ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 8, August 2013

DISCRIMINATING BETWEEN WEIBULL AND LOG-LOGISTIC DISTRIBUTIONS

Elsayed A. Elsherpieny[1], Sahar A. N. Ibrahim[2], Noha U. M. M. Radwan[3]

Professor, Department of Mathematical Statistics, Institute of Statistical Studies and Research, Cairo University, Egypt[1]

Lecture, Department of Mathematical Statistics, Institute of Statistical Studies and Research, Cairo University, Egypt[2]

Postgraduate Student, Department of Mathematical Statistics, Institute of Statistical Studies and Research, Cairo University, Egypt[3]

Abstract*:* Weibull and log-logistic distributions are two popular distributions for analyzing lifetime data. In this paper it is assumed that the data are coming either from Weibull or log-logistic distributions. The maximized likelihood ratio test to discriminate between the two distributions is used. The asymptotic distributions of the logarithm of the ratio of the maximized likelihood are obtained. These asymptotic results are used to estimate the probability of correct selection and the minimum sample size needed to discriminate between the two distributions. Two real data life are analyzed to see how the proposed method works in practice.

Keywords*:* Asymptotic distribution; Weibull distribution; Log-logistic distribution; Maximized likelihood ratio statistic; Probability of correct selection.

I. INTRODUCTION

Choosing the correct or best-fitting distribution for a given data set is an important issue. Most of the times distribution functions may provide a similar data fit but selecting the correct or more nearly correct model is desirable. The problem of choosing the correct model has been attempted by many researchers. Cox (1961) discussed the effect of choosing the wrong model. Cox (1962) tackled the problem of discriminating between the log-normal and the exponential distributions, based on the likelihood function, and derived the asymptotic distribution of the likelihood ratio statistic. Jackson (1968) derived asymptotic results for the case log-normal versus gamma. The case log-normal versus Weibull was addressed by Dumonceaux and Antle (1973). They proposed a certain test and provided its critical values. Pereira (1977) developed another two tests to discriminate between log-normal and Weibull distributions. Bain and Engelhardt (1980) covered the case Weibull versus gamma. Chen (1980) made significant contribution in discrimination problem when using small sample size. Kappenman (1982) studied the probability of correct selection for the pairs Weibull versus log-normal, Weibull versus gamma, and gamma versus log-normal. Firth (1988) discussed the problem of discriminating between the log-normal and gamma distributions. Fearn and Nebenzahl (1991) used the maximum likelihood ratio method in discriminating between the Weibull and gamma distributions. Wiens (1999) discussed the effect of choosing the wrong model through a real data example and by using log-normal and gamma models. Gupta and Kundu (2003) considered the likelihood ratio statistic for discriminating between Weibull and generalized exponential distributions. Gupta and Kundu (2004) discussed the problem of discriminating between the gamma and generalized exponential distributions by using maximized likelihood ratio test. Pascual (2005) discussed the effect of misspecification on the maximum likelihood estimates when discriminating between the log-normal and gamma distribution functions. Kundu and Manglick (2005) used the ratio of the maximized likelihoods in discriminating between log-normal and gamma distributions. Dey and Kundu (2009) considered the problem of discrimination among Weibull, log-normal and generalized exponential distributions. They used the maximized likelihood test to choose the best fitted model. Dey and Kundu (2010) used the maximized likelihood ratio test in the discrimination problem between log-normal and log-logistic distributions.

Some procedures for selecting between distributrions for data of not only complete but also censored have been paid attention by some authors. Siswadi and Quesenberry (1982), when selecting among Weibull, log-normal and gamma distributions, compared the scale invariant, scale shape invariant and maximized likelihood function tests for complete

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 8, August 2013

data and scale invariant and maximized likelihood function tests for Type-I censored data. Kim and Yum (2008) compared the ratio of maximized likelihoods and scale invariant tests for discriminating between the Weibull and log-normal distributions for complete, Type-I and Type–II censored data. Dey and Kundu (2012) considered the maximized likelihood ratio test in choosing between Weibull and log-normal distributions for type-II censored data.

Weibull and log-logistic distributions are two popular distributions for analyzing lifetime data. In this paper, the problem of discriminating between these two distribution functions is considered. A hypothesis testing method is used in which it is assumed that a data are coming either from Weibull or log-logistic distribution. The ratio of the maximized likelihood test is used to discriminate between them. The asymptotic distributions of the logarithm of the ratio of the maximized likelihood are obtained through two theorems. These asymptotic results are used to estimate the probability of correct selection, from which the minimum sample size needed to discriminate between the two distribution functions for a user specified probability of correct selection is obtained. Two real data life are analyzed to see how the proposed method works in practice.

Figures 1 and 3 show the diverse shape of the probability density function (p.d.f.) and cumulative distribution function (c.d.f.) respectively, of Weibull distribution at η=1 and β = 0.5, 1, 1.5, 2, 5. While Figures 2 and 4 show that of log-logistic distribution at ε =1 and σ = 0.5, 1, 2, 4, 8. From these figures, the closeness of the two p.d.f. and c.d.f. functions can be easily visualized. However some of the characteristics of Weibull and log-logistic distributions can be quite different. This can be shown when considering their hazard functions, given in Figures 5 and 6 respectively. Therefore, if the data are coming from any one of them, may be it is modeled by the other one. In addition if the sample size is not very large the problem of choosing the correct distribution becomes more difficult, but it is still very important to make the best decision based on the data at hand.

The rest of the paper is organized as follows. In Section 2, the test statistic is presented. In Section 3, the asymptotic distributions of the test statistic under null hypotheses is obtained. The minimum sample size needed to discriminate between Weibull and log-logistic distributions at a user specified protection level and tolerance level is determined In Section 4. Two real life data sets are analyzed in Section 5. Finally a conclusion is given in Section 6.

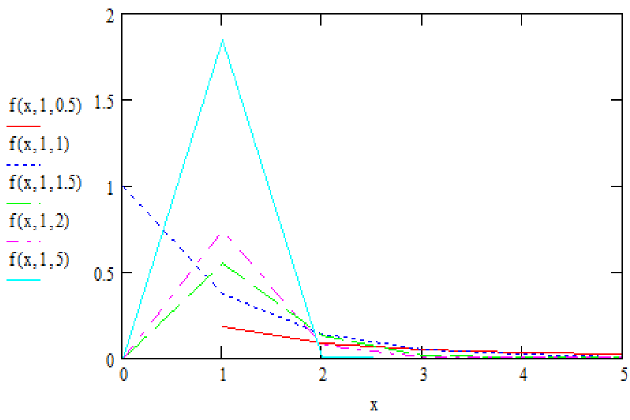


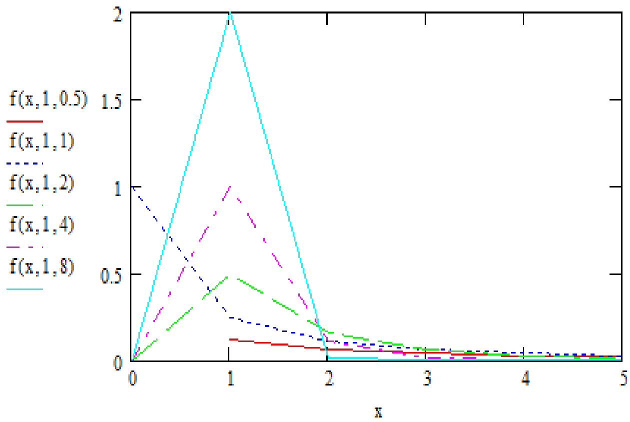
Fig. 1. Density functions of the Weibull distribution at η=1 and β = 0.5, 1, 1.5, 2, 5

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 8, August 2013



Fig, 2. Density functions of the log-logistic distribution at ε =1 and σ = 0.5, 1, 2, 4, 8.

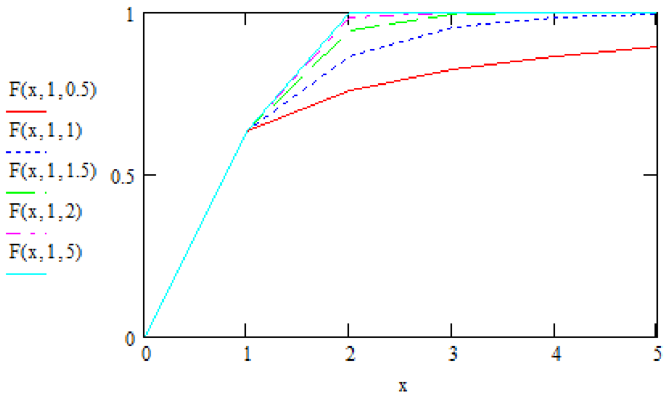


Fig. 3. Cumulative distribution functions of the Weibull at η=1 and β =0.5, 1, 1.5, 2, 5.

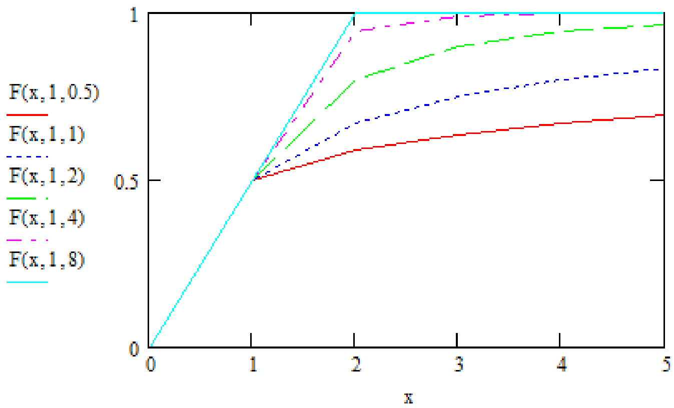


Fig. 4. Cumulative distribution functions of the log-logistic at ε =1 and σ =0.5, 1, 2, 4, 8.

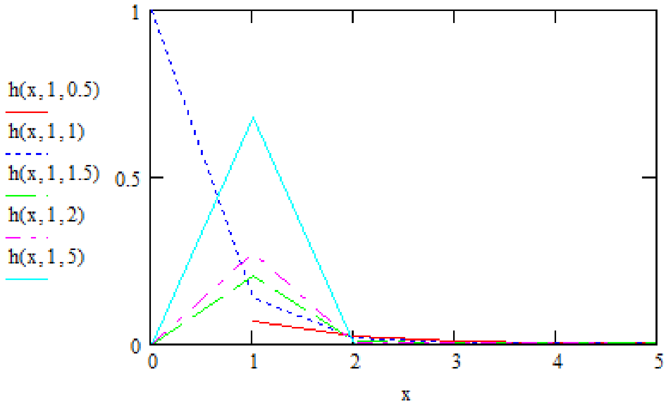


Fig. 5. Hazard functions of the Weibull at η=1 and β =0.5, 1, 1.5, 2, 5.

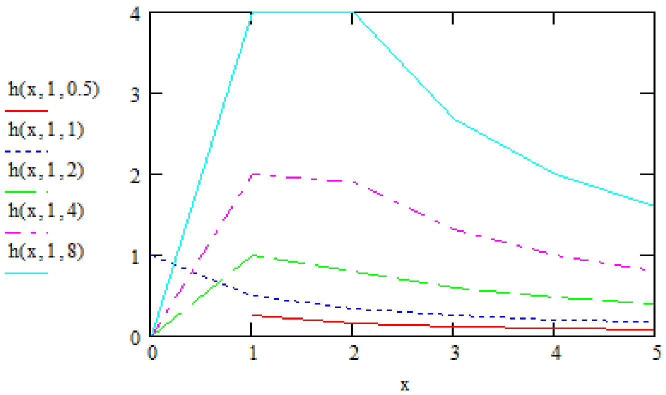


Fig. 6. Hazard functions of the log-logistic at ε =1 and σ =0.5, 1, 2, 4, 8.

II. THE TEST STATISTIC

Suppose $X_1,\ldots,X_n$ are independent and identically distributed random variables from any one of the Weibull or log-logistic distribution functions. A Weibull distribution, denoted by WE(η,β), with scale parameter η > 0 and shape parameter β > 0 has probability density function

$$f_{WE}(x;\eta,\beta)=\beta\eta^{\beta}x^{\beta-1}e^{-(x\eta)^{\beta}}, \qquad \eta,\beta>0 \qquad x>0. \tag{2.1}$$

A log-logistic distribution, denoted by LL(ε,σ), with scale parameter ε > 0 and shape parameter σ > 0 has probability density function

$$f_{LL}(x;\varepsilon,\sigma)=\frac{(\sigma/\varepsilon)(x/\varepsilon)^{\sigma-1}}{\left[1+(x/\varepsilon)^{\sigma}\right]^{2}} \qquad \varepsilon,\sigma>0 \qquad x>0. \tag{2.2}$$

The likelihood functions of WE(η,β) and LL(ε,σ) distributions are given as

$$L_{WE}(\eta,\beta)=\beta^{n}\eta^{n\beta}\prod_{i=1}^{n}x_i^{\beta-1}.\exp\left[-\eta^{\beta}\sum_{i=1}^{n}x_i^{\beta}\right],$$

$$L_{LL}(\varepsilon,\sigma)=\frac{\sigma^{n}\left(\frac{1}{\varepsilon}\right)^{n\sigma}\prod_{i=1}^{n}x_i^{\sigma-1}}{\prod_{i=1}^{n}\left[1+(x_i/\varepsilon)^{\sigma}\right]^{2}}.$$

respectively. The ratio of the maximized likelihood (RML) is defined as

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 8, August 2013

RML= $\dfrac{L_{WE}(\hat{\eta},\hat{\beta})}{L_{LL}(\hat{\varepsilon},\hat{\sigma})}$.

Where $(\hat{\eta},\hat{\beta})$ and $(\hat{\varepsilon},\hat{\sigma})$ are the maximum likelihood estimators of (η,β) and (ε,σ) respectively. The logarithm of RML can be obtained as follows

$$T = \ln RML = \ln \frac{L_{WE}(\hat{\eta},\hat{\beta})}{L_{LL}(\hat{\varepsilon},\hat{\sigma})} = \ln L_{WE}(\hat{\eta},\hat{\beta}) - \ln L_{LL}(\hat{\varepsilon},\hat{\sigma})$$

$$= n\ln\left(\frac{n\hat{\beta}\,\hat{\varepsilon}^{\hat{\sigma}}}{\hat{\sigma}\sum_{i=1}^{n} x_i^{\hat{\beta}}}\right) + (\hat{\beta}-\hat{\sigma})\sum_{i=1}^{n}\ln x_i - n + 2\sum_{i=1}^{n}\ln\left[1+(x_i/\hat{\varepsilon})^{\hat{\sigma}}\right]. \quad (2.3)$$

Where

$$\hat{\eta} = \left(\frac{n}{\sum_{i=1}^{n} x_i^{\hat{\beta}}}\right)^{\frac{1}{\hat{\beta}}}$$

The Weibull distribution is chosen if: T >0. Otherwise the log-logistic distribution is the preferred model.

The exact distributions of T given in Equation (2.3) under the respective parent distributions are difficult to obtain, therefore the asymptotic distributions of T under the null hypotheses must be obtained. It is known that, also the probability of correct selection (PCS) depends on the parent distribution. That is, if the data are originally coming from Weibull distribution the probability of correct selection (PCS_{WE}) is given as,

PCS_{WE} = P (T > 0 / data follow Weibull distribution).

Also, if the data are originally coming from log-logistic distribution the probability of correct selection (PCS_{LL}) is given as,

PCS_{LL} = P (T < 0 /data follow log-logistic distribution).

Therefore, the asymptotic distributions of T can be used to compute the approximate PCS.

III. ASYMPTOTIC DISTRIBUTIONS OF THE TEST STATISTIC UNDER NULL HYPOTHESES

In this section, the asymptotic distribution of T statistic under null hypothesis is obtained in two different cases. Case 1: The data are coming from a Weibull distribution and the alternative they are from a log-logistic distribution. Case 2: The data are coming from a log-logistic distribution and the alternative they are from a Weibull distribution. The results are obtained through two theorems. The following definition and lemmas are needed to follow up the next theorems. The proof of lemmas follows using similar arguments as that of Gupta and Kundu (2003).

Definition

For any Borel measurable function h(.), E_{WE}[h(U)] and V_{WE}[h(U)] denote mean and variance of h(U) under the assumption that U follows WE(.,.). Similarly define E_{LL}[h(U)] and V_{LL}[h(U)] as mean and variance of h(U) under the assumption that U follows LL(.,.). Also if g(.) and h(.) are two Borel measurable functions, define along the same line :

Cov_{WE}[g(U),h(U)]= E_{WE}[g(U)h(U)] – E_{WE}[g(U)]E_{WE}[h(U)],

and similarly

Cov_{LL}[g(U),h(U)]= E_{LL}[g(U)h(U)] – E_{LL}[g(U)]E_{LL}[h(U)],

where U follows WE(.,.) and LL(.,.) respectively. In the following a.s. denote the almost sure convergence.

Lemma 1. Under the assumption that the data are from WE(η, β), and as $n \longrightarrow \infty$, we have:

(i) $\hat{\eta} \longrightarrow \eta$, a.s., $\hat{\beta} \longrightarrow \beta$, a.s., where

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 8, August 2013

$$E_{WE}(\ln\ f_{WE}(X;\eta,\beta)) = \max_{\bar{\eta}\bar{\beta}} E_{WE}(\ln\ f_{WE}(X;\bar{\eta},\bar{\beta})).$$

(ii) $\hat{\varepsilon} \longrightarrow \tilde{\varepsilon}$, a.s., $\hat{\sigma} \longrightarrow \tilde{\sigma}$, a.s., where

$$E_{WE}(\ln\ f_{LL}(X;\tilde{\varepsilon},\tilde{\sigma})) = \max_{\varepsilon,\sigma} E_{WE}(\ln\ f_{LL}(X;\varepsilon,\sigma)).$$

Let us denote

$$T^* = \ln \frac{L_{WE}(\eta,\beta)}{L_{LL}(\tilde{\varepsilon},\tilde{\sigma})}.$$

(iii) $\frac{1}{\sqrt{n}}\left[T - E_{WE}(T)\right]$ is asymptotically equivalent to $\frac{1}{\sqrt{n}}\left[T^* - E_{WE}(T^*)\right]$.

Lemma 2. Under the assumption that the data are from LL(ε,σ), and as $n \longrightarrow \infty$, we have

(i) $\hat{\sigma} \longrightarrow \sigma$, a.s., $\hat{\varepsilon} \longrightarrow \varepsilon$ a.s., where

$$E_{LL}(\ln f_{LL}(X;\varepsilon,\sigma)) = \max_{\bar{\varepsilon},\bar{\sigma}} E_{LL}(\ln f_{LL}(X;\bar{\varepsilon},\bar{\sigma})).$$

(ii) $\hat{\eta} \longrightarrow \tilde{\eta}$, a.s., $\hat{\beta} \longrightarrow \tilde{\beta}$, a.s., where

$$E_{LL}(\ln f_{WE}(X;\tilde{\eta},\tilde{\beta})) = \max_{\eta,\beta} E_{LL}(\ln\ f_{WE}(X;\eta,\beta)).$$

Let us denote

$$T_* = \ln \frac{L_{WE}(\tilde{\eta},\tilde{\beta})}{L_{LL}(\varepsilon,\sigma)}.$$

(iii) $\frac{1}{\sqrt{n}}\left[T - E_{LL}(T)\right]$ is asymptotically equivalent to $\frac{1}{\sqrt{n}}\left[T_* - E_{LL}(T_*)\right]$.

Theorem 1. Under the assumption that the data are from a Weibull distribution, the distribution of T is approximately normally distributed with mean $E_{WE}(T)$ and variance $V_{WE}(T)$, where

$$\frac{1}{n} E_{WE}(T) \approx AM_{WE}(\eta,\beta)$$

$$= \ln(\frac{\beta}{\tilde{\sigma}}) + \frac{\left(\beta - \tilde{\sigma}\right)}{\beta} E_E\left(\ln Z\right) - 1 - \tilde{\sigma}\ln(\eta\,\tilde{\varepsilon}) + 2E_E \ln[(\eta\,\tilde{\varepsilon})^{\tilde{\sigma}} + Z^{\frac{\tilde{\sigma}}{\beta}}], \qquad (3.1)$$

and

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 8, August 2013

$$\frac{1}{n}V_{WE}(T) \approx AV_{WE}(\eta,\beta)$$

$$= \left(\frac{\beta-\tilde{\sigma}}{\beta}\right)^2 V_E(\ln Z) + 1 + 4V_E\left(\ln[(\eta\tilde{\varepsilon})^{\tilde{\sigma}} + Z^{\frac{\tilde{\sigma}}{\beta}}]\right) - 2\left(\frac{\beta-\tilde{\sigma}}{\beta}\right)Cov_E(Z,\ln Z)$$

$$+4\left(\frac{\beta-\tilde{\sigma}}{\beta}\right)Cov_E\left(\ln Z, \ln[(\eta\tilde{\varepsilon})^{\tilde{\sigma}} + Z^{\frac{\tilde{\sigma}}{\beta}}]\right) - 4Cov_E\left(Z, \ln[(\eta\tilde{\varepsilon})^{\tilde{\sigma}} + Z^{\frac{\tilde{\sigma}}{\beta}}]\right). \qquad (3.2)$$

With Z ~ E(1).

Proof

Let us assume that n data points $x_1, x_2, \ldots, x_n$ are obtained from WE(η, β) with scale parameter η and shape parameter β.

Now to obtain $\tilde{\varepsilon}$ and $\tilde{\sigma}$, as defined in Lemma 1 and define

$$h(\varepsilon,\sigma) = E_{WE}\,[\ln(f_{LL}(X;\varepsilon,\sigma)]$$

$$= E_{WE}\left[\ln\sigma - \ln\varepsilon + (\sigma-1)[\ln X - \ln\varepsilon] - 2\ln\left[1+\left(\frac{X}{\varepsilon}\right)^{\sigma}\right]\right]$$

$$= \ln\sigma + \frac{(\sigma-1)}{\beta}E_E(\ln Z) + (\sigma+1)\ln\eta + \sigma\ln\varepsilon - 2E_E\ln[(\eta\varepsilon)^{\sigma} + Z^{\frac{\sigma}{\beta}}]. \qquad (3.3)$$

Where Z ~ E(1). Differentiating Equation(3.3) with respect to σ, then equating by zero we get:

$$\frac{\partial h(\varepsilon,\sigma)}{\partial\sigma} = \frac{1}{\tilde{\sigma}} + \frac{1}{\beta}E_E(\ln Z) + \ln\eta + \ln\tilde{\varepsilon} - 2\ln(\eta\tilde{\varepsilon})E_E\left(\frac{1}{1+(\frac{Z^{\frac{1}{\beta}}}{\eta\tilde{\varepsilon}})^{\tilde{\sigma}}}\right) - 2E_E\left(\frac{\ln Z^{\frac{1}{\beta}}}{1+(\frac{Z^{\frac{1}{\beta}}}{\eta\tilde{\varepsilon}})^{-\tilde{\sigma}}}\right) = 0 \cdot \qquad (3.4)$$

Also differentiating Equation(3.3) with respect to ε, equating with zero we get

$$E_E\left(\frac{1}{1+(\frac{Z^{\frac{1}{\beta}}}{\eta\tilde{\varepsilon}})^{\tilde{\sigma}}}\right) = \frac{1}{2}. \qquad (3.5)$$

Using Equation (3.5) in Equation (3.4) we get:

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 8, August 2013

$$\frac{1}{\tilde{\sigma}}+\frac{1}{\beta}E_E(\ln Z)-2\,E_E\left(\frac{\ln Z^{\frac{1}{\beta}}}{1+(\frac{Z^{\frac{1}{\beta}}}{\eta\tilde{\varepsilon}})^{-\tilde{\sigma}}}\right)=0\cdot \tag{3.6}$$

$\tilde{\sigma}$ and $\tilde{\varepsilon}$ can be obtained by solving Equations (3.5) and (3.6). From these equations, it is clear that $\tilde{\sigma}$ and $\tilde{\varepsilon}$ are both functions of η and β.

To obtain $E_{WE}(T)$ and $V_{WE}(T)$ denote,

$$\lim_{n\to\infty}\frac{1}{n}E_{WE}(T)=AM_{WE}(\eta,\beta),$$

and

$$\lim_{n\to\infty}\frac{1}{n}V_{WE}(T)=AV_{WE}(\eta,\beta)\ .$$

Therefore for large n we get,

$$\frac{1}{n}E_{WE}(T)\approx AM_{WE}(\eta,\beta)$$

$$=\ E_{WE}\left[\ln f_{WE}(X;\eta,\beta)-\ln f_{LL}(X;\tilde{\varepsilon},\tilde{\sigma})\right]$$

$$=\ln\beta+\beta\ln\eta+\frac{(\beta-1)}{\beta}E_E(\ln Z)-(\beta-1)\ln\eta-\eta^{\beta}\frac{1}{\eta^{\beta}}$$

$$-\ln\tilde{\sigma}-\frac{(\tilde{\sigma}-1)}{\beta}E_E(\ln Z)-(\tilde{\sigma}+1)\ln\eta-\tilde{\sigma}\ln\tilde{\varepsilon}+2E_E\ln[(\eta\tilde{\varepsilon})^{\tilde{\sigma}}+Z^{\frac{\tilde{\sigma}}{\beta}}].$$

Then the result of Equation (3.1) is obtained. Also for large n we have,

$$\frac{1}{n}V_{WE}(T)\approx AV_{WE}(\eta,\beta)$$

$$=V_{WE}\left[\ln f_{WE}(X;\eta,\beta)-\ln\ f_{LL}(X;\tilde{\varepsilon},\tilde{\sigma})\right]$$

$$=\left(\beta-\tilde{\sigma}\right)^2V_{WE}(\ln X)+\eta^{2\beta}V_{WE}\left(X^{\beta}\right)+4V_{WE}\ln\left[1+\left(\frac{X}{\tilde{\varepsilon}}\right)^{\tilde{\sigma}}\right]$$

$$-2\eta^{\beta}\left(\beta-\tilde{\sigma}\right)Cov_{WE}(\ln X,X^{\beta})+4\left(\beta-\tilde{\sigma}\right)Cov_{WE}\left(\ln X,\ln\left[1+\left(\frac{X}{\tilde{\varepsilon}}\right)^{\tilde{\sigma}}\right]\right)$$

$$-4\eta^{\beta}Cov_{WE}\left(X^{\beta},\ln\left[1+\left(\frac{X}{\tilde{\varepsilon}}\right)^{\tilde{\sigma}}\right]\right).$$

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 8, August 2013

Then the result of Equation (3.2) is obtained. Using the central limit theorem and using (iii) of Lemma 1, one can easily shows that $\frac{1}{\sqrt{n}}\left[T^{*}-E_{WE}(T^{*})\right]$ is asymptotically normally distributed with mean $E_{WE}(T)$ and variance $V_{WE}(T)$ given in Equations (3.1) and (3.2) respectively. □

Theorem 2. Under the assumption that the data are from log-logistic distribution, the distribution of T is approximately normally distributed with mean $E_{LL}(T)$ and variance $V_{LL}(T)$, where

$$\frac{1}{n}E_{LL}(T)\approx AM_{LL}(\varepsilon,\sigma)$$

$$=\ln\left(\frac{\tilde{\beta}}{\sigma}\right)+\tilde{\beta}\ln(\tilde{\eta}\,\varepsilon)+\frac{(\tilde{\beta}-\sigma)}{\sigma}E_{LL}(\ln Y)-(\varepsilon\,\tilde{\eta})^{\tilde{\beta}}E_{LL}\left(Y^{\frac{\tilde{\beta}}{\sigma}}\right)+2E_{LL}\ln[1+Y], \qquad (3.7)$$

and

$$\frac{1}{n}V_{LL}(T)\approx AV_{LL}(\varepsilon,\sigma)$$

$$=\left(\frac{\tilde{\beta}-\sigma}{\sigma}\right)^{2}V_{LL}(\ln Y)+4V_{LL}\ln[1+Y]+(\varepsilon\tilde{\eta})^{2\tilde{\beta}}V_{LL}\left(Y^{\frac{\tilde{\beta}}{\sigma}}\right)-2(\varepsilon\tilde{\eta})^{\tilde{\beta}}\left(\frac{\tilde{\beta}-\sigma}{\sigma}\right)Cov_{LL}(\ln Y,Y^{\frac{\tilde{\beta}}{\sigma}})$$

$$+4\left(\frac{\tilde{\beta}-\sigma}{\sigma}\right)Cov_{LL}(\ln Y,\ln[1+Y])-4(\varepsilon\tilde{\eta})^{\tilde{\beta}}Cov_{LL}\left(Y^{\frac{\tilde{\beta}}{\sigma}},\ln[1+Y]\right). \qquad (3.8)$$

With $Y\sim LL(1,1)$.

Proof

Let us assume that n data points $x_1,x_2,\ldots,x_n$ are obtained from $LL(\varepsilon,\sigma)$ with scale parameter ε and shape parameter σ. Now to obtain $\tilde{\eta}$ and $\tilde{\beta}$ as defined in Lemma 2 and define

$$g(\eta,\beta)=E_{LL}[\ln(f_{WE}(X;\eta,\beta)]$$

$$=E_{LL}[\ln\beta+\beta\ln\eta+(\beta-1)\ln X-(X\eta)^{\beta}]$$

$$=\ln\beta+\beta\ln\eta+(\beta-1)\left(\frac{1}{\sigma}E_{LL}(\ln Y)+\ln\varepsilon\right)-(\varepsilon\eta)^{\beta}E_{LL}\left(Y^{\frac{\beta}{\sigma}}\right). \qquad (3.9)$$

Where $Y\sim LL(1,1)$.

Differentiating Equation (3.9) with respect to η and equating by zero we get

$$\tilde{\eta}=\frac{1}{\varepsilon}\left(\frac{1}{E_{LL}(Y^{\frac{\tilde{\beta}}{\sigma}})}\right)^{\frac{1}{\tilde{\beta}}}. \qquad (3.10)$$

Also, differentiating Equation (3.9) with respect to β and equating by zero we get:

$$\frac{1}{\tilde{\beta}} - \frac{1}{\sigma} \frac{Cov_{LL}\left(Y^{\frac{\tilde{\beta}}{\sigma}}, \ln Y\right)}{E_{LL}(Y^{\frac{\tilde{\beta}}{\sigma}})} = 0. \tag{3.11}$$

$\tilde{\eta}$ and $\tilde{\beta}$ can be obtained by solving Equations (3.10) and (3.11). From these equations it is clear that $\tilde{\eta}$ and $\tilde{\beta}$ are both functions of ε and σ.

Now to obtain $E_{LL}(T)$ and $V_{LL}(T)$, denote

$$\lim_{n\to\infty} \frac{1}{n} E_{LL}(T) = AM_{LL}(\varepsilon, \sigma),$$

and

$$\lim_{n\to\infty} \frac{1}{n} V_{LL}(T) = AV_{LL}(\varepsilon, \sigma).$$

Therefore for large n we get,

$$\frac{1}{n} E_{LL}(T) \approx AM_{LL}(\varepsilon, \sigma)$$

$$= E_{LL}\left[\ln f_{WE}(X; \tilde{\eta}, \tilde{\beta}) - \ln f_{LL}(X; \varepsilon, \sigma)\right]$$

$$= \ln \tilde{\beta} + \tilde{\beta} \ln \tilde{\eta} + \frac{(\tilde{\beta} - \sigma)}{\sigma} E_{LL}(\ln Y) + \tilde{\beta} \ln \varepsilon - \sigma \ln \varepsilon - (\varepsilon \tilde{\eta})^{\tilde{\beta}} E_{LL}\left(Y^{\frac{\tilde{\beta}}{\sigma}}\right)$$

$$- \ln \sigma + \sigma \ln \varepsilon + 2E_{LL} \ln[1 + Y].$$

Then the result of Equation (3.7) is obtained. Also for large n we have

$$\frac{1}{n} V_{LL}(T) \approx AV_{LL}(\varepsilon, \sigma)$$

$$= V_{LL}[\ln f_{WE}(X; \tilde{\eta}, \tilde{\beta}) - \ln f_{LL}(X; \varepsilon, \sigma)]$$

$$= (\tilde{\beta} - \sigma)^2 V_{LL}(\ln X) + 4V_{LL} \ln\left[1 + \left(\frac{X}{\varepsilon}\right)^{\sigma}\right] + \tilde{\eta}^{2\tilde{\beta}} V_{LL}(X^{\tilde{\beta}}) - 2\tilde{\eta}^{\tilde{\beta}} (\tilde{\beta} - \sigma) Cov_{LL}(\ln X, X^{\tilde{\beta}})$$

$$+ 4(\tilde{\beta} - \sigma) Cov_{LL}\left(\ln X, \ln\left[1 + \left(\frac{X}{\varepsilon}\right)^{\sigma}\right]\right) - 4\tilde{\eta}^{\tilde{\beta}} Cov_{LL}\left(X^{\tilde{\beta}}, \ln\left[1 + \left(\frac{X}{\varepsilon}\right)^{\sigma}\right]\right).$$

Then the result of Equation (3.8) is obtained. Using the central limit theorem and using (iii) of Lemma 2, one can easily shows that $\frac{1}{\sqrt{n}}\left[T_* - E_{LL}(T_*)\right]$ is asymptotically normally distributed with mean $E_{LL}(T)$ and variance $V_{LL}(T)$ given in Equations (3.7) and (3.8) respectively. □

In the next section the minimum sample size required to discriminate between Weibull and log-logistic distribution is obtained. In order to do that, $\tilde{\sigma}, \tilde{\varepsilon}$, $AM_{WE}(\eta, \beta)$, $AV_{WE}(\eta, \beta)$, $\tilde{\eta}$, $\tilde{\beta}$, $AM_{LL}(\varepsilon, \sigma)$ and $AV_{LL}(\varepsilon, \sigma)$ are computed numerically using Equations (3.5), (3.6), (3.1), (3.2), (3.10), (3.11), (3.7) and (3.8) with

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 8, August 2013

Mathcad program. Without lose of generality, we take η =1 and ε =1, while β =0.5, 1, 1.5, 2, 2.5, 3, 5 and σ =0.5, 1, 1.5, 2, 2.5, 3, 4. The results are given in Tables 1 and 2.

Table I. Different values of $AM_{WE}(1,\beta)$, $AV_{WE}(1,\beta)$, $\tilde{\varepsilon}$ and $\tilde{\sigma}$ at η =1 and β =0.5, 1, 1.5, 2, 2.5, 3, 5.

β	$AM_{WE}(1,\beta)$	$AV_{WE}(1,\beta)$	$\tilde{\varepsilon}$	$\tilde{\sigma}$
0.5	-0.049	1.161	0.397	0.718
1	-0.074	2.407	0.630	1.437
1.5	-0.142	3.810	0.735	2.155
2	-0.211	5.358	0.794	2.874
2.5	-0.279	7.021	0.831	3.592
3	-0.348	8.786	0.857	4.311
5	-0.622	12.975	0.912	7.185

Table II. Different values of $AM_{LL}(1,\sigma)$, $AV_{LL}(1,\sigma)$, $\tilde{\eta}$ and $\tilde{\beta}$ at ε =1and σ =0.5, 1, 1.5, 2, 2.5, 3, 4.

σ	$AM_{LL}(1,\sigma)$	$AV_{LL}(1,\sigma)$	$\tilde{\eta}$	$\tilde{\beta}$
0.5	-0.169	8.680	0.164	0.25
1	-0.194	8.839	0.405	0.5
1.5	-0.219	9.006	0.548	0.75
2	-0.245	9.171	0.637	1
2.5	-0.271	9.337	0.697	1.25
3	-0.298	9.505	0.740	1.5
4	-0.355	9.859	0.798	2

IV. DETERMINATION OF SAMPLE SIZE

In this section, a method to determine the minimum sample size needed to discriminate between Weibull and log-logistic distributions is proposed. The same arguments as that given in Gupta and Kundu (2003) are followed. It is known that if two distribution functions are very close, one needs a very large sample size to discriminate between them. While, if they are quite different, then one may not need very large sample size to discriminate between them. Also from a practical point of view, one may not need to differentiate between two so closed distribution functions. Therefore, it is expected that the user will specify before hand the minimum distance D* that he does not want to make the discrimination between two distribution functions if their distance is less than it. This minimum distance is called tolerance limit. Here the Kolmogrov-Smirnov (K-S) distance is used to measure the closeness between the Weibull and log-logistic distributions. Where, the Kolmogrov-Smirnov (K–S) distance between two distribution functions, say F(x) and G(x) is defined as $\sup_x |F(x) - G(x)|$. Also it is expected that the user will specify beforehand the probability of correct selection (PCS) to achieve a certain protection level P*. With the help of K–S distance and PCS the required sample size n is obtained as follows. Considering Case 1 where it is assumed that the data are coming from WE(η,β),

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 8, August 2013

then from Theorem 1, T is approximately normally distributed with mean $E_{WE}(T)$ and variance $V_{WE}(T)$. In this case the PCS_{WE} is given by

$PCS_{WE}(\eta,\beta) = \Pr(T > 0)$

$$\approx \Phi\left(\frac{n \times AM_{WE}(\eta,\beta)}{\sqrt{n \times AV_{WE}(\eta,\beta)}}\right)$$

Where Φ is the distribution function of the standard normal random variable. $AM_{WE}(\eta,\beta)$ and $AV_{WE}(\eta,\beta)$ are given in Equations (3.1) and (3.2) respectively. Therefore, to determine the minimum sample size required to achieve at least P* protection level, solve for n the equation

$$\Phi\left(\frac{n \times AM_{WE}(\eta,\beta)}{\sqrt{n \times AV_{WE}(\eta,\beta)}}\right) = P^*,$$

i.e.,

$$n = \frac{Z_{P^*}{}^2 \times AV_{WE}(\eta,\beta)}{(AM_{WE}(\eta,\beta))^2}. \tag{4.1}$$

Here Z_{P^*} is the 100 P* percentile point of a standard normal distribution. The values of n are obtained and reported in Table 3 where For P*=0.7, η=1 and β=0.5, 1, 1.5, 2, 2.5, 3, 5, with $\tilde{\sigma}$ and $\tilde{\varepsilon}$ as given in Table 1. Also the K-S distances between WE(1, β) and LL($\tilde{\varepsilon}$,$\tilde{\sigma}$) are obtained and presented in Table 3. Considering Case 2 where it is assumed that the data are coming from WE(η,β), then from Theorem 2, T is approximately normally distributed with mean $E_{LL}(T)$ and variance $V_{LL}(T)$. In this case, the PCS_{LL} is given by

$PCS_{LL}(\varepsilon,\sigma) = \Pr(T < 0)$

$$\approx \Phi\left(-\frac{n \times AM_{LL}(\varepsilon,\sigma)}{\sqrt{n \times AV_{LL}(\varepsilon,\sigma)}}\right).$$

Where $AM_{LL}(\varepsilon,\sigma)$ and $AV_{LL}(\varepsilon,\sigma)$ are given in Equations (3.7) and (3.8) respectively. Therefore, to determine the minimum sample size required to achieve at least P* protection level, solve for n the equation

$$\Phi\left(-\frac{n \times AM_{LL}(\varepsilon,\sigma)}{\sqrt{n \times AV_{LL}(\varepsilon,\sigma)}}\right) = P^*,$$

i.e.,

$$n = \frac{Z_{P^*}{}^2 \times AM_{LL}(\varepsilon,\sigma)}{(AV_{LL}(\varepsilon,\sigma))^2}. \tag{4.2}$$

The values of n are obtained and reported in Table 4 where P*= 0.7, ε=1 and σ=0.5, 1, 1.5, 2, 2.5, 3, 4, with $\tilde{\eta}$ and $\tilde{\beta}$ as given in Table 2. Also the K-S distances between LL(1, σ) and WE($\tilde{\eta}$, $\tilde{\beta}$) are obtained and presented in Table 4. From Tables 3 and 4 it can be seen that for a given PCS i.e., p*=0.7, as β and σ increase the sample size decreases. Also as the K-S distance between the two distributions increases the sample size decreases as expected. To resume, if one knows the range of the shape parameter of the null distribution and for a given PCS that achieves a certain protection level P*, then the minimum sample size can be obtained by taking the maximum n obtained from Equations (4.1) and (4.2). But unfortunately in practice the shape parameter may be completely unknown; therefore, the K-S distances can replace the unknown parameters to take the decision. That is, for a given protection level P* and a given pre-specified tolerance limit D*, the minimum sample size can be obtained by taking the maximum n obtained from

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 8, August 2013

Equations (4.1), (4.2). For example, suppose that for a given P*= 0.7 and for β= 0.5 and σ= 0.5, then from Tables 3 and 4 the minimum sample size required to discriminate between Weibull and log-logistic distributions is max(133, 84) =133. On the other hand if β and σ are unknown and suppose that the practitioner wants to discriminate between a Weibull and a log-logistic distribution functions only when the distance between them is greater than or equal to 0.180, i.e., D* ≥ 0.180 and with P* = 0.7. Then from Tables 3 and 4, it is clear that, D* ≥ 0.180 if β ≥ 0.5 and σ ≥ 3. Also, when the null distribution is Weibull, then for the tolerance limit D* ≥ 0.180, one needs n=133 to meet the PCS, P*= 0.7. Similarly when the null distribution is log-logistic then one needs n=29 to meet the same protection level. Finally, the minimum sample size required to discriminate between Weibull and log-logistic distributions with P*= 0.7 and D* ≥ 0.180 is max(133, 29) = 133.

Table 3. The sample size n and the K-S distance for P*=0.7, η=1, β =0.5, 1, 1.5, 2, 2.5, 3, 5 when the null hypothesis is Weibull distribution.

β	0.5	1	1.5	2	2.5	3	5
n	133	121	52	33	25	20	9
K-S	0.187	0.195	0.196	0.208	0.22	0.231	0.277

Table 4. The sample size n and the K-S distance for P*=0.7, ε=1, σ=0.5, 1, 1.5, 2, 2.5, 3, 4 when the null hypothesis is log-logistic distribution.

σ	0.5	1	1.5	2	2.5	3	4
n	84	65	52	42	35	29	22
K-S	0.123	0.135	0.147	0.158	0.17	0.182	0.205

Notice that, Tables 3 and 4 are obtained for the protection level 0.7 but for other protection levels the tables can be easily modified. For example, if we need a sample size corresponding to protection level P*=0.9, then all the entries corresponding to the row of n, must be multiplied by $Z_{0.9}^2 / Z_{0.7}^2$.

V. Data Analysis

For illustrative purposes, two real data sets to discriminate between the Weibull and log-logistic distribution functions are analyzed.

Data Set 1: The first data set (Gupta and kundu(2003)) represent the failure times of 30 air conditions of an airplane (in hours): 23, 261, 87, 7, 120, 14, 62, 47, 225,71, 246, 21, 42, 20, 5, 12, 120, 11, 3, 14, 71, 11, 14, 11, 16, 90, 1, 16, 52, 95.

When the Weibull distribution is used, the MLEs of the different parameters are: $\hat{\eta}$ =0.01827 and $\hat{\beta}$ = 0.85494. Also ln[$L_{WE}(\hat{\eta},\hat{\beta})$] = -152.0068. Similarly when the log-logistic distribution is used, the MLEs of the different parameters are: $\hat{\sigma}$ =1.2015 and $\hat{\varepsilon}$ =26.61693. Also ln[$L_{LL}(\hat{\varepsilon},\hat{\sigma})$]= -152.34578. Consequently T= 0.3389. Therefore, by using the maximum likelihood ratio test to discriminate between Weibull and log-logistic distributions, the Weibull model is chosen for this data set.

Data Set 2: The second data set (Gupta and kundu (2003)) represent the number of million revolutions before failure for each of 23 ball bearings in the life test and they are: 17.88, 28.92, 33.00, 41.52, 42.12, 45.60, 48.80, 51.84, 51.96, 54.12, 55.56, 67.80, 68.44, 68.64, 68.88, 84.12, 93.12, 98.64, 105.12, 105.84, 127.92, 128.04, 173.40.

When the Weibull model is used, the MLEs of η and β are: $\hat{\eta}$ = 0.01217 and $\hat{\beta}$ = 2.10490. Also ln[$L_{WE}(\hat{\eta},\hat{\beta})$]= -113.67899. Similarly, if the log-logistic model is used, the MLEs of ε and σ parameters are: $\hat{\sigma}$ =0.30078 and $\hat{\varepsilon}$ =64.00749 Also ln[$L_{LL}(\hat{\varepsilon},\hat{\sigma})$]= -113.36619. Consequently T= -0.3128. Therefore, by using the maximum likelihood ratio test to discriminate between Weibull and log-logistic distributions, the log-logistic model is chosen for this data set.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 8, August 2013

VI. CONCLUSION

In this paper we consider the problem of discriminating between Weibull and log-logistic distribution functions. It is assumed that a data are coming either from Weibull or log-logistic distribution. The maximized likelihood ratio test to discriminate between them is used. The asymptotic distributions of the logarithm of the ratio of the maximized likelihood are obtained. These asymptotic results are used to estimate the probability of correct selection. The minimum sample size needed to discriminate between the two distribution functions for a user specified probability of correct selection and a tolerance limit based on the distance between the two distributions is calculated. Two real data life are analyzed to see how the proposed method works in practice.

REFERENCES

[1] Bain, L. J., and Engelhardt, M., "Probability of Correct Selection of Weibull versus Gamma based on Likelihood Ratio", Communications in Statistics, Theory and Methods, Vol. 9, pp. 375-381, 1980.
[2] Chen, W. W., "On the Tests of Separate Families of Hypotheses with Small Sample Size", Journal of Statistical Computations and Simulations, vol. 2, 183-187, 1980.
[3] Cox, D. R., "Tests of Separate Families of Hypotheses. Proceeding of the Fourth Berkely Symposium in Mathematical Statistics and Probability", Berkely, University of California Press, pp.105-123, 1961.
[4] Cox, D. R., "Further Results on Tests of Separate Families of Hypotheses", Journal of the Royal Statistical Society, Series B, Vol. 24, pp. 406-424, 1962.
[5] Dey, A. K., and Kundu, D. K., "Discriminating among the Log-normal, Weibull and Generalized Exponential Distributions", IEEE Transactions on Reliability, Vol. 58, no. 3, pp.416-424, 2009.
[6] Dey, A. K., and Kundu, D. K., "Discriminating between the Log-Normal and Log-Logistic Distributions", Communications in Statistics, Theory and Methods, Vol. 39, pp.280 – 292, 2010.
[7] Dey, A. K., and Kundu, D. K., "Discriminating between the Weibull and Log-Normal Distributions for Type-II Censored Data", Statistics, Vol. 46, no. 2, pp.197- 214, 2012.
[8] Dumonceaux, R., and Antle., C.E., "Discriminating between the Log-Normal and Weibull Distribution", Technometrics, Vol.15(4), pp.923-926, 1973.
[9] Fearn, D. H., and Nebenzahl, E., "On the Maximum Llikelihood Ratio Method of Deciding Between the Weibull and Gamma Distributions", Communications in Statistics, Theory and Methods, Vol. 20, 579-593, 1991.
[10] Firth, D., "Multiplicative Errors: Log-Normal or Gamma?", Journal of the Royal Statistical Society, Series B, 2, pp.266-268, 1988.
[11] Gupta, R. D., and Kundu, D. K., "Discriminating between Weibull and Generalized Exponential Distributions", Journal of Computational Statistics and Data Analysis, Vol. 43, pp.179 – 196, 2003.
[12] Gupta, R. D., and Kundu, D. K., "Discriminating between the Gamma and Generalized Exponential Distributions", Journal of Statistical Computation and Simulation, Vol. 74(2), pp.107-121, 2004.
[13] Jackson, O. A. Y., "Some results on tests separate families of hypotheses", Biometrika, Vol. 55, pp.355-363, 1968.
[14] Kappenman, R. F., "On a Method for Selecting a Distributional Model", Communication in Statistics, Theory and Methods, Vol. 11, pp.663-672, 1982.
[15] Kim, J. S., and Yum, B-J., "Selection between Weibull and log-normal distributions: a comparative simulation study", Computational Statistics and Data Analysis, Vol. 53, pp.477 – 485, 2008.
[16] Kundu, D. K., and Manglick, A., "Discriminating between Weibull and Log-Normal Distributions", Naval Research Logistics (NRL), Vol. 51, Issue 6, pp.893–905, 2004.
[17] Kundu, D. K., and Manglick, A., "Discriminating between the Log-Normal and Gamma Distributions", Journal of the Applied Statistical Sciences, Vol. 14, pp.175-187, 2005.
[18] Lawless, J. F., "Statistical Models and Methods for Lifetime Data", John Wiley and Sons, New York, 1982.
[19] Pascual, F. G., "Maximum Likelihood Estimation under Misspecified Log-Normal and Weibull Distributions", Communications in Statistics, Simulation and Computations, Vol. 34, pp.503-524, 2005.
[20] Pereira, B. de B., "A Note on the Consistency and on the Finite Sample Comparisons of Some Tests of Separate Families of Hypotheses", Biometrika, Vol. 64, pp.109-113, 1977.
[21] Siswadi, and Quesenberry, C. P., "Selecting among Weibull, Log-Normal and Gamma Distributions using Complete and Censored Samples", Naval Research Logistics Quarterly, Vol. 29 (4), pp.557-569, 1982.
[22] Stephens, M. A., "EDF statistics for goodness of fit and some comparisons", Journal of the American Statistical Association, Vol.69, pp.730-737, 1974.
[23] Wiens, B. L., "When Log-Normal and Gamma Models Give Different Results: A Case Study", American Statistician, Vol. 53, pp.89-93. 1999.