Comput Stat (2015) 30:1199–1229
DOI 10.1007/s00180-015-0564-z

ORIGINAL PAPER

Maximum likelihood estimation for a special exponential family under random double-truncation

Ya-Hsuan Hu · Takeshi Emura

Received: 23 June 2014 / Accepted: 28 January 2015 / Published online: 12 February 2015

Abstract Doubly-truncated data often appear in lifetime data analysis, where samples are collected under certain time constraints. Nonparametric methods for doubly-truncated data have been studied well in the literature. Alternatively, this paper considers parametric inference when samples are subject to double-truncation. Efron and Petrosian (J Am Stat Assoc 94:824–834, 1999) proposed to fit a parametric family, called the special exponential family, with doubly-truncated data. However, non-trivial technical aspects, such as parameter space, support of the density, and computational algorithms, have not been discussed in the literature. This paper fills this gap by providing the technical aspects, including adequate choices of parameter space as well as support, and reliable computational algorithms. Simulations are conducted to verify the suggested techniques, and real data are used for illustration.

Keywords Fixed point iteration · Newton–Raphson algorithm · Survival analysis · Truncated data

1 Introduction

Statistical inferences for doubly truncated data have been an active research area with a variety of applications. A paper by Efron and Petrosian (1999) first developed inference procedures for doubly truncated data, and analyzed the quasar luminosity data in astronomy. In particular, due to the resolution of telescopes, the luminosity of stars may be undetected if it is either too dim or too bright, leading to double truncation (i.e., both

Y.-H. Hu · T. Emura (✉)
Graduate Institute of Statistics, National Central University, Taoyuan, Taiwan
e-mail: takeshiemura@gmail.com; emura@stat.ncu.edu.tw

Y.-H. Hu
e-mail: shumau6@hotmail.com

lower and upper truncations). Moreira and de Uña-Álvarez (2010) considered doubly truncated data, arising from the childhood cancer study of North Portugal. Other examples of double-truncation can be found in medical settings, including Stovring and Wang (2007), Zhu and Wang (2012) and Moreira et al. (2014). In general, double truncation is very common in fields such as medical, astronomical, population ageing and industrial studies.

Ignoring double-truncation effect leads to biased estimation. For $i = 1, 2, \dots, N$, let y_i be random samples from a density f, and $R_i = [u_i, v_i]$ be random intervals, where u_i and v_i are the left- and right-truncation limits, respectively. Due to some sampling constraints, one can obtain a sample only if y_i falls in the interval R_i. Hence, our samples consist of $\{y_i : y_i \in R_i\}$; nothing is available for $\{y_i : y_i \notin R_i\}$ and the number N. Standard statistics for the observable part $\{y_i : y_i \in R_i\}$, 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. Any proper estimation procedure needs to recover the original distribution f from the observable part.

Double-truncation includes the one-sided truncation (right- or left-truncation only) as a special case. Under left-truncation, one obtains the sample when y_i is large enough compared to the left-truncation limit u_i. In clinical survival analysis, left-truncation is also called 'delayed entry'; the age at disease onset y_i becomes available only if it exceeds the entry age u_i (Andersen and Keiding 2002). Under right-truncation, one obtains the sample when y_i is smaller than the right-truncation limit v_i. Right-truncated data is especially relevant to studies of AIDS (Lagakos et al. 1988; Strzalkowska-Kominiak and Stute 2013). However, double-truncation is essentially different from double-censoring (i.e., both left- and right-censorings) and interval censoring. Double-truncation yields inclusion/exclusion of samples while double-censoring and interval censoring produce incomplete lifetimes for the included samples (Commenges 2002).

Nonparametric methods for doubly truncated data have been studied well in the literature. Efron and Petrosian (1999) proposed the nonparametric maximum likelihood estimator (NPMLE) of the distribution function. Shen (2010) gave an alternative expression of the NPMLE based on inverse truncation probability weighting, and then derived the uniform consistency and weak convergence of the NPMLE. Moreira and de Uña-Álvarez (2010) developed interval estimation by bootstrapping the NPMLE. Emura et al. (2014) derived an explicit standard error estimator of the NPMLE, which is a computationally effective alternative to the bootstrap. Moreira and de Uña-Álvarez (2012) and Moreira and Van Keilegom (2013) considered a nonparametric methodology that gives a smooth kernel density estimator.

Compared to the nonparametric methods, research is much scarcer on parametric methods under double-truncation. Efron and Petrosian (1999) considered parametric maximum likelihood estimation under the so-called special exponential family (SEF). Stovring and Wang (2007) briefly indicated the use of parametric approaches for doubly-truncated data. Emura and Konno (2012a) also mentioned an application of their normal distribution approach under dependent double-truncation.

To the best of our knowledge, no simulation study has been conducted for the aforementioned parametric approaches, and other parametric methods have not been proposed. This paper fills this gap by developing technical details for implementing parametric inference under the SEF that is originally proposed by Efron and Petrosian

(1999). Due to double-truncation, this is not a straightforward application of the maximum likelihood inference. For instance, the likelihood of observed data depends on both the truncation interval and the support of the SEF. Such complicated likelihood functions also motivate reliable computational algorithms for finding the maximum likelihood estimator.

This paper is organized as follows. Section 2 includes the in-depth review of the special exponential family. Section 3 constructs the likelihood function and then introduces numerical algorithms such as fixed-point iteration and Newton–Raphson method. Section 4 performs simulations and Sect. 5 analyzes real data. Section 6 concludes the paper.

2 Special exponential family (SEF)

We revisit the SEF considered by Efron and Petrosian (1999) for fitting doubly-truncated data. Unlike their paper, we explicitly discuss the choices of the parameter space and support, which are necessary for the density to be well-defined.

We assume that the lifetime variable Y follows a continuous distribution with a density

$$f_{\boldsymbol{\eta}}(y) = \exp\left\{\boldsymbol{\eta}^{\mathrm{T}} \cdot \boldsymbol{t}(y) - \phi(\boldsymbol{\eta})\right\}, \quad y \in \mathcal{Y},$$

where $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_k)^{\mathrm{T}} \in \Theta$, $\boldsymbol{t}(y) = (y, y^2, \ldots, y^k)^{\mathrm{T}}$, $\mathcal{Y} \subset \Re$ is the support of Y, and $\Theta \subset \Re^{\mathrm{k}}$ is a parameter space. Here, $\phi(\eta)$ is chosen to make $\int_{\mathcal{Y}} f_\eta(y)\,dy = 1$. It follows that $\phi(\boldsymbol{\eta}) = \log[\int_{\mathcal{Y}} \exp\{\boldsymbol{\eta}^{\mathrm{T}} \cdot \boldsymbol{t}(y)\}\,dy]$, provided $\int_{\mathcal{Y}} \exp\{\boldsymbol{\eta}^{\mathrm{T}} \cdot \boldsymbol{t}(y)\}\,dy < \infty$. For this integral to be finite, both $\mathcal{Y}$ and Θ need to be chosen carefully.

2.1 Example 1: One-parameter SEF ($k = 1$)

A one-parameter SEF corresponds to $t(y) = y$, $\eta_1 = \eta$ and $\eta_2 = \cdots = \eta_k = 0$. The family yields two mutually exclusive cases; $\eta > 0$ and $\eta < 0$.

First, consider the case $\eta > 0$. The parameter space of η is $\Theta = \{\eta : \eta > 0\} = (0, \infty)$. If we let τ_2 be the upper support of Y, then

$$f_\eta(y) = \eta \exp\{\eta(y - \tau_2)\}, \qquad y \in \mathcal{Y} = (-\infty, \tau_2].$$

This is a well-defined density since $\phi(\eta) = \log\{\int_{\mathcal{Y}} \exp(\eta y)\} = -\log \eta + \eta\tau_2$ for all $\eta \in \Theta$. Figure 1 displays the density with $\eta = 1$ or 3. It shows that the upper support τ_2 is necessary such that $\int_{\mathcal{Y}} \exp(\eta y) < \infty$. The corresponding survival function is

$$S_\eta(y) = \int_y^{\tau_2} \eta \exp\{\eta(t - \tau_2)\}dt = 1 - \exp\{\eta(y - \tau_2)\}, \qquad y \leq \tau_2.$$

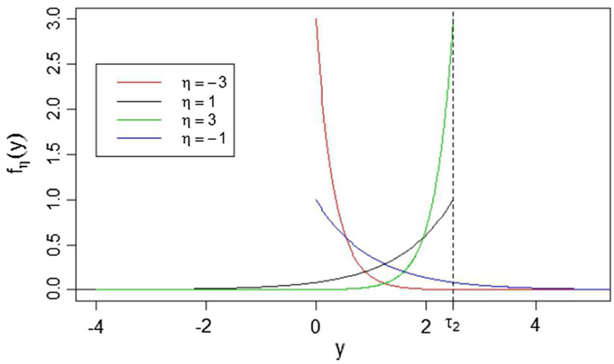

Fig. 1 The density $f_\eta(y)$ of the one-parameter SEF with parameter $\eta \in \{-3, -1, 1, 3\}$. For $\eta > 0$, we choose the support $\mathcal{Y} = (-\infty, \tau_2]$, where $\tau_2 = 2.5$. For $\eta < 0$, we choose the support $\mathcal{Y} = [\tau_1, \infty)$, where $\tau_1 = 0$

Next, consider the case $\eta < 0$. Accordingly, the parameter space of η is $\Theta = \{\eta : \eta < 0\} = (-\infty, 0)$. Then the density becomes

$$f_\eta(y) = -\eta \exp\{\eta(y - \tau_1)\}, \qquad y \in \mathcal{Y} = [\tau_1, \infty),$$

This distribution is a two-parameter exponential distribution if τ_1 is considered as unknown origin (Balakrishnan and Asit Basu 1996). The usual exponential distribution is given by $\tau_1 = 0$ with the density shown in Fig. 1. The corresponding survival function is

$$S_\eta(y) = \int_y^\infty -\eta \exp\{\eta(t - \tau_1)\}dt = \exp\{\eta(y - \tau_1)\}, \qquad y \geq \tau_1.$$

2.2 Example 2: Two-parameter SEF ($k = 2$)

First, consider the case $\eta_2 < 0$. A two-parameter SEF is obtained by setting $\boldsymbol{t}(y) = (y, y^2)^{\mathrm{T}}$, $\boldsymbol{\eta} = (\eta_1, \eta_2)^{\mathrm{T}}$ and $\eta_3 = \cdots = \eta_k = 0$. With $\mu = -\eta_1/2\eta_2$ and $\sigma^2 = -1/2\eta_2$, a normal distribution is obtained (see also Castillo 1994). Then, the density of Y is

$$f_\eta(y) = \exp\left\{\eta_1 y + \eta_2 y^2 + \frac{\eta_1^2}{4\eta_2} - \log\left(\sqrt{\frac{-\pi}{\eta_2}}\right)\right\}, \qquad y \in \mathcal{Y} = (-\infty, \infty).$$

For $\phi(\boldsymbol{\eta}) = \log\{\int_{\mathcal{Y}} \exp(\eta_1 y + \eta_2 y^2)\, dy\} = \log(\sqrt{-\pi/\eta_2}) - \eta_1^2/4\eta_2$ to be well-defined, the parameter space needs to be $\Theta = \{(\eta_1, \eta_2) : \eta_1 \in \Re, \eta_2 < 0\}$. The survival function is

$$S_{\boldsymbol{\eta}}(y) = \int_y^\infty \exp\left\{\eta_1 t + \eta_2 t^2 + \frac{\eta_1^2}{4\eta_2} - \log\left(\sqrt{\frac{-\pi}{\eta_2}}\right)\right\}dt$$

$$= 1 - \Phi\left(\frac{y + \frac{\eta_1}{2\eta_2}}{\sqrt{\frac{-1}{2\eta_2}}}\right), \quad y \in \Re.$$

The case $\eta_2 > 0$ seems less useful in practice. In such a case, the density function is convex, and so the support must be bounded from both below and above. Since such a distribution makes inference complicated and is less practical, we do not discuss this case in this paper.

2.3 Example 3: Cubic SEF ($k = 3$)

A cubic SEF is obtained by setting $\boldsymbol{t}(y) = (y, y^2, y^3)^{\mathrm{T}}$, $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3)^{\mathrm{T}}$ and $\eta_4 = \cdots = \eta_k = 0$. The density of Y can be expressed as

$$f_{\boldsymbol{\eta}}(y) = \exp\left[\eta_1 y + \eta_2 y^2 + \eta_3 y^3 - \phi(\boldsymbol{\eta})\right], \qquad y \in \mathcal{Y},$$

where $\phi(\boldsymbol{\eta}) = \log\{\int_{\mathcal{Y}} \exp(\eta_1 y + \eta_2 y^2 + \eta_3 y^3)\, dy\}$. For $f_{\boldsymbol{\eta}}(y)$ to be well-defined, it is necessary that $\int_{\mathcal{Y}} \exp(\eta_1 y + \eta_2 y^2 + \eta_3 y^3)\, dy < \infty$. To make the integral finite, we consider the parameter space as two mutually exclusive cases with different ranges of Y.

First, if the parameter space is $\Theta = \{(\eta_1, \eta_2, \eta_3) : \eta_1 \in \Re, \eta_2 \in \Re, \eta_3 > 0\}$, then we set the range of Y as $\mathcal{Y} = (-\infty, \tau_2]$, where τ_2 is the upper support of Y. Figure 2a displays $f_{\boldsymbol{\eta}}(y)$ under $\eta_1 = 5$, $\eta_2 = -0.5$ and $\eta_3 \in \{0, 0.005, 0.01, 0.015\}$. Clearly, one needs to set τ_2 such that $\int_{\mathcal{Y}} \exp(\eta_1 y + \eta_2 y^2 + \eta_3 y^3)\, dy < \infty$. The corresponding survival function is

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

Second, if the parameter space is $\Theta = \{(\eta_1, \eta_2, \eta_3) : \eta_1 \in \Re, \eta_2 \in \Re, \eta_3 < 0\}$, then we set the range of Y as $\mathcal{Y} = [\tau_1, \infty)$, where τ_1 is the lower support of Y. Figure 2b shows that one needs to set the lower bound τ_1 such that $\int_{\mathcal{Y}} \exp(\eta_1 y + \eta_2 y^2 + \eta_3 y^3)\, dy < \infty$. The corresponding survival function is

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

Remark 1 The cubic SEF yields a skewed truncated normal distribution, where η_3 determines the degree of skewness. The skew normal density is often fitted for biomedical applications. For instance, Robertson and Allison (2012) fitted the US life table with a negatively skewed distribution, which has an upper bound like τ_2 and hence similar to the cubic SEF with $\eta_3 > 0$. See Mandrekar and Nandrekar (2003) for other skewed data in biomedical studies.

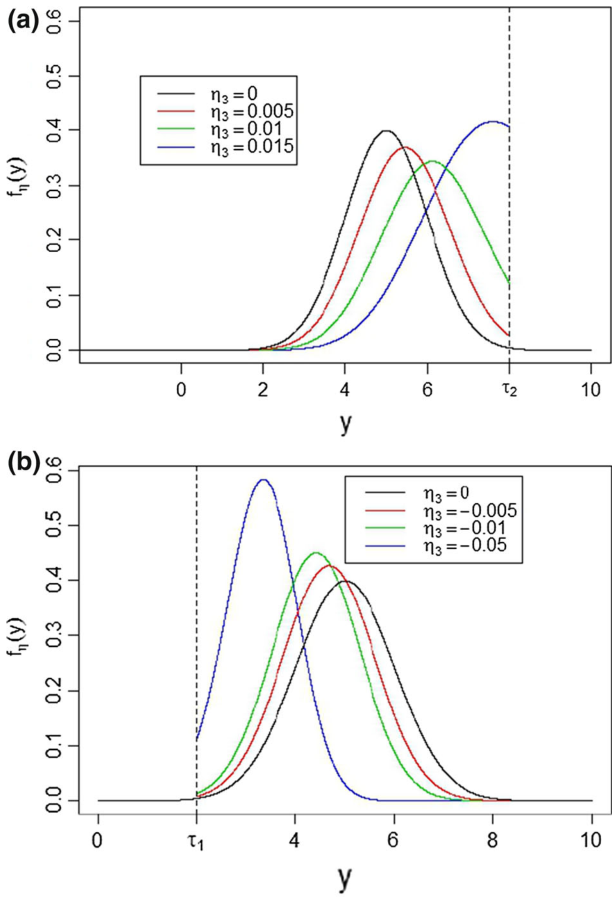

Fig. 2 **a** The density $f_{\boldsymbol{\eta}}(y)$ of cubic SEF with parameters (η_1, η_2, η_3). We set $\eta_1 = 5$, $\eta_2 = -0.5$, $\eta_3 \in \{0, 0.005, 0.01, 0.015\}$, and the support $(-\infty, \tau_2]$, where $\tau_2 = 8$. **b** The density $f_{\boldsymbol{\eta}}(y)$ of cubic SEF with parameters (η_1, η_2, η_3). We set $\eta_1 = 5$, $\eta_2 = -0.5$, $\eta_3 \in \{0, -0.005, -0.01, -0.05\}$, and the support $[\tau_1, \infty)$, where $\tau_1 = 2$

3 Method of estimation

3.1 Likelihood functions

We introduce the likelihood function with doubly truncated data originally proposed by Efron and Petrosian (1999). Especially, we give explicit formulas for the likelihood functions that are not available in Efron and Petrosian (1999).

Let $R_i = [u_i, v_i]$ be the truncation interval, where u_i is the left-truncation limit and v_i is the right-truncation limit. We consider estimation 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_i, subject to $y_i \in R_i$, is

$$f_{\boldsymbol{\eta}}(y_i \,|Y \in R_i\,) = \begin{cases} f_{\boldsymbol{\eta}}(y_i)/F_i(\boldsymbol{\eta}) & \text{if } y_i \in R_i, \\ 0 & \text{if } y_i \notin R_i, \end{cases}$$

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

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

3.1.1 Example 1: One-parameter SEF ($k = 1$)

As in Sect. 2, we consider two mutually exclusive cases: $\eta > 0$ and $\eta < 0$.

First, consider the case $\eta > 0$. As discussed before, we need to set the upper support τ_2 of Y. Whether the right-truncation limit v_i precedes the support τ_2 influences the likelihood for y_i. Define the indicator

$$\delta_i = \begin{cases} 1 & \text{if } \; v_i < \tau_2, \\ 0 & \text{if } \; v_i \geq \tau_2. \end{cases}$$

Then, the log-likelihood function is given by

$$\begin{aligned} \ell(\eta) = \log L(\eta) = n \log \eta + \eta \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \delta_i[\log\{\exp(\eta v_i) - \exp(\eta u_i)\}] \\ - \sum_{i=1}^{n}(1-\delta_i)[\log\{\exp(\eta\tau_2) - \exp(\eta u_i)\}]. \end{aligned} \tag{1a}$$

Next, consider the case $\eta < 0$. In this case, one needs to set the lower support τ_1 of Y. Whether the left-truncation limit u_i exceeds the lower support influences the likelihood for y_i. Define the indicator

$$\delta_i = \begin{cases} 1 & \text{if } \; u_i \geq \tau_1, \\ 0 & \text{if } \; u_i < \tau_1. \end{cases}$$

Then, the log-likelihood function is given by

$$\begin{aligned} \ell(\eta) = \log L(\eta) = n \log(-\eta) + \eta \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \delta_i[\log\{-\exp(\eta v_i) + \exp(\eta u_i)\}] \\ - \sum_{i=1}^{n} (1-\delta_i)[\log\{-\exp(\eta v_i) + \exp(\eta\tau_1)\}]. \end{aligned} \tag{1b}$$

3.1.2 Example 2: Two-parameter SEF ($k = 2$)

Let $\Phi(.)$ denote the cumulative distribution function of the standard normal distribution. Then, the log-likelihood function is given by

$$\ell(\boldsymbol{\eta}) = \log L(\boldsymbol{\eta}) = \sum_{i=1}^{n} \eta_1 y_i + \sum_{i=1}^{n} \eta_2 y_i^2 + \frac{n\eta_1^2}{4\eta_2} + \frac{n}{2}\log(-\eta_2) - \frac{n}{2}\log(\pi)$$
$$- \sum_{i=1}^{n} \log\left\{ \Phi\left(\frac{v_i + \frac{\eta_1}{2\eta_2}}{\sqrt{\frac{-1}{2\eta_2}}} \right) - \Phi\left(\frac{u_i + \frac{\eta_1}{2\eta_2}}{\sqrt{\frac{-1}{2\eta_2}}} \right) \right\}. \tag{2}$$

3.1.3 Example 3: Cubic SEF ($k = 3$)

Consider the case $\eta_3 < 0$ with the parameter space $\Theta = \{(\eta_1, \eta_2, \eta_3) : \eta_1 \in \Re, \eta_2 \in \Re, \eta_3 < 0\}$. Then we set the lower support of Y as τ_1. Whether the left-truncation limit u_i exceeds the lower support influences the likelihood for y_i. Define the indicator

$$\delta_i = \begin{cases} 1 & \text{if } u_i \geq \tau_1, \\ 0 & \text{if } u_i < \tau_1. \end{cases}$$

Then, the log-likelihood function is given by

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

The case of $\eta_3 > 0$ is similar.

To assure some numerical stability, we consider some approximate likelihood derived under conditions: $\tau_1 \leq \min_i(u_i)$ when $\eta_3 < 0$ and $\tau_2 \geq \max_i(v_i)$ when $\eta_3 > 0$. With these conditions, both the left-truncation limit u_i and right-truncation limit v_i are within the support. Then, the approximate log-likelihood function is given by

$$\ell(\boldsymbol{\eta}) = \log L(\boldsymbol{\eta}) = \sum_{i=1}^{n} \left(\eta_1 y_i + \eta_2 y_i^2 + \eta_3 y_i^3 \right)$$
$$- \sum_{i=1}^{n} \log\left\{ \int_{u_i}^{v_i} \exp\left(\eta_1 y + \eta_2 y^2 + \eta_3 y^3 \right) dy \right\}. \tag{3}$$

The other advantage of the approximate likelihood is that one can avoid the prior choice of $\eta_3 > 0$ or $\eta_3 < 0$.

Remark 2 If the imposed conditions do not satisfy, Eq. (3) is an approximation to the exact log-likelihood. The data analysis of Sect. 5 will use the exact log-likelihood. However, the development of the algorithms in Sect. 3.2 and the simulations of Sect. 4 will be done on the approximate log-likelihood in Eq. (3), which gives more stable results over a large number of repetitions.

3.2 Newton–Raphson method

Efron and Petrosian (1999) suggested using the Newton–Raphson algorithm to obtain the maximum likelihood estimator (MLE). In this section, we consider the details about the Newton–Raphson algorithm, including the first- and second-order derivatives of the log-likelihood with respect to the parameters. We also propose a modification of the Newton–Raphson for the three-parameter case.

3.2.1 Example 1: One-parameter SEF

For the case $\eta > 0$, the first- and second-order derivatives of the log-likelihood are

$$\begin{aligned}\frac{\partial}{\partial\eta}\ell(\eta) &= \frac{n}{\eta} + \sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\delta_i\left\{\frac{v_i\exp(\eta v_i) - u_i\exp(\eta u_i)}{\exp(\eta v_i) - \exp(\eta u_i)}\right\}\\ &\quad - \sum_{i=1}^{n}(1-\delta_i)\left\{\frac{\tau_2\exp(\eta\tau_2) - u_i\exp(\eta u_i)}{\exp(\eta\tau_2) - \exp(\eta u_i)}\right\},\\ \frac{\partial^2}{\partial\eta^2}\ell(\eta) &= \frac{-n}{\eta^2} - \sum_{i=1}^{n}\delta_i\left[\frac{v_i^2\exp(\eta v_i) - u_i^2\exp(\eta u_i)}{\exp(\eta v_i) - \exp(\eta u_i)}\right.\\ &\quad \left. - \left\{\frac{v_i\exp(\eta v_i) - u_i\exp(\eta u_i)}{\exp(\eta v_i) - \exp(\eta u_i)}\right\}^2\right]\\ &\quad - \sum_{i=1}^{n}(1-\delta_i)\left[\frac{\tau^2\exp(\eta\tau_2) - u_i^2\exp(\eta u_i)}{\exp(\eta\tau_2) - \exp(\eta u_i)} - \left\{\frac{\tau_2\exp(\eta\tau_2) - u_i\exp(\eta u_i)}{\exp(\eta\tau_2) - \exp(\eta u_i)}\right\}^2\right].\end{aligned}$$

Similarly, for the case $\eta < 0$, one has

$$\begin{aligned}\frac{\partial}{\partial\eta}\ell(\eta) &= \frac{n}{\eta} + \sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\delta_i\left\{\frac{-v_i\exp(\eta v_i) + u_i\exp(\eta u_i)}{-\exp(\eta v_i) + \exp(\eta u_i)}\right\}\\ &\quad - \sum_{i=1}^{n}(1-\delta_i)\left\{\frac{-v_i\exp(\eta v_i) - \tau_1\exp(\eta\tau_1)}{-\exp(\eta v_i) + \exp(\eta\tau_1)}\right\},\\ \frac{\partial^2}{\partial\eta^2}\ell(\eta) &= \frac{-n}{\eta^2} - \sum_{i=1}^{n}\delta_i\left[\frac{-v_i^2\exp(\eta v_i) + u_i^2\exp(\eta u_i)}{-\exp(\eta v_i) + \exp(\eta u_i)}\right.\end{aligned}$$

$$
\left.-\left\{\frac{-v_i\exp(\eta v_i)+u_i\exp(\eta u_i)}{-\exp(\eta v_i)+\exp(\eta u_i)}\right\}^2\right]
$$

$$
-\sum_{i=1}^{n}(1-\delta_i)\left[\frac{-v_i^2\exp(\eta v_i)+\tau^2\exp(\eta\tau_1)}{-\exp(\eta v_i)+\exp(\eta\tau_1)}-\left\{\frac{-v_i\exp(\eta v_i)+\tau_1\exp(\eta\tau_1)}{-\exp(\eta v_i)+\exp(\eta\tau_1)}\right\}^2\right].
$$

Now, for both cases, we can obtain the MLE of η by the Newton–Raphson method. At the $(k+1)$th step of the iteration, the updated parameters are obtained as

$$
\eta^{(k+1)}=\eta^{(k)}-S(\eta^{(k)})/\{\partial S(\eta^{(k)})/\partial\eta\},
$$

for $k=0,1,2,\ldots$, where $S(\eta)=\partial\ell(\eta)/\partial\eta$ is the score function. The iteration procedure then continues until convergence, i.e., until $|\eta^{(k+1)}-\eta^{(k)}|<\varepsilon$ for some small $\varepsilon>0$.

3.2.2 Example 2: Two-parameter SEF

Let $\varphi(.)$ and $\Phi(.)$ denote the probability density function and cumulative distribution function of the standard normal distribution. Following Cohen (1991, pp. 32) we let

$$
H_{i1}(\eta_1,\eta_2)=\frac{\varphi\left(\frac{v_i+\frac{\eta_1}{2\eta_2}}{\sqrt{\frac{-1}{2\eta_2}}}\right)}{\Phi\left(\frac{v_i+\frac{\eta_1}{2\eta_2}}{\sqrt{\frac{-1}{2\eta_2}}}\right)-\Phi\left(\frac{u_i+\frac{\eta_1}{2\eta_2}}{\sqrt{\frac{-1}{2\eta_2}}}\right)},
$$

$$
H_{i2}(\eta_1,\eta_2)=\frac{\varphi\left(\frac{u_i+\frac{\eta_1}{2\eta_2}}{\sqrt{\frac{-1}{2\eta_2}}}\right)}{\Phi\left(\frac{v_i+\frac{\eta_1}{2\eta_2}}{\sqrt{\frac{-1}{2\eta_2}}}\right)-\Phi\left(\frac{u_i+\frac{\eta_1}{2\eta_2}}{\sqrt{\frac{-1}{2\eta_2}}}\right)},
$$

which are recognized as the hazard function of the normal distribution with doubly truncated samples (Sankaran and Sunoj 2004).

The first derivatives of the log-likelihood are

$$
\frac{\partial}{\partial\eta_1}\ell(\boldsymbol{\eta})=\sum_{i=1}^{n}y_i+\frac{n\eta_1}{2\eta_2}+\frac{1}{\sqrt{-2\eta_2}}\left\{\sum_{i=1}^{n}H_{i1}(\eta_1,\eta_2)-\sum_{i=1}^{n}H_{i2}(\eta_1,\eta_2)\right\},
$$

$$
\frac{\partial}{\partial\eta_2}\ell(\boldsymbol{\eta})=\sum_{i=1}^{n}y_i^2-\frac{n\eta_1^2}{4\eta_2^2}+\frac{n}{2\eta_2}-\sum_{i=1}^{n}\left\{H_{i1}(\eta_1,\eta_2)\cdot\left(\frac{-v_i}{\sqrt{-2\eta_2}}-\eta_1\frac{\sqrt{-2\eta_2}}{4\eta_2^2}\right)\right\}
$$

$$
+\sum_{i=1}^{n}\left\{H_{i2}(\eta_1,\eta_2)\cdot\left(\frac{-u_i}{\sqrt{-2\eta_2}}-\eta_1\frac{\sqrt{-2\eta_2}}{4\eta_2^2}\right)\right\}.
$$

The second-order derivatives of the log-likelihood are

$$
\begin{aligned}
\frac{\partial^2}{\partial \eta_1^2}\ell(\boldsymbol{\eta}) &= \frac{n}{2\eta_2} \\
&\quad - \frac{1}{2\eta_2}\sum_{i=1}^{n}\left\{ H_{i1}(\eta_1,\eta_2)\left(\frac{v_i+\frac{\eta_1}{2\eta_2}}{\sqrt{\frac{-1}{2\eta_2}}}\right) - H_{i2}(\eta_1,\eta_2)\left(\frac{u_i+\frac{\eta_1}{2\eta_2}}{\sqrt{\frac{-1}{2\eta_2}}}\right)\right\} \\
&\quad - \frac{1}{2\eta_2}\sum_{i=1}^{n}\{H_{i1}(\eta_1,\eta_2)-H_{i2}(\eta_1,\eta_2)\}^2, \\
\frac{\partial^2}{\partial \eta_2^2}\ell(\boldsymbol{\eta}) &= \sum_{i=1}^{n}\left\{ H_{i1}(\eta_1,\eta_2)\left(\frac{-v_i}{\sqrt{-2\eta_2}} - \eta_1\frac{\sqrt{-2\eta_2}}{4\eta_2^2}\right)\right. \\
&\quad \left. -\ H_{i2}(\eta_1,\eta_2)\left(\frac{-u_i}{\sqrt{-2\eta_2}} - \eta_1\frac{\sqrt{-2\eta_2}}{4\eta_2^2}\right)\right\}^2 \\
&\quad + \sum_{i=1}^{n} H_{i1}(\eta_1,\eta_2)\left\{\left(\frac{v_i+\frac{\eta_1}{2\eta_2}}{\sqrt{\frac{-1}{2\eta_2}}}\right)\left(\frac{-v_i}{\sqrt{-2\eta_2}} - \eta_1\frac{\sqrt{-2\eta_2}}{4\eta_2^2}\right)^2\right. \\
&\quad \left. -\left(\frac{-v_i}{\sqrt{(-2\eta_2)^3}} - \frac{3\eta_1}{\sqrt{(-2\eta_2)^5}}\right)\right\} \\
&\quad - \sum_{i=1}^{n} H_{i2}(\eta_1,\eta_2)\left\{\left(\frac{u_i+\frac{\eta_1}{2\eta_2}}{\sqrt{\frac{-1}{2\eta_2}}}\right)\left(\frac{-u_i}{\sqrt{-2\eta_2}} - \eta_1\frac{\sqrt{-2\eta_2}}{4\eta_2^2}\right)^2\right. \\
&\quad \left. -\left(\frac{-u_i}{\sqrt{(-2\eta_2)^3}} - \frac{3\eta_1}{\sqrt{(-2\eta_2)^5}}\right)\right\} + \frac{n\eta_1^2}{2\eta_2^3} - \frac{n}{2\eta_2^2}, \\
\frac{\partial^2}{\partial \eta_1 \partial \eta_2}\ell(\boldsymbol{\eta}) &= -\frac{n\eta_1}{2\eta_2^2} + (-2\eta_2)^{\frac{-3}{2}}\sum_{i=1}^{n}\{H_{i1}(\eta_1,\eta_2)-H_{i2}(\eta_1,\eta_2)\} \\
&\quad - \frac{1}{\sqrt{-2\eta_2}}\sum_{i=1}^{n}\left[H_{i1}(\eta_1,\eta_2)\left(\frac{v_i+\frac{\eta_1}{2\eta_2}}{\sqrt{\frac{-1}{2\eta_2}}}\right)\left\{\frac{-v_i}{\sqrt{-2\eta_2}} - (-2\eta_2)^{\frac{-3}{2}}\eta_1\right\}\right] \\
&\quad + \frac{1}{\sqrt{-2\eta_2}}\sum_{i=1}^{n}\left[H_{i2}(\eta_1,\eta_2)\left(\frac{u_i+\frac{\eta_1}{2\eta_2}}{\sqrt{\frac{-1}{2\eta_2}}}\right)\left\{\frac{-u_i}{\sqrt{-2\eta_2}} - (-2\eta_2)^{\frac{-3}{2}}\eta_1\right\}\right] \\
&\quad - \frac{1}{\sqrt{-2\eta_2}}\sum_{i=1}^{n}\left[\{H_{i1}(\eta_1,\eta_2)-H_{i2}(\eta_1,\eta_2)\}\right. \\
&\quad \left. \times \left[\begin{array}{l} H_{i1}(\eta_1,\eta_2)\left\{\frac{-v_i}{\sqrt{-2\eta_2}} - (-2\eta_2)^{\frac{-3}{2}}\eta_1\right\} \\ -H_{i2}(\eta_1,\eta_2)\left\{\frac{-u_i}{\sqrt{-2\eta_2}} - (-2\eta_2)^{\frac{-3}{2}}\eta_1\right\}\end{array}\right]\right].
\end{aligned}
$$

Now, one can obtain the MLE by the two-dimensional Newton–Raphson method. For $k = 0, 1, 2, \ldots$, the $(k+1)$th step of the iteration is

$$\begin{bmatrix} \eta_1^{(k+1)} \\ \eta_2^{(k+1)} \end{bmatrix} = \begin{bmatrix} \eta_1^{(k)} \\ \eta_2^{(k)} \end{bmatrix} - J_S^{-1}(\eta_1^{(k)}, \eta_2^{(k)}) \begin{bmatrix} S_1(\eta_1^{(k)}, \eta_2^{(k)}) \\ S_2(\eta_1^{(k)}, \eta_2^{(k)}) \end{bmatrix},$$

where $S_1(\eta_1, \eta_2) = \partial\ell(\eta_1, \eta_2)/\partial\eta_1$, $S_2(\eta_1, \eta_2) = \partial\ell(\eta_1, \eta_2)/\partial\eta_2$ and

$$J_S(\eta_1, \eta_2) = \begin{bmatrix} \frac{\partial^2}{\partial\eta_1^2}\ell(\eta_1, \eta_2) & \frac{\partial^2}{\partial\eta_1\partial\eta_2}\ell(\eta_1, \eta_2) \\ \frac{\partial^2}{\partial\eta_1\partial\eta_2}\ell(\eta_1, \eta_2) & \frac{\partial^2}{\partial\eta_2^2}\ell(\eta_1, \eta_2) \end{bmatrix}$$

is the Hessian matrix. The iteration procedure then continues until convergence, i.e., until $|\eta_j^{(k+1)} - \eta_j^{(k)}| < \varepsilon$ for $j = 1, 2$ and for some small $\varepsilon > 0$.

3.2.3 Example 3: Cubic SEF

For $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3)^{\mathrm{T}}$, we define

$$E_i^k(\boldsymbol{\eta}) = \int_{u_i}^{v_i} y^k \exp\left(\eta_1 y + \eta_2 y^2 + \eta_3 y^3\right) dy, \quad k = 0, 1, \ldots, 6.$$

Since $E_i^k(\boldsymbol{\eta})$ has no closed form, one need to use some routine for numerical integration, such as R **integrate** (R Development Core Team 2014). The first-order derivatives of the log-likelihood are

$$\frac{\partial}{\partial\boldsymbol{\eta}_k}\ell(\boldsymbol{\eta}) = \sum_{i=1}^{n}\left\{y_i^k - E_i^k(\boldsymbol{\eta})/E_i^0(\boldsymbol{\eta})\right\}, \quad k = 1, 2, 3.$$

The second-order derivatives of the log-likelihood are obtained as

$$\frac{\partial^2}{\partial\eta_1^2}\ell(\boldsymbol{\eta}) = \sum_{i=1}^{n}\left[-E_i^2(\boldsymbol{\eta})/E_i^0(\boldsymbol{\eta}) + \{E_i^1(\boldsymbol{\eta})/E_i^0(\boldsymbol{\eta})\}^2\right],$$

$$\frac{\partial^2}{\partial\eta_2^2}\ell(\boldsymbol{\eta}) = \sum_{i=1}^{n}\left[-E_i^4(\boldsymbol{\eta})/E_i^0(\boldsymbol{\eta}) + \{E_i^2(\boldsymbol{\eta})/E_i^0(\boldsymbol{\eta})\}^2\right],$$

$$\frac{\partial^2}{\partial\eta_3^2}\ell(\boldsymbol{\eta}) = \sum_{i=1}^{n}\left[-E_i^6(\boldsymbol{\eta})/E_i^0(\boldsymbol{\eta}) + \{E_i^3(\boldsymbol{\eta})/E_i^0(\boldsymbol{\eta})\}^2\right],$$

$$\frac{\partial^2}{\partial\eta_2\partial\eta_1}\ell(\boldsymbol{\eta}) = \sum_{i=1}^{n}\left[-E_i^3(\boldsymbol{\eta})/E_i^0(\boldsymbol{\eta}) + \{E_i^2(\boldsymbol{\eta})/E_i^0(\boldsymbol{\eta})\}\{E_i^1(\boldsymbol{\eta})/E_i^0(\boldsymbol{\eta})\}\right],$$

$$\frac{\partial^2}{\partial \eta_3 \partial \eta_1} \ell(\boldsymbol{\eta}) = \sum_{i=1}^{n} \left[-E_i^4(\boldsymbol{\eta})/E_i^0(\boldsymbol{\eta}) + \{E_i^3(\boldsymbol{\eta})/E_i^0(\boldsymbol{\eta})\}\{E_i^1(\boldsymbol{\eta})/E_i^0(\boldsymbol{\eta})\} \right],$$

$$\frac{\partial^2}{\partial \eta_3 \partial \eta_2} \ell(\boldsymbol{\eta}) = \sum_{i=1}^{n} \left[-E_i^5(\boldsymbol{\eta})/E_i^0(\boldsymbol{\eta}) + \{E_i^3(\boldsymbol{\eta})/E_i^0(\boldsymbol{\eta})\}\{E_i^2(\boldsymbol{\eta})/E_i^0(\boldsymbol{\eta})\} \right].$$

Now, one can obtain the MLE by a three-dimensional Newton–Raphson method. At the $(k+1)$th step of the iteration, the updated parameters are obtained as

$$\begin{bmatrix} \eta_1^{(k+1)} \\ \eta_2^{(k+1)} \\ \eta_3^{(k+1)} \end{bmatrix} = \begin{bmatrix} \eta_1^{(k)} \\ \eta_2^{(k)} \\ \eta_3^{(k)} \end{bmatrix} - J_S^{-1}\left(\eta_1^{(k)}, \eta_2^{(k)}, \eta_3^{(k)}\right) \begin{bmatrix} S_1(\eta_1^{(k)}, \eta_2^{(k)}, \eta_3^{(k)}) \\ S_2(\eta_1^{(k)}, \eta_2^{(k)}, \eta_3^{(k)}) \\ S_3(\eta_1^{(k)}, \eta_2^{(k)}, \eta_3^{(k)}) \end{bmatrix}, \tag{4}$$

for $k = 0, 1, 2, \ldots$, and $S_j(\eta_1, \eta_2, \eta_3) = \partial \ell(\eta_1, \eta_2, \eta_3)/\partial \eta_j, \quad j = 1, 2, 3$, and

$$J_S(\boldsymbol{\eta}) = \begin{bmatrix} \frac{\partial^2}{\partial \eta_1^2} \ell(\boldsymbol{\eta}) & \frac{\partial^2}{\partial \eta_1 \partial \eta_2} \ell(\boldsymbol{\eta}) & \frac{\partial^2}{\partial \eta_1 \partial \eta_3} \ell(\boldsymbol{\eta}) \\ \frac{\partial^2}{\partial \eta_2 \partial \eta_1} \ell(\boldsymbol{\eta}) & \frac{\partial^2}{\partial \eta_2^2} \ell(\boldsymbol{\eta}) & \frac{\partial^2}{\partial \eta_2 \partial \eta_3} \ell(\boldsymbol{\eta}) \\ \frac{\partial^2}{\partial \eta_3 \partial \eta_1} \ell(\boldsymbol{\eta}) & \frac{\partial^2}{\partial \eta_3 \partial \eta_2} \ell(\boldsymbol{\eta}) & \frac{\partial^2}{\partial \eta_3^2} \ell(\boldsymbol{\eta}) \end{bmatrix}$$

The iteration continues until convergence, i.e., until $|\eta_j^{(k+1)} - \eta_j^{(k)}| < \varepsilon$ for $j = 1, 2, 3$ and for some small $\varepsilon > 0$.

It is well-known that the Newton–Raphson algorithm is sensitive to the initial values, especially in estimating three or more parameters (Knight 2000). In our simulations, the algorithm occasionally diverges when we choose initial values based on samples. To stabilize the algorithm, we propose a modified version for Newton–Raphson method, where the initial values are randomized. Such a randomization scheme is previously used to stabilize a three-dimensional Newton–Raphson method in some other context (Long and Emura 2014). The proposed algorithm is stated as follows.

3.2.4 Randomized Newton–Raphson (RNR) algorithm

Let D_1, D_2, D_3, d_1 and d_2 be some positive tuning parameters.

1. Choose the initial value $\boldsymbol{\eta} = (\eta_1^{(0)}, \eta_2^{(0)}, \eta_3^{(0)})^{\mathrm{T}}$.
2. For $k = 0, 1, 2, \ldots$, continue the iteration of Eq. (4) until convergence, i.e., until $|\eta_j^{(k+1)} - \eta_j^{(k)}| < \varepsilon$ for $j = 1, 2, 3$ and for some small $\varepsilon > 0$. Then $(\eta_1^{(k+1)}, \eta_2^{(k+1)}, \eta_3^{(k+1)})$ is the MLE.
3. If $|\eta_1^{(k+1)} - \eta_1^{(k)}| > D_1$,$|\eta_2^{(k+1)} - \eta_2^{(k)}| > D_2$ or $|\eta_3^{(k+1)} - \eta_3^{(k)}| > D_3$ occurs during the iteration, stop the algorithm. Replace $(\eta_1^{(0)}, \eta_2^{(0)}, \eta_3^{(0)})^{\mathrm{T}}$ with $(\eta_1^{(0)} + u_1, \eta_2^{(0)} + u_2, \eta_3^{(0)})^{\mathrm{T}}$, where $u_1 \sim U(-d_1, d_1)$ and $u_2 \sim U(-d_2, d_2)$, and then return to Step 1.

3.3 Fixed-point iteration method

In order to obtain the MLE of parameters, one can use the fixed-point iteration method (Burden and Faires 2011), which is a very simple iteration algorithm. The algorithm is performed as follows:

(i) One dimensional fixed-point iteration (Burden and Faires 2011, pp. 56–64)
To solve the score function $S(\eta) = 0$, we rewrite $S(\eta) = 0$ as $\eta = g(\eta)$. Then,

1. Choose the initial value $\eta^{(0)}$.
2. Perform the recursive process $\eta^{(k+1)} = g(\eta^{(k)})$.

The iteration procedure continues until convergence, i.e., until $|\eta^{(k+1)} - \eta^{(k)}| < \varepsilon$ for $k = 0, 1, 2, \ldots$ and small $\varepsilon > 0$. Finally, $\eta^{(k+1)}$ is the solution.

(ii) Two dimensional fixed-point iteration (Burden and Faires 2011, pp. 630–638)
To solve the score functions $S_1(\eta_1, \eta_2) = 0$ and $S_2(\eta_1, \eta_2) = 0$, we rewrite them as $\eta_1 = g(\eta_1, \eta_2)$and $\eta_2 = p(\eta_1, \eta_2)$. Then,

1. Choose the initial values $\eta_1^{(0)}$ and $\eta_2^{(0)}$.
2. Perform the recursive process $\eta_1^{(k+1)} = g(\eta_1^{(k)}, \eta_2^{(k)})$ and $\eta_2^{(k+1)} = p(\eta_1^{(k)}, \eta_2^{(k)})$.

The iteration continues until convergence, e.g. $|\eta_1^{(k+1)} - \eta_1^{(k)}| < \varepsilon$ and $|\eta_2^{(k+1)} - \eta_2^{(k)}| < \varepsilon$ for small $\varepsilon > 0$. Finally, $\eta_1^{(k+1)}$ and $\eta_2^{(k+1)}$ are the solutions.

The advantage of the fixed-point iteration is that it does not require the second derivatives of the log-likelihood which are often complicated (e.g., Example 2 of Sect. 3.2). The disadvantage is that the choice of g is not unique (Burden and Faires 2011, pp. 61). Fortunately, a good choice of g is available in one- and two-parameter SEFs. There are many papers implicitly applying the fixed-point algorithm for finding the MLE (e.g., Chen 2009).

3.3.1 Example 1: One-parameter SEF

First, consider the case $\eta > 0$. From the first-order derivative of the log-likelihood, the $(k + 1)$th iteration step of the fixed-point iteration is

$$\frac{1}{\eta^{(k+1)}} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\left\{\frac{v_i\exp(\eta^{(k)}v_i) - u_i\exp(\eta^{(k)}u_i)}{\exp(\eta^{(k)}v_i) - \exp(\eta^{(k)}u_i)}\right\} + \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\left\{\frac{\tau_2\exp(\eta^{(k)}\tau_2) - u_i\exp(\eta^{(k)}u_i)}{\exp(\eta^{(k)}\tau_2) - \exp(\eta^{(k)}u_i)}\right\} - \frac{1}{n}\sum_{i=1}^{n}y_i.$$

The iteration continues until convergence, i.e., until $|\eta^{(k+1)} - \eta^{(k)}| < \varepsilon$ for small $\varepsilon > 0$.

Similarly, for the case $\eta < 0$, one can obtain the MLE by the iteration

$$\frac{1}{\eta^{(k+1)}} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\left\{\frac{-v_i\exp(\eta^{(k)}v_i)+u_i\exp(\eta^{(k)}u_i)}{-\exp(\eta^{(k)}v_i)+\exp(\eta^{(k)}u_i)}\right\}$$
$$+\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\left\{\frac{-v_i\exp(\eta^{(k)}v_i)-\tau_1\exp(\eta^{(k)}\tau_1)}{-\exp(\eta^{(k)}v_i)+\exp(\eta^{(k)}\tau_1)}\right\}-\frac{1}{n}\sum_{i=1}^{n}y_i.$$

The iteration continues until convergence, i.e., until $|\eta^{(k+1)} - \eta^{(k)}| < \varepsilon$ for small $\varepsilon > 0$.

3.3.2 Example 2: Two-parameter SEF

We rewrite the first-order derivatives of the log-likelihood as

$$-\frac{\eta_1}{2\eta_2} = \overline{y} + \frac{1}{n\sqrt{-2\eta_2}}\sum_{i=1}^{n}\{H_{i1}(\eta_1,\eta_2)-H_{i2}(\eta_1,\eta_2)\},$$
$$-\frac{1}{2\eta_2} = \overline{y^2} - \frac{\eta_1^2}{4\eta_2^2} - \frac{1}{n}\sum_{i=1}^{n}\left\{H_{i1}(\eta_1,\eta_2)\left(\frac{-v_i}{\sqrt{-2\eta_2}}-\eta_1\frac{\sqrt{-2\eta_2}}{4\eta_2^2}\right)\right\}$$
$$+\frac{1}{n}\sum_{i=1}^{n}\left\{H_{i2}(\eta_1,\eta_2)\left(\frac{-u_i}{\sqrt{-2\eta_2}}-\eta_1\frac{\sqrt{-2\eta_2}}{4\eta_2^2}\right)\right\}.$$

The re-parameterization ($\mu = -\eta_1/2\eta_2$ and $\sigma^2 = -1/2\eta_2$) gives

$$\mu = \overline{y} + \frac{1}{\sqrt{-2\eta_2}}\frac{1}{n}\sum_{i=1}^{n}\{H_{i1}(\eta_1,\eta_2)-H_{i2}(\eta_1,\eta_2)\},$$
$$\sigma^2 = \overline{y^2} - \mu^2$$
$$-\frac{1}{n}\sum_{i=1}^{n}\left\{H_{i1}(\eta_1,\eta_2)\left(\frac{-v_i}{\sqrt{-2\eta_2}}-\eta_1\frac{\sqrt{-2\eta_2}}{4\eta_2^2}\right)\right\}$$
$$+\frac{1}{n}\sum_{i=1}^{n}\left\{H_{i2}(\eta_1,\eta_2)\left(\frac{-u_i}{\sqrt{-2\eta_2}}-\eta_1\frac{\sqrt{-2\eta_2}}{4\eta_2^2}\right)\right\}.$$

Now, we can obtain the MLE of μ and σ^2 by a slightly modified two-dimensional fixed-point iteration method. The $(k+1)$th step of the iteration becomes

$$\mu^{(k+1)} = \overline{y} + \frac{1}{\sqrt{-2\eta_2^{(k)}}}\frac{1}{n}\sum_{i=1}^{n}\left\{H_{i1}(\eta_1^{(k)},\eta_2^{(k)})-H_{i2}(\eta_1^{(k)},\eta_2^{(k)})\right\},$$

$$\sigma^{2^{(k+1)}} = \overline{y^2} - \{\mu^{(k+1)}\}^2 - \frac{1}{n}\sum_{i=1}^{n}\left\{H_{i1}(\eta_1^{(k)}, \eta_2^{(k)})\left(\frac{-v_i}{\sqrt{-2\eta_2^{(k)}}} - \eta_1^{(k)}\frac{\sqrt{-2\eta_2^{(k)}}}{4\eta_2^{2(k)}}\right)\right\}$$
$$+\frac{1}{n}\sum_{i=1}^{n}\left\{H_{i2}(\eta_1^{(k)}, \eta_2^{(k)})\left(\frac{-u_i}{\sqrt{-2\eta_2^{(k)}}} - \eta_1^{(k)}\frac{\sqrt{-2\eta_2^{(k)}}}{4\eta_2^{2(k)}}\right)\right\}.$$

By the invariance of MLE (Casella and Berger 2002, pp. 319–320), the MLE of $(\eta_1, \eta_2)^{\mathrm{T}}$ is

$$\begin{bmatrix}\eta_1^{(k+1)}\\ \eta_2^{(k+1)}\end{bmatrix} = \begin{bmatrix}\mu^{(k+1)}/\sigma^{2^{(k+1)}}\\ -1/2\sigma^{2^{(k+1)}}\end{bmatrix}.$$

The iteration continues until $|\eta_i^{(k+1)} - \eta_i^{(k)}| < \varepsilon, \quad i = 1, 2$ for small $\varepsilon > 0$.

4 Simulations

We conduct simulations by randomly drawing (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 vector (U, Y, V) subject to the inclusion criterion $U \le Y \le V$. Here, Y follows the SEF and the distribution of (U, V) is chosen such that $P(U \le Y \le V) \approx 0.5$. The details of the data generation schemes are given in "Appendix".

4.1 Simulation results for the one-parameter SEF

First, we consider the case $\eta > 0$, in particular the two cases $\eta = 1$ and 3. These true values are used as initial value $\eta^{(0)}$ in the algorithms. Also, the data-driven initial value for η of the form $\eta^{(0)} = 1/(\tau_2 - \overline{y})$ is considered, where $\overline{y} = \sum_i y_i/n$ and $\tau_2 = y_{(n)}$.

Table 1 compares the performance of the NR (Newton–Raphson) and FPI (fixed point iteration) methods for $\eta > 0$. We observe that the NR and FPI give almost identical estimates and that estimates are not affected by the initial values. For both algorithms, estimates are almost unbiased and the mean squared error (MSE) decreases as the sample size increases. However, the average number of iterations of the NR is much smaller than that of the FPI. Hence, the NR and FPI converges to the same values, but only the speed is different.

The virtually same results are found for the case $\eta > 0$ (Table 2).

4.2 Simulation results for the two-parameter SEF

We use the true parameter values $(\eta_1, \eta_2) = (30, -0.5)$ or $(5, -0.5)$ as the initial value $(\eta_1^{(0)}, \eta_2^{(0)})$. By the relationship $(\eta_1, \eta_2) = (\mu/\sigma^2, -1/2\sigma^2)$, we also

Table 1 Simulation results under a one-parameter SEF with a parameter $\eta > 0$ based on 500 repetitions

	True	Initial value	Method	$E(\hat{\eta})$	$MSE\ (\hat{\eta})$	AI
$n = 100$	$\eta = 3$	$\eta^{(0)} = 3$	FPI	3.0960	0.2041	12.6
			NR	3.0960	0.2042	4.42
		$\eta^{(0)} = \frac{1}{y_{(n)} - \bar{y}}$	FPI	3.0960	0.2042	12.42
			NR	3.0960	0.2042	4.3
$n = 200$	$\eta = 3$	$\eta^{(0)} = 3$	FPI	3.0614	0.1035	12.14
			NR	3.0618	0.1036	4.25
		$\eta^{(0)} = \frac{1}{y_{(n)} - \bar{y}}$	FPI	3.0619	0.1036	12.06
			NR	3.0618	0.1036	4.1
$n = 300$	$\eta = 3$	$\eta^{(0)} = 3$	FPI	3.0426	0.0652	11.86
			NR	3.0426	0.0652	4.12
		$\eta^{(0)} = \frac{1}{y_{(n)} - \bar{y}}$	FPI	3.0426	0.0652	11.62
			NR	3.0426	0.0652	4.0
$n = 100$	$\eta = 1$	$\eta^{(0)} = 1$	FPI	1.0331	0.0226	10.92
			NR	1.0331	0.0226	4.25
		$\eta^{(0)} = \frac{1}{y_{(n)} - \bar{y}}$	FPI	1.0331	0.0226	10.57
			NR	1.0331	0.0226	4.1
$n = 200$	$\eta = 1$	$\eta^{(0)} = 1$	FPI	1.0205	0.0114	10.46
			NR	1.0205	0.0114	4.06
		$\eta^{(0)} = \frac{1}{y_{(n)} - \bar{y}}$	FPI	1.0206	0.0114	10.16
			NR	1.0205	0.0114	3.94
$n = 300$	$\eta = 1$	$\eta^{(0)} = 1$	FPI	1.0143	0.0072	10.18
			NR	1.0143	0.0072	3.96
		$\eta^{(0)} = \frac{1}{y_{(n)} - \bar{y}}$	FPI	1.0144	0.0072	9.95
			NR	1.0143	0.0072	3.89

FPI fixed point iteration, *NR* Newton–Raphson algorithm, $MSE(\hat{\eta}) = E(\hat{\eta} - \eta)^2$, *AI* the average number of iterations until convergence

consider data-driven initial values $(\eta_1^{(0)}, \eta_2^{(0)}) = (\bar{y}/s^2, -1/2s^2)$, where $s^2 = \sum_i (y_i - \bar{y})^2/(n-1)$.

We compare the performance of the FPI and NR methods in Table 3. It can be seen that both the NR and FPI work well and they give almost identical estimates. The estimates are not influenced by the choices of initial values. However, the average number of iterations of the NR is remarkably smaller than that of the FPI. Hence, the NR and FPI converge to the same solution, but the NR converges more quickly than the FPI.

Next, we investigate how well the MLE performs under model misspecifications, especially compared to the distribution-free method. For estimating a true survival probability $S(y) = 0.5$ or 0.75, the parametric estimator $S_{\hat{\eta}}(y)$, where $\hat{\boldsymbol{\eta}} = (\hat{\eta}_1, \hat{\eta}_2)$, and the NPMLE (Efron and Petrosian 1999) are compared. Data are generated from

Table 2 Simulation results under a one-parameter SEF with a parameter $\eta < 0$ based on 500 repetitions

	True	Initial value	Method	$E(\hat{\eta})$	$MSE(\hat{\eta})$	AI
$n = 100$	$\eta = -3$	$\eta^{(0)} = -3$	FPI	−3.0975	0.2337	12.68
			NR	−3.0975	0.2337	4.42
		$\eta^{(0)} = \frac{1}{y_{(1)} - \bar{y}}$	FPI	−3.0976	0.2337	12.24
			NR	−3.0975	0.2337	4.27
$n = 200$	$\eta = -3$	$\eta^{(0)} = -3$	FPI	−3.0440	0.1057	12.24
			NR	−3.0440	0.1057	4.26
		$\eta^{(0)} = \frac{1}{y_{(1)} - \bar{y}}$	FPI	−3.0440	0.1057	11.92
			NR	−3.0440	0.1057	4.13
$n = 300$	$\eta = -3$	$\eta^{(0)} = -3$	FPI	−3.0271	0.0646	11.98
			NR	−3.0271	0.0647	4.14
		$\eta^{(0)} = \frac{1}{y_{(1)} - \bar{y}}$	FPI	−3.0272	0.0647	11.68
			NR	−3.0271	0.0647	4.03
$n = 100$	$\eta = -1$	$\eta^{(0)} = -1$	FPI	−1.0311	0.0256	11.09
			NR	−1.0311	0.0256	4.26
		$\eta^{(0)} = \frac{1}{y_{(1)} - \bar{y}}$	FPI	−1.0311	0.0256	10.55
			NR	−1.0311	0.0256	4.08
$n = 200$	$\eta = -1$	$\eta^{(0)} = -1$	FPI	−1.0144	0.0116	10.55
			NR	−1.0144	0.0116	4.08
		$\eta^{(0)} = \frac{1}{y_{(1)} - \bar{y}}$	FPI	−1.0144	0.0116	10.22
			NR	−1.0144	0.0116	3.96
$n = 300$	$\eta = -1$	$\eta^{(0)} = -1$	FPI	−1.0087	0.0071	10.3
			NR	−1.0087	0.0071	3.97
		$\eta^{(0)} = \frac{1}{y_{(1)} - \bar{y}}$	FPI	−1.0088	0.0071	10.03
			NR	−1.0087	0.0071	3.89

FPI fixed point iteration, *NR* Newton–Raphson algorithm, $MSE(\hat{\eta}) = E(\hat{\eta} - \eta)^2$, *AI* the average number of iterations until convergence

four different models, including the correct two-dimensional SEF, and three other misspecified models. Note that the value of y to make $S(y) = 0.5$ or 0.75 depends on the chosen model (Table 4).

Table 4 compares the parametric estimator with the NPMLE under the four model alternatives. The NPMLE shows almost unbiased results for estimating the true survival function under all the models considered. If the true model is the one-dimensional SEF, the fitted model is misspecified and the parametric estimator exhibits some systematic bias that does not vanish with large samples. However, the MSE still tends to be smaller than that of the NPMLE. This phenomenon is due to the small variance under the one-parameter model. If the true model is the three-parameter SEF, the fitted model is also misspecified, but the bias of the parametric estimator is quite modest. As

Table 3 Simulation results under a two-parameter SEF with parameters (η_1, η_2) based on 500 repetitions

	True (η_1, η_2)	Initial (η_1^0, η_2^0)	Method	$E(\hat{\eta}_1)$	$E(\hat{\eta}_2)$	$MSE(\hat{\eta}_1)$	$MSE(\hat{\eta}_2)$	AI
$n = 100$	(30, −0.5)	(30, −0.5)	FPI	31.329	−0.522	65.378	0.0181	34.6
			NR	31.329	−0.522	65.382	0.0181	5.2
		$(\frac{\bar{y}}{s^2}, \frac{-1}{2s^2})$	FPI	31.329	−0.522	65.380	0.0181	37.7
			NR	31.329	−0.522	65.382	0.0181	6.2
$n = 200$	(30, −0.5)	(30, −0.5)	FPI	30.721	−0.512	31.135	0.0087	31.8
			NR	30.721	−0.512	31.137	0.0087	5.0
		$(\frac{\bar{y}}{s^2}, \frac{-1}{2s^2})$	FPI	30.721	−0.512	31.137	0.0087	35.9
			NR	30.721	−0.512	31.137	0.0087	6.2
$n = 300$	(30, −0.5)	(30, −0.5)	FPI	30.731	−0.512	21.094	0.0059	30.5
			NR	30.731	−0.512	21.096	0.0059	4.9
		$(\frac{\bar{y}}{s^2}, \frac{-1}{2s^2})$	FPI	30.731	−0.512	21.096	0.0059	35.1
			NR	30.731	−0.512	21.096	0.0059	6.1
$n = 100$	(5, −0.5)	(5, −0.5)	FPI	5.216	−0.522	1.851	0.0181	28.2
			NR	5.216	−0.522	1.852	0.0181	5.0
		$(\frac{\bar{y}}{s^2}, \frac{-1}{2s^2})$	FPI	5.217	−0.522	1.851	0.0181	31.3
			NR	5.216	−0.522	1.852	0.0181	6.0
$n = 200$	(5, −0.5)	(5, −0.5)	FPI	5.115	−0.512	0.874	0.0087	25.9
			NR	5.115	−0.512	0.875	0.0087	4.8
		$(\frac{\bar{y}}{s^2}, \frac{-1}{2s^2})$	FPI	5.116	−0.512	0.875	0.0087	29.9
			NR	5.115	−0.512	0.875	0.0087	6.0
$n = 300$	(5, −0.5)	(5, −0.5)	FPI	5.116	−0.512	0.592	0.0059	24.8
			NR	5.116	−0.512	0.592	0.0059	4.7
		$(\frac{\bar{y}}{s^2}, \frac{-1}{2s^2})$	FPI	5.117	−0.512	0.592	0.0059	29.3
			NR	5.116	−0.512	0.592	0.0059	6.0

FPI fixed point iteration, *NR* Newton–Raphson algorithm, $MSE(\hat{\eta}_1) = E(\hat{\eta}_1 - \eta_1)^2$, $MSE(\hat{\eta}_2) = E(\hat{\eta}_2 - \eta_2)^2$, *AI* the average number of iterations until convergence

a result, the MLE still perform better than the NPMLE in terms of the MSE. If the true model is the two-dimensional SEF, then the fitted model is correct and the parametric approach is superior to the NPMLE in terms of both the bias and MSE.

4.3 Simulation results for the cubic SEF

As mentioned in Sect. 3.1, we will perform our simulations on the basis of the approximate log-likelihood in Eq. (3), which gives more stable results over large number of repetitions. The chosen initial values of the NR algorithm are the true values $(\eta_1^{(0)}, \eta_2^{(0)}, \eta_3^{(0)}) = (5, -0.5, 0.005)$ or $(5, -0.5, -0.005)$, and the data-driven initial values $(\eta_1^{(0)}, \eta_2^{(0)}, \eta_3^{(0)}) = (\bar{y}/s^2, -1/2s^2, 0)$. In addition, we also test some wrong

Table 4 Simulation results for comparing the MLE with the NPMLE based on 500 repetitions

True model	True survival	Sample size	Fit by 2-parameter SEF		Fit by NPMLE	
			$Bias\{\hat{S}(y)\}$	$MSE\{\hat{S}(y)\}$	$Bias\{\hat{S}(y)\}$	$MSE\{\hat{S}(y)\}$
1-par SEF	$S(y) = 0.5$	$n = 100$	0.0230	0.0029	0.0000	0.0056
$\eta = -1$	at $y = 4.7$	$n = 200$	0.0230	0.0017	0.0020	0.0040
$\tau_1 = 4$		$n = 300$	0.0230	0.0013	0.0010	0.0027
2-par SEF	$S(y) = 0.5$	$n = 100$	−0.0030	0.0042	−0.0060	0.0052
(η_1, η_2)	at $y = 5.0$	$n = 200$	−0.0020	0.0022	−0.0030	0.0026
$= (5, -0.5)$		$n = 300$	−0.0030	0.0014	−0.0030	0.0017
3-par SEF	$S(y) = 0.5$	$n = 100$	0.0050	0.0047	0.0040	0.0054
(η_1, η_2, η_3)	at $y = 5.4$	$n = 200$	0.0040	0.0022	0.0030	0.0025
$= (5, -0.5, 0.005)$		$n = 300$	0.0060	0.0014	0.0050	0.0016
3-par SEF	$S(y) = 0.5$	$n = 100$	0.0030	0.0038	0.0050	0.0046
(η_1, η_2, η_3)	at $y = 4.7$	$n = 200$	0.0020	0.0017	0.0030	0.0021
$= (5, -0.5, -0.005)$		$n = 300$	0.0020	0.0012	0.0040	0.0013
1-par SEF	$S(y) = 0.75$	$n = 100$	−0.0260	0.0023	−0.0010	0.0030
$\eta = -1$	at $y = 4.3$	$n = 200$	−0.0280	0.0016	0.0000	0.0018
$\tau_1 = 4$		$n = 300$	−0.0280	0.0014	0.0000	0.0012
2-par SEF	$S(y) = 0.75$	$n = 100$	−0.0030	0.0037	−0.0050	0.0058
(η_1, η_2)	at $y = 4.3$	$n = 200$	−0.0020	0.0018	−0.0030	0.0026
$= (5, -0.5)$		$n = 300$	−0.0010	0.0012	−0.0010	0.0015
3-par SEF	$S(y) = 0.75$	$n = 100$	0.0030	0.0026	0.0000	0.0034
(η_1, η_2, η_3)	at $y = 4.7$	$n = 200$	0.0010	0.0014	−0.0010	0.0017
$= (5, -0.5, 0.005)$		$n = 300$	0.0030	0.0009	0.0020	0.0011
3-par SEF	$S(y) = 0.75$	$n = 100$	0.0030	0.0035	0.0010	0.0047
(η_1, η_2, η_3)	at $y = 4.0$	$n = 200$	0.0020	0.0018	0.0010	0.0024
$= (5, -0.5, -0.005)$		$n = 300$	0.0020	0.0012	0.0020	0.0016

$Bias\{\hat{S}(y)\} = E\{\hat{S}(y)\} - S(y)$ for $S(y) = 0.50$ or 0.75
$MSE\{\hat{S}(y)\} = E\{\hat{S}(y) - S(y)\}^2$ for $S(y) = 0.50$ or 0.75
For the 3-parameter SEF (3-par SEF), we choose $\tau_2 = 8$ for $\eta_3 = 0.005$ and $\tau_1 = 2$ for $\eta_3 = -0.005$

initial values $(\eta_1^{(0)}, \eta_2^{(0)}, \eta_3^{(0)}) = (-3, -0.5, 0.005)$ or $(-3, -0.5, -0.005)$ to examine the sensitivity. The randomized Newton-Raphson (RNR) is performed based on the choices $D_1 = 20$, $D_2 = 10$, $D_3 = 1$, $d_1 = 6$ and $d_2 = 0.5$.

Tables 5 compare the performance of the randomized Newton–Raphson (RNR) and R optim routine (R Development Core Team 2014) with $\eta_3 > 0$. The RNR always converges to the solution without being influenced by the choices of initial values. On the other hand, the R optim occasionally yields un-convergence (Table 5). This happens when the data-driven initial value is used. If the initial value is the true value, the RNR and R optim produce very similar results. However, the speed of convergence is extremely faster for the RNR than that of the R optim. In average, the RNR converges within 5 to 7 iterations while the R optim takes nearly 200 iterations.

Table 5 Simulations results under a cubic SEF with parameters $(\eta_1, \eta_2, \eta_3) = (5, -0.5, 0.005)$ and $\tau_2 = 8$ based on 500 repetitions

	Initial value $(\eta_1^{(0)}, \eta_2^{(0)}, \eta_3^{(0)})$	Method	$E(\hat{\eta}_1)$	$E(\hat{\eta}_2)$	$E(\hat{\eta}_3)$	$MSE(\hat{\eta}_1)$	$MSE(\hat{\eta}_2)$	$MSE(\hat{\eta}_3)$	AI
$n = 100$	$(5, -0.5, 0.005)$	RNR	5.247	−0.501	0.0022	43.99	1.63	0.0066	5.8
		Optim	5.247	−0.501	0.0022	43.99	1.63	0.0066	175.4
	$(\frac{\bar{y}}{s^2}, \frac{-1}{2s^2}, 0)$	RNR	5.247	−0.501	0.0022	43.99	1.63	0.0066	7.3
		Optim				Un-convergent			
	$(-3, -0.5, 0.005)$	RNR	5.247	−0.501	0.0022	43.99	1.63	0.0066	7.0
		Optim	3.499	−0.162	0.0192	49.77	1.90	0.0079	212.0
$n = 200$	$(5, -0.5, 0.005)$	RNR	4.998	−0.478	0.0023	22.60	0.84	0.0034	5.5
		Optim	4.994	−0.477	0.0023	22.59	0.84	0.0034	168.1
	$(\frac{\bar{y}}{s^2}, \frac{-1}{2s^2}, 0)$	RNR	4.998	−0.478	0.0023	22.60	0.84	0.0034	7.3
		Optim				Un-convergent			
	$(-3, -0.5, 0.005)$	RNR	4.998	−0.478	0.0023	22.60	0.84	0.0034	7.0
		Optim	3.253	−0.137	0.0194	33.65	1.29	0.0054	221.9
$n = 300$	$(5, -0.5, 0.005)$	RNR	5.019	−0.487	0.0032	14.69	0.54	0.0022	5.2
		Optim	5.018	−0.487	0.0032	14.69	0.54	0.0022	162.1
	$(\frac{\bar{y}}{s^2}, \frac{-1}{2s^2}, 0)$	RNR	5.019	−0.487	0.0032	14.69	0.54	0.0022	5.5
		Optim				Un-convergent			
	$(-3, -0.5, 0.005)$	RNR	5.019	−0.487	0.0032	14.69	0.54	0.0022	5.5
		Optim	2.566	−0.007	0.0272	30.36	1.17	0.0048	216.1

$MSE(\hat{\eta}_1) = E(\hat{\eta}_1 - \eta_1)^2$, $MSE(\hat{\eta}_2) = E(\hat{\eta}_2 - \eta_2)^2$ and $MSE(\hat{\eta}_3) = E(\hat{\eta}_3 - \eta_3)^2$
RNR = Randomized Newton–Raphson algorithm (proposed method)
AI = The average number of iterations until convergence
Un-convergent = At least one repetition does not converge among 500 repetitions
Optim = R optim routine for maximizing the log-likelihood

Table 5 also demonstrates the performance of the estimators $\hat{\eta}_i, i = 1, 2, 3$. As long as the algorithm converges, the estimators are roughly unbiased, and their MSE vanishes as the sample size increase from $n = 100$ to 300.

Table 6 displays the results with $\eta_3 < 0$. The results are virtually same as Table 5.

5 Data analysis

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

We analyzed the childhood cancer data from Moreira and de Uña-Álvarez (2010). Their data include the ages at onset of cancer at a young age within a recruitment period of 5 years (between January 1, 1999 and December 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. Specifically, the data consists of 406 children with $\{(u_i, y_i, v_i) : i =$

Table 6 Simulations results under a cubic SEF with parameters $(\eta_1, \eta_2, \eta_3) = (5, -0.5, -0.005)$ and $\tau_1 = 2$ based on 500 repetitions

	Initial value $(\eta_1^{(0)}, \eta_2^{(0)}, \eta_3^{(0)})$	Method	$E(\hat{\eta}_1)$	$E(\hat{\eta}_2)$	$E(\hat{\eta}_3)$	$MSE(\hat{\eta}_1)$	$MSE(\hat{\eta}_2)$	$MSE(\hat{\eta}_3)$	AI
$n = 100$	(5, −0.5, −0.005)	RNR	5.373	−0.521	−0.0074	50.59	2.20	0.010	5.9
		Optim	5.361	−0.518	−0.0075	50.55	2.19	0.010	183.6
	$(\frac{\bar{y}}{s^2}, \frac{-1}{2s^2}, 0)$	RNR	5.373	−0.521	−0.0074	50.59	2.20	0.010	7.3
		Optim				Un-convergent			
	(−3, −0.5, −0.005)	RNR	5.373	−0.521	−0.0074	50.59	2.20	0.010	7.1
		Optim	4.640	−0.368	−0.0178	49.94	2.22	0.011	195.0
$n = 200$	(5, −0.5, −0.005)	RNR	5.243	−0.523	−0.0052	24.70	1.09	0.005	5.5
		Optim	5.245	−0.523	−0.0052	24.69	1.09	0.005	170.9
	$(\frac{\bar{y}}{s^2}, \frac{-1}{2s^2}, 0)$	RNR	5.243	−0.523	−0.0052	24.70	1.09	0.005	7.3
		Optim				Un-convergent			
	(−3, −0.5, −0.005)	RNR	5.243	−0.523	−0.0052	24.70	1.09	0.005	7.2
		Optim	4.935	−0.458	−0.0097	26.68	1.18	0.006	196.5
$n = 300$	(5, −0.5, −0.005)	RNR	5.282	−0.539	−0.0036	16.87	0.73	0.003	5.3
		Optim	5.338	−0.551	−0.0027	16.95	0.73	0.004	163.4
	$(\frac{\bar{y}}{s^2}, \frac{-1}{2s^2}, 0)$	RNR	5.282	−0.539	−0.0036	16.87	0.73	0.003	7.3
		Optim				Un-convergent			
	(−3, −0.5, −0.005)	RNR	5.282	−0.539	−0.0036	16.87	0.73	0.003	7.3
		Optim	5.080	−0.496	−0.0065	18.62	0.81	0.004	197.3

$MSE(\hat{\eta}_1) = E(\hat{\eta}_1 - \eta_1)^2$, $MSE(\hat{\eta}_2) = E(\hat{\eta}_2 - \eta_2)^2$ and $MSE(\hat{\eta}_3) = E(\hat{\eta}_3 - \eta_3)^2$
RNR = Randomized Newton–Raphson algorithm (proposed method)
AI = The average number of iterations until convergence
Un-convergent = At least one repetition does not converge among 500 repetitions
Optim = R optim routine for maximizing the log-likelihood

$1, \ldots, 406\}$ subject to double-truncation $u_i \le y_i \le v_i$, where y_i is the age (in days) at diagnosis, u_i is the age at the start of the recruitment (January 1, 1999), and $v_i = u_i + 1{,}825$ is the age at the end of the recruitment (December 31, 2003). It is of interest to make statistical inference for the survival function $S(t)$ of the pre-truncated age at diagnosis. The data have been analyzed by Moreira and de Uña-Álvarez (2010) and Emura et al. (2014) based on the NPMLE. In this paper, we provide additional inference results based on the MLE under the SEF.

5.2 Numerical results

The following models are fitted to the data:

Model (a): one-parameter SEF ($\eta_1 > 0$),
Model (b): one-parameter SEF ($\eta_1 < 0$),
Model (c): two-parameter SEF,
Model (d): cubic SEF ($\eta_3 < 0$).

To find the MLE, we apply the NR algorithm under the stopping criterion $\varepsilon = 0.0001$ for all the models.

For the one-parameter SEF with $\eta_1 > 0$, we set $\tau_2 = y_{(n)} = 5{,}474$ such that $\int_{\mathcal{Y}} \exp(\eta_1 y)\, dy < \infty$. The initial value of the NR method is $\eta_1^{(0)} = 1/(\tau_2 - \overline{y}) = 0.0003215$, and the resultant MLE becomes $\hat{\eta}_1 = 0.0000874$ at the 2nd iteration. For the case $\eta_1 < 0$, we set $\tau_1 = y_{(1)} = 6$ such that $\int_{\mathcal{Y}} \exp(\eta_1 y)\, dy < \infty$. The initial value of the NR method is $\eta_1^{(0)} = 1/(\tau_1 - \overline{y}) = -0.000424$, and the MLE becomes $\hat{\eta}_1 = -0.000385$ at the 1st iteration. Figure 3 shows that the NR method successfully searches the global maximum of the log-likelihood function.

For the two-parameter SEF, we perform the NR method with an initial value $(\eta_1^{(0)}, \eta_2^{(0)}) = (\overline{y}/s^2, -1/2s^2) = (0.000875, -1.85 \times 10^{-7})$. The resultant MLE becomes $(\hat{\eta}_1, \hat{\eta}_2) = (0.000771, -1.87 \times 10^{-7})$ at the 2nd iteration. Accordingly, $(\hat{\mu}, \hat{\sigma}) = (-\hat{\eta}_1/2\hat{\eta}_2, \sqrt{-1/2\hat{\eta}_2}) = (2{,}065.1, 1{,}636.6)$. Figure 4 demonstrates that that the MLE indeed attains the global maximum of the log-likelihood function.

For the cubic SEF, we set the lower limit of the support as $\tau_1 = y_{(1)} = 6$, and work on the exact log-likelihood in Example 3 of Sect. 3.1. We appropriately modify the NR steps of Sect. 3.2 to adapt to the exact likelihood. The NR algorithm starts with $(\eta_1^{(0)}, \eta_2^{(0)}, \eta_3^{(0)}) = (\overline{y}/s^2, -1/2s^2, 0) = (0.000875, -1.85 \times 10^{-7}, 0)$. The resultant MLE becomes $(\hat{\eta}_1, \hat{\eta}_2, \hat{\eta}_3) = (-0.00079, 3.38 \times 10^{-7}, -4.87 \times 10^{-11})$ at the 3rd iteration. The MLE attains the global maximum of the log-likelihood function (Fig. 5).

We select a suitable model among the four candidates in terms of Akaike information criterion (AIC) (Akaike 1973), given by AIC $= -2 \log L + 2k$, where L is the maximized value of the likelihood function and k is the number of unknown parame-

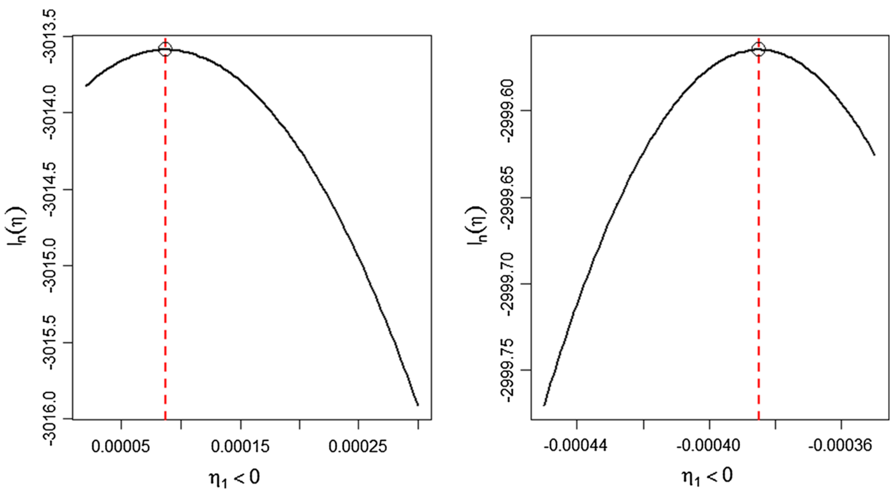

Fig. 3 The log-likelihood under the one-parameter SEF based on the childhood cancer data. The *vertical lines* signify the MLEs $\hat{\eta}_1 = 0.0000874$ (*left*) and $\hat{\eta}_1 = -0.000385$ (*right*)

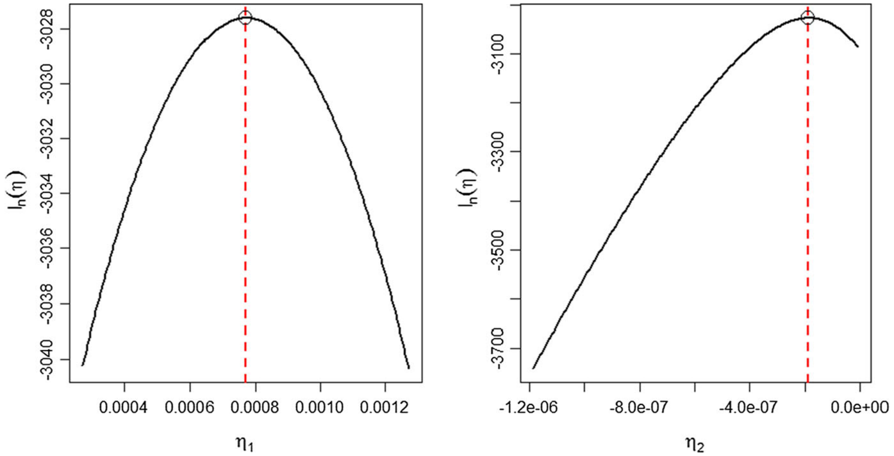

Fig. 4 The log-likelihood under the two-parameter SEF based on the childhood cancer data. The MLEs $\hat{\eta}_1 = 7.71 \times 10^{-4}$ and $\hat{\eta}_2 = -1.87 \times 10^{-7}$ are indicated by *vertical lines*

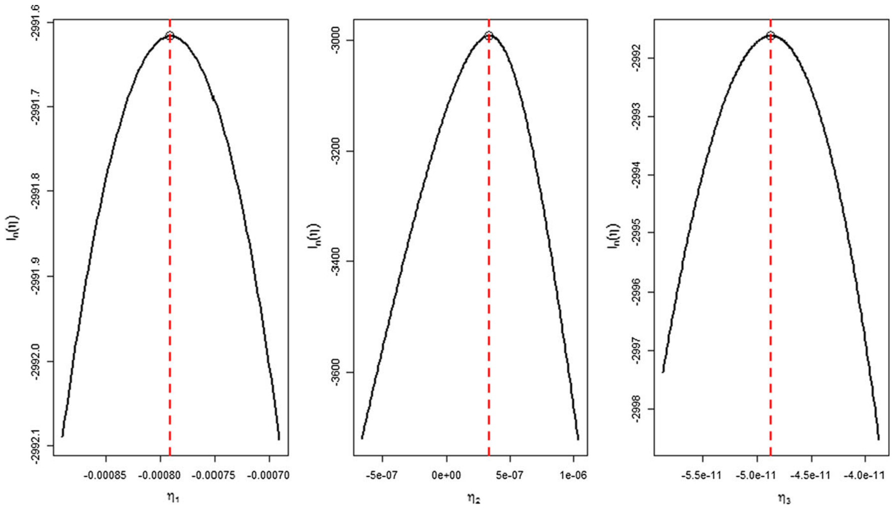

Fig. 5 The log-likelihood under the cubic SEF based on the childhood cancer data. The MLEs $\hat{\eta}_1 = -7.90 \times 10^{-4}$, $\hat{\eta}_2 = 3.38 \times 10^{-7}$, and $\hat{\eta}_3 = -4.87 \times 10^{-11}$ are signified by the *vertical lines*

ters under the fitted model. The preferred model is the one with the minimum AIC value. Table 7 summarizes the results of model selection. The best AIC is attained by the cubic SEF ($k = 3$) with $\eta_3 < 0$. Note that the likelihood ratio test may not work since the candidate models are not nested.

Figure 6 compares the fitted survival curves for the four candidate models with the survival curve of the NPMLE. Unlike the AIC-based comparison, such a comparison gives a graphical way to examine the goodness-of-fit. Here, the survival curves for the parametric estimators are calculated as $S_{\hat{\boldsymbol{\eta}}}(y)$, where $S_{\boldsymbol{\eta}}(y)$ is defined in Sect. 2. To

Table 7 The maximum likelihood inference for the childhood cancer data

Model	$\hat{\eta}_1$	$\hat{\eta}_2$	$\hat{\eta}_3$	log L	AIC	K–S statistic
(a) 1 par. SEF ($\eta_1 > 0$)	8.74×10^{-5}	0	0	−3,013.6	6,029.2	0.206
(b) 1 par. SEF ($\eta_1 < 0$)	-3.85×10^{-4}	0	0	−2,999.6	6,001.1	0.121
(c) 2 par. SEF	7.71×10^{-4}	-1.87×10^{-7}	0	−3,027.6	6,059.2	0.132
(d) Cubic SEF ($\eta_3 < 0$)	-7.90×10^{-4}	3.38×10^{-7}	-4.87×10^{-11}	−2,991.6	5,989.2	0.084

Model (a) = The one-parameter SEF ($\eta_1 > 0$)
Model (b) = The one-parameter SEF ($\eta_1 < 0$)
Model (c) = The two-parameter SEF
Model (d) = The cubic SEF ($\eta_3 < 0$); log L = The maximized log-likelihood
AIC = Akaike information criterion, defined as AIC $= -2 \log L + 2k$; (smaller AIC corresponds to better fit)
K–S statistic = The Kolmogorov–Smirnov distance between the MLE and the NPMLE of the survival function (smaller K–S statistic corresponds to better fit)

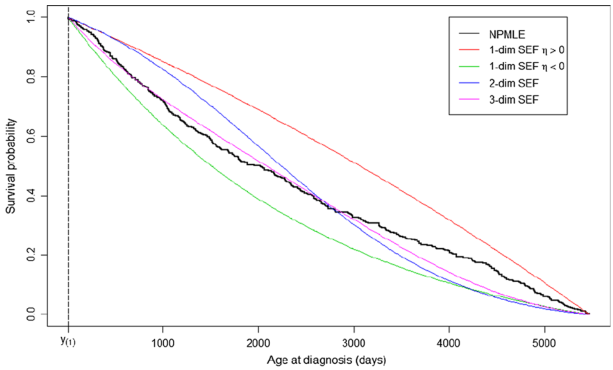

Fig. 6 The survival functions for the childhood cancer data. Five survival curves are calculated using the one-parameter SEF (1-dim SEF $\eta_1 > 0$ and 1-dim SEF $\eta_1 < 0$), two-parameter SEF (2-dim SEF), cubic SEF (3-dim SEF $\eta_3 < 0$) and the NPMLE. The *vertical line* signifies the value $\min_i(y_i) = y_{(1)} = 6$

measure the goodness-of-fit, we use the Kolmogorov–Smirnov (K–S) statistic, which is the maximum vertical distance between fitted survival curve and the NPMLE. We find that the minimum K–S statistic is achieved by the cubic SEF with $\eta_3 < 0$. This is also clear from Fig. 6 that the survival curve for the cubic SEF is quite close to that obtained by the NPMLE.

Since both the AIC and K–S statistic select the cubic SEF with $\eta_3 < 0$ as the best model, we suggest this model for estimating the survival function of the pre-truncated age at diagnosis for childhood cancer.

Figure 7 depicts the estimated density under the cubic SEF with $\eta_3 < 0$, giving a highly skewed density with the lower limit of the support at $\tau_1 = \min_i(y_i) = 6$. This figure reminds the important point that one needs to set the finite support $\mathcal{Y} = [\tau_1, \infty)$ for the density to be well-defined. As mentioned in Remark I, such a skewed truncated normal density is often used for fitting lifetime data.

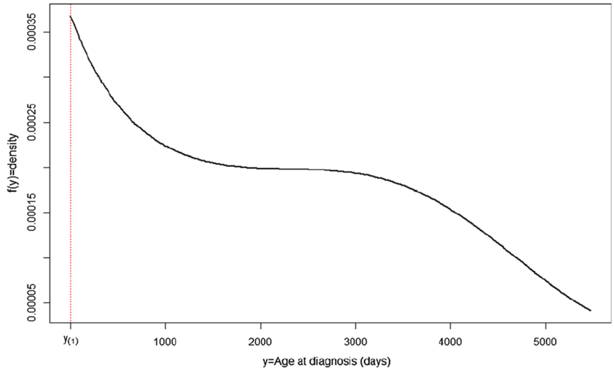

Fig. 7 The estimated density function under the cubic SEF with $\eta_3 < 0$ for the childhood cancer data. The *vertical line* signifies the lower support $\tau_1 = \min_i(y_i) = y_{(1)} = 6$

6 Conclusion and discussion

While the present research may be framed as a typical application of the MLE, we have demonstrated that the adequate implementation requires careful set-ups of parameter space, support, and computational algorithms, especially under double-truncation. For practitioners, such information should not be ignored or simply treated as minor technical matters. In particular, we emphasize that the adequate choice of the support of the density is essential for the SEF to be well-defined and to be fitted with doubly-truncated data. Concretely, whether truncation limits exceed/precede the support of the density would influence the likelihood for observed lifetimes. This paper has clarified these technical aspects of parametric inference under double-truncation, which has not been explicitly discussed in the literature.

As for computational contributions, we have developed the Newton–Raphson algorithm and fixed-point iteration to implement the MLE under the SEF. In particular, we give explicit formulations of the Newton–Raphson and fixed-point iteration algorithms, and compare the performance of the algorithms using simulations. For the three-parameter SEF, the computation is more involved and the convergence of the Newton–Raphson algorithm becomes unstable. In light of these results, we consider an approximated log-likelihood under certain assumptions on the support, and also developed the randomized version of the Newton–Raphson methods. Our simulations demonstrate that the proposed randomized Newton–Raphson always converge to the MLE without influenced by the choice of initial values. Moreover, we see that the proposed algorithm outperforms an existing R package for optimization in terms of both the speed and stability of convergence.

We propose to fit the cubic SEF model to the childhood cancer data after model selection. An important advantage of the estimator with the cubic SEF over the NPMLE is to offer a smooth survival curve. Moreover, it is also possible to obtain the density estimator. On the other hand, the density does not directly follow from the discrete distribution of the NPMLE. An alternative is a smooth kernel density estimator developed by Moreira and de Uña-Álvarez (2012) and Moreira and Van Keilegom (2013).

Goodness-of-fit procedures are clearly important when the present SEF models and other parametric models are fitted to doubly-truncated data. In our data analysis, we assess the goodness-of-fit based on the Kolmogorov–Smirnov distance between the model-based and the model-free survival functions. This approach can be further formalized by developing a hypothesis test to check the fitted model. One needs to be careful for directly applying the goodness-of-fit test for doubly-truncated data by Emura et al. (2014) in which the null hypothesis is completely fixed. That is, this approach does not take into account for the variability of estimating unknown parameters. One straightforward but computationally demanding way is to apply a parametric bootstrap procedure as in Stute et al. (1993). The derivation of asymptotic distribution and the numerical validity (type I error and power) of the bootstrap test remain to be done. A more elaborate but computationally fast method follows the line of Emura and Konno (2012b). More studies on goodness-of-fit procedures under double-truncation are warranted.

Acknowledgments We are grateful to the comments and suggestions from associate editor and two anonymous referees that greatly improved the manuscript. This work is supported by the research grant funded by the National Science Council of Taiwan (NSC 101-2118-M-008-002-MY2) and the Ministry of Science and Technology of Taiwan (MOST 103-2118-M-008-MY2).

Appendix: The data generations for simulations

Data generations for one-parameter SEF

First, consider the case $\eta > 0$. We let

$$U \sim f_U(u) = \eta_u \exp\{\eta_u(u - \tau_2)\}, -\infty < u < \tau_2, \eta_u > 0,$$
$$V \sim f_V(v) = \eta_v \exp\{\eta_v(v - \tau_2)\}, -\infty < v < \tau_2, \eta_v > 0,$$
$$Y \sim f_Y(y) = \eta \exp\{\eta(y - \tau_2)\}, -\infty < y < \tau_2, \eta > 0.$$

In this case, data are generated by using inverse transformations

$$U = \tau_2 + \frac{1}{\eta_u}\log(W_1), \quad V = \tau_2 + \frac{1}{\eta_v}\log(W_2), \quad Y = \tau_2 + \frac{1}{\eta}\log(W_3),$$

where $W_1, W_2, W_3 \sim U(0, 1)$. Then, the inclusion probability is

$$P(U \le Y \le V) = \int_{-\infty}^{\tau_2}\int_{-\infty}^{v}\int_{-\infty}^{y} f_Y(y) f_V(v) f_U(u) du dy dv = \frac{\eta \cdot \eta_v}{(\eta + \eta_u)(\eta + \eta_u + \eta_v)}.$$

$\eta_u = 0$ yields $P(-\infty < Y \le V) = \eta_v/(\eta_v + \eta)$ and $\eta_v = \infty$ yields $P(U \le Y < \tau_2) = \eta/(\eta + \eta_u)$. One can get $P(U \le Y \le V) \approx 0.5$ by making $P(U \le Y) = 0.75$ and $P(Y \le V) = 0.75$.

For instance, with fixed $\eta = 3$, we find η_u and η_v as follows:

1. Set $P(U \le Y) = \eta/(\eta + \eta_u) = 0.75$, and then obtain $\eta_u = 1$.
2. Set $P(Y \le V) = \eta_v/(\eta_v + \eta) = 0.75$, and then obtain $\eta_v = 9$.

Accordingly, the inclusion probability becomes $P(U \leq Y \leq V) = 0.5031447$.

Another case is $\eta < 0$, where the range of Y is $y \in [\tau_1, \infty)$. We consider

$$U \sim f_U(u) = -\eta_u \exp\{\eta_u(u - \tau_1)\}, \tau_1 < u < \infty, \eta_u < 0,$$
$$V \sim f_V(v) = -\eta_v \exp\{\eta_v(v - \tau_1)\}, \tau_1 < v < \infty, \eta_v < 0,$$
$$Y \sim f_Y(y) = -\eta \exp\{\eta(y - \tau_1)\}, \tau_1 < y < \infty, \eta < 0.$$

In this case, data are generated by using inverse transformations

$$U = \tau_1 + \frac{1}{\eta_u}\log(1 - W_1), \quad V = \tau_1 + \frac{1}{\eta_v}\log(1 - W_2), \quad Y = \tau_1 + \frac{1}{\eta}\log(1 - W_3),$$

where $W_1, W_2, W_3 \sim U(0, 1)$. Then, the inclusion probability is

$$P(U \leq Y \leq V) = \int_{\tau}^{\infty}\int_{u}^{\infty}\int_{y}^{\infty} f_Y(y) f_V(v) f_U(u) dv dy du = \frac{\eta \cdot \eta_u}{(\eta + \eta_v)(\eta + \eta_u + \eta_v)}.$$

$\eta_u = \infty$ yields $P(0 \leq Y \leq V) = \eta/(\eta + \eta_v)$ and $\eta_v = 0$ yields $P(U \leq Y < \infty) = \eta_u/(\eta + \eta_u)$. One can get $P(U \leq Y \leq V) \approx 0.5$ by making $P(U \leq Y) = 0.75$ and $P(Y \leq V) = 0.75$.

For instance, with fixed $\eta = -1$, we find η_u and η_v as follows:

1. Set $P(U \leq Y) = \eta_u/(\eta + \eta_u) = 0.75$, and then obtain $\eta_u = -3$.
2. Set $P(Y \leq V) = \eta/(\eta + \eta_v) = 0.75$, and then obtain $\eta_v = -1/3$.

Accordingly, the inclusion probability becomes $P(U \leq Y \leq V) = 0.5235602$.

Data generation for two-parameter SEF

We consider $U \sim N(\mu_u, 1)$, $V \sim N(\mu_v, 1)$ and $Y \sim N(\mu, 1)$. One can obtain $P(U \leq Y \leq V) \approx 0.5$ by making left-truncated percentage is equal to right-truncated percentage. Since the normal distribution is symmetric, we set $\mu_u = \mu - \Delta$ and $\mu_v = \mu + \Delta$. Then,

$$P(U \leq Y \leq V) = \int_{-\infty}^{\infty} \varphi(y - \mu) \cdot \Phi(y - \mu + \Delta) \cdot \{1 - \Phi(y - \mu - \Delta)\} dy.$$

The desired value $\Delta > 0$ is chosen numerically. For instance, with fixed $\mu = 5$, the desired value is Δ= 0.91, which makes $P(U \leq Y \leq V) \approx 0.5$ (Fig. 8). Then, we choose $\mu_u = 4.09$ and $\mu_v = 5.91$. With this setting, we have $P(U \leq Y \leq V) = 0.5076142$.

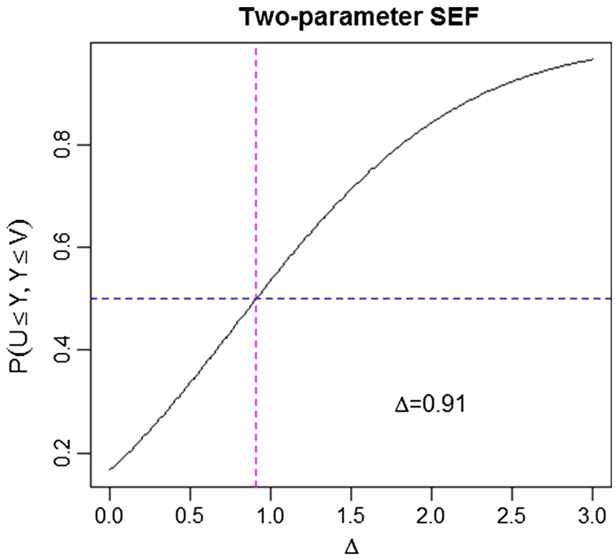

Fig. 8 An example for how to choose the value Δ under the two-parameter SEF

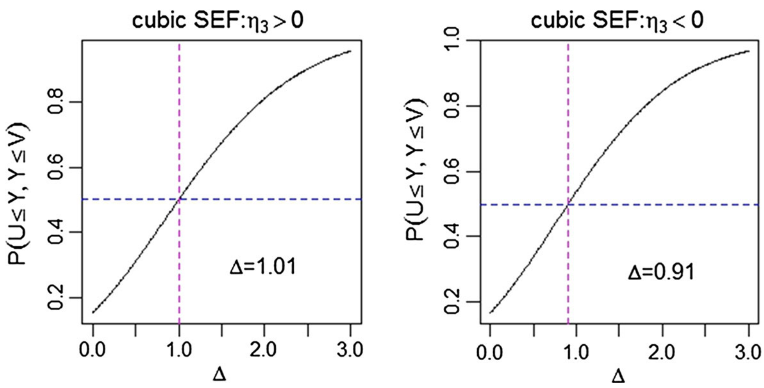

Fig. 9 An example for how to choose the value Δ under the cubic SEF

Data generations for cubic SEF

For the cubic SEF with $\eta_3 > 0$, we consider $U \sim N(\mu_u, 1)$, $V \sim N(\mu_v, 1)$ and

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

where $\phi(\boldsymbol{\eta}) = \log\{\int_{\mathcal{Y}} \exp(\eta_1 y + \eta_2 y^2 + \eta_3 y^3)\, dy\}$. Data are generated by using an inverse transformation, which numerically solves $1 - S_\eta(y) = W$, where $W \sim U(0, 1)$. One can get $P(U \le Y \le V) \approx 0.5$ by making left-truncated and right-truncated percentages equal by setting $\mu_u = \eta_1 - \Delta$, $\mu_v = \eta_1 + \Delta$. Then,

$$P(U \leq Y \leq V) = \int_{-\infty}^{\tau_2} \{1 - \Phi(y - \eta_1 - \Delta)\}\Phi(y - \eta_1 + \Delta) f_\eta(y)\, dy.$$

The desired value $\Delta > 0$ is chosen numerically. For instance, for fixed $\eta_1 = 5$, $\eta_2 = -0.5$, $\eta_3 = 0.005$ and $\tau_2 = 8$, the value of Δ is 1.01 (Fig. 9). Hence, we choose $\mu_u = 3.99$ and $\mu_v = 6.01$. Accordingly, the inclusion probability becomes $P(U \leq Y \leq V) = 0.5035228$.

The other case $\eta_3 < 0$ is similar. Under $\eta_1 = 5$, $\eta_2 = -0.5$, $\eta_3 = -0.005$, and $\tau_1 = 2$, the desired value is $\Delta = 0.91$ (see Fig. 9). Hence, we choose $\mu_u = 4.09$ and $\mu_v = 5.91$. Accordingly, the inclusion probability becomes $P(U \leq Y \leq V) = 0.5027334$.

References

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F. In: Proceedings of the 2nd international symposium on information theory, Akademia Kiado, Budapest, pp 267–281

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

Balakrishnan N, Asit Basu P (1996) The exponential distribution: theory, methods and applications. Taylor & Francis Ltd, USA

Burden RL, Faires JD (2011) Numerical analysis. Cengage Learning, Boston

Chen YH (2009) Weighted Breslow-type and maximum likelihood estimation in semiparametric transformation models. Biometrika 96:235–251

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

Casella G, Berger RL (2002) Statistical inference. Duxbury Thomson Learning, Australia

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

Commenges D (2002) Inference for multi-state models from interval-censored data. Stat Methods Med Res 11:167–182

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

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

Emura T, Konno Y (2012b) A goodness-of-fit tests for parametric models based on dependently truncated data. Comput Stat Data Anal 56:2237–2250

Emura T, Konno Y, Michimae H (2014) Statistical inference based on the nonparametric maximum likelihood estimator under double-truncation. Lifetime Data Anal. doi:10.1007/s10985-014-9297-5

Knight K (2000) Mathematical statistics. Chapman and Hall, Boca Raton

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

Long TH, Emura T (2014) A control chart using copula-based Markov chain models. J Chin Stat Assoc 52:466–496

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 Stati 22:567–583

Moreira C, de Uña-Álvarez J, Van Keilegom I (2014) Goodness-of-fit tests for a semiparametric model under random double truncation. Comput Stat. doi:10.1007/s00180-014-0496-z

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

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

R Development Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, R version 3:2

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

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

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

Stute W, González-Manteiga W, Quindimil MP (1993) Bootstrap based goodness-of-fit-tests. Metrika 40:243–256

Zhu H, Wang MC (2012) Analyzing bivariate survival data with interval sampling and application to cancer epidemiology. Biometrika 99:345–361