Flexible two-piece distributions for right censored survival data

Peer-reviewed author version

EWNETU, Worku Biyadgie; Gijbels, Irene & VERHASSELT, Anneleen (2023) Flexible two-piece distributions for right censored survival data. In: Lifetime data analysis, 29 (1), p. 34-65.

DOI: 10.1007/s10985-022-09574-4
Handle: http://hdl.handle.net/1942/38675

[redacted]

Flexible two-piece distributions for right censored survival data

Worku B. Ewnetu[1,2] · **Irène Gijbels**[1] · **Anneleen Verhasselt**[2]

[redacted]

Abstract An important complexity in censored data is that only partial information on the variables of interest is observed. In recent years, a large family of asymmetric distributions and maximum likelihood estimation for the parameters in that family has been studied, in the complete data case. In this paper, we exploit the appealing family of quantile-based asymmetric distributions to obtain flexible distributions for modelling right censored survival data. The flexible distributions can be generated using a variety of symmetric distributions and monotonic link functions. The interesting feature of this family is that the location parameter coincides with an index-parameter quantile of the distribution. This family is also suitable to characterize different shapes of the hazard function (constant, increasing, decreasing, bathtub and upside-down bathtub or unimodal shapes). Statistical inference is done for the whole family of distributions. The parameter estimation is carried out by optimizing a non-differentiable likelihood function. The asymptotic properties of the estimators are established. The finite-sample performance of the proposed method and the impact of censorship are investigated via simulations. Finally, the methodology is illustrated on two real data examples (times to weaning in breast-fed data and German Breast Cancer data).

Keywords censored data · complete data · flexible distributions · hazard function · maximum likelihood · quantile

1 Introduction

In the analysis of lifetime or time to an event data, we are confronted with a strictly positive variable of interest, which is often asymmetric in nature. An important complexity in censored lifetime data is that only partial information on the variables of interest is observed. In particular, in a right censored data context the survival time of interest is not observed, but only known to exceed the observed time for some of the studied objects. Several classical parametric probability distributions have been used to model both types of lifetime data, complete and incomplete data. They include exponential, Weibull, log-normal, log-logistic, and gamma distributions. See for example, Nelson (1982), Lawless (2003), and Meeker and Escobar (2014), among others.

[1] *Department of Mathematics and Leuven Statistics Research Center (LStat), KU Leuven, Belgium*
[2] *Center for Statistics, Data Science Institute, Hasselt University, Belgium*
E-mail: anneleen.verhasselt@uhasselt.be

However, modelling lifetime data using the aforementioned standard distributions may not give enough flexibility to describe complex data. Furthermore, they are incapable to describe all types of nonmonotonic hazard shapes (Rubio and Hong, 2016; Vallejos and Steel, 2015). To remediate these drawbacks, these distributions have been used as baseline to include a new additional parameter(s), which leads to capture skewness and the tail behavior of the distribution. For example, a power of an exponential distributed random variable gives a two-parameter Weibull distribution; whereas the power of a Weibull distributed random variable leads to a new Weibull family (Marshall and Olkin, 1997; Gupta and Kundu, 2001). A more flexible family, the exponentiated Weibull family has been introduced by Mudholkar and Srivastava (1993). This family has as particular cases the exponentiated exponential family and the Weibull family. The exponentiated Weibull family is able to accommodate monotonic and nonmonotonic hazard rate functions, and can provide better model fits than the classical two-parameter Weibull, gamma, and log-normal distributions (Mudholkar and Srivastava, 1993; Mudholkar et al., 1995; Pal et al., 2006). The log-normal power distribution studied by Reed (2011), also exhibits a variety of shapes for the hazard function, including a bathtub shape. According to Reed (2011), this typical flexible distribution includes the log-normal and the power distribution as special cases.

Recently, Khan and Khosa (2016) proposed a generalized log-logistic (GLL) distribution to handle both monotone and nonmonotone hazard functions. The GLL distribution reduces to a two parameters Weibull distribution in the limit. More recently, mixture Weibull distributions have been proposed by Liao and Liu (2019), including the cure model as a special case. Despite these models are flexible compared to the commonly used parametric models and useful to accommodate a cure component, a drawback is the number of mixtures needs to be chosen.

Log-symmetric distributions such as log-normal, log-logistic, log-Laplace and likes are used for describing distributions of a right-skewed positive random variable. They are basically generated from a distribution of a random variable whose logarithm has a symmetric distribution. On one hand the newly transformed variable via a logarithmic or any other monotonic function, might be right-skewed and left-skewed, meaning that the symmetric property may not be achieved after transformation. On the other hand, as stated by Hougaard (1999), lifetime random variables may also have a left-skewed distribution. In such cases, the log-symmetric family of distributions does not always capture important characteristics of the data that we might want to investigate. For instance, they are not flexible enough to capture complex survival curves such as bathtub hazard shapes which can be observed in time to event data. The purpose of this study is thus to propose a parametric family of distributions for lifetime data which may be used as alternative to the aforementioned distributions in producing flexibility in terms of describing complex survival curves. In fact, our proposed general family includes the existing log-symmetric family as a special case.

So far there are two common methods to construct more flexible probability distributions for lifetime data. The first one is adding one or more shape parameters to an existing underlying distribution with positive support $(0,+\infty)$, and the second method is a transformation of a symmetric distribution, see Marshall and Olkin (1997), Rubio and Hong (2016), and references therein.

Our focus in this paper is on the following family of two-piece asymmetric distributions, in which a density takes the form

$$\tilde{f}_{\alpha}(z;\mu,\phi) = \frac{2\alpha(1-\alpha)}{\phi} \begin{cases} f_0\big((1-\alpha)(\frac{\mu-z}{\phi})\big) & \text{if } z < \mu \\ f_0\big(\alpha(\frac{z-\mu}{\phi})\big) & \text{if } z \geq \mu, \end{cases} \tag{1.1}$$

where $\mu \in \mathbb{R}$ and $\phi \in (0,+\infty)$ are, respectively, the location and scale parameters, and $\alpha \in (0,1)$ is the index-parameter which controls the allocation of mass of the distribution to the left or right of the mode μ. Herein $f_0(.)$ denotes a symmetric around zero unimodal density, which will be referred to as the reference density. This family introduced by Nassiri and Loris (2013) and studied in detail by Gijbels et al. (2019a) is referred to as the family of *quantile-based asymmetric* (QBA) densities, since the location parameter μ coincides with the α-th quantile of the distribution. This is one of the interesting features of this family, allowing for inference about quantiles of the distribution rather than only the mean.

The approach in this paper has two ingredients: (i) a symmetric unimodal reference density f_0 that serves to model a two-piece asymmetric density; (ii) a link function g that provides the transformation from the half real line $(0,+\infty)$ to $\mathbb{R}$. Each pair (f_0,g) leads to a specific flexible model. The built-in flexibility is obtained from the two ingredients, and it allows for various shapes of the hazard function within one and the same parametric family. The assumptions imposed on (f_0,g) are very mild, and statistical inference is carried out in this unified framework, covering a wide variety of flexible distributions. Our contribution to the existing literature is therefore three-fold: (i) we allow for a large family of distributions (by allowing a general symmetric reference density and allowing a general link function), (ii) the statistical inference results obtained hold for all members of this large family, i.e. we prove the properties at once for all members (instead of proving the results for each distribution separately), and (iii) if the index parameter α equals 0.5, then the proposed general family with log-link function reduces to a log-symmetric family based on the reference density f_0; hence, the entire class of log-symmetric densities is a special case of the considered general family.

The rest of the paper is structured in the following manner. The QBA family of distributions in the context of lifetime data is discussed in Section 2. Section 3 focuses on the statistical inference, which includes parameter estimation and asymptotic properties of the estimators. The performance of the proposed methods in finite–samples is illustrated via simulations in Section 4. The fifth section presents the application on two real data examples. Finally, we discuss the results and conclude the findings in Section 6. In the Supplementary Material we provide details on some examples of the GQBA family, proofs for some of the theoretical results, details for a data example, and additional simulation results.

2 Generalized QBA family and lifetime data

Consider a random variable T, typically a lifetime, or time to an event. In right random censoring, one observes the couple

$$(\min\{T,C\}, I\{T \leq C\}) = (Y,\Delta),$$

where C is the censoring random variable, and $\Delta = I\{T \leq C\})$, is the censoring indicator. Note that Δ is a Bernoulli distributed random variable with success probability $P(\Delta = 1)$.

We consider an increasing and differentiable function $g : (0,+\infty) \to \mathbb{R}$. Let Z be a random variable on $\mathbb{R}$ with density given by (1.1) and let $Z = g(T)$. Applying the second approach (i.e. transformation of a symmetric distribution) mentioned in Section 1, the distribution of the lifetime random variable T can be obtained from the distribution of Z. Denote $\mu = g(\eta)$. The probability density function of T is then given by

$$f_\alpha(t;\eta,\phi)=\frac{2\alpha(1-\alpha)g'(t)}{\phi}\begin{cases} f_0\{(1-\alpha)\big(\frac{g(\eta)-g(t)}{\phi}\big)\} & \text{if } t<\eta \\ f_0\{\alpha\big(\frac{g(t)-g(\eta)}{\phi}\big)\} & \text{if } t\geq\eta, \end{cases} \tag{2.1}$$

where $\eta=g^{-1}(\mu)\in\mathbb{R}_0^+=(0,+\infty)$ is the location parameter for T, $\phi\in\mathbb{R}_0^+$, and $\alpha\in(0,1)$. If $\alpha=0.5$, then (2.1) reduces to a class of location-scale families with the underlying symmetric density f_0. In addition, the density function defined in (1.1) is a special case of (2.1) taking the identity link function $g(t)=t$. Hereafter, we refer to (2.1) as the family of *generalized quantile-based asymmetric* (GQBA in short) distributions.

2.1 Basic properties

We begin by establishing some basic quantities for the GQBA family of survival distributions. Let F_0, S_0, h_0 and Λ_0 denote, respectively, the cumulative distribution, survival, hazard, and cumulative hazard functions for the reference symmetric density f_0. The proofs of Theorems 2.1 and 2.2 are deferred to Appendix A.

Theorem 2.1 *Let T be a lifetime random variable with density $f_\alpha(t;\eta,\phi)$, then the cumulative distribution function of T is given by*

$$F_\alpha(t;\eta,\phi)=\begin{cases} 2\alpha F_0\{(1-\alpha)\big(\frac{g(t)-g(\eta)}{\phi}\big)\} & \text{if } t<\eta \\ 2\alpha-1+2(1-\alpha)F_0\{\alpha\big(\frac{g(t)-g(\eta)}{\phi}\big)\} & \text{if } t\geq\eta, \end{cases} \tag{2.2}$$

and the τ-th quantile function of T for any $\tau\in(0,1)$ is given by

$$F_\alpha^{-1}(\tau)=\begin{cases} g^{-1}\{g(\eta)+\frac{\phi}{1-\alpha}F_0^{-1}\big(\frac{\tau}{2\alpha}\big)\} & \text{if } \tau<\alpha \\ g^{-1}\{g(\eta)+\frac{\phi}{\alpha}F_0^{-1}\big(\frac{1+\tau-2\alpha}{2(1-\alpha)}\big)\} & \text{if } \tau\geq\alpha. \end{cases} \tag{2.3}$$

In particular $F_\alpha^{-1}(\alpha)=\eta$, which indicates that the α-th quantile of the distribution equals the location parameter η. It is also seen from (2.2) that

$$\frac{P(T<\eta)}{P(T\geq\eta)}=\frac{\alpha}{1-\alpha}\Leftrightarrow(1-\alpha)P(T<\eta)=\alpha P(T\geq\eta),$$

which illustrates that the index-parameter α controls the allocation of mass to each side of the location parameter η. α less (greater) than 0.5 reveals a right (left) skewed distribution, respectively.

Theorem 2.2 *Let T be a lifetime random variable with density $f_\alpha(t;\eta,\phi)$, then*

i). the survival function of T is

$$S_\alpha(t;\eta,\phi) = \begin{cases} 1-2\alpha S_0\{(1-\alpha)(\frac{g(\eta)-g(t)}{\phi})\} & \text{if } t<\eta \\ 2(1-\alpha)S_0\{\alpha(\frac{g(t)-g(\eta)}{\phi})\} & \text{if } t\geq\eta. \end{cases} \tag{2.4}$$

Note: if $\lim_{t\to+\infty} g(t) = +\infty$ *and* $\lim_{t\to 0} g(t) = -\infty$, *then* $\lim_{t\to+\infty} S_\alpha(t;\eta,\phi) = 0$ *and* $\lim_{t\to 0} S_\alpha(t;\eta,\phi) = 1$;

ii). the hazard function of T *is*

$$h_\alpha(t;\eta,\phi) = \frac{\alpha g'(t)}{\phi} \begin{cases} \frac{2(1-\alpha)h_0\{(1-\alpha)(\frac{g(\eta)-g(t)}{\phi})\}S_0\{(1-\alpha)(\frac{g(\eta)-g(t)}{\phi})\}}{1-2\alpha S_0\{(1-\alpha)(\frac{g(\eta)-g(t)}{\phi})\}} & \text{if } t<\eta \\ h_0\{\alpha(\frac{g(t)-g(\eta)}{\phi})\} & \text{if } t\geq\eta; \end{cases} \tag{2.5}$$

iii). the cumulative hazard function of T *is*

$$\Lambda_\alpha(t;\eta,\phi) = \begin{cases} -\ln\{1-2\alpha S_0[(1-\alpha)(\frac{g(\eta)-g(t)}{\phi})]\} & \text{if } t<\eta \\ -\ln\{2(1-\alpha)\}+\Lambda_0\{\alpha(\frac{g(t)-g(\eta)}{\phi})\} & \text{if } t\geq\eta. \end{cases} \tag{2.6}$$

The GQBA family is suitable to describe all common hazard shapes (constant, increasing, decreasing, bathtub and inverted bathtub or unimodal shapes). This is illustrated in the next section.

2.2 Shape of the hazard function

The hazard function $h_\alpha(\cdot;\eta,\phi)$ is continuous but not differentiable at $t=\eta$. Therefore we investigate its shape to the left and right of η separately, for a concave (and increasing on $\mathbb{R}_0^+$) link function $g(\cdot)$.

For $t\geq\eta$: Let $z=\alpha(\frac{g(t)-g(\eta)}{\phi})$ with $t=g^{-1}(g(\eta)+\frac{\phi}{\alpha}z)$ and rewrite the log-hazard as a function of $z\in\mathbb{R}^+$, denoted by $\tilde{h}_r(z)$

$$\tilde{h}_r(z) = \ln(\alpha)-\ln(\phi)+\ln\left(g'\{g^{-1}(g(\eta)+\frac{\phi}{\alpha}z)\}\right) \ + \ \ln\{h_0(z)\}.$$

Then the derivative of this function is

$$\begin{aligned} \tilde{h}_r'(z) &= \frac{\phi}{\alpha}\frac{g''\{g^{-1}(g(\eta)+\frac{\phi}{\alpha}z)\}}{g'\{g^{-1}(g(\eta)+\frac{\phi}{\alpha}z)\}}\{g^{-1}\}'(g(\eta)\ +\ \frac{\phi}{\alpha}z)\ +\ (\ln h_0)'(z) \\ &= \underbrace{\frac{\phi}{\alpha}}_{>0}\underbrace{g''\{g^{-1}(g(\eta)+\frac{\phi}{\alpha}z)\}}_{\leq 0}\underbrace{\left(\{g^{-1}\}'(g(\eta)\ +\ \frac{\phi}{\alpha}z)\right)^2}_{\geq 0}\ +\ (\ln h_0)'(z). \end{aligned}$$

Therefore the shape of the (log-)hazard depends on the sign of $(\ln h_0)'(z)$.

If f_0 is log-convex (e.g. the Arc-sine distribution) on an interval (l,u), then $h_0(.)$ is monotone decreasing on (l,u) (Bagnoli and Bergstrom, 2005) and therefore $\tilde{h}_r(z)$ is decreasing on (l,u). Therefore the log-hazard is constant or decreasing (not increasing) on (l,u).

If $f_0(.)$ is log-concave (e.g. standard normal, Laplace or logistic distribution) on (l,u), then h_0 is increasing on (l,u)). If $\ln h_0$ is constant on (l,u), then $h_\alpha(\cdot;\eta,\phi)$ is decreasing on (l,u), depending on the magnitude of $(\ln h_0)'(z)$, the link function, α and ϕ the log-hazard can be increasing or decreasing on (l,u).

For $t<\eta$: Denote $z=(1-\alpha)\big(\frac{g(\eta)-g(t)}{\phi}\big)$ and rewrite the log-hazard as a function of $z\in\mathbb{R}^+$, denoted by $\tilde{h}_l(z)$

$$\begin{aligned}\tilde{h}_l(z) = \ln\{2\alpha(1-\alpha)\} \;-\; \ln(\phi) \;+\; \ln\left(g'\big\{g^{-1}\big(g(\eta)-\frac{\phi}{1-\alpha}z\big)\big\}\right)\\ +\; \ln\{h_0(z)\} \;+\; \ln\{S_0(z)\} \;-\; \ln\big(1-2\alpha S_0(z)\big),\end{aligned}$$

where $2\alpha S_0(z)<1$ to obtain a valid hazard value. The first derivative of $\tilde{h}_l(z)$ with respect to z becomes

$$\begin{aligned}\tilde{h}_l'(z) = -\underbrace{\frac{\phi}{1-\alpha}}_{>0}\underbrace{g''\Big\{g^{-1}\big(g(\eta)-\frac{\phi}{1-\alpha}z\big)\Big\}}_{\leq 0}\underbrace{\Big(\big\{g^{-1}\big\}'\big(g(\eta)-\frac{\phi}{1-\alpha}z\big)\Big)^2}_{\geq 0}\\ +\; (\ln h_0)'(z) - h_0(z)\big\{\underbrace{\frac{1}{1-2\alpha S_0(z)}}_{>0}\big\}.\end{aligned}$$

If $\ln h_0$ is constant/decreasing/increasing on (l,u), then the log-hazard can either be increasing or decreasing on (l,u), depending on the magnitude of $h_0(z)$, $S_0(z)$, the link function, α and ϕ. From this it is clear that the link function, the reference density, the scale and index-parameters are characterizing the behavior of the hazard function for the GQBA distributions.

For example, for the GQBA logistic distribution with log-link: $g(t)=\ln(t)$, $h_0(z)=\frac{1}{1+e^{-z}}$. Therefore $h_\alpha(.;\eta,\phi)$ is always decreasing in the right-tail for $\phi>\alpha$. On the other hand, for the left-hand tail ($t<\eta$) of the distribution, when $\phi\geq\alpha/2$ (respectively $\phi<\alpha/2$) the hazard is increasing (respectively decreasing) as $t\to\eta$. See also Fig. 2.1.

2.3 Examples of distributions in the GQBA family

We give some illustrative examples of the GQBA family of survival distributions with their basic characteristics. There are essentially two basic ingredients to formulate the GQBA distribution; the reference symmetric density f_0 and the link function g. In the examples, we consider normal and logistic standard distributions. We, in particular, look at two appealing link functions, the classical logarithmic link $g(t)=\ln(t)$ and the link function $g(t;\lambda)=\ln\big(e^{\lambda t}-1\big)$. The second link function is the logit of the cumulative distribution function of an exponential distribution. That is, $W\sim\text{Expo}(\lambda),\lambda>0,W>0$, with cumulative distribution function $F_\lambda(t)=1-e^{-\lambda t}$, which gives a monotonically increasing link function $g(t;\lambda)=\text{logit}\big(F_\lambda(t)\big)=\ln(e^{\lambda t}-1)$, and its deriva-

tive $g'(t;\lambda) = \dfrac{\lambda e^{\lambda t}}{e^{\lambda t}-1}$. This example stands for an illustration for a whole class of link functions obtained as the logit of the cumulative distribution function of any positive random variable W. Throughout, we refer to this function in general as a logit-link function.

In Table S1 in the Supplementary Material we summarize some characteristics of the GQBA family of asymmetric normal $\mathcal{AN}(\cdot)$ and asymmetric logistic $\mathcal{AL}(\cdot)$ distributions with $g(t) = \ln(t)$. For the logit-link function, for example, the probability density function of $\mathcal{AL}(\eta,\phi,\alpha,\lambda)$ is given by

$$f_\alpha(t;\eta,\phi,\lambda) = \frac{2\alpha(1-\alpha)\lambda e^{\lambda t}}{\phi(e^{\lambda t}-1)} \begin{cases} \dfrac{\left(\frac{e^{\lambda t}-1}{e^{\lambda\eta}-1}\right)^{\left(\frac{1-\alpha}{\phi}\right)}}{\left\{1+\left(\frac{e^{\lambda t}-1}{e^{\lambda\eta}-1}\right)^{\left(\frac{1-\alpha}{\phi}\right)}\right\}^2} & \text{if } t<\eta \\ \dfrac{\left(\frac{e^{\lambda t}-1}{e^{\lambda\eta}-1}\right)^{-\frac{\alpha}{\phi}}}{\left\{1+\left(\frac{e^{\lambda t}-1}{e^{\lambda\eta}-1}\right)^{-\frac{\alpha}{\phi}}\right\}^2} & \text{if } t\geq\eta\ . \end{cases} \tag{2.7}$$

The exponential distribution Expo(λ) arises as a special case of (2.7) when $\alpha = 0.5, \phi = 0.5$ and $\eta = \frac{1}{\lambda}\ln(2)$. Furthermore, the distribution $\mathcal{AL}(\eta,\phi,\alpha,\lambda)$ gives the Logistic–Exponential distribution (Lan and Leemis, 2008) when taking $\alpha = 0.5, \phi = \frac{1}{2\kappa}$ and $\lambda\eta = \ln(2)$, where $\kappa > 0$ is the shape parameter.

In addition to the logarithmic and logit-link functions, other useful link functions satisfying the properties of $g(.)$ can be proposed, and used to generate a variety of distributions within the GQBA family. For instance, Slymen and Lachenbruch (1984) proposed modified Weibull and modified log-logistic distributions using the following link functions

$$g_1(t;\lambda) = \frac{t^\lambda - t^{-\lambda}}{2\lambda}, \quad \lambda > 0,$$
$$g_2(t;\lambda,\beta) = \ln\left\{\ln[1+(\lambda t)^\beta]\right\}, \lambda > 0, \beta > 0.$$

More specifically, the function $g_1(t;\lambda)$ has the important property that $\lim\limits_{\lambda\to 0}\left(\dfrac{t^\lambda - t^{-\lambda}}{2\lambda}\right) = \ln(t)$, which thus generalizes the logarithmic link function by introducing an additional parameter λ.

An important remark to be made is that if a link function depends on an unknown parameter we will estimate this parameter via maximum likelihood estimation (MLE). See Sections 4 and 5. Such a situation occurs with the logit-link function $g(t;\lambda)$, as well as with $g_1(t;\lambda)$ and $g_2(t;\lambda,\beta)$. This is also the reason why we denote the asymmetric logistic density with link $g(t) = \ln(t)$ by $\mathcal{AL}(\eta,\phi,\alpha)$ whereas the asymmetric logistic density with link function $g(t;\lambda) = \ln(e^{\lambda t}-1)$ is denoted by $\mathcal{AL}(\eta,\phi,\alpha,\lambda)$.

In order to illustrate the flexibility of the GQBA family, in particular in terms of the hazard shapes as well as the shapes of survival function, we present in Fig. 2.1 the case of the asymmetric logistic distribution for some parameter values and link function $g(t) = \ln(t)$ for (a) and $g(t;\lambda) = \ln(e^{\lambda t}-1)$ for (b) and (c). As we can see from the plots, different hazard shapes (constant, decreasing, increasing, bathtub, inverted bathtub) are obtained for one particular member of the GQBA family of distributions.

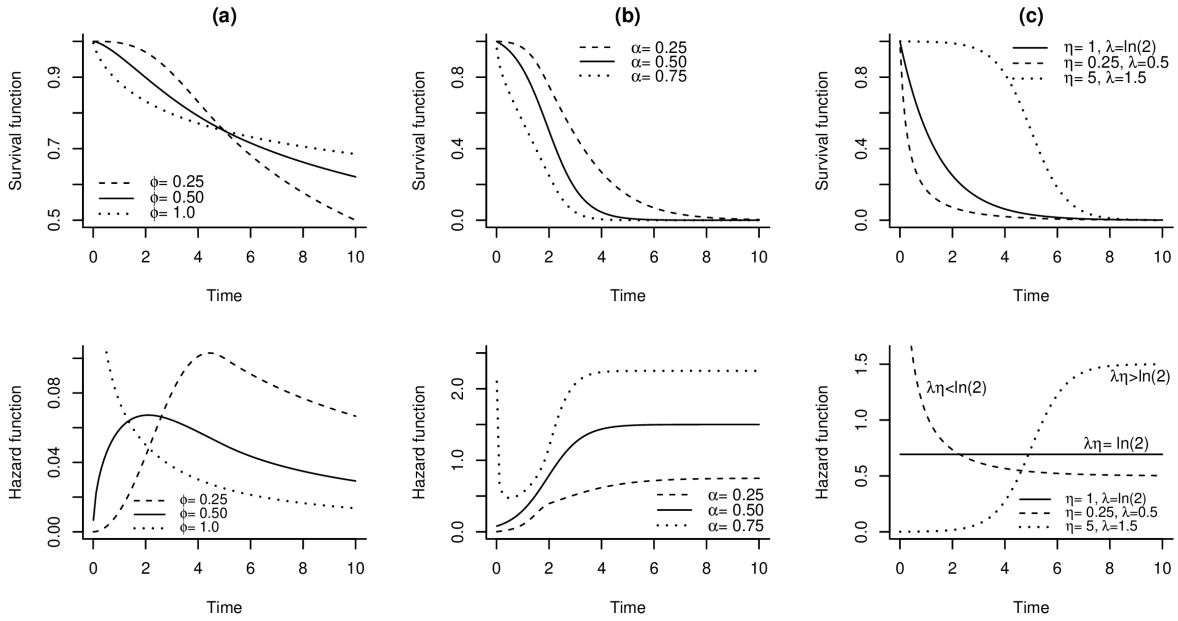

Fig. 2.1 Survival function (first row) and hazard function (second row) for one member of the GQBA family of distributions: (a) $\mathcal{AL}(\eta = 5, \phi, \alpha = 0.25)$, (b) $\mathcal{AL}(\eta = 2, \phi = 0.5, \alpha, \lambda = 1.5)$, and (c) $\mathcal{AL}(\eta, \phi = 0.5, \alpha = 0.5, \lambda)$.

3 Statistical inference

3.1 MLE of the parameters

Let $(y_1, \delta_1), \ldots, (y_n, \delta_n)$ be an i.i.d. sample from (Y, Δ). Let $\ell(y_i, \delta_i; \boldsymbol{\theta}) = \ln f_{Y,\Delta}(y_i, \delta_i; \boldsymbol{\theta})$ denote the log of the joint density function evaluated in the i-th observation. Therefore, the total log-likelihood function of $\boldsymbol{\theta}$ for the n i.i.d. sampled observations is then the sum of $\ell(y_i, \delta_i; \boldsymbol{\theta})$. Under the assumption of *independent and non-informative censoring*, the log-likelihood function with respect to $\boldsymbol{\theta}$ is given by

$$\mathcal{L}_n(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{\delta}) = \sum_{i=1}^{n} \left[\delta_i \ln\{f_\alpha(y_i)\} + (1 - \delta_i) \ln\{S_\alpha(y_i)\} \right]. \tag{3.1}$$

It is observed from (3.1), that the log-likelihood is proportional to the log of the probability density function for an uncensored case and the log of the survival function for a censored case. Consequently, for the inference on the survival time with probability density function as in (2.1), the log-likelihood function of $\boldsymbol{\theta}$ given $(\boldsymbol{y}, \boldsymbol{\delta})$ is written as

$$\begin{aligned}\mathcal{L}_n(\boldsymbol{\theta};\boldsymbol{y},\boldsymbol{\delta}) = &\sum_{i=1}^{n}\delta_i\big\{\ln(2\alpha(1-\alpha))-\ln(\phi)+\ln\{g'(y_i)\}\big\}\\ &+\sum_{i=1}^{n}\delta_i I(y_i<\eta)\ln\big\{f_0\big((1-\alpha)\big[\frac{g(\eta)-g(y_i)}{\phi}\big]\big)\big\}\\ &+\sum_{i=1}^{n}\delta_i I(y_i\geq\eta)\ln\big\{f_0\big(\alpha\big[\frac{g(y_i)-g(\eta)}{\phi}\big]\big)\big\}\\ &+\sum_{i=1}^{n}(1-\delta_i) I(y_i<\eta)\ln\big\{1-2\alpha S_0\big((1-\alpha)\big[\frac{g(\eta)-g(y_i)}{\phi}\big]\big)\big\}\\ &+\sum_{i=1}^{n}(1-\delta_i) I(y_i\geq\eta)\ln\big\{2(1-\alpha)S_0\big(\alpha\big[\frac{g(y_i)-g(\eta)}{\phi}\big]\big)\big\}.\end{aligned} \tag{3.2}$$

The MLE estimator of $\boldsymbol{\theta}=(\eta,\phi,\alpha)^T$ is obtained by maximizing the log-likelihood function given in (3.2) with respect to $\boldsymbol{\theta}$. That is,

$$\widehat{\boldsymbol{\theta}}_n = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathcal{L}_n(\boldsymbol{\theta};\boldsymbol{y},\boldsymbol{\delta}),$$

where $\widehat{\boldsymbol{\theta}}_n=(\widehat{\eta},\widehat{\phi},\widehat{\alpha})^T$ is the MLE estimator of $\boldsymbol{\theta}$ and $\boldsymbol{\Theta}=\mathbb{R}_0^+\times\mathbb{R}_0^+\times(0,1)$ is the parameter space of $\boldsymbol{\theta}$. However, the log-likelihood function is not differentiable with respect to η at $\eta=y_i$, for $i=1,2,...,n$. The log-likelihood function is differentiable with respect to ϕ and α, although the MLE does not have a closed form. For such type of optimization, the estimation is usually done by using direct search or derivative free algorithms. These derivative free algorithms only need the objective functions to be evaluated; meaning that differentiability is not needed during optimization. For example, Bottai (2010) and Su (2016) used the Nelder-Mead simplex algorithm in censored data analysis. In this study we also apply the Nelder-Mead simplex algorithm using the `nloptr` optimization package in the statistical software `R`.

3.2 Asymptotic properties of MLE estimators

The following notations turn out to be useful in the sequel.

$$z_1=(1-\alpha)\big(\frac{g(\eta)-g(y)}{\phi}\big)\Rightarrow y\equiv g_{1\boldsymbol{\theta}}(z_1)=g^{-1}\big\{g(\eta)-\frac{\phi}{1-\alpha}z_1\big\}, \tag{3.3}$$

and

$$z_2=\alpha\big(\frac{g(y)-g(\eta)}{\phi}\big)\Rightarrow y\equiv g_{2\boldsymbol{\theta}}(z_2)=g^{-1}\big\{g(\eta)+\frac{\phi}{\alpha}z_2\big\}. \tag{3.4}$$

The following major assumptions are considered in establishing consistency and asymptotic normality properties of the MLE. Let $\boldsymbol{\theta}_0=(\eta_0,\phi_0,\alpha_0)^T$ be the true value of the parameter $\boldsymbol{\theta}$.

Assumptions

(B1) Let $\boldsymbol{\Theta}_R=[\eta_l,\eta_u]\times[\phi_l,\phi_u]\times[\alpha_l,\alpha_u]$, with $0<\eta_l\leq\eta\leq\eta_u<\infty$, $0<\phi_l\leq\phi\leq\phi_u<\infty$ and $0<\alpha_l\leq\alpha\leq\alpha_u<1$, be a compact subset of $\boldsymbol{\Theta}$, and assume that $\boldsymbol{\theta}_0\in\mathring{\boldsymbol{\Theta}}_R$ with $\mathring{\boldsymbol{\Theta}}_R$ the interior of $\boldsymbol{\Theta}_R$.

(B2) $g(.)\colon \mathbb{R}_0^+ \to \mathbb{R}$ is monotone increasing and differentiable function, such that $\lim_{t\to 0} g(t) = -\infty$ and $\lim_{t\to+\infty} g(t) = +\infty$.

(B3) The symmetric around zero reference density f_0 is bounded, differentiable and satisfies $\int_0^\infty z f_0'(z)dz = -\frac{1}{2}$ or $\lim_{z\to\infty} z f_0(z) = 0$.

(B4) $\lim_{z\to+\infty} z^{r-1} f_0(z) G[g_{l\boldsymbol{\theta}}(z)] = 0$ for $l, r = 1, 2$.

(B5) $\xi_{l,r}(\boldsymbol{\theta}) = \int_0^\infty z^{r-1} f_0(z) dG[g_{l\boldsymbol{\theta}}(z)] < \infty$ for $l, r = 1, 2$.

(B6) $\gamma_r = \int_0^\infty z^{r-1} \frac{\{f_0'(z)\}^2}{f_0(z)} dz < \infty$ for $r = 1, 2, 3$.

(B7) $\gamma_{l,r}(\boldsymbol{\theta}) = \int_0^\infty z^{r-1} \frac{\{f_0'(z)\}^2}{f_0(z)} G[g_{l\boldsymbol{\theta}}(z)]dz < \infty$ for $l = 1, 2$, and $r = 1, 2, 3$.

(B8) $\kappa_{1,r}(\boldsymbol{\theta}) = \int_0^\infty z^{r-1} \frac{\{f_0(z)\}^2}{1-2\alpha S_0(z)} dG[g_{1\boldsymbol{\theta}}(z)] < \infty$ for $r = 1, 2, 3$.

(B9) $\kappa_{2,r}(\boldsymbol{\theta}) = \int_0^\infty z^{r-1} \frac{\{f_0(z)\}^2}{S_0(z)} dG[g_{2\boldsymbol{\theta}}(z)] < \infty$ for $r = 1, 2, 3$.

(B10) $\varphi_{1,r}(\boldsymbol{\theta}) = \int_0^\infty z^{r-1} \frac{f_0(z)S_0(z)}{1-2\alpha S_0(z)} dG[g_{1\boldsymbol{\theta}}(z)] < \infty$ for $r = 1, 2$,

(B11) $\varphi_2(\boldsymbol{\theta}) = \int_0^\infty \frac{\{S_0(z)\}^2}{1-2\alpha S_0(z)} dG[g_{1\boldsymbol{\theta}}(z)] < \infty$.

The first Assumption (B1) is postulated to ensure that any continuous function of $\boldsymbol{\theta}$ in a bounded parameter space $\boldsymbol{\Theta}_R$ is also bounded; which is required in both consistency and asymptotic normality. Assumptions (B2)-(B5) are essentially assumed to ensure that the expected value of the score function has a unique zero at the true value of the parameter $\boldsymbol{\theta}_0 \in \mathring{\boldsymbol{\Theta}}_R$. The other Assumptions (B6)-(B11) are also vital for the existence of the Fisher-information matrix.

All these assumptions are rather mild conditions. We check the conditions for the GQBA logistic family of survival distribution and an exponential censoring distribution. See Example 3.1 and Section S3, in the Supplementary Material.

Proposition 3.1 *Let Assumptions (B2) to (B6) hold for $\boldsymbol{\theta}_0$. Then the expectation of the score function at the true parameter value $\boldsymbol{\theta}_0$ is zero. That is,*

$$\mathrm{E}_{Y,\Delta}\left[\frac{\partial}{\partial\boldsymbol{\theta}}\ell(\boldsymbol{\theta};Y,\Delta)\right]\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}$$

The proofs of Theorems 3.1 and 3.2 are provided in Appendix B, whereas the proofs of Propositions 3.1 and 3.2 are given in Section S2 in the Supplementary Material of the paper.

Theorem 3.1 *Let $\widehat{\boldsymbol{\theta}}_n = (\widehat{\eta}_n, \widehat{\phi}_n, \widehat{\alpha}_n)^T$ denote any sequence of estimators for $\boldsymbol{\theta}$ that maximizes the log-likelihood function (3.2). If Assumptions (B1)-(B5) hold for $\boldsymbol{\theta}_0$, then $\widehat{\boldsymbol{\theta}}_n$ is strongly consistent for $\boldsymbol{\theta}_0$. That is, $\widehat{\boldsymbol{\theta}}_n \xrightarrow{a.s} \boldsymbol{\theta}_0$, as $n \to \infty$.*

Proposition 3.2 *Suppose Assumptions (B2)-(B11) hold for $\boldsymbol{\theta}$, then the Fisher information matrix $\mathcal{I}(\boldsymbol{\theta}) = \left[\mathrm{E}_{Y,\Delta}\left(\frac{\partial}{\partial\theta_i}\ell(\boldsymbol{\theta};Y,\Delta)\cdot\frac{\partial}{\partial\theta_j}\ell(\boldsymbol{\theta};Y,\Delta)\right)\right]_{i,j=1,2,3}$ is finite and given by*

$$\mathcal{I}(\boldsymbol{\theta}) = \begin{bmatrix} 2\alpha(1-\alpha)\left(\frac{g'(\eta)}{\phi}\right)^2\gamma_1 + \mathcal{R}_{11}(\boldsymbol{\theta}) & 0 + \mathcal{R}_{12}(\boldsymbol{\theta}) & \frac{-2g'(\eta)}{\phi}\gamma_2 + \mathcal{R}_{13}(\boldsymbol{\theta}) \\ 0 + \mathcal{R}_{12}(\boldsymbol{\theta}) & \frac{1}{\phi^2}(2\gamma_3 - 1) + \mathcal{R}_{22}(\boldsymbol{\theta}) & -\frac{(1-2\alpha)(2\gamma_3-1)}{\phi\alpha(1-\alpha)} + \mathcal{R}_{23}(\boldsymbol{\theta}) \\ \frac{-2g'(\eta)}{\phi}\gamma_2 + \mathcal{R}_{13}(\boldsymbol{\theta}) & -\frac{(1-2\alpha)(2\gamma_3-1)}{\phi\alpha(1-\alpha)} + \mathcal{R}_{23}(\boldsymbol{\theta}) & \frac{2\gamma_3\left[\alpha^3+(1-\alpha)^3\right]-(1-2\alpha)^2}{\alpha^2(1-\alpha)^2} + \mathcal{R}_{33}(\boldsymbol{\theta}) \end{bmatrix}, \tag{3.5}$$

where

$$\begin{aligned}
\mathcal{R}_{11}(\boldsymbol{\theta}) &= -2\alpha(1-\alpha)\left(\frac{g'(\eta)}{\phi}\right)^2\Big\{(1-\alpha)\left[\gamma_{1,1}(\boldsymbol{\theta}) + 2\alpha\kappa_{1,1}(\boldsymbol{\theta})\right] + \alpha\left[\gamma_{2,1}(\boldsymbol{\theta}) - \kappa_{2,1}(\boldsymbol{\theta})\right]\Big\}, \\
\mathcal{R}_{12}(\boldsymbol{\theta}) &= \frac{2\alpha(1-\alpha)g'(\eta)}{\phi^2}\Big\{\gamma_{1,2}(\boldsymbol{\theta}) - \gamma_{2,2}(\boldsymbol{\theta}) - \xi_{1,1}(\boldsymbol{\theta}) + \xi_{2,1}(\boldsymbol{\theta}) + 2\alpha\kappa_{1,2}(\boldsymbol{\theta}) + \kappa_{2,2}(\boldsymbol{\theta})\Big\}, \\
\mathcal{R}_{13}(\boldsymbol{\theta}) &= \frac{2g'(\eta)}{\phi}\Big\{\alpha\gamma_{1,2}(\boldsymbol{\theta}) + (1-\alpha)\gamma_{2,2}(\boldsymbol{\theta}) + (1-2\alpha)\xi_{1,1}(\boldsymbol{\theta}) - (1-\alpha)\xi_{2,1}(\boldsymbol{\theta}) \\
&\qquad + 2\alpha^2\kappa_{1,2}(\boldsymbol{\theta}) - (1-\alpha)\kappa_{2,2}(\boldsymbol{\theta}) + 2\alpha(1-\alpha)\varphi_{1,1}(\boldsymbol{\theta})\Big\}, \\
\mathcal{R}_{22}(\boldsymbol{\theta}) &= -\frac{2}{\phi^2}\Big\{\alpha\left[\gamma_{1,3}(\boldsymbol{\theta}) - 2\xi_{1,2}(\boldsymbol{\theta}) + 2\alpha\kappa_{1,3}(\boldsymbol{\theta})\right] + (1-\alpha)\left[\gamma_{2,3}(\boldsymbol{\theta}) - 2\xi_{2,2}(\boldsymbol{\theta}) - \kappa_{2,3}(\boldsymbol{\theta})\right]\Big\} \\
&\qquad - \alpha\int_0^\infty f_0(z)G[g_{1\boldsymbol{\theta}}(z)]dz - (1-\alpha)\int_0^\infty f_0(z)G[g_{2\boldsymbol{\theta}}(z)]dz, \\
\mathcal{R}_{23}(\boldsymbol{\theta}) &= -\frac{2}{\alpha(1-\alpha)\phi}\Big\{\alpha^2\gamma_{1,3}(\boldsymbol{\theta}) - (1-\alpha)^2\gamma_{2,3}(\boldsymbol{\theta}) + \alpha(1-3\alpha)\xi_{1,2}(\boldsymbol{\theta}) \\
&\qquad + 2(1-\alpha)^2\xi_{2,2}(\boldsymbol{\theta}) + 2\alpha^3\kappa_{1,3}(\boldsymbol{\theta}) + (1-\alpha)^2\kappa_{2,3}(\boldsymbol{\theta}) + 2\alpha^2(1-\alpha)\varphi_{1,2}(\boldsymbol{\theta}) \\
&\qquad - \alpha^2\int_0^\infty f_0(z)G[g_{1\boldsymbol{\theta}}(z)]dz + (1-\alpha)^2\int_0^\infty f_0(z)G[g_{2\boldsymbol{\theta}}(z)]dz\Big\}, \\
\mathcal{R}_{33}(\boldsymbol{\theta}) &= -\frac{2}{\alpha^2(1-\alpha)^2}\Big\{\alpha^2(1-\alpha)G(\eta) + \alpha^3\gamma_{1,3}(\boldsymbol{\theta}) + (1-\alpha)^3\gamma_{2,3}(\boldsymbol{\theta}) \\
&\qquad + 2\alpha^2(1-2\alpha)\xi_{1,2}(\boldsymbol{\theta}) - 2(1-\alpha)^3\xi_{2,2}(\boldsymbol{\theta}) + 2\alpha^4\kappa_{1,3}(\boldsymbol{\theta}) \\
&\qquad - (1-\alpha)^3\kappa_{2,3}(\boldsymbol{\theta}) + 4\alpha^3(1-\alpha)\varphi_{1,2}(\boldsymbol{\theta}) + 2\alpha^2(1-\alpha)^2\varphi_2^*(\boldsymbol{\theta}) \\
&\qquad + \alpha(1-2\alpha)\int_0^\infty f_0(z)G[g_{1\boldsymbol{\theta}}(z)]dz - (1-\alpha)^3\int_0^\infty f_0(z)G[g_{2\boldsymbol{\theta}}(z)]dz\Big\}.
\end{aligned}$$

Note that $\mathcal{I}(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta} = (\eta, \phi, \alpha)^T$. It is clearly seen that the Fisher information matrix depends on the distribution of the censoring time G, the reference symmetric probability density f_0, and the link function g.

Remark 3.1 *When the censoring distribution has a point mass at $+\infty$ (i.e., there is no censoring), then all $\mathcal{R}_{ij}(\boldsymbol{\theta})$ in Proposition 3.2 are equal zero, and the Fisher information matrix reduces to*

$$\mathcal{I}_{NC}(\boldsymbol{\theta}) = \begin{bmatrix} 2\alpha(1-\alpha)\left(\frac{g'(\eta)}{\phi}\right)^2\gamma_1 & 0 & \frac{-2g'(\eta)}{\phi}\gamma_2 \\ 0 & \frac{1}{\phi^2}(2\gamma_3-1) & -\frac{(1-2\alpha)(2\gamma_3-1)}{\phi\alpha(1-\alpha)} \\ \frac{-2g'(\eta)}{\phi}\gamma_2 & -\frac{(1-2\alpha)(2\gamma_3-1)}{\phi\alpha(1-\alpha)} & \frac{2\gamma_3\left[\alpha^3+(1-\alpha)^3\right]-(1-2\alpha)^2}{\alpha^2(1-\alpha)^2} \end{bmatrix}, \tag{3.6}$$

where the index "NC" refers to non censoring. This is the Fisher information matrix in the complete data case that can be found in Gijbels et al. (2019b). In this special case, we can see that the asymptotic variance of $\widehat{\eta}$ is proportional to $\left(\frac{\phi}{g'(\eta)}\right)^2$, of $\widehat{\phi}$ proportional to ϕ^2, and that the asymptotic variance of $\widehat{\alpha}$ depends only on α.

Example 3.1: Consider the GQBA logistic survival distribution with $g(t) = \ln(t)$, and an exponential censoring distribution. See Table S1. In Section S3 in the Supplementary Material we check the validity of the assumptions for this example. We have that $g'(\eta) = \frac{1}{\eta}$, $\gamma_1 = \frac{1}{6}$, $\gamma_2 = \frac{1}{6} + \frac{\ln(2)}{3}$, and $\gamma_3 = \frac{2}{3} + \frac{\pi^2}{18}$. The Fisher information matrix for these survival and censoring distributions is then given by

$$\mathcal{I}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{1}{3}\alpha(1-\alpha)\left(\frac{1}{\phi\eta}\right)^2 + \mathcal{R}_{11}(\boldsymbol{\theta}) & \mathcal{R}_{12}(\boldsymbol{\theta}) & -\frac{(1+2\ln(2))}{3\phi\eta} + \mathcal{R}_{13}(\boldsymbol{\theta}) \\ \mathcal{R}_{12}(\boldsymbol{\theta}) & \frac{3+\pi^2}{9\phi^2} + \mathcal{R}_{22}(\boldsymbol{\theta}) & -\frac{(1-2\alpha)(3+\pi^2)}{9\phi\alpha(1-\alpha)} + \mathcal{R}_{23}(\boldsymbol{\theta}) \\ -\frac{(1+2\ln(2))}{3\phi\eta} + \mathcal{R}_{13}(\boldsymbol{\theta}) & -\frac{(1-2\alpha)(3+\pi^2)}{9\phi\alpha(1-\alpha)} + \mathcal{R}_{23}(\boldsymbol{\theta}) & \frac{(12+\pi^2)\left[\alpha^3+(1-\alpha)^3\right]-9(1-2\alpha)^2}{9\alpha^2(1-\alpha)^2} + \mathcal{R}_{33}(\boldsymbol{\theta}) \end{bmatrix}.$$

We include this example in the simulation study in Section 4. We also evaluate the Fisher information matrix numerically, with true values of the parameters and different censoring proportions used in the simulation (see Section S3). Note that all $\mathcal{R}_{ij}(\boldsymbol{\theta})$, $i, j = 1, 2, 3$, are easily computable.

Theorem 3.2 *Under Assumptions (B1)-(B11) for $\boldsymbol{\theta}_0$, the MLE $\widehat{\boldsymbol{\theta}}_n$ is asymptotically normally distributed with mean zero vector $\mathbf{0}$ and variance-covariance matrix $\left[\mathcal{I}(\boldsymbol{\theta}_0)\right]^{-1}$, that is,*

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \xrightarrow{D} N_3\left(\mathbf{0}, \left[\mathcal{I}(\boldsymbol{\theta}_0)\right]^{-1}\right), \text{ as } n \to \infty,$$

where $\left[\mathcal{I}(\boldsymbol{\theta}_0)\right]^{-1}$ is the inverse of the Fisher information matrix given in (3.5).

The inverse of the Fisher information matrix is

$$[\mathcal{I}(\boldsymbol{\theta})]^{-1} = \frac{1}{\det\{\mathcal{I}(\boldsymbol{\theta})\}} \begin{bmatrix} \mathcal{I}_{11}^{-1}(\boldsymbol{\theta}) & \mathcal{I}_{12}^{-1}(\boldsymbol{\theta}) & \mathcal{I}_{13}^{-1}(\boldsymbol{\theta}) \\ \mathcal{I}_{12}^{-1}(\boldsymbol{\theta}) & \mathcal{I}_{22}^{-1}(\boldsymbol{\theta}) & \mathcal{I}_{23}^{-1}(\boldsymbol{\theta}) \\ \mathcal{I}_{13}^{-1}(\boldsymbol{\theta}) & \mathcal{I}_{23}^{-1}(\boldsymbol{\theta}) & \mathcal{I}_{33}^{-1}(\boldsymbol{\theta}) \end{bmatrix}, \tag{3.7}$$

where each component of $[\mathcal{I}(\boldsymbol{\theta})]^{-1}$ is given as follows:

$$\begin{aligned}
\mathcal{I}_{11}^{-1}(\boldsymbol{\theta}) &= \mathcal{I}_{22}(\boldsymbol{\theta})\mathcal{I}_{33}(\boldsymbol{\theta}) - \mathcal{I}_{23}^2(\boldsymbol{\theta}), \qquad \mathcal{I}_{12}^{-1}(\boldsymbol{\theta}) = -\big\{\mathcal{I}_{12}(\boldsymbol{\theta})\mathcal{I}_{33}(\boldsymbol{\theta}) - \mathcal{I}_{13}(\boldsymbol{\theta})\mathcal{I}_{23}(\boldsymbol{\theta})\big\} \\
\mathcal{I}_{13}^{-1}(\boldsymbol{\theta}) &= \mathcal{I}_{12}(\boldsymbol{\theta})\mathcal{I}_{23}(\boldsymbol{\theta}) - \mathcal{I}_{13}(\boldsymbol{\theta})\mathcal{I}_{22}(\boldsymbol{\theta}), \qquad \mathcal{I}_{22}^{-1}(\boldsymbol{\theta}) = \mathcal{I}_{11}(\boldsymbol{\theta})\mathcal{I}_{33}(\boldsymbol{\theta}) - \mathcal{I}_{13}^2(\boldsymbol{\theta}) \\
\mathcal{I}_{23}^{-1}(\boldsymbol{\theta}) &= -\big\{\mathcal{I}_{11}(\boldsymbol{\theta})\mathcal{I}_{23}(\boldsymbol{\theta}) - \mathcal{I}_{13}(\boldsymbol{\theta})\mathcal{I}_{12}(\boldsymbol{\theta})\big\}, \quad \mathcal{I}_{33}^{-1}(\boldsymbol{\theta}) = \mathcal{I}_{11}(\boldsymbol{\theta})\mathcal{I}_{22}(\boldsymbol{\theta}) - \mathcal{I}_{12}^2(\boldsymbol{\theta}),
\end{aligned} \tag{3.8}$$

with $\mathcal{I}_{ij}(\boldsymbol{\theta}),\ i,j = 1,2,3$ as in (3.5), and the determinant of the information matrix equals

$$\begin{aligned}
\det\big\{\mathcal{I}(\boldsymbol{\theta})\big\} = \mathcal{I}_{11}(\boldsymbol{\theta})\mathcal{I}_{22}(\boldsymbol{\theta})\mathcal{I}_{33}(\boldsymbol{\theta}) \ - \ \mathcal{I}_{11}(\boldsymbol{\theta})\mathcal{I}_{23}^2(\boldsymbol{\theta}) \ - \ \mathcal{I}_{22}(\boldsymbol{\theta})\mathcal{I}_{13}^2(\boldsymbol{\theta}) \\
- \ \mathcal{I}_{33}(\boldsymbol{\theta})\mathcal{I}_{12}^2(\boldsymbol{\theta}) \ + \ 2\mathcal{I}_{12}(\boldsymbol{\theta})\mathcal{I}_{13}(\boldsymbol{\theta})\mathcal{I}_{23}(\boldsymbol{\theta}).
\end{aligned} \tag{3.9}$$

See Aitken (2017) for details on computation of the inverse of symmetric matrices. The asymptotic variance-covariance matrix can be estimated by $\big[\mathcal{I}(\widehat{\boldsymbol{\theta}}_n)\big]^{-1}$. Given the continuity of the Fisher information matrix under the stated assumptions $\big[\mathcal{I}(\widehat{\boldsymbol{\theta}}_n)\big]^{-1}$ is a consistent estimator of $\big[\mathcal{I}(\boldsymbol{\theta}_0)\big]^{-1}$.

4 Simulation study

We study the finite–sample performance of the MLE, including the impact of the censoring rate. We consider one scenario with a specific survival and censoring distribution and link function. Three other scenarios are studied in Section S4 of the Supplementary Material.

The finite–sample performance of the MLE is measured via Monte Carlo approximations of bias and mean squared error (MSE). More precisely, let $\hat{\theta}_k^{(j)}$ be the MLE of the k-th parameter in the j-th simulated sample, $k = 1,2,3;\ j = 1,2,...,N$, where N is the number of Monte Carlo simulated samples. The criteria used to measure the finite–sample performance are

$$\begin{aligned}
&\text{Bias: } \text{ABias}(\hat{\theta}_k) = \bar{\hat{\theta}}_k - \theta_{0,k}, \\
&\text{Mean squared error: } \text{AMSE}(\hat{\theta}_k) = \frac{1}{N}\sum_{j=1}^{N}\big(\hat{\theta}_k^{(j)} - \theta_{0,k}\big)^2,
\end{aligned} \tag{4.1}$$

where $\bar{\hat{\theta}}_k = \frac{1}{N}\sum_{j=1}^{N}\hat{\theta}_k^{(j)}$, is the average of $\hat{\theta}_k^{(j)}$ across the simulated samples and $\theta_{0,k}$ is the k-th element of the vector $\boldsymbol{\theta}_0$.

4.1 Data generation

We first determine the censoring parameter θ_c for a predefined censoring proportion, say P_c. For a given P_c we solve $P(\Delta = 0) - P_c = 0$,

$$P(\Delta = 0) - P_c = \int_0^{\infty}\{1 - F_{\alpha}(v)\}dG(v) - P_c. \tag{4.2}$$

There is no closed form expression for the solution. To compute the integrals, we apply a numerical integration through the Brent-Dekker algorithm known as the `uniroot()` function in the statistical software `R`. After having determined, for given censoring proportion P_c, the censoring parameter θ_c, the following simulation designs are conducted to generate both the survival time $T \sim \mathcal{AN}(\eta,\phi,\alpha)$ and censoring time $C \sim \text{Expo}(\theta_c)$, and finally we obtain the observed time as $Y = \min(T,C)$.

(i) Generate a random variable $U_i \sim \text{Uniform}(0,1)$, for $i = 1,2,\ldots,n$;

(ii) Generate the survival time using the inverse transform sampling algorithm and log-link function (i.e. $g(t) = \ln(t)$). That is using

$$T_i = F_\alpha^{-1}(U_i) \equiv \begin{cases} g^{-1}\left[g(\eta) + \frac{\phi}{1-\alpha} F_0^{-1}\left(\frac{U_i}{2\alpha}\right)\right] & \text{if } U_i < \alpha, \quad i = 1,\ldots,n \\ g^{-1}\left[g(\eta) + \frac{\phi}{\alpha} F_0^{-1}\left(\frac{1+U_i-2\alpha}{2(1-\alpha)}\right)\right] & \text{if } U_i \geq \alpha; \end{cases}$$

(iii) Generate the censoring time $C_i \sim \text{Expo}(\theta_c)$ considering each nominal censoring percentage $P_c = 25\%, 50\%, 75\%$;

(iv) The observed time is $Y_i = \min(T_i, C_i)$ with censoring indicator $\Delta_i = I(T_i \leq C_i)$;

(v) Repeat the steps (i) - (iv) for $N = 1000$ times for the sample sizes $n = 100, 300, 500$, censoring percentages $P_c = 25\%, 50\%, 75\%$, and the index-parameter values $\alpha = 0.25, 0.5, 0.75$. The other parameters are $\eta = 1$ and $\phi = 0.5$. Hence, in total there are $3 \times 3 \times 3 = 27$ data generating mechanisms.

The survival and hazard functions for the true GQBA family of distribution used in this simulation study is given in Fig. 4.1.

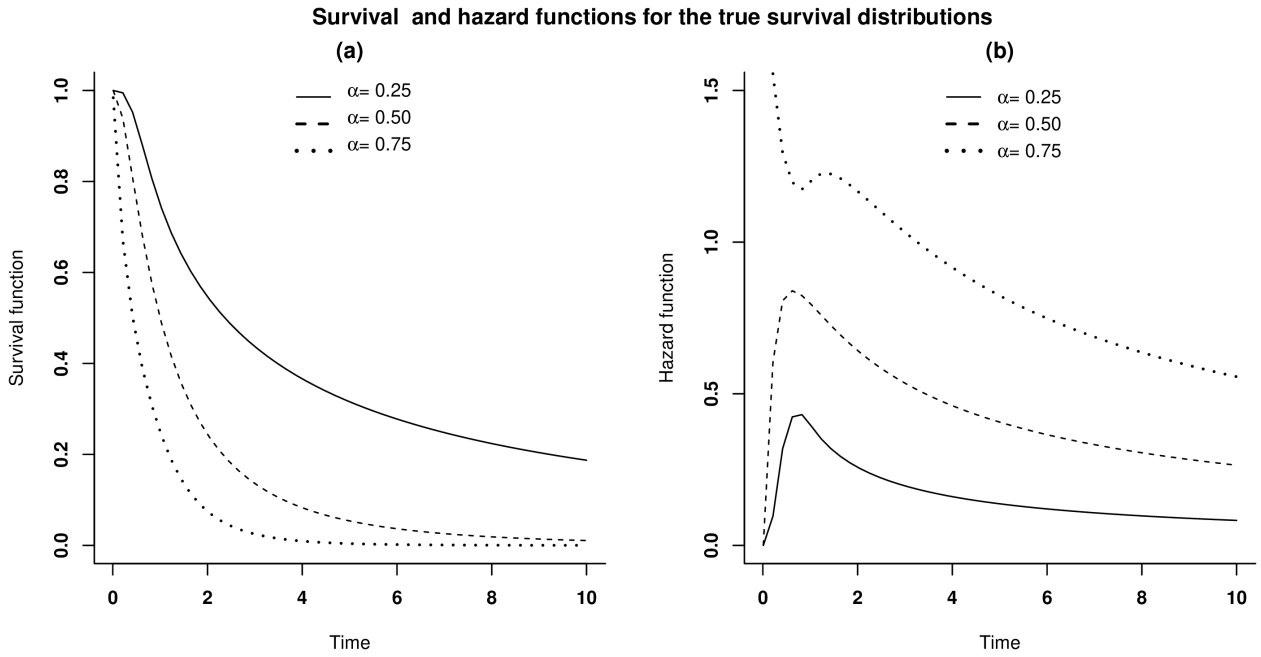

Fig. 4.1 Simulation study. Survival function (a) and hazard function (b) for the true distribution.

4.2 Simulation results

The MLE estimates are obtained using package `nloptr` (version 1.2.1) in `R` with bound constraint Nelder-Mead simplex algorithm. To start the optimization process we use some steps. First, a set

of ten equally spaced starting values for the parameters of the GQBA distributions is chosen (from 0.25 to 5 for η; from 0.15 to 2.5 for ϕ; and from 0.05 to 0.95 for α). The set of initial values that yields the largest log-likelihood value is selected as final starting value. The approximated bias and MSE of the MLE are computed based on the results from 1000 simulated samples.

The approximated bias and approximated MSE of the estimators are presented in Table S5(in the Supplementary Material) and Table 4.1, respectively. Further, the boxplots of the MLE estimates across the censoring proportions, α-value and sample sizes are presented in Fig. 4.2. From these tables and figure it is clear that the bias and approximated MSE increase with censoring proportions. A slight difference is only highlighted in a specific data generating mechanism, that is when $\alpha = 0.75$ and $P_c = 50\%$. From Table 4.1 we can conclude that the MSE is getting smaller as the sample size increases, for a given censoring percentage. Overall, accurate estimates can be achieved for large sample sizes, irrespective of censoring proportion.

Table 4.1 Simulation study. The approximated MSE of the MLE.

		$\alpha = 0.25$			$\alpha = 0.50$			$\alpha = 0.75$		
P_c	n	$\widehat{\eta}$	$\widehat{\phi}$	$\widehat{\alpha}$	$\widehat{\eta}$	$\widehat{\phi}$	$\widehat{\alpha}$	$\widehat{\eta}$	$\widehat{\phi}$	$\widehat{\alpha}$
25%	100	0.225	0.019	0.009	0.154	0.004	0.012	0.354	0.048	0.017
	300	0.047	0.004	0.002	0.036	8.0e-4	0.003	0.087	0.012	0.005
	500	0.024	0.002	0.001	0.020	4.0e-4	0.002	0.043	0.006	0.002
50%	100	0.362	0.024	0.012	0.401	0.019	0.030	0.577	0.112	0.045
	300	0.056	0.005	0.003	0.073	0.003	0.007	0.282	0.063	0.021
	500	0.031	0.002	0.002	0.034	0.001	0.004	0.177	0.038	0.012
75%	100	1.033	0.040	0.032	0.648	0.059	0.078	1.073	0.169	0.117
	300	0.141	0.007	0.007	0.235	0.019	0.030	0.530	0.156	0.067
	500	0.056	0.003	0.003	0.140	0.011	0.017	0.409	0.140	0.051

It is also worth to mention here that the simulation shows better performance for $\alpha = 0.25$ compared to $\alpha = 0.75$. This is due to the fact that the survival distributions are mostly right skewed and the index-parameter α controls the allocation of mass of the distributions as indicated in the paragraph following Theorem 2.1.

In the real data applications (Section 5), we present bootstrap based standard errors (based on 500 bootstrap samples) for the parameter estimates. In order to investigate the validity of such bootstrap based standard errors (based on 200 bootstrap samples), we compare them in Table 4.2 with the standard error within the Monte Carlo simulation study for the setting with 25% censoring. Note that both standard errors are similar and their differences diminish as sample size increases.

In the above we used in the estimation procedure as GQBA distribution the $\mathcal{AN}(\eta, \phi, \alpha)$ distribution, and the log-link function, under which the data were simulated. In a real data setting one might not know which GBQA model and link function is most appropriate. One then could use different models and apply a model selection criterium to select a model. To investigate the quality of such model selection we fit various candidate models. To compare the quality of the fitted models we use the AIC criterion (Akaike information criterion), to choose the most appropriate

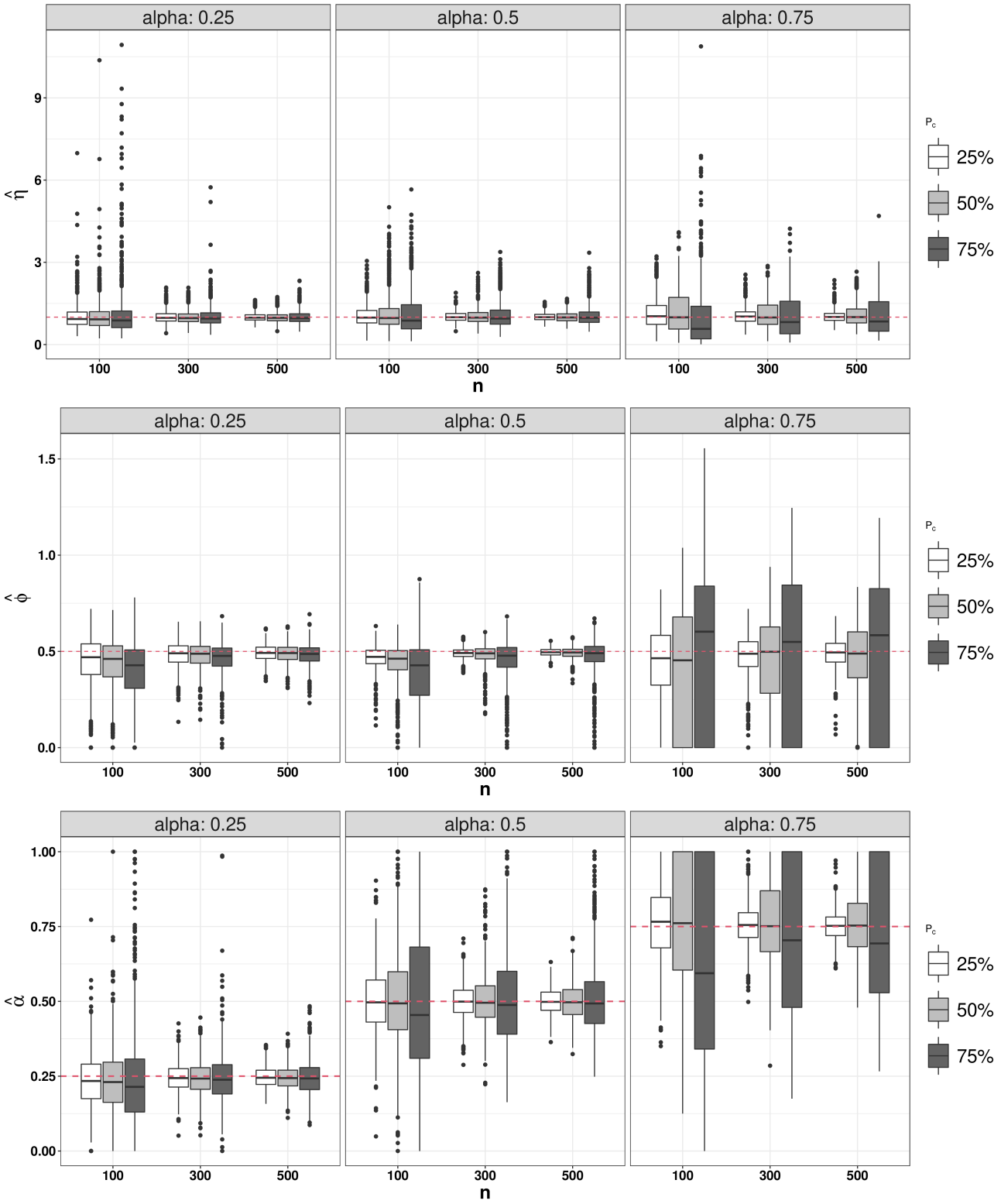

Fig. 4.2 Simulation study. Boxplots of the maximum likelihood estimates.

Table 4.2 Simulation study. Comparison of bootstrap based (Boot) and Monte Carlo (MC) standard errors (SE). Setting when $P_c = 0.25$.

	$\widehat{\eta}$		$\widehat{\phi}$		$\widehat{\alpha}$	
n	MC SE	Boot SE	MC SE	Boot SE	MC SE	Boot SE
$\alpha = 0.25$						
100	0.505	0.618	0.120	0.125	0.090	0.096
300	0.207	0.220	0.058	0.061	0.044	0.046
500	0.146	0.153	0.045	0.044	0.032	0.033
$\alpha = 0.50$						
100	0.380	0.466	0.055	0.082	0.107	0.121
300	0.190	0.200	0.025	0.030	0.055	0.057
500	0.143	0.143	0.021	0.021	0.043	0.043
$\alpha = 0.75$						
100	0.588	0.515	0.211	0.183	0.133	0.126
300	0.299	0.312	0.107	0.108	0.066	0.067
500	0.219	0.219	0.077	0.077	0.050	0.049

distribution(s) among the considered candidate set. More specifically, the AIC value is expressed as

$$\text{AIC} = -2\mathcal{L}_n(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}) + 2p,$$

where $\widehat{\boldsymbol{\theta}}_n$ is the vector of MLE estimates, $\mathcal{L}_n(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}})$ is the maximal log-likelihood value and p is the number of estimated parameters for a given distribution. Alternatively one can use a BIC criterion (Bayesian information criterion) which is defined as AIC but with the term $2p$ replaced by $p\ln(n)$.

Table 4.3 presents the percentage a specific GQBA distribution ($\mathcal{AN}(\eta,\phi,\alpha)$: normal with $g(t) = \ln(t)$; $\mathcal{AN}(\eta,\phi,\alpha,\lambda)$: normal with $g(t;\lambda) = \ln(e^{\lambda t} - 1)$; $\mathcal{AL}(\eta,\phi,\alpha)$: logistic with $g(t) = \ln(t)$; $\mathcal{AL}(\eta,\phi,\alpha,\lambda)$: logistic with $g(t;\lambda) = \ln(e^{\lambda t} - 1)$; $\mathcal{ALa}(\eta,\phi,\alpha)$: Laplace with $g(t) = \ln(t)$; and $\mathcal{ALa}(\eta,\phi,\alpha,\lambda)$: Laplace with $g(t;\lambda) = \ln(e^{\lambda t} - 1)$) is selected as best based on the AIC value over the 1000 simulated samples in the setting with $P_c = 25\%$. Note that the correct model $\mathcal{AN}(\eta,\phi,\alpha,\lambda)$ is selected most of the times and that the percentage that the correct model is selected as best increases with sample size.

In Fig. 4.3 we present the density of the standardized ML estimates (across simulations) along with the theoretical asymptotic distribution of the estimators. This shows that the asymptotic normal density is achievable as the sample size increases.

5 Real data applications

We consider two examples for which we illustrate the use of the proposed GQBA family of distributions. The main intention here is to select a kind of best model distribution in fitting a set of candidate distributions to the data. In addition to the GQBA family, four log-location-scale type

Table 4.3 Simulation study. Percentage (over 1000 simulated samples) a specific distribution is selected as best based on AIC. The correct model is $\mathcal{AN}(\eta = 1, \phi = 0.5, \alpha)$ and $P_c = 25\%$.

n	$\mathcal{AN}(\widehat{\eta},\widehat{\phi},\widehat{\alpha})$	$\mathcal{AL}(\widehat{\eta},\widehat{\phi},\widehat{\alpha})$	$\mathcal{ALa}(\widehat{\eta},\widehat{\phi},\widehat{\alpha})$	$\mathcal{AN}(\widehat{\eta},\widehat{\phi},\widehat{\alpha},\widehat{\lambda})$	$\mathcal{AL}(\widehat{\eta},\widehat{\phi},\widehat{\alpha},\widehat{\lambda})$	$\mathcal{ALa}(\widehat{\eta},\widehat{\phi},\widehat{\alpha},\widehat{\lambda})$
			$\alpha = 0.25$			
100	54.66	16.20	8.83	6.53	3.26	10.52
300	68.56	15.24	0.85	7.13	4.47	3.75
500	72.07	12.82	0.00	7.38	5.68	2.06
			$\alpha = 0.50$			
100	57.44	17.90	5.20	5.68	4.59	9.19
300	70.86	12.09	0.00	6.53	4.96	5.56
500	75.57	8.10	0.00	6.77	7.62	1.93
			$\alpha = 0.75$			
100	67.71	13.91	5.93	5.32	3.87	3.26
300	72.31	11.00	0.24	5.68	7.98	2.78
500	74.73	6.89	0.00	7.38	9.79	1.21

distributions (hereafter we call them classical) such as log-normal, log-logistic, log-Laplace and Weibull distributions are included in the analysis. These distributions are included for illustrating how the proposed GQBA family may give an improved fit compared to some well known existing distributions in time to event analysis. We use the notation log-$\mathcal{N}(\eta,\phi)$ for the log-normal distribution, log-$\mathcal{L}(\eta,\phi)$ for log-logistic and log-$\mathcal{L}a(\eta,\phi)$ for log-Laplace.

For the GQBA family of distributions, we consider four cases for f_0: a Laplace, a normal, a logistic and a Student's-t density with ν degrees of freedom; and for g two link functions: $g(t) = \ln(t)$ and $g(t;\lambda) = \ln(e^{\lambda t} - 1)$. This leads to eight candidate models from the GQBA family. We denote the GQBA Laplace family by $\mathcal{AL}a(.)$ and $\mathcal{AS}$-$t(.,\nu)$ for the GQBA family of Student's-t density with ν degrees of freedom. We further denote $\mathcal{AL}a(.,\lambda)$ and $\mathcal{AS}$-$t(.,\lambda,\nu)$ in case the link function $g(t;\lambda) = \ln(e^{\lambda t} - 1)$ is used. We thus have a total of twelve candidate distributions to describe the behaviour of the data. It may be useful to note that the location parameter μ for the log-location-scale family that appears in most of the standard books is reparametrized by $\mu = \ln(\eta)$. With this parameterization η becomes the scale while ϕ is the shape parameter.

The parameters in all models are estimated by the MLE technique with a Nelder-Mead bound-constrained simplex algorithm via `nloptr` package in `R`. We optimize as follows.

Step 1: The method of moment based estimates are used as starting values to fit the corresponding classical densities (only with η and ϕ parameters). Denote these estimates by $\widehat{\eta}_c$ and $\widehat{\phi}_c$.

Step 2: Estimate the index-parameter α for the GQBA family with $g(t) = \ln(t)$ by fixing $\eta = \widehat{\eta}_c$ and $\phi = \widehat{\phi}_c$. We denote this estimate by $\widehat{\alpha}_c$.

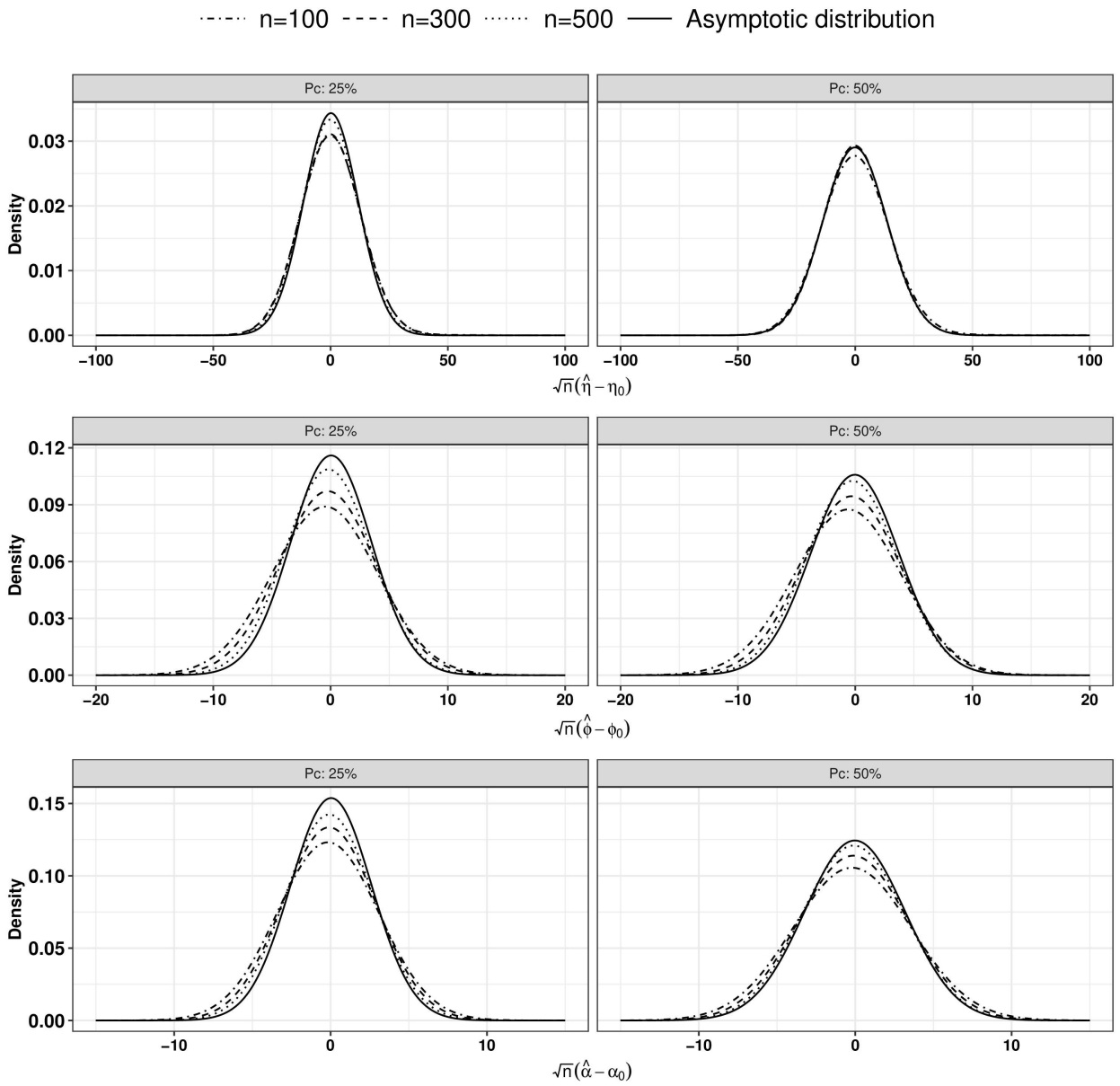

Fig. 4.3 Simulation study. Density plots for the standardized MLE $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ with $\alpha = 0.25$, across each censoring percentage (P_c) and sample size (n). The asymptotic normal density $N(0, \mathcal{I}^{-1}(\boldsymbol{\theta}_{0,k}))$ is depicted by the solid line.

Step 3: Fit the GQBA distributions with link function $g(t) = \ln(t)$ taking $\widehat{\eta}_c$, $\widehat{\phi}_c$ and $\widehat{\alpha}_c$ as starting values in the algorithm. Denote the resulting estimates by $\widehat{\eta}_q$, $\widehat{\phi}_q$, and $\widehat{\alpha}_q$. These are the final MLE estimates for this family of distributions.

Step 4: Estimate λ for the GQBA family with link function $g(t;\lambda) = \ln(e^{\lambda t} - 1)$, by fixing $\eta = \widehat{\eta}_q$, $\phi = \widehat{\phi}_q$, and $\alpha = \widehat{\alpha}_q$. Denote this estimate by $\widehat{\lambda}_q$.

Step 5: Fit the GQBA family with link $g(t;\lambda) = \ln(e^{\lambda t} - 1)$, using $\widehat{\eta}_q$, $\widehat{\phi}_q$, $\widehat{\alpha}_q$, and $\widehat{\lambda}_q$ as starting values. The resulting MLE estimates are the final MLE values for $(\widehat{\eta}_f, \widehat{\phi}_f, \widehat{\alpha}_f, \widehat{\lambda}_f)$ except for $\mathcal{AS}$-$t(\eta, \phi, \alpha, \lambda, \nu)$ that will be obtained in *Step 6*.

Step 6: Estimate ν in $\mathcal{AS}$-$t(\eta, \phi, \alpha, \lambda, \nu)$ by setting the remaining parameters with the MLE values obtained in *Step 5* from the $\mathcal{AN}(\eta, \phi, \alpha, \lambda)$ distribution. Next, the initial values of $\mathcal{AS}$-$t(\eta, \phi, \alpha, \nu)$ (Step 3 in our algorithm) are estimated from an $\mathcal{AS}$-t model but with known degrees of freedom (ν). Denote the resulting ML estimate by $\widehat{\nu}_q$. Finally, $\mathcal{AS}$-$t(\eta, \phi, \alpha, \lambda, \nu)$ can be fitted considering $(\widehat{\eta}_f, \widehat{\phi}_f, \widehat{\alpha}_f, \widehat{\lambda}_f, \widehat{\nu}_q)$ as starting values.

5.1 Example 1: Times to weaning of breast-fed data

In the US National Longitudinal Survey of Youth (NLSY), females aged 14 to 21 in 1979 were interviewed yearly until 1988. The data considered here consists of the information from 927 first-born children to mothers who chose to breast feed their children. The response variable in the data set was duration of breast feeding in weeks, followed by an indicator of whether the breast feeding was completed (i.e., the infant is weaned). The censoring proportion in this dataset is small, only 3.8% of the sample cases are right censored, and the median survival time is 12 weeks. Details on the data can be found in the book by Klein and Moeschberger (2006, Chapter 1), and the dataset is also freely available in the `R` statistical software package `KMsurv`.

The results for all the twelve fitted models for the time to weaning of breast-fed data are summarized in Table 5.1 and the estimated survival and hazard function based on the best model, together with the Kaplan-Meier estimator of the survival function (and their bootstrap based pointwise 95% confidence intervals) are given in Fig. 5.1 (where the estimation is done on a grid of 100 equally spaced time points within the range of uncensored observations). From Fig. 5.1 we can see that the estimated survival function based on the selected (best) GQBA model is very close to the non-parametric Kaplan-Meier estimate and that the confidence intervals of the best GQBA model are narrower than these of the Kaplan-Meier estimate. The former comes of course with the advantage of being a parametric model (with its ease of interpretations). Table 5.1 presents the number of parameters (p) with MLE estimates, the negative of maximum log-likelihood ($-\mathcal{L}_n(\widehat{\boldsymbol{\theta}}_n)$), AIC and BIC values. For these data, the GQBA logistic distribution with logit-link function, $\mathcal{AL}(\eta, \phi, \alpha, \lambda)$ has the lowest AIC and BIC value (e.g., AIC= 6657.20) compared to the other fitted models. Interestingly, the negative maximal log-likelihood and the MLE estimates are exactly the same in the case of models $\mathcal{AN}(\eta, \phi, \alpha)$ and $\mathcal{AS}$-$t(\eta, \phi, \alpha, \nu)$. This is not surprising since the estimate $\widehat{\nu}$ is very large in the latter model. This is also observed for the results of German Breast Cancer data presented in Tables 5.2 and S11 (in the Supplementary Material). A non-parametric bootstrap method (based on 500 bootstrap samples) is used to calculate standard errors. The bootstrap-based standard errors are given in brackets in Table 5.1.

5.2 Example 2: German breast cancer data

The second real data set concerns a subset of the clinical trial data conducted by the German Breast Cancer Study (GBCS) group. The data have been obtained from a 2×2 randomized trial

Table 5.1 Breast-fed data. MLE (bootstrap based standard errors), negative maximal log-likelihood, AIC and BIC values for the fitted models.

Density	p	$\widehat{\eta}$	$\widehat{\phi}$	$\widehat{\alpha}$	$\widehat{\lambda}$	$\widehat{\nu}$	$-\mathcal{L}_n(\widehat{\boldsymbol{\theta}}_n)$	AIC	BIC
Classical densities									
log-$\mathcal{L}(\eta,\phi)$	2	9.81	0.70	–	–	–	3429.32	6862.63	6872.29
		(0.40)	(0.01)						
log-$\mathcal{N}(\eta,\phi)$	2	9.40	1.18	–	–	–	3402.77	6809.55	6819.21
		(0.35)	(0.02)						
log-$\mathcal{L}a(\eta,\phi)$	2	12.00	0.98	–	–	–	3483.78	6971.55	6981.22
		(1.05)	(0.02)						
Weibull(η,ϕ)	2	16.59	0.97	–	–	–	3408.56	6821.13	6830.79
		(0.60)	(0.02)						
GQBA with $g(t)=\ln(t)$									
$\mathcal{AL}(\eta,\phi,\alpha)$	3	20.96	0.28	0.71	–	–	3411.25	6828.50	6842.99
		(3.35)	(0.02)	(0.04)					
$\mathcal{AN}(\eta,\phi,\alpha)$	3	19.87	0.49	0.70	–	–	3386.12	6778.23	6785.89
		(2.56)	(0.05)	(0.05)					
$\mathcal{AL}a(\eta,\phi,\alpha)$	3	16.00	0.44	0.64	–	–	3463.92	6933.84	6948.34
		(1.37)	(0.02)	(0.03)					
$\mathcal{AS}$-$t(\eta,\phi,\alpha,\nu)$	4	19.87	0.49	0.70	–	3.4e+13	3386.12	6780.23	6799.56
		(1.62)	(0.02)	(0.02)		(4.7e+13)			
GQBA with $g(t;\lambda)=\ln(e^{\lambda t}-1)$									
$\mathcal{AL}(\eta,\phi,\alpha,\lambda)$	4	2.00	0.34	0.13	0.16	–	3324.60	**6657.20**	**6676.52**
		(0.43)	(0.22)	(0.05)	(0.03)				
$\mathcal{AN}(\eta,\phi,\alpha,\lambda)$	4	2.06	0.52	0.14	0.10	–	3373.67	6755.34	6774.66
		(2.73)	(0.26)	(0.13)	(0.04)				
$\mathcal{AL}a(\eta,\phi,\alpha,\lambda)$	4	16.00	0.52	0.62	0.02	–	3457.53	6923.06	6942.39
		(1.43)	(0.05)	(0.03)	(0.01)				
$\mathcal{AS}$-$t(\eta,\phi,\alpha,\lambda,\nu)$	5	2.05	0.51	0.14	0.10	40.75	3370.85	6751.71	6775.86
		(4.82)	(0.09)	(0.11)	0.02	(3.8e+16)			

evaluating hormonal treatment and the duration of chemotherapy for node-positive breast cancer patients. The trial was used to compare recurrence-free and overall survival between the different treatment modalities (Schumacher et al., 1994). The event of interest considered in this paper is the time to tumor recurrence (defined as the occurrence of either locoregional or distant recurrence, contralateral tumor, and secondary tumor). There were 686 patients in the study, of whom 299 had an event (tumor recurrence) at the end of the study, and right censoring occurred for the remaining 387 patients. The censoring proportion is 56.4%, of whom 22.1% belong to the hormone therapy treatment group while 34.3% were in the untreated group. Further detailed description of the data can be found in Schumacher et al. (1994) and Sauerbrei et al. (1999). The data are also available in the `survidm` package of the statistical software `R`.

For this dataset, the twelve candidate models are fitted to all patients (without considering treatment type) and also for each hormonal treatment (untreated and treated) groups separately. The results are presented in Table 5.2 for all the twelve fitted models based on data for all patients.

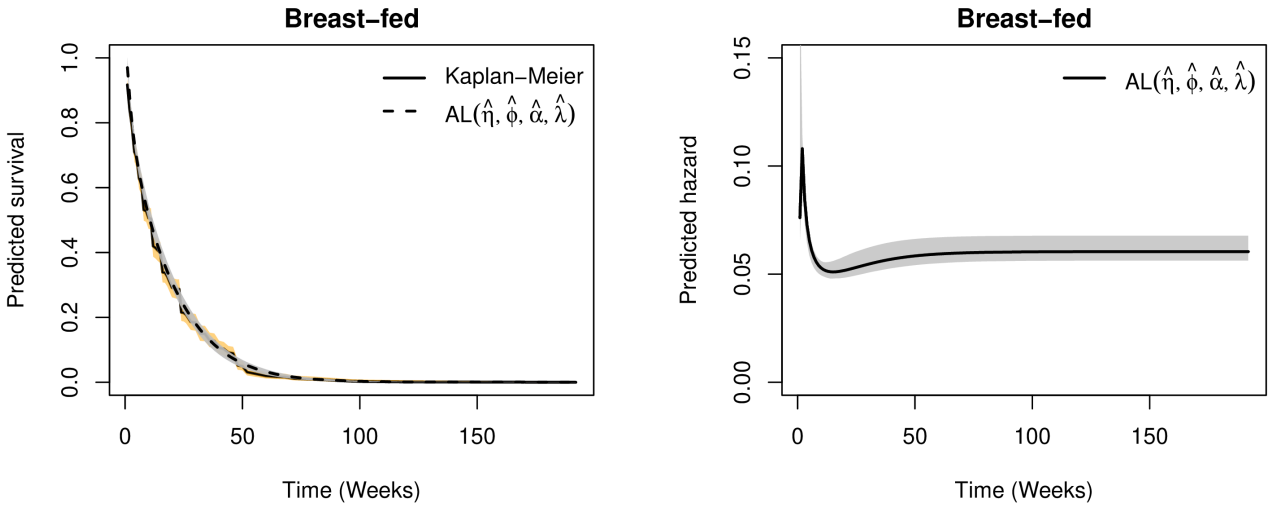

Fig. 5.1 Breast-fed data. Kaplan-Meier and estimated survival (left) and hazard (right) function using the best model with point-wise 95% confidence intervals (yellow for Kaplan-Meier and grey for the best model) .

Table S11 (in the Supplementary Material) presents the results for models fitted to each treatment group separately. From Table 5.2, it is seen that the GQBA logistic and GQBA normal family densities have almost comparable AIC values. However, the GQBA normal with logit-link function $g(t;\lambda) = \ln(e^{\lambda t} - 1)$ gives a slightly smaller AIC value (AIC=1703.64). Based on BIC, the best model is GQBA normal with log-link. Turning to the different treatment groups presented in Table S11 (in the Supplementary Material), the smallest AIC and BIC value for the untreated group are for $\mathcal{AN}(\eta, \phi, \alpha)$ with log-link. For the treated group, the smallest AIC value is also for the GQBA normal with log-link, while the smallest BIC value is for the classical log-normal distribution.

Fig. 5.2 displays the Kaplan-Meier estimates of the survival function together with its estimated counterpart corresponding to the best fitting model (based on AIC), for all patients. The estimated hazard function is also provided for the best model according to AIC. From this figure, it is clearly observed that the best selected parametric model produces similar and coinciding fitted survival function compared to the Kaplan-Meier survival curve and that the confidence intervals of the best GQBA model are narrower then these of the Kaplan-Meier estimate. The estimated survival and hazard function for the treated and untreated group separately are given in Fig. S6 and S5 (in the Supplementary Material).

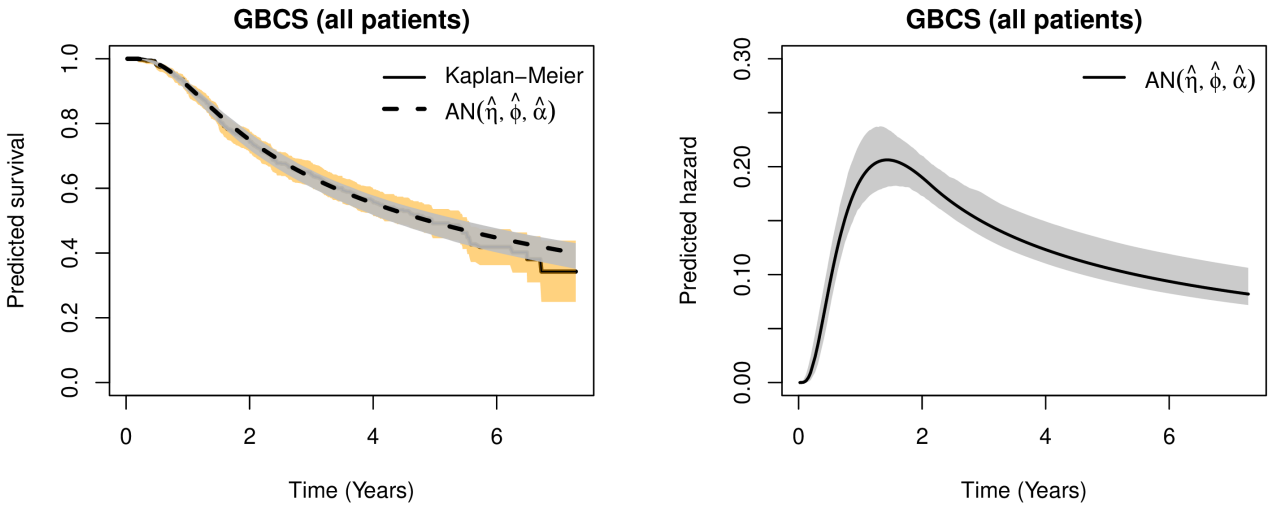

Fig. 5.2 GBC data (all patients). Kaplan-Meier and estimated survival (left) and hazard (right) function using the best model with point-wise 95% confidence intervals (yellow for Kaplan-Meier and grey for the best model) .

Table 5.2 GBC data (all patients). MLE, negative maximal log-likelihood, AIC and BIC values for the fitted models and 95% bootstrap confidence intervals of the estimates.

Density	p	$\widehat{\eta}$	$\widehat{\phi}$	$\widehat{\alpha}$	$\widehat{\lambda}$	$\widehat{\nu}$	$-\mathcal{L}_n(\widehat{\boldsymbol{\theta}}_n)$	AIC	BIC
Classical densities									
log-$\mathcal{L}(\eta,\phi)$	2	4.50	0.65	–	–	–	863.67	1731.35	1740.41
		[4.094, 5.000]	[0.604, 0.698]	–	–	–			
log-$\mathcal{N}(\eta,\phi)$	2	4.58	1.11	–	–	–	854.61	1713.22	1722.28
		[4.157, 5.116]	[1.031, 1.195]	–	–	–			
log-$\mathcal{L}a(\eta,\phi)$	2	4.64	0.93	–	–	–	884.78	1773.56	1782.62
		[4.084, 5.128]	[0.862, 1.001]	–	–	–			
Weibull(η,ϕ)	2	6.19	1.27	–	–	–	873.00	1750.01	1759.07
		[5.643, 6.821]	[1.189, 1.368]	–	–	–			
GQBA with $g(t)=\ln(t)$									
$\mathcal{AL}(\eta,\phi,\alpha)$	3	1.69	0.29	0.21	–	–	849.44	1704.88	1718.48
		[1.384, 2.266]	[0.245, 0.333]	[0.169, 0.289]	–	–			
$\mathcal{AN}(\eta,\phi,\alpha)$	3	2.14	0.55	0.27	–	–	849.09	1704.18	**1717.78**
		[1.601, 3.283]	[0.462, 0.611]	[0.205, 0.391]	–	–			
$\mathcal{AL}a(\eta,\phi,\alpha)$	3	1.48	0.40	0.17	–	–	852.92	1711.84	1725.44
		[1.166, 5.525]	[0.327, 0.492]	[0.124, 0.554]	–	–			
$\mathcal{AS}$-$t(\eta,\phi,\alpha,\nu)$	4	2.04	0.53	0.26	–	47.06	849.07	1706.13	1724.26
		[1.526, 3.283]	[0.395, 0.607]	[0.184, 0.391]	–	[4.966, 1.6e+13]			
GQBA with $g(t;\lambda)=\ln(e^{\lambda t}-1)$									
$\mathcal{AL}(\eta,\phi,\alpha,\lambda)$	4	1.49	0.32	0.17	0.22	–	848.03	1704.05	1722.18
		[1.032, 1.740]	[0.248, 0.390]	[0.096, 0.218]	[0.0001, 0.658]	–			
$\mathcal{AN}(\eta,\phi,\alpha,\lambda)$	4	1.65	0.60	0.20	0.23	–	847.82	**1703.64**	1721.76
		[1.195, 2.144]	[0.478, 0.720]	[0.129, 0.278]	[7.8e-9, 0.502]	–			
$\mathcal{AL}a(\eta,\phi,\alpha,\lambda)$	4	1.23	0.48	0.12	0.41	–	849.92	1707.84	1726.00
		[0.925, 5.525]	[0.363, 0.549]	[0.067, 0.554]	[1.0e-10, 0.981]	–			
$\mathcal{AS}$-$t(\eta,\phi,\alpha,\nu,\lambda)$	5	1.60	0.57	0.19	0.23	24.27	847.71	1705.43	1728.07
		[1.087, 2.017]	[0.400, 0.715]	[0.109, 0.249]	[0.0001, 0.518]	[4.623, 2.92e+13]			

6 Discussion and conclusion

The aim of this paper is to study a family of generalized quantile-based asymmetric distributions when the variable of interest is partially observed, due to right random censoring. We establish the consistency and the asymptotic normality of the MLE in the nonstandard setting of the GQBA family of distributions. The finite–sample performance of the MLE is investigated via a simulation study. A good performance was obtained also under high censoring cases.

The study contributes to the area of lifetime data analysis in proposing flexible distributions for censored data. The flexibility of the GQBA family can be summarized in three fold: first to characterize different shapes of the hazard function; second it includes the commonly used log-symmetric family of distributions as special cases; and third the statistical inference we show is valid for any member of the broad family. The proposed probability distribution is not twice differentiable at the location parameter η and hence some of the standard regularity conditions in MLE based inference may not be applied. Nevertheless, we showed that the MLE of the parameters have good asymptotic properties, and the objective function can be easily optimized with existing derivative free algorithms.

As noted in Fernandez and Steel (1999) the use of MLE for the Student-t with unknown degrees of freedom should be done carefully. We experienced that the estimate of the degrees of freedom ν is sensitive to the starting value, in contrast to the other parameters.

A natural progression of this work, which is being currently investigated is to apply the GQBA family in a flexible regression setting, by allowing the parameters (η, ϕ and α to be a function of a covariate). This will allow us to assess the effect of covariates by varying the quantiles, meaning that the effect of covariates on a specific percentile of the survival distribution will be investigated. Another extension of this work is considering other types of censoring, like left or interval censoring, by adapting the log-likelihood function (3.1) appropriately.

Acknowledgement

The authors are grateful to the editor, and associate editor and reviewers for their valuable comments that led to an improvement of the manuscript. The second author gratefully acknowledge support from Research Grant FWO G0D6619N of the Flemish Science Foundation, and from the C16/20/002 project of the Research Fund KU Leuven.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Aitken, A. C. (2017). *Determinants and Matrices*. Read Books Ltd.

Bagnoli, M., & Bergstrom, T. (2005). Log-concave probability and its applications. *Economic Theory*, 26(2), 445-469.

Barr, D. R., & Davidson, T. (1973). A Kolmogorov-Smirnov test for censored samples. *Technometrics*, 15(4), 739-757.

Bottai, M., & Zhang, J. (2010). Laplace regression with censored data. *Biometrical Journal*, 52(4), 487-503.

Fernandez, C., & Steel, M. F. (1999). Multivariate Student-t regression models: Pitfalls and inference. *Biometrika*, 86(1), 153-167.

Gerds, T. A., & Schumacher, M. (2006). Consistent estimation of the expected Brier score in general survival models with right-censored event times. *Biometrical Journal*, 48(6), 1029-1040.

Gijbels, I., Karim, R. & Verhasselt, A. (2019a). On quantile-based asymmetric family of distributions: properties and inference. *International Statistical Review*, 87(3), 471-504.

Gijbels, I., Karim, R. & Verhasselt, A. (2019b). Quantile estimation in a generalized asymmetric distributional setting. In A. Steland, E. Rafajlowicz, & O. Okhrin (Eds.) *Stochastic Models, Statistics and Their Applications* (pp. 13-40). Springer Proceedings in Mathematics & Statistics.

Graf, E., Schmoor, C., Sauerbrei, W., & Schumacher, M. (1999). Assessment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine*, 18(17-18), 2529-2545.

Gupta, R. D., & Kundu, D. (2001). Exponentiated exponential family: an alternative to Gamma and Weibull distributions. *Biometrical Journal*, 43(1), 117-130.

Hougaard, P. (1999). Fundamentals of survival data. *Biometrics*, 55(1), 13-22.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* , 5(1), 221-233.

Khan, S. A., & Khosa, S. K. (2016). Generalized log-logistic proportional hazard model with applications in survival analysis. *Journal of Statistical Distributions and Applications*, 3(1), 1-18.

Klein, J. P., & Moeschberger, M. L. (2006). *Survival Analysis: Techniques for Censored and Truncated Data.* Springer Science and Business Media.

Koziol, J. A., & Green, S. B. (1976). A Cramer-von Mises statistic for randomly censored data. *Biometrika*, 63(3), 465-474.

Lan, Y., & Leemis, L. M. (2008). The logistic–exponential survival distribution. *Naval Research Logistics*, 55(3), 252-264.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons.

Liao, J. J., & Liu, G. F. (2019). A flexible parametric survival model for fitting time to event data in clinical trials. *Pharmaceutical Statistics*, 18(5), 555-567.

Li, R., & Peng, L. (2017). Assessing quantile prediction with censored quantile regression models. *Biometrics*, 73(2), 517-528.

Marshall, A. W., & Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrika*, 84(3), 641-652.

Meeker, W. Q., & Escobar, L. A. (2014). *Statistical Methods for Reliability Data*. John Wiley & Sons.

Mudholkar, G. S., & Srivastava, D. K. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability*, 42(2), 299-302.

Mudholkar, G. S., Srivastava, D. K., & Freimer, M. (1995). The exponentiated Weibull family: A reanalysis of the bus-motor-failure data. *Technometrics*, 37(4), 436-445.

Nassiri, V., & Loris, I. (2013). A generalized quantile regression model. *Journal of Applied Statistics*, 40(5), 1090-1105.

Nelson, W. (1982). *Applied Life Data Analysis*. John Wiley & Sons.

Pal, M., Ali, M. M., & Woo, J. (2006). Exponentiated Weibull distribution. *Statistica*, 66(2), 139-147.

Reed, W. J. (2011). A flexible parametric survival model which allows a bathtub-shaped hazard rate function. *Journal of Applied Statistics*, 38(8), 1665-1680.

Rubio, F. J., & Hong, Y. (2016). Survival and lifetime data analysis with a flexible class of distributions. *Journal of Applied Statistics*, 43(10), 1794-1813.

Sauerbrei, W., Royston, P., Bojar, H., Schmoor, C., & Schumacher, M. (1999). Modelling the effects of standard prognostic factors in node-positive breast cancer. *British Journal of Cancer*, 79(11), 1752.

Schumacher, M., Bastert, G., Bojar, H., Huebner, K., Olschewski, M., Sauerbrei, W., & Rauschecker, H. F. (1994). Randomized 2 x 2 trial evaluating hormonal treatment and the duration of chemotherapy in node-positive breast cancer patients. German Breast Cancer Study Group. *Journal of Clinical Oncology*, 12(10), 2086-2093.

Slymen, D. J., & Lachenbruch, P. A. (1984). Survival distributions arising from two families and generated by transformations. *Communications in Statistics-Theory and Methods*, 13(10), 1179-1201.

Su, S. (2016). Flexible modelling of survival curves for censored data. *Journal of Statistical Distributions and Applications*, 3(1), 1-20.

Vallejos, C. A., & Steel, M. F. (2015). Objective Bayesian survival analysis using shape mixtures of log-normal distributions. *Journal of the American Statistical Association*, 110(510), 697-710.

Van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press.

Appendix A Proofs of Theorems 2.1 and 2.2 (basic properties)

A.1 Proof of Theorem 2.1

For $t < \eta$, we can write

$$F_\alpha(t;\eta,\phi) = \int_{-\infty}^{t} f_\alpha(s;\eta,\phi)ds = \frac{2\alpha(1-\alpha)}{\phi}\int_0^t g'(s) f_0\big\{(1-\alpha)\big(\frac{g(\eta)-g(s)}{\phi}\big)\big\}ds, \qquad \text{since } s>0$$
$$= 2\alpha F_0\big\{(1-\alpha)\big(\frac{g(t)-g(\eta)}{\phi}\big)\big\},$$

where we used a change of variable $z = (1-\alpha)\big(\frac{g(\eta)-g(s)}{\phi}\big)$. Similarly, for $t \geq \eta$, we have

$$F_\alpha(t;\eta,\phi) = \int_{-\infty}^{t} f_\alpha(s;\eta,\phi)ds = \int_0^\eta f_\alpha(s;\eta,\phi)ds + \int_\eta^t f_\alpha(s;\eta,\phi)ds$$
$$= \frac{2\alpha(1-\alpha)}{\phi}\left[\int_0^\eta g'(s) f_0\big\{(1-\alpha)\big(\frac{g(\eta)-g(s)}{\phi}\big)\big\}ds + \int_\eta^t g'(s) f_0\big\{\alpha\big(\frac{g(s)-g(\eta)}{\phi}\big)\big\}ds\right]$$
$$= 2\alpha - 1 + 2(1-\alpha)F_0\big\{\alpha\big(\frac{g(t)-g(\eta)}{\phi}\big)\big\}, \text{ through a change of variable } z = \alpha\big(\frac{g(s)-g(\eta)}{\phi}\big)$$

Finally, the quantile function of T for any quantile order $\tau \in (0,1)$ given in (2.3) is the inverse of $F_\alpha(t;\eta,\phi)$. □

A.2 Proof of Theorem 2.2

For the survival function given in Theorem 2.2 *(i)*, we use that $S_\alpha(t) = P(T > t) = 1 - F_\alpha(T \leq t)$. Considering this together with the property that $S_0(z) + S_0(-z) = 1$, we can write for $t < \eta$

$$S_\alpha(t;\eta,\phi) = 1 - 2\alpha S_0\big\{(1-\alpha)\big(\frac{g(\eta)-g(t)}{\phi}\big)\big\},$$

and for $t \geq \eta$, we have

$$S_\alpha(t;\eta,\phi) = 2(1-\alpha)S_0\big\{\alpha\big(\frac{g(t)-g(\eta)}{\phi}\big)\big\}.$$

Now, the hazard function for the random variable T is defined by the ratio of the probability density function and the corresponding survival function. Thus,

$$h_\alpha(t;\eta,\phi) = \frac{f_\alpha(t;\eta,\phi)}{S_\alpha(t;\eta,\phi)} = \frac{\alpha g'(t)}{\phi}\begin{cases} \dfrac{2(1-\alpha)h_0\big[(1-\alpha)\big(\frac{g(\eta)-g(t)}{\phi}\big)\big]S_0\big[(1-\alpha)\big(\frac{g(\eta)-g(t)}{\phi}\big)\big]}{1-2\alpha S_0\big[(1-\alpha)\big(\frac{g(\eta)-g(t)}{\phi}\big)\big]} & \text{if } t < \eta \\ h_0\big[\alpha\big(\frac{g(t)-g(\eta)}{\phi}\big)\big] & \text{if } t \geq \eta, \end{cases}$$

where h_0 is the hazard rate function for the underlying distribution with density f_0. Finally, to prove Theorem 2.2 *(iii)* we start from the definition of the cumulative hazard function, and find

$$\begin{aligned}\Lambda_\alpha(t;\eta,\phi) &= \int_0^t h_\alpha(s)ds = -\ln S_\alpha(t;\eta,\phi) \\ &= \begin{cases} -\ln\big\{1-2\alpha S_0\big[(1-\alpha)\big(\dfrac{g(\eta)-g(t)}{\phi}\big)\big]\big\} & \text{if } t < \eta \\ -\ln[2(1-\alpha)] + \Lambda_0\big[\alpha\big(\dfrac{g(t)-g(\eta)}{\phi}\big)\big] & \text{if } t \geq \eta, \end{cases}\end{aligned}$$

where $\Lambda_0(.)$ is the cumulative hazard function associated with the density f_0. □

Appendix B Proofs of Theorems 3.1 and 3.2

B.1 Proof of Theorem 3.1 (consistency)

Since the likelihood function is not differentiable with respect to η at $\eta = y_i$ for all $i = 1, \cdots, n$, the GQBA family does not satisfy the usual regularity conditions to prove the asymptotic properties of MLEs. Huber (1967) established the consistency and asymptotic normality of any sequence of estimators under non-standard conditions. We here follow Huber's method in the right censored data case. Since we assumed independent censoring with non-informative censoring mechanism, we do not have identifiability issues between survival and censoring time distributions. Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space of (Y, Δ) with $\Omega = \mathbb{R}_0^+ \times \{0,1\}$ and $\boldsymbol{\Psi}(\boldsymbol{\theta}; Y, \Delta)$ be a vector of the score function on $\boldsymbol{\Theta}_R \times \Omega$ given as

$$\boldsymbol{\Psi}(\boldsymbol{\theta};Y,\Delta) = \begin{bmatrix}\psi_1(\boldsymbol{\theta};Y,\Delta)\\ \psi_2(\boldsymbol{\theta};Y,\Delta)\\ \psi_3(\boldsymbol{\theta};Y,\Delta)\end{bmatrix} = \begin{bmatrix}\frac{1}{2}\Big(\frac{\partial}{\partial\eta^-}\ell(\boldsymbol{\theta};Y,\Delta) + \frac{\partial}{\partial\eta^+}\ell(\boldsymbol{\theta};Y,\Delta)\Big)\\ \frac{\partial}{\partial\phi}\ell(\boldsymbol{\theta};Y,\Delta)\\ \frac{\partial}{\partial\alpha}\ell(\boldsymbol{\theta};Y,\Delta)\end{bmatrix},$$

where $\frac{\partial}{\partial \eta^-}\ell(\boldsymbol{\theta};Y,\Delta)$ and $\frac{\partial}{\partial \eta^+}\ell(\boldsymbol{\theta};Y,\Delta)$ denote the left-hand and right-hand derivative with respect to η, respectively. The aim is to prove the convergence of any sequence $\widehat{\boldsymbol{\theta}}_n$ such that

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Psi}(\widehat{\boldsymbol{\theta}}_n;Y_i,\Delta_i) \xrightarrow{\text{a.s}} \mathbf{0},$$

converges almost surely to $\boldsymbol{\theta}_0$. We want to give sufficient conditions for the assumptions that have been established in Theorem 2 of Huber (1967, pp.224-225). Sufficient conditions such that Huber's assumptions for the consistency of the estimator $\widehat{\boldsymbol{\theta}}_n$ hold, are:

(H-1). For each fixed $\boldsymbol{\theta} \in \boldsymbol{\Theta}_R$, $\boldsymbol{\Psi}(\boldsymbol{\theta};Y,\Delta)$ is Ω-measurable and separable.

(H-3). The expected value $\lambda(\boldsymbol{\theta}) = \mathrm{E}_{Y,\Delta}\left[\boldsymbol{\Psi}(\boldsymbol{\theta};Y,\Delta)\right]$ exists for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_R$, and has a unique zero at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

(H-2′). As the neighbourhood U_θ of $\boldsymbol{\theta}$ shrinks to $\{\boldsymbol{\theta}_0\}$

$$\mathrm{E}_{Y,\Delta}\left\{u(\boldsymbol{\theta};Y,\Delta)\right\} \to \mathbf{0},$$

where $u(\boldsymbol{\theta};Y,\Delta) = \sup_{\boldsymbol{\theta}_0 \in U_\theta} \|\boldsymbol{\Psi}(\boldsymbol{\theta}_0;Y,\Delta) - \boldsymbol{\Psi}(\boldsymbol{\theta};Y,\Delta)\|$, $\|\cdot\|$ denotes Euclidean norm.

We can prove (H-1) by imposing Assumption (B1), in that we have $\boldsymbol{\Theta}_R$ as a compact subset of the parameter space $\boldsymbol{\Theta}$. It is also well known that a compact space is totally bounded and complete. We also know that a compact space is separable (Van der Vaart, 2000). Furthermore, the score function $\boldsymbol{\Psi}(\boldsymbol{\theta};Y,\Delta)$ is a function of $\boldsymbol{\theta}$ in a compact space $\boldsymbol{\Theta}_R$, and it also depends on the random variables Y and Δ; hence separability is achieved here. Note that, $\boldsymbol{\Psi}(\boldsymbol{\theta};Y,\Delta)$ is a measurable function of (Y,Δ) for each $\boldsymbol{\theta}$ since it is established from a joint probability distribution defined on the measure space $(\Omega,\mathcal{F},\mathcal{P})$.

For the second assumption (H-3), we first need to check the identifiability of the parameters, i.e., for $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, we need to ensure that $\ell(\boldsymbol{\theta};y,\delta) \neq \ell(\boldsymbol{\theta}_0;y,\delta)$. We have that $\exp\left(\ell(\boldsymbol{\theta}_0;y,\delta=1)\right) = f_{\alpha_0}(y;\boldsymbol{\theta}_0)$ and $\exp\left(\ell(\boldsymbol{\theta}_0;y,\delta=0)\right) = S_{\alpha_0}(y;\boldsymbol{\theta}_0) \equiv 1 - F_{\alpha_0}(y;\boldsymbol{\theta}_0)$. In the complete data analysis, Gijbels et al. (2019a) proved the identifiability issue by comparing the mode value in $f_\alpha(y;\boldsymbol{\theta})$ and $f_{\alpha_0}(y;\boldsymbol{\theta}_0)$. Consider $f_\alpha(y;\boldsymbol{\theta}) = f_{\alpha_0}(y;\boldsymbol{\theta}_0)$, which shows that both densities have the same mode $\eta = \eta_0$, since each distribution of GQBA family is unimodal at η. Moreover, $f_\alpha(y;\boldsymbol{\theta}) = f_{\alpha_0}(y;\boldsymbol{\theta}_0)$ implies $S_\alpha(y;\boldsymbol{\theta}) = S_{\alpha_0}(y;\boldsymbol{\theta}_0)$ and also $\alpha = F_\alpha(\eta) = F_{\alpha_0}(\eta_0)) = \alpha_0$, which indicates that $\alpha = \alpha_0$. Therefore, $f_\alpha(\eta;\boldsymbol{\theta}) = f_{\alpha_0}(\eta_0;\boldsymbol{\theta}_0)$ with $\eta = \eta_0$, and $\phi = \phi_0$. Hence, the identifiability of the parameters holds, i.e., $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. In other words, $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \Rightarrow \ell(\boldsymbol{\theta};y,\delta) \neq \ell(\boldsymbol{\theta}_0;y,\delta)$.

The second condition (H-3) we first should check is the existence of the expected score function. Under Assumptions (B2)-(B5) we have a finite expected value of the score function as stated in Proposition 3.1; hence $\lambda(\boldsymbol{\theta})$ exists for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_R$, and has a unique zero at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

For the last assumption (H-2′) we know that $u(\boldsymbol{\theta};Y,\Delta)$ is a continuous function in $\boldsymbol{\theta}$, and $\boldsymbol{\theta}$ belongs to a compact set $\boldsymbol{\Theta}_R$. Hence, $u(\boldsymbol{\theta};Y,\Delta)$ is bounded on $\boldsymbol{\Theta}_R$. It is also immediate from (H-2′) that $\lambda(\boldsymbol{\theta})$ is continuous. □

B.2 Proof of Theorem 3.2 (asymptotic normality)

Similar to the consistency, we also established the asymptotic normality using the conditions of Theorem 3 in Huber (1967, pp.226-227). Huber provides sufficient conditions which assure that every sequence of $\widehat{\boldsymbol{\theta}}_n$ satisfying

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\Psi}(\widehat{\boldsymbol{\theta}}_n;Y_i,\Delta_i) \xrightarrow{P} \mathbf{0},$$

is asymptotically normal.

(N-1). For each fixed $\boldsymbol{\theta} \in \boldsymbol{\Theta}_R$, $\boldsymbol{\Psi}(\boldsymbol{\theta}; Y, \Delta)$ is Ω-measurable and separable.

(N-2). There is a $\boldsymbol{\theta}_0 \in \mathring{\boldsymbol{\Theta}}_R$ such that $\lambda(\boldsymbol{\theta}_0) = \mathbf{0}$.

(N-3). There are strictly positive numbers a, b, c, d_0 such that

(i) $\|\lambda(\boldsymbol{\theta})\| \geq a\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|$ for $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq d_0$,

(ii) $\mathrm{E}_{Y,\Delta}\left[u(Y, \Delta, \boldsymbol{\theta}, d)\right] \leq bd$ for $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| + d \leq d_0$, $d \geq 0$,

(iii) $\mathrm{E}_{Y,\Delta}\left[u(Y, \Delta, \boldsymbol{\theta}, d)^2\right] \leq cd$ for $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| + d \leq d_0$, $d \geq 0$.

Where we define $u(Y, \Delta, \boldsymbol{\theta}, d) = \sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq d} \|\boldsymbol{\Psi}(\boldsymbol{\theta}; Y, \Delta) - \boldsymbol{\Psi}(\boldsymbol{\theta}_0; Y, \Delta)\|$ for $\boldsymbol{\theta}_0 \in \mathring{\boldsymbol{\Theta}}_R$.

(N-4). The expectation $\mathrm{E}_{Y,\Delta}\left[||\boldsymbol{\Psi}(\boldsymbol{\theta}_0; Y, \Delta)||^2\right]$ is finite.

Note that the first assumption (N-1) is the same as (H-1) in the consistency part, and already proved in Section B.1. Furthermore, for the second assumption (N-2), we already proved that $\lambda(\boldsymbol{\theta})$ exists for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_R$ in Proposition 3.1, and has a unique zero at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ in (H-3) of Huber (1967). Since $\mathring{\boldsymbol{\Theta}}_R$ is an interior of $\boldsymbol{\Theta}_R$ and the expectations are always taken with respect to the true distributions $F_{\alpha_0}(y; \eta_0, \phi_0)$ and $G(.)$, we find that $\lambda(\boldsymbol{\theta}_0) = \mathbf{0}$.

The last two assumptions (N-3) and (N-4) involve the existence of the Fisher information matrix. Condition (N-4) holds, i.e., $\mathrm{E}_{Y,\Delta}\left[||\boldsymbol{\Psi}(\boldsymbol{\theta}_0; Y, \Delta)||^2\right] < \infty$, since

$$\begin{aligned}\mathrm{E}_{Y,\Delta}\left[||\boldsymbol{\Psi}(\boldsymbol{\theta}_0; Y, \Delta)||^2\right] &= \mathrm{E}_{Y,\Delta}\left\{\left[\boldsymbol{\Psi}(\boldsymbol{\theta}_0; Y, \Delta)\right]\left[\boldsymbol{\Psi}(\boldsymbol{\theta}_0; Y, \Delta)\right]^T\right\} \\ &= \mathrm{Trace}[\mathcal{I}(\boldsymbol{\theta}_0)] < \infty.\end{aligned}$$

To prove (N-3), a Taylor expansion for $\lambda(\boldsymbol{\theta})$ at the point $\boldsymbol{\theta}_0$ has been used in Gijbels et al. (2019a). The interesting expression in this sense is that since $\lambda(\boldsymbol{\theta})$ is continuously differentiable in any neighbourhood of $\boldsymbol{\theta}_0$, and hence

$$\lambda(\boldsymbol{\theta}) = \lambda(\boldsymbol{\theta}_0) - \mathcal{I}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o\big(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|\big).$$

For further details, see the reference above. □