Computational Statistics and Data Analysis 55 (2011) 3093–3103

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Sensitivity analysis of predictive modeling for responses from the three-parameter Weibull model with a follow-up doubly censored sample of cancer patients

Hafiz M.R. Khan [a,*], Ahmed Albatineh [a], Saeed Alshahrani [a], Nadine Jenkins [b], Nasar U. Ahmed [a]

[a] Department of Epidemiology & Biostatistics, Robert Stempel College of Public Health and Social Work, Florida International University, Miami, FL 33199, USA
[b] UMDNJ University Hospital, University of Medicine & Dentistry of New Jersey, 150 Bergen Street, Newark, NJ 07103, USA

ARTICLE INFO

Article history:
Received 2 February 2011
Received in revised form 23 May 2011
Accepted 23 May 2011
Available online 30 May 2011

Keywords:
Censored sample
Doubly censored sample
Three-parameter Weibull model
Bayesian approach
Predictive inference

ABSTRACT

The purpose of this paper is to derive the predictive densities for future responses from the three-parameter Weibull model given a doubly censored sample. The predictive density for a single future response, bivariate future response, and a set of future responses has been derived when the shape parameter α is unknown. A real data example representing 44 patients who were diagnosed with laryngeal cancer (2000–2007) at a local hospital is used to illustrate the predictive results for the four stages of cancer. The survival days of eight out of the 44 patients could not be calculated as the patients were lost to follow-up. They were the first four and the last four patients' survival days in order. Thus, the recorded data for the survival days of 36 patients composed of 18 male and 18 female patients with cancer of the larynx are used for the predictive analysis. Furthermore, a subgroup level of the male and female patients follow-up data are considered to obtain the future survival days. A sensitivity study of the mean, standard deviation, and 95% highest predictive density (HPD) interval of the future survival days with respect to stages and doses are performed when the shape parameter α is unknown.

1. Introduction

Predictive inference has been playing an important role in healthcare data analysis. Healthcare researchers rely on past observations to analyze and forecast treatment outcomes. Usually, observations recorded in the past can be incomplete for any reason such as lost to follow-up due to migration, early termination of the study, unrelated cause of death, limited resources to conduct a study, or withdrawal from a study due to undesirable effect of a drug. When a data set is incomplete, the healthcare practitioners or researchers need to carefully deal with them, otherwise any underestimate or overestimate of some statistical measures may lead to misinterpretation of the data. There are three types of censored data such as left, right, and doubly censored. Recently several statistical research works have been accomplished based on censored data, for examples, Ahmed and Saleh (1999) and Buhamra et al. (2004, 2007). Doubly censored data are commonly observed in clinical studies, where the first few observations and the last few observations are unavailable from a sequence of observations. For example, there are n patients who are diagnosed with a particular disease in a clinical study. These patients are given optimal doses for a period of time. During the follow-up, the survival time of each patient is recorded in order with $t_1 \leq \cdots \leq t_n$. In this case, the record of survival days of the first few observations and the last few observations may not be available.

* Corresponding author.
E-mail address: hmkhan@fiu.edu (H.M.R. Khan).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2011.05.017

Therefore, they might be removed from the statistical analysis in order to make statistical inferences. This type of data constitutes a doubly censored sample. For more about doubly censored samples, the reader is referred to Sarhan (1955), Khan (2003), Khan et al. (2006), Kambo (1978), Raqab (1995) and Lalitha and Mishra (1996), among others. Doubly censored data from clinical studies may be modeled by the three-parameter Weibull model. Based on a doubly censored sample, it is assumed in this paper that the three-parameter Weibull model specified by the following density is appropriate:

$$f(t|\alpha,\beta,\mu) = \begin{cases} \dfrac{\alpha}{\beta}\left(\dfrac{t-\mu}{\beta}\right)^{\alpha-1} \exp\left\{-\left(\dfrac{t-\mu}{\beta}\right)^{\alpha}\right\}, & \text{for } t \ge \mu;\ \alpha,\beta > 0, \\ 0 & \text{elsewhere,} \end{cases} \tag{1}$$

where μ is the location parameter, α is the shape parameter, and β is the scale parameter. This model is an extension of the two-parameter Weibull model. The additional parameter is the location parameter. For analytical convenience, on writing $\theta = \beta^{\alpha}$, we have

$$f(t|\alpha,\theta,\mu) = \begin{cases} \dfrac{\alpha}{\theta}(t-\mu)^{\alpha-1} \exp\left\{-\dfrac{(t-\mu)^{\alpha}}{\theta}\right\}, & \text{for } t \ge \mu;\ \alpha > 0, \theta > 0, \\ 0 & \text{elsewhere.} \end{cases} \tag{2}$$

Ahmed (1992) discussed an asymptotic estimation of reliability in a life-testing model. Sarhan (1955) obtained the mean and standard deviation of certain populations from singly and doubly censored samples. Kambo (1978) derived the maximum likelihood estimators of the location and scale parameters by using a doubly censored sample. Khan et al. (2006) derived the predictive distributions of future responses on the basis of a doubly censored sample from the two-parameter exponential life testing model. There are several studies on lifetime models related to censoring, for example, Ahmed et al. (2008), Ahmed and Saleh (1999) and Ahsanullah and Ahmed (2001).

The inference about the future responses given a set of observed data is known as predictive inference. In predictive inference, the observed data can be considered as an informative experiment, and the unobserved future data form the future experiment. The goal of predictive inference is to obtain probability statements about the future experiment given the informative experiment. Predictive inference has been considered by many authors, for instance, Mahdi et al. (1998), Thabane (1998), Khan et al. (2006), Bernardo and Smith (1994), Gelman et al. (2004), and Martz and Waller (1982). One may consider the Bayesian approach to derive the predictive inference of future observations. Several authors have utilized the Bayesian approach for determining predictive inference. For example, in the past, Aitchison and Dunsmore (1975), Aitchison and Sculthorpe (1965), and Berger (1985) considered a general Bayesian predictive problem. Geisser (1993) discussed various Bayesian predictive problems for future responses. Moreover, Geisser (1971) discussed the inferential use of predictive distributions. Predictive distributions for linear models which make use of the Bayesian approach have been determined by Dunsmore (1974), and Zellner and Chetty (1965). Additional applications of the Bayesian approach to predictive inference have been considered for example by Khan (2003), Thabane and Haq (2000), Thabane (1998), Khan and Provost (2008), Khan et al. (2007), Gelman et al. (2004), Bernardo and Smith (1994), Sinha (1986), Evans and Nigm (1980a,b), and Martz and Waller (1982).

The objective of this paper is to derive the predictive densities for future responses from the three-parameter Weibull distribution given a doubly censored sample with unknown shape parameter (α). The results will be illustrated by making use of the Bayesian approach with application to larynx cancer patients.

The remainder of the paper is organized as follows: In the case of unknown shape parameter (α), Section 2 discusses the derivation of the likelihood function, the prior density function, the posterior density function, the predictive density function for a single future response, a bivariate future response, and a set of future responses. Section 3 discusses the highest predictive density (HPD) interval. Section 4 describes the estimation of the hyperparameters of prior density. Section 5 includes a real data example of larynx cancer patients to illustrate some of the results.

2. Posterior density

Let $t_1, \ldots, t_n$ be an ordered random sample of size n from model (2), where $t_1 \le \cdots \le t_g$ be the g smallest ordered observations and $t_{\ell+1} \le \cdots \le t_n$ be the $(n-\ell)$ largest ordered observations from the sample. Only the remaining ordered observations $\mathbf{t} = (t_{g+1}, \ldots, t_\ell)$ are used for statistical analysis. It is assumed that the sample data are modeled by the three-parameter Weibull model. The likelihood function of θ, α, and μ for a given doubly censored sample $\mathbf{t} = (t_{g+1}, \ldots, t_\ell)$ is given by

$$\mathrm{L}(\theta,\alpha,\mu|\mathbf{t}) \propto \left[F\left\{\frac{(t_{g+1}-\mu)^{\alpha}}{\theta}\right\}\right]^{g}\left[1-F\left\{\frac{(t_\ell-\mu)^{\alpha}}{\theta}\right\}\right]^{n-\ell}\left[\prod_{i=g+1}^{\ell} f\left\{\frac{(t_i-\mu)^{\alpha}}{\theta}\right\}\right], \quad \text{for } t_i \ge 0;\ \theta \ge 0,$$

where $F\{\frac{(t_{g+1}-\mu)^{\alpha}}{\theta}\} = 1 - \exp\{-\frac{(t_{g+1}-\mu)^{\alpha}}{\theta}\}$.