

ASYMPTOTIC EFFICIENT ESTIMATION OF THE CHANGE POINT WITH UNKNOWN DISTRIBUTIONS¹

BY Y. RITOV

The Hebrew University of Jerusalem

Suppose X_1, \dots, X_n are distributed according to a probability measure under which X_1, \dots, X_n are independent, $X_i \sim F_0$, for $i = 1, \dots, [\theta_n n]$ and $X_i \sim F^{(n)}$ for $i = [\theta_n n] + 1, \dots, n$, where $[x]$ denotes the integer part of x . In this paper we consider the asymptotic efficient estimation of θ_n when the distributions are not known. Our estimator is efficient in the sense that if $F^{(n)} = F_{\eta_n}$, $\eta_n \rightarrow 0$ and $\{F_\eta\}$ is a regular one-dimensional parametric family of distributions, then the estimator is asymptotically equivalent to the best regular estimator.

1. Introduction. Suppose X_1, \dots, X_n are distributed according to the probability measure P_{θ_n, η_n} under which X_1, \dots, X_n are independent, $X_i \sim F_0$, $i = 1, \dots, [\theta_n n]$, and $X_i \sim F_{\eta_n}$, $i = [\theta_n n] + 1, \dots, n$, where $[x]$ denotes the integer part of x .

To make the problem sensible, we assume that $\mathcal{F} = \{F_\eta: \eta \in (-1, 1)\}$ is a regular parametric model, F_η has density f_η , $\eta_n \rightarrow 0$ but $\eta_n n^{1/2} \rightarrow \infty$. Thus, the change point can be estimated consistently, while the number of observations between θ_n and its estimator converges to infinity as $n \rightarrow \infty$. The family \mathcal{F} may be known or unknown.

The nonparametric estimation of the change point was considered recently by Carlstein (1988). His estimators were based on the maximization of a distance between the empirical distribution functions of the observations X_1, \dots, X_k and the observations X_{k+1}, \dots, X_n , respectively. Previous work was either based on parametric assumptions [e.g., Hinkley (1970), Hinkley and Hinkley (1970) and Cobb (1978)] or on a given particular functional of the distribution that distinguishes between the distributions before and after the change point [cf. Darkhovshky (1976)].

In the next section we consider the convergence of the preceding experiment, as $n \rightarrow \infty$, to a particular experiment. In the limit experiment, we observe a diffusion process on $(-\infty, \infty)$ with a constant infinitesimal variance σ^2 , a drift $\frac{1}{2}\sigma^2$ on $(-\infty, \tau)$ and a drift $-\frac{1}{2}\sigma^2$ on (τ, ∞) (σ is considered known in this experiment). The efficient estimator of τ in the limit experiment is described in Section 3. This estimator suggests an asymptotically efficient estimator for the original model. It is assumed in these two sections that the model \mathcal{F} as well as $\{\eta_n\}$ are known. This assumption is dropped in the fourth section, where an adaptive estimator is constructed.

Received November 1988; revised December 1989.

¹This research was done while the author visited the Department of Statistics, University of Pennsylvania.

AMS 1980 subject classifications. 62G05, 62G20.

Key words and phrases. Asymptotic efficiency, limit of experiments, regular estimator.

2. The limit experiment. We call a family of distributions $\mathcal{F} = \{F_\eta\}$, $\eta \in (-c, c)$, regular if the following hold:

1. \mathcal{F} is dominated by a σ -finite measure μ .
2. Let $s_\eta = (dF_\eta/d\mu)^{1/2}$. Then $s_\eta: (-c, c) \rightarrow L_2(\mu)$ is Frechet differentiable with derivative \dot{s}_η .
3. $\int (\dot{s}_\eta(x))^2 d\mu(x) < \infty$.
4. The map $\eta \rightarrow \dot{s}_\eta$ is continuous from $(-c, c)$ to $L_2(\mu)$.

We assume the following:

ASSUMPTION 1. $\mathcal{F} = \{F_\eta\}$ is regular.

ASSUMPTION 2. $\eta_n \rightarrow 0, \eta_n^2 n \rightarrow \infty$ as $n \rightarrow \infty$.

ASSUMPTION 3. $\theta_n = \theta_0 + O(\eta_n^{-2} n^{-1})$ and $\theta_0 \in (0, 1)$.

The sample size needed for testing between F_0 and F_{η_n} with the sum of errors bounded away from 0 and 1 is of order η_n^{-2} . For a given $\{\eta_n\}$, the error in the estimation of θ_n is expected, therefore, to be of order η_n^{-2}/n . To avoid the degeneration of the estimation problem as n increases, we consider a local parametrization, $\theta_n = \theta_0 + \tau_n n^{-1} \eta_n^{-2}$, with θ_0 known. In this section as well as in the next, the sequence η_1, η_2, \dots is known, so that τ_n is the only unknown parameter of the n -th problem.

Let $L_n(\tau)$ be the log-likelihood that the change point is $\theta_0 + \tau n^{-1} \eta_n^{-2}$:

$$L_n(\tau) = -\text{sgn}(\tau) \sum_{i \in J_n(\tau)} \psi_n(X_i),$$

where $\psi_n = \log dF_{\eta_n}/dF_0$,

$$J_n(\tau) = \begin{cases} \{i: [n\theta_0] < i \leq [n\theta_0 + \tau \eta_n^{-2}]\} & \text{for } \tau > 0, \\ \{i: [n\theta_0 + \tau \eta_n^{-2}] < i \leq [n\theta_0]\} & \text{for } \tau \leq 0 \end{cases}$$

and $\text{sgn}(\tau) = 1$ if $\tau > 0$, 0 if $\tau = 0$ and -1 if $\tau < 0$. Note that the cardinality of $J_n(c)$ is $(1 + o(1))|c|\eta_n^{-2}$ for any finite c . Let c be a fixed value, $0 < c < \infty$. It follows from the regularity of \mathcal{F} that for any $\tau, |\tau| < c$,

$$\max \left\{ \left| \frac{dF_{\eta_n}}{dF_0}(X_i) - 1 \right| : i \in J_n(\tau) \right\} \rightarrow_p 0$$

[see Roussas (1972), Lemma 3.5.2, page 56]. A Taylor expansion yields

$$L_n(\tau) = -\text{sgn}(\tau) \left\{ \sum_{i \in J_n(\tau)} \left[\frac{dF_{\eta_n}}{dF_0}(X_i) - 1 \right] - \frac{1}{2}(1 + c_n) \sum_{i \in J_n(\tau)} \left[\frac{dF_{\eta_n}}{dF_0}(X_i) - 1 \right]^2 \right\}$$

uniformly in $\tau \in (-c, c)$, where

$$|c_n| \leq \max \left\{ \left| \left(\frac{dF_{\eta_n}}{dF_0}(X_i) \right)^{-2} - 1 \right| : i \in J_n(\tau) \right\} \rightarrow_p 0.$$

Note that τ governs only the number of terms in the partial sum that defines L_n , while the terms themselves are only functions of the observations and the nonstochastic sequence $\{\eta_n\}$. Hence, here and hereafter, uniformity in τ is relatively simple. Since all the approximations in the proofs in Roussas [(1972), pages 54–66] hold uniformly for partial sums, we can conclude that

$$\begin{aligned} L_n(\tau) &= \text{sgn}(\tau) \left\{ 2\eta_n \sum_{i \in J_n(\tau)} \frac{\dot{s}_0}{s_0}(X_i) - \frac{1}{2}(1 + o_p(1))\eta_n^2 |J_n(\tau)| I(F_0; \mathcal{F}) \right\} \\ &= \text{sgn}(\tau) \left\{ 2\eta_n \sum_{i \in J_n(\tau)} \frac{\dot{s}_0}{s_0}(X_i) - \frac{1}{2}\eta_n^2 |J_n(\tau)| I(F_0; \mathcal{F}) + o_p(1) \right\} \end{aligned}$$

uniformly on $(-c, c)$, where $I(F_0; \mathcal{F}) = 4 \int \dot{s}_0^2(x) d\mu(x)$.

By the Donsker theorem [cf. Billingsley (1968), pages 137 and 138] the distribution of the process $L_n(\tau)$, $-c < \tau < c$, under $P_{\theta_0 + \tau_n(\eta_n^2 n)^{-1}, \eta_n}$, $\tau_n \rightarrow \tau_0$, converges weakly to the distribution of the process $L_{\tau_0}(\cdot)$ where $L_{\tau_0}(\cdot)$ is a diffusion process on $(-c, c)$ whose infinitesimal variance equals $I(F_0; \mathcal{F})$ and its drift in $\frac{1}{2}I(F_0; \mathcal{F})$ on $(-c, \tau_0)$, $-\frac{1}{2}I(F_0; \mathcal{F})$ on (τ_0, c) .

The process $L_{\tau_0}(\cdot)$ is the log-likelihood function of the experiment $\{Q_\tau: \tau \in (-\infty, \infty)\}$, where, under Q_τ , the law of the observation $\mathbf{X}(\cdot)$ is the same as the law of $L_\tau(\cdot)$. A uniform convergence of the log-likelihood functions implies the convergence of the experiments in the sense of Le Cam. That is, for any bounded loss function, the estimation of τ_n is, in the limit, as difficult as the estimation of τ_0 in the limit experiment (see next section).

3. Regular estimators. Let the loss function for the estimation of θ_n by d be $\rho(\beta_n(\theta_n - d))$, where $\gamma_n \equiv \beta_n/(n\eta_n^2) \rightarrow \gamma$ and ρ is a bowl-shaped bounded loss function. Since θ_n is not known, it makes sense to restrict the discussion to “regular” estimators—estimators that behave uniformly in a neighborhood of θ_0 . Using the local parametrization, we call a sequence of estimators $\{\hat{\tau}_n\}$ regular, if the limit distribution of $\hat{\tau}_n - \tau_n$ is independent of the particular sequence $\{\tau_n\}$ of true parameter values, as long as $\limsup |\tau_n| < \infty$. This restricts the discussion of the limit experiment to estimators which are invariant under the shift group [δ is invariant if $\delta(g_a \mathbf{X}) = \delta(\mathbf{X}) - a$, where $g_a(\mathbf{X})(t) = \mathbf{X}(t - a)$].

The invariant measure of the shift group is, of course, the Lebesgue measure. Thus $\hat{\tau}$, the best invariant estimator is defined implicitly by

$$\frac{\int \rho(\gamma(\tau - \hat{\tau})) e^{\mathbf{X}(\tau)} d\tau}{\int e^{\mathbf{X}(\tau)} d\tau} = \text{inf!}.$$

The preceding section shows that the estimator defined by

$$(1) \quad \frac{\int \rho(\gamma_n(\tau - \hat{\tau}_n)) e^{L_n(\tau)} d\tau}{\int e^{L_n(\tau)} d\tau} = \inf!$$

is asymptotically efficient in the following sense. The distribution of any regular estimator is asymptotically equal to the convolution of $\hat{\tau}_n$ with some distribution [Le Cam (1986), page 128]. A standard large deviation argument implies that the estimator

$$(2) \quad \hat{\tau}_n = \frac{\int \tau e^{L_n(\tau)} d\tau}{\int e^{L_n(\tau)} d\tau}$$

is asymptotically efficient for $\rho(t) = t^2$.

4. Unknown \mathcal{F} . We are going to describe next an estimator which is asymptotically equivalent to the estimators defined by (1) and (2), but is not based on an a-priori knowledge of \mathcal{F} . For the sake of simplicity, we need, however, a slightly stronger version of Assumption 3:

ASSUMPTION 4. For some $\varepsilon > 0$ and for some $\theta_0 \in (-2\varepsilon, 1 - 2\varepsilon)$, $\theta_n = \theta_0 + O(\eta_n^{-2} n^{-1})$.

Moreover, we assume that θ_0 (i.e., the approximate location of the change) is known. In practice, this does not introduce much difficulty. In some cases, θ_0 is known a-priori, as indeed it is in the example that brought this problem to our attention. In this example, the change is a short delayed reaction to a stimulus given to a monkey at a known time. In other cases, θ_0 can be taken to be any naive estimator. Formally, θ_0 can be one of the estimators suggested by Carlstein (1988) truncated to the grid $\{i[n^{1/2}\tilde{\eta}_n]^{-2}; i = 1, 2, \dots\}$, where $\tilde{\eta}_n$ is the Kolmogorov distance between the empirical distribution functions of $X_1, \dots, X_{[n/2]}$ and the empirical distribution function of $X_{[n/2]+1}, \dots, X_n$.

Let \mathbb{F}_{n0} and \mathbb{F}_{n1} be the empirical distribution functions of the two sample tails

$$\mathbb{F}_{n0}(\cdot) = [n(\theta_0 - \varepsilon)]^{-1} \sum_1^{[n(\theta_0 - \varepsilon)]} \chi(X_i \leq \cdot)$$

and

$$\mathbb{F}_{n1}(\cdot) = [n(1 - \theta_0 - \varepsilon)]^{-1} \sum_{[n(1 - \theta_0 - \varepsilon)]}^n \chi(X_i \leq \cdot),$$

where $\chi(\cdot)$ is the indicator function.

Let $2 > \alpha > 0$,

$$k_n = \min\left\{ [n^{1/2} \|\mathbb{F}_{n1} - \mathbb{F}_{n0}\|_\infty]^{2-\alpha}, [\|\mathbb{F}_{n1} - \mathbb{F}_{n0}\|_\infty^{-2+\alpha}] \right\}$$

and define $x_0^{(n)}, \dots, x_{k_n}^{(n)}$ by

$$x_j^{(n)} = \inf\{x: F_{n0}(x) > jk_n^{-1}\}, \quad j = 0, \dots, k_n.$$

Let $\hat{p}_{1j}^{(n)} = F_{n1}(x_j^{(n)}) - F_{n1}(x_{j-1}^{(n)})$, $j = 1, \dots, k_n$. Define now

$$\hat{\psi}_n(x) = \sum_{j=1}^{k_n} \chi(x_{j-1}^{(n)} < x \leq x_j^{(n)}) \log(k_n \hat{p}_{1j}^{(n)}) \chi(\hat{p}_{1j}^{(n)} > 0)$$

and

$$\hat{L}_n(\theta) = -\text{sgn}(\theta - \theta_0) \sum_{i \in J_n^*(\theta)} \hat{\psi}_n(X_i),$$

where

$$J_n^*(\theta) = \begin{cases} \{i: [n\theta_0] < i \leq [n\theta]\} & \text{for } \theta > \theta_0, \\ \{i: [n\theta] < i \leq [n\theta_0]\} & \text{for } \theta \leq \theta_0. \end{cases}$$

Note that since we do not know the sequence $\{\eta_n\}$, the argument of \hat{L}_n is θ and not the local parameter τ (recall that τ is the argument of L_n). Since all $\hat{p}_{1j}^{(n)} > 0$ with probability converging to 1 and to simplify notation, we ignore hereafter the possibility that the $\hat{p}_{1j}^{(n)} = 0$.

The following proposition shows that we can replace L_n in (1) and (2) by \hat{L}_n and obtain an efficient estimator. That is, we define the estimator implicitly by

$$(3) \quad \frac{\int_{-M_n}^{M_n} \rho(\beta_n(\theta - \hat{\theta}_n)) e^{\hat{L}_n(\theta)} d\theta}{\int_{-M_n}^{M_n} e^{\hat{L}_n(\theta)} d\theta} = \text{inf!}.$$

for an appropriate sequence $M_n \rightarrow \infty$.

PROPOSITION 1. *Suppose Assumptions 1, 2 and 4 hold. Then for any sequence $\{M_n\}$ such that $M_n k_n^{-\alpha/(4-2\alpha)} \rightarrow 0$,*

$$\sup_{|\tau| < M_n} \left| \hat{L}_n(\theta_0 + \tau(\eta_n^2 n)^{-1}) - L_n(\tau) \right| \rightarrow_p 0.$$

PROOF. Let $p_{0j}^{(n)} = F_0(x_j^{(n)}) - F_0(x_{j-1}^{(n)})$, $p_{1j}^{(n)} = F\eta_n(x_j^{(n)}) - F\eta_n(x_{j-1}^{(n)})$ and

$$\bar{L}_n(\tau) = -\text{sgn}(\tau) \sum_{i \in J_n(\tau)} \bar{\psi}_n(X_i),$$

where

$$\bar{\psi}_n = \sum_{j=1}^{k_n} \chi(x_{j-1}^{(n)} < x \leq x_j^{(n)}) \log\left(\frac{p_{1j}^{(n)}}{p_{0j}^{(n)}}\right).$$

Since

$$\eta_n^{-1} \sum_{j=1}^{k_n} \chi(x_{j-1}^{(n)} < X \leq x_j^{(n)}) \log\left(\frac{p_{1j}^{(n)}}{p_{0j}^{(n)}}\right) \rightarrow_p \frac{\dot{f}_0}{f_0}(X).$$

In $L_2(F_0)$, it is enough to prove that

$$\sup_{|\tau| < M_n} \left| \hat{L}_n(\theta_0 + \tau/(\eta_n^2 n)) - \bar{L}_n(\tau) \right| \rightarrow_p 0.$$

Next note that \mathcal{F} regular implies that $\lim_{\eta \rightarrow 0} \eta^{-1} \|F_\eta - F_0\|_\infty = d_0$ for some $d_0 > 0$. Since $\eta_n n^{1/2} \rightarrow \infty$, we obtain that, for any $\nu > 0$, there is $m < \infty$, such that with probability of at least $1 - \nu$,

$$k_n \in \left\{ [n^{1/2} \eta_n d_0]^{2-\alpha} - m, \dots, [n^{1/2} \eta_n d_0]^{2-\alpha} + m \right\} \\ \cup \left\{ [\eta_n d_0]^{-2+\alpha} - m, \dots, [\eta_n d_0]^{-2+\alpha} + m \right\}.$$

In particular, this implies that we may consider $\{k_n\}$ in the proof as a nonrandom sequence such that

$$(4) \quad k_n^{-1/(2-\alpha)} \min\{n^{1/2} \eta_n d_0, (\eta_n d_0)^{-1}\} \rightarrow 1.$$

Next, note that given \mathbb{F}_{n_0} and \mathbb{F}_{n_1} , $\{\hat{L}_n(\theta_0 + \tau/(\eta_n^2 n)) - \bar{L}_n(\tau) : |\tau| < M_n\}$ is a simple partial sum process with independent terms, $\log(\hat{\psi}_n(X_i)/\bar{\psi}_n(X_i))$. Since the number of the terms for $|\tau| \leq M_n$ is $O(M_n \eta_n^{-2})$, the result follows from Kolmogorov inequality [cf. Shorack and Wellner (1986), page 843] if each term has, given \mathbb{F}_{n_0} and \mathbb{F}_{n_1} , conditional first and second moments which are equal to $o_p(M_n^{-1} \eta_n^2)$. We will prove that these are indeed $O_p(\eta_n^{2-\alpha/2} n^{-\alpha/4})$.

We begin with some preliminaries. Clearly,

$$(5) \quad \max_j |\hat{p}_{1j}^{(n)} - p_{1j}^{(n)}| = O_p(n^{1/2}), \quad \max_j |p_{0j}^{(n)} - k_n^{-1}| = O_p(n^{1/2}).$$

Hence,

$$(6) \quad \max_j |k_n p_{0j}^{(n)} - 1| = o_p(1).$$

By a standard quantile theory,

$$(7) \quad |k_n E p_{0j}^{(n)} - 1| < k_n [n(\theta_0 - \varepsilon)]^{-1}, \\ |nk_n \text{Var}(p_{0j}^{(n)}) - 1| < 2k_n [n(\theta_0 - \varepsilon)]^{-1}.$$

Combining equations (6) and (7), we obtain

$$(8) \quad \sum_{j=1}^{k_n} (k_n p_{0j}^{(n)} - 1)^2 p_{0j}^{(n)} = (1 + o_p(1)) k_n^{-1} \sum_{j=1}^{k_n} (k_n p_{0j}^{(n)} - 1)^2 \\ = O_p(k_n n^{-1}).$$

We develop now some bounds on the distance between the discretized probabilities function. We begin with

$$\begin{aligned}
 D_n &\equiv \sup_{|\eta| < |\eta_n|} \max_j \left| \int_{x_{j-1}^{(n)}}^{x_j^{(n)}} f_\eta(x) dx - \int_{x_{j-1}^{(n)}}^{x_j^{(n)}} f_0(x) dx \right| \\
 &= \sup_{|\eta| < |\eta_n|} \max_j \left| \eta \int_{x_{j-1}^{(n)}}^{x_j^{(n)}} \dot{f}_\eta(x) dx \right| \\
 &\leq \sup_{|\eta| < |\eta_n|} \max_j |\eta| \left[\int_{x_{j-1}^{(n)}}^{x_j^{(n)}} \frac{(\dot{f}_\eta)^2}{f_\eta}(x) dx \right]^{1/2} \left[\int_{x_{j-1}^{(n)}}^{x_j^{(n)}} f_\eta(x) dx \right]^{1/2} \\
 &= o_p(|\eta_n| (k_n^{-1} + D_n)^{1/2}).
 \end{aligned}$$

Solving this relation for D_n yields $D_n = o_p(1) |\eta_n| (\eta_n^2 + k_n^{-1})$. Hence,

$$\sup_{|\eta| < |\eta_n|} \max_j \left| \int_{x_{j-1}^{(n)}}^{x_j^{(n)}} \frac{f_\eta(x) dx}{p_{0j}^{(n)}} - 1 \right| = o_p(1).$$

The distance between the discrete probability measures $(p_{01}^{(n)}, \dots, p_{0k_n}^{(n)})$ and $(p_{11}^{(n)}, \dots, p_{1k_n}^{(n)})$ can be bounded now by

$$\begin{aligned}
 &\eta_n^{-2} \sum_{j=1}^{k_n} \frac{(p_{1j}^{(n)} - p_{0j}^{(n)})^2}{p_{0j}^{(n)}} \\
 (9) \quad &= \eta_n^{-2} \sum_{j=1}^{k_n} \frac{\left\{ \int_{x_{j-1}^{(n)}}^{x_j^{(n)}} (f_{\eta_n}(x) - f_0(x)) dx \right\}^2}{\int_{x_{j-1}^{(n)}}^{x_j^{(n)}} f_0(x) dx} \\
 &= O_p(1).
 \end{aligned}$$

A useful bound of the difference of $\hat{p}_{1j}^{(n)}$ and $p_{1j}^{(n)}$ is given by

$$\begin{aligned}
 &E \left\{ \sum_{j=1}^{k_n} \frac{(\hat{p}_{1j}^{(n)} - p_{1j}^{(n)})^2}{p_{0j}^{(n)}} \middle| \mathbb{F}_{n0} \right\} \\
 &\leq [n(1 - \theta_0 - \varepsilon)]^{-1} \sum_{j=1}^{k_n} \frac{p_{1j}^{(n)}}{p_{0j}^{(n)}} \\
 (10) \quad &= [n(1 - \theta_0 - \varepsilon)]^{-1} \left\{ k_n + \sum_{j=1}^{k_n} \frac{p_{1j}^{(n)} - p_{0j}^{(n)}}{p_{0j}^{(n)}} \right\} \\
 &\leq [n(1 - \theta_0 - \varepsilon)]^{-1} \left\{ k_n + \left(\sum_{j=1}^{k_n} \frac{(p_{1j}^{(n)} - p_{0j}^{(n)})^2}{p_{0j}^{(n)}} \right)^{1/2} \left(\sum_{j=1}^{k_n} \frac{1}{p_{0j}^{(n)}} \right)^{1/2} \right\} \\
 &= O_p(n^{-1} k_n (1 + \eta_n))
 \end{aligned}$$

by (6) and (9). Note that (4), (5) and (9) imply that

$$\begin{aligned} \min_j p_{1j}^{(n)} &\geq \min_j p_{0j}^{(n)} - (1 + o_p(1))O_p(\eta_n) \max_j (p_{0j}^{(n)})^{1/2} \\ &= O_p(k_n^{-1} - \eta_n k_n^{-1/2}) \\ &= O_p(k_n^{-1}). \end{aligned}$$

Hence, (5) implies that $\max_j |\hat{p}_{1j}^{(n)}/p_{1j}^{(n)} - 1| = o_p(1)$, and therefore,

$$(11) \quad \sup_x \left| \frac{\hat{\psi}_n(x)}{\psi_n(x)} - 1 \right| = \max_j \left| \frac{\hat{p}_{1j}^{(n)}}{p_{1j}^{(n)}} - 1 \right| \rightarrow_p 0.$$

We begin with the first moment. We obtain from (11) that

$$(12) \quad 0 \geq \log \frac{\hat{\psi}_n}{\psi_n}(x) - \left\{ \frac{\hat{\psi}_n}{\psi_n}(x) - 1 \right\} \geq \frac{1}{2} (1 + o_p(1)) \left\{ \frac{\hat{\psi}_n}{\psi_n}(x) - 1 \right\}^2.$$

Let $X \sim F\eta_n$. Then,

$$\begin{aligned} (13) \quad E \left\{ \frac{\hat{\psi}_n}{\psi_n}(X) - 1 \middle| \mathbb{F}_{n0}, \mathbb{F}_{n1} \right\} &= \sum_{j=1}^{k_n} (k_n \hat{p}_{1j}^{(n)} p_{0j}^{(n)} - \hat{p}_{1j}^{(n)}) \\ &= \sum_{j=1}^{k_n} \hat{p}_{1j}^{(n)} (k_n p_{0j}^{(n)} - 1) \\ &= \sum_{j=1}^{k_n} (\hat{p}_{1j}^{(n)} - k_n^{-1}) (k_n p_{0j}^{(n)} - 1) \\ &= \sum_{j=1}^{k_n} (\hat{p}_{1j}^{(n)} - p_{1j}^{(n)}) (k_n p_{0j}^{(n)} - 1) \\ &\quad + \sum_{j=1}^{k_n} (p_{1j}^{(n)} - p_{0j}^{(n)}) (k_n p_{0j}^{(n)} - 1) \\ &\quad + \sum_{j=1}^{k_n} (p_{0j}^{(n)} - k_n^{-1}) (k_n p_{0j}^{(n)} - 1). \end{aligned}$$

We can bound the first term on the RHS of (13) by the Cauchy-Schwartz inequality, (8) and (10),

$$\left\{ \sum_{j=1}^{k_n} \frac{(\hat{p}_{1j}^{(n)} - p_{1j}^{(n)})^2}{p_{0j}^{(n)}} \right\}^{1/2} \left\{ \sum_{j=1}^{k_n} (k_n p_{0j}^{(n)} - 1)^2 p_{0j}^{(n)} \right\}^{1/2} = O_p(k_n n^{-1}).$$

The second term on the RHS of (13) can be bounded again by the

Cauchy–Schwartz inequality, (8) and (9),

$$\left\{ \sum_{j=1}^{k_n} \frac{(p_{1j}^{(n)} - p_{0j}^{(n)})^2}{p_{0j}^{(n)}} \right\}^{1/2} \left\{ \sum_{j=1}^{k_n} (k_n p_{0j}^{(n)} - 1)^2 p_{0j}^{(n)} \right\}^{1/2} = O_p(\eta_n k_n^{1/2} n^{-1/2}).$$

Finally, the last term on the RHS of (13) is equal to

$$k_n^{-1} \sum_{j=1}^{k_n} (k_n p_{0j}^{(n)} - 1)^2 = O_p(k_n n^{-1})$$

[compare to (8)]. We conclude that

$$(14) \quad E_{F\eta_n} \left\{ \frac{\hat{\psi}_n}{\psi_n}(X) - 1 \middle| \mathbb{F}_{n0}, \mathbb{F}_{n1} \right\} = O_p(\eta_n k_n^{1/2} n^{-1/2}).$$

Having (12) and (14) being established, we may consider

$$(15) \quad E_{F\eta_n} \left\{ \left(\frac{\hat{\psi}_n}{\psi_n}(X) - 1 \right)^2 \middle| \mathbb{F}_{n0}, \mathbb{F}_{n1} \right\} = \sum_{j=1}^{k_n} \left(\frac{k_n \hat{p}_{1j}^{(n)} p_{0j}^{(n)}}{p_{1j}^{(n)}} - 1 \right)^2 p_{1j}^{(n)};$$

but,

$$(16) \quad E \left\{ \sum_{j=1}^{k_n} \left(\frac{k_n \hat{p}_{1j}^{(n)} p_{0j}^{(n)}}{p_{1j}^{(n)}} - 1 \right)^2 p_{1j}^{(n)} \middle| \mathbb{F}_{n0} \right\} = \sum_{j=1}^{k_n} (k_n p_{0j}^{(n)} - 1)^2 p_{1j}^{(n)} + \sum_{j=1}^{k_n} \frac{(k_n p_{0j}^{(n)})^2}{p_{1j}^{(n)}} \frac{p_{1j}^{(n)}(1 - p_{1j}^{(n)})}{[n(1 - \theta_0 - \varepsilon)]}.$$

By (6) and (9), for some $C > 0$,

$$\max_j p_{1j}^{(n)} \leq \max_j p_{0j}^{(n)} + C(1 + o_p(1))\eta_n = O_p(k_n^{-1}).$$

Hence, (7) implies that the first term on the RHS of (16) is $O_p(k_n n^{-1})$. The second term is clearly $O_p(k_n n^{-1})$ by (6). Hence,

$$(17) \quad E \left\{ \sum_{j=1}^{k_n} \left(\frac{k_n \hat{p}_{1j}^{(n)} p_{0j}^{(n)}}{p_{1j}^{(n)}} - 1 \right)^2 p_{1j}^{(n)} \middle| \mathbb{F}_{n0} \right\} = O_p(k_n n^{-1}).$$

Finally, combining (4), (12), (14) and (17) enables us to conclude that

$$(18) \quad E_{F\eta_n} \log \frac{\hat{\psi}_n}{\psi_n}(X) = O_p(k_n^{1/2} \eta_n n^{-1/2} + k_n n^{-1}) = O_p(n^{-\alpha/4} \eta_n^{2-\alpha/2}).$$

Similarly, we can prove that

$$(19) \quad E_{F_0} \log \frac{\hat{\psi}_n}{\psi_n}(X) = O_p(n^{-\alpha/4} \eta_n^{2-\alpha/2}).$$

To establish the proper order of the second moment, note that

$$(20) \quad 1 - \frac{\bar{\psi}}{\psi}(X) < \log \frac{\hat{\psi}_n}{\psi_n}(X) < \frac{\hat{\psi}_n}{\psi_n}(X) - 1;$$

but it follows from (11) that

$$\frac{1 - \bar{\psi}(x)/\hat{\psi}(x)}{\hat{\psi}(x)/\bar{\psi}(x) - 1} = \frac{\bar{\psi}(x)}{\hat{\psi}(x)} = 1 + o_p(1),$$

uniformly in x , $x_{j-1}^{(n)} < x \leq x_j^{(n)}$. Therefore, (20) implies that

$$(21) \quad E \left\{ \left| \log \frac{\hat{\psi}_n}{\psi_n}(X) \right|^2 \middle| \mathbb{F}_{n0}, \mathbb{F}_{n1} \right\} \leq (1 + o_p(1)) E \left\{ \left| \frac{\hat{\psi}_n}{\psi_n}(X) - 1 \right|^2 \middle| \mathbb{F}_{n0}, \mathbb{F}_{n1} \right\} \\ = O_p(k_n n^{-1}).$$

Together with (18) and (19) this establishes the moments conditions. The proposition follows. \square

Since \mathcal{F} is unknown, one may ask about the ‘‘uniformity’’ of the estimator. In view of the results in Ritov and Bickel (1990), it is clear that no estimator can be uniformly adaptive. Assume, however, that the following hold.

ASSUMPTION 5. $\lim_{\eta \rightarrow 0} \eta^{-1} \|F_\eta - F_0\| \geq c > 0$.

ASSUMPTION 6. There is a sequence $a_n \rightarrow 0$, such that if

- (i) $k_n \leq \eta_n^{-2+\alpha} \vee (n^{1/2} \eta_n)^{2-\alpha}$,
- (ii) $-\infty = x_0^{(n)} \leq x_1^{(n)} \leq \dots \leq x_{k_n}^{(n)} = \infty$,
- (iii) $\lim_{n \rightarrow \infty} \max_j |k_n \{F_0(x_j^{(n)}) - F_0(x_{j-1}^{(n)})\} - 1| = 0$,

then

$$\int \left\{ \frac{\dot{f}_0}{f_0}(x) - \frac{1}{\eta_n} \sum_{j=1}^{k_n} \chi(x_{j-1}^{(n)} < x \leq x_j^{(n)}) \right. \\ \left. \times \log \frac{F\eta_n(x_j^{(n)}) - F\eta_n(x_{j-1}^{(n)})}{F_0(x_j^{(n)}) - F_0(x_{j-1}^{(n)})} \right\}^2 dF_0(x) < a_n.$$

The proof of Proposition 1 establishes the following.

PROPOSITION 2. *The conclusion of Proposition 1 holds uniformly for all \mathcal{F} , $\{\eta_n\}$ and $\{\theta_n\}$ satisfying Assumptions 1–6 for a given c and $\{a_n\}$.*

Acknowledgment. I wish to thank Dr. Shabtai Barash for the discussion that led to this work.

REFERENCES

- BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
- CARLSTEIN, E. (1988). Nonparametric change-point estimation. *Ann. Statist.* **16** 188–197.
- COBB, G. W. (1978). The problem of the Nile: Conditional solution to the change point problem. *Biometrika* **65** 342–251.
- DARKHOVSHKY, B. S. (1976). A non-parametric method for the posteriori detection of the “disorder” time of a sequence of independent random variables. *Theory Probab. Appl.* **21** 178–183.
- HINKLEY, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika* **57** 1–16.
- HINKLEY, D. V. and HINKLEY, E. A. (1970). Inference about the change-point in a sequence of binomial variables. *Biometrika* **57** 477–488.
- LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.
- RITOV, Y. and BICKEL, P. J. (1990). Achieving information bounds in non and semiparametric models. *Ann. Statist.* **18** 925–938.
- ROUSSAS, G. G. (1972). *Contiguity of Probability Measures: Some Applications in Statistics*. Cambridge Univ. Press, Cambridge.
- SHORACK, G. R. and WELLNER, J.A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

DEPARTMENT OF STATISTICS
HEBREW UNIVERSITY
JERUSALEM 91905
ISRAEL