Capturing the Distributional Behaviour of the Maximum Likelihood Estimator of a Changepoint

Venkata K. Jandhyala; Stergios B. Fotopoulos

Biometrika, Vol. 86, No. 1. (Mar., 1999), pp. 129-140.

Biometrika (1999), **86**, 1, *pp.* 129–140

Printed in Great Britain

Capturing the distributional behaviour of the maximum likelihood estimator of a changepoint

By VENKATA K. JANDHYALA

Department of Pure and Applied Mathematics and Program in Statistics, Washington State University, Pullman, Washington 99164, U.S.A.

jandhyala@wsu.edu

AND STERGIOS B. FOTOPOULOS

Department of Management and Systems and Program in Statistics, Washington State University, Pullman, Washington 99164, U.S.A

fotopo@wsu.edu

Summary

We consider the problem of estimating the unknown changepoint in a sequence of time-ordered observations. Upper and lower bounds are derived for the asymptotic distribution of the maximum likelihood estimator and methods of approximation are suggested. A computationally efficient algorithm is presented for deriving the bounds and approximations for the asymptotic probabilities of the maximum likelihood estimator when the parameters before and after the changepoint are unknown. We also show an essentially exponential rate of convergence of the probability distribution of the maximum likelihood estimator from finite samples to the case of infinite samples. We apply the algorithm to the cases of normal and exponential distributions. For the exponential distribution the lower and upper bounds for the right tail probabilities of the maximum likelihood estimator, and the two approximations, are identical. This is not the case for the normal distribution. Finally, we apply our changepoint analysis for the case of the exponential distribution to data on explosions in British coal mines.

Some key words: Maximum likelihood estimator; Maximum of a random walk; Negative drift; Parameter change.

1. Introduction

In this paper, we derive a computationally efficient algorithmic procedure for finding the asymptotic distribution of the maximum likelihood estimator of a changepoint. The problem of maximum likelihood estimation of a changepoint was first considered by Hinkley (1970, 1971, 1972). Subsequently, the problem was considered under various settings by Bhattacharya & Brockwell (1976), Ibragimov & Hasminski (1981), Worsley (1986), Cobb (1978), Seigmund (1988), Bhattacharya (1987), Yao (1987) and Rukhin (1994).

In his pioneering work, Hinkley (1970) discussed the asymptotic distribution theory for the maximum likelihood estimator of a changepoint assuming the parameters both before and after the unknown changepoint to be known. Then, considering the case of unknown parameters, Hinkley (1972) demonstrated that, asymptotically, the probability distributions of the maximum likelihood estimator when the parameters were known and unknown were identical. The result suggests that one need only study the asymptotic distribution for the simpler situation with parameters known. Cobb (1978) derived the conditional distribution of the maximum likelihood estimator given the likelihood shape adjacent to the estimated changepoint. Cobb's conditional inferences are known to be nominally equivalent to certain Bayesian solutions. Worsley (1986) and Siegmund (1988) considered aspects of unconditional confidence interval estimation of a changepoint. Our analysis is also based on unconditional inference.

The asymptotic distribution derived by Hinkley (1970) is complicated from a computational point of view. Also, the computations are based on the assumption that a certain probability distribution function can be approximated by a sum of exponential terms. Hinkley (1970) implemented such an approximation for the case of normal distribution.

Although the case of unknown parameters is of prime interest, in view of the distributional equivalence result of Hinkley (1972) we derive our results and procedures for the case where the parameters are assumed to be known. After presenting random-walk results in § 3, we deal in § 4 with the rate of convergence problem by deriving a large sample approximation for the uniform distance between the finite sample case and the infinite sample case probabilities of the maximum likelihood estimator. The convergence rate is shown to be essentially exponentially fast. Earlier Yao (1987) showed the convergence rate for a related quantity to be only $O(n^{-\frac{1}{2}})$. Then, in § 5, we derive computationally efficient algorithmic procedures for finding the asymptotic distribution of the maximum likelihood estimator.

In § 6, we implement the algorithmic procedure for the normal and exponential distributions. Finally, in § 7, we apply our changepoint analysis to data on time intervals between successive explosions in British coal mines, given in Maguire, Pearson & Wynn (1952), extended and corrected by Jarrett (1979) and analysed by Worsley (1986). Our confidence region is found to be shorter in length than the confidence region derived by Worsley (1986).

For all proofs, readers are referred to the authors' technical report TR97-4 from the Department of Pure and Applied Mathematics, Washington State University.

2. Maximum likelihood estimator

Let $Y_1, \ldots, Y_n$ represent a sequence of time-ordered continuous and independent random variables defined on a common probability space $(\Omega, \mathscr{A}, P)$. Let $F(., \lambda)$, for $\lambda \in \Theta$, be the distribution function, assumed to be absolutely continuous, and let $f(., \lambda)$ denote the probability density function. Let λ change from λ_0 to λ_1, for $\lambda_0 \neq \lambda_1$, at some unknown point v such that

$$f(y_i, \lambda) = \begin{cases} f(y_i, \lambda_0) & \text{for } \lambda_0 \in \Theta,\ i = 1, \ldots, v, \\ f(y_i, \lambda_1) & \text{for } \lambda_1 \in \Theta,\ i = v+1, \ldots, n, \end{cases}$$

where $v \in \{1, \ldots, n-1\}$.

Our problem is to estimate the changepoint v when λ_0 and λ_1 are unknown. Even when λ_0 and λ_1 are known, the problem is known to be quite intractable. We first derive the maximum likelihood estimator of v, along the lines of Hinkley (1972).

For fixed known values of λ_0 and λ_1, the likelihood function is given by

$$L(v) = \prod_{i=1}^{v} f(y_i; \lambda_0) \prod_{i=v+1}^{n} f(y_i; \lambda_1).$$

Following Bhattacharya (1994), the maximum likelihood estimator of ν is given by

$$\hat{\nu}_n = \arg \max_{1 \leqslant j \leqslant n-1} \sum_{i=1}^{j} W_i,$$

where $W_i = \log\{f(Y_i;\, \lambda_0/f(Y_i;\, \lambda_1)\}$, and any non-uniqueness in maximisation is resolved by a suitable convention.

When λ_0 and λ_1 are unknown, we first consider the marginal likelihood $\tilde{L}(\nu)$ given by

$$\tilde{L}(\nu) = \max_{\lambda_0} \prod_{i=1}^{\nu} f(y_i;\, \lambda_0) \max_{\lambda_1} \prod_{i=\nu+1}^{n} f(y_i;\, \lambda_1).$$

The maximum likelihood estimator $\tilde{\nu}_n$ in this case is then given by

$$\tilde{\nu}_n = \arg \max_{1 \leqslant j \leqslant n-1} \tilde{L}(j).$$

Let $\hat{\lambda}_0 | \nu$ and $\hat{\lambda}_1 | \nu$ be the conditional maximum likelihood estimators of λ_0 and λ_1 for a given ν, for $\nu = 1, \ldots, n-1$. Then clearly $\tilde{\nu}_n$ is based on the above conditional maximum likelihood estimator. Hinkley (1972) let $\nu \to \infty$ and $n - \nu \to \infty$ and found that both $\tilde{\nu}_n$ and $\hat{\nu}_n$ have the same asymptotic distribution. We state the relevant theorem of Hinkley (1972, p. 520) in the following.

Theorem 1. *Let the regularity conditions for the consistency of $\hat{\lambda}_0 | \nu$ and $\hat{\lambda}_1 | \nu$ hold. Then $\tilde{\nu}_n$ and $\hat{\nu}_n$ asymptotically have the same probability distribution.*

In view of the above theorem, we can restrict our attention to $\hat{\nu}_n$ in that $\hat{\nu}_n - \nu$ is associated with the maximum of the maxima of two independent random walks, whereas $\tilde{\nu}_n - \nu$ is much more complicated. We begin by noting that

$$\hat{\nu}_n - \nu = \arg \max_{-\nu+1 \leqslant j \leqslant n-\nu-1} \sum_{i=1}^{\nu+j} W_i = \arg \max_{-\nu+1 \leqslant j \leqslant n-\nu-1} \xi(j),$$

where

$$\xi(j) = \sum_{i=1}^{\nu+j} W_i - \sum_{i=1}^{\nu} W_i = \begin{cases} \sum_{i=1}^{j} X_i^* = S_j^* & (j > 0), \\ 0 & (j = 0), \\ \sum_{i=1}^{-j} X_i = S_{-j} & (j < 0), \end{cases}$$

with $X_i = -W_{\nu-i+1}$, for $1 \leqslant i \leqslant \nu - 1$, and $X_i^* = W_{\nu+i}$ for $1 \leqslant i \leqslant n - \nu$. The above shows that the dependence of $\hat{\nu}_n - \nu$ on ν and $n - \nu$ is quite implicit.

Note that the random walks $S = \{S_n : n \geqslant 0\}$ and $S^* = \{S_n^* : n \geqslant 0\}$ are independent of each other and both have negative means. Thus, both walks eventually drift to $-\infty$.

3. Random walk preliminaries

We recall here some of the classical results from the fluctuation theory for random walks, gleaned from references such as Spitzer (1976), Feller (1971), Asmussen (1986) and Betroin & Doney (1994, 1996).

Let X_1 and X_1^* represent the initial random variables associated with the two independent random walks S and S^* respectively. We shall require the following assumptions regarding X_1 and X_1^* which, for convenience, we state in terms of X_1.

Assumption 1. Assume $-\infty \leqslant E(X_1) < 0$.

Assumption 2. The moment generating function $\phi(s) = E(e^{sX_1})$ converges for $0 \leqslant \mathrm{Re}(s) < a$ for some $a > 0$.

Assumption 3. For $s \in \mathbb{R}$ and $s \in [0, a)$, $\phi(s)$ attains a unique minimum at $\tau \in [0, a]$ such that $\phi(\tau) = \gamma < 1$, $\phi'(\tau) = 0$.

Let Assumptions 1*, 2* and 3* denote the corresponding assumptions for X_1^*.

We now introduce notation and results for the random walk S. Similar notation and results with a superscript * hold for S^*.

First let $\tau_0 = \inf\{n \geqslant 1 : S_n \leqslant 0\}$ be the weak descending ladder epoch, and let $\sigma_x = \inf\{n \geqslant 0 : S_n > x\}$, for $x \geqslant 0$, where σ_0 denotes the strict ascending ladder epoch. Let $M_n = \max\{S_j : 0 \leqslant j \leqslant n\}$ be the maximum of the first n partial sums and let $M = \max S_n$ be the overall maximum. For $x \geqslant 0$, we define

(i) $G_n(x) = \mathrm{pr}(M_n \leqslant x) = \mathrm{pr}(\sigma_x > n)$, for $n \geqslant 0$;

(ii) $G(x) = \mathrm{pr}(M \leqslant x)$;

(iii) $u_n(x) = \mathrm{pr}(\tau_0 > n, S_n \in (0, x])$, for $n \geqslant 0$, $x \geqslant 0$.

Define $V_0 = 0$ and let $V_n = G_n(0) = \mathrm{pr}(\sigma_0 > n)$, for $n \geqslant 1$. Note that $u_0(x) = 0$, for $x > 0$, with $u_0(0) = 1$. Then let $q_n = u_n(\infty) = \mathrm{pr}(\tau_0 > n)$, for $n \geqslant 1$.

It is well known from Spitzer's identity that $V_\infty = e^{-B(1)}$, where $B(s) = \sum (s^n b_n)/n$ and $b_n = \mathrm{pr}(S_n > 0)$, for $n \geqslant 1$. Furthermore, as was demonstrated by Embrechts & Hawkes (1982), see also Downham & Fotopoulos (1981), one may deduce the following iterative procedure for the sequence of probabilities $\{q_n, n \geqslant 0\}$:

$$q_0 = 1, \quad nq_n = \sum_{j=0}^{n-1} b_{n-j} q_j \quad (n \geqslant 1). \tag{3·1}$$

We now recall the following proposition due to Stoyan (1976, p. 83).

PROPOSITION 1. *Under Assumptions* 1–3, *the distribution function* $G(x)$ *of* M *satisfies*

$$1 - \alpha_1 \exp(-\vartheta x) \leqslant G(x) \leqslant 1 - \alpha_2 \exp(-\vartheta x) \quad (x \geqslant 0),$$

where

$$\vartheta = \sup\{\theta \in \mathbb{R} : \phi(\theta) \leqslant 1\}, \quad \alpha_1 = \sup_{t \geqslant 0} h(t), \quad \alpha_2 = \inf_{t \geqslant 0} h(t) \quad (0 \leqslant \alpha_1, \alpha_2 \leqslant 1),$$

$$h(t) = \frac{\mathrm{pr}(X_1 > t)}{\int_t^\infty \exp\{\vartheta(x-t)\}\, d\,\mathrm{pr}(X_1 \leqslant x)}.$$

Furthermore, if $m = E(M)$ *and* $1 - \mathrm{pr}(M = 0) = p_M$, *then*

$$\alpha_2/\vartheta \leqslant m \leqslant \alpha_1/\vartheta, \tag{3·2}$$

$$\alpha_2 \leqslant p_M \leqslant \alpha_1. \tag{3·3}$$

4. ASYMPTOTIC THEORY AND THE RESULTS

Let $\hat{v}_\infty$ denote the maximum likelihood estimator of v based on infinite data in which both v and $n - v$ are themselves assumed to be infinitely large. In practice, therefore, we require that the true changepoint be away from both tails of the data for our asymptotic theory to be valid.

THEOREM 2. *Let $\hat{\nu}_\infty$ be the maximum likelihood estimator of the changepoint ν. Then, for $j \in \mathbb{Z}$, we have*

$$\operatorname{pr}(\hat{\nu}_\infty - \nu = j) = \begin{cases} e^{-B^*(1)}[q_j^* - \int_{0^+}^{\infty} \{1 - G(x)\}\, du_j^*(x)] & (j > 0), \\ e^{-B^*(1) - B(1)} & (j = 0), \\ e^{-B(1)}[q_{-j} - \int_{0^+}^{\infty} \{1 - G^*(x)\}\, du_{-j}(x)] & (j < 0). \end{cases}$$

While Theorem 2 provides the probability distribution for $\hat{\nu}_\infty - \nu$, the expressions therein are not easily computable. The main difficulty is that the distribution function $G(x)$ of M cannot be easily evaluated. However, we exploit the results of Proposition 1 and derive computable inequalities for the probability distribution of $\hat{\nu}_\infty - \nu$. Let $\tilde{u}_n = \int e^{-\lambda x}\, du_n(x)$, for $n \geqslant 1$, be the Laplace transform of $u_n(x)$, for $x > 0$, and let $\tilde{u}_0(\lambda) = 1$. From Spitzer's identity, the sequence $\{\tilde{u}_n(\lambda) : n \geqslant 0, \lambda \in \mathbb{R}^+\}$ satisfies the relation

$$\tilde{Q}(s; \lambda) = \sum_{n=0}^{\infty} s^n \tilde{u}_n(\lambda) = \exp\{\tilde{B}(s; \lambda)\},$$

where $\tilde{B}(s; \lambda) = \sum \{s^n \tilde{b}_n(\lambda)\}/n$ and $\tilde{b}_n(\lambda) = E\{e^{-\lambda S_n} I(S_n > 0)\}$. Thus, the following iterative procedure, which parallels that given by (3·1), facilitates the computation of $\{\tilde{u}_n(\lambda), n \geqslant 0, \lambda \in \mathbb{R}^+\}$:

$$\tilde{u}_0(\lambda) = 1, \quad n\tilde{u}_n(\lambda) = \sum_{j=0}^{n-1} \tilde{b}_{n-j}(\lambda)\tilde{u}_n(\lambda) \quad (n \geqslant 1). \tag{4·1}$$

The following theorem then provides computable inequalities for the distribution of $\hat{\nu}_\infty - \nu$. Through examples, we will illustrate subsequently that the inequalities given by the theorem are quite sharp.

THEOREM 3. *For $j \in \mathbb{Z}^+$ and $0 \leqslant \alpha, \alpha^* \leqslant 1$, define $A_j(\alpha)$ and $A_j^*(\alpha^*)$ by*

$$A_j(\alpha) = e^{-B^*(1)}\{q_j^* - \alpha\tilde{u}_j^*(\vartheta)\}, \quad A_j^*(\alpha^*) = e^{-B(1)}\{q_j - \alpha^*\tilde{u}_j(\vartheta^*)\}.$$

Then, under Assumptions 1–3 *and* 1*–3*, *the following inequalities hold for the distribution of $\hat{\nu}_\infty - \nu$. For $j \in \mathbb{Z}^+$, we have*

(i) $A_j(\alpha_1) \leqslant \operatorname{pr}(\hat{\nu}_\infty - \nu = j) \leqslant A_j(\alpha_2)$, (4·2)

(ii) $A_j^*(\alpha_1^*) \leqslant \operatorname{pr}(\hat{\nu}_\infty - \nu = -j) \leqslant A_j^*(\alpha_2^*)$. (4·3)

The following corollary, which provides simpler bounds, follows easily from the two theorems.

COROLLARY 1. *Under Assumptions* 1–3 *and* 1*–3*, *we have, for $j \in \mathbb{Z}^+$,*
(i) $e^{-\{B(1)+B^*(1)\}} q_j^* \leqslant \operatorname{pr}(\hat{\nu}_\infty - \nu = j) \leqslant e^{-B^*(1)} q_j^*$,
(ii) $e^{-\{B(1)+B^*(1)\}} q_j \leqslant \operatorname{pr}(\hat{\nu}_\infty - \nu = -j) \leqslant e^{-B(1)} q_j$.

In our next result, we derive approximations for $\operatorname{pr}(\hat{\nu}_\infty - \nu = \pm k)$ when k is large. The result is largely based on Veraverbeke & Tuegels (1975).

THEOREM 4. *Let Assumptions* 1–3 *and* 1*–3* *hold. Furthermore, let $u(x) = \sum u_n(x)$ and $u^*(x) = \sum u_n^*(x)$. Then, for k large, we have*
(i) $\operatorname{pr}(\hat{\nu}_\infty - \nu = k) \sim c_1^* e^{-\{B(1)+B^*(1)\}} \gamma^{*k} k^{-3/2} \{e^{-B^*(1/\gamma^*)} + \int_{0^+}^{\infty} u(x)\, dV(x)\}$,
(ii) $\operatorname{pr}(\hat{\nu}_\infty - \nu = -k) \sim c_1 e^{-\{B(1)+B^*(1)\}} \gamma^k k^{-3/2} \{e^{-B(1/\gamma)} + \int_{0^+}^{\infty} u^*(x)\, dV(x)\}$,

where

$$V(x)=1-e^{-\tau x}+\tau\int_{0^+}^{x} u\left(\frac{1}{\gamma}, x-y\right)e^{-\tau y}\,dy, \quad u(s,x)=\sum s^n u_n(x)$$

and $V^*(x)$ *is defined similarly to* $V(x)$.

We now investigate the rate of convergence for the probability distribution of $\hat{\nu}_n$. Yao (1987) derived inequalities of $O(n^{-\frac{1}{2}})$ for $\mathrm{pr}(\hat{\nu}_n \neq \hat{\nu}_\infty)$. Here, we approximate the uniform distance between the probabilities for the finite and infinite sample cases. Let $\mathscr{B}$ be the Borel σ-field defined on $\mathbb{R}$ and consider the uniform distance given by

$$\sup_{A\in\mathscr{B}} |\mathrm{pr}(\hat{\nu}_n-\nu\in A)-\mathrm{pr}(\hat{\nu}_\infty-\nu\in A)|. \tag{4·4}$$

PROPOSITION 2. *Let* $\mathscr{B}$ *be the Borel* σ*-field on* $\mathbb{R}$. *Then,*

$$\sup_{A\in\mathscr{B}} |\mathrm{pr}(\hat{\nu}_n-\nu\in A)-\mathrm{pr}(\hat{\nu}_\infty-\nu\in A)| \leqslant \mathrm{pr}(\hat{\nu}_n\neq\hat{\nu}_\infty).$$

Note that the upper bound above is $\mathrm{pr}(\hat{\nu}_n \neq \hat{\nu}_\infty)$, the quantity considered by Yao (1987). The proof of the proposition is based on a coupling argument.

THEOREM 5. *Let Assumptions* 1–3 *and* 1*–3* *hold. Then, for large n and some constant d, the uniform distance in* (4·4) *satisfies the following approximation*:

$$\sup_{A\in\mathscr{B}} |\mathrm{pr}(\hat{\nu}_n-\nu\in A)-\mathrm{pr}(\hat{\nu}_\infty-\nu\in A)| \sim d\min\{\delta^{\nu}\nu^{-3/2}, \delta^{n-\nu}(n-\nu)^{-3/2}\},$$

where $\delta=\min(\gamma,\gamma^*)$.

Note that the order of approximation is negative exponential times $O(n^{-3/2})$ as compared to the $O(n^{-1/2})$ derived by Yao (1987) for the upper bound $\mathrm{pr}(\hat{\nu}_n \neq \hat{\nu}_\infty)$.

5. THE ALGORITHMIC PROCEDURE

In this section, we illustrate how to compute the upper and lower bounds in (4·2) and (4·3). Our algorithm as such may be applied to any distribution that satisfies the required assumptions. While Theorem 3 provides upper and lower bounds for the asymptotic probabilities, we may incorporate (3·2) and (3·3) of Proposition 1 in Theorem 3 and obtain two different approximations for the asymptotic probabilities.

Since (4·2) suggests that $A_j(\alpha_1)\leqslant \mathrm{pr}(\hat{\nu}_\infty-\nu=j)\leqslant A_j(\alpha_2)$ when $0\leqslant\alpha_2\leqslant\alpha_1\leqslant 1$, we are motivated to approximate $\mathrm{pr}(\hat{\nu}_\infty-\nu=j)$ by $A_j(\alpha)$ for a suitable choice of α.

From (3·2) we have $\alpha_2\leqslant m\vartheta\leqslant\alpha_1$. Choosing $\alpha=m\vartheta$ and similarly $\alpha^*=m^*\vartheta^*$, we obtain our first approximation:

$$\mathrm{pr}(\hat{\nu}_\infty-\nu=j)\doteq A_j(m\vartheta), \quad \mathrm{pr}(\hat{\nu}_\infty-\nu=-j)\doteq A_j^*(m^*\vartheta^*). \tag{5·1}$$

Next, from (3.3) we have $\alpha_2\leqslant p_M\leqslant\alpha_1$. This time we let $\alpha=p_M$ and $\alpha^*=p^*_{M^*}$, and we obtain our second approximation:

$$\mathrm{pr}(\hat{\nu}_\infty-\nu=j)\doteq A_j(p_M), \quad \mathrm{pr}(\hat{\nu}_\infty-\nu=-j)\doteq A_j^*(p^*_{M^*}). \tag{5·2}$$

While there is no guarantee that either of the approximations should perform well, subsequent examples show that they both perform extremely well.

Evaluation of the bounds in (4·2) and (4·3) requires computation of $B(1)$, α_1, α_2, ϑ, $\{q_j\}$,

$\{\tilde{u}_j(\vartheta^*)\}$ and their '*' counterparts. We also need to compute m, p_M and m^*, $p^*_{M^*}$ in order to compute the two approximations given by (5·1) and (5·2).

Among the above, we only need to comment on the computation of $\{q_j\}$ and $\{\tilde{u}_j(\vartheta^*)\}$. The computation of $\{q_j\}$ is facilitated by implementing the iterative procedure given by (3·1). Similarly, $\{\tilde{u}_j(\vartheta^*)\}$ may also be computed through procedure (4·1). Similar iterative procedures may be applied to compute $\{q^*_j\}$ and $\{\tilde{u}^*_j(\vartheta)\}$.

In the algorithm below, where there is no confusion, we omit the relevant '*' counterparts.

ALGORITHM

Step S_0. *For specified* λ_0 *and* λ_1, *let* $Y \sim f(., \lambda_0)$ *and* $Y^* \sim f(., \lambda_1)$.

Step S_1. *Derive the probability density functions of* X_1 *and* X^*_1, *where*

$$X_1 = -\log \frac{f(Y;\lambda_0)}{f(Y;\lambda_1)}, \quad X^*_1 = \log \frac{f(Y^*;\lambda_0)}{f(Y^*;\lambda_1)}.$$

Step S_2. *Compute* $\{b_n\}$, $B(1)$ *and* $\{b^*_n\}$, $B^*(1)$: $b_n = \mathrm{pr}(S_n > 0)$ $(n = 1, 2, \ldots)$ *and* $B(1) = \sum b_n/n$.

Step S_3. *Find* ϑ, ϑ^*: $\vartheta = \sup\{\theta \in \mathbb{R} : \phi(\theta) \leqslant 1\}$.

Step S_4. *Compute* $\tilde{b}_n(\vartheta^*)$, $\tilde{b}^*_n(\vartheta)$: $\tilde{b}_n(\vartheta^*) = E\{e^{-\vartheta^* S_n} I(S_n > 0)\}$.

Step S_5. *Evaluate* $h(t)$, $h^*(t)$:

$$h(t) = \frac{\mathrm{pr}(X_1 > t)}{E[\exp\{\vartheta(X_1 - t)\} I(X_1 > t)]}.$$

Step S_6. *Find* α_1, α_2 *and* α^*_1, α^*_2: $\alpha_1 = \sup_{t \geqslant 0} h(t)$ *and* $\alpha_2 = \inf_{t \geqslant 0} h(t)$.

Step S_7. *Compute* $m = E(M)$, $m^* = E(M^*)$: $m = \sum E(S^+_n)/n$.

Step C_1. *Compute* $\{q_j\}$, $\{\tilde{u}_j(\vartheta^*)\}$ *and* $\{q^*_j\}$, $\{\tilde{u}^*_j(\vartheta)\}$: *implement the iterative procedures*

$$q_0 = 1, \quad nq_n = \sum_{j=0}^{n-1} b_{n-j} q_j; \quad \tilde{u}_0(\vartheta^*) = 1, \quad n\tilde{u}_n(\vartheta^*) = \sum_{j=0}^{n-1} \tilde{b}_{n-j}(\vartheta^*) \tilde{u}_j(\vartheta^*).$$

Step C_2. *For* $j = 1, 2, \ldots,$ *compute the lower and upper bounds* $L(j)$ *and* $U(j)$ *and the two approximations* $M_1(j)$ *and* $M_2(j)$: $L(j) = A_j(\alpha_1)$, $U(j) = A_j(\alpha_2)$; $M_1(j) = A_j(m\vartheta)$, $M_2(j) = A_j(p_M)$, *where* $A_j(\alpha) = e^{-B^*(1)}\{q^*_j - \alpha\tilde{u}^*_j(\vartheta)\}$.

Step C_3. *For* $-j = 1, 2, \ldots,$ *compute* $L(j)$, $U(j)$ *and* $M_1(j)$, $M_2(j)$: $L(j) = A^*_j(\alpha^*_1)$, $U(j) = A^*_j(\alpha^*_2)$; $M_1(j) = A^*_j(m^*\vartheta^*)$, $M_2(j) = A^*_j(p^*_{M^*})$.

Note that Steps S_0–S_7 depend on the specific $f(., \lambda)$. Once these steps are implemented, Steps C_1–C_3 are common to all cases.

6. EXAMPLES

6·1. *Normal distribution*

Step S_0. We let $Y \sim N(\lambda_0, \sigma^2)$ and $Y^* \sim N(\lambda_1, \sigma^2)$.

Step S_1. We find easily that $X_1 \sim N(-2\eta^2, 4\eta^2)$ and $X^*_1 \sim N(-2\eta^2, 4\eta^2)$, where $\eta = |\lambda_1 - \lambda_0|/(2\sigma)$. Upon rescaling, without loss of generality, we may assume that both X_1 and X^*_1 follow $N(\mu, 1)$, where $\mu = -\eta$. Since X_1 and X^*_1 are identically distributed, so are the random walks $\{S_n\}$ and $\{S^*_n\}$. Thus, in the steps below, we omit the computations for the '*' counterpart. It follows that the asymptotic probabilities for $\hat{\nu}_n$ are symmetric about the true changepoint ν.

Step S_2. We find $b_n = 1 - \Phi(-n^{\frac{1}{2}}\mu)$, for $n = 1, 2, \ldots$, where $\Phi(.)$ is the cumulative distribution function of the $N(0, 1)$ distribution.

Step S_3. Since $\phi(\theta) = e^{\mu\theta + \frac{1}{2}\theta^2}$, we find $\vartheta = -2\mu$.

Step S_4. We find $\tilde{b}_n(\vartheta^*) = e^{4n\mu^2}\{1 - \Phi(-3n^{\frac{1}{2}}\mu)\}$, for $n = 1, 2, \ldots$.

Step S_5. We obtain

$$h(t) = \frac{1 - \Phi(t - \mu)}{e^{2\mu t}\{1 - \Phi(t + \mu)\}} \quad (t \geqslant 0).$$

Step S_6. Since the above function $h(t)$, for $t \geqslant 0$, is monotone, we find that $\alpha_1 = 1$, which is obtained at $t = \infty$, and $\alpha_2 = \{1 - \Phi(-\mu)\}/\{1 - \Phi(\mu)\}$, which is obtained at $t = 0$.

Step S_7. We have $m = \sum_{n=1}^{\infty} \{(n/2\pi)^{\frac{1}{2}} \exp(-n\mu^2/2) + n\mu\bar{\Phi}(-n^{\frac{1}{2}}\mu)\}/n$.

It only remains to implement the common Steps C_1, C_2 and C_3 of the Algorithm. Table 1 presents the bounds $L(.)$ and $U(.)$ and the two approximations $M_1(.)$ and $M_2(.)$ for various values of δ, where we let $\delta = -\mu$. Also, for each δ, the 'Sum' in Table 1 represents the sum of all probabilities for each column while Table 1 contains the probabilities only for selected values of k. In view of symmetry, we present only the right half of the asymptotic distribution.

Table 1. *Asymptotic probabilities* $\mathrm{pr}(\hat{\nu}_\infty - \nu = \kappa)$ $(\kappa = 0, \pm 1, \pm 2, \ldots)$ *for the maximum likelihood estimate of the changepoint in the case of the normal distribution*

δ	κ	$L(.)$	$M_1(.)$	$M_2(.)$	$U(.)$	δ	κ	$L(.)$	$M_1(.)$	$M_2(.)$	$U(.)$
0·5	0	0·2802	0·2802	0·2802	0·2802	1·0	0	0·6409	0·6409	0·6049	0·6049
	1	0·0672	0·1122	0·1181	0·1204		1	0·0680	0·1121	0·1152	0·1159
	2	0·0468	0·0664	0·0689	0·0699		2	0·0262	0·0377	0·0385	0·0387
	3	0·0331	0·0445	0·0454	0·0459		3	0·0122	0·0153	0·0156	0·0156
	4	0·0242	0·0309	0·0318	0·0321		4	0·0051	0·0068	0·0069	0·0069
	5	0·0181	0·0226	0·0231	0·0234		5	0·0025	0·0032	0·0032	0·0033
	6	0·0138	0·0169	0·0173	0·0175		6	0·0012	0·0016	0·0016	0·0016
	7	0·0107	0·0129	0·0132	0·0133		7	0·0006	0·0008	0·0008	0·0008
	8	0·0084	0·0100	0·0102	0·0103		8	0·0003	0·0004	0·0004	0·0004
	9	0·0066	0·0079	0·0080	0·0081		9	0·0002	0·0002	0·0002	0·0002
	10	0·0053	0·0062	0·0064	0·0064		10	0·0001	0·0001	0·0001	0·0001
	15	0·0019	0·0022	0·0022	0·0022						
	20	0·0007	0·0008	0·0008	0·0008						
	25	0·0003	0·0003	0·0003	0·0003						
	Sum	0·7960	0·9951	1·0213	1·0317		Sum	0·8719	0·9974	1·0063	1·0081
1·5	0	0·8568	0·8568	0·8568	0·8568	2·0	0	0·9531	0·9531	0·9531	0·9531
	1	0·0364	0·0592	0·0599	0·0600		1	0·0136	0·0219	0·0220	0·0220
	2	0·0066	0·0096	0·0097	0·0097		2	0·0009	0·0014	0·0014	0·0014
	3	0·0014	0·0019	0·0020	0·0020		3	0·0001	0·0001	0·0001	0·0001
	4	0·0003	0·0004	0·0004	0·0004						
	5	0·0001	0·0001	0·0001	0·0001						
	Sum	0·9463	0·9994	1·0011	1·0013		Sum	0·9825	0·9999	1·0001	1·0001
2·5	0	0·9874	0·9874	0·9874	0·9874	3·0	0	0·9973	0·9973	0·9973	0·9973
	1	0·0039	0·0062	0·0062	0·0062		1	0·0009	0·0013	0·0013	0·0013
	2	0·0001	0·0001	0·0001	0·0001						
	Sum	0·9954	0·9999	1·0001	1·0001		Sum	0·9990	0·9999	1·0000	1·0000

6·2. *Exponential distribution*

The case of the exponential distribution has not been dealt with in the literature thus far.

Step S_0. We let $Y \sim \mathrm{Ex}(1/\lambda_0)$ and $Y^* \sim \mathrm{Ex}(1/\lambda_1)$. Without loss of generality, we assume $\lambda_1 > \lambda_0$.

Step S_1. Let $f_{X_1}(.)$ and $f_{X_1^*}(.)$ be the probability density functions of X_1 and X_1^* respectively. Then

$$f_{X_1}(x) = \delta e^{-\delta(x+d)}, \quad x \in [-d, \infty), \quad f_{X_1^*}(x^*) = (\delta - 1)e^{-(\delta-1)(d-x^*)}, \quad x^* \in (-\infty, d],$$

where $\delta = \lambda_1/(\lambda_1 - \lambda_0) > 1$ and $d = \log\{\delta/(\delta - 1)\}$.

Step S_2. We have $b_n = \mathrm{pr}(\chi^2_{2n} > 2\delta nd)$, $b_n^* = 1 - \mathrm{pr}\{\chi^2_{2n} > 2n(\delta - 1)d\}$.

Step S_3. We find

$$\phi(\theta) = \frac{\delta}{\delta - \theta}\left(1 - \frac{1}{\delta}\right)^{\theta}.$$

Thus, $\phi(\vartheta) = 1$ yields $\vartheta = 1$. Also,

$$\phi^*(\theta^*) = e^{\theta^* d}\,\frac{(\delta - 1)}{\delta - 1 + \theta^*}$$

and again $\vartheta^* = 1$.

Step S_4. We have

$$\tilde{b}_n(\vartheta^*) = \left(\frac{\delta^2}{\delta^2 - 1}\right)^n \mathrm{pr}\{\chi^2_{2n} > 2n(\delta + 1)d\},$$

$$\tilde{b}_n^*(\vartheta) = \begin{cases} \left\{\dfrac{(\delta-1)^2}{\delta(\delta-2)}\right\}^n \mathrm{pr}\{\chi^2_{2n} < 2(\delta - 2)nd\} & (\delta > 2), \\[2ex] \dfrac{n^n(\log 2)^n}{2^n n!} & (\delta = 2), \\[2ex] \left\{\dfrac{(\delta-1)^2}{\delta(2-\delta)}\right\}^n (-1)^n\left[1 - e^{(2-\delta)nd}\displaystyle\sum_{j=0}^{n-1}\frac{\{(\delta-2)nd\}^j}{j!}\right] & (1 < \delta < 2). \end{cases}$$

In the above, note the singularity in $\tilde{b}_n^*(\vartheta)$ at $\delta = 2$.

Step S_5. We find

$$h(t) = \frac{\delta - 1}{\delta} \quad (t \geqslant 0), \quad h^*(t^*) = \left(\frac{\delta - 1}{\delta}\right)\frac{1 - \exp\{-(\delta - 1)(d - t^*)\}}{\exp(d - t^*) - \exp\{-(\delta - 1)(d - t^*)\}}.$$

It is somewhat surprising that $h(t)$ is independent of t.

Step S_6. We find $\alpha_1 = \alpha_2 = (\delta - 1)/\delta$. Also, $\alpha_1^* = 1$, which is attained at $t^* = d$; and

$$\alpha_2^* = \frac{\delta\{\delta^{\delta-1} - (\delta - 1)^{\delta-1}\}}{\delta^\delta - (\delta - 1)^\delta},$$

which is attained at $t^* = 0$.

Step S_7. We have

$$m = -d \sum_{n=1}^{\infty} \mathrm{pr}(\chi^2_{2n} > 2n\delta d) + \frac{1}{\delta} \sum_{n=1}^{\infty} \mathrm{pr}(\chi^2_{2(n+1)} > 2n\delta d) = \frac{\delta - 1}{\delta}.$$

$$m^* = d \sum_{n=1}^{\infty} \mathrm{pr}\{\chi^2_{2n} < 2n(\delta - 1)d\} - \frac{1}{\delta - 1} \sum_{n=1}^{\infty} \mathrm{pr}\{\chi^2_{2(n+1)} < 2n(\delta - 1)d\}.$$

The common Steps C_1, C_2 and C_3 of the Algorithm may now be implemented. The computations for selected values of δ are summarised in Table 2.

Table 2. *Asymptotic probabilities* $\mathrm{pr}(\hat{\nu}_\infty - \nu = \kappa)$ $(\kappa = 0, \pm 1, \pm 2, \ldots)$ *for the maximum likelihood estimate of the changepoint in the case of the exponential distribution*

δ	κ	$L(.)$	$M_1(.)$	$M_2(.)$	$U(.)$	δ	κ	$L(.)$	$M_1(.)$	$M_2(.)$	$U(.)$
2·4	−20	0·0041	0·0044	0·0046	0·0046	2·0	−20	0·0025	0·0027	0·0028	0·0029
	−10	0·0093	0·0103	0·0108	0·0110		−10	0·0076	0·0085	0·0089	0·0091
	−5	0·0169	0·0197	0·0209	0·0213		−5	0·0167	0·0197	0·0209	0·0213
	−4	0·0196	0·0233	0·0248	0·0254		−4	0·0202	0·0243	0·0260	0·0265
	−3	0·0231	0·0281	0·0303	0·0310		−3	0·0250	0·0309	0·0333	0·0340
	−2	0·0277	0·0353	0·0385	0·0396		−2	0·0318	0·0411	0·0449	0·0461
	−1	0·0336	0·0475	0·0534	0·0554		−1	0·0417	0·0598	0·0672	0·0694
	0	0·1023	0·1023	0·1023	0·1023		0	0·1534	0·1534	0·1534	0·1534
	1	0·0734	0·0734	0·0734	0·0734		1	0·1002	0·1002	0·1002	0·1002
	2	0·0570	0·0570	0·0570	0·0570		2	0·0726	0·0726	0·0726	0·0726
	3	0·0462	0·0462	0·0462	0·0462		3	0·0556	0·0556	0·0556	0·0556
	4	0·0385	0·0385	0·0385	0·0385		4	0·0441	0·0441	0·0441	0·0441
	5	0·0327	0·0327	0·0327	0·0327		5	0·0358	0·0358	0·0358	0·0358
	10	0·0170	0·0170	0·0170	0·0170		10	0·0154	0·0154	0·0154	0·0154
	20	0·0072	0·0072	0·0072	0·0072		20	0·0049	0·0049	0·0049	0·0049
	Sum	0·9487	0·9989	1·0203	1·0275		Sum	0·9431	0·9974	1·0198	1·0264
1·4	−20	0·0001	0·0002	0·0002	0·0002						
	−10	0·0016	0·0019	0·0020	0·0021						
	−8	0·0020	0·0037	0·0039	0·0039						
	−7	0·0042	0·0051	0·0054	0·0055						
	−6	0·0059	0·0072	0·0077	0·0078						
	−5	0·0084	0·0104	0·0111	0·0113						
	−4	0·0123	0·0156	0·0167	0·0169	1·05	−4	0·0002	0·0003	0·0003	0·0003
	−3	0·0187	0·0243	0·0262	0·0266		−3	0·0008	0·0013	0·0014	0·0014
	−2	0·0298	0·0406	0·0443	0·0450		−2	0·0037	0·0058	0·0062	0·0062
	−1	0·0515	0·0783	0·0875	0·0893		−1	0·0190	0·0335	0·0359	0·0360
	0	0·3564	0·3564	0·3564	0·3564		0	0·8074	0·8074	0·8074	0·8074
	1	0·1662	0·1662	0·1662	0·1662		1	0·1180	0·1180	0·1180	0·1180
	2	0·0932	0·0932	0·0932	0·0932		2	0·0243	0·0243	0·0243	0·0243
	3	0·0573	0·0573	0·0573	0·0573		3	0·0058	0·0058	0·0058	0·0058
	4	0·0373	0·0373	0·0373	0·0373		4	0·0015	0·0015	0·0015	0·0015
	5	0·0252	0·0252	0·0252	0·0252		5	0·0004	0·0004	0·0004	0·0004
	6	0·0175	0·0175	0·0175	0·0175		6	0·0001	0·0001	0·0001	0·0001
	7	0·0124	0·0124	0·0124	0·0124						
	8	0·0089	0·0089	0·0089	0·0089						
	10	0·0048	0·0048	0·0048	0·0048						
	20	0·0004	0·0004	0·0004	0·0004						
	Sum	0·9433	0·9963	1·0146	1·0180		Sum	0·9814	0·9987	1·0016	1·0017

6·3. *Discussion of the examples*

In summarising the information in Tables 1 and 2, we concentrate mainly on the evidence provided by the 'Sum' of all probabilities.

Clearly, the upper bound is quite sharp in both cases, compared to the lower bound. Generally, the approximations $M_1(.)$ and $M_2(.)$ provide very good accuracy. For the normal distribution, Table 1 shows that the probabilities based on $M_1(.)$ are extremely accurate, while those based on $M_2(.)$ show a tendency slightly to overestimate the true probabilities, mostly for small changes. When δ is large ($\delta \geqslant 2$), the approximations $M_1(.)$, $M_2(.)$ and the bounds $L(.)$ and $U(.)$ all perform extremely well. Finally, note that the values given by $M_1(.)$ tally very well with those of Hinkley (1970).

In the exponential case, note that by definition $\delta > 1$. While the asymptotic probabilities in Table 2 are not symmetric in this case, note that generally the probability distributions have longer tails, which is especially true for large values of δ ($\delta \geqslant 2$). While both approximations are extremely accurate, $M_1(.)$ is even better than $M_2(.)$.

Furthermore, we highlight a result which is pleasing and noteworthy. As a consequence of $\alpha_1 = \alpha_2$, in the exponential case, the right tail probabilities for both the lower and upper bounds and the two approximations were identical. Thus, subject to minor computational errors, these probabilities are exact. One would not have anticipated this.

7. British coal mine explosions data

The 109 time intervals, in days, between successive explosions in British coal mines between the years 1875 and 1950 were analysed by Maguire et al. (1952). Jarrett (1979) presented an extended dataset covering the years 1851 to 1962, giving 190 data points, and corrected some errors in the data given by Maguire et al. (1952). Maguire et al. (1952) concluded that the time intervals between explosions followed an exponential distribution with a constant mean over time. Cox & Lewis (1966, Ch. 23) reanalysed the data and found strong evidence that the mean did not remain constant in time. Assuming exponential distributions for the time intervals, Worsley (1986) used the likelihood ratio statistic for the data of Maguire et al. (1952) to test for a change in mean at an unknown point with λ_0, the mean before the change, and λ_1, the mean after the change, both unknown. The test was highly significant and he found the maximum likelihood estimate of the changepoint to be $\tilde{v}_{109} = 46$, which corresponds to the year 1890. Based on the distribution of the likelihood ratio statistic, Worsley (1986) constructed a 95% confidence region of $\{36, \ldots, 53\}$ for the unknown changepoint. For the extended dataset of Jarrett (1979), his results were very similar. The estimate of the changepoint year and the ends of the confidence interval all moved just one year later, with $\hat{v}_{190} = 124$ and confidence region $\{116, \ldots, 133\}$. He also applied the conditional solution of Cobb (1978) to the original data of Maguire et al. (1952) and found the corresponding 95% confidence region to be $\{26, \ldots, 39, 41, \ldots, 53\}$.

Here, we apply our asymptotic analysis under the exponential distribution to Jarrett's (1979) extended data. We consider the dataset to be large enough for our asymptotics to hold and thus we let $\tilde{v}_{190} = \tilde{v}_\infty = \tilde{v}$. We obtain $\tilde{v} = 124$, which matches Worsley's result. Further, the conditional estimates of the means are $\hat{\lambda}_{0|124} = 118$ and $\hat{\lambda}_{1|124} = 403$. These estimates are consistent so we may take $\lambda_0 = 118$ and $\lambda_1 = 403$ as known values. This gives $\delta = 1{\cdot}41$ and the corresponding maximum likelihood estimate for the known-means case is $\hat{v} = 124$. Then Table 2, with $\delta = 1{\cdot}40$, provides the probability distribution of $\tilde{v}$. For example, using $M_1(.)$ as the approximate probability distribution of $\tilde{v}$, we find that

$\text{pr}(\tilde{\nu} - \nu = 0) = 0{\cdot}3564$. Furthermore, the shortest 95% confidence region for ν is found to be $\{118, \ldots, 132\}$. For comparison, Cobb's conditional region for the extended data is $\{116, \ldots, 129, 133\}$. The size of our confidence region matches exactly that of Cobb's. However, our unconditional confidence region is biased to the right, whereas Cobb's conditional confidence region is biased to the left. Worsley's region is larger in size by three additional points when compared to both ours and Cobb's.

In summary, our unconditional solution seems as good as the conditional solution of Cobb (1978). However, our solution can be applied more routinely than Cobb's.

ACKNOWLEDGEMENT

The authors thank the editor and an anonymous referee for their insightful comments which have led to an improved presentation of the paper.

REFERENCES

ASMUSSEN, S. (1985). *Applied Probability and Queues.* New York: John Wiley.

BETROIN, J. & DONEY, R. A. (1994). On conditioning a random walk to stay non-negative. *Ann. Prob.* **22**, 2151–67.

BETROIN, J. & DONEY, R. A. (1996). Some asymptotic results for transient random walks. *Adv. Appl. Prob.* **28**, 207–26.

BHATTACHARYA, P. K. (1987). Maximum likelihood estimation of change-point in the distribution of independent random variables: general multiparameter case. *J. Mult. Anal.* **23**, 183–208.

BHATTACHARYA, P. K. (1994). Some aspects of change-point analysis. In *Change-point Problems,* IMS Monograph 23, Ed. E. Carlstein, H. G. Muller and D. Siegmund, pp. 28–56. Hayward, CA: Institute of Mathematical Statistics.

BHATTACHARYA, P. K. & BROCKWELL, P. J. (1976). The minimum of an additive process with applications to signal estimation and storage theory. *Z. Wahr. Verw. Geb.* **37**, 51–75.

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

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

DOWNHAM, S. & FOTOPOULOS, S. (1981). Some inequalities on the distribution of ladder epochs. *J. Appl. Prob.* **18**, 770–5.

EMBRECHTS, P. & HAWKES, P. (1982). A limit theory for the tails of discrete infinitely divisible laws with applications to fluctuation theory. *J. Aust. Math. Soc.* A **32**, 412–22.

FELLER, W. (1971). *An Introduction to Probability Theory and its Applications,* **2**. New York: John Wiley.

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

HINKLEY, D. V. (1971). Inference in two phase regression. *J. Am. Statist. Assoc.* **55**, 736–43.

HINKLEY, D. V. (1972). Time order classification. *Biometrika* **59**, 509–23.

IBRAGIMOV, I. A. & HASMINSKI, R. Z. (1981). *Statistical Estimation: Asymptotic Theory.* New York: Springer-Verlag.

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

MAGUIRE, B. A., PEARSON, E. S. & WYNN, A. H. A. (1952). The time intervals between industrial accidents. *Biometrika* **38**, 168–80.

RUKHIN, A. L. (1994). Asymptotic minimaxity in the change-point problem. In *Change-point Problems,* IMS Monograph 23, Ed. E. Carlstein, H. G. Muller and D. Siegmund, pp. 284–91. Hayward, CA: Institute of Mathematical Statistics.

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

SPITZER, F. (1976). *Principles of Random Walks.* New York: Springer-Verlag.

STOYAN, D. (1976). *Comparison Methods for Queues and Other Stochastic Models.* New York: John Wiley.

VERAVERBEKE, N. & TEUGELS, J. T. (1975). The exponential rate of convergence of the distribution of the maximum of a random walk. *J. Appl. Prob.* **12**, 279–88.

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

YAO, Y. C. (1987). Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. *Ann Statist.* **15**, 1321–8.

[*Received August* 1997. *Revised April* 1998]