Stat Papers (2017) 58:877–909
DOI 10.1007/s00362-015-0730-y

REGULAR ARTICLE

Asymptotic inference for maximum likelihood estimators under the special exponential family with double-truncation

Takeshi Emura[1] · Ya-Hsuan Hu[1] · Yoshihiko Konno[2]

Received: 30 March 2015 / Revised: 20 October 2015 / Published online: 19 December 2015

Abstract Biased sampling affects the inference for population parameters of interest if the sampling mechanism is not appropriately handled. This paper considers doubly-truncated data arising in lifetime data analysis in which samples are subject to both left- and right-truncations. To correct for the sampling bias with doubly-truncated data, maximum likelihood estimator (MLE) has been proposed under a parametric family called the special exponential family (Efron and Petrosian, in J Am Stat Assoc 94:824–834, 1999). However, there is still a lack of justifying the fundamental properties for the MLE, including consistency and asymptotic normality. In this paper, we point out that the classical asymptotic theory for the independent and identically distributed data is not suitable for studying the MLE under double-truncation due to the non-identical truncation intervals. Alternatively, we formalize the asymptotic results under independent but not identically distributed data that suitably takes into account for the between-sample heterogeneity of truncation intervals. We establish the consistency and asymptotic normality of the MLE under a reasonably simple set of regularity conditions. Then, we give asymptotically valid techniques to estimate standard errors and to construct confidence intervals. Simulations are conducted to verify the suggested techniques, and childhood cancer data are used for illustration.

✉ Takeshi Emura
takeshiemura@gmail.com

Ya-Hsuan Hu
shumau6@hotmail.com

Yoshihiko Konno
konno@fc.jwu.ac.jp

1 Graduate Institute of Statistics, National Central University, Taoyuan, Taiwan

2 Department of Mathematical and Physical Sciences, Japan Women's University, Tokyo, Japan

Keywords Asymptotic normality · Central limit theorem · Consistency · Maximum likelihood estimation · Survival analysis · Truncated data

1 Introduction

Biased sampling commonly occurs in astronomy, epidemiology and population ageing studies in which samples are collected under certain constraints. Among the various sampling schemes, recent studies identify the so-called double truncation phenomenon; one can collect a sample only if the variable of interest falls within a certain interval. Since the variable of the observed sample is truncated by the lower and upper truncation limits, the sampled data is said to be "doubly truncated" (Efron and Petrosian 1999). For instance, Moreira and de Uña-Álvarez (2010) considered doubly truncated data arising from the epidemiological study of the childhood cancer in North Portugal. Here, the truncation limits correspond to the 5-year recruitment period during which the samples are ascertained. Other examples of double-truncation are found in similar settings, including Stovring and Wang (2007) and Moreira et al. (2014).

For $i = 1, 2, \ldots$, let y_i^* be a random sample from a density f, let u_i^* be a left-truncation limit and let v_i^* be a right-truncation limit. Suppose that a sample becomes available only if $u_i^* \leq y_i^* \leq v_i^*$ holds. Then, for a fixed sample size n, the available subsamples $y_1, y_2, \ldots, y_n$, subject to the constraints $u_i \leq y_i \leq v_i, i = 1, 2, \ldots, n$, are doubly-truncated data. Concretely, naïve statistics for the doubly-truncated data, such as sample mean and standard deviation, yield biased information about f due to the data loss in the upper- and lower-tails of f. Bias adjustment for the observable part is required to recover the population density f.

Although double truncation is one type of biased sampling, it accommodates both left- and right-truncation as special cases. Under left-truncation only, one obtains the sample when y_i^* is large enough compared to the left-truncation limit u_i^*. Left-truncation is also called 'delayed entry' when the lifetime y_i^* becomes available only if it exceeds the entry time u_i^*, as commonly encountered in biostatistics (Andersen and Keiding 2002; Klein and Moeschberger 2003) , educational research (Emura and Konno 2012), and industrial life testing (e.g., Sect. 2.4 of Lawless 2003). Under right-truncation only, one obtains the sample when y_i^* is smaller than the right-truncation limit v_i^*. Right-truncated data is especially relevant to the incubation time data of AIDS (e.g., Lagakos et al. 1988; Strzalkowska-Kominiak and Stute 2013) and the survival data for centenarians (e.g., Emura and Murotani 2015) in which the samples are ascertained before a fixed time limit. In most cases, statistical methodologies established for doubly-truncated data can be directly applicable to the left- or right-truncated data by setting $v_i = \infty$ or $u_i = -\infty$, respectively. Hence, methodological research on doubly-truncated data accommodates a broad class of data structures in a unified framework.

Recent years, nonparametric procedures for doubly truncated data have been actively studied in the literature. Important contributions include Shen (2010, 2011), Moreira and de Uña-Álvarez (2010, 2012), Emura and Konno (2012), Moreira and Van Keilegom (2013), Austin et al. (2014) and Emura et al. (2015).

Compared to the nonparametric analyses, research is much scarcer on parametric analyses under double-truncation. Efron and Petrosian (1999) proposed the maximum likelihood estimator (MLE) under a parametric family, called the special exponential family (SEF). Following them, Hu and Emura (2015) developed the randomized Newton–Raphson algorithms to obtain the MLE. However, there is still a lack of justifying the fundamental properties for the MLE, such as consistency and asymptotic normality. We aim to fill this gap of the previous two papers.

In this paper, we point out that the classical asymptotic theory for the independent and identically distributed (i.i.d.) data is not suitable for studying the MLE under double-truncation. Alternatively, we formalize the asymptotic results under the independent but not identically distributed (i.n.i.d.) data that take into account for the between-sample heterogeneity of truncation variables. Our mathematical tools include, among others, the Lindeberg–Feller multivariate central limit theorem (CLT) which settles the i.n.i.d. data. We derive a set of the regularity conditions such that the MLE is consistent and asymptotic normal under the SEF. In addition, we give the sufficient conditions that are reasonably interpreted and verified by the user. Then, we give asymptotically verified techniques to obtain standard errors and to construct confidence intervals. The developed techniques are examined by simulations and demonstrated by the childhood cancer data.

The rest of the paper is organized as follows. Section 2 reviews the model. Section 3 introduces the likelihood function. Section 4 gives the asymptotic analysis which is the main proposal of this paper. Section 5 conducts simulations. Section 6 analyzes real data. Section 7 concludes the paper.

2 Special exponential family (SEF)

We review the SEF considered by Efron and Petrosian (1999) for fitting doubly-truncated data. Let $1\{\cdot\}$ be the indicator function. We assume that a random variable Y^* follows the k-dimensional SEF, which is a continuous distribution with a density

$$f_{\boldsymbol{\eta}}(y) = \exp\{\boldsymbol{\eta}^{\mathrm{T}} \cdot \mathbf{t}(y) - \phi(\boldsymbol{\eta})\} 1\{y \in \mathsf{y}\},$$

where $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_k)^{\mathrm{T}} \in \Theta$, $\mathbf{t}(y) = (y, y^2, \ldots, y^k)^{\mathrm{T}}$, $\mathsf{y} \subset \Re$ is the support of Y^*, and $\Theta \subset \Re^{\mathrm{k}}$ is a parameter space. Here, $\phi(\boldsymbol{\eta})$ is a normalizing factor chosen to satisfy $\int_{\mathsf{y}} f_{\boldsymbol{\eta}}(y)dy = 1$. The SEF is a special case of a k-dimensional exponential family (p. 23 of Lehmann and Casella 1998).

The parameter space Θ is called "natural" if $\int_{\mathsf{y}} \exp\{\boldsymbol{\eta}^{\mathrm{T}} \cdot \boldsymbol{t}(y)\}dy < \infty$ for any $\boldsymbol{\eta} \in \Theta$. If Θ is natural, one can interchange the integration and differentiation as follows:

$$\frac{\partial}{\partial \boldsymbol{\eta}} \int\limits_{\mathsf{y}} g(y) \exp\{\boldsymbol{\eta}^{\mathrm{T}} \cdot \mathbf{t}(y)\}dy = \int\limits_{\mathsf{y}} \boldsymbol{t}(y) g(y) \exp\{\boldsymbol{\eta}^{\mathrm{T}} \cdot \mathbf{t}(y)\}dy, \quad \boldsymbol{\eta} \in \Theta, \tag{1}$$

for any function g (Theorem 2.7.1 of Lehmann and Romano 2005). The above identity is fundamental in the subsequent developments.

The cubic SEF (the SEF with $k = 3$) is particularly introduced by Efron and Petrosian (1999), which is obtained by setting $\mathbf{t}(y) = (y, y^2, y^3)^{\mathrm{T}}$, $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3)^{\mathrm{T}}$ and $\eta_4 = \cdots = \eta_k = 0$. The density of Y^* can be expressed as

$$f_{\boldsymbol{\eta}}(y) = \exp\{\eta_1 y + \eta_2 y^2 + \eta_3 y^3 - \phi(\boldsymbol{\eta})\} 1\{y \in \mathsf{y}\},$$

where $\phi(\boldsymbol{\eta}) = \log\{\int_{\mathsf{y}} \exp(\eta_1 y + \eta_2 y^2 + \eta_3 y^3) dy\}$. For the parameter space $\Theta \subset \Re^3$ to be natural, it is necessary that $\int_{\mathsf{y}} \exp(\eta_1 y + \eta_2 y^2 + \eta_3 y^3) dy < \infty$ for any $\boldsymbol{\eta} \in \Theta$.

Hu and Emura (2015) study the following natural parameter spaces:

First, if we consider the parameter space $\Theta^+ = \{(\eta_1, \eta_2, \eta_3) : \eta_1 \in \Re, \eta_2 \in \Re, \eta_3 > 0\}$, then we need to set the support of Y^* as $\mathsf{y}^+ = (-\infty, \tau_2]$, where $\tau_2 < \infty$ is the upper bound of Y^*. The corresponding survival function is

$$S_{\boldsymbol{\eta}}(y) = \int_y^{\tau_2} \exp\{\eta_1 t + \eta_2 t^2 + \eta_3 t^3 - \phi(\boldsymbol{\eta})\} dt, \quad y \le \tau_2.$$

Second, if we consider the parameter space $\Theta^- = \{(\eta_1, \eta_2, \eta_3)\colon \eta_1 \in \Re, \eta_2 \in \Re, \eta_3 < 0\}$, then we need to set the support of Y^* as $\mathsf{y}^- = [\tau_1, \infty)$, where $\tau_1 > -\infty$ is the lower bound of Y^*. The corresponding survival function is

$$S_{\boldsymbol{\eta}}(y) = \int_y^{\infty} \exp\{\eta_1 t + \eta_2 t^2 + \eta_3 t^3 - \phi(\boldsymbol{\eta})\} dt, \qquad y \ge \tau_1.$$

In the above two cases, we do not include the boundary $\eta_3 = 0$ in the parameter spaces for two reasons. First, we make the parameter space open to satisfy a regularity condition, which prevents some potential problem when the MLE reaches the boundary. Second, if $\eta_3 = 0$, then there is no need to set lower or upper bound. In fact, if $\eta_3 = 0$, then it is more natural to consider the normal distribution with $\mu = -\eta_1/2\eta_2$ and $\sigma^2 = -1/2\eta_2$ with a parameter space $\Theta^-_{\eta_3=0} = \{(\eta_1, \eta_2)\colon \eta_1 \in \Re, \eta_2 < 0\}$ (Castillo 1994; Hu and Emura 2015).

Remark 1 One can regard the cubic SEF as a skewed normal density, where η_3 represents a skewing parameter with respect to the symmetric kernel function $\exp(\eta_1 y + \eta_2 y^2)$. For $\eta_3 > 0$ the density is skewed to the right while, for $\eta_3 < 0$ the density is skewed to the left. If one chooses the parameter space Θ^+ for $\eta_3 > 0$, then the density is a skewed to the right and truncated by the upper boundary τ_2. The skewed and truncated density often provides a good fit to biomedical studies; see for instance, Mandrekar and Nandrekar (2003) for liver cirrhosis data and Robertson and Allison (2012) for the US life table data.

3 Likelihood function

Efron and Petrosian (1999) introduced the likelihood function, which corrects for the sampling bias with double-truncation. For $i = 1, 2, \ldots, n$, let $R_i = [u_i, v_i]$ be a truncation interval, where u_i is a left-truncation limit and v_i is a right-truncation limit. They consider the maximum likelihood estimator (MLE) under the SEF when the random samples $y_1, y_2, \ldots, y_n$ are subject to the constraints $y_i \in R_i$, $i = 1, 2, \ldots, n$. The truncated density of Y^*, subject to $Y^* \in R_i$, is

$$f_i(y|\boldsymbol{\eta}) \equiv \frac{f_{\boldsymbol{\eta}}(y)}{F_i(\boldsymbol{\eta})} 1\{y \in R_i\},$$

where $F_i(\boldsymbol{\eta}) = \int_{R_i} f_{\boldsymbol{\eta}}(y)dy$. Hence, the log-likelihood function for data $(y_1, y_2, \ldots, y_n)$ is

$$\ell_n(\boldsymbol{\eta}) = \log\left\{\prod_{i=1}^{n} f_i(y_i|\boldsymbol{\eta})\right\} = \sum_{i=1}^{n}\{\log f_{\boldsymbol{\eta}}(y_i) - \log F_i(\boldsymbol{\eta})\}.$$

Under the cubic SEF, Hu and Emura (2015) consider two cases: $\eta_3 > 0$ and $\eta_3 < 0$. We briefly review their results. First, consider the case $\eta_3 > 0$. As discussed before, the parameter space is $\Theta^+ = \{(\eta_1, \eta_2, \eta_3) : \eta_1 \in \Re, \eta_2 \in \Re, \eta_3 > 0\}$, and the upper bound of Y^* is τ_2. Define $\delta_i = 1\{v_i < \tau_2\}$. Then, the log-likelihood function is given by

$$\begin{aligned}\ell_n(\boldsymbol{\eta}) = &\sum_{i=1}^{n}(\eta_1 y_i + \eta_2 y_i^2 + \eta_3 y_i^3)\\ &-\sum_{i=1}^{n}\delta_i \log\left\{\int_{u_i}^{v_i}\exp(\eta_1 y + \eta_2 y^2 + \eta_3 y^3)dy\right\}\\ &-\sum_{i=1}^{n}(1-\delta_i)\log\left\{\int_{u_i}^{\tau_2}\exp(\eta_1 y + \eta_2 y^2 + \eta_3 y^3)dy\right\}.\end{aligned} \tag{2}$$

Next, in the case $\eta_3 < 0$, the parameter space is $\Theta^- = \{(\eta_1, \eta_2, \eta_3) : \eta_1 \in \Re, \eta_2 \in \Re, \eta_3 < 0\}$, and the lower bound of Y^* is τ_1. Define $\delta_i = 1\{u_i \geq \tau_1\}$. Then, the log-likelihood function is

$$\begin{aligned}\ell_n(\boldsymbol{\eta}) = &\sum_{i=1}^{n}(\eta_1 y_i + \eta_2 y_i^2 + \eta_3 y_i^3)\\ &-\sum_{i=1}^{n}\delta_i \log\left\{\int_{u_i}^{v_i}\exp(\eta_1 y + \eta_2 y^2 + \eta_3 y^3)dy\right\}\\ &-\sum_{i=1}^{n}(1-\delta_i)\log\left\{\int_{\tau_1}^{v_i}\exp(\eta_1 y + \eta_2 y^2 + \eta_3 y^3)dy\right\}.\end{aligned} \tag{3}$$

In their real data analyses, both Efron and Petrosian (1999) and Hu and Emura (2015) show that the cubic SEF (the SEF with $k = 3$) gives the best fit among a pool of models, including the SEF with $k = 1$ and $k = 2$. This implies the great practical value of the cubic SEF for real data analysis (see also Remark I). However, one potential concern for the cubic SEF is the unstability of the MLE due to the rich parameter space. In fact, the mean square error for estimating η_1 and η_2 is remarkably larger under the cubic SEF than the mean square error for estimating η_1 and η_2 under the SEF with $k = 2$ and given $\eta_3 = 0$ [compare Tables 3, 5 and 6 of Hu and Emura (2015)]. This motivates us to study the formal justification of the consistency as well as the convergence rate of the MLE under the k-dimensional SEF, especially for $k \geq 3$.

4 Asymptotic inference

This section develops the asymptotic theory for the MLE and then gives asymptotically valid standard error and confidence interval under the k-dimensional SEF. If we regard the samples $y_1, y_2, \ldots, y_n$ as random variables, we write them as $Y_1, Y_2, \ldots, Y_n$, where Y_i follows the truncated density $f_i(y|\boldsymbol{\eta})$ for $i = 1, 2, \ldots, n$.

4.1 Asymptotic theory

This subsection develops asymptotic properties of the MLE under the k-dimensional SEF

$$f_{\boldsymbol{\eta}}(y) = \exp\{\boldsymbol{\eta}^{\mathrm{T}} \cdot \mathbf{t}(y) - \phi(\boldsymbol{\eta})\} \cdot 1\{y \in \mathsf{y}\}, \quad \boldsymbol{\eta} \in \Theta,$$

where $\mathsf{y} \subset \Re$ is the support of Y^*, $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_k)^{\mathrm{T}}$, $\mathbf{t}(y) = (y, y^2, \ldots, y^k)^{\mathrm{T}}$ and $\phi(\boldsymbol{\eta})$ is chosen to make $\int_{\mathsf{y}} f_{\boldsymbol{\eta}}(y)dy = 1$, that is, $\phi(\boldsymbol{\eta}) = \log[\int_{\mathsf{y}} \exp\{\boldsymbol{\eta}^{\mathrm{T}} \cdot \mathbf{t}(y)\}dy]$. In Sect. 2, we have mentioned that the choice of (Θ, y) is important for the density to be well-behaved. To keep the generality of our theory, we do not explicitly specify the form of (Θ, y). Instead, we will impose the following general assumption:

Assumption (A) The parameter space Θ is open and contains the true parameter point $\boldsymbol{\eta}^0 = (\eta_1^0, \eta_2^0, \ldots, \eta_k^0)^{\mathrm{T}}$. In addition, the parameter space Θ is natural, i.e.,

$$\int_{\mathsf{y}} \exp\{\boldsymbol{\eta}^{\mathrm{T}} \cdot \mathbf{t}(y)\}dy < \infty \text{ for any } \boldsymbol{\eta} \in \Theta.$$

The three natural cases (Θ^-, y^-), (Θ^+, y^+), and $(\Theta^-_{\eta_3=0}, \Re)$ satisfy Assumption A. However, the parameter space $\Theta_{\eta_3=0} = \{(\eta_1, \eta_2) : \eta_1 \in \Re, \eta_2 \in \Re\}$ becomes natural only when the support y is bounded from below and above.

Given samples $y_1, y_2, \ldots, y_n$, the log-likelihood is given by

$$\ell_n(\boldsymbol{\eta}) = \sum_{i=1}^{n} \boldsymbol{\eta}^{\mathrm{T}} \cdot \mathbf{t}(y_i) - \sum_{i=1}^{n} \log\left[\int_{R_i \cap \mathsf{y}} \exp\{\boldsymbol{\eta}^{\mathrm{T}} \cdot \mathbf{t}(y)\}dy\right].$$

Define $\hat{\boldsymbol{\eta}}_n = (\hat{\eta}_{1n}, \hat{\eta}_{2n}, \ldots, \hat{\eta}_{kn})^{\mathrm{T}}$ to be a solution to the score equations

$$\frac{\partial}{\partial \eta_j} \ell_n(\boldsymbol{\eta}) = 0, \quad j = 1, 2, \ldots, k, \tag{4}$$

where $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_k)^{\mathrm{T}} \in \Theta$.

Lemma 1 *Under Assumption (A), if the solution* $\hat{\boldsymbol{\eta}}_n = (\hat{\eta}_{1n}, \hat{\eta}_{2n}, \ldots, \hat{\eta}_{kn})^{\mathrm{T}}$ *to Eq. (4) exists, then it is the MLE, that is,* $\ell_n(\hat{\boldsymbol{\eta}}_n) \geq \ell_n(\boldsymbol{\eta})$ *for any* $\boldsymbol{\eta} \in \Theta$.

The proof of Lemma 1 is given in "Proof of Lemma 1" in Appendix 1 that utilizes the concavity of $\ell_n(\boldsymbol{\eta})$. Although Lemma 1 does not assure the existence of the solution, it will be established under more assumptions.

Define the Fisher information of the ith sample as

$$I_{i,js}(\boldsymbol{\eta}) = E_{\boldsymbol{\eta}}\left\{\frac{\partial}{\partial \eta_j}\log f_i(Y_i|\boldsymbol{\eta}) \cdot \frac{\partial}{\partial \eta_s}\log f_i(Y_i|\boldsymbol{\eta})\right\}, \quad i = 1, 2, \ldots, n, \quad j, s = 1, 2, \ldots, k.$$

Under Assumption (A), it follows that

$$E_{\boldsymbol{\eta}}\left\{\frac{\partial}{\partial \eta_j}\log f_i(Y_i|\boldsymbol{\eta})\right\} = 0, \quad j = 1, 2, \ldots, k, \quad i = 1, 2, \ldots, n.$$

They are the consequence of Eq. (1) with $g(y) = 1\{y \in R_i\}$. Similarly,

$$\begin{aligned} I_{i,js}(\boldsymbol{\eta}) &= E_{\boldsymbol{\eta}}\left\{\frac{\partial}{\partial \eta_j}\log f_i(Y_i|\boldsymbol{\eta}) \cdot \frac{\partial}{\partial \eta_s}\log f_i(Y_i|\boldsymbol{\eta})\right\} \\ &= E_{\boldsymbol{\eta}}\left\{-\frac{\partial^2}{\partial \eta_j \partial \eta_s}\log f_i(Y_i|\boldsymbol{\eta})\right\}, \quad j, s = 1, 2, \ldots, k, \quad i = 1, 2, \ldots, n. \end{aligned}$$

All the expectations are taken with respect to the distribution of Y_i following the truncated density $f_i(y|\boldsymbol{\eta})$ for $i = 1, 2, \ldots, n$.

It is well-known that MLEs with independent and identically distributed (i.i.d.) models have consistency and asymptotic normality under some regularity conditions. However, the doubly-truncated random samples $y_1, y_2, \ldots, y_n$ are independent but not identically distributed (i.n.i.d.) due to the heterogeneity of intervals $[u_i, v_i]$, $i = 1, 2, \ldots, n$. Hence, we need specific justifications for the asymptotic analysis under the i.n.i.d. models.

Asymptotic theory under i.n.i.d. data has been applied to a linear regression model with fixed regressors (p. 21 of Van der Vaart 1998; p. 104 of Lehmann and Romano 2005), lifetime model with fixed censored points (Hoadley 1971), and many others. Nowadays, asymptotic analyses under i.n.i.d. data is less discussed in the literature as they are simply referred to either the Lindeberg–Feller CLT or the Liapounov CLT governing the case of i.n.i.d. data.

The classical theorems of Bradley and Gart (1962), Hoadley (1971) and Philippou and Roussas (1975) cover the asymptotic properties of the MLE under some i.n.i.d. models. In spite of the significant contribution of their papers, it prevents us to directly apply their theorems to the present setting. Firstly, the consistency and asymptotic normality proofs of Bradley and Gart (1962) are largely omitted, which make it difficult to follow how their regularity conditions are utilized in the proofs. This problem is partly due to their paper's dependence on the classical literature in 1940s during which the probability theory was not established. In particular, the Lindeberg–Feller CLT, which should be a standard tool, was not referred in their papers. Second, the regularity conditions given by Hoadley (1971) are fairly technical and less intuitive, though they are weaker than those given by Bradley and Gart (1962). Third, the asymptotic normality theorems of Philippou and Roussas (1975) are stated under the assumption that the MLE is consistent. Rather, we wish to establish both the consistency and asymptotic normality under more intuitive regularity conditions and under the established probability theory. Finally, modern empirical processes techniques for studying the asymptotic properties of the MLE are almost exclusively focuses on the i.i.d. settings (Van der Vaart 1998). In our conclusion, the asymptotic behaviors of the MLE under double-truncation are not straightforwardly derived by referring to existing theorems.

We develop the asymptotic theory by following a general strategy similar to Bradley and Gart (1962) with suitable modifications of their regularity conditions to the present setting. Then, our proofs modify the proofs of consistency and asymptotic normality described by the book of Lehmann and Casella (1998) under i.i.d. cases. Although there are many textbooks describing the asymptotic theory of the MLE, their proofs are not always rigorous or are often limited to the simplistic setting of one-dimensional parameter. We believe that the mathematical treatment of Lehmann and Casella (1998) is a right way to handle our multi-parameter (k-parameter) setting and to clarify how the regularity conditions are utilized in the proof.

Our main mathematical tools for establishing the consistency and asymptotic normality of the MLE are the weak law of large number (WLLN) and the Lindeberg–Feller CLT for i.n.i.d. random variables. Let "$\xrightarrow{P}$" denote "convergence in probability". The following lemma is available in p. 65 of Shao (2003):

Lemma 2 (The WLLN for i.n.i.d. random variables) *Let* $Y_1, Y_2, \ldots$ *be independent random variables with* $E[|Y_i|] < \infty$ *for* $i = 1, 2, \ldots$. *If there is a constant* $p \in [1, 2]$ *such that*

$$\lim_{n\to\infty} \frac{1}{n^p} \sum_{i=1}^{n} E[|Y_i|^p] = 0,$$

then

$$\frac{1}{n} \sum_{i=1}^{n} (Y_i - E[Y_i]) \xrightarrow{P} 0.$$

Let "$\xrightarrow{d}$" denote "convergence in distribution". The so-called Lindeberg–Feller CLT is a version of the CLT for the sum of i.n.i.d. random variables. The following is the multivariate extension of the Lindeberg–Feller CLT available in Van der Vaart (1998).

Lemma 3 (The Lindeberg–Feller multivariate central limit theorem) *Let* $\mathbf{D}_{n,1}, \ldots, \mathbf{D}_{n,n}$ *be independent* k*-dimensional random vectors with finite second moments such that*

$$\sum_{i=1}^{n} E[||\mathbf{D}_{n,i} - E[\mathbf{D}_{n,i}]||^2 1\{||\mathbf{D}_{n,i} - E[\mathbf{D}_{n,i}]|| > \varepsilon\}] \to 0, \quad n \to \infty \tag{5}$$

for every $\varepsilon > 0$, *and*

$$\sum_{i=1}^{n} \mathbf{Cov}(\mathbf{D}_{n,i}) \to \Sigma, \quad n \to \infty.$$

Then, $\sum_{i=1}^{n} (\mathbf{D}_{n,i} - E\mathbf{D}_{n,i}) \xrightarrow{d} N_k(\mathbf{0}, \Sigma)$ *as* $n \to \infty$.

Equation (5) is known as the Lindeberg condition.

We impose the following conditions motivated by Bradley and Gart (1962):

Assumption (B) There exists a $k \times k$ positive definite matrix $I(\boldsymbol{\eta}) = \{I_{js}(\boldsymbol{\eta})\}_{j,s=1,2,\ldots,k}$ such that, as $n \to \infty$,

$$\sum_{i=1}^{n} I_{i,js}(\boldsymbol{\eta})/n \to I_{js}(\boldsymbol{\eta}), \quad j, s \in \{1, 2, \ldots, k\}, \quad \boldsymbol{\eta} \in \Theta.$$

Assumption (C) For $j, s, l \in \{1, 2, \ldots, k\}$, there is a measurable function $M_{jsl}(\cdot)$ such that

$$\left| \frac{\partial^3}{\partial \eta_j \partial \eta_s \partial \eta_l} \log f_i(y|\boldsymbol{\eta}) \right| \le M_{jsl}(y), \quad y \in \mathsf{y}, \quad \boldsymbol{\eta} \in \Theta, \quad i = 1, 2, \ldots, n$$

with $m_{i,jsl} \equiv E_{\boldsymbol{\eta}^0}\{M_{jsl}(Y_i)\} < \infty$ and $m^2_{i,jsl} \equiv E_{\boldsymbol{\eta}^0}\{M_{jsl}(Y_i)^2\} < \infty$. For some m_{jsl} and m^2_{jsl}, it holds that $\sum_{i=1}^{n} m_{i,jsl}/n \to m_{jsl}$ and $\sum_{i=1}^{n} m^2_{i,jsl}/n \to m^2_{jsl}$ as $n \to \infty$.

Assumption (D) For $j, s \in \{1, 2, \ldots, k\}$, there is a measurable function $W_{js}(\cdot)$ such that

$$\left| \frac{\partial^2}{\partial \eta_j \partial \eta_s} \log f_i(y|\boldsymbol{\eta}) \right| \le W_{js}(y), \quad y \in \mathsf{y}, \quad \boldsymbol{\eta} \in \Theta, \quad i = 1, 2, \ldots, n$$

with $w_{i,js} \equiv E_{\boldsymbol{\eta}^0}\{W_{js}(Y_i)\} < \infty$ and $w^2_{i,js} \equiv E_{\boldsymbol{\eta}^0}\{W_{js}(Y_i)^2\} < \infty$. For some w_{js} and w^2_{js}, it holds that $\sum_{i=1}^{n} w_{i,js}/n \to w_{js}$ and $\sum_{i=1}^{n} w^2_{i,js}/n \to w^2_{js}$ as $n \to \infty$.

Assumption (E) For $j \in \{1, 2, \ldots, k\}$, there is a measurable function $A_j(\cdot)$ such that

$$\left| \frac{\partial}{\partial \eta_j} \log f_i(y|\boldsymbol{\eta}) \right| \le A_j(y), \quad y \in \mathsf{y}, \quad \boldsymbol{\eta} \in \Theta, \quad i = 1, 2, \ldots, n$$

with $\sup_y A_j^2(y) < \infty$.

Assumption (B) is an essential requirement that the Fisher information matrix for large sample is stabilized. Here, the large sample Fisher information matrix is reasonably defined as the limit of the average Fisher information matrices for individual samples. Assumption (C) is required for the remainder terms of the Taylor expansion of $\ell_n(\boldsymbol{\eta})$ to be negligible. It plays a fundamental role in proving both consistency and asymptotic normality, analogous to i.i.d. cases. Assumption (D) is also required for the Taylor expansion to work. The bounding functions in Assumption (B)–(D) are necessary to check the condition of the WLLN in Lemma 2. Assumption (E) gives the bounds of score functions which are similar to those of Bradley and Gart (1962) under i.n.i.d. cases. This sort of assumption does not appear under i.i.d. data since it is claimed to be too strong under i.i.d. cases and without truncation (Hoadley 1971). However, such assumptions can be reasonably satisfied under double-truncation since the density is truncated from below and above. Assumption (E) is required to verify the Lindeberg condition in Lemma 3.

Theorem 1 *If Assumptions (A)–(E) hold, then*

(a) Existence and consistency: There exists a solution $\hat{\boldsymbol{\eta}}_n$ to Eq. (4) with probability tending to one, such that $\hat{\boldsymbol{\eta}}_n \xrightarrow{P} \boldsymbol{\eta}^0$ as $n \to \infty$.
(b) Asymptotic normality: $\sqrt{n}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}^0) \xrightarrow{d} N_k(\mathbf{0}, I(\boldsymbol{\eta}^0)^{-1})$ as $n \to \infty$.

The proof of Theorem 1 is given in Appendix 1.

4.2 Standard error and confidence interval

We use Theorem 1 (b) to obtain the standard error $SE(\hat{\eta}_{jn})$ and to construct the confidence interval for η_j. By Assumptions (A) and (B), when n is large, we have the following approximations:

$$
\begin{aligned}
I_{js}(\boldsymbol{\eta}^0) &\approx \frac{1}{n}\sum_{i=1}^{n} I_{i,js}(\boldsymbol{\eta}^0) \\
&= \frac{1}{n}\sum_{i=1}^{n} E_{\boldsymbol{\eta}^0}\left\{\frac{\partial}{\partial \eta_j}\log f_i(Y_i|\boldsymbol{\eta}^0)\cdot\frac{\partial}{\partial \eta_s}\log f_i(Y_i|\boldsymbol{\eta}^0)\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n} E_{\boldsymbol{\eta}^0}\left\{-\frac{\partial^2}{\partial \eta_j \partial \eta_s}\log f_i(Y_i|\boldsymbol{\eta}^0)\right\} \\
&\approx -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial \eta_j \partial \eta_s}\log f_i(Y_i|\boldsymbol{\eta})\bigg|_{\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}_n} \\
&= -\frac{1}{n}\frac{\partial^2}{\partial \eta_j \partial \eta_s}\ell_n(\boldsymbol{\eta})\bigg|_{\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}_n} \equiv \hat{I}_{js}(\hat{\boldsymbol{\eta}}_n).
\end{aligned}
$$

The last term constitutes the observed Fisher information matrix $\hat{I}(\hat{\boldsymbol{\eta}}_n) = \{\hat{I}_{js}(\hat{\boldsymbol{\eta}}_n)\}_{j,s=1,2,\ldots,k}$, which is obtained through the final step of the Newton–Raphson algorithm of Hu and Emura (2015). Hence,

$$
I(\boldsymbol{\eta}^0) \approx \hat{I}(\hat{\boldsymbol{\eta}}_n) = -\frac{1}{n}\frac{\partial^2}{\partial \boldsymbol{\eta}^2}\ell_n(\hat{\boldsymbol{\eta}}_n),
$$

and the standard error is

$$SE(\hat{\eta}_{jn}) = \sqrt{\frac{\{\hat{I}(\hat{\boldsymbol{\eta}}_n)^{-1}\}_{jj}}{n}} = \sqrt{\left\{-\frac{\partial^2}{\partial \boldsymbol{\eta}^2}\ell_n(\boldsymbol{\eta})\right\}^{-1}_{jj}\Bigg|_{\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}_n}}, \quad j \in \{1, 2, \ldots, k\},$$

where $\{\hat{I}(\hat{\boldsymbol{\eta}}_n)^{-1}\}_{jj}$ is the jth diagonal element in the inverse of the observed Fisher information matrix $\hat{I}(\hat{\boldsymbol{\eta}}_n)$. By the normal approximation of $\hat{\boldsymbol{\eta}}_n = (\hat{\eta}_{1n}, \hat{\eta}_{2n}, \ldots, \hat{\eta}_{kn})^{\mathrm{T}}$ due to Theorem 1 (b), we construct the $(1-\alpha)100\ \%$ confidence intervals for η_j:

$$[\hat{\eta}_{jn} - Z_{\alpha/2} \cdot SE(\hat{\eta}_{jn}),\ \hat{\eta}_{jn} + Z_{\alpha/2} \cdot SE(\hat{\eta}_{jn})], \quad j \in \{1, 2, \ldots, k\},$$

where Z_p is the pth upper quantile for $N(0, 1)$.

By the delta method, the standard error of the density estimator is

$$SE\{f_{\hat{\boldsymbol{\eta}}_n}(y)\} = \sqrt{\left\{\frac{\partial}{\partial \boldsymbol{\eta}} f_{\boldsymbol{\eta}}(y)\right\}^{\mathrm{T}} \cdot \left\{-\frac{\partial^2}{\partial \boldsymbol{\eta}^2}\ell_n(\boldsymbol{\eta})\right\}^{-1} \cdot \frac{\partial}{\partial \boldsymbol{\eta}} f_{\boldsymbol{\eta}}(y)\Bigg|_{\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}_n}},$$

where

$$\frac{\partial}{\partial \boldsymbol{\eta}} f_{\boldsymbol{\eta}}(y) = \begin{bmatrix} y - e^1(\boldsymbol{\eta})/e^0(\boldsymbol{\eta}) \\ \vdots \\ y^k - e^k(\boldsymbol{\eta})/e^0(\boldsymbol{\eta}) \end{bmatrix} \cdot f_{\boldsymbol{\eta}}(y),$$

and where

$$e^j(\boldsymbol{\eta}) = \int_{\mathsf{y}} y^j \exp\{\boldsymbol{\eta}^{\mathrm{T}} \cdot \mathbf{t}(y)\} dy, \quad j \in \{0, 1, 2, \ldots, k\}.$$

The $(1-\alpha)100\ \%$ confidence interval for the density $f_{\boldsymbol{\eta}}(y)$ is,

$$[f_{\hat{\boldsymbol{\eta}}_n}(y) - Z_{\alpha/2} \cdot SE\{f_{\hat{\boldsymbol{\eta}}_n}(y)\}, \quad f_{\hat{\boldsymbol{\eta}}_n}(y) + Z_{\alpha/2} \cdot SE\{f_{\hat{\boldsymbol{\eta}}_n}(y)\}].$$

Similarly, the standard error of estimating the survival function is

$$SE\{S_{\hat{\boldsymbol{\eta}}_n}(y)\} = \sqrt{\left\{\frac{\partial}{\partial \boldsymbol{\eta}} S_{\boldsymbol{\eta}}(y)\right\}^{\mathrm{T}} \cdot \left\{-\frac{\partial^2}{\partial \boldsymbol{\eta}^2}\ell_n(\boldsymbol{\eta})\right\}^{-1} \cdot \frac{\partial}{\partial \boldsymbol{\eta}} S_{\boldsymbol{\eta}}(y)\Bigg|_{\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}_n}},$$

where

$$\frac{\partial}{\partial \boldsymbol{\eta}} S_{\boldsymbol{\eta}}(y) = \int_{y \in \mathsf{y},\ t > y} \begin{bmatrix} t - e^1(\boldsymbol{\eta})/e^0(\boldsymbol{\eta}) \\ \vdots \\ t^k - e^k(\boldsymbol{\eta})/e^0(\boldsymbol{\eta}) \end{bmatrix} \cdot f_{\boldsymbol{\eta}}(t) dt.$$

The $(1-\alpha)100\,\%$ confidence interval for the survival function $S_{\eta}(y)$ is

$$[S_{\hat{\eta}_n}(y) - Z_{\alpha/2}\cdot SE\{S_{\hat{\eta}_n}(y)\},\quad S_{\hat{\eta}_n}(y) + Z_{\alpha/2}\cdot SE\{S_{\hat{\eta}_n}(y)\}].$$

4.3 Sufficient conditions

We show that there exist simple sufficient conditions to verify the assumptions in Theorem 1.

Lemma 4 *Assumptions (C)–(E) hold under the following two conditions:*

Assumption (F) The parameter space Θ is bounded.

Assumption (G) The lower support of left-truncation $u_{\inf} \equiv \inf_i(u_i^*)$ and the upper support of right-truncation $v_{\sup} \equiv \sup_i(v_i^*)$ are finite real numbers. In addition, there exist constants $u_0 < v_0$ such that

$$[u_0, v_0] \subset [u_i, v_i] \subset [u_{\inf}, v_{\sup}] \subset \mathsf{y}, \quad i = 1, 2, \ldots.$$

The proof of Lemma 4 is given in "Proofs of Lemma 4" in Appendix 1.

Assumption (F) can be always satisfied since one can choose an arbitrary large bound for each parameter. This sort of technical assumption is often employed in mathematical statistics (e.g., see Example 4.19 of Shao 2003), but it poses no practical restriction.

Assumption (G) is interpreted as the "stability" of truncation intervals among samples. The interval length cannot be too short (should be longer than $v_0 - u_0$) and cannot be too long (should be bounded by $v_{\sup} - u_{\inf}$). In addition, all the intervals must contain the common region $[u_0, v_0]$. Intuitively, if the interval length is too short, then the sample can have extremely high impact on the MLE since the sample inclusion probability $F_i(\boldsymbol{\eta}) = \int_{u_i}^{v_i} f_{\boldsymbol{\eta}}(y)dy$ in the likelihood is too small. In particular, Assumption (G) excludes the extreme case of $u_i = v_i$. Hence, Assumption (G) is a requirement for bounding the effect of individual's likelihood contribution. This sort of stabilizing assumption is common in the context of the "inverse censoring/truncation probability weighting" (Seaman and White 2011).

Example 1 (*Fixed double-truncation*) Truncation intervals are fixed for all subjects, say $u_i^* = u_i = u_0$ and $v_i^* = v_i = v_0$ for $i = 1, 2, \ldots$. Assumption (G) holds when $[u_0, v_0] \subset \mathsf{y}$. Statistical inference under the fixed double-truncation is extensively studied in the classical literature and well summarized in the book of Cohen (1991). See also some recent work of Sankaran and Sunoj (2004)

Example 2 (*Fixed-length double-truncation*) The length of truncation intervals is fixed for all subjects, say $[u_i^*, v_i^*] = [u_i^*, u_i^* + d_0]$, $d_0 > 0$, for $i = 1, 2, \ldots$. If $u_i^* = u_0$ for $i = 1, 2, \ldots$, then this reduces to the fixed double truncation (Example 1). We relax the fixed double truncation by allowing u_i^* to vary on $[a, b]$ where $a < b$ are known. If $[a, b + d_0] \subset \mathsf{y}$ and $b < a + d_0$, then Assumption (G) holds with $u_{\inf} = a$, $u_0 = b$, $v_0 = a + d_0$, and $v_{\sup} = b + d_0$. The condition $b < a + d_0$ guarantees the sufficient follow-up length ($d_0 > b - a$). If d_0 is too short, then the intervals $[u_i, v_i]$, $i = 1, 2, \ldots$, cannot share any common region.

Remark II Under Assumption (G), all the observations y_i, $i = 1, 2, \ldots$, fall in the region $[u_{\text{inf}}, v_{\text{sup}}] \subset \mathsf{y}$. Typical nonparametric estimators encounter the unidentifiability about the population density f since no information is available in the region $\mathsf{y} \cap [u_{\text{inf}}, v_{\text{sup}}]^c$. For instance, if $\mathsf{y}^+ = (-\infty, \tau_2]$ is the support of f and $v_{\text{sup}} < \tau_2$, then, the density f is unidentifiable on the region $\mathsf{y} \cap [u_{\text{inf}}, v_{\text{sup}}]^c = (-\infty, u_{\text{inf}}) \cup (v_{\text{sup}}, \tau_2]$. In this instance, Theorem 2 still verifies the consistency of the MLE under Assumption (G). The key is the strong assumptions that the parametric form of the distribution is known on the entire support $\mathsf{y}^+ = (-\infty, \tau_2]$, and the value of τ_2 is known. If no information about τ_2 is available, Hu and Emura (2015) suggest using the value $\tau_2 = \max_i y_i$. In some special applications, a reasonable value of τ_2 is available, for instance $\tau_2 = 120$ (years of age) for survival analysis of centenarians (those who live beyond the age of 100 years) (see Emura and Murotani 2015).

5 Simulations

We conduct Monte Carlo simulations to examine the numerical validity of the asymptotic results. For each repetition, we generate random triplet (u_i, y_i, v_i), subject to $u_i \le y_i \le v_i$, for $i = 1, 2, \ldots, n$. The data come from the independent random triplet (U^*, Y^*, V^*) subject to the inclusion criterion $U^* \le Y^* \le V^*$. Here, Y^* follows the cubic SEF ($k = 3$) with $\eta_1 = 5$, $\eta_2 = -0.5$, $\eta_3 = 0.005$ or $\eta_3 = -0.005$, and the distribution of (U^*, V^*) is chosen such that $P(U^* \le Y^* \le V^*) = 0.5$ or 0.25. The details of the data generation schemes are given in Appendix 2.

To obtain the MLE $\hat{\boldsymbol{\eta}}_n = (\hat{\eta}_{1n}, \hat{\eta}_{2n}, \hat{\eta}_{3n})^{\mathrm{T}}$, we maximize the log-likelihood in Eqs. (2) or (3) by performing the randomized Newton–Raphson algorithm starting with the initial values $(\eta_1^{(0)}, \eta_2^{(0)}, \eta_3^{(0)}) = (\bar{y}/s^2, -1/2s^2, 0)$, where $\bar{y} = \sum_i y_i/n$ and $s^2 = \sum_i (y_i - \bar{y})^2/(n-1)$. We randomize the initial values for the case of unconvergence. In this way, the algorithm always converges. The details of the algorithm follow Hu and Emura (2015). Based on 1000 repetitions, we evaluate the performance of $\hat{\boldsymbol{\eta}}_n = (\hat{\eta}_{1n}, \hat{\eta}_{2n}, \hat{\eta}_{3n})^{\mathrm{T}}$, $f_{\hat{\boldsymbol{\eta}}}(t)$ and $S_{\hat{\boldsymbol{\eta}}}(t)$, where t is chosen as $S_{\boldsymbol{\eta}}(t) = 0.5$. We also evaluate the standard error and confidence interval for the estimators.

Table 1 displays the results for estimating η_1, η_2 and η_3 under $P(U^* \le Y^* \le V^*) = 0.5$. The estimators are roughly unbiased for the true values, and their standard deviation (SD) vanishes as the sample size increase from $n = 100$ to 300. The standard errors give very good approximations to the SDs of the estimators for the cases of $\eta_3 = 0.005$ while they slightly overestimate the SDs for the cases of $\eta_3 = -0.005$. The overestimation is due to a few standard errors occurring when the MLE is maximized near the boundary $\eta_3 = 0$. The empirical coverage rates of the confidence intervals are reasonably close to the nominal 95 %.

Table 2 displays the results when the truncation effect is heavier (i.e., $P(U^* \le Y^* \le V^*) = 0.25$). Compared to Table 1 (the case of $P(U^* \le Y^* \le V^*) = 0.5$), the SDs of the estimators inflate. The standard errors somewhat overestimate the SDs in all cases. Consequently, the empirical coverage rates of the confidence intervals are slightly larger than the nominal 95 %. The overestimation and over-coverage become negligible when the sample size increases up to $n = 300$.

Table 1 Simulation results for estimating parameters η_1, η_2 and η_3 under the cubic SEF based on 1000 repetitions (under the inclusion probability $P(U^* \le Y^* \le V^*) \approx 0.50$)

(η_1, η_2, η_3)	n	$E(\hat{\eta}_1)$	$SD(\hat{\eta}_1)$	$E[SE(\hat{\eta}_1)]$	95 % Cov
(5, −0.5, 0.005)	100	5.856	7.282	7.436	0.936
	200	5.484	5.146	5.165	0.944
	300	5.378	4.125	4.207	0.951
(5, −0.5, −0.005)	100	4.928	7.550	8.160	0.957
	200	4.981	5.346	5.675	0.947
	300	4.955	4.292	4.608	0.958
(η_1, η_2, η_3)	n	$E(\hat{\eta}_2)$	$SD(\hat{\eta}_2)$	$E[SE(\hat{\eta}_2)]$	95 % Cov
(5, −0.5, 0.005)	100	−0.622	1.397	1.437	0.940
	200	−0.573	0.995	0.998	0.945
	300	−0.561	0.797	0.813	0.946
(5, −0.5, −0.005)	100	−0.422	1.582	1.717	0.954
	200	−0.465	1.121	1.194	0.949
	300	−0.472	0.901	0.970	0.956
(η_1, η_2, η_3)	n	$E(\hat{\eta}_3)$	$SD(\hat{\eta}_3)$	$E[SE(\hat{\eta}_3)]$	95 % Cov
(5, −0.5, 0.005)	100	0.0101	0.089	0.091	0.944
	200	0.0084	0.063	0.063	0.950
	300	0.0081	0.051	0.052	0.949
(5, −0.5, −0.005)	100	−0.0145	0.109	0.119	0.952
	200	−0.0094	0.077	0.083	0.953
	300	−0.0081	0.062	0.067	0.956

If $\eta_1 = 5$, $\eta_2 = -0.5$, and $\eta_3 = 0.005$, then the upper support is $\tau_2 = 8$
If $\eta_1 = 5$, $\eta_2 = -0.5$, and $\eta_3 = -0.005$, then the lower support is $\tau_1 = 2$
95 % Cov The empirical coverage rate of the 95 % confidence intervals

Table 3 displays the results for estimating survival function $S_{\boldsymbol{\eta}}(t)$ and density function $f_{\boldsymbol{\eta}}(t)$ under $P(U^* \le Y^* \le V^*) = 0.5$. The estimators are virtually unbiased in all the cases. The SDs decrease as the sample size n increases from 150 to 300, and they are precisely estimated by the standard errors. Also, the empirical coverage rates are in good agreement with the nominal 95 %. These results are similar even when the truncation effect is heavier (i.e., $P(U^* \le Y^* \le V^*) = 0.25$), except some minor over-coverage of the confidence intervals (Table 4).

6 Data analysis

6.1 The childhood cancer data (Moreira and de Uña-Álvarez 2010)

The childhood cancer dataset (Moreira and de Uña-Álvarez 2010) is analyzed for illustration. The data contains the ages at onset of cancer at a young age (below 15 years) within a recruitment period of 5 years (between January 1, 1999 and December

Table 2 Simulation results for estimating parameters η_1, η_2 and η_3 under the cubic SEF based on 1000 repetitions (under the inclusion probability $P(U^* \le Y^* \le V^*) \approx 0.25$)

(η_1, η_2, η_3)	n	$E(\hat{\eta}_1)$	$SD(\hat{\eta}_1)$	$E[SE(\hat{\eta}_1)]$	95 % Cov
(5, −0.5, 0.005)	100	6.241	8.963	9.783	0.977
	200	5.189	6.494	6.722	0.949
	300	5.334	5.133	5.502	0.954
(5, −0.5, −0.005)	100	5.306	9.266	10.431	0.964
	200	5.228	6.868	7.338	0.964
	300	5.175	5.776	5.958	0.949
(η_1, η_2, η_3)	n	$E(\hat{\eta}_2)$	$SD(\hat{\eta}_2)$	$E[SE(\hat{\eta}_2)]$	95 % Cov
(5, −0.5, 0.005)	100	−0.702	1.725	1.894	0.977
	200	−0.517	1.256	1.304	0.950
	300	−0.549	0.991	1.067	0.950
(5, −0.5, −0.005)	100	−0.504	1.935	2.178	0.962
	200	−0.508	1.429	1.533	0.959
	300	−0.508	1.206	1.245	0.953
(η_1, η_2, η_3)	n	$E(\hat{\eta}_3)$	$SD(\hat{\eta}_3)$	$E[SE(\hat{\eta}_3)]$	95 % Cov
(5, −0.5, 0.005)	100	0.0157	0.110	0.121	0.971
	200	0.0049	0.080	0.083	0.949
	300	0.0072	0.063	0.068	0.954
(5, −0.5, −0.005)	100	−0.0087	0.134	0.150	0.966
	200	−0.0069	0.098	0.105	0.954
	300	−0.0062	0.083	0.086	0.956

If $\eta_1 = 5$, $\eta_2 = -0.5$, and $\eta_3 = 0.005$, then the upper support is $\tau_2 = 8$
If $\eta_1 = 5$, $\eta_2 = -0.5$, and $\eta_3 = -0.005$, then the lower support is $\tau_1 = 2$
95 % Cov The empirical coverage rate of the 95 % confidence intervals

31, 2003). The onset ages are considered as the ages at which children are diagnosed as cancer within the period. However, they do not have any information on children who developed cancer outside the period. Since the time constraint is purely by the design problem, observed data are biased sampling from the target population in which the constraint is completely ignored. The observed samples consist of 406 children with $\{(u_i, y_i, v_i): i = 1, \ldots, 406\}$ subject to $u_i \le y_i \le v_i$, where y_i is the age at diagnosis, u_i is the age at the recruitment start (January 1, 1999), and $v_i = u_i + 1825$ is the age at the recruitment end (December 31, 2003). We make inference for the survival function $S(t)$ of the age at diagnosis.

The data have been analyzed previously. Moreira and de Uña-Álvarez (2010) and Emura et al. (2015) nonparametrically estimated the distribution function and survival function, respectively, based on the NPMLE. Hu and Emura (2015) performed model selection among the pool of parametric models, and concluded that the cubic SEF gives the best fit. In addition, Hu and Emura (2015) demonstrated that the two survival

Table 3 Simulation results for estimating survival function $S_{\eta}(y)$ and density function $f_{\eta}(y)$ under the cubic SEF based on 1000 repetitions (under the inclusion probability $P(U^* \le Y^* \le V^*) \approx 0.50$)

(η_1, η_2, η_3)	n	$E\{S_{\hat{\eta}}(t)\}$	$SD\{S_{\hat{\eta}}(t)\}$	$E[SE\{S_{\hat{\eta}}(t)\}]$	95 % Cov
(5, −0.5, 0.005)	100	0.499	0.071	0.070	0.944
$S_{\eta}(y) = 0.5$	200	0.500	0.049	0.049	0.939
	300	0.502	0.038	0.039	0.947
(5, −0.5, −0.005)	100	0.504	0.062	0.065	0.941
$S_{\eta}(y) = 0.5$	200	0.503	0.044	0.045	0.957
	300	0.502	0.036	0.037	0.949
(η_1, η_2, η_3)	n	$E\{f_{\hat{\eta}}(t)\}$	$SD\{f_{\hat{\eta}}(t)\}$	$E[SE\{f_{\hat{\eta}}(t)\}]$	95 % Cov
(5, −0.5, 0.005)	100	0.367	0.054	0.057	0.969
$f_{\eta}(y) = 0.369$	200	0.367	0.036	0.039	0.967
	300	0.367	0.030	0.031	0.961
(5, −0.5, −0.005)	100	0.430	0.057	0.060	0.961
$f_{\eta}(y) = 0.427$	200	0.428	0.040	0.041	0.947
	300	0.427	0.032	0.033	0.958

If $\eta_1 = 5$, $\eta_2 = -0.5$, and $\eta_3 = 0.005$, then the upper support is $\tau_2 = 8$
If $\eta_1 = 5$, $\eta_2 = -0.5$, and $\eta_3 = -0.005$, then the lower support is $\tau_1 = 2$
95 % Cov The empirical coverage rate of the 95 % confidence intervals

Table 4 Simulation results for estimating survival function $S_{\eta}(y)$ and density function $f_{\eta}(y)$ under the cubic SEF based on 1000 repetitions (under the inclusion probability $P(U^* \le Y^* \le V^*) \approx 0.25$)

(η_1, η_2, η_3)	n	$E\{S_{\hat{\eta}}(t)\}$	$SD\{S_{\hat{\eta}}(t)\}$	$E[SE\{S_{\hat{\eta}}(t)\}]$	95 % Cov
(5, −0.5, 0.005)	100	0.512	0.091	0.097	0.977
$S_{\eta}(y) = 0.5$	200	0.501	0.063	0.067	0.959
	300	0.501	0.051	0.053	0.958
(5, −0.5, −0.005)	100	0.499	0.081	0.087	0.960
$S_{\eta}(y) = 0.5$	200	0.504	0.055	0.060	0.960
	300	0.502	0.045	0.048	0.957
(η_1, η_2, η_3)	n	$E\{f_{\hat{\eta}}(t)\}$	$SD\{f_{\hat{\eta}}(t)\}$	$E[SE\{f_{\hat{\eta}}(t)\}]$	95 % Cov
(5, −0.5, 0.005)	100	0.357	0.076	0.076	0.954
$f_{\eta}(y) = 0.369$	200	0.363	0.047	0.052	0.969
	300	0.365	0.040	0.042	0.964
(5, −0.5, −0.005)	100	0.421	0.072	0.079	0.976
$f_{\eta}(y) = 0.427$	200	0.425	0.050	0.054	0.968
	300	0.426	0.040	0.043	0.963

If $\eta_1 = 5$, $\eta_2 = -0.5$, and $\eta_3 = 0.005$, then the upper support is $\tau_2 = 8$
If $\eta_1 = 5$, $\eta_2 = -0.5$, and $\eta_3 = -0.005$, then the lower support is $\tau_1 = 2$
95 % Cov The empirical coverage rate of the 95 % confidence intervals

curves for the NPMLE and cubic SEF are quite similar. However, their analysis only gives point estimates without precision (e.g., standard error and confidence interval). Here in this paper, we supply the previous results.

6.2 Numerical results

To fit the cubic SEF with the lower boundary $\tau_1 = y_{(1)} = 6$, we maximize the log-likelihood in Eq. (3), and obtain the MLE $\hat{\boldsymbol{\eta}}_n = (\hat{\eta}_{1n}, \hat{\eta}_{2n}, \hat{\eta}_{3n})^{\mathrm{T}} = (-0.00079, 3.38 \times 10^{-7}, -4.87 \times 10^{-11})$ by the Newton–Raphson algorithm as proposed in Hu and Emura (2015). We estimate the survival function by $S_{\hat{\boldsymbol{\eta}}}(y)$ for $y \geq \tau_1 = 6$, which is depicted in Fig. 1. Figure 1 also draws the 95 % confidence intervals for the survival function based on the method of Sect. 4.2.

Figure 1 compares the estimated survival function $S_{\hat{\boldsymbol{\eta}}}(y)$ with the NPMLE. As previously observed by Hu and Emura (2015), the two estimated survival curves are close to one another. Here in our analysis, this observation can be further confirmed since the lower and upper confidence intervals completely bracket the NPMLE. Still, this is not a formal statistical test since we do not consider the variability of the NPMLE.

An important advantage of the cubic SEF over the NPMLE is to provide a density estimator. The NPMLE gives a discrete distribution which cannot offer a continuous density function unless some non-trivial smoothing techniques are applied.

Figure 2 shows the estimated density along with the 95 % confidence intervals. The density has a heavy tail near the lower bound $\tau_1 = \min_i(y_i) = 6$. Biologically, this implies that there is higher risk of developing cancer in early ages, especially before 2-years old. This observation agrees with Emura et al. (2015) who tested the null hypothesis "childhood cancer occurs uniformly over all ages" against the alternative hypothesis "occurrence of childhood cancer decreases as their age increases". Statistically, the heavy tail is due to the effect of having a negative value $\hat{\eta}_3 = -4.87 \times 10^{-11}$.

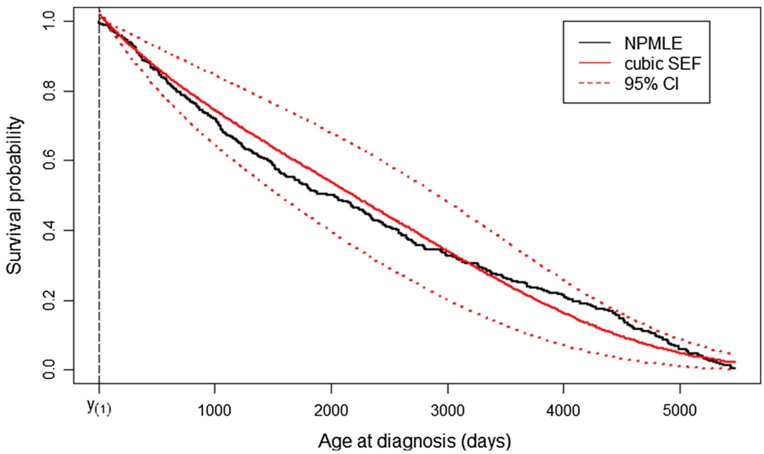

Fig. 1 Estimated survival functions for the childhood cancer data based on the cubic SEF and the NPMLE. *Dotted lines* are the 95 % confidence interval based on the cubic SEF. The *vertical line* signifies the lower boundary $\min_i(y_i) = y_{(1)} = 6$

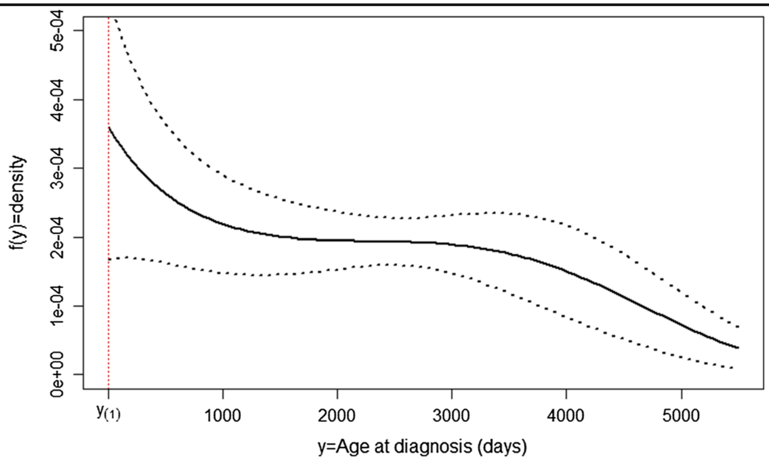

Fig. 2 The estimated density function (with 95 % confidence intervals) under the cubic SEF for the childhood cancer data. The *vertical line* is the lower bound $\tau_1 = \min_i(y_i) = y_{(1)} = 6$

Since the tail of the density is sensitive to the value of $\hat{\eta}_3$, the confidence intervals are wider there.

6.3 Checking regularity conditions

We examine how the regularity conditions for the asymptotic analysis are checked in terms of Assumption (G). The truncation mechanism corresponds to the fixed-length double-truncation (Example 2 of Sect. 4.3), where the follow-up length is fixed at 5 years, $d_0 = 1825$ (days). Assumption (G) requires that the follow-up is sufficiently long, i.e., $d_0 > b - a$, where $[a, b]$ is the support for the distribution of u_i^*'s (ages at the start of recruitment). Since the distribution of v_i^*'s is previously approximated by a uniform distribution on [0, 7300] (Moreira and de Uña-Álvarez 2010), we assume that u_i^*'s are uniformly distributed on [−1825, 5475]. Accordingly, $a = -1825$ and $b = 5475$. Unfortunately, Assumption (G) does not hold since $d_0 < b - a = 7300$. If the study could increase the follow-up length by $d_0 > 7300$, Assumption (G) would hold. This example demonstrates how Assumption (G) is checked and interpreted by the user.

7 Conclusion and discussion

When samples are subject to double-truncation, the asymptotic properties of MLE may not be derived through the classical theories for i.i.d. samples. The problem about how one should treat the non-identical truncation intervals among samples poses a unique problem for double-truncation, which has been missed in the literature. The goal of this paper is to point out the problem and to give a possible solution by deriving the formal asymptotic results under the theories on independent but not identically distributed (i.n.i.d.) random variables. The consistency and asymptotic normality of the MLE under the SEF are established assuming a reasonably simple set of regularity conditions. The simulations show that the standard error and confidence intervals based on our asymptotic theories exhibit desirable performance in finite samples. Utilizing the proposed confidence intervals, our analysis of the childhood cancer data confirms the previously reported findings on the risk of the cancer (e.g., higher risk of developing cancer in early ages).

We derive the set of regularity conditions (Assumptions A–E) such that the MLE is consistent and asymptotic normal under the k-dimensional SEF. Note that the SEF, as a member of the k-dimensional exponential family, satisfies many mathematically convenient properties, such as interchangeability of integration and differentiation (Sect. 2) and the concave property of the log-likelihood (Sect. 4). Consequently, our regularity conditions do not need to impose distributional assumptions, and hence they are more simplified than those required to regulate general parametric models. Importantly, this simplification makes some of our regularity conditions easily verifiable by users. Indeed, we show that part of the regularity conditions are satisfied under a very simple stability condition about the truncation intervals (Sect. 4.3). However, to extend the conditions to general parametric models, one needs to add extra conditions guaranteeing the desired distributional properties, with risk of increasing complexity.

It would be of great interest to examine the efficiency of the MLE. For our asymptotic analysis of the MLE, we have adopted the approach of Efron and Petrosian (1999) who constructed likelihood conditional on the truncation limits. This "conditional" approach has the advantage of being free from the distributional assumptions for the truncation limits. On the other hand, it is often natural to utilize distributional assumptions of the truncation limits into estimation. In particular, the assumptions that the left-truncation limit u_i^* is a realization from a uniform distribution, and the right-truncation limit is $v_i^* = u_i^* + d_0$, where $d_0 > 0$ is a constant, are often plausible in doubly-truncated data (Stovring and Wang 2007; Moreira and de Uña-Álvarez 2010). A related paper is De Uña-álvarez (2004) who constructed a moment-based estimator which is more efficient than the NPMLE when u_i^* follows a uniform distribution, and $v_i^* = u_i^* + d_0$ is a right-censoring limit (instead of right-truncation limit). An attempt to derive more efficient estimators than the MLE has not been made under the parametric models, which is an interesting topic for further investigation.

Acknowledgements We thank the editor Christine H. Müller and two reviewers for their helpful comments that led to improvements of our paper. We are also thankful to Prof. De Uña-álvarez for his comments on the earlier version of our paper. The work of T. Emura was supported by the research grant funded by the Ministry of Science and Technology of Taiwan (MOST 103-2118-M-008-MY2). The work of Y. Konno was partially supported by Grant-in-Aid for Scientific Research(C) (No. 25330043).

Appendix 1: Proofs

Proof of Lemma 1

It suffices to check that the $k \times k$ matrix $\partial^2 \ell_n(\boldsymbol{\eta})/\partial \boldsymbol{\eta}^2$ is negative semi-definite for any $\boldsymbol{\eta} \in \Theta$. Define

$$E_i^j(\boldsymbol{\eta}) = \int\limits_{R_i \cap \mathsf{y}} y^j \exp\{\boldsymbol{\eta}^{\mathrm{T}} \cdot \mathbf{t}(y)\} dy, \quad j = 0, 1, \ldots, 3k.$$

With these notations, the log-likelihood is written as

$$\ell_n(\boldsymbol{\eta}) = \sum_{i=1}^{n} \boldsymbol{\eta}^{\mathrm{T}} \cdot \mathbf{t}(y_i) - \sum_{i=1}^{n} \log E_i^0(\boldsymbol{\eta}).$$

As in Hu and Emura (2015), the score functions are

$$\frac{\partial}{\partial \eta_j} \ell_n(\boldsymbol{\eta}) = \sum_{i=1}^{n} \{y_i^j - E_i^j(\boldsymbol{\eta})/E_i^0(\boldsymbol{\eta})\}, \qquad j = 1, 2, \ldots, k,$$

and the second-order derivatives of the log-likelihood are

$$\begin{aligned} \frac{\partial^2}{\partial \eta_j \partial \eta_s} \ell_n(\boldsymbol{\eta}) &= -\sum_{i=1}^{n} [E_i^{j+s}(\boldsymbol{\eta})/E_i^0(\boldsymbol{\eta}) \\ &\quad - \{E_i^j(\boldsymbol{\eta})/E_i^0(\boldsymbol{\eta})\}\{E_i^s(\boldsymbol{\eta})/E_i^0(\boldsymbol{\eta})\}] \\ &= -\sum_{i=1}^{n} Cov_i(Y^j, Y^s|\boldsymbol{\eta}), \quad j, s = 1, 2, \ldots, k. \end{aligned}$$

Let $\mathbf{Cov}_i\{\mathbf{t}(Y)|\boldsymbol{\eta}\}$ be the covariance matrix whose (j, s) element is $Cov_i(Y^j, Y^s|\boldsymbol{\eta})$, $j, s = 1, 2, \ldots, k$. Then,

$$\frac{\partial^2 \ell_n(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^2} = -\sum_{i=1}^{n} \mathbf{Cov}_i\{\mathbf{t}(Y)|\boldsymbol{\eta}\}.$$

Since the covariance matrices $\mathbf{Cov}_i\{\mathbf{t}(Y)|\boldsymbol{\eta}\}, i = 1, 2, \ldots, n$ are positive semi-definite (see p. 287, Theorem B.2 of Sen and Srivastava 1990), their sum is also positive semi-definite. Hence, $\partial^2 \ell_n(\boldsymbol{\eta})/\partial \boldsymbol{\eta}^2$ is negative semi-definite. □

Proof of Theorem 1 (a): Existence and consistency

Under Assumption (A), one can define a subset of Θ,

$$Q_a = \{\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_k) : ||\boldsymbol{\eta} - \boldsymbol{\eta}^0||^2 \le a^2\},$$

where $||\boldsymbol{\eta}||^2 = \boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\eta}$ and $a > 0$ is a small number, which produces a sphere with center $\boldsymbol{\eta}^0$ and radius a.The surface of Q_a is defined as

$$\partial Q_a = \{\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_k) : ||\boldsymbol{\eta} - \boldsymbol{\eta}^0||^2 = a^2\}.$$

Now, we will show that for any sufficiently small a and for any $\boldsymbol{\eta} \in \partial Q_a$,

$$\lim_{n\to\infty} P\{\ell_n(\boldsymbol{\eta}) < \ell_n(\boldsymbol{\eta}^0)\} = \lim_{n\to\infty} P\left\{\frac{1}{n}\ell_n(\boldsymbol{\eta}) - \frac{1}{n}\ell_n(\boldsymbol{\eta}^0) < 0\right\} = 1.$$

This implies that, with probability tending to one, there exists a local maxima in Q_a, which solves Eq. (4).

By a Taylor expansion, we expand the log-likelihood about the true value $\boldsymbol{\eta}^0$ as

$$\begin{aligned}\ell_n(\boldsymbol{\eta}) &= \ell_n(\boldsymbol{\eta}^0) + \sum_{j=1}^{k}\left\{\frac{\partial}{\partial\eta_j}\ell_n(\boldsymbol{\eta})\bigg|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0}\right\}(\eta_j-\eta_j^0)\\ &\quad+\frac{1}{2!}\sum_{j=1}^{k}\sum_{s=1}^{k}\left\{\frac{\partial^2}{\partial\eta_j\partial\eta_s}\ell_n(\boldsymbol{\eta}\bigg|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0}\right\}(\eta_j-\eta_j^0)(\eta_s-\eta_s^0)\\ &\quad+\frac{1}{3!}\sum_{j=1}^{k}\sum_{s=1}^{k}\sum_{l=1}^{k}\left\{\frac{\partial^3}{\partial\eta_j\partial\eta_s\partial\eta_l}\ell_n(\boldsymbol{\eta})\bigg|_{\boldsymbol{\eta}=\boldsymbol{\eta}^*}\right\}\\ &\quad\times(\eta_j-\eta_j^0)(\eta_s-\eta_s^0)(\eta_l-\eta_l^0),\end{aligned} \tag{6}$$

where $\boldsymbol{\eta}^*$ is on the line between $\boldsymbol{\eta}$ and $\boldsymbol{\eta}^0$. By Assumption (C), there is a measurable function M_{jsl} such that

$$-M_{jsl}(y) \le \frac{\partial^3}{\partial\eta_j\partial\eta_s\partial\eta_l}\log f_i(y|\boldsymbol{\eta}^*) \le M_{jsl}(y), \quad i=1,2,\ldots,n.$$

This implies that

$$\frac{\partial^3}{\partial\eta_j\partial\eta_s\partial\eta_l}\log f_i(y|\boldsymbol{\eta}^*) = \gamma_{jsl}(y|\boldsymbol{\eta}^*)\cdot M_{jsl}(y),$$

for some $\gamma_{jsl}(y|\boldsymbol{\eta}^*) \in [-1, 1]$. Thus

$$\frac{\partial^3}{\partial\eta_j\partial\eta_s\partial\eta_l}\ \ell_n(\boldsymbol{\eta})|_{\boldsymbol{\eta}=\boldsymbol{\eta}^*} = \sum_{i=1}^{n}\gamma_{jsl}(y_i|\boldsymbol{\eta}^*)\cdot M_{jsl}(y_i).$$

Then, we rewrite Eq. (6) to yield

$$\begin{aligned}\frac{1}{n}\ell_n(\boldsymbol{\eta})-\frac{1}{n}\ell_n(\boldsymbol{\eta}^0) &= \frac{1}{n}\sum_{j=1}^{k}\left\{\frac{\partial}{\partial\eta_j}\ell_n(\boldsymbol{\eta})\bigg|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0}\right\}(\eta_j-\eta_j^0)\\ &\quad+\frac{1}{2n}\sum_{j=1}^{k}\sum_{s=1}^{k}\left\{\frac{\partial^2}{\partial\eta_j\partial\eta_s}\ell_n(\boldsymbol{\eta})\bigg|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0}\right\}(\eta_j-\eta_j^0)(\eta_s-\eta_s^0)\\ &\quad+\frac{1}{6n}\sum_{j=1}^{k}\sum_{s=1}^{k}\sum_{l=1}^{k}(\eta_j-\eta_j^0)(\eta_s-\eta_s^0)(\eta_l-\eta_l^0)\sum_{i=1}^{n}\gamma_{jsl}(y_i|\boldsymbol{\eta}^*)\cdot M_{jsl}(y_i)\\ &\equiv S_{n,1}(\boldsymbol{\eta})+S_{n,2}(\boldsymbol{\eta})+S_{n,3}(\boldsymbol{\eta}).\end{aligned}$$

Here, we define

$$\begin{cases} S_{n,1}(\boldsymbol{\eta}) \equiv \frac{1}{n}\sum_{j=1}^{k}\left\{\frac{\partial}{\partial\eta_j}\ell_n(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0}\right\}(\eta_j-\eta_j^0), \\ S_{n,2}(\boldsymbol{\eta}) \equiv \frac{1}{2n}\sum_{j=1}^{k}\sum_{s=1}^{k}\left\{\frac{\partial^2}{\partial\eta_j\partial\eta_s}\ell_n(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0}\right\}(\eta_j-\eta_j^0)(\eta_s-\eta_s^0), \\ S_{n,3}(\boldsymbol{\eta}) \equiv \frac{1}{6n}\sum_{j=1}^{k}\sum_{s=1}^{k}\sum_{l=1}^{k}(\eta_j-\eta_j^0)(\eta_s-\eta_s^0)(\eta_l-\eta_l^0)\sum_{i=1}^{n}\gamma_{jsl}(y_i|\boldsymbol{\eta}^*)\cdot M_{jsl}(y_i). \end{cases}$$

Our target is to prove that, for a sufficiently small a and for any $\boldsymbol{\eta}\in\partial Q_a$,

$$\lim_{n\to\infty}P\left\{\frac{1}{n}\ell_n(\boldsymbol{\eta})-\frac{1}{n}\ell_n(\boldsymbol{\eta}^0)<0\right\}=\lim_{n\to\infty}P\{S_{n,1}(\boldsymbol{\eta})+S_{n,2}(\boldsymbol{\eta})+S_{n,3}(\boldsymbol{\eta})<0\}=1.$$

By Lemma 2 (WLLN) and Assumption (B), one can obtain

$$\frac{1}{n}\frac{\partial}{\partial\eta_j}\ell_n(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0}=\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\eta_j}\log f_i(Y_i|\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0}\xrightarrow{p}0, \tag{7}$$

where we have verified the condition of Lemma 2 with $p=2$ by

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^{n}E\left\{\frac{\partial}{\partial\eta_j}\log f_i(Y_i|\boldsymbol{\eta}^0)\right\}^2$$
$$=\lim_{n\to\infty}\frac{1}{n}\cdot\frac{1}{n}\sum_{i=1}^{n}E\left\{\frac{\partial}{\partial\eta_j}\log f_i(Y_i|\boldsymbol{\eta}^0)\right\}^2=\lim_{n\to\infty}\frac{1}{n}I_{jj}(\boldsymbol{\eta}^0)=0.$$

Note that

$$\begin{aligned}\frac{1}{n}\frac{\partial^2}{\partial\eta_j\partial\eta_s}\ell_n(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0}&=\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\eta_j\partial\eta_s}\log f_i(y_i|\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0}\\&=\frac{1}{n}\sum_{i=1}^{n}\left[\left\{\frac{\partial^2}{\partial\eta_j\partial\eta_s}\log f_i(y_i|\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0}\right\}-\{-I_{i,js}(\boldsymbol{\eta}^0)\}\right]\\&\quad-\frac{1}{n}\sum_{i=1}^{n}I_{i,js}(\boldsymbol{\eta}^0).\end{aligned} \tag{8}$$

By Lemma 2 and Assumptions (B) and (D), Eq. (8) converges in probability to $-I_{js}(\boldsymbol{\eta}^0)$, where we have verified the condition of Lemma 2 with $p=2$ by

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^{n}E\left\{\frac{\partial^2}{\partial\eta_j\partial\eta_s}\log f_i(Y_i|\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0}\right\}^2\le\lim_{n\to\infty}\frac{1}{n}\cdot\frac{1}{n}\sum_{i=1}^{n}w_{i,js}^2$$

$$= \lim_{n\to\infty} \frac{1}{n} \cdot w_{js}^2 = 0.$$

Step 1 $\lim_{n\to\infty} P\{|S_{n,1}(\boldsymbol{\eta})| < ka^3\} = 1$ for any $\boldsymbol{\eta} \in \partial Q_a$:

Since $|\eta_j - \eta_j^0| \le a$ for any $\boldsymbol{\eta} \in \partial Q_a$, we have

$$|S_{n,1}(\boldsymbol{\eta})| \le a \sum_{j=1}^{k} \left| \frac{1}{n} \frac{\partial}{\partial \eta_j} \ell_n(\boldsymbol{\eta})|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0} \right|.$$

This implies

$$\{|S_{n,1}(\boldsymbol{\eta})| < ka^3\} \supset \left\{ a \sum_{j=1}^{k} \left| \frac{1}{n} \frac{\partial}{\partial \eta_j} \ell_n(\boldsymbol{\eta}) \right|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0} \right| < ka^3 \right\}.$$

Thus, we have

$$\lim_{n\to\infty} P\{|S_{n,1}(\boldsymbol{\eta})| < ka^3\} \ge \lim_{n\to\infty} P \left\{ a \sum_{j=1}^{k} \left| \frac{1}{n} \frac{\partial}{\partial \eta_j} \ell_n(\boldsymbol{\eta}) \right|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0} \right| < ka^3 \right\} = 1,$$

where the last equation follows from Eq. (7).

Step 2 $\lim_{n\to\infty} P\{S_{n,2}(\boldsymbol{\eta}) < -ca^2\} = 1$ for some $c > 0$ and for any $\boldsymbol{\eta} \in \partial Q_a$:

$$\begin{aligned} 2S_{n,2}(\boldsymbol{\eta}) =& \frac{1}{n} \sum_{j=1}^{k} \sum_{s=1}^{k} \left\{ \frac{\partial^2}{\partial \eta_j \partial \eta_s} \ell_n(\boldsymbol{\eta}) \right|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0} \right\} (\eta_j - \eta_j^0)(\eta_s - \eta_s^0) \\ =& \sum_{j=1}^{k} \sum_{s=1}^{k} \left[\frac{1}{n} \frac{\partial^2}{\partial \eta_j \partial \eta_s} \ell_n(\boldsymbol{\eta}) \right|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0} - \{-I_{js}(\boldsymbol{\eta}^0)\} \right] (\eta_j - \eta_j^0)(\eta_s - \eta_s^0) \\ & - \sum_{j=1}^{k} \sum_{s=1}^{k} I_{js}(\boldsymbol{\eta}^0)(\eta_j - \eta_j^0)(\eta_s - \eta_s^0) \\ \equiv& B_n(\boldsymbol{\eta}) + B(\boldsymbol{\eta}), \end{aligned} \tag{9}$$

where we define

$$B_n(\boldsymbol{\eta}) \equiv \sum_{j=1}^{k} \sum_{s=1}^{k} \left[\frac{1}{n} \frac{\partial^2}{\partial \eta_j \partial \eta_s} \ell_n(\boldsymbol{\eta}) \right|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0} - \{-I_{js}(\boldsymbol{\eta}^0)\} \right] (\eta_j - \eta_j^0)(\eta_s - \eta_s^0),$$

$$B(\boldsymbol{\eta}) \equiv \sum_{j=1}^{k} \sum_{s=1}^{k} \{-I_{js}(\boldsymbol{\eta}^0)\}(\eta_j - \eta_j^0)(\eta_s - \eta_s^0).$$

For $\boldsymbol{\eta} \in \partial Q_a$, we know that $|\eta_j - \eta_j^0| \le a$ and $|\eta_s - \eta_s^0| \le a$. Thus

$$|B_n(\boldsymbol{\eta})| \le a^2 \sum_{k=1}^{k} \sum_{s=1}^{k} \left| \frac{1}{n} \frac{\partial^2}{\partial \eta_j \partial \eta_s} \ell_n(\boldsymbol{\eta}) \Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0} - \{-I_{js}(\boldsymbol{\eta}^0)\} \right|.$$

By arguments following Eq. (8),

$$\lim_{n\to\infty} P\left(\left| \frac{1}{n} \frac{\partial^2}{\partial \eta_j \partial \eta_s} \ell_n(\boldsymbol{\eta}) \Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0} - \{-I_{js}(\boldsymbol{\eta}^0)\} \right| < \varepsilon \right) = 1,$$

for $\varepsilon > 0$. Letting $\varepsilon = a$,

$$\lim_{n\to\infty} P\left(\sum_{j=1}^{k} \sum_{s=1}^{k} a^2 \left| \frac{1}{n} \frac{\partial^2}{\partial \eta_j \partial \eta_s} \ell_n(\boldsymbol{\eta}) \Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0} - \{-I_{js}(\boldsymbol{\eta}^0)\} \right| < k^2 a^3 \right) = 1. \quad (10)$$

Note that

$$\begin{aligned} B(\boldsymbol{\eta}) &= \sum_{j=1}^{k} \sum_{s=1}^{k} \{-I_{js}(\boldsymbol{\eta}^0)\}(\eta_j - \eta_j^0)(\eta_s - \eta_s^0) = (\boldsymbol{\eta} - \boldsymbol{\eta}^0)^{\mathrm{T}} \{-I(\boldsymbol{\eta}^0)\}(\boldsymbol{\eta} - \boldsymbol{\eta}^0) \\ &= (\boldsymbol{\eta} - \boldsymbol{\eta}^0)^{\mathrm{T}} \{\Gamma \Lambda \Gamma^{\mathrm{T}}\}(\boldsymbol{\eta} - \boldsymbol{\eta}^0) = \{\Gamma^{\mathrm{T}}(\boldsymbol{\eta} - \boldsymbol{\eta}^0)\}^{\mathrm{T}} \cdot \Lambda \cdot \Gamma^{\mathrm{T}}(\boldsymbol{\eta} - \boldsymbol{\eta}^0), \end{aligned}$$

where $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_k)$ is a diagonal matrix of the eigenvalues of $-I(\boldsymbol{\eta}^0)$ and Γ is a orthogonal matrix ($\Gamma\Gamma^{\mathrm{T}} = \mathrm{I}$) whose column i corresponds to the eigenvector of λ_i. We order the eigenvalues such that $\lambda_k \le \cdots \le \lambda_2 \le \lambda_1$ and arrange Γ accordingly. By Assumption (B), we know that $\lambda_1 < 0$. Letting $\boldsymbol{\xi} = \Gamma^{\mathrm{T}}(\boldsymbol{\eta} - \boldsymbol{\eta}^0)$,

$$B(\boldsymbol{\eta}) = \sum\nolimits_{i=1}^{k} \lambda_i \xi_i^2 \le \sum\nolimits_{i=1}^{k} \lambda_1 \xi_i^2 = \lambda_1 \boldsymbol{\xi}^{\mathrm{T}} \boldsymbol{\xi} = \lambda_1 (\boldsymbol{\eta} - \boldsymbol{\eta}^0)^{\mathrm{T}} (\boldsymbol{\eta} - \boldsymbol{\eta}^0) = \lambda_1 a^2. \quad (11)$$

Form Eq. (10), we have

$$\lim_{n\to\infty} P(|B_n(\boldsymbol{\eta})| < k^2 a^3) = \lim_{n\to\infty} P(B_n(\boldsymbol{\eta}) < k^2 a^3) = 1,$$

and from Eq. (11), we know $B(\boldsymbol{\eta}) \le \lambda_1 a^2$. Thus,

$$\lim_{n\to\infty} P\left\{ S_{n,2}(\boldsymbol{\eta}) < \frac{k^2}{2} a^3 + \frac{\lambda_1}{2} a^2 \right\} = 1.$$

There always exist constants $c_0 > 0$ and $a_0 > 0$ such that, for $a < a_0$ and $0 < c < c_0$,

$$\lim_{n\to\infty} P\{S_{n,2}(\boldsymbol{\eta}) < -ca^2\} = 1.$$

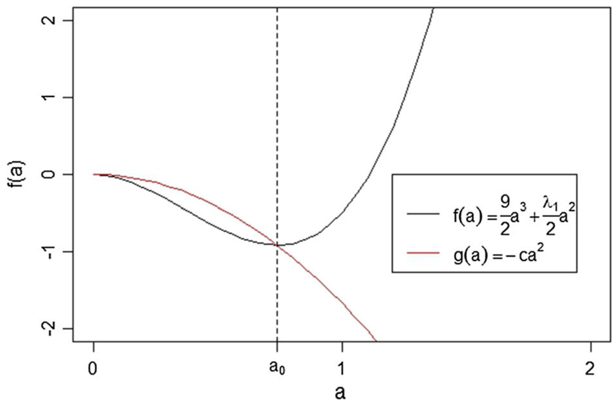

Fig. 3 The sketch of $f(a) = 9a^3/2 + \lambda_1 a^2/2$ and $g(a) = -ca^2$

The idea of choosing c_0 and a_0 is conveniently explained under $k = 3$ as follows: We wish to find a range of a such that $9a^3/2 + \lambda_1 a^2/2 \leq -ca^2$. This is explained in Fig. 3. Concretely,

$$f(a) = \frac{9}{2}a^3 + \frac{\lambda_1}{2}a^2 \Rightarrow f'(a) = \frac{27}{2}a^2 + \lambda_1 a = 0 \quad \Rightarrow \quad a = \frac{-2\lambda_1}{27}$$
$$\Rightarrow f''(a) = 27a + \lambda_1|_{a=-2\lambda_1/27} = -\lambda_1 > 0.$$

Then, $f(a)$ has the local minimum at $a_0 = -2\lambda_1/27 > 0$, and c_0 can be obtained by solving

$$\frac{9a^3}{2} + \frac{\lambda_1 a^2}{2} = -ca^2 \Rightarrow c = -\frac{\lambda_1}{2} - \frac{9a}{2}.$$

Hence, $c_0 = -\lambda_1/2 - 9a_0/2 = -\lambda_1/6 > 0$ as seen in Fig. 3.

The values a_0 and c are chosen such that $f(a) \leq g(a)$ for all $a < a_0$.

Step 3 $\lim_{n\to\infty} P\{|S_{n,3}(\boldsymbol{\eta})| < ba^3\} = 1$ for some $b > 0$ and for any $\boldsymbol{\eta} \in \partial Q_a$:

By Lemma 2 and Assumption (C), we obtain

$$\frac{1}{n}\sum_{i=1}^{n} M_{jsl}(Y_i) = \frac{1}{n}\sum_{i=1}^{n}[M_{jsl}(Y_i) - E\{M_{jsl}(Y_i)\}]$$
$$+\frac{1}{n}\sum_{i=1}^{n} E\{M_{jsl}(Y_i)\} \xrightarrow{p} m_{jsl},$$

where we have verified the condition $p = 2$ of Lemma 2 by

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^{n} E[M_{jsl}(Y_i)^2] = \lim_{n\to\infty}\frac{1}{n}\frac{1}{n}\sum_{i=1}^{n} E[M_{jsl}(Y_i)^2] = \lim_{n\to\infty}\frac{1}{n}\frac{1}{n}\sum_{i=1}^{n} m_{i,jsl}^2 = 0.$$

Then, we obtain

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} M_{jsl}(Y_i) - m_{jsl}\right| < \varepsilon\right\} = 1.$$

Letting $\varepsilon = m_{jsl}$ and by $M_{jsl}(Y_i) > 0$,

$$\begin{aligned}\lim_{n\to\infty} P&\left\{\left|\frac{1}{n}\sum_{i=1}^{n} M_{jsl}(Y_i) - m_{jsl}\right| < m_{jsl}\right\}\\ &= \lim_{n\to\infty} P\left\{\frac{1}{n}\sum_{i=1}^{n} M_{jsl}(Y_i) < 2m_{jsl}\right\} = 1.\end{aligned} \tag{12}$$

When $\boldsymbol{\eta} \in \partial Q_a$, we have $|\eta_j - \eta_j^0|$, $|\eta_s - \eta_s^0|$, $|\eta_l - \eta_l^0| \le a$. Thus,

$$\begin{aligned}|S_{n,3}(\boldsymbol{\eta})| &\le \frac{a^3}{6}\sum_{j=1}^{k}\sum_{s=1}^{k}\sum_{l=1}^{k}\left|\frac{1}{n}\sum_{i=1}^{n}\gamma_{jsl}(y_i|\boldsymbol{\eta}^*)M_{jsl}(y_i)\right|\\ &\le \frac{a^3}{6}\sum_{j=1}^{k}\sum_{s=1}^{k}\sum_{l=1}^{k}\frac{1}{n}\sum_{i=1}^{n} M_{jsl}(y_i).\end{aligned}$$

For any given $a > 0$, it follows from (12) that

$$\begin{aligned}1 &= \lim_{n\to\infty} P\left\{\frac{a^3}{6}\sum_{j=1}^{k}\sum_{s=1}^{k}\sum_{l=1}^{k}\frac{1}{n}\sum_{i=1}^{n} M_{jsl}(Y_i) < \frac{a^3}{6}\sum_{j=1}^{k}\sum_{s=1}^{k}\sum_{l=1}^{k} 2m_{jsl}\right\}\\ &= \lim_{n\to\infty} P\left\{\frac{a^3}{6}\sum_{j=1}^{k}\sum_{s=1}^{k}\sum_{l=1}^{k}\frac{1}{n}\sum_{i=1}^{n} M_{jsl}(Y_i) < \frac{a^3}{3}\sum_{j=1}^{k}\sum_{s=1}^{k}\sum_{l=1}^{k} m_{jsl}\right\}.\end{aligned}$$

This implies the desired result

$$\lim_{n\to\infty} P\{|S_{n,3}(\boldsymbol{\eta})| < ba^3\} = 1, \quad b = \frac{1}{3}\sum_{j=1}^{k}\sum_{s=1}^{k}\sum_{l=1}^{k} m_{jsl}.$$

Combining the results of Steps 1–3, we know that

$$\lim_{n\to\infty} P\left\{S_{n,1}(\boldsymbol{\eta}) + S_{n,2}(\boldsymbol{\eta}) + S_{n,3}(\boldsymbol{\eta}) < ka^3 - ca^2 + ba^3\right\} = 1,$$

and that

$$\lim_{n\to\infty} P\left\{\frac{1}{n}\ell_n(\boldsymbol{\eta}) - \frac{1}{n}\ell_n(\boldsymbol{\eta}^0) < ka^3 - ca^2 + ba^3\right\} = 1.$$

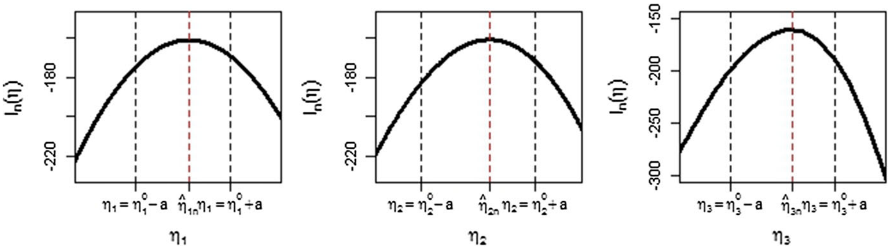

Fig. 4 The occurrence $\{\ell_n(\boldsymbol{\eta}) - \ell_n(\boldsymbol{\eta}^0) < 0\} \subset \{||\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}^0|| \leq a\}$, where $\boldsymbol{\eta}^0 = (\eta_1^0, \eta_2^0, \eta_3^0)$ and $\boldsymbol{\eta} \in \partial Q_a$ for a small $a > 0$

To complete the proof, we choose a such that $ka^3 - ca^2 + ba^3 < 0$, equivalently $a < c/(b+k)$. This is possible by taking a as small as possible. With this choice, there always exists $\hat{\boldsymbol{\eta}}_n$ such that $\{\ell_n(\boldsymbol{\eta}) - \ell_n(\boldsymbol{\eta}^0) < 0\} \subset \{||\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}^0|| \leq a\}$ with probability tending to one. Please see Fig. 4 for our numerical example of $k = 3$ in which the preceding relationship occurs. Therefore, letting $\varepsilon = a$, we have shown the existence of $\hat{\boldsymbol{\eta}}_n$ (with probability tending to one) and consistency simultaneously as

$$\lim_{n\to\infty} P(||\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}|| \leq \varepsilon) \geq \lim_{n\to\infty} P(\ell_n(\boldsymbol{\eta}) - \ell_n(\boldsymbol{\eta}^0) < 0) = 1.$$

Proofs of Theorem 1 (b)

By a Taylor expansion, we expand the first order derivative of log-likelihood function between the MLE $\hat{\boldsymbol{\eta}}_n$ and the true value $\boldsymbol{\eta}^0$ as

$$\begin{aligned} 0 = \frac{\partial}{\partial \eta_j} \ell_n(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0} &+ \sum_{s=1}^{k} \left\{ \frac{\partial^2}{\partial \eta_j \partial \eta_s} \ell_n(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0} \right\} (\hat{\eta}_{sn} - \eta_s^0) \\ &+ \frac{1}{2} \sum_{s=1}^{k} \sum_{l=1}^{k} \left\{ \frac{\partial^3}{\partial \eta_j \partial \eta_s \partial \eta_l} \ell_n(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\tilde{\boldsymbol{\eta}}_n} \right\} (\hat{\eta}_{sn} - \eta_s^0)(\hat{\eta}_{ln} - \eta_l^0), \end{aligned}$$

where $\tilde{\boldsymbol{\eta}}_n$ is on the line between $\hat{\boldsymbol{\eta}}_n$ and $\boldsymbol{\eta}^0$. It follows that

$$\begin{aligned} \frac{\partial}{\partial \eta_j} \ell_n(\boldsymbol{\eta}^0)\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0} = &- \sum_{s=1}^{k} \left\{ \frac{\partial^2}{\partial \eta_j \partial \eta_s} \ell_n(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0} \right\} (\hat{\eta}_{sn} - \eta_s^0) \\ &- \frac{1}{2} \sum_{s=1}^{k} \sum_{l=1}^{k} \left\{ \frac{\partial^3}{\partial \eta_j \partial \eta_s \partial \eta_l} \ell_n(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\tilde{\boldsymbol{\eta}}_n} \right\} (\hat{\eta}_{sn} - \eta_s^0)(\hat{\eta}_{ln} - \eta_l^0). \end{aligned}$$

Multiplying $1/\sqrt{n}$ both sides,

$$\frac{1}{\sqrt{n}} \frac{\partial}{\partial \eta_j} \ell_n(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0} = \sum_{s=1}^{k} \Bigg[-\frac{1}{n} \frac{\partial^2}{\partial \eta_j \partial \eta_s} \ell_n(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0}$$

$$
\left. - \frac{1}{2n} \sum_{l=1}^{k} \left\{ \frac{\partial^3}{\partial \eta_j \partial \eta_s \partial \eta_l} \ell_n(\boldsymbol{\eta}) \right|_{\boldsymbol{\eta}=\tilde{\boldsymbol{\eta}}_n} \right\} (\hat{\eta}_{ln} - \eta_l^0) \Bigg]
\times \sqrt{n}(\hat{\eta}_{sn} - \eta_s^0).
$$

This is written as

$$
T_{n,j}(\boldsymbol{\eta}^0) = \sum_{s=1}^{k} R_{n,js}(\boldsymbol{\eta}^0) \cdot C_{n,s}(\boldsymbol{\eta}^0), \quad j = 1, 2, \ldots, k,
$$

where

$$
\begin{aligned}
T_{n,j}(\boldsymbol{\eta}^0) &\equiv \frac{1}{\sqrt{n}} \frac{\partial}{\partial \eta_j} \ell_n(\boldsymbol{\eta}) \Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial \eta_j} \log f_i(y_i|\boldsymbol{\eta}) \Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0}, \\
R_{n,js}(\boldsymbol{\eta}^0) &\equiv -\frac{1}{n} \frac{\partial^2}{\partial \eta_j \partial \eta_s} \ell_n(\boldsymbol{\eta}) \Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0} \\
&\quad - \frac{1}{2n} \sum_{l=1}^{k} \left\{ \frac{\partial^3}{\partial \eta_j \partial \eta_s \partial \eta_l} \ell_n(\boldsymbol{\eta}) \Big|_{\boldsymbol{\eta}=\tilde{\boldsymbol{\eta}}_n} \right\} (\hat{\eta}_{ln} - \eta_l^0), \\
C_{n,s}(\boldsymbol{\eta}^0) &\equiv \sqrt{n}(\hat{\eta}_{sn} - \eta_s^0).
\end{aligned}
$$

Our target is to prove the convergence of $\mathbf{C}_n = (C_{n,1}, C_{n,2}, \ldots, C_{n,k})^{\mathrm{T}}$.

Step 1 $\mathbf{T}_n(\boldsymbol{\eta}^0) = (T_{n,1}(\boldsymbol{\eta}^0), T_{n,2}(\boldsymbol{\eta}^0), \ldots, T_{n,k}(\boldsymbol{\eta}^0))^{\mathrm{T}} \xrightarrow{d} N_k(\mathbf{0}, I(\boldsymbol{\eta}^0))$.
Let $\mathbf{T}_n(\boldsymbol{\eta}^0) = \sum_{i=1}^{n} \mathbf{D}_{n,i}$, where

$$
\mathbf{D}_{n,i} = \left[\frac{1}{\sqrt{n}} \frac{\partial}{\partial \eta_1} \log f_i(y_i|\boldsymbol{\eta}),\ \frac{1}{\sqrt{n}} \frac{\partial}{\partial \eta_2} \log f_i(y_i|\boldsymbol{\eta}), \ldots,\ \frac{1}{\sqrt{n}} \frac{\partial}{\partial \eta_k} \log f_i(y_i|\boldsymbol{\eta}) \right]^{\mathrm{T}} \Bigg|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0}.
$$

For the Lindeberg–Feller multivariate CLT to be applied, we check the Lindeberg condition in Eq. (5). For any $\varepsilon > 0$,

$$
\begin{aligned}
&\sum_{i=1}^{n} E_{\boldsymbol{\eta}^0}(||\mathbf{D}_{n,i} - E[\mathbf{D}_{n,i}]||^2 1\{||\mathbf{D}_{n,i} - E[\mathbf{D}_{n,i}]|| > \varepsilon\}) \\
&= \sum_{i=1}^{n} E_{\boldsymbol{\eta}^0} \left[\frac{1}{n} \sum_{j=1}^{k} \left\{ \frac{\partial}{\partial \eta_j} \log f_i(Y_i|\boldsymbol{\eta}) \right\}^2 \right. \\
&\quad \left. \times 1 \left\{ \frac{1}{n} \sum_{j=1}^{k} \left\{ \frac{\partial}{\partial \eta_j} \log f_i(Y_i|\boldsymbol{\eta}) \right\}^2 > \varepsilon^2 \right\} \right] \Bigg|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0}.
\end{aligned}
$$

By Assumption (E),

$$\frac{1}{n}\sum_{j=1}^{k}\left\{\frac{\partial}{\partial\eta_j}\log f_i(Y_i|\boldsymbol{\eta}^0)\right\}^2 \le \frac{1}{n}\sum_{j=1}^{k}A_j^2(Y_i) \le \frac{1}{n}\sum_{j=1}^{k}\sup_y A_j^2(y).$$

Hence,

$$1\left\{\frac{1}{n}\sum_{j=1}^{k}\left\{\frac{\partial}{\partial\eta_j}\log f_i(Y_i|\boldsymbol{\eta})\right\}^2 > \varepsilon^2\right\} \le 1\left\{\frac{1}{n}\sum_{j=1}^{k}\sup_y A_j^2(y) > \varepsilon^2\right\}, \quad i=1,2,\ldots,n.$$

It follows that

$$\begin{aligned}
&\sum_{i=1}^{n}E_{\boldsymbol{\eta}^0}(||\mathbf{D}_{n,i}-E\mathbf{D}_{n,i}||^2 1\{||\mathbf{D}_{n,i}-E\mathbf{D}_{n,i}||>\varepsilon\})\\
&\le \sum_{i=1}^{n}E_{\boldsymbol{\eta}^0}\left[\frac{1}{n}\sum_{j=1}^{k}\left\{\frac{\partial}{\partial\eta_j}\log f_i(Y_i|\boldsymbol{\eta})\right\}^2 1\left\{\frac{1}{n}\sum_{j=1}^{k}\sup_y A_j^2(y)>\varepsilon^2\right\}^2\right]\Bigg|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0}\\
&= 1\left\{\frac{1}{n}\sum_{j=1}^{k}\sup_y A_j^2(y)>\varepsilon^2\right\}\sum_{i=1}^{n}E_{\boldsymbol{\eta}^0}\left[\frac{1}{n}\sum_{j=1}^{k}\left\{\frac{\partial}{\partial\eta_j}\log f_i(Y_i|\boldsymbol{\eta})\right\}^2\right]\Bigg|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0}\\
&= 1\left\{\frac{1}{n}\sum_{j=1}^{k}\sup_y A_j^2(y)>\varepsilon^2\right\}\sum_{j=1}^{k}\sum_{i=1}^{n}\frac{1}{n}I_{i,jj}(\boldsymbol{\eta}^0) \to 1\{0>\varepsilon^2\}\sum_{j=1}^{k}I_{jj}(\boldsymbol{\eta}^0)=0,
\end{aligned}$$

where the last convergence follows from Assumptions (B) and (E). Hence, the Lindeberg condition in Lemma 3 holds. In addition, by Assumption (B),

$$\sum_{i=1}^{n}\{\mathbf{Cov}_{\boldsymbol{\eta}^0}(\mathbf{D}_{n,i})\}_{js} = \frac{1}{n}\sum_{i=1}^{n}I_{i,js}(\boldsymbol{\eta}^0) \to I_{js}(\boldsymbol{\eta}^0).$$

By Lemma 3 (the Lindeberg–Feller CLT),

$$\mathbf{T}_n(\boldsymbol{\eta}^0) = \sum\nolimits_{i=1}^{n}\mathbf{D}_{n,i} \xrightarrow{d} \mathbf{T}(\boldsymbol{\eta}^0) \sim N_k(\mathbf{0}, I(\boldsymbol{\eta}^0)).$$

Step 2 $R_{n,js}(\boldsymbol{\eta}^0) \xrightarrow{p} I_{js}(\boldsymbol{\eta}^0)$

Recall that

$$\begin{aligned}
R_{n,js}(\boldsymbol{\eta}^0) \equiv & -\frac{1}{n}\frac{\partial^2}{\partial\eta_j\partial\eta_s}\ell_n(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0}\\
& -\frac{1}{2n}\sum_{l=1}^{k}\left\{\frac{\partial^3}{\partial\eta_j\partial\eta_s\partial\eta_l}\ell_n(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\tilde{\boldsymbol{\eta}}_n}\right\}(\hat{\eta}_{ln}-\eta_l^0).
\end{aligned}$$

By the arguments following Eq. (8),

$$-\frac{1}{n}\frac{\partial^2}{\partial\eta_j\partial\eta_s}\ell_n(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0} \xrightarrow{p} I_{js}(\boldsymbol{\eta}^0).$$

Since $\hat{\boldsymbol{\eta}}_n \xrightarrow{P} \boldsymbol{\eta}^0$ and

$$\begin{aligned}\left|\frac{1}{n}\frac{\partial^3}{\partial\eta_j\partial\eta_s\partial\eta_l}\ell_n(\boldsymbol{\eta})\,\Big|_{\boldsymbol{\eta}=\tilde{\boldsymbol{\eta}}_n}\right| &= \left|\frac{1}{n}\sum_{i=1}^{n}\gamma_{jsl}(Y_i|\tilde{\boldsymbol{\eta}}_n)\cdot M_{jsl}(Y_i)\right| \\ &\le \frac{1}{n}\sum_{i=1}^{n}M_{jsl}(Y_i) \xrightarrow{p} m_{jsl},\end{aligned}$$

by Slutsky's theorem,

$$-\frac{1}{2n}\sum_{l=1}^{k}\left\{\frac{\partial^3}{\partial\eta_j\partial\eta_s\partial\eta_l}\ell_n(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\tilde{\boldsymbol{\eta}}_n}\right\}(\hat{\eta}_{ln}-\eta_l^0) \xrightarrow{p} 0.$$

Hence, we have $R_{n,js}(\boldsymbol{\eta}^0) \xrightarrow{p} I_{js}(\boldsymbol{\eta}^0)$.

Lemma 5 (Lehmann and Casella 1998) *Let* $\mathbf{T}_n = (T_{1n}, T_{2n}, \ldots, T_{kn}) \xrightarrow{d} \mathbf{T} = (T_1, T_2, \ldots, T_k)$. *Suppose that for fixed* j *and* s, *let* R_{jsn} *be a sequence of random variables, where* $R_{jsn} \xrightarrow{p} r_{js}$ *(constants) for which the matrix* $\mathbf{R}$, *with each element* r_{js}, *is nonsingular. Let* $\mathbf{B} = \mathbf{R}^{-1}$ *with each element* b_{js}. *Let* $\mathbf{C}_n = (C_{1n}, C_{2n}, \ldots, C_{kn})$ *be a solution to*

$$\sum_{s=1}^{k}R_{jsn}C_{sn} = T_{jn}, \quad j = 1, 2, \ldots, k,$$

and let $\mathbf{C} = (C_1, C_2, \ldots, C_k)$ *be a solution to*

$$\sum_{s=1}^{k}r_{js}C_s = T_j, \quad j = 1, 2, \ldots, k,$$

given by $C_j = \sum_{s=1}^{k} b_{js}T_s$, $j = 1, 2, \ldots, k$. *Then, if the distribution of* $(T_1, T_2, \ldots, T_k)$ *has a density,*

$$\mathbf{C}_n = (C_{1n}, C_{2n}, \ldots, C_{kn}) \xrightarrow{d} \mathbf{C} = (C_1, C_2, \ldots, C_k), \quad n \to \infty.$$

Combining Steps 1–2 with Lemma 5, $\mathbf{C}_n = \sqrt{n}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}^0)$ converges in distribution to $\mathbf{C}$, a solution to

$$\sum_{s=1}^{k}I_{js}(\boldsymbol{\eta}^0)C_s = T_j(\boldsymbol{\eta}^0), \quad j = 1, 2, \ldots, k,$$

where $\mathbf{T}(\boldsymbol{\eta}^0) = (T_1(\boldsymbol{\eta}^0), T_2(\boldsymbol{\eta}^0), \ldots, T_k(\boldsymbol{\eta}^0)) \sim N_k(\mathbf{0}, I(\boldsymbol{\eta}^0))$. Therefore, we have the desired result $\mathbf{C} = [I(\boldsymbol{\eta}^0)]^{-1} \cdot \mathbf{T}(\boldsymbol{\eta}^0) \sim N_k(\mathbf{0}, [I(\boldsymbol{\eta}^0)]^{-1})$. □

Proofs of Lemma 4

Using the notations of "Proof of Lemma 1" in Appendix 1,

$$\frac{\partial}{\partial \eta_j} \log f_i(y|\boldsymbol{\eta}) = y^j - \frac{E_i^j(\boldsymbol{\eta})}{E_i^0(\boldsymbol{\eta})}, \quad j = 1, 2, \ldots, k.$$

Under Assumption (G), $[u_0, v_0] \subset [u_i, v_i] = R_i \subset \mathsf{y}$. Thus,

$$E_i^0(\boldsymbol{\eta}) = \int_{R_i \cap y} \exp\{\boldsymbol{\eta}^{\mathrm{T}} \cdot \mathbf{t}(y)\} dy \geq \int_{u_0}^{v_0} \exp\{\boldsymbol{\eta}^{\mathrm{T}} \cdot \mathbf{t}(y)\} dy.$$

It follows from Assumption (F) that

$$\inf_{\boldsymbol{\eta} \in \Theta} E_i^0(\boldsymbol{\eta}) \geq \inf_{\boldsymbol{\eta} \in \Theta} \int_{u_0}^{v_0} \exp\{\boldsymbol{\eta}^{\mathrm{T}} \cdot \mathbf{t}(y)\} dy \equiv E_{\mathrm{Inf}}^0 > 0, \quad i = 1, 2, \ldots, n,$$

Similarly, since all the moments exist,

$$\sup_{\eta \in \Theta} |E_i^j(\boldsymbol{\eta})| \leq \sup_{\boldsymbol{\eta} \in \Theta} \int_{\mathsf{y}} |y|^j \exp\{\boldsymbol{\eta}^{\mathrm{T}} \cdot \mathbf{t}(y)\} dy \equiv E_{\mathrm{Sup}}^j < \infty, \quad j = 0, 1, \ldots, 3k,$$

for $i = 1, 2, \ldots, n$. Then, as in "Proof of Lemma 1" in Appendix 1,

$$\begin{aligned} \left| \frac{\partial^2}{\partial \eta_j \partial \eta_s} \log f_i(y|\boldsymbol{\eta}) \right| &\leq \left| \frac{E_i^{j+s}(\boldsymbol{\eta})}{E_i^0(\boldsymbol{\eta})} - \frac{E_i^j(\boldsymbol{\eta})}{E_i^0(\boldsymbol{\eta})} \frac{E_i^s(\boldsymbol{\eta})}{E_i^0(\boldsymbol{\eta})} \right| \\ &\leq \frac{\sup_\eta |E_i^{j+s}(\boldsymbol{\eta})|}{\inf_\eta E_i^0(\boldsymbol{\eta})} + \frac{\sup_\eta |E_i^j(\boldsymbol{\eta})|}{\inf_\eta E_i^0(\boldsymbol{\eta})} \frac{\sup_\eta |E_i^s(\boldsymbol{\eta})|}{\inf_\eta E_i^0(\boldsymbol{\eta})} \\ &\leq \frac{E_{\mathrm{Sup}}^{j+s}}{E_{\mathrm{Inf}}^0} + \frac{E_{\mathrm{Sup}}^j}{E_{\mathrm{Inf}}^0} \frac{E_{\mathrm{Sup}}^s}{E_{\mathrm{Inf}}^0} \equiv W_{js}(y) < \infty. \end{aligned}$$

In this way, one can find all the constant functions $W_{js}(\cdot)$ that satisfy the requirements of Assumption (D). In a similar fashion, Assumption (C) can be checked with

$$\begin{aligned} \left| \frac{\partial^3}{\partial \eta_j \partial \eta_s^\partial \eta_l} \log f_i(y|\boldsymbol{\eta}) \right| \leq &\frac{E_{\mathrm{Sup}}^{j+s+l}}{E_{\mathrm{Inf}}^0} + \frac{E_{\mathrm{Sup}}^{j+s}}{E_{\mathrm{Inf}}^0} \frac{E_{\mathrm{Sup}}^l}{E_{\mathrm{Inf}}^0} + \frac{E_{\mathrm{Sup}}^j}{E_{\mathrm{Inf}}^0} \frac{E_{\mathrm{Sup}}^{l+s}}{E_{\mathrm{Inf}}^0} + \frac{E_{\mathrm{Sup}}^l}{E_{\mathrm{Inf}}^0} \frac{E_{\mathrm{Sup}}^{j+s}}{E_{\mathrm{Inf}}^0} \\ &+2 \frac{E_{\mathrm{Sup}}^j}{E_{\mathrm{Inf}}^0} \frac{E_{\mathrm{Sup}}^s}{E_{\mathrm{Inf}}^0} \frac{E_{\mathrm{Sup}}^l}{E_{\mathrm{Inf}}^0} \equiv M_{jsl}(y) < \infty. \end{aligned}$$

To check Assumption (E), we use $|y^j| \le \max\{|u_i^j|, |v_i^j|\} \le \max\{|u_0^*|^j, |v_0^*|^j\} < \infty$ for $u_i \le y \le v_i$. Then,

$$\left|\frac{\partial}{\partial \eta_j} \log f_i(y|\boldsymbol{\eta})\right| \le |y^j| 1\{u_i \le y \le v_i\} + \frac{\sup_{\boldsymbol{\eta}} E_i^j(\boldsymbol{\eta})}{\inf_{\boldsymbol{\eta}} E_i^0(\boldsymbol{\eta})}$$
$$\le \max\{|u_0^*|^j, |v_0^*|^j\} + \frac{E_{\text{Sup}}^j}{E_{\text{Inf}}^0} \equiv A_j(y).$$

Hence, Assumption (E) holds for the constant function $A_j(\cdot)$.

Appendix 2: Data generations

For the cubic SEF with $\eta_3 > 0$, we consider $U^* \sim N(\mu_u, 1)$, $V^* \sim \min\{N(\mu_v, 1), \tau_2\}$ and

$$Y^* \sim f_{\boldsymbol{\eta}}(y) = \exp[\eta_1 y + \eta_2 y^2 + \eta_3 y^3 - \phi(\boldsymbol{\eta})], \quad y \in \mathsf{y} = (-\infty, \tau_2],$$

where $\phi(\boldsymbol{\eta}) = \log\{\int_{\mathsf{y}} \exp(\eta_1 y + \eta_2 y^2 + \eta_3 y^3) dy\}$. The value Y^* is generated by solving

$$W^* = F_{\boldsymbol{\eta}}(Y^*) = \frac{\int_{-\infty}^{Y^*} \exp[\eta_1 y + \eta_2 y^2 + \eta_3 y^3] dy}{\int_{-\infty}^{\tau_2} \exp[\eta_1 y + \eta_2 y^2 + \eta_3 y^3] dy},$$

where $W^* \sim U(0, 1)$. Under these models, we know $u_{\text{inf}} = \inf_i(u_i^*) = \inf_i(u_i) = -\infty$ and $v_{\text{sup}} = \sup_i(v_i^*) = \sup_i(v_i) = \tau_2$. Under this setting, Assumption (G) does not hold as $u_{\text{inf}} = -\infty$ is not a finite number. In addition, there is a chance that the length $v_i - u_i$ is quite small. The case of $\eta_3 < 0$ is similar. It would be of our interest to study the numerical properties of the MLE under this delicate setting.

We set the sample inclusion probability to be $P(U^* \le Y^* \le V^*) \approx 0.5$ or 0.25 by letting $\mu_u = \eta_1 - \Delta$ and $\mu_v = \eta_1 + \Delta$. First, under $\eta_1 = 5$, $\eta_2 = -0.5$, $\eta_3 = 0.005$ and $\tau_2 = 8$, the value is $\Delta = 1.01$ (Hu and Emura 2015) to meet $P(U^* \le Y^* \le V^*) \approx 0.50$. If we set $\Delta = 0.33$ then $P(U^* \le Y^* \le V^*) \approx 0.25$. Second, under $\eta_1 = 5$, $\eta_2 = -0.5$, $\eta_3 = -0.005$, and $\tau_1 = 2$, we set $\Delta = 0.91$ (Hu and Emura 2015) to meet $P(U^* \le Y^* \le V^*) \approx 0.50$. If we set $\Delta = 0.26$, then $P(U^* \le Y^* \le V^*) \approx 0.25$.

References

Andersen PK, Keiding N (2002) Multi-state models for event history analysis. Stat Methods Med Res 11:91–115

Austin D, Simon DK, Betensky RA (2014) Computationally simple estimation and improved efficiency for special cases of double truncation. Lifetime Data Anal 20(3):335–354

Bradley RA, Gart JJ (1962) The asymptotic properties of ML estimators when sampling from associated population. Biometrika 49:205–214

Cohen AC (1991) Truncated and censored samples. Marcel Dekker, New York

Castillo JD (1994) The singly truncated normal distribution: a non-steep exponential family. Ann Inst Stat Math 46:57–66

De Uña-álvarez J (2004) Nonparametric estimation under length-biased sampling and type I censoring: a moment based approach. Ann Inst Stat Math 56:667–681

Efron B, Petrosian R (1999) Nonparametric methods for doubly truncated data. J Am Stat Assoc 94:824–834

Emura T, Konno Y (2012) Multivariate normal distribution approaches for dependently truncated data. Stat Pap 53:133–149

Emura T, Murotani K (2015) An algorithm for estimating survival under a copula-based dependent truncation model. TEST. doi:10.1007/s11749-015-0432-8

Emura T, Konno Y, Michimae H (2015) Statistical inference based on the nonparametric maximum likelihood estimator under double-truncation. Lifetime Data Anal 21(3):397–418

Hoadley B (1971) Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. Ann Math Stat 42:1977–1991

Hu YH, Emura T (2015) Maximum likelihood estimation for a special exponential family under random double-truncation. Comput Stat. doi:10.1007/s00180-015-0564-z

Klein JP, Moeschberger ML (2003) Survival analysis: techniques for censored and truncated data, 2nd edn. Springer, New York

Lagakos SW, Barraj LM, De Gruttola V (1988) Non-parametric analysis of truncated survival data with application to AIDS. Biometrika 75:515–523

Lawless JF (2003) Statistical models and methods for lifetime data, 2nd edn. Wiley, New York

Lehmann EL, Casella G (1998) Theory of point estimation. Springer, New York

Lehmann EL, Romano JP (2005) Testing statistical hypotheses. Springer, New York

Mandrekar SJ, Nandrekar JN (2003) Are our data symmetric? Stat Methods Med Res 12:505–513

Moreira C, de Uña-Álvarez J (2010) Bootstrapping the NPMLE for doubly truncated data. J Nonparametric Stat 22:567–583

Moreira C, de Uña-Álvarez J (2012) Kernel density estimation with doubly-truncated data. Electron J Stat 6:501–521

Moreira C, de Uña-Álvarez J, Van Keilegom I (2014) Goodness-of-fit tests for a semiparametric model under random double truncation. Comput Stat 29(5):1365–1379

Moreira C, Van Keilegom I (2013) Bandwidth selection for kernel density estimation with doubly truncated data. Comput Stat Data Anal 61:107–123

Philippou A, Roussas G (1975) Asymptotic normality of the maximum likelihood estimate in the independent but not identically distributed case. Ann Inst Stat Math 27:45–55

Robertson HT, Allison DB (2012) A novel generalized normal distribution for human longevity and other negatively skewed data. PLoS ONE 7:e37025

Sankaran PG, Sunoj SM (2004) Identification of models using failure rate and mean residual life of doubly truncated random variables. Stat Pap 45:97–109

Seaman SR, White IR (2011) Review of inverse probability weighting for dealing with missing data. Stat Methods Med Res 22(3):278–295

Sen A, Srivastava M (1990) Regression analysis, theory, methods, and applications. Springer, New York

Shao J (2003) Mathematical statistics. Springer, New York

Shen PS (2010) Nonparametric analysis of doubly truncated data. Ann Inst Stat Math 62:835–853

Shen PS (2011) Testing quasi-independence for doubly truncated data. J Nonparametric Stat 23:1–9

Stovring H, Wang MC (2007) A new approach of nonparametric estimation of incidence and lifetime risk based on birth rates and incidence events. BMC Med Res Methodol 7:53

Strzalkowska-Kominiak E, Stute W (2013) Empirical copulas for consequtive survival data: copulas in survival analysis. TEST 22:688–714

Van der Vaart AW (1998) Asymptotic statistics. Cambridge University Press, Cambridge