Order statistics
for
Decomposable Combinatorial Structures

Jennie C. Hansen
Heriot-Watt University
Edinburgh, SCOTLAND

Summary. In this paper we consider the component structure of decomposable combinatorial objects, both labeled and unlabeled, from a probabilistic point of view. In both cases we show that when the generating function for the components of a structure is a logarithmic function, then the joint distribution of the normalized order statistics of the component sizes of a random object of size n converges to the Poisson-Dirichlet distribution on the simplex $\nabla = \{\{x_i\} : \sum x_i = 1, x_1 \geq x_2 \geq ... \geq 0\}$. This result complements recent results obtained by Flajolet and Soria [9] on the total number of components in a random combinatorial structure.

American Mathematical Society 1980 subject classifications. Primary 60C05; Secondary 60B10.
This research was partially supported by NSF grant DMS-90 099074.

1. Introduction and Examples

A class of combinatorial objects $\mathcal{P}$ is said to be "decomposable" over another class $\mathcal{C}$ of combinatorial objects if each element of $\mathcal{P}$ may be uniquely decomposed into a multiset of elements of $\mathcal{C}$. In the labeled case, such structures are sometimes refered to as partitional complexes. In both the labeled and unlabeled cases, a decomposable structure $(\mathcal{P},\mathcal{C})$ is characterized by equations which relate certain generating functions associated with the classes $\mathcal{P}$ and $\mathcal{C}$. The characterizing equations and some motivating examples are given below. For further discussion, see Goulden and Jackson [11].

In the labeled case, Π_n will denote the set of objects in the class $\mathcal{P}$ which are labeled by the integers $1,2,...,n$ and we have $\mathcal{P}=\bigcup_{n\geq 1}\Pi_n$. In many examples Π_n consists of all labeled graphs on n vertices which satisfy some property, and the objects in the unique decomposition of $\pi\in\Pi_n$ are the connected components of π. Let $C_n\subseteq\Pi_n$ denote the objects in Π_n whose decomposition consists of one object (so $C_n\subseteq\mathcal{C}$). It is a feature of partitional complexes that each "component" in the decomposition of $\pi\in\Pi_n$ can be uniquely identified with some element of $\bigcup_{k=1}^{\infty}C_k$. In particular, a component of π of size k and labeled by $1\leq l_1<l_2<...<l_k\leq n$ can be identified with an element of C_k by replacing the label l_i by i for $1\leq i\leq k$. Set $p_n=|\Pi_n|$ for $n\geq 1, p_0=1$ and $c_k=|C_k|$ for $k\geq 1$, then the following equation between exponential generating functions holds

$$\hat{P}(z)=\sum_{n=0}^{\infty}\frac{p_n z^n}{n!}=\exp\left(\sum_{k=1}^{\infty}\frac{c_k z^k}{k!}\right)=\exp\left(\hat{C}(z)\right). \tag{1}$$

In the unlabeled case, Π_n will denote the set of objects in $\mathcal{P}$ of size n, and $\mathcal{P}=\bigcup_{n=1}^{\infty}\Pi_n$. Again, $C_n\subseteq\Pi_n$ will denote the objects in Π_n whose decomposition consists of one object in $\mathcal{C}$. It follows that $\mathcal{C}=\bigcup_{n=1}^{\infty}C_n$. As before, for $n\geq 1$, set $p_n=|\Pi_n|, c_n=|C_n|$, and $p_0=1$ and $c_0=0$. The following equations between ordinary generating functions holds

$$P(z)=\sum_{n\geq 0}p_n z^n=\prod_{n\geq 1}(1-z^n)^{-c_n}=\exp\left(\sum_{n\geq 1}\frac{C(z^n)}{n}\right) \tag{2}$$

where $C(z)=\sum_{n\geq 1}c_n z^n$. Although we use the same notation in the unlabeled case as in the labeled case, it will always be clear which case we are considering.

To illustrate these constructions, we consider some motivating examples. Other examples can be found in Flajolet and Soria [9].

Example 1. Let Π_n be the set of all permutations of the set $\{1, 2, ..., n\}$. Every element of Π_n has a unique decomposition into cycles, so $\mathcal{C}$ consists of all cycles of permutatons in $\mathcal{P} = \bigcup_{n\geq 1} \Pi_n$ and C_n consists of all permutations in Π_n which consist of one cycle. In this case, $\hat{P}(z) = (1-z)^{-1}$ and $\hat{C}(z) = \log(1/(1-z))$.

Example 2. Let Π_n be the set of all mappings of the set $\{1, 2, ..., n\}$ into $\{1, 2, ..., n\}$. Each mapping in Π_n can be represented as a directed graph on n labeled vertices such that the out-degree of each vertex is one (i.e. in the graph which represents $\pi \in \Pi_n$ there is an edge from i to j if $\pi(i) = j$). The graph representing π decomposes uniquely into connected components which are also digraphs with out-degree one. Thus C_n consists of all mappings of $\{1, 2, ..., n\}$ whose representing graph is connected. In this example, $p_n = n^n$ and $c_n = (n-1)! \sum_{k=0}^{n-1} n^k/k!$ (see Bollabas [4]).

Example 3. Let Π_n equal the set of all monic polynomials of degree n over the finite field $GF(q)$. We consider these polynomials to be unlabeled objects of size n. Each such polynomial can be decomposed uniquely into its irreducible factors over $GF(q)$. In this case C_n consists of all irreducible polynomials in Π_n. Clearly, $p_n = q^n$ and $P(z) = (1-qz)^{-1}$. Moebius inversion can be used to determine $C(z)$, the generating function for $c_n = |C_n|$, in terms of $\log P(z)$.

Now suppose that $(\mathcal{P}, \mathcal{C})$ is some "decomposable" structure, labeled or unlabeled, and let μ_n denote the uniform measure on Π_n, i.e. for each $\pi \in \Pi_n$, $\mu_n(\pi) = 1/p_n$. Various random variables defined on Π_n can be investigated in order to obtain information about the decompostion of a typical element of Π_n. Much attention has been focused on the asymptotic distribution of the number of components in a random element of Π_n (see [3],[5],[9],[10], and [19]). Typically the limiting distribution for the number of components depends on the behaviour of the exponential generating function $\hat{C}(z)$, in the labeled case, or the behaviour of the generating function $C(z)$ in the unlabeled case. In this paper we confine our attention to structures for which the (appropriate) generating function for the sequence $\{c_n\}$ is a logarithmic function. Such structures have been investigated by Flajolet and Soria [9]. Their results and the definition of logarithmic function are summarized below.

To define logarithmic functions, we introduce some notation. For $\varrho > 0$ and $\eta > 0$, let $\Delta(\varrho, \eta) = \{z \in C : |z| < \varrho + \eta, z \notin [\varrho, \varrho + \eta]\}$, i.e. $\Delta(\varrho, \eta)$ is a disc with radius $\varrho + \eta$ which is slit along the interval $[\varrho, \varrho + \eta]$.

Definition. Suppose that $G(z)$ is a generating function which is analytic at 0 and which has a unique dominant singularity $\varrho > 0$ on its circle of convergence. We say that $G(z)$ is a *logarithmic function* with multiplier $\theta > 0$ and constant K if near this singularity $G(z) = \theta \log\left(\frac{1}{1-z/\varrho}\right) + R(z)$ where $R(z)$ is analytic on some set $\Delta(\varrho, \eta)$ and $R(z) = K + o(1)$ as $z \to \varrho$ in Δ.

Proposition 1.1. (Flajolet and Soria)

Suppose that the generating function $G(z)$ is a logarithmic function with radius of convergence $\varrho > 0$, multiplier $\theta > 0$ and constant K, then

$$[z^n] \exp(G(z)) \sim \frac{\varrho^{-n} n^{\theta-1} e^K}{\Gamma(\theta)}$$

where $[z^n] \exp(G(z))$ denotes the coefficient of z^n in the power series $\exp(G(z))$.

Theorem 1.2. (Flajolet and Soria)

(i) Suppose $(\mathcal{P}, \mathcal{C})$ is a labeled decomposable structure and suppose that $\hat{C}(z)$ is a logarithmic function with multiplier $\theta > 0$. For each $n > 0$ and $\pi \in \Pi_n$, define $X_n(\pi)$ to be the number of components in the decomposition of π. Then $\frac{X_n - \theta \log n}{\sqrt{\theta \log n}}$ converges in distribution to a standard normal distribution with mean 0 and variance 1.

(ii) Suppose that $(\mathcal{P}, \mathcal{C})$ is an unlabeled decomposable structure and $C(z)$ is a logarithmic function with multiplier $\theta > 0$. For each $n > 0$ and $\pi \in \Pi_n$, define $Y_n(\pi)$ to be the number of components in the decomposition of π. Then $\frac{Y_n - \theta \log n}{\sqrt{\theta \log n}}$ converges in distribution to the standard normal distribution.

Remark. It can be shown that the examples given above satisfy the hypotheses of Theorem 1.2. (see [9]). For random permutations $\theta = 1$, for random mappings, $\theta = 1/2$ and for random polynomials over $GF(q)$, $\theta = 1$.

Flajolet and Soria's results focus on the *number* of components in the decomposition of a random element of Π_n. In this paper we investigate the asymptotic distribution of the sizes of the *large* components in the decomposition of a random element of Π_n. For $m > 0$ and any $n > 0$, define λ_m on Π_n by setting $\lambda_m(\pi)$ equal to the size of the m^{th} largest "component"in the decomposition of π (set $\lambda_m(\pi)$ equal to zero if the decomposition of π has fewer than m components). The sequence $\{\lambda_m(\pi)\}_{m\geq 1}$ consists of the order statistics of the sizes of the components of π. In order to study the limiting distribution of the order statistic sequence, we introduce an infinite sequence space

$$\nabla = \{\{x_i\} : \sum x_i \leq 1, x_1 \geq x_2 \geq ... \geq 0\}.$$

The space ∇ is a subset of the product space $\prod_{i=1}^{\infty}[0,1]$ and the topology on ∇ is the topology induced from the usual product topology on $\prod_{i=1}^{\infty}[0,1]$. Let $\mathcal{F}$ denote the Borel σ-algebra generated by the topology on ∇. For each $n > 0$, define $L_n : \Pi_n \rightarrow \nabla$ by $L_n(\pi) = \{\frac{\lambda_m(\pi)}{n}\}_{m=1}^{\infty}$. Each map L_n induces a measure $\mu_n \circ L_n^{-1}$ on ∇. Our main result is that for structures which satisfy Flajolet and Soria's logarithmic condition, the measures $\mu_n \circ L_n^{-1}$ converge weakly to a nondegenerate measure on ∇. The limiting distribution is the Poisson-Dirichlet distribution which is described in Section 2.

Our result in the labeled case unifies results which are known for particular examples such as random permutations and random mappings (see [1], [20]). In the unlabeled case, Arratia, Barbour, and Tavare [2] have recently obtained, with error estimates, the Poisson-Dirichlet limit for the order statistics of the degree sequence of a random polynomial over a finite field. Their methods are somewhat specific to the case of random polynomials, whereas our result gives a general criterion for convergence to the Poisson-Dirichlet distribution. Our result and the results of Arratia, Barbour, and Tavare can also be applied to the riffle shuffles studied by Diaconis, McGrath, and Pitman [6]. It is interesting to contrast our results with those of Flajolet and Soria. One consequence of Theorem 1.2. is that the number of components of a random element of Π_n is roughly $\theta \log n$. Our results imply that the largest k components of a random element of Π_n are on the order of n, not $n/\log n$. So the sizes of the components of a random π partition the integer n in a very *uneven* way. For the examples given above even more is known. Furthermore, one can normalize a function which counts components of various sizes and obtain a Brownian motion process as a limit (see [7], [12], and [13]).

2. Poisson-Dirichlet distribution

Before investigating the weak convergence of the measures $\mu_n \circ L_n^{-1}$ on ∇ we must describe a particular distribution on ∇. The Poisson-Dirichlet distribution on ∇ was first considered by Kingman (see [15],[16],[17]). It arises as a limiting distribution for certain models in population genetics. It is also arises as the limiting distribution of a certain urn model (see Hoppe [14]). Although the distribution is completely determined by its finite-dimensional distributions, it is not easy or instructive to describe these distributions explicitly. We outline two alternative descriptions of this distribution.

One way to obtain the Poisson-Dirichlet distribution on ∇ is via a Poisson point process construction (for details, see Aldous [1]). Let γ denote the measure on $(0,\infty)$ with density $f(x) = \theta x^{-1}e^{-x}$, and let $Y_1 > Y_2 > ...$ denote the points in a Poisson point process on $(0,\infty)$ with intensity γ. Let $\Sigma = \sum_{n\geq 1} Y_n$ and let $P_n = Y_n/\Sigma$, then $(P_1, P_2, ...) \in \nabla$ and the distribution of the sequence $(P_1, P_2, ...)$ is the Poisson-Dirichlet(θ) distribution on ∇. A key fact which facilitates calculations involving the Poisson-Dirichlet distribution is that the distribution of Σ and the distribution of the sequence $(P_1, P_2, ...)$ are independent.

The Poisson-Dirichlet distribution can also be obtained by first considering a sequence of i.i.d. random variables $Z_1, Z_2, Z_3, ...$ such that each Z_i has a beta distribution with density $f(z) = \theta(1-z)^{\theta-1}$. Now form the sequence $(W_1, W_2, ...)$ where $W_n = Z_n(1-Z_1)(1-Z_2)\cdots(1-Z_{n-1})$ for $n > 1$ and $W_1 = Z_1$. We note that $(W_1, W_2, ...) \in \tilde{\nabla} = \{\{x_i\} : x_i \geq 0, \sum x_i \leq 1\}$ and the distribution of this sequence, G_θ, on $\tilde{\nabla}$ is called the GEM distribution. Finally, define $V : \tilde{\nabla} \to \nabla$ such that $(V\{x_i\})_k$ is the k^{th} largest term in the sequence $\{x_i\} \in \tilde{\nabla}$, then $G_\theta \circ V^{-1}$, the distribution of the sequence $V((W_1, W_2, ...)) = (Q_1, Q_2, ...)$, is the Poisson-Dirichlet(θ) distribution on ∇. For further details of this construction, see Donnelly and Joyce [8].

In order to prove the convergence results in the next section, we use a principle which follows from the second construction of the Poisson-Dirichlet (θ) distribution given above. It is a consequence of the construction above, that if a sequence of measures

P_n converge weakly to G_θ, then $P_n \circ V^{-1}$ converges weakly to $G_\theta \circ V^{-1}$, the Poisson-Dirichlet(θ) distribution on ∇. Furthermore, the following principle also holds. Define $T : \prod_{i=1}^{\infty}[0,1] \to \tilde{\nabla}$ by $(T\{x_i\})_j = x_j(1-x_1)\cdots(1-x_{j-1})$ and let $\nu_\theta^\infty = \nu_\theta \times \nu_\theta \times \cdots$ denote the product measure on $\prod_{i=1}^{\infty}[0,1]$, (i.e. ν_θ^∞ is the distribution of the sequence $(Z_1, Z_2, ...)$. Then T is a continuous map and $G_\theta = \nu_\theta^\infty \circ T^{-1}$. It follows that if a sequence of measures Q_n converges weakly to ν_θ^∞ on $\prod_{i=1}^{\infty}[0,1]$, then $Q_n \circ T^{-1} \circ V^{-1}$ converges weakly to $G_\theta \circ V^{-1}$, the Poisson-Dirichlet(θ) distribution on ∇. For further discussion of this principle for establishing convergence in distribution can be found in Donnelly and Joyce [8].

3. Main Results.

Our first result is a straightforward consequence of Proposition 1.1.

Proposition 3.1. *Let $(\mathcal{P}, \mathcal{C})$ be a labeled partitional complex, and suppose that the generating function $\hat{C}(z) = \sum_{k=1}^{\infty} \frac{c_k z^k}{k!}$ is a logarithmic function with multiplier $\theta > 0$, constant $K > 0$, and radius of convergence $\varrho > 0$. In addition, suppose that $\frac{c_k \varrho^k}{k!} \sim \frac{\theta}{k}$. Then the measures $\mu_n \circ L_n^{-1}$ converge weakly to the Poisson-Dirichlet(θ) distribution on ∇.*

Proof: The first step, is to define a function $\Phi_n : \Pi_n \to \tilde{\nabla}$ which orders the normalized component sizes of $\pi \in \Pi_n$ according to a size-biassed sampling scheme. We begin by defining a sequence of functions $\phi_1, \phi_2, \phi_3, ...$ on $\mathcal{P}$ as follows. For each $\pi \in \Pi_n$, let $\phi_1(\pi)$ equal the size of the component in π which is labeled by a subset which contains 1. Call this component π_1. If $\pi_1 \neq \pi$, let $a_2 = \inf\{k \in Z : 1 \leq k \leq n, k \notin \pi_1\}$ and define $\phi_2(\pi)$ to be the size of the component in π which contains a_2; otherwise, set $\phi_2(\pi) = 0$. If $\phi_2(\pi) \neq 0$, let π_2 denote the component which contains a_2; otherwise let π_2=$\emptyset$. Proceeding inductively, if $\pi_1 \cup \pi_2 \cup \cdots \cup \pi_{i-1} \neq \pi$, let $a_i = \inf\{1 \leq k \leq n : k \notin \pi_1 \cup \cdots \cup \pi_{i-1}\}$ and let $\phi_i(\pi)$ equal the size of the component in π which contains a_i and label this component π_i. Otherwise, let $\phi_i(\pi) = 0$ and set $\pi_i = \emptyset$. Next, for each $n > 0$, define $\Phi_n : \Pi_n \to \tilde{\nabla}$ by $\Phi_n(\pi) = \{\frac{\phi_m(\pi)}{n}\}_{m=1}^{\infty}$. We note that $\Phi_n(\pi)$ is simply a *rearrangement* of the ordered sequence $L_n(\pi)$ and, in particular, $V \circ \Phi_n(\pi) = L_n(\pi)$. Thus $\mu_n \circ \Phi_n^{-1} \circ V^{-1} = \mu_n \circ V^{-1}$

and suffices, by the principle discussed in Section 2, to show that $\mu_n \circ \Phi_n^{-1}$ converges to the GEM distribution G_θ.

We could proceed to establish the convergence of $\mu_n \circ \Phi_n^{-1}$ to G_θ by checking that the finite-dimensional distributions of $\mu_n \circ \Phi_n^{-1}$ converge to those of G_θ. However, although the GEM distribution is elegantly defined as the distribution of the sequence $(Z_1, Z_2(1-Z_1), ...)$, formulae for its finite dimensional distributions are still somewhat complicated due to the dependence of the variables in the sequence $(Z_1, Z_2(1-Z_1), ..)$. Instead, we introduce another map $\tilde{T} : \tilde{\nabla}_1 \to \prod_{i=1}^{\infty}[0,1]$, where $\tilde{\nabla}_1 = \{\{x_i\} \in \tilde{\nabla} : \sum x_i = 1\}$, which allows us to change the problem to one of checking the convergence of a sequence of measures to the product measure ν_θ^∞ on the product space $\prod_{i=1}^{\infty}[0,1]$. This approach is similar to the approach used by Vershik and Schmidt [20] in their study of random permutations. We define $\tilde{T} : \tilde{\nabla}_1 \to \prod_{i=1}^{\infty}[0,1]$ by $\tilde{T}(\{x_j\})_i = x_i(1 - x_1 - \cdots - x_{i-1})^{-1}$ if $\sum_{j=1}^{i-1} x_j < 1$ and $\tilde{T}(\{x_j\})_i = 1$ otherwise. The key features of $\tilde{T}$ are that $T \circ \tilde{T} = id$ on $\tilde{\nabla}_1$ and $V \circ T \circ \tilde{T} \circ \Phi_n = L_n$. Thus, letting P_n denote the induced measure $\mu_n \circ (\tilde{T} \circ \Phi_n)^{-1}$ on $\prod_{i=1}^{\infty}[0,1]$, we have $P_n \circ (V \circ T)^{-1} = \mu_n \circ L_n^{-1}$. Hence, by the convergence principle discussed in Section 2, it suffices to show that $P_n \to \nu_\theta^\infty$ weakly as $n \to \infty$.

To show that the measures P_n converge weakly to ν_θ^∞, it is enough to show that for any $0 < \alpha_1, \alpha_2, ..., \alpha_m < 1$,

$$P_n^m\left(\prod_{i=1}^{m}[0,\alpha_i]\right) \to \nu_\theta^{\infty,m}\left(\prod_{i=1}^{m}[0,\alpha_i]\right) = \prod_{i=1}^{m}\int_0^{\alpha_i} \theta(1-z)^{\theta-1}dz$$

where P_n^m and $\nu_\theta^{\infty,m}$ are the projections of the measures P_n and ν_θ^∞ on the first m coordinates of the product space $\prod_{i=1}^{\infty}[0,1]$. We do this for the case $m = 2$. The argument can be easily generalized for any value of $m \geq 1$.

Fix $0 < \alpha, \beta < 1$. Then

$$P_n^2([0,\alpha] \times [0,\beta]) = \mu_n(\pi \in \Pi_n : \frac{\phi_1(\pi)}{n} \leq \alpha, \frac{\phi_2(\pi)}{n - \phi_1(\pi)} \leq \beta)$$

$$= \sum_{k=1}^{n\alpha} \sum_{j=1}^{\beta(n-k)} \binom{n-1}{k-1}\binom{n-k-1}{j-1} c_k c_j \frac{p_{n-k-j}}{p_n}$$

$$= \sum_{k=1}^{n\alpha} \sum_{j=1}^{\beta(n-k)} \frac{k}{n}\frac{c_k}{k!}\frac{j}{(n-k)}\frac{c_j}{j!}\left[\frac{n!}{p_n}\frac{p_{n-k}}{(n-k)!}\right]\left[\frac{(n-k)!}{p_{n-k}}\frac{p_{n-k-j}}{(n-k-j)!}\right].$$

In order to determine the limit of this expression, we use the fact that $\frac{c_k \varrho^k}{k!} \sim \frac{\theta}{k}$, by hypothesis, and $p_n \sim \frac{n! \varrho^{-n} n^{\theta-1} e^K}{\Gamma(\theta)}$ by Proposition 1.1. So

$$P_n^2([0,\alpha] \times [0,\beta]) = \sum_{k=1}^{n\alpha} \sum_{j=1}^{\beta(n-k)} \theta^2 \frac{1}{n} \frac{1}{(n-k)} \left(1 - \frac{k}{n}\right)^{\theta-1} \left(1 - \frac{j}{n-k}\right)^{\theta-1} (1 + \epsilon(n,k,j)). \tag{3}$$

For $k, j > n/\log n$ we have $|\epsilon(n,k,j)| \leq \epsilon_n$ and $\epsilon_n \to 0$ as $n \to \infty$. Also, for all $k, j, n > 0$, there is a constant C such that $|1 + \epsilon(n,k,j)| < C$, where C is a constant which does not depend on n, k, or j.

The limit of the sum on the right side of (3) can be computed by noting that

(i)

$$\sum_{k > \frac{n}{\log n}}^{n\alpha} \sum_{j > \frac{n}{\log n}} \theta^2 \frac{1}{n} \frac{1}{(n-k)} \left(1 - \frac{k}{n}\right)^{\theta-1} \left(1 - \frac{j}{n-k}\right)^{\theta-1} (1 + \epsilon(n,k,j))$$
$$\sim \int\int_{\Delta_n} \theta^2 (1-z_1)^{\theta-1}(1-z_2)^{\theta-1} dz_1 dz_2 = I_1(n)$$

(ii)

$$\sum_{k \leq \frac{n}{\log n}} \sum_{j=1}^{\beta(n-k)} \frac{1}{n(n-k)} \left(1 - \frac{k}{n}\right)^{\theta-1} \left(1 - \frac{j}{n-k}\right)^{\theta-1} (1 + \epsilon(n,j,k))$$
$$\leq C \int\int_{\Delta_n'} \theta^2 (1-z_1)^{\theta-1}(1-z_2)^{\theta-1} dz_1 dz_2 = I_2(n)$$

(iii)

$$\sum_{k > \frac{n}{\log n}}^{n\alpha} \sum_{j \leq \frac{n}{\log n}} \theta^2 \frac{1}{n(n-k)} \left(1 - \frac{k}{n}\right)^{\theta-1} \left(1 - \frac{j}{n-k}\right)^{\theta-1} (1 + \epsilon(n,j,k))$$
$$\leq C \int\int_{\Delta_n''} \theta^2 (1-z_1)^{\theta-1}(1-z_2)^{\theta-1} dz_1 dz_2 = I_3(n)$$

where

$$\Delta_n = \{(z_1, z_2) : 1/\log n \leq z_1 \leq \alpha, 1/\log n \leq z_2 \leq \beta\}$$
$$\Delta_n^{'} = \{(z_1, z_2) : 0 \leq z_1 \leq \frac{1}{\log n} + \frac{1}{n}, 0 \leq z_2 \leq \beta + \frac{1}{n(1-\alpha)}\}$$
$$\Delta_n^{''} = \{(z_1, z_2) : \frac{1}{\log n} - \frac{1}{n} \leq z_1 \leq \alpha + \frac{1}{n}, 0 \leq z_2 \leq \frac{1}{\log n} + \frac{1}{n(1-\alpha)}\}.$$

It is easy to check that $I_1(n) \to \int_0^\alpha \int_0^\beta \theta^2 (1-z_1)^{\theta-1}(1-z_2)^{\theta-1} dz_1 dz_2$ and both $I_2(n)$ and $I_3(n)$ go to zero as $n \to \infty$. This completes the proof of the proposition.

The corresponding result for the order statistics of unlabeled structures also depends on the asymptotics of Proposition 1.1, but the proof is more complicated than the proof of Proposition 3.1. In particular, the proof of the result depends on two additional lemmas which we prove after giving the proof of the main result.

Theorem 3.2. *Let $(\mathcal{P}, \mathcal{C})$ be an unlabeled decomposable combinatorial structure and suppose that the generating function $C(z) = \sum_{k=1}^{\infty} c_k z^k$ is a logarithmic function with multiplier $\theta > 0$, constant $K > 0$ and radius of convergence $1 > \varrho > 0$. In addition, suppose $c_k \varrho^k \sim \frac{\theta}{k}$. Then the measures $\mu_n \circ L_n^{-1}$ converge weakly to the Poisson- Dirichlet(θ) distribution on ∇.*

Proof: We follow the general approach of previous proof. In the proof of Theorem 3.1, we constructed a function Φ_n which reordered the normalized component sizes according to a size-biassed permutation. The procedure that determined the new ordering involved examining the labels associated with each component of the structure. In this case, there are no labels attached to the components and there is no direct analogue of Φ_n. Although one can still construct a size-biassed permutation of the components (see [2]), we take a slightly different approach which allows us to to use the methods developed in the proof of the previous theorem. We consider certain labelled structures and show that convergence of a sequence of measures on ∇ associated with these new structures is *equivalent* to the convergence of $\mu_n \circ L_n^{-1}$ to the Poisson-Dirchlet(θ) distribution.

Let $\tilde{\Pi}_n$ denote the set of all partitions of the set $\{1, 2, ..., n\}$ into disjoint subsets . We define the probability measure $\tilde{\mu}_n$ on $\tilde{\Pi}_n$ by

$$\tilde{\mu}_n(\tilde{\pi}) = \frac{1}{n! p_n} \prod_{k=1}^{n} \binom{m_k + c_k - 1}{m_k} (k!)^{m_k} (m_k)!$$

where m_k equals the number of sets in the partition $\tilde{\pi}$ of size k. The vector $(m_1, m_2, ..., m_n)$ is the type vector of $\tilde{\pi}$. Note that the measure $\tilde{\mu}_n$ on $\tilde{\Pi}_n$ is *not* uniform. It has been constructed so that

$$\tilde{\mu}_n(\tilde{\pi} \in \tilde{\Pi}_n : \tilde{\pi} \text{ has type } (m_1, m_2, ..., m_n)) = \mu_n(\pi \in \Pi_n : \alpha_i(\pi) = m_i, 1 \le i \le n). \quad (4)$$

and so that for every $\mathbf{m} = (m_1, m_2, ..., m_n)$ such that $\sum_{k=1}^{n} k m_k = n$, the conditional probability $\mu_n(\cdot | A_{\mathbf{m}})$ is uniform on $A_{\mathbf{m}}$, where

$A_{\mathbf{m}} = \{\tilde{\pi} \in \tilde{\Pi}_n : \tilde{\pi}$ has type $\mathbf{m} = (m_1, m_2, ..., m_n)\}$. Define $\tilde{\lambda}_m : \tilde{\Pi}_n \to Z$ by setting $\tilde{\lambda}_m(\tilde{\pi})$ equal to the size of the m^{th} largest set in the partition $\tilde{\pi}$ with $\lambda_m(\tilde{\pi}) = 0$ if $\tilde{\pi}$ is a partition with fewer than m sets. Define $\tilde{L}_n : \tilde{\Pi}_n \to \tilde{\nabla}$ by $\tilde{L}_n(\tilde{\pi}) = \{\frac{\tilde{\lambda}_m(\tilde{\pi})}{n}\}_{m=1}^{\infty}$. The key observation, which follows from (4) and the definitions of L_n and $\tilde{L}_n$, is that the induced measures $\tilde{\mu}_n \circ \tilde{L}_n^{-1}$ and $\mu_n \circ L_n^{-1}$ are the *same* on ∇ (even though μ_n and $\tilde{\mu}_n$ are measures on *different* structures). Thus it suffices to show that $\tilde{\mu}_n \circ \tilde{L}_n^{-1}$ converges to the Poisson-Dirichlet (θ) distribution (where $\tilde{\mu}_n \circ \tilde{L}_n^{-1}$ is the measure induced on ∇ from $(\tilde{\Pi}_n, \tilde{\mu}_n)$ via the map $\tilde{L}_n$).

As in the proof of Proposition 3.1, we define functions $\phi_1, \phi_2, ...$ on each partition space $\tilde{\Pi}_n$ such that $\phi_1(\tilde{\pi})$ equals the size of the set in $\tilde{\pi}$ which contains 1 and let $\tilde{\pi}_1$ denote the set which contains 1. If $\tilde{\pi}_1 \neq \tilde{\pi}$, let $a_2 = \inf\{1 \leq k \leq n : k \notin \tilde{\pi}_1\}$ and let $\tilde{\pi}_2$ denote the set in $\tilde{\pi}$ which contains a_2. Otherwise, set $\tilde{\pi}_2 = \emptyset$. In both cases, define $\phi_2(\tilde{\pi})$ to be equal to the size of $\tilde{\pi}_2$, etc. Define $\Phi_n : \tilde{\Pi}_n \to \tilde{\nabla}$ by $\Phi_n(\tilde{\pi}) = \{\frac{\phi_m(\tilde{\pi})}{n}\}_{m=1}^{\infty}$. Then $V \circ T \circ \tilde{T} \circ \Phi_n = \tilde{L}_n$, where V, T, and $\tilde{T}$ are as defined in Section 2 and the previous proof. Let $\tilde{P}_n$ denote the measure induced on $\prod_{k=1}^{\infty}[0,1]$ from $(\tilde{\Pi}_n, \tilde{\mu}_n)$ via $\tilde{T} \circ \Phi_n$, then $\tilde{P}_n \circ (V \circ T)^{-1} = \tilde{\mu}_n \circ \tilde{L}_n^{-1}$. We show that $\tilde{P}_n$ converges weakly to ν_θ^∞ as $n \to \infty$ and hence $\tilde{\mu}_n \circ \tilde{L}_n^{-1} = \mu_n \circ L_n^{-1}$ converges to $\nu_\theta^\infty \circ (V \circ T)^{-1}$, the Poisson-Dirichlet($\theta$) distribution on ∇. It suffices to show that the projection measures $\tilde{P}_n^m$ converge weakly to $\nu_\theta^{\infty,m}$ for each $m \geq 1$. We do this for $m = 2$. The calculations, though complicated, can be extended in a straightforward manner to handle the general case.

Since $[0,1] \times [0,1]$ is a compact space, the measures $\{\tilde{P}_n^2\}$ are sequentially compact, i.e. every subsequence $\{\tilde{P}_{n_i}^2\}$ has a further subsequence which converges weakly to some probability measure on $[0,1] \times [0,1]$. Thus, to show that $\tilde{P}_n^2$ converges to $\nu_\theta^{\infty,2}$ it suffices to show that each such convergent subsequence converges to $\nu_\theta^{\infty,2}$. To prove this, we show that for $0 < \alpha < \alpha' < 1$ and $0 < \beta < \beta' < 1$, $\tilde{P}_n^2((\alpha, \alpha'] \times (\beta, \beta'])$ converges to $\nu_\theta^{\infty,2}((\alpha, \alpha'] \times (\beta, \beta'])$. (In general, sets of the form $(\alpha, \alpha'] \times (\beta, \beta']$ are not convergence determining. However, in this case, convergence on these sets implies convergence of the measures since the sequence of measures is sequentially compact and since $\nu_\theta^{\infty,2}$ does not have atoms at 0 or 1.)

Let $0 < \alpha < \alpha' < 1$ and $0 < \beta < \beta' < 1$ be fixed. Then

$$\tilde{P}_n^2((\alpha,\alpha'] \times (\beta,\beta']) = \tilde{\mu}_n(\tilde{\pi} \in \tilde{\Pi}_n : \alpha < \frac{\phi_1(\tilde{\pi})}{n} \le \alpha', \beta < \frac{\phi_2(\tilde{\pi})}{n-\phi_1(\tilde{\pi})} \le \beta')$$
$$= \sum_{k>n\alpha}^{n\alpha'} \sum_{j>\beta(n-k)}^{\beta'(n-k)} \tilde{\mu}_n(\tilde{\pi} : \phi_1(\tilde{\pi}) = k, \phi_2(\tilde{\pi}) = j).$$

We determine an expression for $\tilde{\mu}_n(\tilde{\pi} : \phi_1(\tilde{\pi}) = k, \phi_2(\tilde{\pi}) = j)$. First suppose that $k \neq j$, then

$$\tilde{\mu}_n(\tilde{\pi} : \phi_1(\tilde{\pi}) = k, \phi_2(\tilde{\pi}) = j) =$$
$$\sum_{\substack{m_k, m_j \ge 1 \\ s.t. km_k + jm_j \le n}} \tilde{\mu}_n(\tilde{\pi} : \phi_1(\tilde{\pi}) = k, \phi_2(\tilde{\pi}) = j, \tilde{\alpha}_k(\tilde{\pi}) = m_k, \tilde{\alpha}_j(\tilde{\pi}) = m_j) \tag{5}$$

where $\tilde{\alpha}_i(\tilde{\pi})$ is defined to be the number of sets of size i in $\tilde{\pi}$. Now suppose that $m_k, m_j \ge 1$ are fixed with $km_k + jm_j \le n$, then

$$\tilde{\mu}_n(\tilde{\pi} : \phi_1(\tilde{\pi}) = k, \phi_2(\tilde{\pi}) = j, \tilde{\alpha}_k(\tilde{\pi}) = m_k, \tilde{\alpha}_j(\tilde{\pi}) = m_j)$$
$$= \frac{(n-k-j)!}{n! p_n} \binom{n-1}{k-1} \binom{n-k-1}{j-1} k! j! m_k m_j \cdot$$
$$\times \sum \left[\binom{m_k + c_k - 1}{m_k} \binom{m_j + c_j - 1}{m_j} \prod_{\substack{n \ge l \ge 1 \\ l \neq j,k}} \binom{m_l + c_l - 1}{m_l} \right]$$
$$= \frac{kj}{p_n n(n-k)} (m_k + c_k - 1)(m_j + c_j - 1) \sum \left[\binom{m_k + c_k - 2}{m_k - 1} \binom{m_j + c_j - 2}{m_j - 1} \prod_{\substack{n \ge l \ge 1 \\ l \neq j,k}} \binom{m_l + c_l - 1}{m_l} \right]$$

where the sum is over all type vectors with m_k and m_j fixed and $\sum_{l=1}^n l m_l = n$. This equality simplifies to

$$\tilde{\mu}_n(\tilde{\pi} : \phi_1(\tilde{\pi}) = k, \phi_2(\tilde{\pi}) = j, \tilde{\alpha}_k(\tilde{\pi}) = m_k, \tilde{\alpha}_j(\tilde{\pi}) = m_j)$$
$$= \frac{p_{n-k-j}}{p_n} \frac{kj}{n(n-k)} (m_k + c_k - 1)(m_j + c_j - 1) \tilde{\mu}_{n-k-j}(\tilde{\pi} : \tilde{\alpha}_k(\tilde{\pi}) = m_k - 1, \tilde{\alpha}_j(\tilde{\pi}) = m_j - 1).$$

Substituting this expression into (5), we get

$$\tilde{\mu}_n(\tilde{\pi} : \phi_1(\tilde{\pi}) = k, \phi_2(\tilde{\pi}) = j)$$

$$= \frac{p_{n-k-j}}{p_n} \frac{c_k k}{n} \frac{c_j j}{(n-k)} \sum_{\substack{m_k, m_j \geq 1 \\ s.t. k m_k + j m_j \leq n}} \tilde{\mu}_{n-k-j}(\tilde{\pi} : \tilde{\alpha}_k = m_k - 1, \tilde{\alpha}_j = m_j - 1)$$

$$+ \frac{p_{n-k-j}}{p_n} \frac{c_k k}{n} \frac{j}{(n-k)} \sum_{\substack{m_k, m_j \geq 1 \\ s.t. k m_k + j m_j \leq n}} (m_j - 1) \tilde{\mu}_{n-k-j}(\tilde{\pi} : \tilde{\alpha}_k = m_k - 1, \tilde{\alpha}_j = m_j - 1)$$

$$+ \frac{p_{n-k-j}}{p_n} \frac{k}{n} \frac{c_j j}{(n-k)} \sum_{\substack{m_k, m_j \geq 1 \\ s.t. k m_k + j m_j \leq n}} (m_k - 1) \tilde{\mu}_{n-k-j}(\tilde{\pi} : \tilde{\alpha}_k = m_k - 1, \tilde{\alpha}_j = m_j - 1)$$

$$+ \frac{p_{n-k-j}}{p_n} \frac{k}{n} \frac{j}{(n-k)} \sum_{\substack{m_k, m_j \geq 1 \\ s.t. k m_k + j m_j \leq n}} (m_k - 1)(m_j - 1) \tilde{\mu}_{n-k-j}(\tilde{\pi} : \tilde{\alpha}_k = m_k - 1, \tilde{\alpha}_j = m_j - 1).$$

The sums simplify and we have

$$\begin{aligned} \tilde{\mu}_n(\tilde{\pi} : \phi_1(\tilde{\pi}) = k, \phi_2(\tilde{\pi}) = j) = \frac{c_k k}{n} \frac{c_j j}{(n-k)} \frac{p_{n-k-j}}{p_n} + \frac{p_{n-k-j}}{p_n} \frac{c_k k}{n} \frac{j E_{n-j-k}(\tilde{\alpha}_j)}{(n-k)} \\ + \frac{p_{n-k-j}}{p_n} \frac{k E_{n-k-j}(\tilde{\alpha}_k)}{n} \frac{c_j j}{(n-k)} + \frac{p_{n-k-j}}{p_n} \frac{k j E_{n-k-j}(\tilde{\alpha}_k \tilde{\alpha}_j)}{n(n-k)}. \end{aligned} \tag{6}$$

Now suppose that $k = j$, then Lemma 3.4., which is proved below, establishes that there is a positive constant C_θ such that for all large n and $k > \alpha n$,

$$\tilde{\mu}_n(\tilde{\pi} : \phi_1(\tilde{\pi}) = \phi_2(\tilde{\pi}) = k) \leq \tilde{\mu}_n(\tilde{\alpha}_k \geq 2) \leq C_\theta \left(\frac{1}{k}\right)^{2 \wedge (1+\theta/2)}.$$

Thus

$$\lim_{n \to \infty} \sum_{k > \alpha n} \tilde{\mu}_n(\tilde{\pi} : \phi_1(\tilde{\pi}) = \phi_2(\tilde{\pi}) = k) \leq \lim_{n \to \infty} C_\theta \sum_{k > \alpha n} \left(\frac{1}{k}\right)^{2 \wedge (1+\theta/2)} = 0.$$

It follows from this and equation (6) that

$$
\begin{aligned}
\lim_{n\to\infty} \tilde{P}_n^2((\alpha,\alpha']\times(\beta,\beta']) &= \lim_{n\to\infty} \sum_{k>\alpha n}^{\alpha' n} \sum_{j>\beta(n-k)}^{\beta'(n-k)} \tilde{\mu}_n(\phi_1(\tilde{\pi})=k, \phi_2(\tilde{\pi})=j) \\
&= \lim_{n\to\infty} \sum_{k\alpha n}^{\alpha' n} \sum_{\substack{j>\beta(n-k)\\ s.t. j\neq k}}^{\beta'(n-k)} \frac{kc_k}{n}\frac{jc_j}{(n-k)}\frac{p_{n-k-j}}{p_n} \\
&+ \lim_{n\to\infty} \sum_{k>\alpha n}^{\alpha' n} \sum_{\substack{j>\beta(n-k)\\ s.t. j\neq k}}^{\beta'(n-k)} \frac{kc_k}{n}\frac{jE_{n-k-j}(\tilde{\alpha}_j)}{(n-k)}\frac{p_{n-k-j}}{p_n} \\
&+ \lim_{n\to\infty} \sum_{k>\alpha n}^{\alpha' n} \sum_{\substack{j>\beta(n-k)\\ s.t. j\neq k}}^{\beta'(n-k)} \frac{jc_jkE_{n-k-j}(\tilde{\alpha}_k)}{n(n-k)}\frac{p_{n-k-j}}{p_n} \\
&+ \lim_{n\to\infty} \sum_{k>\alpha n}^{\alpha' n} \sum_{\substack{j>\beta(n-k)\\ s.t. j\neq k}}^{\beta'(n-k)} \frac{kjE_{n-k-j}(\tilde{\alpha}_k\tilde{\alpha}_j)}{n(n-k)}\frac{p_{n-k-j}}{p_n}.
\end{aligned} \tag{7}
$$

The first limit on the right side of (7) can be computed by appealing to the asymptotics for p_n and c_n. In particular,

$$
\begin{aligned}
&\lim_{n\to\infty} \sum_{k>\alpha n}^{\alpha' n} \sum_{\substack{j>\beta(n-k)\\ s.t. j\neq k}}^{\beta'(n-k)} \frac{kc_k}{n}\frac{jc_j}{(n-k)}\frac{p_{n-k-j}}{p_n} \\
&= \lim_{n\to\infty} \sum_{k>\alpha n}^{\alpha' n} \sum_{j>\beta(n-k)}^{\beta'(n-k)} \frac{kc_k\varrho^k}{n}\frac{jc_j\varrho^j}{(n-k)}\frac{p_{n-k-j}\varrho^{n-k-j}}{p_n\varrho^n} \\
&= \lim_{n\to\infty} \sum_{k>\alpha n}^{\alpha' n} \sum_{j>\beta(n-k)}^{\beta'(n-k)} \frac{\theta^2}{n(n-k)}\left(1-\frac{k}{n}\right)^{\theta-1}\left(1-\frac{j}{(n-k)}\right)^{\theta-1}(1+\epsilon(n,j,k)) \\
&= \int_\alpha^{\alpha'}\int_\beta^{\beta'} \theta^2(1-z_1)^{\theta-1}(1-z_2)^{\theta-1}dz_1dz_2
\end{aligned}
$$

as in the proof of Proposition 3.1. Note that in computing this limit, we have included some terms in the sum whose contribution is zero in the limit.

Next, it follows from the asymptotics for p_n and c_n and from bounds which are established below in Lemma 3.3 that there exist positive constants B_1 and B_2 such that

$$
\frac{p_{n-k-j}}{p_n}kc_kjE_{n-k-j}(\tilde{\alpha}_j) \le \frac{B_1(1\vee n^{1-\theta})\varrho^j}{(1-\varrho)}\left(1-\frac{k}{n}\right)^{\theta-1}\left(1-\frac{j}{n-k}\right)^{\theta-1}(1+\epsilon(n,k,j))
$$

$$\frac{p_{n-j-k}}{p_n} jc_j k E_{n-k-j}(\tilde{\alpha}_k) \le \frac{B_1(1 \vee n^{1-\theta})\varrho^k}{(1-\varrho)} \left(1 - \frac{k}{n}\right)^{\theta-1} \left(1 + \frac{j}{n-k}\right)^{\theta-1} (1+\epsilon(n,k,j))$$

$$\frac{p_{n-k-j}}{p_n} kj E_{n-k-j}(\tilde{\alpha}_k \tilde{\alpha}_j) \le \frac{B_2(1 \vee n^{1-\theta})\varrho^{j+k}}{(1-\varrho)^2} \left(1 - \frac{k}{n}\right)^{\theta-1} \left(1 + \frac{j}{n-k}\right)^{\theta-1} (1+\epsilon(n,k,j))$$

where $\alpha n < k \le \alpha' n$ and $\beta(n-k) < j \le \beta'(n-k)$. By hypothesis, there is a constant C such that $|(1+\epsilon(n,k,j))| \le C$ for all $n, j, k > 0$. Thus

$$\begin{aligned}
0 &\le \lim_{n\to\infty} \sum_{k>\alpha n}^{\alpha' n} \sum_{\substack{j>\beta(n-k) \\ s.t. j\neq k}}^{\beta'(n-k)} \frac{p_{n-k-j}}{p_n} \frac{kc_k}{n} \frac{jE_{n-k-j}(\tilde{\alpha}_j)}{(n-k)} \\
&\le \lim_{n\to\infty} \frac{B_1(1 \vee n^{1-\theta})}{(1-\varrho)} \sum_{k>\alpha n}^{\alpha' n} \sum_{j>\beta(n-k)}^{\beta'(n-k)} \frac{C\varrho^j}{n(n-k)} \left(1 - \frac{k}{n}\right)^{\theta-1} \left(1 - \frac{j}{n-k}\right)^{\theta-1} \\
&\le \lim_{n\to\infty} \frac{B_1(1 \vee n^{1-\theta})\varrho^{\beta(1-\alpha)n}}{(1-\varrho)} \sum_{k>\alpha n}^{\alpha' n} \sum_{\beta(n-k)}^{\beta'(n-k)} \frac{C}{n(n-k)} \left(1 - \frac{k}{n}\right)^{\theta-1} \left(1 - \frac{j}{n-k}\right)^{\theta-1} \\
&\le \lim_{n\to\infty} \frac{B_1 C(1 \vee n^{1-\theta})\varrho^{\beta(1-\alpha)n}}{(1-\varrho)} \\
&= 0.
\end{aligned}$$

The last inequality follows since

$$\begin{aligned}
\lim_{n\to\infty} \sum_{k>\alpha n}^{\alpha' n} \sum_{j>\beta(n-k)}^{\beta'(n-k)} &\frac{1}{n(n-k)} \left(1 - \frac{k}{n}\right)^{\theta-1} \left(1 - \frac{j}{n-k}\right)^{\theta-1} \\
&= \int_{\alpha}^{\alpha'} \int_{\beta}^{\beta'} (1-z_1)^{\theta-1}(1-z_2)^{\theta-1} dz_1 dz_2 \\
&\le 1.
\end{aligned}$$

Similarly,

$$\lim_{n\to\infty} \sum_{k>\alpha n}^{\alpha' n} \sum_{\substack{j>\beta(n-k) \\ s.t. j\neq k}}^{\beta'(n-k)} \frac{p_{n-k-j}}{p_n} \frac{jc_j k E_{n-k-j}(\tilde{\alpha}_k)}{n(n-k)} = 0$$

$$\lim_{n\to\infty} \sum_{k>\alpha n}^{\alpha' n} \sum_{\substack{j>\beta(n-k) \\ s.t. j\neq k}}^{\beta'(n-k)} \frac{p_{n-k-j}}{p_n} \frac{kjE_{n-k-j}(\tilde{\alpha}_k \tilde{\alpha}_j)}{n(n-k)} = 0.$$

This completes the proof of the theorem.

It remains to prove the bounds which were used in the proof of Theorem 3.2. To obtain bounds for expectations with respect to $\tilde{\mu}_n$ we construct a transform for computing such expectations. The transform itself is analogous to a transform construction used by Shepp and Lloyd [18] in the study of the cycle structure of random permutations. Let $\Omega = \{\{m_k\} : m_k \in Z, m_k \geq 0, k = 1, 2, ...\}$ and for $0 < z < 1$, define the product measure P_z on Ω so that for each $i \geq 0$ and $k \geq 1$,

$$P_z(\omega \in \Omega : m_k = i) = \binom{i + c_k - 1}{i}(1 - (\varrho z)^k)^{c_k}(\varrho z)^{ki}.$$

In other words, the distribution of the k^{th} coordinate of an element of Ω with respect to P_z is negative binomial with parameters $p = (1 - (\varrho z)^k)$, $q = (\varrho z)^k$ and $r = c_k$. Now define $\nu : \Omega \to Z$ by $\nu(\omega) = \sum_{k=1}^{\infty} km_k$, then ν is finite a.s. with respect to P_z provided $0 < z < 1$. To see this, we compute the probability generating of ν,

$$E_z(u^\nu) = \prod_{k=1}^{\infty} E_z(u^{km_k}) = \prod_{k=1}^{\infty} \frac{(1 - (\varrho z)^k)^{c_k}}{(1 - (\varrho zu)^k)^{c_k}} = \frac{P(\varrho zu)}{P(\varrho z)}.$$

The first equality follows from the independence of the coordinates of elements of Ω with respect to P_z. The last equality follows from the basic equation (2) which characterizes unlabeled structures. Evaluating $P(\varrho zu)/P(\varrho z)$ at $u = 1$ shows that $P_z(\nu < \infty) = 1$, and for $k \geq 0$,

$$P_z(\nu = k) = \frac{p_n \varrho^n}{P(\varrho z)}.$$

The space (Ω, P_z) is useful because we can recover the joint distribution of the variables $\tilde{\alpha}_1, \tilde{\alpha}_2, ..., \tilde{\alpha}_n$ with respect to $\tilde{\mu}_n$ by conditioning on the value of the variable ν. In particular,

$$\begin{aligned} P_z((m_1, m_2, ...)|\nu = n) &= \frac{\prod_{k=1}^{\infty} \binom{m_k + c_k - 1}{m_k}(1 - (\varrho z)^k)^{c_k}(\varrho z)^{km_k}}{p_n \varrho^n (P(\varrho z))^{-1}} \\ &= \frac{(\varrho z)^n \prod_{k=1}^{\infty}(1 - (\varrho z)^k)^{c_k} \prod_{k=1}^{n} \binom{m_k + c_k - 1}{m_k}}{p_n (\varrho z)^n (P(\varrho z))^{-1}} \\ &= \frac{\prod_{k=1}^{n} \binom{m_k + c_k - 1}{m_k}}{p_n} \\ &= \tilde{\mu}_n(\tilde{\alpha}_1 = m_1, ..., \tilde{\alpha}_n = m_n). \end{aligned} \tag{8}$$

Now suppose that $\Psi : \Omega \to R$, then Ψ induces a map $\Psi_n : \tilde{\Pi}_n \to R$ which is defined by $\Psi_n(\tilde{\pi}) = \Psi(\tilde{\alpha}_1(\tilde{\pi}), ..., \tilde{\alpha}_n(\tilde{\pi}), ...)$. It follows that

$$\begin{aligned} E_z(\Psi) &= \sum_{n=0}^{\infty} E_z(\Psi|\nu = n) P_z(\nu = n) \\ &= \sum_{n=1}^{\infty} E_n(\Psi_n) \frac{p_n \varrho^n}{P(\varrho z)} + \frac{\Psi(0)}{P(\varrho z)} \end{aligned}$$

where E_n is the expectation with respect to $\tilde{\mu}_n$ on $\tilde{\Pi}_n$. The second equation follows from (8) and the defintion of Ψ_n. Thus

$$E_n(\Psi_n) = \frac{[z^n]P(\varrho z)}{p_n \varrho^n} E_z(\Psi). \tag{9}$$

So we can compute $E_n(\Psi_n)$ by first computing $E_z(\Psi)$ and then extracting the coefficient of z^n from the product $\frac{P(\varrho z)}{p_n \varrho^n} E_z(\Psi)$. When computing $E_z(\Psi)$ we can exploit the independence of the coordinates in the space Ω with respect to the measure P_z.

Lemma 3.3. *Let the sets $\tilde{\Pi}_n$, the measures $\tilde{\mu}_n$, and the functions $\tilde{\alpha}_m$ be as defined above. Suppose that the hypotheses of Theorem 3.2. hold, and in particular, that there are positive constants A_1, A_2,and A_3 such that $A_1 n^{\theta-1} \leq p_n \varrho^n \leq A_2 n^{\theta-1}$ and $c_n \varrho^n \leq A_3/n$ for all $n \geq 1$. Finally, suppose that $k_1, k_2, ..., k_l$ are integers such that $1 \leq k_i$ for $i = 1, ..., l$ and $k_1 + k_2 + .. + k_l = M' \leq M \leq n - 1$ for some integer $n > 1$. Then there exists a constant B_l which depends on the constants A_1, A_2, A_3, and l such that*

$$E_{n-M}(\tilde{\alpha}_{k_1} \tilde{\alpha}_{k_2} \cdots \tilde{\alpha}_{k_l}) \leq \frac{B_l}{(1-\varrho)^l} \frac{(1 \vee n^{1-\theta})}{k_1 k_2 \cdots k_l}$$

where E_{n-M} denotes the expectation with respect to $\tilde{\mu}_{n-M}$.

Proof: Note that if $k_1 + k_2 + \cdots + k_l > n - M$, then $E_{n-M}(\tilde{\alpha}_{k_1} \cdots \tilde{\alpha}_{k_l}) = 0$. Now suppose that $k_1 + \cdots + k_l \leq n - M$. We use the transform to compute $E_{n-M}(\tilde{\alpha}_{k_1} \cdots \tilde{\alpha}_{k_l})$.

$$\begin{aligned} E_{n-M}(\tilde{\alpha}_{k_1} \cdots \tilde{\alpha}_{k_l}) &= \frac{[z^{n-M}]P(\varrho z)}{p_{n-M} \varrho^{n-M}} E_z(m_{k_1} \cdots m_{k_l}) \\ &= \frac{[z^{n-M}]P(\varrho z)}{p_{n-M} \varrho^{n-M}} \prod_{i=1}^{l} E_{n-M}(m_{k_i}) \\ &= \frac{[z^{n-M}]P(\varrho z)}{p_{n-M} \varrho^{n-M}} \prod_{i=1}^{l} \frac{c_{k_i} (\varrho z)^{k_i}}{(1 - (\varrho z)^{k_i})}. \end{aligned} \tag{10}$$

The second equality follows from the independence of the coordinates $m_{k_1},...,m_{k_l}$ with respect to P_z. The product on the right side of (10) is a power series with positive coefficients. We write $\prod_{i=1}^{l} c_{k_i}(\varrho z)^{k_i}(1-(\varrho z)^{k_i})^{-1} = \sum_{j\geq M'} a_j z^j$. Thus

$$
\begin{aligned}
\frac{[z^{n-M}]P(\varrho z)}{p_{n-M}\,\varrho^{n-M}} \prod_{i=1}^{l} \frac{c_{k_i}(\varrho z)^{k_i}}{(1-(\varrho z)^{k_i})} &= \sum_{j\geq M'}^{n-M} \frac{a_j p_{n-M-j}\,\varrho^{n-M-j}}{p_{n-M}\,\varrho^{n-M}} \\
&\leq \frac{(A_2 \vee 1)}{A_1} \sum_{j\geq M'} a_j (1 \vee (n-M)^{1-\theta}) \\
&\leq \frac{(A_2 \vee 1)}{A_1} (1 \vee n^{1-\theta}) \sum_{j\geq M'} a_j \\
&= \frac{(A_2 \vee 1)}{A_1} (1 \vee n^{1-\theta}) \prod_{i=1}^{l} c_{k_i}\varrho^{k_i}(1-\varrho^{k_i})^{-1} \\
&\leq \frac{(A_2 \vee 1)}{A_1} \left(\frac{A_3}{1-\varrho}\right)^l \frac{(1 \vee n^{1-\theta})}{k_1 \cdots k_l}.
\end{aligned}
$$

The first and third inequalities follow from the asymptotic bounds for p_n and c_n respectively. The lemma is proved with $B_l = \frac{(A_2 \vee 1)}{A_1}(A_3)^l$.

Remarks. Lemma 3.3. is stated in the generality needed for the extension of the 2-dimensional argument given in the proof of Theorem 3.2. to the m-dimensional case.

Lemma 3.4. *Suppose that the hypotheses of Lemma 3.3. are satisfied. In particular, suppose that* A_1, A_2, A_3 *are constants as in Lemma 3.3. Let the sets* $\tilde{\Pi}_n$*, the measures* $\tilde{\mu}_n$*, and the functions* $\tilde{\alpha}_k$ *be defined as above. Let* M *be such that for all* $k \geq M$, $c_k \varrho^k \leq \frac{A_3}{k} \leq \frac{1}{2}$. *Suppose that* n *is such that* $\frac{\log n}{n^{\theta/2}} \leq 1$, *then for all* $k \geq \frac{n}{\log n} \vee M$,

$$
\tilde{\mu}_n(\tilde{\alpha}_k \geq 2) \leq C_\theta (\frac{1}{k})^{2 \wedge (1+\theta/2)}
$$

where C_θ *is a constant that depends only on the constants* A_1, A_2, A_3 *and the constant* θ.

Proof: We know from the transform that

$$\tilde{\mu}_n(\alpha_k \geq 2) = \frac{[z^n]P(\varrho z)}{p_n \varrho^n} \mu_z(\alpha_k \geq 2)$$

$$= \frac{[z^n]P(\varrho z)}{p_n \varrho^n}(1 - (1 - (\varrho z)^k)^{c_k} - c_k(\varrho z)^k(1 - (\varrho z)^k)^{c_k})$$

$$= \frac{[z^n]P(\varrho z)}{p_n \varrho^n}\left(\sum_{j=2}^{c_k}\binom{c_k}{j}(-1)^{j+1}(\varrho z)^{kj} + c_k(\varrho z)^k \sum_{i=1}^{c_k}\binom{c_k}{i}(-1)^{i+1}(\varrho z)^{ik}\right)$$

$$\leq \sum_{j=1}^{n/k}\binom{c_k}{j}\frac{p_{n-kj}}{p_n} + c_k \sum_{i=1}^{n/k-1}\binom{c_k}{i}\frac{p_{n-k(i+1)}}{p_n}.$$

It follows from the asymptotics for p_n and the asymptotics for $c_k \varrho^k$, that

$$\begin{aligned}
\sum_{j=2}^{n/k}\binom{c_k}{j}\frac{p_{n-kj}}{p_n} &\leq \sum_{j=2}^{n/k}(c_k)^j \frac{p_{n-kj}}{p_n} \\
&\leq \sum_{j=2}^{n/k}\left(\frac{A_3}{k}\right)^j \frac{(A_2 \vee 1)}{A_1}(1 \vee n^{1-\theta}) \\
&\leq \frac{(A_2 \vee 1)}{A_1}\left(\frac{A_3}{k}\right)^2 (1 \vee n^{1-\theta})\frac{1}{1 - A_3/k} \\
&\leq \frac{2(A_2 \vee 1)(A_3)^2}{A_1}\left(\frac{1}{k^2} \vee \frac{n^{1-\theta}}{k^2}\right) \\
&\leq C_1\left(\frac{1}{k^2} \vee \frac{1}{n^{\theta/2}k}\right) \\
&\leq C_1\left(\frac{1}{k}\right)^{(2\wedge(1+\theta/2))}
\end{aligned}$$

where $C_1 = \frac{2(A_2 \vee 1)(A_3)^2}{A_1}$. The fourth inequality follows from the assumption that $A_3/k \leq 1/2$. Similarly,

$$\begin{aligned}
c_k \sum_{i=1}^{n/k-1}\binom{c_k}{i}\frac{p_{n-k(i+1)}}{p_n} &\leq \sum_{i=1}^{n/k-1}(c_k)^{i+1}\frac{p_{n-k(i+1)}}{p_n} \\
&\leq \sum_{i=1}^{n/k-1}\left(\frac{A_3}{k}\right)^{i+1}\frac{(A_2 \vee 1)}{A_1}(1 \vee n^{1-\theta}) \\
&\leq C_1\left(\frac{1}{k}\right)^{(2\wedge(1+\theta/2))}.
\end{aligned}$$

So the lemma is proved with $C_\theta = 2C_1$.

References

[1] Aldous,D. Exchangeability and related topics. Springer Verlag Lecture Notes in Math. **1117** (1985).

[2] Arratia, R., Barbour, A.D., and Tavare, S. *On Random Polynomials over Finite Fields*, To appear in Math. Proc. Camb. Phil. Soc. (1993).

[3] Bender, E. *Central and local limit theorems applied to asymptotic enumeration*, Jour. Comb. Theor. A. **15** (1973) 91-111.

[4] Bollabas, B. Random graphs. Academic Press. (1985).

[5] Compton, K.J. *Some methods for computing component distribution probabilities in relational structures,* Discrete Math. **66** (1987) 59-77.

[6] Diaconis, P., McGrath, M., and Pitman, J. *Riffle shuffles, cycles and descents*, To appear in Combinatorica, (1993).

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

[8] Donnelly, P. and Joyce, P. *Continuity and weak convergence of ranked and size-biassed permutations on an infinite simplex*, Stoch. Proc. Appl. **31** (1989) 89-103.

[9] Flajolet, P. and Soria, M., *Gaussian limiting distributions for the number of components in combinatorial structures,* Jour. Comb. Theor. A. **153** (1990) 165-182.

[10] Goncharov, V. *On the field of combinatory analysis,* Amer. Math. Soc. Trans. Ser. 2, **19** (1962) 1-46.

[11] Goulden, I. and Jackson, D. Combinatorial Enumeration. Academic Press. (1983).

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

[13] Hansen, J.C. *Factorization in GF(q)[x] and Brownian motion,* (1991) In Preparation.

[14] Hoppe, F.M. *The sampling theory of neutral alleles and an urn model in population genetics,* Jour. Math. Biol. **25** (1987) 123-159.

[15] Kingman, J.F.C. *Random discrete distributions,* J. Roy. Statist. Soc. B **37** (1975) 1-22.

[16] Kingman, J.F.C. *The population structure associated with Ewens' sampling formula,* Theor. Popul. Biol. **11** (1977) 274-283.

[17] Kingman, J.F.C. *Random partitions in population genetics,* Proc. R. Soc. Lond. A **361** (1978) 1-20.

[18] Shepp, L.A. and Lloyd, S.P. *Ordered cycle lengths in a random permutation*, Trans. Amer. Math. Soc. **121** (1966) 340-357.

[19] Stepanov, V.E. *Limit distributions for certain characteristics of random mappings.* Theor. Probab. Appl. **14** (1969) 612-626.

[20] Vershik, A.M. and Shmidt, A.A. *Limit measures arising in the asymptotic theory of symmetric groups. I.* Theor. Probab. Appl. **22** (1977) 70-85.