Bernoulli **3**(3), 1997, 329–349

Berry–Esseen bounds for statistics of weakly dependent samples

V. BENTKUS[1], F. GÖTZE[1*] and A. TIKHOMIROV[2]

[1]*Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, 33501 Bielefeld 1, Germany*
[2]*Faculty of Mathematics, University of Syktyvkar, Oktyabrskii Prospekt 55, 167001 Syktyvkar, Russia*

We prove Berry–Esseen bounds for a general class of asymptotically normal statistics which are functions of N weakly dependent random variables under easily verifiable conditions. In particular, we show, for some $\delta > 0$, the validity of the bound $O(N^{-1/2}\log^{\delta} N)$ for U-statistics, studentized means, functions of sample means, functionals of empirical distribution functions and linear combinations of order statistics.

Keywords: absolute regularity; asymptotically normal statistics; Berry–Esseen bounds; functionals of empirical distribution functions; functions of sample means; linear combinations of order statistics; mixing; studentized means; U-statistics; weakly dependent random variables

1. Introduction and results

Let $X_1, X_2, \ldots$ be a sequence of random variables taking values in an arbitrary measurable space $(\mathscr{X}, \mathscr{A})$ which is *stationary in the strong sense*. We shall assume that the sequence satisfies an absolute regularity condition with coefficients

$$\beta(m) \overset{\text{def}}{=} \sup_{k\geqslant 1} \mathrm{E} \sup \{|P\{A|\sigma[1, k]\} - P\{A\}|: A \in \sigma[k+m, \infty)\} \to 0,$$

as $m \to \infty$, where $\sigma[a, b]$ denotes the σ-algebra generated by the random variables X_l such that $l \in [a, b]$.

The aim of this paper is to prove Berry–Esseen bounds for a sufficiently large class of statistics of weakly dependent random variables. Let $t = t_N$ be real-valued function of N variables. We shall consider statistics $T = t(X_1, \ldots, X_N)$ which can be represented as

$$T = S + R, \qquad \text{where } S = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} g(X_j), \tag{1.1}$$

for some function $g: \mathscr{X} \to \mathbb{R}$ such that $\mathrm{E}g(X_1) = 0$ and some remainder term $R =$

*To whom correspondence should be addressed. e-mail: goetze@mathematik.uni-bielefeld.de

$R_N(X_1, \ldots, X_N)$. The function g may depend on N; throughout we allow such dependence without explicitly mentioning it. Decomposition (1.1) is just a notational convention, and we shall later impose additional conditions on g and R.

Let

$$\sigma_N^2 \stackrel{\text{def}}{=} \mathrm{E}S^2 = \mathrm{E}g^2(X_1) + 2\sum_{j=1}^{N-1}(1 - jN^{-1})\mathrm{E}g(X_1)g(X_{1+j})$$

denote the variance of S, and let $\rho_s = \mathrm{E}|g(X_1)|^s$.

In the case of independent identically distributed (i.i.d.) random variables $X_1, X_2, \ldots,$ the linear part S of the statistic T is asymptotically normal as $N \to \infty$,

$$P\{S < \sigma_N x\} \to \Phi(x),$$

where $\Phi(x)$ is the standard normal distribution function, provided that

$$\sigma_N^2 \geqslant \Sigma^2 > 0, \qquad \text{for all } N, \tag{1.2}$$

and $\sup_N \rho_{2+\delta} < \infty$, for some fixed $\Sigma > 0$ and $\delta > 0$. In the weakly dependent case the additional condition $\sup_m m^{(1+\varepsilon)(1+1/\delta)}\beta(m) < \infty$, for some $\varepsilon > 0$, ensures the same result (see Ibragimov and Linnik 1971; Ibragimov 1975; Eberlein 1984). Thus the statistic T will be asymptotically normal provided that in addition $R \to 0$ in probability.

In order to prove Berry–Esseen bounds, some stronger conditions are necessary. It is known (Tikhomirov 1980) that the conditions

$$\sup_N \rho_3 \leqslant \rho < \infty \tag{1.3}$$

and

$$\beta(m) \leqslant K \exp\{-\beta m\}, \qquad \text{for all } m \geqslant 1, \tag{1.4}$$

for some $K < \infty$ and $\beta > 0$, together imply

$$\sup_x |P\{S < \sigma_N x\} - \Phi(x)| \leqslant AN^{-1/2}\log^2 N$$

with a constant A depending on K, β, Σ and ρ only. We shall extend this estimate for general nonlinear statistics.

Let $\sigma^{\mathrm{c}}[j, k]$ denote the σ-algebra generated by X_l such that $l \notin [j, k]$ and $1 \leqslant l \leqslant N$. In the case $k < j$, set $\sigma^{\mathrm{c}}[j, k] = \sigma(X_1, \ldots, X_N)$.

Theorem 1.1. *Assume that* (1.2)–(1.4) *hold. Let*

$$R_{j,k} = R_{j,k}(X_1, \ldots, X_{j-1}, X_{k+1}, \ldots, X_N)$$

denote any $\sigma^{\mathrm{c}}[j, k]$*-measurable random variables such that* $R_{j,j-1} = R$, *for all* $1 \leqslant j \leqslant N$, *and let*

$$\gamma \stackrel{\text{def}}{=} \max\{\mathrm{E}^{2/3}|R_{j,k} - R_{j,k-1}|^{3/2}\colon |j-k| \leqslant \log^3 N \text{ and } 1 \leqslant j \leqslant k \leqslant N\}.$$

Then

$$\sup_x |P\{T < \sigma_N x\} - \Phi(x)| \leq AN^{1/2}\log^2 N + A\mathrm{E}|R| + A\gamma\sqrt{N}\log^3 N$$

with a constant A depending only on K, β, Σ and ρ. If the function g is independent of N and

$$\sigma^2 \stackrel{\text{def}}{=} \lim_{N\to\infty} \sigma_N^2 = \mathrm{E}g^2(X_1) + 2\sum_{j=1}^{\infty} \mathrm{E}g(X_1)g(X_{1+j}) > 0,$$

then $\sup_x |P\{T < \sigma x\} - \Phi(x)|$ *satisfies the same estimate with a constant A depending on K, β, σ and ρ only.*

For the proof of the result we apply a modification of the method used by Stein (1972) and Tikhomirov (1980), among others. Theorem 1.1 seems to be the first Berry–Esseen bound for a general class of statistics of dependent samples. In Section 2 we apply Theorem 1.1 to functions of sample means, functionals of the empirical distribution functions, studentized means, *U*-statistics and linear combinations of order statistics. In all these applications the estimation of $\mathrm{E}|R|$ and γ in Theorem 1.1 is quite simple and reduces to the estimation of certain moments. The random variables $R_{j,k}$ may be obtained by the simple rule: 'remove all terms of R involving random variables X_l with l such that $j \leq l \leq k$'.

In the literature only special classes of (nonlinear) statistics of weakly dependent samples have so far been considered: Yoshihara (1976) proved asymptotic normality for a class of *U*-statistics; Denker (1982) and Denker and Keller (1983) proved the asymptotic normality and functional limit theorems for classes of *U*-statistics and von Mises statistics and obtained Berry-Esseen bounds; Yoshihara (1984) obtained a Berry–Esseen bound for *U*-statistics; for results concerning sums see, for example, a review of Sunklodas (1991). Edgeworth expansions for statistics of dependent samples were considered by Götze and Hipp (1983; 1994). The aim of the present paper is to develop a method to prove sufficiently precise Berry–Esseen bounds for a sufficiently general class of statistics such that the previous results are included as particular cases, thus avoiding further extensions for specific statistics.

Our result is similar to a Berry–Esseen bound described by van Zwet (1984) for symmetric statistics of independent samples. In the independent case more precise estimates are known; see, for example, Friedrich (1989), Bolthausen and Götze (1993), Bentkus and Götze (1996) and, for lower estimates, Bentkus *et al.* (1994). We could improve the moment conditions for the nonlinear part of the statistic combining methods developed for independent random variables and those of the present paper, but detailed proofs would require a large amount of routine work. Furthermore, the dependence on $\log N$ in Theorem 1.1 can be improved using modifications of our proofs. In order to avoid technicalities, we do not formulate and prove bounds with better powers of $\log N$. The question whether the $\log N$ factors in Theorem 1.1 are unavoidable remains open. Recently Rio (1996) obtained an $O(N^{-1/2})$ result for sums in the case of φ-mixing stationary bounded sequences. Whether Rio's result for sums holds for the β-mixing case, remains open.

The mixing condition (1.4) is relatively weak. For example, solutions of the Itô equations in Euclidean spaces satisfy this condition and do not satisfy conditions with other stronger (e.g., uniformly) mixing coefficients; see Veretennikov (1987). Heinrich (1992) found a clear and simple condition yielding (1.4) for stationary renewal processes. Condition (1.4) is also fulfilled when the similar condition holds for ψ-mixing coefficients $\psi(m)$ or for φ-mixing coefficients $\varphi(m)$ since $\beta(m) \leqslant \varphi(m) \leqslant \psi(m)$. We have chosen β-mixing as the weakest mixing condition such that the decoupling inequality (1.5) for the distance in the variation holds. The exponential decay of the mixing coefficients in (1.4) is imposed in order to simplify the technicalities.

Let $A, A_1, A_2, \ldots$ denote generic constants which may depend on parameters of interest, such as $K, \beta, \rho, \alpha, \ldots$. Let m denote a natural number such that for a sufficiently large generic constant A, $m \approx A \log N$, for example, $m = [A \log N]$.

Let $\tau, \tau_1, \tau_2, \ldots$ denote a sequence of i.i.d. random variables uniformly distributed on [0, 1], and independent of all other random variables. By E_τ we shall denote the conditional expectation given all random variables but τ.

Let $\hat{\xi}$ denote an independent copy of the random variable ξ, which is also independent of all other random variables appearing in the specific context.

We shall often use the following simple decoupling inequality which allows us to get rid of dependence problems. Assume that ξ is $\sigma(-\infty, k]$-measurable, that η is $\sigma[k+m, \infty)$-measurable and that the random variables ξ and η take values in a Polish space. If $\sup_{u,v} |\varphi(u, v)| \leqslant D$, then

$$|\mathrm{E}\varphi(\xi, \eta) - \mathrm{E}\varphi(\xi, \hat{\eta})| \leqslant D\beta(m). \tag{1.5}$$

Weaker measures of dependence such as α-dependence will not allow inequalities of type (1.5) which we need to treat nonlinear statistics.

2. Applications

In this section we shall apply Theorem 1.1 to various special statistics to show that in fact the conditions can be easily verified. Similar examples were considered in the independent case by Bentkus *et al.* (1997).

2.1. Functions of sample means

Assume for this subsection only that $X_1, X_2, \ldots$ take values in a real separable Banach space $\mathscr{B}$. Consider the statistic

$$T = \sqrt{N}(H(\bar{X}) - H(0))$$

for a function $H: \mathscr{B} \to \mathbb{R}$, where the sample mean $\bar{X} \stackrel{\text{def}}{=} N^{-1}(X_1 + \cdots + X_N)$.

We shall use Taylor's expansion for sufficiently smooth functions $f: \mathscr{B} \to \mathbb{R}$ (see, for example, Cartan 1971),

$$f(x+h) = f(0) + f'(x)h + \cdots + \frac{1}{k!} f^{(k)}(x)h^k + \frac{1}{k!}\mathrm{E}_\tau(1-\tau)^k f^{(k+1)}(x+\tau h)h^{k+1},$$

where $f^{(j)}(x)h^j$ denotes the jth Fréchet derivative of f at point x in the direction h.

Assuming that H is Fréchet differentiable and using Taylor's expansion, we may write $g(x) \stackrel{\text{def}}{=} H'(0)x$, and

$$R \stackrel{\text{def}}{=} \sqrt{N}(H(\bar{X}) - H(0) - H'(0)\bar{X}) = \sqrt{N}\mathrm{E}_\tau(1-\tau)H''(\tau\bar{X})\bar{X}^2.$$

The function g is independent of N, and

$$T = \frac{1}{\sqrt{N}}\sum_{j=1}^{N} g(X_j) + R.$$

Denote

$$M_s \stackrel{\text{def}}{=} \sum_{j=1}^{s} \sup_{x\in B} \|H^{(j)}(x)\|.$$

Theorem 2.1. *Assume that the mixing condition* (1.4) *is fulfilled. Let*

$$\mathrm{E}X_1 = 0, \qquad \rho \stackrel{\text{def}}{=} \mathrm{E}\|X_1\|^3 < \infty.$$

Assume that $\sigma^2 > 0$. *If* $M_3 < \infty$ *then there exists a constant* A *depending only on* $\mathscr{B}$, K, β, ρ, M_3 *and* σ *such that*

$$\delta_N \stackrel{\text{def}}{=} \sup_x |P\{T < \sigma x\} - \Phi(x)| \leqslant CN^{-1/2}\log^4 N.$$

If the Banach space $\mathscr{B}$ *is of type* 2 *then the smoothness condition* $M_3 < \infty$ *may be relaxed to* $M_2 < \infty$.

Remark. A Banach space $\mathscr{B}$ is of type 2 if there exists a constant $C = C(\mathscr{B})$ such that $\mathrm{E}\|\sum_{i=1}^N Y_i\|^2 \leqslant C\sum_{i=1}^N \mathrm{E}\|Y_i\|^2$, for any independent centred random variables Y_i. Finite-dimensional spaces, Hilbert spaces and L_p, l_p, $2 \leqslant p < \infty$, are of type 2.

Proof of Theorem 2.1. We shall derive the result from Theorem 1.1.

The verification of $|(\sigma_N/\sigma) - 1| \leqslant A/N$ is easy. Therefore it is sufficient to prove that

$$\mathrm{E}|R| \leqslant AmN^{-1/2} \tag{2.1}$$

and that

$$\mathrm{E}|R_{j,k} - R_{j,k-1}|^{3/2} \leqslant A(m/N)^{3/2}, \tag{2.2}$$

where

$$R_{j,k} \stackrel{\text{def}}{=} \sqrt{N}(H(\bar{X}_{[j,k]}) - H(0) - H'(0)\bar{X}_{[j,k]}),$$

and

$$\bar{X}_{[j,k]} \stackrel{\text{def}}{=} \bar{X} - N^{-1}\sum_{i\in[j,k]} X_i.$$

While proving (2.1) and (2.2) we may assume that $\|X_j\| \leqslant N^2$ with probability 1. Otherwise we may replace X_j by $X_j\mathbf{1}\{\|X_j\| \leqslant N^2\} - \mathrm{E}X_j\mathbf{1}\{\|X_j\| \leqslant N^2\}$. For example, in the case of (2.1), by such a replacement the error is bounded from above by

$$AN^{-1/2}\sum_{j=1}^{N}\mathrm{E}\|X_j\|\mathbf{1}\{\|X_j\| \geqslant N^2\} = A\sqrt{N}\mathrm{E}\|X_1\|\mathbf{1}\{\|X_1\| \geqslant N^2\} \leqslant AN^{-7/2}.$$

Indeed, we may split $X_j = X_j\mathbf{1}\{\|X_j\| \leqslant N^2\} + X_j\mathbf{1}\{\|X_j\| > N^2\}$, use the representation $R = \sqrt{N}(H(\bar{X}) - H(0) - H'(0)\bar{X})$, apply $|H(u) - H(v)| \leqslant M_1\|u - v\|$ and use the triangle inequality.

Let us prove (2.1). Note that $|R| \leqslant M_2\sqrt{N}\|\bar{X}\|^2$. Thus, in the case of the Banach space $\mathscr{B}$ of type 2, the bound (2.1) is a consequence of Lemma 3.1 below. The proof of (2.1) without the type 2 assumption is slightly more complicated (see Section 3).

Let us prove (2.2). Notice that $\bar{X}_{[j,k-1]} = \bar{X}_{[j,k]} + N^{-1}X_k$ and that

$$|R_{j,k} - R_{j,k-1}| = \sqrt{N}|H(\bar{X}_{[j,k-1]}) - H(\bar{X}_{[j,k]}) - H'(0)N^{-1}X_k|.$$

Expanding in powers of $N^{-1}X_k$, we see that instead of (2.2) it is sufficient to show that

$$\mathrm{E}|H'(\bar{X}_{[j,k]})X_k - H'(0)X_k|^{3/2} \leqslant Am^{3/2}N^{-3/4}. \tag{2.3}$$

Let us write $\bar{X}_{[j,k]} = U + \Delta$, where U is the sum of terms $N^{-1}X_l$ in the sum $\bar{X}_{[j,k]}$ such that $|l - k| > m$, and where Δ denotes the remaining part of the sum $\bar{X}_{[j,k]}$. Expanding in powers of Δ and using $(a_1 + \cdots + a_m)^{3/2} \leqslant \sqrt{m}(a_1^{3/2} + \cdots + a_m^{3/2})$, we reduce (2.3) to

$$\mathrm{E}|H(U)X_k - H'(0)X_k|^{3/2} \leqslant Am^{3/2}N^{-3/4}. \tag{2.4}$$

Using (1.5), we may replace X_k in (2.4) by an independent copy, say $\hat{X}_k$. An expansion in powers of U and an application of the Hölder inequality show that (2.4) follows from

$$\mathrm{E}(H''(\tau U)U\hat{X}_k)^2 \leqslant Am^2N^{-1}. \tag{2.5}$$

If the Banach space is of type 2, we may apply Lemma 3.1 (below) in order to estimate $\mathrm{E}\|U\|^2$, and (2.5) implies the result. Let us continue the proof of (2.2) for arbitrary Banach spaces. Expanding the square in (2.5), we see that (2.5) is a consequence of

$$N^{-1}\sum_i\sum_l |\mathrm{E}H''(\tau U)X_i\hat{X}_kH''(\tau U)X_l\hat{X}_k| \leqslant Am^2, \tag{2.6}$$

where the sums $\sum_i$ and $\sum_l$ are taken over the indices i and l in the sum U. By the triangle inequality we may remove in (2.6) summands with $|i - l| \leqslant m$. Next we may remove from the sum U summands with indices near to i or l. An application of (1.5) will complete the proof. □

2.2. Functionals of empirical distribution functions

The results of this subsection are obvious extensions of those for functions of sample means.

Throughout this subsection we shall use the following notation. Let $\eta_1, \eta_2, \ldots$ denote a sequence of real random variables stationary in the strong sense. Let F denote the distribution function of η_1, and let F_N be the empirical distribution function corresponding to the sample $\eta_1, \dots, \eta_N$. Define the random processes $X_i(t)$, $t \in \mathbb{R}$, $1 \leqslant i \leqslant N$, by

$$X_i(t) \stackrel{\text{def}}{=} \mathbf{1}\{\eta_i < t\} - F(t).$$

Finally, let

$$E_N \stackrel{\text{def}}{=} \sqrt{N}(F_N - F) = (X_1 + \cdots + X_N)/\sqrt{N}$$

denote the empirical process.

Assume that a functional H takes real values and that $H(F)$ and $H(F_N)$ are well defined. Define the statistic

$$T \stackrel{\text{def}}{=} \sqrt{N}(H(F_N) - H(F)).$$

We may write $F_N - F = E_N/\sqrt{N} = \bar{X}$. Introducing the functional

$$G_F(h) \stackrel{\text{def}}{=} H(F + h),$$

we have

$$T = \sqrt{N}(G_F(\bar{X}) - G_F(0)).$$

Let us define derivatives of H via derivatives of G_F as $H^{(s)}(F + x) \stackrel{\text{def}}{=} G_F^{(s)}(x)$. In order to define derivatives of G_F, introduce a Banach space $\mathscr{B}$, which may depend on F and should be chosen in dependence on H and the particular problem. We shall assume that $G_F\colon \mathscr{B} \to \mathbb{R}$ admits Fréchet derivatives, and we shall require that $\mathscr{B}$ contains the sample functions $X_i(t)$ almost surely. Furthermore, we assume that X_i, $\bar{X}$ are well-defined and take values in $\mathscr{B}$.

Denote

$$M_s \stackrel{\text{def}}{=} \sum_{j=1}^{s} \sup_{x \in B} \|H^{(j)}(x)\|.$$

Theorem 2.1 implies the following result.

Theorem 2.2. *Assume that the sequence $\eta_1, \eta_2, \dots$ satisfies the mixing condition* (1.4). *Let*

$$\mathrm{E}X_1 = 0, \qquad \rho \stackrel{\text{def}}{=} \mathrm{E}\|X_1\|^3 < \infty.$$

Define $g(x) \stackrel{\text{def}}{=} H'(F)x$ and assume that $\sigma^2 > 0$. If $M_3 < \infty$, then there exists a constant A depending only on $\mathscr{B}$, K, β, ρ, M_3 and σ such that

$$\sup_x |P\{T < \sigma x\} - \Phi(x)| \leqslant CN^{-1/2} \log^4 N.$$

If the Banach space $\mathscr{B}$ is of type 2 *then the condition $M_3 < \infty$ may be relaxed to $M_2 < \infty$.*

2.3. Linear combinations of order statistics

We will apply the result for functionals of empirical distribution functions using the notation of Section 2.2.

Consider the statistic

$$l_N \stackrel{\text{def}}{=} N^{-1} \sum_{i=1}^{N} c_{iN} \eta_{i:N},$$

where $\eta_{1:N} \leqslant \cdots \leqslant \eta_{N:N}$ denote the order statistics of $\eta_1, \ldots, \eta_N$, and coefficients $c_{1N}, \ldots, c_{NN}$ are generated by a weight function J: $[0, 1] \to \mathbb{R}$,

$$c_{iN} \stackrel{\text{def}}{=} N \int_{(i-1)/N}^{i/N} J(u)\,\mathrm{d}u.$$

Define

$$T \stackrel{\text{def}}{=} \sqrt{N}(l_N - \mu),$$

where

$$\mu = \int_{-\infty}^{\infty} xJ(F(x))\,\mathrm{d}F(x).$$

If $\mathrm{E}|\eta_1| < \infty$, the boundedness of J is sufficient for the following representation (see Govindarajulu and Mason 1983):

$$l_N - \mu = \int_{-\infty}^{\infty} [\Psi(F_N(t)) - \Psi(F(t))]\,\mathrm{d}t,$$

where

$$\Psi(x) = \int_x^1 J(u)\,\mathrm{d}u.$$

Therefore we may write

$$l_N - \mu = H(F_N) - H(F),$$

where

$$H(h) = \int_{-\infty}^{0} [\Psi(h(t)) - \Psi(0)]\,\mathrm{d}t + \int_0^{\infty} \Psi(h(t))\,\mathrm{d}t.$$

Let $\|\cdot\|_p$ denote the norm of the space $L^p(\mathbb{R})$. Let B be the Banach space of functions with norm

$$\|x\| = \|x\|_1 + \|x\|_2 + \|x\|_3.$$

If the function J is smooth, then the functional $h \mapsto T(P + h)$: $B \to \mathbb{R}$ is Fréchet differentiable and

$$M_s \leqslant L_{s-1} \stackrel{\text{def}}{=} \sum_{j=0}^{s-1} \sup_x |J^{(j)}(x)|, \qquad s \leqslant 3.$$

It is easy to verify that the random process X_1 (see Section 2.2) satisfies

$$\mathrm{E}\|X_1\|_p^s \leqslant c(p, s)(\mathrm{E}|\eta_1|^{s/p} + (\mathrm{E}|\eta_1|)^{s/p}), \qquad \text{for all } p \geqslant 1,\ s > 0.$$

Theorem 2.3. *Assume that a stationary sequence* $\eta_1, \eta_2, \ldots$ *satisfies the mixing condition* (1.4). *Let* $\rho \stackrel{\text{def}}{=} \mathrm{E}|\eta_1|^3 < \infty$. *Define* $a_0 = 1$, $a_j \stackrel{\text{def}}{=} 2$, *for* $j > 0$,

$$\sigma^2 \stackrel{\text{def}}{=} \int_{\mathbb{R}^2} J(F(t))J(F(s)) \sum_{j=0}^{\infty} a_j[P\{\eta_1 < t, \eta_{1+j} < s\} - F(t)F(s)]\,\mathrm{d}s\,\mathrm{d}t$$

and assume that $\sigma^2 > 0$. *If* $L_2 < \infty$ *then there exists a constant A depending only on K,* β, L_2 *and* σ *such that*

$$\sup_x |P\{T < \sigma x\} - \Phi(x)| \leqslant CN^{-1/2} \log^4 N.$$

The condition $L_2 < \infty$ *may be relaxed to* $L_1 < \infty$.

Under the condition $L_2 < \infty$, the proof of Theorem 2.3 requires a straightforward application of Theorem 2.2 only. The weaker condition $L_1 < \infty$ implies an inequality for the empirical processes E_N like that characterizing type 2 spaces. Hence, Theorem 2.3 again follows from Theorem 2.2 (cf. Bentkus *et al.* 1997) in this case as well.

2.4. Studentized sample means

In this subsection we shall assume that random variables in the stationary sequence $X_1, X_2, \ldots$ take real values and satisfy the mixing condition (1.4). Assume as well that $\mathrm{E}X_1 = 0$ and $\mathrm{E}X_1^4 \leqslant \rho_4 < \infty$. Denote

$$\bar{X} = N^{-1}(X_1 + \cdots + X_N), \qquad S \stackrel{\text{def}}{=} \sqrt{N}\bar{X}, \qquad \sigma_N^2 = \mathrm{E}S^2,$$

and assume that $\lim_{N\to\infty} \sigma_N^2 \stackrel{\text{def}}{=} \sigma^2 > 0$. Consider the estimator

$$s^2 = \frac{1}{N} \sum_{j=1}^{N} \sum_{l:|j-l|\leqslant m} X_j X_l$$

of σ. It is consistent and asymptotically unbiased provided $m \approx A \log N$, for a sufficiently large constant A. Put $s \stackrel{\text{def}}{=} \sqrt{s^2}$ if $s^2 \geqslant 0$, and $s \stackrel{\text{def}}{=} 0$ if $s^2 < 0$. Introduce the studentized statistic t,

$$t \stackrel{\text{def}}{=} \begin{cases} \dfrac{S}{s}, & s > 0, \\ 0, & s = 0. \end{cases}$$

Theorem 2.4. *There exists a constant* $A = A(K, \beta, \sigma, \rho_4)$ *such that*

$$\sup_x |P\{t < x\} - \Phi(x)| \leqslant AN^{-1/2} \log^{9/2} N.$$

Proof. We shall derive the result from Theorem 1.1. Without loss of generality, we shall assume that $\sigma^2 = 1$.

Notice that

$$P\{|s^2 - 1| \geqslant \varepsilon\} \leqslant A\varepsilon^{-1} m^{3/2} N^{-1/2}, \qquad \text{for any } \varepsilon > 0. \tag{2.7}$$

Indeed, by Chebyshev's inequality, it is sufficient to estimate $\mathrm{E}|s^2 - \sigma^2| \leqslant \mathrm{E}|s^2 - \mathrm{E}s^2| + |\mathrm{E}s^2 - 1|$, and to show that

$$|\mathrm{E}s^2 - 1| \leqslant Am^2 N^{-1}, \qquad \mathrm{E}(s^2 - \mathrm{E}s^2)^2 \leqslant m^3 N^{-1}. \tag{2.8}$$

While proving (2.8) we may assume that $|X_j| \leqslant N$. Otherwise we may replace X_j by $X_j \mathbf{1}\{|X_j| \leqslant N\} - \mathrm{E}X_j \mathbf{1}\{|X_j| \leqslant N\}$. By the triangle inequality, the error is bounded by Am/N. Let us prove, for example, the second inequality in (2.8). By the triangle inequality,

$$\mathrm{E}(s^2 - \mathrm{E}s^2)^2 \leqslant mN^{-2} \sum_{s=0}^{m} \mathrm{E}\left(\sum_{j=1}^{N-s} Z_{j,s}\right)^2, \qquad Z_{j,s} \overset{\text{def}}{=} X_j X_{j+s} - \mathrm{E}X_j X_{j+s}.$$

The sequence $X_j X_{j+s}$, $j = 1, 2, \ldots,$ is stationary and satisfies the mixing condition with coefficients $\beta(m - s)$, for $m \geqslant s$. Thus we may apply Lemma 3.1 (with $2m$ instead of m) and get $\mathrm{E}(\sum_{j=1}^{N-s} Z_{j,s})^2 \leqslant AmN$. Collecting the estimates, we obtain (2.8).

Introduce a C^∞ function such that

$$\theta(x) \overset{\text{def}}{=} \begin{cases} 2 & x \leqslant 0 \\ \dfrac{1}{\sqrt{x}} & \dfrac{1}{2} \leqslant x \leqslant \dfrac{3}{2} \\ 0 & x \geqslant 2. \end{cases}$$

Due to (2.7)

$$\sup_x |P\{t < x\} - P\{S\theta(s^2) < x\}| \leqslant Am^{3/2} N^{-1/2},$$

and it is sufficient to prove that

$$\sup_x |P\{S\theta(s^2) < x\} - \Phi(x)| \leqslant AmN^{-1/2}. \tag{2.9}$$

While proving (2.9) we may assume that $|X_j| \leqslant N^2$, for $1 \leqslant j \leqslant N$. Otherwise we may replace the X_j by their truncated and centred versions because θ is a bounded function such that $|\theta(u) - \theta(v)| \leqslant A|u - v|$.

Write

$$T \overset{\text{def}}{=} S\theta(s^2) = S + R, \qquad \text{with } R \overset{\text{def}}{=} S(\theta(s^2) - 1).$$

In order to apply Theorem 1.1 we have to verify that

$$\mathrm{E}|R| \leqslant Am^2 N^{-1/2}, \qquad \mathrm{E}|R_{j,k} - R_{j,k-1}|^{3/2} \leqslant Am^{9/4} N^{-3/2}, \tag{2.10}$$

for $|j - k| \leqslant \log^3 N$, with

$$R_{j,k} \overset{\text{def}}{=} S_{j,k}(\theta(s_{j,k}^2) - 1),$$

where $S_{j,k}$ denotes the sum S without the summands $N^{-1/2}X_i$, $j \leqslant i \leqslant k$, and $s_{j,k}^2$ denotes the sum s^2 but without the summands $N^{-1}X_iX_l$ such that at least one of $i, l \in \{j, \ldots, k\}$.

Using $\theta(1) = 1$ and $|\theta(u) - \theta(v)| \leqslant A|u - v|$, we have

$$\mathrm{E}|R| \leqslant A\mathrm{E}|S|\,|s^2 - 1| \leqslant Am^2 N^{-1/2},$$

and the bound for $\mathrm{E}|R|$ follows provided we apply Hölder's inequality, then Lemma 3.1 to bound $\mathrm{E}S^2 \leqslant Am^{1/2}$, and use (2.8).

Let us prove the second inequality in (2.10). It is sufficient to show that

$$\mathrm{E}|(S_{j,k} - S_{j,k-1})(\theta(s_{j,k}^2) - 1)|^{3/2} \qquad \text{and } \mathrm{E}|S_{j,k-1}(\theta(s_{j,k}^2) - \theta(s_{j,k-1}^2))|^{3/2} \tag{2.11}$$

are bounded by $Am^{9/4}N^{-3/2}$. Let us estimate the first expectation in (2.11). Notice that $|S_{j,k} - S_{j,k-1}| = N^{-1/2}|X_k|$ and that we may represent $s_{j,k}^2 = q_{j,k} + \Delta_{j,k}$, where $\Delta_{j,k}$ denotes the sum of terms $N^{-1}X_iX_l$ of the sum $s_{j,k}^2$ such that i or l is near to k. Using $|\theta(u) - \theta(v)| \leqslant A|u - v|^{5/6}$ and the triangle inequality, we may neglect $\Delta_{j,k}$, and it is sufficient to show that

$$\mathrm{E}|X_k(\theta(q_{j,k}^2) - 1)^{3/2} \leqslant Am^{9/4}N^{-3/4}.$$

Using (1.5), we may replace X_k by its independent copy, and it remains to show that

$$\mathrm{E}|\theta(q_{j,k}^2) - 1|^{3/2} \leqslant A\mathrm{E}|q_{j,k}^2 - 1|^{3/2} \leqslant Am^{9/4}N^{-3/4}.$$

But this bound may be proved like (2.8).

In order to estimate the second expectation in (2.11) notice that the difference $s_{j,k}^2 - s_{j,k-1}^2$ may contain at most $A \log N$ summands $N^{-1}X_iX_l$ and proceed similarly. □

2.5. U-statistics

For notational simplicity we shall consider U-statistics of second order only, that is

$$U = S + R, \qquad \text{with } S = \frac{1}{\sqrt{N}} \sum_1^N g(X_j),\ R = N^{-3/2} \sum_{1 \leqslant i < j \leqslant N} \psi(X_i, X_j),$$

where the function $\psi\colon \mathscr{X}^2 \to \mathbb{R}^1$ is symmetric, $\psi(x, y) = \psi(y, x)$, and $\mathrm{E}\psi(x, X_j) = 0$, for all $x \in \mathscr{X}$ (functions g and ψ may depend on N).

Theorem 2.5. *Assume that the conditions* (1.2)–(1.4) *are fulfilled and that*

$$\sup_N \sup_{1\leqslant l<j\leqslant N} \mathrm{E}\psi^2(X_l, X_j) + \sup_N \mathrm{E}\psi^2(\hat{X}_1, X_2) \leqslant L < \infty.$$

Then

$$\delta_N \stackrel{\text{def}}{=} \sup_x |P\{U < x\sigma_N\} - \Phi(x)| \leqslant AN^{-1/2}\log^{7/2} N,$$

where the constant A may depend on K, β, ρ, Σ and L only.

Corollary 2.6. *Assume that the conditions of Theorem 2.5 hold and let g be independent of N. Assume that the variance σ^2 is positive. Then*

$$\delta'_N \stackrel{\text{def}}{=} \sup_x |P\{U < \sigma x\} - \Phi(x)| \leqslant AN^{-1/2}\log^{7/2} N,$$

where the constant A may depend on K, β, ρ, σ and L only.

Remark. Yoshihara (1984) proved the bound $\delta'_N = O(N^{-1/2}\log^2 N)$ assuming that g and ψ are independent of N and that $\sup_{l<j} \mathrm{E}|\psi(X_l, X_j)|^3 + \mathrm{E}|\psi(\hat{X}_1, X_2)|^3 < \infty$.

Proof of Theorem 2.5. We shall apply Theorem 1.1. It is sufficient to show that

$$\mathrm{E}|R| \leqslant AmN^{-1/2} \qquad \text{and} \qquad \gamma \leqslant A\sqrt{m}N^{-1}, \tag{2.12}$$

with

$$R_{j,k} \stackrel{\text{def}}{=} N^{-3/2} \sum_{B\cap[j,k]=\varnothing} \psi(X_i, X_l),$$

where the sum is taken over all two-point subsets $B = \{i, l\} \subset \{1, \dots, N\}$ such that $B \cap [j, k] = \varnothing$. Notice that

$$R_{j,k} - R_{j,k-1} = N^{-3/2} \sum_{l\in[1,N]\setminus[j,k]} \psi(X_k, X_l).$$

Therefore the inequalities (2.12) follow from

$$\mathrm{E}\Big|\sum_{1\leqslant i<j\leqslant N} \psi(X_i, X_j)\Big| \leqslant AmN \tag{2.13}$$

and

$$\mathrm{E}\Big|\sum_{l\in[1,N]\setminus[j,k]} \psi(X_k, X_l)\Big|^{3/2} \leqslant A(mN)^{3/4}. \tag{2.14}$$

While proving (2.13) and (2.14) we may and shall assume that $|\psi(x, y)| \leqslant 4N^3$. Indeed, otherwise we may replace $\psi(x, y)$ by

$$\psi(x,\, y)\mathbf{1}\{|\psi(x,\, y)| \leqslant N^3\} - \lambda(x) - \lambda(y) + \mathrm{E}\lambda(\hat{X}_1),$$

with $\lambda(x) \stackrel{\text{def}}{=} \mathrm{E}\psi(x,\, X_1)\mathbf{1}\{|\psi(x,\, X_1)| \leqslant N^3\}$.

Lemma 3.2, together with the Hölder inequality, yields (2.13).

Let us prove (2.14). By the triangle inequality, we may remove from the sum in (2.14) all summands with indices l such that $|l - k| \leqslant m$. Using (1.5), we may replace X_k by its independent copy $\hat{X}_k$. Thus, by the Hölder inequality, (2.14) follows from

$$\mathrm{E}\left(\left(\sum \psi(\hat{X}_k,\, X_l)\right)^2 \middle| \hat{X}_k\right) \leqslant AmN,$$

where the sum is taken over all l such that $|l - k| > m$ and $l \in [1,\, N]\backslash[j,\, k]$. An application of Lemma 3.1 concludes the proof of (2.14) and of the lemma. □

3. Auxiliary results

Lemma 3.1. *Assume* (1.4) *and let the random variables X_i take values in a Banach space $\mathscr{B}$ of type* 2. *Let* $\mathrm{E}X_i = 0$ *and* $\mathrm{P}\{\|X_i\| \leqslant D\} = 1$, *for all* $1 \leqslant i \leqslant N$, *for some* $D < \infty$. *Then*

$$\mathrm{E}\left\|\sum_{i=1}^N X_i\right\|^2 \leqslant 2Cn\sum_{i=1}^N \mathrm{E}\|X_i\|^2 + 2\beta(n)D^2N^3,$$

for any natural n.

Proof. Let us split $\{1,\, \ldots,\, N\} = \Delta_1 \cup \Delta_2 \cup \cdots \cup \Delta_s$ into the union of disjoint subintervals Δ_j of length n with $s \approx N/n$, for $1 \leqslant j \leqslant s - 1$. Denote

$$Y_j = \sum_{i \in \Delta_j} X_i, \qquad Z_1 = \sum_{j:1 \leqslant 2j \leqslant s} Y_{2j}, \qquad Z_2 = \sum_{j:1 \leqslant 2j+1 \leqslant s} Y_{2j+1}.$$

Then $\sum_{j=1}^N X_j = Z_1 + Z_2$, and it is sufficient to estimate $\mathrm{E}\|Z_1\|^2$ and $\mathrm{E}\|Z_2\|^2$. We shall estimate $\mathrm{E}\|Z_1\|^2$ only. Let $\hat{Y}_1,\, \hat{Y}_2,\, \ldots$ denote a sequence of independent copies of $Y_1,\, Y_2,\, \ldots$. Applying (1.5) and estimating $\|X_j\| \leqslant D$, we have

$$\mathrm{E}\|Z_1\|^2 = \mathrm{E}\|Y_2 + Z_1 - Y_2\|^2 \leqslant \mathrm{E}\|\hat{Y}_2 + Z_1 - Y_2\|^2 + D^2N^2\beta(n).$$

Repeating this procedure (with $Y_4,\, Y_6,\, \ldots$ instead of Y_2), we obtain

$$\mathrm{E}\|Z_1\|^2 \leqslant \mathrm{E}\left\|\sum_{j:1 \leqslant 2j \leqslant s} \hat{Y}_{2j}\right\|^2 + D^2N^3\beta(n).$$

Now the definition of Banach spaces of type 2 and the triangle inequality yield

$$\mathrm{E}\left\|\sum_{j:1 \leqslant 2j \leqslant s} \hat{Y}_{2j}\right\|^2 \leqslant Cn\sum_{i=1}^N \mathrm{E}\|X_i\|^2.$$

Collecting the estimates we conclude the proof of the lemma. □

Proof of (2.1) for arbitrary Banach spaces. Expanding into a Taylor series, we may write

$$R = N^{-1/2}\mathrm{E}_\tau \sum_{i=1}^{N} (H'(\tau\bar{X}) - h'(0))X_i,$$

and (2.1) follows from

$$\mathrm{E}\left| \sum_{i=1}^{N} (H'(\tau\bar{X}) - H'(0))X_i \right| \leqslant Am. \tag{3.1}$$

Write $\bar{X}_i = \bar{X} - \Delta$, where $\Delta \overset{\text{def}}{=} N^{-1}\sum_{j\in[i-m,i+m]} X_j$. Expanding in powers of Δ, using the triangle inequality and applying the Hölder inequality, we see that instead of (3.1) it is sufficient to prove

$$\mathrm{E}\left(\sum_{i=1}^{N} \mathrm{E}_\tau (H'(\tau\bar{X}_i) - H'(0))X_i \right)^2 \leqslant Am^2. \tag{3.2}$$

Expanding the square in (3.2), we arrive at

$$\sum_{i=1}^{N}\sum_{j=1}^{N} |\mathrm{E}(H'(\tau\bar{X}_i) - H'(0))X_i(H'(\tau\bar{X}_j) - H'(0))X_j| \leqslant Am^2,$$

and it is sufficient to prove that

$$|\mathrm{E}(H'(\tau\bar{X}_i) - H'(0))X_i(H'(\tau\bar{X}_j) - H'(0))X_j| \leqslant AmN^{-1}, \qquad \text{for } |i-j| \leqslant 2m, \tag{3.3}$$

and

$$|\mathrm{E}(H'(\tau\bar{X}_i) - H'(0))X_i(H'(\tau\bar{X}_j) - H'(0))X_j| \leqslant Am^2N^{-2}, \qquad \text{for } |i-j| > 2m. \tag{3.4}$$

Let us prove (3.3). Split the sum $\bar{X}_i = \bar{X}_{ij} + \Delta_j$, with Δ_j denoting the sum of $N^{-1}X_l$ in $\bar{X}_i$ such that $|l-j| \leqslant m$. We may also write a similar expression for $\bar{X}_j$ with Δ_i. Expanding in powers of Δ_j and of Δ_i, we see that (3.3) follows from

$$|\mathrm{E}(H'(\tau\bar{X}_{ij}) - H'(0))X_i(H'(\tau\bar{X}_{ij}) - H'(0))X_j| \leqslant AmN^{-1}, \qquad \text{for } |i-j| \leqslant 2m.$$

By (1.5) we may replace the pair $\eta \overset{\text{def}}{=} (X_i, X_j)$ by its independent copy, say $(\hat{X}_i, \hat{X}_j)$, use the Taylor expansion and reduce (3.3) to

$$\sum_{l,k} |\mathrm{E}H''(\tau\tau_1\bar{X}_{ij})X_l\hat{X}_i H''(\tau\tau_2\bar{X}_{ij})X_k\hat{X}_k| \leqslant AmN, \qquad |i-j| \leqslant 2m,$$

where the sum is taken over all l and k present as indices in the sum $\bar{X}_{ij}$. A repetition of the previous arguments concludes the proof of (3.3).

The proof of (3.4) is similar to (although more technical than) that of (3.3), and we omit it. □

Lemma 3.2. *Assume the mixing condition* (1.4). *Let* $\sup_{x,y} |\psi(x, y)| \leqslant D$ *and* $\mathrm{E}\psi(x, X_1) = 0$, *for all x. Then*

$$J \stackrel{\text{def}}{=} \mathrm{E}\left(\sum_{1\leqslant i< j\leqslant N} \psi(X_i, X_j)\right)^2 \leqslant Am^2N^2L + AD^2N^4\beta(m),$$

where $L \stackrel{\text{def}}{=} \max_{1\leqslant i< j\leqslant N} \mathrm{E}\psi^2(X_i, X_j)$.

Proof. We may write $J \stackrel{\text{def}}{=} \mathrm{E}(\sum_B \psi(X_i, X_j))^2$, where the sum is taken over all two-point subsets $B = \{i, j\}$ such that $B \subset \{1, \ldots, N\}$. Let B_m denote the m-neighbourhood of $B \subset \{1, \ldots, N\}$, that is, $B_m \stackrel{\text{def}}{=} \{s \in \{1, \ldots, N\}: |i - S| \leqslant m \text{ or } |j - s| \leqslant m\}$. Let $d(B) \stackrel{\text{def}}{=} |i - j|$ be the diameter of B. Then $J \leqslant 2J_1 + 2j_2$, where

$$J_1 \stackrel{\text{def}}{=} \mathrm{E}\left(\sum_{d(B)\leqslant m} \psi(X_i, X_j)\right)^2, \qquad \text{and} \qquad J_2 \stackrel{\text{def}}{=} \mathrm{E}\left(\sum_{d(B)>m} \psi(X_i, X_j)\right)^2.$$

The number of two-point subsets $B \subset \{1, \ldots, N\}$ such that $d(B) \leqslant m$ does not exceed mN. Thus $(a_1 + \cdots + a_s)^2 \leqslant s(a_1^2 + \cdots + a_s^2)$ implies $J_1 \leqslant Am^2N^2L$. Furthermore,

$$J_2 = \sum_{d(B)>m, d(D)>m} \mathrm{E}\psi(X_i, X_j)\psi(X_k, X_l),$$

where $B = \{i, j\}$ and $D = \{k, l\}$ denote two-point subsets of $\{1, \ldots, N\}$. The number of pairs of sets B and D such that $D \subseteq B_m$ is bounded from above by Am^2N^2. Therefore, it is sufficient to show that

$$\sum_{d(B)>m, d(D)>m, D\not\subseteq B_m} \mathrm{E}\psi(X_i, X_j)\psi(X_k, X_l) \leqslant AD^2N^4\beta(m). \tag{3.5}$$

But the relations $d(B) > m$, $d(D) > m$ and $D \not\subseteq B_m$ imply that at least one of k or l, say k, satisfies $|k - i| > m$ and $|k - j| > m$. Thus, by (1.5), we may replace X_k by its independent copy $\hat{X}_k$ and (3.5) follows since $\mathrm{E}\psi(\hat{X}_k, x) = 0$, for all x. □

4. Proof of Theorem 1.1

We shall denote by a a generic, sufficiently small positive constant which may depend on K, β, Σ and ρ only. Also write

$$f(t) \stackrel{\text{def}}{=} \mathrm{E}\exp\{\mathrm{i}tT\}, \qquad \phi(t) \stackrel{\text{def}}{=} \exp\{-t^2/2\}.$$

The proof of the theorem combines the techniques of Tikhomirov (1980) for sums of weakly dependent random variables, and those used by Götze (1991), Bentkus *et al.* (1997) in the i.i.d. case for symmetric statistics.

While proving the theorem we may assume that $|g(X_j)| \leqslant \sqrt{N}$. Otherwise we may replace $g(X_j)$ by

$$g(X_j)\mathbf{1}\{|g(X_j)| \leqslant \sqrt{N}\} - \mathrm{E}g(X_j)\mathbf{1}\{|g(X_j)| \leqslant \sqrt{N}\}.$$

Without loss of generality, we shall assume as well that $\sigma_N^2 = \mathrm{E}S^2 = 1$.

We shall prove that the characteristic function f for $|t| \leqslant a\sqrt{N}/\log^2 N = a\sqrt{N}/m^2$ satisfies the ordinary differential equation,

$$f'(t) = -tf(t) + \varepsilon(t)f(t) + \varepsilon_0(t), \qquad f(0) = 1, \tag{4.1}$$

with some functions ε and ε_0 such that

$$|\varepsilon(t)| \leqslant \frac{Am^2t^2}{\sqrt{N}}, \tag{4.2}$$

and

$$|\varepsilon_0(t)| \leqslant AN^{1/2}|t|m^2\gamma + A\mathrm{E}|R| + AN^{-1/2} + Am|t|N^{-1/2}. \tag{4.3}$$

The equation has the unique solution

$$f(t) = \phi(t)\exp\left\{\int_0^t \varepsilon(u)\,\mathrm{d}u\right\} + \phi(t)\int_0^t \exp\left\{\frac{u^2}{2} + \int_u^t \varepsilon(z)\,\mathrm{d}z\right\}\varepsilon_0(u)\,\mathrm{d}u. \tag{4.4}$$

Let us derive Theorem 1.1 from (4.4). It follows from (4.4) that

$$|f(t) - \phi(t)| \leqslant I_1 + I_2,$$

where

$$I_1 \stackrel{\text{def}}{=} \phi(t)\left|\exp\left\{\int_0^t \varepsilon(u)\,\mathrm{d}u\right\} - 1\right| \leqslant Am^2N^{-1/2}t^2\exp\{-t^2/4\}, \tag{4.5}$$

and

$$I_2 \stackrel{\text{def}}{=} \phi(t)\int_0^t \exp\left\{\frac{u^2}{2} + \int_u^t \varepsilon(z)\,\mathrm{d}z\right\}|\varepsilon_0(u)|\,\mathrm{d}u$$
$$\leqslant A(N^{1/2}m^2 + mN^{-1/2})\gamma\min\{1;\, |t|\} + A(\mathrm{E}|R| + N^{-1/2})\min\{|t|^{-1};\, |t|\}. \tag{4.6}$$

Estimates (4.5) and (4.6), together with Esseen's inequality for characteristic functions, imply the result of the theorem.

In order to prove the inequality in (4.5), apply $|\exp\{z\} - 1| \leqslant |z|\exp\{|z|\}$ and (4.2) on the interval $|t| \leqslant a\sqrt{N}/m^2$ with a sufficiently small a. Similarly, the estimate for I_2 is derived using $|\int_u^t \varepsilon(z)\,\mathrm{d}z| \leqslant aA(t^2 - u^2)$ and $aA < 1/4$.

It remains to prove (4.1)–(4.3). We shall write $B \sim D$ if $B = D + \varepsilon(t)f(t) + \varepsilon_0(t)$, for $|t| \leqslant a\sqrt{N}/m^2$, with some functions ε and ε_0 bounded as in (4.2) and (4.3). Thus we have to prove that $f'(t) \sim -tf(t)$. Differentiating, we have

$$f'(t) = \mathrm{i}\mathrm{E}S\exp\{\mathrm{i}tT\} + \mathrm{i}\mathrm{E}R\exp\{\mathrm{i}tT\} \sim I_3 \stackrel{\text{def}}{=} \frac{\mathrm{i}}{\sqrt{N}}\sum_{j=1}^N \mathrm{E}g(X_j)\exp\{\mathrm{i}tT\},$$

with an error bounded by $\mathrm{E}|R|$.

Define $S_{j,0} \stackrel{\text{def}}{=} S$ and

$$\Delta_{j,1} \stackrel{\text{def}}{=} N^{-1/2} \sum_{l \in \Omega_1} g(X_l) \qquad \text{with} \qquad \Omega_1 \stackrel{\text{def}}{=} \{l: 1 \leqslant l \leqslant N, |l-j| \leqslant m\}.$$

Put $S_{j,1} \stackrel{\text{def}}{=} S - \Delta_{j,1}$. By induction we may define

$$\Delta_{j,s} \stackrel{\text{def}}{=} N^{-1/2} \sum_{l \in \Omega_s} g(X_l) \qquad \text{with} \qquad \Omega_s \stackrel{\text{def}}{=} \{l: 1 \leqslant l \leqslant N, (s-1)m < |l-j| \leqslant sm\},$$

and $S_{j,s} \stackrel{\text{def}}{=} S_{j,s-1} - \Delta_{j,s}$. Furthermore, for a natural $r = A_1 \log N$ (to be chosen later),

$$Q_j \stackrel{\text{def}}{=} R_{j-rm,j+rm}, \qquad \delta_j \stackrel{\text{def}}{=} R - Q_j, \qquad T_{j,s} \stackrel{\text{def}}{=} S_{j,s} + Q_j.$$

An application of $(a_1 + \cdots + a_p)^{3/2} \leqslant \sqrt{p}(a_1^{3/2} + \cdots + a_p^{3/2})$ and of the triangle inequality implies

$$\mathrm{E}|\delta_j|^{3/2} \leqslant Am^3\gamma^{3/2}. \tag{4.7}$$

Taylor's expansion in powers of δ_j and an application of (4.7), together with the Hölder inequality, show that

$$I_3 \sim I_4 \stackrel{\text{def}}{=} \frac{\mathrm{i}}{\sqrt{N}} \sum_{j=1}^{N} \mathrm{E}\, g(X_j) \exp\{\mathrm{i}tT_{j,0}\},$$

with an error bounded by $A|t|m^2N^{1/2}\gamma$.

Splitting $T_{j,0} = T_{j,1} + \Delta_{j,1}$, we may write

$$I_4 = \frac{\mathrm{i}}{\sqrt{N}} \sum_{j=1}^{N} \mathrm{E}\, g(X_j) J_0 \exp\{\mathrm{i}tT_{j,1}\} + \frac{\mathrm{i}}{\sqrt{N}} \sum_{j=1}^{N} \mathrm{E}\, g(X_j) J_1 \exp\{\mathrm{i}tT_{j,1}\},$$

where $J_0 \stackrel{\text{def}}{=} 1$ and $J_s \stackrel{\text{def}}{=} \exp\{\mathrm{i}t\Delta_{j,s}\} - 1$, for $s \geqslant 1$. Repeating the procedure, we obtain (we shall choose $r = A_1 \log N$)

$$f'(t) \sim I_4 = \sum_{s=0}^{r-1} \sum_{j=1}^{N} I(j, s) + \sum_{j=1}^{N} I_1(j, r), \qquad \text{for any natural } r, \tag{4.8}$$

with

$$I(j, s) \stackrel{\text{def}}{=} \frac{\mathrm{i}}{\sqrt{N}} \mathrm{E}\, g(X_j) J_1 \ldots J_s \exp\{\mathrm{i}tT_{j,s+1}\},$$

$$I_1(j, r) \stackrel{\text{def}}{=} \frac{\mathrm{i}}{\sqrt{N}} \mathrm{E}\, g(X_j) J_1 \ldots J_r \exp\{\mathrm{i}tT_{j,r}\}.$$

The expectation in $I(j, s)$ is taken over a function of a set of random variables which does not contain X_l with indices l such that $sm < |l-j| \leqslant (s+1)m$. Thus, by (1.5), we may replace $\exp\{\mathrm{i}tT_{j,s+1}\}$ by its independent copy, and obtain

$$\sum_{s=0}^{r-1}\sum_{j=1}^{N} I(j,\, s) \sim \sum_{s=0}^{r-1}\sum_{j=1}^{N} I_2(j,\, s), \tag{4.9}$$

with

$$I_2(j,\, s) \stackrel{\text{def}}{=} \frac{\mathrm{i}}{\sqrt{N}} \mathrm{E}\, g(X_j) J_1 \ldots J_s \mathrm{E}\exp\{\mathrm{i}tT_{j,s+1}\}.$$

Indeed, $|J_s| \leqslant 2$ and $|g(X_j)| \leqslant \sqrt{N}$, and the error in (4.9) is bounded from above by $A\sum_{s=0}^{r-1} N2^s\beta(m) \leqslant N2^r\beta(m) \leqslant AN^{-1/2}$ since $r = A_1 \log N$ and since we may choose the constant $A = A(A_1)$ in $m = A\log N$ sufficiently large.

We have

$$I_2(j,\, 0) = 0 \qquad \text{since } \mathrm{E}\, g(X_j) = 0. \tag{4.10}$$

Collecting J_l with odd or even l into separate groups, applying the Hölder inequality and then, by (1.5), replacing the multipliers $|J_l| \leqslant 2$ by their independent copies, we obtain

$$(\mathrm{E}|J_1 \ldots J_s|^{3/2})^{2/3} \leqslant \prod_{l=1}^{s} (\mathrm{E}|J_l|^3)^{1/3} + A3^s\beta^{1/3}(m), \qquad s \geqslant 1. \tag{4.11}$$

The choices of m and r and $|J_s| \leqslant |t\Delta_{j,s}|$, together with (4.11), imply

$$(\mathrm{E}|J_1 \ldots J_s|^{3/2})^{2/3} \leqslant A \min\{(a/m)^s;\, (m|t|N^{-1/2})^s\} + AN^{-3}. \tag{4.12}$$

Relation (4.8) implies that

$$f'(t) \sim \sum_{s=1}^{r-1}\sum_{j=1}^{N} I_2(j,\, s), \qquad \text{since } \sum_{j=1}^{N} I_1(j,\, r) \sim 0. \tag{4.13}$$

To prove (4.13), use (4.10) and bound $I_1(j,\, r)$ using (4.12). The error in the transition from (4.8) to (4.13) is bounded by $AN^{-1/2}$.

Denote $T'_{j,s} \stackrel{\text{def}}{=} T_{j,s} + \delta_j = S_{j,s} + R$. Then the relation (4.13) implies

$$f'(t) \sim \sum_{s=1}^{r-1}\sum_{j=1}^{N} I_3(j,\, s), \qquad I_3(j,\, s) \stackrel{\text{def}}{=} \frac{\mathrm{i}}{\sqrt{N}} \mathrm{E}\, g(X_j) J_1 \ldots J_s \mathrm{E}\exp\{\mathrm{i}tT'_{j,s+1}\}. \tag{4.14}$$

Indeed, it suffices to expand in powers of δ_j and to use (4.7) and (4.12). The error in the transition from (4.13) to (4.14) is bounded by $A|t|m^2\sqrt{N}\gamma$.

Define

$$I_4(s) \stackrel{\text{def}}{=} \mathrm{i}\mathrm{E}\, g(X_j) J_1 \ldots J_s$$

and notice that, for $j \in \Gamma_s \stackrel{\text{def}}{=} \{j\colon sm < j < N - sm\}$, the expectation $I_4(s)$ is independent of j since the sequence $X_1, X_2, \ldots$ is stationary. Thus

$$\sum_{j\in\Gamma_s} I_3(j,\, s) = I_4(s) \sum_{j\in\Gamma_s} N^{-1/2}\mathrm{E}\exp\{\mathrm{i}tT'_{j,s+1}\}. \tag{4.15}$$

Using (4.12), we have

$$\sum_{s=1}^{r-1} \sum_{1\leqslant j\leqslant N, j\notin\Gamma_s} N^{-1/2}\mathrm{E}|g(X_j)J_1 \ldots J_s| \leqslant AN^{-1/2}\sum_{s=1}^{r-1} ms(a/m)^s \leqslant AN^{-1/2}. \tag{4.16}$$

Due to (4.15) and (4.16), we derive from (4.14) that

$$f'(t) \sim \sum_{s=1}^{r-1} \sqrt{N} I_4(s) f_s(t), \qquad f_s(t) \stackrel{\text{def}}{=} \frac{1}{N}\sum_{j=1}^{N} \mathrm{E}\exp\{\mathrm{i}tT'_{j,s+1}\}. \tag{4.17}$$

The error in the transition from (4.14) to (4.17) is bounded by $AN^{-1/2}$.

Let us show that

$$|f_s(t) - f(t)| \leqslant AmsN^{-1/2}|tf(t)| + AsN^{-1/2}. \tag{4.18}$$

Define $\mu_{j,s} \stackrel{\text{def}}{=} S - S_{j,s+1}$. Then $T'_{j,s+1} = T - \mu_{j,s}$, and

$$f_s(t) - f(t) = D_1 f(t) + D_2,$$

with

$$D_1 \stackrel{\text{def}}{=} \frac{1}{N}\sum_{j=1}^{N} \mathrm{E}\exp\{-\mathrm{i}t\mu_{j,s}\} - 1,$$

$$D_2 \stackrel{\text{def}}{=} \frac{1}{N}\mathrm{E}\sum_{j=1}^{N}(\xi_j - \mathrm{E}\xi_j)\exp\{\mathrm{i}tT\}, \qquad \xi_j \stackrel{\text{def}}{=} \exp\{-\mathrm{i}t\mu_{j,s}\}.$$

Expanding in powers of $t\mu_{j,s}$, we obtain

$$|D_1| \leqslant \frac{1}{N}\sum_{j=1}^{N} \mathrm{E}|t\mu_{j,s}| \leqslant Ams|t|N^{-1/2}.$$

By the Hölder inequality,

$$|D_2|^2 \leqslant N^{-2}\mathrm{E}\left|\sum_{j=1}^{N}(\xi_j - \mathrm{E}\xi_j)\right|^2 = N^{-2}\sum_{j=1}^{N}\sum_{k=1}^{N} \mathrm{E}(\xi_j - \mathrm{E}\xi_j)(\bar{\xi}_k - \mathrm{E}\bar{\xi}_k) \leqslant As^2N^{-1},$$

since in the last sum summands, say $P_{j,k}$, with indices j and k such that $|j-k| \leqslant 4ms$ satisfy

$$|P_{j,k}| \leqslant (\mathrm{E}|\xi_j - \mathrm{E}\xi_j|^2\mathrm{E}|\xi_k - \mathrm{E}\xi_k|^2)^{1/2} \leqslant At^2(\mathrm{E}|\mu_{j,s}|^2\mathrm{E}|\mu_{j,s}|^2)^{1/2} \leqslant As^2m^{-2},$$

and since $|P_{j,k}| \leqslant A/N$ in the case $|j-k| > 4ms$ (to see this, apply (1.5)). Collecting these estimates, we obtain (4.18).

Relations (4.17), (4.18) and (4.12) together imply

$$f'(t) \sim \sum_{s=1}^{r-1} \sqrt{N} I_4(s) f(t). \tag{4.19}$$

The error in replacing (4.17) by (4.19) is bounded from above by $Am^2t^2N^{-1/2}|f(t)| + Am|t|N^{-1/2}$. By (4.12),

$$\sum_{s=2}^{r-1} |I_4(s)| \leqslant At^2 m^2 N^{-1}$$

and

$$f'(t) \sim \sqrt{N} I_4(1) f(t) = \mathrm{i}\sqrt{N} \mathrm{E} g(X_j) J_1 f(t).$$

A Taylor expansion applied to J_1 and a comparison of the coefficient of t with $-1 = -\mathrm{E}S^2$ show that (4.19) implies that $f'(t) \sim -tf(t)$, with an error bounded by $AN^{-1/2}(1 + m^2 t^2)|f(t)|)$, which concludes the proof of the theorem.

Acknowledgements

The authors gratefully acknowledge the support of the German Science Foundation Sonderforschungsbereich (SFB) 343 in Bielefeld, and of grant N93-011-1454 of the Russian Foundation for Fundamental Research.

We would also like to thank the referees for helpful suggestions concerning the presentation of the results.

References

Bentkus, V. and Götze, F. (1996) The Berry–Esseen bound for Student's statistics. *Ann. Probab.*, **24**, 491–503.

Bentkus, V., Götze, F. and Zitikis, R. (1994) Lower estimates of the convergence rate for *U*-statistics. *Ann. Probab.*, **22**, 1707–1714.

Bentkus, V., Götze, F. and van Zwet, W. (1997) An Edgeworth expansion for symmetric statistics. *Ann. Statist.*, **25**, 851–896.

Bolthausen, E. and Götze, F. (1993) The rate of convergence for multivariate sampling statistics. *Ann. Statist.*, **21**, 1692–1710.

Cartan, H. (1971) *Calcul Differentiel. Formes Differentielles.* Paris: Hermann.

Denker, M. (1982) Statistical decision procedures and ergodic theory. In *Ergodic Theory and Related Topics. Mathematical Results*, Vitte (GDR), 1981, vol. 12, pp. 35–47. Berlin: Akademie-Verlag.

Denker, M. and Keller, G. (1983) On *U*-statistics and von Mises statistics for weakly dependent processes. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **64**, 505–522.

Eberlein, E. (1984) Weak convergence of partial sums of absolutely regular sequences. *Statist. Probab. Lett.*, **2**, 291–293.

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.

Götze, F. and Hipp, C. (1983) Asymptotic expansions for sums of weakly dependent random vectors. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **64**, 211–239.

Götze, F. and Hipp, C. (1994) Asymptotic distribution of statistics in time series. *Ann. Statist.*, **22**, 2062–2088.

Govindarajulu, Z. and Mason, D.M. (1983) A strong representation for linear combinations of order statistics with application to fixed-width confidence intervals for location and scale parameters. *Scand. J. Statist.*, **10**, 97–115.

Heinrich, L. (1992) Bounds for the absolute regularity coefficient of a stationary renewal process. *Yokohama Math. J.*, **40**, 25–33.

Ibragimov, I.A. (1975) A note on the CLT for dependent random variables. *Theory Probab. Appl.*, **20**, 135–141.

Ibragimov, I.A. and Linnik, Yu. V. (1971) *Independent and Stationary Sequences of Random Variables.* Groningen: Wolters-Noordhoff.

Rio, E. (1996) Sur le théorème de Berry–Esseen pour les suites faiblement dpendantes. *Probab. Theory Related Fields*, **104**, 255–282.

Stein, C. (1972) A bound on the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 2, pp. 583–602. Berkeley: University of California Press.

Sunklodas, J. (1991) Approximation of distributions of sums of weakly dependent random variables by the normal distribution. In R.V. Gamkrelidze, Yu.V. Prekhorov and V. Statulevičius (eds), *Limit Theorems of Probability Theory*, Vol. 81, pp. 140–199. Moscow: VINITI.

Tikhomirov A.N. (1980) On the rule of convergence in the central limit theorem for weakly dependent variables. *Theory Probab. Appl.*, **25**, 790–809.

van Zwet W.R. (1984) A Berry–Esseen bound for symmetric statistics. *Z. Wahrscheinlichkeitstheorie. Verw. Geb.*, **66**, 425–440.

Veretennikov A.Yu. (1987) Bounds for the mixing rate in the theory of stochastic equations. *Theory. Probab. Appl.*, **32**, 273–281.

Yoshihara K.I. (1976) Limit behavior for stationary absolutely regular processes. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **35**, 237–252.

Yoshihara K.-I. (1984) The Berry–Esseen theorems for *U*-statistics generated by absolutely regular processes. *Yokohama Math. J.*, **32**, 89–111.

Received May 1995 and revised July 1996