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 $\tau$ has been considered. Two estimates for $\tau$ 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 $A$, $r(t) = a I \left( 0 \leq t \leq \tau \right) + b I(t > \tau)$, 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

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

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

$$ r(t) = \lambda(t) \quad \text{if} \quad 0 \leq t \leq \tau $$

$$ = \lambda_0 \quad \text{if} \quad t > \tau \quad (1.1) $$

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 $\tau$. If one knew $\tau$, items
could be tested up to time $\tau$ 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 $\tau$ 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,

$$\tau(t) = a \ I(0 \leq t \leq \tau) + b \ I(t > \tau)$$  \hspace{1cm} (1.2)

and proposed a consistent estimate for $\tau$ (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 $\tau$. 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 $\tau$. 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 $\tau$

In our model it is reasonable to assume that

$$0 < F(\tau) < 1.$$ \hspace{1cm} (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

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

Let $\xi_p$ 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

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

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

2
Under (1.1) for \( t > r \), \( \log \bar{F}(t) \) is linear in \( t \) and \( \hat{\lambda}_0 \) is an ordinary least square estimate of the slope \( \lambda_0 \) treating \( T_i(t) \) as independent and \( \log \bar{F}_n(T_i(t)) \) as dependent variables. It is well known (vide e.g. Serfling [1980] p. 59) that

\[
\sqrt{n} \sup_t |F_n(t) - F(t)| = O_p(1),
\]

and for \( 0 < p < 1 \)

\[
\sqrt{n} (\hat{\xi}_p - \xi_p) = O_p(1).
\]  

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

\[
\sqrt{n}(y_n(t) - y(t)) = O_p(1).
\]

Now we claim that

\[
\sqrt{n}(\hat{\lambda}_0 - \lambda_0) = O_p(1).
\]  

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

\[
\frac{n}{(k + 1)} \cdot \int_{T_{(n_0),+1}}^{T_{(n_{p_0})}} \phi(F_n(x), x) dF_n(x)
\]

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

**Lemma:** Let \( T_{i_n} = \xi_i + O_p(n^{-1/2}) \) \( i = 1, 2 \) 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 \( \tau \). Let \( h_n = n^{-\frac{1}{4}} \) and \( \varepsilon_n = c(\log n)n^{-\frac{1}{4}} \)

\[
\hat{\tau}_1 = \inf \{ t : y_n(t + h_n) - y_n(t) \leq h_n \hat{\lambda}_0 + \varepsilon_n \}
\]

and

\[
\hat{\tau}_2 = \inf \{ t : \log \bar{F}_n(t) - \log(1 - p_0) \leq \hat{\lambda}_0(\hat{\xi}_{p_0} - t) + \varepsilon_n \}.
\]
To see the motivation for $\hat{r}_1$ 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 $\{y_n(t + h_n) - y_n(t) \leq h_n\lambda_0 + \epsilon_n\}$. We then estimate $r$ as the smallest $t$ for which $H_{0t}$ is accepted. Formally $\hat{r}_1$ is as given above. Motivation of $\hat{r}_2$ is similar.

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

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

$$
y_n(r + \epsilon + h_n) - y_n(r + \epsilon) = y(r + \epsilon + h_n) - y(r + \epsilon) + O(n^{-\frac{1}{2}})$$

$$= \lambda_0 h_n + O(n^{-\frac{1}{2}}) \quad \text{by (1.1), and by (2.5)}$$

$$= \lambda_0 h_n + O(n^{-\frac{1}{2}}). \quad (2.6)$$

Now note that

$$(y_n(r + \epsilon + h_n) - y_n(r + \epsilon) \leq \lambda_0 h_n + \epsilon_n)$$

$$\Rightarrow (\hat{r}_1 \leq r + \epsilon). \quad (2.7)$$

Thus using (2.6) and (2.7) we have

$$P(\hat{r}_1 \leq r + \epsilon) \to 1. \quad (2.8)$$

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

$$y_n(t + h_n) - y_n(t) = \log \bar{F}(t) - \log \bar{F}(t + h_n) + O(n^{-\frac{1}{2}})$$

uniformly in $0 \leq t \leq r - \epsilon$

$$\geq h_n \lambda(x + h_n) + O(n^{-\frac{1}{2}})$$

uniformly in $0 \leq t \leq r - \epsilon$

$$> h_n \lambda_0 + h_n \delta_\epsilon + O(n^{-\frac{1}{2}})$$

uniformly in $0 \leq t \leq r - \epsilon$

where $\delta_\epsilon$ is such that $\lambda(r - \epsilon/2) > \lambda_0 + \delta_\epsilon$

$$> h_n \lambda_0 + h_n \delta_\epsilon + O(n^{-1/2})$$

uniformly in $0 \leq t \leq r - \epsilon \quad (2.9)$

Hence

$$P(\hat{r}_1 \geq r - \epsilon) \to 1. \quad (2.10)$$

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

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

$$\log \bar{F}_n(r + \epsilon) - \log(1 - p_0) = \log \bar{F}(r + \epsilon) - \log(1 - p_0) + O(n^{-\frac{1}{2}})$$

$$= \lambda_0 (\hat{p}_0 - r - \epsilon) + O(n^{-\frac{1}{2}}), \quad \text{(by (2.5)).}$$
Also,
\[(\log \bar{F}_n(\tau + \varepsilon) - \log(1 - p_0) \leq \lambda_0(\hat{\xi}_{p_0} - \tau - \varepsilon) + \varepsilon_n) \Rightarrow (\hat{\tau}_2 \leq \tau + \varepsilon).\]

Hence
\[P(\hat{\tau}_2 \leq \tau + \varepsilon) \to 1. \quad (2.11)\]

Now, by (2.4), uniformly in \(0 \leq x < \tau - \varepsilon,\)
\[
\log \bar{F}_n(x) - \log(1 - p_0) = \log \bar{F}(x) - \log(1 - p_0) + O_p(n^{-\frac{1}{2}})\\
= \int_x^{\hat{\xi}_{p_0}} \lambda(t) dt + O_p(n^{-\frac{1}{2}})\\
\geq (\hat{\xi}_{p_0} - x) \lambda_0 + \delta + O_p(n^{-\frac{1}{2}}) \quad \text{for some } \delta > 0.
\]

Hence
\[P(\hat{\tau}_2 \geq \tau - \varepsilon) \to 1. \quad (2.12)\]

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
\[f(t) = a \exp(-at)I(0 \leq t \leq \tau) + b \exp(-ar - b(t - \tau))I(t > \tau) \quad (3.1)\]

which is p.d.f. for \((a, b, \tau), 0 < a, 0 < b, 0 < \tau.\) (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 \(\tau\) is a change point not in the usual sense (e.g. Hinkley [1970]) where one has a sequence of parameters \(\theta_t\) which change from one value \(\theta_0\) for \(t < \tau\) to another value \(\theta_1\) 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 \(\tau.\)

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 \(\tau\) is \(\hat{\tau} = t_{(n)}\). Obviously \(t_{(n)}\) is not consistent.

If one chooses a compact set of \((a, b, \tau)\)'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 \(\tau\) 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 \( \tau \). 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 \( \tau \). 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

\[
X_n(t) \overset{a.s.}{\rightarrow} X(t)
\]

where \( X(t) \) is a non-stochastic function, \( X(\tau) = 0 \) and \( X(t) \) is monotone in a neighborhood of \( \tau \). 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 \( \tau \)), 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

\[
B_1(T, t) = T, \\
B_2(T, t) = I(T > t), \\
B_3(T, t) = T I(T > t), \\
B_4(T, t) = T^2 I(T > t).
\]

and

\[
\bar{B} = (B_1, B_2, B_3, B_4).
\]

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

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

\[
H(\bar{B}(t)) = (\bar{B}_4(t)/\bar{B}_2(t) - \bar{B}_3^2(t)/\bar{B}_2^2(t))^\frac{1}{2} (\bar{B}_2(t) - \bar{B}_2(t) \log \bar{B}_2(t) - 1)
\]

\[
+ (1 - \bar{B}_2(t))\bar{B}_3(t)/\bar{B}_2(t) + \bar{B}_1(t) \log \bar{B}_2(t).
\]

Let

\[
X_n(0) = 0, \\
X_n(T_{(i)}) = H(\bar{B}(T_{(i)})), \quad i = 1, \ldots, n - 1 \\
X_n(t) = X_n(T_{(n-1)}), \quad \text{for } t \geq T_{(n-1)}
\]

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 \( \tau \) and let \( \hat{\tau}_3 = \) zero of \( X_n(t) \) nearest to \( \hat{\tau}_c \). Fix a neighborhood \([\tau_1, \tau_2]\) of \( \tau \) and consider

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

where

\[
X(t) = H(\psi(t)), \quad \psi(t) = E(\bar{B}(\cdot, t)).
\]
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

\[ Y_n(t) - \sqrt{n}(H(\mathcal{F}(t)) - X(t)) = o_p(1). \]

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

\[ Y_n(\cdot) \xrightarrow{w} Y(\cdot) \]

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

\[ Z_n(h) = Y_n(\tau + n^{-\frac{1}{2}} h) \quad \text{if} \quad |h| \leq \log n \]
\[ = Y_n(\tau + n^{-\frac{1}{2}} \log n) \quad \text{if} \quad h \geq \log n \]
\[ = Y_n(\tau - n^{-\frac{1}{2}} \log n) \quad \text{if} \quad h \leq -\log n. \]

Then

\[ Z_n(\cdot) \xrightarrow{w} Z(\cdot), \]

where

\[ Z(\cdot) = Y(\tau). \quad (3.3) \]

Now let

\[ W_n(h) = \sqrt{n} \ X_n(\tau + hn^{-\frac{1}{2}}) \quad \text{if} \quad |h| \leq \log n \]
\[ = \sqrt{n} \ X_n(\tau + n^{-\frac{1}{2}} \log n) \quad \text{if} \quad h > \log n \]
\[ = \sqrt{n} \ X_n(\tau - n^{-\frac{1}{2}} \log n) \quad \text{if} \quad h < -\log n. \quad (3.4) \]

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

\[ W_n(\cdot) \xrightarrow{w} A(\cdot), \]

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

\[ A(h) = Y(\tau) + h\dot{X}(\tau^+) \quad \text{if} \quad h > 0. \quad (3.5) \]
\[ Y(\tau) - h\dot{X}(\tau^-), \quad \text{if} \quad h < 0 \]

and

\[ A(0) = Y(\tau). \]

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

\[ A_1(f) = \sup \{ t : f(t) = 0 \}, \]

\[ A_2(f) = \inf \{ t : f(t) = 0 \}, \]

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

\[ (A_1(W_n) - A_2(W_n)) \xrightarrow{w} (A_1(A) - A_2(A)) \quad (3.6) \]

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 \( w.p. \to 1 \)

\[ A_1(W_n) \geq \sqrt{n}(\hat{\tau}_3 - \tau) \geq A_2(W_n). \quad (3.7) \]

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 \( i = 1, 2 \)

\[ A_i(A) = \begin{cases} \frac{-Y(\tau)}{\dot{X}(\tau^+)} & \text{if} \quad Y(\tau) \leq 0 \quad \text{and} \quad \dot{X}(\tau^+) > 0 \\ \text{or if} \quad Y(\tau) \geq 0 \quad \text{and} \quad \dot{X}(\tau^+) < 0 \\ \frac{-Y(\tau)}{\dot{X}(\tau^-)} & \text{if} \quad Y(\tau) > 0 \quad \text{and} \quad \dot{X}(\tau^-) > 0 \\ \text{or if} \quad Y(\tau) \leq 0 \quad \text{and} \quad \dot{X}(\tau^-) < 0 \end{cases} \quad (3.8) \]

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

\[ V^2(\tau) = \text{Var}(Y(\tau)) \]

\[ = \sum_{i \leq 4} \sum_{j \leq 4} \frac{\partial^2 Y(\tau)}{\partial B_i \partial B_j} |\mu(\tau)| \frac{\partial H}{\partial B_i} |\mu(\tau)|. \quad (3.9) \]

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 $\hat{r}_4$ which is the m.l.e. with parameter space $r \leq \delta_1, |a - b| \leq \delta_2 > 0$.

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

$$\hat{a} = \frac{(1 - B_2(\hat{r}_3))}{(B_1(\hat{r}_3) - B_3(\hat{r}_3) + \hat{r}_3 B_2(\hat{r}_3))} = H_1(\overline{B}, \hat{r}_3),$$

say, and

$$\hat{b} = \frac{B_2(\hat{r}_3)}{(B_3(\hat{r}_3) - \hat{r}_3 B_2(\hat{r}_3))}.$$  

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 $\delta$-method we have

$$n^{\frac{1}{2}}(\hat{a} - a) = n^{\frac{1}{2}}(H_1(\overline{B}(\hat{r}_3), \hat{r}_3) - H_1(\overline{B}(\mu(r), r)))$$

$$= n^{\frac{1}{2}} \sum_{i=1}^{3} (\overline{B}_i(\hat{r}_3) - \mu_i(r)) \frac{\partial H_1}{\partial \mu_i(r)} + n^{\frac{1}{2}}(\hat{r}_3 - r) \frac{\partial H_1}{\partial r} + o_p(1)$$

$$= n^{\frac{1}{2}} \sum_{i=1}^{3} (\overline{B}_i(\hat{r}_3) - \mu_i(\hat{r}_3)) \frac{\partial H_1}{\partial \mu_i(r)} + n^{\frac{1}{2}}(\hat{r}_3 - r) \left[ \frac{\partial H_1}{\partial r} + \sum_{i=1}^{3} \frac{\partial H_1}{\partial \mu_i(r)} \frac{\partial \mu_i(r)}{\partial r} \right] + o_p(1)$$

$$= W_{3n}(r) + Q_{1n}(h_n) - Q_{1n}(0) + c(r) h_n + o_p(1), \quad (3.10)$$

where

$$h_n = n^{\frac{1}{2}}(\hat{r}_3 - r),$$

$$W_{3n}(t) = n^{\frac{1}{2}} \sum_{i=1}^{3} (\overline{B}_i(t) - \mu_i(t)) \frac{\partial H_1}{\partial \mu_i(t)},$$

$$Q_{1n}(h) = n^{\frac{1}{2}} \sum_{i=1}^{3} (\overline{B}_i(r + n^{-\frac{1}{2}} h) - \mu_i(r + n^{-\frac{1}{2}} h)) \frac{\partial H_1}{\partial \mu_i(r)}$$

and

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

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

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

9
where $\delta$ is the measure degenerate at zero. This implies
\[ \sqrt{n}(\hat{\tau}_3 - \tau) - A_1(W_n(\cdot)) \overset{P}{\to} 0. \]
It follows that
\[ c(\tau) h_n = A_1(W_n(\cdot)) c(\tau) + o_P(1), \]
and using Theorem 8.2 of Billingsly [1968, p. 55] we have
\[ \sup_{|h| < \log n} |Q_{1n}(h) - Q_{1n}(0)| \overset{P}{\to} 0. \tag{3.11} \]
Thus
\[ n^{\frac{1}{2}} (\hat{\tau} - \tau) = W_{3n}(\tau) + c(\tau) A_1(W_n(\cdot)) + o(1) \tag{3.12} \]
It is easy to check by the delta method that
\[ (W_{3n}(\tau), \sqrt{n}(H(\overline{F}(\tau)))) \overset{w}{\to} (X,Y) \]
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 $\mathbb{R} \times C(-\infty, \infty)$, 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 $\mathbb{R} \times C(-\infty, \infty)$, it follows from Prohorov's theorem that
\[ (W_{3n}(\tau), W_n(\cdot)) \overset{w}{\to} (X, A(\cdot)). \tag{3.13} \]
It now follows from (3.8), (3.12) and (3.13) that $\sqrt{n}(\hat{\tau} - \tau)$ converges in distribution to $X + c(\tau) A_1(A(\cdot))$.

The limiting distribution can be calculated from bivariate normal tables. For example if $\hat{X}(\tau^+) > 0 > \hat{X}(\tau^-)$, then for any real "$d$",
\[ \lim P\{\sqrt{n}(\hat{\tau} - \tau) \leq d\} = P\{X - c(\tau) Y(\tau)/\hat{X}(\tau^+) \leq d, Y(\tau) \geq 0\} + P\{X - c(\tau) Y(\tau)/\hat{X}(\tau^-) \leq d, Y(\tau) < 0\}. \]

In case $\hat{\tau}_3$ is replaced by $\hat{\tau}_4$ in the estimate for $\hat{\tau}$, then (3.10) and (3.11) continue to hold and $h_n \overset{P}{\to} 0$ since $(\hat{\tau}_4 - \tau) = o_P(1/n)$. It follows that $\sqrt{n}(\hat{\tau} - \tau)$ 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 \( \hat{r}_2 \) for simulations (the summations below range over \( i = [n p_0] + 1 \) to \( i = [n p_1] \))

\[
\hat{r}_2 = \inf \{ t: \log \mathcal{F}_n(t) - \sum \log \mathcal{F}_n(T(i))/(k + 1) \\
\leq \hat{\lambda}_0(\sum T(i)/(k + 1) - t) + \epsilon_n \}
\]

if the infimum is less than or equal to \( \hat{\epsilon}_{p_0} \)

\[= \hat{\epsilon}_{p_0} \]

otherwise.

For \( \hat{r}_1 \): \( \epsilon_n = .05, \) \( h_n = n^{-\frac{1}{4}}. \)

For \( \hat{r}_2 \): \( \epsilon_n = .05. \)

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

For \( \hat{r}_4 \): \( \delta_1 = 3, \delta_2 = .01. \)

The values of \( p_0, p_1, \epsilon_n, h_n, \delta_1 \) and \( \delta_2 \) are chosen somewhat arbitrarily.

The \( m_i \)'s and the \( R_i \)'s are respectively means and mean square errors; \( R_{3a} \) is the mean square error using the limiting distribution of \( \hat{r}_3 \).

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>( \tau )</th>
<th>( \hat{r}_1 )</th>
<th>( \hat{r}_2 )</th>
<th>( \hat{r}_3 )</th>
<th>( \hat{r}_4 )</th>
<th>( \hat{r}_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
<td>.15</td>
<td>.1061 (.0115)</td>
<td>.1175 (.0122)</td>
<td>.1743 (.0131)</td>
<td>.9799 (.16256)</td>
<td>.0271</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>.10</td>
<td>.0792 (.0094)</td>
<td>.1005 (.0131)</td>
<td>.1890 (.0255)</td>
<td>.9888 (.18319)</td>
<td>.0229</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>.15</td>
<td>.1371 (.0076)</td>
<td>.1920 (.0227)</td>
<td>.1957 (.0194)</td>
<td>.1533 (.0023)</td>
<td>.0199</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>.10</td>
<td>.0955 (.0093)</td>
<td>.1705 (.0401)</td>
<td>.1929 (.0459)</td>
<td>.0994 (.0019)</td>
<td>.0171</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>.20</td>
<td>.1459 (.0159)</td>
<td>.2139 (.0352)</td>
<td>.2728 (.0441)</td>
<td>.3528 (.3273)</td>
<td>.0385</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>.15</td>
<td>.1102 (.0122)</td>
<td>.1923 (.0447)</td>
<td>.2758 (.0697)</td>
<td>.2726 (.2075)</td>
<td>.0346</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>.10</td>
<td>.0883 (.0136)</td>
<td>.1799 (.0585)</td>
<td>.3322 (.1232)</td>
<td>.4270 (.6662)</td>
<td>.0310</td>
</tr>
<tr>
<td>2</td>
<td>.5</td>
<td>.20</td>
<td>.1846 (.0109)</td>
<td>.3404 (.1129)</td>
<td>.3092 (.0793)</td>
<td>.1984 (.0017)</td>
<td>.0568</td>
</tr>
<tr>
<td>2</td>
<td>.5</td>
<td>.15</td>
<td>.1246 (.0084)</td>
<td>.3215 (.1625)</td>
<td>.3289 (.1649)</td>
<td>.1515 (.0018)</td>
<td>.0513</td>
</tr>
<tr>
<td>2</td>
<td>.5</td>
<td>.10</td>
<td>.0839 (.0059)</td>
<td>.3364 (.2383)</td>
<td>.4036 (.3270)</td>
<td>.1090 (.0024)</td>
<td>.0463</td>
</tr>
<tr>
<td>1</td>
<td>.5</td>
<td>.40</td>
<td>.2548 (.0454)</td>
<td>.4279 (.1409)</td>
<td>.5456 (.1766)</td>
<td>.4802 (.2559)</td>
<td>.1538</td>
</tr>
<tr>
<td>1</td>
<td>.5</td>
<td>.30</td>
<td>.1736 (.0285)</td>
<td>.3846 (.1789)</td>
<td>.5517 (.2788)</td>
<td>.4010 (.2448)</td>
<td>.1382</td>
</tr>
<tr>
<td>1</td>
<td>.5</td>
<td>.20</td>
<td>.1120 (.0201)</td>
<td>.3599 (.2339)</td>
<td>.6643 (.5294)</td>
<td>.4218 (.4608)</td>
<td>.1241</td>
</tr>
<tr>
<td>1</td>
<td>.5</td>
<td>.15</td>
<td>.0826 (.0172)</td>
<td>.3367 (.2546)</td>
<td>.7568 (.7140)</td>
<td>.4111 (.4989)</td>
<td>.1176</td>
</tr>
<tr>
<td>1</td>
<td>.5</td>
<td>.10</td>
<td>.0613 (.0125)</td>
<td>.3465 (.3034)</td>
<td>.9064 (.9721)</td>
<td>.5203 (.7823)</td>
<td>.1114</td>
</tr>
</tbody>
</table>

We like to make the following remarks on the simulations.

(1) Of the three estimates \( \hat{r}_1, \hat{r}_2, \hat{r}_3 \), \( \hat{r}_1 \) seems to do best in all the cases. (This is a little surprising for we thought \( \hat{r}_2 \) should do better than \( \hat{r}_1 \)).
(2) The (restricted) m.l.e. $\hat{r}_4$ is best when $a = 2$, $b = .5$ or $a = 3, b = 1$, i.e. when the amount of discontinuity in the density at $r$ is maximum in the cases simulated. A reason for this may be that the Chernoff-Rubin asymptotics (see Chernoff and Rubin [1956]) leading $O_p(1/n)$ errors for $\hat{r}_4$ is valid only when the extent of discontinuity is relatively large.

(3) $\sqrt{n} (\hat{r}_3 - r)$ has an asymptotic distribution with mean zero. So its asymptotic variance may be compared either with simulated variance ($= R_3 - (m_3)^2$) or with the mean square error ($= R_3$). 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.

$$\mu_1(r) = a^{-1} + e^{-ar}(b^{-1} - a^{-1}),$$
$$\mu_2(r) = e^{-ar},$$
$$\mu_3(r) = b^{-1}e^{-ar}(1 + br),$$
$$\mu_4(r) = b^{-2}e^{-ar}(b^2 r^2 + 2br + 2).$$

Let

$$\mu_{(3)}(r) = b^{-3}e^{-ar}(b^3 r^3 + 3b^2 r^2 + 6br + 6),$$
$$\mu_{(4)}(r) = b^{-4}e^{-ar}(b^4 r^4 + 4b^3 r^3 + 12b^2 r^2 + 24br + 24).$$

Then

$$\vartheta_{11}(r) = 2e^{-ar}(b^{-1} - a^{-1}) + 2e^{-ar}(b^{-2} - a^{-2}) + 2a^{-2} - \mu_1^2(r),$$
$$\vartheta_{22}(r) = \mu_2(r) - \mu_2^2(r),$$
$$\vartheta_{33}(r) = \mu_4(r) - \mu_3^2(r),$$
$$\vartheta_{44}(r) = \mu_{(4)}(r) - \mu_4^2(r),$$
$$\vartheta_{12}(r) = \mu_3(r) - \mu_1(r)\mu_2(r),$$
$$\vartheta_{13}(r) = \mu_4(r) - \mu_1(r)\mu_3(r),$$
$$\vartheta_{14}(r) = \mu_{(3)}(r) - \mu_1(r)\mu_4(r),$$
$$\vartheta_{23}(r) = \mu_3(r) - \mu_2(r)\mu_3(r),$$
$$\vartheta_{24}(r) = \mu_4(r) - \mu_2(r)\mu_4(r),$$
$$\vartheta_{34}(r) = \mu_{(3)}(r) - \mu_3(r)\mu_4(r).$$
\[
\frac{\partial H}{\partial B_1} |_{\mu(r)} = -ar,
\]
\[
\frac{\partial H}{\partial B_2} |_{\mu(r)} = e^{ar} (a^{-1} - b^{-1} - 2r - br^2/2) + r(1 + ab^{-1}) + \tau^2(a + b/2) + abr^3/2 + b^{-1} - a^{-1},
\]
\[
\frac{\partial H}{\partial B_3} |_{\mu(r)} = 2e^{ar} - 2 - ar - br - abr^2 + b a^r,
\]
\[
\frac{\partial H}{\partial B_4} |_{\mu(r)} = (b - be^{ar} + abr)/2.
\]

Let

\[
X(t) = K_1(t) \quad \text{for } t \leq r
\]
\[
= K_2(t) \quad \text{for } t > r,
\]

then (vide NRW)

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

Note that

\[
\dot{S}(t) = S^{-1}(t)a(b^{-1} - a^{-1})[\tau - t + b^{-1} - a^{-1} - (b^{-1} - a^{-1}) \exp(-ar + at)] \exp(-ar + at),
\]
\[
\dot{K}_1(t) = \dot{S}(t)(at \exp(-at) - 1 + \exp(-at))
\]
\[
- a^2tS(t) \exp(-at) + at \exp(-at)
\]
\[
+ a(b^{-1} - a^{-1}) \exp(-ar + at),
\]
\[
\dot{K}_2(t) = 1 - (a - b)\tau \exp(-ar - b(t - r))
\]
\[
- b(a^{-1} - (b^{-1} - a^{-1}) \exp(-ar)).
\]

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

\[
\dot{X}(\tau+) = \dot{K}_2(\tau) = b^{-1}[a - b + (abr - a^2r + b - a) \exp(-ar)]
\]
\[
\dot{X}(\tau-) = \dot{K}_1(\tau) = a^{-1}b\dot{K}_2(\tau)
\]
(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 \((\theta_{ij}(\tau))\) the dispersion matrix of \( B(\tau) \) is positive definite; if not then there exist, say, \( \lambda_1, \lambda_2 \) and \( \lambda_3 \) (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 \( \frac{\partial H}{\partial \mu} / \mu(\tau) \neq 0 \). Thus \( V^2(\tau) > 0 \). Now by continuity of \( V^2(t) \) at \( \tau \), \( V^2(t) > 0 \) in a neighborhood of \( \tau \).

REFERENCES


