Bernoulli **23**(4A), 2017, 2720–2745
DOI: 10.3150/16-BEJ825

Inference under biased sampling and right censoring for a change point in the hazard function

YASSIR RABHI[1] and MASOUD ASGHARIAN[2]

[1]*Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Canada. E-mail: rabhi@ualberta.ca*
[2]*Department of Mathematics and Statistics, McGill University, Montréal, Canada. E-mail: masoud@math.mcgill.ca*

Length-biased survival data commonly arise in cross-sectional surveys and prevalent cohort studies on disease duration. Ignoring biased sampling leads to bias in estimating the hazard-of-failure and the survival-time in the population. We address estimating the location of a possible change-point of an otherwise smooth hazard function when the collected data form a biased sample from the target population and the data are subject to informative censoring. We provide two estimation methodologies, for the location and size of the change-point, adapted to two scenarios of the truncation distribution: known and unknown. While the estimators in the first case show gain in efficiency as compared to those in the second case, the latter is more robust to the form of the truncation distribution. In both cases, the change-point estimators can achieve the rate $\mathcal{O}_p(1/n)$. We study the asymptotic properties of the estimates and devise interval-estimators for the location and size of the change, paving the way towards making statistical inference about whether or not a change-point exists. Several simulated examples are discussed to assess the finite sample behavior of the estimators. The proposed methods are then applied to analyze a set of survival data collected on elderly Canadian citizen (aged 65+) suffering from dementia.

Keywords: biased sampling; change point; informative censoring; jump size; left truncation; prevalent cohort survival data; survival with dementia

1. Introduction

Data on survival, or failure time, typically comprise an initiating event, say onset of a disease, and a terminating event, say death. In an ideal situation, incident cases, that is, subjects who have not experienced their initiating event before recruitment, are sampled. The recruited samples are then followed to a terminating event or censoring. Logistic or other constraints may, however, preclude the possibility of recruiting incident cases. A feasible alternative in such circumstances is to recruit subjects who have already experienced the initiating event, the so-called prevalent case, often through cross-sectional sampling. Despite its feasibility, the use of cross-sectional sampling generally leads to selecting subjects with longer survival time. The survival time of subjects so recruited is left truncated. In other words, the survival data collected on prevalent cases form a biased sample from the population of interest. When subjects are further followed-up, that is, cross-sectional sampling with follow-up, identified cases are also subject to right censoring. Such censoring is informative.

1350-7265 © 2017 ISI/BS

Prevalent cohort studies have attracted attention and interest in many areas of research over the past decades. Some recent examples of prevalent cohort studies include [27] in screening and the early detection of disease, [26] in epidemiology and geriatric medicine, and [5] in labor force studies among others. The literature on biased sampling is vast and can be traced as far back as [25], with seminal contributions by [4,9,17]. For applications of biased sampling in other areas, see also [18] in land valuation, [19] in marketing, [22] in genetics and linkage mapping, [7] in applied physics, [2] in prevalent cohort studies on disease duration, [12] in nano physics and [13] in water quality.

In survival analysis, the risk pattern of a disease, mortality, or other events in time is a practical concern for many studies, and are usually explored via the survival or hazard function. When our interest revolves around the risk of failure, however, a careful study of the hazard function can be more helpful as it provides a refined insight into the structural changes of the risk pattern of a time-related event. An abrupt change in the hazard of failure, in particular, can be quite informative about the risk pattern and hence of prime importance in many applications.

The estimation of a change-point in the hazard function has been studied in different settings. Early works focused on the parametric setting. [1] and [16] studied the change-point problem in the nonparametric setting when observations are, only, subject to right censoring. The former used wavelets while the latter employed kernel smoothing. The semi-parametric approach was recently taken by [20], who considered the change-point in the Cox regression model, and [11] who studied the same problem using transformation models. All of these studies have, however, considered only representative samples from the population of interest, and hence limited to incident cases.

When observations form a biased sample from the target population, one needs methods that account for left truncation in addition to informative censoring. Failure to do so leads to inconsistent estimators of the jump-size function, $\Delta(x)$, used to locate the change-point. One can show, for instance in the nonparametric approach, that estimators of the jump-size devised for representative samples tend to $r_b(x) \times \Delta(x)$, where

$$r_b(x) = \frac{\Pr[T \leq x \leq Y]}{\Pr[x \leq Y]} \leq 1, \tag{1.1}$$

T denotes the observed left-truncation and Y denotes possibly censored lifetime. The function $r_b(x)$ represents the proportion of subjects in the sampled population who can be under study at time x out of those who survive beyond x. This proportion is close to zero for small values of x. As a consequence, the closer the change-point is to zero, the more chance it be missed if we ignore biased sampling. It should be noted that the settings of left-censoring and left-truncation with right-censoring are different. While censoring means information exists, but we may not have access to it, truncation, however, means that some part of the sample space is not observable. Formally speaking, let X^*, C and T^* be respectively the lifetime (variable of interest), the censoring-time and the truncation-time. In the left-censoring setting, on each subject we observe $(\max(X^*; C), \delta)$, while in the left-truncation with right-censoring setting, we observe $(\min(X^*; C), T^*, \delta)$ if $\min(X^*; C) \geq T^*$, where $\delta = \mathrm{I}(X^* \geq C)$ is the failure indicator.

We provide nonparametric methods for estimating a possible change-point (location and size) of an otherwise smooth hazard function when the collected data form a biased sample from the target population and the data are subject to informative censoring. Given an appropriate odd function K_*, and a sequence h_n of bandwidth tending to zero, we devise an estimation procedure based on estimating the maximum modulus of the convolution of $K_*(\cdot/h_n)/h_n$ with the hazard function. We introduce two types of estimators adapted to the available information about the truncation distribution: known and unknown. While the estimators in the first case show gain in efficiency, the second type of estimators are more robust to the form of the truncation distribution. The rates of the change-point estimators are of order $\mathcal{O}_p(1/n)$, when $K_*(t) = \mathrm{I}_{[-1,0]}(t) - \mathrm{I}_{(0,1]}(t)$. We further study the asymptotic distributions, in both cases, and devise interval-estimators for the location and size of the change, paving the way towards making statistical inference about whether or not a change-point exists.

The rest of this manuscript is organized as follows. Section 2 includes some preliminary notations and definitions needed in the sequel. The estimation of the location and size of the change-point is discussed in Section 3. In Section 4, we study the asymptotic properties of these estimators. The finite sample behavior of the estimators is studied in Section 5 by means of simulations. In Section 6, we illustrate our methods on a set of survival data collected on elderly Canadian citizens (aged 65+) suffering from dementia. Section 7 includes the proofs of the main theorems and the statements of several lemmas, while the proofs of the latter are given in a supplementary material file [21].

2. Preliminaries

Let X^* (variable of interest) and T^* be two independent nonnegative random variables representing, respectively, the lifetime and the truncation time from the population. In a cross-sectional study, the right-censored prevalent cohort data consist of nonnegative random triples $\{(T_i, Y_i, \delta_i), i = 1, \ldots, n\}$, where $Y = T + \min(R, R_c)$, R and R_c denote the respective residual lifetime and residual censoring-time, and $\delta = I(R \le R_c)$ is the censoring indicator. The pair $(T, X = T + R)$ denote the truncation and lifetime associated with the observed subjects, and arise from the conditional distribution of (T^*, X^*) given that $T^* \le X^*$. Let F, H and G be the respective distribution functions (cdf) of X^*, T^* and R_c. The cdf of the observed lifetime X is

$$F_b(x) = \int_0^x \frac{H(u)}{\int_0^\infty H(u)\, dF(u)}\, dF(u).$$

In many applications, it is reasonable to assume that R_c is independent of (T, X). Throughout this article, we adopt this assumption and assume that F, G and H, with density h, are all continuous functions. In the sequel, we denote by $f^{(i)}(x)$ the ith derivative of a function $f(x)$, and by l_f and u_f the lower and upper bounds of the support of f. Let Q be the distribution of the residual lifetime R. The following assumptions are needed for establishing the asymptotic properties in Section 4:

$$\text{B1:} \quad \begin{cases} \text{(a)} & \text{The density } h \text{ is bounded in its support } [l_H, u_H], \\ \text{(b)} & l_G \wedge l_Q > 0 \text{ and } l_H > 0, \\ \text{(c)} & u_G \le u_Q \text{ with } G(u_G) < 1. \end{cases}$$

The first part of assumption (b) essentially means that there is no immediate failure or censoring at the beginning of the study, while the second part of the assumption means that all subjects recruited to the study are prevalent cases. This condition reflects the setting of the CSHA and in general is reasonable in most prevalent cohort studies (see [2]). Assumption (c) means that the lifetimes of some individuals, who are still alive at the end of the study, will be censored. This is common in the follow-up studies and is due to the limited time of the follow-up.

3. Estimators

Given an appropriate compactly supported odd-function K_*, and a sequence $h_n > 0$ tending to zero, the proposed estimators for the location/size of the change-point, τ and $\Delta(\tau)$, are based on estimating the maximum modulus of the convolution of $K_*(\cdot/h_n)/h_n$ with the hazard function λ. Depending on the information available on the left-truncation d.f. H, we define two types of estimators for τ and

$$\Delta(t) = \lambda\big(t^+\big) - \lambda\big(t^-\big),$$

when H is known and when H is unknown. In the sequel, we assume that the functions K_+, K_-, λ and Λ satisfy the two following sets of conditions:

$$\text{K1:}\quad \begin{cases} \text{(a)} & K_+ \in \mathcal{C}^1\big([-1,0]\big) \text{ with support } (-1,0], \\ \text{(b)} & K_- \in \mathcal{C}^1\big([0,1]\big) \text{ with support } [0,1), \\ \text{(c)} & \displaystyle\int_{-1}^{0} K_+(t)\,dt = \int_0^1 K_-(t)\,dt = 1. \end{cases}$$

$$\text{B2:}\quad \begin{cases} \text{(a)} & \text{The cumulative hazard } \Lambda \text{ is continuous,} \\ \text{(b)} & \text{The hazard function } \lambda \text{ has one discontinuity } \tau, \\ \text{(c)} & \lambda \in \mathcal{C}^1\big([0,\tau]\big) \cap \mathcal{C}^1\big([\tau,\infty]\big). \end{cases}$$

Condition B2(c) means that λ is continuously differentiable except at τ, where the right and the left derivatives of the hazard, $\lambda^{(1)}(\tau^+) = \lim_{x>\tau, x\to\tau} \lambda^{(1)}(x)$ and $\lambda^{(1)}(\tau^-) = \lim_{x<\tau, x\to\tau} \lambda^{(1)}(x)$, exist. Let l_L and u_L denote the lower/upper bounds of the support of $L(y) = \Pr[Y \le y]$, and $U < u_L$ be a positive constant. We note that all functions considered in this manuscript are defined on the set $\mathcal{A} = [l_L, U]$. However, it should be noted that there is no mass below l_L. Note also that the hazard function λ can be decomposed into a smooth part λ_0 and a discontinuous part:

$$\lambda(x) = \lambda_0(x) + \Delta(\tau)\mathrm{I}_{[\tau,U]}(x).$$

Let $\widehat{G}$ be the product-limit estimator of G, $\widehat{W}_1(t,s) = \int_0^s [1 - \widehat{G}(t-u)]\,dH(u)$, $\widehat{W}_0(t,s) = \int_{t-s}^t h(t-u)\,d\widehat{G}(u)$ and

$$\widehat{C}^*(s) = \frac{1}{n}\sum_{i=1}^n \mathrm{I}(Y_i \ge s)\left\{\delta_i \frac{\widehat{W}_1(Y_i,s)}{\widehat{W}_1(Y_i,Y_i)} + (1-\delta_i)\frac{\widehat{W}_0(Y_i,s)}{\widehat{W}_0(Y_i,Y_i)}\right\}. \tag{3.1}$$

In the following definition, we introduce the estimators of the change-point τ and the jump-size Δ, when the truncation distribution H is assumed to be known.

Definition 3.1. *Suppose that H is known, $x \in \mathcal{A}$ and let*

$$K_* = K_+ - K_-. \tag{3.2}$$

The proposed estimators of the jump-size Δ and the change-point τ are:

$$\widehat{\Delta}_1(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K_*\left(\frac{x - Y_i}{h_n}\right) \frac{\delta_i}{\widehat{C}^*(Y_i)}, \tag{3.3}$$

$$\widehat{\tau}_1 = \arg\max_{x \in \mathcal{A}} \left|\widehat{\Delta}_1(x)\right|. \tag{3.4}$$

The estimator $\widehat{\Delta}_1$ is equivalently equal to the difference between the estimators of $\lambda(x^+)$ and $\lambda(x^-)$, defined by

$$\widehat{\lambda}_{1,\pm}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K_\pm\left(\frac{x - Y_i}{h_n}\right) \frac{\delta_i}{\widehat{C}^*(Y_i)}. \tag{3.5}$$

Note that $\widehat{C}^*(x)$ ([14]) is an estimator of $\Pr[T \le x \le Y]$, the proportion of recruited subjects who are at risk to fail at time x. This estimate, essentially, amounts to adapt the at-risk process $\sum_{i=1}^{n} \mathrm{I}(T_i \le s \le Y_i)/n$ to the case where we have information about the truncation distribution H.

The estimation of $\Delta(x)$ and τ is considered over the set $\mathcal{A}$, where we assume that $\tau \in (l_L, U]$. The asymptotic results established in Section 4 are valid for any constant U that fulfills the condition $U < u_L$. In real applications, however, we need $U < Y_{(n)}$, where $Y_{(n)}$ is the maximum observation. The latter estimators are defined when the left-truncation d.f. H can be recovered and parameterized, which is possible in some real data situations (see Figure 4(b) in Section 6). When the truncation d.f. H seems to be difficult to parameterize and is left unspecified, we propose a second type of estimator for Δ and τ in the next definition.

Definition 3.2. *Let $x \in \mathcal{A}$ and $\widehat{C}(x) = \sum_{i=1}^{n} \mathrm{I}(T_i \le x \le Y_i)/n$. The estimators of Δ and τ are respectively*:

$$\widehat{\Delta}_2(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K_*\left(\frac{x - Y_i}{h_n}\right) \frac{\delta_i}{\widehat{C}(Y_i)}, \tag{3.6}$$

$$\widehat{\tau}_2 = \arg\max_{x \in \mathcal{A}} \left|\widehat{\Delta}_2(x)\right|. \tag{3.7}$$

The estimators of the jump size/location in (3.6) and (3.7) generalize those of [16] to the case where $T \neq 0$. Unlike the estimators introduced in Definition 3.1, $\widehat{\Delta}_2$ and $\widehat{\tau}_2$ make no assumption about the possible form of the truncation d.f. H. In this sense, the latter estimators are more robust than their predecessors in (3.3) and (3.4). However, as shown in Remark 4.1 below, $\widehat{\Delta}_1$ which incorporates the available information about H is more efficient that $\widehat{\Delta}_2$.

4. Asymptotic properties

4.1. Rate of the change-point estimator

In the following result, we establish the convergence rate $\mathcal{O}_p(1/n)$ for the change-point estimators $\widehat{\tau}_1$ and $\widehat{\tau}_2$.

Theorem 4.1. *Let $K_*(t) = \mathrm{I}_{[-1,0]}(t) - \mathrm{I}_{(0,1]}(t)$. Suppose that* B1, B2 *and* K1 *hold, the distributions of Y and T are Lipschitz functions and nh_n^2, $(\log n)^2/nh_n \to 0$, as $n \to \infty$ and $h_n \to 0$. Then*

$$|\widehat{\tau}_\ell - \tau| = \mathcal{O}_p(1/n) \qquad (\ell = 1, 2). \tag{4.1}$$

The proof is given in Section 7. We can replace the condition that the d.f. of Y and T are Lipschitz functions by $c(x) = \Pr[T \le x \le Y]$ is a Lipschitz function, which is what we need in the proof. Note that in Section 4.2 and Section 4.3, the rates of $\widehat{\tau}_1$ and $\widehat{\tau}_2$ are of order $\mathcal{O}_p(\log^{\frac{5}{3}}(n)/n)$, which is a little slower than the rate in (4.1). This is essentially caused by the assumption $K_\pm^{(i)}(0) = 0$ $(i = 0, 1, 2)$ in (K2), implying that $K_*(0) = 0$. This condition is needed, in the method followed in Section 4.2 and Section 4.3, for establishing the asymptotic distributions of $\widehat{\tau}_1$ and $\widehat{\tau}_2$ and deriving confidence intervals for τ.

4.2. Asymptotic distributions: *H* is known

An important question that arises in real data applications, after locating a possible change-point τ, is whether such a point is the location of a genuine change within the hazard function. Unlike the continuous part of λ, this particular point is characterized by a nonzero jump size. Our first result establishes the asymptotic normality of $\widehat{\Delta}_1(x)$, for each $x \in \mathcal{A}$.

Theorem 4.2. *Let $x \in \mathcal{A}$. Suppose that assumptions* B1, B2 *and* K1 *hold, and $nh_n^3 \to 0$ and $(\log n)^2/nh_n \to 0$, as $n \to \infty$ and $h_n \to 0$. Then*

$$\sqrt{nh_n}\big[\widehat{\Delta}_1(x) - \Delta(x)\big] \xrightarrow{d} N\left(0, \frac{\lambda(x^-)}{c(x)} \int_0^1 \big[K_-(t)\big]^2 dt + \frac{\lambda(x^+)}{c(x)} \int_{-1}^0 \big[K_+(t)\big]^2 dt\right),$$

where $c(x) = \Pr(T \le x \le Y)$, representing the proportion of recruited subjects who are at risk to fail at time x.

The proof of Theorem 4.2 is given in Section 7. To establish the limit distributions of the estimators $\widehat{\tau}_1$ and $\widehat{\Delta}_1(\widehat{\tau}_1)$, we consider the two sequences of processes

$$\varphi_n(x) = \alpha_n\big[\widehat{\Delta}_1(\tau + x\theta_n h_n) - \widehat{\Delta}_1(\tau)\big] \qquad \text{if } \Delta(\tau) > 0, \tag{4.2}$$

$$\phi_n(x) = -\varphi_n(x) \qquad \text{if } \Delta(\tau) < 0, \tag{4.3}$$

on $[-M, M]$, where M is a positive constant and

$$\alpha_n = (nh_n)^{2/3}, \qquad \theta_n = (nh_n)^{-1/6}. \tag{4.4}$$

The local deviation, $\tau + x\theta_n h_n$, from the change-point τ in (4.2) was considered, in different contexts, by [6], [15] and [16] in order to derive the respective limiting distributions of the estimators of the density mode, the change-point in the regression setting and the hazard discontinuity, when the lifetime data is only subject to right censoring. In what follows, we require that the one-sided kernel functions K_+ and K_- satisfy the additional set of conditions

$$\text{K2:} \quad \begin{cases} \text{(a)} & K_+ \in \mathcal{C}^3\big([-1,0]\big) \text{ and } K_- \in \mathcal{C}^3\big([0,1]\big), \\ \text{(b)} & K_+ \text{ and } K_- \text{ are Lipschitz functions}, \\ \text{(c)} & K_-^{(3)}(0) > 0 \text{ and } K_\pm^{(i)}(0) = 0 \text{ for } i = 0, 1, 2, \end{cases}$$

and assume that when $n \to \infty$ and $h_n \to 0$

$$nh_n^3 \to 0, \qquad (\log n)^3/(nh_n) \to 0. \tag{4.5}$$

In the following result, the sequence φ_n converges weakly to a random parabola with a unique maximum. This leads to the derivation of the asymptotic distributions of the estimators $\widehat{\tau}_1$ and $\widehat{\Delta}_1(\widehat{\tau}_1)$.

Theorem 4.3. *Suppose assumptions* B1–B2, K1–K2 *and* (4.5) *hold. The sequence $\varphi_n(x)$ converges in distribution to the Gaussian process*

$$\varphi(x) = -\frac{\Delta(\tau) K_-^{(3)}(0)}{4!} x^4 + xZ, \tag{4.6}$$

where $Z \sim N(0, \frac{\lambda(\tau^-)}{c(\tau)} \int_0^1 [K_-^{(1)}(t)]^2\, dt + \frac{\lambda(\tau^+)}{c(\tau)} \int_{-1}^0 [K_+^{(1)}(t)]^2\, dt)$.

The proof is given in Section 7. Having established this result, the asymptotic distribution of $\widehat{\tau}_1$ and $\widehat{\Delta}_1(\widehat{\tau}_1)$ can be derived as follows. First, notice that the limiting process $\varphi(x)$ has a unique maximum at $V = [Z \times 6/(\Delta(\tau) K_-^{(3)}(0))]^{1/3}$, since $\Delta(\tau), K_-^{(3)}(0) > 0$, and by Theorem 4.3

$$\max_{x \in [-M,M]} \varphi_n(x) \xrightarrow{d} \max_{x \in [-M,M]} \varphi(x), \tag{4.7}$$

$$\operatorname*{arg\,max}_{x \in [-M,M]} \varphi_n(x) \xrightarrow{d} \operatorname*{arg\,max}_{x \in [-M,M]} \varphi(x). \tag{4.8}$$

Note that $\widehat{\tau}_1 = \arg\max_{x \in [-M,M]} |\widehat{\Delta}_1(\tau + x\theta_n h_n)|$, which follows from Lemma 7.11, and if $V_n = \arg\max_{x \in [-M,M]} \varphi_n(x)$ then

$$\widehat{\tau}_1 = \tau + V_n \theta_n h_n. \tag{4.9}$$

The result (4.7) is equivalent to $\varphi_n(V_n) \xrightarrow{d} \varphi(V)$, and since $\theta_n \varphi(V) \xrightarrow{\Pr} 0$ as $\theta_n \to 0$, then

$$\theta_n \varphi_n(V_n) = \sqrt{nh_n}\big[\widehat{\Delta}_1(\widehat{\tau}_1) - \widehat{\Delta}_1(\tau)\big] \xrightarrow{\Pr} 0. \tag{4.10}$$

The asymptotic distribution of $\widehat{\Delta}_1(\widehat{\tau}_1)$ can then be obtained from (4.10) and Theorem 4.2 in the next corollary.

Corollary 4.1. *Suppose the assumptions of Theorem 4.3 hold. We have*

$$\sqrt{nh_n}\big[\widehat{\Delta}_1(\widehat{\tau}_1) - \Delta(\tau)\big] \xrightarrow{d} N\left(0, \frac{\lambda(\tau^-)}{c(\tau)} \int_0^1 \big[K_-(t)\big]^2 dt + \frac{\lambda(\tau^+)}{c(\tau)} \int_{-1}^0 \big[K_+(t)\big]^2 dt\right).$$

On the other hand, note that (4.8) is equivalent to $V_n \xrightarrow{d} V$, and together with (4.9), we establish the limiting distribution of the change-point estimator $\widehat{\tau}_1$ in the next result.

Corollary 4.2. *Under the assumptions of Theorem 4.3*

$$\frac{\widehat{\tau}_1 - \tau}{\theta_n h_n} \xrightarrow{d} \left[Z \times \frac{6}{\Delta(\tau) K_-^{(3)}(0)}\right]^{1/3},$$

where Z is the normal random variable defined in Theorem 4.3.

For the case $\Delta(\tau) < 0$, the sequence of processes $\phi_n(x)$ converges weakly to the Gaussian process $\phi(x) = -\varphi(x)$. The limiting process $\phi(x)$ has a unique maximum at V, and we obtain the same asymptotic distributions for $\widehat{\tau}_1$ and $\widehat{\Delta}_1(\widehat{\tau}_1)$ as in Corollaries 4.1–4.2.

We can use the asymptotic results in Corollaries 4.1–4.2 to derive interval estimates for τ and $\Delta(\tau)$, when the truncation d.f. H is known. To do so, we first estimate $\lambda(\tau^\pm)$ by $\widehat{\lambda}_{1,\pm}(\widehat{\tau}_1)$ and $c(\tau)$ by $\widehat{C}^*(\widehat{\tau}_1)$. Note that under B1–B2, K1–K2 and (4.5), one can show that

$$\widehat{\lambda}_{1,\pm}(\widehat{\tau}_1) \xrightarrow{\Pr} \lambda\big(\tau^\pm\big),$$

by using Lemma 7.11 and showing, analogously to Lemma 7.4(ii), that $\sup_{x\in\mathcal{A}} E[(\widehat{\lambda}_{1,\pm}(x) - \widetilde{\lambda}_\pm(x))^2] = o(1)$, where $\widetilde{\lambda}_\pm(x) = \int h_n^{-1} K_\pm[(x-t)/h_n]\, d\Lambda(t)$. Also, we have that

$$\widehat{C}^*(\widehat{\tau}_1) \xrightarrow{\Pr} c(\tau),$$

which follows from the uniform convergence result (A.31), in the supplementary material file, and the continuity of the distributions F, H and G. We then estimate the variance of Z and the asymptotic variance in Corollary 4.1, respectively, by

$$\widehat{\sigma}_1^2 = \frac{\widehat{\lambda}_{1,-}(\widehat{\tau}_1)}{\widehat{C}^*(\widehat{\tau}_1)} \int_0^1 \big[K_-^{(1)}(t)\big]^2 dt + \frac{\widehat{\lambda}_{1,+}(\widehat{\tau}_1)}{\widehat{C}^*(\widehat{\tau}_1)} \int_{-1}^0 \big[K_+^{(1)}(t)\big]^2 dt,$$

$$\widehat{\sigma}_2^2 = \frac{\widehat{\lambda}_{1,-}(\widehat{\tau}_1)}{\widehat{C}^*(\widehat{\tau}_1)} \int_0^1 \big[K_-(t)\big]^2 dt + \frac{\widehat{\lambda}_{1,+}(\widehat{\tau}_1)}{\widehat{C}^*(\widehat{\tau}_1)} \int_{-1}^0 \big[K_+(t)\big]^2 dt.$$

The interval estimators of $\Delta(\tau)$ and τ, with confidence level $1-\alpha$, are respectively,

$$\left[\widehat{\Delta}_1(\widehat{\tau}_1) - \frac{\widehat{\sigma}_2}{\sqrt{nh_n}} z_{1-\alpha}, \widehat{\Delta}_1(\widehat{\tau}_1) + \frac{\widehat{\sigma}_2}{\sqrt{nh_n}} z_{1-\alpha}\right], \tag{4.11}$$

$$\left[\widehat{\tau}_1 - \left(\frac{6\times(\theta_n h_n)^3 \widehat{\sigma}_1}{\widehat{\Delta}_1(\widehat{\tau}_1) K_-^{(3)}(0)} z_{1-\alpha}\right)^{1/3}, \widehat{\tau}_1 + \left(\frac{6\times(\theta_n h_n)^3 \widehat{\sigma}_1}{\widehat{\Delta}_1(\widehat{\tau}_1) K_-^{(3)}(0)} z_{1-\alpha}\right)^{1/3}\right], \tag{4.12}$$

where $z_{1-\alpha} = \Phi^{-1}(1-\alpha)$ and Φ is the cdf of a standard normal distribution.

4.3. Asymptotic distributions: *H* is unknown

The next result presents limiting distributions of $\widehat{\Delta}_2(x)$, $\widehat{\tau}_2$ and $\widehat{\Delta}_2(\widehat{\tau}_2)$, when the truncation distribution H is unknown. These results can be derived using the representation (7.25) of $\widehat{\Delta}_2(x)$, in Lemma 7.12, and following similar steps to those in Section 4.2 and Lemmas 7.2–7.11.

Theorem 4.4. 1. *Let $x \in \mathcal{A}$. Suppose that assumptions* B1, B2 *and* K1 *hold, and nh_n^3, $(\log n)^2/nh_n \to 0$, as $n \to \infty$ and $h_n \to 0$, then,*

$$\sqrt{nh_n}\big[\widehat{\Delta}_2(x) - \Delta(x)\big] \xrightarrow{d} N\left(0, \frac{\lambda(x^-)}{c(x)} \int_0^1 \big[K_-(t)\big]^2 dt + \frac{\lambda(x^+)}{c(x)} \int_{-1}^0 \big[K_+(t)\big]^2 dt\right).$$

2. *Under assumptions* B1–B2, K1–K2 *and* (4.5), *we have*

(i) $$\sqrt{nh_n}\big[\widehat{\Delta}_2(\widehat{\tau}_2) - \Delta(\tau)\big] \xrightarrow{d} N\left(0, \frac{\lambda(\tau^-)}{c(\tau)} \int_0^1 \big[K_-(t)\big]^2 dt + \frac{\lambda(\tau^+)}{c(\tau)} \int_{-1}^0 \big[K_+(t)\big]^2 dt\right),$$

(ii) $$\frac{\widehat{\tau}_2 - \tau}{\theta_n h_n} \xrightarrow{d} \left[Z \times \frac{6}{\Delta(\tau) K_-^{(3)}(0)}\right]^{1/3},$$

where Z is the normal variable defined in Theorem 4.3 and θ_n is given in (4.4).

The above results generalize those of [16] to left-truncated and right-censored data. When $T = 0$, the asymptotic results in Theorem 4.4 coincide with the results established by [16] for simple right-censored observations.

Remark 4.1. Notice that the limiting distributions in Theorem 4.4 and those in Theorem 4.2 and Corollaries 4.1–4.2 are the same. However, if h_n tends to zero slower than $\log n/\sqrt{n}$, which is the case in our simulations and data application, there is an efficiency gain in using $\widehat{\Delta}_1$. In fact, under the assumption $(\log n)^2/nh_n^2 \to 0$ as $n \to \infty$, we have

$$(nh_n) \times \big[\mathrm{var}\big(\widehat{\Delta}_2(x)\big) - \mathrm{var}\big(\widehat{\Delta}_1(x)\big)\big] = \kappa(x) h_n + o(h_n), \tag{4.13}$$

where κ is positive and equal to

$$\kappa(x) = \left(f_2\big(x^-, x^-\big) + \int_0^{u_G^-} f_3\big(x^-, x^-, t\big)\, dL_0^G(t) \right) \times \int_0^1 \big[K_-(z)\big]^2\, dz$$
$$+ \left(f_2\big(x^+, x^+\big) + \int_0^{u_G^-} f_3\big(x^+, x^+, t\big)\, dL_0^G(t) \right) \times \int_{-1}^0 \big[K_+(z)\big]^2\, dz.$$

The functions L_0^G, f_2 and f_3 are respectively, defined in Section 7.1 and in the proof of Lemma 7.4 given in the supplementary material file [21]. The result in (4.13) can be obtained using the equations (A.1), (A.4), (A.5), (A.7), (A.8) in the technical appendix (see the supplementary material file [21]) and the representation (7.24) (Lemma 7.12) in Section 7.

Based on the asymptotic results in Theorem 4.4, we may define interval estimators for τ and $\Delta(\tau)$, when H is unknown. Define

$$\widehat{\lambda}_{2,\pm}(x) = \frac{1}{nh_n} \sum_{i=1}^n K_\pm\left(\frac{x - Y_i}{h_n}\right) \frac{\delta_i}{\widehat{C}(Y_i)},$$

which represent one-sided estimators of $\lambda(x^\pm)$. One can show that $\widehat{\lambda}_{2,\pm}(\widehat{\tau}_2) \xrightarrow{\text{Pr}} \lambda(\tau^\pm)$ by using Theorem 4.4(2)(ii) and showing, analogously to Lemma 7.4, that $\sup_{x\in\mathcal{A}} E[(\widehat{\lambda}_{2,\pm}(x) - \widetilde{\lambda}_\pm(x))^2] = o(1)$, with $\widetilde{\lambda}_\pm(x) = \int h_n^{-1} K_\pm[(x-t)/h_n]\, d\Lambda(t)$. We also have $\widehat{C}(x) \xrightarrow{\text{a.s.}} c(x)$ uniformly in $x \in \mathcal{A}$, using the Law of the Iterated Logarithm result in [8] (Lemma 2). This leads to $\widehat{C}(\widehat{\tau}_2) \xrightarrow{\text{Pr}} c(\tau)$ by employing Theorem 4.4(2)(ii) and the continuity of F, H and G. The variance of Z and the asymptotic variance in Theorem 4.4(i) can then be estimated respectively, by $\widehat{\sigma}_3^2$ and $\widehat{\sigma}_4^2$, where $\lambda(\tau^\pm)$ and $c(\tau)$ are replaced by $\widehat{\lambda}_{2,\pm}(\widehat{\tau}_2)$ and $\widehat{C}(\widehat{\tau}_2)$, respectively. Hence, with confidence level $1-\alpha$, the interval estimator of $\Delta(\tau)$ is

$$\left[\widehat{\Delta}_2(\widehat{\tau}_2) - \frac{\widehat{\sigma}_4}{\sqrt{nh_n}} z_{1-\alpha}, \widehat{\Delta}_2(\widehat{\tau}_2) + \frac{\widehat{\sigma}_4}{\sqrt{nh_n}} z_{1-\alpha}\right] \tag{4.14}$$

and that of τ is

$$\left[\widehat{\tau}_2 - \left(\frac{6 \times (\theta_n h_n)^3 \widehat{\sigma}_3}{\widehat{\Delta}_2(\widehat{\tau}_2) K_-^{(3)}(0)} z_{1-\alpha}\right)^{1/3}, \widehat{\tau}_2 + \left(\frac{6 \times (\theta_n h_n)^3 \widehat{\sigma}_3}{\widehat{\Delta}_2(\widehat{\tau}_2) K_-^{(3)}(0)} z_{1-\alpha}\right)^{1/3}\right], \tag{4.15}$$

where $z_{1-\alpha} = \Phi^{-1}(1-\alpha)$.

5. Simulation study

A simulation study was carried out to explore the moderate sample size performances of the estimators $\widehat{\Delta}_\ell$ and $\widehat{\tau}_\ell$ ($\ell = 1, 2$) in Definitions 3.1–3.2.

Example 1. The following Weibull model is considered for the hazard function

$$\lambda(x) = \frac{a_1}{b_1^{a_1}} x^{a_1-1} \mathrm{I}_{(0,\tau]}(x) + \frac{a_2}{b_2^{a_2}} x^{a_2-1} \mathrm{I}_{(\tau,\infty)}(x), \tag{5.1}$$

where λ has a jump at τ. We generate independently the lifetime data x_i^* $(i = 1, 2, \dots)$ from the aforementioned model and the truncation data t_i^* $(i = 1, 2, \dots)$ from a uniform d.f. H. Truncate (exclude) the couple (x_i^*, t_i^*) if $x_i^* < t_i^*$, otherwise keep generating the data until $n = 800$ observations are collected. The resulting sample is $\{(x_i, t_i), i = 1, \dots, n\}$, and define $r_i = x_i - t_i$ $(i = 1, \dots, n)$. We then independently generate n residual censoring data $r_{c,1}, \dots, r_{c,n}$ from a gamma d.f. G. This allows us to define the data $\{(t_i, y_i, \delta_i), i = 1, \dots, n\}$, where $y_i = t_i + \min(r_i, r_{c,i})$ and $\delta_i = \mathrm{I}(r_i \le r_{c,i})$.

Two cases were chosen for the jump location/size $(\tau, \Delta(\tau))$. The first is $(\tau = 1, \Delta(\tau) = 1)$, with model parameters $a_1 = 1.1$, $b_1 = 3$, $a_2 = 2$ and $b_2 = 1.23$; $H = \mathrm{Unif}(0, 2)$ and $G = \Gamma(6.5, 4)$. The second case is $(\tau = 2, \Delta(\tau) = 0.25)$, and the model parameters are: $a_1 = 2$, $b_1 = 4$, $a_2 = 1.2$ and $b_2 = 2.32$, with $H = \mathrm{Unif}(0, 5)$ and $G = \Gamma(17, 4)$. The censoring proportion corresponding to the gamma distributed residual-censoring is approximately 23%, for both cases. We generated 400 replications from the above models.

Note that, in these simulations, we mimic the conditions of the real data example in Section 6 (CSHA-1). As reported in [3], the Weibull distribution seems to fit the lifetimes in the CSHA-1 data, and as shown in Figure 4(b), the truncation times are reasonably uniform. Also, the real data in Section 6 is of size $n = 807$ and the censoring level observed is 22.3%.

One of the issues of kernel estimation is the choice of the bandwidth h_n. In this study, our choice is based on the almost-sure convergence rate of $\widehat{\Delta}_\ell$ $(\ell = 1, 2)$,

$$\widehat{\Delta}_\ell(x) - \Delta(x) = \mathcal{O}\Big(h_n + \sqrt{\log\log n / n h_n^2}\Big), \tag{5.2}$$

which can be obtained from the uniform result (7.21) in Section 7 (case $\ell = 1$) and Corollary 1(b) in [10] (case $\ell = 2$). The optimal rate in (5.2) corresponds to a bandwidth h_n of order $[\log(\log n)/n]^{1/4}$. The window $h_n = [\log(\log n)/n]^{1/4}$ was chosen for the estimation of Δ in the first case. We notice that under the first setting, the interval $(0, 3]$ covers approximately 95% of the lifetime data, while in the second case the range is approximately $(0, 6]$, almost twice the first range. Given that $h_n = [\log(\log n)/n]^{1/4}$ showed satisfactory performance in the estimation of Δ in the first case, we chose a bandwidth $h_n = 2 \times [\log(\log n)/n]^{1/4}$ for the second one. The kernel functions used in (3.3) and (3.6) are $K_-(x) = 20x^3(1 - x)\mathrm{I}_{[0,1]}(x)$ and $K_+(x) = -20x^3(1 + x)\mathrm{I}_{[-1,0]}(x)$.

It should be noted that the censoring level increases near the right endpoints of the ranges $(0, 3]$ and $(0, 6]$, creating artifact jumps. For this reason, we narrowed our investigation of a jump to the range $\mathcal{A} = [l_L, U]$, where U is chosen to be the 90% percentile of the simulated lifetimes, which approximately corresponds to the value $U = 1.6$ for the first case and $U = 4$ for the second case.

According to the simulations results depicted in Tables 1–2 and Figure 1(a)–(d), the change-point estimators $\widehat{\tau}_1$ and $\widehat{\tau}_2$ show satisfactory performances for locating the true change-point τ in both cases. We notice, in Tables 1–2, that $\widehat{\tau}_1$ and $\widehat{\tau}_2$ are much closer to their target τ as compared

Table 1. Summary statistics for the estimation of the location/size of the hazard change-point over 400 simulations; $\tau = 1$ and $\Delta(\tau) = 1$

Estimator	Mean	Median	5-percentile	90-percentile
$\widehat{\tau}_1$	1.03	1.01	0.99	1.05
$\widehat{\Delta}_1(\widehat{\tau}_1)$	1.20	1.20	1.00	1.36
$\widehat{\tau}_2$	1.03	1.02	0.99	1.05
$\widehat{\Delta}_2(\widehat{\tau}_2)$	1.20	1.21	0.98	1.36

to the estimated jump-sizes $\widehat{\Delta}_1(\widehat{\tau}_1)$ and $\widehat{\Delta}_2(\widehat{\tau}_2)$. This can be explained by the rate of change-point estimators, $\mathcal{O}_p(\log^{\frac{5}{3}}(n)/n)$, which is faster than the rate of their estimated sizes, $\mathcal{O}_p(1/(nh_n)^{\frac{1}{2}})$. Figure 1(e), (f), which display the average of the simulations of $n[\widehat{\mathrm{var}}(\widehat{\Delta}_2(x)) - \widehat{\mathrm{var}}(\widehat{\Delta}_1(x))]$ indicate generally smaller variance for the first estimator and reflect the efficiency gain in using $\widehat{\Delta}_1$, as proven in Remark 4.1.

Example 2. In this example, we simulate the estimator of $n[\mathrm{var}(\widehat{\Delta}_2(x)) - \mathrm{var}(\widehat{\Delta}_1(x))]$, in Remark 4.1, to evaluate the efficiency of $\widehat{\Delta}_1(x)$ as compared with $\widehat{\Delta}_2(x)$. We consider two models for the hazard function with change-point; the Gompertz model,

$$\lambda(x) = a_1 b_1 \mathrm{e}^{b_1 x} \mathrm{I}_{(0,\tau]}(x) + a_2 b_2 \mathrm{e}^{b_2 x} \mathrm{I}_{(\tau,\infty)}(x), \tag{5.3}$$

and the Log-logistic model, given by,

$$\lambda(x) = \frac{b_1 x^{b_1 - 1}}{a_1^{b_1} + x^{b_1}} \mathrm{I}_{(0,\tau]}(x) + \frac{b_2 x^{b_2 - 1}}{a_2^{b_2} + x^{b_2}} \mathrm{I}_{(\tau,\infty)}(x). \tag{5.4}$$

Two cases were chosen for each of these two models. In the Gompertz model, the first case is ($\tau = 2$, $\Delta(\tau) = 0.5$) and the parameters are ($a_1 = 0.025$, $b_1 = 1.2$, $a_2 = 6.801$, $b_2 = 0.1$), with $H = \mathrm{Unif}(0, 4)$ and $G = \Gamma(11.5, 4)$ (see Figure 2(a)). The second is ($\tau = 2$, $\Delta(\tau) = -0.75$) with parameters ($a_1 = 0.025$, $b_1 = 1.5$, $a_2 = 0.0023$, $b_2 = 0.5$), where $H = \mathrm{Unif}(0, 13)$ and $G = \Gamma(35, 4)$ (Figure 2(b)). In the Log-logistic model, the first example is ($\tau = 1.5$, $\Delta(\tau) = -0.5$) and the parameters are ($a_1 = 1$, $b_1 = 1.3$, $a_2 = 8.03$, $b_2 = 2$), with $H = \mathrm{Unif}(0, 8)$ and $G = \Gamma(56, 4)$ (Figure 2(c)). The second example is ($\tau = 2$, $\Delta(\tau) = 0.75$) with parameters ($a_1 = 2$, $b_1 = 1$, $a_2 = 0.003$, $b_2 = 1$), where $H = \mathrm{Unif}(0, 5)$ and $G = \Gamma(15, 4)$ (Figure 2(d)).

Table 2. Summary statistics for the estimation of the location/size of the hazard change-point over 400 simulations; $\tau = 2$ and $\Delta(\tau) = 0.25$

Estimator	Mean	Median	5-percentile	90-percentile
$\widehat{\tau}_1$	2.03	2.00	1.89	2.11
$\widehat{\Delta}_1(\widehat{\tau}_1)$	0.30	0.30	0.24	0.36
$\widehat{\tau}_2$	2.06	2.00	1.88	2.12
$\widehat{\Delta}_2(\widehat{\tau}_2)$	0.31	0.37	0.24	0.37

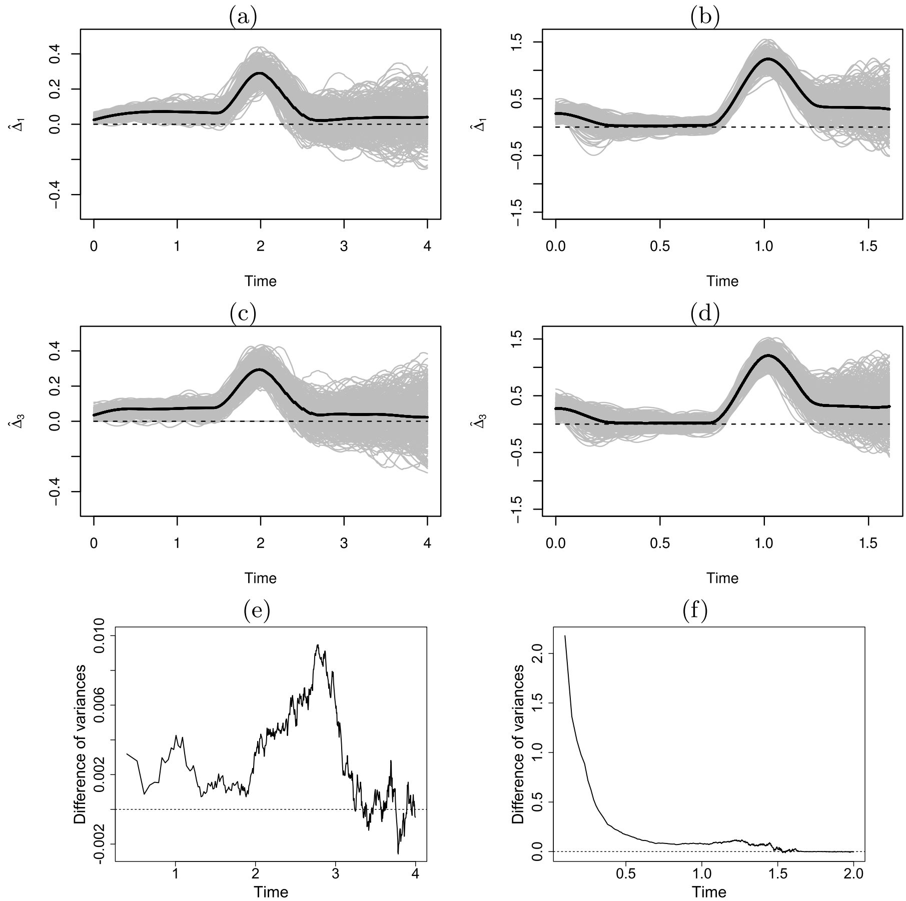

Figure 1. Simulations of the hazard jump-size estimators (a), (b) $\widehat{\Delta}_1$, (c), (d) $\widehat{\Delta}_2$, with the average in solid black. (e), (f) Average of 400 simulations of $n[\widehat{\text{var}}(\widehat{\Delta}_2(x)) - \widehat{\text{var}}(\widehat{\Delta}_1(x))]$. (a), (c), (e) $\tau = 2$ and $\Delta(\tau) = 0.25$, (b), (d), (f) $\tau = 1$ and $\Delta(\tau) = 1$.

The censoring proportion is approximately 21%, in the four cases, and the size of the sample is $n = 800$. Figure 2 displays the curve of the average of 400 simulations of $n[\widehat{\text{var}}(\widehat{\Delta}_2(x)) - \widehat{\text{var}}(\widehat{\Delta}_1(x))]$. The plots in Figure 2(a), (b), (c), (d), of the considered models, indicate that $\widehat{\Delta}_1$ is relatively efficient as compared with $\widehat{\Delta}_2(x)$.

6. Survival with dementia

We apply the methodology described in Sections 3–4 to a set of prevalent cohort survival data collected on elderly Canadians with dementia. In 1991/1992, a nationwide cross-sectional survey was conducted in five regions of Canada among 9008 community-residing persons and 1255

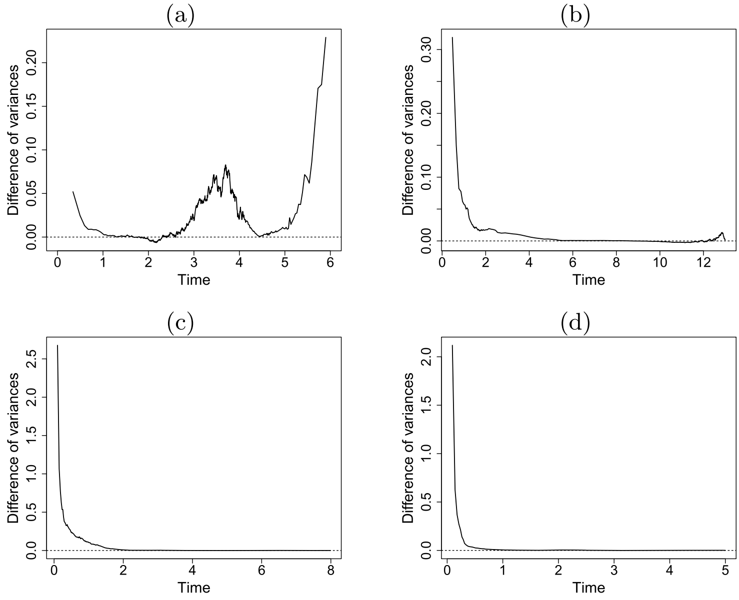

Figure 2. Efficiency comparison: Average of 400 simulations of $n[\widehat{\text{var}}(\widehat{\Delta}_2(x)) - \widehat{\text{var}}(\widehat{\Delta}_1(x))]$ in (a), (b) Gompertz lifetime model, and (c), (d) Log-logistic lifetime model.

institutionalized persons aged 65 and older. The CSHA-1 (Canadian Study of Health and Aging 1) identified 1132 persons with dementia who were followed for a period of 5 years until 1996/1997. The primary purpose of the CSHA-1 was the study of the risk factors for dementia and to determine its prevalence in the Canadian population. Asgharian, M'Lan and Wolfson [2] and Wolfson et al. [26] reported that those patients with missing date of onset or with survival $\geq$ 20 years, who unlikely had dementia, need to be excluded. We then considered a sample of $n = 807$ patients in our statistical analysis, among whom 627 died and 180 were censored during the follow-up. The survival time was defined as the time elapsed from the onset of dementia to death, and the truncation time was the time from disease onset to study recruitment.

The purpose of the present example is to determine whether survival with dementia is a smooth process. This entails that we investigate the possibility of a change-point in the hazard of failure with dementia, and find its location. First, we employed the method suggested in [24] to estimate the truncation d.f. H. Figure 4(b) shows the estimator of H, and indicates that a uniform distribution is a reasonable assumption. The jump-size estimator $\widehat{\Delta}_1$ is then used under the uniformity of H, while in $\widehat{\Delta}_2$ the truncation distribution is not specified. Under the latter assumption, we found that the estimated median survival is 3.95 years by using the NPMLE in [2]. This estimate suggests that half of the patients with dementia died before 4 years.

The choice of the bandwidth, in this application, is based on the asymptotic rate in (5.2). In the simulations, we noticed that the range of 95% of the lifetime data generated from the first Weibull case was approximately (0, 3], and we chose the window $h_n = [\log(\log n)/n]^{1/4}$. In the CSHA-1 data, the range is (0, 14.5] years, that is, almost 5 times the first range. Given that the

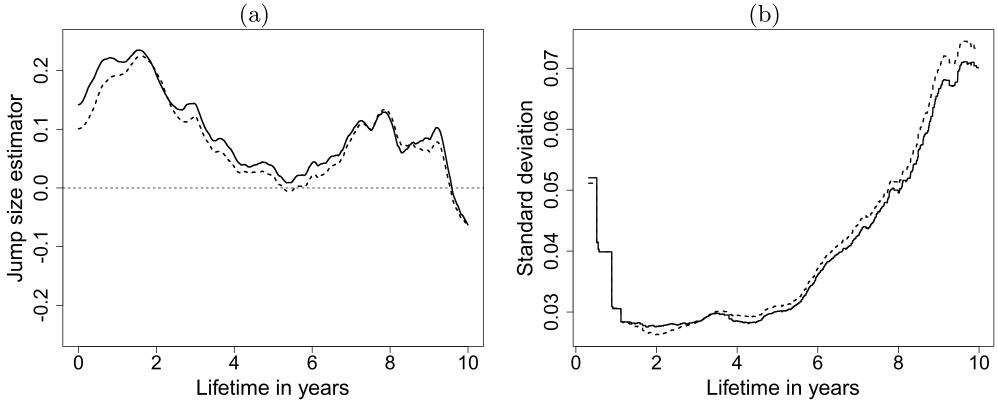

Figure 3. (a) Estimated hazard jump-size of 807 patients with dementia: $\widehat{\Delta}_1$ is the solid line (H is uniform) and $\widehat{\Delta}_2$ is the dashed line (H is unspecified). (b) Estimated standard deviations of $\widehat{\Delta}_1$, in the solid line, and $\widehat{\Delta}_2$, in the dashed line.

Weibull distribution seems to fit the CSHA-1 data, as reported in [3], we chose a bandwidth $h_n = (14.5/3) \times [\log(\log n)/n]^{1/4}$ ($n = 807$) with value $h_n = 1.06$ years. Since the censoring level increases dramatically after 6 years, creating artifact jumps, we narrowed our search of a change-point to the range $\mathcal{A} = [l_L, U] = [0.31, 10]$. The kernel functions used for the estimation of $\lambda(x^-)$, $\lambda(x^+)$ and Δ are $K_-(x) = 20x^3(1-x)\mathrm{I}_{[0,1]}(x)$, $K_+(x) = K_-(-x)$ and $K_* = K_+ - K_-$.

The plot of the jump-size estimator $\widehat{\Delta}_1$ in Figure 3(a), in the solid line, shows a positive maximum $\widehat{\Delta}_1(\widehat{\tau}_1) = 0.238$ around 18–19 months from the onset of dementia. The estimated location of the change point is $\widehat{\tau}_1 = 1.59$. Using the interval estimators in (4.11) and (4.12), when H is assumed uniform, we obtain the 95% confidence intervals $[0.17, 0.29]$ for $\Delta(\tau)$ and $[0.97, 2.12]$ for τ. The estimator $\widehat{\Delta}_2$, in the dashed line, shows a similar behavior to $\widehat{\Delta}_1$ and indicate a positive maximum $\widehat{\Delta}_2(\widehat{\tau}_3) = 0.229$ at $\widehat{\tau}_2 = 1.61$. The truncation distribution H is unspecified in $\widehat{\Delta}_2$ and $\widehat{\tau}_2$. In this case, based on the interval estimators in (4.14) and (4.15), we find the 95% confidence intervals $[0.15, 0.31]$ for $\Delta(\tau)$ and $[1.04, 2.18]$ for τ.

Figure 4(a) displays the plots of the estimators $\widehat{\Delta}_2|T \neq 0$ (solid line), accounting for left-truncation, and $\widehat{\Delta}_2|T = 0$ (dashed line), ignoring the fact that the sampling is biased and the censoring is informative. The curve of the latter estimator (dashed) shows more flatness compared to the curve of the former, since it treats the data as a right-censored random sample from the target population. We can also see that the magnitude of the jumps considerably shrinks when we ignore biased sampling and use $\widehat{\Delta}_2|T = 0$.

In conclusion, both estimation methodologies (for uniform H and unspecified H) are in agreement that an abrupt-change occurs within the hazard function around 18–19 months from the onset of dementia. As indicated by the 95% confidence intervals of the jump location/size obtained above, this change is positive and significant. This implies that the risk of failure, among patients with dementia, experiences a significant sharp-increase between 1 and 2 years from the beginning of the disease. In term of efficiency, the jump-size estimator $\widehat{\Delta}_1$, which incorporates

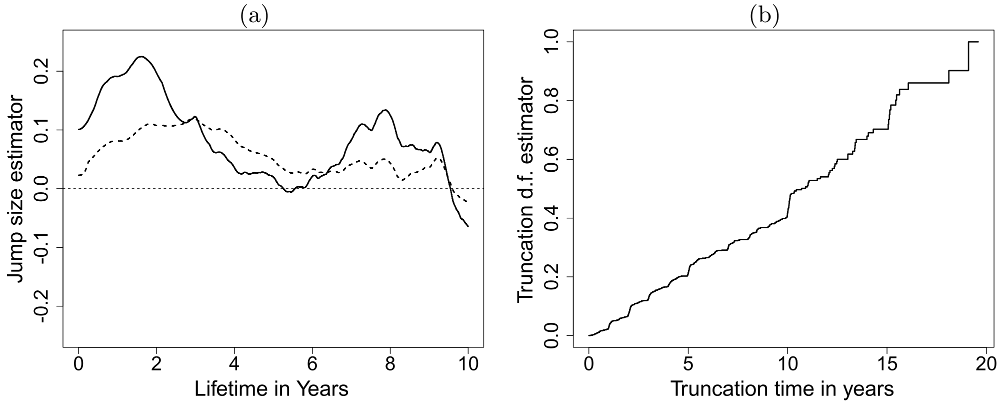

Figure 4. (a) Comparison of the hazard jump-size estimator $\widehat{\Delta}_2|T \neq 0$ (solid line), accounting for truncation, with $\widehat{\Delta}_2|T = 0$ (dashed line), ignoring truncation. (b) Estimator of the truncation distribution H.

the available uniform information of H, shows gain in efficiency in most of the search range $[0.31, 10]$ of the change-point, as shown in Figure 3(b), and reflects the finding in Remark 4.1.

7. Lemmas and proofs

7.1. Notations

We start with some notation from [14] needed to study the asymptotic properties of the estimators (3.3) and (3.4). Define the processes $\widehat{C}(s) = n^{-1}\sum_{i=1}^n C_i(s) = n^{-1}\sum_{i=1}^n \mathrm{I}(T_i \le s \le Y_i)$ and $\widetilde{C}^*(s) = n^{-1}\sum_{i=1}^n C_i^*(s) = n^{-1}\sum_{i=1}^n E[C_i(s)|(Y_i, \delta_i, T_i \le Y_i)]$. One can show that $C_i^*(s)$ can be written as

$$C_i^*(s) = \mathrm{I}(Y_i \ge s)\left\{\delta_i \frac{W_1(Y_i, s)}{W_1(Y_i, Y_i)} + (1-\delta_i)\frac{W_0(Y_i, s)}{W_0(Y_i, Y_i)}\right\}, \tag{7.1}$$

where $W_1(t, s) = \int_0^s [1 - G(t-u)]\, dH(u)$ and $W_0(t, s) = \int_{t-s}^t h(t-u)\, dG(u)$. Note that the latter functions can be estimated respectively, by $\widehat{W}_1(t, s)$ and $\widehat{W}_0(t, s)$, where G is replaced by its Kaplan–Meier estimator $\widehat{G}$.

Let $L^G(x)$ denote the d.f. of $\gamma = \min(R, R_c)$ and $N_i(s) = \mathrm{I}(Y_i \le s, \delta_i = 1)$ $(i = 1, \ldots, n)$. Define $L_1^G(x) = \Pr(\gamma \le x, \delta = 1)$, $L_0^G(x) = \Pr(\gamma \le x, \delta = 0)$, $N^G(s) = n^{-1}\sum_{i=1}^n N_i^G(s) = n^{-1}\sum_{i=1}^n \mathrm{I}(\gamma_i \le s, \delta_i = 0)$ and $C^G(s) = n^{-1}\sum_{i=1}^n C_i^G(s) = n^{-1}\sum_{i=1}^n \mathrm{I}(\gamma_i \ge s)$. For any s, x and $y \in [0, U]$, let

$$M_i(s) = N_i(s) - \int_{(0,s]} C_i(t)\, d\Lambda(t), \tag{7.2a}$$

$$M_i^G(s) = \int_{[0,s]} \frac{1}{\overline{L^G}(t)}\, dN_i^G(t) - \int_{[0,s]} \frac{C_i^G(t)}{\overline{L^G}(t)^2}\, dL_0^G(t), \tag{7.2b}$$

$$J_1(x,y)=\int_{[0,y]}\frac{\xi_1(s,x)}{c(s)}\,d\Lambda(s), \tag{7.2c}$$

$$\begin{aligned}\xi_1(s,x)=&\int_{s\vee x}^{\infty}\frac{W_1(t,s)W_1(t,t-x)}{W_1^2(t,t)}\,dF_1(t)\\&+\int_{s\vee x}^{\infty}\frac{W_0(t,s)W_0(t,t-x)}{W_0^2(t,t)}\,dF_0(t)-\frac{[1-F(s)]W_1(s,s-x)}{\alpha_*}\\&-\int_{s\vee x}^{\infty}\frac{W_0(t,s)h(t-x)[1-G(x)]}{W_0^2(t,t)}\,dF_0(t),\end{aligned} \tag{7.2d}$$

$$\begin{aligned}P_1(s,s')&=\mathrm{Cov}\big(\widetilde{C}^*(s)-\widehat{C}(s),\widetilde{C}^*(s')-\widehat{C}(s')\big)\\&=\frac{[1-F(s\vee s')]W_1(s\vee s',s\wedge s')}{\alpha_*}-\int_{s\vee s'}^{\infty}\frac{W_1(t,s)W_1(t,s')}{W_1(t,t)^2}\,dF_1(t)\\&\quad-\int_{s\vee s'}^{\infty}\frac{W_0(t,s)W_0(t,s')}{W_0(t,t)^2}\,dF_0(t),\end{aligned} \tag{7.2e}$$

where $F_1(t)=\Pr(Y\le t,\delta=1)$ and $F_0(t)=\Pr(Y\le t,\delta=0)$ are the respective sub-distributions of the uncensored and censored observations, while $\overline{L^G}(t)=1-L^G(t^-)$ and $\alpha_*=\Pr[T^*\le X^*]$. The quantity α_* represents the proportion of truncated observation in the population, and is assumed to be positive. Let $M_*^G=\sum_{i=1}^n M_i^G(s)$.

7.2. Lemmas and proofs

First, we note that the proofs of all the lemmas are given in the supplementary material file [21]. In the following, we introduce an approximation for the cumulative hazard estimator

$$\widehat{\Lambda}_1(t)=\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{I}(Y_i\le t,\delta_i=1)}{\widehat{C}^*(Y_i)}. \tag{7.3}$$

The result is similar to the asymptotic representation in [14]. Let $x\in[0,U]$ and

$$\chi_i(x)=\int_0^x\frac{dM_i(s)}{c(s)}-\int_0^x\frac{C_i^*(s)-C_i(s)}{c(s)}\,d\Lambda(s)-\int_0^{u_G^-}J_1(s,x)\,dM_i^G(s). \tag{7.4}$$

Lemma 7.1. *If assumptions* B1 *and* B2(a) *hold, then* $\widehat{\Lambda}_1(x)$ *admits the following representation on* $[0,U]$:

$$\widehat{\Lambda}_1(x)-\Lambda(x)=n^{-1}\sum_{i=1}^{n}\chi_i(x)+r_n(x), \tag{7.5}$$

where $\sup_{0\le x\le U}|r_n(x)| = \mathcal{O}(\log(n)/n)$ *a.s., and the covariance of* $\chi_i(x)$ *and* $\chi_i(y)$ *is*

$$\Sigma(x,y) = \int_0^{x\wedge y}\frac{d\Lambda(s)}{c(s)} - \int_0^x\int_0^y \frac{P_1(s,s')}{c(s)c(s')}\,d\Lambda(s')\,d\Lambda(s) - \int_0^{u_G^-} J_1(s,x)J_1(s,y)\frac{dL_0^G(s)}{[\overline{L^G}(s)]^2}.$$

In what follows, we use $f_\pm$ to express both functions, f_+ and f_-. Define

$$\beta_\pm(x) = \int_{-1}^1 \lambda(x-uh_n)K_\pm(u)\,du - \lambda(x^\pm), \tag{7.6a}$$

$$\sigma_\pm(x) = (nh_n)^{-1}\sum_{i=1}^n\int_{-1}^1\big[\chi_i(x-uh_n)-\chi_i(x)\big]\,dK_\pm(u), \tag{7.6b}$$

$$e_\pm(x) = h_n^{-1}\int_{-1}^1\big[r_n(x-uh_n)-r_n(x)\big]\,dK_\pm(u). \tag{7.6c}$$

Note that $\widehat{\Delta}_1$ can be regarded as the convolution of K_* with respect to $\widehat{\Lambda}_1$. The representation of $\widehat{\Delta}_1$ in the next lemma will help us to derive the asymptotic approximations of $E[\widehat{\Delta}_1(x)]$ and $\mathrm{Var}[\widehat{\Delta}_1(x)]$ in Lemma 7.4, and establishes the asymptotic normality of $\widehat{\Delta}_1(x)$ in Theorem 4.2.

Lemma 7.2. *Let* $\beta_n^* = \beta_+ - \beta_-(x)$, $\sigma_n^* = \sigma_+ - \sigma_-$ *and* $e_n^* = e_+ - e_-$. *Suppose assumptions* B1, B2 *and* K1 *hold, and* $\log(n)/nh_n \to 0$ *as* $n\to\infty$. *The estimators* $\widehat{\lambda}_{1,\pm}$ *and* $\widehat{\Delta}_1$ *have the following representations on* $\mathcal{A}$:

$$\widehat{\lambda}_{1,\pm}(x) = \lambda(x^\pm) + \beta_\pm(x) + \sigma_\pm(x) + e_\pm(x), \tag{7.7}$$

$$\widehat{\Delta}_1(x) = \Delta(x) + \beta_n^*(x) + \sigma_n^*(x) + e_n^*(x), \tag{7.8}$$

where $\sup_{x\in\mathcal{A}}|e_\pm(x)| = \mathcal{O}(\log(n)/nh_n)$ *a.s. and* $\sup_{x\in\mathcal{A}}|e_n^*(x)| = \mathcal{O}(\log(n)/nh_n)$ *a.s.*

Lemma 7.3. *Suppose assumption* B1 *hold. Then*

(i) $c(t)$ *is bounded away from* 0 *for* $t\in[l_L,u_L]$.
(ii) $\widetilde{C}^*(s)$, $P_1(x,s)$ *and* $\xi_1(x,s)$ *are bounded.*

Lemma 7.4. *Let* $x\in\mathcal{A}$. *Suppose assumptions* B1, B2 *and* K1 *hold, then*

(i) $E\big[\widehat{\Delta}_1(x)\big] = \Delta(x) + \mathcal{O}\big(h_n + \log(n)/(nh_n)\big)$.

(ii) $$\mathrm{Var}\big[\widehat{\Delta}_1(x)\big] = \frac{1}{nh_n}\left[\frac{\lambda(x^-)}{c(x)}\int_0^1\big[K_-(u)\big]^2du + \frac{\lambda(x^+)}{c(x)}\int_{-1}^0\big[K_+(u)\big]^2du\right] + \mathcal{O}\left(\frac{\log^2(n)}{(nh_n)^2}\right).$$

The result (ii) *holds uniformly in* $x\in\mathcal{A}$.

Lemma 7.5. *Suppose assumptions* B1, B2 *and* K1 *hold, then*

$$\lim_{n\to\infty} \Pr\big(|\widehat{\tau}_1 - \tau| > h_n\big) = 0.$$

Proof of Theorem 4.1. The proof is by contradiction and is only given for $\widehat{\tau}_1$ with $\Delta(\tau) > 0$, the proof is similar for $\widehat{\tau}_2$ and the case $\Delta(\tau) < 0$. Since $\lim_{n\to\infty} \Pr(|\widehat{\tau}_1 - \tau| \le h_n) = 1$, by Lemma 7.5, it suffices to prove the theorem on the set $E_n^1 = \{\omega : |\widehat{\tau}_1(\omega) - \tau| \le h_n\}$. Suppose $|\widehat{\tau}_1 - \tau| = \mathcal{O}(1/n)$ does not hold, then there exist $\alpha > 0$, $m_n \to \infty$ and a subsequences of the n's, n_k, such that

$$P\big[n_k|\widehat{\tau}_1 - \tau| > m_{n_k}\big] > \alpha$$

for large n_k. Let $E_n^2 = \{\omega : n_k|\widehat{\tau}_1(\omega) - \tau| > m_{n_k}\}$ and $E_n = E_n^1 \cap E_n^2$. *w.l.o.g.* we replace n_k by n. We have for sufficiently large n

$$P[E_n] > \alpha_0, \tag{7.9}$$

with $\alpha_0 = \alpha/2 > 0$. We prove the theorem through the following steps.

Step 1. For $x = \tau, \widehat{\tau}_1$, large n and small h_n, it follows from the representation (7.8) that

$$\big|\widehat{\Delta}_1(x)\big| = \widetilde{\Delta}(x) + \sigma_n^*(x) + e_n^*(x), \tag{7.10}$$

since on E_n^1, $\widetilde{\Delta}(x) = \int_{-\infty}^{+\infty} h_n^{-1} K_*[(x-t)/h_n]\, d\Lambda(t) \simeq \Delta(\tau)$ for $x = \tau, \widehat{\tau}_1$ and sufficiently small n^{-1} and h_n, and $\sup_{\mathcal{A}} |\sigma_n^*(x)|$, $\sup_{\mathcal{A}} |e_n^*(x)| \to 0$ as $n \to \infty$ and $h_n \to 0$.

Step 2. Recall that $\lambda(x) = \lambda_0(x) + \Delta(\tau)\mathrm{I}_{[\tau,U]}(x)$ with $\lambda_0 \in \mathcal{C}^1([0, \mathrm{U}])$. We have

$$\begin{aligned}
\widetilde{\Delta}(x) &= \int_{-1}^{+1} \lambda(x - th_n)K_*(t)\, dt = \int_{-1}^{+1} \big[\lambda_0(x - th_n) + \Delta(\tau)\mathrm{I}_{[\tau,U]}(x - th_n)\big]K_*(t)\, dt \\
&= \int_{-1}^{+1} \big[\lambda_0(x - th_n) - \lambda_0(x)\big]K_*(t)\, dt + \Delta(\tau)\int_{-1}^{+1} \mathrm{I}_{[\tau,U]}(x - th_n)K_*(t)\, dt \\
&= a_n(x) + \Delta(\tau)\int_{(x-U)/h_n}^{(x-\tau)/h_n} K_*(t)\, dt,
\end{aligned}$$

where $\sup_{\mathcal{A}} |a_n(x)| = \mathcal{O}(h_n)$, using the Mean Value theorem. One can then show, for $K_*(t) = \mathrm{I}_{[-1,0]}(t) - \mathrm{I}_{[0,1]}(t)$, that

$$\widetilde{\Delta}(\tau) - \widetilde{\Delta}(\widehat{\tau}_1) = \Delta(\tau)\left[\frac{|\widehat{\tau}_1 - \tau|}{h_n} + 1 - \min\big(1, (U - \widehat{\tau}_1)/h_n\big)\right] + a_n^*, \tag{7.11}$$

with $a_n^* = \mathcal{O}_p(h_n)$. Let P_0 denotes the term inside the brackets in (7.11), we notice that

$$P_0 \ge \frac{|\widehat{\tau}_1 - \tau|}{h_n}. \tag{7.12}$$

Step 3. Analogously to Lemma 7.4 one can show, uniformly in $x \in [\tau - h_n, \tau - h_n]$, that

$$\sigma_n^*(\tau) - \sigma_n^*(x) = \mathcal{O}_p\left(\sqrt{\frac{|x-\tau|}{nh_n}}\right),$$

by using the facts that λ and $c(x) = P[T \le x \le Y]$ are Lipschitz functions on $[0,\tau)$ or $[\tau, U]$, since $\lambda \in \mathcal{C}^1([0,\tau]) \cap \mathcal{C}^1([\tau, U])$ and the d.f. of Y and T are Lipschitz functions. It then follows, on E_n^1, that

$$\sigma_n^*(\tau) - \sigma_n^*(\widehat{\tau}_1) = \mathcal{O}_p\left(\sqrt{\frac{|\widehat{\tau}_1-\tau|}{nh_n}}\right). \tag{7.13}$$

Note that since $\sup_{\mathcal{A}} |\sigma_n^*(x)| = \mathcal{O}_p(1/\sqrt{nh_n})$, $\sup_{\mathcal{A}} |e_n^*(x)| = \mathcal{O}_p(\log(n)/nh_n)$ and $\log^2(n)/nh_n \to 0$ as $n \to \infty$ and $h_n \to 0$ (Theorem's 4.1 assumptions), then

$$e_n^*(\tau) - e_n^*(\widehat{\tau}_1) = \mathcal{O}_p\big(\sigma_n^*(\tau) - \sigma_n^*(\widehat{\tau}_1)\big). \tag{7.14}$$

Step 4. Now, using (7.10) and (7.11), we have

$$\big|\widehat{\Delta}_1(\tau)\big| - \big|\widehat{\Delta}_1(\widehat{\tau}_1)\big| = \Delta(\tau)P_0\left\{1 + \frac{a_n^*}{\Delta(\tau)P_0} + \frac{\sigma_n^*(\tau) - \sigma_n^*(\widehat{\tau}_1)}{\Delta(\tau)P_0} + \frac{e_n^*(\tau) - e_n^*(\widehat{\tau}_1)}{\Delta(\tau)P_0}\right\}. \tag{7.15}$$

Combining the facts that, on E_n, $P_0 \ge |\widehat{\tau}_1 - \tau|/h_n$ and $|\widehat{\tau}_1 - \tau| > m_n/n$, and using (7.13) and (7.14), we obtain for any $\epsilon > 0$ and sufficiently large n

$$\begin{aligned}
&P\left[\left(\frac{|a_n^* + \sigma_n^*(\tau) - \sigma_n^*(\widehat{\tau}_1) + e_n^*(\tau) - e_n^*(\widehat{\tau}_1)|}{\Delta(\tau)P_0} \le \epsilon\right) \cap E_n\right] \\
&\ge P\left[\left(\frac{|a_n^*| + |\sigma_n^*(\tau) - \sigma_n^*(\widehat{\tau}_1)| + |e_n^*(\tau) - e_n^*(\widehat{\tau}_1)|}{\Delta(\tau)|\widehat{\tau}_1 - \tau|} h_n \le \epsilon\right) \cap E_n\right] \\
&\ge P\left[\left(\frac{|a_n^*|}{\Delta(\tau)|\widehat{\tau}_1 - \tau|} h_n \le \frac{\epsilon}{3}\right) \cap \left(\frac{|\sigma_n^*(\tau) - \sigma_n^*(\widehat{\tau}_1)|}{\Delta(\tau)|\widehat{\tau}_1 - \tau|} h_n \le \frac{\epsilon}{3}\right)\right. \\
&\quad \left.\cap \left(\frac{|e_n^*(\tau) - e_n^*(\widehat{\tau}_1)|}{\Delta(\tau)|\widehat{\tau}_1 - \tau|} h_n \le \frac{\epsilon}{3}\right) \cap E_n\right] \\
&= P\left[\left(\frac{|a_n^*|}{\Delta(\tau)|\widehat{\tau}_1 - \tau|} h_n \le \frac{\epsilon}{3}\right) \cap \left(\frac{\sqrt{h_n}}{\Delta(\tau)\sqrt{n|\widehat{\tau}_1 - \tau|}} \le \epsilon_0\right) \cap E_n\right] \\
&\ge P\left[\left(\frac{|a_n^* n h_n|}{\Delta(\tau)m_n} \le \frac{\epsilon}{3}\right) \cap \left(\frac{\sqrt{h_n}}{\Delta(\tau)\sqrt{m_n}} \le \epsilon_0\right) \cap E_n\right],
\end{aligned}$$

where ϵ_0 is an arbitrary positive value. Note that $|a_n^* n h_n| = \mathcal{O}_p(nh_n^2)$ with $nh_n^2 \to 0$ (Theorem's 4.1 assumptions) and $m_n \to \infty$ as $n \to \infty$ and $h_n \to 0$. Thus, for sufficiently large n and small h_n,

$$P\left[\left(\frac{|a_n^* + \sigma_n^*(\tau) - \sigma_n^*(\widehat{\tau}_1) + e_n^*(\tau) - e_n^*(\widehat{\tau}_1)|}{\Delta(\tau)P_0} \le \frac{1}{2}\right) \cap E_n\right] \ge P(E_n)/2. \tag{7.16}$$

On the other hand, since $P_0, \Delta(\tau) > 0$, we have

$$P\big[\big(\big|\widehat{\Delta}_1(\tau)\big| - \big|\widehat{\Delta}_1(\widehat{\tau}_1)\big| > 0\big) \cap E_n\big] \\ \geq P\bigg[\bigg(\frac{|a_n^* + \sigma_n^*(\tau) - \sigma_n^*(\widehat{\tau}_1) + e_n^*(\tau) - e_n^*(\widehat{\tau}_1)|}{\Delta(\tau)P_0} \leq \frac{1}{2}\bigg) \cap E_n\bigg].$$

Consequently, we obtain by (7.9)

$$P\big[\big(\big|\widehat{\Delta}_1(\tau)\big| - \big|\widehat{\Delta}_1(\widehat{\tau}_1)\big| > 0\big) \cap E_n\big] \geq \alpha_0 > 0.$$

This contradicts the fact $|\widehat{\Delta}_1(\widehat{\tau}_1)| = \sup_{\mathcal{A}} |\widehat{\Delta}_1(x)|$, and completes the proof. □

Proof of Theorem 4.2. In the following, we show that the conditions of the central limit theorem of Lindeberg–Feller are fulfilled. Notice that from the proof of Lemma 7.4(i), $\beta_n^*(y) = \mathcal{O}(h_n)$, and since $\sup_{y\in\mathcal{A}} |e_n^*(y)| = \mathcal{O}(\log(n)/nh_n)$ a.s., then the representation (7.8) implies

$$\begin{aligned} &\sqrt{nh_n}\big[\widehat{\Delta}_1(y) - \Delta(y)\big] \\ &\quad= \sqrt{nh_n}\big[\sigma_n^*(y) + \beta_n^*(y) + e_n^*(y)\big] \\ &\quad= \frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \int_{-1}^1 \big[\chi_i(y - uh_n) - \chi_i(y)\big]\, d\big(K_+(u) - K_-(u)\big) \\ &\qquad+ \mathcal{O}\Big(\sqrt{nh_n^3} + \log(n)/\sqrt{nh_n}\Big). \end{aligned} \tag{7.17}$$

The first term of (7.17) is a triangular array of row-wise i.i.d. random variables, and the second term converges almost surely to zero. Define the following function

$$A_{n,i}(y) = \frac{1}{\sqrt{nh_n}} \int_{-1}^1 \big[\chi_i(y - uh_n) - \chi_i(y)\big]\, d\big(K_+(u) - K_-(u)\big).$$

Notice that $E[A_{n,i}(y)] = 0$ follows from the fact that $E[\chi_i(y)] = 0$ $(i = 1, \dots, n)$, and $\sum_{i=1}^n E[A_{n,i}^2(y)] \to \sigma^2$ as $n \to \infty$ follows from Lemma 7.4(ii), where σ^2 is the asymptotic variance in Theorem 4.2. It remains to show that $\sum_{i=1}^n E[A_{n,i}^2(y)\mathrm{I}(|A_{n,i}(y)| > \epsilon)] \to 0$ as $n \to \infty$ for $\epsilon > 0$. First, using equalities (7.2a)–(7.2d) and (7.4), we may write $\chi_i(y - uh_n) - \chi_i(y)$ as

$$\begin{aligned} \chi_i(y - uh_n) - \chi_i(y) &= \frac{\mathrm{I}[Y_i \in (y, y - uh_n], \delta_i = 1]}{c(Y_i)} - \int_y^{y-uh_n} \frac{C_i^*(s)}{c(s)}\, d\Lambda(s) \\ &\quad- \frac{\mathrm{I}[\gamma_i \leq u_G, \delta_i = 0]}{\overline{L^G}(\gamma_i)} \times \int_y^{y-uh_n} \frac{\xi_1(s, \gamma_i)}{c(s)}\, d\Lambda(s) \\ &\quad+ \int_{s<u_G} \bigg[\int_y^{y-uh_n} \frac{\xi_1(v, s)}{c(v)}\, d\Lambda(v)\bigg] \frac{\mathrm{I}(s \leq \gamma_i)}{[\overline{L^G}(s)]^2}\, dL_0^G(s), \end{aligned}$$

where $\gamma_i = R_i \wedge R_{c,i}$ $(i = 1, \dots, n)$. Note that under assumptions B1–B2, $\lambda(t)$ and $1/\overline{L^G}(s)$ are bounded for $t \in [l_L, U]$ and $s < u_G$, and by Lemma 7.3, the functions $1/c(t)$, $C_i^*(t)$ and $\xi_1(t, \cdot)$ are bounded for $t \in (y, y - uh_n) \subseteq [l_L, U]$, for sufficiently small h_n. Hence, $\chi_i(y - uh_n) - \chi_i(y)$ is bounded for every $y \in [l_L, U]$, and therefore, $\sqrt{nh_n} A_{n,i}(y)$ is bounded on $[l_L, U]$. Let m be the upper-bound of $\sqrt{nh_n}|A_{n,i}(y)|$. It follows from equation (A.3), in the supplementary material file, that $E[nh_n A_{n,i}^2(y)] = \mathcal{O}(h_n)$. Thus, using the Chebyshev inequality, we have

$$\begin{aligned} E\big[A_{n,i}^2(y)\mathrm{I}\big(\big|A_{n,i}(y)\big| > \epsilon\big)\big] &\le \frac{m^2}{nh_n} \times \Pr\big\{\big|A_{n,i}(y)\big| > \epsilon\big\} \le \frac{m^2}{nh_n} \times \frac{E[A_{n,i}^2(y)]}{\epsilon^2} \\ &= \frac{m^2}{nh_n} \times \frac{\mathcal{O}(h_n)}{\epsilon^2 nh_n} = \frac{\mathcal{O}(1)}{\epsilon^2 n^2 h_n}, \end{aligned}$$

thus $\sum_{i=1}^n E[A_{n,i}^2(y)I(|A_{n,i}(y)| > \epsilon)] \le \frac{\mathcal{O}(1)}{\epsilon^2 nh_n}$, hence $\sum_{i=1}^n E[A_{n,i}^2(y)I(|A_{n,i}(y)| > \epsilon)] \to 0$ as $n \to \infty$. The desired result then follows from the Lindeberg–Feller theorem. □

The following Lemmas 7.6–7.11 will help us to establish the weak convergence of φ_n in Theorem 4.3 and the results in Corollaries 4.1–4.2.

Lemma 7.6. *Let $x \in [-M, M]$ and define the sequence of processes*

$$\begin{aligned} \psi_n(x) = \alpha_n\big[&\Delta(\tau + x\theta_n h_n) - \Delta(\tau) + \beta_n^*(\tau + x\theta_n h_n) - \beta_n^*(\tau) \\ &+ \sigma_n^*(\tau + x\theta_n h_n) - \sigma_n^*(\tau)\big]. \end{aligned} \tag{7.18}$$

Under the assumptions of Theorem 4.3 we have, uniformly in $x \in \mathcal{A}$,

$$\varphi_n(x) = \psi_n(x) + \mathcal{O}\big(\log(n)/(nh_n)^{1/3}\big) \quad a.s.$$

Note that $\beta_n^*(t)$ and $\sigma_n^*(t)$ are defined in Lemma 7.2.

Lemma 7.7. *Let $x \in [-M, M]$. Under the assumptions of Theorem 4.3, we have*

$$E\big[\psi_n(x)\big] = -\frac{\Delta(\tau)K_-^{(3)}(0)}{4!}x^4 + o(1). \tag{7.19}$$

Lemma 7.8. *Let $x, y \in [-M, M]$. Suppose the assumptions of Theorem 4.3 hold, we have*

$$\begin{aligned} \mathrm{Cov}\big[\psi_n(x), \psi_n(y)\big] = xy\bigg[&\frac{\lambda(\tau^-)}{c(\tau)}\int_0^1 \big[K_-^{(1)}(u)\big]^2\,du + \frac{\lambda(\tau^+)}{c(\tau)}\int_{-1}^0 \big[K_+^{(1)}(u)\big]^2\,du\bigg] \\ &+ o(1). \end{aligned} \tag{7.20}$$

Lemma 7.9. *Let $x_1, \dots, x_l \in [-M, M]$, where $l \ge 1$ is an integer. Under the assumptions of Theorem 4.3, the vector $(\psi_n(x_1) - E[\psi_n(x_1)], \dots, \psi_n(x_l) - E[\psi_n(x_l)])^t$ has an asymptotic normal* *distribution, that is*

$$\big(\psi_n(x_1) - E\big[\psi_n(x_1)\big], \dots, \psi_n(x_l) - E\big[\psi_n(x_l)\big]\big)^t \xrightarrow{d} N(0, \mathcal{V}),$$

where $\mathcal{V} = (v_{i,j})_{1\le i,j\le l}$ *is the covariance matrix, with* $v_{i,j} = \mathrm{Cov}[\psi_n(x_i), \psi_n(x_j)]$.

Lemma 7.10. *Suppose the assumptions of Theorem 4.3 hold, and let* $x \in [-M, M]$. *The sequence of processes* $\psi_n(x) - E[\psi_n(x)]$ *is tight.*

Proof of Theorem 4.3. By Lemmas 7.7–7.10, the sequence of processes $\psi_n(x)$ (defined in Lemma 7.6) converges weakly to

$$\varphi(x) = -\frac{\Delta(\tau)K_-^{(3)}(0)}{4!}x^4 + xZ,$$

where $Z \sim N(0, \frac{\lambda(\tau^-)}{c(\tau)}\int_0^1 [K_-^{(1)}(t)]^2\,dt + \frac{\lambda(\tau^+)}{c(\tau)}\int_{-1}^0 [K_+^{(1)}(t)]^2\,dt)$. Hence, by Lemma 7.6 and Theorem 4.3's assumptions, $\varphi_n(x)$ converges weakly to $\varphi(x)$ as $n \to \infty$. □

Lemma 7.11. *Suppose the assumptions of Theorem 4.3 hold. We have*

$$|\widehat{\tau}_1 - \tau| = \mathcal{O}_p(h_n\theta_n).$$

Theorem 7.1. *Suppose assumptions* B1 *and* B2 *hold, then*

$$\sup_{0\le x\le U}\big|\widehat{\Lambda}_1(x) - \Lambda(x)\big| = \mathcal{O}(\sqrt{\log\log n/n}) \quad a.s. \tag{7.21}$$

7.3. Lemma 7.12

Let

$$\widehat{\Lambda}_2(t) = \frac{1}{n}\sum_{i=1}^n \frac{\mathrm{I}(Y_i \le t, \delta_i = 1)}{\widehat{C}(Y_i)}, \tag{7.22}$$

which is an estimator ([23]) of the cumulative hazard Λ. Define,

$$\widetilde{\chi}_i(x) = \mathrm{I}(Y_i \le x, \delta_i = 1)/c(Y_i) - \int_0^x \big[\mathrm{I}(T_i \le u \le Y_i)/c^2(u)\big]\,dF_1(u)$$

and

$$\widetilde{\beta}_*(x) = \int_{-1}^1 \lambda(x - uh_n)K_*(u)\,du - \Delta(x), \tag{7.23a}$$

$$\widetilde{\sigma}_*(x) = (nh_n)^{-1}\sum_{i=1}^n \int_{-1}^1 \big[\widetilde{\chi}_i(x - uh_n) - \widetilde{\chi}_i(x)\big]\,dK_*(u), \tag{7.23b}$$

$$\widetilde{e}_*(x) = h_n^{-1}\int_{-1}^1 \big[\widetilde{r}_n(x - uh_n) - \widetilde{r}_n(x)\big]\,dK_*(u), \tag{7.23c}$$

where $\widetilde{r}_n$ is the reminder term in the representation

$$\widehat{\Lambda}_2(x) = \Lambda(x) + n^{-1} \sum_{i=1}^{n} \widetilde{\chi}_i(x) + \widetilde{r}_n(x) \tag{7.24}$$

(representation (1.12) in [10]). Note that $\widehat{\Delta}_2$ can be viewed as the convolution of K_* with respect to $\widehat{\Lambda}_2$. The i.i.d. representation of $\widehat{\Delta}_2$, in the next lemma, helps to derive the asymptotic distributions of $\widehat{\Delta}_2(x)$, $\widehat{\tau}_2$ and $\widehat{\Delta}_2(\widehat{\tau}_2)$ in Theorem 4.4, by following similar steps to those in Lemmas 7.2–7.11.

Lemma 7.12. *Suppose assumptions* B1, B2 *and* K1 *hold, and* $\log(n)/nh_n \to 0$ *as* $n \to \infty$ *and* $h_n \to 0$. *The estimator* $\widehat{\Delta}_3$ *admits the following representation on* $\mathcal{A}$

$$\widehat{\Delta}_2(x) = \Delta(x) + \widetilde{\beta}_*(x) + \widetilde{\sigma}_*(x) + \widetilde{e}_*(x), \tag{7.25}$$

where $\sup_{x\in\mathcal{A}} |\widetilde{e}_*(x)| = \mathcal{O}(\log(n)/nh_n)$ *a.s.*

8. Summary and conclusions

We presented methods for estimating the change-point (location and size) of an otherwise smooth hazard function under biased sampling when the observations are subject to informative censoring. Two methodologies were developed for two scenarios: the left-truncation distribution H is known and when it is unknown. The estimators in the first scenario, which incorporate the available information of H, were found more efficient than their counterparts in the second scenario, when H is unknown (Remark 4.1). In both cases, the change-point estimators can achieve the rate $\mathcal{O}_p(1/n)$. Under condition K2(c), needed for the derivation of the asymptotic distribution of the estimators, the rate is, however, slightly slower and is of order $\mathcal{O}_p(\log^{\frac{5}{3}}(n)/n)$. For both scenarios, we studied a sequence of processes, (4.2) and (4.3), operating on a local deviation from the change-point τ, and established its convergence to a random parabola. This particular geometrical shape (parabola) allowed us to obtain the asymptotic distributions of the estimators, and hence, be able to devise interval-estimators for τ and its size $\Delta(\tau)$; necessary for making statistical inference about whether or not an abrupt-change exists within the hazard of failure in real-data applications. An issue, not studied in this paper, is the choice of the smoothing parameter h_n. Although our asymptotic-rate (5.2) have shown satisfactory results in the simulations (Tables 1–2, Figure 1(a)–(d)), the question of finding an adapted method for the choice of the bandwidth h_n, in change-point problem under scenarios considered in this manuscript, is not explored yet and remains open.

Acknowledgements

This work has been supported in part by NSERC of Canada. The data reported in this article were collected as part of the Canadian Study of Health and Aging. The core study was funded by the Senior's Independence Research Program, through the National Health Research and Development Program (NHRDP) of Health Canada [project no. 6606-3954-MC(S)]. Additional funding was provided by Pfizer Canada Incorporated through the Medical Research Council/ Pharmaceutical Manufacturers Association of Canada Health Activity Program, NHRDP [project no. 6603-1417-302(R)], Bayer Incorporated and the British Columbia Health Research Foundation [project no. 38(93-2) and no. 34(96-1)]. The study was coordinated through the University of Ottawa and the Division of Aging and Seniors, Health Canada. The authors thank Professor Christina Wolfson for providing the data. We also like to thank the editor, associate editor and the referees for their comments and questions which led us to improve some results and sharpen some of the rates.

Supplementary Material

Additional technical details: Proofs of lemmas (DOI: 10.3150/16-BEJ825SUPP; .pdf). The proofs of all the lemmas and Theorem 7.1 are provided in the supplementary material.

References

[1] Antoniadis, A., Gijbels, I. and MacGibbon, B. (2000). Non-parametric estimation for the location of a change-point in an otherwise smooth hazard function under random censoring. *Scand. J. Stat.* **27** 501–519. MR1795777

[2] Asgharian, M., M'Lan, C.E. and Wolfson, D.B. (2002). Length-biased sampling with right censoring: An unconditional approach. *J. Amer. Statist. Assoc.* **97** 201–209. MR1947280

[3] Bergeron, P.-J., Asgharian, M. and Wolfson, D.B. (2008). Covariate bias induced by length-biased sampling of failure times. *J. Amer. Statist. Assoc.* **103** 737–742. MR2524006

[4] Cox, D.R. (1969). Some sampling problems in technology. In *New Developments in Survey Sampling* (N. L. Johnson and H. Smith, eds.) 506–527. New York: Wiley.

[5] de Uña-Álvarez, J. (2004). Nonparametric estimation under length-biased sampling and type I censoring: A moment based approach. *Ann. Inst. Statist. Math.* **56** 667–681. MR2126804

[6] Eddy, W.F. (1980). Optimum kernel estimators of the mode. *Ann. Statist.* **8** 870–882. MR0572631

[7] Feuerverger, A. and Hall, P. (2000). Methods for density estimation in thick-slice versions of Wicksell's problem. *J. Amer. Statist. Assoc.* **95** 535–546. MR1803171

[8] Finkelstein, H. (1971). The law of the iterated logarithm for empirical distributions. *Ann. Math. Stat.* **42** 607–615. MR0287600

[9] Fisher, R.A. (1934). The effect of methods of ascertainment upon the estimation of frequencies. *Ann. Hum. Genet.* **6** 13–25.

[10] Gijbels, I. and Wang, J.-L. (1993). Strong representations of the survival function estimator for truncated and censored data with applications. *J. Multivariate Anal.* **47** 210–229. MR1247375

[11] Kosorok, M.R. and Song, R. (2007). Inference under right censoring for transformation models with a change-point based on a covariate threshold. *Ann. Statist.* **35** 957–989. MR2341694

[12] Kvam, P. (2008). Length bias in the measurements of carbon nanotubes. *Technometrics* **50** 462–467. MR2655646

[13] Leiva, V., Barros, M., Paula, G.A. and Sanhueza, A. (2008). Generalized Birnbaum–Saunders distributions applied to air pollutant concentration. *Environmetrics* **19** 235–249. MR2420468

[14] Luo, X. and Tsai, W.Y. (2009). Nonparametric estimation for right-censored length-biased data: A pseudo-partial likelihood approach. *Biometrika* **96** 873–886. MR2767276
[15] Müller, H.-G. (1992). Change-points in nonparametric regression analysis. *Ann*. *Statist*. **20** 737–761. MR1165590
[16] Müller, H.-G. and Wang, J.-L. (1996). An invariance principle for discontinuity estimation in smooth hazard functions under random censoring. *Sankhyā Ser. A* **58** 392–402. MR1659134
[17] Neyman, J. (1955). Statistics; servant of all sciences. *Science* **122** 401–406.
[18] Nowell, C., Evans, M.A. and McDonald, L. (1988). Length-biased sampling in contingent valuation studies. *Land Economics* **64** 367–371.
[19] Nowell, C. and Stanley, R.L. (1991). Length-biased sampling in mall intercept surveys. *J*. *Mark*. *Res*. **28** 475–479.
[20] Pons, O. (2003). Estimation in a Cox regression model with a change-point according to a threshold in a covariate. *Ann*. *Statist*. **31** 442–463. Dedicated to the memory of Herbert E. Robbins. MR1983537
[21] Rabhi, Y. and Asgharian, M. (2016). Supplement to "Inference under biased sampling and right censoring for a change point in the hazard function." DOI:10.3150/16-BEJ825SUPP.
[22] Terwilliger, J.D., Shannon, W.D., Lathrop, G.M., Nolan, J.P., Goldin, L.R., Chase, G.A. and Weeks, D.E. (1997). True and false positive peaks in genomewide scans: Applications of length-biased sampling to linkage mapping. *Am*. *J*. *Hum*. *Genet*. **61** 430–438.
[23] Tsai, W.Y., Jewell, N.P. and Wang, M.C. (1987). A note on the product limit estimator under right censoring and left truncation. *Biometrika* **74** 883–886.
[24] Wang, M.-C. (1991). Nonparametric estimation from cross-sectional survival data. *J*. *Amer*. *Statist*. *Assoc*. **86** 130–143. MR1137104
[25] Wicksell, S.D. (1925). The corpuscle problem, part I. *Biometrika* **17** 84–89.
[26] Wolfson, C., Wolfson, D.B., Asgharian, M., M'lan, C.E., Østbye, T., Rockwood, K. and Hogan, D.B. (2001). A re-evaluation of the duration of survival after the onset of dementia. *N*. *Engl*. *J*. *Med*. **344** 1111–1116.
[27] Zelen, M. (1993). Optimal scheduling of examinations for the early detection of disease. *Biometrika* **80** 279–293. MR1243504

Received August 2014 and revised January 2016