THE INSTITUTE OF STATISTICS

THE CONSOLIDATED UNIVERSITY
OF NORTH CAROLINA

NONPARAMETRIC CHANGE-POINT ESTIMATION

(Revised)

by

E. Carlstein
University of North Carolina
Chapel Hill, NC 27514

Mimeo Series #1595

February 1986

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

NONPARAMETRIC CHANGE-POINT ESTIMATION

(Revised)

by E. Carlstein

University of North Carolina at Chapel Hill

SUMMARY

Consider a sequence of independent random variables $\{X_i: 1 \leq i \leq n\}$ having cdf F for $i \leq \theta n$ and cdf G otherwise. A strongly consistent estimator of the change-point $\theta \in (0,1)$ is proposed. The estimator requires no knowledge of the functional forms or parametric families of F and G. Furthermore, F and G need not differ in their means (or other measure of location). The only requirement is that F and G differ on a set of positive probability. The proof of consistency provides rates of convergence and bounds on the error probability for the estimator. The estimator is applied to two well-known data sets; in both cases it yields results in close agreement with previous parametric analyses. A simulation study is conducted, showing that the estimator performs well even when F and G share the same mean, variance, and skewness.

Key words: Change-point; Estimation; Nonparametric.

AMS 1980 Subject Classification: Primary 62G05; Secondary 60F15.

Running Head: Nonparametric Changepoint Estimator

1. INTRODUCTION

Let $X_1^n, \ldots, X_n^n$ be independent random variables with:

$$X_1^n, \ldots, X_{[\theta n]}^n \text{ identically distributed with cdf } F;$$

$$X_{[\theta n]+1}^n, \ldots, X_n^n \text{ identically distributed with cdf } G;$$

where $[y]$ denotes the greatest integer not exceeding y. The parameter $\theta \in (0,1)$ is the change-point to be estimated. The body of literature addressing this problem is extensive, but most of the work is based upon at least one of the following assumptions:

(i) F and G are known to belong to parametric families (e.g. normal, binomial), or are otherwise known in functional form;

(ii) F and G differ, in particular, in their levels (e.g. mean or median).

Hinkley (1970) and Hinkley & Hinkley (1970) use maximum likelihood to estimate θ in the situation where F and G are from the same parametric family. Hinkley (1972) generalizes this method to the case where F and G may be arbitrary known distributions, or alternatively where a sensible discriminant function (for discriminating between F and G) is known. Smith's (1975) Bayesian approach and Cobb's (1978) conditional solution also require assumptions of type (i). These authors generally suggest that any unknown parameters in F and G can be estimated from the sample, but nevertheless F and G must be specified as functions of those parameters.

At the other extreme, Darkhovshk (1976) presents a nonparametric estimator based on the Mann-Whitney statistic. Although his estimator makes no explicit use of the functional forms of F and G, his asymptotic results require $\int_{-\infty}^{\infty} G(x)\,dF(x) \neq \frac{1}{2}$. This excludes cases where F and G are both symmetric and have a common median. Bhattacharyya & Johnson (1968) give a nonparametric test for the presence of a change-point, but again under the type (ii) assumption that the variables after the change are stochastically larger than those before. (See Shaban (1980) for an annotated bibliography of change-point literature.)

In contrast to assumptions of types (i) and (ii), the estimator studied here does not require any knowledge of F and G; and virtually any salient difference between F and G will ensure detection of the change-point (asymptotically). Specifically, we assume:

(I) The set $\Lambda = \{x \in R : |F(x) - G(x)| > 0\}$ satisfies either

$$\int_\Lambda dF(x) > 0 \quad \text{or} \quad \int_\Lambda dG(x) > 0 .$$

Note that F and G may be discrete, continuous, or mixed. The theoretical results for Darkhovshk's (1976) nonparametric estimator assumed F and G to be continuous. Unlike Cobb (1978), we do not require the supports of F and G to be identical; in fact the supports may be entirely unknown.

The intuition behind the proposed estimator is as follows. For a hypothetical (but not necessarily correct) change-point $t \in (0,1)$, consider the pre-t empirical cdf ${}_t h^n(x)$, which is constructed as if $X_1^n, \ldots, X_{[tn]}^n$ were identically distributed, and the post-t empirical cdf $h_t^n(x)$, which is constructed as if $X_{[tn]+1}^n, \ldots, X_n^n$ were identically distributed. The former estimates the unknown mixture distribution:

$${}_t h(x) = I\{t \leq \theta\}\ F(x) + I\{t > \theta\}\ (\theta\, F(x) + (t-\theta)\ G(x))/t,$$

and the latter similarly estimates:

$$h_t(x) = I\{t \leq \theta\}\ ((\theta-t)\ F(x) + (1-\theta)\ G(x))/(1-t) + I\{t > \theta\}\ G(x) .$$

The total difference between these two distributions is measured by:

$$J_n(t) = \sum_{j=1}^{n} |{}_t h(X_j^n) - h_t(X_j^n)|/n$$

$$= (I\{t \leq \theta\}\ (1-\theta)/(1-t) + I\{t > \theta\}\ \theta/t)\ J_n(\theta) .$$

Note that $J_n(t)$ attains its maximum (over $t \in (0,1)$) at $t = \theta$. Thus a reasonable estimator for θ is the value of t that maximizes:

$$H_n(t) = \sum_{j=1}^{n} \left| {}_t h^n(X_j^n) - h_t^n(X_j^n) \right| / n .$$

In Section 2 the estimator is formally defined and its asymptotic properties are presented. The results include strong consistency, with rates of convergence and bounds on the error probability. Proof of these results is deferred to Section 4. Section 3 investigates the finite-sample behavior of the estimator: First the estimator is calculated on Cobb's (1978) Nile data and on the Lindisfarne Scribes data (see Smith, 1980). In both examples this nonparametric analysis produces results which are nearly identical to the results of the earlier parametric analyses. Then the estimator is tested (via simulation) in a situation where no other change-point estimator can be used: F and G are both unknown and are of different parametric families, but they are both symmetric and share the same mean and variance. Here again the performance of the estimator is quite reasonable.

2. THE ESTIMATOR

In order to formalize the ideas of Section 1, the following additional definitions and assumptions are introduced. For $t \in (0,1)$ the empirical cdf's are:

$${}_t h^n(x) = \sum_{i=1}^{[tn]} I\{X_i^n \le x\} / [tn] ,$$

$$h_t^n(x) = \sum_{i=[tn]+1}^{n} I\{X_i^n \le x\} / (n - [tn]) .$$

Since the accuracy of the approximation of $J_n(t)$ by $H_n(t)$ depends in turn upon the accuracy of ${}_t h^n(\cdot)$ and $h_t^n(\cdot)$ in approximating ${}_t h(\cdot)$ and $h_t(\cdot)$, we must restrict our attention to values of t that will yield reasonably large sample sizes ($[tn]$ and $n - [tn]$) for our empirical cdf's. Moreover, for t such that [tn] (say) is small, the corresponding ${}_th^n(\cdot)$ consists of just a few big steps - regardless of how smooth F and G are. So, even if n is large and $h_t^n(\cdot)$ is devoid of large jumps, the difference $H_n(t)$ will tend to be exaggerated. That is, the change-point estimator will be drawn (erroneously) towards such extremal values of t. This again suggests that we restrict our attention to values of t bounded away from 0 and 1. Specifically, we only consider $t \in [\alpha_n, 1-\alpha_n]$, where $\{\alpha_n: n \geq 1\}$ is such that $n\alpha_n \to \infty$ and $\alpha_n \downarrow 0$ as $n \to \infty$. The change-point estimator is then defined as:

$$\theta_n \in [\alpha_n, 1-\alpha_n] \quad \text{for which } H_n(\theta_n) = \sup_{\alpha_n \leq t \leq 1-\alpha_n} H_n(t) \,.$$

In practice, α_n prohibits estimation of a change that is too close to either end of the data string. The constraint $n\alpha_n \to \infty$ requires the minimum acceptable sample size to grow with n, thus ensuring increasing precision (in estimating $J_n(t)$ by $H_n(t)$) over the whole range of acceptable t-values. On the other hand, since $\alpha_n \downarrow 0$, any $\theta \in (0,1)$ will eventually be fair game. (Darkhovshk (1976) assumes $\theta \in [\alpha, 1-\alpha]$, where $\alpha \in (0,\frac{1}{2})$ is known.)

Notice that $H_n(t)$ depends upon t only through [tn], so for fixed n there are at most n-1 distinct values of $H_n(\cdot)$ to be compared (as t ranges through (0,1)). Therefore an estimator θ_n attaining the supremum does exist and is easily calculated. Also note that θ_n is invariant under strict monotone transformations of the data, since $H_n(t)$ depends on the data only via terms of the form $I\{X_i^n \leq X_j^n\}$.

The fundamental theoretical result for θ_n is

<u>THEOREM</u>: Let $\{a_n: n \geq 1\}$ be s.t. $1 \geq a_n > 0$, $\alpha_n/a_n \to 0$, and $n\alpha_n a_n \to \infty$ as $n \to \infty$. Assume (I) holds. Then, for any $\varepsilon > 0$,

$$P\{|\theta_n - \theta|/a_n > \varepsilon\} \leq c_1 n^2 \exp\{-c_2 \varepsilon^2 n\alpha_n a_n^2\} \; \forall n \geq n_\varepsilon \,,$$

where $c_1 > 0$ and $c_2 > 0$ are constants.

Choosing $a_n \equiv 1$ we obtain the bound

$$P\{|\theta_n - \theta| > \varepsilon\} \leq c_1 n^2 \exp\{-c_2 \varepsilon^2 n\alpha_n\} .$$

Strong consistency of θ_n follows from the Theorem by applying the Borel-Cantelli Lemma. Formally we have

COROLLARY: Let $\{a_n : n \geq 1\}$ be s.t. $1 \geq a_n > 0$, $\alpha_n / a_n \to 0$, and $\sum_{n=1}^{\infty} n^2 \exp\{-\tau n\alpha_n a_n^2\} < \infty \; \forall \tau > 0$. Assume (I) holds. Then

$$|\theta_n - \theta| / a_n \to 0 \quad \text{almost surely as } n \to \infty .$$

If we choose $\alpha_n = n^{-\xi}$ with $0 < \xi < 1$, then the conditions of the Corollary are satisfied by $a_n = n^{-\phi}$ with $0 \leq \phi < \min\{\xi, \frac{1}{2}(1-\xi)\}$.

3. APPLICATIONS

The Nile Data. Cobb (1978) reports the annual volume of discharge from the Nile River for each year from 1871 to 1970. His analyses assume that the observations are independent normal variables with common variance for the whole series. The results he obtains clearly indicate 1898 as the most likely change-point; he cites independent meteorological evidence that this change is real. Figure 1 shows the data X_i along with the criterion function $H_n(t)$ at $t = i/n$. For any reasonable choice of α_n, our nonparametric estimate is $\theta_n = .28$ (i.e. 1898). The function $H_n(t)$ does exhibit the sort of erratic behavior near $t = 0$ and 1 as was discussed in Section 2; note that Cobb (1978) also makes the qualification that the change-point be restricted away from the ends of the data.

The Lindisfarne Scribes Data. The Lindisfarne text (as studied by Smith (1980)) divides into 13 sections. It is assumed that only one scribe was involved in the writing of any one section, and that sections written by any one scribe are consecutive. It is also assumed that a scribe may be characterized by his propensity to use one of two possible grammatical variants: either the "s" or "δ" ending in

FIGURE 1: The Nile Data

i = Year (1871 to 1970).

X_i = Annual volume of discharge ($10^{10} m^3$) .

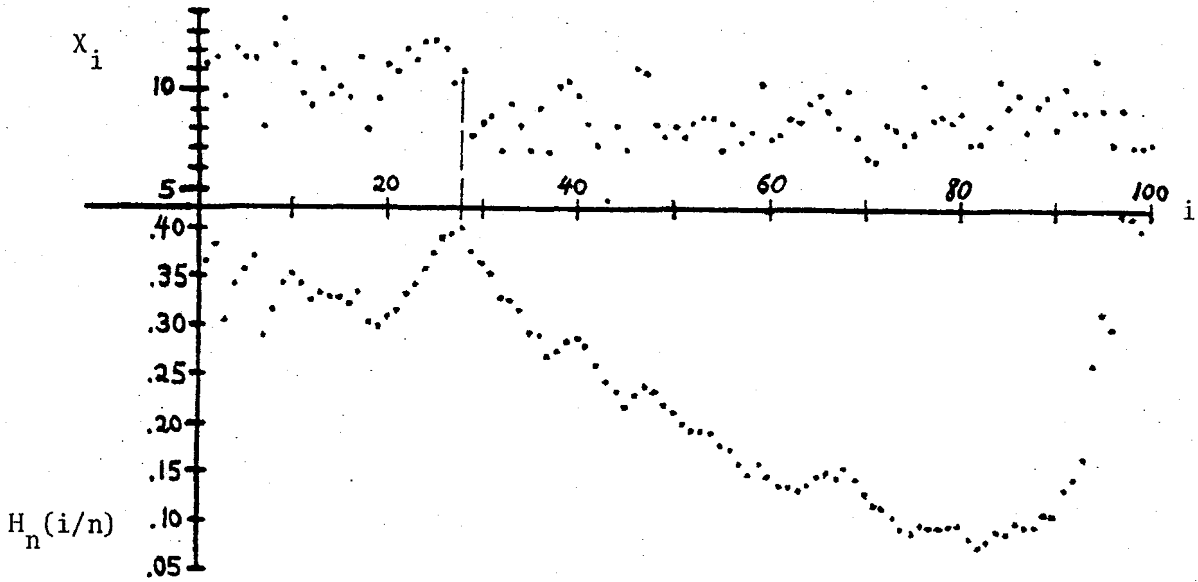

TABLE 1: The Lindisfarne Scribes Data

i = Section of the text.

X_i = Proportion of "s" endings.

i	1	2	3	4	5	6	7	8	9	10	11	12	13
X_i	.571	.722	.705	.800	.538	.756	.813	.807	.854	.864	.850	.810	.800
$H_n(i/n)$	.42	.37	.41	.37	.47	.49	.42	.41	.31	.22	.19	.27	---

TABLE 2: Simulation Study

θ	n	$E\{\theta_n\}$	$E\{\|\theta_n - \theta\|\}$
.4	100	.423*	.101*
.4	200	.402†	.085†

*Estimated empirically by 250 realizations of θ_n; the standard deviation of $E\{\theta_n\}$ is approximately .009 .

†Estimated empirically by 200 realizations of θ_n; the standard deviation of $E\{\theta_n\}$ is approximately .009 .

Note: For calculating θ_n we used $\alpha_n = n^{-\xi}$ with $\xi = .3$.

the present indicative 3rd person singular. Let m_i denote the total number of relevant words in the i^{th} section, and let Y_i denote the number of times that the "s" ending was used in those words. Smith (1980) assumes that the Y_i are independent binomial variables with common parameter p between change-points, and he uses independent beta prior distributions on the p's. His analysis (which entertains the possibility of multiple change-points) arrives at a model with changes of scribe after section 4 and again after section 5.

We take the view that the r.v.s $X_i = Y_i/m_i$ (for all i on one side of $[\theta n]$) share approximately the same distribution. (The distributions cannot be precisely identical since m_i - and thus the supports - vary with i. The approximation should not be of much consequence since in all cases m_i is fairly large (i.e. at least 20).) Table 1 shows the data and the $H_n(t)$ function, which is maximized at $\theta_n = 6/13$. Our non-parametric analysis has the following interpretation: It is clear by inspection that the earlier X_i's are smaller in magnitude but larger in their variability than the later X_i's. In estimating the precise change-point for this transition, our θ_n chooses to group sections 5 and 6 with the pre-change portion of the series. Given the constraint of our analysis that there be exactly one change-point, this choice is quite reasonable: X_5 and X_6 are of relatively small magnitude and large variability. Smith's (1980) approach refines the analysis by explaining (with more change-points and more assumptions) the disorder in the series at sections 4 through 6.

Simulation Study. When the functional forms of F and G are unknown to the statistician, but both distributions are symmetric with the same mean, then no other change-point estimator is appropriate. Such is the case in this example:

F is the distribution with density $f(x) = .697128\, x^2 I\{|x| < 1.291\}$;

G is the N(0,1) distribution.

Actually, F and G also share the same variance in this situation, making it even more difficult for the user to choose an estimator that discriminates between them. Nonetheless, θ_n performs well for moderately large n, as illustrated by the simulation results in Table 2.

4. PROOF OF THEOREM

LEMMA 1: Let $Y_1^n, \ldots, Y_n^n$ be iid with cdf Q, and define:

$$B_n = \{(r,s,\ell,m): \alpha_n \le r-s \quad \text{and} \quad 0 \le m \le s < r \le \ell \le 1\},$$

$$Q_n(x;r,s,\ell,m) = \sum_{i=[sn]+1}^{[rn]} I\{Y_i^n \le x\}/([\ell n] - [mn]),$$

$$d_n(x;r,s,\ell,m) = |Q_n(x;r,s,\ell,m) - (r-s)Q(x)/(\ell-m)|,$$

$$\Delta_n = \sup_{(r,s,\ell,m)\in B_n} \sup_{x\in R} d_n(x;r,s,\ell,m).$$

Then $P\{\Delta_n/a_n > \varepsilon\} \le K_1 n^2 \exp\{-K_2 \varepsilon^2 n\alpha_n a_n^2\} \; \forall n \ge N_\varepsilon$,

where $K_1 > 0$ and $K_2 > 0$ are constants.

PROOF: $d_n(x;r,s,\ell,m) \le d_n(x;r,s,r,s)([rn] - [sn])/([\ell n] - [mn])$

$$+ Q(x)|([rn] - [sn])/([\ell n] - [mn]) - (r-s)/(\ell-m)|$$

$$\le d_n(x;r,s,r,s) + 4/n\alpha_n .$$

So, $P\{\Delta_n/a_n > \varepsilon\} \le P\{\sup_{(r,s)\in C_n} D_n(r,s) > \frac{1}{2} a_n\varepsilon\} + P\{4/n\alpha_n a_n > \frac{1}{2}\varepsilon\}$,

where $C_n = \{(r,s): \alpha_n \le r-s \text{ and } 0 \le s < r \le 1\}$ and $D_n(r,s) = \sup_{x\in R} d_n(x;r,s,r,s)$.

Now $D_n(r,s)$ depends on r and s only through [rn] and [sn], each of which takes on at most n distinct values as (r,s) ranges through C_n. Hence, for $n \ge N_\varepsilon$,

$$P\{\Delta_n/a_n > \varepsilon\} \le n^2 \sup_{(r,s)\in C_n} P\{D_n(r,s) > \tfrac{1}{2} a_n\varepsilon\}$$

$$\le n^2 \sup_{(r,s)\in C_n} K \exp\{-\tfrac{1}{2}([rn] - [sn])\varepsilon^2 a_n^2\}$$

where $K > 0$ is constant.

The last inequality is an application of Dvoretzky, Kiefer, & Wolfowitz (1956); see their Lemma 2 and the discussion after their Theorem 3. Since $[rn] - [sn] \geq \frac{1}{2}\alpha_n n \;\; \forall (r,s) \in C_n$, Lemma 1 is established. □

In order to apply Lemma 1, we must restrict the range of t away from θ, so that there is a sufficient amount of data between $[tn]$ and $[\theta n]$. This constraint will be eliminated later in the argument. Let $A_n = [\alpha_n, \theta-\alpha_n] \cup [\theta+\alpha_n, 1-\alpha_n]$.

LEMMA 2: $P\{\sup_{t\in A_n} |J_n(t) - H_n(t)|/a_n > \varepsilon\} \leq K_3\, n^2 \exp\{-K_4\varepsilon^2\, n\alpha_n\, a_n^2\} \;\; \forall n \geq N'_\varepsilon$, where $K_3 > 0$ and $K_4 > 0$ are constants.

PROOF: For $t \in A_n$ we have:

$$H_n(t) - J_n(t) \leq \sum_{j=1}^{n} (|{}_t h^n(X_j^n) - {}_t h(X_j^n)| + |h_t^n(X_j^n) - h_t(X_j^n)|)/n; \text{ the same bound}$$

applies to $J_n(t) - H_n(t)$.

$$\text{Now } |{}_t h^n(X_j^n) - {}_t h(X_j^n)| \leq I\{t \leq \theta\}\; |Q_n(X_j^n; t,0,t,0) - F(X_j^n)| +$$

$$I\{t > \theta\}\; (|Q_n(X_j^n; \theta,0,t,0) - \theta F(X_j^n)/t| + |Q_n(X_j^n; t,\theta,t,0) - (t-\theta)\; G(X_j^n)/t|),$$

where the Y_i^n's in the definition of $Q_n(\cdot)$ (see Lemma 1 above) have been replaced with X_i^n's that are identically distributed with either distribution F or distribution G. Each of the three terms on the r.h.s. of the last inequality can be bounded above by a term of the form Δ_n, which does not depend on j or t. A similar argument applies to $|h_t^n(X_j^n) - h_t(X_j^n)|$, so that

$$P\{\sup_{t\in A_n} |H_n(t) - J_n(t)|/a_n > \varepsilon\} \leq 2\, K_1 n^2 \exp\{-K_2(\varepsilon/4)^2\, n\alpha_n\, a_n^2\},$$

by Lemma 1. □

Still restricting attention to the set A_n, define $\tilde\theta_n$ as:

$$\tilde\theta_n \in A_n \text{ for which } H_n(\tilde\theta_n) = \sup_{t\in A_n} H_n(t), \quad \text{and} \quad \tilde\theta_n = \theta_n \text{ if } \theta_n \in A_n .$$

Similarly, define $t_n = I\{\theta \leq \frac{1}{2}\}(\theta-\alpha_n) + I\{\theta > \frac{1}{2}\}(\theta+\alpha_n)$, so that:

$t_n \in A_n$ and $J_n(t_n) = \sup_{t \in A_n} J_n(t)$.

LEMMA 3: $P\{|J_n(\tilde{\theta}_n) - J_n(\theta)|/a_n > \varepsilon\} \leq K_5 n^2 \exp\{-K_6 \varepsilon^2 n\alpha_n a_n^2\} \; \forall\, n \geq N''_\varepsilon$, where $K_5 > 0$ and $K_6 > 0$ are constants.

PROOF: $|J_n(\tilde{\theta}_n) - J_n(\theta)| \leq |J_n(\tilde{\theta}_n) - H_n(\tilde{\theta}_n)| + |H_n(\tilde{\theta}_n) - J_n(t_n)| + |J_n(t_n) - J_n(\theta)|$.

Note that either $H_n(\tilde{\theta}_n) \geq J_n(t_n) \geq J_n(\tilde{\theta}_n)$ or $J_n(t_n) \geq H_n(\tilde{\theta}_n) \geq H_n(t_n)$; in either case the second term on the r.h.s. of the above inequality is bounded by $\sup_{t \in A_n} |J_n(t) - H_n(t)|$. Of course the same bound applies to the first term.

The third term is $J_n(\theta) \; (I\{\theta \leq \frac{1}{2}\}/(1-\theta+\alpha_n) + I\{\theta > \frac{1}{2}\}/(\theta+\alpha_n))\alpha_n$. Since $J_n(\theta) \leq \sup_{x \in R} F(x) + \sup_{x \in R} G(x) = 2$, and since the second factor in the above product is also bounded by 2, we obtain

$$P\{|J_n(\tilde{\theta}_n) - J_n(\theta)|/a_n > \varepsilon\} \leq P\{2 \sup_{t \in A_n} |J_n(t) - H_n(t)|/a_n > \tfrac{1}{2}\varepsilon\} + P\{4\alpha_n/a_n > \tfrac{1}{2}\varepsilon\} .$$

Now apply Lemma 2. □

Denote $Z_j^n = |F(X_j^n) - G(X_j^n)|$, $\mu_F = \int_R |F(x) - G(x)|dF(x)$, $\mu_G = \int_R |F(x) - G(x)|dG(x)$, $\mu = \theta\mu_F + (1-\theta)\mu_G$, and $c = \frac{1}{2}\mu$. By assumption (I) we have $\mu > 0$. Note that $J_n(\theta) = \bar{Z}_1^n[\theta n]/n + \bar{Z}_2^n(n-[\theta n])/n$, where $\bar{Z}_1^n = \sum_{j=1}^{[\theta n]} Z_j^n/[\theta n]$ and $\bar{Z}_2^n = \sum_{j=[\theta n]+1}^{n} Z_j^n/(n-[\theta n])$.

LEMMA 4: $P\{|\tilde{\theta}_n-\theta|/a_n > \varepsilon\} \leq K_7 n^2 \exp\{-K_8\varepsilon^2 n\alpha_n a_n^2\} \; \forall n \geq N'''_\varepsilon$, where $K_7 > 0$ and $K_8 > 0$ are constants.

PROOF: For any $t \in (0,1)$: $|\theta-t| > \delta \Rightarrow |J_n(\theta) - J_n(t)|$

$$= (I\{\theta-t \geq 0\}(\theta-t)/(1-t) + I\{t-\theta > 0\}(t-\theta)/t) \; J_n(\theta) > \delta \; J_n(\theta) .$$

Thus $P\{|\tilde{\theta}_n-\theta|/a_n > \varepsilon\} \leq P\{|J_n(\theta) - J_n(\tilde{\theta}_n)|/a_n > \varepsilon \, J_n(\theta)\}$

$$\leq P\{|J_n(\theta) - J_n(\tilde{\theta}_n)|/a_n > \varepsilon c\} + P\{J_n(\theta) < c\} .$$

Since Lemma 3 applies to the first probability on the r.h.s. of this last inequality, it suffices to consider

$$P\{|J_n(\theta) - \mu| > c\} \le P\{|\overline{Z}_1^n[\theta n]/n - \theta\mu_F| > \tfrac{1}{2}c\} + P\{|\overline{Z}_2^n(n - [\theta n])/n - (1-\theta)\mu_G| > \tfrac{1}{2}c\} .$$

The two probabilities in this upper bound are both handled in the same way; we only deal explicitly with the first one. It is dominated by

$$P\{\overline{Z}_1^n|[\theta n]/n-\theta| > c/4\} + P\{\theta|\overline{Z}_1^n - \mu_F| > c/4\} .$$

Of these two probabilities, the first is zero for n sufficiently large (since $\overline{Z}_1^n \le 1$ always holds). Since $\{Z_j^n: 1 \le j \le [\theta n]\}$ are iid and bounded, we can use equation (2.3) of Hoeffding (1963) to dominate the second probability by $2\exp\{-2[\theta n](c/4\theta)^2\}$. Since $\alpha_n a_n^2 \to 0$, this bound can be absorbed into the earlier bound from Lemma 3. □

Finally, for the unrestricted estimator we have

$$P\{|\theta_n-\theta|/a_n > \varepsilon\} \le P\{|\tilde{\theta}_n-\theta|/a_n > \varepsilon\} + P\{|\theta_n-\theta| > \varepsilon\, a_n \cap \theta_n \notin A_n\} .$$

Since $\theta_n \notin A_n \iff |\theta_n-\theta| < \alpha_n$, and $\alpha_n/a_n \to 0$, the last probability is zero for n sufficiently large. Applying Lemma 4 concludes the proof.

5. ACKNOWLEDGMENT

The referees' comments led to numerous improvements in this paper.

REFERENCES

BHATTACHARYYA, G.K. & JOHNSON, R.A. (1968). Non-parametric Tests for Shift at an Unknown Time Point. Ann. Math. Stat., 39, 1731-43.

COBB, G.W. (1978). The Problem of the Nile: Conditional Solution to a Change-Point Problem. Biometrika, 65, 243-51.

DARKHOVSHK, B.S. (1976). A Non-Parametric Method for the *a posteriori* Detection of the "Disorder" Time of a Sequence of Independent Random Variables. Theory of Prob. and Applic., 21, 178-83.

DVORETZKY, A., KIEFER, J., & WOLFOWITZ, J. (1956). Asymptotic Minimax Character of the Sample Distribution Function and of the Classical Multinomial Estimator. Ann. Math. Stat., 27, 642-69.

HINKLEY, D.V. (1970). Inference about the Change-Point in a Sequence of Random Variables. Biometrika, 57, 1-16.

HINKLEY, D.V. (1972). Time-Ordered Classification. Biometrika, 59, 509-23.

HINKLEY, D.V. & HINKLEY, E.A. (1970). Inference About the Change-Point in a Sequence of Binomial Variables. Biometrika, 57, 477-88.

HOEFFDING, W. (1963). Probability Inequalities for Sums of Bounded Random Variables. J.A.S.A., 58, 13-30.

SHABAN, S.A. (1980). Change-Point Problem and Two-Phase Regression: An Annotated Bibliography. Intern. Stat. Rev., 48, 83-93.

SMITH, A.F.M. (1975). A Bayesian Approach to Inference about a Change-Point in a sequence of Random Variables. Biometrika, 62, 407-16.

SMITH, A.F.M. (1980). Change-Point problems: approaches and applications. Trabajos Estadistica, 31, 83-98.