Z. Wahrscheinlichkeitstheorie verw. Gebiete
65, 523–534 (1984)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete


# An Exact Rate of Convergence in the Functional Central Limit Theorem for Special Martingale Difference Arrays

Erich Haeusler

University of Munich, Mathematical Institute, Theresienstrasse 39, D-8000 Munich 2, Federal Republic of Germany

**Summary.** Since the topology of weak convergence of probability distributions on the Borel $\sigma$-field of the space $C=C([0,1])$ is metrizable, it is natural to describe the speed of convergence in weak functional limit theorems by means of an appropriate metric. Using the metric proposed by Prokhorov it is shown that under suitable conditions the rate of convergence in the functional central limit theorem for $C$-valued partial sum processes based on martingale difference arrays is the same as in the special case of row-wise independent random variables where this rate is known to be an optimal one.

## 1. Introduction and Main Results

The speed of convergence in functional central limit theorems may be measured in various ways. A very natural one originated in [10] is by means of a metric for weak convergence of probability distributions on function spaces. Using this approach we give the exact rate of convergence in the functional central limit theorem for partial sum processes based on special martingale difference arrays.

Let us consider a finite sequence $\xi_1, \dots, \xi_k$ of square integrable random variables on the probability space $(\Omega, \mathscr{F}, P)$ and sub-$\sigma$-fields $\mathscr{F}_0 \subset \mathscr{F}_1 \subset \dots \subset \mathscr{F}_k$ of $\mathscr{F}$ such that $\xi_i$ is measurable w.r.t. $\mathscr{F}_i$ for $i=1, \dots, k$. Furthermore, assume that $(\xi_i, \mathscr{F}_i)$ is a martingale difference sequence, viz. $E(\xi_i|\mathscr{F}_{i-1})=0$ for $i=1, \dots, k$, where here and throughout the present paper all equalities and inequalities for random variables are supposed to hold almost surely. For brevity we write $V_j=\sum_{i=1}^{j} E(\xi_i^2|\mathscr{F}_{i-1})$ and assume

$$V_k=\sum_{i=1}^{k} E(\xi_i^2|\mathscr{F}_{i-1})=1. \tag{1}$$

This restriction has been used by several authors; for example, it occurs in some results of the recent paper [3] by Bolthausen which presents sharp rates of convergence in the ordinary central limit theorem for martingales.

The partial sum process $S_{(k)}$ based on $(\xi_i, \mathscr{F}_i)$ is defined by $S_{(k)}(t)=\sum_{i=1}^{j}\xi_i$ if $t=V_j$, $j=0,1,\ldots,k$, and by linear interpolation on the subintervals $[V_{j-1}, V_j]$, $j=1,\ldots,k$. Then $S_{(k)}$ is an almost surely well defined random element taking its values in the space $C=C([0,1])$ of all real valued functions on the unit interval endowed with the topology of uniform convergence and the corresponding Borel $\sigma$-field $\mathscr{B}(C)$.

In functional central limit theory one considers a whole sequence $(\xi_{ni}, \mathscr{F}_{ni})_{1\leqq i\leqq k_n}$, $n\in N$, of martingale difference sequences which is usually called a martingale difference array (m.d.a. for short). The interest is in the asymptotic behaviour of the partial sum processes $S_{(k_n)}$ pertaining to $(\xi_{ni}, \mathscr{F}_{ni})_{1\leqq i\leqq k_n}$ if $n$ tends to infinity. Several sufficient conditions are known for

$$S_{(k_n)} \xrightarrow{\mathscr{L}} B_1 \quad \text{as } n\to\infty, \tag{2}$$

i.e. for convergence in law of $S_{(k_n)}$ to a Brownian motion $B_1$ with time interval $[0,1]$ where convergence in law is to be understood as weak convergence of the distributions $P\circ S_{(k_n)}^{-1}$ (induced from $P$ by $S_{(k_n)}$) to the distribution $W$ of $B_1$, which is the Wiener measure on $\mathscr{B}(C)$. As shown in [10], weak convergence in $C$ is compatible with the now so-called Prokhorov metric $\rho$ which for two probability measures $Q_1$ and $Q_2$ on $\mathscr{B}(C)$ is defined by

$$\rho(Q_1, Q_2)=\inf\{\varepsilon>0: Q_1(B)\leqq Q_2(B^\varepsilon)+\varepsilon \quad \text{for all} \quad B\in\mathscr{B}(C)\}.$$

Here $B^\varepsilon=\{f\in C: d(f,B)<\varepsilon\}$, where $d(f,B)$ denotes the uniform distance of $f$ and $B$. Since (2) is equivalent to

$$\rho(P\circ S_{(k_n)}^{-1}, W)\to 0 \quad \text{as } n\to\infty, \tag{3}$$

the rate of convergence in (3) is a natural measure of the speed in (2). Furthermore, any estimate of $\rho(P\circ S_{(k_n)}^{-1}, W)$ implies estimates of Berry-Esséen-type for real valued functions of $S_{(k_n)}$ and $B_1$ satisfying certain Lipschitz conditions, and also for $|P(S_{(k_n)}\in B)-W(B)|$ if the boundary of $B\in\mathscr{B}(C)$ satisfies a Lipschitz condition; for details, see [4], p. 208.

To obtain a rate in (3) one has to produce a bound of $\rho(P\circ S_{(k_n)}^{-1}, W)$ for each fixed $n$. Consequently, it suffices to deal with a single row of the array $(\xi_{ni}, \mathscr{F}_{ni})$, i.e. to consider a single finite martingale difference sequence $\xi_1,\ldots,\xi_k$ as described at the beginning. Then our main result is the following

**Theorem 1.** *If* (1) *holds and if there exist a* $\delta\in(0,3/2)$ *and a real number* $L_k$ *such that*

$$\sum_{i=1}^{k} E(|\xi_i|^{2+2\delta}|\mathscr{F}_{i-1})\leqq L_k, \tag{4}$$

*then*

$$\rho(P\circ S_{(k)}^{-1}, W)\leqq K_\delta L_k^{1/(3+2\delta)}$$

*for a finite constant* $K_\delta$ *depending only on* $\delta$.

*Remark.* If (1) and (4) are satisfied for some $\delta \geqq 3/2$, then

$$\rho(P \circ S_{(k)}^{-1}, W) \leqq K_\delta (L_k^{1/4\delta} |\log L_k|^{3/4})$$

for a finite constant $K_\delta$ depending only on $\delta$.

For independent variables and $\delta \in (0, 1/2]$ the assertion of Theorem 1 has been established by Borovkov in Theorem 1 of [4], where, of course, $L_k = \sum_{i=1}^{k} E(|\xi_i|^{2+2\delta})$. An example in [1] shows that this result is optimal for all $\delta > 0$. Therefore, no better result can be obtained in the setup of Theorem 1, too. In the martingale case Hall and Heyde have proved estimates under less stringent conditions than (1) and (4), see [6], Theorem 4.5, but their bound turns out to be weaker than the optimal one when specialized to the situation considered here. Borovkov's proof is based on a quantile approximation which seems not to be applicable in the dependent case. Hall and Heyde used the martingale version of the Skorokhod embedding scheme following the approach of Rosenkrantz in [11] in the independent case. Our proof is via the Skorokhod embedding, too, but a truncation is added and an appropriate maximal inequality for martingale difference sequences is employed to achieve the optimal estimate for all $\delta \in (0, 3/2)$ whereas Hall and Heyde's Theorem 4.5 is restricted to the range $\delta \in (0, 1]$. Since it is known that a rate of order $O(n^{-1/4})$ in the functional central limit theorem cannot be obtained by means of Skorokhod's embedding, see [4], p. 224, this method does not allow to prove the assertion of Theorem 1 for $\delta \geqq 3/2$ (observe that $\left(\sum_{i=1}^{n} E(|\xi_{ni}|^5)\right)^{1/6} = \text{const} \cdot n^{-1/4}$ if $\xi_{ni} = n^{-1/2} \xi_i$ for a sequence $(\xi_i)_{i \in N}$ of independent and identically distributed random variables such that $E(\xi_1) = 0$, $E(\xi_1^2) = 1$ and $E(|\xi_1|^5) < \infty$).

For a m.d.a. which arises in the usual way from a single martingale difference sequence Theorem 1 provides the rate $O(n^{-\delta/(3+2\delta)})$ under suitable conditions. If in addition the variables are identically distributed, this result can be strengthened to some extent.

**Theorem 2.** *Let $(\xi_i)_{i \in N}$ be a martingale difference sequence w.r.t. $(\mathscr{F}_i)_{i \geqq 0}$ such that $E(\xi_i^2 | \mathscr{F}_{i-1}) = 1$ and $E(|\xi_i|^{2+2\delta} | \mathscr{F}_{i-1}) \leqq K_0 < \infty$ for some $\delta \in (0, 3/2)$ and all $i \in N$. If the variables $\xi_i$ are identically distributed, then*

$$\rho(P \circ S_{(n)}^{-1}, W) = o(n^{-\delta/(3+2\delta)}) \quad \text{as } n \to \infty$$

*where the partial sum processes $S_{(n)}$ are based on the m.d.a.*

$$(\xi_{ni}, \mathscr{F}_{ni})_{1 \leqq i \leqq n}, n \in N, \quad \text{with} \quad \xi_{ni} = n^{-1/2} \xi_i \quad \text{and} \quad \mathscr{F}_{ni} = \mathscr{F}_i.$$

According to an example in [8], the rate $o(n^{-\delta/(3+2\delta)})$ is optimal even in the special case of independent, identically distributed random variables.

*Example 1. Chain Dependent Random Variables*

Let $(J_n)_{n \geqq 0}$ be a Markov chain with finite state space $I$ and transition matrix $(p_{jk})_{j,k \in I}$. A sequence $(\xi_n)_{n \in N}$ of random variables is chain dependent w.r.t. $(J_n)_{n \geqq 0}$ if for all $j \in I$, $t \in R$ and $n \in N$

$$P(\{J_n=j\}\cap\{\xi_n\leqq t\}|\mathscr{I}_{n-1})=P(\{J_n=j\}\cap\{\xi_n\leqq t\}|J_{n-1})=p_{J_{n-1},j}F_{J_{n-1},j}(t),$$

where $\mathscr{I}_{n-1}$ is the $\sigma$-field generated by $J_0, J_1, \ldots, J_{n-1}, \xi_1, \ldots, \xi_{n-1}$ and where $F_{jk}, j, k\in I$, are given distribution functions. Then

$$P\left(\bigcap_{i=1}^{n}\{\xi_i\leqq t_i\}\,|\,J_0, J_1, \ldots, J_n\right)=\prod_{i=1}^{n}P(\xi_i\leqq t_i\,|\,J_{i-1}, J_i),$$

i.e. the variables $\xi_n$, $n\in N$, are independent given the chain $(J_n)_{n\geqq 0}$. Suppose that $\int tF_{jk}(dt)=0$, $\int t^2 F_{jk}(dt)=1$ and $\int |t|^{2+2\delta}F_{jk}(dt)<\infty$. Then $F_j=\sum_{k\in I}p_{jk}F_{jk}$ is a distribution function for every $j\in I$ such that $\int tF_j(dt)=0$ and $\int t^2 F_j(dt)=1$. Furthermore, $\xi_{ni}=n^{-1/2}\xi_i$, $i=1, \ldots, n$, $n\in N$, and $\mathscr{F}_{ni}=\mathscr{I}_i$, $i=0, 1, \ldots, n$, $n\in N$, form a m.d.a. $(\xi_{ni}, \mathscr{F}_{ni})$ such that (1) and (4) are satisfied for each row:

$$\sum_{i=1}^{n}E(\xi_{ni}^2\,|\,\mathscr{F}_{n,i-1})=n^{-1}\sum_{i=1}^{n}E(\xi_i^2\,|\,\mathscr{I}_{i-1})=n^{-1}\sum_{i=1}^{n}\int t^2 F_{J_{i-1}}(dt)=1$$

and

$$\begin{aligned}&\sum_{i=1}^{n}E(|\xi_{ni}|^{2+2\delta}\,|\,\mathscr{F}_{n,i-1})\\&\quad=n^{-1-\delta}\sum_{i=1}^{n}E(|\xi_i|^{2+2\delta}\,|\,\mathscr{I}_{i-1})=n^{-1-\delta}\sum_{i=1}^{n}\int|t|^{2+2\delta}F_{J_{i-1}}(dt)\\&\quad=n^{-1-\delta}\sum_{i=1}^{n}\sum_{j\in I}p_{J_{n-1},j}\int|t|^{2+2\delta}F_{J_{n-1},j}(dt)\leqq K_\delta n^{-\delta}\end{aligned}$$

where $K_\delta\leqq|J|\max_{j,k\in I}p_{jk}\int|t|^{2+2\delta}F_{jk}(dt)<\infty$. Thus for $0<\delta<3/2$ by Theorem 1

$$\rho(P\circ S_{(n)}^{-1}, W)=O(n^{-\delta/(3+2\delta)})\qquad\text{as}\quad n\to\infty,$$

where $S_{(n)}$ denotes the partial sum process pertaining to $(\xi_{ni}, \mathscr{F}_{ni})_{1\leqq i\leqq n}$.

*Example 2. Stationary Linear Processes*

Following [7] we consider the stationary linear process

$$\xi_n=\sum_{j=0}^{\infty}\beta_j\varepsilon_{n-j},\qquad n\in N,\tag{5}$$

where $\beta_j$, $j\geqq 0$, are real numbers and where $(\varepsilon_j)_{j\in Z}$ is a stationary and ergodic martingale difference sequence w.r.t. the $\sigma$-fields $\mathscr{F}_j$, $j\in Z$, generated by the variables $\varepsilon_i$, $i\leqq j$. Suppose that $E(\varepsilon_i^2|\mathscr{F}_{i-1})=1$ and $E(|\varepsilon_i|^{2+2\delta}|\mathscr{F}_{i-1})\leqq K_0<\infty$ for each $i\in Z$ (by stationarity it is enough to consider $i=0$) and some $\delta\in(0, 3/2)$. If $\sum_{j=0}^{\infty}\beta_j^2<\infty$, then the series in (5) converges almost surely. Let us assume here the more stringent conditions

$$\sum_{j=0}^{\infty}\beta_j\quad\text{and}\quad\sum_{j=1}^{\infty}\left(\sum_{k=j}^{\infty}\beta_k\right)^2\quad\text{are convergent, and}\quad\sigma=\sum_{j=0}^{\infty}\beta_j\neq 0.$$

Put $\gamma_j = \sum_{k=j}^{\infty} \beta_k$ for $j \in N$. Then $\eta_n = \sum_{j=1}^{\infty} \gamma_j \varepsilon_{n-j}$ is well defined and, using the dominated and the monotone convergence theorem and a moment inequality of Burkholder, cf. Theorem 2.11 in [6], we obtain with some finite constant $K$ depending only on $\delta$

$$\begin{aligned}
&E(|\eta_n|^{2+2\delta})\\
&\leqq E\left(\max_{m\in N}\left|\sum_{j=1}^{m} \gamma_j \varepsilon_{n-j}\right|^{2+2\delta}\right) = \lim_{M\to\infty} E\left(\max_{1\leqq m\leqq M}\left|\sum_{j=n-m}^{n-1} \gamma_{n-j}\varepsilon_j\right|^{2+2\delta}\right)\\
&\leqq \limsup_{M\to\infty} K\left[E\left(\left(\sum_{j=n-M}^{n-1} \gamma_{n-j}^2 E(\varepsilon_j^2|\mathscr{F}_{j-1})\right)^{1+\delta}\right) + \mathrm{E}\left(\max_{n-M\leqq j\leqq n-1} |\gamma_{n-j}\varepsilon_j|^{2+2\delta}\right)\right]\\
&\leqq \limsup_{M\to\infty} K\left[\left(\sum_{j=1}^{M} \gamma_j^2\right)^{1+\delta} + \sum_{j=1}^{M} |\gamma_j|^{2+2\delta} E(|\varepsilon_{n-j}|^{2+2\delta})\right]\\
&\leqq K\left[\left(\sum_{j=1}^{\infty}\gamma_j^2\right)^{1+\delta} + K_0(\max_{j\in N}|\gamma_j|^{2\delta})\sum_{j=1}^{\infty}\gamma_j^2\right] < \infty.
\end{aligned}$$

Obviously, $\xi_n = \sigma\varepsilon_n + \eta_n - \eta_{n+1}$ for all $n \in N$. Define the partial sum processes $S_{(n)}$ and $S'_{(n)}$ by $S_{(n)}(t) = \sigma^{-1} n^{-1/2} \sum_{i=1}^{k} \xi_i$ and $S'_{(n)}(t) = n^{-1/2} \sum_{i=1}^{k} \varepsilon_i$ if $t = kn^{-1}$ for $k = 0, 1, \ldots, n$, and by linear interpolation on the intervals $[(k-1)n^{-1}, kn^{-1}]$ for $k = 1, \ldots, n$. By Theorem 2

$$\rho(P \circ S'^{-1}_{(n)}, W) = o(n^{-\delta/(3+2\delta)}) \quad \text{as } n \to \infty,$$

whereas for all $\varepsilon > 0$ by inequality (1.7) in [8]

$$\begin{aligned}
\rho(P \circ S^{-1}_{(n)}, P \circ S'^{-1}_{(n)}) &\leqq 2\varepsilon + P(\sup_{0\leqq t\leqq 1} |S_{(n)}(t) - S'_{(n)}(t)| \geqq 2\varepsilon)\\
&\leqq 2\varepsilon + P\left(\max_{1\leqq k\leqq n}\left|\sum_{i=1}^{k}(\eta_i - \eta_{i+1})\right| \geqq 2|\sigma|\varepsilon n^{1/2}\right)\\
&\leqq 2\varepsilon + P(\max_{1\leqq k\leqq n} |\eta_{k+1} - \eta_1| \geqq 2|\sigma|\varepsilon n^{1/2})\\
&\leqq K[\varepsilon + \varepsilon^{-2-2\delta} n^{-\delta} E(|\eta_1|^{2+2\delta} I(|\eta_1| \geqq |\sigma|\varepsilon n^{1/2}))].
\end{aligned}$$

Putting

$$\varepsilon = n^{-\delta/(3+2\delta)} [n^{-1/(6+4\delta)} + E(|\eta_1|^{2+2\delta} I(|\eta_1| \geqq |\sigma| n^{1/(3+2\delta)}))]^{1/(4+4\delta)}$$

we get $\rho(P \circ S^{-1}_{(n)}, P \circ S'^{-1}_{(n)}) = o(n^{-\delta/(3+2\delta)})$ as $n \to \infty$, hence

$$\rho(P \circ S^{-1}_{(n)}, W) = o(n^{-\delta/(3+2\delta)}) \quad \text{as } n \to \infty.$$

## 2. Proofs

The following martingale inequality is related to those given in [5].

**Lemma 1.** *If $\xi_1, \ldots, \xi_k$ is a martingale difference sequence w.r.t. the non-decreasing σ-fields $\mathscr{F}_0, \mathscr{F}_1, \ldots, \mathscr{F}_k$, then for all $\gamma$, $u$, $v > 0$*

$$P\left(\max_{1\leqq j\leqq k}\left|\sum_{i=1}^{j}\xi_i\right|\geqq\gamma\right)\leqq\sum_{i=1}^{k}P(|\xi_i|>u)$$
$$+2P\left(\sum_{i=1}^{k}E(\xi_i^2|\mathscr{F}_{i-1})>v\right)+2\exp[\gamma u^{-1}(1-\log(\gamma u v^{-1}))].$$

*Proof.* It suffices to show for all $\gamma$, $u$, $v>0$

$$P\left(\max_{1\leqq j\leqq k}\sum_{i=1}^{j}\xi_i\geqq\gamma\right)\leqq\sum_{i=1}^{k}P(\xi_i>u)$$
$$+P\left(\sum_{i=1}^{k}E(\xi_i^2|\mathscr{F}_{i-1})>v\right)+\exp[\gamma u^{-1}(1-\log(\gamma u v^{-1}))].$$

To this end we write $\eta_i=\xi_i I(\xi_i\leqq u)$ and $\zeta_i=\eta_i-E(\eta_i|\mathscr{F}_{i-1})$ for $i=1,\ldots,k$ and note that $\eta_i\leqq\zeta_i$ because of $E(\eta_i|\mathscr{F}_{i-1})\leqq 0$ which follows from

$$0=E(\xi_i|\mathscr{F}_{i-1})=E(\eta_i|\mathscr{F}_{i-1})+E(\xi_i I(\xi_i>u)|\mathscr{F}_{i-1}).$$

Thus

$$P\left(\max_{1\leqq j\leqq k}\sum_{i=1}^{j}\xi_i\geqq\gamma\right)\leqq P\left(\max_{1\leqq j\leqq k}\sum_{i=1}^{j}\eta_i\geqq\gamma\right)+P\left(\bigcup_{i=1}^{k}\{\eta_i\neq\xi_i\}\right)$$
$$\leqq P\left(\max_{1\leqq j\leqq k}\sum_{i=1}^{j}\zeta_i\geqq\gamma\right)+\sum_{i=1}^{k}P(\xi_i>u).$$

To produce a bound of the first summand we set

$$\phi(\lambda,u)=u^{-2}(\exp(\lambda u)-1-\lambda u)\qquad\text{for all }\lambda,u>0$$

and

$$S_j=\exp\left(\lambda\sum_{i=1}^{j}\zeta_i-\phi(\lambda,u)\sum_{i=1}^{j}E(\eta_i^2|\mathscr{F}_{i-1})\right)\qquad\text{for } j=0,1,\ldots,k\ \ (S_0=1).$$

From [9], p. 155, we know that $\exp(\lambda y)\leqq 1+\lambda y+y^2\phi(\lambda,u)$ for all $\lambda,u>0$ and $-\infty<y\leqq u$. This implies

$$E(\exp(\lambda\eta_{j+1})|\mathscr{F}_j)\leqq 1+\lambda E(\eta_{j+1}|\mathscr{F}_j)+\phi(\lambda,u)\,E(\eta_{j+1}^2|\mathscr{F}_j)$$
$$\leqq\exp(\lambda E(\eta_{j+1}|\mathscr{F}_j))\exp(\phi(\lambda,u)\,E(\eta_{j+1}^2|\mathscr{F}_j)),$$

i.e. $E(\exp(\lambda\zeta_{j+1})|\mathscr{F}_j)\leqq\exp(\phi(\lambda,u)\,E(\eta_{j+1}^2|\mathscr{F}_j))$; hence

$$E(S_{j+1}|\mathscr{F}_j)=S_j\exp(-\phi(\lambda,u)\,E(\eta_{j+1}^2|\mathscr{F}_j))\,E(\exp(\lambda\zeta_{j+1})|\mathscr{F}_j)\leqq S_j.$$

Thus $S_0,S_1,\ldots,S_k$ is a nonnegative supermartingale w.r.t. $\mathscr{F}_0,\mathscr{F}_1,\ldots,\mathscr{F}_k$ for all $\lambda,u>0$. Using the maximal inequality given in Proposition II-2-7 of [9] we obtain

$$P\left(\max_{1\leqq j\leqq k}\sum_{i=1}^{j}\zeta_i\geqq\gamma\right)$$
$$\leqq P(\max_{1\leqq j\leqq k}S_j\geqq\exp(\lambda\gamma-\phi(\lambda,u)\,v))+P\left(\sum_{i=1}^{k}E(\eta_i^2|\mathscr{F}_{i-1})>v\right)$$
$$\leqq\exp(\phi(\lambda,u)\,v-\lambda\gamma)+P\left(\sum_{i=1}^{k}E(\eta_i^2|\mathscr{F}_{i-1})>v\right).$$

Minimizing the first summand w.r.t. $\lambda$ gives $\exp(\gamma u^{-1}-(\gamma u^{-1}+vu^{-2})\log(\gamma uv^{-1}+1))$ which is smaller than $\exp[\gamma u^{-1}(1\text{-}\log(\gamma uv^{-1}))]$. This finishes the proof. □

We also need an estimate of the oscillations of a Brownian motion which is well known; see, for example, [6], p. 113.

**Lemma 2.** *If $(B(t))_{t\geqq 0}$ is a Brownian motion, then for all $a, \varepsilon>0$ and $0<\gamma<1$*

$$P(\sup_{\substack{0\leqq s,t\leqq a\\ |s-t|\leqq\gamma}}|B(s)-B(t)|\geqq\varepsilon)\leqq 12(a+1)(2\pi\gamma)^{-1/2}\varepsilon^{-1}\exp(-\varepsilon^2/18\gamma).$$

Theorems 1 and 2 are immediate consequences of the following lemma.

**Lemma 3.** *Under the assumptions of Theorem 1 there exist constants $K_1, K_2\in(0,\infty)$ depending only on $\delta$ such that for $L_k\leqq K_2$*

$$\rho(P\circ S_{(k)}^{-1}, W)\leqq K_1\left\{L_k^{1/(3+2\delta)}|\log L_k|^{-1}+\left[\sum_{i=1}^{k}E(|\xi_i|^{2+2\delta}I(|\xi_i|>L_k^{1/(3+2\delta)}|\log L_k|^{-6}))\right]^{1/(3+2\delta)}\right\}.$$

*Proof.* For the sake of simplicity we write $L$ and $S$ instead of $L_k$ and $S_{(k)}$, respectively. $K$ always denotes some finite constant which may depend on $\delta$ but on nothing else. For brevity we set

$$d=L^{1/(3+2\delta)}|\log L|^{-6},$$

$$\varepsilon=6L^{1/(3+2\delta)}|\log L|^{-1}+\left[\sum_{i=1}^{k}E(|\xi_i|^{2+2\delta}I(|\xi_i|>d))\right]^{1/(3+2\delta)}$$

and for $i=1,\dots,k$

$$\eta_i=\xi_i I(|\xi_i|\leqq d)-E(\xi_i I(|\xi_i|\leqq d)|\mathscr{F}_{i-1})$$

and

$$\zeta_i=\xi_i I(|\xi_i|>d)-E(\xi_i I(|\xi_i|>d)|\mathscr{F}_{i-1}).$$

By construction $\eta_1,\dots,\eta_k$ and $\zeta_1,\dots,\zeta_k$ are martingale difference sequences w.r.t. $\mathscr{F}_0,\mathscr{F}_1,\dots,\mathscr{F}_k$ such that $\xi_i=\eta_i+\zeta_i$. Let the random element $\bar{S}$ in $C$ be defined by $\bar{S}(t)=\sum_{i=1}^{j}\eta_i$ for $t=V_j$, $j=0,1,\dots,k$, and by linear interpolation on the intervals $[V_{j-1},V_j]$, $j=1,\dots,k$. Then

$$\rho(P\circ S^{-1}, W)\leqq\rho(P\circ S^{-1}, P\circ\bar{S}^{-1})+\rho(P\circ\bar{S}^{-1}, W).$$

Using the elementary estimate (1.7) in [8] we obtain

$$\begin{aligned}\rho(P\circ S^{-1}, P\circ\bar{S}^{-1})&\leqq\varepsilon+P(\sup_{0\leqq t\leqq 1}|S(t)-\bar{S}(t)|\geqq\varepsilon)\\&\leqq\varepsilon+P\left(\max_{1\leqq j\leqq k}\left|\sum_{i=1}^{j}\xi_i-\sum_{i=1}^{j}\eta_i\right|\geqq\varepsilon\right)\leqq\varepsilon+P\left(\max_{1\leqq j\leqq k}\left|\sum_{i=1}^{j}\zeta_i\right|\geqq\varepsilon\right).\end{aligned}$$

An application of Lemma 1 with $\varepsilon$, $\varepsilon/2$ and $d^{-2\delta}L$ instead of $\gamma$, $u$ and $v$, respectively, yields

$$P\left(\max_{1\leqq j\leqq k}\left|\sum_{i=1}^{j}\zeta_i\right|\geqq\varepsilon\right)\leqq\sum_{i=1}^{k}P(|\zeta_i|>\varepsilon/2)+K\varepsilon^{-4}d^{-4\delta}L^2,$$

since

$$\sum_{i=1}^{k}E(\zeta_i^2|\mathscr{F}_{i-1})\leqq\sum_{i=1}^{k}E(\xi_i^2 I(|\xi_i|>d)|\mathscr{F}_{i-1})\leqq d^{-2\delta}L \text{ by (4),}$$

i.e. $P\left(\sum_{i=1}^{k}E(\zeta_i^2|\mathscr{F}_{i-1})>Ld^{-2\delta}\right)=0$. For $L$ sufficiently small we have $\varepsilon^{-4}d^{-4\delta}L^2\leqq K\varepsilon$ and

$$\sum_{i=1}^{k}P(|\zeta_i|>\varepsilon/2)\leqq K\varepsilon^{-2-2\delta}\sum_{i=1}^{k}E(|\xi_i|^{2+2\delta}I(|\xi_i|>d))\leqq K\varepsilon.$$

Thus it remains to prove $\rho(P\circ\bar{S}^{-1},W)\leqq K\varepsilon$ for small values of $L$. Here we use the Skorokhod embedding scheme for martingales which enables us to estimate $\rho(P\circ\bar{S}^{-1},W)$ by another application of inequality (1.7) in [8]. Without loss of generality we may assume that the space $(\Omega,\mathscr{F},P)$ is rich enough such that there exist nonnegative random variables $0=\tau_0,\tau_1,\ldots,\tau_k$ and a Brownian motion $B=(B(t))_{t\geqq 0}$ on $(\Omega,\mathscr{F},P)$ with the following properties:

(i) For $j=0,1,\ldots,k$ $T_j=\sum_{i=0}^{j}\tau_i$ is a Markov time w.r.t. $(\mathscr{B}(t))_{t\geqq 0}$ where $\mathscr{B}(t)$ is the $\sigma$-field generated by the variables $B(s)$ for $0\leqq s\leqq t$.

(ii) $\eta_i=B(T_i)-B(T_{i-1})$ for $i=1,\ldots,k$. (6)

(iii) If $\mathscr{B}_i=\{F\in\mathscr{F}:F\cap\{T_i\leqq t\}\in\mathscr{B}(t)$ for all $t\geqq 0\}$ denotes the $\sigma$-field of all events observable before $T_i$, then for $i=1,\ldots,k$

$$E(\tau_i|\mathscr{B}_{i-1})=E(\eta_i^2|\mathscr{B}_{i-1})=E(\eta_i^2|\eta_1,\ldots,\eta_{i-1}). \tag{7}$$

(iv) For any $p>1$ there exists a constant $L_p<\infty$ depending only on $p$ (with $L_2\leqq 4$) such that for $i=1,\ldots,k$

$$E(\tau_i^p|\mathscr{B}_{i-1})\leqq L_pE(|\eta_i|^{2p}|\mathscr{B}_{i-1})=L_pE(|\eta_i|^{2p}|\eta_1,\ldots,\eta_{i-1}). \tag{8}$$

The variables $\tau_0,\tau_1,\ldots,\tau_k$ may be constructed as in [2], Sect. 37. Billingsley's proof is formulated for independent variables only, but it is straightforward to give a martingale version using conditional distribution techniques. In this construction $T_0,T_1,\ldots,T_k$ are Markov times w.r.t. the Brownian motion $B$ whereas a larger filtration is used in Strassen's original Theorem 4.3 in [12]. This point, however, is not essential for our application; we could work with Strassen's theorem also. For the remaining part of the proof we write $\gamma=L^{2/(3+2\delta)}|\log L|^{-3}$, $u=L^{2/(3+2\delta)}|\log L|^{-4}$ and $v=e^{-2}L^{4/(3+2\delta)}|\log L|^{-7}$. Let $B_1$ denote the restriction of the Brownian motion $B$ to the time interval $[0,1]$; then we have

$$\rho(P\circ\bar{S}^{-1}, W)=\rho(P\circ\bar{S}^{-1}, P\circ B_1^{-1})\leqq 2\varepsilon+P(\sup_{0\leqq t\leqq 1}|\bar{S}(t)-B(t)|\geqq 2\varepsilon)$$
$$\leqq 2\varepsilon+P\left(\max_{1\leqq j\leqq k}\left|\sum_{i=1}^{j}\eta_i-B(V_j)\right|\geqq\varepsilon\right)$$
$$+P(\max_{1\leqq j\leqq k}\sup_{V_{j-1}\leqq s,t\leqq V_j}|B(s)-B(t)|\geqq\varepsilon).$$

Lemma 2 implies

$$P(\max_{1\leqq j\leqq k}\sup_{V_{j-1}\leqq s,t\leqq V_j}|B(s)-B(t)|\geqq\varepsilon)$$
$$\leqq P(\sup_{\substack{0\leqq s,t\leqq 1\\|s-t|\leqq\gamma}}|B(s)-B(t)|\geqq\varepsilon)\leqq K\varepsilon^{-1}\gamma^{-1/2}\exp(-\varepsilon^2/18\gamma)\leqq K\varepsilon$$

for $L$ sufficiently small since then by (4)

$$\max_{1\leqq j\leqq k}|V_j-V_{j-1}|=\max_{1\leqq j\leqq k}E(\xi_j^2|\mathscr{F}_{j-1}|)$$
$$\leqq\left(\sum_{i=1}^{k}E(|\xi_i|^{2+2\delta}|\mathscr{F}_{i-1})\right)^{1/(1+\delta)}\leqq L^{1/(1+\delta)}\leqq\gamma.$$

From (6) we get $\sum_{i=1}^{j}\eta_i=B(T_j)$ for $j=1,\ldots,k$, hence

$$P\left(\max_{1\leqq j\leqq k}\left|\sum_{i=1}^{j}\eta_i-B(V_j)\right|\geqq\varepsilon\right)$$
$$\leqq P(\max_{1\leqq j\leqq k}|T_j-V_j|>2\gamma)+P(\{\max_{1\leqq j\leqq k}|B(T_j)-B(V_j)|\geqq\varepsilon\}\cap\{\max_{1\leqq j\leqq k}|T_j-V_j|\leqq 2\gamma\})$$

where the second term on the right hand side is easily handled by another application of Lemma 2. To bound the remaining probability we write taking (7) into account

$$P(\max_{1\leqq j\leqq k}|T_j-V_j|>2\gamma)$$
$$\leqq P\left(\max_{1\leqq j\leqq k}\left|\sum_{i=1}^{j}(\tau_i-E(\tau_i|\mathscr{B}_{i-1}))\right|>\gamma\right)$$
$$+P\left(\max_{1\leqq j\leqq k}\left|\sum_{i=1}^{j}(E(\eta_i^2|\eta_1,\ldots,\eta_{i-1})-E(\xi_i^2|\mathscr{F}_{i-1}))\right|>\gamma\right).$$

Lemma 1 implies

$$P\left(\max_{1\leqq j\leqq k}\left|\sum_{i=1}^{j}(\tau_i-E(\tau_i|\mathscr{B}_{i-1}))\right|>\gamma\right)$$
$$\leqq\sum_{i=1}^{k}P(|\tau_i-E(\tau_i|\mathscr{B}_{i-1})|>u)+2P\left(\sum_{i=1}^{k}E(\tau_i^2|\mathscr{B}_{i-1})>v\right)$$
$$+2\exp[\gamma u^{-1}(1-\log(\gamma u v^{-1}))].$$

Now by (8)

$$\sum_{i=1}^{k} P(|\tau_i - E(\tau_i|\mathscr{B}_{i-1})| > u) \leqq Ku^{-4} \sum_{i=1}^{k} E(\tau_i^4) \leqq Ku^{-4} \sum_{i=1}^{k} E(\eta_i^8)$$

$$\leqq Ku^{-4} \sum_{i=1}^{k} E(\xi_i^8\, I(|\xi_i| \leqq d)) \leqq Ku^{-4} d^{6-2\delta} L \leqq K\varepsilon$$

and also $\exp[\gamma u^{-1}(1 - \log(\gamma u v^{-1}))] \leqq K\varepsilon$ for $L$ small enough. Furthermore

$$\sum_{i=1}^{k} E(\eta_i^4|\mathscr{F}_{i-1}) \leqq 16 \sum_{i=1}^{k} E(\xi_i^4\, I(|\xi_i| \leqq d)|\mathscr{F}_{i-1})$$

$$\leqq \begin{cases} 16 d^{2-2\delta} \sum_{i=1}^{k} E(|\xi_i|^{2+2\delta}|\mathscr{F}_{i-1}) \leqq 16 d^{2-2\delta} L & \text{if } 0 < \delta < 1 \\ 16 \sum_{i=1}^{k} E(\xi_i^4|\mathscr{F}_{i-1}) \leqq 16 L^{1/\delta} & \text{if } 1 \leqq \delta < 3/2 \end{cases}$$

where for $1 < \delta < 3/2$ we have used (1), (4) and Hölder's inequality to obtain

$$\sum_{i=1}^{k} E(\xi_i^4|\mathscr{F}_{i-1}) = \sum_{i=1}^{k} E(|\xi_i|^{(2+2\delta)/\delta} |\xi_i|^{(2\delta-2)/\delta}|\mathscr{F}_{i-1})$$

$$\leqq \left(\sum_{i=1}^{k} E(|\xi_i|^{2+2\delta}|\mathscr{F}_{i-1})\right)^{1/\delta} \left(\sum_{i=1}^{k} E(\xi_i^2|\mathscr{F}_{i-1})\right)^{(\delta-1)/\delta} \leqq L^{1/\delta}.$$

Thus $\sum_{i=1}^{k} E(\eta_i^4|\mathscr{F}_{i-1}) \leqq v/12$ for $0 < \delta < 3/2$ and small values of $L$. Combining (8) with this last conclusion, we get

$$P\left(\sum_{i=1}^{k} E(\tau_i^2|\mathscr{B}_{i-1}) > v\right) \leqq P\left(\sum_{i=1}^{k} E(\eta_i^4|\eta_1, \ldots, \eta_{i-1}) > v/4\right)$$

$$\leqq P\left(\sum_{i=1}^{k} (E(\eta_i^4|\eta_1, \ldots, \eta_{i-1}) - \eta_i^4) > v/12\right)$$

$$+ P\left(\sum_{i=1}^{k} (\eta_i^4 - E(\eta_i^4|\mathscr{F}_{i-1})) > v/12\right)$$

$$\leqq Kv^{-2} \sum_{i=1}^{k} E(\eta_i^8) \leqq Kv^{-2} d^{6-2\delta} L \leqq K\varepsilon.$$

Finally

$$\max_{1 \leqq j \leqq k} \left| \sum_{i=1}^{j} (E(\eta_i^2|\mathscr{F}_{i-1}) - E(\xi_i^2|\mathscr{F}_{i-1})) \right|$$

$$\leqq 2 \sum_{i=1}^{k} E(\xi_i^2\, I(|\xi_i| > d)|\mathscr{F}_{i-1}) \leqq 2 d^{-2\delta} L \leqq \gamma/3$$

for $L$ sufficiently small, hence

$$P\left(\max_{1\leq j\leq k}\left|\sum_{i=1}^{j}(E(\eta_i^2|\eta_1,\ldots,\eta_{i-1})-E(\xi_i^2|\mathscr{F}_{i-1}))\right|>\gamma\right)$$
$$\leq P\left(\max_{1\leq j\leq k}\left|\sum_{i=1}^{j}(E(\eta_i^2|\eta_1,\ldots,\eta_{i-1})-\eta_i^2)\right|>\gamma/3\right)$$
$$+P\left(\max_{1\leq j\leq k}\left|\sum_{i=1}^{j}(\eta_i^2-E(\eta_i^2|\mathscr{F}_{i-1}))\right|>\gamma/3\right).$$

These probabilities are estimated by another application of Lemma 1 with $\gamma$, $u$ and $v$ replaced by $\gamma/3$, $u/3$ and $v/9$, respectively, and now straightforward calculations. This finishes the proof of Lemma 3. □

*Proof of Theorem 1.* For the constants $K_1$, $K_2$ occurring in Lemma 3 which depend only on $\delta$ we may assume w.l.o.g. that $0<K_2<e^{-1}<1<K_1<\infty$. Then for $L_k\leq K_2$ by Lemma 3

$$\rho(P\circ S_{(k)}^{-1}, W)\leq 2K_1 L_k^{1/(3+2\delta)}$$

and for $K_2<L_k$

$$\rho(P\circ S_{(k)}^{-1}, W)\leq 1<K_2^{-1/(3+2\delta)}L_k^{1/(3+2\delta)}.$$

Taking $K_\delta=\max(2K_1, K_2^{-1/(3+2\delta)})$ finishes the proof. □

To prove the remark following Theorem 1 put

$$d=L^{1/4\delta}|\log L|^{-2/3},\qquad \varepsilon=6e^{1/2}192^{1/4}L^{1/4\delta}|\log L|^{3/4},$$
$$\gamma=192^{1/2}eL^{1/2\delta}|\log L|^{1/2},\qquad u=192^{1/2}eL^{1/2\delta}$$

and $v=192\,L^{1/\delta}$ in the proof of Lemma 3. Then for $3/2\leq\delta\leq 3$ all arguments go through unchanged, whereas for $\delta>3$ one has to replace $\sum_{i=1}^{k}E(\eta_i^8)\leq d^{6-2\delta}L$ by $\sum_{i=1}^{k}E(\eta_i^8)\leq L^{3/\delta}$ obtained from (1) and (4) by Hölder's inequality.

*Proof of Theorem 2.* An application of Lemma 3 with $\xi_{n1},\ldots,\xi_{nn}$ and $L_n=K_0 n^{-\delta}$ instead of $\xi_1,\ldots,\xi_k$ and $L_k$, respectively, yields the desired result since

$$|\log L_n|^{-1}\to 0\quad\text{and}\quad n^{1/2}L_n^{1/(3+2\delta)}|\log L_n|^{-6}\to\infty\qquad\text{as } n\to\infty.\quad\square$$

*Acknowledgements.* The author is especially grateful to Prof. Dr. P. Gaenssler for his advice and encouragement during the preparation of the author's doctoral thesis, which the here presented results are part of. He thanks the referees for their helpful comments that improved the presentation.

## References

1. Arak, T.V.: On an estimate of A.A. Borovkov. Theor. Probability Appl. **20**, 372–373 (1975)
2. Billingsley, P.: Probability and Measure. New York: Wiley 1979
3. Bolthausen, E.: Exact convergence rates in some martingale central limit theorems. Ann. Probability **10**, 672–688 (1982)

4. Borovkov, A.A.: On the rate of convergence for the invariance principle. Theor. Probability Appl. **18**, 207–225 (1973)
5. Fuk, D.Kh.: Some probabilistic inequalities for martingales. Siberian Math. J. **14**, 131–137 (1973)
6. Hall, P., Heyde, C.C.: Martingale Limit Theory and Its Application. New York: Academic Press 1980
7. Hall, P., Heyde, C.C.: Rates of convergence in the martingale central limit theorem. Ann. Probability **9**, 395–404 (1981)
8. Komlós, J., Major, P., Tusnády, G.: Weak convergence and embedding. In: Colloquia Mathematica Societatis János Bolyai, 11. Limit Theorems of Probability Theory (P. Révész, Ed.), Keszthely, 1974, pp. 149–165. Amsterdam-London: North-Holland 1974
9. Neveu, J.: Discrete-Parameter Martingales. Amsterdam-Oxford-New York: North Holland and American Elsevier 1975
10. Prokhorov. Yu.V.: Convergence of random processes and limit theorems in probability theory. Theor. Probability Appl. **1**, 157–214 (1956)
11. Rosenkrantz, W.A.: On the rate of convergence for the invariance principle. Trans. Amer. Math. Soc. **129**, 542–552 (1967)
12. Strassen, V.: Almost sure behavior of sums of independent random variables and martingales. In: Proc. Fifth Berkeley Symp. Math. Statist. Probability, Vol. II, Part 1, pp. 315–343. Berkeley and Los Angeles: Univ. of California Press 1967

Received November 15, 1982; in revised form November 15, 1983