Mathematical Proceedings of the Cambridge Philosophical Society

VOL. 139 JULY 2005 PART 1

Math. Proc. Camb. Phil. Soc. (2005), **139**, 1 © 2005 *Cambridge Philosophical Society*
doi:10.1017/S0305004105008583 *Printed in the United Kingdom*

A probabilistic approach to analytic arithmetic on algebraic function fields

BY RICHARD ARRATIA[1], A. D. BARBOUR†[2] and SIMON TAVARÉ[1]

[1]*Department of Mathematics, University of Southern California, Los Angeles, CA* 90089-1113, U.S.A. [2]*Abteilung für Angewandte Mathematik, Universität Zürich, Winterthurerstrasse* 190, *CH*-8057, *Zürich, Switzerland.*

(*Received* 7 *July* 2003; *Revised* 15 *December* 2003)

Abstract

Knopfmacher [**13**] introduced the idea of an additive arithmetic semigroup as a general setting for an algebraic analogue of number theory. Within his framework, Zhang [**19**] showed that the asymptotic distribution of the values taken by additive functions closely resembles that found in classical number theory, in as much as there are direct analogues of the Erdős–Wintner and Kubilius Main Theorems. In this paper, we use probabilistic arguments to show that similar theorems, and their functional counterparts, can be proved in a much wider class of decomposable combinatorial structures.

1. *Introduction*

Knopfmacher [**13**] formalized the idea of an additive arithmetic semigroup $\mathcal{G}$, which he defined to be a free commutative semigroup with identity element 1 having a countable free generating set $\mathcal{P}$ of primes p and a degree mapping $\partial : \mathcal{G} \to \mathbb{Z}_+$ satisfying:

$$(1)\ \partial(ab) = \partial(a) + \partial(b) \text{ for all } a, b \in \mathcal{G};$$
$$(2)\ G(n) < \infty \text{ for each } n \geqslant 0,$$

† Supported in part by Schweizerischer Nationalfondsprojekte Nrs 20-61753.00 and 20-67909.02.

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

where $G(n)$ denotes the number of elements of degree n in $\mathcal{G}$; $\partial(a) = 0$ if and only if $a = 1$. Monic polynomials over a finite field are an example of such a semigroup, with irreducible polynomials as primes and with the degree of the polynomial as ∂. Knopfmacher used additive arithmetic semigroups as a general setting for an algebraic analogue of number theory, an extended treatment of which is to be found in [**14**]. Here, we consider only the properties of additive functions.

A real function f on $\mathcal{G}$ is *additive* if $f(ab) = f(a) + f(b)$ for each coprime pair $a, b \in \mathcal{G}$; f is strongly additive if $f(p^k) = f(p)$ for each $p \in \mathcal{P}$, $k \geqslant 1$, and f is completely additive if $f(p^k) = kf(p)$. For example, with $f(p) = 1$ for all $p \in \mathcal{P}$ and f completely additive, then $f(a)$ is the number of prime factors of a; if instead f is strongly additive, then $f(a)$ is the number of distinct prime factors of a. The asymptotic distribution of the values taken by $f(a)$ when a ranges over all elements such that $\partial(a) = n$ and $n \to \infty$ has been shown to parallel that of additive functions in classical number theory, provided that $\mathcal{G}$ satisfies some variant of Knopfmacher's Axiom $A^{\#}$: that $G(n) = Aq^n(1 + O(e^{-\alpha n}))$ for some $q > 1$ and $\alpha > 0$. In particular, Zhang [**19**] proves analogues of the Erdős–Wintner and Kubilius Main Theorems in classical probabilistic number theory; they are also to be found in [**14**, chapter 7], though under somewhat more restrictive conditions. Zhang's approach is based on generating function arguments, related to those used in classical number theory.

In this paper, we are interested in deriving similar theorems using considerations of a more directly probabilistic nature. We show that both can be established using a simple probabilistic identity imbedded in the structure of an additive arithmetic semigroup, if some additional asymptotic regularity is assumed. Both the identity and the asymptotic regularity are also found in a wide variety of other combinatorial structures, so that our approach greatly broadens the applicability of the two theorems. Furthermore, even in the original context of additive arithmetic semigroups, we are able to weaken some of the conditions imposed in [**19**], though at the cost of assuming somewhat stronger asymptotic regularity.

Our treatment is based on two observations. The first is just that studying the distribution of the values taken by $f(a)$ when a ranges over all elements such that $\partial(a) = n$ is equivalent to studying the probability distribution of $f(a)$ when an element $a \in \mathcal{G}$ is chosen uniformly at random from those having $\partial(a) = n$. This is a trite remark, but it allows us to introduce the language and methods of probability. The second observation is that additive arithmetic semigroups are multisets, in the sense of Arratia and Tavaré [**1**]. A multiset is a set of distinguishable objects of sizes $1, 2, \ldots$, some of which are irreducible. The general object is composed of any finite unordered collection of irreducible objects, with repeats allowed, and its size is the sum of the sizes of these irreducible objects. An additive arithmetic semigroup can thus be seen as a multiset, with primes as irreducible objects and with degree as the size of an object. Thus, for instance, in multiset representation, the element p^2 of an additive arithmetic semigroup becomes the object of size $2\partial(p)$ consisting of two copies of the irreducible object p.

An object of size n from a multiset (additive arithmetic semigroup) is said to have component size spectrum $C^{(n)} := (C_1^{(n)}, \ldots, C_n^{(n)})$ if $C_j^{(n)}$ denotes the number of irreducible objects (prime factors) of size j in its composition, $1 \leqslant j \leqslant n$; of course, size n means that $\sum_{j=1}^{n} jC_j^{(n)} = n$. Now the distribution of $C^{(n)}$, when an object is chosen uniformly and at random from all those of size n, has been extensively studied

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

for multisets; see [**1**], for example. To express their results, we introduce two further definitions. We say that a sequence of random vectors $W^{(n)}$, $n \geqslant 1$, with $W^{(n)} \in \mathbb{Z}_+^n$ a.s., has distributions satisfying the *Conditioning Relation* if

$$\mathcal{L}\big(W^{(n)}\big) = \mathcal{L}\left((Z_1, \ldots, Z_n) \,\middle|\, \sum_{j=1}^{n} jZ_j = n\right),$$

for some sequence $Z_1, Z_2, \ldots$ of independent nonnegative integer valued random variables. Their distributions also satisfy the *Logarithmic Condition* if

$$\lim_{j\to\infty} j\mathbb{P}[Z_j = 1] = \lim_{j\to\infty} j\mathbb{E}Z_j = \theta$$

for some $0 < \theta < \infty$. Then Arratia and Tavaré [**1**] observe that the component size spectra $C^{(n)}$ for multisets satisfy the Conditioning Relation, with the $Z_j \sim NB(m_j, x^j)$ having negative binomial distributions:

$$NB(k,p)\{r\} := (1-p)^k \binom{k+r-1}{r} p^r, \quad r \in \mathbb{Z}_+,$$

where m_j denotes the number of irreducible objects of size j, and x may be taken to be *any* value in $(0,1)$. They also satisfy the Logarithmic Condition if $m_j \sim \theta j^{-1} y^j$ for some $y > 1$, provided that one chooses $x = y^{-1}$. Many, but not all of the examples considered in [**14**] indeed also satisfy the Logarithmic Condition, and we shall take it as a basic condition from now on.

In a multiset, *any* finite unordered collection of irreducible objects (repeats allowed) makes up a permissible object. Many other well-known combinatorial structures can be obtained by imposing restrictions on the collections allowed. For instance, by not allowing repeated choices of the same irreducible object, the class of monic polynomials over a finite field is changed into the class of square-free monic polynomials over the field. Similarly, if cyclic permutations of distinct elements of $\mathbb{N}$ are taken to be the irreducible objects, with size the number of elements in the cycle, then there are strong restrictions on the collections of them which are allowed, when combining them into a permutation. Combinatorial structures which are obtained in this way we refer to as decomposable. The notion of component size spectrum is inherited from the parent multiset, but its distribution is in general very different. However, many of the classical decomposable structures also have component size spectra which satisfy the Conditioning Relation and the Logarithmic Condition, though the random variables $Z_1, Z_2, \ldots$ may no longer be negative binomially distributed; for instance, for random permutations decomposed into cycles, the Z_i have Poisson distributions, and for random square-free polynomials decomposed into irreducible factors they have binomial distributions.

We refer to any decomposable combinatorial structure satisfying the Conditioning Relation and the Logarithmic Condition as a logarithmic combinatorial structure, and we include them in our treatment as well, provided that they also satisfy the *Uniform Logarithmic Condition* (ULC):

$$\varepsilon_{i1} := i\mathbb{P}[Z_i = 1] - \theta \quad \text{satisfies} \quad |\varepsilon_{i1}| \leqslant e(i)c_1; \tag{1·1}$$

$$\varepsilon_{il} := i\mathbb{P}[Z_i = l] \leqslant e(i)c_l, \quad l \geqslant 2; \tag{1·2}$$

$$\text{where } e(i) \downarrow 0 \text{ as } i \to \infty \text{ and } D_1 = \sum_{l\geqslant 1} lc_l < \infty,$$

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

a condition which, for multisets, follows automatically from the Logarithmic Condition (Arratia, Barbour and Tavaré [**4**, proposition 1·1]); and provided also that

$$\sum_{i\geqslant 1} i^{-1}e(i) < \infty, \tag{1·3}$$

which, in the case of multisets satisfying the Logarithmic Condition with $m_j \sim \theta j^{-1}y^j$, requires in addition that

$$\sum_{i\geqslant 1} i^{-1} \sup_{j\geqslant i} |jm_jy^{-j} - \theta| < \infty.$$

These conditions ensure that the following proposition holds [**4**, theorems 3·1 and 3·2], describing the asymptotic behaviour of the component size spectrum: Po (μ) is used to denote the Poisson distribution with mean μ.

Proposition 1·1. *For a logarithmic combinatorial structure satisfying the ULC and* (1·3), *we have*:

(1) $\sum_{j\geqslant 1} \mathbb{P}[Z_j \geqslant 2] < \infty$;

(2) $\lim_{n\to\infty} d_{TV}\big(\mathcal{L}\big(C^{(n)}[1, b(n)]\big), \mathcal{L}(Z[1, b(n)])\big) = 0$ *if* $b(n) = o(n)$;

(3) $\lim_{n\to\infty} d_{TV}\big(\mathcal{L}\big(C^{(n)}[b(n)+1, n]\big), \mathcal{L}\big(C^{*(n)}[b(n)+1, n]\big)\big) = 0$ *if* $b(n) \to \infty$,

where $X[r, s]$ *denotes the vector* $(X_r, X_{r+1}, \ldots, X_s)$ *and* $C^{*(n)}$ *has as distribution the Ewens Sampling Formula with parameter* θ, *obtained through the Conditioning Relation with* $Z_j = Z_j^* \sim \text{Po}\,(\theta/j)$.

Since the additive function f which counts the number of prime factors is just the function $\sum_{j=1}^n C_j^{(n)}$, and that which counts the number of distinct prime factors is just $\sum_{j=1}^n I[C_j^{(n)} \geqslant 1]$, their distributions can be deduced directly from that of $C^{(n)}$. However, the value $f(a)$ at $a \in \mathcal{G}$ for most additive functions depends not only on the component structure of a, but also on which irreducible objects of the different component sizes it is composed of. For instance, if $C_j^{(n)} = 1$, then $f(a)$ contains a contribution $f(p)$ from one of the m_j primes p with $\partial(p) = j$; if $C_j^{(n)} = 2$, there is either a contribution $f(p) + f(p')$ from one of the $\binom{m_j}{2}$ distinct pairs of primes of degree j, or a contribution $f(p^2)$ from a repeated prime p of degree j. Because, in choosing a random instance of a multiset, the particular irreducible objects of each component size that are chosen are also random, there is randomness additional to that of the component structure, and it is carried over into the distribution of $f(a)$; what is more, for an object of size n chosen uniformly at random, the random choices of irreducible objects are conditionally independent, given $C^{(n)}$. This motivates consideration of the following general construct, which can be defined for any logarithmic combinatorial structure, and not just for multisets:

$$X^{(n)} = \sum_{j=1}^n I\big[C_j^{(n)} \geqslant 1\big] U_j\big(C_j^{(n)}\big), \tag{1·4}$$

where the $(U_j(l),\ j, l \geqslant 1)$ are independent random variables which are also independent of $C^{(n)}$.

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

For an additive function f on an additive arithmetic semigroup, $X^{(n)}$ constructed as above indeed models $f(a)$, for randomly chosen objects $a \in \mathcal{G}$ with $\partial(a) = n$, if the distributions of the random variables $U_j(l)$ are specified as follows. The distribution of $U_j(1)$ assigns probability $1/m_j$ to $f(p)$ for each of the m_j primes p of degree j; $U_j(2)$ gives probability $2/m_j(m_j+1)$ to $f(p)+f(p')$ for each of the $\binom{m_j}{2}$ pairs of distinct primes p and p' of degree j, and probability $2/m_j(m_j+1)$ to each $f(p^2)$; and so on. In the example with $f(p)=1$ for all primes p and f completely additive, counting the total number of prime factors, then $U_j(l)=l$ a.s. for all j; if instead f is strongly additive, counting the number of distinct prime factors, then $U_j(l)$ has a more complicated distribution.

Our goal is to describe aspects of the limiting behaviour of $X^{(n)}$, defined in (1·4), for a general logarithmic combinatorial structure $C^{(n)}$ and for certain choices of the random variables $U_j(l)$. In Sections 2 and 3, we investigate choices of the $U_j(l)$ which lead either to the convergence of $X^{(n)}$ in distribution or to a central limit type of behaviour, results analogous to the Erdős–Wintner theorem and Kubilius' Main Theorem in probabilistic number theory. In both cases, our proof consists of showing that only the small components contribute significantly to the result; once this has been shown, Proposition 1·1(2) reduces the problem to that of a sum of independent random variables, to which classical theory can be applied. This general strategy is strongly reminiscient of that used by Kubilius [**16**], though our setting is quite different. The final section concerns circumstances in which the behaviour of $X^{(n)}$ is dominated by that of the large components, and the dependence becomes all important; here, the approximations are formulated in terms of the Ewens Sampling Formula.

Results of the first two kinds were proved by Zhang [**19**], in the particular case of additive arithmetic semigroups. However, our conditions are rather different from his. For additive arithmetic semigroups, the Logarithmic Condition is expressed by requiring that $jm_jx^j \to \theta$ for some $0<\theta<\infty$ and $0<x<1$, a condition involving the numbers m_j of irreducible elements of size j. In contrast, Zhang formulates his conditions in terms of the total numbers $G(n)$ of elements of size $n \geqslant 1$, and one of his main interests is to derive from them information about the asymptotics of the m_j, in analogy to the prime number theorem. These asymptotics are often equivalent to the Logarithmic Condition, together with a rate of convergence, but need not be: there are also cases in which his prime number theorem yields more complicated asymptotics, and his analogue of the Erdős–Wintner and Kubilius Main Theorems are thus valid for a number of additive arithmetic semigroups which do not satisfy the Logarithmic Condition. However, by assuming the Logarithmic Condition and (1·3), we are able to relax Zhang's other conditions, even in the context of additive arithmetic semigroups; for instance, our results are valid whatever the value of $\theta>0$, whereas Zhang's theorems can be applied in our setting only in situations where $\theta \geqslant 1$, and at times only when $\theta=1$: see Section 5.

The classical definition of an additive function also allows f to be complex valued. For complex valued f, both real and imaginary parts are real valued additive functions, and for our purposes such an f can be treated using a two dimensional generalization of (1·4). Even greater generality can be achieved by considering the

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

construction

$$X^{(n)} = \sum_{j=1}^{n} I\big[C_j^{(n)} \geqslant 1\big] U_j\big(C_j^{(n)}\big) \tag{1·5}$$

with independent d-dimensional random vectors $U_j(l) = (U_{j1}(l), \ldots, U_{jd}(l))$, $j, l \geqslant 1$; the main theorems of the paper carry over without difficulty to this setting. Thus if, for example, each prime element in an additive arithmetic semigroup belongs to exactly one of d different classes, one could in principle investigate the asymptotics as $n \to \infty$ of the joint distribution of the numbers of primes in each class.

2. *Convergence*

The first set of results concerns conditions under which the random variables $X^{(n)}$ have a limit in distribution, without normalization. This theorem is our analogue of the Erdős–Wintner theorem in probabilistic number theory. Hereafter, we write U_j for $U_j(1)$.

THEOREM 2·1. *Suppose that a logarithmic combinatorial structure satisfies the Uniform Logarithmic Condition together with* (1·3). *Then $X^{(n)}$ converges in distribution if and only if the series*

$$\sum_{j\geqslant 1} j^{-1}\mathbb{P}[|U_j| > 1]; \quad \sum_{j\geqslant 1} j^{-1}\mathbb{E}\{U_j I[|U_j| \leqslant 1]\} \text{ and } \sum_{j\geqslant 1} j^{-1}\mathbb{E}\{U_j^2 I[|U_j| \leqslant 1]\} \tag{2·1}$$

all converge. If so, then

$$\lim_{n\to\infty} \mathcal{L}\big(X^{(n)}\big) = \mathcal{L}\left(\sum_{j\geqslant 1} I[Z_j \geqslant 1] U_j(Z_j)\right).$$

Proof. The three series (2·1) are equivalent to those of Kolmogorov's three series criterion [**17**, p. 249] for the sum of independent random variables $\sum_{j\geqslant 1} I[Z_j = 1]U_j$, since, from the Logarithmic Condition, $\mathbb{P}[Z_j = 1] \asymp j^{-1}$. By Proposition 1·1(1), it also follows that $\sum_{j\geqslant 1} I[Z_j = 1]U_j$ and $\sum_{j\geqslant 1} I[Z_j \geqslant 1]U_j(Z_j)$ are convergence equivalent. Hence it is enough to show that, for some sequence $b(n) \to \infty$, $X^{(n)}$ and $W_{0,b(n)}(Z)$ are asymptotically close to one another, where, for any $y \in \mathbb{Z}_+^\infty$ and any $0 \leqslant l < m$,

$$W_{l,m}(y) := \sum_{j=l+1}^{m} I[y_j \geqslant 1] U_j(y_j).$$

That this is the case follows from Lemmas 2·2 and 2·3 below.

LEMMA 2·2. *If a logarithmic combinatorial structure satisfies the Uniform Logarithmic Condition together with* (1·3), *and if*

$$\lim_{n\to\infty} n^{-1} \sum_{j=1}^{n} \mathbb{P}[|U_j| > \delta] = 0 \quad \text{for all } \delta > 0, \tag{2·2}$$

then there exists a sequence $b(n) \to \infty$ with $b(n) = o(n)$ such that $X^{(n)}$ and $W_{0,b(n)}(Z)$ are convergence equivalent.

Proof. First, we note that, for any b,

$$X^{(n)} = W_{0b}\big(C^{(n)}\big) + W_{bn}\big(C^{(n)}\big).$$

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

By Proposition 1·1(2),

$$d_{TV}\big(\mathcal{L}\big(W_{0,b(n)}\big(C^{(n)}\big)\big), \mathcal{L}\big(W_{0,b(n)}(Z)\big)\big) = o(1) \text{ as } n \to \infty,$$

for any sequence $b(n)$ such that $b(n) = o(n)$ as $n \to \infty$. Hence $W_{0,b(n)}(C^{(n)})$ and $W_{0,b(n)}(Z)$ are convergence equivalent for any such sequence $b(n)$. It thus remains to show that $W_{b(n),n}(C^{(n)}) \xrightarrow{\mathcal{D}} 0$ for some such sequence $b(n)$.

Now, from Proposition 1·1(3), it follows that

$$d_{TV}\big(\mathcal{L}\big(W_{b(n),n}\big(C^{(n)}\big)\big), \mathcal{L}\big(W_{b(n),n}\big(C^{*(n)}\big)\big)\big) \longrightarrow 0$$

provided only that $b(n) \to \infty$. Furthermore, defining

$$\widetilde{W}_{l,m}(y) \coloneqq \sum_{j=l+1}^{m} I[y_j = 1]U_j \tag{2·3}$$

for any $y \in \mathbb{Z}_+^\infty$ and any $0 \leqslant l < m$, we have

$$\begin{aligned} &d_{TV}\big(\mathcal{L}\big(W_{b(n),n}\big(C^{*(n)}\big)\big), \mathcal{L}\big(\widetilde{W}_{b(n),n}\big(C^{*(n)}\big)\big)\big) \\ &\quad \leqslant \mathbb{P}\left[\bigcup_{j=b(n)+1}^{n} \{C_j^{*(n)} \geqslant 2\}\right] \leqslant c_{\{6·9b\}}\{b(n)\}^{-1} \end{aligned} \tag{2·4}$$

from Lemma 6·9. Hence, so long as $b(n) \to \infty$, $W_{b(n),n}(C^{(n)}) \xrightarrow{\mathcal{D}} 0$ follows, if we can show that $\widetilde{W}_{b(n),n}(C^{*(n)}) \xrightarrow{\mathcal{D}} 0$.

Because of the assumption (2·2), there exists a sequence $\delta_n \to 0$ such that

$$\eta_n \coloneqq n^{-1} \sum_{j=1}^{n} \mathbb{P}[|U_j| > \delta_n] \longrightarrow 0$$

as $n \to \infty$. Thus, defining

$$A_n(b) \coloneqq \bigcup_{j=b+1}^{n} \{\{C_j^{*(n)} = 1\} \cap \{|U_j| > \delta_n\}\},$$

we have

$$\begin{aligned} \mathbb{P}[A_n(b)] &\leqslant \sum_{j=b+1}^{n} \mathbb{P}\big[C_j^{*(n)} = 1\big]\mathbb{P}[|U_j| > \delta_n] \\ &\leqslant \sum_{j=b+1}^{n} c_{\{6·9a\}} j^{-1} \left(\frac{n}{n-j+1}\right)^{1-\theta} \mathbb{P}[|U_j| > \delta_n], \end{aligned} \tag{2·5}$$

from Lemma 6·9. Thus, for any $n/2 < m < n$, it follows that

$$\begin{aligned} \mathbb{P}[A_n(b)] &\leqslant c_{\{6·9a\}}\left\{\frac{n}{b}\left(\frac{1}{n}\sum_{j=1}^{n}\mathbb{P}[|U_j| > \delta_n]\right)\left(\frac{n}{n-m+1}\right)^{1-\theta} + \frac{n^{1-\theta}(n-m+1)^{\theta}}{m\theta}\right\} \\ &\leqslant c_{\{6·9a\}}\left\{\frac{n\eta_n}{b}\left(\frac{n}{n-m}\right) + \frac{2}{\theta}\left(\frac{n-m+1}{n}\right)^{\theta}\right\}. \end{aligned} \tag{2·6}$$

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

Now, if $A_n(b)$ does not occur, then

$$\widetilde{W}_{bn}\big(C^{*(n)}\big) = \overline{W}_{bn} := \sum_{j=b+1}^{n} I\big[C_j^{*(n)} = 1\big] U_j I[|U_j| \leqslant \delta_n]$$

and

$$\mathbb{E}|\overline{W}_{bn}| \leqslant \delta_n \sum_{j=b+1}^{n} \mathbb{P}\big[C_j^{*(n)} = 1\big]. \tag{2·7}$$

Again, from Lemma 6·9, arguing much as above, we thus have

$$\mathbb{E}|\overline{W}_{bn}| \leqslant c_{\{6\cdot 9a\}}\delta_n \left\{ \left(\frac{n}{n-m}\right) \log\left(\frac{n+1}{b+1}\right) + \frac{2}{\theta}\left(\frac{n-m+1}{n}\right)^{\theta} \right\}. \tag{2·8}$$

So pick $b(n) = o(n)$ so large that

$$\eta_n' := \max\{n\eta_n/b(n), \delta_n \log(n/b(n))\} \longrightarrow 0,$$

and then pick $m(n)$ such that $n - m(n) = o(n)$ and yet $n\eta_n'/(n - m(n)) \to 0$; for these choices, it follows from (2·6) and (2·8) that

$$\lim_{n\to\infty} \mathbb{E}|\overline{W}_{b(n),n}| = 0 \quad \text{and} \quad \lim_{n\to\infty} \mathbb{P}\big[\widetilde{W}_{b(n),n}\big(C^{*(n)}\big) \neq \overline{W}_{b(n),n}\big] = 0,$$

and hence that $\widetilde{W}_{b(n),n}(C^{*(n)}) \xrightarrow{\mathcal{D}} 0$, completing the proof.

LEMMA 2·3. *If the three series* (2·1) *converge, or if the Uniform Logarithmic Condition and* (1·3) *hold and* $X^{(n)}$ *converges in distribution, then*

$$\lim_{n\to\infty} n^{-1} \sum_{j=1}^{n} \mathbb{P}[|U_j| > \delta] = 0 \quad \text{for all } \delta > 0.$$

Proof. The first part is standard, using Chebyshev's inequality and Kronecker's lemma; the second relies heavily on technical results whose proofs are deferred to Section 6.

We begin by showing that $\mathcal{L}(X^{(n)})$ is close in total variation to $\mathcal{L}(X^{(b,n)})$, for suitably chosen $b = b(n)$, where

$$X^{(b,n)} := \sum_{j=1}^{b} I\big[C_j^{(b,n)} \geqslant 1\big] U_j\big(C_j^{(b,n)}\big) + \sum_{j=b+1}^{n} \sum_{l=1}^{C_j^{(b,n)}} U_{jl}'. \tag{2·9}$$

Here, $\mathcal{L}(C^{(b,n)})$ is defined by using the Conditioning Relation, but based on the random variables $\widehat{Z}_j^b$, $j \geqslant 1$, given by

$$\widehat{Z}_j^b = Z_j,\ 1 \leqslant j \leqslant b, \quad \text{and} \quad \widehat{Z}_j^b = Z_j^*,\ j \geqslant b+1, \tag{2·10}$$

so that

$$\mathcal{L}(C^{(b,n)}) := \mathcal{L}\left(\big(\widehat{Z}_1^b, \ldots, \widehat{Z}_n^b\big) \,\middle|\, \sum_{j=1}^{n} j\widehat{Z}_j^b = n \right),$$

and the random variables $(U_{jl}',\ j \geqslant 1, l \geqslant 1)$ are independent of one another and of $C^{(b,n)}$, and are such that $\mathcal{L}(U_{jl}) = \mathcal{L}(U_j)$. We do this in two steps, by way of $\widehat{X}^{(b,n)} :=$

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

$\sum_{j=1}^{n} I[C_j^{(b,n)} \geqslant 1] U_j(C_j^{(b,n)})$. First, from Theorem 6·7, if $b(n) \to \infty$ and $n^{-1}b(n) \to 0$, then

$$\lim_{n\to\infty} d_{TV}\big(\mathcal{L}\big(X^{(n)}\big), \mathcal{L}\big(\widehat{X}^{(b(n),n)}\big)\big) = 0;$$

and then

$$d_{TV}\big(\mathcal{L}\big(\widehat{X}^{(b,n)}\big), \mathcal{L}\big(X^{(b,n)}\big)\big) \leqslant \sum_{j=b+1}^{n} \mathbb{P}\big[C_j^{(b,n)} \geqslant 2\big] \leqslant c_{\{6\cdot 9b\}} b^{-1},$$

by Lemma 6·9.

Hence, for any $f \in \mathcal{F}_{BL}$, where

$$\mathcal{F}_{BL} := \{f : \mathbb{R} \to \big[-\tfrac{1}{2}, \tfrac{1}{2}\big];\ \|f'\| \leqslant 1\}, \tag{2·11}$$

it follows that

$$|\mathbb{E}f\big(X^{(n)}\big) - \mathbb{E}f\big(X^{(b,n)}\big)| \leqslant \eta_1(n,b), \tag{2·12}$$

where $\eta_1(n,b)$ is increasing in n for each fixed b, and, if $b(n) \to \infty$ and $n^{-1}b(n) \to 0$, then $\lim_{n\to\infty} \eta_1(n, b(n)) = 0$.

Now let $R^{(b,n)}$ denote a size-biassed choice from $C^{(b,n)}$: that is,

$$\mathbb{P}\big[R^{(b,n)} = j \,|\, C^{(b,n)}\big] = jC_j^{(b,n)}/n. \tag{2·13}$$

Then a simple calculation shows that, for $b+1 \leqslant j \leqslant n$, and for any $c \in \mathbb{Z}_+^\infty$ with $\sum_{j\geqslant 1} jc_j = n$,

$$\mathbb{P}\big[C^{(b,n)} = c \,|\, R^{(b,n)} = j\big] = \mathbb{P}\big[C^{(b,n-j)} + \varepsilon^j = c\big],$$

where ε^j denotes the jth coordinate vector in $\mathbb{Z}_+^\infty$. Hence, for any $f \in \mathcal{F}_{BL}$, the equation

$$\mathbb{E}f\big(X^{(b,n)}\big) = \sum_{j=1}^{n} \mathbb{P}\big[R^{(b,n)} = j\big]\, \mathbb{E}\{f\big(X^{(b,n)}\big) \,|\, R^{(b,n)} = j\}$$

implies that

$$\mathbb{E}f\big(X^{(b,n)}\big) \sum_{j=1}^{n} \mathbb{P}\big[R^{(b,n)} = j\big] = \sum_{j=1}^{n} \mathbb{P}\big[R^{(b,n)} = j\big]\, \mathbb{E}f\big(X^{(b,n-j)} + \widehat{U}_j\big),$$

where $\widehat{U}_j$ is independent of $X^{(b,n-j)}$ and $\mathcal{L}(\widehat{U}_j) = \mathcal{L}(U_j)$. Hence, for any $m \in [b+1, n]$, we have

$$\begin{aligned}\Bigg| \sum_{j=b+1}^{m} \mathbb{P}\big[R^{(b,n)} = j\big]\{\mathbb{E}f\big(X^{(b,n)}\big) - \mathbb{E}f\big(X^{(b,n-j)} + \widehat{U}_j\big)\}\Bigg| \\ \leqslant \mathbb{P}\big[R^{(b,n)} \leqslant b\big] + \mathbb{P}\big[R^{(b,n)} > m\big].\end{aligned} \tag{2·14}$$

If $X^{(n)}$ converges in distribution to some X^∞, then

$$\eta_2(m) := \sup_{n\geqslant m} \sup_{f\in\mathcal{F}_{BL}} |\mathbb{E}f\big(X^{(n)}\big) - \mathbb{E}f\big(X^\infty\big)|$$

exists and satisfies $\lim_{m\to\infty} \eta_2(m) = 0$, by Dudley [**6**, theorem 8·3]. Hence, from (2·14),

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

it follows that if $V^{(b,n)}$ is independent of X^∞ and satisfies

$$\mathbb{P}\big[V^{(b,n)} \in A\big] = \sum_{j=1}^{n} \mathbb{P}\big[R^{(b,n)} = j\big]\,\mathbb{P}[U_j \in A],$$

then

$$\begin{aligned}
&\big|\mathbb{E}f(X^\infty) - \mathbb{E}f\left(X^\infty + V^{(b,n)}\right)\big| \\
&= \left|\sum_{j=1}^{n} \mathbb{P}\left[R^{(b,n)} = j\right]\left\{\mathbb{E}f\left(X^\infty\right) - \mathbb{E}f(X^\infty + \widehat{U}_j)\right\}\right| \\
&\leqslant \mathbb{P}\big[R^{(b,n)} \leqslant b\big] + \mathbb{P}\big[R^{(b,n)} > m\big] \\
&\quad + \left|\sum_{j=b+1}^{m} \mathbb{P}\big[R^{(b,n)} = j\big]\{\mathbb{E}f(X^\infty) - \mathbb{E}f(X^\infty + \widehat{U}_j)\}\right| \\
&\leqslant \mathbb{P}\big[R^{(b,n)} \leqslant b\big] + \mathbb{P}\big[R^{(b,n)} > m\big] + \eta_2(n) + \eta_1(n,b) \\
&\quad + \sum_{j=b+1}^{m} \mathbb{P}\big[R^{(b,n)} = j\big](\eta_2(n-j) + \eta_1(n-j,b)) \\
&\quad + \left|\sum_{j=b+1}^{m} \mathbb{P}\big[R^{(b,n)} = j\big]\{\mathbb{E}f\big(X^{(b,n)}\big) - \mathbb{E}f\big(X^{(b,n-j)} + \widehat{U}_j\big)\}\right| \\
&\leqslant 2\mathbb{P}\big[R^{(b,n)} \leqslant b\big] + 2\mathbb{P}\big[R^{(b,n)} > m\big] + 2\eta_1(n,b) + 2\eta_2(n-m). \qquad (2{\cdot}15)
\end{aligned}$$

Furthermore, from Lemma 6·6,

$$\mathbb{P}\big[R^{(b(n),n)} \leqslant b(n)\big] = n^{-1} \sum_{j=b(n)+1}^{n} j\,\mathbb{E}C_j^{(b(n),n)} \longrightarrow 0$$

provided only that $n^{-1}b(n) \to 0$, and, from Lemma 6·9, whatever the value of b,

$$\mathbb{P}\big[R^{(b,n)} > m\big] = n^{-1} \sum_{j=m+1}^{n} j\mathbb{P}\big[C_j^{(b,n)} = 1\big] \leqslant \theta^{-1} c_{\{6\cdot 9a\}} \left(\frac{n-m+1}{n}\right)^{\theta}.$$

Hence, for any choice of $b(n)$ such that $b(n) \to \infty$ with $b(n) = o(n)$, we can choose $m(n)$ such that $n - m(n) \to \infty$ and that $n - m(n) = o(n)$, and deduce that

$$\lim_{n\to\infty} \big|\mathbb{E}f(X^\infty) - \mathbb{E}f\big(X^\infty + V^{(b(n),n)}\big)\big| = 0,$$

for all $f \in \mathcal{F}_{BL}$. Thus, considering complex exponentials in place of f, it follows easily that $V^{(b(n),n)} \xrightarrow{\mathcal{D}} 0$ [**17**, application 3, p. 210], and hence that

$$\mathbb{P}\big[\,\big|V^{(b(n),n)}\big| > \delta\big] = \sum_{j=1}^{n} \mathbb{P}\big[R^{(b(n),n)} = j\big]\,\mathbb{P}[|U_j| > \delta] \longrightarrow 0,$$

for all $\delta > 0$.

Finally, from the definition of $R^{(b,n)}$, for $b(n) + 1 \leqslant j \leqslant n/2$, we have

$$\mathbb{P}\big[R^{(b(n),n)} = j\big] = \frac{\theta}{n} \frac{\mathbb{P}\big[T_{0n}\big(\widehat{Z}^{b(n)}\big) = n - j\big]}{\mathbb{P}\big[T_{0n}\big(\widehat{Z}^{b(n)}\big) = n\big]} \geqslant n^{-1}\theta(k_-/k_+),$$

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

from Lemma 6·4. Hence we have proved that

$$\lim_{n\to\infty} n^{-1} \sum_{j=b(n)+1}^{n/2} \mathbb{P}[|U_j| > \delta] = 0,$$

and, since $b(n) = o(n)$, the lemma follows.

Theorem 2·1 has a d-dimensional analogue. Since each component of a d-dimensional additive function is a real additive function, the sequence of random vectors $X^{(n)}$ defined in (1·5) has a limit if and only if, for all $1 \leqslant s \leqslant d$, the three series in (2·1) with U_j replaced by U_{js} all converge. It is then not hard to see that this criterion is equivalent to the convergence of the three series

$$\sum_{j\geqslant 1} j^{-1}\mathbb{P}[|U_j| > 1]; \quad \sum_{j\geqslant 1} j^{-1}\mathbb{E}\{U_j I[|U_j| \leqslant 1]\} \quad \text{and} \quad \sum_{j\geqslant 1} j^{-1}\mathbb{E}\{|U_j|^2 I[|U_j| \leqslant 1]\}, \tag{2·16}$$

only the second of which is $\mathbb{R}^d$-valued. For complex valued U_j, the third series can also be replaced by $\sum_{j\geqslant 1} j^{-1}\mathbb{E}\{U_j^2 I[|U_j| \leqslant 1]\}$, recovering the same form as in (2·1).

3. *Slow growth*

In this section, we consider situations in which $X^{(n)}$ converges, after appropriate normalization, to some infinitely divisible limit having finite variance. We assume that

$$\sigma^2(m) := \sum_{j=1}^{m} j^{-1}\mathbb{E}U_j^2 \longrightarrow \infty \text{ as } m \longrightarrow \infty; \quad \sigma^2 \text{ is slowly varying at } \infty, \tag{3·1}$$

where $U_j = U_j(1)$ as before; these conditions are equivalent for additive arithmetic semigroups to condition H of [**19**], the analogue of Kubilius's [**15**] condition H.

Lemma 3·1. *Suppose that a logarithmic combinatorial structure satisfies the Uniform Logarithmic Condition together with* (1·3)*, and that* (3·1) *holds. Then there exists a sequence $b(n) \to \infty$ with $b(n) = o(n)$ such that*

$$\sigma(n)^{-1} W'_{b(n),n}\big(C^{(n)}\big) \xrightarrow{\mathcal{D}} 0,$$

where, for $y \in \mathbb{Z}_+^\infty$,

$$W'_{lm}(y) := \sum_{j=l+1}^{m} I[y_j \geqslant 1]|U_j(y_j)|. \tag{3·2}$$

Proof. As in the proof of Lemma 2·2, we have

$$d_{TV}\big(\mathcal{L}\big(W'_{b(n),n}\big(C^{(n)}\big)\big), \mathcal{L}\big(\widetilde{W}'_{b(n),n}\big(C^{*(n)}\big)\big)\big) \longrightarrow 0$$

as $n \to \infty$, provided only that $b(n) \to \infty$, where

$$\widetilde{W}'_{lm}(y) := \sum_{j=l+1}^{m} I[y_j = 1]|U_j|; \tag{3·3}$$

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

hence we need only consider $\widetilde{W}'_{b(n),n}(C^{*(n)})$. Now, for any $n/2 \leqslant m \leqslant n$, by Lemma 6·9,

$$\mathbb{P}\left[\sum_{j=m+1}^{n} I\big[C_j^{*(n)} = 1\big]|U_j| \neq 0\right] \leqslant \frac{2}{n}\sum_{j=m+1}^{n} c_{\{6\cdot 9a\}}\left(\frac{n}{n-j+1}\right)^{1-\theta} \leqslant 2\theta^{-1}c_{\{6\cdot 9a\}}\left(\frac{n-m+1}{n}\right)^{\theta}, \tag{3·4}$$

so that the sum from $m+1$ to n contributes with asymptotically small probability, provided that $n-m$ is small compared to n. On the other hand, again from Lemma 6·9,

$$\sigma^{-1}(n)\mathbb{E}\left(\sum_{j=b+1}^{m} I\big[C_j^{*(n)} = 1\big]|U_j|\right) \leqslant \sigma^{-1}(n)\sum_{j=b+1}^{m}\mathbb{P}\big[C_j^{*(n)} = 1\big]\,\mathbb{E}|U_j| \leqslant \sigma^{-1}(n)c_{\{6\cdot 9a\}}\left(\frac{n}{n-m+1}\right)^{1-\theta}\sum_{j=b+1}^{m} j^{-1}\mathbb{E}|U_j|, \tag{3·5}$$

and, by the Cauchy–Schwarz inequality,

$$\sum_{j=b+1}^{m} j^{-1}\mathbb{E}|U_j| \leqslant \left\{\left(\sum_{j=b+1}^{m} j^{-1}\right)\left(\sum_{j=b+1}^{m} j^{-1}\mathbb{E}U_j^2\right)\right\}^{1/2} \leqslant \sigma(n)\{\log(n/b)(1-\sigma^2(b)/\sigma^2(n))\}^{1/2}. \tag{3·6}$$

Since σ^2 is slowly varying at ∞, we can pick $\beta(n) \to \infty$, $\beta(n) = o(n)$, in such a way that $\sigma^2(\beta(n))/\sigma^2(n) \to 1$. Hence we can pick $b(n) \to \infty$ with $\beta(n) \leqslant b(n) = o(n)$ in such a way that $\log(n/b(n))(1-\sigma^2(\beta(n))/\sigma^2(n)) \to 0$, and thus so that

$$\eta_n := \{\log(n/b(n))(1-\sigma^2(b(n))/\sigma^2(n))\}^{1/2} \longrightarrow 0. \tag{3·7}$$

Now pick $m = m(n)$ in such a way that $n - m(n) = o(n)$ and also $\{n/(n-m(n))\}^{1-\theta}\eta_n \to 0$. Then, from (3·4)–(3·6), it follows that

$$\sigma^{-1}(n)\widetilde{W}'_{b(n),n}\big(C^{*(n)}\big) \xrightarrow{\mathcal{D}} 0,$$

and the lemma is proved.

Thus, under the conditions of Lemma 3·1, there is a sequence $b(n) \to \infty$ with $b(n) = o(n)$ such that the asymptotic behaviour of the sequence $\sigma^{-1}(n)X^{(n)}$ is equivalent to that of $\sigma^{-1}(n)W_{1,b(n)}(C^{(n)})$, and, by Proposition 1·1(2),

$$d_{TV}\big(\mathcal{L}\big(C^{(n)}[1,b(n)]\big), \mathcal{L}(Z[1,b(n)])\big) = o(1) \text{ as } n \longrightarrow \infty. \tag{3·8}$$

Note also that

$$\mathbb{P}\left[\sup_{m\geqslant 1}\left|\sum_{j=1}^{m} I[Z_j \geqslant 1]U_j(Z_j) - \sum_{j=1}^{m} I[Z_j = 1]U_j\right| > \varepsilon\sigma(n)\right] \leqslant \mathbb{P}\left[\sum_{j=1}^{\infty} I[Z_j \geqslant 2]|U_j(Z_j)| > \varepsilon\sigma(n)\right], \tag{3·9}$$

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

where the infinite sum is finite a.s. by the Borel–Cantelli Lemma, from Proposition 1·1(1). Then one can also define independent Bernoulli random variables $\widehat{Z}_j \sim \mathrm{Be}\,(\theta/j)$ on the same probability space as the Z_j's and U_j's, independent also of the U_j's, in such a way that

$$\sum_{j\geqslant 1} \mathbb{P}[\widehat{Z}_j \neq I[Z_j = 1]] < \infty,$$

because, from (1·3),

$$\sum_{j\geqslant 1} |\mathbb{P}[Z_j = 1] - \theta j^{-1}| \leqslant \sum_{j\geqslant 1} j^{-1} e(j) < \infty.$$

Then we have

$$\mathbb{P}\left[\sup_{m\geqslant 1}\left|\sum_{j=1}^{m} I[Z_j = 1]U_j - \sum_{j=1}^{m} \widehat{Z}_j U_j\right| > \varepsilon\sigma(n)\right] \leqslant \mathbb{P}\left[\sum_{j=1}^{\infty} I[\widehat{Z}_j \neq I[Z_j = 1]]|U_j| > \varepsilon\sigma(n)\right], \tag{3·10}$$

with the infinite sum finite a.s. by the Borel–Cantelli Lemma. Since also $\sigma(n) \to \infty$, the right-hand sides of both (3·9) and (3·10) converge to zero as $n \to \infty$. Finally, as in the proof of Lemma 3·1,

$$\sigma(n)^{-1}\mathbb{E}\left(\sum_{j=b(n)+1}^{n} \widehat{Z}_j|U_j|\right) = \sigma(n)^{-1}\sum_{j=b(n)+1}^{n} \theta j^{-1}\mathbb{E}|U_j| \leqslant \eta_n, \tag{3·11}$$

where η_n is as defined in (3·7), and $\lim_{n\to\infty}\eta_n = 0$. Hence the asymptotic behaviour of $\sigma^{-1}(n)X^{(n)}$ is equivalent to that of

$$\sigma^{-1}(n)\widehat{X}^{(n)}, \quad \text{where} \quad \widehat{X}^{(n)} := \sum_{j=1}^{n} \widehat{Z}_j U_j, \tag{3·12}$$

in the following sense.

Theorem 3·2. *Suppose that a logarithmic combinatorial structure satisfies the Uniform Logarithmic Condition together with* (1·3)*, and that* (3·1) *holds. Then if, for any sequence $M(n)$ of centring constants, either of the sequences $\mathcal{L}(\sigma^{-1}(n)(\widehat{X}^{(n)} - M(n)))$ or $\mathcal{L}(\sigma^{-1}(n)(X^{(n)} - M(n)))$ converges as $n \to \infty$, so too does the other, and to the same limit.*

Note that $\widehat{X}^{(n)}$ is just a sum of independent random variables, with distribution depending only on θ and the distributions of the U_j, to which standard theory can be applied. Note also that the theorem remains true as stated for d-dimensional random vectors $U_j(l)$, if, in (3·1), $\mathbb{E}U_j^2$ is replaced by $\mathbb{E}|U_j|^2$.

As an example, take the following analogue of the Kubilius Main Theorem. Define $\mu_j := \theta j^{-1}\mathbb{E}U_j$ and $M(n) := \sum_{j=1}^{n}\mu_j$.

Theorem 3·3. *Suppose that a logarithmic combinatorial structure satisfies the Uniform Logarithmic Condition together with* (1·3)*, and that* (3·1) *holds. Then the sequence $\sigma^{-1}(n)(X^{(n)} - M(n))$ converges in distribution as $n \to \infty$ if and only if there is a*

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

distribution function K such that

$$\lim_{n\to\infty} \sigma^{-2}(n) \sum_{j=1}^{n} j^{-1} \mathbb{E}\big(U_j^2 I[U_j \leqslant x\sigma(n)\sqrt{\theta}]\big) = K(x) \tag{3·13}$$

for all continuity points x of K; the limit then has characteristic function ψ, where

$$\log \psi(t) = \int (e^{itx} - 1 - itx)\, x^{-2}\, K(dx).$$

Proof. The theorem follows because of the asymptotic equivalence of $\sigma^{-1}(n)X^{(n)}$ and $\sigma^{-1}(n)\widehat{X}^{(n)}$ of Theorem 3·2, together with [**17**, theorem 22·2A]. Writing $Y_j := \widehat{Z}_j U_j - \mu_j$, the necessary and sufficient condition for uniformly asymptotically negligible arrays in the above theorem is that

$$\lim_{n\to\infty} \sigma_1^{-2}(n) \sum_{j=1}^{n} \mathbb{E}\left\{Y_j^2 I[Y_j \leqslant x\sigma_1(n)]\right\} = K(x) \tag{3·14}$$

for all continuity points x of K, where

$$\sigma_1^2(n) := \sum_{j=1}^{n} \operatorname{Var} Y_j = \sum_{j=1}^{n} \theta j^{-1} \mathbb{E}U_j^2 - \sum_{j=1}^{n} \theta^2 j^{-2} (\mathbb{E}U_j)^2.$$

Note that

$$\zeta_n := \sigma^{-2}(n)|\sigma_1^2(n) - \theta\sigma^2(n)| \leqslant \sigma^{-2}(n) \sum_{j=1}^{n} \theta^2 j^{-2} \mathbb{E}U_j^2 = o(1) \tag{3·15}$$

as $n \to \infty$. It then follows from (3·1) that $\lim_{n\to\infty} \sigma^{-2}(n) \max_{1\leqslant j\leqslant n} \operatorname{Var} Y_j = 0$, since $\sigma^{-2}(n)\operatorname{Var} Y_n = 1 - \sigma^2(n-1)/\sigma^2(n) \to 0$ and $\sigma^2(n)$ is increasing in n; hence the random variables $\sigma_1^{-1}(n)Y_j$, $1 \leqslant j \leqslant m$, $m \geqslant 1$, indeed form a uniformly asymptotically negligible array.

To show the equivalence of (3·13) and (3·14), we start by writing

$$\begin{aligned}\mathbb{E}\left\{Y_j^2 I[Y_j \leqslant x\sigma_1(n)]\right\} &= \theta j^{-1} \mathbb{E}\{(U_j - \mu_j)^2 I[U_j \leqslant x\sigma_1(n) + \mu_j]\} \\ &\quad + (1 - \theta j^{-1})\mu_j^2 I[-\mu_j \leqslant x\sigma_1(n)].\end{aligned}$$

Now observe that

$$\sum_{j=1}^{n} \mu_j^2 I[-\mu_j \leqslant x\sigma_1(n)] \leqslant \sum_{j=1}^{n} (\theta j^{-1} \mathbb{E}U_j)^2 \leqslant \theta^2 \sum_{j=1}^{n} j^{-2} \mathbb{E}U_j^2 = o(\sigma^2(n)) \tag{3·16}$$

and that

$$\sum_{j=1}^{n} \mu_j \theta j^{-1} \mathbb{E}\{|U_j| I[U_j \leqslant x\sigma_1(n) + \mu_j]\} \leqslant \theta^2 \sum_{j=1}^{n} j^{-2} (\mathbb{E}U_j)^2 = o(\sigma^2(n))$$

also; hence

$$\begin{aligned}&\lim_{n\to\infty} \sigma_1^{-2}(n) \sum_{j=1}^{n} \mathbb{E}\left\{Y_j^2 I[Y_j \leqslant x\sigma_1(n)]\right\} \\ &\quad = \lim_{n\to\infty} \theta^{-1}\sigma^{-2}(n) \sum_{j=1}^{n} \theta j^{-1} \mathbb{E}\left\{U_j^2 I[U_j \leqslant x\sigma_1(n) + \mu_j]\right\}.\end{aligned}$$

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

Finally, for any $1 \leqslant n' \leqslant n$,

$$\begin{aligned}
&\sigma^{-2}(n) \sum_{j=n'}^{n} j^{-1} \mathbb{E}\{U_j^2 I[U_j \leqslant (x - \eta'_{n'})\sigma(n)\sqrt{\theta}]\} \\
&\quad \leqslant \sigma^{-2}(n) \sum_{j=n'}^{n} j^{-1} \mathbb{E}\left\{U_j^2 I[U_j \leqslant x\sigma_1(n) + \mu_j]\right\} \\
&\quad \leqslant \sigma^{-2}(n) \sum_{j=n'}^{n} j^{-1} \mathbb{E}\{U_j^2 I[U_j \leqslant (x + \eta'_{n'})\sigma(n)\sqrt{\theta}]\} \qquad (3{\cdot}17)
\end{aligned}$$

where

$$\eta'_l := \sup_{j \geqslant l}\{(|\mu_j|/\sigma(j)) + (\zeta_j/2\theta)\} \longrightarrow 0$$

as $l \to \infty$, from (3·15) and (3·16). The equivalence of the convergence in (3·13) and (3·14) at continuity points of K is now immediate.

The approximations in Theorems 3·2 and 3·3 both have process counterparts. Define $W^{(n)}$ and $\widetilde{X}^{(n)}$ for $t \in [0, 1]$ by

$$\widetilde{X}^{(n)}(t) := \sigma^{-1}(n) \sum_{j:\sigma^2(j) \leqslant t\sigma^2(n)} \left(I\left[C_j^{(n)} \geqslant 1\right] U_j\left(C_j^{(n)}\right) - \mu_j\right) \qquad (3{\cdot}18)$$

and

$$W^{(n)}(t) := \sigma^{-1}(n) \sum_{j:\sigma^2(j) \leqslant t\sigma^2(n)} (\widehat{Z}_j U_j - \mu_j). \qquad (3{\cdot}19)$$

Then it follows from Lemma 3·1 and (3·8)–(3·11) that

$$\mathbb{P}\left[\sup_{0 \leqslant t \leqslant 1} |\widetilde{X}^{(n)}(t) - W^{(n)}(t)| > \varepsilon\right] \longrightarrow 0$$

for each $\varepsilon > 0$, so that the whole process $\widetilde{X}^{(n)}$ is asymptotically equivalent to $W^{(n)}$, the normalized partial sum process for a sequence of independent random variables. In particular, if K is the distribution function of the degenerate distribution at 0, the limiting distribution of $\sigma^{-1}(n)X^{(n)}$ is standard normal, the analogue of the Erdős–Kac [**7**, **8**] Theorem, and $\widetilde{X}^{(n)}$ converges to standard Brownian motion. The special case $U_j(l) = l$ a.s. for all j, counting the total number of components, and its analogue which counts the number of distinct components, both come in this category, and we recover the functional central limit theorems of DeLaurentis and Pittel [**5**], Hansen [**10**–**12**], Arratia, Barbour and Tavaré [**2**] and Goh and Schmutz [**9**] as particular examples. The process version of Theorem 3·2 also carries over to d-dimensions.

4. *Regular growth*

In this section, we explore the consequences of replacing the slow growth of $\sigma^2(n)$ in (3·1) by regular variation:

$$\sigma^2(m) := \sum_{j=1}^{m} j^{-1} \mathbb{E}U_j^2 \text{ is regularly varying at } \infty, \quad \text{with exponent } \alpha > 0, \qquad (4{\cdot}1)$$

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

so that, in particular, $\sigma^2(b(n))/\sigma^2(n) \to 0$ for all sequences $b(n) = o(n)$ as $n \to \infty$. Our aim is to approximate $X^{(n)}$ by

$$Y^{*(n)} := \sum_{j=1}^{n} I\big[C_j^{*(n)} = 1\big] U_j, \tag{4·2}$$

which is a standard quantity, defined solely in terms of the U_j's and ESF(θ).

It is actually just as easy to prove a functional version of the approximation. Define the normalized (but not centred) process

$$\overline{X}^{(n)}(t) := \sigma^{-1}(n) \sum_{j:\sigma^2(j)\leqslant t\sigma^2(n)} I\big[C_j^{(n)} \geqslant 1\big] U_j\big(C_j^{(n)}\big), \quad 0 \leqslant t \leqslant 1, \tag{4·3}$$

and a process analogue $\overline{Y}^{*(n)}$ of $Y^{*(n)}$ by

$$\overline{Y}^{*(n)}(t) := \sigma^{-1}(n) \sum_{j:\sigma^2(j)\leqslant t\sigma^2(n)} I\big[C_j^{*(n)} = 1\big] U_j, \quad 0 \leqslant t \leqslant 1. \tag{4·4}$$

Then we have the following result.

THEOREM 4·1. *Suppose that a logarithmic combinatorial structure satisfies the Uniform Logarithmic Condition together with* (1·3), *and that* (4·1) *holds. Then if, for some sequence of centring functions* $\widetilde{M}_n$: $[0,1] \to \mathbb{R}$, *either of* $\mathcal{L}(\overline{Y}^{*(n)} - \widetilde{M}_n)$ *or* $\mathcal{L}(\overline{X}^{(n)} - \widetilde{M}_n)$ *converges, it follows that the other also converges, and to the same limit.*

Proof. The first step is to show that the small components play little part. From Proposition 1·1(2), it follows that

$$d_{TV}\big(\mathcal{L}\big(W'_{1b(n)}\big(C^{(n)}\big)\big), \mathcal{L}\big(W'_{1b(n)}(Z)\big)\big) = o(1) \text{ as } n \longrightarrow \infty, \tag{4·5}$$

where W' is as defined in (3·2), whenever $b(n) = o(n)$ as $n \to \infty$. Then, as in (3·9),

$$\sigma^{-1}(n) \left| W'_{1,b(n)}(Z) - \sum_{j=1}^{b(n)} I[Z_j = 1]|U_j| \right| \xrightarrow{\mathcal{D}} 0, \tag{4·6}$$

whatever the choice of $b(n)$. But now

$$\begin{aligned} &\mathrm{Var}\left(\sum_{j=1}^{b} I[Z_j = 1]|U_j|\right) \\ &= \mathbb{E}\left(\sum_{j=1}^{b} I[Z_j = 1]\mathrm{Var}\,|U_j|\right) + \mathrm{Var}\left(\sum_{j=1}^{b} I[Z_j = 1]\mathbb{E}|U_j|\right) \\ &\leqslant \theta(1+\varepsilon_{01}^*)\sum_{j=1}^{b} j^{-1}(\mathrm{Var}\,|U_j| + \{\mathbb{E}|U_j|\}^2) \qquad (4·7) \\ &= \theta(1+\varepsilon_{01}^*)\sigma^2(b) \qquad (4·8) \end{aligned}$$

and, from the Cauchy–Schwarz inequality as in (3·6),

$$\mathbb{E}\left\{\sum_{j=1}^{b} I[Z_j = 1]|U_j|\right\} \leqslant \{(1+\log b)\sigma^2(b)\}^{1/2} = O\big(\sigma(b)\log^{1/2} b\big). \tag{4·9}$$

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

Combining (4·5)–(4·9), it follows that

$$\sigma^{-1}(n)W'_{1,b(n)}\big(C^{(n)}\big)\big) \xrightarrow{\mathcal{D}} 0 \tag{4·10}$$

provided that $b(n) = O(n^{\zeta})$ for some $\zeta < 1$; and (4·8) and (4·9) then imply that

$$\sigma^{-1}(n)\sum_{j=1}^{b(n)} I\big[C_j^{*(n)} = 1\big]|U_j| \longrightarrow 0 \tag{4·11}$$

also.

Now, for the processes $\overline{X}^{(n)}$ of (4·3) and $\overline{Y}^{*(n)}$ of (4·4), the contributions from indices $j \leqslant b(n)$ are asymptotically negligible, by (4·10) and (4·11). Then, from Proposition 1·1(3), if $b(n) \to \infty$ as $n \to \infty$, it follows that

$$d_{TV}\big(\mathcal{L}\big(C^{(n)}[b(n)+1,n]\big), \mathcal{L}\big(C^{*(n)}[b(n)+1,n]\big)\big) = o(1) \text{ as } n \longrightarrow \infty, \tag{4·12}$$

whereas, from Lemma 6·9,

$$\begin{aligned} d_{TV}\big(\mathcal{L}\big(C^{*(n)}[b(n)+1,n]\big), \mathcal{L}\big(\{I\big[C_j^{*(n)} = 1\big],\, b(n)+1 \leqslant j \leqslant n\}\big)\big) \\ = o(1) \text{ as } n \longrightarrow \infty. \end{aligned} \tag{4·13}$$

Combining (4·10), (4·11), (4·12) and (4·13), it follows that $\overline{X}^{(n)}$ and $\overline{Y}^{*(n)}$ are asymptotically equivalent, as required.

Theorem 4·1 remains true in d-dimensions, if, in (4·1), $\mathbb{E}U_j^2$ is replaced by $\mathbb{E}|U_j|^2$.

Note, however, that $\overline{Y}^{*(n)}$ can only be expected to have a non-degenerate limit if

$$v(n) := \operatorname{Var} Y^{*(n)} \geqslant k\sigma^2(n) \tag{4·14}$$

for some $k > 0$ and for all n. This condition is satisfied if the random variables U_j are centred, or, more generally, if $\operatorname{Var} U_j \geqslant k'\mathbb{E}U_j^2$ for some $k' > 0$ and for all j, since

$$\begin{aligned} \operatorname{Var} Y^{*(n)} &= \operatorname{Var}\left(\sum_{j=1}^{n} I\big[C_j^{*(n)} = 1\big]U_j\right) \\ &= \mathbb{E}\left(\sum_{j=1}^{n} I\big[C_j^{*(n)} = 1\big]\operatorname{Var} U_j\right) + \operatorname{Var}\left(\sum_{j=1}^{n} I\big[C_j^{*(n)} = 1\big]\mathbb{E}U_j\right) \\ &\geqslant k' \sum_{j=\lfloor n/2\rfloor+1}^{\lfloor 3n/4\rfloor} \mathbb{P}\big[C_j^{*(n)} = 1\big]\mathbb{E}U_j^2 \\ &\geqslant k' c_{\{6\cdot10\}} \sum_{j=\lfloor n/2\rfloor+1}^{\lfloor 3n/4\rfloor} j^{-1}\mathbb{E}U_j^2 \geqslant k''\sigma^2(n), \end{aligned} \tag{4·15}$$

for suitable constants k' and k'', by Lemma 6·10 and from (4·1). On the other hand, the dependence between the random variables $C_j^{*(n)}$ can result in $v(n)$ being of smaller order than $\sigma^2(n)$. For instance, if $U_j(s) = sj$ a.s. for all j and s, then $\sigma^2(n) = \frac{1}{2}n(n+1)$ is regularly varying with exponent $\alpha = 2$, but $X^{(n)} - n$ is a.s. zero, and the distribution of $Y^{*(n)} - n \leqslant 0$ has a non-trivial limit. In such circumstances, the non-degenerate normalization for $X^{(n)}$ may not be $\sigma^{-1}(n)$, nor need $Y^{*(n)}$ be appropriate for describing its limiting behaviour.

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

Even when (4·14) holds, so that the asymptotics of $\overline{X}^{(n)}$ are the same as those of $\overline{Y}^{*(n)}$, the limit theory is complicated. For one thing, there is still the dependence between the $C_j^{*(n)}$, which leads to Poisson–Dirichlet approximations [**3**, theorem 3·3], rather than to approximations based on processes with independent increments. For instance, if $U_j = j^\alpha$ a.s. for $j \geqslant 1$, then

$$\overline{Y}^{*(n)}(t) \xrightarrow{\mathcal{D}} \alpha^{1/2} \sum_{j \geqslant 1} L_j^{\alpha/2} I\big[L_j \leqslant t^{1/\alpha}\big], \quad 0 < t \leqslant 1,$$

where $1 > L_1 > L_2 > \cdots$ are the points of a Poisson–Dirichlet process with parameter θ. But, even allowing for this, there is no universal approximation valid for a wide class of U_j sequences, as was the case with slow growth and the Gaussian approximations. For example, take the case in which $\mathbb{E}U_j^2 \sim cj^\alpha$ for some $\alpha > 0$. Then $\sigma^2(n) \sim c\alpha^{-1}n^\alpha$ is of the same order as $\mathbb{E}U_j^2$ for $n/2 < j \leqslant n$, and there is an asymptotically non-trivial probability that one such j will have $C_j^{*(n)} = 1$. Hence the distribution of the sum $Y^{*(n)}$ typically depends in detail on the distributions of the individual U_j's.

5. *Zhang's setting*

Zhang [**19**] proves theorems analogous to Theorems 2·1 and 3·3 for additive arithmetic semigroups under different conditions, specifying the asymptotic behaviour of the total number $G(n)$ of different elements of degree n. For instance, for his counterpart of Theorem 2·1 for additive arithmetic semigroups, he assumes (a little more than) that

$$\sum_{n \geqslant 1} |q^{-n}G(n) - Q(n)| < \infty, \tag{5·1}$$

where $Q(n) = \sum_{i=1}^r A_i n^{\rho_i - 1}$, with $\rho_1 < \rho_2 < \cdots < \rho_r \geqslant 1$ and $A_r > 0$. This condition does not necessarily imply that the Logarithmic Condition is satisfied. In our formulation, applying Theorem 2·1 to multisets, if $\theta_j(x) := jm_j x^j$, then we require that

$$\sum_{i \geqslant 1} i^{-1} \sup_{j \geqslant i} |\theta_j(x) - \theta| < \infty \quad \text{for some } \ 0 < x < 1, \tag{5·2}$$

without any more detailed specification of the exact form of the $\theta_i(x)$.

Translation between the two sorts of conditions is made possible by observing that, if $Z_j \sim \mathrm{NB}\,(m_j, x^j)$ for any $0 < x < 1$, then

$$\begin{aligned}\frac{m_n}{G(n)} &= \mathbb{P}\big[C_n^{(n)} = 1\big] = \mathbb{P}[Z_n = 1] \prod_{j=1}^{n-1} \mathbb{P}[Z_j = 0] / \mathbb{P}[T_{0n}(Z) = n] \\ &= m_n x^n \prod_{j=1}^{n} (1 - x^j)^{m_j} / \mathbb{P}[T_{0n}(Z) = n],\end{aligned}$$

where, here and subsequently, for $y \in \mathbb{Z}_+^n$ and $0 \leqslant r < s \leqslant n$,

$$T_{rs}(y) := \sum_{j=r+1}^{s} jy_j. \tag{5·3}$$

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

Hence, with $x = q^{-1}$, we have

$$G(n)q^{-n} = \mathbb{P}[T_{0n}(Z) = n] \left\{ \prod_{j=1}^{n} (1 - q^{-j})^{m_j} \right\}^{-1}. \tag{5·4}$$

From Theorem 6·1, it follows under (5·2) with $x = q^{-1}$ that

$$n\mathbb{P}[T_{0n}(Z) = n] \sim \theta \mathbb{P}[X_\theta \leqslant 1] > 0,$$

with X_θ as in (6·1), and that

$$\prod_{j=1}^{n} (1 - q^{-j})^{m_j} \sim k \exp\left\{ -\theta \sum_{j=1}^{n} j^{-1} \right\}$$

for some constant k. This then implies that $G(n)q^{-n} \sim k' n^{\theta-1}$, and comparison with the definition of $Q(n)$ in (5·1) identifies ρ_r with θ in cases where both (5·1) and (5·2) are satisfied. Hence, since Zhang assumes that $\rho_r \geqslant 1$ for his counterpart of Theorem 2·1 and $\rho_r = 1$ for that of Theorem 3·3, his theorems require $\theta \geqslant 1$ and $\theta = 1$ respectively, if both (5·1) and (5·2) are satisfied; our conditions impose no restriction on θ, but demand the extra regularity inherent in (5·2). In fact, Zhang [**20**, theorem 1·3] implies that the Uniform Logarithmic Condition and (1·3) both hold with $\theta = \rho_r$ if $\rho_r < 1$ and

$$G(n)q^{-n} = Q(n) + O(n^{-\gamma}) \quad \text{for some } \gamma > 3; \tag{5·5}$$

as a consequence, the basic conditions on the combinatorial structure required for Theorems 2·1 and 3·3 are automatically fulfilled if $\rho_r < 1$ and (5·5) is satisfied.

A more precise description of the values $G(n)$ implied by (5·2) can be derived using size-biassing as in [**1**], giving

$$\begin{aligned} n\mathbb{P}[T_{0n}(Z) = n] &= \sum_{j=1}^{n} g(j)\mathbb{P}[T_{0n}(Z) = n - j] \\ &= \sum_{j=1}^{n} g(j)\mathbb{P}[T_{0,n-j} = n - j] \prod_{l=n-j+1}^{n} (1 - x^l)^{m_l}, \end{aligned}$$

for

$$g(j) := x^j \sum_{l=1:l|j}^{j} l m_l = \theta_j(x) + O\left(x^{j/2}\right) \sim \theta,$$

where $\mathbb{P}[T_{00} = 0]$ is interpreted as 1. This, with (5·4), implies that

$$F(n) = n^{-1} \sum_{j=1}^{n} g(j) F(n - j), \tag{5·6}$$

where $F(n) := G(n)x^n$, $n \geqslant 1$, and $F(0) = 1$. Equation (5·6) gives a recursive formula for $F(n)$, and hence for $G(n)$, in terms of the values of $g(j)$, $1 \leqslant j \leqslant n$, and of $F(j)$, $0 \leqslant j < n$; it also enables generating function methods, such as singularity theory (Odlyzko [**18**, theorem 11·4]), to be applied, in order to deduce properties of the $g(j)$ from those of $G(n)$. Equation (5·6) is at the heart of Zhang's method; under his conditions on the $G(n)$, the solutions $g(j)$ can have non-trivial oscillations

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

[**20**, theorem 1·3] if his $\rho_r \geqslant 1$, in which case the Logarithmic Condition is not satisfied; hence his results cover cases not included in Theorems 2·1 and 3·3.

6. *Details*

In this section, we collect some technical results that were needed in the previous sections. The first of these is essentially theorem 2·6 of [**4**], with the statement adapted so as to give a uniform bound valid for all the processes $C^{(b,n)}$ of the form introduced in the proof of Lemma 2·3. The proof runs exactly as for the original theorem, with precisely the same error bound, since replacing any of the Z_j by the corresponding $Z_j^* \sim \mathrm{Po}\,(\theta/j)$ merely removes terms which would otherwise contribute to the error. The random variable X_θ appearing in the statement of the theorem has density p_θ satisfying

$$p_\theta(x) = \frac{e^{-\gamma\theta}x^{\theta-1}}{\Gamma(\theta)},\ 0 \leqslant x \leqslant 1; \quad \frac{d}{dx}\{x^{1-\theta}p_\theta(x)\} = -\theta x^{-\theta}p_\theta(x-1),\ x > 1,$$

and is such that

$$n^{-1}T_{0n}(Z^*) \xrightarrow{\mathcal{D}} X_\theta, \tag{6·1}$$

where we recall the definition (5·3) of T_{0n}.

THEOREM 6·1. *If the ULC holds and* $\lim_{m\to\infty} m^{-1}B_m = 0$, *then*

$$\max_{0\leqslant v\leqslant B_m}\ \max_{0\leqslant b\leqslant m}\ \sup_{s\geqslant 1} |s\mathbb{P}[T_{vm}(\widehat{Z}^b) = s] - \theta\mathbb{P}[m^{-1}(s-m) \leqslant X_\theta < m^{-1}(s-v)]| \longrightarrow 0$$

as $m \to \infty$, *where* $\widehat{Z}^b$ *is as defined in* (2·10).

We use the following three direct consequences of this theorem.

COROLLARY 6·2. *If the ULC holds, then*

$$\lim_{n\to\infty} \max_{0\leqslant b\leqslant n} |\mathbb{P}[T_{0n}(\widehat{Z}^b) = n]/\mathbb{P}[T_{0n}(Z) = n] - 1| = 0.$$

COROLLARY 6·3. *If the ULC holds and* $\lim_{n\to\infty} n^{-1}b(n) = 0$, *then*

$$\lim_{n\to\infty} \max_{n/2\leqslant s\leqslant n} |\mathbb{P}\big[T_{b(n),n}\big(\widehat{Z}^{b(n)}\big) = s\big]/\mathbb{P}\big[T_{b(n),n}(Z) = s\big] - 1| = 0.$$

COROLLARY 6·4. *If the ULC holds, then there exist* $0 < k_- < k_+ < \infty$ *such that, for all* $n/2 \leqslant s \leqslant n$ *and all* $0 \leqslant b \leqslant n$,

$$n^{-1}k_- \leqslant \mathbb{P}[T_{0n}(\widehat{Z}^b) = s] \leqslant n^{-1}k_+.$$

The fourth corollary is almost the same as Proposition 1·1(2); the proof is as for [**4**, theorem 3·1].

COROLLARY 6·5. *If the ULC holds and* $\lim_{n\to\infty} n^{-1}b(n) = 0$, *then*

$$\lim_{n\to\infty} d_{TV}\big(\mathcal{L}\big(C^{(b(n),n)}[1,b(n)]\big), \mathcal{L}(Z[1,b(n)])\big) = 0.$$

The next lemma makes use of this last result.

LEMMA 6·6. *If the ULC and* (1·3) *hold, and if* $n^{-1}b(n) \to 0$, *then*

$$\lim_{n\to\infty} n^{-1}\mathbb{E}T_{0,b(n)}\big(C^{(b(n),n)}\big) = 0.$$

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

Proof. We bound the expression $n^{-1}\sum_{l\geqslant 1} l\mathbb{P}[T_{0,b(n)}(C^{(b(n),n)}) = l]$ by considering the cases $l \leqslant n/2$ and $l > n/2$ separately. In the latter range, we have

$$n^{-1}\sum_{l=\lfloor n/2\rfloor+1}^{n} l\mathbb{P}\big[T_{0,b(n)}\big(C^{(b(n),n)}\big) = l\big] \leqslant \mathbb{P}\big[T_{0,b(n)}\big(C^{(b(n),n)}\big) > n/2\big] \tag{6·2}$$
$$\leqslant \mathbb{P}\big[T_{0,b(n)}(Z) > n/2\big] + d_{TV}\left(\mathcal{L}\big(C^{(b(n),n)}[1,b(n)]\big), \mathcal{L}(Z[1,b(n)])\right).$$

But now, from the Uniform Logarithmic Condition,

$$\mathbb{P}\big[T_{0,b(n)}(Z) > n/2\big] \leqslant 2n^{-1}\mathbb{E}T_{0,b(n)}(Z)$$
$$\leqslant 2(b(n)/n)\left\{\theta + (c_1 + D_1)\{b(n)\}^{-1}\sum_{j=1}^{b(n)} e(j)\right\} \longrightarrow 0 \tag{6·3}$$

as $n \to \infty$, and $d_{TV}(\mathcal{L}(C^{(b(n),n)}[1,b(n)]), \mathcal{L}(Z[1,b(n)])) \to 0$ by Corollary 6·5.

In the former range, we have

$$n^{-1}\sum_{l=1}^{\lfloor n/2\rfloor} l\mathbb{P}\big[T_{0,b(n)}\big(C^{(b(n),n)}\big) = l\big]$$
$$= n^{-1}\sum_{l=1}^{\lfloor n/2\rfloor} \frac{l\mathbb{P}\big[T_{0,b(n)}(Z) = l\big]\mathbb{P}\big[T_{b(n),n}\big(\widehat{Z}^{b(n)}\big) = n-l\big]}{\mathbb{P}\big[T_{0n}\big(\widehat{Z}^{b(n)}\big) = n\big]},$$

and, since $n/2 \leqslant n - l \leqslant n$, we can apply Corollaries 6·3 and 6·4 to conclude that

$$n^{-1}\sum_{l=1}^{\lfloor n/2\rfloor} l\mathbb{P}\big[T_{0,b(n)}\big(C^{(b(n),n)}\big) = l\big]$$
$$\leqslant k_3 n^{-1}\sum_{l=1}^{\lfloor n/2\rfloor} l\mathbb{P}\big[T_{0,b(n)}(Z) = l\big] \leqslant k_3 n^{-1}\mathbb{E}T_{0,b(n)}(Z),$$

for some $k_3 < \infty$. Convergence to zero now follows from (6·3), and the lemma is proved.

The next result is rather more complicated; it expresses the fact that replacing $C^{(n)}$ by $C^{(b,n)}$ makes little difference throughout.

THEOREM 6·7. *If the ULC and* (1·3) *hold,* $b(n) \to \infty$ *and* $n^{-1}b(n) \to 0$, *then*

$$\lim_{n\to\infty} d_{TV}\left(\mathcal{L}\big(C^{(n)}\big), \mathcal{L}\big(C^{(b(n),n)}\big)\right) = 0.$$

Proof. Writing $c \coloneqq (c_1, \ldots, c_n)$ for the generic element of $\mathbb{Z}_+^n$, we first observe that

$$\sum_{\substack{c:T_{0n}(c)=n\\ T_{0,b(n)}(c)>n/2}} \mathbb{P}\big[C^{(b(n),n)} = c\big] = \mathbb{P}\big[T_{0,b(n)}\big(C^{(b(n),n)}\big) > n/2\big]$$
$$\leqslant 2n^{-1}\mathbb{E}T_{0,b(n)}\big(C^{(b(n),n)}\big) \longrightarrow 0,$$

by Lemma 6·6; in similar fashion, $\mathbb{P}[T_{0,b(n)}(C^{(n)}) > n/2] \to 0$ also. Hence, when comparing $\mathcal{L}(C^{(n)})$ with $\mathcal{L}(C^{(b(n),n)})$, it is enough to look at c such that $T_{0,b(n)}(c) \leqslant n/2$.

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

We thus turn to bounding

$$\sum_{\substack{c:T_{0n}(c)=n \\ T_{0,b(n)}(c)\leqslant n/2}} \left| \mathbb{P}\left[C^{(b(n),n)}=c\right] - \mathbb{P}\left[C^{(n)}=c\right] \right|. \tag{6·4}$$

Now, for any c with $T_{0n}(c)=n$, it follows from the Logarithmic Condition that

$$\mathbb{P}\left[C^{(n)}=c\right] = \mathbb{P}[Z[1,n]=c]/\mathbb{P}[T_{0n}(Z)=n]. \tag{6·5}$$

Then, using Corollary 6·2, the denominator in (6·5) can be replaced by the quantity $\mathbb{P}[T_{0n}(\widehat{Z}^{b(n)})=n]$ for use in (6·4) with only small error, since

$$\lim_{n\to\infty} \sum_{c:T_{0n}(c)=n} \mathbb{P}[Z[1,n]=c] \left| \frac{1}{\mathbb{P}[T_{0n}(Z)=n]} - \frac{1}{\mathbb{P}\left[T_{0n}\left(\widehat{Z}^{b(n)}\right)=n\right]} \right| = 0.$$

Also, writing $l = T_{0,b(n)}(c) \leqslant n/2$, we have

$$\begin{aligned}\mathbb{P}[Z[1,n]=c] = \mathbb{P}[Z[1,b(n)]=c[1,b(n)]]\mathbb{P}\left[T_{b(n),n}(Z)=n-l\right] \\ \times \mathbb{P}\left[Z[b(n)+1,n]=c[b(n)+1,n] \,|\, T_{b(n),n}(Z)=n-l\right],\end{aligned}$$

and the factor $\mathbb{P}[T_{b(n),n}(Z)=n-l]$ can be replaced by $\mathbb{P}[T_{b(n),n}(\widehat{Z}^{b(n)})=n-l]$ for use in (6·4) with only small error, since

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{\mathbb{P}[T_{0n}(\widehat{Z}^{b(n)})=n]} \sum_{\substack{c:T_{0n}(c)=n \\ T_{0,b(n)}(c)\leqslant n/2}} \mathbb{P}[Z[1,n]=c] \\ \times \left| 1 - \frac{\mathbb{P}[T_{b(n),n}(\widehat{Z}^{b(n)})=n-T_{0,b(n)}(c)]}{\mathbb{P}[T_{b(n),n}(Z)=n-T_{0,b(n)}(c)]} \right| = 0\end{aligned}$$

by Corollaries 6·3 and 6·2. Thus, to show that (6·4) is asymptotically small, it is enough to examine

$$\sum_{\substack{c:T_{0n}(c)=n \\ T_{0,b(n)}(c)\leqslant n/2}} \left| \mathbb{P}\left[C^{(b(n),n)}=c\right] - \frac{\mathbb{P}[Z[1,n]=c]\mathbb{P}\left[T_{b(n),n}\left(\widehat{Z}^{b(n)}\right)=n-T_{0,b(n)}(c)\right]}{\mathbb{P}\left[T_{0n}\left(\widehat{Z}^{b(n)}\right)=n\right]\mathbb{P}\left[T_{b(n),n}(Z)=n-T_{0,b(n)}(c)\right]} \right|.$$

Dissecting the formula for $\mathbb{P}[C^{(b(n),n)}=c]$ arising from the Logarithmic Condition, this is just

$$\begin{aligned}&\frac{1}{\mathbb{P}[T_{0n}\left(\widehat{Z}^{b(n)}\right)=n]} \sum_{l=0}^{\lfloor n/2 \rfloor} \mathbb{P}\left[T_{0,b(n)}(Z)=l\right]\mathbb{P}\left[T_{b(n),n}\left(\widehat{Z}^{b(n)}\right)=n-l\right] \\ &\quad\times \sum_{\substack{c:T_{0,b(n)}(c)=l \\ T_{b(n),n}(c)=n-l}} \mathbb{P}\left[Z[1,b(n)]=c[1,b(n)] \,|\, T_{0,b(n)}(Z)=l\right] \\ &\quad\times \left| \frac{\mathbb{P}\left[Z[b(n)+1,n]=c[b(n)+1,n]\right]}{\mathbb{P}\left[T_{b(n),n}(Z)=n-l\right]} - \frac{\mathbb{P}\left[\widehat{Z}^{b(n)}[b(n)+1,n]=c[b(n)+1,n]\right]}{\mathbb{P}\left[T_{b(n),n}\left(\widehat{Z}^{b(n)}\right)=n-l\right]} \right|.\end{aligned} \tag{6·6}$$

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

However, from [**4**, theorem 3·3], if (1·3) is satisfied,

$$\lim_{n\to\infty} \max_{n/2\leqslant s\leqslant n} d_{TV}\big(\mathcal{L}\big(C^{(n)}[b(n)+1,n]\,|\,T_{b(n),n}\big(C^{(n)}\big)=s\big),\\ \mathcal{L}\big(C^{*(n)}[b(n)+1,n]\,|\,T_{b(n),n}\big(C^{*(n)}\big)=s\big)\big)\longrightarrow 0, \quad (6\cdot7)$$

provided that $b(n)\to\infty$; and then

$$\frac{\mathbb{P}\big[C^{(n)}[b(n)+1,n]=c[b(n)+1,n]\big]}{\mathbb{P}\big[T_{b(n),n}\big(C^{(n)}\big)=s\big]}=\frac{\mathbb{P}[Z[b(n)+1,n]=c[b(n)+1,n]]}{\mathbb{P}\big[T_{b(n),n}(Z)=s\big]},$$

and the same equality is true if $C^{(n)}$ is replaced by $C^{*(n)}$ and Z by $\widehat{Z}^{b(n)}$. Hence (6·7) implies that

$$\sum{}^{(l)}\left|\frac{\mathbb{P}[Z[b(n)+1,n]=c[b(n)+1,n]]}{\mathbb{P}[T_{b(n),n}(Z)=n-l]}\right.\\ \left.-\frac{\mathbb{P}\big[\widehat{Z}^{b(n)}[b(n)+1,n]=c[b(n)+1,n]\big]}{\mathbb{P}\big[T_{b(n),n}\big(\widehat{Z}^{b(n)}\big)=n-l\big]}\right|\leqslant\eta(n)\longrightarrow 0,$$

uniformly in $0\leqslant l\leqslant n/2$, where $\sum^{(l)}$ denotes a sum over all $c_{b(n)+1},\dots,c_n$ such that $\sum_{j=b(n)+1}^{n} jc_j=n-l$. Substituting this into (6·6), we have at most

$$\frac{1}{\mathbb{P}\big[T_{0n}\big(\widehat{Z}^{b(n)}\big)=n\big]}\sum_{l=0}^{\lfloor n/2\rfloor}\mathbb{P}\big[T_{0,b(n)}(Z)=l\big]\mathbb{P}\big[T_{b(n),n}\big(\widehat{Z}^{b(n)}\big)=n-l\big]\\ \times\sum_{c[1,b(n)]:\sum_{j=1}^{b(n)}jc_j=l}\mathbb{P}\big[Z[1,b(n)]=c[1,b(n)]\,|\,T_{0,b(n)}(Z)=l\big]\,\eta(n)\\ \leqslant\eta(n)\longrightarrow 0,$$

and the theorem is proved.

In addition, we need some estimates connected with the probabilities $\mathbb{P}[C_j^{(b,n)}=l]$ for $b+1\leqslant j\leqslant n$ and for $l\geqslant 1$; these are, not surprisingly, much the same as the corresponding bounds for $\mathbb{P}[C_j^{*(n)}=l]$. In order to establish these, we first need an upper bound for the probability in Corollary 6·4 which is valid for all $0\leqslant s\leqslant n$.

Lemma 6·8. *If the ULC and* (1·3) *hold, then there exists* $c_{\{6\cdot8\}}<\infty$ *such that*

$$\max_{0\leqslant b\leqslant n}\mathbb{P}[T_{0n}(\widehat{Z}^b)=s]\leqslant c_{\{6\cdot8\}}n^{-\theta}(s+1)^{-(1-\theta)},$$

for all $0\leqslant s\leqslant n$.

Proof. By independence, it is immediate that

$$\mathbb{P}[T_{0n}(\widehat{Z}^b)=s]=\mathbb{P}[T_{0s}(\widehat{Z}^b)=s]\prod_{j=s+1}^{n}\mathbb{P}\big[\widehat{Z}_j^b=0\big], \quad (6\cdot8)$$

with $\mathbb{P}[T_{0s}(\widehat{Z}^b)=s]$ taken to be 1 if $s=0$. Now $\mathbb{P}[Z_j^*=0]=e^{-\theta/j}$, whereas, by the Uniform Logarithmic Condition,

$$\mathbb{P}[Z_j=0]\leqslant 1-j^{-1}\theta+c_1j^{-1}e(j)\leqslant\exp\{-j^{-1}\theta+c_1j^{-1}e(j)\};$$

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

hence, whatever the value of b, we have

$$\mathbb{P}[T_{0n}(\widehat{Z}^b) = s] \leqslant k_1 \exp\left\{-\theta \sum_{j=s+1}^{n} j^{-1}\right\} \mathbb{P}[T_{0s}(\widehat{Z}^b) = s],$$

with $k_1 := \exp\{c_1 \sum_{j \geqslant 1} j^{-1} e(j)\}$. Furthermore, for $s \geqslant 1$,

$$s\mathbb{P}[T_{0s}(Z^*) = s] = \theta \mathbb{P}[T_{0s}(Z^*) < s] \to \theta \mathbb{P}[X_\theta < 1] \quad \text{as } s \to \infty, \tag{6·9}$$

from (6·1) and the special properties of the compound Poisson random variable $T_{0s}(Z^*)$. Hence, because of Corollary 6·2, and remembering also the case $s = 0$, it follows that

$$\mathbb{P}[T_{0n}(\widehat{Z}^b) = s] \leqslant k_1'(s+1)^{-1} \exp\left\{-\theta \sum_{j=s+1}^{n} j^{-1}\right\}$$

for some other constant k_1'. The asymptotics of the harmonic series now complete the proof.

LEMMA 6·9. *If the Uniform Logarithmic Condition and* (1·3) *hold, then there exist constants* $c_{\{6\cdot 9a\}}$ *and* $c_{\{6\cdot 9b\}}$ *such that, for any* $0 \leqslant b \leqslant n$,

$$j\mathbb{P}\big[C_j^{(b,n)} = 1\big] \leqslant c_{\{6\cdot 9a\}} \left(\frac{n}{n-j+1}\right)^{1-\theta}, \quad b+1 \leqslant j \leqslant n;$$

$$\mathbb{P}\left[\bigcup_{j=b+1}^{n} \{C_j^{(b,n)} \geqslant 2\}\right] \leqslant c_{\{6\cdot 9b\}} b^{-1}.$$

Proof. Since $\widehat{Z}_j^b = Z_j^* \sim \mathrm{Po}\,(\theta/j)$ for all $j \geqslant b+1$, it follows that, for such j,

$$\begin{aligned}
\mathbb{P}\big[C_j^{(b,n)} = l\big] &= \frac{\mathbb{P}\big[\widehat{Z}_j^b = l\big]\mathbb{P}\big[T_{0n}(\widehat{Z}^b) - j\widehat{Z}_j^b = n - jl\big]}{\mathbb{P}[T_{0n}(\widehat{Z}^b) = n]} \\
&\leqslant \frac{\mathbb{P}[Z_j^* = l]\mathbb{P}[T_{0n}(\widehat{Z}^b) = n - jl]}{\mathbb{P}[Z_j^* = 0]\mathbb{P}[T_{0n}(\widehat{Z}^b) = n]} \\
&= \frac{1}{l!}\left(\frac{\theta}{j}\right)^l \frac{\mathbb{P}[T_{0n}(\widehat{Z}^b) = n - jl]}{\mathbb{P}[T_{0n}(\widehat{Z}^b) = n]}.
\end{aligned}$$

Hence, applying Corollary 6·2, (6·9) and Lemma 6·8, it follows that there is a constant k_2 such that

$$\mathbb{P}\big[C_j^{(b,n)} = l\big] \leqslant \frac{1}{l!}\left(\frac{\theta}{j}\right)^l k_2 \left(\frac{n+1}{n-jl+1}\right)^{1-\theta}.$$

The first part of the lemma is now immediate.

For the second part, we just need to bound the sum

$$\sum_{j=b+1}^{\lfloor n/2 \rfloor} \sum_{l=2}^{\lfloor n/j \rfloor} \frac{1}{l!}\left(\frac{\theta}{j}\right)^l \left(\frac{n+1}{n-jl+1}\right)^{1-\theta},$$

since $\mathbb{P}[C_j^{(b,n)} = l] = 0$ outside the given ranges of $l \geqslant 2$ and j. For $\theta \geqslant 1$, a bound of order $O(b^{-1})$ is easy. For $\theta < 1$, swap the order of the j and l summations, and

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

then consider the ranges $b+1 \leqslant j \leqslant \lfloor n/2l \rfloor$ and $\lfloor n/2l \rfloor < j \leqslant \lfloor n/l \rfloor$ separately. In the first of these ranges, the final factor is at most $2^{1-\theta}$, giving an upper bound for the sum of $\theta^l 2^{1-\theta} b^{-(l-1)}/\{l!(l-1)\}$; adding over $l \geqslant 2$ thus gives a contribution of order $O(b^{-1})$. In the second j range, we have

$$(1/l!)(\theta/j)^l \leqslant (e/l)^l (2l\theta/n)^l \leqslant (2e\theta/n)^l,$$

while the j sum of the final factor is bounded above by

$$\theta^{-1}(n+1)^{1-\theta}\{1+(1+n/2)^{\theta}\} = O(n);$$

adding over $l \geqslant 2$ gives a contribution of order $O(n^{-1})$, uniformly in $n \geqslant 3e\theta$, and smaller values of n can at worst increase the constant implied by the order symbol. This proves the second part of the lemma.

The final result gives a simple lower bound for $j\mathbb{P}[C_j^{*(n)} = 1]$, valid in $n/2 < j \leqslant 3n/4$.

Lemma 6·10. *There exists a constant $c_{\{6\cdot 10\}} > 0$ such that*

$$j\mathbb{P}\big[C_j^{*(n)} = 1\big] \geqslant c_{\{6\cdot 10\}} \quad \textit{for all} \quad n/2 < j \leqslant 3n/4.$$

Proof. Clearly, in this range of j, $C_j^{*(n)}$ can only take the values 0 or 1. Hence, using the Feller coupling,

$$\mathbb{P}\big[C_j^{*(n)} = 1\big] \geqslant \frac{\theta\Gamma(n)\Gamma(n-j+\theta)}{\Gamma(n-j+1)\Gamma(n+\theta)} \geqslant n^{-1}4^{-\lceil\theta\rceil},$$

which is enough.

REFERENCES

[1] R. Arratia and S. Tavaré. Independent process approximations for random combinatorial structures. *Adv. Math.* **104** (1994), 90–154.

[2] R. Arratia, A. D. Barbour and S. Tavaré. On random polynomials over finite fields. *Math. Proc. Cam. Phil. Soc.* **114** (1993), 347–368.

[3] R. Arratia, A. D. Barbour and S. Tavaré. On Poisson–Dirichlet limits for random decomposable combinatorial structures. *Combin. Probab. Comput.* **8** (1999), 193–208.

[4] R. Arratia, A. D. Barbour and S. Tavaré. Limits of logarithmic combinatorial structures. *Ann. Probab.* **28** (2000), 1620–1644.

[5] J. M. DeLaurentis and B. G. Pittel. Random permutations and Brownian motion. *Pacific J. Math.* **119** (1985), 287–301.

[6] R. M. Dudley. *Probabilities and metrics.* Lecture Notes Series **45**, Aarhus Universitet (1976).

[7] P. Erdős and M. Kac. On the Gaussian law of errors in the theory of additive functions. *Proc. Amer. Math. Soc.* **25** (1939), 206–207.

[8] P. Erdős and M. Kac. The Gaussian law of errors in the theory of additive number-theoretic functions. *Amer. J. Math.* **62** (1940), 738–742.

[9] W. M. Y. Goh and E. Schmutz. Random matrices and Brownian motion. *Combin. Probab. Comput.* **2** (1993), 157–180.

[10] J. C. Hansen. A functional central limit theorem for random mappings. *Ann. Probab.* **17** (1989), 317–332.

[11] J. C. Hansen. A functional central limit theorem for the Ewens Sampling Formula. *J. Appl. Probab.* **27** (1990), 28–43.

[12] J. C. Hansen. Factorization in $F_q[x]$ and Brownian motion. *Combin. Probab. Comput.* **2** (1993), 285–299.

[13] J. Knopfmacher. *Analytic arithmetic of algebraic function fields.* Lecture Notes in Pure and Applied Mathematics **50** (Marcel Dekker, 1979).

[14] J. Knopfmacher and W.-B. Zhang. *Number Theory Arising from Finite Fields: Analytic and Probabilistic Theory.* Lecture Notes in Pure and Applied Mathematics **241** (Marcel Dekker, 2001).

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press

[**15**] J. Kubilius. Probabilistic methods in the theory of numbers. *Uspekhi Mat. Nauk* **11** (1956), 31–66; *Amer. Math. Soc. Transl.* **19** (1962), 47–85.
[**16**] J. Kubilius. Probabilistic methods in the theory of numbers. *Amer. Math. Soc. Transl. Math. Monographs* **11** (1964).
[**17**] M. Loève. *Probability Theory* Vol. I (Springer, 1977).
[**18**] A. M. Odlyzko. Asymptotic enumeration methods. In *Handbook of Combinatorics* (Elsevier, 1995), 1063–1229.
[**19**] W.-B. Zhang. Probabilistic number theory in additive arithmetic semigroups. In: *Analytic Number Theory: Proceedings of a Conference in Honor of Heini Halberstam*, Eds. B. C. Berndt, H. G. Diamond and A. J. Hildebrand. *Progr. Math.* **139** (1996).
[**20**] W.-B. Zhang. The prime element theorem in additive arithmetic semigroups, II. *Illinois J. Math.* **42** (1998), 198–229.

https://doi.org/10.1017/S0305004105008583 Published online by Cambridge University Press