ISSN: 1803-3814

ON TESTING REDUCTION OF A LEFT-CENSORED WEIBULL DISTRIBUTION TO AN EXPONENTIAL SUBMODEL

Michal Fusek

Brno University of Technology
Department of Mathematics, Faculty of Electrical Engineering and Communication
Technická 2848/8, 61600 Brno
Czech Republic
fusekmi@feec.vutbr.cz

Abstract: *When analyzing environmental or chemical data, it is often necessary to deal with left-censored observations. Since the distribution of the observed variable is often asymmetric, the exponential or the Weibull distribution can be used. This paper summarizes statistical model of a multiply left-censored Weibull distribution, and estimation of its parameters and their variances using the expected Fisher information matrix. Since in many situations the Weibull distribution is unnecessarily complicated for data modelling, statistical tests (the Lagrange multiplier test, the likelihood ratio test, the Wald test) for assessing suitability of replacement of the censored Weibull distribution with the exponential submodel are introduced and their power functions are analyzed using simulations.*

Keywords: *Asymptotic tests, multiply left-censored data, Fisher information matrix, maximum likelihood.*

1 Introduction

Censored data occurs frequently in many application areas. When experimental units may not be observed for the full time before failure, the right censoring is considered. When a substance or an attribute being measured is either absent or exists at such a low level that it is not present above the detection limit of a measuring device, the left censoring is used. The right censoring is typically used in survival data analyses [11, 12], and the left censoring is used in analyses of environmental and/or chemical data [10]. Both censoring types can be divided into two categories with regard to the observation period of experimental units or a detection limit of a measuring device. If the detection limit is fixed, the censoring is called Type I censoring or time censoring. In such a situation, the number of censored experimental units is a random variable. In case the number of censored units is fixed, the censoring is called Type II censoring or failure censoring.

When analyzing chemical or environmental data, Type I left-censored data with more than one detection limit is usually present. Therefore, detection limits $d_1 < \cdots < d_k$, $k > 1$, are often considered and we talk about multiply left-censored samples [2, 8]. In such a case only observations above the highest detection limit d_k and the number of observations under the remaining detection limits are available. When dealing with chemical data, two detection limits (a limit of detection and a limit of quantification) are usually present. In such cases, we talk about doubly left-censored data [1, 7].

This paper focuses on Type I multiply left-censored samples. Since interest in statistical analysis of censored data, especially with asymmetric or highly skewed distributions, has increased lately [3, 4, 13, 17, 18], attention will be paid to the censored Weibull distribution. However, despite the fact that the Weibull distribution is very flexible and can be widely used, it can sometimes be unnecessarily complicated for modelling of given data. In such a situation, a much simpler model of the censored exponential distribution (which is a special case of the Weibull distribution) can be utilized [7]. Therefore, a statistical tests for assessing suitability of replacing the Weibull distribution with the exponential distribution will be described and their power will be analyzed.

The following section summarizes a statistical model of the multiply left-censored Weibull distribution, estimation of its parameters using the method of maximum likelihood (ML), and the expected Fisher information matrix (FIM) which can be used for estimating the variability of the estimated parameters. In section three, asymptotic tests with nuisance parameters (the Lagrange multiplier test, the likelihood ratio test, the Wald test) for testing suitability of replacement of the censored Weibull distribution with the exponential submodel are introduced, and their power functions considering various sample sizes, censoring schemes and levels of censoring are analyzed using simulations.

2 Multiply Left-Censored Weibull Distribution

Let $X_1, \ldots, X_n$ be a type I multiply left-censored random sample from the Weibull distribution with scale parameter $\lambda > 0$, shape parameter $\tau > 0$, cumulative distribution function

$$F(x, \lambda, \tau) = \begin{cases} 1 - \exp\left[-\left(\frac{x}{\lambda}\right)^{\tau}\right] & \text{for } x \geq 0, \\ 0 & \text{for } x < 0, \end{cases}$$

and probability density function

$$f(x, \lambda, \tau) = \begin{cases} \frac{\tau}{\lambda^{\tau}} x^{\tau-1} \exp\left[-\left(\frac{x}{\lambda}\right)^{\tau}\right] & \text{for } x \geq 0, \\ 0 & \text{for } x < 0. \end{cases}$$

Let $X_{(1)}, \ldots, X_{(n)}$ be the ordered sample of $X_1, \ldots, X_n$. Moreover, N_i is the number of observations in the interval $(d_{i-1}, d_i\rangle$, and N_0 is the number of uncensored observations $X_{(n-N_0+1)}, \ldots, X_{(n)}$. In order to simplify the mathematical notation, in all the formulas, we put $d_0 = 0$ and $\ln(d_0) = 0$.

The log-likelihood function of the multiply left-censored sample is (see [5])

$$\begin{aligned} l(\lambda, \tau, N_0, \ldots, N_k, X_{(n-N_0+1)}, \ldots, X_{(n)}) = {} & \log\left(\frac{n!}{N_1! \ldots N_k!}\right) + \sum_{i=1}^{k} N_i \log\left[F(d_i, \lambda, \tau) - F(d_{i-1}, \lambda, \tau)\right] \\ & + \sum_{i=n-N_0+1}^{n} \log\left[f(X_{(i)})\right]. \end{aligned} \tag{1}$$

The ML estimates $\widehat{\lambda}$, $\widehat{\tau}$ of parameters λ, τ were estimated by maximization of the log-likelihood function (1) using the Nelder-Mead simplex algorithm [14].

Since (see [15])

$$\sqrt{n}(\widehat{\lambda} - \lambda) \overset{A}{\sim} \mathrm{N}(0, J_{11}^{-1}),$$
$$\sqrt{n}(\widehat{\tau} - \tau) \overset{A}{\sim} \mathrm{N}(0, J_{22}^{-1}),$$

variability of the estimators $\widehat{\lambda}$, $\widehat{\tau}$ can be calculated from the expected FIM

$$\boldsymbol{J}_n = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} = \begin{bmatrix} -\mathrm{E}\frac{\partial^2 l}{\partial \lambda^2} & -\mathrm{E}\frac{\partial^2 l}{\partial \lambda \partial \tau} \\ -\mathrm{E}\frac{\partial^2 l}{\partial \tau \partial \lambda} & -\mathrm{E}\frac{\partial^2 l}{\partial \tau^2} \end{bmatrix}, \tag{2}$$

where

$$\begin{aligned} J_{11} = {} & n \sum_{i=1}^{k} \frac{\tau^2 \left\{d_{i-1}^{\tau} \exp\left[-\left(\frac{d_{i-1}}{\lambda}\right)^{\tau}\right] - d_i^{\tau} \exp\left[-\left(\frac{d_i}{\lambda}\right)^{\tau}\right]\right\}^2}{\lambda^{2\tau+2} \left\{\exp\left[-\left(\frac{d_{i-1}}{\lambda}\right)^{\tau}\right] - \exp\left[-\left(\frac{d_i}{\lambda}\right)^{\tau}\right]\right\}} \\ & - n \frac{\left(d_k^{\tau} \lambda^{\tau} \tau^2 + d_k^{\tau} \lambda^{\tau} \tau - d_k^{2\tau} \tau^2\right) \exp\left[-\left(\frac{d_k}{\lambda}\right)^{\tau}\right]}{\lambda^{2\tau+2}} - n \frac{\tau}{\lambda^2} \exp\left[-\left(\frac{d_k}{\lambda}\right)^{\tau}\right] + \frac{\tau^2 + \tau}{\lambda^{\tau+2}} E_1, \end{aligned}$$

$$\begin{aligned} J_{22} = {} & n \sum_{i=1}^{k} \frac{\left\{d_{i-1}^{\tau} \ln\left(\frac{d_{i-1}}{\lambda}\right) \exp\left[-\left(\frac{d_{i-1}}{\lambda}\right)^{\tau}\right] - d_i^{\tau} \ln\left(\frac{d_i}{\lambda}\right) \exp\left[-\left(\frac{d_i}{\lambda}\right)^{\tau}\right]\right\}^2}{\lambda^{2\tau} \left\{\exp\left[-\left(\frac{d_{i-1}}{\lambda}\right)^{\tau}\right] - \exp\left[-\left(\frac{d_i}{\lambda}\right)^{\tau}\right]\right\}} \\ & - n \frac{\left(d_k^{\tau} \lambda^{\tau} - d_k^{2\tau}\right) \left[\ln\left(\frac{d_k}{\lambda}\right)\right]^2 \exp\left[-\left(\frac{d_k}{\lambda}\right)^{\tau}\right]}{\lambda^{2\tau}} + \frac{n}{\tau^2} \exp\left[-\left(\frac{d_k}{\lambda}\right)^{\tau}\right] + \frac{(\ln \lambda)^2}{\lambda^{\tau}} E_1 - \frac{2 \ln \lambda}{\lambda^{\tau}} E_2 + \frac{1}{\lambda^{\tau}} E_3, \end{aligned}$$

$$\begin{aligned} J_{12} = J_{21} = {} & -n \sum_{i=1}^{k} \frac{\tau \left\{d_{i-1}^{\tau} \exp\left[-\left(\frac{d_{i-1}}{\lambda}\right)^{\tau}\right] - d_i^{\tau} \exp\left[-\left(\frac{d_i}{\lambda}\right)^{\tau}\right]\right\}}{\lambda^{2\tau+1} \left\{\exp\left[-\left(\frac{d_{i-1}}{\lambda}\right)^{\tau}\right] - \exp\left[-\left(\frac{d_i}{\lambda}\right)^{\tau}\right]\right\}} \\ & \times \left\{d_{i-1}^{\tau} \ln\left(\frac{d_{i-1}}{\lambda}\right) \exp\left[-\left(\frac{d_{i-1}}{\lambda}\right)^{\tau}\right] - d_i^{\tau} \ln\left(\frac{d_i}{\lambda}\right) \exp\left[-\left(\frac{d_i}{\lambda}\right)^{\tau}\right]\right\} \\ & + n \frac{\left[d_k^{\tau} \tau \lambda^{\tau} \ln\left(\frac{d_k}{\lambda}\right) + d_k^{\tau} \lambda^{\tau} - d_k^{2\tau} \tau \ln\left(\frac{d_k}{\lambda}\right)\right] \exp\left[-\left(\frac{d_k}{\lambda}\right)^{\tau}\right]}{\lambda^{2\tau+1}} + \frac{n}{\lambda} \exp\left[-\left(\frac{d_k}{\lambda}\right)^{\tau}\right] \\ & + \frac{\tau \ln \lambda - 1}{\lambda^{\tau+1}} E_1 - \frac{\tau}{\lambda^{\tau+1}} E_2, \end{aligned}$$

and, considering Euler's constant $\gamma_e \doteq 0.57722$,

$$E_1 = n\lambda^\tau \sum_{n_0=0}^{n} \sum_{i=n-n_0+1}^{n} \binom{n-1}{i-1} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \binom{n}{n_0} \exp\left[-n_0 \left(\frac{d_k}{\lambda}\right)^\tau\right] \times \left\{1 - \exp\left[-\left(\frac{d_k}{\lambda}\right)^\tau\right]\right\}^{n-n_0},$$

$$E_2 = n\frac{\lambda^\tau}{\tau} \sum_{n_0=0}^{n} \sum_{i=n-n_0+1}^{n} \binom{n-1}{i-1} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \left[\ln\left(\frac{\lambda^\tau}{n-i+j+1}\right) + 1 - \gamma_e\right] \times \binom{n}{n_0} \exp\left[-n_0 \left(\frac{d_k}{\lambda}\right)^\tau\right] \left\{1 - \exp\left[-\left(\frac{d_k}{\lambda}\right)^\tau\right]\right\}^{n-n_0},$$

$$E_3 = n\frac{\lambda^\tau}{\tau^2} \sum_{n_0=0}^{n} \sum_{i=n-n_0+1}^{n} \binom{n-1}{i-1} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \times \left\{\left[\ln\left(\frac{\lambda^\tau}{n-i+j+1}\right)\right]^2 + 2\ln\left(\frac{\lambda^\tau}{n-i+j+1}\right)(1-\gamma_e) + \frac{\pi^2}{6} - 2\gamma_e + \gamma_e^2\right\} \times \binom{n}{n_0} \exp\left[-n_0 \left(\frac{d_k}{\lambda}\right)^\tau\right] \left\{1 - \exp\left[-\left(\frac{d_k}{\lambda}\right)^\tau\right]\right\}^{n-n_0}.$$

More details and the derivation of FIM (2) can be found in [6].

3 Reduction of the Weibull distribution to the Exponential Submodel

There are situations when the Weibull distribution is too complicated for modelling of given data. The problem is that there can be numerical difficulties with estimation of shape parameter τ of the Weibull distribution. However, if $\tau = 1$, then the model of Weibull distribution can be reduced to the exponential submodel where all the calculations are much easier. To assess suitability of replacement of the censored Weibull distribution with the exponential distribution, asymptotic tests with nuisance parameters can be used [16], specifically the Lagrange multiplier (LM) test, the Wald (W) test and the likelihood ratio (LR) test. Next, particular test statistics will be derived and their power functions will be analyzed using simulations.

3.1 Asymptotic Tests with Nuisance Parameters

The null hypothesis H_0 is expressed as a restriction on shape parameter τ of the censored Weibull distribution. Specifically, H_0: $\tau = 1$ is set against the alternative H_1: $\tau \neq 1$, and λ is a nuisance parameter. Therefore, in case the null hypothesis is not rejected at a specified significance level α, the censored exponential distribution can be used instead of the Weibull distribution.

The test statistics are

$$\begin{aligned} LM &= \frac{U_1^2(\widetilde{\lambda}, 1)}{J_{n,22.1}(\widetilde{\lambda}, 1)}, \\ W &= (\widehat{\tau} - 1)^2 J_{n,22.1}(\widehat{\lambda}, \widehat{\tau}), \\ LR &= 2\left[l(\widehat{\lambda}, \widehat{\tau}) - l(\widetilde{\lambda}, 1)\right], \end{aligned} \tag{3}$$

where

$$U_1(\lambda,\tau) = \frac{\partial l}{\partial \tau} = \sum_{i=1}^{k} N_i \frac{d_i^\tau \ln\left(\frac{d_i}{\lambda}\right) \exp\left[-\left(\frac{d_i}{\lambda}\right)^\tau\right] - d_{i-1}^\tau \ln\left(\frac{d_{i-1}}{\lambda}\right) \exp\left[-\left(\frac{d_{i-1}}{\lambda}\right)^\tau\right]}{\lambda^\tau \left\{\exp\left[-\left(\frac{d_{i-1}}{\lambda}\right)^\tau\right] - \exp\left[-\left(\frac{d_i}{\lambda}\right)^\tau\right]\right\}} + N_0 \frac{(1 - \tau \ln \lambda)}{\tau} + \sum_{i=n-N_0+1}^{n} \ln X_{(i)} + \frac{\ln \lambda}{\lambda^\tau} \sum_{i=n-N_0+1}^{n} X_{(i)}^\tau - \frac{1}{\lambda^\tau} \sum_{i=n-N_0+1}^{n} X_{(i)}^\tau \ln X_{(i)}$$

is the score function and $J_{n,22.1}(\lambda,\tau) = n(J_{22} - J_{21}J_{11}^{-1}J_{12})$ is a transformation of the expected FIM (2). The parameters estimated under the null hypothesis are denoted by a tilde, and those estimated under the alternative are denoted by a hat. The test statistics LM, W, LR have asymptotically χ^2 distribution with one degree of freedom [16].

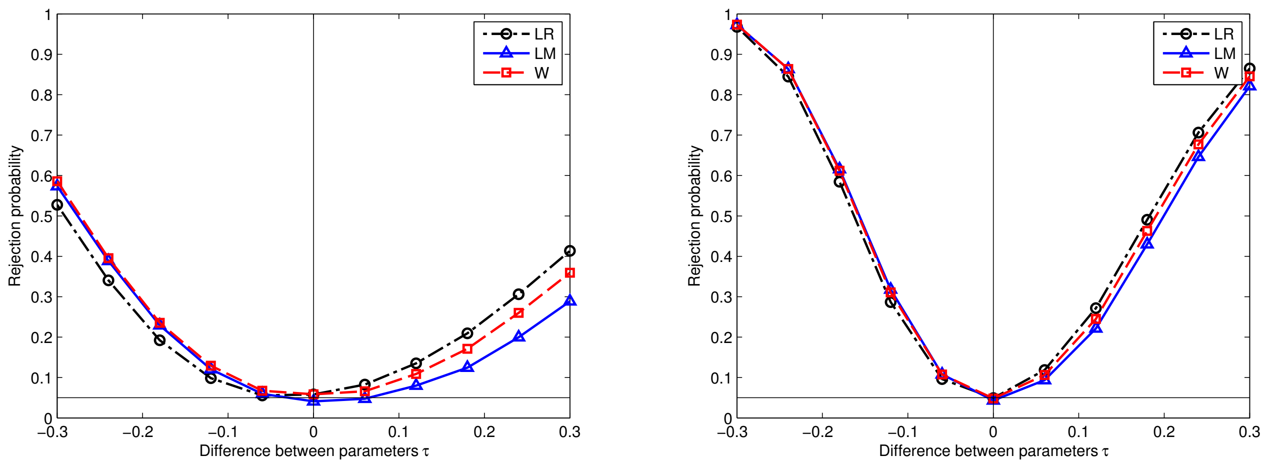
3.2 Power of the Tests

Performance of the above mentioned test statistics (3) was assessed by means of simulated power functions (10,000 repetitions). Since the majority of researchers deal with censored data with one or two detection limits, two levels of censoring were considered. Specifically, single censoring with one detection limit, and double censoring with two detection limits. The detection limits d_i, $i = 1, \ldots, k$, $k = 1, 2$, were selected as quantiles of Weibull distribution using equations $q_i = F(d_i, \lambda, \tau)$, where q_i are given in Table 1. For example, the q_1 given in column "Double" in Table 1 denotes the proportion of doubly censored observations, and describes the given censoring scheme. The censoring scheme "Low" represents the smallest proportion (10%) of censored data, and the censoring scheme "High" represents the largest proportion (90%) of censored data in case of singly and doubly censored samples. Since λ is the scale parameter, and ML estimators are scale invariant, we take $\lambda = 1$ without loss of generality. This fact was also verified using simulations. The power functions were calculated for singly and doubly left-censored samples with size $n = 10, 20, 30, 50, 100$.

Table 1: Quantiles for determination of detection limit values considering single and double censoring and various censoring schemes.

Censoring	Single	Double	
	q_1	q_1	q_2
Low	0.10	0.05	0.10
Medium	0.50	0.25	0.50
High	0.90	0.45	0.90

In case of double censoring, all the test statistics perform poorly for small sample sizes ($n < 30$; not shown in figures). Furthermore, all the test statistics perform very similarly (see Fig. 1) for sample size $n = 100$. Overall, LR test statistic has the highest and LM test statistic the lowest power. When comparing the power functions considering various censoring schemes, Fig. 2 shows what difference between power functions can be expected in case of LR test statistic. Similar behavior was observed for test statistics LM and W (not shown in figures).



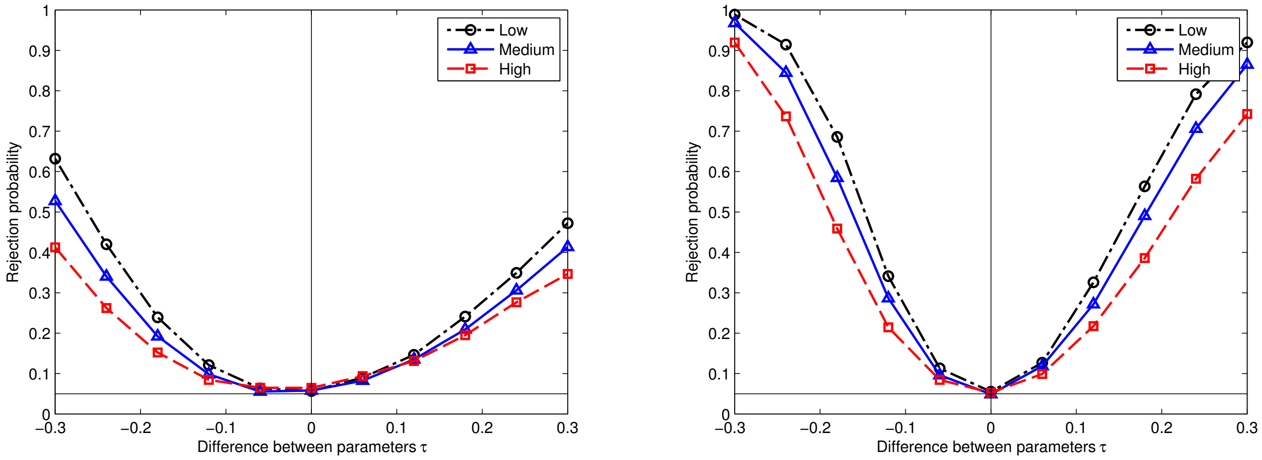
(a) $n = 30$ (b) $n = 100$

Figure 1: Power functions for test statistics (3), double censoring and medium number of censored values.

In case of single censoring, all the test statistics perform poorly for small sample sizes ($n < 30$; not shown in figures). Moreover, when the number of censored values is high, all the tests are practically unusable even for sample size $n = 100$ (see Fig. 3). All the test statistics perform very similarly (see Fig. 4) for sample size $n = 100$ and low/medium number of censored values. Furthermore, LR test statistic has the highest and LM test statistic the lowest power.

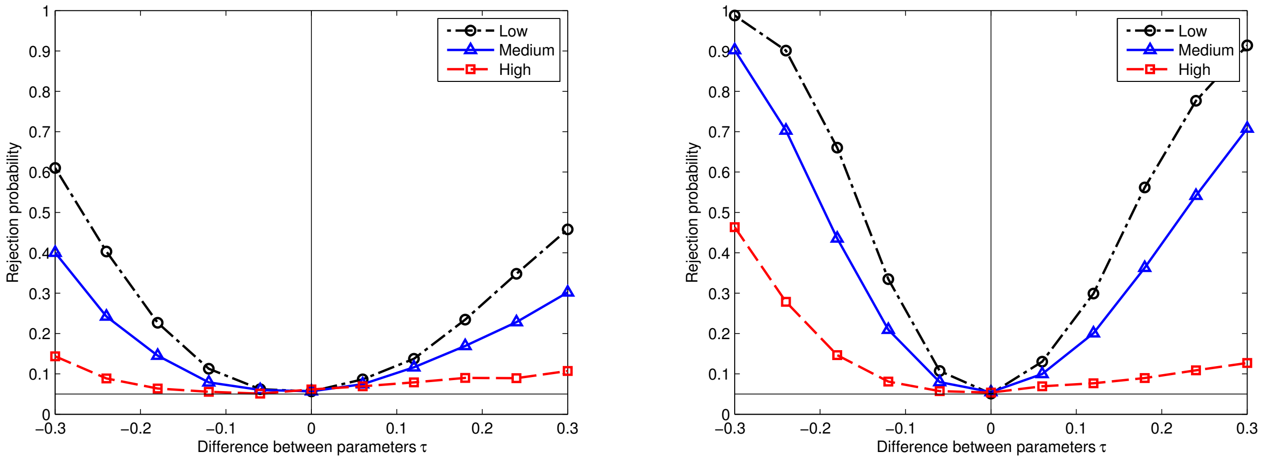
4 Conclusion

This paper dealt with statistical tests for assessing possibility of replacement of the censored Weibull distribution with the exponential submodel in order to obtain numerically less complicated and still sufficient model for describing censored data. For estimation of parameters of the censored Weibull distribution, method of maximum



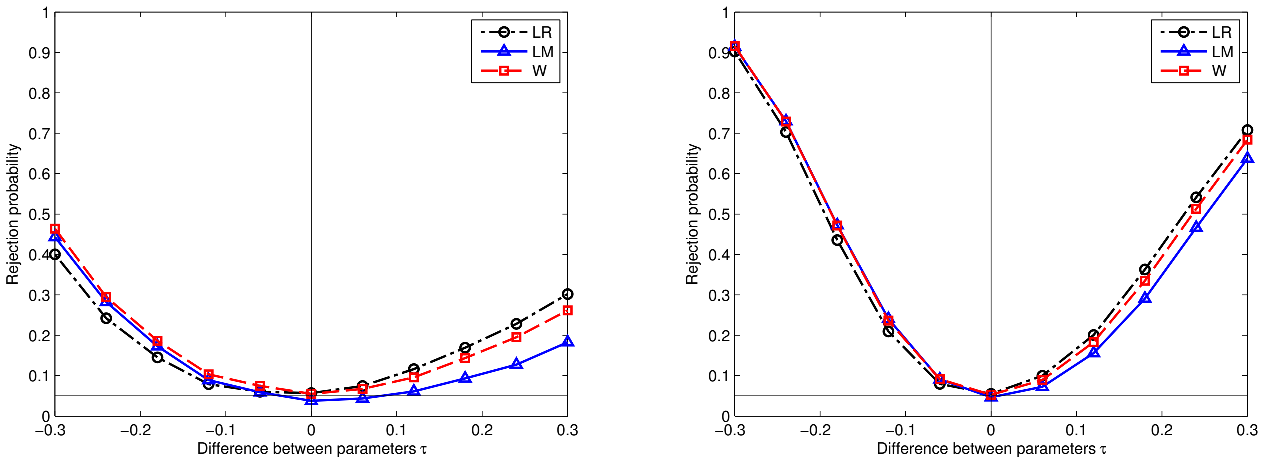
(a) $n = 30$ (b) $n = 100$

Figure 2: Power functions for test statistic LR, double censoring and a various number of censored values.



(a) $n = 30$ (b) $n = 100$

Figure 3: Power functions for test statistic LR, single censoring and a various number of censored values.



(a) $n = 30$ (b) $n = 100$

Figure 4: Power functions for test statistics (3), single censoring and medium number of censored values.

likelihood was used. Moreover, the expected Fisher information matrix for calculating variances of the estimated parameters was presented. Three test statistics for testing suitability of reduction of the Weibull distribution to the exponential distribution were described and their performance was assessed using simulations.

It was found out that performance of all the test statistics is very similar for sample size $n = 100$ considering single and double censoring. Besides that, LR test statistic has the highest power. However, in case of single censoring and a high number of censored values, all the tests perform poorly and are virtually unusable. For example, when $\tau = 1.3$, the rejection probability of the null hypothesis ($\tau = 1$) is 0.13 considering LR test statistic and sample size $n = 100$. All the procedures used were implemented in Matlab environment (version R2015a), and are available upon request.

Acknowledgement: The paper was written with the support of the specific research project FEKT-S-17-4225 (Brno University of Technology).

References

[1] Aboueissa, A.E.-M.A., Stoline, M.R.: Maximum likelihood estimators of population parameters from doubly left-censored samples. Environmetrics **17**, 811–826 (2006).

[2] Aboueissa, A.E.-M.A.: Maximum likelihood estimators of population parameters from multiply censored samples. Environmetrics **20**, 312–330 (2009)

[3] Antweiler, R.C., Taylor, H.E.: Evaluation of statistical treatments of left-censored environmental data using coincident uncensored data sets: I. Summary statistics. Environmental Science & Technology **42**, 3732–3738 (2008)

[4] Antweiler, R.C.: Evaluation of statistical treatments of left-censored environmental data using coincident uncensored data sets: II. Group comparisons. Environmental Science & Technology **49**, 13439–13446 (2015)

[5] Cohen, A.C.: *Truncated and Censored Samples*, Marcel Dekker, New York (1991).

[6] Fusek, M.: *Extreme Value Distributions with Applications*, doctoral thesis (in Czech), Brno University of Technology, Brno (2013).

[7] Fusek, M., Michálek, J.: Statistical methods for analyzing musk compounds concentration based on doubly left-censored samples. International Journal of Mathematical Models and Methods in Applied Sciences **7**(8), 755–763 (2013).

[8] Fusek, M., Michálek, J.: Statistical analysis of type I multiply left-censored samples from exponential distribution. Journal of Statistical Computation and Simulation **85**, 2148–2163 (2015).

[9] Fusek, M., Michálek, J., Vávrová, M.: Evaluation of contamination data with non-detects using censored distributions. Fresenius Environmental Bulletin **24**(11c), 4165–4172 (2015).

[10] Helsel, D.R.: *Statistics for Censored Environmental Data using Minitab and R*, John Wiley and Sons, New York (2012).

[11] Klein, J.P., Moeschberger, M.L.: *Survival Analysis: Techniques for Censored and Truncated Data*, second edn. Springer, New York (2005).

[12] Kotb, M.S., Raqab, M.Z.: Inference and prediction for modified Weibull distribution based on doubly censored samples. Mathematics and Computers in Simulation 132, 195–207 (2017).

[13] Krishnamoorthy, K., Mathew, T., Xu, Z.: Comparison of means of two lognormal distributions based on samples with multiple detection limits. Journal of Occupational and Environmental Hygiene **11**, 538–546 (2014).

[14] Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.E.: Convergence properties of the Nelder–Mead simplex method in low dimensions. SIAM Journal on Optimization **9**, 112–147 (1998).

[15] Lehmann, E.L., Casella, G.: *Theory of Point Estimation*, Springer-Verlag, New York (1998).

[16] Lehmann, E.L., Romano, J.P.: *Testing Statistical Hypotheses*, Springer, New York (2005).

[17] Shoari, N., Dubé, J.-S., Chenouri, S.: Estimating the mean and standard deviation of environmental data with below detection limit observations: Considering highly skewed data and model misspecification. Chemosphere **138**, 599–608 (2015).

[18] Tekindal, M.A., Erdoğan, B.D., Yavuz, Y.: Evaluating left-censored data through substitution, parametric, semi-parametric, and nonparametric methods: A simulation study. Interdisciplinary Sciences: Computational Life Sciences **9**, 153–172 (2017).