Proceedings of the 5^{th} International Conference on Statistics: Theory and Applications (ICSTA'23)
Brunel University, London, United Kingdom - August 03-05, 2023
Paper No. 168
DOI: 10.11159/icsta23.168

Validation of Two-sample Location-scale Model Using Empirical Likelihood Based Statistics

Leonora Pahirko[1], Janis Valeinis[1], Janis Gredzens[1], Madara Krumina[2]

[1]Laboratory of Statistical Research and Data Analysis, Faculty of Physics, Mathematics and Optometry / University of Latvia
[2]Faculty of Physics, Mathematics and Optometry / University of Latvia
Jelgavas iela 3, Riga, Latvia
leonora.pahirko@lu.lv; janis.valeinis@lu.lv; janis.gredzens@lu.lv; madara.krumina@inbox.lv

Abstract - One of the most popular semi-parametric relationships between two distribution functions is the location-scale model. Our goal is to develop a goodness-of-fit testing procedure for such location-scale models. For this purpose we use the plug-in approach by the estimation of the structural parameters and transformation of one of the samples. Further we propose to use the empirical likelihood test for the difference of two distribution functions and probability-probability plots. The effect of the estimation procedure on the limiting chi-square distribution is analysed by some simulation study. For practical applications we use the location-scale appropriate bootstrap resampling method to approximate the limiting distribution of the plug-in test statistic.

Keywords: location-scale model, empirical likelihood, plug-in, Mallows distance, probability-probability plots

1. Introduction

For the two group comparisons the conventional t-test is a powerful tool to detect the differences in locations, if the underlying distribution is normal. Otherwise, the Wilcoxon test is commonly used as an alternative procedure. Numerous non-parametric two-sample tests such as Lepage, Cucconi, Podgor-Gastwirth, Neuhauser among others have been developed to test simultaneously for the differences in locations and scales (see [7] for review). However, these tests are applicable only under the semi-parametric assumption that two distribution functions belong to the same location-scale family of distributions.

Such assumption generally might be true for randomized controlled trials, however within other study designs the location-scale assumption has to be verified. The goodness-of-fit test of the location-scale model itself has not been widely studied in the literature. Doksum and Sievers [2] introduced non-parametric simultaneous confidence bands for the general shift function based on the Kolmogorov-Smirnov statistic, which allow to detect a location-scale model between two samples. Freitag and Munk [3] described the general methodology to obtain a goodness-of-fit tests for the assessment of the validity of structural relationship models, including the location-scale models as well. Hall et.al. [4] developed a test based on the empirical characteristic functions to verify if the distributions belong to a location-scale family. Recently, the validation of a location-scale model was also studied by Subramanian [10] for censored data case using the plug-in empirical likelihood method.

The aim of our work is to develop an alternative procedure that can be used for goodness-of-fit testing of the two-sample location-scale model. We propose first to estimate the location and scale parameters and then to transform the second sample by the location-scale transformation. We use the two-sample empirical likelihood method to test the equality between the first and the transformed sample. Two empirical likelihood-based statistics are considered within this study. The first statistic is based on the difference between the distribution functions at a fixed point, which was introduced in [9], and the second statistic is based on the probability-probability plots at a fixed point that was established in [1]. Both test statistics are implemented in the R package *EL*. Due to the parameter estimation, the considered test statistics are not distribution-free and their limiting distribution converges to a scaled χ^2_1 random variable. Such plug-in effect on the empirical likelihood statistic in the general form was first discussed by Hjort et.al. [6], and various two-sample plug-in empirical likelihood problems were studied in [11].

In this work we propose to use two methods for the location and scale parameter estimation. First method replaces theoretical parameters by the sample estimators. In the second method the parameter estimates are obtained by minimizing the Mallows distance as introduced in [3]. Moreover, Potgieter and Lombard [8] recently proposed two other types of distance metrics based on the empirical characteristic functions for this purpose. We provide comparison of all methods by a small simulation study. We investigate the effect of the parameter estimation on the limiting distribution of the test statistic by the simulations as well, and we propose to use the nonparametric bootstrap resampling method for practical applications.

The structure of the paper is following. In Section 2 the semi-parametric location-scale model along with the two methods for the parameter estimation is considered. The two sample empirical likelihood method for the location-scale model testing is described in Section 3. The bootstrap resampling method is discussed in Section 4. Results of two simulation studies are given in Section 5. And finally, in Section 6 the proposed procedure of the location-scale model testing is used to normalize the data measured on separate plates.

2. Location-scale models and parameter estimation

Let X and Y denote two independent random variables with distribution functions F and G, respectively. The location-scale model between distribution functions F and G is defined as

$$F(x) = G\left(\frac{x-\mu}{\sigma}\right), \quad x \in \mathbb{R}, \tag{1}$$

where $\mu \in \mathbb{R}$ is the location and $\sigma > 0$ is the scale parameter. For the location-scale model the parameters can be expressed as

$$\sigma = \frac{\sigma_X}{\sigma_Y} \quad \text{and} \quad \mu = \mu_X - \sigma\mu_Y, \tag{2}$$

where μ_X, μ_Y, σ_X and σ_Y denote the means and standard deviations of X and Y, respectively. If $\sigma = 1$, the model (1) simplifies to the location model $F(x) = G(x - \mu)$; if $\mu = 0$, the model (1) simplifies to the scale model $F(x) = G(x/\sigma)$.

In order to validate the location-scale model (1) between two distributions, we aim to test the hypothesis

$$H_0 : \text{There exist constants}\, \mu \in \mathbb{R}\, \text{and}\, \sigma > 0\, \text{such that}\, X \overset{\mathcal{D}}{=} \sigma Y + \mu, \tag{3}$$

where $\mathcal{D}$ denotes the equality of random variables in distribution. To check the hypothesis (3), the test based on characteristic functions was introduced in [4]. Their approach was based on the standardization of both samples using the sample means and standard deviations. However, our approach is different. We propose to use the following three-step procedure. First, obtain some consistent estimates $\hat{\mu}$ and $\hat{\sigma}$ of the location and the scale parameters μ and σ. Second, transform one of the samples, say Y, using the location-scale transformation $Y^* = \hat{\sigma}Y + \hat{\mu}$. Finally, test the hypothesis

$$H_0 : F(x) = G^*(x), \quad x \in \mathbb{R}, \tag{4}$$

where $G^*(x)$ denote the distribution function of Y^*.

For the estimation of μ and σ a natural choice is to directly replace the theoretical values in (2) with sample estimators leading to the estimators $\hat{\sigma} = S_X/S_Y$ and $\hat{\mu} = \bar{X} - \hat{\sigma}\bar{Y}$, where $\bar{X}$ and $\bar{Y}$ denote the corresponding sample means and S_X and S_Y denote the corresponding sample standard deviations. Note that the location-scale model can be expressed in terms of quantile functions as

$$F^{-1}(t) = \sigma G^{-1}(t) + \mu, \quad t \in (0,1),\, \mu \in \mathbb{R},\, \sigma > 0, \tag{5}$$

where F^{-1} and G^{-1} are the inverse functions of F and G. The equation (5) was used in [3] to define the Mallows distance as

$$M(F,G) := \int_0^1 \left(F^{-1}(u) - \sigma G^{-1}(u) - \mu\right)^2 \mathrm{d}u. \tag{6}$$

Thus, the estimators $\hat{\mu}$ and $\hat{\sigma}$ can be obtained by minimizing the Mallows distance, in which F and G are replaced by the corresponding empirical distribution functions denoted by F_n and G_m.

Generally, to test whether the hypothesis (4) holds true, we could use any appropriate two-sample goodness-of-fit test statistic based on X and Y^*. However, often the parameter estimation affects the limiting distribution of the test statistic under the null hypothesis. For example, the one-sample Kolmogorov-Smirnov normality test illustrates the effect of the parameter estimation on the limiting distribution, where the Lilliefor's correction is required. The same applies for other test statistics based on the empirical distribution or quantile functions, and in the case of the two-sample Kolmogorov-Smirnov test statistic as well (see [4] for more details).

3. Two-sample empirical likelihood method

In this section we briefly describe the two-sample empirical likelihood method for the difference of distribution functions and for probability-probability plots. Let $X_1, \ldots, X_n$ and $Y_1, \ldots Y_m$ be two independent random samples from distributions F and G, respectively. Let us denote the univariate parameter of interest with Δ and the vector of a location and scale parameters with ν for simplicity, i.e., $\nu = (\mu, \sigma)$. The true parameters are determined by the unbiased estimating equations

$$E\{w_1(X, \theta_0, \Delta_0, \nu_0)\} = 0,$$
$$E\{w_2(Y, \theta_0, \Delta_0, \nu_0)\} = 0,$$

where Δ_0 denotes the true value of Δ, θ_0 is the true value of the nuisance parameter $\theta \in \mathbb{R}$ associated with one of the distribution functions, and $\nu_0 = (\mu_0, \sigma_0)$ denotes the vector of the true location and scale parameter values.

The two-sample empirical likelihood function is defined as

$$\begin{aligned} L(\Delta, \theta, \nu) = n^n m^m \sup_{\theta, p, q} \Bigg\{ & \prod_{i=1}^n p_i \prod_{j=1}^m q_j \ : \\ & p_i \geq 0,\ i = 1, \ldots, n,\ \sum_{i=1}^n p_i = 1,\ \sum_{i=1}^n p_i w_1(X_i, \theta, \Delta, \nu) = 0, \\ & q_j \geq 0,\ j = 1, \ldots, m,\ \sum_{j=1}^m q_j = 1,\ \sum_{j=1}^m q_j w_2(Y_j, \theta, \Delta, \nu) = 0 \Bigg\}, \end{aligned} \tag{7}$$

where $p = (p_1, \ldots, p_n)$ and $q = (q_1, \ldots, q_m)$ denote the probability vectors. Optimal p and q can be found by Lagrange multiplier method if the solution to the maximization problem exists, i.e., if 0 belongs to interior of the convex hull of $w_1(X_1, \theta, \Delta, \nu), \ldots, w_1(X_n, \theta, \Delta, \nu)$ and $w_2(Y_1, \theta, \Delta, \nu), \ldots, w_2(Y_m, \theta, \Delta, \nu)$.

Example 1. Difference of distribution functions at a fixed point. Empirical likelihood method for various differences of two populations was studied by Qin and Zhao [9]. For the location model let the parameter of interest be $\Delta = G\left(\frac{x-\mu}{\sigma}\right) - F(x)$, $x \in \mathbb{R}$, and denote the true nuisance parameter as $\theta_0 = F(x)$. In this case the estimating equations are

$$w_1(X, \theta_0, \Delta_0, \nu_0) = I_{\{X \leq x\}} - \theta_0, \tag{8}$$
$$w_2(Y, \theta_0, \Delta_0, \nu_0) = I_{\{\sigma_0 Y + \mu_0 \leq x\}} - \theta_0 - \Delta_0, \tag{9}$$

where $I_{\{\cdot\}}$ denotes the indicator function. If $\mu_0 = 0$ and $\sigma_0 = 1$, then $\Delta_0 = G(x) - F(x)$, which is the case considered by Qin and Zhao [9].

Example 2. Probability-probability plot at a fixed point. Empirical likelihood for probability-probability plots at a fixed point was studied by Claeskens et.al. [1]. Define the parameter of interest Δ as

$$\Delta = F\left(\sigma G^{-1}(t) + \mu\right),\ t \in (0, 1). \tag{10}$$

The estimating equations in this case are

$$w_1(X, \theta_0, \Delta_0, \nu_0) = I_{\{X \leq \theta_0\}} - \Delta_0, \tag{11}$$
$$w_2(Y, \theta_0, \Delta_0, \nu_0) = I_{\{\sigma_0 Y + \mu_0 \leq \theta_0\}} - t, \tag{12}$$

where $\theta_0 = \sigma_0 G^{-1}(t) + \mu_0$ denotes the true nuisance parameter. If $\mu_0 = 0$ and $\sigma_0 = 1$, then $\Delta_0 = F(G^{-1}(t))$ is the classical probability-probability plot studied in [1]. The functions (11) and (12) are not smooth with respect to θ, in which case Claeskens et.al. [1] suggested to use the smoothed unbiased estimating equations

$$w_1(X, \theta_0, \Delta_0, \nu_0) = H_{b_1}(\theta_0 - X) - \Delta_0, \tag{13}$$
$$w_2(Y, \theta_0, \Delta_0, \nu_0) = H_{b_2}(\theta_0 - (\sigma_0 Y + \mu_0)) - t, \tag{14}$$

where $H_{b_j}(t) = H_j(t/b_j)$ denotes the kernel density estimator, and $b_1 = b_1(n)$, $b_2 = b_2(m)$ are bandwidth sequences converging to zero as $n, m \to \infty$ (see [1] for more detail).

If $\mu_0 = 0$ and $\sigma_0 = 1$, under some regularity conditions it was shown for both examples in [9] and [1] that the empirical likelihood statistic $-2 \log L(\Delta_0, \hat{\theta}(\Delta_0, \nu_0), \nu_0)$ is distribution free and converges to a χ_1^2 random variable, where

$$\hat{\theta}(\Delta, \nu) = \arg\min_{\theta} \left(-2 \log L(\Delta, \theta, \nu)\right).$$

Hjort et.al. [6] introduced the plug-in empirical likelihood method demonstrating the effect of plug-in parameter estimators on the limiting distribution under the null hypothesis. Usually the plug-in empirical likelihood statistic has the scaled χ_1^2 distribution in one-sample case. For various two-sample problems the plug-in empirical likelihood method was extended in [11]. The following result follows from [11] for the parameter of interest Δ as described in Examples 1 and 2 using the estimation for ν.

Theorem 1 *Under some regularity conditions (see [11]) the two-sample plug-in empirical likelihood statistic converges to scaled χ_1^2 distribution, i.e.,*

$$-2 \log L(\Delta_0, \hat{\theta}(\Delta_0, \hat{\nu}), \hat{\nu}) \xrightarrow{d} r\chi_1^2, \tag{15}$$

where r is some scaling constant.

Using Theorem 1 $(1 - \alpha)100\%$ confidence interval for Δ is obtained as

$$\{\Delta : -2 \log L(\Delta, \hat{\theta}(\Delta, \hat{\nu}), \hat{\nu}) < c\}, \tag{16}$$

where c is $1 - \alpha$ quantile of the distribution $r\chi_1^2$ and $\hat{\nu} = (\hat{\mu}, \hat{\sigma})$.

The asymptotic behaviour of the limiting distribution was analysed using simulation study (see Section 5.2), figure 1 shows the comparison of simulated limiting distributions of the test statistics with theoretical χ_1^2 distribution for Δ's described both in Example 1 and 2. The scaling constant r depends on the underlying distribution of samples X and Y, and is complicated to estimate. Instead we propose to use the bootstrap resampling for practical applications.

4. The bootstrap resampling for the empirical likelihood method and construction of simultaneous confidence bands

In this section we provide the description of application of the bootstrap resampling method for two cases. First, the procedure of bootstrap resampling to obtain the limiting distribution of the statistic for practical applications, and second, the procedure to construct the simultaneous confidence bands for the parameter of interest over a partition of some interval.

4.1. The bootstrap procedure for the limiting distribution of the test statistic

We need to use the appropriate location-scale bootstrap resampling. Similar approach for censored data case was described in [10]. The procedure is as follows.

1. Obtain the estimates of location and scale parameters μ and σ using one of the proposed estimation procedures;

2. Fix one of the samples, say X, and obtain the transformed sample X^* by inverse location-scale transformation using the estimates of μ and σ as
$$X^* = \frac{X - \hat{\mu}}{\hat{\sigma}}.$$
The inverse transformation using $\hat{\mu}$ and $\hat{\sigma}$ ensure that the null hypothesis of location-scale model holds true for X and X^*;

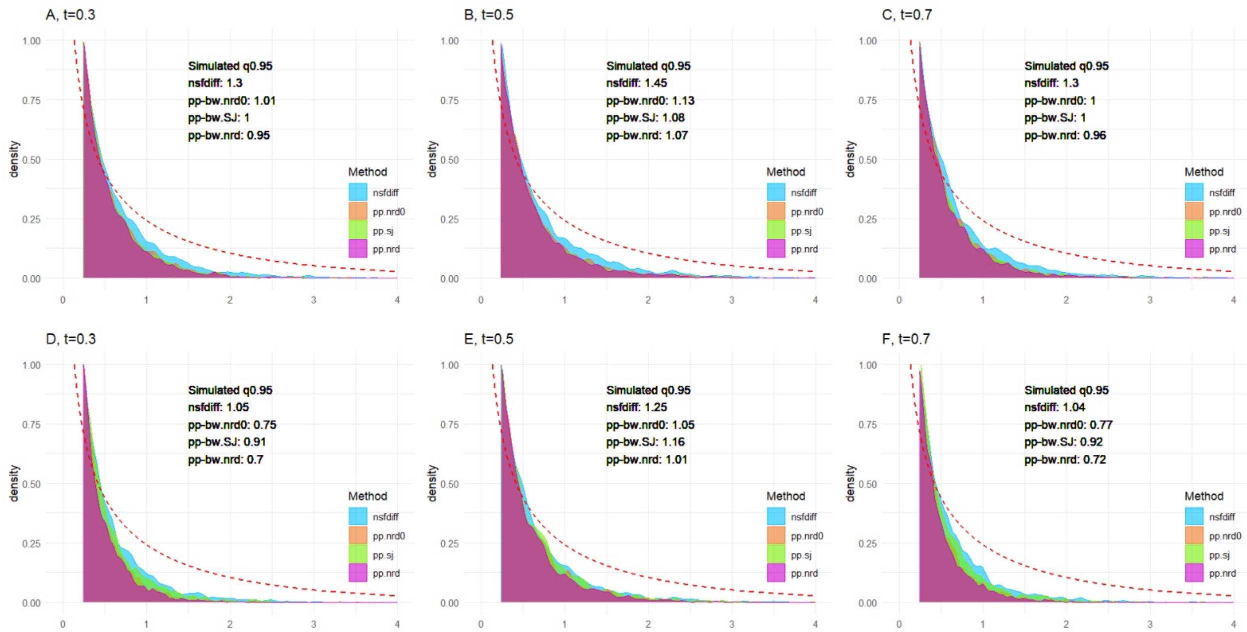


Fig. 1: Comparison of the χ^2_1 distribution (dashed line) with the simulated limiting distributions for the empirical likelihood statistic for the difference of distribution functions (*nsfdiff*) at a fixed point $x = q_t$ and for the probability-probability plots (*pp*) at a fixed point t, where $t = 0.3, 0.5, 0.7$. Samples of size $n = m =$10 000 were drawn 10 000 times from normal distribution family in the upper panels (A, B, C), and from the uniform distribution family in the lower panels (D, E, F) with the true parameters $\mu_0 = 0$ and $\sigma_0 = 1$. Three methods for the smoothing bandwidth selection were used in case of *pp*: *bw.nrd0*, *bw.SJ*, and *bw.nrd*.

3. Draw nonparametric bootstrap samples from X and X^* 10 000 times and calculate the statistic for each pair of samples.

The location-scale hypothesis test is performed by calculating the empirical likelihood statistic between X and Y and comparing to the bootstrapped critical value $q_{1-\alpha}$, which is the empirical $1-\alpha$ quantile of the distribution obtained in step 3.

4.2. Simultaneous confidence bands for the parameter of interest over some interval

The procedure for constructing a simultaneous confidence bands over some interval (a, b) using bootstrap resampling method first was introduced in [5], but was used in [13] as well. The procedure without plug-in parameter estimation is implemented in R package `EL` [12]. To carry out the procedure, define the maximum empirical likelihood estimator as

$$\hat{\Delta} = \arg\max_{\Delta} L(\Delta, \hat{\theta}, \hat{\nu}), \tag{17}$$

where $\hat{\theta} = \hat{\theta}(\Delta, \hat{\nu})$, and calculate $\hat{\Delta}$ for all $x \in (a, b)$. The bootstrap critical value c^* is chosen such that

$$P\left(-2\log L^*(\hat{\Delta}, \hat{\theta}, \hat{\nu}) \leq c^* \quad \text{for} \quad a \leq x \leq b\right) = 1 - \alpha, \tag{18}$$

where L^* is the empirical likelihood function for bootstrapped samples from X and Y. Next, find maximum statistic over the interval (a, b) and repeat the bootstrapping to find the distribution of maxima, which gives us the $1-\alpha$ quantile for construction of simultaneous confidence bands.

By constructing simultaneous confidence bands, we can carry out the goodness-of-fit of the location-scale model over some interval (a, b). For the difference of the distribution functions location-scale model holds if the horizontal line $y = 0$ falls within the confidence bands, for the probability-probability plots we need to check if the diagonal $y = x$ falls within the bands. The graphical approach of the goodness-of-fit testing demonstrates, how exactly both distributions differ and where the location-scale model does not hold in case the hypothesis is rejected.

5. Simulation study results

In this section, we demonstrate two simulation studies. The first study is designed to compare methods for the parameter estimation, the second study demonstrates the empirical quantiles of the limiting distribution for the empirical likelihood statistic with estimated parameters.

5.1. Comparison of two methods for location and scale parameter estimation

Here we demonstrate the results for small simulation study to compare two methods described in Section 2 for estimation of location parameter μ and scale parameter σ. Our goal is to compare two methods for the parameter estimation: direct estimation by sample estimators (2) and estimators obtained by Mallows distance (6). Two samples of size $n = m = 100, 250$ from $N(0,1)$, Student's distributions with 3 and 5 degrees of freedom (t_3 and t_5, respectively) were drawn 10 000 times. We use the exact setup from [8] in order to compare results with two methods (k-L and WI) based on empirical characteristic functions summarized in Table 4 - Table 6 in [8]. The true location and scale parameters in all cases are chosen as $\mu_0 = 0$ and $\sigma_0 = 1$. Estimated standard errors and biases of obtained estimates for all scenarios are summarized in Table 1, the rows corresponding to the methods k-L and WI were chosen from [8] according to the parameters with the best performance. The Mallows distance estimates were obtained using R function `optim` with quasi-Newton method `"BFGS"`.

As can be seen from the simulation results in Table 1, both methods perform similarly for the normal distribution. For t_5 distribution Mallows distance gives slightly better estimates of μ and substantially better estimates of σ. In case of t_3 distribution, Mallows distance performs noticeably better. In comparison with the results from [8], only in some cases their methods give better results that direct estimation and Mallows distance.

Table 1: Standard errors (SE) and biases of 10 000 times calculated estimates of location parameter μ and scale parameter σ by the sample estimators and the Mallows distance for two samples simulated from given distribution of size $n = m = 100, 250$. The true parameter values are $\mu_0 = 0$ and $\sigma_0 = 1$. For methods k-L and WI results added from Potgieter and Lombard [8] for comparison.

		$n = m = 100$				$n = m = 250$			
Distribution	Method	$SE(\hat{\mu})$	$bias(\hat{\mu})$	$SE(\hat{\sigma})$	$bias(\hat{\sigma})$	$SE(\hat{\mu})$	$bias(\hat{\mu})$	$SE(\hat{\sigma})$	$bias(\hat{\sigma})$
$N(0,1)$	Sample estimators	0.142	-0.003	0.101	0.004	0.089	-0.001	0.064	0.001
	Mallows	0.141	0.004	0.108	-0.003	0.089	0.000	0.066	-0.002
	Potgieter & Lombard (k-L)	0.183	0.005	0.161	0.136	0.105	0.001	0.097	0.074
	Potgieter & Lombard (WI)	0.143	0.003	0.105	0.006	0.090	0.000	0.065	0.002
t_3	Sample estimators	0.246	-0.002	0.365	0.049	0.153	-0.001	0.257	0.026
	Mallows	0.198	-0.001	0.155	-0.005	0.127	-0.001	0.092	0.005
	Potgieter & Lombard (k-L)	0.246	-0.003	0.339	0.200	0.118	-0.001	0.128	0.061
	Potgieter & Lombard (WI)	0.175	-0.001	0.148	0.017	0.111	-0.001	0.089	0.008
t_5	Sample estimators	0.180	-0.001	0.173	0.015	0.116	0.000	0.112	0.007
	Mallows	0.170	0.002	0.131	-0.004	0.108	-0.001	0.081	0.001
	Potgieter & Lombard (k-L)	0.242	-0.001	0.312	0.233	0.122	0.000	0.167	0.098
	Potgieter & Lombard (WI)	0.166	0.002	0.131	0.011	0.104	0.000	0.082	0.004

5.2. Simulated quantiles of the limiting distribution for the plug-in empirical likelihood statistic

The limiting distribution of the plug-in empirical likelihood statistic is the scaled χ^2_1 distribution (see Theorem 1). To demonstrate that we conduct a small simulation study and report the 95% critical values corresponding to the empirical 0.95-th quantiles for both statistics given in Example 1 and 2. We simulate two samples of size $n = m =$10 000 under the location-scale model from

1. the normal distribution family with $X \sim N(\mu_X, \sigma^2_X)$ and $Y \sim N(0,1)$,
2. the uniform distribution family with $X \sim U[a,b]$ and $Y \sim U[0,1]$,

where different combinations of parameters μ_X, σ_Y, a and b were chosen. Next, 10 000 times the statistic using samples X and Y^* was calculated and the empirical 0.95-th quantile of the obtained limiting distribution was found for $t = 0.3, 0.5, 0.7$. For the difference of the distribution functions statistic was calculated at the point $x = q_t$, where

q_t denotes t-th quantile of the theoretical distribution, and for the probability-probability plot at the point t (see Table 2 and Figure 1 for the results).

The estimates of μ and σ were obtained using only the direct sample estimators due to high calculation time for the optimisation of the Mallows distance. The statistic values were calculated by the function *EL.statistic* from R library `EL` [12], using the methods *"nsfdiff"* and *"pp"* for the difference of the distribution functions and probability-probability plot at a fixed point, respectively. For probability-probability plot three different smoothing bandwidths were selected using built-in R functions *bw.nrd0*, *bw.nrd* and *bw.SJ*.

Results in Table 2 demonstrates that empirical 0.95-th quantiles are quite similar within the same distribution family with chosen different parameter values. Therefore, the limiting distribution of the statistic with estimated parameters might not be affected by the values of the true location and scale parameters. However, for probability-probability plots we can observe that quantiles differ along with the choice of bandwidth selection method. Empirical quantiles also depends on the point t at which the statistic is calculated.

Table 2: Empirical 0.95-th quantiles of the distribution of 10 000 times simulated empirical likelihood statistic for the difference of the distribution functions (*nsfdiff*) calculated at point $x = q_t$ and the probability-probability plot (*pp*) calculated at point t. For normal distributions $Y \sim N(0, 1)$, for uniform distributions $Y \sim U[0, 1]$, sample size $n = m =$10 000. Three methods were used for bandwidth selection in case of *pp*.

X	$t = 0.3$				$t = 0.5$				$t = 0.7$			
	nsfdiff	*pp*			*nsfdiff*	*pp*			*nsfdiff*	*pp*		
		bw.nrd0	*bw.nrd*	*bw.SJ*		*bw.nrd0*	*bw.nrd*	*bw.SJ*		*bw.nrd0*	*bw.nrd*	*bw.SJ*
$N(0, 1)$	1.30	1.01	0.95	1.00	1.45	1.13	1.07	1.08	1.30	1.00	0.96	1.00
$N(0.5, 1)$	1.35	0.99	0.94	0.98	1.38	1.09	1.04	1.04	1.35	1.00	0.96	0.98
$N(1, 1)$	1.30	1.01	0.96	0.99	1.35	1.09	1.03	1.03	1.34	1.02	0.98	1.00
$N(2, 1)$	1.36	1.01	0.96	0.99	1.41	1.12	1.09	1.09	1.25	1.00	0.96	0.99
$N(0, 0.5^2)$	1.36	1.04	0.98	1.04	1.38	1.08	1.03	1.04	1.29	1.02	0.96	1.02
$N(0, 3.14^2)$	1.32	1.00	0.95	0.99	1.38	1.08	1.04	1.05	1.34	0.97	0.93	0.97
$N(0, 7^2)$	1.33	1.06	1.01	1.05	1.34	1.10	1.05	1.06	1.26	1.05	1.00	1.04
$N(1, 2^2)$	1.30	1.01	0.96	1.01	1.38	1.09	1.04	1.05	1.33	1.01	0.97	1.00
$N(2, 0.5^2)$	1.30	1.04	0.99	1.02	1.41	1.11	1.07	1.07	1.28	1.00	0.95	0.98
$N(5, 5^2)$	1.33	1.01	0.96	0.99	1.41	1.06	1.01	1.01	1.27	1.00	0.96	0.98
$U[0, 1]$	1.05	0.75	0.70	0.91	1.25	1.05	1.01	1.16	1.04	0.77	0.72	0.92
$U[0.5, 1.5]$	1.05	0.75	0.70	0.91	1.25	1.02	0.99	1.14	1.04	0.77	0.72	0.90
$U[-2, -1]$	1.06	0.78	0.72	0.91	1.31	1.05	1.01	1.16	1.01	0.75	0.71	0.91
$U[7, 8]$	1.04	0.78	0.74	0.93	1.28	1.02	0.99	1.13	1.05	0.77	0.72	0.92
$U[-1, 2]$	1.06	0.76	0.71	0.92	1.31	1.03	1.00	1.13	1.04	0.79	0.74	0.92
$U[0.25, 0.75]$	1.08	0.74	0.70	0.90	1.28	1.03	0.99	1.14	1.05	0.75	0.72	0.90
$U[0.4, 0.6]$	1.06	0.76	0.71	0.92	1.28	1.05	1.00	1.17	1.04	0.77	0.72	0.89
$U[0, 6]$	1.04	0.74	0.70	0.92	1.34	1.04	1.00	1.17	1.09	0.76	0.71	0.91
$U[-3, -1]$	1.05	0.76	0.71	0.92	1.28	1.07	1.03	1.18	1.03	0.78	0.73	0.94
$U[-3.14, 3.14]$	1.04	0.76	0.72	0.92	1.25	1.05	1.00	1.18	1.07	0.79	0.74	0.93

6. Application of the method

Potential markers of inflammation were studied in Latvian patients with type 1 diabetes mellitus. Lipopolysaccharide binding protein (LBP) was measured using ELISA kit on three separate plates along with the four other potential markers of inflammation. Numbers of blood samples tested on plate 1, 2 and 3 were 36, 36 and 31, respectively. In this section we illustrate the use of the location-scale goodness-of-fit testing to carry out the normalization of data measured on separate plates.

Variability of LBP measurements across the plates was detected visually (see panel A of figure 2), and confirmed using Kruskal-Wallis test for locations ($p < 0.001$). Moreover, significant differences were observed specifically between plate 1 and 3 and between plate 2 and 3.

In order to use the data for further analysis, it was decided to normalize data on plate 3 with respect to the data on plate 1 and 2. To perform the normalization we consider the following steps: combine observations from plates 1 and

2 in one sample as X and take plate 3 as Y; test, whether the two-sample location-scale model holds between X and Y; if the previous hypothesis is not rejected, transform Y by location-scale transformation using parameter estimates to obtain normalized data on plate 3.

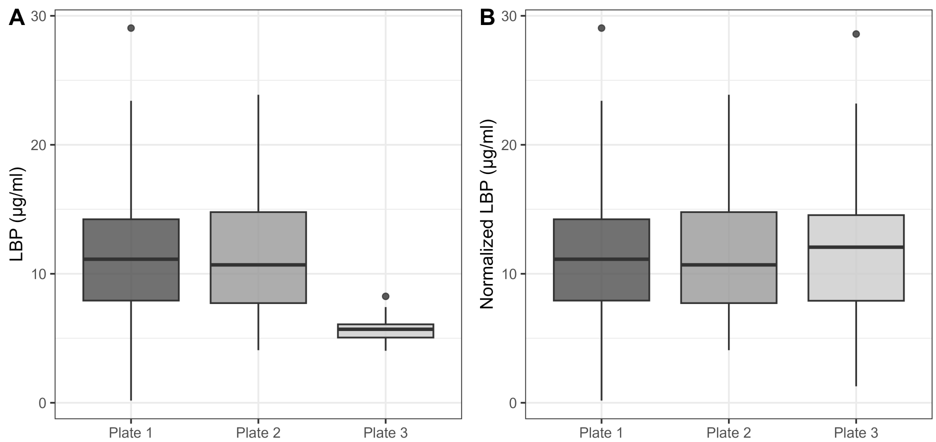


Fig. 2: Boxplots for LBP data by plates before (A) and after (B) the normalization procedure using location-scale transformation with the parameter estimates obtained by direct sample estimators.

The location and scale parameter estimates with 95% bootstrap percentile confidence intervals using sample estimators were $\hat{\mu} = -24.88\,(-43.4, -15.3)$ and $\hat{\sigma} = 6.48\,(4.68, 9.97)$, and using Mallows distance were $\hat{\mu} = -27.25\,(-40.2, -20.2)$ and $\hat{\sigma} = 6.93\,(5.39, 9.48)$. According to the test based on empirical characteristic functions by Hall et.al. [4] the location-scale model between distributions of both samples is not rejected (test statistic was 0.34, 500 times bootstrapped critical value was 2.77, and $p = 0.757$).

For further analysis, we only use the method with the sample estimators to reduce the computation time for bootstrap resampling. We calculate the empirical likelihood statistic for the difference of distribution functions (smoothed version is implemented by the method *fdiff* and is used for better visualization) and for the probability-probability plot to construct pointwise and simultaneous confidence (see figure 3). The method *bw.SJ* was used to select the smoothing bandwidth. We check if the horizontal line $y = 0$ falls within the bands in case of *fdiff* or if the diagonal $y = x$ falls within the bands in case of *pp*. As can be seen in Figure 3 we can not reject the hypothesis of location-scale model between both samples X and Y. Additionally, we can see that the largest deviation from the null hypothesis is observed at the right tail of the distributions. The pointwise interval bounds cross the reference lines approximately at $t = 0.9$.

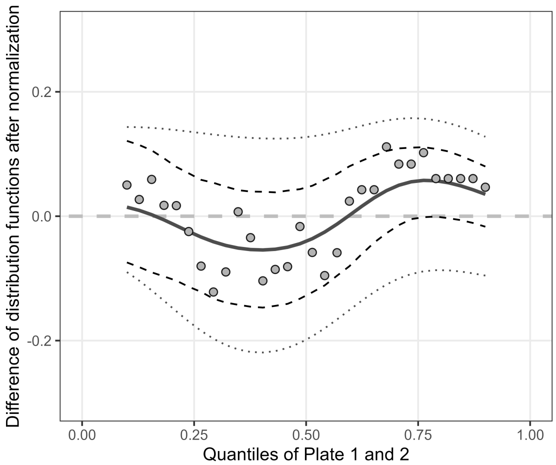
Finally, as the hypothesis of location-scale model is not rejected, we obtain normalized data on plate 3 by applying the location-scale transformation. Observations across the plates are visually homogeneous after the normalization (see panel B in figure 2).

7. Conclusions

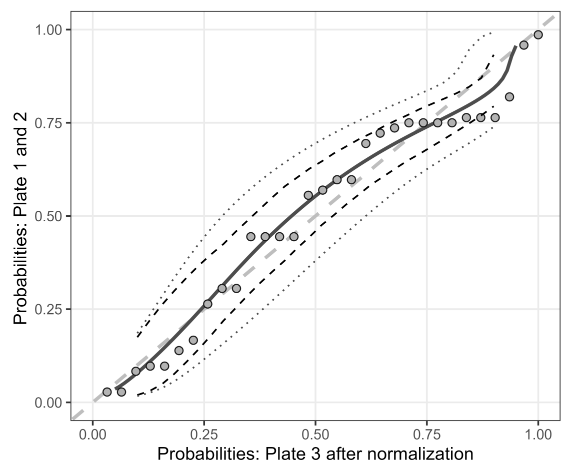
In this study the goodness-of-fit procedure for the two-sample location-scale models was established using the plug-in empirical likelihood method for the difference of the distribution functions and for the probability-probability plots at a fixed point.

Two methods for the estimation of the location and scale parameters were studied within this work. We considered the direct plug-in sample estimators and the method by minimizing the Mallows distance. Recently two new methods based on empirical characteristic functions were introduced in [8]. The comparison of all methods are provided in Table 1. For the future work two additional methods can be proposed by using robust sample estimators and by replacing the empirical distribution functions in the Mallows distance with the smoothed versions. However, in the latter case the choice of the smoothing parameter needs to be considered.

The limiting distribution of the empirical likelihood statistic is affected by the parameter estimation procedure (see Table 2). Simulation study shows that scaling constant depends on the underlying distribution, but might be the same for the different true parameter values. The method for the difference of the distribution functions (*nsfdiff*) does not depend on the choice of the smoothing parameter, which is an advantage over the probability-probability plot method (*pp*).



(a) *fdiff*



(b) *pp*

Fig. 3: Plots for goodness-of-fit testing of location-scale model in LBP data example. The smoothing bandwidth was selected by *bw.SJ*.

We propose to use a graphical illustration of the location-scale model testing. This is convenient for a visual assessment of the differences between two samples to examine at which points exactly and what kind of the discrepancies are observed. Such a graphical tool can be used in addition to other testing procedures, for example, with the test developed in [4].

The proposed method also can be used as a pretest for the validation of the location-scale model between two distributions. If the location-scale model is not rejected, a test for simultaneous detection of the location and scale differences such as Lepage, Cucconi, Podgor-Gastwirth or Neuhauser (see [7] for the review) can be applied. We demonstrate the application of the procedure to test the location-scale model for the laboratory data observed on separate plates. If the variability between two plates are observed and the location-scale model is not rejected, we propose to normalize the data using the location-scale transformation. In practice, more than two separate plates are often used, thus the extension for the normalization across multiple plates can be studied in the future.

Acknowledgements

L.Pahirko's work is funded by the grant of the project "Strengthening of the Capacity of Doctoral Studies at the University of Latvia within the Framework of the New Doctoral Model" (identification No. 8.2.2.0/20/I/006) and partially by the grant of the project No. lzp-2020/1-0138. Data analysed in Section 6 was acquired within the scope of the project "Intestinal inflammation as a potentially modifiable risk factor for complications in type 1 diabetes" (Project No. 1.1.1.2/VIAA/3/19/525).

References

[1] G. Claeskens, B. Jing, L. Peng, and W. Zhou. "Empirical likelihood confidence regions for comparison distributions and ROC curves". In: *Canadian Journal of Statistics* 31.2 (2003), pp. 173–190.

[2] K.A. Doksum and G.L. Sievers. "Plotting with confidence: Graphical comparisons of two populations". In: *Biometrika* 63.3 (1976), pp. 421–434.

[3] G. Freitag and A. Munk. "On Hadamard differentiability in k-sample semiparametric models–with applications to the assessment of structural relationships". In: *Journal of Multivariate Analysis* 94.1 (2005), pp. 123–158.

[4] P. Hall, F. Lombard, and C.J. Potgieter. "A new approach to function-based hypothesis testing in location-scale families". In: *Technometrics* 55.2 (2013), pp. 215–223.

[5] P. Hall and A.B. Owen. “Empirical likelihood confidence bands in density estimation”. In: *Journal of Computational and Graphical Statistics* 2.3 (1993), pp. 273–289.

[6] N.L. Hjort, I.W. McKeague, and I. Van Keilegom. “Extending the Scope of Empirical Likelihood”. In: *The Annals of Statistics* 37.3 (2009), pp. 1079–1111.

[7] M. Marozzi. “Nonparametric simultaneous tests for location and scale testing: a comparison of several methods”. In: *Communications in Statistics-Simulation and Computation* 42.6 (2013), pp. 1298–1317.

[8] C.J. Potgieter and F. Lombard. “Nonparametric two-sample estimation of location and scale parameters from empirical characteristic functions”. In: *Journal of Statistical Computation and Simulation* 86.16 (2016), pp. 3225–3242.

[9] Y.S. Qin and L.C. Zhao. “Empirical likelihood ratio confidence intervals for various differences of two populations”. In: *Journal of Systems Science and Complexity* 13.1 (2000), pp. 23–30.

[10] S. Subramanian. “Function-based hypothesis testing in censored two-sample location-scale models”. In: *Lifetime Data Analysis* 26.1 (2020), pp. 183–213.

[11] J. Valeinis. “Confidence bands for structural relationship models”. PhD thesis. Goettingen, Germany, 2007.

[12] J. Valeinis and E. Cers. *EL: Two-Sample Empirical Likelihood*. R package version 1.2. 2022. URL: `https://CRAN.R-project.org/package=EL`.

[13] J. Valeinis, E. Cers, and J. Cielens. “Two-sample problems in statistical data modelling”. In: *Mathematical modelling and analysis* 15.1 (2010), pp. 137–151.