Statistical methods for analyzing musk compounds concentration based on doubly left-censored samples

Michal Fusek and Jaroslav Michálek

***Abstract*—**This contribution is focused on statistical methods for analyzing the worldwide commonly used synthetic musk compounds. Method of maximum likelihood considering doubly left-censored samples is used for statistical modeling of musk compound concentration. As for model distributions, the exponential and Weibull distributions are considered. The suitability of replacement of Weibull distribution with exponential distribution is explored using the asymptotic tests (Lagrange multiplier test, likelihood ratio test, Wald test). Moreover, using the asymptotic properties of maximum likelihood estimates, methods for comparison of two censored samples from exponential distribution are proposed and applied in analysis of concentrations of musk compounds extracted from the fish samples caught in front of and behind a wastewater treatment plant. The power functions of particular tests are compared by simulations.

***Keywords*—**Musk compounds, maximum likelihood, doubly left-censored sample, Weibull distribution, exponential distribution, Lagrange multiplier, likelihood ratio, Wald test.

I. Introduction

SYNTETIC musk compounds represent a group of organic environmental contaminants because of their persistence, bioaccumulation potential (lipophilic properties) and toxicity (see [2], [6], [16]). They have widespread use as substitutes for natural musks in fragrances, and can be found in a number of consumer products such as laundry detergents, fabric softeners, cleaning agents, and cosmetic and hygiene products (soaps, shampoos, body lotions, perfumes, etc.). Synthetic musk compounds penetrate into the environment primarily through wastewater because of their ineffective removal in wastewater treatment plants (WWTP). An accumulation of these substances in the environment (surface water, sediment) results in their occurrence in food chain, especially in the aquatic ecosystems. These compounds can also be found in human body, e.g. in tissue or body fluids like blood or mother's milk (see e.g. [14]), as a consequence of fish consumption. On that account, it is important to monitor the concentration of musk compounds if fish tissue.

M. Fusek is with the Institute of Mathematics, Faculty of Mechanical Engineering, Brno University of Technology, Technická 2896/2, 616 69 Brno, Czech Republic (phone: +420 541 142 720, fax: +420 541 142 710, e-mail: fusek.m@gmail.com).

J. Michálek is with the Institute of Mathematics, Faculty of Mechanical Engineering, Brno University of Technology, Technická 2896/2, 616 69 Brno, Czech Republic (e-mail: michalek@fme.vutbr.cz).

When analyzing musk compounds, we often have to deal with a situation when the substance is either absent or exists at such a low concentration that it is not present above the detection limit level. Performed chemical analyses do not allow for precise determination of respective concentrations in case the resulting values are found below the limit of detection (LOD), or limit of quantification (LOQ) of the determination method. The LOD is the lowest concentration of a substance that can be distinguished from the absence of that substance in a sample, and LOQ is the lowest concentration at which we can reasonably tell the difference between two different values of concentration. Since two fixed detection limits are present, it is necessary to work with doubly left-censored samples. Thus type I censoring is considered and the number of censored experimental units is a random variable.

Various censoring techniques and statistical analyses of censored data are described in more details in many monographs, e.g. [4], [5]. In many environmental studies, left censoring is based on normal distribution (see [7], [8]). However, the distribution of variables, such as concentration is asymmetric and skewed to the right (see Fig. 1 for example, where the histogram of one typical musk compound concentration — concentration of traseolide — can be found; data are described in [16]). Thus the normal distribution is not the most suitable choice.

For this very reason, this contribution is focused on statistical methods for analyzing left-censored chemical data with asymmetric distribution. Further, let us assume that distribution of the given musk compound concentration follows Weibull distribution, because it is very flexible, and with suitable choice of parameters, it allows describing symmetric and also highly asymmetric distributions. It can be seen from Fig. 2, where the typical histogram of musk compound concentration — concentration of phantolide — is, and also from previous results (see [11], [16]), that distribution of environmental contaminants concentration is highly asymmetric, and it seems that there is a possibility of using exponential distribution, which is a special case of Weibull distribution, instead of Weibull distribution. On that account, statistical tests for validating whether Weibull distribution can be reduced to exponential distribution are proposed in this paper. Recently developed methods for dealing with doubly left-censored samples from exponential distribution (see [9]) and Weibull distribution (see [10]) considering Type I censoring are used.

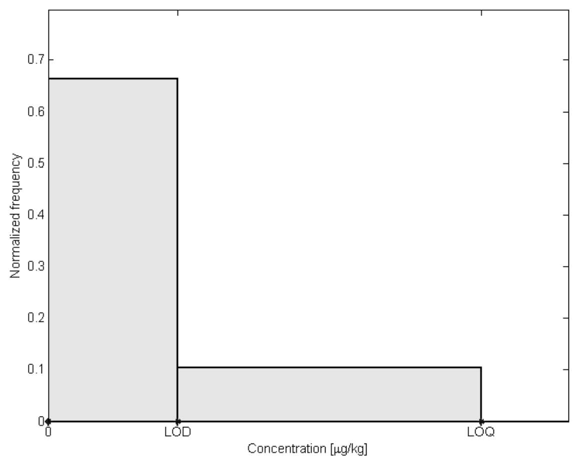

Fig. 1 Histogram of traseolide concentration.

The proposed test statistics will be based on the maximum likelihood (ML) theory (see [12]). Using of ML estimators is rather common in many application areas, e.g. in biometric models of signal transduction process (see [15]) or in measuring service quality (see [1]). In both examples, ML theory is used in a similar way as in this paper; however, censored data are considered in this contribution.

Another reason for the use of ML theory is that, subject to the regularity conditions, estimators have good asymptotic properties. In addition, it is possible to derive test statistics (likelihood ratio, Lagrange multiplier and Wald statistic; see [13]) even in case of presence of unknown nuisance parameters, which is precisely the situation described in this paper.

On the other hand, using ML theory brings also difficulties, because in order to obtain ML estimates, it is often necessary to find a solution of generally rather complicated nonlinear equations. On that account, many authors use numerically less demanding methods like method of moments or L-moments (see e.g. [3]) when looking for estimates of unknown parameters of particular distributions. In this contribution, it was possible to find numerical solution of the likelihood equations using proper algorithms, and the method of moments is used only for obtaining initial (starting) estimates of unknown parameters for solving likelihood equations.

Proposed statistical tests will be compared on real data and their power functions will be simulated in order to choose optimal tests for statistical analysis of particular musk compound concentration.

II. Data

The real sample (see [11], [16]) consists of 60 fish from the carp family, specifically of the European chub (Leuciscus cephalus), which were caught in the Svratka River near the WWTP Brno-Modřice (Czech Republic). Half of them were caught in front of (Group 1), and half of them behind (Group 2) the WWTP. Fish tissue samples (specifically in muscle) were analyzed, and two nitromusk compounds (musk ambrette (AMB), musk tibetene (TIB)), and two polycyclic musk compounds (phantolide (PH), traseolide (TR)) were explored. Fish of approximately the same age were chosen for the analysis.

III. Model of Censored Sample

Firstly, doubly left-censored Weibull distribution for modeling of musk compounds concentration in fish muscle will be considered.

A. Weibull Distribution

Let $X_1,\ldots,X_n$ be a Type I doubly left-censored random sample from Weibull distribution with scale parameter $\lambda > 0$, shape parameter $\tau > 0$, cumulative distribution function (cdf)

$$F(x,\lambda,\tau) = \begin{cases} 1 - \mathrm{e}^{-\left(\frac{x}{\lambda}\right)^{\tau}}, & x \geq 0, \\ 0, & x < 0 \end{cases} \tag{1}$$

and probability density function (pdf)

$$f(x,\lambda,\tau) = \begin{cases} \frac{\tau}{\lambda^{\tau}} x^{\tau-1} \mathrm{e}^{-\left(\frac{x}{\lambda}\right)^{\tau}}, & x \geq 0. \\ 0, & x < 0 \end{cases} \tag{2}$$

Furthermore, let $X_{(1)},\ldots,X_{(n)}$ be the ordered sample of $X_1,\ldots,X_n$. For simplicity, in all the formulas, detection limits will be denoted as $\mathrm{LOD} = d_1, \mathrm{LOQ} = d_2$, and we put $d_0 = 0$. Moreover, let N_1 be the number of observations below the d_1, N_2 be the number of observations in the interval $\left(d_1, d_2\right\rangle$, and N_0 be the number of uncensored observations $X_{(n-N_0+1)},\ldots,X_{(n)}$.

Using the results from [4], the likelihood function of the censored sample can be written in the form of

$$L = \frac{n!}{N_1! N_2!}\left[F(d_1)\right]^{N_1}\left[F(d_2) - F(d_1)\right]^{N_2} \prod_{i=n-N_0+1}^{n} f(X_{(i)}), \tag{3}$$

where the product $\prod_{i=n-N_0+1}^{n} f(X_{(i)})$ equals 1 for $N_0 = 0$. The ML estimating equations for estimating parameters λ and τ can be obtained from the log-likelihood function $l = \log L$ and are of the form of

$$\frac{\partial l}{\partial \lambda} = \sum_{i=1}^{2} N_i H_i^{\lambda} - N_0 \frac{\tau}{\lambda} + \frac{\tau}{\lambda^{\tau+1}} \sum_{i=n-N_0+1}^{n} X_{(i)}^{\tau} = 0 \tag{4}$$

$$\frac{\partial l}{\partial \tau}=\sum_{i=1}^{2} N_i H_i^{\tau}+N_0\frac{1-\tau\log\lambda}{\tau}+\sum_{i=n-N_0+1}^{n}\log X_{(i)} + \frac{\log\lambda}{\lambda^{\tau}}\sum_{i=n-N_0+1}^{n}X_{(i)}^{\tau}-\frac{1}{\lambda^{\tau}}\sum_{i=n-N_0+1}^{n}X_{(i)}^{\tau}\log X_{(i)}=0, \tag{5}$$

where H_i^{λ} (H_i^{τ} respectively), $i=1,2$, are the first derivatives of $\log[F(d_i)-F(d_{i-1})]$ with respect to the parameter λ (τ respectively). The ML estimates $\hat{\lambda}$ and $\hat{\tau}$ of parameters λ and τ can be obtained as a numerical solution of (4) and (5).

The histograms of musk compounds concentrations are highly skewed (see Fig. 2 for example, where phantolide concentration is); therefore, there is an idea of using simpler model and replace Weibull distribution with exponential distribution for modeling of musk compounds concentration.

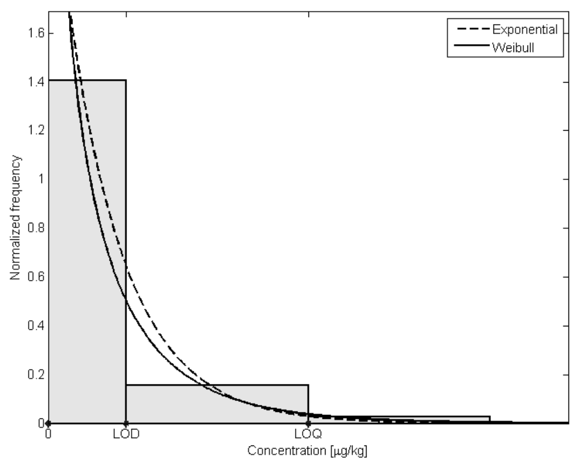

Fig. 2 Histogram of phantolide concentration with exponential (dashed line) and Weibull density (solid line).

In order to confirm suitability of exponential distribution, the asymptotic tests (see e.g. [13]) can be used. The null hypothesis is expressed as a constraint on the value of parameter τ of Weibull distribution (unrestricted model); thus under the null hypothesis $\mathrm{H}_0:\tau=1$ against the alternative $\mathrm{H}_1:\tau\neq 1$, the restricted model of exponential distribution is obtained. Although the parameter λ is not involved in the null hypothesis, it is necessary for description of the probability model. Therefore, it is called nuisance parameter. Let us define

$$U_1(\lambda,\tau)=\frac{\partial l(\lambda,\tau)}{\partial\tau},\, U_2(\lambda,\tau)=\frac{\partial l(\lambda,\tau)}{\partial\lambda}, \tag{6}$$

and

$$J(\lambda,\tau)=\begin{pmatrix} J_{11} & J_{12}\\ J_{21} & J_{22}\end{pmatrix} =\begin{pmatrix} -\frac{1}{n}\mathrm{E}\frac{\partial^2 \partial l(\lambda,\tau)}{\partial\tau^2} & -\frac{1}{n}\mathrm{E}\frac{\partial^2 \partial l(\lambda,\tau)}{\partial\tau\partial\lambda}\\ -\frac{1}{n}\mathrm{E}\frac{\partial^2 \partial l(\lambda,\tau)}{\partial\lambda\partial\tau} & -\frac{1}{n}\mathrm{E}\frac{\partial^2 \partial l(\lambda,\tau)}{\partial\lambda^2}\end{pmatrix}, \tag{7}$$

where J is the expected Fisher information matrix (FIM). As far as ML estimators are concerned, it is necessary to distinguish two different situations. In one case, parameters are estimated without additional conditions specified by the given null hypothesis. Let us denote the estimators as $\hat{\lambda}$ and $\hat{\tau}$. In the other case, parameters are estimated under the null hypothesis, thus $\tau=1$. In that case, log-likelihood function l is maximized with respect to parameter λ only and this estimator is denoted as $\tilde{\lambda}$.

Three asymptotic tests with nuisance parameter are distinguished, specifically the Lagrange multiplier (LM) test, likelihood ratio (LR) test and Wald (W) test. These tests are based on one of the three test statistics, specifically

$$LM=\frac{\tilde{U}_1^2}{n\tilde{J}_{11.2}}, \tag{8}$$

$$LR=2\left[l(\hat{\lambda},\hat{\tau})-l(\tilde{\lambda},1)\right], \tag{9}$$

$$W=(\hat{\tau}-1)^2 n\hat{J}_{11.2}, \tag{10}$$

where each of them have asymptotically χ^2 distribution with 1 degree of freedom. Function $\tilde{U}_1$ is a score function, which can be obtained from (5), evaluated in the parameter estimates under the null hypothesis (denoted by a tilde), and $J_{11.2}=J_{11}-J_{12}J_{22}^{-1}J_{21}$ is a transformation of the expected FIM evaluated in the parameter estimates under the null hypothesis (denoted by a tilde) or without any additional constraints (denoted by a hat). The elements of the expected FIM (7) are of the form of

$$nJ_{11}=-\sum_{i=1}^{2}H_i^{\tau\tau}(\lambda,\tau)\mathrm{E}(N_i)+\frac{1}{\tau^2}\mathrm{E}(N_0) +\frac{(\log\lambda)^2}{\lambda^{\tau}}\mathrm{E}\left(\sum_{i=n-N_0+1}^{n}X_{(i)}^{\tau}\right) -\frac{2\log\lambda}{\lambda^{\tau}}\mathrm{E}\left(\sum_{i=n-N_0+1}^{n}X_{(i)}^{\tau}\log X_{(i)}\right) +\frac{1}{\lambda^{\tau}}\mathrm{E}\left[\sum_{i=n-N_0+1}^{n}X_{(i)}^{\tau}(\log X_{(i)})^2\right], \tag{11}$$

$$nJ_{12} = nJ_{21} = -\sum_{i=1}^{2} H_i^{\lambda\tau}(\lambda,\tau)\mathrm{E}(N_i) + \frac{1}{\lambda}\mathrm{E}(N_0) + \frac{\tau\log\lambda - 1}{\lambda^{\tau+1}}\mathrm{E}\left(\sum_{i=n-N_0+1}^{n} X_{(i)}^{\tau}\right) - \frac{\tau}{\lambda^{\tau+1}}\mathrm{E}\left(\sum_{i=n-N_0+1}^{n} X_{(i)}^{\tau}\log X_{(i)}\right), \tag{12}$$

$$nJ_{22} = -\sum_{i=1}^{2} H_i^{\lambda\lambda}(\lambda,\tau)\mathrm{E}(N_i) - \frac{\tau}{\lambda^2}\mathrm{E}(N_0) + \frac{\tau^2+\tau}{\lambda^{\tau+2}}\mathrm{E}\left(\sum_{i=n-N_0+1}^{n} X_{(i)}^{\tau}\right). \tag{13}$$

Values of the second derivatives $H_i^{\lambda\lambda}$, $H_i^{\lambda\tau}$, $H_i^{\tau\tau}$ of $\log[F(d_i) - F(d_{i-1})]$ with respect to the parameters λ and τ, as well as the expected values

$$\mathrm{E}\left(\sum_{i=n-N_0+1}^{n} X_{(i)}^{\tau}\right), \mathrm{E}\left(\sum_{i=n-N_0+1}^{n} X_{(i)}^{\tau}\log X_{(i)}\right), \mathrm{E}\left[\sum_{i=n-N_0+1}^{n} X_{(i)}^{\tau}\left(\log X_{(i)}\right)^2\right],$$

can be found in [10].

Note that to perform the LM test, only estimation of the parameters subject to the restricted model is required. This is in contrast with the Wald test, which is based on unrestricted estimates, and likelihood ratio test, which requires both restricted and unrestricted estimates.

Further, we use the W test statistic as an example. It can be seen from Table 1 that exponential distribution is suitable for modeling of musk compounds concentration in almost all the cases. There is one exception – concentration of musk tibetene in Group 1 – where the null hypothesis is rejected at the significance level 0.05.

Thus further attention will be paid to the exponential model, because in most of cases it is possible to reduce Weibull model distribution to exponential.

Table 1: Wald test for assessing the suitability of the restricted exponential model against the unrestricted Weibull model; H = 0 (H = 1) denotes that the null hypothesis is not rejected (is rejected) at the significance level 0.05.

Compound	Group 1		Group 2	
	H	p-value	H	p-value
Phantolide	0	0.91	0	0.43
Traseolide	0	0.79	0	0.82
Musk ambrette	0	0.90	0	0.90
Musk tibetene	1	0.00	0	0.90

B. Exponential Distribution

From now on, $X_1, \ldots, X_n$ denotes a Type I doubly left-censored random sample from exponential distribution with scale parameter $\lambda > 0$, cdf

$$F(x,\lambda) = \begin{cases} 1 - \mathrm{e}^{-\frac{x}{\lambda}}, & x \ge 0 \\ 0, & x < 0 \end{cases}, \tag{14}$$

and pdf

$$f(x,\lambda) = \begin{cases} \frac{1}{\lambda}\mathrm{e}^{-\frac{x}{\lambda}}, & x \ge 0 \\ 0, & x < 0 \end{cases}. \tag{15}$$

The ML estimating equation for estimating parameter λ of the exponential distribution can be obtained from the log-likelihood function $l = \log L$, where L is given by (3). The ML estimating equation is of the form of

$$\frac{\partial l}{\partial \lambda} = \sum_{i=1}^{2} N_i H_i - \frac{N_0}{\lambda} + \frac{1}{\lambda^2}\sum_{i=n-N_0+1}^{n} X_{(i)} = 0, \tag{16}$$

where

$$H_i = \frac{d_{i-1}\exp\left(-\frac{d_{i-1}}{\lambda}\right) - d_i\exp\left(-\frac{d_i}{\lambda}\right)}{\lambda^2\left[\exp\left(-\frac{d_{i-1}}{\lambda}\right) - \exp\left(-\frac{d_i}{\lambda}\right)\right]}, i = 1,2. \tag{17}$$

Instead of solving (16), we can also maximize the log-likelihood function *l*.

In next section, test statistics for comparison of two censored samples from exponential distribution are proposed.

IV. Testing Equality of Two Censored Samples from Exponential Distribution

When comparing two censored samples from exponential distribution with parameters λ_1 and λ_2, classical test statistic *T* directly derived from the asymptotic normality of the estimates $\hat{\lambda}_1$ and $\hat{\lambda}_2$ can be used. This approach was applied in [11]. The null hypothesis $\mathrm{H}_0: \lambda_1 - \lambda_2 = 0$ is set against the alternative $\mathrm{H}_1: \lambda_1 - \lambda_2 \ne 0$, where λ_1 (λ_2 respectively) is the expected concentration of musk compound in Group 1 (Group 2 respectively). The test statistic *T* has asymptotically normal distribution $\mathrm{N}(0,1)$ and is of the form of

$$T = \frac{\hat{\lambda}_1 - \hat{\lambda}_2}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}}, \tag{18}$$

where $\hat{\sigma}_1^2$ ($\hat{\sigma}_2^2$ respectively) is the estimated variance of $\hat{\lambda}_1$ ($\hat{\lambda}_2$ respectively). The asymptotic variance $\hat{\sigma}_k^2$ is defined as

$$\hat{\sigma}_k^2 = \hat{J}_k^{-1}, k = 1,2, \tag{19}$$

where $\hat{J}_k$ is the expected FIM, which is derived in [9], evaluated in the parameter estimates and is of the form of

$$\begin{aligned} J_k = {} & \sum_{i=1}^{2} \frac{\left[d_{i-1}\exp\left(-\frac{d_{i-1}}{\lambda}\right) - d_i \exp\left(-\frac{d_i}{\lambda}\right)\right]^2}{\lambda^4 \left[\exp\left(-\frac{d_{i-1}}{\lambda}\right) - \exp\left(-\frac{d_i}{\lambda}\right)\right]} \\ & + \frac{\left(d_2^2 - 2d_2\lambda\right)\exp\left(-\frac{d_2}{\lambda}\right)}{\lambda^4} - \frac{1}{\lambda^2}\exp\left(-\frac{d_2}{\lambda}\right) \\ & + \frac{2}{\lambda^2}\sum_{n_0=0}^{n}\left\{ \binom{n}{n_0} \exp\left(-\frac{n_0 d_2}{\lambda}\right)\left[1 - \exp\left(-\frac{d_2}{\lambda}\right)\right]^{n-n_0} \right. \\ & \left. \times \sum_{i=n-n_0+1}^{n} \binom{n-1}{i-1} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \right\} \end{aligned} \tag{20}$$

The estimate $\hat{\lambda}_k$ of the parameter λ_k, $k = 1,2$, in (18) is obtained as a solution of (16).

Let us try a different approach and have two censored samples $X_{1,1},\ldots,X_{1,n}$ and $X_{2,1},\ldots,X_{2,n}$ from exponential distribution with parameters λ_1, λ_2 with cdf (14) and pdf (15). Furthermore, $X_{(j,1)},\ldots,X_{(j,n)}$, $j = 1,2$, again denotes the ordered sample $X_{j,1},\ldots,X_{j,n}$, and variables $N_{j,i}$ are frequencies corresponding to frequencies N_i, $i = 0,1,2$, from the previous section, where j denotes the number of the sample ($j = 1,2$). The log-likelihood function of two samples is of the form of

$$\begin{aligned} l(\lambda_1,\lambda_2) = {} & \sum_{j=1}^{2} \log\frac{n_j!}{N_{j,1}!N_{j,2}!} + N_{j,1}\log F(d_{j,1},\lambda_j) \\ & + N_{j,2}\log\left[F(d_{j,2},\lambda_j) - F(d_{j,1},\lambda_j)\right] \\ & + \sum_{i=n_j-N_{j,0}+1}^{n_j} \log f(X_{(j,i)}) \end{aligned} \tag{21}$$

Further, the reparametrisation $\lambda_1 = \lambda$ and $\lambda_2 = \lambda + \alpha$ will be used. It allows us to easily describe the power of presented tests as a function of parameter α. Then the log-likelihood function l_R of the new model is of the form of

$$l_R(\lambda,\alpha) = l(\lambda,\lambda+\alpha). \tag{22}$$

The ML estimates $\hat{\lambda}$ and $\hat{\alpha}$ of parameters λ and α can be obtained by maximization of (22).

When we want to compare two censored samples from exponential distribution, the asymptotic tests with nuisance parameter can again be used. The null hypothesis $\mathrm{H}_0 : \alpha = 0$ is set against the alternative $\mathrm{H}_1 : \alpha \neq 0$, thus λ is a nuisance parameter. The three test statistics have asymptotically χ^2 distribution with 1 degree of freedom and are of the form of

$$LM = \frac{\tilde{U}_1^2}{n\tilde{J}_{11.2}}, \tag{23}$$

$$LR = 2\left[l(\hat{\lambda},\hat{\alpha}) - l(\tilde{\lambda},0)\right], \tag{24}$$

$$W = \hat{\alpha}^2 n\hat{J}_{11.2}, \tag{25}$$

where $\tilde{U}_1$ is a score function and is of the form of

$$U_1 = \frac{\partial l_R}{\partial \alpha} = \sum_{i=1}^{2} N_{2,i}H_{2,i} - \frac{N_{2,0}}{\lambda+\alpha} + \frac{1}{(\lambda+\alpha)^2}\sum_{i=n_2-N_{2,0}+1}^{n_2} X_{(2,i)}, \tag{26}$$

with

$$H_{2,i} = \frac{d_{2,i-1}\exp\left(-\frac{d_{2,i-1}}{\lambda+\alpha}\right) - d_{2,i}\exp\left(-\frac{d_{2,i}}{\lambda+\alpha}\right)}{(\lambda+\alpha)^2\left[\exp\left(-\frac{d_{2,i-1}}{\lambda+\alpha}\right) - \exp\left(-\frac{d_{2,i}}{\lambda+\alpha}\right)\right]}, i = 1,2, \tag{27}$$

evaluated in the parameter estimates under the null hypothesis (denoted by a tilde). Quantity $J_{11.2} = J_{11} - J_{12}J_{22}^{-1}J_{21}$ is a transformation of the expected FIM evaluated in the parameter estimates under the null hypothesis (denoted by a tilde) or without any additional constraints (denoted by a hat). The elements of the expected FIM

$$\begin{aligned} J(\lambda,\alpha) &= \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix} \\ &= \begin{pmatrix} -\frac{1}{n}\mathrm{E}\frac{\partial^2 \partial l(\lambda,\tau)}{\partial\alpha^2} & -\frac{1}{n}\mathrm{E}\frac{\partial^2 \partial l(\lambda,\tau)}{\partial\alpha\partial\lambda} \\ -\frac{1}{n}\mathrm{E}\frac{\partial^2 \partial l(\lambda,\tau)}{\partial\lambda\partial\alpha} & -\frac{1}{n}\mathrm{E}\frac{\partial^2 \partial l(\lambda,\tau)}{\partial\lambda^2} \end{pmatrix} \end{aligned} \tag{28}$$

can be found in the appendix.

We can use the above mentioned tests for exploration of the

Table 2: Comparison of the expected concentrations of musk compounds between Group 1 and Group 2 considering various test statistics; H = 0 (H = 1) denotes that the null hypothesis is not rejected (is rejected) at the significance level 0.05.

Compound	T			LM			LR			W		
	H	statistic	p-value	H	statistic	p-value	H	statistic	p-value	H	statistic	p-value
Phantolide	0	0.73	0.47	0	0.60	0.44	0	0,57	0.45	0	0.53	0.47
Traseolide	0	-0.55	0.58	0	0.32	0.57	0	0.31	0.58	0	0.30	0.58
Musk ambrette	0	0.00	1.00	0	0.00	1.00	0	0.00	1.00	0	0.00	1.00
Musk tibetene	1	3,36	0.00	1	15.38	0.00	1	16.71	0.00	1	11.29	0.00

difference in concentration of the musk compounds between Group 1 and Group 2. The test results using test statistics (18), (23), (24), (25) are in Table 2, and estimates of parameters $\hat{\lambda}$, $\hat{\alpha}$, and $\tilde{\lambda}$ together with their standard deviations are in Table 3. All four tests gave equal results and it can be seen that there is no difference between Group 1 and Group 2 in expected concentrations of phantolide, traseolide and musk ambrette at the significance level 0.05. However, there is a significant difference between Group 1 and Group 2 in expected concentrations of musk tibetene. But the last result was obtained under the assumption of exponential distribution of samples, and this assumption was not confirmed by statistical tests at the first station. Therefore, the result is only approximate.

Table 3: Parameter estimates of the exponential model in the comparison of the expected concentrations of musk compounds together with their standard deviations *s*.

Compound	$\hat{\lambda} \pm s(\hat{\lambda})$	$\hat{\alpha} \pm s(\hat{\alpha})$	$\tilde{\lambda} \pm s(\tilde{\lambda})$
Phantolide	0.51±0.068	-0,10±0.088	0.46±0.069
Traseolide	0.67±0.111	0.12±0.170	0.73±0.112
Musk ambrette	0.54±0.096	0.00±0.136	0.54±0.096
Musk tibetene	0.20±0.014	-0.14±0.014	0.14±0.020

If we want to compare the performance of the above mentioned test statistics (18), (23), (24), and (25), it is necessary to compare their power functions.

V. Comparing of Simulated power functions

The power functions of tests based on statistics T, LM, LR, and W were compared by simulations (1000 repetitions). The calculations were carried out in Matlab software (version 7.12, R2011a) considering values of parameter λ from 0.2 to 1.5, which covers the estimated values $\hat{\lambda}$ for particular compounds. In order to assess the influence of censoring level on power of the tests, various LOD and LOQ values were chosen in simulations. Detection limits values were chosen as quantiles of the exponential distribution with cdf (14) using equations $q_{LOD} = F(LOD, \lambda)$ and $q_{LOQ} = F(LOQ, \lambda)$, where $q_{LOD} = 0.05, 0.25, 0.45$ and $q_{LOQ} = 0.10, 0.50, 0.90$. Powers of the tests were calculated for samples of length $n = 30, 100$.

Comparison of the simulated power functions shows (see Fig. 3 to 10) that LM test and LR test are more powerful than Wald test and classical test based on statistic (18). When the number of censored values is small (see Fig. 3, 5, 7), LR test is more powerful than LM test. However, when the number of censored values is large (see Fig. 4, 6, 8), as it is for data used in this paper, LM test is rather more powerful than LR test. The small differences in power functions between LM test and LR test can be observed for small sample size (n = 30) only (see Fig. 3 to 8). When the sample size is large (n = 100), the differences between the four tests are almost negligible (see Fig. 9, 10). It can be seen from Fig. 3 to 10 that power functions are nearly independent of the values of parameter λ.

Simulated power functions can further be used for assessing the tests powers considering various sample sizes and also for choosing a proper sample size when repeating the experiment.

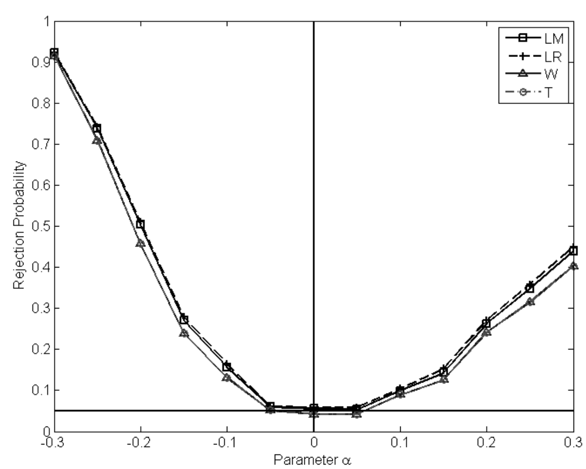

Fig. 3 Power functions of the tests; $\lambda = 0.50$, $n = 30$, $q_{LOD} = 0.05$, $q_{LOQ} = 0.10$.

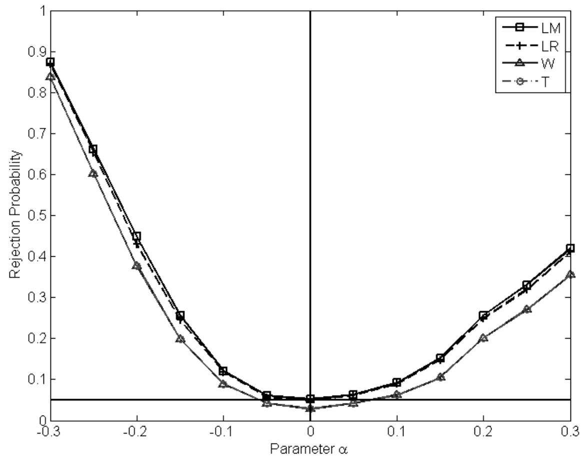

Fig. 4 Power functions of the tests; $\lambda = 0.50$, $n = 30$, $q_{LOD} = 0.45$, $q_{LOQ} = 0.90$.

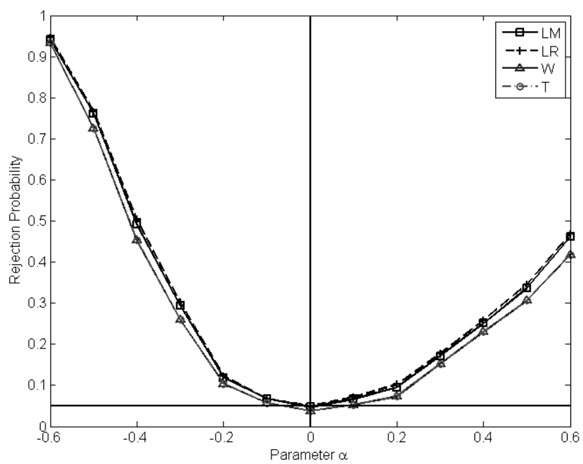

Fig. 5 Power functions of the tests; $\lambda = 1.0$, $n = 30$, $q_{LOD} = 0.05$, $q_{LOQ} = 0.10$.

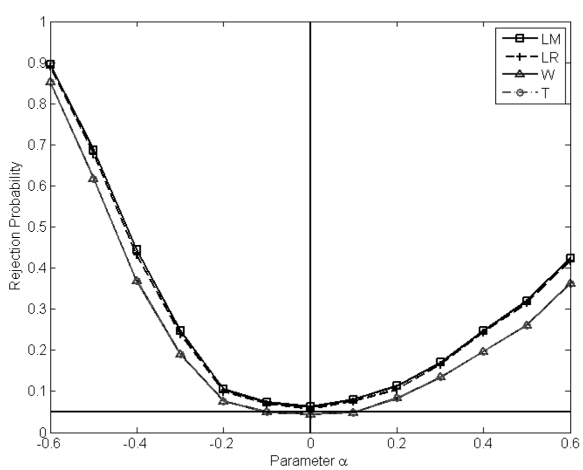

Fig. 6 Power functions of the tests; $\lambda = 1.0$, $n = 30$, $q_{LOD} = 0.45$, $q_{LOQ} = 0.90$.

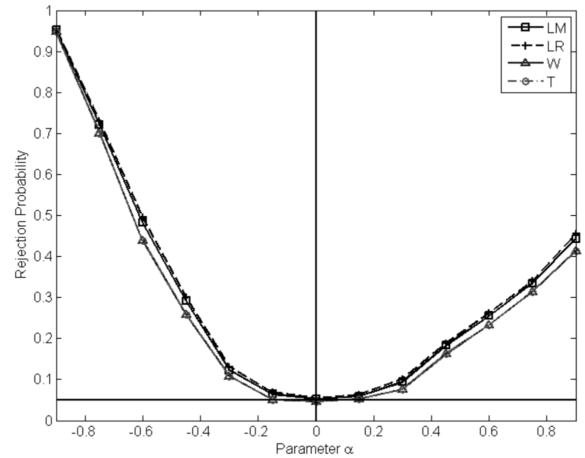

Fig. 7 Power functions of the tests; $\lambda = 1.5$, $n = 30$, $q_{LOD} = 0.05$, $q_{LOQ} = 0.10$.

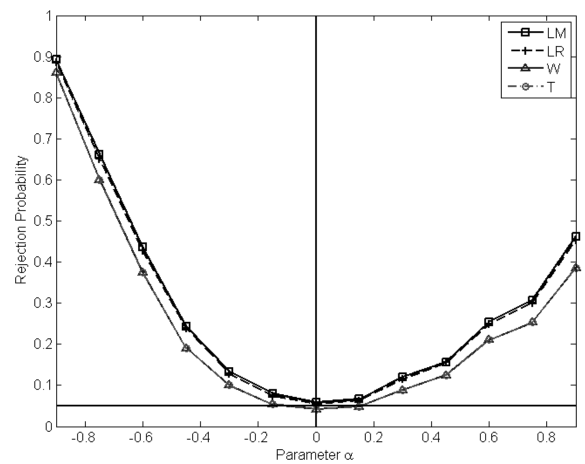

Fig. 8 Power functions of the tests; $\lambda = 1.5$, $n = 30$, $q_{LOD} = 0.45$, $q_{LOQ} = 0.90$.

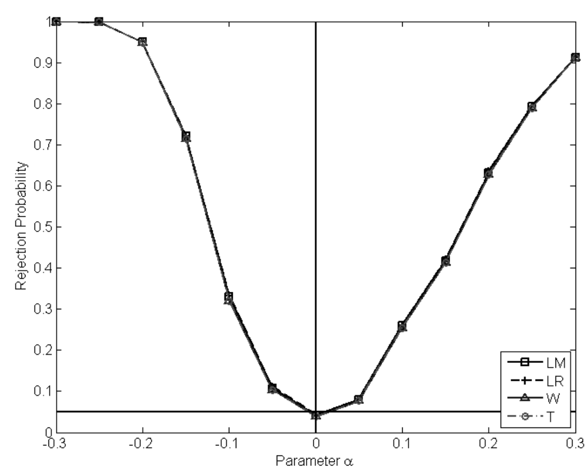

Fig. 9 Power functions of the tests; $\lambda = 0.5$, $n = 100$, $q_{LOD} = 0.05$, $q_{LOQ} = 0.10$.

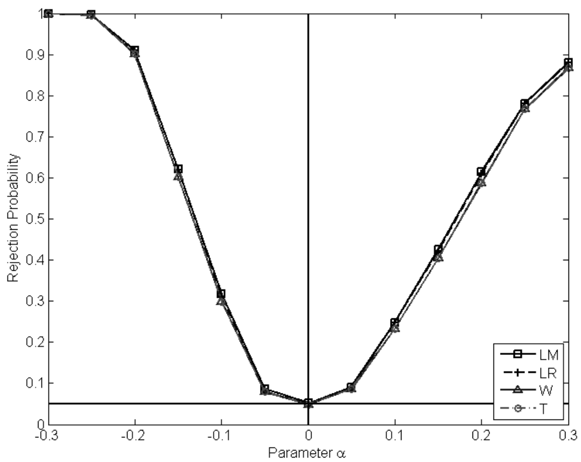

Fig. 10 Power functions of the tests; $\lambda = 0.5$, $n = 100$, $q_{LOD} = 0.45$, $q_{LOQ} = 0.90$.

VI. Conclusion

This contribution was focused on the statistical methods for analyzing musk compounds. The recently developed method for processing of doubly left-censored samples from exponential and Weibull distribution considering Type I censoring was used. Considering the shape of musk compound concentration histograms, exponential distribution as a model distribution was proposed. Three asymptotic tests (Lagrange multiplier test, likelihood ratio test, Wald test) for assessing suitability of the exponential distribution were suggested. Using asymptotic tests and properties of ML estimates, methods for comparison of two censored samples from exponential distribution were proposed and used in analysis of concentrations of musk compounds extracted from the fish samples caught in front of and behind the WWTP. It was discovered that there is no significant difference between Group 1 and Group 2 in expected concentrations of phantolide, traseolide and musk ambrette. However, there is a difference between Group 1 and Group 2 in expected concentrations of musk tibetene. But the last result was obtained under the assumption of exponential distribution of samples which was not confirmed by the statistical test in Group 1. Thus the result is approximate only.

The comparison of power functions of particular tests showed that when analyzing musk compounds concentrations, LM and LR tests should be preferred. The simulated power functions can also be used for assessing the probability, with which the tests detect true differences between two exponential populations considering sample sizes n = 30 and 100.

All the algorithms and Matlab m-files can be obtained from first author.

Appendix

Elements of the expected FIM (28) are of the form of

$$
\begin{aligned}
J_{11} &= J_{12} = J_{21} \\
&= \sum_{i=1}^{2} \frac{\left[d_{2,i-1}\exp\left(-\frac{d_{2,i-1}}{\lambda+\alpha}\right) - d_{2,i}\exp\left(-\frac{d_{2,i}}{\lambda+\alpha}\right)\right]^2}{(\lambda+\alpha)^4\left[\exp\left(-\frac{d_{2,i-1}}{\lambda+\alpha}\right) - \exp\left(-\frac{d_{2,i}}{\lambda+\alpha}\right)\right]} \\
&+ \frac{\left[d_{2,2}^2 - 2d_{2,2}(\lambda+\alpha)\right]\exp\left(-\frac{d_{2,2}}{\lambda+\alpha}\right)}{(\lambda+\alpha)^4} \\
&- \frac{1}{(\lambda+\alpha)^2}\exp\left(-\frac{d_{2,2}}{\lambda+\alpha}\right) + \frac{2}{(\lambda+\alpha)^2}\sum_{n_0=0}^{n_2}\left\{\binom{n_2}{n_0}\right. \\
&\times \exp\left(-\frac{n_0 d_{2,2}}{\lambda+\alpha}\right)\left[1-\exp\left(-\frac{d_{2,2}}{\lambda+\alpha}\right)\right]^{n_2-n_0} \\
&\times \left.\sum_{i=n_2-n_0+1}^{n_2}\binom{n_2-1}{i-1}\sum_{j=0}^{i-1}(-1)^j\binom{i-1}{j}(n_2-i+j+1)^{-2}\right\}
\end{aligned}
\quad , \qquad (29)
$$

$$
\begin{aligned}
J_{22} &= \sum_{i=1}^{2} \frac{\left[d_{1,i-1}\exp\left(-\frac{d_{1,i-1}}{\lambda}\right) - d_{1,i}\exp\left(-\frac{d_{1,i}}{\lambda}\right)\right]^2}{\lambda^4\left[\exp\left(-\frac{d_{1,i-1}}{\lambda}\right) - \exp\left(-\frac{d_{1,i}}{\lambda}\right)\right]} \\
&+ \frac{\left(d_{1,2}^2 - 2d_{1,2}\lambda\right)\exp\left(-\frac{d_{1,2}}{\lambda}\right)}{\lambda^4} - \frac{1}{\lambda^2}\exp\left(-\frac{d_{1,2}}{\lambda}\right) \\
&+ \frac{2}{\lambda^2}\sum_{n_0=0}^{n_1}\left\{\binom{n_1}{n_0}\exp\left(-\frac{n_0 d_{1,2}}{\lambda}\right)\left[1-\exp\left(-\frac{d_{1,2}}{\lambda}\right)\right]^{n_1-n_0}\right. \\
&\times \left.\sum_{i=n_1-n_0+1}^{n_1}\binom{n_1-1}{i-1}\sum_{j=0}^{i-1}(-1)^j\binom{i-1}{j}(n_1-i+j+1)^{-2}\right\} \\
&+ J_{11}
\end{aligned}
\quad . \qquad (30)
$$

Acknowledgment

The paper was written with the support of the specific research project FSI-S-11-3 No. 22160

References

[1] A. Andronov, N. Kolmakova and I. Yatskiv, "A quasi regression model for polytomous data and its application for measuring service quality," *International Journal of Mathematics and Computers in Simulation*, vol. 4 (2), pp. 50–57, 2010.

[2] K. Bester, "Analysis of musk fragrances in environmental samples," *Journal of Chromatography A,* vol. 1216 (3), pp. 470–480, 2009.

[3] D. Bílková, "Lognormal distribution and using L-moment method for estimating its parameters," *International Journal of Mathematical Models and Methods in Applied Sciences*, vol. 6 (1), pp. 30–44, 2012.

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

[5] D. R. Cox and D. Oakes, *Analysis of Survival Data*. New York: Chapman and Hall/CRC, 1984.

[6] R. Draisci, C. Marchiafava, E. Ferretti, L. Palleschi, G. Catellani and A. Anastasio, “Evaluation of musk contamination of freshwater fish in Italy by accelerated solvent extraction and gas chromatography with mass spectrometric detection,” *Journal of Chromatography A*, vol. 814 (1–2), pp. 187–197, 1998.

[7] A. H. El-Shaarawi and D. M. Dolan, “Maximum likelihood estimation of water concentrations from censored data,” *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 46, pp. 1033–1039, 1989.

[8] A. H. El-Shaarawi and S. R. Esterby, “Replacement of censored observations by a constant: An evaluation,” *Water Research*, vol. 26, pp. 835–844, 1992.

[9] M. Fusek and J. Michálek, “Statistical analysis of doubly left-censored samples from exponential distribution,” *In MENDEL 2012, 18th International Conference on Soft Computing, Mendel Journal series*, pp. 564–569, 2012.

[10] M. Fusek and J. Michálek, “Statistical inference of doubly left-censored samples from Weibull distribution,” In *XX International Conference PDMU-2012 Problems of Decision Making under Uncertainties Proceedings - Applied Papers*, pp. 31–40, 2012.

[11] M. Fusek, J. Michálek, L. Zouhar and M. Vávrová, “Statistical Analysis of Musk Compounds Concentrations in Fish Tissue Based on Doubly Left-Censored Samples,” *3rd International Conference on Development, Energy, Environment, Economics (DEEE '12)*, pp. 247–250, 2012.

[12] E. L. Lehmann and G. Casella, *Theory of point estimation*. 2nd ed., New York: Springer-Verlag, 1998.

[13] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses.* 3rd ed., New York: Springer, 2005.

[14] S. Lignell, P. O. Darnerud, M. Aune, S. Cnattingius, J. Hajslova, L. Setkova and A. Glynn, “Temporal trends of synthetic musk compounds in mother's milk and associations with personal use of perfumed products,” *Environ. Sci. Technol.*, vol. 42 (17), pp. 6743–6748, 2008.

[15] D. Prathumwan, Y. Lenbury, P. Satiracoo and C. Rattanakul, “Euler-Maruyama approximation and maximum likelihood estimator for a stochastic differential equation model of the signal transduction process,” *International Journal of Mathematical Models and Methods in Applied Sciences*, vol. 6 (2), pp. 323–331, 2012.

[16] L. Zouhar, M. Vávrová, M. Fusek and J. Michálek, “Evaluation of the water biota contamination by musk compounds,” *7th European Conference on Pesticides and Related Organic Micropollutants in the Environment + 13th Symposium on Chemistry and Fate of Modern Pesticides - Book of Abstracts*, pp. 263–264, 2012.