Computational Statistics and Data Analysis 75 (2014) 96–111

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Lower confidence limit for reliability based on grouped data using a quantile-filling algorithm

Mimi Zhang [a,*], Qingpei Hu [b], Min Xie [a], Dan Yu [b]

[a] *Department of Systems Engineering and Engineering Management, City University of Hong Kong, Hong Kong, China*

[b] *Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China*

ARTICLE INFO

Article history:
Received 5 January 2012
Received in revised form 10 January 2014
Accepted 17 January 2014
Available online 27 January 2014

Keywords:
Data completion
Expectation–maximization algorithm
Incomplete data
Interval estimate
Method of moments

ABSTRACT

The aim of this paper is to propose an approach to constructing lower confidence limits for a reliability function and investigate the effect of a sampling scheme on the performance of the proposed approach. This is accomplished by using a data-completion algorithm and certain Monte Carlo methods. The data-completion algorithm fills in censored observations with pseudo-complete data while the Monte Carlo methods simulate observations for complicated pivotal quantities. The Birnbaum–Saunders distribution, the lognormal distribution and the Weibull distribution are employed for illustrative purpose. The results of three cases of data-analysis are presented to validate the applicability and effectiveness of the proposed methods. The first case is illustrated through simulated data, and the last two cases are illustrated through two real-data sets.

1. Introduction

Censoring schemes and censored data are prevalent in reliability-testing experiments. Various types of censoring schemes, resulting in various types of censored data, have been studied in the literature such as Type-I censoring, Type-II censoring and progressive censoring. The type of censored data to be studied in the present paper is termed as grouped data. A grouped data set contains several groups. The lifetimes of the failed items in each of the groups are only known to be smaller than a certain time point; that is, the lifetimes of the items are only partially known. Grouped data are very common in, e.g., burn-in tests (Ye et al., 2011). In a burn-in test, a number of components are exercised for a period of time before being placed in service. At the end of the burn-in test, we obtain a group of failed components and survival components. We only know the numbers of failed components and survival components; we do not know the exact lifetimes of the failed components. By repeating the preceding procedure for several times with different burn-in durations, we obtain a grouped data set. Grouped data also occur extensively in all kinds of maintenance activities (Taghipour and Banjevic, 2011). In view of the prevalence, developing executable and competent data-analysing methods to handle grouped data is an issue of great interest. To this end, this paper proposes a data-completion algorithm which is termed as the modified quantile-filling (MQF) algorithm. The expectation–maximization (EM) algorithm is an iterative method for maximum likelihood (ML) estimation. The EM algorithm iteratively applies two steps: an expectation step and a maximization step. The expectation step involves taking expectation of a conditional distribution, and the maximization step involves ML estimation based on complete data. However, the EM algorithm is not effective in dealing with grouped data because all the observations in a grouped data set are censored. Each evaluation of the expectation step shall require a numerical approximation to a multiple integral,

* Correspondence to: Department of Systems Engineering and Engineering Management, City University of Hong Kong, Kowloon Tong, Hong Kong, China. Tel.: +852 3442 4604.

E-mail address: mmzhang5-c@my.cityu.edu.hk (M. Zhang).

0167-9473/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2014.01.010

which is burdensome and extremely difficult. A possible solution to overcoming this computational inefficiency is to invoke data-completion methods.

The MQF algorithm fills in partially observed data with plausible values, transforming the incomplete data into pseudo-complete data. It is believed that Rubin (1976) was the first to come up with the idea of data-completion, filling in a missing datum with several plausible values. Recent works on completing missing data can be found in Chen and Sun (2010), Daniel and Kenward (2012) and Hapfelmeier et al. (2012). Grouped data are different from missing data. Grouped data provide incomplete information about the unobserved exact lifetimes, whereas missing data provide no information. Nevertheless, one can treat the unobserved exact lifetimes as missing and fill in them with appropriately simulated values. To estimate the parameters of a linear regression model, Buckley and James (1979) adopted a data-completion technique and replaced censored observations with estimated conditional expectations. Lai and Ying (1991) proved that, under certain conditions, a modified Buckley–James estimator is consistent and asymptotically normal. Wang et al. (2008) related high-dimensional genomic data to survival outcomes by using a doubly penalized Buckley–James method. Andersen and Perme (2010) presented a review of recent work on the application of pseudo observations in survival analysis. Yu and Dai (1996) and Yu and Guo (2001) proposed a quantile-filling algorithm which fills in incomplete data with conditional quantiles and estimates parameters by using the ML method. Yet, the convergence and consistency of parameter estimates have not been proved. The MQF algorithm advanced in this paper fills in incomplete data with conditional quantiles and estimates unknown parameters by the method of moments. When dealing with right-censored data, Jiang (2008) proposed a quantile-filling algorithm by using moment estimation. The convergence and consistency of parameter estimates were verified in that paper. Since a grouped data set is composed of censored lifetime data, invoking the MQF algorithm to deal with grouped data is not only reasonable but also efficient.

Having obtained the parameter estimates by the MQF algorithm, we then construct lower confidence limits for the underlying reliability function. With respect to the performance, we usually require an assurance regarding the minimum value of an index. A point estimator will make little sense if the variance of the estimator is high. Consequently, it is necessary to give a lower limit, such that the value of the index will be larger than the lower limit with certain probability. Heard and Pensky (2006) considered the construction of confidence intervals for reliability when the sample size is relatively small. Yang et al. (2007) proposed a general method for constructing analytically adjusted naive confidence intervals for the Weibull model. When dealing with stress–strength models, Barbiero (2011) constructed confidence intervals for reliability through a parametric bootstrap procedure. In this paper, by means of the Monte Carlo simulation, we propose a numerical method to construct lower confidence limits.

There are three main contributions of this paper. The first contribution is the development of a data-completion algorithm by utilizing quantiles and the method of moments. The second contribution is the calculation of a lower confidence limit for reliability (LCLR) based on a pseudo-complete data set. The third contribution is the assessment of the influence of sampling schemes on the performance of the proposed methods.

The remainder of this paper is organized as follows. In Section 2, we present a general framework of grouped data and algorithms for deriving lower confidence limits. In Section 3, we specify the algorithms via three widely used lifetime distributions. In Section 4, we provide three illustrative examples: a simulation study, an application to aluminium coupons and an application to ball bearings.

2. Analyse grouped data via the MQF algorithm: a general framework

In this section we develop a generic framework for analysing grouped data by using the MQF algorithm. The procedure for calculating an LCLR is illustrated via a heuristic example. We detail the prescriptive structure via three widely used lifetime distributions in the next section.

Consider an experiment in which a population of statistically identical and independent items is placed on a life-test. At the first sampling epoch t_1, n_1 items are randomly drawn without replacement from the population. The number of failed items in the first group is recorded, denoted by f_1. At the second sampling epoch t_2, n_2 items are randomly drawn without replacement from the population; the number of failed items in the second group is recorded, denoted by f_2. Repeating the sampling procedure for k times, we obtain a grouped data set, denoted by $\{(t_i, n_i, f_i),\ i = 1, 2, \ldots, k\}$. Here, t_i is the ith sampling epoch. f_i is the number of failed items in the ith group. n_i, referred to as group size, is the size of the ith group. k is referred to as sampling number. In real-life practice, the sampling number, the sampling epochs and the group sizes are pre-determined by domain experts to balance between various technical and practical requirements.

Assume that the survival times of the items in the population are independent and identically distributed with distribution function $F(t;\ \boldsymbol{\theta})$ in which $\boldsymbol{\theta}$ is a vector of unknown parameters. Define T to be a random variable having the distribution function $F(t;\ \boldsymbol{\theta})$. The MQF algorithm involves two steps: the quantile-filling (QF) step and the moment-estimation (ME) step. The QF step fills in the partially known lifetimes with conditional quantiles, transforming the incomplete grouped data $\{(t_i, n_i, f_i),\ i = 1, 2, \ldots, k\}$ into pseudo-complete data. The ME step estimates the unknown parameters using the method of moments. After m, $m \in \mathbb{Z}^+$, iterations of the MQF algorithm, let $\boldsymbol{t}_i^{(m)} = \left(t_{i,1}^{(m)}, \ldots, t_{i,f_i}^{(m)}, t_{i,f_i+1}^{(m)}, \ldots, t_{i,n_i}^{(m)}\right)$ denote the imputed pseudo-complete lifetime data of the items in the ith group. Let $\hat{\boldsymbol{\theta}}^{(m)}$ denote a moment estimate on the unknown parameter vector $\boldsymbol{\theta}$, calculated based on the pseudo-complete data $\boldsymbol{t}^{(m)} = \left(\boldsymbol{t}_1^{(m)}, \boldsymbol{t}_2^{(m)}, \ldots, \boldsymbol{t}_k^{(m)}\right)$.