

LEFT-CENSORED SAMPLES FROM SKEWED DISTRIBUTIONS: STATISTICAL INFERENCE AND APPLICATIONS

Michal Fusek¹, Jaroslav Michálek²

¹Department of Mathematics, Faculty of Electrical Engineering and Communication, Brno University of Technology, Technická 2848/8, 616 00 Brno, Czech Republic

²Department of Econometrics, Faculty of Military Leadership, University of Defence, Šumavská 4, 662 10 Brno, Czech Republic

Abstract

FUSEK MICHAL, MICHÁLEK JAROSLAV. 2018. Left-Censored Samples from Skewed Distributions: Statistical Inference and Applications. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 66(1): 0245–0252.

Left-censored data occur frequently in many areas. At present, researchers pay attention to skewed censored distributions more frequently. This paper deals with statistical inference of type I multiply left-censored Weibull and exponential distributions. It suggests a computational procedure for calculation of maximum likelihood estimates of the parameters. The expected Fisher information matrix for estimation of variances of estimated parameters is introduced. The estimates are then used for construction of confidence intervals for the expectation using the maximum likelihood method. Asymptotic tests for comparison of distributions (expectations respectively) of two independent left-censored Weibull samples are proposed. Furthermore, asymptotic tests for assessing suitability of reduction of the Weibull distribution to the exponential distribution are introduced. Finally, the left-censored exponential distribution is briefly described. Methods derived in this paper are illustrated on elemental carbon measurements, and can be applied in analysis of real environmental and/or chemical data.

Keywords: asymptotic tests, Fisher information matrix, left-censored data, maximum likelihood, skewed distribution

INTRODUCTION

When analyzing real environmental data, it is often necessary to deal with a situation in which the attribute being measured (for example the concentration) falls below detection limit d of the measuring instrument. Such measurements are called left-censored (El-Shaarawi and Piegorsch, 2012). In case the detection limit is fixed, we talk about type I censoring, and the number of censored experimental units is a random variable. In practice, more detection limits $d_1 < \dots < d_k$, $k > 1$, are often considered, and we talk about multiply left-censored samples (Aboueissa, 2009; Aboueissa and Stoline, 2006; Cohen, 1991; El-Shaarawi and Naderi, 1991; Fusek and Michálek, 2015b). In such a case, only observations above the highest detection limit d_k and the number of observations

under the remaining detection limits are known. Often, $k = 2$; in such cases, we talk about doubly left-censored data (Fusek and Michálek, 2012a, 2012b). Standard statistical methods for analyzing left-censored data are based on the assumption that the measured attribute is a normally distributed random variable. This approach is well described in El-Shaarawi and Dolan (1989), and El-Shaarawi and Naderi (1991). In practice, various ad hoc methods where the censored observations are replaced by an appropriate constant lying between the particular detection limits (Helsel, 2012) are very popular. This approach, which is frequent in many environmental studies, is critically reviewed in El-Shaarawi and Esterby (1992), and Helsel (2006). When the left-censored samples are not normally distributed, various data normalization methods can

be used, for example, a logarithmic transformation in case of log-normal data, or the Box-Cox transformation. Nevertheless, in many cases it is impossible to find a suitable transformation function. In some cases, it is also possible to convert left-censored data to right-censored data, and use nonparametric methods for right-censored samples like the Kaplan-Meier estimator (Shumway *et al.*, 2002). However, the parametric approach to analysis of left-censored data is currently preferred (Aboueissa and Stoline, 2006; Helsel, 2012). It is based on the assumption that the censored sample has a known distribution. In order to estimate unknown parameters of the distribution, the maximum likelihood (ML) method (Barndorff-Nielsen and Cox, 1994; Lehmann and Casella, 1998; Lehmann and Romano, 2005), and the approach based on the regression on order statistics (Helsel, 2012; Shoari *et al.*, 2015) are often used. When analyzing skewed left-censored samples, the log-normal, the Weibull, the exponential, the generalized exponential, and the gamma distributions are usually used as model distributions (Fusek and Michálek, 2012b, 2015b; Gupta and Kundu, 1999; Helsel, 2012; Hogg *et al.*, 2005; Mitra and Kundu, 2008; Schmoyer *et al.*, 1996).

In this paper, attention will be paid to applying the ML method to left-censored samples with skewed distributions. This topic is only partially accounted for in literature (e.g., Mitra and Kundu, 2008; Shoari *et al.*, 2015). On one hand, the ML approach can certainly be used when data are not normally distributed, and the sample size is large enough. Then, under certain general regularity conditions, the distribution of the estimates can be approximated by the normal distribution. On the other hand, likelihood equations can be nonlinear, rather complicated, and their analytic solution is often impossible to find. In such a case, numerical calculations are necessary, and researchers can use a suitable software package. Well elaborated statistical methods for censored environmental data in Minitab and R can be found in Helsel (2012). Nevertheless, in many cases, it is necessary to assess properties of ML estimates for small sample sizes, to estimate variability of these estimates, to test hypotheses on parameters of particular distributions and/or to construct confidence intervals for various parameters. This requires determination of the Fisher information matrix (FIM), which is necessary for description of asymptotic variability of obtained estimates. Asymptotic properties of these estimates and their variability for small sample sizes can be assessed using simulations.

This contribution focuses on reviewing available methods for type I left-censored samples from skewed distributions, specifically from the exponential and the Weibull distributions. Authors prefer the Weibull distribution because of its flexibility and ability to adapt to real data. In Shoari *et al.* (2015), there are critical comments

about low robustness of the Weibull distribution considering a misspecification of the model distribution, especially for high values of skewness γ . However, simulation results showed that when the sample size is small, and skewness of the distribution is high ($\gamma > 2$), there can be numerical difficulties with parameters estimates, which can give an impression of low robustness of the Weibull distribution to the model misspecification. In the next section, the computational procedure for determination of ML estimates of parameters of the Weibull distribution is presented, and the analytical form of the expected FIM is introduced. The asymptotic confidence intervals for the expectation of the left-censored Weibull distribution based on the ML estimates will be derived. After that, a two-sample censored Weibull model will be introduced. Moreover, for comparison of two independent multiply left-censored samples from the Weibull distribution, three asymptotic tests with nuisance parameters (the Lagrange multiplier test, the Wald's test, the likelihood ratio test) will be proposed. Furthermore, methods for comparison of expectations of two independent left-censored samples from the Weibull distribution will be presented. Next, asymptotic tests for testing the suitability of using the exponential distribution instead of the Weibull distribution will be introduced, and the censored exponential distribution will be briefly described. The proposed methods are illustrated on a comparison of two data sets of elemental carbon measurements.

MATERIALS AND METHODS

One-Sample Censored Weibull Model

Let X_1, \dots, X_n be a random sample from the Weibull distribution with scale parameter $\lambda > 0$, shape parameter $\tau > 0$, cumulative distribution function (CDF)

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

and probability density function (PDF)

$$f(x, \lambda, \tau) = \begin{cases} \frac{\tau}{\lambda^\tau} x^{\tau-1} \exp\left[-\left(\frac{x}{\lambda}\right)^\tau\right] & \text{for } x \geq 0, \\ 0 & \text{for } x < 0. \end{cases} \quad (2)$$

Attention will be also paid to characteristics of the Weibull distribution, specifically the expected value

$$\mu(\lambda, \tau) = \lambda \Gamma\left(1 + \frac{1}{\tau}\right), \quad (3)$$

and the skewness

$$\gamma(\lambda, \tau) = \frac{2\Gamma^3\left(1+\frac{1}{\tau}\right) - 3\Gamma\left(1+\frac{1}{\tau}\right)\Gamma\left(1+\frac{2}{\tau}\right) + \Gamma\left(1+\frac{3}{\tau}\right)}{\left[\Gamma\left(1+\frac{2}{\tau}\right) - \Gamma^2\left(1+\frac{1}{\tau}\right)\right]^{3/2}}, \quad (4)$$

where Γ is the gamma function. Furthermore, the type I multiply left-censored ordered sample of X_1, \dots, X_n will be denoted by $X_{(1)}, \dots, X_{(n)}$. The detection limits d_1, \dots, d_k will be considered and we put $d_0 = 0$. Moreover, N_i is the number of observations in the interval (d_{i-1}, d_i) , and N_0 is the number of uncensored observations $X_{(n-N_0+1)}, \dots, X_{(n)}$. The log-likelihood function of the censored sample is given by (Cohen, 1991)

$$\begin{aligned} l(\lambda, \tau, N_0, \dots, N_k, X_{(n-N_0+1)}, \dots, X_{(n)}) \\ = \log\left(\frac{n!}{N_1! \dots N_k!}\right) + \sum_{i=n-N_0+1}^n \log[f(X_{(i)})] \\ + \sum_{i=1}^k N_i \log[F(d_i, \lambda, \tau) - F(d_{i-1}, \lambda, \tau)], \end{aligned} \quad (5)$$

and for $N_0 = 0$ we put $\sum_{i=n-N_0+1}^n \log[f(X_{(i)})] = 0$.

The standard approach to estimation of parameters λ, τ is to derive likelihood equations, and solve them using the Newton-Raphson method. However, there are numerical difficulties with obtaining the solution for low values of the shape parameter τ , i.e. for high values of skewness (4). Therefore, ML estimates $\hat{\lambda}, \hat{\tau}$ of parameters λ, τ were obtained by maximizing the log-likelihood function (5). This is a complex optimization problem which has to be solved numerically. In our case, the Nelder-Mead simplex algorithm (Lagarias *et al.*, 1998) in Matlab (version R2015a) was applied. When using this type of algorithms, it is necessary to select initial values of parameters that need to be estimated. Starting values were selected using the moment estimator (Anděl, 2005) of parameters of the Weibull distribution based on samples in which the censored observations were replaced by a constant lying between the detection limits.

According to Barndorff-Nielsen and Cox (1994), the sample FIM $\tilde{\mathbf{J}}_n$, which is an unbiased estimator of the expected FIM

$$\mathbf{J}_n(\lambda, \tau) = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} = \mathbf{E} \tilde{\mathbf{J}}_n = \begin{bmatrix} \mathbf{E} \tilde{J}_{11} & \mathbf{E} \tilde{J}_{12} \\ \mathbf{E} \tilde{J}_{21} & \mathbf{E} \tilde{J}_{22} \end{bmatrix}, \quad (6)$$

can be calculated using formula

$$\tilde{\mathbf{J}}_n(\lambda, \tau) = \begin{bmatrix} \tilde{J}_{11} & \tilde{J}_{12} \\ \tilde{J}_{21} & \tilde{J}_{22} \end{bmatrix} = \begin{bmatrix} -\frac{\partial^2 l}{\partial \lambda^2} & -\frac{\partial^2 l}{\partial \lambda \partial \tau} \\ -\frac{\partial^2 l}{\partial \tau \partial \lambda} & -\frac{\partial^2 l}{\partial \tau^2} \end{bmatrix}.$$

Since $\tilde{\mathbf{J}}_n(\lambda, \tau) \rightarrow \mathbf{J}_n(\lambda, \tau)$ in probability for $n \rightarrow \infty$, and the determination of the expected FIM can be unnecessarily complicated, the sample FIM is often

used (Aboueissa, 2009) instead of the expected FIM. One major disadvantage of this approach is the rather extensive variability of the sample FIM. Therefore, the expected FIM \mathbf{J}_n will be preferred. Derivation of the exact form of the expected FIM can be found in Fusek (2013). Considering (under mild regularity conditions) the asymptotic properties of the ML estimator $(\hat{\lambda}, \hat{\tau})$ of the parameter vector (λ, τ) , it has, according to Lehmann and Casella (1998), asymptotically normal distribution with the expectation (λ, τ) , and the asymptotic variance matrix

$$\mathbf{J}_n^{-1}(\lambda, \tau) = \begin{bmatrix} J^{11} & J^{12} \\ J^{21} & J^{22} \end{bmatrix}$$

Therefore,

$$\sqrt{n}(\hat{\lambda} - \lambda) \xrightarrow{A} N(0, J^{11}), \quad \text{and} \quad \sqrt{n}(\hat{\tau} - \tau) \xrightarrow{A} N(0, J^{22}).$$

The bias of various estimators of variance of the shape parameter τ is analyzed in Fusek and Michálek (2012b) for doubly left-censored samples. It was shown on simulated data that bias of the asymptotic variance estimator obtained using the expected FIM \mathbf{J}_n is lower than that obtained using the sample FIM $\tilde{\mathbf{J}}_n$ even for small sample sizes. Therefore, the expected FIM is preferable.

Confidence Intervals for Expectation of the Left-Censored Weibull Distribution

When analyzing environmental data, researchers are often interested in estimation of the expected value of an observed variable. Therefore, the point estimate and the asymptotic confidence interval for the expectation (3) of the left-censored Weibull distribution will be introduced. Moreover, it can be used for testing a hypothesis on the expected value of the measured variable.

Using asymptotic properties of the ML estimates, it can be shown that the distribution of the estimate $\mu(\hat{\lambda}, \hat{\tau})$ of the expected value $\mu(\lambda, \tau)$ is under regularity conditions asymptotically normal with the expectation $\mu(\lambda, \tau)$. The asymptotic variance can be obtained by the delta method (Casella and Berger, 2002) using derivatives of the parametric function $\mu(\lambda, \tau)$, and is of the form

$$\begin{aligned} \text{Var}(\mu(\hat{\lambda}, \hat{\tau})) &= \\ &= \begin{bmatrix} \Gamma\left(1+\frac{1}{\tau}\right), -\frac{\lambda \Psi\left(1+\frac{1}{\tau}\right) \Gamma\left(1+\frac{1}{\tau}\right)}{\tau^2} \end{bmatrix} \\ &\quad \mathbf{J}_n^{-1}(\lambda, \tau) \begin{bmatrix} \Gamma\left(1+\frac{1}{\tau}\right), -\frac{\lambda \Psi\left(1+\frac{1}{\tau}\right) \Gamma\left(1+\frac{1}{\tau}\right)}{\tau^2} \end{bmatrix}^T, \end{aligned} \quad (7)$$

where Ψ is the digamma function, and \mathbf{a}^T denotes the transpose of vector \mathbf{a} . Since the distribution of the estimate $\mu(\hat{\lambda}, \hat{\tau})$ is asymptotically normal,

the asymptotic $100(1-\alpha)\%$ confidence interval for $\mu(\lambda, \tau)$ can be easily obtained from (3) and (7), and the lower (LCL) and upper (UCL) $(1-\alpha)\%$ confidence limits for estimate $\hat{\mu}$ can be calculated as

$$\text{UCL}(\hat{\mu}) = \mu(\hat{\lambda}, \hat{\tau}) + z_{1-\frac{\alpha}{2}} \sqrt{\text{Var}(\mu(\hat{\lambda}, \hat{\tau}))},$$

$$\text{LCL}(\hat{\mu}) = \mu(\hat{\lambda}, \hat{\tau}) - z_{1-\frac{\alpha}{2}} \sqrt{\text{Var}(\mu(\hat{\lambda}, \hat{\tau}))},$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution. It is always difficult to assess how well the sample statistic estimates the underlying population value, what its bias is, and what the coverage probabilities of confidence intervals are. For this purpose, a simulation study comparing a) estimates of the expected value, b) the ML-based and bootstrap-based (Efron, 1979) confidence intervals, was carried out in Fusek and Michálek (2016). It was found out that the expected values estimated using the ML and the bootstrap methods are very similar in almost all cases. However, in case of a small sample size, a high censoring level and $\gamma > 0$, ML estimates of the expectation were the best. Despite the fact that in case of a small sample size and a high censoring level, the bootstrap-based confidence intervals were much narrower than the ML-based ones, it was found out that ML-based confidence intervals have better coverage probabilities in case of high censoring. More details on this topic can be found in Fusek and Michálek (2016).

Two-Sample Censored Weibull Model

Let $X_{j,1}, \dots, X_{j,n_j}$, $j = 1, 2$, be two independent censored samples from the Weibull distribution with CDF (1), PDF (2), and parameters $\lambda_1 = \lambda$, $\tau_1 = \tau$ in case of the first sample ($j = 1$), and $\lambda_2 = \lambda + \alpha$, $\tau_2 = \tau + \beta$ in case of the second sample ($j = 2$). Furthermore, the ordered sample of $X_{j,1}, \dots, X_{j,n_j}$ is denoted by $X_{j(1)}, \dots, X_{j(n_j)}$, and frequencies N_{ji} correspond to frequencies N_i , $i = 0, 1, \dots, k$, from previous sections. The log-likelihood function of the two joint censored samples can be formulated using (5), and is of the form

$$l_R(\alpha, \beta, \lambda, \tau) = l(\lambda, \tau, N_{1,0}, \dots, N_{1,k}, X_{1,(n-N_{1,0}+1)}, \dots, X_{1,(n)}) + l(\lambda + \alpha, \tau + \beta, N_{2,0}, \dots, N_{2,k}, X_{2,(n-N_{2,0}+1)}, \dots, X_{2,(n)}). \quad (8)$$

The maximum likelihood estimates $\hat{\alpha}, \hat{\beta}, \hat{\lambda}, \hat{\tau}$ of parameters $\alpha, \beta, \lambda, \tau$ can be obtained by maximization of the log-likelihood function (8). The Nelder-Mead simplex algorithm can be applied again.

Using the FIM for the one-sample Weibull model, we get the expected FIM for the two-sample Weibull model in the form

$$J_n^R(\alpha, \beta, \lambda, \tau) = \begin{pmatrix} J_{11}^R & J_{12}^R & J_{13}^R & J_{14}^R \\ J_{21}^R & J_{22}^R & J_{23}^R & J_{24}^R \\ J_{31}^R & J_{32}^R & J_{33}^R & J_{34}^R \\ J_{41}^R & J_{42}^R & J_{43}^R & J_{44}^R \end{pmatrix} = \begin{pmatrix} -E \frac{\partial^2 l_R}{\partial \alpha^2} & -E \frac{\partial^2 l_R}{\partial \alpha \partial \beta} & -E \frac{\partial^2 l_R}{\partial \alpha \partial \lambda} & -E \frac{\partial^2 l_R}{\partial \alpha \partial \tau} \\ -E \frac{\partial^2 l_R}{\partial \beta \partial \alpha} & -E \frac{\partial^2 l_R}{\partial \beta^2} & -E \frac{\partial^2 l_R}{\partial \beta \partial \lambda} & -E \frac{\partial^2 l_R}{\partial \beta \partial \tau} \\ -E \frac{\partial^2 l_R}{\partial \lambda \partial \alpha} & -E \frac{\partial^2 l_R}{\partial \lambda \partial \beta} & -E \frac{\partial^2 l_R}{\partial \lambda^2} & -E \frac{\partial^2 l_R}{\partial \lambda \partial \tau} \\ -E \frac{\partial^2 l_R}{\partial \tau \partial \alpha} & -E \frac{\partial^2 l_R}{\partial \tau \partial \beta} & -E \frac{\partial^2 l_R}{\partial \tau \partial \lambda} & -E \frac{\partial^2 l_R}{\partial \tau^2} \end{pmatrix}$$

and the analytical form of elements of the expected FIM $J_n^R(\alpha, \beta, \lambda, \tau)$ can be found in Fusek (2013).

Comparison of Two Left-Censored Weibull Samples

In order to compare two independent multiply left-censored samples from the Weibull distribution, asymptotic tests with nuisance parameters can be used (Lehmann and Romano, 2005), specifically the Lagrange multiplier test (LM), the Wald's test (W) and the likelihood ratio test (LR). The null hypothesis H_0 is that distributions of both samples are equal. Thus $H_0: (\alpha, \beta)^T = (0, 0)^T$ is set against the alternative $H_1: (\alpha, \beta)^T \neq (0, 0)^T$ and λ, τ are nuisance parameters. The test statistics are of the form (Anděl, 2005)

$$\begin{aligned} LM &= U_1(0, 0, \tilde{\lambda}, \tilde{\tau}) \left[J_{n,11.2}^R(0, 0, \tilde{\lambda}, \tilde{\tau}) \right]^{-1} U_1^T(0, 0, \tilde{\lambda}, \tilde{\tau}), \\ W &= (\hat{\alpha}, \hat{\beta}) \left[J_{n,11.2}^R(\hat{\alpha}, \hat{\beta}, \hat{\lambda}, \hat{\tau}) \right]^{-1} (\hat{\alpha}, \hat{\beta})^T, \\ LR &= 2 \left[l_R(\hat{\alpha}, \hat{\beta}, \hat{\lambda}, \hat{\tau}) - l_R(0, 0, \tilde{\lambda}, \tilde{\tau}) \right], \end{aligned} \quad (10)$$

where $U_1(\alpha, \beta, \lambda, \tau) = (\partial l_R / \partial \alpha, \partial l_R / \partial \beta)$ is the score function, and

$$J_{n,11.2}^R(\alpha, \beta, \lambda, \tau) = \begin{pmatrix} J_{11}^R & J_{12}^R \\ J_{21}^R & J_{22}^R \end{pmatrix} - \begin{pmatrix} J_{13}^R & J_{14}^R \\ J_{23}^R & J_{24}^R \end{pmatrix} \begin{pmatrix} J_{33}^R & J_{34}^R \\ J_{43}^R & J_{44}^R \end{pmatrix}^{-1} \begin{pmatrix} J_{31}^R & J_{32}^R \\ J_{41}^R & J_{42}^R \end{pmatrix}$$

is the element of the blockwise inversion of the expected FIM (9). 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 two degrees of freedom (Lehmann and Romano, 2005). More information about the performance of test statistics (10) together with their simulated power functions can be found in Fusek and Michálek (2014).

Comparison of Expectations of Two Left-Censored Weibull Samples

For comparison of expectations $\mu_1(\lambda_1, \tau_1)$ and $\mu_2(\lambda_2, \tau_2)$ of two independent multiply left-censored samples from the Weibull distribution, the asymptotic test statistic

$$W_E = \frac{\mu(\hat{\lambda}_1, \hat{\tau}_1) - \mu(\hat{\lambda}_2, \hat{\tau}_2)}{\sqrt{\widehat{\text{Var}}(\mu(\hat{\lambda}_1, \hat{\tau}_1)) + \widehat{\text{Var}}(\mu(\hat{\lambda}_2, \hat{\tau}_2))}}$$

can be used. The expectation $\mu(\lambda, \tau)$ is given by (3), and the variance $\widehat{\text{Var}}(\mu(\hat{\lambda}, \hat{\tau}))$ is given by (7) where all the unknown parameters have to be replaced by their ML estimates. Under the null hypothesis $H_0: \mu_1 = \mu_2$, the statistic W_E has asymptotically normal distribution $N(0,1)$. Power of the test based on the test statistic W_E can be found in Fusek and Michálek (2015a).

Reduction of the Weibull Model to the Exponential Submodel

There are situations when the censored Weibull distribution is too complicated for description of the measured variable, and the exponential distribution would be sufficient. In such a situation, asymptotic tests with nuisance parameters can be used. The null hypothesis is that the Weibull distribution can be simplified to the exponential distribution, specifically $H_0: \tau = 1$ against the alternative $H_1: \tau \neq 1$. The test statistics have under the null hypothesis asymptotically χ^2 distribution with one degree of freedom (Lehmann and Romano, 2005), and are of the form

$$\begin{aligned} LM_{exp} &= \frac{U_1^2(\tilde{\lambda}, 1)}{J_{n,22.1}(\tilde{\lambda}, 1)}, \\ W_{exp} &= (\tilde{\tau} - 1)^2 J_{n,22.1}(\hat{\lambda}, \hat{\tau}), \\ LR_{exp} &= 2 \left[l(\hat{\lambda}, \hat{\tau}) - l(\hat{\lambda}, 1) \right], \end{aligned}$$

where $l(\lambda, \tau)$ is the log-likelihood function (5). Moreover, $J_{n,22.1}(\lambda, \tau) = n(J_{22} - J_{21}J_{11}^{-1}J_{12})$ is a transformation of the FIM (6) and $U_1(\lambda, \tau) = \partial l(\lambda, \tau) / \partial \tau$ is the score function. The parameters estimated under the null hypothesis are again denoted by a tilde, and those estimated under the alternative are denoted by a hat. More details on this topic including powers of particular tests can be found in Fusek and Michálek (2017). An application of this method on real data can be found in Fusek and Michálek (2013).

One-Sample Censored Exponential Model

In this section, attention will be paid to a description of the left-censored exponential distribution, which is a special case of the Weibull distribution for $\tau = 1$. The exponential distribution has skewness $\gamma = 2$, and can play an important role when dealing with skewed distributions because of its simple analytical form. Let us have a type I multiply left-censored random sample X_1, \dots, X_n

from the exponential distribution with scale parameter λ and CDF

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

The ML estimate of parameter λ can be calculated by maximizing the log-likelihood function (5) considering $\tau = 1$. The sample FIM is given by

$$\tilde{J}(\lambda) = -\frac{\partial^2 l}{\partial \lambda^2} \quad (11)$$

and the expected FIM is given by

$$J(\lambda) = E\tilde{J}(\lambda). \quad (12)$$

The derivation of (11) and (12) together with their exact analytical forms can be found in Fusek and Michálek (2015b). In order to study properties of the ML estimate of λ and its variance, simulation study was carried out. The asymptotic variance was compared with its estimates $\sigma^2(\hat{\lambda}) = J^{-1}(\hat{\lambda})$ and $\hat{\sigma}^2(\hat{\lambda}) = \tilde{J}^{-1}(\hat{\lambda})$. Considering the estimates $\sigma^2(\hat{\lambda})$ and $\hat{\sigma}^2(\hat{\lambda})$, 95% confidence intervals for parameter λ were calculated, and their coverage probabilities were determined. It was found out that for sample sizes $n \geq 30$ and less than 50% of censored values, a satisfactory statistical inference about parameter λ is possible. In case the number of censored values is high ($\geq 50\%$), the coverage probabilities for sample sizes $n \leq 30$ based on $\sigma^2(\hat{\lambda})$ are about 1% lower than the specified 95% confidence limit, and those based on $\hat{\sigma}^2(\hat{\lambda})$ are about 2% lower than the specified 95% confidence limit. More results and details of the simulations can be found in Fusek and Michálek (2015b).

Similar methods for comparison of distributions (expectations respectively) of two independent left-censored samples from the exponential distribution were derived, and can be found in Fusek (2013) and Fusek and Michálek (2013).

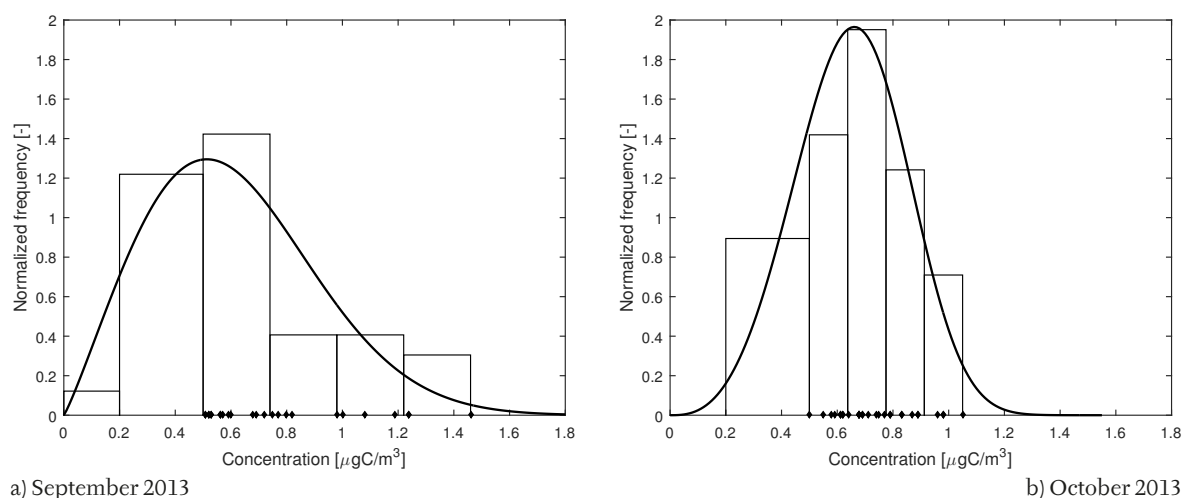
RESULTS

The proposed methods will now be illustrated on real data. Let us have two data samples of elemental carbon concentration measurements from the atmospheric station Křešín u Pacova, Czech Republic (Dvorská *et al.*, 2015). Each of the samples contains 41 measurements from one week in September (first sample) and October 2013 (second sample). The measurements are doubly left-censored with detection limits $d_1 = 0.2$ and $d_2 = 0.5$. There are $N_{1,0} = 25$ uncensored measurements, and $N_{1,1} = 1$, $N_{1,2} = 15$ in case of the first sample, and there are $N_{2,0} = 30$ uncensored measurements, and $N_{2,1} = 0$, $N_{2,2} = 11$ in case of the second sample.

Considering that both samples are independent and have the Weibull distribution with parameters $\lambda_j, \tau_j, j = 1, 2$, the goal is to test the hypothesis that both samples have the same distribution, and

subsequently test the hypothesis that both samples have the same expected value. The assumption of the Weibull distribution was confirmed using Q-Q plots and the χ^2 goodness-of-fit test. Using the ML method, estimates of parameters λ_j , τ_j , expected values $\mu(\lambda_j, \tau_j)$ and their variances $\sigma_j^2 = \text{Var}(\mu(\hat{\lambda}_j, \hat{\tau}_j))$, $j = 1, 2$, can be calculated. The obtained estimates are $\lambda_1 = 0.691$, $\tau_1 = 2.131$, $\mu_1 = 0.611$, $\sigma_1^2 = 0.302$ in case of the first sample, and $\lambda_2 = 0.720$, $\tau_2 = 3.695$, $\mu_2 = 0.650$, $\sigma_2^2 = 0.196$ in case of the second sample. The 95% confidence intervals for expectations are (0.019, 1.203) in case of the first sample, and (0.266, 1.035) in case of the second sample.

In order to test the hypothesis $H_0: \lambda_1 = \lambda_2$ and $\tau_1 = \tau_2$, test statistic (10) can be used. The obtained p -values are $p = 0.015$ for the Lagrange multiplier test, $p = 0.001$ for the likelihood ratio test, and $p = 0.016$ for the Wald's test. Since the null hypothesis is rejected at the significance level of 0.05 in all cases, the distributions of both samples are not identical. Moreover, the hypothesis of equality of expected values of both samples was tested, and $W_E = -0.704$ with p -value $p = 0.481$ were calculated. Therefore, the difference in expectation of both samples has not been proven. Histograms of both samples together with the estimated PDF of the Weibull distribution are in Fig. 1.



1: Histograms of elemental carbon measurements with the Weibull density. Measured values are displayed on the x-axis.

CONCLUSION

This paper focused on the statistical inference of skewed type I left-censored distributions, specifically the exponential and the Weibull distributions. The computational procedure for calculation of ML estimates of parameters of the Weibull distribution was proposed, and the analytical form of the expected FIM was introduced. The asymptotic confidence intervals for the expectation of the left-censored Weibull distribution based on ML estimates were introduced and compared with those obtained by the bootstrap methodology. It was found out that ML method outperforms the bootstrap method in terms of bias of expectations and also in terms of coverage probabilities of the particular confidence intervals in case of the high censoring and skewness. In order to compare two independent left-censored samples from the Weibull distribution, asymptotic tests with nuisance parameters were suggested. For a comparison of expectations, an asymptotic test based on the Wald's test statistic was introduced. The proposed methods were illustrated on comparison of two data sets of elemental carbon concentration measurements. Furthermore, methods for testing the suitability of replacement of the Weibull distribution with the exponential distribution were described. Finally, a model of the left-censored exponential distribution was briefly described. It was shown that using the expected FIM instead of the sample FIM can be beneficial despite the more difficult calculations. More applications of the presented methods can be found in Fusek and Michálek (2013), and Fusek *et al.* (2015), where they were used for assessing the effectiveness of the wastewater treatment plant with respect to the presence of musk compounds in fish tissue samples. All the procedures used were implemented in Matlab environment (R2015a), and are available upon request. Methods described in this paper can be used in the analysis of real environmental and/or chemical data.

Acknowledgement

The paper was written with the support of project No. MO DZRO K-110, University of Defence. The authors would like to thank the editor and the reviewers for a number of good suggestions which helped to improve the manuscript.

REFERENCES

- ABOUEISSA, A. E.-M. A. 2009. Maximum likelihood estimators of population parameters from multiply censored samples. *Environmetrics*, 20(3): 312–330.
- ABOUEISSA, A. E.-M. A. and STOLINE, M. R. 2006. Maximum likelihood estimators of population parameters from doubly left-censored samples. *Environmetrics*, 17(8): 811–826.
- ANĎĚL, J. 2005. *Introduction to mathematical statistics* [in Czech: *Základy matematické statistiky*]. Praha: Matfyzpress.
- BARNDORFF-NIELSEN, O. E. and COX, D. R. 1994. *Inference and asymptotics*. London: Chapman and Hall/CRC.
- CASELLA, G. and BERGER, R. L. 2002. *Statistical inference*. 2nd Edition. Duxbury: Thomson Learning.
- COHEN, A. C. 1991. *Truncated and censored samples*. New York: Marcel Dekker.
- DVORSKÁ, A., SEDLÁK, P., SCHWARZ, J., FUSEK, M., HANUŠ, V., VODIČKA, P. and TRUSINA, J. 2015. Atmospheric station Křešín u Pacova, Czech Republic – a Central European research infrastructure for studying greenhouse gases, aerosols and air quality. *Advances in Science and Research*, 12(1): 79–83.
- EFRON, B. 1979. Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7(1): 1–26.
- EL-SHAARAWI, A. H. and DOLAN, D. M. 1989. Maximum likelihood estimation of water concentrations from censored data. *Canadian Journal of Fisheries and Aquatic Sciences*, 46(6): 1033–1039.
- EL-SHAARAWI, A. H. and ESTERBY, S. R. 1992. Replacement of censored observations by a constant: an evaluation. *Water Research*, 26(6): 835–844.
- EL-SHAARAWI, A. H. and NADERI, A. 1991. Statistical inference from multiply censored environmental data. *Environmental Monitoring and Assessment*, 17(2–3): 339–347.
- EL-SHAARAWI, A. H. and PIEGORSCH, W. W. 2012. *Encyclopedia of environmetrics*. 2nd Edition. Chichester: John Wiley and Sons.
- FUSEK, M. 2013. *Extreme value distributions with applications*. Brno University of Technology. Unpublished doctoral dissertation.
- FUSEK, M. and MICHÁLEK, J. 2012a. Statistical analysis of doubly left-censored samples from exponential distribution. In: *Proceedings of the 18th International Conference on Soft Computing MENDEL 2012*. Brno University of Technology, June 2012. Brno: Faculty of Mechanical Engineering, pp. 564–569.
- FUSEK, M. and MICHÁLEK, J. 2012b. Statistical inference of doubly left-censored samples from Weibull distribution. In: *Proceedings of the 20th International Conference Problems of Decision Making under Uncertainties*. University of Defence, August 2012. Brno: University of Defence, pp. 31–40.
- FUSEK, M. and MICHÁLEK, J. 2013. 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.
- FUSEK, M. and MICHÁLEK, J. 2014. Asymptotic tests for multiply left-censored samples from Weibull distribution. In: *Proceedings of the 20th International Conference on Soft Computing MENDEL 2014*. Brno University of Technology, June 2014. Brno: Faculty of Mechanical Engineering, pp. 317–322.
- FUSEK, M. and MICHÁLEK, J. 2015a. T-test alternative for multiply left-censored samples from Weibull distribution. In: *Proceedings of the 21st International Conference on Soft Computing MENDEL 2015*. Brno University of Technology, June 2015. Brno: Faculty of Mechanical Engineering, pp. 169–174.
- FUSEK, M. and MICHÁLEK, J. 2015b. Statistical analysis of type I multiply left-censored samples from exponential distribution. *Journal of Statistical Computation and Simulation*, 85(11): 2148–2163.
- FUSEK, M., MICHÁLEK, J. and VÁVROVÁ, M. 2015. Evaluation of contamination data with non-detects using censored distributions. *Fresenius Environmental Bulletin*, 24(11): 4165–4172.
- FUSEK, M. and MICHÁLEK, J. 2016. On the confidence intervals for mean of left-censored Weibull distribution. In: *Proceedings of the 22nd International Conference on Soft Computing MENDEL 2016*. Brno University of Technology, June 2016. Brno: Faculty of Mechanical Engineering, pp. 249–254.
- FUSEK, M. and MICHÁLEK, J. 2017. On testing reduction of a left-censored Weibull distribution to an exponential submodel. In: *Proceedings of the 23rd International Conference on Soft Computing MENDEL 2017*. Brno University of Technology, June 2017. Brno: Faculty of Mechanical Engineering, pp. 179–184.
- GUPTA, R. D. and KUNDU, D. 1999. Generalized exponential distribution. *Australian & New Zealand Journal of Statistics*, 41(2): 173–188.
- HELSEL, D. R. 2006. Fabricating data: How substituting values for nondetects can ruin results, and what can be done about it. *Chemosphere*, 65(11): 2434–2439.
- HELSEL, D. R. 2012. *Statistics for censored environmental data using Minitab[®]* and R. New York: John Wiley and Sons.
- HOGG, R. V., MCKEAN, J. W., CRAIG, A. 2005. *Introduction to mathematical statistics*. 6th Edition. Upper Saddle River, NJ: Pearson Education.
- LAGARIAS, J. C., REEDS, J. A., WRIGHT, M. H. and WRIGHT, P. E. 1998. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal on Optimization*, 9(1): 112–147.
- LEHMANN, E. L. and CASELLA, G. 1998. *Theory of point estimation*. 2nd Edition. New York: Springer-Verlag.
- LEHMANN, E. L. and ROMANO, J. P. 2005. *Testing statistical hypotheses*. New York: Springer.

- MITRA, S. and KUNDU, D. 2008. Analysis of left censored data from the generalized exponential distribution. *Journal of Statistical Computation and Simulation*, 78(7): 669–679.
- SCHMOYERI, R. L., BEAUCHAMP, J. J., BRANDT, C. C. and HOFFMAN JR., F. O. 1996. Difficulties with the lognormal model in mean estimation and testing. *Environmental and Ecological Statistics*, 3(1): 81–97.
- SHOARI, N., DUBÉ, J. S. and CHENOURI, S. 2015. 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.
- SHUMWAY, R. H., AZARI, R. S. and KAYHANIAN, M. 2002. Statistical approaches to estimating mean water quality concentrations with detection limits. *Environmental Science & Technology*, 36(15): 3345–3353.

Contact information

Michal Fusek: fusekmi@feec.vutbr.cz

Jaroslav Michálek: jaroslav.michalek@unob.cz