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

ESTIMATION IN CHANGE-POINT HAZARD RATE MODELS WITH RANDOM CENSORSHIP

BY I-SHOU CHANG, CHEN-HSIN CHEN AND CHAO A. HSIUNG

National Central University and Academia Sinica

Hazard rate models with a change-point allowing for random censorship are considered. An estimator of the change-point is proposed by examining a functional of Nelson-Aalen type estimator in the context of counting processes. Consistency and asymptotic distribution of the proposed estimator are established by martingale inequalities and Poisson approximation respectively. The performance of the proposed estimator is compared with that of a constrained maximum likelihood estimator using simulations. Robustness of the proposed estimator is also discussed.

1. Introduction. Let X be a random variable representing the time to some event, for example, the time-to-relapse after remission for leukemia patients. Several authors considered a model for the distribution of X specified by the hazard rate

$$\lambda(x) = \beta + \theta 1_{[\tau,\infty)}(x), \tag{1.1}$$

where 1_S is the indicator function of a set $S, \beta \geq 0, \beta + \theta > 0$ and τ is a change-point parameter.

In particular, Matthews and Farewell (1982) and Matthews, Farewell and Pyke (1985) studied the problem of testing for a constant hazard rate against alternatives with hazard rates involving a single change-point. The former presented a likelihood ratio test, and the latter proposed tests based on maximal score statistics and derived the asymptotic significance levels. Recently, Henderson (1990) suggested some modified likelihood ratio tests and presented an extensive literature review. Loader (1991) derived large deviation approximations to the significance level of the likelihood ratio test by a random change of time scale for the empirical process.

As for the problem of estimation, Nguyen, Rogers and Walker (1984) observed that the likelihood function is unbounded when τ is just before the largest observation and proposed a consistent estimator of the change-point by

---

AMS 1991 Subject Classification: Primary 62F10, Secondary 60G05, 62G20

Key words and phrases: Censoring, change-point, counting processes, hazard rate, Poisson approximation.

examining certain properties of the moments of the mixture density for a complete sample. Later Yao (1986) suggested the use of the maximum likelihood estimator for τ subject to the constraint that τ is not greater than the second largest observation, and obtained consistency and limiting distributions for the proposed estimators under a complete sample.

A related problem has also been worked on by Akman and Raftery (1986), who analyzed a change-point Poisson process and provided point and interval estimates of the change-point. Müller and Wang (1990) considered a different approach to model change-points by using kernel methods in estimating the point of the most rapid change of a continuous hazard function.

In this paper, we take account of random censorship in the model and investigate the estimation procedure in the context of counting processes. We provide an estimator of the change-point by examining a functional of Nelson-Aalen (1978) type estimator. Consistency of the proposed estimator is established by some martingale inequalities and the asymptotic distribution is obtained by Poisson approximation. We would like to point out that the approach in this paper is nonparametric in nature and can be applied to more general change-point models than (1.1). The expected robustness of the proposed estimator will be discussed at the end of this paper.

This paper is organized as follows. Section 2 defines the estimator of the change-point and establishes its asymptotic consistency. The weak convergence is obtained in Section 4. Section 5 and 6 compare the present approach with the constrained maximum likelihood estimates by providing simulation studies, which require estimates of the hazard rates given in Section 3.

2. A Consistent Estimator of the Change-Point. Let $(X_i, C_i), i = 1, \cdots, n$, be a random sample of positive vectors and the X_i and C_i be independent. Assume that the hazard rate of X_i is of the form (1.1), and there are known constants τ_1, τ_2 such that $0 < \tau_1 \leq \tau \leq \tau_2 < \infty$. We assume also that $\theta > 0$. The case $\theta < 0$ can be treated similarly.

Let $T_i = X_i \wedge C_i$ and assume $P\{T_i \geq t\} > 0$, for all $t > 0$. Based on the right censored data $\{(T_i, 1_{\{X_i \leq C_i\}}), i = 1, \cdots, n\}$, the estimation procedures are derived as follows. Let $N_i(t) = 1_{[X_i,\infty)}(t \wedge C_i)$, $H_i(t) = 1_{(0,T_i]}(t)$, and

$$A_n(t) = \int_0^t \left\{ \sum_{i=1}^n H_i(s) \right\}^{-1} d \sum_{i=1}^n N_i(s), t \geq 0, \tag{2.1}$$

be a Nelson-Aalen (1978) type estimator for the cumulative hazard $A(t) =$

$\int_0^t \lambda(s)ds$. Considering the basic martingale

$$M_n(t) = \sum_{i=1}^{n} N_i(t) - \int_0^t \sum_{i=1}^{n} H_i(s)\lambda(s)ds,$$

we know that

$$A_n(t \wedge T_{(n)}) - A(t \wedge T_{(n)}) = \int_0^{t\wedge T_{(n)}} \left\{ \sum_{i=1}^{n} H_i(s) \right\}^{-1} dM_n(s) \tag{2.2}$$

is a mean zero, square integrable martingale, where $T_{(n)} = \max\{T_i, 1 \le i \le n\}$, and

$$E\{A_n(t \wedge T_{(n)}) - A(t \wedge T_{(n)})\}^2 = E\int_0^{t\wedge T_{(n)}} \left\{ \sum_{i=1}^{n} H_i(s) \right\}^{-1} \lambda(s)ds. \tag{2.3}$$

Let $T > \tau_2$, and $g(x) = x^p, 0 \le p \le 1$. Define

$$Y_n(t) = \left\{ \frac{A_n(T) - A_n(t)}{T-t} - \frac{A_n(t) - A_n(0)}{t} \right\} g(t(T-t)), 0 < t < T, \tag{2.4}$$

$$Y(t) = \left\{ \frac{A(T) - A(t)}{T-t} - \frac{A(t) - A(0)}{t} \right\} g(t(T-t)), 0 < t < T. \tag{2.5}$$

Then,

$$\begin{aligned} Y(t) &= A(T)g(t(T-t))/(T-t) - A(t)g(t(T-t))T/\{(T-t)t\} \\ &= \theta\frac{T-\tau}{T-t} g(t(T-t))1_{\{t\le\tau\}} + \theta\frac{\tau}{t} g(t(T-t))1_{\{t>\tau\}}. \end{aligned} \tag{2.6}$$

Observing that $Y(t)$ is increasing on $[0, \tau]$ and decreasing on $[\tau, T]$, we consider the estimator

$$\hat{\tau}_n = \inf\{t \in [\tau_1, \tau_2] : Y_n(t\pm) = \sup Y_n(u)\}, \tag{2.7}$$

with sup hereafter abbreviating for the supremum attained in $[\tau_1, \tau_2]$. Here $Y_n(t\pm)$ denotes the right hand limit or left hand limit at t.

Let $T_{n1} < T_{n2} < \cdots < T_{nu_n}$ be the order statistics of the uncensored T_i's. If $p \ge \frac{1}{2}$, it can be shown that on each interval $[T_{n,i-1}, T_{ni})$, $Y_n(t) = t^{p-1}(T-t)^{p-1}\{tA_n(T) - TA_n(T_{n,i-1})\}$ is increasing. This shows that $\hat{\tau}_n$ equals to some uncensored time T_{ni}, or one of the end-points τ_1 and τ_2.

The consistency of $\hat{\tau}_n$ is established in the following.

THEOREM 2.1. *The estimator $\hat{\tau}_n$ is consistent for τ.*

PROOF. For $\epsilon > 0$ sufficiently small, there exists a constant c_1, which depends only on $\epsilon, \tau_1, \tau_2, \theta$ and p, such that $Y(\tau) - Y(\tau \pm \epsilon) \geq c_1$. Then, for sufficiently large n,

$$\begin{aligned} P\{|\hat{\tau}_n - \tau| > \epsilon\} &\leq P\{|Y(\hat{\tau}_n) - Y(\tau)| > c_1\} \\ &\leq P\{\sup|Y_n(t) - Y(t)| > c_1/2\} \\ &\leq P\{T\tau_1^{p-1}(T-\tau_2)^{p-1} \sup|U_n(t)| \\ &\quad + \tau_2^p (T-\tau_2)^{p-1}|U_n(T)| > c_1/2\}, \end{aligned} \tag{2.8}$$

where $U_n(t) = A_n(t) - A(t)$ is a martingale. We note that the second inequality in the above follows from the fact that Y and Y_n attain their maximum at τ and $\hat{\tau}_n$ respectively. Consequently, it follows from (2.8) that there exist $c_2 > 0, c_3 > 0$, depending on c_1, τ_1, τ_2, T and p such that

$$P\{|\hat{\tau}_n - \tau| > \epsilon\} \leq P\{\sup|U_n(t)| > c_2\} + P\{|U_n(T)| > c_3\}. \tag{2.9}$$

Let the first term on the right of (2.9) be denoted by I and the second term by II. Thus

$$\begin{aligned} I &\leq P\{\sup|U_n(t)| > c_2, \tau_2 \leq T_{(n)}\} + P\{T_{(n)} < \tau_2\} \\ &\leq P\{\sup|U_n(t \wedge T_{(n)})| > c_2\} + \prod_{i=1}^{n} P\{T_i < \tau_2\} \\ &\leq c_2^{-2} E(\sup|U_n(t \wedge T_{(n)})|)^2 + (P\{T_1 < \tau_2\})^n. \end{aligned} \tag{2.10}$$

In view of (2.2), (2.3) and a martingale inequality (Métivier, 1982, p. 60), the first term on the right of (2.10), denoted by I_1, satisfies

$$I_1 \leq 2c_2^{-2} E(U_n(\tau_2 \wedge T_{(n)}))^2 = 2c_2^{-2} E \int_0^{\tau_2 \wedge T_{(n)}} \left\{ \sum_{i=1}^{n} H_i(s) \right\}^{-1} \lambda(s) ds. \tag{2.11}$$

Since $\int_0^{\tau_2 \wedge T_{(n)}} n \{ \sum_{i=1}^n H_i(s)\}^{-1} \lambda(s) ds$ converges to a constant almost surely, (2.11) converges to zero as n tends to ∞. This implies that I tends to zero as n tends to ∞. Next,

$$\begin{aligned} II &\leq P\{|U_n(T)| > c_3, T \leq T_{(n)}\} + P\{T_{(n)} < T\} \\ &\leq c_3^{-2} E(U_n(T \wedge T_{(n)}))^2 + (P\{T_1 < T\})^n \\ &= c_3^{-2} E \int_0^{T \wedge T_{(n)}} \left\{ \sum_{i=1}^{n} H_i(s) \right\}^{-1} \lambda(s) ds + (P\{T_1 < T\})^n. \end{aligned}$$

Then the convergence of II to zero is clearly true as for I. Thus, $\hat{\tau}_n$ is consistent for τ. This completes the proof. ∎

If $\theta < 0$, then $Y(t)$ in (2.5) is decreasing on $[0, \tau]$ and increasing on $[\tau, T]$. Thus, a consistent estimator for τ, in this case, can be obtained by defining

$$\hat{\tau}_n = \inf\{t \in [\tau_1, \tau_2] : Y_n(t\pm) = \inf_{\tau_1 \leq u \leq \tau_2} Y_n(u)\}.$$

3. Consistent Estimators for the Hazard Rates. Following the estimation of the change-point, we can estimate the two hazard rates by considering the score functions. Since the score functions with respect to β and θ under the true parameters have zero expectations and $\hat{\tau}_n$ is a consistent estimator of τ, it suggests that the solutions $\hat{\beta}_n(\hat{\tau}_n)$ and $\hat{\theta}_n(\hat{\tau}_n)$ of $\partial\ell_n(\beta,\theta,\hat{\tau}_n)/\partial\beta = 0$ and $\partial\ell_n(\beta,\theta,\hat{\tau}_n)/\partial\theta = 0$, where ℓ_n is the log likelihood, be reasonable estimators of β and θ respectively. Note that $\partial\ell_n(\beta,\theta,\tau)/\partial\beta = \beta^{-1}\sum_{i=1}^n N_i(\tau) + (\beta+\theta)^{-1}\sum_{i=1}^n \{N_i(\infty) - N_i(\tau)\} - \sum_{i=1}^n T_i = 0$, and $\partial\ell_n(\beta,\theta,\tau)/\partial\theta = (\beta + \theta)^{-1}\sum_{i=1}^n \{N_i(\infty) - N_i(\tau)\} - \sum_{i=1}^n (T_i - \tau)^+ = 0$. Thus, we propose the estimators

$$\hat{\beta}_n(\hat{\tau}_n) = \sum_{i=1}^n N_i(\hat{\tau}_n) / \sum_{i=1}^n (\hat{\tau}_n \wedge T_i), \tag{3.1}$$

$$\hat{\theta}_n(\hat{\tau}_n) = \sum_{i=1}^n \{N_i(\infty) - N_i(\hat{\tau}_n)\} / \sum_{i=1}^n (T_i - \hat{\tau}_n)^+ - \hat{\beta}_n(\hat{\tau}_n). \tag{3.2}$$

THEOREM 3.1. *For $\hat{\beta}_n(\hat{\tau}_n), \hat{\theta}_n(\hat{\tau}_n)$ as above and $\hat{\tau}_n$ as in (2.7), $\hat{\beta}_n(\hat{\tau}_n)$ converges to β and $\hat{\theta}_n(\hat{\tau}_n)$ converges to θ in probability respectively.*

PROOF. Standard arguments show that as n tends to ∞, $\hat{\beta}_n(\tau)$ converges to $P\{X_1 \leq (\tau \wedge C_1)\}/E(\tau \wedge T_1) = \beta$ and $\{\hat{\beta}_n(\tau) + \hat{\theta}_n(\tau)\}$ converges to $[P\{X_1 \leq C_1\} - P\{X_1 \leq (\tau \wedge C_1)\}] / E(T_1 - \tau)^+ = \beta + \theta$ in probability. Finally the consistency is obtained by making use of Theorem 2.1 and continuity via the inequality, for $\delta > 0$,

$$\begin{aligned} &P\Big\{|\sum_{i=1}^n (\hat{\tau}_n \wedge T_i)/n - E(\tau \wedge T_1)| > \delta\Big\} \\ &\leq P\Big\{|\sum_{i=1}^n (\hat{\tau}_n \wedge T_i)/n - E(\tau \wedge T_1)| > \delta, |\hat{\tau}_n - \tau| < \epsilon\Big\} + P\{|\hat{\tau}_n - \tau| > \epsilon\}. \end{aligned} \tag{3.3}$$

4. Asymptotic Distribution of the Change-Point Estimator. The purpose of this section is to obtain the limiting distribution of the change-point estimator $\hat{\tau}_n$. For this, we will first apply the Poisson approximation of empirical process to get a weak convergence theorem for the local process $Q_n(x) \equiv n(Y_n(\tau + \frac{x}{n}) - Y_n(\tau))$, defined in terms of (2.4), and then apply continuous mapping theorem to it to obtain the limiting distribution of $\hat{\tau}_n$.

In order to facilitate the discussion of weak convergences, we will introduce some function spaces. Let $D_0[0,a]$ be the standard D space of real-valued functions on $[0,a]$ equipped with Skorohod topology. (cf. Billingsley (1968)). Let $D_1[-a,a]$ be the space of real-valued functions f defined on $[-a,a]$ such that (i) both $f|_{[0,a]}(\cdot)$ and $g(\cdot)$ are in $D_0[0,a]$, where $g(x) = f|_{[-a,0]}(-x)$, and (ii) $f(0) = 0$. We will give $D_1[-a,a]$ the natural topology adapted from $D_0[0,a]$. Let $D_2[-a,a]$ be the subspace of $D_1[-a,a]$ consisting of piecewise increasing functions. Let $D_3[-a,a]$ be the subspace of $D_2[-a,a]$ consisting of functions being piecewise constant with only jumps of size one. We note that both $D_2[-a,a]$ and $D_3[-a,a]$ are closed in $D_1[-a,a]$.

Let $M_+[-a,a]$ denote the set of non-negative Radon measures on $[-a,a]$, which can be made into a complete, separable metric space. We note that $D_3[-a,a]$ can be viewed as a subset of $M_+[-a,a]$. (cf. Resnick (1987), p. 147). It can be shown that the relative topologies on $D_3[-a,a]$ induced by $M_+[-a,a]$ and $D_1[-a,a]$ are identical.

We are now in a position to present the weak convergence more efficiently.

The following is a weak convergence of local process for the empirical process with censored data. The case with no censoring variable was obtained by Al-Husaini and Elliott (1984) using martingale theory. Our approach seems more elementary and shorter. Let $X_1, X_2, \cdots$ be independent random variables with common distribution function $F(x)$ and density function $f(x)$, which needs not satisfy (1.1). Let $C_1, C_2, \cdots$ be independent random variables with common distribution function $G(x)$ and density function $g(x)$. Assume the C's and X's are independent. Let $\bar{N}_n(t) = \sum_{i=1}^{n} 1_{[X_i,\infty)}(t \wedge C_i)$. For a fixed τ, we define $Z_n(x) = \bar{N}_n(\tau + \frac{x}{n}) - \bar{N}_n(\tau)$. Then for each $a > 0$, Z_n is a random element of $D_3[-a,a]$.

LEMMA 4.1. *Z_n converges weakly to a process Z on $D_3[-a,a]$, where $\{Z(x)|x \in [0,a]\}$ is a Poisson process with intensity $f(\tau+)(1-G(\tau))$, $\{-Z(-x)| x \in [0,a]\}$ is a Poisson process with intensity $f(\tau-)(1-G(\tau))$, and $\{Z(x)|x \in [0,a]\}$ and $\{Z(x)|x \in [-a,0]\}$ are independent.*

PROOF. This lemma is proved by making use of the following weak convergence criterion (4.1) for point processes, which reduces the problem to the weak convergence of finite dimensional distributions of Z_n. We shall first compute the finite dimensional distribution of $Z_n(x), -a \le x \le a$. For $-a \le x_2 < x_1 \le 0, 0 \le y_1 < y_2 \le a, k_1, k_2, l_1, l_2$ being non-negative integers,

$$
\begin{aligned}
&P\{Z_n(x_1) = -k_1, Z_n(x_2) - Z_n(x_1) = -k_2, Z_n(y_1) = l_1, Z_n(y_2) - Z_n(y_1) = l_2\} \\
&= P\Big\{ \sum_{i=1}^{n} 1_{[\tau+\frac{x_1}{n}<X_i\le\tau, X_i\le C_i]} = k_1, \sum_{i=1}^{n} 1_{[\tau+\frac{x_2}{n}<X_i\le\tau+\frac{x_1}{n}, X_i\le C_i]} = k_2, \\
&\quad \sum_{i=1}^{n} 1_{[\tau<X_i\le\tau+\frac{y_1}{n}, X_i\le C_i]} = l_1, \sum_{i=1}^{n} 1_{[\tau+\frac{y_1}{n}<X_i\le\tau+\frac{y_2}{n}, X_i\le C_i]} = l_2 \Big\} \\
&= \frac{n!}{k_1!k_2!l_1!l_2!(n-k_1-k_2-l_1-l_2)!} \big(F(\tau) - F(\tau+\frac{x_1}{n})\big)^{k_1} \big(1 - G(\tau) + o(\frac{1}{n})\big)^{k_1} \cdot \\
&\big(F(\tau+\frac{x_1}{n}) - F(\tau+\frac{x_2}{n})\big)^{k_2} \big(1 - G(\tau) + o(\frac{1}{n})\big)^{k_2} \big(F(\tau+\frac{y_1}{n}) - F(\tau)\big)^{l_1} \cdot \\
&\big(1 - G(\tau) + o(\frac{1}{n})\big)^{l_1} \big(F(\tau+\frac{y_2}{n}) - F(\tau+\frac{y_1}{n})\big)^{l_2} \big(1 - G(\tau) + o(\frac{1}{n})\big)^{l_2} \cdot \\
&\Big[1 - \big(F(\tau) - F(\tau+\frac{x_2}{n})\big)\big(1 - G(\tau) + o(\frac{1}{n})\big) \\
&\quad - \big(F(\tau+\frac{y_2}{n}) - F(\tau)\big)\big(1 - G(\tau) + o(\frac{1}{n})\big)\Big]^{n-k_1-k_2-l_1-l_2},
\end{aligned}
$$

which, as n goes to infinity, converges to

$$
\begin{aligned}
&\frac{1}{k_1!k_2!} \big(f(\tau-)(-x_1)\big)^{k_1} \big(1 - G(\tau)\big)^{k_1} e^{-f(\tau-)(-x_1)(1-G(\tau))} \cdot \\
&\big(f(\tau-)(x_1 - x_2)\big)^{k_2} \big(1 - G(\tau)\big)^{k_2} e^{-f(\tau-)(x_1-x_2)(1-G(\tau))} \cdot \\
&\frac{1}{l_1!l_2!} \big(f(\tau+)y_1\big)^{l_1} \big(1 - G(\tau)\big)^{l_1} e^{-f(\tau+)y_1(1-G(\tau))} \cdot \\
&\big(f(\tau+)(y_2 - y_1)\big)^{l_2} \big(1 - G(\tau)\big)^{l_2} e^{-f(\tau+)(y_2-y_1)(1-G(\tau))}.
\end{aligned}
$$

In general, the finite-dimensional distributions of $\{Z_n(x), -a \le x \le a\}$ converges to those of $\{Z(x), -a \le x \le a\}$.

In order to show that Z_n converges weakly to Z on $D_3[-a,a]$, it suffices to show that for all $h \in C^+([-a,a])$,

$$
\int_{-a}^{a} h(x)dZ_n(x) \text{ converges weakly to } \int_{-a}^{a} h(x)dZ(x). \tag{4.1}
$$

(See e.g. Exercise 3.5.1 of Resnick (1987)). Here $C^+([-a,a])$ is the set of continuous, non-negative functions on $[-a,a]$.

For $h = \sum_{i=1}^{l} c_i 1_{(a_i,b_i]}$, $c_i \ge 0$ and $\{(a_i,b_i], i = 1,\cdots,l\}$ being disjoint intervals in $[-a,a]$, (4.1) holds due to the finite-dimensional convergence of $Z_n(\cdot)$. Now let $h \in C^+([-a,a])$. For $\epsilon > 0$, there exist simple functions $h_\epsilon, \tilde{h}_\epsilon$ of the previous form such that $0 \le h_\epsilon \le h \le \tilde{h}_\epsilon$ and $|h_\epsilon(x) - \tilde{h}_\epsilon(x)| < \epsilon$ for every $x \in [-a, a]$. Applying standard approximation arguments to h_ϵ, h, and $\tilde{h}_\epsilon$, one can show that $\int_{-a}^{a} h(x)dZ_n(x)$ converges to $\int_{-a}^{a} h(x)dZ(x)$ weakly. The proof is thus complete. ∎

In order to get the weak convergence of the local process $n(Y_n(\tau + \frac{x}{n}) - Y_n(\tau))$, we shall first get the weak convergence of the local deviation process $n(A_n(\tau + \frac{x}{n}) - A_n(\tau))$ on $D_2[-a, a]$, where τ is the change-point.

LEMMA 4.2. *$n\{A_n(\tau + \frac{x}{n}) - A_n(\tau)\}$ converges weakly on $D_2[-a, a]$ to $P^{-1}\{X \wedge C > \tau\}Z(x)$, where $Z(x)$ is defined in Lemma 4.1 with $f(\tau+) = (\beta + \theta)e^{-\beta\tau}$, $f(\tau-) = \beta e^{-\beta\tau}$.*

PROOF. $n\{A_n(\tau + \frac{x}{n}) - A_n(\tau)\}$

$$
\begin{aligned}
&= P^{-1}\{X \wedge C > \tau\}\left(\bar{N}_n(\tau + \frac{x}{n}) - \bar{N}_n(\tau)\right) \\
&+ \int_{\tau}^{\tau + \frac{x}{n}} \left(\left\{\frac{\sum_{i=1}^{n} 1_{[0, X_i \wedge C_i]}(s)}{n}\right\}^{-1} - P^{-1}\{X \wedge C > \tau\}\right) d\bar{N}_n(s).
\end{aligned}
\tag{4.2}
$$

Let $I_n(x), II_n(x)$ denote respectively the first and the second term in (4.2). Observe

$$
\begin{aligned}
\sup_{-a \le x \le a} |II_n(x)| \le &\sup_{\tau - \frac{a}{n} \le s \le \tau + \frac{a}{n}} \left| \left\{\frac{\sum_{i=1}^{n} 1_{(0, X_i \wedge C_i]}(s)}{n}\right\}^{-1} \right. \\
&\left. - P^{-1}\{X \wedge C > \tau\} \right| \left(\bar{N}_n(\tau + \frac{a}{n}) - \bar{N}_n(\tau - \frac{a}{n})\right) \\
&\le \max\left(\left\{\frac{\sum_{i=1}^{n} 1_{[\tau - \frac{a}{n} \le X_i \wedge C_i]}}{n}\right\}^{-1} - P^{-1}\{X \wedge C > \tau\}, \right. \\
&\left. \left\{\frac{\sum_{i=1}^{n} 1_{[\tau + \frac{a}{n} \le X_i \wedge C_i]}}{n}\right\}^{-1} - P^{-1}\{X \wedge C > \tau\} \right) \left(\bar{N}_n(\tau + \frac{a}{n}) - \bar{N}_n(\tau - \frac{a}{n})\right).
\end{aligned}
\tag{4.3}
$$

Since the first factor in (4.3) goes to zero almost surely and, by Lemma 4.1, the second factor has a limiting distribution, $\sup_{-a \le x \le a} |II_n(x)|$ converges to zero in probability.

It follows from Lemma 4.1, (4.2) and (4.3) that $n\{A_n(\tau + \frac{x}{n}) - A_n(\tau)\}$ converges weakly to $P^{-1}\{X \wedge C > \tau\}Z(x)$. The proof is complete. ∎

Let $Q_n(x) \equiv n\{Y_n(\tau + \frac{x}{n}) - Y_n(\tau)\}, Q(x) \equiv xv - wZ(x)$ with $v = A(T)\tau^{p-1}(T - \tau)^{p-2}(pT - 2p\tau + \tau) - TA(\tau)\tau^{p-2}(T - \tau)^{p-2}(T - 2\tau)(p - 1), w = T\tau^{p-1}(T - \tau)^{p-1}P^{-1}\{X \wedge C > \tau\}$, and $Z(x)$ being defined in Lemma 4.2. Assume $p \ge \frac{1}{2}$, then $Q_n|_{[-a,a]}$ is a random element of $D_2[-a, a]$.

THEOREM 4.1. *Q_n converges weakly on $D_2[-a, a]$ to Q.*

PROOF. Using the fact that $Y_n(t) = A_n(T)t^p(T-t)^{p-1} - A_n(t)Tt^{p-1}(T-t)^{p-1}$, we know

$$\begin{aligned}
&n\{Y_n(\tau+\frac{x}{n}) - Y_n(\tau)\}\\
&= nA_n(T)\big((\tau+\frac{x}{n})^p(T-\tau-\frac{x}{n})^{p-1} - \tau^p(T-\tau)^{p-1}\big)\\
&- TnA_n(\tau)\big((\tau+\frac{x}{n})^{p-1}(T-\tau-\frac{x}{n})^{p-1} - \tau^{p-1}(T-\tau)^{p-1}\big)\\
&- Tn\big(A_n(\tau+\frac{x}{n}) - A_n(\tau)\big)(\tau+\frac{x}{n})^{p-1}(T-\tau-\frac{x}{n})^{p-1}.
\end{aligned}$$

The weak convergence of $Q_n(x)$ is then obtained by Lemma 4.2 and the convergence in probability of $A_n(t)$. This completes the proof. ∎

Using the arguments of Lindvall (1973), one can extend Theorem 4.1 to get Q_n converges weakly to Q on $D_2(-\infty,\infty)$. Here the domain of definition of an element f in $D_2[-a,a]$ can be extended to $D_2(-\infty,\infty)$ by defining $f(x) = f(a)$ for $x \geq a$ and $f(x) = f(-a)$ for $x \leq -a$. Asymptotic distribution of the change-point estimator $\hat{\tau_n}$ can now be obtained as a consequence of the continuous mapping theorem for weak convergence as follows.

LEMMA 4.3. *Assume $v > 0$. Then with probability one the limit process $Q(x)$ in Theorem 4.1 has a unique supremum point d^*.*

PROOF. Since the supremum of $Q(x)$ must occur at $B_j\pm$, where B_j are jump points of the process $Z(x)$, it suffices to show that for every $j \neq k$, $P\{Q(B_j\pm) = Q(B_k\pm)\} = 0$. This is true because (B_j, B_k) has joint density, and the proof is complete. ∎

For $f \in D_2(-\infty,\infty)$, let $h(f)$ be the smallest number in $\mathcal{R}$ satisfying $f(x) \leq f(h(f)\pm)$ for every x. Lemma 4.3 implies that $h(Q)$ is a well-defined random variable. We note that $\lim_{x\to\pm\infty} Q(x) = -\infty$.

Assume $p \geq \frac{1}{2}$. In this case, $Q_n(x)$ is increasing on each interval $[n(T_{n,i-1}-\tau), n(T_{ni}-\tau))$, which says that $Q_n(\cdot) \in D_2(-\infty,\infty)$. It can be shown by the arguments in Lemma 4.3 that $P[Q \in \{f \in D_2(-\infty,\infty) : h \text{ is discontinuous at } f\}] = 0$. This together with Theorem 4.1 and the continuous mapping theorem implies $h(Q_n)$ converges weakly to $h(Q) = d^*$. Since $P\{n(\tau_1-\tau) \leq h(Q_n) \leq n(\tau_2-\tau)\} \to 1$, as n goes to infinity, we have

COROLLARY 4.1. *Assume that $v > 0$ and $p \geq \frac{1}{2}$, $n(\hat{\tau_n}-\tau)$ converges weakly to d^*, where d^* is the supremum point of $Q(x)$, defined in Lemma 4.3.* ∎

5. Comparisons.

5.1. A Constrained Maximum Likelihood Estimator of τ. In the following we investigate the relative performance of the proposed estimators to a constrained maximum likelihood estimators (CMLE). Let $T_{(i)}$, $1 \le i \le n$, be the order statistics of the uncensored and the censored observations. Then, we may extend the constrained maximum likelihood estimator in Yao(1986) to the case of random censoring by imposing a constraint that $\tau \le T_{(n-1)}$, if $T_{(n)}$ is uncensored, which removes the singularity of the likelihood. In this case,

$$\begin{aligned}
\ell_n(\hat{\beta}_n(\tau), \hat{\theta}_n(\tau), \tau) + S_n &= S_n \log\left[S_n\left\{\sum_{i=1}^n (T_i - \tau)\right\}^{-1}\right], \text{if } \tau < T_{n1}; \text{ or} \\
&= R_n \log\left[R_n\left\{\sum_{i=1}^n (T_i \wedge \tau)\right\}^{-1}\right] \\
&\quad + (S_n - R_n)\log\left[(S_n - R_n)\left\{\sum_{i=1}^n (T_i - \tau)^+\right\}^{-1}\right], \text{ if } \tau \ge T_{n1}; \\
&= S_n \log\left[S_n\left\{\sum_{i=1}^n (T_i \wedge \tau)\right\}^{-1}\right], \text{if } T_{(n-1)} \le \tau < T_{(n)} \text{ and } T_{(n)} \text{ is censored},
\end{aligned}$$

where $\hat{\beta}_n(\tau)$, and $\hat{\theta}_n(\tau)$ are given in (3.1) and (3.2),$S_n = \sum_{i=1}^n N_i(\infty)$ and $R_n = \sum_{i=1}^n N_i(\tau)$. It follows that $\ell_n(\hat{\beta}_n(\tau), \hat{\theta}_n(\tau), \tau)$ is strictly convex in $[T_{(i-1)}, T_{(i)})$ for $2 \le i \le n-1$. Consequently, the maximizer of $\ell_n(\hat{\beta}_n(\tau), \hat{\theta}_n(\tau), \tau)$, denoted by $\tilde{\tau}$, lies either just before or just after a $T_{(i)}$. Thus, $(\tilde{\tau}, \tilde{\beta}, \tilde{\theta})$= $(\tilde{\tau}, \hat{\beta}_n(\tilde{\tau}), \hat{\theta}_n(\tilde{\tau}))$ are the constrained maximum likelihood estimators for (τ, β, θ).

5.2. Simulation Procedures. In the simulation study, we generated 1000 random samples from each of the following two change-point hazard rate models: (a) $\tau = 1, \beta = 1, \theta = 1, \tau_1 = 0.75$ and $\tau_2 = 1.15$, (b) $\tau = 1, \beta = 0.25, \theta = 1.25, \tau_1 = 0.5$ and $\tau_2 = 1.25$. Uniform censoring times were generated in the interval $(0, U)$ with U selected to give expected censoring proportions of 20% and 40%, respectively. We used IMSL subroutine DRNUN to generate pseudorandom numbers. For each censoring proportion, sample sizes of $n = 50$ and 100 were generated.

The value of T for calculating $Y_n(t)$ in (2.5) was set to satisfy $P\{T_i > T\} \approx 0.01$. The proposed estimator $\hat{\tau}$ was computed with p=0, 0.25, 0.5, 0.75, and 1, respectively, in $g(x) = x^p$. For p=0 and 0.25, $\hat{\tau}$ was computed using the IMSL optimization subroutine DBCPOL based on a direct search complex algorithm. All the computations of $(\hat{\tau}, \hat{\beta}, \hat{\theta})$ and $(\tilde{\tau}, \tilde{\beta}, \tilde{\theta})$ are subject to specified bounds τ_1 and τ_2.

The averages (AVE) and the square-roots of mean square errors (RMSE) of the estimators $(\hat{\tau}, \hat{\beta}, \hat{\theta})$ and $(\tilde{\tau}, \tilde{\beta}, \tilde{\theta})$, respectively, in 1000 replications were computed.

5.3. Results. The simulation results pertaining to the evaluation of $(\hat{\tau}(p), \hat{\beta}(p), \hat{\theta}(p))$, only for p=0.5 and 1, and $(\tilde{\tau}, \tilde{\beta}, \tilde{\theta})$ are presented in Table 1. The results for p=0.75, that are similar to those for p=0.5 or 1, are omitted. Since the root mean square errors of $\hat{\tau}(p)$ with $p < 0.5$ are sometimes double those of $\hat{\tau}(0.5)$, $(\hat{\tau}(p), \hat{\beta}(p), \hat{\theta}(p))$ with $p < 0.5$ are not recommended in practice.

Table 1

Parameter estimates and their square-root mean square errors with 1000 replication

(a) Model $\tau = 1$, $\beta = 1$, $u_1 = 0.75$ and $\tau_2 = 1.15$

		Parameters	τ		β		θ	
n	Censoring proportion	Estimators	AVE	RMSE	AVE	RMSE	AVE	RMSE
50	20%	$p = 0.5$	0.981	0.108	0.955	0.195	1.415	0.860
	$(U = 4.1,$	$p = 1$	0.990	0.102	0.959	0.193	1.428	0.874
	$T = 2.4)$	CMLE	0.973	0.110	0.977	0.200	1.275	0.791
50	40%	$p = 0.5$	0.969	0.118	0.978	0.203	1.715	1.610
	$(U = 2.0,$	$p = 1$	0.966	0.117	0.978	0.204	1.691	1.561
	$T = 1.8)$	CMLE	0.971	0.121	0.997	0.216	1.600	2.156
100	20%	$p = 0.5$	0.999	0.096	0.982	0.132	1.250	0.556
	$(U = 4.1,$	$p = 1$	1.010	0.091	0.985	0.130	1.262	0.569
	$T = 2.4)$	CMLE	0.987	0.096	0.989	0.133	1.175	0.513
100	40%	$p = 0.5$	0.997	0.101	0.974	0.141	1.507	1.028
	$(U = 2.0,$	$p = 1$	0.993	0.100	0.973	0.141	1.493	1.003
	$T = 1.8)$	CMLE	0.987	0.106	0.977	0.145	1.438	1.044

(b) Model $\tau = 1$, $\beta = 0.25$, $\theta = 1.25$, $\tau_1 = 0.5$ and $\tau_2 = 1.25$

		Parameters	τ		β		θ	
n	Censoring proportion	Estimators	AVE	RMSE	AVE	RMSE	AVE	RMSE
50	20%	$p = 0.5$	1.031	0.077	0.243	0.079	1.400	0.351
	($U = 7.0$,	$p = 1$	1.055	0.093	0.255	0.080	1.417	0.371
	$T = 3.5$)	CMLE	1.009	0.078	0.251	0.083	1.328	0.321
50	40%	$p = 0.5$	1.053	0.098	0.253	0.083	1.480	0.522
	($U = 3.5$,	$p = 1$	1.048	0.089	0.251	0.081	1.473	0.512
	$T = 2.0$)	CMLE	1.009	0.098	0.244	0.084	1.385	0.438
100	20%	$p = 0.5$	1.026	0.055	0.255	0.057	1.320	0.229
	($U = 7.0$,	$p = 1$	1.054	0.086	0.271	0.066	1.328	0.238
	$T = 3.5$)	CMLE	1.009	0.041	0.256	0.056	1.281	0.214
100	40%	$p = 0.5$	1.036	0.071	0.256	0.062	1.359	0.308
	($U = 3.5$,	$p = 1$	1.030	0.061	0.253	0.061	1.353	0.304
	$T = 2.0$)	CMLE	1.008	0.054	0.249	0.061	1.309	0.276

With both sample sizes $n = 50$ and 100, the biases and the mean square errors of $(\hat{\tau}, \hat{\beta}, \hat{\theta})$ with $p \geq 0.5$ are of the same magnitudes as $(\tilde{\tau}, \tilde{\beta}, \tilde{\theta})$. Moreover, the performance of p=0.5 and 1 is not distinguishable. The biases and the mean square errors of $(\hat{\tau}, \hat{\beta})$ and $(\tilde{\tau}, \tilde{\beta})$ are satisfactory even in the case of 40% moderate censoring. Those of $\hat{\theta}$ and $\tilde{\theta}$ are noticeable even in the case of 20% light censoring, but have improvements when $n = 500$. The mean square errors of $\hat{\theta}$ and $\tilde{\theta}$ greatly increase in heavier censoring case, but not those of the estimators for τ and β.

6. Discussion. We extend the model in Nguyen, Rogers and Walker (1984) to randomly censored data. Their estimator for τ is of rather complicated form based on certain moment properties of a mixture density. Our proposed estimator $\hat{\tau}$ is much easier to implement and can be viewed as a nonparametric counterpart of the estimator resulting from the score process in Matthews, Farewell and Pyke (1985). In fact, the former focuses on the difference between two slopes before and after an uncensored time on the cumulative hazard plot, while the latter process can be interpreted as the difference of parametric estimators of the hazard rates multiplied by a scaling factor.

In comparison with the likelihood-based estimator, the proposed estimator with $p \geq 0.5$ behaves almost as well as the constrained maximum likelihood estimator with respect to negligible differences in their biases and mean square errors. The restricted range $[\tau_1, \tau_2]$ of τ chosen in the proposed procedure is actually needed in all available methods. This restriction is inherent in the consistency of all such estimators for τ. In our simulation study, the estimators $(\hat{\tau}, \hat{\beta}, \hat{\theta})$ and $(\tilde{\tau}, \tilde{\beta}, \tilde{\theta})$ without the restriction of $[\tau_1, \tau_2]$ were also computed. The corresponding biases and mean square errors are all eminently large.

Furthermore, all the results regarding $\hat{\tau_n}$ in this paper are valid when $\lambda(x)$ in (1.1) is replaced by $\tilde{\lambda}(x) = \lambda(x) + \epsilon h(x)$ for some continuous functions $h(x)$ and $\epsilon > 0$, or any hazard rate model which satisfies the property that the maximum point of $Y(x)$ is at the change-point τ. Simulations given in Table 2 also indicate that the proposed change-point estimator of this paper is more favorable than the constrained maximum likelihood estimator under some perturbated change-point models. In summary, under the two-step model (1.1) neither the proposed estimator nor the CMLE is dominant, while the proposed estimator seems more favorable in a larger class of models. This may be an advantage of the present approach.

Table 2

Estimates for the change-point τ under deviated models and their square-root mean square errors with 1000 replications

(a) Model $\tilde{\lambda}(x) = \lambda(x) + \epsilon(x)$ $\tau = 1$, $\beta = 1$, $\theta = 1$, $\tau_1 = 0.75$ and $\tau_2 = 1.15$

		ϵ	1.0		0.5		0.1	
n	Censoring proportion	Estimators	AVE	RMSE	AVE	RMSE	AVE	RMSE
50	20%	$p = 0.5$	0.961	0.124	0.969	0.118	0.973	0.115
		$p = 1$	0.951	0.124	0.969	0.117	0.981	0.111
		CMLE	0.936	0.133	0.955	0.124	0.967	0.117
50	40%	$p = 0.5$	0.947	0.132	0.954	0.126	0.971	0.118
		$p = 1$	0.926	0.138	0.943	0.128	0.964	0.117
		CMLE	0.940	0.133	0.956	0.124	0.975	0.117
100	20%	$p = 0.5$	0.982	0.114	0.989	0.110	0.996	0.100
		$p = 1$	0.968	0.115	0.989	0.105	1.005	0.096
		CMLE	0.934	0.134	0.956	0.122	0.984	0.103
100	40%	$p = 0.5$	0.967	0.125	0.975	0.117	0.994	0.109
		$p = 1$	0.935	0.134	0.954	0.120	0.984	0.107
		CMLE	0.937	0.135	0.957	0.123	0.980	0.113

(b) Model $\tilde{\lambda}(x) = \lambda(x) + \epsilon x$, $\tau = 1$, $\beta = 0.25$, $\theta = 1.25$, $\tau_1 = 0.5$ and $\tau_2 = 1.25$

		ϵ	1.0		0.5		0.1	
n	Censoring proportion	Estimators	AVE	RMSE	AVE	RMSE	AVE	RMSE
		$p = 0.5$	0.994	0.153	1.019	0.117	1.037	0.091
50	20%	$p = 1$	1.011	0.125	1.046	0.107	1.065	0.103
		CMLE	0.887	0.216	0.955	0.148	1.008	0.092
		$p = 0.5$	0.996	0.170	1.019	0.136	1.040	0.097
50	40%	$p = 1$	0.988	0.152	1.029	0.121	1.059	0.102
		CMLE	0.868	0.238	0.937	0.175	1.006	0.104
		$p = 0.5$	1.029	0.119	1.035	0.091	1.032	0.068
100	20%	$p = 1$	1.033	0.101	1.054	0.096	1.062	0.098
		CMLE	0.891	0.196	0.970	0.105	1.007	0.048
		$p = 0.5$	1.021	0.149	1.036	0.112	1.035	0.076
100	40%	$p = 1$	1.003	0.124	1.038	0.098	1.052	0.089
		CMLE	0.863	0.228	0.958	0.131	1.005	0.061

The simulation program and the exact values of T and U used for each model can be provided for interested readers.

Acknowledgements. The authors thank the referee for some helpful comments and acknowledge the support provided by grants from the National Science Council, R.O.C. awarded to each of the authors.

REFERENCES

AALEN, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6**, 701–726.

AKMAN, V.E. AND RAFTERY, A.E. (1986). Asymptotic inference for a change-point Poisson process. *Ann. Statist.* **14**, 1583–1590.

AL-HUSSAINI, A. AND ELLIOTT, R. J. (1984). Convergence of the empirical distribution to the Poisson process. *Stochastics*, **13**, 299–308.

BILLINGSLEY, P. (1968). Convergence of Probability Measures. John Wiley, New York.

HENDERSON, R. (1990). A problem with the likelihood ratio test for a change-point hazard rate model. *Biometrika* **77**, 835–843.

LINDVALL, T. (1973). Weak convergence of probability measures and random functions in the function space $D[0, \infty)$. *J. Appl. Prob.* **10**, 109–121.

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

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., 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.

MÉTIVIER, M. (1982).] *Semimartingale: A Course on Stochastic Processes.* Berlin: De Gruyter.

MÜLLER, 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.

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

RESNICK, S. I. (1987). Extreme values, regular variation, and point processes. *Springer-Verlag.*

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

DEPARTMENT OF MATHEMATICS
NATIONAL CENTRAL UNIVERSITY
CHUNG-LI, TAIWAN, ROC

INSTITUTE OF STATISTICAL SCIENCE
ACADEMIC SINICA
TAIPEI, TAIWAN, ROC