Sankhyā : The Indian Journal of Statistics
1997, Volume 59, Series A, Pt. 3, pp. 392-407

ISOTONIC MAXIMUM LIKELIHOOD ESTIMATION FOR THE CHANGE POINT OF A HAZARD RATE*

By S.N. JOSHI
Indian Statistical Institute, Bangalore
and
STEVEN N. MacEACHERN
The Ohio State University, Ohio

SUMMARY. A hazard rate $\lambda(t)$ is assumed to be of the shape of the "first" part of a "bathtub" model, i.e., $\lambda(t)$ is non-increasing for $t < \tau$ and is constant for $t \geq \tau$. The isotonic maximum likelihood estimator of the hazard rate is obtained and its asymptotic distribution is investigated. This leads to the maximum likelihood estimator and a confidence interval for a new version of the change point parameter. Their asymptotic properties are investigated. Some simulations are reported.

1. **Introduction**

Let F be an absolutely continuous distribution function (d.f.) on $[0, \infty)$ with density f, and let $X_1, X_2, \ldots, X_n$ be a random sample of size n from F. The failure rate function, or hazard rate, of F is denoted by λ, i.e., $\lambda(t) = f(t)/\bar{F}(t)$, where $\bar{F}(t) = 1 - F(t)$ for $t \geq 0$.

We assume that λ has the shape of the "first" part of a "bathtub" model. More precisely, assume that
for some $\tau > 0$

$$\lambda \text{ is nonincreasing }, \lambda(t) > \lambda_0 \text{ for } t < \tau \text{ and } \lambda(t) = \lambda_0 \text{ for } t \geq \tau, \qquad \ldots (1.1)$$

where $\lambda_0 > 0$ is a constant.
τ is sometimes called the change point of the failure rate function (see for example Basu, Ghosh and Joshi, 1988). In reliability and survival analysis τ is an important parameter. Typically, just before τ, the hazard rate is very

Paper received. June 1996; revised May 1997.
AMS (1990) *subject classification:* Primary: 62G07
Key words and phrases. Isotonic mle; failure rate; change point; asymptotic distribution; hypothesis testing; confidence interval.
* Part of the work was done when the first author was visiting the Ohio State University.

high and after τ it is constant, or in other words, it has reached its infimum. Estimating τ, or a point which has an interpretation similar to τ, is an important problem. In a reliability setting, for example, a burn-in period to, or beyond τ, maximizes the expected residual life of a device. Since implementing a longer burn-in costs more, one has traditionally searched for the minimum burn-in that produces the desired result. The length of this burn-in period is τ. Problems such as this have motivated much recent research into changepoint problems. In this paper we define the change point more realistically as below.

For a fixed $\epsilon > 0$ let τ_ϵ be such that

$$\tau_\epsilon = \text{Supremum } \{\text{t} : \lambda(t) \geq (1+\epsilon)\lambda_0\}. \qquad \ldots(1.2)$$

For small ϵ, τ_ϵ has an interpretation similar to that of τ. Under (1.1), a burn-in to time τ_ϵ will result in an expected residual lifetime of no less than $(1+\epsilon)^{-1}$ that of a burn-in to time τ: $E[(X-\tau_\epsilon)|X \geq \tau_\epsilon] \geq (1+\epsilon)^{-1}E[(X-\tau)|X \geq \tau]$. This allows us to obtain most of the benefit of a burn-in while providing us with a quantity which, as we shall see, is easier to estimate.

An additional benefit of our definition of the change point is that it is robust to small departures from the model (1.1), such as a model where $\lambda(t)$ slowly decreases after τ_ϵ to an asymptote $\lambda_0 > 0$. Such departures often arise when populations are mixtures. While the failure rate has not reached its infimum at τ_ϵ, it is only slightly higher than this infimum: τ_ϵ is finite while $\tau = \infty$.

For many practical problems, we find τ_ϵ to be a more compelling definition of a changepoint than τ. To use it, we advocate the choice of the largest ϵ such that a multiple of the hazard rate of $(1+\epsilon)$ and a reduction in expected residual life by a factor of $(1+\epsilon)^{-1}$ are judged relatively unimportant. In this paper we investigate the problem of inference for τ_ϵ.

Muller and Wang (1990) also have an alternative approach to the change point problem. Their parameter of interest is the point of maximum change in the hazard rate, and their estimation procedure involves the use of the estimated derivative of the hazard rate.

Consider the model

$$\lambda(t) = \begin{cases} a & \text{for } t < \tau, \\ b & \text{for } t \geq \tau \end{cases} \qquad \ldots(1.3)$$

This simple model for the hazard rate can be taken as an approximation to situations where λ of model (1.1) decreases to λ_0 rapidly. Though it is simple, it unnecessarily introduces the problems arising out of the discontinuity in the density, and represents a strong parametric assumption. A nonparametric approach avoids these difficulties while still allowing the constraints of model (1.1). The model (1.3) has been considered by many authors; see for example Loader (1991) who discusses inference based on the likelihood process and Ghosh et al. (1992) and (1996) for a Bayesian analysis.

The model of (1.1) was also considered by Ghosh et al. (1988). They have proposed consistent estimators for τ. One of their estimator, namely $\hat{\tau}_1$, is based on the idea of estimating the hazard rate by a simple histogram type estimate and then to use it to locate the change point. When one is estimating a monotone hazard rate, it is natural to demand that the estimate should also be monotone. In this paper we use the isotonic maximum likelihood approach to inference for our version of the change point, namely τ_ϵ.

For the model (1.1) it is reasonable to assume the knowledge of an upper bound for τ, say T_1. We obtain the nonparametric maximum likelihood estimator $\hat{\lambda}^\star$ of the hazard rate λ under the restrictions (see Theorem 3.1)

C1 : λ is nonincreasing, and

C2 : λ is constant from T_1 on.

For some of the results we will impose the following extra condition that ensures the unique definition of $\hat{\lambda}^\star$

C3 : $X_{(n)} \geq T_1$.

This estimator of λ leads naturally to the profile maximum likelihood estimator of τ_ϵ (see Corollary 3.1),

$$\hat{\tau}_\epsilon = \text{ Supremum } \{\text{t} : \hat{\lambda}^\star(t) \geq (1+\epsilon)\hat{\lambda}^\star(T_1)\}. \qquad \ldots (1.4)$$

The asymptotic distribution of $\hat{\lambda}^\star$ is investigated by proving a result similar to Theorem 7.1 of Prakasa Rao (1970) (see Theorem 3.2). The consistency of $\hat{\tau}_\epsilon$ is established (see Remark 3.2) and a confidence interval for τ_ϵ is derived (see Theorem 3.3).

If instead of T_1, only an upper bound p_0, $0 < p_0 < 1$, for $F(\tau)$ is known, one can get an isotonic estimate of λ by replacing T_1 by sample p_0 th quantile and as in (1.4) an estimate $\hat{\tau}_{m,\epsilon}$ (say), of τ_ϵ can be defined. Asymptotic results mentioned above also hold for these estimates. (see Remark 3.4).

Simulations were carried out for the model where the density is a mixture of two exponential densities. In Section 4 we report results of our simulation study comparing a version of $\hat{\tau}_{m,\epsilon}$, to the parametric m.l.e. of τ_ϵ. We also have compared a version of $\hat{\tau}_1$ of Basu, Ghosh and Joshi (1988) and unrestricted isotonic m.l.e. of τ_ϵ. $\hat{\tau}_{m,\epsilon}$ compares quite well with the parametric m.l.e. The performance of the other two estimates is very poor.

2. **Definitions and Preliminaries**

Let $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ be the ordered sample and let $X_{(0)} = 0$. Let F_n be the sample d.f. Let k be such that

$$X_{(k)} < T_1 \leq X_{(k+1)}.$$

Several of the proofs of results that follow rely on a specific implementation of the pool the adjacent violators algorithm (PAVA) representation of the isotonic m.l.e. Barlow *et al.* (1972) provide an excellent discussion of this algorithm.

Let $\hat{\rho}$ and $\hat{\rho}^{\star}$ be defined by

$$\begin{aligned}\hat{\rho}(t) &= 1/(n-i)(X_{(i)} - X_{(i-1)}) \\ &\quad \text{for } X_{(i-1)} < t \leq X_{(i)} \text{ and undefined for } t > X_{(n)},\end{aligned}$$

and

$$\begin{aligned}\hat{\rho}^{\star}(t) &= \hat{\rho}(t) \; for \; t \leq X_{(k)} \\ &= (n-k)/TTTA(X_{(k)}) \;\; \text{for } t > X_{(k)}\end{aligned}$$

where

$$\begin{aligned}TTTA(t) &= \text{total time on test after } t \\ &= \sum (X_i - t) I(X_i > t).\end{aligned}$$

Let $\hat{\lambda}$ be the m.l.e. of λ under the restriction C1 alone (see Section 5.3 of Barlow *et al.*, 1972). Then $\hat{\lambda}$ is the result of applying PAVA to $\hat{\rho}$. $\hat{\lambda}^{\star}$, derived in Theorem 3.1, is the m.l.e. of λ under restrictions C1 and C2. It is later noted (see Remark 3.1) that $\hat{\lambda}^{\star}$ can be obtained by applying PAVA to $\hat{\rho}^{\star}$.

For proofs of our results we need to study the intermediate steps involved in the applications of PAVA. Towards this end we need some more definitions.

Let $\hat{\rho}^{(0)}(t)$ be the result of applying PAVA to $\hat{\rho}(t)$, first for t in $(0, X_{(k)}]$ and then for t in $(X_{(k)}, X_{(n)}]$. The resulting $\hat{\rho}^{(0)}(t)$ is a step function which is nonincreasing on $(0, X_{(k)}]$ and on $(X_{(k)}, X_{(n)}]$, but which may not be monotone on the entire interval $(0,\, X_{(n)}]$. Let $X_{(k)} = X^{(1)} > X^{(2)} > X^{(3)}$... be the points to the left of T_1 at which $\hat{\rho}^{(0)}(t)$ has steps.

We define a finite set of estimators $\hat{\rho}^{(i)}$ for $i = 1, \ldots, I$. The estimator $\hat{\rho}^{(i)}$, defined only if the previous estimator, $\hat{\rho}^{(i-1)}$, is not monotone on $(0,\, X_{(n)}]$, is obtained by applying PAVA to $\hat{\rho}^{(i-1)}$ on the interval $(X^{(i+1)}, X_{(n)}]$. For this i-th stage $X^{(i+1)}$ is defined analogously. This application either produces an estimator that is monotone on $(0,\, X_{(n)}]$, or it forces the violation of the monotonicity condition one step to the left, to $X^{(i+1)}$. Thus, while $\hat{\rho}^{(0)}$ represents the initial estimator, $\hat{\rho}^{(I)} = \hat{\lambda}$, the isotonic estimator, and the intermediate $\hat{\rho}^{(i)}$ represent intermediate stages of our specific implementation of PAVA. More formally, for $i > 0$, $\hat{\rho}^{(i)}$ is defined only if $\hat{\rho}^{(i-1)}(X^{(i)}) < \hat{\rho}^{(i-1)}(T_1)$. In that case, $\hat{\rho}^{(i)}(t) = \hat{\rho}^{(i-1)}(t)$ for all $t \leq X^{(i+1)}$ while for $t > X^{(i+1)}$, $\hat{\rho}^{(i)}$ is the result of the application of PAVA to $\hat{\rho}^{(i-1)}$ on $(X^{(i+1)}, X_{(n)}]$. For $i = 0, 1, 2, \ldots, I^*$, $\hat{\rho}^{\star(i)}(t)$ is obtained analogously.

Several properties of the estimators $\hat{\rho}^{(i)}$ and $\hat{\rho}^{\star(i)}$ should be noted:

(a) $\hat{\rho}^{(i)}(t) = \hat{\rho}^{\star(i)}(t)$ for $t \leq X^{(i+1)}$,

(b) $\hat{\rho}^{(i)}(t) = \hat{\rho}^{(i)}(T_1)$ for $X^{(i+1)} < t \leq T_1$,

and

(c) $\hat{\rho}^{\star(i)}(t) = \hat{\rho}^{\star(i)}(T_1)$ for $t > X^{(i+1)}$.

These properties follow from the sequential application of PAVA to the estimators.

3. Main Results

This section contains the main theoretical results of the paper. First, we derive the maximum likelihood estimator of the hazard under conditions C1 and C2. This is then extended to provide the profile maximum likelihood estimator of τ_ϵ, where the hazard is treated as a nuisance parameter. Second, we obtain the asymptotic distribution of $\hat{\lambda}^\star(t)$ for an arbitrary fixed time point $t < \tau$. We then construct a confidence interval based on this asymptotic distribution.

THEOREM 3.1. *Let C1 and C2 hold. Then if $X_{(n)} \geq T_1$, the maximum likelihood estimator of λ is given by*

$$\hat{\lambda}^\star(t) = \max_{\substack{k \geq v \geq i+1 \\ v=n}} \min_{u \leq i} \{(v-u)/\sum_{j=u}^{v-1}(n-j)(X_{(j+1)} - X_{(j)})\}$$

for $t\epsilon(X_{(i)}, X_{(i+1)}]$ and $i = 0, 1, \ldots, k-1$, and

$$\hat{\lambda}^\star(t) = \min_{u \leq k}(n-u)/\sum_{j=u}^{n-1}(n-j)(X_{(j+1)} - X_{(j)}) \quad \text{for } t > X_{(k)}.$$

If $X_{(n)} < T_1$, any maximum likelihood estimator of λ must satisfy

$$\hat{\lambda}^\star(t) = \max_{v \geq i+1} \min_{u \leq i} \{(v-u)/\sum_{j=u}^{v-1}(n-j)(X_{(j+1)} - X_{(j)})\}$$

for $t \ \epsilon \ (X_{(i)}, X_{(i+1)}]$ and $i = 0, 1, \ldots, n-1$.

In this latter case, the m.l.e. may be extended beyond $X_{(n)}$ in any fashion that is consistent with the monotonicity and constancy assumptions above.

PROOF. The derivation of the maximum likelihood estimator of λ, subject to the restrictions provided in the assumptions is similar to the argument needed under condition C1 alone. The argument involves two parts. The first part shows that the m.l.e. must be constant between order statistics. The second part finds the estimator that maximizes the likelihood within this smaller class of estimators.

Consider any estimator, $\lambda^\star$, of the hazard that is non-increasing and constant on $[T_1, \infty)$. The log-likelihood of the data under this hazard is given by

$$l(\lambda^\star) = \sum_{i=1}^{n} \log(\lambda^\star(X_i)) - \sum_{i=1}^{n} \int_0^{X_i} \lambda^\star(t)dt.$$

Replace $\lambda^\star$ by $\lambda^{\star\star}$ where $\lambda^{\star\star}(t) = \lambda^\star(X_{(i)})$ for $t\epsilon(X_{(i-1)}, X_{(i)}]$, and note that, for every $t, \lambda^\star(t) \geq \lambda^{\star\star}(t)$. Then the difference between the log-likelihoods,

$$l(\lambda^{\star\star}) - l(\lambda^\star) = \sum_{i=1}^{n} \int_0^{X_i} \lambda^\star(t)dt - \sum_{i=1}^{n} \int_0^{X_i} \lambda^{\star\star}(t)dt$$

$$= \sum_{i=1}^{n} \int_{0}^{X_i} (\lambda^\star(t) - \lambda^{\star\star}(t))dt$$

is the sum of non-negative components. Hence, if a hazard exists that satisfies the assumptions above and maximizes the likelihood, it must be piecewise constant on the intervals $(X_{(i)}, X_{(i+1)}]$ for $i = 0, \ldots, n-1$.

Now, assume that $\lambda^\star$ is constant on the intervals $(X_{(i)}, X_{(i+1)}]$ for $i = 0, \ldots, n-1$. Let $X_{(k)}$ denote the largest order statistic that is less than T_1, and define $\lambda_i^\star = \lambda^\star(X_{(i)})$ for $i = 1, \ldots, k$. Also define $\lambda_0^\star = \lambda^\star(X_{(n)})$. The maximization of the likelihood over the values of $\lambda_1^\star \geq \lambda_2^\star \geq \ldots \geq \lambda_k^\star \geq \lambda_0^\star$ coincides with the standard isotonic maximization. See, for example, Marshall and Proschan (1965).

Remark 3.1. The estimator $\hat{\lambda}^\star$ is an isotonic estimator. As such, PAVA (see Barlow *et al.* (1972)) may be used to obtain $\hat{\lambda}^\star$. There are $k+1$ initial groups, the first k of which have weight 1 and the last of which has weight $n-k$. The initial values for the "solution blocks" are $(n-i)(X_{(i+1)} - X_{(i)})$ for $i = 0, \ldots, k-1$ and $\sum_{j=k+1}^{n}(X_{(j)} - X_{(k)})$ for the last block. Applying PAVA with these initial conditions provides an estimate of λ^{-1}.

Corollary 3.1. *The profile maximum likelihood estimator of τ_ϵ under (1.2) and conditions C1 through C3 is*

$$\hat{\tau}_\epsilon = \text{ Supremum } \{\text{t} : \hat{\lambda}^\star(t) \geq (1+\epsilon)\hat{\lambda}^\star(T_1)\}.$$

Proof. Consider the maximization of the likelihood over the parameters (τ_ϵ, λ), subject to the constraints C1 through C3. The likelihood does not depend directly on τ_ϵ. Instead, τ_ϵ specifies an additional restriction on λ. $\hat{\tau}_\epsilon$ defined above satisfies

$$\hat{\lambda}^\star(\tau_\epsilon) \geq (1+\epsilon)\hat{\lambda}^\star(T_1) \qquad \ldots (3.1)$$

Let l_2 represent the likelihood as a function of both parameters τ_ϵ and λ, and note that

$$l_2(\hat{\tau}_\epsilon, \hat{\lambda}^\star) = \sup_{\lambda/(3.1),\text{C1-C3}} l(\hat{\tau}_\epsilon, \lambda) = \sup_{\lambda/\text{C1-C3}} l(\lambda),$$

and so $\hat{\tau}_\epsilon$ is a profile m.l.e. of τ_ϵ. To see that $\hat{\tau}_\epsilon$ is the unique profile m.l.e. of τ_ϵ, notice that $\hat{\lambda}^\star$ does not satisfy (1.2) for any other value of τ_ϵ. Since $l(\hat{\lambda}^\star) > l(\lambda)$ for any other λ satisfying C1 and C2, $l_2(\hat{\tau}_\epsilon, \hat{\lambda}^\star) > l(\lambda)$ for any λ satisfying C1, C2 and (3.1) with $\tau_\epsilon \neq \hat{\tau}_\epsilon$.

We next turn to the derivation of the asymptotic distribution of $\hat{\lambda}^\star(t)$ for t $< \tau$. The result itself is presented in Theorem 3.2. Its proof relies on three lemmas and the eventual application of a theorem from Prakasa Rao (1970).

Prakasa Rao's theorem provides the asymptotic distribution of $\hat{\lambda}$. To derive the asymptotic distribution of $\hat{\lambda}^{\star}$, we investigate the relationship between $\hat{\lambda}$ and $\hat{\lambda}^{\star}$. The first pair of lemmas prepare the way for Lemma 3.3.

LEMMA 3.1. *Assume C3. For any $t < T_1$, if $\hat{\lambda}^{\star}(t) = \hat{\lambda}^{\star}(T_1)$, then $\hat{\lambda}(t) = \hat{\lambda}(T_1)$.*

PROOF. We first establish that whenever both $\hat{\rho}^{(i)}$ and $\hat{\rho}^{\star(i)}$ are defined, $\hat{\rho}^{(i)}(T_1) \geq \hat{\rho}^{\star(i)}(T_1)$. Consider the following mean time on test representation for $\hat{\rho}^{(i)}$ on the interval $(X^{(i+1)}, X_{(n)}]$. Let $t_1, \ldots, t_{m-1}$ denote the steps in $\hat{\rho}^{(i)}$ on this interval; set $t_0 = X^{(i+1)}$ and $t_m = X_{(n)}$. Partition the interval into subintervals $(t_{j-1}, t_j]$ for $j = 1, \ldots, m$. Let A_j denote the total time on test over the jth subinterval, and let n_j denote the number of failures in this interval. Then $\hat{\rho}^{(i)}(t_j) = n_j/A_j$ for $j = 1, \ldots, m$ with $n_1/A_1 > \ldots > n_m/A_m$. Now $\hat{\rho}^{\star(i)}$ is constant on $(X^{(i+1)}, X_{(n)}]$, and so $\hat{\rho}^{\star(i)}(t) = \sum_{j=1}^{m} n_j / \sum_{j=1}^{m} A_j$ for any t in the interval. Hence $\hat{\rho}^{(i)}(t_1) \geq \hat{\rho}^{\star(i)}(t_1)$, with equality only when $m = 1$. Finally, note that $T_1 < t_1$.

Since $\hat{\rho}^{(i)}(t_1) \geq \hat{\rho}^{\star(i)}(t_1)$, if $\hat{\rho}^{\star(i)}$ violates the monotonicity condition for the intervals $(X^{(i+2)}, X^{(i+1)}]$ and $(X^{(i+1)}, X^{(i)}]$, then $\hat{\rho}^{(i)}$ also violates the condition. This implies that whenever $\hat{\rho}^{\star(i)}$ is defined, $\hat{\rho}^{(i)}$ is also defined. As a consequence, if $\hat{\rho}^{\star(i)}(t_1) = \hat{\rho}^{\star(i)}(T_1)$ then $\hat{\rho}^{(i)}(t_1) = \hat{\rho}^{(i)}(T_1)$, implying the conclusion of the lemma.

LEMMA 3.2. *Assume C3. For any $t < T_1$, if $\hat{\lambda}(t) \neq \hat{\lambda}(T_1)$, then $\hat{\lambda}(t) = \hat{\rho}^{(0)}(t)$. Also, if $\hat{\lambda}^{\star}(t) \neq \hat{\lambda}^{\star}(T_1)$, then $\hat{\lambda}^{\star}(t) = \hat{\rho}^{\star(0)}(t)$.*

PROOF. Assume that $\hat{\lambda}(t) \neq \hat{\rho}^{(0)}(t)$. This will only happen if $\hat{\rho}^{(i)}(t) \neq \hat{\rho}^{(0)}(t)$ for some i. In turn, this will only happen when the interval containing t is pooled with another interval when PAVA is applied to $\hat{\rho}$ to obtain $\hat{\lambda}$. Since $\hat{\rho}^{(i)}$ is constant on $(X^{(i+1)}, T_1]$ for each i, this implies that the interval containing t is pooled with the interval containing T_1, and hence that $\hat{\lambda}(t) = \hat{\lambda}(T_1)$. Thus we have established the contrapositive of the first assertion. The proof of the second assertion in the lemma mimics this proof of the first with $\hat{\lambda}^{\star}$ in place of $\hat{\lambda}$ and $\hat{\rho}^{\star}$ in place of $\hat{\rho}$.

The final lemma needed for the derivation of the asymptotic distribution of $\hat{\lambda}^{\star}$ is given below. It shows that the two estimators $\hat{\lambda}$ and $\hat{\lambda}^{\star}$ are asymptotically equivalent for those t $< \tau$ with $\lambda(t) > \lambda_0$.

LEMMA 3.3. *Let (1.1) hold, and further assume that $\lambda(t) > \lambda_0$ for $t < \tau$. Then for each $t < \tau$,*

$$P(\hat{\lambda}(t) = \hat{\lambda}^{\star}(t)) \longrightarrow 1.$$

PROOF. Fix a $t < \tau$. Relying on Lemmas 3.1 and 3.2 when C3 is satisfied, we have that $\hat{\lambda}(t) \neq \hat{\lambda}(T_1)$ implies both $\hat{\lambda}(t) = \hat{\rho}^{(0)}(t)$ and $\hat{\lambda}^{\star}(t) = \hat{\rho}^{\star(o)}(t)$. If in addition t $< X_{(k)}$, then $\hat{\rho}^{\star(o)}(t) = \hat{\rho}^{(o)}(t)$ and so $\hat{\lambda}^{\star}(t) = \hat{\lambda}(t)$. Since $X_{(k)}$ is defined to be the largest order statistic less than T_1,

$$P(X_{(k)} > t) \longrightarrow 1.$$

Also,

$$P(X_{(n)} > T_1) \longrightarrow 1.$$

Further noting that $\hat{\lambda}$ is consistent for λ (see Theorem 4.1 of Marshall and Proschan (1965)) implies that

$$P(\hat{\lambda}(t) \neq \hat{\lambda}(T_1)) \longrightarrow 1,$$

completing the proof.

In view of the above lemma it is clear that for t $< \tau$, the asymptotic behaviour of $\hat{\lambda}^{\star}(t)$ can be studied using that of $\hat{\lambda}(t)$. Before presenting the asymptotic distribution, we need some more notation. Let $a(\cdot)$ be the p.d.f. of the asymptotic distribution obtained in Theorem 7.1 of Prakasa Rao (1970) (see Appendix for more details). Let Δ be a random variable having p.d.f. $a(\cdot)$ Let

$$\varphi(t) = -\left\{ \frac{\bar{F}(t)}{\lambda'(t)\lambda(t)} \right\}^{1/3}$$

The following theorem is an easy consequence of Lemma 3.3 and Theorem 7.1 of Prakasa Rao (1970).

THEOREM 3.2. *Let (1.1) hold, and further, for $t < \tau$ let $\lambda'(t) < 0$. Then*

$$n^{1/3}\varphi(t)(\hat{\lambda}^{\star}(t) - \lambda(t)) \xrightarrow{d} \Delta. \qquad \ldots (3.2)$$

Theorem 3.2 would lead directly to an asymptotic hypothesis test for $\tau_\epsilon = \tau_{\epsilon_0}$ for an arbitrary $\tau_{\epsilon_0} < T_1$, and hence to a confidence interval for τ_ϵ–if only we knew the values $\lambda(\tau_\epsilon) = (1+\epsilon)\lambda_0$ and $\varphi(\tau_{\epsilon_0})$. Unfortunately, this is not the case. Instead, we show in the upcoming Lemma 3.4 that $\hat{\lambda}^{\star}(T_1) \longrightarrow \lambda_0$ in probability with a rate quicker than $n^{-1/3}$. Consequently, when we replace λ_0 with $\hat{\lambda}^{\star}(T_1)$, the asymptotic properties of the test and interval remain unchanged. With the addition of a consistent estimator of $\varphi(\tau_\epsilon)$, and noting that $\varphi(t)$ is near $\varphi(\tau_\epsilon)$ when t is near τ_ϵ, we obtain a hypothesis test and the corresponding confidence interval for τ_ϵ. The remainder of this section presents this argument. The proof of Lemma 3.4 is rather technical, and so is deferred to the appendix.

LEMMA 3.4. *Let (1.1) hold and further assume that for some $t_1 < \tau$*

$$\lambda(t_1) > \lambda_0 \qquad \ldots (3.3)$$

Then

$$\sqrt{n}(\hat{\lambda}^\star(T_1) - \lambda_0) \text{ is } O_p(1).$$

REMARK 3.2. To investigate the consistency of $\hat{\tau}_\epsilon$, begin with consistency (under condition C1) of $\hat{\lambda}$ for λ at all points of continuity of λ. Then use Lemmas 3.3 and 3.4. Conclude that $\hat{\tau}_\epsilon$ is consistent for τ_ϵ unless $\lambda(t) = (1+\epsilon)\lambda_0$ for an open interval of t.

REMARK 3.3. The estimator $\hat{\tau}_\epsilon$ is robust to a violation of C2. If $\lambda(t)$ continues to decrease beyond T_1, then $\hat{\lambda}^\star(T_1) \longrightarrow E[X|X \geq T_1]^{-1}$. In this case, τ_ϵ may be defined as a multiple of this limiting term rather than as a multiple of the infimum of the hazard, and $\hat{\tau}_\epsilon$ will be consistent for this newly defined τ_ϵ.

In order to continue the argument, we need a consistent estimator of $\varphi(\tau_\epsilon)$. There are many consistent estimators of $\varphi(t)$. One such estimator is

$$\hat{\varphi}_1(t) = -\left\{ \frac{\bar{F}_n(t)}{\hat{\lambda}'(t)\hat{\lambda}^\star(t)} \right\}^{1/3}$$

where $\hat{\lambda}'(t)$ is as in (2.2) of Muller and Wang. Using Polya's Theorem and supposing the conditions necessary for (2.7) of Muller and Wang (1990), we may take $\hat{\nu} = \hat{\varphi}_1(\hat{\tau}_\epsilon)$, obtaining a consistent estimator of $\varphi(\tau_\epsilon)$. With $\hat{\nu}$ in place of $\varphi(\tau_\epsilon)$, at τ_ϵ the level of the test based on (3.2) goes to α. For any other value of t, $\lambda(t) \neq \lambda(\tau_\epsilon)$, and so the asymptotic power of the test is 1. Similarly, the 100(1-α) confidence interval obtained through the inversion of the hypothesis test has an asymptotic coverage probability of 1-α.

To formalize the preceeding paragraphs, we define a confidence interval $C_{n,\alpha}$ for τ_ϵ, "centered" at the m.l.e. $\hat{\tau}_\epsilon$ and based on the m.l.e. $\hat{\lambda}^\star$ of λ and a consistent estimator $\hat{\nu}$ of $\varphi(\tau_\epsilon)$ in the following way.

$$C_{n,\alpha} = \{t : \hat{\lambda}^\star(t)\epsilon[(1+\epsilon)\hat{\lambda}^\star(T_1) - n^{-1/3}\hat{\nu}\Delta_{1-\alpha/2}, (1+\epsilon)\hat{\lambda}^\star(T_1) + n^{-1/3}\hat{\nu}\Delta_{\alpha/2}]\}.$$

Theorem 3.3 follows easily from the above discussion and Lemma 3.4.

THEOREM 3.3. *Let (1.1) hold. Further, assume that $\lambda'(\tau_\epsilon) < 0$ and that $\hat{\nu}$ is a consistent estimator of $\varphi(\tau_\epsilon)$. Then*

$$P(\tau_\epsilon \in C_{n,\alpha}) \longrightarrow 1 - \alpha.$$

REMARK 3.4. Suppose we do not know an upper bound T_1 for τ but instead we know an upper bound p_0 for $F(\tau)$ where $0 < p_0 < 1$. We may replace T_1 by the pth sample quantile $\hat{\xi_{p_0}}$ and define $\hat{\lambda}^\star$ accordingly and then get $\hat{\tau}_{m,\epsilon}$. Of course they will not be the m.l.e.'s of λ and τ_ϵ respectively but it can be checked that lemmas 3.1, 3.2,3.3 and 3.4 hold good and hence theorems 3.1, 3.2 and 3.3 follow.

4. **Simulations**

This section describes a simulation study that investigates the accuracy of various estimates of the changepoint. The type of model that we are most interested in is one in which the hazard rate may never reach its infimum. In this instance, estimators which are designed to consistently estimate τ will tend to ∞, or to some upper bound which has been specified for the changepoint. In order to make a relatively fair comparison of various estimation strategies, we have created a set of estimators of τ_ϵ. These estimators are briefly described below.

The simulation is based on a density which is a mixture of two exponential densities. For some $\alpha > 0,\ \beta > 0$ and $0 < p < 1$ (with q=1-p), let the density be given by

$$f(x, \alpha,\ \beta, p) = p\alpha exp(-\alpha x)\ +\ q(\alpha+\beta)exp(-(\alpha+\beta)x).$$

$(\alpha+\beta)/\alpha$ represents the relative hazard rate of the two exponentials. Note that

$$\lambda(t) = \alpha \quad + q\beta/(p\ exp(t\beta)\ +\ q)).$$

Thus the hazard rate decreases to an asymptote at α, the lesser of the two hazard rates. Further assume that $\beta > \alpha\epsilon/q$ so that

$$\tau_\epsilon \ =\ \beta^{-1}log(q(\beta-\alpha\epsilon)/p\alpha\epsilon)\ \ >\ 0.$$

Assuming α and p to be known and β to be the unknown parameter, the m.l.e. of β and then that of τ_ϵ was obtained. Numerical maximization was required to calculate the m.l.e. Note that under the assumption of known α and p, $f(x;\beta)$ is a regular family in the sense that Cramer-Rao type conditions hold and hence the m.l.e. of β and hence that of τ_ϵ is asymptotically normal. (see e.g. Serfling (1980) pp. 144). We call this estimate the parametric m.l.e..

We obtained the unrestricted isotonic m.l.e. of the hazard rate, and using this obtained the lower endpoint of the step at which it just goes above (1+ϵ)×(its infimum). We call this estimate the unrestricted isotonic m.l.e.

Assuming p_0=0.5, (see Remark 3.4), $\hat{\lambda_m}^\star$ is obtained by replacing T_1 by the sample p_0 th quantile. Whenever the step of $\hat{\lambda_m}^\star$ either equal or just above $(1+\epsilon)\hat{\lambda_m}^\star(\hat{\xi_{p_0}})$ had more than one observation, we took the median of these observations as our estimate. We call this estimate, a slight modification of the restricted isotonic m.l.e. that is motivated by robustness concerns, the nonparametric m.l.e

We modified the estimate $\hat{\tau}_1$ of Basu, Ghosh and Joshi (1988) so that it estimates τ_ϵ instead of τ. We call this estimator BGJ1. For sample size n, the window length h_n was taken to be $n^{-1/4}$; the same as there. ϵ_n was taken to be 0.01. The choice of ϵ_n=0.05 was also investigated, but this value resulted in poorer performance. The choice of the other estimate of Basu, Ghosh and Joshi (1988), namely $\hat{\tau}_2$, is not amenable to such a modification.

Table 1. COMPARISONS OF THE NONPARAMETRIC M.L.E. (NP M.L.E.), PARAMETRIC M.L.E. (P M.L.E.), BGJ1 AND UNRESTRICTED ISOTONIC M.L.E. (UI M.L.E.) OF THE CHANGE POINT FOR SAMPLES OF SIZE 50.

α	β	p $\lambda(0)$	τ_ϵ	$F(\tau_\epsilon)$	NP m.l.e.	P m.l.e.	BGJ1	UI m.l.e.
0.5	15	.85	.3107	.2711	.3600	.3670	.1393	.863
		2.75			(.0496)	(.0634)	(.0561)	(3.262)
0.5	20	.85	.2474	.2480	.3489	.2945	.1341	.734
		3.50			(.0589)	(.0418)	(.0400)	(2.642)
0.5	25	.85	.2069	.2328	.3385	.2600	.1229	.725
		4.25			(.0669)	(.0370)	(.0332)	(2.665)
0.5	15	.9	.2799	.2162	.3783	.3305	.1313	.794
		2.00			(.0697)	(.0717)	(.0513)	(2.936)
0.5	20	.9	.2243	.1945	.3643	.2854	.1175	.776
		2.50			(.0803)	(.0619)	(.0385)	(2.954)
0.5	25	.9	.1884	.1801	.3550	.2566	.1101	.769
		3.00			(.0892)	(.0557)	(.0327)	(2.992)
1.0	15	.85	.2644	.3453	.2039	.2735	.1105	.443
		3.25			(.0157)	(.0224)	(.0425)	(.783)
1.0	20	.85	.2127	.3112	.1955	.2390	.1014	.434
		4.00			(.0120)	(.0187)	(.0305)	(.781)
1.0	25	.85	.1791	.2880	.1893	.2092	.0957	.371
		4.75			(.0122)	(.0158)	(.0233)	(.632)
1.0	15	.9	.2335	.2851	.2120	.2410	.1092	.413
		2.50			(.0167)	(.0259)	(.0387)	(.718)
1.0	20	.9	.1896	.2536	.2046	.2150	.1023	.409
		3.00			(.0161)	(.0225)	(.0299)	(.732)
1.0	25	.9	.1606	.2320	.1957	.1952	.0964	.401
		3.50			(.0168)	(.0195)	(.0251)	(.680)
1.5	15	.85	.2372	.4015	.1492	.2145	.0902	.285
		3.75			(.0135)	(.0113)	(.0336)	(.272)
1.5	20	.85	.1924	.3607	.1416	.1980	.0813	.301
		4.50			(.0079)	(.0093)	(.0231)	(.357)
1.5	25	.85	.1629	.3322	.1352	.1816	.0769	.295
		5.25			(.0060)	(.0086)	(.0178)	(.357)
1.5	15	.9	.2064	.3363	.1462	.1795	.0825	.305
		3.00			(.0110)	(.0139)	(.0286)	(.359)
1.5	20	.9	.1693	.2992	.1360	.1654	.0760	.257
		3.50			(.0075)	(.0116)	(.0209)	(.281)
1.5	25	.9	.1444	.2730	.1379	.1570	.0746	.298
		4.00			(.0069)	(.0099)	(.0172)	(.367)

The simulation itself considered several values of α, β and p. In all cases, ϵ was taken to be 0.05. For each combination of parameter values, sample sizes of 50 and 100 were investigated. 1000 data sets were generated for each combination of parameter values and sample size. Table 1 contains the results for sample size 50 and Table 2 for sample size 100. In these tables we report τ_ϵ, means, and mean square errors (in paranthesis) of the above estimates. We also have given values for $\lambda(0)$, and F at τ_ϵ.

Table 2. COMPARISONS OF THE NONPARAMETRIC M.L.E. (NP M.L.E.), PARAMETRIC M.L.E. (P M.L.E.), BGJ1 AND UNRESTRICTED ISOTONIC M.L.E. (UI M.L.E.) OF THE CHANGE POINT FOR SAMPLES OF SIZE 100.

α	β	p $\lambda(0)$	τ_ϵ	$F(\tau_\epsilon)$	NP m.l.e.	P m.l.e.	BGJ1	UI m.l.e.
0.5	15	.85	.3107	.2711	.3603	.3407	.1565	1.3529
		2.75			(.0427)	(.0290)	(.0479)	(6.097)
0.5	20	.85	.2473	.2479	.3412	.2843	.1403	1.240
		3.5			(.0512)	(.0225)	(.0363)	(5.795)
0.5	25	.85	.2069	.2378	.3263	.2305	.1195	1.142
		4.25			(.0565)	(.0114)	(.0263)	(5.218)
0.5	15	.90	.2798	.2162	.3560	.3438	.1328	1.258
		2.0			(.0547)	(.0542)	(.0474)	(5.840)
0.5	20	.90	.2243	.1944	.3408	.2855	.1224	1.142
		2.5			(.0628)	(.0455)	(.0367)	(5.410)
0.5	25	.90	.1883	.1807	.3344	.2271	.1072	1.187
		3.0			(.0754)	(.0231)	(.0231)	(5.914)
1	15	.85	.2644	.3452	.2215	.2849	.1328	.6725
		3.25			(.0117)	(.0147)	(.0350)	(1.458)
1	20	.85	.2127	.3111	.2029	.2388	.1142	.6308
		4.0			(.0105)	(.0116)	(.0258)	(1.355)
1	25	.85	.1791	.2879	.1915	.2045	.1047	.6196
		4.75			(.0107)	(.0086)	(.0211)	(1.369)
1	15	.90	.2335	.2851	.2085	.2528	.1154	.6902
		2.5			(.0141)	(.0206)	(.0342)	(1.598)
1	20	.90	.1896	.2536	.1912	.2201	.1023	.5825
		3.0			(.0128)	(.0173)	(.0267)	(1.252)
1	25	.90	.1601	.2320	.1851	.1833	.0829	.6977
		3.50			(.0129)	(.0127)	(.0214)	(1.765)
1.5	15	.85	.2372	.4015	.1566	.2365	.1001	.356
		3.75			(.0110)	(.0075)	(.0302)	(.473)
1.5	20	.85	.1924	.3607	.1474	.2044	.0899	.357
		4.50			(.0067)	(.0066)	(.0214)	(.501)
1.5	25	.85	.1629	.3322	.1452	.1827	.0900	.339
		5.25			(.0047)	(.0059)	(.0160)	(.495)
1.5	15	.90	.2064	.3363	.1421	.1985	.0882	.323
		3.00			(.0102)	(.0118)	(.0272)	(.462)
1.5	20	.90	.1693	.2992	.1386	.1804	.0771	.317
		3.50			(.0066)	(.0093)	(.0195)	(.452)
1.5	25	.90	.1444	.2730	.1286	.1639	.0730	.311
		4.00			(.0054)	(.0082)	(.0167)	(.451)

Our nonparametric m.l.e. is always below the p_0 th sample quantile $\hat{\xi_{p_0}}$. In order to have a fair comparison all the rest of the estimates were truncated above at $\hat{\xi_{p_0}}$. In all samples $\hat{\xi_{p_0}}$ was observed to be greater than τ_ϵ and thus truncation, if at all, had only a positive effect on their performances.

Substantial biases are sometimes present. In all the cases the unrestricted isotonic m.l.e. tends to overestimate τ_ϵ wheras BGJ1 tends to underestimate it; both by a wide margin. The nonparametric m.l.e. and the parametric m.l.e. do not show a pattern of either positive or negative bias.

As expected, in most of the cases the parametric m.l.e. performs very well. Its performance is aided by the assumption of known values for α and p. Surprisingly, however, the nonparametric m.l.e. is often comparable to the parametric m.l.e., and is ofttimes better. This is particularly true for samples of size 50 with a ratio of $(\alpha+\beta)/\alpha$ smaller than 31. With samples of size 100 and these lower hazard rate ratios, neither estimator shows a clear pattern of dominance.

When the nonparametric m.l.e. is compared to BGJ1, we find a clear pattern. For larger $(\alpha+\beta)/\alpha$ (over 31) BGJ1 has smaller mean square error; for smaller $(\alpha+\beta)/\alpha$ (under 31) the nonparametric m.l.e. has smaller mean square error. These results hold true for both sample sizes investigated.

As an overall conclusion, we find the nonparametric m.l.e. to be an effective estimator of our robust version of the changepoint. It compares quite favorably to its natural competitor, BGJ1 (which we had to create in this work) when the decline in hazard rates is not too precipitous. Remarkably, it is often competitive and sometimes beats the parametric m.l.e., computed with additional knowledge about the density. Additional simulations under a model of the form (1.3), not reported here, also support the choice of the nonparametric m.l.e.

While the nonparametric m.l.e. performs well in the simulations, we believe that the real strengths of the estimator are twofold. Its nonparametric nature removes the necessity of assuming a parametric model, and the principle of maximum likelihood provides a strong motivation for the estimator.

Appendix

Define (see Prakasa Rao (1970))

$$\psi(t) = \frac{1}{2}U_x(t^2,t)U_x(t^2,-t)$$

where $u(x,z) = P[W(t) > t^2 \text{ for some } t > z \mid W(z) = x]$ is a solution of the heat equation $\frac{1}{2}U_{xx} = -U_x$ subjrct to the boundary conditions (i) u(x,z)=1 for $x \geq z^2$ and (ii) $u(x,z) \longrightarrow 0$ as $x \longrightarrow -\infty$. Here U_x denotes the partial derivative of u(x,z) w.r.t. x and W(t) is the Wiener process on (-∞, ∞).
Then the density $a(\cdot)$ is given by

$$a(\cdot) = \frac{1}{2}\psi(\frac{1}{2}x).$$

Regarding Lemma 3.4. To prove this lemma we need some more notation.

If $X_{(n)} \leq T_1$, define $S_{1n}(t) = S_{2n}(t) = 1$ for all t. If $X_{(n)} > T_1$, define $S_{1n}(t)$ and $S_{2n}(t)$ at $t = X_{(i)}$ by

$$S_{1n}(t) = \Sigma I(X_j > t)/n$$

and

$$S_{2n}(t) = \Sigma (X_j - t) I(X_j > t)/n.$$

For $t > X^{(1)}$ define

$$S_{in}(t) = S_{in}(X^{(1)}),$$

and for all other t's $S_{in}(t)$ is defined by linear interpolation.

$$\begin{aligned} \text{Let} \quad h_i(t) &= ES_{in}(t), \\ W_{in}(t) &= \sqrt{n}(S_{in}(t) - h_i(t)), \\ h(t) &= h_1(t)/h_2(t), \\ S_n(t) &= S_{1n}(t)/S_{2n}(t), \text{ and} \\ W_n(t) &= \sqrt{n}(S_n(t) - h(t)) \end{aligned}$$

We also rely on another lemma, Lemma A.1.

LEMMA A.1. *As a process in* $C[t_1, T_1]$, $W_n(t) \xrightarrow{w} G(t)$ *where* $G(t)$ *is a zero mean Gaussian process.*

PROOF OF LEMMA A.1. The proof of Lemma A.1 is similar to the proof of Lemma 2.1 of Ghosh and Joshi (1992), and so we provide only the following sketch of a proof: Note that the weak convergence of the finite dimensional distributions of W_{1n} and W_{2n} can be proved easily by the multivariate central limit theorem. Tightness (and hence the weak convergence) of W_{1n} follows from an application of Theorem 13.1 of Billingsley (1968). Tightness of W_{2n} (and hence the weak convergence) can be proved by checking the moment condition (12.51) and applying Theorem 12.3 of Billingsley (1968). Thus

$$\begin{aligned} W_n(t) &= \sqrt{n}(S_n(t) - h(t)) \\ &= \sqrt{n}(\frac{S_{1n}(t)}{S_{2n}(t)} - \frac{h_1(t)}{h_2(t)}) \\ &= a_1(t)W_{1n}(t) + a_2(t)W_{2n}(t) + R_n(t), \end{aligned}$$

where

$$a_1(t) = 1/h_2(t), a_2(t) = -h_1(t)/h_2^2(t)$$

and

$$\sup_{[t_1, T_1]} R_n(t) \xrightarrow{P} 0.$$

Now note that weak convergence of the finite dimensional distributions of $\Sigma a_i(t)W_{in}(t)$ can be proved easily, and that the tightness follows from the tightness of $W_{1n}(t)$ and $W_{2n}(t)$ and the continuity of $a_1(t)$ and $a_2(t)$. Thus the weak convergence of $W_n(t)$ is established.

Proof of Lemma 3.4. Note that using Lemma A.1 we have for a given $\delta > 0$ that there exist $A_1 > 0$ and n_0 such that

$$P\{\inf_{[t_1,T_1]} \sqrt{n}(S_n(t) - h(t)) > -A_1\} \geq 1 - \delta \text{ for all } n \geq n_0.$$

Now note that $h(t) \geq \lambda_0$ and hence

$$P\{\inf_{[t_1,T_1]} S_n(t) > \lambda_0 - n^{-\frac{1}{2}} A_1\} \geq 1 - \delta \text{ for all } n \geq n_0. \qquad \ldots(A.1)$$

Now using (3.3), Lemma 3.3 and the consistency of $\hat{\lambda}(t_1)$, we have that for $\eta = (\lambda_{(t_1)} + \lambda_0)/2$,

$$P(\hat{\lambda}^\star(t_1) > \lambda_0 + \eta) \longrightarrow 1. \qquad \ldots(A.2)$$

Notice that $\hat{\lambda}^\star$ is obtained by applying PAVA to $\hat{\rho}^\star$ and hence the extent to which it can be "below" $\hat{\rho}^\star$ is "controlled" by $S_n(t)$. When $\hat{\lambda}^\star(t_1) > \lambda_0 + \eta$, we find $\hat{\lambda}^\star(T_1) \geq \inf_{[t_1,T_1]} S_n(t)$. Thus

$$\hat{\lambda}^\star(t_1) > \lambda_0 + \eta$$

and

$$\inf_{[t_1,T_1]} S_n(t) > \lambda_0 - n^{-\frac{1}{2}} A_1$$

imply

$$\hat{\lambda}^\star(T_1) > \lambda_0 - n^{-\frac{1}{2}} A_1.$$

The proof of the lemma is completed by using (A.1) and (A.2).

References

Barlow, R. E., D. J. Bartholomew, J. M. Bremner, and H. D. Brunk (1972). *Statistical Inference under Order Restrictions.* John Wiley and Sons, London.

Basu, A. P. Ghosh, J. K. and Joshi, S. N. (1988)."On estimating change point in a failure rate" *Statistical Decision Theory and Related Topics IV* **Vol. 2** Ed. S. S. Gupta and J. O. Berger, Springer-Verlag, New York, pp. 239-2

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

Ghosh, J. K. and Joshi, S. N. (1992). On the asymptotic distribution of an estimate of the change point in a failure rate, *Comm. in Statistics, Theory and Methods*, 21,pp 3571-88.

GHOSH, J. K., JOSHI, S. N. and MUKHOPADHYAY, C. (1992). A Bayesian approach to the estimation of change point in a hazard rate, *Advances in Reliability Theory*, Ed. A.P.Basu, 1141-170, Elsevier Science Publishers B.V.

GHOSH, J. K., JOSHI, S. N. and MUKHOPADHYAY, C. (1996). Asymptotics of a Bayesian approach to estimating change-point in a hazard rate, *Comm. in Statistics, Theory and Methods*, 25, 12, pp 3147-65.

LOADER, C. R. (1991). Inference for a hazard rate change point. *Biometrika,* **78,** 749-57.

MARSHALL, A. W. and F. PROSCHAN (1965). Maximum likelihood estimation for distributions with monotone failure rate, *Ann. Math. Statist.,* **36,** 69-77.

MULLER, H. G. and WANG, J. L. (1990). Nonparametric analysis of changes in hazard rates for censored survival data: An alternative to change-point models. *Biometrika* **77,** 305-314.

PRAKASA RAO, B. L. S. (1970). Estimation for distributions with monotone failure rate, *Ann. Math. Statist.,* **41,** 507-19.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley. New York.

INDIAN STATISTICAL INSTITUTE
8TH MILE MYSORE ROAD
BANGALORE 560059
INDIA
snj@isibang.ernet.in

THE OHIO STATE UNIVERSITY
COLUMBUS
OHIO 43210
USA