Change-point Problems
IMS Lecture Notes – Monograph Series (Volume 23, 1994)

CHANGE-POINT MODELS FOR HAZARD FUNCTIONS

BY H. G. MÜLLER AND J.-L. WANG
University of California, Davis

A review is presented of parametric and nonparametric models and corresponding estimation procedures for change-points in hazard functions where the data are possibly subject to random censoring. In particular, we discuss nonparametric models and the application of nonparametric smoothing techniques for change-point estimation and estimation of a hazard function when a change-point is present. Preliminary theoretical results are mentioned and a simulation study provides further insight.

1. Introduction. Change in distribution at an unknown time point arises in quality control problems and has been studied extensively. Another related type of problem is a change-point in a hazard function which may occur in medical follow up studies after a major operation, e.g. bone marrow transplantation. There is usually a high initial risk and then the risk settles down to a lower constant long term risk. A simple mathematical model is the following

MODEL 1: PARAMETRIC CHANGE-POINT MODEL. *The hazard function λ of a failure time variable T is of the form*

$$\lambda(t) = \begin{cases} \lambda_1, & t \le \tau, \\ \lambda_2, & t > \tau; \end{cases} \tag{1.1}$$

with constants $\lambda_1, \lambda_2 > 0$.

There are three parameters (λ_1,λ_2 and τ) in this model, τ is called the change-point. We refer to this model as the three-parameter change-point model. A short review of the pertinent literature is given in Section 2. In most of the published work to date the mathematical theories were developed

---

Research supported in part by Air Force Grant AFOSR-89-0386.

AMS 1991 Subject Classification: Primary 62F10, 62F12, 62N05; Secondary 62G07, 62P10.

Key words and phrases: Boundary, censoring, discontinuity, kernel smoothing, nonparametric estimation.

for a time variable T which is observable, i.e., an i.i.d. sample $T_1,\ldots,T_n$ of T is available. This is of course rarely the case in reality, e.g. in the example given in Matthews and Farewell (1982), which was subsequently analyzed by Worsley (1988) and Achcar (1989), data on 33 out of the 84 acute nonlymphoblastic leukemia patients were censored. Of those, 24 were censored at 182 days, when the patients were randomized to an experimental protocol. In another example in Matthews, Farewell and Pyke (1985), 11 out of the 31 advanced non-Hodgkin's lymphoma patients were still alive at the last time of follow up and were thus censored. Matthews and Farewell (1982) claimed that dropping the 24 censored observations at 182 days did not affect significantly the outcome of the likelihood ratio test, and most subsequent work develops theory for the case of observable time variables and in applications the censored observations are either discarded or the likelihood function is modified for censored data. Loader (1991, pp. 751–2) presents a discussion of the effect of censorship. Section 2.3 below contains more discussions on the issue of censoring under Model 1.

It should be noted that Model 1 is only a simplification of and approximation to the true model. More complicated parametric change-point models (possibly allowing for several changes) may be needed in reality. In view of the technical difficulty for even the simplest three-parameter model in (1.1) and the complications due to censoring, nonparametric change-point models may be an attractive alternative.

In this article, we examine two types of nonparametric models and discuss some of the practical issues associated with estimating the change-points and the hazard function based on randomly censored data.

MODEL 2: NONPARAMETRIC CHANGE-POINT MODEL. *Assume that the hazard function* $\lambda \in \mathcal{C}^k([0,\tau]) \cap \mathcal{C}^k([\tau,\infty])$, *for some integer* $k \geq 1$, *and for* $0 \leq j \leq k$ *let* $\lambda_+^{(j)}(x)$ *and* $\lambda_-^{(j)}(x)$ *be the respectively left- and right-hand limit of* $\lambda^{(j)}$, *the* j*th derivative of* λ, *and let* $\Delta_j = \lambda_+^{(j)}(\tau) - \lambda_-^{(j)}(\tau)$, *where we assume without loss of generality that* $\Delta_0 > 0$.

That is, λ is k times continuously differentiable with the exception of the change-point τ, where an isolated discontinuity occurs. Note that this includes Model 1 as a special case where the continuous parts of the hazard function would be constants. It is also possible to extend Model 2 to a more general case where $\Delta_l > 0$ for some $l \geq 1$ and $\Delta_j = 0$, for $0 < j < l$, i.e., change-points in the lth derivative. However, for simplicity we shall assume here $\Delta_0 > 0$. Inference for the more general case was discussed in Müller and Wang (1990b) and models where change-points occur in a derivative of a hazard function were also discussed independently by Antoniadis and Grégoire (1991), who assume in their approach that the location of the change-point is known.

An alternative approach was discussed in Müller and Wang (1990a). Approximate (1.1) by a smooth function, say g. The change-point τ in (1.1) then corresponds roughly to the location of the extremum of the first derivative of g, i.e., τ is the point where the "most rapid change" in g occurs. We refer to this approach as "smooth approximation model".

MODEL 3: SMOOTH APPROXIMATION MODEL. *Assume that the hazard function $\lambda \in C^k([0,\infty])$ for some $k \geq 2$ and there exists a point τ such that $|\lambda^{(1)}(\tau)| > |\lambda^{(1)}(x)|$ for all $x \neq \tau$.*

Thus no actual discontinuity occurs but rather a "point of most rapid change" τ exists. Kernel methods for estimating the change-point τ under random censoring in both Model 2 and Model 3 are discussed in Section 3. For the Smooth Approximation Model (Model 3) τ is estimated via the location $\tilde{\tau}$ where the estimate of $\lambda^{(1)}$ attains its maximum (cf. (3.1) and (3.3)). As for Model 2, as $|\lambda_+(x) - \lambda_-(x)| = \Delta_0 1_{\{x=\tau\}}$, it is natural to estimate τ via the location $\hat{\tau}$ of the maximal difference of one-sided kernel estimates of λ (cf. (3.2) and (3.4)). Preliminary results for these estimates are summarized in Propositions 1–4 in Section 3.

In a practical problem, one might not know which of these models actually applies. The two kernel methods are seen to be surprisingly similar in Section 3.3, and both adapt naturally to the actual change-point model, which will then only affect rates of convergence and asymptotic confidence regions. Once τ is estimated, one can then estimate the hazard function with a modified kernel method which employs boundary kernels to adapt to the estimated change-point (Section 3.4). The results of a simulation study assessing the practical effects of bandwidth and kernel choice are discussed in Section 4.

2. A Review of Parametric Modeling. As mentioned earlier, most of the results for the Parametric Change-point Model (Model 1) are based on a sample of i.i.d. observed failure times $T_1, ..., T_n$ with hazard function λ as specified in (1.1), where $T_{(1)} \leq ... \leq T_{(n)}$ denote the ordered observations. Model 1 was first postulated by Miller (1960) as an alternative to the constant hazard (exponential lifetime) model commonly used in life testing experiments. Miller assumed the change-point τ to be either known or known to be in a specified interval $[a, b]$ and did not consider inference for τ. For the three parameter change-point model the first inference procedure appeared in Matthews and Farewell (1982) and was motivated by the analysis of survival of leukemia patients.

2.1. Testing Hypotheses. It may be of interest to test a constant hazard rate against the change-point alternative for a group subjected to a new therapy. The hypotheses are thus:

$$H_0: \quad \lambda_1 = \lambda_2 \ (\text{or } \tau = 0) \quad \text{vs} \ H_1: \quad \lambda_1 \neq \lambda_2 \ (\text{or } \tau > 0). \tag{2.1}$$

Basically three types of test procedures have been proposed: (1) Likelihood Ratio type tests (LRT) and modified versions; (2) Score test; (3) Bayesian test. We discuss now each of these types of test.

(a) Likelihood Ratio Type Tests. References include Matthews and Farewell (1982), Worsley (1988), Henderson (1990) and Loader (1991). It should be noted that the classical asymptotic results for LRT do not apply here due to the discontinuity of the likelihood function at τ in H_1. The log-likelihood function based on observations $T_1, ..., T_n$ is:

$$l(\lambda_1, \lambda_2, \tau) = N(\tau) ln\lambda_1 + [n - N(\tau)] ln\lambda_2 - \lambda_1 S(\tau) - \lambda_2 [S - S(\tau)], \tag{2.2}$$

where

$$N(t) = \sum_{i=1}^{n} I(T_i \leq t), \tag{2.3}$$

is the number of failures observed up to time t,

$$S(t) = \sum_{i=1}^{n} (T_i \wedge t), \tag{2.4}$$

is the total time on test up to time t, and

$$S = S(\infty) = \sum_{i=1}^{n} T_i. \tag{2.5}$$

Nguyen, Rogers and Walker (1984) observed that the likelihood function in (2.2) is unbounded unless $\lambda_1 \geq \lambda_2$. In this case the MLE for λ_1, λ_2 and τ can be computed by numerical algorithms as in Matthews and Farewell (1982), where the critical regions were simulated for several sample sizes. Otherwise, pseudo-MLE's were considered as an alternative. For fixed τ, (2.2) is maximized by

$$\widehat{\lambda}_1 = \frac{N(\tau)}{S(\tau)}, \ \widehat{\lambda}_2 = \frac{n - N(\tau)}{S - S(\tau)}. \tag{2.6}$$

Substituting (2.6) into (2.2), we have

$$l(\tau) := l(\widehat{\lambda}_1, \widehat{\lambda}_2, \tau) = N(\tau) ln \frac{N(\tau)}{S(\tau)} + [n - N(\tau)] ln \frac{n - N(\tau)}{S - S(\tau)} - n. \tag{2.7}$$

Here, $l(\tau)$ is unbounded if τ tends to $T_{(n)}$. An estimate $\widehat{\tau}$ of τ is obtained by maximizing $l(\tau)$ over a restricted interval $[a, b]$, i.e.,

$$\widehat{\tau} = \arg \max_{a \leq \tau \leq b} l(\tau). \tag{2.8}$$

For example, Yao (1986) used $a = 0, b = T_{(n-1)}$. Worsley (1988) suggested in addition to use respectively the pth and $(1-p)$th sample quantiles for a and b, and Loader (1991) chose a and b by $u = S(a)/S$, and $1 - u = S(b)/S$.

Other methods to cope with the singularities include Matthews and Farewell (1985), who modified the likelihood function by using the probability $F(T_i + \epsilon) - F(T_i - \epsilon)$ instead of the density at T_i, and Worsley (1988), who suggested to artificially censor the largest observation so that $l(\tau)$ is always finite.

Note that for the pseudo (or restricted) MLE's $(\widehat{\lambda}_1, \widehat{\lambda}_2, \widehat{\tau})$ defined in (2.6) and (2.8), the restricted log likelihood ratio statistic for the hypotheses (2.1) is $l(\widehat{\tau})$. Hence H_0 is rejected if $l(\widehat{\tau}) \geq c$, where

$$\alpha = Pr_0\{l(\widehat{\tau}) \geq c\} \tag{2.9}$$

and Pr_0 is obtained under the null hypotheses of an exponential distribution with constant rate λ_0. With Loader's choice of $[a, b]$, the test is invariant under scale transformation and (2.9) is independent of λ_0. The critical value c was then derived by large deviation approximation of the boundary crossing probabilities of a Poisson process with rate λ_0.

Worsley (1988) derived the exact critical values for three situations: (i) $[a, b]$ equal to $[0, T_{(n-1)}]$, (ii) $[a, b]$ equal to [sample pth quantile, sample $(1-p)$th quantile], for $p = .1$, $.2$, and (iii) artificially censored largest observation. Henderson (1990) noted that the LRT is not sufficient and modified it by weighting and standardizing the likelihood ratio $l(\tau)$ at $\tau \in \{T_{(i)}, T_{(i)}- : i = 1, \ldots, n-1\}$. Exact critical values are also derived for this modified LRT.

(b) Score Test. Matthews, Farewell and Pyke (1985) considered tests based on maximal score statistics. They consider a variant of model (1.1) by reparametrizing using (λ, ξ, τ) with $\lambda_1 = \lambda$, and $\lambda_2 = (1-\xi)\lambda$, and testing the hypotheses that

$$H_0^{'} : \ \xi = 0 \ \text{ vs } \ H_a^{'} : \ 0 < \xi < 1. \tag{2.10}$$

Note that (2.10) corresponds to adding the restriction $\lambda_1 \geq \lambda_2$ to model (1.1).

Let $l(\lambda, \xi, \tau)$ denote the log likelihood function in (2.2) with (λ_1, λ_2) replaced by (λ, ξ). The normalized score statistic for a given τ and λ is then

$$Z_n(\tau, \lambda) = \left. \frac{\partial l / \partial \xi}{[E(-\partial^2 l / \partial \xi^2)]^{1/2}} \right|_{\xi=0} \tag{2.11}$$

$$= n^{-1/2} \sum_{i=1}^{n} e^{\lambda\tau/2}[(T_i - \tau)\lambda - 1]1(T_i \geq \tau).$$

Matthews et al. show that for a given λ and any $b < \infty$, the score statistic process $\{Z_n(\tau, \lambda) : 0 \leq \tau \leq b\}$ converges under H_0' weakly to the Ornstein-Uhlenbeck process $\{Z(\tau, \lambda) : 0 \leq \tau \leq b\}$ with mean 0 and covariance $\exp\{-(\lambda/2)|\tau_1 - \tau_2|\}$ which is a Brownian motion normalized to constant variance 1.

Thus if it is known that $a \leq \tau \leq b$, a suitable test for (2.10), when λ is known, is to reject $H_0{}'$ for $M_n(a, b, \lambda) \geq c$ or equivalently $T(c, \lambda) \leq b - a$, where $M_n(a, b, \lambda) := \sup_{a \leq \tau \leq b} Z_n(\tau, \lambda)$ and $T(c, \lambda) := \inf\{\tau \geq 0 : Z(\tau, \lambda) \geq c\}$ is the first passage time and c is the upper α-quantile of $\sup_{a \leq \tau \leq b} Z(\tau, \lambda)$. Approximation of c can be found in Mandl (1962) or Keilson and Ross (1975); a more easily accessible and very accurate approximation is provided by James, James and Siegmund (1987). Further numerical tables can be found in DeLong (1981).

When λ is unknown, the normalized score statistic $Z_n(\tau, \lambda)$ in (2.11) should be replaced by

$$\begin{aligned} \widehat{Z}_n(\tau) &= \left. \frac{\frac{\partial l}{\partial \xi}}{\frac{\partial^2 l}{\partial \xi^2} - \frac{(\partial^2 l/\partial \xi \partial \lambda)^2}{\partial^2 l/\partial \lambda^2}} \right|_{\xi=0, \lambda=\widehat{\lambda}_n} \\ &= (1 - e^{-\lambda\tau})^{-1/2} Z_n(\tau, \lambda)\big|_{\lambda=\widehat{\lambda}_n} \end{aligned} \qquad (2.12)$$

where $\widehat{\lambda}_n = n / \sum T_i$ is the MLE of λ under H_0'.

Setting $\widehat{Z}_n^\star(t) := \widehat{Z}_n(\tau)$ with $t = e^{-\lambda\tau}$, for $0 < t_0 \leq t_1 < 1$, $\widehat{Z}_n^\star(t)$ converges uniformly over $[t_0, t_1]$ to $\widehat{Z}^\star(t) := W^0(t)[t(1-t)]^{-1/2}$ with probability one, where W^0 is the standard Brownian bridge and $\widehat{Z}^\star$ is thus a Brownian bridge normalized to variance 1.

(c) Bayesian Test. Owing to the the drawback that the null distribution of LR type tests and Score tests depends on preassigned bounds a and b for the change-point τ, Yao (1987) proposed another type of test for the change-point alternative, by connecting the classical change-point problem in quality control (mentioned in the beginning of Section 1) with the parametric change-point problem in hazard functions. More precisely, consider the following two problems:

Problem A: Observe $T_1, ..., T_n$ with hazard function λ as in (1.1). We want to test:

$$H_0 : \lambda_1 = \lambda_2 \text{ vs } H_a^\star : \lambda_1 \geq \lambda_2. \qquad (2.13)$$

Problem B: Independent random variables $Y_1, ..., Y_n$ are observed with $Y_1, ..., Y_k$ distributed as an exponential distribution with rate λ_1, and $Y_{k+1}, ..., Y_n$ distributed as an exponential distribution with rate λ_2, where k is the unknown change-point. The problem is to test the hypotheses in (2.13).

Setting $T_{(1)} = 0$, define the normalized spacing to be $D_i = (n - i + 1)(T_{(i)} - T_{(i-1)})$, $i = 1, ..., n$. Denote $K = N(\tau)$ in (2.2), then K is binomial (n, ρ), with $\rho = 1 - e^{-\lambda\tau}$. Yao (1987) showed that $D_1, ..., D_n$ in Problem A play the role of $Y_1, ..., Y_n$ in problem B, with K in Problem A corresponding to k in Problem B. For Problem B with random change-point k, Hsu (1979) generalized a Bayesian test for (2.13) of Kander and Zacks (1966) with uniform prior on k.

Based on the connection between the two problems, Yao (1987) extended Hsu's test to Problem A which rejects H_0 if S_n is large, where $S_n = \sum_{i=1}^n iT_{(i)} / \sum_{i=1}^n T_{(i)}$. Exact levels of significance can be computed based on properties of order statistics and are given in Table 1 of Hsu (1979). Yao (1987) derived the asymptotic normality of S_n under H_0 and local alternatives and showed that the asymptotic Pitman relative efficiency (ARE) of S_n with respect to the score test is $ARE(\tau) = 3\rho(1 - \rho)$. The $ARE(\tau)$ thus attains its maximum .75 at $\rho = 1/2$ (i.e., τ is the median), and it decreases to 0 as ρ tends to 0 or 1. Thus the score test is more efficient than this Bayesian test based on S_n, especially when the change-point occurs either very early or late. The Bayesian test has the advantage that it is computationally simple and may be useful when the possibility of an early or late change-point is excluded.

2.2. Point Estimation and Confidence Sets. The first attempt to estimate the three parameter change-point model (1.1) was implicit in Anderson and Senthilselvan (1982), where model (1.1) appears as a special case of an extended Cox proportional hazards model. Anderson and Senthilselvan (1982) were motivated by a cancer mortality study in which some covariate effects may decay with time, and proposed a two-step proportional hazards model which allows time-varying covariate coefficients. The hazard function for an individual with covariate $z \in \Re^p$ is assumed to be $\lambda(t, z) = \lambda_0(t) \exp(\beta^T(t) z)$, where

$$\beta_j(t) = \begin{cases} \alpha_j, & \text{for } t \leq B \\ \gamma_j, & \text{for } t > B \end{cases} \qquad j = 1, ..., p.$$

This yields a two-step regression model with

$$\lambda(t, z) = \begin{cases} \lambda_0(t) e^{\alpha^T z}, & \text{for } t \leq B \\ \lambda_0(t) e^{\gamma^T z}, & \text{for } t > B, \end{cases}$$

which includes Model (1.1) as a special case with $\lambda_1 = e^{\alpha}$, $\lambda_2 = e^{\gamma}$, $B = \tau$ and $\lambda_0(t) = 1$. Following Cox's (1972) suggestion, the parameters α, γ and B are first estimated using conditional log-likelihood $\tilde{l}(\alpha, \gamma)$ given the value of B; then the baseline hazard function $\lambda_0(t)$, is estimated by the penalized maximum likelihood method, conditioning on the estimates $\widehat{\alpha}, \widehat{\gamma}$, and $\widehat{B}$. However, in the one-sample case as in model (1.1), $\tilde{l}(\alpha, \gamma) = -\sum_{i=1}^{n} ln(n - i + 1)$ is a constant. This method thus fails to produce estimates for the parameters in model (1.1).

(a) Maximum Likelihood Type Estimtors. Independently, Matthews and Farewell (1982) consider MLE's for model (1.1) in the context of deriving the LRT. No distributional results are given. As Nguyen, Rogers and Walker (1984) pointed out later, $l(\lambda_1, \lambda_2, \tau)$ may not be bounded unless $\lambda_1 \geq \lambda_2$. However, for a given τ, λ_1 and λ_2 can still be estimated as in (2.6). Using the observation that the density function corresponding to model (1.1) is a mixture of a truncated (on the right at τ) exponential (with rate λ_1) distribution and an exponential (with rate λ_2) distribution with delay τ, Nguyen, Roger and Walker (1984) construct a stochastic process $\{X_n(t), t \geq 0\}$ for which $X_n(\tau)$ converges to 0. An estimate $\widehat{\tau}$ should thus satisfy $X_n(\widehat{\tau}) = 0$. Using such a $\widehat{\tau}$ for τ, they showed the strong consistency of the estimators for τ, λ_1 and λ_2. The asymptotic distributions were not derived.

Yao (1986) avoids the singularity by restricting the MLE $\widehat{\tau}$ to be in the interval $[0, T_{(n-1)}]$, i.e., $\widehat{\tau} = \arg \sup_{0 \leq \tau \leq T_{(n-1)}} l(\tau)$, with $l(\tau)$ defined in (2.7). He also notices that $\widehat{\tau} \in \{T_{(1)}-, T_{(1)}, ..., T_{(n-1)}-, T_{(n-1)}\}$ so that it is sufficient to maximize $l(\tau)$ over $2(n-1)$ points only. Consistency of $\widehat{\tau}$ was established by connecting this model to the aforementioned model in Problem B. Following arguments by Chernoff and Rubin (1956), Yao showed that $n(\widehat{\tau} - \tau)$ converges in distribution to a similar limit as the MLE in Problem B, two independent random walks depending on λ_1, λ_2, and τ. Also $\sqrt{n}(\widehat{\lambda}_1 - \lambda_1)$, $\sqrt{n}(\widehat{\lambda}_2 - \lambda_2)$ and $n(\widehat{\tau} - \tau)$ are asymptotically independent, and the limiting distributions of the former two are normal .

Pham and Nguyen (1990) extended the result of Yao (1986) by maximizing $l(\tau)$ over a random compact set $[L_n, U_n]$, $0 \leq L_n < U_n < T_{(n)}$. However, they recommend Yao's choice of $L_n = 0$, $U_n = T_{(n-1)}$. Under additional assumptions, they showed the strong consistency of $\widehat{\tau}$ from which the strong consistency of $\widehat{\lambda}_1$ and $\widehat{\lambda}_2$ follows. Limiting distribution results as in Yao (1986) were also derived.

(b) Bayes Estimator. Bayes estimators were proposed in Achcar (1989) where it is assumed that the change-point τ is a discrete random variable with prior probability π_0. The mode of the respective marginal posterior density is

used to estimate τ, λ_1, and λ_2. No asymptotic results are available for these Bayes estimates.

(c) Confidence Regions. Note that the asymptotic distributional results in Yao (1986) and Pham and Nguyen (1990) can be used to construct approximate confidence intervals for λ_1, and λ_2. Loader (1991) adapted the likelihood ratio method of Siegmund (1988) to derive approximate confidence regions of the form

$$I_1 = \{t : l(t) \geq \sup_{a \leq t \leq b} l(t) - c_1\}$$

for τ; the selection of a and b was discussed earlier. Approximate joint confidence regions for τ and jump size $\delta = ln(\lambda_2/\lambda_1)$ were obtained by

$$I_2 = \{(t, \delta) : l(t \mid \delta) \geq \sup_{a \leq t \leq b} l(t) - c_2\}.$$

2.3. Censored Data. To conclude this brief review, we note that very little is known on the parametric hazard change-point model under the important case of random censoring. The Bayes estimator in Achcar (1989) can accommodate fixed (nonrandom) censoring but no sampling results are derived for this Bayes estimator. Worsley (1988) showed that the null distribution of his test statistics is unchanged for Type II censoring. As for other types of censoring as occurring in the leukemia data in Matthews and Farewell (1982), Worsley (1988) simply excluded the 33 (out of 84) censored observations from the analysis following the suggestion of Matthews and Farewell (1982) that moderate censorship has little impact on the null distribution of the likelihood ratio. Notice that this approach may create biases in the procedures even in cases of moderate censoring. Another approach, e.g. in Matthews, Farewell and Pyke (1985), is to adjust the likelihood function by incorporating contributions from censored observations and then to apply sampling results based on the uncensored case. Since the sampling results under censoring may differ from those of the uncensored case such an approach is questionable.

Loader (1991) noted that, under random censoring, the log likelihood functions denoted by $l_c(\lambda_1, \lambda_2, \tau)$ and $l_c(\tau)$ are obtained from (2.2) and (2.7) by replacing $N(t)$ and n respectively by $N_c(t)$ and n_c, where $N_c(t) = \sum_{i=1}^n I(Z_i \leq t, \delta_i = 1)$, is the number of uncensored failures at time t, and $n_c = \sum_{i=1}^n I(\delta_i = 1)$, is the total number of uncensored observations. The ML type estimators $\widehat{\lambda}_1$, $\widehat{\lambda}_2$, and $\widehat{\tau}$ can then be obtained as in (2.6) and (2.8) with this modification, and

$$\widehat{\lambda}_1 = \frac{N_c(\tau)}{S(\tau)}, \qquad \widehat{\lambda}_2 = \frac{n_c - N_c(\tau)}{S - S(\tau)}, \qquad \widehat{\tau} = \arg\max_{a \leq \tau \leq b} l_c(\tau),$$

where $S(\tau)$ and S are defined as before in (2.4) and (2.5).

Loader (1991, p.752) suggested how to approximate the significance level by conditioning on n_c and the censoring status of the largest observation $Z_{(n)}$.

Explicit sampling and asymptotic results for inference procedures based on censored data will be of keen interest.

3. Kernel Methods.

3.1. Change-point Estimates. In this section we discuss estimates for the change-point τ and the hazard function $\lambda(x)$ under both Model 2 and Model 3. The lifetime variables are subject to random censoring in the sense that there is a sequence of i.i.d. censoring variables $C_1, ..., C_n$ and one observes only (X_i, δ_i), $i = 1, ..., n$, where $X_i = \min(T_i, C_i)$, $\delta_i = I(T_i \leq C_i)$. We assume that T_i is independent of C_i, and denote the survival function of X_i, by $\bar{H}(x) = 1 - H(x)$, where $H(x) = P(X_i \leq x)$. The following considerations can be extended to cover change-points in a derivative (cf. Müller and Wang (1990b)).

Let $K \in S \cap M \cap C$ be a kernel function, i.e., a real function which belongs to a class of functions with well-defined support S, to a class M satisfying certain moment conditions, and to a class C satisfying certain smoothness conditions. We will consider the following special classes:

$$M(0) = \{f : \int f(x)dx > 0\},$$

$$M(\nu, k) = \{f : \int f(x)x^j dx = 0, \quad 0 \leq j < k, j \neq \nu,$$

$$\int f(x)x^\nu dx = (-1)^\nu \nu!, \int f(x)x^k dx = \beta_{\nu,k} \neq 0\},$$

$$\text{for integers } \nu, k, \ 0 \leq \nu < k;$$

$$S(q) = \{f : \text{support}(f) = [-q, 1]\};$$

$C(\alpha, \beta) = \{f : \int |df| < \infty, \quad f$ is Lipschitz continuous and $\min(\alpha, \beta)$ times continuously differentiable on $S(q)$, $f^{(j)}(-q) = 0, \quad 0 \leq j \leq \alpha, f^{(\alpha)}(-q) > 0, \quad f^{(j)}(1) = 0, \quad 0 \leq j < \beta\}$, for nonnegative integers α, β.

Note that $C(\alpha, \beta) \subset C(\alpha', \beta')$ if $\alpha \leq \alpha'$ and $\beta \leq \beta'$. We are now in a position to define the required kernel estimators:

$$\hat{\lambda}^{(\nu)}(x) = \frac{1}{b^{\nu+1}} \sum_{i=1}^{n} K_\nu\left(\frac{x - X_{(i)}}{b}\right) \frac{\delta_{(i)}}{n - i + 1} \quad \text{for } \lambda^{(\nu)}(x), \tag{3.1}$$

and

$$\hat{\lambda}_\pm(x) = \frac{1}{b} \sum_{i=1}^{n} K_\pm\left(\frac{x - X_{(i)}}{b}\right) \frac{\delta_{(i)}}{n - i + 1} \Big/ \int K_-(x)dx \quad \text{for } \lambda_\pm(x), \tag{3.2}$$

for $\lambda_{\pm}(x)$, where $\lambda_{+}(x) = \lim_{y \downarrow x} \lambda(y)$, and $\lambda_{-}(x) = \lim_{y \uparrow x} \lambda(y)$. Here, $b = b(n)$ is a sequence of bandwidths (smoothing parameters) and we use kernels satisfying (as minimum requirements) $K_\nu \in S(1) \cap M(\nu, k) \cap C(0,0)$, for a $k > \nu$, $K_+(x) = K_-(-x)$ and $K_- \in S(0) \cap C(0,1) \cap M(0)$. Let $Q = \int K_-(x)dx$. Under the stronger requirement $K_- \in S(0) \cap C(0,0) \cap M(0,2)$, we find $Q = 1$.

In the Smooth Approximation Model (Model 3), $\tau = \arg\max \lambda^{(1)}(x)$, and it is suggested in Müller and Wang (1990a) to estimate τ by

$$\tilde{\tau} = \arg\max \widehat{\lambda}^{(1)}(x); \tag{3.3}$$

in practice, $\tilde{\tau}$ may be determined as a zero of $\widehat{\lambda}^{(2)}(x)$. The obvious estimator for Model 2 is:

$$\widehat{\tau} = \arg\max\{\widehat{\lambda}_+(x) - \widehat{\lambda}_-(x)\}. \tag{3.4}$$

3.2. Asymptotic Properties. The following result is shown in Müller and Wang (1990a).

PROPOSITION 1. *Assume that $\lambda(.)$ is four times continuously differentiable, $\lambda^{(3)}(\tau)\bar{H}(\tau) \neq 0$, $K_1 \in S(1) \cap M(1,3) \cap C(3,3)$, and $b \to 0$, $nb^6 \to \infty$, $nb^9 \to d^2$, where $0 \le d < \infty$, as $n \to \infty$, then*

$$(nb^5)^{1/2}(\tilde{\tau} - \tau) \longrightarrow \mathcal{N}\left(\frac{d}{24} \int K_1^{(1)}(v) v^4 dv \frac{\lambda^{(4)}(\tau)}{\lambda^{(3)}(\tau)}, \frac{\lambda(\tau)}{\lambda^{(3)}(\tau)^2} \frac{\int K_1^{(1)}(v)^2 dv}{\bar{H}(\tau)}\right).$$

Contrasting this with the Model 2 estimator $\widehat{\tau}$, we first show in complete analogy to the proof of Lemma 4.1 in Müller (1992), substituting a result like that given by Yandell (1983) for the uniform convergence of derivatives of hazard kernel estimates:

PROPOSITION 2. *Assume that $\lambda(\cdot)$ is two times continuously differentiable, $K_- \in S(0) \cap M(0) \cap C(\mu, \mu)$ for some $\mu \ge 1$, $b \to 0$, $\frac{nb}{(\log n)^2} \to \infty$, then*

$$|\widehat{\tau} - \tau| = O_p\left([b^{2\mu - 1}/n]^{1/(2\mu)}\right).$$

The following functional limit theorem is proved by verifying tightness and multivariate weak convergence and requires an i.i.d. representation of the kernel estimators $\widehat{\lambda}_{\pm}(x)$

PROPOSITION 3. *Assume that the assumptions of Proposition 2 are satisfied with $\mu = 3$ and that $nb^3 \to d^2$, where $0 \le d < \infty$. Then the processes*

$$\eta_n(x) = (nb)^{2/3}\left[\{\widehat{\lambda}_+(\tau + b(\frac{x}{nb})^{1/6}) - \widehat{\lambda}_-(\tau + b(\frac{x}{nb})^{1/6})\} - \{\widehat{\lambda}_+(\tau) - \widehat{\lambda}_-(\tau)\}\right]$$

converge weakly on $C([-R,R])$ for any $R > 0$ to the continuous Gaussian process

$$\eta(x) = -\frac{1}{24}\Delta_0 x^4 K_-^{(3)}(0) + \Delta_1 x d + \mathcal{Z}x,$$

$x \in [-R,R]$*, where $\mathcal{Z}$ is a random variable with*

$$\mathcal{Z} \sim \mathcal{N}\left(0, \bar{H}(\tau)^{-1}\{\lambda_+(\tau)+\lambda_-(\tau)\}\int K_-^{(1)}(v)^2 dv\right).$$

From Propositions 2 and 3, we infer

PROPOSITION 4. *Assume that the assumptions of Propositions 2 and 3 are satisfied. Then*

$$\left(\frac{n}{b^5}\right)^{1/2}(\hat{\tau}-\tau)^3 \longrightarrow$$

$$\mathcal{N}\left(\frac{6\Delta_1 d}{\Delta_0 K_-^{(3)}(0)}, \left\{\frac{6}{\Delta_0 K_-^{(3)}(0)}\right\}^2 \{\lambda_+(\tau)+\lambda_-(\tau)\}\left\{\frac{\int K_-^{(1)}(v)^2 dv}{\bar{H}(\tau)}\right\}\right).$$

Note that Propositions 3 and 4 provide an invariance principle for estimated change-points, when the jump size Δ_0 is fixed. It has been suggested that kernels which are discontinuous at 0 (and therefore not covered by the assumptions of Propositions 3 and 4) may yield faster rates of convergence. While for hazard functions under random censoring, this is still an open problem at this point, it is to be expected that better rates will not lead to an invariance principle as long as the jump size Δ_0 is fixed, in analogy to Chernoff and Rubin (1956). Invariance principles are then only obtainable in the "contiguous" case where $\Delta_0 \to 0$ as $u \to \infty$.

3.3. Model Adaptive Change-point Estimation. The two estimators $\tilde{\tau}$ and $\hat{\tau}$ for τ appear to be quite different, $\tilde{\tau}$ being motivated by the Smooth Approximation Model (Model 3), whereas $\hat{\tau}$ is motivated by the Nonparametric Change-point Model (Model 2). However, it is easy to demonstrate that these two types of estimators are closely connected and actually are applicable in both models: Assume $\tilde{\tau}$ utilizes a kernel $K_1 \in S(1)\cap M(1,3)\cap C(3,3)$ as required in Proposition 1. Note that we may require K_1 to satisfy $K_1(x) = -K_1(-x)$. Defining $\tilde{K}_+ = K_1 1_{[-1,0]}$, $\tilde{K}_- = -K_1 1_{[0,1]}$, we find that for $\hat{\tau}$ based on $\tilde{K}_+$, $\tilde{K}_-$, $\hat{\tau} = \tilde{\tau}$. Within the framework of Model 2, this follows from $\{1/b\}[\hat{h}_+(x) - \hat{h}_-(x)] = \hat{h}^{(1)}(x)$, noting that $\tilde{K}_+(x) = \tilde{K}_-(-x)$ and $\tilde{K}_- \in S(0)\cap M(0)\cap C(1,3)$, which means that Proposition 2 applies for $\mu = 1$, so that $\tilde{\tau}$ is consistent not only within Model 3 (according to Proposition 1), but also within Model 2.

If Model 2 applies, one may achieve O_p-rates $(\log n)^2/n$ for small bandwidths, see Proposition 2. If Model 3 applies, the O_p-rate is $n^{-2/9}$, which is considerably slower. In this sense the estimator $\tilde{\tau}$ is adaptive to the underlying model. Of course, the construction of confidence regions still requires knowledge of the correct model, and also (for both models) knowledge of auxiliary quantities like $\bar{H}(\tau)$, $\{\lambda_+(\tau)+\lambda_-(\tau)\}$ for Model 2 and $\lambda^{(4)}(\tau)/\lambda^{(3)}(\tau)$, $\lambda(\tau)/\{\lambda^{(3)}(\tau)^2\bar{H}(\tau)\}$ for Model 3. Substituting $\tilde{\tau}$, these quantities can be estimated consistently.

Consider now $\hat{\tau}$ based on $K_- \in S(0) \cap C(3,3) \cap M(0)$, such that $R = \int K_-(x)x\,dx > 0$, $\int K_-(x)x^3 dx \neq 0$; let $K_+(x) = K_-(-x)$, and define $\tilde{K}_1 = \{K_+ - K_-\}$. Then $\tilde{\tau} = \hat{\tau}$, when $\tilde{\tau}$ is based on $\tilde{K}_1$. This follows from $(1/2R)\tilde{K}_1 \in S(1) \cap C(3,3) \cap M(1,3)$. Therefore, $\hat{\tau}$ when based on a smooth kernel is also model adaptive.

3.4. Estimating the Hazard Function. In applications, the problem may not be restricted to estimating τ, but may include the estimation of the hazard function. In this case, assuming Model 2, a two-step procedure is a natural approach: First estimate τ via $\hat{\tau}$ or $\tilde{\tau}$ and an initial bandwidth b_1. Then adapt the hazard function estimate employing a second bandwidth b_2 to this estimated change-point which assumes the role of an endpoint. Within the interval $[\hat{\tau}-b, \hat{\tau}+b]$, special "boundary kernels" with asymmetric support can be used to obtain good global consistency properties in analogy to Section 4 in Müller (1992); unmodified kernel estimators will be strongly biased near the discontinuity at τ and will produce a curve estimate which is deceptively "smooth".

Define a family of kernels $K_+(.,q)$, $0 \le q \le 1$, which should vary "smoothly" in q and satisfy $K(.) = K_-(.,1)$, $K_+(x,q) = K_-(-x,q)$, where $K(.)$ is the (symmetrically supported) kernel to be used in $\hat{\lambda}(.,K)$ outside $[\hat{\tau}-b_2, \hat{\tau}+b_2]$ and $\hat{\lambda}(.,K)$ denotes estimator (3.1) using kernel K. Requiring $K \in S(1) \cap C(\mu,\mu) \cap M(0,2)$, $K(.,q) \in S(q) \cap C(\mu-1,\mu) \cap M(0,2)$, the final estimate of $\lambda(.)$ in Model 2 is then

$$\tilde{\lambda}(x) = \begin{cases} \hat{\lambda}(x,K), & |x-\hat{\tau}| > b_2 \\ \hat{\lambda}(x,K_-(.,q)), & \hat{\tau}-b_2 \le x \le \hat{\tau}, \quad q = (\hat{\tau}-x)/b_2\,, \\ \hat{\lambda}(x,K_+(.,q)), & \hat{\tau} < x \le \hat{\tau}+b_2, \quad q = (x-\hat{\tau})/b_2 \end{cases}$$

where $\hat{\lambda}$ employs bandwidth b_2 and $\hat{\tau}$ is determined with bandwidth b_1.

4. A Simulation Study on Effects of Kernel and Bandwidth. A simulation study was carried out to explore the influence of bandwidth and kernel choice on the mean squared error of the change-point estimate. The following kernels were used: $K_1 = \frac{15}{4}(x^3 - x)$ $K_2 = \frac{315}{32}(x^7 - 3x^5 + 3x^3 - x)$ as kernels for estimating the first derivative as needed for $\tilde{\tau}$, and

$$K_{-1} \equiv 1_{[0,1]}, \quad K_{-2} \equiv (1-x)1_{[0,1]},$$
$$K_{-3} \equiv (1-x)p_3(x)1_{[0,1]}, \quad \text{and } K_{-4} \equiv x^3(1-x)^3 p_4(x)1_{[0,1]},$$

where the linear polynomials p_3, p_4 are chosen in such a way that K_{-3}, $K_{-4} \in M(0,2)$ (which determines the two coefficients uniquely), as kernels for estimating one-sided limits in $\hat{\tau}$.

Note that $K_{-1} \in S(0) \cap C(0,0) \cap M(0)$, $K_{-2} \in S(0) \cap C(0,1) \cap M(0)$, $K_{-3} \in S(0) \cap C(0,1) \cap M(0,2)$, $K_{-4} \in S(0) \cap C(3,3) \cap M(0,2)$, $K_1 \in S(1) \cap C(1,1) \cap M(1,3)$, and $K_2 \in S(1) \cap C(3,3) \cap M(1,3)$. K_{-1} - K_{-4} are therefore one-sided kernels, where K_{-1}, K_{-2} do not satisfy particular moment conditions and K_{-2} is continuous at the endpoint of its support opposite to the possible change-point, whereas K_{-1} is not. K_{-3} and K_{-4} have different smoothness, K_{-3} being discontinuous at the endpoint near the possible change-point whereas K_{-4} is very smooth at both endpoints. Both K_{-3} and K_{-4} satisfy $M(0,2)$ and provide therefore consistent estimates of one-sided limits, in contrast to K_{-1} and K_{-2}. Antisymmetric kernels K_1 and K_2 are both in $M(1,3)$ and therefore yield consistent derivative estimates, but differ in their smoothness at the endpoints, K_2 being very smooth, while K_1 is just continuous on the real line.

The following two cases were considered for the hazard function:

Case A: $\lambda(x) = 0.3, \quad 0 \le x \le 3, \quad \lambda(x) = 0.05, \quad x > 3;$

Case B: $\lambda(x) = 0.1x, \quad 0 \le x \le 3, \quad \lambda(x) = 0.05 + 0.1(x-3), \quad x > 3,$

so that Model 2 applies with a change-point at 3.0. Each simulation was based on 200 Monte Carlo runs and $n = 200$ data. Independent random censoring was applied with exponential distribution ($\lambda = 0.1$). The outcome of one typical sample run for Case A is shown in Figure 1: The estimator $\hat{\tau}$ based on K_{-2} and b_1=1.75 is $\hat{\tau} = 2.96$ for this sample and the modified hazard function estimator $\tilde{\lambda}(.)$ (3.5) was then constructed with boundary kernels $K_-(x,q) = \frac{12}{(1+q)^4}(1-x)(x(2q-1) + (3q^2 - 2q + 1)/2)$, where $K_-(x,0) = K_{-3}(x)$ and $K_-(x,1) = \frac{3}{4}(1-x^2)1_{[-1,1]}$.

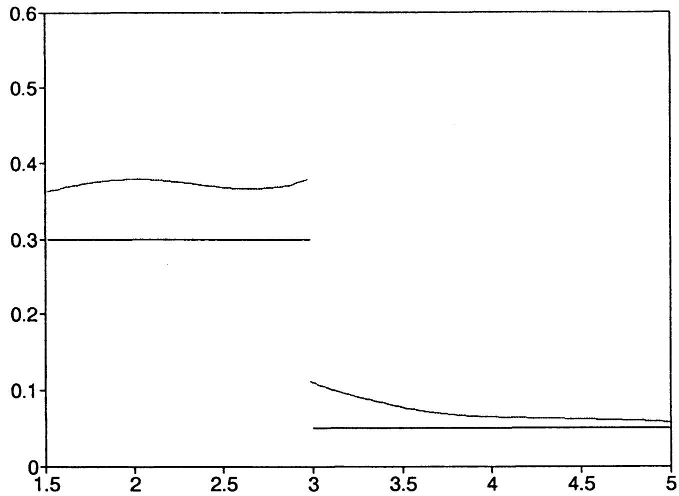

Figure 1: True hazard function (solid) and estimated hazard function (dashed)

The simulation results are given in Table 1 for Case A and in Table 2 for Case B in terms of the mean squared error (MSE) of $\tilde{\tau}$ (with kernels K_1, K_2) and of $\hat{\tau}$ (with kernels $K_{-1} - K_{-4}$).

Table 1: Results of a Monte Carlo study for Case A, based on 200 simulations. Entries are mean squared errors for various change-point estimators, using bandwidths b and kernels K_1 or K_2 for the derivative based estimator $\tilde{\tau}$, $K_{-1}-K_{-4}$ for the estimator $\hat{\tau}$ based on one-sided kernels. For more details and kernels, see text. No result is reported if in more than 15% of all simulations the change-point could not be located.

	Kernel					
b	K_1	K_2	K_{-1}	K_{-2}	K_{-3}	K_{-4}
0.25	0.219	0.249	0.243	0.262	0.263	0.320
0.75	0.066	0.090	0.061	0.098	0.281	0.269
1.25	0.069	0.042	0.067	0.062	0.224	0.216
1.75	0.086	0.052	0.058	0.056	0.143	0.219
2.25	0.097	0.060	0.054	0.058	0.092	0.165
2.75	0.077	0.069	0.024	0.037	0.086	0.175
3.25	0.079	0.052	0.062	0.024	0.100	0.371
3.75	0.286	0.049	0.319	0.026	0.094	0.415
4.25	-	0.106	0.439	0.653	0.077	0.272

Table 2: As Table 1, but for Case B.

	Kernel					
b	K_1	K_2	K_{-1}	K_{-2}	K_{-3}	K_{-4}
0.25	0.085	0.111	0.084	0.114	0.244	0.242
0.75	0.036	0.036	0.027	0.018	0.130	0.177
1.25	-	0.029	0.035	0.020	0.062	0.193
1.76	-	0.024	0.061	0.020	0.024	0.168
2.25	-	-	-	0.037	0.019	0.092
2.75	-	-	-	-	0.012	0.128
3.25	-	-	-	-	0.013	0.166
3.75	-	-	-	-	0.011	0.190
4.25	-	-	-	-	0.011	0.243

Notice that bandwidth and kernel choice strongly influence the MSE of the change-point estimates. Kernel K_2 appears to be consistently better than K_1, and "smooth" kernel K_{-4} is often worse than the other methods. In Case A, discontinuous kernels K_{-1} and K_{-2} appear to be slightly better than continuous kernel K_{-3}, as $K_{-3} \in M(0,2)$ is subject to an additional constraint which increases the variation of the kernel and ultimately the variance of the corresponding estimates. In the more interesting Case B however, kernels K_1, K_2, K_{-1}, and K_{-2} are not useful, as for many bandwidth choices the change-point could not be located within the interval $[2,4]$ on which a change-point was sought; the apparent reason is that exactly these four kernels out of the six kernels do not provide asymptotically unbiased (in the case of K_1, K_2, implicit) estimates of the one-sided limits of $\lambda(.)$. This shows that certain discontinuous kernels do have a disadvantage.

Our recommendation is therefore to use kernel K_{-3}, which places relatively large mass towards the vicinity of a possible change-point and at the same time retains the moment conditions $M(0,2)$. To keep variances down, it is also advisable to choose relatively large bandwidths. It remains an open problem to devise and motivate more specific bandwidth choice procedures for nonparametric hazard change-point estimation.

REFERENCES

ACHCAR, J. A. (1989). Constant hazard against a change-point alternative: a Bayesian approach with censored data. *Comm. Statist. Theor. Meth.* **18**, 3801–3819.

ANDERSON, J. A. and SENTHILSELVAN, A. (1982). A two-step regression model for hazard function. *Appl. Statist.* **31**, 44–51.

ANTONIADIS, A. and GRÉGOIRE, G. (1991) Nonparametric estimation in change-point hazard rate model for censored data; a counting process approach. *Technical Report, Lab Modélisation et Calcul.*, University of Grenoble.

CHERNOFF, H. and RUBIN, H. (1956). The estimation of the location of a discontinuity in density. *Proc. 3rd Berkeley Symp. Math. Statist. Prob.* **1**, 19–37. Berkeley: Univ. of California Press.

COX, D. R. (1972). Regression models and life tables (with Discussion). *J. R. Statist. Soc. B* **34**, 187–202.

DELONG, D. M. (1981). Crossing probabilities for a square root boundary by a Bessel process. *Comm. in Statist., A* **10**, 2197–2213.

HENDERSON, R. (1990). A problem with the likelihood ratio test for a change-point hazard rate model. *Biometrika* **77**, 835–43.

HSU, D. A. (1979). Detecting shifts of parameter in gamma sequences with applications to stock price and air traffic flow analysis. *J. Amer. Statist. Assn.* **74**, 31–40.

JAMES, B., JAMES, K. L. and SIEGMUND, D. O. (1987). Tests for a change-point. *Biometrika* **74**, 71–83.

KANDER, A. and ZACKS, S. (1966). Test procedures for possible changes in parameters of statistical distributions occurring at unknown time points. *Ann. Math. Statist.* **37**, 1196–1210.

KEILSON, J. and ROSS, H. F. (1975). Passage time distribution for Gaussian Markov (Ornstein-Uhlenbeck) statistical processes. *In Selected Tables in Mathematical Statistics* **3**, 233–328. (Institute of Mathematical Statistics, ed.) American Mathematical Society, Providence.

LOADER, C. R. (1991). Inference for a hazard rate change model. *Biometrika* **78** 749–757.

MANDL, P. (1962). On the distribution of the time which the Uhlenbeck process requires to exceed a boundary (in Czechoslovakian with Russian and German summaries). *Apl. Mat.* **7**, 141–148.

MATTHEWS, D. E. and FAREWELL, V. T. (1982). On testing for a constant hazard against a change-point alternative. *Biometrics* **38**, 463–468.

MATTHEWS, D. E. and FAREWELL, V. T. (1985). On a singularity in the likelihood for a change-point hazard rate model. *Biometrika* **72**, 703–704.

MATTHEWS, D. E., FAREWELL, V. T. and PYKE, R. (1985). Asymptotic score-statistic processes and tests for constant hazard against a change-point alternative, *Ann. Statist.* **13**, 583–591.

MILLER, R. G. (1960). Early failure in life testing. *J. Amer. Statist. Assoc.* **55**, 591–602.

MÜLLER, H. G. (1992). Change-points in nonparametric regression analysis. *Ann. Statist.* **20**, 737–761.

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

MÜLLER, H. G. and WANG, J.-L. (1990b). A functional limit theorem for change-points in smooth hazard functions under random censoring. Technical Report.

NGUYEN, H. T., ROGERS, G. S. and WALKER, E. A. (1984). Estimation in change-point hazard rate models. *Biometrika* **71**, 299–304.

PHAM, D. T. and NGUYEN , H. T. (1990) Strong consistency of the maximum likelihood estimator in the change-point hazard rate model. *Statistics* **21**, 203–216.

SIEGMUND, D. O. (1988). Confidence sets in change-point problems. *Int. Statist. Rev.* **56**, 31–48.

WORSLEY, K. J. (1988). Exact percentage points of the likelihood-ratio test for a change-point hazard-rate model. *Biometrics* **44**, 259–263.

YAO, Y. C. (1986) Maximum likelihood estimation in hazard rate model with a change-point. *Comm. Statist. Theor. Meth.* **15**, 2455–2466.

YAO, Y. C. (1987) A note on testing for constant hazard against a change-point alternative. *Ann. Inst. Statist. Math.* **39**, 377–383.

YANDELL , B. S. (1983). Nonparametric inference for rates with censored data. *Ann. Statist.* **11**, 1119–1135.

DIVISION OF STATISTICS
UNIVERSITY OF CALIFORNIA
DAVIS, CA 95616