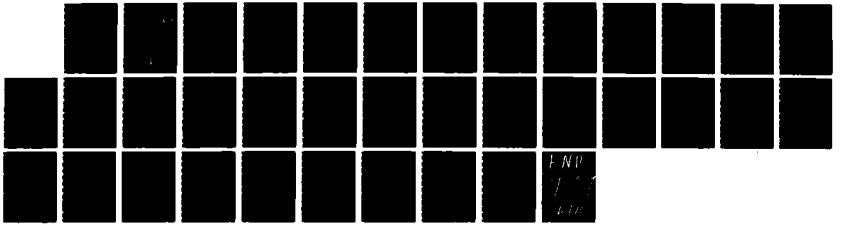
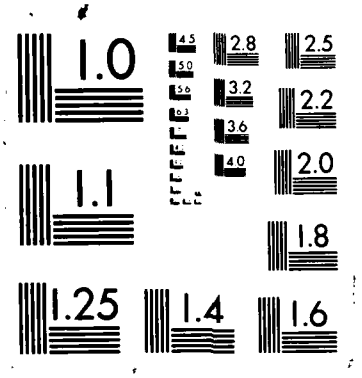
AD-A181 075 CONFIDENCE SETS IN CHANGE-POINT PROBLEMS REVISION(U) 1/1
STANFORD UNIV CA DEPT OF STATISTICS D SIEGMUND MAY 87
TR-2 N00014-87-K-0078

UNCLASSIFIED F/G 12/3 NL




1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6

CONFIDENCE SETS IN CHANGE-POINT PROBLEMS

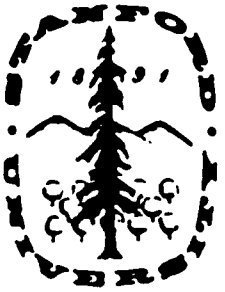
by

David Siegmund
Stanford University

TECHNICAL REPORT NO. 2
MAY 1987

PREPARED UNDER CONTRACT
N00014-87-K-0078 (NR-042-373)
FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA



CONFIDENCE SETS IN CHANGE-POINT PROBLEMS

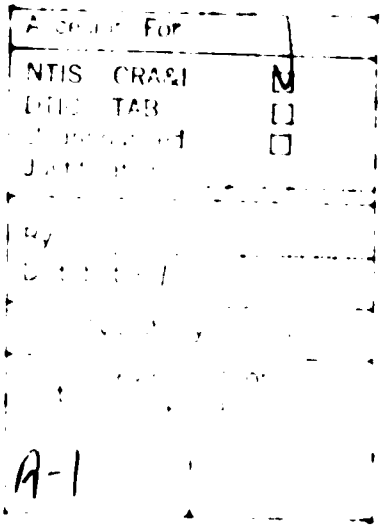
by

David Siegmund
Stanford University

TECHNICAL REPORT NO. 2
MAY 1987

PREPARED UNDER CONTRACT
N00014-87-K-0078 (NR-042-373)
FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA



This is a substantial revision of Technical Report No. 39 of ONR Contract N00014-77-0306, entitled "Confidence Sets for a Change–point."

CONFIDENCE SETS IN CHANGE-POINT PROBLEMS

David Siegmund

Department of Statistics, Sequoia Hall, Stanford University, Stanford, California 94305, USA

Summary.

Several methods are discussed for confidence set estimation of a change–point in a sequence of independent observations from completely specified distributions. The method based on the likelihood ratio statistic is extended to the case of independent observations from an exponential family. Joint confidence sets for the change–point and the parameters of the exponential family are also considered.

Key words: change–point, likelihood ratio, boundary crossing probabilities

1. Introduction.

Let $x_1, x_2, \ldots, x_m$ be independent random variables with $x_1, \ldots, x_j$ having distribution F and $x_{j+1}, \ldots, x_m$ having distribution $G \neq F$. The change–point j, where the distribution shifts from F to G, is an unknown parameter, to be estimated by a confidence set. In general, the distributions F and G may be known, completely unknown, or specified up to an unknown parameter. In this paper I discuss several procedures for the artificial but informative case of completely specified F and G, and then develop more completely a method based on the likelihood ratio statistic for the case where F and G come from a common exponential family of distributions. Precedent for the approach taken here is found in Worsley (1986) and Siegmund (1986).

A distinguishing feature of change–point problems is that the likelihood function is not smooth, even (or perhaps especially) if the process evolves in continuous times. Hence there is no reason to expect maximum likelihood, likelihood ratio, and Bayes estimates from different prior distributions to lead to asymptotically equivalent results. In fact, confidence sets based directly on the maximum likelihood estimator are demonstrably inferior to those obtained by

other methods. See Siegmund (1986) and Ibragimov and Khasminski (1981) for related results in the context of detecting a change in the drift of Brownian motion.

Section 2 is concerned with known F and G. In addition it is assumed that the sequence of observations is actually doubly infinite,$\ldots x_{-1}, x_0, x_1, \ldots$. This additional assumption has little effect if m is large and it is known that j is not close to 1 nor to m, because observations far from the change–point carry little information about the location of the change–point. The virtue of the assumption is that it makes j into a location parameter and provides an exact ancillary statistic: the class of shift invariant events. Five confidence set estimates are discussed. Three are studied by Siegmund (1986), in the context of estimating a change–point in the drift of Brownian motion. The fourth is essentially the suggestion of Cobb (1978), and the fifth has smallest expected size among all shift invariant confidence sets. Section 3 compares the different confidence sets.

Sections 4 and 5 are concerned with the case that F and G are imbedded in a common exponential family, whose parameter θ is unknown. Section 4 develops a method based on the likelihood ratio statistic for obtaining exact confidence sets for j. A new, fairly simple approximation is suggested for the required probability calculation. The approximation is illustrated on the coal mining accident data along the lines discussed by Worsley (1986). In Section 5 the likelihood ratio method is extended to give a joint confidence set for j and a function of the parameters of the exponential family. Technical results are given in two appendices.

2. The Cases of Known F and G.

Let $\mathbb{Z}$ denote the integers and let $j \in \mathbb{Z}$. Let x_n, $n \in \mathbb{Z}$ be a sequence of independent random variables with x_n having the distribution function F or G according as $n \leq j$ or $n > j$. The distributions F and G are assumed known; the change–point j is unknown. Let P_j denote the probability measure induced by this model on the space of infinite sequences $\omega = (x_n, n \in \mathbb{Z})$. Let σ denote the shift operator, i.e., the mapping which takes $\omega = (x_n, n \in \mathbb{Z})$ into $\sigma\omega = (x_{n+1}, n \in \mathbb{Z})$. Note that the family $\{P_j, j \in \mathbb{Z}\}$ is a translation family in the sense that for any event B and $j \in \mathbb{Z}$

$$P_j(B) = P_j(\omega \epsilon B) = P_0(\sigma^{-j}\omega \epsilon B) = P_0(\sigma^j B).$$

Let $z_n = \log\{dG(x_n)/dF(x_n)\}$ denote the log likelihood ratio of x_n, and put

$$\begin{aligned} \tilde{S}_n &= z_1 + \ldots + z_n \qquad (n \geq 1) \\ &= -(z_{n+1} + \ldots + z_0) \quad (n \leq -1) \\ &= 0 \qquad (n = 0) \end{aligned}$$

Let $\ell_i = dP_i/dP_0$ denote the likelihood function at i. By considering the finite sequence x_n, $-N \leq n \leq N$, and then letting $N \to \infty$, one can easily show that $\ell_i = \exp(\tilde{S}_i)$. Under P_0 the log likelihood process $(\tilde{S}_n, n \epsilon \mathbb{Z})$ is a random walk satisfying $\tilde{S}_0 = 0$ and having increments $\tilde{S}_n - \tilde{S}_{n-1}$ with mean $\int \log(dG/dF)dF < 0$ for $n > 0$ and $\int \log(dF/dG)dF > 0$ for $n \leq 0$.

The maximum likelihood estimator for j is the value $\hat{j}$ where the process $(\tilde{S}_n, n \epsilon \mathbb{Z})$ assumes its maximum value. In general this value need not be unique, but to avoid technicalities it is assumed to be so in what follows. In the space of the sufficient statistic $(\tilde{S}_n, n \epsilon \mathbb{Z})$, the sequence $Y_i = \tilde{S}_{\hat{j}+i} - \tilde{S}_{\hat{j}}$, $i \epsilon \mathbb{Z}$, is ancillary.

In the context of estimating a change–point in the drift of a Brownian motion process, Siegmund (1986) compares the following three confidence sets for the change–point j. The first two were discussed earlier by Hinkley (1970, 1972), who, however, made no attempt to establish their relative efficiency.

(i) Since $\hat{j} - j$ is pivotal, if $r = r_\alpha$ is defined by $P_0(|\hat{j}| > r) = \alpha$, then $C_1 = [\hat{j} - r, j + r]$ is a $(1 - \alpha)$ 100% confidence interval.

(ii) Let A_j devote the acceptance region of a size α likelihood ratio test of the hypothesis that the change–point is j, i.e., $A_j = \{\max_n \tilde{S}_n - \tilde{S}_j < \eta\}$, where $\eta = \eta_\alpha$ satisfies $P_j(A_j) = \{P_0(\max_{n \geq 0} \tilde{S}_n < \eta)\}^2 = 1 - \alpha$. Then the set C_2 of $n \epsilon \mathbb{Z}$ such that the observed sample point $\omega \epsilon A_n$ is a $(1 - \alpha)$ 100% confidence set. Since the log likelihood process $(\tilde{S}_n, n \epsilon \mathbb{Z})$ is in general multimodal, this confidence set is not in general an interval.

(iii) A modification of the preceding method which always yields an interval is to define

$$L(R) = \min(\max)\left\{n : \tilde{S}_n \geq \max_i \tilde{S}_i - \eta'\right\},$$

which for suitable $\eta' < \eta$ satisfies

$$P_j(L \leq j \leq R) = P_0(L \leq 0 \leq R) = 1 - 2P_0(R < 0) = 1 - \alpha.$$

The next possibility is essentially the suggestion of Cobb (1978). In analogy with Fisher's (1934) observation that the conditional probability density of the maximum likelihood estimator of a location parameter given the sample spacings, which are ancillary in that case, is the normalized likelihood function, one may show by a direct calculation that

$$P_j(\hat{j} - j = n | Y_i, i \in \mathbb{Z}) = P_0(\hat{j} = n | Y_i, i \in \mathbb{Z}) = \exp\left(\tilde{S}_{\hat{j}_{obs}-n}\right) / \sum_i \exp(\tilde{S}_i), \tag{1}$$

where $\hat{j}_{obs}$ denotes the observed value of $\hat{j}$. Let

$$p_n = \exp(\tilde{S}_n) / \sum \exp(\tilde{S}_i), \ n \in \mathbb{Z}. \tag{2}$$

(iv) It follows from (1) that a confidence set of conditional coverage probability $1 - \alpha$ can be formed as follows. Order the p_n in (2) as $p_{(1)} \geq p_{(2)} \geq \ldots$. Construct the set C_4 by putting the index n_1 corresponding to $p_{(1)}$ in C_4 and continuing to add points $n_2, \ldots, n_k$ corresponding to $p_{(2)}, \ldots, p_{(k)}$ as long as $\sum_{i \leq k} p_{(i)} < 1 - \alpha$. Note that for a Bayesian with a uniform prior on $\mathbb{Z}$,

$$p_n = P(j = n | x_i, i \in \mathbb{Z})$$

and hence the set C_4 is a highest posterior probability credible set for j. In fact, even without the explicit evaluation in (1), one knows from a general theorem of Stein (1965) and Hora and Buehler (1966) that the highhest posterior credible set for j is also a confidence set.

(v) One can also obtain an unconditional confidence set from the formal posterior probabilities $(p_n, n \in \mathbb{Z})$ in (2) as follows: let c be such that

$$P_j\{p_j \geq c\} = P_0\left\{\sum \exp(\tilde{S}_n) \leq c^{-1}\right\} = 1 - \alpha, \tag{3}$$

and $C_5 = \{n : p_n \geq c\}$. Then C_5 is a $(1-\alpha)$ 100% confidence set, which according to a general theorem of Hooper (1982) or alternatively by a simple Neyman–Pearson argument has smallest expected size among all shift equivariant confidence sets.

Remarks. The confidence sets (ii), (iv), and (v) all order the parameter values for inclusion according to the value of the likelihood function. Where they disagree is where to draw the line between inclusion and exclusion. For those who strongly prefer a confidence interval to a possibly disconnected confidence set, (iii) appears to be a reasonable modification of (ii). It is possible to give analogous modifications of (iv) and (v).

Of these five confidence sets, all except for (iv) require computation of a sampling distribution. Approximations are suggested in the following section.

3. Comparisons.

The purpose of this section is to compare the expected size of the various confidence sets proposed in Section 2. Since the case of known G and F is artificially simple and our main goal is insight into the case where G and F contain unknown nuisance parameters, there seems to be little harm in simplifying the technical problems somewhat by assuming that F is $N(0,1)$ and G is $N(\delta,1)$ for a known $\delta > 0$.

Siegmund (1986) considers the computationally simpler case of a Brownian motion process and shows that the length of the confidence interval defined in (i) is substantially longer than the expected size of the confidence sets in (ii) and (iii).

In the present context it can be shown as $\alpha \to 0$ that the expected sizes of the confidence sets in (ii) - (v) are all $\sim 4\delta^{-2}\log\alpha^{-1}$, whereas the length of the interval in (i) is $\sim 8\delta^{-2}\log\alpha^{-1}$. Hence the confidence interval C_1 defined in (i) appears not to be competitive with the others and will not be considered further.

Although Siegmund's (1986) comparison of (ii) and (iii) favors (ii), the difference is not large. In fact there is a transcription error in passing from the first to the second line of the display following (3.15) of Siegmund (1986), and consequently the difference in the numerical example between methods (ii) and (iii) is smaller than stated there. Since one suspects that the rapid fluctuations of Brownian motion may account for some of that difference, and since (iii) is the only remaining interval estimate and is a surrogate for interval modifications of (iv) and (v), it seems reasonable to make a comparison of (ii) and (iii) in the present discrete time setting. Theorem 1 below gives asymptotic expansions as $\alpha \to 0$ of the expected size of the confidence sets (ii) and (iii).

It seems difficult to give comparably precise expansions for (iv) and (v). Hence (ii), (iv), and (v) are compared below in a Monte Carlo experiment, which also shows that the approximations given in Theorem 1 are reasonably accurate.

We begin with approximations for the coverage probability of (ii) and (iii). Let Φ be the *standard normal distribution function* and

$$\nu(x) = 2x^{-2} \exp\left\{-2\sum_{1}^{\infty} n^{-1}\Phi\left(-x\sqrt{n}/2\right)\right\} \qquad (x > 0). \tag{4}$$

For computational purposes it usually suffices to use the small x approximation (Siegmund, 1985, p. 219)

$$\nu(x) = \exp(-\rho x) + o(x^2) \qquad (x \to 0), \tag{5}$$

where $\rho \cong .583$. For the normally distributed $x_n, n\epsilon\, \mathbb{Z}$, under consideration here $\tilde{S}_n = \delta(n\delta/2 - S_n), n = 0, 1, \ldots$, where $S_n = x_1 + \ldots + x_n$. It follows from a classical result of Cramér (cf. Siegmund, 1985, (8.49)) that

$$P_0\left(\max_{n\geq 0} \tilde{S}_n \geq \eta\right) \sim \nu(\delta)\exp(-\eta) \qquad (\eta \to \infty) \tag{6}$$

and hence by (5) for A_j defined in (ii) above

$$P_j(A_j) \cong \{1 - \exp(-\eta - \rho\delta)\}^2. \tag{7}$$

By conditioning on $\max_{n\geq 0} \tilde{S}_n$, one may show for R defined in (iii),

$$P_0(R<0) = P_0\left(\max_{n\leq 0}\tilde{S}_n > \max_{n\geq 0}\tilde{S}_n + \eta'\right) \tag{8}$$

$$\sim \nu(\delta)\exp(-\eta')E_0\left\{\exp\left(-\max_{n\geq 0}\tilde{S}_n\right)\right\}$$

$\eta' \to \infty$. It is possible to compute the expectation on the right hand side of (8) numerically or give a small δ expansion analogous to (5), but for our purposes it seems adequate to pretend that (6) is an equality, which after an integration by parts in (8) leads to the approximation

$$P_j(0 \notin [L,R]) \cong 2\exp(-\eta' - \rho\delta)\{1-\exp(-\rho\delta)/2\}. \tag{9}$$

The following theorem gives an asymptotic expansion as $\alpha \to 0$ of the expected size of C_2 defined in (ii) and $[L,R]$ defined in (iii). It will be convenient to use the notation $\lfloor y \rfloor$ = integer part of y, $|C|$ = number of elements in the set C, and $M = \sup_{n\geq 0}\tilde{S}_n$.

Theorem 1. Let C_2 be the confidence set defined in (ii) and $[L,R]$ the confidence interval defined in (iii). As $\eta \to \infty$

$$E_j|C_2| = 2\lfloor 2\eta/\delta^2 \rfloor + 4/\delta^2$$
$$- 4\delta^{-1}\int_0^\infty \{2P_0(M>x) - P_0^2(M>x)\}dx + o(1),$$

and as $\eta' \to \infty$

$$E_j(R-L) = 2\lfloor 2\eta'/\delta^2 \rfloor + 4/\delta^2$$
$$- 4\delta^{-1}\int_0^\infty \int_0^\infty P_0(M \epsilon dy)\{2P_0(M>x+y) - P_0^2(M>x+y)\}dx + o(1).$$

A proof is sketched in an appendix.

To obtain easily evaluated approximations to the integrals appearing in these expressions, one may again pretend that (6) is an equality and use (5). This leads to

$$E_j|C_2| \cong 2\lfloor 2\eta/\delta^2 \rfloor + 2\delta^{-2}\left(2 - 4e^{-\rho\delta} + e^{-2\rho\delta}\right) \tag{10}$$

and

$$E_j(R-L) \cong 2\lfloor 2\eta'/\delta^2 \rfloor + 2\delta^{-2}\left(2 - 4e^{-\rho\delta} + 3e^{-2\rho\delta} - 2e^{-3\rho\delta}/3\right). \tag{11}$$

Table 1 contains some numerical examples. It indicates that there is essentially no difference between the expected size of the confidence sets (ii) and (iii). On the basis of these results a statistician who strongly prefers a confidence interval to the generally disconnected likelihood ratio confidence set should feel comfortable in imposing that constraint.

Table 1.

Expected Size of Confidence Sets (ii) and (iii)

α	δ	η (7)	$E_0\|C_2\|$ (10)	η' (9)	$E_0(R-L)$ (11)
.1	0.7	2.56	19.1	2.18	17.9
.1	1.0	2.39	8.2	2.08	9.2
.05	0.7	3.27	25.1	2.88	23.9
.05	1.0	3.09	12.2	2.78	11.2
.01	0.7	4.89	37.1	4.49	37.9
.01	1.0	4.71	18.2	4.39	17.2

In the present context of completely specified distributions there is no sampling theory to develop in order to use the confidence set (iv). However, it seems a difficult problem to give a reasonable approximation for the related set defined in (v). A crude approximation to (3) which might be used as the first step in an iterative numerical or Monte Carlo scheme is to replace $\tilde{S}_n$ by a Brownian motion process $\tilde{W}(t)$ with drift $-(\delta^2/2)sgn(t)$ and variance δ^2 and replace the sum in (3) by an integral. One easily sees that the integral over $[0,\infty)$ has the

distribution given by Pollak and Siegmund (1985, Proposition 3). This can be convolved with itself to obtain $\text{pr}[\int_{-\infty}^{\infty} \exp\{\tilde{W}(t)\}dt < c^{-1}] = 2\delta^{-1}\sqrt{c}\ \exp(-4c/\delta^2)K_1(2\delta^{-1}\sqrt{c})$, where K_1 is the modified Bessel function of the second kind.

Table 2 reports the results of 1000 repetition Monte Carlo experiment with $m = 100$ and $j = 50$ to compare the confidence sets C_2, C_4, and C_5. It confirms that the analytic approximation for the expected size of C_2 given in Theorem 1 is reasonably accurate and shows that all three confidence sets have about the same expected size.

Table 2.

Monte Carlo Comparison of C_2, C_4, and C_5

	C_2			C_4		C_5		
α (nominal)	δ	$\hat{\alpha}$	$E_0\lvert C_2\rvert$	$\hat{\alpha}$	$E_0\lvert C_4\rvert$	c	$\hat{\alpha}$	$E_0\lvert C_5\rvert$
.10	.07	.090	18.8	.084	19.5	.010	.092	19.3
.10	1.0	.098	9.6	.085	10.3	.022	.113	9.4
.05	0.7	.041	24.6	.040	25.2	.005	.047	26.0
.05	1.0	.048	12.6	.037	13.2	.011	.052	12.6

Although the confidence sets defined in (ii)-(iv) perform similarly on the average, they can treat individual sets of data differently. Figure 1 displays two simulated log likelihoods with $m = 101, j = 50$, and $\delta = 0.7$. The horizontal line defines the 95% likelihood ratio confidence set (ii). In accordance with the approximation (7) it is drawn 3.27 units below the maximum of the log likelihood function.

In the upper part of Figure 1 the one major peak of the log likelihood is fairly sharp with the consequence that all the confidence sets are about one half their expected size of 25. The confidence interval defined in (iii) has one point less on each end than the likelihood ratio confidence set. The formal Bayes posterior set, C_4, makes a smaller adaptation to the peaked log likelihood; it contains four more points, including the local maximum at 63. The confidence set C_5 is the same as the likelihood ratio confidence set.

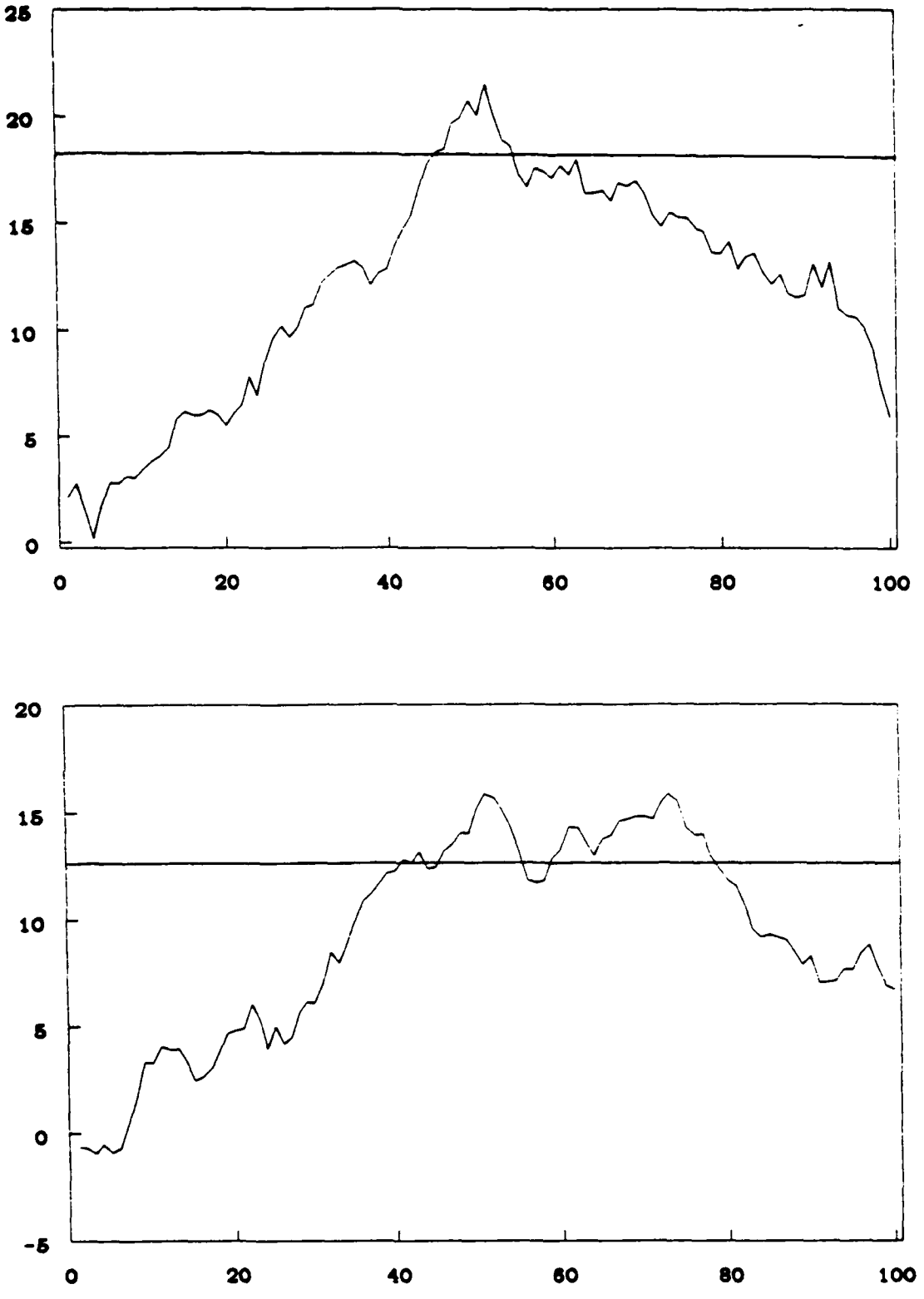


Figure 1. Two Simulated Log Likelihoods.

The lower part of Figure 1 contains a comparatively flat log likelihood with two distinct peaks. The likelihood ratio confidence set contains 33 points. The interval modification is now slightly larger because it contains points of relatively low likelihood: 44, 45, 56-58. Again the formal Bayes posterior set adapts less to the departure of the log likelihood from its expected shape and this time contains four fewer points than the likelihood ratio confidence set.

In general, the interval modification (iii) is usually slightly shorter than the likelihood ratio confidence set but can be considerably larger. The formal Bayes posterior set is usually larger than the likelihood ratio set when both sets are small and smaller when both sets are large. This suggests that there may be recognizable subsets making the conditional coverage probability of the likelihood ratio set differ from its nominal value. The confidence set C_5 can look rather foolish conditionally. If all the p_i are very small and about equal, it can deliver a small, or perhaps empty confidence set while the other methods recognize the data as uninformative and yield large confidence sets. Presumably this occurs with small probability.

Overall the evidence given here does not seem persuasive for choosing among the confidence sets (ii) - (v). A possible conclusion is that in more complex problems one may reasonably use whichever method seems most easily adpated to the problem at hand. When the distributions F and G are unknown, but can be imbedded in a common exponential family, one can use a conditioning argument to obtain exact likelihood ratio confidence sets. This is the subject of the next section.

4. The Likelihood Ratio Method for an Exponential Family.

Now suppose that F and G can be imbedded in an exponential family of the form

$$dF_\theta(x) = \exp\{\theta x - \psi(\theta)\}dF_0(x)$$

relative to some fixed distribution F_0, which without loss of generality can be standardized to have mean 0 and variance 1. Thus for some unknown $\theta_0 \neq \theta_1$ and $j \epsilon \{1, \ldots, m\}$, $x_1, \ldots, x_j$ have distribution F_{θ_0} and $x_{j+1}, \ldots, x_m$ have distribution F_{θ_1}. The probability on the space of $x_1, \ldots, x_m$ will be denoted by P, with the dependence on j, θ_0, and θ_1 suppressed. For

the most part we consider a scalar parameter θ, but with some technical complications the methods described below are generally valid.

Several writers, e.g., Davies (1977), Siegmund (1986), and Worsley (1986), have observed that one can extend the likelihood ratio method (ii) of Section 2 to obtain a confidence set for j in the presence of the unknown nuisance parameters θ_0, θ_1 as follows. Let $H(x) = \sup_\theta\{\theta x - \psi(\theta)\}$, $S_n = x_1 + \ldots + x_n$, and

$$\Lambda_n = nH(n^{-1}S_n) + (m-n)H\{(m-n)^{-1}(S_m - S_n)\}. \tag{12}$$

The likelihood ratio test of the hypothesis that the change–point is j has acceptance region of the form

$$A_j = \left(\max_n \Lambda_n - \Lambda_j \leq k\right).$$

By sufficiency the conditional probability of A_j given (S_j, S_m) does not depend on θ_0, θ_1. Hence if one chooses $k = k(j, \xi_1, \xi_2)$ so that

$$P(A_j | S_j = \xi_1, S_m = \xi_2) = 1 - \alpha$$

for all j, ξ_1, ζ_2, then the set of values j which are accepted by the test is a $(1-\alpha)100\%$ confidence set.

It is not actually necessary to solve for $k(j, \xi_1, \xi_2)$ in order to determine the confidence set. Given S_j and S_m, Λ_j is constant, and hence the confidence set is most easily determined as the set of j for which

$$P\left\{\max_n \Lambda_n \leq (\max_n \Lambda_n)_{obs} | S_j, S_m\right\} \leq 1 - \alpha. \tag{13}$$

Approximations for this conditional probability which seem adequate for many cases are given below.

Bayesian credible sets for the change–point have been considered by Smith (1975) and Raferty and Akman (1986). Although some numerical computation is required, the computa-

tional problems are not particularly onerous. However, the elegant relation of Section 2, where any shift equivariant credible set for the uniform prior was also a confidence set is no longer valid. Results of Stein (1985) lead one to hope that a similar relation is approximately true in the present context; but because the likelihood function is not smooth, a new argument is required to make such a relation precise.

Some close cousins of the likelihood ratio confidence set might also be considered. For example, Worsley's (1986) D_α includes j in the confidence set if the likelihood ratio tests for no change in $[0, j-1]$ and in $[j, m]$ are both accepted at significance levels greater than $1-(1-\alpha)^{1/2} \cong \alpha/2$. Alternatively, Pettitt's (1980) test might be inverted to yield a confidence set. A third possibility is to invert the likelihood ratio test in the conditional model given S_m. It would be interesting to study the expected sizes of these confidence sets along the lines of Section 3, but the computations will be substantially more complicated. At present one can make the following qualitative comparisons. (i) If one considers the boundary crossing problems defined by the likelihood ratio confidence set and Worsley's D_α in the simple case of a normal mean, one sees that for a "typical" sample path Worsley's D_α is more likely to include values of j far from the true one and less likely to include close by values. (ii) Pettitt's test presumably gives smaller confidence sets than the likelihood ratio test for values of j near $m/2$ and larger sets for values of j near 0 and m. See James, James, and Siegmund (1987) for related results about the power of the tests. An objection to the use of Pettitt's test is that for values of j not close to $m/2$ the two factors in the relevant probability (cf. (14) below) are quite unequal with the result that the confidence sets are biased in the direction of $m/2$ and hence give the impression that the change-point is closer to $m/2$ than is actually the case.

Given (S_j, S_m) the random variables $\max_{n \le j} \Lambda_n$ and $\max_{j \le n < m} \Lambda_n$ are conditionally independent, and hence the left hand side of (13) is of the form

$$P\left(\max_{n \le j} \Lambda_n \le a | S_j, S_m\right) P\left(\max_{j \le n < m} \Lambda_n \le a | S_j, S_m\right) \tag{14}$$

These two probabilities present similar computational problems, so it suffices to consider the second one, or equivalently

$$P\left(\max_{j\le n<m}\Lambda_n > a|S_j, S_m\right) \tag{15}$$

In order to evaluate (15) Worsley (1986) in the special case of exponentially distributed observations uses repeated numerical integration, and Siegmund (1986) in the case of normal observations with known variance gives an asymptotic approximation. Related approximations representing different compromises between accruacy and simplicity are suggested below.

Suppose initially that F_θ is a d-variate normal distribution with mean vector θ and identity covariance matrix. The case of an arbitrary, known covariance matrix is easily reduced to this one. Then the probability (15) equals

$$P\left\{\max_{j\le n<m}\frac{\| nS_m/m - S_n\|^2}{2n(1-n/m)} > a\middle| jS_m/m - S_j = \xi\right\}, \tag{16}$$

for which Siegmund (1986) in the case $d = 1$ gives an approximation when j, a, and $\|\xi\|$ are proportional to m, and

$$c^2 = 2a - \|\xi\|^2/j(1-j/m) \tag{17}$$

is a positive multiple of m as $m \to \infty$. A generalization of that argument shows that (16) is

$$\sim \left[1 + c^2\|\xi\|^{-2} j(1-j/m)\right]^{d/2} \nu\left\{c^2 j/(m\|\xi\|) + \|\xi\|/j(1-j/m)\right\} \exp(-c^2/2), \tag{18}$$

where ν is defined in (4) and given approximately by (5). Appendix B gives a version of (18) for exponential families.

One can obtain a simpler and quite general approximation by means of weak convergence arguments to replace the likelihood ratio process Λ_n by $\|B_0(t) - t\|^2 / t(1-t)$, where $t = n/m$ and B_0 is a d dimensional Brownian bridge superimposed on a triangular drift. This approach leads to (18) with $\nu \equiv 1$. Although the approximation is quite conservative, its simplicity and generality make it useful in complicated cases.

One obtains a different simplication of (18) by assuming that $c(n, m)$ satisfies $c \to \infty$ and $c^2/m \to 0$. Then (18) is

$$\sim \xi \quad j(1 \quad m) \exp(-c^2/2) \tag{19}$$

This approximation has the disadvantage that it does not depend on d. We shall see its advantages below. From the simulations in Table 6 of Siegmund (1986) in the case $d = 1$ one can see that (19) is reasonably accurate for the range of j, m, and ξ considered there. Presumably it is less accurate for larger c, smaller m, and/or larger d, but it seems more than adequate for many cases of interest.

For smooth exponential families the approximation (19) takes the form

$$P\left\{\max_{j\le n<m}\Lambda_n \ge a|S_j, S_m\right\} \sim \nu^* \exp[-(a - \Lambda_j)], \tag{20}$$

where $a - \Lambda_j$ is assumed small compared to m and ν^* is a distribution dependent quantity whose exact definition is given in Appendix B. A detailed example involving the exponential distribution is discussed below.

In the normal case, according to the approximations (19) and (5) the confidence set defined by (13) is the set of all τ such that

$$\left\{1 - \exp(-.583[2\Lambda_\tau/\{\tau(1 - \tau/m)\}]^{1/2} - (\max_n \Lambda_n - \Lambda_\tau))\right\}^2 \le 1 - \alpha. \tag{21}$$

Even when one questions the accuracy of (19) or when the data are not normal, the central limit theorem suggests the use of (21) as a first approximation. A better approximation, simulation, or numerical methods can be used to decide whether values of τ on the boderline according to (21) should be included in or excluded from the confidence set.

Note also the formal similarity between (19) and (6). To the extent that $\{n(1 - \tau/m)\}^{1/2}$ is nearly constant over the values τ of interest, e.g., when the likelihood ratio statistic is sharply peaked and hence the confidence set is small, (21) shows that the confidence set consists of those τ for which Λ_τ is within some distance of $\max_n \Lambda_n$, which can be displayed graphically as in Section 2.

Figure 2 shows the log likelihood ratio statistic and the approximate cutoff for a 95% confidence set for the same simulated data as in Figure 1. Qualitatively the cases of known and unknown λ look quite similar. Usually the confidence set is larger in the case of unknown λ, and this is indeed so in the lower plot. However the reverse is true in the upper plot.

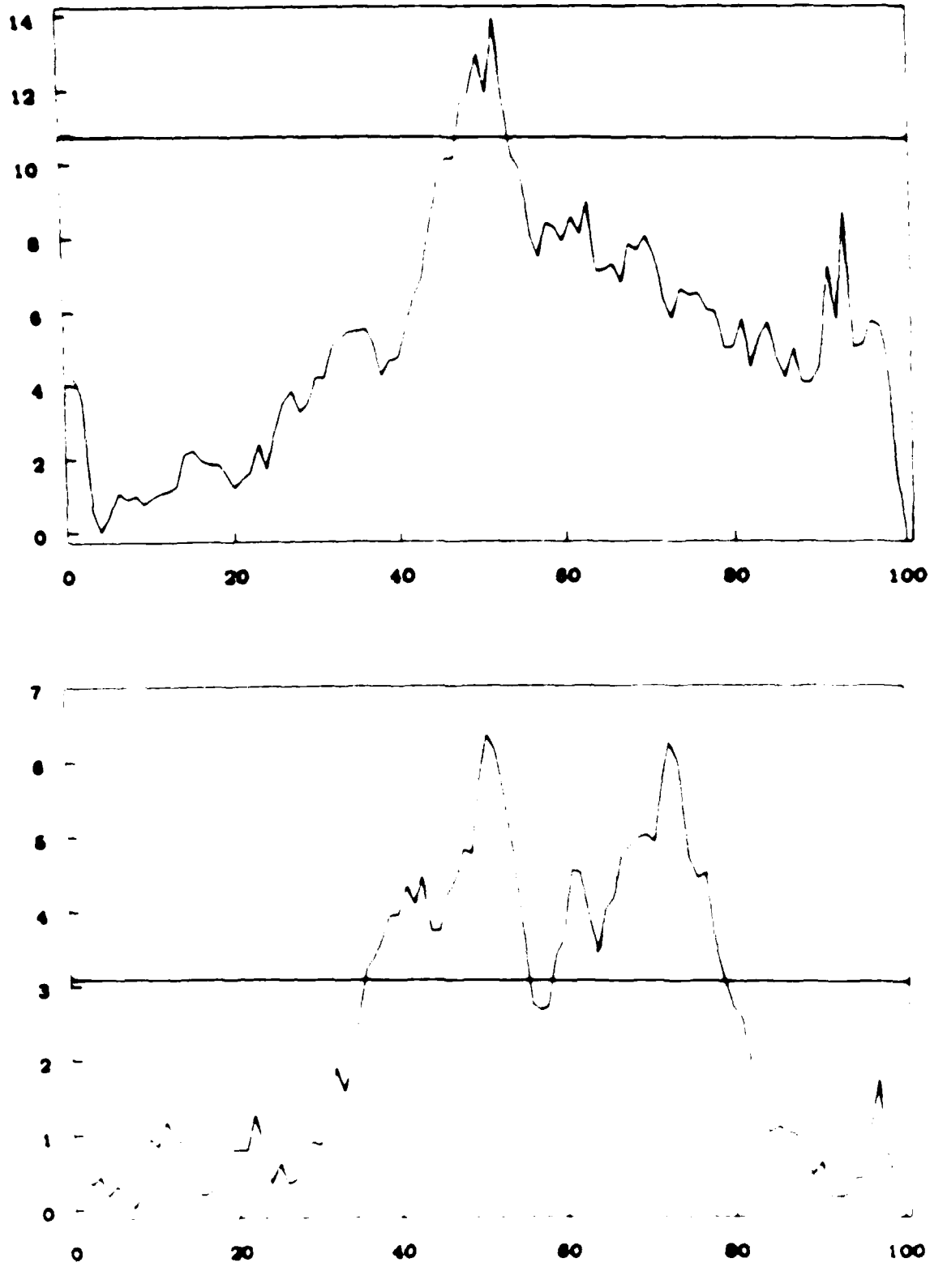


Figure 2 Simulated Log Likelihood Ratio Processes, λ Unknown

presumably because the procedure in effect estimates δ and then acts as if the, in this case large, estimated value is the true one.

As an illustration we consider the British coal mining accident data of Maguire, Pearson and Wynn (1952), as extended and corrected by Jarrett (1979). Worsley (1986) has analysed the original data and determined the likelihood ratio confidence set by numerical computation of (14).

The data are intervals in days between accidents in British coal mines in which at least ten deaths occurred. Jarrett's (1979) data involve $m = 190$ intervals from 15 March, 1851 to 22 March, 1962, a period of 40,549 days. Under the assumption that the intervals $y_1, \ldots, y_m$ are independent and exponentially distributed with a change after the j-th observation in the mean time between accidents, we shall determine a likelihood ratio confidence set for j.

The likelihood ratio statistic is $\max_n \Lambda_n = \max_n\{m \log(W_m/m) - n \log(W_n/n) - (m - n)\log\{(W_m - W_n)/(m - n)\}\}$, where $W_n = y_1 + \cdots + y_n$. For Jarrett's data the maximum value equals 35.6 and is assumed at $n = 124$ in the year 1890. According to (B.13)–(B.15) in Appendix B the likelihood ratio confidence set is given approximately as the set of all j such that

$$1 - c^*(\lambda_1, \lambda_2)\exp\{-(\max_n \Lambda_n - \Lambda_j)\} \le 1 - (\lambda_2/\lambda_1)\exp\{-(\max_n \Lambda_n - \Lambda_j)\} \le 1 - \alpha, \quad (22)$$

where $\lambda_1 = (W_j/j)^{-1}$, $\lambda_2 = \{(W_m - W_j)/(m - j)\}^{-1}$, and $c^*(\cdot)$ is defined in (B.14). The use of (21) assumes that $\lambda_1 > \lambda_2$; if $\lambda_2 > \lambda_1$, then the values of λ_1 and λ_2 should be interchanged. The approximation (22) gives the set $\{116, 117, \ldots, 128, 133\}$ as a 95% confidence set for the change point. This corresponds to the interval from 1887 to 1893 together with an isolated point in 1897.

Application of (22) to the original Maguire, Pearson, and Wynn (1952) data gives precisely the same confidence set which Worsley computed numerically. However, because of discrepancies between the two data sets, the years covered by the two confidence sets are slightly different.

Raferty and Akman (1986) give a flat prior Bayesian analysis of these data. It appears from their calculations and Figure that a highest posterior set estimate for the change point is

essentially the same as the confidence set computed here. Actually Raferty and Akman consider a continuous time Poisson process model and hence allow the change to occur between event times. It is a straightforward matter to adapt the theory developed here to allow for that possibility.

Cobb (1978) has suggested an extension of method (iv) in Section 2 to deal with nuisance parameters, but it contains some arbitrary features which may make it difficult to implement with small or moderate sample sizes. An interesting variant of Cobb's analysis has recently been proposed by Hinkley and Schechtman (1987).

5 Joint Confidence Sets

The likelihood ratio method can also be adapted to give joint confidence sets for the change-point j and some function δ of the parameters θ_0 and θ_1. We begin with the simple case that the x_i are normally distributed with mean θ_0 or θ_1 according as $1 \le i \le j$ or $j < i \le m$ and identity covariance matrix, and take $\delta = \theta_1 - \theta_0$.

The acceptance region of the likelihood ratio test that the parameters are j and δ is

$$A_{j,\delta} = \left\{ \sup_i \Lambda_i - i\delta'(jS_m/m - S_i) - j(1 - j/m)\|\delta\|^2/2 \le c^2/2 \right\}$$

where $\Lambda_i = \|iS_m/m - S_i\|^2/\{2i(1 - i/m)\}$ and $c = c(j, \delta)$ is chosen to satisfy

$$P(A_{j,\delta}) = 1 - \alpha$$

for all j, δ. Note that

$$\sup_i \Lambda_i - i\delta'(jS_m/m - S_i) - j(1 - j/m)\|\delta\|^2/2$$
$$= \sup_i \Lambda_i - \Lambda_j + i\delta'(jS_m/m - S_i) - j(1 - j/m)\delta'\delta/2 + j(1 - j/m)\delta'\delta/2 \qquad (3)$$

and since the first difference on the right hand side is necessarily non-negative, one obtains

$$P(A_{j,\delta}) = P\left\{ i\delta'(jS_m/m - S_i) - j(1 - j/m)\delta'\delta/2 \le c^2/2 - j(1 - j/m)\delta'\delta/2 \right\}$$
$$+ P\left\{ P(A_{i,\delta}) \le i\delta'(jS_m/m - S_i) - j(1 - j/m)\delta'\delta/2 \le c^2/2 - j(1 - j/m)\delta'\delta/2 \right\} \qquad (4)$$

The first term on the right hand side of (24) is exactly $1 - F_d(c^2)$, where F_d denotes the χ^2 distribution with d degrees of freedom. According to (19)

$$P(A^c_{j,\delta} | jS_m/m - S_j = \xi)$$
$$\sim 2\nu[\,\|\xi\|/\{j(1-j/m)\}]\exp\{-c^2/2 + \|\xi - j(1-j/m)\delta\|^2/[2j(1-j/m)]\},$$

provided the exponent is small compared to m as $m \to \infty$. Substitution of this approximation into (24) yields (if $c^2 = o(m)$)

$$P(A^c_{j,\delta}) \cong 1 - F_d(c^2) + \nu(\delta)c^d \exp(-c^2/2)/\{2^{d/2-1}\Gamma(d/2+1)\}. \tag{25}$$

If instead of (19) one uses in (24) the presumably more accurate approximation (18), the integration must be performed numerically.

Using (25) one can easily find an approximate confidence set by trial and error.

An extension of this method to non-normal exponential families requires a consideration of special cases, depending on the parameter δ of interest. The generalization of (23), in an almost obvious notation is

$$\sup_i \Lambda_i - \Lambda_j^{(\delta)} = (\sup_i \Lambda_i - \Lambda_j) + \Lambda_j - \Lambda_j^{(\delta)}.$$

If δ is a function of the difference between the natural parameters of the exponential family, e.g. if the parent populations are Poisson and δ is the ratio of their means, one obtains distributions parameterized by δ by computing probabilities conditionally, given S_m. On the other hand, if the parent distributions are exponential and δ is again the ratio of their means, considerations of invariance of the two sample problem under scale changes shows that unconditional probabilities are appropriate. In either case, using a χ^2 approximation to the distribution of $\Lambda_i - \Lambda_j^{(\delta)}$ in conjunction with (20), one obtains an approximation similar to (25), but with the ν^* appropriate to the distribution under consideration in place of ν. In large samples one may consider replacing ν^* by ν, but some thought must be given to the choice of argument of the function ν.

For the special case of exponentially distributed y's having mean λ^{-1} and $\delta = \lambda_1/\lambda_0$, one

obtains from (B.13)–(B.15) when $\delta > 1$

$$P(A^c_{j,\delta}) \cong 2[1 - \Phi(c)] + 2[\nu^*(\delta) + \delta^{-1}]c\varphi(c), \tag{26}$$

where ν^* is defined in (B.14). When $\delta < 1$ (26) holds with δ^{-1} in place of δ.

Table 3 gives an approximate 90% joint confidence set for j and $\delta = \lambda_1/\lambda_2$ for the British coal mining data.

Table 3

90% Confidence set for (j, δ)

j	δ	j	δ
115	(2.7, 3.9)	124	(2.3, 5.3)
116	(2.5, 4.3)	125	(2.4, 4.9)
117	(2.4, 4.6)	126	(2.4, 5.0)
118	(2.3, 4.8)	127	(2.4, 4.7)
119	(2.4, 4.5)	128	(2.5, 4.4)
120	(2.5, 4.3)	129	(2.7, 4.0)
121	(2.4, 4.7)	130	(2.9, 3.6)
122	(2.4, 4.7)	132	(2.8, 3.9)
123	(2.3, 4.9)	133	(2.5, 4.5)

Acknowledgements

I would like to thank M. Pollak and T. Sellke for very helpful discussions, I. Einott for some programming assistance, and Qiwei Yao for spotting the transcription error in Siegmund (1986), which is mentioned in Section 2.

APPENDIX A

Informal Proof of Theorem 1.

We consider only the confidence interval $[L, R]$. The proof for the likelihood ratio confidence set is similar and somewhat simpler. Since the confidence set is equivariant, it suffices to consider the case $j = 0$. To simplify the notation we shall write P and E instead of P_0 and E_0, η instead of η', S_n instead of $\tilde{S}_n$, and take $\delta = 1$. Recall that $M = \sup_{n\geq 0} S_n$.

For arbitrary $n_0 = 1, 2, \ldots$

$$
\begin{aligned}
E(R-L) &= \sum_{-\infty}^{\infty} P(L \leq n \leq R) = P(L \leq 0 \leq R) + 2\sum_{1}^{\infty} P(L \leq n \leq R) \\
&= P(L \leq 0 \leq R) + 2\sum_{1}^{\infty} \{P(R \geq n) - P(L > n)\} \\
&= 1 + 2\sum_{1}^{\infty} P(R \geq n) + o(1) \qquad \text{as } \eta \to \infty \\
&= 1 + 2n_0 + 2\sum_{n_0+1}^{\infty} P(R \geq n) - 2\sum_{1}^{n_0} \{1 - P(R \geq n)\} + o(1). \qquad (A1)
\end{aligned}
$$

For positive n, by the definition of R

$$
\begin{aligned}
P(R \geq n) &= P\left(\sup_{i<n} S_i \leq \sup_{i\geq n} S_i + \eta\right) \\
&= \int\int_{[-\eta,0)\times[0,\infty)} P\left\{S_n \in d\xi, \max_{i\geq n}(S_i - S_n) \in dy\right\} \\
&\qquad \times P\left(\max_{0\leq i\leq n} S_i \leq \eta + \xi + y | S_n = \xi\right) P\left(\max_{i\leq 0} S_i \leq \eta + \xi + y\right) \\
&= \int\int_{[-\eta,0)\times[0,\infty)} P(S_n \in d\xi) P(M \in dy) P\left(\max_{0\leq i\leq n} S_i \leq \eta + \xi + y | S_n = \xi\right) P(M \leq \eta + \xi + y) \\
&= \int\int_{[0,\infty)\times[0,\infty)} P(S_n \in -\eta + dx) P(M \in dy) P\left(\max_{0\leq i\leq n} S_i \leq x + y | S_n = x - \eta\right) P(M \leq x - y)
\end{aligned}
$$

Let $n_0 = \lfloor 2\eta \rfloor$ and $k = n - n_0$, so

$$P(S_n \in -\eta + dx) = \varphi\left\{\frac{x + k/2 - \eta + n_0/2}{(n_0+k)^{1/2}}\right\}(n_0+k)^{-1/2}dx.$$

It may be shown that the contribution to the two series in (A1) from values of x and k outside the range $|k| \le \eta^{2/3}, |x + k/2| \le \eta^{2/3}$ is negligible, and inside this range

$$P\left(\max_{0\le i\le n_0+k} S_i > x + y | S_{n_0+k} = -\eta + x\right) - P(M > x + y)$$

converges uniformly to 0. Hence for the purpose of evaluating (A1) asymptotically, $P(R \ge n_0 + k)$ may be replaced by

$$\int_0^{\eta^{2/3}} \int_0^{\infty} \varphi\left\{\frac{x + k/2}{(n_0+k)^{1/2}}\right\}(n_0+k)^{-1/2}dx\ P(M \in dy)\{1 - 2P(M > x+y) + P^2(M > x+y)\}.$$

For $k = 0$ this integral converges to $1/2$. The terms in (A1) for $k = \pm 1, \pm 2, \ldots$ may be paired, and after some calculation one obtains

$$\begin{aligned}\sum_{k\ge 1} P(R \ge n_0 + k) - \sum_{k\le 0}\{1 - P(R \ge n_0 + k)\} &= -1/2 \\ &+ \sum_{k\ge 1}[\Phi\{2^{-1}k/(n_0 - k)^{1/2}\} - \Phi\{2^{-1}k/(n_0 + k)^{1/2}\} \\ &- 2n_0^{-1/2}\varphi(2^{-1}/n_0^{1/2}) \int_0^{\infty}\int_0^{\infty} P(M \in dy)\{2P(M > x + y) \\ &- P^2(M > x + y)\}dx + o(1).\end{aligned}$$

A Taylor series expansion, approximation of Riemann sums by integrals, and substitution of the result back into (A1) complete the informal proof of Theorem 1.

APPENDIX B

This appendix is concerned with approximations to boundary crossing probabilities like (15).

Let $x_1, x_2, \ldots,$ be independent random variables with probability density function of the form

$$f_\theta(x) = \exp\{\theta x - \psi(\theta)\} f_0(x),$$

where f_0 is without loss of generality standardized to have mean 0 and variance 1. We shall consider only the case of real x and θ, although the extension to the multivariate case is straightforward. However, the case of vector θ in which some components change at j while others are assumed to remain fixed is substantially more difficult. See James, James, and Siegmund (1986) for the special case of normal observations whose mean changes while the unknown variance does not. One can also handle discrete random variables and continuous time Poisson process. Some remarks about the necessary modification in the argument are given below.

Let $S_n = x_1 + \ldots + x_n, H(x) = \sup_\theta\{\theta x - \psi(\theta)\}, \Lambda_n(\xi, \eta) = nH(\eta/n) + (m-n)H[(\xi - \eta)/(m-n)] - mH(\xi/m), \Lambda_n(\xi) = \Lambda_n(\xi, S_n)$, and define

$$T = \inf\{n : n \geq m_0, \Lambda_n(\xi) \geq a\} \tag{B.1}$$

($= \infty$ if $\Lambda_n(\xi) < a$ for all $m_0 \leq n < m$). Also put $\mu = \psi'(\theta)$. We shall write P_μ to denote dependence of probabilities on the parameter θ. For events A defined in terms of $x_1, \ldots, x_m$ let

$$P_\xi^{(m)}(A) = P_\mu(A | S_m = \xi).$$

By sufficiency this probability does not depend on μ. Theorem B.1 gives approximations for $P_\xi^{(m)}(T < m | S_{m_0} = \eta)$.

In order to describe those approximations let $\hat\theta = \hat\theta(x)$ be defined by $\psi'(\hat\theta) = x$, so $H(x) = \hat\theta x - \psi(\hat\theta)$. Note that $H'(x) = \hat\theta$ and $xH'(x) - H(x) = \psi(\hat\theta)$. For $\mu_1 \neq \mu_2$ let $\hat\theta_1 = \hat\theta(\mu_1)$ and $\hat\theta_2 = \hat\theta(\mu_2)$, and define

$$\tau = \inf\left\{n : (\hat\theta_1 - \hat\theta_2)S_n - n[\psi(\hat\theta_1) - \psi(\hat\theta_2)] \geq b\right\}.$$

Also let

$$\nu^*(\mu_1,\mu_2) = \lim_{b\to\infty} E_{\mu_1} \exp\big(- \{(\hat\theta_1 - \hat\theta_2)S_\tau - \tau[\psi(\hat\theta_1) - \psi(\hat\theta_2)] - b\}\big). \qquad (B.2)$$

The limit indicated in (B.2) exists as a consequence of the renewal theorem. A general method for computing ν^* numerically has been given by Woodroofe (1979). In the special case of normal x's $\nu^*(\mu_1,\mu_2) = \nu\big(|\mu_1 - \mu_2|\big)$, where ν is defined in (4) and given approximately by (5). The case of exponentially distributed x's is discussed below.

Theorem B.1. Assume for fixed $0 < t_0 < 1, a_0 > 0, \xi_0$, and $\eta_0 \neq \xi_0 t_0$ that $m_0 \sim mt_0, a \sim ma_0, \xi \sim m\xi_0$, and $\eta \sim m\eta_0$. Let t^* be defined by

$$t^* H(\eta_0/t_0) + (1-t^*)H\{(\xi_0 - \eta_0 t^*/t_0)/(1-t^*)\} - H(\xi_0) = a_0,$$

and assume that $t_0 < t^* < 1$. Then as $m \to \infty$, for T defined by (B.1)

$$P_\xi^{(m)}\{T < m | S_{m_0} = \eta\} \sim \exp\{-[a - \Lambda_{m_0}(\xi,\eta)]\}$$
$$\times \left\{\frac{(1-t_0)t^* H''[(\xi_0 - \eta_0 t^*/t_0)/(1-t^*)]}{t_0(1-t^*)H''[(\xi_0 - \eta_0)/(1-t_0)]}\right\}^{1/2} \nu^*\big(\eta_0/t_0, (\xi_0 - \eta_0\, t^*/t_0)/(1-t^*)\big) \qquad (B.3)$$

where ν^* is defined in (B.2).

A simpler approximation to $P_\xi^{(m)}\{T < m | S_{m_0} = \eta\}$ is obtained by assuming

$$\infty \longleftarrow a - \Lambda_{m_0}(\xi,\eta) = o(m). \qquad (B.4)$$

In this case $t^* = t_0$, so (B.3) becomes

$$P_\xi^{(m)}\{T < m | S_{m_0} = \eta\} \sim \nu^*\big(\eta_0/t_0, (\xi_0 - \eta_0)/(1-t_0)\big) \exp\{-[a - \Lambda_{m_0}(\xi,\eta)]\}. \qquad (B.5)$$

Complete proofs of (B.3) and (B.5) are quite long and technical. The main idea and some important lemmas are given here. The method is inspired by that of Lai and Siegmund (1977, Section 3), but it differs in several crucial ways.

Let f_n denote the n–fold convolution of $f_0 (n = 1, 2, \ldots)$ and assume that f_n has an integrable characteristic function for some n. The following large deviation approximation for f_n is used repeatedly. (See Borovkov and Rogozin, 1965.) As $n \to \infty$

$$f_n(nx) \sim [H''(x)/2\pi n]^{1/2} \exp[-nH(x)]. \qquad (B.6)$$

Let $Q = \int_{-\infty}^{\infty} P_{\xi'}^{(m)} d\xi'/(2\pi)^{1/2}$, and let L_n denote the likelihood ratio of $x_1, \ldots, x_n$ under Q relative to $P_\xi^{(m)}$ $(n = 1, 2, \ldots, m-1)$. Then

$$L_n = [f_{m-n}(\xi - S_n)/f_m(\xi)]^{-1} \int_{-\infty}^{\infty} [f_{m-n}(\xi' - S_n)/f_m(\xi')] d\xi'/(2\pi)^{1/2}. \qquad (B.7)$$

The following representation is basic.

Lemma B.1. $P_\xi^{(m)}\{T < m | S_{m_0} = \eta\}$

$$= \frac{f_m(\xi)}{f_{m-m_0}(\xi - \eta)} \int_{-\infty}^{\infty} \frac{f_{m-m_0}(\xi' - \eta)}{f_m(\xi')} E_{\xi'}^{(m)} \left[L_T^{-1} 1_{\{T<m\}} | S_{m_0} = \eta\right] d\xi'/(2\pi)^{1/2}. \qquad (B.8)$$

Proof. By Wald's likelihood ratio identity and the definition of Q

$$\begin{aligned} P_\xi^{(m)}\{\eta \le S_{m_0} < \eta + \delta,\ T < m\} &= \int_{\{\eta \le S_{m_0} < \eta+\delta,\ T<m\}} L_T^{-1}\, dQ \\ &= \int_{-\infty}^{\infty} \int_{\{\eta \le S_{m_0} < \eta+\delta,\ T<m\}} L_T^{-1}\, dP_{\xi'}^{(m)}\, d\xi'/(2\pi)^{1/2} \\ &= \int_{-\infty}^{\infty} \int_{\eta}^{\eta+\delta} E_{\xi'}^{(m)} \left[L_T^{-1} 1_{\{T<m\}} | S_{m_0} = \eta'\right] P_{\xi'}^{(m)}(S_{m_0} \in d\eta') d\xi'/(2\pi)^{1/2}. \end{aligned}$$

The desired representation follows by dividing by

$$P_\xi^{(m)}(\eta \le S_{m_0} < \eta + \delta) = \int_{\eta}^{\eta+\delta} \{f_{m_0}(\eta') f_{m-m_0}(\xi - \eta')/f_m(\xi)\} d\eta'$$

and letting $\delta \to 0$.

It follows from (B.6) that

$$\frac{f_m(\xi) m}{f_{m_0}(\eta) f_{m-m_0}(\xi - \eta) m_0} \sim \left\{ \frac{2\pi H''(\xi/m) m(m - m_0)}{H''(\eta/m_0) H''[(\xi - \eta)/(m - m_0)] m_0} \right\}^{1/2} \exp[\Lambda_{m_0}(\xi, \eta)].$$

The measure $\{m_0 f_{m_0}(\eta) f_{m-m_0}(\xi' - \eta)/m f_m(\xi')\} d\xi'$ behaves asymptotically like a normal distribution with mean $m\eta_0/t_0$ and standard deviation proportional to $m^{1/2}$, i.e. like a Dirac delta function at $m\eta_0/t_0$. Hence the right hand side of (B.8) is asymptotic to

$$\exp[\Lambda_{m_0}(\xi, \eta)] \left\{ \frac{m H''(\xi_0)(1 - t_0)}{H''(\eta_0/t_0) H''[(\xi_0 - \eta_0)/(1 - t_0)] t_0} \right\}^{1/2} \times E_{m\eta_0/t_0}^{(m)} \left[L_T^{-1} 1_{\{T \le m_2\}} | S_{m_0} = \eta\right]. \qquad (B.9)$$

The following lemma is useful in approximating the conditional expectation in (B.9).

Lemma B.2. For n proportional to m as $m \to \infty$ except for an event of negligibly small probability under $P^{(m)}_{\eta/t_0}(\cdot|S_{m_0} = \eta)$

$$L_n \sim \exp[\Lambda_n(\xi)] \left\{ \frac{H''(\xi_0)m(m-n)}{H''[(\xi_0 - \eta_0 n/m_0)/(1-n/m)]H''(\eta_0/t_0)n} \right\}^{1/2}. \qquad (B.10)$$

The proof of Lemma B.2 follows from substitution of (B.6) into (B.7), the observation that $mH(\xi'/m) - (m-n)H[(\xi' - S_n)/(m-n)]$ is maximized at $\xi' = mS_n/n$, where it equals $nH(S_n/n)$, and a Laplace type asymptotic expansion of the integral in (B.7).

The Hájek–Rényi–Chow inequality applied to the $P^{(m)}_{m\eta_0/t_0}$-martingale $(S_n - n\eta_0/t_0)/(1 - n/m)$ shows that for any $0 < \varepsilon < 1$

$$P^{(m)}_{\eta/t_0}\{|S_n - n\eta_0/t_0| \geq \lambda + n\varepsilon \quad \text{for some} \quad m_0 < n \leq m(1-\varepsilon)|S_{m_0} = \eta\}$$

can be made arbitrarily small by taking λ sufficiently large, and hence

$$m^{-1}T \longrightarrow t^* \quad \text{in} \quad P^{(m)}_{\eta_0/t_0}(\cdot|S_{m_0} = \eta) - \text{probability}.$$

It follows from (B.10) that

$$L_T \sim \exp[\Lambda_T(\xi)] \left\{ \frac{H''(\xi_0)m(1-t^*)}{H''[(\xi_0 - \eta_0 t^*/t_0)/(1-t^*)]H''(\eta_0/t_0)t^*} \right\}^{1/2} \qquad (B.11)$$

except for an event of negligibly small probability under $P^{(m)}_{\eta/t_0}(\cdot|S_{m_0} = \eta)$.

The proof of Theorem B.1 can be completed by substituting (B.11) into (B.9) and appealing to Hu's (1987) conditional nonlinear renewal theorem, which says that the distribution of the excess over the boundary, $\Lambda_T(\xi) - a$, under the conditional probability $P^{(m)}_{\eta/t_0}$ has the same limit as a suitable random walk approximation to $\Lambda_n(\xi)$ under the unconditional probability P_{η_0/t_0}. See Siegmund (1986, Appendix 2) for an intuitive discussion of nonlinear renewal theory.

Now assume that $y_1, y_2, \ldots, y_m$ are independently and exponentially distributed with mean λ^{-1}. Let $W_n = y_1 + \ldots + y_n$ and $S_n = n - W_n$. Then $\theta = \lambda - 1$ and $\psi(\theta) = \theta - \log(1+\theta)$. Let $\hat{\lambda}_1 = (W_{m_0}/m_0)^{-1}$, $\hat{\lambda}_2 = [(W_m - W_{m_0})/(m - m_0)]^{-1}$ and assume that $\hat{\lambda}_1 > \hat{\lambda}_2$. (The case $\hat{\lambda}_1 < \hat{\lambda}_2$ is similar.) Assuming that (B.4) holds one can use the lack of memory property of

the exponential distribution to obtain

$$P\{\max_{m_0 \le n < m} \Lambda_n \ge a | W_{m_0}, W_m\} \sim \nu^*(\lambda_1/\lambda_2)\exp[-(a - \Lambda_{m_0})], \qquad (B.13)$$

where in this case

$$\nu^*(\delta) = |\log(\delta)/(\delta - 1) - 1|/|\log(\delta)/(1 - \delta^{-1}) - 1| \qquad (B.14)$$

Similarly

$$P\left\{\max_{n \le m_0} \Lambda_n \ge a | W_{m_0}, W_m\right\} \sim \lambda_2\lambda_1^{-1}\exp[-(a - \Lambda_{m_0})]. \qquad (B.15)$$

The details of these evaluations are omitted.

With minor modifications the methods developed here yield likelihood ratio confidence sets for a change–point in the intensity of a continuously observed Poisson process. They also apply to many discrete exponential families, even though the nonlinear renewal theorem used in the proof of Theorem B.1 requires that certain distributions be non–arithmetic. However, these are the distributions of the P_{η_0/t_0}–random walk $(\dot\theta_1 - \dot\theta_2)S_n - n[\psi(\dot\theta_1 - \psi(\theta_2)]$, which usually are non–arithmetic for all but countably many values of η_0, ξ_0, and t_0.

References

Borovkov, A. A. and Rogozin, B. A. (1965). On the multidimensional central limit theorem. *Theory Probab. Applic.* **10**, 55-62.

Cobb, G. W. (1978). The problem of the Nile: conditional solution to a change point problem. *Biometrika* **62**, 243-51.

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

Fisher, R. A. (1934). Two new properties of mathematical likelihood. *Proc. R. Soc. A* **144**, 285-307.

Pollak, M. and Siegmund, D. (1985). A diffusion process and its application to detecting a change in the drift of Brownian motion. *Biometrika* **72**, 267-80.

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

Hinkley, D. V. and Schechtman, E. (1987). Conditional bootstrap methods in the mean shift model. *Biometrika* **74**, 85-94.

Hinkley, D. V. (1972). Time ordered classification. *Biometrika* **59**, 509-523.

Hooper, P. M. (1982). Invariant confidence sets with smallest expected measure. *Ann. Statist.* **10**, 1283-94.

Hora, R. B. and Buehler, R. J. (1966). Fiducial theory and invariant estimation. *Ann. Math. Statist.* **37**, 643-656.

Hu, I. (1987). Nonlinear renewal theory for conditional random walks. University of Pennsylvania Preprint.

Ibragimov, I. A. and Khasminski, R. Z. (1981). *Statistical Estimation: Asymptotic Theory*. Springer Verlag, New York Heidelberg Berlin.

James, B., James, K. L., and Siegmund, D. (1986). Conditional boundary crossing probabilities with applications to change point problems. Stanford University Technical Report.

James, B., James, K. L., and Siegmund, D. (1987). Tests for a change point. *Biometrika* 74, 71-84.

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

Lai, T. L. and Siegmund, D. (1977). A nonlinear renewal theorem with applications to sequential analysis I. *Ann. Statist.* **5**, 946-954.

Maguire, B. A., Pearson, E. S., and Wynn, A. H. A. (1952). The time intervals between industrial accidents. *Biometrika* **38**, 168-80.

Pettitt, A. N. (1980). A simple cumulative sum type statistic for the change point problem with zero-one observations. *Biometrika* **67**, 79-84.

Raferty, A. E. and Akman, V. E. (1986). Bayesian analysis of a Poisson process with a change point. *Biometrika* **73**, 85-90.

Siegmund, D. (1985). *Sequential Analysis: Test and Confidence Intervals*. Springer Verlag New York-Heidelberg-Berlin.

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

Smith, A. F. M. (1975). A Bayesian approach to inference about a change point in a sequence of random variables. *Biometrika* **62**, 407-416.

Stein, C. (1985). On the coverage probability of confidence sets based on a prior distribution. *Banach Center Publications* **1b**, 485-514.

Stein, C. (1965). Approximation of improper prior measures by prior probability measures. *Bernoulli, Bayer, Laplace, Anniversary Volume*. J. Neyman and L. M. Le Cam, eds. Springer-Verlag, New York-Heidelberg-Berlin, 217-240.

Woodroofe M. 1979. Repeated likelihood ratio tests. *Biometrika* **66** 453-463.

Worsley K. J. 1986. Confidence regions and tests for a change point in a sequence of exponential family random variables. *Biometrika* **73** 91-104.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1 REPORT NUMBER: 2	2 GOVT ACCESSION NO.: AD-A181075	3 RECIPIENT'S CATALOG NUMBER
4 TITLE (and Subtitle): CONFIDENCE SETS IN CHANGE-POINT PROBLEMS		5 TYPE OF REPORT & PERIOD COVERED: Technical Report
		6 PERFORMING ORG. REPORT NUMBER
7 AUTHOR(s): David Siegmund		8 CONTRACT OR GRANT NUMBER(s): N00014-87-K-0078
9 PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics - Sequoia Hall, Stanford University, Stanford, California 94305-4065		10 PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR-042-373
11 CONTROLLING OFFICE NAME AND ADDRESS: Statistics & Probability Program Code (411 (SP)), Office of Naval Research, Arlington, Virgina 22217		12 REPORT DATE: May 1987
		13 NUMBER OF PAGES: 30pp.
14 MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15 SECURITY CLASS. (of this report): Unclassified
		15a DECLASSIFICATION/DOWNGRADING SCHEDULE

16 DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17 DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18 SUPPLEMENTARY NOTES

19 KEY WORDS (Continue on reverse side if necessary and identify by block number)

Change-point, likelihood ratio, boundary crossing probabilities.

20 ABSTRACT (Continue on reverse side if necessary and identify by block number)

Several methods are discussed for confidence set estimation of a change-point in a sequence of independent observations from completely specified distributions. The method based on the likelihood ratio statistic is extended to the case of independent observations from an exponential family. Joint confidence sets for the change-point and the parameters of the exponential family are also considered.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

7-87

DTIC