,(\max_{i\geq 1} S_i\leq -x)\ \mathrm{pr}\,(\max_{i\geq 0} S_i+\max_{k\geq 1} V_k\leq -x)\,dx$$
$$\to(2\pi)^{-\frac{1}{2}}\left(\frac{2\mu^2}{\sigma^2}\right)\nu\left(\frac{2\mu}{\sigma}\right), \quad (3\cdot 7)$$

*where $\nu$ is defined in* (2·11).

*Proof.* By considering separately the cases $V>0$ and $V<0$ and observing that, when $V<0$,

$$\max_{k\geq 1}(\Delta kV-\tfrac{1}{2}\Delta^2k^2)=\Delta V-\tfrac{1}{2}\Delta^2=V_1,$$

one can easily see that the left-hand side of (3·7) differs by a term of order $\Delta$ from

$$\Delta^{-1}\int_0^\infty \exp\left(-2\frac{\mu x}{\sigma^2}\right)\mathrm{pr}\,(\max_{i\geq 1} S_i\leq -x)\ \mathrm{pr}\,(\max_{i\geq 0} S_i+V_1\leq -x,\ V\leq 0)\,dx. \quad (3\cdot 8)$$

The change of variable $x = \Delta y$ shows that (3·8) equals

$$\int_0^\infty \exp\left(-2\frac{\mu\Delta y}{\sigma^2}\right) \operatorname{pr}(\max_{i\geq 1} S_i \leq -\Delta y) \operatorname{pr}(\max_{i\geq 0} S_i + V_1 \leq -\Delta y,\ V \leq 0)\, dy. \qquad (3\cdot 9)$$

Since

$$\max_{i\geq 0} S_i \geq 0,$$

$$\operatorname{pr}(\max_{i\geq 0} S_i + V_1 \leq -\Delta y,\ V\leq 0) = \int_{-\infty}^{-y+\frac{1}{2}\Delta} \operatorname{pr}\{\max_{i\geq 0} S_i \leq -\Delta(y+v) + \tfrac{1}{2}\Delta^2\}\,\phi(v)\,dv$$

$$\rightarrow \operatorname{pr}(\max_{i\geq 0} S_i \leq 0) \int_{-\infty}^{-y} \phi(v)\,dv$$

as $\Delta \rightarrow 0$. Hence the limit of (3·9) as $\Delta\rightarrow 0$ equals

$$\operatorname{pr}(\max_{i\geq 1} S_i \leq 0) \operatorname{pr}(\max_{i\geq 0} S_i \leq 0) \int_0^\infty \int_{-\infty}^{-y} \phi(v)\,dv\,dy = (2\pi)^{-\frac{1}{2}}\{\operatorname{pr}(\max_{i\geq 0} S_i \leq 0)\}^2,$$

which is well known to equal $(2\pi)^{-\frac{1}{2}}(2\mu^2/\sigma^2)\nu(2\mu/\sigma)$, for example Siegmund (1985, Corol. 8·44). □

*Remark.* For the most part the additional techniques necessary to turn the preceding outline into a rigorous proof are available in the literature already cited. One novel technical feature of the random field $Z(t, \theta)$ is that for an interval of values of $\theta$ there exists a value of $t$ for which $\mu(t, \theta) = 0$. Our argument breaks down in a small neighbourhood of this set, but since it is a set of planar measure zero, it plays no role in the final result.

## 4. Confidence sets

In this section, assuming that $H_1$ holds, we give a confidence region for $j$ and an approximate joint confidence region for $j$, $\delta = \alpha_0 - \alpha_1$ and $\beta$. Our method is an adaptation of that suggested by Siegmund (1988a). See also Davies (1977), Worsley (1986), Siegmund (1986), and James, James & Siegmund (1988). The required probability calculations can be carried out along the lines developed above, although the details are not obvious and will be discussed in a future paper.

To obtain a confidence region for $j$, we invert the likelihood ratio test of the hypothesis that $j$ is the true change-point and include in the confidence region all values of $j$ accepted by an $\alpha$-level test. In order that the significance level not involve unknown nuisance parameters, we evaluate it conditionally given the sufficient statistic: $\bar{y}_m$, $\bar{y}_j$, $\Sigma\, x_i y_i$, $\Sigma\, y_i^2$, where the sums are over $i = 1, \ldots, m$.

The likelihood ratio statistic is equivalent to

$$\Lambda_j = \max_i \frac{U_m^2(i) - U_m^2(j)}{m\hat{\sigma}^2 - U_m^2(j)}. \qquad (4\cdot 1)$$

Hence a confidence region for the change-point is the set of all $j$ such that

$$\operatorname{pr}_j\left\{\Lambda_j \leq (\Lambda_j)_{\text{obs}} \,|\, \bar{y}_m, \bar{y}_j, \sum_{i=1}^m x_i y_j, \sum_{i=1}^m y_i^2\right\} \leq 1 - \alpha. \qquad (4\cdot 2)$$

The notation $\mathrm{pr}_j$ shows explicitly the dependence of this conditional probability on $j$, and by the absence of other subscripts is an implicit reminder that the probability does not depend on $\alpha_0$, $\alpha_1$, $\beta$ or $\sigma$. To evaluate (4·2) approximately one can use

$$\mathrm{pr}_j\left(\Lambda_j > c^2 \,|\, \bar{y}_m, \bar{y}_j, \sum_{i=1}^m x_i y_i, \sum_{i=1}^m y_i^2\right) \sim 2(1-c^2)^{(m-5)/2}\,\nu(\hat{\delta}_j/[(1-c^2)\{\hat{\sigma}^2 - m^{-1}U_m^2(j)\}]^{\frac{1}{2}}), \tag{4·3}$$

where

$$\hat{\delta}_j = U_m(j)/\{j(1-j/m)D_m(j,j)\}^{\frac{1}{2}} \tag{4·4}$$

and the asymptotic relation holds as $m\to\infty$, $c\to 0$ and $mc^2\to\infty$. Compare Siegmund (1988a, eqn (18)).

Now suppose we are interested in a joint confidence region for $j$, $\delta = \alpha_0 - \alpha_1$ and $\beta$. It is useful to introduce the notation

$$U_m^2(j,\delta,\beta) = (Q_{xxj} + Q_{xxj}^*)\left(\frac{Q_{xyj}+Q_{xyj}^*}{Q_{xxj}+Q_{xxj}^*} - \beta\right)^2 + j\left(1-\frac{j}{m}\right)\{\bar{y}_j - \bar{y}_j^* - \delta - \beta(\bar{x}_j - \bar{x}_j^*)\}^2, \tag{4·5}$$

$$\Lambda_j(\delta,\beta) = U_m^2(j,\delta,\beta)/\{m\hat{\sigma}^2 - U_m^2(j) + U_m^2(j,\delta,\beta)\}. \tag{4·6}$$

If the change-point $j$ were known, $\Lambda_j(\delta,\beta)$ would be a version of the likelihood ratio statistic for testing the specific values $(\delta,\beta)$ against a general alternative. When $j$, $\delta$ and $\beta$ are the true parameter values, the distribution of $\Lambda_j(\delta,\beta)$ is $B\{1, \frac{1}{2}(m-3)\}$.

The log likelihood ratio statistic for the specific value $(j,\delta,\beta)$ against a general alternative is proportional to

$$-\log(1-\Lambda_j) - \log\{1-\Lambda_j(\delta,\beta)\},$$

where $\Lambda_j$ is defined in (4·1) and $\Lambda_j(\delta,\beta)$ in (4·6). Large values of the first term provide evidence against $j$; given a specific $j$, large values of the second term provide evidence against $(\delta,\beta)$. The set of values $(j,\delta,\beta)$ not rejected by an $\alpha$ level test give a $(1-\alpha)$ 100% confidence region for $(j,\delta,\beta)$. The appropriate significance level is given by

$$\begin{aligned}\mathrm{pr}\,\{\Lambda_j + \Lambda_j(\delta,\beta) - \Lambda_j\Lambda_j(\delta,\beta) \geq c^2\} &= \mathrm{pr}\,\{\Lambda_j(\delta,\beta)\geq c^2\} \\ &\quad + \mathrm{pr}\left\{\Lambda_j(\delta,\beta) < c^2,\ \Lambda_j \geq \frac{c^2-\Lambda_j(\delta,\beta)}{1-\Lambda_j(\delta,\beta)}\right\}.\end{aligned} \tag{4·7}$$

The first term on the right-hand side of (4·7) is just

$$(1-c^2)^{\frac{1}{2}(m-3)}. \tag{4·8}$$

The second term can be evaluated by taking the expectation of the conditional probability, given $\bar{y}_j, \bar{y}_m, \Sigma x_i y_i, \Sigma x_i^2$, where the sums are over $i=1,\ldots,m$. For the conditional probability we use the approximation (4·3) with $c^2$ replaced by

$$\{c^2-\Lambda_j(\delta,\beta)\}/\{1-\Lambda_j(\delta,\beta)\}$$

and find after some calculation that the second term is given approximately by

$$(m-3)c^2(1-c^2)^{\frac{1}{2}(m-5)}\nu[\delta/\{\sigma(1-c^2)^{\frac{1}{2}}\}]. \tag{4·9}$$

Substitution of (4·8) and (4·9) into (4·7) yields an approximation to the desired probability.

Since this approximation involves an unknown parameter, it cannot be applied directly. A conservative modification would be to replace $\nu$ in (4·9) by the upper bound of 1. However, this modification would often result in approximately doubling (4·9). A less conservative modification would be to evaluate $\nu$ at a conservatively estimated value, perhaps $\nu(\hat{\delta}/\hat{\sigma})$ or $\nu(\hat{\delta}_j/\hat{\sigma})$, where $\hat{\delta}_j$ is given in (4·4).

## 5. Discussion and open problems

The purpose of this paper has not been to present a polished solution to any particular problem, but rather to indicate that the use of the likelihood ratio statistic for change-point problems in regression models does not present the apparently intractable analytic difficulties which have hindered its application in the past. There remain a large number of open problems, some of which we discuss briefly here.

Many interesting change-point problems, particularly in epidemiology, do not involve changes in the mean of normally distributed data, but rather changes in the rate of occurrence of rare events. In cases where there are no covariates, one can obtain results for the likelihood ratio statistic itself or develop normal approximations (Worsley, 1983b, 1986; Siegmund, 1988a). The situation when there are covariates is more complicated and should be studied.

For normal linear regression, when there are several independent variables, some of which may be random while others are controlled or are a function of time, the situation appears similar in principle to what we have studied here, but the details can be substantially more complex.

The recursive residuals test of Brown et al. (1975) has been widely used as a diagnostic for testing the constancy of regression relations over time. Its principal virtue is the relative simplicity of its sampling distribution under the hypothesis of no change, compared to the likelihood ratio statistic or other suitably normalized maxima of (2·3). However, it is designed without reference to any particular class of alternatives, and its power to detect specific changes has rarely been studied. James et al. (1987) modify the Brown et al. test and show for the simple problem of detecting at most one change in the mean of a normal population with known variance that their modification competes favourably with the likelihood ratio test. However, this situation seems over-simplified, and some investigation is desirable to discover if similar results hold more generally. In her unpublished Ph.D thesis, H.-J. Kim has made a start in this direction by obtaining an approximation for the power of the likelihood ratio test of $H_0$ against $H_1$.

The joint confidence region for $j$, $\alpha_0-\alpha_1$ and $\beta$ discussed in § 4 may provide some insight into a problem posed by Cox (1961): to choose between a simple change-point model without covariates and a regression model with no change-point. One can imbed both of these models in the larger model described by $H_1$, and then decide whether one of the simpler models would suffice by examining the extent to which a confidence region for $j$, $\alpha_0-\alpha_1$ and $\beta$ includes zero as a possible value for $\alpha_0-\alpha_1$ or $\beta$. Of course, this procedure may not supply a clear-cut answer, and in some cases a clear-cut answer may not be possible. The procedure probably is satisfactory if the $x$'s are random, because then there is a relatively small loss of accuracy in estimating $\alpha_0-\alpha_1$ caused by the requirement that one must also estimate $\beta$. However, in cases where the empirical distributions of the $x_i$ for $i \leq j$ and $i > j$ have disjoint supports, this loss of accuracy can be quite considerable. Hence even if there should be a clear answer to the question of the choice of model, the joint confidence region may lack the power to provide it.

For the model described by $H_1$ it should be possible to obtain a confidence region for $j$, $(\alpha_0-\alpha_1)/\sigma$ and $\beta/\sigma$, which unlike the joint region in § 4 is in principle exact. In fact, in many respects it seems more appropriate to measure the magnitude of a change in a normal mean in units of the standard deviation than in the more conventional units of the observations themselves.

## ACKNOWLEDGEMENT

This research was performed with partial support of the National Science Foundation.

## REFERENCES

ALDOUS, D. (1989). *Probability Approximations via the Poisson Clumping Heuristic*. New York: Springer-Verlag.

BICKEL, P. & ROSENBLATT, M. (1973). Two-dimensional random fields. In *Multivariate Analysis—III*, Ed. P. K. Krishnaiah, pp. 3-15. New York: Academic Press.

BROWN, R. L., DURBIN, J. & EVANS, J. M. (1975). Techniques for testing the constancy of regression relationships over time (with discussion). *J. R. Statist. Soc.* B **37**, 149-92.

COX, D. R. (1961). Tests of separate families of hypotheses. *Proc. 4th Berkeley Symp.* **1**, 105-23.

COX, D. R. & LEWIS, P. A. W. (1966). *The Statistical Analysis of a Series of Events*. London: Methuen.

DAVIES, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64**, 247-54.

DAVIES, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74**, 33-43.

HINKLEY, D. V. (1969). Inference about the intersection in two-phase regression. *Biometrika* **56**, 495-504.

HINKLEY, D. V. (1971). Inference in two-phase regression. *J. Am. Statist. Assoc.* **66**, 736-43.

HOGAN, M. L. & SIEGMUND, D. (1986). Large deviations for the maxima of some random fields. *Adv. Appl. Math.* **7**, 2-22.

JAMES, B., JAMES, K. L. & SIEGMUND, D. (1987). Tests for a change-point. *Biometrika* **74**, 71-84.

JAMES, B., JAMES, K. L. & SIEGMUND, D. (1988). Conditional boundary crossing probabilities with applications to change-point problems. *Ann. Prob.* **16**, 825-39.

JARRETT, R. G. (1979). A note on the intervals between coal mining disasters. *Biometrika* **66**, 191-3.

JOHNSTONE, I. & SIEGMUND, D. (1989). On Hotelling's formula for the volume of tubes and Naiman's inequality. *Ann. Statist.* **16**, 184-94.

LEVIN, B. & KLINE, J. (1985). The cusum test of homogeneity with an application to spontaneous abortion epidemiology. *Statist. Med.* **4**, 469-88.

MARONNA, R. & YOHAI, V. J. (1978). A bivariate test for the detection of a systematic change in mean. *J. Am. Statist. Assoc.* **73**, 640-5.

PICKANDS, J. (1979). Upcrossing probabilities for stationary Gaussian processes. *Trans. Am. Math. Soc.* **145**, 51-73.

QUALLS, C. & WATANABE, H. (1973). Asymptotic properties of Gaussian random fields. *Trans. Am. Math. Soc.* **177**, 155-71.

QUANDT, R. E. (1958). The estimation of the parameter of a linear regression system obeying two separate regimes. *J. Am. Statist. Assoc.* **53**, 873-80.

QUANDT, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *J. Am. Statist. Assoc.* **55**, 324-30.

RAFERTY, A. E. & AKMAN, V. E. (1986). Bayesian analysis of a Poisson process with a change-point. *Biometrika* **11**, 701-19.

SIEGMUND, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. New York: Springer-Verlag.

SIEGMUND, D. (1986). Boundary crossing probabilities and statistical applications. *Ann. Statist.* **14**, 361-404.

SIEGMUND, D. (1988a). Confidence sets in change-point problems. *Int. Statist. Rev.* **56**, 31-48.

SIEGMUND, D. (1988b). Approximate tail probabilities for the maxima of some random fields. *Ann. Prob.* **16**, 487-501.

WOODROOFE, M. (1976). Frequentist properties of Bayesian sequential tests. *Biometrika* **63**, 101-10.

WOODROOFE, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*. Philadelphia: SIAM.

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**, 455-64.
WORSLEY, K. J. (1986). Confidence regions and tests for a change-point in a sequence of exponential family variables. *Biometrika* **73**, 91-104.

[*Received November* 1988. *Revised March* 1989]