VĚDECKÉ SPISY VYSOKÉHO UČENÍ TECHNICKÉHO V BRNĚ

Edice Habilitační a inaugurační spisy, sv. 805

ISSN 1213-418X

Michal Fusek

LEFT-CENSORED PROBABILITY DISTRIBUTIONS AND THEIR APPLICATIONS

BRNO UNIVERSITY OF TECHNOLOGY
Faculty of Electrical Engineering and Communication
Department of Mathematics

Ing. Michal Fusek, Ph.D.

LEFT-CENSORED PROBABILITY DISTRIBUTIONS AND THEIR APPLICATIONS

ZLEVA CENZOROVANÁ ROZDĚLENÍ PRAVDĚPODOBNOSTI A JEJICH APLIKACE

ZKRÁCENÁ VERZE HABILITAČNÍ PRÁCE
V OBORU ELEKTROTECHNIKA A KOMUNIKAČNÍ TECHNOLOGIE

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
NAKLADATELSTVÍ VUTIUM

BRNO 2024

KEYWORDS
biogenic amines, left censoring, exponential distribution, maximum likelihood, musk compounds, goodness-of-fit tests, Weibull distribution

KLÍČOVÁ SLOVA
biogenní aminy, cenzorovaní zleva, exponenciální rozdělení, maximální věrohodnost, musk sloučeniny, testy dobré shody, Weibullovo rozdělení

MÍSTO ULOŽENÍ PRÁCE
Fakulta elektrotechniky a komunikačních technologií
Vysokého učení technického v Brně

ISBN 978-80-214-6314-1
ISSN 1213-418X

Contents

ABOUT THE AUTHOR

Ing. Michal Fusek, Ph.D.
Department of Mathematics
Faculty of Electrical Engineering and Communication
Brno University of Technology
Technická 3058/10
616 00 Brno
Czech Republic
Phone: +420 541 146 096
E-mail: fusekmi@vut.cz
ORCID: 0000-0002-9842-8384
Researcher ID: E-4920-2018
SCOPUS ID: 39161155400

Michal Fusek was born in 1984 in Zlín, Czech Republic. He graduated with red diploma in master's degree program called Mathematical engineering at the Faculty of Mechanical Engineering (FME), Brno University of Technology (BUT) in 2009. Then he continued to study a doctoral programme called Applied mathematics at the Institute of Mathematics, FME, BUT, and defended his PhD thesis entitled *Extreme value distributions with applications* in 2013. Since then, he is a part of the academic staff at the Department of Mathematics, Faculty of Electrical Engineering and Communication (FEEC), BUT.

As a part of his pedagogical activities, he taught various mathematical courses at FME, FEEC, Faculty of Information Technology (FIT) and Institute of Forensic Engineering (IFE) since 2009. He is a co-founder of 2 courses, specifically Probability and Statistics (IPT, FIT) and Mathematics (ESMAT, IFE). He is an author or a co-author of a number of teaching materials such as textbooks, exercise books and presentations for university students. He also regularly visits foreign universities under the Erasmus+ programme. He supervised 1 foreign student (University of L'Aquila) who successfully defended their Master's thesis. Since 2021, he is a board member of the master's degree programme Expert Engineering in Traffic at IFE, BUT. He is a member of the organizing and steering committees at the EEICT conference and is responsible for the Mathematical sciences section.

Research activities of the author are focused on statistical methods and their applications in various scientific fields. His work is primarily focused on censored probability distributions and on the theory of extreme value distributions. He participated in 6 research projects and 3 educational projects.

Michal Fusek has published 13 articles in journals with impact factor and one of them was selected for the SKV (Systém Kvalitních Výsledků) in 2023. He has 19 articles indexed in Web of Science (WoS) and 26 articles indexed in Scopus. His articles have 76 citations (excluding self-citations) according to WoS and 85 citations according to Scopus. His H-index is 6 according to WoS and 7 according to Scopus.

1 INTRODUCTION

Censored data occur frequently in many application areas. When performing an experiment, we can sometimes run into a situation that some of the measured values can be reported only as less than some value, greater than some value, or as an interval. In such cases we talk about left-censored, right-censored and interval-censored data. There are two basic types of censoring, specifically the Type I and the Type II censoring.

The Type I censoring (time censoring) typically occurs in experiments that stop at a prespecified time. The censoring level is known in advance and the number of censored values is a random variable. For example, we have a certain number of light bulbs and study if they fail before a prespecified time. The light bulbs that have not failed are Type I right-censored. Another frequently used application of the Type I censoring is as follows. We measure concentrations of a chemical compound in a sample and our instrument (determination method, respectively) is not able to measure the concentrations below a specified (detection) limit with a stated accuracy and precision. Such observations are called Type I left-censored.

The Type II censoring (failure censoring) typically occurs in experiments that stop when a prespecified number of failures are observed. The number of censored values is known in advance and the censoring level is a random variable. As an example, we can modify our light bulb experiment. Now we have a certain number of light bulbs and study at what time a prespecified number of them fails. The light bulbs that have not failed are Type II right-censored. Another application of the Type II censoring is as follows. We conduct a study where the event of interest, e.g. infection with a sexually transmitted disease, has already taken place at the time when the study starts, but the exact time of occurrence of the event is not known. The exact time when the sickness started is Type II left-censored. When analyzing real data, it is often necessary to work with more than one censoring level. In such case we talk about singly, doubly, or even multiply censored data.

The habilitation thesis is focused on statistical methods for Type I multiply left-censored data. A special attention is paid to the censored Weibull distribution, which is very flexible and can be used for modelling of various engineering problems. With regards to the terminology that is usually used in applications of these methods, censoring levels will be called detection limits (DLs).

In general, it is necessary to be cautious when dealing with left-censored data. As pointed out by Helsel (2012), an unsuitable approach can influence the results significantly. When the censored values are ignored, a certain amount of information that can be obtained from the data is lost. Moreover, such an approach yields biased estimates of parameters in the model. For example, when censored values are omitted, the mean concentration of a chemical compound is going to be overestimated. A common practice to circumvent this issue is to replace censored values under the detection limits with a constant (e.g., 0, DL/2, DL/$\sqrt{2}$, $\sqrt{\text{DL}}$, DL), and to analyze data with traditional methods such that the substituted values are assumed to be observed (Guérin et al., 2011; Hoelzer et al., 2014; Munoz et al., 2015; Struciński et al., 2015). More information about how substituting values for censored observations can affect the results can be found in Helsel (2006).

2 MULTIPLY LEFT-CENSORED WEIBULL DISTRIBUTION

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

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

probability density function (pdf)

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

and expected value

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

where Γ is the gamma function.

Let $X_{(1)} \leq \cdots \leq X_{(n)}$ be the ordered sample of $X_1, \ldots, X_n$ which is Type I multiply left-censored with detection limits $d_1, \ldots, d_k$ and we put $d_0 = 0$. Moreover, N_i is the number of observations in the interval $(d_{i-1}, d_i]$, $i = 1, \ldots, k$, and N_0 is the number of uncensored observations $X_{(n-N_0+1)}, \ldots, X_{(n)}$. In order to simplify notation in some formulas, we replace $\log(x)$ by zero in case the natural logarithm is undefined.

Using results from Cohen (1991), the log-likelihood function of the Type I multiply left-censored sample can be written as

$$\begin{aligned} & l(\lambda, \tau, N_0, \ldots, N_k, X_{(n-N_0+1)}, \ldots, X_{(n)}) \\ & \quad = \log\left(\frac{n!}{N_1! \cdots N_k!}\right) + \sum_{i=1}^{k} N_i \log\left\{\exp\left[-\left(\frac{d_{i-1}}{\lambda}\right)^{\tau}\right] - \exp\left[-\left(\frac{d_i}{\lambda}\right)^{\tau}\right]\right\} \\ & \qquad + N_0 \log\left(\frac{\tau}{\lambda^{\tau}}\right) + (\tau - 1) \sum_{i=n-N_0+1}^{n} \log\left(X_{(i)}\right) - \frac{1}{\lambda^{\tau}} \sum_{i=n-N_0+1}^{n} X_{(i)}^{\tau}. \end{aligned} \tag{4}$$

and we put $\sum_{i=n-N_0+1}^{n} X_{(i)}^{\tau} = 0$ for $N_0 = 0$. The ML estimates $\widehat{\lambda}$, $\widehat{\tau}$ of parameters λ, τ can be obtained by maximization of the log-likelihood function (4) using the Nelder-Mead simplex algorithm (Lagarias et al., 1998) in Matlab (version R2022b).

In order to calculate variability of the ML estimates $\widehat{\lambda}$ and $\widehat{\tau}$, the expected Fisher information matrix (FIM) can be used. According to Barndorff-Nielsen and Cox (1994), the expected FIM can be defined (under certain regularity conditions) using formula

$$\boldsymbol{J}_n(\lambda, \tau) = \begin{bmatrix} -\mathrm{E}\frac{\partial^2 l}{\partial \lambda^2} & -\mathrm{E}\frac{\partial^2 l}{\partial \lambda \partial \tau} \\ -\mathrm{E}\frac{\partial^2 l}{\partial \tau \partial \lambda} & -\mathrm{E}\frac{\partial^2 l}{\partial \tau^2} \end{bmatrix} = n\boldsymbol{J}(\lambda, \tau) = n \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix}, \tag{5}$$

where

$$
\begin{aligned}
J_{11} = & \sum_{i=1}^{k} \frac{\tau^2 \left\{ d_{i-1}^{\tau} \exp\left[-\left(\frac{d_{i-1}}{\lambda}\right)^{\tau} \right] - d_i^{\tau} \exp\left[-\left(\frac{d_i}{\lambda}\right)^{\tau} \right] \right\}^2}{\lambda^{2\tau+2} \left\{ \exp\left[-\left(\frac{d_{i-1}}{\lambda}\right)^{\tau} \right] - \exp\left[-\left(\frac{d_i}{\lambda}\right)^{\tau} \right] \right\}} \\
& - \frac{(d_k^{\tau}\lambda^{\tau}\tau^2 + d_k^{\tau}\lambda^{\tau}\tau - d_k^{2\tau}\tau^2) \exp\left[-\left(\frac{d_k}{\lambda}\right)^{\tau} \right]}{\lambda^{2\tau+2}} - \frac{\tau}{\lambda^2} \exp\left[-\left(\frac{d_k}{\lambda}\right)^{\tau} \right] \\
& + \frac{\tau^2 + \tau}{\lambda^2} \sum_{n_0=0}^{n} \sum_{i=n-n_0+1}^{n} \binom{n-1}{i-1} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \\
& \times \binom{n}{n_0} \exp\left[-n_0 \left(\frac{d_k}{\lambda}\right)^{\tau} \right] \left\{ 1 - \exp\left[-\left(\frac{d_k}{\lambda}\right)^{\tau} \right] \right\}^{n-n_0},
\end{aligned}
$$

$$
\begin{aligned}
J_{22} = & \sum_{i=1}^{k} \frac{\left\{ d_{i-1}^{\tau} \log\left(\frac{d_{i-1}}{\lambda}\right) \exp\left[-\left(\frac{d_{i-1}}{\lambda}\right)^{\tau} \right] - d_{i-1}^{\tau} \log\left(\frac{d_i}{\lambda}\right) \exp\left[-\left(\frac{d_i}{\lambda}\right)^{\tau} \right] \right\}^2}{\lambda^{2\tau} \left\{ \exp\left[-\left(\frac{d_{i-1}}{\lambda}\right)^{\tau} \right] - \exp\left[-\left(\frac{d_i}{\lambda}\right)^{\tau} \right] \right\}} \\
& - \frac{(d_k^{\tau}\lambda^{\tau} - d_k^{2\tau}) \left[\log\left(\frac{d_k}{\lambda}\right) \right]^2 \exp\left[-\left(\frac{d_k}{\lambda}\right)^{\tau} \right]}{\lambda^{2\tau}} + \frac{1}{\tau^2} \exp\left[-\left(\frac{d_k}{\lambda}\right)^{\tau} \right] \\
& + [\log(\lambda)]^2 \sum_{n_0=0}^{n} \sum_{i=n-n_0+1}^{n} \binom{n-1}{i-1} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \\
& \times \binom{n}{n_0} \exp\left[-n_0 \left(\frac{d_k}{\lambda}\right)^{\tau} \right] \left\{ 1 - \exp\left[-\left(\frac{d_k}{\lambda}\right)^{\tau} \right] \right\}^{n-n_0} \\
& - \frac{2\log(\lambda)}{\tau} \sum_{n_0=0}^{n} \sum_{i=n-n_0+1}^{n} \binom{n-1}{i-1} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \\
& \times \left[\log\left(\frac{\lambda^{\tau}}{n-i+j+1}\right) + 1 - \gamma_e \right] \\
& \times \binom{n}{n_0} \exp\left[-n_0 \left(\frac{d_k}{\lambda}\right)^{\tau} \right] \left\{ 1 - \exp\left[-\left(\frac{d_k}{\lambda}\right)^{\tau} \right] \right\}^{n-n_0} \\
& + \frac{1}{\tau^2} \sum_{n_0=0}^{n} \sum_{i=n-n_0+1}^{n} \binom{n-1}{i-1} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \\
& \times \left\{ \left[\log\left(\frac{\lambda^{\tau}}{n-i+j+1}\right) \right]^2 + 2\log\left(\frac{\lambda^{\tau}}{n-i+j+1}\right)(1-\gamma_e) \right. \\
& \left. + \frac{\pi^2}{6} - 2\gamma_e + \gamma_e^2 \right\} \binom{n}{n_0} \exp\left[-n_0 \left(\frac{d_k}{\lambda}\right)^{\tau} \right] \left\{ 1 - \exp\left[-\left(\frac{d_k}{\lambda}\right)^{\tau} \right] \right\}^{n-n_0},
\end{aligned}
$$

$$
\begin{aligned}
J_{12} = J_{21} = &- \sum_{i=1}^{k} \frac{\tau \left\{ d_{i-1}^{\tau} \exp\left[-\left(\frac{d_{i-1}}{\lambda}\right)^{\tau} \right] - d_i^{\tau} \exp\left[-\left(\frac{d_i}{\lambda}\right)^{\tau} \right] \right\}}{\lambda^{2\tau+1} \left\{ \exp\left[-\left(\frac{d_{i-1}}{\lambda}\right)^{\tau} \right] - \exp\left[-\left(\frac{d_i}{\lambda}\right)^{\tau} \right] \right\}} \\
&\times \left\{ d_{i-1}^{\tau} \log\left(\frac{d_{i-1}}{\lambda}\right) \exp\left[-\left(\frac{d_{i-1}}{\lambda}\right)^{\tau} \right] - d_i^{\tau} \log\left(\frac{d_i}{\lambda}\right) \exp\left[-\left(\frac{d_i}{\lambda}\right)^{\tau} \right] \right\} \\
&+ \frac{\left[d_k^{\tau} \tau \lambda^{\tau} \log\left(\frac{d_k}{\lambda}\right) + d_k^{\tau} \lambda^{\tau} - d_k^{2\tau} \tau \log\left(\frac{d_k}{\lambda}\right) \right] \exp\left[-\left(\frac{d_k}{\lambda}\right)^{\tau} \right]}{\lambda^{2\tau+1}} + \frac{1}{\lambda} \exp\left[-\left(\frac{d_k}{\lambda}\right)^{\tau} \right] \\
&+ \frac{\tau \log(\lambda) - 1}{\lambda} \sum_{n_0=0}^{n} \sum_{i=n-n_0+1}^{n} \binom{n-1}{i-1} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \\
&\times \binom{n}{n_0} \exp\left[-n_0 \left(\frac{d_k}{\lambda}\right)^{\tau} \right] \left\{ 1 - \exp\left[-\left(\frac{d_k}{\lambda}\right)^{\tau} \right] \right\}^{n-n_0} \\
&- \frac{1}{\lambda} \sum_{n_0=0}^{n} \sum_{i=n-n_0+1}^{n} \binom{n-1}{i-1} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \\
&\times \left[\log\left(\frac{\lambda^{\tau}}{n-i+j+1}\right) + 1 - \gamma_e \right] \\
&\times \binom{n}{n_0} \exp\left[-n_0 \left(\frac{d_k}{\lambda}\right)^{\tau} \right] \left\{ 1 - \exp\left[-\left(\frac{d_k}{\lambda}\right)^{\tau} \right] \right\}^{n-n_0}.
\end{aligned}
$$

Details of the derivation can be found in the habilitation thesis. Considering the asymptotic properties of ML estimator $\widehat{\lambda}$ ($\widehat{\tau}$ respectively), according to Lehmann and Casella (1998), $\sqrt{n}(\widehat{\lambda}-\lambda)$ ($\sqrt{n}(\widehat{\tau}-\tau)$ respectively) has asymptotically normal distribution $\mathrm{N}(0, J^{11})$ ($\mathrm{N}(0, J^{22})$ respectively), where J^{11}, J^{22} are diagonal elements of the variance matrix $\boldsymbol{J}^{-1}(\lambda, \tau)$. The properties of estimators $\widehat{\lambda}$, $\widehat{\tau}$ considering various sample sizes n, various number of detection limits k and various censoring schemes were analyzed in Fusek and Michálek (2019) and can be found in the habilitation thesis.

2.1 Confidence Intervals for Expectation

In general, in order to estimate a statistic (e.g., mean, standard deviation) of a population, a random sample from the population is taken, and the individual statistics are calculated. In the field of environmental sciences or chemistry, it is often of interest to estimate the expected concentration of a chemical compound. In case the concentration can be modeled using the Weibull distribution, the expected value $\mu = \mu(\lambda, \tau)$ can be calculated using formula (3). In order to calculate the estimate $\widehat{\mu}$ of the expected value μ, we can replace λ, τ in (3) by their ML estimates $\widehat{\lambda}$, $\widehat{\tau}$.

However, it is always an issue to assess how well the sample statistic estimates the underlying population value. For this purpose, a confidence interval is used because it provides a range of values which is likely to contain the population parameter of interest. Considering the asymptotic normality of ML estimates, the expected value has the asymptotically normal distribution (Likeš and Machek, 1988), specifically

$$
\mu(\lambda, \tau) \overset{A}{\sim} \mathrm{N}\left(\mu(\lambda, \tau), \mathrm{Var}\left(\mu(\lambda, \tau)\right)\right),
$$

where

$$\begin{aligned}\operatorname{Var}(\mu(\lambda,\tau)) &= \left[\frac{\partial\mu(\lambda,\tau)}{\partial\lambda}, \frac{\partial\mu(\lambda,\tau)}{\partial\tau}\right] \boldsymbol{J}_n^{-1}(\lambda,\tau) \left[\frac{\partial\mu(\lambda,\tau)}{\partial\lambda}, \frac{\partial\mu(\lambda,\tau)}{\partial\tau}\right]^{\mathrm{T}} \\ &= \left[\Gamma\left(1+\frac{1}{\tau}\right), -\frac{\lambda\Psi\left(1+\frac{1}{\tau}\right)\Gamma\left(1+\frac{1}{\tau}\right)}{\tau^2}\right] \boldsymbol{J}_n^{-1}(\lambda,\tau) \\ &\quad \times \left[\Gamma\left(1+\frac{1}{\tau}\right), -\frac{\lambda\Psi\left(1+\frac{1}{\tau}\right)\Gamma\left(1+\frac{1}{\tau}\right)}{\tau^2}\right]^{\mathrm{T}}\end{aligned} \tag{6}$$

and $\Psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function.

Considering $\widehat{\mu} = \mu(\widehat{\lambda}, \widehat{\tau})$ and $\widehat{\operatorname{Var}}(\widehat{\mu}) = \operatorname{Var}\left(\mu(\widehat{\lambda}, \widehat{\tau})\right)$, the asymptotic $(1-\alpha)\%$ confidence interval for μ can be calculated as

$$\left(\widehat{\mu} - z_{1-\frac{\alpha}{2}}\sqrt{\widehat{\operatorname{Var}}(\widehat{\mu})}, \widehat{\mu} + z_{1-\frac{\alpha}{2}}\sqrt{\widehat{\operatorname{Var}}(\widehat{\mu})}\right),$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution $\mathrm{N}(0,1)$. The range and the coverage probability of the confidence interval was assessed using simulations and compared with the bootstrap method in Fusek and Michálek (2016) and can be found in the habilitation thesis.

2.2 Reduction of Weibull to Exponential Distribution

There are situations when the Weibull distribution is too complicated for modelling of given data. In case $\tau = 1$, the model of Weibull distribution can be reduced to the exponential submodel where all the calculations are much easier. To assess suitability of replacement of the censored Weibull distribution with the exponential distribution, asymptotic tests with nuisance parameters can be used (see e.g. Lehmann and Romano, 2005), specifically the Lagrange multiplier (LM) test, the Wald (W) test and the likelihood ratio (LR) test.

The null hypothesis H_0 is expressed as a restriction on the shape parameter τ of the censored Weibull distribution. Specifically, $\mathrm{H}_0 : \tau = 1$ is set against the alternative $\mathrm{H}_1 : \tau \neq 1$, and λ is the nuisance parameter. In case the null hypothesis is not rejected at a specified significance level, the censored exponential distribution can be used instead of the Weibull distribution.

The test statistics are

$$\begin{aligned} LM &= \frac{U_1^2(\widetilde{\lambda}, 1)}{J_{n,22.1}(\widetilde{\lambda}, 1)}, \\ W &= (\widehat{\tau} - 1)^2 J_{n,22.1}(\widehat{\lambda}, \widehat{\tau}), \\ LR &= 2\left[l(\widehat{\lambda}, \widehat{\tau}) - l(\widetilde{\lambda}, 1)\right], \end{aligned} \tag{7}$$

where

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

$$+ N_0 \frac{(1 - \tau \ln \lambda)}{\tau} + \sum_{i=n-N_0+1}^{n} \ln X_{(i)} + \frac{\ln \lambda}{\lambda^\tau} \sum_{i=n-N_0+1}^{n} X_{(i)}^\tau$$

$$- \frac{1}{\lambda^\tau} \sum_{i=n-N_0+1}^{n} X_{(i)}^\tau \ln X_{(i)}$$

is the score function and $J_{n,22.1}(\lambda, \tau) = n(J_{22} - J_{21}J^{11}J_{12})$ is a transformation of the expected FIM (5). The parameters estimated under the null hypothesis are denoted by tilde, and those estimated under the alternative are denoted by hat. Under the null hypothesis, the test statistics (7) have asymptotically χ^2 distribution with one degree of freedom (see e.g. Lehmann and Romano, 2005). The null hypothesis is rejected at a prescribed significance level when the test statistics exceed the critical value of the χ^2 distribution. Performance of the tests using statistics (7) was assessed by means of simulated power functions in Fusek (2017) and can be found in the habilitation thesis.

2.3 Comparison of Two Left-Censored Weibull Samples

In order to compare two independent Type I multiply left-censored samples from the Weibull distribution, we extend the one-sample model from the previous sections to the two-sample model. Let $X_{j,1}, \ldots, X_{j,n_j}$, $j = 1, 2$, be two independent Type I multiply left-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 ($j = 2$). The ordered sample of $X_{j,1}, \ldots, X_{j,n_j}$, $j = 1, 2$, is denoted as $X_{j,(1)} \leq \cdots \leq X_{j,(n_j)}$, and $N_{j,i}$ are frequencies corresponding to frequencies N_i, $i = 0, 1, \ldots, k$, from the one-sample model, where j denotes the sample number.

The log-likelihood function of the two joint censored samples is

$$\begin{aligned}
l_{\mathrm{R}}(\alpha, \beta, \lambda, \tau) = {} & \log\left(\frac{n_1!}{N_{1,1}! \cdots N_{1,k}!}\right) + \sum_{i=1}^{k} N_{1,i} \log\left\{\exp\left[-\left(\frac{d_{i-1}}{\lambda}\right)^\tau\right] - \exp\left[-\left(\frac{d_i}{\lambda}\right)^\tau\right]\right\} \\
& + N_{1,0} \log\left(\frac{\tau}{\lambda^\tau}\right) + (\tau - 1) \sum_{i=n_1-N_{1,0}+1}^{n_1} \log\left(X_{1,(i)}\right) - \frac{1}{\lambda^\tau} \sum_{i=n_1-N_{1,0}+1}^{n_1} X_{1,(i)}^\tau \\
& + \log\left(\frac{n_2!}{N_{2,1}! \cdots N_{2,k}!}\right) + \sum_{i=1}^{k} N_{2,i} \log\left\{\exp\left[-\left(\frac{d_{i-1}}{\lambda + \alpha}\right)^{\tau+\beta}\right] - \exp\left[-\left(\frac{d_i}{\lambda + \alpha}\right)^{\tau+\beta}\right]\right\} \\
& + N_{2,0} \log\left[\frac{\tau + \beta}{(\lambda + \alpha)^{\tau+\beta}}\right] + (\tau + \beta - 1) \sum_{i=n_2-N_{2,0}+1}^{n_2} \log\left(X_{2,(i)}\right) - \frac{1}{(\lambda + \alpha)^{\tau+\beta}} \sum_{i=n_2-N_{2,0}+1}^{n_2} X_{2,(i)}^{\tau+\beta}.
\end{aligned} \tag{8}$$

As in the one-sample model, the ML estimates $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\lambda}$, $\widehat{\tau}$ of parameters α, β, λ, τ can be obtained by maximization of the log-likelihood function (8). The variability of ML estimates $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\lambda}$, $\widehat{\tau}$ can again be calculated from the expected FIM (see the habilitation thesis for more details).

Comparison of Distributions

For comparison of two independent censored samples from Weibull distribution, asymptotic tests with nuisance parameters can be used (see e.g. Lehmann and Casella, 1998), specifically the Lagrange multiplier (LM) test, the Wald (W) test and the likelihood ratio (LR) test. The null hypothesis H_0 is that distributions of both samples are equal. As it was stated at the beginning of this section, parameters α and β describe the difference between distributions of the first and the second sample. In case $\alpha = 0$ and $\beta = 0$, distributions of the two samples are identical. The null hypothesis $\mathrm{H}_0 : (\alpha, \beta)^{\mathrm{T}} = (0, 0)^{\mathrm{T}}$ is set against the alternative $\mathrm{H}_1 : (\alpha, \beta)^{\mathrm{T}} \neq (0, 0)^{\mathrm{T}}$, where λ and τ are nuisance parameters.

The test statistics are

$$\begin{aligned} LM &= \boldsymbol{U}_1(0, 0, \widetilde{\lambda}, \widetilde{\tau}) \left[\boldsymbol{J}^{\mathrm{R}}_{n,11.2}(0, 0, \widetilde{\lambda}, \widetilde{\tau}) \right]^{-1} \boldsymbol{U}_1^{\mathrm{T}}(0, 0, \widetilde{\lambda}, \widetilde{\tau}), \\ W &= (\widehat{\alpha}, \widehat{\beta}) \left[\boldsymbol{J}^{\mathrm{R}}_{n,11.2}(\widehat{\alpha}, \widehat{\beta}, \widehat{\lambda}, \widehat{\tau}) \right] (\widehat{\alpha}, \widehat{\beta})^{\mathrm{T}}, \\ LR &= 2 \left[l_{\mathrm{R}}(\widehat{\alpha}, \widehat{\beta}, \widehat{\lambda}, \widehat{\tau}) - l_{\mathrm{R}}(0, 0, \widetilde{\lambda}, \widetilde{\tau}) \right], \end{aligned} \tag{9}$$

where

$$\boldsymbol{U}_1(\alpha, \beta, \lambda, \tau) = \left(\frac{\partial l_{\mathrm{R}}}{\partial \alpha}, \frac{\partial l_{\mathrm{R}}}{\partial \beta} \right) = (u_1, u_2),$$

$$\begin{aligned} u_1 = &\sum_{i=1}^{k} N_i \frac{(\tau+\beta) \left\{ d_{i-1}^{\tau+\beta} \exp\left[-\left(\frac{d_{i-1}}{\lambda+\alpha} \right)^{\tau+\beta} \right] - d_i^{\tau+\beta} \exp\left[-\left(\frac{d_i}{\lambda+\alpha} \right)^{\tau+\beta} \right] \right\}}{(\lambda+\alpha)^{\tau+\beta+1} \left\{ \exp\left[-\left(\frac{d_{i-1}}{\lambda+\alpha} \right)^{\tau+\beta} \right] - \exp\left[-\left(\frac{d_i}{\lambda+\alpha} \right)^{\tau+\beta} \right] \right\}} - N_0 \frac{\tau+\beta}{\lambda+\alpha} \\ &+ \frac{\tau+\beta}{(\lambda+\alpha)^{\tau+\beta+1}} \sum_{i=n-N_0+1}^{n} X_{(i)}^{\tau+\beta}, \end{aligned}$$

$$\begin{aligned} u_2 = &\sum_{i=1}^{k} N_i \frac{d_i^{\tau+\beta} \log\left(\frac{d_i}{\lambda+\alpha} \right) \exp\left[-\left(\frac{d_i}{\lambda+\alpha} \right)^{\tau+\beta} \right] - d_{i-1}^{\tau+\beta} \log\left(\frac{d_{i-1}}{\lambda+\alpha} \right) \exp\left[-\left(\frac{d_{i-1}}{\lambda+\alpha} \right)^{\tau+\beta} \right]}{(\lambda+\alpha)^{\tau+\beta} \left\{ \exp\left[-\left(\frac{d_{i-1}}{\lambda+\alpha} \right)^{\tau+\beta} \right] - \exp\left[-\left(\frac{d_i}{\lambda+\alpha} \right)^{\tau+\beta} \right] \right\}} \\ &+ N_0 \frac{[1 - (\tau+\beta)\log(\lambda+\alpha)]}{\tau+\beta} + \sum_{i=n-N_0+1}^{n} \log X_{(i)} + \frac{\log(\lambda+\alpha)}{(\lambda+\alpha)^{\tau+\beta}} \sum_{i=n-N_0+1}^{n} X_{(i)}^{\tau+\beta} \\ &- \frac{1}{(\lambda+\alpha)^{\tau+\beta}} \sum_{i=n-N_0+1}^{n} X_{(i)}^{\tau+\beta} \log X_{(i)}, \end{aligned}$$

is the score function and $\boldsymbol{J}^{\mathrm{R}}_{n,11.2}(\alpha, \beta, \lambda, \tau)$ is a transformation of the expected FIM from the two-sample model (see the habilitation thesis for more details). The parameters estimated under the null hypothesis are denoted by tilde and those estimated under the alternative are denoted by hat. Under the null hypothesis, the test statistics (9) have asymptotically χ^2 distribution with two degrees of freedom (see e.g. Lehmann and Romano, 2005). The null hypothesis is rejected at a prescribed significance level when the test statistics exceed

the critical value of the χ^2 distribution. Performance of the tests using statistics (9) was studied using simulated power functions in Fusek and Michálek (2014) and can be found in the habilitation thesis.

Comparison of Expected Values

For comparison of means μ_1, μ_2 of two independent censored samples from Weibull distribution, the test based on Wald's test statistic (see e.g. Lehmann and Romano, 2005) can be used. The null hypothesis $\mathrm{H}_0 : \mu_1 - \mu_2 = 0$ is set against the alternative $\mathrm{H}_1 : \mu_1 - \mu_2 \neq 0$.

The test statistic is

$$W = \frac{\widehat{\mu}_1 - \widehat{\mu}_2}{\sqrt{\widehat{\mathrm{Var}}\,(\widehat{\mu}_1) + \widehat{\mathrm{Var}}\,(\widehat{\mu}_2)}}, \tag{10}$$

where $\widehat{\mu}_i = \mu(\widehat{\lambda}_i, \widehat{\tau}_i)$ can be calculated from (3) and $\widehat{\mathrm{Var}}\,(\widehat{\mu}_i) = \mathrm{Var}\left(\mu(\widehat{\lambda}_i, \widehat{\tau}_i)\right)$ can be calculated from (6) for $i = 1, 2$. Under the null hypothesis, the test statistic (10) is considered to be asymptotically normal $\mathrm{N}(0, 1)$.

In order to compare means of two independent censored samples, there is another option. We can use the asymptotic t-test. Nevertheless, since we deal with censored observations, certain adjustments have to be done. The usual approach in such a situation is based on replacing values between detection limits d_{i-1} and d_i, $i = 1, \ldots, k$, by constants lying between the individual detection limits, often by the midpoint of such an interval (see e.g. El-Shaarawi and Esterby, 1992). The null and the alternative hypotheses remain the same as above and the test statistic is

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{n}}}, \tag{11}$$

where $\overline{X}_1$ ($\overline{X}_2$ respectively) is the sample mean and S_1^2 (S_2^2 respectively) is the sample variance of the first (second respectively) sample. Under the null hypothesis of equal means, the statistic (11) is considered to be asymptotically normal $\mathrm{N}(0, 1)$. Performance of tests based on statistics (10), (11) was studied using simulated power functions in Fusek and Michálek (2015a) and can be found in the habilitation thesis.

2.4 Goodness-of-Fit Tests

When analyzing real data using the parametric approach, it is assumed that data have a specific distribution with cdf $F(x, \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k) \in \boldsymbol{\Theta} \subset \mathbb{R}^k$ is a vector of parameters. In environmental studies, data are typically skewed and various distributions such as the log-normal (Baccarelli et al., 2005; El-Shaarawi, 1989), the gamma (Hrdličková et al., 2008; Singh et al., 2002) and the Weibull (Fusek et al., 2015, 2020; Mbengue et al., 2018) 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 a 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 et al., 2011; Pakyari

and Balakrishnan, 2013; Pakyari and Nia, 2017). In this section, three tests (Kolmogorov-Smirnov, Cramér-von Mises, Anderson-Darling) based on the empirical distribution function (EDF) are considered, and their power is investigated by varying the null and the alternative distributions, the sample size and the degree of censoring.

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 is equivalent to testing that the log-transformed data follow the Gumbel distribution. 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 is equivalent to testing that the log-transformed data follow the normal distribution. Critical values of the EDF statistics can be obtained by means of Monte Carlo simulations. Three test statistics based on the EDF $F_n(x)$ are applied (see d'Agostino and Stephens, 1986, for more details).

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

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

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 $Z_{(N_0+1)} = 1 - F_0(d_2, \widehat{\boldsymbol{\theta}})$.

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 $Z_{(N_0+1)} = 1 - F_0(d_2, \widehat{\boldsymbol{\theta}})$. The empirical significance level as well as the power of the above mentioned tests was studied by means of Monte Carlo simulations in Fusek (2023) for Weibull, lognormal and gamma distributions and the results can be found in the habilitation thesis.

3 APPLICATIONS

When dealing with environmental or microbiological data, measured values are often found below the detection limits of a measurement method, and only the number of values below the detection limits can be determined. There are usually two detection limits. One of them is called the limit of detection (LOD), which is the lowest quantity of a substance that can be distinguished from the absence of that substance (a blank value) within a stated confidence limit. The other one is called the limit of quantification (LOQ), which is the lowest analyte concentration that can be quantitatively detected with a stated accuracy and precision. In such case, we are dealing with Type I doubly left-censored data (Busschaert et al., 2010; Fusek et al., 2015; Pouillot et al., 2013; Shorten et al., 2006; Valero et al., 2017). In this section, statistical methods described in previous sections are used for analyses of environmental data. Two approaches are used. The first one is based on the censored exponential and Weibull distributions. The second one is based on the so-called "replacement method," where all values under the detection limits are replaced by midpoints of intervals $(0, \mathrm{LOD}]$ and $(\mathrm{LOD}, \mathrm{LOQ}]$.

3.1 Musk Compounds

Synthetic aromatic substances or musk compounds are lipophilic contaminants able to accumulate in various components of the environment. They represent a group of persistent

pollutants, and may occur in environmental matrices and human tissues. Synthetic aromatic substances were launched on the market in the early 20th century and the volume of their production has significantly increased in recent years (Luckenbach and Epel, 2005). Since they have potential carcinogenic properties, efforts are currently being made to limit or prohibit their use in many regions worldwide.

In general, musk compounds can be divided into four groups: linear, macrocyclic, polycyclic and nitro musk compounds. The last two groups are used most frequently as substitutes for natural musks in fragrances and personal hygiene products (OSPAR, 2004). Galaxolide (HHCB) and tonalide (AHTN) are examples of the most important polycyclic musk compounds. Musk xylene, musk ketone and musk ambrette are well-known nitro musk compounds (i.e. compounds containing one or more nitro groups in a molecule). The production of nitro musk compounds, that are generally included in a group of substances posing a risk to the environment, has decreased over the last years (Bester, 2009; Rimkus, 1999). By contrast, production of polycyclic synthetic aromatic substances, which are less toxic, has increased because of their frequent use as additives in many personal care products, e.g. soaps, shampoos, deodorants, body lotions, perfumes, cleaning and disinfecting agents, air fresheners and industrial cleaning agents (see e.g. Sumner et al., 2010). Synthetic aromatic substances were also detected in samples of air and dust collected in indoor environments (Regueiro et al., 2009). They often penetrate into the environment through wastewater because of their ineffective removal in the wastewater treatment plant (WWTP), see e.g. Gómez et al. (2006) and references inside. Accumulation of these substances in the environment (surface water, sediment) results in their occurrence in food chain, especially in aquatic ecosystems. A number of studies revealed the presence of musk compounds in tissues of aquatic animals. These compounds can also be found in human body, for example in fat tissue, human milk and blood plasma (see e.g. Lignell et al., 2008; Zlámalová Gargošová et al., 2013), as a consequence of fish consumption.

The goal is:

a) to model musk compound concentrations using methods for censored samples;

b) to evaluate the amount of musk compounds in fish caught upstream and downstream the WWTP;

c) to compare mean concentrations and distributions of concentrations of musk compounds in fish caught upstream and downstream the WWTP.

Data

The sample of aquatic biota consists of 60 fish from the carp family, specifically of the European chub (Leuciscus cephalus), which were caught in the Svratka River, Czech Republic, near the WWTP Brno-Modřice by Morava River Basin Administration employees. Fish were caught on 10th November 2009; half of them came from a watercourse upstream (Group 1), and half of them from a watercourse downstream (Group 2) from the WWTP. The fish were transported to the laboratory of the Institute of Veterinary Hygiene and Ecology of Veterinary and Pharmaceutical University in Brno, and examined by a veterinarian. Relevant characteristics were noted and then muscle, skin and guts were separated. Muscle tissue was selected for the musk compound analysis because it is considered to be representative of all of the

Tab. 1: Musk compounds distribution in fish samples upstream and downstream the WWTP.

	Upstream			Downstream			LOD [μg/kg]	LOQ [μg/kg]
	N_1^a	N_2^b	N_0^c	N_1	N_2	N_0		
PH	18	12	0	23	6	1	0.55	1.82
AMB	28	2	0	28	2	0	1.46	4.88
TR	24	6	0	22	8	0	1.11	3.68
HHCB	3	23	4	0	22	8	8.95	29.83
AHTN	6	17	7	8	4	18	1.98	6.62
MX	22	4	4	23	6	1	0.75	2.50
TIB	27	1	2	28	2	0	0.15	0.51
MK	0	17	13	0	19	11	0.57	1.90

[a] N_1 - the number of values below the LOD

[b] N_2 - number of values in the interval (LOD,LOQ]

[c] N_0 - number of uncensored values

body. The muscle tissue was homogenized (using a blender), subsequently frozen at -20 °C and kept frozen until the analysis. Fish of approximately the same age were chosen for the analysis. As a result, four nitro (musk ambrette - AMB, musk xylene - MX, musk tibetene - TIB, musk ketone - MK), and four polycyclic musk compounds (phantolide - PH, traseolide - TR, galaxolide - HHCB, tonalide - AHTN) were detected. Details of the chemical analyses can be found in Fusek et al. (2015). The LOD and the LOQ were calculated using calibration curves of particular analytes, see Kellner (1998) for more information. Frequencies of censored and uncensored musk compound concentrations are presented in Table 1 together with the specified detection limits.

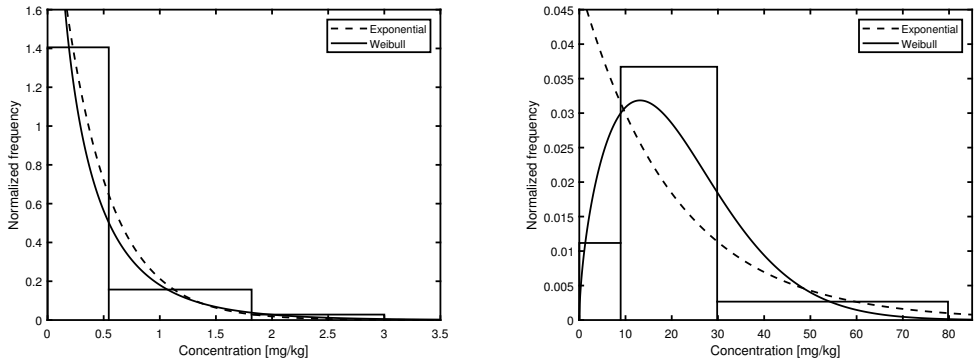

Fig. 1: Histogram (normalized to the pdf) of PH concentrations downstream the WWTP (left) and HHCB concentrations upstream the WWTP (right) with exponential and Weibull densities.

Results

The suitability of the exponential and the Weibull distributions for modelling of BA concentrations was tested using Pearson's χ^2 goodness-of-fit test and Cramér-von Mises test for censored data that was implemented in Matlab (version R2022b). Moreover, graphical analysis of the data was used in order to choose the model distribution. Specifically, Q-Q plots and correspondence between the histogram (normalized to the pdf) and the exponential (Weibull respectively) density was assessed. It was found out that it is possible to use the exponential distribution for modelling of PH, AMB, TR, TIB, MK and the Weibull distribution for modelling of HHCB, MX (see Fig. 1 for an example). In case of AHTN in fish caught downstream the WWTP, neither the exponential nor the Weibull distribution was suitable and we applied the more flexible Weibull distribution just for illustration purposes.

Mean musk compound concentrations were estimated together with their 95% confidence limits using methods for censored exponential and Weibull distributions (see Table 2), and using the replacement method, where all censored values were replaced by midpoints of intervals $(0, \mathrm{LOD}]$ and $(\mathrm{LOD}, \mathrm{LOQ}]$ (see Table 3). It can be seen that both estimates are quite similar. Mean musk compound concentration estimates based on censored distributions are of lower values than those estimated using the replacement method in most cases. In order to assess the estimation quality, a simulation study was carried out in Fusek et al. (2015). The authors showed that behavior of mean and variance estimates based on the censored distribution and the replacement method (using the sample mean and the sample variance) are rather similar in case of low censoring. On the other hand, when the number of censored values is high, performance of the estimates based on the replacement method is not particularly good.

Results of the comparison of mean musk compound concentrations in fish caught upstream and downstream the WWTP are presented in Table 4. There is no significant difference between the two groups in most cases. In case of methods for censored distributions, there is a significant difference in mean concentrations of TIB between fish caught upstream and downstream the WWTP at the significance level of 0.05. On the other hand, the replacement method was not able to reveal the difference between mean concentrations of TIB in fish caught upstream and downstream the WWTP. Another part of the analysis was to compare distributions of musk compound concentrations in fish caught upstream and downstream the WWTP. It was found out that all tests (the likelihood ratio test, the Lagrange multiplier test, and the Wald test) give similar results. There is no significant difference between the two groups at the significance level of 0.05 with the exception of TIB.

To summarize the results, it was found out that the WWTP has no significant influence on concentrations of musk compounds in fish tissue at the significance level of 0.05 with the exception of musk tibetene.

3.2 Biogenic Amines

Biogenic amines (BAs; e.g. histamine, tyramine, fenyletylamine, tryptamine, putrescine, kadaverine, spermine and spermidine) are biologically active organic bases with a low molecular weight which are synthesized by living organisms for their own needs. They pass into food and beverages through ingredients (usually a small amount) and are also generated by microbial decarboxylation of amino acids. The intake of BAs into the body is regulated by a detoxification system composed of monoamine oxidases, diamino oxidases and histidine

Tab. 2: Mean musk compound concentrations (in μg/kg) with their standard deviations (SD) and 95% lower (LCL) and upper (UCL) confidence limits estimated using the censored exponential (PH, AMB, TR, TIB) and Weibull (HHCB, AHTN, MX, MK) distributions.

	Upstream				Downstream			
	Mean	SD	LCL	UCL	Mean	SD	LCL	UCL
PH	0.512	0.107	0.303	0.721	0.411	0.088	0.239	0.583
AMB	0.540	0.136	0.274	0.806	0.540	0.136	0.274	0.806
TR	0.669	0.147	0.381	0.957	0.792	0.170	0.459	1.126
HHCB	21.482	2.676	16.238	26.726	26.008	2.593	20.926	31.091
AHTN	4.664	0.610	3.468	5.860	6.140	0.802	4.568	7.712
MX	0.934	0.405	0.141	1.727	0.544	0.151	0.249	0.839
TIB	0.201	0.040	0.122	0.280	0.057	0.014	0.029	0.085
MK	2.194	0.411	1.389	3.000	1.918	0.361	1.210	2.625

Tab. 3: Mean musk compound concentrations (in μg/kg) with their standard deviations (SD) and 95% lower (LCL) and upper (UCL) confidence limits estimated using the replacement method.

	Upstream				Downstream			
	Mean	SD	LCL	UCL	Mean	SD	LCL	UCL
PH	0.636	0.083	0.474	0.798	0.513	0.085	0.346	0.680
AMB	0.895	0.113	0.673	1.116	0.895	0.113	0.673	1.116
TR	0.921	0.137	0.653	1.189	1.044	0.151	0.747	1.340
HHCB	21.923	2.516	16.991	26.855	25.606	2.347	21.006	30.205
AHTN	4.757	0.559	3.662	5.852	6.170	0.717	4.765	7.575
MX	1.064	0.253	0.568	1.559	0.716	0.124	0.472	0.959
TIB	0.211	0.094	0.027	0.395	0.094	0.012	0.071	0.118
MK	2.232	0.381	1.485	2.980	1.966	0.241	1.493	2.439

Tab. 4: Comparison of mean musk compound concentrations between fish caught upstream and downstream the WWTP using methods for censored distributions with statistic (10) (p-value p_{cen}), and the replacement method with statistic (11) (p-value p_{rep}). Comparison of distributions of musk compound concentrations using the likelihood ratio test (p-value p_{LR}), the Lagrange multiplier test (p-value p_{LM}), the Wald test (p-value p_{W}) and statistics (9).

	Comparison of means		Comparison of distributions		
	p_{cen}	p_{rep}	p_{LR}	p_{LM}	p_{W}
PH	0.47	0.30	0.45	0.44	0.47
AMB	1.00	1.00	1.00	1.00	1.00
TR	0.58	0.55	0.58	0.57	0.58
HHCB	0.22	0.28	0.36	0.33	0.38
AHTN	0.14	0.12	0.28	0.28	0.31
MX	0.37	0.22	0.37	0.42	0.47
TIB	$< 0.01^*$	0.22	$< 0.01^*$	$< 0.01^*$	$< 0.01^*$
MK	0.61	0.56	0.61	0.60	0.61

* - rejection of the hypothesis at the significance level of 0.05

methyl-transferases. High intake of BAs (in general over 100 mg/kg of food) or the presence of factors that reduce the effectiveness of the detoxification system can lead to intoxication which may endanger health and, in some cases, life (Halász et al., 1994; Shalaby, 1996; Silla Santos, 1996; Ten Brink et al., 1990). For example, in case of histamine, there is a legislatively based concentration limit of 100 mg/kg (200 mg/kg, respectively) in fish and fishery products (Commission Regulation EC 2073/2005). Moreover, lower limits for specific BAs have been proposed in literature, e.g. 10 mg/kg of histamine, 80 mg/kg of tyramine, and 3 mg/kg of phenylethylamine (Halász et al., 1994; Ten Brink et al., 1990).

Freshwater and saltwater fish are an important part of our diet owing to a high content of polyunsaturated fatty acids, minerals and other biologically active substances. Nevertheless, fish meat represents a system with very short shelf-life due to very fast post-mortem changes which are related to frequent occurrence of BAs. Frequent contaminants of fish include bacteria from the *Enterobacteriaceae* family and the genera Pseudomonas. In addition, the lactic acid bacteria, e.g. representatives of the genera Lactobacillus and Enterococcus, can also contribute to production of BAs (Apetrei and Apetrei, 2015; Arnold and Brown, 1978; Jaw et al., 2012; Kaale et al., 2011; Prester, 2011; Rawles et al., 1996; Zhang et al., 2010). The content of BAs in fish meat was previously studied in Buňka et al. (2013). They found out that the BA content was higher than 100 mg/kg in approximately 15% of samples. Moreover, in 6 samples the concentrations were so high that they failed to comply with legislative requirements established in Commission Regulation EC 2073/2005. Such high concentrations have a significant impact on food safety, and may endanger health or even life of sensitive individuals, which emphasizes the importance of monitoring BAs in these commodities.

Since in the previous study by Buňka et al. (2013) the concentrations below detection limits were not taken into account, the goal is:

a) to model concentration of BAs using methods for censored samples;

b) to evaluate the amount of BAs in various fish species (Atlantic salmon, Atlantic cod, striped catfish);

c) to compare mean BA concentrations and distributions of concentrations among the species;

d) to determine the risk of exceeding certain BA limits for various fish species.

Data

In total 54 samples of fish commonly consumed in Central Europe were analyzed. There were 18 samples of Atlantic salmon (Salmo salar), 17 samples of Atlantic cod (Gadus morhua), and 19 samples of striped catfish (Pangasius hypophthalmus). The fish were bought in retail stores, stored on ice, and the samples were extracted from commonly consumed parts of the fish. The same parts of the fish muscle tissue were used. The period between buying the fish and start of lyophilization of the samples in the laboratory did not exceed 6 hours, and the samples were stored in a fridge at 2 ± 1 °C. The samples were extracted immediately after the lyophilization.

The extraction and determination of BAs (histamine - HIM, tyramine - TYM, phenylethylamine - PHE, tryptamine - TRM, putrescine - PUT, cadaverine - CAD, spermidine - SPD, spermine - SPN) were carried out according to Buňka et al. (2013). The BA content in samples was determined using high performance liquid chromatography (LabAlliance, State College, USA; Agilent Technologies, Agilent, Palo Alto, USA) after preceding derivatization with dansyl chloride. Every sample was analyzed eight times (2 extracts of each sample, times 2 derivatizations of each extract, times 2 analyses of each derivatized extract). Results (in mg/kg) are expressed for the fresh matter before lyophilization. The LOD and the LOQ were determined according to standard chromatography procedures (Lister, 2005; Wenzl et al., 2016) and in accordance with ISO 17025 (ISO, 2017). Frequencies of censored and uncensored BA concentrations for various fish species are presented in Table 5 together with the specified detection limits.

Results

The suitability of the exponential and the Weibull distributions for modelling of BA concentrations was tested using Pearson's χ^2 goodness-of-fit test and Cramér-von Mises test for censored data that was implemented in Matlab (version R2022b). Moreover, graphical analysis of the data was used in order to choose the model distribution. Specifically, Q-Q plots and correspondence between the histogram (normalized to the pdf) and the exponential (Weibull respectively) density was assessed. It was found out that the Weibull distribution is suitable (despite some anomalies caused by extreme values or missing values in the interval (LOD,LOQ]) for modelling of all the BAs. In addition, it was possible to use the exponential distribution for modelling of TRM, CAD, HIM and TYM. In case of PHE, PUT, SPD and SPN, it was necessary to use the Weibull distribution (see Fig. 2 for an example).

Tab. 5: Biogenic amines distribution in fish samples.

	Atlantic salmon			Atlantic cod			Striped catfish			LOD [mg/kg]	LOQ [mg/kg]
	N_1^a	N_2^b	N_0^c	N_1	N_2	N_0	N_1	N_2	N_0		
TRM	17	1	0	15	2	0	15	4	0	0.13	0.35
PHE	12	2	4	9	1	7	16	1	2	0.06	0.21
PUT	0	0	18	0	0	17	2	1	16	0.16	0.82
CAD	7	1	10	4	1	12	10	1	8	0.09	0.26
HIM	10	2	6	4	1	12	16	3	0	0.11	0.38
TYM	10	0	8	4	0	13	13	2	4	0.01	0.08
SPD	0	0	18	1	1	15	0	0	19	0.13	0.29
SPN	0	0	18	0	0	17	1	0	18	0.02	0.13

[a] N_1 - the number of values below the LOD

[b] N_2 - number of values in the interval (LOD,LOQ]

[c] N_0 - number of uncensored values

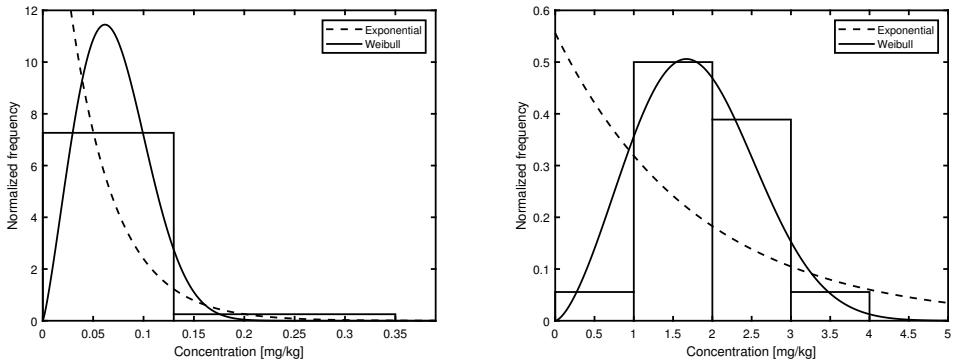

Fig. 2: Histogram (normalized to the pdf) of TRM (left) and SPN (right) concentrations in Atlantic salmon with exponential and Weibull densities.

Estimation of BA Concentrations

Mean BA concentrations were estimated together with their 95% confidence limits using methods for censored exponential and Weibull distributions and the results are in Table 6. It can be seen that concentrations of TRM, PHE, TYM, SPD and SPN are relatively low and do not pose a high risk to consumers' health. Nevertheless, the amount of PUT was relatively high and can have a negative impact on food safety, and ultimately on consumers' health, especially in combination with alcohol consumption, which inhibits the detoxification system in the human body (Shalaby, 1996; Silla Santos, 1996; Ten Brink et al., 1990). Similar conclusions can be made in case of CAD in Atlantic cod and Atlantic salmon and HIM in Atlantic salmon, where the concentrations of BAs are also high.

In Table 7, there are mean BA concentrations together with their 95% confidence limits that were estimated using the replacement method, where all censored values were replaced by

Tab. 6: Mean BA concentrations (in mg/kg) with their standard deviations (SD) and 95% lower (LCL) and upper (UCL) confidence limits estimated using the censored exponential (TRM, CAD, HIM, TYM) and Weibull (PHE, PUT, SPD, SPN) distributions.

	Atlantic salmon				Atlantic cod				Striped catfish			
	Mean	SD	LCL	UCL	Mean	SD	LCL	UCL	Mean	SD	LCL	UCL
TRM	0.045	0.015	0.015	0.074	0.060	0.018	0.024	0.095	0.079	0.021	0.037	0.121
PHE	0.492	0.565	0.000	1.601	0.670	0.512	0.000	1.674	0.860	2.886	0.000	6.517
PUT	38.867	13.787	11.845	65.890	36.367	2.738	31.001	41.732	25.210	6.844	11.795	38.624
CAD	8.335	1.968	4.478	12.192	10.447	2.538	5.474	15.421	0.396	0.094	0.213	0.580
HIM	12.470	2.944	6.700	18.240	3.807	0.928	1.987	5.627	0.059	0.017	0.026	0.092
TYM	5.165	1.218	2.777	7.552	0.504	0.123	0.262	0.745	0.093	0.022	0.049	0.137
SPD	5.136	1.556	2.087	8.186	0.747	0.115	0.522	0.971	1.817	0.307	1.215	2.420
SPN	1.801	0.178	1.452	2.150	0.741	0.088	0.568	0.913	0.690	0.123	0.449	0.930

Tab. 7: Mean BA concentrations (in mg/kg) with their standard deviations (SD) and 95% lower (LCL) and upper (UCL) confidence limits estimated using the replacement method.

	Atlantic salmon				Atlantic cod				Striped catfish			
	Mean	SD	LCL	UCL	Mean	SD	LCL	UCL	Mean	SD	LCL	UCL
TRM	0.075	0.010	0.056	0.094	0.086	0.014	0.058	0.113	0.102	0.017	0.069	0.135
PHE	0.334	0.147	0.046	0.621	0.491	0.152	0.193	0.790	0.298	0.227	0.000	0.743
PUT	38.988	13.065	13.380	64.596	36.479	2.676	31.234	41.724	24.481	4.607	15.451	33.511
CAD	8.335	3.223	2.018	14.652	10.447	3.366	3.850	17.044	0.397	0.142	0.118	0.677
HIM	12.470	5.625	1.446	23.494	3.807	0.979	1.889	5.725	0.085	0.016	0.053	0.117
TYM	5.165	1.860	1.518	8.811	0.504	0.105	0.299	0.709	0.094	0.044	0.008	0.180
SPD	5.515	2.929	0.000	11.257	0.745	0.119	0.511	0.979	1.793	0.364	1.081	2.506
SPN	1.798	0.179	1.447	2.149	0.737	0.092	0.557	0.917	0.693	0.121	0.457	0.930

Tab. 8: Comparison of mean BA concentrations among various fish species using methods for censored distributions with statistic (10) (p-value p_{W}), and the replacement method with statistic (11) (p-value p_{rep}). Comparison of distributions of BA concentrations using the likelihood ratio test (9) (p-value p_{LR}).

	Atlantic salmon vs. Atlantic cod			Atlantic salmon vs. Striped catfish			Atlantic cod vs. Striped catfish		
	p_{W}	p_{rep}	p_{LR}	p_{W}	p_{rep}	p_{LR}	p_{W}	p_{rep}	p_{LR}
TRM	0.53	0.53	0.52	0.20	0.16	0.18	0.50	0.46	0.49
PHE	0.82	0.46	0.57	0.90	0.89	0.40	0.95	0.48	0.06
PUT	0.86	0.85	$< 0.01^*$	0.38	0.30	0.54	0.13	0.02^*	$< 0.01^*$
CAD	0.51	0.65	0.50	$< 0.01^*$	0.01^*	$< 0.01^*$	$< 0.01^*$	$< 0.01^*$	$< 0.01^*$
HIM	$< 0.01^*$	0.13	$< 0.01^*$	$< 0.01^*$	0.03^*	$< 0.01^*$	$< 0.01^*$	$< 0.01^*$	$< 0.01^*$
TYM	$< 0.01^*$	0.01^*	$< 0.01^*$	$< 0.01^*$	$< 0.01^*$	$< 0.01^*$	$< 0.01^*$	$< 0.01^*$	$< 0.01^*$
SPD	$< 0.01^*$	0.10	$< 0.01^*$	0.04^*	0.21	$< 0.01^*$	$< 0.01^*$	$< 0.01^*$	0.01^*
SPN	$< 0.01^*$	$< 0.01^*$	$< 0.01^*$	$< 0.01^*$	$< 0.01^*$	$< 0.01^*$	0.74	0.77	0.17

* - rejection of the hypothesis at the significance level of 0.05

midpoints of intervals $(0, \mathrm{LOD}]$ and $(\mathrm{LOD}, \mathrm{LOQ}]$. In a situation where there are no uncensored values, the ability of estimating mean concentrations is very limited. In case of TRM (and HIM for striped catfish), the replacement method overestimates the mean concentration and underestimates its variability in comparison to the use of censored distribution. When both censored and uncensored values are present, the estimates of mean concentrations are quite similar with the exception of PHE, where the replacement method underestimates the mean concentration.

Comparison of BA Concentrations

Results of the comparison of mean BA concentrations are presented in Table 8. First of all, let us focus on the methods for censored distributions. There is a significant difference in mean concentrations of HIM, TYM and SPD among all three species at the significance level of 0.05. Moreover, in case of CAD, only Atlantic salmon and Atlantic cod have similar mean concentrations. In case of SPN, only Atlantic cod and striped catfish have similar mean concentrations. Furthermore, let us focus on the replacement method, where the equality of mean concentrations was tested using the asymptotic t-test. Contradictory results were obtained in case of comparison of a) Atlantic salmon and Atlantic cod for HIM and SPD, and b) Atlantic salmon and striped catfish for SPD, where the replacement method was not able to reveal the difference between mean BA concentrations. One of the reasons for that could be a low power of the asymptotic test caused by the small sample sizes or a high skewness of the data. Moreover, there is a high variability of the estimate of the mean HIM concentration in Atlantic salmon (see standard deviations in Table 7) which can affect the test results. In case of Atlantic cod and striped catfish for PUT, the high variability of mean concentration estimates (see standard deviations in Table 6) has a significant influence on values of the test statistics (10) which results into non-significant differences between the means.

Another part of the analysis was to compare distributions of BA concentrations among various species. It was found out that all tests (the likelihood ratio test, the Lagrange multiplier test, and the Wald test) give similar results. On that account, only results for the likelihood ratio test are presented in Table 8. It can be seen that the results are very similar to the comparison of mean BA concentrations with one exception. There is a significant difference in distributions of PUT concentrations between Atlantic cod and the other species, even though the mean concentrations among species are similar (see Fig. 3, left). In case of SPN, there is a clear difference in distributions between Atlantic salmon and the other species. Nevertheless, the difference in distributions between Atlantic cod and striped catfish is not significant enough to warrant rejection of the null hypothesis at the significance level of 0.05 (see Fig. 3, right).

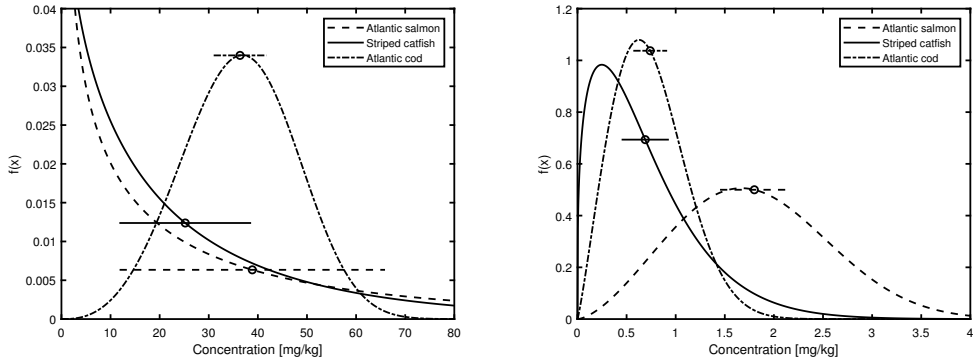

Fig. 3: Probability density functions of PUT (left) and SPN (right) concentrations with mean values (circle) and their confidence intervals (horizontal line).

Risk Probabilities

Once the model distributions of BA concentrations are stated and the unknown parameters in the model are estimated, probabilities of exceeding certain limit values of BA concentrations can be calculated. The risk probability R of exceeding the limit value LV can be approximated using formula

$$R = P(X > LV) \doteq 1 - F(x, \widehat{\lambda}, \widehat{\tau}),$$

where $F(x, \lambda, \tau)$ is cdf (1).

In general, it is difficult to select a specific limit value of BA concentrations that could seriously harm consumers' health. In fact, every BA has its own physiological effect on human body; additionally, each body reacts to exposure to BAs (and other biologically active substances) in a slightly different way. Based on our opinion and also recommendations regarding the food safety in other studies, four limit values of BA concentrations were selected, specifically 3, 10, 22 and 100 mg/kg. According to Halász et al. (1994), Shalaby (1996) and Ten Brink et al. (1990), the limits of 3 and 10 mg/kg are very important especially for PHE and HIM. Higher concentrations can cause vasodilation effects (affect blood pressure and heart activity), headache and/or breathing problems. More serious problems can be expected with increased alcohol consumption and/or when antihistamins are used. Table 9 shows that HIM concentration in Atlantic salmon exceeds the limit value of 3, 10 and 22 mg/kg with probabilities 0.79, 0.45 and 0.17. Additionally, PUT and CAD concentrations over 20 mg/kg can

Tab. 9: Probabilities of exceeding limit values (LV) of BA concentrations for Atlantic salmon (AS), Atlantic cod (AC), and striped catfish (SC).

LV [mg/kg]	3			10			22			100		
	AS	AC	SC	AS	AC	SC	AS	AC	SC	AS	AC	SC
TRM	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
PHE	0.03	0.05	0.03	< 0.01	< 0.01	0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
PUT	0.81	1.00	0.84	0.62	0.99	0.61	0.44	0.89	0.38	0.10	< 0.01	0.03
CAD	0.70	0.75	< 0.01	0.30	0.38	< 0.01	0.07	0.12	< 0.01	< 0.01	< 0.01	< 0.01
HIM	0.79	0.45	< 0.01	0.45	0.07	< 0.01	0.17	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
TYM	0.56	< 0.01	< 0.01	0.14	< 0.01	< 0.01	0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
SPD	0.48	< 0.01	0.17	0.15	< 0.01	< 0.01	0.03	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
SPN	0.07	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

increase the effects of HIM, TYM and PHE on the human body (Halász et al., 1994; Shalaby, 1996; Ten Brink et al., 1990). Table 9 shows that PUT concentration exceeds the limit value of 22 mg/kg with probabilities 0.38 for striped catfish, 0.44 for Atlantic salmon, and 0.89 in case of Atlantic cod. CAD concentration exceeds 22 mg/kg with probabilities 0.07 for Atlantic salmon, and 0.12 for Atlantic cod. The limit of 100 mg/kg is the generally accepted limit for evaluation of food safety not only for individual BAs, but also for the total amount of BA concentrations (Benkerroum, 2016; EFSA, 2011; Halász et al., 1994; Kalač, 2014; Ten Brink et al., 1990). The limit value of 100 mg/kg is exceeded only in case of PUT concentration in Atlantic salmon with probability 0.1, and in striped catfish with probability 0.03.

4 CONCLUSIONS

Type I left-censored data occur frequently in many application areas. This thesis proposed new statistical methods for an analysis of censored data with the Weibull (exponential respectively) distribution, which is very flexible and can be used for modelling of various engineering problems.

Methods for statistical analysis of one data file and also for various comparisons of two independent data files were described and analyzed using simulations. Since it is very important to correctly determine the probability distribution of the analyzed data, goodness-of-fit tests for the Type I left-censored Weibull, lognormal and gamma distributions, which are among the most frequently used distributions for modelling of censored data, were described and analyzed using simulations. In addition, the suggested statistical methods were applied in the real data analysis.

More applications of the described methods can be expected in future and they don't have to be limited to the environmental data. One of the possible applications of the Type I left-censored distributions in the extreme value theory was already suggested and briefly described in the habilitation thesis.

REFERENCES

APETREI, I. M., APETREI, C., 2015. The biocomposite screen-printed biosensor based on immobilization of tyrosinase onto the carboxyl functionalised carbon nanotube for assaying tyramine in fish products. *Journal of Food Engineering*, **149**, 1–8. ISSN: 0260-8774.

ARNOLD, S. H., BROWN, W. D., 1978. Histamine toxicity from fish products. *Advances in Food Research*, **24**, 113–154. ISSN: 0065-2628.

BACCARELLI, A., PFEIFFER, R., CONSONNI, D., PESATORI, A., BONZINI, M., PATTERSON, D. G., 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. ISSN: 0045-6535.

BARNDORFF-NIELSEN, O. E., COX, D. R., 1994. *Inference and Asymptotics*. London: Chapman and Hall/CRC. ISBN: 978-0412494406.

BENKERROUM, N., 2016. Biogenic amines in dairy products: Origin, incidence and control means. *Comprehensive Reviews in Food Science and Food Safety*, **15**(4), 801–826. ISSN: 1541-4337.

BESTER, K., 2009. Analysis of musk fragrances in environmental samples. *Journal of Chromatography A*, **1216**(3), 470–480. ISSN: 0021-9673.

BISPO, R., MARQUES, T. A., 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. ISSN: 0094-9655.

BUŇKA, F., BUDINSKÝ, P., ZIMÁKOVÁ, B., MERHAUT, M., FLASAROVÁ, R., PACHLOVÁ, V., KUBÁŇ, V., BUŇKOVÁ, L., 2013. Biogenic amines occurrence in fish meat sampled from restaurants in region of Czech Republic. *Food Control*, **31**(1), 49–52. ISSN: 0956-7135.

BUSSCHAERT, P., GEERAERD, A. H., UYTTENDAELE, M., VAN IMPE, J. F., 2010. Estimating distributions out of qualitative and (semi)quantitative microbiological contamination data for use in risk assessment. *International Journal of Food Microbiology*, **138**(3), 260–269. ISSN: 0168-1605.

COHEN, A. C., 1991. *Truncated and Censored Samples*. New York: Marcel Dekker. ISBN: 978-0824784478.

COMMISSION REGULATION EC 2073/2005 of 15 November 2005 on microbiological criteria for foodstuffs. *Official Journal of the European Union*, **L338**, 1–26. ISSN: 1977-0677.

D'AGOSTINO, R. B.; STEPHENS, M. A., 1986. *Goodness-of-fit-techniques*. New York: M. Dekker. ISBN: 978-0824774875.

EUROPEAN FOOD SAFETY AUTHORITY - EFSA, 2011. Scientific opinion on risk based control of biogenic amine formation in fermented foods. *EFSA Journal*, **9**(10), 2393. ISSN: 1831-4732.

EL-SHAARAWI, A. H., 1989. Inferences about the mean from censored water quality data. *Water Resources Research*, **25**, 685–690. ISSN: 1944-7973.

EL-SHAARAWI, A. H., ESTERBY, S. R., 1992. Replacement of censored observations by a constant: An evaluation. *Water Research*, **26**(6), 835–844. ISSN: 0043-1354.

FUSEK, M., 2017. On testing reduction of a left-censored Weibull distribution to an exponential submodel. In: MATOUŠEK, R., ed. *Proceedings of 23rd International Conference on Soft Computing – MENDEL 2017*. Brno: VUT Press, 179–184. ISSN: 1803-3814.

FUSEK, M., 2023. Statistical power of goodness-of-fit tests for type I left-censored data. *Austrian Journal of Statistics*, **52**(1), 51–61. ISSN: 1026-597X.

FUSEK, M., MICHÁLEK, J., 2014. Asymptotic tests for multiply left-censored samples from Weibull distribution. In: MATOUŠEK, R., ed. *Proceedings of 20th International Conference on Soft Computing – MENDEL 2014*. Brno: VUT Press, 317–322. ISSN: 1803-3814.

FUSEK, M., MICHÁLEK, J., 2015a. T-test alternative for multiply left-censored samples from Weibull distribution. In: MATOUŠEK, R., ed. *Proceedings of 21st International Conference on Soft Computing – MENDEL 2015*. Brno: VUT Press, 169–174. ISSN: 1803-3814.

FUSEK, M., MICHÁLEK, J., 2016. On the confidence intervals for mean of left- censored Weibull distribution. In: MATOUŠEK, R., ed. *Proceedings of 22nd International Conference on Soft Computing – MENDEL 2016*. Brno: VUT Press, 249–254. ISSN: 1803-3814.

FUSEK, M., MICHÁLEK, J., 2019. Statistical inference for type I multiply left-censored samples from Weibull distribution. *Cybernetics and Systems Analysis*, **55**(4), 590–604.

FUSEK, M., MICHÁLEK, J., BUŇKOVÁ, L., BUŇKA, F., 2020. Modelling biogenic amines in fish meat in Central Europe using censored distributions. *Chemosphere*, **251**, 126390. ISSN: 0045-6535.

FUSEK, M., MICHÁLEK, J., VÁVROVÁ, M., 2015. Evaluation of contamination data with non-detects using censored distributions. *Fresenius Environmental Bulletin*, **24**, 4165–4172. ISSN: 1018-4619.

GÓMEZ, M. J., PETROVIĆ, M., FERNÁNDEZ-ALBA, A. R., BARCELÓ, D., 2006. Determination of pharmaceuticals of various therapeutic classes by solid-phase extraction and liquid chromatography–tandem mass spectrometry analysis in hospital effluent wastewaters. *Journal of Chromatography A*, **1114**(2), 224–233. ISSN: 0021-9673.

GUÉRIN, T., CHEKRI, R., VASTEL, C., SIROT, V., VOLATIER, J.-L., LEBLANC, J.-C., NOËL, L., 2011. Determination of 20 trace elements in fish and other seafood from the French market. *Food Chemistry*, **127**(3), 934–942. ISSN: 0308-8146.

HALÁSZ, A., BARÁTH, Á., SIMON-SARKADI, L., HOLZAPFEL, W., 1994. Biogenic amines and their production by microorganisms in food. *Trends in Food Science and Technology*, **51**, 42–49. ISSN: 0924-2244.

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. ISSN: 0045-6535.

HELSEL, D. R., 2012. *Statistics for censored environmental data using Minitab® and R*. New York: John Wiley and Sons. ISBN: 978-0470479889.

HOELZER, K., POUILLOT, R., VAN DOREN, J. M., DENNIS, S., 2014. Reduction of Listeria monocytogenes contamination on produce – A quantitative analysis of common liquid fresh produce wash compounds. *Food Control*, **46**, 430–440. ISSN: 0956-7135.

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. ISSN: 1352-2310.

ISO, 2017. *ISO Standard No. 17025: General requirements for the competence of testing and calibration laboratories*. Geneva, Switzerland: ISO. ISBN: 978-92-67-10780-6.

JAW, Y.-M., CHEN, Y.-Y., LEE, Y.-C., LEE, P.-H., JIANG, C.-M., TSAI, Y.-H., 2012. Histamine content and isolation of histamine-forming bacteria in fish meal and fish soluble concentrate. *Fisheries Science*, **78**(1), 155–162. ISSN: 0919-9268.

KAALE, L. D., EIKEVIK, T. M., RUSTAD, T., KOLSAKER, K., 2011. Superchilling of food: A review. *Journal of Food Engineering*, **107**, 141–146. ISSN: 0260-8774.

KALAČ, P., 2014. Health effects and occurrence of dietary polyamines: A review for the period 2005–mid 2013. *Food Chemistry*, **161**, 27–39. ISSN: 0308-8146.

KELLNER, R., 1998. *Analytical Chemistry: The Approved Text to the FECS Curriculum Aanalytical Chemistry*. Weinheim: Wiley-VCH. ISBN: 9783527288816.

LAGARIAS, J. C., REEDS, J. A., WRIGHT, M. H., WRIGHT, P. E., 1998. Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM Journal on Optimization*, **9**, 112–147. ISSN: 1052-6234.

LEHMANN, E. L., CASELLA, G., 1998. *Theory of Point Estimation*. 2nd ed. New York: Springer-Verlag. ISBN 978-0387985022.

LEHMANN, E. L., ROMANO, J. P., 2005. *Testing Statistical Hypotheses*. 3rd ed. New York: Springer. ISBN 978-0387988641.

LIGNELL, S., DARNERUD, P.O., AUNE, M., CNATTINGIUS, S., HAJSLOVA, J., SETKOVA, L., GLYNN, A., 2008. Temporal trends of synthetic musk compounds in mother's milk and associations with personal use of perfumed products. *Environmental Science & Technology*, **42**(17), 6743–6748. ISSN: 0013-936X.

LIKEŠ, J., MACHEK, J., 1988. *Mathematical Statistics (In Czech: Matematická statistika)*. 2nd ed. Praha: SNTL.

LISTER, A. S., 2005. Validation of HPLC Methods in Pharmaceutical Analysis. In: AHUJA, S., DONG, M. W., eds. *Handbook of pharmaceutical analysis by HPLC, Volume 6*. Amsterdam: Elsevier, 191–217. ISBN: 978-0120885473.

LUCKENBACH, T., EPEL, D., 2005. Nitromusk and polycyclic musk compounds as long-term inhibitors of cellular xenobiotic defense systems mediated by multidrug transporters. *Environmental Health Perspectives*, **113**(1), 17–24. ISSN: 1552-9924.

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. ISSN: 1352-2310.

MUNOZ, G., GIRAUDEL, J.-L., BOTTA, F., LESTREMAU, F., DÉVIER, M.-H., BUDZINSKI, H., LABADIE, P., 2015. Spatial distribution and partitioning behavior of selected poly- and perfluoroalkyl substances in freshwater ecosystems: A French nationwide survey. *Science of The Total Environment*, **517**, 48–56. ISSN: 0048-9697.

OSPAR COMMISSION, 2004. *The convention for the protection of the marine environment of the North-East Atlantic: Musk xylene and other musks*. London: OSPAR Commission. ISBN: 1-904426-36-0.

PAKYARI, R., BALAKRISHNAN, N., 2013. Testing exponentiality based on Type-I censored data. *Journal of Statistical Computation and Simulation*, **83**, 2369–2378. ISSN: 0094-9655.

PAKYARI, R., NIA, K. R., 2017. Testing goodness-of-fit for some lifetime distributions with conventional Type-I censoring. *Communications in Statistics - Simulation and Computation*, **46**, 2998–3009. ISSN: 0361-0918.

POUILLOT, R., HOELZER, K., CHEN, Y., DENNIS, S., 2013. Estimating probability distributions of bacterial concentrations in food based on data generated using the most probable number (MPN) method for use in risk assessment. *Food Control*, **29**(2), 350–357. ISSN: 0956-7135.

PRESTER, L., 2011. Biogenic amines in fish, fish products and shellfish: a review. *Food Additives and Contaminants: Part A – Chemistry, Analysis, Control, Exposure and Risk Assessment*, **28**(11), 1547–1560. ISSN: 1944-0049.

RAWLES, D. D., FLICK, G. J., MARTIN, R. E., 1996. Biogenic amines in fish and shelfish. *Advances in Food and Nutrition Research*, **39**, 329–365. ISSN: 1043-4526.

REGUEIRO, J., GARCIA-JARES, C., LLOMPART, M., LAMAS, J. P., CELA, R., 2009. Development of a method based on sorbent trapping followed by solid-phase microextraction for the determination of synthetic musks in indoor air. *Journal of Chromatography A*, **1216**(14), 2805–2815. ISSN: 0021-9673.

RIMKUS, G. G., 1999. Polycyclic musk fragrances in the aquatic environment. *Toxicology Letters*, **111**(1-2), 37–56. ISSN: 0378-4274.

SHALABY, A. R., 1996. Significance of biogenic amines to food safety and human health. *Food Research International*, **29**(7), 675–690. ISSN: 0963-9969.

SHORTEN, P. R., PLEASANTS, A. B., SOBOLEVA, T. K., 2006. Estimation of microbial growth using population measurements subject to a detection limit. *International Journal of Food Microbiology*, **108**(3), 369–375. ISSN: 0168-1605.

SILLA SANTOS, M. H., 1996. Biogenic amines: their importance in foods. *International Journal of Food Microbiology*, **29**(2–3), 213–231. ISSN: 0168-1605.

SINGH, A., SINGH, A. K., IACI, R. J., 2002. Estimation of the exposure point concentration term using a gamma distribution. Technology Support Center Issue 2002, EPA/600/R-02/084, EPA Technology Support Center for Monitoring and Site Characterization, Las Vegas, Nevada.

STRUCIŃSKI, P., MORZYCKA, B., GÓRALCZYK, K., HERNIK, A., CZAJA, K., KORCZ, W., MATUSZAK, M., MINORCZYK, M., LYCZEWSKA, M., PRUSS, B., LUDWICKI, J. K. , 2015. Consumer risk assessment associated with intake of pesticide residues in food of plant origin from the retail market in Poland. *Human and Ecological Risk Assessment: An International Journal*, **21**(8), 2036–2061. ISSN: 1080-7039.

SUMNER, N. R., GUITART, C., FUENTES, G., READMAN, J. W., 2010. Inputs and distributions of synthetic musk fragrances in an estuarine and coastal environment; a case study. *Environmental Pollution*, **158**(1), 215–222. ISSN: 0269-7491.

TEN BRINK, B., DAMINK, C., JOOSTEN, H. M., HUIS IN 'T VELD, J. H., 1990. Occurrence and formation of biologically active amines in foods. *International Journal of Food Microbiology*, **11**(1), 73–84. ISSN: 0168-1605.

VALERO, A., ORTIZ, J. C., FONGARO, G., HERNÁNDEZ, M., RODRÍGUEZ-LÁZARO, D., 2017. Definition of sampling procedures for collective-eating establishments based on the distribution of environmental microbiological contamination on food handlers, utensils and surfaces. *Food Control*, **77**, 8–16. ISSN: 0956-7135.

WENZL, T., HAEDRICH, J., SCHAECHTELE, A., ROBOUCH, P., STROKA, J., 2016.

Guidance document on the estimation of LOD and LOQ for measurements in the field of contaminants in feed and food; EUR 28099. Luxembourg: Publications Office of the European Union. ISBN: 978-92-79-61768-3.

ZHANG, W., XIAO, S., SAMARAWEERA, H., LEE, E. J., AHN, D. U., 2010. Improving functional value of meat products. *Meat Science*, **86**(1), 15–31. ISSN: 0309-1740.

ZLÁMALOVÁ GARGOŠOVÁ, H., ČÁSLAVSKÝ, J., VÁVROVÁ, M., 2013. Selected pharmaceuticals and musk compounds in wastewater. In: EINSCHLAG, F. S. G., ed. *Waste Water - Treatment Technologies and Recent Analytical Developments*. Rijeka: InTech, 121–144. ISBN: 978-953-51-0882-5.

ABSTRACT

The habilitation thesis is focused on statistical methods for Type I left-censored data with Weibull distribution and the exponential distribution as its special case. The maximum likelihood method is used for estimation of the unknown parameters. In order to describe variability of the parameters' estimates, the expected Fisher information matrix is derived together with the confidence intervals for the expected value. Performance of the estimators and confidence intervals is analyzed using simulations. Moreover, statistical tests for testing reduction of the censored Weibull distribution to the exponential submodel are proposed and analyzed. In addition, methods for comparison of expected values and also distributions of two independent and identically distributed Type I left-censored Weibull samples are described. Next, the goodness-of-fit tests for the Type I left-censored Weibull, lognormal and gamma distributions are described and analyzed using simulations. The described methods are applied in modelling of musk compounds and biogenic amines concentrations in fish tissue. In addition, Type I left-censored distributions were also applied in the extreme value theory for estimation of the extremal index.

ABSTRAKT

Habilitační práce je zaměřena na statistické metody pro analýzu zleva cenzorovaných dat s cenzorováním typu I, která mají Weibullovo rozdělení nebo exponenciální rozdělení jako speciální případ. Pro odhad parametrů je použita metoda maximální věrohodnosti. Pro popis variability těchto odhadů je odvozena očekávaná Fisherova informační matice a dále intervaly spolehlivosti pro střední hodnotu cenzorovaného Weibullova rozdělení. Chování odhadů a intervalů spolehlivosti je analyzováno pomocí simulací. Jsou navrženy a analyzovány statistické testy pro testování vhodnosti nahrazení cenzorovaného Weibullova rozdělení exponenciálním rozdělením. Dále jsou popsány metody pro porovnání středních hodnot a také rozdělení dvou nezávislých a stejně rozdělených zleva cenzorovaných dat s Weibullovým rozdělením. Jsou zde také popsány a pomocí simulací analyzovány testy dobré shody pro Weibullovo, lognormální a gama rozdělení zleva cenzorovaných dat s cenzorováním typu I. Popsané metody jsou použity při modelování koncentrací musk sloučenin a biogenních aminů v rybí tkáni. Kromě toho byla zleva cenzorovaná rozdělení s cenzorováním typu I použita v teorii extrémních hodnot pro odhad extrémálního indexu.