



Contents lists available at ScienceDirect

# Computational Statistics and Data Analysis

journal homepage: [www.elsevier.com/locate/csda](http://www.elsevier.com/locate/csda)

## Selection between Weibull and lognormal distributions: A comparative simulation study

Jin Seon Kim, Bong-Jin Yum\*

Department of Industrial Engineering, Korea Advanced Institute of Science and Technology, 373-1, Guseong-Dong, Yuseong-Gu, Daejeon, 305-701, Republic of Korea

### ARTICLE INFO

#### Article history:

Received 15 April 2008  
 Received in revised form 7 August 2008  
 Accepted 15 August 2008  
 Available online 20 August 2008

### ABSTRACT

How to select the correct distribution for a given set of data is an important issue, especially when the tail probabilities are of interest as in lifetime data analysis. The Weibull and lognormal distributions are assumed most often in analyzing lifetime data, and in many cases, they are competing with each other. In addition, lifetime data are usually censored due to the constraint on the amount of testing time. A literature review reveals that little attention has been paid to the selection problems for the case of censored samples. In this article, relative performances of the two selection procedures, namely, the maximized likelihood and scale invariant procedures are compared for selecting between the Weibull and lognormal distributions for the cases of not only complete but also censored samples. Monte Carlo simulation experiments are conducted for various combinations of the censoring rate and sample size, and the performance of each procedure is evaluated in terms of the probability of correct selection (PCS) and average error rate. Then, previously unknown behaviors and relative performances of the two procedures are summarized. Computational results suggest that the maximized likelihood procedure can be generally recommended for censored as well as complete sample cases.

© 2008 Elsevier B.V. All rights reserved.

### 1. Introduction

Choosing the correct or best-fitting distribution for a given set of data is an important issue, especially when the tail probability, which is usually sensitive to the assumed model, is of interest as in reliability engineering. Statistical methods for distribution choice include probability plotting (Nelson, 1982), goodness-of-fit (GOF) testing, hypothesis testing (HT), and selection procedures. Probability plotting provides useful information concerning the right distribution. However, it can be subjective and may yield multiple plots which appear to be equally adequate. Concerning the usual GOF and HT approaches, one may prefer the latter when one may not wish to reject the distribution in the null hypothesis without an alternative to take its place. However, the HT approach treats the two distributions asymmetrically, while selection procedures put candidate distributions on equal footings and can deal with more than two distributions at the same time.

Many authors developed selection procedures with respective selection statistics and decision rules. The two well-known procedures are those that are respectively based on the maximized likelihood function (MLF) and scale invariant (SI) selection statistic. For the MLF-based procedure, the reader is referred to Bain and Engelhardt (1980), Kappenman (1982), Gupta and Kundu (2003, 2004), Kundu and Manglick (2004), Kundu et al. (2005), Strupczewski et al. (2006), and Kundu and Raqab (2007), among others. The SI selection procedure was developed in Quesenberry and Kent (1982) for selecting among the exponential, Weibull, lognormal, and gamma distributions. In both procedures, the distribution with the largest value of the selection statistic is selected. For other selection procedures, the reader is referred to Croes et al. (1998), Marshall et al.

\* Corresponding author. Tel.: +82 42 869 3116; fax: +82 42 869 3110.  
 E-mail address: [bjyum@kaist.ac.kr](mailto:bjyum@kaist.ac.kr) (B.-J. Yum).

(2001), Cain (2002), Dick (2004) and Mitosek et al. (2006), among others. In addition, Bayesian selection procedures were developed by Kim et al. (2000), Upadhyay and Peshwani (2003), Araújo and Pereira (2007), etc.

Several authors compared the relative performance of selection procedures. Siswadi and Quesenberry (1982) compared the SI, scale-shape invariant (SSI), and MLF procedures when complete data are available and the SI and MLF procedures when the data are Type-I censored. For other comparative studies, the reader is referred to Kappenman (1989), Pandey et al. (1991), Lee and Pope (2006), Mitosek et al. (2006), and Basu et al. (2008).

A review of the related literature reveals that little attention has been paid to the selection problems for the case of censoring. Exceptions include Siswadi and Quesenberry (1982) for Type-I censored samples, Croes et al. (1998), Cain (2002) and Kim et al. (2000) for Type-II censored samples, and Block and Leemis (2008) for randomly right-censored samples.

In this paper, the ratio of the maximized likelihoods (RML) and SI procedures are compared for discriminating between the Weibull and lognormal distributions for the cases of complete, Type-I censored, and Type-II censored samples (The RML procedure is equivalent to the MLF procedure in the case of two distributions). The Weibull and lognormal distributions are most often assumed and competing with each other in analyzing lifetime data (e.g., see Bain (1978), Chen et al. (1987) and Prendergast et al. (2005) for Time Dependent Dielectric Breakdown (TDDB) data, and Lloyd (1979) and Pinto (1991) for electromigration lifetime data). In such cases, it is highly desirable to have a means to discriminating between the two distributions since there is a significant difference in the low percentiles which are of interest in lifetime data analysis. In addition, lifetime data are usually censored due to the constraint on the amount of testing time, which necessitates an extensive comparative study of selection procedures for the case of censoring. Siswadi and Quesenberry (1982) compared the SI and MLF procedures for Type-I censored samples, but only considered the case where the sample size is 30 and the expected censoring rate is 10%. As can be seen in the simulation results in Section 3, considering some specific cases only could be misleading in understanding the whole behaviors of the two procedures. In this paper, the performances of the RML and SI procedures are evaluated in terms of the probability of correct selection (PCS) for various simulated data sets with different sample sizes and censoring rates.

The rest of this article is organized as follows. In Section 2, the two procedures are described in detail, and Section 3 shows the comparison results obtained from the simulation experiments. Finally, conclusions and guidelines are presented in Section 4.

## 2. RML and SI procedures for discriminating between Weibull and lognormal distributions

The probability density functions (pdfs) of the Weibull and lognormal distributions are respectively given as follows.

$$W(\eta, \beta) : f(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\eta}\right)^\beta\right]$$

where  $\eta > 0$  and  $\beta > 0$  are the scale and shape parameters, respectively, and

$$\text{LN}(\theta, \sigma) : g(t) = \frac{1}{\sqrt{2\pi}\sigma t} \exp\left\{-\frac{[\ln(t/\theta)]^2}{2\sigma^2}\right\}$$

where  $\theta > 0$  and  $\sigma > 0$  are the scale and shape parameters, respectively.

### 2.1. Procedure RML: Maximized likelihood function approach

For the case of complete data, the logarithm of the MLF for the Weibull distribution (i.e.,  $\ln L_W$ ) and that for the lognormal distribution (i.e.,  $\ln L_{\text{LN}}$ ) are given as in Eqs. (1) and (2), respectively.

$$\ln L_W = n \ln \hat{\beta} - n \hat{\beta} \ln \hat{\eta} + (\hat{\beta} - 1) \sum_{i=1}^n \ln t_i - \sum_{i=1}^n (t_i/\hat{\eta})^{\hat{\beta}} \quad (1)$$

$$\ln L_{\text{LN}} = -n \ln \hat{\sigma} - \frac{n}{2} \ln 2\pi - \sum_{i=1}^n \ln t_i - \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^n (\ln t_i - \ln \hat{\theta})^2 \quad (2)$$

where  $\hat{\eta}$ ,  $\hat{\beta}$ ,  $\hat{\theta}$  and  $\hat{\sigma}$  are the maximum likelihood estimators (MLEs) of  $\eta$ ,  $\beta$ ,  $\theta$  and  $\sigma$ , respectively.

For the case of Type-I censoring, let  $t_1, t_2, \dots, t_r$  be the ordered failure times observed until the censoring time  $t_c$ . Then, the logarithm of the MLF for each distribution can be derived as follows except for the common term,  $\ln [n!/(n-r)!]$  (Siswadi and Quesenberry, 1982).

$$\ln L_W = r \ln \hat{\beta} - r \hat{\beta} \ln \hat{\eta} + (\hat{\beta} - 1) \sum_{i=1}^r \ln t_i - \sum_{i=1}^r (t_i/\hat{\eta})^{\hat{\beta}} - (n-r) (t_c/\hat{\eta})^{\hat{\beta}} \quad (3)$$

$$\ln L_{\text{LN}} = -r \ln \hat{\sigma} - \frac{r}{2} \ln 2\pi - \sum_{i=1}^r \ln t_i - \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^r (\ln t_i - \ln \hat{\theta})^2 + (n-r) \ln \left[1 - \Phi\left(\frac{\ln t_c - \ln \hat{\theta}}{\hat{\sigma}}\right)\right] \quad (4)$$

where  $\Phi$  is the cumulative distribution function (cdf) of the standard normal distribution.

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات