UNIVERSITY
OF LATVIA

FACULTY OF PHYSICS, MATHEMATICS AND OPTOMETRY

Māra Delesa-Vēliņa

EMPIRICAL LIKELIHOOD METHOD FOR A LOCATION PARAMETER USING SOME ROBUST ESTIMATORS

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

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

List of designations

$\lfloor . \rfloor$ – the floor function
$\xrightarrow{a.s.}$ – convergence with probability 1
$\xrightarrow{d}$ – convergence in distribution
$\xrightarrow{p}$ – convergence in probability
$\#A$ – number of occurrences of an event A
cdf - cumulative distribution function
E – expectation
E_F – expectation with respect to a distribution F
ecdf - empirical cumulative distribution function
EL – empirical likelihood
i.i.d. – independent and identically distributed
$F_{n,k}$ – F-distribution with n and k degrees of freedom
FBP – finite-sample breakdown point
I_A – indicator function of an event A
log – the natural logarithm
LRT – likelihood ratio test
maxeig(A) – the largest eigenvalue of a symmetric matrix A
mineig(A) – the smallest eigenvalue of a symmetric matrix A
MLE – maximum likelihood estimator
MSE – mean squared error
$o_p(a_n)$ – a random variable X_n such that for a set of constants a_n, $X_n/a_n \xrightarrow{p} 0$
pdf – probability density function
pmf – probability mass function
sgn – the sign function
Var – variance
$X_{(1)}, \ldots, X_{(n)}$ – ordered statistics of a sample $X_1, \ldots, X_n$
δ_a – point mass distribution at a
$\mu_{\alpha\beta}$ – trimmed mean with trimming proportions α, β from the left and the right, respectively
ϕ – density function of the standard normal distribution
Φ – cumulative distribution function of the standard normal distribution
$\mathcal{X}$ – sample space
χ^2_k – chi-square distribution with k degrees of freedom
$\chi^2_{k,1-p}$ – $1-p$ quantile of chi-square distribution with k degrees of freedom

Introduction

Comparing two populations in the classical setting

A common problem in statistical analysis is to compare two populations F_1 and F_2 based on observations of two samples $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$. For example, one might be interested to find whether a new medicament administered to the treatment group is more effective than the placebo given to a control group. Or one may want to test whether two different doses of the same medicament induce similar changes in patients' physiological or biochemical health indicators. The most widely used test in such situations is Student's t-test [40] for comparing the means of two independent normal populations. In case $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ are independent and identically distributed (i.i.d.) and with equal variances, i.e., 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 $H_1 : \mu_1 > \mu_2$ (see, for example, [3, Chapter 9]).

Sample mean has useful theoretical properties: under the normal distribution it is the maximum likelihood estimator of the population mean and hence is asymptotically the most efficient (has the smallest variance) among all unbiased estimators (see, for example, [5, Chapter 10]). Unfortunately, the efficiency of the sample mean decreases quickly under slight departures from normality. Intuitively, a single observation can change the value and standard error of the sample mean by an arbitrary large amount. It can be said that the sample mean is *not robust* to the departures from normality.

At the onset of the theory of estimation the statistical variability was almost exclusively due to measurement errors and as such, was a nuisance to get rid of [18]. It is in this context that Carl Friedrich Gauss introduced the normal distribution and provided an elegant way to describe the behaviour of errors around the arithmetic mean. However, the assumption that the observed data is exactly normally distributed is rarely achievable in most practical applications. The data collected can be sampled from skewed distributions, from distributions with heavy tails (heavier than normal distribution), or it can contain one or several *outliers* (atypical observations deviating from the most of the data).

Regarding the Student's t-test, the presence of outliers or heavy tails inflates the standard error of the mean thus decreasing the power of the test. When distributions differ in skewness, the Student's t-test is not even asymptotically correct [7]. Bernard L. Welch [54] developed an approximate degrees of freedom (ADF) modification that overcomes the problems associated with the inequality of variances, problems associated with outliers and heavy tails remain. When nonnormality occurs simultaneously with the variance heterogeneity and unbalanced sample sizes, the probability of a type I error may considerably differ from the nominal (see, for example, [57] for review).

Comparing multiple populations

Let $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})$, $i = 1, \ldots, k$ be independent samples from k populations $F_1, \ldots, F_k$. The classical method to compare the means of multiple independent populations is the analysis of variance (ANOVA) F-test introduced by Ronald Fisher. ANOVA F-test 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: variance heterogeneity and outliers can break down the results of the test completely when not taken into account properly. As a consequence, ANOVA has problems of controlling the probability of empirical type I error at the specified nominal level. B. Welch [55] proposed a version of the F-test for the heteroscedastic case; similarly as for the two-sample problem, the method is based on the modification of the degrees of freedom of the limiting distribution. However, Welch's ANOVA test is not robust to departures from normality and outliers, especially when the skewness differs among the groups [23].

Empirical likelihood

Empirical likelihood (EL) was introduced by Art B. Owen in 1988 [28]. EL is a nonparametric method that does not require assumptions about the distribution family of the data. A. B. Owen [28] showed that the empirical likelihood 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 [34]. Regarding the two-sample problem, Yonsong Qin and Lincheng Zhao [35] extended the EL method for the difference of two univariate parameters. The properties of EL for some two-sample problems were analysed empirically by Jānis Valeinis et al. [49] and a complementary R program [36] package *EL* [6] was developed. EL method, in an ANOVA-like setting for comparing means of k independent groups, was demonstrated by A. B. Owen in 1991 [30]. A general overview on EL methods can be found in [31]. More recent monographs are written by Mai Zhou [61], and Albert Vexler and Jinhee Yu [50]. The first monograph is devoted to empirical likelihood methods in survival analysis, while the second presents EL methods with applications in biostatistics, especially in light of nonparametric Bayesian inference.

EL is devised by constructing a multinomial distribution on the observed data points, hence it has an advantage that the associated confidence intervals are not necessarily symmetrical and have a data driven shape. For example, EL confidence intervals for the mean tend to be extended in directions where the data is skewed [31]. However, 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]. [43] demonstrated that the finite-sample breakdown point of the length of the EL confidence interval for the mean is $1/n$ based on a sample of size n, namely, a single outlying observation can arbitrarily increase the length of the interval. Thus the problems inherent to the inference for means are relevant to the empirical likelihood setting as well.

A. B. Owen [31] 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 second approach relates to the weighted empirical likelihood. For example, [10] proposed weights depending on an automatic outlier detection procedure based on absolute deviation statistics. Their weighted EL estimator for the mean had a smaller mean squared error (MSE) than the classical EL estimator in simulation settings of normal data contaminated with outliers. [45] consider another weighted EL approach for estimating the common mean of k independent samples, where the weights depend on the sample dispersion estimates. Their method was shown to have an advantage over the classical EL method in simulation setting for data sampled from skewed distributions with heterogeneous variances. [59] further provided a bootstrap-calibration for the weighted empirical likelihood and observed its advantage in the skewed distribution setting.

Robust estimators

The centre of interest of this thesis, however, is the first of Owen's propositions, namely, to study the empirical likelihood method for some *robust estimators*. In particular, we are interested in *robust estimators of location* or *centre* of the data. As for the meaning of the word *robustness*, it signifies "insensitivity to small deviations from the assumptions" [19, 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, when the inefficiency problems associated with the classical estimators (such as sample mean, sample variance and the least squares regression) were started to be studied in a comprehensive manner. J. W. Tukey ([46], cf. [14]) considered *contaminated normal* distributions of the form $(1-\epsilon)N(\mu,\sigma^2)+\epsilon N(\mu,9\sigma^2)$, where ϵ varies from 0 to 1/2 and thus represents a proportion of erroneous observations increased by factor 3. J. W. Tukey compared the efficiency of the mean, median and several other estimators in the contaminated normal distribution setting. As a conclusion, he advocated the use of the 6%-*trimmed mean* that demonstrated good efficiency throughout the whole range of ϵ contaminations.

Trimmed mean is a simple location parameter that is obtained by calculating the arithmetic mean after removing a fixed proportion of the most extreme observations from the sample. [47] proposed a one-sample trimmed t procedure based on t-statistic for trimmed mean and Winsorized square deviations of the sample (in case of Winsorization, the extreme values are shifted towards the middle of the data instead of being removed). This method was extended to the two-sample case by Karen K. Yuen [60]. Stephen M. Stigler provided the conditions for the asymptotic normality of the trimmed mean in [39].

In 1964 P. J. Huber published the seminal paper "Robust Estimation of a Location Parameter" [17], inventing a class of *M-estimators* that, in a sense, is a generalization of the maximum likelihood (MLE) estimators. Consider an MLE estimator $\hat{\theta}=\hat{\theta}(X_1,\dots,X_n)$ of the parameter θ. Then $\hat{\theta}$ is defined as the value that minimizes the minus of the log likelihood function, i.e.,

$$\hat{\theta}=\arg\min_{\theta}\sum_{i=1}^{n}-\log f_\theta(X_i),$$

where f_θ is the probability density function (pdf) of X_is under θ. Huber proposed to replace the function $-\log f_\theta$ with a general ρ-function, i.e.,

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

Equivalently, if ψ is differentiable in θ, $\hat{\theta}$ is the solution of

$$\sum_{i=1}^{n}\psi(X_i,\theta)=0,$$

where $\psi = (\partial/\partial\theta)\rho(x,\theta)$. In the same article, P. J. Huber also set out the conditions for the consistency and asymptotic normality of the newly-established M-estimators. Finally, he demonstrated that there existed a certain 'optimal' M-estimator in a full neighbourhood of the normal distribution. Namely, 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 an M-estimator, now referred to as the *Huber estimator*, defined for a given positive constant c by

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

has the minimax asymptotic variance among all translation-equivariant location estimators. Moreover, Huber estimator is the MLE for the so-called *Huber's least favourable distribution*, which is normal in the middle, but exponential in the tails.

The performance of the Huber estimator is rather sensitive to the behaviour of the underlying distribution F at the points $\pm c$, where the function $\psi(x)$ is not differentiable [17] , and better results could be attained if the ψ-function was smoothed around $\pm c$. The smoothing principle for a Huber estimator was first proposed by Frank Hampel in [12]. For a ψ-function of a general M-estimator, the smoothing principle was presented in 2011 by Frank Hampel et al. [13], where the degree of smoothness depends on the sample size. The smoothed estimators were demonstrated to have smaller MSE than their non-smoothed counterparts in small and moderate sample settings.

Empirical likelihood for robust location estimators in the one-sample case

At first it might seem unnecessary to consider distributional robustness in the case of empirical likelihood, since there are no explicit distributional assumptions involved. However, EL is essentially a likelihood defined on the empirical cdf, thus the robustness properties of the estimators considered are still relevant. A. B. Owen [28] demonstrated that the EL method can be applied to certain M-estimators, including the Huber estimator. [43] gave the expression of a finite-sample breakdown point of the length of the EL confidence interval for the Huber estimator and showed that it is asymptotically equal to 0.5, which is the best attainable value for any estimator.

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. In the context of time series, the dependence can be removed by grouping data in blocks before using the EL method [22], however, such an approach does not seem to be possible for the trimmed mean case. Instead, Gensheng Qin and Min Tsao [33] 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. A different approach was taken in [44], where the EL method for the trimmed mean was constructed as in the case of the means, but a more conservative quantile of the chi-square distribution was used to define the respective confidence interval. However, in a limited simulation study we found that the confidence interval described in [44] does not attain the right coverage probability when the null hypothesis is true.

Aims of the research

The goal of the thesis is to develop empirical likelihood-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 [35]. In particular, consider the smoothed Huber estimator [13].

2. Develop an empirical likelihood method for the difference of two population trimmed means extending the results of G. Qin and M. Tsao [33], and Y. Qin and L. Zhao [35].

3. Develop an ANOVA-like empirical likelihood method to compare the trimmed means of multiple populations, extending the results of G. Qin and L. Tsao [33], and A. B. Owen [30].

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.

Structure of the thesis

The thesis is organized as follows. The preliminaries are given in the first two chapters. In Chapter 1, the theory of the empirical likelihood method is presented. The parametric likelihood is shortly revisited, the empirical likelihood 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 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-based ANOVA test for comparing more than two population trimmed means is presented.

In Chapter 6, the simulation study and data analysis results are presented. 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 research results have been published in three peer-reviewed scientific papers:

- 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 (Indexed in SCOPUS, SCIE (Web of Science))

- 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 (Indexed in SCOPUS, zbMATH)

- 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 (Indexed in SCOPUS, SCIE (Web of Science), zbMATH)

Māra Delesa-Vēliņa proved the asymptotic results, performed the simulation study and data analysis (in cooperation with Luca Greco in [51]), and contributed to the writing and editing of the papers.

Chapter 1

Empirical likelihood method

Empirical likelihood is a nonparametric method of statistical inference that, unlike the parametric maximum likelihood, does not require knowledge about the family of the data distribution. However, there are many similarities between the empirical likelihood and the classical maximum likelihood: both methods allow estimating parameters, constructing likelihood ratio tests and estimating confidence intervals by the test inversion. A nonparametric analogue of Wilks' theorem [58] exists for the empirical likelihood ratio and leads to the same limiting chi-square distribution as in the parametric case. Chapter 1.1 gives the main ideas on the maximum likelihood method.

It is well known that the empirical distribution function is the maximum likelihood estimate of the underlying probability distribution function the sample was taken from. A. B. Owen constructed an empirical likelihood ratio function for distributions and showed in his 1988 paper [28] that it can be used to construct confidence intervals for the sample mean, for the class of M-estimators, and for differentiable statistical functionals. Chapter 1.2 presents the EL method in its simplest case for the mean of independent random vector as introduced in [28].

The information about the parameter of interest $\theta(F)$, associated with the distribution F, is often available in the form of estimating equations. For the one-sample case, a general approach to EL with estimating equations involving smooth estimating functions was developed by J. Qin and J. Lawless [34], and is presented in Chapter 1.4. Y. Qin and L. Zhao [35] generalized the EL method with estimating equations to the two-sample case, and their method is described in Chapter 1.4. It builds upon the results of [34] and also requires smooth estimating functions. This method is essential in developing our new EL-based methods in Chapters 3 and 4.

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) $f(x|\theta)$, $\theta \in \Theta \subset \mathbb{R}^k$. Let $\mathbf{x} = (x_1, \ldots, x_n)$ be the observed sample values. The *likelihood function* is a function of θ defined by

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

The likelihood function provides one of the most popular techniques for deriving estimators, namely, the maximum likelihood estimators (MLE). MLE is the parameter point for which the observed sample is the most likely to occur. A formal definition is given below.

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 $\hat{\theta}(\mathbf{X})$.

If the likelihood function is differentiable in θ_i, the MLE can be found solving

$$\frac{\partial}{\partial \theta_i} L(\theta|\mathbf{x}) = 0, \ i = 1, \dots, k,$$

and finding the global maximum. We list regularity conditions on $f(x|\theta)$ that are essential for some useful properties of the MLE to hold. The regularity conditions relate to the differentiability of $f(x|\theta)$ and the ability to interchange differentiation and integration, and are usually satisfied in most reasonable problems.

Assumption 1.1.1. (Regularity conditions of MLEs) [5, p. 516]

(A1) We observe $X_1, \dots, X_n$, where $X_i \sim f(x|\theta)$ are i.i.d.

(A2) The parameter is identifiable, i.e., if $\theta \neq \theta'$, then $f(x|\theta) \neq f(x|\theta')$.

(A3) The densities $f(x|\theta)$ have common support, and $f(x|\theta)$ is differentiable in θ.

(A4) The parameter space Ω contains an open set ω of which the true parameter value θ_0 is an interior point.

(A5) For every x in sample space $\mathcal{X}$, the density $f(x|\theta)$ is three times differentiable with respect to θ, the third derivative is continuous in θ, and $\int f(x|\theta)dx$ can be differentiated three times under the integral sign.

(A6) For any $\theta_0 \in \Omega$, there exists a positive number c and a function $M(x)$ such that

$$\left| \frac{\partial^3}{\partial \theta^3} \log f(x|\theta) \right| \leq M(x) \ \ \forall x \in \mathcal{X}, \ \ \theta_0 - c < \theta < \theta_0 + c,$$

with $E_{\theta_0}[M(X)] < \infty$.

Properties of MLEs [5, pp. 320, 470, 472]
Let $X_1, \dots, X_n$ be an i.i.d. sample from a population with pdf or pmf $f(x|\theta)$, and let $\hat{\theta}$ be the MLE of θ.

1. *Functional invariance.* For a function $\tau(\theta)$, define the MLE of $\tau(\theta) = \eta$ as a value $\hat{\eta}$ that maximizes
$$L^*(\eta|\mathbf{x}) = \sup_{\theta:\tau(\theta)=\eta} L(\theta|\mathbf{x}).$$
Then for any function $\tau(\theta)$, the MLE of $\tau(\theta)$ is $\tau(\hat{\theta})$.

2. *Consistency.* Let $\tau(\theta)$ be a continuous function of θ. Under the regularity conditions (A1) - (A4) in Assumption 1.1.1 on $f(x|\theta)$, $\tau(\hat{\theta})$ is a *consistent estimator* of $\tau(\theta)$, i.e., for every $\epsilon > 0$ and every $\theta \in \Theta$,
$$\lim_{n\to\infty} P_\theta(|\tau(\hat{\theta}) - \tau(\theta)| \geq \epsilon) = 0.$$

3. *Asymptotic normality and efficiency.* Let $\tau(\theta)$ be a continuous function of θ. Under the regularity conditions (A1) - (A6) in Assumption 1.1.1 on $f(x|\theta)$,

$$\sqrt{n}(\tau(\hat{\theta}) - \tau(\theta)) \xrightarrow{d} N(0, \nu(\theta)) \text{ as } n \to \infty,$$

where $\nu(\theta)$ is the Cramér-Rao Lower Bound

$$\nu(\theta) = \frac{\left(\frac{d}{d\theta}\tau(\theta)\right)^2}{nE_\theta\{(\frac{\partial}{\partial\theta}\log f(x|\theta))^2\}}. \tag{1.2}$$

The quantity in the denominator of (1.2) is called the *Fisher information.*

The maximum likelihood estimators lead to the likelihood ratio method of hypothesis testing. If θ denotes the population parameter, consider the hypotheses $H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta_0^c$, where Θ_0 is some subset of the parameter space Θ and Θ_0^c is its complement.

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

$$\lambda(\mathbf{x}) = \frac{\sup_{\Theta_0} L(\theta|\mathbf{x})}{\sup_{\Theta} L(\theta|\mathbf{x})}. \tag{1.3}$$

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 $0 \leq c \leq 1$.

The connection between the MLEs and LRTs is as follows. If $\hat{\theta}$, an MLE of θ, exists, it is obtained by unrestricted maximization of the likelihood function $L(\theta|\mathbf{x})$ over the entire parameter space. Consider another MLE of θ, call it $\hat{\theta}_0$, which is obtained by assuming the parameter space is Θ_0, that is, maximizing $L(\theta|\mathbf{x})$ over all $\theta \in \Theta_0$. Then the LRT statistic is

$$\lambda(\mathbf{x}) = \frac{L(\hat{\theta}_0|\mathbf{x})}{L(\hat{\theta}|\mathbf{x})}.$$

To evaluate the performance of a hypothesis test, the notions of *power function* and *size* of the test are important. Consider a hypotheses test of $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_0^c$.

Definition 1.1.4. [5, p. 383] Suppose R denotes the rejection region for a hypothesis test. The *power function* of a hypothesis test is the function of θ given by

$$\beta(\theta) = P_\theta(\mathbf{X} \in R).$$

A good test should have a power function close to 1 for values of $\theta \in \Theta_0^c$, and close to 0 for values of $\theta \in \Theta_0$.

Definition 1.1.5. [5, p. 385] For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a *size α test* if

$$\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha.$$

A test is a *level α test* if

$$\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha.$$

It is desirable to find the test with the highest power under H_1 among all size α tests. If such a test exists, it is called the *most powerful test*. In the special case of simple hypothesis $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$, the likelihood ratio test provides the most powerful test.

Theorem 1.1.1. *(Neyman-Pearson lemma) [5, Theorem 8.3.12] Let $X_1, \dots, X_n$ be a random sample from a pdf or pmf $f(x|\theta)$. Consider a test $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$. Let*

$$\lambda(\mathbf{x}) = \frac{f(\mathbf{x}|\theta_0)}{f(\mathbf{x}|\theta_1)} = \frac{L(\theta_0|\mathbf{x})}{L(\theta_1|\mathbf{x})}$$

and $R = \{\mathbf{x} : \lambda(\mathbf{x}) \leq k\}$, where k is a constant such that

$$\alpha = P_{\theta_0}(\mathbf{X} \in R).$$

Suppose we reject H_0 when $\mathbf{x} \in R$, then this test is the most powerful, size α test.

Neyman-Pearson lemma considers simple hypotheses, however, for most practical uses, we are interested in a composite hypothesis. Consider the likelihood ratio test statistic $\lambda(\mathbf{x})$ given by (1.3). Given the data $\mathbf{X} = \mathbf{x}$ is observed, to define a level α test, the constant c must be chosen so that

$$\sup_{\theta \in \Theta_0} P_\theta(\lambda(\mathbf{X}) \leq c) \leq \alpha.$$

The following theorem indicates a general procedure to obtain an approximate value for c in the asymptotic case.

Theorem 1.1.2. *(Asymptotic distribution of LRT, Wilks' theorem) [5, Theorem 10.3.3] Let $X_1, \dots, X_n$ be a random sample from a pdf or pmf $f(x|\theta)$. Under the regularity conditions in Assumption 1.1.1, 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 $p < q$.

Finally, we remark that there is a general equivalence between the hypothesis testing and the interval estimation that allows to construct interval estimates by *test inversion*. The following theorem formalizes this equivalence.

Theorem 1.1.3. *[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

This section presents the basis of empirical likelihood inference through a nonparametric likelihood ratio function as developed by A. B. Owen [28]. We first consider the mean of a scalar random variables and independent random vectors. Denote $F(x-) = P(X < x)$ and so $P(X = x) = F(x) - F(x-)$.

Definition 1.2.1. [31, p. 6] Let $X_1, \ldots, X_n \in \mathbb{R}$. For for $-\infty < x < \infty$, the *empirical cumulative distribution function* (ecdf) of $X_1, \ldots, X_n$ is defined as

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{X_i \leq x}.$$

Definition 1.2.2. [31, p. 6] Let $X_1, \ldots, X_n$ be i.i.d. random variables with unknown cdf F. The *empirical likelihood* $L(F)$ *of the cdf* F is given by

$$L(F) = \prod_{i=1}^{n} (F(X_i) - F(X_i-)) = \prod_{i=1}^{n} p_i, \tag{1.4}$$

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

The value $L(F)$ in (1.4) is the probability of getting exactly the observed sample values $X_1, \ldots, X_n$ from the distribution function F. If F is a continuous distribution, $L(F) = 0$. Thus, to have a positive empirical likelihood, F must belong to a parametric family that is a multinomial distribution on all n observed data values. The empirical likelihood $L(F)$ is maximized by the empirical cumulative distribution function.

Theorem 1.2.1. *[31, Theorem 2.1.] Let $X_1, \ldots, X_n \in \mathbb{R}$ be i.i.d. random variables with a cdf F_0. Let F_n be their ecdf and let F be any distribution function. If $F \neq F_n$, then $L(F) < L(F_n)$.*

Thus it can be said that $F_n(x)$ is a "nonparametric maximum likelihood estimator" of $F(x)$, and, similarly to the parametric likelihood, it is possible to define a *nonparametric likelihood ratio* that can be used a a basis for constructing nonparametric hypothesis tests and confidence intervals.

Definition 1.2.3. The *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. To evaluate the empirical likelihood of θ, we search for the optimal distribution F and choose a value of θ that maximises the likelihood function. Thus the maximisation is carried out both over $p = (p_1, \ldots, p_n)$ and θ. Usually, this is done by the *profiling*, i.e., for a fixed value of θ, optimal weights p are found, thus $p = p(\theta)$.

Definition 1.2.4. The *profile empirical likelihood ratio function* is given by

$$\mathcal{R}(\theta) = \sup\{R(F) | T(F) = \theta, F \in \mathcal{F}\}. \tag{1.5}$$

For the true unknown parameter $\theta_0 = T(F_0)$, empirical likelihood hypothesis test rejects $H_0 : T(F_0) = \theta_0$ when $\mathcal{R}(\theta_0) < r_0$ for some threshold r_0, and 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 are looking for an optimal $F = (p_1, \ldots, p_n)$, 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

$$\mathcal{R}(\mu) = \sup_p \left\{ \prod_{i=1}^n np_i \,|\, \sum_{i=1}^n p_i X_i = \mu^*, p_i \geq 0, \sum_{i=1}^n p_i = 1 \right\}. \tag{1.6}$$

The next theorem provides the basis for the calibration of the confidence intervals for the mean and can be regarded as a nonparametric analogue to Wilks' Theorem 1.1.2.

Theorem 1.2.2. *[31, 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*

$$-2 \log \mathcal{R}(\mu_0) \xrightarrow{d} \chi_1^2 \text{ as } n \to \infty.$$

Two comments can be given regarding Theorem 1.2.2. First, the asymptotic limit of the empirical likelihood is the same χ_1^2 as in the nonparametric case. And second, it is not required for the X_i to be bounded; they only require to have bounded variance, which constrains how fast the sample minimum and sample maximum grow as n increases.

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

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

where $\chi_{1,1-\alpha}^2$ denotes the $1-\alpha$ quantile of χ_1^2 distribution. The interval C_α is an asymptotic coverage interval, i.e.,

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

Now, consider independent d-dimensional random vectors X_i, $d \geq 1$, assuming common distribution F_0. It is convenient to describe distributions by probabilities they attach on sets. Let $F(A)$ denote $P(X \in A)$ for $X \sim F$ and $A \subseteq \mathbb{R}^d$, and let δ_x denote the distribution under which $X = x$ with probability 1. Thus $\delta_x(A) = I_{x \in A}$.

Definition 1.2.5. Let $X_1, \ldots X_n \in \mathbb{R}^d$. The *empirical distribution function of* $X_1, \ldots X_n$ is defined by

$$F_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}. \tag{1.7}$$

Definition 1.2.6. Given $X_1, \dots X_n \in \mathbb{R}^d$ with common distribution function F_0, the *nonparametric likelihood of* F is

$$L(F) = \prod_{i=1}^{n} F(\{X_i\}).$$

Here $F(\{X_i\})$ is the probability of observing a value X_i in a sample from F. Like in the univariate case, the nonparametric likelihood $L(F)$ is maximized by the empirical distribution function F_n.

Now, for the multivariate mean, the empirical likelihood ratio is the same as in (1.6), except that it is now defined on $\mathbb{R}^d$. The confidence region is in form

$$C_{r,n} = \left\{ \sum_{i=1}^{n} p_i X_i \,|\, \prod_{i=1}^{n} np_i > r, p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \right\},$$

and is a subset in $\mathbb{R}^d$. The univariate EL Theorem 1.2.2 generalizes to a vector case.

Theorem 1.2.3. *[31, Theorem 3.2.] Let $X_1, \dots X_n \in \mathbb{R}^d$ be independent random vectors with common distribution F_0 having mean μ_0 and variance-covariance matrix V_0 of rank $q > 0$. Then $C_{r,n}$ is a convex set and*

$$-2 \log \mathcal{R}(\mu_0) \xrightarrow{d} \chi^2_q \text{ as } n \to \infty.$$

1.3 Empirical likelihood for general estimating equations

Consider $X_1, \dots, X_n \in \mathbb{R}^d$ i.i.d. random variables with unknown distribution F, and a p-dimensional parameter θ associated with F. Assume that the information about θ and F is available in the form of $r \geq p$ functionally independent unbiased estimating functions $g_j(X, \theta)$, $j = 1, \dots, r$, where

$$E_F\{g_j(X, \theta)\} = 0.$$

In vector form,

$$g(X, \theta) = (g_1(X, \theta), \dots, g_r(X, \theta))^T,$$

where $E_F\{g(X, \theta)\} = 0$.

In the case θ is the mean, the estimating equation is in the form $g(X, \theta) = X - \theta$. In the case θ is the quantile $\theta_q = F^{-1}(q)$, the estimating equation is in the form $g(X, \theta_q) = I_{X \leq \theta_q} - q$. J. Qin and J. Lawless [34] demonstrated how to use the estimating equations to estimate F and θ under empirical likelihood setting. When $r = p$, the method in [34] is the same as developed by A. B. Owen in [28, 29].

Maximize the likelihood $L(F) = \prod_{i=1}^{n} p_i$ subject to restrictions

$$p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i g(X_i, \theta) = 0. \tag{1.8}$$

For a fixed θ, a unique maximum exists, if θ is inside the convex hull of the points $g(X_1, \theta), g(X_2, \theta), \dots, g(X_n, \theta)$ [34], and the maximum may be found with the help of Lagrange multipliers method. Define

$$H = \sum_{i=1}^{n} \log p_i + \lambda_0 \left(1 - \sum_{i=1}^{n} p_i\right) - n\lambda^T \sum_{i=1}^{n} p_i g(X_i, \theta), \tag{1.9}$$

where $\lambda = (\lambda_1, \ldots, \lambda_r)^T$ and λ_0 are Lagrange multipliers. Taking derivatives of (1.9) with respect to p_i,

$$\frac{\partial}{\partial p_i} H = \frac{1}{p_i} - \lambda_0 - n\lambda^T g(X_i, \theta) = 0,$$

$$\sum_{i=1}^{n} p_i \frac{\partial}{\partial p_i} H = n - \lambda_0 = 0 \Rightarrow \lambda_0 = n,$$

and

$$p_i = \frac{1}{n} \frac{1}{1 + \lambda^T g(X_i, \theta)}.$$

Using the last restriction from (1.8), we get

$$0 = \sum_{i=1}^{n} p_i g(X_i, \theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{g(X_i, \theta)}{1 + \lambda^T g(X_i, \theta)}, \tag{1.10}$$

thus λ can be expressed in terms of θ as follows: from the restriction $0 \leq p_i \leq 1$ observe that $1+\lambda^T g(X_i, \theta) \geq 1/n$ for all i. For a fixed θ, consider $D_\theta = \{\lambda : 1+\lambda^T g(X_i, \theta) \geq 1/n\}$. D_θ is convex and closed, and bounded if 0 is inside the convex hull of $g(X_i, \theta)$'s. Moreover, the derivative of (1.10) with respect to λ,

$$\frac{\partial}{\partial \lambda} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{g(X_i, \theta)}{1 + \lambda^T g(X_i, \theta)} \right\} = -\frac{1}{n} \sum_{i=1}^{n} \frac{g(X_i, \theta) g^T(X_i, \theta)}{(1 + \lambda^T g(X_i, \theta))^2},$$

is negative definite for λ in D_θ provided that $\sum_{i=1}^{n} g(X_i, \theta) g^T(X_i, \theta)$ is positive definite. By the inverse function theorem, $\lambda = \lambda(\theta)$ is a continuous differentiable function in θ.

Empirical likelihood function for a parameter θ is defined as

$$L(\theta) = \prod_{i=1}^{n} \left\{ \left(\frac{1}{n} \right) \frac{1}{1 + \lambda^T(\theta) g(X_i, \theta)} \right\}$$

and the empirical likelihood ratio is

$$R(\theta) = \prod_{i=1}^{n} n p_i = \prod_{i=1}^{n} \frac{1}{1 + \lambda^T g(X_i, \theta)}.$$

It is convenient to work with the empirical log likelihood ratio

$$l(\theta) = -\log R(\theta) = \sum_{i=1}^{n} \log \left(1 + \lambda^T(\theta) g(X_i, \theta) \right). \tag{1.11}$$

Minimizing $l(\theta)$, we can obtain the empirical likelihood estimator for θ,

$$\tilde{\theta} = \arg\min_{\theta} l(\theta).$$

The next Lemma and Theorem provides regularity conditions for the EL inference in the setting with general estimating functions.

Lemma 1.3.1. *[34, Lemma 1] Assume that $E\{g(X_i, \theta) g^T(X_i, \theta)\}$ is positive definite, $(\partial/\partial\theta) g(x, \theta)$ is continuous in the neighbourhood of the true value θ_0, $||(\partial/\partial\theta) g(x, \theta)||$ and $||g(x, \theta)^3||$ are bounded by some integrable function $G(x)$ in this neighbourhood, and the rank of $E\{(\partial/\partial\theta) g(x, \theta_0)\}$ is p. Then, as $n \to \infty$, with probability 1 $l(\theta)$ attains its maximum at some point $\tilde{\theta}$ in the interior of the ball $||\theta - \theta_0|| \leq n^{-1/3}$.*

Theorem 1.3.1. *[34, Theorem 2] In addition to the conditions of Lemma 1.3.1, assume that $(\partial^2/\partial\theta\partial\theta^T)g(x,\theta)$ is continuous in θ in the neighbourhood of the true parameter θ_0, and $||(\partial^2/\partial\theta\partial\theta^T)g(x,\theta)||$ can be bounded by some integrable function $G(x)$ in this neighbourhood. Then the empirical likelihood ratio statistic for testing $H_0 : \theta = \theta_0$*

$$\mathcal{W}(\theta_0) = 2l(\theta_0) - 2l(\tilde{\theta}) \xrightarrow{d} \chi_p^2,$$

as $n \to \infty$ when $H_0 : \theta = \theta_0$ is true, where $l(\theta)$ is given by (1.11).

Theorem 1.3.1 allows to construct empirical likelihood confidence interval for parameter θ. Choose r_0 such that

$$P(\chi_p^2 \leq r_0) = 1 - \alpha.$$

Then the empirical likelihood $1 - \alpha$ confidence interval is in the form

$$C_\alpha = \{\theta | \mathcal{W}(\theta) \leq r_0\},$$

and due to Theorem 1.3.1, its asymptotic coverage ratio is

$$P(\theta_0 \in C_\alpha) = P\{\mathcal{W}(\theta_0) \leq r_0\} = 1 - \alpha.$$

1.4 Empirical likelihood method in general two-sample case

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, $\Delta_0 = \theta_1 - \theta_0$. It was shown by Y. Qin and L. Zhao [35] that under certain regularity conditions, the EL test statistic for the difference Δ_0 of two univariate parameters has the asymptotic χ_1^2 distribution.

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

$$E_{F_1} w_1(X, \theta_0, \Delta_0) = 0, \quad (1.12)$$
$$E_{F_2} w_2(Y, \theta_0, \Delta_0) = 0, \quad (1.13)$$

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

Example 1.4.1. 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 1.4.2. The difference of distribution functions. For a fixed point 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

$$L(F_1, F_2) = \prod_{i=1}^{n_1}(F_1(X_i) - F_1(X_i-))\prod_{j=1}^{n_2}(F_2(Y_j) - F_2(Y_j-)) = \prod_{i=1}^{n_1} p_i \prod_{j=1}^{n_2} q_j, \qquad (1.14)$$

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 empirical likelihood ratio is defined as

$$R(F_1, F_2) = \frac{L(F_1, F_2)}{L(F_{n_1}, F_{n_2})} = \prod_{i=1}^{n_1} n_1 p_i \prod_{j=1}^{n_2} n_2 q_j,$$

and the profile empirical likelihood ratio is in the form

$$\mathcal{R}(\Delta, \theta) = \sup_{p,q}\left\{\prod_{i=1}^{n_1} n_1 p_i \prod_{j=1}^{n_2} n_2 q_j \,|\, \sum_{i=1}^{n_1} p_i w_i(X_i, \theta, \Delta) = 0, \sum_{j=1}^{n_2} q_j w_2(Y_j, \theta, \Delta) = 0\right\}, \qquad (1.15)$$

where $p_i \geq 0$, $\sum_{i=1}^{n_1} p_i = 1$, $q_j \geq 0$ and $\sum_{i=1}^{n_2} q_j = 1$. Thus we first fix Δ and θ, and solve for p_i, q_j, using Lagrange multipliers. Consider

$$\begin{aligned} H(\Delta, \theta) = {} & \sum_{i=1}^{n_1} \log p_i + \sum_{j=1}^{n_2} \log q_j + t_1(\Delta, \theta)(1 - \sum_{i=1}^{n_1} p_i) + t_2(\Delta, \theta)(1 - \sum_{j=1}^{n_2} q_j) \qquad (1.16) \\ & -n_1\lambda_1(\Delta, \theta)\sum_{i=1}^{n_1} p_i w_1(X_i, \theta, \Delta) - n_2\lambda_2(\Delta, \theta)\sum_{j=1}^{n_2} q_j w_2(Y_j, \theta, \Delta), \end{aligned}$$

where $t_1(\Delta, \theta)$, $t_2(\Delta, \theta)$, $\lambda_1(\Delta, \theta)$, $\lambda_2(\Delta, \theta)$ are Lagrange multipliers. Calculating the partial derivatives of (1.16) with respect to p_i's, we have

$$\frac{\partial}{\partial p_i} H = \frac{1}{p_i} - t_1(\Delta, \theta) - n_1\lambda_1(\Delta, \theta) w_1(X_i, \Delta, \theta) = 0,$$

thus

$$\sum_{i=1}^{n_1} p_i \frac{\partial}{\partial p_i} H = n_1 - t_1(\Delta, \theta) = 0 \Rightarrow t_1(\Delta, \theta) = n_1,$$

and

$$p_i = \frac{1}{n_1(1 + \lambda_1(\Delta, \theta) w_1(X_i, \theta, \Delta))}, \; i = 1, \ldots, n_1. \qquad (1.17)$$

Similarly, solving for $(\partial/\partial q_j)H = 0$, we obtain

$$q_j = \frac{1}{n_2(1 + \lambda_2(\Delta, \theta) w_2(Y_j, \theta, \Delta))}, \; j = 1, \ldots, n_2. \qquad (1.18)$$

Using the restrictions on the estimating functions, we obtain

$$\sum_{i=1}^{n_1} \frac{w_1(X_i, \theta, \Delta)}{1 + \lambda_1(\Delta, \theta) w_1(X_i, \theta, \Delta)} = 0, \qquad (1.19)$$

$$\sum_{j=1}^{n_2} \frac{w_2(Y_j, \theta, \Delta)}{1 + \lambda_2(\Delta, \theta) w_2(Y_j, \theta, \Delta)} = 0, \qquad (1.20)$$

from which the Lagrange multipliers λ_1 and λ_2 can be expressed in terms of θ similarly as in the one-sample case (1.10). Inserting p_i and q_j from (1.17) - (1.18) in (1.15) and taking the logarithm, we obtain the empirical log likelihood ratio function as

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

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

$$\sum_{i=1}^{n_1} \frac{\lambda_1(\Delta, \theta)\alpha_1(X_i, \theta, \Delta)}{1 + \lambda_1(\Delta, \theta) w_1(X_i, \theta, \Delta)} + \sum_{j=1}^{n_2} \frac{\lambda_2(\Delta, \theta)\alpha_2(Y_j, \theta, \Delta)}{1 + \lambda_2(\Delta, \theta) w_2(Y_j, \theta, \Delta)} = 0, \tag{1.21}$$

where $\alpha_1 = \partial w_1/\partial\theta$ and $\alpha_2 = \partial w_2/\partial\theta$. The sufficient conditions for the existence of the solution to (1.21) are given by Assumption 1.4.1.

Assumption 1.4.1. [35, 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 neighbourhood 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 neighbourhood, $\alpha_2(Y, \theta, \Delta)$ and $w_2^3(Y, \theta, \Delta)$ are bounded by some integrable function $G_2(Y)$ in this neighbourhood, 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 1.4.1. *[35, Theroem 1] Under Assumption 1.4.1, there exists a root $\hat{\theta}(\Delta)$ of (1.21) such that $\hat{\theta}(\Delta)$ is a consistent estimator 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 [35]. The confidence intervals for the true parameter Δ_0 can be obtained by test inversion and have the form $\{\Delta \,|\, \mathcal{R}(\Delta, \hat{\theta}(\Delta)) > c\}$, where the constant c can be calibrated using Theorem 1.4.1.

Chapter 2

Robust estimation of a location parameter

In this chapter we consider the ideas of robust estimation theory that will be of importance in developing our new robust EL-based methods. In Chapter 2.1 we consider the M-estimators proposed by P. J. Huber [17]. We introduce the model of location and define the scale-equivariant M-estimators for the location parameter. We also report the conditions for the asymptotic normality of a general M-estimator. Finally, we introduce the smoothing principle for the ψ-function of a general M-estimator [13] and give a definition of the smoothed Huber estimator. The smoothing principle is essential to this research, since our new EL-based method for the difference of two M-estimators is applicable to the smooth Huber estimator, but not to the non-smooth version of it.

In Chapter 2.2 we consider another important class of robust estimators, L-estimators, and in particular the trimmed mean. The trimmed mean is an appealing robust location estimator that is obtained by trimming a fixed proportion of the extreme data values. We report the theorem on the asymptotic distribution of the trimmed mean by S. Stigler [39]. The asymptotic distribution of the trimmed mean is more complicated than that of an M-estimator; it is necessary and sufficient that the trimming be done at uniquely defined percentiles of the population distribution to establish its asymptotic normality. This theorem is important to prove the asymptotic results for our new two-sample EL method for the trimmed means.

In Chapter 2.3 we introduce an important property of the estimator used to quantify its robustness, the finite-sample breakdown point (FBP). Informally, the breakdown point of an estimator $\hat{\theta}$ of parameter θ is the largest proportion of atypical points that the data may contain such that $\hat{\theta}$ still measures θ, i.e, the information related to the typical data points. The breakdown point can be defined for the finite-sample and the asymptotic case, but here only the latter will be considered. The definition of a FBP for the length of a confidence interval will be given.

2.1 M-estimators

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables from sample space $\mathcal{X} \subseteq \mathbb{R}$ with a common distribution F_θ, where the unknown parameter θ belongs to some parameter space Θ. In classical statistics, it is assumed that X_i are distributed exactly as F_θ. For example, $\mathcal{X} = \mathbb{R}$, $\Theta = \mathbb{R}$ and F_θ is normally distributed random variable with mean θ and variance equal to one; $\mathcal{X} = [0, \infty]$, $\Theta = (0, \infty)$ and F_θ is exponential distribution with expectation

θ. Consider an MLE estimator $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ of the parameter θ. It follows from (1.1) that $\hat{\theta}$ is the value that minimizes the minus of the log likelihood function, i.e.,

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{n} -\log f_\theta(X_i).$$

If F_θ is known exactly, the MLE is 'optimal' in the sense that it attains the variance lower bound as described in Chapter (1.1). In robust statistics, it is assumed that we know F_θ only 'approximately', and we look for estimators that are 'optimal' at F_θ (e.g., normal distribution) and 'nearly optimal' in the neighbourhood of F_θ (e.g., contaminated normal distribution).

With this goal in mind, Huber [17] proposed to generalize the concept of the maximum likelihood estimation by replacing the function $-\log f_\theta$ with a general *ρ-function*, i.e.,

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{n} \rho(X_i, \theta), \tag{2.1}$$

where ρ is a function on $\mathcal{X} \times \Theta$. Suppose that ρ has a derivative $\psi(x, \theta) = (\partial/\partial\theta)\rho(x, \theta)$, then (2.1) is equivalent to

$$\sum_{i=1}^{n} \psi(X_i, \theta) = 0. \tag{2.2}$$

Definition 2.1.1. [14, p. 101] An estimator defined in the form (2.1) or (2.2) is called *M-estimator*. If G_n is the ecdf generated by the sample, then the solution $\hat{\theta}$ of (2.2) can also be written as $T(G_n)$, where T is the functional given by

$$\int \psi(x, T(G))dG(x) = 0 \tag{2.3}$$

for all distributions G where the integral is defined.

2.1.1 M-estimators of location

In this thesis, we are mainly interested in the estimation of the *location parameter* of a distribution F, the definition of which is given below.

Definition 2.1.2. [26, p. 17] Let $X_1, \ldots X_n$ be i.i.d. random variables with distribution function F that depend on an unknown parameter θ through the model

$$X_i = \theta + u_i, \; i = 1, \ldots, n, \tag{2.4}$$

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 (2.4) is called *the location model*, and θ is referred to as the *location parameter*.

Definition 2.1.3. [26, p. 25] Consider the location model (2.4). Given a function ρ, an *M-estimator of location parameter* θ is defined as

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{n} \rho(X_i - \theta). \tag{2.5}$$

If ρ is differentiable with respect to θ, then $\hat{\theta}$ is the solution to the equation

$$\sum_{i=1}^{n} \psi(X_i - \theta) = 0, \tag{2.6}$$

where $\psi(x, \theta) = (\partial/\partial\theta)\rho(x, \theta)$.

Note that choosing $\rho(x, \theta) = -\log f_\theta(x - \theta)$ and $\psi(x, \theta) = -(\partial/\partial\theta) \log f_\theta(x - \theta)$ in (2.5) and (2.6), respectively, we obtain an MLE of a location parameter θ of f_θ.

Example 2.1.1. *MLE of the standard normal.* If F_θ is $N(0, 1)$, apart from a constant, $\rho(x, \theta) = (x - \theta)^2/2$ and $\psi(x, \theta) = x - \theta$. From (2.3) it follows that $\theta = EX$, the mean.

Example 2.1.2. *MLE of exponential distribution.* The density is $f_\theta(x) = 1/2 \exp(-|x|)$, and we have $\rho(x, \theta) = |x - \theta|$. The derivative of ρ exists for $x \neq 0$ and is given by the sign function,

$$\psi(x, \theta) = \operatorname{sgn}(x - \theta).$$

Note that $\operatorname{sgn}(x) = I_{x>0} - I_{x<0}$, thus (2.5) yields

$$\begin{aligned} \sum_{i=1}^{n} \operatorname{sgn}(X_i - \theta) &= \sum_{i=1}^{n} (I_{X_i>\theta} - I_{X_i<\theta}) = \\ &= \#(X_i > \theta) - \#(X_i < \theta) = 0, \end{aligned}$$

and we have $\#(X_i > \theta) = \#(X_i < \theta)$, thus θ is the median.

Example 2.1.3. *Huber estimator.* For a given positive constant k, the Huber estimator is defined by (2.5) or (2.6) with

$$\rho = \rho_k(x) = \begin{cases} 2kx - k^2, & x > k \\ x^2, & -k \leq x \leq k \\ -2kx - k^2, & x < -k \end{cases} \tag{2.7}$$

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

$$\psi = \psi_k(x) = \begin{cases} k, & x > k \\ x, & -k \leq x \leq k \\ -k, & x < -k. \end{cases} \tag{2.8}$$

Huber [17] 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 any symmetric distribution. The Huber estimator is the MLE for the so-called Huber's least favourable distribution given by the density

$$f_k(x) = \begin{cases} (1 - \epsilon)\phi(k) \exp(-k(x - k)), & x > k \\ (1 - \epsilon)\phi(x), & -k \leq x \leq k \\ (1 - \epsilon)\phi(k) \exp(k(x + k)), & x < -k, \end{cases} \tag{2.9}$$

where k and ϵ are related through the formula

$$2\phi(k)/k - 2\Phi(-k) = \epsilon/(1 - \epsilon), \tag{2.10}$$

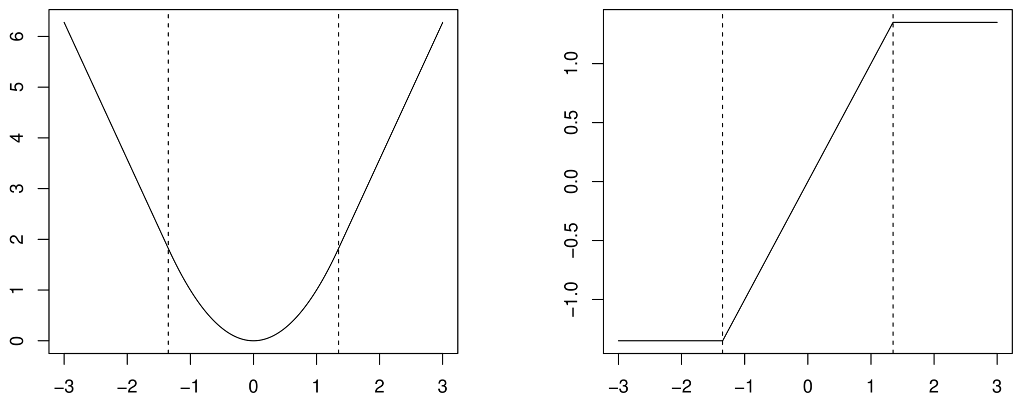


Figure 2.1: ρ-function (2.7) (left) and ψ-function (2.8) (right) of the Huber M-estimator with $k = 1.35$, where $x = k$ is represented by the dotted line.

and ϕ denotes the pdf of standard normal distribution. Huber estimator can be interpreted as an intermediary estimator between the mean and the median. For $k \to 0$, one obtains the sample median, while $k \to \infty$ leads to the sample mean as the limiting cases. Based on extensive simulation results, it was concluded in [17] 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 $k = 1.35$ as recommended in [26] and [31], which corresponds to a 95% efficiency of the Huber estimator compared to the sample mean at the standard normal distribution. In simulation settings when ϵ is known, one may choose the k that satisfies (2.10). $\psi_k(x)$ and $\rho_k(x)$ defining the Huber estimator with $k = 1.35$ are plotted in Figure 2.1.

Example 2.1.4. *Bisquare estimator* is defined by (2.5) and (2.6) with

$$\rho(x) = \begin{cases} 1 - \left[1 - (x/k)^2\right]^3, & |x| < k \\ 1, & |x| \geq k, \end{cases} \tag{2.11}$$

with derivative $\rho'(x) = 6\psi(x)/k^2$, where

$$\psi(x) = x \left[1 - \left(\frac{x}{k}\right)^2\right]^2 I_{|x| \leq k}. \tag{2.12}$$

One may choose $k = 4.68$ which corresponds to 95% efficiency of the bisquare estimator at the standard normal distribution. $\psi(x)$ and $\rho(x)$ defining the bi-square estimator with $k = 4.68$ are plotted in Figure 2.1. The ρ-function of the bi-square estimator is bounded. The ψ-function is not monotone, it is everywhere differentiable, and is zero outside interval $[-k, k]$. Note that the bi-square estimator is not an MLE of any distribution.

The kind of ψ functions that tend to zero at infinity are referred to as "redescending", and the related solutions of (2.6) are called "redescending M-estimators". They can provide increased robustness against heavy tails and large outliers [26].

Remark 2.1.1. Existence of a solution. *Assume ψ is monotone nondecreasing with $\psi(-\infty) < 0 < \psi(\infty)$. Then (2.6) and hence (2.5) always has a solution [26, Theorem*

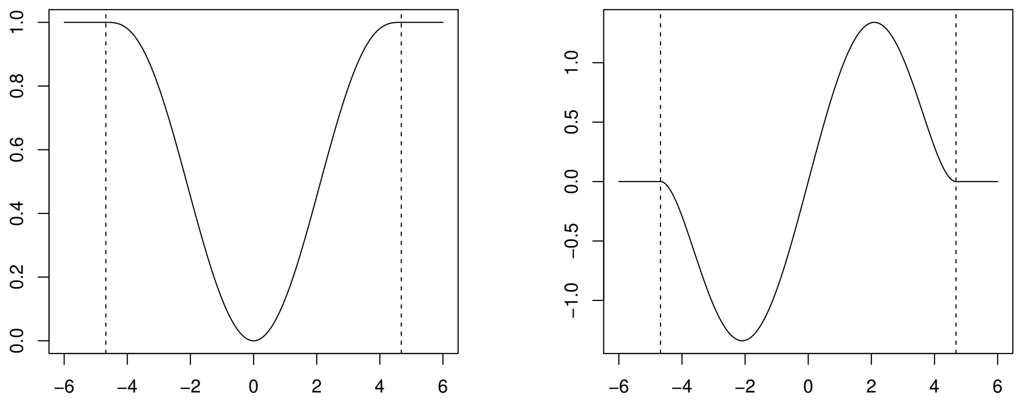


Figure 2.2: ρ-function (2.11) and ψ-function (2.12) of the bisquare M-estimator with $k = 4.68$, where $x = k$ is represented by the dotted line.

10.1]. If ψ is continuous and increasing, the solution is unique, otherwise the set of solutions is a point or an interval.

If ψ is redescending, some solutions of (2.6) may not correspond to the absolute minimum criterion of (2.5). The uniqueness of the asymptotic value of a redescending M-estimator requires that the density $f(x)$ of X_i's is symmetric and unimodal; in other words, $f(x)$ is a decreasing function of $|x|$ [26, Theorem 10.2].

Finally we list the properties of the ρ- and ψ-functions that cover the most cases of interest in the estimation.

Definition 2.1.4. [26, p. 31] A ρ-function will denote a function ρ such that

(R1) $\rho(x)$ is a nondecreasing function of $|x|$.

(R2) $\rho(0) = 0$.

(R3) $\rho(x)$ is increasing for $x > 0$ such that $\rho(x) < \rho(\infty)$.

(R4) if ρ is bounded, it is also assumed that $\rho(\infty) = 1$.

A ψ-function will denote a function ψ which is a derivative of a ρ-function. In particular, it implies

(P1) ψ is odd and $\psi(x) \geq 0$ for $x \geq 0$.

2.1.2 Scale equivariant M-estimators of location

Next we discuss some useful properties of estimators, the *shift and scale equivariance*, as well as *shift invariance*.

Definition 2.1.5. [26, pp. 18, 35] An estimator $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ is called *shift equivariant* if for any constant $c \neq 0$

$$\hat{\theta}(X_1 + c, X_2 + c, \ldots, X_n + c) = \hat{\theta}(X_1, X_2, \ldots, X_n) + c.$$

$\hat{\theta}$ is called *scale equivariant* if

$$\hat{\theta}(cX_1, cX_2, \ldots, cX_n) = |c|\hat{\theta}(X_1, X_2, \ldots, X_n).$$

$\hat{\theta}$ is called *shift invariant* if

$$\hat{\theta}(X_1 + c, X_2 + c, \ldots, X_n + c) = \hat{\theta}(X_1, X_2, \ldots, X_n).$$

Any statistic satisfying the shift invariance and scale equivariance will be called a *dispersion estimator*.

Any location M-estimator $\hat{\theta}$ given by (2.5) or (2.6) is shift equivariant (see, for example, [26]), 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. To obtain scale equivariant location M-estimator, using an auxiliary dispersion estimator is necessary.

Definition 2.1.6. [26, 38] *A scale equivariant M-estimator* $\hat{\theta}$ for the location parameter θ with a previous estimation of dispersion is defined as the solution to the equation

$$\sum_{i=1}^{n} \psi \left(\frac{X_i - \hat{\theta}}{\hat{\sigma}} \right) = 0, \tag{2.13}$$

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

Intuitively, $\hat{\sigma}$ in (2.13) should be robust itself. The classical dispersion estimator, the standard deviation, is not robust. A popular choice for $\hat{\sigma}$ is the *normalized median absolute deviation about the median (MADN).*

Definition 2.1.7. [26, p. 36] The *median absolute deviation about the median (MAD)* is defined by

$$\mathrm{MAD}(X) = \mathrm{MAD}(X_1, \ldots, X_n) = \mathrm{Med}\{|X - \mathrm{Med}(X)|\}, \tag{2.14}$$

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.

Finally, we note that there exists an alternative approach to consider a *location-dispersion* model with two unknown parameters μ and σ that allows constructing simultaneous M-estimators of location and dispersion. However, estimation with a previously computed dispersion estimator is more robust than simultaneous estimation [26, Chapter 2.7.2], thus the simultaneous estimation is not considered in this work.

2.1.3 Asymptotic distribution of M-estimators of location

Except for the mean and the median, there are no explicit expressions of the distributions of M-estimators for finite sample sizes. However, asymptotic approximations of the distribution of a general M-estimator can be established. The next two theorems provide the conditions for the asymptotic normality of the M-estimator of location and M-estimator of location with a preliminary dispersion estimator. In this section, let $X_1, \ldots, X_n$ be i.i.d. random variables with distribution F.

Theorem 2.1.1. *[26, Theorem 10.7] Consider an M-estimator of location parameter θ defined by the equation (2.6) with $\theta \in \mathbb{R}$. Define $\lambda_F(\theta) = \mathrm{E}_F\psi(X_i - \theta)$. Assume that $A = \mathrm{E}_F\psi^2(X_i - \theta) < \infty$ and $B = \lambda'(\theta_F)$ exists and is non-null. Let $\hat{\theta}_n$ be the solution of (2.6) such that $\hat{\theta}_n \xrightarrow{p} \theta_F$. Then the distribution of $\sqrt{n}(\hat{\theta}_n - \theta_F)$ tends to $N(0, \nu)$ as $n \to \infty$ with*

$$\nu = A/B^2.$$

If $\psi'(x - \theta) = (\partial/\partial\theta)\psi(x - \theta)$ exists and verifies for all x, θ

$$|\psi'(x - \theta)| \leq K(x) \text{ with } \mathrm{E}K(x) < \infty,$$

then $B = \mathrm{E}\psi'(X_i - \theta_F)$.

Remark 2.1.2. *For the mean, the existence of A requires the existence of $\mathrm{E}X^2$. In general, if ψ is bounded, A always exists. If ψ' exists, then $\lambda'(\theta) = -\mathrm{E}\psi'(x - \theta)$. If $\lambda'(\theta_F)$ does not exist, $\hat{\theta}_n$ tends to θ_F faster than with the rate $n^{-1/2}$, and there is no asymptotic normality.*

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 scale) [26, 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 2.1.2. *[26, 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

$$a = \mathrm{E}\psi^2\left(\frac{u_i}{\sigma}\right), \; b = \mathrm{E}\psi'\left(\frac{u_i}{\sigma}\right), \; c = \mathrm{E}\psi\left(\frac{u_i}{\sigma}\right)\psi'\left(\frac{u_i}{\sigma}\right). \tag{2.15}$$

Assumption 2.1.2. (Asymptotic normality of an M-estimator with preliminary scale) [26, p. 385]

(A1) Quantities defined in (2.15) exist and $b \neq 0$.

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

(A3) $c = 0$.

Theorem 2.1.3. *[26, Theorem 10.13] Under Assumption 2.1.2,*

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

2.1.4 Smoothed M-estimators

F. Hampel et al. [13] provided a smoothing principle for a ψ-function of an M-estimator, where the degree of smoothness depends on the sample size n, so that the ψ-function is much smoother for small sample n, but asymptotically equivalent. The smoothing principle can be applied to any M-estimator, even if ψ is already smooth.

Definition 2.1.8. [13, p. 325] Consider i.i.d. random variables $X_1, \dots, X_n$ with common 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

$$\tilde{\psi}(x) = \int \psi(x+u) dQ_n(u), \tag{2.16}$$

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

$$\sum_{i=1}^{n} \tilde{\psi}\left(\frac{X_i - t}{\sigma}\right) = 0. \tag{2.17}$$

Remark 2.1.3. *The distribution $Q_n(u)$ of the M-estimators for finite sample sizes cannot be expressed explicitly except for the mean and the median case. However, due to the asymptotic normality of M-estimators in Theorem 2.1.3, 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 the Huber estimator, it is Huber's least favourable distribution with the density f_k from (2.9).*

Proposition 2.1.1. *[13, p. 326] Taking f_k (2.9) as the density of Q_n in (2.16), the $\tilde{\psi}$-function defining the smoothed Huber estimator can be expressed in the explicit form as*

$$\tilde{\psi}_k(x) = k\Phi\left(\frac{x-k}{\sigma_n}\right) - k\left(1 - \Phi\left(\frac{x+k}{\sigma_n}\right)\right) + x\left(\Phi\left(\frac{x+k}{\sigma_n}\right) - \Phi\left(\frac{x-k}{\sigma_n}\right)\right) + \sigma_n\left(\phi\left(\frac{x+k}{\sigma_n}\right) - \phi\left(\frac{x-k}{\sigma_n}\right)\right), \tag{2.18}$$

where $\sigma_n = \sqrt{V/n}$, and k is the tuning constant defining the non-smoothed Huber estimator (2.8).

Simulation study in [13] considered several symmetric distributions both with fixed and unknown scale parameter σ of (2.17): the standard normal distribution, Huber's least favourable distribution with $k = 0.862$, the double exponential distribution, and the Cauchy distribution. For scenarios with an unknown scale parameter, MAD was used as a preliminary dispersion estimator for σ, and the asymptotic variance was set to $V = 2.046$. The value $V = 2.046$ is motivated by the equation (2.10): it is the value of the asymptotic variance of Huber estimator under the Huber's least favourable distribution with $k = 0.862$, that represents $\epsilon = 0.2$ contamination level to the normal distribution (see, for example, Table 1 in [17]). The graph of the $\tilde{\psi}_k$ is depicted in Figure 2.3.

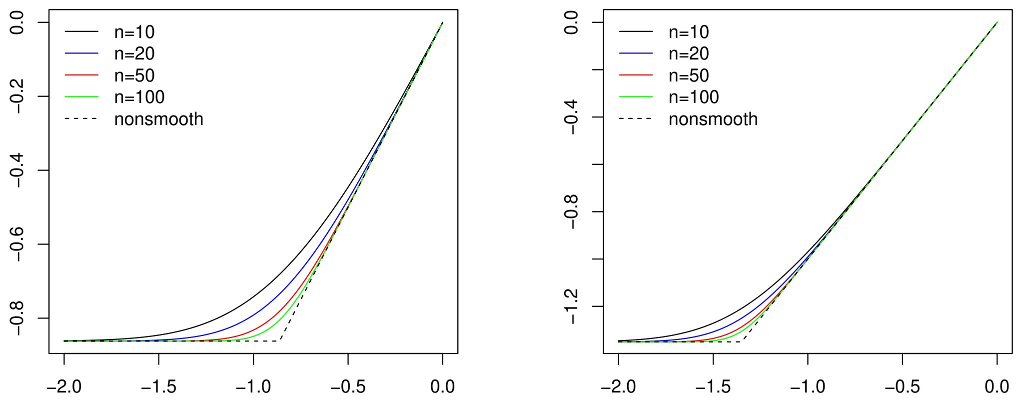


Figure 2.3: ψ-function of the smoothed Huber estimator, $\tilde{\psi}_k$ (2.18). Smoothness depends on the sample size n. Note, that $\tilde{\psi}_k$ is odd; only the portion for x negative is depicted. Left: $\epsilon = 0.2$, $k = 0.862$ and $V = 2.046$. Right: $\epsilon = 0.05$, $k = 1.35$ and $V = 1.256$.

A small-sample simulation study in [13] demonstrated that the smoothed M-estimators performed better than their non-smooth counterparts (in terms of MSE distribution) in all the settings, especially in the tail area of the MSE distribution, the comparative gain in the efficiency being up to 10%.

2.2 Trimmed mean

Definition 2.2.1. [33, 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

$$\bar{X}_{\alpha\beta} = \frac{1}{m} \sum_{i=r}^{s} X_{(i)}, \tag{2.19}$$

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 trimmed mean belongs to the class of the so called *L-statistics* that stands for *linear functions of order statistics*. For several distributions, L-statistics provide good estimators of location and scale parameters.

Definition 2.2.2. [1, p. 227] Suppose $a_{i,n}$ is a double sequence of constants. The statistic

$$L_n = \sum_{i=1}^{n} a_{i,n} X_{(i)}$$

is called an *L-statistic*. When used as an estimator, L_n is referred to as *L-estimator*.

A variety of limiting distributions is possible for L_n. For example, if L_n is a function of a single order statistic, i.e., when $a_{i,n}$ is non-zero for all but one i, the limiting distribution of $X_{(i)}$ depends on how i is related to n, and the limiting distribution might not exist for some extreme order statistics. If many $a_{i,n}$'s are non-zero, L_n is asymptotically normal when the weights are reasonably smooth [1]. The specific results for the asymptotic distribution of the trimmed mean was provided by S. Stigler in [39]. His result is stated below.

First, let

$$A = F_0^{-1}(\alpha) - F_0^{-1}(\alpha-) \text{ and } B = F_0^{-1}(1-\beta) - F_0^{-1}((1-\beta)-) \tag{2.20}$$

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:

$$H(x) = \begin{cases} 0, & x < \xi_\alpha \\ \frac{F_0(x)-\alpha}{1-\alpha-\beta}, & \xi_\alpha \le x \le \xi_{1-\beta} \\ 1, & x > \xi_{1-\beta}. \end{cases} \tag{2.21}$$

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

Theorem 2.2.1. *[39, 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 (2.20), 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 [39] 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*

$$\begin{aligned}
\mathrm{Var}W &= \frac{1}{(1-\alpha-\beta)^2}\Big(\mathrm{Var}Z + (\xi_\alpha - \mu_{\alpha\beta})^2\mathrm{Var}Z_1 \\
&\quad + 2(\xi_\alpha - \mu_{\alpha\beta})(\xi_{1-\beta} - \mu_{\alpha\beta})\mathrm{Cov}(Z_1, Z_2) + (\xi_{1-\beta} - \mu_{\alpha\beta})^2 VarZ_2\Big) \\
&= \frac{1}{(1-\alpha-\beta)^2}\Big(\sigma^2_{\alpha\beta} + \alpha(1-\alpha)(\xi_\alpha - \mu_{\alpha\beta})^2 \\
&\quad - 2\alpha\beta(\xi_\alpha - \mu_{\alpha\beta})(\xi_{1-\beta} - \mu_{\alpha\beta}) + \beta(1-\beta)(\xi_{1-\beta} - \mu_{\alpha\beta})^2\Big) \\
&=: \tau^2_{\alpha\beta},
\end{aligned}$$

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

2.3 Finite-sample breakdown point

Definition 2.3.1. [26, p. 61] Let $\hat{\theta}_n(x)$ be an estimator of $\theta \in \Theta$ defined for samples $\mathbf{x} = \{x_1, \ldots, x_n\}$. The *replacement finite-sample breakdown point* (FBP) of $\hat{\theta}_n$ at x is the largest proportion of $\epsilon_n^*(\hat{\theta}_n, \mathbf{x})$ data points that can be arbitrarily replaced by outliers without $\hat{\theta}_n$ leaving a set which is bounded and also bounded away from the boundary of Θ. More formally, let

$$\mathcal{X}_m = \{y \mid \#(\mathbf{y}) = n, \#(\mathbf{x} \cap \mathbf{y}) = n - m\},$$

i.e., $\mathcal{X}_m$ is the set of all data sets $\mathbf{y}$ of size n having $n - m$ elements in common with $\mathbf{x}$. Then

$$\epsilon_n^*(\hat{\theta}_n, \mathbf{x}) = \frac{m^*}{n},$$

where $m^* = \max\{m \geq 0 : \hat{\theta}_n(\mathbf{y})$ bounded and also bounded away from $\partial\Theta \ \forall \mathbf{y} \in \mathcal{X}_m\}$.

For most cases of interest, $\epsilon_n^*(\hat{\theta}_n)$ does not depend on $\mathbf{x}$.

Example 2.3.1. Equivariant location estimators. FBP is given by

$$\epsilon_n^*(\hat{\theta}_n) \leq \frac{1}{n} \left\lfloor \frac{n-1}{2} \right\rfloor.$$

This bound is attained by M-estimators with an odd and bounded ψ-function (see [26] for proof).

Example 2.3.2. M-estimator with a preliminary dispersion estimator (2.13). The breakdown point cannot be larger than that of the dispersion estimator. For bounded, monotone and symmetric ψ-functions and MAD dispersion estimator it holds

$$\epsilon_n^*(\hat{\theta}_n) = \frac{1}{n} \left\lfloor \frac{n-1}{2} \right\rfloor.$$

Note that the ψ-function of Huber estimator is odd and bounded, and the same holds for the smoothed Huber estimator.

Example 2.3.3. For the α-trimmed mean, $m^* = [n\alpha]$, so that $\epsilon_n^* \approx \alpha$ for large n.

An important property of a confidence interval is its length and whether it is sensitive to outliers in the data. A finite-sample breakdown point can be defined also for the length of a confidence interval.

Definition 2.3.2. [43, p. 138] Let L_n be the length of a confidence interval based on a sample $X_1, \ldots, X_n$, $|X_i| < \infty$. The *finite-sample breakdown point of* L_n is given by

$$\epsilon_n(L_n; X_1, \ldots, X_n) = \frac{1}{n} \min \left\{ m \mid \max_{i_1,\ldots,i_m} \sup_{Y_1,\ldots,Y_m} \{L_n(Z_1, \ldots, Z_n)\} = \infty \right\},$$

where $Z_1, \ldots, Z_n$ is obtained by replacing m data points $X_{i_1}, \ldots, X_{i_m}$ by the arbitrary values $Y_1, \ldots, Y_m$. Here, X_i and Y_i are not random variables.

Definition 2.3.3. [43, p. 138] Define the *finite-sample upper breakdown point* ϵ_n^U of L_n as

$$\epsilon_n^U = \frac{1}{n} \min \left\{ m \mid \sup_{Y_{(1)},\ldots,Y_{(m)}} \{L_n(X_{(1)}, \ldots, X_{(n-m)}, Y_{(1)}, \ldots, Y_{(n)})\} = \infty \right\},$$

where m largest X_is are replaced by m large Y_is that satisfy $X_{(n-m)} < Y_{(1)}, \ldots, < Y_{(m)}$.

Chapter 3

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

The aim of this chapter is to establish the empirical likelihood method for the difference of two M-estimators. The results presented in this chapter have been published in M. Delesa-Vēliņa et al. [52]

In Chapter 3.1 we present the existing results for the EL inference for M-estimators in the one-sample case. A. B. Owen [28] provided conditions under which the empirical likelihood confidence intervals can be constructed for M-estimators. Intuitively, the empirical likelihood inference for the mean is not a robust procedure, since the EL confidence interval borders are impacted by the outlier values in the data set. It was demonstrated formally in [43] that the EL confidence intervals for Huber estimator are in a certain sense (regarding their length) more robust than the EL confidence intervals for the means. The good properties of the EL inference for the Huber estimator in the one-sample case provides the motivation for exploring the corresponding method in the two-sample case.

In Chapter 3.2 the main results of the Chapter are presented: we establish the conditions under which the empirical likelihood ratio can be constructed for a difference of two general M-estimators and show that the smoothed Huber estimator fits in this setting. Note that the two-sample EL setting of Y. Qin and L. Zhao [35] cannot be applied directly to the Huber estimator, since the condition (C1) in Assumption 1.4.1 that necessitates a continuous derivative of the estimating function does not hold for Huber estimator's ψ-function (2.8). Thus the smoothed ψ-function (2.18) is used.

As it was noted in Chapter 2, the M-estimators defined by (2.5) or (2.6) 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 (2.13) are preferred. The difficulty involved is that the real value of the scale parameter σ in (2.13) is not known in practical situations. Thus, the scale parameter is interpreted as an additional nuisance parameter for the EL maximization problem and needs to be dealt with appropriately. One option is to profile the empirical likelihood ratio simultaneously on all the nuisance parameters θ_0, σ_1 and σ_2 [31, Chapter 3.5]. However, this approach is computationally complicated. The second option taken in this work is to use the plug-in empirical likelihood that allows possibly infinite-dimension nuisance parameters in estimating equations. The plug-in EL was formalized for the one-sample case by N. Hjort et al. [16]. J. Valeinis [48] generalized the conditions for the plug-in EL method for the two-sample case.

3.1 Empirical likelihood for M-estimators in the one-sample case

Theorem 3.1.1. *[28, p. 243] Let $T(F)$ be an M-estimator solving the equation (2.3) and let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with common distribution F_0. Consider univariate functions $\psi_{\cdot t}(x)$ and $\psi_{x\cdot}(t)$ given by*

$$\psi_{\cdot t}(x) = \psi(x, t) = \psi_{x\cdot}(t).$$

Assume that (i) $T(F_0) = \tau$ exists and is unique; (ii) $\psi_{\cdot\tau}(x)$ is measurable; (iii) $\operatorname{Var}\{\psi(X_i, \tau)\} > 0$; (iv) $\operatorname{E}\{|\psi_{\cdot\tau}(X_i)|^3\} < \infty$ and (v) $\psi(x, t)$ is a non-increasing function in t for all x in the support of F_0. For a positive $c < 1$ and empirical likelihood ratio $\mathcal{R}(\theta)$ given by (1.5), consider $\mathcal{F}_{c,n} = \{F | \mathcal{R}(\theta) \geq c, F \ll F_n\}$, where $F \ll F_n$ denotes distributions with support in the sample, and

$$S_{c,n} = \bigcup_{F \in \mathcal{F}_{c,n}} \left\{ t \middle| \int \psi(x, t) dF(x) = 0 \right\}.$$

Then $S_{c,n}$ is an interval and

$$P(T(F_0) \in S_{c,n}) \to P\left\{\chi_1^2 \leq -2 \log c\right\} \text{ as } n \to \infty.$$

Assumptions (i)-(v) hold for the ψ-function of the Huber location estimator given by (2.8). Consider an estimating equation $\psi\{(X_i - t)/\hat{\sigma}\}$, where $\hat{\sigma}$ is a preliminary dispersion estimator. Then the profile EL ratio is in the form

$$\mathcal{R}(t) = \sup_p \left\{ \prod_{i=1}^n np_i | p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \psi\left(\frac{X_i - t}{\hat{\sigma}}\right) = 0 \right\}. \tag{3.1}$$

Put $Z_i = \psi(X_i - t)$. Then (3.1) is maximized by

$$p_i = \{n(1 + \lambda Z_i)\}^{-1},$$

and λ is the unique root of

$$\frac{1}{n} \sum_{i=1}^n \frac{Z_i}{1 + \lambda Z_i} = 0$$

in the interval $(-Z_{(n)}^{-1}, -Z_{(1)}^{-1})$. A. B. Owen [28] provides bracketing values for λ as a solution to $\{n(1 + \lambda z)\}^{-1} = 2$, where z takes the values $Z_{(1)}$ and $Z_{(n)}$.

Regarding the confidence interval for the mean, its end points are determined by weighted means of the data. Since all the weights p_i are positive by construction, intuitively, the length of the confidence interval can be greatly influenced by outlier values. It might be interesting to consider the finite-sample breakdown point (FBP) of the length of the EL confidence interval as defined in (2.3.3). M. Tsao and J. Zhou [43] gave a FBP ϵ_n^U of the length of the EL confidence interval for the mean and for the Huber estimator. They proved that $\epsilon_n^U = 1/n$ for the EL confidence interval of the mean, namely, it suffices to arbitrary change one observation in the sample for the interval length to break down. In contrast, EL confidence interval for the Huber estimator is more robust and its breakdown point attains 0.5 when $n \to \infty$.

Theorem 3.1.2. *[43, p. 133] The finite-sample upper breakdown point ϵ_n^U (2.3.3) of $1-\alpha$ EL confidence interval length L_n for Huber estimator is given by*

$$\epsilon_n^U = \min\{m \mid c(m) \geq c\}/n,$$

where n is sample size, $c = \exp(-\chi^2_{1,\alpha}/2)$ and

$$c(m) = \left(\frac{n}{2m}\right)^m \left(\frac{n}{2(n-m)}\right)^{(n-m)}.$$

Note the interpretation of $c(m)$: it is the maximum value of the likelihood ratio $\mathcal{R}(\mu)$ from (3.1) in the case where sample contains m outliers in the upper side. The maximum is attained by weights p_i where $p_i = 1/(2m)$ if the ith observation is an outlier, and $p_i = 1/\{2(n-m)\}$ otherwise. The ϵ_n^U depends on the confidence level and on the sample size n. For example, for 95% confidence interval, $\epsilon_{10}^U = 0.246$, $\epsilon_{20}^U = 0.318$, and $\epsilon_{100}^U = 0.419$ [43].

3.2 Main results

Consider the two-sample problem defined in Chapter 1.4: $X_1, \ldots, X_{n_1}$ 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 (1.12) - (1.13) have the form

$$\begin{aligned} w_1(X, \Delta_0, \theta_0, \sigma_1^0, \sigma_2^0) &= \psi\left(\frac{X-\theta_0}{\sigma_1^0}\right), \\ w_2(Y, \Delta_0, \theta_0, \sigma_1^0, \sigma_2^0) &= \psi\left(\frac{Y-\Delta_0-\theta_0}{\sigma_2^0}\right), \end{aligned} \tag{3.2}$$

where ψ corresponds to a general ψ-function of an M-estimator defined in (2.13), 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, and θ denotes the location parameter for the sample X with the true value θ_0.

The two-sample problem setting in [35] is for fixed estimating functions w_1 and w_2, but in our case (3.2) involves nuisance parameters θ, σ_1, σ_2, V_1 and V_2. In addition, estimating function is dependent on sample size n via V. There are two possible approaches of dealing with the nuisance parameters. First approach is to profile the EL function on several nuisance parameters. Consider a fixed V and profile on θ_0, σ_1 and σ_2 simultaneously (see for example, [31]). The second approach involves the *plug-in empirical likelihood*, that allows estimation of the nuisance parameters under some additional restrictions, see [16] for general assumptions in one sample case. The plug-in EL approach for smooth estimating equations was extended to two-sample case in [48].

We define the profile empirical likelihood function

$$\begin{aligned} \mathcal{R}(\Delta, \theta, \sigma_1, \sigma_2) = n_1^{n_1} n_2^{n_2} \sup_{p,q} \Bigg\{ & \prod_{i=1}^{n_1} p_i \prod_{j=1}^{n_2} q_j : p_i \geq 0, q_j \geq 0, \sum_{i=1}^{n_1} p_i = 1, \\ & \sum_{j=1}^{n_2} q_j = 1, \sum_{i=1}^{n_1} p_i \psi\left(\frac{X_i - \theta}{\sigma_1}\right) = 0, \sum_{j=1}^{n_2} q_j \psi\left(\frac{Y_j - \Delta - \theta}{\sigma_2}\right) = 0. \Bigg\} \end{aligned} \tag{3.3}$$

A unique solution to (3.3) 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 (1.19)-(1.20) with the estimating functions defined by (3.2), i.e., from

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

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

$$\frac{\partial}{\partial\theta}\mathcal{W}(\Delta, \theta, \sigma_1, \sigma_2) = \sum_{i=1}^{n_1} \frac{\lambda_1\psi'\left(\frac{X_i-\theta}{\sigma_1}\right)}{1+\lambda_1\psi\left(\frac{X_i-\theta}{\sigma_1}\right)} + \sum_{j=1}^{n_2} \frac{\lambda_2\psi'\left(\frac{Y_j-\Delta-\theta}{\sigma_2}\right)}{1+\lambda_2\psi\left(\frac{Y_j-\Delta-\theta}{\sigma_2}\right)} = 0, \tag{3.5}$$

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 (2.13):

Assumption 3.2.1.

(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 neighbourhood 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 neighbourhood, $\psi'\left((Y_j-\theta-\Delta)/\hat{\sigma}_2\right)$ and $\psi^3\left((Y_j-\theta-\Delta)/\hat{\sigma}_2\right)$ are bounded by some integrable function $G_2(Y)$ in this neighbourhood, 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 non-zero.

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

Assumption 3.2.2.

(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 3.2.3.

(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 3.2.1. *Assumption 3.2.1 is very similar to Assumption 1.4.1, 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 [34] (see Theorem 1.3.1) and describes the smoothness conditions for the estimating functions. Part (A3) requires that the sample sizes are asymptotically comparable.*

Assumption 3.2.2 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 in Chapter 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 3.2.3 contains technical assumptions for the plug-in empirical likelihood, similarly as in [16] and [48]. 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 Valeinis [48] for the details.

Next, we present Lemma 3.2.1 commenting on the relationship between the Assumptions 3.2.1 - 3.2.3, the main Theorem 3.2.1 that establishes the EL method for the difference of two general M-estimators, and Lemma 3.2.2 that states the conditions under which the smoothed Huber estimator fits in the setting of Theorem 3.2.1. We proceed with the proofs in the next section.

Lemma 3.2.1. *(M. Delesa-Vēliņa et al. [52]) For a general ψ-function of an M-estimator satisfying Assumptions 3.2.1 and 3.2.2, Assumption 3.2.3 holds.*

Theorem 3.2.1. *(M. Delesa-Vēliņa et al. [52]) Assume that the EL maximization problem has a solution $\hat{\theta}(\Delta, \hat{\sigma}_1, \hat{\sigma}_2)$ determined by (3.5). Then, for a general ψ-function of an M-estimator satisfying Assumptions 3.2.1 and 3.2.3, 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 3.2.2. *(M. Delesa-Vēliņa et al. [52]) Let $\tilde{\psi}_k$ be the score function (2.18) defining the smoothed Huber M-estimator and let $\hat{\sigma}_1$ and $\hat{\sigma}_2$ be the mean absolute deviation (MAD) dispersion estimates (2.14) of samples X and Y, respectively. Assume that the underlying distributions F_1 and F_2 of X and Y are symmetric. Then Assumptions 3.2.1 and 3.2.2 hold for $\psi = \tilde{\psi}_k$.*

3.3 Proofs

At first, we present one technical Lemma.

Lemma 3.3.1. *Suppose $1/3 < \eta < 1/2$ and Assumption 3.2.3 is satisfied. Then*

$$\lambda_1(\theta) = O_p(n_1^{-\eta}), \quad \lambda_2(\theta) = O_p(n_2^{-\eta})$$

uniformly around $\theta \in \{\theta : |\theta - \theta_0| \leq cn_1^{-\eta}|\}$, where c is some positive constant.

For the proof of Lemma 3.3.1, see [35].

Proof of Theorem 3.2.1. Denote $\hat{\lambda}_1 = \lambda_1(\Delta, \hat{\theta}, \hat{\sigma}_1, \hat{\sigma}_2)$, $\hat{\lambda}_2 = \lambda_2(\Delta, \hat{\theta}, \hat{\sigma}_1, \hat{\sigma}_2)$.

First, we show that given the root $\hat{\theta} = \theta(\Delta, \hat{\sigma}_1, \hat{\sigma}_2)$ of (3.5), the following holds:

$$\sqrt{n_1}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, \frac{V_1 V_2 (M_1^2 + kM_2^2)}{c_1^2}\right), \tag{3.6}$$

$$\hat{\lambda}_1 = -k\left(\frac{M_2}{M_1}\right)\hat{\lambda}_2 + o_p(n_1^{-1/2}), \tag{3.7}$$

$$\sqrt{n_1}\hat{\lambda}_2 \xrightarrow{d} N\left(0, \frac{M_1^2}{kc_1}\right), \tag{3.8}$$

where

$$c_1 = V_2 M_1^2 + kV_1 M_2^2.$$

Consider

$$Q_1(\theta, \lambda_1, \lambda_2) = \frac{1}{n_1}\sum_{i=1}^{n_1} \frac{\psi\left(\frac{X_i - \theta}{\sigma_1}\right)}{1 + \lambda_1 \psi\left(\frac{X_i - \theta}{\sigma_1}\right)},$$

$$Q_2(\theta, \lambda_1, \lambda_2) = \frac{1}{n_2}\sum_{j=1}^{n_2} \frac{\psi\left(\frac{Y_i - \Delta - \theta}{\sigma_2}\right)}{1 + \lambda_2 \psi\left(\frac{Y_i - \Delta - \theta}{\sigma_2}\right)},$$

$$Q_3(\theta, \lambda_1, \lambda_2) = \lambda_1 \times \frac{1}{n_1}\sum_{i=1}^{n_1} \frac{\psi'\left(\frac{X_i - \theta}{\sigma_1}\right)}{1 + \lambda_1 \psi\left(\frac{X_i - \theta}{\sigma_1}\right)} + \lambda_2 \times \frac{1}{n_1}\sum_{j=1}^{n_2} \frac{\psi'\left(\frac{Y_i - \Delta - \theta}{\sigma_2}\right)}{1 + \lambda_2 \psi\left(\frac{Y_i - \Delta - \theta}{\sigma_2}\right)}.$$

Then we have

$$Q_i(\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2) = 0 \text{ for } i = 1, 2, 3.$$

By Taylor expansion, we have

$$0 = Q_i(\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2) = Q_i(\theta_0, 0, 0) + \frac{\partial Q_i(\theta_0, 0, 0)}{\partial \theta}(\hat{\theta} - \theta_0) + \frac{\partial Q_i(\theta_0, 0, 0)}{\partial \lambda_1}\hat{\lambda}_1 + \frac{\partial Q_i(\theta_0, 0, 0)}{\partial \lambda_2}\hat{\lambda}_2 + O_p(n_1^{-2\eta}), \ i = 1, 2, 3.$$

Hence,

$$Q_i(\theta_0,0,0)+\frac{\partial Q_i(\theta_0,0,0)}{\partial\theta}(\hat\theta-\theta_0)+\frac{\partial Q_i(\theta_0,0,0)}{\partial\lambda_1}\hat\lambda_1+\frac{\partial Q_i(\theta_0,0,0)}{\partial\lambda_2}\hat\lambda_2=o_p(n_1^{-1/2}),\quad i=1,2,3.$$

From conditions (C1) - (C3) of Assumption 3.2.3 it follows

$$\begin{aligned}
&\frac{\partial Q_1(\theta_0,0,0)}{\partial\theta}\to M_1 \text{ a.s.}, && \frac{\partial Q_1(\theta_0,0,0)}{\partial\lambda_1}\to -V_1 \text{ a.s.}, && \frac{\partial Q_1(\theta_0,0,0)}{\partial\lambda_2}=0,\\
&\frac{\partial Q_2(\theta_0,0,0)}{\partial\theta}\to M_2 \text{ a.s.}, && \frac{\partial Q_2(\theta_0,0,0)}{\partial\lambda_1}=0, && \frac{\partial Q_2(\theta_0,0,0)}{\partial\lambda_2}\to -V_2 \text{ a.s.},\\
&\frac{\partial Q_3(\theta_0,0,0)}{\partial\theta}=0, && \frac{\partial Q_3(\theta_0,0,0)}{\partial\lambda_1}\to M_1 \text{ a.s.}, && \frac{\partial Q_3(\theta_0,0,0)}{\partial\lambda_2}\to kM_2 \text{ a.s.}
\end{aligned}$$

Thus

$$\begin{pmatrix}\hat\theta-\theta_0\\ \hat\lambda_1\\ \hat\lambda_2\end{pmatrix}=S^{-1}\begin{pmatrix}Q_1(\theta_0,0,0)\\ Q_2(\theta_0,0,0)\\ 0\end{pmatrix}+o_p(n_1^{-1/2}),$$

where

$$S=\begin{pmatrix}M_1 & -V_1 & 0\\ M_2 & 0 & -V_2\\ 0 & M_1 & kM_2\end{pmatrix}$$

and

$$S^{-1}=\frac{1}{c_1}\begin{pmatrix}V_2M_1 & kV_1M_2 & V_1V_2\\ -kM_2^2 & kM_1M_2 & V_2M_1\\ M_1M_2 & -M_1^2 & V_1M_2\end{pmatrix}.$$

Then we have

$$\hat\theta-\theta_0=\frac{1}{c_1}(V_2M_1Q_1(\theta_0,0,0)+kV_1M_2Q_2(\theta_0,0,0))+o_p(n_1^{-1/2}),$$

$$\hat\lambda_1=-\frac{kM_2}{c_1}(M_2Q_1(\theta_0,0,0)-M_1Q_2(\theta_0,0,0))+o_p(n_1^{-1/2}),$$

$$\hat\lambda_2=\frac{M_1}{c_1}(M_2Q_1(\theta_0,0,0)-M_1Q_2(\theta_0,0,0))+o_p(n_1^{-1/2}).$$

Note that according to Assumption 3.2.3 it holds that

$$\sqrt{n_1}\begin{pmatrix}Q_1(\theta_0,0,0)\\ Q_2(\theta_0,0,0)\end{pmatrix}\xrightarrow{d}N\left(0,\begin{pmatrix}V_1 & 0\\ 0 & k^{-1}V_2\end{pmatrix}\right).$$

The statements (3.6) - (3.8) follow. By Assumption 3.2.1 (A2), $\psi^3((X_i-\hat\theta)/\hat\sigma)$ is bounded by some integrable function $G_1(X)$. Thus $E|\psi((X_i-\hat\theta)/\hat\sigma_1)|^3$ exists, which is equivalent to

$$\sum P(|\psi((X_i-\hat\theta)/\hat\sigma_1)|^3>n_1)<\infty,$$

see, for example, [28]. It follows by the Borel-Cantelli lemma that $|\psi((X_i-\hat\theta)/\hat\sigma_1)|<n_1^{1/3}$ with probability 1. This implies that

$$\max_{1\le i\le n_1}|\psi((X_i-\hat\theta)/\hat\sigma_1)|\le n_1^{1/3}.$$

Thus, using Lemma 3.3.1 with $\eta \in (1/3; 1/2)$ we have

$$\max_{1\le i\le n_1}\left|\hat{\lambda}_1\psi\left(\frac{X_i-\hat{\theta}}{\hat{\sigma}_1}\right)\right| = O_p(n_1^{-\eta})o_p(n_1^{1/3}) = o_p(1),$$

and with $\xi \in \left[0, \hat{\lambda}_1\psi\left(\frac{X_i-\hat{\theta}}{\hat{\sigma}_1}\right)\right]$ by the law of large numbers we have

$$\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{\psi^3\left(\frac{X_i-\hat{\theta}}{\hat{\sigma}_1}\right)}{(1+\xi)^3} = O_p(1).$$

Thus, the following holds:

$$\frac{n_1}{3}\hat{\lambda}_1^3\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{\psi^3\left(\frac{X_i-\hat{\theta}}{\hat{\sigma}_1}\right)}{(1+\xi)^3} = O(n_1)O_p(n_1^{-3\eta})O_p(1) = O_p(n^{-3\eta+1}) = o_p(1).$$

A similar argument can be made for $\hat{\lambda}_2$. Then, using Taylor expansion for $\log(1+x)$, we have

$$\begin{aligned}
&\log\mathcal{R}(\Delta_0,\hat{\theta},\hat{\sigma}_1,\hat{\sigma}_2) = \\
&\quad= -\sum_{i=1}^{n_1}\log\left(1+\hat{\lambda}_1\psi\left(\frac{X_i-\hat{\theta}}{\hat{\sigma}_1}\right)\right) - \sum_{j=1}^{n_2}\log\left(1+\hat{\lambda}_2\psi\left(\frac{Y_j-\Delta_0-\hat{\theta}}{\hat{\sigma}_2}\right)\right) \\
&\quad= -n_1\hat{\lambda}_1S_{1x}(\hat{\theta}) + \frac{n_1}{2}\hat{\lambda}_1^2S_{2x}(\hat{\theta}) - n_2\hat{\lambda}_2S_{1y}(\hat{\theta}) + \frac{n_2}{2}\hat{\lambda}_2^2S_{2y}(\hat{\theta}) + o_p(1),
\end{aligned} \tag{3.9}$$

where

$$S_{1x}(\hat{\theta}) = \frac{1}{n_1}\sum_{i=1}^{n_1}\psi\left(\frac{X-\hat{\theta}}{\hat{\sigma}_1}\right), \qquad S_{2x}(\hat{\theta}) = \frac{1}{n_1}\sum_{i=1}^{n_1}\psi^2\left(\frac{X-\hat{\theta}}{\hat{\sigma}_1}\right),$$

$$S_{1y}(\hat{\theta}) = \frac{1}{n_2}\sum_{j=1}^{n_2}\psi\left(\frac{Y-\Delta_0-\hat{\theta}}{\hat{\sigma}_2}\right), \qquad S_{2y}(\hat{\theta}) = \frac{1}{n_2}\sum_{j=1}^{n_2}\psi^2\left(\frac{Y-\Delta_0-\hat{\theta}}{\hat{\sigma}_2}\right).$$

From (3.4) we have

$$\begin{aligned}
0 &= \sum_{i=1}^{n_1}\frac{\psi\left(\frac{X_i-\hat{\theta}}{\hat{\sigma}_1}\right)}{1+\lambda_1\psi\left(\frac{X_i-\hat{\theta}}{\hat{\sigma}_1}\right)} \\
&= \frac{1}{n_1}\sum_{i=1}^{n_1}\psi\left(\frac{X_i-\hat{\theta}}{\hat{\sigma}_1}\right)\left(1-\hat{\lambda}_1\psi\left(\frac{X_i-\hat{\theta}}{\hat{\sigma}_1}\right)+\frac{\hat{\lambda}_1^2\psi^2\left(\frac{X_i-\hat{\theta}}{\hat{\sigma}_1}\right)}{1+\hat{\lambda}_1\psi\left(\frac{X_i-\hat{\theta}}{\hat{\sigma}_1}\right)}\right) \\
&= S_{1x}(\hat{\theta}) - \hat{\lambda}_1S_{2x}(\hat{\theta}) + \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{\hat{\lambda}_1^2\psi^3\left(\frac{X_i-\hat{\theta}}{\hat{\sigma}_1}\right)}{1+\hat{\lambda}_1\psi\left(\frac{X_i-\hat{\theta}}{\hat{\sigma}_1}\right)}.
\end{aligned}$$

The absolute value of the last term is bound by

$$\frac{1}{n_1}\sum_{i=1}^{n_1}|\psi^3((X_i-\hat{\theta})/\hat{\sigma}_1)||\hat{\lambda}_1|^2|1+\hat{\lambda}_1\psi((X_i-\hat{\theta})/\hat{\sigma}_1)|^{-1} = O(1)O_p(n_1^{-2\eta})O_p(1) = O_p(n_1^{-2\eta}).$$

Thus, it follows (using a similar argument for $\hat{\lambda}_2$) that

$$S_{1x}(\hat{\theta}) = \hat{\lambda}_1 S_{2x}(\hat{\theta}) + O_p(n_1^{-2\eta}), \quad S_{1y}(\hat{\theta}) = \hat{\lambda}_2 S_{2y}(\hat{\theta}) + O_p(n_1^{-2\eta}).$$

Hence from (3.9) we have

$$-2\log \mathcal{R}(\Delta_0, \hat{\theta}, \hat{\sigma}_1, \hat{\sigma}_2) = n_1\hat{\lambda}_1^2 S_{2x}(\hat{\theta}) + n_2\hat{\lambda}_2^2 S_{2y}(\hat{\theta}) + o_p(1).$$

From condition (C3) of Assumption 3.2.3 we have

$$S_{2x}(\hat{\theta}) = V_1 + o_p(1), \quad S_{2y}(\hat{\theta}) = V_2 + o_p(1),$$

and using (3.7)

$$\begin{aligned} -2\log \mathcal{R}(\Delta_0, \hat{\theta}, \hat{\sigma}_1, \hat{\sigma}_2) &= n_1\hat{\lambda}_1^2 S_{2x}(\hat{\theta}) + n_2\hat{\lambda}_2^2 S_{2y}(\hat{\theta}) + o_p(1) \\ &= n_1 k^2 \frac{M_2^2}{M_1^2}\hat{\lambda}_2^2 V_1 + n_2\hat{\lambda}_2^2 V_2 + o_p(1) \\ &= k\left[\sqrt{n_1}\hat{\lambda}_2\right]^2 \left(\frac{kV_1M_2^2 + V_2M_1^2}{M_1^2}\right) + o_p(1). \end{aligned}$$

Using (3.8), we have

$$\sqrt{n_1}\hat{\lambda}_2 \xrightarrow{d} N\left(0, \frac{M_1^2}{k(V_2M_1^2 + kV_1M_2^2)}\right).$$

Then

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

which proves Theorem 3.2.1.

Proof of Lemma 3.2.1. We will present the proof only for the sample X, as for Y the result can be obtained similarly. Condition (A2) of Assumption 3.2.1 states that ψ' is bounded by some integrable function; thus, the expectation exists and condition (C1) of Assumption 3.2.3 holds by the law of large numbers. To prove (C2) and (C3), we follow the technique used in [26, Section 10.6] to establish the asymptotic distribution of the location M-estimators with a preliminary scale. Denote $u_i = X_i - \theta_0$ and $\hat{\sigma}_1 = \sigma_1^0 + \delta$. Expand $\psi(u_i/\hat{\sigma}_1)$ to the second order Taylor series around θ_0:

$$\psi\left(\frac{u_i}{\hat{\sigma}_1}\right) = \psi\left(\frac{u_i}{\sigma_1^0 + \delta}\right) \approx \psi\left(\frac{u_i}{\sigma_1^0}\right) + \frac{u_i}{\sigma_1^0}\psi'\left(\frac{u_i}{\sigma_1^0}\right)\left(1 - \frac{\sigma_1^0}{\hat{\sigma}_1}\right).$$

Summing over i and dividing by $\sqrt{n_1}$, we obtain

$$\frac{1}{\sqrt{n_1}}\sum_{i=1}^{n_1}\psi\left(\frac{u_i}{\hat{\sigma}_1}\right) = \sqrt{n_1}A_{n_1} + \sqrt{n_1}B_{n_1}\left(1 - \frac{\sigma_1^0}{\hat{\sigma}_1}\right), \tag{3.10}$$

where

$$A_{n_1} = \frac{1}{n_1}\sum_{i=1}^{n_1}\psi\left(\frac{u_i}{\sigma_1^0}\right), \quad B_{n_1} = \frac{1}{n_1}\sum_{i=1}^{n_1}\left(\frac{u_i}{\sigma_1^0}\right)\psi'\left(\frac{u_i}{\sigma_1^0}\right).$$

$\mathrm{E}\psi(u_i/\sigma_1^0) = 0$ by the definition of the M-estimator, thus $\sqrt{n_1}A_{n_1} \xrightarrow{d} N(0, V_1)$ by Assumption 3.2.2 condition (B2). According to Assumption 3.2.2 condition (B3), $\sqrt{n_1}B_{n_1}$ tends to a normal distribution by the central limit theorem, and since $(1 - \sigma_1^0/\hat{\sigma}_1) \to 0$ by

Assumption 3.2.2 condition (B1), the second term in the right-hand side of (3.10) tends to zero by Slutsky's lemma. Hence, we obtain Assumption 3.2.3 condition (C2).

Now, we expand $\psi(u_i/\hat{\sigma}_1)$ around θ_0:

$$\psi^2\left(\frac{u_i}{\sigma_1^0+\delta}\right) = \psi^2\left(\frac{u_i}{\sigma_1^0}\right) + 2\psi\left(\frac{u_i}{\sigma_1^0}\right)\psi'\left(\frac{u_i}{\sigma_1^0}\right)\left(\frac{u_i}{\sigma_1^0}\right)\left(1-\frac{\sigma_1^0}{\hat{\sigma}_1}\right).$$

Summing over i and dividing by n_1,

$$\frac{1}{n_1}\sum_{i=1}^{n_1}\psi^2\left(\frac{u_i}{\sigma_1^0+\delta}\right) = C_{n_1} + 2D_{n_1}\left(1-\frac{\sigma_1^0}{\hat{\sigma}_1}\right),$$

where

$$C_{n_1} = \frac{1}{n_1}\sum_{i=1}^{n_1}\psi^2\left(\frac{u_i}{\sigma_1^0}\right), \text{ and } D_{n_1} = \frac{1}{n_1}\sum_{i=1}^{n_1}\psi\left(\frac{u_i}{\sigma_1^0}\right)\psi'\left(\frac{u_i}{\sigma_1^0}\right)\left(\frac{u_i}{\sigma_1^0}\right).$$

By the central limit theorem and Assumption 3.2.2 condition (B2), C_{n_1} tends to V_1, D_{n_1} tends to a constant by condition (B4), and $(1-\sigma_1^0/\hat{\sigma}_1) \to 0$. Hence we obtain (C3).

Proof of Lemma 3.2.2. First, we verify that Assumption 3.2.1 condition (A2) holds. The derivative $\tilde{\psi}_k'$ is continuous due to the general smoothing principle of M-estimators established in (2.16). Next, $0 \le \tilde{\psi}_k'(x) \le 1$ and $0 \le \tilde{\psi}_k^3(x) \le k^3$, thus they are bounded.

Now, we verify that conditions of Assumption 3.2.2 hold. (B1) holds for $\hat{\sigma}_1 = \text{MAD} = Med\{|X - Med(X)|\}$ under mild (smoothness) conditions on the underlying distribution F (see, for example, [11]). (B2) holds because $\tilde{\psi}_k$ with $k < \infty$ is a bounded ψ-function. For F_1 symmetric, θ_0 coincides with the center of symmetry and, since $\tilde{\psi}_k$ is odd, (B3) holds. Next, as $\tilde{\psi}_k'(x) = 0$ for $|x| > k$, for F_1 symmetric (B4) is an expectation of an even and bounded function, hence it is finite.

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 by M. Delesa-Vēliņa et al. [8].

In Chapter 4.1 the EL method for the trimmed means in the one-sample case by G. Qin and M. Tsao [33] are presented. Owen [28] established the EL method for independent observations, while the observations of the trimmed sample are dependent. Using Owen's EL method [28] on the trimmed means directly results in incorrect limiting distribution. Instead, G. Qin and M. Tsao [33] proposed to estimate the EL ratio 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. They showed in a simulation setting that the EL confidence interval for the trimmed mean had better coverage than the confidence interval for the trimmed mean based on the normal approximation.

In Chapter 4.2 our new EL method for the difference of two trimmed means is developed, extending the results of [33] to the two-sample case using the tools of Y. Qin and L. Zhao [35] described in Chapter 1. In Chapter 4.3 the proofs are given.

4.1 Empirical likelihood for trimmed means in the one-sample case

Consider the setting of Chapter 2.2: let $X_1, X_2, \dots, X_n$ be i.i.d. random variables with distribution F_0, and $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ be ordered statistics. Let $\bar{X}_{\alpha\beta}$ be the sample trimmed mean as defined by (2.19), 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 is 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 empirical likelihood ratio for the trimmed mean as

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

Theorem 4.1.1. *[33, Theorem 2.1]*
Assume F_0 is continuous, $F_0'(\xi_\alpha) > 0$ and $F_0'(\xi_{1-\beta}) > 0$, and let $\mu^0_{\alpha\beta}$ be the true value of the trimmed mean $\mu_{\alpha\beta}$. Then

$$-2a \log \mathcal{R}(\mu^0_{\alpha\beta}) \xrightarrow{d} \chi^2_1,$$

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

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

The proof of the Theorem 4.1.1 can be found in [33].

Remark 4.1.1. *From conditions $F_0'(\xi_\alpha) > 0$ and $F_0'(\xi_{1-\beta}) > 0$ it follows that $A = 0$ and $B = 0$ in Theorem 2.2.1, i.e., the trimming is done at unique percentiles of the distribution F_0 (note that there was a typo in [33] demanding $F_0'(\xi_\beta) < 0$ instead of $F_0'(\xi_{1-\beta}) > 0$). Note that $\mu_{\alpha\beta}$ and $\sigma^2_{\alpha\beta}$ are the mean and the variance, respectively, of the truncated distribution of F_0 defined by (2.21). Under the conditions of Theorem [33], it follows from Theorem 2.2.1 and Remark 2.2.1 that the asymptotic distribution of $\bar{X}_{\alpha\beta}$ is normal with mean $\mu_{\alpha\beta}$ and variance $\tau^2_{\alpha\beta}$.*

G. Qin and M. Tsao [33] provided a consistent estimator for scaling constant a by

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

where

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

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

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

4.2 Main results

To obtain a test for the difference of two trimmed means, the idea is to combine the results of [33] and [35]. Consider the two-sample EL problem described in Chapter 1.4, where i.i.d. random variables $X_1, \dots, X_{n_1}$ and $Y_1, \dots, Y_{n_2}$ 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 (1.12) - (1.13) 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., $m_1 = n_1 - \lfloor n_1\alpha \rfloor - \lfloor n_1\beta \rfloor$, $m_2 = n_2 - \lfloor n_2\alpha \rfloor - \lfloor n_2\beta \rfloor$. 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

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

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 1.4, 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 (1.17) - (1.18) we have

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

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

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 log likelihood ratio is defined as

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

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

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

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 1.4.1.*

Theorem 4.2.1. *(M. Delesa-Vēliņa et al. [8]) 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*

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

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

4.3 Proofs

Lemma 4.3.1. *Suppose $1/3 < \eta < 1/2$ and the Assumption 4.2.1 conditions (A2) and (A3) are satisfied. Then*

$$\lambda_1(\mu_t) = O_p(n_1^{-\eta}), \qquad \lambda_2(\mu_t) = O_p(n_2^{-\eta})$$

uniformly around $\mu_t \in \{\mu_t : |\mu_t - \mu_{\alpha\beta 1}| \leq bn_1^{-\eta}|\}$, where b is some positive constant.

Proof. For the proof of Lemma 4.3.1, see [35].

Lemma 4.3.2. *Under the Assumption 4.2.1 condition (A1) the following holds as $n_1 \to \infty$ and $n_2 \to \infty$:*

$$\frac{1}{m_1}\sum_{i=r_1}^{s_1} w_1^2(X_{(i)}, \mu_{\alpha\beta 1}, \Delta) \xrightarrow{p} \sigma_1^2, \quad \frac{1}{m_2}\sum_{j=r_2}^{s_2} w_2^2(Y_{(j)}, \mu_{\alpha\beta 1}, \Delta) \xrightarrow{p} \sigma_2^2, \tag{4.8}$$

$$\sqrt{n_1}\left(\frac{1}{m_1}\sum_{i=r_1}^{s_1} w_1(X_{(i)}, \mu_{\alpha\beta 1}, \Delta)\right) \xrightarrow{d} N(0, \tau_1^2), \tag{4.9}$$

$$\sqrt{n_2}\left(\frac{1}{m_2}\sum_{j=r_2}^{s_2} w_2(Y_{(j)}, \mu_{\alpha\beta 1}, \Delta)\right) \xrightarrow{d} N(0, \tau_2^2). \tag{4.10}$$

Proof. We will prove the lemma for sample X, the proof for Y being equivalent. Note that $1/m_1 \xrightarrow{a.s.} 1/(1-\alpha-\beta)$ as $n_1 \to \infty$, and $\mu_{\alpha\beta 1}$ is the mean of the truncated distribution of F_1 given by (2.21). Thus

$$\frac{1}{m_1}\sum_{i=r_1}^{s_1} w_1^2(X_{(i)}, \mu_{\alpha\beta 1}, \Delta) = \frac{1}{m_1}\sum_{i=r_1}^{s_1}(X_{(i)} - \mu_{\alpha\beta 1})^2$$

provides a consistent estimate for the dispersion σ_1^2 of the truncated distribution of F_1, and we get (4.9). Next,

$$\sqrt{n_1}\left(\frac{1}{m_1}\sum_{i=r_1}^{s_1} w_1(X_{(i)}, \mu_{\alpha\beta 1}, \Delta)\right) = \sqrt{n_1}(\bar{X}_{\alpha\beta} - \mu_{\alpha\beta}) \xrightarrow{d} W$$

and we can use the Theorem 2.2.1 to establish the asymptotic distribution of the trimmed mean. The Assumptions of Theorem 4.1.1 ensure that

$$A = F_1^{-1}(\alpha) - F_1^{-1}(\alpha-) = 0 \text{ and } B = F_1^{-1}(1-\beta) - F_1^{-1}((1-\beta)-) = 0,$$

and the asymptotic distribution W is normal with variance

$$\begin{aligned}\mathrm{Var}W = \frac{1}{1-\alpha-\beta}&((1-\alpha-\beta)\sigma_{\alpha\beta 1}^2 + (\xi_\alpha - \mu_{\alpha\beta 1})^2\alpha(1-\alpha) + (\xi_{1-\beta} - \mu_{\alpha\beta 1})^2\beta(1-\beta) \\ &- 2\alpha\beta(\xi_\alpha - \mu_{\alpha\beta 1})(\xi_{1-\beta} - \mu_{\alpha\beta 1})) = \tau_{\alpha\beta 1}^2.\end{aligned}$$

Lemma 4.3.3. *Denote $\hat\lambda_1 = \lambda_1(\Delta_0, \hat\mu_t)$, $\hat\lambda_2 = \lambda_2(\Delta_0, \hat\mu_t)$. With the root $\hat\mu_t = \mu_t(\Delta_0)$ of (4.7) the following holds as $n_1, n_2 \to \infty$:*

$$\sqrt{n_1}(\hat\mu_t - \mu_{\alpha\beta 1}) \xrightarrow{d} -\frac{1}{\sigma_2^2 + \gamma\sigma_1^2}\left(\sigma_2^2 N\left(0, \tau_1^2\right) + \gamma\sigma_1^2 N(0, \kappa\tau_2^2)\right),$$
$$c\sqrt{n_1}\left(\tau_1^2 + \frac{n_1}{n_2}\tau_2^2\right)^{-1/2}\hat\lambda_2 \xrightarrow{d} N(0,1),$$
$$\hat\lambda_1 = -\frac{m_2}{m_1}\hat\lambda_2 + o_p(n_1^{-1/2}),$$

where $c = \sigma_2^2 + \sigma_1^2 m_2/m_1$ and $\gamma = \lim_{n_1,n_2\to\infty} m_2/m_1$.

Proof. Denote

$$Q_1(\mu_t, \lambda_1, \lambda_2) = \sum_{i=r_1}^{s_1}\frac{w_1(X_{(i)}, \mu_t, \Delta)}{1+\lambda_1 w_1(X_{(i)}, \mu_t, \Delta)}, \quad Q_2(\mu_t, \lambda_1, \lambda_2) = \sum_{j=r_2}^{s_2}\frac{w_2(Y_{(j)}, \mu_t, \Delta)}{1+\lambda_2 w_2(Y_{(j)}, \mu_t, \Delta)},$$
$$Q_3(\mu_t, \lambda_1, \lambda_2) = \sum_{i=r_1}^{s_1}\frac{-\lambda_1}{1+\lambda_1 w_1(X_{(i)}, \mu_t, \Delta)} + \sum_{j=r_2}^{s_2}\frac{-\lambda_2}{1+\lambda_2 w_2(Y_{(j)}, \mu_t, \Delta)}.$$

Then

$$Q_i(\hat\mu_t, \hat\lambda_1, \hat\lambda_2) = 0 \text{ for } i = 1, 2, 3.$$

By Taylor expansion around $(\mu_{\alpha\beta 1}, 0, 0)$,

$$0 = Q_i(\hat\mu_t, \hat\lambda_1, \hat\lambda_2) = Q_i(\mu_{\alpha\beta 1}, 0, 0) + \frac{\partial Q_i(\mu_{\alpha\beta 1}, 0, 0)}{\partial \mu_t}(\hat\mu_t - \mu_{\alpha\beta 1}) + \frac{\partial Q_i(\mu_{\alpha\beta 1}, 0, 0)}{\partial\lambda_1}\hat\lambda_1 + \frac{\partial Q_i(\mu_{\alpha\beta 1}, 0, 0)}{\partial\lambda_2}\hat\lambda_2 + O_p(n_1^{-2\eta}),\ i = 1, 2, 3.$$

calculating the partial derivatives and using equation (4.8) of Lemma 4.3.2 gives

$$\begin{pmatrix}\hat\mu_t - \mu_{\alpha\beta 1}\\ \hat\lambda_1\\ \hat\lambda_2\end{pmatrix} = S^{-1}\begin{pmatrix}Q_1(\mu_{\alpha\beta 1}, 0, 0)\\ Q_2(\mu_{\alpha\beta 1}, 0, 0)\\ 0\end{pmatrix} + o_p(n_1^{-1/2}),$$

where

$$S = \begin{pmatrix}-1 & -\sigma_1^2 & 0\\ -1 & 0 & -\sigma_2^2\\ 0 & -1 & -m_2/m_1\end{pmatrix}.$$

Thus we have

$$\begin{aligned}
\hat\mu_t - \mu_{\alpha\beta 1} &= -\frac{1}{c}(\sigma_2^2 Q_1(\mu_{\alpha\beta 1}, 0, 0) + \frac{m_2}{m_1}\sigma_1^2 Q_2(\mu_{\alpha\beta 1}, 0, 0)) + o_p(n_1^{-1/2}),\\
\hat\lambda_1 &= \frac{m_2}{m_1 c}(Q_2(\mu_{\alpha\beta 1}, 0, 0) - Q_1(\mu_{\alpha\beta 1}, 0, 0)) + o_p(n_1^{-1/2}),\\
\hat\lambda_2 &= -\frac{1}{c}(Q_2(\mu_{\alpha\beta 1}, 0, 0) - Q_1(\mu_{\alpha\beta 1}, 0, 0)) + o_p(n_1^{-1/2}).
\end{aligned}$$

Using Lemma 4.3.2 equation (4.9), we have

$$\sqrt{n_1}\begin{pmatrix}Q_1(\mu_{\alpha\beta 1}, 0, 0)\\ Q_2(\mu_{\alpha\beta 1}, 0, 0)\end{pmatrix} \xrightarrow{d} N\left(0, \begin{pmatrix}\tau_1^2 & 0\\ 0 & \kappa\tau_2^2\end{pmatrix}\right),$$

and we obtain the results of the Lemma 4.3.3.

Proof of Theorem 4.2.1. Using a Taylor expansion for $\log(1+\hat\lambda_i w_i)$ in (4.6), we have

$$\log \mathcal{R}(\Delta_0,\hat\mu_t) = -\hat\lambda_1 \sum_{i=r_1}^{s_1} w_1(X_{(i)},\hat\mu_t,\Delta_0) + \frac{1}{2}\hat\lambda_1^2 \sum_{i=r_1}^{s_1} w_1^2(X_{(i)},\hat\mu_t,\Delta_0)$$
$$-\hat\lambda_2 \sum_{j=r_2}^{s_2} w_2(Y_{(j)},\hat\mu_t,\Delta_0) + \frac{1}{2}\hat\lambda_2^2 \sum_{j=r_2}^{s_2} w_2^2(Y_{(j)},\hat\mu_t,\Delta_0) + r_1 + r_2,$$

where

$$|r_1| \le C_1 \sum_{i=r_1}^{s_1} |\hat\lambda_1 w_1(X_{(i)},\hat\mu_t,\Delta)|^3, \quad |r_2| \le C_2 \sum_{j=r_2}^{s_2} |\hat\lambda_2 w_2(Y_{(j)},\hat\mu_t,\Delta)|^3.$$

From Assumption 4.2.1 condition (A1) it follows (see, for example, [1]) that

$$\max_{r_1\le i\le s_1} |X_{(i)}| = O_p(1).$$

Thus

$$\max_{r_1\le i\le s_1} |w_1(X_{(i)},\hat\mu_t,\Delta)| = \max_{r_1\le i\le s_1} |X_{(i)} - \hat\mu_t| \le \max_{r_1\le i\le s_1} |X_{(i)}| + |\hat\mu_t| = O_p(1). \quad (4.11)$$

Using Lemma 4.3.1 with $1/2 < \eta < 1/3$, (4.8) and (4.11),

$$|r_1| \le C_1 |\lambda_1|^3 \max_{r_1\le i\le s_1} |w_1(X_{(i)},\hat\mu_t,\Delta)| \sum_{i=r_1}^{s_1} w_1(X_{(i)},\hat\mu_t,\Delta)^2$$
$$= O_p(n_1^{-3\eta}) O_p(1) O_p(n_1) = O_p(n_1^{1-3\eta}) = o_p(1),$$

and with the same arguments for r_2, we have

$$|r_2| = o_p(1).$$

From (4.2),

$$\begin{aligned} 0 &= \sum_{i=r_1}^{s_1} \frac{w_1(X_{(i)},\hat\mu_t,\Delta_0)}{1+\hat\lambda_1 w_1(X_{(i)},\hat\mu_t,\Delta_0)} \\ &= \sum_{i=r_1}^{s_1} w_1(X_{(i)},\hat\mu_t,\Delta_0)\left(1 - \hat\lambda_1 w_1(X_{(i)},\hat\mu_t,\Delta_0) + \frac{\hat\lambda_1^2 w_1^2(X_{(i)},\hat\mu_t,\Delta_0)}{1+\hat\lambda_1 w_1(X_{(i)},\hat\mu_t,\Delta_0)}\right), \end{aligned}$$

where the last term

$$\begin{aligned} \sum_{i=r_1}^{s_1} \hat\lambda_1^2 \frac{w_1^3(X_{(i)},\hat\mu_t,\Delta_0)}{1+\hat\lambda_1 w_1(X_{(i)},\hat\mu_t,\Delta_0)} &\le \hat\lambda_1^2 \max_{r_1\le i\le s_1} |w_1(X_{(i)},\hat\mu_t,\Delta_0)| \sum_{i=r_1}^{s_1} \frac{w_1^2(X_{(i)},\hat\mu_t,\Delta_0)}{1+\hat\lambda_1 w_1(X_{(i)},\hat\mu_t,\Delta_0)} \\ &= O_p(n_1^{-2\eta}) O_p(1) O_p(n_1) = O_p(n_1^{1-2\eta}). \end{aligned}$$

Using the same arguments for λ_2, we have

$$\sum_{i=r_1}^{s_1} w_1(X_{(i)},\hat\mu_t,\Delta_0) = \hat\lambda_1 \sum_{i=r_1}^{s_1} w_1^2(X_{(i)},\hat\mu_t,\Delta_0) + O_p(n_1^{1-2\eta}), \quad (4.12)$$

$$\sum_{j=r_2}^{s_2} w_2(Y_{(j)},\hat\mu_t,\Delta_0) = \hat\lambda_2 \sum_{j=r_2}^{s_2} w_2^2(Y_{(j)},\hat\mu_t,\Delta_0) + O_p(n_1^{1-2\eta}). \quad (4.13)$$

Multiplying the both sides of (4.12) and (4.13) with $\hat{\lambda}_1$ and $\hat{\lambda}_2$, respectively, by Lemma 4.3.1 we have

$$\hat{\lambda}_1 \sum_{i=r_1}^{s_1} w_1(X_{(i)}, \hat{\mu}_t, \Delta_0) = \hat{\lambda}_1^2 \sum_{i=r_1}^{s_1} w_1^2(X_{(i)}, \hat{\mu}_t, \Delta_0) + o_p(1),$$
$$\hat{\lambda}_2 \sum_{j=r_2}^{s_2} w_2(Y_{(j)}, \hat{\mu}_t, \Delta_0) = \hat{\lambda}_2^2 \sum_{j=r_2}^{s_2} w_2^2(Y_{(j)}, \hat{\mu}_t, \Delta_0) + o_p(1).$$

Hence,

$$-2 \log \mathcal{R}(\Delta_0, \hat{\mu}_t) = \hat{\lambda}_1^2 \sum_{i=r_1}^{s_1} w_1^2(X_{(i)}, \hat{\mu}_t, \Delta_0) + \hat{\lambda}_2^2 \sum_{j=r_2}^{s_2} w_2^2(Y_{(j)}, \hat{\mu}_t, \Delta_0) + o_p(1).$$

Recalling (4.8), (4.9), (4.10) and Lemma 4.3.3, we have

$$\begin{aligned} -2 \log \mathcal{R}(\Delta_0, \hat{\mu}_t) &= m_1 \hat{\lambda}_1^2 \sigma_1^2 + m_2 \hat{\lambda}_2^2 \sigma_2^2 + o_p(1) \\ &= m_1 \left(\frac{m_2}{m_1} \right)^2 \hat{\lambda}_2^2 \sigma_1^2 + m_2 \hat{\lambda}_2^2 \sigma_2^2 + o_p(1) \\ &= m_2 \hat{\lambda}_2^2 \left(\sigma_2^2 + \frac{m_2}{m_1} \sigma_1^2 \right) + o_p(1) \\ &= \frac{m_2}{c n_1} (\sqrt{n_1} c \hat{\lambda}_2)^2 + o_p(1). \end{aligned}$$

Finally, using Lemma 4.3.3, we obtain as $n_1, n_2 \to \infty$

$$-2 \frac{c n_1}{m_2} \left(\tau_1^2 + \frac{n_1}{n_2} \tau_2^2 \right)^{-1} \mathcal{R}(\Delta_0, \hat{\mu}_t) \xrightarrow{d} [N(0,1)]^2$$

and

$$-2 \frac{n_1 n_2 (m_2 \sigma_1^2 + m_1 \sigma_2^2)}{m_1 m_2 (n_2 \tau_1^2 + n_1 \tau_2^2)} \log \mathcal{R}(\Delta_0, \hat{\mu}_t) = -2 a_2 \log \mathcal{R}(\Delta_0, \hat{\mu}_t) \xrightarrow{d} \chi_1^2.$$

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 results described in this chapter have been previously published in M. Delesa-Vēliņa et al. [51].

Consider the problem of comparing multiple populations: let $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{in_i})$, $i = 1, 2, \dots, k$, 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

$$F = \frac{\sum_{i=1}^{k} n_i(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2/(k-1)}{\sum_{i=1}^{k}(n_i - 1)s_i^2/(N-k)},$$

where

$$\bar{Y}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij} \text{ and } s_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i\cdot})^2$$

are the sample mean and the sample variance of the ith group, respectively, and

$$\bar{Y}_{\cdot\cdot} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}/N$$

is the pooled sample mean. The null hypothesis in (5.1) is rejected at level c if $F > F_{c,k-1,N-k}$, where $F_{c,k-1,N-k}$ is the critical value based on the F distribution with $k-1$ and $N-k$ degrees of freedom.

It is well known that the classical ANOVA F-test can not handle the variance heterogeneity and problems of controlling the probability of type I error arise. Various methods have been proposed to deal with the heterogeneity problem, for example, B. L. Welch [55] 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, [56]). K. Yuen [60] suggested a robust modification to the Welch's test, using trimmed means and Winsorized variances together with Welch

ADF statistics. It was demonstrated in [23] that such approach offers a better control over the probability of type I error of one-way ANOVA under distributions of various degree of skewness and unbalanced sample sizes.

A. B. Owen [30] proposed an empirical likelihood-based ANOVA method for independent groups to test the hypothesis of equality of means. Later, [45] considered an EL-based estimator of the common mean stemming from the EL ANOVA method, and showed in a simulation setting that it is more efficient than the parametric estimator under the variance heterogeneity.

We take advantage of the good robustness properties of the trimmed means and propose an EL-based ANOVA type method to test the hypothesis of equality of the trimmed means. We first present A. B. Owen's EL ANOVA method [30] in Chapter 5.1. We then proceed with the main result on comparing multiple population trimmed means in the EL setting in Chapter 5.2.

5.1 Empirical likelihood ANOVA

For an empirical likelihood approach to ANOVA, we follow the display in [30]. Let observations $Y_{ij} \in \mathbb{R}$, where $i = 1, \dots, k$, $j = 1, \dots, n_i$ and $N = \sum_{i=1}^{k} n_i$ denotes the total number of observations.

A. B. Owen [30] provides two alternative formulations for the empirical likelihood ratio function in ANOVA context. Under the first, suppose $Y_{ij} \sim F_{i0}$ are independent samples from k different distributions. Let F_i denote a candidate for the true unknown distribution F_{i0}, and $v_{ij} = F_i(\{Y_{ij}\})$ denote the probability of Y_{ij} under F_i. The likelihood ratio function is defined by

$$R_k(F_1, \dots, F_k) = \prod_{i=1}^{k} \prod_{j=1}^{n_i} n_i v_{ij}. \tag{5.2}$$

Under the second formulation, consider N random pairs (I, Y), where $I \in \{1, \dots, k\}$ 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 empirical likelihood ratio function can be then expressed as

$$\begin{aligned} R(F) &= \prod_{i=1}^{k} \prod_{j=1}^{n_i} N v_{i\cdot} v_{j|i} \qquad (5.3) \\ &= \left(\prod_{i=1}^{k} \left(\frac{N v_{i\cdot}}{n_i} \right)^{n_i} \right) \left(\prod_{i=1}^{k} \prod_{j=1}^{n_i} n_i v_{j|i} \right). \end{aligned}$$

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 (5.3) becomes equal to one and the maximization of $R(F)$ is only subject to constraints on $v_{j|i}$. Thus the both approaches (5.2) and (5.3) lead to the same likelihood ratio function,

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

The advantage of the second approach is that a triangular version of the empirical likelihood theorem can be used to establish the inference for ANOVA type hypotheses. Under the triangular empirical likelihood theorem, we consider the mean of random variables that are not necessarily i.i.d, but can be arranged into a triangular array structure $Z_{in} \in \mathbb{R}^p$, $i = 1, \ldots, n$.

Theorem 5.1.1. *(Triangular array ELT) [31, Theorem 4.1.] Let $Z_{in} \in \mathbb{R}_p$, $1 \leq i \leq n$, $n \geq n_{min}$ be a triangular array of random vectors. For each n, suppose that $Z_{1n}, \ldots, Z_{nn}$ are independent and have common mean μ_n. Let H_n denote the convex hull of $Z_{1n}, \ldots, Z_{nn}$, and $V_n = (1/n)\sum_{i=1}^n \mathrm{Var}(Z_{in})$. Put $\sigma_{1n} = \mathrm{maxeig}(V_n)$, and $\sigma_{pn} = \mathrm{mineig}(V_n)$. Assume that as $n \to \infty$*

$$P(\mu_n \in H_n) \to 1$$

and

$$\frac{1}{n^2}\sum_{i=1}^{n} E\left(||Z_{in} - \mu_n||^4/\sigma_{1n}^2\right) \to 0,$$

and that for some $c > 0$ and all $n \geq n_{min}$,

$$\frac{\sigma_{pn}}{\sigma_{1n}} \geq c. \tag{5.4}$$

Then, as $n \to \infty$,

$$-2\log \mathcal{R}(\mu_n) \xrightarrow{d} \chi_p^2,$$

where

$$\mathcal{R}(\mu_n) = \max_{v_i}\left\{\prod_{i=1}^{n} nv_i | \sum_{i=1}^{n} v_i(Z_{in} - \mu_n) = 0, v_i \geq 0, \sum_{i=1}^{n} v_i = 1\right\}.$$

For the proof of Theorem 5.1.1, see [31, Chapter 11.3].

To apply Theorem 5.1.1 in ANOVA setting, suppose

$$\mu_{i0} = \int y dF_{i0}(y) \in \mathbb{R}$$

and define

$$\mathcal{R}(\mu_1, \ldots, \mu_k) = \sup_{v_{ij}}\left\{\prod_{i=1}^{k}\prod_{j=1}^{n_i} Nv_{ij} \,|\, v_{ij} \geq 0, \sum_{i=1}^{k}\sum_{j=1}^{n_j} v_{ij} = 1, \right.$$
$$\left. \sum_{j=1}^{n_i} v_{ij}(Y_{ij} - \mu_i) = 0, i = 1, \ldots, k\right\}.$$

Define the auxiliary variables $Z_{ijN} \in \mathbb{R}^k$

$$Z_{ijN} = (0, \ldots, 0, Y_{ij}^T - \mu_i^T, 0, \ldots, 0)^T,$$

where $Y_{ij}^T - \mu_i^T$ are preceded by $(i-1)$ and followed by $(k-i)$ zeros. The matrix V_n in Theorem 5.1.1 is then given by

$$V_N = \frac{1}{N}\begin{pmatrix} n_1\mathrm{Var}(Y_{11}) & 0 & \dots & 0 \\ 0 & n_2\mathrm{Var}(Y_{21}) & \dots & 0 \\ \vdots & \vdots & \ddots & \dots \\ 0 & 0 & \dots & n_k\mathrm{Var}(Y_{k1}) \end{pmatrix}.$$

If each $\mathrm{Var}(Y_{i1})$ is finite and non-singular, the condition (5.4) of Theorem 5.1.1 on eigenvalues holds as long as

$$\lim_{N\to\infty} \frac{\min_i n_i}{\max_i n_i} > 0.$$

The convex hull condition for Z_{ijN} unites k convex hull conditions: for each $i = 1, \dots, k$, the convex hull of Y_{ij} needs to contain μ_{i0}. Thus, under mild conditions

$$-2\log \mathcal{R}(\mu_1, \dots, \mu_k) \xrightarrow{d} \chi_k^2 \text{ as } n \to \infty.$$

The hypothesis (5.1) corresponds to $k-1$ constraints on vector μ_n instead of k. We arrive to the following corollary of the triangular ELT for the null hypothesis (5.1).

Corollary 5.1.1. *(EL ANOVA for the equality of means) [30, p. 1739] Suppose $E(Y_{ij}) = \mu_0$. Let*

$$\mathcal{R}(\mu) = \max_{v_{j|i}} \left\{ \prod_{i=1}^{k}\prod_{j=1}^{n_k} n_i v_{j|i} \,\middle|\, \sum_{j=1}^{n_1} v_{j|1}Y_{1j} = \dots = \sum_{j=1}^{n_k} v_{j|k}Y_{kj} = \mu, \right.$$
$$\left. \sum_{j=1}^{n_k} v_{j|i} = 1, v_{j|i} \geq 0, i = 1, \dots, k \right\} \quad (5.5)$$

and define $n_0 = \min_{1\leq i\leq k} n_i$. If $\mu = \mu_0 + O(n_0^{-1/2})$ and for each $i = 1, \dots k$, $\mathrm{Var}Y_{i1}$ 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 from Theorem 5.1.1 becomes

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

5.2 Main results

Now we present the empirical likelihood ANOVA-type method for the trimmed means.

We are interested in the null hypothesis

$$H_0^T : \mu_{\alpha\beta 1} = \mu_{\alpha\beta 2} = \ldots = \mu_{\alpha\beta k} =: \mu_{\alpha\beta}, \tag{5.6}$$

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

$$\begin{aligned} \bar{Y}_{\alpha\beta i} &= \frac{1}{m_i} \sum_{j=r_i}^{s_i} Y_{i(j)}, \\ S^2_{\alpha\beta i} &= \frac{1}{m_i} \sum_{j=r_i}^{s_i} (Y_{i(j)} - \bar{Y}_{\alpha\beta i})^2. \end{aligned}$$

Analogically to the EL ANOVA setting in (5.5), 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

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

Theorem 5.2.1. *(M. Delesa-Vēliņa et al. [51]) 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 (5.6),*

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

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

$$a_i = \sigma^2_{\alpha\beta i} / ((1-\alpha-\beta)\tau^2_{\alpha\beta i}) . \tag{5.7}$$

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. Let

$$G = \sum_{i=1}^{k}\sum_{j=r_i}^{s_i} \log v_{ij} + \sum_{i=1}^{k} \gamma_i (1 - \sum_{j=r_i}^{s_i} v_{ij}) + \sum_{i=1}^{k} m_i \lambda_i \left(\mu_{\alpha\beta} - \sum_{j=r_i}^{s_i} v_{ij} Y_{i(j)} \right).$$

We maximize G over the choice of v_{ij}, the Lagrange multipliers γ_i, λ_i, and the common trimmed mean $\mu_{\alpha\beta}$. First, note that

$$m_i \lambda_i \left(\mu_{\alpha\beta} - \sum_{j=r_i}^{s_i} v_{ij} Y_{i(j)} \right) = m_i \lambda_i \sum_{j=r_i}^{s_i} v_{ij} \left(\mu_{\alpha\beta} - Y_{i(j)} \right).$$

Consider the derivatives

$$\frac{\partial}{\partial v_{ij}} G = \frac{1}{v_{ij}} - \gamma_i + m_i \lambda_i (\mu_{\alpha\beta} - Y_{i(j)}) = 0. \tag{5.8}$$

Multiplying (5.8) by v_{ij} and summing over j, we get

$$0 = \sum_{j=r_i}^{s_i} \left\{ \frac{\partial}{\partial v_{ij}} G \cdot v_{ij} \right\} = \sum_{j=r_i}^{s_i} 1 - \sum_{j=r_i}^{s_i} v_{ij} \gamma_i + m_i \lambda_i \sum_{j=r_i}^{s_i} v_{ij} (\mu_{\alpha\beta} - Y_{i(j)}),$$

and since the last sum is zero by our constraint condition, we get $\gamma_i = m_i$. Inserting γ_i in (5.8), we have

$$v_{ij}^{-1} = m_i (1 + \lambda_i (Y_{i(j)} - \mu_{\alpha\beta})).$$

Thus, the weights $v_{ij}, i = 1, \ldots, k$, that maximize $R(\mu_{\alpha\beta})$ are given by

$$v_{ij} = \frac{1}{m_i (1 + \lambda_i (Y_{i(j)} - \mu_{\alpha\beta}))}, j = r_i, \ldots, s_i. \tag{5.9}$$

Now, using the constraint $\sum_{j=r_i}^{s_i} (v_{ij} - 1/m_i) = 0$, we get that the Lagrange multiplier λ_i, $i = 1, \ldots k$, is the solution to

$$\sum_{j=r_i}^{s_i} \frac{Y_{i(j)} - \mu_{\alpha\beta}}{1 + \lambda_i (Y_{i(j)} - \mu_{\alpha\beta})} = 0. \tag{5.10}$$

The resulting EL log likelihood function is given by

$$\log \mathcal{R}(\mu_{\alpha\beta}) = - \sum_{i=1}^{k}\sum_{j=r_i}^{s_i} \log \left(1 + \lambda_i (Y_{i(j)} - \mu_{\alpha\beta}) \right),$$

and the maximum empirical likelihood estimator of the common trimmed mean $\mu_{\alpha\beta}$ is given by $(\partial / \partial \mu_{\alpha\beta}) \log \mathcal{R}(\mu_{\alpha\beta}) = 0$. Solve

$$\begin{aligned} 0 &= \frac{\partial}{\partial \mu_{\alpha\beta}} \left\{ - \sum_{i=1}^{k}\sum_{j=r_i}^{s_i} \log \left(1 + \lambda_i (Y_{i(j)} - \mu_{\alpha\beta}) \right) \right\} \\ &= - \sum_{i=1}^{k}\sum_{j=r_i}^{s_i} \left(\frac{\partial \lambda_i}{\partial \mu_{\alpha\beta}} (Y_{i(j)} - \mu_{\alpha\beta}) - \lambda_i \right) \left(1 + \lambda_i (Y_{i(j)} - \mu_{\alpha\beta}) \right)^{-1} \\ &= - \sum_{i=1}^{k}\sum_{j=r_i}^{s_i} \frac{\partial \lambda_i}{\partial \mu_{\alpha\beta}} \left(\frac{Y_{i(j)} - \mu_{\alpha\beta}}{1 + \lambda_i (Y_{i(j)} - \mu_{\alpha\beta})} \right) + \sum_{i=1}^{k}\sum_{j=r_i}^{s_i} \frac{\lambda_i}{1 + \lambda_i (Y_{i(j)} - \mu_{\alpha\beta})}. \end{aligned} \tag{5.11}$$

The first sum in (5.11) is equal to zero due to (5.10), and using (5.9), we have

$$\sum_{i=1}^{k} \lambda_i m_i \sum_{j=r_i}^{s_i} v_{ij} = \sum_{i=1}^{k} m_i \lambda_i = 0. \tag{5.12}$$

It can be shown (see, for example, [31], Chapter 11.2), that under H_0^T,

$$\lambda_i = (\bar{Y}_{\alpha\beta i} - \mu_{\alpha\beta 0})/S^2_{\alpha\beta i} + o_p(n_i^{-1/2}).$$

Substituting this into (5.12), we have

$$\sum_{i=1}^{k} \frac{m_i(\bar{Y}_{\alpha\beta i} - \mu_{\alpha\beta 0})}{S^2_{\alpha\beta i}} = o_p(\sqrt{n_0}),$$

which in turn gives the maximum empirical likelihood estimator of $\mu_{\alpha\beta 0}$ as

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

Now, under H_0^T,

$$-2\log \mathcal{R}(\mu_{\alpha\beta 0}) \quad = \quad 2\sum_{i=1}^{k}\sum_{j=r_i}^{s_i} \log(1 + \lambda_i(Y_{i(j)} - \mu_{\alpha\beta 0})).$$

Denoting $\tilde{Y}_{i(j)} = \lambda_i(Y_{i(j)} - \mu_{\alpha\beta 0})$ and considering the Taylor expansion

$$\log(1 + \tilde{Y}_{i(j)}) = \tilde{Y}_{i(j)} - \tilde{Y}^2_{i(j)}/2 + \eta_{i(j)},$$

we have by standard EL arguments (see, for example, [31, Chapter 11.2]),

$$\begin{aligned}
-2\log \mathcal{R}(\mu_{\alpha\beta 0}) &= 2\sum_{i=1}^{k}\sum_{j=r_i}^{s_i} \tilde{Y}_{i(j)} - \sum_{i=1}^{k}\sum_{j=r_i}^{s_i} \tilde{Y}^2_{i(j)} + 2\sum_{i=1}^{k}\sum_{j=r_i}^{s_i} \eta_{i(j)} \\
&= 2\sum_{i=1}^{k} m_i\lambda_i(\bar{Y}_{\alpha\beta i} - \mu_{\alpha\beta 0}) - \sum_{i=1}^{k}\sum_{j=r_i}^{s_i} \lambda_i^2 (Y_{i(j)} - \mu_{\alpha\beta 0})^2 + o_p(1) \\
&= 2\sum_{i=1}^{k} m_i\lambda_i(\bar{Y}_{\alpha\beta i} - \mu_{\alpha\beta 0}) - \sum_{i=1}^{k} m_i \lambda_i^2 S^2_{\alpha\beta_i} + o_p(1) \\
&= 2\sum_{i=1}^{k} m_i(\bar{Y}_{\alpha\beta i} - \mu_{\alpha\beta 0})^2/S^2_{\alpha\beta i} - \sum_{i=1}^{k} m_i(\bar{Y}_{\alpha\beta i} - \mu_{\alpha\beta 0})^2/S^2_{\alpha\beta i} + O_p(n_0^{-1/2}) \\
&= \sum_{i=1}^{k} m_i(\bar{Y}_{\alpha\beta i} - \mu_{\alpha\beta 0})^2/S^2_{\alpha\beta i} + O_p(n_0^{-1/2}).
\end{aligned} \tag{5.13}$$

Now, consider

$$l_i = m_i(\bar{Y}_{\alpha\beta i} - \mu_{\alpha\beta 0})^2/S^2_{\alpha\beta i}.$$

For each i, given F_{i0} continuous, $F'_{i0}(\xi_\alpha) > 0$ and $F'_{i0}(\xi_{1-\beta}) > 0$, we have by Lemma 4.3.2

$$\sqrt{n_i}(\bar{Y}_{\alpha\beta i} - \mu_{\alpha\beta 0})^2 \xrightarrow{d} N(0, \tau^2_{\alpha\beta i})$$

and

$$S^2_{\alpha\beta i} \xrightarrow{p} \sigma^2_{\alpha\beta i}.$$

Using $m_i/n_i \xrightarrow{a.s.} (1-\alpha-\beta)$, we have for each l_i, $i = 1, \ldots, k$,

$$\frac{\sigma^2_{\alpha\beta i}}{(1-\alpha-\beta)\tau^2_{\alpha\beta i}} l_i = a_i l_i \xrightarrow{d} \chi^2_1$$

as $n_0 \to \infty$. Finally, consider $\sum_{i=1}^{k} a_i l_i$ and note that under H_0^T (5.6) there are $(k-1)$ constraints on the trimmed means instead of k (cf. Corollary 5.1.1). Thus we acquire

$$\sum_{i=1}^{k} a_i l_i \xrightarrow{d} \chi^2_{k-1}$$

as $n_0 \to \infty$, which proves the theorem.

Chapter 6

Simulation and data analysis results

The aim of this chapter is to analyse the performance of the newly-established empirical likelihood methods for robust location estimators presented in Chapters 3 - 5. In particular, we are interested in exploring the performance of the methods 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) will be investigated in a simulation setting regarding the ability to control the type I error and the power of the tests. Examples of applications to real data sets will be provided as well. The newly-established methods will be compared to well-known methods of classical and robust statistics.

The results are organized as follows: Chapter 6.1 deals with the performance of the two-sample methods presented in Chapters 3 - 4. Chapter 6.2 considers the performance of the EL ANOVA method for the trimmed means presented in Chapter 5. Chapter 6.3 shows the application of the methods to some real data sets. Chapter 6.4 presents an overall discussion on the simulation and data analysis results.

The performance of the newly-established EL methods has been analysed before in M. Delesa-Vēliņa et al. [51, 52, 8]. In this chapter the conclusions drawn before are recapitulated and some further comparative analysis is carried out.

6.1 Simulation study for comparing two populations

In this chapter we are interested in the hypotheses

$$H_0 : \Delta_0 = 0,\ H_1 : \Delta_0 \neq 0,$$

where Δ_0 is the difference of two unknown location parameters of interest associated with populations F_1 and F_2, respectively. We consider (i) the difference of two trimmed means, (ii) the difference of two smoothed Huber estimators and, for comparison, (iii) the difference of two means. When comparing the performance of the tests based on these three estimators, it is important to be aware that different hypotheses are considered. Comparison of the tests in known simulation conditions and a careful interpretation of the results is essential to choose an appropriate method in a real-world situation involving data from unknown underlying populations.

The performance of methods (i), (ii) and (iii) has been analysed in detail in the author's publications [52] and [8]. However, it did not include a comparison between the methods based on (i) and (ii), which will be done in this chapter.

6.1.1 EL method for the difference of two smoothed Huber estimators

In M. Delesa-Vēliņa et al. [52], the EL method for the difference of two smoothed Huber estimators was analysed. For the simulation study symmetric double exponential and Huber's least favourable distributions were considered. Note that Lemma 3.2.2 stipulates that the asymptotic results of Theorem 3.2.1 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 (two cases: 50 and 100) both 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 under the departure from H_0. The results of the study are presented in Tables 6.1 - 6.6.

The estimation of the smoothed Huber estimator (2.18) requires an estimate of the asymptotic variance V of the initial non-smooth M-estimator. According to [13], V should behave as a tuning parameter, since it is closely related to k through the equation (2.10). Using simulation analysis, we discovered that this is true for symmetric distributions. However, for asymmetric distributions, we have found that the value of V significantly influences the results.

Thus, regarding the variance V, two situations were distinguished in [52]: first, V was set equal to 2.046 as recommended in [13] (panel *ELHubVF*), and second, V was estimated for the particular distribution using Monte Carlo simulations (panel *ELHubVE*). The asymptotic variances of the non-smoothed Huber estimator under F_1 and F_2 are reported in tables as V_1 and V_2, respectively. MAD was used as a preliminary estimator of the scale parameter σ of the underlying distribution, as required by Lemma 3.2.2. For comparison, hypothesis tests regarding the difference of means were included, namely, Student's t-test (panel t) and the empirical likelihood test for the difference of means of Example 1.4.1 (panel *EL Means*). The results for the EL-based methods were computed using R package *EL* [6], functions *EL.means* and *EL.Huber*.

Table 6.1: Simulated 95% quantiles of the test statistic $-2\log R(\Delta_0, \hat{\theta}, \hat{\sigma}_1, \hat{\sigma}_2)$ under H_0 for various distributions: $F_1 = \text{Gamma}(a = 5; s = 1)$, $F_2 = \text{Gamma}(a = 1; s = 1/5)$ (Gamma model); $F_1 = F_2 = \text{doublexp}(0; 1)$ (doublexp model); $F_1 = F_2 = \text{Hlf}(0; 1)$ (Hlf model) using 10,000 replications

n	Gamma			doublexp			Hlf		
	EL Means	EL HubVF	EL HubVE	EL Means	EL HubVF	EL HubVE	EL Means	EL HubVF	EL HubVE
50	4.176	5.832	4.169	4.085	4.014	3.965	3.890	3.841	3.870
100	3.952	6.464	3.998	3.951	3.925	3.936	3.861	3.895	3.874
500	3.875	10.676	3.895	3.837	3.853	3.853	3.770	3.764	3.813
1000	3.897	14.766	3.948	3.725	3.696	3.706	3.766	3.794	3.807

In Table 6.1, the simulated 95% quantile of the test statistic $-2\log \mathcal{R}(\Delta_0, \hat{\theta}, \hat{\sigma}_1, \hat{\sigma}_2)$ under H_0 for various distributions for the three EL-based methods under $H_0 : \Delta_0 = 0$ is reported. The simulated 95% quantile should converge to the respective χ^2_1 quantile, which is approximately equal to 3.841. We can see that for the double exponential, and Huber's least favourable distributions, the simulated quantile is close to the theoretical one. Regarding the gamma distribution, the quantile diverges rapidly for the smoothed Huber estimator with fixed V.

Table 6.2: Empirical coverage of 95% confidence intervals for $\Delta_0 = 0$. $F_1 = \text{doublexp}(0, \sigma)$; $F_2 = \text{doublexp}(0, 1)$, $n = 100$. The asymptotic variance of the non-smooth Huber estimator under F_1 and F_2 simulated using 10,000 replications is reported as V_1 and V_2, respectively. The tests considered are Student's t-test (t), EL test for the difference of means (*EL Means*), EL test for the difference of two smoothed Huber estimators with V_1, V_2 set to constant 2.046 (*ELHubVF*), and EL test for the difference of two smoothed Huber estimators with V_1, V_2 estimated by Monte Carlo simulations (*ELHubVE*). The scale estimator of Huber estimators is MAD.

σ	V_1	V_2	t	EL Means	ELHubVF	ELHubVE
0.1	0.01	1.29	0.951	0.945	0.947	0.947
0.2	0.05	1.29	0.952	0.947	0.949	0.948
0.5	0.33	1.29	0.953	0.949	0.950	0.950
1	1.31	1.29	0.951	0.948	0.949	0.948
2	5.25	1.29	0.950	0.947	0.948	0.948
5	32.8	1.29	0.952	0.947	0.946	0.947

Table 6.3: Empirical coverage of 95% confidence intervals for $\Delta_0 = 0$. $F_1 = \text{Hlf}(0, \sigma)$, $F_2 = \text{Hlf}(0, 1)$, $n = 100$. The asymptotic variance of the non-smooth Huber estimator under F_1 and F_2 simulated using 10,000 replications is reported as V_1 and V_2, respectively. The tests considered are Student's t-test (t), EL test for the difference of means (*EL Means*), EL test for the difference of two smoothed Huber estimators with V_1, V_2 set to constant 2.046 (*ELHubVF*), and EL test for the difference of two smoothed Huber estimators with V_1, V_2 estimated by Monte Carlo simulations (*ELHubVE*). The scale estimator of Huber estimators is MAD.

σ	V_1	V_2	t	EL Means	ELHubVF	EL HubVE
0.1	0.02	2.07	0.949	0.943	0.946	0.946
0.2	0.08	2.07	0.950	0.945	0.948	0.948
0.5	0.52	2.07	0.949	0.947	0.948	0.948
1	2.06	2.07	0.952	0.948	0.951	0.951
2	8.25	2.07	0.949	0.945	0.949	0.948
5	51.6	2.07	0.951	0.945	0.949	0.946

The results on the empirical coverage in [52] were as follows.:

1. For the symmetrical distributions, double exponential and Huber's least favourable distribution from Tables 6.2 and 6.3, respectively, all methods give similar results, the empirical coverage being close to the nominal 95%. This holds regardless of the degree of the variance heterogeneity.

2. Regarding the uncontaminated gamma distribution from Table 6.4), 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, see Tables 6.5 and 6.6, respectively, the new EL method based on the V estimated has overall better empirical coverage than Student's t-test and EL test for the difference of means.

Table 6.4: Empirical coverage of 95% confidence intervals for $\Delta_0 = 0$ when $F_1 = \text{Gamma}(a = \sigma, s = 1)$; $F_2 = \text{Gamma}(a = 1, s = 1/\sigma)$. The asymptotic variance of the non-smooth Huber estimator under F_1 and F_2 simulated using 10,000 replications is reported as V_1 and V_2, respectively. The tests considered are Student's t-test (t), EL test for the difference of means (*EL Means*), EL test for the difference of two smoothed Huber estimators with V_1, V_2 set to constant 2.046 (*ELHubVF*), and EL test for the difference of two smoothed Huber estimators with V_1, V_2 estimated by Monte Carlo simulations (*ELHubVE*). The scale estimator of Huber estimators is MAD.

σ	n	V_1	V_2	t	EL Means	ELHubVF	ELHubVE
3	50	3.3	7.3	0.942	0.938	0.894	0.935
	100			0.946	0.944	0.879	0.941
4	50	4.1	12.7	0.945	0.938	0.894	0.938
	100			0.946	0.945	0.867	0.943
5	50	5.8	20.2	0.944	0.941	0.887	0.941
	100			0.945	0.946	0.862	0.946
6	50	6.8	28.9	0.941	0.935	0.883	0.936
	100			0.944	0.944	0.858	0.944
7	50	8.2	39.2	0.941	0.938	0.881	0.938
	100			0.948	0.950	0.851	0.949
10	50	11.8	81.1	0.939	0.937	0.877	0.937
	100			0.943	0.944	0.840	0.944
20	50	24.6	318.3	0.936	0.935	0.866	0.935
	100			0.945	0.945	0.832	0.945

Table 6.5: Empirical coverage of 95% confidence intervals when $F_1 = (1-\epsilon)\text{Gamma}(a = \sigma, s = 1) + \epsilon\text{Unif}[0, 50]$; $F_2 = \text{Gamma}(a = 1, s = 1/\sigma)$ with $\epsilon = 0.06$. The asymptotic variance of the non-smooth Huber estimator under F_1 and F_2 simulated using 10,000 replications is reported as V_1 and V_2, respectively. The tests considered are Student's t-test (t), EL test for the difference of means (*EL Means*), EL test for the difference of two smoothed Huber estimators with V_1, V_2 set to constant 2.046 (*ELHubVF*), and EL test for the difference of two smoothed Huber estimators with V_1, V_2 estimated by Monte Carlo simulations (*ELHubVE*). The scale estimator of Huber estimators is MAD.

σ	n	V_1	V_2	t	EL Means	ELHubVF	ELHubVE
3	50	4.2	8.6	0.839	0.646	0.781	0.842
	100			0.564	0.405	0.668	0.757
4	50	5.7	14.9	0.845	0.723	0.779	0.841
	100			0.632	0.534	0.649	0.757
5	50	7.0	23.4	0.852	0.780	0.776	0.845
	100			0.695	0.644	0.657	0.768
6	50	8.3	34.5	0.859	0.818	0.773	0.856
	100			0.742	0.714	0.659	0.787
7	50	9.7	46.6	0.871	0.853	0.782	0.870
	100			0.787	0.781	0.663	0.820
10	50	13.0	93.5	0.893	0.901	0.785	0.904
	100			0.854	0.869	0.666	0.876
20	50	24.9	375.1	0.927	0.930	0.803	0.930
	100			0.935	0.942	0.709	0.942

4. The method based on $V = 2.046$ gives inconsistent results for the skewed gamma distribution with or without contamination (see Tables 6.4 - 6.6).

Table 6.6: Empirical coverage of 95% confidence intervals for $\Delta_0 = 0$ when $F_1 = (1 - \epsilon)\text{Gamma}(a = \sigma, s = 1) + \epsilon\text{Unif}[0, 50]$; $F_2 = \text{Gamma}(a = 1, s = 1/\sigma)$ with $\epsilon = 0.2$. The asymptotic variance of the non-smooth Huber estimator under F_1 and F_2 simulated using 10,000 replications is reported as V_1 and V_2, respectively. The tests considered are Student's t-test (t), EL test for the difference of means (*EL Means*), EL test for the difference of two smoothed Huber estimators with V_1, V_2 set to constant 2.046 (*ELHubVF*), and EL test for the difference of two smoothed Huber estimators with V_1, V_2 estimated by Monte Carlo simulations (*ELHubVE*). The scale estimator of Huber estimators is MAD.

σ	n	V_1	V_2	t	EL Means	ELHubVF	ELHubVE
3	50	8.7	8.0	0.146	0.056	0.353	0.108
	100			0.005	0.002	0.109	0.007
4	50	10.2	13.5	0.211	0.119	0.405	0.162
	100			0.016	0.010	0.156	0.017
5	50	13.4	21.3	0.285	0.199	0.459	0.232
	100			0.037	0.026	0.206	0.037
6	50	16.0	31.2	0.373	0.307	0.511	0.328
	100			0.076	0.065	0.253	0.078
7	50	16.8	42.0	0.448	0.404	0.556	0.420
	100			0.136	0.125	0.299	0.140
10	50	22.0	82.6	0.653	0.655	0.659	0.660
	100			0.401	0.417	0.450	0.424
20	50	36.8	344.3	0.911	0.924	0.834	0.924
	100			0.908	0.924	0.760	0.925

Table 6.7: Simulated empirical power for H_0: $\Delta_0 = 0$ when F_1 = (1-ϵ)Gamma(a = 5; s =1)+ $\epsilon Unif[0, 50]$; F_2 = Gamma(a = 1; s = $1/\sigma$), $\epsilon = 0$ and $\epsilon = 0.06$. MAD is used as the scale parameter for the smoothed Huber M-estimator. For Huber variance estimation, $k = 1.35$ and $V = 2.046$ are used with 10,000 replications.

		$\epsilon = 0$				$\epsilon = 0.06$			
σ	n	t	EL Means	EL HubVF	EL HubVE	t	EL Means	EL HubVF	EL HubVE
5	50	0.055	0.049	0.048	0.049	0.160	0.203	0.102	0.143
	100	0.057	0.057	0.051	0.056	0.309	0.354	0.131	0.227
6	50	0.155	0.206	0.075	0.211	0.049	0.068	0.038	0.063
	100	0.314	0.375	0.097	0.379	0.049	0.061	0.033	0.062
7	50	0.466	0.551	0.296	0.559	0.113	0.117	0.162	0.183
	100	0.809	0.855	0.521	0.857	0.151	0.138	0.267	0.315
8	50	0.769	0.835	0.603	0.839	0.264	0.249	0.414	0.422
	100	0.980	0.987	0.875	0.988	0.436	0.397	0.682	0.699
9	50	0.923	0.956	0.828	0.957	0.457	0.416	0.658	0.642
	100	0.999	0.999	0.982	0.999	0.725	0.692	0.916	0.908

Finally, we performed an empirical power analysis similar to [25]. We generated 10,000 samples from $F_1 = (1 - \epsilon)\text{Gamma}(a = 5; s = 1) + \epsilon\text{Unif}[0, 50]$ and $F_2 = \text{Gamma}(a = 1; s = 1/\sigma)$ distributions. For the approximation of the null distribution, $\epsilon = 0$ and $\sigma = 5$ were chosen. Thus, for the power analysis, we used the simulated critical values from Table 6.7 (panel *Gamma*) along with the different values of the parameters $\epsilon \in \{0, 0.06\}$ and $\sigma \in \{5, 6, 7, 8, 9\}$.

The results of the power analysis are presented in Table 6.7. In the case with no contamination (panel $\epsilon = 0$), the EL method for the difference of the means and the EL method for the difference of two smoothed Huber estimators (with V simulated) have similar power and outperform the t-test. When the parameter $V = 2.046$, the EL method for the difference of smoothed Huber estimators has substantially lower power.

In the case involving uniform contamination (panel $\epsilon = 0.06$), we see that the EL method for the difference of two means is not very robust and has power similar to t test. The EL method for two smoothed Huber estimators with simulated V has the highest power, slightly outperforming the EL method when V is fixed. However, these results have to be viewed in light of Table 6.1, where the 5% critical values diverge for larger sample sizes when V is fixed.

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 [13]. 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.

6.1.2 EL method for the difference of two trimmed means

The EL-based method for the comparison of two trimmed means was considered in detail in M. Delesa-Vēliņa et al. [8]. The simulation study involved the following various aspects of violation of the classical assumptions: underlying distributions of various shapes (heavy tails, outliers, skewness), as well as unbalanced sample sizes and variance heterogeneity combined. A simulation study was designed to evaluate the performance of the method in terms of the empirical level and power under various sample sizes.

First, the empirical level (empirical probability of a type I error) of the test was explored as a function of the sample size for two equal underlying distributions. $N = 5,000$ data sets of sample sizes $n_1 \in \{9, 12, 15, \ldots, 75\}$ were generated and the results of empirical rejection rates at $\alpha = 0.05$ nominal significance level were considered. Six types of distributions were considered: the standard normal distribution, heavy-tailed t_2-distribution, skewed χ^2_3 and χ^2_1 distributions, as well as two contaminated normal distributions, $0.95N(0;1) + 0.05N(0;25)$ and $0.9N(0;1) + 0.05N(0;100)$. A similar approach was used in [9]. Two sample size scenarios were considered: equal sample sizes $n_1 = n_2$ and the case where the second sample size is double the first, $n_2 = 2n_1$.

Next, the power of the asymptotic tests under various location differences Δ_0 was investigated, 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 an approach has been previously used in [9]. Note that in the case of the standard normal distribution, δ is equal to 1. For the power analysis, the same distributions were considered as for the empirical level analysis. Both balanced sample size $n_1 = n_2 = 50$ and unbalanced sample size $n_1 = 50$, $n_2 = 100$ scenarios were considered. Finally, the robustness of the tests under the normal and the skewed χ^2_3 distribution with various degrees of variance heterogeneity and unbalanced sample sizes was explored, similarly as in [23] and [21].

Methods included for the comparison were: Student's t-test, Welch's test [54] and EL test (Example 1.4.1) – for the comparison of means; Yuen's test [60] with bootstrap-t approximation [57, 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. The test statistics of the alternative methods considered in the simulation study can be found in the references given. Regarding the tests for the trimmed means, two trimming versions were considered: 10% and 20% trimming. The EL test for the difference of two smoothed Huber estimators was not included in the study [8], but is added to the results below. Regarding the smoothed Huber estimators, two versions of the test – with the asymptotic variance V of the initial non-smoothed Huber estimator fixed to 2.046 (panel *ELHubVF*) and V estimated by 10,000 Monte Carlo simulations (panel *ELHubVE*) – will be considered. We exclude the comparison with the EL ANOVA method for the trimmed means from the results below and comment on it in Chapter 6.2.

The results for the EL-based methods were computed using *R* package *EL* [6], functions *EL.means* and *EL.Huber*. For the Yuen's test, the *R* package *WRS2* [24] function *yuenbt* was used.

Simulations of the empirical level

Consider the results of the empirical level simulations in Figures 6.1 - 6.2. Horizontal dotted lines have been added to 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 the standard normal distribution, Student's t-test, Welch's test and Yuen's test simulated level are close to the nominal already for very small samples ($n = 9$). The EL-based tests converge (i.e., reach the nominal test level) for the total sample size $n_1+n_2 > 40$. Note, that the two versions of the test for the difference of two smoothed Huber estimators yield identical results.

For the heavy-tailed t_2-distribution, EL tests for the trimmed means converge approximately for $n_1 + n_2 > 70$. Tests for the difference of two smoothed Huber estimators converge slightly faster, at $n_1 + n_2 > 50$. The EL tests based on the trimmed means converge faster with 10% than with 20% trimming. The EL test for the means is oversized even for large samples. Student's t-test and Welch's test are somewhat undersized even for large samples, and Yuen's test is slightly undersized for very small samples.

Regarding the 5% contaminated normal distribution, all the tests converge for $n_1 + n_2 > 60$, except the EL test for the means which is oversized. Similarly, as for the $N(0,1)$ distribution, the EL test for 10% trimmed means converges faster than the EL test for the 20% trimmed means. For small samples, the test for two smoothed Huber estimators is slightly oversized, while the Student's t-test and Welch's tests are undersized.

Regarding the 10% contaminated normal distribution, the pattern is somewhat similar to that of the t_2 distribution. However, Student's t and Welch's test are considerably undersized for small sample sizes. The test for the difference of smoothed Huber estimators is slightly oversized and converges only for $n_1 + n_2 > 100$.

For the skewed χ^2_3 distribution, the EL tests for 10% trimmed means converge for $n_1+n_2 > 40$, while the EL test for the 20% trimmed means and the EL test for the means converge for $n_1+n_2 > 80$. The test for smoothed Huber estimators with V estimated acts similarly to EL test for the means, both being slightly oversized even for large sample sizes. However, the version with V fixed does not converge to the nominal level and is considerably oversized. Yuen's test is slightly undersized for very small samples and converges for $n_1 + n_2 > 40$.

Finally, for the very skewed χ^2_1 distribution, tests require larger sample sizes to converge, most tests converging for $n_1 + n_2 > 80$. The EL test is slightly oversized even for large samples, but tests based on Huber estimators have a wrong empirical level fluctuating around 0.15 for any sample size. Interestingly, the level of EL test for the trimmed

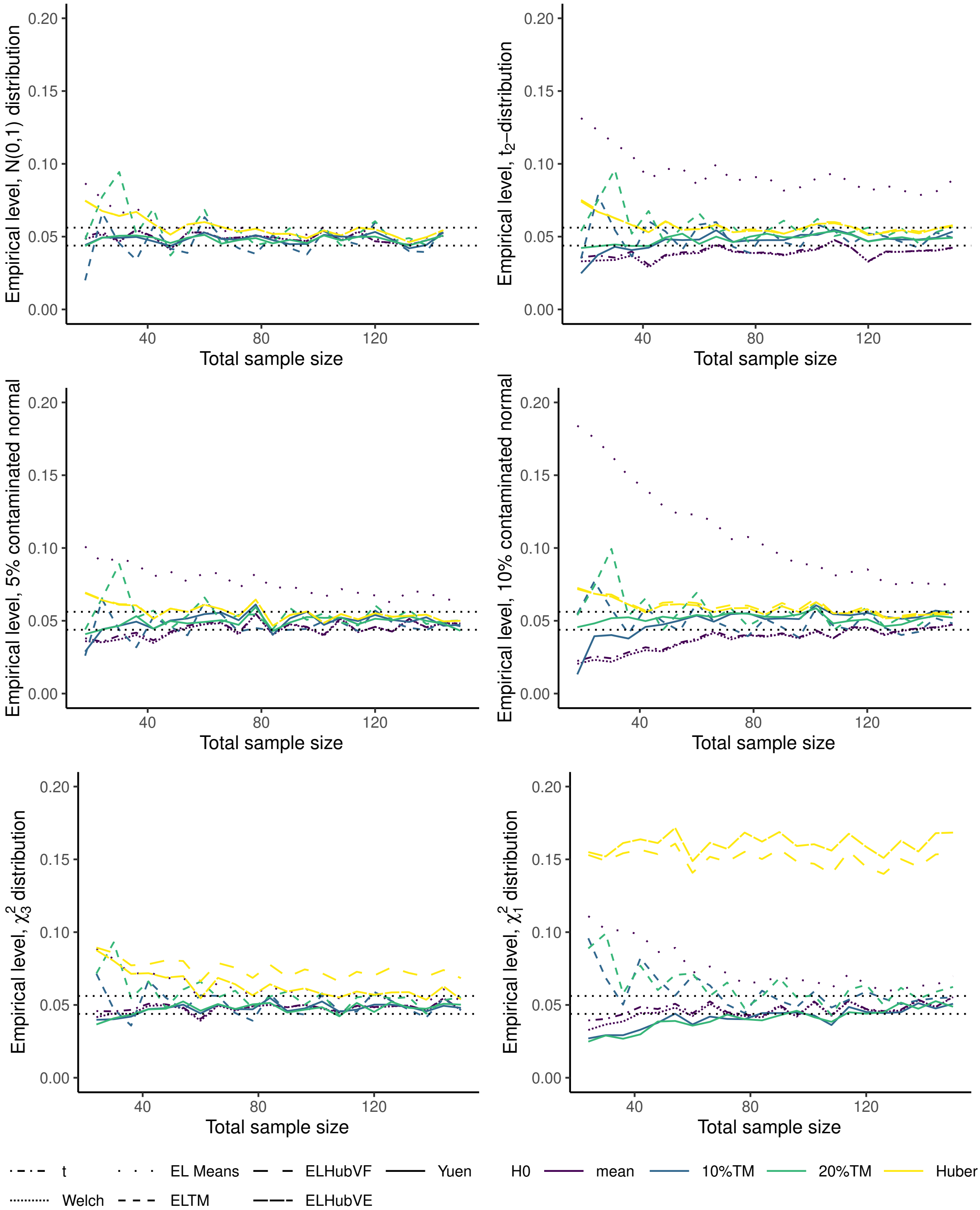


Figure 6.1: Empirical level of the tests as a function of the total sample size $n_1 + n_2$ for various distributions, balanced sample sizes. Top: $N(0, 1)$ distribution (left), t_2-distribution (right). Middle: 5% contaminated normal distribution $0.95N(0, 1) + 0.05N(0, 25)$ (left), 10% contaminated 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.

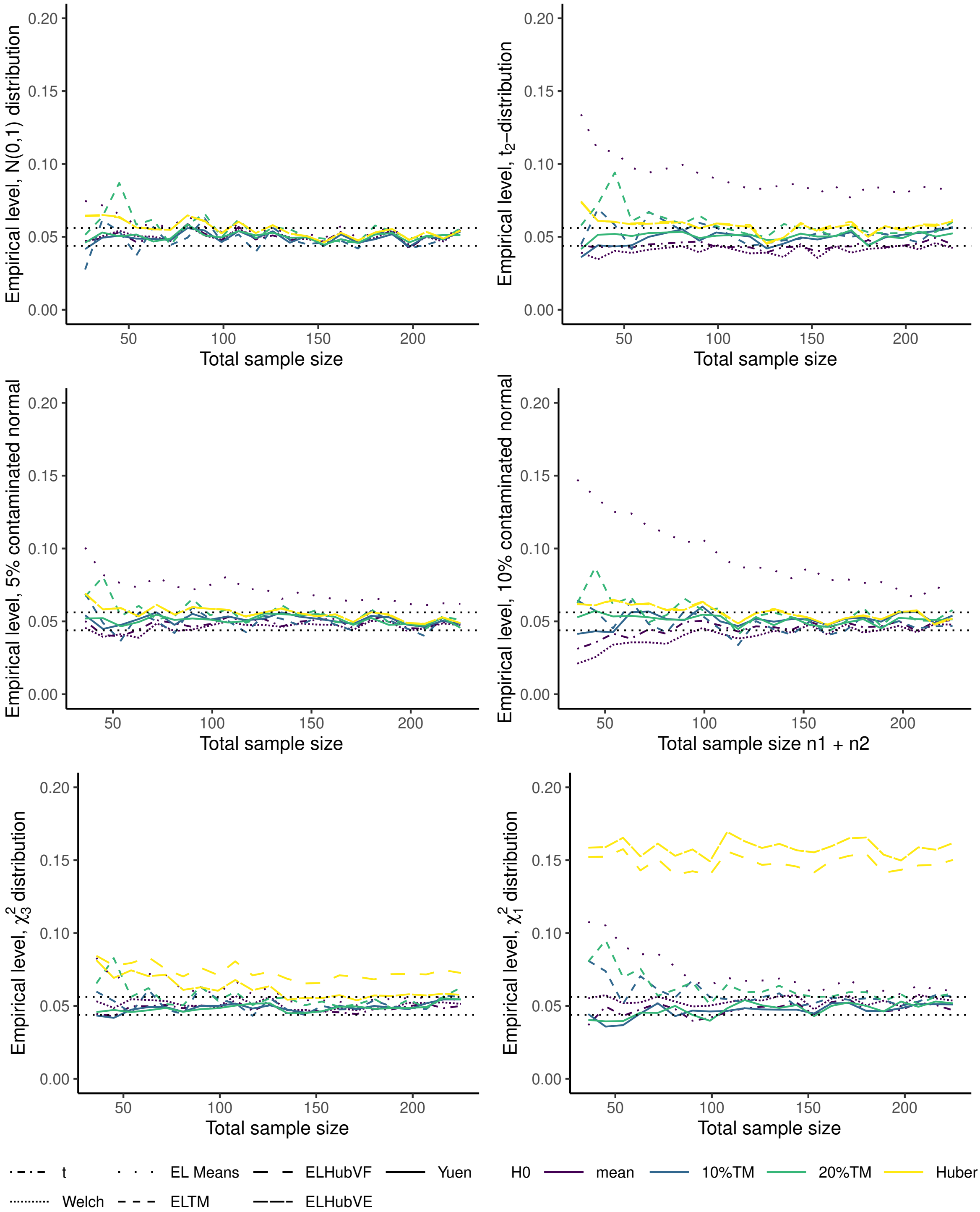


Figure 6.2: Empirical level of the tests as a function of the total sample size $n_1 + n_2$ for various distributions, unbalanced sample sizes $n_2 = 2n_1$. Top: $N(0,1)$ distribution (left), t_2-distribution (right). Middle: 5% contaminated normal distribution $0.95N(0,1) + 0.05N(0,25)$ (left), 10% 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.

means converges to the nominal 0.05 level from above, while the levels of the rest of the tests converge from below. The 10% trimmed means seem to converge slightly faster than their 20% trimmed counterpart. Yuen's test is undersized for $n_1 + n_2 < 80$.

The empirical level simulation results for the unbalanced sample size $n_2 = 2n_1$ were very similar to the balanced sample case, see Figure 6.2. In this scenario, some difference between t-test and Welch's test results appear. In general, we may observe that the test convergence to the empirical level is not seriously impacted by the differences in the sample sizes, but rather depends on the total sample size under homogeneous variance condition.

Power simulations

Next, consider the power simulation results in Figures 6.3 - 6.4. 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. The exceptions are the EL test for the means, that has the empirical type I error higher than the nominal level for the heavy-tailed t_2 and contaminated normal distributions, and the test for two smoothed Huber estimators that has a wrong level at χ^2_1 distribution. The corresponding power results should be interpreted carefully.

For the standard normal distribution, all the tests have similar performance, tests based on the trimmed means having a slightly lower power. Note that the t-test and Welch's test have practically the same power for this sample size for all distributions considered. Regarding the t_2-distribution, tests based on the trimmed means outperform the tests based on the means considerably, moreover, the 20% trimming performs better than the 10% trimming, where EL test based on the trimmed means performs slightly better than Yuen's test. The test for the difference of two smoothed Huber estimators has similar power to the test based on 10% trimmed means.

Regarding the contaminated normal distributions, the tests based on the means lose power substantially, especially in the more severe contamination setting. For the less severe contamination setting, the EL test for the trimmed means has slightly larger power than Yuen's test, and the 20% trimming performs somewhat better than the 10% trimming. For the more severe contamination setting, the EL test for 10% trimmed means seems to perform slightly better than Yuen's test with 20% trimming, while there seem to be no differences between the tests at the 10% trimming level.

Regarding the χ^2_3 distribution, all tests based on the trimmed means have similar power, slightly larger than the power of the tests based on the means. The test for smoothed Huber estimators with V fixed has a wrong level at H_0, while the test with V estimated performs similarly to the tests based on the means.

Finally, for the very skewed χ^2_1 distribution, the tests for the means have substantially lower power than the tests based on the trimmed means. The tests for the 20% trimmed means perform slightly better than the tests for the 10% trimmed means, and the method based on the Yuen's test for the trimmed means performs slightly better than the EL tests based on the trimmed means. The tests based on Huber estimators have a wrong level at H_0, resulting in considerably higher power than other tests at small location differences.

The results for the unbalanced sample sizes scenario in Figure 6.4 are very similar. We can observe that the power becomes essentially 1 for a somewhat smaller location difference Δ_0 when compared to the scenario of equal sample sizes, most probably due to a larger total sample size in this case. Some differences can be observed between Student's t-test and Welch's test, especially, for χ^2_1 distribution. The power curves of the t-test and

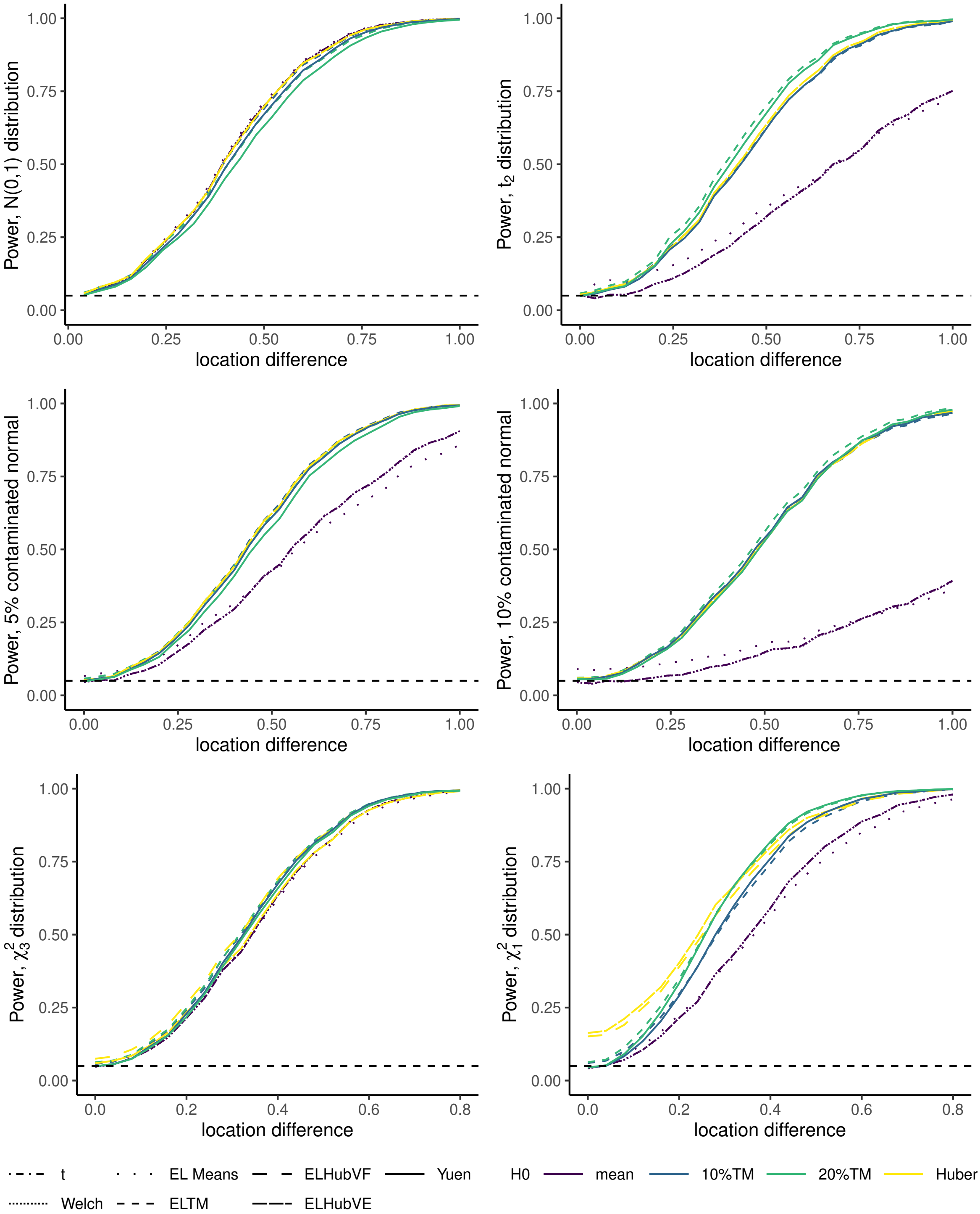


Figure 6.3: Power of the tests as a function of location difference Δ_0 for various distributions, balanced sample sizes $n_1 = n_2 = 50$. Top: $N(0, 1)$ distribution (left), t_2-distribution (right). Middle: 5% contaminated normal distribution $0.95N(0, 1) + 0.05N(0, 25)$ (left), 10% contaminated normal distribution $0.9N(0, 1) + 0.1N(0, 100)$ (right). Bottom: χ_3^2 distribution (left), χ_1^2 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.

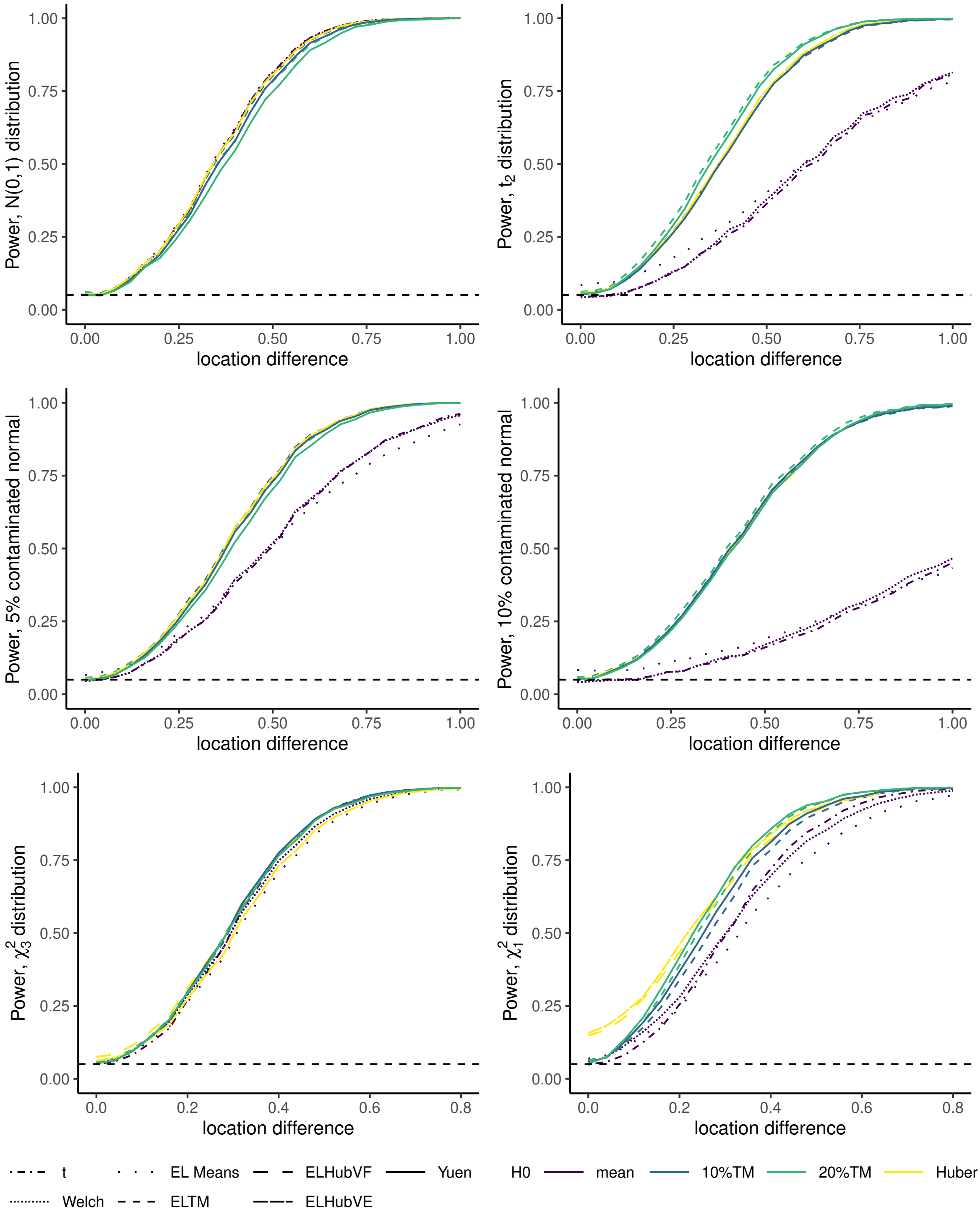


Figure 6.4: Power of the tests as a function of location difference Δ_0 for various distributions, unbalanced sample sizes $n_1 = 50$, $n_2 = 100$. Top: $N(0,1)$ distribution (left), t_2-distribution (right). Middle: 5% contaminated normal distribution $0.95N(0,1) + 0.05N(0,25)$ (left), 10% contaminated 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.

Welch's test are very close to the methods based on the trimmed means in χ_1^2 case, and Welch's test outperforms Student's t-test for small location differences. As in the balanced sample size case, tests based on Huber estimators have higher power for small Δ_0 than other methods due to the wrong test level at $\Delta_0 = 0$.

Results on robustness to variance heterogeneity

Simulation results regarding the robustness of the tests under the normal and χ_3^2 distribution with various degrees of heterogeneity and unbalanced sample sizes is presented here. 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 variances for the comparison. Each data set was generated $N = 5,000$ times and the empirical type I error at 0.05 significance level was recorded.

To ensure that the null hypothesis remains true for all settings of variance heterogeneity being considered, χ_3^2 variates were standardized to have the theoretical location parameter 0 and standard deviation 1 prior to scaling to the desired variance ratio. Namely, the real population value of the mean (for the tests involving means), of the trimmed mean (for the tests involving trimmed means) or of Huber estimator (for the test for two smoothed Huber estimators) under χ_3^2 distribution was subtracted from the data, and then it was multiplied by $\sqrt{1/6}$.

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). The average type I error for each group is reported in Tables 6.8 and 6.9, as well as visually in Figures 6.5 and 6.6. We were able to do the grouping as the pattern of results was the same in each of the groups, the deviations from the level (where such occurred) being more extreme for the larger degree of variance heterogeneity.

Finally, we evaluate the test performance by the 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 $0.5\alpha \leq \hat{\alpha} \leq 1.5\alpha$. The interval is depicted by the dotted lines in the Figures 6.5 and 6.6. We count the number of designs where each test passes the robustness condition, i.e., where it yields an empirical type I error in the interval $[0.025, 0.075]$ (Tables 6.8 and 6.9).

First, consider the results regarding the standard normal distribution in Table 6.8 and Figure 6.5. For large sample designs, all tests, except Student's t-test, are robust by Bradley's liberal criterion. For small sample designs, the EL test for the means fails to be robust for one design (with negative variance pairing), while the classical t-test and EL test for 20% trimmed means fail to be robust for most of the settings. The rest of the tests are robust in all the situations by Bradley's criterion. Analysing the results in more detail, Student's t-test is undersized for positive variance pairings and oversized for negative variance pairings, regardless of the sample size. Welch's test overcomes the negative effects of variance heterogeneity and shows results close to the nominal level in all the settings. The EL test for the means is oversized for small samples, but close to the nominal level for large samples. The results of EL test for the difference of two smoothed

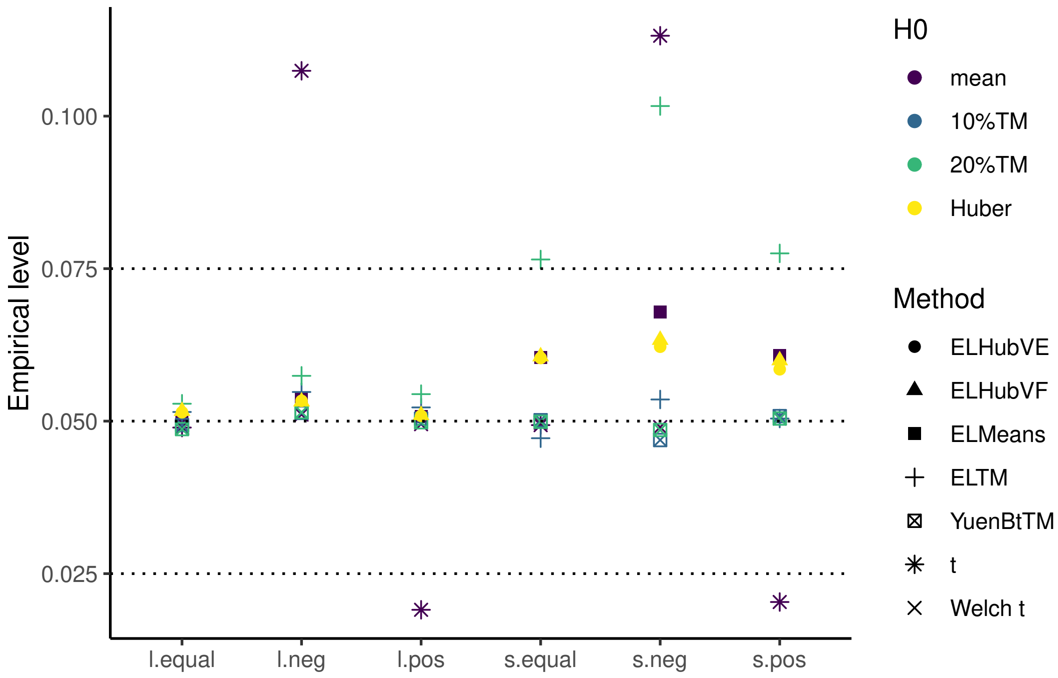


Figure 6.5: Empirical level of tests under unbalanced and heterogeneous designs for standard normal distribution. The results are grouped by the design types. Labels l and s indicate large and small sample designs, respectively, label *pos* indicates positive variance and sample size pairings, label *neg* indicates negative variance and sample size pairings, and label *equal* refers to equal variances. The dotted horizontal lines indicate the Bradley's criterion of robustness at 0.05 level, the interval $[0.025, 0.075]$. 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). 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.

Table 6.8: Empirical level of tests under unbalanced and heterogeneous designs for the $N(0,1)$ distribution. 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).

	Empirical level of tests						No. robust conditions	
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

Huber estimators, regardless of estimating V or not, are similar to that of the EL test for the means. Regarding the EL test for the difference of trimmed means, the test with 20% trimming is very oversized for small samples, while being close tot he level for large samples. The test with 10% trimming is close to the nominal level for both small and large samples. Yuen's test results are close to the nominal level in all settings.

Consider the results for χ^2_3 distribution in Table 6.9 and Figure 6.6. As expected, Student's t-test is again undersized for positive variance pairings and oversized for the negative pairings. Welch's test fails to control the probability of a type I error for small-sample heterogeneous settings, but performs quite well for the large sample settings. The EL test for the means fails to be robust for most of the heterogeneous small sample designs, and is also slightly oversized for the large sample designs. The EL test for the difference of 10% trimmed means is robust for all the settings, while the EL test for the difference of 20% trimmed means is seriously oversized for the small sample settings.

Considering the EL test for the difference of two smoothed Huber estimators, the version with V fails to be robust for small sample sizes, and performs worse than the EL method for the means for large sample sizes. The version with V estimated has the empirical level close to the nominal for large sample sizes, but is still oversized for small sample sizes. Finally, Yuen's test controls the rate of a type I error well both for small and large sample heterogeneous settings. For small samples, Yuen's test with the 20% trimming seems to yield better results than with the 10% trimming, while for large samples, 10% trimming performs slightly better.

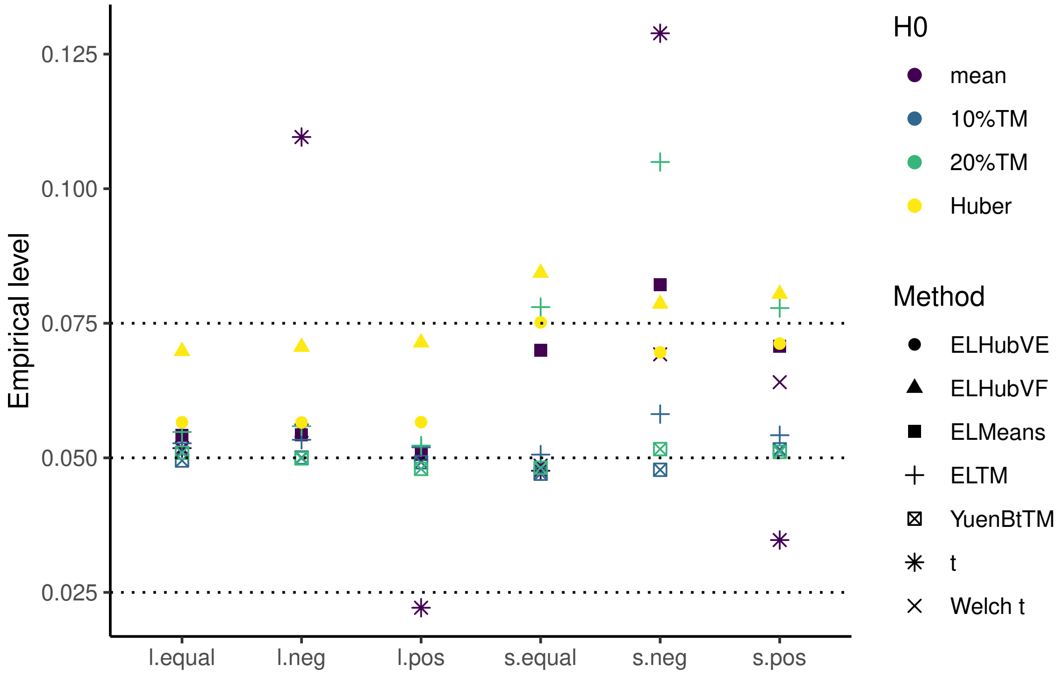


Figure 6.6: Empirical level of tests under unbalanced and heterogeneous designs for χ^2_3 distribution. The results are grouped by the design types. Labels *l* and *s* indicate large and small sample designs, respectively, label *pos* indicates positive variance and sample size pairings, label *neg* indicates negative variance and sample size pairings, and label *equal* refers to equal variances. The dotted horizontal lines indicate the Bradley's criterion of robustness at 0.05 level, the interval $[0.025, 0.075]$. 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). 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.

Table 6.9: Empirical level of tests and robustness by Bradley's criterion under unbalanced and heterogeneous designs for the χ^2_3 distribution. 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).

	Empirical level of tests						No. robust conditions	
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

6.2 Simulation study for EL-based ANOVA method for the trimmed means

The performance of the empirical likelihood-based ANOVA method for the trimmed means described in Chapter 5 has been analysed in M. Delesa-Vēliņa et al. [51, 52] in detail, the results of the studies are described below.

The study in [51] explored the properties of EL ANOVA method for the trimmed means with 5%, 10% and 20% symmetric trimming. The empirical level (empirical probability of a type I error) of the method under various skewed distributions was considered. For comparison, tests for the difference of means were included in the study, specifically the classical ANOVA F-test (panel *F-test*), Welch heteroscedastic ANOVA F-test [55] (panel *Welch*), and EL ANOVA test for the means [30] (panel *EL*). Finally, Yuen's test, which is a Welch-type procedure based on the trimmed means and Winsorized variances (panel *Yuen*, see [57, Table 7.1] for details), was performed. It was shown in [23] that Yuen's test provides a better control of the probability of the type I error for one-way ANOVA situations involving unbalanced designs and skewed distributions. For Welch's test and Yuen's test, the *R* function *t1way* from *R* package *WRS2* [24] was used. For the EL related methods, the author's custom-made *R* functions were used.

The study in [51] involved a comparison of three groups, $k = 3$, of equal size, $n \in \{20, 30, 40, 50, 100, 200, 500\}$. The potential effect of the shape of the distributions on the empirical probability of type I error was explored. Several skewed distributions with and without variance heterogeneity were considered.

For the scenario with homogeneous variances, we considered χ^2_3 distribution, the log-normal distribution with normal mean $\mu = 0$ and normal scale $\sigma = 1$, gamma distribution with shape parameter $a = 2$ and scale parameter $\sigma = 1$, and the skew-normal distribution with location parameter $\xi = 0$, scale parameter $\omega = 1$, and slant parameter $\alpha = 1$ (see [2] for details on skew-normal distribution).

For the scenario with heterogeneous variances, we further transform the data simulated from the three independent skewed distributions as to have the ratios between variances to be either 1:4:9 or 1:1:36. To ensure that the relevant H_0^T (5.6) of equal trimmed means or H_0 (5.1) of equal means is true, before altering the variances, we centre the data using the theoretically determined trimmed means (when using tests for the comparison of trimmed means) or means (when using tests for the comparison of means). A similar approach was used in [20], where (trimmed) means of two independent skewed populations were compared. We use 10,000 Monte Carlo simulations to calculate the empirical level of the tests performed at the nominal 0.05 significance level.

The results of the study are presented in Tables 6.10 - 6.12. Table 6.10 presents the empirical level of the tests for the homogeneous variances scenario. We can see that for all distributions the empirical rejection rates of EL ANOVA test for the trimmed means converge to the nominal level 0.05. For small samples, the rejection rates are closer to the nominal level when the trimming proportion is lower. Note that EL ANOVA test for 5% and 10% trimmed means has rejection rates similar to or closer to nominal than Welch's test in all settings. However, the rates of EL ANOVA test for the trimmed means are further from the nominal than those of Yuen's test for all trimming proportions. Among the tests for the means, the results of the EL ANOVA test for means are the closest to the nominal significance level.

Regarding the heterogeneous scenarios in Tables 6.11 and 6.12, we can observe that the empirical level of the EL ANOVA test for the trimmed means converges to the nominal level when n is large. The rejection rates of EL ANOVA test for 5% and 10% trimmed means are always closer to nominal than those of F-test. They are similar and, in many cases, closer to the nominal rate than those of Welch's test for χ_3^2, Gamma and lognormal distributions. For small samples, the EL ANOVA test is considerably oversized, especially with 20% trimmed means. For the skew-normal distribution (which is the least skewed from all the distributions considered), EL ANOVA for the means performs the best from all the methods considered.

The results for the heterogeneous variances scenario suggest that Yuen's test performs best among the tests for the comparison of the trimmed means, while the EL ANOVA test performs best among the tests for the comparison of means. Among EL methods, the EL method for the means performs better than the EL method for the trimmed means for small sample sizes, except for the lognormal distribution, where the EL ANOVA method with 10% trimmed means performs better.

The study in [8] provided an additional insight in the performance of the EL ANOVA method for the trimmed means in case of comparing two groups. The design of the study is described in Chapter 6.1.2 and for the results see [8]. Considering the empirical level simulations in [8, Figure 1 – Figure 2], the EL ANOVA test for the trimmed means converges to the nominal level under H_0 of equal trimmed means for all the distributions considered. The convergence pattern was similar, although not identical, to that of the EL method for the difference of two trimmed means. For small samples, the method based on 10% trimmed means has the empirical rates closer to the nominal than the method based on 20% trimmed means, which is consistent with the results for three groups in [51]. For all nonnormal cases the EL ANOVA method based on the trimmed means is closer to the nominal level than the EL method for the means, except for the moderately skewed χ_3^2 distribution under the balanced sample design (see [8, Figure 1 bottom left]), where the two methods are very close for large sample sizes.

Table 6.10: Empirical level of various tests for the equality of means and trimmed means of three independent skewed distributions with homogeneous variances. For methods involving trimmed means, symmetric trimming proportions $\alpha = \beta = c$ are used.

χ^2_3									
				Trimming level					
				$c = 5\%$		$c = 10\%$		$c = 20\%$	
n	F-test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.047	0.076	0.052	0.050	0.079	0.050	0.080	0.049	0.090
30	0.048	0.070	0.054	0.054	0.055	0.053	0.075	0.054	0.079
40	0.044	0.062	0.052	0.049	0.063	0.048	0.063	0.050	0.069
50	0.048	0.058	0.049	0.046	0.047	0.047	0.060	0.049	0.067
100	0.049	0.055	0.051	0.050	0.056	0.050	0.056	0.049	0.056
200	0.051	0.055	0.053	0.050	0.053	0.051	0.053	0.051	0.056
500	0.051	0.049	0.049	0.048	0.050	0.050	0.051	0.051	0.052
Lognormal ($\mu = 0, \sigma = 1$)									
				Trimming level					
				$c = 5\%$		$c = 10\%$		$c = 20\%$	
n	F test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.044	0.073	0.047	0.040	0.069	0.040	0.070	0.040	0.081
30	0.045	0.069	0.050	0.048	0.049	0.046	0.068	0.046	0.072
40	0.044	0.063	0.049	0.046	0.062	0.046	0.062	0.045	0.066
50	0.045	0.065	0.054	0.049	0.048	0.047	0.059	0.046	0.062
100	0.049	0.059	0.055	0.049	0.055	0.049	0.057	0.049	0.057
200	0.050	0.053	0.050	0.045	0.048	0.046	0.049	0.048	0.052
500	0.051	0.053	0.052	0.051	0.052	0.050	0.050	0.052	0.054
Gamma ($\alpha = 2, \sigma = 1$)									
				Trimming level					
				$c = 5\%$		$c = 10\%$		$c = 20\%$	
n	F test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.052	0.078	0.050	0.050	0.077	0.052	0.079	0.052	0.096
30	0.049	0.069	0.053	0.052	0.053	0.050	0.070	0.052	0.080
40	0.050	0.062	0.052	0.052	0.064	0.052	0.067	0.053	0.074
50	0.050	0.060	0.051	0.048	0.048	0.050	0.062	0.052	0.069
100	0.052	0.057	0.052	0.051	0.057	0.050	0.056	0.048	0.056
200	0.052	0.056	0.053	0.052	0.055	0.053	0.055	0.052	0.055
500	0.049	0.052	0.051	0.050	0.051	0.049	0.051	0.051	0.052
Skew-normal ($\xi = 0, \omega = 1, \alpha = 1$)									
				Trimming level					
				$c = 5\%$		$c = 10\%$		$c = 20\%$	
n	F-test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.055	0.077	0.049	0.049	0.077	0.051	0.083	0.054	0.099
30	0.048	0.065	0.050	0.051	0.051	0.050	0.068	0.051	0.078
40	0.049	0.061	0.049	0.049	0.062	0.051	0.065	0.051	0.073
50	0.051	0.058	0.049	0.049	0.050	0.050	0.061	0.052	0.071
100	0.055	0.055	0.052	0.051	0.056	0.050	0.056	0.052	0.060
200	0.052	0.051	0.049	0.050	0.052	0.051	0.054	0.052	0.056
500	0.046	0.048	0.047	0.047	0.048	0.048	0.049	0.048	0.049

Regarding the power analysis in [8, Figure 3 – Figure 4], the EL ANOVA test for the trimmed means has power comparable to that of the classical methods based on the means when the distribution is normal, and higher power for all the other distributions, except the moderately skewed χ^2_3 distribution where the powers are similar. Regarding the robustness of the test level to the variance heterogeneity and unbalanced sample sizes, see [8, Table 1 – 2], EL ANOVA method is robust for the large sample cases both in skewed

Table 6.11: Empirical level of various tests for the equality of means and trimmed means of three independent skewed distributions with the ratios between variances being 1:4:9. For methods involving trimmed means, symmetric trimming proportions $\alpha = \beta = c$ are used.

						Trimming level			
				$c = 5\%$		$c = 10\%$		$c = 20\%$	
χ^2_3									
n	F-test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.086	0.101	0.071	0.064	0.096	0.061	0.094	0.062	0.109
30	0.083	0.088	0.071	0.060	0.062	0.064	0.084	0.064	0.091
40	0.080	0.075	0.062	0.060	0.072	0.055	0.072	0.054	0.078
50	0.079	0.067	0.057	0.050	0.051	0.051	0.066	0.052	0.071
100	0.079	0.063	0.058	0.055	0.060	0.055	0.060	0.054	0.062
200	0.085	0.060	0.057	0.057	0.059	0.054	0.058	0.054	0.058
500	0.073	0.052	0.051	0.053	0.054	0.052	0.054	0.051	0.053
Lognormal ($\mu = 0, \sigma = 1$)									
						Trimming level			
				$c = 5\%$		$c = 10\%$		$c = 20\%$	
n	F test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.110	0.146	0.113	0.078	0.106	0.070	0.106	0.066	0.114
30	0.109	0.131	0.111	0.063	0.064	0.065	0.088	0.063	0.090
40	0.100	0.115	0.100	0.066	0.081	0.062	0.083	0.059	0.084
50	0.098	0.110	0.099	0.060	0.060	0.060	0.074	0.059	0.077
100	0.097	0.090	0.083	0.058	0.063	0.055	0.062	0.055	0.063
200	0.082	0.071	0.069	0.051	0.055	0.052	0.055	0.054	0.058
500	0.077	0.061	0.060	0.051	0.053	0.052	0.053	0.051	0.054
Gamma ($\alpha = 2, \sigma = 1$)									
						Trimming level			
				$c = 5\%$		$c = 10\%$		$c = 20\%$	
n	F test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.086	0.099	0.070	0.061	0.093	0.062	0.097	0.065	0.111
30	0.083	0.080	0.063	0.058	0.062	0.056	0.079	0.060	0.090
40	0.079	0.074	0.061	0.058	0.073	0.056	0.073	0.059	0.083
50	0.079	0.068	0.057	0.049	0.050	0.053	0.066	0.058	0.075
100	0.078	0.059	0.055	0.053	0.060	0.053	0.060	0.053	0.061
200	0.078	0.057	0.054	0.053	0.055	0.053	0.057	0.053	0.057
500	0.079	0.052	0.050	0.049	0.051	0.050	0.050	0.052	0.053
Skew-normal ($\xi = 0, \omega = 1, \alpha = 1$)									
						Trimming level			
				$c = 5\%$		$c = 10\%$		$c = 20\%$	
n	F-test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.080	0.079	0.048	0.048	0.081	0.052	0.085	0.054	0.106
30	0.075	0.069	0.050	0.051	0.054	0.053	0.073	0.057	0.085
40	0.080	0.060	0.049	0.050	0.064	0.050	0.069	0.053	0.075
50	0.076	0.060	0.050	0.050	0.051	0.051	0.064	0.053	0.073
100	0.079	0.053	0.049	0.048	0.054	0.048	0.055	0.051	0.059
200	0.078	0.054	0.051	0.051	0.054	0.051	0.054	0.052	0.056
500	0.075	0.049	0.047	0.048	0.049	0.048	0.049	0.048	0.049

χ^2_3 and normal distribution setting. However, for the small sample cases, only the EL ANOVA method with 10% trimmed means appears to be robust.

Table 6.12: Empirical level of various tests for the equality of means and trimmed means of three independent skewed distributions with the ratios between variances being 1:1:36. For methods involving trimmed means, symmetric trimming proportions $\alpha = \beta = c$ are used.

$\boldsymbol{\chi^2_3}$

				Trimming level					
				$c = 5\%$		$c = 10\%$		$c = 20\%$	
n	F-test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.124	0.090	0.067	0.059	0.087	0.056	0.088	0.056	0.102
30	0.119	0.080	0.064	0.056	0.058	0.058	0.078	0.058	0.086
40	0.116	0.070	0.058	0.055	0.067	0.053	0.067	0.052	0.072
50	0.112	0.063	0.052	0.048	0.049	0.051	0.063	0.052	0.069
100	0.111	0.061	0.055	0.053	0.059	0.051	0.058	0.050	0.058
200	0.113	0.059	0.056	0.054	0.057	0.053	0.056	0.052	0.056
500	0.102	0.053	0.053	0.053	0.054	0.050	0.052	0.050	0.052

Lognormal ($\boldsymbol{\mu = 0, \sigma = 1}$)

				Trimming level					
				$c = 5\%$		$c = 10\%$		$c = 20\%$	
n	F test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.168	0.126	0.101	0.071	0.095	0.062	0.095	0.058	0.103
30	0.166	0.118	0.098	0.055	0.056	0.060	0.081	0.057	0.084
40	0.153	0.104	0.089	0.060	0.076	0.061	0.077	0.057	0.080
50	0.148	0.095	0.086	0.052	0.053	0.056	0.068	0.055	0.072
100	0.136	0.080	0.075	0.054	0.061	0.053	0.062	0.054	0.061
200	0.119	0.064	0.062	0.050	0.054	0.049	0.052	0.051	0.055
500	0.112	0.056	0.055	0.052	0.053	0.050	0.051	0.052	0.054

Gamma ($\boldsymbol{\alpha = 2, \sigma = 1}$)

				Trimming level					
				$c = 5\%$		$c = 10\%$		$c = 20\%$	
n	F test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.123	0.089	0.064	0.058	0.089	0.059	0.093	0.060	0.107
30	0.122	0.079	0.061	0.057	0.059	0.055	0.077	0.058	0.086
40	0.116	0.071	0.057	0.056	0.069	0.054	0.070	0.055	0.077
50	0.113	0.066	0.055	0.051	0.052	0.051	0.064	0.053	0.070
100	0.110	0.059	0.053	0.053	0.058	0.052	0.057	0.051	0.060
200	0.109	0.054	0.051	0.051	0.054	0.050	0.054	0.051	0.055
500	0.108	0.052	0.051	0.049	0.050	0.050	0.051	0.050	0.051

Skew-normal ($\boldsymbol{\xi = 0, \omega = 1, \alpha = 1}$)

				Trimming level					
				$c = 5\%$		$c = 10\%$		$c = 20\%$	
n	F-test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.113	0.077	0.050	0.047	0.080	0.049	0.083	0.054	0.103
30	0.107	0.067	0.051	0.050	0.052	0.052	0.074	0.054	0.083
40	0.112	0.060	0.049	0.049	0.063	0.051	0.067	0.054	0.074
50	0.107	0.058	0.048	0.049	0.050	0.051	0.064	0.055	0.074
100	0.107	0.054	0.048	0.048	0.054	0.049	0.056	0.052	0.060
200	0.106	0.053	0.051	0.052	0.054	0.051	0.054	0.053	0.057
500	0.106	0.048	0.048	0.050	0.051	0.050	0.051	0.049	0.051

6.3 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. Chapter 6.3.1 deals with the two-sample case, while Chapter 6.3.2 treats the ANOVA case.

6.3.1 Comparing two populations

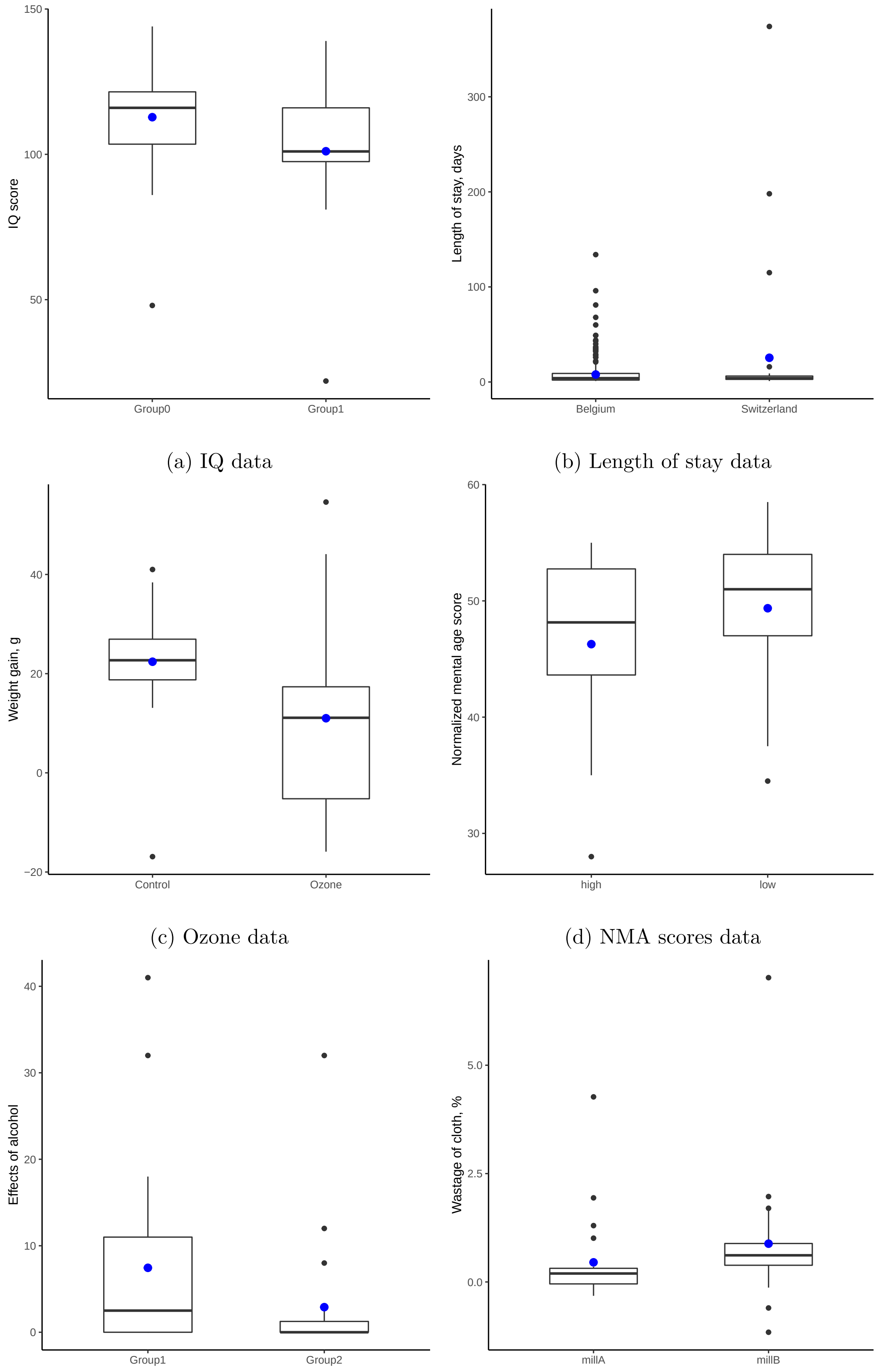
We consider the EL test for the difference of two trimmed means with 10%, 15% and 20% trimming, panel *EL TM*, and the EL test for the difference of two smoothed Huber estimators with $V = 2.046$ in (2.16) fixed as suggested in [13], panel *ELHubVF*, and V estimated for each sample separately by the nonparametric bootstrap method with 10,000 resamples (see, for example, [53] for the reference), panel *ELHubVE*. For computation of the smoothed Huber estimator, MAD was used as a preliminary robust estimator of the scale parameter σ of the underlying distribution as required by Lemma 3.2.2. In [52], we noted that the choice of the scale estimator is important: choosing SD as the scale estimator yielded p-values closer to the methods based on the means.

For the comparison, we report the results of Student's t-test, Welch's test [54], Wilcoxon rank sum test (see, for example, [32]), EL test for the means from Example 1.4.1 (panel *EL Means*), and Yuen's test for the trimmed means [60] with 10%, 15% and 20% symmetric trimming (panel *Yuen*). We give a short description of each of the datasets along with the descriptive statistics in Table 6.13 and visual representation in Figure 6.7. The p-values of the tests are given in Table 6.14, the confidence intervals for the difference of the location parameters in Table 6.15 and the respective interval lengths in Table 6.16.

Table 6.13: Descriptive statistics for the data sets considered in Chapter 6.3. n denotes sample size, κ denotes the estimate of the coefficient of skewness, *SHub* denotes smoothed Huber estimate with $k = 0.862$ and *TM* denotes sample trimmed mean with the given trimming rate.

data set	group	n	$\bar{x}$	SD	Med	MAD	κ	SHub	TM10	TM15	TM20
IQ	Group0	79	112.8	14.3	116.0	11.9	-1.2	116.0	113.6	113.8	114.1
	Group1	15	101.1	27.0	101.0	14.8	-1.5	101.3	104.2	104.1	104.2
LOS	Belgium	315	7.9	13.4	4.0	3.0	5.1	3.9	5.0	4.7	4.4
	Switzerland	32	25.5	74.6	4.0	3.0	3.8	4.1	4.8	4.4	4.2
ozone	Control	23	22.4	10.8	22.7	6.5	-1.9	23.0	23.1	23.2	23.3
	Ozone	22	11.0	19.0	11.1	11.9	0.5	11.1	9.7	9.2	9.2
NMA	high	18	46.3	7.7	48.1	7.0	-0.9	48.0	46.9	47.2	47.5
	low	21	49.4	6.9	51.0	5.9	-0.6	50.6	49.9	50.1	50.4
alcohol	Group1	20	7.4	11.3	2.5	3.7	1.9	2.5	4.8	4.1	3.9
	Group2	20	2.9	7.5	0.0	0.0	3.2	0.4	0.9	0.4	0.2
cotton	millA	22	0.5	1.0	0.2	0.2	2.8	0.3	0.2	0.2	0.2
	millB	22	0.9	1.5	0.6	0.4	3.0	0.7	0.7	0.7	0.6

IQ data set This data set was analysed in [15] and the author's publication [52]. The IQ scores of 94 children aged 5 years are provided. Fifteen children have mothers suffering from postnatal depression (group 1), whereas 79 children have healthy mothers (group 0). The null hypothesis of interest is that there is no difference between the locations of the IQ distributions of the two groups of children. Most of the IQ values are between 80 and 144, except two small values, 22 and 48, corresponding to one child in each group, respectively. The standard deviation (SD) in group 1 is almost twice of that of group 0,



(a) IQ data

(b) Length of stay data

(c) Ozone data

(d) NMA scores data

(e) Alcohol data

(f) Cotton data

Figure 6.7: Grouped boxplots for the two-sample data sets analyzed in the chapter. The blue dots represent the sample mean of the group.

Table 6.14: p-values of the two-sample tests for the difference of two location parameters for the data sets considered in Chapter 6.3.

method	IQ	LOS	LOS*	ozone	NMA	alcohol	cotton
Student's t	0.016	<0.001	0.933	0.017	0.194	0.142	0.277
Welch	0.122	0.192	0.953	0.019	0.198	0.143	0.278
Wilcoxon	0.058	0.632	0.912	0.003	0.190	0.047	0.010
EL Means	0.052	0.023	0.951	0.025	0.170	0.131	0.243
ELHubVF	0.068	0.792	0.252	0.009	0.183	0.192	0.029
ELHubVE	0.052	0.449	0.210	0.024	0.170	0.130	0.003
ELTM 10%	0.093	0.816	0.141	0.011	0.239	0.034	0.046
Yuen 10%	0.082	0.808	0.135	0.008	0.229	0.073	0.034
ELTM 15%	0.080	0.688	0.266	0.001	0.295	0.002	0.031
Yuen 15%	0.087	0.669	0.250	0.004	0.290	0.038	0.019
ELTM 20%	0.082	0.782	0.391	<0.001	0.263	0.003	<0.001
Yuen 20%	0.065	0.779	0.405	0.004	0.287	0.076	0.001

Table 6.15: Two-tailed 95% confidence intervals for the difference of two location estimators stemming from the two-sample tests considered in Chapter 6.3.

method	IQ	LOS	LOS*	ozone	NMA	alcohol	cotton
Student's t	(2.3, 21.2)	(-27, -8.2)	(-5.5, 5.1)	(2.2, 20.6)	(-1.6, 7.8)	(-1.6, 10.7)	(-1.2, 0.4)
Welch	(-3.5, 26.9)	(-44.5, 9.3)	(-8, 7.5)	(2, 20.8)	(-1.7, 7.9)	(-1.6, 10.7)	(-1.2, 0.4)
Wilcoxon	(0, 18)	(-1, 1)	(-1, 1)	(5.9, 20.6)	(-1.7, 7.5)	(0, 6)	(-0.7, -0.1)
EL Means	(-0.1, 29.4)	(-55.2, -1.3)	(-12, 4.2)	(1.6, 19.8)	(-1.3, 7.8)	(-1.5, 11.2)	(-1.4, 0.3)
ELHubVF	(-0.8, 24.4)	(-1.9, 1.6)	(-0.8, 2.0)	(3.5, 20.9)	(-1.4, 7.8)	(-1.8, 7.7)	(-0.8, -0.1)
ELHubVE	(-0.1, 29.4)	(-13.0, 2.6)	(-1.9, 4.7)	(1.6, 19.8)	(-1.3, 7.8)	(-1.5, 11.2)	(-0.8, -0.2)
ELTM 10%	(-1.6, 19.8)	(-2.8, 1.9)	(-0.3, 2)	(3.5, 22.6)	(-2, 8.2)	(0.3, 8.4)	(-0.9, 0)
Yuen 10%	(-1.3, 20)	(-1.8, 2.3)	(-0.3, 2.1)	(3.9, 23.1)	(-2, 8)	(-0.4, 8.2)	(-0.8, 0)
ELTM 15%	(-1.1, 20.5)	(-1.3, 1.6)	(-0.5, 1.8)	(5.8, 23.1)	(-2.6, 8.6)	(1.1, 6.9)	(-0.9, 0)
Yuen 15%	(-1.7, 21.2)	(-1.1, 1.6)	(-0.5, 1.8)	(5.1, 22.9)	(-2.6, 8.3)	(0.2, 7.2)	(-0.8, -0.1)
ELTM 20%	(-1.2, 16.6)	(-1.1, 1.4)	(-0.6, 1.6)	(7.2, 24)	(-1.9, 9.2)	(0.9, 7.5)	(-0.7, -0.2)
Yuen 20%	(-0.8, 20.6)	(-1.1, 1.5)	(-0.7, 1.7)	(5.3, 22.9)	(-2.6, 8.3)	(-0.5, 7.8)	(-0.7, -0.2)

Table 6.16: The lengths of the confidence intervals for the difference of two location estimators stemming from the two-sample tests considered in Chapter 6.3.

method	IQ	LOS	LOS*	ozone	NMA	alcohol	cotton
Student's t	18.89	18.78	10.59	18.48	9.46	12.29	1.58
Welch	30.42	53.84	15.50	18.86	9.56	12.35	1.59
Wilcoxon	18.00	2.00	2.00	14.70	9.20	6.00	0.58
EL Means	29.46	53.87	16.18	18.15	9.13	12.74	1.73
ELHubVF	25.14	3.41	2.72	17.40	9.28	9.49	0.77
ELHubVE	29.47	15.51	6.60	18.14	9.13	12.73	0.59
ELTM 10%	21.38	4.64	2.30	19.10	10.20	8.13	0.84
Yuen 10%	21.36	4.17	2.37	19.21	9.94	8.55	0.80
ELTM 15%	21.63	2.82	2.28	17.31	11.19	5.82	0.87
Yuen 15%	22.81	2.71	2.28	17.77	10.90	6.97	0.76
ELTM 20%	17.84	2.53	2.21	16.82	11.06	6.56	0.45
Yuen 20%	21.37	2.57	2.34	17.57	10.95	8.24	0.48

27.0 and 14.3, respectively. The data sets are of similar skewness. The presence of outliers and heterogeneous variances suggests using hypotheses tests for robust estimators.

Student's t-test rejects the H_0 with $p = 0.016$, while Welch's test does not, having $p = 0.12$. The EL test for the means, Wilcoxon test and the EL test for the smoothed Huber estimators with V estimated have p-values just above 0.05. Tests based on trimmed means do not reject the H_0 at 0.05 level, the results given by the EL method and Yuen's test are similar. The EL method gives slightly shorter confidence intervals for 15% and 20% trimming.

Length of stay (LOS) data set This data set was analysed previously in [25] and author's publication [52]. The first sample contains 315 lengths of stay (LOS) in days for patients hospitalized in Belgium during 1988 for certain disorders of the nervous system. The second sample contains 32 LOS of patients hospitalized during the same year in Switzerland for the same types of illnesses. The Switzerland sample contains two extreme values: 374 and 198 days, respectively. A derived data set, called LOS*, was obtained by removing the two extreme Switzerland sample LOS values. We consider the null hypothesis H_0 that there is no difference between the location of the LOS distributions in Belgium and Switzerland. Both data sets are very skewed, the skewness coefficient being 5.1 and 3.8 for Belgium and Switzerland, respectively, and the SD is almost six times larger in the smallest group, Switzerland. This suggests that tests for the trimmed means are preferable. The use of the EL Huber test with V fixed should not be advised, as our simulation study showed that for skewed data sets it is oversized.

Student's t-test and the EL test for the means reject H_0 at 0.05 level, while none of the tests based on robust estimators reject H_0 and have much larger p-values. Neither does Welch's test reject H_0, however, its confidence interval is much larger than those associated with the tests based on robust estimators. Yuen's test and the EL test for the trimmed means yield similar p-values for the same trimming proportion, also the confidence intervals are of similar length. We can see that the p-values for the two versions of Huber test are quite different, version with V estimated yielding a smaller value and a larger confidence interval.

After removing the two extreme observations from the Switzerland data sample (LOS* data set), none of the tests reject H_0 at 0.05 level, yielding the same conclusion as the tests based on the robust estimators with the raw data. However, removing observations manually involves subjective judgement, thus doing tests on the raw data are preferable.

Ozone data set The experimental group consisted of 22 rats kept in an ozone environment for 7 days, while a control group of 23 rats were kept in an ozone-free environment. The weight gains of the rats, in grams, were registered. We are interested in H_0 that there were no differences in weight gain between the experimental and control groups. The data set was used in [57]. The data sets are of similar size, while the standard deviations differ almost two times, SD=19.8 in the experimental group and SD=10.8 in the control group. The control group is negatively skewed (mainly due to one small outlier), while the experimental group is positively skewed (one large outlier).

The tests based on means (Student's t-test, EL means test, Welch's test) reject the H_0 at 0.05 level. *ELHubVE* gives a result similar to the EL test for the means. The rest of the tests based on robust location measures (except the EL test for 10% trimmed means) and Wilcoxon's test reject the H_0 at 0.01 level. Interestingly, also Wilcoxon sum of ranks test rejects at 0.01 level, although this test is not recommended in the cases where groups differ in skewness. EL tests for trimmed means have slightly shorter confidence intervals than Yuen's test for the trimmed means. The test for the smoothed Huber estimator with V estimated has larger confidence intervals than those of the tests for the trimmed means, but shorter than those of the tests based on the means.

Normalized mental-age (NMA) scores data set This data set was considered in [32]. The data reports the normalized mental age scores of children suffering from phenylketonuria (PKU), i.e., a genetic condition of not being able to metabolize the protein phenylalanine. It has been suggested that an elevated level of serum phenylalanine increases a child's likelihood of mental deficiency, however, early dietary treatment can be efficient in normalizing the serum phenylalanine levels. 39 children with PKU condition who had received an early treatment were included in the data set: 21 children having a low average daily serum phenylalanine levels at the age of two (group *low*), and 18 children having a high exposure (group *high*). The mental age scores normalized to 48 months are reported, the mean scores being 46.3 months for the low group, and 49.4 months for the high group. We are interested in H_0 that the NMA scores are equal for the two groups. The data sets are of similar size, have similar standard deviations and both are mildly negatively skewed. The results show that none of the tests reject H_0 at 0.05 level, however, the p-values of the tests based on trimmed means are slightly higher. The test for the smoothed Huber estimators yields p-values closer to that of the methods based on the means, $p = 0.183$ and $p = 0.170$ for the fixed and estimated variance, respectively. All the tests yield similar confidence interval lengths,the EL test for the trimmed means being comparatively the largest.

Alcohol data set This data set was analysed in [57] and reports the hangover symptoms the morning after consuming equal amounts of alcohol in a laboratory. Group 1 was a control group and group 2 was formed of sons of alcoholic fathers. Both data sets have equal size (20 participants) and both are severely positively skewed. Group 2 contains many zero values resulting in a median value of zero, consequently, MAD being zero as well. Thus it is possible to compute the smoothed Huber estimates only using the SD as the preliminary estimator of variance, in which case it is not considered robust.

None of the tests for means reject the H_0 at 0.1 level, neither the test based on the smoothed Huber estimators. In contrary, the EL test for trimmed means rejects the score equality at 0.05 level with 10% trimming, and at 0.01 level for 15% and 20% trimming. Also, Wilcoxon test rejects H_0 at 0.05 level. Interestingly, Yuen's test for the trimmed means fail to reject H_0 with 10% and 20% trimming even at 0.05 level. Moreover, the confidence intervals for the EL test for the trimmed means are narrower when compared to those of Yuen's test. Wilcox [57, p. 164] notes, that the shapes of the distribution are similar to exponential, and even with 20% trimming, Yuen's test might yield inaccurate probability coverage.

Cotton data set The example has been used in [38] and contains data on wastage due to defects in cloth in two Levi Strauss garment factories. The measure is expressed in the percentage with respect to the wastage calculated by the computerized layouts of patterns on the cloth. The measure can be negative since the workers can do better than the computer by laying out the patterns by hand. The data sets (labelled *millA* and *millB*) are of equal size (22 observations) and have similar standard deviations. Both data sets contain several outliers and are considerably positively skewed (2.8 and 3.0, respectively), suggesting that the tests based on the means will lack power.

In fact, none of the tests based on the means reject H_0 at 0.05 level, the p-values being 0.28 for Student's t-test and Welch's test, and 0.24 for the EL means test. The tests for 10% and 15% trimmed means reject H_0 at 0.05 level, while the Wilcoxon test, the test for 20% trimmed means and the test for smoothed Huber estimator with V estimated reject the H_0 at 0.01 level. When comparing Yuen's method with the EL based test for the trimmed means, the length of the confidence intervals are similar. The shortest confidence intervals are given by the tests for 20% trimmed mean, followed by the tests for the smoothed Huber estimator with V estimated.

6.3.2 EL-based ANOVA method for the trimmed means

To illustrate the use of the EL ANOVA method for the trimmed means, we consider Oslo transect data set [37]. This data set includes 360 observations corresponding to different plants collected along a 120 km transect running through the city of Oslo, Norway, and was previously analysed in [42] and the author's publication [51]. The concentrations of 25 chemical elements found in the plants were recorded together with factors that may influence the mineral concentration. Except for not including two chemical elements, Au and Na, this data is available within *R* package *rrcov* [41] as *OsloTransect* data set.

We analyse the remaining 23 concentrations of chemical elements as the response variables, and the lithology as a group variable with four levels, see Table 6.17 for the details. After removing the observations with missing values, we are left with 332 observations. The box plots of the data are given in Figures 6.8 - 6.9. We can see that for most of the element concentrations data sets are skewed and contain outliers. Moreover, the sample sizes are not balanced.

Table 6.17: Names of the lithology groups in Oslo Transect data

Key	Lithological group	No. observations
CAMSED	Cambro-Silurian sedimentary rocks	98
GNEISS_O	Precambrian gneisses – Oslo	89
GNEISS_R	Precambrian gneisses – Randsfjord	32
MAGM	Magmatic rocks of the Oslo Rift	113

We consider the EL ANOVA method for the trimmed means with 10%, 15% and 20% symmetrical trimming (panel *ELT*). For comparison, we provide the classical ANOVA F-test, Welch's heteroscedastic ANOVA F-test [55] (panel *Welch*), EL ANOVA test for the means [30] (panel *EL*), and Yuen's test [57, Table 7.1] (panel *Yuen*) for 10%, 15% and 20% symmetrically trimmed means.

We report the respective p-values of the tests in Table 6.18. We note that, for each trimming strategy, the p-values from the EL ANOVA for trimmed means and Yuen's test are very similar. In addition, the p-values from the EL ANOVA for means and Welch's heteroscedastic F-test are also very similar.

We comment on some of the chemical elements for which the tests based on the means and the tests based on the trimmed means give different conclusions. For Cr, La and Pb, the F-test gives large p-values not rejecting H_0 at 0.05 significance level. Welch's test and EL test for the mean, as well as the tests based on trimmed means all reject at 0.01 level (except for La, where tests based on 10% trimmed means have $p > 0.1$). Cr, La and Pb concentrations are quite skewed in some of the groups, thus inflating individual group standard deviations and leading to non-significant F-test. Welch's and EL test can deal with the skewness quite well, just as the tests based on the trimmed means. Ti is significant according to F-test at 0.01 level, but not significant according to tests based on the trimmed means. Concentrations of Ti data contain outliers but are not very skewed. Thus the tests based on the trimmed means mitigate the outlier effect. Sb is is an interesting case where Welch's test and EL tests reject H_0 at 0.01 level, while other

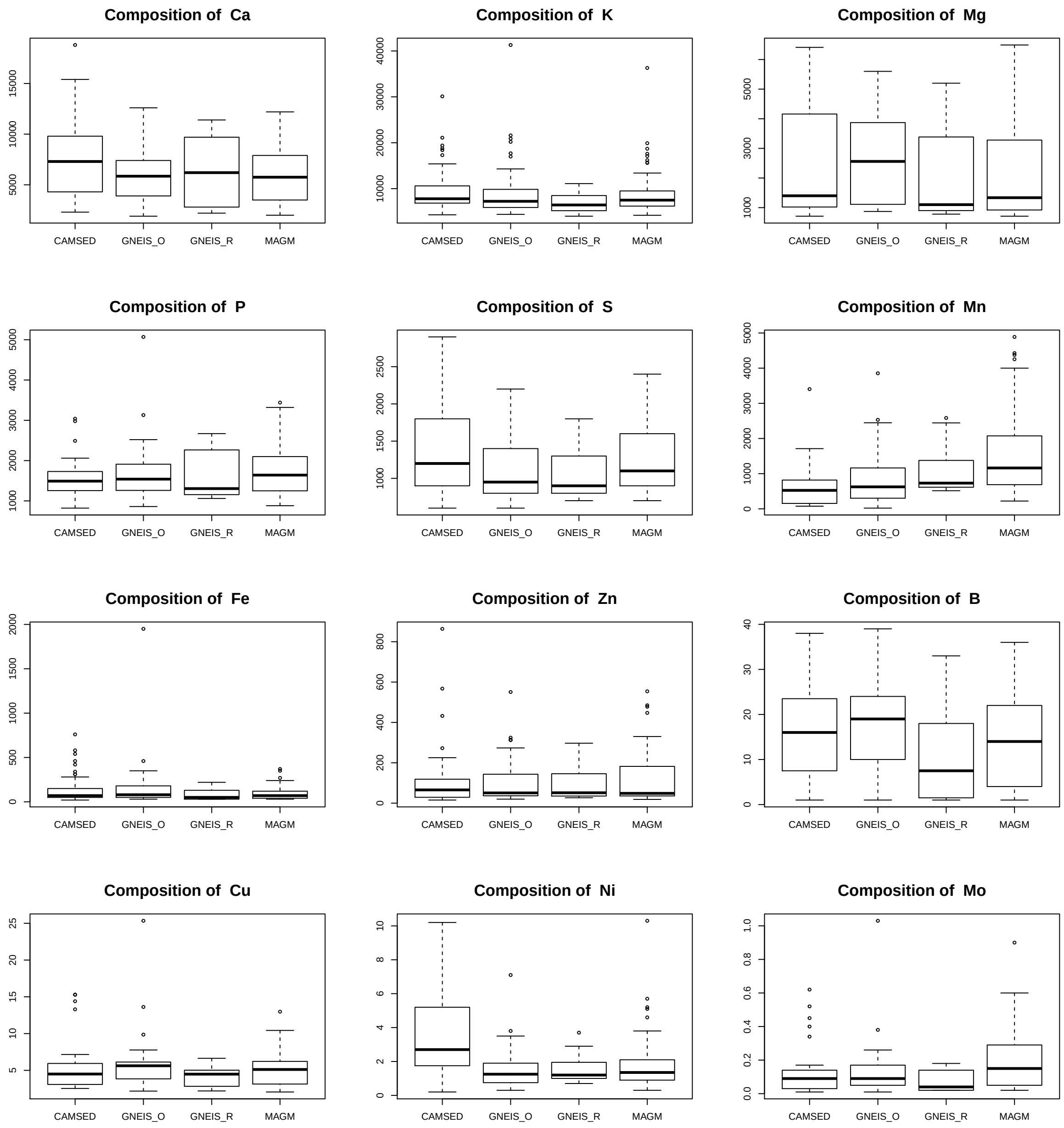


Figure 6.8: Box plots of the Oslo transect data [37] chemical elements (macronutrients and essential micronutrients) grouped by lithology type . Concentration units are mg/kg.

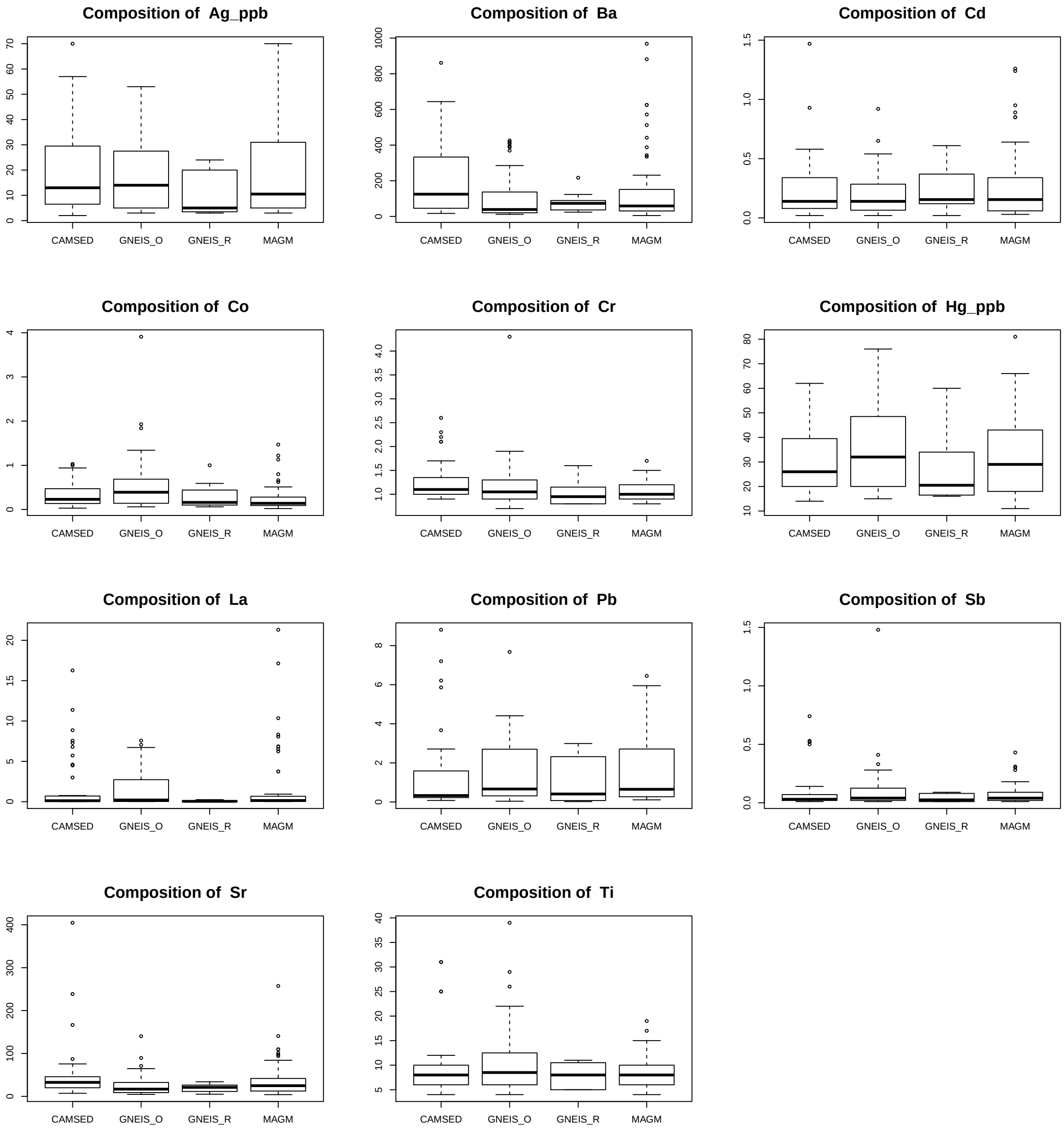


Figure 6.9: Box plots of the Oslo transect data [37] chemical elements (trace elements) grouped by lithology type. Concentrations units are mg/kg, except for Ag and Hg, for which the units are µg/kg.

Table 6.18: p-values from tests of equality of means and trimmed means of 23 chemical element concentrations in plants collected along the Oslo Transect [37]. Symmetric trimming, $\alpha_i = \beta_i = c$, $i = 1, \ldots, 4$.

				Trimming level					
				$c = 5\%$		$c = 10\%$		$c = 20\%$	
Element	F-test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
Ag	0.260	0.102	0.088	0.220	0.228	0.421	0.405	0.739	0.731
B	0.077	0.085	0.073	0.095	0.092	0.121	0.112	0.179	0.161
Ba	0.015	0.009	0.006	0.030	0.031	0.019	0.017	0.003	0.002
Ca	0.149	0.192	0.177	0.218	0.223	0.315	0.311	0.423	0.410
Cd	0.081	0.046	0.038	0.094	0.095	0.054	0.048	0.031	0.023
Co	<0.001	0.011	0.008	0.001	0.001	<0.001	<0.001	<0.001	<0.001
Cr	0.167	0.001	<0.001	0.001	0.001	0.001	0.001	0.005	0.002
Cu	0.440	0.264	0.244	0.664	0.672	0.768	0.762	0.756	0.750
Fe	0.026	0.017	0.013	0.035	0.036	0.022	0.019	0.043	0.034
Hg	0.308	0.287	0.268	0.351	0.373	0.191	0.184	0.397	0.383
K	0.473	0.279	0.259	0.500	0.515	0.531	0.523	0.583	0.567
La	0.275	<0.001	<0.001	0.005	0.005	0.134	0.101	0.008	0.005
Mg	0.241	0.228	0.213	0.279	0.276	0.379	0.374	0.573	0.562
Mn	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Mo	0.017	0.002	0.001	0.017	0.016	0.035	0.029	0.166	0.151
Ni	<0.001	<0.001	<0.001	<0.001	<0.001	0.001	<0.001	0.017	0.012
P	0.284	0.252	0.236	0.395	0.398	0.435	0.428	0.584	0.571
Pb	0.524	<0.001	<0.001	0.008	0.009	0.007	0.005	0.001	0.001
S	0.584	0.550	0.535	0.701	0.719	0.779	0.778	0.811	0.807
Sb	0.164	0.007	0.005	0.211	0.223	0.189	0.186	0.246	0.204
Sr	0.139	0.073	0.064	0.179	0.187	0.222	0.218	0.101	0.088
Ti	0.005	0.007	0.005	0.057	0.055	0.092	0.085	0.084	0.069
Zn	0.884	0.800	0.792	0.966	0.968	0.973	0.972	0.965	0.965

tests yield $p > 0.1$. The concentrations for this element are extremely skewed, around one third of the observations taking the smallest possible concentration value 0.01mg/kg and containing some extreme outliers. The tests based on the trimmed means mitigate the outlier effect and are not significant.

These various situations show again that for each particular situation the researcher needs to be clear about the aims – what hypothesis needs to be tested and whether the related assumptions hold.

6.4 Discussion

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 symmetrical $N(0, 1)$ 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). It should be remarked that 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 Huber test version with V estimated by simulations converged to the nominal 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 [52], where it was concluded that fixing $V = 2.046$ yields empirical coverage lower than the 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 [52], however, the interpretation of this result may lay in the degree of skewness of the distributions analysed. Consider the estimate of the population skewness parameter $\hat{\kappa}$ for some of the distributions considered, obtained by $N = 1,000,000$ Monte Carlo simulations: for χ^2_1 distribution $\hat{\kappa} = 2.83$, for χ^2_3, $\hat{\kappa} = 1.61$, while for Gamma($a = 1, s = 1/20$), the most skewed setting considered in [52], $\hat{\kappa} = 2.01$.

In general, the empirical level of the EL test for the trimmed means for very small samples (under 20) fluctuates greatly (consider, for example, the empirical levels for contaminated normal distribution in Figure 6.1). There might be two reasons for these fluctuations. First, the way the constant a_2 in Theorem 4.2.1 is estimated, see Remark 4.2.2. It is clear that for small samples, the changes in the estimate of a_2 are sensitive to small changes in the sample sizes n_1, n_2. Second, it has been pointed out by Owen [31] that for small samples, an approximation of the limiting law for the empirical log likelihood ratio by F distribution can be more precise than an approximation by the chi-square distribution. Namely, if the limiting distribution is χ^2_k, then the confidence intervals and p-values for a sample of size n can be calculated, using the $F_{k,n-1}$ distribution. F-distribution approximation, however, was not considered by Qin and Tsao for their EL method for the trimmed means [33]. The tests based on the Huber estimators are more stable for small sample sizes, nevertheless, their accuracy could probably still be improved by using the F-approximation.

Regarding the power of the tests, the power of the EL test for the difference of trimmed means is close to the power of Student's t-test under the standard normal distribution, while exceeding it considerably for nonnormal distribution settings, except for the moderately skewed χ^2_3 distribution. Moreover, the new test has a comparable power to that 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, in case of the $N(0,1)$ and 5% contaminated normal distribution 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 consequence of an incorrect test level at H_0.

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 was robust only for large sample sizes.

It should also be remarked that despite the data-adaptive nature of the EL method, the EL test for the difference of the means is not robust for most nonnormal distributions and samples of small to moderate size (this was also noted in [57]). Moreover, it has much lower power. Also for the real data sets considered, EL means test mostly yields p-value closer to Student's t and Welch's test than to the tests based on robust estimators.

Based on these observations, the new EL-based test for the difference of trimmed means can be recommended for use in moderate and large sample settings under departures from normality, such as heavy tails, presence of outliers and variance heterogeneity. For small samples, 10% trimming should be preferred to 20% trimming, or Yuen's test for the trimmed means should be used. In case there are heavy tails or outliers in the data, for small samples, EL test for the Huber's estimator with V estimated can be recommended for use in preference to the EL test for the trimmed means as it is less sensitive to actual sample size value. However, it is not recommended to use Huber's test in case of very skewed data sets or situations when skewness is combined with variance heterogeneity in small samples.

In practice, considering the data examples of Chapter 6.3, the EL test for the difference of trimmed means yields p-values similar to the robust Yuen's test for the trimmed means. As expected, the tests based on the trimmed means can lead to the opposite conclusion about H_0 than the tests based on the means. We observe that the p-values of the EL test and Yuen's test for the difference of trimmed means are quite close, and this is true also for small sized samples (such as *cotton*, *ozone*, or *nma*) and 20% trimming. The confidence intervals of the EL-based tests are somewhat shorter. Regarding the test for the difference of two Huber estimators, we note 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. In cases of moderate skewness (such as *NMA, ozone, IQ*), the Huber test with V estimated yields p-values close to the EL test for the means.

Finally, regarding the newly-established EL ANOVA test for the trimmed means, the simulations with $k = 3$ groups sampled from skewed distributions with equal variances show that the test converges to the empirical level for all distributions. 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. Similarly, as in the two-sample case, the EL ANOVA results for the trimmed means for small samples could probably be improved by using an appropriate F-approximation of the limiting law.

Simulations with $k = 2$ 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 $k = 3$ case, the $k = 2$ case reveals that EL ANOVA test 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 $k = 2$ give only a limited view on the behaviour of ANOVA-like methods. The data of Oslo transect [37] analysed in Chapter 6.3 shows that in practical situations the p-values given by the EL based test for the trimmed means are quite close to Yuen-Welch test. It should be noted, however, that the Oslo data set fits rather in a large-sample setting, the total number of observations for the four groups being 332.

Conclusions

The main aims of the thesis research have been achieved. New empirical likelihood-based methods for comparing two and more independent populations based on robust location parameter estimators were 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-like 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 [35], 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's estimator fits under the established conditions. The smoothing principle established by F. Hampel et al. [12] is important, since it allows for constructing smooth EL estimating functions essential for the conditions to hold.

We generalized the one-sample EL for the trimmed means by G. Qin and M. Tsao [33] 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 [39]. In the case of the EL ANOVA-like method for the trimmed means, it was demonstrated that 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 EL ANOVA for the equality of means established by A. B. Owen in [30].

A large 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 heterogeneity do not hold. It can be said that the 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. Some improvement for small-sample cases could probably be gained, using an F approximation for the limiting chi-square laws, see discussion in Chapter 6.4. It should be noted that the power of the methods considerably exceeds that of the methods based on the means for the most of the distributional settings considered.

As an extension of the thesis research, one might consider the difference of other M-estimators than the smoothed Huber estimator, for example, the bi-square M-estimator that was demonstrated to have good robustness properties in [13]. Note, that EL-based methods considered in this thesis research are based on smoothed estimating functions. An alternative EL approach exists that is based on non-smooth criterion functions developed by [27]. 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. [52]

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. [51], [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] P. Hall and A. H. Welsh. Convergence for estimates of parameters of regular variation. *The Annals of Statistics*, 12(3):1079–1084, 1984.

[12] 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.

[13] 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.

[14] F. R. Hampel, E. M. Roncheti, P. J. Rousseeuw, and A. W. Stahel. *Robust Statistics. The Approach Based on Influence Functions*. Wiley, New York, 1986.

[15] S. Heritier, E. Cantoni, S. Copt, and M. P. Victoria Feser. *Robust Methods in Biostatistics*. Wiley, 2009.

[16] 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.

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

[18] P. J. Huber. The 1972 Wald Lecture Robust Statistics: A Review. *The Annals of Mathematical Statistics*, 43(4):1041 – 1067, 1972.

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

[20] H. J. Keselman, R. R. Wilcox, R. K. Kowalchuk, and S. Olejnik. Comparing trimmed or least squares means of two independent skewed populations. *Biometrical Journal*, 44(4):478–489, 2002.

[21] H. J. Keselman, R. R. Wilcox, A. R. Othman, and K. Fradette. Trimming, transforming statistics, and bootstrapping: Circumventing the biasing effects of heterescedasticity and nonnormality. *Journal of Modern Applied Statistical Methods*, 1(2):288–309, 2002.

[22] Y. Kitamura. Empirical likelihood methods with weakly dependent processes. *The Annals of Statistics*, 25(5):2084–2102, 1997.

[23] 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.

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

[25] A. Marazzi. Bootstrap tests for robust means of asymmetric distributions with unequal shapes. *Computational Statistics & Data Analysis*, (49):503–528, 2002.

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

[27] 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.

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

[29] A. B. Owen. Empirical likelihood confidence regions. *The Annals of Statistics*, 18:90–120, 1990.

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

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

[32] M. Pagano and K. Gauvreau. *Principles of Biostatistics*. Duxbury, Pacific Grove, CA, 2nd edition, 2000.

[33] 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.

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

[35] 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.

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

[37] C. Reimann, A. Arnoldussen, R. Boyd, T.E. Finne, F. Koller, and Oe. Nordgulen. Element contents in leaves of four plant species (birch, mountain ash, fern and spruce) along anthropogenic and geogenic concentration gradients. *Science of the Total Environment*, 377:416–433, 2007.

[38] R. G. Staudte and S. J. Sheather. *Robust Estimation and Testing*. Wiley, New York, 1990.

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

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

[41] V. Todorov and P. Filzmoser. An object-oriented framework for robust multivariate analysis. *Journal of Statistical Software*, 32(3):1–47, 2009.

[42] V. Todorov and P. Filzmoser. Robust statistic for the one-way MANOVA. *Computational Statistics & Data Analysis*, 54(1):37–48, 2010.

[43] M. Tsao and J. Zhou. On the robustness of empirical likelihood ratio confidence intervals for location. *The Canadian Journal of Statistics*, 29(1):129–140, 2001.

[44] M. Tsao and J. Zhou. A nonparametric confidence interval for the trimmed mean. *Journal of Nonparametric Statistics*, 14(6):665–673, 2002.

[45] W. Tsao and C. Wu. Empirical likelihood inference for a common mean in presence of heteroscedasticity. *The Canadian Journal of Statistics*, 34(1):45–59, 2006.

[46] J. W. Tukey. A survey of sampling from contaminated distributions. In I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, and H. B. Mann, editors, *Contributions to probability and statistics: Essays in Honor of Harold Hotelling*, pages 448–485. Stanford University Press, Redwood City, CA, 1960.

[47] J. W. Tukey and D. H. McLaughlin. Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/winsorization 1. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 25(3):331–352, 1963.

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

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

[50] A. Vexler and J. Yu. *Empirical Likelihood Methods in Biomedicine and Health*. Chapman and Hall/CRC, Boca raton, 2018.

[51] 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.

[52] 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.

[53] L. Wasserman. *All of Nonparametric Statistics*. Springer, New York, 2006.

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

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

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

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

[58] S. S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9:60–62, 1938.

[59] C. Wu and Y. Yan. Empirical likelihood inference for two-sample problems. *Statistics and Its Interface*, 5, 2012.

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

[61] M. Zhou. *Empirical Likelihood Method in Survival Analysis*. Chapman and Hall/CRC biostatistics series. Chapman and Hall/CRC, New York, 2016.

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 step on the path of a 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.