The Annals of Statistics
1989, Vol. 17, No. 1, 170–183

A BERRY–ESSEEN BOUND FOR FUNCTIONS OF INDEPENDENT RANDOM VARIABLES[1]

BY KARL O. FRIEDRICH

Institut für Mathematische Stochastik der Albert-Ludwigs-Universität

The rate of convergence in the central limit theorem for functions of independent random variables is studied in a unifying approach. The basic result sharpens and extends a theorem of van Zwet. Applications to U-, L- and R-statistics are also given, improving or extending the results of Helmers and van Zwet, Helmers and Hušková, Does and van Es and Helmers.

1. Introduction. It is well known that many statistics occurring in estimation and testing problems behave asymptotically like sums of independent random variables. In particular, their distribution may often be approximated by a normal one. As this is of theoretical and practical importance, limit theorems and certain asymptotic properties of higher order have been the subject of many articles.

The next step after proving asymptotic normality is to establish bounds for the rate of convergence. Berry–Esseen bounds of order $O(n^{-1/2})$ have been obtained for several classes of U-, L- and R-statistics, but the methods of proof and the statements themselves were adjusted to the individual structure of these examples. Significant progress towards a general Berry–Esseen theory was achieved by van Zwet (1984), who studied the rate of convergence for symmetric functions of i.i.d. random variables.

The basic result of the present paper (Theorem 2.1) is another unifying Berry–Esseen type theorem, improving van Zwet's result in two directions: The restriction to symmetric statistics is dispensed with, and the moment condition itself is relaxed.

The first type of extension is not primarily of mathematical significance [see van Zwet (1984), where its possibility was indicated], but it is important from the point of view of application: The theorem applies to functions of independent but not necessarily identically distributed random variables under relatively simple moment conditions, and no additional assumption concerning the function or the sample model is required.

More interesting from a theoretical point of view is the second improvement. Though it applies to a broad range of very different statistics, the theorem nevertheless provides sharp estimates in various particular situations. It even improves or generalizes some results that were proved earlier for such special types of statistics. Applied to U-statistics, it leads to a moment of order exactly $\frac{5}{3}$ which has been conjectured in Korolyuk and Borovskikh (1985).

Received July 1986; revised May 1988.
[1]Work supported by the Deutsche Forschungsgemeinschaft, Grant Wi 49/18-1.
AMS 1980 *subject classification*. Primary 60F05.
Key words and phrases. Berry–Esseen bound, U-statistic, linear combination of order statistics, R-statistic.

As in most papers on Berry–Esseen bounds, the proof is based on Esseen's smoothing lemma. The difference with van Zwet's proof is the use of a reverse martingale structure instead of Hoeffding's expansion and the use of a new technique in the analysis of the characteristic function of the statistic.

To be more concrete, let us consider a U-statistic of order 2, defined as

$$U_n = \sum\sum_{1 \le i < j \le n} h(X_i, X_j), \qquad n \in \mathbb{N}, \tag{1.1}$$

where $h: \mathbb{R}^2 \to \mathbb{R}$ is a symmetric function and $X_1, X_2, \dots$ are i.i.d. random variables. Assume that

$$Eh(X_1, X_2) = 0, \qquad EH(X_1, X_2)^2 < \infty, \tag{1.2}$$

and define

$$\hat{U}_n = (n-1) \sum_{i=1}^{n} g(X_i), \qquad g(x) = E(h(X_1, X_2) | X_1 = x). \tag{1.3}$$

A crucial point in proving a Berry–Esseen theorem for U_n is handling the difference of the statistic and its projection, i.e.,

$$\Delta = U_n - \hat{U}_n = \sum\sum_{1 \le i < j \le n} \big(h(X_i, X_j) - g(X_i) - g(X_j)\big). \tag{1.4}$$

The same problem arises for general functions of independent random variables, but in this case Δ may not have the simple sum decomposition of (1.4) anymore.

Defining

$$\Delta_i = E(\Delta | X_i, \dots, X_n) - E(\Delta | X_{i+1}, \dots, X_n), \qquad i = 1, \dots, n,$$

$$\Delta_{i,j} = E(\Delta_i | X_i, X_j, \dots, X_n) - E(\Delta_i | X_i, X_{j+1}, \dots, X_n), \qquad 1 \le i < j \le n,$$

the decomposition (1.4) can be generalized to

$$\Delta = \sum\sum_{1 \le i < j \le n} \Delta_{i,j}, \tag{1.5}$$

which reduces to (1.4) since in that case

$$\Delta_{i,j} = h(X_i, X_j) - g(X_i) - g(X_j) \quad \text{a.s., } 1 \le i < j \le n.$$

The conditional expectations appearing so far are defined for a general statistic as well if its first moment exists. From the definitions it follows directly that Δ_i, $i = 1, \dots, n$, and $\Delta_{i,j}$, $j = i+1, \dots, n$ for fixed i with $1 \le i \le n$, are reverse martingale differences.

The difference from van Zwet's approach is the following: Using the reverse martingale structure one has to deal with far fewer terms than with Hoeffding's expansion, which implicitly contains it. Later on it will be seen that van Zwet's key assumption may be interpreted as a moment bound for the reverse martingale differences. Therefore their use clarifies the proof and the result, whereas Hoeffding's decomposition appears to be a roundabout way. Indeed, it seems possible to simplify van Zwet's proof and simultaneously sharpen his theorem to comparable strength as in Helmers and van Zwet (1982), adapting the techniques used there.

In the present paper we use a different expansion for the characteristic function leading to a somewhat stronger result. The idea is to continue the work which Chan and Wierman (1977) began and which has led to new and better Berry–Esseen type results for U-statistics. These improvements are based on the approximation of the characteristic function $\varphi_0 \equiv E\exp(itU_n)$ in the critical region $t \notin I_n = [-n^{\varepsilon}, n^{\varepsilon}]$, $0 < \varepsilon < \frac{1}{2}$, by $\varphi_m \equiv E\exp(itS_m)$, where S_m denotes the projection defined by

$$S_m \equiv (n-1)\sum_{j=1}^{m} g(X_j) + E(U_n|X_{m+1},\dots,X_n)$$

for suitable $0 < m < n$. Helmers and van Zwet succeeded in proving the Berry–Esseen result by choosing $m < n$ depending on $t \notin I_n$ and for all $t \in I_n$ using the same (linear) projection $S_n = \hat{U}_n$. Here we shall use a more refined procedure.

Splitting $\varphi_0 - \varphi_n$ into $\sum_j(\varphi_{j-1} - \varphi_j)$ and writing

$$\varphi_{j-1} - \varphi_j = E\exp(itS_j)\big(\exp(it\Delta_j) - 1\big)$$

we can expand the second factor in the stochastically small quantities $\Delta_j = S_{j-1} - S_j$. Another decomposition of S_j into the reverse martingale differences $\Delta_{i,k}$ for the linear expansion terms yields—via a telescoping argument as above —a sum of products whose factors are either sufficiently small or independent such that the resulting estimate can be integrated over the desired range $[-Cn^{1/2}, Cn^{1/2}]$.

Thus many small steps seem to be better than a few large steps in estimating $\varphi_0 - \varphi_m$.

The application of Theorem 2.1 to specific statistics is mainly a technical problem: One has to determine some conditional expectations which define the random variables appearing in the bound and then to estimate their moments. This can be done without difficulties in the case of U-statistics of a higher order, and the result sharpens that obtained by Helmers and van Zwet (1982). Theorem 2.1 can also be used to slightly improve a recent result of van Es and Helmers (1986) for elementary symmetric polynomials.

The final conclusion concerning L- and R-statistics (Theorem 4.1) requires some additional checks and computations, which are omitted here. For a detailed proof, see Friedrich (1985). Only the representation of the random variables appearing in the bound is given, as this result may be used to derive other versions of the theorem, replacing smoothness of the weights by stronger conditions on the distributions.

In the case of L-statistics, Theorem 4.1 weakens the assumptions of Helmers and Hušková (1984). For R-statistics, it may be looked at as an extension of Does (1982) to the non-i.i.d. case under a somewhat stronger condition. Of course, for the special case of the null hypothesis a better result can be given. See Bolthausen (1984) for such a result based on the method of Charles Stein.

2. Main result. Suppose we are given independent and real random variables $X_1, \dots, X_n$ and a measurable function $\tau: \mathbb{R}^n \to \mathbb{R}$, $n \in \mathbb{N}$, and put

$$T = \tau(X_1, \dots, X_n). \tag{2.1}$$

This is the general form of a statistic with which we shall be concerned. At the moment we only assume that $ET = 0$ for convenience, but later we will need $E|T|^p < \infty$ for some $p \geq \frac{3}{2}$. Define

$$\hat{T} = \sum_{j=1}^{n} T_j, \qquad T_j = E\left[T|X_j\right], \qquad j = 1, \dots, n, \tag{2.2}$$

$$\Delta = T - \hat{T} \tag{2.3}$$

and

$$S_m = \sum_{j=1}^{m} T_j + E\left[T|X_{m+1}, \dots, X_n\right] \quad \text{for } 0 < m < n. \tag{2.4}$$

With $S_0 = T$, $S_n = \hat{T}$, it then follows that

$$\Delta_m = S_{m-1} - S_m \quad \text{a.s., } m = 1, \dots, n. \tag{2.5}$$

The main drawback, from which the variables $\Delta_1, \dots, \Delta_n$ and $\Delta_{j,k}$, $1 \leq j < k \leq n$ suffer, is that they depend on the special form of T and the order of $X_1, \dots, X_n$. To avoid this disadvantage, we define

$$\begin{aligned} \check{X}_j &= \left(X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_n\right), \\ \hat{T}_j &= T_j + E\left[T|\check{X}_j\right], \qquad D_j[T] = T - \hat{T}_j \end{aligned} \tag{2.6}$$

for $j = 1, \dots, n$, and furthermore for $1 \leq j < k \leq n$,

$$\begin{aligned} \check{X}_{jk} &= \left(X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_{k-1}, X_{k+1}, \dots, X_n\right), \\ \hat{T}_{jk} &= E\left[T|\check{X}_j\right] + E\left[T|\check{X}_k\right] - E\left[T|\check{X}_{jk}\right], \\ D_{jk}[T] &= T - \hat{T}_{jk}. \end{aligned} \tag{2.7}$$

The relation of these quantities to the Δ's is given by the equalities

$$\begin{aligned} \Delta_j &= E\left[D_j[T]|X_j, \dots, X_n\right] \quad \text{a.s., for } j = 1, \dots, n, \\ \Delta_{jk} &= E\left[D_{jk}[T]|X_j, X_k, \dots, X_n\right] \quad \text{a.s., } 1 \leq j < k \leq n, \end{aligned} \tag{2.8}$$

which follow directly from the definitions.

Though it will not be used in the present paper, it is interesting to note that $\hat{T}_1, \dots, \hat{T}_n$ and $\hat{T}_{jk}$, $j \neq k$, can be defined as projections, if T has a finite variance. Because of $ET = 0$, this can be drawn from Hájek's projection lemma in the first case. In the second case and with a slight generalization, we find $\hat{T}_{jk}$ to be (in L_2) the best approximation of T by a sum of functions of the dependent random variables $\check{X}_j$ and $\check{X}_k$, $j \neq k$. Some projection results that go further were stated by Rüschendorf (1985).

Our main result now reads as follows:

THEOREM 2.1. *Suppose* $ET = 0$, $0 < \sigma^2 = \mathrm{Var}(\hat{T}) < \infty$, *and define*

$$\gamma_0 = \sigma^{-3} \sum_{j=1}^{n} E|T_j|^3,$$

$$\gamma_1 = \sigma^{-1} \max_{1 \le j \le n} \left(E|T_j|^3\right)^{1/3},$$

$$\gamma_{2,p} = \sigma^{-p} \max_{1 \le j \le n} E|D_j[T]|^p \quad \text{for } p \ge 1,$$

$$\gamma_{3,p} = \sigma^{-1} \max_{1 \le j < k \le n} \left(E|D_{jk}[T]|^p\right)^{1/p} \quad \text{for } p \ge 1.$$

Then:

(a) *If* $\frac{3}{2} \le p < 2$, *there exists a constant* $C \in \mathbb{R}$, *such that*

$$\sup_{x \in \mathbb{R}} |P(\sigma^{-1}T \le x) - \Phi(x)| \le C\left(\gamma_0 + \frac{1}{2-p} n^{3/2}\gamma_1\gamma_{2,p} + n^2\gamma_1^2\gamma_{3,3/2}\right),$$

where Φ *denotes the distribution function of the standard normal distribution.*

(*b*) *The estimate remains true for* $p = 2$, *if* $1/(2-p)$ *is replaced by* $\log n$.

(*c*) *In both* (*a*) *and* (*b*), $\gamma_{2,p}$ *may be replaced by* $n\gamma_{3,p}^p$.

REMARK 1. The term γ_0 results from $\hat{T}$, and it is the usual Berry–Esseen bound for sums of independent random variables. To achieve the order of convergence $O(n^{-1/2})$ for the general statistic T, one has to show that not only γ_0, but also γ_1 is of this order, which is a somewhat stronger requirement.

REMARK 2. Moreover, one has to establish a set of inequalities, such as

$$\gamma_{3,3/2} \le Cn^{-3/2} \tag{2.9}$$

and

$$\gamma_{2,p} \le Cn^{-3/2} \quad \text{or} \quad \gamma_{3,p}^p \le Cn^{-5/2} \tag{2.10}$$

for some $p \in [\frac{3}{2}, 2)$. Alternatively, one can combine (2.9) with

$$\gamma_{2,2} \le C(\log n)^{-1}n^{-3/2} \quad \text{or} \quad \gamma_{3,2}^2 \le C(\log n)^{-1}n^{-5/2} \tag{2.11}$$

instead of (2.10).

REMARK 3. Because of the moment inequality $\gamma_{3,p} \ge \gamma_{3,q}$ for $p \ge q \ge 1$, (2.9) and the second relation in (2.10) are ensured by

$$\gamma_{3,5/3} \le Cn^{-3/2}. \tag{2.12}$$

This form is adequate for U-statistics, but it may be harder to verify in general. In the case of L- and R-statistics for instance, we will use (2.9) and (2.11).

REMARK 4. (2.9)–(2.12) are the counterparts of van Zwet's condition

$$1 + E(E(T|X_1,\dots, X_{n-2}))^2 - 2E(E(T|X_1,\dots, X_{n-1}))^2 \le Bn^{-3}$$

for a $B \in \mathbb{R}$. Note that under van Zwet's general assumption $ET^2 = 1$, $ED_{n-1,n}[T]^2$ coincides with the left-hand side.

REMARK 5. In the case of a U-type-statistic

$$U = \sum\sum_{1 \le i_1 < \cdots < i_r \le n} Y_{i_1,\dots,i_r} \tag{2.13}$$

with $Y_{i_1,\dots,i_r} = h_{i_1,\dots,i_r}(X_{i_1},\dots, X_{i_r})$ and measurable functions $h_{i_1,\dots,i_r}\colon \mathbb{R}^r \to \mathbb{R}$, $1 \le i_1 < \cdots < i_r \le n$, it follows that

$$D_{jk}[U] = \sum\sum_{\substack{1 \le i_1 < \cdots < i_r \le n \\ \{i_1,\dots,i_r\} \supseteq \{j,k\}}} D_{jk}\left[Y_{i_1,\dots,i_r}\right] \quad \text{a.s., } 1 \le j < k \le n. \tag{2.14}$$

Employing Minkowski's inequality, (2.12) is now guaranteed by

$$\max_{1 \le j_1 < \cdots < j_r \le n} \sigma_n^{-1}\left(E|Y_{j_1,\dots,j_r}|^{5/3}\right)^{3/5} \le Bn^{1/2-r}, \qquad B \in \mathbb{R}, \tag{2.15}$$

where σ_n^2 denotes the variance of the projection of U. This also proves the result for U-statistics mentioned in the Introduction.

REMARK 6. For sequences of U-statistics where the kernel length tends to infinity, a detailed analysis of the random variables in (2.14) may provide a better result. As an example, we mention elementary symmetric polynomials

$$S_n^{(k)} = \binom{n}{k}^{-1} \sum\sum_{1 \le i_1 < \cdots < i_k \le n} X_{i_1} \cdots X_{i_k}, \tag{2.16}$$

where $X_1,\dots, X_n$ are i.i.d. with $\mu = EX_1 \ne 0$, $0 < \sigma^2 = \mathrm{Var}(X_1)$, $E|X_1|^3 < \infty$. Here we get

$$D_{n-1,n}[S_n^{(k)}] = \binom{n}{k}^{-1}\binom{n-2}{k-2}(X_{n-1} - \mu)(X_n - \mu)S_{n-2}^{(k-2)} \quad \text{a.s.} \tag{2.17}$$

The application of Theorem 2.1 leads to the rate of convergence $O(k/n^{1/2})$ provided $k = O(n^{1/2}/\log n)$, therefore improving the result of van Es and Helmers (1986). These authors also demonstrated that this bound is sharp and that $k = o(n^{1/2})$ is needed to prove asymptotic normality.

3. Proof of Theorem 2.1. For abbreviation, we first write

$$Y_j = (X_j,\dots, X_n), \qquad j = 1,\dots, n,$$

$$W_{k,j} = E(\Delta_k|X_k, X_{j+1},\dots, X_n), \qquad 1 \le k \le j \le n.$$

Note that we have $S_n = S_{n-1} = \hat{T}$, and therefore $\Delta_n = 0$. Also put $S_r = \hat{T}$, $\Delta_r = 0$, $W_{k,r} = 0$ and $\Delta_{k,r} = 0$ for $k = 1, \dots, n-1$, $r > n$.

The reverse martingale structure of the terms in (1.5) will be used throughout the proof. On the one hand, it will be exploited in some conditioning arguments, on the other hand, it leads to useful moment-bounds.

Applying Lemma 1 of Chatterji (1969), we find for each $p \in [1,2]$,

$$\begin{aligned} &E|S_m - S_l|^p \le 2^{2-p} \sum_{j=m+1}^{l} E|\Delta_j|^p, \qquad 0 \le m < l \le n, \\ (3.1)\quad &E|W_{k,l} - W_{k,m}|^p \le 2^{2-p} \sum_{j=l+1}^{m} |\Delta_{k,j}|^p, \qquad 1 \le k < l < m \le n. \end{aligned}$$

In the cases $m = 0$, $l = n$ or $l = k$, $m = n$, these relations are inequalities for the absolute moments of Δ, respectively, Δ_k. Employing Jensen's inequality, it follows from (2.8) that it suffices to prove (a) and (b).

As the bound remains the same for all statistics $T_\alpha = \alpha T$, $\alpha \in \mathbb{R}$, we have to prove the theorem only for $\sigma^2 = 1$. Applying Esseen's smoothing lemma, we obtain

$$\begin{aligned} \delta &= \sup_{x \in \mathbb{R}} |P(T \le x) - \Phi(x)| \\ &\le \frac{1}{\pi} \int_{-M}^{M} |t|^{-1} |Ee^{it\hat{T}} - e^{-t^2/2}|\, dt + \frac{24}{\pi\sqrt{2\pi}} M^{-1} \qquad (3.2) \\ &\quad + \frac{1}{\pi} \int_{-M}^{M} |t|^{-1} |Ee^{itT} - Ee^{it\hat{T}}|\, dt \\ &= \delta_1 + \delta_2 + \delta_3, \end{aligned}$$

where M may be chosen arbitrarily. To estimate the first two terms, we proceed as in the proof of the Berry–Esseen theorem for sums of independent random variables [cf. Feller (1971), page 544]. Let us agree to prove the theorem for a constant $C \ge 6$. Then we may assume that $\gamma_0 < \frac{1}{6}$, because δ cannot exceed 1. With $M = 8/(9\gamma_0)$ Feller arrived at

$$\delta_1 + \delta_2 \le 6\gamma_0. \qquad (3.3)$$

Moreover, we may adapt some of his intermediate results, namely,

$$|Ee^{itT_j}| \le e^{-\kappa_j t^2} \quad \text{for } |t| \le M,\ j = 1, \dots, n,$$

with $\kappa_j = \frac{1}{2}\operatorname{Var}(T_j) - \frac{3}{8}E|T_j|^3 M$, and

$$\prod_{\substack{j=1 \\ j \ne k}}^{n} |Ee^{itT_j}| \le \exp\left(-t^2 \sum_{\substack{j=1 \\ j \ne k}}^{n} \kappa_j\right) \le e^{-t^2/8} \quad \text{for } |t| \le M, 1 \le k \le n. \qquad (3.4)$$

It does not affect the generality of our theorem to assume $\kappa_1 \geq \cdots \geq \kappa_n$. Then from (3.4) we conclude that

$$\prod_{\substack{j=1\\ j\neq k}}^{r} |Ee^{itT_j}| \leq \exp\left(-\frac{1}{8}\frac{r-1}{n-1}t^2\right) \leq C_0 \exp\left(-\frac{r}{8n}t^2\right) \tag{3.5}$$

for $t \in [-M, M]$, $1 \leq k \leq r \leq n$, with $C_0 = e^{1/10}$.

After these preparations, we turn to δ_3. In the first step we show how the decomposition of Δ may be used to achieve an estimate for the difference of the characteristic functions of T and $\hat{T}$.

Let us define $m(k) = \max\{r \in \mathbb{N} | rk < n\}$, $S_r = \hat{T}$, $\Delta_r = 0$, $W_{k,r} = 0$, $\Delta_{k,r} = 0$ for $k = 1, \dots, n-1$, $r > n$, and

$$Z_{1k}(t) = Ee^{itS_k}\left(e^{it\Delta_k} - 1 - it\Delta_k\right),$$

$$Z_{2,k,j}(t) = E\left(e^{itS_{jk}} - e^{itS_{(j+1)k}}\right)W_{k,jk},$$

$$Z_{3,k,j,l}(t) = Ee^{itS_{(j+1)k}}\Delta_{k,l}$$

for $1 \leq k \leq n-1$, $1 \leq j \leq m(k)$, $l = jk+1, \dots, (j+1)k$. Simple computations show

$$Ee^{itT} - Ee^{it\hat{T}} = \sum_{k=1}^{n-1} Z_{1k}(t) + it\sum_{k=1}^{n-1} Ee^{itS_k}\Delta_k \tag{3.6}$$

and

$$\begin{aligned} Ee^{itS_k}\Delta_k &= Ee^{itS_k}W_{k,k} - Ee^{itS_{(m(k)+1)k}}W_{k,(m(k)+1)k} \\ &= \sum_{j=1}^{m(k)} \left(Ee^{itS_{jk}}W_{k,jk} - Ee^{itS_{(j+1)k}}W_{k,(j+1)k}\right) \\ &= \sum_{j=1}^{m(k)} E\left(e^{itS_{jk}} - e^{itS_{(j+1)k}}\right)W_{k,jk} \\ &\quad + \sum_{j=1}^{m(k)} Ee^{itS_{(j+1)k}}\left(W_{k,jk} - W_{k,(j+1)k}\right) \\ &= \sum_{j=1}^{m(k)} Z_{2,k,j}(t) + \sum_{j=1}^{m(k)}\sum_{l=jk+1}^{(j+1)k} Z_{3,k,j,l}(t). \end{aligned} \tag{3.7}$$

Our next task is to estimate the various terms just obtained. To begin with Z_{1k}, we use the independence of $X_1, \dots, X_n$, (3.5) for $r = k$ and $|e^{ix} - 1 - ix| \leq 2|x|^p$

for $p \in [1,2]$, $x \in \mathbb{R}$, and find

$$\begin{aligned} |Z_{1k}(t)| &= \left|\prod_{j=1}^{k-1} Ee^{itT_j}\right| |Ee^{itE[S_k|Y_k]}(e^{it\Delta_k} - 1 - it\Delta_k)| \\ &\le 2C_0 E|\Delta_k|^p |t|^p e^{-(k/8n)t^2} \\ &\le 2C_0 \gamma_{2,p} |t|^p e^{-(k/8n)t^2} \quad \text{for } |t| \le M,\ k = 1,\dots,n-1, \end{aligned} \tag{3.8}$$

with $Y_k = (X_k, \dots, X_n)$.

As for $Z_{2,k,j}$, $1 \le k \le n-1$, $1 \le j \le m(k)$, we have the identities

$$\begin{aligned} Z_{2,k,j}(t) &= Ee^{itS_{jk}}(1 - e^{it(S_{(j+1)k} - S_{jk})}) W_{k,jk} \\ &= \left\{\prod_{\substack{l=1 \\ l \ne k}}^{jk} Ee^{itT_l}\right\} Ee^{it(T_k + E(T|Y_{jk+1}))}(1 - e^{it(S_{(j+1)k} - S_{jk})}) W_{k,jk} \\ &= \left\{\prod_{\substack{l=1 \\ l \ne k}}^{jk} Ee^{itT_l}\right\} E(e^{itT_k} - 1) e^{itE(T|Y_{jk+1})}(1 - e^{it(S_{(j+1)k} - S_{jk})}) W_{k,jk}. \end{aligned}$$

Here the first step is obvious from the definition, the second follows, because $W_{k,jk}$ depends only on $(X_k, X_{jk+1}, \dots, X_n)$ and $S_{(j+1)k} - S_{jk}$ is independent of $X_1, \dots, X_{jk}$, whereas the third equality follows from $E[W_{k,jk}|Y_{jk+1}] = 0$ a.s. Using (3.5) and Hölder's inequality we now get

$$\begin{aligned} |Z_{2,k,j}(t)| &\le C_0 e^{-(jk/8n)t^2} E|e^{itT_k} - 1|\,|1 - e^{it(S_{(j+1)k} - S_{jk})}|\,|W_{k,jk}| \\ &\le C_0 e^{-(jk/8n)t^2} \left(E|e^{itT_k} - 1|^q |1 - e^{it(S_{(j+1)k} - S_{jk})}|^q\right)^{1/q} \left(E|W_{k,jk}|^p\right)^{1/p} \end{aligned} \tag{3.9}$$

with $q = p/(p-1)$. Exploiting the independence of the sample once more and then using $|e^{ix} - 1| \le |x|$ and $|e^{ix} - 1| \le 2|x|^{p-1}$ for $x \in \mathbb{R}$, $p \in [1,2]$, we obtain

$$\begin{aligned} &\left(E|e^{itT_k} - 1|^q |1 - e^{it(S_{(j+1)k} - S_{jk})}|^q\right)^{1/q} \\ &\quad = \left(E|e^{itT_k} - 1|^q\right)^{1/q} \left(E|1 - e^{it(S_{(j+1)k} - S_{jk})}|^q\right)^{1/q} \\ &\quad \le 2|t|^p \left(E|T_k|^q\right)^{1/q} \left(E|S_{jk} - S_{(j+1)k}|^p\right)^{1-1/p}. \end{aligned} \tag{3.10}$$

Because of $q \le 3$ for $p \ge 3/2$, the moment inequality entails

$$\left(E|T_k|^q\right)^{1/q} \le \left(E|T_k|^3\right)^{1/3} \le \gamma_1. \tag{3.11}$$

From (1.5), (2.8) and Jensen's inequality we conclude that

$$\left(E|W_{k,jk}|^p\right)^{1/p} \le \left(E|\Delta_k|^p\right)^{1/p} \le \left(E|D_k[T]|^p\right)^{1/p} \le \gamma_{2,p}^{1/p} \tag{3.12}$$

and (3.9)–(3.12) together with (2.11) yield

$$|Z_{2,k,j}(t)| \le 4C_0 \gamma_1 \gamma_{2,p} k^{1-1/p} |t|^p e^{-(jk/8n)t^2} \quad \text{for } |t| \le M. \tag{3.13}$$

To estimate the second sum on the right-hand side of (3.7), note that we have $E(\Delta_{k,l}|X_k, Y_{l+1}) = 0$ a.s. and $E(\Delta_{k,l}|Y_l) = 0$ a.s. for $1 \le k < l \le n$. Proceeding as above, we thus get

$$|Ee^{itS_r}\Delta_{k,l}| = \left|\left\{\prod_{\substack{j=1\\ j\neq k}}^{l-1} Ee^{itT_j}\right\} Ee^{it(T_k+T_l+E[S_r|Y_{l+1}])}\Delta_{k,l}\right|$$

$$(3.14)\qquad = \left|\left\{\prod_{\substack{j=1\\ j\neq k}}^{l-1} Ee^{itT_j}\right\} E(e^{itT_k}-1)(e^{itT_l}-1)e^{itE[S_r|Y_{l+1}]}\Delta_{k,l}\right|$$

$$\le C_0 e^{-((l-1)/8n)t^2}\big(E|e^{itT_k}-1|^3|e^{itT_l}-1|^3\big)^{1/3}\big(E|\Delta_{k,l}|^{3/2}\big)^{2/3}$$

$$\le C_0\big(E|T_k|^3\big)^{1/3}\big(E|T_l|^3\big)^{1/3}\big(E|\Delta_{k,l}|^{3/2}\big)^{2/3}t^2e^{-((l-1)/8n)t^2}$$

$$\le C_0\gamma_1^2\gamma_{3,3/2}t^2e^{-((l-1)/8n)t^2} \quad \text{for } |t|\le M,\ 1\le k<l\le n,\ r\ge l.$$

Summarizing (3.6)–(3.8), (3.13) and (3.14) and inserting

$$\int_0^\infty x^t e^{-\lambda x^2}\,dx = \frac{1}{2}\lambda^{-(t+1)/2}\Gamma\left(\frac{t+1}{2}\right), \qquad t>-1,\ \lambda>0,$$

we arrive at

$$\delta_3 \le \sum_{k=1}^{n-1}\int_{-M}^{M}|t|^{-1}|Z_{1,k}(t)|\,dt + \sum_{k=1}^{n-1}\sum_{j=1}^{m(k)}\int_{-M}^{M}|Z_{2,k,j}(t)|\,dt$$

$$+\sum_{k=1}^{n-1}\sum_{j=1}^{m(k)}\sum_{l=jk+1}^{(j+1)k}\int_{-M}^{M}|Z_{3,k,j,l}(t)|\,dt$$

$$\le 4C_0\gamma_{2,p}\sum_{k=1}^{n-1}\int_0^\infty t^{p-1}e^{-(k/8n)t^2}\,dt$$

$$+4C_0\gamma_1\gamma_{2,p}\sum_{k=1}^{n-1}\sum_{j=1}^{m(k)}k^{1-1/p}\int_{-\infty}^{\infty}|t|^pe^{-(jk/8n)t^2}\,dt$$

$$+C_0\gamma_1^2\gamma_{3,3/2}\sum_{k=1}^{n-1}\sum_{j=1}^{m(k)}\sum_{l=jk+1}^{\min(n,(j+1)k)}\int_{-\infty}^{\infty}t^2e^{-((l-1)/8n)t^2}\,dt$$

$$\le 2C_0\Gamma\left(\frac{p}{2}\right)8^{p/2}n^{p/2}\gamma_{2,p}\sum_{k=1}^{n-1}k^{-p/2}$$

$$+4C_0 8^{(p+1)/2}\Gamma\left(\frac{p+1}{2}\right)\gamma_1\gamma_{2,p}n^{(p+1)/2}\sum_{k=1}^{n-1}k^{1-1/p-(p+1)/2}\sum_{j=1}^{m(k)}j^{-(p+1)/2}$$

$$+C_0\Gamma\left(\frac{3}{2}\right)8^{3/2}n^{3/2}\gamma_1^2\gamma_{3,3/2}\sum_{k=1}^{n-1}\sum_{l=k+1}^{n}(l-1)^{-3/2}.$$

Note that

$$1-\frac{1}{p}-\frac{p+1}{2} = \frac{1}{2p}(p-1)(2-p)-1 > \frac{(2-p)}{8}-1 \quad \text{for } p\in[3/2,2).$$

The assertion follows now from (3.2) and (3.3) after applying well-known estimates for geometric series.

4. Application to *L*- and *R*-statistics. Let $X_1, \ldots, X_n$ be independent and real random variables and $h, g_1, \ldots, g_n$: $\mathbb{R} \to \mathbb{R}$ be measurable functions, and put

$$V_j = h(X_j), \qquad Y_j = g_j(X_j), \qquad j = 1, \ldots, n.$$

The distribution functions $F_1, \ldots, F_n$ of $V_1, \ldots, V_n$ are assumed to be continuous. Suppose $a(1), \ldots, a(n) \in \mathbb{R}$ are constants, put $u(x) = 1$ for $x \geq 0$ and $u(x) = 0$ for $x < 0$ and define

$$\begin{aligned} r_j &= \sum_{k=1}^{n} u(V_j - V_k), \qquad j = 1, \ldots, n, \\ T &= \sum_{j=1}^{n} a(r_j) Y_j. \end{aligned} \tag{4.1}$$

The statistics which are our objective may be regarded as special examples of this general type of random variables:

(a) If h is the identity and $g_1, \ldots, g_n$ are constants, then T is a simple linear rank statistic.

(b) In the special case $h(x) = |x|$, $g_1(x) = \cdots = g_n(x) = \operatorname{sgn}(x)$ for $x \in \mathbb{R}$, T defines a signed rank statistic.

(c) For a given L-statistic, a statistic of the above type can be defined, which is distributed according to the same law. [Hint: Use i.i.d. random variables from the uniform distribution in (0, 1), if the continuity assumption is not fulfilled.]

The conditions for the scores $a(1), \ldots, a(n)$ are written using their first and second differences:

$$\begin{aligned} b(k) &= a(k+1) - a(k), \qquad k = 1, \ldots, n-1, \\ d(k) &= a(k) - 2a(k+1) + a(k+2), \qquad k = 1, \ldots, n-2. \end{aligned}$$

As we are mainly interested in the case of unbounded scores, we impose conditions that look similar to those used by Chernoff and Savage (1958). If $\delta > 0$ is given, we say that (C1) is fulfilled, if there exists a constant $A \in \mathbb{R}$, such that

(C1)(a) $$|a(k)| \leq A\left[\frac{k}{n}\left(1 - \frac{k-1}{n}\right)\right]^{-\delta}, \qquad k = 1, \ldots, n,$$

(b) $$|b(k)| \leq An^{-1}\left[\frac{k}{n}\left(1 - \frac{k}{n}\right)\right]^{-(1+\delta)}, \qquad k = 1, \ldots, n-1,$$

(c) $$|d(k)| \leq An^{-2}\left[\frac{k}{n}\left(1 - \frac{k+1}{n}\right)\right]^{-(2+\delta)}, \qquad k = 1, \ldots, n-2.$$

Note that δ is assumed to be positive only in order to exclude some uninteresting cases. In the theorem to follow it has to be bounded from above, which is more restrictive.

As for $Y_1, \dots, Y_n$ we need some moment conditions. Let

$$\|Y\|_p = (E|Y|^p)^{1/p} \quad \text{for } p \in [1, \infty)$$

and

$$\|Y\|_\infty = \sup\{c \in \mathbb{R} \mid P(|Y| > c) < 1\}$$

denote the p-norms of a random variable Y, as usual. We will need an estimate

$$\text{(C2)} \qquad \mu_p = \max_{1 \le j \le n} \|Y_j\|_p \le B n^{-1/2}$$

with a constant $B \in \mathbb{R}$, where p will be coupled with a δ in a suitable way.

In the third condition, the distribution functions $F_1, \dots, F_n$ are related to their mean

$$G(x) = n^{-1} \sum_{j=1}^{n} F_j(x), \qquad x \in \mathbb{R}.$$

(C3) There exists a constant $C \in \mathbb{R}$, such that

$$F_j(x) \le CG(x), \qquad 1 - F_j(x) \le C(1 - G(x)) \quad \text{for all } x \in \mathbb{R},\ j = 1, \dots, n.$$

In multisample problems, (C3) may easily be checked even under fixed alternatives, i.e., nonidentical distributions. For if m denotes the number of observations in the smallest subpopulation and if we have $m/n \ge \lambda > 0$, we find that (C3) is fulfilled with $C = \lambda^{-1}$.

As before, we use the variance of the projection $\hat{T}$ as a norming factor for T. Then our result reads as follows.

THEOREM 4.1. *Let* (C1)–(C3) *be fulfilled with* $\delta > 0$, $p \in (4, \infty]$ *such that* $\delta + 1/p < 1/4$. *Then there exists a constant* $K \in \mathbb{R}$, *such that* $\sigma^2 = \operatorname{Var}(\hat{T}) > 0$ *implies*

$$\sup_{x \in \mathbb{R}} |P(\sigma^{-1}(T - ET) \le x) - \Phi(x)| \le K n^{-1/2}.$$

REMARK 1. The constant K does not depend on the sample-size n, but it is a function of A, B, C and the parameters δ and p. From the proof it becomes clear that the last dependence only comes about via the term $\frac{1}{4} - \delta - 1/p$, which has to be bounded away from 0.

REMARK 2. The variance condition is adjusted for series $(T_n)_{n \in \mathbb{N}}$, which possess nondegenerate limit distributions.

REMARK 3. In the case of a simple linear rank statistic, Theorem 4.1 may be applied with $p = \infty$ and $0 < \delta < \frac{1}{4}$. Condition (C2) says, that the maximum of the regression constants has to behave like $O(n^{-1/2})$ to achieve this rate of convergence. This is somewhat stronger than Does' (1982) condition.

REMARK 4. The result may directly be compared with that of Helmers and Hušková (1984) on L-statistics. These authors also need a technical condition for the pseudoinverse of the common distribution function of the sample.

In order to apply Theorem 2.1, one has first to determine the terms used there. For the statistic (4.1), this is easily achieved using relative ranks of $V_1, \dots, V_n$ in suitable subsamples. Define

$$r_{jk} = \sum_{\substack{s=1\\ s\neq k}}^{n} u(V_j - V_s), \qquad 1 \le j, k \le n,\ j \neq k,$$

$$r_{jkl} = \sum_{\substack{s=1\\ s\neq k,\, l}}^{n} u(V_j - V_s), \qquad 1 \le j, k, l \le n,\ j \neq k, l,\ k < l,$$

$$q_k(x, y) = u(x - y) - F_k(x), \qquad x \in \mathbb{R},\ k = 1, \dots, n.$$

Then one gets the following representations, which also hold in situations other than those of Theorem 4.1.

LEMMA. *Let T be given as in* (4.1) *and suppose $E|T| < \infty$. Then*

(a) $$E[T|X_k] - ET = E[a(r_k)Y_k|X_k] - Ea(r_k)Y_k + \sum_{\substack{j=1\\ j\neq k}}^{n} E\left[b(r_{jk})Y_j q_k(V_j, V_k)|X_k\right] \quad a.s., 1 \le k \le n,$$

(b) $$D_{jk}[T] = b(r_{jk})Y_j q_k(V_j, V_k) - E\left[b(r_{jk})Y_j q_k(V_j, V_k)|\check{X}_j\right] + b(r_{kj})Y_k q_j(V_k, V_j) - E\left[b(r_{kj})Y_k q_j(V_k, V_j)|\check{X}_k\right] + \sum_{\substack{l=1\\ l\neq k,\, j}}^{n} d(r_{lkj})Y_l q_j(V_l, V_j) q_k(V_l, V_k) \quad a.s., 1 \le k < j \le n.$$

The proofs of (a) and (b) are similar. Distinguishing the cases $V_l < \min(V_j, V_k)$, $\min(V_j, V_k) \le V_l < \max(V_j, V_k)$ and $\max(V_j, V_k) \le V_l$ for $1 \le j < k \le n$, $l \neq j, k$, one gets

$$a(r_l) = d(r_{ljk})u(V_l - V_k)u(V_l - V_j) + a(r_{ljk} + 1)\big(u(V_l - V_k) + u(V_l - V_j)\big) + a(r_{ljk})\big(1 - u(V_l - V_k) - u(V_l - V_j)\big).$$

Similarly, one finds $a(r_j) = a(r_{jk}) + b(r_{jk})u(V_j - V_k)$ for $1 \le j, k \le n$, $j \neq k$. The assertions then follow evaluating the expressions on the left-hand sides.

Note that D_{jk} is a linear operator and that $D_{jk}[Z] = 0$, $j \neq k$ for any random variable Z that is independent of X_j or X_k.

To complete the proof of Theorem 4.1 as an application of Theorem 2.1, one has next to establish inequalities for the moments of the random variables given in the lemma. These long and tedious calculations are omitted here. Detailed proof is given in Friedrich (1985). See also Müller-Funk (1979) for the technique of using results on empirical processes.

It is interesting to note that under the assumptions of Theorem 4.1, the inequalities (2.9) and (2.11), but not (2.12) can be proved. This also shows that the statement of Theorem 2.1, which at first sight appears to be circumstantial, is justified.

Acknowledgments. This work is part of my Ph.D. thesis, which was submitted to the Mathematical Faculty of the University of Freiburg im Breisgau, West Germany.

I am grateful to Professor Dr. H. Witting for his stimulating interest in this problem. Dr. U. Müller-Funk has read the original text and given much constructive advice. My thanks also go to a referee, an Associate Editor and Professor van Zwet for their valuable suggestions and remarks.

REFERENCES

BOLTHAUSEN, E. (1984). An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. verw. Gebiete* **66** 379–386.

CHAN, Y.-K. and WIERMAN, J. (1977). On the Berry-Esseen theorem for U-statistics. *Ann. Probab.* **5** 136–139.

CHATTERJI, S. D. (1969). An L^p-convergence theorem. *Ann. Math. Statist.* **40** 1068–1070.

CHERNOFF, H. and SAVAGE, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *Ann. Math. Statist.* **29** 972–994.

DOES, R. J. M. M. (1982). Berry–Esseen theorems for simple linear rank statistics under the null-hypothesis. *Ann. Probab.* **10** 982–991.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**, 2nd ed. Wiley, New York.

FRIEDRICH, K. O. (1985). Berry–Esseen Abschätzungen für nichtparametrische Statistiken. Ph.D. thesis, Albert–Ludwigs Universität, Freiburg im Breisgau.

HELMERS, R. and HUŠKOVÁ, M. (1984). A Berry–Esseen bound for L-statistics with unbounded weight functions. In *Asymptotic Statistics 2, Proc. of the Third Prague Symposium on Asymptotic Statistics* (P. Mandl and M. Hušková, eds.) 93–101. North-Holland, Amsterdam.

HELMERS, R. and VAN ZWET, W. R. (1982). The Berry–Esseen bound for U-statistics. In *Statistical Decision Theory and Related Topics III* (S. S. Gupta and J. O. Berger, eds.) **1** 497–512. Academic, New York.

KOROLYUK, V. S. and BOROVSKIKH, YU. V. (1985). Approximation of nondegenerate U-statistics. *Theory Probab. Appl.* **30** 439–450.

MÜLLER-FUNK, U. (1979). Non-parametric sequential tests for symmetry. *Z. Wahrsch. verw. Gebiete* **46** 325–342.

RÜSCHENDORF, L. (1985). Projections and iterative procedures. In *Multivariate Analysis VI* (P. R. Krishnaiah, ed.) 485–493. North-Holland, Amsterdam.

VAN ES, A. J. and HELMERS, R. (1986). Elementary symmetric polynomials of increasing order. Report MS-R8601, Centre for Mathematics and Computer Science, Amsterdam.

VAN ZWET, W. R. (1984). A Berry–Esseen bound for symmetric statistics. *Z. Wahrsch. verw. Gebiete* **66** 425–440.

INSTITUT FÜR MATHEMATISCHE STOCHASTIK
DER ALBERT-LUDWIGS-UNIVERSITÄT
D-7800 FREIBURG IM BREISGAU
FEDERAL REPUBLIC OF GERMANY