MATHEMATICAL CENTRE TRACTS 151

HIGHER ORDER ASYMPTOTICS FOR SIMPLE LINEAR RANK STATISTICS

R.J.M.M. DOES

MATHEMATISCH CENTRUM AMSTERDAM 1982

1980 Mathematics subject classification: Primary: 62E20
Secondary: 60F05, 60E10, 62G10, 62G20

ISBN 90 6196 243 9

CONTENTS

PREFACE

Except for a few typing errors this tract is a copy of the author's thesis. The research for it has been carried out at the Department of Mathematical Statistics of the Mathematical Centre at Amsterdam and completed at the Data Processing Department of the University of Limburg at Maastricht.

It is a pleasure to acknowledge the many helpful discussions with my thesis advisor Professor W.R. van Zwet; I greatly appreciated his encouragement and constructive criticism during this study.

I am grateful to the Mathematical Centre for publishing this monograph in the present form.

CHAPTER 1

INTRODUCTION

1.1. HIGHER ORDER ASYMPTOTICS AND SIMPLE LINEAR RANK STATISTICS

As higher order asymptotics is the topic of this study, we shall start by providing a rough outline of this area. For reasons of simplicity we restrict ourselves at this point to the classical case of sums of independent and identically distributed random variables. For a complete and much more eloquent account one should consult e.g. FELLER (1971) Chapter XVI, PETROV (1975) Chapters V and VI and for the multidimensional case BHATTACHARYA & RAO (1976).

According to the central limit theorem standardized sums of independent and identically distributed random variables with finite variance are asymptotically normally distributed. The first question that is discussed in second order asymptotics is that of the rate of convergence to this normal limit.

<u>THEOREM 1.1.1</u>. *Let* $Y_1, Y_2, \ldots$ *be independent and identically distributed random variables such that* $EY_1 = 0$, $EY_1^2 = 1$ *and* $E|Y_1|^3 = \rho < \infty$. *Furthermore let* F_N *stand for the distribution function of the normalized sum* $N^{-\frac{1}{2}} \Sigma_{j=1}^{N} Y_j$. *Then*

$$(1.1.1) \qquad \sup_{x \in \mathbb{R}} |F_N(x) - \Phi(x)| \leq \frac{A\rho}{\sqrt{N}} ,$$

where A *is an universal constant and* Φ *denotes the standard normal distribution function.*

<u>PROOF</u>. See e.g. PETROV (1975), Theorem V 2.4. □

Theorem 1.1.1 is called the Berry-Esseen theorem for sums of independent and identically distributed random variables. It follows that a Berry-Esseen bound is of the order $O(N^{-\frac{1}{2}})$.

The logical next step is to go beyond limit theorems and investigate higher order terms of the distribution functions of asymptotically normal sums of random variables. For the special case of sums of independent and identically distributed random variables we may follow the account given in Section 2 of VAN ZWET (1977).

Let $Y_1, Y_2, \ldots$ be independent and identically distributed random variables with $EY_1 = 0$, $EY_1^2 = 1$, $EY_1^3 = \mu_3$ and $EY_1^4 = \mu_4$. Furthermore, let F and ψ_1 denote the distribution function and the characteristic function of Y_1 respectively. Moreover, we assume that Cramér's Condition (C) is satisfied; i.e.

$$\limsup_{|t|\to\infty} |\psi_1(t)| < 1. \tag{1.1.2}$$

We remark that (1.1.2) is satisfied if F has an absolutely continuous component. Define

$$F_N(x) = P(N^{-\frac{1}{2}} \sum_{j=1}^{N} Y_j \le x), \qquad \text{for } -\infty < x < \infty.$$

<u>THEOREM 1.1.2</u>. *Suppose that* $EY_1^4 = \mu_4 < \infty$ *and* (1.1.2) *are satisfied. Then*

$$\sup_{x\in\mathbb{R}} |F_N(x) - \tilde{F}_N(x)| = o(N^{-1}) \qquad \text{as } N \to \infty, \tag{1.1.3}$$

where

$$\tilde{F}_N(x) = \Phi(x) - \phi(x)\left\{\frac{\mu_3}{6\sqrt{N}}(x^2-1) + \frac{(\mu_4-3)}{24N}(x^3-3x) + \frac{\mu_3^2}{72N}(x^5 - 10x^3 + 15x)\right\}, \tag{1.1.4}$$

and ϕ *denotes the standard normal density.*

A result like (1.1.3) is called an Edgeworth expansion with remainder $o(N^{-1})$ for the distribution function F_N. Note that $\mu_3 N^{-\frac{1}{2}}$ and $(\mu_4-3)N^{-1}$ are the third and fourth cumulants of $N^{-\frac{1}{2}} \Sigma_{j=1}^N Y_j$ respectively. The proof of Theorem 1.1.3 is based on an application of the smoothing lemma (cf. ESSEEN (1945)).

LEMMA 1.1.3. (Smoothing lemma). *Let* M *be a positive number,* F *a distribution function on* $\mathbb{R}$ *and* $\tilde{F}$ *a differentiable function of bounded variation on* $\mathbb{R}$ *with* $\tilde{F}(-\infty) = 0$, $\tilde{F}(\infty) = 1$ *and* $|\tilde{F}'| \le M$. *Define the Fourier-Stieltjes transforms* $\psi(t) = \int \exp(itx)dF(x)$ *and* $\tilde{\psi}(t) = \int \exp(itx)d\tilde{F}(x)$. *Then there exists a constant* C *such that for every* $T > 0$

$$(1.1.5) \qquad \sup_{x\in\mathbb{R}} |F(x) - \tilde{F}(x)| \le \frac{1}{\pi}\int_{-T}^{T} \left|\frac{\psi(t)-\tilde{\psi}(t)}{t}\right| dt + \frac{CM}{T}.$$

With this in mind, the proof now proceeds as follows. Let ψ_N denote the characteristic function of $N^{-\frac{1}{2}} \Sigma_{j=1}^{N} Y_j$ and choose $T = N \log N$. It follows from the fact that $\mu_4 < \infty$, that for $N \to \infty$ and $|t| = o(N^{\frac{1}{2}})$,

$$\log \psi_N(t) = -\tfrac{1}{2}t^2 - \frac{i\mu_3}{6\sqrt{N}} t^3 + \frac{(\mu_4-3)}{24N} t^4 + o(\frac{t^4}{N})$$

and this expansion is easily converted into

$$\psi_N(t) = \tilde{\psi}_N(t) + o(N^{-1}t^4e^{-\frac{1}{4}t^2}),$$

where

$$(1.1.6) \qquad \tilde{\psi}_N(t) = e^{-\frac{1}{2}t^2}\left\{1 - \frac{i\mu_3}{6\sqrt{N}} t^3 + \frac{(\mu_4-3)}{24N} t^4 - \frac{\mu_3^2}{72N} t^6\right\}.$$

For any sufficiently small $\delta > 0$ this expansion for $\psi_N(t)$ remains valid for $|t| \le \delta\sqrt{N}$ because

$$|\psi_N(t)| \le \{1 - \frac{t^2}{3N}\}^N \le e^{-t^2/3},$$

for $|t| \le \delta\sqrt{N}$. Hence

$$\int_{-\delta\sqrt{N}}^{\delta\sqrt{N}} \left|\frac{\psi_N(t)-\tilde{\psi}_N(t)}{t}\right| dt = o(N^{-1}).$$

From the definition of $\tilde{\psi}_N$ in (1.1.6) and $\mu_4 < \infty$ it follows that

$$\int_{|t|\ge\delta\sqrt{N}} \left|\frac{\tilde{\psi}_N(t)}{t}\right| dt = o(N^{-1}).$$

Finally, Cramér's Condition (C) (cf. (1.1.2)) guarantees that

$$\int_{\delta\sqrt{N}\le|t|\le N\log N} \left|\frac{\psi_N(t)}{t}\right| dt = o(N^{-1}).$$

Since $\tilde{\psi}_N$ is the Fourier-Stieltjes transform of $\tilde{F}_N$ (cf. (1.1.4)) the proof of Theorem 1.1.2 is complete.

In recent years much effort has been devoted to extending the theory of higher order asymptotics from sums of independent and identically distributed random variables to the estimators and test statistics that interest statisticians. In this study we deal with the problem of obtaining Berry-Esseen bounds and Edgeworth expansions for simple linear rank statistics.

Let $X_1,X_2,\ldots,X_N$ be independent random variables with continuous distribution functions $F_1,F_2,\ldots,F_N$ respectively. If $X_{1:N} < X_{2:N} < \ldots < X_{N:N}$ denotes the sequence $X_1,X_2,\ldots,X_N$ arranged in increasing order then the rank R_{jN} of X_j is defined by $X_j = X_{R_{jN}:N}$, $j = 1,2,\ldots,N$. For specified vectors of real numbers $c_N = (c_{1N},c_{2N},\ldots,c_{NN})$ (regression constants) and $a_N = (a_{1N},a_{2N},\ldots,a_{NN})$ (scores)

$$(1.1.7) \qquad T_N = \sum_{j=1}^{N} c_{jN} a_{R_{jN}N}$$

is called a simple linear rank statistic. It may be used for testing the null-hypothesis H_0: $F_1 = F_2 = \ldots = F_N$ against certain classes of alternatives indicated by the choice of regression constants and scores. This test is distributionfree and under the null-hypothesis the random vector $(R_{1N}, R_{2N},\ldots,R_{NN})$ equals each permutation of the numbers $1,2,\ldots,N$ with probability $1/N!$.

As an example, we consider the problem of regression in location. Take a random sample of size N of independent observations $X_1,X_2,\ldots,X_N$, where X_j has density f_j, $j = 1,2,\ldots,N$. The null-hypothesis H_0 is $f_1 = f_2 = \ldots = f_N = f$, where the common density f is arbitrary. The one-sided alternative of interest will be $f_j(x) = f(x-\Delta c_j)$ for $j = 1,2,\ldots,N$ with $\Delta > 0$. If we assume that f is absolutely continuous with derivative f' such that $\int|f'| < \infty$ and define

$$(1.1.8) \qquad a_{jN} = E\left\{-\frac{f'(X_{j:N})}{f(X_{j:N})}\right\},$$

then the test with critical region

$$\sum_{j=1}^{N} c_j a_{R_{jN}N} \geq k$$

is the locally most powerful rank test for H_0 against $\{\Pi f(x-\Delta c_j), \Delta > 0\}$ at the level of significance determined by the choice of k.

More generally, for almost all well-known linear rank tests, the scores are generated by a function J on (0,1) in either one of the following two ways

(1.1.9) (exact scores) $a_{jN} = EJ(U_{j:N}), \quad j = 1,2,\ldots,N,$

(1.1.10) (approximate scores) $a_{jN} = J(\frac{j}{N+1}), \quad j = 1,2,\ldots,N.$

Here $U_{j:N}$ denotes the j-th order statistic in a random sample of size N from the uniform distribution on (0,1). Scores given by (1.1.9) occur in statistics yielding locally most powerful rank tests (cf. (1.1.8)). Those given by (1.1.10) have the appeal of simplicity.

Well-known special cases of (1.1.7) are linear rank statistics for the two-sample problem, where $F_j = F$, $c_{jN} = 0$, for $j = 1,2,\ldots,n$ and $F_j = G$, $c_{jN} = 1$, for $j = n+1,\ldots,N$ and Spearman's rank correlation coefficient ρ_N which, under the null-hypothesis of independence, is distributed as T_N under H_0 with $c_{jN} = a_{jN} = j$ for $j = 1,2,\ldots,N$.

The distribution of a simple linear rank statistic is determined by the following three entities: first, the distribution functions $F_1,F_2,\ldots$ $\ldots,F_N$ of the observations $X_1,X_2,\ldots,X_N$, second, the regression constants $c_{1N},c_{2N},\ldots,c_{NN}$ and finally, the scores $a_{1N},a_{2N},\ldots,a_{NN}$.

Define, for each $N \geq 2$,

(1.1.11) $T_N^* = \dfrac{T_N - ET_N}{\sigma(T_N)}$,

(1.1.12) $F_N^*(x) = P(T_N^* \leq x), \quad \text{for } -\infty < x < \infty.$

In asymptotic theory one studies the behavior of the distribution function F_N^* for large values of N both under the null-hypothesis as well as under various types of alternatives.

Many authors have established the asymptotic normality of T_N^* under different sets of conditions (see Section 1.2 for a review); i.e.

(1.1.13) $$\lim_{N\to\infty} \sup_{x\in\mathbb{R}} |F_N^*(x) - \Phi(x)| = 0,$$

where Φ denotes the standard normal distribution function. This first order result justifies the use of the normal approximation to compute the critical value and the power of a simple linear rank test for large values of the sample size N. Furthermore, it enables us to find the limiting power of the test under contiguous alternatives and to make a comparison with other tests on that basis.

The next question is to obtain the rate of convergence in (1.1.13). Under some regularity conditions this will typically be of the order $N^{-\frac{1}{2}}$; i.e.

$$\sup_{x\in\mathbb{R}} |F_N^*(x) - \Phi(x)| = \mathcal{O}(N^{-\frac{1}{2}}).$$

Quite often, however, one needs more precise information than asymptotic normality can provide. To achieve this one needs an asymptotic expansion for F_N^* with a uniform remainder of order $o(N^{-1})$. Such expansions will typically be Edgeworth expansions; i.e. they are of the type

$$\tilde{F}_N(x) = \Phi(x) - \phi(x)\{\frac{\kappa_{3N}}{6}(x^2-1) + \frac{\kappa_{4N}}{24}(x^3-3x) + \frac{\kappa_{3N}^2}{72}(x^5-10x^3+15x)\},$$

where ϕ denotes the standard normal density, the quantities κ_{3N} and κ_{4N} are the leading terms in asymptotic expansions for the third and fourth cumulants of T_N^* (cf. (1.1.11)). To establish such an expansion one has to compute $\tilde{F}_N$ and prove

$$\sup_{x\in\mathbb{R}} |F_N^*(x) - \tilde{F}_N(x)| = o(N^{-1}), \quad \text{as } N \to \infty.$$

An excellent review of the developments in the area of Edgeworth expansions in nonparametric statistics was given by BICKEL (1974). He notes that these expansions are of interest on various grounds:

(1) Taking one of the two terms of the expansion frequently improves the basic normal approximation strikingly. Examples of this phenomenon may be found in FIX & HODGES (1955) and in Section 3.5 of the present study.

(2) The higher order terms give some qualitative insight into the regions of unreliability of first order results. For instance, the third and fourth cumulants typically correct for skewness and kurtosis.

(3) The expansions can be used to discriminate between procedures equivalent to first order. For a complete account one should consult HODGES & LEHMANN (1970).

(4) Last but not least, the probabilistic problems involved are very challenging.

The organization of this study is as follows. In this chapter we review the literature on simple linear rank statistics. Furthermore, we formulate a lemma which will be a basic tool in the Chapters 2 through 4. Chapter 2 is devoted to the problem of establishing Berry-Esseen theorems for simple linear rank statistics under the null-hypothesis. In Chapter 3 we establish Edgeworth expansions for these statistics under the null-hypothesis, whereas Chapter 4 deals with asymptotic expansions under contiguous alternatives.

1.2. RELATION TO PREVIOUS WORK

The concept of a simple linear rank statistic is essential in nonparametric theory. For an introduction to this subject the reader is referred to LEHMANN (1975). More advanced text books were written by HÁJEK & ŠIDÁK (1967) and PURI & SEN (1971). The latter deals with multivariate problems.

The first central problem concerning simple linear rank statistics is their asymptotic normality. This problem was discussed in increasing generality in long series of papers beginning with HOTELLING & PABST (1936). They prove asymptotic normality for Spearman's rank correlation coefficient under the null-hypothesis of independence. Their work was generalized to statistics of the form (1.1.7) by WALD & WOLFOWITZ (1944). The method of proof consists of showing that the moments of T_N^* (cf. (1.1.11)) converge to those of the standard normal distribution. NOETHER (1949) showed that the condition on the scores could be weakened. In a formulation later given by HOEFFDING (1951) Noether's condition is

$$(1.2.1) \qquad \lim_{N\to\infty} \frac{\max_{1\le j\le N}(a_{jN}-\bar{a}_N)^2}{\Sigma_{j=1}^N (a_{jN}-\bar{a}_N)^2} = 0,$$

where $\bar{a}_N = N^{-1}\,\Sigma_{j=1}^N\, a_{jN}$. The series of every more general proofs of asymptotic normality of T_N^* under the null-hypothesis culminated in a paper by HÁJEK (1961) providing necessary and sufficient conditions. For exact scores his result reads as follows.

THEOREM 1.2.1. *Suppose that the regression constants satisfy Noether's condition (cf.* (1.2.1)*). Let* J(t), 0 < t < 1, *be square integrable such that* $\int\{J(t)-\bar{J}\}^2dt > 0$, *where* $\bar{J} = \int J(t)dt$. *Under the null-hypothesis*

$$(1.2.2) \qquad T_N = \sum_{j=1}^{N} c_{jN} a_{R_{jN}N}$$

is asymptotically normal (ET_N, σ_N^2), *where* $a_{jN} = EJ(U_{j:N})$ *(cf.* (1.1.9)*)*, $\sigma_N^2 = \Sigma_{j=1}^N (c_{jN}-\bar{c}_N)^2 \int\{J(t)-\bar{J}\}^2dt$ *and* $\bar{c}_N = N^{-1} \Sigma_{j=1}^N c_{jN}$.

Parallel to this development an attack was made on the related but somewhat easier problem of proving asymptotic normality for samples from a finite population (the two-sample case under the null-hypothesis). Among others we mention the papers of MADOW (1948) and DWASS (1955). For this problem, necessary and sufficient conditions were given by ERDÖS & RÉNYI (1959). They derive a representation for the characteristic function for random sampling without replacement which is not only useful to prove asymptotic normality but also for establishing Berry-Esseen bounds and Edgeworth expansions as will become clear in the sequel.

For more general statistics of the form $\Sigma_{j=1}^N d(j,R_{jN})$, where d(i,j), i,j = 1,2,...,N are N^2 real numbers, sufficient conditions for asymptotic normality under the null-hypothesis were established in HOEFFDING (1951). MOTOO (1957) proved asymptotic normality for these statistics under a Lindeberg type condition which appears to come close to being necessary. The paper of FRASER (1956) deals with a vector form of the Wald-Wolfowitz-Noether-Hoeffding theorem.

Under the alternative the problem of asymptotic normality is much more difficult to handle. The first attempt to treat the asymptotic distribution in the two-sample case in any generality was made by DWASS (1956). He was successful for a polynomial scores generating function J only (cf. (1.1.9), (1.1.10)). The first general paper for the two-sample case was written by CHERNOFF & SAVAGE (1958) and this work formed the basis for all further developments until 1968, excepting those based on the contiguity approach. Their work was generalized by GOVINDARAJULU, LECAM & RAGHAVACHARI (1967). By a totally different approach, based on properties of the empirical process, PYKE & SHORACK (1968) were able to give somewhat shorter proofs of the results in CHERNOFF & SAVAGE (1958) and to weaken the smoothness and tail behavior conditions on the scores generating function.

The line of development initiated by CHERNOFF & SAVAGE (1958) culminates in a paper of HÁJEK (1968). In this paper Hājek introduces the method of projection and a variance inequality. With these two tools he was able to prove asymptotic normality for simple linear rank statistics (cf. (1.1.7)) under fixed alternatives and for broad classes of regression constants and scores generating functions. Hājek's results were extended to scores generating functions which are not absolutely continuous by DUPAČ & HÁJEK (1969). HOEFFDING (1973) showed that centering at the asymptotic mean instead of the mean is permissible if one of Hājek's conditions is strengthened slightly. DUPAČ (1970) modified the theorems of his article with Hājek by allowing a different centering and generalized one of the results in HÁJEK (1968) to the case of unit step scores generating functions.

A different type of approach is to prove asymptotic normality for simple linear rank statistics under contiguous sequences of alternatives only. The concept of continguity is due to LECAM (1960) (see also HÁJEK & ŠIDÁK (1967), Chapter VI) and is basic for the most successful asymptotic theory of hypothesis testing. It ensures that the power does not tend to the significance level α or to one as the sample size tends to infinity. For these alternatives HÁJEK (1962) proved asymptotic normality in a situation where the regression constants satisfy Noether's condition (cf. (1.2.1)) and the scores generating function is square integrable. For a detailed discussion ans further results until the mid-sixties, the reader is referred to HÁJEK & ŠIDÁK (1967), Chapters V and VI.

The first result for the rate of convergence for a linear rank statistic was derived by STOKER (1954). For the Wilcoxon statistic he found that, under the null-hypothesis, the rate of convergence to the normal distribution is $O((m+n)^{-\frac{1}{2}+\varepsilon})$, where $\varepsilon > 0$ and m, n denote the sample sizes. For sampling from a finite population BIKELIS (1969) proved a Berry-Esseen bound of the order $O(N^{-\frac{1}{2}})$ using the representation of the characteristic function as given in ERDÖS & RÉNYI (1959). In a more general paper which covers the two-sample problem under the null-hypothesis as a special case von BAHR (1972) derived the same result. A general attempt to treat simple linear rank statistics was made by JUREČKOVÁ & PURI (1975). They consider approximate scores and show, both under the null-hypothesis and under certain contiguous location alternatives that the rate of convergence is of order $O(N^{-\frac{1}{2}+\varepsilon})$, for any $\varepsilon > 0$. Under the null-hypothesis their result is valid for scores generating functions having a bounded first derivative and under the alternative they assume the boundedness of the fourth derivative.

In two related papers von BAHR (1976) and HO & CHEN (1978) consider statistics of the form $\Sigma_{j=1}^{N} X(j,R_{jN})$ where $X = \{X(i,j): 1 \le i,j \le N\}$ is a square matrix of random variables with independent row vectors. These authors obtain as a by-product Berry-Esseen bounds of order $\mathcal{O}(N^{-\frac{1}{2}})$, under the null-hypothesis, for simple linear rank statistics with bounded scores and regression constants. This natural bound of order $\mathcal{O}(N^{-\frac{1}{2}})$ was also proved under the null-hypothesis, under certain contiguous alternatives as well as under certain fixed alternatives in HUŠKOVÁ (1977a, 1979a, 1979b). She considers bounded scores only, which excludes an important special case, namely the normal quantile function. Related results for bounded scores but without the optimal rate of convergence $\mathcal{O}(N^{-\frac{1}{2}})$ were proved in BERGSTRÖM & PURI (1977) and SERFLING (1980).

Under the null-hypothesis MASON (1981) obtains a rate of convergence of order $\mathcal{O}(N^{-\frac{1}{2}+\varepsilon})$, where $\varepsilon > 0$, when the scores generating function tends to infinity in the neighborhood of 0 and 1. In Chapter 2 we shall prove, that under the null-hypothesis, the natural bound of order $\mathcal{O}(N^{-\frac{1}{2}})$ holds even when the scores generating functions do not remain bounded. Our theorems include the normal quantile function, Φ^{-1} (see also DOES (1981a)). Finally, we note that in HUŠKOVÁ (1977b) rates of convergence of order $\mathcal{O}(N^{-\frac{1}{2}+\varepsilon})$ for the distribution functions of quadratic rank statistics to the chi-squared distribution are derived.

Turning now to Edgeworth expansions, we note that the first general result for rank statistics was obtained in the one-sample problem by ALBERS, BICKEL & VAN ZWET (1976). They establish the Edgeworth expansion under the hypothesis of symmetry as well as under contiguous location alternatives. Some extensions of this paper can be found in ALBERS (1974). The same programme was carried out for the two-sample problem in BICKEL & VAN ZWET (1978). Under the null-hypothesis the same result was proved by ROBINSON (1978). Recently ROBINSON (1980) has established an asymptotic expansion for rank tests for several samples. Extension of these results to the case of simple linear rank statistics remained an open problem (cf. BICKEL (1974) and HUŠKOVÁ (1977a)).

To establish Edgeworth expansions for simple linear rank statistics a bound on its characteristic function is obviously required to play the role of Cramér's Condition (C) in the case of sums of independent and identically distributed random variables (cf. (1.1.2)). This problem was solved by

VAN ZWET (1980). In Section 1.3 we reproduce a version of his theorem. In Chapter 3 we derive an Edgeworth expansion with remainder $o(N^{-1})$ for simple linear rank statistics under the null-hypothesis. The theorem is proved for a wide class of scores generating functions which includes the normal quantile function (see also DOES (1981b). In the last chapter we consider contiguous alternatives. Our theorem in this case is valid for bounded scores only. In a forthcoming paper (DOES (1982)) we shall establish asymptotic expansions under contiguous alternatives with a different standardization than is used in Chapter 4.

The reader may have noticed that we have not discussed U-statistics although some linear rank statistics are of this type. For those who are interested in these statistics an introduction is found in LEHMANN (1975). The best result concerning Berry-Esseen bounds is established in HELMERS & VAN ZWET (1981). An Edgeworth expansion can be found in CALLAERT, JANSSEN & VERAVERBEKE (1980).

1.3. A BASIC LEMMA AND SOME NOTATION

Let $X_1, X_2, \ldots, X_N$ be independent random variables with probability density functions $f_1, f_2, \ldots, f_N$ respectively. If $X_{1:N} < X_{2:N} < \ldots < X_{N:N}$ denotes the sequence of $X_1, X_2, \ldots, X_N$ arranged in increasing order, then the rank R_{jN} of X_j is defined by $X_j = X_{R_{jN}:N}$, $j = 1,2,\ldots,N$. For sequences of real numbers $c_{1N}, c_{2N}, \ldots, c_{NN}$ and $a_{1N}, a_{2N}, \ldots, a_{NN}$

$$(1.3.1) \qquad T_N = \sum_{j=1}^{N} c_{jN} a_{R_{jN}N}$$

is called a simple linear rank statistic (cf. (1.1.7)).

Define

$$(1.3.2) \qquad \bar{c}_N = \frac{1}{N} \sum_{j=1}^{N} c_{jN}, \quad \bar{a}_N = \frac{1}{N} \sum_{j=1}^{N} a_{jN}$$

and for $\zeta > 0$, let $\gamma(a_{1N}, a_{2N}, \ldots, a_{NN}; \zeta)$ denote the Lebesgue measure λ of the ζ-neighborhood of the set $\{a_{1N}, a_{2N}, \ldots, a_{NN}\}$, thus

$$(1.3.3.) \qquad \gamma(a_{1N}, a_{2N}, \ldots, a_{NN}; \zeta) = \lambda\{x: \exists j \ |x - a_{jN}| < \zeta\}.$$

Let us suppose that the regression constants satisfy

$$(1.3.4) \qquad \sum_{j=1}^{N} c_{jN} = 0, \qquad \sum_{j=1}^{N} c_{jN}^2 = 1.$$

Furthermore, define

$$(1.3.5) \qquad \psi_N(t) = Ee^{it(T_N - ET_N)}.$$

In this section we consider the behavior of the characteristic function ψ_N of the centered simple linear rank statistic T_N for large values of the argument. We quote a result from VAN ZWET (1980) which makes it possible to obtain Edgeworth expansions for statistics of the form (1.3.1) both under the null-hypothesis and under contiguous alternatives.

LEMMA 1.3.1. *Let* $\bar{c}_N = 0$ *and* $\Sigma c_{jN}^2 = 1$. *Suppose that there exist positive numbers* C, a, A *and* δ, *a density* f *and a sequence* $\varepsilon_N \downarrow 0$ *such that*

$$(1.3.6) \qquad \sum_{j=1}^{N} |c_{jN}|^k \le CN^{1-k/2},$$

for some $k > 2$,

$$(1.3.7) \qquad \sum_{j=1}^{N} |a_{jN} - \bar{a}_N|^m \ge aN, \qquad \sum_{j=1}^{N} |a_{jN} - \bar{a}_N|^n \le AN,$$

for some $n > 2$ *and* $0 < m < n$,

$$(1.3.8) \qquad \gamma(a_{1N}, a_{2N}, \ldots, a_{NN}; \zeta) \ge \delta N \zeta \quad \textit{for some } \zeta \ge N^{-3/2} \log N,$$

$$(1.3.9) \qquad \sum_{j=1}^{N} \int_{-\infty}^{\infty} \frac{\{f_j(x) - f(x)\}^2}{f(x)} \, dx \le \varepsilon_N N.$$

Then there exist positive numbers b, B *and* β *depending only on* C, a, A, δ *and the sequence* ε_N *and such that*

$$(1.3.10) \qquad |\psi_N(t)| \le BN^{-\beta \log N} \quad \textit{for } \log N \le |t| \le bN^{3/2}.$$

PROOF. This lemma is an immediate consequence of Theorem 2.1 and the comments in Section 3 of VAN ZWET (1980). □

To conclude this chapter we introduce some notation which will be used throughout this study. The symbols $\mathcal{O}$, o denote the well-known Landau symbols; i.e. as $N \to \infty$, $b_N = \mathcal{O}(d_N)$ denotes the boundedness of b_N/d_N and $b_N = o(d_N)$ means that b_N/d_N tends to zero. Furthermore, $b_N \sim d_N$ stands for b_N/d_N tends to one, as $N \to \infty$.

The standard normal distribution function will be denoted by Φ and its density by ϕ. In this study

$$\sum_{(j,k)\neq}\sum \quad \text{or} \quad \sum\sum_{(j,k,\ell)\neq}\sum \quad \text{etc.}$$

denote the summation over all non-negative distinct integers j, k or j, k, ℓ etc. over a certain range which will be specified.

CHAPTER 2

BERRY-ESSEEN THEOREMS UNDER THE NULL-HYPOTHESIS

2.1. INTRODUCTION AND BERRY-ESSEEN THEOREMS

In this chapter we limit ourselves to the null-hypothesis; i.e. $X_1,X_2,\ldots,X_N$ are independent and identically distributed random variables with a common continuous distribution function F. If $X_{1:N} < X_{2:N} < \ldots < X_{N:N}$ denotes the sequence $X_1,X_2,\ldots,X_N$ arranged in increasing order, then the rank R_{jN} of X_j is defined by $X_j = X_{R_{jN}:N}$ and the antirank D_{jN} is defined by $X_{D_{jN}} = X_{j:N}$, $j = 1,2,\ldots,N$. For specified vectors of real numbers $c_N = (c_{1N}, c_{2N},\ldots,c_{NN})$ (regression constants) and $a_N = (a_{1N},a_{2N},\ldots,a_{NN})$ (scores)

$$(2.1.1) \qquad T_N = \sum_{j=1}^{N} c_{jN}a_{R_{jN}N}$$

is called a simple linear rank statistic.

The purpose of this chapter is to obtain precise information about the rate of convergence of the distribution functions of T_N to their normal limit. We shall establish Berry-Esseen bounds of order $O(N^{-\frac{1}{2}})$ for these statistics. These have been derived in DOES (1981a); the present chapter contains the results of that paper. Our two theorems allow unbounded scores generating functions and include the important special case of the normal quantile function. For bounded scores and under some additional assumptions on the regression constants Berry-Esseen bounds of order $O(N^{-\frac{1}{2}})$, under the null-hypothesis, were obtained by von BAHR (1976), HUŠKOVÁ (1977a, 1979b) and HO & CHEN (1978).

Throughout this chapter we make the following assumption about the regression constants.

ASSUMPTION (2A). *The regression constants* $c_{1N},c_{2N},\ldots,c_{NN}$ *satisfy*

$$\sum_{j=1}^{N} c_{jN} = 0, \qquad \sum_{j=1}^{N} c_{jN}^2 = 1 \quad \text{and} \quad \sum_{j=1}^{N} |c_{jN}|^3 = O(N^{-\frac{1}{2}}).$$

Note that Assumption (2A) implies that $ET_N = 0$.

We also need a condition which ensures that the derivative of a function does not oscillate too wildly near 0 and 1 (cf. Appendix 2 of ALBERS, BICKEL & VAN ZWET (1976)).

CONDITION R_r. *For real* $r > 0$, *a function* h *on* (0,1) *is said to satisfy condition* R_r *if* h *is twice continuously differentiable on* (0,1) *and*

$$\limsup_{t\to 0,1} t(1-t)\left|\frac{h''(t)}{h'(t)}\right| < 1 + \frac{1}{r}.$$

The scores $a_{1N}, a_{2N}, \ldots, a_{NN}$ are generated by a function $J(t)$, $0 < t < 1$ in either one of the following two ways

$$(2.1.2) \qquad \text{(approximate scores)} \qquad a_{jN} = J\left(\frac{j}{N+1}\right), \qquad j = 1,2,\ldots,N,$$

$$(2.1.3) \qquad \text{(exact scores)} \qquad a_{jN} = EJ(U_{j:N}), \qquad j = 1,2,\ldots,N.$$

Here $U_{j:N}$ denotes the j-th order statistic in a random sample of size N from the uniform distribution on (0,1). For almost all well-known linear rank tests the scores are one of these two types. Taking $\bar{a}_N = N^{-1} \Sigma_{j=1}^N a_{jN}$, we find that the variance σ_N^2 of T_N (cf. (2.1.1)) is given by

$$(2.1.4) \qquad \sigma_N^2 = \sigma^2(T_N) = \frac{1}{N-1} \sum_{j=1}^{N} (a_{jN} - \bar{a}_N)^2$$

(see e.g. Theorem II 3.1.c of HÁJEK & ŠIDÁK (1967)).

Define for each $N \geq 2$,

$$(2.1.5) \qquad T_N^* = \sigma_N^{-1}\, T_N$$

and

$$(2.1.6) \qquad F_N^*(x) = P(T_N^* \leq x) \qquad \text{for } -\infty < x < \infty.$$

To formulate our theorems we need some smoothness assumptions for the scores generating function J.

ASSUMPTION (2B). *The scores generating function* J *satisfies*

$$\int_0^1 J(t)dt = 0, \quad \int_0^1 J^2(t)dt = 1 \quad \text{and} \quad \int_0^1 |J(t)|^3 dt < \infty.$$

ASSUMPTION (2C). *The scores generating function* J *is continuously differentiable on* (0,1). *There exist positive numbers* $\Gamma > 0$ *and* $\alpha < 5/4$ *such that its first derivative* J' *satisfies*

$$|J'(t)| \le \Gamma\{t(1-t)\}^{-\alpha} \quad \text{for } t \in (0,1).$$

For exact scores Theorem 2.1.1 provides a Berry-Esseen theorem for the distribution function F_N^* (cf. (2.1.6)) of T_N^* (cf. (2.1.5)). Theorem 2.1.2 deals with the case of approximate scores.

THEOREM 2.1.1. *Take* $a_{jN} = EJ(U_{j:N})$ *for* $j = 1,2,\ldots,N$. *Assume that Assumptions* (2A) *and* (2B) *are satisfied and that*

$$\sum_{j=1}^{N} \sigma^2(J(U_{j:N})) = \mathcal{O}(N^{\frac{1}{2}}(\log N)^{-2}). \tag{2.1.7}$$

Then

$$\sup_{x\in\mathbb{R}} |F_N^*(x) - \Phi(x)| = \mathcal{O}(N^{-\frac{1}{2}}). \tag{2.1.8}$$

THEOREM 2.1.2. *Take* $a_{jN} = J(j/(N+1))$ *for* $j = 1,2,\ldots,N$. *Assume that the regression constants satisfy Assumption* (2A) *and that the scores generating function* J *satisfies Condition* R_1, *Assumptions* (2B) *and* (2C). *Then*

$$\sup_{x\in\mathbb{R}} |F_N^*(x) - \Phi(x)| = \mathcal{O}(N^{-\frac{1}{2}}). \tag{2.1.9}$$

We shall show in Lemma 2.2.1 that, if the scores generating function J satisfies Assumption (2C), then Condition (2.1.7) in Theorem 2.1.1 is fulfilled. It follows that the assumptions of Theorem 2.1.2 imply those of Theorem 2.1.1. We note that Assumption (2B) reduces to a norming assumption if J satisfies Assumption (2C). Furthermore, we note that Theorems 2.1.1 and 2.1.2 both allow scores generating functions tending to infinity in the neighborhood of 0 and 1 at the rate of $\{t(1-t)\}^{-1/4+\varepsilon}$ for $\varepsilon > 0$.

Section 2.2 contains a number of preliminary lemmas. The proofs of the theorems are contained in Section 2.3.

2.2. PRELIMINARIES

The aim in this section is threefold. In the first place we obtain bounds for moments of functions of order statistics. For this we shall draw heavily on the results in Appendix 2 of ALBERS, BICKEL & VAN ZWET (1976). Secondly, we consider the behavior of the characteristic function of T_N^* (cf. (2.1.5)) for large values of the argument. We shall prove a lemma which can be reduced to a special case of Theorem 2.1 of VAN ZWET (1980). Finally, we prove a technical lemma needed in the proof of Theorem 2.1.2. From this point on we shall suppress the index N whenever it is possible: in particular we shall write a_j and c_j instead of a_{jN} and c_{jN}.

LEMMA 2.2.1. *If* J *satisfies Assumption* (2C)*, then there exists a number* $\delta \in (0,\frac{1}{4})$*, such that uniformly for integers* $1 \le k \le \ell \le N$*,*

$$(2.2.1) \qquad \sum_{j=k}^{\ell} E\{J(U_{j:N}) - J(\tfrac{j}{N+1})\}^2 = O(\{\tfrac{k}{N+1}\}^{-\frac{1}{2}+2\delta} + \{\tfrac{N+1-\ell}{N+1}\}^{-\frac{1}{2}+2\delta}),$$

$$(2.2.2) \qquad \sum_{j=1}^{N} \sigma^2(J(U_{j:N})) = O(N^{\frac{1}{2}-2\delta}).$$

If J *also satisfies Condition* R_1*, then uniformly in* k *and* ℓ*,*

$$(2.2.3) \qquad \sum_{j=k}^{\ell} \{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}^2 = O(N^{-1}\{\tfrac{k}{N+1}\}^{-3/2+2\delta} + N^{-1}\{\tfrac{N+1-\ell}{N+1}\}^{-3/2+2\delta}).$$

Finally, if in addition J *satisfies Assumption* (2B)*, then*

$$(2.2.4) \qquad \sum_{j=1}^{N} \{J(\tfrac{j}{N+1}) - \tfrac{1}{N}\sum_{i=1}^{N} J(\tfrac{i}{N+1})\}^2 = N + O(N^{\frac{1}{2}-2\delta}).$$

PROOF. Without loss of generality, we suppose that Assumption (2C) holds for $\alpha \in (1,5/4)$ and we take $\delta = 5/4 - \alpha$. Let h be a function on (0,1) with $h'(t) \equiv \Gamma\{t(1-t)\}^{-5/4+\delta}$ and write $\lambda_j = j/(N+1)$. Since h satisfies Condition R_2, Lemma A.2.3 of ALBERS, BICKEL & VAN ZWET (1976) yields

$$E\{h(U_{j:N}) - h(\lambda_j)\}^2 = O\left(\frac{\{\lambda_j(1-\lambda_j)\}^{-3/2+2\delta}}{N}\right),$$

uniformly in j. Because $|J'(t)| \le h'(t)$ we have $|J(s) - J(t)| \le |h(s) - h(t)|$ for every $s,t \in (0,1)$ and hence

$$E\{J(U_{j:N}) - J(\lambda_j)\}^2 = O\left(\frac{\{\lambda_j(1-\lambda_j)\}^{-3/2+2\delta}}{N}\right),$$

uniformly in j. Now (2.2.1) follows by summation and (2.2.2) is implied by (2.2.1) as $\sigma^2(J(U_{j:N})) \le E\{J(U_{j:N}) - J(\lambda_j)\}^2$.

If J also satisfies Condition R_1 then, in view of (A.2.11) in ALBERS, BICKEL & VAN ZWET (1976), we have

$$(2.2.5) \qquad |EJ(U_{j:N}) - J(\lambda_j)| = \mathcal{O}\left(\frac{\lambda_j(1-\lambda_j)+|J'(\lambda_j)|}{N}\right) = \mathcal{O}\left(\frac{\{\lambda_j(1-\lambda_j)\}^{-5/4+\delta}}{N}\right),$$

uniformly in j; (2.2.3) follows by summation.

If J also satisfies Assumption (2B), then

$$\left|\frac{1}{N}\sum_{j=1}^{N} J(\tfrac{j}{N+1})\right| = \left|\frac{1}{N}\sum_{j=1}^{N}\{J(\tfrac{j}{N+1}) - EJ(U_{j:N})\}\right| = \mathcal{O}(N^{-3/4-\delta}),$$

because of (2.2.5). Furthermore, in view of (2.2.1) and (2.2.5),

$$\left|\sum_{j=1}^{N} J^2(\tfrac{j}{N+1}) - N\right| = \left|\sum_{j=1}^{N}\{J^2(\tfrac{j}{N+1}) - EJ^2(U_{j:N})\}\right|$$

$$\le \sum_{j=1}^{N} E\{J(U_{j:N}) - J(\tfrac{j}{N+1})\}^2 + 2\sum_{j=1}^{N} |J(\tfrac{j}{N+1})|\,|EJ(U_{j:N}) - J(\tfrac{j}{N+1})|$$

$$= \mathcal{O}(N^{\frac{1}{2}-2\delta}),$$

which proves (2.2.4) and the lemma. □

We now consider the behavior of the characteristic function of T_N^* for large values of the argument. Let

$$(2.2.6) \qquad \psi_N^*(t) = Ee^{itT_N^*}.$$

<u>LEMMA 2.2.2</u>. *Suppose that the assumptions of either Theorem* 2.1.1 *or Theorem* 2.1.2 *are satisfied. Then there exist positive numbers* B, β *and* γ *such that*

$$(2.2.7) \qquad |\psi_N^*(t)| \le BN^{-\beta \log N},$$

for $\log N \le |t| \le \gamma N^{\frac{1}{2}}$ *and* $N = 2,3,\ldots$.

<u>PROOF</u>. The present lemma is essentially a special case of Lemma 1.3.1, where (2.2.7) is proved for $\log N \le |t| \le \gamma N^{3/2}$. Since we are concerned with independent and identically distributed random variables $X_1, X_2, \ldots, X_N$ - which

we may assume to be uniformly distributed without loss of generality - Condition (1.3.9) of this lemma is clearly satisfied. Moreover, it is easy to see that Condition (1.3.8) is superfluous in our case since we are only concerned with values of $|t| \leq \gamma N^{\frac{1}{2}}$. It follows from Assumption (2A) that for $k = 3$ Condition (1.3.6) is satisfied. Hence the existence of positive numbers c and C such that

$$(2.2.8) \qquad \sum_{j=1}^{N} (a_j - \bar{a})^2 \geq cN, \qquad \sum_{j=1}^{N} |a_j - \bar{a}|^3 \leq CN$$

suffices to prove the present lemma.

For exact scores $a_j = EJ(U_{j:N})$, Assumption (2B) and (2.1.7) imply that $\bar{a} = \int J = 0$ and

$$\sum_{j=1}^{N} a_j^2 = \sum_{j=1}^{N} EJ^2(U_{j:N}) - \sum_{j=1}^{N} \sigma^2(J(U_{j:N})) = N - O(N^{\frac{1}{2}}),$$

$$\sum_{j=1}^{N} |a_j|^3 \leq \sum_{j=1}^{N} E|J(U_{j:N})|^3 = N \int_0^1 |J(t)|^3 dt$$

and (2.2.8) follows. For approximate scores $a_j = J(j/(N+1))$, (2.2.8) is a consequence of Assumptions (2B) and (2C) (cf. also (2.2.4)). □

Let $[x]$ denote the largest integer not exceeding x. Define $m = [N^{1/3}]$ and $I = \{1,2,\ldots,m,N-m+1,\ldots,N-1,N\}$. Take $\delta \in (0,\frac{1}{4})$ as in Lemma 2.2.1.

<u>LEMMA 2.2.3</u>. *If Assumptions* (2A) *and* (2C) *are satisfied, then*

$$(2.2.9) \qquad E\Big(\sum_{j\in I} c_{D_j} J\big(\tfrac{j}{N+1}\big)\Big)^4 = O(N^{-2/3-8\delta/3}).$$

<u>PROOF</u>. According to Assumption (2A), $\Sigma c_j = 0$, $\Sigma c_j^2 = 1$, $\Sigma |c_j|^3 = O(N^{-\frac{1}{2}})$ and $\Sigma c_j^4 \leq \max|c_j| \Sigma |c_j|^3 = O(N^{-2/3})$. Hence, straightforward computation shows that for distinct $i,j,h,k \in I$,

$$Ec_{D_i}^4 = O(N^{-5/3}), \quad Ec_{D_i}^3 c_{D_j} = O(N^{-8/3}), \quad Ec_{D_i}^2 c_{D_j}^2 = O(N^{-2}),$$

$$Ec_{D_i}^2 c_{D_j} c_{D_h} = O(N^{-3}), \quad Ec_{D_i} c_{D_j} c_{D_h} c_{D_k} = O(N^{-4}).$$

Assumption (2C) ensures that for $\ell = 1,2,3,4$,

$$(2.2.10)\qquad \frac{1}{N}\sum_{j\in I}\left|J\left(\frac{j}{N+1}\right)\right|^{\ell} \sim \int_0^{N^{-2/3}} \{|J(t)|^{\ell} + |J(1-t)|^{\ell}\}dt$$

$$= O(N^{-2/3+\ell/6-2\ell\delta/3}).$$

Direct computation of the left-hand side of (2.2.9) now produces the result of the lemma. □

2.3. PROOFS OF THE THEOREMS

To establish a Berry-Esseen theorem one usually invokes Esseen's smoothing lemma (cf. Lemma 1.1.3), which implies that for all $\gamma > 0$

$$(2.3.1)\qquad \sup_{x\in \mathbb{R}} |F_N^*(x) - \Phi(x)| \leq \frac{1}{\pi}\int_{-\gamma N^{\frac{1}{2}}}^{\gamma N^{\frac{1}{2}}} \frac{|\psi_N^*(t)-e^{-\frac{1}{2}t^2}|}{|t|}\,dt + O(N^{-\frac{1}{2}}),$$

where ψ_N^* denotes the characteristic function T_N^* (cf. (2.2.6)).

It follows from Lemma 2.2.2 that in order to prove Theorems 2.1.1 and 2.1.2 it is sufficient to show that

$$(2.3.2)\qquad \int_{|t|\leq \log N} \frac{|\psi_N^*(t)-e^{-\frac{1}{2}t^2}|}{|t|}\,dt = O(N^{-\frac{1}{2}}).$$

We first prove Theorem 2.1.1. Let $R = (R_1, R_2, \ldots, R_N)$ and $D = (D_1, D_2, \ldots, D_N)$ denote the vectors of ranks and antiranks respectively and define

$$(2.3.3)\qquad S_N = \sum_{j=1}^{N} c_j J(U_j) = \sum_{j=1}^{N} c_{D_j} J(U_{j:N}),$$

where $U_1, U_2, \ldots, U_N$ are independent and uniformly distributed random variables on (0,1). Since the vector of order statistics is independent of R, we have

$$(2.3.4)\qquad E(S_N|R) = \sum_{j=1}^{N} c_{D_j} EJ(U_{j:N}) = T_N$$

and it follows that

$$E(e^{itT_N}(S_N-T_N)) = E(E(e^{itT_N}(S_N-T_N)|R)) = E(e^{itT_N}E(S_N-T_N|R)) = 0.$$

Hence

$$(2.3.5) \qquad Ee^{itS_N} = Ee^{itT_N} + \mathcal{O}(t^2E(S_N-T_N)^2)$$

and because of (2.3.4), Assumption (2B) and (2.1.7)

$$(2.3.6) \qquad \begin{aligned} E(S_N-T_N)^2 &= ES_N^2 - ET_N^2 = 1 - \frac{1}{N-1}\sum_{j=1}^{N}\{EJ(U_{j:N})\}^2 \\ &= \frac{1}{N-1}\sum_{j=1}^{N}\sigma^2(J(U_{j:N})) - \frac{1}{N-1} = \mathcal{O}(N^{-\frac{1}{2}}(\log N)^{-2}). \end{aligned}$$

As S_N is a sum of independent random variables with $ES_N = 0$, $\sigma^2(S_N) = 1$ and $\Sigma|c_j|^3E|J(U_j)|^3 = \mathcal{O}(N^{-\frac{1}{2}})$ (cf. Assumptions (2A) and (2B)), we may apply Lemma V 2.1 of PETROV (1975) to obtain that for $|t| \le \log N$,

$$(2.3.7) \qquad |Ee^{itS_N} - e^{-\frac{1}{2}t^2}| = \mathcal{O}(N^{-\frac{1}{2}}|t|^3e^{-t^2/3}).$$

Finally we note that (2.3.6) implies that

$$(2.3.8) \qquad \sigma_N^2 = \sigma^2(T_N) = 1 + \mathcal{O}(N^{-\frac{1}{2}}(\log N)^{-2}).$$

Combining (2.3.5) through (2.3.8) we arrive at (2.3.2) and the proof of Theorem 2.1.1 is complete.

We now turn to the proof of Theorem 2.1.2. To distinguish simple linear rank statistics with exact scores and with approximate scores we define

$$(2.3.9) \qquad T_N' = \sum_{j=1}^{N} c_j J\left(\frac{R_j}{N+1}\right) = \sum_{j=1}^{N} c_{D_j} J\left(\frac{j}{N+1}\right)$$

and

$$(2.3.10) \qquad T_N = \sum_{j=1}^{N} c_{D_j} EJ(U_{j:N}).$$

Because of Lemma 2.2.1, the assumptions of Theorem 2.1.2 imply those of Theorem 2.1.1. and we may therefore conclude from the proof of Theorem 2.1.1 that

$$(2.3.11)\qquad \int_{|t|\le \log N} \frac{|Ee^{itT_N}-e^{-\frac{1}{2}t^2}|}{|t|}\,dt = O(N^{-\frac{1}{2}}).$$

A Taylor expansion yields

$$(2.3.12)\qquad Ee^{itT_N} = Ee^{itT_N'} + itEe^{itT_N'}(T_N-T_N') + O(t^2E(T_N-T_N')^2).$$

In the situation of Theorem 2.1.2 the scores generating function satisfies both Assumption (2C) and Condition R_1, so that we may apply Lemma 2.2.1 to find that for some $\delta \in (0,\frac{1}{4})$,

$$(2.3.13)\qquad \begin{aligned} E(T_N-T_N')^2 &= E(\sum_{j=1}^{N} c_{D_j}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\})^2 \\ &= \frac{1}{N}\sum_{j=1}^{N}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}^2 - \frac{1}{N(N-1)}\sum\sum_{(i,j)\neq}\{EJ(U_{i:N}) - J(\tfrac{i}{N+1})\} \\ &\quad\cdot\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\} = \frac{1}{N-1}\sum_{j=1}^{N}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}^2 \\ &\quad - \frac{1}{N(N-1)}(\sum_{j=1}^{N}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\})^2 = O(N^{-\frac{1}{2}-2\delta}). \end{aligned}$$

Define $m = [N^{1/3}]$ and $I = \{1,2,\ldots,m,N-m+1,\ldots,N-1,N\}$ as in Section 2.2. Repeating the argument of (2.3.13) for restricted sums we find

$$(2.3.14)\qquad \begin{aligned} &E|\sum_{j=m+1}^{N-m} c_{D_j}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}| \\ &\le \{E(\sum_{j=m+1}^{N-m} c_{D_j}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\})^2\}^{\frac{1}{2}} = O(N^{-\frac{1}{2}-2\delta/3}). \end{aligned}$$

Combining (2.3.11) through (2.3.14) we see that, in order to prove (2.3.2) we have to show that

$$(2.3.15)\qquad \int_{|t|\le \log N} |E(e^{itT_N'}\sum_{j\in I} c_{D_j}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\})|\,dt = O(N^{-\frac{1}{2}}).$$

We note that (2.3.13) and (2.3.14) imply that

$$(2.3.16)\qquad E(\sum_{j\in I} c_{D_j}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\})^2 = O(N^{-\frac{1}{2}-2\delta}).$$

Let $\Omega = \{D_j : j \in I\}$ be the set of antiranks D_j with indices in I and let $\omega = \{d_j : j \in I\}$ be a possible realization of Ω. We have

$$E(e^{itT'_N} \sum_{j\in I} c_{D_j}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\})$$

$$(2.3.17) \qquad = E\{E(\exp\{it \sum_{j=m+1}^{N-m} c_{D_j} J(\tfrac{j}{N+1})\} \mid \Omega) E(\exp\{it \sum_{j\in I} c_{D_j} J(\tfrac{j}{N+1})\}$$

$$\cdot \sum_{j\in I} c_{D_j}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\} \mid \Omega)\}.$$

Conditionally on $\Omega = \omega$, $\Sigma_{j=m+1}^{N-m} c_{D_j} J(j/(N+1))$ is distributed as a simple linear rank statistic for sample size $N-2m$ based on a set of regression constants $\{c_1, c_2, \ldots, c_N\} \backslash \{c_{d_j} : j \in I\}$ and having a scores generating function

$$J_N(x) = J(\frac{m+(N-2m+1)x}{N+1}) \qquad \text{for } x \in (0,1).$$

We write this simple linear rank statistic as

$$(2.3.18) \qquad T_{\omega N} = \sum_{j=1}^{M} b_j J_N\left(\frac{Q_j}{M+1}\right),$$

where $M = N-2m$, $\{b_1, b_2, \ldots, b_M\} = \{c_1, c_2, \ldots, c_N\} \backslash \{c_{d_j} : j \in I\}$, $Q_1, Q_2, \ldots, Q_M$ are the ranks of $V_1, V_2, \ldots, V_M$, which are independent and uniformly distributed random variables on (0,1). Define

$$(2.3.19) \qquad S_{\omega N} = \sum_{j=1}^{M} b_j J_N(V_j).$$

<u>LEMMA 2.3.1</u>. *Under the assumptions of Theorem* 2.1.2 *we have*

$$(2.3.20) \qquad E(T_{\omega N} - S_{\omega N})^2 = (1 + \{\sum_{j\in I} c_{d_j}\}^2) O(N^{-2/3-4\delta/3}).$$

<u>PROOF</u>.

$$E(T_{\omega N} - S_{\omega N})^2 = \sum_{j=1}^{M} b_j^2 E(J_N\left(\frac{Q_1}{M+1}\right) - J_N(V_1))^2$$

$$+ \sum_{(i,j)\neq} \sum b_i b_j E(J_N\left(\frac{Q_1}{M+1}\right) - J_N(V_1))(J_N\left(\frac{Q_2}{M+1}\right) - J_N(V_2)).$$

Because $\Sigma_{j=1}^{M} b_j^2 \le 1$ and

$$|\sum_{(i,j)\neq}\sum b_i b_j| = |(\sum_{j=1}^{M} b_j)^2 - \sum_{j=1}^{M} b_j^2| = |(\sum_{j\in I} c_{d_j})^2 - \sum_{j=1}^{M} b_j^2|$$

$$\le 1 + (\sum_{j\in I} c_{d_j})^2,$$

the Cauchy-Schwarz inequality yields

$$E(T_{\omega N} - S_{\omega N})^2 \le (2 + \{\sum_{j\in I} c_{d_j}\}^2) E(J_N\left(\frac{Q_1}{M+1}\right) - J_N(V_1))^2.$$

Furthermore we have

$$E(J_N\left(\frac{Q_1}{M+1}\right) - J_N(V_1))^2 = \frac{1}{M}\sum_{j=1}^{M} E(J_N(V_{j:M}) - J_N(\frac{j}{M+1}))^2,$$

where $V_{1:M} < V_{2:M} < \ldots < V_{M:M}$ denote the order statistics of $V_1, V_2, \ldots, V_M$. We note that $|J_N'(t)|$ is bounded above by

$$h_N'(t) = (\frac{N-2m+1}{N+1}) h'(\frac{m+(N-2m+1)t}{N+1}),$$

where h is defined as in the proof of Lemma 2.2.1. Since h_N satisfies Condition R_2, we can argue as in the proof of Lemma 2.2.1 to show that

$$\sum_{j=1}^{M} E\{J_N(V_{j:M}) - J_N(\frac{j}{M+1})\}^2 = O(\frac{1}{M}\sum_{j=1}^{M} \frac{j}{M+1}(1 - \frac{j}{M+1})\{h_N'(\frac{j}{M+1})\}^2)$$

$$= O\left(\int_0^1 t(1-t)\{h_N'(t)\}^2 dt\right) = O\left(\int_{\frac{m}{N+1}}^{1-\frac{m}{N+1}} \{t(1-t)\}^{-3/2+2\delta} dt\right)$$

$$= O(N^{1/3-4\delta/3})$$

and the proof of the lemma is complete. □

It follows from Lemma 2.3.1 that

$$(2.3.21) \qquad \left|Ee^{itT_{\omega N}} - Ee^{itS_{\omega N}}\right| = O(|t|N^{-1/3-2\delta/3}\{1 + (\sum_{j\in I} c_{d_j})^2\}^{\frac{1}{2}}).$$

Since $S_{\omega N}$ is a sum of independent random variables, with variance (cf. Assumption (2A))

$$(2.3.22)\qquad \tau_\omega^2 = \sigma^2(S_{\omega N}) = (1 - \sum_{j\in I} c_{d_j}^2)\sigma^2(J_N(V_1)),$$

Lemma V 2.1 of PETROV (1975) together with Assumptions (2A) and (2B) yield

$$(2.3.23)\qquad \left|Ee^{it(S_{\omega N}-ES_{\omega N})} - e^{-\frac{1}{2}\tau_\omega^2 t^2}\right| = O(N^{-\frac{1}{2}}|t|^3),$$

for all $|t| \leq \log N$. Furthermore, in view of Assumptions (2A), (2B) and (2C),

$$\left|Ee^{it(S_{\omega N}-ES_{\omega N})} - Ee^{itS_{\omega N}}\right| \leq |t|\,|ES_{\omega N}|$$

$$(2.3.24)\qquad = |t|\left|\sum_{j=1}^{M} b_j \int_0^1 J_N(t)dt\right| \leq |t|\frac{N}{M}\left|\sum_{j\in I} c_{d_j}\right|\left|\int_{\frac{m}{N+1}}^{1-\frac{m}{N+1}} J(t)dt\right|$$

$$\leq 2|t|\frac{mN}{M}\max_{1\leq j\leq N}|c_j|\left|\int_0^{\frac{m}{N+1}}\{J(t)+J(1-t)\}dt\right| = O(|t|N^{-1/3-2\delta/3}).$$

Defining

$$(2.3.25)\qquad \tau_N^2 = MN^{-1}\sigma^2(J_N(V_1))$$

and noting that

$$\left|e^{-\frac{1}{2}\tau_\omega^2 t^2} - e^{-\frac{1}{2}\tau_N^2 t^2}\right| \leq \frac{1}{2}|\tau_\omega^2 - \tau_N^2|t^2,$$

we combine (2.3.21) through (2.3.24) to arrive at

$$(2.3.26)\qquad \left|Ee^{itT_{\omega N}} - e^{-\frac{1}{2}\tau_N^2 t^2}\right| = O(|t|N^{-1/3-2\delta/3}\{1 + (\sum_{j\in I} c_{d_j})^2\}^{\frac{1}{2}} + t^2|\tau_\omega^2 - \tau_N^2|)$$

for all $|t| \leq \log N$ and uniformly in ω. Substituting this result in (2.3.17), we obtain by repeated use of the Cauchy-Schwarz inequality

$$E e^{itT'_N} \sum_{j\in I} c_{D_j}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}$$

$$= e^{-\frac{1}{2}\tau_N^2 t^2} E(\exp\{it \sum_{j\in I} c_{D_j} J(\tfrac{j}{N+1})\} \sum_{j\in I} c_{D_j}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\})$$

(2.3.27)

$$+ O(E|\sum_{j\in I} c_{D_j}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}|(|t|N^{-1/3-2\delta/3}\{1 + (\sum_{j\in I} c_{D_j})^2\}^{\frac{1}{2}}$$

$$+ t^2|1 - \sum_{j\in I} c_{D_j}^2 - \tfrac{M}{N}|))$$

$$= e^{-\frac{1}{2}\tau_N^2 t^2} E(\{1 + it \sum_{j\in I} c_{D_j} J(\tfrac{j}{N+1})\} \sum_{j\in I} c_{D_j}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\})$$

$$+ O(\{E(\sum_{j\in I} c_{D_j}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\})^2\}^{\frac{1}{2}}[\{t^2(E\{\sum_{j\in I} c_{D_j} J(\tfrac{j}{N+1})\}^4)^{\frac{1}{2}}$$

$$+ |t|N^{-1/3-2\delta/3}\{E(1 + (\sum_{j\in I} c_{D_j})^2)\}^{\frac{1}{2}} + t^2\{E(\sum_{j\in I} c_{D_j}^2 - \tfrac{2m}{N})^2\}^{\frac{1}{2}}]),$$

for all $|t| \leq \log N$.

We note that Assumptions (2A) and (2C), (2.2.5) and (2.2.10) imply

$$|E\{1 + it \sum_{j\in I} c_{D_j} J(\tfrac{j}{N+1})\} \sum_{j\in I} c_{D_j}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}|$$

(2.3.28) $$= |t|\left|\frac{1}{N-1} \sum_{j\in I} J(\tfrac{j}{N+1})\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\} - \frac{1}{N(N-1)}\{\sum_{j\in I} J(\tfrac{j}{N+1})\}\right.$$

$$\left.\cdot \sum_{j\in I} \{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}\right| = O(|t|N^{-\frac{1}{2}-2\delta}).$$

Furthermore, we obtain by applying Lemma 2.2.3

(2.3.29) $$\{E(\sum_{j\in I} c_{D_j} J(\tfrac{j}{N+1}))^4\}^{\frac{1}{2}} = O(N^{-1/3-4\delta/3}).$$

Finally,

(2.3.30) $$\{E(1 + \{\sum_{j\in I} c_{D_j}\}^2)\}^{\frac{1}{2}} = 1 + O(N^{-2/3})$$

and

$$(2.3.31)\qquad \{E(\sum_{j\in I} c_{D_j}^2 - \frac{2m}{N})^2\}^{\frac{1}{2}} = O(N^{-2/3})$$

according to Assumption (2A). Combining (2.3.16) and (2.3.28) through (2.3.31) and substituting these results in the right-hand side of (2.3.27) we find that

$$(2.3.32)\qquad E(e^{itT_N'} \sum_{j\in I} c_{D_j}\{EJ(U_{j:N}) - J(\frac{j}{N+1})\}) = O(|t|N^{-\frac{1}{2}-2\delta}),$$

for all $|t| \leq \log N$. To conclude we note that it follows from (2.2.4) that

$$(2.3.33)\qquad \sigma^2(T_N') = \frac{1}{N-1}\sum_{j=1}^{N} (J(\frac{j}{N+1}) - \frac{1}{N}\sum_{i=1}^{N} J(\frac{i}{N+1}))^2 = 1 + O(N^{-\frac{1}{2}-2\delta}).$$

We see that the proof of Theorem 2.1.2 is complete by combining (2.3.1), Lemma 2.2.2, (2.3.11) through (2.3.14), (2.3.32) and (2.3.33). □

CHAPTER 3

EDGEWORTH EXPANSIONS UNDER THE NULL-HYPOTHESIS

3.1. INTRODUCTION AND EDGEWORTH EXPANSION

In the present chapter we investigate third order approximations for the distribution functions of simple linear rank statistics under the null-hypothesis. We shall establish Edgeworth expansions for these statistics with remainder $o(N^{-1})$. This chapter contains the results of DOES (1981b) where these expansions have been derived.

Let $X_1, X_2, \ldots, X_N$ be independent and identically distributed random variables with a common continuous distribution function F. If $X_{1:N} < X_{2:N} < \ldots < X_{N:N}$ denotes the sequence $X_1, X_2, \ldots, X_N$ arranged in increasing order, then the rank R_{jN} of X_j is defined by $X_j = X_{R_{jN}:N}$ and the antirank D_{jN} is defined by $X_{D_{jN}} = X_{j:N}$, $j = 1,2,\ldots,N$. We consider the simple linear rank statistic

$$(3.1.1) \qquad T_N = \sum_{j=1}^{N} c_{jN} J\left(\frac{R_{jN}}{N+1}\right) = \sum_{j=1}^{N} c_{D_{jN}N} J\left(\frac{j}{N+1}\right),$$

where $c_{1N}, c_{2N}, \ldots, c_{NN}$, $N = 1,2,\ldots$, is a triangular array of regression constants and J is a scores generating function defined on (0,1).

Throughout this chapter we make the following assumptions.

<u>ASSUMPTION (3A)</u>. *The regression constants* $c_{1N}, c_{2N}, \ldots, c_{NN}$ *satisfy*

$$\sum_{j=1}^{N} c_{jN} = 0, \qquad \sum_{j=1}^{N} c_{jN}^2 = 1 \quad \textit{and} \quad \max_{1\le j\le N} |c_{jN}| = O(N^{-\frac{1}{2}}).$$

This assumption implies that $ET_N = 0$.

<u>ASSUMPTION (3B)</u>. *The scores generating function* J *is three times differentiable on* (0,1) *and*

$$(3.1.2)\qquad \limsup_{t\to 0,1} t(1-t)\left|\frac{J''(t)}{J'(t)}\right| < 2;$$

there exist positive numbers $\Gamma > 0$ *and* $\alpha < 3 + 1/14$ *such that its third derivative* J''' *satisfies*

$$(3.1.3)\qquad |J'''(t)| \le \Gamma\{t(1-t)\}^{-\alpha} \quad \textit{for } t \in (0,1).$$

Furthermore

$$(3.1.4)\qquad \int_0^1 J(t)dt = 0, \quad \int_0^1 J^2(t)dt = 1.$$

We note that (3.1.2) ensures that the function J' does not oscillate too wildly near 0 and 1 (see also Appendix 2 of ALBERS, BICKEL & VAN ZWET (1976)). Assumption (3.1.4) can be made without loss of generality.

Taking

$$(3.1.5)\qquad \bar{J} = \frac{1}{N}\sum_{j=1}^{N} J(\tfrac{j}{N+1}),$$

we know that the variance σ_N^2 of T_N (cf. (3.1.1)) is given by

$$(3.1.6)\qquad \sigma_N^2 = \sigma^2(T_N) = \frac{1}{N-1}\sum_{j=1}^{N} (J(\tfrac{j}{N+1}) - \bar{J})^2$$

(see e.g. Theorem II 3.1.c of HÁJEK & ŠIDÁK (1967)). Define, for each $N \ge 2$,

$$(3.1.7)\qquad T_N^* = \sigma_N^{-1}\, T_N$$

and

$$(3.1.8)\qquad F_N^*(x) = P(T_N^* \le x) \quad \text{for } -\infty < x < \infty.$$

Furthermore define for each $N \ge 2$ and real x, the function $\tilde{F}_N$ by

$$(3.1.9)\qquad \tilde{F}_N(x) = \Phi(x) - \phi(x)\left\{\frac{\kappa_{3N}}{6}(x^2-1) + \frac{\kappa_{4N}}{24}(x^3-3x) + \frac{\kappa_{3N}^2}{72}(x^5-10x^3+15x)\right\},$$

where the quantities κ_{3N} and κ_{4N} are given by

$$(3.1.10)\qquad \kappa_{3N} = \sum_{j=1}^{N} c_{jN}^3\left\{\int_0^1 J^3(t)dt\right\}$$

and

$$\kappa_{4N} = \sum_{j=1}^{N} c_{jN}^4 \left\{\int_0^1 J^4(t)dt - 3\right\} - \frac{3}{N}\left\{\int_0^1 J^4(t)dt - 1\right\}. \tag{3.1.11}$$

Our theorem reads as follows.

THEOREM 3.1.1. *If the Assumptions* (3A) *and* (3B) *are satisfied, then as* $N \to \infty$

$$\sup_{x\in \mathbb{R}} |F_N^*(x) - \tilde{F}_N(x)| = o(N^{-1}). \tag{3.1.12}$$

We note that κ_{3N} and κ_{4N} (cf. (3.1.10) and (3.1.11) are asymptotic expressions for the third and fourth cumulants of T_N^* where terms of order $o(N^{-1})$ have been neglected. Hence $\tilde{F}_N$ may be said to constitute a genuine Edgeworth expansion for F_N^*. We should also point out that Theorem 3.1.1 allows scores generating functions tending to infinity in the neighborhood of 0 and 1 at the rate of $\{t(1-t)\}^{-1/14+\varepsilon}$ for some $\varepsilon > 0$. It is clear that this includes the normal quantile function, Φ^{-1}.

Whenever we shall suppose in the remainder of this chapter that (3.1.3) in Assumption (3B) is satisfied, we shall tacitly and without loss of generality assume that $\alpha \in (3,3+1/14)$ and define $\delta = 3+1/14-\alpha$. Hence from now on we replace (3.1.3) in Assumption (3B) by

$$|J'''(t)| \le \Gamma\{t(1-t)\}^{-(3+1/14)+\delta} \quad \text{for } t \in (0,1), \tag{3.1.13}$$

where

$$0 < \delta < 1/14. \tag{3.1.14}$$

The reason for making the obviously superfluous assumption that $\alpha > 3$ and hence that $\delta < 1/14$, is that some of our intermediate results hold only in that case.

In Section 3.2 we prove a number of preliminaries. The proof of Theorem 3.1.1 is contained in Section 3.3. In Section 3.4 we compare our results with those in BICKEL & VAN ZWET (1978) for the two-sample linear rank statistics. Finally, in the last section we discuss briefly the numerical aspects of our expansions.

3.2. PRELIMINARY LEMMAS

The aim of this section is threefold. In the first place we approximate $(N-1)\sigma_N^2$ (cf. (3.1.6)) by an integral. Secondly we study the behavior of the characteristic function of T_N^* (cf. (3.1.7)) for large values of the argument. To this end we shall provide a lemma which is a special case of Theorem 2.1 of VAN ZWET (1980). Finally we prove two technical lemmas, the purpose of which will become clear in Section 3.3.

As in Chapter 2 we shall suppress the index N whenever it is possible: in particular we shall write c_j, R_j and D_j instead of c_{jN}, R_{jN} and D_{jN}. Let $U_1, U_2, \ldots, U_N$ be independent and uniformly distributed random variables on $(0,1)$ and $U_{1:N} < U_{2:N} < \ldots < U_{N:N}$ the corresponding uniform order statistics.

LEMMA 3.2.1. *If* J *satisfies Assumption* (3B), *then*

$$\sum_{j=1}^{N} \left(J\left(\tfrac{j}{N+1}\right) - \bar{J}\right)^2 = N + O(N^{1/7-2\delta}), \tag{3.2.1}$$

with δ *as in* (3.1.13) *and* (3.1.14).

PROOF. Take δ as in (3.1.13) and (3.1.14), let h be a function on (0,1) with $h'(t) \equiv \Gamma\{t(1-t)\}^{-15/14+\delta}$ and write $\lambda_j = j/(N+1)$. Since

$$\limsup_{t\to 0,1} t(1-t)\left|\frac{h''(t)}{h'(t)}\right| < \frac{3}{2},$$

Lemma A.2.3 of ALBERS, BICKEL & VAN ZWET (1976) yields

$$E\{h(U_{j:N}) - h(\lambda_j)\}^2 = O\left(\frac{\{\lambda_j(1-\lambda_j)\}^{-8/7+2\delta}}{N}\right),$$

uniformly in j. Because $|J'(t)| \le h'(t)$ we have $|J(s)-J(t)| \le |h(s)-h(t)|$ for every $s,t \in (0,1)$ and hence

$$E\{J(U_{j:N}) - J(\lambda_j)\}^2 = O\left(\frac{\{\lambda_j(1-\lambda_j)\}^{-8/7+2\delta}}{N}\right), \tag{3.2.2}$$

uniformly in j. As J satisfies (3.1.2), we also have, in view of (A.2.11) in ALBERS, BICKEL & VAN ZWET (1976),

$$(3.2.3)\qquad |EJ(U_{j:N}) - J(\lambda_j)| = O\left(\frac{\lambda_j(1-\lambda_j)+|J'(\lambda_j)|}{N}\right) = O\left(\frac{\{\lambda_j(1-\lambda_j)\}^{-15/14+\delta}}{N}\right),$$

uniformly in j. Since $\int J = 0$ and $\delta \in (0,1/14)$ (cf. (3.1.4) and (3.1.14)), it follows that

$$(3.2.4)\qquad \left|\frac{1}{N}\sum_{j=1}^{N} J(\lambda_j)\right| = \left|\frac{1}{N}\sum_{j=1}^{N} \{J(\lambda_j) - EJ(U_{j:N})\}\right| = O(N^{-13/14-\delta}).$$

Furthermore, in view of (3.2.2) and (3.2.3) and since $\int J^2 = 1$ and $\delta \in (0,1/14)$

$$\left|\sum_{j=1}^{N} J^2(\lambda_j)-N\right| = \left|\sum_{j=1}^{N} \{J^2(\lambda_j) - EJ^2(U_{j:N})\}\right|$$

$$\le \sum_{j=1}^{N} E\{J(U_{j:N}) - J(\lambda_j)\}^2 + 2\sum_{j=1}^{N} |J(\lambda_j)|\,|EJ(U_{j:N}) - J(\lambda_j)|$$

$$= O(N^{1/7-2\delta}),$$

which proves the lemma. □

We now consider the behavior of the characteristic function of T_N^* for large values of the argument. Let

$$(3.2.5)\qquad \psi_N^*(t) = Ee^{itT_N^*}.$$

<u>LEMMA 3.2.2</u>. *If Assumptions* (3A) *and* (3B) *are satisfied, then there exist positive numbers* B, β *and* γ *such that*

$$(3.2.6)\qquad |\psi_N^*(t)| \le BN^{-\beta \log N},$$

for $\log N \le |t| \le \gamma N^{3/2}$ *and* $N = 2,3,\ldots$.

<u>PROOF</u>. The present lemma is a special case of Lemma 1.3.1. Since we are concerned with independent and identically distributed random variables - which we may assume to be uniformly distributed without loss of generality - Condition (1.3.9) of this lemma is clearly satisfied. Moreover, Assumption (3B) guarantees that there exists a positive fraction of the scores which are at a distance of at least $N^{-3/2}\log N$ apart from each other, so Condition (1.3.8)

is also fulfilled. It follows from Assumption (3A) that for $k = 4$ Condition (1.3.6) of Lemma 1.3.1 is satisfied. Hence the existence of positive numbers c and C such that

$$(3.2.7) \qquad \sum_{j=1}^{N} (J(\tfrac{j}{N+1}) - \bar{J})^2 \geq cN, \qquad \sum_{j=1}^{N} (J(\tfrac{j}{N+1}) - \bar{J})^4 \leq CN$$

suffices to prove the present lemma. However (3.2.7) is a consequence of Assumption (3B) (cf. (3.2.1)). □

Let [x] denote the largest integer not exceeding x. Define $m = [N^{8/15}]$ and $I = \{1,2,\ldots,m,N-m+1,\ldots,N-1,N\}$.

<u>LEMMA 3.2.3</u>. *If Assumptions* (3A) *and* (3B) *are satisfied, then*

$$(3.2.8) \qquad E|\sum_{j\in I} c_{D_j} J(\tfrac{j}{N+1})|^5 = O(N^{-1-7\delta/3}),$$

$$(3.2.9) \qquad \{\frac{1}{N-2m} \sum_{j=m+1}^{N-m} J(\tfrac{j}{N+1})\}^2 E(\sum_{j=m+1}^{N-m} c_{D_j})^2 = O(N^{-4/3-14\delta/15}),$$

with δ *as in* (3.1.13) *and* (3.1.14).

<u>PROOF</u>. According to Assumption (3A) $\Sigma c_j = 0$, $\Sigma c_j^2 = 1$ and

$$\sum_{j=1}^{N} |c_j|^k \leq \max_{1\leq j\leq N} |c_j|^{k-2} \sum_{j=1}^{N} c_j^2 = O(N^{1-k/2}),$$

for $k > 2$. It follows that for distinct $i,j,h,g,k,\ell \in I$

$Ec_{D_i}^6 = O(N^{-3})$, $\qquad Ec_{D_i}^5 c_{D_j} = O(N^{-4})$, $\qquad Ec_{D_i}^4 c_{D_j}^2 = O(N^{-3})$,

$Ec_{D_i}^3 c_{D_j}^3 = O(N^{-3})$, $\qquad Ec_{D_i}^4 c_{D_j} c_{D_h} = O(N^{-4})$, $\qquad Ec_{D_i}^3 c_{D_j}^2 c_{D_h} = O(N^{-4})$,

$Ec_{D_i}^2 c_{D_j}^2 c_{D_h}^2 = O(N^{-3})$, $\qquad Ec_{D_i}^3 c_{D_j} c_{D_h} c_{D_g} = O(N^{-5})$,

$Ec_{D_i}^2 c_{D_j}^2 c_{D_h} c_{D_g} = O(N^{-4})$, $\qquad Ec_{D_i}^2 c_{D_j} c_{D_h} c_{D_g} c_{D_k} = O(N^{-5})$,

$Ec_{D_i} c_{D_j} c_{D_h} c_{D_g} c_{D_k} c_{D_\ell} = O(N^{-6})$.

Furthermore, Hölder's inequality yields

(3.2.10) $$E\left|\sum_{j\in I} c_{D_j} J(\tfrac{j}{N+1})\right|^5 \le \{E(\sum_{j\in I} c_{D_j} J(\tfrac{j}{N+1}))^6\}^{5/6}.$$

In view of (3.1.13) and (3.1.14) we have for $k = 1,2,\dots,6$

(3.2.11) $$\frac{1}{N}\sum_{j\in I} |J(\tfrac{j}{N+1})|^k = O\left(\int_0^{\frac{m}{N+1}} \{t(1-t)\}^{-k/14+k\delta} dt\right) = O(\{\tfrac{m}{N+1}\}^{1-k/14+k\delta}).$$

Direct computation of the right-hand side of (3.2.10) produces (3.2.8). Since $\Sigma c_j = 0$, $Ec_{D_j}^2 = N^{-1}$ and $Ec_{D_i}c_{D_j} = -\{N(N-1)\}^{-1}$ for $i \ne j$, we have

$$E(\sum_{j=m+1}^{N-m} c_{D_j})^2 = E(\sum_{j\in I} c_{D_j})^2 = O(\tfrac{m}{N}).$$

Similarly we find that, in view of (3.2.4) and (3.2.11),

(3.2.12) $$\left|\frac{1}{N}\sum_{j=m+1}^{N-m} J(\tfrac{j}{N+1})\right| = \left|\frac{1}{N}\sum_{j\in I} J(\tfrac{j}{N+1})\right| + O(N^{-13/14-\delta}) = O(\{\tfrac{m}{N}\}^{13/14+\delta})$$

and the lemma follows. □

To conclude this section we prove

<u>LEMMA 3.2.4</u>. *If Assumption* (3A) *is satisfied, then for any* $\gamma < 1$ *and* $N \to \infty$

(3.2.13) $$P(\sum_{j\in I} c_{D_j}^2 \ge 1-\gamma) = O(N^{-22/15}).$$

<u>PROOF</u>. Since $E(\Sigma_{j\in I}\, c_{D_j}^2) = 2mN^{-1}$ and

$$E(\sum_{j\in I} c_{D_j}^2 - \tfrac{2m}{N})^2 = \frac{2m(N-2m)}{N(N-1)}\,(\sum_{j=1}^{N} c_j^4 - \tfrac{1}{N}),$$

the Bienaymé-Chebyshev inequality ensures that for every $\gamma < 1$

$$P(|\sum_{j\in I} c_{D_j}^2 - \tfrac{2m}{N}| \ge \tfrac{1-\gamma}{2}) \le \frac{4}{(1-\gamma)^2}\, E(\sum_{j\in I} c_{D_j}^2 - \tfrac{2m}{N})^2 = O(N^{-22/15}).$$

The lemma follows because $mN^{-1} \to 0$ as $N \to \infty$. □

3.3. PROOF OF THEOREM 3.1.1

To prove Theorem 3.1.1 we start with an application of Esseen's smoothing lemma (cf. Lemma 1.1.3), which implies that for all $\gamma > 0$

$$(3.3.1)\qquad \sup_{x\in\mathbb{R}} |F_N^*(x)-\tilde{F}_N(x)| \le \frac{1}{\pi}\int_{-\gamma N^{3/2}}^{\gamma N^{3/2}} \frac{|\psi_N^*(t)-\tilde{\psi}_N(t)|}{|t|}dt + O(N^{-3/2}),$$

where ψ_N^* denotes the characteristic function of T_N^* (cf. (3.2.5)) and $\tilde{\psi}_N$ denotes the Fourier-Stieltjes transform of $\tilde{F}_N$ (cf. (3.1.9)), i.e.

$$(3.3.2)\qquad \tilde{\psi}_N(t) = \int_{-\infty}^{\infty} e^{itx}d\tilde{F}_N(x) = e^{-\frac{1}{2}t^2}\{1 - \frac{\kappa_{3N}}{6}it^3 + \frac{\kappa_{4N}}{24}t^4 - \frac{\kappa_{3N}^2}{72}t^6\}.$$

The derivative of $\tilde{\psi}_N$ is uniformly bounded and also

$$\left|\frac{d\psi_N^*(t)}{dt}\right| \le E|T_N^*| \le 1.$$

Because $\psi_N^*(0) = \tilde{\psi}_N(0) = 1$, we have

$$(3.3.3)\qquad \int_{|t|\le N^{-3/2}} \frac{|\psi_N^*(t)-\tilde{\psi}_N(t)|}{|t|}dt = O(N^{-3/2}).$$

Similarly, Lemma 3.2.2 and (3.3.2) ensure that

$$(3.3.4)\qquad \int_{\log N\le|t|\le\gamma N^{3/2}} \frac{|\psi_N^*(t)-\tilde{\psi}_N(t)|}{|t|}dt = O(N^{-3/2}).$$

From (3.3.1), (3.3.3) and (3.3.4) it follows that, in order to prove Theorem 3.1.1 it suffices to show that

$$(3.3.5)\qquad \int_{t\in A} \frac{|\psi_N^*(t)-\tilde{\psi}_N(t)|}{|t|}dt = o(N^{-1}),$$

where $A = \{t: N^{-3/2} \le |t| \le \log N\}$.

To solve this problem we use a conditioning argument. We take δ as in (3.1.13) and (3.1.14) and define $m = [N^{8/15}]$ and $I = \{1,2,\ldots,m,N-m+1,\ldots \ldots,N-1,N\}$ as in Section 3.2. Let $\Omega = \{D_j : j \in I\}$ be the set of antiranks D_j with indices in I and let $\omega = \{d_j : j \in I\}$ be a possible realization of Ω.

Finally define

$$(3.3.6)\qquad Z_N = \sum_{j\in I} c_{D_j} J(\tfrac{j}{N+1}).$$

Because $(T_N - Z_N)$ and Z_N are conditionally independent given Ω, we have

$$\psi_N^*(t) = Ee^{itT_N^*} = E[E(e^{it\sigma_N^{-1}(T_N-Z_N)} \mid \Omega)E(e^{it\sigma_N^{-1}Z_N} \mid \Omega)]$$

(3.3.7)

$$= E[E(e^{it\sigma_N^{-1}\{(T_N-Z_N)-E(T_N-Z_N|\Omega)\}} \mid \Omega)e^{it\sigma_N^{-1}E(T_N-Z_N|\Omega)}E(e^{it\sigma_N^{-1}Z_N}|\Omega)].$$

We note that conditionally on $\Omega = \omega$, $T_N - Z_N = \Sigma_{j=m+1}^{N-m} c_{D_j}J(j/(N+1))$ is distributed as a simple linear rank statistic for sample size $N-2m$ based on a set of regression constants $\{c_1,c_2,\ldots,c_N\}\backslash\{c_{d_j}: j \in I\}$ and having a scores generating function

$$(3.3.8)\qquad J_N(t) = J(\tfrac{m+(N-2m+1)t}{N+1}) \quad \text{for } t \in (0,1).$$

We write this simple linear rank statistic as

$$(3.3.9)\qquad T_{\omega N} = \sum_{j=1}^{M} b_j J_N\left(\frac{Q_j}{M+1}\right),$$

where $M = N-2m$, $\{b_1,b_2,\ldots,b_M\} = \{c_1,c_2,\ldots,c_N\}\backslash\{c_{d_j}: j \in I\}$, $Q_1,Q_2,\ldots Q_M$ are the ranks of $V_1,V_2,\ldots,V_M$, which are independent and uniformly distributed random variables on $(0,1)$.

Define for $j = 1,2,\ldots,M$

$$(3.3.10)\qquad \hat{V}_j = E\left(\frac{Q_j}{M+1} \mid V_j\right) = \frac{1}{M+1} + \frac{M-1}{M+1} V_j$$

and let $S_{\omega N}$ be a three-term Taylor expansion of $T_{\omega N}$, viz.

$$(3.3.11)\qquad S_{\omega N} = \sum_{j=1}^{M} b_j\{J_N(\hat{V}_j) + J_N'(\hat{V}_j)\left(\frac{Q_j}{M+1} - \hat{V}_j\right) + \frac{1}{2} J_N''(\hat{V}_j)\left(\frac{Q_j}{M+1} - \hat{V}_j\right)^2\}.$$

We shall approximate $(T_{\omega N} - ET_{\omega N})$ by $(S_{\omega N}-ES_{\omega N})$ and for this we need

<u>LEMMA 3.3.1</u>. *Under the Assumptions* (3A) *and* (3B) *we have, uniformly in* ω,

(3.3.12) $$\sigma^2(T_{\omega N} - S_{\omega N}) = (1 + \{\sum_{j\in I} c_{d_j}\}^2)\mathcal{O}(N^{-2-14\delta/15}),$$

with δ *as in* (3.1.13) *and* (3.1.14).

PROOF. Let, for $j = 1,2,\ldots,M$,

$$Y_j = J_N\left(\frac{Q_j}{M+1}\right) - \{J_N(\hat{V}_j) + J_N'(\hat{V}_j)\left(\frac{Q_j}{M+1} - \hat{V}_j\right) + \frac{1}{2} J_N''(\hat{V}_j)\left(\frac{Q_j}{M+1} - \hat{V}_j\right)^2\}.$$

Because $\Sigma_{j=1}^{M} b_j^2 \le 1$ and

$$|\sum_{(j,k)\neq}\sum b_j b_k| = |(\sum_{j=1}^{M} b_j)^2 - \sum_{j=1}^{M} b_j^2| = |(\sum_{j\in I} c_{d_j})^2 - \sum_{j=1}^{M} b_j^2|$$
$$\le 1 + (\sum_{j\in I} c_{d_j})^2,$$

the Cauchy-Schwarz inequality yields

$$\sigma^2(T_{\omega N}-S_{\omega N}) \le E(T_{\omega N}-S_{\omega N})^2 = E(\sum_{j=1}^{M} b_j Y_j)^2 = \sum_{j=1}^{M} b_j^2 EY_1^2$$
$$+ \sum_{(j,k)\neq}\sum b_j b_k EY_1 Y_2 \le EY_1^2 + (1 + \{\sum_{j\in I} c_{d_j}\}^2)E|Y_1 Y_2|$$
$$\le (2 + \{\sum_{j\in I} c_{d_j}\}^2)EY_1^2.$$

Define $r(t) = \{t(1-t)\}^{-1}$. By Taylor's theorem, (3.3.8), (3.1.13) and the convexity of the function $r(t)$ we see that

$$EY_1^2 \le \frac{1}{36} E\left(\frac{Q_1}{M+1} - \hat{V}_1\right)^6 \sup_{0\le\eta\le 1} \{J_N'''(\eta \frac{Q_1}{M+1} + (1-\eta)\hat{V}_1)\}^2$$
$$\le \frac{\Gamma^2}{36} E\left(\frac{Q_1}{M+1} - \hat{V}_1\right)^6 \{r^{6+1/7-2\delta}\left(\frac{m+Q_1}{N+1}\right) + r^{6+1/7-2\delta}\left(\frac{m+(M+1)\hat{V}_1}{N+1}\right)\}.$$

The independence of the vector of ranks $(Q_1,Q_2,\ldots,Q_M)$ and the vector of order statistics $(V_{1:M},V_{2:M},\ldots,V_{M:M})$ and Lemma A.2.3 of ALBERS, BICKEL & VAN ZWET (1976) imply

$$E\left(\frac{Q_1}{M+1}-\hat{V}_1\right)^6 r^{6+1/7-2\delta}\left(\frac{m+Q_1}{M+1}\right) = \left(\frac{M-1}{M+1}\right)^6 E\left(\frac{Q_1-1}{M-1}-V_1\right)^6$$

$$(3.3.13) \qquad \cdot r^{6+1/7-2\delta}\left(\frac{m+Q_1}{N+1}\right) \le \frac{1}{M}\sum_{j=1}^{M} E\left(V_{j:M}-\frac{j-1}{M-1}\right)^6 r^{6+1/7-2\delta}\left(\frac{m+j}{N+1}\right)$$

$$= \mathcal{O}\left(\frac{1}{M^4}\sum_{j=1}^{M} r^{-3}\left(\frac{j}{M+1}\right) r^{6+1/7-2\delta}\left(\frac{m+j}{N+1}\right)\right) = \mathcal{O}(N^{-2-14\delta/15}).$$

Furthermore, the conditional distribution of $Q_1 - 1$ given V_1 is binomial with parameters $M-1$ and V_1 and by application of a recursion formula for the central moments of this distribution (cf. JOHNSON & KOTZ (1969), p. 52) we find

$$E(\{Q_1 - E(Q_1|V_1)\}^6|V_1) = \mathcal{O}(\{MV_1(1-V_1)\}^3 + MV_1(1-V_1)).$$

Hence,

$$E\left(\frac{Q_1}{M+1}-V_1\right)^6 r^{6+1/7-2\delta}\left(\frac{m+(M+1)\hat{V}_1}{N+1}\right) = \mathcal{O}\left(E\left(\left\{\frac{V_1(1-V_1)}{M}\right\}^3 + \frac{V_1(1-V_1)}{M^5}\right)\right.$$

$$(3.3.14) \qquad \left.\cdot r^{6+1/7-2\delta}\left(\frac{m+(M+1)\hat{V}_1}{N+1}\right)\right) = \mathcal{O}(N^{-2-14\delta/15}).$$

Combining (3.3.13) and (3.3.14) we find that $EY_1^2 = \mathcal{O}(N^{-2-14\delta/15})$. This proves the lemma. □

It follows from Lemma 3.3.1, (3.1.6) and (3.2.7) that

$$\left|Ee^{it\sigma_N^{-1}(T_{\omega N}-ET_{\omega N})} - Ee^{it\sigma_N^{-1}(S_{\omega N}-ES_{\omega N})}\right|$$

$$(3.3.15) \qquad \le |t|\sigma_N^{-1}E|T_{\omega N}-ET_{\omega N}-S_{\omega N}+ES_{\omega N}| = \mathcal{O}(|t|N^{-1-7\delta/15}\{1+(\sum_{j\in I} c_{d_j})^2\}^{\frac{1}{2}}),$$

uniformly in t and ω.

Our next task is to evaluate $E\exp\{it\sigma_N^{-1}(S_{\omega N}-ES_{\omega N})\}$. The technique for doing this resembles that in HELMERS (1980). Let χ be the indicator function of $(0,\infty)$ and define

$$S_1 = \sum_{j=1}^{M} b_j(J_N(\hat{V}_j) - EJ_N(\hat{V}_j)) = \sum_{j=1}^{M} b_j\tilde{J}_N(\hat{V}_j),$$

$$S_2 = \frac{1}{M+1} \sum\sum_{(j,k)\neq} b_j J_N'(\hat{V}_j)(\chi(V_j - V_k) - V_j),$$

(3.3.16)
$$S_3 = \frac{1}{2(M+1)^2} \sum\sum_{(j,k)\neq} b_j\{J_N''(\hat{V}_j)(\chi(V_j-V_k) - V_j)^2 - EJ_N''(\hat{V}_j)(\chi(V_j-V_k) - V_j)^2\},$$

$$S_4 = \frac{1}{2(M+1)^2} \sum\sum\sum_{(j,k,\ell)\neq} b_j J_N''(\hat{V}_j)(\chi(V_j-V_k) - V_j)(\chi(V_j-V_\ell) - V_j).$$

It is easy to see that $S_{\omega N} - ES_{\omega N} = \Sigma_{\nu=1}^4 S_\nu$ and $ES_\nu = 0$ for $\nu = 1,2,3,4$. First of all we compute a number of moments.

<u>LEMMA 3.3.2</u>. *Under the Assumptions* (3A) *and* (3B) *we have, uniformly in* ω,

(3.3.17)
$$E|S_2|^3 = \mathcal{O}(N^{-13/10-7\delta/5}), \quad ES_3^2 = \mathcal{O}(N^{-22/15-14\delta/15}),$$
$$ES_4^2 = \mathcal{O}(N^{-7/5-14\delta/15}),$$

with δ *as in* (3.1.13) *and* (3.1.14).

<u>PROOF</u>. By applying Hölder's inequality we obtain $E|S_2|^3 \leq \{ES_2^4\}^{3/4}$. Let for distinct j and k, $h(V_j,V_k) = J_N'(\hat{V}_j)(\chi(V_j-V_k) - V_j)$. Define $h(x,x) = 0$ for all $0 < x < 1$. Direct computation of ES_2^4 shows that

(3.3.18)

$$ES_2^4 = \frac{1}{(M+1)^4} E\{\sum_{j=1}^M b_j \sum_{k=1}^M h(V_j,V_k)\}^4$$

$$= \frac{1}{(M+1)^4} [\sum_{j=1}^M b_j^4\{\sum_{r=1}^M \sum_{s=1}^M \sum_{t=1}^M \sum_{u=1}^M Eh(V_1,V_r)h(V_1,V_s)h(V_1,V_t)h(V_1,V_u)\}$$

$$+ 4 \sum\sum_{(j,k)\neq} b_j^3 b_k\{\sum_{r=1}^M \sum_{s=1}^M \sum_{t=1}^M \sum_{u=1}^M Eh(V_1,V_r)h(V_1,V_s)h(V_1,V_t)h(V_2,V_u)\}$$

$$+ 3 \sum\sum_{(j,k)\neq} b_j^2 b_k^2\{\sum_{r=1}^M \sum_{s=1}^M \sum_{t=1}^M \sum_{u=1}^M Eh(V_1,V_r)h(V_1,V_s)h(V_2,V_t)h(V_2,V_u)\}$$

$$+ 6 \sum\sum\sum_{(j,k,\ell)\neq} b_j^2 b_k b_\ell\{\sum_{r=1}^M \sum_{s=1}^M \sum_{t=1}^M \sum_{u=1}^M Eh(V_1,V_r)h(V_1,V_s)h(V_2,V_t)h(V_3,V_u)\}$$

$$+ \sum\sum\sum\sum_{(j,k,\ell,n)\neq} b_j b_k b_\ell b_n\{\sum_{r=1}^M \sum_{s=1}^M \sum_{t=1}^M \sum_{u=1}^M Eh(V_1,V_r)h(V_2,V_s)h(V_3,V_t)h(V_4,V_u)\}].$$

To bound the right-hand side of (3.3.18) we note that an expectation in (3.3.18) equals zero if at least one of the indices (r,s,t,u) occurs only once. With the aid of the Cauchy-Schwarz inequality the non-zero expectations may be bounded by either $Eh^4(V_1,V_2)$, $\{Eh^4(V_1,V_2)\}^{\frac{1}{2}}Eh^2(V_1,V_2)$ or $\{Eh^2(V_1,V_2)\}^2$ and we obtain

$$\begin{aligned} ES_2^4 &= O(\{\sum_{j=1}^{M} b_j^4\}N^{-2}Eh^4(V_1,V_2) \\ &+ \{\sum_{(j,k)\neq}\sum b_j^3 b_k\}\{N^{-2}\{Eh^4(V_1,V_2)\}^{\frac{1}{2}}Eh^2(V_1,V_2) + N^{-3}Eh^4(V_1,V_2)\} \\ &+ \{\sum_{(j,k)\neq}\sum b_j^2 b_k^2 + \sum_{(j,k,\ell)\neq}\sum\sum b_j^2 b_k b_\ell + \sum_{(j,k,\ell,n)\neq}\sum\sum\sum b_j b_k b_\ell b_n\} \\ &\cdot\{N^{-2}\{Eh^2(V_1,V_2)\}^2 + N^{-3}Eh^4(V_1,V_2)\}). \end{aligned} \tag{3.3.19}$$

In view of (3.3.8) and Assumption (3B) we have, for $1 \le k \le 4$,

$$\begin{aligned} E|h^k(V_1,V_2)| &= E|J_N'(\hat{V}_1)|^k|\chi(V_1-V_2) - V_1|^k \\ &\le 2\,E\,|J_N'(\hat{V}_1)|^k V_1(1-V_1) \le 2E\left|J'\left(\frac{m+1+(M-1)V_1}{N+1}\right)\right|^k V_1(1-V_1) \\ &= O\left(\int_{\frac{m+1}{N+1}}^{1-\frac{m+1}{N+1}} \{t(1-t)\}^{-15k/14+k\delta}\{\frac{(N+1)t-(m+1)}{M-1}\}\{\frac{(M+m)-(N+1)t}{M-1}\}dt\right) \\ &= O\left(\int_{\frac{m}{N}}^{1-\frac{m}{N}} \{t(1-t)\}^{1-15k/14+k\delta}dt\right) = O(N^{k/2-14/15-7k\delta/15}). \end{aligned} \tag{3.3.20}$$

According to Assumption (3A) and the fact that $\{b_1,b_2,\ldots,b_M\} = \{c_1,c_2,\ldots \ldots,c_N\}\backslash\{c_{d_j}\colon j \in I\}$, we have

$$|\sum_{j=1}^{M} b_j| = |\sum_{j\in I} c_{d_j}| = O\left(\frac{m}{N^{\frac{1}{2}}}\right) = O(N^{1/30})$$

and similarly

$$\sum_{j=1}^{M} b_j^4 = O(N^{-1}), \qquad |\sum\sum_{(j,k)\neq} b_j^3 b_k| = O(N^{-7/15}),$$

(3.3.21) $$\sum\sum_{(j,k)\neq} b_j^2 b_k^2 = 1 + O(N^{-7/15}), \qquad |\sum\sum\sum_{(j,k,\ell)\neq} b_j^2 b_k b_\ell| = O(N^{1/15}),$$

$$|\sum\sum\sum\sum_{(j,k,\ell,n)\neq} b_j b_k b_\ell b_n| = O(N^{2/15}).$$

Combining (3.3.19) through (3.3.21) we find that $ES_2^4 = O(N^{-26/15-28\delta/15})$ and hence $E|S_2|^3 = O(N^{-13/10-7\delta/5})$. In the same way one can obtain the other two assertions in (3.3.17). □

Define, for real t and $N \geq 2$,

(3.3.22) $$\rho_N(t) = Ee^{it\sigma_N^{-1}(S_{\omega N}-ES_{\omega N})}$$

and

(3.3.23) $$\rho_{1N}(t) = Ee^{it\sigma_N^{-1}S_1}\{1 + \frac{it}{\sigma_N}(S_2+S_3+S_4) + \frac{(it)^2}{2\sigma_N^2} S_2^2\}.$$

The next lemma shows that ρ_N can be approximated by ρ_{1N}.

<u>LEMMA 3.3.3</u>. *Under the Assumptions* (3A) *and* (3B) *we have, uniformly for* $|t| \leq \log N$ *and* ω

(3.3.24) $$|\rho_N(t) - \rho_{1N}(t)| = O(t^2 N^{-17/15-14\delta/15}),$$

with δ *as in* (3.1.13) *and* (3.1.14).

<u>PROOF</u>. Repeated use of Lemma XV 4.1 of FELLER (1971) yields

$$|\rho_N(t)-\rho_{1N}(t)| = O(t^2\sigma_N^{-2}E|S_2||S_3+S_4| + t^2\sigma_N^{-2}E(S_3^2+S_4^2) + |t|^3\sigma_N^{-3}E|S_2|^3).$$

From (3.1.6) and (3.2.7) it follows that for all sufficiently large N there exist positive numbers $\varepsilon_1 \leq \varepsilon_2$ such that $\varepsilon_1 \leq \sigma_N^2 \leq \varepsilon_2$. Lemma 3.3.2 produces the desired result. □

Clearly, our next task is to evaluate the right-hand side of (3.3.23) and we start with the leading term. According to (3.3.16) $S_1 = \Sigma_{j=1}^M b_j\tilde{J}_N(\hat{V}_j)$.

We have $E\tilde{J}_N(\hat{V}_1) = 0$ and for all sufficiently large N, there exist positive numbers $\gamma_1 \leq \gamma_2$ such that $\gamma_1 \leq E\tilde{J}_N^2(\hat{V}_1) \leq \gamma_2$ (cf. (3.1.4)). In the sequel we shall assume

$$(3.3.25) \qquad \sum_{j \in I} c_{d_j}^2 < 1 - \gamma$$

for some $\gamma \in (0,1)$, to guarantee that

$$(3.3.26) \qquad \gamma\gamma_1 \leq \sigma^2(S_1) \leq \gamma_2 .$$

Finally we note that Assumptions (3A) and (3B) imply that $\Sigma_{j=1}^M b_j^4 = O(N^{-1})$ and that the random variable $\tilde{J}_N(\hat{V}_1)$ has a finite 14-th absolute moment. It follows from the classical theory of Edgeworth expansions for sums of independent and non-identically distributed random variables (see e.g. Lemma VI 4.11 in PETROV (1975)) that

$$(3.3.27) \qquad \begin{aligned} &|E \exp\{itS_1/\sigma(S_1)\} - e^{-\frac{1}{2}t^2}\{1 - \frac{it^3}{6\sigma^3(S_1)} \sum_{j=1}^M b_j^3 E\tilde{J}_N^3(\hat{V}_1) \\ &+ \frac{t^4}{24\sigma^4(S_1)} \sum_{j=1}^M b_j^4\{E\tilde{J}_N^4(\hat{V}_1) - 3[E\tilde{J}_N^2(\hat{V}_1)]^2\} \\ &- \frac{t^6}{72\sigma^6(S_1)} \{\sum_{j=1}^M b_j^3 E\tilde{J}_N^3(\hat{V}_1)\}^2\}| = o(N^{-1}(t^4 + |t|^9)e^{-\frac{1}{2}t^2}), \end{aligned}$$

uniformly for $|t| \leq \log N$ and ω for which (3.3.25) is satisfied. Replacing t by $t_N = t\sigma(s_1)/\sigma_N$ and expanding $\exp\{-\frac{1}{2}t_N^2\}$ we find that uniformly for $|t| \leq \log N$ and ω for which (3.3.25) is satisfied

$$(3.3.28) \qquad \begin{aligned} &|Ee^{it\sigma_N^{-1}S_1} - e^{-\frac{1}{2}t^2}\{1 - \frac{it^3}{6\sigma_N^3} \sum_{j=1}^M b_j^3 E\tilde{J}_N^3(\hat{V}_1) + \frac{t^4}{24\sigma_N^4} \sum_{j=1}^M b_j^4\{E\tilde{J}_N^4(\hat{V}_1) \\ &- 3[E\tilde{J}_N^2(\hat{V}_1)]^2\} - \frac{t^6}{72\sigma_N^6} \{\sum_{j=1}^M b_j^3 E\tilde{J}_N^3(\hat{V}_1)\}^2 + \frac{t^2}{2\sigma_N^2} (\sigma_N^2 - \sigma^2(S_1)) \\ &+ \frac{t^4}{8\sigma_N^4} (\sigma_N^2 - \sigma^2(S_1))^2 - \frac{it^5}{12\sigma_N^5} (\sigma_N^2 - \sigma^2(S_1)) \sum_{j=1}^M b_j^3 E\tilde{J}_N^3(\hat{V}_1)\}| \\ &= o(N^{-1}(t^4 + |t|^9)e^{-\frac{1}{2}t^2}) + O(|\sigma_N^2 - \sigma^2(S_1)|^3 |t| P_1(t) e^{-\theta t^2}) \\ &+ O(N^{-1}|\sigma_N^2 - \sigma^2(S_1)| |t| P_2(t) e^{-\theta t^2}), \end{aligned}$$

where $0 < \theta < \frac{1}{2}$ and P_1 and P_2 are fixed polynomials.

We now turn to the remaining terms on the right in (3.3.23). Let

$$(3.3.29) \qquad \mu_N(t) = Ee^{it\tilde{J}_N(\hat{V}_1)}$$

denote the characteristic function of $\tilde{J}_N(\hat{V}_1)$, so that

$$(3.3.30) \qquad Ee^{it\sigma_N^{-1}S_1} = \prod_{j=1}^{M} \mu_N\left(\frac{b_j t}{\sigma_N}\right).$$

From the Assumptions (3A) and (3B) it follows by Taylor expansion that for distinct integers $\ell_1,\ldots,\ell_n$ where $1 \le n \le 4$

$$(3.3.31) \qquad \prod_{\nu=1}^{n} \mu_N\left(\frac{b_{\ell_\nu} t}{\sigma_N}\right) = 1 - \frac{t^2}{2\sigma_N^2}\{\sum_{\nu=1}^{n} b_{\ell_\nu}^2\}E\tilde{J}_N^2(\hat{V}_1) + O(N^{-3/2}|t|^3),$$

uniformly for $|t| \le \log N$ and ω for which (3.3.25) is satisfied.

In the last two lemmas of this section we summarize the results we need.

LEMMA 3.3.4. *Under the Assumptions* (3A) *and* (3B) *we have, uniformly for* $|t| \le \log N$ *and* ω *for which* (3.3.25) *is satisfied*

$$(3.3.32) \qquad \begin{aligned} &|E(e^{it\sigma_N^{-1}S_1}S_2) - Ee^{it\sigma_N^{-1}S_1}\{\frac{it}{\sigma_N}ES_1S_2 + \frac{(it)^2}{2\sigma_N^2}ES_1^2S_2 \\ &- \frac{(it)^3}{4N\sigma_N^3}[E\tilde{J}_N^4(\hat{V}_1) - \{E\tilde{J}_N^2(\hat{V}_1)\}^2]\}| = O(N^{-1-\varepsilon}|t|P(t)e^{-\theta t^2}), \end{aligned}$$

$$(3.3.33) \qquad |E(e^{it\sigma_N^{-1}S_1}S_3) - Ee^{it\sigma_N^{-1}S_1}\{\frac{it}{\sigma_N}ES_1S_3\}| = O(N^{-1-\varepsilon}|t|P(t)e^{-\theta t^2}),$$

$$(3.3.34) \qquad |E(e^{it\sigma_N^{-1}S_1}S_4)| = O(N^{-1-\varepsilon}|t|P(t)e^{-\theta t^2}),$$

where $0 < \theta < \frac{1}{2}$, $\varepsilon > 0$ and P is a fixed polynomial.

PROOF. Because the statements (3.3.32) through (3.3.34) are all proved in essentially the same manner, we shall only prove the first statement, by way of an example. An application of Lemma XV 4.1 of FELLER (1971) shows

$$|\exp\{it\sigma_N^{-1}(b_j\tilde{J}_N(\hat{V}_j) + b_k\tilde{J}_N(\hat{V}_k))\} - 1 - \frac{it}{\sigma_N}(b_j\tilde{J}_N(\hat{V}_j) + b_k\tilde{J}_N(\hat{V}_k))$$
$$- \frac{(it)^2}{2\sigma_N^2}(b_j\tilde{J}_N(\hat{V}_j) + b_k\tilde{J}_N(\hat{V}_k))^2 - \frac{(it)^3}{6\sigma_N^3}(b_j\tilde{J}_N(\hat{V}_j) + b_k\tilde{J}_N(\hat{V}_k))^3|$$
$$\leq \frac{t^4}{\sigma_N^4}(b_j\tilde{J}_N(\hat{V}_j) + b_k\tilde{J}_N(\hat{V}_k))^4.$$

It follows that

$$E\ \exp\{it\sigma_N^{-1}(b_j\tilde{J}_N(\hat{V}_j) + b_k\tilde{J}_N(\hat{V}_k))\}J_N'(\hat{V}_j)(\chi(V_j-V_k) - V_j)$$

$$= E[J_N'(\hat{V}_j)(\chi(V_j-V_k) - V_j)][\frac{it}{\sigma_N}(b_j\tilde{J}_N(\hat{V}_j) + b_k\tilde{J}_N(\hat{V}_k))$$

(3.3.35)
$$+ \frac{(it)^2}{2\sigma_N^2}(b_j^2\tilde{J}_N^2(\hat{V}_j) + 2b_jb_k\tilde{J}_N(\hat{V}_j)\tilde{J}_N(\hat{V}_k) + b_k^2\tilde{J}_N^2(\hat{V}_k))$$

$$+ \frac{(it)^3}{6\sigma_N^3}(b_j^3\tilde{J}_N^3(\hat{V}_j) + 3b_j^2b_k\tilde{J}_N^2(\hat{V}_j)\tilde{J}_N(\hat{V}_k) + 3b_jb_k^2\tilde{J}_N(\hat{V}_j)\tilde{J}_N^2(\hat{V}_k) + b_k^3\tilde{J}_N^3(\hat{V}_k))]$$

$$+ O(t^4E|J_N'(\hat{V}_j)(\chi(V_j-V_k) - V_j)|\{b_j^4\tilde{J}_N^4(\hat{V}_j) + b_k^4\tilde{J}_N^4(\hat{V}_k)\}).$$

We note that it is easy to check that

$$E[J_N'(\hat{V}_j)(\chi(V_j-V_k) - V_j)][\sum_{\ell\neq j,k}(\frac{it}{\sigma_N}b_\ell\tilde{J}_N(\hat{V}_\ell) + \frac{(it)^2}{2\sigma_N^2}b_\ell^2\tilde{J}_N^2(\hat{V}_\ell)$$

(3.3.36)
$$+ \frac{(it)^2}{\sigma_N^2}b_\ell\tilde{J}_N(\hat{V}_\ell)\{b_j\tilde{J}_N(\hat{V}_j) + b_k\tilde{J}_N(\hat{V}_k) + \frac{1}{2}\sum_{n\neq j,k,\ell}b_n\tilde{J}_N(\hat{V}_n)\})$$

$$- \frac{(it)^3}{6\sigma_N^3}b_j^3\tilde{J}_N^3(\hat{V}_j)] = 0$$

and hence

$$E\ \exp\{it\sigma_N^{-1}(b_j\tilde{J}_N(\hat{V}_j)+b_k\tilde{J}_N(\hat{V}_k))\}J_N'(\hat{V}_j)(\chi(V_j-V_k)-V_j)$$

$$=E[J_N'(\hat{V}_j)(\chi(V_j-V_k)-V_j)][\frac{it}{\sigma_N}S_1+\frac{(it)^2}{2\sigma_N^2}S_1^2$$

$$+\frac{(it)^3}{6\sigma_N^3}\{3b_j^2b_k\tilde{J}_N^2(\hat{V}_j)\tilde{J}_N(\hat{V}_k)+3b_jb_k^2\tilde{J}_N(\hat{V}_j)\tilde{J}_N^2(\hat{V}_k)+b_k^3\tilde{J}_N^3(\hat{V}_k)\}]$$

$$+\mathcal{O}(t^4E|J_N'(\hat{V}_j)(\chi(V_j-V_k)-V_j)|\{b_j^4\tilde{J}_N^4(\hat{V}_j)+b_k^4\tilde{J}_N^4(\hat{V}_k)\}).$$

From (3.3.31) it follows that for distinct integers $1\le j,k\le M$ and $|t|\le\log N$

$$\prod_{\ell\ne j,k}\mu_N\left(\frac{b_\ell t}{\sigma_N}\right)=Ee^{it\sigma_N^{-1}S_1}\{1+\frac{t^2}{2\sigma_N^2}(b_j^2+b_k^2)E\tilde{J}_N^2(\hat{V}_1)+\mathcal{O}(N^{-3/2}|t|^3)\},\tag{3.3.37}$$

uniformly for $|t|\le\log N$ and ω for which (3.3.25) is satisfied. Hence, combining (3.3.35) through (3.3.37) and Assumption (3A), we find after some algebra

$$E(e^{it\sigma_N^{-1}S_1}S_2)=\sum_{(j,k)\ne}\sum\frac{b_j}{M+1}\prod_{\ell\ne j,k}\mu_N\left(\frac{b_\ell t}{\sigma_N}\right)$$

$$\cdot E\ \exp\{it\sigma_N^{-1}(b_j\tilde{J}_N(\hat{V}_j)+b_k\tilde{J}_N(\hat{V}_k))\}J_N'(\hat{V}_j)(\chi(V_j-V_k)-V_j)$$

$$=Ee^{it\sigma_N^{-1}S_1}[\frac{it}{\sigma_N}ES_1S_2+\frac{(it)^2}{2\sigma_N^2}ES_1^2S_2$$

$$+\frac{(it)^3}{6\sigma_N^3}\sum_{(j,k)\ne}\sum E[J_N'(\hat{V}_j)(\chi(V_j-V_k)-V_j)][\frac{3b_j^3b_k}{M+1}\tilde{J}_N^2(\hat{V}_j)\tilde{J}_N(\hat{V}_k)\tag{3.3.38}$$

$$+\frac{3b_j^2b_k^2}{M+1}\tilde{J}_N(\hat{V}_j)\tilde{J}_N^2(\hat{V}_k)+\frac{b_jb_k^3}{M+1}\tilde{J}_N^3(\hat{V}_k)]+\frac{(it)^3}{2\sigma_N^3}\sum_{(j,k)\ne}\sum\frac{b_jb_k(b_j^2+b_k^2)}{M+1}$$

$$\cdot\{E\tilde{J}_N(\hat{V}_k)J_N'(\hat{V}_j)(\chi(V_j-V_k)-V_j)\}\{E\tilde{J}_N^2(\hat{V}_1)\}]$$

$$+\mathcal{O}(N^{-3/2}t^4e^{-\theta t^2}E\{|\tilde{J}_N(\hat{V}_1)|+\tilde{J}_N^4(\hat{V}_1)\}\{|J_N'(\hat{V}_1)|+|J_N'(\hat{V}_2)|\}),$$

uniformly for $|t| \le \log N$ and ω for which (3.3.25) is satisfied. From Assumption (3B) and (3.3.8) it follows that (see also (3.3.20))

$$
\begin{aligned}
&E|\tilde{J}_N^2(\hat{V}_1)\tilde{J}_N(\hat{V}_2)J_N'(\hat{V}_1)| = O(N^{1/10-7\delta/5});\\
&E|\tilde{J}_N(\hat{V}_1)J_N'(\hat{V}_1)| = O(N^{1/15-14\delta/15});\\
(3.3.39)\qquad &E|\tilde{J}_N^4(\hat{V}_1)J_N'(\hat{V}_1)| = O(N^{1/6-7\delta/3});\quad E\tilde{J}_N^2(\hat{V}_1) = O(1);\\
&E(\tilde{J}_N^4(\hat{V}_1) + |\tilde{J}_N^3(\hat{V}_1)| + |\tilde{J}_N(\hat{V}_1)|)|J_N'(\hat{V}_2)| = O(N^{1/30-7\delta/15}).
\end{aligned}
$$

Finally we obtain by partial integration

$$
\begin{aligned}
&E\tilde{J}_N(\hat{V}_1)\tilde{J}_N^2(\hat{V}_2)J_N'(\hat{V}_1)(\chi(V_1-V_2)-V_1)\\
(3.3.40)\qquad &= -\tfrac{1}{2}\left(\tfrac{M+1}{M-1}\right)^2 E\tilde{J}_N^4(\hat{V}_1) + \tfrac{1}{2}\left(\tfrac{M+1}{M-1}\right)^3\{E\tilde{J}_N^2(\hat{V}_1)\}^2.
\end{aligned}
$$

Combining (3.3.38) through (3.3.40) and (3.3.21) we arrive at (3.3.32). □

<u>LEMMA 3.3.5</u>. *Under the Assumptions* (3A) *and* (3B) *we have, uniformly for* $|t| \le \log N$ *and* ω *for which* (3.3.25) *is satisfied,*

$$
\begin{aligned}
&|E(e^{it\sigma_N^{-1}S_1}S_2^2) - Ee^{it\sigma_N^{-1}S_1}\{ES_2^2 + \frac{it}{\sigma_N}ES_1S_2^2\\
(3.3.41)\qquad &+ \frac{(it)^2}{4N\sigma_N^2}[E\tilde{J}_N^4(\hat{V}_1) - \{E\tilde{J}_N^2(\hat{V}_1)\}^2]\}| = O(N^{-1-\varepsilon}|t|P(t)e^{-\theta t^2}),
\end{aligned}
$$

where $0 < \theta < \frac{1}{2}$, $\varepsilon > 0$ *and* P *is a fixed polynomial.*

<u>PROOF</u>. The proof of the statement (3.3.41) is similar to that of Lemma 3.3.4 and we shall only provide a sketch. Throughout, all order symbols will be uniform for $|t| \le \log N$ and ω for which (3.3.25) is satisfied. Let, for distinct j and k, $h(V_j,V_k) = J_N'(\hat{V}_j)(\chi(V_j-V_k) - V_j)$. Direct computation of $E(\exp\{it\sigma_N^{-1}S_1\}S_2^2)$ shows

$$E(e^{it\sigma_N^{-1}S_1}S_2^2) = E(e^{it\sigma_N^{-1}S_1}\{\sum_{j=1}^{M}\frac{b_j}{M+1}\sum_{k\neq j}h(V_j,V_k)\}^2)$$

$$= \frac{1}{(M+1)^2}[\sum_{j=1}^{M} b_j^2\{\sum_{r\neq j} E\exp\{it\sigma_N^{-1}(b_j\tilde{J}_N(\hat{V}_j)+b_r\tilde{J}_N(\hat{V}_r))\}h^2(V_j,V_r)$$

$$\cdot\prod_{\ell\neq j,r}\mu_N\left(\frac{b_\ell t}{\sigma_N}\right) + \sum_{r\neq j}\sum_{s\neq j,r} E\exp\{it\sigma_N^{-1}(b_j\tilde{J}_N(\hat{V}_j)+b_r\tilde{J}_N(\hat{V}_r))\}$$

$$\cdot\exp\{it\sigma_N^{-1}b_s\tilde{J}_N(\hat{V}_s)\}h(V_j,V_r)h(V_j,V_s)\prod_{\ell\neq j,r,s}\mu_N\left(\frac{b_\ell t}{\sigma_N}\right)\}$$

(3.3.42)
$$+ \sum_{(j,k)\neq}\sum b_j b_k\{E\exp\{it\sigma_N^{-1}(b_j\tilde{J}_N(\hat{V}_j)+b_k\tilde{J}_N(\hat{V}_k))\}h(V_j,V_k)h(V_k,V_j)$$

$$\cdot\prod_{\ell\neq j,k}\mu_N\left(\frac{b_\ell t}{\sigma_N}\right) + \sum_{r\neq j,k} E\exp\{it\sigma_N^{-1}(b_j\tilde{J}_N(\hat{V}_j)+b_k\tilde{J}_N(\hat{V}_k)+b_r\tilde{J}_N(\hat{V}_r))\}$$

$$\cdot[h(V_j,V_r)h(V_k,V_r)+2h(V_j,V_r)h(V_k,V_j)]\prod_{\ell\neq j,k,r}\mu_N\left(\frac{b_\ell t}{\sigma_N}\right)$$

$$+ \sum_{r\neq j,k}\sum_{s\neq j,k,r} E\exp\{it\sigma_N^{-1}(b_j\tilde{J}_N(\hat{V}_j)+b_r\tilde{J}_N(\hat{V}_r))\}h(V_j,V_r)$$

$$\cdot E\exp\{it\sigma_N^{-1}(b_k\tilde{J}_N(\hat{V}_k)+b_s\tilde{J}_N(\hat{V}_s))\}h(V_k,V_s)\prod_{\ell\neq j,k,r,s}\mu_N\left(\frac{b_\ell t}{\sigma_N}\right)\}].$$

Using Lemma XV 4.1 of FELLER (1971), we expand all six exponents in the right-hand side of (3.3.42) (cf. (3.3.35)). From (3.3.31) it follows that for distinct integers $\ell_1,\ldots,\ell_n$ where $1 \le n \le 4$ we have (cf. (3.3.37))

(3.3.43)
$$\prod_{\ell\neq\ell_1,\ldots,\ell_n}\mu_N\left(\frac{b_\ell t}{\sigma_N}\right) = Ee^{it\sigma_N^{-1}S_1}\{1+\frac{t^2}{2\sigma_N^2}\{\sum_{\nu=1}^{n} b_{\ell_\nu}^2\}E\tilde{J}_N^2(\hat{V}_1)$$
$$+ O(N^{-3/2}|t|^3)\}.$$

With the aid of (3.3.43) and the expansions of the exponents we proceed as in (3.3.38). For example, the term involving $h(V_j,V_r)h(V_k,V_r)$ on the right in (3.3.42) equals

$$\frac{1}{(M+1)^2} \sum\sum\sum_{(j,k,r)\neq} b_j b_k \prod_{\ell\neq j,k,r} \mu_N\left(\frac{b_\ell t}{\sigma_N}\right) E \exp\{it\sigma_N^{-1} b_j \tilde{J}_N(\hat{V}_j)\}$$

$$\cdot \exp\{it\sigma_N^{-1}(b_k \tilde{J}_N(\hat{V}_k) + b_r \tilde{J}_N(\hat{V}_r))\} h(V_j, V_r) h(V_k, V_r)$$

$$= Ee^{it\sigma_N^{-1}S_1}\{ \sum\sum\sum_{(j,k,r)\neq} \frac{b_j b_k}{(M+1)^2} [Eh(V_j, V_r) h(V_k, V_r)$$

$$+ \frac{it}{\sigma_N} ES_1 h(V_j, V_r) h(V_k, V_r) + \frac{(it)^2}{2\sigma_N^2} E(b_j^2 \tilde{J}_N^2(\hat{V}_j) + b_k^2 \tilde{J}_N^2(\hat{V}_k) + b_r^2 \tilde{J}_N^2(\hat{V}_r)$$

$$+ 2b_j b_k \tilde{J}_N(\hat{V}_j) \tilde{J}_N(\hat{V}_k) + 2b_j b_r \tilde{J}_N(\hat{V}_j) \tilde{J}_N(\hat{V}_r) + 2b_k b_r \tilde{J}_N(\hat{V}_k) \tilde{J}_N(\hat{V}_r))$$

$$\cdot h(V_j, V_r) h(V_k, V_r)] + \frac{t^2}{2\sigma_N^2} \sum\sum\sum_{(j,k,r)\neq} \frac{b_j b_k}{(M+1)^2} (b_j^2 + b_k^2 + b_r^2)$$

$$\cdot E\tilde{J}_N^2(\hat{V}_1) Eh(V_j, V_r) h(V_k, V_r)\}$$

$$+ O(N^{-3/2} |t|^3 e^{-\theta t^2} E\{|\tilde{J}_N^3(\hat{V}_1)| + |\tilde{J}_N(\hat{V}_1)| + |\tilde{J}_N^3(\hat{V}_3)| + |\tilde{J}_N(\hat{V}_3)| + 1\}$$

$$\cdot |J_N'(\hat{V}_1) J_N'(\hat{V}_2)|).$$

From the Assumptions (3A) and (3B) and (3.3.8) we are able to calculate these sums (cf. (3.3.21) and (3.3.39)). Note that by partial integration we have

$$E\tilde{J}_N(\hat{V}_1) \tilde{J}_N(\hat{V}_2) h(V_1, V_3) h(V_2, V_3) = \frac{1}{4}\left(\frac{M+1}{M-1}\right)^2 [E\tilde{J}_N^4(\hat{V}_1) - \{E\tilde{J}_N^2(\hat{V}_1)\}^2].$$

Following this programme, we finally arrive at

$$\frac{1}{(M+1)^2} \sum\sum\sum_{(j,k,r)\neq} b_j b_k \prod_{\ell\neq j,k,r} \mu_N\left(\frac{b_\ell t}{\sigma_N}\right) E\, \exp\{it\sigma_N^{-1} b_j \tilde{J}_N(\hat{V}_j)\}$$

$$\cdot \exp\{it\sigma_N^{-1}(b_k\tilde{J}_N(\hat{V}_k) + b_r\tilde{J}_N(\hat{V}_r))\} h(V_j,V_r) h(V_k,V_r)$$

$$= Ee^{it\sigma_N^{-1}S_1}\{\sum\sum\sum_{(j,k,r)\neq} \frac{b_j b_k}{(M+1)^2}[Eh(V_j,V_r)h(V_k,V_r)$$

$$+ \frac{it}{\sigma_N} ES_1 h(V_j,V_r)h(V_k,V_r)] + \frac{(it)^2}{4\sigma_N^2} \sum\sum_{(j,k)\neq} \frac{b_j^2 b_k^2}{N}[E\tilde{J}_N^4(\hat{V}_1) - \{E\tilde{J}_N^2(\hat{V}_1)\}^2]\}$$

$$+ O(N^{-1-\varepsilon}|t|P(t)e^{-\theta t^2}),$$

where $0 < \theta < \frac{1}{2}$, $\varepsilon > 0$ and P is a fixed polynomial. All other terms in the right-hand side of (3.3.42) can be handled in the same way. □

From Lemmas 3.3.4 and 3.3.5 it follows that uniformly for $|t| \leq \log N$ and ω for which (3.3.25) is satisfied (cf. (3.3.23)),

$$\rho_{1N}(t) = Ee^{it\sigma_N^{-1}S_1}\{1 + \frac{(it)^2}{2\sigma_N^2}[2ES_1S_2 + 2ES_1S_3 + ES_2^2]$$

$$+ \frac{(it)^3}{2\sigma_N^3}[ES_1^2S_2 + ES_1S_2^2] - \frac{(it)^4}{8N\sigma_N^4}[E\tilde{J}_N^4(\hat{V}_1) - \{E\tilde{J}_N^2(\hat{V}_1)\}^2]\}$$

$$+ O(N^{-1-\varepsilon}|t|P(t)e^{-\theta t^2}),$$

where $\varepsilon > 0$, $0 < \theta < \frac{1}{2}$ and P is a fixed polynomial. Using (3.3.26), Lemmas 3.3.1 and 3.3.2, as well as the fact that $ES_1S_4 = 0$, we obtain

$$2ES_1S_2 + 2ES_1S_3 + ES_2^2 = \sigma^2(S_{\omega N}) - \sigma^2(S_1) + O(N^{-17/15-14\delta/15})$$

(3.3.44)

$$= \sigma^2(T_{\omega N}) - \sigma^2(S_1) + (1 + \{\sum_{j\in I} c_{d_j}\}^2)^{\frac{1}{2}} O(N^{-1-7\delta/15}),$$

uniformly for ω satisfying (3.3.25). Writing $h(V_1,V_2) = J_N'(\hat{V}_1)(\chi(V_1-V_2) - V_1)$ as before, we find by repeated use of Assumptions (3A) and (3B) (cf. (3.3.20), (3.3.21) and (3.3.39)) that, uniformly for ω satisfying (3.3.25),

$$ES_1^2S_2 + ES_1S_2^2 = \frac{A_{1N}}{N} \sum_{(j,k)\neq}\sum b_j^2 b_k + O(N^{-1-\varepsilon}),$$

where $\varepsilon > 0$ and

$$A_{1N} = E\tilde{J}_N^2(\hat{V}_1)h(V_2,V_1) + 2E\tilde{J}_N(\hat{V}_1)\tilde{J}_N(\hat{V}_2)h(V_1,V_2) + 2E\tilde{J}_N(\hat{V}_1)h(V_1,V_3)h(V_2,V_3). \tag{3.3.45}$$

It follows that uniformly for $|t| \le \log N$ and ω satisfying (3.3.25),

$$\rho_{1N}(t) = Ee^{it\sigma_N^{-1}S_1}\{1 + \frac{(it)^2}{2\sigma_N^2}[\sigma^2(T_{\omega N}) - \sigma^2(S_1)] + \frac{(it)^3}{2\sigma_N^3}\frac{A_{1N}}{N}\sum_{(j,k)\neq}\sum b_j^2 b_k - \frac{(it)^4}{8N\sigma_N^4}[E\tilde{J}_N^4(\hat{V}_1) - \{\tilde{E}J_N^2(\hat{V}_1)\}^2]\} + O(N^{-1-\varepsilon}|t|P(t)e^{-\theta t^2}(1 + \{\sum_{j\in I} c_{d_j}\}^2)^{\frac{1}{2}}), \tag{3.3.46}$$

where $\varepsilon > 0$, $0 < \theta < \frac{1}{2}$ and P is a fixed polynomial.

Let us return to our starting point (3.3.7). Choose $\gamma \in (0,1)$ and define the event $B = \{\Sigma_{j\in I} c_{D_j}^2 < 1-\gamma\}$ (cf. (3.3.25)). According to Lemma 3.2.4, $P(B^c) = O(N^{-22/15})$, so

$$\psi_N^*(t) = Ee^{itT_N^*} = E[\chi(B)E(e^{it\sigma_N^{-1}\{T_N-Z_N-E(T_N-Z_N|\Omega)\}}|\Omega) \cdot e^{it\sigma_N^{-1}E(T_N-Z_N|\Omega)}E(e^{it\sigma_N^{-1}Z_N}|\Omega)] + O(N^{-22/15}).$$

From Lemma 3.2.3 it follows that $E|Z_N|^5 = O(N^{-1-7\delta/3})$ and $E(E(T_N-Z_N|\Omega))^2 = O(N^{-4/3-14\delta/15})$. Hence by Taylor expansion we obtain

$$\psi_N^*(t) = Ee^{itT_N^*} = E[\chi(B)E(e^{it\sigma_N^{-1}\{T_N-Z_N-E(T_N-Z_N|\Omega)\}}|\Omega)$$

$$\cdot\{1 + \frac{it}{\sigma_N}\{E(Z_N|\Omega) + E(T_N-Z_N|\Omega)\} + \frac{(it)^2}{2\sigma_N^2}\{E(Z_N^2|\Omega) + 2E(Z_N|\Omega)E(T_N-Z_N|\Omega)\} + \frac{(it)^3}{6\sigma_N^3}E(Z_N^3|\Omega) + \frac{(it)^4}{24\sigma_N^4}E(Z_N^4|\Omega)\}] + O(N^{-22/15}) + O([t^2+|t|^5]N^{-1-7\delta/3}), \tag{3.3.47}$$

uniformly for $|t| \le \log N$. In view of (3.3.15), (3.3.22) and (3.3.24) we have, uniformly for $|t| \le \log N$ and ω satisfying (3.3.25)

$$E(e^{it\sigma_N^{-1}\{T_N - Z_N - E(T_N - Z_N|\Omega=\omega)\}} \mid \Omega=\omega) = E(e^{it\sigma_N^{-1}(T_{\omega N} - ET_{\omega N})})$$

$$(3.3.48) \qquad = \rho_N(t) + O(|t|N^{-1-7\delta/15}(1+\{\sum_{j\in I} c_{d_j}\}^2)^{\frac{1}{2}})$$

$$= \rho_{1N}(t) + O(N^{-1-\varepsilon}|t|P(t)(1+\{\sum_{j\in I} c_{d_j}\}^2)^{\frac{1}{2}}),$$

where $\varepsilon > 0$ and P is a fixed polynomial.

Before substituting this in (3.3.47) we shall provide uniform bounds for the quantities $\sigma_N^2 - \sigma^2(T_{\omega N})$ and $\sigma^2(T_{\omega N}) - \sigma^2(S_1)$. Theorem II 3.1.c of HÁJEK & ŠIDÁK (1967) and Assumption (3A) imply that

$$\sigma^2(T_{\omega N}) = \frac{1}{M-1}\,(1 - \sum_{j\in I} c_{d_j}^2 - \frac{1}{M}(\sum_{j\in I} c_{d_j})^2)\sum_{j=1}^{M}(J_N(\tfrac{j}{M+1}) - \bar{J}_N)^2,$$

where (cf. (3.3.8))

$$\bar{J}_N = \frac{1}{M}\sum_{j=1}^{M} J_N(\tfrac{j}{M+1}) = \frac{1}{M}\sum_{j=m+1}^{N-m} J(\tfrac{j}{N+1}).$$

It follows from (3.2.12) that $|\bar{J}_N| = O(N^{-13/30-7\delta/15})$ and from Assumption (3A) that $|\Sigma_{j\in I} c_{d_j}| = O(N^{1/30})$, hence

$$(3.3.49) \qquad \sigma^2(T_{\omega N}) = \frac{1}{M-1}\,(1 - \sum_{j\in I} c_{d_j}^2)\sum_{j=1}^{M} J_N^2(\tfrac{j}{M+1}) + O(N^{-13/15-14\delta/15}),$$

uniformly in ω. Furthermore we know from (3.2.4) that $|\bar{J}| = O(N^{-13/14-\delta})$, so in view of (3.1.6) and Assumption (3B) we have

$$|\sigma_N^2 - \sigma^2(T_{\omega N})| = |\frac{1}{N-1}\sum_{j=1}^{N} J^2(\tfrac{j}{N+1}) - \frac{1}{M-1}(1 - \sum_{j\in I} c_{d_j}^2)\sum_{j=1}^{M} J_N^2(\tfrac{j}{M+1})|$$

$$(3.3.50) \qquad + O(N^{-13/15-14\delta/15})$$

$$= |\frac{1}{N-1}\sum_{j\in I} J^2(\tfrac{j}{N+1}) + \frac{1}{M-1}(\sum_{j\in I} c_{d_j}^2 - \frac{2m}{N})\sum_{j=1}^{M} J_N^2(\tfrac{j}{M+1})|$$

$$+ O(N^{-13/15-14\delta/15}) = O(N^{-2/5-14\delta/15}),$$

uniformly in ω.

To obtain the second bound, we argue as in Lemma 3.2.1 with J and h(t) replaced by J_N and $h_N(t) = h((N+1)^{-1}(m+(M+1)t))$ to conclude that

$$\frac{1}{M}\sum_{j=1}^{M} J_N^2\left(\frac{j}{M+1}\right) = EJ_N^2(V_1) + O(N^{-14/15-14\delta/15}).$$

One easily verifies that $|EJ_N^2(V_1) - E\tilde{J}_N^2(\hat{V}_1)| = O(N^{-13/15-14\delta/15})$ and together with (3.3.49) and (3.3.16) this yields

$$|\sigma^2(T_{\omega N}) - \sigma^2(S_1)| = O(N^{-13/15-14\delta/15}), \tag{3.3.51}$$

uniformly in ω.

We now substitute the random versions of (3.3.48), (3.3.46) and (3.3.28) in (3.3.47). Using (3.3.50) and (3.3.51) we find after straightforward computations that uniformly for $|t| \le \log N$

$$\begin{aligned}
\psi_N^*(t) &= E[\chi(B)\{e^{-\frac{1}{2}t^2}(1 - \frac{it^3}{6\sigma_N^3}(\sum_{j=1}^N c_j^3 - \sum_{j\in I} c_{D_j}^3)E\tilde{J}_N^3(\hat{V}_1) \\
&+ \frac{t^4}{24\sigma_N^4}[\sum_{j=1}^N c_j^4[E\tilde{J}_N^4(\hat{V}_1) - 3\{E\tilde{J}_N^2(\hat{V}_1)\}^2] - \frac{3}{N}[E\tilde{J}_N^4(\hat{V}_1) - \{E\tilde{J}_N^2(\hat{V}_1)\}^2]] \\
&- \frac{t^6}{72\sigma_N^6}(\sum_{j=1}^N c_j^3)^2\{E\tilde{J}_N^3(\hat{V}_1)\}^2 + \frac{t^2}{2\sigma_N^2}[\sigma_N^2 - \sigma^2(T_N - Z_N|\Omega)] \\
&+ \frac{t^4}{8\sigma_N^4}[\sigma_N^2 - \sigma^2(T_N-Z_N|\Omega)]^2 - \frac{it^5}{12\sigma_N^5}\sum_{j=1}^N c_j^3 E\tilde{J}_N^3(\hat{V}_1)[\sigma_N^2 - \sigma^2(T_N-Z_N|\Omega)] \\
&+ \frac{it^3}{2\sigma_N^3}\frac{A_{1N}}{N}\sum_{j\in I} c_{D_j} + o(N^{-1}|t|P(t)e^{-\theta t^2}) \\
&+ O(N^{-1-\varepsilon}|t|P(t)\{1 + (\sum_{j\in I} c_{D_j})^2\}^{\frac{1}{2}})\}][1 + \frac{it}{\sigma_N}[E(Z_N|\Omega) + E(T_N-Z_N|\Omega)] \\
&- \frac{t^2}{2\sigma_N^2}[E(Z_N^2|\Omega) + 2E(Z_N|\Omega)E(T_N-Z_N|\Omega)] - \frac{it^3}{6\sigma_N^3}E(Z_N^3|\Omega) \\
&+ \frac{t^4}{24\sigma_N^4}E(Z_N^4|\Omega)] + O(N^{-22/15} + |t|P(t)N^{-1-7\delta/3}),
\end{aligned} \tag{3.3.52}$$

where $\varepsilon > 0$, $0 < \theta < \frac{1}{2}$ and P is a fixed polynomial.

A few more facts are needed to complete our calculation of $\psi_N^*(t)$. First we note that for $a = (m+1)(N+1)^{-1} = O(N^{-7/15})$, Assumption (3B) and (3.3.8) imply that

$$\int_0^a \{|J(t)|^k + |J(1-t)|^k\}dt = O(N^{-7/15+k/30-7k\delta/15}),$$

for k = 1,2,3,4 and hence

$$|EJ_N(\hat{V}_1)| = O(N^{-13/30-7\delta/15}),$$

(3.3.53)

$$E\tilde{J}_N^2(\hat{V}_1) = \frac{N+1}{M-1}\int_a^{1-a} J^2(t)dt + O(N^{-13/15-14\delta/15}),$$

$$E\tilde{J}_N^3(\hat{V}_1) = \frac{N+1}{M-1}\int_a^{1-a} J^3(t)dt - 3(\frac{N+1}{M-1})^2\left\{\int_a^{1-a} J^2(t)dt\right\}\left\{\int_a^{1-a} J(t)dt\right\}$$
$$+ O(N^{-13/10-7\delta/3}),$$

$$E\tilde{J}_N^4(\hat{V}_1) = \frac{N+1}{M-1}\int_a^{1-a} J^4(t)dt + O(N^{-13/30-7\delta/15}).$$

Furthermore, Lemma 3.2.3 yields

(3.3.54)
$$E(\sigma_N^2 - \sigma^2(T_N-Z_N|\Omega)) = E(E(Z_N^2|\Omega)) + 2E(E(Z_N|\Omega)E(T_N-Z_N|\Omega))$$
$$+ O(N^{-4/3-14\delta/15}).$$

Combining (3.3.53) and (3.3.54) with (3.3.52) it follows after some computations and repeated use of Assumptions (3A) and (3B) that, uniformly for $N^{-3/2} \le |t| \le \log N$,

(3.3.55)
$$\psi_N^*(t) = e^{-\frac{1}{2}t^2}\{1 - \frac{it^3}{6\sigma_N^3}\kappa_{3N} + \frac{t^4}{24\sigma_N^4}\kappa_{4N} - \frac{t^6}{72\sigma_N^6}\kappa_{3N}^2\}$$
$$+ o(N^{-1}|t|P(t)e^{-\theta t^2}) + O(N^{-1-\varepsilon}|t|P(t)),$$

where $\varepsilon > 0$, $0 < \theta < \frac{1}{2}$, P is a fixed polynomial and κ_{3N} and κ_{4N} are given by (3.1.10) and (3.1.11).

To conclude the proof of Theorem 3.1.1 we note that (3.2.1) implies

$$\sigma_N^2 = 1 + O(N^{-6/7-2\delta}).$$

Substituting this in (3.3.55) we obtain (3.3.5) with $\tilde{\psi}_N$ as in (3.3.2) and the proof of the theorem is complete.

3.4. TWO-SAMPLE LINEAR RANK STATISTICS

In this section we compare our results with the expansions for the two-sample linear rank statistics in BICKEL & VAN ZWET (1978). Let $1 \le n \le N$, $\lambda = nN^{-1}$ and assume that $\varepsilon \le \lambda \le 1-\varepsilon$ for some fixed $\varepsilon \in (0,\frac{1}{2})$ and all N. Define $c_j = (1-\lambda)/\{N\lambda(1-\lambda)\}^{\frac{1}{2}}$, $j = 1,2,\ldots,n$ and $c_j = -\lambda/\{N\lambda(1-\lambda)\}^{\frac{1}{2}}$, $j = n+1,\ldots$ $\ldots,N$. It is easy to check that in this case the c_j's satisfy Assumption (3A) and

$$\sum_{j=1}^{N} c_j^3 = \frac{1-2\lambda}{\{N\lambda(1-\lambda)\}^{\frac{1}{2}}}, \qquad \sum_{j=1}^{N} c_j^4 = \frac{1-3\lambda+3\lambda^2}{N\lambda(1-\lambda)} .$$

Taking a scores generating function J which satisfies Assumption (3B), we define the two-sample linear rank statistic as in (3.1.1). For the distribution F_N^* of the standardized version of this statistic Theorem 3.1.1 provides an Edgeworth expansion with remainder $o(N^{-1})$: if

$$\begin{aligned}\tilde{F}_N(x) = \Phi(x) - \phi(x)\Bigg\{&\frac{1-2\lambda}{6\{N\lambda(1-\lambda)\}^{\frac{1}{2}}}\left(\int_0^1 J^3(t)\,dt\right)(x^2-1)\\ &+ \frac{1}{24N\lambda(1-\lambda)}\left[(1-6\lambda+6\lambda^2)\int_0^1 J^4(t)\,dt - 3(1-2\lambda)^2\right](x^3-3x)\\ &+ \frac{(1-2\lambda)^2}{72N\lambda(1-\lambda)}\left(\int_0^1 J^3(t)\,dt\right)^2(x^5-10x^3+15x)\Bigg\},\end{aligned} \tag{3.4.1}$$

then

$$\sup_{x\in\mathbb{R}} |F_N^*(x) - \tilde{F}_N(x)| = o(N^{-1}), \quad \text{as } N \to \infty.$$

BICKEL & VAN ZWET (1978) consider the two-sample linear rank statistic T'_N for an arbitrary vector of scores $a = (a_1, a_2, \ldots, a_N)$, i.e.

$$(3.4.2) \qquad T'_N = \sum_{j=1}^{N} a_j V_j,$$

where

$$V_j = \begin{cases} 1 & \text{if } 1 \le D_j \le n \\ 0 & \text{otherwise,} \end{cases}$$

for $j = 1,2,\ldots,N$ and where $D_1, D_2, \ldots, D_N$ denote the antiranks. In their paper they establish asymptotic expansions for the distribution function of T'_N under the null-hypothesis as well as under contiguous alternatives. A related paper is that of ROBINSON (1978) which deals only with the null-hypothesis.

In order to compare the results in BICKEL & VAN ZWET (1978) with Theorem 3.1.1 in the present chapter we introduce the following assumption on the scores a_j.

ASSUMPTION (3C). *Let* $a_j = J(j/(N+1))$ *for* $j = 1,2,\ldots,N$. *This scores generating function* J *is twice continuously differentiable on* $(0,1)$ *and*

$$(3.4.3) \qquad \limsup_{t \to 0,1} t(1-t)\left|\frac{J''(t)}{J'(t)}\right| < 2;$$

there exist positive numbers $K > 0$ *and* $0 < \beta < 1/6$ *such that its first derivative* J' *satisfies*

$$(3.4.4) \qquad |J'(t)| \le K\{t(1-t)\}^{-7/6+\beta} \quad \text{for } t \in (0,1).$$

Furthermore

$$(3.4.5) \qquad \int_0^1 J(t)dt = 0, \qquad \int_0^1 J^2(t)dt = 1.$$

LEMMA 3.4.1. *If* $\varepsilon \le \lambda \le 1-\varepsilon$ *for some fixed* $\varepsilon \in (0,\frac{1}{2})$ *and Assumption* (3C) *is satisfied, then as* $N \to \infty$

$$(3.4.6) \qquad \sup_{x \in \mathbb{R}} \left| P\left(\frac{T'_N - ET'_N}{\sigma(T'_N)} \le x\right) - \tilde{F}_N(x)\right| = o(N^{-1}),$$

where $\tilde{F}_N$ *is defined in* (3.4.1).

PROOF. The present lemma is almost an immediate consequence of Corollary 2.1 of BICKEL & VAN ZWET (1978). Assumption (3C) guarantees that there exists a positive fraction of the scores which are at a distance of at least $N^{-3/2}\log N$ apart from each other. Furthermore, in view of Lemma 3.2.1 and Appendix 2 of ALBERS, BICKEL & VAN ZWET (1976), Assumption (3C) yields that

$$\sum_{j=1}^{N} a_j = O(N^{1/6-\beta}), \qquad \sum_{j=1}^{N} a_j^2 = N + O(N^{1/3-2\beta}),$$

$$\sum_{j=1}^{N} a_j^3 = N\int_0^1 J^3(t)dt + O(N^{\frac{1}{2}-3\beta}),$$

$$\sum_{j=1}^{N} a_j^4 = N\int_0^1 J^4(t)dt + O(N^{2/3-4\beta}).$$

Substituting this in the expansion $\tilde{R}(x,\bar{\lambda})$ (cf. (2.56) in BICKEL & VAN ZWET (1978)) and standardizing T_N' with the exact variance $\sigma^2(T_N')$ the result follows. □

For the two-sample case Lemma 3.4.1 is clearly a better result than Theorem 3.1.1, as was to be expected. Roughly speaking, Assumption (3B) in Theorem 3.1.1 requires a bit more smoothness than Assumption (3C) in Lemma 3.4.1; it also requires $\int|J|^{14+\varepsilon} < \infty$ instead of $\int|J|^{6+\varepsilon} < \infty$, where $\varepsilon > 0$. For practical purposes, however, Assumption (3B) is already quite satisfactory. It is gratifying to find that the expansions in the two results coincide. We note that some numerical examples are contained in BICKEL & VAN ZWET (1978).

3.5. FINITE SAMPLE COMPUTATIONS

In the preceding sections of this chapter we have derived Edgeworth expansions with remainder $o(N^{-1})$ for the distribution functions of simple linear rank statistics. In this section we investigate the performance of these expansions as approximations for the finite sample distributions of one special statistic, namely Spearman's rank correlation coefficient ρ_N. In particular we compare our expansions with the usual normal approximation. As noted in Chapter 1 we know that, under the null-hypothesis of independence, Spearman's rank correlation coefficient ρ_N is distributed as

$$(3.5.1) \qquad T_N^* = \frac{12}{N(N+1)(N-1)^{\frac{1}{2}}} \sum_{j=1}^{N} jR_j - \frac{3(N+1)}{(N-1)^{\frac{1}{2}}} .$$

From Theorem 3.1.1 it follows that, as $N \to \infty$

$$(3.5.2) \qquad F_N^*(x) = P(T_N^* \le x) = \tilde{F}_N(x) + o(N^{-1}),$$

where

$$(3.5.3) \qquad \tilde{F}_N(x) = \Phi(x) + \phi(x)\left\{\frac{9N^2-21}{100N(N^2-1)} + \frac{1}{10N}\right\}(x^3 - 3x).$$

We note that the third cumulant is zero because the scores generating function is symmetric.

In OLDS (1938) the exact distribution of T_N^* under the null-hypothesis was given for N = 2 through 7. The same results, together with the exact distribution for N = 8, were obtained by KENDALL, KENDALL & BABINGTON SMITH (1939). The 5% significance levels for N = 11 through 30 were derived in OLDS (1949). Further extensions of the exact distribution of Spearman's rank correlation coefficient under the hypothesis of independence were given in DAVID, KENDALL & STUART (1951). They established the exact distribution for N = 9 and 10 and showed that the formal Edgeworth expansions including the N^{-3} term would be quite satisfactory in practice for $N \ge 10$.

In Table 3.5.1 a comparison of the Edgeworth expansion $\tilde{F}_N$ and the normal approximation Φ with the exact distribution F_N^* is made for sample sizes N = 5, 10 and 20 and various values of the argument. We note that $\tilde{F}_N$ is truncated at 0 and 1. Furthermore, we note that for N = 20 we have employed a Monte-Carlo simulation based on 90,000 samples to estimate the exact distribution function F_N^*.

Inspection of Table 3.5.1 shows that the agreement between the estimated exact distribution function F_{20}^* and the expansion $\tilde{F}_{20}$ is almost perfect. It also shows that the expansion performs much better than the normal approximation. For N = 5 and 10 the agreement between F_N^* and $\tilde{F}_N$ is reasonable but not nearly as good as for N = 20. This is due to the fact that the probabilities of single values are still rather large for such small values of N; one can't expect to approximate a distribution function with large jumps by a continuous one in a satisfactory manner. To overcome this problem, we have employed a continuity correction. In Table 3.5.2 we summarize the results with this continuity correction for N = 5 and 10. Inspection of this table shows that the approximations $\tilde{F}_N$ are much improved; for sample size N = 10 the expansion $\tilde{F}_{10}$ performs quite well. It also shows that the expansions provide much better approximations than the usual normal approximation.

TABLE 3.5.1

Comparison of the exact distribution function with the Edgeworth expansion and normal approximation for N = 5, 10 and 20.

x	F_5^*	$\tilde{F}_5$	F_{10}^*	$\tilde{F}_{10}$	F_{20}^*	$\tilde{F}_{20}$	Φ
-3.0	0.0000	0.0000	0.0000	0.0000	0.0006	0.0006	0.0013
-2.8	0.0000	0.0000	0.0001	0.0005	0.0015	0.0015	0.0026
-2.6	0.0000	0.0000	0.0011	0.0022	0.0033	0.0034	0.0047
-2.4	0.0000	0.0027	0.0036	0.0054	0.0067	0.0068	0.0082
-2.2	0.0000	0.0086	0.0101	0.0112	0.0123	0.0125	0.0139
-2.0	0.0083	0.0188	0.0195	0.0207	0.0217	0.0217	0.0228
-1.8	0.0417	0.0347	0.0367	0.0353	0.0358	0.0356	0.0359
-1.6	0.0667	0.0577	0.0569	0.0563	0.0553	0.0555	0.0548
-1.4	0.1167	0.0888	0.0893	0.0849	0.0824	0.0828	0.0808
-1.2	0.1750	0.1285	0.1240	0.1219	0.1178	0.1185	0.1151
-1.0	0.2250	0.1766	0.1740	0.1678	0.1637	0.1632	0.1587
-0.8	0.2583	0.2321	0.2240	0.2222	0.2172	0.2170	0.2119
-0.6	0.3417	0.2938	0.2920	0.2842	0.2794	0.2793	0.2743
-0.4	0.3917	0.3601	0.3540	0.3525	0.3472	0.3485	0.3446
-0.2	0.4750	0.4293	0.4330	0.4251	0.4217	0.4229	0.4207
0.0	0.5250	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
0.2	0.6083	0.5707	0.5810	0.5749	0.5759	0.5771	0.5793
0.4	0.6583	0.6399	0.6460	0.6475	0.6506	0.6515	0.6554
0.6	0.7417	0.7062	0.7200	0.7158	0.7190	0.7207	0.7257
0.8	0.7750	0.7679	0.7760	0.7778	0.7821	0.7830	0.7881
1.0	0.8250	0.8234	0.8350	0.8322	0.8363	0.8368	0.8413
1.2	0.8833	0.8715	0.8760	0.8781	0.8821	0.8815	0.8849
1.4	0.9333	0.9112	0.9169	0.9151	0.9174	0.9172	0.9192
1.6	0.9583	0.9423	0.9431	0.9437	0.9450	0.9445	0.9452
1.8	0.9917	0.9653	0.9666	0.9647	0.9647	0.9644	0.9641
2.0	1.0000	0.9812	0.9805	0.9793	0.9791	0.9783	0.9772
2.2	1.0000	0.9914	0.9913	0.9888	0.9879	0.9875	0.9861
2.4	1.0000	0.9973	0.9964	0.9946	0.9935	0.9932	0.9918
2.6	1.0000	1.0000	0.9992	0.9978	0.9966	0.9966	0.9953
2.8	1.0000	1.0000	0.9999	0.9995	0.9984	0.9985	0.9974
3.0	1.0000	1.0000	1.0000	1.0000	0.9994	0.9994	0.9987

TABLE 3.5.2.

Comparison of the exact distribution function with the Edgeworth expansion and normal distribution after a continuity correction, for N = 5 and 10

x	F_5^*	$\tilde{F}_5$	Φ	F_{10}^*	$\tilde{F}_{10}$	Φ
-3.0	0.0000	0.0000	0.0019	0.0000	0.0000	0.0014
-2.8	0.0000	0.0000	0.0035	0.0001	0.0005	0.0026
-2.6	0.0000	0.0009	0.0062	0.0011	0.0024	0.0049
-2.4	0.0000	0.0052	0.0107	0.0036	0.0054	0.0082
-2.2	0.0000	0.0130	0.0179	0.0101	0.0119	0.0146
-2.0	0.0083	0.0259	0.0287	0.0195	0.0207	0.0228
-1.8	0.0417	0.0452	0.0446	0.0367	0.0369	0.0374
-1.6	0.0667	0.0722	0.0668	0.0569	0.0563	0.0548
-1.4	0.1167	0.1076	0.0968	0.0893	0.0879	0.0835
-1.2	0.1750	0.1515	0.1357	0.1240	0.1219	0.1151
-1.0	0.2250	0.2035	0.1841	0.1740	0.1724	0.1631
-0.8	0.2583	0.2623	0.2420	0.2240	0.2222	0.2119
-0.6	0.3417	0.3264	0.3085	0.2920	0.2902	0.2803
-0.4	0.3917	0.3944	0.3821	0.3540	0.3525	0.3446
-0.2	0.4750	0.4646	0.4602	0.4330	0.4319	0.4279
0.0	0.5250	0.5354	0.5398	0.5000	0.5000	0.5000
0.2	0.6083	0.6056	0.6179	0.5810	0.5816	0.5864
0.4	0.6583	0.6736	0.6915	0.6460	0.6475	0.6554
0.6	0.7417	0.7377	0.7580	0.7200	0.7217	0.7318
0.8	0.7750	0.7965	0.8159	0.7760	0.7778	0.7881
1.0	0.8250	0.8485	0.8643	0.8350	0.8367	0.8457
1.2	0.8833	0.8924	0.9032	0.8760	0.8781	0.8849
1.4	0.9333	0.9278	0.9332	0.9169	0.9181	0.9219
1.6	0.9583	0.9548	0.9554	0.9431	0.9437	0.9452
1.8	0.9917	0.9741	0.9713	0.9666	0.9663	0.9655
2.0	1.0000	0.9870	0.9821	0.9805	0.9793	0.9772
2.2	1.0000	0.9948	0.9893	0.9913	0.9895	0.9867
2.4	1.0000	0.9991	0.9938	0.9964	0.9946	0.9918
2.6	1.0000	1.0000	0.9965	0.9992	0.9980	0.9956
2.8	1.0000	1.0000	0.9981	0.9999	0.9995	0.9974
3.0	1.0000	1.0000	0.9995	1.0000	1.0000	0.9987

CHAPTER 4

ASYMPTOTIC EXPANSIONS UNDER CONTIGUOUS ALTERNATIVES

4.1. INTRODUCTION AND ASYMPTOTIC EXPANSION

In the preceding chapter we have derived Edgeworth expansions with remainder $o(N^{-1})$ for the distribution functions of simple linear rank statistics under the null-hypothesis. In the present chapter we turn to the case of contiguous alternatives. For simplicity we shall limit our study to contiguous location alternatives. Extension of the result to general contiguous alternatives is possible.

Hence, let $X_1, X_2, \ldots, X_N$ be independent random variables with probability density functions $f(x-\theta_{1N}), f(x-\theta_{2N}), \ldots, f(x-\theta_{NN})$, where $\max_{1\le j\le N} |\theta_{jN}| = O(N^{-\frac{1}{2}})$. We consider the simple linear rank statistic

$$(4.1.1) \qquad T_N = \sum_{j=1}^{N} c_{jN} J\left(\frac{R_{jN}}{N+1}\right),$$

where $c_{1N}, c_{2N}, \ldots, c_{NN}$ is a triangular array of regression constants, R_{jN} the rank of X_j among $(X_1, X_2, \ldots, X_N)$, $j = 1,2,\ldots,N$, $N = 1,2,\ldots$, and J is a scores generating function defined on $(0,1)$.

Throughout this chapter we make the following assumptions.

<u>ASSUMPTION (4A)</u>. *The regression constants* $c_{1N}, c_{2N}, \ldots, c_{NN}$ *satisfy*

$$\sum_{j=1}^{N} c_{jN} = 0, \qquad \sum_{j=1}^{N} c_{jN}^2 = 1 \quad \text{and} \quad \max_{1\le j\le N} |c_{jN}| = O(N^{-\frac{1}{2}}).$$

<u>ASSUMPTION (4B)</u>. *The location parameters* $\theta_{1N}, \theta_{2N}, \ldots, \theta_{NN}$ *satisfy*

$$\sum_{j=1}^{N} \theta_{jN} = 0, \qquad \sum_{j=1}^{N} \theta_{jN}^2 = 1 \quad \text{and} \quad \max_{1\le j\le N} |\theta_{jN}| = O(N^{-\frac{1}{2}}).$$

ASSUMPTION (4C). *The density function* f *is absolutely continuous on* $\mathbb{R}$ *with Radon-Nikodym derivative* f'. *This derivative* f' *is bounded and* $\int |f'| < \infty$.

ASSUMPTION (4D). *The scores generating function* J *is three times differentiable on* (0,1) *and its third derivative* J''' *satisfies a Lipschitz condition of the order* $0 < \alpha \leq 1$ *on* [0,1]. *Furthermore*

$$\int_0^1 J(t)dt = 0, \quad \int_0^1 J^2(t)dt = 1.$$

We note that Assumption (4A) is the same as Assumption (3A). Assumptions (4B) and (4C) appear quite satisfactory for practical purposes. It is easily seen that Assumption (4C) implies that f is bounded (cf. Lemma I.2.4.a of HÁJEK & ŠIDÁK (1967)). Since we have not required finite Fisher information, the sequence of alternatives may not be contiguous. This is not of great importance, however, because the rank tests satisfying Assumptions (4A) and (4D) will not be consistent against these alternatives.

Finally we note that Assumption (4D) is rather restrictive because it does not allow scores generating functions tending to infinity in the neighborhood of 0 and 1. In particular the case $J = \Phi^{-1}$, i.e. normal scores, are not covered.

Define for each $N \geq 2$

$$T_N^* = \frac{T_N - ET_N}{\sigma(T_N)}, \tag{4.1.2}$$

$$F_N^*(x) = P(T_N^* \leq x) \quad \text{for } -\infty < x < \infty \tag{4.1.3}$$

Furthermore, define, for each $N \geq 2$ and real x, the function $\tilde{F}_N$ by

$$\tilde{F}_N(x) = \Phi(x) - \phi(x)\left\{\frac{\tilde{\kappa}_{3N}}{6}(x^2-1) + \frac{\kappa_{4N}}{24}(x^3-3x) + \frac{\kappa_{3N}^2}{72}(x^5-10x^3+15x)\right\}, \tag{4.1.4}$$

where the quantities $\tilde{\kappa}_{3N}$, κ_{3N} and κ_{4N} are given by

$$\tilde{\kappa}_{3N} = \sum_{j=1}^{N} c_{jN}^3 \left\{ \int_0^1 J^3(t)dt + 3\theta_{jN} \int_0^1 (J^2(t)-1)J'(t)f(F^{-1}(t))dt \right.$$

$$(4.1.5) \qquad \left. - 3\int_0^1 J^3(t)dt \sum_{k=1}^{N} c_{kN}^2\theta_{kN} \int_0^1 J(t)J'(t)f(F^{-1}(t))dt \right\}$$

$$+ \frac{3}{N}\sum_{j=1}^{N} c_{jN}\theta_{jN} \int_0^1 (1-3J^2(t))J'(t)f(F^{-1}(t))dt,$$

$$(4.1.6) \qquad \kappa_{3N} = \sum_{j=1}^{N} c_{jN}^3 \int_0^1 J^3(t)dt,$$

$$(4.1.7) \qquad \kappa_{4N} = \sum_{j=1}^{N} c_{jN}^4 \left\{ \int_0^1 J^4(t)dt - 3 \right\} - \frac{3}{N}\left\{ \int_0^1 J^4(t)dt - 1 \right\}.$$

Note that κ_{3N} is merely the leading term in $\tilde{\kappa}_{3N}$, and that κ_{3N} and κ_{4N} are the same as in (3.1.10) and (3.1.11).

THEOREM 4.1.1. *If the Assumptions* (4A), (4B), (4C) *and* (4D) *are satisfied, then as* $N \to \infty$,

$$(4.1.8) \qquad \sup_{x\in\mathbb{R}} |F_N^*(x) - \tilde{F}_N(x)| = o(N^{-1}).$$

Section 4.2 contains two preliminary lemmas. In Section 4.3 we prove our theorem. Some comments are given in Section 4.4.

4.2. PRELIMINARIES

In this section we shall present two lemmas which we shall need in the next section. From this point on we shall suppress the index N whenever possible: in particular we shall write c_j and θ_j for c_{jN} and θ_{jN}.

LEMMA 4.2.1. *Let* $\{Y_N\}$ *and* $\{Z_N\}$ *be two sequences of random variables with finite second moments and let*

$$Y_N^* = \frac{Y_N - EY_N}{\sigma(Y_N)}, \qquad Z_N^* = \frac{Z_N - EZ_N}{\sigma(Z_N)}$$

denote the standardized versions. Suppose that there exists a number $\delta > 0$ *such that*

$$(4.2.1) \qquad \sigma^2(Y_N - Z_N) = O(N^{-\delta}) \quad \text{as } N \to \infty$$

and that

$$(4.2.2) \qquad \liminf_N \sigma^2(Y_N) > 0.$$

Then

$$(4.2.3) \qquad \sigma^2(Y_N^* - Z_N^*) = O(N^{-\delta}) \quad \text{as } N \to \infty.$$

PROOF. Direct computation shows

$$\sigma^2(Y_N^* - Z_N^*) = 2 - 2\operatorname{cov}(Y_N^*, Z_N^*) = 2 - \frac{2\operatorname{cov}(Y_N, Z_N)}{\sigma(Y_N)\sigma(Z_N)}$$

$$= 2 + \frac{\sigma^2(Y_N - Z_N) - \sigma^2(Y_N) - \sigma^2(Z_N)}{\sigma(Y_N)\sigma(Z_N)} = \frac{\sigma^2(Y_N - Z_N) - (\sigma(Y_N) - \sigma(Z_N))^2}{\sigma(Y_N)\sigma(Z_N)}.$$

Because of (4.2.1) and (4.2.2) we see that $\liminf_N \sigma^2(Z_N) > 0$ and

$$\sigma^2(Y_N^* - Z_N^*) \le \frac{\sigma^2(Y_N - Z_N)}{\sigma(Y_N)\sigma(Z_N)} = O(N^{-\delta}).$$

□

Let us now consider the behavior of the characteristic function of $T_n - ET_N$ for large values of the argument. The following lemma is again a special case of Theorem 2.1 of VAN ZWET (1980).

LEMMA 4.2.2. *Suppose that the assumptions of Theorem* 4.1.1 *are satisfied. Then there exist positive numbers* B, β *and* γ *such that*

$$(4.2.4) \qquad |Ee^{it(T_N - ET_N)}| \le BN^{-\beta \log N},$$

for $\log N \le |t| \le \gamma N^{3/2}$ *and* $N = 2,3,\ldots$.

PROOF. In view of Lemma 1.3.1 and Lemma 3.2.2 it remains to be shown that

$$\sum_{j=1}^{N} \int_{-\infty}^{\infty} \frac{\{f(x-\theta_j) - f(x)\}^2}{f(x)}\, dx = o(N),$$

as $N \to \infty$. It follows from Section 3 of VAN ZWET (1980) that this is implied by

$$(4.2.5) \qquad \int_{-\infty}^{\infty} \{\frac{1}{N} \sum_{j=1}^{N} (f(x-\theta_j)-f(x))^2\}^{\frac{1}{2}} dx = o(1).$$

Assumptions (4B) and (4C) imply that, for a fixed positive constant A,

$$\int_{-\infty}^{\infty} \{\frac{1}{N} \sum_{j=1}^{N} (f(x-\theta_j)-f(x))^2\}^{\frac{1}{2}} dx = \int_{-\infty}^{\infty} \{\frac{1}{N} \sum_{j=1}^{N} \left(\int_{x-\theta_j}^{x} f'(y)dy \right)^2\}^{\frac{1}{2}} dx$$

$$\leq \int_{-\infty}^{\infty} \int_{x-AN^{-\frac{1}{2}}}^{x+AN^{-\frac{1}{2}}} |f'(y)| dy dx = 2AN^{-\frac{1}{2}} \int_{-\infty}^{\infty} |f'(y)| dy = O(N^{-\frac{1}{2}}). \qquad \square$$

4.3. PROOF OF THEOREM 4.1.1.

The standard approach to prove results like Theorem 4.1.1 is to start with an application of Esseen's smoothing lemma (cf. Lemma 1.1.3) which implies that for all $\gamma > 0$

$$(4.3.1) \qquad \sup_{x\in \mathbb{R}} |F_N^*(x)-\tilde{F}_N(x)| \leq \frac{1}{\pi} \int_{-\gamma N^{3/2}}^{\gamma N^{3/2}} \frac{|\psi_N^*(t)-\tilde{\psi}_N(t)|}{|t|} dt + O(N^{-3/2}),$$

where ψ_N^* denotes the characteristic function of T_N^* (cf. (4.1.2)), i.e.

$$(4.3.2) \qquad \psi_N^*(t) = Ee^{itT_N^*}$$

and $\tilde{\psi}_N$ denotes the Fourier-Stieltjes transform of $\tilde{F}_N$ (cf. (4.1.4)), i.e.

$$(4.3.3) \qquad \tilde{\psi}_N(t) = \int_{-\infty}^{\infty} e^{itx} d\tilde{F}_N(x) = e^{-\frac{1}{2}t^2}\{1 - \frac{\tilde{\kappa}_{3N}}{6} it^3 + \frac{\kappa_{4N}}{24} t^4 - \frac{\kappa_{3N}^2}{72} t^6\}.$$

Define, for $j = 1,2,\ldots,N$,

$$(4.3.4) \qquad \rho_j = \frac{R_j}{N+1} = \frac{1}{N+1} \sum_{k=1}^{N} \chi(X_j - X_k),$$

where

$$\chi(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

(4.3.5) $$\zeta_j = E(\rho_j | X_j) = \frac{1}{N+1}\,[1 + \sum_{k\neq j} F(X_j - \theta_k)].$$

Let S_N be a four-term Taylor expansion of T_N, viz.

(4.3.6) $$S_N = \sum_{j=1}^{N} c_j\{J(\zeta_j) + (\rho_j - \zeta_j)J'(\zeta_j) + \tfrac{1}{2}(\rho_j - \zeta_j)^2 J''(\zeta_j) + \tfrac{1}{6}(\rho_j - \zeta_j)^3 J'''(\zeta_j)\},$$

and let us split the random variable $S_N - ES_N$ into a number of terms. Define,

$$I_1 = \sum_{j=1}^{N} c_j(J(\zeta_j) - EJ(\zeta_j)) = \sum_{j=1}^{N} c_j\tilde{J}(\zeta_j),$$

$$I_2 = \frac{1}{N+1} \sum\sum_{(j,k)\neq} c_j J'(\zeta_j)(\chi(X_j - X_k) - F(X_j - \theta_k)),$$

$$I_3 = \frac{1}{2(N+1)^2} \sum\sum\sum_{(j,k,\ell)\neq} c_j J''(\zeta_j)(\chi(X_j - X_k) - F(X_j - \theta_k))(\chi(X_j - X_\ell) - F(X_j - \theta_\ell)),$$

(4.3.7) $$I_4 = \frac{1}{2(N+1)^2} \sum\sum_{(j,k)\neq} c_j\{J''(\zeta_j)(\chi(X_j - X_k) - F(X_j - \theta_k))^2 - EJ''(\zeta_j)(F(X_j - \theta_k) - F^2(X_j - \theta_k))\},$$

$$I_5 = \frac{1}{6(N+1)^3} \sum\sum\sum\sum_{(j,k,\ell,n)} c_j J'''(\zeta_j)(\chi(X_j - X_k) - F(X_j - \theta_k))(\chi(X_j - X_\ell) - F(X_j - \theta_\ell))(\chi(X_j - X_n) - F(X_j - \theta_n)),$$

$$I_6 = \frac{1}{2(N+1)^3} \sum\sum\sum_{(j,k,\ell)\neq} c_j J'''(\zeta_j)(\chi(X_j - X_k) - F(X_j - \theta_k))^2(\chi(X_j - X_\ell) - F(X_j - \theta_\ell)),$$

$$I_7 = \frac{1}{6(N+1)^3} \sum_{(j,k)\neq} \sum c_j \{J'''(\zeta_j)(\chi(X_j - X_k) - F(X_j - \theta_k))^3$$

$$- EJ'''(\zeta_j) F(X_j - \theta_k)(1 - F(X_j - \theta_k))(1 - 2F(X_j - \theta_k))\},$$

so that $S_N - ES_N = \Sigma_{\nu=1}^{7} I_\nu$. Our first task is to derive an asymptotic expansion for

(4.3.8) $\tau_N^2 = \sigma^2(S_N)$.

Define

(4.3.9) $U_j = F(X_j - \theta_j)$,

$j = 1, 2, \ldots, N$, so that $U_1, U_2, \ldots, U_N$ are independent and uniformly distributed random variables on $(0,1)$.

LEMMA 4.3.1. *If the assumptions of Theorem* 4.1.1 *are satisfied, then*

(4.3.10) $\tau_N^2 = EI_1^2 + O(N^{-1}) = 1 + 2 \sum_{j=1}^{N} c_j^2 \theta_j \int_0^1 J(t) J'(t) f(F^{-1}(t)) dt + O(N^{-1})$.

PROOF. In view of Assumptions (4B) and (4C) Taylor expansion yields

(4.3.11) $F(x - \theta_k) = F(x) - \theta_k f(x) + O(N^{-1})$,

uniformly in x and k. Replacing x by $X_j - \theta_j$ and θ_k by $\theta_k - \theta_j$, for $j \neq k$, it follows from (4.3.5), (4.3.9) and Assumption (4B) that

$$\zeta_j = U_j + \theta_j f(F^{-1}(U_j)) + O(N^{-1}),$$

uniformly in j and with probability 1. Combining this, the fact that f is bounded (cf. Assumption (4C)) and Assumption (4D), it follows by another Taylor expansion that, uniformly in j and with probability 1,

(4.3.12) $J(\zeta_j) = J(U_j) + \theta_j f(F^{-1}(U_j)) J'(U_j) + O(N^{-1}) = J(U_j) + O(N^{-\frac{1}{2}})$,

(4.3.13) $J'(\zeta_j) = J'(U_j) + \theta_j f(F^{-1}(U_j)) J''(U_j) + O(N^{-1}) = J'(U_j) + O(N^{-\frac{1}{2}})$,

(4.3.14) $J''(\zeta_j) = J''(U_j) + O(N^{-\frac{1}{2}})$, (4.3.15) $$J'''(\zeta_j) = J'''(U_j) + O(N^{-\alpha/2}),$$

where $0 < \alpha \le 1$.

From (4.3.11) through (4.3.15) and Assumption (4D) it follows after some computations that

$$EI_1^2 = \sum_{j=1}^{N} c_j^2 E\tilde{J}^2(\zeta_j) = 1 + 2\sum_{j=1}^{N} c_j^2\theta_j \int_0^1 J(t)J'(t)f(F^{-1}(t))dt + O(N^{-1}),$$
$$\sum_{\nu=3}^{7} EI_\nu^2 = O(N^{-1}).$$

To show that $EI_2^2 = O(N^{-1})$ we use the technique employed in the proof of Lemma 4.3.3. Since $\tau_N^2 = \sigma^2(\Sigma_{\nu=1}^7 I_\nu)$, the result follows. □

The following lemma will enable us to show that $T_N - S_N$ is of negligible order for our purposes.

LEMMA 4.3.2. *If the assumptions of Theorem* 4.1.1 *are satisfied, then*

(4.3.16) $$\sigma^2(T_N - S_N) = O(N^{-2-\alpha}).$$

PROOF. Since the third derivative J''' satisfies a Lipschitz condition of order α we have, in view of Assumption (4A) and the Cauchy-Schwarz inequality,

$$\sigma^2(T_N-S_N) \le E(T_N-S_N)^2$$
$$\le E\{\sum_{j=1}^{N} |c_j||\rho_j-\zeta_j|^3 \sup_{0\le\eta_j\le 1} |J'''(\eta_j\rho_j + (1-\eta_j)\zeta_j) - J'''(\zeta_j)|\}^2$$
$$= \sum_{j=1}^{N} c_j^2\, O(E\sum_{j=1}^{N} |\rho_j-\zeta_j|^{6+2\alpha}) = O(\sum_{j=1}^{N} E|\rho_j-\zeta_j|^{6+2\alpha}).$$

Define for $X_j = x$, $Y_k = \chi(x-X_k) - F(x-\theta_k)$ for $1 \le (j,k)\ne \le N$. An application of the Marcinkievitz, Zygmund & Chung inequality (cf. CHUNG (1951)) shows that for all $r \ge 1$

$$E|\sum_{k\ne j} Y_k|^{2r} \le C(N-1)^{r-1} \sum_{k\ne j} E|Y_k|^{2r} \le C(N-1)^r,$$

where C depends only on r. It follows that for $j = 1,2,\ldots,N$

$$E|\rho_j - \zeta_j|^{6+2\alpha} = \frac{1}{(N+1)^{6+2\alpha}} E[E(|\sum_{k\neq j} \chi(X_j - X_k) - F(X_j - \theta_k)|^{6+2\alpha}|X_j)]$$
$$\leq CN^{-3-\alpha},$$

which proves the lemma. □

From Lemmas 4.3.1 and 4.3.2 it follows that

$$(4.3.17)\qquad \sigma^2(T_N) = 1 + 2\sum_{j=1}^{N} c_j^2\theta_j \int_0^1 J(t)J'(t)f(F^{-1}(t))dt + O(N^{-1}) = 1 + O(N^{-\frac{1}{2}}).$$

This fact, Lemma 4.2.2 and (4.3.3) ensure that (cf. Section 3 in VAN ZWET (1980))

$$(4.3.18)\qquad \int_{\log N \leq |t| \leq \gamma N^{3/2}} \frac{|\psi_N^*(t) - \tilde{\psi}_N(t)|}{|t|} dt = O(N^{-3/2}),$$

for some $\gamma > 0$, as $N \to \infty$. From (4.3.1) and (4.3.18) we see that, in order to prove Theorem 4.1.1, it suffices to show that

$$(4.3.19)\qquad \int_{|t| \leq \log N} \frac{|\psi_N^*(t) - \tilde{\psi}_N(t)|}{|t|} dt = o(N^{-1}).$$

To prove this we start with the computation of a number of moments.

<u>LEMMA 4.3.3</u>. *If the assumptions of Theorem* 4.1.1 *are satisfied, then*

$$(4.3.20)\qquad \begin{aligned} &E|I_2|^3 = O(N^{-9/8}), \qquad && E(\sum_{\nu=3}^{6} I_\nu^2) = O(N^{-3/2}), \\ &EI_7^2 = O(N^{-3}), && EI_1 I_6 = O(N^{-1-\alpha/2}), \end{aligned}$$

where $0 < \alpha \leq 1$ *(cf. Assumption* (4D)*).*

<u>PROOF</u>. For distinct j and k, let $h(X_j, X_k) = J'(\zeta_j)(\chi(X_j - X_k) - F(X_j - \theta_k))$. Define $h(x,x) = 0$ for all x. Direct computation of EI_2^4 shows that

$$EI_2^4 = \frac{1}{(N+1)^4}\Big[\sum_{j=1}^{N} c_j^4\Big\{\sum_{r=1}^{N}\sum_{s=1}^{N}\sum_{t=1}^{N}\sum_{u=1}^{N} Eh(X_j,X_r)h(X_j,X_s)h(X_j,X_t)$$

$$\cdot h(X_j,X_u)\Big\} + 4\sum\sum_{(j,k)\neq} c_j^3 c_k\Big\{\sum_{r=1}^{N}\sum_{s=1}^{N}\sum_{t=1}^{N}\sum_{u=1}^{N} Eh(X_j,X_r)h(X_j,X_s)$$

$$\cdot h(X_j,X_t)h(X_k,X_u)\Big\} + 3\sum\sum_{(j,k)\neq} c_j^2 c_k^2\Big\{\sum_{r=1}^{N}\sum_{s=1}^{N}\sum_{t=1}^{N}\sum_{u=1}^{N} Eh(X_j,X_r)$$

(4.3.21)
$$\cdot h(X_j,X_s)h(X_k,X_t)h(X_k,X_u)\Big\} + 6\sum\sum\sum_{(j,k,\ell)\neq} c_j^2 c_k c_\ell$$

$$\cdot\Big\{\sum_{r=1}^{N}\sum_{s=1}^{N}\sum_{t=1}^{N}\sum_{u=1}^{N} Eh(X_j,X_r)h(X_j,X_s)h(X_k,X_t)h(X_\ell,X_u)\Big\}$$

$$+ \sum\sum\sum\sum_{(j,k,\ell,n)\neq} c_j c_k c_\ell c_n\Big\{\sum_{r=1}^{N}\sum_{s=1}^{N}\sum_{t=1}^{N}\sum_{u=1}^{N} Eh(X_j,X_r)h(X_k,X_s)$$

$$\cdot h(X_\ell,X_t)h(X_n,X_u)\Big\}\Big].$$

To bound the right-hand side of (4.3.21) we note that an expectation in (4.3.21) equals zero if at least one of the indices (r,s,t,u) occurs only once. In view of Assumption (4D) all non-zero expectations are bounded. According to Assumption (4A) we have

$$\sum_{j=1}^{N} c_j^4 = O(N^{-1}), \qquad \sum\sum_{(j,k)\neq} |c_j^3 c_k| = O(1) \qquad \sum\sum_{(j,k)\neq} c_j^2 c_k^2 = O(1).$$

Hence the first three terms in the right-hand side of (4.3.21) are $O(N^{-2})$.

Similarly, the fourth term equals

$$\frac{6}{(N+1)^4}\sum\sum\sum_{(j,k,\ell)\neq} c_j^2 c_k c_\ell\Big\{\sum\sum_{(r,t)\neq} Eh^2(X_j,X_r)h(X_k,X_t)h(X_\ell,X_t)$$

$$+ 2Eh(X_j,X_r)h(X_j,X_t)h(X_k,X_t)h(X_\ell,X_r)\Big\} + O(N^{-2}),$$

Define $U_1,U_2,\ldots,U_N$ by (4.3.9) and $\tilde{h}(U_j,U_k) = J'(U_j)(\chi(U_j-U_k) - U_j)$ for $j\neq k$. Since f is bounded and $\max|\theta_j| = O(N^{-\frac{1}{2}})$ by Assumptions (4C) and (4B) respectively, we see that

(4.3.22) $\quad P(\chi(X_j-X_k) \neq \chi(U_j-U_k)) = O(N^{-\frac{1}{2}}),$

uniformly in j and k. Invoking (4.3.11) and (4.3.13) we find that

$$h(X_j, X_k) = \tilde{h}(U_j, U_k) + O(N^{-\frac{1}{2}}),$$

with probability $1 - O(N^{-\frac{1}{2}})$ and uniformly in j and k. Since h and $\tilde{h}$ are bounded, this implies, uniformly in j, k, r and t

$$Eh^2(X_j, X_r)h(X_k, X_t)h(X_\ell, X_t) = E\tilde{h}^2(U_1, U_4)\tilde{h}(U_2, U_5)\tilde{h}(U_3, U_5) + O(N^{-\frac{1}{2}}),$$

$$Eh(X_j, X_r)h(X_j, X_t)h(X_k, X_t)h(X_\ell, X_r)$$

$$= E\tilde{h}(U_1, U_4)\tilde{h}(U_1, U_5)\tilde{h}(U_2, U_5)\tilde{h}(U_3, U_4) + O(N^{-\frac{1}{2}}).$$

Note that the leading terms on the right of these expansions do not depend on the indices j, k, r and t. According to Assumption (4A) we have

$$\Big|\sum\sum\sum_{(j,k,\ell)\neq} c_j^2 c_k c_\ell\Big| = O(1), \qquad \sum\sum\sum_{(j,k,\ell)\neq} |c_j^2 c_k c_\ell| = O(N).$$

Combining these results we find that the fourth term is $O(N^{-3/2})$.

The non-zero expectations in the last term are of the form $Eh(X_j, X_r)h(X_k, X_r)h(X_\ell, X_r)h(X_n, X_r)$ or $Eh(X_j, X_r)h(X_k, X_r)Eh(X_\ell, X_s)h(X_n, X_s)$ where $(j,k,\ell,n,r,s)\neq$. By the same arguments as above we obtain

$$\frac{1}{(N+1)^4}\sum\sum\sum\sum\sum_{(j,k,\ell,n,r)\neq} c_j c_k c_\ell c_n Eh(X_j, X_r)h(X_k, X_r)h(X_\ell, X_r)h(X_n, X_r)$$

$$= O(N^{-3/2}).$$

From (4.3.11) and (4.3.13) we find after straightforward computations that, uniformly for distinct j, k and r,

$$Eh(X_j, X_r)h(X_k, X_r) = E\tilde{h}(U_1, U_3)\tilde{h}(U_2, U_3) + \theta_j A_1 + \theta_k A_2 + \theta_r A_3 + O(N^{-1}),$$

where the constants A_1, A_2 and A_3 are finite and do not depend on j, k and r. According to Assumptions (4A) and (4B) we have

$$|\sum\sum\sum\sum_{(j,k,\ell,n)\neq} c_j c_k c_\ell c_n| = O(1), \quad |\sum\sum\sum\sum_{(j,k,\ell,n)\neq} c_j\theta_j c_k c_\ell c_n| = O(N^{-\frac{1}{2}}),$$

$$|\sum\sum\sum\sum_{(j,k,\ell,n)\neq} c_j\theta_j c_k\theta_k c_\ell c_n| = O(1), \quad \sum\sum\sum\sum_{(j,k,\ell,n)\neq} |c_j c_k c_\ell c_n| = O(N^2).$$

Combining these results we find

$$\frac{1}{(N+1)^4} \sum\sum\sum\sum\sum\sum_{(j,k,\ell,n,r,s)\neq} c_j c_k c_\ell c_n Eh(X_j,X_r)h(X_k,X_r)Eh(X_\ell,X_s)h(X_n,X_s)$$

$$= O(N^{-3/2}).$$

Hence $EI_2^4 = O(N^{-3/2})$ and so $E|I_2|^3 = O(N^{-9/8})$.

The proof of the other assertions in (4.3.20) is easier and is therefore omitted. □

Define, for real t and $N \geq 2$ (cf. (4.3.6) and (4.3.8)),

$$(4.3.23) \qquad S_N^* = \frac{S_N - ES_N}{\tau_N},$$

$$(4.3.24) \qquad \mu_N^*(t) = Ee^{itS_N^*}$$

and

$$(4.3.25) \qquad \mu_{1N}(t) = Ee^{it\tau_N^{-1}I_1}\{1 + \frac{it}{\tau_N}(I_2 + I_3 + I_4 + I_5 + I_6) + \frac{(it)^2}{2\tau_N^2} I_2^2\}.$$

The next lemma shows that ψ_N^* can be approximated by μ_{1N}.

LEMMA 4.3.4. *If the assumptions of Theorem 4.1.1 are satisfied, then*

$$(4.3.26) \qquad |\psi_N^*(t) - \mu_{1N}(t)| = O(|t|\{N^{-1-\alpha/2} + (1+t^2)N^{-9/8}\}),$$

uniformly for $|t| \leq \log N$.

PROOF. It follows from Lemmas 4.2.1, 4.3.1 and 4.3.2 that

$$(4.3.27) \qquad |\psi_N^*(t) - \mu_N^*(t)| \leq |t|E|T_N^* - S_N^*| \leq |t|\sigma(T_N^* - S_N^*) = O(|t|N^{-1-\alpha/2}).$$

Furthermore, repeated use of Lemma XV 4.1 of FELLER (1971) yields

$$|\mu_N^*(t)-\mu_{1N}(t)| = O\Big(\frac{|t|}{\tau_N} E|I_7| + \frac{t^2}{\tau_N^2} E|I_2||I_3+I_4+I_5+I_6| \tag{4.3.28}$$
$$+ \frac{t^2}{\tau_N^2} E(I_3^2+I_4^2+I_5^2+I_6^2) + \frac{|t|^3}{\tau_N^3} E|I_2|^3\Big).$$

Applying Lemmas 4.3.1 and 4.3.3 we find that the right-hand side of (4.3.28) equals $O(|t|(1+t^2)N^{-9/8})$, uniformly for $|t| \leq \log N$. Combining this result with (4.3.27) we arrive at (4.3.26). □

It is clear that our next step is to evaluate the right-hand side of (4.3.25). The technique for doing this resembles that for $E \exp\{it\sigma_N^{-1}(S_{\omega n} - ES_{\omega N})\}$ in the preceding chapter (cf. (3.3.11)). We start with the leading term. According to (4.3.7) $I_1 = \Sigma_{j=1}^N c_j\tilde{J}(\zeta_j)$. We have $E\tilde{J}(\zeta_j) = 0$ and for all sufficiently large N, there exist positive constants $\gamma_1 \leq \gamma_2$ such that $\gamma_1 \leq E\tilde{J}^2(\zeta_j) \leq \gamma_2$ (cf. (4.3.12) and Assumption (4D)), $j = 1,2,\ldots,N$. Hence for all sufficiently large N we have $\gamma_1 \leq \sigma^2(I_1) \leq \gamma_2$. We also note that Assumptions (4A) and (4D) imply that $\max|c_j| = O(N^{-\frac{1}{2}})$ and that the random variable $\tilde{J}(\zeta_j)$ has finite absolute moments of any order, $j = 1,2,\ldots,N$. It follows from the classical theory of Edgeworth expansions for sums of independent and non-identically distributed random variables (see e.g. Lemma VI 4.11 in PETROV (1975)) that

$$|E \exp\{itI_1/\sigma(I_1)\} - e^{-\frac{1}{2}t^2}\{1 - \frac{it^3}{6\sigma^3(I_1)} \sum_{j=1}^N c_j^3\, E\tilde{J}^3(\zeta_j)$$
$$+ \frac{t^4}{24\sigma^4(I_1)} \sum_{j=1}^N c_j^4[E\tilde{J}^4(\zeta_j) - 3\{E\tilde{J}^2(\zeta_j)\}^2]$$
$$- \frac{t^6}{72\sigma^6(I_1)} \{\sum_{j=1}^N c_j^3\, E\tilde{J}^3(\zeta_j)\}^2\}|$$
$$= o(N^{-1}(t^4+|t|^9)e^{-\frac{1}{2}t^2}),$$

uniformly for $|t| \leq \log N$. Replacing t by $t_N = t\sigma(I_1)\tau_N^{-1}$, expanding $\exp\{-\frac{1}{2}t_N^2\}$ and using Lemma 4.3.1, we find that, uniformly for $|t| \leq \log N$,

$$(4.3.29)\quad \Big| Ee^{it\tau_N^{-1}I_1} - e^{-\frac{1}{2}t^2}\{1 - \frac{it^3}{6\tau_N^3}\sum_{j=1}^N c_j^3\, E\tilde{J}^3(\zeta_j) + \frac{t^4}{24\tau_N^4}\sum_{j=1}^N c_j^4[E\tilde{J}^4(\zeta_j) - 3\{E\tilde{J}^2(\zeta_j)\}^2] - \frac{t^6}{72\tau_N^6}\{\sum_{j=1}^N c_j^3\, E\tilde{J}^3(\zeta_j)\}^2\} + \frac{t^2}{2\tau_N^2}(\tau_N^2 - \sigma^2(I_1))\}\Big| = o(N^{-1}|t|P(t)e^{-\theta t^2}),$$

where $0 < \theta < \frac{1}{2}$ and P is a fixed polynomial.

We now turn to the remaining terms in the right-hand side of (4.3.25). Let

$$(4.3.30)\quad \nu_{jN}(t) = Ee^{it\tilde{J}(\zeta_j)},$$

denote the characteristic function of $\tilde{J}(\zeta_j)$, $j = 1,2,\ldots,N$, so that

$$(4.3.31)\quad Ee^{it\tau_N^{-1}I_1} = \prod_{j=1}^N \nu_{jN}\left(\frac{c_j t}{\tau_N}\right).$$

From Assumptions (4A) and (4D) it follows by Taylor expansion that for distinct integers $\ell_1,\ldots,\ell_n$, where $1 \le n \le 4$,

$$(4.3.32)\quad \prod_{j=1}^n \nu_{\ell_j N}\left(\frac{c_{\ell_j} t}{\tau_N}\right) = 1 - \frac{t^2}{2\tau_N^2}\sum_{j=1}^n c_{\ell_j}^2\, E\tilde{J}^2(\zeta_{\ell_j}) + O(|t|^3 N^{-3/2}),$$

uniformly for $|t| \le \log N$.

In the last lemma we summarize the results we need.

<u>LEMMA 4.3.5</u>. *If the assumptions of Theorem* 4.1.1 *are satisfied then, uniformly for* $|t| \le \log N$,

$$(4.3.33)\quad \Big| Ee^{it\tau_N^{-1}I_1}I_2 - Ee^{it\tau_N^{-1}I_1}\{\frac{it}{\tau_N}EI_1I_2 + \frac{(it)^2}{2\tau_N^2}EI_1^2I_2 - \frac{(it)^3}{4N\tau_N^3}\Big[\int_0^1 J^4(s)ds - 1\Big]\}\Big| = O(N^{-3/2}|t|P(t)e^{-\theta t^2}),$$

$$(4.3.34)\qquad |Ee^{it\tau_N^{-1}I_1}I_4 - Ee^{it\tau_N^{-1}I_1}\{\frac{it}{\tau_N}EI_1I_4\}| = O(N^{-3/2}|t|P(t)e^{-\theta t^2}),$$

$$(4.3.35)\qquad |Ee^{it\tau_N^{-1}I_1}(I_3+I_5+I_6)| = O(N^{-1-\alpha/2}|t|P(t)e^{-\theta t^2}),$$

$$(4.3.36)\qquad \begin{aligned}&|Ee^{it\tau_N^{-1}I_1}I_2^2 - Ee^{it\tau_N^{-1}I_1}\{EI_2^2+\frac{it}{\tau_N}EI_1I_2^2\\&+\frac{(it)^2}{4N\tau_N^2}\Big[\int_0^1 J^4(s)ds-1\Big]\}| = O(N^{-1-\alpha/2}|t|P(t)e^{-\theta t^2}),\end{aligned}$$

where $0 < \theta < \frac{1}{2}$, $0 < \alpha \le 1$ and P is a fixed polynomial,

PROOF. Because the statements (4.3.33) through (4.3.36) are all proved in essentially the same manner and because these statements resemble those in Lemma 3.3.4 and Lemma 3.3.5, we shall only prove the first statement, by way of an example. An application of Lemma XV 4.1 of FELLER (1971) shows that for distinct $1 \le j,k \le N$

$$\begin{aligned}&|\exp\{it\tau_N^{-1}(c_j\tilde{J}(\zeta_j)+c_k\tilde{J}(\zeta_k))\} - 1 - \frac{it}{\tau_N}(c_j\tilde{J}(\zeta_j)+c_k\tilde{J}(\zeta_k))\\&-\frac{(it)^2}{2\tau_N^2}(c_j\tilde{J}(\zeta_j)+c_k\tilde{J}(\zeta_k))^2-\frac{(it)^3}{6\tau_N^3}(c_j\tilde{J}(\zeta_j)+c_k\tilde{J}(\zeta_k))^3|\\&\le\frac{t^4}{\tau_N^4}(c_j\tilde{J}(\zeta_j)+c_k\tilde{J}(\zeta_k))^4.\end{aligned}$$

It follows from Assumptions (4A) and (4D) that

$$E\exp\{it\tau_N^{-1}(c_j\tilde{J}(\zeta_j)+c_k\tilde{J}(\zeta_k))\}J'(\zeta_j)(\chi(X_j-X_k)-F(X_j-\theta_k))$$

$$(4.3.37)\qquad \begin{aligned}&=EJ'(\zeta_j)(\chi(X_j-X_k)-F(X_j-\theta_k))[\frac{it}{\tau_N}(c_j\tilde{J}(\zeta_j)+c_k\tilde{J}(\zeta_k))\\&+\frac{(it)^2}{2\tau_N^2}(c_j^2\tilde{J}^2(\zeta_j)+2c_jc_k\tilde{J}(\zeta_j)\tilde{J}(\zeta_k)+c_k^2\tilde{J}^2(\zeta_k))+\frac{(it)^3}{6\tau_N^3}(c_j^3\tilde{J}^3(\zeta_j)\\&+3c_j^2c_k\tilde{J}^2(\zeta_j)\tilde{J}(\zeta_k)+3c_jc_k^2\tilde{J}(\zeta_j)\tilde{J}^2(\zeta_k)+c_k^3\tilde{J}^3(\zeta_k))]+O(N^{-2}t^4).\end{aligned}$$

Note that

$$EJ'(\zeta_j)(\chi(X_j-X_k)-F(X_j-\theta_k))[\sum_{\ell\neq j,k}\{\frac{it}{\tau_N}c_\ell\tilde{J}(\zeta_\ell)+\frac{(it)^2}{2\tau_N^2}c_\ell^2\tilde{J}^2(\zeta_\ell)$$

$$+\frac{(it)^2}{\tau_N^2}c_\ell\tilde{J}(\zeta_\ell)[c_j\tilde{J}(\zeta_j)+c_k\tilde{J}(\zeta_k)+\frac{1}{2}\sum_{n\neq j,k,\ell}c_n\tilde{J}(\zeta_n)]\}$$

$$-\frac{(it)^3}{6\tau_N^3}c_j^3\tilde{J}^3(\zeta_j)]=0$$

and hence that (4.3.37) equals

$$E\exp\{it\tau_N^{-1}(c_j\tilde{J}(\zeta_j)+c_k\tilde{J}(\zeta_k))\}J'(\zeta_j)(\chi(X_j-X_k)-F(X_j-\theta_k))$$

(4.3.38)
$$=EJ'(\zeta_j)(\chi(X_j-X_k)-F(X_j-\theta_k))[\frac{it}{\tau_N}I_1+\frac{(it)^2}{2\tau_N^2}I_1^2$$
$$+\frac{(it)^3}{6\tau_N^3}(3c_j^2c_k\tilde{J}^2(\zeta_j)\tilde{J}(\zeta_k)+3c_jc_k^2\tilde{J}(\zeta_j)\tilde{J}^2(\zeta_k)+c_k^3\tilde{J}^3(\zeta_k))]$$
$$+O(N^{-2}t^4).$$

From (4.3.32) it follows that for distinct integers $1\leq j,k\leq N$ and uniformly for $|t|\leq\log N$

(4.3.39)
$$\prod_{\ell\neq j,k}\nu_{\ell N}\left(\frac{c_\ell t}{\tau_N}\right)=Ee^{it\tau_N^{-1}I_1}\{1+\frac{t^2}{2\tau_N^2}(c_j^2E\tilde{J}^2(\zeta_j)+c_k^2E\tilde{J}^2(\zeta_k))$$
$$+O(|t|^3N^{-3/2}).$$

Combining (4.3.38), (4.3.39), (4.3.22), (4.3.29), (4.3.11), (4.3.12) and (4.3.13) with Assumptions (4A) through (4D) we find after some algebra

$$Ee^{it\tau_N^{-1}I_1}I_2=\sum_{(j,k)\neq}\sum\frac{c_j}{N+1}\prod_{\ell\neq j,k}\nu_{\ell N}\left(\frac{c_\ell t}{\tau_N}\right)$$

(4.3.40)
$$\cdot E\exp\{it\tau_N^{-1}(c_j\tilde{J}(\zeta_j)+c_k\tilde{J}(\zeta_k))\}J'(\zeta_j)(\chi(X_j-X_k)-F(X_j-\theta_k))=$$

$$(4.3.40)\quad = Ee^{it\tau_N^{-1}I_1}\Big[\frac{it}{\tau_N} EI_1I_2 + \frac{(it)^2}{2\tau_N^2} EI_1^2I_2 + \frac{(it)^3}{2N\tau_N^3} EJ(U_1)J'(U_1)J^2(U_2)(\chi(U_1-U_2) - U_1)\Big] + O(N^{-3/2}|t|P(t)e^{-\theta t^2}).$$

Finally we obtain by partial integration

$$(4.3.41)\quad EJ(U_1)J'(U_1)J^2(U_2)(\chi(U_1-U_2) - U_1) = -\frac{1}{2}\left[\int_0^1 J^4(s)ds - 1\right].$$

From (4.3.40) and (4.3.41) we conclude (4.3.33). □

From Lemma 4.3.5 it follows that uniformly for $|t| \le \log N$ (cf. (4.3.25))

$$\mu_{1N}(t) = Ee^{it\tau_N^{-1}I_1}\{1 + \frac{(it)^2}{2\tau_N^2}[2EI_1I_2 + 2EI_1I_4 + EI_2^2]$$

$$(4.3.42)\quad + \frac{(it)^3}{2\tau_N^3}[EI_1^2I_2 + EI_1I_2^2] - \frac{(it)^4}{8N\tau_N^4}\left[\int_0^1 J^4(s)ds - 1\right]\}$$

$$+ O(N^{-1-\alpha/2}|t|P(t)e^{-\theta t^2}),$$

where $0 < \theta < \frac{1}{2}$, $0 < \alpha \le 1$ and P is a fixed polynomial. Using Lemma 4.3.3 as well as the fact that $EI_1I_3 = EI_1I_5 = 0$, we obtain

$$(4.3.43)\quad 2EI_1I_2 + 2EI_1I_4 + EI_2^2 = \tau_N^2 - \sigma^2(I_1) + O(N^{-9/8} + N^{-1-\alpha/2}),$$

Let, for distinct j and k, $h(X_j,X_k) = J'(\zeta_j)(\chi(X_j-X_k) - F(X_j-\theta_k))$. We find by repeated use of Assumptions (4A) through (4D) that

$$EI_1^2I_2 = \sum_{(j,k)\neq}\sum \frac{c_j^2c_k}{N+1}\{E\tilde{J}^2(\zeta_j)h(X_k,X_j) + 2E\tilde{J}(\zeta_j)\tilde{J}(\zeta_k)h(X_j,X_k)\}$$

$$(4.3.44)\quad = -\sum_{j=1}^N \frac{c_j\theta_j}{N}\left\{\int_0^1 [tJ''(t) + 2J^2(t)J'(t)]f(F^{-1}(t))dt - \int_0^1 J^2(t)\int_t^1 J''(s)f(F^{-1}(s))dsdt\right\} + O(N^{-3/2}),$$

$$EI_1I_2^2 = \sum\sum_{(j,k)\neq} \frac{c_j^3}{(N+1)^2} E\tilde{J}(\zeta_j)h^2(X_j,X_k) + \sum\sum_{(j,k)\neq} \frac{c_j^2 c_k}{(N+1)^2}$$

$$\cdot \{E\tilde{J}(\zeta_k)h^2(X_j,X_k) + 2E\tilde{J}(\zeta_j)h(X_j,X_k)h(X_k,X_j)$$

$$+ 2\sum_{\ell\neq j,k} E\tilde{J}(\zeta_j)h(X_j,X_\ell)h(X_k,X_\ell)\}$$

(4.3.45)

$$+ \sum\sum\sum_{(j,k,\ell)\neq} \frac{c_j c_k c_\ell}{(N+1)^2} \{2E\tilde{J}(\zeta_\ell)h(X_j,X_k)h(X_k,X_\ell)$$

$$+ E\tilde{J}(\zeta_\ell)h(X_j,X_\ell)h(X_k,X_\ell)\}$$

$$= \sum_{j=1}^{N} \frac{c_j\theta_j}{N}\left\{\int_0^1 [tJ''(t) + J'(t) - J^2(t)J'(t)]f(F^{-1}(t))dt\right.$$

$$\left. - \int_0^1 J^2(t)\int_t^1 J''(s)f(F^{-1}(s))dsdt\right\} + O(N^{-3/2}).$$

Substituting (4.3.43) through (4.3.45) in (4.3.42) we have that uniformly for $|t| \le \log N$,

$$\mu_{1N}(t) = Ee^{it\tau_N^{-1}I_1}\{1 - \frac{t^2}{2\tau_N^2}[\tau_N^2 - \sigma^2(I_1)]$$

(4.3.46)

$$- \frac{it^3}{2\tau_N^3}\sum_{j=1}^{N}\frac{c_j\theta_j}{N}\int_0^1 (1 - 3J^2(s))J'(s)f(F^{-1}(s))ds$$

$$- \frac{t^4}{8N\tau_N^4}\left[\int_0^1 J^4(s)ds - 1\right] + O(\{N^{-9/8} + N^{-1-\alpha/2}\}|t|P(t)e^{-\theta t^2}),$$

where $0 < \theta < \frac{1}{2}$, $0 < \alpha \le 1$ and P is a fixed polynomial.

A few more facts are needed to complete our proof of Theorem 4.1.1. First we note that Assumptions (4C) and (4D) imply that (cf. 4.3.12)

$$EJ(\zeta_j) = \theta_j \int_0^1 J'(t)f(F^{-1}(t))dt + O(N^{-1}),$$

(4.3.47) $$EJ^2(\zeta_j) = 1 + O(N^{-\frac{1}{2}}),$$

$$(4.3.47)\qquad EJ^3(\zeta_j) = \int_0^1 J^3(t)dt + 3\theta_j \int_0^1 J^2(t)J'(t)f(F^{-1}(t))dt + O(N^{-1}),$$

$$EJ^4(\zeta_j) = \int_0^1 J^4(t)dt + O(N^{-\frac{1}{2}}),$$

uniformly in j. Now we substitute (4.3.29), (4.3.10) and (4.3.47) in (4.3.46) and find after some computations that, uniformly for $|t| \le \log N$.

$$\mu_{1N}(t) = e^{-\frac{1}{2}t^2}\{1 - \frac{it^3}{6}\tilde{\kappa}_{3N} + \frac{t^4}{24}\kappa_{4N} - \frac{t^6}{72}\kappa_{3N}^2\} + o(N^{-1}|t|P(t)e^{-\theta t^2}),$$

where $0 < \theta < \frac{1}{2}$, P is a fixed polynomial and the quantities $\tilde{\kappa}_{3N}$, κ_{3N} and κ_{4N} are given by (4.1.5), (4.1.6) and (4.1.7). In view of Lemma 4.3.4 and our starting point (4.3.19) this result completes the proof of Theorem 4.1.1. □

4.4. COMMENTS

In this section we provide a discussion of Theorem 4.1.1. First of all we note that the standardization of T_N with ET_N and $\sigma(T_N)$ is not realistic because under the alternative these quantities are unknown. However, it is possible to prove a modification of Theorem 4.1.1 since we can replace these quantities by their asymptotic expressions. Let λ_N denote the asymptotic expectation of T_N and ω_N the asymptotic standard deviation.

Define, for each $N \ge 2$,

$$(4.4.1)\qquad G_N(x) = P\left(\frac{T_N-\lambda_N}{\omega_N} \le x\right) \quad \text{for } -\infty < x < \infty.$$

The problem is now to establish an asymptotic expansion with remainder $o(N^{-1})$ for the distribution function G_N. We know that for each $N \ge 2$ and real x (cf. (4.1.3))

$$G_N(x) = F_N^*\left(\frac{x\omega_N+\lambda_N-ET_N}{\sigma(T_N)}\right).$$

Using this identity and applying Theorem 4.1.1 we find that (cf. (4.1.4))

$$(4.4.2)\qquad \sup_x |G_N(x) - \tilde{F}_N\left(\frac{x\omega_N+\lambda_N-ET_N}{\sigma(T_N)}\right)| = o(N^{-1}).$$

From (4.4.2) it follows that we need expansions for $\omega_N \sigma^{-1}(T_N)$ and $(\lambda_N - ET_N) \cdot \sigma^{-1}(T_N)$ with remainder terms of order $o(N^{-1})$. It will be clear that to establish such expansions we have to impose a stronger assumption on the density f than Assumption (4C). In a forthcoming paper (DOES (1982)), we shall establish these expansions.

We conclude this section with a discussion of Assumption (4C) concerning the density function f. It is remarkable that the strength of this assumption is comparable to what is needed for proving asymptotically normality under these alternatives (cf. HÁJEK & ŠIDÁK (1967) Theorem VI 2.4). Moreover, if we compare this assumption with assumptions in related papers on asymptotic expansions both in nonparametric statistics (cf. ALBERS, BICKEL & VAN ZWET (1976) Section 3 and BICKEL & VAN ZWET (1978) Section 4) and in parametric statistics (cf. CHIBISOV (1973a, 1973b) and PFANZAGL (1973, 1974, 1979)) we find that in all papers much stronger assumptions on the density functions are made. Apparently this is due to the rather restrictive assumption concerning the scores generating function and to the exact standardization employed in the present chapter.

REFERENCES

[1] ALBERS, W. (1974), *Asymptotic Expansions and the Deficiency Concept in Statistics*, Mathematical Centre Tracts 58, Amsterdam.

[2] ALBERS, W., P.J. BICKEL & W.R. VAN ZWET (1976), *Asymptotic expansions for the power of distribution free tests in the one-sample problem*, Ann. Statist. 4, 108-156.

[3] BAHR, B. von (1972), *On sampling from a finite set of independent random variables*, Z. Wahrsch. Verw. Gebiete 24, 279-286.

[4] BAHR, B. von (1976), *Remainder term estimate in a combinatorial limit theorem*, Z. Wahrsch. Verw. Gebiete 35, 131-139.

[5] BERGSTRÖM, H. & M.L. PURI (1977), *Convergence and remainder terms in linear rank statistics*, Ann. Statist. 5, 671-680.

[6] BHATTACHARYA, R.N. & R.R. RAO (1976), *Normal Approximation and Asymptotic Expansions*, Wiley, New York.

[7] BICKEL, P.J. (1974), *Edgeworth expansions in nonparametric statistics*, Ann. Statist. 2, 1-20.

[8] BICKEL, P.J. & W.R. VAN ZWET (1978), *Asymptotic expansions for the power of distributionfree tests in the two-sample problem*, Ann. Statist. 6 937-1004.

[9] BIKELIS, A. (1969), *On the estimation of the remainder term in the central limit theorem for samples from finite populations*, (in Russian), Studia Sci. Math. Hungar. 4, 345-354.

[10] CALLAERT, H., P. JANSSEN & N. VERAVERBEKE (1980), *An Edgeworth expansion for U-statistics*, Ann. Statist. 8, 299-312.

[11] CHERNOFF, H. & I.R. SAVAGE (1958), *Asymptotic normality and efficiency of certain nonparametric test statistics*, Ann. Math. Statist. 29, 972-994.

[12] CHIBISOV, D.M. (1973a), *Asymptotic expansions for Neyman's* C(α) *tests*, Proc. Second Japan-USSR Symp. Probab. Theory, Lecture Notes in Math. 330, Springer-Verlag, Berlin, 16-45.

[13] CHIBISOV, D.M. (1973b), *An asymptotic expansion for a class of estimators containing maximum likelihood estimators*, Theory Probab. Appl. 18, 295-303.

[14] CHUNG, K.L. (1951), *The strong law of large numbers*, Proc. Second Berkeley Symp. Math. Statist. Probab., J. Neyman (ed.), University of Calfornia Press, Berkeley, 341-352.

[15] DAVID, S.T., M.G. KENDALL & A. STUART (1951), *Some questions of distribution in the theory of rank correlation*, Biometrika 38, 131-140.

[16] DOES, R.J.M.M. (1981a), *Berry-Esseen theorems for simple linear rank statistics under the null-hypothesis*, Mathematical Centre Report SW 71, Amsterdam, to appear in the Annals of Probability.

[17] DOES, R.J.M.M. (1981b), *An Edgeworth expansion for simple linear rank statistics under the null-hypothesis*, Mathematical Centre Report SW 78, Amsterdam.

[18] DOES, R.J.M.M. (1982), *Asymptotic expansions for simple linear rank statistics under contiguous alternatives*, Manuscript in preparation.

[19] DUPAČ, V. & J. HÁJEK (1969), *Asymptotic normality of simle linear rank statistics under alternatives II*, Ann. Math. Statist. 40, 1992-2017.

[20] DUPAČ, V. (1970), *A contribution to the asymptotic normality of simple linear rank statistics*, Nonparametric Techniques in Statistical Inference, M.L. Puri (ed.), Cambridge University Press, London, 75-88.

[21] DWASS, M. (1955), *On the asymptotic normality of some statistics used in non-parametric tests*, Ann. Math. Statist. 26, 334-339.

[22] DWASS, M. (1956), *The large-sample power of rank order tests in the two-sample problem*, Ann. Math. Statist. 27, 352-374.

[23] ERDÖS, P. & A. RÉNYI (1959), *On the central limit theorem for samples from a finite population*, Publ. Math. Inst. Hungar. Acad. Sci., Vol. 4, 49-61.

[24] ESSEEN, C.G. (1945), *Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law*, Acta Math. 77, 1-125.

[25] FELLER, W. (1971), *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd edition, Wiley, New York.

[26] FIX, E. & J.L. HODGES (1955), *Significance probabilities of the Wilcoxon test*, Ann. Math. Statist. 26, 301-312.

[27] FRASER, D.A.S. (1956), *A vector form of the Wald-Wolfowitz-Hoeffding theorem*, Ann. Math. Statist. 27, 540-543.

[28] GOVINDARAJULU, Z., L. LECAM & M. RAGHAVACHARI (1967), *Generalizations of theorems of Chernoff and Savage on the asymptotic normality of test statistics*, Proc. Fifth Berkeley Symp. Math. Statist. Probab., Vol. I, L. LeCam and J. Neyman (eds), University of California Press, Berkeley, 609-638.

[29] HÁJEK, J. (1961), *Some extensions of the Wald-Wolfowitz-Noether theorem*, Ann. Math. Statist. 32, 506-523.

[30] HÁJEK, J. (1962), *Asymptotically most powerful rank-order tests*, Ann. Math. Statist. 33, 1124-1147.

[31] HÁJEK, J. & Z. ŠIDÁK (1967), *Theory of Rank Tests*, Academic Press, New York.

[32] HÁJEK, J. (1968), *Asymptotic normality of simple linear rank statistics under alternatives*, Ann. Math. Statist. 39, 325-346.

[33] HELMERS, R. (1980), *Edgeworth expansions for linear combinations of order statistics with smooth weight functions*, Ann. Statist. 8, 1361-1374.

[34] HELMERS, R. & W.R. VAN ZWET (1981), *The Berry-Esseen bound for U-statistics*, Mathematical Centre Report SW 75, Amsterdam, to appear in Statistical Decision Theory and Related Topics III, S.S. Gupta and J.O. Berger (eds), Academic Press, New York

[35] HO, S.T. & L.H.Y. CHEN (1978), *An* L_p *bound for the remainder in a combinatorial central limit theorem*, Ann. Probab. 6, 231-249.

[36] HODGES, J.L. & E.L. LEHMANN (1970), *Deficiency*, Ann. Math. Statist. 41, 783-801.

[37] HOEFFDING, W. (1951), *A combinatorial central limit theorem*, Ann. Math. Statist. 22, 558-566.

[38] HOEFFDING, W. (1973), *On the centering of a simple linear rank statistic*, Ann. Statist. 1, 54-66.

[39] HOTELLING, H. & M.R. PABST (1936), *Rank correlation and tests of significance involving no assumption of normality*, Ann. Math. Statist. 7, 29-43.

[40] HUŠKOVÁ, M. (1977a), *The rate of convergence of simple linear rank statistics under hypothesis and alternatives*, Ann. Statist. 5, 658-670.

[41] HUŠKOVÁ, M. (1977b), *Rates of convergence of quadratic rank statistics*, J. Multivariate Anal. 7, 63-73.

[42] HUŠKOVÁ, M. (1979a), *The rate of convergence of simple linear rank statistics under alternatives*, Contributions to Statistics (J. Hájek Memorial Volume), J. Jurečková (ed.), Academia, Prague, 99-108.

[43] HUŠKOVÁ, M. (1979b), *The Berry-Esseen theorem for rank statistics*, Comment. Math. Univ. Carol. 20, 399-415.

[44] JOHNSON, N.L. & S. KOTZ (1969), *Distributions in Statistics: Discrete Distributions*, Wiley, New York.

[45] JUREČKOVÁ, J. & M.L. PURI (1975), *Order of normal approximation for rank test statistics distribution*, Ann. Probab. 3, 526-533.

[46] KENDALL, M.G., S.F.H. KENDALL & B. BABINGTON SMITH (1939), *The distribution of Spearman's coefficient of rank correlation in a universe in which all rankings occur an equal number of times*, Biometrika 30, 251-273.

[47] LECAM, L. (1960), *Locally asymptotically normal families of distributions*, Univ. Calif. Publ. Statist. 3, 37-98.

[48] LEHMANN, E.L. (1975), *Nonparametrics: statistical methods based on ranks*, Holden-Day, San Francisco.

[49] MADOW, W.G. (1948), *On the limiting distributions of estimates based on samples from finite universes*, Ann. Math. Statist. 19, 535-545.

[50] MASON, D.M. (1981), *On the use of a statistic based on sequential ranks to prove limit theorems for simple linear rank statistics*, Ann. Statist. 9, 424-436.

[51] MOTOO, M. (1957), *On the Hoeffding's combinatorial central limit theorem*, Ann. Inst. Statist. Math. 8, 145-154.

[52] NOETHER, G.E. (1949), *On a theorem by Wald and Wolfowitz*, Ann. Math. Statist. 20, 455-458.

[53] OLDS, E.G. (1938), *Distributions of sums of squares of rank differences for small numbers of individuals*, Ann. Math. Statist. 9, 133-148.

[54] OLDS, E.G. (1949), *The 5% significance levels for sums of squares of rank differences and a correction*, Ann. Math. Statist. 20, 117-118.

[55] PETROV, V.V. (1975), *Sums of Independent Random Variables*, Springer-Verlag, Berlin.

[56] PFANZAGL, J. (1973), *Asymptotic expansions related to minimum contrast estimators*, Ann. Statist. 1, 993-1026.

[57] PFANZAGL, J. (1974), *Asymptotically optimum estimation and test procedures*, Proc. Prague Symp. Asympt. Statist., Vol. I, J. Hájek (ed.), Charles University, Prague, 201-274.

[58] PFANZAGL, J. (1979), *First order efficiency implies second order efficiency*, Contributions to Statistics (J. Hájek Memorial Volume), J. Jurečková (ed.), Academia, Prague, 167-196.

[59] PURI, M.L. & P.K. SEN (1971), *Nonparametric Methods in Multivariate Analysis*, Wiley, New York.

[60] PYKE, R. & G.R. SHORACK (1968), *Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorems*, Ann. Math. Statist. 39, 755-771.

[61] ROBINSON, J. (1978), *An asymptotic expansion for samples from a finite population*, Ann. Statist. 6, 1005-1011.

[62] ROBINSON, J. (1980), *An asymptotic expansion for permutation tests with several samples*, Ann. Statist. 8, 851-864.

[63] SERFLING, R.J. (1980), *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

[64] STOKER, D.J. (1954), *An upper bound for the deviation between the distribution of Wilcoxon's test statistic for the two-sample problem and its limiting normal distribution for finite samples I & II*, Indag. Math. 16, 599-614.

[65] VAN ZWET, W.R. (1977), *Asymptotic expansions for the distribution functions of linear combinations of order statistics*, Statistical Decision Theory and Related Topics II, S.S. Gupta and D.S. Moore (eds), Academic Press, New York, 421-437.

[66] VAN ZWET, W.R. (1980), *On the Edgeworth expansion for the simple linear rank statistic*, Mathematical Centre Report SW 70, Amsterdam, to appear in the Proceedings of the Colloquium on Nonparametric Statistical Inference, I. Vincze (ed.), Budapest.

[67] WALD, A. & J. WOLFOWITZ (1944), *Statistical tests based on permutations of the observations*, Ann. Math. Statist. 15, 358-372.

AUTHOR INDEX

SUBJECT INDEX