UNIVERSITY
OF LATVIA

Summary of
Doctoral Thesis

Māra Delesa-Vēliņa

EMPIRICAL LIKELIHOOD METHOD FOR A LOCATION PARAMETER USING SOME ROBUST ESTIMATORS

Rīga 2022

UNIVERSITY
OF LATVIA

FACULTY OF PHYSICS, MATHEMATICS AND OPTOMETRY

Māra Delesa-Vēliņa

EMPIRICAL LIKELIHOOD METHOD FOR A LOCATION PARAMETER USING SOME ROBUST ESTIMATORS

SUMMARY OF DOCTORAL THESIS

Submitted for the degree of Doctor of mathematics
Subfield of Probability theory and mathematical statistics

Rīga 2022

The doctoral thesis was carried out at the Chair of Mathematical analysis, Department of Mathematics, Faculty of Physics, Mathematics and Optometry, University of Latvia, from 2012 to 2021.

The thesis contains the introduction, six chapters, conclusion, theses and bibliography.

Form of the thesis: dissertation in mathematics, probability theory and mathematical statistics.

Supervisor: *Dr. math.* Jānis Valeinis, Professor, Leading Researcher, University of Latvia

Reviewers:

1. *Ph. D.* Mihai Giurcanu, Research Associate Professor, University of Chicago, USA
2. *Ph. D.* Tõnu Kollo, Professor Emeritus, University of Tartu, Estonia
3. *Dr. habil. math.* Aleksandrs Šostaks, Leading Researcher, University of Latvia, Latvia

The thesis will be defended at the public session of the Doctoral Committee of Mathematical Sciences, University of Latvia, on March 18, 2022.

The thesis is available at the Library of the University of Latvia, Kalpaka blvd. 4.

This thesis is accepted for the commencement of the degree of Doctor of mathematics on November 11, 2021, by the Doctoral Committee of Mathematical Sciences, University of Latvia.

Chairman of the Doctoral Committee: *Dr. math.*, prof. Andrejs Cibulis
Secretary of the Doctoral Committee: *Dr. math.*, doc. Sergejs Smirnovs

© University of Latvia, 2022
© Māra Delesa-Vēliņa, 2022

ISBN 978-9934-18-781-0
ISBN 978-9934-18-782-7 (PDF)

Abstract

In this research empirical likelihood methods for comparing two and multiple independent populations based on robust location estimators are developed. Empirical likelihood (EL) is a nonparametric statistics method that does not require the normality assumption of the data. New asymptotic results are proven for the following empirical likelihood-based methods. 1. The difference of two M-estimators (in particular, two smoothed Huber estimators), 2. the difference of two trimmed means and 3. EL-based ANOVA method for comparing multiple trimmed means. A simulation study was designed and data examples were analysed showing that the newly-established methods provide a comparable alternative to the classical procedures when the data is normally distributed, demonstrating similar power and ability to control the type I error. In addition, the methods have good robustness properties, having an advantage over the classical procedures when the assumption of normality does not hold.

Keywords: empirical likelihood; robust statistics; M-estimator; smoothed M-estimator; trimmed mean; two-sample problem; EL ANOVA

Contents

Introduction

Motivation of the research

A common problem in statistical analysis is to compare two populations F_1 and F_2 based on observations of two samples $X_1, \dots, X_{n_1}$ and $Y_1, \dots, Y_{n_2}$. For example, one might be interested to find whether a new medicament administered to the treatment group is more effective than placebo given to a control group. The most widely used test in such situations is Student's t-test [28]. In case $X_1, \dots, X_{n_1}$ and $Y_1, \dots, Y_{n_2}$ are independent and identically distributed (i.i.d.) from $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$, respectively, Student's t-test is optimal in the sense that it is the size α likelihood ratio test for the hypotheses $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 > \mu_2$ [3, Chapter 9].

However, the assumption that the observed data is exactly normally distributed is rarely achievable in practical applications. The data collected can be sampled from skewed distributions, from distributions with heavy tails, or it can contain one or several *outliers* (atypical observations deviating from the most of the data). The presence of outliers or heavy tails inflates the standard error of the mean thus decreasing the power of the Student's t-test. When distributions differ in skewness, the Student's t-test is not even asymptotically correct [7]. Bernard L. Welch [33] proposed a modification to Student's t-test based on approximate degrees of freedom (ADF) for normal data with variance heterogeneity. In case of nonnormality the problems persist for the so-called Welch's test.

Let $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{in_i})$, $i = 1, \dots, k$ be independent samples from k populations $F_1, \dots, F_k$. For the k-sample case, the classical method to compare the means is the analysis of variance (ANOVA) F-test. It is based on the normality assumption and the equality of variances across the k groups. It is well known that ANOVA F-test cannot handle violations of these assumptions. B. L. Welch [34] proposed a version of F-test for the heterogeneous variances case, however, it is not robust to departures from normality and outliers, especially when the skewness differs among the groups.

Empirical likelihood (EL) was introduced by Art B. Owen in 1988 [20]. EL is a nonparametric method that does not require assumptions about the distribution family of the data. A. B. Owen [20] showed that the EL ratio statistic for an estimator $\theta(F)$ expressed as a function of an unknown distribution F has a limiting chi-square distribution. Analogically to the parametric likelihood case, EL allows estimating parameters, constructing confidence intervals, and hypothesis tests. For the one-sample case, a general framework based on smooth estimating equations was provided by Jin Qin and Jerry Lawless in 1994 [24]. Regarding the two-sample problem, Yongsong Qin and Lincheng Zhao [25] extended the EL method for the difference of two univariate parameters in 2000. The properties of EL for some two-sample problems were analyzed empirically by Jānis Valeinis et al. [30] and a complementary R program [26] package *EL* [6] was developed. EL method, in ANOVA-like setting for comparing means of k independent groups, was demonstrated by A. B. Owen in 1991 [21]. A general overview on EL methods can be found in [22].

EL is defined by constructing a multinomial distribution on the observed data points. The presence of outliers can greatly lengthen the EL confidence intervals for the mean in direction of their placement in the sample, and therefore the resulting coverage probabilities of the interval estimates might be incorrect [10]. A. B. Owen [22] discussed two approaches towards a more robust empirical likelihood: first, using estimators $\theta(F)$ that are *more robust than the mean*, and second, to *construct a more robust likelihood function*.

The centre of interest of this thesis is the first of the propositions, namely, to study the EL method for some *robust estimators*. In particular, we are interested in *robust estimators* of *location* or *centre* of the data. Note that *robustness* signifies "insensitivity to small deviations from the assumptions" [15, p. 2]. The main concern of robust statistics is the *distributional robustness*, i.e., the behaviour of the methods when the true underlying distribution deviates slightly from the assumed (usually normal) model.

The discipline of robust statistics developed in 1960s with the work of John W. Tukey and Peter J. Huber. In 1964 P. J. Huber published the seminal paper "Robust Estimation of a Location Parameter" [14], inventing a class of *M-estimators* that in a sense is a generalization of the maximum likelihood (MLE) estimators. Consider an estimator $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ of the parameter θ. P. J. Huber proposed to define $\hat{\theta}$ using a general ρ- or ψ-function:

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{n} \rho(X_i, \theta) \text{ or } \sum_{i=1}^{n} \psi(X_i, \theta) = 0,$$

where the second formulation can be used if ρ is differentiable in θ with $\psi = (\partial/\partial\theta)\rho(x, \theta)$.

P. J. Huber acquired the conditions for the consistency and asymptotic normality of the M-estimators and showed that there existed an 'optimal' M-estimator in the neighbourhood of the normal distribution. He considered the class of contaminated distributions $P_\epsilon = (1-\epsilon)\Phi + \epsilon H$, where Φ is the standard normal cumulative distribution function (cdf) and H is a cdf of any symmetric distribution. Then the *Huber estimator*, defined by

$$\psi(x) = \max(-c, \min(c, x))$$

for a given $c > 0$, has the minimax asymptotic variance among all translation-equivariant location estimators. Frank Hampel et al. [12] provided a smoothing principle for a ψ-function of a general M-estimator and showed that the smoothed estimators have smaller MSE than their non-smoothed counterparts in small and moderate sample settings.

Another well-known robust location estimator is the trimmed mean that is obtained by calculating the arithmetic mean after removing a fixed proportion of the most extreme observations from the sample. Karen K. Yuen [37] provided a robust test for comparing two population trimmed means, based on a t type statistic.

For the robust location estimators discussed above, empirical likelihood methods for the one-sample case exist. EL method can be applied to certain M-estimators, including the Huber estimator [20]. Regarding the trimmed mean, a key assumption for the classical EL approach is the independence of the observations, however, the trimmed sample consists of dependent observations. As a solution, Gengsheng Qin and Min Tsao [23] defined the EL ratio directly for the trimmed sample and proved that the limiting distribution was a scaled chi-square. They demonstrated that the EL confidence interval for the trimmed mean is more accurate than the confidence interval based on the normal approximation in a skewed distribution simulation setting.

Aims of the research

The goal of the thesis is to develop EL-based methods for comparing two or more independent populations using some well-established robust location parameter estimators. Given the good robustness properties of the trimmed mean and the Huber estimator in the one-sample case, they are good candidates for establishing robust methods also in the two-sample and ANOVA case. The aims of the research are as follows:

1. Develop an empirical likelihood method for the difference of two location M-estimators using the results of Y. Qin and L. Zhao [25]. In particular, consider the smoothed Huber estimator [12].

2. Develop an empirical likelihood method for the difference of two population trimmed means by extending the results of G. Qin and M. Tsao [23], using the approach of Y. Qin and L. Zhao [25].

3. Develop an ANOVA-like empirical likelihood method to compare the trimmed means of multiple populations, extending the results of G. Qin and M. Tsao [23], and A. B. Owen [21].

4. Develop a simulation study comparing the performance of the newly-established empirical likelihood methods for robust location parameter estimators with some widely used classical and robust methods.

5. Study the applications of the newly-developed methods on real data sets comparing with some classical and robust alternatives.

The thesis is organized as follows. The first two chapters treat the preliminaries. In Chapter 1, the theory of the EL method is presented. The maximum likelihood is shortly revisited, the EL function is introduced, and details on the EL estimation for the one- and two-sample cases with smooth unbiased estimating equations are presented. In Chapter 2, theory regarding robust location estimation is presented. M-estimators, smoothed M-estimators, and trimmed means are defined and their properties are described.

The original theoretical results of the author are presented in Chapters 3, 4 and 5. In Chapter 3, we present the EL method for the difference of two smoothed M-estimators. We give the conditions under which the EL ratio can be constructed for a difference of general M-estimators, and show that the Huber estimator fits in this setting. In Chapter 4, the empirical likelihood method for the difference of two trimmed means is presented. In Chapter 5, the EL ANOVA-like test for comparing more than two population trimmed means is presented.

In Chapter 6, the simulation study and data analysis results are presented. In particular, the empirical level and the power of the tests when sampling from various distributions is explored. The newly-developed EL methods are compared with some well-known classical and robust methods. Finally, the conclusions and the theses of the doctoral research are given.

Approbation of the results and contribution of the author

The doctoral thesis research has been presented in twelve scientific conferences (see Appendix *Conferences*): eleven international conferences, **C1-C10, C12**, and one national conference in Latvia, **C11**. The original results have been published in three Scopus/Web of Science indexed scientific papers (see Appendix *Author's publications*). Māra Delesa-Vēliņa proved the asymptotic results, performed the simulation study and data analysis, and contributed to the writing and editing of the papers.

Chapter 1. Empirical likelihood method

1.1 Maximum likelihood method

Definition 1.1.1. [5, p. 315] Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample of independent and identically distributed random variables (i.i.d.) from a population with probability density function (pdf) or probability mass function (pmf) $f(x|\theta_1, \ldots, \theta_k)$, $\theta \in \Theta \subset \mathbb{R}^k$. Let $\mathbf{x} = (x_1, \ldots, x_n)$ be the observed sample values. The *likelihood function* is a function of θ

$$L(\theta|\mathbf{x}) = L(\theta_1, \ldots, \theta_k | x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i|\theta_1, \ldots, \theta_k). \tag{1.1}$$

Definition 1.1.2. [5, p. 316] For each observed sample $\mathbf{x}$, let $\hat{\theta}(\mathbf{x})$ be a parameter value at which $L(\theta|\mathbf{x})$ attains its maximum as a function of θ, with $\mathbf{x}$ held fixed. An *MLE estimator* of the parameter θ based on a sample $\mathbf{X}$ is $\hat{\theta}(\mathbf{X})$.

Under certain regularity conditions [5, p. 516] on $f(x|\theta)$, the MLE estimators are functionally invariant, consistent, and asymptotically normal and efficient.

Definition 1.1.3. [5, p. 375] The likelihood ratio test statistic for testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_0^c$ is

$$\lambda(\mathbf{x}) = \frac{\sup_{\Theta_0} L(\theta|\mathbf{x})}{\sup_{\Theta} L(\theta|\mathbf{x})}. \tag{1.2}$$

A *likelihood ratio test* (LRT) is any test that has a rejection region R of the form $R = \{\mathbf{x} : \lambda(\mathbf{x}) \leq c\}$, where $0 \leq c \leq 1$.

Theorem 1.1.1. *(Wilks' theorem) [5, Theorem 10.3.3] Let $X_1, \ldots, X_n$ be a random sample from a pdf or pmf $f(x|\theta)$. Under certain regularity conditions on $f(x|\theta)$ [5, p. 516], if $\theta \in \Theta_0$ and $n \to \infty$, then*

$$-2 \log \lambda(\mathbf{X}) \xrightarrow{d} \chi^2_{q-p},$$

where the degrees of freedom of the chi-square distribution are determined by the number of free parameters q specified by $\theta \in \Theta$ and the number of free parameters p specified by $\theta \in \Theta_0$, where $p < q$.

There is a general equivalence between the hypothesis testing and the interval estimation that allows to construct interval estimates by *test inversion*.

Theorem 1.1.2. *[5, Theorem 9.2.2] For each $\theta_0 \in \Theta$, let $A(\theta_0)$ be the acceptance region of a level α test of $H_0 : \theta = \theta_0$. For each $\mathbf{x} \in \mathcal{X}$, define a set $C(\mathbf{x})$ in the parameter space by*

$$C(\mathbf{x}) = \{\theta_0 | \mathbf{x} \in A(\theta_0)\}.$$

Then the random set $C(\mathbf{X})$ is a $1 - \alpha$ confidence set. Conversely, let $C(\mathbf{X})$ be a $1 - \alpha$ confidence set. For any $\theta_0 \in \Theta$, define

$$A(\theta_0) = \{\mathbf{x} | \theta_0 \in C(\mathbf{x})\}.$$

Then $A(\theta_0)$ is the acceptance region of a level α test of $H_0 : \theta = \theta_0$.

1.2 Empirical likelihood method

A detailed exposure of empirical likelihood function can be found in [20]. For a probability distribution function F, denote $F(x-) = P(X < x)$ and so $P(X = x) = F(x) - F(x-)$.

Definition 1.2.1. [22, p. 6] Let $X_1, \ldots, X_n$ be i.i.d. random variables with a common unknown distribution F. The *empirical likelihood* $L(F)$ *of the cumulative distribution function* F is given by

$$L(F) = \prod_{i=1}^{n} (F(X_i) - F(X_i-)) = \prod_{i=1}^{n} p_i, \tag{1.3}$$

where $p_i = P(X = X_i)$ and $\sum_{i=1}^{n} p_i = 1$.

Theorem 1.2.1. *[22, Theorem 2.1.] Let $X_1, \ldots, X_n \in \mathbb{R}$ be i.i.d. random variables with a common cdf F_0. Let $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{X_i \leq x}$ for $-\infty < x < \infty$ be their ecdf and let F be any distribution function. If $F \neq F_n$, then $L(F) < L(F_n)$.*

Definition 1.2.2. [22, p. 10] The *empirical (or nonparametric) likelihood ratio for a distribution* F is given by

$$R(F) = \frac{L(F)}{L(F_n)} = \prod_{i=1}^{n} np_i.$$

Suppose we are interested in some parameter θ expressed as a real-valued functional T on distributions, i.e., $\theta = T(F)$, $F \in \mathcal{F}$ where $\mathcal{F}$ is a set of distributions.

Definition 1.2.3. [22, p. 11] The *profile empirical likelihood ratio function* is given by

$$\mathcal{R}(\theta) = \sup\{R(F) | T(F) = \theta, F \in \mathcal{F}\}. \tag{1.4}$$

Empirical likelihood hypothesis test rejects $H_0 : T(F_0) = \theta_0$ if $\mathcal{R}(\theta_0) < r_0$ for some threshold r_0. The empirical likelihood confidence region for the true unknown parameter $\theta_0 = T(F_0)$ is in the form $\{\theta | \mathcal{R}(\theta) \geq r_0\}$.

Example 1.2.1. Consider a hypothesis test about the population mean $\mu^* = E_F X_i$:

$$H_0 : \mu = \mu^*, \; H_1 : \mu \neq \mu^*.$$

In the functional form, $\mu^* = \int x dF(x)$, $F \in \mathcal{F}$. $\mathcal{F}$ is a class of multinomial distributions placing nonnegative weights on the observations X_i. Thus for a fixed μ^* we optimize $F = (p_1, \ldots, p_n)$, where $p_i \geq 0$, and $\sum_{i=1}^{n} p_i = 1$. The functional form under F becomes $\sum_{i=1}^{n} p_i X_i = \mu^*$, and the profile empirical likelihood function is given by

$$\mathcal{R}(\mu) = \sup_p \left\{ \prod_{i=1}^{n} np_i \,|\, \sum_{i=1}^{n} p_i X_i = \mu^*, p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \right\}. \tag{1.5}$$

Theorem 1.2.2. *[22, Theorem 2.2.] Let $X_1, \ldots, X_n$ be i.i.d. random variables with common distribution function F_0. Let $\mu_0 = E_{F_0} X_i$, and suppose that $0 < \mathrm{Var} X_i < \infty$. Then $-2 \log R(\mu_0) \xrightarrow{d} \chi_1^2$ as $n \to \infty$.*

A $1-\alpha$ confidence interval for the mean is given by

$$C_\alpha = \{\mu \in \mathbb{R} | -2\log \mathcal{R}(\mu) \leq \chi^2_{1,1-\alpha}\},$$

where $\chi^2_{1,1-\alpha}$ denotes the $1-\alpha$ quantile of χ^2_1 distribution. The interval C_α is an asymptotic coverage interval, i.e.,

$$P(\mu_0 \in C_\alpha) \to (1-\alpha) \text{ as } n \to \infty.$$

1.3 Empirical likelihood method in the two-sample case

Empirical likelihood method for various differences of univariate parameters of two populations was introduced by Y. Qin and L. Zhao [25]. Consider the two-sample problem, where $X_1, \ldots, X_{n_1}$ are i.i.d. random variables with unknown distribution F_1, and $Y_1, \ldots, Y_{n_2}$ are i.i.d. random variables with unknown distribution F_2. Let θ_0 and θ_1 be univariate parameters associated with the distributions F_1 and F_2, respectively. We are interested in the difference of the parameters, $\Delta_0 = \theta_1 - \theta_0$. Assume that the information about F_1, F_2, θ_0 and θ_1 is given by two estimating functions $w_1(X, \theta_0, \Delta_0)$ and $w_2(Y, \theta_0, \Delta_0)$ satisfying

$$E_{F_1} w_1(X, \theta_0, \Delta_0) = 0, \; E_{F_2} w_2(Y, \theta_0, \Delta_0) = 0, \tag{1.6}$$

where Δ_0 is the true parameter of interest and θ_0 is considered a nuisance parameter.

Example 1.3.1. The difference of means. Denote $\theta_0 = \int x dF_1(x)$, $\theta_1 = \int y dF_2(y)$ and $\Delta_0 = \int y dF_2(y) - \int x dF_1(x)$. The estimating functions have the following form

$$w_1(X, \theta_0, \Delta_0) = X - \theta_0, \quad w_2(Y, \theta_0, \Delta_0) = Y - \theta_0 - \Delta_0.$$

Example 1.3.2. The difference of distribution functions. For a given t_0, $0 < t_0 < 1$, consider $\theta_0 = F_1(t_0)$, $\theta_1 = F_2(t_0)$, and $\Delta_0 = F_2(t_0) - F_1(t_0)$. Then

$$w_1(X, \theta_0, \Delta_0) = I_{X \leq t_0} - \theta_0, \quad w_2(Y, \theta_0, \Delta_0) = I_{Y \leq t_0} - \theta_0 - \Delta_0.$$

In the two-sample case, the empirical likelihood function is defined as

$$L(F_1, F_2) = \prod_{i=1}^{n_1} (F_1(X_i) - F_1(X_i-)) \prod_{j=1}^{n_2} (F_2(Y_j) - F_2(Y_j-)) = \prod_{i=1}^{n_1} p_i \prod_{j=1}^{n_2} q_j, \tag{1.7}$$

where $p_i = P(X = X_i)$ and $q_j = P(Y = Y_i)$. $L(F_1, F_2)$ has maximum value $n_1^{-n_1} n_2^{-n_2}$, i.e., it is maximized when F_1 and F_2 are the respective empirical cumulative distribution functions F_{n_1} and F_{n_2}. Thus the profile empirical likelihood ratio is in the form

$$\mathcal{R}(\Delta, \theta) = \sup_{p,q} \left\{ \prod_{i=1}^{n_1} n_1 p_i \prod_{j=1}^{n_2} n_2 q_j \,|\, \sum_{i=1}^{n_1} p_i w_i(X_i, \theta, \Delta) = 0, \sum_{j=1}^{n_2} q_j w_2(Y_j, \theta, \Delta) = 0 \right\}, \tag{1.8}$$

where $p_i \geq 0$, $\sum_{i=1}^{n_1} p_i = 1$, $q_j \geq 0$ and $\sum_{i=1}^{n_2} q_j = 1$. To solve for p_i, q_j in (1.8) for a fixed Δ and θ , method of Lagrange multipliers can be used (see [25] for details), and we have

$$p_i = \frac{1}{n_1(1+\lambda_1(\Delta,\theta)w_1(X_i,\theta,\Delta))},\ i=1,\dots,n_1, \tag{1.9}$$

$$q_j = \frac{1}{n_2(1+\lambda_2(\Delta,\theta)w_2(Y_j,\theta,\Delta))},\ j=1,\dots,n_2. \tag{1.10}$$

The Lagrange multipliers $\lambda_1 = \lambda_1(\Delta,\theta)$ and $\lambda_2 = \lambda_2(\Delta,\theta)$ can be determined by solving

$$\sum_{i=1}^{n_1} \frac{w_1(X_i,\theta,\Delta)}{1+\lambda_1(\Delta,\theta)w_1(X_i,\theta,\Delta)} = 0,\ \sum_{j=1}^{n_2} \frac{w_2(Y_j,\theta,\Delta)}{1+\lambda_2(\Delta,\theta)w_2(Y_j,\theta,\Delta)} = 0. \tag{1.11}$$

Inserting p_i and q_j from (1.9) - (1.10) in (1.8) and taking the logarithm, we obtain the empirical log likelihood ratio function as

$$\log \mathcal{R}(\Delta,\theta) = -\sum_{i=1}^{n_1} \log(1+\lambda_1(\Delta,\theta)w_1(X_i,\theta,\Delta)) - \sum_{j=1}^{n_2} \log(1+\lambda_2(\Delta,\theta)w_2(Y_j,\theta,\Delta)). \tag{1.12}$$

To solve for $\hat{\theta}(\Delta)$ that maximizes $\mathcal{R}(\Delta,\theta)$, set $(\partial/\partial\theta)\{\log \mathcal{R}(\Delta,\theta)\} = 0$, and obtain

$$\sum_{i=1}^{n_1} \frac{\lambda_1(\Delta,\theta)\alpha_1(X_i,\theta,\Delta)}{1+\lambda_1(\Delta,\theta)w_1(X_i,\theta,\Delta)} + \sum_{j=1}^{n_2} \frac{\lambda_2(\Delta,\theta)\alpha_2(Y_j,\theta,\Delta)}{1+\lambda_2(\Delta,\theta)w_2(Y_j,\theta,\Delta)} = 0, \tag{1.13}$$

where $\alpha_1 = \partial w_1/\partial\theta$ and $\alpha_2 = \partial w_2/\partial\theta$.

Assumption 1.3.1. [25, p. 26]

(C1) $\theta_0 \in \Omega$, and Ω is an open interval.

(C2) $E_{F_1} w_1^2(X,\theta,\Delta) > 0$ and $E_{F_2} w_2^2(Y,\theta,\Delta) > 0$, $\alpha_1(X,\theta,\Delta)$ and $\alpha_2(Y,\theta,\Delta)$ are continuous in the neighborhood of θ_0, $\alpha_1(X,\theta,\Delta)$ and $w_1^3(X,\theta,\Delta)$ are bounded by some integrable function $G_1(X)$ in this neighborhood, $\alpha_2(Y,\theta,\Delta)$ and $w_2^3(Y,\theta,\Delta)$ are bounded by some integrable function $G_2(Y)$ in this neighborhood, and $E_{F_1}\alpha_1(X,\theta,\Delta)$ and $E_{F_2}\alpha_2(Y,\theta,\Delta)$ are non-zero.

(C3) $n_2/n_1 \to k$ (as $n_1, n_2 \to \infty$) and $0 < k < \infty$.

Theorem 1.3.1. *[25, Theroem 1] Under Assumption 1.3.1, there exists a root $\hat{\theta}(\Delta)$ of* (1.13) *such that $\hat{\theta}(\Delta)$ is a consistent estimate of θ_0, $\mathcal{R}(\Delta,\theta)$ attains its maximum at $\hat{\theta}(\Delta)$, and*

$$-2\log \mathcal{R}(\Delta_0, \hat{\theta}(\Delta_0)) \xrightarrow{d} \chi_1^2 \ \text{as } n_1, n_2 \to \infty.$$

The proof can be found in [25]. The confidence intervals for the true parameter Δ_0 can be obtained by test inversion and have the following form $\{\Delta \,|\, \mathcal{R}(\Delta, \hat{\theta}(\Delta)) > c\}$, where the constant c can be calibrated using Theorem 1.3.1.

Chapter 2. Robust estimation of a location parameter

In this chapter we consider the *model of location* and the related *estimators of location*. We define the M-estimator of location (in particular, the Huber estimator) and the trimmed mean, and we give the properties there-of. A detailed exposition of robust estimation of location can be found in [18].

2.1 M-estimators of location

Definition 2.1.1. [18, p. 17] Let $X_1, \dots X_n$ be i.i.d. random variables with distribution function F that depend on an unknown parameter θ through the model

$$X_i = \theta + u_i,\ i = 1, \dots, n, \tag{2.1}$$

where the errors u_i are i.i.d and have the distribution function F_0, and $F_0(u) = 1 - F_0(-u)$. The model (2.1) is called *the location model*, and θ is referred to as the *location parameter*.

Definition 2.1.2. [18, p. 25] Consider the location model (2.1). Given a function ρ, an *M-estimator of location parameter* θ is defined as

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{n} \rho(X_i - \theta). \tag{2.2}$$

If ρ is differentiable in θ, and $\psi(x, \theta) = (\partial/\partial\theta)\rho(x, \theta)$, then $\hat{\theta}$ is the solution to

$$\sum_{i=1}^{n} \psi(X_i - \theta) = 0. \tag{2.3}$$

Example 2.1.1. *MLE of the location parameter.* Note that choosing $\rho(x, \theta) = -\log f_\theta(x - \theta)$ and $\psi(x, \theta) = -(\partial/\partial\theta)\log f_\theta(x - \theta)$ in (2.2) and (2.3), respectively, we obtain an MLE of a location parameter θ of f_θ. If F_θ is $N(0, 1)$, apart from a constant, $\rho(x, \theta) = (x - \theta)^2/2$ and $\psi(x, \theta) = x - \theta$, and we obtain θ is the mean. If F_θ is the exponential distribution with the density $f_\theta(x) = 1/2\exp(-|x|)$, we have $\rho(x, \theta) = |x - \theta|$, and we obtain the median.

Example 2.1.2. *Huber estimator.* For a given positive constant k, the Huber estimator is defined by (2.2) or (2.3) with

$$\rho = \rho_k(x) = \begin{cases} 2kx - k^2, & x > k \\ x^2, & -k \le x \le k \\ -2kx - k^2, & x < -k \end{cases} \tag{2.4}$$

with derivative $2\psi_k(x)$, where

$$\psi = \psi_k(x) = \begin{cases} k, & x > k \\ x, & -k \le x \le k \\ -k, & x < -k. \end{cases} \tag{2.5}$$

The Huber estimator is the maximum likelihood estimator for the Huber's least favorable distribution given by the density

$$f_k(x) = \begin{cases} (1-\epsilon)\phi(k)\exp(-k(x-k)), & x > k \\ (1-\epsilon)\phi(x), & -k \le x \le k \\ (1-\epsilon)\phi(k)\exp(k(x+k)), & x < -k, \end{cases} \tag{2.6}$$

where k and ϵ are related through the formula

$$2\phi(k)/k - 2\Phi(-k) = \epsilon/(1-\epsilon), \tag{2.7}$$

where ϕ and Φ denote the pdf and cdf of standard normal distribution.

For $k \to 0$ one obtains the sample median, while $k \to \infty$ leads to the sample mean as the limiting cases. P. J. Huber [14] proved that this estimator has minimax asymptotic variance among the class of contaminated distributions $P_\epsilon = (1-\epsilon)\Phi + \epsilon H$, where Φ is the standard normal cdf and H is a cdf of a symmetric distribution. It was advocated in [14] that Huber estimator is not too sensitive to the choice of k, and that any value of k between 1 and 2 yields satisfactory results for all contamination rates $\epsilon < 0.2$. A common proposal is to take $k = 1.35$, which corresponds to a 95% efficiency of the Huber estimator compared to the sample mean at the standard normal distribution [18, 22].

Any location M-estimator $\hat{\theta}$ given by (2.2) or (2.3) is shift equivariant [18], however, it is not necessarily scale equivariant. A lack of scale equivariance can create problems, since the estimator value may be heavily dependent on the measurement units.

Definition 2.1.3. *A scale equivariant M-estimator* for the location parameter θ with a previous estimation of dispersion is defined as the solution to the equation

$$\sum_{i=1}^{n} \psi\left(\frac{X_i - \theta}{\hat{\sigma}}\right) = 0, \tag{2.8}$$

where $\hat{\sigma}$ is a previously computed dispersion estimator.

Intuitively, the dispersion estimator $\hat{\sigma}$ in (2.8) should be robust itself. A popular robust choice for $\hat{\sigma}$ is the *normalized median absolute deviation about the median (MADN)*.

Definition 2.1.4. [18, p. 36] The *median absolute deviation about the median (MAD)* is defined by

$$\mathrm{MAD}(X) = \mathrm{MAD}(X_1, \ldots, X_n) = \mathrm{Med}\{|X - \mathrm{Med}(X)|\}, \tag{2.9}$$

where Med denotes the sample median. *The normalized MAD (MADN)* is defined as

$$\mathrm{MADN}(X) = \mathrm{MAD}(X)/0.6745,$$

where the choice of the constant 0.6745 is motivated by the fact that at the standard normal distribution MAD is equal to 0.6745, thus the MADN is equal to the standard deviation.

For each n let $\hat{\sigma}_n$ be a dispersion estimator and denote $\hat{\theta}_n$ the solution (assumed unique) of

$$\sum_{i=1}^{n} \psi\left(\frac{X_i - \theta}{\hat{\sigma}_n}\right) = 0.$$

Assumption 2.1.1. (Consistency of an M-estimator of location with a preliminary dispersion estimate) [18, p. 385]

(A1) ψ is monotone and bounded with a bounded derivative.

(A2) There exists σ such that $\hat{\sigma}_n \xrightarrow{p} \sigma$.

(A3) The equation $\mathrm{E}(\psi(X_i - \theta)/\sigma) = 0$ has a unique solution θ_0.

Theorem 2.1.1. *[18, Theorem 10.12] If Assumption 2.1.1 holds, then*

$$\hat{\theta}_n \xrightarrow{p} \theta_0.$$

Define $u_i = X_i - \theta_0$ and

$$a = \mathrm{E}\psi^2\left(\frac{u_i}{\sigma}\right),\ b = \mathrm{E}\psi'\left(\frac{u_i}{\sigma}\right),\ c = \mathrm{E}\psi\left(\frac{u_i}{\sigma}\right)\psi'\left(\frac{u_i}{\sigma}\right). \tag{2.10}$$

Assumption 2.1.2. (Asymptotic normality of M-estimators with preliminary scale) [18, p. 385]

(A1) Quantities defined in (2.10) exist and $b \neq 0$.

(A2) $\sqrt{n}(\hat{\sigma}_n - \sigma)$ converges to some distribution.

(A3) $c = 0$.

Theorem 2.1.2. *[18, Theorem 10.13] Under Assumption 2.1.2,*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \nu) \text{ with } \nu = \sigma^2 \frac{a}{b^2}.$$

Definition 2.1.5. [12, p. 325] Consider i.i.d. random variables $X_1, \ldots, X_n$ from a distribution $F_{\theta,\sigma}$ with uni-modal symmetric density

$$f_{\theta,\sigma}(x) = \frac{1}{\sigma} f\left(\frac{x - \theta}{\sigma}\right),$$

and consider a score function $\tilde{\psi}(x)$ of a general ψ-function of an M-estimator

$$\tilde{\psi}(x) = \int \psi(x + u)\, dQ_n(u), \tag{2.11}$$

where Q_n is the distribution of the initial non-smooth M-estimator based on n i.i.d observations from an assumed underlying distribution. Then the *smoothed M-estimator* of the location parameter θ is defined as a solution t of

$$\sum_{i=1}^{n} \tilde{\psi}\left(\frac{X_i - t}{\sigma}\right) = 0. \tag{2.12}$$

Remark 2.1.1. *Due to the asymptotic normality of M-estimators in Theorem 2.1.2, Q_n can be approximated by $N(0, V/n)$, where V is the asymptotic variance of the initial non-smooth M-estimator. For the maximum likelihood estimators, Q_n may be chosen as the corresponding distribution under which the maximum likelihood estimator is derived. For Huber estimator, it is Huber's least favourable distribution f_k from* (2.6)*.*

Proposition 2.1.1. *[12, p. 326] Taking the f_k from* (2.6) *as the density of Q_n in* (2.11)*, the $\tilde{\psi}$-function defining the smoothed Huber estimator can be expressed in the explicit form as*

$$\begin{aligned}\tilde{\psi}_k(x) &= k\Phi\left(\frac{x-k}{\sigma_n}\right) - k\left(1 - \Phi\left(\frac{x+k}{\sigma_n}\right)\right)\\ &+ x\left(\Phi\left(\frac{x+k}{\sigma_n}\right) - \Phi\left(\frac{x-k}{\sigma_n}\right)\right) + \sigma_n\left(\phi\left(\frac{x+k}{\sigma_n}\right) - \phi\left(\frac{x-k}{\sigma_n}\right)\right), \end{aligned} \tag{2.13}$$

where $\sigma_n = \sqrt{V/n}$ and k is the tuning constant defining the non-smoothed Huber estimator (2.5)*.*

2.2 Trimmed mean

Definition 2.2.1. [23, p. 2199] Let $X_1, X_2, \ldots, X_n$ be i.i.d. random sample from population F_0 and let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be ordered statistics. The trimmed mean is defined as

$$\bar{X}_{\alpha\beta} = \frac{1}{m}\sum_{i=r}^{s} X_{(i)}, \tag{2.14}$$

where $0 \leq \alpha < 1/2$, $0 \leq \beta < 1/2$ are trimming proportions from the left and the right side, respectively, $r = \lfloor n\alpha \rfloor + 1$, $s = n - \lfloor n\beta \rfloor$, and $m = n - \lfloor n\alpha \rfloor - \lfloor n\beta \rfloor$.

The result for the asymptotic distribution of the trimmed mean was provided by S. Stigler in [27]. Let

$$A = F_0^{-1}(\alpha) - F_0^{-1}(\alpha-) \text{ and } B = F_0^{-1}(1-\beta) - F_0^{-1}((1-\beta)-) \tag{2.15}$$

represent the jumps of F_0^{-1} at the trimming proportions. For any $0 < p < 1$ denote $\xi_p := F_0^{-1}(p)$ and introduce a distribution function $H(x)$ obtained by truncating F_0 as follows:

$$H(x) = \begin{cases} 0, & x < \xi_\alpha \\ \frac{F_0(x)-\alpha}{1-\alpha-\beta}, & \xi_\alpha \leq x \leq \xi_{1-\beta} \\ 1, & x > \xi_{1-\beta}. \end{cases} \tag{2.16}$$

Let $\mu_{\alpha\beta}$ and $\sigma^2_{\alpha\beta}$ denote the mean and the variance of the distribution H, respectively.

Theorem 2.2.1. *[27, p. 473] Let $0 < \alpha < 1 - \beta < 1$ and $n \to \infty$. Then*

$$\sqrt{n}(\bar{X}_{\alpha\beta} - \mu_{\alpha\beta}) \xrightarrow{d} W, \text{ where}$$

$$W = \frac{1}{1-\alpha-\beta}\left[Z + (\xi_\alpha - \mu_{\alpha\beta})Z_1 + (\xi_{1-\beta} - \mu_{\alpha\beta})Z_2 - A\max(0, Z_1) + B\max(0, Z_2)\right],$$

A and B are the quantities defined in (2.15)*, the random variable Z is* $N(0,(1-\alpha-\beta)\sigma^2_{\alpha\beta})$*, Z is independent from the random vector* (Z_1, Z_2)*, and* (Z_1, Z_2) *is* $N(0, C)$*, where*

$$C = \begin{pmatrix} \alpha(1-\alpha) & -\alpha\beta \\ -\alpha\beta & \beta(1-\beta) \end{pmatrix}.$$

For the proof of the Theorem 2.2.1, see [27] or [1].

Remark 2.2.1. *If* $A = 0$ *and* $B = 0$ *in Theorem 2.2.1 (in other words, the trimming is done at uniquely defined percentiles of distribution* F_0*), the asymptotic distribution W of the trimmed mean has a simpler form. In such case,* $EW = 0$ *and*

$$\mathrm{Var}W = \frac{1}{(1-\alpha-\beta)^2}\Big(\sigma^2_{\alpha\beta} + \alpha(1-\alpha)(\xi_\alpha - \mu_{\alpha\beta})^2 - 2\alpha\beta(\xi_\alpha - \mu_{\alpha\beta})(\xi_{1-\beta} - \mu_{\alpha\beta}) + \beta(1-\beta)(\xi_{1-\beta} - \mu_{\alpha\beta})^2\Big) =: \tau^2_{\alpha\beta},$$

thus $\sqrt{n}(\bar{X}_{\alpha\beta} - \mu_{\alpha\beta}) \xrightarrow{d} N(0, \tau^2_{\alpha\beta})$.

Chapter 3. Empirical likelihood method for the difference of two location M-estimators

The aim of this chapter is to establish the two-sample empirical likelihood method for the difference of two M-estimators. The new results presented in this chapter have been published in M. Delesa-Vēliņa et al. [32].

A. B. Owen [20] provided assumptions under which the empirical likelihood confidence intervals can be constructed for M-estimates, and showed that these hold for the Huber estimator. However, the two-sample EL method described in Chapter 1.3 cannot be applied directly to Huber estimator, since the condition (C1) in Assumption 1.3.1 that necessitates a continuous derivative of the estimating function does not hold for Huber estimator's ψ-function (2.5). Thus the smoothing principle of the ψ-function described in Chapter 2 is used.

As it was noted in Chapter 2, the M-estimates defined by (2.2) or (2.3) are not scale-equivariant and the results may depend on the measurement units to a large extent. Thus the scale-equivariant M-estimators defined by (2.8) are preferred. However, the real value of the scale parameter σ in (2.8) is not known in practical situations. Thus the scale parameter is interpreted as an additional nuisance parameter for the EL maximization problem. We use the plug-in empirical likelihood that allows possibly infinite-dimension nuisance parameters in the estimating equations. The plug-in EL was formalized for the one-sample case by Nils L. Hjort et al. [13]. J. Valeinis [29] generalized the conditions for the plug-in EL method for the two-sample case.

3.1 Main results

Consider the two-sample problem defined in Chapter 1.3: $X_1, \ldots, X_{n_1}$ are i.i.d. random variables with unknown distribution F_1, and $Y_1, \ldots, Y_{n_2}$ are i.i.d. random variables with unknown distribution F_2, and we are interested in the difference of two M-estimators θ_0 and θ_1 of the samples X and Y, respectively. The estimating functions in (1.6) - (1.6) have the form

$$\begin{aligned} w_1(X, \Delta_0, \theta_0, \sigma_1^0, \sigma_2^0) &= \psi\left(\frac{X-\theta_0}{\sigma_1^0}\right), \\ w_2(Y, \Delta_0, \theta_0, \sigma_1^0, \sigma_2^0) &= \psi\left(\frac{Y-\Delta_0-\theta_0}{\sigma_2^0}\right), \end{aligned} \tag{3.1}$$

where ψ corresponds to a general ψ-function of an M-estimator defined in (2.8), σ_1 and σ_2 are the scale parameters for the samples X and Y with true values of σ_1^0 and σ_2^0, respectively, θ denotes the location parameter for the sample X with the true value θ_0.

The two-sample problem setting in Chapter 1.3 is for fixed estimating functions w_1 and w_2, but in our case (3.1) involves additional nuisance parameters θ, σ_1,

σ_2, and V_1 and V_2 indirectly. We define the profile EL function

$$\mathcal{R}(\Delta,\theta,\sigma_1,\sigma_2) = n_1^{n_1} n_2^{n_2} \sup_{p,q} \Bigg\{ \prod_{i=1}^{n_1} p_i \prod_{j=1}^{n_2} q_j \,|\, p_i \geq 0, q_j \geq 0, \sum_{i=1}^{n_1} p_i = 1,$$
$$\sum_{j=1}^{n_2} q_j = 1, \sum_{i=1}^{n_1} p_i \psi\left(\frac{X_i-\theta}{\sigma_1}\right) = 0, \sum_{j=1}^{n_2} q_j \psi\left(\frac{Y_j-\Delta-\theta}{\sigma_2}\right) = 0 \Bigg\}. \quad (3.2)$$

A unique solution to (3.2) exists, provided that 0 is both inside the convex hull of the $w_1(X_i,\Delta,\theta,\sigma_1,\sigma_2)$'s and the convex hull of the $w_2(Y_j,\Delta,\theta,\sigma_1,\sigma_2)$'s. The maximum may be found by using the standard Lagrange multipliers method, where the Lagrange multipliers now depend not only on Δ and θ, but also on the nuisance parameters σ_1 and σ_2, i.e., $\lambda_1 = \lambda_1(\Delta,\theta,\sigma_1,\sigma_2)$ and $\lambda_2 = \lambda_2(\Delta,\theta,\sigma_1,\sigma_2)$. Lagrange multipliers can be determined in therms of $\Delta(\theta)$ from the equations (1.11)-(1.11) with the estimating functions defined by (3.1), i.e., from

$$\sum_{i=1}^{n_1} \frac{\psi\left(\frac{X_i-\theta}{\sigma_1}\right)}{1+\lambda_1\psi\left(\frac{X_i-\theta}{\sigma_1}\right)} = 0, \quad \sum_{j=1}^{n_2} \frac{\psi\left(\frac{Y_j-\Delta-\theta}{\sigma_2}\right)}{1+\lambda_2\psi\left(\frac{Y_j-\Delta-\theta}{\sigma_2}\right)} = 0. \quad (3.3)$$

We define the empirical log likelihood ratio (multiplied by minus two) as

$$\mathcal{W}(\Delta,\theta,\sigma_1,\sigma_2) = -2\log\mathcal{R}(\Delta,\theta,\sigma_1,\sigma_2) =$$
$$= 2\sum_{i=1}^{n_1} \log\left(1+\lambda_1\psi\left(\frac{X_i-\theta}{\sigma_1}\right)\right) + 2\sum_{j=1}^{n_2} \log\left(1+\lambda_2\psi\left(\frac{Y_j-\Delta-\theta}{\sigma_2}\right)\right).$$

To find an estimator $\hat{\theta} = \hat{\theta}(\Delta,\sigma_1,\sigma_2)$ for θ that maximizes $\mathcal{R}(\Delta,\theta,\sigma_1,\sigma_2)$, set

$$\frac{\partial}{\partial\theta}\mathcal{W}(\Delta,\theta,\sigma_1,\sigma_2) = \sum_{i=1}^{n_1} \frac{\lambda_1\psi'\left(\frac{X_i-\theta}{\sigma_1}\right)}{1+\lambda_1\psi\left(\frac{X_i-\theta}{\sigma_1}\right)} + \sum_{j=1}^{n_2} \frac{\lambda_2\psi'\left(\frac{Y_j-\Delta-\theta}{\sigma_2}\right)}{1+\lambda_2\psi\left(\frac{Y_j-\Delta-\theta}{\sigma_2}\right)} = 0, \quad (3.4)$$

where $\psi' = (\partial/\partial\theta)\psi$.

Let $\hat{\sigma}_1$ and $\hat{\sigma}_2$ be two estimators for the scale parameters σ_1 and σ_2 respectively. We present the assumptions for a general ψ-function of M-estimator defined in (2.8):

Assumption 3.1.1.

(A1) $\theta_0 \in \Omega$ and Ω is an open interval.

(A2) $\mathrm{E}\psi^2\left((X_i-\theta)/\hat{\sigma}_1\right) > 0$, $\mathrm{E}\psi^2\left((Y_j-\theta-\Delta)/\hat{\sigma}_2\right) > 0$, $\psi'\left((X_i-\theta)/\hat{\sigma}_1\right)$, $\psi'\left((Y_j-\theta-\Delta)/\hat{\sigma}_2\right)$ are continuous in the neighborhood of θ_0, $\psi'\left((X_i-\theta)/\hat{\sigma}_1\right)$ and $\psi^3\left((X_i-\theta)/\hat{\sigma}_1\right)$ are bounded by some integrable function $G_1(X)$ in this neighborhood, $\psi'\left((Y_j-\theta-\Delta)/\hat{\sigma}_2\right)$ and

$\psi^3\left((Y_j-\theta-\Delta)/\hat{\sigma}_2\right)$ are bounded by some integrable function $G_2(Y)$ in this neighborhood, and $\mathrm{E}\psi'\left((X_i-\theta)/\hat{\sigma}_1\right)$, $\mathrm{E}\psi'\left((Y_j-\theta-\Delta)/\hat{\sigma}_2\right)$ are nonzero.

(A3) $n_2/n_1 \to k$ (as $n_1, n_2 \to \infty$) and $0 < k < \infty$.

Assumption 3.1.2.

(B1) $\hat{\sigma}_1 \xrightarrow{p} \sigma_1^0$, $\hat{\sigma}_2 \xrightarrow{p} \sigma_2^0$.

(B2) $\mathrm{E}\psi^2\left(\frac{X_i-\theta_0}{\sigma_1^0}\right) = V_1 < \infty$, $\mathrm{E}\psi^2\left(\frac{Y_j-\theta_0-\Delta_0}{\sigma_2^0}\right) = V_2 < \infty$.

(B3) $\mathrm{E}\left(\left(\frac{X_i-\theta_0}{\sigma_1^0}\right)\psi'\left(\frac{X_i-\theta_0}{\sigma_1^0}\right)\right) = 0$, $\mathrm{E}\left(\left(\frac{Y_j-\theta_0-\Delta_0}{\sigma_2^0}\right)\psi'\left(\frac{Y_j-\theta_0-\Delta_0}{\sigma_2^0}\right)\right) = 0$.

(B4) $\mathrm{E}\left(\left(\frac{X_i-\theta_0}{\sigma_1^0}\right)\psi\left(\frac{X_i-\theta_0}{\sigma_1^0}\right)\psi'\left(\frac{X_i-\theta_0}{\sigma_1^0}\right)\right) < \infty$,
$\mathrm{E}\left(\left(\frac{Y_j-\theta_0-\Delta_0}{\sigma_2^0}\right)\psi\left(\frac{Y_j-\theta_0-\Delta_0}{\sigma_2^0}\right)\psi'\left(\frac{Y_j-\theta_0-\Delta_0}{\sigma_2^0}\right)\right) < \infty$.

Assumption 3.1.3.

(C1) ${n_1}^{-1}\sum_{i=1}^{n_1}\psi'\left(\frac{X_i-\theta_0}{\hat{\sigma}_1}\right) \xrightarrow{p} M_1$,
${n_2}^{-1}\sum_{j=1}^{n_2}\psi'\left(\frac{Y_j-\Delta_0-\theta_0}{\hat{\sigma}_2}\right) \xrightarrow{p} M_2$.

(C2) $\frac{1}{\sqrt{n_1}}\sum_{i=1}^{n_1}\psi\left(\frac{X_i-\theta_0}{\hat{\sigma}_1}\right) \xrightarrow{d} U_1$, where $U_1 \sim N(0, V_1)$,
$\frac{1}{\sqrt{n_2}}\sum_{j=1}^{n_2}\psi\left(\frac{Y_j-\theta_0-\Delta_0}{\hat{\sigma}_2}\right) \xrightarrow{d} U_2$, where $U_2 \sim N(0, V_2)$.

(C3) $n_1^{-1}\sum_{i=1}^{n_1}\psi^2\left(\frac{X_i-\theta_0}{\hat{\sigma}_1}\right) \xrightarrow{p} V_1$,
${n_2}^{-1}\sum_{j=1}^{n_2}\psi^2\left(\frac{Y_j-\theta_0-\Delta_0}{\hat{\sigma}_2}\right) \xrightarrow{p} V_2$.

Remark 3.1.1. *Assumption 3.1.1 is very similar to Assumption 1.3.1, except that now the conditions need to hold for the estimating functions with the nuisance parameters $\hat{\sigma}_1$ and $\hat{\sigma}_2$. Part (A1) states that the true parameter θ_0 should be in an open interval. Part (A2) was also used in [24] and describes the smoothness conditions for the estimating functions. Part (A3) requires that the sample sizes are asymptotically comparable.*

Assumption 3.1.2 is necessary to establish the asymptotic distribution of the location M-estimator with a preliminary scale, see Assumptions 2.1.1 and 2.1.2. (B1) holds for a suitable scale estimator under mild (smoothness) conditions on the underlying distribution. (B2) holds for a bounded ψ-function. (B3) holds for F_1, F_2 symmetric and ψ odd.

Assumption 3.1.3 contains technical assumptions for the plug-in empirical likelihood, similarly as in [13] and [29]. It allows establishing the limiting distribution of the plug-in EL ratio assuming that the solution to the EL maximisation problem exists. To establish the existence of the solution, stronger assumptions would be necessary, since it would require almost sure convergence instead of convergence in probability of the nuisance parameter estimators in (B1), see J. Valeinis [29] for the details.

Lemma 3.1.1 comments on the relationship between the Assumptions 3.1.1 - 3.1.3, the main Theorem 3.1.1 that establishes the EL method for the difference of two general M-estimators, and Lemma 3.1.2 that states the conditions under which the smoothed Huber estimator fits in the setting of Theorem 3.1.1.

Lemma 3.1.1. *(M. Delesa-Vēliņa et al. [32]) For a general ψ-function of an M-estimator satisfying Assumptions 3.1.1 and 3.1.2, Assumption 3.1.3 holds.*

Theorem 3.1.1. *(M. Delesa-Vēliņa et al. [32]) Assume that the EL maximization problem has a solution $\hat{\theta}(\Delta, \hat{\sigma}_1, \hat{\sigma}_2)$ determined by* (3.4). *Then, for a general ψ-function of an M-estimator satisfying Assumptions 3.1.1 and 3.1.3, as $n_1, n_2 \to \infty$,*

$$-2 \log \mathcal{R}(\Delta_0, \hat{\theta}(\Delta_0, \hat{\sigma}_1, \hat{\sigma}_2), \hat{\sigma}_1, \hat{\sigma}_2) \xrightarrow{d} \chi_1^2.$$

Lemma 3.1.2. *(M. Delesa-Vēliņa et al. [32]) Let $\tilde{\psi}_k$ be the score function* (2.13) *defining the smoothed Huber M-estimator and let $\hat{\sigma}_1$ and $\hat{\sigma}_2$ be the mean absolute deviation (MAD) dispersion estimates* (2.9) *of samples X and Y, respectively. Assume that the underlying distributions F_1 and F_2 of X and Y are symmetric. Then Assumptions 3.1.1 and 3.1.2 hold for $\psi = \tilde{\psi}_k$.*

Remark 3.1.2. *For the proofs, see M. Delesa-Vēliņa et al. [32].*

Chapter 4. Empirical likelihood method for the difference of trimmed means

In this Chapter a new empirical likelihood method for the difference of two trimmed means is developed. The results provided in this Chapter have been previously published in M. Delesa-Vēliņa et al. [8].

In chapter 4.1, preliminary results on the EL method for the trimmed means in the one-sample case are given. A. B. Owen's EL method was established for independent observations, while the observations of the trimmed sample are dependent. As a solution G. Qin and M. Tsao [23] proposed to estimate the EL ratio directly for the trimmed sample and consequently established the impact of the dependence on the limiting distribution of the EL ratio, obtaining a scaled chi-square distribution.

In Chapter 4.2 a new EL-based inference method for the difference of two trimmed means is developed, extending the results of G. Qin and M. Tsao [23] to the two-sample case using the tools of Y. Qin and L. Zhao [25] described in Chapter 1.

4.1 Empirical likelihood method for the trimmed mean in the one-sample case

Consider the setting of Chapter 2.2: let $X_1, X_2, \ldots, X_n$ be i.i.d. with distribution function F_0, and $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be ordered statistics. Let $\bar{X}_{\alpha\beta}$ be the sample trimmed mean as defined by (2.14), i.e.,

$$\bar{X}_{\alpha\beta} = \frac{1}{m}\sum_{i=r}^{s} X_{(i)},$$

where $0 < \alpha < 1/2$, $0 < \beta < 1/2$ are trimming proportions from the left and the right side, respectively, $r = \lfloor n\alpha \rfloor + 1$, $s = n - \lfloor n\beta \rfloor$, and m be the effective sample size, $m = n - \lfloor n\alpha \rfloor - \lfloor n\beta \rfloor$.

According to Theorem 2.2.1, the asymptotic value of the sample trimmed mean $\bar{X}_{\alpha\beta}$ is

$$\mu_{\alpha\beta} = \frac{1}{1-\alpha-\beta}\int_{\xi_\alpha}^{\xi_{1-\beta}} x dF_0.$$

Let weights $p_i = 0$ for $i < r$ and $i > s$, $p_i \geq 0$ for $r \leq i \leq s$ and $\sum_{i=r}^{s} p_i = 1$. Define the profile empirical likelihood ratio for the trimmed mean as

$$\mathcal{R}(\mu_{\alpha\beta}) = \sup\left\{\prod_{i=r}^{s} mp_i : p_i \geq 0, \sum_{i=r}^{s} p_i = 1, \sum_{i=r}^{s} p_i X_{(i)} = \mu_{\alpha\beta}\right\}.$$

Theorem 4.1.1. *[23, Theorem 2.1]*

Assume F_0 is continuous, $F_0'(\xi_\alpha) > 0$ and $F_0'(\xi_{1-\beta}) > 0$, and let $\mu_{\alpha\beta}^0$ be the true value of the trimmed mean $\mu_{\alpha\beta}$. Then

$$-2a \log \mathcal{R}(\mu_{\alpha\beta}^0) \xrightarrow{d} \chi_1^2,$$

where

$$a = \sigma^2_{\alpha\beta}/((1-\alpha-\beta)\tau^2_{\alpha\beta}),$$

$$\sigma^2_{\alpha\beta} = \frac{1}{(1-\alpha-\beta)}\int_{\xi_\alpha}^{\xi_{1-\beta}} x^2 dF_0(x) - \mu^2_{\alpha\beta}, \tag{4.1}$$

and

$$\tau^2_{\alpha\beta} = \frac{1}{(1-\alpha-\beta)^2}((1-\alpha-\beta)\sigma^2_{\alpha\beta} + \beta(1-\beta)(\xi_{1-\beta}-\mu_{\alpha\beta})^2 - 2\alpha\beta(\xi_\alpha-\mu_{\alpha\beta})(\xi_{1-\beta}-\mu_{\alpha\beta}) + \alpha(1-\alpha)(\xi_\alpha-\mu_{\alpha\beta})^2). \tag{4.2}$$

[23] provided a consistent estimator for the scaling constant a by

$$\hat{a} = \hat{\sigma}^2_{\alpha\beta}/((1-\alpha-\beta)\hat{\tau}^2_{\alpha\beta}),$$

where

$$\hat{\sigma}^2_{\alpha\beta} = \frac{1}{(1-\alpha-\beta)}\int_{\hat{\xi}_\alpha}^{\hat{\xi}_{1-\beta}} x^2 dF_n(x) - \bar{X}^2_{\alpha\beta}, \tag{4.3}$$

$$\hat{\tau}^2_{\alpha\beta} = \frac{1}{(1-\alpha-\beta)^2}((1-\alpha-\beta)\hat{\sigma}^2_{\alpha\beta} + \beta(1-\beta)(\hat{\xi}_{1-\beta}-\bar{X}_{\alpha\beta})^2 - 2\alpha\beta(\hat{\xi}_\alpha-\bar{X}_{\alpha\beta})(\hat{\xi}_{1-\beta}-\bar{X}_{\alpha\beta}) + \alpha(1-\alpha)(\hat{\xi}_\alpha-\bar{X}_{\alpha\beta})^2), \tag{4.4}$$

$\hat{\xi}_p = \inf\{x : F_n(x) \geq p\}$ for any $0 < p < 1$, and $F_n(x)$ is the empirical distribution function.

4.2 Main results

Consider the two-sample EL problem described in Chapter 1.3 where i.i.d. random variables $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ have unknown distribution functions F_1 and F_2, respectively.

We are interested in the difference of two trimmed means with trimming proportions $0 < \alpha < 1/2$, $0 < \beta < 1/2$. Thus for (1.6) - (1.6) consider the parameters

$$\theta_0 = \frac{1}{1-\alpha-\beta}\int_{\xi_\alpha}^{\xi_{1-\beta}} x dF_1 =: \mu_{\alpha\beta 1}, \quad \theta_1 = \frac{1}{1-\alpha-\beta}\int_{\xi_\alpha}^{\xi_{1-\beta}} y dF_2 =: \mu_{\alpha\beta 2},$$

and

$$\Delta_0 = \mu_{\alpha\beta 2} - \mu_{\alpha\beta 1}.$$

Consider the respective sample means

$$\bar{X}_{\alpha\beta} = \frac{1}{m_1}\sum_{i=r_1}^{s_1} X_{(i)}, \quad \bar{Y}_{\alpha\beta} = \frac{1}{m_2}\sum_{j=r_2}^{s_2} Y_{(j)},$$

where $r_1 = \lfloor n_1\alpha \rfloor + 1$, $s_1 = n_1 - \lfloor n_1\beta \rfloor$, $r_2 = \lfloor n_2\alpha \rfloor + 1$, $s_2 = n_2 - \lfloor n_2\beta \rfloor$, and m_1 and m_2 are the effective sample sizes after trimming, i.e., $m_1 = n_1 - \rfloor n_1\alpha \rfloor - \lfloor n_1\beta \rfloor$,

$m_2 = n_2 - \lfloor n_2\alpha \rfloor - \lfloor n_2\beta \rfloor$. Similarly as in Theorem 4.1.1, let weights $p_i = 0$ for $i < r_1$, $i > s_1$, and $q_j = 0$ for $j < r_2$ and $j > s_2$. Define the estimating functions

$$w_1(X, \mu_{\alpha\beta 1}, \Delta_0) = X - \mu_{\alpha\beta 1}, \quad w_2(Y, \mu_{\alpha\beta 1}, \Delta_0) = Y - \Delta_0 - \mu_{\alpha\beta 1}.$$

Finally, define the profile empirical likelihood ratio function for the difference Δ of the trimmed means as

$$\mathcal{R}(\Delta, \mu_t) = \sup_{p_i, q_j} \Big\{ \prod_{i=1}^{m_1} m_1 p_i \prod_{j=1}^{m_2} m_2 q_j | p_i \geq 0, q_j \geq 0, \sum_{i=r_1}^{s_1} p_i = 1, \sum_{j=r_2}^{s_2} q_j = 1, \sum_{i=r_1}^{s_1} p_i w_1(X_{(i)}, \mu_t, \Delta) = 0, \sum_{j=r_2}^{s_2} q_j w_2(Y_{(j)}, \mu_t, \Delta) = 0 \Big\}, \quad (4.5)$$

where μ_t is considered as a nuisance parameter and has the real value $\mu_{\alpha\beta 1}$. This setting is similar to the one described in Chapter 1.3, with a distinction that additional restrictions $p_i = 0$ for $i < r_1$, $i > s_1$, and $q_j = 0$ for $j < r_2$, $j > s_2$ are added. A unique solution of (4.5) exists, provided that 0 is inside the convex hull of the points $w_1(X_{(i)}, \mu_t, \Delta)$'s and $w_2(Y_{(j)}, \mu_t, \Delta)$'s, $r_1 \leq i \leq s_1$, $r_2 \leq j \leq s_2$, and may be found using the Lagrange multipliers method. Similarly to (1.9) - (1.10) we have

$$p_i = \frac{1}{m_1(1 + \lambda_1 w_1(X_{(i)}, \mu_t, \Delta))}, \quad i = r_1, \ldots, s_1,$$

$$q_j = \frac{1}{m_2(1 + \lambda_2 w_2(Y_{(j)}, \mu_t, \Delta))}, \quad j = r_2, \ldots, s_2,$$

where the Lagrange multipliers $\lambda_1 = \lambda_1(\mu_t, \Delta)$ and $\lambda_2 = \lambda_2(\mu_t, \Delta)$ can be determined in terms of μ_t by the equations

$$\sum_{i=r_1}^{s_1} \frac{w_1(X_{(i)}, \mu_t, \Delta)}{1 + \lambda_1 w_1(X_{(i)}, \mu_t, \Delta)} = 0, \quad \sum_{j=r_2}^{s_2} \frac{w_2(Y_{(j)}, \mu_t, \Delta)}{1 + \lambda_2 w_2(Y_{(j)}, \mu_t, \Delta)} = 0.$$

The empirical likelihood profile log ratio is defined as

$$\begin{aligned} \mathcal{W}(\Delta, \mu_t) &= -2 \log \mathcal{R}(\Delta, \mu_t) \quad (4.6) \\ &= 2 \sum_{i=r_1}^{s_1} \log(1 + \lambda_1 w_1(X_{(i)}, \mu_t, \Delta)) + 2 \sum_{j=r_2}^{s_2} \log(1 + \lambda_2 w_2(Y_{(j)}, \mu_t, \Delta)). \end{aligned}$$

To find an estimator $\hat{\mu}_t = \hat{\mu}_t(\Delta)$ for the nuisance parameter μ_t that maximizes $\mathcal{R}(\Delta, \mu_t)$ for a fixed parameter Δ, set $(\partial/\partial\mu_t)\mathcal{W}(\Delta, \mu_t) = 0$. Noting that the derivatives of w_1 and w_2 with respect to μ_t are equal to -1, we obtain the empirical likelihood equation

$$\frac{\partial}{\partial \mu_t} \mathcal{W}(\Delta, \mu_t) = \sum_{i=r_1}^{s_1} \frac{-\lambda_1}{1 + \lambda_1 w_1(X_{(i)}, \mu_t, \Delta)} + \sum_{j=r_2}^{s_2} \frac{-\lambda_2}{1 + \lambda_2 w_2(Y_{(j)}, \mu_t, \Delta)} = 0. \quad (4.7)$$

Assumption 4.2.1.

(A1) F_1, F_2 is continuous, $F_1'(\xi_\alpha) > 0$, $F_1'(\xi_{1-\beta}) > 0$, $F_2'(\xi_\alpha) > 0$, $F_2'(\xi_{1-\beta}) > 0$.

(A2) $\mu_{\alpha\beta 1} \in \Omega$, where Ω is an open interval.

(A3) $n_2/n_1 \to k$ as $n_1, n_2 \to \infty$, and $0 < k < \infty$.

Remark 4.2.1. *Assumption 4.2.1 condition (A1) comes from Theorem 4.1.1 and ensures that the samples are trimmed so that the corresponding percentiles of the population distributions F_1 and F_2 are uniquely defined. Notice that it is assumed that the trimming proportions α and β are positive. To allow α or β to be equal to zero, an additional condition $E(X^2) < \infty$ and $E(Y^2) < \infty$ should be imposed, and the proof of Theorem 4.2.1 would require a slight change. Conditions (A2) and (A3) are inherited from assumptions for the EL method in the general two sample case, Assumption 1.3.1.*

Theorem 4.2.1. *(M. Delesa-Vēliņa et al. [32]) Under Assumption 4.2.1 there exists a root $\hat{\mu}_t(\Delta_0)$ of (4.7) such that $\hat{\mu}_t(\Delta_0)$ is a consistent estimator for $\mu_{\alpha\beta 1}$, $\mathcal{R}(\Delta_0, \mu_t)$ attains its local maximum value at $\hat{\mu}_t(\Delta_0)$, and*

$$-2a_2 \log \mathcal{R}(\Delta_0, \hat{\mu}_t(\Delta_0)) \xrightarrow{d} \chi_1^2$$

as $n_1, n_2 \to \infty$, with the scaling constant

$$a_2 = \frac{n_1 n_2 (m_2 \sigma_1^2 + m_1 \sigma_2^2)}{m_1 m_2 (n_2 \tau_1^2 + n_1 \tau_2^2)},$$

where ($\sigma_1^2 = \sigma_{\alpha\beta 1}^2$, $\tau_1^2 = \tau_{\alpha\beta 1}^2$) and ($\sigma_2^2 = \sigma_{\alpha\beta 2}^2$, $\tau_2^2 = \tau_{\alpha\beta 2}^2$) are the parameters defined in (4.1) and (4.2), associated with the underlying distribution functions F_1 and F_2, respectively.

Remark 4.2.2. *A consistent estimator for the scaling constant a_2 from Theorem 4.2.1 is provided by*

$$\hat{a}_2 = \frac{n_1 n_2 (m_2 \hat{\sigma}_1^2 + m_1 \hat{\sigma}_2^2)}{m_1 m_2 (n_2 \hat{\tau}_1^2 + n_1 \hat{\tau}_2^2)},$$

where the parameter estimators $\hat{\sigma}_1^2$, $\hat{\tau}_1^2$, and $\hat{\sigma}_2^2$, $\hat{\tau}_2^2$ are defined as in the one-sample case in (4.3) and (4.4) with the empirical distributions $F_{n_1}(x)$, $F_{n_2}(y)$, and the trimmed means $\bar{X}_{\alpha\beta}$, $\bar{Y}_{\alpha\beta}$, respectively.

Remark 4.2.3. *An approximate $1-p$ confidence interval for the true difference of trimmed means Δ_0 can be obtained by test inversion and has the form*

$$\{\Delta : -2\hat{a}_2 \log \mathcal{R}(\Delta, \hat{\mu}_t(\Delta)) \leq \chi^2_{1,1-p}\},$$

where $\chi^2_{1,1-p}$ denotes the $1-p$ quantile of the χ_1^2 distribution.

Proof. For the proof, see M. Delesa-Vēliņa et al. [8].

Chapter 5. Empirical likelihood-based ANOVA method for the trimmed means

The goal of this chapter is to develop an empirical likelihood based ANOVA method for comparing multiple population trimmed means. The new results described in this chapter have been previously published in Delesa-Velina et al. [31].

Consider the problem of comparing more than two populations: let $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})$, $i = 1, 2, \ldots, k$, be independent random samples from k different distributions with population means μ_i. The classical approach is to test the null hypothesis of equal population means

$$H_0 : \mu_1 = \ldots = \mu_k =: \mu. \tag{5.1}$$

Under the assumption of equal variances (homoscedasticity) and normally distributed data in each group, i.e. $Y_{ij} \sim N(\mu_i, \sigma)$, one can use the classical ANOVA F test.

It is well known that the ANOVA F-test can not handle the variance heterogeneity since the problems of controlling the probability of type I error arise. B. L. Welch [34] proposed an approximate degrees of freedom (ADF) type procedure that can deal with variance heterogeneity for normally distributed data. However, problems still arise when the variance heterogeneity appears in combination with nonnormal data and unbalanced sample designs (see, for example, [35]). K. Yuen et al. [37] suggested a robust modification to the Welch's test using trimmed means and Winsorized variances together with ADF statistics. It was demonstrated in [16] that such an approach offers a better control over the probability of a type I error for one-way ANOVA under distributions of various degree of skewness and unbalanced sample sizes. A. B. Owen [21] proposed an empirical likelihood-based ANOVA method for independent groups to test the hypothesis of equality of means. We take advantage of the good robustness properties of the trimmed means in the one-sample case and propose an EL-based ANOVA type method to test the hypothesis of equality of more than two trimmed means.

We first present the A. B. Owen's EL ANOVA method in Chapter 5.1. The main result on comparing multiple population trimmed means in the empirical likelihood setting is presented in Chapter 5.2.

5.1 Empirical likelihood-based ANOVA method

To present the EL ANOVA method, we follow [21]. Let observations $Y_{ij} \in \mathbb{R}$, where $i = 1, \ldots, k$, $j = 1, \ldots, n_i$ and $N = \sum_{i=1}^{k} n_i$ denotes the total number of observations.

Consider N random pairs (I, Y), where $I \in \{1, \ldots, k\}$ and $Y \in \mathbb{R}^d$. The observation Y_{ij} is represented by a pair where $I = i$ and $Y = Y_{ij}$. Let F be a distribution on (I, Y) pairs. The data are not i.i.d from F, because I in each pair is a non-random categorical predictor. Define the likelihood

$$L(F) = \prod_{i=1}^{n} \prod_{j=1}^{n_i} v_{ij},$$

where $v_{ij} = F\{(i, Y_{ij})\}$. The weights v_{ij} can be factorized into $v_{ij} = v_{j|i}v_{i\cdot}$, where $v_{i\cdot} = \sum_{j=1}^{n_i} v_{ij}$, and $v_{j|i} = v_{ij}/v_{i\cdot}$. The EL ratio function can be then expressed as

$$\begin{aligned} R(F) &= \prod_{i=1}^{k}\prod_{j=1}^{n_i} N v_{i\cdot} v_{j|i} \\ &= \left(\prod_{i=1}^{k}\left(\frac{N v_{i\cdot}}{n_i}\right)^{n_i}\right)\left(\prod_{i=1}^{k}\prod_{j=1}^{n_i} n_i v_{j|i}\right). \end{aligned} \tag{5.2}$$

In ANOVA analysis, we are usually interested in F only through $v_{j|i}$, thus we can take $v_{i\cdot} = n_i/N$. The first product in (5.2) becomes equal to one and the maximization of $R(F)$ is only subject to constraints on $v_{j|i}$.This leads to the likelihood ratio function

$$R(F) = R_k(F_1, \ldots F_k).$$

A triangular array EL theorem [22, Theorem 4.1.] can be used to establish the inference for ANOVA type hypotheses.

Proposition 5.1.1. *(EL ANOVA for the equality of means) [21, p. 1739] Suppose $E(Y_{ij}) = \mu_0$. Let*

$$\mathcal{R}(\mu) = \max_{v_{j|i}}\left\{\prod_{i=1}^{k}\prod_{j=1}^{n_k} n_i v_{j|i} \,\middle|\, \sum_{j=1}^{n_1} v_{j|1}Y_{1j} = \ldots = \sum_{j=1}^{n_k} v_{j|k}Y_{kj} = \mu, \right.$$
$$\left. \sum_{j=1}^{n_k} v_{j|i} = 1, v_{j|i} \geq 0, i = 1, \ldots, k\right\} \tag{5.3}$$

and define $n_0 = \min_{1\leq i\leq k} n_i$. If $\mu = \mu_0 + O(n_0^{-1/2})$ and for each $i = 1, \ldots k$, $\mathrm{Var}Y_{i1}$ is finite and nonzero, then

$$-2\log\max_{\mu}\mathcal{R}(\mu) = \sum_{i=1}^{k} n_i(\bar{Y}_{i\cdot} - \hat{\mu})^2/s_i^2 + O_p(n_0^{-1/2}) \xrightarrow{d} \chi_k^2$$

as $n_0 \to \infty$, where $\bar{Y}_{i\cdot} = n_i^{-1}\sum_j Y_{ij}$, $s_i^2 = n_i^{-1}\sum_j(Y_{ij} - \bar{Y}_{i\cdot})^2$, and $\hat{\mu}$ is the EL estimator of the common mean μ_0 given by

$$\hat{\mu} = \frac{\sum_{i=1}^{k} n_i\bar{Y}_{i\cdot}/s_i^2}{\sum_{i=1}^{k} n_i/s_i^2}.$$

Note that $\hat{\mu}$, the EL estimator of the common mean μ_0, is not the mean of all Y_{ij} as in the classical ANOVA case. Instead, $\hat{\mu}$ weights the group means inversely to the group variances. The convex hull condition for the EL ANOVA case is

$$\min_j Y_{ij} \leq \mu_i \leq \max_j Y_{ij},\ i = 1, \ldots, k.$$

5.2 Main results

We are interested in the null hypothesis

$$H_0^T : \mu_{\alpha\beta 1} = \mu_{\alpha\beta 2} = \ldots = \mu_{\alpha\beta k} =: \mu_{\alpha\beta}, \tag{5.4}$$

where

$$\mu_{\alpha\beta i} = \frac{1}{1-\alpha-\beta} \int_{\xi_\alpha}^{\xi_{1-\beta}} x dF_{i0},$$

and $\mu_{\alpha\beta}$ represents the common population trimmed mean.

Let $Y_{i(1)}, Y_{i(2)}, \ldots, Y_{i(n_i)}$ denote the order statistics of the ith sample, $i = 1, \ldots, k$. Set $r_i = \lfloor n_i\alpha \rfloor + 1$ and $s_i = n_i - \lfloor n_i \beta \rfloor$, where $0 < \alpha < 1/2$ and $0 < \beta < 1/2$ represent the proportion of the observations trimmed from the left and the right tails, respectively. Then $m_i = n_i - \lfloor n_i\alpha \rfloor - \lfloor n_i\beta \rfloor$ is the effective sample size after trimming of the ith group. The group-specific sample trimmed means and trimmed variances are given by

$$\bar{Y}_{\alpha\beta i} = \frac{1}{m_i} \sum_{j=r_i}^{s_i} Y_{i(j)}, \quad S^2_{\alpha\beta i} = \tfrac{1}{m_i} \textstyle\sum_{j=r_i}^{s_i} (Y_{i(j)} - \bar{Y}_{\alpha\beta i})^2.$$

Analogically to the EL ANOVA setting in (5.3), we are only interested in the weights conditioned on the ith sample, $v_{j|i}$. For the sake of simplicity, we will write v_{ij} instead of $v_{j|i}$ from now on. Next, we use the same idea as developed in Chapter 4, defining the EL ratio function directly over the trimmed samples, forcing weights $v_{ij} = 0$ for all $i = 1, \ldots, k$ and $j < r_i$, $j > s_i$. Thus define the EL ratio as

$$R(\mu_{\alpha\beta}) = \sup_{v_{ij}} \left\{ \prod_{i=1}^{k} \prod_{j=r_i}^{s_i} m_i v_{ij}, \sum_{j=r_i}^{s_i} v_{ij} = 1, \sum_{j=r_i}^{s_i} v_{ij}(Y_{i(j)} - \mu_{\alpha\beta}) = 0, i = 1, \ldots, k \right\}.$$

Theorem 5.2.1. *(M. Delesa-Vēliņa et al. [31]) Let $\mu_{\alpha\beta 0}$ be the common population trimmed mean. Assume that F_{i0} is continuous, $F'_{i0}(\xi_\alpha) > 0$ and $F'_{i0}(\xi_{1-\beta}) > 0$ for each $i = 1, \ldots, k$. If $\mu_{\alpha\beta i} = \mu_{\alpha\beta 0} + O(n_0^{-1/2})$, $i = 1, \ldots, k$, where $n_0 = \min_{1\le i\le k} n_i$, then under H_0^T (5.4),*

$$\sum_{i=1}^{k} a_i l_i := \sum_{i=1}^{k} a_i m_i (\bar{Y}_{\alpha\beta i} - \bar{Y}_{\alpha\beta})^2 / S^2_{\alpha\beta i} + O_p(n_0^{-1/2}) \xrightarrow{d} \chi^2_{(k-1)}$$

as $n_0 \to \infty$, where $\bar{Y}_{\alpha\beta}$ is the EL estimator of the common trimmed mean,

$$\bar{Y}_{\alpha\beta} = \frac{\sum_{i=1}^{k} \bar{Y}_{\alpha\beta i} m_i / S^2_{\alpha\beta i}}{\sum_{i=1}^{k} m_i / S^2_{\alpha\beta i}} + o_p(n_0^{-1/2}),$$

and the scaling factors are given by

$$a_i = \sigma^2_{\alpha\beta i} / ((1-\alpha-\beta)\tau^2_{\alpha\beta i}) \ . \tag{5.5}$$

The quantities $\sigma^2_{\alpha\beta i}$ and $\tau^2_{\alpha\beta i}$ for the ith trimmed sample are given by (4.1) *and* (4.2)*.*

Proof. The proof can be found in [31].

Chapter 6. Simulation and data analysis results

In this chapter, performance of the newly-established empirical likelihood methods for robust location estimators is analyzed in situations where the classical assumptions regarding the normality and variance homogeneity do not hold. The effects of the shape of the distribution (skewness, heavy tails or outliers with or without variance heterogeneity) are investigated in simulation setting regarding the ability to control the type I error, empirical coverage of confidence intervals and the power of the tests. The application of the methods to some real data sets are considered. The newly-established methods are compared to well-known methods of classical and robust statistics.

The performance of the newly-established EL methods has been analyzed before in author's publications [31], [32] and [8]. In this chapter the conclusions drawn before are recapitulated, and some further comparative analysis is carried out. The results for the EL-based methods were computed using `R` package `EL` [6] as well author's custom-made `R` functions.

6.1 EL method for the difference of two smoothed Huber estimators

In Delesa-Vēliņa et. al. [32], the difference of two smooth Huber estimators was considered. The simulation study involved data generated from symmetrical heavy-tailed distributions, double exponential and Huber's least favorable distribution. Note that Lemma 3.1.2 stipulates that the asymptotic results of Theorem 3.1.1 hold for the smoothed Huber estimator if the underlying distributions F_1 and F_2 are symmetric. However, we were interested to evaluate the effects of departure from the symmetry assumption empirically, since skewed distributions are common in practical settings, thus asymmetric gamma distribution with and without uniformly distributed contamination was considered as well. Data sets of equal sample size (50 and 100) with and without variance heterogeneity were considered. The focus of the simulation study was the empirical coverage of 95% confidence intervals and the power of the tests.

Computation of the smoothed M-estimate (2.11) requires the asymptotic variance V of the initial non-smooth M-estimator. Regarding the estimation of V, two variants of the test for the difference of two smoothed Huber estimators were analyzed: first, V was set equal to 2.046 as recommended in [12], and second, V was estimated for the particular distribution using Monte Carlo simulations. MAD was used as a preliminary estimate of the scale parameter σ of the underlying distribution as required by Lemma 3.1.2. For comparison, hypothesis tests regarding the difference of means were included, namely, Student's t-test and EL test for the difference of means of Example 1.3.1.

Regarding the results of the empirical coverage, the results were as follows. 1. For the symmetrical distributions (double exponential and Huber's least favorable distribution) all methods gave similar results, the empirical coverage being close to the nominal 95%. This held regardless of the degree of the variance heterogeneity. 2. Regarding the uncontaminated gamma distribution, the empirical coverage of the methods based on the means and the method based on the smoothed

Huber estimator with V estimated was again close to the nominal 95%. However, for the method based on the smoothed Huber estimator with $V = 2.046$, the empirical coverage was lower, being only 0.879 for moderate shape difference ($\sigma = 3$) and 0.832 for large shape difference ($\sigma = 20$) when $n = 100$. 3. For gamma distributions with 6% or 20% of contamination, the new EL method based on the V estimated had overall better empirical coverage than Student's t-test and EL test for the difference of means. 4. The EL method based on $V = 2.046$ gave inconsistent results.

The empirical power was analyzed sampling from two gamma distributions of differing shapes and means, with and without uniformly distributed contamination. In the case with no contamination, the EL method for the difference of the means and the EL method for the difference of two smoothed Huber estimators with V estimated had similar power and outperformed the t-test. With $V = 2.046$, the EL method for the difference of smoothed Huber estimators had substantially lower power. In the case involving uniform contamination, the EL method for the difference of two means was not very robust and had power similar to t test. The EL method for two smoothed Huber estimators with simulated V had the highest power.

Based on these findings, it was concluded that for symmetrical distributions, the asymptotic variance V of the initial non-smooth Huber estimator can be considered a tuning parameter, and can be fixed to a constant as recommended in [12]. However, for skewed distributions it is not the case and estimating V should be preferable. In practical situations V could be estimated using the nonparametric bootstrap method.

6.2 EL method for the difference of two trimmed means

The EL method for the comparison of two trimmed means was considered in M. Delesa-Vēliņa et al. [8]. The simulation study involved various aspects of violation of the classical assumptions: underlying distributions of various shapes, as well as unbalanced sample sizes and variance heterogeneity combined. The empirical level and the power of the tests was analyzed. The standard normal distribution, t_2-distribution, skewed χ^2_3 and χ^2_1 distributions, as well as two contaminated normal distributions were considered under balanced ($n_1 = n_2$) and unbalanced ($n_2 = 2n_1$) sample size scenarios. ,

Methods included for the comparison were: Student's t-test, Welch's test [33] and EL test (Example 1.3.1) – for the comparison of means; Yuen's test [37] with bootstrap-t approximation [36, Table 5.6] – for the comparison of the trimmed means, as well as the EL ANOVA method for comparing trimmed means described in Chapter 5. Regarding the trimmed means, two trimming versions were considered: 10% and 20% trimming. For Yuen's test, the *R* package *WRS2* [17] function *yuenbt* was used.

Comparison with the EL test for the difference of two smoothed Huber estimators was not included in the study in Delesa-Velina et al. [8], but has been included in the thesis and is reported below. Regarding the smoothed Huber estimators, two versions of the test – with the asymptotic variance V of the initial non-smoothed Huber estimate fixed to 2.046 (panel *ELHubVF*) and V estimated by 10,000 Monte Carlo simulations (panel *ELHubVE*) were considered. We exclude the comparison with the EL ANOVA method for the trimmed means from the results below and comment on it in Chapter 6.3. We present here only the results for the balanced sample size scenario.

As the first aspect, the empirical level was explored as a function of the sample size for two equal underlying distributions, see Figure 6.1. Horizontal dotted lines indicate the simulation error as two standard deviation intervals around the nominal level, the standard deviation being calculated as $\sqrt{\alpha(1-\alpha)/5000}$ yielding the interval $(0.047, 0.053)$. For both newly-established EL two-sample methods, the simulation results confirm the convergence of empirical level to the nominal under the null hypothesis for $N(0,1)$ and t_2 distributions, as well as for contaminated normal distributions. For heavy-tailed distributions, such as contaminated normal and t_2-distribution, the convergence of the EL tests for the trimmed means and smoothed Huber estimators is considerably faster than that of the tests based on the means. In some settings, the EL test for the difference of 10% trimmed means converges faster to the empirical level than the test for the 20% trimmed means, and thus it would be preferable to use 10% trimming in small samples (under 30). The EL-based methods for the comparison of the trimmed means converge to the nominal level more slowly than Yuen's test with a bootstrap-t approximation.

For skewed distributions, the results depend on the test used. The EL method for the trimmed means converge to the nominal level, although more slowly than Yuen's test. For the moderately skewed χ^2_3 distribution, the EL Huber test with V estimated converged to the nominal level less quickly than the tests based on the trimmed means, while the version with V fixed did not converge to the nominal level at all. This result is concordant with findings in Velina et. al. [32], where it was concluded that fixing $V = 2.046$ yields empirical coverage lower than nominal when the underlying distributions are skewed and of differing shapes. For the very skewed χ^2_1 distribution, the empirical level of the tests based on Huber estimators did not converge to the nominal at all. This might seem in contrast to the results of [32], however, the interpretation of this result may lay in the degree of skewness of the distributions – none of the distributions considered in [32] were as skewed as χ^2_1.

As the second aspect, the power of the tests was investigated under various location differences Δ_0, where $\Delta_0 = j \cdot 0.04 \cdot \delta$, $j = 1, \ldots, 25$. The value $\delta = F^{-1}(0.841) - F^{-1}(0.5)$ was chosen as the difference between the 84.13% and 50% percentile of the underlying distribution F being considered, thus allowing to compare power analysis results between different types of distributions. Such approach has been previously used in [9]. The same distributions as for the empirical level simulations are considered. Sample size $n_1 = 50$ was chosen as being sufficient for most of the tests to control the empirical type I error for the distributions considered.

Power simulation results are presented in Figure 6.2. Regarding the newly-established EL test for the difference of trimmed means, its power close to that of Student's t-test under the standard normal distribution, while exceeds it considerably for most of the nonnormal distribution settings, except for the moderately skewed χ^2_3 distribution. Moreover, the new test has a comparable power to that

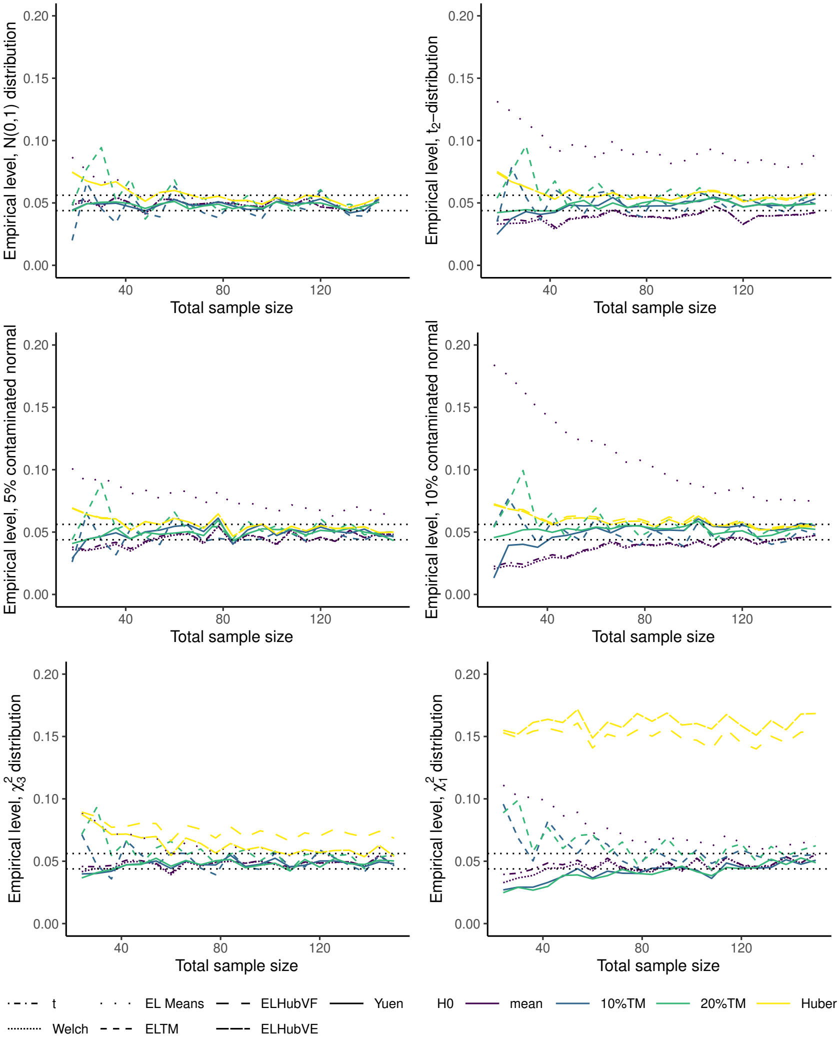

Figure 6.1: Empirical level of the tests as a function of the total sample size $n_1 + n_2$ for various distributions, balanced sample sizes. Top: $N(0,1)$ distribution (left), t_2-distribution (right). Middle: 5% contaminated normal distribution $0.95N(0,1) + 0.05N(0,25)$ (left), 10% contaminated normal distribution $0.9N(0,1) + 0.1N(0,100)$ (right). Bottom: χ^2_3 distribution (left), χ^2_1 distribution (right). Tests considered: Student's t-test (t), Welch's test (Welch), Yuen's test for the trimmed means with bootstrap-t approximation (Yuen), EL test for the means (EL Means), EL test for the difference of trimmed means (ELTM), EL test for the difference of smoothed Huber estimators, with $V = 2.046$ fixed (ELHubVF) and V simulated (ELHubVE). Colour indicates the relevant hypothesis tested – violet for methods comparing means, blue and green for 10% and 20% trimmed means, respectively, yellow for smoothed Huber estimators.

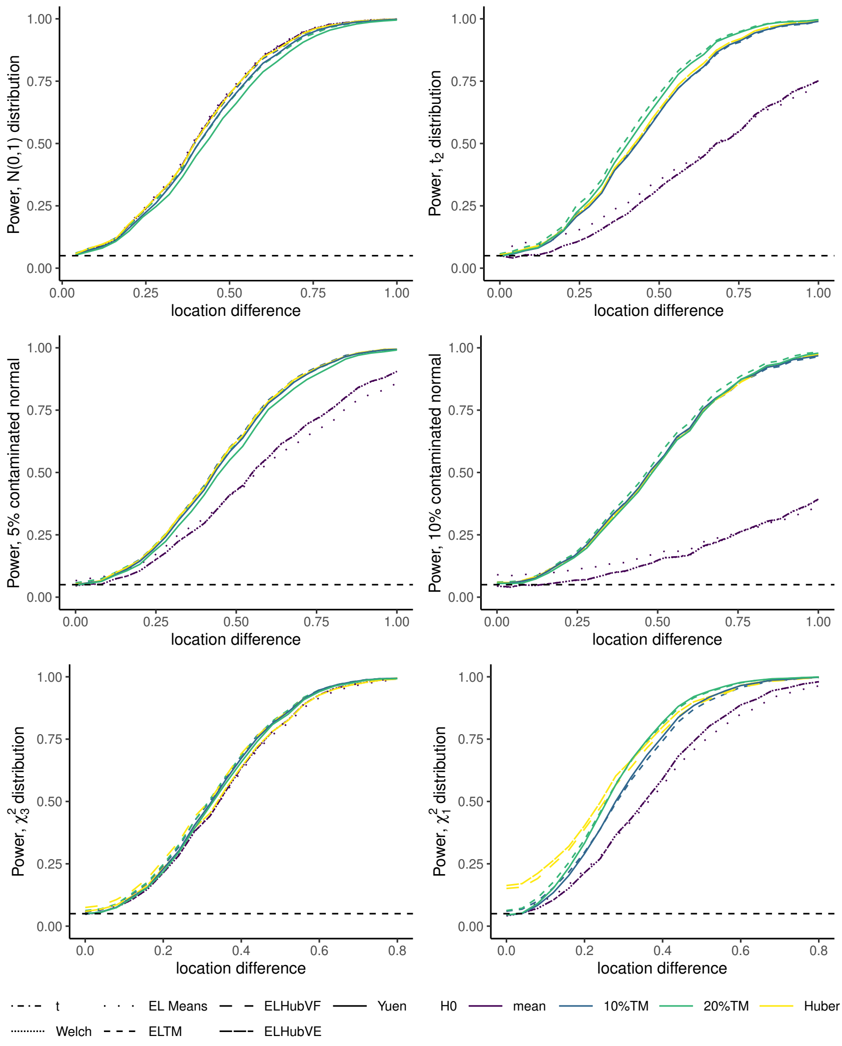

Figure 6.2: Power of the tests as a function of location difference Δ_0 for various distributions, balanced sample sizes $n_1 = n_2 = 50$. Top: $N(0,1)$ distribution (left), t_2-distribution (right). Middle: 5% contaminated normal distribution $0.95N(0,1) + 0.05N(0,25)$ (left), 10% contaminated normal distribution $0.9N(0,1) + 0.1N(0,100)$ (right). Bottom: χ^2_3 distribution (left), χ^2_1 distribution (right). Tests considered: Student's t-test (t), Welch's test (Welch), Yuen's test for the trimmed means with bootstrap-t approximation (Yuen), EL test for the means (EL Means), EL test for the difference of trimmed means (ELTM), EL test for the difference of smoothed Huber estimators, with $V = 2.046$ fixed (ELHubVF) and V simulated (ELHubVE). Color indicates the relevant hypothesis tested – violet for methods comparing means, blue and green for 10% and 20% trimmed means, respectively, yellow for smoothed Huber estimators.

of Yuen's test, in some cases even exceeding it. The power of the test of two smoothed Huber estimators is similar to that of the test of the trimmed means, and in case of the $N(0,1)$ and 5% contaminated normal distribution it exceeds most of the tests based on the trimmed means. The power of the test at the very skewed χ^2_1 distribution is higher than the rest of the tests, but this is a result of an incorrect test level at H_0.

Table 6.1: Empirical level of tests and robustness by Bradley's criterion under unbalanced and heterogeneous designs for standard normal and χ^2_3 distributions. Small sample designs (panel *small*) and large sample designs (panel *large*) are presented separately. *pos* indicates positive variance and sample size pairings, *neg* indicates negative pairings, and *equal* refers to equal variances. Tests considered are: Student's t-test (t), Welch's test (Welch), Yuen's test for the trimmed means with bootstrap-t approximation (Yuen), EL test for the means (EL Means), EL test for the difference of trimmed means (ELTM), EL test for the difference of smoothed Huber estimators, with $V = 2.046$ fixed (ELHubVF) and V simulated (ELHubVE).

Standard normal distribution

	Empirical level of tests						# robust	
sample size	small			large			small	large
variance pairings	equal	pos	neg	equal	pos	neg		
t	0.049	0.020	0.113	0.049	0.019	0.107	4	2
Welch t	0.050	0.050	0.049	0.049	0.050	0.051	10	10
EL Means	0.060	0.061	0.068	0.050	0.051	0.054	9	10
ELTM 10%	0.047	0.050	0.054	0.052	0.052	0.055	10	10
ELTM 20%	0.077	0.077	0.102	0.053	0.054	0.057	2	10
Yuen 10%	0.050	0.051	0.047	0.049	0.050	0.051	10	10
Yuen 20%	0.050	0.050	0.049	0.049	0.050	0.051	10	10
ELHubVF	0.061	0.060	0.063	0.052	0.051	0.053	10	10
ELHubVE	0.060	0.058	0.062	0.051	0.051	0.053	10	10

χ^2_3 distribution

	Empirical level of tests						# robust	
sample size	small			large			small	large
variance pairings	equal	pos	neg	equal	pos	neg		
t	0.048	0.035	0.129	0.052	0.022	0.110	6	2
Welch	0.049	0.064	0.069	0.052	0.051	0.055	9	10
EL Means	0.070	0.071	0.082	0.054	0.051	0.054	6	10
ELTM 10%	0.051	0.054	0.058	0.053	0.052	0.053	10	10
ELTM 20%	0.078	0.078	0.105	0.055	0.052	0.056	3	10
Yuen 10%	0.047	0.052	0.048	0.050	0.049	0.050	10	10
Yuen 20%	0.048	0.051	0.052	0.051	0.048	0.050	10	10
ELHubVF	0.084	0.080	0.079	0.070	0.071	0.071	1	9
ELHubVE	0.075	0.071	0.070	0.057	0.057	0.057	6	10

As the third aspect, the robustness of the tests under the normal and χ^2_3 distribution with various degrees of heterogeneity combined with unbalanced sample sizes was analyzed. The following unbalanced designs were considered: two small sample designs, $(n_1, n_2) = (15, 25)$ and $(n_1, n_2) = (25, 35)$, and two large-sample designs, $(n_1, n_2) = (80, 120)$ and $(n_1, n_2) = (160, 240)$. For the degree of heterogeneity, the ratio of the variances of the two populations were chosen 1:16 and 1:36, and the equal case 1:1 for comparison. Three possible unbalanced design and variance pairing conditions were considered – positive, where the largest variance is associated with the largest sample size, negative, where the smallest variance is associated with the largest sample size, and equal for the comparison. To ensure that the null hypothesis remained true for all the settings of variance heterogeneity being considered, χ^2_3 variates were standardized to have the theoretical location parameter 0 and standard deviation 1 prior to scaling to the desired variance ratio. Instead of reporting the result of each simulation experiment, the results are grouped over (i) small and large sample designs, and (ii) variance pairing conditions (positive, negative and equal). We evaluated the test performance by Bradley's liberal criterion for robustness [4]. Namely, the test is considered robust if its empirical type I error $\hat{\alpha}$ falls into the interval $0.5\alpha \leq \hat{\alpha} \leq 1.5\alpha$. We counted the number of designs where each test passes the robustness condition, i.e., where it yields an empirical type I error in the interval $[0.025, 0.075]$.

The results are reported in Table 6.1. The robustness by Bradley's criterion of the new test for the difference of 10% trimmed means was confirmed for all unbalanced and variance heterogeneity designs both for the normal and the χ^2_3 distribution. However, the EL test for 20% trimmed means failed to be robust for most of the small sample settings, despite showing good results for the large sample settings. Regarding the tests based on the smoothed Huber estimators, they were robust to heterogeneity and unbalanced sample sizes under the normal distribution. For the skewed χ^2_3 distribution, the test can be considered robust only for large sample sizes, the version with V estimated being more precise than the version with V fixed.

6.3 EL-based ANOVA method for the trimmed means

The performance of the empirical likelihood ANOVA method for the trimmed means has been analyzed previously in M. Delesa-Vēliņa et al. [31, 32] in detail. The study in [31] explored the properties of EL ANOVA method for the trimmed means with 5%, 10% and 20% symmetric trimming. The empirical probability of a type I error of the method under various skewed distributions was considered. For comparison, the classical ANOVA F-test, Welch heteroscedastic ANOVA F-test [34] and EL ANOVA test for the means [21], as well as Yuen's test [36, Table 7.1] for the trimmed means were included in the study.

The study involved a comparison of three groups of equal sizes ranging from $n = 20$ to $n = 500$. We recorded the rate of empirical type I errors under H_0 sampling from χ^2_3 distribution, lognormal distribution, gamma distribution with shape parameter $a = 2$ and scale parameter $\sigma = 1$, and the skew-normal distribution [2] with location parameter $\xi = 0$, scale parameter $\omega = 1$, and slant parameter $\alpha = 1$. For the scenario with heterogeneous variances, we further transformed the simulated data as to have the ratios between variances to be either 1:4:9 or 1:1:36.

See [31] for the results of the study. Regarding the newly-established EL ANOVA test for the trimmed means, the simulations with three groups sampled from skewed distributions with equal variances show that the test converges to the empirical level for all distributions considered. Scenario with heterogeneous variances suggests that the test is rather oversized for small sample sizes (below 100). For large sample sizes, the test converges to the empirical level. For all heterogeneous settings, the EL-based ANOVA for the trimmed means is more robust than the classical F-test, having the empirical rejection rates closer to the nominal level.

The study in M. Delesa-Vēliņa et al. [8] provided additional insight in the performance of the EL ANOVA method for the trimmed means in case of comparing two groups. Simulations with two groups suggest that the ANOVA-like EL test for the trimmed means converges to the empirical level also when data is sampled from heavy tailed distributions or distributions containing outliers. In addition, it has good power properties, exceeding the power of the methods based on the means when the data distributions are not normal. Similarly as in the case of three groups, the case of two groups reveals that EL ANOVA test for the trimmed means is not robust to the combination of variance heterogeneity and skewness for small sample sizes. It should be noted, however, that the simulation results with two groups give only a limited view on the behavior of ANOVA-like methods.

6.4 Analysis of data sets

We explore a number of real data sets exhibiting various departures from normality. We are interested in testing the null hypothesis of equal location parameters of two or more populations, using the newly-established EL-based methods as well as some well known classical and robust methods for the comparison. See thesis Chapter 6.3 as well as M. Delesa-Vēliņa et al. [31, 32, 8] for the data analysis results.

For the two-sample methods, we observed that the tests based on the trimmed means could lead to the opposite conclusion about H_0 in comparison to the tests based on the means. We observed that the p-values of the EL test and Yuen's test for the difference of trimmed means were quite close, and this was true also for small sized samples. The confidence intervals of the EL-based tests were somewhat shorter. Regarding the test for the difference of two smoothed Huber estimators, we noted that the p-values can be quite different depending on the value of the asymptotic variance V, especially if the underlying distribution is substantially skewed. This is in line with our simulation results, and once again suggests that using EL test for the two smoothed Huber estimators should be avoided for very skewed data sets. In cases of moderate skewness the Huber test with V estimated yields p-values close to the EL test for the means.

Regarding the ANOVA methods, we noted that, for each trimming proportion, the p-values from the EL ANOVA for trimmed means and Yuen's test are very similar. Interestingly, the p-values from the EL ANOVA for means and Welch's heteroscedastic ANOVA test were also very similar between themselves. The tests based on the trimmed means could lead to different conclusions in comparison to the tests based on the means.

Conclusions

The main aims of the research have been achieved. New EL-based methods for comparing two and more populations based on robust location parameter estimators have been developed:

1. An EL-based method for comparing two location M-estimators;

2. An EL-based method for comparing two population trimmed means;

3. An EL-based ANOVA method for comparing more than two population trimmed means.

The conditions for the use of the methods were established and the asymptotic results were proven. Using the approach of Y. Qin and L. Zhao [25], it was shown that under particular conditions the limiting law of the EL log likelihood ratio for the difference of two M-estimators is the χ_1^2 distribution, similarly as in the case of the difference of the means. It was shown that the smoothed Huber estimator fits under the established conditions. The smoothing principle provided by F. Hampel et al. [11] is important, since it allows to construct smooth EL estimating functions essential for the conditions to hold.

We generalized the one-sample EL for the trimmed mean by G. Qin and M. Tsao [23] to the two-sample and ANOVA case. The limiting law of the EL log likelihood ratio for the difference of the trimmed means is a scaled χ_1^2, and is essentially related to the asymptotic distribution of the trimmed mean established by S. Stigler [27]. In the case of the EL ANOVA-like method for the trimmed means, there are scaling constants involved for each of the k populations, and the resulting limiting law is χ_{k-1}^2. This result is related to the EL ANOVA for the equality of means established by A. B. Owen in [21].

Simulation study was realized to explore the behaviour of the methods when sampling from various types of probability distributions, especially when the classical assumptions of normality and variance equality do not hold. We observe that EL methods based on the trimmed means are robust to distributional skewness, heavy tails, outliers and variance heterogeneity combined with unbalanced sample sizes, in a sense that the empirical type I error converges to the nominal level. For the difference of smoothed Huber estimators, however, the robustness was not confirmed for very skewed distributions, but held for distributions with moderate skewness, heavy tails and outliers. It should be noted that the power of the methods considerably exceeds that of the methods based on the means.

For extension of the thesis research, there are several options. First, one might consider the difference of other M-estimators than the smoothed Huber estimator. Second, one might consider the EL method for the non-smooth criterion functions developed by in [19]. Their approach has the potential of wider application, however, it has a slower theoretical convergence rate. To the best of our knowledge, the comparison of the smooth and the non-smooth approaches for the two-sample and ANOVA problems has not been done and would be of interest in the future.

Theses

1. **Empirical likelihood method for comparing two location M-estimators was developed, conditions for the application of the method were established and the asymptotic results were proven. It was shown that the conditions hold for the difference of two smoothed Huber estimators. Simulation study showed that the method has good robustness properties when sampling from distributions containing outliers or heavy tails.**

   Simulation study showed that the method version with asymptotic variance parameter V confirms the robustness of the level of the test (i.e., the empirical level of the test is close to the nominal) when sampling from symmetric, heavy tailed and moderately skewed distributions. This method has a higher power than the methods based on the means when normality does not hold. This method is robust to the combination of variance heterogeneity and unbalanced sample design for normal distribution settings, and for large sample settings also for chi squared distribution settings. [32]

2. **Empirical likelihood method for comparing two trimmed means was developed and the asymptotic results were proven. Simulation study when sampling from symmetric, heavy tailed and skewed distributions confirmed the good robustness properties of the method.**

   The empirical level of the new test is robust and, moreover, it has higher power than the classical tests when sampling from skewed or heavy-tailed distributions. EL test for the difference of 10% trimmed means was robust to the combination of variance heterogeneity and unbalanced sample sizes both for normal and chi squared distribution settings. [8]

3. **Empirical likelihood-based ANOVA method for comparing more than two population trimmed means was developed and the asymptotic results were proven. Simulation study involving skewed distributions demonstrated the good robustness properties of the method in comparison to the classical F-test.**

   Simulation study with three groups involving skewed distributions confirmed that the test level was robust. The test empirical level is closer to the nominal than that of the classical F-test when the variances are not equal.

   Simulation study with two groups showed that the new method has higher power than the ANOVA methods based on the means when the underlying distribution is severely skewed, contains outliers or is heavy-tailed. The EL ANOVA method for 10% trimmed means is robust to combination of unbalanced sample sizes and variance heterogeneity both in normal and chi squared distribution settings when the sample size is large. [31], [8]

Author's publications

P1 M. Velina, J. Valeinis, L. Greco, G. Luta. Empirical Likelihood-Based ANOVA for Trimmed Means. *International Journal of Environmental Research and Public Health.* 13(10):953, 2016. https://doi.org/10.3390/ijerph13100953

P2 M. Velina, J. Valeinis, G. Luta. Empirical Likelihood-Based Inference for the Difference of Two Location Parameters Using Smoothed M-Estimators. *Journal of Statistical Theory and Practice* 13(34), 2019. https://doi.org/10.1007/s42519-019-0037-8

P3 M. Delesa-Vēliņa, J. Valeinis, G. Luta. Comparing Two Independent Populations Using a Test Based on Empirical Likelihood and Trimmed Means. *Lithuanian Mathematical Journal* 61: 199-–216, 2021. https://doi.org/10.1007/s10986-021-09516-x

List of conferences

C1 J. Valeinis, M. Vēliņa, G. Luta. Empirical likelihood-based inference for the difference of smoothed Huber estimators. 11th International Conference on Robust Statistics, Valladolid, Spain, 2011.

C2 M. Vēliņa, J. Valeinis, G. Luta. Empirical likelihood-based methods for the difference of two trimmed means. 12th International Conference on Robust Statistics, Burlington, Vermont, USA, 2012.

C3 M. Vēliņa, J. Valeinis. Empirical likelihood based robust inference for trimmed mean. 18th International Conference on Mathematical Modeling and Analysis, Tartu, Estonia, 2013.

C4 M. Vēliņa, J. Valeinis, G. Luta. Robust inference using empirical likelihood based ANOVA methods. 13th International Conference on Robust Statistics, St. Petersburg, Russia, 2013.

C5 M. Vēliņa. Methods of robust statistics, using empirical likelihood method. 72nd Scientific Conference of University of Latvia, Riga, Latvia, 2014.

C6 M. Vēliņa, J. Valeinis. Empirical likelihood based robust ANOVA inference. 11th International Vilnius Conference on Probability & Mathematical Statistics, Vilnius, Lithuania, 2014.

C7 M. Vēliņa, J. Valeinis, R. Nedovis, G. Luta. A comparison of robust empirical likelihood-based ANOVA methods. 14th International Conference on Robust Statistics, Halle, Germany, 2014.

C8 M. Vēliņa. Robust empirical likelihood function for two and more samples. 73rd Scientific Conference of University of Latvia, Riga, Latvia, 2015.

C9 M. Vēliņa, J. Valeinis. Applications of robust ANOVA methods. 20th International Conference on Mathematical Modeling & Analysis, Sigulda, Latvia, 2015.

C10 M. Vēliņa, J. Valeinis. Two-sample empirical likelihood in the presence of nuisance parameters. European Meeting of Statisticians, Amsterdam, Netherlands, 2015.

C11 M. Vēliņa. Robust empirical likelihood inference for two sample location problem. 12th conference of Latvian Mathematics Society, Ventspils, Latvia, 2018.

C12 M. Delesa-Vēliņa. Empirical likelihood inference for trimmed means. 78th Scientific Conference of University of Latvia, Riga, Latvia, 2020.

Bibliography

[1] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja. *A First Course in Order Statistics*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia, 2008.

[2] A. Azzalini. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, (12):171–178, 1985.

[3] P. J. Bickel and K. A. Doksum. *Mathematical Statistics: Basic Ideas and Selected Topics, Volumes I-II Package*. Texts in Statistical Science. Chapman and Hall/CRC, New York, 1st edition, 2015.

[4] J. V. Bradley. Robustness? *British Journal of Mathematical and Statistical Psychology*, 31:144–152, 1978.

[5] G. Casella and L. R. Berger. *Statistical Inference*. Brooks/Cole, Australia, 2nd edition, 2002.

[6] E. Cers and J. Valeinis. *EL: Two-sample Empirical Likelihood*, 2011. R package version 1.0.

[7] N. A. Cressie and H. J. Whitford. How to use the two sample t-Test. *Biometrical Journal*, 28(2):131–148, 1986.

[8] M. Delesa-Vēliņa, J. Valeinis, and G. Luta. Comparing two independent populations using a test based on empirical likelihood and trimmed means. *Lithuanian mathematical Journal*, 61, 2021.

[9] R. Fried and H. Dehling. Robust nonparametric tests for the two-sample location problem. *Statistical Methods & Applications*, 20(2):409–422, 2011.

[10] N. L. Glenn and Y. Zhao. Weighted empirical likelihood estimates and their robustness properties. *Computational Statistics & Data Analysis*, 51:5130–5141, 6 2007.

[11] F. R. Hampel. On the philosophical foundations of statistics: Bridges to Huber's work, and recent results. In H. Rieder, editor, *Robust statistics, data analysis, and computer intensive methods. In honor of Peter Huber's 60th birthday*, pages 185–196. Springer, New York, 1996.

[12] F. R. Hampel, C. Hennig, and E. A. Ronchetti. A smoothing principle for the Huber and other location M-estimators. *Computational Statistics & Data Analysis*, 55:324–337, 2011.

[13] N. L. Hjort, I. W. McKeague, and I. Van Keilegom. Extending the scope of empirical likelihood. *The Annals of Statistics*, 37(3):1079–1111, 2009.

[14] P. J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73 – 101, 1964.

[15] P. J. Huber and E. Ronchetti. *Robust Statistics.* Wiley, New York, 2nd edition, 2009.

[16] L. M. Lix and H. J. Keselman. To trim or not to trim: Tests of location equality under heteroscedasticity and nonnormality. *Educational and Psychological Measurement*, 58(3):409–429, 1998.

[17] P. Mair and R. Wilcox. Robust Statistical Methods in R Using the WRS2 Package. *Behavior Research Methods*, 52:464–488, 2020.

[18] R. A. Maronna, R. D. Martin, and V. J. Yohai. *Robust Statistics. Theory and Methods (with R).* Wiley, New York, 2nd edition, 2018.

[19] E. M. Molanes Lopez, I. Van Keilegom, and N. Veraverbeke. Empirical likelihood for non-smooth criterion functions. *Scandinavian Journal of Statistics*, 36(3):413–432, 2009.

[20] A. B. Owen. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249, 1988.

[21] A. B. Owen. Empirical likelihood for linear models. *The Annals of Statistics*, 19(4):1725 – 1747, 1991.

[22] A. B. Owen. *Empirical Likelihood.* Chapman & Hall/CRC, New York, 2001.

[23] G. Qin and M. Tsao. Empirical likelihood ratio confidence interval for the trimmed mean. *Communications in Statistics - Theory and Methods*, 31(12):2197–2208, 2002.

[24] J. Qin and J. F. Lawless. Empirical likelihood and general estimating equations. *The Annals of Statistics*, 22(1):300 – 325, 1994.

[25] Y. Qin and L. Zhao. Empirical likelihood ratio confidence intervals for various differences of two populations. *Journal of Systems Science and Mathematical Sciences*, pages 23–30, 2000.

[26] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2020.

[27] S. M. Stigler. The asymptotic distribution of the trimmed mean. *The Annals of Statistics*, 1(3):472–477, 1973.

[28] Student. The probable error of a mean. *Biometrika*, 6(1):1–25, 1908.

[29] J. Valeinis. *Confidence bands for structural relationship models.* PhD thesis, University of Goettingen, Goettingen, 2007.

[30] J. Valeinis, E. Cers, and J. Cielens. Two-sample problems in statistical data modelling. *Mathematical Modelling and Analysis*, 15(1):137–151, 2010.

[31] M. Vēliņa, J. Valeinis, L. Greco, and G. Luta. Empirical likelihood-based ANOVA for trimmed means. *International Journal of Environmental Research and Public Health*, 13(10):1–13, 2016.

[32] M. Vēliņa, J. Valeinis, and G. Luta. Likelihood-based inference for the difference of two location parameters using smoothed M-estimators. *Journal of Statistical Theory and Practice*, 13, 2019.

[33] B. L. Welch. The generalization of "Student's" problem when several different population variances are involved. *Biometrika*, 34(1):28–35, 1947.

[34] B. L. Welch. On the comparison of several mean values: An alternative approach. *Biometrika*, 38:330–336, 1951.

[35] R. R. Wilcox. Comparing the means of two independent groups. *Biometrical Journal*, 32(7):771–780, 1990.

[36] R. R. Wilcox. *Introduction to Robust Estimation and Hypothesis Testing.* Elsevier Academic Press, London, 4th edition, 2017.

[37] K. K. Yuen. The two-sample trimmed t for unequal population variances. *Biometrika*, 61:165–170, 1974.

Acknowledgements

I would like to express my gratitude to Dr. math. Jānis Valeinis for introducing me to the scientific and pedagogical work – his optimism and dedication motivated me to choose the path of the researcher. I would like to thank Dr. biostat. George Luta for the excellent idea to explore the field of robust statistics. I thank Dr. math., prof. Inese Bula and Dr. phys., prof. Gunta Krūmiņa for trusting me and allowing to develop myself as a teaching assistant.

Warm thanks to my PhD colleagues Leonora Pahirko and Tatjana Pladere for cheering me up in difficult moments. Thanks to my classmate Agnese Melbārde for persuading me that it is the statistics studies that will bring the richest opportunities in our lives. I thank my large family and especially my dear husband Žils Deless-Vēliņš – for the patience and encouragement, for providing the perspective, and for the faith in my doctoral work that exceeded mine.

And throughout I am thankful for the music of voices, strings, winds, keyboards and all other kinds that along with the statistics form the right harmony in my life.