Statistics & Probability Letters 47 (2000) 115–124

STATISTICS & PROBABILITY LETTERS

www.elsevier.nl/locate/stapro

A more general central limit theorem for *m*-dependent random variables with unbounded *m*

Joseph P. Romano[a,*], Michael Wolf[b]

[a] *Department of Statistics, Stanford University, Stanford, CA 94305, USA*
[b] *Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Madrid, Spain*

Received June 1998; received in revised form July 1999

Abstract

In this article, a general central limit theorem for a triangular array of *m*-dependent random variables is presented. Here, *m* may tend to infinity with the row index at a certain rate. Our theorem is a generalization of previous results. Some examples are given that show that the generalization is useful. In particular, we consider the limiting behavior of the sample mean of a combined sample of independent long-memory sequences, the limiting behavior of a spectral estimator, and the moving blocks bootstrap distribution. The examples make it clear the consideration of asymptotic behavior with the amount of dependence *m* increasing with *n* is useful even when the underlying processes are weakly dependent (or even independent), because certain natural statistics that arise in the analysis of time series have this structure. In addition, we provide an example to demonstrate the sharpness of our result.

Keywords: Central limit theorem; *m*-dependent random variables

1. Introduction

Central limit theorems for *m*-dependent random variables (with *m* fixed) have been proved by Hoeffding and Robbins (1948), Diananda (1955), Orey (1958) and Bergstrom (1970). Berk (1973) proved a theorem for the case of a triangular array with unbounded *m*, that is, *m* may be a function of the row index and tend to infinity at a certain rate. However, his theorem is somewhat limited by the fact that it only allows for a moderate amount of positive dependence. The variance of the sum of the random variables in the *n*th row eventually needs to be of the order of the corresponding sample size. Our theorem is intended to extend Berks' result and allow for stronger dependence structures as well. The point of this paper is to present the theorem and illustrate it with a number of examples. We would like to note, however, that it will be a necessary tool for developing inferential procedures for moderately dependent and long-range-dependent random variables; these are research topics under current investigation.

* Corresponding author. Tel.: +1-650-723-6326; fax: +1-650-725-8977.
E-mail address: romano@stat.stanford.edu (J.P. Romano)

0167-7152/00/$ - see front matter
PII: S0167-7152(99)00146-7

Several other authors have also considered the problem of limit theorems for triangular arrays of weakly dependent random variables, such as Krieger (1984), Samur (1984), and Peligrad (1996). While these authors consider more general dependence structures based on mixing notions, their results do not imply our result because in essence they assume that the mixing coefficients across rows are decaying uniformly in the row index. In our theorem (as in Berk's), the strength of dependence in the nth row of the array increases with n. Stronger dependence, however, is allowed in Rio (1995), in particular, see his Corollary 1. But, his result still imposes stronger assumptions than our result in some cases.

The paper is organized as follows. Section 2 contains the main theorem. Some examples that show that our generalizations are useful are presented in Section 3. These include: the sample mean of a combined sample of independent long-memory sequences; estimates of the variance of the sample mean, and the corresponding spectral estimates; the moving blocks bootstrap distribution. The sharpness of our result is demonstrated in Section 4. The proof of the theorem appears in the appendix.

2. The theorem

Theorem 2.1. *Let $\{X_{n,i}\}$ be a triangular array of mean zero random variables. For each $n = 1,2,\ldots$ let $d = d_n$, $m = m_n$, and suppose $X_{n,1},\ldots,X_{n,d}$ is an m-dependent sequence of random variables. Define*

$$B^2_{n,k,a} \equiv \operatorname{Var}\left(\sum_{i=a}^{a+k-1} X_{n,i}\right),$$

$$B^2_n \equiv B_{n,d,1} \equiv \operatorname{Var}\left(\sum_{i=1}^{d} X_{n,i}\right).$$

Assume the following conditions hold. For some $\delta > 0$ and some $-1 \leqslant \gamma < 1$:

$$E|X_{n,i}|^{2+\delta} \leqslant \Delta_n \quad \text{for all } i, \tag{1}$$

$$B^2_{n,k,a}/(k^{1+\gamma}) \leqslant K_n \quad \text{for all } a \text{ and for all } k \geqslant m, \tag{2}$$

$$B^2_n/(dm^{\gamma}) \geqslant L_n, \tag{3}$$

$$K_n/L_n = O(1), \tag{4}$$

$$\Delta_n/L_n^{(2+\delta)/2} = O(1), \tag{5}$$

$$m^{1+(1-\gamma)(1+2/\delta)}/d \to 0. \tag{6}$$

Then, $B_n^{-1}(X_{n,1} + \cdots + X_{n,d}) \Rightarrow N(0,1)$.

Remark 2.1. Note that our theorem extends the previous result by Berk (1973) in two ways. Berk essentially proved this theorem for the special case $\gamma = 0$. Condition (ii) of his theorem corresponds to assumption (2) of our theorem with γ replaced by 0. The greater generality of our theorem is needed to accommodate stronger dependence structures. For example, it can handle the situation of $\operatorname{Var}(X_{n,1} + \cdots + X_{n,d}) \sim d^{1+\gamma}$, for positive $\gamma < 1$. Note that for $\gamma > 0$ the condition on m in (6) becomes weaker. It is therefore desirable to have this extension, rather than making $\gamma = 0$ work by a proper standardization of the $X_{n,i}$ sequence. Second, unlike Berk's theorem, our conditions permit the bounding constants in (1)–(3) to depend on the row index n. On the one hand, this allows for greater ease of using the theorem. For example, to satisfy the moment bound

(1), one otherwise might have to worry about a proper standardization of the $X_{n,i}$ in the first place. On the other hand, and more importantly, it might result in weaker conditions on m again. If the bound L_n in (3) tends to infinity with n, then the power $(2+\delta)/2$ in (5) can play a crucial rule.

We will illustrate our theorem with some examples in the next section. They show that the greater generality of our theorem is useful. Example 3.1 utilizes the extension for $\gamma > 0$. Example 3.2 benefits from the fact that the moment bounds Δ_n, K_n, and L_n may depend on the row index n.

3. Examples

Example 3.1 (*Averaging long-memory processes*). In the recent literature, there has been an increasing interest in time series exhibiting long memory. Let $\{Z_i\}$ be a stationary time series. Let $\overline{Z}_n = n^{-1}\sum_{i=1}^n Z_i$, the sample mean of the observations $Z_1,\dots,Z_n$. Assuming the variance of $\overline{Z}_n$ decays to zero proportional to $n^{-\alpha}$ for some $0<\alpha\leqslant 1$, that is, assuming

$$\mathrm{Var}(\overline{Z}_n) \simeq c_{\mathrm{var}} n^{-\alpha} \quad \text{for a positive constant } c_{\mathrm{var}}, \tag{7}$$

we have short-range dependence in the case $\alpha=1$ and long-range dependence (or long memory) in the case $\alpha<1$. For a general discussion of long-memory processes, see Beran (1994).

It is usual to characterize a long-memory process by the number $H=1-\alpha/2$, the so-called Hurst number; note that $0.5<H<1$. It is well known that the sample mean of a long-memory process can have a quite different asymptotic behavior compared to the sample mean of a short-memory process. To be specific, consider the case $Z_i=G(Y_i)$, where $\{Y_i\}$ is a stationary Gaussian process and $G(\cdot)$ is a sufficiently smooth function. We assume, without loss of generality, that Z_i has mean zero. It turns out that the limiting behavior of $\overline{Z}_n$ depends on the Hermite rank m_{Her} of $G(\cdot)$; see, for example, Chapter 3 of Beran (1994). Beran uses the notation m for this rank, but we will employ m_{Her} to avoid confusion with the parameter m indicating m-dependence. If $1-1/(2m_{\mathrm{Her}})<H<1$, it can be shown that $n^{-\beta}(Z_1+\cdots+Z_n)$ converges to a nondegenerate limiting distribution, where $\beta=1+m_{\mathrm{Her}}(H-1)>0.5$. The limiting distribution is normal if $m_{\mathrm{Her}}=1$, but non-normal otherwise.

In the latter case, we therefore have a non-normal limiting distribution for the properly standardized sample mean. We ask under which conditions we can get limiting normality if we combine independent long-memory sequences of this kind. For instance, this situation may arise when the results of independent experiments are pooled.

Let $\{Z_i^1\},\{Z_i^2\},\dots$ be independent long-memory processes having the same distribution as $\{Z_i\}$ defined above (with $1-1/(2m_{\mathrm{Her}})<H<1$ and $m_{\mathrm{Her}}>1$). For $l=1,\dots,h$ and $j=1,\dots,m$, let $X_{n,(l-1)m+j}=Z_j^l$. Finally, let $d=hm$. This means that the $X_{n,i}$ sequence is obtained by concatenating h independent Z_i sequences, each one of length m. Obviously, $\{X_{n,i}\}$ is m-dependent. Note that we could easily extend this discussion to the case where we include m_j data points from the process $\{Z_i^j\}$ by considering $m=\min_{1\leqslant j\leqslant h}\{m_j\}$. Under which conditions on h and m will we get limiting normality as d tends to infinity? The trivial cases are $m\equiv 1$, where we will get it by the standard CLT, and $h\equiv 1$, where we will not get it for the aforementioned reasons. Hence, let us focus on the scenario of both m and h tending to infinity, applying Theorem 2.1.

We assume a finite $2+\delta$ moment to ensure bound (1). For $k\geqslant m$ and m sufficiently large, we will have $B_{n,k,a}^2\leqslant 2c_{\mathrm{var}}k^{1+\gamma}$ due to (7). Here, $\gamma=1-\alpha>0$. So we can take $K_n=2c_{\mathrm{var}}$ in inequality (2). Next, note that by (7) again for large d, $B_n^2\geqslant 0.5c_{\mathrm{var}}hm^{1+\gamma}=0.5c_{\mathrm{var}}dm^{\gamma}$. Hence, we can take $L_n=0.5c_{\mathrm{var}}$ in inequality (3). Bounds (4) and (5) are now trivially satisfied. Since $d=hm$, the rate condition (6) simplifies to $m^{(1-\gamma)(1+2/\delta)}/h\to 0$. Typically, γ is not known, but it can be estimated from the data; see Chapters 4–7 of Beran (1994), for example. Another approach would be to use the conservative choice $\gamma=0$ in practice.

Example 3.2 (*Variance of the sample mean and spectral estimation*). Focus is on the problem of estimating the variance of the (standardized) sample mean of a strictly stationary time series. Assume that $\{\dots, Z_{-1}, Z_0, Z_1, \dots\}$ is an infinite sequence of strictly stationary random variables having, without loss of generality, mean zero. The sample mean based on observing the finite segment $Z_1, \dots, Z_n$ is defined as $\overline{Z}_n = n^{-1}\sum_{i=1}^n Z_i$. The variance of $n^{1/2}\overline{Z}_n$ is given by

$$\sigma_n^2 \equiv \mathrm{Var}(n^{1/2}\overline{Z}_n) = \gamma(0) + \sum_{i=1}^{n-1} 2\left(1 - \frac{i}{n}\right)\gamma(i),$$

where $\gamma(i) = \mathrm{Cov}(Z_1, Z_{1+i})$ is the ith autocovariance of the underlying time series. Note that $\sigma_n^2 \to 2\pi f(0)$, where $f(\cdot)$ is the spectral density function of the Z_i process. Thus, the problem of estimating σ_n^2 is asymptotically equivalent to estimating the spectral density function at zero. The considerations below apply equally well to the problem of spectral density estimation. A naive estimate of σ_n^2 would be given by

$$\hat{\sigma}^2_{n,\mathrm{naiv}} = \hat{\gamma}(0) + 2\sum_{i=1}^{n-1}\left(1 - \frac{i}{n}\right)\hat{\gamma}(i),$$

where $\hat{\gamma}(i)$ is an estimate of the ith autocovariance defined as

$$\hat{\gamma}(i) = \frac{1}{n}\sum_{j=1}^{n-i} Z_j Z_{j+i}.$$

The estimator $\hat{\sigma}^2_{n,\mathrm{naiv}}$ is also known as 2π times the periodogram at frequency zero in the time-series literature. However, it is well known that this naive estimator is inconsistent, for example, see Section 10.3 of Brockwell and Davis (1991). This is basically due to the fact that estimates $\hat{\gamma}(i)$ for i close to $n-1$ are not reliable, since they are based on very few observations. To ensure consistency, the higher-order autocovariance estimates need to be downweighted in a sufficient way. A very simple downweighting scheme is to assign weight zero to all $\hat{\gamma}(i)$ for $i > m$, where m is a truncation point depending on the sample size n, and weight one to all $\hat{\gamma}(i)$ for all $i \leqslant m$. As before, we implicitly assume the dependence of m on n to be understood rather than using the notation m_n. The truncation estimator of σ_n^2 is then defined as

$$\hat{\sigma}^2_{n,\mathrm{Trunc}} = \hat{\gamma}(0) + 2\sum_{i=1}^{m}\hat{\gamma}(i). \tag{8}$$

It is well known that under suitable conditions this estimator will be consistent and asymptotically normal for general stationary sequences; for example, see Sections 10.4 and 10.5 of Brockwell and Davis (1991). The point of this example is to give a very simple alternative proof using Theorem 2.1 for the special case of i.i.d. random variables $\{Z_i\}$. Of course, this may be considered a toy example. The point is that even in the simple i.i.d. case there are certain restrictions at what rate the truncation point m may increase with the sample size n. Naturally, we assume that the practitioner trying to estimate the variance of the sample mean assumes stationarity of the underlying sequence, not knowing about the actual independence structure. Obviously, in this case we have $\sigma_n^2 = \sigma^2 = \mathrm{Var}(Z_1)$ for all n. Note that this example could be easily extended to the somewhat more interesting case where $\{Z_i\}$ is a sequence of $\tilde{m}$-dependent random variables. This would come at the expense of an additional computational burden only.

To apply the theorem, it will be helpful to rearrange the terms in (8) in the following way:

$$n\hat{\sigma}^2_{n,\mathrm{Trunc}} = X_{n,1} + X_{n,2} + \cdots + X_{n,n}, \tag{9}$$

where $X_{n,i} = Z_i^2 + 2Z_iZ_{i+1} + 2Z_iZ_{i+2} + \cdots + 2Z_iZ_{i+m}$. We have multiplied $\hat{\sigma}^2_{n,\mathrm{Trunc}}$ by n here for notational convenience later on. Note that we define $Z_j \equiv 0$ for $j > n$. This arrangement is useful for our purposes, since the $X_{n,i}$ are easily seen to be m-dependent random variables. Note that if we considered the case of $\tilde{m}$-dependent Z_i, the $X_{n,i}$ would be $(m+\tilde{m})$-dependent random variables.

Proposition 3.1. *Let Z_i be a sequence of i.i.d. random variables with mean zero and variance σ^2. Assume that, for some $\delta > 0$,*

- $E|Z_i|^{4+2\delta} \leqslant \Gamma < \infty$,
- $m^{2+2/\delta}/n \to 0$.

Then, $(n/m)^{1/2}(\hat{\sigma}^2_{n,\mathrm{Trunc}} - \sigma^2) \Rightarrow \mathrm{N}(0, 4\sigma^4)$.

Proof. We need to check the conditions of Theorem 2.1. We claim that they are satisfied if we take $\gamma = 0$. Note that, for this example, $d = n$ in the notation of the theorem. Recall from (9) that our setup is for $n\hat{\sigma}_{n,\mathrm{Trunc}}$, rather than $\hat{\sigma}_{n,\mathrm{Trunc}}$. With the definition $\| Y \|_p = (E|Y|^p)^{1/p}$, for $p > 0$, we first note that

$$\begin{aligned}
\|X_{n,i}\|_{2+\delta} &= \| Z_i^2 + 2Z_iZ_{i+1} + \cdots + 2Z_iZ_{i+m} \|_{2+\delta} \\
&\leqslant \| Z_i^2 \|_{2+\delta} + \| Z_i \|_{2+\delta}\, 2\, \| Z_{i+1} + \cdots + Z_{i+m} \|_{2+\delta} \quad \text{(due to independence)} \\
&\leqslant \Gamma^{1/(2+\delta)} + \Gamma^{1/(2+\delta)} 2C_{2+\delta} m^{1/2}.
\end{aligned}$$

The last inequality follows from Lemma A.1 which states a moment bound for independent random variables and is given in the appendix. Here, $C_{2+\delta}$ is a constant that only depends on $2+\delta$. Therefore,

$$E|X_{n,i}|^{2+\delta} \leqslant 3^{2+\delta} C_{2+\delta}^{2+\delta} \Gamma m^{(2+\delta)/2} \equiv \Delta_n. \tag{10}$$

The important fact is that $\Delta_n = \mathrm{O}(m^{(2+\delta)/2})$.

Next, we have to look at the $B^2_{n,k,a}$, the variance of $X_{n,a} + \cdots + X_{n,a+k-1}$. Since the time series $\{Z_i\}$ is assumed to be i.i.d., we may take, without loss of generality, $a = 1$. Define

$$s_{k,l} = \sum_{i=1}^{k-l} \mathrm{Cov}(X_{n,i},\ X_{n,i+l}).$$

Looking at the variance of a block of size $k \geqslant m$, we then have

$$B^2_{n,k,1} = \mathrm{Var}\left(\sum_{i=1}^{k} X_{n,i} \right) = s_{k,0} + 2\sum_{l=1}^{k-1} s_{k,l}.$$

Since the Z_i are i.i.d. mean zero random variables, it is easy to see that $\mathrm{Cov}(X_{n,i}, X_{n,i+l}) = 0$ for $l \geqslant 1$, which implies that $s_{k,l} = 0$ for $l \geqslant 1$. On the other hand,

$$\mathrm{Cov}(X_{n,i}, X_{n,i}) = \kappa^2 + 4\min(m, n-i)\sigma^4,$$

where $\kappa^2 = \mathrm{Var}(Z_1^2)$ and $\sigma^2 = \mathrm{Var}(Z_1)$. With $\Lambda_1 = \max(\kappa^2, 4\sigma^4)$, we therefore get

$$B^2_{n,k,1} = s_{k,0} \leqslant mk\Lambda_1. \tag{11}$$

Hence, since $\gamma \geqslant 0$, we may choose $K_n = m\Lambda_1$. By the same reasoning, it follows that, for $\Lambda_2 = \min(\kappa^2, 4\sigma^4)$,

$$B_n^2 = B^2_{n,n,1} = s_{n,0} \geqslant \tfrac{1}{2} mn\Lambda_2. \tag{12}$$

To be more precise, since we need this result for the limiting variance of $\hat{\sigma}^2_{n,\mathrm{Trunc}}$,

$$B_n^2 = nm\left(\frac{\kappa^2}{m} + 4\frac{m-1}{m}\sigma^4 \right) + \mathrm{o}(1). \tag{13}$$

For our choice $\gamma = 0$, we can let $L_n = \frac{1}{2} m \Lambda_2$ and so condition (4) is trivially satisfied. To check condition (5), it follows by our choice of L_n and the definition of Δ_n in (10) that

$$\Delta_n / L_n^{(2+\delta)/2} = O(1).$$

Conditions (4) and (5) together, therefore, are met for $\gamma = 0$. Thus, finally, condition (6) is satisfied for $m^{2+2/\delta}/n \to 0$. Recall that $d = n$ for this example. By (13) it then follows that the proper normalizing constant for $n\hat{\sigma}^2_{n,\mathrm{Trunc}}$ is $(nm)^{-1/2}$, so the proper normalizing constant for $\hat{\sigma}^2_{n,\mathrm{Trunc}}$ has to be $(n/m)^{1/2}$. Eq. (13) also tells us that the limiting variance is $4\sigma^4$. □

Example 3.3 (*Moving blocks bootstrap*)**.** The moving blocks bootstrap was introduced by Künsch (1989) and Liu and Singh (1992) as an extension of the classical bootstrap by Efron (1979). It resamples blocks of data at a time, rather than single data points. The extension is needed in the case of dependent observations, where Efron's bootstrap fails. Conditional on observing a sequence $Z_1, \ldots, Z_n$, the moving blocks method generates pseudo sequences $Z_1^*, \ldots, Z_n^*$ by sampling l blocks of size b and concatenating them. Specifically, suppose that $b = b_n$ is a fixed sequence of integers satisfying $b_n \to \infty$. Then, $l = l_n$ is chosen to be the smallest integer satisfying $b_n l_n \geqslant n$. Now, given the observed data $Z_1, \ldots Z_n$, there are $n - b + 1$ blocks of size b, namely

$$(Z_1, \ldots, Z_b), (Z_2, \ldots, Z_{b+1}), \ldots, (Z_{n-b+1}, \ldots, Z_n).$$

Conditional on the observed data, sample l of the blocks with replacement to get a total of bl bootstrap observations, say $Z^*_{n,1}, \ldots, Z^*_{n,bl}$, though we usually just consider $Z^*_{n,1}, \ldots, Z^*_{n,n}$ is the case bl is not exactly an integer.

The point of this example is that, conditional on $Z_1, \ldots, Z_n$, a pseudo sequence $Z^*_{n,1}, \ldots, Z^*_{n,n}$ clearly will be b-dependent. Consider the inference problem of trying to construct a confidence interval for $\mu = E(Z_i)$. Let $\overline{Z}_n = \sum_{i=1}^n Z_i/n$. If the sampling distribution of $n^{1/2}(\overline{Z}_n - \mu)$ were known, a confidence statement could be made. Asymptotically, under moment and mixing assumptions, and stationary of the Z's, the limiting distribution of $n^{1/2}(\overline{Z}_n - \mu)$ will be normal with mean 0 and asymptotic variance

$$\sigma^2 = \mathrm{Var}(Z_1) + 2 \sum_{i=1}^{\infty} \mathrm{Cov}(Z_1, Z_{1+i}).$$

The bootstrap approximation says we can estimate this distribution by the distribution of $n^{1/2}(\overline{Z}_n^* - \overline{Z}_n)$, conditional on $Z_1, \ldots, Z_n$, where $\overline{Z}_n^*$ is the average of the bootstrap sample: $\overline{Z}_n^* = \sum_{i=1}^n Z^*_{n,i}/n$. Hence, the limiting behavior of this distribution can be analyzed by appealing to our theorem, because we are dealing with a normalized average of b_n-dependent variables (and the analysis is all conditional on $Z_1, \ldots, Z_n$ with the randomness coming from the block selection). It is well known that to ensure asymptotic consistency of the moving blocks method, the block size b has to tend to infinity with the sample size n. This is exactly the set-up of our theorem. Indeed, the theorem can be applied with $X_{n,i} = Z^*_{n,i} - E(Z^*_{n,i} | Z_1, \ldots, Z_n)$, $d_n = n$, $m_n = b_n$, $\gamma = 0$, and δ depending on the moments assumed about the Z's. Since the properties of the moving blocks bootstrap have already been analyzed elsewhere, we will not give any alternative proofs. Here, we merely point out the structure of the moving blocks process $\{Z_i^*\}$, thereby validating the utility of our results. It should be clear, however, that asymptotic normality of the bootstrap distribution may ensure because of the independence of the blocks in the resampling scheme, even if $n^{1/2}(\overline{Z}_n - \mu)$ is not asymptotically normal.

4. Sharpness of the theorem

The purpose of this section is to demonstrate the sharpness of our result, meaning that the rate condition (6) cannot be relaxed. Let Y_n follow a $\mathrm{Bin}(n, p_n)$ distribution with $p_n = \lambda/n^\beta$. Here, $\lambda > 0$ and $0 < \beta \leqslant 1$.

Obviously, we assume that n is big enough to ensure $p_n \leqslant 1$. It is well known that $(Y_n - np_n)/[np_n(1-p_n)]^{1/2}$ converges to a standard normal distribution iff $\beta < 1$. The "if" part follows from the Berry–Esseen theorem and the "only if" part from the fact that in case $\beta = 1$ we have a standardized Poisson limit.

To put this scenario in the framework of our theorem, let $Z_{n,1}, \dots, Z_{n,n}$ be i.i.d. according to a Bin$(1, p_n)$ distribution. Let $m = \lfloor n^\beta/\lambda \rfloor + 1$ and for $l = 1, \dots, n$ and $j = 1, \dots, m$ define $X_{n,(l-1)m+j} = Z_{n,j} - p_n$. Hence, we generate the $X_{n,i}$ sequence by including each centered $Z_{n,j}$ term m times, yielding a sequence of length $d = nm$. This complication is necessary to ensure condition (5) below.

Condition (1) is trivially satisfied with $\Delta_n = 1$ for any $\delta > 0$. With $\gamma = 0$, we can choose $K_n = 2$ for inequality (2) and $L_n = 0.5$ for inequality (3). These choices automatically take care of bounds (4) and (5). Keeping in mind that we can pick δ arbitrarily large, the rate condition (6) yields that for our choice $\gamma = 0$ it is sufficient to have $m^2/d \to 0$ or $n^{\beta-1} \to 0$ or $\beta < 1$. By the above discussion this condition is also necessary, since it is easy to see that $B_n^{-1}(X_{n,1} + \cdots + X_{n,d})$ is equal to $(Y_n - np_n)/[np_n(1-p_n)]^{1/2}$.

Remark 4.1. Berk (1973) also gave an example to demonstrate the sharpness of his result. Since his theorem can be considered a special case of ours, the sharpness of our result alternatively follows from Berk's example.

Acknowledgements

We thank an anonymous referee for helpful comments and suggestions that have resulted in an improved presentation of the paper.

Appendix A. Proof of the theorem

Proof of Theorem 2.1. In the proof we will need a result for bounding moments of m-dependent sequences. We will state it as a corollary of the following lemma, which implicitly is given in Chow and Teicher (1978) and deals with independent sequences.

Lemma A.1. *Let $\{Y_i\}$ be an independent sequence of mean zero random variables. Assume $E|Y_i|^q \leqslant \Delta$ for some $q \geqslant 2$ and all i.*

Then,

$$E\left|\sum_{i=1}^{n} Y_i\right|^q \leqslant C_q^q \Delta n^{q/2},$$

where C_q is a positive constant depending only upon q.

Proof. See Theorem 2 and Corollary 2 in Section 10.3 of Chow and Teicher (1978).

Corollary A.1. *Let $\{X_i\}$ be an m-dependent sequence of mean zero random variables. Assume $E|X_i|^q \leqslant \Delta$ for some $q \geqslant 2$ and all i.*

Then, for all $n \geqslant 2m$,

$$E\left|\sum_{i=1}^{n} X_i\right|^q \leqslant C_q^q \Delta (4mn)^{q/2},$$

where C_q is a positive constant depending only upon q.

Proof. Define $t = \lfloor n/m \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part. Now split $X_1 + \cdots + X_n$ into t blocks of size m and a remainder block: $X_1 + \cdots + X_n \equiv A_1 + \cdots + A_t + A_{t+1}$. Due to m-dependence, the odd-numbered blocks are independent of each other, as are the even-numbered blocks. This allows us to apply Lemma A.1:

$$\left\| \sum_{i=1}^{n} X_i \right\|_q \leqslant \left\| \sum_{i \text{ odd}} A_i \right\|_q + \left\| \sum_{i \text{ even}} A_i \right\|_q \quad \text{(by Minkowski)}$$

$$\leqslant 2C_q m(\Delta)^{1/q}(t/2+1)^{1/2} \quad \text{(by Lemma A.1 and Minkowski).}$$

But, this is equivalent to

$$E\left| \sum_{i=1}^{n} X_i \right|^q \leqslant C_q^q 2^q m^q \Delta (t/2+1)^{q/2}$$

$$\leqslant C_q^q 2^q m^q \Delta (t)^{q/2}$$

$$\leqslant C_q^q 2^q \Delta (mn)^{q/2}$$

$$= C_q^q \Delta (4mn)^{q/2}. \qquad \square$$

We are now able to prove the theorem. The main idea of the proof follows Berk (1973), but we need some modifications, since our theorem is more general.

For each n, we choose an integer $p = p_n > 2m$ so that

$$\lim_{n\to\infty} m/p = 0, \qquad \lim_{n\to\infty} p^{1+(1-\gamma)(1+2/\delta)}/d = 0. \tag{14}$$

This can be done, for example, by remembering assumption (6) and choosing p to be the smallest integer greater than $2m$ and greater than $m^{1/2}d^{1/2\xi}$, where ξ is equal to $1 + (1-\gamma)(1+2/\delta)$. Next, define integers $t = t_n$ and $q = q_n$ by $d = pt + q$, $0 \leqslant q < p$. The main idea of the proof is to split the sum $X_{n,1} + \cdots + X_{n,d}$ into alternate blocks of length $p - m$ (the big blocks) and m (the little blocks). This is a common approach to proving central limit theorems for dependent random variables, and is attributed to Markov in Bernstein (1927). Let

$$U_{n,i} = X_{n,(i-1)p+1} + \cdots + X_{n,ip-m}, \quad 1 \leqslant i \leqslant t,$$

$$V_{n,i} = X_{n,ip-m+1} + \cdots + X_{n,ip}, \quad 1 \leqslant i \leqslant t,$$

$$U_{n,t+1} = X_{n,tp+1} + \ldots + X_{n,d}.$$

By definition, $X_{n,1} + \cdots + X_{n,d} = \sum_{i=1}^{t+1} U_{n,i} + \sum_{i=1}^{t} V_{n,i}$. Since the $X_{n,i}$ are m-dependent and $p > 2m$, $\{U_{n,i}\}$ and $\{V_{n,i}\}$ are each independent sequences. It is easily seen that the difference between $B_n^{-1}(X_{n,1} + \cdots + X_{n,d})$ and $B_n^{-1}\sum_{i=1}^{t+1} U_{n,i}$ has variance approaching zero. Indeed,

$$\operatorname{Var}\left(B_n^{-1} \sum_{i=1}^{t} V_{n,i} \right) = B_n^{-2} \sum_{i=1}^{t} \operatorname{Var}(V_{n,i})$$

$$\leqslant B_n^{-2} t \left[\sup_i \operatorname{Var}(V_{n,i}) \right]$$

$$\leqslant B_n^{-2} t K_n m^{1+\gamma} \quad \text{(by assumption (2))}$$

$$\leqslant B_n^{-2} (d/p) K_n m^{1+\gamma}$$

$$\leqslant \frac{K_n}{L_n} \frac{m}{p} \to 0 \quad \text{(by assumptions (3) and (4)).}$$

Hence, provided they exist, the asymptotic distributions of the two quantities $B_n^{-1}\sum_{i=1}^d X_{n,i}$ and $B_n^{-1}\sum_{i=1}^{t+1} U_{n,i}$ are the same, and the goal now is to show that $B_n^{-1}\sum_{i=1}^{t+1} U_{n,i} \Rightarrow \mathrm{N}(0,1)$.

In order to apply assumption (3) again, we will first establish that $B_n^{-2}\mathrm{Var}(\sum_{i=1}^{t+1} U_{n,i})$ tends to one, or, equivalently, that $B_n^{-2}\mathrm{Cov}(\sum_{i=1}^{t+1} U_{n,i}, \sum_{i=1}^{t} V_{n,i})$ tends to zero. Note first that $\mathrm{Cov}(U_{n,i}, V_{n,j}) = 0$ unless $j = i$ or $i-1$. Furthermore,

$$\begin{aligned}|\mathrm{Cov}(U_{n,i}, V_{n,j})| &= |E(U_{n,i}V_{n,j})| \\ &\leqslant [\mathrm{Var}(U_{n,i})\mathrm{Var}(V_{n,i})]^{1/2} \quad \text{(by Cauchy–Schwarz)} \\ &\leqslant K_n(mp)^{(1+\gamma)/2} \quad \text{(by assumption (2)).}\end{aligned}$$

Combining these two facts, we obtain

$$\left|\mathrm{Cov}\left(\sum_{i=1}^{t+1} U_{n,i}, \sum_{i=1}^{t} V_{n,i}\right)\right| \leqslant 2tK_n(mp)^{(1+\gamma)/2}$$

and finally,

$$\begin{aligned}B_n^{-2}\left|\mathrm{Cov}\left(\sum_{i=1}^{t+1} U_{n,i}, \sum_{i=1}^{t} V_{n,i}\right)\right| &\leqslant 2\frac{K_n}{L_n}\frac{t}{dm^\gamma}(mp)^{(1+\gamma)/2} \quad \text{(by assumption (3))} \\ &\leqslant 2\frac{K_n}{L_n}\frac{1}{pm^\gamma}(mp)^{(1+\gamma)/2} \\ &= 2\frac{K_n}{L_n}\left(\frac{m}{p}\right)^{(1-\gamma)/2} \to 0 \quad \text{(by assumption (4) and since } \gamma < 1).\end{aligned}$$

By Lyapounov's theorem, it will now suffice to verify that $\sum_{i=1}^{t+1} E|U_{n,i}|^{2+\delta}/B_n^{2+\delta}$ tends to zero. By Corollary A.1,

$$E|U_{n,i}|^{2+\delta} \leqslant C_{2+\delta}^{2+\delta}\Delta_n(4pm)^{(2+\delta)/2}, \quad 1 \leqslant i \leqslant t+1,$$

and therefore

$$\sum_{i=1}^{t+1} E|U_{n,i}|^{2+\delta}/B_n^{2+\delta} \leqslant \text{Const.}\,\Delta_n(d/p+1)(pm)^{(2+\delta)/2}/B_n^{2+\delta}.$$

By assumption (3), finally,

$$\begin{aligned}\Delta_n(d/p)(pm)^{(2+\delta)/2}/B_n^{2+\delta} &\leqslant \Delta_n L_n^{-(2+\delta)/2}\frac{d}{p}\left(\frac{pm}{dm^\gamma}\right)^{(2+\delta)/2} \\ &\leqslant \Delta_n L_n^{-(2+\delta)/2}\left(\frac{p}{d}\right)^{\delta/2} m^{(1-\gamma)(2+\delta)/2} \\ &= \Delta_n L_n^{-(2+\delta)/2} p^{\delta/2+(1-\gamma)(2+\delta)/2} d^{-\delta/2}\left(\frac{m}{p}\right)^{(1-\gamma)(2+\delta)/2} \\ &= \mathrm{O}(1)AB \quad \text{(by assumption (5)),}\end{aligned}$$

where $A \equiv p^{\delta/2+(1-\gamma)(2+\delta)/2}d^{-\delta/2}$ and $B \equiv (m/p)^{(1-\gamma)(2+\delta)/2}$. The second condition on p in (14) implies that A tends to zero. The first condition on p in (14), together with the fact that $\gamma < 1$, imply that B tends to zero as well. □

References

Beran, J., 1994. Statistics for Long-Memory Processes. Chapman & Hall, New York.

Bergstrom, H., 1970. A comparison method for distribution functions of sums of independent and dependent random variables. Theor. Probab. Appl. 15, 430–457.

Berk, K.N., 1973. A central limit theorem for m-dependent random variables with unbounded m. Ann. Probab. 1, 352–354.

Bernstein, S., 1927. Sur l'extension du théorème du calcul des probabilités aux sommes de quantités dependantes. Math. Ann. 97, 1–59.

Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Models, 2nd Edition. Springer, New York.

Chow, Y.S., Teicher, H., 1978. Probability Theory. Springer, New York.

Diananda, P.H., 1955. The central limit theorem for m-dependent variables. Proc. Combin. Philos. Soc. 51, 92–95.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. Ann. Statist. 7, 1–26.

Hoeffding, W., Robbins, H., 1948. The central limit theorem for dependent random variables. Duke Math. J. 15, 773–780.

Krieger, H., 1984. A new look at Bergstrom's theorem on convergence in distribution for sums of dependent random variables. Israel J. Math. 47, 32–64.

Künsch, H.R., 1989. The jackknife and the bootstrap for general stationary observations. Ann. Statist. 17, 1217–1241.

Liu, R.Y., Singh, K., 1992. Moving blocks jackknife and bootstrap capture weak dependence. In: LePage, Billard (Eds.), Exploring the Limits of Bootstrap. Wiley, New York.

Orey, S.A., 1958. Central limit theorems for m-dependent random variables. Duke Math. J. 25, 543–546.

Peligrad, M., 1996. On the asymptotic normality of sequences of weak dependent random variables. J. Theoret. Probab. 9, 703–715.

Rio, E., 1995. About the Lindeberg method for strongly mixing sequences. European Ser. Appl. Ind. Math. Probab. Statist. 1, 35–61.

Samur, J., 1984. Convergence of sums of mixing triangular arrays of random vectors with stationary rows. Ann. Probab. 12, 390–426.