The Annals of Probability
1996, Vol. 24, No. 4, 2098–2127

WEAK CONVERGENCE FOR WEIGHTED EMPIRICAL PROCESSES OF DEPENDENT SEQUENCES

BY QI-MAN SHAO[1] AND HAO YU[2]

National University of Singapore and University of Western Ontario

In this paper we establish weak convergence theorems for weighted empirical processes of strong mixing, ρ-mixing and associated sequences. We apply these results to obtain weak convergence of integral functionals of empirical processes and of mean residual life processes in reliability theory. To carry out the proofs, we develop two Rosenthal-type inequalities for strong mixing and associated sequences.

1. Introduction. Let $\{X_n,\ n \geq 1\}$ be a sequence of random variables with common distribution function F. Then the empirical distribution function F_n of $X_1, \dots, X_n$ is defined by $F_n(x) = n^{-1}\sum_{i=1}^{n} I(X_i \leq x),\ -\infty < x < \infty$, where $I(A)$ is the usual indicator function of the set A. The nth empirical process β_n is defined by $\beta_n(x) = n^{1/2}(F_n(x) - F(x)), -\infty < x < \infty$.

Let Q be the quantile function of F, defined by

$$Q(t) = F^{-1}(t) = \inf\{x\colon F(x) \geq t\}, \qquad 0 < t \leq 1.$$
$$Q(0) = Q(0+).$$

That is, the quantile function Q as defined here is the left continuous inverse of the right continuous distribution function F. If F is continuous, then Q satisfies

$$Q(t) = F^{-1}(t) = \inf\{x\colon F(x) = t\}, \qquad F(Q(t)) = t \in [0, 1]. \tag{1.1}$$

Hence $U_n = F(X_n)$ for all $n \geq 1$ are uniform $[0, 1]$ distributed. The uniform empirical distribution function of $U_1, \dots, U_n$ is defined by

$$E_n(t) = F_n(Q(t)) = \frac{1}{n}\sum_{i=1}^{n} I(U_i \leq t), \qquad 0 \leq t \leq 1, \tag{1.2}$$

and the nth uniform empirical process is given by

$$\{\alpha_n(t),\ 0 \leq t \leq 1\} = \{n^{1/2}(E_n(t) - t),\ 0 \leq t \leq 1\}, \qquad n = 1, 2, \dots. \tag{1.3}$$

Received June 1995; revised March 1996.

[1]Supported by a National University of Singapore Research Project.

[2]Supported by an NSERC Canada Grant to M. Csörgő and an NSERC Grant at the University of Western Ontario.

AMS 1991 *subject classifications*. 60B10, 60F15, 60E15, 60F17.

Key words and phrases. Weak convergence, weighted empirical processes, integral functionals, mixing sequences, associated sequences, mean residual life processes, Rosenthal-type inequalities.

Thus, in terms of $U_i = F(X_i)$, $i = 1, \dots, n$, we have for any continuous distribution function F,

$$\{\beta_n(Q(t)),\ 0 \le t \le 1\} = \{\alpha_n(t),\ 0 \le t \le 1\}, \qquad n = 1, 2, \dots. \tag{1.4}$$

This implies that all theorems proved for α_n will hold automatically for β_n as well, simply by letting $y = F(x)$ in (1.4). So, we will mainly be concerned with the (weighted) uniform empirical process α_n in this paper.

Let q be a positive weight function on $(0, 1)$; that is, $\inf_{\delta \le t \le 1-\delta} q(t) > 0$ for all $0 < \delta < 1/2$. Define the weighted uniform empirical process as $\{\alpha_n(t)/q(t),\ 0 < t < 1\}$. When $\{U_n,\ n \ge 1\}$ is a sequence of independent r.v.'s uniformly distributed on $[0, 1]$, starting off with Rényi (1953), Chibisov (1964) and O'Reilly (1974), there has been considerable interest in the asymptotic behavior of weighted uniform empirical processes. For an insightful view of this subject we refer to Csörgő, Csörgő, Horváth and Mason (1986a), Shorack and Wellner (1986) and Csörgő and Horváth (1993), as well as to the references in these works. For the sake of easy reference, we restate a theorem of Csörgő, Csörgő, Horváth and Mason (1986a) as follows. For a shorter and more direct proof we refer to Csörgő and Horváth (1986), as well as to Csörgő and Horváth [(1993), Chapter 4], where both proofs are presented in complete detail.

THEOREM A. *We assume that q is positive and continuous on $(0, 1)$, and is nondecreasing in a neighborhood of 0 and nonincreasing in a neighborhood of 1. Then*

$$I(q, \lambda) = \int_0^1 \frac{1}{t(1-t)} \exp\left(-\frac{\lambda q^2(t)}{t(1-t)}\right) dt < \infty \tag{1.5}$$

for all $\lambda > 0$ if and only if, as $n \to \infty$,

$$\alpha_n(\cdot)/q(\cdot) \to_{\mathscr{D}} B(\cdot)/q(\cdot) \quad \textit{in } D[0, 1], \tag{1.6}$$

where $\{B(t),\ 0 \le t \le 1\}$ is a Brownian bridge and $D = D[0, 1]$ is the usual D space on $[0, 1]$ with the Skorokhod J_1-topology [cf. Billingsley (1968)].

While the problem of weak convergence for weighted empirical processes of independent sequences has been intensively studied in recent years, there are only a few studies concerned with the counterpart for dependent sequences [cf. Yu (1993a)]. In the latter case the limit process is changed from being a Brownian bridge due to the appearance of covariances among observations. Namely, under certain conditions, we have

$$\alpha_n(\cdot) \to_{\mathscr{D}} B^*(\cdot) \quad \text{in } D[0, 1], \tag{1.7}$$

where $B^*(\cdot)$ is a zero-mean Gaussian process specified by $B^*(0) = B^*(1) = 1$ and

$$\begin{aligned} EB^*(s)B^*(t) &= s \wedge t - st \\ &\quad + \sum_{k=2}^{\infty} \big\{\mathrm{Cov}(I(U_1 \le s), I(U_k \le t)) \\ &\qquad + \mathrm{Cov}(I(U_k \le s), I(U_1 \le t))\big\}. \end{aligned} \tag{1.8}$$

The main objective of this paper is to establish weak convergence for weighted empirical processes of strictly stationary observations under mixing and associated dependence assumptions and to apply the results to studying the weak convergence of integral functionals of empirical processes and of mean residual life processes in reliability. Our results on weak convergence for weighted empirical processes of stationary associated sequences are partially based on Yu (1993a), where they are successfully applied to obtain weak convergence for weighted quantile processes of stationary associated sequences as well as to establish a unified asymptotic theory for empirical reliability and concentration processes of associated sequences. For such a theory in the i.i.d. case we refer to Csörgő, Csörgő and Horváth (1986).

This paper is organized as follows. Section 2 presents a basic theorem and its corollary for stationary sequences and the main results on weak convergence for weighted empirical processes of stationary mixing and associated sequences. The strong consistency and weak approximation for mean residual life processes in reliability are given in Section 3. Two Rosenthal-type inequalities for α-mixing and associated sequences are stated and proved in Section 4. All other proofs are carried out in Section 5.

2. Main results. We first give the following basic theorem for a stationary sequence of uniform $[0, 1]$ random variables. By stationarity we mean that the joint distribution of $U_{i+1}, \dots, U_{i+m}$ does not depend on i for any fixed positive integer m.

THEOREM 2.1. *Let $\{U_n,\ n \geq 1\}$ be a stationary sequence of uniform $[0,1]$ random variables. Assume that for all $0 \leq s, t \leq 1$ and $n \geq 1$ we have the following conditions*:

(A1) $E|\alpha_n(t) - \alpha_n(s)|^p \leq C_1(|t-s|^{p_1} + n^{-p_2/2}|t-s|^{r_1})$ *for some* $C_1 > 0$, $p > 2$, $p_1 > 1$, $0 \leq r_1 \leq 1$ *and* $p_2 > 1 - r_1$;

(A2) $E(\alpha_n(t) - \alpha_n(s))^2 \leq C_2|t-s|^{r_2}$ *for some* $C_2 > 0$ *and* $0 < r_2 \leq 1$.

If we have

$$\alpha_n(\cdot) \to_{\mathscr{D}} B^*(\cdot) \quad \text{in } D[0,1] \tag{2.1}$$

with the Gaussian process $B^*(\cdot)$ defined by (1.8), *then*

$$\alpha_n(\cdot)/q(\cdot) \to_{\mathscr{D}} B^*(\cdot)/q(\cdot) \quad \text{in } D[0,1], \tag{2.2}$$

where q is an arbitrary weight function such that, for some $C > 0$ and $\beta > 1/2$,

$$q(t) \geq C(t(1-t))^{\mu}(\log 1/(t(1-t)))^{\beta} \quad \text{for all } 0 < t < 1 \tag{2.3}$$

and

$$\mu = \min\left(\frac{p_1}{p}, \frac{r_1 + p_2}{p + p_2}, \frac{r_2}{2}\right). \tag{2.4}$$

REMARK 2.1. By using a standard argument [cf. Theorem 12.2 and (22.18) in Billingsley (1968)], one can easily verify that $\{\alpha_n(t),\ 0 \leq t \leq 1\}$ is tight by the condition (A1). Hence to prove (2.1), one only needs to show that any finite dimensional distribution of $\{\alpha_n(t),\ 0 \le t \le 1\}$ will converge to that of $\{B^*(t),\ 0 \le t \le 1\}$ and the series in (1.8) converges absolutely.

REMARK 2.2. The weight function q used in Theorem A is usually called a Chibisov–O'Reilly weight function. If we write $q(t) = (t(1-t)\log\log(1/(t(1-t))))^{1/2}g(t)$, then, necessarily, $g(t) \to \infty$ as $t \to 0$ or $t \to 1$. Thus our weight function q of (2.3) can be compared to a Chibisov–O'Reilly weight function by taking μ in (2.4) close to 1/2 or exactly 1/2 for properly chosen $p,\ p_1,\ p_2,\ r_1$ and r_2. In fact, Theorems 2.2, 2.3 and 2.4 show this possibility of taking $\mu = 1/(2+\varepsilon)$ for some $\varepsilon > 0$ in cases of mixing and associated sequences. In particular, our sharpest rate of $\mu = 1/2$ is obtained for ρ-mixing under a stronger mixing decay rate. In most cases, however, $\mu < 1/2$. We note in passing that if for a general weight function q we have $\int_0^1 q^{-2}(t)\,dt < \infty$, then we have (1.5) as well for all $\lambda > 0$. That is, q is then necessarily a Chibisov–O'Reilly weight function.

REMARK 2.3. If $\mu = (r_1 + p_2)/(p + p_2) < \min(p_1/p, r_2/2)$ in (2.4), then from the proof of Theorem 2.1, one can relax the restriction on β from $\beta > 1/2$ to $\beta > 1/(p+p_2) = (1-\mu)/(p-r_1)$. Moreover, in the case of $\mu \ge 1/(p+1-r_1)$, one can use a simple sufficient condition $\int_0^1 q^{-1/\mu}(t)\,dt < \infty$ to replace (2.3).

A direct application of Theorems 2.1 is to obtain weak convergence for integral functionals of α_n. For example, we consider the integral functional

$$\Delta_n(t) = \int_0^t \alpha_n(s)\,dQ(s), \qquad 0 \le t \le 1,$$

and its approximating Gaussian counterpart

$$\Delta(t) = \int_0^t B^*(s)\,dQ(s), \qquad 0 \le t \le 1.$$

This function plays a central role in weak approximation theory for empirical total time on test, mean residual life, empirical Lorenz and Goldie concentration processes which are of interest in reliability and economic concentration theories. The reader may refer to Csörgő, Csörgő, Horváth and Mason (1986b) and Csörgő, Csörgő, and Horváth (1986).

COROLLARY 2.1. *Under the conditions of Theorem* 2.1, *if*

$$\int_0^1 (t(1-t))^{\mu}(\log 1/(t(1-t)))^{\beta}\,dQ(t) < \infty, \tag{2.5}$$

then

$$\Delta_n(\cdot) \to_{\mathscr{D}} \Delta(\cdot) \quad \text{in } D[0,1].$$

REMARK 2.4. Assume that F is the distribution function of a random variable X. Then condition (2.5) is slightly stronger than the existence of the $(1/\mu)$th moment of X. Indeed, on extending the discussion in the Appendix of Hoeffding (1973), we see that (2.5) implies $E|X|^{1/\mu} < \infty$. This is not necessarily true conversely, but $E|X|^{1/\mu}(\log(1+|X|))^{(1+\beta)/\mu+\delta} < \infty$, with any $\delta > 0$, implies (2.5).

Theorem 2.1 enables us to establish weak convergence for weighted empirical processes of stationary mixing and associated sequences. We first introduce the following dependence notions.

Let $(\Omega, \mathscr{F}, P)$ be a probability space and $\mathscr{F}_1$ and $\mathscr{F}_2$ be two σ-algebras contained in $\mathscr{F}$. Define the following measures of dependence between $\mathscr{F}_1$ and $\mathscr{F}_2$:

$$\rho(\mathscr{F}_1, \mathscr{F}_2) = \sup_{\substack{X \in L_2(\mathscr{F}_1) \\ Y \in L_2(\mathscr{F}_2)}} \frac{|\operatorname{Cov}(X, Y)|}{(\operatorname{Var} X \cdot \operatorname{Var} Y)^{1/2}}$$

and

$$\alpha(\mathscr{F}_1, \mathscr{F}_2) = \sup_{\substack{A \in \mathscr{F}_1 \\ B \in \mathscr{F}_2}} |P(A \cap B) - P(A)\, P(B)|.$$

Let $\{X_n, n \geq 1\}$ be a sequence of real-valued random variables on $(\Omega, \mathscr{F}, P)$, and let $\mathscr{F}_n^m = \sigma(X_i, n \leq i \leq m)$ be σ-algebras generated by the indicated random variables and put

$$\rho(n) = \sup_{k \geq 1} \rho(\mathscr{F}_1^k, \mathscr{F}_{n+k}^\infty), \quad \text{and} \quad \alpha(n) = \sup_{k \geq 1} \alpha(\mathscr{F}_1^k, \mathscr{F}_{n+k}^\infty).$$

The sequence $\{X_n,\ n \geq 1\}$ is said to be ρ-mixing or α-mixing, according as $\rho(n) \to 0$ or $\alpha(n) \to 0$ as $n \to \infty$, respectively. It is well known that $\alpha(n) \leq \rho(n)$.

A finite collection of random variables $X_1, \dots, X_n$ is said to be associated if for any two coordinatewise nondecreasing functions f, g: $R^n \to R$,

$$\operatorname{Cov}(f(X_1, \dots, X_n),\ g(X_1, \dots, X_n)) \geq 0$$

whenever this covariance is defined. An infinite family of random variables is associated if every finite subfamily is associated. This definition was introduced by Esary, Proschan and Walkup (1967) and has found several applications in reliability theory [cf. Barlow and Proschan (1981) and Yu (1993a)].

THEOREM 2.2. *Let $\{U_n,\ n \geq 1\}$ be a stationary α-mixing sequence of uniform* $[0, 1]$ *random variables. If*

$$\alpha(n) = O(n^{-\theta-\varepsilon}) \quad \text{for some } \theta \geq 1 + \sqrt{2} \text{ and } \varepsilon > 0, \tag{2.6}$$

then we have

$$\alpha_n(\cdot)/q(\cdot) \to_{\mathscr{D}} B^*(\cdot)/q(\cdot) \quad \text{in } D[0, 1]$$

for any weight function q satisfying $q(t) \geq C(t(1-t))^{(1-1/\theta)/2}$ for some $C > 0$.

THEOREM 2.3. *Let* $\{U_n,\ n \ge 1\}$ *be a stationary* ρ*-mixing sequence of uniform* $[0,1]$ *random variables. Suppose that the series in* (1.8) *converges absolutely. If*

$$\sum_{n=1}^{\infty} \rho(2^n) < \infty, \tag{2.7}$$

then for any $\varepsilon > 0$ *we have*

$$\alpha_n(\cdot)/q(\cdot) \to_{\mathscr{D}} B^*(\cdot)/q(\cdot) \quad \textit{in } D[0,1]$$

for any weight function q *satisfying* $q(t) \ge C(t(1-t))^{1/(2+\varepsilon)}$ *for some* $C > 0$.

If, in addition,

$$\sum_{n=1}^{\infty} \rho^{2/p}(2^n) < \infty \quad \textit{for some } p > 2, \tag{2.8}$$

then we have

$$\alpha_n(\cdot)/q(\cdot) \to_{\mathscr{D}} B^*(\cdot)/q(\cdot) \quad \textit{in } D[0,1]$$

for any weight function q *satisfying* $q(t) \ge C(t(1-t))^{1/2}(\log 1/(t(1-t)))^{\beta}$ *for some* $C > 0$ *and* $\beta > 1/2$.

THEOREM 2.4. *Let* $\{U_n, n \ge 1\}$ *be a stationary associated sequence of uniform* $[0,1]$ *random variables. If*

$$\operatorname{Cov}(U_1, U_n) = O(n^{-\nu-\varepsilon}) \quad \textit{for some } \nu \ge \big(3+\sqrt{33}\big)/2 \textit{ and } \varepsilon > 0, \tag{2.9}$$

then we have

$$\alpha_n(\cdot)/q(\cdot) \to_{\mathscr{D}} B^*(\cdot)/q(\cdot) \quad \textit{in } D[0,1]$$

for any weight function q *satisfying* $q(t) \ge C(t(1-t))^{(1-3/\nu)/2}$ *for some* $C > 0$.

REMARK 2.5. If we choose $\nu = (3+\sqrt{33})/2$ in (2.9), then the covariance restriction $\operatorname{Cov}(U_1, U_n) = O(n^{-4.373-\varepsilon})$ is weaker than that in Theorem 2.2 of Yu (1993b), where $\operatorname{Cov}(U_1, U_n) = \mathrm{O}(n^{-7.5-\varepsilon})$.

Based on Theorems 2.2, 2.3 and 2.4, we have the following corollaries that are Corollary 2.1-type analogs for mixing and associated sequences.

COROLLARY 2.2. *Under the conditions of Theorem* 2.2, *if*

$$\int_{-\infty}^{\infty} |x|^{2/(1-1/\theta)}\, dF(x) < \infty,$$

then

$$\Delta_n(\cdot) \to_{\mathscr{D}} \Delta(\cdot) \quad \textit{in } D[0,1].$$

COROLLARY 2.3. *Let* $\{U_n,\ n \geq 1\}$ *be a stationary* ρ*-mixing sequence of uniform* $[0,1]$ *random variables. If* (2.7) *holds and for some* $\varepsilon > 0$ *we have*

$$\int_{-\infty}^{\infty} |x|^{2+\varepsilon}\, dF(x) < \infty,$$

then

$$\Delta_n(\cdot) \to_{\mathscr{D}} \Delta(\cdot) \quad in\ D[0,1].$$

If, in addition, (2.8) *holds and*

$$\int_0^1 (t(1-t))^{1/2}(\log 1/(t(1-t)))^{\beta}\, dQ(t) < \infty,$$

then we have

$$\Delta_n(\cdot) \to_{\mathscr{D}} \Delta(\cdot) \quad in\ D[0,1].$$

COROLLARY 2.4. *Let* $\{U_n, n \geq 1\}$ *be a stationary associated sequence of uniform* $[0,1]$ *random variables. Assume that* (2.9) *holds and*

$$\int_{-\infty}^{\infty} |x|^{2/(1-3/\nu)}\, dF(x) < \infty.$$

Then

$$\Delta_n(\cdot) \to_{\mathscr{D}} \Delta(\cdot) \quad in\ D[0,1].$$

3. An application to mean residual life processes. In this section we apply Theorems 2.2, 2.3 and 2.4 and Corollaries 2.2, 2.3 and 2.4 to obtain the asymptotic normality of mean residual life processes of reliability theory. In the following text we assume that X is a nonnegative random variable. Its distribution function F (usually called the lifetime distribution function) is assumed to be continuous and $EX < \infty$. In reliability theory the mean residual the lifetime function at age $x \geq 0$ is defined by

$$M_F(x) = E(X - x | X > x) = \frac{1}{1-F(x)} \int_x^{\infty} (1 - F(t))\, dt.$$

The empirical counterpart of M_F, denoted by $M_n(s)$, is defined by

$$M_n(x) = M_{F_n}(x) = \frac{1}{1-F_n(x)} \int_x^{\infty} (1 - F_n(t))\, dt.$$

The strong consistency and normal approximation for the mean residual life process $M_n - M_F$ based on i.i.d. random variables have been studied by many authors. We refer to Yang (1978), Burke, Csörgő and Horváth (1981) and Csörgő, Csörgő and Horváth (1986). However, the i.i.d. assumption may not be realistic in reliability. As Barlow and Proschan (1981) pointed out, dependence in reliability is frequently modeled in terms of associated random variables. Thus it is natural to ask whether corresponding results hold for associated sequences. Fortunately this turns out to be possible by using tend approaches for weighted empirical processes of associated sequences. Here we should point out that one may mistakenly think that the weak convergence of (normalized) mean residual life process [see (3.2)] can be obtained by using the continuous mapping theorem in Billingsley (1968). In fact the mean residual life function $M_F(x)$ in the second term of (3.2) will often tend to infinity as x approaches $\inf\{x: F(x) = 1\}$. Consequently, the continuous mapping theorem cannot be applied.

Setting $T_F = \inf\{x: F(x) = 1\}$, we have the following strong consistency result for $M_n - M_F$.

THEOREM 3.1. *Let $\{X_n,\ n \geq 1\}$ be a sequence of associated random variables having the same distribution F. If $T < T_F$ and*

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \operatorname{Cov}\left(X_n, \sum_{i=1}^{n} X_i\right) < \infty, \tag{3.1}$$

then we have, as $n \to \infty$,

$$\sup_{0 \leq x \leq T} |M_n(x) - M_F(x)| \to 0 \quad \text{a.s.}$$

Next, we present the weak approximation results for the normalized mean residual life process $z_n(\cdot)$, which is defined by

$$\begin{aligned} z_n(x) &= n^{1/2}(M_n(x) - M_F(x)) \\ &= n^{1/2}(1 - F_n(x))^{-1}\left(-\int_x^{\infty}(F_n(t) - F(t))\,dt + M_F(x)(F_n(x) - F(x))\right) \\ &= (1 - F_n(x))^{-1}\left(-\int_{F(x)}^{1} \alpha_n(t)\,dQ(t) + M_F(x)\alpha_n(F(x))\right) \end{aligned} \tag{3.2}$$

for $0 \leq x < \max_{1 \leq i \leq n} X_i$. Hence its approximating Gaussian process will be

$$Z(x) = (1 - F(x))^{-1}\left(-\int_{F(x)}^{1} B^*(t)\,dQ(t) + M_F(x)B^*(F(x))\right), \qquad 0 \leq x < T_F.$$

THEOREM 3.2. *Let $\{X_n,\ n \geq 1\}$ be a stationary associated sequence of random variables. Assume that*

$$\operatorname{Cov}(F(X_1), F(X_n)) = O(n^{-\nu-\varepsilon}) \quad \textit{for some } \nu \geq \left(3 + \sqrt{33}\right)/2 \textit{ and } \varepsilon > 0.$$

(i) *If $T < T_F$ and $EX^{2/(1-3/\nu)} < \infty$, then*

$$z_n(\cdot) \to_{\mathscr{D}} Z(\cdot) \quad \textit{in } D[0, T].$$

(ii) *If $EX^{2/(1-3/\nu)} < \infty$, then*

$$(1 - F_n(Q(\cdot))) \cdot z_n(Q(\cdot)) \to_{\mathscr{D}} (1 - F(Q(\cdot))) \cdot Z(Q(\cdot)) \quad \textit{in } D[0, 1].$$

The results of Theorem 3.2 hold also for α-mixing sequences by applying Theorem 2.2 and Corollary 2.2, and for ρ-mixing sequences by applying Theorem 2.3 and Corollary 2.3. In the following we present these results without proof.

THEOREM 3.3. *Let $\{X_n,\ n \geq 1\}$ be a stationary α-mixing sequence of random variables. Assume that*

$$\alpha(n) = O(n^{-\theta-\varepsilon}) \quad \text{for some } \theta \geq 1+\sqrt{2} \text{ and } \varepsilon > 0.$$

(i) *If $T < T_F$ and $EX^{2/(1-1/\theta)} < \infty$, then*

$$z_n(\cdot) \to_{\mathscr{D}} Z(\cdot) \quad \text{in } D[0, T].$$

(ii) *If $EX^{2/(1-1/\theta)} < \infty$, then*

$$(1 - F_n(Q(\cdot))) \cdot z_n(Q(\cdot)) \to_{\mathscr{D}} (1 - F(Q(\cdot))) \cdot Z(Q(\cdot)) \quad \text{in } D[0, 1].$$

THEOREM 3.4. *Let $\{X_n,\ n \geq 1\}$ be a stationary ρ-mixing sequence of random variables. Assume that* (2.7) *holds and $EX^{2+\varepsilon} < \infty$ for some $\varepsilon > 0$, or* (2.8) *holds and $EX^2(\log(1+X))^{\delta} < \infty$ for some $\delta > 3$.*

(i) *For $T < T_F$ we have*

$$z_n(\cdot) \to_{\mathscr{D}} Z(\cdot) \quad \text{in } D[0, T].$$

(ii) *We have*

$$(1 - F_n(Q(\cdot))) \cdot z_n(Q(\cdot)) \to_{\mathscr{D}} (1 - F(Q(\cdot))) \cdot Z(Q(\cdot)) \quad \text{in } D[0, 1].$$

4. Moment inequalities. Moment inequalities of partial sums play a very important role in various proofs of limit theorems, for example, in the proofs of our theorems given in Sections 2 and 3. We have the Marcinkiewicz–Zygmund inequality and the Rosenthal inequality for independent random variables and the Burkholder inequality for martingales. Some nice moment inequalities are also available for dependent random variables. One can refer to Yokoyama (1980) for α-mixing sequences, Shao (1995) for ρ-mixing sequences and Birkel (1988a) for associated sequences. However, the known moment inequalities for α-mixing and associated sequences do not meet our need for proofs related to empirical processes. In this section we develop two Rosenthal-type inequalities for α-mixing and associated sequences, which are of their own interest.

THEOREM 4.1. *Let $2 < p < r \leq \infty$, $2 < v \leq r$ and $\{X_n,\ n \geq 1\}$ be an α-mixing sequence of random variables with $EX_n = 0$ and $\|X_n\|_r := (E|X_n|^r)^{1/r} < \infty$. Assume that*

$$\alpha(n) \leq C\, n^{-\theta} \quad \text{for some } C > 0 \text{ and } \theta > 0. \tag{4.1}$$

Then, for any $\varepsilon > 0$, there exists $K = K(\varepsilon, r, p, v, \theta, C) < \infty$ such that

$$E|S_n|^p \leq K\Big((n\, C_n)^{p/2} \max_{i \leq n} \|X_i\|_v^p + n^{(p-(r-p)\theta/r)\vee(1+\varepsilon)} \max_{i \leq n} \|X_i\|_r^p\Big), \tag{4.2}$$

where $C_n = (\sum_{i=0}^{n}(i+1)^{2/(v-2)}\alpha(i))^{(v-2)/v}$. In particular, for any $\varepsilon > 0$,

$$E|S_n|^p \leq K\Big(n^{p/2} \max_{i \leq n} \|X_i\|_v^p + n^{1+\varepsilon} \max_{i \leq n} \|X_i\|_r^p\Big) \tag{4.3}$$

if $\theta > v/(v-2)$ *and* $\theta \geq (p-1)r/(r-p)$, *and*

$$E|S_n|^p \leq K\, n^{p/2} \max_{i\leq n} \|X_i\|_r^p \tag{4.4}$$

if $\theta \geq pr/(2(r-p))$.

THEOREM 4.2. *Let* $2 < p < r \leq \infty$, *let* f *be an absolutely continuous function satisfying* $\sup_{x\in R} |f'(x)| \leq B$ *and let* $\{X_n,\ n \geq 1\}$ *be a sequence of associated random variables with* $Ef(X_n) = 0$ *and* $\|f(X_n)\|_r := (E|f(X_n)|^r)^{1/r} < \infty$. *Let*

$$u(n) = \sup_{i\geq 1} \sum_{j:\ |j-i|\geq n} \mathrm{Cov}(X_j, X_i) < \infty, \qquad n \geq 0.$$

Assume that

$$u(n) \leq C\, n^{-\theta} \quad \text{for some } C > 0 \text{ and } \theta > 0. \tag{4.5}$$

Then, for any $\varepsilon > 0$, *there exists* $K = K(\varepsilon, r, p, \theta) < \infty$ *such that*

$$\begin{aligned} E\left|\sum_{i=1}^n f(X_i)\right|^p \leq K\biggl(& n^{1+\varepsilon} \max_{i\leq n} E|f(X_i)|^p \\ &+ \biggl(n \max_{i\leq n} \sum_{j=1}^n |\,\mathrm{Cov}(f(X_i), f(X_j))|\biggr)^{p/2} \\ &+ n^{(r(p-1)-p+\theta(p-r))/(r-2)\vee(1+\varepsilon)} \\ &\times \max_{i\leq n} \|f(X_i)\|_r^{r(p-2)/(r-2)} (B^2 C)^{(r-p)/(r-2)}\biggr). \end{aligned} \tag{4.6}$$

In particular, we have

$$\begin{aligned} E\left|\sum_{i=1}^n f(X_i)\right|^p \leq K\biggl(& n^{1+\varepsilon} \max_{i\leq n} E|f(X_i)|^p \\ &+ \biggl(n \max_{i\leq n} \sum_{j=1}^n |\,\mathrm{Cov}(f(X_i), f(X_j))|\biggr)^{p/2} \\ &+ n^{1+\varepsilon} \max_{i\leq n} \|f(X_i)\|_r^{r(p-2)/(r-2)} (B^2 C)^{(r-p)/(r-2)}\biggr) \end{aligned} \tag{4.7}$$

if $\theta \geq (r-1)(p-2)/(r-p)$ and

$$\begin{aligned} E\left|\sum_{i=1}^n f(X_i)\right|^p \leq K\biggl(& n^{1+\varepsilon} \max_{i\leq n} E|f(X_i)|^p \\ &+ \biggl(n \max_{i\leq n} \sum_{j=1}^n |\,\mathrm{Cov}(f(X_i), f(X_j))|\biggr)^{p/2} \\ &+ n^{p/2} \max_{i\leq n} \|f(X_i)\|_r^{r(p-2)/(r-2)} (B^2 C)^{(r-p)/(r-2)}\biggr) \end{aligned} \tag{4.8}$$

if $\theta \geq r(p-2)/(2(r-p))$.

REMARK 4.1. Equation (4.4) coincides with the result of Yokoyama (1980), but with a less restrictive mixing rate.

REMARK 4.2. Equation (4.8) is a generalization of Theorems 1 and 2 of Birkel (1988a). Note that by Example 1 of Birkel (1988a), the condition $\theta \geq r(p-2)/(2(r-p))$ cannot be improved.

Let us first give the so-called C_r inequality as follows:

$$\forall\ x,\ y \geq 0, \qquad (x+y)^r \leq C_r(x^r + y^r), \qquad r > 0,$$

where $C_r = \max\{1, 2^{r-1}\}$. In the following text, we will use this inequality all the time without mentioning it. To prove Theorem 4.1, we need the following lemma.

LEMMA 4.1. *Let $\{\xi_i,\ 1 \leq i \leq n\}$ be a sequence of random variables and let $\mathscr{F}_i = \sigma(\xi_j,\ j \leq i)$. Then, for any $p \geq 2$, there exists a constant $D = D(p)$ such that*

$$\begin{aligned} E\left|\sum_{i=1}^n \xi_i\right|^p \leq D\Biggl(\biggl(\sum_{i=1}^n E\xi_i^2\biggr)^{p/2} &+ \sum_{i=1}^n E|\xi_i|^p + n^{p-1}\sum_{i=1}^n E|E(\xi_i|\mathscr{F}_{i-1})|^p \\ &+ n^{p/2-1}\sum_{i=1}^n E|E(\xi_i^2|\mathscr{F}_{i-1}) - E\xi_i^2|^{p/2}\Biggr). \end{aligned} \tag{4.9}$$

PROOF. Let $\eta_i = \xi_i - E(\xi_i|\mathscr{F}_{i-1})$ for $1 \leq i \leq n$. Then $\{\eta_i, \mathscr{F}_i,\ 1 \leq i \leq n\}$ is a martingale difference sequence. By the Burkholder (1973) inequality, there is a $D = D(p) < \infty$ such that

$$\begin{aligned} E\left|\sum_{i=1}^n \eta_i\right|^p &\leq D\Biggl(E\biggl(\sum_{i=1}^n E(\eta_i^2|\mathscr{F}_{i-1})\biggr)^{p/2} + \sum_{i=1}^n E|\eta_i|^p\Biggr) \\ &\leq 2^p D\Biggl(\biggl(\sum_{i=1}^n E\xi_i^2\biggr)^{p/2} + \sum_{i=1}^n E|\xi_i|^p \\ &\qquad + E\biggl(\sum_{i=1}^n |E(\xi_i^2|\mathscr{F}_{i-1}) - E\xi_i^2|\biggr)^{p/2}\Biggr) \\ &\leq 2^{2p} D\Biggl(\biggl(\sum_{i=1}^n E\xi_i^2\biggr)^{p/2} + \sum_{i=1}^n E|\xi_i|^p \\ &\qquad + n^{p/2-1}\sum_{i=1}^n E|E(\xi_i^2|\mathscr{F}_{i-1}) - E\xi_i^2|^{p/2}\Biggr). \end{aligned} \tag{4.10}$$

On the other hand, it is easy to see that

$$E\left|\sum_{i=1}^n \xi_i\right|^p \leq 2^p\Biggl(E\left|\sum_{i=1}^n \eta_i\right|^p + n^{p-1}\sum_{i=1}^n E|E(\xi_i|\mathscr{F}_{i-1})|^p\Biggr).$$

This proves (4.9) by the inequalities above. □

PROOF OF THEOREM 4.1. For the sake of convenience of statement, we assume that $\{X, X_n,\ n \ge 1\}$ is a strictly stationary sequence of strong mixing random variables. By a recent result of Rio (1993) [cf. also Peligrad (1994)], there is $D_1 = D_1(v)$ such that

$$ES_n^2 \le D_1\, n\, C_n \|X\|_v^2. \tag{4.11}$$

We shall prove (4.2) by induction on n. Suppose that, for each $1 \le k < n$,

$$E|S_k|^p \le K\big((k\, C_k)^{p/2}\|X\|_v^p + k^{(p-(r-p)\theta/r)\vee(1+\varepsilon)}\|X\|_r^p\big). \tag{4.12}$$

We now prove that (4.12) is still true for $k = n$. Let a, $0 < a < 1/2$, be a constant that will be specified later and let $m = [an] + 1$. Define

$$\xi_i = \sum_{j=1+2(i-1)m}^{n\wedge(2i-1)m} X_j \quad \text{and} \quad \eta_i = \sum_{j=1+(2i-1)m}^{n\wedge 2im} X_j \quad \text{for } 1 \le i \le k_n := 1+[n/(2m)].$$

Clearly, we have

$$E|S_n|^p \le 2^{p-1}\left(E\left|\sum_{i=1}^{k_n} \xi_i\right|^p + E\left|\sum_{i=1}^{k_n} \eta_i\right|^p\right) := 2^{p-1}(I_1 + I_2).$$

Let $\mathscr{F}_i = \sigma(\xi_j,\ j \le i)$. It follows from Lemma 4.1 that there is D_2 such that $D_2 \ge (2\, D_1)^{p/2}$ and

$$\begin{aligned} I_1 \;\le\; & D_2\Bigg(\sum_{i=1}^{k_n} E|\xi_i|^p + \left(\sum_{i=1}^{k_n} E\xi_i^2\right)^{p/2} + k_n^{p-1}\sum_{i=1}^{k_n} E|E(\xi_i|\mathscr{F}_{i-1})|^p \\ & \qquad + k_n^{p/2-1}\sum_{i=1}^{k_n} E|E(\xi_i^2|F_{i-1}) - E\xi_i^2|^{p/2}\Bigg) \\ := \;& D_2\left(\sum_{i=1}^{k_n} E|\xi_i|^p + I_{1,1} + I_{1,2} + I_{1,3}\right). \end{aligned} \tag{4.13}$$

In terms of (4.11), we have

$$\begin{aligned} I_{1,1} &\le (D_1\, k_n\, m\, C_m\, \|X\|_v^2)^{p/2} \\ &\le (2\, D_1\, n\, C_n)^{p/2}\|X\|_v^p \le D_2(n\, C_n)^{p/2}\|X\|_v^p. \end{aligned} \tag{4.14}$$

To estimate $I_{1,3}$, we write

$$Y_i = E(\xi_i^2|\mathscr{F}_{i-1}) - E\xi_i^2.$$

Then, by the Davydov (1970) inequality

$$\begin{aligned} E|Y_i|^{p/2} &= E|Y_i|^{p/2-1}\operatorname{sgn}(Y_i)Y_i = E\big(|Y_i|^{p/2-1}\operatorname{sgn}(Y_i)(\xi_i^2 - E\xi_i^2)\big) \\ &= \sum_{2(i-1)m<j,\, l\le n\wedge(2i-1)m} E|Y_i|^{p/2-1}\operatorname{sgn}(Y_i)(X_j X_l - EX_j X_l) \\ &\le 12 \sum_{2(i-1)m<j,\, l\le n\wedge(2i-1)m} \alpha^{1-(p-2)/p-2/r}(m)(E|Y_i|^{p/2})^{(p-2)/p}\|X_j X_l\|_{r/2} \\ &\le 12\, m^2 \alpha^{2/p-2/r}(m)(E|Y_i|^{p/2})^{(p-2)/p}\|X\|_r^2 \end{aligned}$$

and hence

$$E|Y_i|^{p/2} \le 12^{p/2}\, m^p \alpha^{1-p/r}(m)\|X\|_r^p, \tag{4.15}$$

which, together with (4.1), yields

$$\begin{aligned} I_{1,3} &\le k_n^{p/2}\, 12^p\, m^p \alpha^{1-p/r}(m)\|X\|_r^p \\ &\le 24^p n^p \alpha^{1-p/r}(m)\|X\|_r^p \\ &\le C\, 24^p\, n^p\, m^{(p-r)\theta/r}\, \|X\|_r^p \\ &\le C\, 24^p\, a^{(p-r)\theta/r}\, n^{p+(p-r)\theta/r}\, \|X\|_r^p \\ &\le C\, 24^p\, a^{(p-r)\theta/r}\, n^{(p+(p-r)\theta/r)\vee(1+\varepsilon)}\, \|X\|_r^p. \end{aligned} \tag{4.16}$$

Similar to (4.15), we have

$$E|E(\xi_i|\mathscr{F}_{i-1})|^p \le 12^p\, m^p \alpha^{1-p/r}(m)\|X\|_r^p. \tag{4.17}$$

Therefore

$$\begin{aligned} I_{1,2} &\le k_n^p\, 12^p\, m^p \alpha^{1-p/r}(m)\|X\|_r^p \\ &\le C\, 24^p\, a^{(p-r)\theta/r}\, n^{(p+(p-r)\theta/r)\vee(1+\varepsilon)}\, \|X\|_r^p. \end{aligned} \tag{4.18}$$

Putting the inequalities above together yields

$$\begin{aligned} I_1 \le D_2\bigg(&\sum_{i=1}^{k_n} E|\xi_i|^p + D_2(n\,C_n)^{p/2}\|X\|_v^p \\ &+ 2C\, 24^p\, a^{(p-r)\theta/r}\, n^{(p+(p-r)\theta/r)\vee(1+\varepsilon)}\, \|X\|_r^p\bigg). \end{aligned}$$

Similarly,

$$\begin{aligned} I_2 \le D_2\bigg(&\sum_{i=1}^{k_n} E|\xi_i|^p + D_2(n\,C_n)^{p/2}\|X\|_v^p \\ &+ 2C\, 24^p\, a^{(p-r)\theta/r}\, n^{(p+(p-r)\theta/r)\vee(1+\varepsilon)}\, \|X\|_r^p\bigg). \end{aligned}$$

Consequently, we have

$$\begin{aligned} E|S_n|^p \le 2^{p-1} D_2\bigg(&\sum_{i=1}^{k_n} E|\xi_i|^p + \sum_{i=1}^{k_n} E|\eta_i|^p + 2D_2(n\,C_n)^{p/2}\|X\|_v^p \\ &+ 4C\, 24^p\, a^{(p-r)\theta/r}\, n^{(p+(p-r)\theta/r)\vee(1+\varepsilon)}\, \|X\|_r^p\bigg). \end{aligned} \tag{4.19}$$

Now we let

$$a = (2^{p+4} D_2)^{-1/\varepsilon - p/(p-2)} \quad \text{and} \quad K = 2^{p+1} D_2(D_2 + 2C\, 24^p\, a^{(p-r)\theta/r}).$$

By (4.19) and induction hypothesis (4.12), we get

$$
\begin{aligned}
&E|S_n|^p \\
&\quad\le 2^p D_2\big(k_n\, K\big((mC_m)^{p/2}\|X\|_v^p + m^{(p+(p-r)\theta/r)\vee(1+\varepsilon)}\|X\|_r^p\big) \\
&\qquad\qquad + D_2(n\,C_n)^{p/2}\|X\|_v^p + 2C\,24^p\,a^{(p-r)\theta/r}\,n^{(p+(p-r)\theta/r)\vee(1+\varepsilon)}\,\|X\|_r^p\big) \\
&\quad\le 2^p D_2\,(n/m)\,K\big((mC_n)^{p/2}\|X\|_v^p + m^{(p+(p-r)\theta/r)\vee(1+\varepsilon)}\|X\|_r^p\big) \\
&\qquad + 2^p D_2\big(D_2 + 2C\,24^p\,a^{(p-r)\theta/r}\big) \\
&\qquad \times \big((n\,C_n)^{p/2}\|X\|_v^p + n^{(p+(p-r)\theta/r)\vee(1+\varepsilon)}\,\|X\|_r^p\big) \\
&\quad\le 2^p D_2\,K\big(a^{(p-2)/2}(nC_n)^{p/2}\|X\|_v^p + a^{\varepsilon} n^{(p+(p-r)\theta/r)\vee(1+\varepsilon)}\|X\|_r^p\big) \\
&\qquad + (1/2)\,K\big((n\,C_n)^{p/2}\|X\|_v^p + n^{(p+(p-r)\theta/r)\vee(1+\varepsilon)}\,\|X\|_r^p\big) \\
&\quad\le (1/2)K\big((nC_n)^{p/2}\|X\|_v^p + n^{(p+(p-r)\theta/r)\vee(1+\varepsilon)}\|X\|_r^p\big) \\
&\qquad + (1/2)K\big((nC_n)^{p/2}\|X\|_v^p + n^{(p+(p-r)\theta/r)\vee(1+\varepsilon)}\|X\|_r^p\big) \\
&\quad= K\big((nC_n)^{p/2}\|X\|_v^p + n^{(p+(p-r)\theta/r)\vee(1+\varepsilon)}\|X\|_r^p\big).
\end{aligned}
$$

This proves that (4.12) remains valid for $k = n$, as desired. □

The proof of Theorem 4.2 is based on the following lemmas.

LEMMA 4.2. *Let $\{X_i,\ 1 \le i \le n\}$ be associated random variables and let M and J be two subsets of $\{1, 2, \ldots, n\}$. If g: $R^{\#M} \to R$ and h: $R^{\#J} \to R$ are partially differentiable with bounded partial derivatives, then*

$$
\begin{aligned}
&|\operatorname{Cov}(g(X_i,\ i \in M),\ h(X_j, j \in J))| \\
&\qquad \le \sum_{i\in M}\sum_{j\in J} \|\partial g/\partial t_i\|_\infty \|\partial h/\partial t_i\|_\infty \operatorname{Cov}(X_i, X_j).
\end{aligned}
\tag{4.20}
$$

For the proof, see Lemma 3.1 of Birkel (1988b).

LEMMA 4.3. *Let $2 \le p < r \le \infty$, let f be an absolutely continuous function satisfying $\sup_{x\in R}|f'(x)| \le B$ and let $\{X_i,\ 1 \le i \le n\}$ be associated random variables with $\|f(X_i)\|_r = (E|f(X_i)|^r)^{1/r} < \infty$. Then, for any two subsets M and J of $\{1, 2, \ldots, n\}$,*

$$
\begin{aligned}
&\left|\operatorname{Cov}\left(\left|\sum_{i\in M} f(X_i)\right|, \left|\sum_{j\in J} f(X_j)\right|^{p-1}\right)\right| \\
&\qquad \le p\left(E\left|\sum_{j\in J} f(X_j)\right|^p\right)^{(r-1)(p-2)/r_p} \left(\sum_{i\in M}\|f(X_i)\|_r\right)^{r(p-2)/r_p} \\
&\qquad\quad \times \left(B^2\sum_{i\in M}\sum_{j\in J}\operatorname{Cov}(X_i, X_j)\right)^{(r-p)/r_p},
\end{aligned}
\tag{4.21}
$$

where $r_p = r(p-1) - p$.

PROOF. Let $A > 0$ be fixed. Put $g(t_i,\ i \in M) = |\sum_{i\in M} f(t_i)|$ and

$$h(t_j,\ j \in J) = \Big|\sum_{j\in J} f(t_j)\Big|^{p-1} I_{\{|\sum_{j\in J} f(t_j)|\le A\}} + A^{p-1} I_{\{|\sum_{j\in J} f(t_j)|>A\}}.$$

It is easy to see that

$$\begin{aligned}\bigg|\mathrm{Cov}\bigg(\Big|\sum_{i\in M} f(X_i)\Big|, \Big|\sum_{j\in J} f(X_j)\Big|^{p-1}\bigg)\bigg| \\ \le |\mathrm{Cov}(g,h)| + \bigg|\mathrm{Cov}\bigg(g, \Big|\sum_{j\in J} f(X_j)\Big|^{p-1} - h\bigg)\bigg|,\end{aligned} \tag{4.22}$$

$$\|\partial g(t_i)/\partial t_i\|_\infty \le B \quad \text{and} \quad \|\partial h(t_j)/\partial t_j\|_\infty \le (p-1)A^{p-2} B.$$

Therefore, by Lemma 4.2,

$$|\mathrm{Cov}(g,h)| \le (p-1)A^{p-2} B^2 \sum_{i\in M}\sum_{j\in J} \mathrm{Cov}(X_i, X_j). \tag{4.23}$$

As for the second term on the right-hand side of (4.22), by Hölder's inequality,

$$\begin{aligned}&\bigg|\mathrm{Cov}\bigg(g, \Big|\sum_{j\in J} f(X_j)\Big|^{p-1} - h\bigg)\bigg| \\ &= \bigg|\mathrm{Cov}\bigg(g, \bigg(\Big|\sum_{j\in J} f(X_j)\Big|^{p-1} - A^{p-1}\bigg) I_{\{|\sum_{j\in J} f(X_j)|>A\}}\bigg)\bigg| \\ &\le \max\bigg\{E\bigg(g\Big|\sum_{j\in J} f(X_j)\Big|^{p-1} I_{\{|\sum_{j\in J} f(X_j)|>A\}}\bigg), \\ &\qquad Eg\, E\Big|\sum_{j\in J} f(X_j)\Big|^{p-1} I_{\{|\sum_{j\in J} f(X_j)|>A\}}\bigg\} \\ &\le \max\bigg\{\sum_{i\in M} E\bigg(|f(X_i)|\Big|\sum_{j\in J} f(X_j)\Big|^{p-1} I_{\{|\sum_{j\in J} f(X_j)|>A\}}\bigg), \\ &\qquad \sum_{i\in M} E|f(X_i)|\, E\Big|\sum_{j\in J} f(X_j)\Big|^{p-1} I_{\{|\sum_{j\in J} f(X_j)|>A\}}\bigg\} \\ &\le \sum_{i\in M} \|f(X_i)\|_r \bigg(E\Big|\sum_{j\in J} f(X_j)\Big|^{(p-1)r/(r-1)} I_{\{|\sum_{j\in J} f(X_j)|>A\}}\bigg)^{(r-1)/r} \\ &\le A^{(p-r)/r} \sum_{i\in M} \|f(X_i)\|_r \bigg(E\Big|\sum_{j\in J} f(X_j)\Big|^{p}\bigg)^{(r-1)/r}.\end{aligned} \tag{4.24}$$

Then, by choosing

$$A = \bigg(\sum_{i\in M} \|f(X_i)\|_r \bigg(E\Big|\sum_{j\in J} f(X_j)\Big|^{p}\bigg)^{(r-1)/r} \Big/ B^2 \sum_{i\in M}\sum_{j\in J} \mathrm{Cov}(X_i, X_j)\bigg)^{r/r_p},$$

(4.22), (4.23) and (4.24) imply (4.21). □

LEMMA 4.4. *Let $0 < \alpha < 1$, $0 < \beta < 1$ and x, a, b, c be nonnegative numbers. If*

$$x \le a + b\,x^{1-\alpha} + c\,x^{1-\beta}, \tag{4.25}$$

then

$$x \le 2a + (4b)^{1/\alpha} + (4c)^{1/\beta}. \tag{4.26}$$

PROOF. It is well known that

$$s^{\theta} t^{1-\theta} \le s + t \quad \text{for any } s \ge 0,\ t \ge 0 \text{ and } 0 \le \theta \le 1. \tag{4.27}$$

Thus, by (4.25)

$$x \le a + b^{1/\alpha}\,4^{(1-\alpha)/\alpha} + x/4 + c^{1/\beta}\,4^{(1-\beta)/\beta} + x/4,$$

which yields (4.26) immediately. □

We are now ready to prove Theorem 4.2.

PROOF OF THEOREM 4.2. Let

$$T_n = \sum_{i=1}^{n} f(X_i) \quad \text{and} \quad r_p = r(p-1) - p.$$

For the sake of convenience of statement, we assume that $\{X, X_n,\ n \ge 1\}$ is a strictly stationary sequence of associated random variables. We shall prove (4.6) by induction on n. Suppose that, for each $1 \le k < n$,

$$\begin{aligned} E|T_k|^p \le K\Bigg(& k^{1+\varepsilon} E|f(X)|^p + \Bigg(n \max_{i \le k} \sum_{j=1}^{k} |\operatorname{Cov}(f(X_i), f(X_j))|\Bigg)^{p/2} \\ & + k^{(r_p + \theta(p-r))/(r-2) \vee (1+\varepsilon)} \|f(X)\|_r^{r(p-2)/(r-2)} (B^2 C)^{(r-p)/(r-2)}\Bigg). \end{aligned} \tag{4.28}$$

We now prove that (4.28) is still true for $k = n$. Let a, $0 < a < 1/2$, be a constant that will be specified later and let $m = [an] + 1$. Define

$$\xi_l = \sum_{j=1+2(l-1)m}^{n \wedge (2l-1)m} f(X_j),$$

$$\eta_l = \sum_{j=1+(2l-1)m}^{n \wedge 2lm} f(X_j) \quad \text{for } 1 \le l \le k_n := 1 + [n/(2m)],$$

$$T_{1,n} = \sum_{l=1}^{k_n} \xi_l \quad \text{and} \quad T_{2,n} = \sum_{l=1}^{k_n} \eta_l.$$

Clearly, we have

$$E|T_n|^p \le 2^{p-1}\big(E|T_{1,n}|^p + E|T_{2,n}|^p\big)$$

and

$$
\begin{aligned}
E|T_{1,n}|^p &\le \sum_{l=1}^{k_n} E|\xi_l||T_{1,n}|^{p-1} \\
&\le 2^{p-2}\sum_{l=1}^{k_n} E|\xi_l|^p + 2^{p-2}\sum_{l=1}^{k_n} E|\xi_l||T_{1,n}-\xi_l|^{p-1} \qquad (4.29)\\
&:= 2^{p-2}(I_{1,1}+I_{1,2}).
\end{aligned}
$$

From Lemma 4.3, it follows that

$$
\begin{aligned}
I_{1,2} &\le \sum_{l=1}^{k_n} E|\xi_l|\, E|T_{1,n}-\xi_l|^{p-1} \\
&\quad + \sum_{l=1}^{k_n} p(E|T_{1,n}-\xi_l|^p)^{(r-1)(p-2)/r_p} \\
&\qquad \times (m\,\|f(X)\|_r)^{r(p-2)/r_p}(B^2\, m\, u(m))^{(r-p)/r_p} \\
&\le 2^{p-2}\sum_{l=1}^{k_n} E|\xi_l|^p + 2^{p-2}E|T_{1,n}|^{p-1}\sum_{l=1}^{k_n} E|\xi_l| \qquad (4.30)\\
&\quad + p\, m\, C_1 \sum_{l=1}^{k_n}(E|T_{1,n}-\xi_l|^p)^{1-(r-2)/r_p} \\
&\le 2^{p-2}I_{1,1} + 2^{p-2}(E|T_{1,n}|^p)^{(p-1)/p}\sum_{l=1}^{k_n} E|\xi_l| \\
&\quad + p\, 2^{p-1}\, m\, C_1 \sum_{l=1}^{k_n}(E|\xi_l|^p)^{1-(r-2)/r_p} \\
&\quad + p\, 2^{p-1}\, m\, k_n\, C_1\,(E|T_{1,n}|^p)^{1-(r-2)/r_p},
\end{aligned}
$$

where $C_1 = \|f(X)\|_r^{r(p-2)/r_p}(B^2\, u(m))^{(r-p)/r_p}$. By (4.27), we have

$$
m\, C_1 \sum_{l=1}^{k_n}\left(E|\xi_l|^p\right)^{1-(r-2)/r_p} \le \sum_{l=1}^{k_n} E|\xi_l|^p + k_n\,(m\, C_1)^{r_p/(r-2)},
$$

which, together with (4.30), yields

$$
\begin{aligned}
I_{1,2} &\le p2^p I_{1,1} + 2^{p-2}(E|T_{1,n}|^p)^{(p-1)/p}\sum_{l=1}^{k_n} E|\xi_l| \qquad (4.31)\\
&\quad + p\, 2^{p-1}(m\, k_n\, C_1)^{r_p/(r-2)} + p\, 2^{p-1}(m\, k_n\, C_1)(E|T_{1,n}|^p)^{1-(r-2)/r_p}.
\end{aligned}
$$

In what follows D stands for a generic positive constant that depends only on p and r; however, it may take different values in each appearance. From

(4.29), (4.31) and Lemma 4.4, it follows that

$$E|T_{1,n}|^p \le D\left\{I_{1,1} + \left(\sum_{l=1}^{k_n}(E\xi_l^2)^{1/2}\right)^p + (m\,k_n\,C_1)^{r_p/(r-2)}\right\}.$$

Put

$$C_2 = \|f(X)\|_r^{r(p-2)/(r-2)}(B^2\,C)^{(r-p)/(r-2)}.$$

Note that

$$ET_n^2 = \sum_{i=1}^n\sum_{j=1}^n \operatorname{Cov}\big(f(X_i), f(X_j)) \le n\,\max_{i\le n}\sum_{j=1}^n|\operatorname{Cov}(f(X_i),\ f(X_j))| := n\,a_n.$$

Then, by (4.5),

$$\begin{aligned}E|T_{1,n}|^p &\le D\left\{\sum_{l=1}^{k_n}E|\xi_l|^p + k_n^p(m\,a_n)^{p/2}\right.\\ &\qquad\left.+ (m\,k_n)^{r_p/(r-2)}\|f(X)\|_r^{r(p-2)/(r-2)}(B^2\,u(m))^{(r-p)/(r-2)}\right\}\\ &\le D\left\{\sum_{l=1}^{k_n}E|\xi_l|^p + k_n^p(m\,a_n)^{p/2} + (m\,k_n)^{r_p/(r-2)}m^{\theta(p-r)/(r-2)}\,C_2\right\}.\end{aligned}$$

Similarly, we have

$$E|T_{2,n}|^p \le D\left\{\sum_{l=1}^{k_n}E|\eta_l|^p + k_n^p(m\,a_n)^{p/2} + (m\,k_n)^{r_p/(r-2)}m^{\theta(p-r)/(r-2)}\,C_2\right\}.$$

Therefore,

$$\begin{aligned}E|T_n|^p \le D\Bigg\{&\sum_{l=1}^{k_n}E|\xi_l|^p + \sum_{l=1}^{k_n}E|\eta_l|^p + k_n^p(m\,a_n)^{p/2}\\ &+ (m\,k_n)^{r_p/(r-2)}m^{\theta(p-r)/(r-2)}\,C_2\Bigg\}.\end{aligned}\tag{4.32}$$

Without loss of generality, we assume $0 < \varepsilon < \min(1, p/2 - 1)$. Let

$$a = (16D)^{-1/\varepsilon} \quad\text{and}\quad K = 2D\big((1+1/a)^p + 2^{r_p/(r-2)}\,a^{\theta(p-r)/(r-2)}\big).$$

By (4.32) and the induction hypothesis (4.28), we have

$$\begin{aligned}E|T_n|^p &\le 2\,D\,k_n\,K\left(m^{1+\varepsilon}E|f(X)|^p + (m\,a_n)^{p/2} + m^{(r_p+\theta(p-r))/(r-2)\vee(1+\varepsilon)}C_2\right)\\ &\quad + D\big\{k_n^p(m\,a_n)^{p/2} + (m\,k_n)^{r_p/(r-2)}m^{\theta(p-r)/(r-2)}\,C_2\big\}\\ &\le 2\,D\,(4a)^\varepsilon\,K\left(n^{1+\varepsilon}E|f(X)|^p + (n\,a_n)^{p/2} + n^{(r_p+\theta(p-r))/(r-2)\vee(1+\varepsilon)}C_2\right)\\ &\quad + D(1+1/a)^p(n\,a_n)^{p/2}\\ &\quad + D2^{r_p/(r-2)}a^{\theta(p-r)/(r-2)}\,n^{(r_p+\theta(p-r))/(r-2)\vee(1+\varepsilon)}C_2\end{aligned}$$

$$\le (K/2)\big(n^{1+\varepsilon}E|f(X)|^p + (n\,a_n)^{p/2} + n^{(r_p+\theta(p-r))/(r-2)\vee(1+\varepsilon)}C_2\big)$$
$$+ (K/2)\big((n\,a_n)^{p/2} + n^{(r_p+\theta(p-r))/(r-2)\vee(1+\varepsilon)}C_2\big)$$
$$\le K\big(n^{1+\varepsilon}E|f(X)|^p + (n\,a_n)^{p/2} + n^{(r_p+\theta(p-r))/(r-2)\vee(1+\varepsilon)}C_2\big).$$

This proves that (4.28) holds for $k = n$. The proof is now complete. □

5. Proofs.

PROOF OF THEOREM 2.1. By Theorem 4.2 in Billingsley (1968), it is sufficient to prove that for any $\varepsilon > 0$,

$$\lim_{\theta\to 0}\limsup_{n\to\infty} P\Big\{\sup_{0<t\le\theta}|\alpha_n(t)/q(t)| \ge \varepsilon\Big\} = 0, \tag{5.1}$$

$$\lim_{\theta\to 0}\limsup_{n\to\infty} P\Big\{\sup_{1-\theta\le t<1}|\alpha_n(t)/q(t)| \ge \varepsilon\Big\} = 0, \tag{5.2}$$

$$\lim_{\theta\to 0} P\Big\{\sup_{0<t\le\theta}|B^*(t)/q(t)| \ge \varepsilon\Big\} = 0 \tag{5.3}$$

and

$$\lim_{\theta\to 0} P\Big\{\sup_{1-\theta\le t<1}|B^*(t)/q(t)| \ge \varepsilon\Big\} = 0. \tag{5.4}$$

Note that

$$P\Big\{\sup_{0<t\le\theta}|\alpha_n(t)/q(t)| \ge \varepsilon\Big\} \le \sum_{j=1}^{\infty} P\Big\{\sup_{\theta 2^{-j}<t\le\theta 2^{-j+1}}|\alpha_n(t)/q(t)| \ge \varepsilon\Big\}$$
$$\le \sum_{j=1}^{\infty} P\Big\{\sup_{\theta 2^{-j}<t\le\theta 2^{-j+1}}|\alpha_n(t)| \ge \varepsilon q(\theta 2^{-j})\Big\}.$$

Hence, (5.1) can be rewritten as

$$\lim_{\theta\to 0}\limsup_{n\to\infty} P\Big\{\sup_{0<t\le\theta}|\alpha_n(t)/q(t)| \ge \varepsilon\Big\} \le \limsup_{\theta\to 0}\ \limsup_{n\to\infty}\sum_{j=1}^{\infty} B_{j,n}, \tag{5.5}$$

where $B_{j,n} = P\{\sup_{0\le t\le\theta 2^{-j+1}}|\alpha_n(t)| \ge \varepsilon q(\theta 2^{-j})\}$.

In the following discussion, C is a positive constant independent of θ, j and n, and may take different values in each appearance. Put

$$\varepsilon_j = \varepsilon q(\theta 2^{-j}),$$

$$G_n = \{j\colon n^{1/2}\theta 2^{-j+1} \le \varepsilon_j/2\} \quad\text{and}\quad H_n = \{j\colon n^{1/2}\theta 2^{-j+1} > \varepsilon_j/2\}.$$

By the following inequality [cf. (22.17) in Billingsley (1968)], for any $0 \le s \le t \le s+h \le 1$,

$$|\alpha_n(t) - \alpha_n(s)| \le |\alpha_n(s+h) - \alpha_n(s)| + n^{1/2}h, \tag{5.6}$$

we have for $j \in G_n$

$$
\begin{aligned}
B_{j,n} &\le P\{|\alpha_n(\theta 2^{-j+1})| + n^{1/2}\theta 2^{-j+1} \ge \varepsilon_j\} \\
&\le P\{|\alpha_n(\theta 2^{-j+1})| \ge \varepsilon_j/2\} \\
&\le C\,\varepsilon_j^{-2}(\theta 2^{-j})^{r_2} \\
&\le C\,\varepsilon^{-2}(\log(2^j/\theta))^{-2\beta}
\end{aligned} \tag{5.7}
$$

by (A2), (2.3) and (2.4). Hence (5.7) implies that

$$
\limsup_{\theta\to 0}\ \limsup_{n\to\infty} \sum_{j\in G_n} B_{j,n} = 0. \tag{5.8}
$$

Note that in the case of $\mu < r_2/2$, (5.8) holds also for $\beta = 0$.

Write

$$
\Delta := \Delta_{j,n} = \frac{1}{4}\frac{\varepsilon_j}{n^{1/2}} = \frac{\varepsilon}{4}\frac{q(\theta 2^{-j})}{n^{1/2}}. \tag{5.9}
$$

When $j \in H_n$, using (5.6) again, we obtain

$$
\begin{aligned}
B_{j,n} &\le P\Big\{\max_{1\le i\le \theta 2^{-j+1}/\Delta} |\alpha_n(i\Delta)| \ge \varepsilon_j/2\Big\} \\
&\qquad + P\Big\{\max_{0\le i\le \theta 2^{-j+1}/\Delta}\ \sup_{i\Delta<t\le(i+1)\Delta} |\alpha_n(t)-\alpha_n(i\Delta)| \ge \varepsilon_j/2\Big\} \\
&\le P\Big\{\max_{1\le i\le \theta 2^{-j+1}/\Delta} |\alpha_n(i\Delta)| \ge \varepsilon_j/2\Big\} \\
&\qquad + P\Big\{\max_{0\le i\le \theta 2^{-j+1}/\Delta} |\alpha_n((i+1)\Delta)-\alpha_n(i\Delta)| + \Delta n^{1/2} \ge \varepsilon_j/2\Big\} \\
&\le P\Big\{\max_{1\le i\le \theta 2^{-j+1}/\Delta} |\alpha_n(i\Delta)| \ge \varepsilon_j/2\Big\} \\
&\qquad + P\Big\{\max_{0\le i\le \theta 2^{-j+1}/\Delta} |\alpha_n((i+1)\Delta)-\alpha_n(i\Delta)| \ge \varepsilon_j/4\Big\} \\
&\le 3P\Big\{\max_{1\le i\le \theta 2^{-j+2}/\Delta} |\alpha_n(i\Delta)| \ge \varepsilon_j/8\Big\}.
\end{aligned} \tag{5.10}
$$

From (A1) it follows that, for all $0 \le i < k \le \theta 2^{-j+2}/\Delta$,

$$
\begin{aligned}
E|\alpha_n(k\Delta)-\alpha_n(i\Delta)|^p &\le C_1\big(((k-i)\Delta)^{p_1} + n^{-p_2/2}((k-i)\Delta)^{r_1}\big) \\
&\le C_1\big(((k-i)\Delta)^{p_1} + n^{-p_2/2}(k-i)\Delta^{r_1}\big).
\end{aligned}
$$

Thus, by a theorem of Móricz (1982), there is a constant C, depending only on C_1 and p_1, such that

$$
\begin{aligned}
&E\max_{0\le i\le \theta 2^{-j+2}/\Delta} |\alpha_n(i\Delta)|^p \\
&\quad\le C\big((\theta 2^{-j+2}/\Delta)^{p_1}\Delta^{p_1} + n^{-p_2/2}(\theta 2^{-j+2}/\Delta)\Delta^{r_1}\log^p(\theta 2^{-j+2}/\Delta)\big) \\
&\quad\le C\big((\theta 2^{-j})^{p_1} + n^{-p_2/2}\theta 2^{-j}\Delta^{r_1-1}\log^p(\theta 2^{-j+2}/\Delta)\big).
\end{aligned} \tag{5.11}
$$

Since $p_2 > 1 - r_1$ in $(A1)$, $l(x) = \log^p(x)/x^{-1+r_1+p_2} \le C$ for all $x \ge 1$. Thus from (2.3), (2.4), (5.9), (5.11) and the fact that $\theta 2^{-j+4} n^{1/2}/\varepsilon_j \ge 8$, we conclude that for $j \in H_n$,

$$
\begin{aligned}
P\Big\{&\max_{0\le i\le \theta 2^{-j+2}/\Delta} |\alpha_n(i\Delta)| \ge \varepsilon_j/8\Big\}\\
&\le C\varepsilon_j^{-p}\big((\theta 2^{-j})^{p_1} + n^{-p_2/2}\theta 2^{-j}\Delta^{r_1-1}\log^p(\theta 2^{-j+2}/\Delta)\big)\\
&\le C\varepsilon_j^{-p}\big((\theta 2^{-j})^{p_1} + \varepsilon_j^{r_1-1}(\theta 2^{-j}) n^{(1-r_1-p_2)/2}\log^p(\theta 2^{-j+4}n^{1/2}/\varepsilon_j)\big)\\
&\le C\varepsilon_j^{-p}\big((\theta 2^{-j})^{p_1} + \varepsilon_j^{-p_2}(\theta 2^{-j})^{r_1+p_2} l(\theta 2^{-j+4}n^{1/2}/\varepsilon_j)\big)\\
&\le C\big(\varepsilon_j^{-p}(\theta 2^{-j})^{p_1} + \varepsilon_j^{-p-p_2}(\theta 2^{-j})^{r_1+p_2}\big)\\
&\le C\varepsilon^{-p-p_2}(\log(2^j/\theta))^{-p\beta}.
\end{aligned}
$$

This, together with (5.10), proves that

$$
\limsup_{\theta\to 0}\limsup_{n\to\infty}\sum_{j\in H_n} B_{j,n} = 0. \tag{5.12}
$$

Note that in the case of $\mu < p_1/p$, (5.12) is true for $\beta > 1/(p+p_2)$. The proof of (5.1) is now complete by (5.5), (5.8) and (5.12). Similarly one can also prove (5.2).

By (2.1) we have for any $0 \le s,\ t \le 1$,

$$
\alpha_n(t) - \alpha_n(s) \to_{\mathscr{D}} B^*(t) - B^*(s).
$$

Then by (A2) and Theorem 5.3 in Billingsley (1968),

$$
E(B^*(t) - B^*(s))^2 \le \liminf_{n\to\infty} E(\alpha_n(t) - \alpha_n(s))^2 \le C_2|t-s|^{r_2}. \tag{5.13}
$$

Thus, based on the fact that $\{B^*(t),\ 0 \le t \le 1\}$ is a Gaussian process, we have

$$
E(B^*(t) - B^*(s))^4 \le CE^2(B^*(t) - B^*(s))^2 \le C|t-s|^{2r_2}
$$

for all $0 \le s, t \le 1$. Applying Theorem 12.2 in Billingsley (1968) directly, we can immediately get that (5.3) and (5.4) are true. This completes our proof of Theorem 2.1. □

PROOF OF COROLLARY 2.1. First we verify that $\{\Delta_n(t),\ n \ge 1\}$ and $\Delta(t)$ are well defined on [0, 1]. By Schwarz's inequality, (A2), (2.4) and (2.5), for $0 \le t \le 1$,

$$
\begin{aligned}
E\Delta_n^2(t) &= \int_0^t\int_0^t E\alpha_n(u)\alpha_n(v)\,dQ(u)\,dQ(v)\\
&\le \left(\int_0^1 E^{1/2}(\alpha_n(t))^2\,dQ(t)\right)^2\\
&\le 2^{r_2}C_2\left(\int_0^1 (t(1-t))^{r_2/2}\,dQ(t)\right)^2 < \infty,
\end{aligned}
$$

where the last inequality follows from $\alpha_n(0) = \alpha_n(1) = 0$, $E(\alpha_n(t))^2 \le C_2 t^{r_2}$ for $0 \le t \le 1/2$, and $E(\alpha_n(t))^2 \le C_2(1-t)^{r_2}$ for $1/2 \le t \le 1$. Similarly, using (5.13) in conjunction with (A2), we have

$$\begin{aligned} E\Delta^2(t) &= \int_0^t\int_0^t EB^*(u)B^*(v)\,dQ(u)\,dQ(v) \\ &\le \left(\int_0^1 E^{1/2}(B^*(t))^2\,dQ(t)\right)^2 \\ &\le 2^{r_2}C_2\left(\int_0^1 (t(1-t))^{r_2/2}\,dQ(t)\right)^2 < \infty. \end{aligned}$$

This shows that $\{\Delta_n(t),\ 0 \le t \le 1;\ n \ge 1\}$ and $\{\Delta(t),\ 0 \le t \le 1\}$ are square integrable processes. Now we have, for any $\theta > 0$,

$$\sup_{0<t\le\theta} |\Delta_n(t)| \le \sup_{0<t\le\theta} |\alpha_n(t)/q^*(t)| \int_0^1 q^*(t)\,dQ(t)$$

and

$$\sup_{1-\theta\le t<1} |\Delta_n(1) - \Delta_n(t)| \le \sup_{1-\theta\le t<1} |\alpha_n(t)/q^*(t)| \int_0^1 q^*(t)\,dQ(t),$$

where $q^*(t) = (t(1-t))^{\mu}(\log 1/(t(1-t)))^{\beta}$. Thus (5.1), (5.2) and (2.5) imply for any $\varepsilon > 0$,

$$\begin{aligned} &\lim_{\theta\to 0}\limsup_{n\to\infty} P\Big\{\sup_{0<t\le\theta}|\Delta_n(t)| \ge \varepsilon\Big\} \\ &\qquad = \lim_{\theta\to 0}\limsup_{n\to\infty} P\Big\{\sup_{1-\theta\le t<1}|\Delta_n(1)-\Delta_n(t)| \ge \varepsilon\Big\} = 0. \end{aligned} \tag{5.14}$$

Similarly, (5.3), (5.4) and (2.5) imply for any $\varepsilon > 0$,

$$\lim_{\theta\to 0} P\Big\{\sup_{0<t\le\theta}|\Delta(t)| \ge \varepsilon\Big\} = \lim_{\theta\to 0} P\Big\{\sup_{1-\theta\le t<1}|\Delta(1)-\Delta(t)| \ge \varepsilon\Big\} = 0. \tag{5.15}$$

Hence Corollary 2.1 follows from Theorem 2.1 and Theorem 4.2 in Billingsley (1968). □

PROOF OF THEOREM 2.2. Let θ and ε be as in (2.6). Since $\theta \ge 1+\sqrt{2}$, we can take $r = \infty$, v and p in Theorem 4.1 such that

$$\frac{2(\theta+\varepsilon)}{\theta+\varepsilon-1} < v < \frac{2\theta}{\theta-1} \quad \text{and} \quad v < \theta+1 < p < \theta+1+\varepsilon. \tag{5.16}$$

Therefore, by (4.3) and (4.11), for any $0 < \eta < (p-1-\theta)/\theta$, there is $K < \infty$ such that for any $0 \le s, t \le 1$,

$$E\left|\sum_{i=1}^n (I(U_i \le t) - I(U_i \le s) - (t-s))\right|^p \le K\big(n^{p/2}|t-s|^{p/v} + n^{1+\eta/2}\big)$$

and

$$E\left|\sum_{i=1}^{n}(I(U_i \le t) - I(U_i \le s) - (t-s))\right|^2 \le K\,n|t-s|^{2/v},$$

which follows that

$$E|\alpha_n(t) - \alpha_n(s)|^p \le K\bigl(|t-s|^{p/v} + n^{-(p-2-\eta)/2}\bigr)$$

and

$$E|\alpha_n(t) - \alpha_n(s)|^2 \le K|t-s|^{2/v}.$$

Hence (A1) and (A2) hold for $p_1 = p/v > 1$, $p_2 = p - 2 - \eta > 1$, $r_1 = 0$ and $r_2 = 2/v$. Noting that $0 < \eta < (p-1-\theta)/\theta$, it is easy to see from (5.16) that

$$\min\left(\frac{p_1}{p}, \frac{r_1 + p_2}{p + p_2}, \frac{r_2}{2}\right) \ge \min\left(\frac{1}{v}, \frac{p-2-\eta}{2p-2}\right) > \frac{1}{2}\left(1 - \frac{1}{\theta}\right). \tag{5.17}$$

By Theorem 1 of Shao (1986), (2.1) holds. This proves Theorem 2.2 by Theorem 2.1. □

PROOF OF THEOREM 2.3. From Theorem 2 of Shao (1986) (cf. Remark 1 there) it follows that (2.1) holds. By Theorem 1.1 of Shao (1995), we have for $p \ge 2$,

$$\begin{aligned}
&E\left|\sum_{i=1}^{n}(I(U_i \le t) - I(U_i \le s) - (t-s))\right|^p \\
&\quad\le C\left(n^{p/2}\exp\left(C\sum_{i=0}^{[\log_2 n]}\rho(2^i)\right)(E(I(U_1 \le t) - I(U_1 \le s) - (t-s))^2)^{p/2}\right. \\
&\qquad\qquad \left. + n\exp\left(K\sum_{i=0}^{[\log_2 n]}\rho^{2/p}(2^i)\right)E|I(U_1 \le t) - I(U_1 \le s) - (t-s)|^p\right) \\
&\quad\le C\left(n^{p/2}\exp\left(C\sum_{i=0}^{[\log_2 n]}\rho(2^i)\right)|t-s|^{p/2} + n\exp\left(C\sum_{i=0}^{[\log_2 n]}\rho^{2/p}(2^i)\right)|t-s|\right),
\end{aligned}$$

where $[x]$ denotes the integer part of x. Clearly, under the condition (2.7), we have for $p \ge 2$ and any $0 < \delta < \min\{\varepsilon(p-1)/(2(2+\varepsilon)), (p-2)/2\}$,

$$\begin{aligned}
&E\left|\sum_{i=1}^{n}(I(U_i \le t) - I(U_i \le s) - (t-s))\right|^p \\
&\quad\le C\left(n^{p/2}|t-s|^{p/2} + n\exp\left(C\sum_{i=0}^{[\log_2 n]}\rho^{2/p}(2^i)\right)|t-s|\right) \\
&\quad\le C\bigl(n^{p/2}|t-s|^{p/2} + n^{1+\delta}|t-s|\bigr),
\end{aligned}$$

since $\exp(C\sum_{i=0}^{[\log_2 n]}\rho^{2/p}(2^i))$ is a slowly varying function. This, in turn, gives us for $p > 2$,

$$E|\alpha_n(t) - \alpha_n(s)|^p \le C\big(|t-s|^{p/2} + n^{-(p-2-2\delta)/2}|t-s|\big)$$

and for $p = 2$,

$$E|\alpha_n(t) - \alpha_n(s)|^2 \le C|t-s|.$$

Thus (A1) and (A2) are satisfied for $p_1 = p/2$, $p_2 = p - 2 - 2\delta > 0$ and $r_1 = r_2 = 1$. Hence by (2.4),

$$\mu = (1 + p - 2 - 2\delta)/(p + p - 2 - 2\delta) > 1/(2+\varepsilon). \tag{5.18}$$

On the other hand, under the condition (2.8), we have for $p \ge 2$,

$$E\left|\sum_{i=1}^{n}(I(U_i \le t) - I(U_i \le s) - (t-s))\right|^p \le C\big(n^{p/2}|t-s|^{p/2} + n|t-s|\big),$$

which implies that for $p \ge 2$,

$$E|\alpha_n(t) - \alpha_n(s)|^p \le C\big(|t-s|^{p/2} + n^{-(p-2)/2}|t-s|\big).$$

Thus (A1) and (A2) are satisfied for $p_1 = p/2$, $p_2 = p - 2$ and $r_1 = r_2 = 1$. Obviously $\mu = 1/2$. Now the proof of Theorem 2.3 is complete. □

Before proving Theorem 2.4, we need the following lemma.

LEMMA 5.1. *Let X and Y be associated random variables with a common uniform distribution over* $[0,1]$. *Then for any* $0 \le s < t \le 1$,

$$|\operatorname{Cov}(I(s < X \le t),\ I(s < Y \le t))| \le 4(t-s)^{1/3}(\operatorname{Cov}(X,Y))^{1/3}.$$

PROOF. Let $u = \operatorname{Cov}(X,Y)$ and $h(x) = I(s < x \le t)$. It suffices to show that

$$Eh(X)h(Y) - (t-s)^2 \le 4(t-s)^{1/3}u^{1/3} \tag{5.19}$$

and

$$(t-s)^2 - Eh(X)h(Y) \le 3(t-s)^{2/3}u^{1/3}. \tag{5.20}$$

If $u \ge (t-s)^2$, then (5.19) holds obviously. When $u < (t-s)^2$, put

$$a = \begin{cases} u^{1/4}, & \text{if } (t-s)^4 \le u < (t-s)^2,\\ (u/(t-s))^{1/3}, & \text{if } u < (t-s)^4 \end{cases}$$

and define

$$f_1(x) = \begin{cases} 0, & \text{if } x < s-a,\\ 1 + (x-s)/a, & \text{if } s-a \le x < s,\\ 1, & \text{if } s \le x \le t,\\ 1 + (t-x)/a, & \text{if } t < x \le t+a,\\ 0, & \text{if } x > t+a. \end{cases} \tag{5.21}$$

By Lemma 4.2, we have

$$
\begin{aligned}
&Eh(X)h(Y) - (t-s)^2 \\
&\quad \le Ef_1(X)f_1(Y) - (t-s)^2 \\
&\quad = \operatorname{Cov}(f_1(X), f_1(Y)) + Ef_1(X)Ef_1(Y) - (t-s)^2 \\
&\quad \le |\operatorname{Cov}(f_1(X), f_1(Y))| + (t-s+a)^2 - (t-s)^2 \\
&\quad \le a^{-2}\operatorname{Cov}(X, Y) + 2a(t-s) + a^2 \\
&\quad = \begin{cases} 2u^{1/2} + 2u^{1/4}(t-s), & \text{if } (t-s)^4 \le u < (t-s)^2, \\ 3u^{1/3}(t-s)^{2/3} + (u/(t-s))^{2/3}, & \text{if } u < (t-s)^4, \end{cases} \\
&\quad \le \begin{cases} 2u^{1/3}(t-s)^{1/3} + 2u^{1/4}(t-s)^{2/3}u^{1/12}, & \text{if } (t-s)^4 \le u < (t-s)^2, \\ 3u^{1/3}(t-s)^{2/3} + u^{1/3}(t-s)^{4/3}/(t-s)^{2/3}, & \text{if } u < (t-s)^4, \end{cases} \\
&\quad \le 4u^{1/3}(t-s)^{1/3},
\end{aligned}
$$

as desired.

As for (5.20), it is obviously true if $u \ge (t-s)^4/8$. When $u < (t-s)^4/8$, let $a = (u/(t-s))^{1/3} < (t-s)/2$ and define

$$
f_2(x) = \begin{cases} 0, & \text{if } x < s, \\ 1 + (x-s-a)/a, & \text{if } s \le x < s+a, \\ 1, & \text{if } s+a \le x \le t-a, \\ 1 + (t-a-x)/a, & \text{if } t-a < x \le t, \\ 0, & \text{if } x > t. \end{cases} \tag{5.22}
$$

Then, we have

$$
\begin{aligned}
(t-s)^2 - Eh(X)h(Y) &\le (t-s)^2 - Ef_2(X)f_2(Y) \\
&= (t-s)^2 - Ef_2(X)Ef_2(Y) - \operatorname{Cov}(f_2(X), f_2(Y)) \\
&\le (t-s)^2 - (t-s-a)^2 + |\operatorname{Cov}(f_2(X), f_2(Y))| \\
&\le 2a(t-s) + a^{-2}\operatorname{Cov}(X, Y) \\
&= 3u^{1/3}(t-s)^{2/3}.
\end{aligned}
$$

This proves (5.20). □

PROOF OF THEOREM 2.4. Let ν and ε be as in (2.9). Clearly, (2.9) implies

$$
u(n) = \max_{i \le n} \sum_{j:\ |j-i| \ge n} \operatorname{Cov}(U_i, U_j) = O(n^{-(\nu-1+\varepsilon)}).
$$

Thus, we can take $p = \nu + 1 + \varepsilon$ and $r = \infty$ in Theorem 4.2. Since $p > (5 + \sqrt{33})/2$, we can find an η such that

$$
0 < \eta < \min\{(p^2 - 5p - 2)/(2p),\ \varepsilon/\nu\}. \tag{5.23}
$$

Therefore, by (2.9) and (4.7), for such an η and for any absolutely continuous function f with $|f| \le 1$ and $\sup|f'| \le B$, where $B \ge 1$, there is $K < \infty$ such that

$$E\left|\sum_{i=1}^{n} f(U_i)\right|^p \le K\left(n^{1+\eta}B^2 + \left(n \max_{i\le n}\sum_{j=1}^{n} |\operatorname{Cov}(f(U_i),\ f(U_j))|\right)^{p/2}\right). \tag{5.24}$$

To use (5.24) to derive (A1), we have to use two absolutely continuous functions f_1 and f_2 defined by (5.21) and (5.22), respectively, but with a common $0 < a < (t-s)/2$ that will be specified later. Clearly,

$$\sup_{x\in R}|f_k'(x)| \le 1/a \quad \text{and} \quad E|f_k(U_n) - Ef_k(U_n) - z_n|^v \le 2a \qquad \text{for } v \ge 1 \text{ and } k = 1, 2, \tag{5.25}$$

where $z_n = I(s < U_n \le t) - (t-s)$. In addition, $f_2(x) \le I(s < x \le t) \le f_1(x)$ and

$$0 \le f_1(x) - f_2(x) \le I(s-a < x < s+a) + I(t-a < x < t+a),$$

which yield

$$\left|\sum_{i=1}^{n} z_i\right| \le 4na + \left|\sum_{i=1}^{n}(f_1(U_i) - Ef_1(U_i))\right| + \left|\sum_{i=1}^{n}(f_2(U_i) - Ef_2(U_i))\right|. \tag{5.26}$$

From stationarity, (5.25) and Lemma 4.2 it follows that

$$\begin{aligned}
&\max_{i\le n}\sum_{j=1}^{n} |\operatorname{Cov}(f_k(U_i), f_k(U_j))| \\
&\quad\le \operatorname{Var}(f_k(U_1)) + 2\sum_{i=2}^{n} |\operatorname{Cov}(f_k(U_1), f_k(U_i))| \\
&\quad\le K_1\left\{Ez_1^2 + a + \sum_{i=2}^{n}(a + |Ez_1z_i|) \wedge a^{-2}\operatorname{Cov}(U_1, U_i)\right\} \\
&\quad\le K_1\left\{\sum_{i=1}^{n}|Ez_1z_i| + \sum_{i\le a^{-3/(\nu+\varepsilon)}} a + \sum_{i>a^{-3/(\nu+\varepsilon)}} a^{-2}\, i^{-\nu-\varepsilon}\right\} \\
&\quad\le K_1\left\{\sum_{i=1}^{n}|Ez_1z_i| + a^{1-3/(\nu+\varepsilon)}\right\}.
\end{aligned}$$

Here, and in the sequel, K_1 is a generic positive constant that depends only on ν and ε; however, it may take different values in each appearance. Thus, by (5.24) and (5.26),

$$\begin{aligned}
E\left|\sum_{i=1}^{n} z_i\right|^p &\le 3^{p-1}\left((4na)^p + \sum_{k=1}^{2} E\left|\sum_{i=1}^{n}(f_k(U_i) - Ef_k(U_i))\right|^p\right) \\
&\le K_1\left((na)^p + n^{1+\eta}a^{-2} + \sum_{k=1}^{2}\left(n\max_{i\le n}\sum_{j=1}^{n}|\operatorname{Cov}(f_k(U_i), f_k(U_j))|\right)^{p/2}\right) \\
&\le K_1\left((na)^p + n^{1+\eta}a^{-2} + (na^{1-3/(\nu+\varepsilon)})^{p/2} + \left(n\sum_{i=1}^{n}|Ez_1z_i|\right)^{p/2}\right).
\end{aligned}$$

Consequently, by choosing $a = n^{-(p-1-\eta)/(p+2)}$, we have

$$E\left|\sum_{i=1}^{n} z_i\right|^p \le K_1\left\{n^{(3+\eta)p/(p+2)} + \left(n^{1-(1-3/(\nu+\varepsilon))(p-1-\eta)/(p+2)}\right)^{p/2} + \left(n\sum_{i=1}^{n}|Ez_1z_i|\right)^{p/2}\right\}$$
$$\le K_1\left\{n^{(3+\eta)p/(p+2)} + \left(n\sum_{i=1}^{n}|Ez_1z_i|\right)^{p/2}\right\}$$

and hence

$$E|\alpha_n(t)-\alpha_n(s)|^p \le K_1\left(n^{-p(p-4-2\eta)/(2(p+2))} + \left(\sum_{i=1}^{n}|Ez_1z_i|\right)^{p/2}\right). \tag{5.27}$$

On the other hand, from Lemma 5.1 we get

$$\begin{aligned} E(\alpha_n(t)-\alpha_n(s))^2 &\le 2\sum_{i=1}^{n}|Ez_1z_i| \\ &\le 2(t-s)^{1-3/(\nu+\eta)}\sum_{i=1}^{\infty}|Ez_1z_i|^{3/(\nu+\eta)} \\ &\le K_1(t-s)^{1-3/(\nu+\eta)}\sum_{i=1}^{\infty}(t-s)^{1/(\nu+\eta)}\operatorname{Cov}(U_1,U_i)^{1/(\nu+\eta)} \\ &\le K_1(t-s)^{1-2/(\nu+\eta)}. \end{aligned} \tag{5.28}$$

Therefore, (A2) is satisfied for $r_2 = 1-2/(\nu+\eta)$. Substituting (5.28) into (5.27), we obtain

$$E|\alpha_n(t)-\alpha_n(s)|^p \le K_1\left(n^{-p(p-4-2\eta)/(2(p+2))} + (t-s)^{(1-2/(\nu+\eta))p/2}\right).$$

Hence (A1) holds for $p_1 = (1-2/(\nu+\eta))p/2 > 1$, $p_2 = p(p-4-2\eta)/(p+2) > 1$ [by (5.23)] and $r_1 = 0$. It is easy to see that

$$\begin{aligned} \min\left(\frac{p_1}{p}, \frac{r_1+p_2}{p+p_2}, \frac{r_2}{2}\right) &= \min\left(\frac{1}{2}\left(1-\frac{2}{\nu+\eta}\right), \frac{p-4-2\eta}{2(p-1-\eta)}\right) \\ &> \frac{1}{2}\left(1-\frac{3}{\nu}\right). \end{aligned} \tag{5.29}$$

Similar to the proof of Theorem 2.2 of Yu (1993b), (2.1) holds (cf. Remark 2.1). This proves Theorem 2.4 by Theorem 2.1. □

The proofs of Corollaries 2.2, 2.3 and 2.4 are similar to that of Corollary 2.1. Note that we use (5.17), (5.18) and (5.29), in conjunction with Remark 2.4, to derive moment conditions in each corollary. The details are omitted.

PROOF OF THEOREM 3.1. First we have

$$M_n(x)-M_F(x) = (1-F_n(x))^{-1}\left(-\int_x^{\infty}(F_n(t)-F(t))\,dt + M_F(x)(F_n(x)-F(x))\right).$$

Hence, Theorem 3.1 follows by Theorem 2.1 of Yu (1993b) if we can show that

$$I_n = \int_0^\infty |F_n(t) - F(t)|\, dt \to 0 \quad \text{a.s.}$$

By (3.1), we know that

$$EX = \int_0^\infty (1 - F(t))\, dt < \infty.$$

Therefore, for $\varepsilon > 0$ arbitrarily small, we can choose $\beta > 0$ so large that

$$I^{(1)}(\beta) = \int_\beta^\infty (1 - F(t))\, dt < \varepsilon/2.$$

Thus

$$I_n \le I^{(1)}(\beta) + I_n^{(2)}(\beta) + I_n^{(3)}(\beta),$$

where

$$I_n^{(2)}(\beta) = n^{-1} \sum_{i=1}^n \int_\beta^\infty I(X_i > t)\, dt$$

and

$$\begin{aligned} I_n^{(3)}(\beta) &= \int_0^\beta |F_n(t) - F(t)|\, dt \\ &\le \beta \sup_{0 \le x < \infty} |F_n(x) - F(x)| \to 0 \quad \text{a.s.} \end{aligned}$$

by (3.1) and Theorem 2.1 of Yu (1993b).

Note that $\int_\beta^\infty I(X > t)\, dt$ is an absolutely continuous and nondecreasing function of X with $E \int_\beta^\infty I(X > t)\, dt = \int_\beta^\infty (1 - F(t))\, dt = I^{(1)}(\beta)$. Hence $\{\int_\beta^\infty I(X_n > t)\, dt, n \ge 1\}$ is a sequence of associated random variables and

$$\operatorname{Cov}\left(\int_\beta^\infty I(X_i > t)\, dt, \int_\beta^\infty I(X_j > t)\, dt \right) \le \operatorname{Cov}(X_i, X_j) \quad \text{for all } i, j = 1, 2, \dots,$$

by Remark 4 of Yu (1993b). This shows that, by (3.1),

$$\sum_{n=1}^\infty \frac{1}{n^2} \operatorname{Cov}\left(\int_\beta^\infty I(X_n > t)\, dt, \sum_{i=1}^n \int_\beta^\infty I(X_i > t)\, dt \right) < \infty.$$

Thus, applying Theorem 2 of Birkel (1989), we get

$$\begin{aligned} I_n^{(2)}(\beta) &\to E \int_\beta^\infty I(X > t)\, dt \quad \text{a.s.} \\ &= I^{(1)}(\beta). \end{aligned}$$

Therefore, $\limsup_{n\to\infty} I_n \le \varepsilon$ a.s. for all small ε and this completes the proof of Theorem 3.1. □

PROOF OF THEOREM 3.2. Part (i) follows from Theorem 2.4 and Corollary 2.4.

Part (ii) follows from Theorem 2.4 and Corollary 2.4, except that we have to verify that $EX^{2/(1-3/\nu)} < \infty$ implies

$$\int_0^1 (1/q(t))^{2/(1-3/\nu)}\, dt < \infty, \tag{5.30}$$

where $1/q(t) = (M_F(Q(t)) + Q(t)) = (1-t)^{-1}\int_t^1 Q(u)\, du$. In fact, by (3.2),

$$z_n(Q(t)) = (1 - F_n(Q(t)))^{-1}\left(-\int_t^1 \alpha_n(s)\, Q(s) + M_F(Q(t))\alpha_n(t)\right)$$

and the function $M_F(Q(t))$ in the second term above is bounded by $0 \le M_F(Q(t)) \le M_F(Q(t)) + Q(t)$. In addition, by integrating by parts,

$$M_F(Q(t)) = -Q(t) + (1-t)^{-1}\int_t^1 Q(u)\, du.$$

It is easy to check that $q(t)$ is a nonincreasing function on (0, 1). Finally, by Hardy's inequality [cf. Hardy, Littlewood and Pólya (1959), page 240], we have

$$\begin{aligned}\int_0^1 \left(\frac{1}{t}\int_0^t Q(1-y)\, dy\right)^{1/\mu} dt &\le \left(\frac{1}{1-\mu}\right)^{1/\mu}\int_0^1 Q^{1/\mu}(y)\, dy \\ &= \left(\frac{1}{1-\mu}\right)^{1/\mu} EX^{1/\mu} < \infty,\end{aligned}$$

where $\mu = (1 - 3/\nu)/2$. This shows that (5.30) is true. Hence our proof of Theorem 3.2 is now complete. □

Acknowledgments. The authors thank Professor Miklós Csörgő for his careful reading of this paper. Thanks also are due the referee and an Associate Editor for their valuable suggestions.

REFERENCES

BARLOW, R. E. and PROSCHAN, F. (1981). *Statistical Theory of Reliability and Life Testing: Probability Models*. To Begin with Publisher, Silver Spring, MD.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BIRKEL, T. (1988a). Moment bounds for associated sequences. *Ann. Probab.* **16** 1184–1193.

BIRKEL, T. (1988b). On the convergence rate in the central limit theorem for associated processes. *Ann. Probab.* **16** 1685–1698.

BIRKEL, T. (1989). A note on the strong law of large numbers for positively dependent random variables. *Statist. Probab. Lett.* **7** 17–20.

BURKE, M. D., CSÖRGŐ, S. and HORVÁTH, L. (1981). Strong approximations of some biometric estimates under random censorship. *Z. Wahrsch. Verw. Gebiete* **56** 87–112.

BURKHOLDER, D. L. (1973). Distribution function inequalities for martingales. *Ann. Probab.* **1** 19–42.

CHIBISOV, D. (1964). Some theorems on the limiting behaviour of empirical distribution functions. *Selected Transl. Math. Statist. Probab.* **9** 147–156.

CSÖRGŐ, M., CSÖRGŐ, S. and HORVÁTH, L. (1986). *An Asymptotic Theory for Empirical Reliability and Concentration Processes. Lecture Notes in Statist.* **33**. Springer, New York.

CSÖRGŐ, M., CSÖRGŐ, S., HORVÁTH, L. and MASON, D. M. (1986a). Weighted empirical and quantile processes. *Ann. Probab.* **14** 31–85.

CSÖRGŐ, M., CSÖRGŐ, S., HORVÁTH, L. and MASON, D. M. (1986b). Normal and stable convergence of integral functions of the empirical distribution function. *Ann. Probab.* **14** 86–118.

CSÖRGŐ, M. and HORVÁTH, L. (1986). Approximations of weighted empirical and quantile processes. *Statist. Probab. Lett.* **4** 275–280.

CSÖRGŐ, M. and HORVÁTH, L. (1993). *Weighted Approximations in Probability and Statistics*. Wiley, Chichester.

DAVYDOV, YU. A. (1970). The invariance principle for stationary processes. *Theory Probab. Appl.* **15** 487–498.

ESARY, J., PROSCHAN, F. and WALKUP, D. (1967). Association of random variables with applications. *Ann. Math. Statist.* **38** 1466–1474.

HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1959). *Inequalities*, 2nd ed. Cambridge Univ. Press.

HÖEFFDING, W. (1973). On the centering of a simple linear rank statistic. *Ann. Statist.* **1** 54–66.

MÓRICZ, F. (1982). A general moment inequality for the maximum of partial sums of single series. *Acta Sci. Math.* **44** 67–75.

O'REILLY, N. (1974). On the weak convergence of empirical processes in sup-norm metrics. *Ann. Probab.* **2** 642–651.

PELIGRAD, M. (1994). Convergence of stopped sums of weakly dependent random variables. Preprint.

RÉNYI, A. (1953). On the theory of order statistics. *Acta Math. Sci. Hung.* **4** 191–232.

RIO, E. (1993). Covariance inequalities for strongly mixing processes. *Ann. Inst. H. Poincaré* **29** 587–597.

SHAO, Q.-M. (1986). Weak convergence of multidimensional empirical processes for strong mixing sequences. *Chinese Ann. Math. Ser. A* **7** 547–552.

SHAO, Q.-M. (1995). Maximal inequality for partial sums of ρ-mixing sequences. *Ann. Probab.* **23** 948–965.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

YANG, G. L. (1978). Estimation of a biometric function. *Ann. Statist.* **6** 112–116.

YOKOYAMA, R. (1980). Moment bounds for stationary mixing sequences. *Z. Wahrsch. Verw. Gebiete* **52** 45–57.

YU, H. (1993a). Weak convergence for empirical and quantile processes of associated sequences with applications to reliability and economics. Ph.D. dissertation, Carleton Univ., Ottawa.

YU, H. (1993b). A Glivenko–Cantelli lemma and weak convergence for empirical processes of associated sequences. *Probab. Theory Related Fields* **95** 357–370.

DEPARTMENT OF MATHEMATICS
NATIONAL UNIVERSITY OF SINGAPORE
SINGAPORE 119260
REPUBLIC OF SINGAPORE
E-MAIL: matsqm@leonis.nus.sg

DEPARTMENT OF STATISTICAL
AND ACTUARIAL SCIENCES
UNIVERSITY OF WESTERN ONTARIO
LONDON, ONTARIO N6A 5B7
CANADA
E-MAIL: hyu@fisher.stats.uwo.ca