Change-Point Problems
IMS Lecture Notes – Monograph Series (Volume 23, 1994)

ON ALMOST SURE BEHAVIOR OF CHANGE-POINT ESTIMATORS

BY YI-CHING YAO, DAWEI HUANG, AND RICHARD DAVIS

Colorado State University, Queensland University of Technology, and Colorado State University

We propose a natural setting for the simple change-point problem which is particularly useful for studying almost sure convergence properties of change-point estimators. It is shown that each member of a class of nonparametric estimators is within $O(1)$ of the true change-point almost surely if the smaller of the pre-change and post-change sample sizes is greater than a constant times $\log n$ where n is the total sample size. This improves upon the known result on nonparametric estimators which gives $O(n^a)$ almost sure rate of convergence with $a > 1/2$ under the restrictive assumption that the pre-change and post-change sample sizes are of the same order of magnitude. We also consider the simple normal mean shift model and show that the finite-sample maximum likelihood estimator converges almost surely to the infinite-sample counterpart if the smaller of the pre-change and post-change sample sizes is greater than a constant times $\log n$. It is found that the best possible constant equals $2\delta^{-2}\sigma^2$ where δ is the amount of change in mean and σ^2 is the common variance.

1. Introduction. Let P and Q be two different unknown probability distributions on a measurable space E and consider the following array of random variables. For $n = 2, 3, \ldots$, let $X_{1,n}, X_{2,n}, \ldots, X_{n,n}$ be independent random variables such that $X_{i,n}$ has distribution P or Q according to whether $i \le \tau_n$ or $i > \tau_n$ where $\tau_n (1 \le \tau_n < n)$ is the unknown time point (change-point) at which the distribution of the random variables changes from P to Q. In a parametric setting, Hinkley (1970) showed that the maximum likelihood estimator (mle) $\hat{\tau}_n$ of τ_n satisfies $\hat{\tau}_n - \tau_n = O_p(1)$. Carlstein (1988) proposed a class of nonparametric estimators of τ_n (denoted by τ_n^*) and established (without specifying the probabilistic relation between the rows of the array) that $\tau_n^* - \tau_n = O(n^a)$ a.s. for any $a > 1/2$, assuming that $\tau_n/n \to \theta \in (0,1)$. Recently, Dümbgen (1991) considered a more general framework and proposed a class of nonparametric estimators (including Carlstein's estimators as special

---

AMS 1991 Subject Classification: Primary 62G05, Secondary 60F15.

Key words and phrases: Change-point problem, nonparametric estimation.

cases) and showed that $\tau_n^* - \tau_n = O_p(1)$, again assuming $\tau_n/n \to \theta \in (0,1)$. (Dümbgen also obtained weak convergence results for the case that $P = P_n$ and $Q = Q_n$ depend on the sample size n and approach a common distribution as $n \to \infty$.)

By embedding naturally the rows of the array into an infinite sequence of independent random variables, we show in this note that $\tau_n^* - \tau_n = O(1)$ a.s. as long as $\liminf_{n\to\infty}[\tau_n \wedge (n - \tau_n)]/\log n > K$ (for some large constant K) where $a \wedge b := \min\{a, b\}$, and τ_n^* is any one of Dümbgen's estimators. Specifically, consider a two-sided infinite sequence of independent random variables $\{\ldots, X_{-1}, X_0, X_1, \ldots\}$ such that X_i has distribution P or Q according as $i \le 0$ or $i > 0$. For each pair of positive integers (k, m), consider the sample $X_{-k+1}, \ldots, X_m$ (but the values of k and m are not known). This sample is equivalent to the sample $X_{1,n}, \ldots, X_{n,n}$ with $n = k + m$ and $\tau_n = k$. In other words, the rows $\{X_{1,n}, \ldots, X_{n,n}\}, n = 2, 3, \ldots$ are embedded into the infinite sequence $\{\ldots, X_{-1}, X_0, X_1, \ldots\}$ by aligning the rows with respect to the sequence of reference points $\{\tau_n\}$. For convenience, we shall treat k and m as functions of the sample size n, i.e. $k = k(n), m = m(n)$ and $n = k + m$.

THEOREM 1. *There exists a $K > 0$ (depending only on C_0 and $\mathcal{D}$ given below) such that if*

$$\liminf_{n\to\infty}(k \wedge m)/\log n > K, \tag{1.1}$$

then $\tau_n^* = O(1)$ a.s. where $\tau_n^* = \arg\max\ \{N_n(D_n^j) : -k < j < m\}$ and N_n is a seminorm satisfying condition (A) below and defined on the space $\mathcal{M}$ of all finite signed measures on E and $D_n^j := D^j_{(k(n),m(n))} := w((j+k)/n)(Q_n^j - P_n^j)$ with $w(x) = [x(1-x)]^{1/2}, P_n^j := \Sigma_{i=-k+1}^{j}\delta_{X_i}/(j+k)$ and $Q_n^j := \Sigma_{i=j+1}^{m}\delta_{X_i}/(m-j)$ (empirical distributions).

Condition (A): (i) There is a Vapnik-Cervonenkis class $\mathcal{D}$ of measurable subsets of E such that

$$N_n(\nu) \le \|\nu\| := \sup\{|\nu(D)| : D \in \mathcal{D}\} \qquad \text{for all } \nu \in \mathcal{M}. \tag{1.2}$$

(ii) There is a constant $C_0 > 0$ such that

$$\liminf_{n\to\infty} N_n(Q - P) > C_0 \qquad \text{a.s.} \tag{1.3}$$

The reader may consult Pollard (1984) for the definition and properties of a Vapnik-Cervonenkis class. Note that if $N_n = \|\cdot\|$ with $\|\cdot\|$ as defined in (i), then (ii) is automatically satisfied. We refer the reader to Dümbgen (1991) for more details concerning this class of nonparametric estimators and various special cases including Carlstein's (1988) and Darkhovskhi's (1976) estimators.

The proof of Theorem 1 is given in Section 3. In Section 2, a similar result is given for the mle of the change-point in the normal mean shift model. More precisely, it is shown that the mle converges a.s. to a random variable which is the mle based on the entire (infinite) sample if $\liminf_{n\to\infty}(k \wedge m)/\log n > 2\delta^{-2}\sigma^2$ where σ^2 is the (common) variance and δ is the amount of change in mean (See Theorem 2 below). Finally, some concluding remarks are contained in Section 4.

2. The Normal Mean Shift Model. In this section, we prove the following theorem concerning the behavior of the mle $\hat{\tau}_n$ of the change-point (which is 0) based on the sample $X_i, i = -k(n)+1, \ldots, m(n)$ under the assumption that $P = N(\mu, \sigma^2)$ and $Q = N(\mu+\delta, \sigma^2)$ with μ and $\delta \neq 0$ unknown but σ^2 known.

THEOREM 2. *If there exist an $\epsilon > 0$ and an increasing subsequence $\{n(l) : l = 1, 2, \ldots\}$ such that $\limsup_{l\to\infty} k(n(l))/\log n(l) < 2(1-\epsilon)\delta^{-2}\sigma^2$ and $n(l+1) - n(l) = O(\{n(l)\}^\epsilon)$ as $l \to \infty$, then $\hat{\tau}_n \neq O(1)$ a.s. If $\liminf_{n\to\infty}(k \wedge m)/\log n > 2\delta^{-2}\sigma^2$, then $\hat{\tau}_n \to \hat{\tau}_\infty$ a.s. where $\hat{\tau}_\infty$ is the mle of the change-point based on the entire (infinite) sample with μ and δ known. In particular, $\hat{\tau}_n = O(1)$ a.s.*

Since we may rescale the X_i to have variance 1, we shall assume, without loss of generality, that $\sigma^2 = 1$. For $-k < j < m$, let

$$L_n(j) := -\,[\Sigma_{i=-k+1}^{j}(X_i - \bar{X}_{-k,j})^2 + \Sigma_{i=j+1}^{m}(X_i - \bar{X}_{j,m})^2] \\ + [\Sigma_{i=-k+1}^{0}(X_i - \bar{X}_{-k,0})^2 + \Sigma_{i=1}^{m}(X_i - \bar{X}_{0,m})^2]$$

where $\bar{X}_{i,j} := \bar{X}_{j,i} := (X_{i+1} + \cdots + X_j)/(j-i)$ for $i < j$. Then $\hat{\tau}_n = \arg\max\{L_n(j) : -k < j < m\}$. Define $Z_i = X_i - \mu - \delta$ for all i, and $e_i = X_i - \mu$ for $i \le 0$ and $e_i = X_i - \mu - \delta$ for $i > 0$. Then the e_i are iid $N(0,1)$ and $Z_i = e_i - \delta 1_{\{i\le 0\}}$ where 1_S denotes the indicator of the set S. Now we have

$$\begin{aligned} L_n(j) &= (k+j)\bar{Z}_{-k,j}^2 + (m-j)\bar{Z}_{j,m}^2 - k\bar{Z}_{-k,0}^2 - m\bar{Z}_{0,m}^2 \\ &= (k+j)^{-1}[k\bar{e}_{-k,0} + j\bar{e}_{0,j} - (k+j^-)\delta]^2 + (m-j)^{-1}[m\bar{e}_{0,m} - j\bar{e}_{0,j} + j^-\delta]^2 \\ &\quad - k\bar{e}_{-k,0}^2 - m\bar{e}_{0,m}^2 + 2k\delta\bar{e}_{-k,0} - k\delta^2 \end{aligned} \tag{2.1}$$

where $x^- := x1_{\{x<0\}}$. So,

$$\begin{aligned} L_n(j) &= -kj(k+j)^{-1}(\delta^2 + \bar{e}_{-k,0}^2 - 2\delta\bar{e}_{-k,0} + 2\delta\bar{e}_{0,j} - 2\bar{e}_{-k,0}\bar{e}_{0,j}) && (2.2)\\ &\quad + j^2(k+j)^{-1}\bar{e}_{0,j}^2 + (m-j)\bar{e}_{j,m}^2 - m\bar{e}_{0,m}^2 \qquad \text{for } j > 0, \\ L_n(j) &= jm(m-j)^{-1}(\delta^2 + \bar{e}_{0,m}^2 + 2\delta\bar{e}_{0,m} - 2\delta\bar{e}_{j,0} - 2\bar{e}_{0,m}\bar{e}_{j,0}) && (2.3)\\ &\quad + j^2(m-j)^{-1}\bar{e}_{j,0}^2 + (k+j)\bar{e}_{-k,j}^2 - k\bar{e}_{-k,0}^2 \qquad \text{for } j \le 0. \end{aligned}$$

LEMMA 1. *For fixed $\epsilon > 0$, with probability 1,*

$$\max_{0 \le i < n} (n-i)\bar{e}_{i,n}^2 < (2+\epsilon)\log n \qquad \textit{for large } n.$$

PROOF. Using a boundary crossing argument for standard Brownian motion $\{W_t\}$ along the lines of the proof of Lemma 3.1 of Venkatraman (1992), it can be shown that for $b > 0$,

$$\Pr(\max_{1 \le t \le n} W_t/t^{1/2} \ge b) \le 2(1-\Phi(b)) + (\pi/2)^{1/2}\phi(b) + 2^{-1}b\phi(b)\log n \quad (2.4)$$

where Φ and ϕ denote the standard normal distribution and density, respectively. It follows that $\Pr(\max_{0 \le i < n}(n-i)\bar{e}_{i,n}^2 \ge (2+\epsilon)\log n)$ is bounded by twice the right hand side of (2.4) with $b = [(2+\epsilon)\log n]^{1/2}$, which along with the Borel-Cantelli Lemma yields the desired result.

LEMMA 2. *If there exist an $\epsilon > 0$ and an increasing subsequence $\{n(l) : l = 1, 2, \ldots\}$ such that $\limsup_{l\to\infty} k(n(l))/\log n(l) < 2(1-\epsilon)\delta^{-2}$ and $n(l+1) - n(l) = O(\{n(l)\}^\epsilon)$ as $l \to \infty$, then $\hat{\tau}_n \ne O(1)$ a.s.*

PROOF. We first assume that $k(n(l)) \to \infty$ as $l \to \infty$. Choose $\epsilon_1 > \epsilon_2 > \epsilon$ and $\epsilon_3 > 0$ such that

$$k(n(l)) < 2(1-\epsilon_1)\delta^{-2}\log n(l) \ \text{ for large } l$$

and

$$-2(1-\epsilon_1)(1+\epsilon_3)/(1-\epsilon_2) + 2 - \epsilon_3 > 0.$$

By (2.2), w.p.1, for large n,

$$L_n(m-1) > -k(1+\epsilon_3)\delta^2 + e_m^2 - O(\log\log n).$$

For each $r = 1, 2, \ldots$, let $l(r) := \max\{l : n(l) \le r^{1/(1-\epsilon_2)}\}$. Then for large r, $n(l(r+1)) - n(l(r)) \ge (n(l(r+1)))^\epsilon$. To see this, note that

$$n(l(r)) \le r^{1/(1-\epsilon_2)} < n(l(r)+1),$$

$$n(l(r+1)) \le (r+1)^{1/(1-\epsilon_2)} < n(l(r+1)+1),$$

and so for large r,

$$\begin{aligned} n(l(r+1)) - n(l(r)) &\ge \{(r+1)^{1/(1-\epsilon_2)} - r^{1/(1-\epsilon_2)}\} \\ &\quad - \{n(l(r+1)+1) - n(l(r+1))\} \\ &\ge (1-\epsilon_2)^{-1} r^{\epsilon_2/(1-\epsilon_2)} - C(n(l(r+1)))^\epsilon \\ &\ge (n(l(r+1)))^\epsilon \end{aligned}$$

where $C > 0$ is a constant. Since $k + m = n$ and $k(n(l)) < 2(1-\epsilon_1)\delta^{-2}\log n(l)$ (for large l), $m(n(l(r)))$ must be all distinct for large r. Write $n_1(r) = n(l(r)), k_1(r) = k(n_1(r))$ and $m_1(r) = m(n_1(r))$. By the Borel-Cantelli Lemma and a simple bound on the standard normal tail probability, it follows that $\Pr(e^2_{m_1(r)} > (2-\epsilon_3)\log r \text{ i.o.}) = 1$. Fix a sample path for which $e^2_{m_1(r)} > (2-\epsilon_3)\log r$ i.o. Suppose $\hat{\tau}_n = O(1)$. Choose a $C_1 > 0$ so that $|\hat{\tau}_n| \leq C_1$. Then $\max_j L_n(j) = \max_{|j|\leq C_1} L_n(j)$ is bounded from above. But the following inequalities occur infinitely often:

$$\begin{aligned} L_{n_1}(m_1 - 1) &> -k_1(1+\epsilon_3)\delta^2 + (2-\epsilon_3)\log r - O(\log\log r^{1/(1-\epsilon_2)}) \\ &> -2(1-\epsilon_1)(1+\epsilon_3)\log r^{1/(1-\epsilon_2)} + (2-\epsilon_3)\log r - O(\log\log r) \\ &= \{-2(1-\epsilon_1)(1+\epsilon_3)/(1-\epsilon_2) + 2 - \epsilon_3\}\log r - O(\log\log r) \\ &> \epsilon_4 \log r \end{aligned}$$

for some $\epsilon_4 > 0$. This contradiction completes the proof for the case that $\lim_{l\to\infty} k(n(l)) = \infty$. In case k is bounded on some sub-subsequence, it is easy to show that along this sub-subsequence, $\hat{\tau}_n \neq O(1)$ a.s., completing the proof of Lemma 2.

REMARK. It might be tempting to conjecture that $\hat{\tau}_n \neq O(1)$ a.s. whenever $\liminf_{n\to\infty}(k \wedge m)/\log n < 2\delta^{-2}$ (assuming $\sigma^2 = 1$). This is not true as the following example shows. Fix $0 < \epsilon_1 < \epsilon < 1$ and set $\epsilon_2 = \epsilon/(1-\epsilon)$. Consider the subsequence $n(l) = [l^{1+\epsilon_2}]$ and $k(n(l)) = [2(1-\epsilon_1)\delta^{-2}\log n(l)]$. Note that $n(l+1) - n(l) = O((n(l))^{\epsilon})$. By (2.4),

$$\Sigma_{l=1}^{\infty}\Pr\{\max_{0\leq j<m(l)}(m-j)\bar{e}^2_{j,m} > 2(1-\epsilon_3)\log m(l)\} < \infty$$

for any $\epsilon_3 < \epsilon$. By the Borel-Cantelli Lemma, w.p.1, for large l

$$\max_{0\leq j<m(l)}(m-j)\bar{e}^2_{j,m} \leq 2(1-\epsilon_3)\log m(l).$$

With this fact, it can be shown, along the lines of the proof of Lemma 3 below, that $\hat{\tau}_{n(l)} = O(1)$ a.s.

LEMMA 3. *If* $\liminf_{n\to\infty}(k\wedge m)/\log n > 2\delta^{-2}$, *then* $\hat{\tau}_n = O(1)$ a.s.

PROOF. Since we may consider separately the two subsequences $\{n'\}$ and $\{n''\}$ with $k(n') \leq n'/2$ and $k(n'') > n''/2$, we shall assume without loss of generality that $k(n) \leq n/2$ for all n.

Choose $0 < \epsilon < 1$ in such a way that $k > (2+\epsilon)\delta^{-2}\log n$ for large n. Since $L_n(0) = 0$, it suffices to show that w.p.1 there exists a positive integer d (depending on the sample path) such that for large n $L_n(j) < 0$ for all j with $-k < j < m$ and $|j| > d$. We break up the proof into several cases.

Case 1: $C_2 \log n \le j < m$ with $C_2 = \epsilon^{-1}\delta^{-2}(4+\epsilon)(2+\epsilon/2)$. By (2.2) and the law of the iterated logarithm, w.p.1 for $C_2 \log n \le j < m$ and for large n,

$$L_n(j) = -kj(k+j)^{-1}(\delta^2 + o(1)) + O(\log\log n) + (m-j)\bar{e}_{j,m}^2 + O(\log\log n).$$

By Lemma 1, w.p.1 $(m-j)\bar{e}_{j,m}^2 < (2+\epsilon/5)\log n$ for $C_2 \log n \le j < m$ and for large n. But,

$$\begin{aligned} kj(k+j)^{-1}\delta^2 \ge & kC_2 \log n(k + C_2 \log n)^{-1}\delta^2 \\ \ge & \begin{cases} \epsilon(4+\epsilon)^{-1}C_2(\log n)\delta^2, & \text{if } k \ge 4^{-1}\epsilon C_2 \log n, \\ 4(4+\epsilon)^{-1}k\delta^2, & \text{if } k < 4^{-1}\epsilon C_2 \log n. \end{cases} \end{aligned}$$

In either case, we get $kj(k+j)^{-1}\delta^2 \ge (2+\epsilon/4)\log n$. This proves that w.p.1 $L_n(j) < 0$ for $C_2 \log n \le j < m$ and for large n.

Case 2: $0 < j < C_2 \log n$. Note that $k(k+j)^{-1} \ge C_3$ for large n where $C_3 = \delta^{-2}(2+\epsilon)/\{\delta^{-2}(2+\epsilon) + C_2\}$. By (2.2), w.p.1 there exists $d_1 > 0$ (depending on the sample path) such that for large n for $d_1 < j < C_2 \log n$,

$$L_n(j) \le -C_3 j\delta^2/2 + m(\bar{e}_{j,m}^2 - \bar{e}_{0,m}^2).$$

But,

$$\begin{aligned} m(\bar{e}_{j,m}^2 - \bar{e}_{0,m}^2) &= m(\bar{e}_{j,m} + \bar{e}_{0,m})j(m-j)^{-1}(\bar{e}_{0,m} - \bar{e}_{0,j}) \\ &\le jC_3\delta^2/4 \end{aligned}$$

for $j > d_2$ for some (sample-path-dependent) d_2. So, $L_n(j) < 0$ for $d_1 \vee d_2 < j < C_2 \log n$ for large n.

Case 3: $-k < j < -C_4 \log n$ where $C_4 = 3(2+\epsilon)\delta^{-2}$. By Lemma 1 and (2.3), w.p.1, for large n, for $-k < j < -C_4 \log n$,

$$L_n(j) < j\delta^2/3 + (2+\epsilon)\log n < 0.$$

Case 4: $-C_4 \log n \le j < -C_5 \log n$ where $C_5 = (2+\epsilon/2)\delta^{-2}$. By Lemma 1 (applied to $(k+j)\bar{e}_{-k,j}^2$ in (2.3)), w.p.1 for large n, for $-C_4 \log n \le j < -C_5 \log n$,

$$L_n(j) < j(1-\epsilon/8)\delta^2 + (2+\epsilon/8)\log n < 0.$$

Case 5: $-C_5 \log n \le j < 0$. We have $k + j > 2^{-1}\epsilon\delta^{-2}\log n$. By (2.3), w.p.1 there exists $d_3 > 0$ such that

$$L_n(j) < j2^{-1}\delta^2 + k(\bar{e}_{-k,j}^2 - \bar{e}_{-k,0}^2)$$

for $-C_5 \log n \leq j < -d_3$ for large n. But,

$$\begin{aligned} k(\bar{e}^2_{-k,j} - \bar{e}^2_{-k,0}) &= -kj(k+j)^{-1}(\bar{e}_{-k,j} + \bar{e}_{-k,0})(\bar{e}_{-k,0} - \bar{e}_{j,0}) \\ &< -j4^{-1}\delta^2 \end{aligned}$$

for $j < -d_4$ for large n (for some sample-path-dependent $d_4 > 0$). This completes Case 5 and completes the proof.

PROOF OF THEOREM 2. By Lemmas 2 and 3, it remains to show that $\hat{\tau}_n \to \hat{\tau}_\infty$ a.s. if $\liminf_{n\to\infty}(k \wedge m)/\log n > 2\delta^{-2}$. Note that $\hat{\tau}_\infty =$ arg $\max\{L_\infty(j) : -\infty < j < \infty\}$ where

$$L_\infty(j) = \begin{cases} \Sigma_{i=1}^{j}[(X_i - \mu - \delta)^2 - (X_i - \mu)^2], & \text{for } j > 0, \\ \Sigma_{i=j+1}^{0}[(X_i - \mu)^2 - (X_i - \mu - \delta)^2], & \text{for } j \leq 0. \end{cases}$$

Since for each j, $L_n(j) \to L_\infty(j)$ as $n \to \infty$ a.s. and since $\hat{\tau}_n = O(1)$ a.s., it follows that $\hat{\tau}_n \to \hat{\tau}_\infty$ a.s.

REMARK. The mle based on the infinite sample has a simple minimax property with respect to the $0-1$ loss. Suppose that $X_i, -\infty < i < \infty$ are independent and X_i has distribution P or Q according to whether $i \leq \tau$ or $i > \tau$ where P and Q are known and the change-point τ is unknown. Let $\hat{\tau}$ denote the mle. Then $\max_\tau \Pr_\tau(\tau' \neq \tau) \geq \max_\tau \Pr_\tau(\hat{\tau} \neq \tau)$ for any estimator τ'. Note that $\Pr_\tau(\hat{\tau} \neq \tau)$ is independent of τ (an equalizer rule). However, the mle based on a finite sample is not minimax in general since it is not an equalizer rule.

3. Proof of Theorem 1.

The following notation is adopted in this section: $\mathcal{L}(X)$ denotes the distribution of X; For $r < s$ define

$$S(r,s) := -S(s,r) := \Sigma_{i=r+1}^{s}[\delta_{X_i} - \mathcal{L}(X_i)]$$

and set

$$\begin{aligned} t &:= t(j,n) := t(j,k(n),m(n)) := (j+k)/n \\ \theta_n &:= t(0,k,m) = k/n \\ \rho_n(t) &:= \{(1-\theta_n)[t/(1-t)]^{1/2}\} \wedge \{\theta_n[(1-t)/t]^{1/2}\} \\ \Delta_n^j &:= \rho_n(t(j,n))(Q-P) \qquad \text{(note that } D_n^j \text{ estimates } \Delta_n^j) \\ B_n^j &:= D_n^j - \Delta_n^j \\ &= [t/(1-t)]^{1/2}n^{-1}S(j,m) - [(1-t)/t]^{1/2}n^{-1}S(-k,j). \end{aligned}$$

Note that (1.1) implies

$$\theta_n \wedge (1-\theta_n) \geq K \log n/n \qquad \textit{for large } n. \tag{3.1}$$

LEMMA 4. *There exists a $K_1 > 0$ (depending only on $\mathcal{D}$) such that w.p.1*

$$\sup_{-k<j<m} \|n^{-1}S(0,j)\| \le K_1(\log\log n/n)^{1/2} \qquad \textit{for large } n;$$

$$\|B_n^0\| \le K_1(\log\log n/n)^{1/2} \qquad \textit{for large } n;$$

$$\sup_{-k<j<m} \|(m-j)^{-1/2}S(j,m)\| \le K_1(\log n)^{1/2} \qquad \textit{for large } n;$$

$$\sup_{-k<j<m} \|(j+k)^{-1/2}S(-k,j)\| \le K_1(\log n)^{1/2} \qquad \textit{for large } n.$$

This lemma can be easily established along the lines of the proof of Lemma 2 of Dümbgen (1991).

PROOF OF THEOREM 1. Without loss of generality, we assume that $\theta_n \le 1/2$. The first part of the proof follows closely that of Proposition 1 of Dümbgen (1991). By (1.2) and Lemma 4, w.p.1 for large n,

$$|N_n(D_n^0) - w(\theta_n)N_n(Q-P)| \le \|B_n^0\| \le K_1(\log\log n/n)^{1/2}.$$

Since by (1.3) and (3.1), w.p.1 there is a sample-path-dependent $\epsilon > 0$ such that for large n, $w(\theta_n)N_n(Q-P) > 2^{-1/2}\theta_n^{1/2}(C_0+\epsilon) \gg (\log\log n/n)^{1/2}$, we have

$$N_n(D_n^0) > 2^{-1/2}C_0\theta_n^{1/2} \qquad \text{for large } n \text{ a.s.} \tag{3.2}$$

Next, we show that w.p.1 for large n for all $-k < j < m$,

$$N_n(D_n^j) - N_n(D_n^0) \le \|B_n^j - B_n^0\| - C_0[\rho_n(\theta_n) - \rho_n(t)]. \tag{3.3}$$

Since

$$\begin{aligned} N_n(D_n^j) &= N_n(B_n^j - B_n^0 + B_n^0 + \rho_n(t)w(\theta_n)^{-1}\Delta_n^0) \\ &= N_n(B_n^j - B_n^0 + [\rho_n(\theta_n) - \rho_n(t)]w(\theta_n)^{-1}B_n^0 + \rho_n(t)w(\theta_n)^{-1}D_n^0) \\ &\le \|B_n^j - B_n^0\| + [\rho_n(\theta_n) - \rho_n(t)]w(\theta_n)^{-1}\|B_n^0\| + \rho_n(t)w(\theta_n)^{-1}N_n(D_n^0), \end{aligned}$$

we have w.p.1 for large n for all $-k < j < m$

$$\begin{aligned} &N_n(D_n^j) - N_n(D_n^0) \\ &\quad \le \|B_n^j - B_n^0\| - [\rho_n(\theta_n) - \rho_n(t)][w(\theta_n)^{-1}N_n(D_n^0) - w(\theta_n)^{-1}\|B_n^0\|] \\ &\quad \le \|B_n^j - B_n^0\| - [\rho_n(\theta_n) - \rho_n(t)][N_n(Q-P) - 2w(\theta_n)^{-1}\|B_n^0\|]. \end{aligned}$$

Now, (3.3) follows from (1.3) and the fact that $\|B_n^0\|/w(\theta_n) = o(1)$ a.s. (by (1.1) and Lemma 4).

By (3.3), the theorem will be proved if we can show that w.p.1 there exists a $d > 0$ such that for large n for all j with $-k < j < m$ and $|j| > d$,

$$\|B_n^j - B_n^0\| < C_0[\rho_n(\theta_n) - \rho_n(t)]. \tag{3.4}$$

For $t > \theta_n$ (i.e. $j > 0$),

$$\begin{aligned} \rho_n(\theta_n) - \rho_n(t) &= \theta_n\{[(1-\theta_n)/\theta_n]^{1/2} - [(1-t)/t]^{1/2}\} \\ &\geq \theta_n(1-\theta_n)^{1/2}(\theta_n^{-1/2} - t^{-1/2}) \\ &= \theta_n^{1/2}(1-\theta_n)^{1/2}t^{-1/2}(t-\theta_n)/(t^{1/2}+\theta_n^{1/2}) \\ &\geq 2^{-3/2}\theta_n^{1/2}(t-\theta_n)/t. \end{aligned} \tag{3.5}$$

For $t < \theta_n \leq 1/2$,

$$\begin{aligned} \rho_n(\theta_n) - \rho_n(t) &= (1-\theta_n)\{[\theta_n/(1-\theta_n)]^{1/2} - [t/(1-t)]^{1/2}\} \\ &\geq (1-\theta_n)(1-t)^{-1/2}(\theta_n^{1/2} - t^{1/2}) \\ &\geq 2^{-1}(\theta_n - t)/(\theta_n^{1/2} + t^{1/2}) \\ &\geq 4^{-1}\theta_n^{-1/2}(\theta_n - t). \end{aligned} \tag{3.6}$$

We now prove (3.4) under the assumption that $3^{-1} \leq \theta_n \leq 2^{-1}$. For $|t-2^{-1}| > 4^{-1}$, $\rho_n(\theta_n) - \rho_n(t) > C_6$ for some constant $C_6 > 0$. But by Lemma 4, $\|B_n^j\| \leq 2K_1(\log n/n)^{1/2}$ for large n a.s. So, $\|B_n^j - B_n^0\| << \rho_n(\theta_n) - \rho_n(t)$ for all $|t-2^{-1}| > 4^{-1}$ for large n a.s. For $1/4 \leq t \leq 3/4$, write

$$\begin{aligned} B_n^j - B_n^0 = &\{[t/(1-t)]^{1/2} - [\theta_n/(1-\theta_n)]^{1/2}\}n^{-1}S(0,m) \\ &+ \{[(1-\theta_n)/\theta_n]^{1/2} - [(1-t)/t]^{1/2}\}n^{-1}S(-k,0) \\ &+ [w(t)^{-1} - w(\theta_n)^{-1}]n^{-1}S(j,0) + w(\theta_n)^{-1}n^{-1}S(j,0). \end{aligned} \tag{3.7}$$

Clearly, there exists a constant $C_7 > 0$ such that the norm of the sum of the first three terms on the right hand side of (3.7) is bounded by

$$\begin{aligned} C_7|t-\theta_n|n^{-1}&\{\|S(0,m)\| + \|S(-k,0)\| + \|S(j,0)\|\} \\ &\leq C_7|t-\theta_n|\,3K_1(\log\log n/n)^{1/2} \\ &\leq 2^{-1}C_0(\rho_n(\theta_n) - \rho_n(t)) \end{aligned}$$

for all $1/4 \leq t \leq 3/4$ for large n a.s. The norm of the last term is bounded for some constant $C_8 > 0$ by

$$C_8|t-\theta_n|\,\|j^{-1}S(j,0)\| \leq C_8|t-\theta_n|K_1(\log\log j/j)^{1/2} \tag{3.8}$$

for all $|j| > d_1$ (for some sample-path-dependent $d_1 > 0$) for large n a.s. The right hand side of (3.8) is bounded by $2^{-1}C_0(\rho_n(\theta_n) - \rho_n(t))$ if d_1 is chosen sufficiently large. This completes the proof of (3.4) for the case that $1/3 \le \theta_n \le 1/2$ for all n.

Therefore, it suffices to consider the case that $\theta_n \le 1/3$ for all n. We consider the following cases separately.

Case 1: $t \ge 1/2$. We have

$$\begin{aligned} B_n^j - B_n^0 =& [t/(1-t)]^{1/2} n^{-1} S(j,m) - [(1-t)/t]^{1/2} n^{-1} S(-k,j) \\ &- [\theta_n/(1-\theta_n)]^{1/2} n^{-1} S(0,m) + [(1-\theta_n)/\theta_n]^{1/2} n^{-1} S(-k,0). \end{aligned}$$

By Lemma 4, w.p.1 for large n for all $t \ge 1/2$,

$$\begin{aligned} \|[(1-t)/t]^{1/2} n^{-1} S(-k,j)\| &\le \|n^{-1} S(-k,0)\| + \|n^{-1} S(0,j)\| \\ &\le 2K_1 (\log\log n/n)^{1/2}, \\ \|[\theta_n/(1-\theta_n)]^{1/2} n^{-1} S(0,m)\| &\le K_1 (\log\log n/n)^{1/2}, \\ \|[(1-\theta_n)/\theta_n]^{1/2} n^{-1} S(-k,0)\| &\le \|n^{-1/2} k^{-1/2} S(-k,0)\| \\ &\le K_1 (\log\log n/n)^{1/2}, \\ \|[t/(1-t)]^{1/2} n^{-1} S(j,m)\| &\le \|n^{-1/2} (m-j)^{-1/2} S(j,m)\| \\ &\le K_1 (\log n/n)^{1/2}. \end{aligned}$$

But by (3.5),

$$\rho_n(\theta_n) - \rho_n(t) \ge 2^{-3/2} (K \log n/n)^{1/2} (t-\theta_n) \ge 2^{-3/2} (K \log n/n)^{1/2} 6^{-1}.$$

Thus, if K is sufficiently large, then (3.4) holds for all $t \ge 1/2$ for large n a.s.

Case 2: $\theta_n < t < 1/2$. Write

$$\begin{aligned} B_n^j - B_n^0 =& \{[t/(1-t)]^{1/2} - [\theta_n/(1-\theta_n)]^{1/2}\} n^{-1} S(0,m) \\ &+ \{[(1-\theta_n)/\theta_n]^{1/2} - [(1-t)/t]^{1/2}\} n^{-1} S(-k,0) - [t(1-t)]^{-1/2} n^{-1} S(0,j). \end{aligned} \tag{3.9}$$

Since $0 \le [t/(1-t)]^{1/2} - [t/(1-\theta_n)]^{1/2} \le C_9 t^{1/2} (t-\theta_n)$ for some constant $C_9 > 0$, we have

$$\begin{aligned} |[t/(1-t)]^{1/2} - [\theta_n/(1-\theta_n)]^{1/2}| &\le |[t/(1-\theta_n)]^{1/2} \\ &\quad - [\theta_n/(1-\theta_n)]^{1/2}| + C_9 t^{1/2} (t-\theta_n) \\ &\le (t-\theta_n)/[(1-\theta_n)^{1/2} (t^{1/2} + \theta_n^{1/2})] + C_9 t^{1/2} (t-\theta_n) \\ &\le C_{10} (t-\theta_n)/t^{1/2} \end{aligned}$$

for some $C_{10} > 0$. Thus, w.p.1 for large n for $\theta_n < t < 1/2$, the norm of the first term on the right hand side of (3.9) is bounded by

$$C_{10} K_1 (t-\theta_n) t^{-1/2} (\log\log n/n)^{1/2}. \tag{3.10}$$

The ratio of (3.10) to the right-most side of (3.5) is bounded by a constant times

$$(t/\theta_n)^{1/2}(\log\log n/n)^{1/2} = [(j+k)/k]^{1/2}(\log\log n/n)^{1/2} = o(1).$$

Next, since

$$\begin{aligned}|[(1-\theta_n)/\theta_n]^{1/2} - [(1-t)/t]^{1/2}|\\ &\leq (1-\theta_n)^{1/2}|\theta_n^{-1/2} - t^{-1/2}| + t^{-1/2}|(1-t)^{1/2} - (1-\theta_n)^{1/2}|\\ &\leq C_{11}(t-\theta_n)\theta_n^{-1/2}t^{-1}\end{aligned}$$

for some $C_{11} > 0$, the norm of the second term on the right hand side of (3.9) is bounded by

$$\begin{aligned}C_{11}(t-\theta_n)\theta_n^{1/2}t^{-1}\|k^{-1}S(-k,0)\|\\ &\leq C_{11}K_1(t-\theta_n)\theta_n^{1/2}t^{-1}(\log\log k/k)^{1/2}\end{aligned}$$

which is much smaller than the right-most side of (3.5) when n (and hence k) is large. The norm of the last term on the right hand side of (3.9) equals

$$\|[t(1-t)]^{-1/2}(t-\theta_n)j^{-1}S(0,j)\| \leq C_{12}t^{-1/2}(t-\theta_n)(\log\log j/j)^{1/2} \quad (3.11)$$

for $j > d_2$ (for some sample-path-dependent d_2) for large n a.s. where C_{12} is a constant. For $j > d_2$, the ratio of the right hand side of (3.11) to that of (3.5) is bounded by a constant times

$$\begin{aligned}(t/\theta_n)^{1/2}(\log\log j/j)^{1/2} &= [(k+j)/k]^{1/2}(\log\log j/j)^{1/2}\\ &\leq 2\{1+(j/k)^{1/2}\}(\log\log j/j)^{1/2}\end{aligned}$$

which can be made arbitrarily small by choosing a sufficiently large d_2. This completes Case 2.

Case 3: $\log n/n \leq t < \theta_n$. Again, we bound each of the three terms on the right hand side of (3.9). The norm of the first term is bounded by a constant times

$$|t-\theta_n|\,t^{-1/2}\|n^{-1}S(0,m)\| \leq K_1|t-\theta_n|\,t^{-1/2}(\log\log n/n)^{1/2} \quad (3.12)$$

for $\log n/n \leq t < \theta_n$ for large n a.s. The ratio of the right hand side of (3.12) to that of (3.6) is $o(1)$ since $(\theta_n/t)^{1/2} = O((n/\log n)^{1/2})$. The norm of the second term is bounded by a constant times

$$(\theta_n - t)t^{-1/2}\|k^{-1}S(-k,0)\| \leq K_1(\theta_n - t)t^{-1/2}(\log\log k/k)^{1/2}$$

which is again dominated by the right hand side of (3.6), since

$$(\theta_n/t)^{1/2}(\log\log k/k)^{1/2} = (\log\log k/(nt))^{1/2} = o(1).$$

The norm of the last term is bounded by a constant times

$$t^{-1/2}(\theta_n - t)\|j^{-1}S(j,0)\| \le K_1\theta_n^{-1/2}(\theta_n - t)(\theta_n/t)^{1/2}(\log\log|j|/|j|)^{1/2} \quad (3.13)$$

for $j < -d_3$ (for some $d_3 > 0$) for large n a.s. Noting that $|j| = n(\theta_n - t)$ and considering the two cases $t < \theta_n/2$ (implying that $|j| \ge n\theta_n/2$) and $t \ge \theta_n/2$, it can be shown that for $j < -d_3$, the ratio of the right hand side of (3.13) to that of (3.6) can be made arbitrarily small by choosing a large d_3. Therefore it follows that (3.4) holds for all t with $\log n/n \le t < \theta_n$ and $j < -d_3$, for large n a.s.

Case 4: $t < \log n/n$. We assume that $K > 2$ so that $t < \theta_n/2$. By (3.6),

$$\rho_n(\theta_n) - \rho_n(t) \ge 8^{-1}\theta_n^{1/2} \ge 8^{-1}(K\log n/n)^{1/2}$$

for large n. The norm of the first term on the right hand side of (3.9) is bounded by a constant times $\theta_n^{1/2}\|n^{-1}S(0,m)\| = o(\theta_n^{1/2})$ for large n. The sum of the last two terms equals

$$\begin{aligned}&[(1-\theta_n)/\theta_n]^{1/2}n^{-1}S(-k,0) + [w(t)^{-1}\\ &\quad - ((1-t)/t)^{1/2}]n^{-1}S(-k,0) - w(t)^{-1}n^{-1}S(-k,j)\end{aligned} \quad (3.14)$$

The norm of the first term of (3.14) is bounded by $\theta_n^{1/2}\|k^{-1}S(-k,0)\| = o(\theta_n^{1/2})$ for large n. The norm of the second term is bounded by a constant times $t^{1/2}\|n^{-1}S(-k,0)\| = o(\theta_n^{1/2})$ for large n a.s. The norm of the last term is bounded by a constant times $n^{-1/2}\|(j+k)^{-1/2}S(-k,j)\| \le K_1 n^{-1/2}(\log n)^{1/2}$ for large n a.s. So if K is sufficiently large, (3.4) holds for all $t < \log n/n$ for large n a.s. The proof is complete.

4. Some Concluding Remarks. We have proved that $\tau_n^* = O(1)$ a.s. if $\liminf_{n\to\infty}(k \wedge m)/\log n > K$ for a large constant K (depending only on C_0 and $\mathcal{D}$ in Condition (A)) where τ_n^* is an estimator of Dümbgen's (1991). It would be interesting to characterize the best possible K as we did for the mle in the normal case. It would also be interesting to know whether τ_n^* actually converges a.s. to a random variable (which may be regarded as an infinite-sample estimator of the change-point). Dümbgen (1991) obtained the limiting distribution of τ_n^* for some particular seminorms with $E = R$ (the real line). It seems that the almost sure limit of τ_n^* can be obtained in these cases (perhaps under a stronger condition than $\liminf_{n\to\infty}(k\wedge m)/\log n > K$).

Another interesting question is to find a necessary and sufficient condition on the growth rate of $k \wedge m$ as $n \to \infty$ under which there exists a nonparametric estimator which is within $O(1)$ of the change-point a.s.

Dümbgen also studied Bayes estimators (for the case $E = R$ only) and obtained their limiting distributions. While we have not investigated their behavior, it appears that these estimators should have essentially the same almost sure convergence properties as those considered in Theorem 1.

Finally, we mention a recent result of Ferger and Stute (1992), who proved (without specifying the probabilistic relation between the rows of the array of random variables) that $\tilde{\tau}_n - \tau_n = O(\log n)$ almost surely where $\tau_n = [n\theta]$ for some $0 < \theta < 1$ and $\tilde{\tau}_n$ belongs to a class of variants of Darkhovskhi's (1976) estimator. Clearly, $O(\log n)$ can be improved to $O(1)$ by embedding the rows of the array into an infinite sequence of independent random variables as discussed in Section 1.

Acknowledgement. The work of R. Davis was partially supported by NSF Grant DMS-9100392.

REFERENCES

CARLSTEIN, E. (1988). Nonparametric estimation of a change-point. *Ann. Statist.* **16**, 188–197.

DARKHOVSKHI, B. S. (1976). A non-parametric method for the a posteriori detection of the "disorder" time of a sequence of independent random variables. *Theory Probab. Appl.* **21**, 178–183.

DÜMBGEN, L. (1991). The asymptotic behavior of some nonparametric change-point estimators. *Ann. Statist.* **19** 1471–1495.

FERGER, D. and STUTE, W. (1992). Convergence of changepoint estimators. *Stochastic Process. Appl.* **42** 345–351.

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

POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

VENKATRAMAN, E. S. (1992). Consistency results in multiple change-point problems. Technical Report No. 24, Department of Statistics, Stanford University.

DEPARTMENT OF STATISTICS
COLORADO STATE UNIVERSITY
FORT COLLINS, CO 80523

SCHOOL OF MATHEMATICS
QUEENSLAND UNIVERSITY OF TECHNOLOGY
GPO BOX 2434
BRISBANE, QUEENSLAND 4001
AUSTRALIA

DEPARTMENT OF STATISTICS
COLORADO STATE UNIVERSITY
FORT COLLINS, CO 80523