THE INSTITUTE OF STATISTICS

THE CONSOLIDATED UNIVERSITY OF NORTH CAROLINA

NONPARAMETRIC CHANGE-POINT ESTIMATION

Edward Carlstein
University of North Carolina
Chapel Hill, North Carolina 27514

Mimeo Series #1573

January 1985

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

NONPARAMETRIC CHANGE-POINT ESTIMATION

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 distribution F for $i \leq \theta n$ and distribution G otherwise. An estimator of the change-point $\theta \in (0,1)$ is proposed. The estimator, which is itself based upon a density estimate, 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 central tendency). The only requirements are that F and G satisfy certain regularity conditions, and that their corresponding densities differ on a set of nonzero Lebesgue measure. Almost sure convergence of the estimator towards θ is proved.

Some key words: Change-point; Density Estimation; Nonparametric.

1. INTRODUCTION

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

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

$X_{[\theta n]+1}^n, \ldots, X_n^n$ identically distributed with distribution 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), 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. Cobb's (1978) conditional solution also requires F and G to be known. 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. Smith (1975) takes a Bayesian approach also requiring assumptions of type (i).

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 some of the 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 (beyond some regularity conditons); and virtually any salient difference between F and G will ensure detection of the change-point (asymptotically). Specifically, we assume:

(I) F and G are both absolutely continuous, with corresponding densities f and g that are uniformly continuous on $\mathbb{R}$;

(II) the set $\Lambda = \{x \in \mathbb{R}: |f(x) - g(x)| > 0\}$ has nonzero Lebesgue measure.

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 density estimate ${}_t h^n(x)$, which is constructed as if $X_1^n, \ldots, X_{[tn]}^n$ were identically distributed, and the post-t density estimate $h_t^n(x)$, which is constructed as if $X_{[tn]+1}^n, \ldots, X_n^n$ were identically distributed.

The former estimates the unknown mixture density:

$$ {}_t h(x) = \mathbb{I}\{t \le \theta\}\ f(x) + \mathbb{I}\{t > \theta\}\ (\theta\, f(x) + (t-\theta)\, g(x))/t\ , $$

and the latter similarly estimates:

$$ h_t(x) = \mathbb{I}\{t \le \theta\}\ ((\theta-t)\, f(x) + (1-\theta)\, g(x))/(1-t) + \mathbb{I}\,\{t > \theta\}\, g(x)\ . $$

The integrated absolute difference between these two densities is:

$$ \begin{aligned} J_\mu(t) &= \int_{-\infty}^{\infty} |{}_t h(x) - h_t(x)|\ d\mu(x) \\ &= (\mathbb{I}\{t \le \theta\}\ (1-\theta)/(1-t) + \mathbb{I}\{t > \theta\}\,\theta/t)\, J_\mu(\theta)\ , \end{aligned} $$

where μ is any non-negative measure not depending on t. Note that $J_\mu(t)$ attains its maximum (over $t \in (0,1)$) at $t=\theta$. Thus a reasonable estimator for θ is the value of t that maximizes:

$$ H_\mu(t) = \int_{-\infty}^{\infty} |{}_t h^n(x) - h_t^n(x)|\ d\mu(x)\ . $$

Although in principle many choices of μ are possible, here μ will be the empirical probability measure $\mu((-\infty,x]) = \sum_{i=1}^{n} \mathbb{I}\,\{X_i^n \le x\}/n$. In practice this means that for fixed t we only have to evaluate ${}_t h^n(\cdot) - h_t^n(\cdot)$ at the n sample points. If we had chosen μ to be Lebesgue measure, we would have to actually calculate (or approximate) the area between the two density estimates. A fixed grid of finitely many weighted x-values is another possibility for μ, but it seems more natural to let the sample determine the x-values.

The parameterization of the change-point in terms of $\theta \in (0,1)$ provides an increasing sample size on both sides of the change. Hence the collections $\{X_1^n, \ldots, X_n^n\}$ and $\{X_1^m, \ldots, X_m^m\}$ $(n < m)$ may be thought of as containing n variables in common (these occurring towards the "middle" of the latter collection, and having entirely different subscripts there than in the former collection).

Finally, no attempt is made here to discuss in general the subject of density estimation. For this particular application, it is convenient to use a framework similar to that of Nadaraya (1965). Hinkley (1972) suggests substituting a density estimate for the true density in his likelihood-based change-point analysis. For a bibliography of density estimation, see Wertz & Schneider (1979).

2. THE ESTIMATOR

In order to formalize the ideas of Section 1, the following additional definitions and assumptions are introduced.

Both ${}_t h^n(\cdot)$ and $h_t^n(\cdot)$ are kernel density estimates, for $t \in (0,1)$:

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

$$ h_t^n(x) = \sum_{i=[tn]+1}^{n} K((x-X_i^n)/b(n-[tn]))/(n-[tn])\, b(n-[tn]) \,, $$

where: $K(\cdot)$ is a density function with bounded variation and continuous derivative on $\mathbb{R}$; $\{b(n):\ n \geq 1\}$ is such that $b(n) > 0$ and $b(n) \downarrow 0$ as $n \to \infty$. The integrated absolute difference between these density estimates is:

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

which is itself an estimate of:

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

Since the accuracy of the approximation of $J_n(t)$ by $H_n(t)$ depends in turn upon the accuracy of ${}_th^n(\cdot)$ and $h_t^n(\cdot)$ in approximating ${}_th(\cdot)$ 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 density estimates. Specifically, we only consider $t \in [\alpha_n, 1-\alpha_n]$, where $\{\alpha_n: n \geq 1\}$ is such that $\alpha_n n \to \infty$ and $\alpha_n \downarrow 0$ as $n \to \infty$. The change-point estimator is then defined as:

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

In practice, α_n prohibits estimation of a change that is too close to either end of the data string. The constraint $\alpha_n n \to \infty$ requires the minimum acceptable sample size to grow with n, thus ensuring increasing precision (in estimating $J_n(t)$ with $H_n(t)$) 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 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.

3. CONSISTENCY

For θ_n defined as in Section 2, and for $\theta \in (0,1)$, the following consistency result holds. Its proof comprises the rest of this Section.

THEOREM. If (I) and (II) of Section 1 hold, and $\{\alpha_n\}$ and $\{b(n)\}$ satisfy the additional condition:

$$\sum_{n=1}^{\infty} n^2 \exp\{-\gamma\alpha_n n b^2(n)\} < \infty \quad \text{for all} \quad \gamma > 0\,;$$

then $\theta_n \to \theta$ almost surely as $n \to \infty$.

(Selecting $\alpha_n = n^{-\tau}$, $0 < \tau < 1$, and $b(n) = n^{-\beta}$, $0 < \beta < \frac{1}{2}(1-\tau)$, fulfills all the requirements on these sequences.)

PROOF. Since θ_n is calculated by maximizing $H_n(t)$ over $t \in [\alpha_n, 1-\alpha_n]$, it is not surprising that we must show the convergence of ${}_t h^n(x)$ and $h_t^n(x)$ towards ${}_t h(x)$ and $h_t(x)$ to be uniform in t (as well as in x). The convergence of these density estimates is most easily dealt with by breaking them up into sub-density-estimates that are based on identically distributed variables (i.e. variables that are all on the same side of θ). The sub-density-estimates may be expected to converge to either $f(\cdot)$ or $g(\cdot)$ (rather than a mixture density like ${}_t h(\cdot)$ or $h_t(\cdot)$). For obtaining convergence of the sub-density-estimates, uniformly in t and x, the following Lemma will be applied. (Proof of the Lemma is deferred to Section 4.)

LEMMA. Let $Y_1^n, \ldots, Y_n^n$ be independent and identically distributed with absolutely continuous distribution Q and density q that is uniformly continuous. Let $K(\cdot)$, $\{\alpha_n\}$, $\{b(n)\}$ satisfy all the requirements for the Theorem. Let

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

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

Then:

$$\sup_{(r,s,\ell,m)\in B_n} \sup_{x\in\mathbb{R}} |q_n(x;r,s,\ell,m) - (r-s)\,q(x)/(\ell-m)| \to 0 \text{ almost}$$

surely as $n \to \infty$.

This Lemma will now be used to show that $H_n(t)$ approximates $J_n(t)$ uniformly well (in t). However, we will restrict the range of t away from θ, so that the sample size is sufficient for those sub-density-estimates based on data between [tn] and [θn]. This constraint will be eliminated later in the argument.

PROPOSITION 1. Let $A_n = [\alpha_n, \theta-\alpha_n] \cup [\theta+\alpha_n, 1-\alpha_n]$. Under the conditions of the Theorem, $\sup_{t\in A_n} |H_n(t) - J_n(t)| \to 0$ almost surely.

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

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

and $|{}_t h^n(X_j^n) - {}_t h(X_j^n)| \le \mathbb{I}\{t \le \theta\}\ |q_n(X_j^n;t,0,t,0) - f(X_j^n)|$ +

$\mathbb{I}\{t > \theta\}\ (|q_n(X_j^n;\theta,0,t,0) - \theta\, f(X_j^n)/t| + |q_n(X_j^n;t,\theta,t,0) - (t-\theta)\ g(X_j^n)/t|)$, where the Y_i^n's in the definition of $q_n(\cdot)$ (see Lemma above) have been replaced with X_i^n's that are identically distributed with either density f or density g. Each of the three terms on the right side of the last inequality can be bounded above by a term of the form:

$$\sup_{(r,s,\ell,m)\in B_n} \sup_{x\in\mathbb{R}} |q_n(x;r,s,\ell,m) - (r-s)q(x)/(\ell-m)| ,$$

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 $\sup_{t\in A_n} (H_n(t) - J_n(t))$ is bounded above by a variable that converges to zero almost surely (by the Lemma). Using analogous logic on $J_n(t) - H_n(t)$ proves Proposition 1.

Still restricting attention to the set A_n, define $\tilde{\theta}_n$ as:

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

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

$t_n \in A_n$ and $J_n(t_n) = \sup_{t\in A_n} J_n(t)$. Thus $\tilde{\theta}_n$ maximizes $H_n(\cdot)$ over A_n, just as θ_n maximizes $H_n(\cdot)$ over $[\alpha_n, 1-\alpha_n]$; and t_n maximizes $J_n(\cdot)$ over A_n, just as θ maximizes $J_n(\cdot)$ over $[\alpha_n, 1-\alpha_n]$. (Note that t_n is deterministic even though $J_n(\cdot)$ is a random function of t.) Since $H_n(\cdot)$ approximates $J_n(\cdot)$ uniformly well (for $t \in A_n$), by Proposition 1, the maximizing value of $J_n(\cdot)$ (namely $J_n(\theta)$) is close to the value of $J_n(\cdot)$ at the t-value which maximizes $H_n(\cdot)$ (namely $t = \tilde{\theta}_n$). This is formalized by:

PROPOSITION 2. Under the conditions of the Theorem,

$$|J_n(\tilde{\theta}_n) - J_n(\theta)| \to 0 \text{ almost surely.}$$

PROOF. $|J_n(\tilde{\theta}_n) - J_n(\theta)| \leq$

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

Proposition 1 eliminates the first term. For the second, observe 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 Proposition 1 applies again. The third term is equal to:

$$(1 - \mathbb{I}\{\theta \leq \tfrac{1}{2}\}\ (1-\theta)/(1-\theta+\alpha_n) - \mathbb{I}\{\theta > \tfrac{1}{2}\}\theta/(\theta+\alpha_n))\ J_n(\theta) .$$

And since $J_n(\theta) \leq \sup_{x\in\mathbb{R}} f(x) + \sup_{x\in\mathbb{R}} g(x) < \infty$, Proposition 2 is proved.

Next we wish to show that Proposition 2 implies almost sure convergence of $\tilde{\theta}_n$ to θ . But to do this we must first establish that for some constant $c > 0$: $J_n(\theta) \geq c$ for all $n \geq n(\omega)$, for almost all $\omega \in \Omega$. To this end denote $Z_j^n = |f(X_j^n) - g(X_j^n)|$, and write:

$$J_n(\theta) = ([\theta n]/n) \sum_{j=1}^{[\theta n]} Z_j^n/[\theta n] + ((n-[\theta n])/n) \sum_{j=[\theta n]+1}^{n} Z_j^n/(n-[\theta n]) .$$

Since $\{Z_j^n : 1 \leq j \leq [\theta n]\}$ are bounded, independent, and identically distributed, and so are $\{Z_j^n : [\theta n]+1 \leq j \leq n\}$, the Strong Law of Large Numbers yields:

$$J_n(\theta) \to \theta \int_{\Lambda\cap\Lambda_1} |f(x)-g(x)|\ f(x)dx + (1-\theta) \int_{\Lambda\cap\Lambda_2} |f(x) - g(x)|\ g(x)dx$$

$$= \rho \quad \text{almost surely ,}$$

where: $\Lambda_1 = \{x \in \mathbb{R}: f(x) > 0\}$, $\Lambda_2 = \{x \in \mathbb{R}: g(x) > 0\}$, and $\rho > 0$. The strict positivity of ρ follows from this logic: If $\rho = 0$, then $m\{\Lambda \cap \Lambda_1\} = m\{\Lambda \cap \Lambda_2\} = 0$ (where $m\{\cdot\}$ is Lebesgue measure). But $\Lambda \subseteq \Lambda_1 \cup \Lambda_2$, so that $\Lambda = (\Lambda \cap \Lambda_1) \cup (\Lambda \cap \Lambda_2)$ and hence $m\{\Lambda\} \leq m\{\Lambda \cap \Lambda_1\} + m\{\Lambda \cap \Lambda_2\} = 0$, which contradicts (II). Putting $c = \frac{1}{2}\rho$ (say) establishes the required boundedness below for $J_n(\theta)$.

Now for almost all $\omega \in \Omega$, there exists for given $\varepsilon > 0$ a $\delta_\varepsilon = c\varepsilon > 0$ such that: $|J_n(\theta) - J_n(t)| < \delta_\varepsilon \Rightarrow |\theta - t| < \varepsilon$, provided $n \geq n(\omega)$ and $t \in (0,1)$. (To see this, observe that when $n \geq n(\omega)$: $|\theta - t| \geq \varepsilon$ implies

$$|J_n(\theta) - J_n(t)| = (\mathbb{I}\{\theta - t \geq \varepsilon\}(\theta - t)/(1-t) + \mathbb{I}\{t - \theta \geq \varepsilon\}(t-\theta)/t)\ J_n(\theta) \geq \varepsilon c) .$$

And by Proposition 2 we can have $|J_n(\theta) - J_n(\tilde{\theta}_n)| < \delta_\varepsilon$ whenever $n \geq N(\varepsilon,\omega)$, for almost all $\omega \in \Omega$. This shows that $\tilde{\theta}_n \to \theta$ almost surely.

Finally we must extend the convergence to apply to the unrestricted estimator θ_n . Consider an ω on which $\tilde{\theta}_n \to \theta$ holds. For given $\varepsilon > 0$ we have $|\tilde{\theta}_n - \theta| < \varepsilon$ for all $n \geq n(\omega,\varepsilon)$. Also $\alpha_n < \varepsilon$ for n sufficiently large. For such n, θ_n either falls within $(\theta - \alpha_n, \theta + \alpha_n)$, or else falls in A_n (in which case $\theta_n = \tilde{\theta}_n$); either way $|\theta_n - \theta| < \varepsilon$.

4. PROOF OF LEMMA

This proof shows how the standard consistency arguments for density estimates (see for example Nadaraya (1965)) may be adapted to obtain uniformity in $(r,s,\ell,m) \in B_n$.

Write $\Delta_n(x;r,s,\ell,m) = |q_n(x;r,s,\ell,m) - \mathbb{E}\{q_n(x;r,s,\ell,m)\}|$ and $V_n = \sup_{(r,s,\ell,m)\in B_n} \sup_{x\in\mathbb{R}} \Delta_n(x;r,s,\ell,m)$. Since $([rn]-[sn])/([\ell n]-[mn]) \le 1$, and $b([\ell n]-[mn]) \ge b(n)$, we have:

$$\Delta_n(x;r,s,\ell,m) \le \left|\int_{-\infty}^{\infty} K((x-u)/b([\ell n]-[mn]))\, dF_n(u;r,s) - \int_{-\infty}^{\infty} K((x-u)/b([\ell n]-[mn]))\, dQ(u)\right|/b(n) ,$$

where $F_n(u;r,s)$ puts mass $1/([rn]-[sn])$ at each of $Y^n_{[sn]+1}, \ldots, Y^n_{[rn]}$. The regularity conditions on Q and K permit integration by parts, yielding: $\Delta_n(x;r,s,\ell,m) \le M\, D_n(r,s)/b(n)$, where M is a finite constant involving the total variation of K, and $D_n(r,s) = \sup_{v\in\mathbb{R}} |F_n(v;r,s) - Q(v)|$. Now, $D_n(r,s)$ depends on r and s only through [rn] and [sn], each of which takes on at most n+1 distinct values as (r,s,ℓ,m) ranges through B_n. Hence, for $\varepsilon>0$:

$$\mathrm{pr}\{V_n > \varepsilon\} \le (n+1)^2 \sup_{(r,s,\ell,m)\in B_n} \mathrm{pr}\{D_n(r,s) > \varepsilon\, b(n)/M\} .$$

Furthermore, by Lemma 2 of Dvoretzky, Kiefer, & Wolfowitz (1956):

$$\begin{aligned}\mathrm{pr}\{D_n(r,s) > \varepsilon\, b(n)/M\} &\le \beta \exp\{-2([rn]-[sn])\varepsilon^2 b^2(n)/M^2\} \\ &\le \beta \exp\{-\gamma_\varepsilon\, \alpha_n n\, b^2(n)\} ,\end{aligned}$$

where β is a finite positive constant and $\gamma_\varepsilon>0$. So $V_n \to 0$ almost surely, by the Borel-Cantelli Lemma.

On the other hand,

$$|\mathbb{E}\{q_n(x;r,s,\ell,m)\} - (r-s)\, q(x)/(\ell-m)| \le$$

$$|\int_{-\infty}^{\infty} K(t/b([\ell n]-[mn]))(q(x-t)-q(x))dt|([rn]-[sn])/([\ell n]-[mn])b([\ell n]-[mn])$$

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

The term within the last absolute value can be bounded by $4/n\alpha_n$, so we only need to deal with:

$$(\int_{|t|\le\delta} K(t/b)\,|\,q(x-t)-q(x)\,|\,dt + \int_{|t|>\delta} K(t/b)\,|\,q(x-t)-q(x)\,|\,dt)/b$$

$$\le \sup_{|t|\le\delta} |q(x-t)-q(x)| + 2A \int_{|t|>\delta/b} K(t)dt\ ,$$

where A is a finite constant bounding $q(\cdot)$, and the argument of $b(\cdot)$ has been suppressed. Let $\varepsilon>0$ be given. By q's uniform continuity, there exists a $\delta_\varepsilon > 0$ such that $\sup_{x\in\mathbb{R}} \sup_{|t|\le\delta_\varepsilon} |q(x-t)-q(x)| < \frac{1}{2}\varepsilon$.

And, for $n \ge n_\varepsilon$ we can have $\alpha_n n$ large enough so that $b = b([\ell n]-[mn]) \le b(\alpha_n n-1)$ is small enough to make $2A \int_{|t|>\delta_\varepsilon/b} K(t)dt < \frac{1}{2}\varepsilon$ for all $(r,s,\ell,m)\in B_n$.

I thank J. Stephen Marron for several helpful discussions relating to this work.

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.

NADARAYA, E.A. (1965). On Non-Parametric Estimates of Density Functions and Regression Curves. Theory of Prob. and Applic., 10, 186-90.

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. <u>Biometrika</u>, 62, 407-16.

WERTZ, W. & SCHNEIDER, B. (1979). Statistical Density Estimation: a Bibliography. <u>Intern. Stat. Rev</u>., 47, 155-75.