First Seminar on
Reliability Theory and its Applications
27-28, May 2015

Estimation of $P(X > Y)$ Using Imprecise Data in the Lindley Distribution

Pak, A. [1]

Department of Computer Sciences, Shahrekord University

Abstract

Classical estimation procedures of the stress-strength parameter $R = Pr(X > Y)$ are based on precise data. However, in real world situations, some collected data might be imprecise and are represented in the form of fuzzy numbers. In this paper, we obtain the maximum likelihood estimation of the parameter R when X and Y are independent Lindley random variables, and the available data are reported in the form of fuzzy numbers. A Monte Carlo simulation study is carried out in order to assess the accuracy of the proposed method.

Keywords: Stress-Strength model, Fuzzy data analysis, Maximum likelihood estimation.

1 Introduction

Extensive research has been conducted on the stressstrength model. This model involves two independent random variables X and Y, and the parameter of interest is the probability $R = P(X > Y)$. A comprehensive account of this topic is given by Kotz et al. (2003). The developments in this field covered a variety of data types including complete data, censored data as well as data with explanatory variables. However, in real world situations, the results of an experimental performance can not always be recorded or measured precisely, but each observable event may only be identified with a fuzzy subset of the sample space. Our aim in this paper is to develop an inferential procedure for the stress-strength model in the situation where the stress measurements and the strength measurements are both in terms of fuzzy numbers. We will construct maximum likelihood estimation for the stress-strength reliability assuming two independent samples from Lindley distribution. In Section 2, We first introduce a generalized likelihood function based

[1] abbas.pak1982@gmail.com

on fuzzy data and then discuss the maximum likelihood estimation of the parameter R. A Monte Carlo simulation study is presented in Section 3, in order to assess the accuracy of the proposed method. For a review about the main definitions of fuzzy sets see Pak et al. (2014) and the references therein.

We use the following notation. A Lindley distribution with the parameter θ, will be denoted by $Lindley(\theta)$ and the corresponding probability density function is as follows;

$$f(x;\theta) = \frac{\theta^2}{1+\theta}(1+x)e^{-\theta x}; \qquad x>0;\ \theta>0. \tag{1}$$

2 Maximum likelihood estimation

Let the strength X and stress Y follow $Lindley(\theta_1)$ and $Lindley(\theta_2)$, respectively, and they are independent. Then, it can be easily shown that

$$\begin{aligned} R &= Pr(Y<X) \\ &= 1 - \frac{\theta_1^2[\theta_1^2(\theta_1+1)+\theta_2(\theta_1+1)(\theta_1+3)+\theta_2^2(2\theta_2+3)+\theta_2^2]}{(\theta_1+1)(\theta_2+1)(\theta_1+\theta_2)^3}. \end{aligned} \tag{2}$$

Suppose that partial information about the stress and strength are available in the form of fuzzy numbers $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{y}}$ with the Borel measurable membership functions $\mu_{\tilde{\boldsymbol{x}}}(\boldsymbol{x})$ and $\mu_{\tilde{\boldsymbol{y}}}(\boldsymbol{y})$. Then, the corresponding observed-data log likelihood function can be obtained as:

$$\begin{aligned} L_O(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}};\theta_1,\theta_2) &= n\log\left(\frac{\theta_1^2}{1+\theta_1}\right) + \sum_{i=1}^{n}\log\int(1+x)e^{-\theta_1 x}\mu_{\tilde{x}_i}(x)dx \\ &+ m\log\left(\frac{\theta_2^2}{1+\theta_2}\right) + \sum_{j=1}^{m}\log\int(1+y)e^{-\theta_2 y}\mu_{\tilde{y}_j}(y)dy. \end{aligned}$$

To compute the maximum likelihood estimate (MLE) of R, we need to compute the MLEs of θ_1 and θ_2, say $\hat{\theta}_1$ and $\hat{\theta}_2$, respectively. The MLE $\hat{R}$ of R can then be obtained by substituting $\hat{\theta}_k$ in place of θ_k, in (2.1) for $k=1$ and 2.
Since the observed fuzzy data $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{y}}$ can be viewed as incomplete specifications of the complete data vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively, the EM algorithm is applicable to obtain the MLEs of the unknown parameters.

To perform the E-step of the algorithm, we need to compute the conditional expectation of the complete-data log-likelihood function conditionally on the observed data $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{y}}$ as follows:

$$\begin{aligned} & n\log\left(\frac{\theta_1^2}{1+\theta_1}\right) + m\log\left(\frac{\theta_2^2}{1+\theta_2}\right) \\ - & \ \theta_1\sum_{i=1}^{n}E_{\theta_1^{(h)}}(X_i \mid \tilde{x}_i) - \theta_2\sum_{j=1}^{m}E_{\theta_2^{(h)}}(Y_j \mid \tilde{y}_j) \end{aligned} \tag{3}$$

where

$$E_{\theta_1^{(h)}}(X_i \mid \tilde{x}_i) = \frac{\int x(1+x)e^{-\theta_1^{(h)}x}\mu_{\tilde{x}_i}(x)dx}{\int(1+x)e^{-\theta_1^{(h)}x}\mu_{\tilde{x}_i}(x)dx}, \qquad i=1,...,n,$$

$$E_{\theta_2^{(h)}}(Y_j \mid \tilde{y}_j) = \frac{\int y(1+y)e^{-\theta_2^{(h)}y}\mu_{\tilde{y}_j}(y)dy}{\int (1+y)e^{-\theta_2^{(h)}y}\mu_{\tilde{y}_j}(y)dy}, \quad j = 1, ..., m.$$

The M-step of the algorithm involves maximizing (2.2) with respect to θ_1 and θ_2, which yields

$$\begin{aligned} \theta_1^{(h+1)} &= \frac{1}{2}(\alpha_h - 1) + \left[(1-\alpha_h)^2 + 8\alpha_h\right], \\ \theta_2^{(h+1)} &= \frac{1}{2}(\beta_h - 1) + \left[(1-\beta_h)^2 + 8\beta_h\right] \end{aligned}$$

where

$$\alpha_h = \frac{n}{\sum_{i=1}^{n} E_{\theta_1^{(h)}}(X_i \mid \tilde{x}_i)}, \qquad \beta_h = \frac{m}{\sum_{j=1}^{m} E_{\theta_2^{(h)}}(Y_j \mid \tilde{y}_j)}.$$

The MLEs of θ_1 and θ_2 can be obtained by repeating the E-step and M-step until convergence occurs.

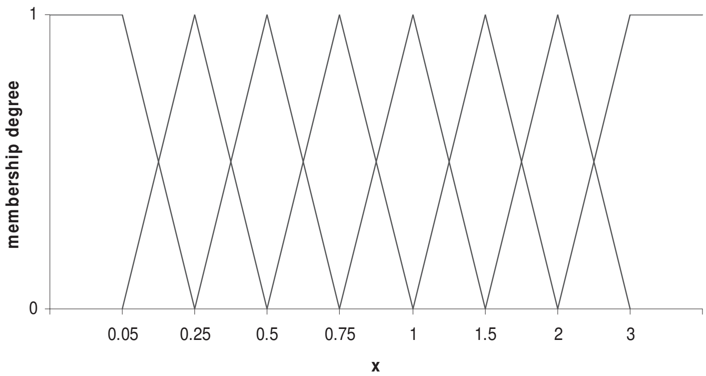

Figure 1: Fuzzy information system used to encode the simulated data

3 Simulation study

In order to assess the accuracy of the proposed method, we have carried out a Monte Carlo simulation study. First, for different sample sizes and a set of parameter values, namely $(\theta_1, \theta_2) = (1.0, 2.0)$, we have generated random samples from Lindley distribution. Then, each realization of the random samples was fuzzified using the fuzzy information system shown in Fig.1 and the estimate of the parameters θ_1, θ_2 and R for the fuzzy samples were computed using the maximum likelihood procedure. The average values (AV) and mean squared errors (MSE) of the ML estimates over 1000 replications are presented in Table 1. From the experiments, we found that the performance of the ML estimates are quite satisfactory and as the sample size increases, the MSEs of the estimates decrease as expected.

Table 1: AVs and MSEs of the ML estimates $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{R}$ for different sample sizes.

(n,m)	$\hat{\theta}_1$		$\hat{\theta}_2$		$\hat{R}$	
	AV	MSE	AV	MSE	AV	MSE
(20,20)	1.1931	0.0827	2.2180	0.1721	0.7025	0.0136
(20,30)	1.1827	0.0640	2.2039	0.1432	0.7003	0.0113
(30,20)	1.1838	0.0644	2.2057	0.1454	0.6891	0.0081
(30,30)	1.1210	0.0492	2.1863	0.1197	0.6678	0.0069
(30,50)	1.0731	0.0313	2.1625	0.0893	0.6653	0.0052
(50,30)	1.0690	0.0307	2.1597	0.0822	0.6672	0.0057
(50,50)	1.0248	0.0238	2.1139	0.0517	0.6319	0.0031

References

[1] S. Kotz, Y. Lumelskii, and M. Pensky, The Stress-Strength Model and its Generalizations: Theory and Applications. Singapore: World scientific, 2003.

[2] A., Pak, G.H., Parham, and M., Saraj, Inferences on the Competing Risk Reliability Problem for Exponential Distribution Based on Fuzzy Data. IEEE Transactions on Reliability, 63(1), 2014, 1-10.