Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 137 (2007) 2832–2849
journal of statistical planning and inference
www.elsevier.com/locate/jspi

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

Gianfranco Adimari*, Paolo Preo

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

Received 30 June 2004; received in revised form 26 January 2006; accepted 17 October 2006
Available online 21 December 2006

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, type II or randomly censored data. It is illustrated by considering inference for log-location-scale models. In particular, we focus on the log-normal and the Weibull models and we tackle 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.

Keywords: Censoring; Empirical likelihood; Estimating function; *M*-estimator; Profile likelihood; Robustness

1. Introduction

Let $y_1, \dots, y_n$ 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 $F(y;\theta) = \mathrm{pr}\{Y \leqslant y;\ \theta\}$ denotes the cumulative distribution function (c.d.f.) which depends on an unknown parameter θ belonging to some set $\Theta \subseteq \mathbb{R}^s$, $s \geqslant 1$. 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 $E_\theta\{\alpha(Y;\theta)\} = \int \alpha(y;\theta) f(y;\theta)\,\mathrm{d}y = 0$, for every $\theta \in \Theta$. Examples include maximum likelihood estimators, 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 typically not robust since it may be very sensitive to outliers and small deviations of the data distribution from the assumed model. For instance,

* Corresponding author. Tel.: +39 4982 74173; fax: +39 4982 74170.
E-mail address: gianfranco.adimari@unipd.it (G. Adimari).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2006.10.008

the level of the gross-error sensitivity (Hampel et al., 1986, Section 4.2), which specifies the extent to which a single outlier can influence an estimator, is generally unbounded for the maximum likelihood estimator.

Robust 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 to yield 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.

In lifetime data analysis, lifetimes $y_1, \ldots, y_n$ are typically 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 $\delta_i = I(y_i \leqslant t_i)$, with $I(\cdot)$ being 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, m $(< n)$ being a fixed integer and $y_{(m)}$ the mth 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

$$\sum_{i=1}^{n}\{\delta_i \varDelta(z_i;\theta) + (1-\delta_i)\varDelta_c(z_i;\theta)\} = 0, \tag{2}$$

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

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

where $\eta(z_i,\delta_i;\theta)=\delta_i\alpha(z_i;\theta)+(1-\delta_i)\alpha_c(z_i;\theta)$ and $\alpha_c(z_i;\theta)=E_\theta\{\alpha(Y;\theta)|Y>z_i\}$. 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. Robustness properties of the 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 (Owen, 1988, 1990) and estimating functions that define James-type robust M-estimators. For censored data, 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 in the paper presents some desirable features. First, it is quite general and applies to several practical situations with type I, type II or randomly censored data. Second, unlike the normal approximation-based approach, it 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, having shape which is determined automatically by the data.

We illustrate the proposed method by considering inference for log-location-scale models. In particular, we focus on the log-normal and the Weibull models and we tackle 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