Journal of Mathematical Research with Applications
May, 2020, Vol. 40, No. 3, pp. 320–330
DOI:10.3770/j.issn:2095-2651.2020.03.009
Http://jmre.dlut.edu.cn

Pseudo-Maximum Likelihood Estimation in the Hazards Cure Model with a Single Change Point for Current Status Data

Bing WANG, Xiaoguang WANG*
Department of Mathematics, Dalian University of Technology, Liaoning 116024*, P. R. China*

Abstract The change-point hazards model has received much attention, since it can not only display the impacts of treatments or medical breakthroughs more directly, but also provide the time point when those impacts occur. In this paper, we propose the single change-point hazards model for current status survival data with long-term survivors and investigate the estimation for the proposed model. Large-sample properties of the estimators are established. Simulation studies are carried out to evaluate the finite-sample performance of the estimation.

Keywords current status data; change-point hazard model; cure fraction; pseudo-maximum likelihood

MR(2010) Subject Classification 62N01; 62N02; 62N03

1. Introduction

In clinical trials, it is a common situation that the treatment effect is not immediately reflected but takes a while or will disappear after a time lag. Detecting this time lag is very important. The hazard model with a change point in time is a widespread research tool to deal with this situation and is given by:

$$\lambda(t) = \beta + \theta I(t > \tau), \tag{1.1}$$

where the initial hazard rate β and the change point τ are positive constants, the jump θ can be either positive or negative, which reflects a higher or lower hazard rate, and $\beta + \theta > 0$ is needed. Model (1.1) corresponds to a piecewise exponential distribution where the jump occurs at τ.

This piecewise constant hazards model was first proposed by [1] to discuss whether the new therapy is effective. The existence of the change point and $\theta < 0$ indicate that the new therapy works. Loader [2] discussed the inference about the likelihood ratio test statistic for the existence of the change point. The estimation of the change point for right censored data was investigated by [3]. Zhao et al. [4] extended it to the situation where there exist long-term survivors. The change-point estimation for current status data still needs discussion. Othus et al. [5] studied the cure model where the latency part is similar to (1.1), however they focused on the change-point effects in covariates.

Received June 6, 2019; Accepted October 10, 2019
Supported by the National Natural Sciences Foundation of China (Grant No. 11471065).
* Corresponding author
E-mail address: wangxg@dlut.edu.cn (Xiaoguang WANG)

The observation in current status data consists of a monitor time and the knowledge of whether the failure occurs before the monitor time. Current status data would appear in cross-sectional studies when the event of interest is a mile-stone event, such as onset of chronic disease or occurrence of first pregnancy. The exact time-to-event data might be collected, but it would cost too much for precision. In this case, current status data are still preferable [6].

In this article, we focus on modeling the current status survival data with long term survivors. Compared with [4], we deal with different types of data. We apply the Iterative Convex Minorant (ICM) method to obtain the nonparametric maximum likelihood estimator (NPMLE) instead of the Kaplan-Meier method. The NPMLE calculated by the ICM algorithm can guarantee the consistency which is helpful to achieve the consistency of the uncured rate's estimator.

The rest of the article is organized as follows. In Section 2, we outline the notations and model descriptions of the single change-point hazard model with a cure fraction for current status data. Details of pseudo-maximum likelihood estimation are presented. Large-sample properties are investigated in Section 3. Simulation results are reported in Section 4.

2. Model and estimation

Let T be the failure time of interest, and C be the random monitor time. For current status data, the observation $\boldsymbol{O}$ consists of (C, δ) where $\delta = I(T \leq C)$ is an indicator function. To describe the long-term survivor, we introduce the cure indicator η: $\eta = 0$ if the subject is cured, and $\eta = 1$ otherwise. Define $p \equiv P(\eta = 1)$ to be the probability of being uncured. Denote by T^* the failure time of uncured patient. Then we can obtain that

$$F(t) = P(T \leq t|\eta = 1)P(\eta = 1) + P(T \leq t|\eta = 0)P(\eta = 0) = pP(T^* \leq t) = pF^*(t),$$

where $F(t)$ and $F^*(t)$ are the cumulative distribution functions (c.d.f.) of T and T^*, respectively. And the hazard function of T has the form

$$\lambda(t) = \frac{f(t)}{1 - F(t)} = \frac{pf^*(t)}{1 - pF^*(t)}, \tag{2.1}$$

where $f(t)$ and $f^*(t)$ are density functions of T and T^*, respectively.

Assume that the hazard function of T^* is specified as

$$\lambda^*(t) = \beta + \theta I, \quad t > \tau.$$

Then, we can obtain the density function f^* and c.d.f. F^* are, respectively,

$$f^*(t) = \lambda_0(t)e^{-\int_0^t \lambda_0(u)\mathrm{d}u} = \begin{cases} \beta e^{-\beta t}, & 0 \leq t \leq \tau, \\ (\beta + \theta)e^{-\beta t - \theta(t-\tau)}, & t > \tau, \end{cases}$$

and

$$F^*(t) = \begin{cases} 1 - e^{-\beta t}, & 0 \leq t \leq \tau, \\ 1 - e^{-\beta t - \theta(t-\tau)}, & t > \tau. \end{cases}$$

By (2.1),

$$\lambda(t)=\begin{cases}\dfrac{p\beta\exp(-\beta t)}{1-p+p\exp(-\beta t)}, & 0\le t\le\tau,\\ \dfrac{p(\beta+\theta)\exp\{-\beta t-\theta(t-\tau)\}}{1-p+p\exp\{-\beta t-\theta(t-\tau)\}}, & t>\tau.\end{cases}$$

Let P_n be the empirical probability measure, and $\mathbb{E}_n f \equiv \int f\mathrm{d}P_n$. Now, according to the preceding instructions, the log-likelihood function based on the observations (C_i,δ_i) can be written as

$$\begin{aligned}l_n(\boldsymbol{\phi}) =&\sum_{i=1}^n \delta_i(\log p+\log F^*(C_i))+(1-\delta_i)\log(1-pF^*(C_i))\\ =&\sum_{i=1}^n[\delta_i\{\log p+I(C_i\le\tau)\log(1-e^{-\beta C_i})+I(C_i>\tau)\log(1-e^{-\beta C_i-\theta(C_i-\tau)})\}+\\ &(1-\delta_i)\{I(C_i\le\tau)\log(1-p+pe^{-\beta C_i})+I(C_i>\tau)\log(1-p+pe^{-\beta C_i-\theta(C_i-\tau)})\}]\\ \equiv& n\mathbb{E}_n l(\boldsymbol{\mu},\boldsymbol{\nu}),\end{aligned}\tag{2.2}$$

where $\boldsymbol{\phi}=(\beta,\theta,p,\tau)^T$, $\boldsymbol{\mu}=(\beta,\theta)^T$, $\boldsymbol{\nu}=(p,\tau)^T$ and

$$\begin{aligned}l(\boldsymbol{\mu},\boldsymbol{\nu}) =&\delta\{\log p+I(C\le\tau)\log(1-e^{-\beta C})+I(C>\tau)\log(1-e^{-\beta C-\theta(C-\tau)})\}+\\ &(1-\delta)\{I(C\le\tau)\log(1-p+pe^{-\beta C})+I(C>\tau)\log(1-p+pe^{-\beta C-\theta(C-\tau)})\}.\end{aligned}$$

From (2.2), $l_n(\boldsymbol{\phi})$ is not continuous with respect to τ. Hence, the sufficient conditions for consistency of the maximum likelihood estimator (MLE) are not met. It is not appropriate to implement MLE. Thus we apply the pseudo-likelihood approach to overcome this difficulty.

The pseudo-likelihood approach was proposed by [7] and further studied by others including [8,9]. The key idea is to replace the true (but unknown) "nuisance" parameters p and τ in (2.2) by their consistent estimators $\hat{p}_n$ and $\hat{\tau}_n$, and then treat the log-likelihood function $l_n((\beta,\theta,\hat{p}_n,\hat{\tau}_n)^T)$, called the pseudo log-likelihood function, as a usual likelihood function of β and θ to generate the pseudo-MLE $(\hat{\beta}_n,\hat{\theta}_n)$ of (β,θ).

As in [10], the consistent estimator of p can be obtained by

$$\hat{p}_n=\hat{F}_n(C_{(n)}),\tag{2.3}$$

where $C_{(n)}=\max\{C_i;i=1,\dots,n\}$, and $\hat{F}_n(t)$ denotes the nonparametric estimate of the c.d.f. of failure times which is achieved by the ICM algorithm. The ICM algorithm proposed by [11] is fast in computing the NPMLE of the distribution function for current status data without covariates. We will show that $\hat{p}_n$ is consistent in Section 3.

As is common in change-point models, we suppose the existence of bounds τ_1 and τ_2 such that $0<\tau_1\le\tau\le\tau_2<\infty$ (see [1–4]). Note that the cumulative hazard function of T is

$$\Lambda(t)=\int_0^t\lambda(u)\mathrm{d}u=\begin{cases}-\log(1-p+pe^{-\beta t}), & 0\le t\le\tau,\\ -\log(1-p+pe^{-\beta t-\theta(t-\tau)}), & t>\tau.\end{cases}$$

Let

$$\tilde{\Lambda}(t)\equiv-\log[\frac{1}{p}\{e^{-\Lambda(t)}-1+p\}]=\begin{cases}\beta t, & 0\le t\le\tau,\\ \beta t+\theta(t-\tau), & t>\tau,\end{cases}$$

which is a piecewise linear function of t. Define

$$X(t) \equiv \{\frac{\tilde{\Lambda}(D) - \tilde{\Lambda}(t)}{D-t} - \frac{\tilde{\Lambda}(t) - \tilde{\Lambda}(0)}{t}\}\{t(D-t)\}, \tag{2.4}$$

for $0 < t < D$, where $D > \tau_2$. Then

$$X(t) = \tilde{\Lambda}(D)t - D\tilde{\Lambda}(t) = \theta(D-\tau)tI(t \leq \tau) + \theta\tau(D-t)I(t > \tau),$$

which is increasing (decreasing) on $[0, \tau]$ and decreasing (increasing) on $(\tau, D]$ if $\theta > 0$ ($\theta < 0$). Define $X_n(t)$ to be the empirical version of (2.4) with the unknown cumulative hazard function $\Lambda(t)$ and p replaced by the estimator $\hat{\Lambda}_n(t) = -\log\{1 - \hat{F}_n(t)\}$ and $\hat{p}_n$ in (2.3), respectively. Then an estimator of τ is given by

$$\hat{\tau}_n = \begin{cases} \inf\{t \in [\tau_1, \tau_2] : X_n(t\pm) = \sup\limits_{u \in [\tau_1, \tau_2]} X_n(u)\}, & \text{if } \theta > 0, \\ \inf\{t \in [\tau_1, \tau_2] : X_n(t\pm) = \inf\limits_{u \in [\tau_1, \tau_2]} X_n(u)\}, & \text{if } \theta < 0. \end{cases} \tag{2.5}$$

Through the asymptotic properties of $\hat{\Lambda}_n(t)$ and $\hat{p}_n$, we can also establish the consistency of $\hat{\tau}_n$.

3. Asymptotic results

For the sake of presentation, we introduce the notations and only consider the situation $\theta > 0$. Let $P_0 = P_{\boldsymbol{\phi}_0}$ be the true probability measure, where the subscript 0 implies the true parameters and let $\mathbb{E}_0$ denote the expectation of the random variables. Define the parameter spaces for $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ as $\boldsymbol{\Theta}_1 = \{\boldsymbol{\mu} : \beta \geq a_1, \theta \geq a_2\}$ and $\boldsymbol{\Theta}_2 = \{\boldsymbol{\nu} : |\tau - \tau_0| \leq b_1, |p - p_0| \leq b_2\}$, respectively, where a_1, a_2, b_1 and b_2 are some positive constants. The first partial derivatives of $l(\boldsymbol{\mu}, \boldsymbol{\nu})$ with respect to $\boldsymbol{\mu}$ are

$$\begin{aligned}\dot{l}_\beta(\boldsymbol{\mu}, \boldsymbol{\nu}) =& \delta\{I(C \leq \tau)\frac{C\exp(-\beta C)}{1 - \exp(-\beta C)} + I(C > \tau)\frac{C\exp\{-\beta C - \theta(C-\tau)\}}{1 - \exp\{-\beta C - \theta(C-\tau)\}}\} - \\ & (1-\delta)\{I(C \leq \tau)\frac{Cp\exp(-\beta C)}{1 - p + p\exp(-\beta C)} + I(C > \tau)\frac{(C-\tau)p\exp\{-\beta C - \theta(C-\tau)\}}{1 - p + p\exp\{-\beta C - \theta(C-\tau)\}}\},\end{aligned}$$

and

$$\dot{l}_\theta(\boldsymbol{\mu}, \boldsymbol{\nu}) = \delta I(C > \tau)\frac{(C-\tau)\exp\{-\beta C - \theta(C-\tau)\}}{1 - \exp\{-\beta C - \theta(C-\tau)\}} - (1-\delta)I(C > \tau)\frac{(C-\tau)p\exp\{-\beta C - \theta(C-\tau)\}}{1 - p + p\exp\{-\beta C - \theta(C-\tau)\}}.$$

The main results on asymptotic properties of $\hat{p}_n$, $\hat{\tau}_n$, $\hat{\beta}_n$ and $\hat{\theta}_n$ are presented in the next five theorems. In order to describe the theorems, we need to define the right extreme τ_{F^*} of T^* by

$$\tau_{F^*} = \sup\{t \geq 0 : \ F^*(t) < 1\},$$

and the right extreme τ_G of C by

$$\tau_G = \sup\{t \geq 0 : \ G(t) < 1\},$$

where G is the c.d.f. of C.

Theorem 3.1 *Suppose that $0 < p < 1$ and that F is continuous at τ_G in case $\tau_G < \infty$. Then $\hat{p}_n \to p$ in probability as $n \to \infty$ if and only if $\tau_{F^*} \leq \tau_G$.*

Proof By the conclusion after Lemma 4.1 in page 78 of [11],

$$\sup_{t\in[0,\infty]}|\hat{F}_n(t)-F(t)|\to 0$$

in probability. Since $C_{(n)}\le\infty$ almost surely, we obtain that $|\hat{F}_n(C_{(n)})-F(C_{(n)})|\to 0$ in probability. Define $H_n(t)=\sum_{i=1}^n I(C_i\ge t)$. When $t<\tau_G$, $\frac{H_n(t)}{n}\to 1-G(t)$ almost surely by the strong law of large numbers, and $1-G(t)>0$. Hence, $H_n(t)\to\infty$, which implies $C_{(n)}\to\tau_G$ in probability. If $\tau_G\le\infty$, $\hat{F}_n(C_{(n)})\to F(\tau_G)$ in probability. If $\tau_G=\infty$, $\hat{F}_n(C_{(n)})\to p=F(\tau_G)$ in probability. Note that

$$\tau_{F^*}=\sup\{t:F^*(t)<1\}=\sup\{t:F(t)<p\}.$$

Hence $F(\tau_G)=p$ if and only if $\tau_G\ge\tau_{F^*}$, and then the theorem follows. $\Box$

Theorem 3.2 *Assume that F is continuous at τ_G in case $\tau_G<\infty$ and $\tau_{F^*}>D$. Then the estimator $\hat{\tau}_n$ of τ defined in (2.5) is consistent.*

Proof Define:

$$X_n(t)=\{\frac{\hat{\tilde{\Lambda}}_n(D)-\hat{\tilde{\Lambda}}_n(t)}{D-t}-\frac{\hat{\tilde{\Lambda}}_n(t)-\hat{\tilde{\Lambda}}_n(0)}{t}\}t(D-t),\quad 0<t<D,$$

and

$$X_n^0(t)=\{\frac{\hat{\tilde{\Lambda}}_n^0(D)-\hat{\tilde{\Lambda}}_n^0(t)}{D-t}-\frac{\hat{\tilde{\Lambda}}_n^0(t)-\hat{\tilde{\Lambda}}_n^0(0)}{t}\}t(D-t),\quad 0<t<D,$$

where

$$\hat{\tilde{\Lambda}}_n(t)=-\log[-\frac{1}{\hat{p}_n}\{\exp(-\hat{\Lambda}_n(t))-1+\hat{p}_n\}],$$

$\hat{\Lambda}_n(t)$ is obtained by $\hat{F}_n$, and

$$\hat{\tilde{\Lambda}}_n^0(t)=-\log[-\frac{1}{p}\{\exp(-\hat{\Lambda}_n)-1+p\}].$$

Notice that $\tilde{\Lambda}(0)=\hat{\tilde{\Lambda}}_n^0(0)=\Lambda(0)=0$. Then $X(t)=t\tilde{\Lambda}(D)-D\tilde{\Lambda}(t)$. For any $\varepsilon>0$, let $c_1\in(0,\min\{X(\tau)-X(\tau-\varepsilon),X(\tau)-X(\tau+\varepsilon)\})$ relying on ε, τ_1, τ_2, p, θ. Then, if $|t-\tau|>\varepsilon$, we have $X(\tau)-X(\tau+\varepsilon)>c_1$. Noting that $X_n(t)$ attains its maximum at $\hat{\tau}_n$, for sufficiently large n, we have

$$\begin{aligned}
&P_0(|\hat{\tau}_n-\tau|>\varepsilon)\le P_0(X(\tau)-X(\hat{\tau}_n)|>c_1)\\
&\quad\le P_0(|X_n(\hat{\tau})-X(\hat{\tau}_n)|+|X(\tau)-X_n(\tau)|>c_1)\\
&\quad=P_0(|X_n(\hat{\tau}_n)-X(\hat{\tau}_n)|+|X(\tau)-X_n(\tau)|>c_1,\sup_{\tau_1<t<\tau_2}|X_n(t)-X(t)|>\frac{c_1}{2})+\\
&\quad\ \ P_0(|X_n(\hat{\tau}_n)-X(\hat{\tau}_n)|+|X(\tau)-X_n(\tau)|>c_1,\sup_{\tau_1<t<\tau_2}|X_n(t)-X(t)|\le\frac{c_1}{2})\\
&\quad\le P_0(\sup_{\tau_1<t<\tau_2}|X_n(t)-X(t)|>\frac{c_1}{2})\\
&\quad\le P_0(\sup_{\tau_1<t<\tau_2}|X_n(t)-X_n^0(t)|>\frac{c_1}{4})+P_0(\sup_{\tau_1<t<\tau_2}|X_n^0(t)-X(t)|>\frac{c_1}{4})\\
&\quad\le P_0(D\sup_{\tau_1<t<\tau_2}|U_n^0(t)|+\tau_2U_n^0(D)>\frac{c_1}{4})+P_0(\sup_{\tau_1<t<\tau_2}|X_n(t)-X_n^0(t)|>\frac{c_1}{4}).
\end{aligned}$$

We can obtain the last inequality by

$$
\begin{aligned}
X_n^0(t) - X(t) =& t\{\hat{\tilde{\Lambda}}_n^0(D) - \tilde{\Lambda}(D)\} - D\{\hat{\tilde{\Lambda}}_n^0(t) - \tilde{\Lambda}(t)\} \\
=& tU_n^0(D) - DU_n^0(t) \le t|U_n^0(D)| + D|U_n^0(t)|,
\end{aligned}
$$

where $U_n^0(t) = \hat{\tilde{\Lambda}}_n^0(t) - \tilde{\Lambda}(t)$. Consequently, there exist $c_2 > 0$ and $c_3 > 0$, depending on c_1, τ_1, τ_2, D and q, such that

$$
\begin{aligned}
&P_0(|\hat{\tau}_n - \tau| > \varepsilon) \le P_0(\sup_{\tau_1<t<\tau_2} |U_n^0(t)| > c_2) + P_0(U_n^0(D) > c_3) + P_0(\sup_{\tau_1<t<\tau_2} |X_n(t) - X_n^0(t)| > \frac{c_1}{4}) \\
&= I_1 + I_2 + I_3. \qquad (3.1)
\end{aligned}
$$

From the definition of $\tilde{\Lambda}(t)$, we find that

$$
\begin{aligned}
|U_n^0(t)| =& |\log(-\frac{1}{p}\{e^{-\hat{\Lambda}_n(t)} - 1 + p\}) - \log(-\frac{1}{p}\{e^{-\Lambda(t)} - 1 + p\})| \\
=& |\frac{e^{-\alpha(t)}}{e^{-\alpha(t)} - 1 + p}| \cdot |\hat{\Lambda}_n(t) - \Lambda(t)|, \qquad (3.2)
\end{aligned}
$$

where $\alpha(t)$ is between $\hat{\Lambda}_n(t)$ and $\Lambda(t)$. Thus, $\exp(-\alpha(t))$ lies on the segment between $\hat{S}(t) = 1 - \hat{F}_n(t)$ and $S_n(t) = 1 - F(t) = 1 - pF^*(t)$. For current status data, according to [11], $\sup_{t\in[0,\tau_{F^*}]} |\hat{F}_n(t) - F(t)| \to 0$ almost surely for $\tau_{F^*} \le \tau_G$. Thus for any $\alpha < 1 - pF^*(D)$,

$$
\exp(-\alpha(t)) > [1 - F(D)] - \alpha = [1 - pF^*(D)] - \alpha = \phi(D),
$$

provided that $\tau_{F^*} > D$. It follows (3.2) that

$$
|U_n^0(t)| \le \frac{1}{\phi(D) - 1 + p}|\hat{\Lambda}_n(t) - \Lambda(t)| = \frac{1}{\phi(D) - 1 + p}|U_n(t)|.
$$

By the likelihood function $l_n(\boldsymbol{\phi})$, there exists $c_4 > 0$ relying on $c_1, c_2, \tau_1, \tau_2, D, q, p$ and F^*, satisfying

$$
\begin{aligned}
I_1 \le& P_0(\sup_{\tau_1<t<\tau_2} |U_n(t)| > c_4, \tau_2 \le C_{(n)}) + P_0(C_{(n)} < \tau_2) \\
\le& P_0(\sup_{\tau_1<t<\tau_2} |U_n(t \wedge C_{(n)})| > c_4) + \prod_{i=1}^{n} P(C_i < \tau_2). \qquad (3.3)
\end{aligned}
$$

We know that

$$
\hat{\Lambda}_n(t \wedge C_{(n)}) - \Lambda(t \wedge C_{(n)}) = \log \frac{1 - \hat{F}_n(t \wedge C_{(n)})}{1 - F(t \wedge C_{(n)})}. \qquad (3.4)
$$

By $P_0(\lim_{n\to\infty} \sup_{t\in\mathbb{R}} |\hat{F}_n(t) - F(t)| = 0) = 1$ obtained by [11], the first term on the right side of last inequality of (3.3) converges to 0 as $n \to \infty$. Next, by (3.1), $I_2 \le P_0(|U_n(D)| > c_3, C_{(n)} \ge D) + P_0(C_{(n)} < D)$. Similarly, I_2 converges to 0 as $n \to \infty$.

In order to prove $I_3 \to 0$, we rewrite $X_n(t)$ and $X_n^0(t)$ as

$$
X_n(t) = t\{\hat{\tilde{\Lambda}}_n(t) - \hat{\tilde{\Lambda}}_n(t)\} - (D - t)\hat{\tilde{\Lambda}}_n(t),
$$

and

$$
X_n^0(t) = t\{\hat{\tilde{\Lambda}}_n^0(D) - \hat{\tilde{\Lambda}}_n^0(t)\} - (D - t)\hat{\tilde{\Lambda}}_n^0(t).
$$

By (3.3) and (3.4),

$$
\begin{aligned}
I_3 \leq & P_0(\sup_{\tau_1<t<\tau_2}|X_n(t)-X_n^0(t)|>\frac{c_1}{4}, C_{(n)}\geq D)+P_0(C_{(n)}<D)\\
\leq & P_0(2\sup_{\tau_1<t<\tau_2}|\hat{\hat{\Lambda}}_n^0(t)-\hat{\hat{\Lambda}}_n^0(t)|\tau_2>\frac{c_4}{8})+\\
& P_0(\sup_{\tau_1<t<\tau_2}|\hat{\hat{\Lambda}}_n^0(t)-\hat{\hat{\Lambda}}_n^0(t)|(D-\tau_2)>\frac{c_4}{8})+P_0(C_{(n)}<D)\\
= & I_{31}+I_{32}+\prod_{i=1}^{n}P(C_i<D).
\end{aligned}
$$

We can see that

$$
I_{31}\leq P_0(|\log(\hat{p}_n)-\log(p)|+\sup_{\tau_1<t<\tau_2}\big|\log\frac{e^{\hat{\Lambda}_n(t)}-1+\hat{p}_n}{e^{\hat{\Lambda}_n(t)}-1+p}\big|>\frac{c_4}{8}).
$$

Since $\hat{p}_n$ converges to p in probability, and $\sup_{0<t<D}|\hat{\Lambda}_n(t)-\Lambda(t)|\to 0$, we have $I_{31}\to 0$. Similarly, $I_{32}\to 0$. This completes the proof of Theorem 3.2. $\square$

Theorem 3.3 *Suppose that $\hat{p}_n$ and $\hat{\tau}_n$ are obtained by (2.3) and (2.4), respectively. Then $\mathbb{E}_n\dot{l}_{\boldsymbol{\mu}}(\hat{\boldsymbol{\mu}}_n,\hat{\boldsymbol{\nu}}_n)=o_{p^*}(n^{-1/2})$ almost surely, where $\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu})$ denotes the first partial derivative of $l(\boldsymbol{\mu},\boldsymbol{\nu})$ with respect to $\boldsymbol{\mu}$, and $\hat{\boldsymbol{\mu}}_n$ converges in probability to $\boldsymbol{\mu}_0$.*

Proof To prove the consistency of the pseudo estimator $\hat{\boldsymbol{\mu}}$, we first need to prove

$$
\sup_{\boldsymbol{\mu}\in\boldsymbol{\Theta}_1,|\boldsymbol{\nu}-\boldsymbol{\nu}_0|\leq\eta_n}|\mathbb{E}_n\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu})-\mathbb{E}_0\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu}_0)|=o_p(1)
$$

for every sequence $\{\eta_n\}\downarrow 0$. Since

$$
|\mathbb{E}_n\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu})-\mathbb{E}_0\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu}_0)|\leq|(\mathbb{E}_n-\mathbb{E}_0)\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu})|+|\mathbb{E}_0(\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu})-\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu}_0))|,
$$

and $\mathbb{E}_0\ddot{l}_{\boldsymbol{\mu}\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu})$ obviously tends to zero when $|\boldsymbol{\nu}-\boldsymbol{\nu}_0|\leq\alpha\downarrow 0$, where $\ddot{l}_{\boldsymbol{\mu}\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu})$ denotes $\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu})\dot{l}_{\boldsymbol{\mu}}^T(\boldsymbol{\mu},\boldsymbol{\nu})$. We need to show that the class of functions $F_\alpha\equiv\{\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu}):\boldsymbol{\mu}\in\boldsymbol{\Theta}_1\subset\mathbb{R}^2,|\boldsymbol{\nu}-\boldsymbol{\nu}_0|\leq\eta_n\}$ is a VC-class for some $\eta_n>0$, where $\boldsymbol{\Theta}_1=\{\boldsymbol{\mu}=(\beta,\theta)^T:\beta\geq A_1,\theta\geq A_2\}$. This implies that the uniform strong law of large numbers holds, i.e., $\sup_{f\in F_\alpha}(\mathbb{E}_n-\mathbb{E}_0)f\xrightarrow{p}0$; see [12], Chap. 2.6-2.7, for details. Let $F_{1\alpha}=\{I(C\leq\tau):|\tau-\tau_0|\leq\alpha_1\}$. Then the VC-indexes of the class of functions $F_{1\alpha}$ is 2 by Example 2.6.1 of [12]. Thus the class of functions

$$
\{I(C_i\leq\tau)\frac{C_i(1-\delta_i)p\exp(-\beta C_i)}{1-p+p\exp(-\beta C_i)}:\boldsymbol{\mu}\in\boldsymbol{\Theta}_1,|\boldsymbol{\nu}-\boldsymbol{\nu}_0|\leq\alpha\}
$$

is Donsker by Lemma 2.6.18 and Example 2.10.8 of [12], because $(1-\delta_i)(1-p)/(1-p+p\exp(-\beta C_i))$ is bounded. It is similar to show that the other classes of functions are also Donsker. Thus the class of functions F_α is VC-class by applying Example 2.10.7 and Theorem 2.10.6 of [12]. Since $\boldsymbol{\mu}_0$ is the unique solution to $\mathbb{E}_0\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu}_0)=0$, $\hat{\boldsymbol{\nu}}_n\to\boldsymbol{\nu}_0$, and $\hat{\boldsymbol{\mu}}_n$ is the unique solution of $\mathbb{E}_n\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\hat{\boldsymbol{\nu}}_n)$, $\hat{\boldsymbol{\mu}}_n$ converges in probability to $\boldsymbol{\mu}_0$. $\square$

Theorem 3.4 *Under the conditions in Theorem 3.3, $\sqrt{n}(\hat{\boldsymbol{\mu}}_n-\boldsymbol{\mu}_0)=O_p(1)$.*

Proof We first verify the stochastic equicontinuity condition:

$$|\sqrt{n}(\mathbb{E}_n-\mathbb{E}_0)\{\dot{l}_{\boldsymbol{\mu}}(\hat{\boldsymbol{\mu}}_n,\hat{\boldsymbol{\nu}}_n)-\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_0,\boldsymbol{\nu}_0)\}|=o_p(1). \tag{3.5}$$

Let $F_\gamma=\{\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu})-\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_0,\boldsymbol{\nu}_0):|\boldsymbol{\mu}-\boldsymbol{\mu}_0|\le\gamma,|\boldsymbol{\nu}-\boldsymbol{\nu}_0|\le\gamma\}$. Similar to the proof of Theorem 3, we can show that F_γ is a VC-class. Thus (3.5) follows from Lemma 3.1.1 of [8], together with $\mathbb{E}_0(\dot{l}_{\boldsymbol{\mu}}(\hat{\boldsymbol{\mu}}_n,\hat{\boldsymbol{\nu}}_n)-\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_0,\boldsymbol{\nu}_0))^2=o_p(1)$. Next, since $\mathbb{E}_0\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu})<\infty$ and $\mathbb{E}_0\ddot{l}_{\boldsymbol{\mu}\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu})<\infty$, we obtain that for $(\boldsymbol{\mu},\boldsymbol{\nu})\in D_n$,

$$\begin{aligned}&|\mathbb{E}_0\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\nu})-\mathbb{E}_0\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_0,\boldsymbol{\nu}_0)-\mathbb{E}_0\ddot{l}_{\boldsymbol{\mu}\boldsymbol{\mu}}(\boldsymbol{\mu}_0,\boldsymbol{\nu}_0)(\boldsymbol{\mu}-\boldsymbol{\mu}_0)-\mathbb{E}_0\ddot{l}_{\boldsymbol{\mu}\boldsymbol{\nu}}(\boldsymbol{\mu}_0,\boldsymbol{\nu}_0)(\boldsymbol{\nu}-\boldsymbol{\nu}_0)|\\ &\quad=o(|\boldsymbol{\mu}-\boldsymbol{\mu}_0|)+o(|\boldsymbol{\nu}-\boldsymbol{\nu}_0|),\end{aligned}$$

where $D_n=\{(\boldsymbol{\mu},\boldsymbol{\nu}):|\boldsymbol{\mu}-\boldsymbol{\mu}_0|\le\eta_n\downarrow 0,|\boldsymbol{\nu}-\boldsymbol{\nu}_0|\le cn^{1/2}\}$ for some constant c, and $\mathbb{E}_n\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_0,\boldsymbol{\nu}_0)$ converges in distribution to a normal random variable by the central limit theorem. Thus $\sqrt{n}|\hat{\boldsymbol{\mu}}_n-\boldsymbol{\mu}|=O_p(1)$ by Theorem 3.1.3 of [8]. □

Theorem 3.5 *Under the conditions in Theorem 3.3,* $\sqrt{n}(\hat{\boldsymbol{\mu}}_n-\boldsymbol{\mu}_0)$ *is asymptotically normal with mean 0 and variance* $\{\mathbb{E}_0\ddot{l}_{\boldsymbol{\mu}\boldsymbol{\mu}}(\boldsymbol{\mu}_0,\boldsymbol{\nu}_0)\}^{-2}\boldsymbol{V}$*, where* $\boldsymbol{V}=\mathrm{Var}\{\boldsymbol{\Lambda}_1+P_0\ddot{l}_{\boldsymbol{\mu}\boldsymbol{\nu}}(\boldsymbol{\mu}_0,\boldsymbol{\nu}_0)\boldsymbol{\Lambda_2}\}$*, and* $\boldsymbol{\Lambda}_1$ *and* $\boldsymbol{\Lambda_2}$ *are random vectors satisfying*

$$\sqrt{n}\begin{bmatrix}(\mathbb{E}_n-\mathbb{E}_0)\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_0,\boldsymbol{\nu}_0)\\ \hat{\boldsymbol{\nu}}_n-\boldsymbol{\nu}_0\end{bmatrix}\xrightarrow{d}\begin{bmatrix}\boldsymbol{\Lambda_1}\\ \boldsymbol{\Lambda_2}\end{bmatrix}.$$

Proof By the consistency of $\hat{p}_n$ and $\hat{\tau}_n$ together with the Slutsky's theorem and the central limit theorem, we can show that

$$\sqrt{n}\begin{bmatrix}(\mathbb{E}_n-\mathbb{E}_0)\dot{l}_{\boldsymbol{\mu}}(\boldsymbol{\mu}_0,\boldsymbol{\nu}_0)\\ \hat{\boldsymbol{\nu}}-\boldsymbol{\nu}_0\end{bmatrix}\xrightarrow{d}\boldsymbol{\Lambda}=\begin{bmatrix}\boldsymbol{\Lambda_1}\\ \boldsymbol{\Lambda_2}\end{bmatrix}$$

holds for normally distributed Λ_1 with mean zero and positive variance. Hence by Corollary 3.1.2 of [8], $\sqrt{n}(\hat{\boldsymbol{\mu}}_n-\boldsymbol{\mu}_0)$ is asymptotically normal with mean 0 and variance $\{\mathbb{E}_0\ddot{l}_{\boldsymbol{\mu}\boldsymbol{\mu}}(\boldsymbol{\mu}_0,\boldsymbol{\nu}_0)\}^{-2}\boldsymbol{V}$. □

4. Simulation results

To access the finite-sample performance of the estimation, we do many simulations through different settings. Particularly, we want to explore the impacts of the jump size, the change-point location and the monitor time.

The data are simulated from the hazard function (2.1) and the corresponding distribution function is

$$F(t)=\begin{cases}1-p+pe^{-\beta t}, & 0\le t\le\tau,\\ 1-p+pe^{-\beta t-\theta(t-\tau)}, & t>\tau,\end{cases}$$

where $p=0.8$, and $\beta=1$. We set $\theta\in\{0.5,\ 1,\ 1.5\}$ to check the influence of the jumping size, and set $\tau\in\{0.5,\ 1,\ 1.5\}$ to see the effect of the change-point location. The change-point search range (τ_1,τ_2) is set to $(0.25,1.75)$. The failure time $\tilde{T}_i$ is generated by solving $F(t)=u_i$ numerically, where $u_i\sim U(0,1)$. The c.d.f. of the monitor time follows an exponential distribution and the left censored rate is 50%. The sample size $n\in\{200,\ 400,\ 800\}$. All results are based on

500 replications. We compare our estimation with the MLE suggested by [7]. They obtain the estimators as follows: with a fixed τ, let $\hat{\boldsymbol{\xi}}_n(\tau)$ be the value of $\boldsymbol{\xi}=(\beta,\theta,p)$ maximizing $l_n(\boldsymbol{\phi})$. Then τ is estimated by

$$\hat{\tau}_n=\inf\{\tau\in[\tau_1,\tau_2]:\max(l_n(\hat{\boldsymbol{\xi}}_n(\tau),\tau),l_n(\hat{\boldsymbol{\xi}}_n(\tau\pm),\tau\pm))=\sup_{\tau\in[\tau_1,\tau_2]}l_n(\hat{\boldsymbol{\xi}}_n(\tau),\tau)\}.$$

Then the maximum likelihood estimator of $\boldsymbol{\xi}$ is obtained as $\hat{\boldsymbol{\xi}}_n=\hat{\boldsymbol{\xi}}_n(\hat{\tau}_n)$.

θ	n		MLE				PMLE			
			$\hat{\tau}_n$	$\hat{p}_n$	$\hat{\beta}_n$	$\hat{\theta}_n$	$\hat{\tau}_n$	$\hat{p}_n$	$\hat{\beta}_n$	$\hat{\theta}_n$
0.5	200	bias	0.121	0.008	0.113	0.476	0.041	0.027	0.107	0.270
		sd	0.479	0.066	0.295	0.757	0.330	0.064	0.269	0.707
	400	bias	0.081	0.005	0.035	0.325	0.036	0.011	0.037	0.090
		sd	0.472	0.044	0.235	0.737	0.317	0.038	0.200	0.496
	800	bias	0.075	0.001	0.062	0.279	0.037	0.005	0.035	0.197
		sd	0.304	0.030	0.172	0.668	0.288	0.032	0.161	0.348
1	200	bias	0.112	0.009	0.136	0.369	0.013	0.047	0.137	0.268
		sd	0.454	0.063	0.263	0.779	0.319	0.065	0.213	0.732
	400	bias	0.078	0.005	0.040	0.203	0.001	0.044	0.055	0.168
		sd	0.405	0.048	0.219	0.676	0.307	0.055	0.191	0.532
	800	bias	0.046	0.001	0.053	0.213	0.021	0.002	0.036	0.155
		sd	0.275	0.030	0.173	0.648	0.207	0.033	0.160	0.370
1.5	200	bias	0.084	0.001	0.035	0.410	0.047	0.005	0.102	0.189
		sd	0.412	0.064	0.274	0.854	0.290	0.076	0.278	0.748
	400	bias	0.034	0.004	0.040	0.310	0.021	0.001	0.001	0.082
		sd	0.302	0.043	0.219	0.834	0.260	0.044	0.240	0.518
	800	bias	0.034	0.006	0.054	0.154	0.013	0.003	0.030	0.142
		sd	0.219	0.027	0.177	0.428	0.183	0.030	0.135	0.342

Table 1 Simulation results for the change point τ, uncure rate p, hazard rate β and jump size θ (τ=0.5)

Table 1 shows the empirical biases and sample standard deviations (sd) of the estimators considering $\theta\in\{0.5,1,1.5\}$, $\tau=0.5$, and the sample size $n\in\{200,\ 400,\ 800\}$. From Table 1, the results indicate that both MLE and PMLE perform reasonably well. More specifically, the biases and sd of the change-point estimator are smaller with a larger jump size. The performances of the other estimators are not affected by the jump size. Additionally, the proposed method PMLE provides smaller biases and sd in most cases than MLE. Tables 2 and 3 display the results with $\tau=1,1.5$, respectively. As in Table 1, the same conclusion can be obtained from the results in Tables 2 and 3. And by comparing Tables 1–3, we obtain that the change-point location has no obvious influence on the estimation.

5. Discussion

In this paper, we suggest the PMLE to handle the single change-point hazards cure model for current status data. To obtain consistent estimators of the "nuisance" parameters, uncured rate and change point, we apply ICM method to calculate the NPMLE of the distribution function.

θ	n		MLE				PMLE			
			$\hat{\tau}_n$	$\hat{p}_n$	$\hat{\beta}_n$	$\hat{\theta}_n$	$\hat{\tau}_n$	$\hat{p}_n$	$\hat{\beta}_n$	$\hat{\theta}_n$
0.5	200	bias	0.046	0.004	0.037	0.226	0.069	0.040	0.081	0.253
		sd	0.460	0.070	0.262	0.918	0.414	0.069	0.206	0.721
	400	bias	0.030	0.002	0.002	0.290	0.044	0.016	0.029	0.203
		sd	0.415	0.053	0.178	0.740	0.341	0.044	0.153	0.508
	800	bias	0.060	0.001	0.061	0.259	0.027	0.003	0.055	0.118
		sd	0.312	0.032	0.174	0.679	0.288	0.041	0.136	0.340
1	200	bias	0.032	0.008	0.058	0.144	0.063	0.037	0.093	0.224
		sd	0.454	0.064	0.232	0.923	0.336	0.062	0.199	0.744
	400	bias	0.014	0.007	0.002	0.335	0.026	0.025	0.036	0.174
		sd	0.324	0.052	0.188	0.776	0.284	0.046	0.146	0.540
	800	bias	0.043	0.005	0.057	0.232	0.026	0.005	0.039	0.155
		sd	0.232	0.030	0.173	0.618	0.207	0.040	0.131	0.365
1.5	200	bias	0.022	0.019	0.076	0.235	0.035	0.031	0.082	0.120
		sd	0.437	0.057	0.238	0.917	0.320	0.053	0.215	0.718
	400	bias	0.013	0.006	0.016	0.167	0.011	0.018	0.025	0.020
		sd	0.309	0.040	0.178	0.828	0.250	0.034	0.145	0.548
	800	bias	0.026	0.001	0.052	0.199	0.022	0.003	0.040	0.133
		sd	0.213	0.027	0.164	0.434	0.180	0.032	0.135	0.350

Table 2 Simulation results for the change point τ, uncure rate p, hazard rate β and jump size θ (τ=1)

θ	n		MLE				PMLE			
			$\hat{\tau}_n$	$\hat{p}_n$	$\hat{\beta}_n$	$\hat{\theta}_n$	$\hat{\tau}_n$	$\hat{p}_n$	$\hat{\beta}_n$	$\hat{\theta}_n$
0.5	200	bias	0.105	0.008	0.039	0.361	0.045	0.039	0.091	0.192
		sd	0.618	0.059	0.232	0.950	0.426	0.051	0.197	0.778
	400	bias	0.130	0.001	0.065	0.295	0.041	0.002	0.037	0.113
		sd	0.412	0.044	0.164	0.820	0.275	0.041	0.146	0.578
	800	bias	0.084	0.001	0.065	0.295	0.041	0.002	0.037	0.213
		sd	0.360	0.034	0.164	0.638	0.275	0.041	0.125	0.378
1	200	bias	0.051	0.013	0.058	0.316	0.048	0.027	0.070	0.155
		sd	0.597	0.058	0.235	0.916	0.353	0.061	0.198	0.766
	400	bias	0.015	0.002	0.007	0.378	0.058	0.013	0.013	0.058
		sd	0.417	0.045	0.159	0.748	0.252	0.042	0.144	0.576
	800	bias	0.056	0.004	0.064	0.190	0.031	0.002	0.040	0.178
		sd	0.312	0.032	0.162	0.648	0.211	0.040	0.122	0.371
1.5	200	bias	0.036	0.013	0.064	0.255	0.041	0.032	0.084	0.141
		sd	0.552	0.062	0.237	0.938	0.302	0.059	0.194	0.802
	400	bias	0.067	0.006	0.075	0.187	0.053	0.014	0.065	0.089
		sd	0.423	0.046	0.171	0.753	0.275	0.043	0.167	0.598
	800	bias	0.036	0.006	0.046	0.198	0.028	0.011	0.025	0.180
		sd	0.202	0.037	0.153	0.643	0.234	0.038	0.126	0.365

Table 3 Simulation results for the change point τ, uncure rate p, hazard rate β and jump size θ (τ=1.5)

Compared with the MLE, the PMLE possesses a smaller bias and standard deviation. The simulation studies illustrate that the proposed method can effectively deal with the change-point problem in current status data.

References

[1] D. E. MATTHEWS, V. T. FAREWELL. *On testing for a constant hazard against a change-point alternative*. Biometrics, 1982, **38**(2): 463–468.

[2] C. R. LOADER. *Inference for a hazard rate change point*. Biometrika, 1991, **78**(4): 749–757.

[3] I. GIJBELS, Ü. GÜRLER. *Estimation of a change point in a hazard function based on censored data*. Lifetime Data Anal., 2003, **9**(4): 395–411.

[4] Xiaobing ZHAO, Xianyi WU, Xian ZHOU. *A change-point model for survival data with long-term survivors*. Statist. Sinica, 2009, **19**(1): 377–390.

[5] M. OTHUS, Yi LI, R. TIWARI. *Change-point cure models with application to estimating the change-point effect of age of diagnosis among prostate cancer patients*. J. Appl. Stat., 2012, **39**(4): 901–911.

[6] A. J. ROSSINI, A. A. TSIATIS. *A semiparametric proportional odds regression model for the analysis of current status data*. J. Amer. Statist. Assoc., 1996, **91**(434): 713–721.

[7] G. GONG, F. SAMANIEGO. *Pseudo maximum likelihood estimation: theory and applications*. Ann. Statist., 1981, **9**(4): 861–869.

[8] Huilin HU. *Large sample theory for pseudo-maximum likelihood estimates in semiparametric models*. Thesis (Ph.D.)-University of Washington, 1998.

[9] Jian HUANG. *Efficient estimation for the proportional hazards model with interval censoring*. Ann. Statist., 1996, **24**(2): 540–568.

[10] R. A. MALLER, S. ZHOU. *Estimating the proportion of immunes in a censored sample*. Biometrika, 1992, **79**(4): 731–739.

[11] P. GROENEBOOM, J. WELLNER. *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Birkhäuser Verlag, Basel, 1992.

[12] A. W. VAART, J. A. WELLNER. *Weak Convergence and Empirical Processes: with Applications to Statistics*. Springer-Verlag, New York, 1996.