The Annals of Statistics
1993, Vol. 21, No. 4, 1692–1710

THE RATE OF CONVERGENCE FOR MULTIVARIATE SAMPLING STATISTICS

BY ERWIN BOLTHAUSEN AND FRIEDRICH GÖTZE

Universität Zürich and Universität Bielefeld

A Berry–Esseen theorem for the rate of convergence of general nonlinear multivariate sampling statistics with normal limit distribution is derived via a multivariate extension of Stein's method. The result generalizes in particular previous results of Bolthausen for one-dimensional linear rank statistics, one-dimensional results of van Zwet and Friedrich for general functions of independent random elements and provides convergence bounds for general multivariate sampling statistics without restrictions on the sampling proportions.

1. Introduction. Let $\mathscr{A}$ be a class of bounded measurable functions $\mathbb{R}^k \to \mathbb{R}$. For $f \in \mathscr{A}$ and $\delta > 0$ define $f_\delta^+(x) = \sup\{f(x+y)\colon |y| \le \delta\}$, $f_\delta^- = -(-f)_\delta^+$ and $\omega(f,\delta) = \int (f_\delta^+ - f_\delta^-)\,d\Phi$ where Φ is the standard normal distribution. Furthermore, let $|f|_\infty = \inf_a \sup_b |f(a) - f(b)|$. Assume that (a) $\mathscr{A}$ is closed under supremum and affine transformations, that is, $f \in \mathscr{A}$ implies $f_\delta^{+/-} \in \mathscr{A}$ and $f \circ M \in \mathscr{A}$ when $M\colon \mathbb{R}^k \to \mathbb{R}^k$ is affine and $\delta > 0$. Furthermore there exists $\gamma > 0$ such that for every $\varepsilon > 0$

$$\sup\{\omega(f,\varepsilon)\colon f \in \mathscr{A}\} \le \gamma\varepsilon.$$

An example where (a) is satisfied is the class of indicator functions of measurable convex sets in $\mathbb{R}^k$ (with $\gamma \le 2\sqrt{k}$). See Corollary 3.2, page 24 of Bhattacharya and Rao (1986) and use Stirling's formula.

Let $n, N \in \mathbb{N}$ satisfying $n \le N$ and π denote a random permutation on the integers $A_N = \{1,\dots,N\}$ which is uniformly distributed. For a function $t\colon A_N^n \to \mathbb{R}^k$, we define

$$T(\pi) \triangleq t(\pi(1),\dots,\pi(n)).$$

We always assume $ET = 0$.

Such a statistic is called *linear* if it is of the form

$$T(\pi) = \sum_{i=1}^n a(i,\pi(i)) \quad \text{where } a\colon A_n \times A_N \to \mathbb{R}^k$$

satisfies

$$\sum_{j=1}^N a(i,j) = 0 \qquad i \in A_n. \tag{1.1}$$

Received September 1989; revised August 1992.

AMS 1991 *subject classifications*. 60F05, 62E20.

Key words and phrases. Berry–Esseen theorem, multivariate central limit theorem, rank statistics, sampling statistics.

The covariance matrix of such a linear statistic is easy to calculate

$$\operatorname{cov}(T) = \frac{1}{N-1}\sum_{i=1}^{n}\sum_{j=1}^{N} a(i,j)^T a(i,j) - \frac{1}{N(N-1)}\sum_{j=1}^{N} c(j)^T c(j),$$

where $c(j) \triangleq \sum_{i=1}^{n} a(i,j)$.

If $N > n$, then a linear statistic determines the matrix $a(i,j)$ if (1.1) is required. In the case $N = n$, this is not true. However, one may then assume that $c(j) = 0$ for all j [by subtracting $c(j)$ from $a(i,j)$ which does not alter the statistics]. If in the case $N = n$, $c(j) = 0$ holds for all j, then the $a(i,j)$ are uniquely determined by the linear statistics. We will always assume that the matrix $a(i,j)$ is given in this canonical form.

If T is a (possibly nonlinear) statistic and $T^0 = \sum_{j=1}^{N} a(j, \pi(j))$ is a linear one, then we define

$$\beta_3(T^0) \triangleq \sum_{i=1}^{n} \frac{1}{N} \sum_{j=1}^{N} |a(i,j)|^3,$$

$$\delta(T,T^0) \triangleq \frac{1}{n}\sum_{j=1}^{n} E\left|(T - T^0)(\pi \circ \tau_j) - (T - T^0)(\pi)\right|,$$

where τ_j, $1 \le j \le n$, are random transpositions independently of π which transpose j with a uniformly chosen element of A_N and $|a(i,j)|$ denotes the Euclidean norm of the vector $a(i,j)$.

THEOREM 1. *There exists a constant C_k which depends only on k such that for any T and any linear statistic T^0 for which* $\operatorname{cov}(T^0) = k \times k$*-identity and any $f \in \mathscr{A}$,*

$$\left|Ef(T) - \int f\,d\Phi\right| \le C_k |f|_\infty \Big[(1+\gamma)\Big(\beta_3(T^0) + \sqrt{\delta(T,T^0)}\Big) + E|T - T^0| + \delta_k\big(\log n\,\delta(T,T^0)\big)^{2/3}\Big]$$

where $\delta_1 = 0$ and $\delta_k = 1$ for $k \ge 2$.

REMARK. The presence of the last summand in the bound for $k > 1$ is a bit annoying and this is certainly not optimal. Remark, however, that if $\sqrt{\delta}$ is small compared with $(\log n)^{-2}$, its contribution is negligible. In typical application, one has

$$\beta_3(T^0) = 0(n^{-1/2}), \qquad E|T - T^0| = 0(n^{-1/2}), \qquad \delta(T,T^0) = 0(n^{-1}),$$

so one can forget about this contribution.

One choice of T^0 is the L_2-projection $p(T)$ of T onto the space of linear statistics. This is easy to calculate. Let

$$\alpha(i,j) \triangleq E(T | \pi(i) = j).$$

As $ET = 0$, we have $\Sigma_j \alpha(i,j) = 0$ for all i. A simple calculation yields

$$p(T) = \sum_{j=1}^{n} a_0(j, \pi(j))$$

with

$$a_0(i,j) = \frac{n-1}{n}\alpha(i,j) \quad \text{if } n = N,$$
$$a_0(i,j) = \frac{N-1}{N}\alpha(i,j) + \frac{N-1}{N(N-n)}\sum_{l=1}^{n}\alpha(l,j) \quad \text{if } N > n.$$

From the form of the bound in our theorem, it is obvious that $p(T)$ may not be the optimal choice. However, from a computational point of view, $p(T)$ is the easiest approximation to handle.

The theorem implies and extends a number of results in the literature. Let us look at some special cases.

$N = n$ is the pure permutation case. Specializing further to $k = 1$ and $T = \Sigma_i a(i, \pi(i))$ the estimate is the same as that given in Bolthausen (1984). The result, therefore, gives a multivariate and nonlinear extension of the estimates derived there. It also extends partially error estimates in the CLT for expectations of smooth functions of bivariate permutation statistics (using Stein's method) by Barbour and Eagleson (1986). (The main emphasis in this paper is however on statistics which are not well approximated by linear statistics.)

For $n < N, T$ is a function of the (ordered) sample $\pi(1), \dots, \pi(n)$ out of A_N. Our result generalizes convergence results on finite population U-statistics by Zhao and Chen (1987) who required that n/N is bounded from below and above.

The case $N \to \infty$ corresponds to the i.i.d. situation. In fact, a simple approximation scheme leads to a corollary for independent random variables:

Let E denote a Polish space with Borel field $\mathscr{E}$, $X_1, \dots, X_n$ be independent $(E, \mathscr{E})$-valued random variables and let t: $E^n \to \mathbb{R}^k$ denote a measurable mapping such that

$$T = t(X_1, \dots, X_n)$$

is integrable.

We again assume $ET = 0$. For f_j: $E \to \mathbb{R}^k$ with $Ef_j(X_j) = 0$ we shall consider linear statistics of the form $\Sigma_j f_j(X_j)$. If T^0 is such a linear statistic, let

$$\beta_3(T^0) \triangleq \sum_{i=1}^{n} E|f_i(X_i)|^3,$$
$$\delta(T, T^0) \triangleq \frac{1}{n}\sum_{j=1}^{n} E\Big|(T - T^0)(X_1, \dots, X_j, \dots, X_n) - (T - T^0)\big(X_1, \dots, \hat{X}_j, \dots, X_n\big)\Big|$$

where $\hat{X}_j$ is an independent copy of X_j.

THEOREM 2. *The same result as Theorem* 1 *holds true* (*with the same* C_k) *but with the modified definitions of* β_3 *and* δ.

REMARK. The easiest choice for T^0 again is the projection $T^0 = \sum_{j=1}^n E(T|X_j)$ but as the difference between T and T^0 enters via the L_1-norm and other more complicated expressions, this may not be the optimal one. Theorem 2 is a corollary of Theorem 1. We will give a proof in Section 4.

For one-dimensional statistics T related results have been proved by van Zwet (1984) and Friedrich (1989) using Fourier methods under moment conditions on *second* order differences of $T - T^0$. This result has been extended to multivariate statistics in Götze (1991) using non-Fourier methods.

One might wonder why we did not start with Theorem 2 and used it to prove results on rank statistics under the hypothesis. But this approach leads to unnatural conditions on moments of derivatives of score functions in the upper bound of Theorem 1.

The paper is organized as follows: In Section 2 we present some preliminary lemmas. In Section 3 Theorem 1 is proved, starting with an outline of the main strategy. Section 4 contains a proof of Theorem 2 and in the last section, we present an application to order statistics.

2. Preliminary lemmas. We will use the multivariate extension of the Stein method given in Götze (1991) together with the combinatorial arguments in Bolthausen (1984) and Schneller (1989).

For $h \in \mathscr{A}$ and $0 \le t \le 1$ introduce

$$\chi_t(x|h) \triangleq \int \Big(h(y) - h\big(t^{1/2}y + (1-t)^{1/2}x\big)\Big)\Phi(dy),$$

$$\psi_t(x|h) \triangleq \frac{1}{2}\int_t^1 \chi_s(x|h)\frac{ds}{1-s},$$

$-\chi_0(x|h) = h(x) - \Phi(h)$ and $-\chi_t(x|h)$ is a smooth approximation of h for small $t > 0$.

If $\varphi\colon \mathbb{R}^k \to \mathbb{R}$ is smooth, we denote by $\varphi^{(j)}$ the jth derivative and by $\varphi^{(j)} \cdot x_1 \cdot x_2 \cdot \ldots \cdot x_j$ the evaluation in the directions $x_1, \ldots, x_j \in \mathbb{R}^k$. Furthermore, we denote by $|\varphi^{(j)}|$ the supremum norm of this j-linear form with respect to the Euclidean metric on $\mathbb{R}^k$.

The following generalization of Stein's differential equation Stein (1972, 1986) to $\mathbb{R}^k$ was introduced in Götze (1981) [see also Barbour (1988)]. It is a diffusion equation of Ornstein–Uhlenbeck type.

LEMMA 1.

(i) $\Delta\psi_t(x|h) - \psi_t^{(1)}(x|h) \cdot x = -\chi_t(x|h)$ *for* $x \in \mathbb{R}^k$ *and* $0 < t \le 1$, *where* Δ *is the Laplace operator*.

(ii) *There are absolute constants* $c_j > 0$ *depending on the dimension k only, such that for* $t > 0$,

$$\sup_x \big|\psi_t^{(j)}(x|h)\big| \le c_j|h|_\infty A_j(t),$$

where $A_j(t) \triangleq t^{-(j-2)/2}$, $j > 2$, $A_2(t) \triangleq -\log t$ *and* $A_1(t) \triangleq A_0(t) \triangleq 1$. *For* $k = 1$ *we have* $A_2(t) \triangleq 1$.

(iii) *Let* Q *denote a finite signed measure on* $\mathbb{R}^k$ *with* $Q(\mathbb{R}^k) = 0$. *Then*

$$\left|\int \psi_t^{(j)}(x|h)Q(dx)\right| \le c_j A_j(t) \sup\left\{\left|\int h(xs + a)Q(dx)\right| : 0 \le s \le 1, a \in \mathbb{R}^k\right\}$$

for $j \in \mathbb{N}_0$.

PROOF. The proof follows by an easy calculation [see Götze (1991), Lemma 2.1]. For the proof of (ii) in the case $k = 1$ notice that $f \triangleq \psi_t'$ is continuous, bounded and satisfies Stein's equation $f'(x) - f(x)x = -\chi_t(x|h)$. Thus it equals the solution (2.9) of Stein (1972) and therefore $f'(x) = \psi_t''(x|h)$ is uniformly bounded by $c|h|_\infty$. □

LEMMA 2. *Assume* (a) *and let* Q *be a probability distribution on* $\mathbb{R}^k$. *If* $a > 0$, *then*

$$\sup_{f \in \mathscr{A}} \left|\int f(dQ - d\Phi)\right| \le \frac{4}{3} \sup_{h \in \mathscr{A}} \left|\int \chi_{\varepsilon^2}(\cdot|h)\, dQ\right| + \frac{5}{2} a \frac{\varepsilon}{1 - \varepsilon^2} \gamma$$

where a^2 *is the* $\frac{7}{8}$*-quantile of the* χ^2*-distribution with* k *degrees of freedom.*

PROOF. This is Lemma 2.11 of Götze (1991). The proof is an easy consequence of standard smoothing inequalities [e.g., Bhattacharya and Ranga Rao (1986), Lemma 1.14]. □

We introduce a slight modification of a construction of Schneller (1989). Let M denote the set of eight-tuples $\mathbf{i} = (i_1, \dots, i_8) \in A_n \times A_N^7$ satisfying

$$\begin{aligned}
&i_1 = i_2 \quad \Leftrightarrow \quad i_3 = i_4 \quad \Leftrightarrow \quad i_5 = i_6 \quad \Leftrightarrow \quad i_7 = i_8,\\
&i_1 = i_3 \quad \Leftrightarrow \quad i_5 = i_7\\
&i_1 = i_4 \quad \Leftrightarrow \quad i_6 = i_7\\
&i_2 = i_3 \quad \Leftrightarrow \quad i_5 = i_8\\
&i_2 = i_4 \quad \Leftrightarrow \quad i_6 = i_8.
\end{aligned} \tag{2.1}$$

If $\mathbf{i} \in M$, we fix permutations $u(\mathbf{i})$, $v(\mathbf{i})$, $s(i)$ of A_N with the following properties: $u(\mathbf{i})$ is the product of at most four transpositions of the elements $i_1, \dots, i_8$ and satisfies $i_1 \to i_7$, $i_3 \to i_5$, $i_2 \to i_8$, $i_4 \to i_6$ and $u(\mathbf{i})$ leaves the elements outside $\{i_1, \dots, i_8\}$ fixed. Under $v(\mathbf{i})$: $i_1 \to i_4$, $i_2 \to i_3$ and $v(\mathbf{i})$ leaves the elements outside $\{i_1, \dots, i_4\}$ fixed. $s(\mathbf{i})$ just transposes i_1 and i_2. Using (2.1) one checks that such permutations always exist.

There is a slight subtlety in the representation of $u(\mathbf{i})$ and $v(\mathbf{i})$ as products of transpositions. If $i_1, \dots, i_8$ are all disjoint, one can put

$$u(\mathbf{i}) \triangleq \tau_{i_1, i_7} \circ \tau_{i_3, i_5} \circ \tau_{i_2, i_8} \circ \tau_{i_4, i_6},$$

where $\tau_{i,j}$ is the transposition of i and j. However, the r.h.s. above does not always define $u(\mathbf{i})$ in an appropriate way. If, for example, $i_1 = i_5$, $i_3 \neq i_7$, then it sends i_1 to i_3. However, we can partition M into pairwise disjoint and nonvoid sets M_r, $M = \bigcup_r M_r$, which are defined by inequalities and equalities among the i_k, $1 \leq k \leq r$, such that on each of the M_r, we can fix a permutation $(\sigma_r(1), \dots, \sigma_r(8))$ of $(1, \dots, 8)$ such that

$$u(\mathbf{i}) = \tau_{i_{\sigma(1)}, i_{\sigma(2)}} \circ \tau_{i_{\sigma(3)}, i_{\sigma(4)}} \circ \tau_{i_{\sigma(5)}, i_{\sigma(6)}} \circ \tau_{i_{\sigma(7)}, i_{\sigma(8)}} \tag{2.2}$$

defines $u(\mathbf{i})$ such that it has the desired properties. The exact form is of no importance for us [see Schneller (1989) for details]. Similar remarks apply to $v(\mathbf{i})$ (whose representation is, of course, simpler).

Let $\mathbf{I} = (I_1, \dots, I_8)$ be a uniformly distributed random variable with values in M and π_1 be a uniformly distributed random permutation of A_N which is independent of $\mathbf{I}$. We put

$$\pi_2 = \pi_1 \circ u(\mathbf{I}), \qquad \pi_3 = \pi_2 \circ v(\mathbf{I}), \qquad \pi_4 = \pi_3 \circ s(\mathbf{I}).$$

Furthermore, let

$$J_k = \pi_1(I_{4+k}) \qquad 1 \leq k \leq 4,$$
$$J_{4+k} = \pi_1(I_k) \qquad 1 \leq k \leq 4.$$

The following lemma is proved in Schneller (1989).

LEMMA 3.

(i) π_2, π_3, π_4 *are uniformly distributed.*

(ii) π_4 *and* I_1 *are independent and for* $1 \leq k \leq 3$, π_k *and* $\{I_1, \dots, I_{2^{3-k}}, J_1, \dots, J_{2^{3-k}}\}$ *are independent.*

(iii) $(I_1, \pi_k(I_1))$ *is uniformly distributed on* $A_n \times A_N$ *for* $1 \leq k \leq 4$. *For* $2 \leq l \leq 8$, $(I_l, \pi_k(I_l))$ *is uniformly distributed in* A_N^2 $(1 \leq k \leq 4)$.

Let S be any mapping from the set of permutations of A_N to $\mathbb{R}$.

LEMMA 4. *There is a universal number K such that for* $1 \leq i, j \leq 4$,

$$E\left|S(\pi_i) - S(\pi_j)\right|$$
$$\leq K\left\{E\left|S(\pi_1) - S(\pi_1 \circ \tau_{I_1, I_2})\right| + E\left|S(\pi_1) - S(\pi_1 \circ \tau_{I_2, I_3})\right|\right\}.$$

PROOF. Obviously, it suffices to prove the bound for $j = i + 1$, $1 \leq i \leq 3$. We take $i = 1$, which is the most complicated case. Also, it suffices to estimate $E(|S(\pi_1) - S(\pi_2)| 1_{\mathbf{I} \in M_r})$, for each set M_r of the above mentioned partition of M. On M_r, $u(\mathbf{i})$ is defined by (2.2). Remarking that $\pi_1 \circ \tau_{I_k, I_l}$ is uniformly distributed for $1 \leq k, l \leq 8$, we see that $E|S(\pi_1) - S(\pi_2)|$ can be estimated by

$$K\left\{E\left|S(\pi_1) - S(\pi_1 \circ \tau_{I_1, I_2})\right| + E\left|S(\pi_1) - S(\pi_1 \circ \tau_{I_2, I_3})\right|\right\}. \qquad \square$$

3. Proof of Theorem 1. Let c always denote a generic constant > 0, depending only on the dimension k, but which may vary from formula to formula. Of course, we may assume that $|f|_\infty \le 1$ for all $f \in \mathscr{A}$.

If $A, B, C > 0$; $n, N \in \mathbb{N}$, $n \le N$, let $\mathscr{T}_{n,N}(A, B, C)$ denote the set of mappings t: $A_N^n \to \mathbb{R}^k$ satisfying $ET = 0$ and for which there exists a linear statistic T^0 such that

$$\beta_3(T^0) \le A, \qquad E|T - T^0| \le B, \qquad \delta(T, T^0) \le C, \qquad \operatorname{cov}(T^0) = id.$$

As $\beta_3(T^0) \ge cn^{-1/2}$, we assume $A \ge n^{-1/2}$. Let

$$\delta \triangleq \delta_{n,N}(A, B, C) \triangleq \sup\left\{\left|Ef(T) - \int f\,d\Phi\right| : f \in \mathscr{A}, t \in \mathscr{T}_{n,N}(A, B, C)\right\}.$$

Of course, δ also depends on the class $\mathscr{A}$, but this is considered to be fixed. We shall estimate δ in terms of n, A, B, C and itself, thus obtaining a recursion relation.

More precisely, we will show that

$$\begin{aligned}\delta(A, B, C) \le c\Big[(1+\gamma)(A + \sqrt{C}) + B + \delta_k(C \log n)^{2/3}\Big]& \\ + \frac{1}{2c_1}\delta(c_1A, c_1B, c_1C),& \end{aligned} \tag{3.1}$$

where $\delta_1 = 0$, $\delta_k = 1$ for $k \ge 2$, and c, c_1 are positive constants. Iterating this inequality and using $\delta \le 1$ proves the theorem.

Before starting with a detailed derivation of (3.1), we give an outline of the strategy. The first step is a fairly standard smoothing which in our case is tailored in such a way that it tallies nicely with the multivariate version of the Stein approach: From Lemma 3 we get for $1/\sqrt{n} < \varepsilon < 1$

$$\delta \le c \sup\{|E\chi_{\varepsilon^2}(T|f)| : f \in \mathscr{A}, t \in \mathscr{T}_{n,N}(A, B, C)\} + c\gamma\varepsilon. \tag{3.2}$$

The ε is chosen later.

The smoothing with χ_{ε^2} is particularly useful as the Stein equation is, by Lemma 1(i) just

$$E\chi_{\varepsilon^2}(T|f) = E\,\Delta\psi_{\varepsilon^2}(T|f) - E\psi_{\varepsilon^2}^{(1)}(T|f)\cdot T. \tag{3.3}$$

The idea of the Stein method is that the second summand on the right-hand side, by a Taylor expansion, approximately cancels with the first one.

There are, however, a number of problems. First, one would like to replace T by a linear version T_0. One step is trivial: If $t \in \mathscr{T}_{n,N}(A, B, C)$, and T^0 is chosen accordingly, then by Lemma 1(ii)

$$\left|E\psi_{\varepsilon^2}^{(1)}(T|f)\cdot T - E\psi_{\varepsilon^2}^{(1)}(T|f)\cdot T^0\right| \le cE|T - T^0| \le cB. \tag{3.4}$$

Handling the nonlinearity inside $\psi_{\varepsilon^2}^{(1)}(\cdot|f)$ is more delicate and must be taken care of together with other things. Writing $T^0 = T^0(\pi) = \sum_{i=1}^n a(i, \pi(i))$ where

π is uniformly distributed on the set of injective mappings $A_n \to A_N$, we have

$$E\psi_{\varepsilon^2}^{(1)}(T|f)\cdot T^0 = \sum_{i=1}^{n} E\psi_{\varepsilon^2}^{(1)}(T(\pi)|f)\cdot a(i,\pi(i)),$$

which we can write in more compact form as

$$nE\big(\psi_{\varepsilon^2}^{(1)}(T(\pi)|f)\cdot a(I,\pi(I))\big).$$

If T were linear, $T = T_0$, we would like to write $T = T - a(I,\pi(I)) + a(I,\pi(I))$ and expand $\psi_{\varepsilon^2}^{(1)}(T|f)$ with Taylor around $T - a(I,\pi(I))$. However, a moment's reflection reveals that this term is not independent of $a(I,\pi(I))$ and so a more complicated procedure is necessary. In fact the combinatorial construction in Section 2 was introduced exactly for this purpose. Coming back to the more general case where T is nonlinear, one immediately sees that writing $T = T_0 + (T - T_0)$ and expanding $\psi_{\varepsilon^2}^{(1)}(\ |f)$ around T_0 would be much too crude. A closer look however reveals that it suffices to have the nonlinear part far enough out of the way that it does not meddle with the combinatorial manipulations alluded to above. This is made precise in the following way.

Let $\pi_1, \pi_2, \pi_3, \pi_4$ be the random permutations constructed in Section 2, and define

$$T_i \triangleq t(\pi_i), \qquad T_i^0 \triangleq T^0(\pi_i), \qquad R_i \triangleq T_i - T_i^0, \qquad 1 \le i \le 4,$$
$$\Delta T_i \triangleq T_{i+1} - T_i, \qquad \Delta T_i^0 \triangleq T_{i+1}^0 - T_i^0, \qquad \Delta R_i \triangleq R_{i+1} - R_i$$

for $i = 1, 2, 3$.

Of course, we can write everything in terms of T_4:

$$E\psi_{\varepsilon^2}^{(1)}(T_4|f)\cdot T_4^0 = E\psi_{\varepsilon^2}^{(1)}\big(T_4^0 + R_4|f\big)\cdot T_4^0,$$

and we now replace R_4 by R_1 estimating the difference by

LEMMA 5.

$$\big|E\psi_{\varepsilon^2}^{(1)}(T_4|f)\cdot T_4^0 - E\psi_{\varepsilon^2}^{(1)}\big(T_4^0 + R_1|f\big)\cdot T_4^0\big| \le c\big(1 + \delta_k(C\log n)^{2/3}\big).$$

We will prove this below.

The reason for this replacement is that R_1 is independent of crucial terms in the combinatorial manipulations. We develop according to the Taylor rule with integral remainder. According to the probabilists' abhorrence of writing integral signs, we express the integral remainder with an expectation over an independent random variable η distributed uniformly on [0, 1]:

$$\begin{aligned} E\psi_{\varepsilon^2}^{(1)}\big(T_4^0 + R_1|f\big)\cdot T_4^0 &= nE\psi_{\varepsilon^2}^{(1)}\big(T_4^0 + R_1|f\big)\cdot a(I_1,J_1)\\ &= nE\psi_{\varepsilon^2}^{(2)}\big(T_2^0 + R_1|f\big)\cdot a(I_1,J_1)\cdot \Delta T_3^0\\ &\quad + nE\Big[\big\{\psi_{\varepsilon^2}^{(2)}\big(T_2^0 + \Delta T_2^0 + \eta\,\Delta T_3^0 + R_1|f\big)\\ &\qquad - \psi_{\varepsilon^2}^{(2)}\big(T_2^0 + R_1|f\big)\big\}\cdot a(I_1,J_1)\cdot \Delta T_3^0\Big]\\ &\triangleq L_1 + L_2 \quad \text{say,} \end{aligned}$$

where we used the fact that $E[\psi_{\varepsilon^2}^{(1)}(T_3^0 + R_1|f) \cdot a(I_1, J_1)] = 0$ because of Lemma 3(ii).

L_1 is treated by the following:

LEMMA 6.

$$\left| L_1 - E\,\Delta\psi_{\varepsilon^2}(T_2|f) \right| \le c\varepsilon^{-1}C.$$

The most delicate part of the proof is the estimation of L_2. In L_2, essentially the third derivative of ψ_{ε^2} enters, whose maximal absolute value is of order ε^{-1}. We have to use the fact that $\psi_{\varepsilon^2}^{(3)}$ is large only on a small part of $\mathbb{R}^k$. This is the place where we have to use a bootstrapping argument leading to a recursion:

LEMMA 7.

$$L_2 \le cA\left(1 + \varepsilon^{-1}\delta_{n,N}\left(cA, B + \frac{c}{\sqrt{n}}, C + \frac{c}{\sqrt{n}}\right)\right).$$

Combining (3.2)–(3.4) and the Lemmas 5, 6 and 7 we obtain

$$\delta_{n,N}(A,B,C) \le c\left[\gamma\varepsilon + B + \delta_k(C\log n)^{2/3} + \varepsilon^{-1}C + A\right] + c_1\varepsilon^{-1}A\delta_{n,N}\left(c_1A, B + \frac{c_1}{\sqrt{n}}, C + \frac{c_1}{n}\right),$$

where $c_1 > 0$ is a constant depending only on k. We may assume $c_1 \ge 2$. As A was assumed to be $\ge c/\sqrt{n}$, we may also assume $B \ge c_1/\sqrt{n}$ and as ε will be chosen $\ge c/\sqrt{n}$, we may also assume $C \ge c_1/n$. Therefore, we get

$$\delta(A,B,C) \le c\left[\gamma\varepsilon + B + \delta_k(C\log n)^{2/3} + \varepsilon^{-1}C + A\right] + c_1\varepsilon^{-1}A\delta(c_1A, c_1B, c_1C),$$

we choose $\varepsilon = \max(2c_1^2A, \sqrt{C})$ and obtain

$$\delta(A,B,C) \le c\left[(1+\gamma)(A + \sqrt{C}) + B + \delta_k(C\log n)^{2/3}\right] + \frac{1}{2c_1}\delta(c_1A, c_1B, c_1C),$$

which is (3.1). Therefore the theorem is proved. □

It remains to prove the three lemmas.

PROOF OF LEMMA 5.

$$\begin{aligned}
&\left| E\psi_{\varepsilon^2}^{(1)}(T_4|f) \cdot T_4^0 - E\psi_{\varepsilon^2}^{(1)}(T_4^0 + R_1|f) \cdot T_4^0 \right| \\
&\quad \le \left| E\left[1_{\{|T_4^0| \ge y\}}\left(\psi_{\varepsilon^2}^{(1)}(T_4^0 + R_4|f) - \psi_{\varepsilon^2}^{(1)}(T_4^0 + R_1|f)\right) \cdot T_4^0\right]\right| \\
&\qquad + \left| E\left[1_{\{|T_4^0| < y\}}\psi_{\varepsilon^2}^{(2)}(T_4^0 + R_1 + \eta(R_4 - R_1)|f) \cdot T_4^0 \cdot (R_4 - R_1)\right]\right|,
\end{aligned}$$

where $y \geq 0$ will be chosen later on and η is a uniformly distributed random variable on [0, 1] which is independent of all r.v. defined so far.

Using Lemma 1(ii) and Chebyshev's inequality the first summand on the r.h.s. is

$$\leq cE\big(|T_4^0| 1_{\{|T_4^0| \geq y\}}\big) \leq cE|T_4^0|^3/y^2$$
$$\leq cy^{-2} \quad \text{using Lemma A1 of the Appendix,}$$

and the second is by Lemma 4 and

$$\varepsilon \geq 1/\sqrt{n} \quad \leq cy \log(n) E|R_4 - R_1| \leq cy \log(n)\delta(T, T^0) \leq cy \log(n)C.$$

We choose y to minimize the sum of the bounds leading to the required inequality for $k \geq 2$.

Note that by Lemma 1(ii) for $k = 1$, $|\psi_t^{(2)}(x)|$ remains bounded and we have in that case no ($\log n$) contribution. □

Proof of Lemma 6. Since $(a(I_1, J_1), \Delta T_3^0)$ and $T_2^0 + R_1$ are independent, we obtain

$$L_1 = nE\psi_{\varepsilon^2}^{(2)}(T_2^0 + R_1|f) \cdot E\big(a(I_1, J_1) \cdot \Delta T_3^0\big),$$

which by an easy calculation

$$\begin{aligned} &= E\,\Delta\psi_{\varepsilon^2}(T_2^0 + R_1|f) \\ &= E\,\Delta\psi_{\varepsilon^2}(T_2|f) \\ &\quad + nE\psi_{\varepsilon^2}^{(3)}(T_2^0 + R_1 + \eta\,\Delta R_1|f) \cdot a(I_1, J_1) \cdot a(I_1, J_1) \cdot \Delta R_1. \end{aligned}$$

By Lemma 1(ii) and Lemma 3 the second summand on the right-hand side is dominated in absolute value by

$$c\varepsilon^{-1} nE|\Delta R_1| E|a(I_1, J_1)|^2 \leq c\varepsilon^{-1}\delta(T, T^0).$$

Therefore,

$$|L_1 - E\,\Delta\psi_t(T_2|f)| \leq c\varepsilon^{-1}C. \qquad \square$$

Proof of Lemma 7. Consider the conditional expectation given $I_1 = i_1, \dots, I_4 = i_4$, $J_1 = j_1, \dots, J_4 = j_4$. This leaves $a(I_1, J_1)$, ΔT_3^0 and ΔT_2^0 fixed. The term R_1 depends on π_1 which is independent of $I_1, \dots, J_4$. We write

$$T_2^0 + R_1 = T_1^0 + R_1 + \Delta T_1^0 = T_1 + \Delta T_1^0.$$

Unfortunately, ΔT_1^0 is not fixed (it still depends on π_1). To overcome this difficulty, we introduce $\tilde{T} \triangleq T_2^0 + R_1$ as a new function of π_1 (keeping $I_1 = i_1, \dots, J_4 = j_4$ fixed).

Let $\tilde{E}\tilde{T}$ denote the conditional expectation of $\tilde{T}$ given $I_1 = i_1, \dots, J_4 = j_4$. We show that $\tilde{T}$ suitably rescaled belongs to $\mathcal{T}_{n,N}(A', B', C')$ where A', B', C' are slightly larger than A, B, C and do not depend on $i_1, \dots, j_4$. Notice that ΔT_1^0 is a sum of terms of the form $a(i, j)$, $a(i, \pi(i))$, $a(\pi^{-1}(j), j)$ and $a(\pi^{-1}(j), \pi(i))$ where $i, j \in \{i_1, \dots, j_4\}$ and $\pi = \pi_1$. Subtracting $\tilde{E}\,\Delta T_1^0$, the first term cancels and we arrive at a sum of centered summands of the other three types, which we denote by S_1, S_2 and S_3, respectively. Here S_1 and S_2 are already linear statistics. For S_1, this is clear. As for S_2 this can be seen by the following calculation:

$$a(\pi^{-1}(j), j) - \tilde{E}a(\pi^{-1}(j), j) = a(\pi^{-1}(j), j) - \frac{1}{N}\sum_{l=1}^{n} a(l, j)$$
$$= \sum_{r=1}^{n} a(r, j)\left(1_{\pi(r)=j} - \frac{1}{N}\right).$$

Let Σ denote the covariance matrix of $T_1^0 + S_1 + S_2$.

We first show that Σ is close to the identity matrix. To see this, it is sufficient to estimate $\tilde{E}|S_1|^2$ and $\tilde{E}|S_2|^2$. We have

$$\tilde{E}|a(i, \pi(i))|^2 = \frac{1}{N}\sum_{j=1}^{N} |a(i, j)|^2 \le \left(\sum_{i=1}^{n} \frac{1}{N}\sum_{j=1}^{N} |a(i, j)|^3\right)^{2/3} \le A^{2/3}$$

and

$$\tilde{E}|a(\pi^{-1}(j), j)|^2 = \frac{1}{N}\sum_{i=1}^{n} |a(i, j)|^2 \le A^{2/3},$$

so

$$\tilde{E}|a(\pi^{-1}(j), j) - \tilde{E}a(\pi^{-1}(j), j)|^2 \le A^{2/3}.$$

Therefore Σ is arbitrary close to the identity matrix if $A \le A_0$ when A_0 is chosen small enough.

We approximate $\hat{T} \triangleq \Sigma^{-1/2}(\tilde{T} - \tilde{E}\tilde{T})$ by the linear statistic $\hat{T}^0 \triangleq \Sigma^{-1/2}(T_1^0 + S_1 + S_2)$. Then

$$\beta_3(S_1) \le c \sup_i \frac{1}{N}\sum_{j=1}^{N} |a(i, j)|^3 \le c\beta_3(T_1^0) \le cA,$$

$$\beta_3(S_2) \le c \sup_j \sum_{r=1}^{n} \frac{1}{N}\sum_{s=1}^{N} |a(r, j)|^3 \left|1_{s=j} - \frac{1}{N}\right|^3$$
$$\le cA.$$

Of course, we can assume that A is smaller than a fixed constant A_0. $A \le A_0$ and $T \in \mathscr{T}_{n,N}(A, B, C)$ together imply

$$\beta_3(\hat{T}^0) \le cA.$$

Write

$$\hat{T} - \hat{T}^0 = \Sigma^{-1/2}(T_1 - T_1^0) + \Sigma^{-1/2}S_3.$$

In order to prove that $\hat{T} \in \mathscr{T}_{n,N}(cA, B + c/\sqrt{n}, C + c/n)$, it suffices to observe that

$$E|S_3| \le c \sup_{i,j} E\big|a(\pi^{-1}(j), \pi(i))\big|$$

$$= c \sup_{i,j}\left[\frac{1}{N}|a(i,j)| + \frac{1}{N(N-1)}\sum_{l\neq i}\sum_{r\neq j}|a(l,r)|\right]$$

$$\le cn^{-1/2} \quad \left(\text{since } |a(i,j)|^2/N \le k\right)$$

and that

$$\frac{1}{n}\sum_{l=1}^{n} E|S_3(\pi\circ\tau_l) - S_3(\pi)| \le cn^{-1},$$

which is as straightforward, too.

In order to obtain a recursion we approximate the distribution of $\tilde{T} - E\tilde{T}$ again by a normal distribution with mean zero and covariance Σ, say Φ_Σ. Let $\mathscr{C}$ denote the Σ-algebra generated by $I_1, \dots, I_4, J_1, \dots, J_4$. Then we may write

$$\begin{aligned} L_2 = nE\Big\{\Big[&E\big(\psi_t^{(3)}(\tilde{T} + b|f) - \tilde{\psi}_\Sigma(b)|\mathscr{C}\big) + E\big(\tilde{\psi}_\Sigma(b)|\mathscr{C}\big)\Big] \\ &\times a(I_4, J_1)\cdot \Delta T_3^0 \cdot \left(\Delta T_2^0 + \eta\,\Delta T_3^0\right)\Big\} \\ = L_3 + L_4 \quad &\text{say,}\end{aligned} \tag{3.5}$$

where $b \triangleq E\tilde{T} + \eta'(\Delta T_2^0 + \eta\,\Delta T_3^0)$ denotes a $\mathscr{C}$-measurable r.v., η' is uniformly distributed in [0, 1] and independent of all r.v. introduced so far, and

$$\tilde{\psi}_\Sigma(x) \triangleq \int \psi_t^{(3)}(x + y|f)\Phi_\Sigma(dx).$$

Since Σ is $\mathscr{C}$-measurable but approximates the $k \times k$-identity uniformly provided that $\beta_3(T^0) \le A \le A_0$ with A_0 sufficiently small we conclude that $\tilde{\psi}_\Sigma(x)$ is uniformly bounded with respect to x and Σ. Thus

$$L_4 \le cA.$$

Using Lemma 1(iii) [with $j = 3$ and $Q(dx) = P(\tilde{T} \in dx) - \Phi_\Sigma(dx)$] together with the invariance of $\mathscr{A}$ under shifts and the linear transformation $\Sigma^{1/2}$ we obtain

$$L_3 \le c\varepsilon^{-1}A\delta_{n,N}\left(cA, B + \frac{c}{\sqrt{n}}, C + \frac{c}{n}\right).$$

Implementing these estimates into (3.5) proves Lemma 7. □

4. Proof of Theorem 2. Let $X_j, j = 1, \dots, n$ be defined on $(\mathscr{A}, \mathscr{B}, P)$ taking values in E. Since E is a polish space there exists finite σ-fields $\mathscr{A}_L \subset \mathscr{E}$ (partitions of E) which form a monotone increasing sequence with $\sigma(\bigcup_L \mathscr{A}_L) = \mathscr{E}$.

Define

$$t^L(k_1,\dots,k_n) \triangleq \mathbb{E}\big(t_n(X_1,\dots,X_n)|X_j \in B_{kj}^L,\ j=1,\dots,n\big),$$

where $\mathscr{A}_L$ is given by the partition $E = B_1^L \dot\cup \cdots \dot\cup B_{n_L}^L$ and $1 \le k_j \le n_L$. Here t^L is considered as a function of r.v. $K_1,\dots,K_n$ taking values in the discrete probability space $A_{N_L}^n$ with probabilities

$$\Pi_j P(K_j = k_j) \triangleq \Pi_j P\big(X_j \in B_{k_j}^L\big).$$

By Jensen's inequality, the conditions of Theorem 2 together with the martingale limit theorem show that for $a(j,l) \triangleq E(T|X_j \in B_l^L) = E(t^L|K_j = l)$

$$\lim_{L\to\infty} \sum_{j=1}^n E\big(b^T a(j,K_j)\big)^p = E\big(b^T E(T|X_j)\big)^p, \quad p = 1,2,3,\ b \in \mathbb{R}^k$$

and that $\beta_3(T^0)$, $E|T - T^0|$ and $\delta(T,T_0)$ are limits of the corresponding discrete quantities.

As remarked at the beginning of Section 3, we may assume $\sup_{f\in\mathscr{A}}|f|_\infty \le 1$. If $f \in \mathscr{A}$, we have similarly as in Lemma 2 for $\varepsilon > 0$,

$$\begin{aligned} &\left|Ef(T) - \int f\,d\Phi\right| \\ &\quad\le (1 + o(\varepsilon^0)) \sup_{h\in\mathscr{A}} \left|E\chi_{\varepsilon^2}(T|h) - \Phi(\chi_{\varepsilon^2}(\cdot|h)\right| + 0(\varepsilon) \\ &\quad\le (1 + o(\varepsilon^0)) \limsup_{L\to\infty} \sup_{h\in\mathscr{A}} \left|E\chi_{\varepsilon^2}(t^L(h) - \Phi(\chi_{\varepsilon^2}(\cdot|h))\right| + 0(\varepsilon). \end{aligned}$$

As $\chi_{\varepsilon^2}(\cdot|h)$ is a convex combination of elements in $\mathscr{A}$ we obtain by letting $\varepsilon \to 0$,

$$\sup_{f\in\mathscr{A}}\left|Ef(T) - \int f\,d\Phi\right| \le \limsup_{L} \sup_{f\in\mathscr{A}}\left|Ef(t^L) - \int f\,d\Phi\right|.$$

Finally we may choose for every L an integer $N \in \mathbb{N}$ large enough such that $P(X_j \in B_{kj}^L) = m_{k_j}/N + 0(1/N)$, $j = 1,\dots,n_L$, $m_{k_j} \in \mathbb{N}$ and a function $t^{L,N}$: $A_N^n \to \mathbb{R}^k$ with $t^{L,N}(A_N^n) = t^L(A_{n_L}^n)$. Let π denote now a random permutation on A_N with uniform distribution. As $N \to \infty$, the laws of $(\pi(1),\dots,\pi(n))$ converges in variation to that of $(K_1,\dots,K_n)$.

Furthermore, the corresponding quantities for $a(j,l)$, $\beta_3(T^0)$, $E|T - T^0|$ and $\delta(T,T_0)$ for the r.v. $t^{L,N}$ converge to those of t^L as $N \to \infty$ (and therefore to those of T as $L \to \infty$).

Thus Theorem 2 can be deduced from Theorem 1. □

5. A Sampling version of linear order statistics. Sampling versions of limit theorems for all kind of statistics have become important in recent years in connection with bootstrap and jackknife procedures, see, for example, Wu (1990). Our main Theorem 1 can be used quite routinely to obtain such results. We illustrate this by deriving a sampling version of a Berry–Esseen theorem for linear functionals of order statistics which appears to be new.

For a random permutation π on A_N, and $n < N$, we define the ordered sample $r(\pi) = (r_1(\pi), \dots, r_n(\pi))$ by the requirement $1 \le r_1 < r_2 < \cdots < r_n \le N$ and $\{r_1, \dots, r_n\} = \{\pi(1), \dots, \pi(n)\}$. Of course, we can also regard r just as a random subset of n elements in A_N.

For a real matrix $A = (a(i,k))_{i \in A_n, k \in A_N}$, we define

$$T(\pi) \triangleq \sum_{i=1}^{n} a(i, r_i(\pi)).$$

It should be remarked that T is not linear in the sense of Section 1 and in fact it appears to be unavoidable to assume a kind of second order differentiability of A. In order to keep the calculation as simple as possible, we assume these derivatives to be bounded. Let

$$\begin{aligned}
\Delta &= \max_{i,k} |a(i,k)|, \\
\Delta_1 &= \max_{1 \le i \le n-1} \max_{k} |a(i+1,k) - a(i,k)|, \\
\Delta_2 &= \max_{i} \max_{1 \le k \le N-1} |a(i,k+1) - a(i,k)|, \\
\Delta_{12} &= \max_{1 \le i \le n-1} \max_{1 \le k \le N-1} |a(i+1,k+1) \\
&\qquad + a(i,k) - a(i+1,k) - a(i,k+1)|,
\end{aligned}$$

To simplify things further, we assume $ET = 0$, $\operatorname{var}(T) = 1$,

$$\begin{aligned}
\Delta &\le K n^{-1/2}, \\
\Delta_1 &\le K n^{-3/2}, \\
\Delta_2 &\le K n^{-1/2} N^{-1}, \\
\Delta_{12} &\le K n^{-3/2} N^{-1},
\end{aligned} \tag{5.1}$$

for some constant $K > 0$. We want to prove:

THEOREM 3. *Under the assumptions* (5.1),

$$\sup_x |P(T \le x) - \Phi(x)| \le c(K)/\sqrt{n}.$$

REMARK. The above setting is natural only when n stays away from N, $n \le \lambda N$, say, for some $\lambda \in (0,1)$. For example, if

$$T = \sum_{i=1}^{n} c_i r_i,$$

an easy calculation yields

$$\operatorname{var}(T) = \frac{\operatorname{var}(r_1)}{n-1} \left[n \sum_{i,j} (i \wedge j) c_i c_j - (\bar{c})^2 \right],$$

where $\bar{c} = \sum_{i=1}^{n} i c_i$. If $n \le \lambda N$, $\lambda \in (0,1)$, then $\operatorname{var}(r_1) \sim (N/n)^2$. In order to have $\operatorname{var}(T) = 1$, one needs $\max_i c_i = O(1/N\sqrt{n})$ if all c_i are of about the same size and if not $n\sum_{i,j}(i \wedge j)c_i c_j - (\bar{c})^2 = o(n^4)$ which would be a degenerate situation. So we see that (5.1) is a natural condition if one is willing to use boundedness conditions at all. If, however, n is close to N, then $\operatorname{var}(r_1)$ is of course much smaller than $(N/n)^2$. T may then be written as a function of the subset $\{1,\dots,N\} \setminus \{r_1,\dots,r_n\}$ and our Theorem 3 gives a bound which typically is of order $(N-n)^{-1/2}$. We leave the details to the reader.

PROOF OF THEOREM 3. In order to apply our main Theorem 1, we have to find an appropriate linear statistic. This linear statistic is

$$T_0 = \sum_{i=1}^{n} \alpha(i, \pi(i)),$$

where $\alpha(i,k) = \beta(i) - \gamma(k)$, $1 \le i \le n$, $1 \le k \le N$. The description of β and γ is given in terms of a random sampling $1 \le s_1 < \cdots < s_n \le N$ which, for later purposes, is assumed to be independent of π. We write $\langle\ \rangle$ for the expectation on this s-sampling.

If $k \in A_N$, let

$$I(k) \triangleq \min\{j \in A_n : s_j \ge k\}.$$

If $s_n < k$, we set $I(k) \triangleq n+1$. We define

$$\gamma(k) \triangleq \left\langle a\big(I(k), s_{I(k)}\big) + \sum_{j=I(k)}^{n-1} \big(a(j+1, s_j) - a(j, s_j)\big) \right\rangle,$$

$$\beta(i) \triangleq \left\langle \sum_{j=1}^{n-1} \big(a(j+1, s_j) - a(j, s_j)\big) \right\rangle.$$

With these definitions, we get

$$T(\pi) - T_0(\pi) = \sum_{i=1}^{n} \left\{ \langle a\big(I(r_i), s_{I(r_i)}\big) - a\big(I(r_i), r_i\big)\rangle + \left\langle \sum_{j \in [i, I(r_i)]} d(i,j) \right\rangle \right\},$$

where

$$d(i,j) = a(j+1, s_j) + a(j, r_i) - a(j+1, r_i) - a(j, s_j)$$

and for $a, b \in \mathbb{N}$,

$$\sum_{j \in [a,b]} = \begin{cases} \sum_{j=a}^{b-1}, & \text{if } a < b, \\ 0, & \text{if } a = b, \\ -\sum_{j=b}^{a-1}, & \text{if } b < a. \end{cases}$$

We write

$$U(\pi) \triangleq \sum_{i=1}^{n} \left(\langle a\left(I(r_i), s_{I(r_i)}\right) - a\left(I(r_i), r_i\right)\rangle\right),$$

$$V(\pi) \triangleq \sum_{i=1}^{n} \left\langle \sum_{j \in [i, I(r_i)]} d(i,j) \right\rangle.$$

The reader should keep in mind that $I(k)$ depends on our sampling s.

Obviously

$$|U(\pi)| \leq \sum_{i=1}^{n} \langle |r_i - s_{I(r_i)}| \rangle \Delta_2,$$

$$|V(\pi)| \leq \sum_{i=1}^{n} \left\langle |i - I(r_i)| \left(|s_i - r_i| \vee |s_i - s_{I(r_i)}|\right) \right\rangle \Delta_{12};$$

$r_i - s_{I(r_i)}$ is stochastically of the order of one spacing of the sample, that is, of order N/n. It is easily checked that

$$E\langle |r_i - s_{I(r_i)}|^2 \rangle = O\left(\frac{N^2}{n^2}\right)$$

and similarly

$$E\left\langle |i - I(r_i)|^2 \left(|s_i - r_i| \vee |s_i - s_{I(r_i)}|\right)^2 \right\rangle = O(N^2). \tag{5.2}$$

Therefore

$$E|T - T_0|^2 \leq c\left\{N^2\Delta_2^2 + n^2N^2\Delta_{12}^2\right\} \leq c(K)/n. \tag{5.3}$$

We get $\operatorname{var}(T_0) = 1 + O(1)$ and can as well assume that $\operatorname{var}(T_0) = 1$, as is required.

$$\beta_3(T_0) \leq c(K)/\sqrt{n} \tag{5.4}$$

is an easy consequence of $\Delta \leq K/n$, $\Delta_1 \leq K/n^{3/2}$.

The estimation of $\delta(T, T_0)$ is slightly more delicate. If $j \in A_n$, then $r' = (r_1', \dots, r_n') = r(\pi \circ \tau_j)$ is r with probability n/N or, with probability $(N - n)/N$, is obtained by taking one of the r_k out and replacing it by an element which is uniformly chosen among the elements $x \notin \{r_1, \dots, r_n\}$. If, for example, $x > r_k$, $r_l < x < r_{l+1}$, say, where $l \geq k$, then $r_i' = r_i$ for $i < k$, $r_i' = r_{i+1}$ for $k \leq i < l$, $r_l' = x$, $r_i' = r_i$ for $i > l$, and similarly when $x < r_k$. In the above case, we have

$$\left| \sum_{i=1}^{n} \left\{ a\left(I(r_i'), s_{I(r_i')}\right) - a\left(I(r_i'), r_i'\right) - a\left(I(r_i), s_{I(r_i)}\right) + a\left(I(r_i), r_i\right) \right| \right.$$

$$= \left| a\left(I(x), s_{I(x)}\right) - a\left(I(x), x\right) - a\left(I(r_k), s_{I(r_k)}\right) - a\left(I(r_k), r_k\right) \right|$$

$$\leq \Delta_2\left(|s_{I(x)} - x| + |s_{I(r_k)} - r_k|\right).$$

If, for fixed x, r_k, that is, fixed π and τ_j, we take the s-average $\langle \ \rangle$, it is easily seen that

$$\langle |s_{I(x)} - x| + |s_{I(r_k)} - r_k| \rangle = O\left(\frac{N}{n}\right)$$

and therefore

$$\frac{1}{n}\sum_{j=1}^{n} E\left|U(\pi) - U(\pi \circ \tau_j)\right| \le c(K)\Delta_2 \frac{N}{n} \le c(K) n^{-3/2}, \tag{5.5}$$

which is better than required.

We come to the estimation of

$$\frac{1}{n}\sum_{j=1}^{n} E\left|V(\pi \circ \tau_j) - V(\pi)\right|.$$

We fix first j as before and set $\pi(j) = r_k$, $\pi \circ \tau_j = \pi'$ and the ordered sample by $(r_1', \dots, r_n')$, where

$$\{r_1', \dots, r_n'\} = \{r_1, \dots, r_{k-1}, r_{k+1}, \dots, r_n, x\}.$$

We keep the sample s fixed for the moment. For $i \ne k$, r_i is either r_i', r_{i+1}' or r_{i-1}'. If $x = r_k'$, we see that

$$\left| \sum_{i:\, i \ne k} \sum_{t \in [i, I(r_i)]} d(i,t) - \sum_{i:\, i \ne k} \sum_{t \in [i, I(r_i')]} d'(i,t) \right|$$
$$\le c \sum_i \big(d(i,i) \vee d(i,i+1) \vee d(i,i-1)\big).$$

Taking $E\langle\ \rangle$ expectation, this is

$$\le c(K) n\sqrt{n}\, \frac{N}{n} (n^{-3/2} N^{-1}) = c(K) n^{-1}.$$

By the estimates (5.1) and (5.2), we have

$$E\left\langle \sum_{t \in [k, I(k)]} |d(k,t)| \right\rangle \le c(K) n^{-3/2}.$$

Therefore, we get

$$\frac{1}{n}\sum_{j=1}^{n} E\left|V(\pi \circ \tau_j) - V(\pi)\right| \le c(K) n^{-1}.$$

Together with (5.5), this proves

$$\delta(T, T_0) \le c(K) n^{-1}.$$

With (5.3), (5.4) and Theorem 1, this proves Theorem 3. □

APPENDIX

In this Appendix we prove a moment inequality which might be of independent interest by means of simplified arguments from the proof of Theorem 1.

LEMMA A1. *For a linear statistic $T^0 \in \mathbb{R}^k$ we have*

$$E|T^0|^3 \le c\Big(\big(\operatorname{tr}\operatorname{cov}(T^0)\big)^{3/2} + \beta_3(T^0)\Big)$$

for some constant $c > 0$.

PROOF. For $h(x) \triangleq |x|^3$, $x \in \mathbb{R}^k$, where $|\cdot|$ denotes the Euclidean norm, notice that Lemma 1(i) holds and that the bound in Lemma 1(ii) can be replaced by

$$\big|\psi_t^{(j)}(x|h)\big| \le c\big(1 + |x|^{2-j}\big) \tag{A2}$$

for $0 \le j \le 3$ and $t > 0$, since $h(x)$ admits three uniformly bounded derivatives in $\mathbb{R}^k \setminus \{0\}$.

By Jensen's inequality we have for fixed $t = 0.1$, say, (using the notations of the proof of Theorem 1)

$$\begin{aligned}(1-t)^{3/2} Eh(T_4^0) - Eh(S) &\le Eh\big((1-t)^{1/2}T_4^0 + t^{1/2}S\big) - Eh(S)\\ &= -E\chi_t(T_4^0|h)\\ &= -\big(E\,\Delta\psi_t(T_4^0|h) - E\psi^{(1)}(T_4^0|h)\cdot T_4^0\big),\end{aligned} \tag{A3}$$

using Lemma 1(i), where S is independent of all other r.v. and has normal distribution with mean zero and covariance $\operatorname{cov}(T^0)$.

Similar arguments to those in following (3.5) immediately yield [exchanging the argument T_4^0 of $\psi^{(1)}$ by T_2^0 which is independent of $a(I_1, J_1)$]

$$\begin{aligned}&\big|E\,\Delta\psi_t(T_4^0|h) - nE\psi^{(1)}(T_4^0|h)\cdot a(I_1,J_1)\big|\\ &\quad\le n\big|E\psi_t^{(3)}\big(T_2^0 + \eta'(\Delta T_2^0 + \eta\,\Delta T_3^0)|h\big)\\ &\qquad\times a(I_1,J_1)\cdot\Delta T_3^0\cdot\big(\Delta T_2^0 + \eta\,\Delta T_3^0\big)\big|\\ &\quad\le c\beta_3(T^0) \quad \text{by (A2)}.\end{aligned} \tag{A4}$$

Here η', η are uniformly distributed in $[0,1]$ and independent of all other r.v. Thus (A3) and (A4) together imply

$$\begin{aligned}Eh(T^0) &\le c\big(\beta_3(T^0) + Eh(S)\big)\\ &\le c\Big(\beta_3(T^0) + \big(\operatorname{tr}\operatorname{cov}(T^0)\big)^{3/2}\Big),\end{aligned}$$

proving the lemma. □

REFERENCES

BARBOUR, A. D. (1990). Stein's method for diffusion approximations. *Probab. Theory Related Fields* **84** 297–322.

BARBOUR, A. D. and EAGLESON, G. K. (1986). Random association of symmetric arrays. *Stochastic Anal. Appl.* **4** 239–281.

BHATTACHARYA, R. N. and RANGA RAO, R. (1986). *Normal Approximation and Asymptotic Expansions*. Wiley, New York.

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

FRIEDRICH, K. O. (1989). A Berry–Esseen bound for functions of independent random variables. *Ann. Statist.* **17** 170–183.

GÖTZE, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19** 724–739.

SCHNELLER, W. (1989). Edgeworth expansions for linear rank statistics. *Ann. Statist.* **17** 1103–1123.

STEIN, C. (1972). 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. Probab.* **2** 583–602. Univ. California Press, Berkeley.

STEIN, C. (1986). *Approximate Computation of Expectations*. IMS, Hayward, CA.

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

WU, C. F. J. (1990). On the asymptotic properties of the jackknife histogram. *Ann. Statist.* **18** 1438–1452.

ZHAO, L. and CHEN, X (1987). Berry–Esseen bounds for finite-population U-statistics. *Scientia Sinica Ser. A* **2** 113–127.

INSTITUT FÜR ANGEWANDTE MATHEMATIK
UNIVERSITÄT ZÜRICH
8001 ZÜRICH
SWITZERLAND

FAKULTÄT FÜR MATHEMATIK
UNIVERSITÄT BIELEFELD
4800 BIELEFELD 1
GERMANY