Robust confidence intervals for log-location-scale models with right censored data

G. Adimari, P.Preo

2003.15

Dipartimento di Scienze Statistiche
Università degli Studi
Via C. Battisti 241-243
35121 Padova

Settembre 2003

ROBUST CONFIDENCE INTERVALS FOR LOG-LOCATION-SCALE MODELS WITH RIGHT CENSORED DATA*

Gianfranco Adimari† and Paolo Preo

Dipartimento di Scienze Statistiche, Università di Padova
via C. Battisti 241/243, 35121 Padova - Italy

Abstract

In this paper we combine empirical likelihood and estimating functions for censored data to obtain robust confidence regions for the parameters and more generally for functions of the parameters of distributions used in lifetime data analysis. The proposed method works with type I or type II or randomly censored data. It is illustrated referring to inference for log-location-scale models. In particular, we focus on the log-normal and the Weibull models and we consider the problem of constructing robust confidence regions (or intervals) for the parameters of the model, as well as for quantiles and values of the survival function. The usefulness of the method is demonstrated through a Monte Carlo study and by examples on two lifetime data sets.

Some key words: Censoring; Empirical likelihood; Estimating function; *M*-estimator; Profile likelihood; Robustness.

**Running title*: ROBUST CONFIDENCE INTERVALS FOR CENSORED DATA

†Corresponding author; e-mail: gianfranco.adimari@unipd.it

1 Introduction

Let $y_1, \ldots, y_n$ be n independent observations on a real-valued random variable Y. Consider a general parametric model $\{F(y;\theta);\ \theta \in \Theta\}$ for Y, where $F(y;\theta) = \text{pr}\{Y \leq y;\theta\}$ denotes the cumulative distribution function (c.d.f.) which depends on an unknown parameter θ belonging to some set $\Theta \subseteq \mathbb{R}^s$, $s \geq 1$. Let $f(y;\theta)$ and $S(y;\theta) = 1 - F(y;\theta)$, respectively, denote the probability density and the survival function corresponding to $F(y;\theta)$.

Inference about θ is often based on unbiased estimating functions. In particular, a general M-estimator for θ is a solution, with respect to θ, of the equation

$$\sum_{i=1}^{n} \alpha(y_i;\theta) = 0, \tag{1}$$

where $\alpha(\cdot;\cdot)$ is a suitable estimating function from $\mathbb{R}\times\Theta$ to $\mathbb{R}^s$, which satisfies $E\{\alpha(Y;\theta)\} = 0$. Examples include maximum likelihood estimators (when $\alpha(y;\theta) = (\partial/\partial\theta)\log f(y;\theta)$), generalized method of moment estimators, quasi-likelihood estimators and many robust estimators.

The maximum likelihood estimator is efficient if the model is correctly specified but is tipically not robust since it may be very sensitive to outliers and small deviations of the data distribution from the assumed model. The level of the gross-error sensitivity specifies the extent to which a single outlier can influence an estimator and it is generally unbounded for the maximum likelihood estimator.

Robust techniques offer the advantage of taking into account automatically possible deviations between the hypothesised model and the actual data distribution. Sacrifying little efficiency at the assumed model such techniques provide protection against outliers and model violation (see Huber, 1981, and Hampel *et al.*, 1986, as general references). Flexibility in choosing the shape of the function $\alpha(\cdot;\cdot)$ in (1) allows for the construction of estimates with desirable robustness properties. In particular, robust M-estimators with bounded gross-error sensitivity correspond to bounded estimating functions. For an optimal robust estimator, the function α may be chosen yielding an estimator whose gross-error sensitivity has a specified bound and within that bound retains as much as possible of the efficiency of the maximum likelihood estimator.

For estimating functions $\alpha(\cdot;\cdot)$ which are not obtained by differentiating a likelihood (or some other) function, it is still possible to develop likelihood ratio-type pivots for testing hypoteses and constructing confidence regions. This may be done by the device of empirical likelihood (Owen, 1988, 1990). In particular, combining empirical likelihood and estimating functions that define robust M-estimators for θ gives rise to confidence regions that, in some way, preserve the robustness features of the associated estimators. Typically such regions have coverage level that is little sensitive with respect to small departures of the data distribution from the specified model. Furthermore, they are generally more accurate than those based on the asymptotic normality of the corresponding M-estimator. We refer to, Adimari (1997), Tsao and Zhou (2001) and Adimari and Ventura (2002) for an analysis of a few specific cases.

In lifetime data analysis, lifetimes $y_1, \ldots, y_n$ are tipically right censored by some censoring values $t_1, \ldots, t_n$ so that one observes (z_i, δ_i), $i = 1, \ldots, n$, where $z_i = \min(y_i, t_i)$ and $\delta_i = I(y_i \leq t_i)$, with $I(\cdot)$ the indicator function. Usual censoring schemes are type I (fixed) censoring, random censoring and type II censoring. In type I censoring, censoring values are assumed to be equal to some fixed constant t, whereas in random censoring they are independent realizations of some random variable, independent of Y. In type II censoring, $t_i = y_{(m)}$ for all i, being m ($< n$) a fixed integer and $y_{(m)}$ the m-th order statistic of $y_1, \ldots, y_n$.

For all the above-mentioned censoring schemes and even more general schemes (see Kalbfleisch and Prentice, 1980, Section 5.2), the log-likelihood function for θ is

$$\ell(\theta) = \sum_{i=1}^{n} \{\delta_i \log f(z_i; \theta) + (1 - \delta_i) \log S(z_i; \theta)\} ,$$

and the maximum likelihood estimate is generally obtained by solving the likelihood equation

$$\sum_{i=1}^{n} \{\delta_i \Delta(z_i; \theta) + (1 - \delta_i) \Delta_c(z_i; \theta)\} = 0 , \qquad (2)$$

where $\Delta(y; \theta) = (\partial/\partial\theta) \log f(y; \theta)$ is the score function and $\Delta_c(y; \theta) = (\partial/\partial\theta) \log S(y; \theta)$. Since $\Delta_c(z_i; \theta) = E\{\Delta(Y; \theta) | Y > z_i\}$ under mild regularity conditions, equation (2) suggests that the natural estimating equation analogue of (1) for censored data is

$$\sum_{i=1}^{n} \eta(z_i, \delta_i; \theta) = 0 , \qquad (3)$$

where $\eta(z_i, \delta_i; \theta) = \delta_i \alpha(z_i; \theta) + (1 - \delta_i) \alpha_c(z_i; \theta)$ and $\alpha_c(z_i; \theta) = E\{\alpha(Y; \theta) | Y > z_i\}$. It follows that robust M-estimators for censored data can be readily obtained using (3) with $\alpha(\cdot; \cdot)$ a robust estimating function for the uncensored case.

James (1986) studies the asymptotic behaviour of the estimators obtained from (3), and we will refer such estimators as James-type M-estimators. The properties of robust estimators within the class of James-type M-estimators are considered by Masarotto and Peracchi (1991), by Akritas *et al.* (1993) in the context of type II censored data, and by Basak (1993) in the context of randomly censored data. In particular, the latter two point out that finding the optimal α-function within the James-class is equivalent to finding the optimal α-function in the uncensored case.

The purpose of the present paper is to combine empirical likelihood and estimating functions that define James-type robust M-estimators. As for the uncensored case, this allows us to obtain likelihood ratio-type pivots for inference with a certain level of protection against outlying observations or small deviations of the data distribution from the assumed model. In particular, we are interested in obtaining robust confidence regions for the parameters and more generally for functions of the parameters of distributions used in lifetime data analysis.

The approach described here presents some desirable features. First, it is quite general and applies to several practical situations with type I or type II or randomly censored data. As we will discuss later, in such situations empirical likelihood confidence regions have a theoretical justification under regularity conditions similar to those stated for the uncensored case. Second, unlike the normal approximation-based approach, the empirical likelihood one does not require the estimation of the asymptotic variance of any statistic. Due to censoring, variance estimates might be rather complicated or unstable. Third, it is range preserving and the corresponding confidence regions are not subject to predetermined symmetry constraints.

We illustrate the proposed method referring to inference for log-location-scale models. In particular, we focus on the log-normal and the Weibull models and we consider the problem of constructing robust confidence regions (intervals, for a scalar parameter of interest) for the parameters of the model, as well as for quantiles and values of the survival function. The usefulness of the method is demonstrated through a Monte Carlo study and by examples on two lifetime data sets. In particular, the Monte Carlo study refers to inference for quantiles and highlights the properties of the empirical likelihood method compared with the classical (parametric) likelihood method. It shows that the empirical likelihood based confidence intervals tend to have coverage levels sufficiently accurate, in general, and reasonably stable under small deviations from the assumed model.

The rest of the paper is organized as follows. In Section 2 we briefly review the empirical likelihood method to obtain confidence regions for parameters defined through estimating functions. Section 3 describes the method proposed in this paper. Simulation results are given in Section 4 whereas applications are presented in Section 5. Some technical issues are addressed in the Appendix.

2 Empirical likelihood and estimating functions

The empirical likelihood function is a nonparametric tool wich allows to obtain tests and confidence regions for parameters expressed as functionals of an unknown distribution. The empirical likelihood method works by profiling a multinomial likelihood supported on the sample. Let F be a generic distribution on the real line and θ the parameter of interest. In this section θ is any general parameter expressed as an M-functional associated with the function $\alpha(y;\theta)$, i.e. $\theta = \theta(F)$ is a root of

$$\int \alpha(y;\theta) dF(y) = 0.$$

We consider the more general case where θ has dimension $s > 1$.

Let $F_0(\cdot)$ denote the unknown c.d.f. of Y and $\theta_0 = \theta(F_0)$ the true parameter value. Based on the (uncensored) sample $y_1, \ldots, y_n$, the empirical likelihood ratio for θ is then defined as follows

$$R(\theta) = \max_{p_i} \prod_{i=1}^{n} np_i \ ,$$

where p_i-weights satisfy $p_i \geq 0 \ \forall i$, $\sum_{i=1}^{n} p_i = 1$ and $\sum_{i=1}^{n} \alpha(y_i;\theta) p_i = 0$. A Lagrangian argument leads to

$$R(\theta) = \prod_{i=1}^{n} \{1 + \lambda' \alpha(y_i;\theta)\}^{-1},$$

when the origin is inside the convex hull of the points $\alpha(y_1;\theta), \ldots, \alpha(y_n;\theta)$. The Lagrangian multiplier $\lambda = \lambda(\theta)$ satisfies (is the unique root of)

$$\sum_{i=1}^{n} \frac{\alpha(y_i;\theta)}{1 + \lambda' \alpha(y_i;\theta)} = 0 \ . \tag{4}$$

Thus, the empirical log-likelihood ratio statistic is

$$l(\theta) = -2 \log R(\theta) = 2 \sum_{i=1}^{n} \log\{1 + \lambda' \alpha(y_i;\theta)\} \tag{5}$$

if θ is such that the origin is inside the convex hull of $\alpha(y_1;\theta),\ldots,\alpha(y_n;\theta)$; otherwise, it is adequate to set $l(\theta)=+\infty$.

Owen (1990) proves under mild conditions (essentially it is required that the covariance matrix of $\alpha(Y;\theta_0)$ is non-singular and finite) that

$$l(\theta_0)\xrightarrow{d}\chi^2_s,$$

where χ^2_s denotes a chi-square distribution with s degrees of freedom. An approximate $(1-\gamma)$-level confidence region for θ may therefore be obtained, in an usual way, as the set of points θ such that $l(\theta)\le c_s(\gamma)$, where $c_s(\gamma)$ is defined such that $\mathrm{pr}\{\chi^2_s > c_s(\gamma)\}=\gamma$.

Under additional regularity conditions (see Qin and Lawless, 1994, 1995), profile empirical likelihood ratio statistics from $R(\theta)$ can also be used to obtain confidence regions for subsets of components or smooth functions of θ. In particular, if the interest focuses on $\tau=h(\theta)$, where $h(\cdot)$ is a smooth $\mathbb{R}^r$-valued function $(r<s)$, the profile empirical log-likelihood ratio statistic for τ is

$$l_P(\tau)=-2\log\{\sup_{\theta:h(\theta)=\tau} R(\theta)\}=\inf_{\theta:h(\theta)=\tau} l(\theta)$$

and $l_P(\tau_0)\xrightarrow{d}\chi^2_r$, with $\tau_0=h(\theta_0)$. Thus a confidence region for τ with asymptotic coverage $(1-\gamma)$ is the set $\{\tau: l_P(\tau)\le c_r(\gamma)\}$.

A general framework for empirical likelihood based on estimating functions is discussed in Qin and Lawless (1994), whereas algorithms for computing profile empirical likelihoods are discussed in Owen (1990) and Hall and La Scala (1990). For an updated overview of the empirical likelihood method and its applications, the reader is referred to the monograph by Owen (2001).

3 The proposed method

To illustrate the proposed method we consider robust inference for log-location-scale models. In particular, we will refer to the log-normal and the Weibull distributions which are frequently used in lifetime models in medical and biological sciences as well as in engineering. In lifetime data analysis, one often consider the logarithm of the lifetime variable so that the resulting transformed lifetime belongs to a location-scale family. The features of location-scale families may facilitate inference. From now on, we assume that the random variable Y describes the logarithm of lifetimes and that the z_i's values are the transformed observed lifetimes.

Let $\theta = (\mu, \sigma)$. We focus on a location-scale family with densities $f(y; \mu, \sigma) = \sigma^{-1} f_*((y - \mu)/\sigma)$, where $f_*(\cdot)$ denotes the density of the standardized ($\mu = 0, \sigma = 1$) element of the family. Let $S_*(\cdot)$ denote the survival function corresponding to $f_*(\cdot)$. For complete data, classical robust M-estimators for the location and scale parameters are defined through the estimating equation (1) with

$$\alpha(y; \theta) = \begin{pmatrix} \alpha_1(y; \mu, \sigma) \\ \alpha_2(y; \mu, \sigma) \end{pmatrix} = \begin{pmatrix} \alpha_1\left(\frac{y-\mu}{\sigma}\right) \\ \alpha_2\left(\frac{y-\mu}{\sigma}\right) \end{pmatrix} \tag{6}$$

for suitable (bounded) real-valued functions $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$. For a symmetric location-scale family of densities, function α_1 will be an odd and α_2 an even function, tipically. For censored data, the related James-type location-scale estimators are obtained from (3) with

$$\eta(z, \delta; \theta) = \begin{pmatrix} \eta_1(z, \delta; \mu, \sigma) \\ \eta_2(z, \delta; \mu, \sigma) \end{pmatrix} = \begin{pmatrix} \delta\alpha_1\left(\frac{z-\mu}{\sigma}\right) + (1-\delta)\alpha_{1c}\left(\frac{z-\mu}{\sigma}\right) \\ \delta\alpha_2\left(\frac{z-\mu}{\sigma}\right) + (1-\delta)\alpha_{2c}\left(\frac{z-\mu}{\sigma}\right) \end{pmatrix}$$

and

$$\begin{aligned} \alpha_{1c}\left(\tfrac{z-\mu}{\sigma}\right) &= E\left\{\alpha_1\left(\tfrac{Y-\mu}{\sigma}\right) | Y > z\right\} = \left\{\int_z^{+\infty} \alpha_1\left(\tfrac{y-\mu}{\sigma}\right) f(y; \mu, \sigma) dy\right\} / S(z; \mu, \sigma), \\ \alpha_{2c}\left(\tfrac{z-\mu}{\sigma}\right) &= E\left\{\alpha_2\left(\tfrac{Y-\mu}{\sigma}\right) | Y > z\right\} = \left\{\int_z^{+\infty} \alpha_2\left(\tfrac{y-\mu}{\sigma}\right) f(y; \mu, \sigma) dy\right\} / S(z; \mu, \sigma). \end{aligned} \tag{7}$$

The above expectations give the contribution of a censored observation to the estimating function η and, in general, must be evaluated numerically.

By equations (4) and (5), when the origin is inside the convex hull of the points $\eta(z_1, \delta_1; \mu, \sigma), \ldots, \eta(z_n, \delta_n; \mu, \sigma)$, the empirical log-likelihood ratio statistic for (μ, σ) based on η and the sample $\{(z_i, \delta_i),\ i = 1, \ldots, n\}$ is

$$l(\mu, \sigma) = 2 \sum_{i=1}^{n} \log\{1 + \lambda' \eta(z_i, \delta_i; \mu, \sigma)\} ,$$

where $\lambda = \lambda(\mu, \sigma)$ satisfies

$$\sum_{i=1}^{n} \frac{\eta(z_i, \delta_i; \mu, \sigma)}{1 + \lambda' \eta(z_i, \delta_i; \mu, \sigma)} = 0 .$$

Profile empirical log-likelihood ratio statistics for a single component or some function $\tau = h(\mu, \sigma)$ of (μ, σ) can be derived from $l(\mu, \sigma)$. For instance, the profile empirical log-likelihood ratio statistic for the scale parameter σ is

$$l_P(\sigma) = \inf_{\mu} l(\mu, \sigma) .$$

Often interest focuses on the q-th quantile of the distribution of Y, that is on $\tau = h(\mu, \sigma) = \mu + y_q^* \sigma$, where y_q^* denotes the corresponding quantile of the standardized element of the location-scale family. Therefore, the profile empirical log-likelihood ratio statistic for the q-th quantile is

$$l_P(\tau) = \inf_{\sigma} l(\tau - y_q^* \sigma, \sigma) \ .$$

The profile empirical log-likelihood ratio statistic for the survival function evaluated at some fixed point y, may be obtained in a similar way. Assume $f_*(y) > 0$ and let $\tau = h(\mu, \sigma) = S_*((y - \mu)/\sigma)$. Then the profile empirical log-likelihood ratio statistic for τ is

$$l_P(\tau) = \inf_{\sigma} l(y - \sigma S_*^{-1}(\tau), \sigma) \ .$$

When data are subject to type I or random censoring (or some combination of them), the pairs (z_i, δ_i), $i = 1, \ldots, n$, are independent realisations of some random variable (Z, δ). Moreover, it can be shown that $E\{\eta(Z, \delta; \mu, \sigma)\} = E\{\alpha(Y; \mu, \sigma)\} = 0$. As a consequence, results by Owen(1990) and Qin and Lawless (1994, 1995) directly apply, so that the chi-square approximation for the distribution of the empirical log-likelihood ratio statistics is justified under the same regularity conditions necessary in the uncensored case. Under type II censoring, the pairs $(z_i, \delta_i), \ldots, (z_n, \delta_n)$ are no longer independent. Thus the above-mentioned results cannot be used directly. However, in the Appendix we argue that the chi-square approximation is still justified under certain regularity conditions.

Therefore, sets as $\{(\mu, \sigma) : l(\mu, \sigma) \leq c_2(\gamma)\}$, $\{\sigma : l_P(\sigma) \leq c_1(\gamma)\}$ and $\{\tau : l_P(\tau) \leq c_1(\gamma)\}$ are approximate $(1 - \gamma)$-level confidence regions for the pair (μ, σ), for the scale parameter σ and for τ (a quantile or the survival function $S(y)$), respectively. For a scalar parameter of interest, such regions are tipically intervals.

Observe that a confidence interval for the quantile of the distribution of the lifetime variable is simply given by $\exp(\cdot)$ applied to each value in the confidence interval for the quantile of the distribution of the log-lifetimes. Similarly, a confidence interval for the survival function of the lifetime variable, evaluated at y, coincides with the confidence interval for the survival function of the log-lifetime variable evaluated at $\log(y)$.

Next we specialize the method to inference for the log-normal and the Weibull models by choosing appropriate location-scale M-estimators.

3.1 Inference for the log-normal distribution

Let Y have a normal distribution, so that $\exp(Y)$ is log-normal. For complete data, a well-known simultaneous location-scale robust estimator is the Huber estimator (Huber, 1964, Proposal 2) defined by (6) with

$$\begin{aligned} \alpha_1(u) &= \psi(u; k_1) \\ \alpha_2(u) &= \psi^2(u; k_2) - a_{k_2} , \end{aligned} \tag{8}$$

for appropriate constants k_1, k_2 and a_{k_2}. The function

$$\psi(u;k) = \begin{cases} k & \text{if } u > k \\ u & \text{if } |u| \le k \\ -k & \text{if } u < -k , \end{cases} \tag{9}$$

is the so-called Huber ψ-function, which defines the optimal bounded influence M-estimator for location at the normal model with known scale. The value of k represents the extent to which one wishes to bound the influence of possible outliers in the sample and is tipically choosen in the range $[1, 1.5]$. The constant a_{k_2} in (8), which ensures unbiasedness (i.e. Fisher consistency) at the normal model, is the expectation of $\psi^2(Y; k_2)$ when the underlying distribution is the standard normal. We have $a_{k_2} = 2k_2^2(1 - \Phi(k_2)) - 2k_2\phi(k_2) + 2\Phi(k_2) - 1$, where $\Phi(\cdot)$ denotes the c.d.f. of the standard normal and $\phi(\cdot)$ its density. Note that when $k_1 = k_2 = +\infty$ the estimating function (8) reduces to the score function.

Extension of the Huber estimator to the censored case is easy. In fact, for the normal model the expectations in (7) can be computed explicitly and do not require numerical integration. In particular, we have

$$\alpha_{1c}(u) = \begin{cases} k_1 & \text{if } u > k_1 \\ \dfrac{k_1(1 - \Phi(k_1)) + \phi(u) - \phi(k_1)}{1 - \Phi(u)} & \text{if } |u| \le k_1 \\ \dfrac{k_1 \Phi(u)}{1 - \Phi(u)} & \text{if } u < -k_1 \end{cases}$$

and

$$
\alpha_{2c}(u) = \begin{cases} k_2^2 - a_{k_2} & \text{if } u > k_2 \\ \dfrac{(a_{k_2}+1)/2 + u\phi(u) - \Phi(u)}{1 - \Phi(u)} - a_{k_2} & \text{if } |u| \le k_2 \\ \dfrac{(a_{k_2} - k_2^2)\Phi(u)}{1 - \Phi(u)} & \text{if } u < -k_2 \,. \end{cases}
$$

Alternative location-scale robust estimators can be considered by choosing differently the function $\alpha_1(\cdot)$ in (8). For instance, the Hampel estimator is obtained by setting that $\alpha_1(\cdot)$ has the shape of the "three-part" function (Hampel *et al.*, 1986, p.150). Extension of the Hampel estimator to the censored case, not given here, is easy too.

3.2 Inference for the Weibull distribution

Let Y follows a smallest extreme values (SEV) distribution, so that $\exp(Y)$ has a Weibull distribution with scale and shape parameters equal to $\exp(\mu)$ and $1/\sigma$, respectively. The standard SEV distribution has density $f_*(u) = \exp\{u - \exp(u)\}$ and survival function $S_*(u) = \exp\{-\exp(u)\}$. For complete data, the score function is

$$
\Delta(y;\theta) = \begin{pmatrix} \Delta_1(y;\mu,\sigma) \\ \Delta_2(y;\mu,\sigma) \end{pmatrix} = \begin{pmatrix} \exp\left(\frac{y-\mu}{\sigma}\right) - 1 \\ \frac{y-\mu}{\sigma}\left\{\exp\left(\frac{y-\mu}{\sigma}\right) - 1\right\} - 1 \end{pmatrix} .
$$

By analogy with the structure of the Huber estimator, a possible simple location-scale robust estimator for the SEV model is then defined by (6) with

$$
\begin{aligned} \alpha_1(u) &= \varphi_1(u;k_{1b}) - a_1 \\ \alpha_2(u) &= \varphi_2(u;k_{2a},k_{2b}) - a_2 \,, \end{aligned} \qquad (10)
$$

for appropriate constants k_{1b}, k_{2a}, k_{2b}, a_1, a_2, and

$$
\begin{aligned} \varphi_1(u;k_{1b}) &= \begin{cases} \exp(k_{1b}) & \text{if } u > k_{1b} \\ \exp(u) & \text{if } u \le k_{1b} \end{cases} \\ \varphi_2(u;k_{2a},k_{2b}) &= \begin{cases} k_{2b}\{\exp(k_{2b}) - 1\} & \text{if } u > k_{2b} \\ u\{\exp(u) - 1\} & \text{if } k_{2a} \le u \le k_{2b} \\ k_{2a}\{\exp(k_{2a}) - 1\} & \text{if } u < k_{2a} \end{cases} \end{aligned} \qquad (11)
$$

The constants a_1 and a_2 ensure unbiasedness and have expression

$$a_1 = 1 - S_*(k_{1b})$$
$$a_2 = k_{2a}\{\exp(k_{2a}) - 1\} + (k_{2a}+1)S_*(k_{2a}) - (k_{2b}+1)S_*(k_{2b}).$$

The threshold values k_{1b}, k_{2a} and k_{2b} can be determined by imposing

$$S_*(k_{1b}) = 1 - \Phi(k_1), \quad 1 - S_*(k_{2a}) = \Phi(-k_2),$$
$$S_*(k_{2b}) = 1 - \Phi(k_2),$$

where k_1 and k_2 are the threshold values that one would choose for the Huber estimator.

Also in this case, extension to the censored setting is easy: the contribution of a censored observation to the estimating function can be computed explicitly. In particular, we have

$$\alpha_{1c}(u) = \begin{cases} \exp(k_{1b}) - a_1 & \text{if } u > k_{1b} \\ \exp(u) + 1 - \dfrac{S_*(k_{1b})}{S_*(u)} - a_1 & \text{if } u \leq k_{1b} \end{cases}$$

and

$$\alpha_{2c}(u) = \begin{cases} k_{2b}\{\exp(k_{2b}) - 1\} - a_2 & \text{if } u > k_{2b} \\ \{u\exp(u) + 1\} - (k_{2b}+1)\dfrac{S_*(k_{2b})}{S_*(u)} - a_2 & \text{if } k_{2a} \leq u \leq k_{2b} \\ k_{2a}\{\exp(k_{2a}) - 1\} + (k_{2a}+1)\dfrac{S_*(k_{2a})}{S_*(u)} - (k_{2b}+1)\dfrac{S_*(k_{2b})}{S_*(u)} - a_2 & \text{if } u < k_{2a} \end{cases}.$$

4 Monte Carlo evidence

In this section we report some results of a Monte Carlo study designed to compare the empirical likelihood-based procedure (ELLR) proposed in this paper, with the classical method based on the log-likelihood ratio statistic (LLR). The objective of the study is both to evaluate the accuracy of the confidence regions when the model is correctly specified and to assess the stability of the coverage levels under small departures from the assumed model.

The study refers to inference for quantiles of the log-normal and the Weibull distributions. In particular, Tables 1-6 give the empirical coverage probabilities of the $(1-\gamma)$-confidence intervals for the first decile, the median and the ninth decile of the distribution of the log-lifetimes, for some nominal levels $1-\gamma$, two sample sizes and different censoring rates π. Each simulation experiment is based on 5000 trials. Type I censoring is considered. The log-lifetimes are generated from the contamination model $F_\varepsilon = (1-\varepsilon)F + \varepsilon G$, where G

denotes the contaminating distribution and ε the contamination percentage which is set at 0% or at 5%.

For the log-normal case (Tables 1-3), without loss of generality we set $F = \Phi$, and we report the results for the situation where G is the c.d.f. of a normal with $\mu = 0$ and $\sigma = 5$. The ELLR intervals are based on the Huber estimator defined by (8) and (9) with $k_1 = k_2 = 1.1$. For the Weibull case (Tables 4-6), we set F to be the c.d.f. of a standard SEV distribution, and we report the results for the situation where G is the cumulative distribution function of a normal random variable with $\mu = -4$ and $\sigma = 5$. The ELLR intervals are based on the estimator defined by (10) and (11), with k_{1b}, k_{2a} and k_{2b} satisfying $S_*(k_{1b}) = 1 - \Phi(1.1)$, $1 - S_*(k_{2a}) = \Phi(-1.1)$ and $S_*(k_{2b}) = 1 - \Phi(1.1)$.

For the empirical coverage probabilities given in Tables 1-6, the estimates of the standard errors can be computed through the binomial formula. Their value decreases with the empirical coverage value, varying approximately between 0.0042 and 0.0014 on the interval $[0.90, 0.99]$.

The following conclusions emerge. When there is no contamination ($\varepsilon = 0$), both for the log-normal and the Weibull case, the ELLR approach performs similarly to the LLR one and yields confidence intervals with accurate coverage probabilities, even for relatively small sample size, i.e. when the expected number of uncensored observations is 12 ($n = 20$ and $\pi = 0.4$). Moreover, the coverage probabilities of the ELLR confidence intervals are reasonably stable under small contaminations. On the contrary, the LLR confidence intervals can be strongly affected by small amounts of contamination in the data distribution. In particular, the lack of robustness for the LLR procedure seems remarkable when the parameter of interest is an extreme quantile (see Tables 1, 3, 4 and 6, which refer to the first and the ninth decile). Similar conclusions arise under type II or random censoring and, for the log-normal case, if we use the Hampel estimator.

5 Examples

We present two examples based on real data sets. These examples illustrate some applications of the proposed method and its usefulness compared with the classical (parametric) likelihood ratio based approach.

Example 5.1 We consider the data from Stablein *et al.* (1981). The data we use are survival times (in days) of 45 patients suffering from gastric cancer which received a combined treatment of both chemotherapy and radiation therapy. There are 7 censored observations; so the observed censoring rate is near to 16%. An explorative analysis suggests the use of the Weibull model for the lifetimes. Therefore, we consider the estimator discussed in Section 3.2, with $k_{1b} = k_{2b} = \log(-\log\{\Phi(-1.1)\})$ and $k_{2a} = \log(-\log\{\Phi(1.1)\})$, and make inference for the scale parameter σ of the distribution of the log-lifetime.

Figure 1 gives the profile empirical log-likelihood ratio statistic (ELLR) and the profile log-likelihood ratio statistic (LLR) for σ. Figure 1 also shows the effect of replacing the largest complete observation by an even larger value and demonstrates the benefits of using the ELLR approach. When the largest log-lifetime 7.22 is replaced by 10, the profile log-likelihood ratio statistic shifts remarkably, whereas this does not occur for the empirical log-likelihood ratio statistic. Clearly, the impact of such a shift on the inference may be significant. Observe, in particular, that the 0.95-level confidence interval for σ based on the LLR approach and the modified sample does not include the value 1. This means that the likelihood ratio test based on the modified sample would lead to reject the hypothesis $H_0 : \sigma = 1$, at the 5% significance level. Such a hypothesis, which refers to the choice of an exponential model for the lifetime variable, is accepted when the likelihood ratio test is based on the original sample or if we use the empirical log-likelihood ratio statistic.

Example 5.2 We consider the data from Klein and Moeschberger (1997, Section 1.9), which refer to patients with advanced acute myelogenous leukemia. The data we use are leukemia-free survival times (in months) of 51 patients which received autologous (auto)bone marrow transplant in which, after high doses of chemoterapy, their own marrow was reinfused to replace their destroyed immune system. There are 23 censored observations; so the observed censoring rate is near to 45%. An explorative analysis suggests the use of the log-normal model for the lifetime. Therefore, we use the Huber estimator discussed in Section 3.1, with $k_1 = k_2 = 1.5$, and we consider inference for the ninth decile and the survival function (evaluated at $y = 5$) of the distribution of the log-lifetimes.

Figure 2 gives the profile empirical log-likelihood ratio statistic and the profile log-likelihood ratio statistic for the ninth decile of the distribution of the log-lifetimes and shows the effect of replacing the smallest complete observation by an even smaller value. As for the Example 5.1, when the smallest log-lifetime -0.418 is replaced by -3, the profile log-likelihood ratio statistic shifts, whereas the empirical log-likelihood ratio statistic shows to be stable. This behaviour may be also observed in Figure 3 and confirms the robustness features of the ELLR approach. Figure 3 gives the profile empirical log-likelihood ratio statistic and the profile log-likelihood ratio statistic for the survival function of the distribution of the log-lifetimes, evaluated at $y = 5$.

Appendix

In this appendix we examine the asymptotic behaviour of the empirical log-likelihood ratio statistic $l(\theta)$ and the profile empirical log-likelihood ratio statistic $l_P(\tau)$, obtained by estimating equations of the form (3), under type II censoring. We consider the model $\{F(y;\theta);\ \theta \in \Theta\}$ for Y (a lifetime or a log-lifetime variable), and denote by θ_0 the true parameter value. The profile empirical log-likelihood ratio statistic refers to the parameter $\tau = h(\theta)$, where $h(\cdot)$ is a $\mathbb{R}^r$-valued function which admits the Taylor expansion

$$h(\theta) = \tau_0 + V_0(\theta - \theta_0) + o(\|\theta - \theta_0\|) \ , \tag{12}$$

with $\tau_0 = h(\theta_0)$. In (12), V_0 denotes the Jacobian matrix (assumed to be of full rank r) calculated at θ_0. We first give a general result on the asymptotic behaviour of the empirical log-likelihood ratio statistics. We then establish sufficient conditions under which this result can be applied to data subject to type II censoring.

Step 1: general result. We assume that, for n large enough, there exists a neighbourhood $\mathcal{B}$ of θ_0 where, with probability one, the empirical log-likelihood ratio statistic $l(\theta)$ exists (finite) and has expression

$$l(\theta) = -2 \log R(\theta) = 2 \sum_{i=1}^{n} \log\{1 + \lambda' \eta(z_i, \delta_i; \theta)\}, \tag{13}$$

with $\lambda = \lambda(\theta)$ satisfying

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\eta(z_i, \delta_i; \theta)}{1 + \lambda' \eta(z_i, \delta_i; \theta)} = 0 \ . \tag{14}$$

For $\theta \in \mathcal{B}$, let $\rho = \rho(\theta) = \|\lambda\|$. Then $\lambda = \rho\xi$ where $\xi = \rho^{-1}\lambda$ and $\|\xi\| = 1$. Following Owen (1990), one can show that

$$\rho \leq \frac{|u'\bar{\eta}(\theta)|}{\xi'\bar{\Omega}(\theta)\xi - \bar{\nu}(\theta)|u'\bar{\eta}(\theta)|}, \tag{15}$$

where $u = (1, \dots, 1)'$, $\quad \bar{\Omega}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \eta(z_i, \delta_i; \theta)\eta'(z_i, \delta_i; \theta)$, $\quad \bar{\eta}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \eta(z_i, \delta_i; \theta)$ and $\bar{\nu}(\theta) = \max_{1 \le i \le n} \|\eta(z_i, \delta_i; \theta)\|$. We assume that

A1) $n^{1/2}\bar{\eta}(\theta_0)$ is asymptotically normal with zero mean and covariance matrix $\Omega(\theta_0)$ non-singular and finite,

and that, for any sequence of random points θ such that $\|\theta - \theta_0\| = O(n^{-1/2})$ almost surely (a.s.),

A2) $\bar{\eta}(\theta) = \bar{\eta}(\theta_0) + \{J(\theta_0) + o_p(1)\}(\theta - \theta_0)$, for some non-singular and finite matrix $J(\theta_0)$,

A3) $\bar{\Omega}(\theta) = \Omega(\theta_0) + o_p(1)$,

A4) $\bar{\nu}(\theta) = o(n^{1/2})$ a.s.,

A5) $\frac{1}{n}\sum_{i=1}^{n} \|\eta(z_i, \delta_i; \theta)\|^2 = O_p(1)$.

Let ω be the smallest eingenvalue of $\Omega(\theta_0)$. Then, by (15) and assumptions A1-A4, for a sequence of random points θ such that $\|\theta - \theta_0\| = O(n^{-1/2})$ a.s., we have

$$\rho \le \frac{O_p(n^{-1/2})}{\xi'\Omega(\theta_0)\xi + o_p(1)} \,.$$

It follows that $\lambda = O_p(n^{-1/2})$ because $\xi'\Omega(\theta_0)\xi \ge \omega > 0$.

Observe that, by assumption A4, $\max_i |\lambda'\eta(z_i\delta_i; \theta)| = O_p(n^{-1/2})o(n^{1/2}) = o_p(1)$. Then we use the expansion

$$\{1 + \lambda'\eta(z_i, \delta_i; \theta)\}^{-1} = 1 - \lambda'\eta(z_i, \delta_i; \theta) + \frac{\{\lambda'\eta(z_i, \delta; \theta)\}^2}{1 + \lambda'\eta(z_i, \delta_i; \theta)}$$

in equation (14) to obtain

$$\begin{aligned} 0 &= \bar{\eta}(\theta) - \left\{\frac{1}{n}\sum_{i=1}^{n} \eta(z_i, \delta_i; \theta)\eta'(z_i, \delta_i; \theta)\right\}\lambda + \frac{1}{n}\sum_{i=1}^{n} \frac{\{\lambda'\eta(z_i, \delta_i; \theta)\}^2}{1 + \lambda'\eta(z_i\delta_i; \theta)}\eta(z_i, \delta_i; \theta) \\ &= \bar{\eta}(\theta) - \bar{\Omega}(\theta)\lambda + \text{Remainder}. \end{aligned}$$

By assumptions A4-A5, the Remainder has norm bounded by

$$\frac{1}{n}\sum \|\eta(z_i, \delta_i; \theta)\|^3 \|\lambda\|^2 |1 + \lambda'\eta(z_i, \delta_i; \theta)|^{-1} = o_p(n^{1/2})O_p(n^{-1})O_p(1) = o_p(n^{-1/2}),$$

so that

$$\lambda = \Omega^{-1}(\theta_0)\bar{\eta}(\theta) + o_p(n^{-1/2}), \tag{16}$$

using also assumption A3 and the fact that $\lambda = O_p(n^{-1/2})$.

In a similar way we use the McLaurin series expansion of $\log(1+x)$ in the right-hand side of equation (13). We can write

$$l(\theta) = 2\sum_{i=1}^{n} \lambda'\eta(z_i, \delta_i; \theta) - \lambda' \left\{ \sum_{i=1}^{n} \eta(z_i, \delta_i; \theta)\eta'(z_i, \delta_i; \theta) \right\} \lambda + \frac{2}{3}\sum_{i=1}^{n} \frac{\{\lambda'\eta(z_i, \delta_i; \theta)\}^3}{(1+\beta_i)^3},$$

where $|\beta_i| < |\lambda'\eta(z_i, \delta_i; \theta)|$. It is easy to establish that $\sum_{i=1}^{n}\{\lambda'\eta(z_i, \delta_i; \theta)\}^3/(1+\beta_i)^3 = o_p(1)$. Hence, using (16) and the fact that $\bar{\eta}(\theta) = O_p(n^{-1/2})$, we obtain

$$l(\theta) = n\bar{\eta}'(\theta)\Omega^{-1}(\theta_0)\bar{\eta}(\theta) + o_p(1). \tag{17}$$

Finally, by assumption A2,

$$l(\theta) = w_n(\theta) + o_p(1), \tag{18}$$

where

$$w_n(\theta) = n\left\{\bar{\eta}_*(\theta_0) + \theta - \theta_0\right\}' \Omega_*^{-1}(\theta_0) \left\{\bar{\eta}_*(\theta_0) + \theta - \theta_0\right\},$$

with $\bar{\eta}_*(\cdot) = J^{-1}(\theta_0)\bar{\eta}(\cdot)$ and $\Omega_*(\theta_0) = \{J'(\theta_0)\Omega^{-1}(\theta_0)J(\theta_0)\}^{-1}$.

Result (18) holds for any sequence of random points θ such that $\|\theta - \theta_0\| = O(n^{-1/2})$ a.s. and can be used to show that

$$\inf_{\theta \in \mathcal{C}_n : V_0(\theta - \theta_0) = 0} l(\theta) = \min_{\theta \in \mathcal{C}_n : V_0(\theta - \theta_0) = 0} w_n(\theta) + o_p(1)\ , \tag{19}$$

for any sequence of balls $\mathcal{C}_n = \{\theta : \|\theta - \theta_0\| \le Cn^{-1/2}\}$, where $C > 0$ is some constant. Indeed, suppose that (19) does not hold. Then there exist a sequence of balls $\mathcal{C}_n^\circ$, an $\epsilon > 0$ and a $\kappa > 0$ such that

$$\mathrm{pr}\{|\inf_{\theta \in \mathcal{C}_n^\circ : V_0(\theta - \theta_0) = 0} l(\theta) - \min_{\theta \in \mathcal{C}_n^\circ : V_0(\theta - \theta_0) = 0} w_n(\theta)| > \epsilon\} > \kappa\ ,$$

infinitely often as $n \to \infty$. Let $\tilde{\theta}_l = \arg \inf_{\theta \in \mathcal{C}_n^\circ : V_0(\theta - \theta_0) = 0} l(\theta)$, $\tilde{\theta}_w = \arg \min_{\theta \in \mathcal{C}_n^\circ : V_0(\theta - \theta_0) = 0} w_n(\theta)$ and

$$\tilde{\theta} = \begin{cases} \tilde{\theta}_l & \text{if} \quad w_n(\tilde{\theta}_w) \ge l(\tilde{\theta}_l) \\ \tilde{\theta}_w & \text{if} \quad w_n(\tilde{\theta}_w) < l(\tilde{\theta}_l). \end{cases}$$

Then,

$$\mathrm{pr}\{|l(\tilde{\theta}) - w_n(\tilde{\theta})| > \epsilon\} \ge \mathrm{pr}\{|\inf_{\theta \in \mathcal{C}_n^\circ : V_0(\theta - \theta_0) = 0} l(\theta) - \min_{\theta \in \mathcal{C}_n^\circ : V_0(\theta - \theta_0) = 0} w_n(\theta)| > \epsilon\} > \kappa$$

infinitely often as $n \to \infty$. This finding contradicts result (18), since $\|\tilde{\theta} - \theta_0\| = O(n^{-1/2})$ a.s. by construction. Hence, result (18) implies result (19), so that, within any sequence of balls $\mathcal{C}_n$, the infimum of $l(\theta)$ subject to the constraint $V_0(\theta - \theta_0) = 0$ may be approximated through the minimum of the function $w_n(\theta)$ subject to the same constraint.

On the other hand, a Lagrangian argument leads to

$$\min_{\theta: V_0(\theta-\theta_0)=0} w_n(\theta) = n\left\{V_0\bar{\eta}_*(\theta_0)\right\}'\left\{V_0\Omega_*(\theta_0)V_0'\right\}^{-1}\left\{V_0\bar{\eta}_*(\theta_0)\right\} \tag{20}$$

and

$$\hat{\theta} = \arg\min_{\theta: V_0(\theta-\theta_0)} w_n(\theta) = \theta_0 - \bar{\eta}_*(\theta_0) + \Omega_*(\theta_0)V_0'\left\{V_0\Omega_*(\theta_0)V_0'\right\}^{-1} V_0\bar{\eta}_*(\theta_0).$$

By assumption A1, $\|\hat{\theta} - \theta_0\| = O_p(n^{-1/2})$. Therefore, for each $\epsilon > 0$, there exist a constant $C_\epsilon > 0$ and an integer n_ϵ such that

$$\Pr\{\|\hat{\theta} - \theta_0\| \leq C_\epsilon n^{-1/2}\} > 1 - \epsilon$$

when $n > n_\epsilon$. This means that we can choose a sequence of balls $\mathcal{C}_n^* = \{\theta : \|\theta - \theta_0\| \leq C^* n^{-1/2}\}$ such that, for n large enough,

$$\hat{\theta} = \arg\min_{\theta \in \mathcal{C}_n^*: V_0(\theta-\theta_0)=0} w_n(\theta) \tag{21}$$

with arbitrarily high probability. For such a sequence of balls and n large enough, we have

$$\begin{aligned}
\inf_{\theta \in \mathcal{C}_n^*: h(\theta)=\tau_0} l(\theta) &= \inf_{\theta \in \mathcal{C}_n^*: V_0(\theta-\theta_0)=0} l(\theta) \\
&= \min_{\theta \in \mathcal{C}_n^*: V_0(\theta-\theta_0)=0} w_n(\theta) + o_p(1) \\
&= \min_{\theta: V_0(\theta-\theta_0)=0} w_n(\theta) + o_p(1) \\
&= n\left\{V_0\bar{\eta}_*(\theta_0)\right\}'\left\{V_0\Omega_*(\theta_0)V_0'\right\}^{-1}\left\{V_0\bar{\eta}_*(\theta_0)\right\} + o_p(1),
\end{aligned}$$

using (12), (18), (20) and (21). Since $n^{1/2}V_0\bar{\eta}_*(\theta_0)$ is asymptotically normal with zero mean and covariance matrix $V_0\Omega_*(\theta_0)V_0'$, it follows that, under assumptions A1-A5, $l_P(\tau_0) = \inf_{\theta: h(\theta)=\tau_0} l(\theta)$ converges to a χ_r^2, at least when the infimum is taken within a suitable sequence of balls centered at θ_0. Moreover, from (17), we have that $l(\theta_0) \xrightarrow{d} \chi_s^2$, under assumption A1 and assumptions A3-A5 restricted to $\theta = \theta_0$.

Step 2: result for type II censoring. In the following we give some conditions sufficient to ensure that assumptions A1-A5 hold when data are subject to type II censoring.

Observe that for a type II right-censored sample, the function $\bar{\eta}(\theta)$ takes the form

$$\bar{\eta}(\theta) = \frac{1}{n}\sum_{i=1}^{m} \alpha(y_{(i)};\theta) + \frac{n-m}{n}\alpha_c(y_{(m)};\theta),$$

where $y_{(1)}, \dots, y_{(m)}$ denote the first m ($< n$) order statistics of $y_1, \dots, y_n$. Similarly,

$$\bar{\Omega}(\theta) = \frac{1}{n}\sum_{i=1}^{m} \alpha(y_{(i)};\theta)\alpha'(y_{(i)};\theta) + \frac{n-m}{n}\alpha_c(y_{(m)};\theta)\alpha_c'(y_{(m)};\theta).$$

Let ζ be the p-th quantile of $f(y;\theta_0)$. We assume that

B1) $m/n \to p$ as $n \to \infty$,

B2) $f(\cdot;\theta_0)$ and $\alpha(\cdot;\theta_0)$ are continuous at ζ, with $f(\zeta;\theta_0) > 0$,

B3) $\partial\alpha_c(y;\theta_0)/\partial y$ exists at $y = \zeta$.

Moreover, we assume that there exists a compact neighborhood $\mathcal{B}^\circ$ of θ_0 such that

B4) $\alpha(y;\theta)$ is continuous in $\theta \in \mathcal{B}^\circ$ for every y,

B5) for $\theta \in \mathcal{B}^\circ$ and all y, $\|\alpha(y;\theta)\| \le g(y)$, for some function $g(\cdot)$ such that $\int g^2(y)f(y;\theta_0)dy < \infty$,

B6) $\alpha_c(y;\theta)$ is continuous in $[\zeta - \epsilon_1, \zeta + \epsilon_1] \times \mathcal{B}^\circ$, for some $\epsilon_1 > 0$.

Then, by Theorem 1 of Bhattacharyya (1985), $n^{1/2}\bar{\eta}(\theta_0)$ is asymptotically normal with mean 0 and covariance matrix

$$\begin{aligned}\Omega(\theta_0) &= \int_{-\infty}^{\zeta} \alpha(y;\theta_0)\alpha'(y;\theta_0)f(y;\theta_0)dy \\ &+ q^{-1}\left(\int_{-\infty}^{\zeta} \alpha(y;\theta_0)f(y;\theta_0)dy\right)\left(\int_{-\infty}^{\zeta} \alpha(y;\theta_0)f(y;\theta_0)dy\right)',\end{aligned}$$

with $q = 1 - p$. Moreover, by the continuity of the function

$$\Omega(\theta) = \int_{-\infty}^{\zeta} \alpha(y;\theta)\alpha'(y;\theta)f(y;\theta_0)dy + q\alpha_c(\zeta;\theta)\alpha_c'(\zeta;\theta)$$

at θ_0 and an application of Theorem 2 of Bhattacharyya (1985) to the matrix-valued function $\bar{\Omega}(\theta)$, we have

$$\bar{\Omega}(\theta) = \Omega(\theta_0) + o(1) \quad \text{a.s.}$$

for any sequence of random points θ that converges to θ_0 almost surely. On the other hand, for $||\theta - \theta_0|| = o(1)$ a.s. and n sufficiently large,

$$\bar{\nu}(\theta) \leq \max\{\max_{1\leq i\leq n} ||\alpha(y_i;\theta)||, ||\alpha_c(y_{(m)};\theta)||\} \leq \max\{\max_{1\leq i\leq n} g(y_i), ||\alpha_c(y_{(m)};\theta)||\} = o(n^{1/2}) \text{ a.s.},$$

by using Lemma 2 of Owen (1990), the continuity of $\alpha_c(y;\theta)$ at $(\zeta;\theta_0)$ and the fact that $y_{(m)} \to \zeta$ with probability one. We have also

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} ||\eta(z_i,\delta_i;\theta)||^2 &\leq \frac{1}{n}\sum_{i=1}^{m} ||\alpha(y_{(i)};\theta)||^2 + \frac{n-m}{n}||\alpha_c(y_{(m)};\theta)||^2 \\ &\leq \frac{1}{n}\sum_{i=1}^{n} g^2(y_i) + \frac{n-m}{n}||\alpha_c(y_{(m)};\theta)||^2 = O_p(1), \end{aligned}$$

again for $||\theta - \theta_0|| = o(1)$ a.s. and n large enough. Therefore, conditions B1-B6 are sufficient to make the assumptions A1 and A3-A5 hold. It follows that, under type II censoring and conditions B1-B6, $l(\theta_0) \xrightarrow{d} \chi^2_s$.

Further regularity conditions are necessary to ensure that assumption A2 holds. In the following we use an upper dot to denote the derivative of a function with respect to θ. We assume that

B7) $\dot{\alpha}_c(y;\theta)$ exists in a neighborhood of θ_0, for all y in a neighborhood of ζ,

B8) $\dot{\alpha}(y;\theta)$ exists in a neighborhood of θ_0, for all y,

B9) conditions B4-B6 are satisfied by the functions $\dot{\alpha}(y;\theta)$ and $\dot{\alpha}_c(y;\theta)$ and the matrix $\int_{-\infty}^{\zeta} \dot{\alpha}(y;\theta_0)f(y;\theta_0)dy + q\dot{\alpha}_c(\zeta;\theta_0)$ is non-singular and finite.

Then, a Taylor expansion of $\bar{\eta}(\theta)$ around θ_0 yields $\bar{\eta}(\theta) = \bar{\eta}(\theta_0) + \dot{\bar{\eta}}(\theta^{\bullet})(\theta - \theta_0) + o(||\theta - \theta_0||)$, where

$$\dot{\bar{\eta}}(\theta) = \frac{1}{n}\sum_{i=1}^{m} \dot{\alpha}(y_{(i)};\theta) + \frac{n-m}{n}\dot{\alpha}_c(y_{(m)};\theta)$$

and $\theta^{\bullet}$ lies on the line segment between θ and θ_0. Hence, for any sequence of random points θ such that $||\theta - \theta_0|| = O(n^{-1/2})$ a.s., we have

$$\bar{\eta}(\theta) = \bar{\eta}(\theta_0) + \{J(\theta_0) + o_p(1)\}(\theta - \theta_0) , \tag{22}$$

where $J(\theta) = \int_{-\infty}^{\zeta} \dot{\alpha}(y;\theta)f(y;\theta_0)dy + q\dot{\alpha}_c(\zeta;\theta)$, by an application of Theorem 2 of Bhattacharyya (1985) to the matrix-valued function $\dot{\bar{\eta}}(\theta)$. It follows that, under type II censoring and conditions B1-B9, $l_P(\tau_0) = \inf_{\theta:h(\theta)=\tau_0} l(\theta)$ converges to a χ^2_r, at least when the infimum is taken within a suitable sequence of balls centered at θ_0.

Conditions B7-B9 facilitate the proof but are stronger than necessary. Actually, it can be shown that, under the following weaker conditions,

B7a) $\dot{\alpha}_c(y;\theta)$ exists in a neighborhood of θ_0 (for all y in a neighborhood of ζ) and $\dot{\alpha}_c(y;\theta_0)$ is continuous at $y = \zeta$,

B8a) (Lipschitz condition) for every θ_1 and θ_2 in a neighborhood of θ_0 and all y, $\|\alpha(y;\theta_1) - \alpha(y;\theta_2)\| \leq g_\circ(y)\|\theta_1 - \theta_2\|$, for some function $g_\circ(\cdot)$ bounded in a neighborhood of ζ and such that $\int g_\circ(y) f(y;\theta_0) dy < \infty$,

B9a) the function $\Lambda(\theta) = \int_{-\infty}^{\zeta} \alpha(y;\theta) f(y;\theta_0) dy \; + \; q\alpha_c(\zeta;\theta)$ is differentiable at θ_0 with non-singular derivative matrix $\dot{\Lambda}(\theta_0)$,

(22) holds with $J(\theta_0) = \dot{\Lambda}(\theta_0)$, for any sequence of random points θ such that $\|\theta - \theta_0\| = O(n^{-1/2})$ a.s..

Table 1: *Empirical coverage probabilities of the confidence intervals for the first decile of the distribution of the log-lifetimes. Log-normal case.*

		n=20		n=40		
$1-\gamma$		π=0.2	π=0.4	π=0.2	π=0.4	π=0.6
0.99	ELLR	0.984	0.982	0.990	0.989	0.986
	LLR	0.986	0.989	0.988	0.991	0.987
0.95	ELLR	0.943	0.942	0.952	0.945	0.945
	LLR	0.941	0.948	0.949	0.948	0.945
0.90	ELLR	0.892	0.894	0.907	0.894	0.892
	LLR	0.885	0.896	0.890	0.898	0.893

($\varepsilon = 0$)

		n=20		n=40		
$1-\gamma$		π=0.2	π=0.4	π=0.2	π=0.4	π=0.6
0.99	ELLR	0.984	0.982	0.987	0.987	0.985
	LLR	0.853	0.876	0.762	0.774	0.844
0.95	ELLR	0.944	0.937	0.944	0.943	0.939
	LLR	0.776	0.787	0.665	0.673	0.728
0.90	ELLR	0.896	0.887	0.896	0.897	0.881
	LLR	0.717	0.722	0.607	0.610	0.650

($\varepsilon = 0.05$)

Table 2: *Empirical coverage probabilities of the confidence intervals for the median of the distribution of the log-lifetimes. Log-normal case.*

		n=20		n=40		
$1-\gamma$		π=0.2	π=0.4	π=0.2	π=0.4	π=0.6
0.99	ELLR	0.991	0.989	0.990	0.990	0.993
	LLR	0.985	0.988	0.990	0.990	0.989
0.95	ELLR	0.945	0.949	0.950	0.949	0.957
	LLR	0.940	0.946	0.948	0.948	0.951
0.90	ELLR	0.896	0.905	0.898	0.897	0.909
	LLR	0.883	0.899	0.890	0.895	0.904

($\varepsilon = 0$)

		n=20		n=40		
$1-\gamma$		π=0.2	π=0.4	π=0.2	π=0.4	π=0.6
0.99	ELLR	0.991	0.989	0.990	0.990	0.989
	LLR	0.991	0.990	0.992	0.993	0.978
0.95	ELLR	0.948	0.949	0.951	0.950	0.951
	LLR	0.948	0.952	0.961	0.960	0.918
0.90	ELLR	0.899	0.901	0.902	0.906	0.905
	LLR	0.900	0.908	0.913	0.919	0.852

($\varepsilon = 0.05$)

Table 3: *Empirical coverage probabilities of the confidence intervals for the ninth of the distribution of the log-lifetimes. Log-normal case.*

		n=20		n=40		
$1-\gamma$		π=0.2	π=0.4	π=0.2	π=0.4	π=0.6
0.99	ELLR	0.986	0.994	0.993	0.993	0.994
	LLR	0.986	0.990	0.990	0.990	0.992
0.95	ELLR	0.952	0.962	0.952	0.957	0.962
	LLR	0.942	0.945	0.948	0.946	0.947
0.90	ELLR	0.907	0.920	0.906	0.909	0.918
	LLR	0.895	0.893	0.899	0.891	0.895

($\varepsilon = 0$)

		n=20		n=40		
$1-\gamma$		π=0.2	π=0.4	π=0.2	π=0.4	π=0.6
0.99	ELLR	0.989	0.991	0.989	0.993	0.994
	LLR	0.948	0.917	0.860	0.814	0.772
0.95	ELLR	0.958	0.960	0.953	0.958	0.968
	LLR	0.870	0.854	0.765	0.718	0.688
0.90	ELLR	0.913	0.911	0.902	0.915	0.923
	LLR	0.809	0.786	0.700	0.653	0.630

($\varepsilon = 0.05$)

Table 4: *Empirical coverage probabilities of the confidence intervals for the first decile of the distribution of the log-lifetimes. Weibull case.*

		n=20		n=40		
$1-\gamma$		π=0.2	π=0.4	π=0.2	π=0.4	π=0.6
0.99	ELLR	0.986	0.985	0.986	0.989	0.990
	LLR	0.988	0.993	0.989	0.990	0.991
0.95	ELLR	0.950	0.945	0.948	0.947	0.951
	LLR	0.948	0.950	0.954	0.949	0.948
0.90	ELLR	0.895	0.896	0.897	0.895	0.904
	LLR	0.891	0.903	0.901	0.895	0.898

($\varepsilon = 0$)

		n=20		n=40		
$1-\gamma$		π=0.2	π=0.4	π=0.2	π=0.4	π=0.6
0.99	ELLR	0.993	0.982	0.983	0.982	0.981
	LLR	0.930	0.929	0.894	0.876	0.889
0.95	ELLR	0.953	0.935	0.940	0.928	0.925
	LLR	0.835	0.823	0.776	0.759	0.749
0.90	ELLR	0.903	0.879	0.886	0.866	0.861
	LLR	0.763	0.740	0.695	0.678	0.659

($\varepsilon = 0.05$)

Table 5: *Empirical coverage probabilities of the confidence intervals for the median of the distribution of the log-lifetimes. Weibull case.*

		n=20		n=40		
$1-\gamma$		π=0.2	π=0.4	π=0.2	π=0.4	π=0.6
0.99	ELLR	0.990	0.986	0.989	0.992	0.994
	LLR	0.987	0.985	0.989	0.990	0.990
0.95	ELLR	0.952	0.946	0.955	0.949	0.960
	LLR	0.946	0.944	0.951	0.947	0.951
0.90	ELLR	0.905	0.893	0.908	0.897	0.920
	LLR	0.895	0.890	0.905	0.897	0.892

($\varepsilon = 0$)

		n=20		n=40		
$1-\gamma$		π=0.2	π=0.4	π=0.2	π=0.4	π=0.6
0.99	ELLR	0.988	0.988	0.985	0.989	0.991
	LLR	0.987	0.988	0.984	0.991	0.989
0.95	ELLR	0.948	0.950	0.940	0.945	0.953
	LLR	0.936	0.953	0.926	0.951	0.946
0.90	ELLR	0.894	0.896	0.890	0.895	0.910
	LLR	0.877	0.899	0.866	0.900	0.897

($\varepsilon = 0.05$)

Table 6: *Empirical coverage probabilities of the confidence intervals for the ninth of the distribution of the log-lifetimes. Weibull case.*

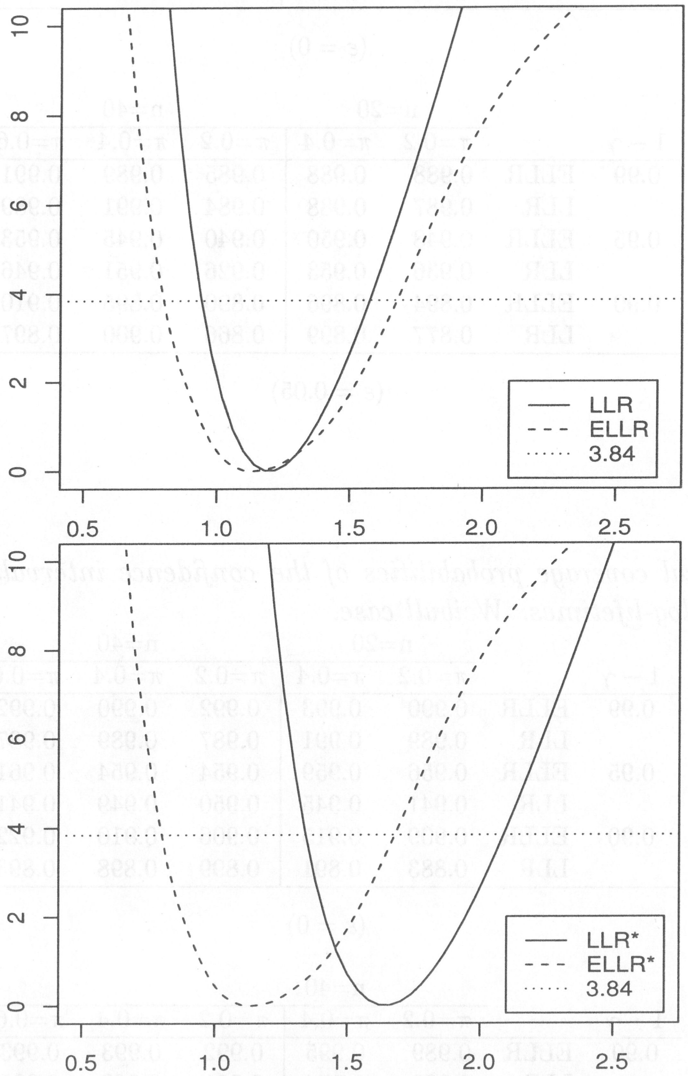
		n=20		n=40		
$1-\gamma$		π=0.2	π=0.4	π=0.2	π=0.4	π=0.6
0.99	ELLR	0.990	0.993	0.992	0.990	0.992
	LLR	0.989	0.991	0.987	0.989	0.987
0.95	ELLR	0.956	0.959	0.954	0.954	0.961
	LLR	0.941	0.945	0.950	0.949	0.941
0.90	ELLR	0.909	0.915	0.906	0.910	0.922
	LLR	0.883	0.891	0.899	0.898	0.893

($\varepsilon = 0$)

				n=40		
$1-\gamma$		π=0.2	π=0.4	π=0.2	π=0.4	π=0.6
0.99	ELLR	0.989	0.995	0.992	0.993	0.993
	LLR	0.983	0.974	0.981	0.963	0.925
0.95	ELLR	0.954	0.967	0.957	0.956	0.962
	LLR	0.932	0.916	0.924	0.886	0.824
0.90	ELLR	0.910	0.924	0.905	0.913	0.928
	LLR	0.875	0.859	0.859	0.808	0.751

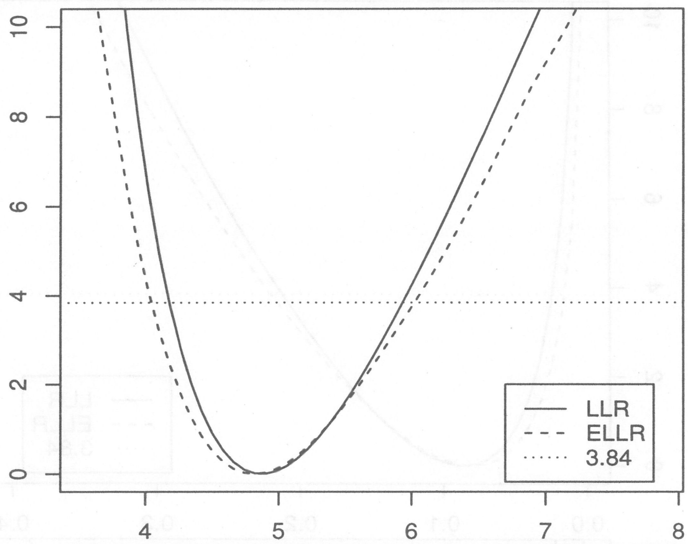
($\varepsilon = 0.05$)

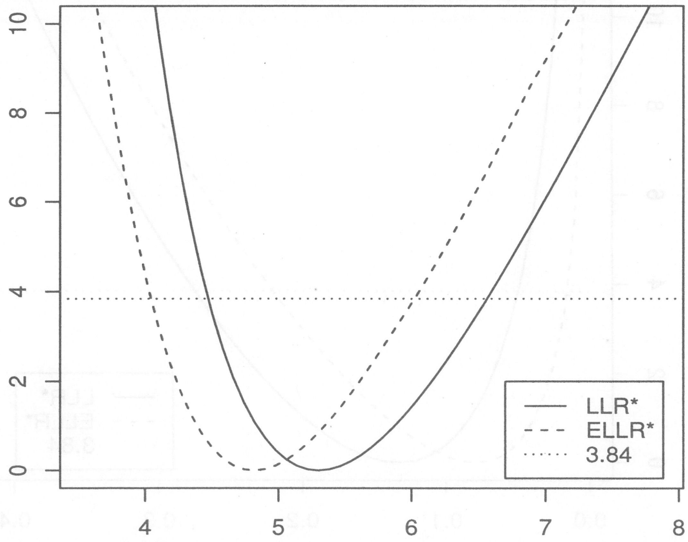
Figure 1: Profile empirical log-likelihood ratio statistic (ELLR) and profile log-likelihood ratio statistic (LLR) for the scale parameter σ of the distribution of the log-lifetimes. Gastric cancer data.



(*) the largest complete log-lifetime is changed from 7.22 to 10.

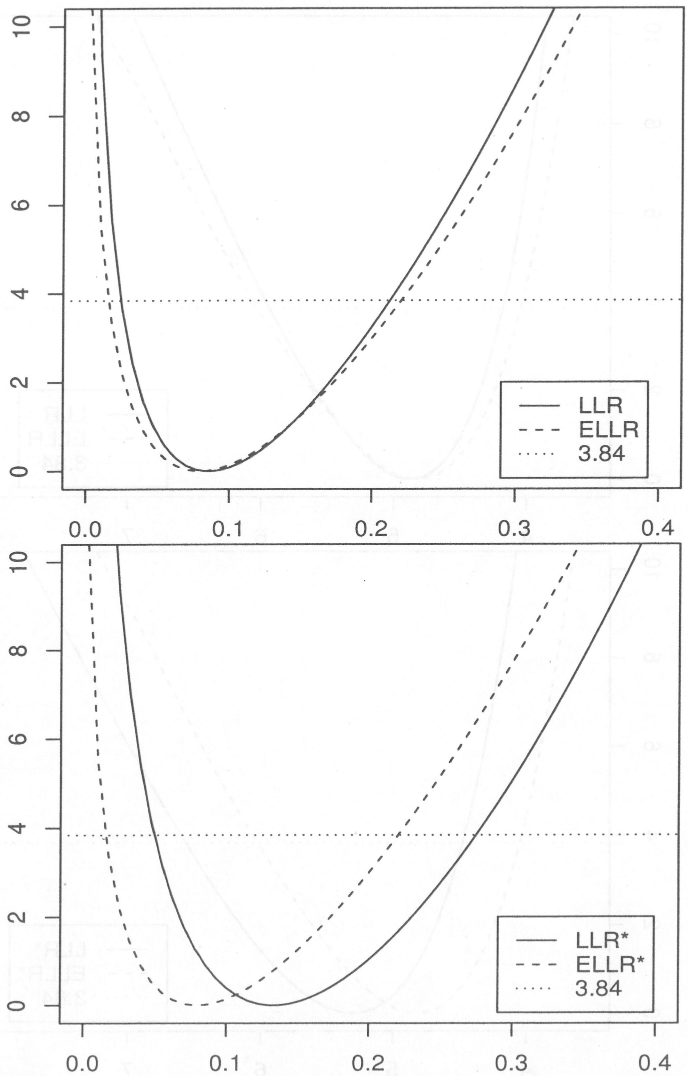
Figure 2: Profile empirical log-likelihood ratio statistic (ELLR) and profile log-likelihood ratio statistic (LLR) for the ninth decile of the distribution of the log-lifetimes. Autologous marrow trasplant data.





(*) the smallest complete log-lifetime is changed from −0.4185 to −3.

Figure 3: Profile empirical log-likelihood ratio statistic (ELLR) and profile log-likelihood ratio statistic (LLR) for the survival function of the distribution of the log-lifetimes, $S(y)$, evaluated at $y = 5$. Autologous marrow trasplant data.



(*) the smallest complete log-lifetime is changed from -0.4185 to -3.

References

Adimari, G. (1997), On the empirical likelihood ratio for smooth functions of M-functionals, *Scand. J. Statist.*, **24**, 47-59.

Adimari, G., Ventura, L. (2002), Quasi-likelihood from M-estimators: a numerical comparison with empirical likelihood, *Statist. Meth. App.*, **11**, 175-185.

Akritas, M.G., Basak, I., Lee, M.H., (1993), Robust estimation under type II censoring, *Austral. J. Stastist.*, **35**, 303-318.

Basak, I. (1993), A note on James' type robust estimators, *Comm. Statist. - Theory Meth.*, **22**, 1069-1084.

Bhattachryya, G.K., (1985), The asymptotics of maximum likelihood and related estimators based on type II censored data, *J. Amer. Statist. Assoc.*, **80**, 398-404.

Hall, P., La Scala, B. (1990), Methodology and algorithms of empirical likelihood, *Internat. Statist. Rev.*, **58**, 109-127.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A. (1986), *Robust Statistics. The Approach Based on Influence Functions*, J.Wiley, New York.

Huber, P.J. (1964), Robust estimation of a location parameter, *Ann. Math. Stat.*, **35**, 73-101.

Huber,P.J. (1981), *Robust Statistics*, J.Wiley, New York.

James, I.R. (1986), On estimating equations with censored data, *Biometrika*, **73**, 35-42.

Kalbfleisch, J.D., Prentice, R.L. (1980), *The statistical analysis of failure time data*, J.Wiley, New York.

Klein, J.P., Moeschberger, M.L. (1997), *Survival analysis - Techniques for censored and truncated data*, Springer, New York.

Masarotto, G., Peracchi, M. (1991), Stimatori di tipo M per dati censurati con una applicazione alla stima robusta del parametro di una distribuzione esponenziale, *Statistica*, **4**, 505-520.

Owen, A.B. (1988), Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, **75**, 237-49.

Owen, A.B. (1990), Empirical likelihood ratio confidence regions, *Ann. Statist.*, **18**, 90-120.

Owen, A.B. (2001), *Empirical likelihood*, Chapman and Hall, London.

Qin, J., Lawless, J. (1994), Empirical likelihood and general estimating equations, *Ann. Statist.*, **23**, 300-325.

Qin, J., Lawless, J. (1995), Estimating equations, empirical likelihood and constraints on parameters, *Canad. J. Statist.*, **23**, 145-159.

Stablein, D.M., Carter, W.H., Novak, J.W. (1981), Analysis of survival data with non-proportional hazard functions, *Control. Clin. Trials*, **2**, 149-159.

Tsao, M., Zhou, J. (2001), On the robustness of empirical likelihood ratio confidence intervals for location, *Canad. J. Statist.*, **29**, 129-140.