J. Appl. Prob. **52,** 1146–1155 (2015)
Printed in England

ON DEGENERATE SUMS OF m-DEPENDENT VARIABLES

SVANTE JANSON,* *Uppsala University*

Abstract

It is well known that the central limit theorem holds for partial sums of a stationary sequence (X_i) of m-dependent random variables with finite variance; however, the limit may be degenerate with variance 0 even if $\mathrm{var}(X_i) \neq 0$. We show that this happens only in the case when $X_i - \mathbb{E}X_i = Y_i - Y_{i-1}$ for an $(m-1)$-dependent stationary sequence (Y_i) with finite variance (a result implicit in earlier results), and give a version for block factors. This yields a simple criterion that is a sufficient condition for the limit not to be degenerate. Two applications to subtree counts in random trees are given.

Keywords: m-dependent; stationary sequence; block factor; random tree

2010 Mathematics Subject Classification: Primary 60G10
Secondary 60F05; 60C05

1. Introduction and results

Consider a strictly stationary sequence $(X_k)_{-\infty}^{\infty}$ of m-dependent random variables for some $m \geq 1$, and suppose that the variables have finite variance, i.e. $\mathbb{E}X_k^2 < \infty$. (Recall that m-dependence means that $(X_k)_{k\leq 0}$ is independent of $(X_k)_{k\geq m+1}$.)

Let $S_n := \sum_{i=1}^n X_i$. A simple standard calculation using stationarity and m-dependence yields, for $n \geq m$,

$$\begin{aligned}\mathrm{var}(S_n) &= \sum_{i,j=1}^{n} \mathrm{cov}(X_i, X_j)\\ &= n\,\mathrm{var}(X_0) + 2\sum_{k=1}^{m}(n-k)\,\mathrm{cov}(X_0, X_k)\\ &= n\sigma^2 - 2\sum_{k=1}^{m} k\,\mathrm{cov}(X_0, X_k), \qquad (1)\end{aligned}$$

where

$$\sigma^2 := \mathrm{var}(X_0) + 2\sum_{k=1}^{m}\mathrm{cov}(X_0, X_k) = \mathrm{cov}\Big(X_0, \sum_{k=-m}^{m} X_k\Big). \qquad (2)$$

In particular,

$$\mathrm{var}(S_n) = n\sigma^2 + O(1). \qquad (3)$$

It is obvious from (3) that $\sigma^2 \geq 0$. If we have the strict inequality $\sigma^2 > 0$, then $\mathrm{var}(S_n)$ grows linearly; moreover, the classic central limit theorem for m-dependent variables by Hoeffding

Received 4 September 2014; revision received 17 October 2014.
* Postal address: Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden. Email address: svante@math.uu.se

and Robbins (1948) and Diananda (1995) (see also Bradley (2007a, Theorem 10.8)) shows that

$$\frac{S_n - \mathbb{E}S_n}{\sqrt{n}} = \frac{S_n - n\mathbb{E}X_0}{\sqrt{n}} \xrightarrow{\text{D}} N(0, \sigma^2), \tag{4}$$

where '$\xrightarrow{\text{D}}$' denotes convergence in distribution. In the exceptional case $\sigma^2 = 0$, however, $\operatorname{var}(S_n)$ is bounded; more precisely, (1) shows that $\operatorname{var}(S_n)$ is constant for all $n \geq m$. In this case, (4) still holds, with the limit 0, but is a triviality. (See Corollary 1 below for the limit of S_n without normalization in this case.)

The purpose of this paper is to study this exceptional case further, and show that it really is exceptional and only occurs in very special cases.

A well-known trivial example with $\sigma^2 = 0$ is obtained by taking an independent and identically distributed (i.i.d.) sequence $(Y_k)_{-\infty}^{\infty}$ (with $\mathbb{E}Y_k^2 < \infty$) and defining $X_k := Y_k - Y_{k-1}$; see, for example, Ibragimov and Linnik (1971, Section 18.1). This sequence is obviously 1-dependent and $S_n = Y_n - Y_0$ with $\operatorname{var}(S_n) = 2\operatorname{var}(Y_0)$, $n \geq 1$, so $\operatorname{var}(S_n)$ is constant and $\sigma^2 = 0$. (This can also be seen from (2), using $\operatorname{var}(X_0) = 2\operatorname{var}(Y_0)$ and $\operatorname{cov}(X_0, X_1) = -\operatorname{var}(Y_0)$.)

In fact, the following theorem (which is implicit in Bradley (2007a, Theorem 8.6) but deserves to be made more explicit) shows that this trivial example is the only example when $m = 1$ (apart from adding a constant), and that a similar result holds for $m > 1$.

Theorem 1. *Let $(X_k)_{-\infty}^{\infty}$ be a strictly stationary sequence of m-dependent variables with finite variance and let $\sigma^2 := \lim_{n\to\infty} n^{-1} \operatorname{var}(S_n)$, which is also given by (2). If $\sigma^2 = 0$, then there exists a strictly stationary sequence $(Y_k)_{-\infty}^{\infty}$ of $(m-1)$-dependent variables with finite variance, and a constant μ, such that*

$$X_k = Y_k - Y_{k-1} + \mu \quad \textit{almost surely (a.s.)} \tag{5}$$

The random variables Y_k are a.s. unique up to an additive constant.

Conversely, for any such sequence $(Y_k)_{-\infty}^{\infty}$ and any μ, (5) yields a strictly stationary m-dependent sequence $(X_k)_{-\infty}^{\infty}$ with $\sigma^2 = 0$.

Taking expectations in (5) yields $\mu = \mathbb{E}X_k$.

Remark 1. Theorem 1 holds also for weakly stationary sequences $(X_k)_{-\infty}^{\infty}$, with $(Y_k)_{-\infty}^{\infty}$ weakly stationary. (Recall that 'weakly stationary' just means that the means and covariances are translation invariant.)

The existence of a (weakly) stationary sequence $(Y_k)_{-\infty}^{\infty}$ such that (5) holds was shown by Leonov (1961) under much weaker conditions than m-dependence: $(X_k)_{-\infty}^{\infty}$ (weakly) stationary, $\operatorname{cov}(X_0, X_n) \to 0$ as $n \to \infty$, and $\liminf_{n\to\infty} \operatorname{var}(S_n) < \infty$. See also Robinson (1960), Ibragimov and Linnik (1971, Theorem 18.2.2), and Bradley (2007a, Theorem 8.6). The $(m-1)$-dependence of $(Y_k)_{-\infty}^{\infty}$ when $(X_k)_{-\infty}^{\infty}$ is m-dependent follows from Bradley (2007a, Theorem 8.6(B)(e)), but does not seem to have been stated explicitly earlier.

For completeness, we give a direct proof of Theorem 1 in Section 2. (The same proof applies to the weakly stationary version, see Remark 1.)

Remark 2. More generally, a theorem by Schmidt (1977, Lemma 11.7), in the version given by Bradley (2007b, Theorem 19.9), implies that even without the assumption of finite variance, if $(X_k)_{-\infty}^{\infty}$ is a strictly stationary and m-dependent sequence such that the family of partial sums S_n are tight, then the conclusion (5) (with $\mu = 0$) holds for some strictly stationary $(m-1)$-dependent sequence $(Y_k)_{-\infty}^{\infty}$.

Note that (5) implies that

$$S_n - \mathbb{E}S_n = S_n - n\mu = Y_n - Y_0 \quad \text{a.s.}, \tag{6}$$

where $Y_n \stackrel{\mathrm{D}}{=} Y_0$ and Y_n and Y_0 are independent when $n \geq m$. We denote equality in distribution by '$\stackrel{\mathrm{D}}{=}$'. An immediate consequence of Theorem 1 is that in the exceptional case $\sigma^2 = 0$, the centered partial sums $S_n - \mathbb{E}S_n$ converge in distribution without normalization. Of course, the limit is in general not normal, so there is no central limit theorem in this case. (For example, X_n may be integer-valued, and then so is S_n.) We state this in detail; see Section 2 for proofs of this and other results.

Corollary 1. *Let $(X_k)_{-\infty}^{\infty}$ be a strictly stationary sequence of m-dependent variables with finite variance, and let σ^2 be given by (2). If $\sigma^2 = 0$ then $S_n - \mathbb{E}S_n$ has the same distribution for all $n \geq m$; more precisely, if Y_k is as in (5) and Y_0' is an independent copy of Y_0, then $S_n - \mathbb{E}S_n \stackrel{\mathrm{D}}{=} Y_0 - Y_0'$.*

Hence, assuming that $\operatorname{var}(X_0) > 0$, $(S_n - \mathbb{E}S_n)/\operatorname{var}(S_n)^{1/2}$ converges in distribution as $n \to \infty$ also in the $\sigma^2 = 0$ case, but then the limit is normal only if each Y_k is normal.

Remark 3. If $\sigma^2 = 0$ and $m = 1$, then (5) holds with independent Y_k. Hence, by a theorem by Cramér, see, for example, Feller (1971, Theorem XV.8.1), each Y_k is normal if and only if X_k is normal (and then $\{X_k, Y_k : k \in \mathbb{Z}\}$ are jointly normal). For $m > 1$ this does not hold. For example, if $U_k \sim U(0,1)$ and $\xi_k \sim N(0,1)$, $k \in \mathbb{Z}$, all independent, then $Y_k := \operatorname{sign}(U_k - U_{k+1})|\xi_k|$ is a sequence of 1-dependent normal variables that are not jointly normal, and the 2-dependent random variables $X_k := Y_k - Y_{k-1}$ are not normal although by (6), $S_n \sim N(0,2)$ is for $n \geq 2$. (A simple calculation yields $\mathbb{E}X_k^2 = 2 + 4/3\pi$ and $\mathbb{E}X_k^4 = 12 + 32/3\pi \neq 3(\mathbb{E}X_k^2)^2$.)

Stationary m-dependent sequences usually appear as *block factors*. We say that (X_k) is an ℓ-block factor if there is an i.i.d. sequence $(\xi_k)_{-\infty}^{\infty}$ and a (measurable) function $f : \mathbb{R}^{\ell} \to \mathbb{R}$ such that $X_k = f(\xi_k, \ldots, \xi_{k+\ell-1})$. Note that every such sequence (X_k) is strictly stationary and $(\ell - 1)$-dependent. (However, there are m-dependent sequences that are not block factors; see Aaronson *et al.* (1989) and Burton *et al.* (1993).)

For block factors, Theorem 1 takes the following form.

Theorem 2. *Let $X_k = f(\xi_k, \ldots, \xi_{k+\ell-1})$ be an ℓ-block factor for some $\ell \geq 1$, where $(\xi_k)_{-\infty}^{\infty}$ is an i.i.d. sequence. Suppose that X_k has finite variance and let $\sigma^2 := \lim_{n\to\infty} n^{-1} \operatorname{var}(S_n)$. If $\sigma^2 = 0$, then there exists a function $g : \mathbb{R}^{\ell-1} \to \mathbb{R}$ and a constant μ such that the $(\ell-1)$-block factor $Y_k := g(\xi_{k+1}, \ldots, \xi_{k+\ell-1})$ has finite variance and*

$$X_k = Y_k - Y_{k-1} + \mu \quad \text{a.s.}$$

The function g is a.s. unique up to an additive constant.

The converse is obvious in this theorem too.

Corollary 2. *Let $X_k = f(\xi_k, \ldots, \xi_{k+\ell-1})$ be an ℓ-block factor with finite variance, where $(\xi_k)_{-\infty}^{\infty}$ is an i.i.d. sequence. If $\sigma^2 = 0$, then there exists a function $g : \mathbb{R}^{\ell-1} \to \mathbb{R}$ such that, for every $n \geq 1$,*

$$S_n - \mathbb{E}S_n = g(\xi_{n+1}, \ldots, \xi_{n+\ell-1}) - g(\xi_1, \ldots, \xi_{\ell-1}) \quad \text{a.s.} \tag{7}$$

https://doi.org/10.1239/jap/1450802758 Published online by Cambridge University Press

Remark 4. The contrapositive form of Corollary 2 yields a simple criterion. If we can find, for some $n \geq \ell$, a set of values of $\xi_1, \ldots, \xi_{\ell-1}$ and $\xi_{n+1}, \ldots, \xi_{n+\ell-1}$ of positive probability such that S_n is not an a.s. constant function of $\xi_\ell, \ldots, \xi_n$, then (7) cannot hold and, thus, $\sigma^2 > 0$.

Corollary 2 and its reformulation in Remark 4 are useful in applications, to show that the asymptotic variance $\sigma^2 > 0$. We give two such applications in Section 3, taken from Holmgren and Janson (2015) and Janson (2014); these applications were the motivation for this study.

Remark 5. The central limit theorem for m-dependent variables has been generalized to much more general mixing sequences under various conditions, see, for example, Ibragimov and Linnik (1971) and Bradley (2007a). For example, if $(X_k)_{-\infty}^{\infty}$ is strictly stationary with finite variances and ρ-mixing, then either

(i) $\operatorname{var}(S_n) = nh(n)$ for some slowly varying function $h(n)$, or

(ii) $\operatorname{var}(S_n)$ is bounded, and converges to some finite limit.

Moreover, in Remark 5(i), a central limit theorem holds under further conditions; see Ibragimov (1975) and Bradley (2007a, Theorems 11.2 and 11.4) (but not in general, Bradley (1980), Bradley (2007c, Chapter 34)).

In Remark 5(ii), there is, by the result by Leonov (1961) mentioned above, a representation as in (5); however, we do not know any useful consequences similar to Corollary 2 and Remark 4 in this generality and we leave it as an open problem to find generalizations of the results above that can be used to show $\sigma^2 > 0$. A typical example of Remark 5(ii) is $X_k = \xi_k - \sum_{j=1}^{\infty} 2^{-j}\xi_{k+j}$ with $(\xi_k)_{-\infty}^{\infty}$ i.i.d. $N(0, 1)$, where we have the representation (5) with $Y_k = -\sum_{j=0}^{\infty} 2^{-j}\xi_{k+1+j}$.

2. Proofs

Proof of Theorem 1. As stated in the introduction, Theorem 1 follows from Bradley (2007a, Theorem 8.6), but we also give a direct proof for completeness. (The proof is similar, but simpler in this special case.)

It is obvious that if $(Y_k)_{-\infty}^{\infty}$ is strictly stationary and $(m-1)$-dependent, then $(X_k)_{-\infty}^{\infty}$ defined by (5) is strictly stationary and m-dependent. Furthermore, (5) implies (6) and, thus, $\operatorname{var}(S_n) = \operatorname{var}(Y_n) + \operatorname{var}(Y_0) = 2\operatorname{var}(Y_0)$ when $n \geq m$; hence, $\sigma^2 = 0$ by (3).

To prove the converse we may assume that $\mathbb{E}X_k = 0$. Define $S_{k,n} := \sum_{i=k}^{n} X_i$ for $-\infty < k \leq n < \infty$. The assumption that $\sigma^2 = 0$ implies, by (3) and stationarity, that $\mathbb{E}S_{k,n}^2 = \operatorname{var}(S_{k,n})$ is bounded. (In fact, by (1) it is constant for all (k, n) with $n - k \geq m - 1$.)

We claim first that for every k, the sequence $S_{k,n}$ converges weakly in L^2 as $n \to \infty$, and there exists a random variable $Z_k \in L^2$ such that

$$S_{k,n} \xrightarrow{\mathrm{w}} Z_k \quad \text{as } n \to \infty, \tag{8}$$

where '$\xrightarrow{\mathrm{w}}$' denotes weak convergence. In fact, since the sequence $(S_{k,n})_{n \geq k}$ is bounded in L^2 and the unit ball of L^2 is weakly compact, it suffices to show that $\mathbb{E}(WS_{k,n})$ converges as $n \to \infty$ for every fixed $W \in L^2$; moreover, it suffices to verify this for a dense set of W. We consider two special cases:

(i) if $\mathbb{E}(WX_j) = 0$ for all j, then $\mathbb{E}(WS_{k,n}) = 0$ for all n, and the convergence is trivial;

(ii) if $W = X_j$ for some j, then $\mathbb{E}(WS_{k,n})$ is constant for all $n \geq \max(j + m, k)$, by m-dependence, and again the convergence is trivial.

Hence, $\mathbb{E}(WS_{k,n})$ converges also when W is a linear combination of variables of the type (i) or (ii). But the set of such linear combinations is dense in L^2, which proves (8).

Similarly (or by reflecting the indices and replacing X_k by X_{-k}), for every $k \in \mathbb{Z}$ there exists a random variable $Y_k \in L^2$ such that

$$S_{-n,k} \xrightarrow{\mathrm{w}} Y_k \quad \text{as } n \to \infty. \tag{9}$$

Since $S_{-n,k} - S_{-n,k-1} = X_k$ for $-n < k$, it follows that $Y_k - Y_{k-1} = X_k$, so (5) holds (with $\mu = \mathbb{E}(X_0) = 0$). Furthermore, $(Y_k)_{-\infty}^{\infty}$ is stationary by (9) and the stationarity of $(X_k)_{-\infty}^{\infty}$. It remains to show that $(Y_k)_{-\infty}^{\infty}$ is $(m-1)$-dependent.

We note first that for any k, as $n \to \infty$, by (9) and (8),

$$S_{-n,k} + S_{k+1,n} \xrightarrow{\mathrm{w}} Y_k + Z_{k+1}. \tag{10}$$

On the other hand, $S_{-n,k} + S_{k+1,n} = S_{-n,n}$ (when $n > |k|$) and, thus, for every $j \in \mathbb{Z}$ and every $n > \max(|k|, m + |j|)$, using m-dependence and (2),

$$\begin{aligned}
\mathbb{E}(X_j(S_{-n,k} + S_{k+1,n})) &= \mathbb{E}(X_j S_{-n,n}) \\
&= \operatorname{cov}(X_j, S_{-n,n}) \\
&= \sum_{i=-n}^{n} \operatorname{cov}(X_j, X_i) \\
&= \sum_{i=j-m}^{j+m} \operatorname{cov}(X_j, X_i) \\
&= \sigma^2 \\
&= 0.
\end{aligned} \tag{11}$$

Combining (10) and (11), we see that $\mathbb{E}(X_j(Y_k + Z_{k+1})) = 0$ for every j. Summing over j, we find that $\mathbb{E}(S_{\ell,n}(Y_k + Z_{k+1})) = 0$ for all ℓ and n, and, thus, by (10) again, $\mathbb{E}(Y_k + Z_{k+1})^2 = 0$. Hence, $Y_k + Z_{k+1} = 0$ a.s., i.e.

$$Y_k = -Z_{k+1} \quad \text{a.s.} \tag{12}$$

For $-\infty \le k \le n \le \infty$, let $\mathcal{F}_{k,n}$ denote the σ-field generated by $\{X_i\}_{i=k}^{n}$. Write $W \in \mathcal{F}_{k,n}$ if the random variable W is $\mathcal{F}_{k,n}$-measurable. Then $S_{-n,k} \in \mathcal{F}_{-n,k} \subseteq \mathcal{F}_{-\infty,k}$, and, thus, from (9), we have

$$Y_k \in \mathcal{F}_{-\infty,k}. \tag{13}$$

Similarly, $Z_k \in \mathcal{F}_{k,\infty}$. By (12), this also yields

$$Y_k \in \mathcal{F}_{k+1,\infty}. \tag{14}$$

Since $(X_k)_{-\infty}^{\infty}$ is m-dependent, the σ-fields $\mathcal{F}_{-\infty,k}$ and $\mathcal{F}_{k+m+1,\infty}$ are independent. Hence, from (13) and (14) it follows that $\{Y_j : j \le k\}$ and $\{Y_j : j \ge k+m\}$ are independent for every k, which is the desired $(m-1)$-dependence.

Finally, we consider the uniqueness of Y_k. It is obvious that we may replace Y_k by $Y_k + C$ for any constant C. For the converse, we may assume that $\mathbb{E}X_k = 0$ so $\mu = 0$. If (5) holds then

$$S_{k,n} = Y_n - Y_{k-1}, \tag{15}$$

and it follows, by (3) applied to $(Y_n)_{-\infty}^{\infty}$, that

$$\operatorname{var}\biggl(\frac{1}{n}\sum_{j=k+1}^{k+n} S_{k,j} + Y_{k-1}\biggr) = O(n^{-1}). \tag{16}$$

Thus, $Y_{k-1} - \mathbb{E}Y_{k-1}$ is the limit in L^2 of the means $-(1/n)\sum_{j=k+1}^{k+n} S_{k,j}$ and is thus a.s. determined by $(X_j)_{-\infty}^{\infty}$.

Remark 6. We use weak convergence in L^2 in (8) and (9), following Leonov (1961) who used weak convergence of a subsequence in a much more general situation. (It is easy to modify the proof by Leonov (1961) to show weak convergence of the full sequence under the conditions there too. Above we have used a simpler version for the m-dependent case.) Strong (norm) convergence does not hold: from (15), we have $\|S_{k,n} - Z_k\| = \|S_{k,n} + Y_{k-1}\| = \|Y_n\|$ which is constant and does not tend to 0 (except in the trivial case $Y_n = 0$ when $X_k = 0$ a.s.). However, assuming that $\mathbb{E}X_k = 0$ and choosing Y_k with $\mathbb{E}Y_k = 0$, (16) shows that the Cesàro means $T_{k,n} := (n+1)^{-1}\sum_{j=k}^{k+n} S_{k,j}$ converge to $Z_k = -Y_{k-1}$ in L^2, i.e. $\|T_{k,n} - Z_k\| \to 0$ and, similarly, $(n+1)^{-1}\sum_{j=k-n}^{k} S_{k,j} \to Y_k$ in L^2; see the proof of Bradley (2007a, Theorem 8.6). (This can be used to give an alternative proof of Theorem 1, using strong Cesàro convergence instead of weak convergence and completing the proof as above.) Furthermore, the strong law of large numbers for stationary m-dependent sequences implies that $T_{k,n} \to Z_k$ a.s., while from (15), it follows that $S_{k,n}$ does not converge a.s. (except when $Z_i = 0$).

Proof of Corollary 1. By Theorem 1, (5) holds and, thus, (6) holds, which shows that $S_n - \mathbb{E}S_n \overset{\mathrm{D}}{=} Y_0 - Y_0'$ when $n \ge m$. In particular, for $n \ge m$, $\operatorname{var}(S_n) = 2\operatorname{var}(Y_0)$ and, hence, $\operatorname{var}(S_n) = 0$ only if Y_0 is degenerate (a.s. constant), and then each X_k is degenerate. Finally, by the theorem by Cramér mentioned in Remark 3, $Y_0 - Y_0'$ is normal if and only if Y_0 has a normal distribution.

Proof of Theorem 2. Let Y_k and Z_k be as in the proof of Theorem 1. For $-\infty \le k \le n \le \infty$, let $\bar{\mathcal{F}}_{k,n}$ denote the σ-field generated by $\{\xi_i\}_{i=k}^{n}$ and all sets of probability 0. (The latter technicality is because Y_k and Z_k are defined only a.s.) Then $X_k \in \bar{\mathcal{F}}_{k,k+\ell-1}$ so $S_{k,n} \in \bar{\mathcal{F}}_{k,n+\ell-1}$ and, thus, $Y_k \in \bar{\mathcal{F}}_{-\infty,k+\ell-1}$ and $Z_k \in \bar{\mathcal{F}}_{k,\infty}$. Since $Y_k = -Z_{k+1}$ by (12); thus,

$$Y_k \in \bar{\mathcal{F}}_{-\infty,k+\ell-1} \cap \bar{\mathcal{F}}_{k+1,\infty} = \bar{\mathcal{F}}_{k+1,k+\ell-1},$$

where the latter equality follows (e.g. by considering conditional expectations) because the variables ξ_i are independent.

Hence, $Y_k = g(\xi_{k+1}, \dots, \xi_{k+\ell-1})$ for some function g (independent of k because of stationarity). The result now follows from Theorem 1.

Proof of Corollary 2. This is an immediate consequence of Theorem 2 and (6).

3. Applications

We sketch two applications of the results above; more details and background are given in Holmgren and Janson (2015) and Janson (2014). In both applications we consider a random rooted tree $\mathcal{T}_n$ with n nodes (with different distributions in the two cases) and let, for a fixed rooted tree T, $n_T(\mathcal{T}_n)$ be the number of nodes $v \in \mathcal{T}_n$ such that the fringe subtree consisting of v and all its descendants is isomorphic to T. (We consider only trees T in the family $\mathfrak{T}^*$ of trees that can appear as fringe subtrees in $\mathcal{T}_n$ for some n; otherwise, $n_T(\mathcal{T}_n)$ is identically 0 for all n.) In the cases studied here, these numbers are asymptotically normal for fixed T as $n \to \infty$:

$$\frac{n_T(\mathcal{T}_n) - n\mu_T}{\sqrt{n}} \xrightarrow{\text{D}} \zeta_T, \tag{17}$$

where $\zeta_T \sim N(0, \sigma_T^2)$ for some $\mu_T > 0$ and $\sigma_T^2 \geq 0$. Moreover, this holds jointly for all T with the limit variables ζ_T jointly normal, with convergence of variances and covariances. We use the results above to show that the limit distribution is not degenerate: $\sigma_T^2 > 0$ for each $T \in \mathfrak{T}^*$ and, moreover, the covariance matrix of $\zeta_{T_1}, \dots, \zeta_{T_N}$ is positive definite, for any finite number of trees $T_1, \dots, T_N \in \mathfrak{T}^*$. Equivalently, if

$$F(\mathcal{T}_n) = \sum_{i=1}^{N} a_j n_{T_j}(\mathcal{T}_n) \tag{18}$$

for some distinct trees $T_1, \dots, T_N \in \mathfrak{T}^*$ and real numbers $a_1, \dots, a_N$, not all 0, then

$$\lim_{n\to\infty} \frac{\operatorname{var} F(\mathcal{T}_n)}{n} = \operatorname{var}\biggl(\sum_{j=1}^{N} a_j \zeta_{T_j}\biggr) > 0. \tag{19}$$

Example 1. (*Binary search trees, Holmgren and Janson (2015).*) A *binary search tree* is a binary tree with a *key* stored at each node. It is constructed from a sequence of (distinct) keys by putting the first key, say x_1, in the root and sending all subsequent keys less than x_1 to the left subtree and the keys greater than x_1 to the right subtree, constructing the subtrees recursively in the same way.

We may assume that the keys are $1, \dots, n$; then, a binary search tree is a binary tree with the nodes labelled $1, \dots, n$ (where n is the size of the tree). Let $\mathcal{T}_n$ be a uniformly random binary search tree with n nodes; this can be constructed by taking the keys $1, \dots, n$ in (uniformly) random order.

We use a modification of this construction by Devroye (1991), (2002). Let $U_1, \dots, U_n$ be i.i.d. random variables with $U_i \sim U(0, 1)$, order the indices $1, \dots, n$ so that the variables U_i are in increasing order and construct the binary search tree $\mathcal{T}_n$ as above using this sequence. (Thus, for example, the root is labelled by the index i such that U_i is the smallest of $U_1, \dots, U_n$.) It is not difficult to see that then the fringe subtrees of $\mathcal{T}_n$ are the trees defined in the same way by the subsequences $U_i, \dots, U_j$ (with $1 \leq i \leq j \leq n$) such that U_{i-1} and U_{j+1} both are smaller than all of $U_i, \dots, U_j$; here, we define $U_0 = U_{n+1} = 0$.

Hence, if $T \in \mathfrak{T}^*$, where now $\mathfrak{T}^*$ is the family of all binary trees, and T has $|T| = k$ nodes, then

$$n_T(\mathcal{T}_n) = \sum_{i=0}^{n-k} f_T(U_i, \dots, U_{i+k+1}) \tag{20}$$

for some indicator $f_T(x_1, \dots, x_{k+2})$ on $[0, 1]^{k+2}$ (depending only on the order relations between $x_1, \dots, x_{k+2}$). For convenience, we ignore the boundary terms in (20), which are asymptotically negligible; we let $(U_i)_{-\infty}^{\infty}$ be i.i.d. with $U_i \sim U(0, 1)$ and then

$$n_T(\mathcal{T}_n) = \sum_{i=1}^{n-k-1} f_T(U_i, \dots, U_{i+k+1}) + O(1), \tag{21}$$

where the sum is a sum of m-dependent variables of the type studied in this paper. Given a function F as in (18), we let $\ell := \max_j |T_j| + 2$ and define

$$f(x_1, \dots, x_\ell) := \sum_j a_j f_{T_j}(x_1, \dots, x_{|T_j|+2}).$$

Then (21) implies that

$$F(\mathcal{T}_n) = \sum_{i=1}^{n-\ell} f(U_i, \dots, U_{i+\ell-1}) + O(1) = S_{n-\ell} + O(1), \tag{22}$$

where $S_n = \sum_{i=1}^n X_i$ with $X_i = f(U_i, \dots, U_{i+\ell-1})$ an ℓ-block factor as in Theorem 2. Hence, the central limit theorem for m-dependent variables (Hoeffding and Robbins (1948) and Diananda (1955)) yields asymptotic normality of $F(\mathcal{T}_n)$, i.e. (17) with joint convergence for several $T \in \mathfrak{T}^*$ and convergence of first and second moments; this is the method by Devroye (1991). We can now also show that (19) holds.

We may suppose that $a_1, \dots, a_N$ all are nonzero, and that $T_1, \dots, T_N$ are ordered with $|T_1| \le |T_2| \le \dots$, so no T_j is a proper subtree of T_1. Let $n > 3\ell$, and consider the event that $U_1 < U_2 < \dots < U_n$; this generates a tree $\mathcal{T}_n = T'$ that is a path to the right from the root. By permuting $U_\ell, \dots, U_{\ell+k}$, where $k = |T_1|$, leaving all other U_i unchanged, we may instead generate a tree T'' that is a path to the right of length $n-k$, with a copy of T_1 attached to the ℓth vertex. Then $n_{T_1}(T'') = n_{T_1}(T') + 1$, but $n_{T_j}(T'') = n_{T_j}(T')$ for $2 \le j \le N$, since except for the new copy of T_1 in T'', the subtrees that appear or disappear when we change T' to T'' are either too small or too large to be a T_j. Hence, by (18), $F(T') \ne F(T'')$, and this holds also if we ignore the boundary trees and consider S_n as in (22), and it follows by Corollary 2, see Remark 4, that (19) holds. (The proof just given was our first proof that $\sigma^2 > 0$ in this case. The proof given in Holmgren and Janson (2015) is actually slightly different and does not use the results in this paper; it uses instead a shortcut based on a special symmetry property.)

Example 2. (*Conditioned Galton–Watson trees, Janson (2014).*) A Galton–Watson tree $\mathcal{T}$ is the tree version of a Galton–Watson process. It is defined by a nonnegative integer-valued random variable ξ which describes the number of children of each node. We assume that $\mathbb{E}\xi = 1$ (a *critical* Galton–Watson process) and $\mathbb{E}\xi^2 < \infty$. The conditioned Galton–Watson tree $\mathcal{T}_n$ is the random tree $\mathcal{T}$ conditioned to have exactly n nodes. It is well known that several standard types of random trees can be defined in this way, with suitable ξ; see, for example, Janson (2012). For simplicity, we assume that $\mathbb{P}(\xi = k) > 0$ for every $k \ge 0$, and let $\mathfrak{T}^*$ be the family of all ordered rooted trees. (The general case is studied in Janson (2014) with a minor variation of the argument below. The result is the same as long as ξ attains at least two positive integers with positive probability, except that $\mathfrak{T}^*$ only consists of trees, where all outdegrees may be attained by ξ, but in the case when $\xi \in \{0, r\}$ for some integer r, we have to exclude the $T = \bullet$ case, the tree of size 1, because $n_\bullet(\mathcal{T}_n)$ is deterministic.)

Let $\xi_1, \xi_2, \dots$ be an i.i.d. sequence of copies of ξ, and let $Z_n := \sum_{i=1}^n \xi_i$. The degree sequence of the nodes in $\mathcal{T}_n$, taken in depth-first order, is $(\xi_1, \dots, \xi_n)$ conditioned on this being the degree sequence of a tree; up to a cyclic shift this is the same as conditioning on $Z_n = n-1$ and it follows that

$$n_T(\mathcal{T}_n) \overset{\mathrm{D}}{=} \left(\sum_{i=1}^n f_T(\xi_i, \dots, \xi_{i+k-1 \bmod n}) \,\middle|\, Z_n = n-1 \right)$$

https://doi.org/10.1239/jap/1450802758 Published online by Cambridge University Press

for a suitable indicator function $f_T\colon \mathbb{N}^k \to \{0,1\}$, where $k = |T|$. Given F as in (18), we let $\ell := \max_j |T_j|$ and $f(x_1,\dots,x_\ell) := \sum_j a_j f_{T_j}(x_1,\dots,x_{|T_j|})$; then, again ignoring some boundary terms,

$$F(\mathcal{T}_n) \overset{\mathrm{D}}{=} \Biggl(\sum_{i=1}^{n-\ell} f(\xi_i,\dots,\xi_{i+\ell-1}) \Biggm| Z_n = n-1\Biggr) + O(1) = (S_n \mid Z_n = n-1) + O(1).$$

In this case, we thus have a conditioned version of the sum S_n, and asymptotic normality follows by a method by Le Cam (1958) and Holst (1981); see also Janson (2014). The proof in Janson (2014) shows that the asymptotic variance σ^2 is given by

$$\sigma^2 = \lim_{n\to\infty} \frac{1}{n}\operatorname{var}(S_n - \alpha Z_n),$$

where the constant α is chosen such that $\operatorname{cov}(S_n - \alpha Z_n, Z_n)/n \to 0$. Let $\tilde{S}_n := S_n - \alpha Z_n$. Then $\tilde{S}_n - \mathbb{E}\tilde{S}_n = \sum_{i=1}^n X_i$, where

$$X_i := f(\xi_i,\dots,\xi_{i+\ell-1}) - \alpha\xi_i + \beta, \tag{23}$$

with β chosen such that $\mathbb{E}X_i = 0$. If $\sigma^2 = 0$, we may apply Corollary 2 to (X_i) and $(\tilde{S}_n)$.

Take first $\xi_i = j$ for all $i \le n+\ell-1$, for some $j > 0$. Then $(\xi_i,\dots,\xi_{i+k-1})$ is never the degree sequence of a tree, so $f_T(\xi_i,\dots,\xi_{i+|T|-1}) = 0$ and $f(\xi_i,\dots,\xi_{i+\ell-1}) = 0$; hence, (23) reduces to $X_i = -\alpha j + \beta$, and (7) yields $n(-\alpha j + \beta) = 0$. Hence, $-\alpha j + \beta = 0$ for every $j > 0$ and, thus, $\alpha = \beta = 0$. Consequently, (23) simplifies to $X_i := f(\xi_i,\dots,\xi_{i+\ell-1})$.

We may again assume that $|T_1| \le |T_2| \le \dots \le |T_N|$ and $a_1 \ne 0$. Let $n > 2\ell$ and assume that $(\xi_{\ell+1},\dots,\xi_{\ell+|T_1|})$ equals the degree sequence of T_1, while all other $\xi_i = 2$, say, for $i \le n+\ell-1$. The only substrings of $\xi_1,\dots,\xi_{n+\ell-1}$ that are degree sequences of trees are $(\xi_{\ell+1},\dots,\xi_{\ell+|T_1|})$ and some of its substrings, corresponding to T_1 and its subtrees. It follows that $\tilde{S}_n - \mathbb{E}\tilde{S}_n = a_1 \ne 0$, which contradicts (7). This contradiction proves $\sigma^2 > 0$, i.e. (19).

Acknowledgements

Research partly supported by the Knut and Alice Wallenberg Foundation. I thank Richard Bradley for several valuable comments.

References

Aaronson, J., Gilat, D., Keane, M. and de Valk, V. (1989). An algebraic construction of a class of one-dependent processes. *Ann. Prob.* **17,** 128–143.

Bradley, R. C. (1980). A remark on the central limit question for dependent random variables. *J. Appl. Prob.* **17,** 94–101.

Bradley, R. C. (2007a). *Introduction to Strong Mixing Conditions*, Vol. 1. Kendrick Press, Heber City, UT.

Bradley, R. C. (2007b). *Introduction to Strong Mixing Conditions*, Vol. 2. Kendrick Press, Heber City, UT.

Bradley, R. C. (2007c). *Introduction to Strong Mixing Conditions*, Vol. 3. Kendrick Press, Heber City, UT.

Burton, R. M., Goulet, M. and Meester, R. (1993). On 1-dependent processes and k-block factors. *Ann. Prob.* **21,** 2157–2168.

Devroye, L. (1991). Limit laws for local counters in random binary search trees. *Random Structures Algorithms* **2,** 303–315.

Devroye, L. (2002). Limit laws for sums of functions of subtrees of random binary search trees. *SIAM J. Comput.* **32,** 152–171.

Diananda, P. H. (1955). The central limit theorem for m-dependent variables. *Proc. Camb. Phil. Soc.* **51,** 92–95.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd edn. John Wiley, New York.

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

Holmgren, C. and Janson, S. (2015). Limit laws for functions of fringe trees for binary search trees and random recursive trees. *Electron. J. Prob.* **20,** 51 pp.

Holst, L. (1981). Some conditional limit theorems in exponential families. *Ann. Prob.* **9,** 818–830.

Ibragimov, I. A. (1975). A note on the central limit theorems for dependent random variables. *Theory Prob. Appl.* **20,** 135–141.

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

Janson, S. (2012). Simply generated trees, conditioned Galton–Watson trees, random allocations and condensation. *Prob. Surveys* **9,** 103–252.

Janson, S. (2014). Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton–Watson trees. *Random Structures Algorithms* 10.1002/rsa.20568.

Le Cam, L. (1958). Un théorème sur la division d'un intervalle par des points pris au hasard. *Publ. Inst. Statist. Univ. Paris* **7,** 7–16.

Leonov, V. P. (1961). On the dispersion of time-dependent means of a stationary stochastic process. *Theory Prob. Appl.* **6,** 87–93.

Robinson, E. A. (1960). Sums of stationary random variables. *Proc. Amer. Math. Soc.* **11,** 77–79.

Schmidt, K. (1977). *Cocycles on Ergodic Transformation Groups* (Macmillan Lectures Math. **1**). Macmillan Company of India, Delhi.