Reprinted from ADVANCES IN MATHEMATICS
All Rights Reserved by Academic Press, New York and London

Vol. 104, No. 1, March 1, 1994
Printed in Belgium

Independent Process Approximations for Random Combinatorial Structures

RICHARD ARRATIA AND SIMON TAVARÉ*

Department of Mathematics, University of Southern California, Los Angeles, California 90089-1113

DEDICATED TO THE MEMORY OF MARK KAC, WHO SOUGHT OUT INDEPENDENCE IN ANALYSIS AND NUMBER THEORY

Many random combinatorial objects have a component structure whose joint distribution is equal to that of a process of mutually independent random variables, conditioned on the value of a weighted sum of the variables. It is interesting to compare the combinatorial structure directly to the independent discrete process, without renormalizing. The quality of approximation can often be conveniently quantified in terms of total variation distance, for functionals which observe part, but not all, of the combinatorial and independent processes. Among the examples are combinatorial assemblies (e.g., permutations, random mapping functions, and partitions of a set), multisets (e.g., polynomials over a finite field, mapping patterns and partitions of an integer), and selections (e.g., partitions of an integer into distinct parts, and square-free polynomials over finite fields). We consider issues common to all the above examples, including equalities and upper bounds for total variation distances, existence of limiting processes, heuristics for good approximations, the relation to standard generating functions, moment formulas and recursions for computing densities, refinement to the process which counts the number of parts of each possible type, the effect of further conditioning on events of moderate probability, large deviation theory and nonuniform measures on combinatorial objects, and the possibility of getting useful results by overpowering the conditioning. © 1994 Academic Press, Inc.

Contents.

* The authors were supported in part by NSF Grant DMS90-05833. We thank Andrew Barbour, Béla Bollobás, Persi Diaconis, Jennie Hansen, Lars Holst, Jim Pitman, and Gian-Carlo Rota for helpful comments on earlier drafts of this paper.

0001-8708/94 $9.00
Copyright © 1994 by Academic Press, Inc.
All rights of reproduction in any form reserved.

1. INTRODUCTION

We consider random combinatorial objects which can be described in terms of their component structure. For an object of weight n, denote the component structure by

$$\mathbf{C} \equiv \mathbf{C}(n) \equiv (C_1(n), C_2(n), \ldots, C_n(n)),$$

where $C_i \equiv C_i(n)$ is the number of components of size i. Since iC_i is the total weight in components of size i, we have

$$C_1 + 2C_2 + \cdots + nC_n = n.$$

For each fixed n, by choosing an object of weight n at random, with all possibilities equally likely, we view $\mathbf{C}(n)$ as a $\mathbb{Z}_+^n$-valued stochastic process, whose coordinates $C_i(n)$, $i = 1, \ldots, n$, are *dependent*, nonnegative integer-valued random variables. This paper considers combinatorial objects for which the joint distribution of $\mathbf{C}(n)$ can be expressed as the joint distribution of *independent* random variables $Z_1, Z_2, \ldots, Z_n$ conditioned on the value of a particular weighted sum.

There are at least three broad classes of combinatorial structures which have this description in terms of conditioning an independent process. The first class is assemblies of labelled structures on $[n] \equiv \{1, 2, \ldots, n\}$ see Foata [23], Joyal [38]. This class includes permutations, decomposed into cycles; mappings, decomposed into connected components; graphs, decomposed into connected components, and partitions of a finite set. The second class is multisets, i.e., unordered samples taken with replacement. This class includes partitions of an integer; random mapping patterns; and monic polynomials over a finite field, decomposed into monic irreducible factors. The third class is selections, i.e., unordered samples taken without replacement, including partitions of an integer into parts of distinct sizes, and square-free polynomials.

The detailed description of any of the above examples is given in terms of a sequence of nonnegative integers $m_1, m_2, \ldots$. For assemblies, let m_i be the number of labelled structures on a set of size i, for $i = 1, 2, \ldots$; permutations have $m_i = (i-1)!$, mappings have $m_i = (i-1)!\,(1 + i + i^2/2 + \cdots + i^{i-1}/(i-1)!)$, and partitions of a set have $m_i = 1$. For multisets and selections, let m_i be the number of objects of weight i; partitions of an integer have $m_i = 1$, and the factorizations of monic polynomials over a finite field have m_i equal to the number of monic, irreducible polynomials of degree i.

For $\mathbf{a} \equiv (a_1, a_2, \ldots, a_n) \in \mathbb{Z}_+^n$, consider the number $N(n, \mathbf{a})$ of objects of total weight n, having a_i components of size i, for $i = 1$ to n. For assemblies, the generalization of Cauchy's formula for permutations is the enumeration

$$N(n, \mathbf{a}) \equiv |\{\text{assemblies on } [n]: \mathbf{C} = \mathbf{a}\}|$$
$$= \mathbf{1}(a_1 + 2a_2 + \cdots + na_n = n)\, n! \prod_1^n \frac{m_i^{a_i}}{(i!)^{a_i} a_i!}. \tag{1}$$

For multisets,

$$N(n, \mathbf{a}) \equiv |\{\text{multisets of weight } n: \mathbf{C} = \mathbf{a}\}|$$
$$= \mathbf{1}(a_1 + 2a_2 + \cdots + na_n = n) \prod_1^n \binom{m_i + a_i - 1}{a_i}. \tag{2}$$

For selections,

$$N(n, \mathbf{a}) \equiv |\{\text{selections of weight } n: \mathbf{C} = \mathbf{a}\}|$$
$$= \mathbf{1}(a_1 + 2a_2 + \cdots + na_n = n) \prod_1^n \binom{m_i}{a_i}. \tag{3}$$

Let $p(n)$ denote the total number of structures of weight n, to wit

$$p(n) = \sum_{\mathbf{a} \in \mathbb{Z}_+^n} N(n, \mathbf{a}). \tag{4}$$

For permutations, $p(n) = n!$; for mappings, $p(n) = n^n$; for graphs, $p(n) = 2^{\binom{n}{2}}$; for partitions of a set $p(n) = B_n$, the Bell number; for partitions of an integer, $p(n)$ is the standard notation; and for monic polynomials over a field with q elements, $p(n) = q^n$.

A *random structure* is understood as follows. Fix a constant n, and choose one of the $p(n)$ structures at random, with each possibility equally likely. This makes $\mathbf{C}(n)$ a stochastic process with values in $\mathbf{Z}_+^n$, whose distribution is determined by

$$\mathbb{P}(\mathbf{C}(n) = \mathbf{a}) \equiv \frac{N(n, \mathbf{a})}{p(n)}, \qquad \mathbf{a} \in \mathbb{Z}_+^n. \tag{5}$$

In Section 2 below, we show that there are independent random variables $Z_1, Z_2, \ldots$ such that the combinatorial distribution (5) is equal to the joint distribution of $(Z_1, Z_2, \ldots, Z_n)$ conditional on the event $\{T_n = n\}$, where

$$T_n \equiv Z_1 + 2Z_2 + \cdots + nZ_n.$$

Explicitly, for all $\mathbf{a} \in \mathbb{Z}_+^n$

$$\mathbb{P}(\mathbf{C}(n) = \mathbf{a}) = \mathbb{P}((Z_1, Z_2, \ldots, Z_n) = \mathbf{a} \mid T_n = n). \tag{6}$$

Assemblies, multisets, and selections are not the only places where (6) arises in combinatorics. For example, the distribution of the counts of the factor degrees of the characteristic polynomial of a uniformly chosen random matrix over a finite field also satisfies (6); see Hansen and Schmutz [62].

It is fruitful to compare the combinatorial structure directly to the independent discrete process, without renormalizing. The quality of approximation can be usefully quantified in terms of total variation distance between the restrictions of the dependent and independent processes to a subset of the possible coordinates. We carry this out in Section 3. Bounds and limit theorems for natural functionals which depend on the coordinates, albeit weakly on those outside a subset, are then easily obtained as corollaries. For examples of this in the context of random polynomials over finite fields, and random permutations and random mappings, see Arratia, Barbour, and Tavaré [5] and Arratia and Tavaré [3].

The comparison of combinatorial structures to independent processes, with and without further conditioning, has a long history. Perhaps the best known example is the representation of the multinomial distribution with parameters n and $p_1, \ldots, p_k$ as the joint law of independent Poisson random variables with means $\lambda p_1, \ldots, \lambda p_k$, conditional on their sum being equal to n.

Holst [34] provides an approach to urn models that unifies multinomial, hypergeometric, and Pólya sampling. The joint laws of the dependent counts of the different types sampled are represented, respectively, as the joint distribution of independent Poisson, negative binomial, and binomial random variables, conditioned on their sum. See also Holst [35, 36]. The quality of such approximations is assessed using metrics, including the total variation distance, by Stam [53] and Diaconis and Freedman [13].

The books by Kolchin, Sevast'yanov, and Chistyakov [40] and Kolchin [39] use the representation of combinatorial structures, including random permutations and random mappings, in terms of independently distributed random variables, conditioned on the value of their sum. However, the Kolchin technique requires that the independent variables be *identically* distributed. The number of components C_i of size i is the number of random variables which take on the value i.

Shepp and Lloyd [52] study random permutations using a conditional relation almost identical to (6), with $\mathbb{E}Z_i = x^i/i$ and $x = x(n)$, except that they condition on n being the value of an infinite sum $Z_1 + 2Z_2 + \cdots$, which of course entails that $Z_{n+1} = Z_{n+2} = \cdots = 0$, and requires $x < 1$. Variants on the Shepp and Lloyd technique are discussed by Diaconis and Pitman [14], are effectively exploited to prove functional central limit theorems for two combinatorial assemblies by Hansen [29, 30], and are used as a convenient tool for moment calculations by Watterson [58] and Hansen [31]. A related technique, coupled with an observation of Levin [41], is used by Fristedt [24, 25] to study random partitions of a set and random partitions of an integer.

1.1. *Notation*

There are several types of asymptotic relations used in this paper. For sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \sim b_n$ for the asymptotic relation $a_n/b_n \to 1$ as $n \to \infty$. We write $a_n \asymp b_n$ if there are constants $0 < c_0 \leqslant c_1 < \infty$ such that $c_0 b_n \leqslant a_n \leqslant c_1 b_n$ for all sufficiently large n. We write $a_n \approx b_n$ to denote that $\log a_n \sim \log b_n$. Finally, we say that $a_n \doteq b_n$ if a_n and b_n are approximately equal in some heuristic sense deliberately left vague.

For $r \in \mathbb{Z}_+ \equiv \{0, 1, 2, \ldots\}$, we denote the rising factorial $y_{(r)}$ by $y_{(0)} = 1$, $y_{(r)} = y(y+1)\cdots(y+r-1)$ and the falling factorial $y_{[r]}$ by $y_{[0]} = 1$, $y_{[r]} = y(y-1)\cdots(y-r+1)$. We also write $\mathbb{N} \equiv \{1, 2, \ldots\}$, $\mathbb{R}_+ \equiv [0, \infty)$.

We write $X_n \to_p X$ if X_n converges to X in probability, $X_n \Rightarrow X$ if X_n converges to X in distribution, and $X \stackrel{\mathrm{d}}{=} Y$ if X and Y have the same distribution. We use $\mathbf{1}$ to denote indicator functions, so that $\mathbf{1}(A) = 1$ if A is true and $\mathbf{1}(A) = 0$ otherwise.

2. Independent Random Variables Conditioned on a Weighted Sum

2.1. *The Combinatorial Setup*

Common to the enumerations (1) through (3) is the form

$$N(n, \mathbf{a}) \equiv |\{\mathbf{C} = \mathbf{a}\}| = \mathbf{1}(a_1 + 2a_2 + \cdots + na_n = n) f(n) \prod_1^n g_i(a_i), \qquad (7)$$

with $f(n) = n!$ for assemblies, and $f(n) \equiv 1$ for multisets and selections. To see that (7) involves independent random variables conditioned on a weighted sum, view the right hand side as a product of three factors. First, the indicator function, which depends on both n and $\mathbf{a}$, corresponds to conditioning on the value of a weighted sum. Second, the factor $f(n)$ does not depend on $\mathbf{a}$, and hence disappears from conditional probabilities. The product form of the third factor corresponds to n mutually independent, but not identically distributed, random variables.

The distribution of a random assembly, multiset, or selection $\mathbf{C}(n)$ given in (5) can now be expressed in the following form. For $\mathbf{a} \in \mathbb{Z}_+^n$,

$$\mathbb{P}(\mathbf{C}(n) = \mathbf{a}) = \mathbf{1}(a_1 + 2a_2 + \cdots + na_n = n) \frac{f(n)}{p(n)} \prod_1^n g_i(a_i). \tag{8}$$

Given functions $g_1, g_2, \ldots$ from $\mathbb{Z}_+$ to $\mathbb{R}_+$, and a constant $x > 0$, let $Z_1, Z_2, \ldots$ be independent nonnegative integer valued random variables with distributions given by

$$\mathbb{P}(Z_i = k) = c_i g_i(k) x^{ik}, \qquad k = 0, 1, 2, \ldots. \tag{9}$$

The above definition, in which $c_i \equiv c_i(x)$ is the normalizing constant, makes sense if and only if the value of x and the functions g_i are such that

$$c_i \equiv \left(\sum_{k \geq 0} g_i(k) x^{ik} \right)^{-1} \in (0, \infty). \tag{10}$$

For assemblies, $g_i(k) = (m_i/i!)^k/k!$, so that (10) is satisfied for all $x > 0$. Defining $\lambda_i \equiv m_i x^i/i!$, we see that $c_i = \exp(-\lambda_i)$ and Z_i is Poisson with mean and variance

$$\mathbb{E} Z_i = \operatorname{var}(Z_i) = \lambda_i \equiv \frac{m_i x^i}{i!}. \tag{11}$$

For multisets, $g_i(k) = \binom{m_i + k - 1}{k}$, so the summability condition (10) is satisfied if and only if $x < 1$. For $x \in (0, 1)$, we have $c_i = (1 - x^i)^{m_i}$ and Z_i has the negative binomial distribution with parameters m_i and x^i given by

$$\mathbb{P}(Z_i = k) = \binom{m_i + k - 1}{k} (1 - x^i)^{m_i} x^{ik}, \qquad k = 0, 1, 2, \ldots,$$

with mean and variance

$$\mathbb{E} Z_i = \frac{m_i x^i}{1 - x^i}, \qquad \operatorname{var}(Z_i) = \frac{m_i x^i}{(1 - x^i)^2}. \tag{12}$$

In the special case $m_i = 1$, this is just the geometric distribution, and in general Z_i is the sum of m_i independent random variables each with the geometric distribution $\mathbb{P}(Y = k) = (1 - x^i) x^{ik}$ for $k \geq 0$.

For selections, $g_i(k) = \binom{m_i}{k}$, which is zero for $k > m_i$, so that (10) is satisfied for all $x > 0$. We see that $c_i = (1 + x^i)^{-m_i}$, by writing

$$\mathbb{P}(Z_i = k) = c_i \binom{m_i}{k} x^{ik} = \binom{m_i}{k} \left(\frac{x^i}{1 + x^i} \right)^k \left(\frac{1}{1 + x^i} \right)^{m_i - k}.$$

Thus, with $p_i = x^i/(1+x^i)$, the distribution of Z_i is binomial with parameters m_i and p_i, with mean and variance

$$\mathbb{E}Z_i = m_i p_i = \frac{m_i x^i}{1+x^i}, \qquad \operatorname{var}(Z_i) = m_i p_i (1-p_i) = \frac{m_i x^i}{(1+x^i)^2}. \tag{13}$$

2.2. *Conditioning on Weighted Sums in General*

In order to give a proof of (6) which will also serve in Section 6 on process refinements, and Section 8 on large deviations, we generalize to a situation that handles weighted sums with an arbitrary finite index set. We assume that I is a finite set, and for each $\alpha \in I$, $g_\alpha: \mathbb{Z}_+ \to \mathbb{R}_+$ is given. We assume that w is a given weight function with values in $\mathbb{R}$ or more generally $\mathbb{R}^d$, so that for $\alpha \in I$, $w(\alpha)$ is the weight of α. For the combinatorial examples in Section 1, we had $I = \{1, 2, ..., n\}$, and a one-dimensional space of weights, with $w(i) = i$. For $\mathbf{a} \in \mathbb{Z}_+^I$ with coordinates $a_\alpha \equiv a(\alpha)$, we use vector dot product notation for the weighted sum

$$\mathbf{w} \cdot \mathbf{a} \equiv \sum_{\alpha \in I} a(\alpha)\, w(\alpha).$$

Furthermore, we assume that we are given a target value t such that there exists a normalizing constant $f(I, t)$ so that the formula

$$\mathbb{P}(\mathbf{C}_I = \mathbf{a}) = \mathbf{1}(\mathbf{w} \cdot \mathbf{a} = t)\, f(I, t) \prod_{\alpha \in I} g_\alpha(a_\alpha), \qquad \mathbf{a} \in \mathbb{Z}_+^I \tag{14}$$

defines a probability distribution for a stochastic process $\mathbf{C}_I$ with values in $\mathbb{Z}_+^I$. The distribution of $\mathbf{C}(n)$ given by (8) is a special case of (14) with $t = n$ and $f(I, t) = f(n)/p(n)$.

Assume that for some value $x > 0$ there exist normalizing constants $c_\alpha \equiv c_\alpha(x) \in (0, \infty)$, such that for each $\alpha \in I$,

$$\mathbb{P}(Z_\alpha = k) = c_\alpha(x)\, g_\alpha(k)\, x^{w(\alpha)k}, \qquad k = 0, 1, 2, ... \tag{15}$$

defines a probability distribution on $\mathbb{Z}_+$. In case $d > 1$, so that $w(\alpha) = (w_1(\alpha), ..., w_d(\alpha))$, we take $x \equiv (x_1, ..., x_d) \in (0, \infty)^d$, and $x^{w(\alpha)k}$ denotes the product $x_1^{w_1(\alpha)k} \cdots x_d^{w_d(\alpha)k}$. Define the weighted sum T by

$$T \equiv T_I \equiv \sum_{\alpha \in I} w(\alpha)\, Z_\alpha. \tag{16}$$

It should now be clear that the following is a generalization of (6).

THEOREM 1. *Let $\mathbf{Z}_I \equiv (Z_\alpha)_{\alpha \in I}$ have independent coordinates Z_α with distribution given by* (15)*, and let $\mathbf{C}_I$ have the distribution given by* (14)*. Then*

$$\mathbf{C}_I \stackrel{\mathrm{d}}{=} (\mathbf{Z}_I \mid T = t), \tag{17}$$

and hence for any $B \subset I$, the processes restricted to indices in B satisfy

$$\mathbf{C}_B \stackrel{\mathrm{d}}{=} (\mathbf{Z}_B \mid T = t). \tag{18}$$

Furthermore, the normalizing constants and the conditioning probability are related by

$$\mathbb{P}(T = t) = f(I, t)^{-1} x^t \prod_{\alpha \in I} c_\alpha(x). \tag{19}$$

Remark. The distribution of $\mathbf{Z}_I$, and hence that of $T \equiv \mathbf{w} \cdot \mathbf{Z}_I$, depends on x, so the left side $\mathbb{P}(T = t)$ of (19) is a function of x.

Proof. The distribution of $\mathbf{Z}_I$ is given by

$$\mathbb{P}(\mathbf{Z}_I = \mathbf{a}) = \prod_{\alpha \in I} (c_\alpha g_\alpha(a_\alpha) x^{w(\alpha) a(\alpha)}) = x^{\mathbf{w} \cdot \mathbf{a}} \prod_{\alpha \in I} c_\alpha \prod_{\alpha \in I} g_\alpha(a_\alpha),$$

for $\mathbf{a} \in \mathbb{Z}_+^I$, so that if $\mathbf{w} \cdot \mathbf{a} = t$ then

$$\mathbb{P}(\mathbf{Z}_I = \mathbf{a}) = x^t \prod c_\alpha f(I, t)^{-1} \mathbb{P}(\mathbf{C}_I = \mathbf{a}). \tag{20}$$

The conditional distribution of $\mathbf{Z}_I$ given $\{T = t\}$ is given by

$$\begin{aligned}
\mathbb{P}(\mathbf{Z}_I = \mathbf{a} \mid T = t) &= \frac{\mathbf{1}(t = \mathbf{w} \cdot \mathbf{a})\, \mathbb{P}(\mathbf{Z}_I = \mathbf{a})}{\mathbb{P}(T = t)} \\
&= \frac{x^t (\prod c_\alpha) f(I, t)^{-1} \mathbb{P}(\mathbf{C}_I = \mathbf{a})}{\mathbb{P}(T = t)} \qquad (21) \\
&= \frac{x^t (\prod c_\alpha) f(I, t)^{-1} \mathbb{P}(\mathbf{C}_I = \mathbf{a})}{\sum_{\mathbf{b} \in \mathbb{Z}_+^I} x^t (\prod c_\alpha) f(I, t)^{-1} \mathbb{P}(\mathbf{C}_I = \mathbf{b})} \\
&= \frac{\mathbb{P}(\mathbf{C}_I = \mathbf{a})}{\sum_{\mathbf{b}} \mathbb{P}(\mathbf{C}_I = \mathbf{b})} \\
&= \mathbb{P}(\mathbf{C}_I = \mathbf{a}), \qquad \mathbf{a} \in \mathbb{Z}_+^I. \qquad (22)
\end{aligned}$$

The equality between (21) and (22), for any $\mathbf{a}$ for which $\mathbb{P}(\mathbf{C}_I = \mathbf{a}) > 0$, establishes (19). ∎

For the combinatorial objects in Section 1, $I = \{1, 2, ..., n\}$, and $w(i) = i$. For this case T reduces to

$$T \equiv T_n \equiv Z_1 + 2Z_2 + \cdots + nZ_n. \tag{23}$$

In the case of assemblies, corresponding to (1) and (11), the distribution of Z_i is Poisson(λ_i), and (19) reduces to

$$\mathbb{P}(T_n = n) = \frac{p(n)}{n!} x^n \exp(-\lambda_1 - \cdots - \lambda_n), \tag{24}$$

where $\lambda_i = m_i x^i/i!$ and $x > 0$. In the case of multisets, corresponding to (2) and (12), Z_i is distributed like the sum of m_i independent geometric (x^i) random variables, and (19) reduces to

$$\mathbb{P}(T_n = n) = p(n)\, x^n \prod_1^n (1 - x^i)^{m_i}, \tag{25}$$

for $0 < x < 1$. In the case of selections, corresponding to (3) and (13), the distribution of Z_i is binomial $(m_i, x^i/(1 + x^i))$, so that (19) reduces to

$$\mathbb{P}(T_n = n) = p(n)\, x^n \prod_1^n (1 + x^i)^{m_i}, \tag{26}$$

for $x > 0$.

3. Total Variation Distance

A useful way to establish that the independent process $\mathbf{Z}_n \equiv (Z_1, Z_2, ..., Z_n)$ is a good approximation for the dependent combinatorial process $\mathbf{C}(n)$ is to focus on a subset B of the possible component sizes, and give an upper bound on the total variation distance between the two processes, both restricted to B. Theorem 3 below shows how this total variation distance for these two processes reduces to the total variation distance between two one-dimensional random variables.

Here is a quick review of the relevant features of total variation distance. For two random elements X and Y of a finite or countable space S, the total variation distance between X and Y is defined by

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

Properly speaking this should be referred to as the distance between the distribution $\mathcal{L}(X)$ of X and the distribution $\mathcal{L}(Y)$ of Y written, for example, as $d_{TV}(\mathcal{L}(X), \mathcal{L}(Y))$. Throughout this paper we use the simpler notation, except in Section 8 which involves changes of measure.

Many authors, following the tradition of analysis of signed measures, omit the factor of 1/2. Using the factor of 1/2, we have that $d_{TV}(X, Y) \in [0, 1]$, and furthermore, d_{TV} is identical to the Prohorov metric, providing the underlying metric on S assigns distance $\geqslant 1$ between any two distinct points. In particular, a sequence of random elements X_n in a discrete space S converges in distribution to X if and only if $d_{TV}(X_n, X) \to 0$.

Another characterization of total variation distance is

$$d_{TV}(X, Y) = \max_{A \subset S} (\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)),$$

and in the discrete case, a necessary and sufficient condition that the maximum be achieved by A is that $\{s: \mathbb{P}(X=s) > \mathbb{P}(Y=s)\} \subset A \subset \{s: \mathbb{P}(X=s) \geqslant \mathbb{P}(Y=s)\}$.

The most intuitive description of total variation distance is in terms of coupling. A "coupling" of X and Y is a probability measure on S^2 whose first and second marginals are the distributions of X and Y, respectively. Less formally, a coupling of X and Y is a recipe for constructing X and Y simultaneously on the same probability space, subject only to having given marginal distributions for X and for Y. In terms of all possible coupling measures on S^2,

$$d_{TV}(X, Y) = \min_{\text{couplings}} \mathbb{P}(X \neq Y). \tag{27}$$

The minimum above is achieved, but in general there is not a unique optimal coupling. In fact a discrete coupling achieves $\mathbb{P}(X \neq Y) = d_{TV}(X, Y)$, if and only if, for all $s \in S$, $\mathbb{P}(X = Y = s) = \min(\mathbb{P}(X=s), \mathbb{P}(Y=s))$. Intuitively, if $d_{TV}(X, Y)$ is small, then X and Y are nearly indistinguishable from a single observation; formally, for any statistical test to decide whether X or Y is being observed, the sum of the type I and type II errors is at least $1 - d_{TV}(X, Y)$.

Upper bounds on the total variation distance between a combinatorial process and a simpler process are useful because these upper bounds are inherited by functionals of the processes. If $h: S \to T$ is a deterministic map between countable spaces, and X and Y are random elements of S, so that $h(X)$ and $h(Y)$ are random elements of T, then

$$d_{TV}(h(X), h(Y)) \leqslant d_{TV}(X, Y). \tag{28}$$

Theorem 3 below, and its refinement, Theorem 5 in Section 6, both describe combinatorially interesting cases in which equality holds in (28). It is natural to ask when, in general, such equality holds. The following elementary theorem provides an answer.

THEOREM 2. *In the discrete case, equality holds in* (28) *if and only if the sign of* $\mathbb{P}(X=s) - \mathbb{P}(Y=s)$ *depends only on* $h(s)$, *in the non-strict sense that* $\forall a, b \in S$,

$$h(a) = h(b) \text{ implies } (\mathbb{P}(X=a) - \mathbb{P}(Y=a))(\mathbb{P}(X=b) - \mathbb{P}(Y=b)) \geqslant 0.$$

Proof. Consider the proof of (28), namely

$$2d_{TV}(h(X), h(Y)) = \sum_{r \in T} |\mathbb{P}(h(X)=r) - \mathbb{P}(h(Y)=r)|$$

$$= \sum_{r} \left| \sum_{a \in S : h(a)=r} (\mathbb{P}(X=a) - \mathbb{P}(Y=a)) \right| \tag{29}$$

$$\leqslant \sum_{r} \sum_{a: h(a)=r} |\mathbb{P}(X=a) - \mathbb{P}(Y=a)| \tag{30}$$

$$= 2d_{TV}(X, Y).$$

Since the inequality in (30) holds term by term in the outer sums, equality holds overall if and only if equality holds for each r. This in turn is equivalent to the condition that for each r, there are no terms of opposite sign in the inner sum in (29). ∎

Diaconis and Pitman [14] view "sufficiency" as a key concept. In the context above, $h: S \to T$ is a sufficient statistic for discriminating between the distributions of X and Y in S, if the likehood ratio depends only on h, i.e., if there is a function $f: T \to \mathbb{R}$ such that for all $s \in S$, $\mathbb{P}(X=s) = f(h(s))\,\mathbb{P}(Y=s)$. Taking a sufficient statistic preserves total variation distance, as observed by Stam [53]. This is also a special case of Theorem 2, in which a product is nonnegative because it is a square: $(\mathbb{P}(X=a) - \mathbb{P}(Y=a))(\mathbb{P}(X=b) - \mathbb{P}(Y=b)) = (f(h(a)) - 1)(f(h(b)) - 1)\mathbb{P}(Y=a)\,\mathbb{P}(Y=b) \geqslant 0$ *whenever* $h(a) = h(b)$.

THEOREM 3. *Let I be a finite set, and for $\alpha \in I$, let C_α and Z_α be $\mathbb{Z}_+$ valued random variables, such that the Z_α are mutually independent. Let $\mathbf{w} = (w(\alpha))_{\alpha \in I}$ be a deterministic weight function on I with values in some linear space, let $T = \sum_{\alpha \in I} w(\alpha)\, Z_\alpha$, and let t be such that $\mathbb{P}(T=t) > 0$. For $B \subset I$, we use the notation $\mathbf{C}_B \equiv (C_\alpha)_{\alpha \in B}$ and $\mathbf{Z}_B \equiv (Z_\alpha)_{\alpha \in B}$ for random elements of $\mathbb{Z}_+^B$. Define*

$$R \equiv R_B \equiv \sum_{\alpha \in B} w(\alpha)\, Z_\alpha, \qquad S \equiv S_B \equiv \sum_{\alpha \in I - B} w(\alpha)\, Z_\alpha,$$

so that $T = R + S$ and R and S are independent. If

$$C_I \stackrel{\mathrm{d}}{=} (\mathbf{Z}_I \mid T = t), \tag{31}$$

then

$$d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) = d_{TV}((R_B \mid T=t), R_B). \tag{32}$$

Proof. We present two proofs, since it is instructive to contrast them. Note that not only are R and S independent, but also that R is a function of $\mathbf{Z}_B$, and $\mathbf{Z}_B$ and S are independent. For $\mathbf{a} \in \mathbb{Z}_+^B$, write $\mathbf{w} \cdot \mathbf{a} \equiv \sum_{\alpha \in B} w(\alpha)\, a(\alpha)$.

$$
\begin{aligned}
d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) &= \frac{1}{2} \sum_{\mathbf{a} \in \mathbb{Z}_+^B} |\mathbb{P}(\mathbf{Z}_B = \mathbf{a} \mid T = t) - \mathbb{P}(\mathbf{Z}_B = \mathbf{a})| \\
&= \frac{1}{2} \sum_r \sum_{\mathbf{a}: \mathbf{w} \cdot \mathbf{a} = r} \left| \frac{\mathbb{P}(\mathbf{Z}_B = \mathbf{a},\, r + S = t)}{\mathbb{P}(T = t)} - \mathbb{P}(\mathbf{Z}_B = \mathbf{a}) \right| \\
&= \frac{1}{2} \sum_r \sum_{\mathbf{a}: \mathbf{w} \cdot \mathbf{a} = r} \left| \frac{\mathbb{P}(\mathbf{Z}_B = \mathbf{a})\, \mathbb{P}(r + S = t)}{\mathbb{P}(T = t)} - \mathbb{P}(\mathbf{Z}_B = \mathbf{a}) \right| \\
&= \frac{1}{2} \sum_r \left| \frac{\mathbb{P}(R = r)\, \mathbb{P}(r + S = t)}{\mathbb{P}(T = t)} - \mathbb{P}(R = r) \right| \\
&= \frac{1}{2} \sum_r \left| \frac{\mathbb{P}(R = r,\, r + S = t)}{\mathbb{P}(T = t)} - \mathbb{P}(R = r) \right| \\
&= \frac{1}{2} \sum_r |\mathbb{P}(R = r \mid T = t) - \mathbb{R}(R = r)| \\
&= d_{TV}((R \mid T = t), R).
\end{aligned}
$$

Here is a second proof of Theorem 3, viewed as a corollary of Theorem 2, with the functional h on $\mathbb{Z}_+^B$ defined by $h(\mathbf{a}) = \mathbf{w} \cdot \mathbf{a}$. We need only observe that h is a sufficient statistic since $\mathbb{P}(\mathbf{Z}_B = \mathbf{a} \mid T = t) = \mathbb{P}(\mathbf{Z}_B = \mathbf{a})\, \mathbb{P}(S = t - h(\mathbf{a}))/\mathbb{P}(T = t)$. ∎

For the sake of calculations of total variation distance between a combinatorial process and its independent process approximation, the most useful form for the conclusion of Theorem 3 is

$$
\begin{aligned}
d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) &= \frac{1}{2} \sum_r \left| \frac{\mathbb{P}(R = r)\, \mathbb{P}(r + S = t)}{\mathbb{P}(T = t)} - \mathbb{P}(R = r) \right| \\
&= \frac{1}{2} \sum_r \mathbb{P}(R = r) \left| \frac{\mathbb{P}(S = t - r)}{\mathbb{P}(T = t)} - 1 \right|. \qquad (33)
\end{aligned}
$$

In the usual combinatorial case, where $t = n$ and $T = Z_1 + 2Z_2 + \cdots + nZ_n$, this gives

$$
d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) = \frac{1}{2}\, \mathbb{P}(R > n) + \frac{1}{2} \sum_{r=0}^{n} \mathbb{P}(R = r) \left| \frac{\mathbb{P}(S = n - r)}{\mathbb{P}(T = n)} - 1 \right|. \qquad (34)
$$

There are two elementary observations that point to strategies for giving upper bounds on total variation distance. First, for discrete random elements we have in general

$$\begin{aligned} d_{TV}(X, Y) &\equiv \tfrac{1}{2} \sum_{s \in S} |\mathbb{P}(X=s) - \mathbb{P}(Y=s)| \\ &= \sum_{s \in S} (\mathbb{P}(X=s) - \mathbb{P}(Y=s))^+ \\ &= \sum_{s \in S} (\mathbb{P}(X=s) - \mathbb{P}(Y=s))^-, \end{aligned}$$

where the notation for positive and negative parts is such that, for real x, $x = x^+ - x^-$, and $|x| = x^+ + x^-$. In the context of (33) this is useful in the following form. Let $A \subseteq I$. Then

$$\begin{aligned} d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) &= \sum_r \mathbb{P}(R=r) \left(1 - \frac{\mathbb{P}(S=t-r)}{\mathbb{P}(T=t)}\right)^+ \\ &\leqslant \mathbb{P}(R \notin A) + \sup_{r \in A} \left(1 - \frac{\mathbb{P}(S=t-r)}{\mathbb{P}(T=t)}\right)^+. \end{aligned} \tag{35}$$

Specializing to the case where the weighted sum R is real valued, and $A = \{0, 1, 2, ..., k\}$, the truncation level k is chosen much larger than $\mathbb{E}R$, so that large deviation theory can be used to bound $\mathbb{P}(R > k)$, but not too large, so that $\mathbb{P}(S=t-r)/\mathbb{P}(T=t)$ can be controlled to show it is close to one.

The second elementary observation, which is proved and exploited in Arratia and Tavaré [2], is that the denominator in (33) can be replaced by any constant $c > 0$, at the price of at most a factor of 2, in the sense that for independent R and S such that $\mathbb{P}(R+S=t) > 0$,

$$\frac{1}{2} \sum_r \mathbb{P}(R=r) \left| \frac{\mathbb{P}(S=t-r)}{\mathbb{P}(R+S=t)} - 1 \right| \leqslant \sum_r \mathbb{P}(R=r) \left| \frac{\mathbb{P}(S=t-r)}{c} - 1 \right|.$$

By using this, for example with $c = \mathbb{P}(S=t)$, giving an upper bound on the total variation distance for combinatorial process approximations is reduced to showing that the density of S is relatively constant.

Lower bounds for variation distance are often more difficult to obtain, but it is worth noting that in the combinatorial setup, since $\{R_B > n\} \subseteq \{\mathbf{C}_B \neq \mathbf{Z}_B\}$, we have, without the factor 1/2 suggested by (34),

$$d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) \geqslant \mathbb{P}(R_B > n). \tag{36}$$

4. Heuristics for Useful Approximation

Recall first that for $B \subset [n]$, we have $\mathbf{C}_B \stackrel{\mathrm{d}}{=} (\mathbf{Z}_B \mid T_n = n)$. If $d_{TV}(\mathbf{C}_B, \mathbf{Z}_B)$ is small, the approximation of $\mathbf{C}_B$ by $\mathbf{Z}_B$ is useful. Probabilistic intuition suggests that conditioning on $T_n = n$ does not change the distribution of $\mathbf{Z}_B$ by much, provided that the event $\{T_n = n\}$ is relatively likely. This in turn corresponds to a choice of $x = x(n)$ for which $\mathbb{E}T_n$ is approximately n. Let $\sigma_n^2 \equiv \mathrm{var}(T_n)$, and let $\sigma_B^2 = \mathrm{var}(R_B)$. Intuition then suggests that *if*

$$\frac{n - \mathbb{E}(T_n)}{\sigma_n} \text{ is not large} \tag{37}$$

and

$$\frac{\mathbb{E}R_B}{\sigma_n} \text{ and } \frac{\sigma_B}{\sigma_n} \text{ are small} \tag{38}$$

then $d_{TV}(\mathbf{C}_B, \mathbf{Z}_B)$ *is small.*

While our main focus is on the appropriate choice of x, we also discuss below the appropriate choice of B for examples including permutations, mappings, graphs, partitions of sets, and partitions of integers.

There is an important qualitative distinction between cases in which the appropriate x is constant, and those in which x varies with n. If x does not depend on n, then a single independent process $\mathbf{Z} = (Z_1, Z_2, \ldots)$ may be used to approximate $\mathbf{C}(n) \equiv (C_1(n), \ldots, C_n(n))$, which we identify with $(C_1(n), \ldots, C_n(n), 0, 0, \ldots) \in \mathbb{Z}_+^\infty$. Under the usual product topology on $\mathbb{Z}_+^\infty$, we have that $\mathbf{C}(n) \Rightarrow \mathbf{Z}$ if, and only if, for every fixed b, $\mathbf{C}_b(n) \equiv (C_1(n), \ldots, C_b(n)) \Rightarrow \mathbf{Z}_b \equiv (Z_1, \ldots, Z_b)$ as random elements in $\mathbb{Z}_+^b$. Since the metric on $\mathbb{Z}_+^b$ is discrete, we conclude that $\mathbf{C}_b(n) \Rightarrow \mathbf{Z}_b$ if, and only if, for each fixed b, $d_{TV}(\mathbf{C}_b(n), \mathbf{Z}_b) \to 0$. For cases where x, and hence $\mathbf{Z}$, varies with n, it makes no sense to write $\mathbf{C}(n) \Rightarrow \mathbf{Z}$. However, it is still useful to be able to estimate $d_{TV}(\mathbf{C}_B(n), \mathbf{Z}_B(n))$.

We discuss first considerations involved in the choice of x and B, and then heuristics for predicting the accuracy of approximation.

4.1. *Choosing the Free Parameter x*

It is convenient to discuss the three basic types of combinatorial structure separately.

4.1.1. *Assemblies*

It follows from (11) that

$$\mathbb{E}T_n \equiv \sum_{i=1}^{n} i\mathbb{E}Z_i = \sum_{i=1}^{n} \frac{m_i x^i}{(i-1)!}, \tag{39}$$

while

$$\sigma_n^2 = \sum_{i=1}^{n} i^2 \mathbb{E} Z_i = \sum_{i=1}^{n} \frac{i^2 m_i x^i}{i!}. \tag{40}$$

In the case of permutations, we take $x=1$ to see that $\mathbb{E}T_n = n$, and $\sigma_n^2 = n(n+1)/2$. In Arratia and Tavaré [2] it is proved that $d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) \to 0$ as $n \to \infty$, with $B = B(n)$, if and only if $|B| = o(n)$.

For the class of assemblies which satisfy the additional condition

$$\frac{m_i}{i!} \sim \frac{\kappa y^i}{i} \quad \text{as} \quad i \to \infty, \tag{41}$$

where $y > 0$ and $\kappa > 0$ are constants, we see that

$$\frac{\mathbb{E}T_n}{n} \to \begin{cases} 0, & \text{if } 0 < x < y^{-1} \\ \kappa, & \text{if } x = y^{-1} \\ \infty, & \text{if } x > y^{-1}. \end{cases}$$

Hence the only fixed x that ensures that $\mathbb{E}T_n \asymp n$ is $x = y^{-1}$, in which case

$$\mathbb{E}T_n \sim n\kappa, \qquad \sigma_n \sim n\sqrt{\frac{\kappa}{2}}. \tag{42}$$

For the example of random mappings,

$$m_i = e^i (i-1)!\, \mathbb{P}(\mathrm{Po}(i) < i),$$

where $\mathrm{Po}(i)$ denotes a Poisson random variable with mean i, Harris [33], Stepanov [56]. It follows that we must take $x = 1/e$, and, from the Central Limit Theorem, $\kappa = 1/2$. In this case $\mathbb{E}T_n \sim n/2$ and $\sigma_n \sim n/2$.

For the example of random graphs, with all $2^{\binom{n}{2}}$ graphs equally likely, the fact that the probability of being connected tends to 1 means that the constant vector $(0, 0, ..., 0, 1) \in \mathbb{Z}_+^n$ is a good approximation, in total variation distance, to $\mathbf{C}(n)$. This is a situation in which the equality $\mathbf{C}(n) =^{\mathrm{d}} (\mathbf{Z}_n \mid T_n = n)$ yields no useful approximation. With x chosen so that $\mathbb{E}T_n = n$, and $B = \{1, 2, ..., n-1\}$, we have that $d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) \to 0$, but only because both distributions are close to that of the process that is identically 0 on $\mathbb{Z}_+^B$.

For partitions of a set, which is discussed further in Subsection 5.2 and Section 10, with $x = x(n)$ being the solution of $xe^x = n$, and $B = \{1, 2, ..., b\} \cup \{c, c+1, ..., n\}$ where $b \equiv b(n)$ and $c \equiv c(n)$, the heuristic (37) suggests that $d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) \to 0$ if and only if both $(x-b)/\sqrt{\log n} \to \infty$ and $(c-x)/\sqrt{\log n} \to \infty$. For B of the complementary form $B = \{b, b+1, ..., c\}$

with $b<c$ both within a bounded number of $\sqrt{\log n}$ of x, the heuristic suggests that $d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) \to 0$ if, and only if, $(c-b) = o(\sqrt{\log n})$. Sachkov [51] and Fristedt [24] have partial results in this area.

4.1.2. *Multisets*

Using (12) we see that

$$\mathbb{E}T_n = \sum_{i=1}^{n} \frac{im_i x^i}{1-x^i}, \tag{43}$$

while

$$\sigma_n^2 = \sum_{i=1}^{n} \frac{i^2 m_i x^i}{(1-x^i)^2}. \tag{44}$$

If the multiset construction satisfies the additional hypothesis that

$$m_i \sim \frac{\kappa y^i}{i} \qquad \text{as} \quad i \to \infty, \tag{45}$$

where $y>1$ and $\kappa>0$ is fixed, a similar analysis shows that the only fixed x that ensures that $\mathbb{E}T_n \asymp n$ is $x = y^{-1}$, in which case the asymptotics for $\mathbb{E}T_n$ and σ_n are the same as those in (42).

The first example that satisfies the hypothesis in (45) is the multiset in which $p(n) = q^n$ for some integer $q \geqslant 2$. In this case the m_i satisfy

$$q^n = \sum_{j \mid n} j m_j, \tag{46}$$

so that by the Möbius inversion formula we have

$$m_n = \frac{1}{n} \sum_{k \mid n} \mu(n/k)\, q^k, \tag{47}$$

where $\mu(\cdot)$ is the Möbius function, defined by

$$\mu(n) = (-1)^k \qquad \text{if} \quad n \text{ is the product of } k \text{ distinct primes}$$

$$\mu(n) = 0 \qquad \text{otherwise.}$$

It follows from (46) that

$$q^i - \frac{q}{q-1} q^{i/2} \leqslant i m_i \leqslant q^i,$$

so that (45) holds with $\kappa = 1$, $y = q$. This construction arises in the study of necklaces (see Metropolis and Rota [44, 45], for example), in card shuffling (Diaconis, McGrath and Pitman [15]), and, for q a prime power, in factoring polynomials over $GF(q)$, a finite field of q elements. In this last case m_i is the number of irreducible monic polynomials over $GF(q)$; see Lidl and Niederreiter [42], for example.

Another example concerns random mapping patterns. Let t_n denote the number of rooted trees with n unlabelled points, and set $T(x) = \sum_{n=1}^{\infty} t_n x^n$. Otter [49] showed that $T(x)$ has radius of convergence $\rho = 0.3383...$, from which Meir and Moon [43] established that

$$m_i \sim \frac{\rho^{-i}}{2i}.$$

Hence (45) applies with $\kappa = 1/2$, $y = \rho^{-1}$.

For an example in which x varies with n, we consider random partitions of the integer n. In this case $m_i \equiv 1$. Taking $x = e^{-c/\sqrt{n}}$ and using (43), we see that

$$\begin{aligned} n^{-1}\mathbb{E}T_n &= \sum_{i=1}^{n} \frac{n^{-1/2} i \exp(-ic/\sqrt{n})}{1-\exp(-ic/\sqrt{n})} \frac{1}{\sqrt{n}} \\ &\to \int_0^{\infty} \frac{ye^{-cy}}{1-e^{-cy}}\, dy \\ &= \frac{1}{c^2} \int_0^1 \frac{-\log(1-v)}{v}\, dv \\ &= \frac{\pi^2}{6c^2}. \end{aligned}$$

Hence to satisfy $\mathbb{E}T_n \sim n$, we choose $c = \pi/\sqrt{6}$, so that

$$x = \exp(-\pi/\sqrt{6n}). \tag{48}$$

From (44), it follows by a similar calculation that

$$\begin{aligned} n^{-3/2}\sigma_n^2 &\to \int_0^{\infty} \frac{y^2 e^{-cy}}{(1-e^{-cy})^2}\, dy \\ &= \frac{1}{c^3} \int_0^1 \left(\frac{-\log(1-v)}{v} \right)^2 dv \\ &= \frac{2}{c}, \end{aligned}$$

so that

$$\sigma_n^2 \sim \frac{2\sqrt{6}}{\pi} n^{3/2}. \tag{49}$$

For sets of the form $B = \{1, 2, ..., b\} \cup \{c, c+1, ..., n\}$ where $0 \leqslant b \equiv b(n)$ and $c \equiv c(n) \leqslant n$, the heuristic in (37) and (38) suggests that $d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) \to 0$ if, and only if, both $b = o(\sqrt{n})$ and $c/\sqrt{n} \to \infty$. For B of the complementary form $B = \{b, b+1, ..., c\}$ with $b < c$ both of the order of $\sqrt{n}$, the heuristic suggests that $d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) \to 0$ if, and only if, $(c-b) = o(\sqrt{n})$. See Fristedt [25] and Goh and Schmutz [26] for related results.

4.1.3. *Selections*

In this case, it follows from (13) that

$$\mathbb{E}T_n = \sum_{i=1}^{n} \frac{im_i x^i}{1+x^i}, \tag{50}$$

while

$$\sigma_n^2 = \sum_{i=1}^{n} \frac{i^2 m_i x^i}{(1+x^i)^2}. \tag{51}$$

If the selection construction satisfies the additional hypothesis (45), then, just as for the assembly and multiset constructions, we take $x = y^{-1}$, and (42) holds once more. As an example, for square-free factorizations of polynomials over a finite field with q elements, we have $y = q$, $\kappa = 1$, $x = q^{-1}$.

For an example in which x varies with n, we consider once more random partitions of the integer n with all parts distinct, which is the selection construction with $m_i \equiv 1$. Taking $x = e^{-d/\sqrt{n}}$, and using (50) we see that

$$\begin{aligned} n^{-1}\mathbb{E}T_n &= \sum_{i=1}^{n} \frac{n^{-1/2} i \exp(-id/\sqrt{n})}{1+\exp(-id/\sqrt{n})} \frac{1}{\sqrt{n}} \\ &\to \int_0^\infty \frac{ye^{-dy}}{1+e^{-dy}}\, dy \\ &= \frac{1}{d^2} \int_0^1 \frac{-\log v}{1+v}\, dv \\ &= \frac{\pi^2}{12d^2}. \end{aligned}$$

Hence to satisfy $\mathbb{E}T_n \sim n$, we pick $d = \pi/\sqrt{12}$, so that

$$x = \exp(-\pi/\sqrt{12n}). \tag{52}$$

From (44), it follows by a similar calculation that

$$n^{-3/2}\sigma_n^2 \to \int_0^\infty \frac{y^2 e^{-dy}}{(1+e^{-dy})^2}\,dy$$

$$= \frac{1}{d^3}\int_0^1 \left(\frac{-\log v}{1+v}\right)^2 dv$$

$$= \frac{2}{d}.$$

For the choice of x in (52), we see that

$$\sigma_n^2 \sim \frac{4\sqrt{3}}{\pi} n^{3/2}. \tag{53}$$

To see how easy the heuristic for choosing x can be, consider partitions of the integer n with all parts distinct and odd. Compared to the above calculations, we are simply leaving out every other term, so that $n^{-1}\mathbb{E}T_n \to \pi^2/(24d^2)$, and we prescribe using $x = \exp(-\pi/\sqrt{24n})$. As with unrestricted partitions, using the appropriate x for either partitions with distinct parts or partitions with distinct odd parts, we believe that the unconditioned process $\mathbf{Z}_B$ is a good approximation for the combinatorial process $\mathbf{C}_B$, in the total variation sense, if and only if $b/\sqrt{n}$ is small and $c/\sqrt{n}$ is large, for $B = \{1, 2, ..., b\} \cup \{c, c+1, ..., n\}$. For B of the complementary form $B = \{b, b+1, ..., c\}$ with $b < c$ both of the order of $\sqrt{n}$, the heuristic suggests that $d_{TV}(\mathbf{C}_B, \mathbf{Z}_B)$ is small if, and only if, $(c-b)$ is small relative to $\sqrt{n}$.

4.2. *A Quantitative Heuristic*

In several examples, the $\mathbb{Z}_+$-valued random variables T_n, appropriately centered and rescaled, converge in distribution to a continuous limit X having a density f on $\mathbb{R}$. For illustration, we describe the important class of cases in which

$$\frac{T_n}{n} \Rightarrow X. \tag{54}$$

A local limit heuristic suggests the approximation

$$\mathbb{P}(T_n = n) \doteq \frac{f(1)}{n}, \tag{55}$$

where the sense of the approximation $\doteq$ is deliberately vague. Assuming that B is small, so that $R/n \to_p 0$, we also have $S/n \Rightarrow X$. For $0 \leqslant k \leqslant n$, the local limit heuristic gives

$$\mathbb{P}(S=n-k) \doteq \frac{1}{n} f\left(1-\frac{k}{n}\right),$$

and a Taylor expansion further simplifies this to

$$\mathbb{P}(S=n-k) \doteq \frac{1}{n}\left(f(1)-\frac{k}{n} f'(1-)\right). \tag{56}$$

Using these approximations in the total variation formula (33) gives

$$\begin{aligned} d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) &= \frac{1}{2} \sum_{k=0}^{n} \mathbb{P}(R=k) \left|1-\frac{\mathbb{P}(S=n-k)}{\mathbb{P}(T_n=n)}\right| + \frac{1}{2}\mathbb{P}(R>n) \\ &\doteq \frac{1}{2} \sum_{k \geqslant 0} \mathbb{P}(R=k) \left|1-\frac{n^{-1}(f(1)-n^{-1}kf'(1-))}{n^{-1}f(1)}\right| \\ &= \frac{1}{2} \frac{|f'(1-)|}{f(1)} \frac{\mathbb{E}\,|R|}{n}. \end{aligned}$$

However, this approximation ignores the essential feature $d_{TV}(\mu, \nu) = \frac{1}{2}|\mu-\nu|$, where the signed measure $\mu-\nu$ has net mass zero. Thus, even though $f(1)/n$ is the natural approximation for $\mathbb{P}(T_n=n)$, it is important to use a more complicated heuristic in which the approximation for T is the convolution of the distribution of R and our approximation for the distribution of S. Thus

$$\begin{aligned} \mathbb{P}(T=n) &= \sum_{k=0}^{n} \mathbb{P}(R=k)\,\mathbb{P}(S=n-k) \\ &\doteq \sum_{k \geqslant 0} \mathbb{P}(R=k) \frac{1}{n}\left(f(1)-\frac{k}{n} f'(1-)\right) \\ &= \frac{1}{n}\left(f(1)-\frac{\mathbb{E}R}{n} f'(1-)\right). \end{aligned} \tag{57}$$

Using this approximation

$$\begin{aligned} d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) &= \frac{1}{2} \sum_{k=0}^{n} \mathbb{P}(R=k) \left|1-\frac{\mathbb{P}(S=n-k)}{\mathbb{P}(T_n=n)}\right| + \frac{1}{2}\mathbb{P}(R>n) \\ &\doteq \frac{1}{2} \sum_{k \geqslant 0} \mathbb{P}(R=k) \left|1-\frac{n^{-1}(f(1)-n^{-1}kf'(1-))}{n^{-1}(f(1)-n^{-1}\mathbb{E}Rf'(1-))}\right| \end{aligned}$$

$$= \frac{1}{2} \sum_{k \geq 0} \mathbb{P}(R=k) \left| \frac{n^{-1}(k - \mathbb{E}R) f'(1-)}{f(1) - n^{-1} \mathbb{E}R f'(1-)} \right|$$

$$= \frac{1}{2n} |f'(1-)| \, \mathbb{E}\,|R - \mathbb{E}R| \, |f(1) - n^{-1}\mathbb{E}R f'(1-)|^{-1}$$

$$\doteq \frac{1}{2n} \frac{|f'(1-)|}{f(1)} \mathbb{E}\,|R - \mathbb{E}R|. \tag{58}$$

As a plausibility check, we note that the alternative approximation using $\mathbb{P}(T_n = n) \doteq (1/n) f(1)$ and $S \doteq T - \mathbb{E}R$, so that $\mathbb{P}(S = n-k) \doteq \mathbb{P}(T = n + \mathbb{E}R - k) \doteq (1/n) f(1 - (k - \mathbb{E}R)/n)$, also satisfies the convolutional property, and leads to the same first order result as (58).

One possible specific interpretation of the approximation in (58) would be the following pair of statements, giving a decay rate for d_{TV}, for fixed B, as $n \to \infty$.

If $T_n/n \Rightarrow X$, and X has density f with $f'(1-) \neq 0$, then

$$d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) \sim \frac{1}{2} \frac{|f'(1-)|}{f(1)} \frac{\mathbb{E}\,|R - \mathbb{E}R|}{n}. \tag{59}$$

If $T_n/n \Rightarrow X$, and X has density f with $f'(1-) = 0$, then

$$d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) = o\left(\frac{1}{n}\right). \tag{60}$$

For the more general case in which there are constants s_n such that

$$\frac{T_n - n}{s_n} \Rightarrow X,$$

where X has density f, these statements are to be replaced by

$$d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) \sim \frac{1}{2} \frac{|f'(0-)|}{f(0)} \frac{\mathbb{E}\,|R - \mathbb{E}R|}{s_n}, \qquad \text{if } f'(0-) \neq 0, \tag{61}$$

and

$$d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) = o\left(\frac{1}{s_n}\right), \qquad \text{if } f'(0-) = 0. \tag{62}$$

For partitions of an integer and for partitions of a set, a good choice for s_n is the standard deviation σ_n with asymptotics given by (49) and (160), and X is normally distributed, so that (62) should apply.

Observe that for two fixed sets B, B' the approximation in (59) or (61) has a corollary the statement that if $f'(0-) \neq 0$ then as $n \to \infty$,

$$\frac{d_{TV}(\mathbf{C}_B, \mathbf{Z}_B)}{d_{TV}(\mathbf{C}_{B'}, \mathbf{Z}_{B'})} \to \frac{\mathbb{E}\,|R_B - \mathbb{E}R_B|}{\mathbb{E}\,|R_{B'} - \mathbb{E}R_{B'}|}.$$

By the Cauchy–Schwarz inequality, $\mathbb{E}\,|R_B - \mathbb{E}R_B| \leqslant \sigma_B$, so another rigorous version of the heuristic in (58) would be the statement that as $n \to \infty$, $d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) = O(\sigma_B/\sigma_n)$ uniformly in B; that is,

$$\limsup_{n \to \infty} \sup_{B \subset [n]} \left(d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) \frac{\sigma_n}{\sigma_B} \right) < \infty. \tag{63}$$

Note that (63) is not embarrassed by the largest possible B, namely $B = [n]$, since $d_{TV}(\cdot, \cdot) \leqslant 1$.

4.3. *Examples with a Limit Process*: *The Logarithmic Class*

The previous section suggests that the limit law of T_n/n plays a key role in analyzing the accuracy of the approximation of certain combinatorial structures by independent processes. The logarithmic class consists of those assemblies which satisfy (41), and those multisets and selections which satisfy (45). All of these, with the appropriate constant choice of x, satisfy

$$i\mathbb{E}Z_i \to \kappa,\ i\mathbb{P}(Z_i = 1) \to \kappa \qquad \text{for some} \quad \kappa > 0. \tag{64}$$

Lemma 2 below shows that, for Z_i satisfying (64), and $T_n = Z_1 + 2Z_2 + \cdots + nZ_n$, the limit distribution of T_n/n depends only on the parameter κ.

Let d_W be the L_1 Wasserstein distance between distributions, which can be defined, in the same spirit as (27), by

$$d_W(X, Y) = \min_{\text{couplings}} \mathbb{E}\,|X - Y|.$$

For $\mathbb{Z}^+$-valued random variables, d_W is easily computed via

$$d_W(X, Y) = \sum_{i \geqslant 1} |\mathbb{P}(X \geqslant i) - \mathbb{P}(Y \geqslant i)|,$$

and when X is stochastically larger than Y, so that the absolute values above do nothing, this further simplifies to $d_W(X, Y) = \mathbb{E}X - \mathbb{E}Y$. Note that for integer-valued random values, $d_W \geqslant d_{TV}$.

Let $\tilde{Z}_i$ be Bernoulli with parameter $\kappa/i \wedge 1$, and let Z_i^* be Poisson with mean κ/i. It is easy to check that the condition (64) is equivalent to $d_W(Z_i, \tilde{Z}_i) = o(1/i)$. Since $d_W(\tilde{Z}_i, Z_i^*) = o(1/i)$, the triangle inequality implies that the condition (64) is also equivalent to $d_W(Z_i, Z_i^*) = o(1/i)$.

For the class of assemblies that satisfy the condition (41), we use $x = y^{-1}$ and $\mathbb{E}Z_i = m_i x^i/i!$, so that $\mathbb{E}Z_i \sim \kappa/i$. Lemma 1 applies directly; for Poisson random variables (64) is equivalent to $\mathbb{E}Z_i \sim \kappa/i$, so Lemma 2 also applies. For multisets and selections satisfying the hypothesis (45), it is easy to show that (64) holds.

LEMMA 1. *If Z_j are independent Poisson random variables with $\mathbb{E}Z_j = \lambda_j \sim \kappa/j$ for some constant $\kappa > 0$, and $T_n = \sum_{j=1}^n jZ_j$, then*

$$n^{-1}T_n \Rightarrow X_\kappa, \qquad n \to \infty \tag{65}$$

and X_κ has Laplace transform

$$\psi(s) \equiv \mathbb{E}e^{-sX_\kappa} = \exp\left(-\kappa \int_0^1 (1-e^{-sx})\frac{dx}{x}\right). \tag{66}$$

Proof. By direct calculation,

$$\begin{aligned}\log \mathbb{E}e^{-sT_n/n} &= -\sum_{j=1}^n \lambda_j(1-e^{-js/n}) \\ &= -\sum_{j=1}^n \frac{\kappa}{j}(1-e^{-js/n}) + \sum_{j=1}^n \left(\frac{\kappa}{j} - \lambda_j\right)(1-e^{-js/n}).\end{aligned}$$

Clearly, the first term on the right converges to $-\kappa \int_0^1 (1-e^{-sx})(dx/x)$. That the second term is $o(1)$ follows by observing that $\lambda_j - \kappa/j = o(j^{-1})$, and comparing to the first sum. ∎

LEMMA 2. *For $i = 1, 2, \ldots$, let Z_i be nonnegative integer-valued random variables satisfying the conditions in* (64). *If $T_n = \sum_{j=1}^n jZ_j$, then*

$$n^{-1}T_n \Rightarrow X_\kappa, \qquad n \to \infty \tag{67}$$

and X_κ has the Laplace transform given in (66).

Proof. Construct independent Bernoulli random variables $\tilde{Z}_i = Z_i \wedge 1$. Clearly $\tilde{Z}_i \leqslant Z_i$ and $\mathbb{P}(Z_i = 1) \leqslant \mathbb{E}\tilde{Z}_i \leqslant \mathbb{E}Z_i$. It follows that $i\mathbb{E}\tilde{Z}_i \to \kappa$. Therefore

$$i\,|\mathbb{E}Z_i - \mathbb{E}\tilde{Z}_i| = i(\mathbb{E}Z_i - \mathbb{E}\tilde{Z}_i) \to 0.$$

Hence if $\tilde{T}_n = \tilde{Z}_1 + \cdots + n\tilde{Z}_n$,

$$\mathbb{E}\left|\frac{T_n}{n} - \frac{\tilde{T}_n}{n}\right| \to 0. \tag{68}$$

It remains to show that $n^{-1}\tilde{T}_n \Rightarrow X_\kappa$.

For $i = 1, 2, \ldots$, let Z_i^* be independent Poisson random variables satisfying $p_i \equiv \mathbb{E}Z_i^* = \mathbb{E}\tilde{Z}_i \sim \kappa/i$. We may construct Z_i^* in such a way that for each i

$$\mathbb{E}\,|\tilde{Z}_i - Z_i^*| = d_W(\tilde{Z}_i, Z_i^*),$$

where d_W denotes Wasserstein L_1 distance. But if X is Bernoulli with parameter p and Y is Poisson with parameter p, then a simple calculation shows that $d_W(X, Y) = 2(p - 1 + e^{-p}) \leqslant p^2$. Hence

$$n^{-1}\mathbb{E}\,|\tilde{T}_n - T_n^*| \leqslant n^{-1} \sum_{i=1}^{n} ip_i^2 \to 0.$$

It follows that $n^{-1}\tilde{T}_n$ has the same limit law as $n^{-1}T_n^*$, which is that of X_κ, by Lemma 1. ∎

The random variable X_κ has appeared in several guises before, not least as part of the description of the density of points in a Poisson–Dirichlet process. See Watterson [60], Vershik and Shmidt [57], Ignatov [37], and Griffiths [28], Ethier and Kurtz [17] and the references contained therein. For our purposes, it is enough to record that the density $g(\cdot)$ of X_κ is known explicitly on the interval [0, 1],

$$g(z) = \frac{e^{-\gamma\kappa}}{\Gamma(\kappa)} z^{\kappa-1}, \qquad 0 \leqslant z \leqslant 1, \tag{69}$$

where γ is Euler's constant. From (69) follows the fact that

$$\frac{g'(1-)}{g(1)} = \kappa - 1. \tag{70}$$

We may now combine the previous results with (58) and (42) to rephrase the asymptotic behavior of $d_{TV}(\mathbf{C}_B, \mathbf{Z}_B)$ in (59) and (60) as follows. For any assembly satisfying (41), or for any multiset or selection satisfying (45), we should have the following decay rates, for any fixed B, as $n \to \infty$.

In the case $\kappa \neq 1$

$$d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) \sim \frac{1}{2}\,|\kappa - 1|\,\frac{\mathbb{E}\,|R - \mathbb{E}R|}{n}. \tag{71}$$

In the case $\kappa = 1$

$$d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) = o\left(\frac{1}{n}\right). \tag{72}$$

For a class of examples known as the Ewens sampling formula, described in Subsection 5.1, and for B of the form $B = \{1, 2, \ldots, b\}$, (71) is proved in Arratia, Stark, and Tavaré [7]. The analogous result for random mappings, in which $\kappa = 1/2$, and other assemblies that can be approximated by the Ewens sampling formula, may also be found there. For the corresponding results for multisets and selections, see Stark [55].

The statement (72) has been established for random permutations by Arratia and Tavaré [2], where it is shown inter alia that for $B=\{1, 2, ..., b\}$, $d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) \leqslant F(n/b)$, where $\log F(x) \sim -x \log x$ as $x \to \infty$. For the case of random polynomials over a finite field, Arratia, Barbour, and Tavaré [5] established that $d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) = O(b \exp(-cn/b))$, where $c = \frac{1}{2}\log(4/3)$.

Among the class of assemblies in the logarithmic class, weak convergence (in $\mathbb{R}^\infty$) of the component counting process to the appropriate Poisson process has been established for random permutations by Goncharov [27], for random mappings by Kolchin [63], and for the Ewens sampling formula by Arratia, Barbour, and Tavaré [4]. For multisets in the logarithmic class, this has been established for random polynomials by Diaconis, McGrath, and Pitman [15] and Arratia, Barbour, and Tavaré [5], and for random mapping patterns by Mutafciev [47].

5. Non-uniqueness in the Choice of the Parameter x

An appropriate choice of $x = x(n)$ for good approximation is not unique. An obvious candidate is that x which maximizes $\mathbb{P}(T_n = n)$, which is also that x for which $\mathbb{E}T_n = n$. This can be seen by differentiating $\log \mathbb{P}(T_n = n)$ in formulas (24)–(26) and comparing to $\mathbb{E}T_n$ from formulas (11)–(13); at the general level this is the observation that $\mathbb{P}(T = t)$ in (19) is maximized by that x for which $\mathbb{E}T = t$. Nevertheless, the obvious candidate is not always the best one.

We discuss here two qualitatively different examples: the logarithmic class, and partitions of a set.

5.1. *The Ewens Sampling Formula*

The central object in the logarithmic class is the Ewens sampling formula (ESF). This is the family of distributions with parameter $\kappa > 0$ given by (6), where the Z_i are independent Poisson random variables with $\mathbb{E}Z_i = \kappa/i$, or more generally, with

$$\lambda_i \equiv \mathbb{E}Z_i = \frac{\kappa x^i}{i}, \tag{73}$$

the conditional distribution being unaffected by the choice of $x > 0$. For $\kappa = 1$, the ESF is the distribution of cycle counts for a uniformly chosen random permutation. For $\kappa \neq 1$, the ESF can be viewed as the nonuniform measure on permutations with sampling bias proportional to $\kappa^{\#\text{cycles}}$; see Section 8 for details. The ESF arose first in the context of population genetics (Ewens [18]), and is given explicitly by

$$\mathbb{P}(C_1(n) = a_1, ..., C_n(n) = a_n) = \mathbf{1}\left(\sum_{l=1}^{n} la_l = n\right) \frac{n!}{\kappa_{(n)}} \prod_{i=1}^{n} \left(\frac{\kappa}{i}\right)^{a_i} \frac{1}{a_i!}. \tag{74}$$

The ESF corresponds to (41) with $y=1$ and the asymptotic relation in i replaced by equality. It is useful in describing all assemblies, multisets, and selections in the logarithmic class; see Arratia, Barbour and Tavaré [6] for further details.

For irrational κ the ESF cannot be realized as a uniform measure on a class of combinatorial objects. For rational $\kappa = r/s$ with integers $r>0$, $s>0$, there are at least two possibilities. First, comparing (6) with $\mathbb{E}Z_i = \kappa/i$, and (11) with $\mathbb{E}Z_i = m_i x^i/i!$, for any choice $x>0$, we take $x = 1/s$ to see that the ESF is the uniform measure on the assembly with $m_i = r(i-1)!\, s^{i-1}$. One interpretation of this is permutations on integers, enriched by coloring each cycle with one of r possible colors, and coloring each element of each cycle, except the smallest, with one of s colors. For a second construction, we use a device from Stark [54]. Consider permutations of ns objects, in which all cycle lengths must be multiples of s. Formally, this is the assembly on ns objects, with $m_i = (i-1)!\, \mathbf{1}(s \mid i)$, so that $(C_1, C_2, \ldots, C_{ns}) \stackrel{\mathrm{d}}{=} (Z_1, Z_2, \ldots, Z_{ns} \mid Z_1 + 2Z_2 + \cdots + nsZ_{ns} = ns)$, where Z_i is Poisson with $\mathbb{E}Z_i = \mathbf{1}(s \mid i)/i$. Since those C_i and Z_i for which s does not divide i are identically zero, we consider $C_i^* \equiv C_{is}$, $Z_i^* \equiv Z_{is}$, and $T_n^* \equiv Z_1^* + 2Z_2^* + \cdots + nZ_n^* = (1/s)(Z_1 + 2Z_2 + \cdots + nsZ_{ns})$. We have $(C_1^*, \ldots, C_n^*) \stackrel{\mathrm{d}}{=} (Z_1^*, \ldots, Z_n^* \mid T_n^* = n)$, and the Z_i^* are independent Poisson with $\mathbb{E}Z_i^* = 1/(si)$. Thus the distribution of $(C_1^*(n), \ldots, C_n^*(n))$ is the ESF with $\kappa = 1/s$. To change this to $\kappa = r/s$, we need only color each cycle with one of r possible colors, so that $m_i = r(i-1)!\, \mathbf{1}(s \mid i)$, $\mathbb{E}Z_i = r\mathbf{1}(s/i)/i$, and $\mathbb{E}Z_i^* = r/(si)$. To summarize our second construction of the ESF with $\kappa = r/s$, let $C_i^*(n)$ be the number of cycles of length si in a random permutation of ns objects, requiring that all cycle lengths be multiples of s, and assigning one of r possible colors to each cycle.

For comparing the ESF to the unconditioned, independent process $(Z_1, \ldots, Z_n)$ it is interesting to consider the role of varying x. The choice $x=1$ in (73), so that $\mathbb{E}Z_i = \kappa/i$, yields $\mathbb{E}T_n = \kappa n$, and $\sigma_n \sim n\sqrt{\kappa/2}$. In the case $\kappa \neq 1$ the discrepancy between $\mathbb{E}T_n$ and the goal n is a bounded multiple of σ_n. This is close enough for good approximation, in the sense that $(C_1(n), \ldots, C_n(n), 0, \ldots) \Rightarrow (Z_1, Z_2, \ldots)$. This, together with a $O(b/n)$ bound on $d_{TV}((C_1(n), \ldots, C_b(n)), (Z_1, \ldots, Z_b))$ that is uniform in $1 \leqslant b \leqslant n$, is proved in Arratia, Barbour, and Tavaré [4] by exploiting a coupling based on Feller [19]. This coupling provided even stronger information whose utility is discussed in Arratia and Tavaré [3]. Barbour [8] showed that the $O(b/n)$ bound above cannot be replaced by $o(b/n)$ for $x=1$, $\kappa \neq 1$.

For the case of independent Z_i' which are Poisson with means varying with n given by

$$\mathbb{E}Z_i' = \mathbb{E}C_i(n) = \frac{\kappa}{i} \frac{n(n-1)\cdots(n-i+1)}{(\kappa+n-i)\cdots(\kappa+n-1)},$$

Barbour [8] showed that $d_{TV}((C_1(n), ..., C_b(n)), (Z_1', ..., Z_b')) = O((b/n)^2)$, uniformly in $1 \leqslant b \leqslant n$. Observe that with this choice of Poisson parameters, $\mathbb{E}T_n' \sim \kappa n$ but it *not* the case that $(C_1(n), ..., C_n(n)) \stackrel{d}{=} (Z_1', ..., Z_n' \mid T_n' = n)$.

If we are willing to use coordinates $Z_i \equiv Z_i(n)$ whose means vary with n, we can still have the conditional relation (6) by using $x = x(n)$ in (73). An appealing family of choices is given by $x = \exp(-c/n)$, since this yields for $c \neq 0$

$$\mathbb{E}T_n = \sum_{i=1}^{n} i\lambda_i = \sum_{i=1}^{n} i \frac{\kappa}{i} e^{-ic/n} \sim n \frac{\kappa(1-e^{-c})}{c}. \tag{75}$$

By choosing $c \equiv c(\kappa)$ as the solution of $\kappa = c/(1-e^{-c})$, we can make $\mathbb{E}T_n \sim n$, and this should provide a closer approximation than the choice $c = 0$, $x = 1$. However, an even better choice of c is available. We explore this in the next subsection.

5.2. *More Accurate Approximations to the Logarithmic Class*

For assemblies, multisets, and selections in the logarithmic class discussed in Subsection 4.3, as well as for the ESF, the choice of x proportional to $\exp(-c/n)$ is interesting. In this situation, the limit law of T_n/n depends only on the parameters κ and c. Properties of this limit law lead to an optimal choice for c.

The following lemma applies to assemblies that satisfy the condition (41), and to the ESF by taking $m_i = \kappa(i-1)!$, $y = 1$, the m_i not necessarily being integers.

LEMMA 3. *Assume that $m_i \geqslant 0$ satisfies $m_i/i! \sim \kappa y^i/i$ for constants $y \geqslant 1$, $\kappa > 0$, and set $x = e^{-c/n}y^{-1}$ for constant $c \in \mathbb{R}$. If $Z_j \equiv Z_j(n)$ are independent Poisson random variables with $\mathbb{E}Z_j = m_j x^j/j!$, and $T_n = \sum_{j=1}^{n} jZ_j$, then*

$$n^{-1}T_n \Rightarrow X_{\kappa, c}, \qquad n \to \infty \tag{76}$$

and $X_{\kappa, c}$ has Laplace transform

$$\psi_c(s) \equiv \mathbb{E}e^{-sX_{\kappa,c}} = \exp\left(-\kappa \int_0^1 (1-e^{-sx}) \frac{e^{-cx}}{x}\, dx\right). \tag{77}$$

Proof. As in Lemma 1, calculate the limit of the log Laplace transform. ∎

Next we prove that the same limit law holds for multisets or selections satisfying the hypothesis (45).

LEMMA 4. *Assume that the multiset (or selection) satisfies* (45): $m_i \sim \kappa y^i/i$ *for constants* $y \geq 1$, $\kappa > 0$, *and set* $x = e^{-c/n} y^{-1}$. *If* $Z_j \equiv Z_j(n)$ *are independent negative binomial random variables with parameters* m_j *and* x^j *(respectively, binomial with parameters* m_j *and* $x^j/(1+x^j)$*) and* $T_n = \sum_{j=1}^n jZ_j$, *then*

$$n^{-1} T_n \Rightarrow X_{\kappa, c}, \qquad n \to \infty \tag{78}$$

and $X_{\kappa, c}$ *has the Laplace transform given in* (77).

Remark. For the case of multisets, we assume that $x < 1$.

Proof. Observe first that in either case, if $b = o(n)$, then $n^{-1} \mathbb{E} T_{0b} \to 0$, so that $n^{-1} T_{0b} \to_P 0$ as $n \to \infty$. Let $\tilde{Z}_j$ be independent Poisson random variables with $\mathbb{E}\tilde{Z}_j = m_j x^j$, and write $\tilde{T}_n = \sum_{j=1}^n j\tilde{Z}_j$, $\tilde{T}_{bn} = \sum_{j=b+1}^n j\tilde{Z}_j$. We show that for $b = o(n)$, T_{bn}/n and $\tilde{T}_{bn}/n$ have the same limit law, which complete the proof since by Lemma 3, $\tilde{T}_{bn}/n \Rightarrow X_{\kappa, c}$. We will use the notation NB, Po, and Geom to denote the negative binomial, Poisson, and geometric distributions with the indicated parameters.

For the multiset case, notice that

$$\begin{aligned} d_{TV}(T_{bn}, \tilde{T}_{bn}) &\leq d_{TV}((Z_{b+1}, \ldots, Z_n), (\tilde{Z}_{b+1}, \ldots, \tilde{Z}_n)) \\ &\leq \sum_{b+1}^{n} d_{TV}(Z_j, \tilde{Z}_j). \end{aligned}$$

To estimate each summand, we have

$$\begin{aligned} d_{TV}(Z_j, \tilde{Z}_j) &= d_{TV}(\mathrm{NB}(m_j, x^j), \mathrm{Po}(m_j x^j)) \\ &\leq m_j d_{TV}(\mathrm{Geom}(x^j), \mathrm{Po}(x^j)) \\ &\leq 2 m_j x^{2j}. \end{aligned} \tag{79}$$

The bound in (79) follows from the fact that $d_{TV}(\mathrm{Geom}(p), \mathrm{Be}(p)) = p^2$ and $d_{TV}(\mathrm{Be}(p), \mathrm{Po}(p)) = p(1 - e^{-p}) \leq p^2$, so that $d_{TV}(\mathrm{Geom}(p), \mathrm{Po}(p)) \leq d_{TV}(\mathrm{Geom}(p), \mathrm{Be}(p)) + d_{TV}(\mathrm{Be}(p), \mathrm{Po}(p)) \leq 2p^2$, a result we apply with $p = x^j$. Hence

$$d_{TV}(T_{bn}, \tilde{T}_{bn}) \leq 2 \sum_{j=b+1}^{n} (m_j x^j) x^j = O(y^{-b}/b).$$

Choosing $b \to \infty$, $b = o(n)$ completes the proof for multisets.

For the selection case, (79) may be replaced by

$$d_{TV}(Z_j, \tilde{Z}_j) \leq m_j d_{TV}(\mathrm{Be}(x^j/(1+x^j)), \mathrm{Po}(x^j)) \leq 2 m_j x^{2j}.$$

The last estimate following from the observation that $d_{TV}(\mathrm{Be}(p/(1+p)), \mathrm{Be}(p)) = p^2/(1+p)$, so that $d_{TV}(\mathrm{Be}(p/(1+p)), \mathrm{Po}(p)) \leqslant d_{TV}(\mathrm{Be}(p/(1+p)), \mathrm{Be}(p)) + d_{TV}(\mathrm{Be}(p), \mathrm{Po}(p)) \leqslant 2p^2$, which we apply with $p = x^j$. This completes the proof. ∎

The random variable X_κ of Subsection 4.3 is the special case $c = 0$ of $X_{\kappa, c}$. Further, for $c \neq 0$,

$$\mathbb{E} X_{\kappa, c} = \kappa \frac{1 - e^{-c}}{c}$$

and

$$\mathrm{Var}\, X_{\kappa, c} = \kappa \frac{1 - (1+c)\, e^{-c}}{c^2}.$$

The density g_c of $X_{\kappa, c}$ may be found from the density g of X_κ by observing that the log Laplace transforms, given by (66) and (77), are related by

$$\psi_c(s) = \frac{\psi(c+s)}{\psi(c)}$$

so that

$$g_c(z) = e^{-cz} g(z)/\psi(c), \qquad z \geqslant 0.$$

In particular, from (69),

$$g_c(z) = \frac{e^{-\gamma\kappa} e^{-cz} z^{\kappa - 1}}{\Gamma(\kappa)\, \psi(c)}, \qquad 0 \leqslant z \leqslant 1. \tag{80}$$

From (80) the value of c that maximizes the density $g_c(z)$ for fixed $z \in [0, 1]$ is the c that maximizes $-cz - \log \psi(c)$, just as suggested by large deviation theory. This c is the solution of the equation

$$cz = \kappa(1 - e^{-cz}).$$

Using $z = 1$, we see from the heuristic (55) that choosing c to be the solution of $c = \kappa(1 - e^{-c})$ asymptotically maximizes $\mathbb{P}(T_n = n)$; and from (75), this also makes $\mathbb{E} T_n \sim n$.

However, the heuristic in (59) and (60) suggests that better approximation should follow from choosing c so that $g_c'(1-) = 0$. From (80) and (70), we get

$$c = \frac{g'(1-)}{g(1)} = \kappa - 1. \tag{81}$$

For this choice of c we have $g_c'(1-)=0$, and

$$\frac{g_c''(1-)}{g_c(1)}=1-\kappa. \tag{82}$$

A second order approximation in the spirit of Section 4 then leads us to the following heuristic: for any fixed B, *in the case* $\kappa \neq 1$

$$d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) \asymp \frac{\sigma_B^2}{n^2}; \tag{83}$$

In the case $\kappa = 1$

$$d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) = o\left(\frac{1}{n^2}\right). \tag{84}$$

For the case $B=[b]\equiv\{1, 2, ..., b\}$, extensive numerical computations using the recurrence methods described in Section 9 support these conjectures for several of the combinatorial examples discussed earlier. In these cases, the bound in (83) is of order $(b/n)^2$. Finding the asymptotic form of this rate seems to be a much harder problem, since it seems to depend heavily on the value of κ.

5.3. *Further Examples*

The class of partitions of a set provides another example to show that the choice of x for good approximation is partly a matter of taste. In this example, $m_i \equiv 1$, so that

$$\mathbb{E}T_n = \sum_{i=1}^{n} \frac{im_i x^i}{i!} = x \sum_{i=0}^{n-1} \frac{x^i}{i!}.$$

One choice of x would be the exact solution x^* of the equation $\mathbb{E}T_n = n$, but this choice is poor since the definition of x^* is complicated. A second choice which is more usable is to take $x = x'$, the solution of the equation $xe^x = n$. This is based on the observation that $\mathbb{E}T_n \sim xe^x$, provided $x = o(n)$. The solution x' has the form (cf. de Bruijn [11, p. 26])

$$x' = \log n - \log\log n + \frac{\log\log n}{\log n} + \frac{1}{2}\left(\frac{\log\log n}{\log n}\right)^2 + O\left(\frac{\log\log n}{\log^2 n}\right).$$

For set partitions, with either x^* or x' in the role of x, we have $\sigma_n^2 \sim x^2 e^x \sim n\log n$, and we can check that $|n - \mathbb{E}T_n| = O(\sqrt{n\log n})$ is satisfied using $x = x'$. This corresponds to checking the condition in (37). Comparing the condition $\mathbb{E}T_n \sim n$ with the condition that $n - \mathbb{E}T_n = O(\sigma_n)$ required by (37), we see that in the logarithmic class the former is too restrictive while for set partitions it is not restrictive enough.

6. Refining the Combinatorial and Independent Processes

6.1. *Refining and Conditioning*

Although the refinements considered in this section are complicated in notation, the ingredients—including geometric and Bernoulli random variables and the counting formulas (89)–(92)—are simpler than their unrefined counterparts.

The dependent random variables $C_i \equiv C_i(n)$, which count the number of components of weight i in a randomly selected object of total weight n, may be refined as

$$C_i = \sum_{j=1}^{m_i} D_{ij}.$$

Here we suppose that the m_i possible structures of weight i have been labelled $1, 2, ..., m_i$, and $D_{ij} \equiv D_{ij}(n)$ counts the number of occurrences of the jth object of weight i. The independent random variable Z_i can also be refined, as

$$Z_i = \sum_{j=i}^{m_i} Y_{ij},$$

where the Y_{ij} are mutually independent, and for each i, $Y_{i1}, Y_{i2}, ..., Y_{im_i}$ are identically distributed. For assemblies, multisets, and selections, respectively, the distribution of Y_{ij} is Poisson $(x^i/i!)$ for $x > 0$, geometric (x^i) for $0 < x < 1$, or Bernoulli $(x^i/(1+x^i))$ for $x > 0$. If the choice of parameter x is taken as a function of n, then one can view Y_{ij} as $Y_{ij}(n)$. For assemblies, with $x > 0$,

$$\mathbb{P}(Y_{ij} = k) = \exp(-x^i/i!)\frac{(x^i/i!)^k}{k!}, \qquad k = 0, 1, ... \tag{85}$$

For multisets, with $0 < x < 1$,

$$\mathbb{P}(Y_{ij} = k) = (1 - x^i)\, x^{ik}, \qquad k = 0, 1, ..., \tag{86}$$

whereas for selections, with $x > 0$, we have

$$\mathbb{P}(Y_{ij} = k) = \frac{1}{1+x^i}\mathbf{1}(k=0) + \frac{x^i}{1+x^i}\mathbf{1}(k=1). \tag{87}$$

For the full refined processes corresponding to a random object of size n we denote the combinatorial process by

$$\mathbf{D}(n) \equiv (D_{ij}(n),\ 1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant m_i),$$

and the independent process by

$$\boldsymbol{Y}(n) \equiv (Y_{ij},\ 1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant m_i).$$

The weighted sum $T_n = \sum_1^n iZ_i$ is of course a weighted sum of the refined independent Y's, since

$$T_n = \sum_{i=1}^{n} \sum_{j=1}^{m_i} iY_{ij}.$$

THEOREM 4. *For assemblies, multisets, and selections, if* $\mathbb{P}(T_n = n) > 0$, *then the refined combinatorial process, for a uniformly chosen object of weight n, is equal in distribution to the independent process* $\boldsymbol{Y}(n)$, *conditioned on the event* $\{T_n = n\}$, *that is,*

$$\mathbf{D}(n) \stackrel{\mathrm{d}}{=} (\boldsymbol{Y}(n) \mid T_n = n).$$

Proof. Just as (6) is a special case of Theorem 1 with $t = n$, so is this. Imagine first the special case of (6) with each $m_i \equiv 1$, and then replicate m_i-fold the index i and its corresponding function g_i and normalizing constant c_i. The case $m_i = 0$ for some i is allowed. We have index set

$$I = \{\alpha = (i, j) : 1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant m_i\} \tag{88}$$

and weight function w given by $w(\alpha) = i$ for $\alpha = (i, j) \in I$.

The reader should be convinced by now, but for the record, here are the details. For $\mathbf{b} \equiv (b(\alpha))_{\alpha \in I} \in \mathbb{Z}_+^I$, write $\mathbf{b} \cdot \mathbf{w} \equiv \sum_I w(\alpha)\, b(\alpha)$. Consider the number $R(n, \mathbf{b})$ of objects of total weight $\mathbf{b} \cdot \mathbf{w} = n$, having $b_\alpha \equiv b(\alpha)$ components of type α, for $\alpha \in I$. For assemblies, the refined generalization of Cauchy's formula is that

$$\begin{aligned} R(n, \mathbf{b}) &\equiv |\{\text{assemblies on } [n] : \mathbf{D} = \mathbf{b}\}| \\ &= \mathbf{1}(\mathbf{b} \cdot \mathbf{w} = n)\, n! \prod_{\alpha \in I} \frac{1}{(i!)^{b(\alpha)}\, b(\alpha)!}, \end{aligned} \tag{89}$$

where $i = w(\alpha) =$ the first coordinate of α. For multisets,

$$\begin{aligned} R(n, \mathbf{b}) &\equiv |\{\text{multisets of weight } n : \mathbf{D} = \mathbf{b}\}| \\ &= \mathbf{1}(\mathbf{b} \cdot \mathbf{w} = n), \end{aligned} \tag{90}$$

while for selections,

$$\begin{aligned} R(n, \mathbf{b}) &\equiv |\{\text{selections of weight } n : \mathbf{D} = \mathbf{b}\}| \\ &= \mathbf{1}(\mathbf{b} \cdot \mathbf{w} = n) \prod_1^n \binom{1}{b_\alpha}. \end{aligned} \tag{91}$$

These examples have the form

$$R(n, \mathbf{b}) \equiv |\{\mathbf{D} = \mathbf{b}\}| = \mathbf{1}(\mathbf{b} \cdot \mathbf{w} = n) f(n) \prod_{\alpha \in I} g_\alpha(b_\alpha), \tag{92}$$

with $f(n) = n!$ for assemblies and $f(n) \equiv 1$ for multisets and selections. With $p(n)$ given by (4), we have the refined analysis of the total number of structures of weight n:

$$p(n) = \sum_{\mathbf{b} \in \mathbb{Z}_+^I} R(n, \mathbf{b}). \tag{93}$$

Picking an object of weight n uniformly defines the refined combinatorial distribution

$$\mathbb{P}(\mathbf{D}(n) = \mathbf{b}) \equiv \frac{R(n, \mathbf{b})}{p(n)} = \mathbf{1}(\mathbf{b} \cdot \mathbf{w} = n) \frac{f(n)}{p(n)} \prod_I g_\alpha(b_\alpha). \tag{94}$$

Observe that with multisets, $g_\alpha(k) = 1$ for $k \in \mathbb{Z}_+$; with selections $g_\alpha(k) = \binom{1}{k} = \mathbf{1}(k = 0 \text{ or } 1)$; and with assemblies, if $\alpha = (i, j)$, then $g_\alpha(k) = (1/i!)^k/k!$, for $k \in \mathbb{Z}_+$. Now apply Theorem 1 with $\mathbf{D}_I$ in the role of $\mathbf{C}_I$, $Y_{ij} \equiv Y_\alpha$ in the role of Z_α, and $t = n$. ∎

Remark. It would be reasonable to consider (89) through (92) as the basic counting formulas, with (1) through (3) as corollaries derived by summing, and to consider the Poisson, geometric, and Bernoulli distributions in (86) as the basic distributions, with the Poisson, negative binomial, and binomial distribution in (11) through (13) derived by convolution.

6.2. *Total Variation Distance*

Since the refined combinatorial process $\mathbf{D}(n)$ and the refined independent process $\mathbf{Y}(n)$ are related by conditioning on the value of a weighted sum of the Y's, Theorem 3 applies. For $K \subset I$, where I is given by (88), write $\mathbf{D}_K$ and $\mathbf{Y}_K$ for our refined processes, restricted to indices in K. Write

$$R'_K \equiv \sum_{\alpha \in K} w(\alpha)\, Y_\alpha, \qquad S'_K \equiv \sum_{\alpha \in I - K} w(\alpha)\, Y_\alpha,$$

so that $T \equiv T_n = R'_K + S'_K$.

THEOREM 5.

$$d_{TV}(\mathbf{D}_K, \mathbf{Y}_K) = d_{TV}((R'_K \mid T = n), R'_K). \tag{95}$$

Proof. This is a special case of Theorem 3, with the independent process $\mathbf{Y}(n) \equiv \mathbf{Y}_I$ playing the role of $\mathbf{Z}_I$ and $\mathbf{D}(n) \equiv \mathbf{D}_I$ playing the role

of $\mathbf{C}_I$. Theorem 4 is used to verify that the hypothesis (31) is satisfied, in the form $\mathbf{D}_I \stackrel{\mathrm{d}}{=} (\mathbf{Y}_I \mid T=n)$. ∎

For the special case where $B \subset \{1, ..., n\}$ and $K = \{\alpha = (i,j) \in I: i \in B\}$, denote the restriction of the refined combinatorial process, restricted to sizes in B, by $\mathbf{D}_{B^*} \equiv \mathbf{D}_K$, so that

$$\mathbf{D}_{B^*} \equiv (D_{ij}, i \in B, 1 \leqslant j \leqslant m_i),$$

and similarly define $\mathbf{Y}_{B^*}$. In this special case, $R'_K = R_B \equiv \sum_{i \in B} iZ_i$ is the weighted sum, restricted to B, for the unrefined process, so (95) reduces to

$$d_{TV}(\mathbf{D}_{B^*}, \mathbf{Y}_{B^*}) = d_{TV}((R_B \mid T=n), R_B). \tag{96}$$

Furthermore, by Theorem 3 applied to the unrefined case, with $I = \{1, ..., n\}$ and $w(i) = i$, we see that $d_{TV}((R_B \mid T=n), R_B)$ is equal to $d_{TV}(\mathbf{C}_B, \mathbf{Z}_B)$.

We have here a most striking example of the situation analyzed in Theorem 2, where taking functionals doesn't change a total variation distance. Namely, there is a functional $g: \mathbb{Z}_+^I \to \mathbb{Z}_+^n$, which "unrefines," and the functional $h: \mathbb{Z}_+^B \to \mathbb{Z}_+$ discussed in our second proof of Theorem 3, such that

$$g(\mathbf{D}_{B^*}) = \mathbf{C}_B, \quad g(\mathbf{Y}_{B^*}) = \mathbf{Z}_B, \quad h(\mathbf{C}_B) \stackrel{\mathrm{d}}{=} (R_B \mid T=n), \quad \text{and} \quad h(\mathbf{Z}_B) = R_B,$$

so that, a priori via (28),

$$d_{TV}(\mathbf{D}_{B^*}, \mathbf{Y}_{B^*}) \geqslant d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) \geqslant d_{TV}((R_B \mid T=n), R_B). \tag{97}$$

Perhaps the result in (96), which shows that equality holds throughout (97), is surprising.

7. Conditioning on Events of Moderate Probability

We consider random combinatorial structures conditioned on some event. Given that we have a good approximation by another process, this other process, conditioned on the same event, may yield a good approximation to the conditioned combinatorial structure. The conditioning event must have moderate probability, large relative to the original approximation error. In contrast, if the conditioning event is very unlikely then the approximating process must also be changed, as discussed in Section 8 on large deviations.

7.1. *Bounds for Conditioned Structures*

In this subsection, we consider bounds on total variation distance that are inherited from an existing approximation, after additional conditioning is applied.

THEOREM 6. *Let $A \subseteq B \subseteq [n]$, and let $h: \mathbb{Z}_+^B \to \{0, 1\}$ be measurable with respect to coordinates in A. Let $\mathbf{Z}_B$, and $\mathbf{C}_B$ be arbitrary processes with values in $\mathbb{Z}_+^B$, and let $\mathbf{Z}_A$ and $\mathbf{C}_A$ denote their respective restrictions to coordinates in A. Let*

$$\mathbf{C}_B^* \stackrel{\mathrm{d}}{=} (\mathbf{C}_B \mid h(\mathbf{C}_B) = 1),$$

and

$$\mathbf{Z}_B^* \stackrel{\mathrm{d}}{=} (\mathbf{Z}_B \mid h(\mathbf{Z}_B) = 1).$$

Write $p = \mathbb{P}(h(\mathbf{Z}_B) = 1)$, $q = \mathbb{P}(h(\mathbf{C}_B) = 1)$, $d_B = d_{TV}(\mathbf{C}_B, \mathbf{Z}_B)$, $d_A = d_{TV}(\mathbf{C}_A, \mathbf{Z}_A)$, and assume that $p > 0$ and $q > 0$. Then

$$d_{TV}(\mathbf{C}_B^*, \mathbf{Z}_B^*) \leqslant \frac{1}{2}\left|1 - \frac{q}{p}\right| + \frac{d_B}{p} \tag{98}$$

$$\leqslant \frac{1}{p}\left(\frac{d_A}{2} + d_B\right) \tag{99}$$

$$\leqslant \frac{3}{2}\frac{d_B}{p}. \tag{100}$$

Proof. The second to last inequality follows from the relation $|p - q| \leqslant d_A$, and is useful when this is the extent of our ability to estimate q. The last inequality follows simply from the fact that $d_A \leqslant d_B$. To establish the first inequality, we have

$$\begin{aligned}
d_{TV}(\mathbf{C}_B^*, \mathbf{Z}_B^*) &= \frac{1}{2}\sum_{\mathbf{a} \in \mathbb{Z}_+^B} |\mathbb{P}(\mathbf{C}_B^* = \mathbf{a}) - \mathbb{P}(\mathbf{Z}_B^* = \mathbf{a})| \\
&= \frac{1}{2}\sum_{\mathbf{a}:\, h(\mathbf{a}) = 1} \left|\frac{\mathbb{P}(\mathbf{C}_B = \mathbf{a})}{q} - \frac{\mathbb{P}(\mathbf{Z}_B = \mathbf{a})}{p}\right| \\
&= \frac{1}{2}\sum_{\mathbf{a}:\, h(\mathbf{a}) = 1} \left|\mathbb{P}(\mathbf{C}_B = \mathbf{a})\left(\frac{1}{q} - \frac{1}{p}\right) + \frac{\mathbb{P}(\mathbf{C}_B = \mathbf{a}) - \mathbb{P}(\mathbf{Z}_B = \mathbf{a})}{p}\right| \\
&\leqslant \frac{1}{2}\left|\frac{1}{q} - \frac{1}{p}\right| \sum_{\mathbf{a}:\, h(\mathbf{a}) = 1} \mathbb{P}(\mathbf{C}_B = \mathbf{a}) \\
&\quad + \frac{1}{2p}\sum_{\mathbf{a}:\, h(\mathbf{a}) = 1} |\mathbb{P}(\mathbf{C}_B = \mathbf{a}) - \mathbb{P}(\mathbf{Z}_B = \mathbf{a})|
\end{aligned}$$

$$= \frac{1}{2}\left|\frac{1}{q} - \frac{1}{p}\right| q + \frac{1}{2p} \sum_{\mathbf{a}:\, h(\mathbf{a})=1} |\mathbb{P}(\mathbf{C}_B = \mathbf{a}) - \mathbb{P}(\mathbf{Z}_B = \mathbf{a})|$$

$$\leqslant \frac{1}{2}\left|\frac{1}{q} - \frac{1}{p}\right| q + \frac{1}{2p} \sum_{\mathbf{a}} |\mathbb{P}(\mathbf{C}_B = \mathbf{a}) - \mathbb{P}(\mathbf{Z}_B = \mathbf{a})|$$

$$= \frac{1}{2}\left|1 - \frac{q}{p}\right| + \frac{d_B}{p}. \quad \blacksquare$$

Remark. While the theorem above uses the notation C_B and Z_B to suggest applications where one process is obtained from an independent process by conditioning, no such structure is required. An arbitrary discrete space S, together with an arbitrary functional $h: S \to \{0, 1\}$, may be encoded in terms of $S = \mathbb{Z}_+^2$, with $A = \{1\}$ and $B = \{1, 2\}$, so that h depends only on the first coordinate. Thus Theorem 6 applies to discrete random objects in general.

7.2. *Examples*

7.2.1. *Random Permutations*

In this case, the Z_i are independent Poisson distributed random variables, with $\lambda_i \equiv \mathbb{E} Z_i = 1/i$. In Arratia and Tavaré [2] it is proved that for $1 \leqslant b \leqslant n$, the total variation distance $d_b(n)$ between $(C_1(n), ..., C_b(n))$ and $(Z_1, ..., Z_b)$ satisfies $d_b(n) \leqslant F(n/b)$ where

$$F(x) \equiv \sqrt{2\pi m}\, \frac{2^{m-1}}{(m-1)!} + \frac{1}{m!} + 3\left(\frac{x}{e}\right)^{-x}, \qquad \text{with} \quad m \equiv \lfloor x \rfloor$$

$$\sim \left(\frac{2e}{\lfloor x - 1 \rfloor}\right)^{\lfloor x-1 \rfloor} \tag{101}$$

as $x \to \infty$. To get an approximation result for derangements, we use the functional h having $h((a_1, ..., a_b)) = \mathbf{1}(a_1 = 0)$, with $A = \{1\}$ and $B = \{1, 2, ..., b\}$. This makes $\mathbf{C}_B^*$ the process counting cycles of size at most b in a randomly chosen derangement, and $\mathbf{Z}_B^* = (Z_1^*, Z_2^*, ..., Z_b^*) \stackrel{\mathrm{d}}{=} (0, Z_2, ..., Z_b)$. The total variation distance $d_b^*(n)$ between $\mathbf{C}_B^*$ and $\mathbf{Z}_B^*$ satisfies $d_b^*(n) \leqslant (3/2)\, eF(n/b)$, simply by using (100).

Changing random permutations to random derangements is a special case of conditioning on some fixed conditions of the form $C_i(n) = c_i$, $i \in A$, for given constants c_i, with $A \subseteq B \subseteq \{1, 2, ..., b\}$. In this situation, all the Z_i^* are mutually independent, $Z_i^* \equiv c_i$ for $i \in A$, and for $i \notin A$, $Z_i^* =^{\mathrm{d}} Z_i$ is Poisson with mean $1/i$. Here, Theorem 6 yields the bound $d_b^*(n) \leqslant 3/(2p)\, F(n/b)$, where $p = \mathbb{P}(Z_i = c_i\, \forall i \in A)$. Theorem 3 in Arratia and Tavaré

[2] gives a different upper bound, namely $d_b^*(n) \leqslant F((n-s)/b) + 2be((n-s)/(be))^{-(n-s)/b}$, where $s = \sum_{i \in A} ic_i$. Either of these two upper bounds may be smaller, depending on the situation given by A, b, and the c_i.

For a more complicated conditioning in which the Z_i^* are not mutually independent, consider random permutations on n objects conditional on having at least one cycle of length two or three. Here, Z_2^* and Z_3^* are dependent, although the *pair* (Z_2^*, Z_3^*) and the variables $Z_1^*, Z_4^*, Z_5^*, \ldots$ are mutually independent. With $A = \{2, 3\} \subseteq B = \{1, 2, \ldots, b\}$, we have $p = \mathbb{P}(Z_2 + Z_3 > 0) = 1 - e^{-5/6}$ and $d_b^*(n) \leqslant 3/(2p)\, F(n/b)$. Thus, for example, with $b = 3$, the probability that a random permutation of n objects is a derangement, given that $C_2(n) + C_3(n) > 0$, can be approximated by $\mathbb{P}(Z_1^* = 0) = 1/e$, with error at most $3/(2p)\, F(n/3)$. Similarly, the probability that a random permutation of n objects has a cycle of length 2, given that $C_2(n) + C_3(n) > 0$, can be approximated by $\mathbb{P}(Z_2^* > 0) = \mathbb{P}(Z_2 > 0 \mid Z_2 + Z_3 > 0) = (1 - e^{-1/2})/(1 - e^{-5/6})$, with error again at most $3/(2p)\, F(n/3)$.

The next example shows how to approximate easily the small component counts for 2-regular graphs by exploiting a decoupling result for the Ewens sampling formula with parameter $\kappa = 1/2$.

7.2.2. *2-Regular Graphs*

The combinatorial structure known as "2-regular graphs" is the assembly in which components are undirected cycles on three or more points, so that

$$m_i = \tfrac{1}{2}(i-1)!\, \mathbf{1}\{i \geqslant 3\}. \tag{102}$$

Let $C_i^*(n)$ be the number of components of size i in a random 2-regular graph on n points. A process that corresponds to this, with the condition $\mathbf{1}\{i \geqslant 3\}$ removed, is the Ewens sampling formula with parameter $\kappa = 1/2$ described in Subsection 5.1. Observe that

$$\mathbf{C}^*(n) \stackrel{\mathrm{d}}{=} (\mathbf{C}(n) \mid C_1(n) = C_2(n) = 0).$$

The bound

$$d_{TV}((C_1, \ldots, C_b), (Z_1, \ldots, Z_b)) \leqslant \frac{2b}{n}$$

is known from results of Arratia, Barbour, and Tavaré [4]. We are interested in how this translates into a bound on

$$d_b^* \equiv d_{TV}((C_3^*, \ldots, C_b^*), (Z_3, \ldots, Z_b)).$$

With $A=\{1,2\}$, $B=\{1,2,\ldots,b\}$, $d_A \leqslant 4/n$, $d_B \leqslant 2b/n$, $p=\mathbb{P}(Z_1=Z_2=0)=e^{-3/4}$, the inequality in (99) guarantees that

$$\begin{aligned} d_b^* &\leqslant \frac{1}{p}\left(\frac{d_A}{2}+d_B\right) \\ &\leqslant e^{3/4}\left(\frac{2}{n}+\frac{2b}{n}\right) \\ &= e^{3/4}\frac{2(b+1)}{n}. \end{aligned}$$

For an example that shows the conditioning event can have probability tending to zero, consider 2-regular graphs conditioned on having no cycles of size less than or equal to $t \equiv t(n) \geqslant 2$. The previous example is the special case $t=2$. For $b>t$, we have

$$(C_{t+1}^*,\ldots,C_b^*) \stackrel{\mathrm{d}}{=} (C_{t+1},\ldots,C_b \mid C_1=\cdots=C_t=0).$$

Now $d_A \leqslant 2t/n$, $d_B \leqslant 2b/n$, and

$$p=\mathbb{P}(Z_1=\cdots=Z_t=0)=\exp\left(-\frac{1}{2}(1+\cdots+1/t)\right)\geqslant\frac{1}{\sqrt{et}},$$

so (99) establishes that

$$\begin{aligned} d_b^* &\leqslant \frac{1}{p}\left(\frac{d_A}{2}+d_B\right) \\ &\leqslant \sqrt{et}\left(\frac{t}{n}+\frac{2b}{n}\right). \end{aligned}$$

This provides a useful bound provided that $\sqrt{t}\,b/n$ is small. Note that both t and b may grow with n, as long as $t \leqslant b$. For example, conditional on no cycles of length less than or equal to $t=\lfloor n^{2/3-\varepsilon}\rfloor$ this approximation successfully describes the distribution of the k smallest cycles, for fixed k as $n\to\infty$, by using $b=n^{2/3}$. See Arratia and Tavaré [3, Theorem 7] for related details.

8. Large Deviation Theory

8.1. *Biasing the Combinatorial and Independent Processes*

A guiding principle of large deviation theory is that unlikely events of the form $\{U \geqslant u\}$ or $\{U \leqslant u\}$ or $\{U=u\}$, where the target u is far from $\mathbb{E}U$,

can be studied by changing the measure $\mathbb{P}$ to another measure $\mathbb{P}_\theta$ defined by

$$\frac{d\mathbb{P}_\theta}{d\mathbb{P}} = \frac{\theta^U}{\mathbb{E}\theta^U}. \tag{103}$$

Observe that for $\theta = 1$, the new measure $\mathbb{P}_\theta$ coincides with the original measure $\mathbb{P}$, regardless of the choice of U. The parameter θ is chosen so that the average value of U under the new measure is u, i.e., $\mathbb{E}_\theta U = u$. In the literature on large deviations and statistical mechanics (cf. Ellis [16]), the notation is usually $\theta \equiv e^\beta$, and our normalizing factor $\mathbb{E}\theta^U$ is expressed as the Laplace transform of the $\mathbb{P}$-distribution of U, parameterized by β.

For the case of a combinatorial process $\mathbf{C}(n) = (C_1(n), ..., C_n(n))$, with the total number of components

$$K \equiv K_n \equiv C_1(n) + \cdots + C_n(n)$$

in the role of U, this says to change from the measure $\mathbb{P}$, which makes all possible structures equally likely, to the measure $\mathbb{P}_\theta$, which selects a structure with bias proportional to $\theta^{\#\text{components}}$. The Ewens sampling formula discussed in Subsection 5.1 is exactly this in the case of random permutations, with κ playing the role of θ. This may easily be verified by comparing (74) to Cauchy's formula, the special case $\kappa = 1$ of (74), in which the equality of normalizing constants, with $\mathbb{E}\kappa^{K_n} = \kappa_{(n)}$, expresses a well known identity for Stirling numbers of the first kind.

Theorem 1 showed that many a combinatorial process is equal in distribution to a process of independent random variables, conditioned on the value of a weighted sum. The next theorem asserts that this form is preserved by the change of measure from large deviation theory, provided that U is also a weighted sum.

As in the discussion before Theorem 1, the weight function $\mathbf{u}$, just like the weight function $\mathbf{w}$, can take values in $\mathbb{R}$ or $\mathbb{R}^d$. In case the weights $\mathbf{u}$, and hence the random variable U, takes values in $\mathbb{R}^d$ with $d > 1$, we take $\theta > 0$ to mean that $\theta = (\theta_1, ..., \theta_d) \in (0, \infty)^d$, and with $U = (U_1, ..., U_d)$, θ^U represents the product $\theta_1^{U_1} \cdots \theta_d^{U_d}$.

THEOREM 7. *Let I be a finite set, and for $\alpha \in I$, let C_α and Z_α be $\mathbb{Z}_+$-valued random variables. Let $\mathbf{w} = (w(\alpha))_{\alpha \in I}$ and $\mathbf{u} = (u(\alpha))_{\alpha \in I}$ be deterministic weight functions on I, with real values for $\mathbf{u}$, let $T = \mathbf{w} \cdot \mathbf{Z}_I \equiv \sum_{\alpha \in I} w(\alpha) Z_\alpha$, and let $U = \mathbf{u} \cdot \mathbf{C}_I$. Let $\mathbb{P}$ be a probability measure and t be a constant such that, under $\mathbb{P}$ the Z_α are mutually independent, $\mathbb{P}(T = t) > 0$, and $\mathbf{C}_I \overset{\mathrm{d}}{=} (\mathbf{Z}_I \mid T = t)$. Let $\theta > 0$ be any constant such that the random variable $Y \equiv \theta^{\mathbf{u} \cdot \mathbf{Z}_I}$ has $\mathbb{E}Y < \infty$. Let $\mathbb{P}_\theta$, restricted to the sigma-field*

generated by $\mathbf{C}_I$, *be given by* (103). *Let* $\mathbb{P}_\theta$, *restricted to the sigma-field generated by* $\boldsymbol{Z}_I$, *be given by*

$$\frac{d\mathbb{P}_\theta}{d\mathbb{P}} = \frac{Y}{\mathbb{E}Y},$$

so that the Z_α *are mutually independent under* $\mathbb{P}_\theta$ *with*

$$\mathbb{P}_\theta(Z_\alpha = k) = \frac{\theta^{u(\alpha)k}}{\mathbb{E}\theta^{u(\alpha)Z_\alpha}}\mathbb{P}(Z_\alpha = k), \qquad k \geqslant 0. \tag{104}$$

Then under $\mathbb{P}_\theta$, $\mathbf{C}_I \stackrel{\mathrm{d}}{=} (\mathbf{Z}_I \mid T = t)$, *that is,*

$$\mathbb{P}_\theta(\mathbf{C}_I = \mathbf{a}) = \mathbb{P}_\theta(\mathbf{Z}_I = \mathbf{a} \mid T = t), \tag{105}$$

for $\mathbf{a} \in \mathbb{Z}_+^I$.

Proof. For $\mathbf{a} \in \mathbb{Z}_+^I$,

$$\begin{aligned}\mathbb{P}_\theta(\mathbf{C}_I = \mathbf{a}) &= (\mathbb{E}\theta^U)^{-1}\,\theta^{\mathbf{u}\cdot\mathbf{a}}\mathbb{P}(\mathbf{C}_I = \mathbf{a}) \\ &= (\mathbb{E}\theta^U)^{-1}\,\theta^{\mathbf{u}\cdot\mathbf{a}}\mathbb{P}(\mathbf{Z}_I = \mathbf{a} \mid T = t) \\ &= (\mathbb{E}\theta^U)^{-1}\,\mathbb{P}(T = t)^{-1}\,\theta^{\mathbf{u}\cdot\mathbf{a}}\mathbf{1}(\mathbf{w}\cdot\mathbf{a} = t)\,\mathbb{P}(\mathbf{Z}_I = \mathbf{a}). \end{aligned} \tag{106}$$

Now

$$\mathbb{P}_\theta(\mathbf{Z}_I = \mathbf{a}) = (\mathbb{E}\theta^{\mathbf{u}\cdot\boldsymbol{Z}_I})^{-1}\,\theta^{\mathbf{u}\cdot\mathbf{a}}\mathbb{P}(\boldsymbol{Z}_I = \mathbf{a})$$

so that

$$\mathbb{P}_\theta(\mathbf{Z}_I = \mathbf{a} \mid T = t) = (\mathbb{E}\theta^{\mathbf{u}\cdot\mathbf{Z}_I})^{-1}\,\mathbb{P}_\theta(T = t)^{-1}\,\theta^{\mathbf{u}\cdot\mathbf{a}}\mathbf{1}(\mathbf{w}\cdot\mathbf{a} = t)\,\mathbb{P}(\boldsymbol{Z}_I = \mathbf{a}). \tag{107}$$

Comparing (106) and (107), we see both expressions are probability densities on $\mathbb{Z}_+^I$ which are proportional to the same function of $\mathbf{a}$, and hence they are equal. From this it also follows that the normalizing constants are equal, which is written below with the combinatorial generating function on the left, and the three factors determined by independent random variables on the right:

$$\mathbb{E}\theta^U = \mathbb{E}\theta^{\mathbf{u}\cdot\boldsymbol{Z}_I}\frac{\mathbb{P}_\theta(T = t)}{\mathbb{P}(T = t)}. \quad \blacksquare \tag{108}$$

For the case $U = K_n$, the total number of components, the $\mathbb{P}_\theta$ measure corresponds to the following generalization of (11) through (13). For assemblies, multisets, or selections, chosen with probability proportional to

$\theta^{\#\text{ components}}$, $\mathbf{C}(n) \stackrel{\mathrm{d}}{=} ((Z_1, \ldots, Z_n) \mid Z_1 + 2Z_2 + \cdots + nZ_n = n)$ where the Z_i are mutually independent. With $\theta, x > 0$, for assemblies we have

$$Z_i \text{ is Poisson } \left(\frac{m_i \theta x^i}{i!}\right), \tag{109}$$

whereas for multisets we require $x \leqslant 1$, $\theta x < 1$ and then

$$Z_i \text{ is negative binomial } (m_i, \theta x^i).$$

Finally, for selections

$$Z_i \text{ is binomial } \left(m_i, \frac{\theta x^i}{1 + \theta x^i}\right).$$

In the general case, where $U = \mathbf{u} \cdot \mathbf{C}(n)$ is a weighted sum of component counts, so that the selection bias is $\theta^{\mathbf{u} \cdot \mathbf{C}(n)}$, each factor θ in (109) above is replaced by $\theta^{u(i)}$. Furthermore, we observe that Theorems 3, 4, and 5 apply to $\mathbb{P}_\theta$ in place of $\mathbb{P}$. For the refinements in Section 6, for assemblies, multisets, and selections respectively, the distribution of Y_{ij} is Poisson $(\theta^{u(i)} x^i / i!)$, Geometric $(\theta^{u(i)} x^i)$, or Bernoulli $(\theta^{u(i)} x^i / (1 + \theta^{u(i)} x^i))$.

An example where such a bias is well known is the random graph model $\mathscr{G}_{n,p}$; see Bollobás [10]. This corresponds to picking a labelled graph on n vertices, where each of the potential edges is independently taken with probability p; the unbiased case with all $2^{\binom{n}{2}}$ graphs equally likely is given by $p = 1/2$. We need something like the refined setup of Section 6 to be able to keep track of components in terms of the number of edges in addition to the number of vertices. Using the full refinement of Section 6, D_{ij} counts the number of components on i vertices having the jth possible structure, for $j = 1, \ldots, m_i$, in some fixed enumeration of these. The weight function should be $u(i, j) = \#$edges in the jth possible structure on i vertices. With $\theta = p/(1-p)$, the $\mathbb{P}_\theta$ law of $\mathbf{D}(n)$ is a description of $\mathscr{G}_{n,p}$. A more natural refinement for this example, intermediate between $\mathbf{C}$ and $\mathbf{D}$, would be the process $\mathbf{A}$ with $A_{ik} = \sum_{j: u(i,j) = k} D_{ij}$, the number of components with i vertices and k edges, for $k = i - 1, \ldots, \binom{i}{2}$. As in (96) and (97), the total variation distances are insensitive to the amount of refining. Presumably there are interesting results about random graphs that could easily be deduced from estimates of the total variation distance in Theorem 5.

One form of the general large deviation heuristic is that for a process $\mathbf{C}$, conditioned on the event $\{U \geqslant u\}$ where U is a functional of the process and $u > \mathbb{E}U$, the $\mathbb{P}$-law of the conditioned process is nicely approximated by the $\mathbb{P}_\theta$-law of $\mathbf{C}$, where θ is chosen so that $\mathbb{E}_\theta U = u$. We are interested in the special case where the functional U is a weighted sum, and the distribution of $\mathbf{C}$ under $\mathbb{P}$ is that of an independent process $\mathbf{Z}$ conditioned on the value of another weighted sum T. In this case, Theorem 3 yields

a direct quantitative handle on the quality of approximation by the $\mathbb{P}_\theta$-distribution of the independent process, provided we condition on the event $\{U=u\}$ instead of the event $\{U \geqslant u\}$.

THEOREM 8. *Assume the hypotheses and notation of Theorems* 3 *and* 7 *combined. For* $B \subset I$ *write* $U_B \equiv \sum_{\alpha \in B} u(\alpha) Z_\alpha$, *so that* $U_I \equiv \mathbf{u} \cdot \mathbf{Z}_I$. *Write* $\mathscr{L}_\theta$ *for distributions governed by* $\mathbb{P}_\theta$, *so that the conclusion of Theorem* 7 *may be written*

$$\mathscr{L}_\theta(\mathbf{C}_I) = \mathscr{L}_\theta(\mathbf{Z}_I \mid T=t),$$

and Theorem 3 *states that for* $B \subset I$

$$d_{TV}(\mathscr{L}_\theta(\mathbf{C}_B), \mathscr{L}_\theta(\mathbf{Z}_B)) = d_{TV}(\mathscr{L}_\theta(R_B \mid T=t), \mathscr{L}_\theta(R_B)). \tag{110}$$

Assume that u *is such that* $\mathbb{P}(U=u)>0$. *Then under the further conditioning on* $U=u$,

$$\begin{aligned} &d_{TV}(\mathscr{L}_1(\mathbf{C}_B \mid U=u), \mathscr{L}_\theta(\mathbf{Z}_B)) \\ &\quad = d_{TV}(\mathscr{L}_\theta((U_B, R_B) \mid U_I=u, T=t), \mathscr{L}_\theta((U_B, R_B))). \end{aligned} \tag{111}$$

Proof. Observe first that

$$\mathscr{L}_1(\mathbf{C}_I \mid U=u) = \mathscr{L}_\theta(\mathbf{C}_I \mid U=u), \tag{112}$$

so that it suffices to prove (111) with the subscript θ appearing on all four distributions, i.e.,

$$\begin{aligned} &d_{TV}(\mathscr{L}_\theta(\mathbf{C}_B \mid U=u), \mathscr{L}_\theta(\mathbf{Z}_B)) \\ &\quad = d_{TV}(\mathscr{L}_\theta((U_B, R_B) \mid U_I=u, T=t), \mathscr{L}_\theta((U_B, R_B))). \end{aligned} \tag{113}$$

Observe next that this is a special case of Theorem 3, but with two-component weights $w^*(\alpha) \equiv (u(\alpha), w(\alpha))$ in the role of $w(\alpha)$. For example, in the usual combinatorial case, with $I=[n]$ and $w(i)=i$, and further specialized to $U=K_n=$ the total number of components, so that $u(i)=1$, we have that $\mathbf{w}^*$ takes values in $\mathbb{R}^2$, with $w^*(i)=(1, i)$. ∎

Discussion. The proof of the previous theorem helps make it clear that the free parameter x, such that $\mathscr{L}((Z_1, \ldots, Z_n) \mid T_n=n)$ does not vary with x, is analogous to the parameter θ, such that relation (112) holds. With this perspective, the discussion of an appropriate choice of x in Section 4 and Subsection 5.2 is simply giving details in some special cases of the general large deviation heuristic. Note that T_n is a sufficient statistic for x, while U is a sufficient statistic for θ.

There are three distributions involved in the discussion above: the first is $\mathscr{L}(\mathbf{C}_I \mid U=u)$, corresponding to a combinatorial distribution conditioned on the value of a weighted sum U, the second is $\mathscr{L}_\theta(\mathbf{C}_I)$, which is a biased version of the combinatorial distribution, and the third is $\mathscr{L}_\theta(\mathbf{Z}_I)$, which governs an independent process. Theorem 3, used with Theorem 7,

compares the second and third of these; Theorem 8 above compares the first and third of these; and the following theorem completes the triangle, by comparing the first and second distributions.

THEOREM 9. *In the setup of Theorem* 8, *for* $B \subset I$,

$$d_{TV}(\mathcal{L}_1(\mathbf{C}_B \mid U=u), \mathcal{L}_\theta(\mathbf{C}_B)) = d_{TV}(\mathcal{L}_\theta((U_B, R_B) \mid U_I=u, T=t), \mathcal{L}_\theta((U_B, R_B) \mid T=t)). \quad (114)$$

Proof. By Theorem 7, together with (112), the left side of (114) is equal to $d_{TV}(\mathcal{L}_\theta(\mathbf{Z}_B \mid U_I=u, T=t), \mathcal{L}_\theta(\mathbf{Z}_B \mid T=t))$. We modify the second proof of Theorem 3 as follows: replace $\mathbb{P}$ by $\mathbb{P}_\theta$, use two-component weights, replace the original conditioning $T=t$ by $U_I=u$, and then further condition on $\{T=t\}$. Explicitly, the functional h on $\mathbb{Z}_+^B$ defined by $h(\mathbf{a}) = \sum_{\alpha \in B} a(\alpha)(u(\alpha), w(\alpha))$ is a sufficient statistic, and the sign of $\mathbb{P}_\theta(\mathbf{Z}_B = \mathbf{a} \mid U_I=u, T=t) - \mathbb{P}_\theta(\mathbf{Z}_B = \mathbf{a} \mid T=t)$ is equal to the sign of $\mathbb{P}_\theta((U_B, R_B) = h(\mathbf{a}) \mid U_I=u, T=t) - \mathbb{P}_\theta((U_B, R_B) = h(\mathbf{a}) \mid T=t)$, i.e., the sign depends on $\mathbf{a}$ only through the value of $h(\mathbf{a})$. ∎

Observe that Theorem 8 contains Theorem 3 as a special case, by taking weights $u(\alpha) \equiv 0$ and target $u=0$, so that $\mathbb{P}_\theta = \mathbb{P}$ and the extra conditioning event $\{U=u\}$ has probability one.

8.2. *Heuristics for Good Approximation of Conditioned Combinatorial Structures*

The following applies to weighted sums U in general, but to be concrete we present the special case $U=K_n$. Let $K \equiv K_n$ be the total number of components of some assembly, multiset, or selection of total weight n, and let some deterministic target $k \equiv k(n)$ be given. The goal is to describe an independent process to approximate $\mathbf{C}(n)$, conditioned on the event $\{K \geqslant k\}$, in case k is large compared to $\mathbb{E}K$; or conditioned on the event $\{K \leqslant k\}$, in the opposite case; or more simply, conditioned on the event $\{K=k\}$. We accomplish this by picking the free parameters θ and x in (109) so that simultaneously $\mathbb{E}(Z_1 + \cdots + Z_n)$ is close to k and $\mathbb{E}T_n$ is close to n.

For example, to study random permutations on n objects, conditional on having at least $5 \log n$ cycles, or conditional on having exactly $\lfloor 5 \log n \rfloor$ cycles, or conditional on having at most $0.3 \log n$ cycles, we propose using $x=1$, and $\theta=5$ or 0.3. The independent process with this choice of parameter should be a good approximation for both the conditioned random permutations and for the Ewens sampling formula. As a corollary, the Ewens sampling formula should be a good approximation for the conditioned permutations; see Arratia, Barbour, and Tovaré [6].

For assemblies, multisets, and selections in the logarithmic class discussed in Subsection 4.3, in which $\mathbb{E}Z_i \sim \kappa/i$, biasing by θ^K yields $\mathbb{E}_\theta Z_i \sim \kappa\theta/i$, so

that the Ewens sampling formula with parameter $\kappa\theta$ is a useful approximation for the biased measures. In particular, the heuristics (71) and (72) should apply in the following form: for fixed $B \subseteq [n]$, *in the case* $\kappa\theta \neq 1$,

$$d_{TV}(\mathscr{L}_\theta(\mathbf{C}_B), \mathscr{L}_\theta(\mathbf{Z}_B)) \sim \frac{1}{2}|\kappa\theta - 1| \frac{\mathbb{E}_\theta |R_B - \mathbb{E}_\theta R_B|}{n}; \tag{115}$$

in the case $\kappa\theta = 1$,

$$d_{TV}(\mathscr{L}_\theta(\mathbf{C}_B), \mathscr{L}_\theta(\mathbf{Z}_B)) = o\left(\frac{1}{n}\right). \tag{116}$$

For random permutations, for which $\kappa = 1$, with $B = \{1, 2, ..., b\}$ the bound

$$d_{TV}(\mathscr{L}_\theta(\mathbf{C}_B), \mathscr{L}_\theta(\mathbf{Z}_B)) \leqslant c(\theta)\frac{b}{n}$$

was established via a particular coupling in Arratia, Barbour, and Tavaré [4], and the asymptotic relation (115) has been established by Arratia, Stark, and Tavaré [7].

To show how the parameters x and θ may interact, we consider random permutations with $k(n)$ further away from $\log n$. Assume that $k(n)$ is given such that as $n \to \infty$,

$$k/\log n \to \infty, \qquad k/n \to 0.$$

Then we would take

$$\theta \equiv \theta(n) = \frac{k}{\log(n/k)}, \qquad x \equiv x(n) = e^{-\theta/n}. \tag{117}$$

Observe that $\theta/n \to 0$, so that $x \to 1$ and $1 - x \sim \theta/n$, and $\theta \to \infty$, so that $x^n = \exp(-\theta) \to 0$. Hence

$$\mathbb{E}T_n = \theta \sum_1^n x^i \sim \theta \sum_0^\infty x^i = \theta \frac{1}{1-x} \sim n$$

and

$$\mathbb{E}K_n = \theta \sum_1^n \frac{x^i}{i} \sim -\theta \log(1-x) \sim \theta \log\left(\frac{n}{\theta}\right) \sim k.$$

With this choice of parameters θ and x the independent Poisson process $(Z_1, Z_2, ...)$ should be a good approximation for random permutations, conditioned either on having exactly k cycles, or on having at least k cycles.

9. The Generating Function Connection and Moments

In this section, we relate the probabilistic technique to the more conventional one based on generating functions; Wilf [61]. One reason for

this is to provide a simple method, based on an idea of Shepp and Lloyd [52], for calculating moments of component counts for combinatorial structures. A second reason is to provide a framework within which detailed estimates and bounds for total variation distances can be obtained by using the results of Theorem 3 and 8, together with analytic techniques such as Darboux's method or the transfer methods of Flajolet and Odlyzko [20, 21].

Throughout, we let $p(n, k)$ be the number of objects of weight n having k components, so that $p(n) = \sum_{k=1}^{n} p(n, k)$ is the number of objects of weight n. Finally, recall that m_i is the number of available structures for a component of size i.

9.1. *Assemblies*

We form the exponential generating functions

$$\hat{P}(s, \theta) \equiv 1 + \sum_{n=1}^{\infty} \left(\sum_{k=1}^{n} p(n, k)\, \theta^k \right) \frac{s^n}{n!}, \tag{118}$$

$$\hat{P}(s) \equiv 1 + \sum_{n=1}^{\infty} p(n) \frac{s^n}{n!} = \hat{P}(s, 1), \tag{119}$$

and

$$\hat{M}(s) \equiv \sum_{n=1}^{\infty} m_n \frac{s^n}{n!}. \tag{120}$$

For assemblies, (1) gives

$$p(n, k) = \sum_{\mathbf{a}} N(n, \mathbf{a}) = \sum_{\mathbf{a}} n! \prod_{j=1}^{n} \left(\frac{m_j}{j!} \right)^{a_j} \frac{1}{a_j!},$$

where $\sum_{\mathbf{a}}$ is over $\{\mathbf{a} \in \mathbb{Z}_+^n : \sum i a_i = n, \sum a_i = k\}$. It follows that

$$\begin{aligned} \hat{P}(s, \theta) &= 1 + \sum_{n=1}^{\infty} \sum_{k=1}^{n} \sum_{\mathbf{a}} \prod_{j=1}^{n} \left(\frac{\theta m_j s^j}{j!} \right)^{a_j} \frac{1}{a_j!} \\ &= \prod_{j=1}^{\infty} \exp\left(\frac{\theta m_j s^j}{j!} \right) \\ &= \exp(\theta \hat{M}(s)). \end{aligned} \tag{121}$$

Equation (121) is the well-known exponential generating function relation for assemblies (cf. Foata [23]), which has as a special case the relationship

$$\hat{P}(s) = \exp(\hat{M}(s)). \tag{122}$$

Recall from Section 8 that in studying large deviations of K_n, the number of components in the structure of total weight n, we were led to the

measure $\mathbb{P}_\theta$ corresponding to sampling with probability proportional to θ^{K_n}. It follows from (1) that there is a normalizing constant $p_\theta(n)$ such that

$$p_\theta(n)\,\mathbb{P}_\theta(\mathbf{C}(n)=\mathbf{a}) = \theta^{a_1+\cdots+a_n} N(n,\mathbf{a})$$
$$= n!\,x^{-n}\prod_{j=1}^{n}\left(\frac{\theta m_j x^j}{j!}\right)^{a_j}\frac{1}{a_j!}\,\mathbf{1}\left(\sum_{l=1}^{n} l a_l = n\right)$$

for any $x>0$. Clearly,

$$p_\theta(n) = \sum_{k=1}^{n} p(n,k)\,\theta^k$$
$$= n!\,[s^n]\,\hat{P}(s,\theta) \tag{123}$$
$$= p(n)\,\mathbb{E}(\theta^{K_n}), \tag{124}$$

where $\mathbb{E}\equiv\mathbb{E}_1$ denotes expectation with respect to the uniform measure $\mathbb{P}\equiv\mathbb{P}_1$, corresponding to $\theta=1$.

Next we explore the connection with the probability generating function (pgf) of the random variable $T_n\equiv\sum_{j=1}^{n} jZ_j$, where the Z_j are independent Poisson distributed random variables with mean

$$\mathbb{E}_\theta Z_j \equiv \theta\lambda_j = \theta\frac{m_j x^j}{j!}.$$

Recall that the pgf of a Poisson-distributed random variable Z with mean λ is

$$\mathbb{E}_\theta s^Z \equiv \sum_{j=0}^{\infty}\mathbb{P}_\theta(Z=j)\,s^j = \exp(-\lambda(1-s)),$$

so using the independence of the Z_j,

$$\mathbb{E}_\theta s^{T_n} = \mathbb{E}_\theta s^{\sum_{j=1}^{n} jZ_j}$$
$$= \prod_{j=1}^{n}\mathbb{E}_\theta(s^j)^{Z_j}$$
$$= \exp\left(-\theta\sum_{j=1}^{n}\lambda_j(1-s^j)\right).$$

Thus

$$\mathbb{P}_\theta(T_n=n) = [s^n]\,\mathbb{E}_\theta s^{T_n}$$
$$= \exp\left(-\theta\sum_{j=1}^{n}\lambda_j\right)[s^n]\exp\left(\theta\sum_{j=1}^{n}\lambda_j s^j\right)$$
$$= \exp\left(-\theta\sum_{j=1}^{n}\lambda_j\right)[s^n]\exp\left(\theta\sum_{j=1}^{\infty}\lambda_j s^j\right)$$

$$= \exp\left(-\theta \sum_{j=1}^{n} \lambda_j\right) [s^n] \exp(\theta \hat{M}(sx))$$

$$= \exp\left(-\theta \sum_{j=1}^{n} \lambda_j\right) [s^n] \hat{P}(sx, \theta),$$

using (121) at the last step. Thus, via (123),

$$\mathbb{P}_\theta(T_n = n) = \exp\left(-\theta \sum_{j=1}^{n} \lambda_j\right) \frac{x^n p_\theta(n)}{n!}, \tag{125}$$

as can also be calculated from (24) and (108) for the special case $U = K_n$.

The next result gives a simple expression for the joint moments of the component counts. We use the notation $y_{[n]}$ to denote the falling factorial $y(y-1)\cdots(y-n+1)$.

LEMMA 5. *For $(r_1, ..., r_b) \in \mathbb{Z}_+^b$ with $m = r_1 + 2r_2 + \cdots + br_b$, we have*

$$\mathbb{E}_\theta \prod_{j=1}^{b} (C_j(n))_{[r_j]} = \mathbf{1}(m \leqslant n)\, x^{-m} \frac{n!}{p_\theta(n)} \frac{p_\theta(n-m)}{(n-m)!} \prod_{j=1}^{b} \left(\frac{\theta m_j x^j}{j!}\right)^{r_j}. \tag{126}$$

Proof. The key step is the substitution of $a_1, ..., a_b$ for $a_1 - r_1, ..., a_b - r_b$ in the third equality below. For $m \leqslant n$, we have

$$\begin{aligned}
\mathbb{E}_\theta \sum_{j=1}^{b} (C_j(n))_{[r_j]} &= \sum_{a_j \geqslant r_j,\, j=1,\, ...,\, b} \sum_{a_{b+1},\, ...,\, a_n:\, \sum j a_j = n} (a_1)_{[r_1]} \cdots (a_b)_{[r_b]} \frac{n!}{x^n p_\theta(n)} \\
&\quad \times \prod_{j=1}^{n} \left(\frac{\theta m_j x^j}{j!}\right)^{a_j} \frac{1}{a_j!} \\
&= \frac{n!}{x^n p_\theta(n)} \prod_{j=1}^{b} \left(\frac{\theta m_j x^j}{j!}\right)^{r_j} \sum \sum \prod_{j=1}^{b} \left(\frac{\theta m_j x^j}{j!}\right)^{a_j - r_j} \\
&\quad \times \frac{1}{(a_j - r_j)!} \prod_{j=b+1}^{n} \left(\frac{\theta m_j x^j}{j!}\right)^{a_j} \frac{1}{a_j!} \\
&= \frac{n!}{x^n p_\theta(n)} \prod_{j=1}^{b} \left(\frac{\theta m_j x^j}{j!}\right)^{r_j} \sum_{a_1,\, ...,\, a_n:\, \sum j a_j = n - m} \\
&\quad \times \prod_{j=1}^{n} \left(\frac{\theta m_j x^j}{j!}\right)^{a_j} \frac{1}{a_j!} \\
&= \frac{n!}{x^n p_\theta(n)} \prod_{j=1}^{b} \left(\frac{\theta m_j x^j}{j!}\right)^{r_j} \frac{x^{n-m} p_\theta(n-m)}{(n-m)!}. \quad \blacksquare
\end{aligned}$$

Remark. If $\{Z_i\}$ are mutually independent Poisson random variables with $\mathbb{E}_\theta Z_i = \theta m_i x^i/i!$, then the product term on the right of Eq. (126) is precisely $\mathbb{E}_\theta \prod_{j=1}^b (Z_j)_{[r_j]}$.

Remark. In the special case of permutations, in which $m_i = (i-1)!$ and $p(n) = n!$, the normalizing constant $p_\theta(n)$ is given by $p_\theta(n) = \theta(\theta+1)\cdots(\theta+n-1)$, and Eq. (126) reduces to

$$\mathbb{E}_\theta \prod_{j=1}^{b} (C_j(n))_{[r_j]} = \mathbf{1}(m \leqslant n) \binom{\theta+n-m-1}{n-m} \binom{\theta+n-1}{n}^{-1} \prod_{j=1}^{b} \left(\frac{\theta}{j}\right)^{r_j},$$

a result of Watterson [58, 59].

9.2. *Multisets*

For multisets, the (ordinary) generating functions are

$$P(s, \theta) \equiv 1 + \sum_{n=1}^{\infty} \left(\sum_{k=1}^{n} p(n, k)\, \theta^k \right) s^n, \tag{127}$$

$$P(s) \equiv 1 + \sum_{n=1}^{\infty} p(n)\, s^n = P(s, 1), \tag{128}$$

and

$$M(s) \equiv \sum_{n=1}^{\infty} m_n s^n. \tag{129}$$

In this case, using (2) gives

$$p(n, k) = \sum_{\mathbf{a}} N(n, \mathbf{a}) = \sum_{\mathbf{a}} \prod_{j=1}^{n} \binom{m_j + a_j - 1}{a_j},$$

the sum $\sum_{\mathbf{a}}$ being over $\{\mathbf{a} \in \mathbb{Z}_+^n : \sum i a_i = n, \sum a_i = k\}$. It follows that

$$\begin{aligned} P(s, \theta) &= 1 + \sum_{n=1}^{\infty} \sum_{k=1}^{n} \sum_{\mathbf{a}} \prod_{i=1}^{n} \binom{m_i + a_i - 1}{a_i} (\theta s^i)^{a_i} \\ &= \prod_{i=1}^{\infty} (1 - \theta s^i)^{-m_i} \qquad (130) \\ &= \exp\left(-\sum_{i=1}^{\infty} m_i \log(1 - \theta s^i) \right) \\ &= \exp\left(\sum_{i=1}^{\infty} m_i \sum_{j=1}^{\infty} \frac{(\theta s^i)^j}{j} \right) \\ &= \exp\left(\sum_{j=1}^{\infty} \frac{\theta^j}{j} \sum_{i=1}^{\infty} m_i s^{ij} \right) \\ &= \exp\left(\sum_{j=1}^{\infty} \frac{\theta^j}{j} M(s^j) \right). \qquad (131) \end{aligned}$$

See Flajolet and Soria [22], for example.

Under the measure $\mathbb{P}_\theta$, there is a normalizing constant $p_\theta(n)$ such that

$$\begin{aligned} p_\theta(n)\,\mathbb{P}_\theta(\mathbf{C}(n)=\mathbf{a}) &= \prod_{i=1}^{n} \binom{m_i+a_i-1}{a_i} \theta^{a_i} \mathbf{1}\left(\sum_{l=1}^{n} l a_l = n\right) \\ &= x^{-n} \prod_{l=1}^{n} (1-\theta x^l)^{-m_l} \prod_{i=1}^{n} \binom{m_i+a_i-1}{a_i} \\ &\quad \times (1-\theta x^i)^{m_i} (\theta x^i)^{a_i} \mathbf{1}\left(\sum_{l=1}^{n} l a_l = n\right), \end{aligned}$$

for any $0<x<1$. Indeed,

$$p_\theta(n) = p(n)\,\mathbb{E}_1(\theta^{K_n}) = [s^n]\,P(s,\theta), \tag{132}$$

where $p_\theta(0) \equiv 1$.

In this case, the relevant Z_j are independent negative binomial random variables with parameters m_i and θx^i and pgf

$$\mathbb{E}_\theta s^{Z_i} = \left(\frac{1-\theta x^i}{1-\theta x^i s}\right)^{m_i}.$$

Using the independence of the Z_j once more, the pfg of T_n may be found as

$$\begin{aligned} \mathbb{E}_\theta s^{T_n} &= \prod_{i=1}^{n} \mathbb{E}_\theta (s^i)^{Z_i} \\ &= \prod_{i=1}^{n} \left(\frac{1-\theta x^i}{1-\theta (xs)^i}\right)^{m_i} \\ &= \left(\prod_{i=1}^{n} (1-\theta x^i)^{m_i}\right) \prod_{i=1}^{n} (1-\theta (xs)^i)^{-m_i}. \end{aligned} \tag{133}$$

Using (130), we see that

$$\begin{aligned} \mathbb{P}_\theta(T_n = n) &= [s^n]\,\mathbb{E}_\theta s^{T_n} \\ &= \left(\prod_{i=1}^{n} (1-\theta x^i)^{m_i}\right) [s^n] \exp\left(-\sum_{i=1}^{n} m_i \log(1-\theta (xs)^i)\right) \\ &= \left(\prod_{i=1}^{n} (1-\theta x^i)^{m_i}\right) [s^n] \exp\left(-\sum_{i=1}^{\infty} m_i \log(1-\theta (xs)^i)\right) \\ &= \left(\prod_{i=1}^{n} (1-\theta x^i)^{m_i}\right) [s^n]\,P(sx,\theta), \end{aligned}$$

so that

$$\mathbb{P}_\theta(T_n = n) = \left(\prod_{i=1}^{n} (1 - \theta x^i)^{m_i} \right) x^n p_\theta(n). \tag{134}$$

Equation (134) can also be calculated from (24) and (108) for the special case $U = K_n$.

In order to calculate moments of the component counts $\mathbf{C}(n)$, it is convenient to use a variant on a theme of Shepp and Lloyd [52]. We assume that $M(s)$ has positive radius of convergence, R. As above, let $Z_1, Z_2, \ldots$ be mutually independent negative binomial random variables, Z_i having parameters m_i and θx^i, where $0 < x < \min\{R, 1, \theta^{-1}\}$. Let $T_\infty \equiv \sum_{i=1}^{\infty} iZ_i$. Note that T_∞ is almost surely finite, because

$$\mathbb{E}_\theta T_\infty = \sum_{i=1}^{\infty} \frac{im_i \theta x^i}{1 - \theta x^i} \leqslant \frac{\theta x}{1 - \theta x} M'(x) < \infty.$$

The distribution of T_∞ follows from (130), (133), and monotone convergence since

$$\mathbb{E}_\theta s^{T_\infty} = \frac{P(sx, \theta)}{P(x, \theta)}.$$

Hence

$$\mathbb{P}_\theta(T_\infty = n) = x^n p_\theta(n)/P(x, \theta), \qquad n = 0, 1, \ldots. \tag{135}$$

Further, for $\mathbf{a} \in \mathbb{Z}_+^n$ and $\mathbf{Z}(n) \equiv (Z_1, \ldots, Z_n)$

$$\mathbb{P}_\theta(\mathbf{C}(n) = \mathbf{a}) = \mathbb{P}_\theta(\mathbf{Z}(n) = \mathbf{a} \mid T_\infty = n). \tag{136}$$

This follows from the statement (109) that

$$\mathbb{P}_\theta(\mathbf{C}(n) = \mathbf{a}) = \mathbb{P}_\theta(\mathbf{Z}(n) = \mathbf{a} \mid T_n = n),$$

and the observation that

$$\begin{aligned}
\mathbb{P}_\theta(\mathbf{Z}(n) = \mathbf{a} \mid T_n = n) &= \frac{\mathbb{P}_\theta(\mathbf{Z}(n) = \mathbf{a}, T_n = n)}{\mathbb{P}_\theta(T_n = n)} \\
&= \frac{\mathbb{P}_\theta(\mathbf{Z}(n) = \mathbf{a}, T_n = n)\, \mathbb{P}_\theta(Z_{n+1} = Z_{n+2} = \cdots = 0)}{\mathbb{P}_\theta(T_n = n)\, \mathbb{P}_\theta(Z_{n+1} = Z_{n+2} = \cdots = 0)} \\
&= \frac{\mathbb{P}_\theta(\mathbf{Z}(n) = \mathbf{a}, T_\infty = n)}{\mathbb{P}_\theta(T_\infty = n)} \\
&= \mathbb{P}_\theta(\mathbf{Z}(n) = \mathbf{a} \mid T_\infty = n).
\end{aligned}$$

Now let $\Phi: \mathbb{Z}_+^\infty \to \mathbb{R}$, and set $\mathbf{C}_n \equiv (C_1(n), ..., C_n(n), 0, 0, ...)$ with $\mathbf{C}_0 \equiv (0, 0, ...)$. The aim is to find an easy way to compute $\mathbb{E}_\theta^n(\Phi) = \mathbb{E}_\theta \Phi(\mathbf{C}_n)$. It is convenient to use the notation $\mathbb{E}_{x,\theta}$ to denote expectations computed under the independent negative binomial measure with parameters x and θ. Shepp and Lloyd's method in the present context is the observation, based on (135) and (136), that $\mathbb{E}_{x,\theta}(\Phi \mid T_\infty = n) = \mathbb{E}_\theta^n(\Phi)$, so that

$$\begin{aligned}\mathbb{E}_{x,\theta}(\Phi) &= \sum_{n=0}^{\infty} \mathbb{E}_{x,\theta}(\Phi \mid T_\infty = n)\, \mathbb{P}_\theta(T_\infty = n) \\ &= \sum_{n=0}^{\infty} \mathbb{E}_\theta^n(\Phi)\, x^n p_\theta(n)/P(x,\theta). \end{aligned} \tag{137}$$

This leads to the result that

$$\mathbb{E}_\theta^n(\Phi) = \frac{[x^n]\, \mathbb{E}_{x,\theta}(\Phi)\, P(x,\theta)}{p_\theta(n)}. \tag{138}$$

For $r \geqslant 1$, $jr \leqslant n$, we use this method to calculate the falling factorial moments $\mathbb{E}_\theta(C_j(n))_{[r]}$. This determines all moments, since $C_j(n)_{[r]} \equiv 0$ if $jr > n$. In this case $\Phi(x_1, x_2, ...) = (x_j)_{[r]}$, and

$$\begin{aligned}\mathbb{E}_{x,\theta}(\Phi) &= \mathbb{E}_{x,\theta}(Z_j)_{[r]} \\ &= \frac{\Gamma(m_j + r)}{\Gamma(m_j)} \left(\frac{\theta x^j}{1 - \theta x^j}\right)^r.\end{aligned}$$

Hence we have

$$\begin{aligned}\mathbb{E}_\theta^n(\Phi) &= \frac{\Gamma(m_j + r)}{p_\theta(n)\, \Gamma(m_j)} [x^n]\, P(x,\theta) \left(\frac{\theta x^j}{1 - \theta x^j}\right)^r \\ &= \frac{\theta^r \Gamma(m_j + r)}{p_\theta(n)\, \Gamma(m_j)} [x^{n-rj}]\, P(x,\theta) \sum_{l=0}^{\infty} \binom{r + l - 1}{l} \theta^l x^{jl} \\ &= \frac{\theta^r \Gamma(m_j + r)}{p_\theta(n)\, \Gamma(m_j)} \sum_{l=0}^{\lfloor n/j \rfloor - r} \binom{r + l - 1}{l} \theta^l p_\theta(n - jr - jl) \\ &= \frac{\Gamma(m_j + r)}{p_\theta(n)\, \Gamma(m_j)} \sum_{m=r}^{\lfloor n/j \rfloor} \binom{m - 1}{r - 1} \theta^m p_\theta(n - jm). \end{aligned} \tag{139}$$

Remark. See Hansen [31] for related material. The Shepp and Lloyd method can also be used in the context of assemblies, for which (135) holds with

$$\mathbb{P}_\theta(T_\infty = n) = \frac{x^n}{n!} p_\theta(n)/\hat{P}(x,\theta), \qquad n = 0, 1, ... \tag{140}$$

This provides another proof of Lemma 5. See Hansen [29] for the case of random mappings, and Hansen [30] for the case of the Ewens sampling formula.

9.3. *Selections*

The details for the case of selections are similar to those for multisets. Most follow by replacing θ and m_i by $-\theta$ and $-m_i$, respectively, in the formulas for multisets. First, we have from (3)

$$p(n,k)=\sum_{\mathbf{a}} N(n,\mathbf{a})=\sum_{\mathbf{a}}\prod_{j=1}^{n}\binom{m_j}{a_j},$$

the sum $\sum_{\mathbf{a}}$ being over $\{\mathbf{a}\in\mathbb{Z}_+^n : \sum ia_i=n, \sum a_i=k\}$. Therefore

$$P(s,\theta)=1+\sum_{n=1}^{\infty}\sum_{k=1}^{n}\sum_{\mathbf{a}}\prod_{i=1}^{n}\binom{m_i}{a_i}(\theta s^i)^{a_i}$$

$$=\prod_{i=1}^{\infty}(1+\theta s^i)^{m_i} \tag{141}$$

$$=\exp\left(\sum_{j=1}^{\infty}\frac{(-1)^{j-1}\theta^j}{j}M(s^j)\right), \tag{142}$$

the last following just as (131) followed from (130). See Flajolet and Soria [22], for example.

Under the measure $\mathbb{P}_\theta$, there is a normalizing constant $p_\theta(n)$ such that

$$p_\theta(n)\,\mathbb{P}(\mathbf{C}(n)=\mathbf{a})=\prod_{i=1}^{n}\binom{m_i}{a_i}\theta^{a_i}\mathbf{1}\left(\sum_{l=1}^{n}la_l=n\right)$$

$$=x^{-n}\prod_{l=1}^{n}(1+\theta x^l)^{m_l}\prod_{i=1}^{n}\binom{m_i}{a_i}(1+\theta x^i)^{-m_i}(\theta x^i)^{a_i}$$

$$\times\mathbf{1}\left(\sum_{l=1}^{n}la_l=n\right),$$

for any $x>0$; $p_\theta(n)$ is given in (132) once more. In this case, the Z_j are independent binomial random variables with pgf

$$\mathbb{E}_\theta s^{Z_i}=\left(\frac{1+\theta x^i s}{1+\theta x^i}\right)^{m_i}, \tag{143}$$

and the pgf of T_n is

$$\mathbb{E}_\theta s^{T_n}=\left(\prod_{i=1}^{n}(1+\theta x^i)^{-m_i}\right)\prod_{i=1}^{n}(1+\theta(xs)^i)^{m_i}. \tag{144}$$

It follows from (130) that

$$\mathbb{P}_\theta(T_n = n) = \left(\prod_{i=1}^{n} (1+\theta x^i)^{-m_i} \right) [s^n]\, P(sx, \theta),$$

so that

$$\mathbb{P}_\theta(T_n = n) = \left(\prod_{i=1}^{n} (1+\theta x^i)^{-m_i} \right) x^n p_\theta(n). \tag{145}$$

The joint moments of the counts can be calculated using the Shepp and Lloyd construction once more. In particular, Eqs. (135) and (136) hold, and we can apply (138) with $\mathbb{E}_{x,\theta}(\Phi)$ denoting expectation with respect to independent binomial random variables $Z_1, Z_2, \ldots$ with distribution determined by (143).

As an example, we use this method to calculate $\mathbb{E}_\theta(C_j(n))_{[r]}$ for $r \geqslant 1$, $jr \leqslant n$. Since

$$\begin{aligned} \mathbb{E}_{x,\theta}(\Phi) &= \mathbb{E}_{x,\theta}(Z_j)_{[r]} \\ &= (m_j)_{[r]} \left(\frac{\theta x^j}{1+\theta x^j} \right)^r, \end{aligned}$$

from (138) we have

$$\begin{aligned} \mathbb{E}^n_\theta(\Phi) &= \frac{(m_j)_{[r]}}{p_\theta(n)} [x^n]\, P(x,\theta) \left(\frac{\theta x^j}{1+\theta x^j} \right)^r \\ &= \frac{(m_j)_{[r]}}{p_\theta(n)} \sum_{m=r}^{\lfloor n/j \rfloor} \binom{m-1}{r-1} (-1)^{m-r} \theta^m p_\theta(n-jm). \end{aligned} \tag{146}$$

9.4. *Recurrence Relations and Numerical Methods*

We saw in Theorems 3 and 8 that for any $B \subseteq [n]$, the total variation distance between $\mathbf{C}_B$ and $\mathbf{Z}_B$ can be expressed in terms of the distributions of random variables S_B and R_B defined by

$$S_B = \sum_{i \in [n]-B} iZ_i, \tag{147}$$

and

$$R_B = \sum_{i \in B} iZ_i \equiv S_{[n]-B}. \tag{148}$$

Specifically,

$$\begin{aligned} d_{TV}(\mathscr{L}_\theta(\mathbf{C}_B), \mathscr{L}_\theta(\mathbf{Z}_B)) &= \frac{1}{2} \mathbb{P}_\theta(R_B > n) \\ &\quad + \frac{1}{2} \sum_{r=0}^{n} \mathbb{P}_\theta(R_B = r) \left| \frac{\mathbb{P}_\theta(S_B = n-r)}{\mathbb{P}_\theta(T_n = n)} - 1 \right|. \end{aligned} \tag{149}$$

A direct attack on estimation of $d_{TV}(\mathscr{L}_\theta(\mathbf{C}_B), \mathscr{L}_\theta(\mathbf{Z}_B))$ can be based on a generating function approach to the asymptotics (for large n) of the terms in (149). In the setting of assemblies, this uses the result before (125) for $\mathbb{P}_\theta(T_n = n)$, and the fact that for $k \geqslant 0$

$$\begin{aligned}\mathbb{P}_\theta(S_B = n-k) &= [s^{n-k}] \exp\left(-\theta \sum_{i \in [n]-B} \lambda_i (1-s^i)\right) \\ &= \exp\left(-\theta \sum_{i \in [n]-B} \lambda_i\right) [s^{n-k}] \exp\left(\theta \sum_{i \in [n]-B} \lambda_i s^i + \theta \sum_{i>n} \lambda_i s^i\right) \\ &= \exp\left(-\theta \sum_{i \in [n]-B} \lambda_i\right) [s^{n-k}]\, \hat{P}(sx, \theta) \exp\left(-\theta \sum_{i \in B} \lambda_i s^i\right).\end{aligned} \tag{150}$$

For applications of this technique, see Arratia, Stark, and Tavaré [7] and Stark [55].

It is also useful to have a recursive method for calculating the distribution of R_B for any $B \subseteq [n]$. Clearly, for assemblies

$$\mathbb{E}_\theta s^{R_B} = \exp\left(-\sum_{i \in B} \theta \lambda_i\right) \exp\left(\sum_{i \in B} \theta \lambda_i s^i\right). \tag{151}$$

Write

$$G_B(s) = \sum_{i \in B} \theta \lambda_i s^i,$$

and

$$F_B(s) = \exp G_B(s) \equiv \sum_{k=0}^{\infty} q_B(k)\, s^k,$$

with $q_B(0) \equiv 1$. Differentiating with respect to s shows that $sF_B'(s) = sG_B'(s)\, F_B(s)$ (cf. Pourahmadi [50]), and equating coefficients of s^k gives

$$k q_B(k) = \sum_{i=1}^{k} g_B(i)\, q_B(k-i), \qquad k = 1, 2, \ldots,$$

where

$$g_B(i) = \theta i \lambda_i\, \mathbf{1}(i \in B). \tag{152}$$

Since $p_B(k) \equiv \mathbb{P}_\theta(R_B = k) = \exp(-G_B(1))\, q_B(k)$, we find that

$$kp_B(k) = \sum_{i=1}^{k} g_B(i)\, p_B(k-i), \qquad k = 1, 2, \ldots \tag{153}$$

with $p_B(0) = \exp(-G_B(1))$. The relation (153) has been exploited in the case of uniform permutations ($\theta = 1$, $\lambda_i = 1/i$) by Arratia and Tavaré [2].

For multisets, the analog of (150) is

$$\begin{aligned}
\mathbb{P}_\theta(S_B = n-k) &= \Big(\prod_{i \in [n]-B} (1-\theta x^i)^{m_i} \Big) [s^{n-k}] \prod_{i \in [n]-B} (1-\theta(xs)^i)^{-m_i} \\
&= \Big(\prod_{i \in [n]-B} (1-\theta x^i)^{m_i} \Big) [s^{n-k}] \prod_{i \in [n]-B} (1-\theta(xs)^i)^{-m_i} \\
&\quad \times \prod_{i > n} (1-\theta(xs)^i)^{-m_i} \\
&= \Big(\prod_{i \in [n]-B} (1-\theta x^i)^{m_i} \Big) [s^{n-k}]\, P(sx, \theta) \prod_{i \in B} (1-\theta(xs)^i)^{m_i}.
\end{aligned} \tag{154}$$

To develop a recursion for $p_B(k) \equiv \mathbb{P}_\theta(R_B = k)$, we can use logarithmic differentiation; cf. Apostol [1, Theorem 14.8]. First, we have

$$\mathbb{E} s^{R_B} = \prod_{i \in B} (1-\theta x^i)^{m_i} \prod_{i \in B} (1-\theta x^i s^i)^{-m_i}. \tag{155}$$

Define

$$G_B(s) = \sum_{i \in B} m_i s^i,$$

and

$$F_B(s) = \prod_{i \in B} (1-\theta x^i s^i)^{-m_i} \equiv \sum_{k=0}^{\infty} q_B(k)\, s^k,$$

with $q_B(0) = 1$. Then

$$\log F_B(s) = \sum_{j=1}^{\infty} \frac{\theta^j}{j} G_B((xs)^j).$$

Differentiating with respect to s and simplifying shows that

$$sF_B'(s) = \Big(\sum_{i \geq 1} g_B(i)\, s^i \Big) F_B(s),$$

where

$$g_B(i) = x^i \sum_{k \mid i} k m_k \theta^{i/k} \mathbf{1}(k \in B). \tag{156}$$

Equating coefficients of s^k gives

$$k q_B(k) = \sum_{i=1}^{k} g_B(i)\, q_B(k-i), \qquad k = 1, 2, \ldots.$$

Since $p_B(k) \equiv \mathbb{P}_\theta(R_B = k) = \prod_{i \in B} (1 - \theta x^i)^{m_i} q_B(k)$, it follows that

$$k p_B(k) = \sum_{i=1}^{k} g_B(i)\, p_B(k-i), \qquad k = 1, 2, \ldots \tag{157}$$

with $p_B(0) = \prod_{i \in B} (1 - \theta x^i)^{m_i}$.

For selections, we have the following identity, valid for $k \geqslant 0$:

$$\mathbb{P}_\theta(S_B = n-k) = \left(\prod_{i \in [n]-B} (1 + \theta x^i)^{-m_i} \right) [s^{n-k}]\, P(sx, \theta) \prod_{i \in B} (1 + \theta (xs)^i)^{-m_i}.$$

If we define $p_B(k) \equiv \mathbb{P}_\theta(R_B = k)$, then we obtain

$$k p_B(k) = \sum_{i=1}^{k} g_B(i)\, p_B(k-i), \qquad k = 1, 2, \ldots, \tag{158}$$

where

$$g_B(i) = -x^i \sum_{k \mid i} k m_k (-\theta)^{i/k} \mathbf{1}(k \in B),$$

and

$$p_B(0) = \prod_{i \in B} (1 + \theta x^i)^{-m_i}.$$

10. Proofs by Overpowering the Conditioning

The basic strategy for making the relation $\mathbf{C}_I \stackrel{\mathrm{d}}{=} (\mathbf{Z}_I \mid T = t)$ into a useful approximation is to pick the free parameter x in the distribution of $\mathbf{Z}_I$ so that the conditioning is not severe, i.e., so that $\mathbb{P}(T = t)$ is not too small. It is sometimes possible to get useful upper bounds on events involving the combinatorial process $\mathbf{C}_I$ by combining upper bounds on the probability of the same event for the independent process, together with lower bounds for $\mathbb{P}(T = t)$. The formal description of this strategy is given by the following lemma.

LEMMA 6. *Assume that* $\mathbf{C}_I \stackrel{d}{=} (\mathbf{Z}_I \mid T=t)$ *and that* h *is a nonnegative functional of these processes, i.e.,*

$$h: \mathbb{Z}_+^I \to \mathbb{R}_+.$$

Then

$$\mathbb{E}h(\mathbf{C}_I) \leqslant \frac{\mathbb{E}h(\mathbf{Z}_I)}{\mathbb{P}(T=t)}. \tag{159}$$

Proof.

$$\mathbb{E}h(\mathbf{C}_I) = \frac{\mathbb{E}(h(\mathbf{Z}_I)\,\mathbf{1}(T=t))}{\mathbb{P}(T=t)} \leqslant \frac{\mathbb{E}h(\mathbf{Z}_I)}{\mathbb{P}(T=t)}. \quad \blacksquare$$

10.1. *Example*: *Partitions of a Set*

Recall that partitions of a set is the assembly with $m_i = 1$ for all i. Following the discussion in Subsection 5.3 we take $x \equiv x(n) = \log n - \log\log n + \cdots$ to be the solution of $xe^x = n$, so that for $i = 1, 2, \ldots, n$, Z_i is Poisson distributed, with mean and variance $\lambda_i = x^i/i!$. With this choice of x, we have

$$\mathbb{E}T_n = \sum_1^n i\lambda_i \sim xe^x = n$$

and

$$\sigma_n^2 \equiv \operatorname{var}(T_n) = \sum_1^n i^2\lambda_i \sim n\log n. \tag{160}$$

By combining (24) with the asymptotics for the Bell numbers given in Moser and Wyman [46], and simplifying, we see that

$$\mathbb{P}(T_n = n) \sim \frac{1}{\sqrt{2\pi n \log n}} \sim \frac{1}{\sqrt{2\pi}\,\sigma_n}, \tag{161}$$

which is easy to remember, since it agrees with what one would guess from the local central limit heuristic.

Write $U_n = Z_1 + Z_2 + \cdots + Z_n$, so that the total number of blocks K_n satisfies $K_n \stackrel{d}{=} (U_n \mid T_n = n)$. Harper [32] proved that K_n is asymptotically normal with mean n/x and variance n/x^2. We observe that this contrasts with the unconditional behavior: U_n is asymptotically normal with mean n/x, like K_n, and variance n/x, unlike K_n. Since U_n is Poisson, it has equal mean and variance. Harper's result says that conditioning on $T_n = n$ reduces the variance of U_n by a factor asymptotic to $\log n$.

Note that Z_1 is Poisson with parameter $x \sim \log n$, and hence the distribution of Z_1 is asymptotically normal with mean and variance $\log n$. Note also that the Poisson parameters $\lambda_i = x^i/i!$ are themselves proportional to $\mathbb{P}(Z_1 = i)$; in fact for $i \geqslant 1$

$$\lambda_i = e^x \mathbb{P}(Z_1 = i) = \frac{n}{x} \mathbb{P}(Z_1 = i).$$

We can use the normal approximation for Z_1 to see that, for fixed $a < b$, as $n \to \infty$,

$$\sum_{a\sqrt{\log n} < i - \log n < b\sqrt{\log n}} \lambda_i \sim \frac{n}{\log n} \int_a^b \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du.$$

Informally, the relatively large values of λ_i occur for i within a few $\sqrt{\log n}$ of $\log n$.

10.1.1. *The Size of a Randomly Selected Block*

A result similar to the following appears as Corollary 3.3 in DeLaurentis and Pittel [12]. The size $D \equiv D_n$ of "a randomly selected component" of a random assembly on n elements is defined by a two step procedure: first pick a random assembly, then pick one of its K_n components, each with probability $1/K_n$. The same definition applies to the case of random multisets or selections of weight n.

Given $1 \leqslant b \leqslant n$, consider the functional $h: \mathbb{Z}_+^n \to [0, 1]$ defined by

$$h(\mathbf{a}) = \left(\sum_{i \leqslant b} a_i\right) \Big/ \left(\sum_{i \leqslant n} a_i\right),$$

with $h(0, 0, ..., 0)$ defined to be 1. The distribution of the size of a randomly selected component is determined by

$$\mathbb{P}(D_n \leqslant b) = \mathbb{E}h(\mathbf{C}(n)).$$

Define $U_b = Z_1 + \cdots + Z_b$, so that $h(Z_1, ..., Z_n) = U_b/U_n$ and

$$\mathbb{P}(D_n \leqslant b) = \mathbb{E}h((Z_1, ..., Z_n) \mid T_n = n) = \mathbb{E}\left(\frac{U_b}{U_n} \,\middle|\, T_n = n\right).$$

Let $\varepsilon > 0$ and $\rho > 1$ be given. Let $1 \leqslant b \leqslant n$ such that

$$q \equiv \mathbb{P}(Z_1 \leqslant b) \in [2\varepsilon, 1 - 2\varepsilon]. \tag{162}$$

Now for all $n \geqslant n(\varepsilon, \rho)$ we have $\mathbb{E}U_b > \varepsilon n/\log n$ and $\mathbb{E}U_b/\mathbb{E}U_n \in [q/\rho, q\rho]$. Large deviation theory says that for $\rho > 1$ there is a constant $c = c(\rho) > 0$

such that if Y is Poisson with parameter λ, then $\mathbb{P}(Y/\lambda \leqslant 1/\rho) \leqslant \exp(-\lambda c)$ and $\mathbb{P}(Y/\lambda \geqslant \rho) \leqslant \exp(-\lambda c)$. (In fact, the optimal c is given by $c(\rho) = \min(1 + \rho \log \rho - \rho, 1 - \rho^{-1} \log \rho - \rho^{-1})$, with the two terms in the minimum corresponding respectively to large deviations above the mean and below the mean.) Putting these together, using the large deviation bounds once with U_b as Y and a second time with U_n as Y, we have for $n \geqslant n(\varepsilon, \rho)$

$$\mathbb{P}\left(\frac{U_b}{U_n} \notin [q/\rho^3, q\rho^3]\right) \leqslant 2 \exp(-c(\rho)\, \varepsilon n/\log n).$$

Since the functional h takes values in $[0, 1]$, this proves, for $n \geqslant n(\varepsilon, \rho)$,

$$|\mathbb{P}(D_n \leqslant b) - q| \leqslant q(\rho^3 - 1) + 2 \exp(-c(\rho)\, \varepsilon n/\log n)/\mathbb{P}(T_n = n). \quad (163)$$

In terms of Lemma 6, the above argument involves the functional h^* defined by $h^*(\mathbf{a}) = \mathbf{1}(h(\mathbf{a}) \notin [q/\rho^3, q\rho^3])$. The inequality (163) not only proves that D_n is asymptotically normal with mean and variance $\log n$, but also provides an upper bound on the Prohorov distance between the distributions of D_n and Z_1.

10.1.2. *The Size of the Block Containing a Given Element*

In the case of assemblies, it is possible that someone describing "a randomly selected component" has in mind the component containing a randomly selected element, where the element and the assembly are chosen independently. This includes, for example, the case where the element is deterministically chosen, say it is always 1. Let D_n^* be the size of the component containing 1, in a random assembly on the set $[n]$.

The two notions of "a randomly selected component" can be very far apart. For example, with random permutations, D_n^* is uniformly distributed over $\{1, 2, ..., n\}$, while the size D_n of a randomly cycle is such that $\log D_n/\log n$ is approximately uniformly distributed over $[0, 1]$. For random partitions of a set, the argument below proves that D_n and D_n^* are close in distribution, because both distributions are close to Poisson with parameter x, where $xe^x = n$.

Given $1 \leqslant b \leqslant n$, consider the functional $g: \mathbb{Z}_+^n \to [0, 1]$ defined by

$$g(\mathbf{a}) = \frac{1}{n} \sum_{i \leqslant b} i a_i.$$

The distribution of the size of the component containing a given element is determined by

$$\mathbb{P}(D_n^* \leqslant b) = \mathbb{E}g(\mathbf{C}(n)).$$

Define $R_b = Z_1 + 2Z_2 + \cdots + bZ_b$, so that $g(Z_1, ..., Z_n) = R_b/n$ and

$$\mathbb{P}(D_n^* \leqslant b) = \mathbb{E}g((Z_1, ..., Z_n) \mid T_n = n) = \mathbb{E}(R_b/n \mid T_n = n).$$

With ε, ρ, b, n, and q as above in (162), and with the same $c(\rho)$ as above but with a different $n(\varepsilon, \rho)$, for all $n \geqslant n(\varepsilon, \rho)$ we have $\mathbb{E}U_b > \varepsilon n/\log n$ and $\mathbb{E}R_b/n \in [q/\rho, q\rho]$. Large deviation theory says that, with $\lambda = \mathbb{E}U_b$ as the mean of an unweighted sum of independent Poissons, the weighted sum $Y = R_b$ satisfies $\mathbb{P}(Y/\mathbb{E}Y \leqslant 1/\rho) \leqslant \exp(-\lambda c)$ and $\mathbb{P}(Y/\mathbb{E}Y \geqslant \rho) \leqslant \exp(-\lambda c)$. Putting these together, we have for $n \geqslant n(\varepsilon, \rho)$

$$\mathbb{P}\left(\frac{R_b}{n} \notin [q/\rho^2, q\rho^2]\right) \leqslant 2\exp(-c(\rho)\,\varepsilon n/\log n).$$

Since the functional g takes values in $[0, 1]$, this proves, for $n \geqslant n(\varepsilon, \rho)$,

$$|\mathbb{P}(D_n^* \leqslant b) - q| \leqslant q(\rho^2 - 1) + 2\exp(-c(\rho)\,\varepsilon n/\log n)/\mathbb{P}(T_n = n). \quad (164)$$

10.1.3. *The Number of Distinct Block Sizes*

Odlyzko and Richmond [48] prove that the number J_n of distinct block sizes in a random partition of the set $[n]$ is asymptotic to $e\log n$ in expectation and in probability. A stronger result can easily be proved by overwhelming the conditioning.

Informally, our argument is that for $1 \leqslant i \leqslant (e-\varepsilon)\log n$, the Poisson parameter $\lambda_i = x^i/i!$ is large, so that $\mathbb{P}(Z_i = 0)$ is very small, in fact small enough to overwhelm the conditioning on $\{T_n = n\}$, so that $\mathbb{P}(C_i(n) = 0)$ is also very small, and we can conclude $\mathbb{P}(C_i(n) = 0$ for any $i \leqslant (e-\varepsilon)\log n) \to 0$. This accounts for at least $(e-\varepsilon)\log n$ distinct block sizes. On the other side, $\sum_{i \geqslant (e+\varepsilon)\log n} \mathbb{E}Z_i$ is small, hence for some $k = k(\varepsilon)$, $\mathbb{P}(Z_i > 0$ for at least k values of $i \geqslant (e+\varepsilon)\log n)$ is very small, in fact small enough to overwhelm the conditioning (using roughly $k = 1/(2\varepsilon)$.) We conclude $\mathbb{P}(C_i(n) > 0$ for at least k values $i \geqslant (e+\varepsilon)\log n) \to 0$. Our result, that for any $\varepsilon > 0$, $\mathbb{P}(C_i(n) = 0$ for any $i \leqslant (e-\varepsilon)\log n$, or $C_i(n) > 0$ for at least k values $i \geqslant (e+\varepsilon)\log n) \to 0$, implies but is not implied by the result that $J_n/\log n \to e$ in probability. Furthermore, the bounds supplied by Theorem 10 below imply that $J_n/\log n \to e$ in the rth mean for every $1 \leqslant r < \infty$. The result that $\mathbb{P}(C_1(n) = 0) \to 0$ was proved in Sachkov [51].

In a little more detail, observe that $\mathbb{P}(Z_1 = 0) = \exp(-\lambda_1) = e^{-x} = x/n \sim \log n/n$, which is smaller than the conditioning probability, given by (161), by a factor on the order of $\sqrt{n}/(\log n)^{3/2}$. The preceding argument is given in Sachkov [51]. The Poisson parameters increase rapidly, so $\mathbb{P}(Z_2 = 0) = \exp(-\lambda_2) = \exp(-x^2/2) = (x/n)^{x/2}$, which decays faster than any power of n.

For a more careful analysis of the boundary where the Poisson parameter λ_i changes from large to small, write $i = (x+d)\,e$, where

$d = o(x)$. Recall $x \sim \log n$. Using Stirling's formula, and writing $\approx$ for logarithmically asymptotic, we have $\lambda_i = x^i/i! \sim (xe/i)^i/\sqrt{2\pi i} = (x/(x+d))^i/\sqrt{2\pi i} \approx \exp(-id/x - \log\sqrt{i}) \approx \exp(-ed - (1/2)\log\log n)$, so that the critical boundary for i, corresponding to $d = -(1/2e)\log\log n$, is at $c(n) \equiv xe - (1/2)\log\log n$. On the left side of this boundary the argument via overwhelming the conditioning shows that $\mathbb{P}(C_i(n) = 0$ for any $i < ex - (3/2 + \varepsilon)\log\log n) \to 0$. The argument is very asymmetric between left and right: on the left, where λ_i is large, we use $\mathbb{P}(Z_i = 0) = \exp(-\lambda_i)$, gaining the use of an exponential; while on the right, where λ_i is small, we use $\mathbb{P}(Z_i > 0) < \lambda_i$. Thus in Theorem 10, the left boundary a is an extra $(1+\varepsilon)\log\log n$ below $c(n)$, while the right boundary b is an extra $\varepsilon \log n$ above $c(n)$.

The results of the above discussion are summarized by the following

THEOREM 10. *For partitions of a set of size n, for $\varepsilon > 0$, there are with high probability blocks of every size $i \leqslant (e-\varepsilon)\log n$, and not many blocks of size $i \geqslant (e+\varepsilon)\log n$. More precisely, for any $r < \infty$ there exists $k = k(\varepsilon, r) < \infty$ so that, as $n \to \infty$,*

$$\mathbb{P}(C_1(n) = 0) = O((\log n)^{3/2}/\sqrt{n}),$$

while for $a \equiv ex - (3/2 + \varepsilon)\log\log n$

$$\mathbb{P}(C_i(n) = 0 \text{ for any } 2 \leqslant i \leqslant a) \leqslant \frac{1}{\mathbb{P}(T_n = n)} \sum_2^a e^{-\lambda_i} = o(n^{-r}),$$

and

$$\mathbb{P}\left(\sum_{i \geqslant b \equiv (e+\varepsilon)\log n} C_i(n) \geqslant k\right) = O\left(\frac{1}{\mathbb{P}(T_n = n)}\left(\sum_{i \geqslant b} \lambda_i\right)^k\right) = o(n^{-r}),$$

where $xe^x = n$, $\lambda_i = x^i/i!$, and $\mathbb{P}(T_n = n)$ satisfies (161).

Proof. Most of the proof is contained in the informal discussion before the theorem. For the second statement, it remains to check that $\sum_2^a \exp(-\lambda_i) = o(n^{-r})$ for any r, which follows from an upper bound on the first and last terms of the sum, which has at most n terms, together with the observation that the $\lambda_2 < \lambda_3 < \cdots < \lambda_{\lfloor x \rfloor} \geqslant \cdots \geqslant \lambda_{\lfloor a \rfloor}$. For the third statement, we are merely using the estimate, for $Y = \sum_{i \geqslant b} Z_i$, which is Poisson with small parameter λ, that $\mathbb{P}(Y \geqslant k) = O(\lambda^k)$ as $\lambda \to 0$. Note that $\mathbb{E}Y \approx \mathbb{E}Z_{\lceil b \rceil} \equiv \lambda_{\lceil b \rceil} \approx (xe/b)^b \approx (1+\varepsilon/e)^{-b} < n^{-\varepsilon}$. ∎

The above argument by overwhelming the conditioning is crude but easy to use because it gives away a factor of $\mathbb{P}(T_n = n)$, when in fact the event $\{T_n = n\}$ is approximately independent of the events involving $\{Z_i > 0\}$ for large i. An effective way to quantify and handle this approximate

independence is the total variation method outlined in Sections 3 and 4. Sachkov [51] analyzed the size L_n of the largest block of a random partition, and gave its approximate distribution. Writing $L_n = h(\mathbf{C}(n))$ where $h(a_1, ..., a_n) = \max(i: a_i > 0)$, Sachkov's result can be paraphrased as $d_{TV}(L_n, h(\mathbf{Z}_n)) \to 0$. Note that the number J_n of distinct block sizes satisfies $J_n \leqslant L_n$ always. Using $B = \{i \leqslant n: i > ex - 2 \log\log n\}$, for example, it should be possible to prove that $d_{TV}(\mathbf{C}_B, \mathbf{Z}_B) \to 0$. Then, by comparison of $J_n = h(\mathbf{C}(n))$ with $h(Z_1, ..., Z_n) = \sum \mathbf{1}(Z_i > 0)$, it would follow that, with centering constants $c(n) \equiv \mathrm{ex} - (1/2e) \log\log n$, the family of random variable $\{J_n - c(n)\}$ is tight, and the family $\{L_n - J_n\}$ is tight; and for each family, along a subsequence $n(k)$ there is convergence in distribution if and only if the $c(n(k)) \bmod 1$ converge.

11. Dependent Process Approximations

For the logarithmic class of structures discussed in Subsections 4.3, 5.1, and 5.2, we have seen that the Ewens sampling formula (ESF) plays a crucial role. In the counting process for large components of logarithmic combinatorial structures, there is substantial dependence; an appropriate comparison object is the dependent process of large components in the ESF. For example, in Arratia, Barbour, and Tavaré [5] it is shown that the process of counts of factors of large degree in a random polynomial over a finite field is close in total variation to the process of counts of large cycles in a random permutation, corresponding to the ESF with parameter $\theta = 1$. In Arratia, Barbour, and Tavaré [6], Stein's method is used to establish an analogous result for all the logarithmic class, and somewhat more generally. The basic technique involving Stein's method specialized to the compound Poisson is described in Barbour, Holst, and Janson [9, Chap. 10].

Once such bounds are available, it is a simple matter to establish approximation results, with bounds, for other interesting functionals of the large component counts of the combinatorial process. For example, the Poisson–Dirichlet and GEM limits for random polynomials are established with metric bounds in Arratia, Barbour, and Tavaré [5]. Poisson–Dirichlet limits for the logarithmic class are also discussed by Hansen [31].

References

1. T. M. Apostol, "An Introduction to Analytic Number Theory," Springer-Verlag, New York, 1976.
2. R. Arratia and S. Tavaré, The cycle structure of random permutations, *Ann. Probab.* **20** (1992), 1567–1591.

3. R. Arratia and S. Tavaré, Limit theorems for combinatorial structures via discrete process approximations, *Random Structures Algorithms* **3** (1992), 321–345.
4. R. Arratia, A. D. Barbour, and S. Tavaré, Poisson process approximations for the Ewens Sampling Formula, *Ann. Appl. Probab.* **2** (1992), 519–535.
5. R. Arratia, A. D. Barbour, and S. Tavaré, On random polynomials over finite fields, *Math. Proc. Cambridge Philos. Soc.* **114** (1993), 347–368.
6. R. Arratia, A. D. Barbour, and S. Tavaré, Logarithmic combinatorial structures, in preparation.
7. R. Arratia, D. Stark, and S. Tavaré, Total variation asymptotics for Poisson process approximations of logarithmic combinatorial assemblies, *Ann. Probab.*, to appear.
8. A. D. Barbour, Refined approximations for the Ewens sampling formula, *Random Structures Algorithms* **3** (1992), 267–276.
9. A. D. Barbour, L. Holst, and S. Janson, "Poisson Approximation," Oxford Univ. Press, Oxford, 1992.
10. B. Bollobás, "Random Graphs," Academic Press, New York, 1985.
11. N. G. de Bruijn, "Asymptotic Methods in Analysis," Dover, New York, 1981; republication of 1958 edition, North-Holland, Amsterdam.
12. J. M. DeLaurentis, and B. G. Pittel, Counting subsets of the random partition and the 'Brownian bridge' process, *Stochastic Process. Appl.* **15** (1983), 155–167.
13. P. Diaconis and D. Freedman, Finite exchangeable sequences, *Ann. Probab.* **8** (1980), 745–764.
14. P. Diaconis and J. W. Pitman, Unpublished lecture notes, Statistics Department, University of California, Berkeley, 1986.
15. P. Diaconis, M. McGrath, and J. W. Pitman, Riffle shuffles, cycles and descents, *Combinatorica* (1994), in press.
16. R. S. Ellis, "Entropy, Large Deviations, and Statistical Mechanics," Springer, Berlin, 1985.
17. S. N. Ethier and T. G. Kurtz, "Markov Processes: Characterization and Convergence," Wiley, New York, 1986.
18. W. J. Ewens, The sampling theory of selectively neutral alleles, *Theoret. Population Biol.* **3** (1972), 87–112.
19. W. Feller, The fundamental limit theorems in probability, *Bull. Amer. Math. Soc.* **51** (1945), 800–832.
20. P. Flajolet and A. M. Odlyzko, Singularity analysis of generating functions, *SIAM J. Discrete Math.* **3** (1990), 216–240.
21. P. Flajolet and A. M. Odlyzko, Random mapping statistics, *in* "Proceedings, Eurocrypt, 1989" (J.-J. Quisquater, Ed.), pp. 329–354, Lecture Notes in Comput. Sci., Vol. 434, Springer-Verlag, New York/Berlin, 1990.
22. P. Flajolet and M. Soria, Gaussian limiting distributions for the number of components in combinatorial structures, *J. Combin. Theory Ser. A* **53** (1990), 165–182.
23. D. Foata, "La série génératrice exponentielle dans les problèmes d'énumérations," Press Univ. Montreal, 1974.
24. B. Fristedt, The structure of random partitions of large sets, preprint.
25. B. Fristedt, The structure of random partitions of large integers, *Trans. Amer. Math. Soc.* **337** (1993), 703–735.
26. W. M. Y. Goh and E. Schmutz, The number of distinct parts in a random partition integer, *J. Comb. Theory A*, in press.
27. V. L. Goncharov, Some facts from combinatorics, *Izv. Akad. Nauk SSSR Ser. Mat.* **8** (1944), 3–48; on the field of combinatory analysis, *Transl. Amer. Math. Soc.* **19** (1944), 1–46.
28. R. C. Griffiths, On the distribution of points in a Poisson–Dirichlet process, *J. Appl. Probab.* **25** (1988), 336–345.

29. J. C. HANSEN, A functional central limit theorem for random mappings, *Ann. Probab.* **17** (1989), 317–332.
30. J. C. HANSEN, A functional central limit theorem for the Ewens Sampling Formula, *J. Appl. Probab.* **27** (1990), 28–43.
31. J. C. HANSEN, Order statistics for decomposable combinatorial structures, *Random Structures Algorithms*, in press.
32. L. H. HARPER, Stirling behavior is asymptotically normal, *Ann. Math. Statist.* **38** (1967), 410–414.
33. B. HARRIS, Probability distributions related to random mappings, *Ann. Math. Statist.* **31** (1960), 1045–1062.
34. L. HOLST, A unified approach to limit theorems for urn models, *J. Appl. Probab.* **16** (1979), 154–162.
35. L. HOLST, Two conditional limit theorems with applications, *Ann. Statist.* **7** (1979), 551–557.
36. L. HOLST, Some conditional limit theorems in exponential families, *Ann. Probab.* **9** (1981), 818–830.
37. T. IGNATOV, On a constant arising in the asymptotic theory of symmetric groups, and on Poisson–Dirichlet measures, *Theory Probab. Appl.* **27** (1982), 136–147.
38. A. JOYAL, Une théorie combinatoire des séries formelles, *Adv. Math.* **42** (1981), 1–82.
39. V. F. KOLCHIN, "Random Mappings," Optimization Software, Inc., New York, 1986.
40. V. F. KOLCHIN, B. A. SEVAST'YANOV, AND V. P. CHISTYAKOV, "Random Allocations," Wiley, New York, 1978.
41. B. LEVIN, A representation for multinomial cumulative distribution functions, *Ann. Statist.* **9** (1981), 1123–1126.
42. R. LIDL AND H. NIEDERREITER, "Introduction to Finite Fields and their Applications," Cambridge Univ. Press, London/New York, 1986.
43. A. MEIR AND J. W. MOON, On random mapping patterns, *Combinatorica* **4** (1984), 61–70.
44. N. METROPOLIS AND G.-C. ROTA, Witt vectors and the algebra of necklaces, *Adv. Math.* **50** (1983), 95–125.
45. N. METROPOLIS AND G.-C. ROTA, The cyclotomic identity, *in* "Contemporary Mathematics," Vol. 34, pp. 19–27, Amer. Math. Soc., Providence, RI, 1984.
46. L. MOSER AND M. WYMAN, An asymptotic formula for the Bell numbers, *Trans. Roy. Soc. Canada* **49** (1955), 49–53.
47. L. R. MUTAFCIEV, Limit theorems for random mapping patterns, *Combinatorica* **8** (1988), 345–356.
48. A. M. ODLYZKO AND L. B. RICHMOND, On the number of distinct block size in partitions of a set, *J. Combin. Theory Ser. A* **38** (1985), 170–181.
49. R. OTTER, The number of trees, *Ann. of Math.* **49** (1948), 583–599.
50. M. POURAHMADI, Taylor expansion of $\exp(\sum_{k=0}^{\infty} a_k z^k)$ and some applications, *Amer. Math. Monthly* **91** (1984), 303–307.
51. V. N. SACHKOV, Random partitions of sets, *Theory Probab. Appl.* **19** (1974), 184–190.
52. L. A. SHEPP AND S. P. LLOYD, Ordered cycle lengths in a random permutation, *Trans. Amer. Math. Soc.* **121** (1966), 340–357.
53. A. J. STAM, Distance between sampling with and without replacement, *Statist. Neerlandica* **32** (1978), 81–91.
54. D. STARK, Unpublished Ph.D. Thesis, Department of Mathematics, University of Southern California, 1994.
55. D. STARK, Total variation asymptotics for independent process approximations of logarithmic multisets and selections, preprint.
56. V. E. STEPANOV, Limit distributions for certain characteristics of random mappings, *Theory Probab. Appl.* **14** (1969), 612–626.

57. A. M. VERSHIK AND A. A. SHMIDT, Limit measures arising in the theory of groups, I, *Theory Probab. Appl.* **22** (1977), 79–85.
58. G. A. WATTERSON, The sampling theory of selectively neutral alleles, *Adv. in Appl. Probab.* **6** (1974), 463–488.
59. G. A. WATTERSON, Models for the logarithmic species abundance distributions, *Theoret. Population Biol.* **6** (1974), 217–250.
60. G. A. WATTERSON, The stationary distribution of the infinitely many alleles diffusion model, *J. Appl. Probab.* **13** (1976), 639–651.
61. H. S. WILF, "Generatingfunctionology," Academic Press, San Diego, CA, 1990.
62. J. C. HANSEN AND E. SCHMUTZ, How random is the characteristic polynomial of a random matrix? *Math. Proc. Camb. Phil. Soc.* **114** (1993), 507–515.
63. V. F. KOLCHIN, A problem of the allocation of particles in cells and random mappings, *Theory. Probab. Appl.* **21** (1976), 48–63.

Printed by Catherine Press, Ltd., Tempelhof 41, B-8000 Brugge, Belgium