Combinatorics, Probability and Computing (1997) 6, 87–105
Copyright © 1997 Cambridge University Press

Explicit Limits of Total Variation Distance in Approximations of Random Logarithmic Assemblies by Related Poisson Processes

DUDLEY STARK†

Department of Mathematics, University of Melbourne, Parkeville, VIC 3052, Australia

Received 24 February 1995; revised 18 September 1995

Assemblies are labelled combinatorial objects that can be decomposed into components. Examples of assemblies include set partitions, permutations and random mappings. In addition, a distribution from population genetics called the Ewens sampling formula may be treated as an assembly. Each assembly has a size n, and the sum of the sizes of the components sums to n. When the uniform distribution is put on all assemblies of size n, the process of component counts is equal in distribution to a process of independent Poisson variables Z_i conditioned on the event that a weighted sum of the independent variables is equal to n. Logarithmic assemblies are assemblies characterized by some $\theta > 0$ for which $i\mathbb{E}Z_i \to \theta$. Permutations and random mappings are logarithmic assemblies; set partitions are not a logarithmic assembly. Suppose $b = b(n)$ is a sequence of positive integers for which $b/n \to \beta \in [0,1]$. For logarithmic assemblies, the total variation distance $d_b(n)$ between the laws of the first b coordinates of the component counting process and of the first b coordinates of the independent processes converges to a constant $H(\beta)$. An explicit formula for $H(\beta)$ is given for $\beta \in (0,1]$ in terms of a limit process which depends only on the parameter θ. Also, it is shown that $d_b(n) \to 0$ if and only if $b/n \to 0$, generalizing results of Arratia, Barbour and Tavaré for the Ewens sampling formula. Local limit theorems for weighted sums of the Z_i are used to prove these results.

1. Introduction

Assemblies are labelled combinatorial objects that may be decomposed into components. Examples are: permutations, whose components consist of cycles; mappings of a finite set into itself, whose components are cycles of rooted trees; and partitions, whose components are simply blocks with no additional structure. Assemblies are constructed by partitioning an underlying set of n elements into blocks and putting some structure on each block. The number of possible structures of size i is denoted by m_i. For permutations $m_i = (i-1)!$; for mappings on a finite set $m_i = (i-1)! \sum_{j=0}^{i-1} \frac{i^j}{j!}$; $m_i = 1$; for 2 regular graphs $m_i = \frac{1}{2}(i-1)!\mathbf{1}(i \geq 3)$. Given an assembly, the number of components of size i will be denoted by $C_i(n)$. The number of assemblies with total size n will be denoted by $p(n)$.

† Most of the work in this paper was done as part of the author's PhD dissertation at the University of Southern California. Thanks to thesis advisor Richard Arratia for his guidance.

https://doi.org/10.1017/S0963548396002775 Published online by Cambridge University Press

If $p(n) > 0$, and all assemblies of size n are chosen uniformly, then the component structure of the random assembly is described by the component counting process

$$\mathbf{C}(n) = (C_1(n), C_2(n), \ldots, C_n(n), 0, 0, \ldots).$$

The law of $\mathbf{C}(n)$ is then given for any $\mathbf{a} \in \mathbb{Z}_+^n$ by

$$\mathbb{P}(\mathbf{C}(n) = \mathbf{a}) = \frac{N(n, \mathbf{a})}{p(n)},$$

where $N(n, \mathbf{a})$ is the the number of assemblies for which $\mathbf{C}(n) = \mathbf{a}$. The number $N(n, \mathbf{a})$ is given by a generalization of Cauchy's formula for permutations:

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

For any $x > 0$, if $Z_1, Z_2, \ldots$ are independent Poisson variables having means

$$\mathbb{E}Z_i \equiv \lambda_i = \frac{m_i x^i}{i!}, \tag{1}$$

then

$$\Big(C_1(n), C_2(n), \ldots\Big) \stackrel{\mathrm{d}}{=} \Big(Z_1, Z_2, \ldots | Z_1 + 2Z_2 + \cdots = n\Big). \tag{2}$$

We let the first b coordinates of $\mathbf{C}(n)$ be

$$\mathbf{C}_b(n) = \Big(C_1(n), C_2(n) \ldots, C_b(n)\Big)$$

and let $b \to \infty$ in such a way that $b/n \to \beta \in [0, 1]$. Under these assumptions the distribution of $\mathbf{C}_b$ will be compared with the distibution of

$$\mathbf{Z}_b \equiv (Z_1, Z_2, \ldots, Z_b)$$

having an appropriate value of x. In this comparison total variation distance, a commonly used metric for probability distributions, will be used. Given two random elements X and Y on a countable space S, the total variation distance between them is given by

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

The total variation distance between $\mathbf{C}_b(n)$ and $\mathbf{Z}_b$ will be denoted by

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

In this paper it will be shown that for the class of logarithmic assemblies, described below, if

$$b/n \to \beta \in [0, 1],$$

then d_b converges to a constant $H(\beta)$. The function $H(\beta)$ has the property that $H(\beta) = 0$ if and only if $\beta = 0$; it follows that $d_b \to 0$ if and only if $b/n \to 0$.

The logarithmic class of assemblies, to which the results of this paper apply, are characterized by the existence of some $\theta > 0$ for which the parameter x in (1) can be fixed in a way such that

$$i\,\mathbb{E}Z_i \sim \theta. \tag{3}$$

https://doi.org/10.1017/S0963548396002775 Published online by Cambridge University Press

Examples of logarithmic assemblies include random mappings, for which $\theta = 1/2, x = e^{-1}$; permutations, for which $\theta = 1, x = 1$; and 2-regular graphs, for which $\theta = 1/2$, $x = 1$. Set partitions, on the other hand, are not a logarithmic assembly.

Distributions $\mathbf{C}(n)$ which are not those of random logarithmic combinatorial assemblies still may be treated as such when they satisfy (2) and (3). The Ewens sampling formula, which occurs in population genetics, is an important example of such a distribution. It is defined to be the measure given by (2), for which the Z_i are Poisson with means satisfying

$$i\,\mathbb{E}Z_i = \theta.$$

The distributions of random logarithmic assemblies may be looked upon as being perturbations of the Ewens sampling formula.

To prove the results in this paper, a definition of logarithmic assembly requiring more than (2) and (3) will be needed. All of the logarithmic assemblies listed above satisfy the stronger definition of logarithmic assembly that will be used. For this paper, logarithimic combinatorial assemblies are those assemblies satisfying (2), and satisfying (4) and (5) below.

The first requirement for an assembly to be logarithmic is that

$$\sum_{j=1}^{\infty} |\lambda_j - \theta/j| < \infty. \tag{4}$$

Other conditions are made on the function

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

We require $f(z)$ to be analytically continuable to a Δ-domain, which is defined for parameters $\eta > 0$ and $0 < \phi < \pi/2$ to be the set

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

Furthermore, $f(z)$ must be of the form

$$f(z) = K(1-z)^{-\theta} + O((1-z)^{-\theta+\delta}) \tag{5}$$

as $z \to 1$ within Δ for some $\delta > 0$.

Requiring $f(z)$ to be of form (5) transfers to knowledge of the behavior of the coefficients of $f(z)$. The following theorem is taken from [7].

Theorem 1.1. *Let α be a real number. If f is analytic on $\Delta \setminus \{1\}$ and $f(z) = O((1-z)^{-\alpha})$ as $z \to 1$ in Δ, then $[z^n]f(z) = O(n^{\alpha-1})$.*

An estimate on the coefficients of $(1-z)^{-\theta}$, also found in [7], will also be needed:

$$[z^n](1-z)^{-\theta} = n^{\theta-1}/\Gamma(\theta) + O(n^{\theta-2}). \tag{6}$$

Theorem 1.2 is the main result of this paper. The densities $g_\beta(z)$ and the defective densities $h_\beta(z)$ used in the statement of Theorem 1.2 are defined in Lemma 2.1.

https://doi.org/10.1017/S0963548396002775 Published online by Cambridge University Press

Theorem 1.2. *For a logarithmic assembly, if* $b/n \to \beta \in (0,1]$, *then*

$$d_b \to H(\beta) \equiv \frac{1}{2}\left(\int_0^\infty g_\beta(x)\left|1-\frac{h_\beta(1-x)}{g_1(1)}\right| dx + \beta^\theta \frac{g_\beta(1)}{g_1(1)}\right).$$

Theorem 1.2 was conjectured by Arratia and Tavaré [2] from local limit theorem heuristics which will be made precise by Proposition 3.1 in Section 3.

Theorem 1.2 does not allow for the case $\beta = 0$. However, Corollary 1.1 shows that $\beta = 0$ may be included when Theorem 1.2 is combined with results of Arratia, Barbour and Tavaré for the Ewens sampling formula.

Corollary 1.1. *For a logarithmic assembly,* $d_b(n) \to 0$ *if and only if* $b/n \to 0$.

Proof. Suppose $b/n \to 0$. For any $\beta \in (0,1]$, if

$$c(n) = b(n) + \lfloor \beta n \rfloor$$

where $\lfloor x \rfloor$ is the integer part of x, then $c(n)$ is such that $b(n) \le c(n)$ and $c(n)/n \to \beta$. Observing that $d_b \le d_c$, we have

$$\limsup_n d_b(n) \le \lim_n d_c(n) = H(\beta).$$

It is shown in [4] that for the Ewens sampling formula

$$d_c(n) \le k(\theta)\, c/n \tag{7}$$

for an explicit constant $k(\theta)$, implying

$$H(\beta) \le \lim_n k(\theta)\, c/n = k(\theta)\, \beta.$$

Since β is arbitrary, $\limsup_n d_b(n) = 0$.

Conversely, if b/n does not converge to 0, then we may find a $\beta \in (0,1]$ and a subsequence of the positive integers n_k for which

$$b(n_k)/n_k \to \beta.$$

For this subsequence $d_b(n_k) \to H(\beta) > 0$. □

Theorem 1.2 and Corollary 1.1 may be written more compactly if we let X_β and Y_β be independent variables with distributions μ_β and ν_β and define $X_1 = X_\beta + Y_\beta$.

Theorem 1.3. *For* $b = b(n)$ *such that* $b/n \to \beta \in [0,1]$,

$$d_b \to H(\beta) \equiv d_{\mathrm{TV}}\left(\mathscr{L}X_\beta, \mathscr{L}\left(X_\beta | X_1 = 1\right)\right).$$

In Section 5 we will show that X_β in Theorem 1.3 may be taken to be a real valued process X_β in Skorohod $D[0,1]$.

Results on the behaviour of d_b as $b/n \to 0$ are proven in [5]. For any logarithmic assembly with parameter θ, for b fixed, there is a constant $c(\theta, b)$ for which $d_b \sim c(\theta, b)\, b/n$. Furthermore, when $b(n) \to \infty$ in such a way that $b = o(n/\log n)$, there is a constant $c(\theta)$ for which the total variation distance d_b has asymptotic behaviour $d_b \sim c(\theta)\, b/n$.

https://doi.org/10.1017/S0963548396002775 Published online by Cambridge University Press

2. Related limiting distributions

If variables T_{lm} are defined by

$$T_{lm} = \sum_{i=l+1}^{m} iZ_i,$$

then equation (2) can be written as

$$\mathbf{C} \stackrel{\mathrm{d}}{=} \left(Z_1, Z_2, \ldots, Z_n | T_{0n} = n\right)$$

and d_b written as

$$d_b = \sum_{r=0}^{\infty} \mathbb{P}(T_{0b} = r) \left| 1 - \frac{\mathbb{P}(T_{bn} = n - r)}{\mathbb{P}(T_{0n} = n)} \right| \tag{8}$$

(see [3]). The limiting distributions of the variables $\frac{1}{n} T_{0b}$ and $\frac{1}{n} T_{bn}$ as $b/n \to \beta$ are characterized in Lemma 2.1.

Lemma 2.1. *If $b/n \to \beta \in (0, 1]$, then the sequence of variables $\frac{1}{n} T_{0b}$ converges weakly to a distribution μ_β having Laplace transform*

$$\int e^{-st} \mu_\beta(ds) = \exp\left\{ \theta \int_0^\beta (e^{-tu} - 1)\, u^{-1} du \right\}. \tag{9}$$

The distribution μ_β has a density $g_\beta(z)$ which on the interval $(0, \beta]$ equals

$$g_\beta(z) = e^{-\gamma\theta} \beta^{-\theta} \Gamma(\theta)^{-1} z^{\theta - 1}, \qquad z \in (0, \beta] \tag{10}$$

where γ is Euler's constant, and satisfies the integral equation

$$g_\beta(z) = \frac{\theta}{z} \int_{z-\beta}^{z} g_\beta(t) dt \tag{11}$$

for all $z > 0$.

The sequence of variables $\frac{1}{n} T_{bn}$ converges weakly to a distribution ν_β having Laplace transform

$$\int e^{-st} \nu_\beta(ds) = \exp\left\{ \theta \int_\beta^1 (e^{-tu} - 1)\, u^{-1} du \right\}. \tag{12}$$

The distribution ν_β has an atom at 0 of size β^θ and a defective density $h_\beta(z)$ which is identically 0 on $0 \le z < \beta$ and satisfies the integral equation

$$h_\beta(z) = \frac{\theta}{z} \left(\beta^\theta \mathbf{1}_{[\beta, 1]}(z) + \int_{z-1}^{z-\beta} h_\beta(t) dt \right) \tag{13}$$

for all $z > 0$.

Proof. The proof of the lemma will be given first for the variables $\frac{1}{n} T_{bn}$. To show that the Laplace transforms of the variables $\frac{1}{n} T_{bn}$ converge weakly to a measure ν_β with Laplace

https://doi.org/10.1017/S0963548396002775 Published online by Cambridge University Press

transform given by (12), we calculate

$$\begin{aligned}\log \mathbb{E}e^{-sT_{bn}/n} &= -\sum_{j=b+1}^{n} \lambda_j(1-e^{-js/n}) \\ &= -\sum_{j=b+1}^{n} \theta/i(1-e^{-js/n}) + \sum_{j=b+1}^{n} (\theta/i-\lambda_i)(1-e^{-js/n}).\end{aligned}$$

The first sum converges to $-\theta\int_\beta^1(1-e^{-sx})\frac{dx}{x}$, while the second is dominated by $\sum_{b+1}^n |\theta/i-\lambda_i|$ and converges to 0.

The size of the atom at 0 of the distribution ν_β is given by

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(T_{bn}=0) &= \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{j=b+1}^{n}\{Z_j=0\}\right) \\ &= \lim_{n\to\infty} \prod_{j=b+1}^{n} e^{-\lambda_j} \\ &= \lim_{n\to\infty} \exp\left(-\sum_{j=b+1}^{n} \theta/j\right) \exp\left(\sum_{j=b+1}^{n} (\theta/j-\lambda_j)\right) \\ &= \lim_{n\to\infty} e^{-\theta\log(n/b)+o(1)} \\ &= \beta^\theta. \end{aligned} \tag{14}$$

Now suppose that the measure ν'_β is defined as having an atom at 0 of size β^θ and having defective density $h_\beta(z)$ as described in the statement of the lemma. It will be shown that in fact $\nu'_\beta = \nu_\beta$. From (13) it follows that $h_\beta(z)$ is bounded.

$$\psi(t) = \beta^\theta + \int_0^\infty e^{-tz} h_\beta(z)dz,$$

we calculate

$$\begin{aligned}\psi'(t) &= \int_0^\infty -z h_\beta(z) e^{-tz} dz \\ &= \int_0^\infty -\theta\left(\beta^\theta \mathbf{1}_{[\beta,1]}(z) + \int_{z-1}^{z-\beta} h_\beta(s)ds\right) e^{-tz}dz \\ &= -\theta\beta^\theta \int_\beta^1 e^{-tz}dz - \theta\int_0^\infty \int_{z-1}^{z-\beta} h_\beta(s)e^{-tz}ds\,dz \\ &= -\theta\beta^\theta \int_\beta^1 e^{-tz}dz - \theta\int_0^\infty \int_{s+\beta}^{s+1} h_\beta(s)e^{-tz}dz\,ds \\ &= -\theta\beta^\theta \int_\beta^1 e^{-tz}dz + \theta\int_0^\infty h_\beta(s)t^{-1}(e^{-t(s+1)}-e^{-t(s+\beta)})ds \\ &= -\theta\beta^\theta \int_\beta^1 e^{-tz}dz - \theta\int_0^\infty h_\beta(s)e^{-ts}ds \int_\beta^1 e^{-tz}dz \\ &= -\theta\left(\int_\beta^1 e^{-tz}dz\right)\psi(t). \end{aligned} \tag{15}$$

https://doi.org/10.1017/S0963548396002775 Published online by Cambridge University Press

Solving equation (15) shows that

$$C \exp\left\{\theta \int_\beta^1 (e^{-tu} - 1)\, u^{-1} du\right\}$$

for some real number $C > 0$, and hence that $\nu'_\beta = C\nu_\beta$. The distributions ν'_β and ν_β have atoms of the same size at 0, from which $\nu'_\beta = \nu_\beta$.

The proof of the lemma for the variables $\frac{1}{n} T_{0b}$ is similar to the proof for $\frac{1}{n} T_{bn}$. Defining μ'_β to be a measure having density $g_\beta(z)$ as in the statement of the lemma, it can be shown as above that $\mu'_\beta = C\mu_\beta$ for some $C > 0$. To show that $C = 1$, it suffices to prove that $\lim_{n\to\infty} \mathbb{P}(T_{0b}/n \le b/n) = \mu'_\beta(0, \beta]$. For any $1 \le i \le b$,

$$\begin{aligned}
\mathbb{P}(T_{0b}/n = i/n) &= [z^i]\mathbb{E}z^{T_{0b}} \\
&= [z^i] \prod_{j=1}^{b} \mathbb{E}z^{jZ_j} \\
&= [z^i] \prod_{j=1}^{b} \exp\left(-\lambda_j + \lambda_j z^j\right) \\
&= \exp\left(-\sum_{j=1}^{b} \lambda_j\right)\ [z^i] \exp\left(\sum_{j=1}^{b} \lambda_j z^j\right) \\
&= \exp\left(-\sum_{j=1}^{b} \lambda_j\right)\ [z^i] \exp\left(\sum_{j=1}^{\infty} \lambda_j z^j\right) \\
&= \exp\left(-\sum_{j=1}^{b} \lambda_j\right)\ [z^i] f(z) \\
&= \exp\left(-\sum_{j=1}^{b} \lambda_j\right)\ [z^i] \left(K(1-z)^{-\theta} + \Psi(z)\right) \\
&= \exp\left(-\sum_{j=1}^{b} \lambda_j\right)\ \left(\frac{K i^{\theta-1}}{\Gamma(\theta)} + E_i\right). \qquad (16)
\end{aligned}$$

Theorem 1.1 and expansion (6) show that $E_i = O(i^{\theta-\delta\wedge 1-1})$. We now calculate

$$\begin{aligned}
\mathbb{P}(T_{0b}/n \le b/n) &= \sum_{j=1}^{b} \mathbb{P}(T_{0b}/n = j/n) \\
&= \exp\left(-\sum_{j=1}^{b} \lambda_j\right) \sum_{j=1}^{b} \left(\frac{K j^{\theta-1}}{\Gamma(\theta)} + E_j\right) \\
&= \exp\left(\sum_{j=1}^{b} \theta/j - \lambda_j\right) \exp\left(-\sum_{j=1}^{b} \theta/j\right) \left(\frac{K}{\Gamma(\theta)} \sum_{j=1}^{b} j^{\theta-1} + K^{-1} \sum_{j=1}^{b} E_j\right).
\end{aligned}$$

https://doi.org/10.1017/S0963548396002775 Published online by Cambridge University Press

Because of (4) and (5) we can apply Abel's limit theorem (see [1], for example) to the function $f(z)(1-z)^{\theta}$ to see that

$$K = \exp\left(\sum_{j=1}^{\infty} \lambda_j - \theta/j\right). \tag{17}$$

Using the bound

$$\sum_{j=1}^{b} E_j = O\left(\log b\, b^{\theta-\delta\wedge 1}\right)$$

and the estimate

$$\sum_{j=1}^{b} 1/j - \gamma - \log b = O(b^{-1}) \tag{18}$$

(see [8]), and letting n, and therefore b, tend to ∞, shows that

$$\lim_{n\to\infty} \mathbb{P}(T_{0b}/n \le b/n) = \tau^{-1}\, \mathrm{e}^{-\tau\gamma}\, \gamma(\tau)^{-1} = \int_0^{\beta} g_{\beta}(z)\, \mathrm{d}z.$$

Therefore $C = 1$ and $\mu_{\beta} = \mu'_{\beta}$. □

Remark. An immediate consequence of Lemma 2.1 is that $g_{\beta}(z)$ is bounded and continuous on any interval of the form $[\epsilon, r]$ for $\epsilon, r > 0$ while $h_{\beta}(z)$ is bounded on $\mathbb{R}$, and continuous except at the points β and 1.

3. A local limit proposition

In this section a local limit result, Proposition 3.1, will be proved. It describes the behavior in the limit of the probability densities

$$g_n(z) = \mathbb{P}\left(T_{0b}/n = z\right) \quad \text{and} \quad h_n(z) = \mathbb{P}\left(T_{bn}/n = z\right).$$

and the recursions satisfied by $g_n(z)$ (resp. $h_n(z)$) will be exploited. Fourier transform methods will not be used; a difficulty in applying such methods here would be that the characteristic functions of ν_{β} and μ_{β} are not generally integrable. It may be shown by differentiating their probability generating functions that the densities $g_n(z)$ and $h_n(z)$ satisfy recursions

$$g_n(z) = \frac{1}{z} \sum_{j=1}^{b} g_n(z - j/n)\, \frac{j\lambda_j}{n} \tag{19}$$

and

$$h_n(z) = \frac{1}{z} \sum_{j=b+1}^{n} h_n(z - j/n)\, \frac{j\lambda_j}{n} \tag{20}$$

for $z > 0$.

Proposition 3.1. *The following statement is true:*

$$|\mathbb{P}(T_{0n} = n) - g_1(1)/n| = o(n^{-1}).$$

https://doi.org/10.1017/S0963548396002775 Published online by Cambridge University Press

In addition, there exist positive real numbers $R_{1,n}(i) \geq 0$ *for* $n \geq 1$ *and* $1 \leq i \leq n$ *such that the following statements are true:*

$$\left| \mathbb{P}(T_{0b} = i) - \frac{g_\beta(i/n)}{n} \right| \leq R_{1,n}(i),$$

$$\sum_{i=1}^{n} R_{1,n}(i) = o(1),$$

$$R_{1,n}(n) = o(n^{-1}).$$

In addition, there exist positive real numbers $R_{2,n}(i) \geq 0$ *for* $n \geq 1$ *and* $1 \leq i \leq n$ *such that these two statements are true:*

$$\left| \mathbb{P}(T_{bn} = i) - \frac{h_\beta(i/n)}{n} \right| \leq R_{2,n}(i),$$

$$\max_{1 \leq i \leq n} R_{2,n}(i) = o(n^{-1}).$$

In the proof of Proposition 3.1 the methods used to define $R_{1,n}(i)$ and $R_{2,n}(i)$ will depend on the value of i. In particular, the function

$$b^*(n) \equiv b(n) \wedge \lfloor \beta n \rfloor$$

will be a dividing line between two methods for defining $R_{1,n}(i)$. For i in the range $1 \leq i \leq b^*$, we will define $R_{1,n}(i)$ explicitly; for i in the range $b^* < i \leq n$, $R_{1,n}(i)$ will be defined inductively from the similarity of (19) and (11). Similarly, for i in the range $1 \leq i < b$, $R_{2,n}(i)$ will be set equal to 0 and for i in the range $b \leq i \leq n$ $R_{2,n}(i)$ will be defined inductively from the similarity between (20) and (13).

The proof of Proposition 3.1 constitutes the remainder of this section.

Lemma 3.1. *$R_{1,n}(i)$ may be defined for $1 \leq i \leq b^*$ so that the following statements hold:*

$$|\mathbb{P}(T_{0b}/n = i/n) - g_\beta(i/n)/n| \leq R_{1,n}(i),$$

$$\sum_{i=1}^{b^*} R_{1,n}(i) = o(1).$$

In addition:

$$|\mathbb{P}(T_{0n} = n) - g_1(1)/n| = o(n^{-1}).$$

Proof. The proof begins by observing that by setting $b = i = n$ in (16) we obtain the estimate

$$\mathbb{P}(T_{0n} = n) = O(n^{-1}). \tag{21}$$

Next, substituting $z = i/n$ with $1 \leq i \leq \lfloor \beta n \rfloor$ in $g_\beta(z)$ gives

$$g_\beta(i/n)/n = e^{-\gamma\theta}\Gamma(\theta)^{-1}(b/\beta n)^\theta b^{-\theta} i^{\theta-1}. \tag{22}$$

https://doi.org/10.1017/S0963548396002775 Published online by Cambridge University Press

Subtracting (22) from (16) one obtains the identity

$$\mathbb{P}(T_{0b}/n = i/n) - g_\beta(i/n)/n = -\mathbb{P}(T_{0b} = i)\Lambda_n + K^{-1}E_i \exp(-\gamma\theta - \theta \log b)\,(b/\beta n)^\theta$$

where Λ_n is defined as

$$\begin{aligned}\Lambda_n &\equiv K^{-1}(b/\beta n)^\theta \exp\left(\sum_{j=1}^{b} \lambda_j - \gamma\theta - \theta \log b\right) - 1 \\ &= K^{-1}(b/\beta n)^\theta \exp\left(\sum_{j=1}^{b} \lambda_j - \theta/j\right) \exp\left(\sum_{j=1}^{b} \theta/j - \gamma\theta - \theta \log b\right) - 1.\end{aligned}$$

The identity (17) and the estimate (18) show that

$$|\Lambda_n| = o(1). \tag{23}$$

If we define $R_{1,n}(i)$ to be

$$R_{1,n}(i) \equiv \mathbb{P}(T_{0b} = i)|\Lambda_n| + K^{-1}|E_i| \exp(-\gamma\theta - \theta \log b)\,(b/\beta n)^\theta, \tag{24}$$

then $R_{1,n}(i)$ satisfies the first statement of the lemma. Letting $b = i = n$, the estimates (23), (21) and

$$E_n = O(n^{\theta - \delta \wedge 1 - 1})$$

show that

$$R_{1,n}(n) = o(n^{-1}),$$

proving the third statement of the lemma.

To complete the lemma, we sum (24) over i from 1 to b^*. Doing so will produce two terms, the first of which is bounded by $|\Lambda_n|$. As for the second term,

$$\begin{aligned}\sum_{i=1}^{b^*} K^{-1}|E_i| \exp(-\gamma\theta - \theta \log b)\,(b/\beta n)^\theta &= O\left(b^{-\theta} \sum_{i=1}^{b^*} i^{\theta - \delta \wedge 1 - 1}\right) \\ &= O\left(b^{-\theta} \log b^* (b^* + 1)^{\theta - \delta \wedge 1}\right) \\ &= o(1). \qquad \square\end{aligned}$$

Remark. The calculation

$$\begin{aligned}\mathbb{P}(T_{0b}/n = 1/n) &= \mathbb{P}\left(\bigcap_{j \neq 1} \{Z_j = 0\} \cup \{Z_1 = 1\}\right) \\ &= \lambda_1 e^{-\sum_{j=1}^{b} \lambda_j} \asymp n^{-\theta}\end{aligned}$$

shows that if $\theta \leq 1$, then it is not possible to choose $R_{1,n}(i)$ in such a way that $R_{1,n}(i) =$

https://doi.org/10.1017/S0963548396002775 Published online by Cambridge University Press

$o(n^{-1})$ for all $1 \leq i \leq n$. However, Proposition 3.1 shows $R_{2,n}(i)$ may be chosen so that $R_{2,n}(i) = o(n^{-1})$ uniformly over i, $1 \leq i \leq n$.

Before defining $R_{1,n}(i)$ for $i > b^*$, we need to examine solutions to recursions of the form

$$y_i = \frac{\alpha}{n} \sum_{j=1}^{n} y_{i-j} + c \tag{25}$$

for $i \geq 1$, where c and α are positive constants and $y_i = 0$ for $i < 0$. Though $R_{1,n}(i)$ for $b^* < i \leq n$ and $R_{2,n}(i)$ for $b \leq i \leq n$ will be solutions to recursions of the more general form

$$y_i = \frac{\alpha}{n} \sum_{j=m}^{m'} y_{i-j} + c \tag{26}$$

where $1 \leq m < m' \leq n$ are integers, bounds for the solution of (25) will also be bounds for solutions of (26).

Lemma 3.2. *Let c,α be positive real numbers. Assuming that $y_i = 0$ for $i \leq 0$ and that y_i is defined recursively by* (25)*, then for $1 \leq i \leq n$,*

$$y_i \leq e^{\alpha} c.$$

Proof. An induction argument shows that

$$y_i = (1 + \alpha/n)^{i-1} c \leq e^{\alpha} c. \qquad \square$$

Lemma 3.3. *For $n \geq 1$ a sequence $R_{1,n}(i)$ may be defined for $b^* + 1 \leq i \leq n$ such that*

$$|\mathbb{P}(T_{0b}/n = i/n) - g_{\beta}(i/n)/n| \leq R_{1,n}(i),$$

$$\sum_{i=b^*+1}^{n} R_{1,n}(i) = o(1),$$

$$R_{1,n}(n) = o(n^{-1}).$$

Proof. automatically from the second. We will define $R_{1,n}(i)$ inductively and then show that

$$\sup_{b^* < i \leq n} R_{1,n}(i) = o(n^{-1}),$$

which implies both the second and third statements. The constant

$$L \equiv (\min_{n}(b^*/n))^{-1}$$

serves as a bound on z^{-1} for $z = i/n$ where i is in the range $b^* < i \leq n$.

Assuming that $R_{1,n}(j)$ has been defined for $j < i$, the following bound will help to define

https://doi.org/10.1017/S0963548396002775 Published online by Cambridge University Press

$R_{1,n}(i)$. We suppose $b^* < i \le n$ is fixed and let $z = i/n$.

$$\begin{aligned}
|g_\beta(z)/n - g_n(z)| &= \left| \frac{\theta}{nz} \int_{z-\beta}^{z} g_\beta(t)dt - \frac{1}{nz} \sum_{j=1}^{b} j\lambda_j g_n(z - j/n) \right| \\
&\le \left| \frac{\theta}{nz} \int_{z-\beta}^{z-b/n} g_\beta(t)dt \right| \\
&\quad + \left| \frac{\theta}{nz} \int_{z-b/n}^{z} g_\beta(t)dt - \frac{\theta}{n^2 z} \sum_{j=1}^{b} g_\beta(z - j/n) \right| \\
&\quad + \left| \frac{\theta}{nz} \sum_{j=1}^{b} g_\beta(z - j/n)/n - \frac{\theta}{nz} \sum_{j=1}^{b} g_n(z - j/n) \right| \\
&\quad + \left| \frac{1}{nz} \sum_{j=1}^{b} \theta g_n(z - j/n) - \frac{1}{nz} \sum_{j=1}^{b} j\lambda_j g_n(z - j/n) \right| \\
&\le \frac{\theta L}{n} \sup_z \left(\left| \int_{z-\beta}^{z-b/n} g_\beta(t)dt \right| \right) \qquad (27) \\
&\quad + \frac{\theta L}{n} \sum_{j=1}^{n-1} \left| \int_{j/n}^{(j+1)/n} g_\beta(t)dt - \frac{1}{n} g_\beta(j/n) \right| \qquad (28) \\
&\quad + \frac{\theta L}{n} \sum_{j=i-b^*}^{i-1} R_{1,n}(i-j) \qquad (29) \\
&\quad + \frac{\theta L}{n} \sum_{j=1}^{i-b^*-1} R_{1,n}(i-j) \\
&\quad + \frac{L}{n} \sup_z \sum_{j=1}^{n} |\theta - j\lambda_j| \; g_n(z - j/n). \qquad (30)
\end{aligned}$$

We define $R_{1,n}(i)$ inductively to be the last sum of terms above, immediately satisfying the first statement of the lemma, and define $\alpha = \theta L$ and c_n to be

$$c_n = \text{expression(27)} + \text{expression(28)} + \text{expression(29)} + \text{expression(30)}.$$

Applying Lemma 3.2 we have the upper bound

$$R_{1,n}(i) \le e^\alpha c_n$$

independently of i.

We will now show that each of (27)–(30) are $o(n^{-1})$. Expression(29) was shown to be $o(n^{-1})$ in Lemma 3.1. Expression(27) is $o(n^{-1})$ because $b/n \to \beta$ and $g_\beta(z)$ is integrable.

To show (28) is $o(n^{-1})$, separate arguments are used for j in the ranges $j \le \lfloor \beta n \rfloor - 1$ and $j > \lfloor \beta n \rfloor - 1$. For j in the second range,

$$\frac{j}{n} > \frac{\lfloor \beta n \rfloor}{n} - \frac{1}{n} \ge \beta - \frac{2}{n}.$$

https://doi.org/10.1017/S0963548396002775 Published online by Cambridge University Press

By the remark after Lemma 2.1, $g_\beta(z)$ is bounded and continuous for $z \geq \beta - \frac{2}{n}$ whenever $\beta - 2/n > 0$, so for large enough n the sum of terms for j in the second range only is $o(1/n)$. For j in the first range ,

$$\frac{j+1}{n} \leq \frac{\lfloor \beta n \rfloor}{n} \leq \beta$$

for which we have a closed form for $g_\beta(z)$; it is enough to show that

$$\sum_{j=1}^{n-1} \left| \int_{j/n}^{(j+1)/n} x^{\theta-1} dx - \frac{1}{n}\left(\frac{j}{n}\right)^{\theta-1} \right| \to 0.$$

As $f(x) = x^{\theta-1}$ is monotone, we have

$$\begin{aligned} \sum_{j=1}^{n-1} \left| \int_{j/n}^{(j+1)/n} x^{\theta-1} dx - \frac{1}{n}\left(\frac{j}{n}\right)^{\theta-1} \right| &= \left| \sum_{j=1}^{n-1} \int_{j/n}^{(j+1)/n} x^{\theta-1} dx - \sum_{j=1}^{n-1} \frac{1}{n}\left(\frac{j}{n}\right)^{\theta-1} \right| \\ &= \left| \int_{1/n}^{1} x^{\theta-1} dx - \sum_{j=1}^{n-1} \frac{1}{n}\left(\frac{j}{n}\right)^{\theta-1} \right|. \end{aligned}$$

Dominated convergence yields

$$\lim_{n\to\infty} \sum_{j=1}^{n-1} \frac{1}{n}\left(\frac{j}{n}\right)^{\theta-1} = \int_0^1 x^{\theta-1} dx = \theta^{-1}.$$

To show expression (30) is of order $o(n^{-1})$, seperate arguments are used for the ranges $1 \leq j \leq \sqrt{n}$ and $\sqrt{n} < j \leq n$. Defining

$$\bar{\theta} \equiv \sup_j j\lambda_j,$$

observe that (19) implies

$$g_n(z) \leq \frac{\bar{\theta} L}{n} \sum_{j=1}^{b} g_n(z - j/n) \leq \frac{\bar{\theta} L}{n}.$$

The sum over the range $1 \leq j \leq \sqrt{n}$ is be bounded by

$$\begin{aligned} \frac{L}{n} \sup_z \sum_{1\leq j\leq\sqrt{n}} |\theta - j\lambda_j|\, g_n(z - j/n) &\leq \frac{L}{n}(\theta + \bar{\theta}) \sup_z \sum_{1\leq j\leq\sqrt{n}} g_n(z - j/n) \\ &\leq \frac{L}{n}(\theta + \bar{\theta}) \frac{\bar{\theta} L}{\sqrt{n}} \\ &= o(n^{-1}). \end{aligned}$$

The sum over the range $\sqrt{n} < j \leq n$ is bounded by

$$\frac{L}{n} \sup_z \sum_{\sqrt{n}<j\leq n} |\theta - j\lambda_j|\, g_n(z - j/n) \leq \frac{L}{n} \sup_{\sqrt{n}<j\leq n} |\theta - j\lambda_j| = o(n^{-1}).$$

This completes the proof of the lemma. □

https://doi.org/10.1017/S0963548396002775 Published online by Cambridge University Press

The statements about $R_{2,n}(i)$ will be proven in a similar way.

Lemma 3.4. *For $n \geq 1$, a sequence $R_{2,n}(i)$ may be defined for $0 < i \leq n$ such that the following statements are true:*

$$|\mathbb{P}(T_{bn}/n = i/n) - h_\beta(i/n)/n| \leq R_{2,n}(i)$$

$$\max_{0<i\leq n} R_{2,n}(i) = o(n^{-1}).$$

Proof. For $0 < i < b$,

$$\mathbb{P}(T_{bn}/n = i/n) = h_\beta(i/n)/n = 0;$$

therefore it suffices to take $R_{2,n}(i) = 0$ for i in that range. For $b \leq i \leq n$, Lemma 3.2 will be used. Letting $z = i/n$, the following bound is used to define $R_{2,n}(i)$:

$$\begin{aligned}
|h_\beta(z)/n - h_n(z)| &= \left| \frac{\theta}{nz} \int_{z-1}^{z-\beta} h_\beta(t)dt + \frac{\theta\beta^\theta}{nz} - \frac{1}{nz} \sum_{j=b+1}^{n} j\lambda_j h_n(z - j/n) \right| \\
&\leq \left| \frac{\theta\beta^\theta}{nz} - \frac{\theta}{nz} h_n(0) \right| \\
&\quad + \left| \frac{\theta}{nz} \int_{z-b/n}^{z-\beta} h_\beta(t)dt \right| \\
&\quad + \left| \frac{\theta}{nz} \int_{z-1}^{z-b/n} h_\beta(t)dt - \frac{\theta}{n^2 z} \sum_{j=b+1}^{i-1} h_\beta(z - j/n) \right| \\
&\quad + \left| \frac{\theta}{nz} \sum_{j=b+1}^{i-1} h_\beta(z - j/n)/n - \frac{\theta}{nz} \sum_{j=b+1}^{i-1} h_n(z - j/n) \right| \\
&\quad + \left| \frac{1}{nz} \sum_{j=b+1}^{i-1} \theta h_n(z - j/n) - \frac{1}{nz} \sum_{j=b+1}^{i-1} j\lambda_j h_n(z - j/n) \right| \\
&\leq \frac{\theta L}{n} \left| \beta^\theta - \mathbb{P}(T_{bn} = 0) \right| \qquad (31) \\
&\quad + \frac{\theta L}{n} \sup_z \left(\left| \int_{z-b/n}^{z-\beta} h_\beta(t)dt \right| \right) \qquad (32) \\
&\quad + \frac{\theta L}{n} \sum_{j=1}^{n} \left| \int_{j/n}^{(j+1)/n} h_\beta(t)dt - h_\beta(j/n)/n \right| \qquad (33) \\
&\quad + \frac{\theta L}{n} \sum_{j=b+1}^{i-b} R_{2,n}(i - j) \\
&\quad + \frac{L}{n} \sup_z \sum_{j=1}^{b} |\theta - j\lambda_j| \, h_n(z - j/n). \qquad (34)
\end{aligned}$$

https://doi.org/10.1017/S0963548396002775 Published online by Cambridge University Press

$R_{2,n}(i)$ are defined recursively as in Lemma 3.3. The second statement of the lemma is proven by showing that each of (31)–(34) is $o(n^{-1})$ as in the proof of Lemma 3.3. Expression (33) is $o(n^{-1})$ because $h_\beta(z)$ is bounded and discontinuous only at $z = \beta$ and $z = 1$. We have already shown that (31) is $o(n^{-1})$, and the arguments that show (27) and (30) are $o(n^{-1})$, respectively, show (32) and (34) are $o(n^{-1})$. □

This completes the proof of Proposition 3.1.

4. Convergence of total variation distance

Now that Proposition 3.1 has been proven, it will be used to prove Theorem 1.2. The primary method for simplifying equation (8) involves writing d_b in the form

$$d_b = \frac{1}{2}\sum_{i=0}^{\infty} \mathbb{P}(T_{0b} = i)\,|S_{1,n,i} + S_{2,n,i}| + o(1) \tag{35}$$

and then using the triangle inequality

$$\Big|\, |S_{1,n,i} + S_{2,n,i}| - |S_{1,n,i}| \,\Big| \le |S_{2,n,i}|.$$

If

$$\mathscr{S}_n \equiv \sum_{i=0}^{\infty} \mathbb{P}(T_{0b} = i)\,|S_{2,n,i}| = o(1),$$

then $|S_{1,n,i} + S_{2,n,i}|$ may be replaced by $|S_{1,n,i}|$ in (35). First, weak convergence is used to handle terms in (8) with $i > n$.

Lemma 4.1. $d_b = \frac{1}{2}\sum_{i=0}^{n} \mathbb{P}(T_{0b} = i)\left|1 - \frac{\mathbb{P}(T_{bn}=n-i)}{\mathbb{P}(T_{0n}=n)}\right| + \frac{1}{2}\int_1^\infty g_\beta(t)dt + o(1).$

Proof. Rewriting d_b as

$$\begin{aligned} d_b &= \frac{1}{2}\sum_{i=0}^{n} \mathbb{P}(T_{0b} = i)\left|1 - \frac{\mathbb{P}(T_{bn} = n-i)}{\mathbb{P}(T_{0n} = n)}\right| + \frac{1}{2}\sum_{i=n+1}^{\infty} \mathbb{P}(T_{0b} = i) \\ &= \frac{1}{2}\sum_{i=0}^{n} \mathbb{P}(T_{0b} = i)\left|1 - \frac{\mathbb{P}(T_{bn} = n-i)}{\mathbb{P}(T_{0n} = n)}\right| + \frac{1}{2}\,\mathbb{P}(T_{0b}/n > 1) \end{aligned}$$

and an application of the Portmanteau lemma proves Lemma 4.1. □

It is convenient to define d_b^* as

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

Lemma 4.2.

$$d_b^* = \frac{1}{2}\sum_{i=0}^{n} \mathbb{P}(T_{0b} = i)\left|1 - \frac{\mathbb{P}(T_{bn} = n-i)}{g_1(1)/n}\right| + o(1).$$

https://doi.org/10.1017/S0963548396002775 Published online by Cambridge University Press

Proof. Writing d_b^* in the form

$$d_b^* = \frac{1}{2}\sum_{i=0}^{n} \mathbb{P}(T_{0b}=i)\left|1-\frac{\mathbb{P}(T_{bn}=n-i)}{g_1(1)/n}+\frac{\left(\mathbb{P}(T_{0n}=n)-g_1(1)/n\right)\,\mathbb{P}(T_{bn}=n-i)}{\mathbb{P}(T_{0n}=n)\,g_1(1)/n}\right|$$

shows that

$$d_b^* = \frac{1}{2}\sum_{i=0}^{n} \mathbb{P}(T_{0b}=i)\left|1-\frac{\mathbb{P}(T_{bn}=n-i)}{g_1(1)/n}\right| + \mathscr{S}_n,$$

where by the triangle inequality

$$\begin{aligned} |\mathscr{S}_n| &\le \frac{1}{2}\sum_{i=0}^{n} \frac{|\mathbb{P}(T_{0n}=n)-g_1(1)/n|\;\mathbb{P}(T_{0b}=i)\,\mathbb{P}(T_{bn}=n-i)}{\mathbb{P}(T_{0n}=n)\,g_1(1)/n} \\ &= \frac{1}{2}\frac{|\mathbb{P}(T_{0n}=n)-g_1(1)/n|}{g_1(1)/n} \\ &= o(1). \end{aligned}$$

□

Lemma 4.3.

$$d_b^* = \frac{1}{2}\sum_{i=0}^{n-1} \mathbb{P}(T_{0b}=i)\left|1-\frac{h_\beta(1-i/n)}{g_1(1)}\right| + \frac{1}{2}\beta^\theta\frac{g_\beta(1)}{g_1(1)} + o(1).$$

Proof. Rewriting Lemma 4.2 in the form

$$\begin{aligned} d_b^* &= \frac{1}{2}\sum_{i=0}^{n-1} \mathbb{P}(T_{0b}=i)\left|1-\frac{h_\beta(1-i/n)}{g_1(1)}-\frac{\mathbb{P}(T_{bn}=n-i)-h_\beta(1-i/n)/n}{g_1(1)/n}\right| \\ &\quad + \frac{1}{2}\mathbb{P}(T_{0b}=n)\left|1-\frac{\mathbb{P}(T_{bn}=0)}{g_1(1)/n}\right| + o(1) \end{aligned} \tag{36}$$

shows that

$$\begin{aligned} d_b^* &= \\ &\frac{1}{2}\sum_{i=0}^{n-1} \mathbb{P}(T_{0b}=i)\left|1-\frac{h_\beta(1-i/n)}{g_1(1)}\right| + \frac{1}{2}\mathbb{P}(T_{0b}=n)\left|1-\frac{\mathbb{P}(T_{bn}=0)}{g_1(1)/n}\right| + \mathscr{S}_n + o(1) \end{aligned}$$

$$\begin{aligned} |\mathscr{S}_n| &\le \frac{1}{2}\sum_{i=0}^{n-1} \mathbb{P}(T_{0b}=i)\frac{R_{2,n}(n-i)}{g_1(1)/n} \\ &\le \frac{1}{2}\frac{\max_{0<i\le n} R_{2,n}(i)}{g_1(1)/n} \\ &= o(1). \end{aligned} \tag{37}$$

Expression (36) converges to $\frac{1}{2}\beta^\theta\frac{g_\beta(1)}{g_1(1)}$ because $\mathbb{P}(T_{0b}=n)\sim g_\beta(1)/n$ by Proposition 3.1 and $\mathbb{P}(T_{bn}=0)\sim\beta^\theta$ by (14). □

https://doi.org/10.1017/S0963548396002775 Published online by Cambridge University Press

Lemma 4.4.

$$d_b^* = \frac{1}{2}\sum_{i=1}^{n} \frac{g_\beta(i/n)}{n}\left|1 - \frac{h_\beta(1-i/n)}{g_1(1)}\right| + \frac{1}{2}\beta^\theta \frac{g_\beta(1)}{g_1(1)} + o(1).$$

Proof. From Lemma 4.3,

$$\begin{aligned} d_b^* &= \frac{1}{2}\sum_{i=1}^{n} \frac{g_\beta(i/n)}{n}\left|1 - \frac{h_\beta(1-i/n)}{g_1(1)}\right| + \frac{1}{2}\beta^\theta \frac{g_\beta(1)}{g_1(1)} + o(1) + \mathscr{S}_n + \\ &\quad \frac{1}{2}\mathbb{P}(T_{0b} = 0)\left|1 - \frac{h_\beta(1)}{g_1(1)}\right| - \frac{1}{2}\mathbb{P}(T_{0b} = n)\left|1 - \frac{h_\beta(0)}{g_1(1)}\right| \end{aligned}$$

where, using the fact that $h_\beta(z)$ is bounded,

$$\begin{aligned} |\mathscr{S}_n| &\le \frac{1}{2}\sum_{i=1}^{n} R_{1,n}(i)\left|1 - \frac{h_\beta(1-i/n)}{g_1(1)}\right| \\ &\le \frac{1}{2}\left(\max_{0\le i\le n}\left|1 - \frac{h_\beta(1-i/n)}{g_1(1)}\right|\right)\sum_{i=1}^{n} R_{1,n}(i) \qquad (38) \\ &= o(1). \end{aligned}$$

From Proposition 3.1 it follows that

$$\mathbb{P}(T_{0b} = n) = o(1).$$

To prove

$$\mathbb{P}(T_{0b} = 0) = o(1)$$

a calculation is carried out similar to the one made in the remark after Lemma 3.1. □

Letting $n \to \infty$ in Lemma 4.4, a dominated convergence argument gives

$$d_b^* = \frac{1}{2}\int_0^1 g_\beta(x)\left|1 - \frac{h_\beta(1-x)}{g_1(1)}\right| dx + \frac{1}{2}\beta^\theta \frac{g_\beta(1)}{g_1(1)} + o(1). \qquad (39)$$

Combining (39) with Lemma 4.1 completes the proof of Theorem 1.2.

5. A related process

In this section it will shown that for the Ewens sampling formula $\frac{1}{n}T_{0,\lfloor\beta n\rfloor}$ converges to a real valued process X_β in Skorohod $D[0,1]$.

For each k-tuple of points $t_1, t_2, \ldots, t_k$ in $[0,1]$ and sets $A_1, A_2, \ldots, A_k$ in the Borel σ-field $\mathscr{R}$, define a measure $\tilde{\mu}_{t_1\ldots t_k}$ on $(\mathbb{R}^k, \mathscr{R}^k)$ by

$$\tilde{\mu}_{t_1\ldots t_k}(A_1,\ldots,A_k) = \lim_{n\to\infty}\mathbb{P}\left(\frac{T_{0,\lfloor nt_1\rfloor}}{n} \in A_1, \ldots, \frac{T_{0,\lfloor nt_k\rfloor}}{n} \in A_k\right).$$

Each of the sequences of variables $T_{0,\lfloor nt_i\rfloor}/n$ is tight, and therefore $\tilde{\mu}_{t_1,\ldots,t_k}$ is a probability measure.

To demonstrate the existence of a process in $D[0,1]$ with finite dimensional distributions $\tilde{\mu}_{t_1\ldots t_2}$ Theorem 15.7 of [6] will be used:

https://doi.org/10.1017/S0963548396002775 Published online by Cambridge University Press

Theorem 5.1. *There exists in $D[0,1]$ a random element with finite dimensional distributions $\mu_{t_1\ldots t_k}$, provided these distributions are consistent; provided*

$$\int_{\mathbb{R}^3} |\zeta - \zeta_1|^\gamma |\zeta_2 - \zeta|^\gamma d\mu_{t_1 t t_2}(\zeta_1, \zeta, \zeta_2) \leq (F(t_2) - F(t_1))^{2\alpha} \tag{40}$$

for $t_1 \leq t \leq t_2$, where $\gamma \geq 0$, $\alpha > 1/2$, and F is a nondecreasing, continuous function on $[0,1]$; and provided

$$\lim_{h \downarrow 0} \mu_{t,t+h} \{(\zeta_1, \zeta) : |\zeta_1 - \zeta_2| \geq \epsilon\} = 0, \tag{41}$$

for $0 \leq t < 1$.

The hypotheses of Theorem 5.1 will now be verified for $\tilde{\mu}$. To prove that condition (40) holds, begin by defining $X_n(t) = \frac{1}{n} T_{0,\lfloor nt \rfloor}$ and calculate

$$\begin{aligned}
\mathbb{E}(|X_n(t) - X_n(t_1)|)^2 &= n^{-2} \sum_{i=\lfloor nt_1 \rfloor + 1}^{\lfloor nt \rfloor} \theta i + \left(n^{-1}\theta(\lfloor nt \rfloor - \lfloor nt_1 \rfloor)\right)^2 \\
&= \frac{\theta}{2}\left(\frac{\lfloor nt \rfloor}{n}\frac{\lfloor nt \rfloor + 1}{n} - \frac{\lfloor nt_1 \rfloor}{n}\frac{\lfloor nt_1 \rfloor + 1}{n}\right) + \\
&\quad \theta^2\left(\frac{\lfloor nt \rfloor}{n} - \frac{\lfloor nt_1 \rfloor}{n}\right)^2.
\end{aligned}$$

The observation that for all $0 \leq x \leq y \leq 1$

$$y(y + 1/n) - x(x + 1/n) = (y - x)(y + x + 1/n) \leq 3(y - x),$$

produces the bound

$$\mathbb{E}\left(|X_n(t) - X_n(t_1)|\right)^2 \leq 3\left(\theta \vee \theta^2\right)\left(\frac{\lfloor nt \rfloor}{n} - \frac{\lfloor nt_1 \rfloor}{n}\right).$$

This and a similar result for $\mathbb{E}|X_n(t_2) - X_n(t)|$ imply that

$$\begin{aligned}
\mathbb{E}|X_n(t) - X_n(t_1)|^2 |X_n(t_2) - X_n(t)|^2 &= \mathbb{E}|X_n(t) - X_n(t_1)|^2 \mathbb{E}|X_n(t_2) - X_n(t)|^2 \\
&\leq 9(\theta^2 \vee \theta^4)\left(\frac{\lfloor nt_2 \rfloor}{n} - \frac{\lfloor nt \rfloor}{n}\right)\left(\frac{\lfloor nt \rfloor}{n} - \frac{\lfloor nt_1 \rfloor}{n}\right)
\end{aligned}$$

which, by examining the cases $t_2 - t_1 \geq 1/n$ and $t_2 - t_1 < 1/n$ as in the proof of Theorem 16.1 of Billingsley, establishes the bound

$$\mathbb{E}|X_n(t) - X_n(t_1)|^2 |X_n(t_2) - X_n(t)|^2 \leq 36(\theta^2 \vee \theta^4)(t_2 - t_1)^2$$

independently of n.

Now, for every triple t, t_1, t_2 the sequence of variables

$$U_n = X_n(t) - X_n(t_1)$$

is uniformly integrable, as is the sequence of variables

$$V_n = X_n(t_2) - X_n(t),$$

and therefore their expectations converge to the expectations of their corresponding

https://doi.org/10.1017/S0963548396002775 Published online by Cambridge University Press

limiting distributions. Also, since U_n and V_n are mutually independent for all n their limiting distributions are also mutually independent. Using Fatou's lemma for convergence in distribution gives

$$\begin{aligned}\int_{\mathbb{R}^3} |\zeta-\zeta_1|^\gamma |\zeta_2-\zeta|^\gamma d\mu_{t_1tt_2}(\zeta_1,\zeta,\zeta_2) &\leq \liminf_n \mathbb{E}U_n^2 \liminf_n \mathbb{E}V_n^2 \\ &\leq \liminf_n \mathbb{E}U_n^2 \mathbb{E}V_n^2 \\ &\leq 36(\theta^2 \vee \theta^4)(t_2-t_1)^2.\end{aligned}$$

Therefore, condition (40) holds with with $\gamma = 2$, $\alpha = 1$ and $F(t) = 6(\theta \vee \theta^2)t$.

To prove condition (41) let $t \in [0,1]$ and $h > 0$ as in condition (41). For any $\epsilon > 0$ the Portmanteau lemma shows

$$\begin{aligned}\mu_{t,t+h}\{(\zeta_1,\zeta) : |\zeta_1-\zeta_2| \geq \epsilon\} &\leq \liminf_n \mathbb{P}(|X_n(t+h)-X_n(t)| > \epsilon) \\ &\leq \epsilon^{-1} \liminf_n \mathbb{E}|X_n(t+h)-X_n(t)| \\ &= \epsilon^{-1} \liminf_n \mathbb{E}\frac{T_{\lfloor tn\rfloor,\lfloor (t+h)n\rfloor}}{n} \\ &\leq \epsilon^{-1} \liminf_n \bar{\theta}(hn+1)/n \\ &= \epsilon^{-1}\bar{\theta}h.\end{aligned}$$

Letting $h \to 0$ now shows that condition (41) holds.

References

[1] Ahlfors, L. (1979) *Complex Analysis.* McGraw-Hill.

[2] Arratia, R. and Tavaré, S. (1992) The cycle structure of random permutations. *Ann. Probab.* **20** 1567–1591.

[3] Arratia, R. and Tavaré, S. (1994) Independent process approximations for random combinatorial structures. *Adv. Math.* **104** 90–154.

[4] Arratia, R., Barbour, A. D. and Tavaré, S. (1992) Poisson process approximations for the Ewens Sampling Formula. *Ann. Appl. Probab.* **2** 519–535.

[5] Arratia, R., Stark, D. and Tavaré, S. (1995) Total Variation Asymptotics for independent Poisson approximations of the Ewens sampling formula and related assemblies. To appear in *Ann. Probab.*

[6] Billingsley, P. (1968) *Convergence of Probability Measures.* John Wiley & Sons.

[7] Flajolet, P. and Odlyzko, A. (1990) Singularity Analysis of Generating Functions. *SIAM J. Disc. Math.* **3** 216–240.

[8] Graham R. L., Knuth D. E. and Patashnik O. (1989) *Concrete Mathematics.* Addison-Wesley.

https://doi.org/10.1017/S0963548396002775 Published online by Cambridge University Press