The Pennsylvania State University
The Graduate School

LOGARITHMIC QUANTILE ESTIMATION AND ITS APPLICATIONS TO NONPARAMETRIC FACTORIAL DESIGNS

A Dissertation in
Statistics
by
Lucia Tabacu

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2014

The dissertation of Lucia Tabacu was reviewed and approved* by the following:

Manfred Denker
Professor of Mathematics
Dissertation Advisor and Co-Chair of Committee

Michael Akritas
Professor of Statistics
Co-Chair of Committee

John Fricks
Associate Professor of Statistics

Guodong Pang
Assistant Professor of Industrial Engineering

David Hunter
Professor of Statistics and Department Head

*Signatures are on file in the Graduate School.

Abstract

In this dissertation we prove an almost sure weak limit theorem for simple linear rank statistics for samples with continuous distribution functions. As a corollary, the result extends to samples with ties and the vector version of an almost sure central limit theorem for vectors of linear rank statistics. Moreover, we derive a weak convergence result for some quadratic forms. These results are then applied to quantile estimation and to hypothesis testing for nonparametric statistical designs, here demonstrated by the c-sample problem, where the samples may be dependent. In general, the method is known to be comparable to the bootstrap and other nonparametric methods (Thangavelu [97], and Fridline [47]) and we confirm this finding for the c-sample problem.

This dissertation also contains a study of longitudinal data originally analyzed by Lumley [75] using odds ratio and later by Brunner, Domhof, and Langer [21] using rank statistics. The quantile estimation procedure developed in this dissertation is well adapted to this situation and shows similar results as in Brunner et al. [21] with the advantage of minimal assumptions on the distributions.

In order to develop the theory necessary to derive these new methods, we further develop Thangavelu's first ideas in [97] and set the framework for a decision theory based on the almost sure convergence property. Although this method seems to be similar to bootstrap it goes beyond these ideas because it is based on almost sure behavior and not on distributional behavior. The same novelty occurs when stating and proving the results on rank statistics and their quadratic forms; some ideas are similar to the distributional approach in Brunner and Denker [20] and Brunner et al. in [21], but the essential new idea is to deal with almost sure terms of approximating statistics, similar to the well known Slutsky result in the distributional theory, using results by Lifshits [73] and others.

Table of Contents

List of Figures **vi**

List of Tables **vii**

Acknowledgments **ix**

Chapter 1
Introduction **1**

Chapter 2
Almost Sure Weak Convergence **10**
2.1 Almost Sure Limit Theorems 10
2.2 General Results in Almost Sure Limit Theory 14
2.3 Examples of Almost Sure Limit Theorems 19

Chapter 3
Statistical Applications of Almost Sure Limit Theorems **24**
3.1 Statistical Decision Theory based on Logarithmic Quantiles 24
3.2 Established Results 31
3.2.1 Hypothesis Testing based on ASCLT 31
3.2.2 Almost Sure Confidence Intervals for the Correlation Coefficient 36

Chapter 4
The Almost Sure Central Limit Theorem for Rank Statistics **40**
4.1 Introduction to Linear Rank Statistics 40
4.2 The Main Result and Remarks 42
4.3 Proofs of Theorem 4.2.1 and Corollary 4.2.3 46

Chapter 5
Factorial Designs I **58**
5.1 The Almost Sure Weak Convergence of the Kruskal-Wallis Statistic . . . 59
5.2 Simulation Results . . . 64

Chapter 6
Factorial Designs II **73**
6.1 Introduction to Nonparametric Factorial Designs . . . 73
6.2 The Shoulder Tip Pain Study . . . 76
6.3 Technical Details . . . 84

Chapter 7
Conclusion **91**
7.1 Summary . . . 91
7.2 Open Problems . . . 93

Bibliography **94**

List of Figures

3.1 Type II error curves for t-test and ASCLT-method 1; sample size 40, level 0.1 (courtesy of K. Thangavelu) 35
3.2 Type II error curves for t-test and ASCLT-method 1; sample size 40, level 0.05 (courtesy of K. Thangavelu) 35
3.3 Type II error curves for t-test and ASCLT-method 1; sample size 50, level 0.1 (courtesy of K. Thangavelu) 36
3.4 Type II error curves for t-test and ASCLT-method 1; sample size 50, level 0.05 (courtesy of K. Thangavelu) 36

List of Tables

5.1 Averaged empirical logarithmic α-quantiles and the squared estimated standard errors for three independent samples with distribution Exp(3) . . . 66

5.2 Averaged empirical logarithmic α-quantiles and the squared estimated standard errors for three independent samples with distribution $\mathcal{N}(2,1)$ . . . 66

5.3 The level of significance for three $\mathcal{N}(0,1)$ dependent samples; 200 simulations and 20 permutations, different sample sizes and different values of α . . . 68

5.4 The level of significance for three Exp(4) dependent samples; 200 simulations and 20 permutations, different sample sizes and different values of α . . . 68

5.5 The level of significance for three $\mathcal{N}(2,1)$ independent samples; 200 simulations and 20 permutations, different sample sizes and different values of α . . . 68

5.6 The level of significance for three Exp(3) independent samples; 200 simulations and 20 permutations, different sample sizes and different values of α . . . 69

5.7 Power for three dependent samples from a normal distribution with different means at level $\alpha = 10\%$; 200 simulations and 20 permutations . . . 69

5.8 Power for three dependent samples from a normal distribution with different means at level $\alpha = 5\%$; 200 simulations and 20 permutations 69

5.9 Power for three dependent samples from a normal distribution with different means at level $\alpha = 1\%$; 200 simulations and 20 permutations 70

5.10 Power for three dependent samples from an exponential distribution with different means at level $\alpha = 10\%$; 200 simulations and 20 permutations . . . 70

5.11 Power for three dependent samples from an exponential distribution with different means at level $\alpha = 5\%$; 200 simulations and 20 permutations 70
5.12 Power for three dependent samples from an exponential distribution with different means at level $\alpha = 1\%$; 200 simulations and 20 permutations 70
5.13 Power for three independent samples from a normal distribution with different means at level $\alpha = 10\%$; 200 simulations and 20 permutations 70
5.14 Power for three independent samples from a normal distribution with different means at level $\alpha = 5\%$; 200 simulations and 20 permutations 71
5.15 Power for three independent samples from a normal distribution with different means at level $\alpha = 1\%$; 200 simulations and 20 permutations 71
5.16 Power for three independent samples from an exponential distribution with different means at level $\alpha = 10\%$; 200 simulations and 20 permutations 71
5.17 Power for three independent samples from an exponential distribution with different means at level $\alpha = 5\%$; 200 simulations and 20 permutations 71
5.18 Power for three independent samples from an exponential distribution with different means at level $\alpha = 1\%$; 200 simulations and 20 permutations 71

6.1 The value of the statistic and the empirical logarithmic p-value for per=100 permutations 82
6.2 The p-values for the ANOVA-type statistics; taken from Brunner et al. [21], page 191 83

Acknowledgments

I would like to thank my adviser for his excellent supervision and support.

Chapter 1

Introduction

In this dissertation, we develop a new statistical theory for linear rank statistics based on almost sure central limit theorems. Rank statistics are an important part of classical nonparametric statistics because of their advantages in applications and good statistical properties. Ranks form a maximal invariant statistic under the transformation group of strictly monotone mappings (Lehmann and Romano[71], chapter 6), are often locally optimal (Hajek, Sidak and Sen [53]), have minimal distributional assumptions, have good robustness properties, and are widely applicable. Rank statistics have existed as long as many parametric statistical method. Bradley [18] mentions that John Arbuthnott used the sign test in 1710 to determine the birth rates of boys and girls. The modern beginning of the rank tests theory goes back to 1930's with papers by Hotelling and Pabst [59], Friedman [48], Kendall [63], Smirnov [94], Wald and Wolfowitz [100], Wilcoxon [101], and others. By the 1970's, the theory of rank statistics was well established and collected in the books by Puri and Sen [86], Puri [85] and later on Puri and Sen [87], and others. The representation of rank statistics as stochastic integrals (see Denker and Rösler [35]) and the theory of nonparametric hypotheses of Akritas and Arnold [1] allowed a considerable extension of the rank statistics theory in the 1990's. This led to new rank test procedures for complex factorial designs (see the survey article by Brunner and Puri [23]). This dissertation builds upon this theory. Many of the rank tests developed in the 1990's require estimation of covariance matrices of specific structures, and this introduces additional model assumptions that are hard to check in practice. As an example, the nonparametric Behrens-

Fisher problem requires covariance estimation, but Thangavelu [97] showed that it can be solved effectively using an estimation procedure based on an almost sure central limit theorem. Using ideas from Thangavelu [97], we were able to develop a general theory for rank statistics based on almost sure limit theorems. The theory of nonparametric factorial designs that we set in this dissertation is independent of specific additional assumptions on distributions and correlation matrices, which makes applications more robust and reliable. We will explain this in the following paragraphs.

The almost sure limit theorems (ASLT) are based on a new type of summation procedure for a variety of statistics to obtain their asymptotic distribution. Although the first ideas of an almost sure limit theorem may be traced back to Lévy [72], it first appeared in two papers by Brosamler [19] and Schatte [91], and later in Lacey and Philipp [69]. There are many applications that motivated the development of almost sure limit theorems: testing random number generators by considering only one sequence of numbers or investigating the empirical distribution function of different statistics in estimation, prediction, or hypothesis testing problems. The application of the almost sure limit theorems to statistical functionals is a fairly new subject.

It has been recently observed by Thangavelu [97] and Fridline [47] that the logarithmic summation of data may be used for quantile estimation in practice. The philosophy behind this procedure resembles Efron's bootstrap method [44], but resampling is not needed in the almost sure method. In this dissertation, we extend the applicability of the almost sure quantile estimation to rank statistics. We call it logarithmic quantile estimation (LQE). We develop the method for general rank models as defined in Brunner and Denker [20], but adapted to the needs of the almost sure concept. Let $X_1, ..., X_N$ be a sample of random vectors, and for each $n \leq N$ let T_n be a statistic based on $X_1, ..., X_n$. The logarithmic average of the sequence T_n has the form

$$\widehat{G}_N(t) = \frac{1}{C_N} \sum_{n=1}^{N} \frac{1}{n} \mathbb{I}(T_n \leq t),$$

where C_N is chosen to make $\widehat{G}_N$ an empirical distribution function, and where

$\mathbb{I}(C)$ denotes the indicator function of the set C. Note that $C_N \asymp \ln N$ (defined by $\frac{C_N}{\ln N} \to 1$ as $N \to \infty$), which is responsible for the name *log average*. In fact, the usage of $\widehat{G}_N$ is as the classical empirical distribution functions. More details, like the corresponding Glivenko-Cantelli theorem, can be found in Thangavelu [97]. The empirical $\alpha-$quantile of $\widehat{G}_N$ can then be used in hypothesis testing, for example a typical rejection region may look like $\{X \in \mathbb{R}^{dN} : |T_N| \geq \widehat{t}_\alpha^{(N)}\}$ with $\widehat{G}_N(\widehat{t}_\alpha^{(N)}) = \alpha$.

Let us give a brief overview of some important results on almost sure limit theorems, which will form the base for the validity of good test procedures. After the work of Brosamler [19], Schatte [91] and Lacey and Philipp [69] was published, many important results have been obtained for both independent and dependent random variables, for random vectors and stochastic processes as well as their almost sure functional versions. The simplest form of the almost sure central limit theorem is

$$\lim_{n\to\infty} \frac{1}{\ln n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}\left(\frac{S_k(\omega)}{\sqrt{k}} < x\right) = \Phi(x) \text{ for almost all } \omega \in \Omega,$$

and all numbers x, where $X_1(\omega), X_2(\omega), ...$ are independent and identically distributed random variables defined on a probability space $(\Omega, \mathcal{B}, P)$ with $E(X_1) = 0, E(X_1^2) = 1$, $S_k(\omega) = X_1(\omega) + ... + X_k(\omega)$, $\mathbb{I}$ is the indicator function and Φ is the standard normal distribution function. A starting point for learning more about the almost sure limit theorems are the survey paper by Berkes [10], and on the more mathematical side, the lecture notes by Lifschits [74]. Berkes and Csáki [11] obtained general results that extend weak limit theorems for independent random variables to their almost sure versions. Applications of their results are the almost sure limit theorems for partial sums, maxima of partial sums, extremes (see also Fahrner and Stadtmüller [46], Cheng et al. [29], Peng et al. [84] discuss some practical issues of their result based on Thangavelu's method [97]), empirical distribution functions, U-statistics (see also Holzmann, Koch and Min [55]), local times, return times, and Darling-Erdös type limit theorems (see also Berkes and Weber [14]). Berkes and Dehling [12] proved a general almost sure central limit theorem for independent, not necessarily identically distributed random variables. Peligrad and Shao [83] considered the almost sure central limit theorem for weakly dependent random variables (see also Hurelbaatar [60], [61], Matula [78]), Lifshits [73] obtained the almost sure limit theorem for sums of random vectors, and Chuprunov and Fazekas [32] treated the case of Pearson statistic. Bercu [9] studied the almost sure limit theorem for vector martingale transform and applied it to cumulative prediction and estimation errors associated to the stable autoregressive processes and branching processes with immigration. Thangavelu [97] proved an almost sure central limit theorem for the two-sample linear rank statistics and developed a hypothesis testing procedure based on the almost sure convergence. He applied his method to problems like testing for the mean in the parametric one-sample problem, testing for the equality of the means in the parametric two-sample problem, and for the nonparametric Behrens-Fisher problem. As a result he showed that the LQE method is better than bootstrap and almost as good as the t-test for the two sample problem. For the nonparametric two-sample Behrens-Fisher problem he compared the LQE method with the methods in Babu and Padmanabhan [7], Reiczigel, Zakariàs and Rózsa [88] and Brunner and Munzel [22]. It is shown that the LQE-method performs stably over various distributions, is comparable to the other methods and often preferable. Later, Denker and Fridline [37] obtained the almost sure version of Cramér's theorem. Using this result, Fridline [47] showed the almost sure central limit theorem for the correlation coefficient and applied the almost sure version of Cramér's theorem to obtain confidence intervals for the population correlation coefficient. It turns out that the LQE method is superior to bootstrap for these statistics.

All these results show that the LQE method should be developed further, in particular, for nonparametric designs when bootstrap methods are difficult to apply. Steland [96] has obtained a bootstrap method for simple rank statistics using the von Mises method, assuming complete independence. Here we are interested in the general result, when samples are not identically distributed and hence the use of resampling becomes questionable. The LQE method does not have this restriction, and we show that it provides good results. As a special design we chose the c-sample problem regardless of whether the samples are independent or not: we show that the test based on the LQE method provides better coverage probability than the classical Kruskal-Wallis test. We also show that in the dependent situation the LQE test has a satisfactory performance. For other designs, we obtained similar results. Besides this advantage, the LQE-method estimates quantiles directly from the data, not using the asymptotic distribution, hence it also does not use any estimation of unknown variances or covariances or eigenvalues of covariance matrices. It is also applicable when asymptotic covariance matrices become degenerate.

The second application given in this dissertation is concerned with a real data set. Statistical decision theory for longitudinal data has attracted much attention over the past 3 decades. Various model assumptions for the parametric and nonparametric case have been proposed and investigated, with only moderate success due to lack of information about the structure of the finite time series involved. Modeling the joint distribution of the time series is a major draw back. A minor problem here is the estimation of unknown parameters and the speed of approximation to the distribution which is used to determine the quantiles for the testing procedure. In order to overcome these difficulties, Akritas and Arnold [1] developed a complete theory of nonparametric hypothesis testing in factorial designs. This was complemented by Akritas, Arnold and Brunner in [2] among other work. As reasonable test statistics rank methods were proposed, either rank transform or linear rank statistics under Wilcoxon scores. It is argued that for repeated measure designs or longitudinal data the gap in efficacy compared to parametric procedures is not as important to close as it is to avoid errors occurring by model misspecification. This leads to a simple statistical model where distributional assumptions are kept to a minimum of conditions. Such procedures are given by rank statistics, which reduces the statistical decision problem to the original Bernoulli idea of making decisions on coin tossing. The test statistics encountered in previous work is based on Wilcoxon scores which is sufficiently general (see Compagnone and Denker [33] how to use Wilcoxon scores to improve efficiency of procedures compared to other score functions). The program can be carried out in full generality due to the central limit theorem for linear rank statistics under arbitrary dependencies in Brunner and Denker [20]. Using this method and its refinements by Munzel [79] and others (see [23] for an overview) a nearly complete evaluation of problems involving nonparametric factorial designs has been established over the years.

Chapter 6 of this dissertation adds to this endeavor and introduces a new method of hypothesis testing for longitudinal data motivated by the study of Lumley [75] on a "shoulder tip pain" data set and subsequentially re-investigated in Brunner, Domhof and Langer [22]. In this study, patients were scored at 6 time instances on different days according to their subjective feeling of pain levels. The group was stratified according to gender and treatment. It turned out that the corresponding Wald-type test statistic was badly approximated by its limiting distribution while the ANOVA-type statistic gave good results using the Box-Welch-Satterhwaite approximation (see [22], p. 72 and 191). The second statistic can only be used by making additional assumptions on the distribution of the data, while the first method requires a huge data set. It is therefore desirable to have a procedure which is independent of such additional assumptions.

The new method we have introduced in Denker and Tabacu [39] may well serve to overcome this unpleasant feature. Logarithmic quantile estimation is based on almost sure central limit theorems and has, for the first time, been set up in [39] in the general framework of simple linear rank statistics. This result parallels the one in Brunner and Denker [20] in as much as the assumptions are the same; the only change which has to be made is that the observations are defined on some fixed probability space (due to the a.s. behavior). It is evident that the result persists in the presence of ties; the ranks have to be interpreted as mid-ranks as in Munzel [79]. Previous results on the logarithmic quantile method are concerned with two sample linear rank statistics and the Behrens-Fisher problem (Thangavelu [97]), the comparison of the logarithmic quantile method and the t-test (Thangavelu [97]), the correlation coefficient in Fridline [47], and the c-sample problem under dependencies in Denker and Tabacu [39]. In all cases, it was found that the logarithmic quantile method works well as a nonparametric procedure with minimal assumptions on distributions. In this dissertation, we set up the testing procedure for longitudinal data based on the results in [20] and [39]. We also evaluate the test statistics for the "shoulder tip pain" study. It turns out that we can assure the findings in [22] with negligible changes due to the generality of this method.

We give an overview of the organization and of the results of this dissertation. Since our results rely heavily on the theory of almost sure limit theorems, we begin in Chapter 2 explaining the necessary background. Recalling the classical theorems of Brosamler [19], Schatte [91] and Lacey-Philipp [69] leads to the groundbreaking result of Berkes and Csáki [11] for applications of the theory for statistics. Reviewing these results, it is evident that our theorem on rank statistics is entirely new and covered only in very special cases by the results in [11]. In Chapter 2, we include some general results from Berkes and Csáki [11] and their applications to different statistics like empirical distribution functions and U-statistics. The difference between Cesaro summation and logarithmic average in the context of almost sure weak convergence is also explained. These theorems can be generalized by considering different weights than the logarithmic ones. The idea that we pursue here is that in the general framework of an almost sure weak convergence, it is enough to consider only one sequence of random variables for statistical testing, and not multiple sequences like in the classical central limit theorems. In this chapter, we justify that the almost sure central limit theorem for rank statistics fills a gap in the literature, and its properties are worth being studied from an application point of view.

In Chapter 3, we rely on Thangavelu's initial ideas [97] but also extend and improve his results. We develop a general foundation for decision theory based on almost sure limit theorems. If for any chosen statistic, an almost sure limit theorem and a weak convergence theorem hold with the same limiting distribution, normal or not, known variance or not, then the logarithmic empirical distribution function provides a correct approximation to the distribution function. This also leads to a correct approximation of the true quantiles of the test statistic considered, and it can be used for hypothesis testing. The partial sum of a sequence of i.i.d random variables or vectors, normalized by $\sqrt{n}$ is a good example for this framework. Berkes and Csaki [11] first noticed that for a statistic of independent random variables satisfying a weak convergence theorem, it is possible to obtain its almost sure weak limit theorem. This idea supports our framework for hypothesis testing and makes it a natural assumption. In this chapter, we also include some of the results that Thangavelu [97] and Fridline [47] obtained in their dissertations. Thangavelu [97] tested parametric and nonparametric hypotheses using the empirical logarithmic quantiles. In the parametric context, he considered normally distributed samples and tested the null hypothesis that the mean is zero, and in the nonparametric case he considered the Behrens-Fisher problem as an application of the two sample rank test. Fridline [47] gave a different interpretation to the almost sure limit theorem. He obtained the almost sure weak convergence of Cramer's theorem and used it to obtain confidence intervals for the correlation population.

Chapter 4 is the central part of this dissertation. It contains the main result upon which all special rank tests in this thesis are based. The almost sure central limit theorem for simple linear rank statistics is proved for a sequence of independent random vectors with possible dependent components and continuous distribution for each random variable. In practice the observations have continuous and discrete values. Thus, our result can be extended to these type of random variables by following the same type of argument from our proof and ideas from Munzel [79]. Munzel [79] has shown that the weak convergence of the simple linear rank statistics holds when the random variables have ties. His result is an extension of the result from Brunner and Denker [20]. The continuous distribution functions are replaced with the normalized versions since they adjust for ties. The proof of this extended version of our theorem is not included in this thesis. Our theorem is sufficient to obtain the result for quadratic forms. It follows from Lifshits' [73] almost sure limit theorem for random vectors. When we have to deal with problems that involve more than one sample or multiple groups of subjects, we consider a simple linear rank statistic for each sample or group and use the same decomposition for the vector of rank statistics. We obtain the almost sure weak convergence for this vector of rank statistics using Fridline's lemma [47] that holds for random vectors.

Chapter 5 contains the first application of our results. The Kruskal-Wallis statistic for testing the equality of the distribution functions for c-samples, independent or dependent, becomes a function of linear rank statistics. For each of the c samples, a linear rank statistic can be defined, and under the null hypothesis the Kruskal-Wallis statistic can be expressed as a quadratic form of linear rank statistics. Using Lifshits [73] and the ideas of proof of our theorem, the Kruskal-Wallis statistic satisfies an almost sure limit theorem, and the logarithmic empirical quantiles are used to test the null hypothesis. For the case of independent samples, we provide two small data sets and compare the p-values obtained using the logarithmic quantiles, the chi-squared approximation and permutation test. For dependent samples simulated from exponential and normal distributions we calculated the level of significance and the power for different sample sizes.

We continue with the second application of our method based on almost sure weak convergence in Chapter 6. A longitudinal data set is considered and analyzed as a nonparametric factorial design, and for each rank test corresponding to main factors and their interactions the p-values are obtained using the logarithmic empirical quantiles. Here the tests are the rank versions of the ANOVA-type statistics and expressed as quadratic forms of the estimators of the relative marginal effects. We were able to write each test statistic in the form of the logarithmic average and prove an almost sure weak convergence for each of them. The p-values calculated are similar to the previous results obtained by Brunner et al [21]. In the last chapter, we summarize the ideas developed in this dissertation and discuss future projects.

Chapter 2

Almost Sure Weak Convergence

This chapter is dedicated to almost sure weak convergence. We start with some basic definitions and equivalent ways of expressing this concept. We consider the example of the partial sum of i.i.d random variables as the simplest form of an almost sure central limit theorem and its functional almost sure version. From an ergodic point of view, it is known that these theorems hold under the logarithmic averaging. Here we present some general results from Berkes and Csáki [11] that hold under logarithmic summation with different weights and show how these results can be applied to various statistics as the empirical distribution function, U-statistics or maxima of partial sums. We also introduce the almost sure limit theorem for random vectors (see Lifschits [73]), since it will be used extensively in the next chapters.

2.1 Almost Sure Limit Theorems

The general idea of the almost sure central limit theorem can be traced back to Lévy [72] and later to Brosamler [19] and Schatte [91]. These theorems are based on the almost sure weak convergence of empirical measures. We begin recalling definitions and results related to the weak convergence of measures in a general framework.

Let E be a metric space, $\mathcal{E}$ be the Borel $\sigma-$field and $\nu, \mu, \mu_1, \mu_2, ...$ be finite measures on $(E, \mathcal{E})$. Let $C_b(E) = \{f : E \to \mathbb{R}$ is continuous and bounded$\}$ and $\mathcal{M}_{\leq 1}(E) = \{\mu | \mu$ is finite and $\mu(E) \leq 1\}$. $\mathcal{M}_{\leq 1}(E)$ becomes a topological space such that the weak convergence of measures is the same as the convergence in this topology. This topology is defined by the open neighborhood basis of a measure ν as follows:

$$U(f_1, ..., f_k; \epsilon) := \{\mu \in \mathcal{M}_{\leq 1}(E) : \left| \int f_i d\nu - \int f_i d\mu \right| < \epsilon, 1 \leq i \leq k\},$$

where $\epsilon > 0$ and $f_1, ..., f_k \in C_b(E)$ (see Billingsley [15] or Klenke [64] for details).

Definition 2.1.1. We say that the sequence of measures μ_n converges weakly to μ, and we denote it by $\mu_n \Rightarrow \mu$ if $\int_E f d\mu_n \to \int_E f d\mu$, for all $f \in C_b(E)$.

The following theorem characterizes weak convergence and is a standard result (see Klenke [64], page 253 or Billingsley [15], page 16).

Theorem 2.1.2. (The portmanteau theorem of weak convergence)
Let $\mu, \mu_1, \mu_2, ... \in \mathcal{M}_{\leq 1}(E)$. The following conditions are equivalent:
(i) $\mu_n \Rightarrow \mu$,
(ii) $\int_E f d\mu_n \to \int_E f d\mu$ for all bounded Lipschitz continuous f,
(iii) $\int_E f d\mu_n \to \int_E f d\mu$ for all bounded uniformly continuous f,
(iv) $\int_E f d\mu_n \to \int_E f d\mu$ for all bounded measurable f continuous at μ-a.e. $x \in E$
(v) $\limsup_n \mu_n(F) \leq \mu(F)$ for all closed F,
(vi) $\liminf_n \mu_n(G) \geq \mu(G)$ for all open G,
(vii) $\mu_n(A) \to \mu(A)$ for all A in $\mathcal{E}$ with the property $\mu(\partial A) = 0$, where ∂A denotes the boundary of A.

Next we continue with definitions and examples of the almost sure central limit theorems. The first almost sure central limit theorem for a type of martingales was stated without proof in Lévy's book in 1937 ([72], page 270). In 1988 the almost sure limit theorems were rediscovered independently by Brosamler [19] and Schatte [91], and since then a vast literature on this subject has been developed. An almost sure limit theorem, also called a pointwise limit theorem, is equivalent to the almost sure weak convergence of the empirical measures associated with a sequence of random variables. More precisely, for the general form of the almost sure limit theorem we consider $(\xi_n)_{n\geq 1}$ a sequence of random variables defined on a common probability space $(\Omega, \mathcal{F}, P)$ with values in a metric space, and the sequence of the normalized sums

$$\zeta_k = \frac{1}{B_k}\sum_{j=1}^{k}\xi_j, \text{ where the constant } B_k > 0.$$

Define the empirical measures

$$Q_n(\omega) = \frac{1}{\gamma_n}\sum_{k=1}^{n} b_k \delta_{\zeta_k(\omega)}, \tag{2.1.1}$$

where $\omega \in \Omega$, δ_x is the point mass at x and $(b_k)_{k\geq 1}$ is a sequence of positive numbers such that

$$\gamma_n = \sum_{k=1}^{n} b_k \to \infty, n \to \infty. \tag{2.1.2}$$

The almost sure limit theorem is a statement of the form

$$Q_n(\omega) \Rightarrow G \text{ for almost all } \omega \in \Omega, \tag{2.1.3}$$

where $\Rightarrow$ denotes the weak convergence of measures and G is a probability measure (see Definition 2.1.1). Alternatively we write $P(Q_n \Rightarrow G) = 1$.

The simplest form of an almost sure central limit theorem for partial sums appears in the work of Brosamler [19], Schatte [91], Lacey and Philipp [69] and states that

$$\lim_{n\to\infty}\frac{1}{\log n}\sum_{k=1}^{n}\frac{1}{k}\mathbb{I}\left(\frac{S_k(\omega)}{\sqrt{k}} < x\right) = \Phi(x), \text{ for almost all } \omega \in \Omega, \tag{2.1.4}$$

and all numbers x, where $X_1(\omega), X_2(\omega), ...$ are independent and identically distributed random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ with $E(X_1) = 0, E(X_1^2) = 1$, $S_k(\omega) = X_1(\omega) + ... + X_k(\omega)$, $\mathbb{I}$ is the indicator function and Φ is the standard normal distribution function. Here and in the sequel $\log n$ is the natural logarithm of n.

Note that using Theorem 2.1.2, equation (2.1.4) can take different forms. It can be equivalently expressed as: there is a $P-$null set $N \subset \Omega$ such that for all

$\omega \in N^c$ we have

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} f\left(\frac{S_k(\omega)}{\sqrt{k}}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(u) e^{-\frac{1}{2}u^2} du,$$

for all bounded measurable functions which are continuous a.e.,
or

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}_A\left(\frac{S_k(\omega)}{\sqrt{k}}\right) = \frac{1}{\sqrt{2\pi}} \int_A e^{-\frac{1}{2}u^2} du,$$

for all Borel sets $A \subseteq \mathbb{R}$ with $\lambda(\partial A) = 0$,
or

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} \delta_{S_k(\omega)/\sqrt{k}} = N(0,1),$$

where $N(0,1)$ is the standard normal distribution on $\mathbb{R}$, δ_x is the point mass at $x \in \mathbb{R}$ and the convergence is weak convergence of measures on $\mathbb{R}$.

Under the same assumptions as in (2.1.4), Lacey and Philipp [69] obtained a functional version of the almost sure central limit theorem for the partial sums (see also Brosamler [19], with additional assumptions on the moments of the random variables). We state their result as follows: $P-$a.s.

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} \delta_{s_k(\cdot,\omega)} = W, \tag{2.1.5}$$

where "the broken line process" on $[0,1]$ is

$$s_n(t,\omega) = \begin{cases} \frac{S_k(\omega)}{\sqrt{n}}, & \text{if } t = \frac{k}{n}, k = 0,1,...,n \\ \text{linear in between}, & \text{otherwise} \end{cases}$$

and W is the Wiener measure on the space $C[0,1]$ of continuous functions on the unit interval (see Billingsley [15], page 86). The convergence type is weak convergence of measures. This functional almost sure central limit theorem corresponds to Donsker's version of the classical central limit theorem for partial sums. Note that (2.1.5) also implies (2.1.4). To see this, rephrase (2.1.5): $P-$a.s.

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} \delta_{\phi \circ s_k(\cdot,\omega)} = \phi(W),$$

for all measurable functions $\phi : C[0,1] \to \mathbb{R}$ which are continuous $W-$a.e. Here $\phi(W)$ denotes the image measure of W under ϕ. For the particular choice $\phi(x) = x(1)$, with $x \in C[0,1]$ we have that

$$\phi(s_n(\cdot,\omega)) = s_n(1,\omega) = \frac{S_n(\omega)}{\sqrt{n}} \text{ and } \phi(W) = W_1$$

with $W_1 \sim N(0,1)$ and we obtained (2.1.4). Kolmogorov's theorem shows the existence of a stochastic process W_t having Wiener measure as its distribution over C[0, 1] (see Billingsley [15], pages 86-90).

2.2 General Results in Almost Sure Limit Theory

In Section 2.1, we introduced the almost sure central limit theorem for partial sums and its almost sure functional version. In this section we will continue presenting other almost sure limit theorems that have applications to statistics. Before doing this, we provide some intuitive aspects that motivate the logarithmic summation in the almost sure weak convergence.

We start from the central limit theorem (CLT) for partial sums

$$\lim_{n\to\infty} P\left(\frac{S_n}{\sqrt{n}} < x\right) = \Phi(x), \text{ for every } x.$$

Following this, we may ask if it is possible to obtain results that hold for almost every realization of the random variables in S_n (see Schatte [91]). This becomes important because, as Brosamler [19] observed, "one would only have to check one (typical) sequence, rather than many sequences as in tests based on the classical

CLT". Rephrasing it, does

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\left(\frac{S_i}{\sqrt{i}} < x\right)$$

exist with probability 1 for every x? A simple application of the strong law of large numbers will not answer our question since the random variables $\mathbb{I}\left(\frac{S_i}{\sqrt{i}} < x\right)$ are not independent. The answer is negative as Schatte [91] showed

$$P\left(\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \left| \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}\left(\frac{S_k}{\sqrt{k}} < x\right) - \Phi(x) \right| = 0\right) = 0$$

but a weaker statement under logarithmic summation is true

$$P\left(\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \left| \frac{1}{\log n} \sum_{k=1}^{n} \mathbb{I}\left(\frac{S_k}{\sqrt{k}} < x\right) - \Phi(x) \right| = 0\right) = 1.$$

The difference between the ordinary (Cesàro) summation and the logarithmic average is illustrated in the case of the random walk for a sequence of i.i.d. random variables with continuous symmetric distribution (Berkes [10], Hörmann [58], Peligrad and Révész [82] discuss this example). Even though

$$P(S_n > 0) = \frac{1}{2}, \text{ for all } n,$$

it is not true that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}(S_k > 0) = \frac{1}{2} \text{ a.s.}$$

Actually one can prove that

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}(S_k > 0) = 0$$

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}(S_k > 0) = 1$$

and Erdös and Hunt [45] showed that

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}(S_k > 0) = \frac{1}{2} \text{ a.s.}$$

While the first version of the almost sure central limit theorem was given in terms of logarithmic average, which comes naturally from the ergodic theory perspective (see Hörmann [58]), this is not the only way to obtain almost sure limit theorems. As we will see in the next sections possible generalizations of the almost sure limit theorems include summation with different weights or considering other statistics. For example Peligrad and Révész [82] obtained

$$\lim_{n\to\infty} \frac{1}{D_n} \sum_{k=1}^{n} \frac{1}{d_k} \mathbb{I}\left(\frac{S_k(\omega)}{\sqrt{k}} < x\right) = \Phi(x) \text{ a.s. for all x} \in \mathbb{R} \tag{2.2.1}$$

with $D_n = \sum_{k=1}^{n} d_k$ and $d_k = \frac{(\log k)^\alpha}{k}$, for $\alpha > -1$ and Berkes and Csáki [11] considered 2.2.1 with the weights $d_k = \frac{\exp((\log k))^\alpha}{k}$, for $0 \leq \alpha < 1/2$.
Results from classical summation theory show that the almost sure limit theorems become "stronger" depending on the weights that they use. One can find details on these properties of the almost sure theory in Berkes and Csáki [11], Hörmann [58], [57] and the references in there.

Almost sure limit theorems were obtained for different statistics than the partial sums, and an overview of these results is contained in Berkes [10]. Berkes and Csáki [11] is the first article that introduces almost sure results for a general form of statistics and unifies ideas from before. They show that under some technical assumptions, every weak limit theorem for independent random variables has an equivalent almost sure version. We continue recalling some of their almost sure general results and in the next section we show how they apply to different statistics.

Theorem 2.2.1. (Berkes and Csáki [11]) Let $X_1, X_2, ...$ be a sequence of independent random variables satisfying the weak limit theorem

$$f_k(X_1, X_2, ..., X_k) \xrightarrow{D} G,$$

where $f_k : \mathbb{R}^k \to \mathbb{R}(k = 1, 2, ...)$ are measurable functions and G is a distribution function. Assume that for each $1 \leq k < l$ there exists a measurable function $f_{k,l} : \mathbb{R}^{l-k} \to \mathbb{R}$ such that

$$E\left(|f_l(X_1, ..., X_l) - f_{k,l}(X_{k+1}, ..., X_l)| \wedge 1\right) \leq A\frac{c_k}{c_l} \tag{2.2.2}$$

whith a constant $A > 0$ and a positive, nondecreasing sequence c_n satisfying $c_n \to \infty$, $\frac{c_{n+1}}{c_n} = O(1)$. Put $d_k = \log(\frac{c_{k+1}}{c_k})$ and $D_n = \sum_{k=1}^n d_k$. Then we have

$$\lim_{n\to\infty} \frac{1}{D_n} \sum_{k=1}^n d_k \mathbb{I}(f_k(X_1, ..., X_k) < x) = G(x) \text{ a.s. for any } x \in C_G,$$

where C_G denotes the set of continuity points of G. The result remains valid if we replace the weight sequence (d_k) by any (d_k^*) such that $0 \leq d_k^* \leq d_k, \sum d_k^* = \infty$.

The next theorem assumes a weaker condition than (2.2.2) and states the equivalence between the almost sure central limit theorem and an weighted weak result. We need to define $\log_+ x = \log x$ if $x \geq 1$ and 0 otherwise.

Theorem 2.2.2. (Berkes and Csáki [11]) Let $X_1, X_2, ...$ be independent random variables, $f_k : \mathbb{R}^k \to \mathbb{R}(k = 1, 2, ...)$measurable functions and assume that for each $1 \leq k < l$ there exists a measurable function $f_{k,l} : \mathbb{R}^{l-k} \to \mathbb{R}$ such that

$$E\left(|f_l(X_1, ..., X_l) - f_{k,l}(X_{k+1}, ..., X_l)| \wedge 1\right) \leq C(\log_+ \log_+(\frac{c_l}{c_k}))^{-(1+\varepsilon)}$$

for some constants $C > 0, \varepsilon > 0$, and a positive, nondecreasing sequence c_n satisfying $c_n \to \infty$, $\frac{c_{n+1}}{c_n} = O(1)$. Put $d_k = \log(\frac{c_{k+1}}{c_k})$ and $D_n = \sum_{k=1}^n d_k$. Then for any distribution function G the expressions

$$\lim_{n\to\infty} \frac{1}{D_n} \sum_{k=1}^n d_k \mathbb{I}(f_k(X_1, ..., X_k) < x) = G(x) \text{ a.s. for any } x \in C_G$$

and

$$\lim_{n\to\infty} \frac{1}{D_n} \sum_{k=1}^n d_k P(f_k(X_1, ..., X_k) < x) = G(x) \text{ for any } x \in C_G$$

are equivalent. The result remains valid if we replace the weight sequence d_k by any d_k^* such that $0 \le d_k^* \le d_k, \sum d_k^* = \infty$.

This type of almost sure limit theorems can be extended to stochastic processes and in this case the functionals f_k depend on a process $\{X(t), t \ge 0\}$ with independent increments.

Theorem 2.2.3. (Berkes and Csáki [11]) Let $\{X(t), t \ge 0\}$ be a process with $X(0) = 0$ and independent increments and let $\xi_1, \xi_2, \ldots$ be random variables such that ξ_k is measurable with respect to $\sigma\{X(t), 0 \le t \le k\}$. Assume that for each $1 \le k < l$ there exists a random variable $\xi_{k,l}$ measurable with respect to $\sigma\{X(t') - X(t) : k \le t \le t' \le l\}$ such that

$$E(|\xi_l - \xi_{k,l}| \wedge 1) \le C(\log_+ \log_+(c_l/c_k))^{-(1+\varepsilon)}$$

for some constants $C > 0, \varepsilon > 0$ and a positive, nondecreasing sequence (c_n) with $c_n \to \infty, c_{n+1}/c_n = O(1)$. Put $d_k = \log(c_{k+1}/c_k), D_n = \sum_{k=1}^n d_k$. Then for any distribution function G the expressions

$$\lim_{n\to\infty} \frac{1}{D_n} \sum_{k=1}^n d_k \mathbb{I}(\xi_k < x) = G(x) \text{ a.s. for any } x \in C_G$$

and

$$\lim_{n\to\infty} \frac{1}{D_n} \sum_{k=1}^n d_k P(\xi_k < x) = G(x) \text{ for any } x \in C_G$$

are equivalent.

The next result obtained by Lifshits [73], [74] is a multivariate almost sure weak convergence for the partial sum of a sequence of random vectors that we will use in Chapter 3 and Chapter 4. We follow the notation introduced in (2.1.1)-(2.1.3).

Theorem 2.2.4. (Lifshits [73], [74]) Let $\mathcal{X}$ be a separable Banach space and X_j be a sequence of independent $\mathcal{X}$-valued random variables. Consider

$$\zeta_k = \frac{1}{B_k} \sum_{j=1}^k X_j - A_k,$$

where $A_k \in \mathcal{X}$ and the sequence $B_k > 0$ is increasing and $B_k \to \infty$. Assume that

$$\zeta_k \xrightarrow{D} G, k \to \infty,$$

and for some $\epsilon > 0$

$$\sup_k E(\log_+ \log_+ ||\zeta_k||)^{1+\epsilon} < \infty. \tag{2.2.3}$$

Then the almost sure limit theorem with limit law G holds for every bounded weight sequence b_k which satisfies $b_k \le C \log(\frac{B_k}{B_{k-1}}), k \ge 2$ and $C > 0$.

2.3 Examples of Almost Sure Limit Theorems

We introduce the almost sure limit theorems for statistics other than partial sums. The results that we present here are taken from Berkes and Csáki [11] and they are direct applications of the theorems of Berkes and Csáki [11].

Example 2.3.1. *(Berkes and Csáki [11]) Empirical distribution functions*

Let $X_1, X_2, ...$ be i.i.d. random variables with continuous distribution function F. Consider the empirical distribution function $F_n(x) = \frac{1}{n}\sum_{k=1}^n \mathbb{I}(X_k < x)$ and the Kolmogorov-Smirnov statistic $\beta_n := \sup_x |F_n(x) - F(x)|$. It is known that

$$\sqrt{n}\beta_n = \sqrt{n}\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \xrightarrow{D} \sup_{0\le x\le 1} |B^{(0)}(x)| =: \beta,$$

where $B^{(0)}(t) = B(t) - tB(1)$, $0 \le t \le 1$ is the Brownian bridge and $B(t), t \ge 0$ is a standard Brownian motion and (Kolmogorov [65])

$$P(\beta \ge d) = 2\sum_{k=1}^{\infty}(-1)^{k-1}e^{-2k^2d^2}$$

Thus, we have

$$P(\sqrt{n}\beta_n \le d) \to \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2d^2} = 1 + 2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2d^2} \tag{2.3.1}$$

Since our goal is to obtain an almost sure limit theorem for the Kolmogorov-Smirnov statistic, Theorem 2.2.1 can be used. Take $c_k = \sqrt{k}$ and

$$f_l(x_1, ..., x_l) = \frac{1}{\sqrt{l}} \sup_x \left| \sum_{i=1}^{l} (\mathbb{I}(x_i < x) - F(x)) \right|,$$

$$f_{k,l}(x_{k+1}, ..., x_l) = \frac{1}{\sqrt{l}} \sup_x \left| \sum_{i=k+1}^{l} (\mathbb{I}(x_i < x) - F(x)) \right|.$$

Then we have

$$E\left(|f_l(X_1, ..., X_l) - f_{k,l}(X_{k+1}, ..., X_l)|\right) \leq E\left|\frac{T_k}{\sqrt{l}}\right| \leq A\left(\frac{k}{l}\right)^{1/2}, \tag{2.3.2}$$

where $A > 0$ is a constant and

$$T_k = \sup_x \left| \sum_{i=1}^{k} (\mathbb{I}(X_i < x) - F(x)) \right|.$$

The inequality in (2.3.2) holds since $E\left|\frac{T_k}{\sqrt{k}}\right|$ is bounded. This can be shown using the inequality (Dvoretzky et al. [43], Lemma 2)

$$P\left(\frac{T_k}{\sqrt{k}} \geq d\right) \leq Ce^{-2d^2}, \text{ where } C \text{ is a positive constant .}$$

For $c_k = \sqrt{k}$ it follows that $d_k := \log(\frac{c_{k+1}}{c_k}) \sim c\frac{1}{k}$ and $D_n \sim c\frac{1}{\log n}$, where $c > 0$ is a constant (in general $a_n \sim b_n$ means that $a_n/b_n \to 1$ as $n \to \infty$). The assumptions in Theorem 2.2.1 are met as equations (2.3.1) and (2.3.2) show, and we can obtain the almost sure limit theorem for the empirical distribution function:

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}(\sqrt{k}\beta_k < x) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2x^2} \text{ a.s. for any } x.$$

Example 2.3.2. *(Berkes and Csáki [11]) U-statistics*

Let $X_1, X_2, ...$ be an i.i.d. sequence of random variables, $m \geq 1$ an integer and $h(x_1, ..., x_m)$ a symmetric measurable function such that $Eh^2(X_1, ..., X_m) < \infty$.

For $n \geq m$, a U-statistic is defined as follows

$$U_n = \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < ... < i_m \leq n} h(X_{i_1}, ..., X_{i_m}).$$

The function h is called the kernel of the U-statistic and the integer m is called the degree of the kernel h. Set $\theta = Eh(X_1, ..., X_m)$ and for $1 \leq j \leq m$, let

$$h_j(x_1, ..., x_j) = Eh(x_1, ..., x_j, X_{j+1}, ..., X_m), \quad \xi_j = \text{Var } h_j(X_1, ..., X_j).$$

It is known that (see Serfling [92], page 182)

$$0 = \xi_0 \leq \xi_1 \leq ... \leq \xi_m = \text{Var } h(X_1, ..., X_m) < \infty.$$

It is also known that (see Denker [36], Koroljuk and Borovskich [66], Lee [70])

$$n^{c/2}(U_n - \theta) \xrightarrow{D} F,$$

where F is a nondegenerate limit distribution and $c \geq 1$ is the critical parameter defined as the smallest integer such that $\xi_c > 0$. For any $1 \leq k < l$, define

$$f_l(x_1, ..., x_l) = \frac{l^{c/2}}{\binom{l}{m}} \sum_{1 \leq i_1 < ... < i_m \leq l} \tilde{h}(x_{i_1}, ..., x_{i_m}),$$

$$f_{k,l}(x_{k+1}, ..., x_l) = \frac{l^{c/2}}{\binom{l}{m}} \sum_{k+1 \leq i_1 < ... < i_m \leq l} \tilde{h}(x_{i_1}, ..., x_{i_m}),$$

where $\tilde{h}(X_1, ..., X_m) = h(X_1, ..., X_m) - \theta$. It can be shown that for $k \leq l$ and some positive constant C,

$$E(|f_l(X_1, ..., X_l) - f_{k,l}(X_{k+1}, ..., X_l)|)^2 \leq C(k/l)$$

The assumptions in Theorem 2.2.1 are met and we obtain the almost sure limit theorem for U-statistics:

$$\lim_{n \to \infty} \frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}(k^{c/2}(U_k - \theta) < x) = F(x) \text{ a.s. for any } x \in C_F. \tag{2.3.3}$$

Note that Holzmann, Koch, Min [55] also deals with the almost sure limit theorem for U-statistics. The authors of [55] relaxed the condition $Eh^2(X_1, ..., X_m) < \infty$ by replacing it with $E|h_k(X_1, ..., X_k)|^{2k/(2k-c)} < \infty$, for $k = c, ..., m$.

Example 2.3.3. *(Berkes and Csáki [11]) Maxima of partial sums*

Let $X_1, X_2, ...$ be independent random variables with partial sums $S_n = \sum_{k=1}^n X_k$ and $S_n^* = \max_{k \leq n} S_k$. Assume that

$$E\left(\log_+ \log_+ \left|\frac{S_n}{b_n}\right|\right)^{1+\delta} \leq K_1 \text{ and } E\left(\log_+ \log_+ \left|\frac{S_n^*}{b_n}\right|\right)^{1+\delta} \leq K_2, n = 1, 2, ...$$

where (b_n) is a positive numerical sequence, $K_1, K_2 > 0, \delta > 0$. Assume also that $b_n \uparrow \infty, b_{n+1}/b_n = O(1)$ and

$$\frac{S_n^*}{b_n} \xrightarrow{D} G, \text{ for some distribution function } G.$$

To obtain an almost sure result, we use Theorem 2.2.2 with $c_k = b_k$ and

$$f_l(x_1, ..., x_l) = \frac{1}{b_l} \max_{i \leq l}(x_1 + ... + x_i)$$

$$f_{k,l}(x_{k+1}, ..., x_l) = \begin{cases} \frac{1}{b_l} \max_{k+1 \leq i \leq l}(x_{k+1} + ... + x_i), & \text{if } i_0 > k \\ 0, & \text{if } i_0 \leq k \end{cases}$$

where $1 \leq i_0 \leq l$ is the smallest integer such that $s_i = x_1 + ... + x_i, 1 \leq i \leq l$ attains its maximum. It is proved that

$$E\left(|f_l(X_1, ..., X_l) - f_{k,l}(X_{k+1}, ..., X_l)| \wedge 1\right) \leq E\left(\left|\frac{S_k}{b_l}\right| \wedge 1\right) + E\left(\left|\frac{S_k^*}{b_l}\right| \wedge 1\right)$$

$$\leq \text{const.} \left(\log_+ \log_+ \frac{b_l}{b_k}\right)^{-(1+\delta)}$$

Therefore, Theorem 2.2.2 implies

$$\lim_{n\to\infty} \frac{1}{D_n} \sum_{k=1}^n d_k \mathbb{I}(\frac{S_k^*}{b_k} < x) = G(x) \text{ a.s. for any } x \in C_G$$

where d_k and D_k are defined as in Theorem 2.2.2.

Remark 2.3.4. There is no derivation of an almost sure central limit theorem for linear rank statistics in the literature, except the case when the rank statistics have a representation as U-statistics (see [70] and [66] for examples). We fill this gap in this dissertation, reducing the problem to Lifshits Theorem 2.2.4 although this is not straightforward as Examples 2.3.1-2.3.3.

Chapter 3

Statistical Applications of Almost Sure Limit Theorems

Almost sure limit theorems offer the chance to evaluate statistical methods in different ways. One way that we pursue here is the quantile estimation. Another idea would be to evaluate the quality of an estimator sequentially based on almost sure convergence and stabilization. We do not pursue this second method in this dissertation. Based on the logarithmic quantile estimation, we develop the general framework for statistical decision theory. We define the empirical logarithmic quantile as the inverse of the empirical distribution function of logarithmic type and show that it converges almost surely to the true quantile. We also propose a test that converges to the significance level under the null hypothesis and its power convergences to 1 under the alternative. Applications and numerical results for hypothesis testing and confidence intervals are taken from Thangavelu [97] and from Fridline [47].

3.1 Statistical Decision Theory based on Logarithmic Quantiles

This section serves as an introduction to some decision theoretic topics connected with the almost sure central limit theorem and its quantile estimation. Since we are only interested in hypothesis testing, we restrict to this issue and discuss test procedures with asymptotically correct significance levels.

Let $(T_n)_{n\geq 1}$ be a sequence of real valued statistics defined on the same measurable space $(\Omega, \mathcal{B})$ and let $\mathcal{P}$ be a family of probability measures on $\mathcal{B}$. Let $H_0, H_1 \subset \mathcal{P}$ denote the hypothesis and alternative respectively. We make the following assumptions:

1. Under $P \in H_0$, the law of T_n converges to some distribution G_P as $n \to \infty$.

2. Under $P \in H_0$, T_n satisfies an almost sure weak convergence theorem towards the same distribution G_P as in 1. That is almost surely

$$\lim_{N\to\infty} C_N^{-1} \sum_{n=1}^{N} \frac{1}{n} \mathbb{I}(T_n \leq t) = G_P(t) \quad \forall t \in C_P \tag{3.1.1}$$

   where C_P denotes the set of continuity points of G_P and $\mathbb{I}$ denotes the indicator function.

3. Under $P \in H_1$, the statistics T_n diverge or accumulate only at values not in the support of any G_P with $P \in H_0$.

Example 3.1.1. Let $\mathcal{P}$ be a class of distributions with finite second moment. Assume that H_0 is the subclass of distributions with zero mean (it is assumed that this subclass is non-empty). Let $T_n = n^{-1/2}(X_1 + ... + X_n)$ denote the partial sum of i.i.d. random variables normalized by $\sqrt{n}$. By the theorem of Lacey and Philipp [69] it follows that under H_0

$$\lim_{N\to\infty} C_N^{-1} \sum_{n=1}^{N} \frac{1}{n} \mathbb{I}(T_n \leq t) = \mathcal{N}(0, \sigma^2)$$

almost surely, where $\sigma^2 \geq 0$ is some constant, depending on $P \in H_0$. Clearly, the distribution of T_n converges weakly to the same normal distribution. However, if $P \in H_1$ then by the strong law of large numbers $\frac{1}{\sqrt{n}} T_n$ converges almost surely to the expectation $\mu(P) = E_P(X_1)$ which is non-zero. Hence T_n diverges almost surely.

Example 3.1.2. Let $\mathcal{P}$ be a class of multivariate distributions and H_0 be the subclass of those distributions with zero mean. Let $\mathbf{X}_i$ be an i.i.d. sequence of random vectors in $\mathbb{R}^d$ with covariance matrix Σ. Let $\mathbf{S}_n = \frac{1}{\sqrt{n}}(\mathbf{X}_1 + ... + \mathbf{X}_n)$ be the partial sum of the random vectors $\mathbf{X}_i$ normalized by $\sqrt{n}$. Under the null hypothesis, the distribution of $\mathbf{S}_n$ converges to the multivariate normal distribution $\mathcal{N}_d(0, \Sigma)$. By Theorem 2.2 of Ibragimov and Lifshits [62] it follows that under H_0,

$$\lim_{N\to\infty} C_N^{-1} \sum_{n=1}^{N} \frac{1}{n} \mathbb{I}(\mathbf{S}_n \leq \mathbf{t}) = G_{\mathbf{X}}(\mathbf{t}) = G_P(t), \tag{3.1.2}$$

where $G_{\mathbf{X}}$ is the distribution function of $\mathbf{X} \sim \mathcal{N}_d(0, \Sigma)$.
Let the quadratic form $\mathbf{T}_n := \mathbf{S}_n^T A \mathbf{S}_n$, where A is a $d \times d$ matrix such that $A^T = A$. Under H_0 and if $(A\Sigma)^2 = (A\Sigma)^3$ and $\mathrm{tr}(A\Sigma) = r$ (see Mathai and Provost [77], Corollary 5.1.2a), for $r \leq d$ it follows that

$$\mathbf{T}_n \xrightarrow{D} \chi_r^2.$$

The almost sure central limit theorem for the quadratic forms $\mathbf{T}_n$ is easily established using Lifshits [74]. Consider the continuous function $f : \mathbb{R}^d \to \mathbb{R}, f(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$, $\mathbf{x} \in \mathbb{R}^d$. We want to show that under H_0,

$$\lim_{N\to\infty} C_N^{-1} \sum_{n=1}^{N} \frac{1}{n} \mathbb{I}(f(\mathbf{S}_n) \leq \mathbf{t}) = G_{f(\mathbf{X})}(\mathbf{t}). \tag{3.1.3}$$

It is well known that the relation (3.1.2) is equivalent to (see Lifshits [74])

$$\lim_{N\to\infty} C_N^{-1} \sum_{n=1}^{N} \frac{1}{n} g(\mathbf{S}_n) = \int_{\mathbb{R}^d} g(u) dG_{\mathbf{X}}(u), \tag{3.1.4}$$

for every bounded continuous function $g : \mathbb{R}^d \to \mathbb{R}$. To obtain (3.1.3) we need to show that

$$\lim_{N\to\infty} C_N^{-1} \sum_{n=1}^{N} \frac{1}{n} h(f(\mathbf{S}_n)) = \int_{\mathbb{R}^d} h(u) dG_{\mathbf{f(X)}}(u),$$

for every bounded continuous function h. This holds since we can consider $g = h \circ f$ bounded continuous function in equation (3.1.4) and since $\int_{\mathbb{R}^d} h \circ f(u) dG_X(u) =$

$\int_{\mathbb{R}^d} h(u)dG_{f(X)}(u)$. Hence, we obtained that

$$\lim_{N\to\infty} C_N^{-1} \sum_{n=1}^{N} \frac{1}{n}\mathbb{I}(\mathbf{S}_n^T A \mathbf{S}_n \leq \mathbf{t}) = G_{\mathbf{X}^T A \mathbf{X}}(\mathbf{t}) = G_P(t), \tag{3.1.5}$$

where $G_{\mathbf{X}^T A \mathbf{X}}$ is the distribution function of a chi-squared random variable with r degrees of freedom.

Now, under the alternative hypothesis, by the multivariate strong law of large numbers and a continuous mapping theorem it follows that $\frac{1}{n}\mathbf{T}_n$ converges a.s. and so $\mathbf{T}_n$ diverges almost surely.

Definition 3.1.3. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables defined on the same measurable space. Let $T_n = T_n(X_1, ..., X_n)$ be a sequence of statistics where T_n is a function of $X_1, ..., X_n$. Then the empirical distribution function of logarithmic type is defined by (see (3.1.1))

$$\widehat{G}_N(t) = \frac{1}{C_N} \sum_{n=1}^{N} \frac{1}{n}\mathbb{I}(T_n \leq t), \text{ for } t \in \mathbb{R}.$$

It is clear (see Thangavelu [97]) that $\widehat{G}_N$ is an empirical distribution function and $\widehat{G}_N$ converges uniformly a.s. to G_P by assumption 1 if $P \in H_0$. We define the inverse of the empirical distribution function of logarithmic type as

$$\widehat{G}_N^{-1}(\alpha) = \begin{cases} \sup\{t|\widehat{G}_N(t) = 0\}, \text{ for } \alpha = 0 \\ \sup\{t|\widehat{G}_N(t) < \alpha\}, \text{ for } \alpha \in (0,1) \\ \inf\{t|\widehat{G}_N(t) = 1\}, \text{ for } \alpha = 1. \end{cases} \tag{3.1.6}$$

Definition 3.1.4. For $\alpha \in [0, 1]$ and $N \in \mathbb{N}$, the logarithmic (α, N)-quantile of the sequence of statistics $(T_n)_{n\geq 1}$ is defined by

$$\widehat{t}_\alpha^{(N)} = \widehat{G}_N^{-1}(\alpha). \tag{3.1.7}$$

It can be shown (Thangavelu [97]) that if G_P is a continuous distribution function, then under $P \in H_0$

$$\lim_{N\to\infty} \widehat{t}_\alpha^{(N)} = t_\alpha(P) \quad P - a.s., \tag{3.1.8}$$

where $t_\alpha(P) = G_P^{-1}(\alpha)$ is the true quantile.

The next results state general decision rules for testing hypotheses.

Proposition 3.1.5. For $P \in H_0$ with a continuous limit distribution G_P,

$$I_\alpha^{(N)} = \left[T_N - \widehat{t}_{1-\alpha}^{(N)}, T_N - \widehat{t}_\alpha^{(N)}\right]$$

is a random interval with the property that

$$P(0 \in I_\alpha^{(N)}) \to 1 - 2\alpha, \text{ when } N \to \infty.$$

Proof.

$$P(0 \in I_\alpha^{(N)}) = P(\widehat{t}_\alpha^{(N)} \leq T_N \leq \widehat{t}_{1-\alpha}^{(N)}).$$

Since a.s. the quantiles converge to the true quantiles, for given $\eta > 0$, there is N_0 such that for $N \geq N_0$, $|\widehat{t}_\alpha^{(N)} - t_\alpha(P)| < \eta$ and similarly for the $1-\alpha$ quantiles. Therefore,

$$\limsup_{N\to\infty} P(0 \in I_\alpha^{(N)}) \leq \limsup_{N\to\infty} P(t_\alpha(P) - \eta \leq T_N \leq t_{1-\alpha}(P) + \eta) =$$

$$= G_P(t_{1-\alpha}(P) + \eta) - G_P(t_\alpha(P) - \eta).$$

Letting $\eta \to 0$ shows

$$\limsup_{N\to\infty} P(0 \in I_\alpha^{(N)}) \leq 1 - 2\alpha.$$

Similarly,

$$\liminf_{N\to\infty} P(0 \in I_\alpha^{(N)}) \geq \liminf_{N\to\infty} P(t_\alpha(P) + \eta \leq T_N \leq t_{1-\alpha}(P) - \eta) =$$

$$= G_P(t_{1-\alpha}(P) - \eta) - G_P(t_\alpha(P) + \eta).$$

Letting $\eta \to 0$ shows

$$\liminf_{N\to\infty} P(0 \in I_\alpha^{(N)}) \geq 1 - 2\alpha.$$

□

Corollary 3.1.6. Consider testing the hypothesis H_0 vs. H_1. The test φ_N with $\varphi_N = 1$ iff $T_N \notin [\widehat{t}_\alpha^{(N)}, \widehat{t}_{1-\alpha}^{(N)}]$ satisfy

$$\lim_{N\to\infty} P(\varphi_N = 1) = 2\alpha, \quad P \in H_0.$$

There is another criteria to distinguish between H_0 and H_1 whose idea goes back to Thangavelu [97].

Lemma 3.1.7. (1). Under $P \in H_0$,

$$P(0 \in [\widehat{t}_\alpha^{(N)}, \widehat{t}_{1-\alpha}^{(N)}]) \to \begin{cases} 1, & \text{if } G_P(0) \in (\alpha, 1-\alpha) \\ 0, & \text{else.} \end{cases}$$

(2). Under $P \in H_1$ and if 0 is in the support of some $G_{P'}$ for $P' \in H_0$,

$$P(0 \in [\widehat{t}_\alpha^{(N)}, \widehat{t}_{1-\alpha}^{(N)}]) \to 0, \text{ when } N \to \infty.$$

Proof. (1). Suppose that $0 < \widehat{t}_\alpha^{(N)} = \sup\{t | \widehat{G}_N(t) < \alpha\}$. It implies that $\widehat{G}_N(0) < \alpha$ and $G_P(0) = \lim_{N\to\infty} \widehat{G}_N(0) \leq \alpha$, contradiction with $\alpha < G_P(0)$. Now suppose that $\widehat{t}_{1-\alpha}^{(N)} < 0$. It follows that $\widehat{G}_N(0) \geq 1-\alpha$ and $G_P(0) = \lim_{N\to\infty} \widehat{G}_N(0) \geq 1-\alpha$, contradiction with $G_P(0) < 1-\alpha$.
(2). $P(0 \in [\widehat{t}_\alpha^{(N)}, \widehat{t}_{1-\alpha}^{(N)}]) = P(\alpha \leq \widehat{G}_N(0) \leq 1-\alpha) \to 0$, when $N \to \infty$ by assumption 3 in the beginning of Section 2. □

Below we need a one-sided version of the foregoing statements.

Lemma 3.1.8. (1). Under $P \in H_0^a$, for $a \in \mathbb{R}$

$$P(a \in [\widehat{t}_{1-\alpha}^{(N)}, \infty]) \to \begin{cases} 1, \text{ if } G_P(a) > 1-\alpha \\ 0, \text{ otherwise.} \end{cases}$$

(2). Under $P \in H_1^a$ and if a is in the support of some $G_{P'}$ for $P' \in H_0$,

$$P(a \in [\widehat{t}_{1-\alpha}^{(N)}, \infty]) \to 0, \text{ when } N \to \infty.$$

Theorem 3.1.9. Let T_N be a sequence of statistics as in the beginning of Section 2. Then, for $\alpha > 0$, the test ϕ_N given by

$$\phi_N(T_N) = \begin{cases} 1, \text{ if } T_N > \widehat{t}_{1-\alpha}^{(N)} \\ 0, \text{ otherwise} \end{cases}$$

has the property that under $P \in H_0$,

$$\lim_{N\to\infty} P(\phi_N(T_N) = 1) = \alpha$$

and under $P \in H_1$, for infinitely many N we have a.s.

$$\phi_N(T_N) = 1.$$

Proof. Let $P \in H_0$. Then

$$P(T_N > \widehat{t}_{1-\alpha}^{(N)}) = P(T_N > \widehat{t}_{1-\alpha}^{(N)}; |\widehat{t}_{1-\alpha}^{(N)} - t_{1-\alpha}(P)| \le \eta) +$$

$$+P(T_N > \widehat{t}_{1-\alpha}^{(N)}; |\widehat{t}_{1-\alpha}^{(N)} - t_{1-\alpha}(P)| > \eta) \le P(T_N > t_{1-\alpha} - \eta) +$$

$$+P(|\widehat{t}_{1-\alpha}^{(N)} - t_{1-\alpha}(P)| > \eta) \to 1 - G(t_{1-\alpha}(P) + \eta), \text{ when } N \to \infty$$

and for $\eta \to 0$

$$\limsup_{N\to\infty} P(T_N > \widehat{t}_{1-\alpha}^{(N)}) \le \alpha.$$

Replace η by $-\eta$ and get the claim for $P \in H_0$.
Let $P \in H_1$. The claim follows from the fact that $T_N \to \infty$, since then $T_N \ge T_n$

for all $n \le N$ for infinitely many N. But then $\widehat{t}_{1-\alpha}^{(N)} < T_N$, proving the claim. $\square$

3.2 Established Results

In his dissertation, Thangavelu [97] introduced the idea of making almost sure decisions and using extensive simulation studies he applied these methods to the parametric one-sample test for mean assuming the normal distribution of the sample, the parametric two-sample test for the equality of the means with and without equal variances assuming the normal distribution of the samples, and nonparametric two-sample test with and without equal variances. Fridline [47] calculated confidence intervals for the correlation coefficient using the almost sure version of Cramer's theorem.

3.2.1 Hypothesis Testing based on ASCLT

In the following, we will present some of these results from Thangavelu's dissertation [97]. We start with Thangavelu's solution to the nonparametric Behrens-Fisher problem based on the ASCLT-test methods of rank statistics.

Definition 3.2.1. Let $X_{i1}, ..., X_{in_i}$ where $i = 1, 2$ be i.i.d. random variables such that $X_{i1} \sim F_i$, $i = 1, 2$. Then the nonparametric Behrens-Fisher problem is to test the hypothesis

$$H_0 : p = \frac{1}{2} \text{ vs } H_1 : p \neq \frac{1}{2}, \tag{3.2.1}$$

where $p = P(X_{11} < X_{21}) + \frac{1}{2}P(X_{11} = X_{21})$ is the relative treatment effect.

To state the ASCLT for the two-sample linear rank statistics, let X_1,..., X_m, X_{m+1},..., X_n be i.i.d. random variables such that the first m r.v. are distributed as F and the remaining of $n - m$ r.v. are distributed as G and the weighted average $H_n(x) = \frac{1}{n}(mF(x) + (n - m)G(x))$. The empirical weighted average is defined by

$$\widehat{H}_n(x) = \frac{1}{n}(m\widehat{F}(x) + (n - m)\widehat{G}(x)) = \frac{1}{n}\sum_{k=1}^{n} c(x - X_k),$$

where c is the normalized counting function such that $c(u) = 0, \frac{1}{2}$ or 1 when $u < 0, u = 0$ or $u > 0$. Let $\widehat{F}$, $\widehat{G}$ be the empirical sample distributions

$$\widehat{F}(x) = \frac{1}{m}\sum_{k=1}^{m} c(x - X_k)$$

$$\widehat{G}(x) = \frac{1}{n-m}\sum_{k=m+1}^{n} c(x - X_k).$$

The two-sample linear rank statistics is defined as

$$T_n = \sum_{i=1}^{n} a_i J(\widehat{H}_n(X_i)),$$

where $1 \leq m(n) = m < n$ such that $\frac{m(n)}{n} \to \lambda \in (0,1)$, $J : (0,1) \to \mathbb{R}$ is absolutely continuous score function and $a_i = \begin{cases} 1, & 1 \leq i \leq m \\ 0, & m < i \leq n \end{cases}$.

Theorem 3.2.2. (Thangavelu [97]) Under some assumptions, the two-sample linear rank statistic satisfies the ASCLT

$$\frac{1}{\log n}\sum_{k=1}^{n} \frac{1}{k}\mathbb{I}(k^{-1/2}(T_k - m\int J(H_k(t))dF_{m_k}) \leq t) \to \Phi_\sigma(t)\ P\text{-a.s.}$$

Thangavelu proposed two methods to test for the hypothesis in (3.2.1), and they are based on a small sample approximation.

Algorithm 1: Let $n = \min(n_1, n_2)$. Assume that the two samples are permuted independently "nper" times and let $\mathbf{x}_i^{*p}$ for $i = 1, 2$, $p = 1, 2, .., \text{nper}$ be the pth permuted sample vector. For each permuted sample, calculate

$$SS_k^{*p} = \sqrt{k}(\overline{R}_{1\cdot,k}^{*p} - 0.5) \text{ for } k = 1, 2, ..., n,$$

where $\overline{R}_{1\cdot,k}^{*p} = \frac{\sum_{j=1}^{k} R_{1j,k}^{*p}}{k}$ and $R_{ij,k}^{*p} = \frac{1}{2} + \sum_{l=1}^{2}\sum_{m=1}^{k} c(X_{ij}^{*p} - X_{lm}^{*p})$, $k = 1, 2, ..., n$, $i = 1, 2$, $j = 1, 2, ..., k$ denote the overall mid-rank of the observations (X_{11}^{*p},...,X_{1k}^{*p}, X_{21}^{*p},..., X_{2k}^{*p}).

For each permutation $p = 1, ..., \text{nper}$, calculate $\overline{SS}^{*p} = \frac{\sum_{k=1}^{n} SS_k^{*p}}{n}$. The empirical quantiles can be computed by

$$\widehat{t}_{\alpha}^{*p,(n)} = \max\{t| \frac{1}{C_n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}(SS_k^{*p} < t) \leq \alpha\}$$

$$\widehat{t}_{1-\alpha}^{*p,(n)} = \max\{t| \frac{1}{C_n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}(SS_k^{*p} < t) \leq 1 - \alpha\}$$

Let $\bar{t}_{\alpha} = \frac{\sum_{p=1}^{\text{nper}} \widehat{t}_{\alpha}^{*p,(n)}}{\text{nper}}$ and $\bar{t}_{1-\alpha} = \frac{\sum_{p=1}^{\text{nper}} \widehat{t}_{1-\alpha}^{*p,(n)}}{\text{nper}}$. Thangavelu proposed to reject H_0 in (3.2.1) if $0 \notin [\bar{t}_{\alpha}, \bar{t}_{1-\alpha}]$.

Algorithm 2: Define the transformed quantiles

$$\widehat{t}_{\alpha,trans.}^{*p,(n)} = \frac{\overline{SS}^{*p} - \widehat{t}_{1-\alpha}^{*p,(n)}}{\sqrt{n}}$$

$$\widehat{t}_{1-\alpha,trans.}^{*p,(n)} = \frac{\overline{SS}^{*p} - \widehat{t}_{\alpha}^{*p,(n)}}{\sqrt{n}}$$

and

$$\bar{t}_{\alpha,trans.} = \frac{\sum_{p=1}^{nper} \widehat{t}_{\alpha,trans.}^{*p,(n)}}{nper},$$

$$\bar{t}_{1-\alpha,trans.} = \frac{\sum_{p=1}^{nper} \widehat{t}_{1-\alpha,trans.}^{*p,(n)}}{nper}.$$

For some constant coefficients $k_{n_1,n_2,2\alpha}$ and $\lambda_{n_1,n_2,2\alpha}$, compute the formulas

$$\bar{t}_{\alpha,fin.} = \frac{\bar{t}_{\alpha,trans.} - 2\alpha\lambda_{n_1,n_2,2\alpha}(\bar{t}_{\alpha,trans.} + \bar{t}_{1-\alpha,trans.})}{k_{n_1,n_2,2\alpha}}$$

$$\bar{t}_{1-\alpha,fin.} = \frac{\bar{t}_{1-\alpha,trans.} - 2\alpha\lambda_{n_1,n_2,2\alpha}(\bar{t}_{\alpha,trans.} + \bar{t}_{1-\alpha,trans.})}{k_{n_1,n_2,2\alpha}}$$

The hypothesis in (3.2.1) is rejected if $\overline{R}_{1\cdot} - 0.5 \notin [\bar{t}_{\alpha,fin.}, \bar{t}_{1-\alpha,fin.}]$, where $\overline{R}_{1\cdot} = \frac{\sum_{j=1}^{n_1} R_{1j}}{n_1}$ and R_{ij} is the overall midrank of X_{ij}.

The next application from Thangavelu's thesis [97] is the modified ASCLT-based test of $H_0 : \mu = 0$ vs $H_a : \mu \neq 0$ for a small normally distributed $N(\mu, \sigma^2)$ sample size.

Let $\mathbf{x}^i$ be the i^{th} permuted sample vector of $\mathbf{x}$ for $i = 1, ...,$nper, where nper is the total number of permutations of the sample we are interested in. For each permuted sample, the weighted partial mean is defined as

$$SS_n^i = n\frac{\overline{x}_n^i}{\sqrt{n}}, n = 1, ..., N,$$

where $\overline{x}_n^i$ is the mean of the partial i^{th} permuted sample $(x_1^i, ..., x_n^i), n = 1, ..., N$. The average of the weighted partial means

$$\overline{SS}_n^i = \frac{1}{N}\sum_{n=1}^{N} SS_n^i$$

is used to estimate the quantiles $\widehat{q}_\alpha^{i,(N)}$ and $\widehat{q}_{1-\alpha}^{i,(N)}$:

$$\widehat{q}_\alpha^{i,(N)} = \max\left\{q | \frac{1}{C_N}\sum_{n=1}^{N}\frac{1}{n}\mathbb{I}(SS_n^i < q) \leq \alpha\right\},$$

$$\widehat{q}_{1-\alpha}^{i,(N)} = \max\left\{q | \frac{1}{C_N}\sum_{n=1}^{N}\frac{1}{n}\mathbb{I}(SS_n^i < q) \leq 1-\alpha\right\}$$

Calculate

$$\overline{q}_\alpha = \frac{1}{\text{nper}}\sum_{i=1}^{\text{nper}}\widehat{q}_\alpha^{i,(N)},$$

$$\overline{q}_{1-\alpha} = \frac{1}{\text{nper}}\sum_{i=1}^{\text{nper}}\widehat{q}_{1-\alpha}^{i,(N)}$$

and reject H_0 if

$$0 \notin [\overline{q}_\alpha, \overline{q}_{1-\alpha}].$$

The type II error curves of the t-test and ASCLT-method 1 for $N = 40$ at significance levels $2\alpha = 5\%$ and $2\alpha = 10\%$, based on samples generated from a normal distribution with variance 1 are presented in Figures 3.1-3.2. Thangavelu [97] considered the total number of simulation runs for each result to be 10000 and for the ASCLT-method 1 he took $nper = 2000$. The t-test is the optimal test for testing the hypothesis H_0. In Figures 3.1-3.4 the ASCLT method 1 shows good power properties when compared to the t-test.

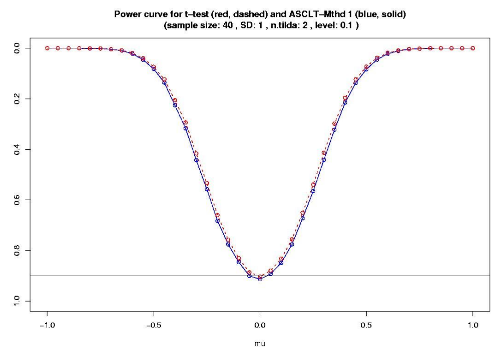

Figure 3.1. Type II error curves for t-test and ASCLT-method 1; sample size 40, level 0.1 (courtesy of K. Thangavelu)

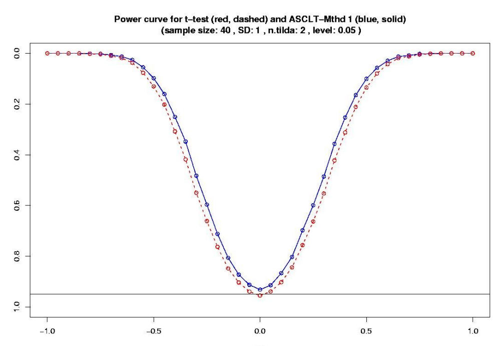

Figure 3.2. Type II error curves for t-test and ASCLT-method 1; sample size 40, level 0.05 (courtesy of K. Thangavelu)

The type II error curves of the t-test and ASCLT-method 1 for $N = 50$ at significance levels $2\alpha = 5\%$ and $2\alpha = 10\%$, based on samples generated from a normal distribution with variance 1 are presented in Figures 3.3-3.4.

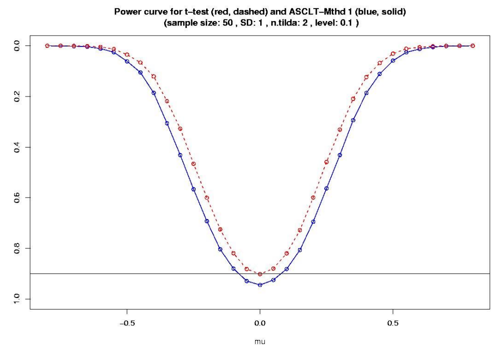

Figure 3.3. Type II error curves for t-test and ASCLT-method 1; sample size 50, level 0.1 (courtesy of K. Thangavelu)

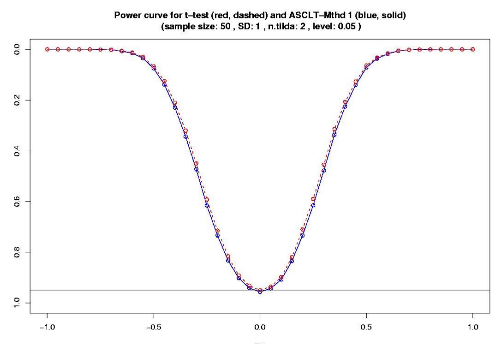

Figure 3.4. Type II error curves for t-test and ASCLT-method 1; sample size 50, level 0.05 (courtesy of K. Thangavelu)

3.2.2 Almost Sure Confidence Intervals for the Correlation Coefficient

Here we introduce the almost sure version of Cramer's theorem as considered in Denker and Fridline [37]. It is known that Cramer's theorem (also known as the delta method, see Lehmann and Romano [71], page 436) is used in statistics to reduce the variation and to construct narrower confidence intervals. Using the almost sure version of Cramer's theorem, Fridline [47] developed confidence intervals for the population correlation coefficient.

Theorem 3.2.3. (Denker and Fridline [37]) Let $g : \mathbb{R}^d \to \mathbb{R}^k$ be a differentiable function in a neighborhood of some $\mu \in \mathbb{R}$ and its derivative $\dot{g}$ be continuous at μ. Let $X_n, n \geq 1$ be a sequence of $\mathbb{R}^d$-valued random vectors satisfying the almost sure weak convergence theorem

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{n=1}^{N} \frac{1}{n} \mathbb{I}(\sqrt{n}(X_n - \mu) \leq t) = G_X(t) \text{ for } t \in C_{G_X} \text{ a.s.},$$

where G_X is the distribution function of some random variable X and C_{G_X} is the set of continuity points of G_X. If there is a sequence $\mathbb{N}_0 = \{n_k : k \in \mathbb{N}\}$ of integers such that

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{n=1; n \notin \mathbb{N}_0}^{N} \frac{1}{n} = 0$$

and

$$\lim_{k\to\infty} X_{n_k} = \mu \text{ a.s. },$$

then

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{n=1}^{N} \frac{1}{n} \mathbb{I}(\sqrt{n}(g(X_n) - g(\mu)) \leq t) = G_{\dot{g}(\mu)X}(t) \text{ for } t \in C_{G_{\dot{g}(\mu)X}} \text{ a.s.}$$

Next, we will introduce some definitions and the almost sure theorem for the correlation coefficient. Let $(X_n, Y_n)_{n\geq 1}$ be a sequence of i.i.d. vectors. The population correlation coefficient is given by

$$\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

and the sample correlation coefficient is defined as

$$r_n = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}.$$

Theorem 3.2.4. (Fridline [47]) Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be a random sample from a bivariate population such that $E(X_1^4) < \infty$ and $E(Y_1^4) < \infty$. Then

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}(\sqrt{k}(r_k - \rho) \leq t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi\gamma^2}} \exp(-\frac{u^2}{2\gamma^2}) du,$$

where

$$\gamma^2 = \frac{1}{4}\rho^2 \left(\frac{b_{33}}{\sigma_x^4} + 2\frac{b_{34}}{\sigma_x^2\sigma_y^2} + \frac{b_{44}}{\sigma_y^4} \right) - \rho \left(\frac{b_{35}}{\sigma_x^3\sigma_y} + \frac{b_{45}}{\sigma_x\sigma_y^3} \right) + \frac{b_{55}}{\sigma_x^2\sigma_y^2}$$

and

$$b_{33} = Var((X_1 - \mu_x)^2 - \sigma_x^2), b_{44} = Var((Y_1 - \mu_y)^2 - \sigma_y^2),$$

$$b_{55} = Var((X_1 - \mu_x)(Y_1 - \mu_y) - \sigma_{xy}), b_{34} = Cov((X_1 - \mu_x)^2, (Y_1 - \mu_y)^2),$$

$$b_{35} = Cov((X_1 - \mu_x)^2, (X_1 - \mu_x)(Y_1 - \mu_y)), b_{45} = Cov((Y_1 - \mu_y)^2, (X_1 - \mu_x)(Y_1 - \mu_y)).$$

If the distribution of the sample is bivariate normal, then $\gamma^2 = (1 - \rho^2)^2$. Note that Theorem 3.2.4 is the a.s. extension of the asymptotic result for the correlation coefficient and the advantage in this case is that γ^2 is not necessarily estimated. Combining Theorem 3.2.4 and Cramer's a.s. Theorem 3.2.3, we have the following:

Theorem 3.2.5. (Fridline [47]) Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be a random sample from a bivariate population such that $E(X_1^4) < \infty$ and $E(Y_1^4) < \infty$. Then

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}(\frac{\sqrt{k}}{2}(\log\frac{1+r_k}{1-r_k} - \log\frac{1+\rho}{1-\rho}) \leq t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi\tau^2}} \exp(-\frac{u^2}{2\tau^2}) du,$$

where

$$\tau^2 = \frac{\rho^2}{4(1-\rho^2)^2} \left(\frac{b_{33}}{\sigma_x^4} + 2\frac{b_{34}}{\sigma_x^2\sigma_y^2} + \frac{b_{44}}{\sigma_y^4} \right) - \frac{\rho}{(1-\rho^2)^2} \left(\frac{b_{35}}{\sigma_x^3\sigma_y} + \frac{b_{45}}{\sigma_x\sigma_y^3} \right) + \frac{b_{55}}{(1-\rho^2)^2\sigma_x^2\sigma_y^2}$$

If the distribution of the sample is bivariate normal, then $\tau^2 = 1$.

We can define the empirical distribution function

$$\widetilde{J}_N(t) = \frac{1}{C_N} \sum_{n=1}^{N} \frac{1}{n} \mathbb{I}(\frac{\sqrt{n}}{2}(\log\frac{1+r_n}{1-r_n} - \log\frac{1+\rho}{1-\rho}) \leq t)$$

and calculate confidence intervals based on it.

Definition 3.2.6. (Fridline [47]) The ASCLT-derived confidence interval for ρ

using the variance stabilizing technique is

$$I_\alpha^{(N)} = \left[\frac{\exp(2(z_N + \frac{\widetilde{t}_{1-\alpha}^{(N)}}{\sqrt{N}})) - 1}{\exp(2(z_N + \frac{\widetilde{t}_{1-\alpha}^{(N)}}{\sqrt{N}})) + 1}, \frac{\exp(2(z_N + \frac{\widetilde{t}_{\alpha}^{(N)}}{\sqrt{N}})) - 1}{\exp(2(z_N + \frac{\widetilde{t}_{\alpha}^{(N)}}{\sqrt{N}})) + 1} \right],$$

where $z_N = \log \frac{1+\widehat{\rho}}{1-\widehat{\rho}}$ and $\widetilde{t}_\alpha^{(N)} = \widetilde{J}_N^{-1}(\alpha)$.

When a sequence of statistics satisfies a weak convergence theorem and an almost sure limit theorem, both converging to the same distribution, one can obtain a type of approximation of the asymptotic distribution based on the logarithmic empirical distribution function. This leads to the definition of the logarithmic empirical quantile of the test statistic which convergences to the true quantile for continuous distributions. This allows for the development of hypothesis testing based on logarithmic quantile estimation. Two applications of this method were given here: the two-sample linear rank statistics considered by Thangavelu [97] and the correlation coefficient investigated by Fridline [47]. We presented the algorithm that was used by Thangavelu [97] for hypothesis testing problems and the method used by Fridline [47] to obtain confidence intervals for the population correlation. All these applications are based on the almost sure limit theorem for the statistic under consideration. Our goal is to extend these ideas and algorithms to general linear rank statistics. In Chapter 4, we take a first step in this direction and prove the almost sure central limit theorem for linear rank statistics.

Chapter 4

The Almost Sure Central Limit Theorem for Rank Statistics

This chapter is the core of the new theory, because it contains the main theorem upon which all special tests in this dissertation are based. We introduce the general model as in Brunner and Denker [20], but with the necessary changes required by the almost sure convergence. The proof of the theorem is based on a decomposition of the linear rank statistic and second moment estimates as in Brunner and Denker [20]. For some terms of this decomposition, we obtained an almost sure weak convergence using well known results from Berkes and Csáki [11]. The corollary that follows the main result shows how the logarithmic average of a sequence of linear rank statistics can be used to approximate the distribution function of the statistic under consideration. The proof of this corollary can be easily obtained from the main theorem and from Brunner and Denker [20].

4.1 Introduction to Linear Rank Statistics

We begin with a few historical facts and basic definitions connected to our theorem. In general, a simple linear rank statistic is given by

$$T = \sum_{i=1}^{N} c_i a_N(R_i),$$

where $c_1, ..., c_N \in \mathbb{R}$ are called regression constants, $R_1, ..., R_N$ are the ranks of N independent random variables $X_1, ..., X_N$ with distributions $F_1, ..., F_N$. The scores $a_N(i)$ are obtained from a function $h : (0, 1) \to \mathbb{R}$ such that for every $i = 1, ..., N$, $a_N(i) = h(\frac{i}{N+1})$ or $a_N(i) = Eh(U_N^{(i)})$, where $U_N^{(i)}$ is the ith order statistic in a sample of N random variables from the uniform distribution on (0, 1). For particular cases of score functions we obtain well-known test statistics. The Wilcoxon two-sample statistic is obtained with the Wilcoxon scores $a_N(i) = i$ and the regression constants $c_i = 1$ or 0, depending on which sample we consider. The van der Waerden scores $a_N(i) = \Phi^{-1}(\frac{i}{N+1})$ (here Φ^{-1} is the inverse of the standard normal distribution function) or the Fisher-Yates-Terry-Hoeffding scores $a_N(i) = E(Z_{(i)})$, where $Z_{(1)}, ..., Z_{(N)}$ denotes an ordered sample of N i.i.d. normal random variables, give rise to the corresponding two-sample linear rank statistics with the same name (see Denker [36] for more examples). An alternative way of representing a simple linear rank statistics is by stochastic integrals. If we let $h(\frac{i}{N+1}) = a_N(i)$ and define

$$\widehat{H}(t) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{I}(X_i \leq t)$$

and

$$\widehat{F}(t) = \frac{1}{N}\sum_{i=1}^{N}c_i\mathbb{I}(X_i \leq t),$$

then $R_i = N\widehat{H}(X_i)$ and T can be written as

$$T = N\int_{-\infty}^{\infty} h\left(\frac{N}{N+1}\widehat{H}(t)\right) d\widehat{F}(t).$$

Since our result is based on ideas from classical distribution theory, we continue our presentation with known results on the asymptotic normal distribution of linear rank statistics. The asymptotic normality of the simple linear rank statistics for two samples was obtained by Chernoff and Savage [30], and further generalized by Govindarajulu, LeCam, Raghavachari [50] with fewer assumptions on the function that generates the scores. Hájek [52] extended the results to general regression constants and more relaxed assumptions on the score generating function. His idea for proving the result was based on the projection method of approximating the linear rank statistic by a sum of independent random variables using conditional expectations. Under slightly stronger assumptions, Hoeffding [54] showed that Hájek's result [52] holds even if the linear rank statistic is centered by a simpler constant and not necessarily by its expectation.

In the multivariate case, a sequence of random vectors (usually of equal dimension) is considered and the test statistics are based on ranks over each coordinate (see Puri and Sen [87], Ruymgaart and van Zuijlen [90] for a general theory) or ranks over all observations. We will focus on methods that consider overall ranks. The idea of dealing with overall ranks can be traced back to Govindarajulu [51], who considered ranks over all paired dependent observations and then generalized by Brunner and Neumann [28] to a hierarchical design. Thompson [98] showed the asymptotic normality of linear rank statistics under dependent data using the overall ranks and a modified projection method. Thompson [99] proposed rank tests for univariate and multivariate designs under dependent data. Thompson [99] considered that the overall ranks may increase the power and there is no loss of information when compared to the traditional ranking methods (like in Friedman or Wilcoxon tests). Brunner and Denker [20] extended the result on the asymptotic normality of linear rank statistics for a sequence of random vectors with unequal dimensions. The proof of our result is based on the asymptotic normal distribution given in Brunner and Denker [20]. The almost sure central limit theorem that we obtain in this dissertation contributes to the general asymptotic theory of linear rank statistics and has interesting practical applications.

4.2 The Main Result and Remarks

We begin this section stating the model assumptions. Since they are different from the standard literature as in Brunner and Denker [20] and subsequently in Brunner and Puri [23], Brunner and Puri [25], Akritas and Arnold [1], and Akritas, Arnold and Brunner [2] to name a few, we need to state the notation in as much as it differs from those references. Note that in the general model (Brunner and Denker [20]) for each n, an array of independent random vectors $\mathbf{X}_i(n) = (X_{i1}(n), ..., X_{im_i(n)}(n))$, with $i = 1, 2, ..., n$ and $n \in \mathbb{N}$ was defined on, possibly, different probability spaces. Since we are heading for an almost sure type result, we need to consider the model on a common probability space. That is why our model requires a sequence of independent random vectors $\mathbf{X}_i = (X_{i1}, ..., X_{im_i}),\ i = 1, 2, ...$ with continuous marginal distributions

$$F_{ij}(x) = P(X_{ij} \leq x), x \in \mathbb{R}, j = 1, ..., m_i.$$

Note that in Munzel [79] this condition of continuity was shown to be unnecessary. Likewise the theorem below also holds when ties are present, as it is well known that one can replace ranks by midranks. For simplicity, we keep the commonly used assumption of having no ties. Note that we allow dependence of the coordinates of the random vectors, each vector may have a different dependence structure. This general approach excludes the use of the bootstrap method. As is well known (see Brunner and Denker [20] and subsequently Brunner and Puri [23], for example), this relaxation of the classical assumptions can be used for a large class of designs, for example repeated measure designs or time series observations. In order to set up the notation for rank statistics, we introduce the following notations.

For $n \geq 1$, let $N(n) = \sum_{i=1}^{n} m_i$ denote the number of observations involved in the vectors $\mathbf{X}_1, ..., \mathbf{X}_n$ and let $\lambda_{ij}^{(n)}$ $(1 \leq j \leq m_i, i \geq 1)$ be (known) regression constants which are assumed to satisfy

$$\max_{1 \leq i \leq n,\ 1 \leq j \leq m_i} |\lambda_{ij}^{(n)}| = 1. \tag{4.2.1}$$

Define

$$H^{(i)}(x) = \sum_{j=1}^{m_i} F_{ij}(x), \qquad \widehat{H}^{(i)}(x) = \sum_{j=1}^{m_i} \mathbb{I}(X_{ij} \leq x) \tag{4.2.2}$$

$$F^{(i,n)}(x) = \sum_{j=1}^{m_i} \lambda_{ij}^{(n)} F_{ij}(x), \qquad \widehat{F}^{(i,n)}(x) = \sum_{j=1}^{m_i} \lambda_{ij}^{(n)} \mathbb{I}(X_{ij} \leq x) \tag{4.2.3}$$

$$H_n(x) = \frac{1}{N(n)} \sum_{i=1}^{n} H^{(i)}(x), \qquad \widehat{H}_n(x) = \frac{1}{N(n)} \sum_{i=1}^{n} \widehat{H}^{(i)}(x) \tag{4.2.4}$$

$$F_n(x) = \frac{1}{N(n)} \sum_{i=1}^{n} F^{(i,n)}(x), \qquad \widehat{F}_n(x) = \frac{1}{N(n)} \sum_{i=1}^{n} \widehat{F}^{(i,n)}(x). \tag{4.2.5}$$

The simple linear rank statistic that we are interested in is defined by

$$L_n(J) = \int_{-\infty}^{\infty} J\left(\frac{N(n)}{N(n)+1}\widehat{H}_n\right) d\widehat{F}_n = \frac{1}{N(n)} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \lambda_{ij}^{(n)} J\left(\frac{R_{ij}(n)}{N(n)+1}\right), \quad (4.2.6)$$

where $R_{ij}(n)$ denotes the rank of X_{ij} among all random variables $\{X_{kl} : 1 \leq k \leq n, 1 \leq l \leq m_k\}$ and $J : (0,1) \to \mathbb{R}$ denotes an (absolutely continuous) score function. Let

$$T_n(J) = L_n(J) - \int_{-\infty}^{\infty} J(H_n)dF_n, \quad (4.2.7)$$

$$s_n^2(J) = N(n)^2 E(T_n(J)^2), \quad (4.2.8)$$

$$B_n(J) = \int_{-\infty}^{\infty} J(H_n)d(\widehat{F}_n - F_n) + \int_{-\infty}^{\infty} J'(H_n)(\widehat{H}_n - H_n)dF_n, \quad (4.2.9)$$

$$\sigma_n^2(J) = N(n)^2 \mathrm{Var}(B_n(J)). \quad (4.2.10)$$

The asymptotic normality of the linear rank statistics for independent random vectors with varying dimension introduced above was proved in Brunner and Denker ([20], Theorem 3.1).

The main result of this note is an almost sure central limit theorem for the statistics defined in (4.2.7).

Theorem 4.2.1. *Let $J : (0,1) \to \mathbb{R}$ be a twice differentiable score function with bounded second derivative and let $\lambda_{ij}^{(n)}$ be regression constants satisfying (4.2.1). Then the rank statistics (4.2.7) satisfies the almost sure central limit theorem, that is*

$$\lim_{N\to\infty} \frac{1}{\ln N} \sum_{n=1}^{N} \frac{1}{n} \mathbb{I}\left(\frac{N(n)}{\sigma_n(J)} T_n(J) \leq t\right) = \Phi(t)$$

provided
(a) $\sigma_n(J)$ defined in (4.2.10) satisfies, for some $M > 0, \gamma > 0$

$$\frac{\sigma_m(J)}{\sigma_n(J)} \geq M \left(\frac{m}{n}\right)^{\gamma}, \text{ for } m \geq n \quad (4.2.11)$$

(b) For $n_k = \min\{j : N(j) \geq k^2\}$ *one has that*

$$\sum_{k=1}^{\infty} \left(\frac{\max_{1 \leq i \leq n_k} m_i}{\sigma_{n_k}(J)} \right)^2 < \infty, \tag{4.2.12}$$

$$(\log n_k) \frac{\max_{1 \leq i \leq n_k} m_i}{\sigma_{n_k}(J)} \to 0, \tag{4.2.13}$$

and for $n_k \leq j < n_{k+1}$ *and a constant* K*,*

$$\max_{1 \leq \lambda \leq j} m_\lambda \leq K \max_{1 \leq i \leq n_k} m_i. \tag{4.2.14}$$

Remark 4.2.2. Note that the essential assumptions in the theorem are (a) and (b). (a) is a condition on the growth of the variances of the asymptotically equivalent statistics B_n (see Brunner and Denker [20] for details). This condition is easily verified in many examples, e.g. if $\sigma_n^2(J) = O(n)$. (b) is in fact a condition on the maximal allowable dimensions of the vectors X_i. If all m_i=1 then condition (b) is trivially satisfied, the same is true if $\max_{i \geq 1} m_i < \infty$.

The next corollary is a form of the theorem which can be used for hypothesis testing.

Corollary 4.2.3. *Assume that in Theorem 4.2.1,*

$$\sigma_n^2(J) = a_n^2 \sigma^2 + o(a_n)$$

where $\sigma^2 > 0$ *and* a_n *satisfies (4.2.11) when replacing* $\sigma_n(J)$ *by* a_n*.*
Then under (b), the statistics $\frac{N(n)}{a_n} T_n(J)$ *satisfies the central limit theorem and the almost sure central limit theorem, that is for* $t \in \mathbb{R}$

$$\lim_{n \to \infty} P\left(\frac{N(n)}{a_n} T_n(J) \leq t \right) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{t} e^{-\frac{u^2}{2\sigma^2}} du$$

and

$$\lim_{n \to \infty} \frac{1}{\ln n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I} \left(\frac{N(k)}{a_k} T_k(J) \leq t \right) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{t} e^{-\frac{u^2}{2\sigma^2}} du \ \ a.s.$$

The next remark states the properties under which hypothesis testing for $L_n(J)$ is possible.

Remark 4.2.4. Let $H_0 : \frac{N(n)}{\sigma_n(J)}\left(\int J(H_n)dF_n - c\right) \to 0$ as $n \to \infty$ for some constant c. Then under the null hypothesis $Q_n(J) := \frac{N(n)}{\sigma_n(J)}(L_n(J) - c)$ is asymptotically normal and satisfies the almost sure central limit theorem.
For $\alpha > 0$, let $\hat{t}_\alpha$ denote the empirical α-quantile of the empirical distribution function $\widehat{G}_N$. Then, under the null hypothesis,

$$I_\alpha^{(N)} = \left[Q_N(J) - \hat{t}_{1-\alpha}^{(N)}, Q_N(J) - \hat{t}_\alpha^{(N)}\right]$$

is a random interval with the property that

$$P(0 \in I_\alpha^{(N)}) \to 1 - 2\alpha.$$

If $H_1 : \frac{N(n)}{\sigma_n(J)}(\int J(H_n)dF_n - d) \to 0$ for $d \neq c$ and if $\frac{N(n)}{\sigma_n(J)} \to \infty$, then $|Q_n(J)| \to \infty$ a.s. It follows that under these conditions the power of the test approaches 1 under the alternative.

4.3 Proofs of Theorem 4.2.1 and Corollary 4.2.3

In this section, we give the proofs of Theorem 4.2.1 and Corollary 4.2.3 stated in Section 4.2. The proof of Theorem 4.2.1 uses additional lemmas that are stated and proved below. We use the notations and the assumptions that were introduced in Section 4.2.

Proof of Theorem 4.2.1. The proof of the theorem follows from a standard decomposition (Brunner and Denker [20]): Taylor expansion of J around $H_n(t)$ and integration by parts yields

$$\frac{N(n)}{\sigma_n(J)}T_n(J) = \frac{N(n)}{\sigma_n(J)}B_n(J) + \frac{N(n)}{\sigma_n(J)}C_1(n) - \frac{N(n)}{\sigma_n(J)}C_2(n) + \frac{N(n)}{\sigma_n(J)}C_3(n),$$

where

$$C_1(n) = \int_{-\infty}^{\infty} J'(H_n)(\widehat{H}_n - H_n)d(\widehat{F}_n - F_n), \tag{4.3.1}$$

$$C_2(n) = \frac{1}{N(n)+1}\int_{-\infty}^{\infty} J'(H_n)\widehat{H}_n d\widehat{F}_n, \tag{4.3.2}$$

$$C_3(n) = \frac{1}{2}\int_{-\infty}^{\infty} J''(\theta(H_n))\left(\frac{N(n)}{N(n)+1}\widehat{H}_n - H_n\right)^2 d\widehat{F}_n, \tag{4.3.3}$$

and $\theta(H_n) \in [H_n, \frac{N(n)}{N(n)+1}\widehat{H}_n] \cup [\frac{N(n)}{N(n)+1}\widehat{H}_n, H_n]$.

By Lemmas 4.3.2, 4.3.3, 4.3.5 it follows that

$$\frac{N(n)}{\sigma_n(J)}C_1(n) - \frac{N(n)}{\sigma_n(J)}C_2(n) + \frac{N(n)}{\sigma_n(J)}C_3(n) \to 0 \text{ a.s. when } n \to \infty. \tag{4.3.4}$$

By Lemma 2.2 in Fridline [47], Lemma 4.3.6 and (4.3.4), we obtain the almost sure central limit theorem for the statistics $T_n(J)$

$$\lim_{n\to\infty} \frac{1}{\ln n}\sum_{k=1}^{n} \frac{1}{k}\mathbb{I}\left(\frac{N(k)}{\sigma_k(J)}T_k(J) \le t\right) = \Phi(t).$$

□

Next, we give the proofs of the lemmas that were used in the proof of Theorem 4.2.1. The proofs of these lemmas rely on the following facts:

(1). The Borel-Cantelli lemma.
(2). The estimation of variances of sums is done by estimating covariances of summands uniformly if they do not vanish, multiplied by the number of non-vanishing covariances.
We shall use these without further mentioning.

Remark 4.3.1. We define

$$\begin{aligned}\phi^j_{iluv}(s,t) &= \lambda_{il}(J'(H_j(s))\mathbb{I}(t \le s) - J'(H_j(s))F_{uv}(s) - \\ &- \int J'(H_j(x))\mathbb{I}(t \le x)dF_{il}(x) + \int J'(H_j(x))F_{uv}(x)dF_{il}(x)).\end{aligned}$$

We list two properties of the functions ϕ^n_{ijkl} that are used repeatedly in the proofs of the following lemmas,

$$E(\phi^j_{iluv}(X_{il}, X_{uv})) = 0 \text{ if } i \neq u$$
$$E(\phi^j_{iluv}(X_{il}, X_{uv})\phi^j_{i'l'u'v'}(X_{i'l'}, X_{u'v'})) = 0,$$

if one of the indices i, i', u, u' is different from the others, where X_{il} is the l-th component of the vector $\mathbf{X}_i$.

Lemma 4.3.2.
$$\frac{N(n)}{\sigma_n(J)} C_1(n) \longrightarrow 0 \ a.s. \ when \ n \longrightarrow \infty.$$

Proof. We use the estimate of the second moment of $\frac{N(n)}{\sigma_n(J)} C_1(n)$ from Brunner and Denker [20]

$$E\left(\frac{N(n)}{\sigma_n(J)} C_1(n)\right)^2 = O\left(\frac{N(n)^2 ||J'||^2_\infty}{\sigma^2_n(J) N(n)^4}\left(\sum_{i=1}^n m_i^2\right)^2\right) = O\left(\left(\frac{\max_{1\leq i\leq n} m_i}{\sigma_n(J)}\right)^2\right). \tag{4.3.5}$$

For the subsequence n_k defined in Theorem 4.2.1, for every $\epsilon > 0$ and by Chebychev's inequality and relation (4.3.5) it follows

$$P\left(\frac{N(n_k)}{\sigma_{n_k}(J)} |C_1(n_k)| > \epsilon\right) \leq \frac{1}{\epsilon^2} E\left(\frac{N(n_k)}{\sigma_{n_k}(J)} C_1(n_k)\right)^2 \leq \frac{1}{\epsilon^2}\left(\frac{\max_{1\leq i\leq n_k} m_i}{\sigma_{n_k}(J)}\right)^2. \tag{4.3.6}$$

Using assumption (4.2.12) and relation (4.3.6) it follows that

$$\sum_{k=1}^\infty P\left(\frac{N(n_k)}{\sigma_{n_k}(J)} |C_1(n_k)| > \epsilon\right) \leq \sum_{k=1}^\infty \frac{1}{\epsilon^2}\left(\frac{\max_{1\leq i\leq n_k} m_i}{\sigma_{n_k}(J)}\right)^2 < \infty,$$

and by the Borel-Cantelli lemma,

$$\frac{N(n_k)}{\sigma_{n_k}(J)} C_1(n_k) \longrightarrow 0 \text{ a.s. when } k \ \to \infty.$$

So far we proved that Lemma 4.3.2 holds for the subsequence n_k. In order to prove that it holds for the whole sequence, it is necessary to show that what happens between the subsequence points does not influence the convergence. Thus, let

$j \in \mathbb{R}$ such that $n_k \le j < n_{k+1}$. The goal is to show that $\frac{N(j)}{\sigma_j(J)} C_1(j)$ converges to zero a.s. when $j \to \infty$. Notice that $C_1(j)$ defined in (4.3.1) can be decomposed as follows

$$\begin{aligned} C_1(j) &= \frac{1}{N(j)^2} \sum_{i=1}^{j} \sum_{l=1}^{m_i} \sum_{u=1}^{j} \sum_{v=1}^{m_u} \lambda_{il} (J'(H_j(X_{il})) \mathbb{I}(X_{uv} \le X_{il}) - \\ & \quad -J'(H_j(X_{il})) F_{uv}(X_{il}) - \int J'(H_j(t)) \mathbb{I}(X_{uv} \le t) dF_{il}(t) \\ & \quad + \int J'(H_j(t)) F_{uv}(t) dF_{il}(t)). \end{aligned}$$

Thus,

$$C_1(j) = \frac{1}{N(j)^2} \sum_{i=1}^{j} \sum_{l=1}^{m_i} \sum_{u=1}^{j} \sum_{v=1}^{m_u} \phi^j_{iluv}(X_{il}, X_{uv}) = \frac{1}{N(j)^2} \Phi^{j,j}_{1,1}(j),$$

where we defined

$$\begin{aligned} \phi^j_{iluv}(s,t) &= \lambda_{il}(J'(H_j(s)) \mathbb{I}(t \le s) - J'(H_j(s)) F_{uv}(s) - \\ & - \int J'(H_j(x)) \mathbb{I}(t \le x) dF_{il}(x) + \int J'(H_j(x)) F_{uv}(x) dF_{il}(x)) \end{aligned}$$

and

$$\Phi^{b,d}_{a,c}(j) = \sum_{i=a}^{b} \sum_{l=1}^{m_i} \sum_{u=c}^{d} \sum_{v=1}^{m_u} \phi^j_{iluv}(X_{il}, X_{uv}), \text{ for } a, b, c, d \in \mathbb{N}, a < b, c < d.$$

The term $\frac{N(j)}{\sigma_j(J)} C_1(j)$ can be estimated as

$$\left| \frac{N(j)}{\sigma_j(J)} C_1(j) \right| \le \left| \frac{N(j)}{\sigma_j(J)} C_1(j) - \frac{N(n_k)}{\sigma_{n_k}(J)} C_1(n_k) \right| + \left| \frac{N(n_k)}{\sigma_{n_k}(J)} C_1(n_k) \right|.$$

Since we showed that $\frac{N(n_k)}{\sigma_{n_k}(J)} C_1(n_k) \longrightarrow 0$ a.s. when $k \to \infty$, it is left to prove that

$$\left| \frac{N(j)}{\sigma_j(J)} C_1(j) - \frac{N(n_k)}{\sigma_{n_k}(J)} C_1(n_k) \right| \longrightarrow 0 \text{ a.s. when } j \to \infty.$$

Now

$$\frac{N(j)}{\sigma_j(J)}C_1(j) - \frac{N(n_k)}{\sigma_{n_k}(J)}C_1(n_k) = \frac{1}{\sigma_j(J)N(j)}\Phi^{j,j}_{1,1}(j) - \frac{1}{\sigma_{n_k}(J)N(n_k)}\Phi^{n_k,n_k}_{1,1}(n_k) =$$
$$= \frac{1}{\sigma_j(J)N(j)}\Phi^{n_k,n_k}_{1,1}(j) - \frac{1}{\sigma_{n_k}(J)N(n_k)}\Phi^{n_k,n_k}_{1,1}(n_k) + \frac{1}{\sigma_j(J)N(j)}\Phi^{n_k,j}_{1,n_k+1}(j) +$$
$$+\frac{1}{\sigma_j(J)N(j)}\Phi^{j,n_k}_{n_k+1,1}(j) + \frac{1}{\sigma_j(J)N(j)}\Phi^{j,j}_{n_k+1,n_k+1}(j).$$

Let

$$\mathrm{A}(j) = \frac{1}{\sigma_j(J)N(j)}\Phi^{n_k,n_k}_{1,1}(j) - \frac{1}{\sigma_{n_k}(J)N(n_k)}\Phi^{n_k,n_k}_{1,1}(n_k) = A_1(j) + A_2(j),$$

where

$$\begin{aligned} A_1(j) &= \frac{1}{\sigma_j(J)N(j)}\left(\Phi^{n_k,n_k}_{1,1}(j) - \Phi^{n_k,n_k}_{1,1}(n_k)\right), \\ A_2(j) &= \left(\frac{1}{\sigma_j(J)N(j)} - \frac{1}{\sigma_{n_k}(J)N(n_k)}\right)\Phi^{n_k,n_k}_{1,1}(n_k). \end{aligned}$$

We will show that $A_1(j)$ and $A_2(j)$ converge to zero a.s. using the Borel-Cantelli lemma. By Remark 4.3.1 we estimate the second moment of $A_1(j)$ as

$$E(A_1^2(j)) = O\left(\frac{||J'||^2}{\sigma_j^2(J)N(j)^2}\left(\frac{N(j)-N(n_k)}{N(j)}\right)^2\left((\max_{1\le i\le n_k} m_i)N(n_k)\right)^2\right).$$

Now, for some constant $C > 0$,

$$\sum_{k=1}^{\infty}\sum_{n_k\le j<n_{k+1}} P(|A_1(j)|\ge\epsilon) \le$$
$$\sum_{k=1}^{\infty}\sum_{n_k\le j<n_{k+1}} \frac{1}{\epsilon^2}\frac{||J'||^2}{\sigma_j^2(J)N(j)^2}\left((\max_{1\le i\le n_k} m_i)N(n_k)\right)^2\left(\frac{N(j)-N(n_k)}{N(j)}\right)^2 \le$$
$$\le C\frac{||J'||^2}{\epsilon^2}\sum_{k=1}^{\infty}\left(\frac{\max_{1\le i\le n_k} m_i}{\sigma_{n_k}(J)}\right)^2 < \infty,$$

which follows from assumption (4.2.12) and the facts that $N(n_k) \le N(j) < N(n_{k+1})$, $\sigma_{n_k}(J) \le \sigma_j(J) < \sigma_{n_{k+1}}(J)$, $N(n_{k+1}) - N(n_k) \ge n_{k+1} - n_k$ and $k^2 \le$

$N(n_k) < (k+1)^2$. Thus $A_1(j) \longrightarrow 0$ a.s. when $j \to \infty$.
For $A_2(j)$, notice that

$$|A_2(j)| = \left| \frac{\sigma_{n_k}(J)N(n_k)}{\sigma_j(J)N(j)} - 1 \right| \left| \frac{N(n_k)}{\sigma_{n_k}(J)} C_1(n_k) \right| \to 0 \text{ a.s. when } j \to \infty.$$

Let

$$B(j) = \frac{1}{\sigma_j(J)N(j)} \Phi^{n_k,j}_{1,n_k+1}(j).$$

We will show that $B(j) \longrightarrow 0$ a.s. when $j \to \infty$ using Borel-Cantelli. An estimate for the second moment of $B(j)$ is obtained using the Remark 4.3.1 and assumption (4.2.14)

$$E(B(j))^2 \leq \frac{16||J'||^2}{\sigma_j^2(J)N(j)^2} \left(\sum_{\lambda=n_k+1}^{j} m_\lambda^2 \right) \left(\sum_{i=1}^{n_k} m_i^2 \right)$$
$$\leq K \frac{16||J'||^2}{\sigma_j^2(J)N(j)^2} (N(j) - N(n_k)) (\max_{1 \leq i \leq n_k} m_i)^2 N(n_k).$$

Now,

$$\sum_{k=1}^{\infty} \sum_{n_k \leq j < n_{k+1}} P(|B(j)| \geq \epsilon) \leq$$
$$\leq \frac{16||J'||^2}{\epsilon^2} K \sum_{k=1}^{\infty} \left(\frac{\max_{1 \leq i \leq n_k} m_i}{\sigma_{n_k}(J)} \right)^2 \frac{(N(n_{k+1}) - N(n_k))^2}{N(n_k)} \leq$$
$$\leq \frac{32^2||J'||^2}{\epsilon^2} K \sum_{k=1}^{\infty} \left(\frac{\max_{1 \leq i \leq n_k} m_i}{\sigma_{n_k}(J)} \right)^2 < \infty,$$

so $B(j)$ converges to zero a.s. when $j \to \infty$.

Let

$$C(j) = \frac{1}{\sigma_j(J)N(j)} \Phi^{j,n_k}_{n_k+1,1}(j).$$

It can be shown that $C(j)$ converges to zero a.s. when $j \to \infty$ in the same way as for $B(j)$.
Let

$$D(j) = \frac{1}{\sigma_j(J)N(j)} \Phi^{j,j}_{n_k+1,n_k+1}(j).$$

Using the same techniques it can be shown that $D(j)$ converges to zero a.s. An estimate of the second moment of $D(j)$ is expressed as

$$E(D(j))^2 = O\left(\frac{||J'||^2}{\sigma_j^2(J)N(j)^2}\left(\sum_{i=n_k+1}^{j} m_i^2\right)^2\right).$$

By the Borel-Cantelli lemma and the assumption (4.2.14), it follows that

$$\sum_{k=1}^{\infty}\sum_{n_k\le j<n_{k+1}} P(|D(j)|\ge\epsilon)\le$$
$$\le\frac{||J'||^2}{\epsilon^2}\sum_{k=1}^{\infty}\sum_{n_k\le j<n_{k+1}}\left(\frac{\max_{1\le i\le n_{k+1}} m_i}{\sigma_{n_k}(J)}\right)^2\frac{(N(j)-N(n_k))^2\,\sigma_{n_k}^2(J)}{\sigma_j^2(J)N(j)^2}\le$$
$$\le\frac{||J'||^2}{\epsilon^2}\sum_{k=1}^{\infty}\left(\frac{\max_{1\le i\le n_{k+1}} m_i}{\sigma_{n_k}(J)}\right)^2\frac{((k+2)^2-k^2)^3}{k^4}<\infty,$$

thus $D(j)$ converges to zero a.s. when $j\to\infty$. □

Lemma 4.3.3.

$$\frac{N(n)}{\sigma_n(J)}C_2(n)\longrightarrow 0 \;\; a.s. \;\; when\;\; n\longrightarrow\infty.$$

Proof. Recall that $C_2(n)$ was defined in (4.3.2). Then using the assumption (4.2.12) it follows that

$$\left|\frac{N(n)}{\sigma_n(J)}C_2(n)\right| =$$
$$\left|\frac{1}{N(n)(N(n)+1)\sigma_n(J)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{n}\sum_{l=1}^{m_k} J'(H_n(X_{ij}))\mathbb{I}(X_{kl}\le X_{ij})\right|\le$$
$$\le ||J'||\frac{N(n)}{(N(n)+1)\sigma_n(J)}\longrightarrow 0 \text{ a.s. when } n\longrightarrow\infty.$$

□

Lemma 4.3.4.

$$D_n=\sup_{t\in\mathbb{R}}\left|\frac{1}{N(n)}\sum_{k=1}^{n}\sum_{l=1}^{m_k}(\mathbb{I}(X_{kl}\le t)-F_{kl}(t))\right|=O(a_n)\;\; a.s.$$

Proof. We shall use Singh's theorem and lemma [93]. Define

$$Y_k(t) = \frac{1}{m_k}\sum_{l=1}^{m_k} \mathbb{I}(X_{kl} \le t), \tag{4.3.7}$$

$$F_k(t) = EY_k(t) = \frac{1}{m_k}\sum_{l=1}^{m_k} F_{kl}(t). \tag{4.3.8}$$

Using (4.3.7) and (4.3.8), D_n can be expressed as

$$D_n = \sup_{t\in\mathbb{R}} \left| \sum_{k=1}^{n} w_k \left(Y_k(t) - F_k(t)\right) \right|,$$

where the weights are $w_k = \frac{m_k}{N(n)}$.
Using Singh's lemma [93], for $a_n \ge \frac{1}{N(n)}\sqrt{\sum_{k=1}^n m_k^2}$ the following inequality holds

$$P(D_n \ge a_n) < \frac{4a_n N(n)^2}{\sum_{k=1}^n m_k^2} \exp\left\{-2\left(\frac{a_n^2 N(n)^2}{\sum_{k=1}^n m_k^2} - 1\right)\right\}.$$

Using Singh's theorem [93], for any sequence $a_n \ge \frac{1}{N(n)}\sqrt{\sum_{k=1}^n m_k^2}$ such that $\sum_{n=1}^\infty \left\{\frac{a_n N(n)^2}{\sum_{k=1}^n m_k^2} \exp\left(-2\left(\frac{a_n^2 N(n)^2}{\sum_{k=1}^n m_k^2}\right)\right)\right\} < \infty$, it follows that

$$D_n = O(a_n) \quad \text{with probability 1.}$$

Take $a_n = b_n \frac{1}{N(n)}\sqrt{\sum m_k^2}$ with $b_n \sim c\sqrt{\log n}$ and $\sum_{n=1}^\infty \frac{b_n N(n)}{\sqrt{\sum_{k=1}^n m_k^2}} \exp(-2b_n^2) < \infty$. □

Lemma 4.3.5. *:*

$$\frac{N(n)}{\sigma_n(J)} C_3(n) \longrightarrow 0 \ \textit{a.s. when} \ n \longrightarrow \infty.$$

Proof. $C_3(n)$ defined in (4.3.3) can be written as

$$\begin{aligned} C_3(n) &= \frac{1}{2N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i} \lambda_{ij} J''(\theta(H_n(X_{ij}))) \times \\ &\times \left(\frac{1}{N(n)+1}\sum_{k=1}^{n}\sum_{l=1}^{m_k} \mathbb{I}(X_{kl} \le X_{ij}) - \frac{1}{N(n)}\sum_{k=1}^{n}\sum_{l=1}^{m_k} F_{kl}(X_{ij})\right)^2. \end{aligned}$$

Thus,

$$\left|\frac{N(n)}{\sigma_n(J)}C_3(n)\right| = \frac{1}{2\sigma_n(J)}|\sum_{i=1}^{n}\sum_{j=1}^{m_i}\lambda_{ij}J''(\theta(H(X_{ij})))(\frac{1}{N(n)+1}\sum_{k=1}^{n}\sum_{l=1}^{m_k}\mathbb{I}(X_{kl}\leq X_{ij}) -$$

$$-\frac{1}{N(n)}\sum_{k=1}^{n}\sum_{l=1}^{m_k}F_{kl}(X_{ij}))^2| \leq \frac{1}{2\sigma_n(J)}||J''||\sum_{i=1}^{n}\sum_{j=1}^{m_i}(\frac{1}{N(n)+1}\sum_{k=1}^{n}\sum_{l=1}^{m_k}\mathbb{I}(X_{kl}\leq X_{ij}) -$$

$$-\frac{1}{N(n)}\sum_{k=1}^{n}\sum_{l=1}^{m_k}F_{kl}(X_{ij}))^2 \leq \frac{1}{2\sigma_n(J)}||J''||\sum_{i=1}^{n}\sum_{j=1}^{m_i}2(\frac{1}{(N(n)+1)^2} +$$

$$+(\frac{1}{N(n)}\sum_{k=1}^{n}\sum_{l=1}^{m_k}(\mathbb{I}(X_{kl}\leq X_{ij}) - F_{kl}(X_{ij})))^2 = \frac{1}{\sigma_n(J)}||J''||\frac{N(n)}{(N(n)+1)^2} +$$

$$+\frac{1}{\sigma_n(J)}||J''||\sum_{i=1}^{n}\sum_{j=1}^{m_i}(\frac{1}{N(n)}\sum_{k=1}^{n}\sum_{l=1}^{m_k}(\mathbb{I}(X_{kl}\leq X_{ij}) - F_{kl}(X_{ij})))^2.$$

Using Lemma 4.3.4 it follows that

$$\left|\frac{N(n)}{\sigma_n(J)}C_3(n)\right| \leq \frac{1}{\sigma_n(J)}||J''||\frac{N(n)}{(N(n)+1)^2} + \frac{1}{\sigma_n(J)}||J''||\sum_{i=1}^{n}\sum_{j=1}^{m_i}D_n^2 \leq$$

$$\leq \frac{1}{\sigma_n(J)}||J''||\frac{N(n)}{(N(n)+1)^2} + \frac{1}{\sigma_n(J)}||J''||N(n)\frac{b_n^2\max m_k}{N(n)} \longrightarrow 0,$$

since $\frac{b_n^2 \max m_k}{\sigma_n(J)} \longrightarrow 0$ by (4.2.13) and choosing b_n of order $\log n$. □

Lemma 4.3.6.

$$\lim_{n\to\infty}\frac{1}{\ln n}\sum_{k=1}^{n}\frac{1}{k}\mathbb{I}\left(\frac{N(k)}{\sigma_k(J)}B_k(J)\leq t\right) = \Phi(t).$$

Proof. The term $B_n(J)$ defined in (4.2.9) can be expanded as

$$\begin{aligned} B_n(J) &= \frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\lambda_{ij}(J(H_n(X_{ij})) - \int J(H_n(x))dF_{ij}(x) + \\ &+ \int J'(H_n(x))\widehat{H}_n(x)dF_{ij}(x) - \int J'(H_n(x))H_n(x)dF_{ij}(x)). \quad (4.3.9)\end{aligned}$$

If we let

$$\begin{aligned}\sum_{i=1}^{n}\alpha_i &= \sum_{i=1}^{n}\sum_{j=1}^{m_i}\lambda_{ij}\left(J(H_n(X_{ij}))-\int J(H_n(x))dF_{ij}(x)\right)=\\ &= \sum_{i=1}^{n}\sum_{j=1}^{m_i}\lambda_{ij}\left(J(H_n(X_{ij}))-E(J(H_n(X_{ij})))\right)\end{aligned}$$

and

$$\begin{aligned}\sum_{i=1}^{n}\beta_i &= \sum_{i=1}^{n}\sum_{j=1}^{m_i}\lambda_{ij}\left(\int J'(H_n(x))\widehat{H}_n(x)dF_{ij}(x)-\int J'(H_n(x))H_n(x)dF_{ij}(x)\right)=\\ &= \sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{n}\sum_{l=1}^{m_k}\frac{1}{N(n)}\lambda_{ij}\int J'(H_n(x))\left(\mathbb{I}(X_{kl}\le x)-F_{kl}(x)\right)dF_{ij}(x)=\\ &= \sum_{k=1}^{n}\sum_{l=1}^{m_k}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{1}{N(n)}\lambda_{ij}\int J'(H_n(x))\left(\mathbb{I}(X_{kl}\le x)-F_{kl}(x)\right)dF_{ij}(x)=\\ &= \sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{n}\sum_{l=1}^{m_k}\frac{1}{N(n)}\lambda_{kl}\int J'(H_n(x))\left(\mathbb{I}(X_{ij}\le x)-F_{ij}(x)\right)dF_{kl}(x)=\\ &= \sum_{i=1}^{n}\left(\sum_{j=1}^{m_i}\frac{1}{N(n)}\sum_{k=1}^{n}\sum_{l=1}^{m_k}\lambda_{kl}\int J'(H_n(x))\left(\mathbb{I}(X_{ij}\le x)-F_{ij}(x)\right)dF_{kl}(x)\right)\end{aligned}$$

and

$$\begin{aligned}\xi_i &= \sum_{j=1}^{m_i}(\lambda_{ij}\left(J(H_n(X_{ij}))-\int J(H_n(x))dF_{ij}(x)\right)+\\ &+ \frac{1}{N(n)}\sum_{k=1}^{n}\sum_{l=1}^{m_k}\lambda_{kl}\int J'(H_n(x))\left(\mathbb{I}(X_{ij}\le x)-F_{ij}(x)\right)dF_{kl}(x)),\end{aligned}$$

then the relationship (4.3.9) can be rewritten as

$$N(n)B_n(J)=\sum_{i=1}^{n}\alpha_i+\sum_{i=1}^{n}\beta_i=\sum_{i=1}^{n}\xi_i,$$

and

$$\frac{N(n)B_n(J)}{\sigma_n(J)}=\frac{1}{\sigma_n(J)}\sum_{i=1}^{n}\xi_i. \tag{4.3.10}$$

The proof of the lemma follows from Theorem 1 of Berkes and Csáki [11]. This result is an almost sure central limit theorem for an independent sequence of random variables. Since in (4.3.10) $\frac{N(n)B_n(J)}{\sigma_n(J)}$ is expressed as the partial sum of the independent random variables ξ_i, it is left to check the assumptions in their theorem. First, we need to have a convergence in distribution, which is given by Theorem 3.1 in Brunner and Denker [20]

$$\frac{N(n)}{\sigma_n(J)}B_n(J) = \frac{1}{\sigma_n(J)}\sum_{i=1}^{n}\xi_i \xrightarrow{D} \mathcal{N}(0,1).$$

In Theorem 1 of Berkes and Csáki [11] we put $f_l(x_1,..,x_l) = \frac{1}{\sigma_l(J)}\sum_{i=1}^{l} x_i$ and $f_{k,l}(x_1,..,x_{l-k}) = \frac{1}{\sigma_l(J)}\sum_{i=1}^{l-k} x_i$ where $1 \le k \le l$ and put $c_l = l^\gamma$ where γ is as in (4.2.11). By Cauchy-Schwarz and using (4.2.10) we conclude that

$$E\left(\left|\frac{1}{\sigma_l(J)}\sum_{i=1}^{k}\xi_i\right| \wedge 1\right) \le \frac{1}{M}\left(\frac{k}{l}\right)^{\gamma}.$$

Thus, the theorem applies and we conclude that

$$\lim_{n\to\infty}\frac{1}{\ln n}\sum_{k\le n}\frac{1}{k}\mathbb{I}\left(\frac{1}{\sigma_k(J)}\sum_{i=1}^{k}\xi_i < x\right) = \Phi(x) \text{ a.s. for any } \; x\in\mathbb{R}.$$

□

Proof of Corollary 4.2.3

See Brunner and Denker [20] for the first part. The second part is a special case of Theorem 4.2.1.

In this chapter, we presented the proof of the almost sure weak convergence for simple linear rank statistics. The proof is divided into a couple of lemmas that show the almost sure convergence and the almost sure weak convergence of the terms involved in the decomposition of the linear rank statistic. The almost sure convergence of some of the terms is proved using the second moment estimates obtained in Brunner and Denker [22] and the Borel-Cantelli lemma. The almost sure weak convergence of one term is given by the theorem of Berkes and Csáki

[11]. Once we have the almost sure central limit theorem for linear rank statistics, we are ready to apply the ideas from Chapter 3 to hypothesis testing problems. In the next chapter, we consider the Kruskal-Wallis test statistic and calculate its quantiles based on the logarithmic quantile estimation.

Chapter 5

Factorial Designs I

The best known test for the nonparametric $c-$sample problem is the Kruskal-Wallis test for c independent samples of i.i.d. random variables $(X_{ij})_{1\leq i\leq n_j, 1\leq j\leq c}$. This test was proposed by Kruskal [67] and by Kruskal and Wallis [68]. The test statistic is given by

$$H = \frac{12}{N(N+1)} \sum_{j=1}^{c} \frac{1}{n_j} R_j^2 - 3(N+1),$$

where n_j is the sample size of the jth sample, $N = \sum_{j=1}^{c} n_j$, and $R_j = \sum_{i=1}^{n_j} R_{ij}$ where R_{ij} is the rank of X_{ij} in the overall ranking. Under the null hypothesis the distribution of the statistic can be explicitly computed since each rank combination has the probability $\prod_{i=1}^{c} \frac{n_i!}{N!}$. We use this below, for the data set from Neuhäuser [81] given in Example 5.1.3. It has been shown in Brunner and Denker [20] that the Kruskal-Wallis test can be also used for dependent samples, provided some information is known for the correlation between samples. Here we show that the logarithmic quantile estimation does not rely on this information.

In this chapter, we show that the almost sure limit theorem for the Kruskal-Wallis statistic can be obtained from the main result in Chapter 4, and it can be used for quantile estimation. For three independent samples we calculated the logarithmic empirical quantiles and compared them with the chi-squared quantiles given by the asymptotic distribution of the Kruskal-Wallis statistic. We also analyzed small real data sets from Boos [16] and Neuhäuser [81]. For three dependent samples we performed simulation studies to calculate the significance level and the power of the Kruskal-Wallis test based on logarithmic quantile estimation.

5.1 The Almost Sure Weak Convergence of the Kruskal-Wallis Statistic

The test statistic that we are using here is the classical Kruskal-Wallis statistic. In the case of dependent samples, the distribution of the test statistic and its asymptotic distribution under the null hypothesis are not known, and it is not possible to derive a statistical decision from this. In order to apply the logarithmic quantile estimation, we first need to show that the Kruskal-Wallis statistic is a particular statistic derived from the simple linear rank statistic $T_n(J)$ as defined in (4.2.7) and that it satisfies an almost sure central limit theorem. Let $\mathbf{X}_i = (X_{i1}, ..., X_{ic})$ $(1 \leq i \leq n)$ be independent random vectors such that the vectors $(X_{1k}, X_{2k}, ..., X_{nk})'$, for $k = 1, 2, ..., c$, are possibly dependent random variables with continuous marginal distribution functions $F_k(t) = P(X_{ik} \leq t)$ for $k = 1, ..., c$. We use the Wilcoxon scores $J(t) = t$ so that $J'(t) = 1$ and $\sigma_n = \sqrt{n}$. Then with the additional fixed index l that corresponds to the l-th sample, definitions (4.2.2)-(4.2.9) can be written as

$$\lambda_{ij}^{(l)} = \begin{cases} 1, j = l \\ 0, j \neq l \end{cases}$$

where $1 \leq i \leq n, 1 \leq j \leq c$,

$$\begin{aligned} H_n(t) &= \frac{1}{N(n)} \sum_{i=1}^{n} \sum_{k=1}^{c} F_k(t) = \frac{1}{c} \sum_{k=1}^{c} F_k(t), \hat{H}_n(t) = \frac{1}{N(n)} \sum_{i=1}^{n} \sum_{k=1}^{c} \mathbb{I}(X_{ik} \leq t) \\ F_n^{(l)}(t) &= \frac{1}{N(n)} \sum_{i=1}^{n} F_l(t) = \frac{1}{c} F_l(t), \qquad \hat{F}_n^{(l)}(t) = \frac{1}{N(n)} \sum_{i=1}^{n} \mathbb{I}(X_{il} \leq t) \end{aligned}$$

and

$$T_n^{(l)} = L_n^{(l)} - \int H_n(t) dF_n^{(l)} = \frac{1}{N(n)(N(n)+1)} \sum_{i=1}^{n} R_{il} - \frac{1}{c^2} \sum_{j=1}^{c} \int F_j(t) dF_l(t)$$

$$\begin{aligned} B_n^{(l)} &= \int H_n(t) d(\hat{F}_n^{(l)} - F_n^{(l)})(t) + \int (\hat{H}_n(t) - H_n(t)) dF_n^{(l)}(t) = \\ &\frac{1}{N(n)c} \sum_{i=1}^{n} \sum_{j=1}^{c} F_j(X_{il}) + \frac{1}{N(n)c} \sum_{i=1}^{n} \sum_{j=1}^{c} \int \mathbb{I}(X_{ij} \le t) dF_l(t) \\ &- \frac{2}{c^2} \sum_{j=1}^{c} \int F_j(t) dF_l(t) = \frac{1}{N(n)c} \sum_{i=1}^{n} \left(\sum_{j=1}^{c} F_j(X_{il}) + \right. \\ &\left. + \sum_{j=1}^{c} \int \mathbb{I}(X_{ij} \le t) dF_l(t) - 2 \sum_{j=1}^{c} \int F_j(t) dF_l(t) \right). \end{aligned}$$

Define the independent c-dimensional vectors

$$\begin{aligned} &\xi_i = (\xi_{ik})_{1 \le k \le c} \\ &= \left(\sum_{j=1}^{c} F_j(X_{ik}) + \sum_{j=1}^{c} \int \mathbb{I}(X_{ij} \le t) dF_k(t) - 2 \sum_{j=1}^{c} \int F_j(t) dF_k(t) \right)_{1 \le k \le c} \end{aligned} \quad (5.1.1)$$

and obtain

$$\frac{1}{\sigma_n} \sum_{i=1}^{n} \xi_i = \left(\frac{N(n)cB_n^{(1)}}{\sigma_n}, ..., \frac{N(n)cB_n^{(c)}}{\sigma_n} \right). \quad (5.1.2)$$

In the following, we will obtain an almost sure central limit theorem for the vector $(\frac{N(n)c}{\sigma_n} T_n^{(1)}, ..., \frac{N(n)c}{\sigma_n} T_n^{(c)})$ that holds under the null and alternative hypothesis. Under the null hypothesis, we then show that the Kruskal-Wallis statistic is written as a function of $T_n^{(1)}, ..., T_n^{(c)}$. In order to show the almost sure central limit theorem for the vectors $(\frac{N(n)c}{\sigma_n} T_n^{(1)}, ..., \frac{N(n)c}{\sigma_n} T_n^{(c)})$, it is sufficient to show the almost sure central limit theorem for the vectors in (5.1.1) and the statistics in (5.1.2) since we may argue as in the proof of Theorem 4.2.1. In order to show the almost sure central limit theorem we need to check the assumptions in Lifshits' Theorem 2.2.4.

The first assumption in Theorem 2.2.4 is to have distributional convergence for the vectors ξ_i. We need to assume that the vectors ξ_i's have a finite covariance matrix $\Sigma_i, i \ge 1$ such that

$$\frac{\Sigma_1 + ... + \Sigma_n}{n} \to \Sigma \text{ as } n \to \infty, \quad (5.1.3)$$

where Σ is a $c \times c$ matrix. Note that this is essentially a condition on the dependencies of the coordinates of the independent random vectors. Also note that $E(\xi_i) = 0$ and since all coordinates of the random vectors ξ_i are bounded, they satisfy the Lindeberg condition. Now, by the multivariate central limit theorem it follows that

$$\zeta_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \xrightarrow{D} N(0, \Sigma).$$

The second assumption is

$$b_k \leq c_1 \log\left(\frac{\sigma_k}{\sigma_{k-1}}\right) = c_1 \log \sqrt{\frac{k}{k-1}},$$

for some constant $c_1 > 0$ and where b_k appears in the formula of the empirical measures $Q_n = \frac{1}{\gamma_n} \sum_{k=1}^{n} b_k \delta_{\zeta_k}$ and $\gamma_n := \sum_{k=1}^{n} b_k$. It can be verified, for example using the mean value theorem, that

$$\frac{1}{n} \leq 2 \log \sqrt{\frac{n}{n-1}},$$

so we can take $b_n = \frac{1}{n}$ and $\gamma_n \sim \log(n)$.

The last assumption is that for some $\epsilon > 0$ it holds that

$$\sup_k E(\log_+ \log_+ ||\zeta_k||)^{1+\epsilon} < \infty.$$

It is easy to see that

$$E(\log_+ \log_+ ||\zeta_k||)^2 \leq E||\zeta_k||^2 = \frac{1}{k} \sum_{i=1}^{k} E(\xi_{i1}^2 + ... + \xi_{ic}^2) \leq \frac{1}{k} kc(2c)^2 = 4c^3,$$

since the vectors ξ_i are independent, have expectation zero and are bounded ($|\xi_{ik}| \leq 2c$ for every i and k). This shows that the sequence $(T_n^{(1)}, ..., T_n^{(c)})_{n \geq 1}$ satisfies the almost sure central limit theorem with limiting distribution function G_X for some normal random vector X. It follows for a continuous function $f : \mathbb{R}^c \to \mathbb{R}$ that

$$\frac{1}{\ln n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}\left(f(\frac{N(k)c}{\sigma_k} T_k^{(1)}, ..., \frac{N(k)c}{\sigma_k} T_k^{(c)}) \leq \mathbf{t} \right) \to G_{f(X)}(\mathbf{t}),$$

where $G_{f(X)}$ denotes the distribution function of $f(X)$. In particular, this applies to $f(\mathbf{x}) = \mathbf{x}^T\mathbf{x}$.

It is left to show that under the null hypothesis the Kruskal-Wallis statistics is such a function of $T_n^{(1)}, ..., T_n^{(c)}$. In this case, for a fixed index l the model is described by:

$$H_n(t) = F(t), \qquad \hat{H}_n(t) = \frac{1}{N(n)} \sum_{i=1}^{n} \sum_{j=1}^{c} \mathbb{I}(X_{ij} \leq t),$$

$$F_n^{(l)}(t) = \frac{1}{c} F(t), \qquad \hat{F}_n^{(l)}(t) = \frac{1}{N(n)} \sum_{i=1}^{n} \mathbb{I}(X_{il} \leq t),$$

$$B_n^{(l)}(J) = \frac{1}{N(n)c} \sum_{i=1}^{n} \left(cF(X_{il}) - \sum_{j=1}^{c} F(X_{ij}) \right),$$

$$T_n^{(l)} = L_n^{(l)} - \int H_n(t) dF_n^{(l)} = \frac{1}{N(n)(N(n)+1)} \sum_{i=1}^{n} R_{il} - \frac{1}{2c}.$$

Then, given that the Kruskal-Wallis statistic is defined by

$$F_n^{KW} = \frac{12}{N(n)(N(n)+1)} \frac{1}{n} \sum_{l=1}^{c} R_l^2 - 3(N(n)+1) \tag{5.1.4}$$

where

$$R_l = \sum_{i=1}^{n} R_{il},$$

it is possible to rewrite it as

$$F_n^{KW} = \frac{12N(n)(N(n)+1)}{n} \sum_{k=1}^{c} \left(T_n^{(k)}\right)^2. \tag{5.1.5}$$

Notice that under the null hypothesis, the asymptotic distribution of the Kruskal-Wallis statistic calculated for c dependent samples is not known. Still, an almost sure central limit theorem holds for the Kruskal-Wallis statistics. If (5.1.3) holds, then by the relationships (5.1.4) and (5.1.5) it follows that,

$$\frac{1}{\ln n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}\left(\frac{kc^3}{12(kc+1)} F_k^{KW} \leq t \right) \to G_{X^TX}(t),$$

where $X \sim \mathcal{N}(0, \Sigma)$ and $t \in \mathbb{R}$.

Remark 5.1.1. Observe that in the classical case, when we consider random vectors with independent coordinates (c independent samples), the Kruskal-Wallis statistic converges in distribution and an almost sure limit theorem holds under the null hypothesis. In particular,

$$F_n^{KW} \xrightarrow{D} \chi^2_{c-1} \Rightarrow \frac{nc^3}{12(nc+1)} F_n^{KW} \longrightarrow \frac{c^2}{12}\chi^2_{c-1} \tag{5.1.6}$$

$$\frac{1}{\ln n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}\left(\frac{kc^3}{12(kc+1)} F_k^{KW} \le t\right) \to G_{X^T X}(t). \tag{5.1.7}$$

Here it is left to show that $X^T X \sim \frac{c^2}{12}\chi^2_{c-1}$. If we take the expectation in (5.1.7), then

$$\frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} P\left(\frac{kc^3}{12(kc+1)} F_k^{KW} \le t\right) \to G_{X^T X}(t)$$

and using (5.1.6) it follows $G_{X^T X} = G_{\frac{c^2}{12}\chi^2_{c-1}}$.

Remark 5.1.2. Let us summarize the results that we obtained for the Kruskal-Wallis statistic. When considering c dependent samples, the asymptotic distribution under the null hypothesis is not known but an almost sure limit theorem holds. Note that the limiting distribution of the almost sure central limit theorem exists but it cannot be calculated. From an applied point of view, the existence of the almost sure central limit theorem allows us to use the empirical quantile estimation method described in Remark 4.2.4 for statistical decisions. In the independent c-sample case, the limiting distribution of the weak convergence and of the almost sure central limit theorem exists, and it can be calculated under the null hypothesis.

The Kruskal-Wallis statistic is used to test the null hypothesis

$$H_0 : F_1 = ... = F_c \Rightarrow H_0 : \int_{-\infty}^{\infty} F_j dF_k = \frac{1}{2} \text{ for every } j, k = 1, 2, ..., c. \tag{5.1.8}$$

The Kruskal-Wallis test rejects H_0 if and only if

$$\frac{nc^3}{12(nc+1)} F_n^{KW} > \widehat{t}^{(n)}_{1-\alpha}.$$

This rejection region provides an asymptotic α-level test for which the power tends to one under the alternative.

In the following example, we present two small data sets that were analyzed using the Kruskal-Wallis statistic when the samples are independent. We calculated the p-values using the logarithmic quantile estimation. For comparison reasons we also provide the p-values obtained from the asymptotic distribution and the permutation tests.

Example 5.1.3. We give two examples which are taken from the literature.

1. Neuhäuser [81] reports on a data set where persons in four groups were differently motivated for their jobs. Their level of concentration during work time was measured by scores. The data set is reproduced from Zöfel [103] on page 141 in Neuhäuser [81]. There were 28 people divided in 4 groups, a control group and 3 groups with different motivations. The application of the asymptotic chi-squared statistic gave the p-value 0.0018, while the exact permutation test showed a p-value of 0.0003. We applied our test and found a p-value of 0.0001, where we used an appropriate R–code as explained in Section 5.2.
2. Boos [16] studies the data set for a study appearing in Nation et al. [80]. The actual data set is contained in Boos ([16], page 1018). Twenty-seven rats were randomly assigned to 3 groups (including a control group with no cadmium in their diet, and 2 other groups where the diet contained 1, resp. 5, milligrams). The Kruskal-Wallis statistic was reported to have a value of 8.17 and its p-value < 0.02. The data represents the role of cadmium in the promotion of emotional reactivity (see Nation et al. [80]). Our test statistics gives 0.04 as a p-value using the R–code as in Section 5.2.

5.2 Simulation Results

In this section, we study simulations for the three-sample problem in Section 5.1. First, for three independent samples we perform simulation studies to show that the empirical logarithmic quantiles are good approximations of the asymptotic chi-squared quantiles of the Kruskal-Wallis statistic. Second, in case of three dependent samples we investigate by simulations the type I error and the power of the test given in (5.1.8).

The computation of the empirical logarithmic quantiles requires the following adjustments. First, for better results we deleted the first five terms in the sum of the empirical logarithmic distribution since their contribution is dominating. Secondly, since the empirical logarithmic distribution for a general statistic $T_n = T_n(X_1, ..., X_n)$ is not symmetric and the rejection or acceptance region might depend on the order of the observations, we considered a number of random permutations of the observations and calculated the quantities of the permuted sequence of independent vectors. Now the empirical logarithmic α-quantiles can be computed by

$$\widehat{t}_{\alpha}^{(n)} = \frac{\sum_{i=1}^{\text{per}} \widehat{t}_{\alpha}^{*i,(n)}}{\text{per}},$$

where "per" is the number of permutations that we want to consider and $\widehat{t}_{\alpha}^{*i,(n)}$ is the empirical logarithmic α-quantile for permutation i and is given by

$$\widehat{t}_{\alpha}^{*i,(n)} = \max\{t | \frac{1}{C_n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}(T_k^{*i} < t) \leq \alpha\},$$

where $T_k^{*i} = T_k(X_{\tau_i(1)}, ..., X_{\tau_i(k)})$ and τ_i is the i-th permutation of $\{1, 2, ..., n\}$.

For three independent samples, we use the almost sure central limit theorem given in (5.1.7). More precisely, using the method described above we compute the empirical logarithmic α-quantile for the statistic $\frac{27n}{12(3n+1)} F_n^{KW}$ and compare it to the asymptotic α-quantile given by $\frac{9}{12}\chi^2(2)$.

For this purpose, we run 500 simulations, consider 1000 observations in each sample and permute independently each sample 100 times. We generate random observations from both a normal and an exponential distribution. The numerical results of our simulation study are presented in Tables 5.1 and 5.2. Note that all quantiles are approximations to the unknown true distribution of the statistics.The table indicates in particular, that the logarithmic quantile method is distributionally stable.

Table 5.1. Averaged empirical logarithmic α-quantiles and the squared estimated standard errors for three independent samples with distribution Exp(3)

sample size/α	1%	s.e.	5%	s.e.	10%	s.e.
n=100	3.765	1.424	3.263	1.082	2.851	0.867
n=500	4.265	1.030	3.544	0.738	3.022	0.527
n=1000	4.237	0.721	3.444	0.506	2.886	0.365
$\frac{9}{12}\chi^2(2)$	6.908		4.494		3.454	

Table 5.2. Averaged empirical logarithmic α-quantiles and the squared estimated standard errors for three independent samples with distribution $\mathcal{N}(2,1)$

sample size/α	1%	s.e.	5%	s.e.	10%	s.e.
n=100	3.757	1.152	3.257	0.865	2.842	0.694
n=500	4.188	0.619	3.485	0.451	2.978	0.341
n=1000	4.228	0.592	3.439	0.420	2.884	0.313
$\frac{9}{12}\chi^2(2)$	6.908		4.494		3.454	

Similar results were obtained for other distributions: gamma, uniform and Cauchy. Also note that the simulation shows that the logarithmic quantiles match the asymptotic quantiles sufficiently well. In order to decide which approximation is better, we approximated the true significance level associated to the two test procedures

$$\frac{27 \times 1000}{12(3 \times 1000 + 1)} F_{1000}^{KW} > \widehat{t}_{0.9}^{(1000)}$$

and

$$\frac{27 \times 1000}{12(3 \times 1000 + 1)} F_{1000}^{KW} > \frac{9}{12}\chi^2(2)$$

by counting the number of rejections among 500 simulations. It is found that, for $\alpha = 0.9$ (a test with significance level of 10%), these covering probabilities are 0.108 for the LQE method and 0.12 for the asymptotic quantile method. This shows that the logarithmic quantile method is (in this study) better suited than the classical method.

We now turn to the second simulation problem. In the case of three dependent samples, the goal is to test whether the samples have the same distribution using the rejection rule for (5.1.8). We consider samples from a normal and an exponential distribution. We calculate the type I error and the power for different levels of α and different sample sizes. Next we briefly describe the algorithms we used to simulate two dependent samples from normal and exponential distributions and to form three dependent samples we add one more independent sample for simplicity. Note that our simulation study is an indication of what can be expected from a detailed analysis of the performance of the test.

To generate two dependent samples from a normal distribution we will simulate dependent bivariate normal random variables with specific parameters and population correlation coefficient ρ. We use the algorithm from Gentle ([49], page 197). We start by generating a matrix $\mathbf{X}_{(n\times 2)}$ of independent standard normal random variables (n independent copies of (X_1, X_2), where X_1 and X_2 are independent standard normal). Then we consider the covariance matrix Σ of the vector (X_1, X_2) given by (note that in this case the correlation coefficient $\rho_{X_1X_2}$is equal to $\sigma_{X_1X_2}$)

$$\Sigma = \begin{pmatrix} 1 & \sigma_{X_1X_2} \\ \sigma_{X_1X_2} & 1 \end{pmatrix}.$$

Using the Cholesky decomposition we obtain the matrix $\mathbf{T}_{(2\times 2)}$ such that $\mathbf{T}'\mathbf{T} = \Sigma$. Now $\mathbf{Y} = \mathbf{XT}'$ gives n independent copies of bivariate vectors (Y_1, Y_2), where Y_1 and Y_2 are dependent standard normal with correlation coefficient $\rho_{X_1X_2}$.

To generate two dependent samples from an exponential distribution, we will simulate dependent bivariate exponential random variables with specific parameters and correlation coefficient. We use the Marshall-Olkin method described in Devroye ([41], page 585). Start with the generation of three independent uniform random variables U, V, S on $[0, 1]$ and then construct $X_1 = \min\{-\frac{\ln U}{\lambda_1}, -\frac{\ln V}{\lambda_3}\}$ and $X_2 = \min\{-\frac{\ln S}{\lambda_2}, -\frac{\ln V}{\lambda_3}\}$. In this way we obtain a bivariate vector (X_1, X_2) with $X_1 \sim \text{Exp}(\lambda_1 + \lambda_3)$, $X_2 \sim \text{Exp}(\lambda_2 + \lambda_3)$ and the correlation coefficient is $\rho_{X_1X_2} = \frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3}$.

The results of our simulation studies are given in Tables 5.3–5.18. We start with the simulated significance level for different cases.

	n=30	n=50	n=80	n=100	n=150	n=200
1%	0	0	0	0	0	0
5%	0.015	0.03	0.03	0.02	0.025	0.025
10%	0.055	0.03	0.075	0.085	0.065	0.08

Table 5.3. The level of significance for three $\mathcal{N}(0,1)$ dependent samples; 200 simulations and 20 permutations, different sample sizes and different values of α

	n=30	n=50	n=80	n=100	n=150	n=200
1%	0.005	0	0	0	0	0
5%	0.02	0.02	0.04	0.035	0.01	0.025
10%	0.045	0.065	0.09	0.055	0.035	0.06

Table 5.4. The level of significance for three Exp(4) dependent samples; 200 simulations and 20 permutations, different sample sizes and different values of α

	n=30	n=50	n=80	n=100	n=150	n=200
1%	0	0	0	0	0	0
5%	0.025	0.04	0.05	0.05	0.045	0.04
10%	0.065	0.1	0.095	0.08	0.125	0.125

Table 5.5. The level of significance for three $\mathcal{N}(2,1)$ independent samples; 200 simulations and 20 permutations, different sample sizes and different values of α

	n=30	n=50	n=80	n=100	n=150	n=200
1%	0	0	0	0	0	0
5%	0.025	0.03	0.04	0.035	0.04	0.05
10%	0.07	0.085	0.08	0.08	0.105	0.105

Table 5.6. The level of significance for three Exp(3) independent samples; 200 simulations and 20 permutations, different sample sizes and different values of α

Note that the test is conservative and strongly conservative at 1% level. A correction factor of 0.9 and a larger number of simulations will increase some significance levels. Thangavelu's thesis [97] has an extensive simulation study and a detailed analysis of small sample properties. Also, we noticed that 20 random permutations are seemingly sufficient.

We also compute the power of the test for different distributions, sample sizes and different significance levels.

μ_1	μ_2	μ_3	n=30	n=50	n=80	n=100	n=150	n=200
0	1	0	1	1	1	1	1	1
1	1	0	0.955	1	1	1	1	1
0	0.5	0	0.44	0.83	0.975	0.995	1	1
0	0.2	0	0.075	0.12	0.215	0.295	0.45	0.59

Table 5.7. Power for three dependent samples from a normal distribution with different means at level $\alpha = 10\%$; 200 simulations and 20 permutations

μ_1	μ_2	μ_3	n=30	n=50	n=80	n=100	n=150	n=200
0	1	0	0.95	1	1	1	1	1
1	1	0	0.825	0.97	1	1	1	1
0	0.5	0	0.165	0.55	0.895	0.95	1	1
0	0.2	0	0.015	0.04	0.08	0.08	0.205	0.38

Table 5.8. Power for three dependent samples from a normal distribution with different means at level $\alpha = 5\%$; 200 simulations and 20 permutations

Note that the statistical power is close to 0 when the difference between the means is negligible and the sample size is small. We can also observe that it increases when the sample size and the significance level increase.

μ_1	μ_2	μ_3	n=30	n=50	n=80	n=100	n=150	n=200
0	1	0	0.155	0.81	1	1	1	1
1	1	0	0.18	0.72	0.97	1	1	1
0	0.5	0	0.005	0.025	0.1	0.195	0.615	0.835
0	0.2	0	0	0.005	0	0	0.01	0.025

Table 5.9. Power for three dependent samples from a normal distribution with different means at level $\alpha = 1\%$; 200 simulations and 20 permutations

μ_1	μ_2	μ_3	n=30	n=50	n=80	n=100	n=150	n=200
0.25	0.2	0.25	0.055	0.165	0.245	0.25	0.435	0.595
1	1	2	0.575	0.805	0.97	0.98	1	1
0.5	1	0.5	0.625	0.89	0.995	1	1	1
1	1	0.75	0.16	0.275	0.355	0.49	0.59	0.74

Table 5.10. Power for three dependent samples from an exponential distribution with different means at level $\alpha = 10\%$; 200 simulations and 20 permutations

μ_1	μ_2	μ_3	n=30	n=50	n=80	n=100	n=150	n=200
0.25	0.2	0.25	0.035	0.04	0.085	0.145	0.235	0.405
1	1	2	0.355	0.665	0.915	0.955	0.995	1
0.5	1	0.5	0.38	0.715	0.985	1	1	1
1	1	0.75	0.04	0.115	0.215	0.285	0.43	0.515

Table 5.11. Power for three dependent samples from an exponential distribution with different means at level $\alpha = 5\%$; 200 simulations and 20 permutations

μ_1	μ_2	μ_3	n=30	n=50	n=80	n=100	n=150	n=200
0.25	0.2	0.25	0.005	0.005	0.02	0.01	0.065	0.105
1	1	2	0.24	0.4	0.34	0.36	0.655	0.945
0.5	1	0.5	0.01	0.085	0.295	0.535	0.785	0.985
1	1	0.75	0.005	0.01	0.015	0.015	0.05	0.135

Table 5.12. Power for three dependent samples from an exponential distribution with different means at level $\alpha = 1\%$; 200 simulations and 20 permutations

μ_1	μ_2	μ_3	n=30	n=50	n=80	n=100	n=150	n=200
0	1	0	0.975	1	1	1	1	1
1	1	0	0.965	1	1	1	1	1
0	0.5	0	0.445	0.775	0.94	0.975	1	1
0	0.2	0	0.11	0.22	0.26	0.335	0.485	0.605

Table 5.13. Power for three independent samples from a normal distribution with different means at level $\alpha = 10\%$; 200 simulations and 20 permutations

μ_1	μ_2	μ_3	n=30	n=50	n=80	n=100	n=150	n=200
0	1	0	0.905	0.99	1	1	1	1
1	1	0	0.885	1	1	1	1	1
0	0.5	0	0.265	0.51	0.805	0.885	0.975	0.995
0	0.2	0	0.06	0.09	0.145	0.25	0.325	0.445

Table 5.14. Power for three independent samples from a normal distribution with different means at level $\alpha = 5\%$; 200 simulations and 20 permutations

μ_1	μ_2	μ_3	n=30	n=50	n=80	n=100	n=150	n=200
0	1	0	0.185	0.71	0.905	0.935	1	1
1	1	0	0.12	0.62	0.945	0.97	1	1
0	0.5	0	0.005	0.045	0.145	0.19	0.58	0.76
0	0.2	0	0	0.005	0.005	0	0.005	0.14

Table 5.15. Power for three independent samples from a normal distribution with different means at level $\alpha = 1\%$; 200 simulations and 20 permutations

μ_1	μ_2	μ_3	n=30	n=50	n=80	n=100	n=150	n=200
0.25	0.2	0.25	0.095	0.25	0.305	0.385	0.46	0.635
1	1	2	0.6	0.91	0.985	0.99	1	1
0.5	1	0.5	0.595	0.845	0.99	0.995	1	1
1	1	0.75	0.22	0.245	0.37	0.45	0.69	0.755

Table 5.16. Power for three independent samples from an exponential distribution with different means at level $\alpha = 10\%$; 200 simulations and 20 permutations

μ_1	μ_2	μ_3	n=30	n=50	n=80	n=100	n=150	n=200
0.25	0.2	0.25	0.045	0.1	0.165	0.195	0.335	0.44
1	1	2	0.475	0.715	0.93	0.96	0.995	1
0.5	1	0.5	0.39	0.695	0.93	0.99	1	1
1	1	0.75	0.065	0.145	0.27	0.28	0.59	0.605

Table 5.17. Power for three independent samples from an exponential distribution with different means at level $\alpha = 5\%$; 200 simulations and 20 permutations

μ_1	μ_2	μ_3	n=30	n=50	n=80	n=100	n=150	n=200
0.25	0.2	0.25	0	0.005	0.02	0.03	0.12	0.15
1	1	2	0.015	0.12	0.32	0.425	0.77	0.925
0.5	1	0.5	0.005	0.07	0.315	0.34	0.75	0.93
1	1	0.75	0	0.005	0.005	0.03	0.055	0.15

Table 5.18. Power for three independent samples from an exponential distribution with different means at level $\alpha = 1\%$; 200 simulations and 20 permutations

In this chapter, we showed that the Kruskal-Wallis statistic can be expressed as a quadratic form of the linear rank statistics and an almost sure weak convergence holds under the null hypothesis. We included the simulation results for the p-values using the logarithmic quantile estimation for two small data sets with three independent samples. For three dependent samples, we presented results on the significance level and the power of the test statistics. The next chapter deals with a different application of the logarithmic quantile estimation. We consider the shoulder tip pain study as an unbalanced nonparametric factorial design and obtain p-values using the logarithmic quantile estimation.

Chapter 6

Factorial Designs II

In this chapter, we show how to apply our general theory for designs other than the c-sample problem. We chose a data set of longitudinal observations first investigated by Lumley [75] and later on by Brunner, Domhof, Langer [21]. This longitudinal clinical study recorded pain scores on 41 patients measured at 6 time points. We considered the overall ranks of the observations and expressed the nonparametric hypotheses in terms of distribution functions. The appropriate test statistics are the rank versions of the ANOVA-type statistics. The results that we obtained using the logarithmic quantile estimation are comparable to the ones obtained by Brunner et al. [21] using the Box-approximation [17]. The logarithmic empirical quantile procedure has the advantage that it does not require the estimation of the covariance matrix as it does for the F asymptotic distribution approximation used by Brunner et al. [21]. We also derived an almost sure limit theorem for the statistics used to test the nonparametric hypotheses for each main factor and interaction.

6.1 Introduction to Nonparametric Factorial Designs

As mentioned earlier, the proposed new quantile estimation method can be used for many designs; however, each design has to be worked out separately. Thus, we begin this section discussing the theoretical and practical development of the nonparametric factorial designs. The nonparametric factorial designs were proposed as a way to relax the assumptions of linearity and normality of the errors in the classical ANOVA model. These designs are analyzed using rank test statistics and they go beyond the classical Friedman, Kruskal-Wallis tests. There were many ideas and attempts to develop the theory of nonparametric designs, but most of them were based on assumptions that are not realistic in practice. Many of the existing procedures did not include interaction effects or consider all distribution functions equal in the model such that the covariance matrices become simpler (see Brunner and Puri [23] for more details).

A different approach was taken by Conover and Iman [34] by proposing the "rank transform method". This method consists of replacing the observations with their ranks in a parametric test statistics and assuming that these tests follow the same asymptotic distribution. This is true for some particular models, but it does not hold for more complex designs. As a simple example, we consider the Kruskal-Wallis statistic and its parametric correspondent statistic. Let $(X_{ij})_{1\leq i\leq c, 1\leq j\leq n_i}$ be independent samples of independent random variables with distribution functions $F_1(x)$,..., $F_c(x)$. If all F_i are normal distributions with equal variance σ^2 and different means μ_i, then the test statistic for testing the equality of means μ_i is given by the one-way analysis of variance

$$F = \frac{1}{c-1}\sum_{i=1}^{c} n_i(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2 / \frac{1}{n-c}\sum_{i=1}^{c}\sum_{j=1}^{n_i} n_i(X_{ij} - \bar{X}_{i\cdot})^2,$$

where $\bar{X}_{i\cdot}$ is the mean of the observations in the ith sample and $\bar{X}_{\cdot\cdot}$ is the mean of the whole sample $n = \sum_{i=1}^{c} n_i$. When $n_i \to \infty, n_i/n \to p_i \in (0,1)$, c is fixed, then the denominator converges to σ^2 and the statistic F converges to $\chi^2_{c-1}/(c-1)$. The Kruskal-Wallis statistic is defined in the same way with $\bar{X}_{i\cdot}$ and $\bar{X}_{\cdot\cdot}$ being replaced by their rank versions $\bar{R}_{i\cdot}$ and $\bar{R}_{\cdot\cdot}$ and its asymptotic distribution is χ^2_{c-1}. Hora and Conover [56] considered the rank transform method for the two-way layout and proved that the F statistic used to test for the main effects with normal data has the same limiting null distribution when the observations are replaced by ranks. Akritas [4] introduced the "asymptotic version of the rank transform" and investigated this approach for the balanced and unbalanced nested models and the balanced and unbalanced two-way layout with and without interaction. Akritas [5],

[6] studied the rank transform method for repeated measures designs and showed the limitation and applicability of this method under different conditions and for specific hypotheses. Akritas [5], [6] concluded that the limitation of this method is due to the nonlinear behavior of the rank transformation of the data.

A breakthrough idea was introduced in Akritas and Arnold [1], where the nonparametric hypothesis of no main effects, no interaction effects, and no factor effects in analysis of variance and repeated measures designs were developed. Later on Akritas, Arnold and Brunner [2] extended this concept to unbalanced factorial designs. We present the example of a two-factor design as considered in [1], [2]. Let the kth observation in cell (i, j) be given by

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk},$$

where $i = 1, ..., a$ levels of factor A, $j = 1, ..., b$ levels of factor B, $k = 1, ..., n_{ij}$ and $\sum_i \alpha_i = 0$, $\sum_j \beta_j = 0$, $\sum_i \gamma_{ij} = 0$, $\sum_j \gamma_{ij} = 0$. The observations Y_{ijk} are considered independent, and the model is specified by the distribution of the errors e_{ijk}. If the errors follow a normal distribution, then we have the classical ANOVA case, and if they follow a general distribution, we have a robust parametric model or semiparametric model. For a pure nonparametric model, we assume

$$Y_{ijk} \sim F_{ij},$$

where F_{ij} is the normalized distribution function (defined in 6.2.1). In Akritas and Arnold [1], the distribution function is decomposed as

$$F_{ij}(y) = M(y) + A_i(y) + B_j(y) + C_{ij}(y),$$

where $\sum_i A_i = \sum_j B_j = 0$, $\sum_i C_{ij} = 0$ for all j, and $\sum_j C_{ij} = 0$ for all i. This decomposition is always possible since we can take $M = F_{\cdot\cdot}$, $A_i = F_{i\cdot} - M$, $B_j = F_{\cdot j} - M$, and $C_{ij} = F_{ij} - F_{i\cdot} - F_{\cdot j} + M$. The nonparametric hypotheses for the main factor effect A, B and the interaction effects AB are given in terms of distribution functions as follows

$$H_0(A) : A_i = F_{i\cdot} - F_{\cdot\cdot} = 0, \text{ for every } i = 1, ..., a,$$

$$H_0(B): B_j = F_{\cdot j} - F_{\cdot\cdot}, \text{ for every } j = 1, ..., b,$$

$$H_0(AB): C_{ij} = F_{ij} - F_{i\cdot} - F_{\cdot j} + F_{\cdot\cdot} = 0, \text{ for every } i = 1, ..., a \text{ and } j = 1, ..., b.$$

They are the nonparametric counterparts of the parametric hypotheses

$$H_0(A): \alpha_i = 0 \qquad H_0(B): \beta_j = 0 \qquad H_0(AB): \gamma_{ij} = 0.$$

The hypothesis of "no simple factor A effect" $\alpha_i + \gamma_{ij} = 0$ also has a nonparametric equivalent as $H_0(A|B): A_i + C_{ij} = F_{ij} - F_{\cdot j} = 0$, for every i, j. This hypothesis tests for the significance of the factor A effect on the response, either through the main effect or interaction. The same hypothesis can be defined for $H_0(B|A)$. The nonparametric hypotheses are more general than the parametric hypotheses. The ideas developed in Akritas and Arnold [1], Akritas, Arnold and Brunner [2] had a significant contribution on the development of nonparametric factorial designs. The nonparametric hypotheses, expressed only in terms of distribution functions look simple, but they have a meaningful statistical interpretation; they are invariant under monotone transformation and do not assume homoskedasticity.

6.2 The Shoulder Tip Pain Study

We introduce the shoulder tip pain study as a nonparametric factorial design and provide its analysis using the logarithmic quantile method described in the previous chapters. In this study, 41 patients had undergone a laparoscopic surgery and during the surgery gas accumulated in the abdomen causing shoulder pain. It is important to understand if a specific suction procedure to remove the abdominal gas may reduce the shoulder pain. To test this hypothesis, a group of 22 patients received the treatment, and the remaining 19 patients were the control group. Each group was divided by gender since pain sensitivity may depend on the gender of the patients. Shoulder pain scores from 1 (low pain) to 5 (high pain) were recorded for each patient at 6 time points (morning and evening for 3 days) after the surgery.

A first analysis of this clinical trial based on cumulative odds ratio was given by Lumley [75] and later on, Brunner et al. [21] analyzed this study as a nonparametric factorial design and provided p-values for the main factors (treatment

(A), gender (B), time (T)) and their interactions (AB, AT, BT, ABT). We start by describing the nonparametric model that was considered in Brunner et al. [21]. Let $\mathbf{X}_{ik} = (X_{ikj}), 1 \leq i \leq 4, 1 \leq k \leq n_i, 1 \leq j \leq t$ be the recorded scores at $t = 6$ time points of patient k within group i. We assume that the vectors $\mathbf{X}_{ik}$ are independent but their components may be dependent. Note that $n_1 = 14$ is the number of patients in the treatment-female group, $n_2 = 8$ is the number of patients in the treatment-male group, $n_3 = 11$ is the number of patients in the control-female group and $n_4 = 8$ is the number of patients in the control-male group. We denote by $n = \sum_{i=1}^{4} n_i$ the total number of subjects and by $N = nt$ the total number of observations. In nonparametric models, the hypotheses and the test statistics are defined in terms of distribution functions and relative marginal effects. We will continue with the definitions of these terms.

We define by

$$F_{ij}(x) = \frac{1}{2}\left(P(X_{ikj} \leq x) + P(X_{ikj} < x)\right) \tag{6.2.1}$$

the normalized version of the distribution function (see Ruymgaart [89], Brunner and Puri [24]) of X_{ikj} and its estimator

$$\widehat{F}_{ij}(x) = \frac{1}{n_i}\sum_{k=1}^{n_i} c(x - X_{ikj}),$$

where $c(u)$ is the normalized counting function such that $c(u) = 0, \frac{1}{2}$ or 1 when $u < 0, u = 0$ or $u > 0$. Note that the normalized version of the distribution functions provides a unified method for handling continuous and discrete data. Denote by

$$\mathbf{F} = (F_{ij})'_{1\leq i\leq 4, 1\leq j\leq t} \tag{6.2.2}$$

the vector of all 24 marginal distributions.

The relative marginal effects are defined as

$$p_{ij} = \int H dF_{ij}, \text{ for every } 1 \leq i \leq 4, 1 \leq j \leq t, \tag{6.2.3}$$

where

$$H(x) = \frac{1}{N}\sum_{i=1}^{4}\sum_{j=1}^{t} n_i F_{ij}(x)$$

is the weighted average of all distribution functions in the model. The relative marginal effects will not be used to formulate nonparametric hypotheses, since they depend on the sample sizes n_i through H. In general, the relative marginal effect p_{ij} shows the tendency of the distribution F_{ij} with respect to the weighted average H (for a detailed explanation see Brunner et al. [21], page 38). If we assume $H(x) = x$, then $p_{ij} = \int x dF_{ij}(x)$ becomes the expectation (if the integral exists), and in this sense the relative marginal effect can be interpreted as a generalized expectation. The estimators of the relative marginal effects are given by

$$\widehat{p}_{ij} = \int \widehat{H} d\widehat{F}_{ij} = \frac{1}{n_i}\sum_{k=1}^{n_i} \widehat{H}(X_{ikj}) = \frac{1}{N}\frac{1}{n_i}\sum_{k=1}^{n_i}\left(R_{ikj} - \frac{1}{2}\right), \tag{6.2.4}$$

where

$$\widehat{H}(x) = \frac{1}{N}\sum_{i=1}^{4}\sum_{j=1}^{t}\sum_{k=1}^{n_i} c(x - X_{ikj}),$$

and

$$R_{ikj} = \frac{1}{2} + N\widehat{H}(X_{ikj}) = \frac{1}{2} + \sum_{u=1}^{4}\sum_{v=1}^{t}\sum_{l=1}^{n_i} c(X_{ikj} - X_{ulv})$$

is the (mid)-rank of X_{ikj} among all N observations. In case of ties R_{ikj} is the midrank of X_{ikj}. Denote by

$$\widehat{\mathbf{p}} = (\widehat{p}_{ij})'_{1\leq i\leq 4, 1\leq j\leq t}$$

the vector of all 24 estimated relative treatment effects. The general form of a nonparametric hypothesis is stated as

$$H_0^F : \mathbf{CF} = 0, \tag{6.2.5}$$

where $\mathbf{C}$ is a suitable chosen contrast matrix and $\mathbf{F}$ is given in (6.2.2). The concept of nonparametric hypothesis based on distribution functions appeared in Akritas and Arnold [1] and further developed by Akritas et al. [2], [3]. In some sense it is similar to the hypothesis in linear models $H_0^{\mu} : \mathbf{C}\mu = 0$, where μ is a vector of expectations. We always have

$$H_0^F : \mathbf{CF} = 0 \Rightarrow H_0^{\mu} : \mathbf{C}\mu = 0,$$

since $\mathbf{C}\mu = \mathbf{C}\int xd\mathbf{F} = \int xd(\mathbf{CF})$.

The rank test statistics used to test the nonparametric hypotheses (6.2.5) are quadratic forms like Wald-type statistics and ANOVA-type statistics that involve the estimators of the relative marginal effects. The Wald-type statistic is given by

$$Q_n(\mathbf{C}) = n\widehat{\mathbf{p}}'\mathbf{C}'[\mathbf{C}\widehat{\mathbf{V}}_n\mathbf{C}']^{-}\mathbf{C}\widehat{\mathbf{p}},$$

where $[\mathbf{C}\widehat{\mathbf{V}}_n\mathbf{C}']^{-}$ is the generalized inverse of the covariance matrix $\mathbf{C}\widehat{\mathbf{V}}_n\mathbf{C}'$ and $\widehat{\mathbf{V}}_n$ is an estimator of the block-diagonal matrix $\mathbf{V}_n$ (see Brunner et al. [21], page 77) given by

$$\mathbf{V}_n = \mathrm{Cov}(\sqrt{n}\overline{\mathbf{Y}}.) = \bigoplus_{i=1}^{4} \frac{n}{n_i}\mathbf{W}_i, \tag{6.2.6}$$

$$\overline{\mathbf{Y}}. = (\overline{\mathbf{Y}}'_{1\cdot}, \overline{\mathbf{Y}}'_{2\cdot}, \overline{\mathbf{Y}}'_{3\cdot}, \overline{\mathbf{Y}}'_{4\cdot})',$$

$$\overline{\mathbf{Y}}'_{i\cdot} = \frac{1}{n_i}\sum_{k=1}^{n_i}\mathbf{Y}_{ik} = \frac{1}{n_i}\sum_{k=1}^{n_i}(H(X_{ik1}), ..., H(X_{ikt}))', i = 1, 2, 3, 4$$

$$\mathbf{W}_i = \mathrm{Cov}(\mathbf{Y}_{i1}), i = 1, 2, 3, 4.$$

The estimator $\widehat{\mathbf{V}}_n$ is given by

$$\widehat{\mathbf{V}}_n = \bigoplus_{i=1}^{4} \frac{n}{n_i}\widehat{\mathbf{W}}_i,$$

where

$$\widehat{\mathbf{W}}_i = \frac{1}{N^2(n_i - 1)}\sum_{k=1}^{n_i}(\mathbf{R}_{ik} - \overline{\mathbf{R}}_{i\cdot})(\mathbf{R}_{ik} - \overline{\mathbf{R}}_{i\cdot})'$$

is the sample covariance matrix of $\frac{1}{N}\mathbf{R}_{ik}$ with $\mathbf{R}_{ik} = (R_{ik1}, ..., R_{ikt})'$ the vector of the ranks R_{iks} of X_{iks} among all N observations and $\overline{\mathbf{R}}_{i\cdot} = \frac{1}{n_i}\sum_{k=1}^{n_i}\mathbf{R}_{ik}$, $i = 1, 2, 3, 4$ and $k = 1, ..., n_i$. It is known that under the null hypothesis (6.2.5) the Wald-type statistic has asymptotically a χ_f^2 distribution with $f = \mathrm{rank}(\mathbf{C})$ degrees of freedom. This test can be used when f is very small (1 or 2) or when the sample size of each group is very large (min $n_i > 200$) (see Akritas and Brunner [3] for simulation studies).

The Wald-type statistic is not appropriate for the analysis of the shoulder tip pain study and we will use the ANOVA-type statistic (see Brunner and Puri [24], Brunner et al. [21], Brunner et al. [27]). To introduce the ANOVA-type statistic we need an equivalent formulation of the null hypothesis (6.2.5) as

$$H_0^F : \mathbf{MF} = 0, \tag{6.2.7}$$

where $\mathbf{M} = \mathbf{C}'(\mathbf{CC}')^{-}\mathbf{C}$ is a projection matrix and $(\mathbf{CC}')^{-}$ is the generalized inverse of $\mathbf{CC}'$. Note that the matrix $\mathbf{M}$ does not depend on the choice of the generalized inverse $(\mathbf{CC}')^{-}$ and its elements are real numbers. The two formulations of the null hypothesis in (6.2.7) and (6.2.5) are equivalent since $\mathbf{C}'(\mathbf{CC}')^{-}$ is a generalized inverse of $\mathbf{C}$. The ANOVA-type statistic is defined by

$$Q_n(\mathbf{M}) = n\widehat{\mathbf{p}}'\mathbf{M}\widehat{\mathbf{p}}. \tag{6.2.8}$$

It is known that under the following assumptions

(a) $\min n_i \to \infty$, for every $1 \le i \le 4$,
(b) $\frac{n}{n_i} \le N_0 < \infty$, for every $1 \le i \le 4$,
(c) $\rho_m(i) \ge \rho_0 > 0$, $\rho_m(i)$ is the smallest characteristic root of $\mathbf{W}_i$ for $1 \le i \le 4$,

and under the null hypothesis (6.2.7) the ANOVA-type statistic $Q_n(\mathbf{M})$ has asymptotically, the same distribution as the random variable $\sum_{i=1}^{4}\sum_{j=1}^{t}\lambda_{ij}U_{ij}$, where $U_{ij} \sim \chi_1^2$ are independent random variables and λ_{ij}'s are eigenvalues of $\mathbf{MV}_n\mathbf{M}$, where the matrix $\mathbf{V}_n$ is defined in (6.2.6). The difficulty in using this test statistic comes from the fact that the eigenvalues λ_{ij} are not known and they need to be estimated.

Following Box's idea [17], Brunner, Dette, and Munk [26] proposed approximating the distribution of the $\sum_{i=1}^{4}\sum_{j=1}^{t}\lambda_{ij}U_{ij}$ by a scaled chi-squared distribution $g\chi_f^2$ such that the first two moments of these distributions be equal. This approximation procedure concludes that

$$F_n(\mathbf{M}) := \frac{1}{\mathrm{tr}(\mathbf{M}\widehat{\mathbf{V}}_n)} Q_n(\mathbf{M}) = \frac{n\widehat{\mathbf{p}}'\mathbf{M}\widehat{\mathbf{p}}}{\mathrm{tr}(\mathbf{M}\widehat{\mathbf{V}}_n)} \sim F(\widehat{f}, \infty),$$

where

$$\widehat{f} = \frac{(\mathrm{tr}(\mathbf{M}\widehat{\mathbf{V}}_n))^2}{\mathrm{tr}(\mathbf{M}\widehat{\mathbf{V}}_n\mathbf{M}\widehat{\mathbf{V}}_n)}.$$

There are other forms of this approximation procedure and they depend on the nonparametric design considered (see Brunner et al. [24], [21]). The logarithmic quantile estimation introduced in the previous chapters is another way of approximating the distribution of the ANOVA-type statistic and avoids the estimation of the covariance matrix $\mathbf{V}_n$. We describe and compare the performance of our procedure with the results in [21] for the shoulder tip pain study. In order to use the logarithmic quantile estimation for the statistics defined in (6.2.8), we need to verify the almost sure weak convergence of (6.2.9) towards the limiting distribution. This proof is deferred to Section 6.3. Next, we explain the method of calculating the logarithmic quantiles for our statistics and in Table 1 we provide the p-values that we obtained.

The logarithmic average of the sequence $Q_n(\mathbf{M})$ has the form

$$\widehat{G}_N(t) = \frac{1}{C_N}\sum_{n=1}^{N}\frac{1}{n}\mathbb{I}(Q_n(\mathbf{M}) \leq t), \tag{6.2.9}$$

where $C_N = \sum_{n=1}^{N}\frac{1}{n}$ and where $\mathbb{I}(C)$ denotes the indicator function of the set C. $\widehat{G}_N$ is an empirical distribution function and the empirical $\alpha-$quantile of $\widehat{G}_N$ can be used in hypothesis testing, for example a typical rejection region may look like $\{X \in \mathbb{R}^{dN} : |Q_N(\mathbf{M})| \geq \widehat{t}_\alpha^{(N)}\}$ with $\widehat{G}_N(\widehat{t}_\alpha^{(N)}) = \alpha$. Since the empirical logarithmic distribution for the general statistic $Q_n(\mathbf{M}) = Q_n(\mathbf{M})(X_1, ..., X_n)$ is not symmetric and the rejection or acceptance region might depend on the order of the observations, we considered a number of random permutations of the observations and calculated the quantities of the permuted sequence of independent vectors. Now, the empirical logarithmic α-quantiles can be computed by

$$\widehat{t}_{\alpha}^{(N)} = \frac{\sum_{i=1}^{\text{per}} \widehat{t}_{\alpha}^{*i,(N)}}{\text{per}},$$

where "per" is the number of permutations that we considered and $\widehat{t}_{\alpha}^{*i,(N)}$ is the empirical logarithmic α-quantile for permutation i and is given by

$$\widehat{t}_{\alpha}^{*i,(N)} = \max\{t | \frac{1}{C_N} \sum_{n=1}^{N} \frac{1}{n} \mathbb{I}(Q_n(\mathbf{M})^{*i} < t) \leq \alpha\},$$

where $Q_n(\mathbf{M})^{*i} = Q_n(\mathbf{M})(X_{\tau_i(1)}, ..., X_{\tau_i(n)})$ and τ_i is the i-th permutation of {1, 2,..., N}.

Our simulation results using the logarithmic quantile estimation are presented in Table 6.1 and the results from [21] using the F distribution approximation are given in Table 6.2 below.

factor	$Q_n(\mathbf{M})$	p-value
A (treatment)	15.06499	0.032
B (gender)	0.04251446	0.8746
T (time)	1.817526	0.1029
AB	0.03291587	0.8596
AT	1.994102	0.0774
BT	0.6149522	0.4785
ABT	0.2351336	0.7698

Table 6.1. The value of the statistic and the empirical logarithmic p-value for per=100 permutations

Note that we obtained larger p-values than in Brunner et al. [21] and at level 0.05 we conclude that the time effect and the interaction treatment-time effect are not significant, which is contrary to the results in [21]. However, the result of a 10% and 7.7% p-value are acceptable to confirm [21].

factor	$F_n(\mathbf{M})$	$\hat{f}_1$	$\hat{f}_0$	p-value
A (treatment)	16.401	1.0	21.86	0.0005
B (gender)	0.046	1.0	21.86	0.8317
T (time)	3.382	2.7	∞	0.0212
AB	0.036	1.0	21.86	0.8516
AT	3.711	2.7	∞	0.0140
BT	1.144	2.7	∞	0.3273
ABT	0.438	2.7	∞	0.7054

Table 6.2. The p-values for the ANOVA-type statistics; taken from Brunner et al. [21], page 191

Next, we present details regarding the nonparametric null hypothesis and the test statistic for each factor and each interaction in the shoulder tip pain study. The null hypotheses can be expressed as in equation (6.2.5), where the contrast matrix $\mathbf{C}$ takes particular forms (see [24]):

$$\begin{aligned}
H_0^F(A) &: \mathbf{C}_A\mathbf{F} = 0, & \mathbf{C}_A &= \mathbf{P}_a \otimes \frac{1}{b}\mathbf{1}_b' \otimes \frac{1}{t}\mathbf{1}_t', \\
H_0^F(B) &: \mathbf{C}_B\mathbf{F} = 0, & \mathbf{C}_B &= \frac{1}{a}\mathbf{1}_a' \otimes \mathbf{P}_b \otimes \frac{1}{t}\mathbf{1}_t', \\
H_0^F(AB) &: \mathbf{C}_{AB}\mathbf{F} = 0, & \mathbf{C}_{AB} &= \mathbf{P}_a \otimes \mathbf{P}_b \otimes \frac{1}{t}\mathbf{1}_t', \\
H_0^F(T) &: \mathbf{C}_T\mathbf{F} = 0, & \mathbf{C}_T &= \frac{1}{a}\mathbf{1}_a' \otimes \frac{1}{b}\mathbf{1}_b' \otimes \mathbf{P}_t, \\
H_0^F(AT) &: \mathbf{C}_{AT}\mathbf{F} = 0, & \mathbf{C}_{AT} &= \mathbf{P}_a \otimes \frac{1}{b}\mathbf{1}_b' \otimes \mathbf{P}_t, \\
H_0^F(BT) &: \mathbf{C}_{BT}\mathbf{F} = 0, & \mathbf{C}_{BT} &= \frac{1}{a}\mathbf{1}_a' \otimes \mathbf{P}_b \otimes \mathbf{P}_t, \\
H_0^F(ABT) &: \mathbf{C}_{ABT}\mathbf{F} = 0, & \mathbf{C}_{ABT} &= \mathbf{P}_a \otimes \mathbf{P}_b \otimes \mathbf{P}_t.
\end{aligned}$$

Here $a = 2, b = 2, t = 6$, $\otimes$ denotes the Kronecker product of matrices, $\mathbf{1}_t = (1, ..., 1)'$ denotes the t-dimensional vector of 1's, $\mathbf{P}_t = \mathbf{I}_t - \frac{1}{t}\mathbf{J}_t$ is a t-dimensional projection matrix of rank $t-1$, with $\mathbf{I}_t$ the t-dimensional unit matrix and $\mathbf{J}_t = \mathbf{1}_t\mathbf{1}_t'$. The ANOVA-type statistics are calculated according to the formula given in (6.2.8):

$$Q_n^A(\mathbf{M}) = \frac{n}{4t}\left(\sum_{j=1}^{t}\widehat{p}_{1j} + \sum_{j=1}^{t}\widehat{p}_{2j} - \sum_{j=1}^{t}\widehat{p}_{3j} - \sum_{j=1}^{t}\widehat{p}_{4j}\right)^2,$$

$$Q_n^B(\mathbf{M}) = \frac{n}{4t}\left(\sum_{j=1}^{t}\widehat{p}_{1j} - \sum_{j=1}^{t}\widehat{p}_{2j} + \sum_{j=1}^{t}\widehat{p}_{3j} - \sum_{j=1}^{t}\widehat{p}_{4j}\right)^2,$$

$$Q_n^{AB}(\mathbf{M}) = \frac{n}{4t}\left(\sum_{j=1}^{t}\widehat{p}_{1j} - \sum_{j=1}^{t}\widehat{p}_{2j} - \sum_{j=1}^{t}\widehat{p}_{3j} + \sum_{j=1}^{t}\widehat{p}_{4j}\right)^2,$$

$$Q_n^{T}(\mathbf{M}) = \frac{n}{4t}\sum_{i=1}^{t}\sum_{j=1,j>i}^{t}(S_i - S_j)^2,$$

where $S_j = \widehat{p}_{1j} + \widehat{p}_{2j} + \widehat{p}_{3j} + \widehat{p}_{4j}$,

$$Q_n^{AT}(\mathbf{M}) = \frac{n}{4t}\sum_{i=1}^{t}\sum_{j=1,j>i}^{t}(S_i - S_j)^2,$$

where $S_j = \widehat{p}_{1j} + \widehat{p}_{2j} - \widehat{p}_{3j} - \widehat{p}_{4j}$,

$$Q_n^{BT}(\mathbf{M}) = \frac{n}{4t}\sum_{i=1}^{t}\sum_{j=1,j>i}^{t}(S_i - S_j)^2,$$

where $S_j = \widehat{p}_{1j} - \widehat{p}_{2j} + \widehat{p}_{3j} - \widehat{p}_{4j}$,

$$Q_n^{ABT}(\mathbf{M}) = \frac{n}{4t}\sum_{i=1}^{t}\sum_{j=1,j>i}^{t}(S_i - S_j)^2,$$

where $S_j = \widehat{p}_{1j} - \widehat{p}_{2j} - \widehat{p}_{3j} + \widehat{p}_{4j}$ and $\widehat{p}_{ij}$, for $i = 1, 2, 3, 4$ are the estimators of the relative marginal effects defined in (6.2.4).

6.3 Technical Details

Note that the p-values that are shown in Table 6.1 were obtained using the logarithmic quantile estimation for the test statistics $Q_n^A(\mathbf{M})$,..., $Q_n^{ABT}(\mathbf{M})$. Before applying the logarithmic quantile estimation, it is necessary to obtain an almost sure limit theorem for each test statistic. This can be shown by considering particular forms of the simple linear rank statistic introduced in equation (4.2.7). We use the basic notation from Chapter 4, equations (4.2.1)-(4.2.10). Define

$$H_n(x) = \frac{1}{N}\sum_{i=1}^{4}\sum_{j=1}^{t} n_i F_{ij}(x),$$

$$\widehat{H}_n(x) = \frac{1}{N}\sum_{i=1}^{4}\sum_{k=1}^{n_i}\sum_{j=1}^{t} \mathbb{I}(X_{ikj} \le x),$$

For a fixed $1 \le l \le t$ and for a fixed $1 \le v \le 4$, define the following

$$F_{v,n}^{(l)}(x) = \frac{1}{N}\sum_{i=1}^{4}\sum_{j=1}^{t} \lambda_{ij}^{(n,v,l)} n_i F_{ij}(x) = \frac{n_v}{N} F_{vl}(x),$$

$$\widehat{F}_{v,n}^{(l)}(x) = \frac{1}{N}\sum_{i=1}^{4}\sum_{k=1}^{n_i}\sum_{j=1}^{t} \lambda_{ij}^{(n,v,l)} \mathbb{I}(X_{ikj} \le x) = \frac{1}{N}\sum_{k=1}^{n_v} \mathbb{I}(X_{vkl} \le x),$$

where

$$\lambda_{ij}^{(n,v,l)} = \begin{cases} 1, & \text{if } i = v, j = l \\ 0, & \text{otherwise,} \end{cases}$$

and the simple linear rank statistics

$$T_{v,n}^{(l)} = \frac{1}{N(N+1)}\sum_{k=1}^{n_v} R_{vkl} - \frac{n_v}{N}\int_{-\infty}^{\infty} H_n dF_{vl}.$$

It can be easily seen that under the null hypothesis, the test statistics used in Section 6.2 become

$$\begin{aligned} Q_n^A(\mathbf{M}) &= \frac{n(nt+1)^2}{4t}\left(\frac{1}{n_1}\sum_{j=1}^{t} T_{1,n}^{(j)} + \frac{1}{n_2}\sum_{j=1}^{t} T_{2,n}^{(j)} - \right. \\ &\left. -\frac{1}{n_3}\sum_{j=1}^{t} T_{3,n}^{(j)} - \frac{1}{n_4}\sum_{j=1}^{t} T_{4,n}^{(j)}\right)^2, \end{aligned}$$

$$\begin{aligned} Q_n^B(\mathbf{M}) &= \frac{n(nt+1)^2}{4t}\left(\frac{1}{n_1}\sum_{j=1}^{t} T_{1,n}^{(j)} - \frac{1}{n_2}\sum_{j=1}^{t} T_{2,n}^{(j)} + \right. \\ &\left. +\frac{1}{n_3}\sum_{j=1}^{t} T_{3,n}^{(j)} - \frac{1}{n_4}\sum_{j=1}^{t} T_{4,n}^{(j)}\right)^2, \end{aligned}$$

$$\begin{aligned} Q_n^{AB}(\mathbf{M}) &= \frac{n(nt+1)^2}{4t}\left(\frac{1}{n_1}\sum_{j=1}^{t}T_{1,n}^{(j)} - \frac{1}{n_2}\sum_{j=1}^{t}T_{2,n}^{(j)} - \right. \\ &\left. -\frac{1}{n_3}\sum_{j=1}^{t}T_{3,n}^{(j)} + \frac{1}{n_4}\sum_{j=1}^{t}T_{4,n}^{(j)}\right)^2, \end{aligned}$$

$$\begin{aligned} Q_n^{T}(\mathbf{M}) &= \frac{n(nt+1)^2}{4t}\sum_{l=1}^{t}\sum_{j=1,j>l}^{t}\left(\frac{1}{n_1}T_{1,n}^{(l)} + \frac{1}{n_2}T_{2,n}^{(l)} + \frac{1}{n_3}T_{3,n}^{(l)} + \right. \\ &\left. +\frac{1}{n_4}T_{4,n}^{(l)} - \frac{1}{n_1}T_{1,n}^{(j)} - \frac{1}{n_2}T_{2,n}^{(j)} - \frac{1}{n_3}T_{3,n}^{(j)} - \frac{1}{n_4}T_{4,n}^{(j)}\right)^2, \end{aligned}$$

$$\begin{aligned} Q_n^{AT}(\mathbf{M}) &= \frac{n(nt+1)^2}{4t}\sum_{l=1}^{t}\sum_{j=1,j>l}^{t}\left(\frac{1}{n_1}T_{1,n}^{(l)} + \frac{1}{n_2}T_{2,n}^{(l)} - \frac{1}{n_3}T_{3,n}^{(l)} - \right. \\ &\left. -\frac{1}{n_4}T_{4,n}^{(l)} - \frac{1}{n_1}T_{1,n}^{(j)} - \frac{1}{n_2}T_{2,n}^{(j)} + \frac{1}{n_3}T_{3,n}^{(j)} + \frac{1}{n_4}T_{4,n}^{(j)}\right)^2, \end{aligned}$$

$$\begin{aligned} Q_n^{BT}(\mathbf{M}) &= \frac{n(nt+1)^2}{4t}\sum_{l=1}^{t}\sum_{j=1,j>l}^{t}\left(\frac{1}{n_1}T_{1,n}^{(l)} - \frac{1}{n_2}T_{2,n}^{(l)} + \frac{1}{n_3}T_{3,n}^{(l)} - \right. \\ &\left. -\frac{1}{n_4}T_{4,n}^{(l)} - \frac{1}{n_1}T_{1,n}^{(j)} + \frac{1}{n_2}T_{2,n}^{(j)} - \frac{1}{n_3}T_{3,n}^{(j)} + \frac{1}{n_4}T_{4,n}^{(j)}\right)^2, \end{aligned}$$

$$\begin{aligned} Q_n^{ABT}(\mathbf{M}) &= \frac{n(nt+1)^2}{4t}\sum_{l=1}^{t}\sum_{j=1,j>l}^{t}\left(\frac{1}{n_1}T_{1,n}^{(i)} - \frac{1}{n_2}T_{2,n}^{(l)} - \frac{1}{n_3}T_{3,n}^{(l)} + \right. \\ &\left. +\frac{1}{n_4}T_{4,n}^{(l)} - \frac{1}{n_1}T_{1,n}^{(j)} + \frac{1}{n_2}T_{2,n}^{(j)} + \frac{1}{n_3}T_{3,n}^{(j)} - \frac{1}{n_4}T_{4,n}^{(j)}\right)^2. \end{aligned}$$

Under the null hypothesis, the almost sure weak convergence of these test statistics holds. The almost sure weak convergence of the vector

$$\mathbf{T}_n = \left(\frac{\sqrt{n}(N+1)}{n_i}T_{i,n}^{(j)}\right)_{1\le i\le 4, 1\le j\le t}$$

can be obtained as in Chapter 4 by showing the almost sure weak convergence of the vector

$$\mathbf{B}_n = \frac{(N+1)}{\sqrt{n}N}\left(\frac{N}{\frac{n_i}{n}}B_{i,n}^{(j)}\right)_{1\le i\le 4, 1\le j\le t},$$

where for a fixed $1 \le l \le t$ and $1 \le v \le 4$,

$$B_{v,n}^{(l)} = \frac{1}{N}\sum_{k=1}^{n_v} H_n(X_{vkl}) - 2\frac{n_v}{N}\int H_n dF_{vl} + \frac{n_v}{N^2}\sum_{i=1}^{4}\sum_{k=1}^{n_i}\sum_{j=1}^{t}\int I(X_{ikj} \le x)dF_{vl}(x).$$

The almost sure weak convergence of $\mathbf{B}_n$ is obtained using Lifschits' Theorem 2.2.4. For this we need to show that $\mathbf{B}_n$ can be expressed as a sum of independent vectors and check the assumptions in Theorem 2.2.4. For a fixed n, we let $\gamma_i = \frac{n}{n_i}$ and assume there is λ_0, N_0 independent of n with:

(1) $\max_n \max_{1\le i\le 4} \gamma_i \le N_0 < \infty$,

(2) $1 - \lambda_0 \le \max_n \frac{n_i}{n} \le \lambda_0 < 1$ for $1 \le i \le 4$,

(3) as $n \to \infty$, $\gamma_i = \gamma_i(n) \to \tilde{\gamma}_i$.

Now, since $n_1 > n_3 > n_2 = n_4$, it follows that $\mathbf{B}_n$ can be written as

$$\mathbf{B}_n = \frac{(N+1)}{N}\frac{1}{\sqrt{n}}\sum_{k=1}^{n_1}\mathbf{Z}_k, \tag{6.3.1}$$

where

$$\begin{aligned}\mathbf{Z}_k &= (z_{ikj})_{1\le i\le 4, 1\le j\le t} = \\ &= ((\xi_{ikj} + u_{ikj} + y_{ikj} + w_{ikj}) - E(\xi_{ikj} + u_{ikj} + y_{ikj} + w_{ikj}))_{1\le i\le 4, 1\le j\le t}.\end{aligned}$$

For a fixed $1 \le l \le t$ and $1 \le k \le n_1$ we define

$$\xi_{1kl} = \frac{n}{n_1}H_n(X_{1kl}) + \frac{n}{N}\sum_{j=1}^{t}\int I(X_{1kj} \le x)dF_{1l}(x),$$

$$\xi_{ikl} = \frac{n}{N}\sum_{j=1}^{t}\int I(X_{ikj} \le x)dF_{il}(x), \text{ for } i = 2, 3, 4.$$

For a fixed $1 \le l \le t$ we define

$$u_{ikl} = \frac{n}{N}\sum_{j=1}^{t}\int I(X_{ikj} \le x)dF_{il}(x),\ \text{for } 1 \le k \le n_2, i = 1,3,4,$$

$$u_{ikl} = 0,\ \text{for } n_2 + 1 \le k \le n_1, i = 1,3,4,$$

$$u_{2kl} = \frac{n}{n_2}H_n(X_{2kl}) + \frac{n}{N}\sum_{j=1}^{t}\int I(X_{2kj} \le x)dF_{2l}(x),\ \text{for } 1 \le k \le n_2,$$

$$u_{2kl} = 0,\ \text{for } n_2 + 1 \le k \le n_1.$$

For a fixed $1 \le l \le t$ we define

$$y_{ikl} = \frac{n}{N}\sum_{j=1}^{t}\int I(X_{ikj} \le x)dF_{il}(x),\ \text{for } 1 \le k \le n_3, i = 1,2,4,$$

$$y_{ikl} = 0,\ \text{for } n_3 + 1 \le k \le n_1, i = 1,2,4,$$

$$y_{3kl} = \frac{n}{n_3}H_n(X_{3kl}) + \frac{n}{N}\sum_{j=1}^{t}\int I(X_{3kj} \le x)dF_{3l}(x),\ \text{for } 1 \le k \le n_3,$$

$$y_{3kl} = 0,\ \text{for } n_3 + 1 \le k \le n_1$$

For a fixed $1 \le l \le t$ we define

$$v_{ikl} = \frac{n}{N}\sum_{j=1}^{t}\int I(X_{ikj} \le x)dF_{il}(x),\ \text{for } 1 \le k \le n_4, i = 1,2,3,$$

$$v_{ikl} = 0,\ \text{for } n_4 + 1 \le k \le n_1$$

$$v_{4kl} = \frac{n}{n_4}H_n(X_{4kl}) + \frac{n}{N}\sum_{j=1}^{t}\int I(X_{4kj} \le x)dF_{4l}(x),\ \text{for } 1 \le k \le n_4,$$

$$v_{4kl} = 0,\ \text{for } n_4 + 1 \le k \le n_1.$$

Now we check the assumptions in Theorem 2.2.4. The first condition is the weak convergence of $\mathbf{B}_n$. We make the assumption that the independent vectors $\mathbf{Z}_k$ have a finite covariance matrix Σ_k, $1 \le k \le n_1$ such that

$$\frac{\Sigma_1 + ... + \Sigma_{n_1}}{n_1} \to \Sigma \text{ as } n_1 \to \infty.$$

Note that $E(\mathbf{Z}_k) = 0$ and since they are bounded, they satisfy the Lindeberg condition. Now, by the multivariate central limit theorem it follows

$$\zeta_{n_1} := \frac{1}{\sqrt{n_1}} \sum_{k=1}^{n_1} \mathbf{Z}_k \to \mathbf{N}(0, \Sigma) \text{ as } n_1 \to \infty$$

and then

$$\mathbf{B}_{n_1\gamma_1} = \frac{n_1\gamma_1 t + 1}{n_1\gamma_1 t} \frac{1}{\sqrt{n_1\gamma_1}} \sum_{k=1}^{n_1} \mathbf{Z}_k \to \frac{1}{\sqrt{\tilde{\gamma}}_1} \mathbf{N}(0, \Sigma) \text{ as } n_1 \to \infty.$$

The second assumption is that for some $\epsilon > 0$ it holds

$$\sup_{n_1} E(\log_+ \log_+ ||\zeta_{n_1}||)^{1+\epsilon} < \infty.$$

It is easy to see that for $\epsilon = 1$ we have

$$E(\log_+ \log_+ ||\zeta_{n_1}||)^2 \leq E(||\zeta_{n_1}||)^2 = \frac{1}{n_1} \sum_{k=1}^{n_1} E((Z_{1k1})^2 + ... + (Z_{4kt})^2) \leq$$

$$\leq \frac{1}{n_1} \sum_{k=1}^{n_2} 6 \left((\frac{n}{n_1} + 4)^2 + ... + (\frac{n}{n_4} + 4)^2 \right) + \frac{1}{n_1} \sum_{k=n_2+1}^{n_3} 6 \left((\frac{n}{n_1} + 4)^2 + (\frac{n}{n_3} + 4)^2 \right)$$

$$+ \frac{1}{n_1} \sum_{k=n_3+1}^{n_1} 6(\frac{n}{n_1} + 4)^2 \leq (N_0 + 4)^2 (24\lambda_0 N_0 + 12N_0 + 6(1 - N_0(\lambda_0 - 1)))$$

According to Theorem 2.2.4 it holds for every vector $\mathbf{x}$

$$\lim_{n_1 \to \infty} \frac{1}{\log n_1} \sum_{k=1}^{n_1} \frac{1}{k} \mathbb{I} \left(\frac{1}{\sqrt{k}} \sum_{i=1}^{k} \mathbf{Z}_i \leq \mathbf{x} \right) = G_X(\mathbf{x}) \text{ a.s. },$$

where $X \sim \mathbf{N}(0, \Sigma)$.

By Fridline's lemma ([47], Lemma 2.2) we have

$$\lim_{n_1 \to \infty} \frac{1}{\log n_1} \sum_{k=1}^{n_1} \frac{1}{k} \mathbb{I}(\mathbf{B}_{k\gamma_1} \leq \mathbf{x}) = G_X(\mathbf{x}\sqrt{\tilde{\gamma}_1}) \text{ a.s.},$$

where $X \sim \mathbf{N}(0, \Sigma)$ and $\mathbf{B}_{k\gamma_1} = \frac{k\gamma_1 t+1}{k\gamma_1 t} \frac{1}{\sqrt{k\gamma_1}} \sum_{i=1}^{k} \mathbf{Z}_i$.

Following the same idea as in Chapter 4, for any continuous function $f : \mathbb{R}^{24} \to \mathbb{R}$, it holds

$$\lim_{n_1\to\infty} \frac{1}{\log n_1} \sum_{k=1}^{n_1} \frac{1}{k} \mathbb{I}(f(\mathbf{B}_{k\gamma_1}) \le \mathbf{x}) = G_{f(X)}(\mathbf{x}\sqrt{\tilde{\gamma}_1}). \tag{6.3.2}$$

Using (6.3.2) with specific functions we obtain the almost sure central limit theorem for the statistics $Q_n^A(\mathbf{M}), ..., Q_n^{ABT}(\mathbf{M})$ introduced in Section 6.2. Finally we arrive at the following result.

Proposition 6.3.1. *Under the assumptions*
(1) $\rho_m(i) \ge \rho_0 > 0$, *where* $\rho_m(i)$ *is the smallest characteristic root of* $\boldsymbol{W}_i$ *for* $1 \le i \le 4$,
(2) $\max_n \max_{1\le i\le 4} \gamma_i \le N_0 < \infty$,
(3) $1 - \lambda_0 \le \max_n \frac{n_i}{n} \le \lambda_0 < 1$ *for* $1 \le i \le 4$,
(4) as $n \to \infty$, $\gamma_i = \gamma_i(n) \to \tilde{\gamma}_i$,
(5) $\frac{\Sigma_1 + ... + \Sigma_{n_1}}{n_1} \to \Sigma$ *as* $n_1 \to \infty$,
the statistics $Q_n(\boldsymbol{M})$ *used in Section 6.2 satisfy*

$$\lim_{N\to\infty} \frac{1}{C_N} \sum_{n=1}^{N} \frac{1}{n} \mathbb{I}(Q_n(\boldsymbol{M}) \le t) = G(t) \ \ a.s.,$$

where G *is the distribution of* $\sum_{i=1}^{4} \sum_{j=1}^{t} \lambda_{ij} U_{ij}$ *defined in Section 6.2.*

In this chapter, we analyzed the "shoulder tip pain" data set using the almost sure quantile estimation introduced in the previous chapters. We considered the overall ranks of the observations and used the theory of the nonparametric factorial designs to express the hypotheses testing in terms of distribution functions. The test statistics that we considered were quadratic forms of the linear rank statistics and for each test statistic we obtained an almost sure weak convergence. This allowed us to compute the empirical logarithmic quantiles.

Chapter 7

Conclusion

7.1 Summary

In this dissertation, we proposed an alternative method to the classical decision theory. This approach is based on the almost sure limit theorems for sequences of statistics. These theorems can be viewed as uniform strong law of large numbers, a Glivenko-Cantelli type theorems. They can also be interpreted as the almost sure convergence of the empirical distribution function of a sequence of statistics. This empirical distribution function has a logarithmic form and its inverse can be used to approximate, almost surely, the true quantiles of the statistic under consideration. We studied the performance of this method for simple linear rank statistics. First, we obtained the almost sure central limit theorem for simple linear rank statistics. We used the same model that was proposed in Brunner and Denker [20] but with the necessary modification needed for the almost sure convergence. This was the main result upon which the statistical applications were based.

The second part of this thesis dealt with the applications to quantile estimation. We studied the performance of the logarithmic quantile estimation using the Kruskal-Wallis statistic with independent or dependent samples and the "shoulder tip pain study". The Kruskal-Wallis statistic is a quadratic form of linear rank statistics and its almost sure convergence holds under the null hypothesis. Two small data sets with independent samples were analyzed using the Kruskal-Wallis test and the p-values obtained with the logarithmic quantile method were compared to the ones calculated from the asymptotic chi-squared distribution and the permutation test. When the samples are dependent, the simulated significance level was found to be conservative or strongly conservative. Under different alternatives, the simulated power was found to be close to one when the sample sizes increased and a larger difference in the means was considered. The advantages of using the logarithmic quantile estimation method is two fold. It allows us to estimate quantiles using only the observations in the sample and not using the asymptotic distribution of the test statistic. When the samples are dependent, the asymptotic distribution of the Kruskal-Wallis test requires information on the correlation of the different samples which must be estimated. In this case the quantile estimation performs well without the estimation of any correlation or covariance matrix.

Another application that we developed was based on the "shoulder tip pain study". This clinical data set was analyzed as a nonparametric factorial design where the hypotheses were stated in terms of distribution functions. The test statistic for each hypothesis was given by a rank version of the ANOVA type statistic. We showed that an almost sure limit theorem holds for each test statistic and the p-values were obtained using the logarithmic empirical quantile procedure. The logarithmic quantile estimation can be applied to these designs once they are modeled in the form of an almost sure limit theorem. While in the Kruskal-Wallis example the samples were equal, the shoulder tip pain study contains four groups of subjects with unequal sample sizes. This data set is an example of an unbalanced design that can be treated in the context of the logarithmic quantile procedure. It is an open question how this method performs for designs with more complicated structures than the shoulder study. The existing methods of analyzing these type of designs require a modification of the ANOVA statistic such that its asymptotic distribution can be approximated by an F distribution. This modification of the test statistic involves estimation of a covariance matrix. When using the logarithmic empirical quantile method, we do not need to estimate covariance matrices or to know the asymptotic distribution of the test statistic.

In general, the logarithmic quantile estimation is a new method of estimating quantiles based only on the observations that are in the sample and not using the asymptotic distribution of the test statistic. The advantage of this method is that it avoids estimation of covariance matrices, correlation and variances.

7.2 Open Problems

One possibility of extending the present research is to investigate 'robustness' issues for the procedure. For rank statistics this means that independence is relaxed, and leads one to consider the random variables under mixing assumptions. This involves some distributional limit theory for weakly dependent processes. There is some work known by Peligrad and Shao [83], but there are many outstanding issues.

A second area of problems that one can address is the practical implementation of the new method. The method is so far reliable, but needs to be investigated under different designs. Nonparametric ANOVA models are a next step to investigate, beyond the c-sample problem. Another important class is longitudinal data. A natural way to proceed is similar to the book by Brunner, Domhof and Langer [21].

Regression problems can be treated by logarithmic quantile estimation as well. Since many results are a direct consequence of a central limit theorem, its generalization to almost sure versions seems a reasonable next step. For correlation analysis, some results are known, but the general theory is not yet developed.

A fourth area of research concerns Poisson limit distribution. There is no almost sure version of this result. It is not clear which form it will take, but it may turn out to be useful in other areas of mathematics, like dynamical systems. It may also be connected to an almost sure local limit theorem, the analogue for a normal distribution would read as an almost sure version of the classical deMoivre theorem as in Denker and Koch [38].

Bibliography

[1] Akritas, M. G., Arnold, S. F. (1994), Fully nonparametric hypotheses for factorial designs I: multivariate repeated measures designs, *Journal of the American Statistical Association*, 89, 336-343.

[2] Akritas, M. G., Arnold, S. F., Brunner, E. (1997), Nonparametric hypotheses and rank statistics for unbalanced factorial designs, *Journal of the American Statistical Association*, 92, 258–265.

[3] Akritas, M. G., Brunner, E. (1997), A unified approach to rank tests for mixed models, *Journal of Statistical Planning and Inference*, 61, 249-277.

[4] Akritas, M. G. (1990), The rank transform method in some two-factor designs, *Journal of the American Statistical Association*, Vol. 85, No. 409, 73-78.

[5] Akritas, M. G. (1991), Limitations of the rank transform procedure: a study of repeated measures designs, part I, *Journal of the American Statistical Association*, 86:414, 457-460.

[6] Akritas, M. G. (1993), Limitations of the rank transform procedure: a study of repeated measures designs, part II, *Statistics* & *Probability Letters*, 17, 149-156.

[7] Babu, G. J., Padmanabhan, A. R. (2002), Re-sampling methods for the nonparametric Behrens-Fisher problem, *Sankhy*ā*: The Indian Journal of Statistics, Series A*, 64, 678-692.

[8] Bagdonavičius, V., Kruopis, J., Nikulin, M. (2011), *Non-parametric tests for complete data*, John Wiley Sons, Inc.

[9] Bercu, B., Cénac, P., Fayolle, G. (2009), On the almost sure central limit theorem for vector martingales: convergence of moments and statistical applications, *Journal of Applied Probability* 46, 151-169.

[10] Berkes, I. (1998), Results and problems related to the pointwise central limit theorem, *Asymptotic Methods in Probability and Statistics*, B. Szyszkowicz (Editor), 59-96.

[11] Berkes, I., Csáki, E. (2001), A universal result in almost sure central limit theory, *Stochastic Processes and their Applications*, 94, 105-134.

[12] Berkes, I., Dehling, H. (1993), Some limit theorems in log density, *The Annals of Probability*, 21, 1640-1670.

[13] Berkes, I., Dehling, H. (1994), On the almost sure (a.s.) central limit theorem for random variables with infinite variance, *Journal of Theoretical Probability*, Vol. 7, No. 3, 667-680.

[14] Berkes, I., Weber, M. (2006), Almost sure versions of the Darling-Erdős theorem, *Statistics* & *Probability Letters* 76, 280-290.

[15] Billingsley, P. (1999), *Convergence of probability measures*, second edition, John Willey and Sons, Inc.

[16] Boos, D. D. (1986), Comparing K populations with linear rank statistics, *Journal of the American Statistical Association*, Vol. 81, No. 396, 1018-1025.

[17] Box, G. E. (1954), Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification, *The Annals of Mathematical Statistics*, 25, 290-302.

[18] Bradley, J. V. (1968), *Distribution-free statistical tests*, Prentice-Hall, Inc.

[19] Brosamler, G. A. (1988), An almost everywhere central limit theorem, *Mathematical Proceedings of the Cambridge Philosophical Society*, 104, 561-574.

[20] Brunner, E., Denker, M. (1994), Rank statistics under dependent observations and applications to factorial designs, *Journal of Statistical Planning and Inference*, 42, 353-378.

[21] Brunner, E., Domhof, S., Langer, F. (2002), *Nonparametric Analysis of Longitudinal Data in Factorial Experiments*, Wiley Series in Probability and Statistics.

[22] Brunner, E., Munzel,U. (2000), The nonparametric Behrens-Fisher problem: asymptotic theory and a small-sample approximation, *Biometrical Journal*, 42, 17-23.

[23] Brunner, E., Puri, M. L. (1996), Nonparametric methods in design and analysis of experiments, *Handbook of Statistics* Vol. 13, Elsevier Science B. V., Amsterdam, 631-703.

[24] Brunner, E., Puri, M. L. (2001), Nonparametric methods in factorial designs, *Statistical Papers* 42, 1-52.

[25] Brunner, E., Puri, M. L. (2002), A class of rank-score tests in factorial designs, *Journal of Statistical Planning and Inference*, 103, 331-360.

[26] Brunner, E., Dette, H., Munk, A. (1997), Box-type approximation in nonparametric factorial designs, *Journal of the American Statistical Association* 92, 1494-1502.

[27] Brunner, E., Munzel, U., Puri, M. L. (1999), Rank-score tests in factorial designs with repeated measures, *Journal of Multivariate Analysis* 70, 286-317.

[28] Brunner, E., Neumann, N. (1982), Rank tests for correlated random variables, *Biometrical Journal*, Vol. 24, No. 4, 373-389.

[29] Cheng, S., Peng, L., Qi, Y. (1998), Almost sure convergence in extreme value theory, *Math. Nachr.* 190, 43-50.

[30] Chernoff, H., Savage, I. R., (1958), Asymptotic normality and efficiency of certain nonparametric test statistics, *The Annals of Mathematical Statistics*, Vol. 29, No. 4, 972-994.

[31] Chuprunov, A., Fazekas, I. (2002), Almost sure versions of some functional limit theorems, *Journal of Mathematical Sciences*, Vol. 111, No. 3, 3528-3536.

[32] Chuprunov, A., Fazekas, I. (2004), Almost sure limit theorems for the Pearson statistic, *Theory Probab. Appl.*, 48, 14-147.

[33] Compagnone, D., Denker, M. (1996), Nonparametric tests for scale and location, *Journal of Nonparametric Statistics* 7, 123-154.

[34] Conover, W. J., Ronald, L. Iman (1981), Rank transform as a bridge between parametric and nonparametric statistics, *The American Statistician*, 35:3, 124-129.

[35] Denker, M., Rösler, U. (1985), Some contributions to Chernoff-Savage theorems, *Statistics & Decisions* 3, 49-75.

[36] Denker, M. (1985), *Asymptotic Distribution Theory in Nonparametric Statistics*, Vieweg Advanced Lectures in Mathematics.

[37] Denker, M., Fridline, M. (2010), The almost sure version of Cramer's theorem, *Dependence in probability, analysis and number theory*, Kendrick Press, Heber City, UT, 195-201.

[38] Denker, M., Koch, S. (2002), Almost sure local limit theorems, *Statistica Neerlandica*, Vol. 56, No. 2, 143-151.

[39] Denker, M., Tabacu, L. (2014), Logarithmic quantile estimation for rank statistics, *Journal of Statistical Theory and Practice*, DOI: 10.1080/15598608.2014.886312

[40] Denker, M., Tabacu, L. (2013), Testing longitudinal data by logarithmic quantiles, Preprint.

[41] Devroye, L. (1986), *Non-uniform random variate generation*, Springer-Verlag.

[42] Dudley, R. M. (2002), *Real analysis and probability*, Cambridge University Press.

[43] Dvoretzky, A., Kiefer, J., Wolfowitz, J. (1956), Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator, *The Annals of Mathematical Statistics*, 27, 642-669.

[44] Efron, B. (1979), Bootstrap methods: another look at the jackknife, *The Annals of Statistics*, 7, 1-26.

[45] Erdös, P., Hunt, G. A. (1953), Changes of signs of sums of random variables, *Pacific Journal of Mathematics* 3, 673-687.

[46] Fahrner, I., Stadtmüller, U. (1998), On almost sure max-limit theorems, Statistics & Probability Letters 37, 229-236.

[47] Fridline, M. (2009), Almost sure confidence intervals for the correlation coefficient, *Ph.D. thesis*, Case Western Reserve University, 2009.

[48] Friedman, M. (1937), The use of ranks to avoid the assumption of normality implicit in the analysis of variance, *Journal of the American Statistical Association*, Vol. 32, No. 200, 675-701.

[49] Gentle, J. (2003), *Random number generation and Monte Carlo methods*, Springer.

[50] Govindarajulu, Z., LeCam, L., Raghavachari, M. (1966), Generalizations of theorems of Chernoff and Savage on the asymptotic normality of test statistics, *Proc. Fifth Berkeley Symp. Math. Statit. Prob.* 1, 609-638.

[51] Govindarajulu, Z. (1975), Robustness of Mann-Whitney-Wilcoxon test to dependence in the variables, *Studia Scientiarum Mathematicarum Hungarica* 10, 39-45.

[52] Hájek, J. (1968), Asymptotic normality of simple linear rank statistics under alternatives, *The Annals of Mathematical Statistics*, Vol. 39, No. 2, 325-346.

[53] Hájek, J., Sidak, Z., Sen, P. (1999), *Theory of rank tests*, Academic Press, Second Edition.

[54] Hoeffding, W. (1973), On the centering of a simple linear rank statistic, *The Annals of Statistics*, Vol. 1, No. 1, 54-66.

[55] Holzmann, H., Koch, S. and Min, A. (2004), Almost sure limit theorems for U-statistics, *Statistics & Probability Letters*, 69, 261-269.

[56] Hora, S. C., Conover, W. J. (1984), The F statistic in the two-way layout with rank-score transformed data, *Journal of the American Statistical Association*, Vol. 79, No. 387, 668-673.

[57] Hörmann, S. (2007), Critical behavior in almost sure central limit theory, *Journal of Theoretical Probability*, Vol. 20, Issue 3, 613-636.

[58] Hörmann, S. (2005), Optimal averaging procedures in almost sure central limit theory, *Metodološki zvezki*, Vol. 2, No. 2, 271-282.

[59] Hotelling, H., Pabst, M. R. (1936), Rank correlation and tests of significance involving no assumptions of normality, *The Annals of Mathematical Statistics*, Vol. 7, No. 1, 29-43.

[60] Hurelbaatar, G. (1995), A strong approximation for logarithmic averages of partial sums of random variables, *Periodica Mathematics Hungarica*, Vol 31 (3), 189-200.

[61] Hurelbaatar, G. (1997), Almost sure limit theorems for dependent random variables, *Studia Scientiarum Mathematicarum Hungarica*, Vol. 33, Issue 1-3, 167-175.

[62] Ibragimov, I. A., Lifshits, M. (1999), On almost sure limit theorems, *Theory Probab. Appl.*, 44, 254-272.

[63] Kendall, M. G. (1938), A new measure of rank correlation, *Biometrika* 30, 81-93.

[64] Klenke, A. (2008), *Probability Theory-A Comprehensive Course*, Springer.

[65] Kolmogorov, A. N. (1933), Sulla determinizione empirica di una legge di distribuzione, *Giorn. Ist. Ital. Attuari*, Vol. 4, 83-91.

[66] Koroljuk, V. S., Borovskich, Yu. V. (1994), *Theory of U-Statistics*, Kluwer Academic Publishers.

[67] Kruskal, W. H. (1952), A nonparametric test for the several sample problem, *The Annals of Mathematical Statistics*, Vol. 23, No. 4, 525-540.

[68] Kruskal, W. H., Wallis, W. A. (1952), Use of ranks in one-criterion variance analysis, *Journal of the American Statistical Association*, Vol. 47, No. 260, 583-621.

[69] Lacey, M. T., Philipp, W. (1990), A note on the almost sure central limit theorem, *Statistics & Probability Letters*, 9, 201-205.

[70] Lee, A. J. (1990), *U-Statistics: Theory and Practice*, Dekker.

[71] Lehmann, E. L., Romano, J. P. (2005), *Testing statistical hypotheses*, Springer, Third Edition.

[72] Lévy, P. (1937), *Théorie de l'addition des variables aléatoires*, Gauthier-Villars, Paris.

[73] Lifshits, M. A. (2002), The almost sure limit theorem for sums of random vectors, *Journal of Mathematical Sciences*, Vol. 109, No. 6, 2166-2178.

[74] Lifshits, M. A. (2001), Lecture notes on almost sure limit theorems, *Publications IRMA, Lille*, 54, No. 8, 1-23.

[75] Lumley, T. (1996), Generalized estimation equations for ordinal data: A note on working correlation structures, *Biometrics* 52, 354-361.

[76] Mann, H. B., Whitney, D. R. (1947), On a test of whether one of two random variables is stochastically larger than the other, *The Annals of Mathematical Statistics*, Vol. 18, No. 1, 50-60.

[77] Mathai, A. M., Provost, S. B. (1992), *Quadratic forms in random variables: theory and applications*, Marcel Dekker, Inc.

[78] Matula, P. (2005), On almost sure limit theorems for positively dependent random variables,*Statistics & Probability Letters* 74, 59-66.

[79] Munzel, U. (1999), Linear rank score statistics when ties are present, *Statistics & Probability Letters*, 41, 389-395.

[80] Nation, J. R., Bourgeois, A. E., Clark, D. E., Baker, D. M., Hare, M. F. (1984), The effects of oral cadmium exposure on passive avoidance performance in the adult rat, *Toxicology Letters* 20, 41-47.

[81] Neuhäuser, M. (2012), *Nonparametric statistical tests: a computational approach*, CRC Press, Taylor & Francis Group.

[82] Peligrad, M., Révész, P. (1991), On the almost sure central limit theorem, *Almost Everywhere Convergence II*, A. Bellow and R. Jones, eds., Academic Press, New York, 209-225.

[83] Peligrad, M. and Shao, Q. M. (1995), A note on the almost sure central limit theorem for weakly dependent random variables, *Statistics & Probability Letters* 22, 131-136.

[84] Peng, Z., Li, J., Nadarajah, S. (2009), Almost sure convergence of extreme order statistics, *Electronic Journal of Statistics*, Vol. 3, 546-556.

[85] Puri, M. L. (1969), Nonparametric techniques in statistical inference, *Proceedings of the first international symposium on nonparametric techniques held at Indiana University*, June 1969, Edited by Puri, M. L.

[86] Puri, M. L., Sen, P. K. (1971), *Nonparametric methods in multivariate analysis*, John Wiley & Sons, Inc.

[87] Puri, M. L., Sen, P. K. (1985) *Nonparametric methods in general linear models*, John Willey & Sons. Inc.

[88] Reiczigel, J., Zakariàs, I., Rózsa, L. (2005), A bootstrap test of stochastic equality of two populations, *The American Statistician*, 59, 156-161.

[89] Ruymgaart, F. H. (1980), A unified approach to the asymptotic distribution theory of certain midrank statistics, *Statistique non Parametrique Asymptotique*, J.P. Raoult (Editor), Lecture Notes on Mathematics, No. 821, Springer, 1-18.

[90] Ruymgaart, F. H., van Zuijlen, M. C. A. (1978), Asymptotic normality of multivariate linear rank statistics in the non-i.i.d. case, *The Annals of Statistics*, Vol. 6, No. 3, 588-602.

[91] Schatte, P. (1988), On strong versions of the central limit theorem, *Mathematische Nachrichten*, 137, 249-256.

[92] Serfling, R. (1980), *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

[93] Singh, R. S. (1975), On the Glivenko-Cantelli theorem for weighted empiricals based on independent random variables, *The Annals of Probability*, 3, 371-374.

[94] Smirnov, N. V. (1939), On the estimation of the discrepancy between empirical curves of distribution for two independent samples, (Russian) *Bull. Moscow Univ.* 2, 3-16.

[95] Stadtmüller, U. (2002), Almost sure versions of distributional limit theorems for certain order statistics, *Statistics* & *Probability Letters* 58, 413-426.

[96] Steland, A. (1998), Bootstrapping rank statistics, *Metrika* 47, 251-264.

[97] Thangavelu, K. (2005), Quantile estimation based on the almost sure central limit theorem, *Ph.D. thesis*, Göttingen University.

[98] Thompson, G. L. (1990), Asymptotic distribution of rank statistics under dependency with multivariate application, *Journal of Multivariate Analysis*, Vol. 33, 183-211.

[99] Thompson, G. L. (1991), A unified approach to rank tests for multivariate and repeated measures designs, *Journal of the American Statistical Association*, Vol. 86, No. 414, 410-419.

[100] Wald, A., Wolfowitz, J. (1940), On a test whether two samples are from the same population, *The Annals of Mathematical Statistics*, Vol. 11, No. 2, 147-162.

[101] Wilcoxon, F. (1945), Individual comparisons by ranking methods, *Biometrics*, Vol. 1, No. 6, 80-83.

[102] Zhongquan, T., Zuoxiang, P., Nadarajah, S. (2007), Almost sure convergence of sample range, *Extremes* 10, 225-233.

[103] Zöfel, P. (1992), *Univariate Varianzanalysen*. Uni-Taschenbuch 1663, G. Fischer Verlag, Jena, Stuttgart.

Vita

Lucia Tabacu

Research Interests: nonparametric statistics: testing procedures, longitudinal data, and factorial designs; almost sure limit theory; mathematical statistics

Education

- **Doctor of Philosophy, Statistics**
  The Pennsylvania State University, August 2014
  Adviser: Professor Manfred Denker

- **Master of Science, Mathematics**
  University of Bucharest, Romania, February 2007
  Concentration: stochastic calculus, applied statistics and optimization

- **Bachelor of Science, Mathematics**
  University of Bucharest, Romania, June 2005

Experience

- **Research Assistant** , Penn State University, Fall 2012-Spring 2013
  NSF grant "Probabilistic Analysis in Dynamical Systems" – quantile estimation and almost sure limit theory (under supervision of Professor Manfred Denker)

- **Research Assistant**, Penn State University, Summer 2012
  Stochastic modeling and computation related to systems of protein-based molecular motors (under supervision of Professor John Fricks)

- **Instructor**, Penn State University, Fall 2011-Spring 2012
  Elementary Probability (Math/Stat 318)
  Applied Statistics in Science (Math/Stat 319)

- **Teaching Assistant**, Penn State University, Fall 2008-Summer 2011, Fall 2013-Summer 2014

- **Consultant**, Statistical Consulting Center, Penn State University, Spring 2010-Summer 2010

- **Part-time consultant**, VMB Partners, and **private tutor in mathematics**, Bucharest, Romania, March 2007- February 2008

- **Assistant Statistician**, National Institute of Statistics, Bucharest, Romania, March-June 2008