The Annals of Probability
1995, Vol. 23, No. 3, 1347–1388

TOTAL VARIATION ASYMPTOTICS FOR POISSON PROCESS APPROXIMATIONS OF LOGARITHMIC COMBINATORIAL ASSEMBLIES[1]

BY RICHARD ARRATIA, DUDLEY STARK AND SIMON TAVARÉ

University of Southern California

Assemblies are the decomposable combinatorial constructions characterized by the exponential formula for generating functions: $\Sigma p(n)s^n/n! = \exp(\Sigma m_i s^i/i!)$. Here $p(n)$ is the total number of constructions that can be formed from a set of size n, and m_n is the number of these structures consisting of a single component. Examples of assemblies include permutations, graphs, 2-regular graphs, forests of rooted or unrooted trees, set partitions and mappings of a set into itself. If an assembly is chosen uniformly from all possibilities on a set of size n, the counts $C_i(n)$ of components of size i are jointly distributed like independent nonidentically distributed Poisson variables Z_i conditioned on the event $Z_1 + 2Z_2 + \cdots + nZ_n = n$. We consider assemblies for which the process of component-size counts has a nontrivial limit distribution, without renormalizing. These include permutations, mappings, forests of labelled trees and 2-regular graphs, but not graphs and not set partitions. For some of these assemblies, the distribution of the component sizes may be viewed as a perturbation of the Ewens sampling formula with parameter θ.

We consider $d_b(n)$, the total variation distance between $(Z_1, \ldots, Z_b)$ and $(C_1(n), \ldots, C_b(n))$, counting components of size at most b. If the generating function of an assembly satisfies a mild analytic condition, we can determine the decay rate of $d_b(n)$. In particular, for $b = b(n) = o(n/\log n)$ and $n \to \infty$, $d_b(n) = o(b/n)$ if $\theta = 1$ and $d_b(n) \sim c(b)b/n$ if $\theta \neq 1$. The constant $c(b)$ is given explicitly in terms of the m_i: $c(b) = |1 - \theta|\mathbb{E}|T_{0b} - \mathbb{E}T_{0b}|/(2b)$, where $T_{0b} = Z_1 + 2Z_2 + \cdots + bZ_b$. Finally, we show that for $\theta \neq 1$ there is a constant c_θ such that $c(b) \sim c_\theta b$ as $b \to \infty$. Our results are proved using coupling, large deviation bounds and singularity analysis of generating functions.

1. Introduction. We consider assemblies, which are combinatorial constructions on a finite set in which the set is partitioned into blocks (i.e., subsets) and each block is assigned additional structure to form a "component." The defining property of an assembly is that the number m_i of possible structures available to form a component from a given block of i elements does not vary with either the choice of elements within the i-block or with the remaining components.

Received December 1993; revised October 1994.

[1]The authors were supported in part by NSF Grant DMS-90-05833.

AMS 1991 *subject classifications.* 60C05, 60F17, 05A05, 05A16.

Key words and phrases. Singularity analysis, Ewens sampling formula, assemblies, species, random mappings, forests, permutations, Poisson approximation, functional limit theorems.

For a given assembly, for $n \geq 1$ and $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{Z}_+^n$ we define $N(n, \mathbf{a})$ to be the number of constructions on n points that have a_i components of size i, $i = 1, 2, \dots, n$. For any $\mathbf{a}$ satisfying $a_1 + 2a_2 + \cdots + na_n = n$, we have

$$N(n, \mathbf{a}) = \left(n! \prod_{i=1}^{n} \left(\frac{1}{i!} \right)^{a_i} \frac{1}{a_i!} \right) \left(\prod_{i=1}^{n} m_i^{a_i} \right).$$

The first factor is the number of ways to partition the n-set into blocks of the specified sizes, and the second factor is the number of ways to assign the additional structure to make blocks into components. It follows that for any $\mathbf{a} \in \mathbb{Z}^n$,

$$(1) \qquad N(n, \mathbf{a}) = \mathbf{1}(a_1 + 2a_2 + \cdots + na_n = n) n! \prod_{1}^{n} \left(\frac{m_i}{i!} \right)^{a_i} \frac{1}{a_i!}.$$

The total number of constructions of size n is

$$(2) \qquad p(n) = \sum_{\mathbf{a}} N(n, \mathbf{a}).$$

The characterizing relation for the class of assemblies is the exponential formula

$$(3) \qquad 1 + \sum_{n \geq 1} p(n) s^n / n! = \exp\left(\sum_{i \geq 1} m_i s^i / i! \right),$$

which can be proved directly from (1) and (2). Assemblies are treated in Foata (1974), where they are called "abelian partitional complexes," and in Joyal (1981), where they are called "assemblies of species."

Examples include permutations, for which the components are cycles; graphs, for which the components are connected components in the usual sense; 2-regular graphs, (i.e., graphs in which each vertex has degree 2), for which the components are undirected cycles of length 3 or more; forests, whose components are trees; set partitions, whose components are simply blocks, having no additional structure; and mappings of a finite set into itself, for which the components may be thought of as cycles of rooted labelled trees. Permutations have $m_i = (i-1)!$, $p(n) = n!$; graphs have $p(n) = 2^{n(n-1)/2}$; 2-regular graphs have $m_i = \mathbf{1}(i > 2)(i-1)!/2$; forests of unrooted trees have $m_i = i^{i-2}$, while forests of rooted trees have $m_i = i^{i-1}$; set partitions have $m_i = 1$ for all i and $p(n) = B_n$, the nth Bell number; finally, random mappings have $p(n) = n^n$ and $m_i = (i-1)! \sum_{j=0}^{i-1} i^j / j!$.

We denote the component structure of an assembly on n points by $\mathbf{C} \equiv \mathbf{C}(n) = (C_1(n), C_2(n), \dots, C_n(n))$, where $C_i(n)$ is the number of components of size i. By fixing n with $p(n) > 0$ and choosing uniformly from the $p(n)$ possibilities, we get a stochastic process $\mathbf{C}(n)$ with values in $\mathbb{Z}_+^n$, whose distribution is specified by

$$\mathbb{P}(\mathbf{C}(n) = \mathbf{a}) = \frac{N(n, \mathbf{a})}{p(n)}, \qquad \mathbf{a} \in \mathbb{Z}_+^n.$$

It is natural to define $C_i(n) \equiv 0$ for $i > n$ and to identify $\mathbf{C}(n)$ with the infinite dimensional process $(C_1(n), C_2(n), \dots) = (C_1(n), \dots, C_n(n), 0, 0, \dots)$ with values in $\mathbb{Z}_+^{\mathbb{N}}$. This paper deals with assemblies where the sequence $m_1, m_2, \dots$ is such that a fixed process $(Z_1\, Z_2, \dots)$ can be used to approximate $\mathbf{C}(n)$ for large n.

2. One moment suffices. For combinatorial assemblies, the joint distribution of the component counts is so constrained that convergence of the single first moment $\mathbb{E}C_1(n)$ to a nonzero, finite limit implies distributional convergence for the entire process. The precise statement of this is Theorem 1 below. There is a complementary result given by Corollary 1, that convergence in distribution of $C_1(n)$ to *any* limit implies proper convergence of $\mathbb{E}C_1(n)$, so the limit of $C_1(n)$ is Poisson, and furthermore the entire process converges in distribution, to a prescribed Poisson process. The hypotheses of Theorem 1 and Corollary 1 are given in terms of a simple case, involving just $C_1(n)$, and a more general case involving $C_i(n)$ for all i in a set with greatest common divisor 1. Example 2 at the end of this section shows that these hypotheses cannot be substantially weakened.

Given $x \in (0, \infty)$, let $Z_1, Z_2, \dots$ be independent Poisson random variables with parameters

$$\lambda_i \equiv \mathbb{E}Z_i = \frac{m_i x^i}{i!}. \tag{4}$$

LEMMA 1. *Assume that $x \in (0, \infty)$ and that, as $n \to \infty$,*

$$\frac{xp(n)}{np(n-1)} \to 1. \tag{5}$$

Then the combinatorial process converges in distribution in $\mathbb{R}^\infty$ to the independent Poisson process: $(C_1(n), C_2(n), \dots) \Rightarrow (Z_1, Z_2, \dots)$. This is equivalent to the condition: for all $b = 1, 2, \dots$, as $n \to \infty$,

$$(C_1(n), C_2(n), \dots, C_b(n)) \Rightarrow (Z_1, Z_2, \dots, Z_b). \tag{6}$$

Furthermore, all moments of the combinatorial process converge to those of the Poisson process.

PROOF. We use the method of moments to prove convergence of the finite-dimensional distributions. The joint falling factorial moments of $(C_1(n), \dots, C_b(n))$ are given by

$$\mathbb{E}\prod_{j=1}^{b} (C_j(n))_{r_j} = \mathbf{1}(m \le n)\, x^{-m} \frac{n!}{p(n)} \frac{p(n-m)}{(n-m)!} \prod_{j=1}^{b} (\lambda_j)^{r_j}, \tag{7}$$

for $(r_1, \dots, r_b) \in \mathbb{Z}_+^b$ with $m = r_1 + 2r_2 + \cdots + br_b$, and where we use the notation $(y)_r$ to denote the falling factorial $y(y-1)\cdots(y-r+1)$ [cf. Arratia and Tavaré (1994), (126)]. From this formula it can be seen that the joint moments of $(C_1(n),\dots,C_b(n))$ converge to those of $(Z_1,\dots,Z_b)$ if, and only if, $xp(n) \sim np(n-1)$. Thus (5) implies that all moments converge, and hence $(C_1(n),\dots,C_b(n)) \Rightarrow (Z_1,\dots,Z_b)$. □

Lemma 1 can be useful when $p(n)$ is known directly. It provides a unifying perspective on the convergence of various assemblies for which the process convergence is already well known. For example, permutations have $p(n) = n!$, so that $p(n)/(np(n-1)) = 1$ and Lemma 1 applies with $x = 1$; mappings have $p(n) = n^n$, so that $p(n+1)/((n+1)p(n)) = (1+1/n)^n \to e$ and Lemma 1 applies with $x = 1/e$; forests of labelled trees have asymptotics given by Tannery's formula $p(n) \sim \sqrt{e}\, n^{n-2}$ [see Moon (1970), page 29], so that Lemma 1 applies with $x = 1/e$. For other examples, Flajolet and Soria (1990) use analytic hypotheses on generating functions to derive asymptotics for $p(n)$, which imply (5); we discuss this further in Section 4.2. Examples of assemblies to which Lemma 1 does not apply include all graphs, for which $p(n) = 2^{n(n-1)/2}$ so that $p(n+1)/(np(n)) = 2^n/n \to \infty$, and set partitions, for which $p(n)$, the nth Bell number, satisfies $p(n)/(np(n-1)) \sim 1/\log(n) \to 0$.

THEOREM 1. *Assume that $p(n) > 0$ for all sufficiently large n. If $m_1 > 0$ and*

$$\mathbb{E}C_1(n) \to l_1 \in (0,\infty), \tag{8}$$

then with x defined by the requirement that $l_1 = m_1 x$, Lemma 1 applies, all moments converge, and the combinatorial process converges in distribution to the independent Poisson process with parameters given by (4). More generally, if I is a finite set of positive integers with greatest common divisor 1 and

$$\text{for all } i \in I, \qquad \mathbb{E}C_i(n) \to l_i \in (0,\infty),$$

then there is a unique choice of $x \in (0,\infty)$ such that for all $i \in I$, $l_i = \lambda_i$. Using this x, Lemma 1 applies, all moments converge and the combinatorial process converges in distribution to the independent Poisson process.

PROOF. For any n such that $p(n) > 0$,

$$\mathbb{E}C_i(n) = \mathbf{1}(i \le n)\, x^{-i} \frac{n!}{p(n)} \frac{p(n-i)}{(n-i)!} \lambda_i, \tag{9}$$

which is the special case of (7) with $b = m = i$, $(r_1,\dots,r_b) = (0,\dots,0,1)$. Writing $h(n) \equiv p(n)x^n/n!$, the hypothesis (8) implies that $h(n-1)/h(n) \to 1$, so that Lemma 1 applies, proving the first part of this theorem.

For the second part, let $a(i)$ be integers such that $\sum_{i\in I} i a(i) = 1$. Fix one particular $j \in I$ and define x by the requirement that $l_j = \lambda_j \equiv m_j x^j/j!$. For $i \in I$, let $r(i) \equiv l_i/\lambda_i$, so that $r(j) = 1$, and the hypothesis $\mathbb{E}C_i(n) \to l_i$ implies that $h(n-i)/h(n) \to r(i)$, for each $i \in I$. We can write $h(n-1)/h(n)$ as a telescoping product to see that $h(n-1)/h(n) \to r \equiv \prod_{i\in I} r(i)^{a(i)}$. In more detail, the product has $\sum |a(i)|$ factors. These factors are of the form

$h(n+k-i)/h(n+k) \to r(i)$ if $a(i) > 0$, and of the form $h(n+k+i)/h(n+k) \to r(i)^{-1}$ if $a(i) < 0$. We exploit a telescoping product again to see

$$1 = \lim \frac{h(n-j)}{h(n)} = \lim \frac{h(n-1)}{h(n)} \frac{h(n-2)}{h(n-1)} \cdots \frac{h(n-j)}{h(n-j+1)} = r^j$$

so that $r = 1$ and hence (5) holds. □

COROLLARY 1. *Assume that an assembly is such that $m_1 > 0$ and $C_1(n)$ converges in distribution to some nonzero random variable or, more generally, that for some set $I \subset \{1, 2, \dots\}$ with $\gcd(I) = 1$, for each $i \in I$, $m_i > 0$ and $C_i(n)$ converges in distribution to some nonzero random variable. Then the hypotheses of Theorem 1 are satisfied, so that all moments converge and the entire combinatorial process converges in joint distribution to the appropriate Poisson limit.*

PROOF. We will show that, if $C_i(n)$ converges in distribution, then $\mathbb{E}C_i(n) \to l_i \in (0, \infty)$. Applying this, either for $i = 1$ or else for all i in a set with greatest common divisor 1, shows that the hypotheses of Theorem 1 are satisfied.

Fix i with $m_i > 0$ and assume that $C_i(n) \Rightarrow U$ with density $f(a) = \mathbb{P}(U = a)$. Let $\hat{p}(n)$ be the number of assemblies on n points having no components of size i. Note that

$$\mathbb{P}(C_i(n) = a) = \frac{(n)_{ia}(m_i)^a}{(i!)^a a!} \frac{\hat{p}(n-ai)}{p(n)}, \tag{10}$$

for $a = 0, 1, 2, \dots$.

Thus, if $f(a) > 0$, then as $n \to \infty$,

$$p(n) \sim n^{ia}\hat{p}(n-ia)\left(\frac{m_i}{i!}\right)^a \frac{1}{a!f(a)}. \tag{11}$$

If we also have $f(a+1) > 0$, then

$$p(n) \sim n^{i(a+1)}\hat{p}(n-ia-i)\left(\frac{m_i}{i!}\right)^{a+1} \frac{1}{(a+1)!f(a+1)}, \tag{12}$$

and (11) with n replaced by $n - i$ gives

$$p(n-i) \sim (n-i)^{ia}\hat{p}(n-i-ia)\left(\frac{m_i}{i!}\right)^a \frac{1}{a!f(a)}. \tag{13}$$

Now (12) divided by (13) yields

$$\frac{p(n)}{p(n-i)} \sim n^i \frac{m_i}{i!} \frac{f(a)}{(a+1)f(a+1)}. \tag{14}$$

By formula (9), this implies that as $n \to \infty$, $\mathbb{E}C_i(n) \to (a+1)f(a+1)/f(a) \in (0, \infty)$. In order to show that for some a, both $f(a) > 0$ and $f(a+1) > 0$, we assume that $a > 0$, $f(a) > 0$ and $f(a+1) = 0$, and derive a contradiction.

From (10) with $a + 1$ in place of a, we get

$$\frac{n^{i(a+1)}\hat{p}(n - ai - i)}{p(n)} \to 0. \tag{15}$$

Combining (13) with (15) yields $p(n - i)n^i/p(n) \to 0$, and applying this with n replaced by $n - i, n - 2i, \dots, n - ai + i$ and multiplying, shows that $p(n - ia)n^{ia}/p(n) \to 0$; this uses $a > 0$. Since $\hat{p}(n - ia) \le p(n - ia)$, we have

$$\frac{\hat{p}(n - ia)n^{ia}}{p(n)} \to 0,$$

which combined with (10) implies $\mathbb{P}(C_i(n) = a) \to 0$, contradicting $f(a) > 0$. □

REMARK (Periodicity). The hypothesis in Theorem 1, that $p(n) > 0$ for all sufficiently large n, is equivalent to the assumption that

$$d \equiv \gcd\{i: m_i > 0\}$$

satisfies $d = 1$. By analogy with Markov chains, we refer to assemblies which satisfy this as *aperiodic*. The general case can be reduced to this one by reindexing to create a new "assembly" which is aperiodic, and translating results back to the original assembly. This can be done for the results of all the theorems in this paper. In detail, given a sequence $m_1, m_2, \dots,$ for which $d > 1$, and an appropriate choice x, for which $\lambda_i \equiv m_i x^i/i!$ might satisfy $\mathbb{E}C_i(n) \to \lambda_i$, we cook up a new "assembly," indicated by primes, such that $\lambda'_i = \lambda_{id}$. This means

$$\frac{m'_i(x')^i}{i!} \equiv \lambda'_i = \lambda_{id} \equiv \frac{m_{id}x^{id}}{(id)!}, \quad \text{hence} \quad x' = x^d, \qquad m'_i = m_{id}\frac{i!}{(id)!}.$$

We refer to the new system as an "assembly" in quotation marks because the m'_i need not be integers, so there need not be a combinatorial interpretation in terms of labelled structures. The generating function relation in (3) serves to define the $p(n)$ for this more general "assembly." Equivalently, we may use (26) below as a more general description of "assemblies," requiring only that for all i, $\lambda_i \ge 0$. Our old assembly, with period d, has $p(n) = 0$ unless $n = md$ for an integer m, and $C_i(n) \equiv 0$ unless $d|i$, so that the entire distribution of our old assembly is described via

$$(C_d(md), C_{2d}(md), \dots, C_{md}(md)) =_d (C'_1(m), C'_2(m), \dots, C'_m(m))$$

whenever $p(n) > 0$ or, equivalently, $p'(m) > 0$.

EXAMPLE 1 (Permutations with only even cycles). This is the assembly with $m_i = (i - 1)!\mathbf{1}(2|i)$ and period $d = 2$. With $x = 1$, we have $\lambda_i = i^{-1}\mathbf{1}(2|i)$ and we cook up the artificial assembly with $\lambda'_i = \lambda_{2i} = 1/(2i)$, $x' = 1$ and $m'_i = (2i - 1)!i!/(2i)! = (i - 1)!/2$. This new assembly *is* the Ewens sampling formula, described in Section 4, with parameter $\theta = 1/2$. Note that m_i' fails to be an integer for $i = 1, 2$. Theorem 2 below, or the Feller coupling as used in Arratia, Barbour and Tavaré (1992), establishes an independent Poisson process limit, and translating back to the original assembly, this result may be written

$$(C_2(2m), C_4(2m), \dots) \Rightarrow (Z_2, Z_4, \dots),$$

where the Z_i are independent, Poisson distributed, with $\mathbb{E}Z_{2i} = 1/(2i)$.

EXAMPLE 2 (Permutations with only even cycles and fixed points). This is the assembly with $m_i = (i-1)!\mathbf{1}(2|i \text{ or } i = 1)$: it is aperiodic and *does not* satisfy the conclusion of Theorem 1. It does satisfy $\mathbb{E}C_i(n) \to \lambda_i$ for all *even* i, but for odd integers i, the limit of $\mathbb{E}C_i(n)$ does not exist. In terms of the proof of Theorem 1, we have $h(n-2)/h(n) \to 1$, but the asymptotic decay of $h(n)$ has different constant coefficients along the odd and even integers. In fact, for any $\theta > 0$ and for any assembly with $\lambda_i = \theta/i$ for i even and $0 < \lambda \equiv \lambda_1 + \lambda_3 + \cdots < \infty$, it can be shown that $h(2m+1)/h(2m) \to \tanh(\lambda) < 1$. An easy way to do this is to compute in terms of independent Poisson random variables; start with the weighted sum T_{0n} defined by (27) and the identity

$$\mathbb{P}(T_{0n} = n) = h(n)\exp(-(\lambda_1 + \lambda_2 + \cdots + \lambda_n)) \tag{16}$$

and end up looking at $\mathbb{P}(K \text{ odd})/\mathbb{P}(K \text{ even}) = \tanh(\lambda)$, where $K = Z_1 + Z_3 + \cdots$ is Poisson with $\mathbb{E}K = \lambda$. This example also is such that $C_1(n)$ does not converge in distribution, but $(C_2(n), C_4(n), \dots) \Rightarrow (Z_2, Z_4, \dots)$, with the same limit as the previous example.

3. Total variation convergence. The restriction of $\mathbf{C}(n)$ to its first b coordinates, which specify the counts of components of sizes up to b, is denoted by $\mathbf{C}_b(n) \equiv (C_1(n), C_2(n), \dots, C_b(n))$. We want to measure how well the process $\mathbf{C}_b(n)$ is approximated by a process of independent Poisson random variables $\mathbf{Z}_b \equiv (Z_1, Z_2, \dots, Z_b)$ with appropriate parameters λ_i. A natural metric for measuring how well $\mathbf{Z}_b$ approximates $\mathbf{C}_b(n)$ is total variation distance. For two random elements X and Y on a finite or countable space S, the total variation distance between X and Y is defined to be

$$d_{\mathrm{TV}}(X, Y) = \sup_{A \subset S} |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|.$$

It is well known that

$$d_{\mathrm{TV}} = \tfrac{1}{2} \sum_{s \in S} |\mathbb{P}(X = s) - \mathbb{P}(Y = s)|.$$

Furthermore, if we define

$$T_{mn} = \sum_{j=m+1}^{n} jZ_j, \qquad 0 \le m < n,$$

then it may be shown [Arratia and Tavaré (1994)] that

$$d_b(n) \equiv d_b \equiv d_{\mathrm{TV}}(\mathbf{C}_b(n), \mathbf{Z}_b)$$

satisfies

$$d_b(n) = \frac{1}{2} \sum_{r=0}^{\infty} \mathbb{P}(T_{0b} = r) \left| 1 - \frac{\mathbb{P}(T_{bn} = n - r)}{\mathbb{P}(T_{0n} = n)} \right|. \tag{17}$$

For each fixed n, $d_b(n)$ is increasing as a function of b. This is not immediate from (17), but is clear from the observation that the functional $h: \mathbb{R}^{b+1} \to \mathbb{R}^b$ which "forgets" the $(b + 1)$st coordinate can be used to express $\mathbf{C}_b = h(\mathbf{C}_{b+1})$, $\mathbf{Z}_b = h(\mathbf{Z}_{b+1})$.

Since $\mathbf{C}_b(n)$ and $\mathbf{Z}_b$ are discrete, (6) is equivalent to $d_b(n) \to 0$ as $n \to \infty$, for any fixed b. However, the convergence in (6) is too weak for many purposes, and it is necessary to show that $(C_1(n), \dots, C_b(n))$ is close to $(Z_1, \dots, Z_b)$ even when b grows with n. The main results in the paper, described in Theorem 3 in Section 6, imply that $d_b(n) \to 0$ for any $b = b(n)$ satisfying $b = o(n/\log n)$.

To illustrate the power of such estimates, we discuss functional central limit theorems, such as those proved for permutations by DeLaurentis and Pittel (1985), for the Ewens sampling formula by Hansen (1990) and for random mappings by Hansen (1989). These results involve the asymptotic behavior of the process $B_n(\cdot)$ with values in $D[0,1]$ defined by

$$B_n(t) = (\theta \log n)^{-1/2} \left(\sum_{i \le n^t} C_i(n) - \theta t \log n \right), \tag{18}$$

for $0 \le t \le 1$. In order for the centering by $\theta t \log n$ to be appropriate, one must have

$$\sum_{i \le b} \lambda_i = \theta \log b + o\left(\sqrt{\log b}\right), \tag{19}$$

as $b \to \infty$. Now suppose, for some fixed $\alpha \in (0,1)$, one knew only that $d_b(n) \to 0$ for $b = \lfloor n^\alpha \rfloor$; this is the situation using (49), for any $\alpha < \varepsilon/(1 + \varepsilon)$. By comparing $B_n(t)$ with the same renormalized sum of the *independent* Poisson random variables Z_i, it follows easily that $B_n(\cdot)$, restricted to $[0, \alpha]$, converges to standard Brownian motion restricted to $[0, \alpha]$. To get convergence to Brownian motion on $[0,1]$ requires only slightly more work, the key ingredient being provided in Lemma 2. On the other hand, convergence to Brownian motion carries substantially *less* information than is carried by results of the form $d_b(n) \to 0$ with $b = b(n)$ growing. Roughly speaking, the Brownian motion result is sensitive to $(C_1(n), C_2(n), \dots)$ only through sums over blocks of the form $i \in (n^\alpha, n^\beta)$ for fixed $0 \le \alpha < \beta \le 1$.

LEMMA 2. *Assume that an assembly satisfies*

$$\bar{\theta} = \sup_{i \ge 1} i\lambda_i < \infty, \qquad \underline{\theta} = \liminf_{i \to \infty} i\lambda_i > 0 \tag{20}$$

and that $d_b(n) \to 0$ for some choice of $b = b(n)$ with $b = o(n)$, $\log(n/b) = o(\sqrt{\log n})$. Then there is a coupling that satisfies

$$(21) \qquad R_n \equiv (\log n)^{-1/2} \sum_{j \le n} |C_j(n) - Z_j| \to_P 0$$

as $n \to \infty$.

PROOF. For any coupling, for any $1 \le b \le n$,

$$(22) \qquad (\log n)^{1/2} R_n \le \sum_{j \le b} |C_j(n) - Z_j| + \sum_{b < j \le n} C_j(n) + \sum_{b < j \le n} Z_j.$$

Now fix the choice of $b = b(n)$. Fix a coupling of $(Z_j, j \ge 1)$ and the processes $(C_1(n), \dots, C_n(n))$ for $n \ge 1$ such that, for all n,

$$\mathbb{P}\big((C_1(n), \dots, C_b(n)) \ne (Z_1, \dots, Z_b)\big) = d_b(n).$$

In this coupling, the first term of (22) converges in probability to 0 because

$$\mathbb{P}\left(\sum_{j \le b} |C_j(n) - Z_j| > \varepsilon \right) \le \mathbb{P}\big((C_1(n), \dots, C_b(n)) \ne (Z_1, \dots, Z_b)\big)$$
$$= d_b(n) \to 0.$$

For the third term of (22), the first condition in (20), together with the upper bound on $\log(n/b)$, implies

$$(23) \qquad \mathbb{E} \sum_{b < j \le n} Z_j = \sum_{b < j \le n} \lambda_j \le \bar{\theta} \sum_{b < j \le n} \frac{1}{j} = O\left(\log\left(\frac{n}{b}\right)\right) = o\left(\sqrt{\log n}\right).$$

Finally, some hard work is needed to show that the second term of (22), which is the mean number of components of size greater than b, is comparable to the third. Recall from (9) and (16) that

$$\mathbb{E} C_j(n) = \frac{x^{-j} n! p(n-j)}{p(n)(n-j)!} \lambda_j \equiv \frac{h(n-j)}{h(n)} \lambda_j,$$

where

$$h(n) = \frac{p(n) x^n}{n!} = \exp\left(\sum_{j=1}^{n} \lambda_j \right) \mathbb{P}(T_{0n} = n).$$

We show in Lemma 3 that if $\underline{\theta} = \liminf i\lambda_i > 0$ and $\bar{\theta} = \sup i\lambda_i < \infty$, then there are constants $c_1 < c_2 \in (0, \infty)$ such that $c_1 \le n\mathbb{P}(T_{0n} = n) \le c_2$ for all sufficiently large n. Hence, for n sufficiently large,

$$\sum_{b < j \le n/2} \mathbb{E} C_j(n) \le \sum_{b < j \le n/2} \frac{c_2 \exp(\sum_{i \le n-j} \lambda_i) n}{(n-j) c_1 \exp(\sum_{l \le n} \lambda_l)} \lambda_j \le 2 \frac{c_2}{c_1} \sum_{b < j \le n/2} \lambda_j.$$

Since $\sum_{n/2 \le j \le n} C_j(n) \le 2$, we conclude from (23) that $\mathbb{E} \sum_{j > b} C_j(n) = o(\sqrt{\log n})$ as $n \to \infty$. Hence $R_n \to_P 0$, by Markov's inequality. □

This error estimate allows us to prove Corollary 2 below. It is immediately applicable to assemblies, such as random mappings, that satisfy the hypotheses of Theorem 3. For random mappings, one can check (19) by noting that $\lambda_i - 1/(2i) = O(i^{-3/2})$ as $i \to \infty$.

COROLLARY 2. *Assume that an assembly satisfies* (19) *and the conditions of Lemma* 2. *Then the rescaled process* $B_n(\cdot)$ *in* (18) *converges weakly in* $D[0,1]$ *to a standard Brownian motion.*

PROOF. Define the functional $h_n\colon \mathbb{R}^\infty \to D[0,1]$ by the requirement that $B_n(\cdot) = h_n(\mathbf{C}(n))$. Note first that the functional $h_n(\mathbf{Z})$ converges weakly to standard Brownian motion, by application of the functional central limit theorem for the Poisson process [cf. Ethier and Kurtz (1986), page 263] and assumption (19). The error introduced in approximating the functional of dependent component counts by that functional of the independent Poisson random variables is $h_n(\mathbf{C}(n)) - h_n(\mathbf{Z})$, with sup norm

$$\sup_{0 \le t \le 1} (\theta \log n)^{-1/2} \left| \sum_{j \le n^t} (C_j(n) - Z_j) \right| \le \theta^{-1/2} R_n \to_P 0,$$

using Lemma 2. □

Another simple corollary is a version of the functional central limit theorem which separates the contributions from even- and odd-sized components. The standard functional central limit theorem in Corollary 2 is implied by, but does not imply, the refined one with even-sized and odd-sized components separated. Assume that the assembly satisfies (20) and that

$$\sum_{\substack{i \le b,\\ i = j \bmod 2}} \lambda_i = \frac{\theta_j}{2} \log b + o\left(\sqrt{\log b}\right), \qquad j = 1,2, \tag{24}$$

as $b \to \infty$. Define $\theta = (\theta_1 + \theta_2)/2$ and for $t \in [0,1]$ set

$$B_n^j(t) = (\theta \log n)^{-1/2} \left(\sum_{\substack{i \le n^t;\\ i = j \bmod 2}} C_i(n) - \theta_j t \log n \right), \qquad j = 1,2. \tag{25}$$

COROLLARY 3. *Assume that an assembly satisfies* (24) *and the conditions of Lemma* 2. *Then the rescaled process* $(B_n^1(\cdot), B_n^2(\cdot))$ *in* (25) *converges weakly in* $D[0,1] \times D[0,1]$ *to* $(\sqrt{\theta_1/(2\theta)}\, B^1, \sqrt{\theta_2/(2\theta)}\, B^2)$, *where* B^1 *and* B^2 *are independent standard Brownian motions.*

PROOF. Define the functional $h_n'\colon \mathbb{R}^\infty \to D[0,1] \times D[0,1]$ by the requirement that $(B_n^1(\cdot), B_n^2(\cdot)) = h_n'(\mathbf{C}(n))$. Just as in Corollary 2, the functional

$h_n'(\mathbf{Z})$ converges weakly to $(\sqrt{\theta_1/(2\theta)}\,B^1, \sqrt{\theta_2/(2\theta)}\,B^2)$. The error in this approximations is $h_n'(\mathbf{C}(n)) - h_n'(\mathbf{Z})$ with sup norm at most

$$(\theta \log n)^{-1/2} \sup_{0 \le t \le 1} \left(\left| \sum_{\substack{j \le n^t, \\ j \text{ even}}} (C_j(n) - Z_j) \right| + \left| \sum_{\substack{j \le n^t, \\ j \text{ odd}}} (C_j(n) - Z_j) \right| \right)$$

$$\le \theta^{-1/2} R_n \to_P 0,$$

using Lemma 2 once more. □

Arratia and Tavaré (1992b) give a wide variety of other corollaries of $d_b(n) \to 0$ for growing b, in the context of random permutations and random mappings. The methods illustrated here can also be used to give estimates on the rate of convergence in the functional limit theorems; see Arratia, Barbour and Tavaré (1993, 1994) for more on this topic.

4. The Ewens sampling formula and its perturbations. The distribution of the component counts $(C_1(n), \dots, C_n(n))$ of an assembly is determined by the relationship

(26) $$(C_1(n), C_2(n), \dots, C_n(n)) =_d (Z_1, Z_2, \dots, Z_n | T_{0n} = n),$$

where

(27) $$T_{0n} = Z_1 + 2Z_2 + \cdots + nZ_n$$

and the Z_i are independent Poisson random variables with means $\mathbb{E} Z_i = \lambda_i = m_i x^i / i!$, as given in (4). The equality (26) holds for any n for which $p(n) > 0$ or, equivalently, $\mathbb{P}(T_{0n} = n) > 0$, and for any $x > 0$. See Arratia and Tavaré (1994) for details and extensions. It is convenient to define the generating function $\log f(z)$ of the sequence $\lambda_1, \lambda_2, \dots$ by

(28) $$f(z) = \exp\left(\sum_{j=1}^{\infty} \lambda_j z^j \right).$$

Using (3) and (4) one sees that f satisfies $1 + \sum_{n \ge 1} p(n) z^n = f(z/x)$; we have scaled f so that it has a singularity at $z = 1$.

The Ewens sampling formula [Ewens (1972)] with parameter $\theta > 0$, abbreviated in what follows to ESF(θ), is the distribution of the process of cycle counts of permutations of n objects, choosing a permutation with probability proportional to $\theta^{\#\text{cycles}}$. This distribution is given by (26), where

(29) $$j\lambda_j = \theta, \qquad j = 1, 2, \dots,$$

and the generating function of the λ_j is

$$f(z) \equiv f_\theta(z) = (1 - z)^{-\theta}.$$

The distributions of the assemblies to which our results apply may be thought of as perturbations of the Ewens sampling formula. We say that these assemblies are in the "logarithmic class." The notion of perturbation may take several different forms, which we now discuss.

4.1. *Bounds on* $i\lambda_i$. One notion of perturbation of ESF(θ) is to suppose that the Poisson parameters λ_i satisfy condition (20). In Theorem 2 we establish that under this condition we have weak convergence of processes in $\mathbb{R}^\infty$, as in (6). In particular, if for some choice of $x > 0$,

$$\lim_{i\to\infty} i\lambda_i \equiv \lim_{i\to\infty} im_i x^i/i! = \theta \in (0,\infty),$$

then we may take $\underline{\theta} = \theta$ and $\overline{\theta} = k\theta$ for some $k \geq 1$.

Various special cases of the result in (6) are already in the literature. Perhaps the most celebrated example concerns random permutations, established by Goncharov (1944) and Kolchin (1971). This is the case of ESF(1), with $x = 1$, $\theta = 1$.

For random mappings, the assembly with

$$m_i = (i-1)!(1 + i + i^2/2 + \cdots + i^{i-1}/(i-1)!),$$

the finite-dimensional convergence in (6) was established by Kolchin (1976, 1986). It can be shown that using $x = e^{-1}$ yields $\lambda_i < 1/(2i)$ and $\lambda_i \sim 1/(2i)$, so that Theorem 2 applies with $\theta = \overline{\theta} = \underline{\theta} = 1/2$.

Another example to which the result applies is 2-regular graphs. This is the assembly with

$$m_i = \mathbf{1}(i > 2)(i-1)!/2,$$

so that Theorem 2 applies with $x = 1$, $\overline{\theta} = \underline{\theta} = \theta = 1/2$.

For another example, consider transforming an assembly by, for example, coloring each component of even size with one of two colors. This yields $m'_i = m_i(1 + \mathbf{1}(i \text{ even}))$ so that if the original λ_i satisfy $i\lambda_i \to \theta$, then the transformed values λ'_i satisfy $\limsup i\lambda'_i = 2\theta = 2\liminf i\lambda'_i$. Such examples clearly satisfy the hypotheses of Theorem 2, so that process convergence obtains once more.

4.2. *Conditions on* $\log f$, *the generating function of the* λ_j. Another notion of perturbation of ESF(θ) involves the hypothesis that the generating function f in (28) satisfies

$$\log f(z) = -\theta \log(1-z) + R(z), \tag{30}$$

where $R(z)$ is analytic in $\Delta_0 \equiv \{z: |z| \leq 1 + \eta,\ z \notin [1, 1+\eta]\}$ for some $\eta > 0$ and satisfies $R(z) = K + o(\log^{-1}(1-z))$ when $z \to 1$ in Δ_0. [The condition on R stated here differs slightly from that given below (2.2) of Flajolet and Soria (1990); the version given here is from Flagolet (personal communication).] The condition (30), combined with Corollary 1 of Flajolet and Odlyzko (1990a), is enough to show that

$$i\lambda_i \to \theta \quad \text{as } i \to \infty, \tag{31}$$

so that Theorem 2 applies directly.

Flajolet and Soria (1990) use condition (30) to deduce asymptotics for $p(n)$ which imply (5) and hence (6) holds. They also show, inter alia, that the examples of random permutations, random mappings and 2-regular graphs satisfy the required conditions on $\log f$. Hansen (1994) shows that under (28) and (31) the process of largest components has a Poisson–Dirichlet limit.

4.3. *Conditions on the exponential generating function f.* Our final notion of perturbation of ESF(θ) involves analytic conditions on f itself. Motivated by the work of Flajolet and Odlyzko (1990a), we define a Δ-domain with parameters $\eta > 0$ and $0 < \phi < \pi/2$ as the set

$$\Delta \equiv \Delta(\eta, \phi) = \{z \in \mathbb{C}: |z| \le 1 + \eta, |\arg(z - 1)| \ge \phi\}. \tag{32}$$

We assume f is analytic on $\Delta \setminus \{1\}$, where $\Delta = \Delta(\eta_0, \phi)$ for some $\eta_0 > 0$, $0 < \phi < \pi/2$, and that there are constants $\delta > 0$, $\theta > 0$, $K > 0$, such that

$$f(z) = K(1 - z)^{-\theta}\left\{1 + O\left((1 - z)^{\delta}\right)\right\}$$

as $z \to 1$ in Δ. If, in addition, we assume that $\sup_{i \ge 1} i\lambda_i < \infty$, then we establish in Theorem 3 precise asymptotics for $d_b(n)$ as $n \to \infty$ with $b = o(n/\log n)$.

For random mappings, it is known that $f(ez) = (1 - t(z))^{-1}$, where $t(z)$ is the exponential generating function for rooted labelled trees, and that

$$t(z) = 1 - \sqrt{2(1 - ez)} + O(1 - ez);$$

see Flajolet and Odlyzko (1990b), for example. It follows that

$$f(z) = \frac{1}{\sqrt{2}}(1 - z)^{-1/2}\left\{1 + O\left((1 - z)^{1/2}\right)\right\}.$$

The conditions of Theorem 3 are satisfied with $\bar{\theta} = \theta = 1/2$ and $\delta = 1/2$.

For 2-regular graphs, we have

$$\begin{aligned} f(z) &= \exp\left(-\tfrac{1}{2}\log(1 - z) - z/2 - z^2/4\right) \\ &= e^{-3/4}(1 - z)^{-1/2}\{1 + O(1 - z)\}. \end{aligned}$$

Theorem 3 applies once more, with $\bar{\theta} = \theta = 1/2$ and $\delta = 1$.

4.4. *A hybrid ESF.* In this section, we define a hybrid version of the ESF that is useful in delimiting the various conditions and assumptions in our results. We define the hybrid Ewens sampling formula with parameters $\theta_1, \theta_2 > 0$, denoted by ESF(θ_1, θ_2), as the "assembly" determined by

$$\lambda_i = \frac{1}{i}\left(\theta_1 \mathbf{1}(i \text{ odd}) + \theta_2 \mathbf{1}(i \text{ even})\right). \tag{33}$$

The standard case is ESF(θ) = ESF(θ, θ). The generating function of the λ_i is

$$\sum_{i \ge 1} \lambda_i z^i = -\frac{\theta_1 + \theta_2}{2}\log(1 - z) + \frac{(\theta_1 - \theta_2)}{2}\log(1 + z),$$

so that

$$f(z) \equiv f_{\theta_1, \theta_2}(z) = (1 - z)^{-(\theta_1 + \theta_2)/2}(1 + z)^{(\theta_1 - \theta_2)/2}. \tag{34}$$

By construction, Theorem 2 applies, and finite-dimensional convergence as in (6) obtains for all values of $\theta_1 > 0$, $\theta_2 > 0$. Note that the analytic condition (30) does not apply for any $\theta_1 \neq \theta_2$, because $\lim_{i\to\infty} i\lambda_i$ does not exist. If $\theta_1 > \theta_2$ and $\theta_1 - \theta_2$ is an even integer, than f_{θ_1,θ_2} satisfies the analytic conditions after (32), so that asymptotics for $d_b(n)$ can be obtained from Theorem 3 even when (20) holds, but (31) does not.

5. Convergence of finite-dimensional distributions. If $Z_1, Z_2, \ldots$ are independent Poisson random variables with $\mathbb{E}Z_i = \lambda_i$ and $T = \sum_{i=1}^n iZ_i$, then $q(k) = \mathbb{P}(T = k)$, with $q(k) = 0$ for $k < 0$, satisfies the recurrence

$$kq(k) = \sum_i g_i q(k-i) = \sum_{i=1}^{k} g_i q(k-i), \qquad k = 1,2,\ldots, \tag{35}$$

where

$$g_i = i\lambda_i \mathbf{1}(1 \le i \le n).$$

See Arratia and Tavaré [(1994), (153)], for example. We use this recursion for three different versions of T. We assume

$$\bar{\theta} = \sup_{i\ge 1} i\lambda_i < \infty \tag{36}$$

and

$$\underline{\theta} = \liminf_{i\to\infty} i\lambda_i > 0. \tag{37}$$

Then we have the following theorem.

THEOREM 2. *Under conditions* (36) *and* (37), $d_b(n) \to 0$ *for every fixed* b; *that is*, $(C_1(n), \ldots, C_b(n)) \Rightarrow (Z_1, \ldots, Z_b)$ *as* $n \to \infty$.

The theorem is proved using several lemmas. We write $a(n) \asymp b(n)$ to denote that the lim inf and lim sup of the ratio are strictly positive and finite.

LEMMA 3. *Under conditions* (36) *and* (37), $\mathbb{P}(T_{0n} = n) \asymp n^{-1}$.

PROOF. Apply (35) to $T = T_{0n}$ to get, under condition (36),

$$kq(k) = \sum_{i=1}^{k} g_i q(k-i) \le \bar{\theta}\sum_{i=1}^{k} q(k-i) = \bar{\theta}\mathbb{P}(T_{0n} < k) \le \bar{\theta}. \tag{38}$$

With $k = n$, this supplies the upper bound needed for this lemma. For the lower bound, fix $\varepsilon > 0$ and $m \ge 1$ such that $i\lambda_i \ge \varepsilon$ for all $i \ge m$. This can be achieved under (37). Let Z_i^* be independent Poisson random variables with means $\bar{\theta}/i$, $i = m, \dots, n$, and let $T^* = \sum_{i=m}^n iZ_i^*$. For $n \geq m$,

$$\begin{aligned} nq(n) &= \sum_{i=1}^{n} g_i q(n-i) \geq \sum_{i=m}^{n} g_i q(n-i) \\ &\geq \varepsilon \sum_{i=m}^{n} q(n-i) = \varepsilon \mathbb{P}(T_{0n} \leq n-m) \\ &\geq \varepsilon \mathbb{P}(T^* + T_{0,m-1} \leq n-m), \end{aligned}$$

the last inequality following because there is a coupling with $T^* \geq T_{m-1,n}$. As noted after Corollary 4, there is a random variable $X_{\bar{\theta}}$ having a density, with $T^*/n \Rightarrow X_{\bar{\theta}}$, so that from (60),

$$P(T^* + T_{0,m-1} \leq n-m) \to \mathbb{P}(X_{\bar{\theta}} \leq 1) = \frac{e^{-\gamma\bar{\theta}}}{\Gamma(\bar{\theta}+1)} > 0.$$

This completes the proof. □

LEMMA 4. *For $1 \leq m \leq i \leq n$ and $0 < \varepsilon \leq 1$, let Z_i' be independent Poisson random variables with means ε/i and let $T' = \sum_{i=m}^n iZ_i'$. Then*

$$\mathbb{P}(T' = k) \leq \mathbb{P}(T' = 0), \qquad k = 1, \dots, n, \tag{39}$$

and for $u = 0, 1, \dots,$

$$|\mathbb{P}(T' = k) - \mathbb{P}(T' = k-u)| \leq \frac{(1+\varepsilon)}{k} u \mathbb{P}(T' = 0), \qquad k = 1, 2, \dots, n. \tag{40}$$

PROOF. In (35) applied to T', with $q(k) = \mathbb{P}(T' = k)$, we have $g_i = \varepsilon \mathbf{1}(m \leq i \leq n)$, so that for $1 \leq k \leq n$,

$$q(k) = \frac{\varepsilon}{k} \sum_{l=0}^{k-m} q(l) \leq \frac{1}{k} \sum_{l=0}^{k-1} q(l).$$

This says that each value is at most the average of the previous values, so that, by induction on k, (39) holds.

To establish (40), use (35) again to see that for $k = 1, \dots, n$,

$$\begin{aligned} kq(k) - (k-u)q(k-u) &= \sum_i g_i (q(k-i) - q(k-u-i)) \\ &= \varepsilon \sum_{i=m}^{k} (q(k-i) - q(k-u-i)) \\ &= \varepsilon \sum_{j=0}^{u-1} q(k-m-j). \end{aligned}$$

Hence

$$k(q(k) - q(k-u)) = -uq(k-u) + \varepsilon \sum_{j=0}^{u-1} q(k-m-j),$$

so that from (39),

$$\begin{aligned} k|q(k) - q(k-u)| &\le uq(k-u) + \varepsilon \sum_{j=0}^{u-1} q(k-m-j) \\ &\le (1+\varepsilon)u\mathbb{P}(T' = 0). \end{aligned} \tag{41}$$

This completes the proof of the lemma. □

LEMMA 5. *Let Z_i be independent Poisson random variables whose means λ_i satisfy* (36) *and* (37). *Fix $0 < \varepsilon \le 1$ and $m \in [b+1, n/4]$, with $i\lambda_i \ge \varepsilon$ for all $i \ge m$. Then for $0 \le u \le n/4$, $3n/4 < k \le n$,*

$$\begin{aligned} &|\mathbb{P}(T_{bn} = k) - \mathbb{P}(T_{bn} = k-u)| \\ &\quad \le \frac{2\bar{\theta}u}{n}[1 + 2(1+\varepsilon)(1+\log n)]\exp\left(-\varepsilon \sum_m^n \frac{1}{i}\right) + \frac{8\varepsilon(1+\varepsilon)u}{n^2}. \end{aligned} \tag{42}$$

PROOF. Let Z_i' be independent Poisson random variables with means $\lambda_i' = \mathbf{1}(m \le i \le n)\varepsilon/i$ and let Z_i'' be independent Poisson random variables with means $\lambda_i'' = (\lambda_i - \lambda_i')\mathbf{1}(i > b) \ge 0$, independent of the Z_i'. Define $T' = \sum_{i=1}^n iZ_i'$ and $T'' = \sum_{i=1}^n iZ_i''$. Then $T_{bn} =_d T' + T''$ with independent nonnegative summands, so

$$\mathbb{P}(T_{bn} = k) = \sum_{i=0}^k \mathbb{P}(T' = k-i)\mathbb{P}(T'' = i).$$

Hence for $u \ge 0$,

$$\begin{aligned} &|\mathbb{P}(T_{bn} = k) - \mathbb{P}(T_{bn} = k-u)| \\ &\quad \le \sum_{l=0}^k |\mathbb{P}(T' = l) - \mathbb{P}(T' = l-u)|\mathbb{P}(T'' = k-l) \\ &\quad = \sum_{0 \le l \le u-1} + \sum_{u \le l < 3n/4} + \sum_{3n/4 \le l \le k} \\ &\quad \equiv \Sigma_1 + \Sigma_2 + \Sigma_3. \end{aligned} \tag{43}$$

For the first sum, use Lemma 4 and (38) applied to T'' to see that

$$
\begin{aligned}
\Sigma_1 &= \sum_{l=0}^{u-1} \mathbb{P}(T' = l)\mathbb{P}(T'' = k - l) \\
&\le \mathbb{P}(T' = 0) \sum_{l=0}^{u-1} \mathbb{P}(T'' = k - l) \\
&\le \mathbb{P}(T' = 0) \sum_{l=0}^{u-1} \frac{\bar{\theta}}{k - l} \qquad (44)\\
&\le \mathbb{P}(T' = 0) u \frac{\bar{\theta}}{k - (u - 1)} \\
&\le \frac{2\bar{\theta}u}{n} \mathbb{P}(T' = 0).
\end{aligned}
$$

For the second sum, use Lemma 4 to bound the first factor and (38) applied to T'' to bound the second factor by $\bar{\theta}/(k - l)$, to get for $u > 0$,

$$
\begin{aligned}
\Sigma_2 &\equiv \sum_{u \le l < 3n/4} |\mathbb{P}(T' = l) - \mathbb{P}(T' = l - u)|\mathbb{P}(T'' = k - l) \\
&\le u(1 + \varepsilon)\mathbb{P}(T' = 0) \sum_{u \le l < 3n/4} \frac{\bar{\theta}}{(k - l)l} \qquad (45)\\
&\le u(1 + \varepsilon)\mathbb{P}(T' = 0) \frac{4\bar{\theta}(1 + \log n)}{n}.
\end{aligned}
$$

For $u = 0$, we have $\Sigma_2 = 0$, and so (45) applies once more.

For the third sum, note that $\Sigma_l \mathbb{P}(T'' = k - l) \le 1$, so that

$$
\begin{aligned}
\Sigma_3 &\equiv \sum_{3n/4 \le l \le k} |\mathbb{P}(T' = l) - \mathbb{P}(T' = l - u)|\mathbb{P}(T'' = k - l) \\
&\le \max_{3n/4 \le l \le k} |\mathbb{P}(T' = l) - \mathbb{P}(T' = l - u)|.
\end{aligned}
$$

Applying (41) to T' to get the first line below, and then the bound from (38) with ε playing the role of $\bar{\theta}$, we have for $l \ge 3n/4$,

$$
\begin{aligned}
|\mathbb{P}(T' = l) - \mathbb{P}(T' = l - u)| &\le \frac{u}{l} q(l - u) + \frac{\varepsilon}{l} \sum_{j=0}^{u-1} q(l - m - j) \\
&\le \frac{u}{l} \frac{\varepsilon}{l - u} + \frac{\varepsilon}{l} \sum_{j=0}^{u-1} \frac{\varepsilon}{l - m - j} \\
&\le \frac{8\varepsilon u}{n^2} + \frac{8\varepsilon^2 u}{n^2},
\end{aligned}
$$

so that

$$\Sigma_3 \le \frac{8\varepsilon(1+\varepsilon)u}{n^2}. \tag{46}$$

Finally,

$$\mathbb{P}(T' = 0) = \mathbb{P}(Z'_m = \cdots = Z'_n = 0) = \exp\left(-\sum_{j=m}^{n} \varepsilon/j\right). \tag{47}$$

Combining (44), (45), (46) and (47) completes the proof. □

PROOF OF THEOREM 2. From (50), we have for any $l_n \ge 0$,

$$d_b(n) \le \mathbb{P}(T_{0b} > l_n) + \frac{1}{\mathbb{P}(T_{0n} = n)} \sum_{r=0}^{l_n} \mathbb{P}(T_{0b} = r)|\mathbb{P}(T_{0n} = n) - \mathbb{P}(T_{bn} = n - r)|.$$

We need to choose $l_n \to \infty$ so that $\mathbb{P}(T_{0b} > l_n) \to 0$. From Lemma 3, we know that $\mathbb{P}(T_{0n} = n) \asymp n^{-1}$, so if we show that l_n may be chosen so that uniformly in $0 \le r \le l_n$, $|\mathbb{P}(T_{0n} = n) - \mathbb{P}(T_{bn} = n - r)| = o(n^{-1})$, then the second term is also $o(1)$ as $n \to \infty$. However,

$$\begin{aligned}
|P(T_{0n} = n) - \mathbb{P}(T_{bn} = n - r)|
&= \left|\sum_{i=0}^{\infty} \mathbb{P}(T_{0b} = i)\mathbb{P}(T_{bn} = n - i) - \mathbb{P}(T_{bn} = n - r)\right| \\
&\le \sum_{i=0}^{\infty} \mathbb{P}(T_{0b} = i)|\mathbb{P}(T_{bn} = n - i) - \mathbb{P}(T_{bn} = n - r)| \\
&\le \mathbb{P}(T_{0b} > l_n) + \sum_{i=0}^{l_n} \mathbb{P}(T_{0b} = i)|\mathbb{P}(T_{bn} = n - i) - \mathbb{P}(T_{bn} = n - r)|.
\end{aligned}$$

We may suppose that $l_n \le n/4$. For $0 \le i,\ r \le l_n$, we apply Lemma 5 with $u = |r - i| \le l_n$, $k = n - (i \wedge r)$, to get the uniform bound

$$\begin{aligned}
&|\mathbb{P}(T_{bn} = n - i) - \mathbb{P}(T_{bn} = n - r)| \\
&\quad \le \frac{2\bar{\theta} l_n[1 + 2(1+\varepsilon)(1 + \log n)]}{n} \exp\left(-\varepsilon \sum_{m}^{n} \frac{1}{i}\right) \\
&\qquad + \frac{8\varepsilon(1+\varepsilon)l_n}{n^2},
\end{aligned} \tag{48}$$

so that uniformly in $0 \le r \le l_n$,

$$|\mathbb{P}(T_{0n} = n) - \mathbb{P}(T_{bn} = n - r)| = O\left(\mathbb{P}(T_{0b} > l_n) + \frac{l_n \log n}{n^{1+\varepsilon}}\right).$$

We shall see in Lemma 9 that if we take $l_n = b \log n$, then for every fixed k,

$$\mathbb{P}(T_{0b} > l_n) = o\big((b/n)^k\big).$$

For this choice of l_n, it follows that indeed

$$\sum_{r=0}^{l_n} \mathbb{P}(T_{0b} = r)|\mathbb{P}(T_{0n} = n) - \mathbb{P}(T_{bn} = n - r)| = o(n^{-1}),$$

completing the proof of Theorem 2. □

REMARK. The proof of Theorem 2 provides an explicit upper bound for $d_b(n)$ for some b that vary with n. To see this, fix m_0 so that $i\lambda_i \geq \varepsilon$ for all $i \geq m_0$ and let m be given by $m = \max(m_0, b + 1)$. Then the factor in (47) is of order $(n/b)^\varepsilon$, so that the upper bound in (48) yields

$$d_b(n) = O\left(\frac{b^{1+\varepsilon} \log^2 n}{n^\varepsilon}\right). \tag{49}$$

Thus if $b = o(n^{\varepsilon/(1+\varepsilon)} \log^{-2/(1+\varepsilon)} n)$, then $d_b(n) \to 0$.

6. The main results. This section states our main result, Theorem 3, which concerns the asymptotic behavior of $d_b(n)$ for large n. The proof, which is given in Sections 7–10 below, starts with the fact that

$$d_b(n) = \sum_{r=0}^{\infty} \mathbb{P}(T_{0b} = r)\left(1 - \frac{\mathbb{P}(T_{bn} = n - r)}{\mathbb{P}(T_{0n} = n)}\right)^+. \tag{50}$$

Large deviation analysis is used in Section 7 to show that the terms in (50) with r large enough are insignificant. To simplify the significant terms of (50), with r small, we need precise local limit approximations to the densities of T_{0b}, T_{bn} and T_{0n}, all of which are weighted sums of independent Poisson random variables. This involves careful analysis of coefficients in generating functions like (28), $f(z) = \exp(\sum\lambda_i z^i)$. Each of these terms is expanded in partial Taylor series. Faa di Bruno's formula is used to bound derivatives, the maximum modulus principle is used to bound remainder terms and singularity analysis is applied to obtain asymptotic behavior. The rest of this section gives some easy results which can be used to simplify the asymptotics in (53), and a simple local limit heuristic which makes it easy to guess the result of Theorem 3.

Recall from (32) that a Δ-domain with parameters $\eta > 0$ and $0 < \phi < \pi/2$ is the set

$$\Delta \equiv \Delta(\eta, \phi) = \{z \in \mathbb{C}: |z| \leq 1 + \eta, |\arg(z - 1)| \geq \phi\}.$$

THEOREM 3. *Assume f is analytic on $\Delta \setminus \{1\}$, where $\Delta = \Delta(\eta_0, \phi)$ for some $\eta_0 > 0$, $0 < \phi < \pi/2$, and that there are constants $\delta > 0$, $\theta > 0$, $K > 0$ such that*

$$f(z) = K(1 - z)^{-\theta}\big\{1 + O\big((1 - z)^\delta\big)\big\} \tag{51}$$

as $z \to 1$ *in* Δ. *Assume further that* (36) *holds*: $\sup i\lambda_i < \infty$. *If*

$$b = o(n/\log n), \tag{52}$$

then

$$d_b(n) = \frac{|1-\theta|}{2n}\mathbb{E}|T_{0b} - \mathbb{E}T_{0b}| + o\left(\frac{b}{n}\right). \tag{53}$$

The proof of this theorem is organized into 13 lemmas, concluding at the end of Section 10.

The assumption $\sup i\lambda_i < \infty$ implies that $\mathbb{E}|T_{0b} - \mathbb{E}T_{0b}| = O(b)$, so that Theorem 3 implies that $d_b(n) = O(b/n)$ for any θ and $d_b(n) = o(b/n)$ in case $\theta = 1$. For the case of uniformly chosen permutations, which has $\theta = 1$, Arratia and Tavaré (1992a) show that $d_b(n)$ converges superexponentially fast as a function of b/n, so that the upper bound $d_b(n) = o(b/n)$ is not sharp. If b is fixed, $\mathbb{E}T_{0b} > 0$ and $\theta \neq 1$, then (53) provides the asymptotics for $d_b(n)$ as $n \to \infty$. The next corollary observes that in Theorem 3, for $\theta \neq 1$, not only is the upper bound $d_b(n) = O(b/n)$ sharp, but also that the error bound in (53) is negligible in comparison to the first term, which thus supplies the asymptotic decay rate for $d_b(n)$ when $b \to \infty$.

COROLLARY 4. *Assume the conditions of Theorem* 3 *obtain and that* $\theta \neq 1$. *For fixed* b,

$$d_b(n) \sim \frac{|1-\theta|}{2n}\mathbb{E}|T_{0b} - \mathbb{E}T_{0b}|. \tag{54}$$

Furthermore, if $\mathbb{E}|T_{0b} - \mathbb{E}T_{0b}| \asymp b$, *then for any choice of* $b = b(n)$ *with* $b = o(n/\log n)$, (54) *holds.*

To apply Corollary 4, it is convenient to have the following lemma.

LEMMA 6. *Assume that* $\liminf i\lambda_i = \underline{\theta} > 0$ *and* $\sup i\lambda_i = \bar{\theta} < \infty$. *As* $b \to \infty$, $\mathbb{E}|T_{0b} - \mathbb{E}T_{0b}| \asymp b$.

PROOF. For the upper bound, note that

$$\mathbb{E}|T_{0b} - \mathbb{E}T_{0b}| \le 2\mathbb{E}T_{0b} = 2\sum_{j=1}^{b} j\lambda_j \le 2\bar{\theta}b.$$

For a lower bound, we proceed as follows. Since $\liminf i\lambda_i = \underline{\theta} > 0$, we can choose i_0 such that $i\lambda_i > \underline{\theta}/2$ for all $i \ge i_0$. Letting $S = \sum_{b/2 \le i \le b} Z_i$, we observe that, for $b \ge 2i_0$, S has a Poisson distribution with mean

$$\mathbb{E}S \ge \frac{1}{2}\underline{\theta}\sum_{b/2 \le i \le b}\frac{1}{i} \ge \frac{1}{2}\underline{\theta}\log\left(\frac{3}{2}\right) \equiv \xi.$$

Letting S' be a Poisson random variable with mean ξ, we see that

$$\begin{aligned}
\mathbb{P}(T_{0b} - \mathbb{E}T_{0b} \ge b) &= \mathbb{P}\left(\sum_1^b iZ_i \ge b + \mathbb{E}T_{0b}\right) \\
&\ge \mathbb{P}\left(\sum_{b/2}^b iZ_i \ge b + \mathbb{E}T_{0b}\right) \\
&\ge \mathbb{P}(bS/2 \ge b + \mathbb{E}T_{0b}) \\
&= \mathbb{P}(S \ge 2 + 2\mathbb{E}T_{0b}/b) \\
&\ge \mathbb{P}(S \ge 2 + 2\bar{\theta}) \\
&\ge \mathbb{P}(S' \ge 2 + 2\bar{\theta}) \equiv c(\underline{\theta}, \bar{\theta}) > 0.
\end{aligned}$$

Hence for $b \ge 2i_0$,

$$b^{-1}\mathbb{E}|T_{0b} - \mathbb{E}T_{0b}| \ge \mathbb{P}(T_{0b} - \mathbb{E}T_{0b} \ge b) \ge c(\underline{\theta}, \bar{\theta}) > 0. \qquad \square$$

The asymptotics in Corollary 4 can be further simplified with additional assumptions about the sequence $\{\lambda_i\}$. For $\theta > 0$, let X_θ be the random variable with mean θ and Laplace transform determined by

$$-\log \mathbb{E}\exp(-uX_\theta) = \int_0^1 \frac{1 - \exp(-ux)}{x}\theta\, dx, \qquad u > 0. \tag{55}$$

Lemma 7. *Assume that* $\bar{\theta} = \sup i\lambda_i < \infty$ *and fix* $\theta \in (0, \infty)$. *The following are equivalent*:

$$\lim_{b \to \infty} \frac{1}{b} \sum_{i=1}^b i\lambda_i = \theta, \tag{56}$$

$$b^{-1}T_{0b} \Rightarrow U \quad \textit{and} \quad \theta = \mathbb{E}U, \tag{57}$$

$$b^{-1}T_{0b} \Rightarrow X_\theta. \tag{58}$$

Proof. We show that (57) ⇒ (56) and that (56) ⇒ (58); clearly, (58) ⇒ (57). Suppose then that (57) holds. Since

$$\mathbb{E}T_{0b}^2 = \sum_{i \le b} i^2\lambda_i + \left(\sum_{i \le b} i\lambda_i\right)^2 \le \sum_{i \le b} i\bar{\theta} + b^2\bar{\theta}^2 \le b^2\bar{\theta}(1 + \bar{\theta}),$$

the sequence $b^{-1}T_{0b}$ is uniformly integrable. Hence $\mathbb{E}T_{0b}/b = \sum_{i \le b} i\lambda_i \to \mathbb{E}U \equiv \theta$ and (56) holds.

Now suppose that (56) holds. Let μ_b be the measure on Borel subsets of $[0, 1]$ given by

$$\mu_b = \sum_{i \le b} \frac{i}{b}\lambda_i \delta_{i/b},$$

so that for $y \in (0, 1]$,

$$\mu_b([0, y]) = \sum_{i/b \le y} \frac{i}{b}\lambda_i = y\frac{1}{by} \sum_{i \le by} i\lambda_i \to \theta y.$$

Since sets of the form $[0, y]$ are convergence determining, we have $\mu_b \Rightarrow \mu \equiv \theta$ times Lebesgue measure on $[0, 1]$. Since $(1 - e^{-uy})/y$ is a bounded continuous function of y,

$$-\log \mathbb{E} e^{-uT_{0b}/b} = \sum_{i=1}^{b} \lambda_i (1 - e^{-iu/b})$$
$$= \int_0^1 \frac{1 - e^{-uy}}{y} \mu_b(dy) \to \int_0^1 \frac{1 - e^{-uy}}{y} \mu(dy)$$

as $b \to \infty$. Hence $b^{-1} T_{0b} \Rightarrow U =_d X_\theta$. Since $b^{-1} T_{0b}$ is uniformly integrable, $\mathbb{E} U = \theta \in (0, \infty)$. This completes the proof. □

COROLLARY 5. *Assume the conditions of Theorem* 3 *and that as* $b \to \infty$, $b^{-1} T_{0b} \Rightarrow X_\theta$. *If* $\theta \neq 1$, $b \to \infty$ *and* $b = o(n/\log n)$, *then*

$$d_b(n) \sim \frac{|1 - \theta|}{2} \mathbb{E} |X_\theta - \theta| \frac{b}{n}. \tag{59}$$

PROOF. Uniform integrability shows that

$$\mathbb{E} |T_{0b} - \mathbb{E} T_{0b}|/b \to \mathbb{E} |X_\theta - \mathbb{E} X_\theta| = \mathbb{E} |X_\theta - \theta|.$$

We may now apply Corollary 4 to complete the proof. □

The evaluation of $\mathbb{E}|X_\theta - \theta|$ as an explicit function of θ is possible, in principle, using expressions for the density g of X_θ given by Ignatov (1982) and Griffiths (1988). In particular,

$$g(x) = \frac{e^{-\gamma\theta}}{\Gamma(\theta)} x^{\theta - 1}, \qquad 0 < x \leq 1, \tag{60}$$

where γ is Euler's constant. In case $\theta \leq 1$, it follows from the fact that $\mathbb{E}|X_\theta - \theta| = 2\mathbb{E}(\theta - X_\theta)^+$ that

$$\mathbb{E} |X_\theta - \theta| = 2 \int_0^\theta (\theta - x) g(x)\, dx = 2 \frac{e^{-\gamma\theta}}{\Gamma(\theta)} \frac{\theta^\theta}{1 + \theta}.$$

The example of random mappings has $\theta = 1/2$ and if $b, n \to \infty$, we see that

$$d_b(n) \sim \frac{e^{-\gamma/2}}{3\sqrt{2\pi}} \frac{b}{n} \doteq 0.0996 \frac{b}{n}.$$

For the Ewens sampling formula, Arratia, Barbour and Tavaré (1992) proved an upper bound of the form $d_b(n) \leq c(\theta) b/n$ for all $1 \leq b \leq n$ for some explicit constant $c(\theta)$. Compared with this upper bound, Corollary 4 is weaker in that it requires $b = o(n/\log n)$ and does not provide an upper bound with an explicit constant, but stronger in that it provides asymptotics when $\theta \neq 1$. For perturbations of the Ewens sampling formula, such as random mappings, Corollary 4 provides the first proof that, for some $b = b(n)$ tending to infinity, $d_b(n) = O(b/n)$.

Before proving the theorem, we give a heuristic derivation of the form of the asymptotics in (53). See Arratia and Tavaré [(1994), Section 4.2] for further details. We noted above that if $i\lambda_i \to \theta$ as $i \to \infty$, then $T_{0n}/n \Rightarrow X_\theta$, a random variable with density g satisfying $g'(1-)/g(1) = \theta - 1$. The local limit heuristic then approximates $\mathbb{P}(T_{bn} = n-k)$ by $n^{-1}(g(1) - kg'(1-)/n)$, to get

$$
\begin{aligned}
\mathbb{P}(T_{0n} = n) &= \sum_{k=0}^{n} \mathbb{P}(T_{0b} = k)\mathbb{P}(T_{bn} = n-k) \\
&\doteq \sum_{k\geq 0} \mathbb{P}(T_{0b} = k)\frac{1}{n}\left(g(1) - \frac{k}{n}g'(1-)\right) \qquad (61)\\
&= \frac{1}{n}\left(g(1) - \frac{\mathbb{E}T_{0b}}{n}g'(1-)\right).
\end{aligned}
$$

Using this approximation and the result in (17), we see that, for some cutoff $L_n \to \infty$ with n,

$$
\begin{aligned}
d_b(n) &= \frac{1}{2}\sum_{k=0}^{L_n} \mathbb{P}(T_{0b} = k)\left|1 - \frac{\mathbb{P}(T_{bn} = n-k)}{\mathbb{P}(T_{0n} = n)}\right| + \frac{1}{2}\mathbb{P}(T_{0b} > L_n) \\
&\doteq \frac{1}{2}\sum_{k\geq 0} \mathbb{P}(T_{0b} = k)\left|\frac{n^{-1}(k - \mathbb{E}T_{0b})g'(1-)}{g(1) - n^{-1}\mathbb{E}T_{0b}g'(1-)}\right| \\
&\doteq \frac{1}{2n}\frac{|g'(1-)|}{g(1)}\mathbb{E}|T_{0b} - \mathbb{E}T_{0b}|.
\end{aligned}
$$

7. Large deviations of T_{0b}. In this section, we assume that conditions (36) and (37) are satisfied. Think of L_n as the dividing line between terms of (50) that are shown by singularity analysis to have the proper asymptotics and those that are shown by large deviation theory to be insignificant. We define

$$L_n \equiv \min(b \log n, b^{2/3}n^{1/3}).$$

Note that $L_n = b \log n$ when $1 \leq b \leq n(\log n)^{-3}$ and $L_n = b^{2/3}n^{1/3}$ when $b \geq n(\log n)^{-3}$. Recall that for Theorem 1, we assume $b = o(n/\log n)$.

LEMMA 8. *For any $b \geq 1$ and $w > 0$, $\mathbb{P}(T_{0b} \geq bw) \leq (\bar{\theta}e/w)^w$.*

PROOF. Using Chebyshev's inequality, for any $\beta \geq 0$,

$$
\begin{aligned}
\mathbb{P}(T_{0b} \geq bw) &= \mathbb{P}(\exp(\beta T_{0b}) \geq \exp(\beta bw)) \qquad (62)\\
&\leq \mathbb{E}\exp(\beta T_{0b} - \beta bw).
\end{aligned}
$$

We calculate

$$\begin{aligned}\log \mathbb{E} \exp(\beta T_{0b}) &= \log \prod_{j=1}^{b} \mathbb{E} \exp(\beta_j Z_j)\\ &= \log \prod_{j=1}^{b} \exp\big(-\lambda_j + \lambda_j \exp(\beta_j)\big)\\ &= \sum_{j=1}^{b} \big(-\lambda_j + \lambda_j \exp(\beta_j)\big)\\ &= \sum_{j=1}^{b} \int_0^{\beta} j\lambda_j \exp(jt)\, dt\\ &\le \sum_{j=1}^{b} \int_0^{\beta} \bar{\theta} \exp(jt)\, dt\\ &\le \bar{\theta} b \int_0^{\beta} \exp(bt)\, dt\\ &= \bar{\theta}(\exp(\beta b) - 1)\\ &\le \theta \exp(\beta b).\end{aligned}$$

Thus, for any $\beta \ge 0$,

$$\log \mathbb{P}(T_{0b} \ge bw) \le \bar{\theta} e^{\beta b} - \beta b w.$$

Using $\beta = b^{-1} \log(w/\bar{\theta})$ completes the proof. □

Lemma 9 shows that the terms of (50) indexed by $r > L_n$ make a net contribution which is $o(b/n)$. Our overall strategy for proving Theorem 1 involves omitting these terms from $d_b(n)$, and Lemma 10 shows that the corresponding contribution to the right side of (53) is negligible. Lemma 6 is needed to deduce Corollary 1 from Theorem 1.

LEMMA 9. *For any k, as $n \to \infty$, with $1 \le b = o(n)$,*

$$\sum_{r > L_n} \mathbb{P}(T_{0b} = r)\left(1 - \frac{\mathbb{P}(T_{bn} = n - r)}{\mathbb{P}(T_{0n} = n)}\right)^{+} \le \mathbb{P}(T_{0b} > L_n) = o\left(\left(\frac{b}{n}\right)^{k}\right).$$

PROOF. Since the expression involving positive parts is at most 1, the first inequality is obvious. We apply Lemma 8 in two cases. In the first case, with $L_n = b \log n$, we have $w = \log n$, so that

$$\log \mathbb{P}(T_{0b} > L_n) \le w \log(\bar{\theta} e/w) \sim -w \log w = -(\log \log n) \log n,$$

which, for any fixed k, tends to minus infinity faster than $\log n^{-k} = -k \log n$. In the second case, for large b, we have $L_n = bw = b^{2/3} n^{1/3}$, so that $w = (n/b)^{1/3} \to \infty$. Thus

$$\log \mathbb{P}(T_{0b} > L_n) \le w \log(\bar{\theta} e/w) \sim -w \log w = -\tfrac{1}{3} w \log(n/b),$$

which, for any fixed k, goes to minus infinity faster than $\log(b/n)^k = -k\log(n/b)$. □

LEMMA 10. *For any k, as $n \to \infty$,*

$$\frac{|1-\theta|}{n}\mathbb{E}\big(|T_{0b} - \mathbb{E}T_{0b}|\mathbf{1}_{\{T_{0b} > L_n\}}\big) = o\left(\left(\frac{b}{n}\right)^k\right).$$

PROOF. Use the Cauchy–Schwarz inequality to see that

$$\mathbb{E}\big(|T_{0b} - \mathbb{E}T_{0b}|\mathbf{1}_{\{T_{0b} > L_n\}}\big) \le \sqrt{\mathbb{E}(T_{0b} - \mathbb{E}T_{0b})^2\mathbb{P}(T_{0b} > L_n)}\,.$$

However,

$$\mathbb{E}(T_{0b} - \mathbb{E}T_{0b})^2 = \sum_{j=1}^{b} j^2\lambda_j \le \bar{\theta}\sum_{j=1}^{b} j \le \bar{\theta}b^2 \le \bar{\theta}n^2,$$

and from Lemma 9, $\mathbb{P}(T_{0b} > L_n) = o((b/n)^{2k})$. □

8. Analysis of $d_b(n)$ via generating functions. Singularity analysis gives asymptotic information on the coefficients of a given generating function, based on the behavior near the singularity of smallest modulus. Flajolet and Odlyzko (1990a) prove several "transfer" theorems in which behavior of the generating function near the singularity of smallest modulus transfer into similar properties of coefficients; o, O and $\sim$ behavior of the generating function near the singularity transfer to similar coefficient behavior. They prove these theorems by calculating

$$[z^n]f(z) = \frac{1}{2\pi i}\oint\frac{f(z)}{z^{n+1}}\,dz$$

about suitably designed contours lying within the Δ-domain and containing the origin. Theorem 4 appears as Theorem 1 of Flajolet and Odlyzko (1990a), while Theorem 5 is their Corollary 2.

THEOREM 4. *Let α be a real number. If if is analytic on $\Delta\setminus\{1\}$ and $f(z) = O((1-z)^{-\alpha})$ as $z \to 1$ in Δ, then $[z^n]f(z) = O(n^{\alpha-1})$.*

THEOREM 5. *Suppose $\alpha \notin \{0, -1, -2, -3, \dots\}$. If f is analytic in $\Delta\setminus\{1\}$ and $f(z) \sim K(1-z)^{-\alpha}$ as $z \to 1$ in Δ, then $[z^n]f(z) \sim K/(\Gamma(\alpha))n^{\alpha-1}$.*

The asymptotic expansion of the coefficient $[z^n]$ of z^n in $f_\theta(z)$ will be needed later. Flajolet and Odlyzko (1990a) show that

$$[z^n]f_\theta(z) \sim \frac{n^{\theta-1}}{\Gamma(\theta)}\left(1 - \frac{\theta(1-\theta)}{2n}\right) + O(n^{\theta-3}). \tag{63}$$

Define

$$(64)\qquad d_b^* = \sum_{r=0}^{L_n} \mathbb{P}(T_{0b} = r)\left(1 - \frac{\mathbb{P}(T_{bn} = n - r)}{\mathbb{P}(T_{0n} = n)}\right)^+.$$

We see from Lemma 9 and (50) that

$$(65)\qquad d_b = d_b^* + o\left(\frac{b}{n}\right).$$

We rewrite the probabilities under the positive part sign in terms of generating functions. First,

$$\begin{aligned}
\mathbb{P}(T_{0n} = n) &= [z^n]\mathbb{E}z^{T_{0n}}\\
&= [z^n]\exp\left(-\sum_{j=1}^{n} \lambda_j + \lambda_j z^j\right)\\
&= \exp\left(-\sum_{j=1}^{n} \lambda_j\right)[z^n]\exp\left(\sum_{j=1}^{n} \lambda_j z^j\right)\\
&= \exp\left(-\sum_{j=1}^{n} \lambda_j\right)[z^n]\exp\left(\sum_{j=1}^{\infty} \lambda_j z^j\right)\\
&= \exp\left(-\sum_{j=1}^{n} \lambda_j\right)[z^n]f(z).
\end{aligned}$$

Defining $h(z) = \exp(-\sum_{j=1}^{b}\lambda_j z^j)$, we have

$$\begin{aligned}
\mathbb{P}(T_{bn} = n - r) &= [z^{n-r}]\mathbb{E}z^{T_{bn}}\\
&= [z^{n-r}]\exp\left(-\sum_{j=b+1}^{n} \lambda_j + \lambda_j z^j\right)\\
&= \exp\left(-\sum_{j=b+1}^{n} \lambda_j\right)[z^{n-r}]\exp\left(\sum_{j=b+1}^{n} \lambda_j z^j\right)\\
&= \exp\left(-\sum_{j=b+1}^{n} \lambda_j\right)[z^{n-r}]\exp\left(\sum_{j=b+1}^{\infty} \lambda_j z^j\right)\\
&= \exp\left(-\sum_{j=b+1}^{n} \lambda_j\right)[z^{n-r}]h(z)f(z).
\end{aligned}$$

Hence we can write (64) in the form

$$(66)\qquad d_b^* = \sum_{r=0}^{L_n} \mathbb{P}(T_{0b} = r)\left(1 - \frac{\exp\left(\sum_{j=1}^{b}\lambda_j\right)[z^{n-r}]f(z)h(z)}{[z^n]f(z)}\right)^+.$$

Our basic strategy for simplifying (66), which is used repeatedly, starting with (79), is to write d_b^* in the form

$$d_b^* = \sum_{r=0}^{L_n} \mathbb{P}(T_{0b} = r)(S_{1,n,r} + S_{2,n,r})^+, \tag{67}$$

where $S_{2,n,r}$ are asymptotically insignificant and may be ignored. We state this idea in the form of a lemma. It is applied when $M_{b,n}$ is $o(b/n)$.

LEMMA 11. *Define* $M_{b,n} = \max_{0 \le r \le L_n} |S_{2,n,r}|$. *Then*

$$d_b^* = \sum_{r=0}^{L_n} \mathbb{P}(T_{0b} = r)(S_{1,n,r})^+ + R_n,$$

where $|R_n| \le M_{b,n}$.

PROOF. Since $|(x+y)^+ - x^+| \le |y|$, we see that

$$|(S_{1,n,r} + S_{2,n,r})^+ - (S_{1,n,r})^+| \le M_{b,n}. \tag{68}$$

Summing against weights $\mathbb{P}(T_{0b} = r)$, we get

$$|R_n| = \left| d_b^* - \sum_{r=0}^{L_n} \mathbb{P}(T_{0b} = r)(S_{1,n,r})^+ \right| \le M_{b,n},$$

establishing the result. □

9. Expansion of $h(z)$ in finite Taylor series. We expand the analytic function $h(z) = \exp(-\sum_{j=1}^{b} \lambda_j z^j)$ in a finite Taylor series, with remainder, about $z = 1$:

$$\begin{aligned} h(z) &= \sum_{k=0}^{N} \frac{h^{(k)}(1)}{k!}(z-1)^k + R_N(z)(z-1)^{N+1} \\ &= h(1) + \sum_{k=1}^{N} \frac{h^{(k)}(1)}{k!}(z-1)^k + R_N(z)(z-1)^{N+1}, \end{aligned} \tag{69}$$

where $N \equiv N(n)$ may grow with n. Finite Taylor series expansion in $\mathbb{C}$ have the property that when the original functions are analytic, the remainder term $R_N(z)$ is analytic as well [Ahlfors (1979)]. We first bound $h^{(k)}(1)$ and then bound $R_N(z)$ on an appropriate disc in the complex plane.

To bound $h^{(k)}(1)$, $k \ge 1$, we use Faa di Bruno's formula:

$$\frac{d^k(\phi \circ g)}{dz^k} = \sum_{a_1 + 2a_2 + \cdots + ka_k = k} \frac{k!}{a_1! \cdots a_k!} \frac{d^m \phi}{dz^m} \left(\frac{g'}{1!}\right)^{a_1} \cdots \left(\frac{g^{(k)}}{k!}\right)^{a_k}, \tag{70}$$

where $m = a_1 + a_2 + \cdots + a_k$; see Gradshteyn and Ryzhik [(1980), page 19] for example. As noted by Harris (1960) in the context of random mappings, (70) is similar in form to (1). In our case, $h = \phi \circ g$, with $\phi(z) = \exp(-z)$ and $g(z) = \sum_{j=1}^{b} \lambda_j z^j$.

LEMMA 12. *For $k \geq 0$,*

$$|h^{(k)}(1)| \leq \exp\left(-\sum_{j=1}^{b} \lambda_j\right) B_k(\rho b)^k,$$

where B_k is the kth Bell number and

$$\rho = \max\{1, \bar{\theta}\}.$$

PROOF. Define $(t)_i = t(t-1)(t-2)\cdots(t-i+1)$, for $i > 0$. Then for $k > 0$,

$$\begin{aligned} \left.\frac{d^k g}{dz^k}\right|_{z=1} &= \sum_{j=1}^{b} (j)_k \lambda_j \\ &\leq \sum_{j=1}^{b} (j)_k \frac{\bar{\theta}}{j} \\ &= \bar{\theta} \sum_{j=1}^{b} (j-1)_{k-1} \\ &= \frac{\bar{\theta}(b)_k}{k} \\ &\leq \bar{\theta} b^k. \end{aligned} \tag{71}$$

Observing that, for $m \equiv \sum a_i$, $\bar{\theta}^m \leq \rho^m \leq \rho^k$ when $m \leq k$, we obtain

$$\begin{aligned} |h^{(k)}(1)| \leq &\sum_{a_1+2a_2+\cdots+ka_k=k} \frac{k!}{a_1!\cdots a_k!(1!)^{a_1}\cdots(k!)^{a_k}} \\ &\times \exp\left(-\sum_{j=1}^{b}\lambda_j\right)(\bar{\theta}b)^{a_1}\cdots(\bar{\theta}b^k)^{a_k} \\ \leq &\exp\left(-\sum_{j=1}^{b}\lambda_j\right)\rho^k b^k \sum_{a_1+2a_2+\cdots+ka_k=k} \frac{k!}{a_1!\cdots a_k!(1!)^{a_1}\cdots(k!)^{a_k}} \\ = &\exp\left(-\sum_{j=1}^{b}\lambda_j\right) B_k(\rho b)^k. \end{aligned}$$

□

Now we analyze the behavior of the remainder term $R_N(z)$. Define

$$D_\eta = \{z \in \mathbb{C}: |z| \leq 1 + \eta\}. \tag{72}$$

LEMMA 13. $\sup_{z \in D_\eta} |R_N(z)| \leq \eta^{-(N+1)}((eb)^{\bar{\theta} e^{\eta b}} + \exp(e^{\rho \eta b} - 1))$.

PROOF. We use the maximum modulus principle: If g is a function defined and continuous on a closed bounded set D and analytic on the interior of D, then the maximum of $|g(z)|$ on D is assumed on the boundary of D:

$$\begin{aligned}
\sup_{z\in D_\eta}|R_N(z)| &= \sup_{|z|=1+\eta}|R_N(z)| \\
&= \sup_{|z|=1+\eta}\left|\frac{h(z)-\sum_{k=0}^{N}(h^{(k)}(1)/k!)(z-1)^k}{(z-1)^{N-1}}\right| \\
&\le \eta^{-(N+1)}\left(\sup_{|z|=1+\eta}|h(z)| + \sum_{k=0}^{N}\exp\left(-\sum_{j=1}^{b}\lambda_j\right)\frac{(\rho\eta b)^k B_k}{k!}\right) \\
&\le \eta^{-(N+1)}\left(\sup_{|z|=1+\eta}|h(z)| + \sum_{k=0}^{\infty}\frac{(\rho\eta b)^k B_k}{k!}\right) \\
&= \eta^{-(N+1)}\left(\sup_{|z|=1+\eta}|h(z)| + \exp(e^{\rho\eta b}-1)\right). \qquad (73)
\end{aligned}$$

Analyzing the first term of (73),

$$\begin{aligned}
\sup_{|z|=1+\eta}|h(z)| &= \sup_{|z|=1+\eta}\left|\exp\left(-\sum_{j=1}^{b}\lambda_j z^j\right)\right| \\
&\le \exp\left(\sum_{j=1}^{b}\lambda_j(1+\eta)^j\right) \\
&\le \exp\left(\sum_{j=1}^{b}\frac{\bar\theta}{j}e^{\eta b}\right) \\
&\le \exp\left(\bar\theta e^{\eta b}(1+\log b)\right) \\
&= (eb)^{\bar\theta e^{\eta b}},
\end{aligned}$$

completing the proof. □

LEMMA 14. *Assume that ψ and r are analytic in $\Delta(\eta,\phi)\setminus\{1\}$ and $|\psi(z)|\le K_1|1-z|^{-\alpha}$ in $\Delta(\eta,\phi)\setminus\{1\}$. Then there is a universal constant $W_{l,\phi}$ such that, for all $|\alpha|\le l$ and $n\ge 2|\alpha|+4$,*

$$|[z^n]r(z)\eta(z)| \le K_1 \sup_{|z|\le 1+\eta}|r(z)|n^{\alpha-1}\left[W_{l,\phi} + \frac{\eta^{-\alpha}}{(1+\eta)^n}n^{1-\alpha}\right] \qquad (74)$$

and

$$W_{l,\phi} \le A\left(1+\left(\frac{l}{\cos\phi}\right)^{l+2}\right), \qquad (75)$$

where $A=(2+4/\cos\phi)^2$.

PROOF. In (2.15) of the proof of Theorem 4, Flajolet and Odlyzko (1990a) give the following uniform bound. There is a function $J(\alpha, \phi) < \infty$ such that if f is analytic and satisfies $|f(z)| \le K|1-z|^{-\alpha}$ in $\Delta(\eta, \phi) \setminus \{1\}$, then for all $n \ge 2|\alpha| + 4$,

$$|[z^n]f(z)| < Kn^{\alpha-1}\left(5 + \frac{J(\alpha, \phi)}{\pi} + \frac{\eta^{-\alpha}}{(1+\eta)^n} n^{-\alpha+1}\right),$$

where

$$J(\alpha, \phi) = \int_1^\infty t^{-\alpha}\left(1 + \frac{t\cos\phi}{2|\alpha|+4}\right)^{-2|\alpha|-4} dt.$$

We apply this with

$$f(z) = r(z)\psi(z), \qquad K = K_1 \sup_{|z| \le 1+\eta} |r(z)|.$$

To establish (75), observe first that $\sup_{|\alpha| \le l} J(\alpha, \phi) = \sup_{0 \le \beta \le l} J(-\beta, \phi)$ and that the integrand in $J(-\beta, \phi)$ has its maximum at the point

$$t = \frac{2\beta(\beta+2)}{(\beta+4)\cos\phi},$$

with value

$$\left(\frac{\beta}{\cos\phi}\right)^\beta \left(\frac{\beta+4}{2\beta+4}\right)^{\beta+4} \le \left(\frac{\beta}{\cos\phi}\right)^\beta. \tag{76}$$

We estimate the size of $J(-\beta, \phi)$ by breaking the integral into two pieces—the first over $t \in (1, t_0^2)$, the second over $t \in (t_0^2, \infty)$—for $t_0 = (2\beta + 4)/\cos\phi$. The second integral is

$$\int_{t_0^2}^\infty t^\beta\left(1 + \frac{t}{t_0}\right)^{-2\beta-4} dt \le t_0^{2\beta+4}\int_{t_0^2}^\infty t^{-\beta-4}\, dt = \frac{t_0^{-2}}{\beta+3} \le \frac{\cos^2\phi}{48} \le 1,$$

whereas the first integral is bounded by t_0^2 times the maximum value of the integrand over $(0, t_0^2)$. From (76), this is at most

$$t_0^2\left(\frac{\beta}{\cos\phi}\right)^\beta \le \left(2 + \frac{4}{\cos\phi}\right)^2 \max\left(\left(\frac{\beta}{\cos\phi}\right)^\beta, \left(\frac{\beta}{\cos\phi}\right)^{\beta+2}\right).$$

This completes the proof. □

10. The asymptotics of $d_b(n)$. In the following section, we take

$$N \equiv N(n) = \lceil \log n \rceil \tag{77}$$

and recall from (52) that $b = o(n/\log n)$ and that $L_n = \min(b \log n, b^{2/3}n^{1/3})$ so that $L_n = o(n)$. From condition (51) we can write

$$f(z) = K(1-z)^{-\theta} + \Psi(z), \tag{78}$$

where Ψ is analytic in $\Delta \setminus \{1\}$ and is $O((1-z)^{\delta-\theta})$ as $z \to 1$ in Δ. We begin by removing the $\Psi(z)$ terms from the numerator of (66).

LEMMA 15.

$$d_b^* = o\left(\frac{b}{n}\right) + \sum_{r=0}^{L_n} \mathbb{P}(T_{0b} = r) \times \left(\frac{[z^n]K(1-z)^{-\theta} - \exp\left(\sum_{j=1}^{b}\lambda_j\right)[z^{n-r}]\left(h(z)K(1-z)^{-\theta}\right)}{[z^n]f(z)}\right)^+ .$$

PROOF. Use (78) to write (66) in the form

$$d_b^* = \sum_{r=0}^{L_n} \mathbb{P}(T_{0b} = r) \times \left(\frac{[z^n]K(1-z)^{-\theta} - \exp\left(\sum_{j=1}^{b}\lambda_j\right)[z^{n-r}]\left(h(z)K(1-z)^{-\theta}\right)}{[z^n]f(z)} + \frac{[z^n]\Psi(z) - \exp\left(\sum_{j=1}^{b}\lambda_j\right)[z^{n-r}](h(z)\Psi(z))}{[z^n]f(z)}\right)^+ .$$

In order to use Lemma 11 we must show that

$$(79)\qquad \max_{0\le r\le L_n}\left|\frac{[z^n]\Psi(z) - \exp\left(\sum_{j=1}^{b}\lambda_j\right)[z^{n-r}](h(z)\Psi(z))}{[z^n]f(z)}\right| = o\left(\frac{b}{n}\right).$$

We can use (69) and $h(1) = \exp(-\sum_{j=1}^{b}\lambda_j)$ to conclude that (79) is dominated by the sum of three terms:

$$(80)\qquad \max_{0\le r\le L_n}\left|\frac{[z^n]\Psi(z) - [z^{n-r}]\Psi(z)}{[z^n]f(z)}\right|;$$

$$(81)\qquad +\max_{0\le r\le L_n}\left|\frac{\exp\left(\sum_{j=1}^{b}\lambda_j\right)[z^{n-r}]\left\{\Psi(z)\left(\sum_{k=1}^{N}\left(h^{(k)}(1)/k!\right)(z-1)^k\right)\right\}}{[z^n]f(z)}\right|;$$

$$(82)\qquad +\max_{0\le r\le L_n}\left|\frac{\exp\left(\sum_{j=1}^{b}\lambda_j\right)[z^{n-r}]\left\{\Psi(z)R_N(z)(z-1)^{N+1}\right\}}{[z^n]f(z)}\right|.$$

We now show that each of the expressions (80)–(82) is $o(b/n)$.

Define $\Phi(z) = \Psi(z)(1-z)$ and notice that $\Phi(z) = O((1-z)^{\delta-\theta+1})$. Using the identity $1 - z^r = (1-z)(1 + z + \cdots + z^{r-1})$, we obtain $\forall\ r \in [0, L_n]$,

$$\begin{aligned}\left|[z^n]\Psi(z) - [z^{n-r}]\Psi(z)\right| &= \left|[z^n]\Phi(z) + \cdots + [z^{n-r+1}]\Phi(z)\right| \\ &\le r \max_{0\le l\le r-1}\left|[z^{n-l}]\Phi(z)\right| \\ &\le L_n \max_{0\le l\le L_n}\left|[z^{n-l}]\Phi(z)\right| \\ &= O\left(L_n n^{\theta-\delta-2}\right).\end{aligned}$$

The last step follows from Theorem 2 and the fact that $n/(n - L_n) = O(1)$. Using Theorem 5 to get the asymptotics for $[z^n]f(z)$, it follows that

$$\max_{0 \le r \le L_n} \left| \frac{n}{b} \frac{[z^n]\Psi(z) - [z^{n-r}]\Psi(z)}{[z^n]f(z)} \right| = O\left(\frac{nL_n n^{\theta-\delta-2}}{bn^{\theta-1}} \right)$$
$$= O\left(\frac{n^{-\delta}L_n}{b} \right) \to 0.$$

Next we consider expression (81). Using the result of Lemma 12, the numerator is bounded above by

$$\max_{0 \le r \le L_n} \sum_{k=1}^{N} \frac{\rho^k B_k b^k}{k!} |[z^{n-r}]\Psi(z)(1-z)^k|. \tag{83}$$

We now apply the result of Lemma 14 to each term in the absolute value signs, taking $\psi(z) = \Psi(z)(1-z)^k$ and $r(z) \equiv 1$, $\alpha = \theta - \delta - k$, $k = 1, 2, \dots, N$. The second term in brackets on the right of (74) yields an $O(1)$ bound, uniformly in k and r, which follows by taking η fixed and noting that the exponential growth of $(1+\eta)^{n-L_n}$ relative to n dominates $\sup_{k \le N} n^{-\alpha-1} = O(n^{\log n})$. For the first term, $W_{l,\phi}$, we take a bound depending on k: if $k \ge \theta - \delta$, then $|\alpha| = k - \theta + \delta \le k + \delta \equiv l$, so that from (75),

$$W_{l,\phi} \le A\left(\frac{k+\delta}{\cos\phi} \right)^{(k+\delta+2)}.$$

Using the fact that $B_k \le k!$, we see that n/b times expression (83) is at most

$$K_1 \rho n^{\theta-\delta-1} \sum_{k=1}^{N} \left(\frac{b\rho}{n} \right)^{k-1} \max_{0 \le r \le L_n} \left(1 - \frac{r}{n} \right)^{\theta-\delta-k-1} \big(W_{k+\delta,\phi} + O(1) \big).$$

Since $L_n \le n/2$, it follows that $\max_{0 \le r \le L_n}(1 - r/n)^{\theta-\delta-k-1} \le \max(1, 2^{k+1+\delta-\theta})$. Using this and the bound (75), we see that, for $k \ge \theta - \delta$, the kth term in the series above is bounded by

$$u_k = 2\left(\frac{2\rho b}{n} \right)^{k-1} \left(\frac{k+\delta}{\cos\phi} \right)^{k+\delta+2},$$

which satisfies

$$\frac{u_{k+1}}{u_k} = O\left(\frac{Nb}{n} \right) \to 0,$$

using (52) and (77). We conclude that the $k = 1$ term of (83) dominates. Combining this bound with the asymptotic behavior of $[z^n]f(z)$ determined by Theorem 5 [namely, $[z^n]f(z) \sim K_1 n^{\theta-1}/\Gamma(\theta)$], we see that indeed

$$\frac{n}{b}\{\text{expression (81)}\} = O(n^{-\delta}) \to 0.$$

Finally we consider the term (82). First,

$$\begin{aligned}&\{\text{expression (82)}\}\\ (84)\qquad &\le \frac{\exp\left(\sum_{j=1}^{b}\bar\theta/j\right)}{[z^n]f(z)} \max_{0\le r\le L_n} |[z^{n-r}]R_N(z)\Psi(z)(1-z)^{N+1}|\\ &\le \frac{(eb)^{\bar\theta}}{[z^n]f(z)} \max_{0\le r\le L_n} |[z^{n-r}]R_N(z)\Psi(z)(1-z)^{N+1}|.\end{aligned}$$

We now use Lemma 14 with

$$r(z) = R_N(z), \qquad \psi(z) = \Psi(z)(1-z)^{N+1},$$

so that from (78), $\alpha = \theta - \delta - N - 1$. Combining (84) with the results of Lemmas 13 and 14, we see that

$$\begin{aligned}&\{\text{expression (82)}\}\\ (85)\qquad &\le \frac{(eb)^{\bar\theta}}{[z^n]f(z)} \max_{0\le r\le L_n} K_1\eta^{-(N+1)}\left((eb)^{\bar\theta e^{\eta b}} + \exp(e^{\rho\eta b}-1)\right)\\ &\quad\times(n-r)^{\alpha-1}\left[W_{l,\phi} + \frac{\eta^{-\alpha}}{(1+\eta)^{n-r}}(n-r)^{1-\alpha}\right],\end{aligned}$$

where $l = |\theta - \delta - N - 1|$. There are essentially two cases to consider: b fixed and $b \to \infty$ as $n \to \infty$.

In the first case, we may take $\eta = \eta_0$, determined by the Δ-domain of f. As in the previous part of the proof, the rightmost term in brackets in (85) has an $O(1)$ bound uniform in $0 \le r \le L_n$. Since $n - L_n \ge n/2$, we may use (75) and the fact that $[z^n]f(z) \sim Kn^{\theta-1}/\Gamma(\theta)$ to deduce that (85) is bounded by a term of the form

$$\left(\frac{CN}{n}\right)^N \times \text{a polynomial in } n,$$

for some positive constant C. Since $N \asymp \log n$, we see that this term is $o(n^{-r})$ for any r. It follows that indeed

$$\frac{n}{b}\{\text{expression (82)}\} = o(1).$$

In the case $b \to \infty$, we may take $\eta = 1/b$ and assume that n is so large that $\eta \le \eta_0$. The second term in brackets in (85) is

$$\begin{aligned}\frac{\eta^{-\alpha}}{(1+\eta)^{n-r}}(n-r)^{1-\alpha} &\le \frac{n^{1-\alpha}b^{\alpha}(1+\eta)^r}{(1+\eta)^n}\\ (86)\qquad &\le n(1+\eta)^{L_n}2^{-n/b}\left(\frac{b}{n}\right)^{\alpha}\\ &\le n^2 2^{-n/b}\left(\frac{b}{n}\right)^{\alpha},\end{aligned}$$

the last two inequalities following from the fact that $(1+\eta)^n \geq 2^{n/b}$ and $(1+\eta)^{L_n} \leq n$. Hence the right side of (85) is bounded above by

$$\frac{(en)^{\bar{\theta}}}{[z^n]f(z)} K_1 n^{\alpha-1} 2^{1-\alpha} b^{N+1}\left((en)^{\bar{\theta}e} + \exp(e^{\rho}-1)\right) \times \left[W_{l,\phi} + n^2 2^{-n/b}\left(\frac{b}{n}\right)^{\alpha}\right], \tag{87}$$

since $n - L_n \geq n/2$. Using the bound for $W_{l,\phi}$ once more, the first term above is of the form

$$\left(\frac{CbN}{n}\right)^N \times \text{a polynomial in } n,$$

which is $o(n^{-r})$ for any r since $bN/n \to 0$ and $N \asymp \log n$. The second term is of the form

$$2^{-n/b} \times \text{a polynomial in } n,$$

which is also $o(n^{-r})$ for any r since $\log n = o(n/b)$. Thus in case $b \to \infty$,

$$\frac{n}{b}\{\text{expression (82)}\} = o(1),$$

completing the proof of the lemma. □

In Lemma 17, we need to bound $\max_{0 \leq r \leq L_n} |[z^{n-r}](1-z)^{-\alpha}|$, with $\alpha = \theta, \theta-1, \ldots, \theta-N-1$. This is accomplished in the following lemma.

LEMMA 16. *For* $1 \leq m \leq n$,

$$|[z^m](1-z)^{-\alpha}| \leq m^{\alpha-1} \leq \left(\frac{n}{m}\right) n^{\alpha-1} \qquad (0 < \alpha < 1)$$
$$\leq e^{\alpha-1} n^{\alpha-1} \qquad (1 \leq \alpha)$$

and if $\alpha = \theta - k$, $\theta > 0$, $k = 1, 2, \ldots, m-1$, *then*

$$|[z^m](1-z)^{-\alpha}| \leq k!\max(\theta^k, 1)\left(\frac{n}{m-k}\right)^{k+1} e^{\theta} n^{\alpha-1}.$$

PROOF. For $\alpha > 0$, we have

$$|[z^m](1-z)^{-\alpha}| = \left|\binom{-\alpha}{m}\right| = \frac{\alpha}{1}\frac{\alpha+1}{2}\cdots\frac{\alpha+m-1}{m}$$
$$= \prod_1^m \left(1 - \frac{1-\alpha}{j}\right) \leq \exp\left(-(1-\alpha)\left(\sum_1^m \frac{1}{j}\right)\right).$$

Depending on whether $\alpha < 1$ or not, we bound the harmonic sum below or above by $\log m$ or $1 + \log m$. For the case $\alpha = \theta - k$, we use falling factorials to write

$$\left|\binom{-\theta+k}{m}\right| = \left|\frac{(-\theta+k)_k}{(m)_k}\frac{(-\theta)_{m-k}}{(m-k)!}\right|.$$

In absolute value, the first factor is bounded by $k!\max(\theta^k, 1)(m-k)^{-k}$, and using the first part of this lemma, the second factor is bounded by $ne^{\theta}n^{\theta-1}/(m-k)$. □

Now we remove the $\Psi(z)$ terms in the denominator.

LEMMA 17.

$$d_b^* = o\left(\frac{b}{n}\right) + \sum_{r=0}^{L_n} \mathbb{P}(T_{0b} = r) \times \left(\frac{[z^n](1-z)^{-\theta} - \exp\left(\Sigma_{j=1}^b \lambda_j\right)[z^{n-r}]\left(h(z)(1-z)^{-\theta}\right)}{[z^n](1-z)^{-\theta}}\right)^+ .$$

PROOF. Rewrite Lemma 15 as

$$d_b^* = o\left(\frac{b}{n}\right) + \sum_{r=0}^{L_n} \mathbb{P}(T_{0b} = r) \times \left(\frac{[z^n]K(1-z)^{-\theta} - \exp\left(\Sigma_{j=1}^b \lambda_j\right)[z^{n-r}]\left(h(z)K(1-z)^{-\theta}\right)}{[z^n]K(1-z)^{-\theta}} - [z^n]\Psi(z)\frac{[z^n]K(1-z)^{-\theta} - \exp\left(\Sigma_{j=1}^b \lambda_j\right)[z^{n-r}]\left(h(z)K(1-z)^{-\theta}\right)}{[z^n]K(1-z)^{-\theta}[z^n]f(z)}\right)^+ .$$

In this case, to use Lemma 11 we must show that

$$\max_{0\le r\le L_n}\left|[z^n]\Psi(z)\frac{[z^n]K(1-z)^{-\theta} - \exp\left(\Sigma_{j=1}^b \lambda_j\right)[z^{n-r}]\left(h(z)K(1-z)^{-\theta}\right)}{[z^n]K(1-z)^{-\theta}[z^n]f(z)}\right| \tag{88}$$

is $o(b/n)$. Expression (88) is dominated by the sum of three terms:

$$\max_{0\le r\le L_n}\left|[z^n]\Psi(z)\frac{[z^n]K(1-z)^{-\theta} - [z^{n-r}]K(1-z)^{-\theta}}{[z^n]K(1-z)^{-\theta}[z^n]f(z)}\right|; \tag{89}$$

$$+\max_{0\le r\le L_n}\left|[z^n]\Psi(z)\frac{\Sigma_{k=1}^N\left(B_k(\rho b)^k/k!\right)[z^{n-r}]K(1-z)^{-\theta+k}}{[z^n]K(1-z)^{-\theta}[z^n]f(z)}\right|; \tag{90}$$

$$+\max_{0\le r\le L_n}\left|[z^n]\Psi(z)\frac{\exp\left(\Sigma_{j=1}^b \lambda_j\right)[z^{n-r}]\left(R_N(z)(1-z)^{-\theta+N+1}\right)}{[z^n]K(1-z)^{-\theta}[z^n]f(z)}\right|. \tag{91}$$

To bound (89) we use the binomial expansion: for all $r \in [0, L_n]$,

$$(1 - r/n)^{\theta - 1} = 1 + O(r/n). \tag{92}$$

It then follows from (63) that

$$\begin{aligned}
[z^n](1-z)^{-\theta} - [z^{n-r}](1-z)^{-\theta} &= \frac{1}{\Gamma(\theta)}\left(n^{\theta-1} - (n-r)^{\theta-1}\right) + O(n^{\theta-2}) \\
&= \frac{n^{\theta-1}}{\Gamma(\theta)}\left(1 - \left(1 - \frac{r}{n}\right)^{\theta-1}\right) + O(n^{\theta-2}) \\
&= O(rn^{\theta-2}) \\
&= O(L_n n^{\theta-2}).
\end{aligned}$$

Hence from Theorem 4,

$$\begin{aligned}
\frac{n}{b}\{\text{expression (89)}\} &= O\left(\frac{n}{b} n^{\theta-\delta-1} \frac{L_n n^{\theta-2}}{n^{\theta-1} n^{\theta-1}}\right) \\
&= O\left(\frac{L_n}{b} n^{-\delta}\right) \to 0.
\end{aligned}$$

The argument to show that expression (90) is $o(b/n)$ is similar to that used to analyze (81), using Lemma 16. The argument to show that expression (91) is $o(b/n)$ is analogous to that used to analyze (82). We omit further details. □

Next we remove the higher terms in the expansion of h from the numerator in Lemma 17.

LEMMA 18.

$$d_b^* = o\left(\frac{b}{n}\right) + \sum_{r=0}^{L_n} \mathbb{P}(T_{0b} = r) \times \left(\frac{[z^n](1-z)^{-\theta} - [z^{n-r}]\{(1-z)^{-\theta} - \mathbb{E}T_{0b}(1-z)^{-\theta+1}\}}{[z^n](1-z)^{-\theta}}\right)^{+}.$$

PROOF. We rewrite Lemma 17 in the form

$$d_b^* = o\left(\frac{b}{n}\right) + \sum_{r=0}^{L_n} \mathbb{P}(T_{0b} = r) \times \left(\left\{\frac{\left([z^n](1-z)^{-\theta} - \exp\left(\sum_{j=1}^{b}\lambda_j\right)[z^{n-r}]h(1)(1-z)^{-\theta}\right)}{[z^n](1-z)^{-\theta}} + \frac{\exp\left(\sum_{j=1}^{b}\lambda_j\right)[z^{n-r}]\{h'(1)\}(1-z)^{-\theta+1}}{[z^n](1-z)^{-\theta}}\right\}\right.$$

$$-\left\{\frac{\exp(\Sigma_{j=1}^b\lambda_j)\Sigma_{k=2}^N[z^{n-r}](-1)^k(h^{(k)}(1)/k!)(1-z)^{-\theta+k}}{[z^n](1-z)^{-\theta}}\right.$$
$$\left.\left.+\frac{\exp(\Sigma_{j=1}^b\lambda_j)(-1)^{N+1}[z^{n-r}]\big(R_N(z)(1-z)^{-\theta+N+1}\big)}{[z^n](1-z)^{-\theta}}\right\}\right)^+.$$

The expressions in large braces play the roles of $S_{1,n,r}$ and $S_{2,n,r}$ in Lemma 11. Now $\max_{0\le r\le L_n}|S_{2,n,r}|$ is dominated by the sum of two terms:

$$\max_{0\le r\le L_n}\left|\frac{\max(\Sigma_{j=1}^b\lambda_j)\Sigma_{k=2}^N[z^{n-r}](-1)^k(h^{(k)}(1)/k!)(1-z)^{-\theta+k}}{[z^n](1-z)^{-\theta}}\right|; \tag{93}$$

$$+\max_{0\le r\le L_n}\left|\frac{\exp(\Sigma_{j=1}^b\lambda_j)[z^{n-r}]R_N(z)(1-z)^{-\theta+N+1}}{[z^n](1-z)^{-\theta}}\right|. \tag{94}$$

Now

$$\{\text{expression (93)}\}\le\max_{0\le r\le L_n}\frac{\Sigma_{k=2}^N(B_k/k!)(\rho b)^k|[z^{n-r}](1-z)^{-\theta+k}|}{|[z^n](1-z)^{-\theta}|},$$

which is $o(b/n)$ by an argument similar to that used to analyze expression (90). That expression (94) is also $o(b/n)$ follows from an argument akin to that applied to expression (91). We have shown that

$$d_b^*=o\left(\frac{b}{n}\right)+\sum_{r=0}^{L_n}\mathbb{P}(T_{0b}=r)$$
$$\times\left(\frac{[z^n](1-z)^{-\theta}-\exp(\Sigma_{j=1}^b\lambda_j)[z^{n-r}]h(1)(1-z)^{-\theta}}{[z^n](1-z)^{-\theta}}\right.$$
$$\left.+\left(\frac{\exp(\Sigma_{j=1}^b\lambda_j)[z^{n-r}]h'(1)(1-z)^{-\theta+1}}{[z^n](1-z)^{-\theta}}\right)\right)^+.$$

To complete the proof we use the fact that $h(1)=\exp(-\Sigma_{j=1}^b\lambda_j)$ and

$$h'(1)=\left(-\sum_{j=1}^b j\lambda_j\right)\exp\left(-\sum_{j=1}^b\lambda_j\right)=-\mathbb{E}T_{0b}\exp\left(-\sum_{j=1}^b\lambda_j\right).\qquad\square$$

Using the expansion in (63), we can simplify d_b further. Define

$$G_{n,\theta}\equiv\Gamma(\theta)[z^n](1-z)^{-\theta}-n^{\theta-1}.$$

By (63), for fixed $\theta>0$,

$$G_{n,\theta}=-\tfrac{1}{2}\theta(1-\theta)n^{\theta-2}+O(n^{\theta-3}). \tag{95}$$

LEMMA 19.

$$d_b^* = \sum_{r=0}^{L_n} \mathbb{P}(T_{0b} = r)\left(\frac{n^{\theta-1} - (n-r)^{\theta-1} - (\theta-1)\mathbb{E}T_{0b}(n-r)^{\theta-2}}{n^{\theta-1}}\right)^+ + o\left(\frac{b}{n}\right).$$

PROOF. Write Lemma 18 in the form

$$d_b^* = \sum_{r=0}^{L_n} \mathbb{P}(T_{0b} = r)\left(\frac{n^{\theta-1} - (n-r)^{\theta-1} - (\theta-1)\mathbb{E}T_{0b}(n-r)^{\theta-2}}{\Gamma(\theta)[z^n](1-z)^{-\theta}} + \frac{G_{n,\theta} - G_{n-r,\theta} - (\theta-1)\mathbb{E}T_{0b}G_{n-r,\theta-1}}{\Gamma(\theta)[z^n](1-z)^{-\theta}}\right)^+ + o\left(\frac{b}{n}\right)$$

and apply Lemma 11 again. In this case it suffices to show that

$$\max_{0\le r\le L_n}\left|\frac{G_{n,\theta} - G_{n-r,\theta}}{\Gamma(\theta)[z^n](1-z)^{-\theta}}\right| = o\left(\frac{b}{n}\right) \tag{96}$$

and

$$\max_{0\le r\le L_n}\left|\frac{\mathbb{E}T_{0b}G_{n-r,\theta-1}}{\Gamma(\theta)[z^n](1-z)^{-\theta}}\right| = o\left(\frac{b}{n}\right). \tag{97}$$

First, we need a bound on the numerator of (96). Using (95) and (92), we have for all $0 \le r \le L_n$,

$$\begin{aligned}|G_{n,\theta} - G_{n-r,\theta}| &= \tfrac{1}{2}| - \theta(1-\theta)n^{\theta-2} + (1-\theta)\theta(n-r)^{\theta-2}| + O(n^{\theta-3})\\ &= \tfrac{1}{2}\theta|1-\theta|n^{\theta-2}|1 - (1-r/n)^{\theta-2}| + O(n^{\theta-3})\\ &= O(L_n n^{\theta-3}),\end{aligned}$$

implying that

$$\begin{aligned}\frac{n}{b}\{\text{expression (96)}\} &= O\left(\frac{n}{b}\frac{L_n n^{\theta-3}}{n^{\theta-1}}\right)\\ &= O\left(\frac{L_n}{bn}\right) \to 0.\end{aligned}$$

The numerator of (97) is bounded by

$$\mathbb{E}T_{0b}G_{n-r,\theta-1} = O(bn^{\theta-3})$$

so that

$$\frac{n}{b}\{\text{expression (97)}\} = O\left(\frac{n}{b}\frac{bn^{\theta-3}}{n^{\theta-1}}\right) = O\left(\frac{1}{n}\right) \to 0.$$

We have now seen that

$$\begin{aligned} d_b^* = \sum_{r=0}^{L_n} \mathbb{P}(T_{0b}=r)\left(\frac{n^{\theta-1}-(n-r)^{\theta-1}-(\theta-1)\mathbb{E}T_{0b}(n-r)^{\theta-2}}{\Gamma(\theta)[z^n](1-z)^{-\theta}}\right)^+ \\ + o\left(\frac{b}{n}\right). \end{aligned} \tag{98}$$

Finally, we substitute $n^{\theta-1}$ in place of $\Gamma(\theta)[z^n](1-z)^{-\theta}$ in the denominator of (98). This makes a difference of at most $O(1/n)$, since $\Gamma(\theta)[z^n](1-z)^{-\theta} = n^{\theta-1}(1+O(1/n))$ and $d_b^* \le 1$. □

LEMMA 20.

$$d_b^* = \frac{1}{n}\sum_{r=0}^{L_n} \mathbb{P}(T_{0b}=r)((\theta-1)(r-\mathbb{E}T_{0b}))^+ + o\left(\frac{b}{n}\right).$$

PROOF. We write Lemma 19 in the form

$$\begin{aligned} d_b^* &= \sum_{r=0}^{L_n} \mathbb{P}(T_{0b}=r)\left(1-\left(1-\frac{r}{n}\right)^{\theta-1} - (\theta-1)\frac{\mathbb{E}T_{0b}}{n}\left(1-\frac{r}{n}\right)^{\theta-2}\right)^+ + o\left(\frac{b}{n}\right) \\ &= \sum_{r=0}^{L_n} \mathbb{P}(T_{0b}=r)\left(\frac{\theta-1}{n}(r-\mathbb{E}T_{0b}) - \left((\theta-1)\frac{r}{n} - \left(1-\left(1-\frac{r}{n}\right)^{\theta-1}\right)\right)\right. \\ &\qquad \left. + (\theta-1)\mathbb{E}T_{0b}\frac{1-(1-r/n)^{\theta-2}}{n}\right)^+ + o\left(\frac{b}{n}\right). \end{aligned}$$

We use the binomial expansions

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

and

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

to see first that

$$\frac{n}{b}\max_{0\le r\le L_n}\left|(\theta-1)\frac{r}{n} - \left(1-\left(1-\frac{r}{n}\right)^{\theta-1}\right)\right| = O\left(\frac{n}{b}\frac{L_n^2}{n^2}\right) = O\left(\frac{L_n^2}{nb}\right) \to 0$$

and then that

$$\frac{n}{b}\max_{0\le r\le L_n}\left|(\theta-1)\frac{\mathbb{E}T_{0b}}{n}\left(1-\left(1-\frac{r}{n}\right)^{\theta-2}\right)\right| = O\left(\frac{L_n}{n}\right)\to 0,$$

proving the lemma. □

PROOF OF THEOREM 3. We have seen that

$$\begin{aligned} d_b &= d_b^* + o\left(\frac{b}{n}\right)\\ &= \frac{1}{n}\sum_{r=0}^{L_n}\mathbb{P}(T_{0b}=r)((\theta-1)(r-\mathbb{E}T_{0b}))^+ + o\left(\frac{b}{n}\right)\\ &= \frac{1}{n}\sum_{r=0}^{\infty}\mathbb{P}(T_{0b}=r)((\theta-1)(r-\mathbb{E}T_{0b}))^+ + o\left(\frac{b}{n}\right)\\ &= \frac{1}{n}\mathbb{E}((\theta-1)(T_{0b}-\mathbb{E}T_{0b}))^+ + o\left(\frac{b}{n}\right), \end{aligned}$$

the first equality following from (65), the second from Lemma 20 and the third from Lemma 10. The proof is completed by noting that $\mathbb{E}(T_{0b}-\mathbb{E}T_{0b}) = 0$, which implies that

$$\mathbb{E}((\theta-1)(T_{0b}-\mathbb{E}T_{0b}))^+ = \mathbb{E}((\theta-1)(T_{0b}-\mathbb{E}T_{0b}))^-,$$

so that

$$\mathbb{E}((\theta-1)(T_{0b}-\mathbb{E}T_{0b}))^+ = \tfrac{1}{2}|\theta-1|\mathbb{E}|T_{0b}-\mathbb{E}T_{0b}|. \qquad \square$$

11. Discussion. The asymptotics in Theorem 3 are proved for the case $b = o(n/\log n)$. It is shown in Arratia, Barbour and Tavaré (1992) that for ESF(θ), $d_b(n) \to$ iff $b = o(n)$. We conjecture that the conclusions of Theorem 3 remain valid assuming only that $b = o(n)$ rather than $b = o(n/\log n)$. For the case $b(n) = \lfloor n/\log\log n\rfloor$ and $\theta = 1/2$, we computed values of $d_b(n)$ for ESF(θ) using a variant of the method described in Arratia and Tavaré (1992a). The results are given in Table 1, together with the ratio of $d_b(n)$ to its value predicted from the right-hand side of (59). These results support our conjecture.

We have seen in (34) that asymptotics of the form $d_b(n) \sim cb/n$ are available from Theorem 3 for some cases in which (20) holds, but (31) does not. There are other assemblies, not satisfying Theorem 3, for which $d_b(n)$ does not appear to decay like cb/n. The simplest examples are provided by the hybrid Ewens sampling formula structure determined by (33), with $\theta_2 > \theta_1$. In Table 2, we present the values of $d_1(n)$ for the special case $\theta_1 + \theta_2 = 2$, for four values of θ_1. Intuitively, we might expect that $d_1(n) = o(n^{-1})$, in accordance with the heuristic behind Theorem 3. However, com-

TABLE 1
Values of $d_b(n)$ for $b = \lfloor n/\log\log n \rfloor$

n	b	$d_b(n)$	Ratio*
5,000	2,334	0.171	3.68
10,000	4,503	0.153	3.42
20,000	8,722	0.138	3.18
25,000	10,798	0.134	3.11
50,000	20,996	0.122	2.90
100,000	40,925	0.111	2.72
200,000	79,938	0.102	2.55
400,000	156,423	0.093	2.40

*Ratio = $d_b(n)$/(predicted value), from (59).

TABLE 2
Values of $d_1(n)$ for hybrid ESF with $\theta_1 + \theta_2 = 2$

n	$\theta_1 = 0.75$	$\theta_1 = 0.50$	$\theta_1 = 0.25$	$\theta_1 = 0.01$
10	0.0659	0.1354	0.1457	0.00982
50	0.0187	0.0631	0.1069	0.00974
100	0.0111	0.0452	0.0929	0.00971
250	0.0056	0.0289	0.0767	0.00966
500	0.0033	0.0206	0.0660	0.00963
1000	0.0020	0.0146	0.0567	0.00959
2000	0.0012	0.0104	0.0485	0.00956
3000	0.0009	0.0085	0.0442	0.00954
5000	0.0006	0.0066	0.0393	0.00951

pletely different behavior is shown in the table. An examination of the two-term local limit heuristic used in (61) shows that

$$|\mathbb{P}(T_{1n} = n - k) - \mathbb{P}(T_{1n} = n - k - 1)| = O(n^{-\theta_1}),$$

leading us to conjecture that $d_1(n) = O(n^{-\theta_1})$ as well. This is consistent with the numerical evidence in Table 2.

Finally, we note that results analogous to Theorem 3 for combinatorial multisets and selections in the logarithmic class have been established by Stark (1994).

Acknowledgments. This paper is based on part of the Ph.D. research of Dudley Stark. We thank the referees and an Associate Editor for comments that have substantially improved the presentation of the results.

REFERENCES

AHLFORS, L. (1979). *Complex Analysis*. McGraw-Hill, New York.

ARRATIA, R. and TAVARÉ, S. (1992a). The cycle structure of random permutations. *Ann. Probab.* **20** 1567–1591.

ARRATIA, R. and TAVARÉ, S. (1992b). Limit theorems for combinatorial structures via discrete process approximations. *Random Structures Algorithms* **3** 321–345.

ARRATIA, R. and TAVARÉ, S. (1994). Independent process approximations for random combinatorial structures. *Adv. Math.* **104** 90–154.

ARRATIA, R., BARBOUR, A. D. and TAVARÉ, S. (1992). Poisson process approximations for the Ewens sampling formula. *Ann. Appl. Probab.* **2** 519–535.

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

ARRATIA, R., BARBOUR, A. D. and TAVARÉ, S. (1994). Logarithmic combinatorial structures. Unpublished manuscript.

DELAURENTIS, J. M. and PITTEL, B. (1985). Random permutations and Brownian motion. *Pacific J. Math.* **119** 287–301.

ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence.* Wiley, New York.

EWENS, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoret. Population Biol.* **3** 87–112.

FLAJOLET, P. and ODLYZKO, A. (1990a). Singularity analysis of generating functions. *SIAM J. Disc. Math.* **3** 216–240.

FLAJOLET, P. and ODLYZKO, A. (1990b). Random mapping statistics. *Proc. Eurocrypt '89. Lecture Notes in Computer Sci.* **434** 329–354. Springer, Berlin.

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.

FOATA, D. (1974). *La Série Génératrice Exponentielle dans les Problèmes d'Énumeration.* Univ. Montréal Press.

GONCHAROV, V. L. (1944). Some facts from combinatorics. *Izv. Akad. Nauk. SSSR. Ser. Mat.* **8** 3–48. (See also: On the field of combinatory analysis. *Trans. Amer. Math. Soc.* **19** 1–46.)

GRADSHTEYN, I. S. and RYZHIK, I. M. (1980). *Tables of Integrals, Series, and Products.* Academic Press, New York.

GRIFFITHS, R. C. (1988). On the distribution of points in a Poisson–Dirichlet process. *J. Appl. Probab.* **25** 336–345.

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. (1994). Order statistics for decomposable combinatorial structures. *Random Structures Algorithms* **5** 517–533.

HARRIS, B. (1960). Probability distributions related to random mappings. *Ann Math. Statist.* **31** 1045–1062.

IGNATOV, T. (1982). On a constant arising in the asymptotic theory of symmetric groups, and on Poisson–Dirichlet measures. *Theory Probab. Appl.* **27** 136–147.

JOYAL, A. (1981). Une théorie combinatoire des séries formelles. *Adv. Math.* **42** 1–82.

KOLCHIN, V. F. (1971). A problem of the allocation of particles in cells and cycles of random permutations. *Theory Probab. Appl.* **16** 74–90.

KOLCHIN, V. F. (1976). A problem of the allocation of particles in cells and random mappings. *Theory Probab. Appl.* **21** 48–63.

KOLCHIN, V. F. (1986). *Random Mappings.* Optimization Software, Inc., New York.

MOON, J. W. (1970). *Counting Labelled Trees. Canadian Mathematical Monographs* **1**. Clowes and Sons, London.

STARK, D. (1994). Total variation asymptotics for independent process approximations of logarithmic multisets and selections. Unpublished manuscript.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTHERN CALIFORNIA
LOS ANGELES, CALIFORNIA 90089-1113