Austrian Journal of Statistics
January 2023, Volume 52, 51–61.
http://www.ajs.or.at/
doi:10.17713/ajs.v52i1.1348

Statistical Power of Goodness-of-Fit Tests for Type I Left-Censored Data

Michal Fusek
Brno University of Technology

Abstract

Type I doubly left-censored data often arise in environmental studies. In this paper, the power of the most frequently used goodness-of-fit tests (Kolmogorov-Smirnov, Cramér-von Mises, Anderson-Darling) is studied considering various sample sizes and degrees of censoring. Attention is paid to testing of the composite hypothesis that the data has a specific distribution with unknown parameters, which are estimated using the maximum likelihood method. Performance of the tests is assessed by means of Monte Carlo simulations for several distributions, specifically the Weibull, lognormal and gamma distributions, which are among the most frequently used distributions for modelling of environmental data. Finally, the tests are used for identification of the distribution of musk concentrations if fish tissue.

Keywords: censored data, goodness-of-fit test, empirical power, type I left-censoring.

1. Introduction

Type I left-censored data are often present in practice, especially in environmental studies where an observed variable (e.g. a concentration of some chemical) frequently fall below detection limits of a measuring instrument (El-Shaarawi and Piegorsch 2012; Helsel 2012). Since there are usually two detection limits (the limit of detection and the limit of quantification) that are fixed and known in advance, we have to deal with doubly left-censored data. The parametric approach is usually used assuming that data has a specific distribution. In environmental studies, data are typically skewed and various distributions such as the lognormal (El-Shaarawi 1989; Baccarelli, Pfeiffer, Consonni, Pesatori, Bonzini, Patterson, Bertazzi, and Landi 2005), gamma (Singh, Singh, and Iaci 2002; Hrdličková, Michálek, Kolář, and Veselý 2008) and Weibull (Fusek, Michálek, and Vávrová 2015; Mbengue, Fusek, Schwarz, Vodička, Holubová Šmejkalová, and Holoubek 2018; Fusek, Michálek, Buňková, and Buňka 2020) distributions are often used. Since selecting an unsuitable probability distribution can lead to biased estimates and potentially misleading inferences, goodness-of-fit tests are of great importance. There are several goodness-of-fit tests available in the literature based on a complete sample and an excellent overview on this topic can be found in D'Agostino and Stephens (1986). Nevertheless, there has been relatively little work done on the problem of goodness-of-fit for Type I censored data and attention was usually paid only to right censoring

(Bispo, Marques, and Pestana 2011; Pakyari and Balakrishnan 2013; Pakyari and Nia 2017). In this paper, we focus on Type I left-censored data and three tests (Kolmogorov-Smirnov, Cramér-von Mises, Anderson-Darling) based on the empirical distribution function (EDF). The power of the tests is investigated by varying the null and the alternative distributions for various sample sizes and degrees of censoring. It can bring readers valuable information about the type II error that can be expected when having a dataset with a specific size and a number of censored values. A similar study was carried out by Bispo *et al.* (2011) for right-censored data, although, only for the completely specified alternative distributions. This is usually not the case when modelling real data. On that account, this paper is focused on a more complex problem of testing a composite hypothesis where the unknown parameters of the tested distributions have to be estimated. As far as we know, no similar analysis has been done before for the Type I left-censored data. The paper is organized as follows. Section 2 describes the statistical inference for selected Type I doubly left-censored probability distributions (Weibull, lognormal, gamma). Section 3 describes the goodness-of-fit tests used in this paper, and their performance is assessed using simulations in Section 4. Real data example is given in Section 5 and all the findings are summarized in the final Section 6.

2. Type I left-censored distributions

Let $X_1, \dots, X_n$ be a random sample from a distribution with cumulative distribution function (CDF) $F(x, \boldsymbol{\theta})$ and probability density function (PDF) $f(x, \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k) \in \boldsymbol{\Theta} \subset \mathbb{R}^k$ is a vector of parameters. The sample $X_1, \dots, X_n$ arranged in ascending order is denoted by $X_{(1)} \leq \cdots \leq X_{(n)}$. Two detection limits d_1, d_2 are considered and N_i, $i = 1, 2$, is the number of observations in the interval $(d_{i-1}, d_i]$, where we put $d_0 = 0$. The number of uncensored observations $X_{(n-N_0+1)}, \dots, X_{(n)}$ is denoted by N_0. The log-likelihood function of the censored sample is given by (Cohen 1991)

$$l(\boldsymbol{\theta}, N_0, N_1, N_2, X_{(n-N_0+1)}, \dots, X_{(n)}) = \log\left(\frac{n!}{N_1! N_2!}\right) + \sum_{i=1}^{2} N_i \log\left[F(d_i, \boldsymbol{\theta}) - F(d_{i-1}, \boldsymbol{\theta})\right] + \sum_{i=n-N_0+1}^{n} \log\left[f(X_{(i)}, \boldsymbol{\theta})\right], \quad (1)$$

and for $N_0 = 0$ we put $\sum_{i=n-N_0+1}^{n} \log\left[f(X_{(i)}, \boldsymbol{\theta})\right] = 0$. The maximum likelihood (ML) estimate $\widehat{\boldsymbol{\theta}}$ of parameter $\boldsymbol{\theta}$ can be obtained by maximizing the log-likelihood function (1) using, for example, the Nelder-Mead simplex algorithm in Matlab (version R2020b). Another option is to derive likelihood equations and solve them using the Newton-Raphson method.

In this paper, following probability distributions are used:

1) The Weibull distribution with parameter $\boldsymbol{\theta} = (\lambda, \tau) \in (0, \infty) \times (0, \infty)$, and CDF

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

where λ is the scale parameter and τ is the shape parameter.

2) The lognormal distribution with parameter $\boldsymbol{\theta} = (\mu, \sigma) \in (-\infty, \infty) \times (0, \infty)$, and CDF

$$F(x, \boldsymbol{\theta}) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} \int\limits_0^x \frac{1}{t} \exp\left[-\frac{(\log(t)-\mu)^2}{2\sigma^2}\right] \mathrm{d}t & \text{for } x > 0, \\ 0 & \text{for } x \leq 0, \end{cases} \quad (3)$$

where μ is the location parameter and σ is the scale parameter of the variable's natural logarithm.

3) The gamma distribution with parameter $\boldsymbol{\theta} = (\lambda, \kappa) \in (0, \infty) \times (0, \infty)$, and CDF

$$F(x, \boldsymbol{\theta}) = \begin{cases} \frac{1}{\lambda^\kappa \Gamma(\kappa)} \int\limits_0^x t^{\kappa-1} \exp\left(-\frac{t}{\lambda}\right) \mathrm{d}t & \text{for } x > 0, \\ 0 & \text{for } x \leq 0, \end{cases}$$

where λ is the scale parameter and κ is the shape parameter.

4) The Gumbel distribution with parameter $\boldsymbol{\theta} = (\mu, \sigma) \in (-\infty, \infty) \times (0, \infty)$, and CDF

$$F(x, \boldsymbol{\theta}) = 1 - \exp\left[-\exp\left(\frac{x - \mu}{\sigma}\right)\right] \text{ for } x \in \mathbb{R}, \tag{4}$$

where μ is the location parameter and σ is the scale parameter.

5) The normal distribution with parameter $\boldsymbol{\theta} = (\mu, \sigma) \in (-\infty, \infty) \times (0, \infty)$, and CDF

$$F(x, \boldsymbol{\theta}) = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{x} \exp\left[-\frac{(t - \mu)^2}{2\sigma^2}\right] \mathrm{d}t \text{ for } x \in \mathbb{R}, \tag{5}$$

where μ is the location parameter and σ is the scale parameter.

3. Goodness-of-fit test statistics

Let $X_1, \ldots, X_n$ be a random sample from a distribution with CDF $F(x)$. We consider a problem of testing a composite hypothesis

$$\mathrm{H}_0 : F(x) \in \left\{F_0(x, \boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^k\right\},$$

where F_0 is a CDF of a known parametric family. In case $\boldsymbol{\theta}$ is fully specified, then H_0 is a simple hypothesis and the distribution theory of EDF statistics is well developed. When $\boldsymbol{\theta}$ is unknown, it can be replaced by its estimate $\widehat{\boldsymbol{\theta}}$, and distributions of EDF statistics depend on the tested distribution, the estimated parameters and the sample size. It is well known fact (D'Agostino and Stephens 1986) that in case the unknown components in $\boldsymbol{\theta}$ are location or scale parameters, distributions of EDF statistics do not depend on the true values of the unknown parameters, and depend only on the tested distribution and on the sample size. When the unknown component in $\boldsymbol{\theta}$ is the shape parameter, distributions of EDF statistics depend on the true value of this parameter. In our case, it was possible to transform the distributions depending on the shape parameter to another distributions depending on the location and scale parameters only. Specifically, if a random variable X has the Weibull distribution, then $\log(X)$ has the Gumbel distribution. Therefore, testing the null hypothesis that the data follow the Weibull distribution (2) is equivalent to testing that the log-transformed data follow the Gumbel distribution with location parameter $\mu = \log(\lambda)$, scale parameter $\sigma = 1/\tau$, and CDF (4). Moreover, a random variable X has the lognormal distribution if $\log(X)$ has the normal distribution. For that reason, testing the null hypothesis that the data follow the lognormal distribution (3) is equivalent to testing that the log-transformed data follow the normal distribution with mean μ, standard deviation σ, and CDF (5).

Critical values of the EDF statistics can be obtained by means of Monte Carlo simulations using the following steps:

1) Generate a Type I doubly left-censored sample $X_1, \ldots, X_n$ with a pre-chosen sample size n and detection limits from the distribution being tested. Detection limits are selected as quantiles of the tested distribution depending on the degree of censoring.

2) Calculate the ML estimates of the unknown parameters of the selected distribution.

3) Calculate the EDF statistic.

4) Repeat steps 1–3 a large number of times and determine the $(1-\alpha)$th quantile of the test statistic as the required critical value of that goodness-of-fit statistic.

Three test statistics based on the EDF $F_n(x)$ are applied (see D'Agostino and Stephens 1986, for more details).

3.1. Kolmogorov-Smirnov statistic

The Kolmogorov-Smirnov (KS) statistic is defined by

$$D = \sup_{d_2 \le x < \infty} |F_n(x) - F_0(x)|$$

with the useful alternative form for computational purposes

$$D = \max_{n-N_0+1 \le i \le n} \left\{ \left| \frac{i}{n} - F_0\left(x_{(i)}, \widehat{\boldsymbol{\theta}}\right) \right|, \left| F_0\left(x_{(i)}, \widehat{\boldsymbol{\theta}}\right) - \frac{i-1}{n} \right|, \left| F_0\left(d_2, \widehat{\boldsymbol{\theta}}\right) - \frac{n-N_0}{n} \right| \right\}.$$

3.2. Cramér-von Mises statistic

The Cramér-von Mises (CM) statistic is defined by

$$W^2 = n \int_{d_2}^{\infty} [F_n(x) - F_0(x)]^2 \, \mathrm{d}F_0(x)$$

with an alternative form for computational purposes

$$W^2 = \sum_{i=1}^{N_0+1} \left(Z_{(i)} - \frac{2i-1}{2n} \right)^2 + \frac{N_0+1}{12n^2} + \frac{n}{3} \left(Z_{(N_0+1)} - \frac{N_0+1}{n} \right)^3,$$

where $Z_{(i)} = 1 - F_0(x_{(n-i+1)}, \widehat{\boldsymbol{\theta}})$, $i = 1, \ldots, N_0$, and $Z_{(N_0+1)} = 1 - F_0(d_2, \widehat{\boldsymbol{\theta}})$.

3.3. Anderson-Darling statistic

The Anderson-Darling (AD) statistic is a modification of the CM statistic placing more weight in the tails of the underlying distribution. It is defined by

$$A^2 = n \int_{d_2}^{\infty} \frac{[F_n(x) - F_0(x)]^2}{F_0(x)[1 - F_0(x)]} \, \mathrm{d}F_0(x)$$

with an alternative form for computational purposes

$$\begin{aligned} A^2 = & -\frac{1}{n} \sum_{i=1}^{N_0+1} (2i-1) \left[\log(Z_{(i)}) - \log(1 - Z_{(i)}) \right] - 2 \sum_{i=1}^{N_0+1} \log(1 - Z_{(i)}) \\ & - \frac{1}{n} \left[(N_0 + 1 - n)^2 \log(1 - Z_{(N_0+1)}) - (N_0+1)^2 \log(Z_{(N_0+1)}) + n^2 Z_{(N_0+1)} \right], \end{aligned}$$

where again $Z_{(i)} = 1 - F_0(x_{(n-i+1)}, \widehat{\boldsymbol{\theta}})$, $i = 1, \ldots, N_0$, and $Z_{(N_0+1)} = 1 - F_0(d_2, \widehat{\boldsymbol{\theta}})$.

4. Simulation study

The empirical significance level as well as the power of the above mentioned tests was studied by means of Monte Carlo simulations. Tested models included the Weibull (denoted by Wbl(λ, τ)), lognormal (denoted by LN(μ, σ)) and gamma (denoted by Gam(λ, κ)) distributions as these are among the most frequently used distributions when modelling censored environmental data. In case of the Weibull (lognormal respectively) distribution, the previously described transformation to the Gumbel (normal respectively) distribution was applied. The power of the goodness-of-fit tests was estimated by the proportion of the correct rejections of the null hypothesis at the significance level of $\alpha = 0.05$. The power of each statistic was simulated from 100,000 replications considering sample sizes $n = 10, 20, 30, 50, 100, 200$ and censoring schemes $c_1, \ldots, c_4$, which represent proportions of censored observations between 10% and 70% (with a step of 20%). For example, c_3 represents 50% of censored values. Detection limits d_1, d_2 were selected as quantiles of the particular distribution using equations $q_i = F(d_i, \boldsymbol{\theta})$, $i = 1, 2$, where $q_2 = 0.1, 0.3, 0.5, 0.7$ and $q_1 = q_2/2$. Critical values of the test statistics were obtained by means of Monte Carlo simulations using 2,000,000 repetitions. When the alternative model is the model from which the data are simulated, the rejection probabilities give the power of the tests. In case the null hypothesis is true, it is expected that the statistics maintain the type I error rate. Overall, differences between the nominal level of 0.05 and the actual levels were very small for various censoring schemes and sample sizes, which shows a reliable performance of the goodness-of-fit statistics for left-censored data (see Tables 1–3).

Table 1: Estimated power for various alternatives and censoring schemes when testing for the Weibull distribution; $n = 30, 100$

Alt. model	Stat.	$n = 30$				$n = 100$			
		c_1	c_2	c_3	c_4	c_1	c_2	c_3	c_4
Wbl(1,0.5)	KS	0.0504	0.0493	0.0485	0.0499	0.0506	0.0505	0.0501	0.0500
	CM	0.0508	0.0494	0.0500	0.0488	0.0502	0.0503	0.0506	0.0515
	AD	0.0505	0.0493	0.0501	0.0490	0.0500	0.0496	0.0505	0.0506
Wbl(1,2)	KS	0.0493	0.0497	0.0496	0.0506	0.0501	0.0506	0.0511	0.0505
	CM	0.0498	0.0498	0.0500	0.0496	0.0509	0.0499	0.0510	0.0503
	AD	0.0496	0.0494	0.0498	0.0496	0.0507	0.0503	0.0514	0.0498
LN(0,0.5)	KS	0.2083	0.1679	0.1487	0.1324	0.5902	0.4462	0.3812	0.3503
	CM	0.2396	0.1891	0.1751	0.1612	0.6835	0.5303	0.4731	0.4238
	AD	0.2396	0.2044	0.1866	0.1631	0.6906	0.5607	0.4919	0.4130
LN(0,2)	KS	0.2073	0.1701	0.1481	0.1318	0.5854	0.4464	0.3858	0.3514
	CM	0.2399	0.1901	0.1759	0.1605	0.6819	0.5328	0.4793	0.4249
	AD	0.2388	0.2063	0.1860	0.1624	0.6892	0.5623	0.4970	0.4131
Gam(0.5,0.5)	KS	0.0696	0.0643	0.0628	0.0619	0.1278	0.1074	0.0992	0.0911
	CM	0.0720	0.0639	0.0617	0.0545	0.1469	0.1181	0.1069	0.0842
	AD	0.0726	0.0615	0.0597	0.0529	0.1546	0.1176	0.1053	0.0819
Gam(0.5,2)	KS	0.0627	0.0605	0.0588	0.0559	0.0952	0.0848	0.0768	0.0746
	CM	0.0676	0.0637	0.0628	0.0611	0.1093	0.0921	0.0869	0.0841
	AD	0.0683	0.0664	0.0653	0.0621	0.1141	0.0968	0.0913	0.0847

Table 1 shows that power of all the test statistics is very low when data generated from the gamma distribution are tested for the Weibull distribution. Similar behavior is observed when data generated from Gam(0.5,2) are tested for the lognormal distribution (Table 2), and data generated from LN(0,0.5) and Wbl(1,2) are tested for the gamma distribution (Table 3). It is caused by the fact that it is very difficult to distinguish between the gamma, lognormal and Weibull distributions in some cases (see Fig. 1).

As expected, the statistical power of the studied tests increases with increase in the sample size, and decreases with the increasing proportion of censored values. The relation between

Table 2: Estimated power for various alternatives and censoring schemes when testing for the lognormal distribution; $n = 30, 100$

		$n=30$				$n=100$			
Alt. model	Stat.	c_1	c_2	c_3	c_4	c_1	c_2	c_3	c_4
LN(0,0.5)	KS	0.0491	0.0492	0.0507	0.0497	0.0507	0.0500	0.0497	0.0499
	CM	0.0487	0.0500	0.0502	0.0484	0.0505	0.0495	0.0494	0.0494
	AD	0.0485	0.0503	0.0508	0.0493	0.0509	0.0497	0.0497	0.0484
LN(0,2)	KS	0.0488	0.0496	0.0500	0.0508	0.0506	0.0496	0.0500	0.0492
	CM	0.0490	0.0495	0.0501	0.0491	0.0515	0.0498	0.0498	0.0497
	AD	0.0490	0.0495	0.0494	0.0493	0.0512	0.0500	0.0500	0.0499
Wbl(1,0.5)	KS	0.2188	0.1530	0.1219	0.0921	0.6180	0.4565	0.3461	0.2335
	CM	0.2636	0.1757	0.1195	0.0731	0.7501	0.5660	0.3809	0.1962
	AD	0.2673	0.1796	0.1227	0.0781	0.7870	0.6016	0.4099	0.2253
Wbl(1,2)	KS	0.2175	0.1543	0.1228	0.0919	0.6177	0.4552	0.3459	0.2320
	CM	0.2635	0.1758	0.1198	0.0735	0.7493	0.5653	0.3820	0.1954
	AD	0.2685	0.1792	0.1237	0.0786	0.7858	0.6003	0.4102	0.2235
Gam(0.5,0.5)	KS	0.3763	0.2675	0.2014	0.1336	0.8906	0.7563	0.6158	0.4062
	CM	0.4778	0.3221	0.1965	0.1012	0.9664	0.8753	0.6633	0.3301
	AD	0.5026	0.3394	0.2137	0.1169	0.9803	0.9077	0.7223	0.4100
Gam(0.5,2)	KS	0.1250	0.0970	0.0831	0.0696	0.3406	0.2421	0.1916	0.1414
	CM	0.1432	0.1065	0.0814	0.0589	0.4249	0.2961	0.2079	0.1227
	AD	0.1410	0.1052	0.0815	0.0600	0.4476	0.3110	0.2159	0.1314

Table 3: Estimated power for various alternatives and censoring schemes when testing for the gamma distribution; $n = 30, 100$

		$n=30$				$n=100$			
Alt. model	Stat.	c_1	c_2	c_3	c_4	c_1	c_2	c_3	c_4
Gam(0.5,0.5)	KS	0.0501	0.0507	0.0496	0.0505	0.0493	0.0506	0.0501	0.0505
	CM	0.0497	0.0516	0.0515	0.0512	0.0499	0.0498	0.0510	0.0509
	AD	0.0501	0.0514	0.0504	0.0509	0.0501	0.0494	0.0506	0.0507
Gam(0.5,2)	KS	0.0501	0.0505	0.0507	0.0501	0.0499	0.0497	0.0497	0.0512
	CM	0.0502	0.0504	0.0499	0.0502	0.0507	0.0498	0.0501	0.0504
	AD	0.0501	0.0502	0.0502	0.0498	0.0511	0.0503	0.0502	0.0509
LN(0,0.5)	KS	0.0963	0.0883	0.0829	0.0768	0.2097	0.1706	0.1521	0.1419
	CM	0.1061	0.0967	0.0933	0.0904	0.2486	0.2033	0.1848	0.1709
	AD	0.1152	0.1068	0.1019	0.0961	0.2640	0.2213	0.1956	0.1712
LN(0,2)	KS	0.5613	0.4790	0.4177	0.3751	0.9746	0.9395	0.8968	0.8565
	CM	0.6307	0.5401	0.4826	0.4389	0.9893	0.9678	0.9395	0.9002
	AD	0.6286	0.5504	0.4891	0.4293	0.9893	0.9701	0.9409	0.8901
Wbl(1,0.5)	KS	0.1385	0.1229	0.1133	0.1064	0.3723	0.3148	0.2817	0.2664
	CM	0.1610	0.1424	0.1348	0.1324	0.4465	0.3814	0.3469	0.3266
	AD	0.1648	0.1501	0.1396	0.1311	0.4568	0.4006	0.3612	0.3209
Wbl(1,2)	KS	0.0833	0.0722	0.0686	0.0645	0.1741	0.1391	0.1224	0.1011
	CM	0.0887	0.0743	0.0674	0.0546	0.2068	0.1610	0.1300	0.0896
	AD	0.0830	0.0693	0.0640	0.0516	0.2078	0.1611	0.1287	0.0892

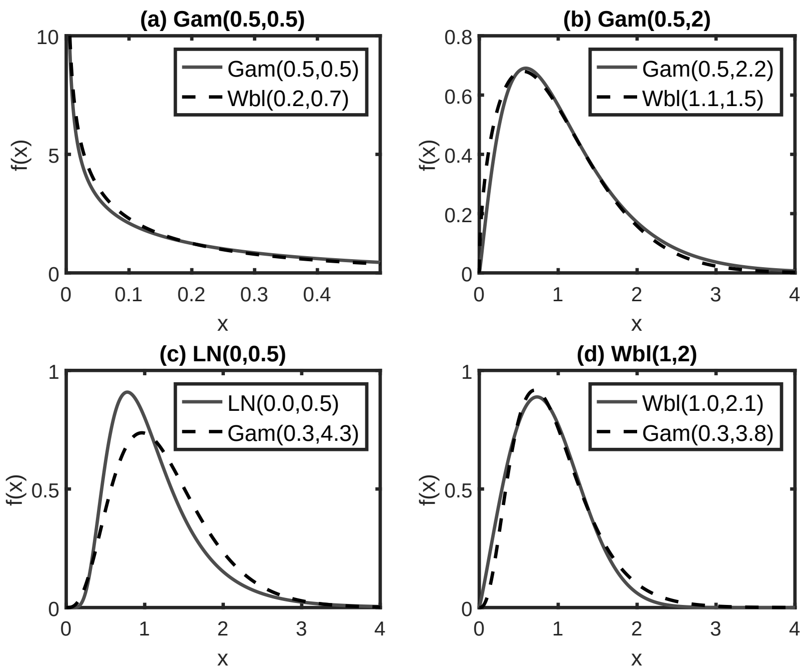

Figure 1: Densities of some alternative distributions (solid lines) compared with densities of the fitted distributions (dashed lines) for $n = 30$ and 30% of censored values

the power of the test and the sample size is visualized for several cases in Fig. 2. Note that the AD test seems to have the best performance for most cases in comparison to the CM and KS tests. In general, there is not much difference between the powers of the AD and CM tests and the KS test usually performs the worst. Let us look more closely at the tests' performance with regards to the proportion of censored values in a sample. If there is at most 50% of censored values (schemes c_1–c_3) in a sample, the AD test can be used without much hesitation. Nevertheless, when the proportion of censored values is high (70%; scheme c_4), the KS and CM tests sometimes outperform the AD test. In case of testing data with sample size $n < 100$ for the lognormal distribution, the KS test has the highest power (i.e. the lowest type II error), see Fig. 3. When testing lognormally distributed data with sample size $n > 50$ for the Weibull distribution and/or data generated from LN(0,2) for the gamma distribution, the CM test has the highest power. There are few other cases in which the KS test performs the best, specifically when data generated from Gam(0.5,0.5) are tested for the Weibull distribution and/or when data generated from Wbl(1,2) are tested for the gamma distribution. Similarly, when data generated from Wbl(1,0.5) are tested for the gamma distribution, the CM test performs the best. Nevertheless, in these three cases, the differences in tests' powers are very small and they can be considered negligible from the practical point of view.

5. Distribution of musk concentrations

Fusek *et al.* (2015) analyzed various musks in fish tissue. We can take a look, for example, at tonalide, which is a synthetic polycyclic musk. Fusek *et al.* (2015) assumed that tonalide concentrations formed an independent and identically distributed Type I left-censored random sample with the Weibull distribution. There were two detection limits, specifically $d_1 = 1.9846$ μg/kg, $d_2 = 6.6154$ μg/kg, and $N_1 = 6$ non-detectable values, $N_2 = 17$ non-quantifiable values and following observed values:

$$7.1215,\ 7.2008,\ 7.4243,\ 7.993,\ 8.2360,\ 10.5176,\ 15.1702.$$

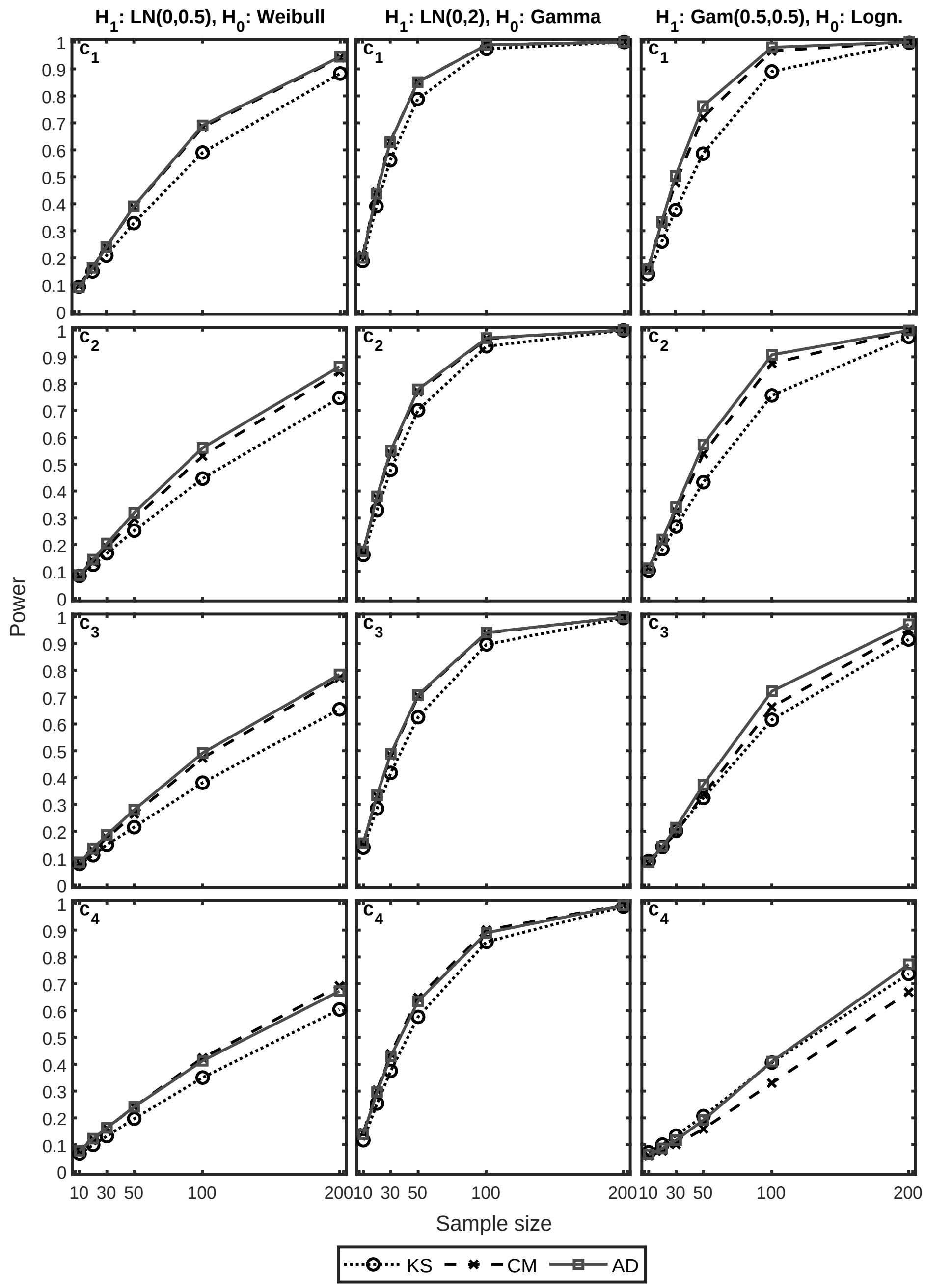

Figure 2: Estimated power for various alternatives as a function of the sample size and the proportion of censored values c_1 (10%), c_2 (30%), c_3 (50%), c_4 (70%) when testing for the Weibull (left), gamma (middle) and lognormal (right) distributions

The data consists of little bit more than 70% of censored values which roughly corresponds to censoring scheme c_4. Tables 1–3 ($n = 30$, censoring scheme c_4) shows that when testing for the Weibull distribution, the lowest type II error (i.e. the highest power) occurs mostly in case of the AD test. When testing for the lognormal and/or gamma distribution, the lowest type II error occurs mostly in case of the KS test. Table 4 presents the test statistics and corresponding p-values when testing for the Weibull, lognormal and gamma distributions. All the p-values are quite large, strongly supporting the corresponding null hypothesis of the Weibull distribution. Moreover, p-values associated with the gamma distribution are slightly larger than those corresponding to other distributions, suggesting that the gamma distribution can be more suitable for fitting the data. This example illustrates, among other things, that choosing the right model distribution can be difficult and sometimes there are alternatives that

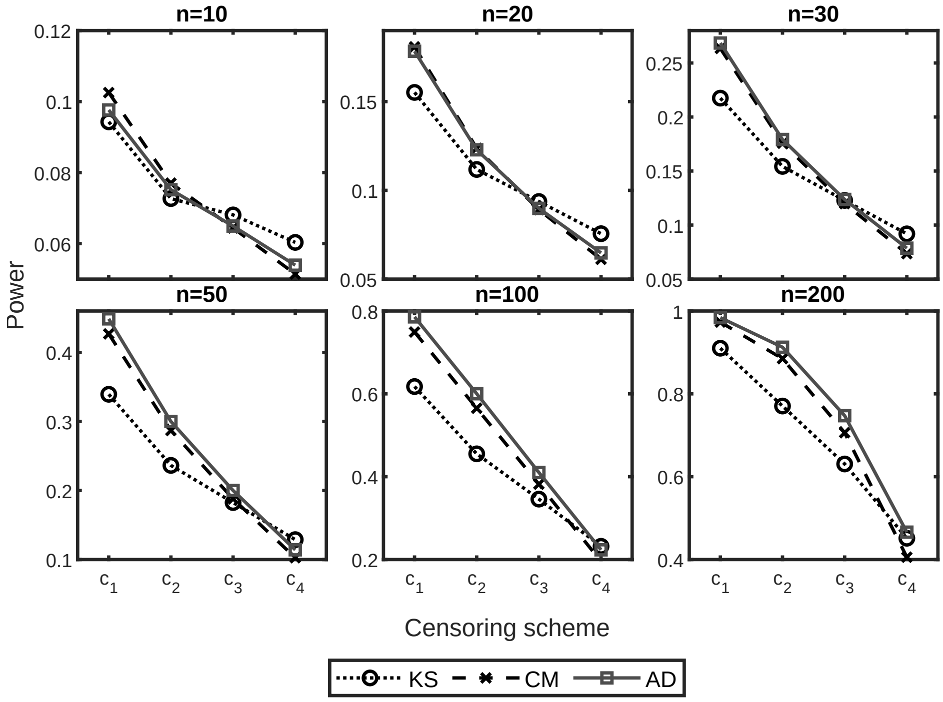

Figure 3: Estimated power for Wbl(1,2) alternative as a function of the sample size n and the proportion of censored values c_1 (10%), c_2 (30%), c_3 (50%), c_4 (70%) when testing for lognormal distribution

Table 4: Test statistics (Stat.) and the corresponding p-values for tonalide concentrations when testing for the Weibull, lognormal and gamma distributions

	AD		CM		KS	
Distribution	Stat.	p-value	Stat.	p-value	Stat.	p-value
Weibull	0.2445	0.5151	0.0271	0.5243	0.1315	0.4827
Lognormal	0.0637	0.5415	0.0043	0.3886	0.0690	0.3686
Gamma	0.1602	0.6421	0.0128	0.5366	0.1014	0.5388

give similar results. For example, in Stadlober, Hübnerová, Michálek, and Kolář (2012), daily averages of PM10 were modelled using two different approaches, specifically the multiple linear regression with square root transformation and the generalized linear model with gamma distribution and log-link. It was shown that both approaches are suitable and give similar results.

Finally, it should be pointed out that musk concentrations can be viewed as compositions. Such an approach may be considered in subsequent research. Nevertheless, many values are below the detection limit, which causes problems for the compositional approach (Lubbe, Filzmoser, and Templ 2021).

6. Conclusion

This paper studied statistical powers of the three frequently used goodness-of-fit tests (Kolmogorov-Smirnov, Cramér-von Mises, Anderson-Darling) for Type I doubly left-censored data in case the distribution function $F_0(x, \boldsymbol{\theta})$ in the hypothesis is not fully specified. Performance of the statistics was assessed using simulations for several alternative distributions (Weibull, lognormal, gamma) considering various sample sizes and degrees of censoring. In that way readers can get and utilize information about the type II error that can be expected when having a dataset with a specific size and a number of censored values. The Monte Carlo simulation study showed that all test statistics works very well in terms of recovering the nominal level of significance of 0.05 even for high degrees of censoring and small sample sizes. Despite the fact that the Anderson-Darling test seems to have the best performance in most cases and it is usually the recommended test by many statisticians, a different test may be preferred for a specific combination of the sample size and the number of censored values. If there is at most 50% of censored values in a sample, the Anderson-Darling test can be recommended. Nevertheless, when the proportion of censored values is very high (70%), the Kolmogorov-Smirnov and Cramér-von Mises tests can sometimes outperform the Anderson-Darling test. For example, in case of testing data with sample size $n < 100$ for the lognormal distribution, the Kolmogorov-Smirnov test has the highest power (i.e. the lowest type II error). Moreover, when testing lognormally distributed data with sample size $n > 50$ for the Weibull distribution or data generated from LN(0,2) for the gamma distribution, the Cramér-von Mises test has the highest power. Sometimes, the alternative distributions are hard to be distinguished from the model under the null hypothesis. For example, the power of all test statistics is very low when data generated from the gamma distribution are tested for the Weibull distribution. In addition, theoretical results were applied to tonalide concentration data. It was shown that various distributions could be used for modelling of tonalide concentration and the Kolmogorov-Smirnov test has the lowest type II error in most cases when testing for the lognormal or gamma distributions. Results of this study can be utilized for identification of the correct distribution of Type I left-censored data, and, above all, for assessing the power and the type II error of the goodness-of-fit tests for various alternatives, sample sizes and degrees of censoring, which is an often discussed topic, especially in case of highly censored data.

References

Baccarelli A, Pfeiffer R, Consonni D, Pesatori A, Bonzini M, Patterson DG, Bertazzi P, Landi M (2005). "Handling of Dioxin Measurement Data in Presence of Non-Detectable Values: Overview of Available Methods and Their Application in the Seveso Chloracne Study." *Chemosphere*, **60**, 898–906.

Bispo R, Marques TA, Pestana D (2011). "Statistical Power of Goodness-of-Fit Tests Based on the Empirical Distribution Function for Type-I Right-Censored Data." *Journal of Statistical Computation and Simulation*, **82**, 173–181.

Cohen AC (1991). *Truncated and Censored Samples.* Marcel Dekker.

D'Agostino RB, Stephens MA (1986). *Goodness-of-Fit Techniques.* Marcel Dekker.

El-Shaarawi AH (1989). "Inferences about the Mean from Censored Water Quality Data." *Water Resources Research*, **25**, 685–690.

El-Shaarawi AH, Piegorsch WW (2012). *Encyclopedia of Environmetrics.* 2nd ed. John Wiley and Sons.

Fusek M, Michálek J, Buňková L, Buňka F (2020). "Modelling Biogenic Amines in Fish Meat in Central Europe Using Censored Distributions." *Chemosphere*, **251**, 126390.

Fusek M, Michálek J, Vávrová M (2015). "Evaluation of Contamination Data with Non-Detects Using Censored Distributions." *Fresenius Environmental Bulletin*, **24**, 4165–4172.

Helsel DR (2012). *Statistics for Censored Environmental Data Using Minitab and R.* John Wiley and Sons.

Hrdličková Z, Michálek J, Kolář M, Veselý V (2008). "Identification of Factors Affecting Air Pollution by Dust Aerosol PM10 in Brno City, Czech Republic." *Atmospheric Environment*, **42**, 8661–8673.

Lubbe S, Filzmoser P, Templ M (2021). "Comparison of Zero Replacement Strategies for Compositional Data with Large Numbers of Zeros." *Chemometrics and Intelligent Laboratory Systems*, **210**, 104248.

Mbengue S, Fusek M, Schwarz J, Vodička P, Holubová Šmejkalová A, Holoubek I (2018). "Four Years of Highly Time Resolved Measurements of Elemental and Organic Carbon at a Rural Background Site in Central Europe." *Atmospheric Environment*, **182**, 335–346.

Pakyari R, Balakrishnan N (2013). "Testing Exponentiality Based on Type-I Censored Data." *Journal of Statistical Computation and Simulation*, **83**, 2369–2378.

Pakyari R, Nia KR (2017). "Testing Goodness-of-Fit for some Lifetime Distributions with Conventional Type-I Censoring." *Communications in Statistics - Simulation and Computation*, **46**, 2998–3009.

Singh A, Singh AK, Iaci RJ (2002). *Estimation of the Exposure Point Concentration Term Using a Gamma Distribution.* U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-02/084 (NTIS PB2003-106618).

Stadlober E, Hübnerová Z, Michálek J, Kolář M (2012). "Forecasting of Daily PM10 Concentrations in Brno and Graz by Different Regression Approaches." *Austrian Journal of Statistics*, **41**(4), 287–310.

Affiliation:

Michal Fusek
Department of Mathematics
Faculty of Electrical Engineering and Communication
Brno University of Technology
Technická 2848/8, 616 00 Brno, Czech Republic
E-mail: fusekmi@vut.cz

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 52 *Submitted:* 2021-05-12
January 2023 *Accepted:* 2021-09-02