

LIMITS OF LOGARITHMIC COMBINATORIAL STRUCTURES

BY R. ARRATIA,¹ A. D. BARBOUR² AND S. TAVARÉ¹

University of Southern California, Universität Zürich and
University of Southern California

Under very mild conditions, we prove that the limiting behavior of the component counts in a decomposable logarithmic combinatorial structure conforms to a single, unified pattern, which includes functional central limit theorems, Erdős–Turán laws, Poisson–Dirichlet limits for the large components and Poisson approximation in total variation for the total number of components. Our approach is entirely probabilistic, and the conditions can readily be verified in practice.

1. Introduction. Many well-known combinatorial structures (for example, permutations, polynomials over a finite field) are made up of elementary components (cycles, irreducible factors) of different sizes. This paper is concerned with the common statistical properties of the numbers of components of different sizes in a structure chosen uniformly at random from all those of size n , in the limit as $n \rightarrow \infty$. The structures we consider have two common properties. First, if the number of components of size i in a structure of size n is denoted by $C_i^{(n)}$, $1 \leq i \leq n$, then the joint distribution of $(C_1^{(n)}, \dots, C_n^{(n)})$ satisfies the *conditioning relation*,

$$(1.1) \quad \begin{aligned} \text{(CR): } & \mathbb{P}[C_n^{(n)} = c_1, \dots, C_n^{(n)} = c_n] \\ &= \mathbb{P}\left[Z_1 = c_1, \dots, Z_n = c_n \mid \sum_{i=1}^n iZ_i = n\right] \end{aligned}$$

for some sequence $(Z_i, i \geq 1)$ of independent random variables on \mathbb{Z}_+ . The requirement that $\sum_{i=1}^n iC_i^{(n)} = n$ simply reflects the fact that the size of the structure is n . Second, the Z_i have distributions satisfying the *logarithmic condition*

$$(1.2) \quad \text{(LC): } \lim_{i \rightarrow \infty} i\mathbb{P}[Z_i = 1] = \lim_{i \rightarrow \infty} i\mathbb{E}Z_i = \theta$$

for some $\theta > 0$. Further examples of such structures are square free polynomials, necklaces, mappings, mapping patterns and characteristic polynomials; certain nonuniform measures, such as the Ewens sampling formula [Ewens (1972)], are also covered. The most commonly studied examples belong to one of three classes, in which the Z_i 's all have Poisson distributions (assemblies), negative binomial distributions (multisets) or binomial distributions (selections); of those mentioned above, random characteristic polynomials are the

Received September 1998, revised June 2000.

¹Supported in part by NSF Grant DMS-96-26412.

²Supported in part by Schweizerischer NF Projekt 20-50686.97.

AMS 1991 subject classification. 20B25.

only exception. The Ewens sampling formula, which we denote by $C^{*(n)}$, has $Z_i^* \sim \text{Po}(\theta/i)$, the case $\theta = 1$ resulting from the uniform distribution over permutations.

Limit theorems for various aspects of the component structure have been proved separately for different instances of logarithmic combinatorial structures. A Poisson–Dirichlet limit for the normalized sizes of the r largest cycles in a uniform random permutation was proved by Vershik and Shmidt (1977), for components of a random mapping by Aldous (1985), for components of a random mapping pattern by Mutafciiev (1990) and for all assemblies and multisets satisfying a certain complex analytic condition by Hansen (1994); a functional central limit theorem for the component sizes was proved for uniform random permutations by De Laurentis and Pittel (1985), for mappings by Hansen (1989), for the Ewens sampling formula by Hansen (1990), for polynomials by Arratia, Barbour and Tavaré (1993) and by Hansen (1993) and for characteristic polynomials by Goh and Schmutz (1993) and Hansen and Schmutz (1993). Here, we show that separate treatment of the individual instances is unnecessary. These and other limit theorems hold for *all* structures which satisfy the conditioning relation and very slight strengthenings of the logarithmic condition.

The conditioning relation is used in the following way. For any $y \in \mathbb{Z}_+^n$, we define

$$(1.3) \quad y[r, s] = (y_r, y_{r+1}, \dots, y_s); \quad K_{rs}(y) = \sum_{i=r+1}^s y_i; \quad T_{rs}(y) = \sum_{i=r+1}^s iy_i.$$

Then the conditioning relation implies that

$$(1.4) \quad \mathbb{P}[C^{(n)}[1, b] = y[1, b]] = \frac{\mathbb{P}[Z[1, b] = y[1, b]]\mathbb{P}[T_{bn}(Z) = n - T_{0b}(y)]}{\mathbb{P}[T_{0n}(Z) = n]}$$

and

$$(1.5) \quad \begin{aligned} & \mathbb{P}[C^{(n)}[b + 1, n] = y[b + 1, n]] \\ &= \frac{\mathbb{P}[Z[b + 1, n] = y[b + 1, n]]\mathbb{P}[T_{0b}(Z) = n - T_{bn}(y)]}{\mathbb{P}[T_{0n}(Z) = n]}. \end{aligned}$$

Thus probabilities for the dependent sequence $C^{(n)}$ are replaced by probabilities only involving the independent sequence $(Z_i, i \geq 1)$, at the cost of needing to handle point probabilities for the random variables $T_{vm}(Z)$, for appropriate choices of v and m . In (1.4), we show that if $b \ll n$ and $T_{0b}(y) \ll n$, then

$$(1.6) \quad \mathbb{P}[T_{bn}(Z) = n - T_{0b}(y)] \approx \mathbb{P}[T_{0n}(Z) = n],$$

implying that the distributions of $C^{(n)}[1, b]$ and $Z[1, b]$ are close: the counts of small components are almost independent. For the large components, this is clearly not the case. Instead, we divide (1.5) by the same equation, but with

$C^{*(n)}$ for $C^{(n)}$ and Z^* for Z , and show that if also $b \gg 1$ then

$$(1.7) \quad \begin{aligned} \mathbb{P}[T_{0n}(Z) = n] &\approx \mathbb{P}[T_{0n}(Z^*) = n], \\ \mathbb{P}[T_{0b}(Z) = n - T_{bn}(y)] &\approx \mathbb{P}[T_{0b}(Z^*) = n - T_{bn}(y)]. \end{aligned}$$

This shows that the distributions of $C^{(n)}[b+1, n]$ and $C^{*(n)}[b+1, n]$ are as close to each other as are those of $Z[b+1, n]$ and $Z^*[b+1, n]$; and these are close to one another for $b = b(n)$ large enough, by the logarithmic condition. The various limit theorems for the joint distribution of the component counts are then deduced from those for independent random variables, using the approximation of the small components, or from those for the Ewens sampling formula $C^{*(n)}$, using the approximation of the big components.

The key to the argument is thus some form of local limit approximation for the distribution of the random variables $T_{vm}(Z) = \sum_{i=v+1}^m iZ_i$, which is enough to justify (1.6) and (1.7). But although $T_{vm}(Z)$ is a sum of independent random variables, it is not approximately normally distributed. Indeed, for Z^* ,

$$(1.8) \quad n^{-1}T_{0n}(Z^*) \xrightarrow{\mathcal{L}} X_\theta,$$

where $\mathcal{L}(X_\theta)$ has a probability density p_θ on \mathbb{R}_+ satisfying

$$(1.9) \quad p_\theta(x) = \theta x^{-1} \mathbb{P}[x-1 \leq X_\theta < x].$$

The distribution of X_θ and its density p_θ were given by Vervaat [(1972), Theorem 4.7.7], and the weak convergence is elementary using Laplace transforms, as in Arratia and Tavaré [(1994), Lemma 1]. The density function in (1.9) satisfies an equation of the type $uf'(u) + af(u) + bf(u-1) = 0$; asymptotics for the solutions of such equations are studied in Hildebrand and Tenenbaum (1993).

We use Q_θ to denote the Lévy concentration function of $\mathcal{L}(X_\theta)$, and note that $Q_\theta(x) = O(x^{\bar{\theta}})$ as $x \rightarrow 0$, where $\bar{\theta} = \min\{\theta, 1\}$. We extend this result, by showing that $n^{-1}T_{bn}(Z) \xrightarrow{\mathcal{L}} X_\theta$ for all Z satisfying the uniform logarithmic condition below, and for all $b = b(n) = o(n)$ as $n \rightarrow \infty$. We then establish an approximate version of (1.9) for $n^{-1}T_{bn}(Z)$, showing that

$$(1.10) \quad \mathbb{P}[T_{bn}(Z) = s] \approx (\theta/s) \mathbb{P}[s-n \leq T_{bn}(Z) < s-b]$$

for $s = [nx]$, thus providing the required local limit approximation.

An advantage of our approach, as opposed to those based on the singularity analysis of Flajolet and Soria (1990), is the unity of argument and great generality which results, as well as the transparency of the conditions needed for the theorems. In contrast, the complex analytic approaches typically require conditions to be satisfied that can be verified in the well-known examples, but which are difficult to express directly in terms of the basic parameters of the structures. For example, Hansen's (1994) Poisson–Dirichlet approximation presupposes that such a condition is satisfied, a condition which is shown in Arratia, Barbour and Tavaré (1999) and in Theorem 3.4 in fact to be spurious. A further advantage of our method is that it can be adapted to give bounds

on the accuracy of the approximations [Arratia, Barbour and Tavaré (2000)]. With the notable exception of one-dimensional limiting approximations, where Fourier methods have a long and celebrated history, and have now been very successfully applied to component counts by Hwang (1998a, b), rates of approximation are not easily obtained by complex variable arguments.

Four main approximations are proved in this paper. The first two, Poisson–Dirichlet limits (Theorem 3.4) and the functional central limit theorem for component counts (Theorem 3.5), have already been mentioned. In addition, we prove a functional version of the Erdős–Turán (1967) law [Theorem 3.6; for previous work in special cases see Nicolas (1984, 1985)] and a Poisson approximation in total variation for the total number of components [Theorem 3.7; see also Hwang (1999)]. For all of these results, we need a slightly stronger version of the logarithmic condition, in the form of the *uniform logarithmic condition*,

$$(1.11) \quad (\text{ULC}): \varepsilon_{i1} = i\mathbb{P}[Z_i = 1] - \theta \quad \text{satisfies } |\varepsilon_{i1}| \leq e(i)c_1;$$

$$(1.12) \quad \varepsilon_{il} = i\mathbb{P}[Z_i = l] \leq e(i)c_l, \quad l \geq 2, \\ \text{where } e(i) \downarrow 0 \text{ as } i \rightarrow \infty \text{ and } D_1 = \sum_{l \geq 1} lc_l < \infty.$$

Some extra condition on the $e(i)$ or on the c_l is needed at times; the assumption that $\sum_{i \geq 1} i^{-1}e(i) < \infty$ is enough for all that we consider in this paper. For assemblies, multisets and selections, the uniform logarithmic condition is in fact no stronger than the logarithmic condition, as is shown in the following proposition, which is proved, together with two other technical results, in Section 4.

PROPOSITION 1.1. *All assemblies, multisets and selections that satisfy the logarithmic condition also satisfy the uniform logarithmic condition.*

2. Preparation. This section lays the groundwork for the theorems of Section 3, by proving the necessary properties of the distribution of the $T_{vm}(Z)$, as discussed in the Introduction. In particular, the convergence in distribution of $n^{-1}T_{bn}(Z)$ is proved in Theorem 2.4, and the formal version of the local limit approximation (1.10) in Theorem 2.6. The uniform logarithmic condition is assumed throughout, even when not explicitly mentioned.

We use two distances between probability distributions P and Q , the total variation distance

$$d_{TV}(P, Q) = \sup_A \{|P(A) - Q(A)|\}$$

and, for distributions on \mathbb{R} , the Kolmogorov distance

$$d_K(P, Q) = \sup_x \{|P(-\infty, x] - Q(-\infty, x]|\}.$$

We start by showing that the Z and Z^* sequences have similar distributions for large indices. We interpret $1/0$ as $+\infty$, for example when $b = 0$ in the following lemma.

LEMMA 2.1. *For any $0 \leq b \leq n$,*

$$d_{TV}(\mathcal{L}(Z[b + 1, n]), \mathcal{L}(Z^*[b + 1, n])) = O\left(b^{-1} + \sum_{i=b+1}^n i^{-1}e(i)\right).$$

PROOF. It is immediate from the uniform logarithmic condition that

$$d_{TV}(\mathcal{L}(Z_i), \mathcal{L}(Z_i^*)) = O(i^{-2} + i^{-1}e(i)),$$

and the lemma follows by making independent couplings of the pairs (Z_i, Z_i^*) , and using the triangle inequality. \square

COROLLARY 2.2. *There exists a sequence β_n with $\lim_{n \rightarrow \infty} n^{-1}\beta_n = 0$ such that*

$$(2.1) \quad \Delta_n := d_{TV}(\mathcal{L}(Z[\beta_n + 1, n]), \mathcal{L}(Z^*[\beta_n + 1, n])) \rightarrow 0.$$

If $\sum_{i \geq 1} i^{-1}e(i) < \infty$, then (2.1) holds for any sequence β_n such that $\lim_{n \rightarrow \infty} \beta_n = \infty$.

PROOF. Since $\sum_{i=b+1}^n i^{-1}e(i) \leq e(b) \log(n/b)$, it suffices to choose β_n large enough that $\lim_{n \rightarrow \infty} e(\beta_n) \log(n/\beta_n) = 0$. \square

We now complement these distributional approximations with a comparison of densities, for which we need the following definition. Set

$$(2.2) \quad \kappa := \min\{i: i^{-1}(\theta + D_1e(i)) < 1/2\},$$

and note that $\mathbb{P}[Z_i = 0] \geq 1/2$ for all $i \geq \kappa$.

LEMMA 2.3. *For any $0 \leq b \leq n$ and $y \in \mathbb{Z}_+^n$ such that $y[b + 1, n] \in \{0, 1\}^{n-b}$, we have*

$$(2.3) \quad \frac{\mathbb{P}[Z[b + 1, n] = y[b + 1, n]]}{\mathbb{P}[Z^*[b + 1, n] = y[b + 1, n]]} \geq 1 - K_\theta \left\{ b^{-1} + \sum_{i=b+1}^n e(i)\{i^{-1} + y_i\} \right\},$$

for some K_θ not depending on b, n or y . Furthermore, if $b \geq \kappa$,

$$(2.4) \quad \frac{\mathbb{P}[Z[b + 1, n] = y[b + 1, n]]}{\mathbb{P}[Z^*[b + 1, n] = y[b + 1, n]]} \leq \exp\left(K'_\theta \sum_{i=b+1}^n \{e(i)\{i^{-1} + y_i\} + i^{-1}y_i\}\right),$$

for some K'_θ .

PROOF. Using the uniform logarithmic condition, we have

$$\begin{aligned}
 \frac{\mathbb{P}[Z[b+1, n] = y[b+1, n]]}{\mathbb{P}[Z^*[b+1, n] = y[b+1, n]]} &= \prod_{i=b+1}^n \left\{ \frac{\mathbb{P}[Z_i = 0]^{1-y_i} \mathbb{P}[Z_i = 1]^{y_i}}{e^{-\theta/i} (\theta/i)^{y_i}} \right\} \\
 (2.5) \quad &\geq \prod_{i=b+1}^n \left\{ \left(1 - D_1 i^{-1} e(i) - \theta i^{-2} (\theta + D_1 e(1))\right)_+ (1 - D_1 \theta^{-1} e(i))_+^{y_i} \right\} \\
 &\geq 1 - \sum_{i=b+1}^n \left\{ D_1 e(i) (i^{-1} + \theta^{-1} y_i) + \theta i^{-2} (\theta + D_1 e(1)) \right\},
 \end{aligned}$$

proving the first inequality. For the second, we have

$$\frac{i \mathbb{P}[Z_i = 1]}{\theta \mathbb{P}[Z_i = 0]} \leq \frac{1 + D_1 e(i)}{\mathbb{P}[Z_i = 0]} \leq (1 + D_1 e(i)) \left\{ 1 + 2i^{-1} (\theta + D_1 e(i)) \right\},$$

since $(1 - x)^{-1} \leq 1 + 2x$ in $x \leq 1/2$, and the remainder of the proof follows from (2.5). \square

On the basis of these comparisons, the convergence in distribution of $m^{-1} \times T_{vm}(Z)$ to X_θ can be established.

THEOREM 2.4. *If the uniform logarithmic condition is satisfied, then, for any sequence B_m such that $\lim_{m \rightarrow \infty} m^{-1} B_m = 0$, it follows that*

$$\lim_{m \rightarrow \infty} \max_{0 \leq v \leq B_m} d_K \left(\mathcal{L}(m^{-1} T_{vm}(Z)), \mathcal{L}(X_\theta) \right) = 0,$$

where X_θ is as in (1.8).

PROOF. Assume without loss of generality that $B_m \geq \beta_m$, where β_m is as in Corollary 2.2. For each m , construct Z and Z^* on the same probability space $(\widehat{\Omega}_m, \widehat{\mathcal{F}}_m, \widehat{\mathbb{P}}_m)$ in such a way that

$$A_m := \{Z[B_m + 1, m] \neq Z^*[B_m + 1, m]\}$$

satisfies $\lim_{m \rightarrow \infty} \widehat{\mathbb{P}}_m[A_m] = 0$; this is possible, by Corollary 2.2. Define

$$\begin{aligned}
 X_m &= m^{-1} \left\{ T_{v, B_m}(Z) + T_{B_m, m}(Z) \right\}; \\
 Y_m &= m^{-1} \left\{ T_{0, B_m}(Z^*) + T_{B_m, m}(Z^*) \right\},
 \end{aligned}$$

and note that, on A_m^c , from the uniform logarithmic condition,

$$m^{-1} \widehat{\mathbb{E}}_m |X_m - Y_m| = m^{-1} \widehat{\mathbb{E}}_m |T_{v, B_m}(Z) - T_{0, B_m}(Z^*)| \leq (\theta + D_1 e(1)) m^{-1} B_m.$$

Now apply Lemma 4.2 with $X = X_m$, $Y = Y_m$, $A = A_m$ and $Z = X_\theta$, the proof being complete because $Y_m \xrightarrow{\mathcal{D}} X_\theta$, in view of (1.8). \square

We now turn to point probabilities for the random variables $T_{vm}(Z)$, which we need to control in order to make precise the argument sketched in the Introduction. The first step is to prove a crude general bound for such probabilities. The argument derives from that of Theorem 4.4 of Arratia, Barbour and Tavaré (1999).

LEMMA 2.5. *If $\lim_{n \rightarrow \infty} n^{-1}B_n = 0$, then*

$$\lim_{n \rightarrow \infty} \max_{0 \leq b \leq B_n} \max_{k \geq 0} \mathbb{P}[T_{bn}(Z) = k] = 0.$$

PROOF. Write $\beta'_n = \max\{B_n, \beta_n\}$. Then, since $T_{bn}(Z) = T_{b, \beta'_n}(Z) + T_{\beta'_n, n}(Z)$ and the two summands are independent, it follows for all $b \leq B_n$ that

$$\max_{k \geq 0} \mathbb{P}[T_{bn}(Z) = k] \leq \max_{k \geq 0} \mathbb{P}[T_{\beta'_n, n}(Z) = k].$$

However, now

$$\max_{k \geq 0} \mathbb{P}[T_{\beta'_n, n}(Z) = k] \leq \max_{k \geq 0} \mathbb{P}[T_{\beta'_n, n}(Z^*) = k] + \Delta_n,$$

where $\lim_{n \rightarrow \infty} \Delta_n = 0$ from Corollary 2.2, and

$$\max_{k \geq 0} \mathbb{P}[T_{\beta'_n, n}(Z^*) = k] \leq \exp\{-\bar{\theta}(h(n) - h(\beta'_n))\} \sim (\beta'_n/n)^{\bar{\theta}} \rightarrow 0$$

as $n \rightarrow \infty$, by Arratia, Barbour and Tavaré [(1999), Lemma 3.2]: here and subsequently, $h(m) = \sum_{i=1}^m i^{-1}$ denotes the m th harmonic number. \square

In the course of the proof of the local limit approximation below, we need not only the bound in Lemma 2.5, but also a similar bound for sums defined as for the $T_{vm}(Z)$, but with one term of the sum missing. So for any $0 \leq v < i \leq m$, define

$$(2.6) \quad T_{vm}^{(i)}(Z) = \sum_{\substack{j=v+1 \\ j \neq i}}^m jZ_j.$$

By independence, it is immediate that

$$\mathbb{P}[T_{vm}^{(i)}(Z) = r] \mathbb{P}[Z_i = l] \leq \mathbb{P}[T_{vm}(Z) = r + il]$$

for any $r, l \geq 0$, leading to the simple bounds

$$(2.7) \quad \mathbb{P}[T_{vm}^{(i)}(Z) = r] \leq 2\mathbb{P}[T_{vm}(Z) = r]$$

whenever $i \geq \kappa$, and, for all i ,

$$(2.8) \quad \max_{r \geq 0} \mathbb{P}[T_{vm}^{(i)}(Z) = r] \leq c(\kappa) \max_{s \geq 0} \mathbb{P}[T_{vm}(Z) = s],$$

where

$$(2.9) \quad c(\kappa) := \max \left\{ 2, \left[\min_{1 \leq j < \kappa} \max_{l \geq 0} \mathbb{P}[Z_j = l] \right]^{-1} \right\}.$$

Thus point probabilities for the random variables $T_{vm}^{(i)}(Z)$ can be bounded in terms of point probabilities for the $T_{vm}(Z)$; in particular, (2.8) can be combined with Lemma 2.5.

THEOREM 2.6. *If the uniform logarithmic condition holds and $\lim_{m \rightarrow \infty} m^{-1} B_m = 0$, then*

$$\lim_{m \rightarrow \infty} \max_{0 \leq v \leq B_m} \sup_{s \geq 1} \left| s \mathbb{P}[T_{vm}(Z) = s] - \theta \mathbb{P}[m^{-1}(s - m) \leq X_\theta < m^{-1}(s - v)] \right| = 0.$$

PROOF. Temporarily, write W for $T_{vm}(Z)$ and W_i for $T_{vm}^{(i)}(Z)$. Then, by conditioning on the value l taken by Z_i , we have

$$(2.10) \quad \begin{aligned} \mathbb{E}\{Z_i I[W = s]\} &= \sum_{i \geq 1} l \mathbb{P}[Z_i = l] \mathbb{P}[W_i = s - il] \\ &= i^{-1} \theta \mathbb{P}[W_i = s - i] + i^{-1} \sum_{l \geq 1} l \varepsilon_{il} \mathbb{P}[W_i = s - il], \end{aligned}$$

where the ε_{il} are as defined in the uniform logarithmic condition. Multiplying by i and adding over $v + 1 \leq i \leq m$ thus immediately gives

$$(2.11) \quad s \mathbb{P}[W = s] = \theta \sum_{i=v+1}^m \mathbb{P}[W_i = s - i] + \sum_{i=v+1}^m \sum_{l \geq 1} l \varepsilon_{il} \mathbb{P}[W_i = s - il].$$

Now the probabilities $\mathbb{P}[W_i = s - i]$ can be replaced in the first sum by $\mathbb{P}[W = s - i]$, together with an appropriate correction, since, again conditioning on the value l taken by Z_i , we have

$$(2.12) \quad \begin{aligned} \mathbb{P}[W = s - i] &= \mathbb{P}[W_i = s - i] \\ &\quad + i^{-1} \theta (1 + \varepsilon_{il}) \left\{ \mathbb{P}[W_i = s - 2i] - \mathbb{P}[W_i = s - i] \right\} \\ &\quad + i^{-1} \theta \sum_{l \geq 2} \varepsilon_{il} \left\{ \mathbb{P}[W_i = s - (l + 1)i] \right. \\ &\quad \quad \left. - \mathbb{P}[W_i = s - i] \right\}. \end{aligned}$$

Thus it follows, by combining (2.11) and (2.12), that the quantity

$$(2.13) \quad \left| s \mathbb{P}[T_{vm}(Z) = s] - \theta \mathbb{P}[s - m \leq T_{vm}(Z) < s - v] \right|$$

is bounded in terms of three sums involving the ε_{il} and point probabilities for the random variables $T_{vm}^{(i)}(Z)$; it is for the latter that (2.7) and (2.8) are used.

In order to show that the three sums are asymptotically negligible, we begin by observing that we can choose a sequence $w_m \geq \kappa$ such that $w_m \rightarrow \infty$ and yet

$$(2.14) \quad \lim_{m \rightarrow \infty} \left\{ h(w_m) + \sum_{i=1}^{w_m} e(i) \right\} \max_{0 \leq v \leq B_m} \max_{r \geq 0} \mathbb{P}[T_{vm}(Z) = r] = 0,$$

in view of Lemma 2.5. Taking the first sum, for any $s \geq 1$ and $0 \leq v \leq B_m$, we have

$$\begin{aligned} & \sum_{i=v+1}^m \sum_{l \geq 1} l \varepsilon_{il} \mathbb{P}[W_i = s - il] \\ & \leq \left\{ \sum_{i=1}^{w_m} + \sum_{i=w_m}^m \right\} e(i) \sum_{l \geq 1} l c_l \mathbb{P}[W_i = s - il] \\ & \leq c(\kappa) D_1 \sum_{i=1}^{w_m} e(i) \max_{0 \leq v \leq B_m} \max_{r \geq 0} \mathbb{P}[T_{vm}(Z) = r] \\ & \quad + 2 \sum_{i=w_m}^m e(i) \sum_{l \geq 1} l c_l \mathbb{P}[T_{vm}(Z) = s - il], \end{aligned}$$

where the last line follows from (2.7) and (2.8); the first contribution is then negligible because of (2.14), and the second because it is bounded by $2D_1 e(w_m)$, since $\sum_{i \geq 0} \mathbb{P}[T_{vm}(Z) = s - il] \leq 1$. The third sum is negligible because

$$\begin{aligned} & \left| \sum_{i=v+1}^m i^{-1} \sum_{l \geq 2} \varepsilon_{il} \{ \mathbb{P}[W_i = s - (l+1)i] - \mathbb{P}[W_i = s - i] \} \right| \\ & \leq \left\{ \sum_{i=1}^{w_m} + \sum_{i=w_m}^m \right\} i^{-1} e(i) \sum_{l \geq 2} c_l \left\{ \mathbb{P}[T_{vm}^{(i)}(Z) = s - (l+1)i] \right. \\ & \quad \left. + \mathbb{P}[T_{vm}^{(i)}(Z) = s - i] \right\} \\ & \leq c(\kappa) D_1 \sum_{i=1}^{w_m} e(i) \max_{0 \leq v \leq B_m} \max_{r \geq 0} \mathbb{P}[T_{vm}(Z) = r] + 4w_m^{-1} e(w_m) \sum_{l \geq 2} c_l, \end{aligned}$$

which tends to 0 much as before, and for the second sum we have

$$\begin{aligned} & \left| \sum_{i=v+1}^m i^{-1} (1 + \varepsilon_{il}) \{ \mathbb{P}[W_i = s - 2i] - \mathbb{P}[W_i = s - i] \} \right| \\ & \leq \left\{ \sum_{i=1}^{w_m} + \sum_{i=w_m}^m \right\} i^{-1} (1 + e(i) c_1) \\ & \quad \times \left\{ \mathbb{P}[T_{vm}^{(i)}(Z) = s - 2i] + \mathbb{P}[T_{vm}^{(i)}(Z) = s - i] \right\}, \end{aligned}$$

with the first contribution no larger than

$$2(1 + e(1)c_1)c(\kappa)h(w_m) \max_{0 \leq v \leq B_m} \max_{r \geq 0} \mathbb{P}[T_{vm}(Z) = r],$$

which is negligible by (2.14), and the second contribution at most $4w_m^{-1}(1 + e(1)c_1) \rightarrow 0$.

This shows that the difference (2.13) is uniformly small in the prescribed ranges of v and s . The final step is to note that

$$\begin{aligned} & \max_{0 \leq v \leq B_m} \sup_{s \geq 1} \left| \mathbb{P}[s - m \leq T_{vm}(Z) < s - v] \right. \\ & \quad \left. - \mathbb{P}[m^{-1}(s - m) \leq X_\theta < m^{-1}(s - v)] \right| \\ & \leq 2 \max_{0 \leq v \leq B_m} d_K(\mathcal{L}(m^{-1}T_{vm}(Z)), \mathcal{L}(X_\theta)) \rightarrow 0, \end{aligned}$$

by Theorem 2.4, concluding the proof. \square

REMARK. Note that (2.13) is exactly zero for $Z = Z^*$:

$$(2.15) \quad s\mathbb{P}[T_{vm}(Z^*) = s] = \theta\mathbb{P}[s - m \leq T_{vm}(Z^*) < s - v].$$

The two following corollaries translate the result of Theorem 2.6 into the forms most useful for making precise the arguments sketched in the Introduction. The first of them requires no proof.

COROLLARY 2.7. *If $\lim_{n \rightarrow \infty} n^{-1}B_n = 0$, then, for any $x > 0$,*

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \max_{0 \leq b \leq B_n} \max_{s: |sn^{-1} - x| < \eta} |s\mathbb{P}[T_{bn}(Z) = s] - \theta\mathbb{P}[x - 1 \leq X_\theta < x]| \\ & \leq 2\theta Q_\theta(\eta). \end{aligned}$$

COROLLARY 2.8. *For any $0 < \varepsilon < 1 < M < \infty$,*

$$(2.16) \quad \lim_{m \rightarrow \infty} \max_{[\varepsilon m] < s \leq [Mm]} \left| 1 - \frac{\mathbb{P}[T_{0m}(Z) = s]}{\mathbb{P}[T_{0m}(Z^*) = s]} \right| = 0.$$

In particular,

$$\lim_{n \rightarrow \infty} \frac{\mathbb{P}[T_{0n}(Z) = n]}{\mathbb{P}[T_{0n}(Z^*) = n]} = 1.$$

PROOF. Apply Theorem 2.5 to approximate both $s\mathbb{P}[T_{0m}(Z) = s]$ and $s\mathbb{P}[T_{0m}(Z^*) = s]$, and note that, for $[\varepsilon m] < s \leq [Mm]$,

$$(2.17) \quad \begin{aligned} & \mathbb{P}[m^{-1}(s - m) \leq X_\theta < m^{-1}s] \\ & \geq \min\{\mathbb{P}[X_\theta < \varepsilon], \min_{1 \leq x \leq M} \mathbb{P}[x - 1 \leq X_\theta < x]\}. \end{aligned} \quad \square$$

3. Main results. Our proofs of the limit theorems are based on two fundamental results: that the small components are jointly distributed like their independent counterparts, and the large components like those of the Ewens sampling formula. The first part of this section consists of proving these assertions.

THEOREM 3.1 (The small components). *If the uniform logarithmic condition holds and $\lim_{n \rightarrow \infty} n^{-1}b_n = 0$, then*

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

PROOF. It is immediate from (1.4) that

$$\begin{aligned} & d_{TV}(\mathcal{L}(C^{(n)}[1, b]), \mathcal{L}(Z[1, b])) \\ (3.1) \quad &= \sum_{r \geq 0} \mathbb{P}[T_{0b}(Z) = r] \left\{ 1 - \frac{\mathbb{P}[T_{bn}(Z) = n - r]}{\mathbb{P}[T_{0n}(Z) = n]} \right\}_+ \end{aligned}$$

Now, writing $b = b_n = o(n)$ and applying Corollary 2.7 with $x = 1$ and $\eta = \sqrt{b/n}$, it follows that

$$\lim_{n \rightarrow \infty} \max_{0 \leq r \leq \lfloor \sqrt{bn} \rfloor} |(n - r)\mathbb{P}[T_{bn}(Z) = n - r] - \theta\mathbb{P}[X_\theta < 1]| = 0,$$

and in particular that $\lim_{n \rightarrow \infty} n\mathbb{P}[T_{0n}(Z) = n] = \theta\mathbb{P}[X_\theta < 1]$; hence

$$(3.2) \quad \lim_{n \rightarrow \infty} \sum_{r=0}^{\lfloor \sqrt{bn} \rfloor} \mathbb{P}[T_{0b}(Z) = r] \left\{ 1 - \frac{\mathbb{P}[T_{bn}(Z) = n - r]}{\mathbb{P}[T_{0n}(Z) = n]} \right\}_+ = 0.$$

Since also, from Markov’s inequality and the uniform logarithmic condition,

$$\sum_{r > \lfloor \sqrt{bn} \rfloor} \mathbb{P}[T_{0b}(Z) = r] \leq (bn)^{-1/2} \mathbb{E}T_{0b}(Z) \leq (\theta + D_1 e(1))(b/n)^{1/2} \rightarrow 0,$$

the theorem follows. \square

THEOREM 3.2 (The large components). *If the uniform logarithmic condition holds and $\lim_{n \rightarrow \infty} b_n = \infty$, $\lim_{n \rightarrow \infty} n^{-1}b_n = 0$, then*

$$\begin{aligned} & \limsup_{n \rightarrow \infty} d_{TV}(\mathcal{L}(C^{(n)}[b_n + 1, n]), \mathcal{L}(C^{*(n)}[b_n + 1, n])) \\ &= O\left(\limsup_{n \rightarrow \infty} \sum_{i=b_n+1}^n i^{-1}e(i)\right). \end{aligned}$$

If $\sum_{i \geq 1} i^{-1}e(i) < \infty$, or if, in general, $b_n \geq \beta_n$ for all n sufficiently large, then

$$\lim_{n \rightarrow \infty} d_{TV}(\mathcal{L}(C^{(n)}[b_n + 1, n]), \mathcal{L}(C^{*(n)}[b_n + 1, n])) = 0.$$

PROOF. Appealing to (1.5), and suppressing the index n where possible, we have

$$\begin{aligned}
 & d_{TV}(\mathcal{L}(C[b+1, n]), \mathcal{L}(C^*[b+1, n])) \\
 &= \sum_y \mathbb{P}[C^*[b+1, n] = y[b+1, n]] \\
 &\quad \times \left\{ 1 - \frac{\mathbb{P}[C[b+1, n] = y[b+1, n]]}{\mathbb{P}[C^*[b+1, n] = y[b+1, n]]} \right\}_+ \\
 (3.3) \quad &= \sum_y \mathbb{P}[C^*[b+1, n] = y[b+1, n]] \\
 &\quad \times \left\{ 1 - \frac{\mathbb{P}[Z[b+1, n] = y[b+1, n]]}{\mathbb{P}[Z^*[b+1, n] = y[b+1, n]]} \right. \\
 &\quad \left. \times \frac{\mathbb{P}[T_{0b}(Z) = n-t] \mathbb{P}[T_{0n}(Z^*) = n]}{\mathbb{P}[T_{0b}(Z^*) = n-t] \mathbb{P}[T_{0n}(Z) = n]} \right\}_+,
 \end{aligned}$$

where t is shorthand for $T_{bn}(y)$ and the y -sum runs over \mathbb{Z}_+^n with $T_{bn}(y) \leq n$.

Fix any $0 < \varepsilon < 1 < M < \infty$. Then, for t such that $\varepsilon b \leq n-t \leq Mb$, we have from (2.15) that

$$\begin{aligned}
 \mathbb{P}[T_{bn}(C^*) = t] &= \frac{\mathbb{P}[T_{bn}(Z^*) = t] \mathbb{P}[T_{0b}(Z^*) = n-t]}{\mathbb{P}[T_{0n}(Z^*) = n]} \\
 (3.4) \quad &= \frac{n \mathbb{P}[T_{bn}(Z^*) < t-b]}{t \mathbb{P}[T_{0n}(Z^*) < n]} \mathbb{P}[T_{0b}(Z^*) = n-t] \\
 &\geq \frac{\mathbb{P}[n^{-1}T_{bn}(Z^*) < 1 - n^{-1}(M+1)b]}{\mathbb{P}[n^{-1}T_{0n}(Z^*) < 1]} \\
 &\quad \times \mathbb{P}[T_{0b}(Z^*) = n-t].
 \end{aligned}$$

Thus, since $b = b_n \rightarrow \infty$ and $b_n = o(n)$, it follows from (1.8) that

$$\begin{aligned}
 & \liminf_{n \rightarrow \infty} \mathbb{P}[T_{bn}(C^*) \in [n - Mb, n - \varepsilon b]] \\
 & \geq \lim_{m \rightarrow \infty} \mathbb{P}[m^{-1}T_{0m}(Z^*) \in [\varepsilon, M]] = \mathbb{P}[\varepsilon \leq X_\theta \leq M],
 \end{aligned}$$

and hence that

$$(3.5) \quad \limsup_{n \rightarrow \infty} \mathbb{P}[T_{bn}(C^*) \notin [n - Mb, n - \varepsilon b]] \leq \mathbb{P}[X_\theta \notin [\varepsilon, M]].$$

Thus the contribution to (3.3) from elements y for which $T_{bn}(y) \notin [n - Mb, n - \varepsilon b]$ can be made arbitrarily small, by choosing ε to be small and M to be large.

For the contributions to (3.3) from y such that $T_{bn}(y) \in [n - Mb, n - \varepsilon b]$, we use Corollary 2.8 to give

$$\lim_{n \rightarrow \infty} \min_{[\varepsilon b] < n-t \leq [Mb]} \frac{\mathbb{P}[T_{0b}(Z) = n-t]}{\mathbb{P}[T_{0b}(Z^*) = n-t]} = 1$$

and

$$\lim_{n \rightarrow \infty} \mathbb{P}[T_{0n}(Z^*) = n] / \mathbb{P}[T_{0n}(Z) = n] = 1;$$

for $y \in \mathbb{Z}_+^n$ with $y[b + 1, n] \in \{0, 1\}^{n-b}$, Lemma 2.3 gives

$$\liminf_{n \rightarrow \infty} \frac{\mathbb{P}[Z[b + 1, n] = y[b + 1, n]]}{\mathbb{P}[Z^*[b + 1, n] = y[b + 1, n]]} \geq 1 - O\left(\sum_{i=b+1}^n e(i)\{i^{-1} + y_i\}\right),$$

and a simple calculation gives

$$\mathbb{P}[C^*[b + 1, n] \notin \{0, 1\}^{n-b}] \leq b^{-1}\theta^2;$$

finally,

$$\begin{aligned} \sum_y \mathbb{P}[C^*[b + 1, n] = y[b + 1, n]] \sum_{i=b+1}^n e(i)y_i \\ = \sum_{i=b+1}^n e(i)\mathbb{E}C_i^* = O\left(\sum_{i=b+1}^n i^{-1}e(i)\right), \end{aligned}$$

from Lemma 4.1. Collecting these facts, it follows that

$$\begin{aligned} \limsup_{n \rightarrow \infty} d_{TV}(\mathcal{L}(C[b + 1, n]), \mathcal{L}(C^*[b + 1, n])) \\ \leq \mathbb{P}[X_\theta \notin [\varepsilon, M]] + O\left(\limsup_{n \rightarrow \infty} \sum_{i=b+1}^n i^{-1}e(i)\right), \end{aligned}$$

for any $0 < \varepsilon < 1 < M < \infty$, and the theorem is proved. \square

These two theorems, describing the behavior of the small and the large components separately, are enough for proving any limit theorems which are essentially governed by the behavior either of the small or the large components alone, and, since the two ranges often overlap substantially, this covers most applications. However, for Theorem 3.7, the joint distribution of all the components simultaneously is required in an essential way; in order to cope with this, we use a conditional variant of Theorem 3.2.

THEOREM 3.3 (The large components, conditional form). *For any combinatorial structure satisfying the uniform logarithmic condition,*

$$\begin{aligned} \lim_{n \rightarrow \infty} \max_{n/2 \leq l \leq n} d_{TV}(\mathcal{L}(C[\beta_n + 1, n] | T_{\beta_n, n}(C) = l), \\ \mathcal{L}(C^*[\beta_n + 1, n] | T_{\beta_n, n}(C^*) = l)) = 0, \end{aligned}$$

where β_n is as defined in Corollary 2.2.

PROOF. We write b for β_n throughout, and suppress the index n where possible. For any $y \in \{0, 1\}^n$ such that $\sum_{i=b+1}^n iy_i = l$, we first use Lemma 2.3 and Theorem 2.6 to show that

$$\begin{aligned}
 (3.6) \quad & \frac{\mathbb{P}[C[b+1, n] = y[b+1, n] | T_{bn}(C) = l]}{\mathbb{P}[C^*[b+1, n] = y[b+1, n] | T_{bn}(C^*) = l]} \\
 &= \frac{\mathbb{P}[Z[b+1, n] = y[b+1, n]] \mathbb{P}[T_{bn}(Z^*) = l]}{\mathbb{P}[Z^*[b+1, n] = y[b+1, n]] \mathbb{P}[T_{bn}(Z) = l]} \\
 (3.7) \quad & \geq 1 - (D_1/\theta) \sum_{i=b+1}^l e(i)y_i - O(\eta(l, n)) \\
 & \geq 1 - (D_1/\theta) \left\{ \sum_{i=b+1}^{\lfloor l/2 \rfloor} e(i)y_i + e(\lfloor l/2 \rfloor) \sum_{i=\lfloor l/2 \rfloor+1}^l y_i \right\} - O(\eta(l, n)),
 \end{aligned}$$

where $\lim_{n \rightarrow \infty} \max_{n/2 \leq l \leq n} \eta(l, n) = 0$. We thus find that, for any $A \subset \mathbb{Z}_+^{n-b}$,

$$\begin{aligned}
 (3.8) \quad & \mathbb{P}[C[b+1, n] \in A | T_{bn}(C) = l] \\
 & \geq \mathbb{P}[C^*[b+1, n] \in A | T_{bn}(C^*) = l] \\
 & \quad - \mathbb{P} \left[\bigcup_{i=b+1}^n \{C_i^* \geq 2\} \mid T_{bn}(C^*) = l \right] \\
 & \quad - (C_1/\theta) \sum_{i=b+1}^{\lfloor l/2 \rfloor} e(i) \mathbb{P}[C_i^* = 1 | T_{bn}(C^*) = l] - O(\eta(l, n)),
 \end{aligned}$$

since, if $T_{bn}(C^*) = l$, then $\sum_{i=\lfloor l/2 \rfloor+1}^l C_i^* \leq 1$. Hence there are two remaining elements in (3.8) to be shown to be small.

First, using the logarithmic relation and (2.15) and recalling the definition (2.6) of $T_{bn}^{(i)}(Z)$, we observe that, for $i \leq l/2$,

$$\begin{aligned}
 \mathbb{P}[C_i^* = 1 | T_{bn}(C^*) = l] &= \frac{\mathbb{P}[Z_i^* = 1] \mathbb{P}[T_{bn}^{(i)}(Z^*) = l - i]}{\mathbb{P}[T_{bn}(Z^*) = l]} \\
 &\leq e^{-\theta/i} \frac{\theta}{i} \frac{e^{\theta/i} \mathbb{P}[T_{bn}(Z^*) < l - i - b]}{\mathbb{P}[T_{bn}(Z^*) < l - b]} \left(\frac{l}{l - i} \right) = O(i^{-1})
 \end{aligned}$$

uniformly in $n/2 \leq l \leq n$, by Theorem 2.4, so that

$$(3.9) \quad \lim_{n \rightarrow \infty} \max_{n/2 \leq l \leq n} (D_1/\theta) \sum_{i=b+1}^{\lfloor l/2 \rfloor} e(i) \mathbb{P}[C_i^* = 1 | T_{bn}(C^*) = l] = 0.$$

Then, by similar estimates,

$$\begin{aligned}
 \mathbb{P}[C_i^* = r | T_{bn}(C^*) = l] &= \frac{e^{-\theta/i} \left(\frac{\theta}{i}\right)^r \mathbb{P}[T_{bn}^{(i)}(Z^*) = l - ir]}{r! \left(\frac{\theta}{i}\right)^r \mathbb{P}[T_{bn}(Z^*) = l]} \\
 (3.10) \qquad &\leq \frac{1}{r!} \left(\frac{\theta}{i}\right)^r \frac{\mathbb{P}[T_{bn}(Z^*) < l - ir - b]}{\mathbb{P}[T_{bn}(Z^*) < l - b]} \left(\frac{l/r}{(l/r) - i}\right).
 \end{aligned}$$

If $i \leq l/2r$, we bound (3.10) by $2\theta^r/r!i^r$; if $l/2r < i < l/r$, we bound by $((2r\theta)^r/r!l^r) \left((l/r)/(l/r - i)\right)$, and if $i = l/r$, we bound by $((r\theta)^r/r!l^r)(l/\mathbb{P}[T_{bn}(Z^*) < l - b])$. Adding over the range $i \geq b + 1$, this gives

$$\begin{aligned}
 &\mathbb{P}\left[\bigcup_{i=b+1}^n \{C_i^* = r\} \mid T_{bn}(C^*) = l\right] \\
 &\leq \frac{2\theta^r}{r!b^{r-1}} + \frac{(2e\theta)^r}{r^{3/2}l^{r-1}} \log(l/r) + \frac{(e\theta)^r}{r^{1/2}l^{r-1}\mathbb{P}[T_{0n}(Z^*) < n/4]},
 \end{aligned}$$

uniformly in $n/2 \leq l \leq n$, for all n so large that $\beta_n < n/4$. Hence, adding over $r \geq 2$, it follows that

$$(3.11) \quad \max_{n/2 \leq l \leq n} \mathbb{P}\left[\bigcup_{i=b+1}^n \{C_i^* \geq 2\} \mid T_{bn}(C^*) = l\right] = O(b^{-1} + n^{-1} \log n) \rightarrow 0$$

as $n \rightarrow \infty$. Putting (3.9) and (3.11) into (3.8) gives the theorem. \square

Armed with these fundamental approximation theorems, we can now proceed to the main applications. The first is a Poisson–Dirichlet limit for the large components in local form; the traditional distributional limit theorem is a direct consequence of this more detailed result, by Scheffé’s theorem. The argument is much as in Arratia, Barbour and Tavaré [(1999), Theorem 4.5], but our conditions are now rather weaker. Define $L_j^{(n)}$ to be the size of the j th largest component of $C^{(n)}$, so that, if $n \geq l_1 > l_2 > \dots > l_r \geq 1$ are such that $t = \sum_{j=1}^r l_j \leq n$, then the events $\{L_1^{(n)} = l_1, \dots, L_r^{(n)} = l_r\}$ and $\{C^{(n)}[l_r + 1, n] = y^l[l_r + 1, n]; C_{l_r}^{(n)} \geq 1\}$ are the same, where

$$y_i^l = 1, \quad 1 \leq i \leq r - 1; \quad y_j^l = 0, \quad j \in [l_r + 1, n] \setminus \{l_1, \dots, l_r\}.$$

THEOREM 3.4 (Poisson–Dirichlet local limit theorem). *Fix any $r \geq 1$, and suppose that $1 > x_1 > x_2 > \dots > x_r > 0$ satisfy $0 < 1 - \sum_{i=1}^r x_i \neq mx_r$ for any integer $m \geq 1$. Then, if the uniform logarithmic condition holds,*

$$\lim_{n \rightarrow \infty} n^r \mathbb{P}[L_1^{(n)} = [nx_1], \dots, L_r^{(n)} = [nx_r]] = f_\theta^{(r)}(x_1, \dots, x_r),$$

where $f_\theta^{(r)}$ is the joint density of the first r components of the PD(θ) Poisson–Dirichlet process.

PROOF. Suppressing the index n where possible, and writing $l_i = [nx_i]$, $1 \leq i \leq r$, $b = l_r - 1$ and $t = \sum_{i=1}^r l_i$, it follows from the conditioning relation that

$$\begin{aligned}
 & \frac{\mathbb{P}[L_1 = l_1, \dots, L_r = l_r]}{\mathbb{P}[C^*[l_r + 1, n] = y^l[l_r + 1, n]; C_{l_r}^* = 1]} \\
 (3.12) \quad &= \sum_{s \geq 1} \left\{ \frac{\mathbb{P}[Z[l_r + 1, n] = y^l[l_r + 1, n]; Z_{l_r} = s]}{\mathbb{P}[Z[l_r + 1, n] = y^l[l_r + 1, n]; Z_{l_r} = 1]} \right. \\
 & \quad \times \left. \frac{\mathbb{P}[T_{0b}(Z) = n - t - (s - 1)l_r]}{\mathbb{P}[T_{0b}(Z^*) = n - t]} \frac{\mathbb{P}[T_{0n}(Z^*) = n]}{\mathbb{P}[T_{0n}(Z) = n]} \right\}
 \end{aligned}$$

Note that, for n sufficiently large, at most $x_r^{-1}(1 - \sum_{i=1}^r x_i) + 2$ values of s give nonzero contributions to the sum. Now, from Lemma 2.3, we have

$$\lim_{n \rightarrow \infty} \frac{\mathbb{P}[Z[l_r + 1, n] = y^l[l_r + 1, n]; Z_{l_r} = 1]}{\mathbb{P}[Z^*[l_r + 1, n] = y^l[l_r + 1, n]; Z_{l_r}^* = 1]} = 1$$

and

$$\lim_{n \rightarrow \infty} \frac{\mathbb{P}[Z[l_r + 1, n] = y^l[l_r + 1, n]; Z_{l_r} = s]}{\mathbb{P}[Z^*[l_r + 1, n] = y^l[l_r + 1, n]; Z_{l_r}^* = 1]} = 0, \quad s \geq 1;$$

furthermore, from Corollary 2.8, $\lim_{n \rightarrow \infty} \mathbb{P}[T_{0n}(Z^*) = n] / \mathbb{P}[T_{0n}(Z) = n] = 1$. Finally, it follows from Corollary 2.8 that

$$\lim_{n \rightarrow \infty} \mathbb{P}[T_{0b}(Z) = n - t] / \mathbb{P}[T_{0b}(Z^*) = n - t] = 1$$

and that

$$(3.13) \quad \lim_{n \rightarrow \infty} \mathbb{P}[T_{0b}(Z) = n - t - ml_r] / \mathbb{P}[T_{0b}(Z^*) = n - t] < \infty$$

for $1 \leq m \leq [x_r^{-1}(1 - \sum_{i=1}^r rx_i)]$. Thus we have established that

$$\lim_{n \rightarrow \infty} \frac{\mathbb{P}[L_1 = [nx_1], \dots, L_r = [nx_r]]}{\mathbb{P}[C^*[l_r + 1, n] = y^l[l_r + 1, n]; C_{l_r}^* = 1]} = 1,$$

and the theorem now follows from the corresponding theorem for C^* , proved in Arratia, Barbour and Tavaré [(1999), Theorem 3.3]. \square

Note that if $z = x_r^{-1}(1 - \sum_{i=1}^r x_i)$ is an integer, then (3.13) is no longer true for $m = z$.

For the next limit theorem, the FCLT for the component counts, we define the process W_n by

$$(3.14) \quad W_n(t) = \{\theta \log n\}^{-1/2} \sum_{i=1}^{[n^t]} (C_i^{(n)} - \mathbb{E}Z_i), \quad 0 \leq t \leq 1.$$

THEOREM 3.5 (FCLT for the component counts). *If the uniform logarithmic condition holds, and, in addition, either $\sum_{s \geq 1} s^2 c_s < \infty$ or $e(i) = o(\{\log i\}^{-1/2})$ as $i \rightarrow \infty$, then $W_n \Rightarrow W$ in $D[0, 1]$, where W is standard Brownian motion. In the latter case, $\sum_{i=1}^{\lfloor nt \rfloor} \mathbb{E} Z_i$ can be replaced in the definition of W_n by $\theta t \log n$.*

PROOF. The “small components” argument reduces the problem to a FCLT for the independent random variables Z_i , and the conditions additional to the uniform logarithmic condition are only needed to prove this FCLT.

Choose $b_n \rightarrow \infty$ in such a way that $b_n \geq \beta_n$ for all n , $b_n = o(n)$ and $\log(n/b_n) = o(\{\log n\}^{1/2})$ as $n \rightarrow \infty$. Then

$$\lim_{n \rightarrow \infty} d_{TV}(\mathcal{L}(C[b_n + 1, n]), \mathcal{L}(C^*[b_n + 1, n])) = 0$$

by Theorem 3.2 and the definition of β_n in Corollary 2.2, and

$$\lim_{n \rightarrow \infty} d_{TV}(\mathcal{L}(C[1, b_n]), \mathcal{L}(Z[1, b_n])) = 0$$

by Theorem 3.1. Furthermore,

$$\mathbb{E} \left\{ \sum_{i=b_n+1}^n C_i^* \right\} = O(\log(n/b_n))$$

from Lemma 4.1, so that

$$\{\log n\}^{-1/2} \sum_{i=b_n+1}^n C_i^* \xrightarrow{\mathcal{D}} 0,$$

and replacing C_i^* by Z_i and using the logarithmic condition to bound the expectations shows that

$$\{\log n\}^{-1/2} \sum_{i=b_n+1}^n Z_i \xrightarrow{\mathcal{D}} 0$$

also. Hence, to prove the theorem, it is enough to show that $W'_n \Rightarrow W$ in $D[0, 1]$, where

$$(3.15) \quad W'_n(t) = \{\theta \log n\}^{-1/2} \sum_{i=1}^{\lfloor nt \rfloor} (Z_i - \mathbb{E} Z_i)$$

is the partial sum process for a sequence of independent random variables.

The theorem stated comes from showing that

$$(3.16) \quad \sup_{0 \leq t \leq 1} |W'_n(t) - W''_n(t)| \xrightarrow{\mathcal{D}} 0,$$

where

$$(3.17) \quad W''_n(t) = \{\theta \log n\}^{-1/2} \sum_{i=1}^{\lfloor nt \rfloor} (I[Z_i = 1] - \mathbb{P}[Z_i = 1]),$$

and then applying a FCLT for independent bounded random variables to W''_n ; note that

$$(3.18) \quad \begin{aligned} \sum_{i=1}^{[n^t]} \text{Var}(I[Z_i = 1]) &= \sum_{i=1}^{[n^t]} \mathbb{P}[Z_i = 1] + O(1) \\ &= \theta t \log n + O\left(\sum_{i=1}^{[n^t]} i^{-1} e(i)\right). \end{aligned}$$

To prove (3.16) if $\sum_{s \geq 1} s^2 c_s < \infty$, use Kolmogorov's inequality, observing that

$$\begin{aligned} \text{Var}\left\{\sum_{i=1}^n (Z_i - I[Z_i = 1])\right\} &\leq \sum_{i=1}^n \mathbb{E}(Z_i - I[Z_i = 1])^2 \\ &= \sum_{i=1}^n \sum_{s \geq 2} s^2 \mathbb{P}[Z_i = s] \\ &\leq \sum_{i=1}^n i^{-1} e(i) \sum_{s \geq 2} s^2 c_s = o(\log n). \end{aligned}$$

If $e(i) = o(\{\log i\}^{-1/2})$, note instead that

$$\begin{aligned} \mathbb{E}\left\{\sum_{i=1}^n |Z_i - I[Z_i = 1]|\right\} &= \sum_{i=1}^n \sum_{s \geq 2} s \mathbb{P}[Z_i = s] \\ &\leq D_1 \sum_{i=1}^n i^{-1} e(i) = o(\{\log n\}^{1/2}), \end{aligned}$$

whence also, using (3.18),

$$\sum_{i=1}^{[n^t]} \mathbb{E} Z_i = \sum_{i=1}^{[n^t]} \mathbb{P}[Z_i = 1] + o(\{\log n\}^{1/2}) = \theta t \log n + o(\{\log n\}^{1/2}).$$

This completes the proof. \square

The Erdős–Turán law states that the logarithm of the order of a uniform random permutation is asymptotically normally distributed; if

$$O_r(y) = \text{l.c.m.}\{i: 1 \leq i \leq r, y_i \geq 1\}$$

for any vector y , then

$$(3.19) \quad \left\{\frac{1}{3}\theta \log^3 n\right\}^{-1/2} \left(\log O_n(C^{(n)}) - \frac{1}{2}\theta \log^2 n\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Here, we prove a functional form of their theorem for general logarithmic combinatorial structures. Define

$$(3.20) \quad U_n(t) = \left\{\frac{1}{3}\theta \log^3 n\right\}^{-1/2} \left(\log O_{[n^t]}(C^{(n)}) - \frac{1}{2}\theta t^2 \log^2 n\right),$$

and write $U(t) = W(t^3)$, where W is standard Brownian motion.

THEOREM 3.6 (Erdős–Turán, functional form). *If the uniform logarithmic condition holds and in addition $\sum_{i \geq 1} i^{-1}e(i) < \infty$, then $U_n \Rightarrow U$ in $D[0, 1]$.*

PROOF. The “large components” argument reduces the problem to the corresponding theorem for the Ewens sampling formula, proved in Barbour and Tavaré [(1994), Theorem 1.3]. The condition $\sum_{i \geq 1} i^{-1}e(i) < \infty$ implies that for any sequence $b_n \rightarrow \infty$, however slowly,

$$(3.21) \quad \lim_{n \rightarrow \infty} d_{TV}(\mathcal{L}(C[b_n + 1, n]), \mathcal{L}(C^*[b_n + 1, n])) = 0,$$

where, as usual, we suppress indices n where possible. Then, for any b ,

$$(3.22) \quad \sup_{0 \leq t \leq 1} \left| \log O_{[nt]}(C[b + 1, n]) - \log O_{[nt]}(C) \right| \leq \sum_{i=1}^b C_i \log i.$$

This inequality is used to ensure that, for an appropriate choice of $b = b_n$, $\log O_{[nt]}(C)$ and $\log O_{[nt]}(C[b + 1, n])$ are asymptotically equivalent, the same of course being true with C^* for C , and (3.21) implies the equivalence of $\log O_{[nt]}(C[b + 1, n])$ and $\log O_{[nt]}(C^*[b + 1, n])$.

To exploit (3.22), we first use (4.1) to show that, if $b \leq n/2$, then

$$\sum_{i=1}^b \mathbb{E}C_i^* \log i \leq 2\theta \sum_{i=1}^b i^{-1} \log i = O(\log^2 b).$$

Then, by Theorem 3.1, if $b_n = o(n)$ as $n \rightarrow \infty$,

$$\lim_{n \rightarrow \infty} d_{TV}(\mathcal{L}(C[1, b_n]), \mathcal{L}(Z[1, b_n])) = 0,$$

so that $\mathcal{L}(\sum_{i=1}^{b_n} C_i \log i)$ and $\mathcal{L}(\sum_{i=1}^{b_n} Z_i \log i)$ are asymptotically equivalent, and the logarithmic condition implies that $\sum_{i=1}^{b_n} \mathbb{E}Z_i \log i = O(\log^2 b)$ also. Hence, for the particular choice $b_n = \exp\{\sqrt{\log n}\}$, we have

$$(3.23) \quad \{\log n\}^{-3/2} \sum_{i=1}^{b_n} C_i^* \log i \xrightarrow{\mathcal{D}} 0; \quad \{\log n\}^{-3/2} \sum_{i=1}^{b_n} C_i \log i \xrightarrow{\mathcal{D}} 0.$$

Applying (3.21), (3.22) and (3.23), it thus suffices to show that $U_n^* \Rightarrow U$, where U_n^* is defined as in (3.20), but with C^* for C ; and this is proved in Barbour and Tavaré [(1994), Theorem 1.3].

Our final example concerns the total number of components, $K_{0n}(C^{(n)}) = \sum_{i=1}^n C_i^{(n)}$. The FCLT, Theorem 3.4, already gives a normal approximation for $K_{0n}(C^{(n)})$, under a slight strengthening of the uniform logarithmic condition. Here, we sharpen the mode of approximation to total variation.

THEOREM 3.7. *If the uniform logarithmic condition holds, then*

$$(3.24) \quad \lim_{n \rightarrow \infty} d_{TV}(\mathcal{L}(K_{0n}(C^{(n)})), \mathcal{L}(K_{0n}(Z))) = 0.$$

If either $\sum_{s \geq 1} s^2 c_s < \infty$ or $e(i) = o((\log i)^{-1/2})$, then

$$(3.25) \quad \lim_{n \rightarrow \infty} d_{TV}(\mathcal{L}(K_{0n}(C^{(n)})), \text{Po}(\theta \log n)) = 0.$$

PROOF. Because total variation is a very strong metric, we have to take careful account of both small and large components; the techniques used up to now to prove convergence in distribution, which involve showing that the contribution from one or other end is negligible, are no longer adequate. Instead, suppressing the index n , we write

$$(3.26) \quad K_{0n}(C) = K_{0b}(C) + K_{bn}(C),$$

for suitably chosen $b = b_n$, use the small components approximation to show that $K_{0b}(C)$ and $K_{0b}(Z)$ are equivalent, and then show that, *conditionally on the small components*, the distribution of $K_{bn}(C)$ is still close to that of $K_{bn}(Z)$, this last being established by way of C^* and Z^* , using the large components approximation.

First, for β_n as defined in Corollary 2.2, we have

$$(3.27) \quad \lim_{n \rightarrow \infty} d_{TV}(\mathcal{L}(C[1, \beta_n]), \mathcal{L}(Z[1, \beta_n])) = 0.$$

Then, from Theorem 3.3, it follows that

$$(3.28) \quad \lim_{n \rightarrow \infty} \sup_{n/2 \leq l \leq n} d_{TV}(\mathcal{L}(C[\beta_n + 1, n] \mid T_{\beta_n, n}(C) = l), \mathcal{L}(C^*[\beta_n + 1, n] \mid T_{\beta_n, n}(C^*) = l)) = 0,$$

whereas

$$(3.29) \quad \begin{aligned} \mathbb{P}[T_{\beta_n, n}(C) < n/2] &= \mathbb{P}[T_{0, \beta_n}(C) > n/2] \\ &\leq 2n^{-1} \mathbb{E}T_{0, \beta_n}(Z) \\ &\quad + d_{TV}(\mathcal{L}(C[1, \beta_n]), \mathcal{L}(Z[1, \beta_n])) \rightarrow 0, \end{aligned}$$

from the uniform logarithmic condition and (3.27). Finally, by the conditioning relation,

$$(3.30) \quad \mathcal{L}(C[\beta_n + 1, n] \mid T_{\beta_n, n}(C) = l) = \mathcal{L}(Z[\beta_n + 1, n] \mid T_{\beta_n, n}(Z) = l).$$

In view of (3.28) and (3.30) with Z^* for Z , the main item still to be considered is $\mathcal{L}(K_{bn}(Z^*) \mid T_{bn}(Z^*) = l)$ for $b = \beta_n$ and $n/2 \leq l \leq n$; the rest is just tidying.

Now the unconditional distribution of $K_{bn}(Z^*)$ is $\text{Po}(\lambda_{bn})$, where

$$(3.31) \quad \lambda_{bn} = \sum_{i=b+1}^n \theta/i \sim \theta \log(n/b).$$

Conditional on $K_{bn}(Z^*) = s$, $Z^*[b+1, n]$ has the multinomial MN ($s; p_{b+1}, \dots, p_n$) distribution, where $p_r = \theta/(r\lambda_{bn})$, $b+1 \leq r \leq n$. Thus, conditional on $K_{bn}(Z^*) = s$, $T_{bn}(Z^*)$ has the distribution of $W_s := \sum_{j=1}^s U_j$, where the U_j

are independent and identically distributed and $\mathbb{P}[U_j = r] = p_r, b+1 \leq r \leq n$. Hence, using Bayes' theorem, we deduce that, for $n/2 \leq l \leq n$,

$$\begin{aligned}
 & \mathbb{P}[K_{bn}(Z^*) = s \mid T_{bn}(Z^*) = l] \\
 (3.32) \quad &= \frac{\text{Po}(\lambda_{bn})\{s\} \mathbb{P}[W_s = l]}{\mathbb{P}[T_{bn}(Z^*) = l]} \\
 &= \text{Po}(\lambda_{bn})\{s\} \frac{(s\theta/\lambda_{bn})\mathbb{P}[W_{s-1} \leq l - b - 1]}{\theta\mathbb{P}[T_{bn}(Z^*) \leq l - b - 1]},
 \end{aligned}$$

where the denominator comes from (2.15) and the numerator is because

$$(3.33) \quad \mathbb{E}\{W_s I[W_s = l]\} = s \sum_{r=b+1}^n r p_r \mathbb{P}[W_{s-1} = l - r].$$

Hence, since $(s/\lambda_{bn})\text{Po}(\lambda_{bn})\{s\} = \text{Po}(\lambda_{bn})\{s - 1\}$, and writing $\lambda_1 = \lambda_{bn} - \lambda_{bn}^{3/4}$, $\lambda_2 = \lambda_{bn} + \lambda_{bn}^{3/4}$, we find that

$$\begin{aligned}
 & d_{TV}(\mathcal{L}(K_{bn}(Z^*)|T_{bn}(Z^*) = l), 1 + \text{Po}(\lambda_{bn})) \\
 (3.34) \quad & \leq \text{Po}(\lambda_{bn})\{[0, \lambda_1] \cup (\lambda_2, \infty)\} \\
 & + \sup_{0 \leq b \leq \beta_n} \sup_{s \in (\lambda_1, \lambda_2)} \left\{ \frac{d_K(\mathcal{L}(W_s), \mathcal{L}(T_{bn}(Z^*)))}{\mathbb{P}[n^{-1}T_{bn}(Z^*) \leq 1/2 - (b+1)/n]} \right\} \rightarrow 0,
 \end{aligned}$$

uniformly in $n/2 \leq l \leq n$, by Corollary 4.3 and Theorem 2.4, since $\beta_n + 1 < n/4$ for all n sufficiently large. Combining (3.28), (3.30) and (3.34), it follows that

$$(3.35) \quad \lim_{n \rightarrow \infty} \sup_{n/2 \leq l \leq n} d_{TV}(\mathcal{L}(K_{\beta_n, n}(C)|T_{\beta_n, n}(C) = l), 1 + \text{Po}(\lambda_{\beta_n, n})) = 0;$$

furthermore, from (3.29) and (3.30), this also implies that

$$(3.36) \quad \lim_{n \rightarrow \infty} \sup_{n/2 \leq l \leq n} d_{TV}(\mathcal{L}(K_{\beta_n, n}(C)|T_{\beta_n, n}(C) = l), \mathcal{L}(K_{\beta_n, n}(Z))) = 0.$$

To conclude the proof, let $p_{kt}(Y)$ denote $\mathbb{P}[K_{0b}(Y) = k, T_{0b}(Y) = t]$; then, again with $b = \beta_n$, we have

$$\begin{aligned}
 & 2d_{TV}(\mathcal{L}(K_{0n}(C)), \mathcal{L}(K_{0n}(Z))) \\
 & \leq \sum_{k \geq 0} \sum_{t \geq 0} \sum_{s \geq 0} \left| \mathbb{P}[K_{0b}(C) = k, T_{0b}(C) = t, K_{bn}(C) = s] \right. \\
 & \quad \left. - \mathbb{P}[K_{0b}(Z) = k, T_{0b}(Z) = t] \mathbb{P}[K_{bn}(Z) = s] \right| \\
 & \leq \sum_{k \geq 0} \sum_{t \geq 0} p_{kt}(C) \sum_{s \geq 0} \left| \mathbb{P}[K_{bn}(C) = s \mid K_{0b}(C) = k, T_{0b}(C) = t] \right. \\
 & \quad \left. - \mathbb{P}[K_{bn}(Z) = s] \right| \\
 & + 2d_{TV}(\mathcal{L}(K_{0b}(C), T_{0b}(C)), \mathcal{L}(K_{0b}(Z), T_{0b}(Z))).
 \end{aligned}$$

The latter contribution is negligible, by (3.27). In the former, by the conditional relation, we have

$$\mathbb{P}[K_{bn}(C) = s \mid K_{0b}(C) = k, T_{0b}(C) = t] = \mathbb{P}[K_{bn}(C) = s \mid T_{bn}(C) = n - t],$$

leading to a contribution of at most

$$2 \sup_{n/2 \leq l \leq n} d_{TV}(\mathcal{L}(K_{bn}(C) \mid T_{bn}(C) = l), \mathcal{L}(K_{bn}(Z))) + \mathbb{P}[T_{0b}(C) > n/2],$$

which is also negligible, by (3.36) and (3.29). This proves (3.24).

For the last part, we apply the Stein–Chen method to show that

$$(3.37) \quad \lim_{n \rightarrow \infty} d_{TV}(\mathcal{L}(K_{0n}(Z)), \text{Po}(\theta \log n)) = 0.$$

If $\sum_{s \geq 1} s^2 c_s < \infty$, we can use Barbour and Hall [(1984), Theorem 4, inequality (4.4)] directly, since then $\sum_{1 \geq 1} (\mathbb{E}Z_i)^2 < \infty$ and $\sum_{i=1}^n \mathbb{E}\{Z_i(Z_i - 1)\} = o(\log n)$. If not, for any bounded $g: \mathbb{Z}_+ \rightarrow \mathbb{R}$, writing $\lambda = \sum_{i=1}^n \mathbb{P}[Z_i = 1]$, $K = K_{0n}(Z)$ and $K_i = K - Z_i$, we have

$$\begin{aligned} & \mathbb{E}\{\lambda g(K + 1) - K g(K)\} \\ &= \sum_{i=1}^n \left\{ \mathbb{P}[Z_i = 1] \mathbb{E}(g(K + 1) - g(K_i + 1)) - \sum_{r \geq 2} \mathbb{P}[Z_i = r] \mathbb{E}g(K_i + r) \right\}, \end{aligned}$$

so that

$$(3.38) \quad \begin{aligned} & |\mathbb{E}\{\lambda g(K + 1) - K g(K)\}| \\ & \leq \sum_{i=1}^n \left\{ \mathbb{P}[Z_i = 1] \mathbb{E}Z_i M_1(g) + \mathbb{P}[Z_i \geq 2] M_0(g) \right\}, \end{aligned}$$

where $M_0(g) = \sup_{j \geq 1} |g(j)|$ and $M_1(g) = \sup_{j \geq 1} |g(j + 1) - g(j)|$. Now $\lambda \sim \theta \log n$, and the test functions g appearing in the Stein–Chen argument for total variation approximation satisfy $M_0(g) \leq \lambda^{-1/2}$ and $M_1(g) \leq \lambda^{-1}$. Thus, for these functions, (3.38) is of order $\{\log n\}^{-1/2} \sum_{i=1}^n i^{-1} e(i)$, by the uniform logarithmic condition, and (3.37) follows under the additional condition on $e(i)$.

4. Technical complements.

PROOF OF PROPOSITION 1.1. We show that if the $\{Z_i\}$ are Poisson, negative binomial or binomial random variables which satisfy the logarithmic condition, then they also satisfy the uniform logarithmic condition.

When $Z_i \sim \text{Po}(\theta_i)$ or $Z_i \sim \text{Bi}(m_i, p_i)$, observe that the logarithmic condition implies that, for $s \geq 2$,

$$i \mathbb{P}[Z_i = s] \leq \frac{i}{s!} (\mathbb{E}Z_i)^s \leq \frac{\theta_*^s}{s! i^{s-1}} \leq i^{-1} c_s,$$

where $c_s = \theta_*^s / s!$ and $\theta_* = \sup_{i \geq 1} i \mathbb{E}Z_i < \infty$; the uniform logarithmic condition follows automatically, with $e(i) = \max(i^{-1}, \sup_{j \geq i} |j \mathbb{P}[Z_j = 1] - \theta|)$.

When $Z_i \sim \text{NB}(m_i, p_i)$, the argument is somewhat more complicated. If $m_i \geq 1$ and $s \geq 2$, we have

$$i\mathbb{P}[Z_i = s] \leq \begin{cases} \frac{i(m_i p_i)^s}{s!} \left(1 + \frac{s}{m_i}\right)^s \leq \frac{i}{s!} (2m_i p_i)^s, & \text{if } s \leq m_i; \\ ip_i^s 2^{m_i+s-1} \leq i(4p_i)^s \leq i(4m_i p_i)^s, & \text{if } s > m_i, \end{cases}$$

and $i\mathbb{P}[Z_i = s] \leq im_i p_i^s$ if $m_i < 1$; thus, whatever the value of m_i , we have

$$i\mathbb{P}[Z_i = s] \leq i(4\theta_*/i)^s + \theta_* p_i^{s-1},$$

with θ_* as before. Now the logarithmic condition implies that

$$\mathbb{P}[Z_i = 2]/\mathbb{P}[Z_i = 1] = \frac{1}{2}(m_i + 1)p_i \rightarrow 0$$

as $i \rightarrow \infty$, so that $p_i \rightarrow 0$ also. Hence, for $s \geq 2$ and $i > i_1 = \max(8\theta_*, i_0)$, where $i_0 = \max\{i: p_i > 1/2\}$, we have $i\mathbb{P}[Z_i = s] \leq e_0(i)c_s$ with $e_0(i) = \max(i^{-1}, p_i)$ and $c_s = 5\theta_* 2^{-(s-2)}$; hence $|i\mathbb{P}[Z_i = s] - \delta_{s1}\theta| \leq e(i)c_s$ for all $i > i_1$ and $s \geq 1$, with $e(i) = \sup_{j \geq i} \max(|j\mathbb{P}[Z_j = 1] - \theta|, e_0(j))$. The extension to all $i \geq 1$ is immediate, because $\mathbb{E}Z_i^2 < \infty$ for all i . \square

LEMMA 4.1. *For the Ewens sampling formula, we have, for any $0 \leq b \leq n$,*

- (i) $\sum_{i=b+1}^n \mathbb{E}C_i^{*(n)} \leq 2\theta \log(n/b) + 1$;
- (ii) $\sum_{i=b+1}^n e(i)\mathbb{E}C_i^{*(n)} \leq 2\theta \sum_{i=b+1}^{[n/2]} i^{-1}e(i) + e([n/2])$.

PROOF. It follows from the Feller coupling [Barbour and Tavaré (1994), Proposition 1.1] that

$$(4.1) \quad \mathbb{E}C_i^{*(n)} \leq \frac{\theta}{i} + \frac{\theta}{\theta + n - i} \leq \frac{2\theta}{i}, \quad 1 \leq i \leq n/2,$$

whereas, since $\sum_{i=1}^n iC_i^{*(n)} = n$, it is always the case that $\sum_{i=[n/2]+1}^n \mathbb{E}C_i^{*(n)} \leq 1$. The lemma now follows immediately. \square

LEMMA 4.2. *Let X, Y and Z be random variables. Let copies \widehat{X} and \widehat{Y} of X and Y be constructed on a common probability space $(\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{\mathbb{P}})$, and let $A \in \widehat{\mathcal{F}}$ be such that $V = |\widehat{X} - \widehat{Y}|I[A^c]$ satisfies $\widehat{\mathbb{E}}V < \infty$. Let Q_Z denote the concentration function of $\mathcal{L}(Z)$. Then*

$$d_K(\mathcal{L}(X), \mathcal{L}(Y)) = O\left(Q_Z\left(\sqrt{\widehat{\mathbb{E}}V}\right) + d_K(\mathcal{L}(Y), \mathcal{L}(Z)) + \widehat{\mathbb{P}}(A)\right).$$

PROOF. The argument is routine: for instance, given any $t \in \mathbb{R}$ and $\varepsilon > 0$, it follows that

$$\begin{aligned} \mathbb{P}[X \leq t] &\leq \widehat{\mathbb{P}}[\widehat{Y} \leq t + \varepsilon] + \widehat{\mathbb{P}}[|\widehat{X} - \widehat{Y}| > \varepsilon] \\ &\leq d_K(\mathcal{L}(Y), \mathcal{L}(Z)) + \mathbb{P}[Z \leq t] + Q_Z(\varepsilon) + \widehat{\mathbb{P}}[A] + \varepsilon^{-1}\widehat{\mathbb{E}}V \\ &\leq 2d_K(\mathcal{L}(Y), \mathcal{L}(Z)) + \mathbb{P}[Y \leq t] + Q_Z(\varepsilon) + \widehat{\mathbb{P}}[A] + \varepsilon^{-1}\widehat{\mathbb{E}}V; \end{aligned}$$

choose $\varepsilon = \sqrt{\widehat{\mathbb{E}}V}$, and note that $x = O(Q_Z(x))$ as $x \rightarrow 0$. \square

COROLLARY 4.3. For any $0 \leq b < n$, $s \geq 1$, let $W_s^{bn} = \sum_{i=1}^s U_i$, where $(U_l, l \geq 1)$ are independent and identically distributed with $\mathbb{P}[U_l = r] = \theta/(r\lambda_{bn})$, $b + 1 \leq r \leq n$, and where $\lambda_{bn} = \sum_{i=b+1}^n \theta/i$; let $S \sim \text{Po}(\lambda_{bn})$ be independent of the U_l . Then, if $B_n = o(n)$ as $n \rightarrow \infty$,

$$\lim_{n \rightarrow \infty} \max_{0 \leq b \leq \beta_n} \max_{|s - \lambda_{bn}| \leq \lambda_{bn}^{3/4}} d_K(W_s^{bn}, W_S^{bn}) = 0.$$

PROOF. Apply Lemma 4.2 with $X = n^{-1}W_s^{bn}$, $Y = n^{-1}W_S^{bn}$, $A = \emptyset$ and $Z = X_\theta$, noting that

$$\begin{aligned} n^{-1}\mathbb{E}\left|\sum_{l=1}^s U_l - \sum_{l=1}^S U_l\right| &\leq n^{-1}\mathbb{E}|S - s|\mathbb{E}U_1 \\ &\leq \{\mathbb{E}|S - \mathbb{E}S| + |\mathbb{E}S - s|\}\theta(1 - b/n)\lambda_{bn}^{-1} \\ &= O\left(\lambda_{bn}^{-1/2} + \lambda_{bn}^{-1}|s - \lambda_{bn}|\right). \end{aligned}$$

For $|s - \lambda_{bn}| \leq \lambda_{bn}^{3/4}$, this gives

$$d_K(\mathcal{L}(W_s^{bn}), \mathcal{L}(W_S^{bn})) = O(Q_\theta(\lambda_{bn}^{-1/8})) + d_K(\mathcal{L}(n^{-1}W_S^{bn}), \mathcal{L}(X_\theta)).$$

But W_S^{bn} has the same compound Poisson distribution as $T_{bn}(Z^*)$, and

$$\lim_{n \rightarrow \infty} \max_{0 \leq b \leq B_n} d_K(\mathcal{L}(n^{-1}T_{bn}(Z^*)), \mathcal{L}(X_\theta)) = 0,$$

by Theorem 2.4. Since also $\lambda_{B_n, n} \rightarrow \infty$ as $n \rightarrow \infty$, the proof is complete. \square

REFERENCES

ALDOUS, D. J. (1985). Exchangeability and related topics. *Lecture Notes in Math.* **17** 1–198. Springer, New York.

ARRATIA, R., BARBOUR, A. D. and TAVARÉ, S. (1993). On random polynomials over a finite field. *Math. Proc. Cambridge Philos. Soc.* **114** 347–368.

ARRATIA, R., BARBOUR, A. D. and TAVARÉ, S. (1999). On Poisson–Dirichlet limits for random decomposable combinatorial structures. *Combin. Probab. Comput.* **8** 193–208.

ARRATIA, R., BARBOUR, A. D. and TAVARÉ, S. (2000). Logarithmic combinatorial structures. Unpublished manuscript.

- ARRATIA, R. and TAVARÉ, S. (1994). Independent process approximations for random combinatorial structures. *Adv. Math.* **104** 90–154.
- BARBOUR, A. D. and HALL, P. G. (1984). On the rate of Poisson convergence. *Math. Proc. Cambridge Philos. Soc.* **95** 473–480.
- BARBOUR, A. D. and TAVARÉ, S. (1994). A rate for the Erdős–Turán law. *Combin. Probab. Comput.* **3** 167–176.
- DE LAURENTIS, J. M. and PITTEL, B. (1985). Random permutations and Brownian motion. *Pacific J. Math.* **119** 287–301.
- ERDÖS, P. and TURÁN, P. (1967). On some problems of statistical group theory III. *Acta Math. Acad. Sci. Hungar.* **18** 309–320.
- EWENS, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoret. Population Biol.* **3** 87–112.
- FLAJOLET, P. and SORIA, M. (1990). Gaussian limiting distributions for the number of components in combinatorial structures. *J. Combin. Theory Ser. A* **53** 165–182.
- GOH, W. M. Y. and SCHMUTZ, E. (1993). Random matrices and Brownian motion. *Combin. Probab. Comput.* **2** 157–180.
- HANSEN, J. C. (1989). A functional central limit theorem for random mappings. *Ann. Probab.* **17** 317–332.
- HANSEN, J. C. (1990). A functional central limit theorem for the Ewens Sampling Formula. *J. Appl. Probab.* **27** 28–43.
- HANSEN, J. C. (1993). Factorization in $GF(q^m)[x]$ and Brownian motion. *Combin. Probab. Comput.* **2** 285–299.
- HANSEN, J. C. (1994). Order statistics for decomposable combinatorial structures. *Random Structures Algorithms* **5** 517–533.
- HANSEN, J. C. and SCHMUTZ, E. (1993). How random is the characteristic polynomial of a random matrix? *Math. Proc. Cambridge Philos. Soc.* **114** 507–515.
- HILDEBRAND, A. and TENENBAUM, G. (1993). On a class of differential-difference equations arising in number theory. *J. Anal. Math.* **61** 145–179.
- HWANG, H.-K. (1998a). On convergence rates in the central limit theorems for combinatorial structures. *European J. Combin.* **19** 329–343.
- HWANG, H.-K. (1998b). A Poisson * geometric convolution law for the number of components in unlabelled combinatorial structures. *Combin. Probab. Comput.* **7** 89–110.
- HWANG, H.-K. (1999). Asymptotics of Poisson approximation to random discrete distributions: an analytic approach. *Adv. Appl. Probab.* **31** 448–491.
- MUTAFICIEV, L. R. (1990). Large components and cycles in a random mapping pattern. In *Random Graphs '87* (M. Karoński, J. Jaworski and A. Ruciński, eds.) 189–202. Wiley, New York.
- NICOLAS, J.-L. (1984). A Gaussian law on $F_q[X]$. *Colloq. Math. Soc. János Bolyai* **34** 1127–1162.
- NICOLAS, J.-L. (1985). Distribution statistique de l'ordre d'un élément du groupe symétrique. *Acta Math. Hungar.* **45** 69–84.
- VERSHIK, A. M. and SHMIDT, A. A. (1977). Limit measures arising in the theory of groups. I. *Theory Probab. Appl.* **22** 79–85.
- VERVAAT, W. (1972). *Success Epochs in Bernoulli Trials with Applications in Number Theory*. Math. Centrum, Amsterdam.

R. ARRATIA AND S. TAVARÉ
 DEPARTMENT OF MATHEMATICS
 UNIVERSITY OF SOUTHERN CALIFORNIA
 LOS ANGELES, CALIFORNIA 90089-1113
 E-MAIL: rarratia@math.usc.edu

A. D. BARBOUR
 ABTEILUNG FÜR ANGEWANDTE
 MATHEMATIK
 UNIVERSITÄT ZÜRICH
 WINTERTHURERSTRASSE 190
 CH-8057, ZÜRICH
 SWITZERLAND