On Estimating Change Point in a Failure Rate

by

A.P. Basu *
University of Missouri - Columbia
J.K. Ghosh
Indian Statistical Institute
and
S.N. Joshi
Purdue University

Technical Report #86-42

Department of Statistics
Purdue University

1986

* The research of the first author was partly carried out at the Indian Statistical Institute, Calcutta. This research was supported in part by a grant from the Graduate Research Council of the Graduate School, University of Missouri - Columbia, in part by the National Science Foundation under grant No. INT-8615383, and in part by UMC multipurpose Arthritis Center grant.

◇

On Estimating Change Point in a Failure Rate

by

A.P. Basu
University of Missouri - Columbia

J.K. Ghosh
Indian Statistical Institute

and

S.N. Joshi
Purdue University

ABSTRACT

Let F be a life distribution function (d.f.) with density f and failure rate r. It is assumed that f is the first part of a "bath-tub" model, that is, $r(t)$ is nonincreasing for $t \leq \tau$ and is constant for $t > \tau$.

In this paper the problem of estimating the change point or threshold τ has been considered. Two estimates for τ have been proposed and their consistency have been proved.

Nguyen, Rogers and Walker [1984] considered a specific parametric case where, with $I(A)$ denoting the indicator function of A, $r(t) = a\ I\ (0 \leq t \leq \tau) + b\ I(t > \tau)$, and proposed a consistent estimate. We have obtained the asymptotic distribution of their estimate using a new method which may have applications to other problems. We also propose a maximum likelihood estimate restricted to lie in a suitable compact set.

We report some simulations comparing the performance of these four estimates.

INTRODUCTION

In reliability theory a widely accepted procedure is to apply "burn-in" techniques to screen out defective items and improve the lifetimes of remaining surviving units.

Formally, let $T_1, T_2, \ldots, T_n$ be a random sample from a lifetime distribution with d.f. $F(t)$ and density $f(t)$. The hazard rate $r(t)$ is defined as

$$r(t) = f(t)/\overline{F}(t) \text{ where } \overline{F}(t) = 1 - F(t).$$

We assume that $r(t)$ is a truncated "bath-tub" model i.e.

$$\begin{aligned} r(t) &= \lambda(t) \quad \text{if} \quad 0 \leq t \leq \tau \\ &= \lambda_0 \quad \text{if} \quad t > \tau \end{aligned} \tag{1.1}$$

where $\lambda(t)$ is nonincreasing and $\lambda(\tau) \geq \lambda_0$ with equality only if $\lambda(t)$ is strictly decreasing in $(\tau - \delta, \tau]$ for some $\delta > 0$. We wish to estimate the threshold τ. If one knew τ, items could be tested up to time τ and only survivors sold. This would be one way of screening. In our experience, screening in such situations is usually provided in a different way by subjecting items to a shock, thermal or electrical, and selling only survivors.

In Section 2 we propose two estimates for τ and prove their consistency.

It is of interest to study how our semiparametric estimates perform in specific parametric models. Nguyen et. al. [1984], hence forth abbreviated as NRW, have considered such a model, namely,

$$r(t) = a\ I(0 \leq t \leq \tau) + b\ I(t > \tau) \tag{1.2}$$

and proposed a consistent estimate for τ (when $a > b$, (1.2) is a special case of (1.1)). We have also introduced a restricted maximum likelihood estimate (m.l.e.) for purpose of comparison.

We carried out some simulations in Section 4 for model (1.2) (with $a > b$) for various values of the parameters a, b, and τ. When $F(\tau)$ is small i.e. when change takes place early in the lifetime our estimates of Section 2 perform well as compared to the NRW estimate or the m.l.e. .

In Section 3 we have obtained the asymptotic distribution of NRW estimate of τ. Our method for getting the asymptotic distribution would also apply to M-estimates with kernels that do not satisfy the usual conditions of differentiability or monotonicity but possess expectation having properties similar to those of the function $X(\cdot)$ of section 3. It should also be observed that the rescaling technique used in our method is due to Prakasa Rao ([1968], [1986]).

2. TWO NEW ESTIMATES OF τ

In our model it is reasonable to assume that

$$0 < F(\tau) < 1. \tag{2.1}$$

Moreover, an upper bound p_0 to $F(\tau)$ is assumed to be known, $p_0 < 1$; this would be a weak assumption in most practical situations.

Let $F_n(t)$ be the empirical d.f. of $T_1, T_2, \ldots, T_n$ and

$$y_n(t) = -\log \overline{F}_n(t),\ y(t) = -\log \overline{F}(t).$$

Let ξ_p and $\hat{\xi}_p$ denote p-th population and sample quantiles respectively.

Let p_1 be such that $p_0 < p_1 < 1$. Let k be the number of order statistics between $T_{([np_0])}$ and $T_{([np_1])}$ and let

$$\hat{\lambda}_0 = \frac{\sum T_{(i)} \log \overline{F}_n(T_{(i)})/(k+1) - (\sum T_{(i)}/(k+1))(\sum \log \overline{F}_n(T_{(i)})/(k+1))}{(\sum T_{(i)}^2/(k+1)) - (\sum T_{(i)}/(k+1))^2}$$

and the summations range over $i = [np_0] + 1$ to $i = [np_1]$.

Under (1.1) for $t > \tau$, $\log \overline{F}(t)$ is linear in t and $\hat{\lambda}_0$ is an ordinary least square estimate of the slope λ_0 treating $T_{(i)}$ as independent and $\log \overline{F}_n(T_{(i)})$ as dependent variables. It is well known (vide e.g. Serfling [1980] p. 59) that

$$\sqrt{n}\mathrm{Sup}_t|F_n(t) - F(t)| = Op(1), \tag{2.2}$$

and for $0 < p < 1$

$$\sqrt{n}(\hat{\xi}_p - \xi_p) = Op(1). \tag{2.3}$$

Using (2.2) it is easy to see that uniformly in $t \leq \xi_p$, $p < 1$, we have

$$\sqrt{n}(y_n(t) - y(t)) = Op(1). \tag{2.4}$$

Now we claim that

$$\sqrt{n}(\hat{\lambda}_0 - \lambda_0) = O_p(1). \tag{2.5}$$

Note that $\hat{\lambda}_0$ can be expressed as a continuous function of quantities of the form

$$\frac{n}{(k+1)} \cdot \int_{T_{([np_0]+1)}}^{T_{([np_1])}} \phi(F_n(x), x)dF_n(x)$$

each of which can be handled using the following lemma and (2.5) can be proved.

Lemma: Let $T_{in} = \xi_i + O_p(n^{-1/2})$ $i = 1, 2$ and let $\phi(x, y)$ be such that for some $M > 0$ and for some $0 < \delta_1 < \delta_2 < 1$

(i) $|\phi(x_1, y) - \phi(x_2, y)| \leq M|x_1 - x_2|$ for all $\xi_1 \leq y \leq \xi_2$ and all $\delta_1 \leq x_1, x_2 \leq \delta_2$,

(ii) $\phi(x, y)$ is bounded in $\xi_1 \leq y \leq \xi_2$, $\delta_1 \leq x \leq \delta_2$.

Then

$$\int_{T_{1n}}^{T_{2n}} \phi(F_n(x), x)dF_n(x) = \int_{\xi_1}^{\xi_2} \phi(F(x), x)dF(x) + O_p(n^{-1/2}).$$

The proof of the lemma is not hard and hence omitted.

Now we define our estimates of τ. Let $h_n = n^{-\frac{1}{4}}$ and $\varepsilon_n = c(\log n)n^{-\frac{1}{2}}$

$$\hat{\tau}_1 = \ \mathrm{Inf}\ \{t : y_n(t + h_n) - y_n(t) \leq h_n\hat{\lambda}_0 + \varepsilon_n\}$$

and

$$\hat{\tau}_2 = \ \mathrm{Inf}\ \{t : \log \overline{F}_n(t) - \log(1 - p_0) \leq \hat{\lambda}_0(\hat{\xi}_{p_0} - t) + \varepsilon_n\}.$$

To see the motivation for $\hat{\tau}_1$ note that $y_n(t+h_n) - y_n(t))/h_n$ is an estimate of the hazard rate $r(t)$ at t. For each fixed t, we test $H_{0t}: \ r(t) = \lambda_0$ vs $H_{1t}: \ r(t) > \lambda_0$, using the acceptance region $\{y_n(t+h_n) - y_n(t) \le h_n\hat{\lambda}_0 + \epsilon_n\}$. We then estimate τ as the smallest t for which H_{0t} is accepted. Formally $\hat{\tau}_1$ is as given above. Motivation of $\hat{\tau}_2$ is similar.

Theorem 1: Let (1.1) and (2.1) hold. Then $\hat{\tau}_1$ and $\hat{\tau}_2$ are consistent for τ.

Proof: Note that for sufficiently small $\varepsilon > 0$ and for sufficiently large n, $\tau + h_n + \varepsilon < \xi_{p_0}$. Hence using (2.4)

$$\begin{aligned} y_n(\tau+\varepsilon+h_n) - y_n(\tau+\varepsilon) &= y(\tau+\varepsilon+h_n) - y(\tau+\varepsilon) + Op(n^{-\frac{1}{2}}) \\ &= \lambda_0 h_n + Op(n^{-\frac{1}{2}}) \qquad \text{by (1.1), and by (2.5)} \\ &= \hat{\lambda}_0 h_n + Op(n^{-\frac{1}{2}}). \end{aligned} \tag{2.6}$$

Now note that

$$\begin{aligned} &(y_n(\tau+\varepsilon+h_n) - y_n(\tau+\varepsilon) \le \hat{\lambda}_0 h_n + \varepsilon_n) \\ &\Rightarrow (\hat{\tau}_1 \le \tau + \varepsilon). \end{aligned} \tag{2.7}$$

Thus using (2.6) and (2.7) we have

$$P(\hat{\tau}_1 \le \tau + \varepsilon) \to 1. \tag{2.8}$$

Now for sufficiently small $\varepsilon > 0$ we have $\tau - \varepsilon > 0$, hence using (2.4), we have

$$\begin{aligned} y_n(t+h_n) - y_n(t) &= \log \overline{F}(t) - \log \overline{F}(t+h_n) + Op(n^{-\frac{1}{2}}) \\ &\qquad \text{uniformly in } 0 \le t \le \tau - \varepsilon \\ &\ge h_n \lambda(x+h_n) + Op(n^{-\frac{1}{2}}) \\ &\qquad \text{uniformly in } 0 \le t \le \tau - \varepsilon \\ &> h_n\lambda_0 + h_n\delta_\varepsilon + Op(n^{-\frac{1}{2}}) \\ &\qquad \text{uniformly in } 0 \le t \le \tau - \varepsilon \\ &\qquad\qquad \text{where } \delta_\varepsilon \text{ is such that } \lambda(\tau - \varepsilon/2) > \lambda_0 + \delta_\varepsilon \\ &> h_n\lambda_0 + h_n\delta_\varepsilon + O_p(n^{-1/2}) \\ &\qquad \text{uniformly in } 0 \le t \le \tau - \varepsilon \end{aligned} \tag{2.9}$$

Hence

$$P(\hat{\tau}_1 \ge \tau - \varepsilon) \to 1. \tag{2.10}$$

The relations (2.8) and (2.10) prove the consistency of $\hat{\tau}_1$.

Now for sufficiently small $\varepsilon > 0$, $\tau + \varepsilon < \xi_{p_0}$ hence using (2.4), we have

$$\begin{aligned} \log \overline{F}_n(\tau+\varepsilon) - \log(1-p_0) &= \log \overline{F}(\tau+\varepsilon) - \log(1-p_0) + Op(n^{-\frac{1}{2}}) \\ &= \hat{\lambda}_0(\hat{\xi}p_0 - \tau - \varepsilon) + Op(n^{-\frac{1}{2}}), \qquad \text{(by (2.5))}. \end{aligned}$$

Also,

$$(\log \overline{F}_n(\tau+\varepsilon) - \log(1-p_0) \leq \hat{\lambda}_0(\hat{\xi}_{p_0} - \tau - \varepsilon) + \varepsilon_n) \Rightarrow (\hat{\tau}_2 \leq \tau + \varepsilon).$$

Hence

$$P(\hat{\tau}_2 \leq \tau + \varepsilon) \to 1. \tag{2.11}$$

Now, by (2.4), uniformly in $0 \leq x < \tau - \varepsilon$,

$$\begin{aligned}\log \overline{F}_n(x) - \log(1-p_0) &= \log \overline{F}(x) - \log(1-p_0) + Op(n^{-\frac{1}{2}}) \\ &= \int_x^{\xi_{p_0}} \lambda(t)dt + Op(n^{-\frac{1}{2}}) \\ &\geq (\hat{\xi}_{p_0} - x)\hat{\lambda}_0 + \delta + Op(n^{-\frac{1}{2}}) \qquad \text{for some } \delta > 0.\end{aligned}$$

Hence

$$P(\hat{\tau}_2 \geq \tau - \varepsilon) \to 1. \tag{2.12}$$

Consistency of $\hat{\tau}_2$ follows from (2.11) and (2.12).

3. A PARAMETRIC EXAMPLE AND SOME PARAMETRIC ESTIMATES

The density specified by (1.2) has the form

$$\begin{aligned} f(t) &= a\exp(-at)I(0 \leq t \leq \tau) \\ &\quad + b\exp(-a\tau - b(t-\tau))I(t > \tau) \end{aligned} \tag{3.1}$$

which is p.d.f. for (a, b, τ), $0 < a$, $0 < b$, $0 < \tau$. (If $a > b$, this is a special case of (1.1) but we will not assume this now.) If $a = b$ or $\tau = 0$ we have identifiability problems.

Note that τ is a change point not in the usual sense (e.g. Hinkley [1970]) where one has a sequence of parameters θ_t which change from one value θ_0 for $t < \tau$ to another value θ_1 for $t \geq \tau$.

Note that the density in (3.1) can be written as a mixture of a right truncated exponential and an untruncated exponential with the mixing proportion depending on the parameters a and τ.

Let $t_{(1)} \leq t_{(2)} \leq \ldots \leq t_{(n)}$ be the ordered sample. If one sets any arbitrary value to the parameter "a" and sets $\hat{b} = 1/(t_{(n)} - \hat{\tau})$ where $t_{(n-1)} < \hat{\tau} < t_{(n)}$, then the likelihood at $\hat{a}$, $\hat{b}$, $\hat{\tau}$ may be made as large as we please making $\hat{\tau}$ as close to $t_{(n)}$ as needed. One may therefore say that in a sense a maximum likelihood estimate of τ is $\hat{\tau} = t_{(n)}$. Obviously $t_{(n)}$ is not consistent.

If one chooses a compact set of (a, b, τ)'s as the parameter space and imposes identifiability conditions like $\tau \geq \delta_1 > 0$, $|a - b| \geq \delta_2 > 0$ then Wald's general result (vide Wald [1949]) implies that m.l.e. $\hat{\tau}$ is consistent. Since τ is a point of discontinuity of the density, the general theory of Chernoff and Rubin [1956] ensures that $|\hat{\tau} - \tau|$ is in fact $Op(n^{-1})$ (better than usual $Op(n^{-\frac{1}{2}})$).

Either not being aware of the above result or because they do not want to impose any conditions on the parameter space, NRW proposed a new estimate $\hat{\tau}_3$ of τ. Using the n observations $t_1, \ldots, t_n$ they construct a kernel $X_n(t)$ such that the solution of $X_n(t) = 0$ provides a consistent estimate of τ. The construction of $X_n(t)$ is ingenious but apart from providing a consistent estimate the kernel does not seem to have any attractive properties. Existence of a consistent solution is a consequence of the fact that

$$X_n(t) \xrightarrow{\text{a.s.}} X(t)$$

where $X(t)$ is a non-stochastic function, $X(\tau) = 0$ and $X(t)$ is monotone in a neighborhood of τ. They show that $X_n(t) = 0$ has a "consistent" solution $\hat{\tau}_3$ which is their estimate. Since $X_n(t)$ is neither monotone nor sufficiently smooth (e.g., not differentiable at τ), it is hard to get the asymptotic distribution of $\sqrt{n}(\hat{\tau}_3 - \tau)$. In fact NRW fail to find it.

Now we derive the limiting distribution of $\hat{\tau}_3$. Let T be a r.v. having density $f(t)$ of (3.1).

Let

$$\begin{aligned} B_1(T,t) &= T, \\ B_2(T,t) &= I(T > t), \\ B_3(T,t) &= TI(T > t), \\ B_4(T,t) &= T^2 I(T > t). \end{aligned}$$

and

$$\underset{\sim}{B} = (B_1, B_2, B_3, B_4).$$

Let $T_1, \ldots, T_n$ be i.i.d. with density $f(t)$ and

$$\overline{B}_j(t) = n^{-1} \sum_{i=1}^{n} B_j(T_i, t) \quad j = 1, \ldots, 4,$$

$$\begin{aligned} H(\overline{\underset{\sim}{B}}(t)) = {} & (\overline{B}_4(t)/\overline{B}_2(t) - \overline{B}_3^2(t)/\overline{B}_2^2(t))^{\frac{1}{2}}(\overline{B}_2(t) - \overline{B}_2(t)\log \overline{B}_2(t) - 1) \\ & + (1 - \overline{B}_2(t))\overline{B}_3(t)/\overline{B}_2(t) + \overline{B}_1(t)\log \overline{B}_2(t). \end{aligned}$$

Let

$$\begin{aligned} X_n(0) &= 0, \\ X_n(T_{(i)}) &= H(\overline{\underset{\sim}{B}}(T_{(i)})), \quad i = 1, \ldots, n-1 \\ X_n(t) &= X_n(T_{(n-1)}), \qquad \text{for } t \geq T_{(n-1)} \end{aligned}$$

and $X_n(\cdot)$ at other points be defined by linear interpolation.

Let $\hat{\tau}_3$ be defined formally as follows. Choose a $\sqrt{n}$-consistent estimate $\hat{\tau}_c$ of τ and let $\hat{\tau}_3 =$ zero of $X_n(t)$ nearest to $\hat{\tau}_c$. Fix a neighborhood $[\tau_1, \tau_2]$ of τ and consider

$$Y_n(t) = \sqrt{n}(X_n(t) - X(t)), \ \tau_1 \leq t \leq \tau_2$$

where

$$X(t) = H(\underset{\sim}{\mu}(t)), \ \underset{\sim}{\mu}(t) = E(\underset{\sim}{B}(\cdot, t)).$$

It can be checked from the appendix that

(i) $X(\tau) = 0$,

(ii) the derivatives $\dot{X}(t)$ and $\ddot{X}(t)$ exist for $t > \tau$ and $t < \tau$ and are continuous, (3.2)

(iii) $\dot{X}(\tau+)$ and $\dot{X}(\tau-)$ exist and are of the same sign (vide Appendix).

It can be checked using estimates like in (13.6) of Billingsley [1968, p. 104] and bounds on the derivatives of H, that

$$Y_n(t) - \sqrt{n}(H(\overline{\underset{\sim}{B}}(t)) - X(t)) = o_p(1).$$

From this one checks via the delta method that the finite dimensional distributions of $Y_n(\cdot)$ converge to a multivariate normal distribution. Tightness is proved by checking a condition analogous to (13.17) of Billingsley [1968, p. 106]. From these considerations it follows that

$$Y_n(\cdot) \xrightarrow{\text{W}} Y(\cdot)$$

where $Y(\cdot)$ is a zero mean Gaussian process.

Consider now a rescaled process in $C(-\infty, \infty)$ (with the topology of uniform convergence on compacts. For tightness in $C(-\infty, \infty)$, see Sen [1981]). Let

$$\begin{aligned} Z_n(h) &= Y_n(\tau + n^{-\frac{1}{2}} h) && \text{if} \quad |h| \le \log n \\ &= Y_n(\tau + n^{-\frac{1}{2}} \log n) && \text{if} \quad h \ge \log n \\ &= Y_n(\tau - n^{-\frac{1}{2}} \log n) && \text{if} \quad h \le -\log n. \end{aligned}$$

Then

$$Z_n(\cdot) \xrightarrow{\text{W}} Z(\cdot),$$

where

$$Z(\cdot) = Y(\tau). \tag{3.3}$$

Now let

$$\begin{aligned} W_n(h) &= \sqrt{n}\, X_n(\tau + h n^{-\frac{1}{2}}) && \text{if} \quad |h| \le \log n \\ &= \sqrt{n}\, X_n(\tau + n^{-\frac{1}{2}} \log n) && \text{if} \quad h > \log n \\ &= \sqrt{n}\, X_n(\tau - n^{-\frac{1}{2}} \log n) && \text{if} \quad h < -\log n. \end{aligned} \tag{3.4}$$

Then using (3.2), (3.3), and (3.4) it is easy to see that

$$W_n(\cdot) \xrightarrow{\text{W}} A(\cdot),$$

where $A(\cdot)$ is a Gaussian process on $C(-\infty, \infty)$ with the representation

$$A(h) = Y(\tau) + h\dot{X}(\tau+) \quad \text{if} \quad h > 0 \tag{3.5}$$

$$= Y(\tau) - h\dot{X}(\tau-), \quad \text{if} \quad h < 0$$

and

$$A(0) = Y(\tau).$$

Let, for $f \in C(-\infty, \infty)$,

$$A_1(f) = \text{Sup}\{t : f(t) = 0\},$$
$$A_2(f) = \text{Inf}\{t : f(t) = 0\},$$

if f has at least one zero and $A_1(f)$ and $A_2(f)$ equal to a constant, say, c otherwise. Note that A_1 and A_2 are measurable and continuous on a set which contains $A(\cdot)$ with probability one, hence

$$(A_1(W_n) - A_2(W_n)) \xrightarrow{W} (A_1(A) - A_2(A)) \tag{3.6}$$

and $A_i(W_n) \xrightarrow{W} A_i(A)$ for $i = 1, 2$. Using (3.2) and (3.5) it is easy to see that $(A_1(A) - A_2(A))$ is degenerate at zero, hence we have $w.p. \to 1$

$$A_1(W_n) \geq \sqrt{n}(\hat{\tau}_3 - \tau) \geq A_2(W_n). \tag{3.7}$$

Thus $\sqrt{n}(\hat{\tau}_3 - \tau)$ has split normal distribution of $A_1(A)$ (or of $A_2(A)$). It is easy to see that $w.p.1$, and $i = 1, 2$

$$\begin{aligned} A_i(A) &= \frac{-Y(\tau)}{\dot{X}(\tau+)} && \text{if} \quad Y(\tau) \leq 0 \text{ and } \dot{X}(\tau+) > 0 \\ & && \text{or if} \quad Y(\tau) \geq 0 \text{ and } \dot{X}(\tau+) < 0 \\ &= \frac{-Y(\tau)}{\dot{X}(\tau-)} && \text{if } Y(\tau) \geq 0 \text{ and } \dot{X}(\tau-) > 0 \\ & && \text{or if } Y(\tau) \leq 0 \text{ and } \dot{X}(\tau-) < 0. \end{aligned} \tag{3.8}$$

Hence d.f. of $\sqrt{n}(\hat{\tau}_3 - \tau)$ converges weakly to $G(t)$ where for $t > 0$

$$G(-t) = \Phi(-t|\dot{X}(\tau-)| \,/\, V(\tau)),$$

$$1 - G(t) = \Phi(-t|\dot{X}(\tau+)| \,/\, V(\tau)),$$

and

$$\begin{aligned} V^2(\tau) &= \text{Var}(Y(\tau)) \\ &= \sum_{1 \leq i,} \sum_{j \leq 4} \vartheta ij(\tau) \frac{\partial H}{\partial B_i}|_{\underset{\sim}{\mu}(\tau)} \frac{\partial H}{\partial B_i}|_{\underset{\sim}{\mu}(\tau)}. \end{aligned} \tag{3.9}$$

It can be shown that $V^2(\tau) > 0$. See appendix for further details. This completes the derivation of limiting distribution of $\hat{\tau}_3$.

For simulations, we also have compared $\hat{\tau}_4$ which is the m.l.e. with parameter space $\tau \le \delta_1$, $|a-b| \le \delta_2 > 0$.

Estimates of a and b and their limiting distributions: For each $\tau > 0$, formal differentiation of the likelihood function w.r.t. a and b yields $\hat{a}(\tau)$ and $\hat{b}(\tau)$, one can plug in an estimate of τ say $\hat{\tau}_3$ and get $\hat{a}$ and $\hat{b}$ the estimates of a and b respectively, it can be seen (vide NRW) that

$$\begin{aligned}\hat{a} &= (1-\overline{B}_2(\hat{\tau}_3))/(\overline{B}_1(\hat{\tau}_3) - \overline{B}_3(\hat{\tau}_3) + \hat{\tau}_3\overline{B}_2(\hat{\tau}_3)) \\ &= H_1(\underset{\sim}{\overline{B}}, \hat{\tau}_3),\end{aligned}$$

say, and

$$\hat{b} = \overline{B}_2(\hat{\tau}_3)/(\overline{B}_3(\hat{\tau}_3) - \hat{\tau}_3\overline{B}_2(\hat{\tau}_3)).$$

The following is the sketch of the derivation of the limiting distribution of $\hat{a}$. Limiting distribution of $\hat{b}$ can be handled in a similar manner.

Using δ-method we have

$$\begin{aligned}n^{\frac{1}{2}}(\hat{a}-a) &= n^{\frac{1}{2}}(H_1(\underset{\sim}{\overline{B}}(\hat{\tau}_3), \hat{\tau}_3) - H_1(\underset{\sim}{\mu}(\tau), \tau)) \\ &= n^{\frac{1}{2}}\sum_{i=1}^{3}(\overline{B}_i(\hat{\tau}_3) - \mu_i(\tau))\frac{\partial H_1}{\partial \mu_i(\tau)} + n^{\frac{1}{2}}(\hat{\tau}_3 - \tau)\frac{\partial H_1}{\partial \tau} + o_p(1) \\ &= n^{\frac{1}{2}}\sum_{i=1}^{3}(\overline{B}_i(\hat{\tau}_3) - \mu_i(\hat{\tau}_3))\frac{\partial H_1}{\partial \mu_i(\tau)} \\ &\quad + n^{\frac{1}{2}}(\hat{\tau}_3 - \tau)\left[\frac{\partial H_1}{\partial \tau} + \sum_{i=1}^{3}\frac{\partial H_1}{\partial \mu_i(\tau)}\frac{\partial \mu_i(\tau)}{\partial \tau}\right] + o_p(1) \\ &= W_{3n}(\tau) + \mathcal{Q}_{1n}(h_n) - \mathcal{Q}_{1n}(0) + c(\tau)h_n + o_p(1),\end{aligned} \tag{3.10}$$

where

$$\begin{aligned}h_n &= n^{\frac{1}{2}}(\hat{\tau}_3 - \tau), \\ W_{3n}(t) &= n^{\frac{1}{2}}\sum_{i=1}^{3}(\overline{B}_i(t) - \mu_i(t))\frac{\partial H_1}{\partial \mu_i(t)}, \\ \mathcal{Q}_{1n}(h) &= n^{\frac{1}{2}}\sum_{i=1}^{3}(\overline{B}_i(\tau + n^{-\frac{1}{2}}h) - \mu_i(\tau + n^{-\frac{1}{2}}h))\frac{\partial H_1}{\partial \mu_i(\tau)}\end{aligned}$$

and

$$c(\tau) = \frac{\partial H_1}{\partial \tau} + \sum_{i=1}^{3}\frac{\partial H_1}{\partial \mu_i(\tau)}\frac{\partial \mu_i(\tau)}{\partial \tau}.$$

From (3.6), (3.7) and the remark following (3.6), it has been proved earlier that

$$\sqrt{n}(\hat{\tau}_3 - \tau) - A_1(W_n(\cdot)) \xrightarrow{W} \delta,$$

where δ is the measure degenerate at zero. This implies

$$\sqrt{n}(\hat{\tau}_3 - \tau) - A_1(W_n(\cdot)) \xrightarrow{\text{P}} 0.$$

It follows that

$$c(\tau)h_n = A_1(W_n(\cdot))c(\tau) + op(1),$$

and using Theorem 8.2 of Billingsly [1968, p. 55] we have

$$\text{Sup}_{|h|<\log n} |\mathcal{Q}_{1n}(h) - \mathcal{Q}_{1n}(0)| \xrightarrow{\text{P}} 0. \tag{3.11}$$

Thus

$$n^{\frac{1}{2}}(\hat{a} - a) = W_{3n}(\tau) + c(\tau)A_1(W_n(\cdot)) + op(1) \tag{3.12}$$

It is easy to check by the delta method that

$$(W_{3n}(\tau), \sqrt{n}(H(\underset{\sim}{\overline{B}}(\tau)))) \xrightarrow{\text{w}} (X, Y)$$

where (X, Y) has a bivariate normal distribution with mean zero and easily computable dispersion matrix. We may take Y to be equal to $Y(\tau) = A(0)$ without loss of generality; define $A(\cdot)$ as before. We first note that $(W_{3n}(\tau), W_n(\cdot))$ is a sequence of random variables taking values in $R \times C(-\infty, \infty)$, equipped with the product topology. Tightness follows from the tightness of the marginal distribution of $W_{3n}(\tau)$ and $W_n(\cdot)$. It is also easy to see that the joint distribution of $W_{3n}(\tau)$ and $W_n(T_1), \ldots, W_n(t_k)$ converge to that of $(X, A(t_1), \ldots, A(t_k))$. Since these finite dimensional distributions determine the distribution on $R \times C(-\infty, \infty)$, it follows from Prohorov's theorem that

$$(W_{3n}(\tau), W_n(\cdot)) \xrightarrow{\text{w}} (X, A(\cdot)). \tag{3.13}$$

It now follows from (3.8), (3.12) and (3.13) that $\sqrt{n}(\hat{a} - a)$ converges in distribution to $X + c(\tau)A_1(A(\cdot))$.

The limiting distribution can be calculated from bivariate normal tables. For example if $\dot{X}(\tau+) > 0 > \dot{X}(\tau-)$, then for any real "$d$",

$$\begin{aligned} \lim P\{\sqrt{n}(\hat{a} - a) \leq d\} &= P\{X - c(\tau)Y(\tau)/\dot{X}(\tau+) \leq d,\ Y(\tau) \geq 0\} \\ &\quad + P\{X - c(\tau)Y(\tau)/\dot{X}(\tau-) \leq d,\ Y(\tau) < 0\}. \end{aligned}$$

In case $\hat{\tau}_3$ is replaced by $\hat{\tau}_4$ in the estimate for $\hat{a}$, then (3.10) and (3.11) continue to hold and $h_n \xrightarrow{\text{P}} 0$ since $(\hat{\tau}_4 - \tau) = 0_p(1/n)$. It follows that $\sqrt{n}\ (\hat{a} - a)$ has the same limiting normal distribution as $W_{3n}(\tau)$. This last fact has been noted by Nguyen and Pham [1987].

4. SIMULATION RESULTS

We obtained 100 samples each of size 100 and carried out simulations with $p_0 = .50$, $p_1 = .90$. We used the following "smoother" version of $\hat{\tau}_2$ for simulations (the summations below range over $i = [np_0] + 1$ to $i = [np_1]$)

$$\hat{\tau}_2 = \inf\{t\colon \log \overline{F}_n(t) - \sum \log \overline{F}_n(T_{(i)})/(k+1) \le \hat{\lambda}_0\left(\sum T_{(i)}/(k+1) - t\right) + \varepsilon_n\}$$

if the infimum is less than or equal to $\hat{\xi}_{p_0}$

$$= \hat{\xi}_{p_0} \qquad \text{otherwise.}$$

For $\hat{\tau}_1$: $\varepsilon_n = .05$, $h_n = n^{-\frac{1}{4}}$.

For $\hat{\tau}_2$: $\varepsilon_n = .05$.

$\hat{\tau}_3$ is the solution of $X_n(\cdot)$ nearest to zero.

For $\hat{\tau}_4$: $\delta_1 = 3$, $\delta_2 = .01$.

The values of p_0, p_1, ε_n, h_n, δ_1 and δ_2 are chosen somewhat arbitrarily.

The mi's and the Ri's are respectively means and mean square errors; R_{3a} is the mean square error using the limiting distribution of $\hat{\tau}_3$.

			$\hat{\tau}_1$	$\hat{\tau}_2$	$\hat{\tau}_3$	$\hat{\tau}_4$	$\hat{\tau}_3$
a	b	τ	m1(R1)	m2(R2)	m3(R3)	m4(R4)	R_{3a}
3	2	.15	.1061(.0115)	.1175(.0122)	.1743(.0131)	.9799(1.6256)	.0271
3	2	.10	.0792(.0094)	.1005(.0131)	.1890(.0255)	.9888(1.8319)	.0229
3	1	.15	.1371(.0076)	.1920(.0227)	.1957(.0194)	.1533(.0023)	.0199
3	1	.10	.0955(.0093)	.1705(.0401)	.1929(.0459)	.0994(.0019)	.0171
2	1	.20	.1459(.0159)	.2139(.0352)	.2728(.0441)	.3528(.3273)	.0385
2	1	.15	.1102(.0122)	.1923(.0447)	.2758(.0697)	.2726(.2075)	.0346
2	1	.10	.0883(.0136)	.1799(.0585)	.3322(.1323)	.4270(.6662)	.0310
2	.5	.20	.1846(.0109)	.3404(.1129)	.3092(.0793)	.1984(.0017)	.0568
2	.5	.15	.1246(.0084)	.3215(.1625)	.3289(.1649)	.1515(.0018)	.0513
2	.5	.10	.0839(.0059)	.3364(.2383)	.4036(.3270)	.1090(.0024)	.0463
1	.5	.40	.2548(.0454)	.4279(.1409)	.5456(.1766)	.4802(.2559)	.1538
1	.5	.30	.1736(.0285)	.3846(.1789)	.5517(.2788)	.4010(.2448)	.1382
1	.5	.20	.1120(.0201)	.3599(.2339)	.6643(.5294)	.4218(.4608)	.1241
1	.5	.15	.0826(.0172)	.3367(.2546)	.7568(.7140)	.4111(.4989)	.1176
1	.5	.10	.0613(.0125)	.3465(.3034)	.9064(.9721)	.5203(.7823)	.1114

We like to make the following remarks on the simulations.

(1) Of the three estimates $\hat{\tau}_1$, $\hat{\tau}_2$, $\hat{\tau}_3$, $\hat{\tau}_1$ seems to do best in all the cases. (This is a little surprising for we thought $\hat{\tau}_2$ should do better than $\hat{\tau}_1$).

(2) The (restricted) m.l.e. $\hat{\tau}_4$ is best when $a = 2$, $b = .5$ or $a = 3, b = 1$, i.e. when the amount of discontinuity in the density at τ is maximum in the cases simulated. A reason for this may be that the Chernoff-Rubin asymptotics (see Chernoff and Rubin [1956]) leading $0_p(1/n)$ errors for $\hat{\tau}_4$ is valid only when the extent of discontinuity is relatively large.

(3) $\sqrt{n}\,(\hat{\tau}_3 - \tau)$ has an asymptotic distribution with mean zero. So its asymptotic variance may be compared either with simulated variance ($= R3 - (m3)^2$) or with the mean square error ($= R3$). The asymptotic value provides good approximation to the simulated variance but not to the simulated mean square error, because the bias isn't negligible in all cases.

APPENDIX

Here we prove some results mentioned in Section 3.

$$\begin{aligned}
\mu_1(\tau) &= a^{-1} + e^{-a\tau}(b^{-1} - a^{-1}),\\
\mu_2(\tau) &= e^{-a\tau},\\
\mu_3(\tau) &= b^{-1}e^{-a\tau}(1 + b\tau),\\
\mu_4(\tau) &= b^{-2}e^{-a\tau}(b^2\tau^2 + 2b\tau + 2).
\end{aligned}$$

Let

$$\begin{aligned}
\mu_{(3)}(\tau) &= b^{-3}e^{-a\tau}(b^3\tau^3 + 3b^2\tau^2 + 6b\tau + 6),\\
\mu_{(4)}(\tau) &= b^{-4}e^{-a\tau}(b^4\tau^4 + 4b^3\tau^3 + 12b^2\tau^2 + 24b\tau + 24).
\end{aligned}$$

Then

$$\begin{aligned}
\vartheta_{11}(\tau) &= 2e^{-a\tau}(b^{-1} - a^{-1}) + 2e^{-a\tau}(b^{-2} - a^{-2}) + 2a^{-2} - \mu_1^2(\tau),\\
\vartheta_{22}(\tau) &= \mu_2(\tau) - \mu_2^2(\tau),\\
\vartheta_{33}(\tau) &= \mu_4(\tau) - \mu_3^2(\tau),\\
\vartheta_{44}(\tau) &= \mu_{(4)}(\tau) - \mu_4^2(\tau),\\
\vartheta_{12}(\tau) &= \mu_3(\tau) - \mu_1(\tau)\mu_2(\tau),\\
\vartheta_{13}(\tau) &= \mu_4(\tau) - \mu_1(\tau)\mu_3(\tau),\\
\vartheta_{14}(\tau) &= \mu_{(3)}(\tau) - \mu_1(\tau)\mu_4(\tau),\\
\vartheta_{23}(\tau) &= \mu_3(\tau) - \mu_2(\tau)\mu_3(\tau),\\
\vartheta_{24}(\tau) &= \mu_4(\tau) - \mu_2(\tau)\mu_4(\tau),\\
\vartheta_{34}(\tau) &= \mu_{(3)}(\tau) - \mu_3(\tau)\mu_4(\tau),
\end{aligned}$$

$$\frac{\partial H}{\partial B_1}|_{\mu(\tau)} = -a\tau,$$
$$\frac{\partial H}{\partial B_2}|_{\mu(\tau)} = e^{a\tau}(a^{-1} - b^{-1} - 2\tau - b\tau^2/2) + \tau(1 + ab^{-1}) + \tau^2(a + b/2) + ab\tau^3/2 + b^{-1} - a^{-1},$$
$$\frac{\partial H}{\partial B_3}|_{\mu(\tau)} = 2e^{a\tau} - 2 - a\tau - b\tau - ab\tau^2 + b\tau e^{a\tau},$$
$$\frac{\partial H}{\partial B_4}|_{\mu(\tau)} = (b - be^{a\tau} + ab\tau)/2.$$

Let

$$X(t) = K_1(t) \quad \text{for } t \le \tau$$
$$= K_2(t) \quad \text{for } t > \tau,$$

then (vide NRW)

$$K_1(t) = S(t)(\text{ at } \exp(-at) - 1 + \exp(-at)) + (1 - \exp(-at))(t + a^{-1} + (b^{-1} - a^{-1})\exp(-a\tau + at)) - (a^{-1} - a^{-1}\exp(-a\tau) + b^{-1}\exp(-a\tau))at,$$
$$S^2(t) = a^{-2} + (b^{-1} - a^{-1})[2(\tau - t + b^{-1}) - (b^{-1} - a^{-1})\exp(-a\tau + at)] \times \exp(-a\tau + at),$$
$$K_2(t) = t + (a - b)\tau b^{-1}\exp(-a\tau - b(t - \tau)) - (a\tau + b(t - \tau))(a^{-1} - a^{-1}\exp(-a\tau) + b^{-1}\exp(-a\tau)).$$

Note that

$$\dot{S}(t) = S^{-1}(t)a(b^{-1} - a^{-1})[\tau - t + b^{-1} - a^{-1} - (b^{-1} - a^{-1})\exp(-a\tau + at)]\exp(-a\tau + at),$$

$$\dot{K}_1(t) = \dot{S}(t)(at\exp(-at) - 1 + \exp(-at)) - a^2tS(t)\exp(-at) + at\exp(-at) + a(b^{-1} - a^{-1})\exp(-a\tau + at),$$

$$\dot{K}_2(t) = 1 - (a - b)\tau\exp(-a\tau - b(t - \tau)) - b(a^{-1} - (b^{-1} - a^{-1})\exp(-a\tau)).$$

Note that $\dot{K}_1(t)$ and $\dot{K}_2(t)$ are continuously differentiable in a neighborhood of τ hence $\dot{X}(t)$ and $\ddot{X}(t)$ exist for $\tau_1 \le t < \tau$ and $\tau < t \le \tau_2$ and are continuous. It is easy to see that

$$\dot{X}(\tau+) = \dot{K}_2(\tau) = b^{-1}[a - b + (ab\tau - a^2\tau + b - a)\exp(-a\tau)]$$
$$\dot{X}(\tau-) = \dot{K}_1(\tau) = a^{-1}b\dot{K}_2(\tau)$$

(Thus $\dot{X}(\tau+)\dot{X}(\tau-) = a^{-1}b\dot{K}_2^2(\tau) > 0$ for $a \neq b$, $a > 0$, $b > 0$, $\tau > 0$). Asymptotic mean square error of $\hat{\tau}$ is

$$\frac{1}{2n}(\dot{X}^{-2}(\tau+) + \dot{X}^{-2}(\tau-))V^2(\tau).$$

Now note that $((\vartheta_{ij}(\tau)))$ the dispersion matrix of $\underset{\sim}{B}(\tau)$ is positive definite; if not then there exist, say, λ_1, λ_2 and λ_3 (not all zero) such that

$$\lambda_1 + \lambda_2 T + \lambda_3 T^2 = \text{ constant a.e. on } T > \tau$$

and

$$T = \text{ constant a.e. on } T < \tau$$

which is not possible. Also $\frac{\partial H}{\partial B_1}/_{\mu(\tau)} \neq 0$. Thus $V^2(\tau) > 0$. Now by continuity of $V^2(t)$ at τ, $V^2(t) > 0$ in a neighborhood of τ.

REFERENCES

[1] Billingsley, P. (1968) *Convergence of Probability Measures*, Wiley, New York.

[2] Chernoff, H. and Rubin H. (1956) "The estimation of the location of a discontinuity in density", *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* **1**, pp.19-37, University of California Press.

[3] Hinkley, D.V. (1970) "Inference about change-point in a sequence of random variables", *Biometrika* **57**, pp.1-17.

[4] Nguyen, H.T. and Pham, T.D. (1987) "Maximum-Likelihood estimation in the change-point hazard-rate model" [Preliminary Report], *The Institute of Mathematical Statistics Bulletin* **16**, No. 1 pp.36.

[5] Nguyen, H.T., Rogers, G.S. and Walker, E.A. (1984) "Estimation in change-point hazard rate models", *Biometrika* **71**, pp.299-304.

[6] Prakasa Rao, B.L.S. (1968) "Estimation of the location of the cusp of a continuous density", *Ann. Math. Statist.* **39**, pp.76-87.

[7] Prakasa Rao, B.L.S. (1986) "Asymptotic theory of estimation in non-linear regression", to appear in *Proceedings of the Conference in honor of M. Joshi.*

[8] Rubin, H. (1961) "The estimation of discontinuities in multivariate densities, and related problems in stochastic processes", *Proceedings of the Fourth Berkeley Symposium*

on Mathematical Statistics and Probability **1**, University of California Press, pp.563-574.

[9] Sen, P.K. (1981) *Sequential Nonparametrics*, Wiley, New York.

[10] Serfling, R.J. (1980) *Approximation Theorems of Mathematical Statistics*, Wiley, New York.