The Annals of Statistics
1989, Vol. 17, No. 3, 1103–1123

EDGEWORTH EXPANSIONS FOR LINEAR RANK STATISTICS

BY WALTER SCHNELLER[1]

Technische Universität Berlin

An Edgeworth expansion of first order is established for general linear rank statistics under the null hypothesis with a remainder term that is usually of order n^{-1}. Furthermore, corresponding results for the second order are formulated, but not proved here. The proof for the first order is based on Stein's method and on an extension of a combinatorial method of Bolthausen. It is also shown that conditions of van Zwet imply up to a small factor our conditions for the validity of Edgeworth expansions. Moreover, our proof for the first order also provides us with a result about Edgeworth expansions for smooth functions.

1. Introduction. Let $A = (a_{ij})$ be an $n \times n$-matrix of real numbers. Let

$$\mu_A = \sum_{i,j} a_{ij}/n, \qquad \sigma_A^2 = \sum_{i,j} \check{a}_{ij}^2/(n-1), \qquad \hat{a}_{ij} = \check{a}_{ij}/\sigma_A,$$

where

$$\check{a}_{ij} = a_{ij} - n^{-1}\sum_l a_{il} - n^{-1}\sum_k a_{kj} + n^{-2}\sum_{k,l} a_{kl}.$$

Furthermore, let

$$\beta_A = \sum_{i,j} \hat{a}_{ij}^3 \quad \text{and} \quad \delta_A = \sum_{i,j} \hat{a}_{ij}^4.$$

If π is uniformly distributed on the set $\mathscr{P}_n$ of permutations of $\{1,\dots,n\}$, then asymptotic normality of the linear rank statistic

$$\mathscr{T}_A = \Big(\sum_i a_{i\pi(i)} - \mu_A\Big)/\sigma_A = \sum_i \hat{a}_{i\pi(i)}$$

has been proved under various conditions by Hoeffding (1951), Motoo (1957) and others [see also Schneller (1988)]. Results on the rate of convergence have been obtained, e.g., by Does (1982) (for the case $a_{ij} = e_i d_j$), Ho and Chen (1978) and most successfully by Bolthausen (1984). He proved the existence of an absolute constant $K > 0$ such that

$$\sup_{z\in\mathbb{R}} |\mathscr{F}_A(z) - \Phi(z)| \le K\beta_A/n, \tag{1.1}$$

where $\mathscr{F}_A$ is the distribution function of $\mathscr{T}_A$ and Φ is the standard normal distribution function.

Received July 1987; revised October 1988.

[1]Now at SEL Software GmbH.

AMS 1980 *subject classifications.* Primary 60F05, 62E20.

Key words and phrases. Edgeworth expansion, linear rank statistic, Stein's method.

The purpose of this paper is to establish Edgeworth expansions of first order for $\mathscr{F}_A$ and $E(q(\mathscr{T}_A))$ where q: $\mathbb{R} \to \mathbb{R}$ is sufficiently smooth. To be more precise, let us consider for a moment $E(q(\mathscr{T}_A))$ with q only bounded. In order to establish Edgeworth expansions for this term we have to assume some smoothness of the function q or of the distribution of $\mathscr{T}_A$.

In Theorem 2.1 we assume that q has in addition a bounded first and second derivative. But we impose no smoothness condition on the distribution of $\mathscr{T}_A$.

In Theorem 2.4 we consider $\mathscr{F}_A$ (i.e., the functions $q = 1_{(-\infty, z]}$, $z \in \mathbb{R}$) and assume in principle that the second differences of $\mathscr{F}_A$ fulfill a boundedness condition. In both theorems the remainder term is of order δ_A/n. We note that in very many cases δ_A/n is of order n^{-1}.

Unfortunately the condition (2.5) of Theorem 2.4 is not very practicable though it is very natural and almost necessary [see Remark 2.11(a)]. But for the case $a_{ij} = e_i d_j$ we are able to verify this condition under the practicable conditions of van Zwet (1982) [see Theorem 2.12(a)].

Furthermore, for the case of the distribution function $\mathscr{F}_A$ we formulate corresponding results for Edgeworth expansions of second order [see Theorem 2.7 and Theorem 2.12(b)]. But we do not prove these results and refer the reader for complete proofs to the thesis of the author [Schneller (1987)]. In addition, some remarks on these proofs may be found at the end of this paper (see Section 7). We note that the result of Theorem 2.12(b) contains the result of Does (1983) [see Remark 2.18(c)] and in the two-sample case (i.e., $a_{ij} = e_i d_j$ with $e_1 = \cdots = e_m = 0$, $e_{m+1} = \cdots = e_n = 1$) this theorem is comparable (up to a factor n^{ε}) to the results of Bickel and van Zwet (1978) and Robinson (1978).

Section 2 contains our results. In Section 3 we introduce our two main methods, namely the method of Stein (1972) and an extension of the combinatorial method of Bolthausen (1984). Using these two methods we prove the basic equation (3.12). This equation gives a kind of Edgeworth expansion for $E(q(\mathscr{T}_A))$ where q is differentiable and bounded. From this equation we deduce Theorem 2.1 in Section 4 and Theorem 2.4 in Section 5. In Section 5 we have to substitute the functions $1_{(-\infty, z]}$ by convenient smooth functions (see Lemma 5.2). We note that the main difference between these two sections is, roughly speaking, the different treatment of the two terms $|x|\,|q(x+y) - q(x)|$ and $|q'(x+y) - q'(x)|$. In Section 4 we simply apply the mean value theorem to both terms [see (4.8)], while in Section 5 we need a result like (1.1) (see Proposition 5.7) for the first and condition (2.5) for the second of these two terms (cf. Proof of Lemma 5.3).

The straightforward use of the mean value theorem in (4.8) reveals that the result of Theorem 2.1 is surely not optimal. We have not tried to improve it, since the emphasis of this paper lies more on Edgeworth expansions for $\mathscr{F}_A$.

In Section 6 we establish the condition (2.5) of Theorem 2.4 using a result of van Zwet (1982). Under the conditions (2.13)–(2.15), this result gives an estimation of the characteristic function of $\mathscr{T}_A$ for arguments t with $c_1 \log n \le |t| \le c_2 n^{3/2}$ [cf. also (6.6)]. From this estimation we deduce (2.5) essentially with the help of Lemma 6.3.

2. The results. First, we need some notation. Given $F: \mathbb{R} \to \mathbb{R}$ and $y \in \mathbb{R}$, we define

$$\|F\| = \sup\{|F(z)| : z \in \mathbb{R}\}$$

and the second difference of F related to y by

$$\Delta_y^2 F(z) = [F(z+2y) - F(z+y)] - [F(z+y) - F(z)], \qquad z \in \mathbb{R}.$$

The interpolating polynomial to F of degree 2 at the points z, $z+y$ and $z+2y$ is

$$P_y^2(x; z, F) = F(z) + [F(z+y) - F(z)](x-z)y^{-1} + \Delta_y^2 F(z)\tfrac{1}{2}(x-z-y)(x-z)y^{-2}, \qquad x, z \in \mathbb{R}.$$

For the matrix A and $l \in \mathbb{N}$ we define $\hat{A} = (\hat{a}_{ij})$ and

F_A = distribution function of $T_A = \sum_i a_{i\pi(i)}$,

$D_A = (\delta_A/n)^{1/2}, \qquad E_A = \left(\sum_{i,j}|\hat{a}_{ij}|^5/n\right)^{1/3}$,

$M(l, A)$ = set of all $(n-l) \times (n-l)$ matrices, which can be obtained from A by cancelling l rows and l columns,

$N(l, A) = \bigcup\{M(r, A): 0 \le r \le l \wedge (n-1)\}$.

Finally, the expansions are

$$e_{1,A}(x) = \Phi(x) - \psi(x)\tfrac{1}{6}\lambda_{1,A}(x^2 - 1),$$
$$e_{2,A}(x) = \Phi(x) - \psi(x)\left\{\tfrac{1}{6}\lambda_{1,A}(x^2-1) + \tfrac{1}{24}\lambda_{2,A}(x^3 - 3x) + \tfrac{1}{72}\lambda_{1,A}^2(x^5 - 10x^3 + 15x)\right\},$$

where

$$\psi = \Phi', \qquad \lambda_{1,A} = n^{-1}\sum_{i,j}\hat{a}_{ij}^3$$

and

$$\lambda_{2,A} = n^{-1}\sum_{i,j}\hat{a}_{ij}^4 + 3n^{-1} - 3n^{-2}\sum_{i,j,k}\left(\hat{a}_{ij}^2\hat{a}_{ik}^2 + \hat{a}_{ij}^2\hat{a}_{kj}^2\right).$$

Here is our first result:

2.1. THEOREM. *There exist positive numbers K_1, K_2 and K_3 such that for all A satisfying $\sigma_A > 0$ and all*

(2.2) $\quad q \in \mathscr{D} = \{g: \mathbb{R} \to \mathbb{R}$: *g is twice differentiable and g, g', g'' are bounded*$\}$,

we have

(2.3) $$\left|E(q(\mathscr{T}_A)) - \int q e_{1,A}' \, dx\right| \le (K_1\|q\| + K_2\|q'\| + K_3\|q''\|)D_A^2.$$

The next result deals with Edgeworth expansions of first order for the distribution function $\mathscr{F}_A$.

2.4. THEOREM. *There exist positive numbers K_4 and K_5 such that for all A satisfying $\sigma_A > 0$ and the condition*

there exists a positive constant C_1 such that

$$|\Delta_y^2 F_B(z)| \le C_1\left(D_A^2 + y^2\right) \tag{2.5}$$

for all $z \in \mathbb{R}, 0 \le y \le D_A$ and $B \in N(8, \hat{A})$,

we have

$$\|\mathscr{F}_A - e_{1,A}\| \le (K_4 C_1 + K_5) D_A^2. \tag{2.6}$$

The corresponding result of this theorem for the second order is

2.7. THEOREM. *There exist positive numbers K_6, K_7 and K_8 such that for all A satisfying $\sigma_A > 0$ and the conditions*

there exists a positive constant C_2 such that

$$|F_B(x) - P_{E_A}^2(x; z, F_B)| \le C_2\left(E_A^3 + (x - z)^3\right) \tag{2.8}$$

for all $z \in \mathbb{R}$, $z \le x \le z + 3E_A$ and $B \in N(16, \hat{A})$,

there exists a positive constant C_3 such that

$$(|z| + 1)\Delta_y^2 F_B(z)| \le C_3\left(E_A^2 + y^2\right) \tag{2.9}$$

for all $z \in \mathbb{R}, 0 \le y \le E_A$ and $B \in N(16, \hat{A})$,

we have

$$\|\mathscr{F}_A - e_{2,A}\| \le (K_6 C_2 + K_7 C_3 + K_8) E_A^3. \tag{2.10}$$

2.11. REMARKS. (a) The conditions (2.5), (2.8) and (2.9) are analogous to those necessary and sufficient conditions which Bickel and Robinson (1982) used to establish Edgeworth expansions for the i.i.d. case. Following their arguments on page 502, one can easily deduce from (2.10) the condition (2.8) for $B = \hat{A}$ and from (2.6) the condition (2.5) for $B = \hat{A}$ (with new C_1, C_2).

However, note that these arguments do not show (2.8) [(2.5)] for general $B \in N(16, \hat{A})$ [$B \in N(8, \hat{A})$].

(b) It seems likely, at least to the present author, that Theorems 2.4 and 2.7 remain correct, if we assume (2.5), (2.8) and (2.9) only for $\hat{A}$ instead of $B \in N(8, \hat{A})$ [$B \in N(16, \hat{A})$]. However, a proof eludes me.

In the special case where $a_{ij} = e_i d_j$, we can replace (2.5), (2.8) and (2.9) by the conditions of van Zwet (1982). If we define

$$\bar{e} = \sum_i e_i/n, \qquad \bar{d} = \sum_j d_j/n, \qquad x^+ = x \vee 0 \quad \text{for } x \in \mathbb{R}$$

and write λ for the Lebesgue measure, we have

2.12. THEOREM. *Suppose that there exist positive numbers e, E, d, D and δ such that*

$$\sum_i |e_i - \bar{e}|^r \geq en, \qquad \sum_i |e_i - \bar{e}|^k \leq En \tag{2.13}$$
for some $k > 2$ and $0 < r < k$,

$$\sum_j |d_j - \bar{d}|^m \geq dn, \qquad \sum_j |d_j - \bar{d}|^s \leq Dn \tag{2.14}$$
for some $s > 2$ and $0 < m < s$,

$$\lambda\big(\{x\colon |x - d_j| < \zeta \text{ for some } 1 \leq j \leq n\}\big) \geq \delta n \zeta \tag{2.15}$$
for some $\zeta \geq n^{-3/2} \log n$.

(a) *Then there exist positive numbers $\mathscr{K}_1$ and $\mathscr{K}_2$ depending only on e, E, d, D, δ and r, k, m, s such that*

$$\begin{aligned} \|\mathscr{F}_A - e_{1,A}\| &\leq \mathscr{K}_1 (\log n)^2 D_A^2 \\ &\leq \mathscr{K}_2 (\log n)^2 n^{-1 + ((4/k)-1)^+ + ((4/s)-1)^+}. \end{aligned} \tag{2.16}$$

(b) *Let $\varepsilon > 0$. Then there exists positive numbers $\mathscr{K}_3$ and $\mathscr{K}_4$ depending only on e, E, d, D, δ, r, k, m, s and ε such that*

$$\begin{aligned} \|\mathscr{F}_A - e_{2,A}\| &\leq \mathscr{K}_3 n^{\varepsilon} E_A^3 \\ &\leq \mathscr{K}_4 n^{-(3/2) + \varepsilon + ((5/k)-1)^+ + ((5/s)-1)^+}. \end{aligned} \tag{2.17}$$

2.18. REMARKS. (a) For $[((5/k) - 1)^+ - ((4/k) - 1)^+] + [((5/s) - 1)^+ - ((4/s) - 1)^+] < \frac{1}{2}$ we can deduce the second estimation of $\|\mathscr{F}_A - e_{1,A}\|$ in (2.16) from (2.17). In this case the factor $(\log n)^2$ is superfluous.

(b) Let

$$d_j = E\big(J(U_{j:n})\big), \qquad j = 1, \dots, n \text{ (exact scores)},$$

where $J\colon (0,1) \to \mathbb{R}$ is an integrable function and $U_{j:n}$ denotes the jth order statistic in a random sample of size n from the uniform distribution on $(0,1)$.

If J is nonconstant, continuously differentiable and satisfies $\int_0^1 |J(t)|^s\, dt < \infty$ for some $s > 2$, then $n_e(J) = \sup\{m \in \mathbb{N}\colon \sum_1^m |d_j - \bar{d}| = 0\} < \infty$, and (2.14) (with this s!) and (2.15) are fulfilled for all $n > n_e(J)$ with constants d, D and δ depending only on J and s. For a proof see Schneller (1987), proof of Theorem 4.4.23(a), (b).

(c) Let

$$d_j = J\left(\frac{j}{n+1}\right), \qquad j = 1, \dots, n \text{ (approximate scores)},$$

where $J\colon (0,1) \to \mathbb{R}$ is a function.

If J is nonconstant, continuously differentiable and satisfies

$$|J'(t)| \leq \Gamma\big(t(1-t)\big)^{-1-(1/s)+\beta} \quad \text{for all } t \in (0,1), \tag{2.19}$$
for some $\Gamma > 0$, $s > 2$ and $0 < \beta < \dfrac{1}{s}$,

then $n_a(J) = \sup\{m \in \mathbb{N}: \sum_1^m |d_j - \bar{d}| = 0\} < \infty$, and (2.14) (with this s!) and (2.15) are fulfilled for all $n > n_a(J)$ with constants d, D and δ depending only on J, s and β. For a proof see Schneller (1987), proof of Theorem 4.4.13(a), (b).

Part (d) of this theorem of Schneller (1987) contains an extension of Does [(1983), Theorem 2.1]. Does uses expansions which are slightly different from ours. Roughly speaking, he uses in his expansions integrals of J whereas we use the corresponding Riemann-sums. It is shown in Schneller (1987), proof of Theorem 4.4.13(d) that, if we assume in addition $k \geq 4$ in (2.13) and $s \geq 4$ in (2.19), then the difference between these two expansions is $O(n^{-(3/2)+3((1/s)-\beta)})$. Combining this with the rate of (2.17), we obtain a result with a better convergence rate and with weaker assumptions for J and e_i than Does. He obtained the rate $o(n^{-1})$ and assumed especially for J that

$$\limsup_{t \to 0,1} t(1-t)|J''(t)/J'(t)| < 2 \quad \text{and} \quad |J'''(t)| \leq \bar{\Gamma}(t(1-t))^{-3-(1/14)+\beta}$$

with $\bar{\Gamma}$, t and β as in (2.19). But from the last inequality (2.19) follows with $s = 14$.

Finally, we remark that the constants introduced in this section remain fixed throughout the paper. In contrast $c, c_1, c_2, \dots$ denote positive constants which depend only on the formula where they appear.

3. Proof of the basic equation. In this and the next (the next but one) section we prove Theorem 2.1 (Theorem 2.4). For that we fix an $n \times n$ matrix A with $\sigma_A > 0$. Of course, we may assume $a_{ij} = \hat{a}_{ij}$. Furthermore, let $q: \mathbb{R} \to \mathbb{R}$ be a function which is assumed to be bounded and differentiable throughout this section.

The essence of Stein's method is that if we define

$$f(x) = (\Theta q)(x) = \psi(x)^{-1} \int_{-\infty}^{x} (q(y) - \Phi(q))\psi(y)\, dy \tag{3.1}$$

[$\Phi(q)$ denotes the standard normal expectation of q], then we obtain the differential equation

$$f'(x) - xf(x) = q(x) - \Phi(q), \qquad x \in \mathbb{R}, \tag{3.2}$$

and thus for any random variable ξ we have

$$E(q(\xi)) - \Phi(q) = E(f'(\xi)) - E(\xi f(\xi)). \tag{3.3}$$

Therefore, in order to estimate $E(q(\xi)) - \Phi(q)$, we can estimate the expression $E(f'(\xi)) - E(\xi f(\xi))$.

Assume for a moment that $\xi = S_n = n^{-1/2}(X_1 + \cdots + X_n)$ where $X_1, \dots, X_n$ are i.i.d. with $E(X_i) = 0$, $V(X_i) = 1$. Then we have

$$E(f'(S_n)) - E(S_n f(S_n)) = E(f'(S_n)) - \sqrt{n}\, E(X_n f(S_n)). \tag{3.4}$$

Define $S_{n-1}^n = n^{-1/2}(X_1 + \cdots + X_{n-1})$. Now, in order to prove the classical CLT or Berry–Esseen theorem, we have to carry out a Taylor expansion of first order of f about S_{n-1}^n in $E(X_n f(S_{n-1}^n + n^{-1/2}X_n))$ and then apply the independence of S_{n-1}^n and X_n [for more details see, e.g., Bolthausen (1984), Section 2]. For Edgeworth expansions we have to use Taylor expansions of higher order of f and f' [for more details see, e.g., Schneller (1987), Chapter 1, especially Section 3 or the related paper of Barbour (1986)].

However, in our case we have $\xi = T_A$ and therefore there is a priori no comparable independence. For that reason we use an extension of Bolthausen's combinatorial method. This method yields a "bit," but for us enough independence.

For an intuitive understanding of this method it is best to inspect Table 2 from right to left and consider the results of Lemma 3.5 (a)–(d) as well as (3.6)–(3.8). In order to prove our analogue of (3.4), we need the independence of π_4 and I_1. This is done in (3.14) where we obtain $E(T_A f(T_A)) = nE(a_{I_1J_1} f(T_4))$. For the next step we need the independence of $a_{I_1J_1}$ and a part of T_4. (I_1, J_1) and π_4 are *not* independent, but (I_1, J_1) and π_3 are, which is achieved through an "exchange" of J_1 and J_2. Thus $a_{I_1J_1}$ and the part T_3 of $T_4 = T_3 + (T_4 - T_3)$ are independent. For the next step we need the independence of $T_4 - T_3 = a_{I_1J_1} + a_{I_2J_2} - a_{I_1J_2} - a_{I_2J_1}$ and a part of T_3. (I_1, I_2, J_1, J_2) and π_3 are *not* independent, but (I_1, I_2, J_1, J_2) and π_2 are, which is achieved through an "exchange" of the blocks (J_1, J_2) and (J_3, J_4). Now it is clear that we find our last independence statement through an "exchange" of the blocks $(J_1, \dots, J_4)$ and $(J_5, \dots, J_8)$.

After these considerations we give the explicit construction which starts with the last step of the above considerations and ends with the first step of these considerations. Let

$$N = \{1, \dots, n\},$$

$$\begin{aligned} M = \{\underline{i} = (i_1, \dots, i_8) \in N^8\colon\ & \underline{i} \text{ satisfies the equivalences:} \\ & (t1) i_1 = i_2 \Leftrightarrow i_3 = i_4; \\ & (u1) i_1 = i_2 \Leftrightarrow i_7 = i_8; \\ & (u2) i_3 = i_4 \Leftrightarrow i_5 = i_6; \\ & (u3) i_l = i_k \Leftrightarrow i_{l+6} = i_{k+2} \text{ for } l = 1,2;\ k = 3,4\}. \end{aligned}$$

For each $\underline{i} \in M$ we fix once for all permutations $u(\underline{i})$, $t(\underline{i})$ and $s(\underline{i})$ of N with properties as described in Table 1.

We remark that the map $\underline{i} \to u(\underline{i})$ is well defined because of $(u1)$, $(u2)$, $(u3)$ and that the map $\underline{i} \to t(\underline{i})$ is well defined because of $(t1)$.

As the reader will see, we have not defined all values of $u(\underline{i})$ and $t(\underline{i})$ explicitly. The reason for that is that we do not need an explicit definition and that we would have to consider a lot of cases for the explicit definition of these values (e.g., $i_4 = i_5$, then $[u(\underline{i})](i_5) = i_6$; but $i_4 \neq i_5$, then $[u(\underline{i})](i_5) \neq i_6$).

Furthermore, we define a random element $\underline{I} = (I_1, \dots, I_8)$ that is uniformly distributed on M and a random permutation π_1 that is uniformly distributed on

TABLE 1
Definition of the permutations $u(\underline{i})$, $t(\underline{i})$ and $s(\underline{i})$

	$u(\underline{i})$	$t(\underline{i})$	$s(\underline{i})$
i_1 i_2	i_7 i_8	i_4 i_3	i_2 i_1
i_3 i_4	i_5 i_6	Values in $\{i_1, \dots, i_4\}$	
i_5 i_6 i_7 i_8	Values in $\{i_1, \dots, i_8\}$	Remain fixed	
$N \setminus \{i_1, \dots, i_8\}$			

$\mathscr{P}_n$ and independent of $\underline{I}$. Finally, we define

$$\pi_2 = \pi_1 \circ u(\underline{I}), \qquad \pi_3 = \pi_2 \circ t(\underline{I}), \qquad \pi_4 = \pi_3 \circ s(\underline{I}),$$
$$J_1 = \pi_1(I_5), \qquad J_2 = \pi_1(I_6), \qquad J_3 = \pi_1(I_7), \qquad J_4 = \pi_1(I_8),$$
$$J_5 = \pi_1(I_1), \qquad J_6 = \pi_1(I_2), \qquad J_7 = \pi_1(I_3), \qquad J_8 = \pi_1(I_4).$$

Using the definition of u, t and s, we see in Table 2 how $\pi_1, \dots, \pi_4$ map $I_1, \dots, I_8$. As usual we define that $\sigma(X)$ is the σ-field generated by the random vector X and $f \in \sigma(X)$ means that f is measurable relative to $\sigma(X)$.

Further results are:

3.5. LEMMA. (a) *π_1, π_2, π_3 and π_4 have the same law and are independent of $\underline{I}$.*

(b) *π_1 and $(I_1, \dots, I_4, J_1, \dots, J_4)$ are independent.*

(c) *π_2 and (I_1, I_2, J_1, J_2) are independent.*

(d) *π_3 and (I_1, J_1) are independent.*

(e) *$(I_l, \pi_k(I_l))$ is uniformly distributed on N^2 for all $1 \le l \le 8$, $1 \le k \le 4$.*

PROOF. (a) For all $\pi \in \mathscr{P}_n$ and $\underline{i} \in M$ we have

$$P(\pi_2 = \pi, \underline{I} = \underline{i}) = P\big(\pi_1 = \pi \circ u(\underline{i})^{-1}, \underline{I} = \underline{i}\big) = \frac{1}{n!} P(\underline{I} = \underline{i}).$$

Thus summation over all $\underline{i} \in M$ gives first the law of π_2 and then the independence of π_2 and $\underline{I}$. The assertion for π_3 and π_4 follows analogously.

(b) For all $\pi \in \mathscr{P}_n$ and $\underline{i} \in M$ we have

$$\begin{aligned} &P((I_1, \dots, I_4, J_1, \dots, J_4) = \underline{i}, \pi_1 = \pi) \\ &\quad = P\big(\underline{I} = \big(i_1, \dots, i_4, \pi^{-1}(i_5), \dots, \pi^{-1}(i_8)\big)\big) P(\pi_1 = \pi) \\ &\quad = P(\underline{I} = \underline{i}) P(\pi_1 = \pi). \end{aligned}$$

TABLE 2
Values of $I_1, \dots, I_8$ under $\pi_1, \pi_2, \pi_3, \pi_4$

	π_1	π_2	π_3	π_4
I_1	J_5	J_3	J_2	J_1
I_2	J_6	J_4	J_1	J_2
I_3	J_7	J_1	Random variables $\in \sigma(I_1, \dots, I_4, J_1, \dots, J_4)$	Same as π_3
I_4	J_8	J_2		
I_5	J_1	Random variables $\in \sigma(I_1, \dots, I_8, J_1, \dots, J_8)$	Same as π_2	Same as π_2
I_6	J_2			
I_7	J_3			
I_8	J_4			

Now summation over all $\pi \in \mathscr{P}_n$ gives that $(I_1, \dots, J_4)$ and $\underline{I}$ have the same law, from which the assertion follows.

(c), (d) Cf. proof of (b) and Table 2.

(e) From $\underline{i} \in M \Leftrightarrow \underline{i} + (1,1,\dots,1) \in M \pmod n$ we conclude that $I_1, \dots, I_8$ are uniformly distributed over N. Thus (e) follows easily using (a). □

Next, we define

$$T_k = \sum_j a_{j\pi_k(j)} \quad \text{for } k = 1,2,3,4,$$

$$\Delta T_k = T_{k+1} - T_k \quad \text{for } k = 1,2,3.$$

From Table 2 and the definition of $u(\underline{i})$, $t(\underline{i})$ and $s(\underline{i})$ we obtain

$$\Delta T_1 = \sum_{l=1}^{8} \left(a_{I_l \pi_2(I_l)} - a_{I_l \pi_1(I_l)}\right) \in \sigma(I_1, \dots, I_8, J_1, \dots, J_8), \tag{3.6}$$

$$\Delta T_2 = \sum_{l=1}^{4} \left(a_{I_l \pi_3(I_l)} - a_{I_l \pi_2(I_l)}\right) \in \sigma(I_1, \dots, I_4, J_1, \dots, J_4), \tag{3.7}$$

$$\Delta T_3 = a_{I_1 J_1} + a_{I_2 J_2} - a_{I_1 J_2} - a_{I_2 J_1} \in \sigma(I_1, I_2, J_1, J_2). \tag{3.8}$$

Furthermore, a simple calculation gives

$$E(a_{I_1 J_1}) = 0, \qquad nE(a_{I_1 J_1}\,\Delta T_3) = 1, \tag{3.9}$$

$$n\left\{E(a_{I_1 J_1}\,\Delta T_3\,\Delta T_2) + E\left(a_{I_1 J_1}(\Delta T_3)^2/2\right)\right\} = E(T_A^3)/2. \tag{3.10}$$

Now we are able to prove the following basic equation.

3.11. LEMMA. *Let f and q be connected as in* (3.1). *Then*

$$E(q(T_A)) - \Phi(q) + \tfrac{1}{2}E(T_A^3)E(T_A f'(T_A)) = R(q), \tag{3.12}$$

where

$$R(q) = \tfrac{1}{2}E(T_A^3)nE\left(a_{I_1J_1}\,\Delta T_3\int_0^1 (f''(T_2+\Delta T_2+t\,\Delta T_3)-f''(T_2))\,dt\right)$$
$$-nE\left(a_{I_1J_1}\,\Delta T_3\,\Delta T_2\int_0^1 (f''(T_1+\Delta T_1+t\,\Delta T_2)-f''(T_1))\,dt\right)$$
$$-nE\left(a_{I_1J_1}(\Delta T_3)^2\int_0^1 (1-t)(f''(T_2+\Delta T_2+t\,\Delta T_3)-f''(T_2))\,dt\right).$$

PROOF. The equation (3.3) with $\xi = T_A$ gives

$$E(q(T_A)) - \Phi(q) = E(f'(T_A)) - E(T_A f(T_A)). \tag{3.13}$$

Now, using the independence of π_4 and I_1 we obtain

$$E(T_A f(T_A)) = E(T_4 f(T_4)) = nE(a_{I_1J_1} f(T_4)). \tag{3.14}$$

A Taylor expansion of f about T_3 yields further

$$= nE(a_{I_1J_1} f(T_3)) + nE(a_{I_1J_1}\,\Delta T_3\, f'(T_3))$$
$$+nE\left(a_{I_1J_1}(\Delta T_3)^2\int_0^1 (1-t)(f''(T_2+\Delta T_2+t\,\Delta T_3)-f''(T_2))\,dt\right)$$
$$+\frac{n}{2}E\left(a_{I_1J_1}(\Delta T_3)^2 f''(T_2)\right).$$

The first summand is zero [cf. Lemma 3.5(d), (3.9)] and the last summand gives [cf. Lemma 3.5(c), (3.8)]

$$\frac{n}{2}E\left(a_{I_1J_1}(\Delta T_3)^2 f''(T_2)\right) = \frac{n}{2}E\left(a_{I_1J_1}(\Delta T_3)^2\right)E(f''(T_2)).$$

For the second summand we have

$$nE(a_{I_1J_1}\,\Delta T_3\, f'(T_3))$$
$$= nE(a_{I_1J_1}\,\Delta T_3\, f'(T_2)) + nE(a_{I_1J_1}\,\Delta T_3\,\Delta T_2\, f''(T_1))$$
$$+nE\left(a_{I_1J_1}\,\Delta T_3\,\Delta T_2\int_0^1 (f''(T_1+\Delta T_1+t\,\Delta T_2)-f''(T_1))\,dt\right)$$
$$= E(f'(T_2)) + nE(a_{I_1J_1}\,\Delta T_3\,\Delta T_2)E(f''(T_1)) + \text{last summand}.$$

[Cf. Lemma 3.5(b), (c) and (3.7)–(3.9).]

Thus, using (3.10) we conclude

$$\begin{aligned} E(T_A f(T_A)) &= E(f'(T_A)) + \tfrac{1}{2}E(T_A^3)E(f''(T_A)) \\ &\quad + \text{second term of } R(q) + \text{third term of } R(q). \end{aligned} \tag{3.15}$$

The degree of differentiation of f in the term $E(f''(T_A))$ will now be reduced by the following consideration:

$$
\begin{aligned}
E(T_A f'(T_A)) &= nE\big(a_{I_1J_1} f'(T_4)\big) \\
&= nE\big(a_{I_1J_1} f'(T_3)\big) + nE\big(a_{I_1J_1}\, \Delta T_3\, f''(T_2)\big) \\
&\quad + nE\left(a_{I_1J_1}\, \Delta T_3 \int_0^1 \big(f''(T_2 + \Delta T_2 + t\,\Delta T_3) - f''(T_2)\big)\, dt\right) \\
&= E(f''(T_2)) \\
&\quad + nE\left(a_{I_1J_1}\, \Delta T_3 \int_0^1 \big(f''(T_2 + \Delta T_2 + t\,\Delta T_3) - f''(T_2)\big)\, dt\right).
\end{aligned}
$$

Implementing this in (3.15) and using (3.13) we obtain the lemma. □

4. Proof of Theorem 2.1. In addition to the assumptions of the last section ($a_{ij} = \hat{a}_{ij}$ for all i, j and q is bounded, differentiable), we have to make some further assumptions.

First, we assume that $q \in \mathscr{D}$; see (2.2). Furthermore, we may assume $\beta_A \le \varepsilon_0 n$ and $n \ge n_0$ for arbitrary but fixed $0 < \varepsilon_0 \le 1$ and $n_0 \ge 9$. These constants will be specified later in Lemma 4.1. For $\beta_A > \varepsilon_0 n$ or $n < n_0$ we obtain (2.3) from

$$|E(q(\mathscr{T}_A)) - \int q e'_{1,A}\, dx| \le \|q\|(2 + \beta_A/n)$$

and $(\beta_A/n)^2 \le D_A^2$, $0 < c \le \beta_A$, if we take K_1 large enough.

Moreover, we must assume $|a_{ij}| \le 1$ for all i, j. We do not show here how this truncation is established and refer the reader to Schneller (1987), Chapter 3, Section 4. For the basic ideas of this truncation one may also consult Bolthausen (1984), pages 382 and 383. We mention that for this truncation we eventually have to reduce the above ε_0 and increase the above n_0.

We divide the following proof of Theorem 2.1 in two parts. The first part is

4.1. LEMMA. *There exist positive numbers c_1, c_2 and c_3 such that*

$$|R(q)| \le (c_1\|q\| + c_2\|q'\| + c_3\|q''\|) D_A^2.$$

PROOF. Let $\bar{I} = (I_1, \dots, I_4)$ and $\bar{i} = (i_1, \dots, i_4)$. Furthermore, let $\underline{J}$, $\bar{J}$, $\underline{j}$, $\bar{j}$ be defined analogously to $\underline{I}$, $\bar{I}$, $\underline{i}$, $\bar{i}$. We have to show that there exist constants $c_4, \dots, c_9$ such that

$$
\begin{aligned}
&\Big|E\big(f''(T_1 + \Delta T_1 + t\,\Delta T_2) - f''(T_1) | \underline{I} = \underline{i},\ \underline{J} = \underline{j}\big)\Big| \\
&\qquad \le (c_4\|q\| + c_5\|q'\| + c_6\|q''\|) E\big(|\Delta T_1| + |\Delta T_2| \,\big|\, \underline{I} = \underline{i},\ \underline{J} = \underline{j}\big) \\
&\qquad \text{for all } 0 \le t \le 1 \quad \text{and} \quad \underline{i}, \underline{j} \in N^8 \quad \text{with } P\big(\underline{I} = \underline{i},\ \underline{J} = \underline{j}\big) > 0,
\end{aligned} \tag{4.2}
$$

$$
\begin{aligned}
&\Big|E\big(f''(T_2 + \Delta T_2 + t\,\Delta T_3) - f''(T_2) | \bar{I} = \bar{i},\ \bar{J} = \bar{j}\big)\Big| \\
&\qquad \le (c_7\|q\| + c_8\|q'\| + c_9\|q''\|) E\big(|\Delta T_2| + |\Delta T_3| \,|\, \bar{I} = \bar{i},\ \bar{J} = \bar{j}\big) \\
&\qquad \text{for all } 0 \le t \le 1 \quad \text{and} \quad \bar{i}, \bar{j} \in N^4 \quad \text{with } P\big(\bar{I} = \bar{i},\ \bar{J} = \bar{j}\big) > 0.
\end{aligned} \tag{4.3}
$$

Since the proofs of (4.2) and (4.3) are very similar, we only show (4.2). For that we fix the quantities $\underline{i}$, $\underline{j}$ and t.

Now we look at the conditional distribution of T_1 given $\underline{I} = \underline{i}$, $\underline{J} = \underline{j}$. T_1 depends only on π_1 and the conditional distribution of π_1 is easy to describe: π_1 takes any permutation φ which satisfies $\varphi(i_k) = j_{k+4}$ for $1 \le k \le 4$ and $\varphi(i_k) = j_{k-4}$ for $5 \le k \le 8$ with equal probability. Therefore, T_1 given $\underline{I} = \underline{i}$, $\underline{J} = \underline{j}$ has the same law as

$$\sum_{(i,j) \in S} a_{ij} + T_B,$$

where

$$S = \{(i_k, j_{k+4}) \colon 1 \le k \le 4\} \cup \{(i_k, j_{k-4}) \colon 5 \le k \le 8\}$$

and B is the $(n-l) \times (n-l)$ matrix which is obtained from A by cancelling the rows $i_1, \dots, i_8$ and the columns $j_1, \dots, j_8$. Using the notation

$$a = \sum_{(i,j) \in S} a_{ij}, \qquad p = E\left(\Delta T_1 | \underline{I} = \underline{i}, \underline{J} = \underline{j}\right),$$

$$r = E\left(\Delta T_2 | \underline{I} = \underline{i}, \underline{J} = \underline{j}\right)$$

[note $\Delta T_1, \Delta T_2 \in \sigma(\underline{I}, \underline{J})!$], it remains to show

$$\begin{aligned} &|E(f''(T_B + a + p + tr) - f''(T_B + a))| \\ &\quad \le (c_{10}\|q\| + c_{11}\|q'\| + c_{12}\|q''\|)(|p| + |r|). \end{aligned} \tag{4.4}$$

In order to prove (4.4) we need the estimates

$$\|f\| \le 4\|q\| \quad \text{and} \quad \|f'\| \le 4\|q\| \tag{4.5}$$

[see Erickson (1974), Proposition 2.1]. Moreover, differentiation of (3.2) gives

$$f''(x) = f(x) + xf'(x) + q'(x), \qquad x \in \mathbb{R}. \tag{4.6}$$

From (4.6) and (4.5) we obtain

$$\begin{aligned} |f''(x+y) - f''(x)| &\le |y| 8\|q\| + |x|\,|f'(x+y) - f'(x)| \\ &\quad + |q'(x+y) - q'(x)|. \end{aligned} \tag{4.7}$$

Using (3.2), (4.5) and $|x| \le 1 + x^2$ we estimate further

$$\begin{aligned} &\le |y| 12\|q\|\{1 + x^2\} + |x|\,|q(x+y) - q(x)| + |q'(x+y) - q'(x)| \\ &\le |y|(12\|q\| + \|q'\| + \|q''\|)\{1 + x^2\}. \end{aligned} \tag{4.8}$$

Therefore, we have

$$\begin{aligned} &|E(f''(T_B + a + p + tr) - f''(T_B + a))| \\ &\quad \le (12\|q\| + \|q'\| + \|q''\|)(|p| + |r|)\left\{1 + E\left((T_B + a)^2\right)\right\}. \end{aligned}$$

It remains to estimate $E((T_B + a)^2)$. If we take the $0 < \varepsilon_0 \le 1$ in $\beta_A \le \varepsilon_0 n$ small enough and the n_0 in $n \ge n_0$ great enough, we can deduce as in Bolthausen

(1984), page 385 after (3.11),

$$(4.9) \qquad |\mu_B| \leq 1, \qquad \tfrac{1}{2} \leq \sigma_B^2 \leq \tfrac{3}{2} \quad \text{and} \quad \beta_B \leq c\beta_A \quad \text{for } B \in N(l, A).$$

[The last inequality is needed in (5.4).] Using $|a_{ij}| \leq 1$ we find further $|a| \leq 8$. Therefore, we obtain

$$(4.10) \qquad E\big((T_B + a)^2\big) \leq 2E\big(T_B^2\big) + 2a^2 = 2\big(\sigma_B^2 + \mu_B^2\big) + 2a^2 \leq c. \qquad \square$$

In the second part of this section we prove

4.11. LEMMA. *There exist positive numbers c_1 and c_2 such that*

$$\left|\tfrac{1}{2}E(T_A)E(T_A f'(T_A)) - \left(\tfrac{1}{2}\sum_{i,j} a_{ij}^3\right)\Phi(xf'(x))\right| \leq (c_1\|q\| + c_2\|q'\|)D_A^2.$$

We remark that an easy calculation using (3.2) shows

$$\Phi(xf'(x)) = \tfrac{1}{3}\int q(x)(3x - x^3)\psi(x)\,dx,$$

so that on the left-hand side of the basic equation (3.12) we have $E(q(T_A)) - \int qe'_{1,A}\,dx$ up to a term smaller than a constant times D_A^2.

PROOF OF LEMMA 4.11. We have

$$\left|E(T_A^3)E(T_A f'(T_A)) - \left(\frac{1}{n}\sum_{i,j} a_{ij}^3\right)\Phi(xf'(x))\right|$$

$$\leq E(|T_A f'(T_A)|)\left|E(T_A^3) - \frac{1}{n}\sum_{i,j} a_{ij}^3\right| + \frac{1}{n}\beta_A|E(T_A f'(T_A)) - \Phi(xf'(x))|$$

$$= A_1 + A_2.$$

Because $E(T_A^3) = n[(n-1)(n-2)]^{-1}\Sigma_{i,j}a_{ij}^3$ (see Hájek and Šidák (1967), Chapter 2, page 82, problem 27) and (4.5) we obtain

$$A_1 \leq c_3\|q\|\beta_A/n^2 \leq c_4\|q\|\beta_A^2/n^2 \leq c_4\|q\|D_A^2.$$

For the estimation of A_2 we show the inequality

$$(4.12) \qquad |E(T_A f'(T_A)) - \Phi(xf'(x))| \leq (c_5\|q\| + c_6\|q'\|)\beta_A/n.$$

In order to prove (4.12) we define $\mathscr{f}(x) = (\Theta\varphi)(x)$ where $\varphi(x) = xf'(x)$. Proceeding as in Erickson (1974) and using (4.5) we find $\|\mathscr{f}\| \leq 3\|f'\| \leq 12\|q\|$. From this and (3.2) for $\mathscr{f}$ we conclude further $|\mathscr{f}'(x)| \leq |x|16\|q\| + 4\|q\|$ for all $x \in \mathbb{R}$. Thus, using the estimation

$$|(x+y)f'(x+y) - xf'(x)| \leq |y|(8\|q\| + \|q'\|)\{1 + x^2\}, \qquad x, y \in \mathbb{R}$$

[cf. proof of (4.7) and (4.8)], and again (3.2) for $\mathscr{f}$, we have

$$(4.13) \qquad |\mathscr{f}'(x+y) - \mathscr{f}'(x)| \leq |y|(c_7\|q\| + \|q'\|)(1 + x^2)(1 + |y|).$$

Now we obtain (4.12) if we proceed first of all as in Bolthausen (1984), page 383 bottom and page 384 [with (4.13) instead of Bolthausen's (2.5)] and then argue as in our proof of Lemma 4.1. Note that Bolthausen's $|\Delta T_1|$ and $|\Delta T_2|$ are bounded since we have $|a_{ij}| \le 1$ for all i, j. □

5. Proof of Theorem 2.4. As in the last section we may assume $a_{ij} = \hat{a}_{ij}$ and $|a_{ij}| \le 1$ for all i, j; $\beta_A \le \varepsilon_0 n$ and $n \ge n_0$ for fixed $0 < \varepsilon_0 \le 1$ and $n_0 \in \mathbb{N}$.

In order to apply the results of Section 3, we must replace the discontinuous functions $1_{(-\infty, z]}$, $z \in \mathbb{R}$, in $\|F_A - e_{1,A}\|$ by functions q_z, $z \in \mathbb{R}$, that are bounded and differentiable. We define for $z \in \mathbb{R}$,

$$q_z(x) = \begin{cases} 1, & \text{for } x \le z, \\ 1 - \frac{1}{2}((x - z)/D_A)^2, & \text{for } z \le x \le z + D_A, \\ \frac{1}{2}((z + 2D_A - x)/D_A)^2, & \text{for } z + D_A \le x \le z + 2D_A, \\ 0, & \text{for } z + 2D_A \le x. \end{cases}$$

Note that we have

$$\text{(5.1)} \quad q_z'(x + y) - q_z'(x) = y\left(\frac{1}{D_A^2} \int_0^1 \left(1_{(z+D_A, z+2D_A]} - 1_{(z, z+D_A]}\right)(x + sy)\, ds \right)$$

for all $x, y, z \in \mathbb{R}$. In this section (5.1) together with (2.5) will play the role of $|q'(x + y) - q'(x)| \le |y| \, \|q''\|$ of the last section.

The further use of q_z instead of $1_{(-\infty, z]}$ will be justified in the following lemma. This lemma needs (2.5) (but only for $B = \hat{A}$) for the first time.

5.2. Lemma.

$$\|F_A - e_{1,A}\| \le \sup_{z \in \mathbb{R}} \left| E(q_z(T_A)) - \int q_z e_{1,A}'\, dx \right| + (C_1 + 1) D_A^2.$$

Proof. We use the abbreviation $D = D_A$. Then we have

$$|F_A(z) - E(q_{z-D}(T_A))| = \left| \int_{(z-D, z]} \frac{1}{2} \left(\frac{x - (z - D)}{D} \right)^2 F_A(dx) - \int_{(z, z+D]} \frac{1}{2} \left(\frac{x - (z + D)}{D} \right)^2 F_A(dx) \right|.$$

Partial integration of each integral gives

$$= \left| F_A(z) - \int_{z-D}^{z} \frac{x - (z - D)}{D^2} F_A(x)\, dx + \int_z^{z+D} \frac{x - (z + D)}{D^2} F_A(x)\, dx \right|.$$

Substitution of $y = z - x$ ($y = x - z$) in the first (second) integral leads to

$$= \left| \int_0^D \frac{D-y}{D^2} \Delta_y^2 F_A(z-y)\, dy \right|$$
$$\le \int_0^D \frac{D-y}{D^2} C_1(D^2 + y^2)\, dy \le C_1 D^2.$$

Similar computations give

$$\left| e_{1,A}(z) - \int q_{z-D} e'_{1,A}\, dx \right| = \left| \int_0^D \frac{D-y}{D^2} \Delta_y^2 e_{1,A}(z-y)\, dy \right|.$$

Now, using $\|\Delta_y^2 e_{1,A}\| \le y^2 \|e''_{1,A}\| \le y^2(1 + \beta_A/n)$ and $\beta_A/n \le \varepsilon_0 \le 1$, we obtain the lemma. □

The rest of the proof of Theorem 2.4 has many parallels to that of Theorem 2.1.

5.3. LEMMA. *There exist positive numbers c_1 and c_2 such that*

$$\sup_{z \in \mathbb{R}} |R(q_z)| \le (c_1 C_1 + c_2) D_A^2.$$

PROOF. We fix $z \in \mathbb{R}$. Therefore, for simplicity we drop the index z of q_z and f_z and the index A of D_A.

We adopt the proof of Lemma 4.1 up to (4.8) (with $c.\|q\| + c.\|q'\| + c.\|q''\|$ substituted by $c.C_1 + c.$). From (4.8), (5.1) and $|q'(x)| \le (1/D)1_{(z,z+2D]}$ we conclude

$$\begin{aligned}
&|E(f''(T_B + a + p + tr) - f''(T_B + a))| \\
&\quad \le 12(|p| + |r|)\Big\{1 + E\big((T_B + a)^2\big) \\
&\qquad + \frac{1}{D}\int_0^1 E\big(|T_B + a| 1_{(z,z+2D]}(T_B + a + sp + str)\big)\, ds \\
&\qquad + \frac{1}{D^2}\int_0^1 \big|E\big((1_{(z+D,z+2D]} - 1_{(z,z+D]})(T_B + a + sp + str)\big)\big|\, ds\Big\} \\
&\quad = 12(|p| + |r|)\{1 + A_1 + A_2 + A_3\}.
\end{aligned}$$

A_1 is estimated in (4.10), so that it remains to estimate A_2 and A_3.

For A_2 we use (1.1), the following Proposition 5.7, $|a| \le 8$ and (4.9):

$$\begin{aligned}
&E\big(|T_B + a| 1_{(z,z+2D]}(T_B + a + sp + str)\big) \\
&\quad = E\big(|\sigma_B \mathcal{T}_B + \mu_B + a| 1_{(z,z+2D]}(\sigma_B \mathcal{T}_B + \mu_B + a + sp + str)\big) \qquad (5.4)\\
&\quad \le c_3\big((\beta_B/(n-l)) + D\big) \\
&\quad \le c_4 D.
\end{aligned}$$

This yields $A_2 \le c_4$. Finally, (2.5) gives

$$\left|E\left(\left(1_{(z+D,\, z+2D]} - 1_{(z,\, z+D]}\right)(T_B + a + sp + str)\right)\right| \le \|\Delta_D^2 F_B\| \le 2C_1 D^2$$

and so we have $A_3 \le 2C_1$. This proves the lemma. □

In order to complete the proof of Theorem 2.4 we need

5.5. LEMMA. *There exists a positive constant c such that*

$$\sup_{z \in \mathbb{R}} \left| \tfrac{1}{2} E(T_A) E(T_A f_z'(T_A)) - \left(\tfrac{1}{2} \sum_{i,j} a_{ij}^3 \right) \Phi(x f_z'(x)) \right| \le c D_A^2.$$

PROOF. If we proceed as in the proof of Lemma 4.11, we see that it remains to prove the following analogue of (4.12):

$$\sup_{z \in \mathbb{R}} |E(T_A f_z'(T_A)) - \Phi(x f_z'(x))| \le c\beta_A / n. \tag{5.6}$$

Using (3.2) and the fact that the functions $q_z(x)$ and $\Phi(q_z) - x f_z(x)$ are monotone decreasing between 0 and 1, we can deduce (5.6) from the following general result which is interesting in its own right and which was already needed for proving (5.4).

5.7. PROPOSITION. *There exists a positive number c such that for all A with* $\sigma_A > 0$,

$$\sup_{z \in \mathbb{R}} \left|E\left(|\mathscr{T}_A| 1_{(-\infty, z]}(\mathscr{T}_A)\right) - \Phi\left(|x| 1_{(-\infty, z]}(x)\right)\right| \le c\beta_A / n.$$

SKETCH OF PROOF. Proceed as in Bolthausen (1984) with the quantities

$$h(x) = h_{z,\lambda}(x) = |x| \{((1 + (z - x)/\lambda) \wedge 1) \vee 0\},$$
$$f(x) = f_{z,\lambda}(x) = (\Theta h_{z,\lambda})(x)$$

and note that Bolthausen's T_A is our $\mathscr{T}_A$.

We make two remarks. First, as in the proof of (4.12) we find positive numbers c_1, c_2, c_3 with $\|f\| \le c_1$ and $|f'(x)| \le c_2|x| + c_3$ for all $x \in \mathbb{R}$. Instead of Bolthausen's (2.5) we, therefore, have to use

$$|f'(x+y) - f'(x)| \le |y| \left\{ c_4(1 + x^2)(1 + |y|) + \frac{|x|}{\lambda} \int_0^1 1_{(z,\, z+\lambda]}(x + sy)\, ds \right\}.$$

Our second remark concerns Bolthausen's (3.6). In order to prove our analogue we define $a'_{ij} = \hat{a}_{ij} 1_{\{|\hat{a}_{ij}| \le 1/2\}}(\hat{a}_{ij})$, $\bar{a}_{ij} = \hat{a}'_{ij}$ and $\Gamma = \{(i, j) : |\hat{a}_{ij}| > \frac{1}{2}\}$. We need the estimation

$$\left|E\left(|\mathscr{T}_A| 1_{(-\infty, z]}(\mathscr{T}_A)\right) - E\left(|T_{\bar{A}}| 1_{(-\infty, z]}(T_{A'})\right)\right| \le c_5 \beta_A / n. \tag{5.8}$$

From this we obtain the following analogue of Bolthausen's (3.6):

$$\begin{aligned}
&\sup_z |E(|\mathscr{T}_A| 1_{(-\infty, z]}(\mathscr{T}_A)) - \Phi(|x| 1_{(-\infty, z]}(x))| \\
&\quad \le \sup_z |E(|T_{\bar{A}}| 1_{(-\infty, z]}(T_{A'})) - \Phi(|x| 1_{(-\infty, z]}(x))| + c_5 \beta_A / n \\
&\quad \le \sup_z |E(|T_{\bar{A}}| 1_{(-\infty, z]}(T_{\bar{A}})) - \Phi(|x| 1_{(-\infty, z]}(x))| \\
&\qquad + \sup_z |\Phi(|x| 1_{(-\infty, (z-\mu_{A'})/\sigma_{A'}]}(x)) - \Phi(|x| 1_{(-\infty, z]}(x))| + c_5 \beta_A / n.
\end{aligned}$$

Clearly, the second summand is of order β_A/n [cf. Bolthausen's (3.2) and (3.3)] and therefore our analogue is established with the exception of (5.8):

$$\begin{aligned}
&|E(|\mathscr{T}_A| 1_{(-\infty, z]}(\mathscr{T}_A)) - E(|T_{\bar{A}}| 1_{(-\infty, z]}(T_{A'}))| \\
&\quad \le |E(|\mathscr{T}_A| 1_{(-\infty, z]}(\mathscr{T}_A)) - E(|T_{\bar{A}}| 1_{(-\infty, z]}(\mathscr{T}_A))| \\
&\qquad + |E(|T_{\bar{A}}| 1_{(-\infty, z]}(\mathscr{T}_A)) - E(|T_{\bar{A}}| 1_{(-\infty, z]}(T_{A'}))| \\
&\quad \le E(|\mathscr{T}_A - T_{\bar{A}}|) + E(|T_{\bar{A}}| 1_{\{\mathscr{T}_A \ne T_{A'}\}}) = B_1 + B_2.
\end{aligned}$$

Furthermore, we have

$$\begin{aligned}
B_1 &\le E(|\mathscr{T}_A - T_{A'}|) + E(|\sigma_{A'} T_{\bar{A}} + \mu_{A'} - T_{\bar{A}}|) \\
&\le \sum_i E(|\hat{a}_{i\pi(i)} - a'_{i\pi(i)}|) + c_6 \beta_A / n \\
&= (1/n) \sum_{(i, j) \in \Gamma} |\hat{a}_{ij}| + c_6 \beta_A / n \le c_7 \beta_A / n, \\
B_2 &\le \sum_i E(|T_{\bar{A}}| 1_\Gamma(i, \pi(i))) \\
&= \sum_{i, j} (1/n) 1_\Gamma(i, j) E(|T_{\bar{A}}| \mid \pi(i) = j).
\end{aligned}$$

If we argue as in Bolthausen (1984), page 385, lines 3–14 (cf. also our proof of Lemma 4.1), we find a constant c_8 with $E(|T_{\bar{A}}| \mid \pi(i) = j) \le c_8$ for all i, j. Therefore, we have $B_2 \le c_8 |\Gamma| / n \le 8 c_8 \beta_A / n$. □

5.9. Remark to Proposition 5.7. In Schneller (1987), Chapter 3, Section 6, a complete, but a bit different proof of Proposition 5.7 may be found for the case $|\hat{a}_{ij}| \le 1$.

6. Proof of Theorem 2.12(a). In order to prove Theorem 2.12(a) we use a result of van Zwet [(1982), Theorem 2.1], which gives an estimation of the characteristic function of $\mathscr{T}_A$ under the conditions (2.13)–(2.15). From this we will deduce (2.5) with $C_1 \approx (\log n)^2$.

We need some preliminaries. Let U be a distribution function on $\mathbb{R}$ which has a density u that is infinitely differentiable and has a support which is contained in $[-1,1]$. It follows [cf. Feller (1971), Chapter 15, Section 4, Lemma 4] that

$$(6.1)\qquad |\hat{U}(t)| = o(|t|^{-n}) \quad \text{for } |t| \to \infty \text{ and all } n \in \mathbb{N},$$

where we denote by $\hat{G}$ the characteristic function of a distribution with distribution function G.

From (6.1) we conclude

$$(6.2)\qquad \int_{\mathbb{R}} |t|^n |\hat{U}(t)|\,dt < \infty, \qquad \int_{\mathbb{R}} |\hat{U}(t)|^n\,dt < \infty \quad \text{for all } n \in \mathbb{N}.$$

Using this we can prove

6.3. LEMMA. *Let F be a distribution function on $\mathbb{R}$ and*

$$U_\theta(x) = U(x/\theta) \quad \text{for } x \in \mathbb{R} \text{ and } \theta > 0.$$

Then

$$(6.4)\qquad |\Delta_y^2 F(z)| \le (y^2 + \theta) \int_{\mathbb{R}} (1 + |t|) |\hat{F}(t) \hat{U}_\theta(t)|\,dt$$

for all $z \in \mathbb{R}$, $y > 0$ and $\theta > 0$.

PROOF. We use a technique introduced by von Bahr (1967), Section 3. Let F_θ be the convolution of F and U_θ. Then F_θ has a density f_θ and

$$F_\theta(x - \theta) \le F(x) \le F_\theta(x + \theta) \quad \text{for all } x \in \mathbb{R}, \theta > 0.$$

Using this and the Plancherel identity we obtain for $y \ge 2\theta$,

$$\begin{aligned}
&F(z + 2y) - 2F(z + y) + F(z)\\
&\quad \le F_\theta(z + 2y + \theta) - 2F_\theta(z + y - \theta) + F_\theta(z + \theta)\\
&\quad \le \left| \int_{\mathbb{R}} \left(1_{(z+y-\theta,\, z+2y+\theta]}(t) - 1_{(z+\theta,\, z+y-\theta]}(t) \right) f_\theta(t)\,dt \right|\\
&\quad = \frac{1}{2\pi} \left| \int_{\mathbb{R}} \left(\int_{z+y-\theta}^{z+2y+\theta} e^{-its}\,ds - \int_{z+\theta}^{z+y-\theta} e^{-its}\,ds \right) \hat{F}_\theta(t)\,dt \right|\\
&\quad \le \frac{1}{2\pi} \int_{\mathbb{R}} \frac{1}{|t|} |e^{-it(z+2y)} - 2e^{-it(z+y)} + e^{-itz}|\, |\hat{F}_\theta(t)|\,dt\\
&\qquad + \theta \int_{\mathbb{R}} |\hat{F}_\theta(t)|\,dt\\
&\quad \le y^2 \int_{\mathbb{R}} |t|\, |\hat{F}_\theta(t)|\,dt + \theta \int_{\mathbb{R}} |\hat{F}_\theta(t)|\,dt\\
&\quad \le (y^2 + \theta) \int_{\mathbb{R}} (1 + |t|) |\hat{F}(t) \hat{U}_\theta(t)|\,dt =: R.
\end{aligned}$$

For the last inequality we used $\hat{F}_\theta = \hat{F}\hat{U}_\theta$. In the case $y < 2\theta$ we proceed similarly. Furthermore, we obtain $-(F(z + 2y) - 2F(z + y) + F(z)) \le R$ by completely analogous arguments. □

The assumption (2.5) contains conditions for $B \in N(8, \hat{A})$. For this reason we need the estimation of van Zwet not only for $\hat{\mathscr{F}}_A$ but also for $\hat{F}_B$, $B \in N(8, \hat{A})$.

6.5. LEMMA. *Let $A = (e_i d_j)$ be an $n \times n$ matrix fulfilling the conditions* (2.13)–(2.15) *and $n \geq n_1 = \max\{n_0, 10, 32/\delta\}$, $\beta_A \leq \varepsilon_0 n$, where n_0 and ε_0 are chosen such that $\frac{1}{2} \leq \sigma_B^2 \leq \frac{3}{2}$ holds for $B \in N(8, \hat{A})$ [cf. (4.9)].*

Then there exist positive numbers b_1, b_2, b_3 and b_4 depending only on e, E, d, D, δ and r, k, m, s such that

$$(6.6) \quad |\hat{F}_B(t)| \leq b_1 n^{-b_2 \log n} \quad \text{for } b_3 \log n \leq |t| \leq b_4 n^{3/2} \text{ and } B \in N(8, \hat{A}).$$

PROOF. Use Theorem 2.1 of van Zwet (1982) and proceed as in the proof of Schneller (1987), Proposition 4.3.9, pages 148–151. We remark that Bolthausen (1984) uses, in the proof of our (4.9), essentially $|\hat{a}_{ij}| \leq 1$. If one uses $\Sigma_i \hat{a}_{ij} = 0 = \Sigma_j \hat{a}_{ij}$ and $|\hat{a}_{ij}| \leq 1 + |\hat{a}_{ij}|^3$ instead of $|\hat{a}_{ij}| \leq 1$, one can show our (4.9) as well [for more details see Schneller (1987), Chapter 3, Section 3]. □

Our last lemma uses the estimation of (6.6) in order to estimate the right-hand side of (6.4).

6.7. LEMMA. *Let $A = (e_i d_j)$ be as in Lemma 6.5. Then there exists a positive number c depending only on e, E, d, D, δ and r, k, m, s such that*

$$(6.8) \qquad \int_{\mathbb{R}} (1 + |t|)|\hat{F}_B(t)\hat{U}_{D_A^2}(t)|\, dt \leq c(\log n)^2 \quad \text{for all } B \in N(8, \hat{A}).$$

PROOF. It suffices to prove (6.8) with $|t|$ instead of $1 + |t|$. Using the abbreviation $\theta = D_A^2$ and the constants $b_1, \dots, b_4$ from Lemma 6.5 we have

$$\begin{aligned}
&\int_{\mathbb{R}} |t|\, |\hat{F}_B(t)\hat{U}_\theta(t)|\, dt \\
&\quad \leq 2\left\{ \int_0^{b_3 \log n} t\, dt + \int_{b_3 \log n}^{b_4 n^{3/2}} t|\hat{F}_B(t)|\, dt + \int_{b_4 n^{3/2}}^{\infty} t|\hat{U}_\theta(t)|\, dt \right\} \\
&\quad = I_1 + I_2 + I_3.
\end{aligned}$$

From Lemma 6.5 we obtain $I_2 \to 0$ as $n \to \infty$ so that $I_1 + I_2 \leq c_1(\log n)^2$. Furthermore, using $\theta \geq 1/(4n)$ and (6.2) yields

$$\begin{aligned}
I_3 &= \frac{1}{\theta^2} \int_{b_4 n^{3/2}\theta}^{\infty} v|\hat{U}(v)|\, dv \qquad (v = t\theta) \\
&\leq 16n^2 \int_{c_2 n^{1/2}}^{\infty} v|\hat{U}(v)|\, dv \qquad \left(c_2 = \frac{b_4}{4} \right) \\
&\leq 16n^2 \int_{c_2 n^{1/2}}^{\infty} \frac{v^6}{c_2^5 n^{5/2}} |\hat{U}(v)|\, dv \\
&\leq 16 c_2^{-5} n^{-1/2} \int_{\mathbb{R}} v^6 |\hat{U}(v)|\, dv \to 0 \quad \text{as } n \to \infty.
\end{aligned}$$

□

For $n \times n$ matrices A with $n \geq n_1$ and $\beta_A \leq \varepsilon_0 n$ the estimation (2.16) now follows from the Lemmas 6.3 and 6.7, from Theorem 2.4 and $D_A^2 \leq cn^{-1+((4/k)-1)^+ +((4/s)-1)^+}$ where c depends only on e, E, d, D, δ and r, k, m, s. For the other matrices we obtain (2.16), if we choose $\mathscr{K}_1$ large enough (cf. proof of Theorem 2.1).

7. Some remarks on the proofs for the second order. The proof of Theorem 2.7 is completely given in Schneller (1987), Chapter 3, Section 8. This proof has a structure which is similar to that of Theorem 2.4, but it is much more extensive than the proof of Theorem 2.4. We mention some essential differences.

1. The set M has to be defined as a subset of N^{16} so that the maps $M \ni \underline{i} \to u(\underline{i})$, $t(\underline{i})$ and $v(\underline{i})$, with $v(\underline{i})$ as follows, are well defined. The permutation $v(\underline{i})$ leaves the numbers outside $\{i_1, \dots, i_{16}\}$ fixed and maps $i_1, \dots, i_8$ according to

$$\begin{matrix} i_1 \to i_{13} & i_3 \to i_{15} & i_5 \to i_9 & i_7 \to i_{11} \\ i_2 \to i_{14} & i_4 \to i_{16} & i_6 \to i_{10} & i_8 \to i_{12} \end{matrix}.$$

With these definitions we define analogously five random permutations $\pi_1, \dots, \pi_5$.

2. In order to obtain the right expansion in the analogue of equation (3.12) and for the estimation of the corresponding remainder terms $R(q_z)$, we need an analogue of Proposition 5.7 for $E(\mathscr{T}_A^2 1_{(-\infty, z]}(\mathscr{T}_A))$ and an Edgeworth expansion of first order for $E(\mathscr{T}_A 1_{(-\infty, z]}(\mathscr{T}_A))$.
3. The condition (2.8) is used mainly to prove the analogue of Lemma 5.2. For the rest of the proof one needs only the weaker condition [cf. Bickel and Robinson (1982), Lemma]

$$|\Delta_y^3 F_B(z)| \leq C_2'\left(E_A^3 + y^3\right) \quad \text{for all } z \in \mathbb{R}, 0 \leq y \leq E_A \text{ and } B \in N(16, \hat{A}), \tag{2.8'}$$

where

$$\Delta_y^3 F_B(z) = \Delta_y^2 F_B(z+y) - \Delta_y^2 F_B(z).$$

Our next remarks concern the proof of Theorem 2.12(b). This proof is also very similar to that of Theorem 2.12(a) and is completely given in Schneller (1987), Chapter 4, Sections 1 and 3. The appearance of the factor n^ε [instead of e.g., $(\log n)^3$] has its reason in establishing an estimation analogous to that of the term I_3 in the proof of Lemma 6.7. For that we need $b_4 n^{3/2}\theta \geq n^\varepsilon$ and, therefore, we have to take $\theta = n^\varepsilon E_A^3 \geq n^\varepsilon / n^{3/2}$ instead of $\theta = E_A^3$.

Acknowledgments. This paper is based on a part of the author's thesis written at the Technische Universität Berlin. The author wishes to thank Professor Dr. E. Bolthausen for suggesting the problem and for stimulating discussions. Furthermore, the author is indebted to the Associate Editor and the referees for their various suggestions which improved the original manuscript.

REFERENCES

BARBOUR, A. D. (1986). Asymptotic expansions based on smooth functions in the central limit theorem. *Probab. Theory Related Fields* **72** 289–303.

BICKEL, P. J. and ROBINSON, J. (1982). Edgeworth expansions and smoothness. *Ann. Probab.* **10** 500–503.

BICKEL, P. J. and VAN ZWET, W. R. (1978). Asymptotic expansions for the power of distribution-free tests in the two-sample problem. *Ann. Statist.* **6** 937–1004.

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

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

DOES, R. J. M. M. (1983). An Edgeworth expansion for simple linear rank statistics under the null-hypothesis. *Ann. Statist.* **11** 607–624.

ERICKSON, R. V. (1974). L_1 bounds for asymptotic normality of m-dependent sums using Stein's technique. *Ann. Probab.* **2** 522–529.

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

HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Academic, New York.

HO, S. T. and CHEN, L. H. Y. (1978). An L_p bound for the remainder in a combinatorial central limit theorem. *Ann. Probab.* **6** 231–249.

HOEFFDING, W. (1951). A combinatorial central limit theorem. *Ann. Math. Statist.* **22** 558–566.

MOTOO, M. (1957). On the Hoeffding's combinatorial central limit theorem. *Ann. Inst. Statist. Math.* **8** 145–154.

ROBINSON, J. (1978). An asymptotic expansion for samples from a finite population. *Ann. Statist.* **6** 1005–1011.

SCHNELLER, W. (1987). Edgeworth-Entwicklungen für lineare Rangstatistiken. Ph.D. thesis, Technische Universität Berlin.

SCHNELLER, W. (1988). A short proof of Motoo's combinatorial central limit theorem using Stein's method. *Probab. Theory Related Fields* **78** 249–252.

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.

VAN ZWET, W. R. (1982). On the Edgeworth expansion for the simple linear rank statistic. In *Colloq. Math. Soc. János Bolyai* (B. V. Gnedenko, M. L. Puri and I. Vincze, eds.) **32** 889–909. North-Holland, Amsterdam.

VON BAHR, B. (1967). Multi-dimensional integral limit theorems. *Ark. Mat.* **7** 71–88.

SEL SOFTWARE GMBH
LORENZWEG 5
1000 BERLIN 42
WEST GERMANY