



**HAL**  
open science

# Asymptotic properties of M-estimators based on estimating equations and censored data in semi-parametric models with multiple change points

Salim Bouzebda, Anouar Abdeldjaoued Ferfache

► **To cite this version:**

Salim Bouzebda, Anouar Abdeldjaoued Ferfache. Asymptotic properties of M-estimators based on estimating equations and censored data in semi-parametric models with multiple change points. Australian Journal of Mathematical Analysis and Applications, 2021, 497 (2), pp.124883. 10.1016/j.jmaa.2020.124883 . hal-03090193

**HAL Id: hal-03090193**

**<https://hal.science/hal-03090193>**

Submitted on 2 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# Asymptotic properties of $M$ -estimators based on estimating equations and censored data in semi-parametric models with multiple change points

Salim BOUZEBDA\* and Anouar Abdeldjaoued FERFACHE†  
Laboratoire de Mathématiques Appliquées de Compiègne  
Université de Technologie de Compiègne

## Abstract

*Statistical models with multiple change points in presence of censored data are used in many fields; however, the theoretical properties of  $M$ -estimators of such models have received relatively little attention. The main purpose of the present work is to investigate the asymptotic properties of  $M$ -estimators of the parameters of a multiple change-point model for a general class of models in which the form of the distribution can change from segment to segment and in which, possibly, there are parameters that are common to all segments, **in the setting of a known number of change points**. Consistency of the  $M$ -estimators of the change points is established and the rate of convergence is determined. The asymptotic normality of the  $M$ -estimators of the parameters of the within-segment distributions is established. Since the approaches used in the complete data models are not easily extended to multiple change-point models in the presence of censoring, we have used some general results of Kaplan-Meier integrals. **We investigate the performance of the methodology for small samples through a simulation study.***

**Key words:** Semiparametric inference; multiple change-points; change-point fraction; common parameter; consistency; convergence rate;  $M$ -estimators;  $Z$ -estimators; censored data; Kaplan Meier integrals; Argmax theorem; Central limit theorem.  
**Mathematics Subject Classification :** 62F12; 62F03; 62G20; 60F05; 62N02; 62E20; 62P20.

## 1 Introduction and motivations

In major real data investigation, the stationarity assumption has been frequently used. However, in practice, time series entail in their dependence structure and therefore modelling non-stationary processes using stationary methods to capture their time-evolving dependence aspects most likely result in a crude approximation. Change-point detection plays a critical role in such situation. Notice that the problem of change-points in a sequence of random variables has a long history. Early work on this problem can be found in Page (1954, 1955, 1957) who investigated quality control problems and proposed a sequential scheme for identifying changes in the mean of a sequence of independent random variables. Over time, methods in change point analysis have been developed to address data analytic questions in fields ranging from biology to finance, and in many cases such methodology has become standard. The statistical community now enjoys a vast literature on change point analysis where many of the most natural and common questions have received at least some attention. For a broader presentation of the field of change-point analysis along with statistical applications, we refer the reader to the monographs by Brodsky and Darkhovsky (1993), Csörgő and Horváth (1997), Chen and Gupta (2000), Wu (2005) and Pons (2018), just to cite a few. We refer to the paper of Lee (2010) for a list of comprehensive bibliography of books and research papers on this topic. The problem of detecting abrupt changes has been discussed intensively in a time series context, we may refer to Jandhyala *et al.* (2013) and Aue and Horváth (2013) for a review of the literature. Recent references on the subject include Chen (2019), Chu and Chen (2019), Garreau and Arlot (2018), Tan and Zhang (2019), Nkurunziza and Fu (2019), Qian *et al.* (2019) and El Ktaibi and Ivanoff (2019). Compared to single change-point detection, multiple change-points detection is a much more challenging problem. Work on detection for multiple change-points began in the 1980s (e.g., Vostrikova (1981), Yin (1988), Yao (1988)). There exists a rich literature devoted to this field, we refer to Truong *et al.* (2020) for review of change-point and some extensions. For the censored setting, there are only a few papers dealing with detection of changes, for single change-point, we refer to Stute (1996) who provided an estimator of the change point based on the  $U$ -statistics. Gombay and Liu (2000), Hušková and Neuhaus (2004), Al-Awadhi and Aly (2005), Wang and Zheng (2012) have considered tests

---

\*e-mail adress: salim.bouzebda@utc.fr

†e-mail adress: anouar-abdeldjaoued.ferfache@utc.fr

procedures for change-point. He (2015, 2017) considered the multiple change-points for particular distributions. To our best knowledge there the case where the change occurs for the two variables, i.e., the censored variable and the censorship variable in general setting was not investigated in the literature up to present. Notice that multiple change-points problem occurs for the survival function due to hazard change according to evolving time. For example, a cancer survival function can change abruptly or smoothly at a few time points. For example, Kim *et al.* (2020) applied their method to find the change-points for leukemia survival data and identified the change-points. However multiple change-points problems are not much considered due to its computational complexity and theoretical difficulty. Hušková and Neuhaus (2004) have investigated the problem of single change when the censored variables are assumed to be independent but not necessarily identically distributed. While the body of work about the change-point constitutes a rich literature, it mainly deals with the inference of a single change in a short or moderate sized sequence. Detecting multiple change-points in a very long sequence has emerged as an important problem that has attracted more and more attention recently, we refer to Niu *et al.* (2016). There is a large literature on the change-point problem and their applications and it is not the purpose of the present paper to survey this extensive literature.

The main purpose of the present work is to consider a general framework and the characterization of the asymptotic properties of semi-parametric  $M$ -estimators based on censored data in models with multiple change-points, this generalization is far from being trivial and harder to control the estimator of Kaplan-Meier of each sample, which form a basically unsolved open problem in the literature. We aim at filling this gap in the literature by combining results He and Severini (2010) with techniques handling the Kaplan Meier integrals. However, as will be seen later, the problem requires much more than “simply” combining ideas from the existing results. In fact, delicate mathematical derivations will be required to cope with Kaplan Meier integrals in our context.

We start by giving some notations and definitions that are needed for the forthcoming sections. Let  $X_1, \dots, X_n$  be  $n$  independent random variables censoring by  $n$  independent random variables  $C_1, \dots, C_n$  respectively, where  $X_i$  and  $C_i$  are independent for all  $i$ , so we observe

$$(Y_i = X_i \wedge C_i, \delta_i = \mathbb{I}_{\{X_i \leq C_i\}}), \quad \text{for } 1 \leq i \leq n.$$

Survival data in clinical trials or failure time data in reliability studies, for example, are often subject to such censoring. To be more specific, many statistical experiments result in incomplete samples, even under well-controlled conditions. For example, clinical data for surviving most types of disease are usually censored by other competing risks to life which result in death. We suppose that there exists unknown change points  $n_1, \dots, n_k$ , such that

$$0 = n_0 < n_1 < \dots < n_k < n_{k+1} = n,$$

where for each  $j = 1, \dots, k + 1$ ,  $(X_{n_{j-1}+1}, C_{n_{j-1}+1}), \dots, (X_{n_j}, C_{n_j})$ , are i.i.d. with distribution function depending on  $j$ . Here, we consider semi-parametric change-points models in which the distribution function of  $X_{n_{j-1}+1}, \dots, X_{n_j}$  is parametric. We suppose that the theoretical distribution  $F_{n_j^0}(\cdot) =: F(\alpha^0, \theta_j^0, \cdot)$  of  $X_i$ ,  $i = 1, \dots, n$ , depends on the real common parameter  $\alpha^0$  for all  $j = 1, \dots, k + 1$  and the real within-segment  $\theta_j^0$ , for each  $j = 1, \dots, k + 1$  which are assumed to be unknown. In this model, there are  $k$  real change points  $n_1^0, \dots, n_k^0$  but unknown, where the number of change point  $k$  is assumed to be known. We estimate the unknown parameters  $n_j$ ,  $\alpha$  and  $\theta_j$ ,  $j = 1, \dots, k + 1$  by maximizing the estimating equations defined by:

$$\ell \equiv \ell(\alpha, \theta_1, \dots, \theta_{k+1}, n_1, \dots, n_k) = \sum_{j=1}^{k+1} \frac{(n_j - n_{j-1})}{n} \int_{\mathbb{R}} g_j(\alpha, \theta_j, x) d\widehat{F}_{n_j}(x), \quad (1.1)$$

where  $1 - \widehat{F}_{n_j}(\cdot)$  is the usual Kaplan-Meier product limit estimator of  $1 - F_{n_j}(\cdot)$  introduced by Kaplan and Meier (1958) and defined by

$$1 - \widehat{F}_{n_j}(x) = \prod_{i=n_{j-1}+1}^{n_j} \left(1 - \frac{d_i}{n_i}\right)^{\mathbb{I}_{\{Y_{(i)} \leq x\}}}, \quad (1.2)$$

where

$$r_i = \sum_{k=n_{j-1}+1}^{n_j} \mathbb{I}_{\{Y_{(i)} \leq Y_k\}}$$

and

$$d_i = \sum_{k=n_{j-1}+1}^{n_j} \mathbb{I}_{\{Y_{(i)} = Y_k, \delta_k = 1\}},$$

denoting the number of individuals still at risk at time  $Y_{(i)}$  and the number of deaths at time  $Y_{(i)}$  respectively, and  $Y_{(i)}$  denotes the order statistic of  $Y_{n_{j-1}+1}, \dots, Y_{n_j}$  and  $\mathbb{1}_E$  denoting the indicator function of  $E$ . For each sample  $X_{n_{j-1}+1}, \dots, X_{n_j}$ ,  $j = 1, \dots, k+1$ , and  $g_j(\cdot)$  is a given measurable function from  $\Upsilon \times \Theta_j \times \mathbb{R}$  to  $\mathbb{R}$ ;  $\Upsilon$  and  $\Theta_j$  are the parameter spaces of  $\alpha$ ,  $\theta_j$  for  $j = 1, \dots, k+1$ , respectively. Simple calculation gives

$$\ell(\alpha, \theta_1, \dots, \theta_{k+1}, n_1, \dots, n_k) = \frac{1}{n} \sum_{j=1}^{k+1} \sum_{i=n_{j-1}+1}^{n_j} \frac{g_j(\alpha, \theta_j, Y_i) \delta_i}{S_C^{n_j}(Y_i^-)}, \quad (1.3)$$

where  $S_C^{n_j}(\cdot)$  is the Kaplan-Meier product limit estimator of  $1 - G_{n_j}(\cdot)$ , for each sample  $C_{n_{j-1}+1}, \dots, C_{n_j}$ ,  $j = 1, \dots, k+1$ .

Our result is generalization for the work of [He and Severini \(2010\)](#) in the sense that we consider the  $M$ -estimation in the censored data setting. [He and Severini \(2010\)](#) investigated statistical models with multiple change-points and established the theoretical properties of the maximum likelihood estimators. Their results are not directly applicable here since we consider more general framework. These results are not only useful in their own right but essential to establish the theoretical properties of our estimators. Under no censoring, there are a number of results available on the asymptotic properties of parameter estimators in change-point models with  $g_j(\alpha, \theta_j, x) = \log f_j(\alpha, \theta_j, x)$ . See, for example, [Hinkley \(1970, 1972\)](#), [Hinkley and Hinkley \(1970\)](#), [Bhattacharya \(1987\)](#), [Fu and Curnow \(1990a,b\)](#), [Jandhyala and Fotopoulos \(2001, 1999\)](#) and [Hawkins \(2001\)](#); the two monographs [Chen and Gupta \(2000\)](#) and [Csörgő and Horváth \(1997\)](#), and for the  $M$ -estimators we refer to [Hušková \(1996\)](#). In [Gombay and Horváth \(1994\)](#), a maximum-likelihood-type statistic is proposed for testing a sequence of observations for no change in the parameter against a possible change, this work is extended to the semi-parametric setting in [Bouzebda and Keziou \(2013\)](#) and [Bouzebda \(2014\)](#). It is worth noticing that  $M$ -estimators include the least squares estimators, several robust version of means and notably their predecessor, the maximum likelihood estimate (MLE) with  $g_j(\alpha, \theta_j, \cdot) = \log f_j(\alpha, \theta_j, \cdot)$ ,  $f(\cdot)$  being the probability density function. Strong consistency of  $M$ -estimators can be verified as that of the MLEs, and it is possible to avoid the differentiability condition of the density function  $f_j(\alpha, \theta_j, x)$  in the MLE case. This approach was first employed by [Wald \(1949\)](#) and later extended, for example, by [LeCam \(1953\)](#), [Kiefer and Wolfowitz \(1956\)](#), [Bahadur \(1967\)](#), [Huber \(1967\)](#), [Pfanzagl \(1969\)](#) and [Perlman \(1972\)](#) among others. Asymptotic properties of Huber's  $M$ -estimators based on complete data are well understood nowadays and can be found, for example, in [Huber \(1981\)](#) and [van der Vaart \(1998\)](#), among others.

In the presence of censoring very little is known about the general large sample properties of  $M$ -estimators. [Reid \(1981\)](#) derived the influence function and the asymptotic normality of a truncated type  $M$ -estimator. (Some modifications are required in Reid's arguments, cf. [Andersen et al. \(2012\)](#)). [Oakes \(1986\)](#) considered  $M$ -estimators with  $g_j(\alpha, \theta_j, \cdot) = \log f_j(\alpha, \theta_j, \cdot)$  and called them approximate MLEs since the corresponding  $M$ -estimators are no longer the MLEs. [Borgan \(1984\)](#) studied the asymptotic properties of the MLE. Another type of  $M$ -estimator, based on the cumulative hazard function and aiming at inclusion of the MLEs under censoring is discussed in [Hjort \(1985\)](#). [Wang \(1995\)](#) has established the strong consistency of this type of estimators under general conditions which can be applied to parametric, semi-and non-parametric models.

The main objective of our paper is to provide a full theoretical justification of the consistency of  $M$ -estimators of the parameters of a general class of multiple change-points models and gives the asymptotic distribution of the parameters of the within-segment distributions. This requires the effective application of large sample theory techniques, which were developed for the empirical processes, refer to Section 4 where we have used results from the work of [Pakes and Pollard \(1989\)](#).

The article is structured as follows. Section 2 is devoted to the statement of our notations and assumptions. In Section 3, the asymptotic properties of our estimators are derived. The general theory of the  $Z$ -estimators is considered in Section 4. In Section 5, we specify the estimation procedure for the maximum likelihood. The finite sample performance of the latter is illustrated by means of Monte Carlo simulations in Section 6. Some concluding remarks are given in Section 7. To avoid interrupting the flow of the presentation, all mathematical developments are relegated to Section 8. Section 9 gives some basic definitions and preliminaries needed to state our results.

## 2 Notation and assumptions

In this section, we introduce notation needed to state the asymptotic results in Section 3. The parameter spaces  $\Upsilon$  and  $\Theta_j$  are the subset of  $\mathbb{R}^d$  and  $\mathbb{R}^{d_j}$  respectively. Let

$$\begin{aligned}\lambda_j &= \frac{n_j}{n}, \text{ for any } j = 1, \dots, k, \\ \lambda_j^0 &= \frac{n_j^0}{n}, \text{ for any } j = 1, \dots, k, \\ \boldsymbol{\lambda} &= (\lambda_1, \lambda_2, \dots, \lambda_k), \\ \boldsymbol{\lambda}^0 &= (\lambda_1^0, \lambda_2^0, \dots, \lambda_k^0), \\ \boldsymbol{\theta} &= (\theta_1, \theta_2, \dots, \theta_{k+1}), \\ \boldsymbol{\theta}^0 &= (\theta_1^0, \theta_2^0, \dots, \theta_{k+1}^0), \\ \boldsymbol{\phi} &= (\alpha, \theta_1, \theta_2, \dots, \theta_{k+1}), \\ \boldsymbol{\phi}^0 &= (\alpha^0, \theta_1^0, \theta_2^0, \dots, \theta_{k+1}^0), \\ S_{F_{n_j^0}}(\cdot) &= 1 - F_{n_j^0}(\cdot), \\ S_{G_{n_j^0}}(\cdot) &= 1 - G_{n_j^0}(\cdot).\end{aligned}$$

We have for each  $j = 1, \dots, k$ ,

$$1 - H_{n_j^0}(\cdot) = (1 - G_{n_j^0}(\cdot))(1 - F_{n_j^0}(\cdot)).$$

Let  $\tau_{F_{n_j^0}}(\cdot)$  (resp.  $\tau_{G_{n_j^0}}(\cdot)$ ) be the upper bound of the support of  $F_{n_j^0}(\cdot)$  (resp.  $G_{n_j^0}(\cdot)$ ). Note that  $\boldsymbol{\lambda}^0$  is taken to be a constant vector as  $n$  goes to infinity. Let  $\boldsymbol{\Lambda}$  be the set of the configurations of change-points and  $\boldsymbol{\Phi}$  the set of parameters,

$$\begin{aligned}\boldsymbol{\Lambda} &= \{(\lambda_1, \lambda_2, \dots, \lambda_k) : \lambda_j = \frac{n_j}{n}, j = 1, \dots, k, 0 < n_1 < \dots < n_k < n\}, \\ \boldsymbol{\Phi} &= \Theta_1 \times \Theta_2 \times \dots \times \Theta_{k+1} \times \Upsilon.\end{aligned}$$

The criterion function computed over the segment  $j$  of  $\boldsymbol{\lambda}$  is defined by

$$\mathcal{G}_n(Y_j, \theta_j, \alpha) = \frac{(n_j - n_{j-1})}{n} \int_{\mathbb{R}} g_j(\alpha, \theta_j, x) d\widehat{F}_{n_j}(x).$$

Consequently, we can rewrite the function  $\ell$  given in (1.1) as

$$\ell = \sum_{j=1}^{k+1} \mathcal{G}_n(Y_j, \theta_j, \alpha).$$

Estimators of all change-points, all within-segment parameters and the common parameter are defined by maximization of the function  $\ell$  in  $\boldsymbol{\Lambda} \times \boldsymbol{\Phi}$ , i.e.,

$$(\widehat{\alpha}, \widehat{\theta}_1, \dots, \widehat{\theta}_{k+1}, \widehat{n}_1, \dots, \widehat{n}_k) = \underset{0 < n_1 < n_2 < \dots < n; \theta_j \in \Theta_j, 1 \leq j \leq k+1, \alpha \in \Upsilon}{\operatorname{argmax}} \ell. \quad (2.1)$$

For a given configuration  $\boldsymbol{\lambda}$ ,  $(\widehat{\theta}_j(\lambda_j), \widehat{\alpha}(\lambda_j))$  maximizes  $\mathcal{G}_n(Y_j, \theta_j, \alpha)$ . We can remark that, when  $\boldsymbol{\lambda} = \boldsymbol{\lambda}^0$ , the estimate of  $(\boldsymbol{\theta}^0, \alpha^0)$  obtained by maximizing  $\ell(\alpha, \theta_1, \dots, \theta_{k+1}, n_1^0, \dots, n_k^0)$  converge to  $(\boldsymbol{\theta}^0, \alpha^0)$  under the Assumptions 2.1 and the first part of the Assumption 2.2 for complete data, by the result of van der Vaart (1998) and by add the first part of Assumption 2.5, we get the convergence for censored data by the result of Wang (1995). In the case where the change point fraction  $\boldsymbol{\lambda}^0$  is unknown, the  $M$ -estimators  $(\widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\theta}}, \widehat{\alpha})$  is the value of  $(\boldsymbol{\lambda}, \boldsymbol{\theta}, \alpha)$  that maximizes  $\ell(\alpha, \theta_1, \dots, \theta_{k+1}, n_1^0, \dots, n_k^0)$  in  $\boldsymbol{\Lambda} \times \boldsymbol{\Phi}$ . Thus  $(\widehat{\theta}_j, \widehat{\alpha}) \stackrel{\text{def}}{=} (\widehat{\theta}_j(\widehat{\lambda}_j), \widehat{\alpha}(\widehat{\lambda}_j))$  is the  $M$ -estimator of  $(\theta_j^0, \alpha_j^0)$  computed in the segment  $j$  of the estimated configuration of change-points  $\widehat{n}_j$ , refer for similar arguments to Lavielle and Ludeña (2000). Let us introduce

$$L^0(\alpha, \theta_1, \dots, \theta_{k+1}) = \sum_{j=1}^{k+1} \frac{(n_j^0 - n_{j-1}^0)}{n} \int_{\mathbb{R}} g_j(\alpha, \theta_j, x) dF_{n_j^0}(x), \quad (2.2)$$

where  $F_{n_j^0}(\cdot)$  (respectively  $G_{n_j^0}(\cdot)$ ) is the true function of distribution for the sample  $X_{n_{j-1}^0+1}, \dots, X_{n_j^0}$  (resp.  $C_{n_{j-1}^0+1}, \dots, C_{n_j^0}$ ),  $j = 1, \dots, k+1$ . The following decomposition will play an instrumental role in the proofs of Theorem 3.1 and Theorem 3.3. Define a function  $\mathbf{W}'$  by

$$\begin{aligned} \mathbf{W}' &= \sum_{j=1}^{k+1} \sum_{i=1}^{k+1} \frac{n_{ji}}{n} \left\{ \int [g_j(\alpha, \theta_j, x) - g_i(\alpha^0, \theta_i^0, x)] dF_{n_i^0} \right\} \\ &\quad + \frac{1}{n} \sum_{j=1}^{k+1} \sum_{i=n_{j-1}^0+1}^{n_j} \left\{ \frac{g_j(\alpha, \theta_j, Y_i) \delta_i}{S_C^{n_j}(Y_i^-)} - \mathbb{E}(g_j(\alpha, \theta_j, X_i)) \right\} \\ &\quad - \frac{1}{n} \sum_{j=1}^{k+1} \sum_{i=n_{j-1}^0+1}^{n_j^0} \left\{ \frac{g_j(\alpha^0, \theta_j^0, Y_i) \delta_i}{S_C^{n_j^0}(Y_i^-)} - \mathbb{E}(g_j(\alpha^0, \theta_j^0, X_i)) \right\}, \end{aligned} \quad (2.3)$$

where  $n_{ji}$  is the number of observations of the interested variables in the set

$$[n_{j-1} + 1, n_j] \cap [n_{i-1}^0 + 1, n_i^0],$$

for  $i, j = 1, \dots, k+1$ . We obviously have that

$$\operatorname{argmax}_{0 < n_1 < n_2 < \dots < n; \theta_j \in \Theta_j, 1 \leq j \leq k+1, \alpha \in \Upsilon} \ell = \operatorname{argmax}_{0 < n_1 < n_2 < \dots < n; \theta_j \in \Theta_j, 1 \leq j \leq k+1, \alpha \in \Upsilon} \mathbf{W}';$$

thus, the  $M$ -estimators may be defined as the maximizers of  $\mathbf{W}'$  rather than as the maximizers of  $\ell$ . Our idea is to replace EKM  $S_C^{n_j}(\cdot)$  in (2.3) by the theoretical survival function  $S_{G_{n_j^0}}(\cdot)$  and to prove the difference between the EKM based on the estimated survival function and the EKM based on the theoretical survival function is negligible, in probability, as  $n$  goes to infinity, see (8.3). Notice that  $S_C^{n_j^0}(\cdot)$  converges to  $S_{G_{n_j^0}}(\cdot)$ , so we can replace the EKM, at the price of some complicated calculations. Let  $b(\alpha, \theta_j, \alpha^0, \theta_i^0)$  be defined by

$$\begin{aligned} b(\alpha, \theta_j, \alpha^0, \theta_i^0) &= \mathbb{E}(g_j(\alpha, \theta_j, X_i)) - \mathbb{E}(g_i(\alpha^0, \theta_i^0, X_i)) \\ &= \int_{\mathbb{R}} [g_j(\alpha, \theta_j, x) - g_i(\alpha^0, \theta_i^0, x)] dF_{n_i^0}(x), \end{aligned} \quad (2.4)$$

for  $i, j = 1, \dots, k+1$ . We substitute  $\mathbf{W}'$  by  $\mathbf{W}$  after replacing the EKM by its true survival function and we define

$$\mathbf{W} = \mathbf{W}_1 + \mathbf{W}_2,$$

where

$$\mathbf{W}_1 = \sum_{j=1}^{k+1} \sum_{i=1}^{k+1} \frac{n_{ji}}{n} b(\alpha, \theta_j, \alpha^0, \theta_i^0) \quad (2.5)$$

and

$$\begin{aligned} \mathbf{W}_2 &= \frac{1}{n} \sum_{j=1}^{k+1} \sum_{i=n_{j-1}^0+1}^{n_j} \left\{ \frac{g_j(\alpha, \theta_j, Y_i) \delta_i}{S_{G_{n_j^0}}(Y_i^-)} - \mathbb{E}(g_j(\alpha, \theta_j, X_i)) \right\} \\ &\quad - \frac{1}{n} \sum_{j=1}^{k+1} \sum_{i=n_{j-1}^0+1}^{n_j^0} \left\{ \frac{g_j(\alpha^0, \theta_j^0, Y_i) \delta_i}{S_{G_{n_j^0}}(Y_i^-)} - \mathbb{E}(g_j(\alpha^0, \theta_j^0, X_i)) \right\}. \end{aligned}$$

Alternatively, we may write

$$\begin{aligned} \mathbf{W}_2 &= \frac{1}{n} \sum_{j=1}^{k+1} \sum_{i=1}^{k+1} \left\{ \sum_{t \in \bar{n}_{ji}} \left[ \frac{g_j(\alpha, \theta_j, Y_t) \delta_t}{S_{G_{n_j^0}}(Y_t^-)} - \mathbb{E}(g_j(\alpha, \theta_j, X_t)) \right] \right. \\ &\quad \left. - \sum_{t \in \bar{n}_{ji}} \left[ \frac{g_i(\alpha^0, \theta_i^0, Y_t) \delta_t}{S_{G_{n_i^0}}(Y_t^-)} - \mathbb{E}(g_i(\alpha^0, \theta_i^0, X_t)) \right] \right\}, \end{aligned} \quad (2.6)$$

where

$$\tilde{n}_{ji} = [n_{j-1} + 1, n_j] \cap [n_{i-1}^0 + 1, n_i^0].$$

We note that in the particular case where  $g_j(\cdot) = \log f_j(\cdot)$ , we get  $\mathbf{W}_1$  is a weighted sum of the negative Kullback-Leibler distances, and  $\mathbf{W}_2 \rightarrow 0$  as  $n \rightarrow 0$ , by applying Proposition 9.1.

In our analysis, the following assumptions will be needed.

**Assumption 2.1** 1. Assume that for  $j = 1, \dots, k + 1$ ,

$$g_{j+1}(\alpha^0, \theta_{j+1}^0, x) \neq g_j(\alpha^0, \theta_j^0, x)$$

on a set of non-zero measure.

2. For any  $j = 1, \dots, k + 1$ , any  $\alpha, \theta_j$ ; for  $i = 1, \dots, k + 1$ ,

$$\int_{\mathbb{R}} (g_j(\alpha, \theta_j, x)) dF_{n_i^0}(x) \leq \int_{\mathbb{R}} (g_i(\alpha^0, \theta_i^0, x)) dF_{n_i^0}(x).$$

The first part of this assumption guarantees that the distributions in two neighboring segments are different. Clearly, this is required for the change-points to be well defined, and the second part is to ensure that the expectation of the function associates with the true parameters is the maximum in the true sample, when we consider the particular case  $g_j(\cdot) = \log f_j(\cdot)$ , this assumption comes directly from the distance of Kullback-Leibler, for further details, we refer to [He and Severini \(2010\)](#), or when the function  $g(\cdot)$  is independent of the index  $j$ , i.e., the same function of all segments for example when the variables are assumed to be from normal distribution and there is a change in variances and having the same mean, or conversely, so we have all parameters are in the same set, i.e.,  $\theta_j \in \Theta$  for any  $j = 1, 2, \dots, k + 1$ , **for the uncensored case**, another example if the variables are assumed to follow the Weibull's distribution. In the  $M$ -estimation theory, this condition is required to ensure that the true parameters are the points that maximizes the criterion function. For more details see also [van der Vaart and Wellner \(1996\)](#).

**Assumption 2.2** Assume that

1. for  $j = 1, \dots, k + 1$ ,  $\theta_j$  and  $\theta_j^0$  are contained in  $\Theta_j$ , where  $\Theta_j$  is a compact subset of  $\mathbb{R}^{d_j}$ ;  $\alpha$  and  $\alpha^0$  are contained in  $\Upsilon$ , where  $\Upsilon$  is a compact subset of  $\mathbb{R}^d$ ; here  $d, d_1, \dots, d_{k+1}$  are non-negative integers.
2.  $\ell(\alpha, \theta)$  is second-order continuously differentiable with respect to  $\alpha, \theta$ , and there is an interchangeability of integration and differentiation in (2.2).

Compactness of the parameter space is used to insure that the maximum is achievable and to establish the consistency of the  $M$ -estimators of

$$\frac{n_1}{n}, \dots, \frac{n_k}{n}, \theta_1, \dots, \theta_{k+1}, \alpha,$$

for discussions and details on this condition and its necessity in general model, the reader can refer to [Huber \(1981\)](#) for complete data and [Wang \(1995\)](#) for censored data. Differentiability of the given function is used to justify some Taylor series expansions, interchangeability of integration and differentiation is technical assumption used for the variance expression in (4.5). The second part of the Assumption 2.2 ensures the existence of the variance of the  $M$ -estimates. Both parts of Assumption 2.2 are relatively weak and are essentially the same as conditions used in parametric models for censored data without change-points, see [Wang \(1999\)](#).

**Assumption 2.3** Assume that

1. for any  $j = 1, \dots, k + 1$  and any integers  $s, t$  satisfying  $0 \leq s < t \leq n$ ,

$$\mathbb{E} \left( \max_{\theta_j \in \Theta_j, \alpha \in \Upsilon} \left[ \sum_{i=s+1}^t \left( \sum_{z=1}^{k+1} \frac{g_j(\alpha, \theta_j, Y_i) \delta_i}{S_{G_{n_z}}(Y_i^-)} \mathbb{I}_{\{n_{z-1}+1 \leq i \leq n_z\}} - \mathbb{E}(g_j(\alpha, \theta_j, X_i)) \right) \right]^2 \right) \leq C(t-s)^r,$$

where  $r < 2$  and  $C$  is a constant.

2. for any  $j = 1, \dots, k + 1$  and any integers  $s, t$  satisfying  $n_{j-1}^0 \leq s < t \leq n_j^0$ ,

$$\mathbb{E} \left( \max_{\theta_j \in \Theta_j, \alpha \in \Upsilon} \left[ \sum_{i=s+1}^t \left( \sum_{z=1}^{k+1} \frac{g_j(\alpha, \theta_j, Y_i) \delta_i}{S_{G_{n_j^0}}(Y_i^-)} \mathbb{I}_{\{n_{z-1}+1 \leq i \leq n_z\}} - \frac{g_j(\alpha^0, \theta_j^0, Y_i) \delta_i}{S_{G_{n_j^0}}(Y_i^-)} - b(\alpha, \theta_j, \alpha^0, \theta_j^0) \right) \right]^2 \right) \leq D(t-s)^r,$$

where  $b(\alpha, \theta_j, \alpha^0, \theta_j^0)$  is introduced in equation (2.4),  $r < 2$  and  $D$  is a constant.

Parts 1 and 2 of Assumption 2.3 are technical requirements on the behavior of the function  $g_j(\cdot)$  between and within segments, respectively. This condition is used to ensure that the information regarding the within- and between-segment parameters grows quickly enough to establish consistency and asymptotic normality of the parameter estimators. Note that where  $g_j(\cdot) = \log f_j(\cdot)$  these conditions are relatively weak; it is easy to check that they are satisfied by at least all distributions in the exponential family, for more details refer to He and Severini (2010).

**Assumption 2.4** 1. The parameter  $\phi^0$  is the unique root of  $\rho(\phi) = 0$ .  
2. The matrix  $C(\phi^0)$  defined in (4.3) is finite.

**Assumption 2.5** 1. Assume that (R1), in the appendix, hold for  $\tau_{F_{n_j^0}}$  and  $\tau_{G_{n_j^0}}$  for any  $j = 1, 2, \dots, k + 1$ .  
2. Assume that (R2) and (R3), in the appendix, hold for any  $j = 1, 2, \dots, k + 1$  when we replace  $\varphi$  by  $\psi_{j(l)}$ ,  $1 \leq l \leq d + d_1 + \dots + d_{k+1}$ ,  $\gamma_0(\cdot)$  by  $\gamma_{j0}(\cdot)$ ,  $H_1(\cdot)$  by  $H_{j1}(\cdot)$ ,  $C(x)$  by  $C_j(x)$  and  $F(\cdot)$  by  $F_{n_j^0}(\cdot)$ .

**Assumption 2.6** Assume that for every  $j = 1, \dots, k$  and for  $t > 0$ ;  $S_C^{n_j}(t) > 0$  and  $S_{G_{n_j^0}}(t) > 0$ .

The first part of the Assumption 2.4 is quite classical condition in the  $Z$ -estimation theory. The second part is used to justify the existence of variance-covariance expression. We use the Assumption 2.5 for the SLLN and CLT of each true sub-sample in the presence of censoring. Assumption 2.6 is imposed to justify the finiteness of some expressions when we have  $S_C^{n_j}(\cdot)$  and  $S_{G_{n_j^0}}(\cdot)$  in the denominator for each  $j$ .

### 3 Asymptotic results

In this section, we establish the consistency of the  $M$ -estimators by using the argmax theorem in van der Vaart and Wellner (1996). For reader convenience, let us recall the basic idea. If the argmax functional is continuous with respect to some metric on the space of the criterion functions, then convergence in distribution of the criterion functions will imply the convergence in distribution of their points of maximum, the  $M$ -estimators, to the maximum of the limit criterion function. So in this section we will give our first main result; the weak consistency of the estimators  $\hat{\alpha}, \hat{\theta}_1, \dots, \hat{\theta}_{k+1}, \hat{\lambda}_1, \dots, \hat{\lambda}_k$ , which it will be considered as an initial step for the next results, where we will treat the rate of convergence and the asymptotic distribution of the estimators  $\hat{\alpha}, \hat{\theta}_1, \dots, \hat{\theta}_{k+1}$ . The results presented in this section extends and complements the theory of He and Severini (2010) in several ways. On the first hand, when all the data are observed and the criterion function is replaced by the probability density function, i.e.,  $g_j(\cdot) = \log f_j(\cdot)$ , our Theorem 3.1 becomes their Theorem 2.1 and our Theorem 3.3 becomes their Theorem 2.2. On the other hand, we consider the censored data setting in semi-parametric models that is quite different from the framework of the last mentioned reference. Let us recall that the estimators  $(\hat{\alpha}, \hat{\theta}_1, \dots, \hat{\theta}_{k+1}, \hat{n}_1, \dots, \hat{n}_k)$  are defined in equation (2.1).

The following theorem gives the consistency of the model's parameters estimators  $(\hat{\alpha}, \hat{\theta}_1, \dots, \hat{\theta}_{k+1}, \hat{n}_1, \dots, \hat{n}_k)$ .

**Theorem 3.1 (Consistency)** Under Assumption 3.1, part 1 of Assumption 3.2, part 1 of Assumption 3.3 and Assumption 3.6, we have, as  $n \rightarrow \infty$ ,

$$\hat{\lambda}_i \xrightarrow{\mathbb{P}} \lambda_i^0, \quad \hat{\theta}_j \xrightarrow{\mathbb{P}} \theta_j^0 \quad \text{and} \quad \hat{\alpha} \xrightarrow{\mathbb{P}} \alpha^0,$$

where

$$\hat{\lambda}_i = \frac{\hat{n}_i}{n}$$

for  $i = 1, \dots, k$  and  $j = 1, \dots, k + 1$ .

**Remark 3.2** It is worth noting that  $\hat{n}_i$ ,  $i = 1, \dots, k$  are not consistent. Here we consider the consistency of the change point fractions  $\hat{\lambda}_i$ ,  $i = 1, \dots, k$ , in a similar spirit as in [Hinkley \(1970\)](#). The weak consistency of the parameters  $\hat{\alpha}$  and  $\hat{\theta}_j$ ,  $j = 1, \dots, k+1$  is based on the classical  $M$ -estimators techniques for the censored data in the complex setting of the multiple change-points models.

The proof of this theorem is based on the proof of Theorem 3.1 in [He and Severini \(2010\)](#). The proof of Theorem 3.1 is captured in the forthcoming Sect. 8.

The following theorem give the convergence rate of the estimator  $\hat{\lambda}_1, \dots, \hat{\lambda}_k$  the change-points coefficients  $\lambda_1, \dots, \lambda_k$ .

**Theorem 3.3 (Convergence rate)** Under Assumption 3.1, part 1 of Assumption 3.2, Assumption 3.3 and Assumption 3.6, we have

$$\lim_{\eta \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P} \left( n \left\| \hat{\lambda} - \lambda^0 \right\|_{\infty} \geq \eta \right) = 0,$$

where

$$\hat{\lambda} = (\hat{\lambda}_1, \dots, \hat{\lambda}_k), \quad \left\| \hat{\lambda} - \lambda^0 \right\|_{\infty} = \max_{1 \leq j \leq k} \left| \hat{\lambda}_j - \lambda_j^0 \right|.$$

That is, for  $i = 1, 2, \dots, k$ ,

$$\hat{\lambda}_i - \lambda_i^0 = O_{\mathbb{P}}(n^{-1}).$$

The proof of this theorem is based of the proof of Theorem 2.2 in [He and Severini \(2010\)](#). The proof of Theorem 3.3 is captured in the forthcoming Sect. 8.

**Remark 3.4** The proof of the asymptotic distribution of  $\hat{\lambda}_1, \dots, \hat{\lambda}_k$ , should require a complex methodology, and we leave this problem open for future research.

**Remark 3.5** In the comparison of the nonparametric regression estimators, [Korostel'ev and Tsybakov \(1993\)](#) argued that the minimax approach is one of the correct ways. [Raimondo \(1998\)](#) considered the sharp change-point problem as an extension of earlier problems in change-point analysis related to the nonparametric regression. [Raimondo \(1998\)](#) proposed a test function for the local regularity of a signal that characterizes such a point as a global maximum and developed a suboptimal wavelet estimator. [Goldenshluger et al. \(2008\)](#) considered the problem of nonparametric estimation of signal change-points from indirect and noisy observations, where the estimation problem is analyzed in a general minimax framework. The authors provide lower bounds for minimax risks and propose rate-optimal estimation procedures, one can refer to the last reference for more details on the subject. [Shiryayev \(2016\)](#) considered the change-point quickest detection problem for Brownian motion. The minimax test proposed by [Lorden \(1971\)](#), is used to solve this problem. An original complete and remarkable proof of the CUSUM statistics optimality is constructed and given in detail. [Pergamenchtchikov and Tartakovsky \(2019\)](#) established very general conditions for some models under which the weighted Shiryayev-Roberts procedure is asymptotically optimal, in the minimax sense. In the setting of the multiple change-points when the number of change-points is known, [Bai and Perron \(1998\)](#) obtained the rate  $1/n$  in the multiple linear regression setting, even the least-squares estimator is consistent with the optimal rate  $1/n$ ; see [Hao et al. \(2013\)](#) and the references therein. Using the maximum likelihood estimators, [He and Severini \(2010\)](#), obtained the same rate, while in the nonparametric maximum likelihood approach [Dumbgen et al. \(1991\)](#) showed that the optimal rate is  $1/n$  in the single change-point setting, which is generalized by [Zou et al. \(2014a\)](#) when they fixed the number of change-points. Notice that the rate  $1/n$  obtained in Theorem 3.3 is the minimax rate when the number of change-points is known. The rate convergence  $1/n$  of the estimated change-points fractions plays a crucial role to obtain standard root- $n$  asymptotic normality of the estimated parameter  $\hat{\phi}$ .

## 4 Z-estimators

In this section, we give the  $Z$ -estimators of  $\phi$  when the functions  $g_j(\cdot)$  are differentiable in  $\phi$ , in two step the first step is maximizing the equation (1.1) in  $n_j$ ,  $j = 1, 2, \dots, k$ , and in the second step, we find the solution to the estimating equation given by

$$\begin{aligned} \rho_n(\alpha, \theta_1, \dots, \theta_{k+1}) &= \frac{\partial \ell}{\partial \phi} = \sum_{j=1}^{k+1} \frac{(\hat{n}_j - \hat{n}_{j-1})}{n} \int_{\mathbb{R}} \psi_j(\alpha, \theta_j, x) d\hat{F}_{\hat{n}_j}(x), \\ &= \sum_{j=1}^{k+1} (\hat{\lambda}_j - \hat{\lambda}_{j-1}) \rho_{\hat{n}_j}(\alpha, \theta_j), \end{aligned} \quad (4.1)$$

where  $\widehat{n}_j$  is the maximizers of  $n_j$  and  $\psi_j(\alpha, \theta_j, x) = \frac{\partial g_j(\alpha, \theta_j, x)}{\partial \phi_i}$ ,  $i = 1, \dots, k+2$ , from  $\Upsilon \times \Theta_j \times \mathbb{R}$  to  $\mathbb{R}^{d+d_1+\dots+d_{k+1}}$ ; satisfies

$$\rho(\alpha^0, \theta_1^0, \dots, \theta_{k+1}^0) = \sum_{j=1}^{k+1} \frac{(n_j^0 - n_{j-1}^0)}{n} \int_{\mathbb{R}} \psi_j(\alpha^0, \theta_j^0, x) dF_{n_j^0}(x) = 0,$$

and, for each  $j = 1, 2, \dots, k+1$ ,

$$\rho_{n_j^0}(\alpha^0, \theta_j^0) = \int_{\mathbb{R}} \psi_j(\alpha^0, \theta_j^0, x) dF_{n_j^0}(x) = 0.$$

Let

$$\begin{aligned} \rho_n^0(\alpha, \theta_1, \dots, \theta_{k+1}) &= \frac{\partial \ell^0}{\partial \phi} = \sum_{j=1}^{k+1} \frac{(n_j^0 - n_{j-1}^0)}{n} \int_{\mathbb{R}} \psi_j(\alpha, \theta_j, x) d\widehat{F}_{n_j^0}(x). \\ &= \sum_{j=1}^{k+1} (\lambda_j^0 - \lambda_{j-1}^0) \rho_{n_j^0}^0(\alpha, \theta_j). \end{aligned}$$

Notice that  $Z$ -estimators include the maximum likelihood estimators, when

$$\psi_j(\phi, x) = \frac{\partial \log f_j(\phi, x)}{\partial \phi},$$

where  $f(\cdot)$  is the density function, generalized method of moment estimators when

$$\psi_j(\phi, x) = h(x) - \mathbb{E}_{\phi} h(x),$$

for some function  $h(\cdot)$ , asymptotic properties are given in [Huber \(1981\)](#), [Serfling \(1980\)](#), [van der Vaart and Wellner \(1996\)](#) and [van der Vaart \(1998\)](#) among others. For the censored data, the case

$$\psi_j(\phi, x) = \frac{\partial \log f_j(\phi, x)}{\partial \phi},$$

no longer correspond to the maximum likelihood estimators. [Oakes \(1986\)](#) referred to this particular type of  $Z$ -estimator as the approximate maximum likelihood estimators and points out its computational and potential robustness advantages over the classical maximum likelihood estimators. [Wang \(1999\)](#) has established the strong consistency of this type of estimators. The asymptotic normality is obtained, under restrictive conditions, by [Reid \(1981\)](#). [Wang \(1999\)](#) established general asymptotic normality results, which are comparable to those in [Cramér \(1946\)](#), [Huber \(1967\)](#) and subsequent work, he provided the influence curves of a  $Z$ -estimator. In this section, we give the asymptotic results and the rate of convergence of  $Z$ -estimators under censored data in models with multiple change-points, after approximating the points of change and giving the general conditions for the asymptotic normality, similar to those considered in [Wang \(1999\)](#). The main hurdle for the full development of the asymptotic properties of  $Z$ -estimators is the work of [Stute \(1995\)](#) obtained the most general CLT for  $\int \varphi d\widehat{F}_n$  with an arbitrary function  $\varphi(\cdot)$ . For any  $j = 1, 2, \dots, k+1$ , let  $\psi_{j(l)}(\alpha, \theta_j, \cdot)$  denote the  $l$ -th component of  $\psi_j(\alpha, \theta_j, \cdot)$ . Replace  $\varphi(\cdot)$  by  $\psi_{j(l)}(\alpha, \theta_j, \cdot)$  in (9.3) and (9.5),  $H_0(\cdot)$  (resp.  $H_1(\cdot)$ ,  $H_{pn}(\cdot)$ ) by  $H_{j0}(\cdot)$  (resp.  $H_{j1}(\cdot)$ ,  $H_{j,pn}(\cdot)$ ) in (9.1) and (4.2) where

$$H_{j,pn}(y) = \frac{1}{n_j^0 - n_{j-1}^0} \sum_{i=n_{j-1}^0}^{n_j^0} \mathbb{I}_{\{Y_i \leq y, \delta_i = p\}}, \quad \text{for } p = 0, 1, \quad (4.2)$$

$H(\cdot)$  (resp.  $F(\cdot)$ ,  $G(\cdot)$ ) by  $H_{n_j^0}(\cdot)$  (resp.  $F_{n_j^0}(\cdot)$ ,  $G_{n_j^0}(\cdot)$ ),  $C(\cdot)$  by  $C_j(\cdot)$  in (9.3) and (9.4) and denote the corresponding  $\gamma_i(\cdot)$ 's and  $U$  by  $\gamma_{ji(l)}(\cdot)$ ,  $i = 0, 1, 2$  and  $U(\psi_{j(l)})$  respectively. It now follows from Proposition 9.2, and the multivariate central limit theorem that,

$$\sqrt{n} \int_{\mathbb{R}} \psi_j(\alpha, \theta_j, x) d(\widehat{F}_{n_j^0} - F_{n_j^0})(x)$$

converges in distribution to a multivariate normal distribution with zero mean and covariance matrix  $C_j(\psi_j, \alpha, \theta_j, F_{n_j^0}, G_{n_j^0})$ , whose  $(i, l)$ -entry is

$$\begin{aligned} C_{j(i)}(\psi_j, \alpha, \theta_j, F_{n_j^0}, G_{n_j^0}) &= \mathbb{E}(U(\psi_{j(i)})U(\psi_{j(l)})) \\ &= \mathbb{E} \left\{ [\psi_{j(i)}(\alpha, \theta_j, Y)\gamma_{j0(i)}(Y)\delta + \gamma_{j1(i)}(Y)(1 - \delta) - \gamma_{j2(i)}(Y) \int_{\mathbb{R}} \psi_{j(i)}(\alpha, \theta_j, x) dF(x)] \right. \\ &\quad \left. [\psi_{j(l)}(\alpha, \theta_j, Y)\gamma_{j0(l)}(Y)\delta + \gamma_{j1(l)}(Y)(1 - \delta) - \gamma_{j2(l)}(Y) - \int_{\mathbb{R}} \psi_{j(l)}(\alpha, \theta_j, x) dF(x)] \right\}. \end{aligned}$$

Let

$$C(\phi) = \sum_{j=1}^{k+1} (\lambda_j^0 - \lambda_j^0) C_j(\psi_j, \alpha, \theta_j, F_{n_j^0}, G_{n_j^0}), \quad (4.3)$$

and

$$\frac{\partial}{\partial \phi} \psi_j(\alpha, \theta_j, x) = \left( \frac{\partial}{\partial \phi_l} \psi_{j(i)}(\alpha, \theta_j, x) \right)_{il},$$

denote the  $(d + d_1 + \dots + d_{k+1}) \times (d + d_1 + \dots + d_{k+1})$  derivative matrix of  $\psi$  with respect to  $\phi$ , let  $\Gamma_{F_{n_j^0}}(t)$  and  $\Gamma(t)$  denote the  $(d + d_1 + \dots + d_{k+1}) \times (d + d_1 + \dots + d_{k+1})$  matrix with

$$\begin{aligned} \Gamma_{F_{n_j^0}}(t) &= \int \frac{\partial}{\partial \phi} \psi_j(\alpha, \theta_j, x) |_{\phi=t} dF_{n_j^0}(x), \\ \Gamma(t) &= \sum_{j=1}^{k+1} \frac{n_j^0 - n_{j-1}^0}{n} \Gamma_{F_{n_j^0}}(t), \end{aligned} \quad (4.4)$$

$$\Sigma = [\Gamma(\phi^0)]^{-1} C(\phi^0) [\Gamma(\phi^0)^\top]^{-1}, \quad (4.5)$$

where  $A^\top$  denotes the transpose of a matrix  $A$ .

The following theorem gives the consistency of  $\hat{\phi}$ .

**Theorem 4.1** *Under the Assumptions of Theorem 3.3, the function  $\rho(\cdot)$  is continuous and for every  $\epsilon > 0$ , for  $n \rightarrow \infty$ ,*

$$\begin{aligned} \sup_{\phi \in \Phi} \|\rho_n^0(\phi) - \rho(\phi)\| &\xrightarrow{\mathbb{P}} 0, \\ \inf_{\phi: \|\phi^0\| \geq \epsilon} \|\rho(\phi)\| &> 0 = \|\rho(\phi^0)\|. \end{aligned}$$

Then any sequence of estimators  $\hat{\phi}$  such that  $\rho_n(\hat{\phi}) = o_{\mathbb{P}}(1)$  converges in probability to  $\phi^0$ .

The proof of Theorem 4.1 is captured in the forthcoming Sect. 8.

The conditions of the last theorem are given in [van der Vaart \(1998\)](#) when the data are complete and without change in distribution, here we give the conditions under the presence of censoring where we use the Kaplan-Meier integral, the first condition of this theorem is satisfied when the families

$$\mathcal{F}_j = \{\psi_j(\alpha, \theta_j, \cdot), \alpha \in \Upsilon, \theta_j \in \Theta_j\}$$

are Glivenko-Cantelli and the functions  $F_{n_j^0}(\cdot)$  are continuous for each  $j = 1, 2, \dots, k + 1$  for more detail see [Stute \(1995\)](#) and [Bae and Kim \(2003\)](#), compactness of the set  $\Phi$  and the continuity of  $\psi_j(\cdot)$  for any  $j = 1, 2, \dots, k + 1$  with the first part of Assumption 2.4 implies the condition 2 of Theorem 4.1.

In the next theorem, we will give weaker conditions than those in the previous theorem, these conditions are introduced in [Pakes and Pollard \(1989\)](#). Note that the first condition is to insure the estimator  $\hat{\phi}$  is taken as any value that comes close enough to provide a global minimum for  $\|\rho_n(\cdot)\|$ , since  $\phi^0$  is included in the set over which the minimum is taken,  $\|\rho_n(\hat{\phi})\|$  cannot be much bigger than  $\|\rho_n(\phi^0)\|$ . If the quantity  $\rho_n(\phi^0)$  is eventually close to zero, the second assumption on  $\rho(\phi^0)$  implies that  $\rho_n(\hat{\phi})$  must also get close to zero. If small values of  $\|\rho_n(\phi)\|$  can occur only near  $\phi^0$ , this forces  $\hat{\phi}$  to be close to  $\phi^0$  by the third condition.

**Theorem 4.2** *Under the following conditions*

(i)

$$\|\rho_n(\hat{\phi})\| \leq o_{\mathbb{P}}(1) + \inf_{\phi \in \Phi} \|\rho_n(\phi)\|;$$

(ii)

$$\rho_n(\phi^0) = o_{\mathbb{P}}(1);$$

(iii)

$$\sup_{\|\phi^0\| > \eta} \|\rho_n(\phi)\|^{-1} = O_{\mathbb{P}}(1) \quad \text{for each } \eta > 0.$$

Then any sequence of estimators  $\widehat{\phi}$  such that  $\rho_n(\widehat{\phi}) = o_{\mathbb{P}}(1)$  converges in probability to  $\phi^0$ .

The proof of Theorem 4.2 is captured in the forthcoming Sect. 8.

The next theorem gives conditions under which  $\widehat{\phi}$ , which is now assumed to converge in probability to  $\phi^0$ , satisfies a central limit theorem like a  $Z$ -estimator. The argument breaks naturally into two steps. First we establish  $\sqrt{n}$ -consistency by means of a comparison between  $\|\rho_n^0(\widehat{\phi})\|$  and  $\|\rho_n^0(\phi^0)\|$ . Informally stated, the new equicontinuity condition (iii) implies that

$$\|\rho(\phi)\| \leq O_{\mathbb{P}}(\|\rho_n(\phi)\|) + O_{\mathbb{P}}(\|\rho_n(\phi^0)\|) + o_{\mathbb{P}}(n^{-1/2})$$

uniformly near  $\phi^0$ . Since  $\widehat{\phi}$  comes close to minimizing  $\|\rho_n(\cdot)\|$ , the quantity  $\|\rho_n(\widehat{\phi})\|$  cannot be much larger than  $\|\rho_n(\phi^0)\|$ , which is of order  $O_{\mathbb{P}}(n^{-1/2})$ . Approximate linearity of  $\rho(\cdot)$  in a neighborhood of  $\phi^0$  transfers the same rate of convergence to  $\widehat{\phi} - \phi^0$ . The argument for the second step need only values of  $\phi$  in a  $O_{\mathbb{P}}(n^{-1/2})$  neighborhood of  $\phi^0$  (see page 1040 in Pakes and Pollard (1989)). The combination of conditions (ii) and (iii) shows that  $\rho_n^0(\cdot)$  is uniformly well approximated by a linear function  $L_n(\cdot)$ . The  $\phi_n^*$  that minimizes  $\|L_n(\cdot)\|$  has an explicit form, from which asymptotic normality of  $\sqrt{n}(\phi_n^* - \phi^0)$  is easily established. A comparison between  $\|\rho_n(\phi_n^*)\|$  and  $\|\rho_n^0(\widehat{\phi})\|$  shows that  $\widehat{\phi}$  must lie within  $O_{\mathbb{P}}(n^{-1/2})$  of  $\phi_n^*$ , which implies the desired central limit theorem.

The following theorem provides the central limit theorem for the estimator  $\widehat{\phi}$ .

**Theorem 4.3** Let  $\widehat{\phi}$  be a consistent estimator of  $\phi^0$ , under the Assumptions of Theorem 3.3, Assumption 2.4 and

- (i)  $\|\rho_n^0(\widehat{\phi})\| \leq o_{\mathbb{P}}(n^{-1/2})$ ;
- (ii)  $\rho(\cdot)$  is differentiable at  $\phi^0$  with a derivative matrix  $\Omega$  of full rank;
- (iii) for every sequence  $\eta_n$  of positive numbers that converges to zero,

$$\sup_{\|\phi - \phi^0\| < \eta_n} \frac{\|\rho_n^0(\phi) - \rho(\phi) - \rho_n^0(\phi^0)\|}{n^{-1/2} + \|\rho_n^0(\phi)\| + \|\rho(\phi)\|} = o_{\mathbb{P}}(1);$$

- (iv)  $\phi^0$  is an interior point of  $\Phi$ ,

then we have, as  $n \rightarrow \infty$ ,

$$\sqrt{n}(\widehat{\phi} - \phi^0) \xrightarrow{D} N(0, (\Omega^{-1})C(\phi^0)(\Omega^{-1})^{\top}).$$

The proof of Theorem 4.3 is captured in the forthcoming Sect. 8.

From Proposition 9.2 the central limit theorem follows. Note that if we can interchange between the integration and differentiation in (4.5), we take

$$\Omega = \Gamma(\phi^0).$$

The proof of Theorem 4.3 is similar to the proof in Pakes and Pollard (1989) but in our case,  $\rho_n^0(\cdot)$  (resp  $\rho_{n_j^0}^0(\cdot)$ ,  $j = 1, 2, \dots, k+1$ ) is not available, we have only  $\rho_n(\cdot)$  (respectively  $\rho_{\widehat{n}_j}(\cdot)$ ,  $j = 1, 2, \dots, k+1$ ), the result expression (8.4) in Lemma 8.4 gives us the asymptotic equivalence when  $n$  is large enough. The condition (i) and (iii) are automatically fulfilled when

(i)'

$$\left\| \rho_{n_j^0}^0(\alpha, \theta_j) \right\| \leq o_{\mathbb{P}}(n^{-1/2}), j = 1, 2, \dots, k+1;$$

(iii)'

$$\sup_{\|(\alpha, \theta_j) - (\alpha^0, \theta_j^0)\| < \eta_n} \left\| \rho_{n_j^0}^0(\alpha, \theta_j) - \rho_{n_j^0}(\alpha, \theta_j) - \rho_{n_j^0}^0(\alpha^0, \theta_j^0) \right\| = o_{\mathbb{P}}(n^{-1/2}), j = 1, 2, \dots, k+1.$$

Note that for the conditions (i) (resp (i)') and (iii) (resp (iii)') which they are assumed for  $\rho_n^0(\cdot)$  (respectively  $\rho_{n_j^0}^0(\cdot)$ ) the same under result in Lemma 8.4, we can show this conditions are required also for  $\rho_n(\cdot)$  (respectively  $\rho_{\widehat{n}_j}(\cdot)$ ) and conversely. In the next theorem, we give the asymptotic normality of  $\sqrt{n}(\widehat{\phi} - \phi^0)$  for  $\widehat{\phi}$  as an  $M$ -estimator or  $Z$ -estimator the proof is much similar.

**Theorem 4.4 (Asymptotic normality)** Under part 2 of Assumption 2.2 for  $\phi$  in a neighborhood of  $\phi^0$ , and let  $\Gamma(\phi^0)$  defined in (4.4) be a finite and non-singular  $(d + d_1 + \dots + d_{k+1}) \times (d + d_1 + \dots + d_{k+1})$  matrix. Assume that the assumptions of Lemma 8.5 with part 1 of Assumption 2.5 hold for

$$s(\phi, x) = \left( \frac{\partial}{\partial \phi_l} \psi_{j(i)}(\alpha, \theta_j, x) \right)_{il}, \quad 1 \leq i, l \leq d + d_1 + \dots + d_{k+1},$$

for any  $j$ , and part 2 of Assumption 2.5. Under Assumption 2.3 and Assumption 2.4, any sequence of  $Z$ -estimates  $\hat{\phi}$  satisfying

$$\hat{\phi} \xrightarrow{\mathbb{P}} \phi^0$$

is asymptotically normal with

$$\sqrt{n}(\hat{\phi} - \phi^0) \xrightarrow{D} N(0, \Sigma),$$

where  $\Sigma$  is defined in (4.5).

The proof of Theorem 4.4 is captured in the forthcoming Sect. 8.

**Remark 4.5** Change-point detection has received enormous attention due to the emergence of an increasing amount of temporal data. In the present work, we are mainly concerned with the estimation of the model parameters. We have assumed that the number of changes in the sample is known, which is not the case in real application. Without the need to know the number of change-points in advance, Zou et al. (2014b) proposed a nonparametric maximum likelihood approach to detecting multiple change-points. It worth to notice that the determination of the number of change-points  $k$  in a dataset has been crucial to multiple change-points analysis for long times. It is often approached as a model selection problem, since  $k$  drives the model dimension. we can use the binary segmentation (BinSeg) method proposed in Vostrikova (1981), which is a “top down” procedure, in the sense that one tests all the data to determine if there is at least one change-point and iterates the procedure in the intervals immediately to the “left” and “right” of the most recently detected change-point. This procedure is widely used motivated by the low computational complexity and the is conceptually easy to implement compared to the Exhaustive Search as described by Niu et al. (2016) in Section 3.1. Each stage of BinSeg involves search for a single change-point, which means that if a given segment contains multiple change-points in certain unfavourable configurations, BinSeg may fail to perform adequately on it, as it attempts to fit the “wrong” model. Fryzlewicz (2014) shows that relatively restrictive theoretical assumptions are needed for BinSeg to offer near-optimal performance in terms of the accuracy of estimation of the change-point locations, refer to Korkas and Fryzlewicz (2017) and Fryzlewicz (2018). In the last reference a new solution is proposed giving a ‘tail-greedy’, bottom-up transform for one-dimensional data, which results in a nonlinear but conditionally orthonormal, multiscale decomposition of the data with respect to an adaptatively chosen unbalanced Haar wavelet basis, which avoids the disadvantages of the classical divisive BinSeg. When the number of changes is unknown, Lavielle (1999), Lavielle and Ludeña (2000) proposed its estimation by minimizing a penalized contrast function. Very recently, Zou et al. (2020) proposed a data-driven selection criterion that is applicable to most kinds of popular change-point detection methods, including in particular the binary segmentation and the optimal partitioning algorithms. The main idea is to select the number of change-points that minimizes the squared prediction error, which measures the fit of a specified model for a new sample. The authors investigated a unified parametric framework which includes classical univariate or multivariate location and scale problems, ordinary least-squares, generalized linear models, and many others as special cases, provided that the corresponding objective (likelihood or loss) function can be recast into their asymptotically equivalent least-squares problems. In Zou et al. (2014c), the number of change-points is determined by the Bayesian information criterion and the locations of the change-points can be estimated via the dynamic programming algorithm and the use of the intrinsic order structure of the likelihood function. Under some general conditions, Zou et al. (2014c) showed that the new method provides consistent estimation with an optimal rate. We refer to the last reference for more discussions. For more details, we refer to Truong et al. (2020), where the authors presented a selective survey of algorithms for the offline detection of multiple change-points.

## 5 Maximum likelihood estimators

In this section we will consider the maximum likelihood estimators in models with multiple change points in censored data framework. To unburden our notation a bit, we assume that the censoring variables  $C$  are independent and identically distributed with distribution function  $G(\cdot)$  and density function  $g(\cdot)$ , with respect to the Lebesgue measure  $\lambda$ . Let the lifetime  $X$  and the censoring time  $C$  be positive continuous random variables assumed to be independent. Recall that, the distribution function of the lifetime  $X$  is  $F(\alpha, \theta, \cdot)$  with density function  $f(\alpha, \theta, \cdot)$ , with respect to the Lebesgue measure  $\lambda$ , where  $\alpha$  and

$\theta$  are the unknown parameters to be estimated. In the random censorship from the right model, one observes the pairs  $(Y, \delta)$ , where  $Y = \min(X, C)$  and  $\delta = \mathbb{I}\{X \leq C\}$ . Let  $(Y_i, \delta_i)$ ,  $1 \leq i \leq n$ , denote a random sample of  $(Y, \delta)$  that one observes, and  $Y_{(1)} < \dots < Y_{(m)}$  denote the  $m$  distinct ordered values of  $Y$ 's. When there are ties among the  $Y$ 's, we have  $m < n$ . The likelihood function for this sample is given by

$$\mathcal{L}(\alpha, \theta) = \prod_{i=1}^n f_{Y, \Delta}(\alpha, \theta, \delta_i, y_i),$$

where  $f_{Y, \Delta}(\cdot)$  is the density function of the couple  $(Y, \Delta)$  with respect to the product measure  $\lambda \otimes \mu$  with  $\lambda$  is the measure of Lebesgue and  $\mu$  is the counting measure on the set  $\{0, 1\}$ . The likelihood function can be rewritten as follows

$$\mathcal{L}(\alpha, \theta) = \prod_{i=1}^n f_{Y, \Delta}(\alpha, \theta, \delta_i, y_i) = \prod_{i=1}^n (f(\alpha, \theta, \delta_i, y_i) G(y_i))^{\delta_i} (g(y_i) (1 - F(\alpha, \theta, \delta_i, y_i)))^{1 - \delta_i}. \quad (5.1)$$

By the hypothesis that the distribution of the censored data is independent of the unknown parameters  $\alpha$  and  $\theta$  so the maximization of  $(\alpha, \theta) \mapsto \mathcal{L}(\alpha, \theta)$  is equivalent to the maximization of the pseudo-likelihood given by

$$L(\alpha, \theta) = \prod_{i=1}^n (f(\alpha, \theta, y_i))^{\delta_i} (1 - F(\alpha, \theta, y_i))^{1 - \delta_i}. \quad (5.2)$$

Now, we consider model with known  $k$  change in the distribution, i.e.,

$$X_i \sim F(\alpha, \theta, x), \quad n_{j-1} + 1 \leq i \leq n_j, j = 1, \dots, k + 1; i = 1, \dots, n.$$

In this case, the likelihood function given in (5.2), can be written as follows

$$L(\alpha, \theta_1, \dots, \theta_{k+1}, n_1, \dots, n_k) = \prod_{j=1}^{k+1} \prod_{i=n_{j-1}+1}^{n_j} (f(\alpha, \theta_j, y_i))^{\delta_i} (1 - F(\alpha, \theta_j, y_i))^{1 - \delta_i},$$

which implies that the log-likelihood function is given by

$$\begin{aligned} \ell &\equiv \ell(\alpha, \theta_1, \dots, \theta_{k+1}, n_1, \dots, n_k) \\ &= \frac{1}{n} \sum_{j=1}^{k+1} \sum_{i=n_{j-1}+1}^{n_j} \{\delta_i \log f(\alpha, \theta_j, y_i) + (1 - \delta_i) \log(1 - F(\alpha, \theta_j, y_i))\}, \end{aligned} \quad (5.3)$$

where  $F(\alpha, \theta_j, y) > 0$  for all  $j = 1, \dots, k + 1$ . The maximization is taken with respect to the vector  $(\alpha, \theta_1, \dots, \theta_{k+1}, n_1, \dots, n_k)$ , so the multiplication by the factor  $1/n$  does not affect the optimization problem, which is needed for asymptotic results.

Although only two examples will be given here, they stand as archetypes for a variety of parametric families that can be investigated in a similar way. Let us specify the log-likelihood function for the exponential and Gaussian random variables.

## Exponential distribution

We consider the following model

$$\begin{aligned} X_i &\sim \text{Exp}(\theta_j), \quad n_{j-1} + 1 \leq i \leq n_j, j = 1, \dots, k + 1; i = 1, \dots, n. \\ C_i &\sim \text{Exp}(\beta_j), \quad n_{j-1} + 1 \leq i \leq n_j, j = 1, \dots, k + 1; i = 1, \dots, n, \end{aligned} \quad (5.4)$$

where  $\beta = (\beta_1, \dots, \beta_{k+1})$  is assumed to be known. The log-likelihood function is given by

$$\begin{aligned} \ell(\theta_1, \dots, \theta_{k+1}, n_1, \dots, n_k) &= \frac{1}{n} \sum_{j=1}^{k+1} \sum_{i=n_{j-1}+1}^{n_j} \{\delta_i \log(\theta_j e^{-\theta_j y_i}) + (1 - \delta_i) \log(e^{-\theta_j y_i})\} \\ &= \frac{1}{n} \sum_{j=1}^{k+1} \sum_{i=n_{j-1}+1}^{n_j} \{\delta_i \log(\theta_j) - \delta_i \theta_j y_i - (1 - \delta_i) \theta_j y_i\} \\ &= \frac{1}{n} \sum_{j=1}^{k+1} \sum_{i=n_{j-1}+1}^{n_j} \{\delta_i \log(\theta_j) - \theta_j y_i\}, \end{aligned} \quad (5.5)$$

where  $y_i$  are the observed values.

## Normal distribution

We now consider the uncensored case, where the variables are normal with change only in mean from segment to segment and fixed variance, this means that the change occurs only in  $\theta_j$  and  $\alpha \equiv 1$ , i.e.,

$$X_i \sim \mathcal{N}(\theta_j, 1), \quad n_{j-1} + 1 \leq i \leq n_j, j = 1, \dots, k + 1; i = 1, \dots, n. \quad (5.6)$$

The log-likelihood function in this case is given by

$$\ell(\theta_1, \dots, \theta_{k+1}, n_1, \dots, n_k) = -\frac{1}{n} \sum_{j=1}^{k+1} \sum_{i=n_{j-1}+1}^{n_j} \frac{(x_i - \theta_j)^2}{2}.$$

## 6 Numerical results

This section is concerned with the evaluation of the finite sample performance of the proposed estimation procedure using the the maximum likelihood in (5.2) with samples of different sizes and different censoring rate. We provide numerical illustrations regarding the bias, the variance and the root mean-squared error RMSE. The computing program codes were implemented in R. In our simulation, we choose one sample of 1000 observations with 10 change-points, i.e.,  $k = 10$  with true location;

$$\lambda^0 = (50, 150, 240, 390, 470, 580, 630, 680, 780, 930)$$

and the true within-parameter

$$\theta^0 = (1, 5, 11, 1, 6, 12, 5, 2, 10, 4, 12).$$

We will consider different intensities of censoring in the sample. The censoring random variables  $C_1, \dots, C_n$  are generated from distribution depending on some parameter  $\beta = (\beta_1, \dots, \beta_{11})$  calibrated to attain the desired censoring rate (5%, 10% or 30%). The three scenarios of the censoring rate (proportion) (cr) are given as follows.

(i) cr = 5%, with censoring random variables

$$C_i \sim \text{Exp}(\beta_j), \quad n_{j-1} + 1 \leq i \leq n_j, \quad j = 1, \dots, k + 1; \quad i = 1, \dots, n,$$

where  $\beta = (0.05, 0.3, 0.6, 0.05, 0.3, 0.6, 0.3, 0.1, 0.5, 0.2, 0.6)$ ;

(ii) cr = 10%, with censoring random variables

$$C_i \sim \text{Exp}(\beta_j), \quad n_{j-1} + 1 \leq i \leq n_j, \quad j = 1, \dots, k + 1; \quad i = 1, \dots, n,$$

where  $\beta = (0.1, 0.6, 1.2, 0.1, 0.7, 1.3, 0.6, 0.2, 1, 0.4, 1.3)$ ;

(iii) cr = 30%, with censoring random variables

$$C_i \sim \text{Exp}(\beta_j), \quad n_{j-1} + 1 \leq i \leq n_j, \quad j = 1, \dots, k + 1; \quad i = 1, \dots, n,$$

where  $\beta = (0.4, 2, 5, 0.4, 3, 5, 2, 1, 4.3, 1.7, 5)$ .

The simulation results are reported in the following Tables 1-5.

cr=5%					
Parameter	True value	Mean	BIAS	SD	RMSE
$n_1$	50	58,776	8,776	10,385	13,597
$n_2$	150	149,804	-0,196	4,129	4,134
$n_3$	240	247,018	7,018	5,495	8,913
$n_4$	330	330,128	0,128	3,794	3,796
$n_5$	410	411,926	1,926	7,234	7,486
$n_6$	520	533,803	13,803	6,391	15,211
$n_7$	610	609,971	-0,029	4,402	4,402
$n_8$	710	712,272	2,272	4,332	4,892
$n_9$	820	820,384	0,384	2,86	2,886
$n_{10}$	930	944,531	14,531	4,047	15,084
$\theta_1$	5	4,914	-0,086	0,846	0,851
$\theta_2$	3	3,024	0,024	0,353	0,354
$\theta_3$	1	1,062	0,062	0,132	0,146
$\theta_4$	6	6,244	0,244	0,78	0,817
$\theta_5$	2	2,085	0,085	0,287	0,3
$\theta_6$	7	6,27	-0,73	0,798	1,081
$\theta_7$	3	3,007	0,007	0,396	0,396
$\theta_8$	1	1,02	0,02	0,119	0,121
$\theta_9$	8	8,061	0,061	0,908	0,91
$\theta_{10}$	2	2,122	0,122	0,221	0,252
$\theta_{11}$	7	7,324	0,324	1,096	1,143

Table 1: Maximum likelihood estimator for censored case sample size 1000, with censoring rate 5%.

cr=10%					
Parameter	True value	Mean	BIAS	SD	RMSE
$n_1$	50	58,449	8,449	9,93	13,038
$n_2$	150	149,689	-0,311	4,266	4,278
$n_3$	240	247,778	7,778	5,48	9,515
$n_4$	330	330,051	0,051	3,77	3,77
$n_5$	410	412,538	2,538	7,926	8,322
$n_6$	520	533,86	13,86	6,1	15,143
$n_7$	610	609,764	-0,236	4,661	4,667
$n_8$	710	712,84	2,84	4,668	5,464
$n_9$	820	820,383	0,383	3,115	3,138
$n_{10}$	930	944,569	14,569	4,048	15,121
$\theta_1$	5	4,91	-0,09	0,85	0,854
$\theta_2$	3	3,024	0,024	0,358	0,359
$\theta_3$	1	1,069	0,069	0,137	0,153
$\theta_4$	6	6,259	0,259	0,797	0,838
$\theta_5$	2	2,093	0,093	0,298	0,312
$\theta_6$	7	6,267	-0,733	0,816	1,096
$\theta_7$	3	3,019	0,019	0,409	0,409
$\theta_8$	1	1,027	0,027	0,124	0,127
$\theta_9$	8	8,067	0,067	0,937	0,94
$\theta_{10}$	2	2,118	0,118	0,224	0,253
$\theta_{11}$	7	7,349	0,349	1,113	1,167

Table 2: Maximum likelihood estimator for censored case sample size 1000, with censoring rate 10%.

cr=30%					
Parameter	True value	Mean	BIAS	SD	RMSE
$n_1$	50	58,604	8,604	9,513	12,827
$n_2$	150	149,229	-0,771	4,947	5,006
$n_3$	240	249,918	9,918	5,239	11,216
$n_4$	330	329,484	-0,516	4,45	4,48
$n_5$	410	415,445	5,445	10,287	11,639
$n_6$	520	533,671	13,671	5,708	14,815
$n_7$	610	609,431	-0,569	5,442	5,472
$n_8$	710	715,239	5,239	5,553	7,635
$n_9$	820	819,935	-0,065	3,832	3,832
$n_{10}$	930	944,803	14,803	3,938	15,318
$\theta_1$	5	4,933	-0,067	0,929	0,932
$\theta_2$	3	3,041	0,041	0,418	0,42
$\theta_3$	1	1,086	0,086	0,149	0,172
$\theta_4$	6	6,354	0,354	0,908	0,975
$\theta_5$	2	2,117	0,117	0,323	0,343
$\theta_6$	7	6,278	-0,722	0,936	1,182
$\theta_7$	3	3,045	0,045	0,462	0,464
$\theta_8$	1	1,047	0,047	0,14	0,148
$\theta_9$	8	8,153	0,153	1,078	1,089
$\theta_{10}$	2	2,1	0,1	0,24	0,26
$\theta_{11}$	7	7,477	0,477	1,268	1,354

Table 3: Maximum likelihood estimator for censored case sample size 1000, with censoring rate 30%.

After we consider the case of complete data, i.e.,  $Y_i = X_i$  and  $\delta_i = 1$  for all  $i = 1, \dots, n$  in the same model given in (5.4), the log-likelihood in (5.5) is written in this form

$$\ell(\theta_1, \dots, \theta_{k+1}, n_1, \dots, n_k) = \frac{1}{n} \sum_{j=1}^{k+1} \sum_{i=n_{j-1}+1}^{n_j} \{\log(\theta_j) - \theta_j y_i\},$$

with the same true location  $\lambda^0$  and the same true within-parameters  $\theta^0$ , we have the following results in Table 4.

The uncensored case with exponential distribution					
Parameter	True value	Mean	BIAS	SD	RMSE
$n_1$	50	58,994	8,994	10,59	13,894
$n_2$	150	149,901	-0,099	3,926	3,927
$n_3$	240	246,438	6,438	5,414	8,412
$n_4$	330	330,205	0,205	3,65	3,656
$n_5$	410	411,314	1,314	6,511	6,642
$n_6$	520	533,732	13,732	6,444	15,169
$n_7$	610	610,012	0,012	4,312	4,312
$n_8$	710	711,716	1,716	3,96	4,316
$n_9$	820	820,473	0,473	2,861	2,9
$n_{10}$	930	944,533	14,533	4,084	15,096
$\theta_1$	5	4,904	-0,096	0,835	0,841
$\theta_2$	3	3,021	0,021	0,341	0,342
$\theta_3$	1	1,057	0,057	0,13	0,143
$\theta_4$	6	6,23	0,23	0,745	0,78
$\theta_5$	2	2,081	0,081	0,281	0,292
$\theta_6$	7	6,27	-0,73	0,78	1,068
$\theta_7$	3	3,001	0,001	0,388	0,388
$\theta_8$	1	1,015	0,015	0,116	0,117
$\theta_9$	8	8,042	0,042	0,88	0,881
$\theta_{10}$	2	2,126	0,126	0,219	0,253
$\theta_{11}$	7	7,287	0,287	1,057	1,095

Table 4: Maximum likelihood estimator for uncensored case sample size 1000, for complete data.

Finally, consider the case of normal distribution for complete data model given in (5.6), with sample size 1000 with 10 change-points, with true location given by

$$\lambda = (70, 160, 250, 340, 440, 540, 630, 730, 820, 920)$$

and the true within-parameter is given

$$\theta = (-5, 3, 0, 4, -1, 3, -3, 10, 4, -2, 0).$$

The results are reported in Table 5.

The uncensored case with normal distribution					
Parameter	True value	Mean	BIAS	SD	RMSE
$n_1$	70	69,999	-0,001	0,024	0,024
$n_2$	160	186,845	26,845	13,54	30,066
$n_3$	250	255,279	5,279	9,215	10,621
$n_4$	340	340,029	0,029	0,376	0,377
$n_5$	440	450,289	10,289	16,02	19,04
$n_6$	540	540,721	0,721	3,609	3,68
$n_7$	630	630	0	0	0
$n_8$	730	730,394	0,394	2,201	2,236
$n_9$	820	820,008	0,008	0,17	0,17
$n_{10}$	920	943,747	23,747	10,356	25,907
$\theta_1$	-5	-5,002	-0,002	0,167	0,167
$\theta_2$	3	2,354	-0,646	0,299	0,711
$\theta_3$	0	0,341	0,341	0,449	0,563
$\theta_4$	4	3,937	-0,063	0,158	0,17
$\theta_5$	-1	-0,588	0,412	0,444	0,606
$\theta_6$	3	2,894	-0,106	0,279	0,298
$\theta_7$	-3	-2,988	0,012	0,134	0,135
$\theta_8$	10	9,903	-0,097	0,198	0,22
$\theta_9$	4	3,968	-0,032	0,142	0,146
$\theta_{10}$	-2	-1,62	0,38	0,173	0,417
$\theta_{11}$	0	-0,011	-0,011	0,177	0,177

Table 5: Maximum likelihood estimator for complete data with sample size 1000, for complete data.

From tables and figures, the best results are obtained when the data is complete, and the results in the censoring case are satisfactory when the censoring rate is moderate 5% and 10% and the performance are deteriorated when the censoring rate increase. The following figures are computed for the three rates of censoring and for complete data for model given in (5.4) with 1000 replicate from samples with sizes from 1000 to 10000 i.e., size = (100, 50, 120, 90, 90, 110, 80, 100, 90, 100, 70)\* $k$ ;  $k = 1, \dots, 10$ , and true location given by  $\lambda = (100, 150, 270, 360, 450, 560, 640, 740, 830, 930)$  By inspecting Figures 1-6, one can see that as in any other inferential context, the greater the sample size, the better. In the literature, it is commonly used two or three changes in the sample for the finite sample experiments. In the present simulations, we have optimized the likelihood criterion with respect to 21 parameters ( $n_1, \dots, n_{10}, \theta_1, \dots, \theta_{11}$ ) simultaneously, including 10 changes in the sample, which has a computational cost. This can be circumvented by using the penalized likelihood criterion. In order to extract methodological recommendations for the use of the procedures proposed in this work, it will be interesting to conduct extensive Monte Carlo experiments to compare our procedures with other scenarios presented in the literature, but this would go well beyond the scope of the present paper.

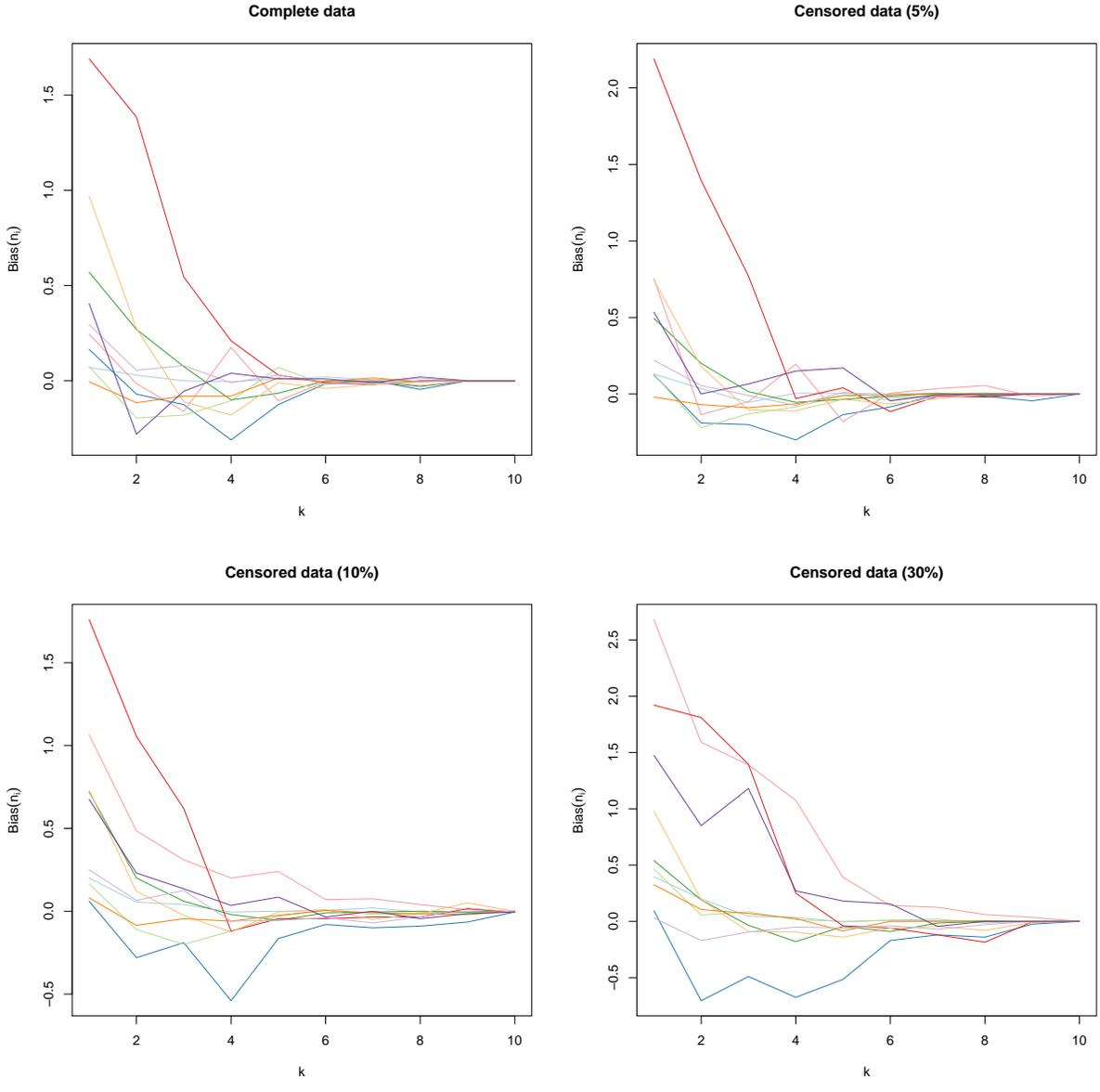


Figure 1: Bias of  $\hat{n}_j, j = 1, \dots, 10$ .

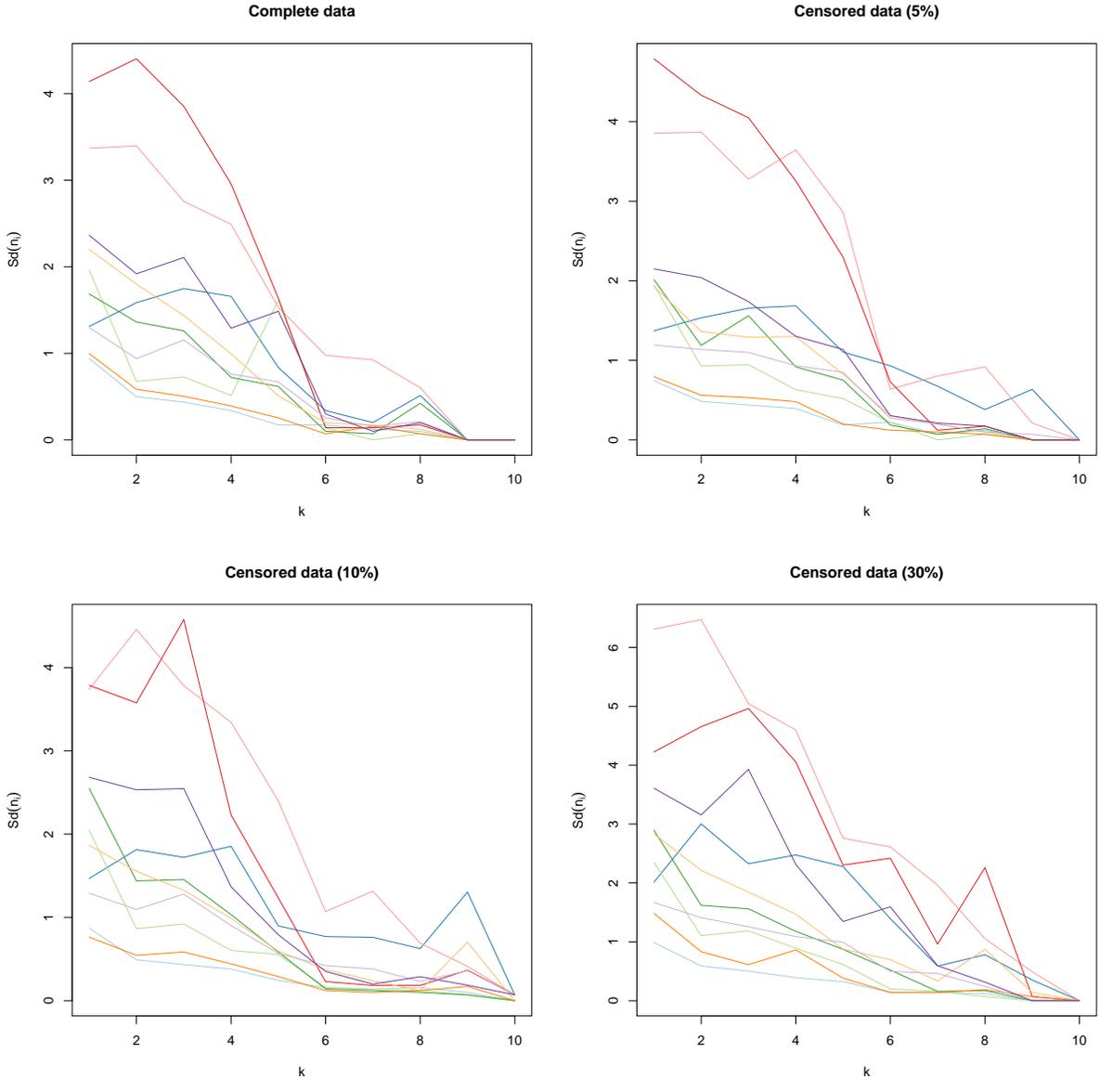


Figure 2: Standard deviation of  $\hat{n}_j, j = 1, \dots, 10$ .

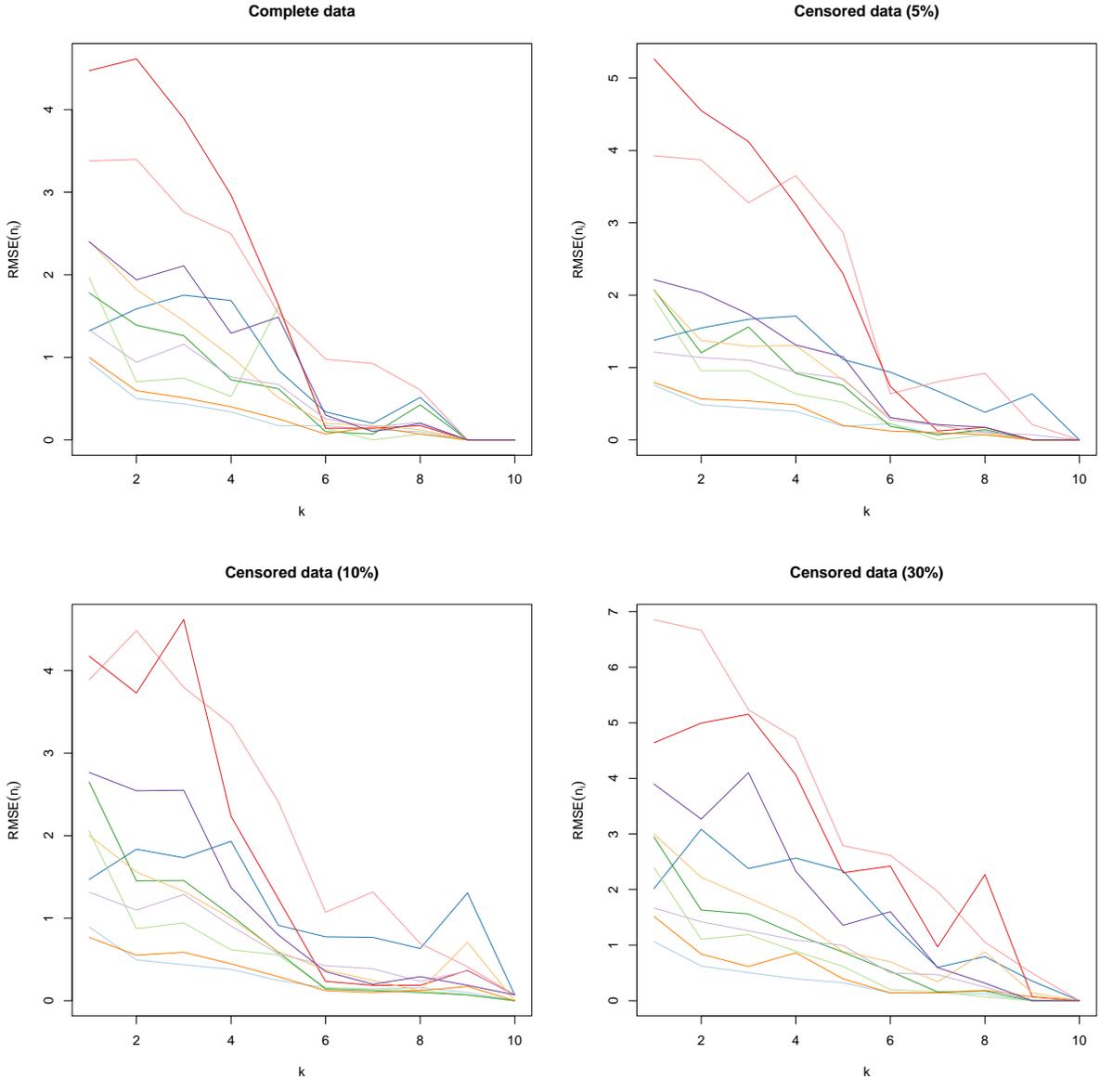


Figure 3: Root of MSE of  $\hat{\eta}_j, j = 1, \dots, 10$ .

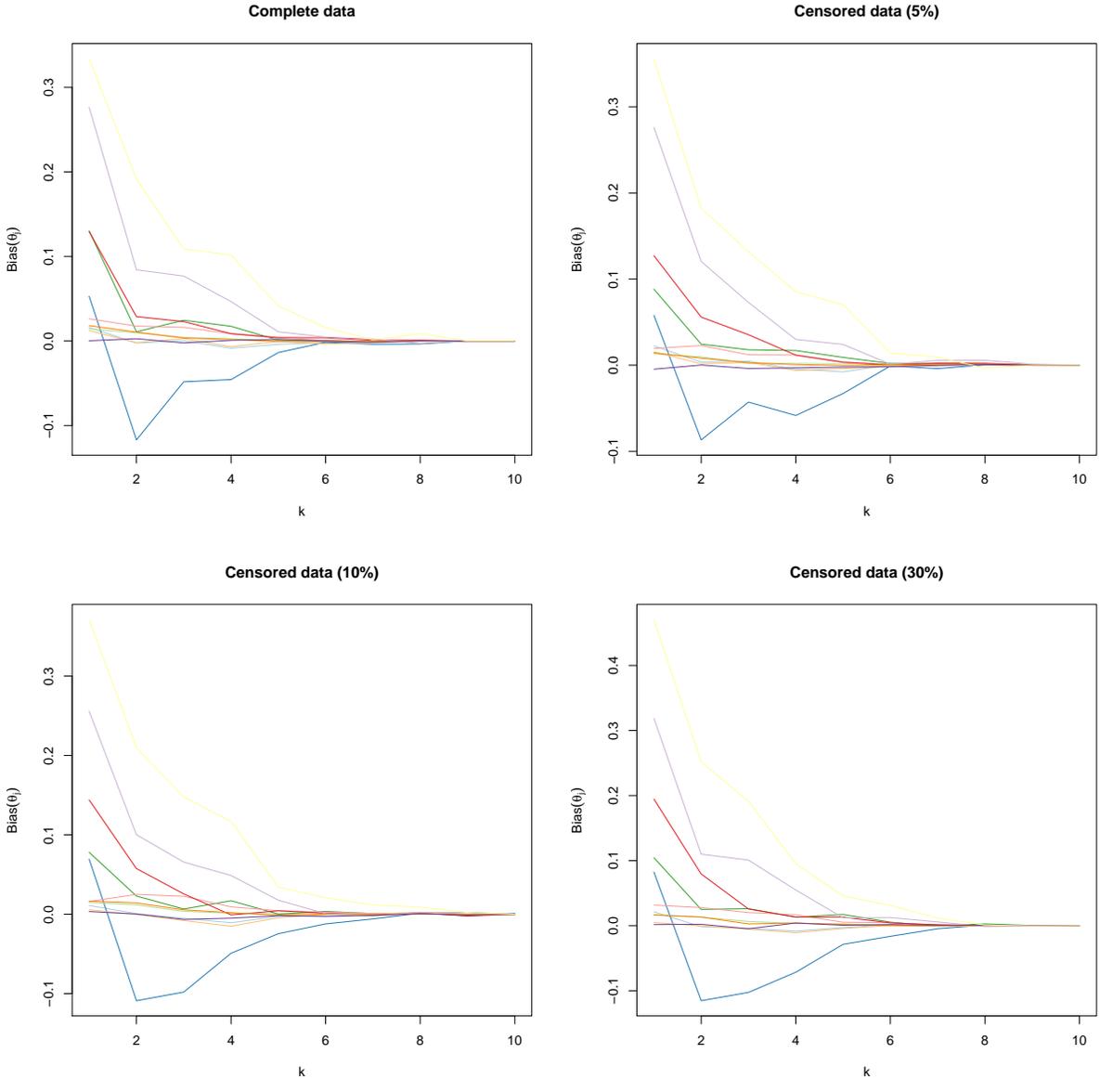


Figure 4: Bias of  $\hat{\theta}_j, j = 1, \dots, 11$ .

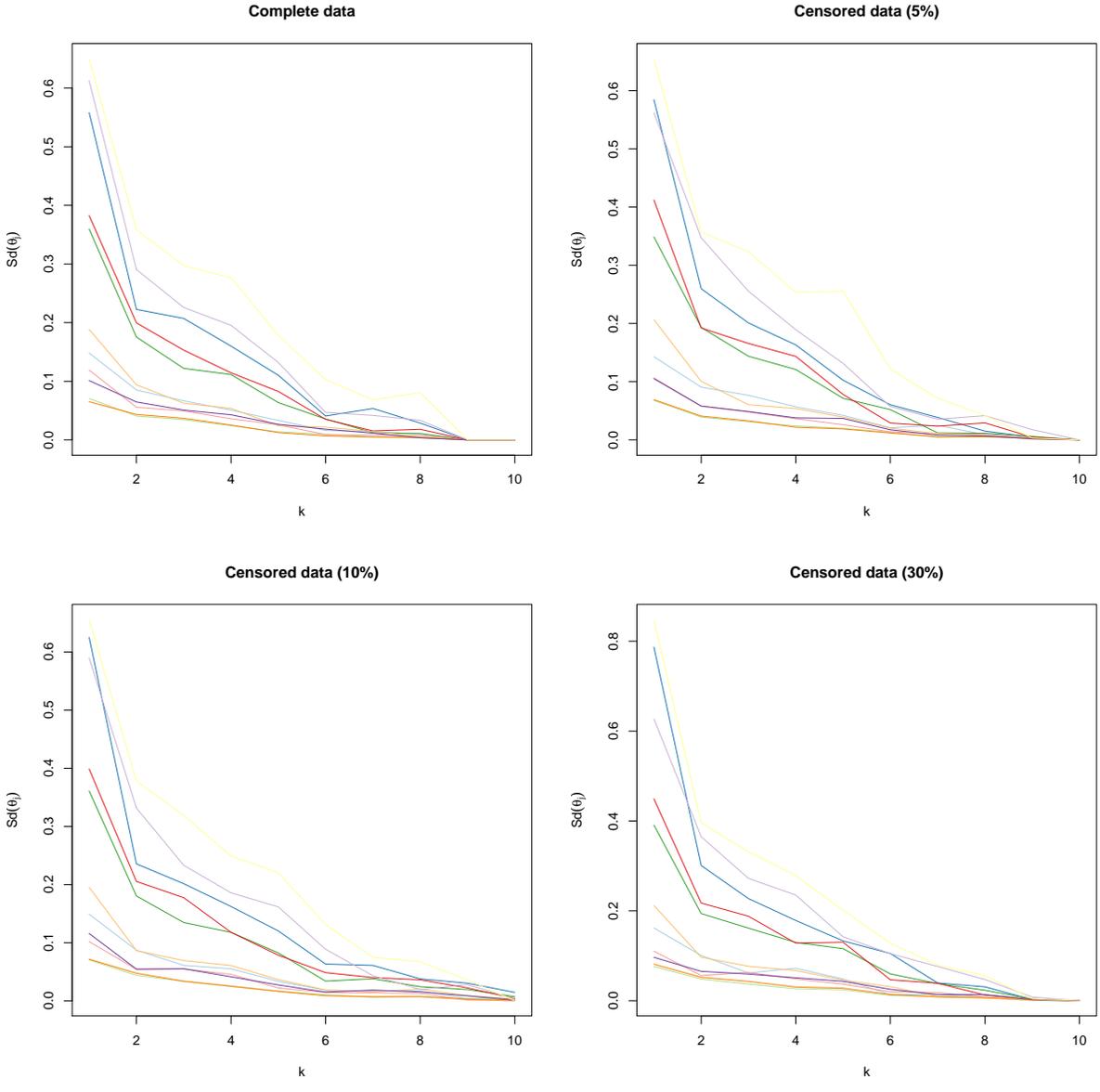


Figure 5: Standard deviation of  $\hat{\theta}_j$ ,  $j = 1, \dots, 11$ .

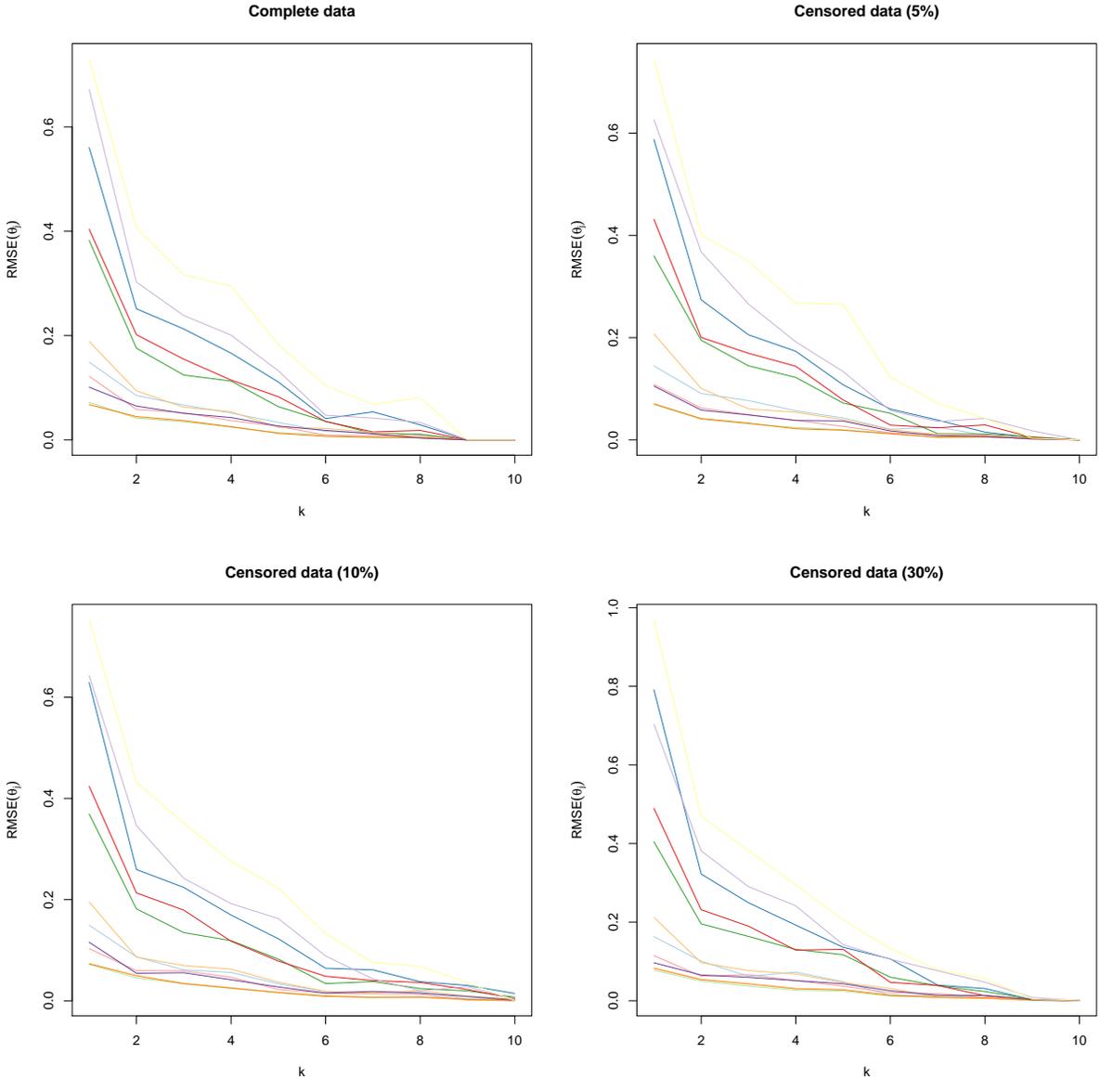


Figure 6: Root of MSE of  $\hat{\theta}_j, j = 1, \dots, 11$ .

## 7 Concluding remarks

Some important problems in the analysis of multiple change-point models were not considered here. One is that the asymptotic distribution of the  $M$ -estimator of the vector of change points was not considered, see for example [Hinkley \(1970\)](#) for a treatment of this problem in a single change-point model and [Döring \(2011\)](#) for multiple change points. Thus, this is essentially a separate research topic. However, the asymptotic properties obtained in this paper are necessary for the establishment of the asymptotic distribution of the  $M$ -estimator of the vector of change points in this model. This will be a subject of investigation for future work.

Another important problem is to extend the results of this paper to the case in which the number of change points is not known and must be determined from the data. Another direction of research is that the methods and arguments in this paper can be extended to other types of incomplete data (e.g. truncation, double censoring, interval censoring etc.) or data subject to sampling bias, where the Kaplan-Meier product-limit estimate  $\widehat{F}_{n_j}(\cdot)$  will be replaced by an appropriate estimate, usually the non-parametric maximum likelihood estimate of the true lifetime distribution function. Such an extension is straightforward whenever, for the suitable choice of  $\widehat{F}_{n_j}(\cdot)$ , the CLT of  $\int_{\mathbb{R}} \varphi(x) d\widehat{F}_{n_j}(x)$  have been established for an arbitrary function  $\varphi(\cdot)$ . It would be interesting to cleanly extend the results to this, but this would require further theory which are out of the scope of the present article. Change point estimation is a classical problem in mathematical statistics which, with its broad range of applications in learning problems, has started to gain attention in the machine learning community. An important question is how to apply our findings in such problems. **Finally, the optimization problems become computationally complex when the number of parameter is large, it will be interesting to consider the penalized version of the likelihood function to alleviate such difficulties.**

## 8 Mathematical developments

This section is devoted to the proofs of our results. The previously defined notation continues to be used below.

The proof of [Theorem 3.1](#) will based on the [Lemma 8.1](#) and [Lemma 8.2](#). The following lemma gives a bound for the term  $\mathbf{W}_1$  given in [equation \(2.5\)](#).

**Lemma 8.1** *Under the Assumption 3.1 and the first part of Assumption 3.2, there exist two positive constants  $C_1 > 0$  and  $C_2 > 0$  such that, for any  $\lambda$  and  $\phi$ , we have*

$$\mathbf{W}_1 \leq -\max\{C_1 \|\lambda - \lambda^0\|_{\infty}, C_2 \varrho(\phi, \phi^0)\},$$

where

$$\|\lambda - \lambda^0\|_{\infty} = \max_j |\lambda_j - \lambda_j^0| \quad \text{and} \quad \varrho(\phi, \phi^0) = \max_j |b(\alpha, \theta_j, \alpha^0, \theta_j^0)|.$$

### Proof of [Lemma 8.1](#)

The proof of this lemma follows the similar arguments used in the proof of [Lemma 3.1](#) in [He and Severini \(2010\)](#). Recall that

$$\begin{aligned} b(\alpha, \theta_j, \alpha^0, \theta_j^0) &= \mathbb{E}(g_j(\alpha, \theta_j, X_i)) - \mathbb{E}(g_i(\alpha^0, \theta_j^0, X_i)) \\ &= \int_{\mathbb{R}} [g_j(\alpha, \theta_j, x) - g_i(\alpha^0, \theta_j^0, x)] dF_{n_i^0}(x). \end{aligned}$$

Let us define, for  $i = 1, 2, \dots, k$ ,

$$h_i(\beta, \phi^0) = \sup_{1 \leq j \leq k} \sup_{\theta_j \in \Theta_j} \sup_{\alpha \in \Upsilon} [\beta b(\alpha, \theta_j, \alpha^0, \theta_{i+1}^0) + (1 - \beta) b(\alpha, \theta_j, \alpha^0, \theta_i^0)],$$

where  $\beta \in [0, 1]$ . We have

$$h_i(0, \phi^0) = h_i(1, \phi^0) = 0 \quad \text{for } i = 1, 2, \dots, k.$$

One can check that  $h_i(\beta, \phi^0)$  is a convex function with respect to  $\beta$  for any  $i = 1, 2, \dots, k$ . Let

$$H_i(\phi^0) = 2h_i(1/2, \phi^0).$$

It follows from the [Assumption 2.1](#) that  $H_i(\phi^0) < 0$ . If we let

$$\overline{H}(\phi^0) = \max_{1 \leq i \leq k} H_i(\phi^0),$$

then we have  $\overline{H}(\phi^0) < 0$ . Let

$$\Delta_{\lambda}^0 = \min_{1 \leq j \leq k-1} |\lambda_{j+1}^0 - \lambda_j^0|.$$

Consider the change-point configuration  $\lambda$  in such a way that

$$\|\lambda - \lambda^0\|_{\infty} \leq \Delta_{\lambda}^0/4.$$

For any  $j = 1, 2, \dots, k$ , there are two cases: a candidate change-point fraction  $\lambda_j$  may be on the left or on the right of the true change-point fraction  $\lambda_j^0$ . For any  $j$  with  $\lambda_j$  on the right of  $\lambda_j^0$ , we have that  $\lambda_{j-1} \leq \lambda_j^0 \leq \lambda_j$ . Then

$$\mathbf{W}_1 \leq \frac{n_{j,j+1}}{n} b(\alpha, \theta_j, \alpha^0, \theta_{j+1}^0) + \frac{n_{jj}}{n} b(\alpha, \theta_j, \alpha^0, \theta_j^0).$$

If we define

$$\beta_{j,j+1} = \frac{n_{j,j+1}}{n_{j,j+1} + n_{jj}},$$

the case  $\|\lambda - \lambda^0\|_{\infty} \leq \Delta_{\lambda}^0/4$  gives that  $\beta_{j,j+1} \leq 1/2$  and

$$\mathbf{W}_1 \leq (\lambda_j - \lambda_j^0) \overline{H}(\phi^0).$$

For any  $j$  with  $\lambda_j$  on the left of  $\lambda_j^0$ , we have that  $\lambda_j \leq \lambda_j^0 \leq \lambda_{j+1}$ . Similarly, we define

$$\beta_{j,j-1} = \frac{n_{j,j-1}}{n_{j,j-1} + n_{jj}},$$

we get  $\beta_{j,j-1} \leq 1/2$  and

$$\mathbf{W}_1 \leq (\lambda_j^0 - \lambda_j) \overline{H}(\phi^0).$$

Therefore, if  $\|\lambda - \lambda^0\|_{\infty} \leq \Delta_{\lambda}^0/4$ , we readily obtain that

$$\mathbf{W}_1 \leq \|\lambda - \lambda^0\|_{\infty} \overline{H}(\phi^0).$$

On the other hand, we have

$$\mathbf{W}_1 \leq \min_{1 \leq j \leq k+1} b(\alpha, \theta_j, \alpha^0, \theta_j^0) \frac{n_{jj}}{n} = - \max_{1 \leq j \leq k+1} |b(\alpha, \theta_j, \alpha^0, \theta_j^0)| \frac{n_{jj}}{n}.$$

For any  $j$ , we have  $\frac{n_{jj}}{n} \geq \Delta_{\lambda}^0/2$ , so we infer that

$$\mathbf{W}_1 \leq -\frac{1}{2} \Delta_{\lambda}^0 \varrho(\phi, \phi^0).$$

Now, consider the other case of change-point fraction configuration  $\lambda$ , where

$$\|\lambda - \lambda^0\|_{\infty} > \Delta_{\lambda}^0/4.$$

It is obvious that there exists a pair of integers  $(i, j)$  such that  $n_{ij} \geq n\Delta_{\lambda}^0/4$ ,  $n_{i,j+1} \geq n\Delta_{\lambda}^0/4$  and  $n_{ij} \geq n_{i,j+1}$ . Let

$$\beta_{i,j+1} = \frac{n_{i,j+1}}{n_{i,j+1} + n_{ij}}.$$

For any  $\phi$ , we have

$$\begin{aligned} \mathbf{W}_1 &\leq \frac{n_{i,j+1} + n_{ij}}{n} [\beta_{i,j+1} b(\alpha, \theta_i, \alpha^0, \theta_{j+1}^0) + (1 - \beta_{i,j+1}) b(\alpha, \theta_i, \alpha^0, \theta_j^0)] \\ &\leq \frac{1}{2} \left( \frac{\Delta_{\lambda}^0}{2} \right)^2 \overline{H}(\phi^0). \end{aligned}$$

Combining the results from the two cases of  $\|\lambda - \lambda^0\|_{\infty} \leq \Delta_{\lambda}^0/4$  and  $\|\lambda - \lambda^0\|_{\infty} > \Delta_{\lambda}^0/4$ , it follows that

$$\mathbf{W}_1 \leq \frac{1}{2} \left( \frac{\Delta_{\lambda}^0}{2} \right)^2 \overline{H}(\phi^0) \|\lambda - \lambda^0\|_{\infty},$$

and

$$\mathbf{W}_1 \leq -\frac{\Delta_\lambda^0}{2} \min \left[ \varrho(\phi, \phi^0), -\frac{\Delta_\lambda^0}{4} \overline{H}(\phi^0) \right]. \quad (8.1)$$

Note that (8.1) can be simplified. Let us define

$$\rho(\phi, \phi^0) = \max_{1 \leq j \leq k+1} \sup_{\theta_j \in \Theta_j} \sup_{\alpha \in \Upsilon} |b(\alpha, \theta_j, \alpha^0, \theta_j^0)|.$$

It follows from the inequality (8.1) that we have

$$\mathbf{W}_1 \leq -\frac{\Delta_\lambda^0}{2} \rho(\phi, \phi^0) \min \left[ \frac{\varrho(\phi, \phi^0)}{\rho(\phi, \phi^0)}, -\frac{\Delta_\lambda^0}{4} \overline{H}(\phi^0) / \rho(\phi, \phi^0) \right].$$

If  $-\frac{\Delta_\lambda^0}{4} \overline{H}(\phi^0) / \rho(\phi, \phi^0) \leq 1$ , then we infer that

$$\mathbf{W}_1 \leq (\Delta_\lambda^0 / 2)^2 (\varrho(\phi, \phi^0) / \rho(\phi, \phi^0)) (\overline{H}(\phi^0) / 2).$$

If  $-\frac{\Delta_\lambda^0}{4} \overline{H}(\phi^0) / \rho(\phi, \phi^0) > 1$ , we readily obtain

$$\mathbf{W}_1 \leq -(\Delta_\lambda^0 / 2) \varrho(\phi, \phi^0).$$

Letting

$$C_2 = \min\{(\Delta_\lambda^0 / 2)^2 |\overline{H}(\phi^0)| / (2\rho(\phi, \phi^0)), \Delta_\lambda^0 / 2\},$$

inequality (8.1) implies that

$$\mathbf{W}_1 \leq -C_2 \varrho(\phi, \phi^0).$$

Setting

$$C_1 = (\Delta_\lambda^0 / 2)^2 |\overline{H}(\phi^0)| / 2,$$

we finally have the desired result.  $\square$

The following lemma describes between-segment properties and within-segment properties of the model.

**Lemma 8.2** *Under the Assumption 2.6, part 1 and 2 of the Assumption 2.3 respectively, it follows that*

(I) *For any  $j = 1, 2, \dots, k+1$ , any  $0 \leq m_1 < m_2 \leq n$  and any positive number  $\epsilon > 0$ , there exists a constant  $A_j$ , independent of  $\epsilon$ , and a constant  $r > 2$ , such that*

$$\begin{aligned} & \mathbb{P} \left( \max_{m_1 \leq s < t \leq m_2, \theta_j \in \Theta_j, \alpha \in \Upsilon} \left| \sum_{i=s+1}^t \left( \sum_{z=1}^{k+1} \frac{g_j(\alpha, \theta_j, Y_i) \delta_i}{S_{G_{n_z^0}}(Y_i^-)} \mathbb{I}_{\{n_{z-1}+1 \leq i \leq n_z\}} - \mathbb{E}(g_j(\alpha, \theta_j, X_i)) \right) \right| > \epsilon \right) \\ & \leq A_j \frac{(m_2 - m_1)^r}{\epsilon^2}. \end{aligned} \quad (8.2)$$

(II) *For any  $j = 1, 2, \dots, k+1$  and any positive number  $\epsilon > 0$ , there exist a constant  $B_j$ , independent of  $\epsilon$ , and a constant  $r > 2$ , such that*

$$\begin{aligned} & \mathbb{P} \left( \max_{n_{j-1}^0 \leq s < t \leq n_j^0, \theta_j \in \Theta_j, \alpha \in \Upsilon} \left[ \sum_{i=s+1}^t \left( \sum_{z=1}^{k+1} \frac{g_j(\alpha, \theta_j, Y_i) \delta_i}{S_{G_{n_z^0}}(Y_i^-)} \mathbb{I}_{\{n_{z-1}+1 \leq i \leq n_z\}} \right. \right. \right. \\ & \left. \left. \left. - \frac{g_j(\alpha^0, \theta_j^0, Y_i) \delta_i}{S_{G_{n_j^0}}(Y_i^-)} - b(\alpha, \theta_j, \alpha^0, \theta_j^0) \right) \right] > \epsilon \right) \leq B_j \frac{(n_j^0 - n_{j-1}^0)^r}{\epsilon^2}. \end{aligned} \quad (8.3)$$

## Proof of Lemma 8.2

By the fact that all variables at hand are independent and keeping the part 1 of the Assumption 2.1 in mind, equation (8.2) can be achieved by induction with respect to  $m_2$ . The induction method is similar to the one used in [Móricz et al. \(1982\)](#), so its proof is omitted here. Using part 2 of the Assumption 2.1, equation (8.3) can be proved similarly by the same induction method. For more details, we can refer to [He and Severini \(2010\)](#).  $\square$

### Proof of Theorem 3.1

Let us introduce the following notation

$$\begin{aligned}\Lambda &= \{(\lambda_1, \lambda_2, \dots, \lambda_k) : \lambda_j = \frac{n_j}{n}, j = 1, \dots, k; 0 < n_1 < \dots < n_k < n\}, \\ \Lambda_\eta &= \{\lambda \in \Lambda : \|\lambda - \lambda^0\|_\infty > \eta\}, \\ \Phi &= \Theta_1 \times \Theta_2 \times \dots \times \Theta_{k+1} \times \Upsilon, \\ \Phi_\eta &= \{\phi \in \Phi : \varrho(\phi, \phi^0) > \eta\}.\end{aligned}$$

Then, for any  $\eta > 0$ , it follows from an application of Lemma 8.1 that

$$-\max_{\lambda \in \Lambda_\eta, \phi \in \Phi} \mathbf{W}_1 \geq C_1 \eta \quad \text{and} \quad -\max_{\lambda \in \Lambda, \phi \in \Phi_\eta} \mathbf{W}_1 \geq C_2 \eta.$$

Therefore, we readily obtain that

$$\begin{aligned}\mathbb{P}(\|\lambda - \lambda^0\|_\infty > \eta) &\leq \mathbb{P}\left(\max_{\lambda \in \Lambda_\eta, \phi \in \Phi} \mathbf{W} > 0\right) \leq \mathbb{P}\left(\max_{\lambda \in \Lambda_\eta, \phi \in \Phi} \mathbf{W}_2 > -\max_{\lambda \in \Lambda_\eta, \phi \in \Phi} \mathbf{W}_1\right) \leq \mathbb{P}\left(\max_{\lambda \in \Lambda_\eta, \phi \in \Phi} |\mathbf{W}_2| > C_1 \eta\right) \\ &\leq \mathbb{P}\left(\max_{\lambda \in \Lambda_\eta, \phi \in \Phi} \sum_{j=1}^{k+1} \frac{1}{n} \left| \sum_{i=n_{j-1}+1}^{n_j} \left\{ \frac{g_j(\alpha, \theta_j, Y_i) \delta_i}{S_{G_{n_j^0}}(Y_i^-)} - \mathbb{E}(g_j(\alpha, \theta_j, X_i)) \right\} \right| > \frac{C_1 \eta}{2}\right) \\ &+ \mathbb{P}\left(\sum_{j=1}^{k+1} \frac{1}{n} \left| \sum_{i=n_{j-1}^0+1}^{n_j^0} \left\{ \frac{g_j(\alpha^0, \theta_j^0, Y_i) \delta_i}{S_{G_{n_j^0}}(Y_i^-)} - \mathbb{E}(g_j(\alpha^0, \theta_j^0, X_i)) \right\} \right| > \frac{C_1 \eta}{2}\right) \\ &\leq \sum_{j=1}^{k+1} \mathbb{P}\left(\max_{0 \leq n_{j-1} < n_j \leq n, \theta_j \in \Theta_j, \alpha \in \Upsilon} \frac{1}{n} \left| \sum_{i=n_{j-1}+1}^{n_j} \left\{ \frac{g_j(\alpha, \theta_j, Y_i) \delta_i}{S_{G_{n_j^0}}(Y_i^-)} - \mathbb{E}(g_j(\alpha, \theta_j, X_i)) \right\} \right| > \frac{C_1 \eta}{2(k+1)}\right) \\ &+ \sum_{j=1}^{k+1} \mathbb{P}\left(\frac{1}{n} \left| \sum_{i=n_{j-1}^0+1}^{n_j^0} \left\{ \frac{g_j(\alpha^0, \theta_j^0, Y_i) \delta_i}{S_{G_{n_j^0}}(Y_i^-)} - \mathbb{E}(g_j(\alpha^0, \theta_j^0, X_i)) \right\} \right| > \frac{C_1 \eta}{2(k+1)}\right).\end{aligned}$$

It follows from Lemma 8.2 that, as  $n \rightarrow +\infty$ ,

$$\mathbb{P}(\|\lambda - \lambda^0\|_\infty > \eta) \leq 2 \left[ \frac{2(k+1)}{C_1 \eta} \right]^2 \left( \sum_{j=1}^{k+1} A_j \right) n^{r-2} \rightarrow 0.$$

For the estimator  $\widehat{\phi}$ , we obtain in a similar way that

$$\begin{aligned}\mathbb{P}(\varrho(\widehat{\phi}, \phi^0) > \eta) &\leq \mathbb{P}\left(\max_{\lambda \in \Lambda, \phi \in \Phi_\eta} \mathbf{W} > 0\right) \\ &\leq \sum_{j=1}^{k+1} \mathbb{P}\left(\max_{0 \leq n_{j-1} < n_j \leq n, \theta_j \in \Theta_j, \alpha \in \Upsilon} \frac{1}{n} \left| \sum_{i=n_{j-1}+1}^{n_j} \left\{ \frac{g_j(\alpha, \theta_j, Y_i) \delta_i}{S_{G_{n_j^0}}(Y_i^-)} - \mathbb{E}(g_j(\alpha, \theta_j, X_i)) \right\} \right| > \frac{C_2 \eta}{2(k+1)}\right) \\ &+ \sum_{j=1}^{k+1} \mathbb{P}\left(\frac{1}{n} \left| \sum_{i=n_{j-1}^0+1}^{n_j^0} \left\{ \frac{g_j(\alpha^0, \theta_j^0, Y_i) \delta_i}{S_{G_{n_j^0}}(Y_i^-)} - \mathbb{E}(g_j(\alpha^0, \theta_j^0, X_i)) \right\} \right| > \frac{C_2 \eta}{2(k+1)}\right).\end{aligned}$$

Once more, an application of Lemma 8.2 shows, as  $n \rightarrow +\infty$ , that

$$\mathbb{P}(\varrho(\widehat{\phi}, \phi^0) > \eta) \rightarrow 0.$$

Noting the fact that  $b(\alpha, \theta_j, \alpha^0, \theta_j^0) = 0$  if and only if  $\alpha = \alpha^0$  and  $\theta_j = \theta_j^0$ , for  $j = 1, \dots, k+1$ , completes the proof of Theorem 3.1.  $\square$

### Proof of Theorem 3.3

Let us first define, for any  $\eta > 0$ ,

$$\Lambda_{\eta,n} = \{\boldsymbol{\lambda} \in \Lambda : n\|\boldsymbol{\lambda} - \boldsymbol{\lambda}^0\|_\infty \geq \eta\}.$$

Making use of the consistency of the change point fraction  $\widehat{\boldsymbol{\lambda}}$ , we need to consider only the observations in  $\tilde{n}_{j,j-1}$ ,  $\tilde{n}_{j,j}$  and  $\tilde{n}_{j,j+1}$  for all  $j$  in equation (2.6). Therefore, we have

$$\begin{aligned} & \mathbb{P}\left(n\|\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^0\|_\infty \geq \eta\right) \\ & \leq \sum_{j=1}^{k+1} \mathbb{P}\left(\max_{\boldsymbol{\lambda} \in \Lambda_{\eta,n}, \phi \in \Phi} \left\{ \frac{1}{n} \sum_{t \in \tilde{n}_{j,j}} \left[ \frac{g_j(\alpha, \theta_j, Y_t) \delta_t}{S_{G_{n_j^0}}(Y_t^-)} - \mathbb{E}(g_j(\alpha, \theta_j, X_t)) \right] \right. \right. \\ & \quad \left. \left. - \frac{1}{n} \sum_{t \in \tilde{n}_{j,j}} \left[ \frac{g_j(\alpha^0, \theta_j^0, Y_t) \delta_t}{S_{G_{n_j^0}}(Y_t^-)} - \mathbb{E}(g_j(\alpha^0, \theta_j^0, X_t)) \right] + \frac{1}{3(k+1)} \mathbf{W}_1 \right\} > 0\right) \\ & + \sum_{j=2}^{k+1} \mathbb{P}\left(\max_{\boldsymbol{\lambda} \in \Lambda_{\eta,n}, \phi \in \Phi} \left\{ \frac{1}{n} \sum_{t \in \tilde{n}_{j,j-1}} \left[ \frac{g_j(\alpha, \theta_j, Y_t) \delta_t}{S_{G_{n_j^0}}(Y_t^-)} - \mathbb{E}(g_j(\alpha, \theta_j, X_t)) \right] \right. \right. \\ & \quad \left. \left. - \frac{1}{n} \sum_{t \in \tilde{n}_{j,j-1}} \left[ \frac{g_{j-1}(\alpha^0, \theta_{j-1}^0, Y_t) \delta_t}{S_{G_{n_{j-1}^0}}(Y_t^-)} - \mathbb{E}(g_{j-1}(\alpha^0, \theta_{j-1}^0, X_t)) \right] + \frac{1}{3k} \mathbf{W}_1 \right\} > 0\right) \\ & + \sum_{j=1}^k \mathbb{P}\left(\max_{\boldsymbol{\lambda} \in \Lambda_{\eta,n}, \phi \in \Phi} \left\{ \frac{1}{n} \sum_{t \in \tilde{n}_{j,j+1}} \left[ \frac{g_j(\alpha, \theta_j, Y_t) \delta_t}{S_{G_{n_j^0}}(Y_t^-)} - \mathbb{E}(g_j(\alpha, \theta_j, X_t)) \right] \right. \right. \\ & \quad \left. \left. - \frac{1}{n} \sum_{t \in \tilde{n}_{j,j+1}} \left[ \frac{g_{j+1}(\alpha^0, \theta_{j+1}^0, Y_t) \delta_t}{S_{G_{n_{j+1}^0}}(Y_t^-)} - \mathbb{E}(g_{j+1}(\alpha^0, \theta_{j+1}^0, X_t)) \right] + \frac{1}{3k} \mathbf{W}_1 \right\} > 0\right) \\ & \equiv \sum_{j=1}^{k+1} I_{1j} + \sum_{j=2}^{k+1} I_{2j} + \sum_{j=1}^k I_{3j}. \end{aligned}$$

First, consider the probability formulas  $I_{1j}$  in the above equation for any  $j = 1, 2, \dots, k+1$ . The consistency of  $\widehat{\boldsymbol{\lambda}}$  allows us to restrict our attention to the case  $n_{jj} > \frac{1}{2}(n_j^0 - n_{j-1}^0)$ . For this case, we have that

$$\mathbf{W}_1 \leq \frac{n_j^0 - n_{j-1}^0}{2n} b(\alpha, \theta_j, \alpha^0, \theta_j^0).$$

Therefore, we readily obtain that

$$\begin{aligned} I_{1j} & \leq \mathbb{P}\left(\max_{n_{j-1}^0 \leq s < t \leq n_j^0, \theta_j \in \Theta_j, \alpha \in \Upsilon} \left[ \sum_{i=s+1}^t \left( \sum_{z=j-1}^{j+1} \frac{g_j(\alpha, \theta_j, Y_i) \delta_i}{S_{G_{n_z^0}}(Y_i^-)} \mathbb{I}_{\{n_{z-1}+1 \leq i \leq n_z\}} \right. \right. \right. \\ & \quad \left. \left. - \frac{g_j(\alpha^0, \theta_j^0, Y_i) \delta_i}{S_{G_{n_j^0}}(Y_i^-)} - b(\alpha, \theta_j, \alpha^0, \theta_j^0) \right) \right] > \frac{a(n_j^0 - n_{j-1}^0)}{6(k+1)}) \\ & \leq \frac{(n_j^0 - n_{j-1}^0)^r}{(n_j^0 - n_{j-1}^0)^2} (6(k+1))^2 = n^{r-2} (\lambda_j^0 - \lambda_{j-1}^0)^{r-2} (6(k+1))^2, \end{aligned}$$

where

$$a = \max_{\theta_j \in \Theta_j, \alpha \in \Upsilon} |b(\alpha, \theta_j, \alpha^0, \theta_j^0)|.$$

Equation (8.3) can then be applied to show that  $I_{1j} \rightarrow 0$  as  $n, \eta \rightarrow \infty$ . Next, we consider the probability formula  $I_{2j}$  for any  $j = 2, \dots, k+1$ . In this case, we can see that

$$\lambda_{j-1} < \lambda_{j-1}^0.$$

We infer readily

$$\begin{aligned}
I_{2j} &\leq \mathbb{P} \left( \max_{\lambda \in \Lambda_{\eta, n}, \phi \in \Phi} \left\{ \frac{1}{n} \sum_{t \in \tilde{n}_{j,j-1}} \left[ \frac{g_j(\alpha, \theta_j, Y_t) \delta_t}{S_{G_{n_j^0}}(Y_t^-)} - \mathbb{E}(g_j(\alpha, \theta_j, X_t)) \right] + \frac{1}{6k} \mathbf{W}_1 \right\} > 0 \right) \\
&\quad + \mathbb{P} \left( \max_{\lambda \in \Lambda_{\eta, n}, \phi \in \Phi} \left\{ -\frac{1}{n} \sum_{t \in \tilde{n}_{j,j-1}} \left[ \frac{g_{j-1}(\alpha, \theta_{j-1}, Y_t) \delta_t}{S_{G_{n_{j-1}^0}}(Y_t^-)} - \mathbb{E}(g_{j-1}(\alpha, \theta_{j-1}, X_t)) \right] + \frac{1}{6k} \mathbf{W}_1 \right\} > 0 \right) \\
&\equiv I_{2j}^{(1)} + I_{2j}^{(2)}.
\end{aligned}$$

Notice that  $I_{2j}^{(1)}$  and  $I_{2j}^{(2)}$  can be handled in the same way, so we just show how to handle  $I_{2j}^{(1)}$ . Only two cases have to be considered.

If  $n_{j-1}^0 - n_{j-1} \leq \eta$ , then

$$\begin{aligned}
I_{2j}^{(1)} &\leq \mathbb{P} \left( \max_{n_{j-1} \leq s < t \leq n_{j-1}^0, \theta_j \in \Theta_j, \alpha \in \Upsilon} \left| \sum_{i=s+1}^t \left[ \frac{g_j(\alpha, \theta_j, Y_t) \delta_t}{S_{G_{n_j^0}}(Y_t^-)} - \mathbb{E}(g_j(\alpha, \theta_j, X_t)) \right] \right| > \frac{C_1 \eta}{6k} \right) \\
&\leq \frac{(n_{j-1}^0 - n_{j-1})^r}{(C_1 \eta)^2} (6k)^2 \\
&\leq \eta^{r-2} \left( \frac{6k}{C_1} \right)^2.
\end{aligned}$$

Equation (8.2) of Lemma 8.2 gives that  $I_{2j}^1 \rightarrow 0$ , as  $n, \eta \rightarrow \infty$ . If  $n_{j-1}^0 - n_{j-1} > \eta$ , for the other case, then we have

$$\mathbf{W}_1 \leq -C_1 \frac{(n_{j-1}^0 - n_{j-1})}{n}.$$

Therefore, we infer that

$$\begin{aligned}
I_{2j}^{(1)} &\leq \mathbb{P} \left( \max_{n_{j-1} \leq s < t \leq n_{j-1}^0, \theta_j \in \Theta_j, \alpha \in \Upsilon} \left| \sum_{i=s+1}^t \left[ \frac{g_j(\alpha, \theta_j, Y_t) \delta_t}{S_{G_{n_j^0}}(Y_t^-)} - \mathbb{E}(g_j(\alpha, \theta_j, X_t)) \right] \right| > \frac{C_1 (n_{j-1}^0 - n_{j-1})}{6k} \right) \\
&\leq (n_{j-1}^0 - n_{j-1})^{r-2} \left( \frac{6k}{C_1} \right)^2,
\end{aligned}$$

which converges to zero as  $n, \eta \rightarrow \infty$ , by equation (8.2) of Lemma (8.2).  $I_{3j}$  can be handled in a similar way as  $I_{2j}$ . Therefore the proof of Theorem 3.3 is complete.  $\square$

The following lemma establishes that the difference between the Kaplan Meier based on estimated proportion of the sample and the true one is asymptotically negligible.

**Lemma 8.3** Assume that, for  $i = 1, 2, \dots, k$ ,

$$\widehat{\lambda}_i - \lambda_i^0 = o_{\mathbb{P}}(1).$$

We have for each  $i = 1, 2, \dots, k$

$$(1 - \widehat{F}_{\widehat{\eta}_i}(x)) = (1 - F_{\eta_i^0}(x)) + o_{\mathbb{P}}(1).$$

### Proof of Lemma 8.3

For every  $\epsilon > 0$  there exist  $\eta' > 0$  and  $\eta'' > 0$  such that

$$\begin{aligned}
& \mathbb{P} \left( \sup_{x \leq \tau_{F_{n_j^0}}} |\widehat{F}_{\widehat{n}_j}(x) - \widehat{F}_{n_j^0}(x)| > \epsilon \right) \\
&= \mathbb{P} \left( \sup_{x \leq \tau_{F_{n_j^0}}} |\widehat{F}_{\widehat{n}_j}(x) - \widehat{F}_{n_j^0}(x)| > \epsilon, \widehat{n}_{j-1} = n_{j-1}^0, \widehat{n}_j = n_j^0 \right) \\
&\quad + \mathbb{P} \left( \sup_{x \leq \tau_{F_{n_j^0}}} |\widehat{F}_{\widehat{n}_j}(x) - \widehat{F}_{n_j^0}(x)| > \epsilon, \widehat{n}_{j-1} \neq n_{j-1}^0, \widehat{n}_j \neq n_j^0 \right) \\
&\quad + \mathbb{P} \left( \sup_{x \leq \tau_{F_{n_j^0}}} |\widehat{F}_{\widehat{n}_j}(x) - \widehat{F}_{n_j^0}(x)| > \epsilon, \widehat{n}_{j-1} \neq n_{j-1}^0, \widehat{n}_j = n_j^0 \right) \\
&\quad + \mathbb{P} \left( \sup_{x \leq \tau_{F_{n_j^0}}} |\widehat{F}_{\widehat{n}_j}(x) - \widehat{F}_{n_j^0}(x)| > \epsilon, \widehat{n}_{j-1} = n_{j-1}^0, \widehat{n}_j \neq n_j^0 \right) \\
&\leq 2\mathbb{P}(\widehat{n}_{j-1} \neq n_{j-1}^0) + 2\mathbb{P}(\widehat{n}_j \neq n_j^0) \\
&\leq 2\mathbb{P}(|\widehat{\lambda}_{j-1} - \lambda_{j-1}^0| > \eta') + 2\mathbb{P}(|\widehat{\lambda}_j - \lambda_j^0| > \eta'') \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.
\end{aligned}$$

Hence the proof is complete.  $\square$

The following lemma gives the approximation of the Kaplan Meier integral based on the estimated proportion of the sample.

**Lemma 8.4** *For any  $j = 1, \dots, k+1$ , under the conditions of Theorem 3.3 and the result of Lemma 8.3 we have*

$$\int_{\mathbb{R}} \psi_j(\alpha, \theta_j, x) d\widehat{F}_{\widehat{n}_j}(x) - \int_{\mathbb{R}} \psi_j(\alpha, \theta_j, x) d\widehat{F}_{n_j^0}(x) = O_{\mathbb{P}} \left( \frac{1}{n} \right).$$

As a consequence of this lemma, for every  $\phi \in \Phi$ , we have that

$$\begin{aligned}
\rho_n(\alpha, \theta_1, \dots, \theta_{k+1}) &= \sum_{j=1}^{k+1} (\widehat{\lambda}_j - \widehat{\lambda}_{j-1}) \int_{\mathbb{R}} \psi_j(\alpha, \theta_j, x) d\widehat{F}_{\widehat{n}_j}(x) \\
&= \sum_{j=1}^{k+1} (\lambda_j^0 - \lambda_{j-1}^0 + O_{\mathbb{P}}(n^{-1})) \left( \int_{\mathbb{R}} \psi_j(\alpha, \theta_j, x) d\widehat{F}_{n_j^0}(x) + O_{\mathbb{P}}(n^{-1}) \right) \\
&= \rho_n^0(\alpha, \theta_1, \dots, \theta_{k+1}) + O_{\mathbb{P}}(n^{-1}).
\end{aligned} \tag{8.4}$$

## Proof of Lemma 8.4

We have

$$\begin{aligned}
& \int_{\mathbb{R}} \psi_j(\alpha, \theta_j, x) d\widehat{F}_{\widehat{n}_j}(x) - \int_{\mathbb{R}} \psi_j(\alpha, \theta_j, x) d\widehat{F}_{n_j^0}(x) \\
&= \frac{\sum_{i=\widehat{n}_{j-1}+1}^{\widehat{n}_j} (n_j^0 - n_{j-1}^0) \frac{\psi_j(\alpha, \theta_j, Y_i) \Delta_i}{S_{C^j}(Y_i^-)} - \sum_{i=n_{j-1}^0+1}^{n_j^0} (\widehat{n}_j - \widehat{n}_{j-1}) \frac{\psi_j(\alpha, \theta_j, Y_i) \Delta_i}{S_{C^j}(Y_i^-)}}{(n_j^0 - n_{j-1}^0)(\widehat{n}_j - \widehat{n}_{j-1})} \\
&= \mathbb{I}_{\{\widehat{n}_j \leq n_j^0, n_{j-1}^0 \leq \widehat{n}_{j-1}\}} \left[ \sum_{i=n_{j-1}^0+1}^{\widehat{n}_{j-1}} \frac{-\psi_j(\alpha, \theta_j, Y_i) \Delta_i}{(n_j^0 - n_{j-1}^0) \left( S_{G_{n_j^0}}(Y_i^-) + o_{\mathbb{P}}(1) \right)} \right. \\
&\quad + \sum_{i=\widehat{n}_{j-1}+1}^{\widehat{n}_j} \frac{((n_j^0 - n_{j-1}^0) - (\widehat{n}_j - \widehat{n}_{j-1})) \psi_j(\alpha, \theta_j, Y_i) \Delta_i}{(\widehat{n}_j - \widehat{n}_{j-1})(n_j^0 - n_{j-1}^0) \left( S_{G_{n_j^0}}(Y_i^-) + o_{\mathbb{P}}(1) \right)} \\
&\quad \left. + \sum_{i=\widehat{n}_j+1}^{n_j^0} \frac{-\psi_j(\alpha, \theta_j, Y_i) \Delta_i}{(n_j^0 - n_{j-1}^0) \left( S_{G_{n_j^0}}(Y_i^-) + o_{\mathbb{P}}(1) \right)} \right] \\
&+ \mathbb{I}_{\{\widehat{n}_j \leq n_j^0, \widehat{n}_{j-1} < n_{j-1}^0\}} \left[ \sum_{i=\widehat{n}_{j-1}+1}^{n_{j-1}^0} \frac{\psi_j(\alpha, \theta_j, Y_i) \Delta_i}{(\widehat{n}_j - \widehat{n}_{j-1}) \left( S_{G_{n_j^0}}(Y_i^-) + o_{\mathbb{P}}(1) \right)} \right. \\
&\quad + \sum_{i=n_{j-1}^0+1}^{\widehat{n}_j} \frac{((n_j^0 - n_{j-1}^0) - (\widehat{n}_j - \widehat{n}_{j-1})) \psi_j(\alpha, \theta_j, Y_i) \Delta_i}{(\widehat{n}_j - \widehat{n}_{j-1})(n_j^0 - n_{j-1}^0) \left( S_{G_{n_j^0}}(Y_i^-) + o_{\mathbb{P}}(1) \right)} \\
&\quad \left. + \sum_{i=\widehat{n}_j+1}^{n_j^0} \frac{-\psi_j(\alpha, \theta_j, Y_i) \Delta_i}{(n_j^0 - n_{j-1}^0) \left( S_{G_{n_j^0}}(Y_i^-) + o_{\mathbb{P}}(1) \right)} \right] \\
&+ \mathbb{I}_{\{n_j^0 < \widehat{n}_j, n_{j-1}^0 \leq \widehat{n}_{j-1}\}} \left[ \sum_{i=n_{j-1}^0+1}^{\widehat{n}_{j-1}} \frac{-\psi_j(\alpha, \theta_j, Y_i) \Delta_i}{(n_j^0 - n_{j-1}^0) \left( S_{G_{n_j^0}}(Y_i^-) + o_{\mathbb{P}}(1) \right)} \right. \\
&\quad + \sum_{i=\widehat{n}_{j-1}+1}^{n_j^0} \frac{((n_j^0 - n_{j-1}^0) - (\widehat{n}_j - \widehat{n}_{j-1})) \psi_j(\alpha, \theta_j, Y_i) \Delta_i}{(\widehat{n}_j - \widehat{n}_{j-1})(n_j^0 - n_{j-1}^0) \left( S_{G_{n_j^0}}(Y_i^-) + o_{\mathbb{P}}(1) \right)} \\
&\quad \left. + \sum_{i=n_j^0+1}^{\widehat{n}_j} \frac{\psi_j(\alpha, \theta_j, Y_i) \Delta_i}{(\widehat{n}_j - \widehat{n}_{j-1}) \left( S_{G_{n_j^0}}(Y_i^-) + o_{\mathbb{P}}(1) \right)} \right]
\end{aligned}$$

$$\begin{aligned}
& + \mathbb{I}_{\{n_j^0 < \widehat{n}_j, \widehat{n}_{j-1} < n_{j-1}^0\}} \left[ \sum_{i=\widehat{n}_{j-1}+1}^{n_{j-1}^0} \frac{\psi_j(\alpha, \theta_j, Y_i) \Delta_i}{(\widehat{n}_j - \widehat{n}_{j-1}) \left( S_{G_{n_j^0}}(Y_i^-) + o_{\mathbb{P}}(1) \right)} \right. \\
& + \sum_{i=n_{j-1}^0+1}^{n_j^0} \frac{((n_j^0 - n_{j-1}^0) - (\widehat{n}_j - \widehat{n}_{j-1})) \psi_j(\alpha, \theta_j, Y_i) \Delta_i}{(\widehat{n}_j - \widehat{n}_{j-1})(n_j^0 - n_{j-1}^0) \left( S_{G_{n_j^0}}(Y_i^-) + o_{\mathbb{P}}(1) \right)} \\
& \left. + \sum_{i=n_j^0+1}^{\widehat{n}_j} \frac{\psi_j(\alpha, \theta_j, Y_i) \Delta_i}{(\widehat{n}_j - \widehat{n}_{j-1}) \left( S_{G_{n_j^0}}(Y_i^-) + o_{\mathbb{P}}(1) \right)} \right].
\end{aligned}$$

An application of Theorem 3.3 gives the desired result.  $\square$

### Proof of Theorem 4.1

For every  $\epsilon > 0$  there exists  $\eta > 0$ , such that we have

$$\begin{aligned}
& \mathbb{P} \left( \|\widehat{\phi} - \phi^0\| > \epsilon \right) \\
& \leq \mathbb{P} \left( \|\rho(\widehat{\phi}) - \rho(\phi^0)\| > \eta \right) \\
& \leq \mathbb{P} \left( \|\rho(\widehat{\phi}) - \rho_n^0(\widehat{\phi}) + \rho_n(\widehat{\phi}) - \rho_n^0(\widehat{\phi}) + \rho_n(\widehat{\phi}) - \rho(\phi^0)\| > \eta \right) \\
& \leq \mathbb{P} \left( \sup_{\phi \in \Phi} \|\rho_n^0(\phi) - \rho(\phi)\| > \frac{\eta}{3} \right) + \mathbb{P} \left( \|\rho_n(\widehat{\phi}) - \rho_n^0(\widehat{\phi})\| > \frac{\eta}{3} \right) \\
& \quad + \mathbb{P} \left( \|\rho_n(\widehat{\phi}) - \rho(\phi^0)\| > \frac{\eta}{3} \right),
\end{aligned}$$

the assumptions of Theorem 4.1 combined with the relation (8.4) show that the last term converges in probability to zero as  $n$  converges to infinity.  $\square$

### Proof of Theorem 4.2

Let us first take  $\epsilon > 0$  and  $\eta > 0$  fixed constants. Condition (ii) implies that there exists a finite  $\mathfrak{M}$ , such that for large value of  $n$ , we have

$$\mathbb{P} \left( \sup_{\|\phi - \phi_0\| > \eta} \|\rho_n(\phi)\|^{-1} > \mathfrak{M} \right) < \epsilon.$$

Notice that the parameter  $\widehat{\phi}$  satisfies

$$\rho_n(\widehat{\phi}) = O_{\mathbb{P}}(1),$$

so we readily obtain

$$\mathbb{P} \left( \|\rho_n(\widehat{\phi})\|^{-1} > \mathfrak{M} \right) \longrightarrow 1.$$

It follows that, with probability of at least  $1 - \epsilon$  for all  $n$  large enough,

$$\|\rho_n(\widehat{\phi})\|^{-1} > \mathfrak{M} \geq \sup_{\|\phi - \phi_0\| > \eta} \|\rho_n(\phi)\|^{-1}.$$

These inequalities force  $\widehat{\phi}$  to lie within a distance  $\eta$  of  $\phi^0$ , that is,

$$\mathbb{P} \left( \|\widehat{\phi} - \phi^0\| > \eta \right) \leq \epsilon.$$

Since  $\epsilon$  and  $\eta$  can be chosen arbitrarily close to zero, the asserted convergence in probability is established.  $\square$

### Proof of Theorem 4.3

We will follow the proof of [Pakes and Pollard \(1989\)](#). First we prove  $\sqrt{n}$ -consistency. The assumed consistency allows us to choose a sequence  $\eta_n$  that converge to zero slowly enough to ensure that

$$\mathbb{P} \left( \|\widehat{\phi} - \phi^0\| > \eta_n \right) \longrightarrow 0.$$

With probability tending to one for this sequence, the supremum in the condition (iii) runs over a range that includes the random value  $\widehat{\phi}$ . Thus we have

$$\|\rho_n^0(\widehat{\phi}) - \rho(\widehat{\phi}) - \rho_n^0(\phi^0)\| \leq o_{\mathbb{P}}(n^{-1/2}) + o_{\mathbb{P}}(\|\rho_n^0(\widehat{\phi})\|) + o_{\mathbb{P}}(\|\rho(\widehat{\phi})\|).$$

By the triangle inequality, the left-hand side is larger than

$$\|\rho(\widehat{\phi})\| - \|\rho_n^0(\widehat{\phi})\| - \|\rho_n^0(\phi^0)\|.$$

Thus we obtain

$$\|\rho(\widehat{\phi})\|[1 - o_{\mathbb{P}}(1)] \leq o_{\mathbb{P}}(n^{-1/2}) + \|\rho_n^0(\widehat{\phi})\|[1 + o_{\mathbb{P}}(1)] + \|\rho_n^0(\phi^0)\|.$$

From conditions (i) and the asymptotic normality of  $\sqrt{n}\rho_n^0(\phi^0)$  it follows that

$$\|\rho(\widehat{\phi})\| = O_{\mathbb{P}}(n^{-1/2}).$$

The differentiability condition (ii) implies the existence of a positive constant  $C$  for which, near  $\phi^0$ , (recall that  $\rho(\phi^0) = 0$ ), we have

$$\|\rho(\phi)\| \geq C\|\phi - \phi^0\|.$$

In particular, we infer that

$$\|\widehat{\phi} - \phi^0\| = O_{\mathbb{P}} \left( \|\rho(\widehat{\phi})\| \right) = O_{\mathbb{P}} \left( n^{-1/2} \right).$$

Next, we establish asymptotic normality of  $\sqrt{n}(\widehat{\phi} - \phi^0)$ , by arguing that  $\rho_n^0(\phi)$  is very well approximated by the linear function

$$L_n(\phi) = \Omega(\phi - \phi^0) + \rho_n^0(\phi^0)$$

within a  $O_{\mathbb{P}}(n^{-1/2})$  neighborhood of  $\phi^0$ . More precisely, we need the approximation error to be of order  $o_{\mathbb{P}}(n^{-1/2})$  at  $\widehat{\phi}$  and at the  $\phi_n^*$  that maximizes  $\|L_n(\cdot)\|$  globally. This follows directly from (ii) and (iii) together with the  $\sqrt{n}$ -consistency results already established

$$\begin{aligned} \left\| \rho_n^0(\widehat{\phi}) - L_n(\widehat{\phi}) \right\| &\leq \left\| \rho_n^0(\widehat{\phi}) - \rho(\widehat{\phi}) - \rho_n^0(\phi^0) \right\| \\ &\quad + \|\rho(\widehat{\phi}) - \Omega(\widehat{\phi} - \phi^0)\| \\ &\leq o_{\mathbb{P}}(n^{-1/2}) + o_{\mathbb{P}}(\|\rho_n^0(\widehat{\phi})\|) + o_{\mathbb{P}}(\|\rho(\widehat{\phi})\|) \\ &\quad + o_{\mathbb{P}}(\|\widehat{\phi} - \phi^0\|) \\ &= o_{\mathbb{P}}(n^{-1/2}). \end{aligned}$$

To correspond to a minimum of  $\|L_n(\cdot)\|$ , the vector  $\Omega(\phi_n^* - \phi^0)$  must be equal to the projection of  $-\rho_n^0(\phi^0)$  onto the column space of  $\Omega$ . Hence, we obtain

$$\sqrt{n}(\phi_n^* - \phi^0) = -\sqrt{n}(\Omega^{\top}\Omega)^{-1}\Omega^{\top}\rho_n^0(\phi^0).$$

The right-hand side has the asymptotic normal distribution specified in the statement of the theorem. Consequently

$$\phi_n^* = \phi^0 + O_{\mathbb{P}}(n^{-1/2}).$$

Because  $\phi^0$  is in the interior point of  $\Phi$  this implies that  $\phi_n^*$  lies in  $\Phi$  with probability tending to one. From the differentiability condition (ii) and condition (iii), we readily obtain that

$$\|\rho_n^0(\phi_n^*)\| = O_{\mathbb{P}}(n^{-1/2}).$$

Then we can argue as for  $\widehat{\phi}$  to deduce that

$$\|\rho_n^0(\phi_n^*) - L_n(\phi_n^*)\| = o_{\mathbb{P}}(n^{-1/2}).$$

We now know that  $\rho_n^0$  and  $L_n$  are close at both  $\widehat{\phi}$ , which almost minimizes  $\|\rho_n^0\|$ , and  $\phi_n^*$ , which minimizes  $\|L_n\|$ . This forces  $\widehat{\phi}$  to come close to minimizing  $\|L_n\|$ . That is,

$$\|L_n(\phi_n^*)\| = \|L_n(\widehat{\phi})\| + o_{\mathbb{P}}(n^{-1/2}).$$

So we have

$$\|L_n(\phi_n^*)\|^2 = \|L_n(\widehat{\phi})\|^2 + o_{\mathbb{P}}(n^{-1}),$$

the across product term being absorbed into  $o_{\mathbb{P}}(n^{-1})$  because  $\|L_n(\phi_n^*)\|$  is of order  $O_{\mathbb{P}}(n^{-1/2})$ . The quadratic form of  $\|L_n(\phi)\|^2$  has the simple expansion

$$\|L_n(\phi)\|^2 = \|L_n(\phi_n^*)\|^2 + \|\Omega(\phi - \phi_n^*)\|^2,$$

about its global minimum. Put  $\phi$  equal to  $\widehat{\phi}$ , then equate the two expressions for  $\|L_n(\widehat{\phi})\|^2$  to deduce that

$$\|\Omega(\widehat{\phi} - \phi_n^*)\|^2 = o_{\mathbb{P}}(n^{-1/2}).$$

Since the matrix  $\Omega$  has full rank, this is equivalent to

$$\sqrt{n}(\widehat{\phi} - \phi^0) = \sqrt{n}(\phi_n^* - \phi^0) + o_{\mathbb{P}}(1),$$

from which the asserted central limit theorem follows.  $\square$

If we replace conditions (i) by (i)' and (iii) by (iii)' in Theorem 4.3 we will obtain the same result of Theorem (3.3) in Pakes and Pollard (1989) under each true sub sample, we get  $L_n(\phi)$  is sum of  $k + 1$  linear function given by

$$L_{n_j^0}(\phi) = \Gamma_{F_{n_j^0}}(\phi^0)(\phi - \phi^0) + \rho_{n_j^0}^0(\phi^0), j = 1, 2, \dots, k + 1.$$

For notation ease, we put  $\phi$  in function for each subsample because there is no influence for other parameters to the ones we are working on.

The following lemma gives the convergence of the Kaplan Meier integrals.

**Lemma 8.5** *Let  $s(\phi, x)$  be any real function with, for any  $j = 1, 2, \dots, k + 1$ ,*

$$\int_{\mathbb{R}} |s(\phi^0, x)| dF_{n_j^0}(x) < \infty.$$

*Assume that the condition (R1) (in the appendix) with replacement of the functions  $H(\cdot)$ ,  $F(\cdot)$  and  $G(\cdot)$  by the functions  $H_{n_j^0}(\cdot)$ ,  $F_{n_j^0}(\cdot)$  and  $G_{n_j^0}(\cdot)$  respectively for each  $j = 1, 2, \dots, k + 1$ , holds for*

$$\varphi(x) = s(\phi^0, x).$$

*For any sequence  $\widehat{\phi} \xrightarrow{\mathbb{P}} \phi^0$ , it follows that, for any  $j = 1, 2, \dots, k + 1$ ,*

$$\int_{\mathbb{R}} s(\widehat{\phi}, x) d\widehat{F}_{\widehat{n}_j}(x) \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} s(\phi^0, x) dF_{n_j^0}(x),$$

*provided that any one of the following conditions holds, for any  $j = 1, 2, \dots, k + 1$ ,*

(i)  $s(\phi, x)$  is continuous at  $\phi^0$  uniformly in  $x$ .

(ii)

$$\int_{\mathbb{R}} \sup_{\{\phi: |\phi - \phi^0| \leq \beta\}} |s(\phi, x) - s(\phi^0, x)| dF_{n_j^0}(x) = h_{\beta} \rightarrow 0 \text{ as } \beta \rightarrow 0.$$

(iii)  $s(\cdot, \cdot)$  is continuous in  $x$  for  $\phi$  in a neighborhood of  $\phi^0$ , and

$$\lim_{\phi \rightarrow \phi^0} \|s(\phi, \cdot) - s(\phi^0, \cdot)\|_V = 0.$$

(iv)  $\int_{\mathbb{R}} s(\phi, x) dF_{n_j^0}(x)$  is continuous at  $\phi = \phi^0$ , and  $s$  is continuous in  $x$  for  $\phi$  in a neighborhood of  $\phi^0$ , and

$$\lim_{\phi \rightarrow \phi^0} \|s(\phi, \cdot) - s(\phi^0, \cdot)\|_V < \infty.$$

(v)  $\int_{\mathbb{R}} s(\phi, x) dF_{n_j^0}(x)$  is continuous at  $\phi = \phi^0$ , and

$$\int_{\mathbb{R}} s(\phi, x) d\widehat{F}_{n_j^0}(x) \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} s(\phi, x) dF_{n_j^0}(x) < \infty,$$

uniformly for  $\phi$  in a neighborhood of  $\phi^0$ .

### Proof of Lemma 8.5

The proof of this lemma is based on the Lemma 8.4 and Lemma 1 in Wang (1999).  $\square$

### Proof of Theorem 4.4

Note that  $\rho_n(\phi)$  is differentiable in  $\phi$  by the conditions imposed on  $\psi_j(\cdot)$ . The multivariate mean value theorem thus implies that

$$\rho_n(\widehat{\phi}) = \rho_n(\phi^0) + \left( \sum_{j=1}^{k+1} (\widehat{\lambda}_j - \widehat{\lambda}_{j-1}) \Gamma_{\widehat{F}_{\widehat{n}_j}}(\xi_n) \right) (\widehat{\phi} - \phi^0),$$

where

$$\|\xi_n - \phi^0\| \leq \|\widehat{\phi} - \phi^0\|$$

and recall that  $\|\cdot\|$  is the Euclidean norm. By using the fact that

$$\rho_n(\widehat{\phi}) = 0, \rho_{n_j^0}(\alpha^0, \theta_j^0) = 0$$

in combination with Lemma 8.4, we infer that

$$\begin{aligned} \sqrt{n}(\widehat{\phi} - \phi^0) &= - \left( \sum_{j=1}^{k+1} (\widehat{\lambda}_j - \widehat{\lambda}_{j-1}) \Gamma_{\widehat{F}_{\widehat{n}_j}}(\xi_n) \right)^{-1} \left[ \sum_{j=1}^{k+1} (\widehat{\lambda}_j - \widehat{\lambda}_{j-1}) \left( \sqrt{n} \left\{ \int \psi_j(\alpha^0, \theta_j^0, x) d\widehat{F}_{n_j^0}(x) \right. \right. \right. \\ &\quad \left. \left. \left. - \int \psi_j(\alpha^0, \theta_j^0, x) dF_{n_j^0}(x) \right\} + O_{\mathbb{P}}(n^{-1/2}) \right) \right]. \end{aligned}$$

Once more, Lemma 8.4 implies that we have

$$\begin{aligned} \rho_n(\phi^0) &= \sum_{j=1}^{k+1} (\widehat{\lambda}_j - \widehat{\lambda}_{j-1}) \int \psi_j(\alpha^0, \theta_j^0, x) d\widehat{F}_{\widehat{n}_j}(x) \\ &= \sum_{j=1}^{k+1} (\widehat{\lambda}_j - \widehat{\lambda}_{j-1}) \left[ \int \psi_j(\alpha^0, \theta_j^0, x) d\widehat{F}_{n_j^0}(x) + O_{\mathbb{P}}(n^{-1}) \right]. \end{aligned}$$

By Theorem 3.3, we have entries of

$$\sum_{j=1}^{k+1} (\widehat{\lambda}_j - \widehat{\lambda}_{j-1}) \Gamma_{\widehat{F}_{\widehat{n}_j}}(\xi_n)$$

converges in probability to the entries of  $\Gamma(\phi^0)$ . The theorem now follows from combining Proposition 9.2, Theorem 3.3 and Slutsky's theorem.  $\square$

## 9 Appendix

In the sequel of this section, we use a notation similar to that used in Wang (1999) including some changes absolutely necessary for our setting. We present, for the convenience of the reader, the random censorship model in Section 1 without change points. Let  $F(\cdot)$  denote the lifetime distribution of  $X$  and  $G(\cdot)$  the censoring distribution of  $C$ . Assume the independence of  $X$  and  $C$ , which implies that the distribution  $H(\cdot)$  of the observation  $Y = \min(X, C)$  satisfies

$$1 - H(\cdot) = (1 - F(\cdot))(1 - G(\cdot)).$$

## 9.1 SLLN and CLT for Kaplan Meier integrals

Let us begin by introducing some results on the Kaplan Meier integrals playing a central role in this study. For any specified real function  $\varphi(\cdot)$ , we state in this section the strong law of large numbers (SLLN) and the central limit theorem (CLT) for the Kaplan Meier integral

$$\int_{\mathbb{R}} \varphi(x) d\widehat{F}_n(x).$$

Such results constitute the main tools to study the limiting behavior of  $M$ -( $Z$ )-estimates in the next sections. For any distribution function  $L(\cdot)$ , let

$$\tau_L = \sup\{x : L(x) < 1\}$$

denote the upper bound of the support of  $L(\cdot)$ . Let

$$\Delta L(x) = L(x) - L(x^-)$$

denote the probability mass of  $L(\cdot)$  at  $x$ . Since one can only observe data in the range of  $[0, \tau_H]$ , it is possible to estimate  $\int_{\mathbb{R}} \varphi(x) dF(x)$  consistency only if  $\tau_F = \tau_H$  or if  $\varphi(x)$  is zero for  $x \geq \tau_H$ . The specific requirement for strong consistency is formulated in the following condition :

(R1) at least one of (i) or (ii) below holds:

(i) For some  $u < \tau_H$ ,  $\varphi(x) = 0$  for  $u < x \leq \tau_H$ .

(ii)  $\tau_F \leq \tau_G$ , where equality may hold except when  $G(\cdot)$  is continuous at  $\tau_F$  and

$$\Delta F(\tau_F) > 0.$$

Note that (R1) (ii) implies  $\tau_F = \tau_H$ , and is the necessary and sufficient condition so that  $F(\cdot)$  can be estimated consistently on its entire support. Such a requirement can be dispensed with only the fact that the function  $\varphi(\cdot)$  satisfies the requirement (R1) (i) which then results in a truncated Kaplan-Meier integral. Note that only one of the two, but not both, conditions in (i) and (ii) need to hold for (R1). We state in the next proposition the strong consistency of

$$\int_{\mathbb{R}} \varphi(x) d\widehat{F}_n(x),$$

which follows from the condition (R1), Theorem 1.1 and Corollary 1.2 of [Stute and Wang \(1993\)](#). Note that the original strong law in [Stute and Wang \(1993\)](#) requires further that  $F(\cdot)$  and  $G(\cdot)$  have no common point of discontinuity. Such a restriction was later discovered to be dispensable, see [Stute \(1995\)](#) for details.

**Proposition 9.1 (Strong law of Large Numbers)** *Under the condition (R1) and for any function  $\varphi(\cdot)$  fulfilling*

$$\int_{\mathbb{R}} |\varphi(x)| dF(x) < \infty,$$

*it follows, with probability one, that*

$$\int_{\mathbb{R}} \varphi(x) d\widehat{F}_n(x) \rightarrow \int_{\mathbb{R}} \varphi(x) dF(x).$$

*Moreover, under (R1) (ii), it follows, with probability one, that*

$$\sup_{-\infty < x \leq \tau_H} |\widehat{F}_n(x) - F(x)| \rightarrow 0.$$

Proposition 9.1 essentially implies that the law of large numbers for censored data hold under the same condition, namely the integrability of  $\varphi(\cdot)$ , as for the uncensored case. The CLT however requires a little more than the uncensored case. Denote  $m(y) = \mathbb{P}(\delta = 1 | Y = y)$  and denote the subdistribution functions for the censored and uncensored observations, respectively, by

$$\begin{aligned} H_0(y) &= \mathbb{P}(Y \leq y, \delta = 0) = \int_{-\infty}^y (1 - m(t)) dH(t) = \int_{-\infty}^y (1 - F(t)) dG(t), \\ H_1(y) &= \mathbb{P}(Y \leq y, \delta = 1) = \int_{-\infty}^y m(t) dH(t) = \int_{-\infty}^y (1 - G(t^-)) dF(t), \end{aligned} \tag{9.1}$$

and let the corresponding empirical estimates be denoted by

$$H_{pn}(y) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{Y_i \leq y, \delta_i = p\}}, \quad \text{for } p = 0, 1. \quad (9.2)$$

Note that

$$H_0(\cdot) + H_1(\cdot) = H(\cdot).$$

The asymptotic representation of  $\int_{\mathbb{R}} \varphi(x) d\widehat{F}_n(x)$  as a sum of i.i.d. variables defined in (9.5) and (9.7), is based upon the following expressions

$$\begin{aligned} \gamma_0(x) &= \exp \left\{ \int_{\mathbb{R}} \frac{\mathbb{I}_{\{y < x\}} dH_0(y)}{1 - H(y)} \right\}, \\ \gamma_1(x) &= [1 - H(x)]^{-1} \int_{\mathbb{R}} \mathbb{I}_{\{y < x\}} \varphi(x) \gamma_0(y) dH_1(y), \\ \gamma_2(x) &= \int_{\mathbb{R}} \varphi(z) \gamma_0(z) C(x \wedge z) dH_1(z), \end{aligned} \quad (9.3)$$

where

$$C(x) = \int_{\mathbb{R}} \frac{\mathbb{I}_{\{y < x\}} dH_0(y)}{[1 - H(y)]^2} = \int_{\mathbb{R}} \frac{\mathbb{I}_{\{y < x\}} dG(y)}{[1 - F(y)][1 - G(y)]^2}, \quad (9.4)$$

refer to [Stute \(1995\)](#) for more details. Let  $U$  denote the random variable defined by

$$U = \varphi(Y) \gamma_0(Y) \delta + \gamma_1(Y) (1 - \delta) - \gamma_2(Y) - \int_{\mathbb{R}} \varphi(x) dF(x). \quad (9.5)$$

It turns out that  $\mathbb{E}(U) = 0$ . The variance of  $U$  depends on  $\varphi(\cdot)$ ,  $F(\cdot)$  and  $G(\cdot)$  and is given by

$$\begin{aligned} \sigma^2(\varphi, F, G) &= \text{Var}(U) \\ &= \int_{\mathbb{R}} \varphi^2(y) \gamma_0^2(y) dH_1(y) - \int_{\mathbb{R}} \gamma_1^2(y) dH_0(y) \\ &\quad - \left( \int_{\mathbb{R}} \varphi(x) dF(x) \right)^2 + \int_{\mathbb{R}} \frac{\gamma_1^2(y) [1 - m(y)]^2}{1 - H(y)} \Delta H(y) dH(y). \end{aligned} \quad (9.6)$$

Clearly, the last integral vanishes for a continuous  $H(\cdot)$ . The additional requirements for the asymptotic normality of  $\int_{\mathbb{R}} \varphi(x) d(\widehat{F}_n(x) - F(x))$  are

(R2)

$$\mathbb{E}[\varphi(Y) \gamma_0(Y) \delta]^2 = \int_{\mathbb{R}} \varphi^2(y) \gamma_0^2(y) dH_1(y) < \infty,$$

(R3)

$$\int_{\mathbb{R}} |\varphi(x)| C^{1/2}(x) dF(x) < \infty.$$

For more discussion of these conditions see [Wang \(1999\)](#). We now present the asymptotic normality results of

$$\int \varphi(x) d(\widehat{F}_n(x) - F(x)),$$

which follow from Theorem 1 of [Stute \(1995\)](#) and (R1).

**Proposition 9.2 (Central limit theorem)** *Assume that the conditions (R1)-(R3) are satisfied. Then we have the following representation*

$$\int_{\mathbb{R}} \varphi(x) d(\widehat{F}_n - F)(x) = n^{-1} \sum_{i=1}^n U_i + o_{\mathbb{P}}(n^{-1/2}), \quad (9.7)$$

where the  $U_i$ s are i.i.d. copies of the variable  $U$  by replacing the  $Y$  and  $\delta$  in (9.5) by  $Y_i$  and  $\delta_i$ , respectively. Thus, for  $\sigma^2(\varphi, F, G)$  defined in (9.6), we have the following convergence in distribution, as  $n \rightarrow \infty$ ,

$$n^{1/2} \int_{\mathbb{R}} \varphi(x) d(\widehat{F}_n - F)(x) \rightarrow N(0, \sigma^2(\varphi, F, G)). \quad (9.8)$$

For continuous distribution function  $H(\cdot)$ , the asymptotic variance in (9.8) becomes

$$\sigma^2(\varphi, F, G) = \int_{-\infty}^{\infty} \frac{\left( \int_x^{\infty} \varphi'(t)[1 - F(t)]dt \right)^2}{[1 - H(x)]^2} dH_1(x). \quad (9.9)$$

The last equality in (9.9) follows from (9.1). A variance estimate can be obtained by replacing  $F(\cdot)$ ,  $H_1(\cdot)$  and  $H(\cdot)$  respectively by their empirical estimates, for more details we refer the reader to Wang (1999).

## Acknowledgement

The authors are indebted to the Editor-in-Chief, Associate Editor and the referee for their very valuable comments, suggestions careful reading of the article which led to a considerable improvement of the manuscript.

## References

- Al-Awadhi, F. and Aly, E.-E. A. A. (2005). On the performance of logrank tests in change point problems for randomly censored data. *J. Stat. Theory Appl.*, **4**(3), 292–302.
- Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (2012). *Statistical models based on counting processes*. Springer Science & Business Media.
- Aue, A. and Horváth, L. (2013). Structural breaks in time series. *J. Time Series Anal.*, **34**(1), 1–16.
- Bae, J. and Kim, S. (2003). The uniform law of large numbers for the Kaplan-Meier integral process. *Bull. Austral. Math. Soc.*, **67**(3), 459–465.
- Bahadur, R. R. (1967). Rates of convergence of estimates and test statistics. *Ann. Math. Statist.*, **38**, 303–324.
- Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, pages 47–78.
- Bhattacharya, P. K. (1987). Maximum likelihood estimation of a change-point in the distribution of independent random variables: general multiparameter case. *J. Multivariate Anal.*, **23**(2), 183–208.
- Borgan, Ø. (1984). Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scandinavian Journal of Statistics*, pages 1–16.
- Bouzebda, S. (2014). Asymptotic properties of pseudo maximum likelihood estimators and test in semi-parametric copula models with multiple change points. *Math. Methods Statist.*, **23**(1), 38–65.
- Bouzebda, S. and Keziou, A. (2013). A semiparametric maximum likelihood ratio test for the change point in copula models. *Stat. Methodol.*, **14**, 39–61.
- Brodsky, B. E. and Darkhovsky, B. S. (1993). *Nonparametric methods in change-point problems*, volume 243 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht.
- Chen, H. (2019). Sequential change-point detection based on nearest neighbors. *Ann. Statist.*, **47**(3), 1381–1407.
- Chen, J. and Gupta, A. K. (2000). *Parametric statistical change point analysis*. Birkhäuser Boston, Inc., Boston, MA.
- Chu, L. and Chen, H. (2019). Asymptotic distribution-free change-point detection for multivariate and non-Euclidean data. *Ann. Statist.*, **47**(1), 382–414.
- Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton Mathematical Series, vol. 9. Princeton University Press, Princeton, N. J.
- Csörgő, M. and Horváth, L. (1997). *Limit theorems in change-point analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester. With a foreword by David Kendall.

- Döring, M. (2011). Convergence in distribution of multiple change point estimators. *J. Statist. Plann. Inference*, **141**(7), 2238–2248.
- Dumbgen, L. *et al.* (1991). The asymptotic behavior of some nonparametric change-point estimators. *The Annals of Statistics*, **19**(3), 1471–1495.
- El Ktaibi, F. and Ivanoff, B. G. (2019). Bootstrapping the empirical distribution of a stationary process with change-point. *Electron. J. Stat.*, **13**(2), 3572–3612.
- Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *Ann. Statist.*, **42**(6), 2243–2281.
- Fryzlewicz, P. (2018). Tail-greedy bottom-up data decompositions and fast multiple change-point detection. *Ann. Statist.*, **46**(6B), 3390–3421.
- Fu, Y.-X. and Curnow, R. N. (1990a). Locating a changed segment in a sequence of Bernoulli variables. *Biometrika*, **77**(2), 295–304.
- Fu, Y.-X. and Curnow, R. N. (1990b). Maximum likelihood estimation of multiple change points. *Biometrika*, **77**(3), 563–573.
- Garreau, D. and Arlot, S. (2018). Consistent change-point detection with kernels. *Electron. J. Stat.*, **12**(2), 4440–4486.
- Goldenshluger, A., Juditsky, A., Tsybakov, A. B., and Zeevi, A. (2008). Change-point estimation from indirect observations. I. Minimax complexity. *Ann. Inst. Henri Poincaré Probab. Stat.*, **44**(5), 787–818.
- Gombay, E. and Horváth, L. (1994). An application of the maximum likelihood test to the change-point problem. *Stochastic Process. Appl.*, **50**(1), 161–171.
- Gombay, E. and Liu, S. (2000). A nonparametric test for change in randomly censored data. *Canad. J. Statist.*, **28**(1), 113–121.
- Hao, N., Niu, Y. S., and Zhang, H. (2013). Multiple change-point detection via a screening and ranking algorithm. *Statist. Sinica*, **23**(4), 1553–1572.
- Hawkins, D. M. (2001). Fitting multiple change-point models to data. *Comput. Statist. Data Anal.*, **37**(3), 323–341.
- He, C. (2017). Bayesian multiple change-point estimation for exponential distribution with truncated and censored data. *Comm. Statist. Theory Methods*, **46**(12), 5827–5839.
- He, C. B. (2015). Parameter estimation of Weibull distribution with multiple change points for truncated and censored data. *Appl. Math. J. Chinese Univ. Ser. A*, **30**(2), 127–138.
- He, H. and Severini, T. A. (2010). Asymptotic properties of maximum likelihood estimators in models with multiple change points. *Bernoulli*, **16**(3), 759–779.
- Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika*, **57**, 1–17.
- Hinkley, D. V. (1972). Time-ordered classification. *Biometrika*, **59**, 509–523.
- Hinkley, D. V. and Hinkley, E. A. (1970). Inference about the change-point in a sequence of binomial variables. *Biometrika*, **57**, 477–488.
- Hjort, N. L. (1985). Discussion of the paper by andersen, p. k. and borgan,  $\phi$ . *Scand. J. Statist.*, **12**(2), 97–158. With discussion and a reply by the authors.
- Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66)*, Vol. I: Statistics, pages 221–233. Univ. California Press, Berkeley, Calif.
- Huber, P. J. (1981). *Robust statistics*. John Wiley & Sons, Inc., New York. Wiley Series in Probability and Mathematical Statistics.
- Hušková, M. (1996). Tests and estimators for the change point problem based on  $M$ -statistics. *Statist. Decisions*, **14**(2), 115–136.
- Hušková, M. and Neuhaus, G. (2004). Change point analysis for censored data. *J. Statist. Plann. Inference*, **126**(1), 207–223.

- Jandhyala, V., Fotopoulos, S., MacNeill, I., and Liu, P. (2013). Inference for single and multiple change-points in time series. *J. Time Series Anal.*, **34**(4), 423–446.
- Jandhyala, V. K. and Fotopoulos, S. B. (1999). Capturing the distributional behaviour of the maximum likelihood estimator of a changepoint. *Biometrika*, **86**(1), 129–140.
- Jandhyala, V. K. and Fotopoulos, S. B. (2001). Rate of convergence of the maximum likelihood estimate of a change-point. *Sankhyā Ser. A*, **63**(2), 277–285.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, **53**, 457–481.
- Kiefer, J. and Wolfowitz, J. (1956). Sequential tests of hypotheses about the mean occurrence time of a continuous parameter Poisson process. *Naval Res. Logist. Quart.*, **3**, 205–219 (1957).
- Kim, J., Cheon, S., and Jin, Z. (2020). Bayesian multiple change-points estimation for hazard with censored survival data from exponential distributions. *J. Korean Statist. Soc.*, **49**(1), 15–31.
- Korkas, K. K. and Fryzlewicz, P. (2017). Multiple change-point detection for non-stationary time series using wild binary segmentation. *Statist. Sinica*, **27**(1), 287–311.
- Korostel'ev, A. P. and Tsybakov, A. B. (1993). *Minimax theory of image reconstruction*, volume 82 of *Lecture Notes in Statistics*. Springer-Verlag, New York.
- Lavielle, M. (1999). Detection of multiple changes in a sequence of dependent variables. *Stochastic Process. Appl.*, **83**(1), 79–102.
- Lavielle, M. and Ludeña, C. (2000). The multiple change-points problem for the spectral distribution. *Bernoulli*, **6**(5), 845–869.
- LeCam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. *Univ. California Publ. Statist.*, **1**, 277–329.
- Lee, T.-S. (2010). Change-point problems: bibliography and review. *J. Stat. Theory Pract.*, **4**(4), 643–662.
- Lorden, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Statist.*, **42**, 1897–1908.
- Móricz, F. A., Serfling, R. J., and Stout, W. F. (1982). Moment and probability bounds with quasisuperadditive structure for the maximum partial sum. *Ann. Probab.*, **10**(4), 1032–1040.
- Niu, Y. S., Hao, N., and Zhang, H. (2016). Multiple change-point detection: a selective overview. *Statist. Sci.*, **31**(4), 611–623.
- Nkurunziza, S. and Fu, K. (2019). Improved inference in generalized mean-reverting processes with multiple change-points. *Electron. J. Stat.*, **13**(1), 1400–1442.
- Oakes, D. (1986). An approximate likelihood procedure for censored data. *Biometrics*, **42**(1), 177–182.
- Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, **41**, 100–115.
- Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, **42**, 523–527.
- Page, E. S. (1957). On problems in which a change in a parameter occurs at an unknown point. *Biometrika*, **44**(1/2), 248–252.
- Pakes, A. and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica*, **57**(5), 1027–1057.
- Pergamenchtchikov, S. and Tartakovsky, A. G. (2019). Asymptotically optimal pointwise and minimax change-point detection for general stochastic models with a composite post-change hypothesis. *J. Multivariate Anal.*, **174**, 104541, 20.
- Perlman, M. D. (1972). On the strong consistency of approximate maximum likelihood estimators. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971)*, Vol. I: *Theory of statistics*, pages 263–281.
- Pfanzagl, J. (1969). Consistent estimation of a location parameter in the presence of an incidental scale parameter. *Ann. Math. Statist.*, **40**, 1353–1357.

- Pons, O. (2018). *Estimations and tests in change-point models*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ.
- Qian, G., Wu, Y., and Xu, M. (2019). Multiple change-points detection by empirical Bayesian information criteria and Gibbs sampling induced stochastic search. *Appl. Math. Model.*, **72**, 202–216.
- Raimondo, M. (1998). Minimax estimation of sharp change points. *Ann. Statist.*, **26**(4), 1379–1397.
- Reid, N. (1981). Influence functions for censored data. *Ann. Statist.*, **9**(1), 78–92.
- Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. John Wiley & Sons, Inc., New York. Wiley Series in Probability and Mathematical Statistics.
- Shiryaev, A. N. (2016). On the minimax optimality of CUSUM statistics in change point problems for Brownian motion. *Teor. Veroyatn. Primen.*, **61**(4), 837–844.
- Stute, W. (1995). The statistical analysis of Kaplan-Meier integrals. In *Analysis of censored data (Pune, 1994/1995)*, volume 27 of *IMS Lecture Notes Monogr. Ser.*, pages 231–254. Inst. Math. Statist., Hayward, CA.
- Stute, W. (1996). Change-point problems under random censorship. *Statistics*, **27**(3-4), 255–266.
- Stute, W. and Wang, J.-L. (1993). The strong law under random censorship. *Ann. Statist.*, **21**(3), 1591–1607.
- Tan, L. and Zhang, Y. (2019). M-estimators of U-processes with a change-point due to a covariate threshold. *J. Bus. Econom. Statist.*, **37**(2), 248–259.
- Truong, C., Oudre, L., and Vayatis, N. (2020). Selective review of offline change point detection methods. *Signal Processing*, **167**, 107299.
- van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.
- van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics.
- Vostrikova, L. J. (1981). Discovery of “discord” in multidimensional random processes. *Dokl. Akad. Nauk SSSR*, **259**(2), 270–274.
- Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statistics*, **20**, 595–601.
- Wang, J. and Zheng, M. (2012). Wavelet detection of change points in hazard rate models with censored dependent data. *J. Nonparametr. Stat.*, **24**(3), 765–781.
- Wang, J.-L. (1995). M-estimators for censored data: strong consistency. *Scand. J. Statist.*, **22**(2), 197–205.
- Wang, J.-L. (1999). Asymptotic properties of M-estimators based on estimating equations and censored data. *Scand. J. Statist.*, **26**(2), 297–318.
- Wu, Y. (2005). *Inference for change-point and post-change means after a CUSUM test*, volume 180 of *Lecture Notes in Statistics*. Springer, New York.
- Yao, Y.-C. (1988). Estimating the number of change-points via Schwarz’ criterion. *Statist. Probab. Lett.*, **6**(3), 181–189.
- Yin, Y. Q. (1988). Detection of the number, locations and magnitudes of jumps. *Comm. Statist. Stochastic Models*, **4**(3), 445–455.
- Zou, C., Yin, G., Feng, L., and Wang, Z. (2014a). Nonparametric maximum likelihood approach to multiple change-point problems. *Ann. Statist.*, **42**(3), 970–1002.
- Zou, C., Yin, G., Feng, L., and Wang, Z. (2014b). Nonparametric maximum likelihood approach to multiple change-point problems. *Ann. Statist.*, **42**(3), 970–1002.
- Zou, C., Yin, G., Feng, L., and Wang, Z. (2014c). Nonparametric maximum likelihood approach to multiple change-point problems. *Ann. Statist.*, **42**(3), 970–1002.
- Zou, C., Wang, G., and Li, R. (2020). Consistent selection of the number of change-points via sample-splitting. *Ann. Statist.*, **48**(1), 413–439.