The Annals of Statistics
1974, Vol. 2, No. 1, 1-20

SPECIAL INVITED PAPER

EDGEWORTH EXPANSIONS IN NONPARAMETRIC STATISTICS[1]

BY P. J. BICKEL

University of California, Berkeley

This is a survey of recent work on Edgeworth expansions for (M) estimates, rank tests and some other statistics arising in nonparametric models. A Berry–Esséen theorem for U-statistics which seems to be new is also proved.

1. Introduction. During the past 25 years various procedures which are not sensitive to certain departures from normality have been evolved and investigated. The study of such methods is loosely referred to as nonparametric statistics. One broad category of such procedures is that of the distribution free tests such as the permutation t test, the rank tests of Wilcoxon, Kruskal–Wallis, Spearman and Kendall, and the omnibus tests such as the two sample Smirnov test. All of these are discussed in the monograph of Hájek and Šidák [26]. Another major category is that of the various robust estimates such as those discussed in the recent Princeton study [2].

Most of the theoretical work done on these procedures has been devoted to obtaining large sample properties by establishing first order limit theorems for the statistics on which these procedures are based. In this paper I intend to discuss what is known about higher order approximations to the distribution of these statistics. In the main I shall limit myself to discussion of results obtained since the general review paper by D. Wallace which appeared in this journal in 1958, [57].

Suppose that we are given a sequence of statistics $\{T_N\}$, $N \geqq 1$, where N usually denotes sample size. In accordance with [57] we shall say that the distribution function F_N of T_N possesses an asymptotic expansion valid to $(r+1)$ terms if there exist functions $A_0, \cdots, A_r$ such that

$$(1.1) \qquad \left|F_N(x) - A_0(x) - \sum_{j=1}^r \frac{A_j(x)}{N^{j/2}}\right| = o(N^{-r/2}) .$$

If,

$$(1.2) \qquad \sup_x \left|F_N(x) - A_0(x) - \sum_{j=1}^r \frac{A_j(x)}{N^{j/2}}\right| = o(N^{-r/2})$$

Received January 1973; revised May 1973.

[1] This research was supported by the Office of Naval Research, Contract N00014-69-A-0200-1038.

AMS 1970 *subject classifications.* Primary 62G05, 10, 20, 30, 35; Secondary 60F05.

Key words and phrases. Edgeworth, Cornish–Fisher, expansions, Berry–Esseen bounds, rank tests, (M) estimates, goodness of fit tests, U-statistics, deficiency.

we shall say the expansion is uniformly valid to $(r + 1)$ terms. (This is not quite in accord with Wallace who requires the remainder to be $O(N^{-(r+1)/2})$ but is more convenient and in accord with [19].) An expansion valid to one term is just an ordinary limit theorem. It is sometimes convenient to consider expansions in which the A_j also depend on N. They are then, of course, no longer uniquely defined.

These higher order terms are of interest on various grounds.

(1) Taking one or two terms of the expansion frequently improves the basic approximation A_0 strikingly. Examples of this phenomenon may be found in Hodges and Fix [28] and Thompson, Govindarajulu and Doksum [55].

(2) The higher order terms give some qualitative insight into regions of unreliability of first order results. For instance, when the limit A_0 is normal the higher order terms A_1 and A_2 typically correct for skewness and kurtosis.

(3) The expansions can be used to discriminate between procedures equivalent to first order, as for example in Hodges and Lehmann's work on deficiency [30].

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

Expansions of the type (1.1) and (1.2) are not the only ones of interest. Density functions and frequency functions of lattice random variables can sometimes be expanded. Extreme and intermediate tail probabilities can also sometimes be expanded (see for example [21], pages 517–520, [13] and [37]), and as P. Huber pointed out to me, the approximation to the power function of tests so obtained can be much more satisfactory than that based on the Edgeworth expansion. However, at least to date, the principal method used has been that of saddle point approximation which seems to require more intimate knowledge of the characteristic function of F_N than is usually available. In any case few if any such expansions appear to be available in nonparametric problems. Thus, we limit ourselves to discussion of expansions of types (1.1) ("Edgeworth") and the related expansions of F_N^{-1} ("Cornish–Fisher"). We shall deal primarily with expansions in which A_0 is the normal distribution. General results are available here for linear rank statistics (Section 2) and M estimates (Section 3) and partial results for linear combinations of order statistics and U-statistics (Section 4). What is known in nonnormal limiting situations is discussed briefly in Section 5.

2. The Berry–Esséen method and linear rank statistics. Suppose that a sequence $\{T_N\}$, $N \geqq 1$, of random variables tends to a standard normal distribution. If we let

$$\rho_N(t) = E(e^{itT_N}) \tag{2.1}$$

then we are asserting that there is a version of $\log \rho_N$ such that as $N \to \infty$,

$$\log \rho_N(t) \to -\frac{t^2}{2} . \tag{2.2}$$

Suppose that we have an asymptotic expansion of $\log \rho_N$ of the form,

$$\log \rho_N(t) = -\frac{t^2}{2} + \frac{P_1(it)}{N^{\frac{1}{2}}} + \cdots + \frac{P_r(it)}{N^{r/2}} + o(N^{-r/2}), \tag{2.3}$$

where the P_j are polynomials of order $\leqq j + 2$ which vanish at 0. Such a development is plausible if the T_N have cumulants $K_{j,N}$, such that $K_{1,N} = 0$, $K_{2,N} = 1$, $K_{j,N} = O(N^{-(j-2)/2})$, $j \geqq 3$, and which themselves admit asymptotic expansions in powers of $N^{-\frac{1}{2}}$. Thus if

$$K_{j,N} = \sum_{l=0}^{r-j+2} \frac{K_j^{(l)}}{N^{(j+l-2)/2}} + o(N^{-r/2}) \tag{2.4}$$

we should have,

$$P_k(it) = \sum_{j=3}^{k+2} \frac{K_j^{(k+2-j)}}{j!} (it)^j . \tag{2.5}$$

This is typically true although it sometimes requires a separate proof. The prototypical such T_N are, of course, standardized sums of independent identically distributed random variables. For more on expansions of the log characteristic function in terms of cumulants we refer the reader to the discussion in [57] and on pages 221–230 of [12]. Now, (2.3) corresponds to

$$\rho_N(t) = e^{-t^2/2}\left(1 + \sum_{j=1}^{r} \frac{Q_j(it)}{N^{j/2}}\right) + o(N^{-r/2}) \tag{2.6}$$

where

$$Q_1(it) = P_1(it)$$

$$Q_2(it) = P_2(it) + \frac{[P_1(it)]^2}{2}$$

and so on.

Normal Fourier inversion suggests that if

$$Q_j(it) = \sum_{k\geqq 1} a_{jk}(it)^k$$

then

$$F_N(x) = \Phi(x) - \phi(x)\left[\sum_{j=1}^{r} \frac{1}{N^{j/2}} \sum_{k\geqq 1} a_{jk} N_{k-1}(x)\right] + o(N^{-r/2}) \tag{2.7}$$

where Φ is the standard normal cdf, ϕ is the standard normal density and the N_k are Hermite polynomials defined by

$$\frac{d^k\phi(x)}{dx^k} = (-1)^k N_k(x)\phi(x) . \tag{2.8}$$

This formal step cannot, of course, be justified in general. It fails for instance if T_N is the standardized sum of independent identically distributed lattice random variables. The passage *is* valid if the weak (2.6) can be replaced by

$$\int_{-MN^{r/2}}^{MN^{r/2}} \left\{\left|\rho_N(t) - e^{-t^2/2}\left(1 + \sum_{j=1}^{r} \frac{Q_j(it)}{N^{j/2}}\right)\right| \Big/ |t|\right\} dt = o(N^{-r/2}) \tag{2.9}$$

for every $M < \infty$. An equivalent useful form of (2.9) is

$$\int_{-\varepsilon N^{\frac{1}{2}}}^{\varepsilon N^{\frac{1}{2}}} \left\{ \left| \rho_N(t) - e^{-t^2/2}\left(1 + \sum_{j=1}^{r} \frac{Q_j(it)}{N^{j/2}}\right) \right| \Big/ |t| \right\} dt = o(N^{-r/2}) \tag{2.10}$$

and

$$\int_{\{\varepsilon N^{\frac{1}{2}} \leq |t| \leq MN^{r/2}\}} \frac{|\rho_N(t)|}{|t|}\, dt = o(N^{r/2})$$

for some $\varepsilon > 0$ and every $M < \infty$. That (2.9) suffices follows from a famous lemma of Berry and Esséen whose statement and proof may be found in Feller [21], Chapter 16, page 510.

The validity of (2.9) and hence of (2.7) to order $1/N$ ($r = 2$) has been established for linear rank statistics both under the hypothesis of symmetry and under contiguous location alternatives by Albers, Bickel, and van Zwet [1]. A similar expansion for the two sample Wilcoxon statistic under the null hypothesis was established earlier by Rogers [48]. Expansions for general two sample rank statistics to order $1/N$ both under the hypothesis and contiguous location alternatives are in preparation [6]. Here is a selection of the results of these papers.

Let $X_1, \cdots, X_N$ be independent identically distributed with common cdf G and density g. Let $Z_{1:N} < \cdots < Z_{N:N}$ denote the ordered $|X_j|$. Define ranks $R_1, \cdots, R_N$ by

$$|X_{R_j}| = Z_{j:N} .$$

Let

$$\begin{aligned} \varepsilon_j &= 1 \qquad \text{if} \quad X_{R_j} > 0 \\ &= -1 \qquad \text{otherwise} , \end{aligned}$$

and suppose that $a_{1N}, \cdots, a_{NN}$ are given constants.

Define

$$T_N = \sum_{j=1}^{N} \frac{a_{jN}\,\varepsilon_j}{\sigma_N} \tag{2.11}$$

where

$$\sigma_N^2 = \sum_{j=1}^{N} a_{jN}^2 . \tag{2.12}$$

For simplicity suppose there exists a function J on $(0, 1)$ such that

$$a_{jN} = E(J(U_{j:N})) \tag{2.13}$$

where $U_{1:N} < \cdots < U_{N:N}$ are the order statistics of a sample of size N from the uniform distribution on $(0, 1)$. All of the usual statistics for testing the hypothesis that g is symmetric about 0, including the sign, Wilcoxon and normal scores tests can be put in this form. Hájek and Šidák [26] provide an extensive discussion of these procedures as well as the two sample tests we shall mention.

If g is symmetric about 0 the ε_j are independent with $P[\varepsilon_j = 1] = \frac{1}{2}$. The statistic T_N is then a sum of independent nonidentically distributed random variables, and

$$\rho_N(t) = \prod_{j=1}^{N} \cos \frac{ta_{jN}}{\sigma_N} . \tag{2.14}$$

If $\int_0^1 J^4(t)\,dt < \infty$, Taylor expansion of (2.14) yields

$$\log \rho_N(t) = -\frac{t^2}{2} - 2\frac{(it)^4}{4!}\sum_{j=1}^N \frac{a_{jN}^4}{\sigma_N^4} + o\left(\frac{1}{N}\right) \tag{2.15}$$
$$= -\frac{t^2}{2} - \frac{(it)^4}{12N}\frac{\int_0^1 J^4(t)\,dt}{(\int_0^1 J^2(t)\,dt)^2} + o\left(\frac{1}{N}\right).$$

If J is in addition continuously differentiable and nonconstant it is shown in [1] that (2.10) holds and hence that

$$\Phi(x) + \frac{\int_0^1 J^4(t)\,dt}{12N(\int_0^1 J^2(t)\,dt)^2}\,\phi(x)H_3(x)$$

is a uniformly valid expansion for F_N to three terms. In particular this proves the validity of the expansions used by Fellingham and Stoker [22] for the Wilcoxon test and by Thompson *et al.* [55] for the normal scores test up to terms of order smaller than $1/N$. Thompson *et al.* noted that the approximation using exact cumulants suggested by the first identity in (2.15) is better than the expansion suggested by the second identity while Fellingham and Stoker only considered the approximation using exact cumulants, with continuity correction. The exact cumulant Edgeworth expansion in both cases did provide substantial improvement over the normal approximation for $N = 10 - 20$ although the latter seems satisfactory for all practical purposes. It is not yet known whether the Edgeworth expansion for statistics such as the normal scores is valid to more than three terms. It seems clear that the expansion to order $1/N^2$ for the Wilcoxon with continuity correction used by Fellingham and Stoker can be justified by a local limit expansion and application of the Euler–Maclaurin formula. Local limit theorems for the two sample Wilcoxon statistic were developed by Rogers [48].

If g is not symmetric about 0 the ε_j are no longer independent. However by conditioning on $|X_1|, \cdots, |X_N|$ Albers, Bickel and van Zwet arrive at the following representation for ρ_N,

$$\rho_N(t) = E\{\textstyle\prod_{j=1}^N [P_{jN}\exp[ita_{jN}/\sigma_N] + (1 - P_{jN})\exp[-ita_{jN}/\sigma_N]]\} \tag{2.16}$$

where

$$P_{jN} = \frac{g(Z_{j:N})}{g(Z_{j:N}) + g(-Z_{j:N})}.$$

From this representation it may be shown that if $\int_0^1 J^4(t)\,dt < \infty$ and J is continuously differentiable and nonconstant then

$$\int_{-bN^{\frac{1}{2}}}^{bN^{\frac{1}{2}}} \{|\rho_N(t) - \tilde\rho_N(t)|/|t|\}\,dt \leqq cN^{-\frac{3}{2}}$$

for b, c depending on g where

$$\tilde\rho_N(t) = E\left\{\exp\left[itK_{1N} - \frac{t^2}{2}K_{2N}\right]\left(1 + \frac{(it)^3}{6}K_{3N} + \frac{(it)^4}{24}K_{4N} + \frac{(it)^6}{72}K_{3N}^2\right)\right\} \tag{2.17}$$

and

$$K_{1N} = \sum_{j=1}^{N} \frac{a_{jN}}{\sigma_N} (2P_{jN} - 1)$$

$$K_{2N} = 4 \sum_{j=1}^{N} \frac{a_{jN}^3}{\sigma_N^{\ 2}} P_{jN}(1 - P_{jN})$$

$$K_{3N} = 8 \sum_{j=1}^{N} \frac{a_{jN}^3}{\sigma_N^{\ 3}} P_{jN}(1 - P_{jN})(1 - 2P_{jN})$$

$$K_{4N} = 16 \sum_{j=1}^{N} \frac{a_{jN}^4}{\sigma_N^{\ 4}} P_{jN}(1 - P_{jN})(1 - 6P_{jN} + 6P_{jN}^2)$$

are the cumulants of T_N.

Further expansion for fixed alternatives appears to depend on the development of the theory of Edgeworth expansion for linear combinations of order statistics. However, if we permit g to depend on N in such a way that g is contiguous to a symmetric density, then K_{1N} is to first order a constant, and further expansion is possible. Specifically suppose that

$$g_N(x) = f(x - \theta_N) \tag{2.18}$$

where f is a fixed density symmetric about 0 and $\theta_N = \theta/N^{\frac{1}{2}}$. It is then shown in [1] under some regularity conditions on f, as well as the previously specified conditions on J, that for some b, c depending on f and J

$$\int_{-bN^{\frac{3}{2}}}^{bN^{\frac{3}{2}}} \{|\tilde{\rho}_N(t) - \gamma_N(t)|/|t|\} \, dt \leqq cN^{-\frac{5}{4}} \tag{2.19}$$

where

$$\gamma_N(t) = \exp\left[it\tilde{K}_{1N} - \frac{t^2}{2}\tilde{K}_{2N}\right]\left(1 + \frac{(it)^3}{6}\tilde{K}_{3N} + \frac{(it)^4}{24}\tilde{K}_{4N}\right) \tag{2.20}$$

and

$$\tilde{K}_{1N} = -\theta_N \sum_{j=1}^{N} \frac{a_{jN}}{\sigma_N} E_0(\psi_1(Z_{j:N}))$$
$$- \frac{\theta_N^{\ 3}}{3\sigma_N} \sum_{j=1}^{N} a_{jN} E_0[\tfrac{1}{2}\psi_3(Z_{j:N}) - 3\psi_1\psi_2(Z_{j:N}) + \tfrac{3}{2}\psi_1^{\ 3}(Z_{j:N})]$$

$$\tilde{K}_{2N} = 1 - \theta_N^{\ 2} \sum_{j=1}^{N} \frac{a_{jN}^2}{\sigma_N^{\ 2}} E_0(\psi_1(Z_{j:N}))^2 + \frac{\theta_N^{\ 2}}{\sigma_N^{\ 2}} \operatorname{Var}_0 \left(\sum_{j=1}^{N} a_{jN}\psi_1(Z_{j:N})\right)$$

$$\tilde{K}_{3N} = 2\theta_N \sum_{j=1}^{N} \frac{a_{jN}^3}{\sigma_N^{\ 3}} E_0(\psi_1(Z_{j:N}))$$

$$\tilde{K}_{4N} = -2 \sum_{j=1}^{N} \frac{a_{jN}^4}{\sigma_N^{\ 4}}$$

where

$$\psi_j(x) = \frac{f^{(j)}}{f}(x)$$

and the subscript 0 indicates that calculation is carried out under f. The $\tilde{K}_{jN}$ may be shown to be the leading terms in the expansion of the cumulants of T_N

under g_N. Berry's lemma can be applied to yield as a uniformly valid expansion for $F_N(t)$ to three terms

(2.21) $$\Phi(y_N) - \phi(y_N)\left\{\frac{\tilde{K}_{3N}}{6} N_2(y_N) + \frac{\tilde{K}_{4N}}{24} N_3(y_N)\right\}$$ where

$$y_N = \frac{t - \tilde{K}_{1N}}{(\tilde{K}_{2N})^{\frac{1}{2}}}.$$

This is not strictly speaking an expansion of the type we have been considering since N enters into the approximation in a complicated fashion. However, the expansion can be used in this form, for instance, to study power under normal alternatives since in this case

$$\psi_j(x) = (-1)^j H_j(x)$$

and moments of order statistics from the half normal distribution are available (cf. [34]).

If J' is defined and continuous on [0, 1] and f satisfies some mild regularity conditions, integral approximations to the $\tilde{K}_{jN}$ can be shown to hold, and a uniformly valid expansion to three terms as defined in Section 1 can be provided. This is adequate for the Wilcoxon but not the normal scores test. If we consider the distribution of the latter under normal alternatives it turns out that the $\tilde{K}_{1N}$ term does not admit an expansion of the form $A + B/N$ with A, B fixed, but rather requires a term of the form $(B \log \log N)/N$. As noted by Wallace, expansions of the type (2.7) can validly be inverted to yield expansions for percentiles (Cornish–Fisher) and hence expansions for the power functions of the rank statistics T_N. Agreement between the power function expansions for the normal scores and Wilcoxon tests obtained from (2.21) and (2.15) for normal and logistic alternatives appears to agree well with the Monte Carlo figures of Thompson *et al.* [55]. However, agreement with the Monte Carlo figures of Arnold [3] for the power function of the Wilcoxon test under Cauchy alternatives seems unsatisfactory.

In [30] Hodges and Lehmann introduced the notion of *deficiency* of a procedure with respect to an equally efficient competitor. For tests of equal level α, the deficiency is crudely defined as the limit of the difference in sample sizes required to reach equal power for the same alternative. The power functions expansions obtained in [1] are used to calculate the deficiency of the normal scores test with respect to the t test for normal alternatives. This turns out to be infinite but of the order of $\log \log N$. The results of [1] can also be used to establish that the permutation t test has deficiency 0 with respect to the t test under normal alternatives.

Suppose now that we have two samples $X_1, \cdots, X_m, Y_1, \cdots, Y_n$, $N = m + n$, the first sample being distributed with common density f, the second with common density g. Let $Z_{1:N} < \cdots < Z_{N:N}$ be the order statistics of the pooled sample and define

$$\begin{aligned} \varepsilon_j &= 1 \quad \text{if} \quad Z_{j:N} = Y_k \quad \text{for some} \quad k \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

A two sample linear rank statistic standardized under the null hypothesis is then given by

$$T_N = \sum_{j=1}^N a_{jN}\left(\varepsilon_j - \frac{n}{N}\right)\Big/\tau_N^2 \tag{2.22}$$

where the a_{jN} are specified scores

$$\tau_N^2 = [\sum_{j=1}^N (a_{jN} - \bar{a}_N)^2]\,\frac{mn}{N(N-1)} \tag{2.23}$$

and

$$\bar{a}_N = \frac{1}{N}\sum_{j=1}^N a_{jN}\,.$$

Suppose again that the a_{jN} are given by (2.13). Using a representation of the characteristic function ρ_N of T_N related to one due to Erdös and Rényi [20] and the Berry lemma, Bickel and van Zwet [6] obtain a uniformly valid expansion for the distribution function F_N of T_N to three terms if $f = g$, n/N stays bounded away from 0 and 1, $\int_0^1 J^4(t)\,dt < \infty$, and J is nonconstant and has continuous derivative. In this case,

$$F_N(x) = \Phi(x) - \phi(x)\left\{\frac{K_{3N}^*}{6}H_2(x) + \frac{K_{4N}^*}{24}H_3(x) + \frac{[K_{3N}^*]^2}{72}H_5(x)\right\} + o\left(\frac{1}{N}\right) \tag{2.24}$$

where the K_{jN}^* are the cumulants of T_N. Essentially this result was obtained by Rogers in [48] for the Wilcoxon statistic. Formal expansions were previously considered by Hodges and Fix [28]. A Berry–Esséen bound was obtained by Stoker [53]. Expansions of the power function and deficiency calculations are in progress [6]. Formal expansions of the power function were considered by Witting [58] using moment expansions due to Sundrum [54]. More Monte Carlo studies of the power functions of the two sample tests are desirable. Figures are available for the Savage test [17] when f and g are exponential densities and for the Wilcoxon and normal scores test under normal alternatives [34], [35], [41].

There are several open problems in this area. Two which I find interesting are:

(1) The extension of these results to tests of independence such as Spearman's ρ and Kendall's τ.

(2) The establishment of valid expansions for fixed alternatives.

3. Multivariate Edgeworth expansions and (M) estimates. A significant development in the theory of asymptotic expansions occurred in 1961 with the appearance of Ranga Rao's thesis on Edgeworth expansions and Berry–Esséen bounds for sums of independent random vectors. Since then there has been considerable development in the field. Some results typical of the most recent state of the art and many references to older work may be found in Bhattacharya's paper [5] in which the following theorem is announced.

Let $\{X^{(r)} = (X_1^{(r)}, \cdots, X_k^{(r)})\}$ be a sequence of independent identically distributed k dimensional random vectors. Suppose that

$$(3.1)\qquad E(X_i^{(1)}) = 0\,, \qquad i = 1, \cdots, k$$
$$E(X_i^{(1)}X_j^{(1)}) = \delta_{ij}\,, \qquad 1 \leqq i \leqq j \leqq k\,.$$

Let

$$(3.2)\qquad \rho(u) = E(e^{iuX^{(1)}})$$

where $u = (u_1, \cdots, u_k)$ and $uX^{(1)}$ is the inner product of u and $X^{(1)}$. As usual consider the formal expansion of $\rho^N(u/N^{\frac{1}{2}})e^{|u|^2/2}$ where $|u|^2 = \sum_{i=1}^k u_i^2$, as a power series in $N^{-\frac{1}{2}}$

$$(3.3)\qquad e^{|u|^2/2}\rho^N\left(\frac{u}{N^{\frac{1}{2}}}\right) = 1 + \sum_{j=1}^{\infty} \frac{P_j(iu)}{N^{j/2}}$$

where the P_j are polynomials whose coefficients depend on the cumulants of $X^{(1)}$. Define polynomials $\tilde{P}_j$ on R^k by the property that $(2\pi)^{-k/2}e^{-|t|^2/2}\tilde{P}_j(t)$ has $e^{-|u|^2/2}P_j(iu)$ as its Fourier transform. For any $A \subset R^k$, let $(\partial A)^\varepsilon$ be the set of all points within a distance ε of the boundary of A, i.e.,

$$(3.4)\qquad (\partial A)^\varepsilon = \{x \in R^k : \exists y \in A,\, z \notin A \ni |x - y| < \varepsilon,\, |x - z| < \varepsilon\}\,.$$

Let $\mathscr{A}(\Phi : d, \varepsilon_0)$ be the class of all Borel sets A such that

$$\Phi((\partial A)^\varepsilon) \leqq d\varepsilon\,, \qquad 0 < \varepsilon \leqq \varepsilon_0$$

where Φ is the standard multivariate normal product probability measure on R^k.

We need Cramér's condition

$$(\mathrm{C})\qquad \limsup_{|u|\to\infty} |\rho(u)| < 1\,.$$

THEOREM (Remark 1, page 255 of [5]). *Suppose that* $E|X_j^{(1)}|^s < \infty$, $1 \leqq j \leqq k$, *for some* $s \geqq 3$, *the* $X^{(j)}$ *are as above and that condition* (C) *holds. Let* $S_N = \sum_{j=1}^N X^{(j)}$. *Then, for every* $d > 0$,

$$(3.5)\qquad \sup\left\{\left|P\left[\frac{S_N}{N^{\frac{1}{2}}} \in A\right] - (2\pi)^{-k/2}\int\cdots\int_A e^{-|t|^2/2} \times\left[1 + \sum_{j=1}^{s-2}\frac{\tilde{P}_j(t)}{N^{j/2}}\right]dt\right| : A \in \mathscr{A}(\Phi : d, \varepsilon_0)\right\} = o(N^{(s-2)/2})\,.$$

By making a linear transformation of the variables this result can obviously be extended to the case that $X^{(1)}$ has a specified nonsingular covariance matrix. These results have been applied in a variety of problems involving expansions of multivariate distributions connected with normal variables. An interesting paper along these lines which also faces the problem of computation of the $\tilde{P}_j(t)$ is that of Chambers [10].

In this section we review the work of Linnik and Mitrofanova [38], [56] and Čibišov [11] who employed results of this type to obtain asymptotic expansions for maximum likelihood estimates, and the related work of Pfanzagl [45], [46] and Michel and Pfanzagl [40]. The work is of interest from the point of view of robust estimation since the same technique yields expansions for Huber's (M) estimates [32], [33].

Let

$$X_j = \theta + E_j\,, \qquad 1 \leqq j \leqq N$$

where the E_j are independent identically distributed with density f. An (M) estimate (scale known) of θ, for given ψ, is by definition, any solution $\hat\theta$ of the equation

$$\sum_{j=1}^{N} \psi(X_j - \hat\theta) = 0\,. \tag{3.6}$$

For the estimation to make sense we suppose

$$E_0(\psi(X_1 - \theta)) = 0\,. \tag{3.7}$$

Condition for consistency and asymptotic normality of such estimates are given in [32] and [33].

Linnik and Mitrofanova [38], in the tradition of Cramér [12], obtained expansions for a solution of (3.6) when $\psi = -f'/f$. It is easy to see in the light of [33] how their conditions should be modified to yield expansions for (M) estimates. It should be noted that [38] has many obscure points and, in particular, it seems to me that the appeal to Ranga Rao's theorem [47] at a crucial point in [38] is inadequate. However, I believe application of the more sophisticated theorem of Bhattacharya that was stated above will carry the proof through.

The main idea which was already used by Haldane and Smith [27] and Shenton and Bowman [9] for formal cumulant expansions of maximum likelihood estimates is to expand the likelihood equation beyond the customary two terms.

$$\begin{aligned} 0 = N^{-\frac{1}{2}} \sum_{j=1}^{N} \psi(X_j - \theta) - \left\{\frac{1}{N} \sum_{j=1}^{N} \psi'(X_j - \theta)\right\} N^{\frac{1}{2}}(\hat\theta - \theta) + \cdots \\ + N^{-(k-1)/2} \frac{(-1)^k}{k!} \left\{\frac{1}{N} \sum_{j=1}^{N} \psi^{(k)}(X_j - \theta)\right\} N^{k/2}(\hat\theta - \theta)^k + R_{Nk}\,. \end{aligned} \tag{3.8}$$

Using the expansion to two terms and suitable conditions on the derivatives of ψ the first step is to show that large deviations of a suitable root of (3.6) are very unlikely and hence that R_{Nk} which is governed by $N^{\frac{1}{2}}(\hat\theta - \theta)^{k+1}$ can be bounded by something only slightly larger than $N^{-k/2}$. The next step is to consider the equation

$$\begin{aligned} 0 = N^{-\frac{1}{2}} \sum_{j=1}^{N} \psi(X_j - \theta) - \left\{\frac{1}{N} \sum_{j=1}^{N} \psi'(X_j - \theta)\right\} N^{\frac{1}{2}}(t - \theta) + \cdots \\ + N^{-(k-1)/2} \frac{(-1)^k}{k!} \left\{\frac{1}{N} \sum_{j=1}^{N} \psi^{(k)}(X_j - \theta)\right\} N^{k/2}(t - \theta)^k\,. \end{aligned} \tag{3.9}$$

The solution $t = \hat\theta_1$ of this equation can be expanded in an asymptotic expansion in $N^{-\frac{1}{2}}$ whose leading term is $N^{-\frac{1}{2}} \sum_{j=1}^{N} \psi(X_j - \theta)/E_\theta(\psi'(X_1 - \theta))$ and whose coefficients are polynomials in $\xi_0, \cdots, \xi_k$ where

$$\xi_r = \frac{1}{N^{\frac{1}{2}}} \sum_{j=1}^{N} [\psi^{(r)}(X_j - \theta) - E_\theta(\psi^{(r)}(X_j - \theta))] . \tag{3.10}$$

Then one shows that $\hat{\theta}$ and $\hat{\theta}_1^{(k)}$, the sum of the first k terms in the expansion of $\hat{\theta}_1$, differ to an order that matters only on a set of relatively negligible probability. Then one applies a theorem such as Bhattacharya's to the event $[N^{\frac{1}{2}}(\hat{\theta}_1^{(k)} - \theta) < x]$ which indeed depends only on $(\xi_0, \cdots, \xi_k)$. Finally there is the problem of expanding the multivariate integrals appearing in the multivariate Edgeworth theorem since these depend on N (since $\hat{\theta}_1^{(k)}$ is a polynomial in powers of $N^{-\frac{1}{2}}$ as well as in the ξ_j). The result is an expansion of the type (2.7). It is formally clear that the coefficients should agree with those obtained by using the formal expansions of the cumulants in powers of $N^{-\frac{1}{2}}$ from [27] and then proceeding to get a formal Edgeworth expansion from the formal Charlier expansion as in (2.4) and (2.5). However, this has not been checked to my knowledge.

Mitrofanova [42] extended the work of [38] to maximum likelihood estimates of a vector parameter. Unfortunately, as was noted by Pfanzagl [46], her proof contains very serious gaps. A salvage operation however seems both possible and worthwhile. In particular this should yield valid expansions for (M) estimates when scale is estimated (as it normally would be). Čibišov's announcement [11] is essentially an extension of the work of [38] to maximum likelihood estimation of a single parameter under rather simple conditions.

Pfanzagl [46] and Michel and Pfanzagl [40] have used a different approach which though much simpler for the case of a single parameter does not appear to generalize. The idea similar to that used by Huber in [32] and earlier by H. E. Daniels [14] is to compare the events $[\hat{\theta} < x]$ and $[\sum_{j=1}^{N} \psi(X_j - x) < 0]$. For increasing ψ the two events are essentially the same. In general even for functions of the form $\psi(x, \theta)$, under suitable conditions, one can argue that the difference of the two events has negligible probability for $x = \theta + a/N^{\frac{1}{2}}$ with $|a|$ bounded. But to $P[\sum_{j=1}^{N} \psi(X_j - x) < 0]$ one can apply the classical univariate expansions for sums of independent identically distributed random variables and then use suitable expansions in $(x - \theta)/N^{\frac{1}{2}}$ of the cumulants of $\psi(X_1 - x)$. This method has the advantage of enabling one to deal with ψ functions which are not very smooth such as those introduced by Huber [32]. There seems at present, however, to be no way of dealing with (M) estimates in which scale is estimated simultaneously when the functions defining the estimates cannot be expanded along the lines of [33].

Pfanzagl [46] gives a variety of applications to parametric models of the univariate expansions mentioned above. There have been hardly any numerical studies of the applicability of these expansions. An interesting example, however, is Barnett's work [4] in which he shows that the (formal) expansion is relatively poor when applied to the maximum likelihood estimate of location for a Cauchy sample.

4. Other classes of asymptotically normal statistics. There has been little success so far in validating expansions or even establishing Berry–Esséen bounds of order $1/N^{\frac{1}{2}}$ for general classes of statistics known to be asymptotically normally distributed, other than the ones we have discussed.

Mr. S. Bjerve in work towards a Berkeley thesis has shown that trimmed means admit valid Edgeworth expansions and is in the process of explicitly calculating the coefficients for comparison with the published distributions of the Princeton project [2]. His method employs special properties of the trimmed means and does not carry over to more general estimates. Further work on systematic statistics which can also be handled by elementary means is intended. Even formal work seems surprisingly scarce here. In this connection I would like to mention [16] in which expansions are obtained for the cumulants of single order statistics.

The only theoretical result on rates of convergence for general linear combinations of order statistics known to me is due to Rosenkrantz and O'Reilly [43] who establish various bounds of Berry–Esséen type for the error committed by using the normal approximation to the distribution of a linear combination of order statistics. None of these bounds is of smaller order than $N^{-\frac{1}{4}}$ where N is the sample size. This limitation appears due to the Skorokhod embedding method which they employ. This order is, of course, incorrect for all cases in which sharp bounds are available, i.e., trimmed means (including the mean) and systematic statistics. I conjecture that under mild conditions the "right" order is $N^{-\frac{1}{2}}$.

In 1948 Hoeffding [31] introduced the interesting class of U-statistics, which includes among its members the Wilcoxon two sample statistic. As another illustration of the power of the Fourier technique in a nonstandard situation we shall prove under rather strong conditions that the normal approximation to the distribution of a U-statistic of order 2 is valid to order $N^{-\frac{1}{2}}$. Our method can be adapted to yield the $N^{-\frac{1}{2}}$ bound for the one and two sample Wilcoxon statistic as well as Kendall's τ. (In fact fixed alternative asymptotic expansions for these statistics can be obtained using a combination of the methods of the appendix and those of [1].) The method should also extend to von Mises statistics [56] of order 1 and hence to linear combinations of order statistics. However we are unable to get $N^{-\frac{1}{2}}$ bounds for U-statistics with unbounded kernels. Bounds of order $N^{-r/2}$, $r < 1$, have been obtained by Grams and Serfling in [25] by a different technique. Asymptotic expansions in general seem out of reach. Here is the statement of our theorem. The proof is given in an appendix.

Let $R_1, \cdots, R_N$ be a sample from the uniform distribution on (0, 1). Let ψ be a measurable real-valued function on the closed unit square such that $|\psi| \leqq M < \infty$ (say). Suppose moreover that ψ is symmetric, $\psi(u, v) = \psi(v, u)$ and that

$$\int_0^1 \int_0^1 \psi(u, v)\, du\, dv = 0 . \tag{4.1}$$

Let

$$T_N = \frac{1}{\sigma_N} \sum_{i<j} \psi(R_i, R_j) \tag{4.2}$$

where

$$\sigma_N^2 = \frac{N(N-1)}{2} \int_0^1 \int_0^1 \psi^2(u, v)\, du\, dv + N(N-1)(N-2) \int_0^1 \gamma^2(u)\, du \tag{4.3}$$

and

$$\gamma(u) = \int_0^1 \psi(u, v)\, dv\,. \tag{4.4}$$

THEOREM 4.1. *If the preceding assumptions hold and γ does not vanish identically, then there exists a constant C depending on ψ but not N such that*

$$\sup_x |P[T_N \leqq x] - \Phi(x)| \leqq \frac{C}{N^{\frac{1}{2}}}$$

where Φ is the standard normal cumulative distribution function.

A new approach has recently been advanced by Stein [52] which does not rely on Fourier analytic methods. Using his method he is able to show that the error committed in applying the normal approximation to the sum of the first N of a stationary sequence of bounded m dependent random variables is of order $N^{-\frac{1}{2}}$. The possibility of applying his method to some of the classes we have considered should be investigated.

5. Expansions for statistics with nonnormal limiting distributions. The omnibus goodness of fit and two sample tests such as those of Kolmogorov–Smirnov and Cramér–von Mises and the Pearson χ^2 test do not have limiting normal distributions. The Russian school of probability theorists has had considerable success in obtaining expansions for the distribution of the Kolmogorov–Smirnov test statistics under the null hypothesis. The methods employed at first used explicit representations of the null distribution. An account of results of this type due to Chan Li–Tsien may be found in Gnedenko, Korolyuk, Skorokhod [23]. The most definitive expansion for the one-sided goodness of fit statistic was given by Lauwerier [36]. Subsequently, the problems were treated as special cases of more general problems of first passage times of random walks (cf. for example Borovkov [7] in which the two sample Smirnov statistic is treated). An account of the latest results and extensive references may be found in Borovkov [8]. Since none of the first order limiting distributions under contiguous alternatives for these statistics have been tabled or extensively studied it is not surprising that there has been no work on asymptotic expansions for the power.

There has recently been some interest in obtaining Berry–Esséen type bounds for the difference between the distribution of the Cramér–von Mises goodness of fit statistic under the null hypothesis and its well known limit distribution. However, the methods used by Rosenkrantz in [49] and Sawyer in [50] (cf. also Orlov [44]) use the Skorokhod embedding and not surprisingly obtain bounds which are of order strictly worse than $N^{-\frac{1}{2}}$ where N is the sample size. In an announcement of results without proofs [15] D. Darling obtained a representation for the characteristic function of the von Mises statistic which he employed to get an asymptotic expansion of the characteristic function to two terms for fixed argument. I do not know whether this approach can be refined to yield the kind of estimates which permit us to apply Berry's lemma.

Finally, I want to mention the recent Chicago thesis of Yarnold [59] in which he obtained asymptotic expansions for the distribution of Pearson's χ^2 statistic. Since χ^2 is a smooth function of the multinomial frequencies we might expect that the theorems on multivariate Edgeworth series should apply. Unfortunately the vector of multinomial frequencies is a normalized sum of independent identically distributed random vectors taking their values in a lattice, Cramér's condition (C) does not hold and in fact the formal Edgeworth expansion is invalid. However, it is possible to use the well-known local limit expansion for the multinomial probability and then sum up over all points in the appropriate region. This is an improvement over the χ^2 approximation but almost as complicated as calculation of the exact probabilities. Moreover, it does not yield a form which is sufficiently tractable analytically to settle long outstanding questions about the relative performance of the χ^2 and likelihood ratio tests. Results which are manageable in this area would be interesting but seem hard.

6. Appendix (Proof of Theorem 4.1). Let

$$S_N = \frac{(N-1)}{\sigma_N} \sum_{i=1}^N \gamma(R_i) \tag{6.1}$$

$$\Delta_N = T_N - S_N \tag{6.2}$$

$$\phi_N(t) = E(e^{itT_N}) \tag{6.3}$$

$$\eta(t) = E(e^{it\gamma(R_1)}) \tag{6.4}$$

$$\tilde{\phi}_N(t) = E(e^{itS_N}) = \eta^N\left(\frac{t(N-1)}{\sigma_N}\right). \tag{6.5}$$

The crux of the argument is to show that there exists $\varepsilon_1 > 0$ and a constant D_1 both independent of N such that

$$\int_{-\varepsilon_1 N^{\frac{1}{2}}}^{\varepsilon_1 N^{\frac{1}{2}}} \frac{|\phi_N(t) - \tilde{\phi}_N(t)|}{|t|}\, dt \leqq D_1 N^{-\frac{1}{2}}. \tag{6.6}$$

Since it is well known that there exists $\varepsilon_2 > 0$ and a constant D_2 both independent of N such that

$$\int_{-\varepsilon_2 N^{\frac{1}{2}}}^{\varepsilon_2 N^{\frac{1}{2}}} \frac{|\tilde{\phi}_N(t) - e^{-t^2/2}|}{|t|}\, dt \leqq D_2 N^{-\frac{1}{2}},$$

it follows that if $\varepsilon = \min(\varepsilon_1, \varepsilon_2)$, $D = D_1 + D_2$,

$$\int_{-\varepsilon N^{\frac{1}{2}}}^{\varepsilon N^{\frac{1}{2}}} \frac{|\phi_N(t) - e^{-t^2/2}|}{|t|}\, dt \leqq D N^{-\frac{1}{2}}, \tag{6.7}$$

and the theorem follows from (6.6) and the usual Berry–Esséen argument.

To prove (6.6) we need the following lemmas.

LEMMA 6.1. *Let* $\{\xi_j\}$, $1 \leqq j \leqq n$ *be a sequence of martingale summands, i.e.,*

$$E(\xi_j \mid \xi_1, \cdots, \xi_{j-1}) = 0 , \qquad 1 \leqq j \leqq n .$$

Let $W_n = \sum_{j=1}^n \xi_j$. *Define* $m_{n,k} = \max_{1\leqq j\leqq n} E(\xi_j^{2k})$, $k \geqq 1$. *Then, for* $k \leqq n$,

$$E(W_n^{2k}) \leqq n^k m_{n,k}(4ek)^k . \tag{6.8}$$

REMARKS. (1) An estimate similar to (6.8) has been obtained by Dharmadhikari, Fabian and Jogdeo [18] with $m_{n,k}$ replaced by $(1/n)\sum_{j=1}^n E(\xi_j^{2k})$. However, their bound grows with k as 2^{k^2} which is quite inadequate for our purposes. We note that our technique readily establishes,

$$E(W_n^{2k}) \leqq n^k m_{n,k}(k)^{2k}$$

for all k, n but even this is inadequate.

(2) The example of ξ_j i.i.d. normal random variables with mean 0 shows that our bound is comparatively sharp. Also see the remark on Lemma 6.2.

Our main interest in Lemma 6.1 is in its application to

LEMMA 6.2. *Under the conditions of Theorem* 4.1, *if* $k \leqq N$,

$$E(\Delta_N^{2k}) \leqq \sigma_N^{-2k} N^{2k}(3M)^{2k}(4ek)^{2k} . \tag{6.9}$$

REMARK. The order of magnitude of the coefficient of $\sigma^{N-2k}N^{2k}$ in (6.9) is quite sharp. Thus if $\psi(x, y) = \frac{3}{4}$ if x and y are both $\geqq \frac{1}{2}$, $= -\frac{1}{4}$ otherwise

$$\sigma_N \Delta_N = \sum_{i<j} \eta_i \eta_j = \frac{1}{2}\left[(\sum_{i=1}^N \eta_i)^2 - \frac{N}{4}\right] \tag{6.10}$$

where the η_i are independent and equal $\pm\frac{1}{2}$ with equal probability $\frac{1}{2}$. It is easy to see that

$$E(\sigma_N \Delta_N)^{2k} \geqq 8^{-2k}\{2^{-2k+1}E(U_N^{4k}) - N^{2k}\} \tag{6.11}$$

where $U_N = \sum_{i=1}^N \varepsilon_i$ and $\varepsilon_i = \pm 1$ with probability $\frac{1}{2}$. Since,

$$E(U_N^{4k}) = \sum_{t_1+\cdots+t_N=2k} \frac{4k!}{2t_1! \cdots 2t_N!} ,$$

$$E(U_N^{4k}) \geqq \binom{N}{2k}\frac{4k!}{2^{2k}} \geqq A(kN)^{2k}\left(1 - \frac{(2k-1)}{N}\right)^{2k}\left(\frac{4}{e}\right)^{2k}$$

for some universal constant A and hence,

$$(kN)^{-2k}E(\sigma_N \Delta_N)^{2k} \geqq c\rho^k$$

for all N and $k \leqq aN$, $a < \frac{1}{2}$ where c and ρ depend on a but not on k and N. Then the ratio between $E(\sigma_N \Delta_N)^{2k}$ and the estimate given by (6.9) is (relatively) negligible.

PROOF OF LEMMA 6.1. The proof is by induction on n for fixed k. Note first that

$$E(\xi_1 + \cdots + \xi_k)^{2k} \leqq k^{2k} m_{k,k} \tag{6.12}$$

and hence the induction hypothesis holds for $n = k$. Suppose it is true for $n = l \geqq k$. Then

$$E(W_{l+1}^{2k}) = E(W_l^{2k}) + \sum_{j=2}^{2k} \binom{2k}{j} E(W_l^{2k-j}\xi_{l+1}^j) \tag{6.13}$$

by the martingale hypothesis. By induction and the Hölder inequality we obtain

$$\begin{aligned} E(W_l^{2k-j}\xi_{l+1}^j) &\leqq [c_k l^k m_{l,k}]^{1-j/2k}[m_{l+1,k}]^{j/2k} \\ &\leqq (c_k l^k m_{l+1,k})(c_k^{1/2k} l^{\frac{1}{2}})^{-j} \end{aligned} \tag{6.14}$$

where $c_k = (4ek)^k$. By elementary estimates (6.13) and (6.14) yield

$$\begin{aligned} E(W_{l+1}^{2k}) &\leqq c_k l^k m_{l+1,k}\left(1 + \frac{4k^2}{lc_k^{1/k}} \sum_{j=0}^{2k-2} \binom{2k-2}{j} (c_k^{1/2k} l^{\frac{1}{2}})^{-j}\right) \\ &\leqq c_k l^k m_{l+1,k}\left(1 + \frac{k}{le}\left(1 + \frac{1}{2(ekl)^{\frac{1}{2}}}\right)^{2k-2}\right) \\ &\leqq c_k l^k m_{l+1,k}\left(1 + \frac{k}{l}\right) \end{aligned} \tag{6.15}$$

for $k \leqq l$. Since $(1 + k/l) \leqq ((l+1)/l)^k$ the hypothesis is verified for $n = l + 1$ and the result follows.

PROOF OF LEMMA 6.2. Begin by noting that

$$\sigma_N \Delta_N = \sum_{j=1}^N \xi_j \tag{6.16}$$

where

$$\xi_j = \sum_{i=1}^{j-1} [\psi(R_i, R_j) - \gamma(R_i) - \gamma(R_j)] \tag{6.17}$$

and that the ξ_j are martingale summands. Moreover, note that

$$E(\xi_j^{2k}) = E(E[\sum_{i=1}^{j-1} (\psi(R_i, R_j) - \gamma(R_i) - \gamma(R_j))^{2k} \mid R_j]) \tag{6.18}$$

and that given R_j the summands $\eta_i = (\psi(R_i, R_j) - \gamma(R_i) - \gamma(R_j))$, $i = 1, \cdots, j-1$ are also martingale summands (in fact i.i.d.). Since

$$E(\psi(R_1, R_2) - \gamma(R_1) - \gamma(R_2))^{2k} \leqq (3M)^{2k} \tag{6.19}$$

we can apply Lemma 6.1 twice in succession to obtain Lemma 6.2.

LEMMA 6.3. *Under the conditions of the theorem,*

$$|E(e^{itS_N}\Delta_N)| \leqq 3M^3t^2 \frac{N^4}{\sigma_N^3} |\eta|^{N-2}\left(\frac{t}{\sigma_N}(N-1)\right) \tag{6.20}$$

$$|E(e^{itS_N}\Delta_N^j)| \leqq \left(\frac{N^2}{\sigma_N}\right)^j \left(\frac{3M}{2}\right)^j |\eta|^{N-2j}\left((N-1)\frac{t}{\sigma_N}\right) \quad \textit{for } j \geqq 1\,. \tag{6.21}$$

PROOF. To prove (6.20) we calculate

$$\begin{aligned} E(\Delta_N e^{itS_N}) = \frac{N(N-1)}{2\sigma_N}\, &\eta^{N-2}\left(\frac{t}{\sigma_N}(N-1)\right) \\ &\times E\left(\exp\left[\frac{it(N-1)}{\sigma_N}(\gamma(R_1) + \gamma(R_2))\right]\right. \\ &\left.\times (\psi(R_1, R_2) - \gamma(R_1) - \gamma(R_2))\right). \end{aligned} \tag{6.22}$$

Since $\psi(R_1, R_2) - \gamma(R_1) - \gamma(R_2)$ and $\gamma(R_1), \gamma(R_2)$ are uncorrelated we can write

$$\begin{aligned}
&\left|E\left(\exp\left[\frac{it(N-1)}{\sigma_N}(\gamma(R_1)+\gamma(R_2))\right](\psi(R_1, R_2)-\gamma(R_1)-\gamma(R_2))\right)\right| \\
&\quad = \left|E\left[\left(\exp\left[\frac{it(N-1)}{\sigma_N}(\gamma(R_1)+\gamma(R_2))\right]-1\right)\right.\right. \\
&\qquad \left.\left.\times (\psi(R_1, R_2)-\gamma(R_1)-\gamma(R_2))\right]\right| \qquad (6.23) \\
&\quad \leqq \frac{t^2}{2}\frac{(N-1)^2}{\sigma_N^{\ 2}} E[(\gamma(R_1)+\gamma(R_2))^2|\psi(R_1, R_2)-\gamma(R_1)-\gamma(R_2)|] \\
&\quad \leqq 6M^3t^2\frac{(N-1)^2}{\sigma_N^{\ 2}},
\end{aligned}$$

and (6.20) follows.

Similarly,

$$\begin{aligned}
(6.24)\qquad &\sigma_N^{\ j}E(\Delta_N^{\ j}e^{itS_N}) \\
&\quad = \textstyle\sum_{\{(a_1,b_1),\cdots,(a_j,b_j)\}} E(e^{itS_N}[\prod_{i=1}^{j}(\psi(R_{a_i}, R_{b_i})-\gamma(R_{a_i})-\gamma(R_{b_i}))]) .
\end{aligned}$$

Applying elementary inequalities we obtain

$$\begin{aligned}
&|\sigma_N^{\ j}E(\Delta_N^{\ j}e^{itS_N})| \\
(6.25)\qquad &\quad \leqq \frac{N^{2j}}{2^j}|\eta|^{N-2j}\left((N-1)\frac{t}{\sigma_N}\right)E|\psi((R_1, R_2)-\gamma(R_1)-\gamma(R_2))|^j \\
&\quad \leqq \left(\frac{3M}{2}\right)^j N^{2j}|\eta|^{N-2j}\left(\frac{(N-1)t}{\sigma^N}\right).
\end{aligned}$$

The lemma follows.

We proceed with the proof of (6.6). Since

$$|\phi_N(t)-\tilde{\phi}_N(t)| = |E(e^{itS_N}(e^{it\Delta_N}-1))|$$

we have for any k,

$$(6.26)\qquad |\phi_N(t)-\tilde{\phi}_N(t)| \leqq \left|\textstyle\sum_{j=1}^{2k-1}\frac{(it)^j}{j!}E(e^{itS_N}\Delta_N^{\ j})\right| + \frac{t^{2k}}{(2k)!}E(\Delta_N^{\ 2k}) .$$

From (6.26), (6.9) and (6.20),

$$(6.27)\qquad |\phi_N(t)-\tilde{\phi}_N(t)| \leqq \left(3M^3\frac{N^4}{\sigma_N^{\ 3}}|\eta|^{N-2}\left(\frac{t}{\sigma_N}(N-1)\right)t + 8e^2\frac{N^2}{\sigma_N^{\ 2}}M^2\right)t^2 .$$

Since there exists $\theta > 0$ such that $\sigma_N^{\ 2} \geqq \theta^2N^3$ for all N we conclude that

$$\begin{aligned}
(6.28)\qquad &\int_{-N^{\frac{1}{2}}}^{N^{\frac{1}{2}}}\frac{|\phi_N(t)-\tilde{\phi}_N(t)|}{|t|}dt \\
&\quad \leqq \frac{3N^{-\frac{1}{2}}M^3}{\theta^3}\int_{-N^{\frac{1}{2}}}^{N^{\frac{1}{2}}}|t|^2|\eta|^{N-2}\left(\frac{tN^{-\frac{1}{2}}}{\theta}\right)dt + \frac{8e^2M^2}{\theta^2}N^{-\frac{1}{2}} \\
&\quad \leqq FN^{-\frac{1}{2}}
\end{aligned}$$

where F is a constant depending on ψ but not N.

Let

$$\varepsilon = \frac{\theta p}{24Me}, \qquad p < 1 \tag{6.29}$$

$$k = \left\{\left(\left[\frac{1}{2}\frac{\log N}{|\log p|}\right] + 1\right) \wedge N\right\}.$$

If $|t| \leqq \varepsilon N^{\frac{1}{2}}$, by Lemma 6.2 for this k and N sufficiently large,

$$\frac{t^{2k}}{(2k)!} E(\Delta_N{}^{2k}) \leqq \frac{\varepsilon^{2k} N^k}{(2k)!} \cdot \frac{N^{2k}}{\sigma_N{}^{2k}} k^{2k}(12eM)^{2k} \tag{6.30}$$

$$\leqq \binom{4k}{2k} 2^{-4k} p^{2k} < N^{-1}.$$

To complete the argument note that for p sufficiently small, there exists $\tau > 0$ such that for $|t| \leqq \varepsilon N^{\frac{1}{2}}$,

$$\log |\eta| \left(\frac{(N-1)t}{\sigma_N}\right) \leqq -\frac{\tau t^2}{N}. \tag{6.31}$$

Applying (6.31) and (6.21) we conclude that for $N^{\frac{1}{2}} \leqq |t| \leqq \varepsilon N^{\frac{1}{2}}$, $j < 2k$,

$$|E(e^{itS_N} \Delta_N{}^j)| \leqq \theta^{-j} N^{j/2} \left(\frac{3M}{2}\right)^j \exp\left[-\tau N^{\frac{1}{2}}\left(1 - \frac{4k}{N}\right)\right]. \tag{6.32}$$

Hence for $N^{\frac{1}{2}} \leqq |t| \leqq \varepsilon N^{\frac{1}{2}}$ with k, ε given by (6.29),

$$\left|\sum_{j=1}^{2k-1} \frac{(it)^j}{j!} E(e^{itS_N} \Delta_N{}^j)\right| \leqq eN^{2k} \left(\frac{3M}{2\theta}\right)^{2k} \exp\left[-\tau N^{\frac{1}{2}}\left(1 - \frac{4k}{N}\right)\right] \tag{6.33}$$

$$= o\left(\frac{1}{N}\right)$$

uniformly for $|t|$ as above. Combining (6.28), (6.30) and (6.33), (6.6) and the theorem follows.

REFERENCES

[1] ALBERS, W., BICKEL, P. J. and VAN ZWET, W. R. (1972). Asymptotic expansions for the power of distribution free tests in the one sample problem. To be submitted to *Ann. Statist.*

[2] ANDREWS, D. F., BICKEL, P. J., HAMPEL, F. R., ROGERS, W. H. and TUKEY, J. W. (1972). *Robust Estimation: Survey and Advances.* Princeton Univ. Press.

[3] ARNOLD, H. J. (1965). Small sample power for the one sample Wilcoxon test. *Ann. Math. Statist.* **36** 1767–1778.

[4] BARNETT, V. (1966). Evaluation of maximum likelihood estimate when likelihood has multiple roots. *Biometrika* **53** 151–165.

[5] BHATTACHARYA, R. N. (1971). Rates of convergence and asymptotic expansions. *Ann. Math. Statist.* **42** 241–259.

[6] BICKEL, P. J. and VAN ZWET, W. R. (1973). Asymptotic expansions for the power of distribution free tests in the two sample problem. In preparation.

[7] BOROVKOV, A. A. (1962). On the problem of two samples. *Selected Transl. Math. Statist. Prob.* **5** 285–307.

[8] BOROVKOV, A. A. (1970). Limit theorems for random walks with boundaries. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **3** 19–30.

[9] BOWMAN, K. and SHENTON, L. R. (1963). Higher moments of maximum likelihood estimates. *J. Roy. Statist. Soc. Ser. B* **25** 305–317.

[10] CHAMBERS, J. M. (1967). On methods of asymptotic approximation for multivariate distributions. *Biometrika* **54** 367–384.

[11] ČIBIŠOV, D. M. (1972). Asymptotic expansions for maximum likelihood estimates. *Teor. Verojatnost. i Primenen.* **17** 387–388.

[12] CRAMÉR, H. (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press.

[13] DANIELS, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25** 631–650.

[14] DANIELS, H. E. (1960). The asymptotic efficiency of a maximum likelihood estimator. *Proc. Fourth Berkeley Symp. Math. Statist.* **1** 151–164.

[15] DARLING, D. (1960). Sur Les Théorèmes de Kolmogorov-Smirnov. *Theor. Probability Appl.* **5** 356–361.

[16] DAVID, F. N. and JOHNSON, N. L. (1954). Statistical treatment of censored data, Part I. *Biometrika* **41** 228–240.

[17] DAVIES, R. B. (1971). Rank tests for Lehmann's alternative. *J. Amer. Statist. Assoc.* **66** 879–883.

[18] DHARMADHIKARI, S. W., FABIAN, V. and JOGDEO, K. (1968). Bounds on the moments of martingales. *Ann. Math. Statist.* **39** 1719–1723.

[19] ERDELYI, A. (1956). *Asymptotic Expansions.* Dover, New York.

[20] ERDÖS, P. and RÉNYI, A. (1950). On the central limit theorem for samples from a finite population. *Publ. Math. Inst. Hungar. Acad. Sci.* **4** 49–61.

[21] FELLER, W. (1966). *An Introduction to Probability Theory and its Applications*, **2**. Wiley, New York.

[22] FELLINGHAM, S. A. and STOKER, D. J. (1964). An approximation for the exact distribution of Wilcoxon test for symmetry. *J. Amer. Statist. Assoc.* **59** 899–905.

[23] GNEDENKO, B. V., KOROLYUK, V. and SKOROKHOD, A. V. (1960). Asymptotic expansions in probabilty theory. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **2** 153–169.

[24] GOVINDARAJULU, Z. and HAYNAM, G. (1956) Exact power of Mann-Whitney test for exponential and rectangular alternatives. *Ann. Math. Statist.* **37** 945–953.

[25] GRAMS, W. F. and SERFLING, R. J. (1973). Convergence rates for *U*-statistics and related statistics. *Ann. Statist.* **1** 153–160.

[26] HÁJEK, J. and ŠIDÁK, A. (1967). *Theory of Rank Tests.* Academic Press, New York.

[27] HALDANE, J. B. S. and SMITH, S. M. (1956). Sampling distribution of a maximum likelihood estimate. *Biometrika* **43** 96–103.

[28] HODGES, J. L. and FIX, E. (1955). Significance probabilities of the Wilcoxon test. *Ann. Math. Statist.* **26** 301–312.

[29] HODGES, J. L. and LEHMANN, E. L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* **34** 598–611.

[30] HODGES, J. L. and LEHMANN, E. L. (1970). Deficiency. *Ann. Math. Statist.* **41** 783–801.

[31] HOEFFDING, W. (1948). A class of statistics with asymptotically normal distributions. *Ann. Math. Statist.* **19** 293–325.

[32] HUBER, P. J. (1964). Robust estimation of location. *Ann. Math. Statist.* **35** 73–101.

[33] HUBER, P. J. (1965). Behavior of maximum likelihood estimates under non-standard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 221–233.

[34] KLOTZ, J. (1963). Small sample power and efficiency of the one sample Wilcoxon and normal scores tests. *Ann. Math. Statist.* **34** 624–632.

[35] KLOTZ, J. (1964). On the normal scores two sample rank test. *J. Amer. Statist. Assoc.* **59** 652–654.

[36] LAUWERIER, H. A. (1963). The asymptotic expansions of the statistical distribution of N. V. Smirnov. *Z. Warscheinlichkeitstheorie und Verw. Gebiete* **2** 61–68.

[37] LINNIK, J. (1960). On the probability of large deviations for the sums of independent random variables. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **2** 289–306.

[38] LINNIK, J. and MITROFANOVA, N. (1965). Some asymptotic expansions for maximum likelihood estimates. *Sankhyā A* **27** 73–82.

[39] MAAG, U. and DICAIRE, G. (1971). On Kolmogorov-Smirnov type one sample statistics. *Biometrika* **58** 653–656.

[40] MICHEL, R. and PFANZAGL, J. (1971). Accuracy of normal approximations for minimum contrast estimates. *Z. Warscheinlichkeitstheorie und Verw. Gebiete* **18** 73–84.

[41] MILTON, R. C. (1970). *Rank Order Probabilities*. Wiley, New York.

[42] MITROFANOVA, N. (1967). Asymptotic expansions for maximum likelihood estimates of a vector parameter. *Theor. Probability Appl.* **12** 364–372.

[43] O'REILLY, N. and ROSENKRANTZ, W. (1972). Applications of the Skorokhod representation to rates of convergence for linear combinations of order statistics. *Ann. Math. Statist.* **43** 1204–1212.

[44] ORLOV, A. I. (1971). Estimates of speed of convergence to their limit distributions of certain statistics. *Theor. Probability Appl.* **16** 526.

[45] PFANZAGL, J. (1971) Berry-Esséen bound for minimum contrast estimates. *Metrika* **17** 82–91.

[46] PFANZAGL, J. (1973). Asymptotic expansions related to minimum contrast estimators. *Ann. Statist.* **1** 993–1026.

[47] RANGA RAO, R. (1961). On the central limit theorem in R^k. *Bull. Amer. Math. Soc.* **67** 359–361.

[48] ROGERS, W. F. (1971). Exact null distributions and asymptotic expansions for rank test statistics. Technical Report, Stanford Univ.

[49] ROSENKRANTZ, W. (1969). A rate of convergence for the von Mises functional. *Trans. Amer. Math. Soc.* **139** 329–337.

[50] SAWYER, S. (1972). Rates of convergence for some functionals in probability. *Ann. Math. Statist.* **43** 273–284.

[51] SAZONOV, V. (1969). Improvement of a convergence rate estimate. *Theor. Probability Appl.* **14** 649–651.

[52] STEIN, C. (1970). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **2**.

[53] 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** 599–606, 607–614.

[54] SUNDRUM, R. (1954). A further approximation to the distribution of Wilcoxon's statistic. *J. Roy. Statist. Soc. Ser. B* **16** 266–268.

[55] THOMPSON, R., GOVINDARAJULU, Z. and DOKSUM, K. (1967). Distribution and power of the absolute normal scores test. *J. Amer. Statist. Assoc.* **62** 966–975.

[56] VON MISES, R. (1947). Differentiable statistical functions. *Ann. Math. Statist.* **18** 309–348.

[57] WALLACE, D. (1958). Asymptotic approximations to distributions. *Ann. Math. Statist.* **29** 635–654.

[58] WITTING, H. (1960). A generalized Pitman efficiency. *Ann. Math. Statist.* **31** 405–414.

[59] YARNOLD, J. (1968). The accuracy of seven approximations for the null distribution of the chi-square goodness of fit statistic. Thesis, Univ. of Chicago.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720