Universality of asymptotically Ewens measures on partitions[*]

James Y. Zhao[†]

Abstract

We give a criterion for functionals of partitions to converge to a universal limit under a class of measures that "behaves like" the Ewens measure. Various limit theorems for the Ewens measure, most notably the Poisson-Dirichlet limit for the longest parts, the functional central limit theorem for the number of parts, and the Erdős-Turán limit for the product of parts, extend to these asymptotically Ewens measures as easy corollaries. Our major contributions are: (1) extending the classes of measures for which these limit theorems hold; (2) characterising universality by an intuitive and easily-checked criterion; and (3) providing a new and much shorter proof of the limit theorems by taking advantage of the Feller coupling.

1 Introduction

Let $\mathcal{P}_n$ be the partitions of $n \in \mathbb{N}$, which we represent by $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathcal{P}_n$, where α_i is the number of parts of size i, so that $\alpha_1 + 2\alpha_2 + \cdots + n\alpha_n = n$. Let

$$\mathbb{P}_n^\theta(\alpha) = \frac{n!}{\theta(\theta+1)\cdots(\theta+n-1)} \prod_{i=1}^n \frac{\theta^{\alpha_i}}{i^{\alpha_i}\alpha_i!} \tag{1.1}$$

be the Ewens (16) measure on $\mathcal{P}_n$ with parameter $\theta > 0$. The Ewens measure should be interpreted as a random permutation weighted by the number of cycles, under which the α_i are asymptotically independent Poisson with parameter θ/i. Any other probability measure $\mathbb{P}_n$ on $\mathcal{P}_n$ can be written as a perturbation of the Ewens measure with parameter $\theta > 0$, that is,

$$\mathbb{P}_n(\alpha) = \mathbb{P}_n^{\theta,\eta}(\alpha) = \frac{\eta(\alpha)}{Z_n^{\theta,\eta}} \prod_{i=1}^n \frac{\theta^{\alpha_i}}{i^{\alpha_i}\alpha_i!}, \tag{1.2}$$

where $\eta : \mathcal{P}_n \to \mathbb{R}^+$ is (any multiple of) the Radon-Nikodym derivative of $\mathbb{P}_n$ with respect to the Ewens measure $\mathbb{P}_n^\theta$, and $Z_n^{\theta,\eta}$ is a normalising constant. Since n, θ and η uniquely determine the measure, it will be convenient to use the notation $\mathbb{P}_n^{\theta,\eta}$ to specify it.

[†]Stanford University, USA. E-mail: jyzhao@math.stanford.edu

We will study partitions by their preimages under the (17) coupling, which we define as the map $\{0,1\}^{\mathbb{N}} \to \mathcal{P}_n$ which sends a binary sequence $(\xi_1, \xi_2, \ldots) \in \{0,1\}^{\mathbb{N}}$ to a partition $(\alpha_1(n), \ldots, \alpha_n(n)) \in \mathcal{P}_n$, where

$$\alpha_i(n) = \sum_{j=1}^{n-i} \xi_j (1 - \xi_{j+1}) \cdots (1 - \xi_{j+i-1}) \xi_{i+j} + \xi_{n-i+1}(1 - \xi_{n-i+2}) \cdots (1 - \xi_n) \qquad (1.3)$$

is the number of gaps between 1s of length $i - 1$ in the string $(\xi_1, \xi_2, \ldots, \xi_n, 1)$. It was shown by (1) that if $\mathbb{P}_F^\theta$ is a measure on $\{0,1\}^{\mathbb{N}}$ under which the ξ_i are independent Bernoulli variables with probability of success $\theta/(\theta + i - 1)$, then the measure induced on $\mathcal{P}_n$ by the Feller coupling is the Ewens measure $\mathbb{P}_n^\theta$. In particular, we can embed $\mathbb{P}_n^\theta$ within $\mathbb{P}_F^\theta$ so that $\alpha_i = \alpha_i(n)$, simultaneously for all $n \in \mathbb{N}$.

For any $m, n \in \mathbb{N}$, $(\alpha_1(n), \ldots, \alpha_m(n))$ is a partition of some integer $k \le n$. Then, for any sequence of measures $\mathbb{P}_n^{\theta, \eta}$ on $\mathcal{P}_n$, $n \in \mathbb{N}$, define the *weights*

$$\eta_{n,m} = \eta(\alpha_1(n), \ldots, \alpha_m(n)), \qquad (1.4)$$

where η is the (unscaled) Radon-Nikodym derivative of $\mathbb{P}_k^{\theta, \eta}$ with respect to $\mathbb{P}_k^\theta$, for the appropriate choice of $k = \alpha_1(n) + 2\alpha_2(n) + \cdots + m\alpha_m(n)$. If we interpret the Radon-Nikodym derivative as weights, then $\eta_{n,m}$ should be interpreted as the weight of a partition of n considering only parts of size at most m. Note that since the weights are defined in terms of the $\alpha_i(n)$, they are random variables under the Feller measure $\mathbb{P}_F^\theta$.

We call a sequence of measures $\mathbb{P}_n^{\theta, \eta}$ *asymptotically Ewens* when:

1. The limits $\lim_{n \to \infty} \eta_{n,n}$ and $\lim_{m \to \infty} \lim_{n \to \infty} \eta_{n,m}$ exist and agree in $L^1(\mathbb{P}_F^\theta)$; and
2. The common limit η_∞ satisfies $0 < \mathbb{E}_F^\theta[\eta_\infty] < \infty$.

These measures are important in a variety of applications and include many extensively studied measures as special cases. Essentially, they generalise the *logarithmic combinatorial structures* of Arratia, Barbour and Tavaré (4) by removing the conditioning relation. We will prove this claim in Section 4, while Section 2 contains background on these measures and the associated limit theorems.

We conclude the introduction by stating our main theorem, the proof of which is presented in Section 3, and showing how the Ewens measure limit theorems extend to asymptotically Ewens measures as easy corollaries.

Theorem 1.1. *Suppose $X_n : \mathcal{P}_n \to (\mathcal{X}, || \cdot ||)$ is a sequence of deterministic functions on partitions of n (and therefore they are functions on $\{0,1\}^n$ by the Feller coupling), taking values in some normed space, such that for any fixed $d \in \mathbb{N}$,*

$$\lim_{n \to \infty} \max_{\xi_1, \ldots, \xi_n} \left\| X_n(\xi_1, \ldots, \xi_n) - X_n(1, \ldots, 1, \xi_{d+1}, \ldots, \xi_n) \right\| = 0. \qquad (1.5)$$

If $X_n \overset{d}{\longrightarrow} X$ under the Ewens measure with parameter θ, then $X_n \overset{d}{\longrightarrow} X$ under any asymptotically Ewens measure with parameter θ.

The criterion (1.5) has a very intuitive interpretation: if the functional does not depend on the first finitely many Feller variables, then it has a universal limit for any asymptotically Ewens measure, depending only on the parameter θ.

Corollary 1.2 (Poisson-Dirichlet)**.** *Let $L_{n,k}$ be the length of the kth longest part of a partition of n, and let $L_n = (L_{n,1}, L_{n,2}, \ldots)$. Under any asymptotically Ewens measure with parameter θ, L_n/n converges in distribution in $L^1(\mathbb{N})$ to $PD(\theta)$, the Poisson-Dirchlet measure with parameter θ (27, 28).*

Proof. Consider strings $(1, \ldots, 1, \xi_{i+1}, \ldots, \xi_n)$, for $i = d, d-1, \ldots, 0$. For each decrement of i, either the partition is unchanged, or a 1-part is deleted and some other part length increases by 1, which does not change the order of parts and thus changes L_n by at most 2 in the $L^1(\mathbb{N})$ norm. Since there are d decrements from $i = d$ to $i = 0$,

$$\max_{\xi_1, \ldots, \xi_n} \left\| L_n(\xi_1, \ldots, \xi_n) - L_n(1, \ldots, 1, \xi_{d+1}, \ldots, \xi_n) \right\|_1 \leq 2d. \tag{1.6}$$

Hence, $X_n = L_n/n$ satisfies the conditions of Theorem 1.1 applied to the normed space $L^1(\mathbb{N})$, so the result follows from the Poisson-Dirichlet limit for the Ewens measure (27, 28, 37). $\qquad\square$

Corollary 1.3 (CLT). *Let $\nu_{n,t} = \alpha_1 + \cdots + \alpha_{\lfloor n^t \rfloor}$, $0 \leq t \leq 1$, be the number of parts of size at most n^t in a partition of n. Under any asymptotically Ewens measure with parameter θ, $(\nu_{n,t} - \theta t \log n)/\sqrt{\theta \log n}$ converges in distribution in $\mathcal{D}[0,1]$, the Skorohod space of right-continuous left-limit functions on $[0,1]$, to the standard Brownian motion W_t.*

Proof. Consider strings $(1, \ldots, 1, \xi_{i+1}, \ldots, \xi_n)$, for $i = d, d-1, \ldots, 0$. For each decrement of i, either the partition is unchanged, or a 1-part is deleted and some other part length increases by 1, which changes $\nu_{n,t}$ by at most 1. Since there are d decrements,

$$\max_{\xi_1, \ldots, \xi_n} \left| \nu_{n,t}(\xi_1, \ldots, \xi_n) - \nu_{n,t}(1, \ldots, 1, \xi_{d+1}, \ldots, \xi_n) \right| \leq d. \tag{1.7}$$

Letting $\nu_n \in \mathcal{D}[0,1]$ be the sample path of $\nu_{n,t}$ for $0 \leq t \leq 1$,

$$\max_{\xi_1, \ldots, \xi_n} \left\| \nu_n(\xi_1, \ldots, \xi_n) - \nu_n(1, \ldots, 1, \xi_{d+1}, \ldots, \xi_n) \right\|_\infty \leq d. \tag{1.8}$$

Hence, $X_n = (\nu_n - \theta t \log n)/\sqrt{\theta \log n}$ satisfies the conditions of Theorem 1.1 applied to the normed space $\mathcal{D}[0,1]$, so the result follows from the functional central limit theorem for the Ewens measure (10, 24). $\qquad\square$

Corollary 1.4 (Erdős-Turán). *Let $O_{n,t} = lcm\{i \leq n^t : \alpha_i > 0\}$ be the least common multiple (or product) of the parts of size at most n^t in a partition of n. Under any asymptotically Ewens measure with parameter θ, $\left(\log O_{n,t} - \theta t^2 (\log n)^2/2 \right)/\sqrt{\theta (\log n)^3/3}$ converges in distribution in $\mathcal{D}[0,1]$ to W_{t^3}, where W_t is the standard Brownian motion.*

Proof. The partitions given by $(\xi_1, \ldots, \xi_n)$ and $(1, \ldots, 1, \xi_{d+1}, \ldots, \xi_n)$ differ in at most d parts. Since deleting a part of size ℓ, adding a part of size m, or replacing ℓ by m changes the logarithm of the least common multiple (or product) by at most $\log \max(\ell, m)$,

$$\max_{\xi_1, \ldots, \xi_n} \left| \log O_{n,t}(\xi_1, \ldots, \xi_n) - \log O_{n,t}(1, \ldots, 1, \xi_{d+1}, \ldots, \xi_n) \right| \leq d \log n^t. \tag{1.9}$$

Then, letting $O_n \in \mathcal{D}[0,1]$ be the sample path of $O_{n,t}$ for $0 \leq t \leq 1$,

$$\max_{\xi_1, \ldots, \xi_n} \left\| \log O_n(\xi_1, \ldots, \xi_n) - \log O_n(1, \ldots, 1, \xi_{d+1}, \ldots, \xi_n) \right\|_\infty \leq d \log n. \tag{1.10}$$

Hence, $X_n = \left(\log O_n - \theta t^2 (\log n)^2/2 \right)/\sqrt{\theta (\log n)^3/3}$ satisfies the conditions of Theorem 1.1, so the result follows from the functional Erdős-Turán limit for the Ewens measure (14, 15, 6). The same result holds for the product of parts. $\qquad\square$

2 Background

2.1 Measures on Partitions

Measures on partitions arise naturally from combinatorial objects which consist of components of various sizes. For example, cycles of a random permutation, irreducible factors of a random polynomial or Jordan blocks of a random matrix are all described by measures on partitions when one cares about only the sizes of those components. More such examples are given by Arratia, Barbour and Tavaré (2, 5).

The most basic example is a uniformly random permutation, which corresponds to our $\mathbb{P}_n^{\theta,\eta}$ when $\theta = 1$ and $\eta = 1$ identically, and has been the subject of extensive study since the 19th century. The generalisation to $\theta > 0$, still with $\eta = 1$ identically, was introduced by Ewens (16) to model propagation of alleles in population genetics, and represents a random permutation weighted by the number of cycles, or perhaps more intuitively, a permutation formed in a Markov process where cycles are added at a rate θ (26).

The further generalisation to weights $\eta(\alpha) = \zeta_1(\alpha_1)\zeta_2(\alpha_2)\cdots\zeta_n(\alpha_n)$, with certain limiting conditions on the ζ_i, was introduced by Arratia, Barbour and Tavaré (4). Their *logarithmic combinatorial structures* generalise the *decomposable combinatorial structures* of Flajolet and Soria (18), which are measures on partitions induced by the uniform measure on families of combinatorial objects determined by sizes of components. We will prove in Section 4 that logarithmic combinatorial structures are exactly asymptotically Ewens measures with weight function in the form above.

An important subclass of logarithmic combinatorial structures are measures with weights $\eta(\alpha) = \zeta_1^{\alpha_1}\zeta_2^{\alpha_2}\cdots\zeta_n^{\alpha_n}$, where ζ_i are constants with $\sum_i |\zeta_i - 1|/i < \infty$. This is the asymptotically Ewens case of the *multiplicatively weighted measures* of Betz and Ueltschi's (2009) combinatorial model for Bose-Einstein condensation. In this model, n points are determined by both their positions and a permutation that describes their trajectories, with the energy function (Hamiltonian) of the system depending particularly on the presence of long cycles. Betz, Ueltschi and Velenik (8) analysed the asymptotic behaviour of the cycle lengths and the normalising constant (partition function) in the asymptotically Ewens case, as well as two other cases where the weights diverge. Ercolani and Ueltschi (13) continued this work, extending the cycle length analysis to a much more comprehensive set of parameter regions.

The further specialisation $\zeta_i = (1-q^i)/(1-t^i)$, for parameters $0 < q < 1$ and $0 < t < 1$, are the *Macdonald polynomial measures* of Diaconis and Ram's (2010) probabilistic interpretation of Macdonald polynomials. This measure is the stationary distribution of a Markov chain on partitions which corresponds to the Macdonald operator on symmetric polynomials.

Some examples of asympotically Ewens measures which are not logarithmic combinatorial structures include many restricted permutations, such as permutations with more even cycles than odd cycles, permutations whose squares have fixed points, or permutations with an even number of cycles; for any underlying measure that is a logarithmic combinatorial structure, these restrictions are asymptotically Ewens.

There are some examples of measures, such as the a-riffle shuffle measures of Diaconis, McGrath and Pitman (11), and the restricted permutations studied by Lugo (30), which are not asymptotically Ewens by our current definition, but behave similarly in the sense that they follow the Poisson-Dirichlet limit, as discussed in more detail in Section 2.2. Generalising our measures to include these examples would be an interesting direction for future work.

Finally, there are many measures which are not asymptotically Ewens in any sense, such as the uniform measure on partitions, Pitman's (1992) two-parameter family of

measures (although this family includes the Ewens measure as a special case), and the induced measure on partitions from various measures on permutations such as the Plancherel measure and its generalisation, the Schur measures of Okounkov (31).

2.2 The Poisson-Dirichlet Limit

The parameter θ in the Ewens measure corresponds to a rate of formation of new parts (26); indeed, θ is the global rate of mutation in Ewens' (1972) original genetic model. This insight extends to the asymptotically Ewens case, where the rate of formation of new parts, appropriately scaled, converges to the parameter θ. This is the intuitive reason why we expect the limit theorems to be universal: with new parts being added at the same rate, the relative sizes of parts should behave similarly.

The key notion that captures this behaviour is the Poisson-Dirichlet limit: the largest parts, normalised by $\frac{1}{n}$, converge in distribution on $L^1(\mathbb{N})$ to a limit known as the Poisson-Dirichlet measure with parameter θ. This was first studied by Kingman (27), who described it via the Dirichlet distribution on $L^1(\mathbb{N})$, and Watterson (37), who found an explicit density.

Historically, Golomb (20) was the first to calculate the expected value of the longest cycle of a uniformly random permutation, Shepp and Lloyd (34) found the distributions of the kth longest cycles, and Kingman (27, 28) and Watterson (37) found the joint distribution of longest cycles under the Ewens measure. Hansen (25) proved the Poisson-Dirichlet limit for decomposable combinatorial structures, while the extension to logarithmic combinatorial structures was made by Arratia, Barbour and Tavaré (3).

We further generalise the Poisson-Dirichlet limit to asymptotically Ewens measures. However, there are still many other measures which satisfy the Poisson-Dirichlet limit, such as the largest prime factors of a random integer (29), the a-riffle shuffle measures of Diaconis, McGrath and Pitman (11), and the restricted permutations studied by Lugo (30). We expect there to be a fundamental reason why we observe the same limit in these cases, although what that reason should be is currently beyond our grasp.

The Poisson-Dirichlet distribution has a two-parameter generalisation (33), which is the limit of the ordered parts of Pitman's (1992) two-parameter family of measures on partitions. Since Pitman's measures are a direct generalisation of the Ewens measure, it seems plausible that our result could be extended in this direction.

2.3 Other Limit Theorems

It is classical that the number of i-cycles in a uniformly random permutation are asymptotically independent Poisson with parameter $1/i$. The usual proof is by generating functions, which also works for multiplicative weights $\eta(\alpha) = \prod_i \zeta_i^{\alpha_i}$, where the parameter becomes $\theta \zeta_i / i$. Such a proof is given by Betz, Ueltschi and Velenik (8); see also the book of Arratia, Barbour and Tavaré (5) for a thorough treatment of generating function techniques in this setting. For logarithmic combinatorial structures, Arratia, Barbour and Tavaré (4) prove that the number of parts of size i are asymptotically independent, although this result is in some sense one of the defining assumptions of logarithmic combinatorial structures.

The total number of parts was first studied by Goncharov (21), who found a central limit theorem for the number of cycles in a uniformly random permutation. The functional central limit theorem as seen in Corollary 1.3 was first proved by DeLaurentis and Pittel (10) for the uniform permutation case, and extended to the Ewens measure by Hansen (24). Flajolet and Soria (18) proved a central limit theorem for decomposable combinatorial structures, and the two theorems were unified by Arratia, Barbour and Tavaré (4), who proved a functional central limit theorem for logarithmic combinatorial

structures.

The asymptotic moments of the smallest parts were derived by Shepp and Lloyd (34). They have not been the subject of extensive study; some facts which are known about them are listed in the book of Arratia, Barbour and Tavaré (5). The shortest parts depend heavily on the first few Feller variables, and do not fall under the scope of our universality theorem.

It is also possible to canonically order the parts under the Ewens measure by the order in which they appear in the *Chinese restaurant coupling*. This limit is called the Griffiths-Engen-McCloskey measure (23), and behaves similarly to the longest parts; in fact its order statistics exactly follow the Poisson-Dirichlet measure. This limit theorem does not generalise to asymptotically Ewens measures due to the a lack of a canonical order.

The lowest common multiple of parts is a statistic of interest when the partition is induced by a permutation, as it is the group order of the permutation. Erdős and Turán (14, 15) found a central limit theorem for the logarithm of the lowest commmon multiple in the uniform permutation case. Their proof, and all subsequent proofs, worked via the product of parts, in particular showing that the product satisfies the same central limit theorem. The generalisation to the Ewens measure was proved by Barbour and Tavaré (6), who also proved the functional form in Corollary 1.4, while the extension to logarithmic combinatorial structures was made by Arratia, Barbour and Tavaré (4).

2.4 Universality

For logarithmic combinatorial structures, Arratia, Barbour and Tavaré (4) prove

$$\lim_{n\to\infty} \big|\big| \mathcal{L}_n^{\theta,\eta}(\alpha_{d_n}, \ldots, \alpha_n) - \mathcal{L}_n^\theta(\alpha_{d_n}, \ldots, \alpha_n) \big|\big|_{TV} = 0, \tag{2.1}$$

where $\mathcal{L}_n^{\theta,\eta}$ and $\mathcal{L}_n^\theta$ are the laws of $(\alpha_{d_n}, \ldots, \alpha_n)$ under $\mathbb{P}_n^{\theta,\eta}$ and $\mathbb{P}_n^\theta$ respectively, and d_n is any sequence satisfying $d_n \to \infty$ and $d_n/n \to 0$. They also proved the other limit theorems above, but most of their proofs ran in parallel to (2.1) instead of directly invoking it.

This left open the question of a simple criterion to determine whether a functional is universal, as well as the question of whether asymptotic independence of the α_i is a necessary condition. Our model answers both of these questions, removing the requirement for the α_i to be asymptotically independent, and also giving an easily-checked criterion for universality.

3 Proof of Main Theorem

Suppose $\mathbb{P}_n^{\theta,\eta}$ is an asymptotically Ewens measure with parameter $\theta > 0$ and X_n satisfies the conditions of Theorem 1.1. It suffices (9) to prove that for any bounded, uniformly continuous function $f : (\mathcal{X}, ||\cdot||) \to (\mathbb{R}, |\cdot|)$, we have $\mathbb{E}_n^{\theta,\eta}\big[f(X_n)\big] \sim \mathbb{E}_n^\theta\big[f(X_n)\big]$ as $n \to \infty$, where $a_n \sim b_n$ means $a_n/b_n \to 1$.

Since $\eta_{n,n}$ is (a multiple of) the Radon-Nikodym derivative,

$$\mathbb{E}_n^{\theta,\eta}\big[f(X_n)\big] = \frac{Z_n^\theta}{Z_n^{\theta,\eta}} \mathbb{E}_F^\theta\big[\eta_{n,n} f(X_n)\big]. \tag{3.1}$$

Since f is bounded and η is asymptotically Ewens, $(\eta_{n,n} - \eta_{n,m})f(X_n) \to 0$ in $L^1(\mathbb{P}_F^\theta)$ as $n \to \infty$ then $m \to \infty$, hence

$$\lim_{n\to\infty} \mathbb{E}_F^\theta\big[\eta_{n,n} f(X_n)\big] = \lim_{m\to\infty} \lim_{n\to\infty} \mathbb{E}_F^\theta\big[\eta_{n,m} f(X_n)\big]. \tag{3.2}$$

For any integer $d < n$, the event $\{\alpha_i(n) \neq \alpha_i(d)\}$ can be decomposed into a union of two events: either $(\xi_{d-i+1}, \ldots, \xi_{d+1}) = (1, 0, \ldots, 0)$, with probability

$$\mathbb{E}_F^\theta\big[\xi_{d-i+1}(1 - \xi_{d-i+2}) \cdots (1 - \xi_{d+1})\big] \leq \mathbb{E}_F^\theta\big[\xi_{d-i+1}\big] = \tfrac{\theta}{d-i+\theta}, \tag{3.3}$$

or $(\xi_{d-i+2}, \ldots, \xi_n, 1)$ contains a substring $(1, 0, \ldots, 0, 1)$ with $i-1$ zeroes, with probability bounded above by

$$\sum_{k=2}^{n-d} \mathbb{E}_F^\theta\big[\xi_{d-i+k}\xi_{d+k}\big] + \mathbb{E}_F^\theta\big[\xi_{n-i+1}\big] = \sum_{k=2}^{n-d} \tfrac{\theta}{d-i+k-1+\theta}\tfrac{\theta}{d+k-1+\theta} + \tfrac{\theta}{n-i+\theta}. \tag{3.4}$$

Using telescoping series, we have $\mathbb{P}_F^\theta[\alpha_i(n) \neq \alpha_i(d)] \leq (2\theta + \theta^2)/(d - i + \theta)$. Hence, the event $E \equiv \{\exists i \leq m : \alpha_i(n) \neq \alpha_i(d)\}$ has $\mathbb{P}_F^\theta$ probability at most $(2\theta + \theta^2)m/(d - m + \theta)$, which converges to 0 as $d \to \infty$. Note that

$$\eta_{n,m}f(X_n) = \eta_{d,m}f(X_n) + (\eta_{n,m} - \eta_{d,m})f(X_n)\mathbb{1}_E. \tag{3.5}$$

Since $\eta_{n,m}$ and $\eta_{d,m}$ converge in $L^1(\mathbb{P}_F^\theta)$, they are uniformly integrable. Since $\mathbb{1}_E \to 0$ in probability in the limit $n \to \infty$ then $d \to \infty$ then $m \to \infty$, $(\eta_{n,m} - \eta_{d,m})f(X_n)\mathbb{1}_E \to 0$ in probability and therefore in $L^1(\mathbb{P}_F^\theta)$, hence

$$\lim_{m \to \infty} \lim_{n \to \infty} \mathbb{E}_F^\theta\big[\eta_{n,m}f(X_n)\big] = \lim_{m \to \infty} \lim_{d \to \infty} \lim_{n \to \infty} \mathbb{E}_F^\theta\big[\eta_{d,m}f(X_n)\big]. \tag{3.6}$$

Uniform continuity of f and (1.5) imply $\mathbb{E}_F^\theta\big[f(X_n)\big|\xi_1, \ldots, \xi_d\big] \sim \mathbb{E}_F^\theta\big[f(X_n)\big]$ uniformly as $n \to \infty$. Since f is bounded and $\eta_{d,m}$ is uniformly integrable,

$$\lim_{n \to \infty} \mathbb{E}_F^\theta\big[\eta_{d,m}f(X_n)\big] = \lim_{n \to \infty} \mathbb{E}_F^\theta\big[\eta_{d,m}\mathbb{E}_F^\theta\big[f(X_n)\big|\xi_1, \ldots, \xi_d\big]\big] \tag{3.7}$$
$$= \lim_{n \to \infty} \mathbb{E}_F^\theta\big[\eta_{d,m}\big]\mathbb{E}_F^\theta\big[f(X_n)\big].$$

Finally, as $n \to \infty$ then $d \to \infty$ then $m \to \infty$,

$$\mathbb{E}_F^\theta\big[\eta_{d,m}\big] \sim \mathbb{E}_F^\theta\big[\eta_{n,n}\big] = \sum_{\alpha \in \mathcal{P}_n} \frac{\eta(\alpha)}{Z_n^\theta} \prod_{i=1}^n \frac{\theta^{\alpha_i}}{i^{\alpha_i}\alpha_i!} = \frac{Z_n^{\theta,\eta}}{Z_n^\theta}. \tag{3.8}$$

We have used the assumption that this quantity has a positive, finite limit. The chain of asymptotic equivalences in (3.1), (3.2), (3.6), (3.7) and (3.8) completes the proof. $\quad\square$

4 Logarithmic Combinatorial Structures

A *uniform logarithmic combinatorial structure* (4) is a sequence of measures $\mathbb{P}_n$ on $\mathcal{P}_n$, $n \in \mathbb{N}$, such that:

1. (*Conditioning Relation*) There is some sequence of independent random variables $Y_1, Y_2, \ldots$ such that $\mathbb{P}_n(\alpha) = \mathbb{P}\big[\forall i \leq n, Y_i = \alpha_i \big| \sum_{i \leq n} iY_i = n\big]$; and

2. (*Uniform Logarithmic Condition*) Each variable Y_i satisfies $\big|i\mathbb{P}[Y_i = 1] - \theta\big| \leq e_i$ and $i\mathbb{P}[Y_i = \ell] \leq e_i c_\ell$ for $\ell \geq 2$, where e_i and c_ℓ are vanishing sequences such that e_i/i and ℓc_ℓ are summable.

Lemma 4.1. *Any uniform logarithmic combinatorial structure $\mathbb{P}_n$ can be written as $\mathbb{P}_n^{\theta,\eta}$ for $\eta(\alpha) = \prod_i \zeta_i(\alpha_i)$, where $\zeta_i(0) = 1$, $\big|\zeta_i(1) - 1\big| \leq e_i/\theta$ and $\zeta_i(\ell) \leq i^{\ell-1}\ell!e_i c_\ell/\theta^\ell$ for $\ell \geq 2$. As before, e_i and c_ℓ are vanishing sequences with e_i/i and ℓc_ℓ summable. Additionally, with $c_0 = 0$ and $c_1 = 1$, we can insist that for each i, $\mathbb{1}_{\{\ell \leq 1\}} + i^{\ell-1}\ell!e_i c_\ell/\theta^\ell$ is monotonic increasing in $\ell \geq 0$.*

Proof. Let $p_i = \mathbb{P}[Y_i = 0]$, and let $\zeta_i(\ell) = \mathbb{P}[Y_i = \ell](i/\theta)^\ell \ell!/p_i$. By the conditioning relation, $\mathbb{P}_n = \mathbb{P}_n^{\theta, \eta}$, $\zeta_i(0) = 1$. By the uniform logarithmic condition, $|p_i\zeta_i(1) - 1| \le e_i/\theta$ and $\zeta_i(\ell) \le i^{\ell-1}\ell!e_ic_\ell/p_i\theta^\ell$ for $\ell \ge 2$.

Note that $1 - p_i = \sum_{\ell \ge 1} \mathbb{P}[Y_i = \ell] \le (\theta + Ce_i)/i$, where $C = \sum_\ell c_\ell < \infty$. In particular, $p_i \to 1$ as $i \to \infty$, so

$$|\zeta_i(1) - 1| \le (1 - p_i)\zeta_i(1) + |p_i\zeta_i(1) - 1| = O(\tfrac{1}{i}) + O(e_i). \tag{4.1}$$

Thus, $e_i' = \max\left(e_i/p_i, \theta|\zeta_i(1) - 1|, 1/i\right)$ is a vanishing sequence such that e_i'/i is summable, and we have the required inequalities $|\zeta_i(1) - 1| \le e_i'/\theta$ and $\zeta_i(\ell) \le i^{\ell-1}\ell!e_i'c_\ell/\theta^\ell$ for $\ell \ge 2$. It remains to replace c_ℓ by c_ℓ' satisfying the desired monotonicity condition.

Let $c_0' = 0$, $c_1' = 1$, and $c_2' = \max\left(c_2, \sup_i(\theta^2 + \theta e_i')/(2ie_i')\right)$, which is finite since $e_i' \ge 1/i$. For $3 \le \ell \le 2\theta$, let $c_\ell' = \max(c_\ell, \theta c_{\ell-1}')$, and for $\ell > 2\theta$, let $c_\ell' = \max\left(c_\ell, \frac{1}{2}c_{\ell-1}'\right)$. There are $\ell_0 = \lfloor 2\theta \rfloor < \ell_1 < \ell_2 < \cdots$ such that for $\ell_0 \le \ell < \ell_1$, $c_\ell' = 2^{\ell_0 - \ell}c_{\ell_0}'$, and for $\ell_k \le \ell < \ell_{k+1}$, $c_\ell' = 2^{\ell_k - \ell}c_{\ell_k}$, hence

$$\sum_{\ell \ge 2\theta} \ell c_\ell' \le \left(\sum_{j \ge 0}(1 + j)2^{-j}\right)\left(\ell_0 c_{\ell_0}' + \sum_{k \ge 1} \ell_k c_{\ell_k}\right) < \infty. \tag{4.2}$$

We also have $c_\ell \le c_\ell'$ and the monotonicity condition by construction. $\qquad\square$

Theorem 4.2. *Uniform logarithmic combinatorial structures are asymptotically Ewens.*

Proof. Let $\alpha_i(\infty) = \sum_{j=1}^\infty \xi_j(1 - \xi_{j+1})\cdots(1 - \xi_{j+i-1})\xi_{i+j}$ be the number of gaps between 1s of size $i - 1$ in the string $(\xi_1, \xi_2, \ldots)$, which are independent and Poisson with parameter θ/i (1).

We will prove $\eta_{n,m}$ and $\eta_{n,n}$ are uniformly integrable and converge in probability to $\eta_\infty = \prod_i \zeta_i(\alpha_i(\infty))$, which has positive and finite expectation, by defining several intermediate weights and proving a sequence of asymptotic equivalences between them.

- Let $\eta_{\infty,m} = \prod_{i=1}^m \zeta_i(\alpha_i(\infty))$.
- Let $\eta_{n,m}^+$ be weights for $\zeta_i^+(\ell) = \mathbb{1}_{\{\ell \le 1\}} + i^{\ell-1}\ell!e_ic_\ell/\theta^\ell$, and let $\eta_{n,m}^-$ be weights for $\zeta^-(0) = 1$, $\zeta^-(1) = \max(1 - e_i/\theta, 0)$ and $\zeta^-(\ell) = 0$ for $\ell \ge 2$. By Lemma 4.1, we have $\eta_{n,m}^- \le \eta_{n,m} \le \eta_{n,m}^+$, and $\eta_{n,m}^+$ is monotonic increasing in each $\alpha_i(n)$.
- Let $\tilde{\eta}_{n,m} = \prod_{i=1}^m \zeta_i(\tilde{\alpha}_i(n))$ be the weight given by replacing $\alpha_i(n)$ by $\tilde{\alpha}_i(n)$, where $\tilde{\alpha}_i(n)$ is the number of gaps between 1s of length $i - 1$ in the string $(\xi_1, \ldots, \xi_n, 0)$.

For brevity, all limits are implicitly $n \to \infty$ with m fixed, then $m \to \infty$. We also omit writing the measure explicitly; the only measure used is $\mathbb{P}_F^\theta$.

Step 1: Positivity and finiteness of limit. Let $C = \sum_\ell c_\ell < \infty$. Since the $\alpha_i(\infty)$, $i \in \mathbb{N}$, are independent and Poisson with parameter θ/i, we can calculate

$$\mathbb{E}[\eta_\infty^+] = \prod_{i=1}^\infty \mathbb{E}[\zeta_i^+(\alpha_i(\infty))] = \prod_{i=1}^\infty e^{-\frac{\theta}{i}}\left(1 + \frac{\theta + Ce_i}{i}\right) \tag{4.3}$$

$$\le \prod_{i=1}^\infty \left(1 + \frac{Ce_i}{i}\right) = \exp\sum_{i=1}^\infty \log\left(1 + \frac{Ce_i}{i}\right). \tag{4.4}$$

We have used the inequality $1 + x + y \le e^x(1 + y)$. Since e_i/i is summable, so is the series above, hence $\mathbb{E}[\eta_\infty^+] < \infty$. A similar calculation shows that $\mathbb{E}[\eta_\infty^-] > 0$. Since $\eta_\infty^- \le \eta_\infty \le \eta_\infty^+$, it follows that $0 < \mathbb{E}[\eta_\infty] < \infty$.

Step 2: Convergence in probability. We will show that

$$\eta_{n,n} \stackrel{p}{\sim} \tilde{\eta}_{n,n} \stackrel{p}{\sim} \tilde{\eta}_{n,m} \stackrel{p}{\sim} \eta_{n,m} \stackrel{p}{\sim} \eta_{\infty,m} \stackrel{p}{\sim} \eta_\infty, \tag{4.5}$$

where $X_n \overset{p}{\sim} Y_n$ means $X_n/Y_n \overset{p}{\to} 1$. The relation $\overset{p}{\sim}$ is clearly transitive; also observe that $\eta_{n,n} \overset{p}{\sim} \eta_\infty$ implies $\eta_{n,n} \overset{p}{\to} \eta_\infty$, since for any $\epsilon > 0$, there exists some M such that $\mathbb{P}[\eta_\infty > M] < \frac{\epsilon}{2}$, and there exists some N depending on M such that for all $n > N$, $\mathbb{P}[|\eta_{n,n}/\eta_\infty - 1| > \frac{\epsilon}{M}] < \frac{\epsilon}{2}$.

Observe that for each fixed n, $\alpha_i(n)$ and $\tilde{\alpha}_i(n)$ are equal for all i except one value $i^*(n)$ where $\alpha_{i^*(n)}(n) = \tilde{\alpha}_{i^*(n)}(n) + 1$, and given $i^*(n) = i$, $\tilde{\alpha}_i(n)$ has the same law as an independent copy of $\alpha_i(n-i)$. Furthermore, $\alpha_i(n-i) \neq 0$ implies either $\alpha_i(\infty) \neq 0$, with probability $1 - e^{-\theta/i}$, or $i^*(n-i) = i$, with probability $\frac{1}{n-i}$. Hence,

$$\mathbb{P}[\tilde{\alpha}_{i^*(n)}(n) \neq 0] = \frac{1}{n}\sum_{i=1}^n \mathbb{P}[\alpha_i(n-i) \neq 0] \leq \frac{1}{n}\sum_{i=1}^n \left(1 - e^{-\theta/i} + \frac{1}{n-i}\right). \tag{4.6}$$

This expression vanishes as $n \to \infty$, hence $\tilde{\alpha}_{i^*(n)}(n) = 0$ with high probability (this means that the part that differs between α and $\tilde{\alpha}$ does not appear elsewhere in α).

Since $i^*(n)$ is the size of the final gap between 1s that appears in $(\xi_1, \ldots, \xi_n, 1)$,

$$\mathbb{P}[i^*(n) = k] = \mathbb{P}[\xi_n = 0, \ldots, \xi_{n-k+2} = 0, \xi_{n-k+1} = 1] \tag{4.7}$$
$$= \frac{n-1}{\theta+n-1}\frac{n-2}{\theta+n-2}\cdots\frac{n-k+1}{\theta+n-k+1}\frac{\theta}{\theta+n-k}. \tag{4.8}$$

This vanishes for fixed k as $n \to \infty$, so $i^*(n) \to \infty$ in probability. Since e_i vanishes as $i \to \infty$, for any fixed $\epsilon > 0$, $e_{i^*(n)} < \epsilon$ with high probability. Hence, with high probability,

$$\left|\frac{\eta_{n,n}}{\tilde{\eta}_{n,n}} - 1\right| = \left|\frac{\zeta_{i^*(n)}(\tilde{\alpha}_{i^*(n)}(n) + 1)}{\zeta_{i^*(n)}(\tilde{\alpha}_{i^*(n)}(n))} - 1\right| = \left|\frac{\zeta_{i^*(n)}(1)}{\zeta_{i^*(n)}(0)} - 1\right| \leq \frac{e_{i^*(n)}}{\theta} \leq \frac{\epsilon}{\theta}. \tag{4.9}$$

This proves that $\eta_{n,n} \overset{p}{\sim} \tilde{\eta}_{n,n}$.

For $\tilde{\eta}_{n,n} \overset{p}{\sim} \tilde{\eta}_{n,m}$, observe that with high probability, $\alpha_i(\infty) \leq 1$ for all $i > m$. Picking m so that $e_i \leq \frac{\theta}{2}$ for all $i > m$, and using the inequality $|\log(1-x)| \leq 2\log(1+x)$ for $x \geq \frac{1}{2}$, with high probability,

$$\left|\log\frac{\tilde{\eta}_{n,n}}{\tilde{\eta}_{n,m}}\right| \leq \sum_{i>m} |\log\zeta_i(\tilde{\alpha}_i(n))| \leq 2\sum_{i>m}\log\zeta_i^+(\alpha_i(\infty)). \tag{4.10}$$

This is twice the tail of a series whose sum is $\log\eta_\infty^+$, which vanishes since $\eta_\infty^+ < \infty$ almost surely, hence $\tilde{\eta}_{n,n}/\tilde{\eta}_{n,m} \overset{p}{\to} 1$. For the same reason, we have $\eta_\infty/\eta_{\infty,m} \overset{as}{\to} 1$. Finally, since $\mathbb{P}[\tilde{\alpha}_i(n) = \alpha_i(n) = \alpha_i(\infty)] \to 1$ for any fixed i, $\mathbb{P}[\tilde{\eta}_{n,m} = \eta_{n,m} = \eta_{\infty,m}] \to 1$.

Step 3: Uniform integrability. Since $\eta_{n,m} \leq \eta_{n,m}^+ \leq \eta_{n,n}^+$, it suffices to prove uniform integrability of $\eta_{n,n}^+$. Let $p_{n,i} = \mathbb{P}[i^*(n) = i]$. By Lemma 4.1 and independence of $\alpha_i(\infty)$,

$$\mathbb{E}[\eta_{n,n}^+] = \sum_{i=1}^n p_{n,i}\mathbb{E}[\eta_{n,n}^+|i^*(n) = i] = \sum_{i=1}^n p_{n,i}\mathbb{E}\left[\zeta_i^+(\tilde{\alpha}_i(n) + 1)\prod_{j\neq i}\zeta_j^+(\tilde{\alpha}_j(n))\right] \tag{4.11}$$

$$\leq \sum_{i=1}^n p_{n,i}\mathbb{E}[\zeta_i^+(\alpha_i(\infty) + 1)]\prod_{j\neq i}\mathbb{E}[\zeta_j^+(\alpha_j(\infty))] \tag{4.12}$$

$$= \mathbb{E}[\eta_\infty^+]\sum_{i=1}^n p_{n,i}\frac{\mathbb{E}[\zeta_i^+(\alpha_i(\infty) + 1)]}{\mathbb{E}[\zeta_i^+(\alpha_i(\infty))]}. \tag{4.13}$$

Let $C = \sum_\ell c_\ell < \infty$ and $D = \sum_\ell \ell c_\ell < \infty$. Since $\alpha_i(\infty)$ is Poisson with parameter θ/i, $\mathbb{E}[\zeta_i^+(\alpha_i(\infty) + 1)] = 1 + De_i/\theta$ and $\mathbb{E}[\zeta_i^+(\alpha_i(\infty))] = 1 + (\theta + Ce_i)/i \geq 1$, hence

$$\sum_{i=1}^n p_{n,i}\frac{\mathbb{E}[\zeta_i^+(\alpha_i(\infty) + 1)]}{\mathbb{E}[\zeta_i^+(\alpha_i(\infty))]} \leq \sum_{i=1}^n p_{n,i}\left(1 + \frac{De_i}{\theta}\right) = 1 + \frac{D}{\theta}\mathbb{E}[e_{i^*(n)}]. \tag{4.14}$$

We previously showed that $e_{i*(n)}$ is bounded and vanishes in probability, so (4.14) converges to 1, hence $\limsup \mathbb{E}[\eta_{n,n}^+] \leq \mathbb{E}[\eta_\infty^+]$. But $\eta_{n,n}^+ \geq \tilde{\eta}_{n,n}^+$, while $\tilde{\eta}_{n,n}^+ \to \eta_\infty^+$ in L^1 by dominated convergence, thus

$$\limsup_{n\to\infty} \mathbb{E}\big[\big|\eta_{n,n}^+ - \tilde{\eta}_{n.n}^+\big|\big] = \limsup_{n\to\infty} \mathbb{E}[\eta_{n,n}^+] - \lim_{n\to\infty} \mathbb{E}[\tilde{\eta}_{n,n}^+] \leq 0. \tag{4.15}$$

Hence, $\eta_{n,n}^+ \to \eta_\infty^+$ in L^1 and is therefore uniformly integrable. $\qquad \square$

5 Conclusion

We studied measures on partitions in terms of perturbations of the Ewens measure, and showed that under a certain condition, many limit theorems for the Ewens measure are universal. This unifies the proofs of the limit theorems under a single universality theorem, while simultaneously extending the class of measures for which they hold beyond the previous frontier of logarithmic combinatorial structures. An interesting direction for future work is to better understand the asymptotically Ewens condition, which will perhaps allow more cases to be unified under the universality theorem.

References

[1] R. Arratia, A.D. Barbour and S. Tavaré (1992). Poisson process approximations for the Ewens sampling formula. *Ann. Appl. Probab.* **2** 519–535. MR-1177897

[2] R. Arratia, A.D. Barbour and S. Tavaré (1997). Random combinatorial structures and prime factorizations. *Notices Amer. Math. Soc.* **44** 903–910. MR-1467654

[3] R. Arratia, A.D. Barbour and S. Tavaré (1999). On Poisson-Dirichlet limits for random decomposable combinatorial structures. *Combin. Probab. Comput.* **8** 193–208. MR-1702562

[4] R. Arratia, A.D. Barbour and S. Tavaré (2000). Limits of logarithmic combinatorial structures. *Ann. Probab.* **28** 1620–1644. MR-1813836

[5] R. Arratia, A.D. Barbour and S. Tavaré (2003). *Logarithmic Combinatorial Structures: A Probabilistic Approach*. European Mathematical Society, Zürich. MR-2032426

[6] A.D. Barbour and S. Tavaré (1994). A rate for the Erdős-Turán law. *Combin. Probab. Comput.* **3** 167–176. MR-1288438

[7] V. Betz and D. Ueltschi (2009). Spatial random permutations and infinite cycles. *Comm. Math. Phys.* **285** 469–501. MR-2461985

[8] V. Betz, D. Ueltschi and Y. Velenik (2011). Random permutations with cycle weights. *Ann. Appl. Probab.* **21** 312–331. MR-2759204

[9] P. Billingsley (1968). *Convergence of Probability Measures*. Wiley, New York. MR-1700749

[10] J.M. DeLaurentis and B.G. Pittel (1985). Random permutations and Brownian motion. *Pacific J. Math.* **119** 287–301. MR-0803120

[11] P. Diaconis, M. McGrath and J. Pitman (1995). Riffles shuffles, cycles, and descents. *Combinatorica* **15** 11–29. MR-1325269

[12] P. Diaconis and A. Ram (2010). A probabilistic interpretation of the Macdonald polynomials. Preprint.

[13] N.M. Ercolani and D. Ueltschi (2011). Cycle structure of random permutations with cycle weights. Preprint.

[14] P. Erdős and P. Turán (1965). On some problems of a statistical group-theory I. *Probab. Theory Relat. Fields* **4** 175–186. MR-0184994

[15] P. Erdős and P. Turán (1967). On some problems of a statistical group-theory III. *Acta Math. Acad. Sci. Hungar.* **18** 309–320. MR-0215908

[16] W.J. Ewens (1972). The sampling theory of selectively neutral alleles. *Theoret. Population Biol.* **3** 87–112. MR-0325177

[17] W. Feller (1945). The fundamental limit theorems in probability. *Bull. Amer. Math. Soc.* **51** 800–832. MR-0013252

[18] P. Flajolet and M. Soria (1990). Gaussian limiting distributions for the number of components in combinatorial structures. *J. Combin. Theory Ser. A* **53** 165–182. MR-1041444

[19] B. Fristedt (1993). The structure of random permutations of large integers. *Trans. Amer. Math. Soc.* **337** 703–735. MR-1094553

[20] S.W. Golomb (1964). Random permutations. *Bull. Amer. Math. Soc.* **70** 747.

[21] V.L. Goncharov (1942). Sur la distribution des cycles dans les permutations. *Comptus Rendus (Doklady) de l'Académie des Sciences de l'URSS* **35** 267–269.

[22] V.L. Goncharov (1944). On the field of combinatory analysis. *Izvestiya Akademii Nauk SSSR* **8** 3–48. Translated to English (1962) *Amer. Math. Soc. Transl.* **19** 1–46.

[23] R.C. Griffiths (1979). On the distribution of allele frequencies in a diffusion model. *Theoret. Population Biol.* **15** 140–158. MR-0528914

[24] J.C. Hansen (1990). A functional central limit theorem for the Ewens sampling formula. *J. Appl. Probab.* **27** 28–43. MR-1039182

[25] J.C. Hansen (1994). Order statistics for decomposable combinatorial structures. *Random Structures Algorithms* **5** 517–533. MR-1293077

[26] F.M. Hoppe (1984). Pólya-like urns and the Ewens' sampling formula. *J. Math. Biol.* **20** 91–94. MR-0758915

[27] J.F.C. Kingman (1975). Random discrete distributions. *J. Roy. Statist. Soc. Ser. B* **37** 1–22. MR-0368264

[28] J.F.C. Kingman (1977). The population structure associated with the Ewens sampling formula. *Theoret. Populations Biol.* **11** 274–283. MR-0682238

[29] D.E. Knuth and L. Trabb Pardo (1976). Analysis of a simple factorization algorithm. *Theoret. Comput. Sci.* **3** 321–348. MR-0498355

[30] M. Lugo (2009). Profiles of permutations. *Electron. J. Combin.* **16** 99. MR-2529808

[31] A. Okounkov (2001). Infinite wedge and random partitions. *Selecta Math.* **7** 57–81. MR-1856553

[32] J. Pitman (1992). The two-parameter generalization of Ewens' random partition structure. Technical Report 345, Department of Statistics, UC Berkeley.

[33] J. Pitman and M. Yor (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25** 855–900. MR-1434129

[34] L. A. Shepp and S. P. Lloyd (1966). Ordered cycle lengths in a random permutation. *Trans. Amer. Math. Soc.* **121** 340–357. MR-0195117

[35] A.M. Vershik and A.A. Schmidt (1977). Limit measures arising in the theory of groups I. *Theory Probab. Appl.* **22** 72–88. MR-0448476

[36] A.M. Vershik and A.A. Schmidt (1978). Limit measures arising in the theory of groups II. *Theory Probab. Appl.* **23** 42–54. MR-0483019

[37] G. A. Watterson (1976). The stationary distribution of the infinitely-many neutral alleles diffusion model. *J. Appl. Probab.* **13** 639–651. MR-0504014