On total variation approximations for random assemblies

Eugenijus Manstavičius†

Faculty and Institute of Mathematics and Informatics, Vilnius University, Vilnius, Lithuania

We prove a total variation approximation for the distribution of component vector of a weakly logarithmic random assembly. The proof demonstrates an analytic approach based on a comparative analysis of the coefficients of two power series.

Keywords: Labelled combinatorial structure, set construction, random structure, component vector, analytic method

1 Introduction

The purpose of the present note is to draw reader's attention to an analytic approach proposed by the author in [8] and refined in [9]. In contrast to the celebrated Flajolet-Odlyzko method (see [4]), it allows to obtain asymptotic formulas for the ratio of coefficients of two power series when separately the coefficients do not have a regular asymptotic behavior as their index increases. Moreover, our approach does not require an analytic continuation of the series outside the disk. To be more concrete, we apply the approach to estimate the total variation distance between the distribution of component vector of a random assembly and that of a random vector with independent coordinates. Thus, this note continues the author's investigations carried out in [10] and [11]. The Stein's method applied by R. Arratia, A.D. Barbour and S. Tavaré [1] and the Flajolet-Odlyzko method (see, e.g., J. Hansen [6] and D. Stark [13], [14]) have led to such type results under some regularity conditions satisfied by the coefficients of the involved series. Instead of regularity, we assume only appropriate upper and lower bounds.

For simplicity, we confine ourselves to the decomposable combinatorial structures called *assemblies* which are just set-constructions for labelled combinatorial structures (see [1] and [5]). If $\mathcal{Q}$ is an initial class of structures κ_i, $i \geq 1$, with the exponential generating series

$$Q(z) = \sum_{j\geq 1} \frac{q_j}{j!} z^j,$$

where q_j denotes the number of structures in $\mathcal{Q}$ of size j, then an assembly σ is a finite collection of κ_i relabelled in an appropriate way. So, starting from cycles of points one obtains permutations and

†Email: `eugenijus.manstavicius@mif.vu.lt`.

from a finite relabelled sample of connected graphs one gets a labelled graph. If $\mathcal{A}_n$ denotes the class of assemblies of size n, $\mathcal{A}_0 = \{\emptyset\}$, then all information about enumeration of the structures lays in the formal equality

$$\sum_{n=0}^{\infty} \frac{|\mathcal{A}_n|}{n!} z^n = \exp\big\{Q(z)\big\}.$$

Denote $\ell(\bar{s}) = 1s_1 + \cdots + ns_n$, where $\bar{s} = (s_1, \ldots, s_n) \in \mathbb{Z}_+^n$. Then the component vector $\bar{k}(\sigma) = \big(k_1(\sigma), \ldots, k_n(\sigma)\big)$, where $k_j(\sigma) \geq 0$ is the number of components of size j in σ, satisfies the relation $\ell\big(\bar{k}(\sigma)\big) = n$ for $\sigma \in \mathcal{A}_n$.

To begin the probability theory, one can take each structure $\sigma \in \mathcal{A}_n$ with equal probability or involve some weights depending on the sizes of the components. In this case, given a sequence $w_j \geq 0$, $1 \leq j \leq n$, we may set

$$w(\sigma) = \prod_{j=1}^{n} w_j^{k_j(\sigma)}, \quad 0^0 := 1, \qquad W_n = \sum_{\sigma \in \mathcal{A}_n} w(\sigma),$$

$$\nu_n^{(w)}(\{\sigma\}) = w(\sigma) W_n^{-1}, \quad \sigma \in \mathcal{A}_n,$$

and extend the latter additively on all subsets of $\mathcal{A}_n$. Note that $\nu_n^{(w)}$ belongs to the so-called class of multiplicative measures (see A. M. Vershik [15]). For instance, in the case of permutations, if $w_j \equiv \theta > 0$, some fixed constant, this leads to the Ewens Sampling Formula. Manipulating with the choice $w_j \in \{0,1\}$, in this way we define probabilities on the subsets of permutations with some restrictions on the cycle lengths, that is, we arrive to the so-called Λ-permutations (see [17]). For a great variety of combinatorial assemblies and for some unexplained facts about them, we refer to the book [1].

The component vector $\bar{k}(\sigma)$, $\sigma \in \mathcal{A}_n$, has the following distribution:

$$\nu_n^{(w)}(\bar{k}(\sigma) = \bar{s}) := \nu_n^{(w)}\big(\{\sigma : \ \bar{k}(\sigma) = \bar{s}\}\big) = \mathbf{1}\{\ell(\bar{s}) = n\} \frac{n!}{W_n} \prod_{j=1}^{n} \frac{1}{s_j!} \left(\frac{q_j w_j}{j!}\right)^{s_j},$$

where $\bar{s} = (s_1, \ldots, s_n) \in \mathbb{Z}_+^n$. This leads to the *Conditioning Relation*

$$\nu_n^{(w)}(\bar{k}(\sigma) = \bar{s}) = P\big(\bar{\xi} = \bar{s} | \ell(\bar{\xi}) = n\big), \tag{1}$$

where $\bar{\xi} := (\xi_1, \ldots, \xi_n)$ and $\xi_j, j \geq 1$, are mutually independent Poisson random variables (r.vs) defined on some probability space $\{\Omega, \mathcal{F}, \mathcal{P}\}$ with $\mathbf{E}\xi_j = x^j q_j w_j / j!$, $j \geq 1$, where $x > 0$ is an arbitrary number. If $q_j w_j = 0$ the corresponding ξ_j becomes the degenerated at the zero point r.v. Note that an appropriate choice of x allows to reduce analysis of various classes of structures taken with weighted probabilities to a unified scheme. To simplify the notation, we will use the following definition.

Definition 1 *Let $\mathcal{A}_n$ be a set of assemblies of size $n \geq 1$ and μ_n be a probability measure on $\mathcal{A}_n$. We call the pair $\big(\mathcal{A}_n, \mu_n\big)$ weakly logarithmic if*

$$\mu_n(\bar{k}(\sigma) = \bar{s}) = P\big(\bar{\xi} = \bar{s} | \ell(\bar{\xi}) = n\big), \quad \bar{s} \in \mathbb{Z}_+^n, \tag{2}$$

where $\bar{\xi} = (\xi_1, \ldots, \xi_n)$ is a random vector with mutually independent Poisson r.vs with the mean values $\mathbf{E}\xi_j =: \lambda_j \geq 0$, holds and, for some $n_0 \in \mathbb{N}$ and positive constants Θ, θ, θ_1, the following conditions are satisfied:

$$j\lambda_j \leq \Theta, \quad 1 \leq j \leq n; \tag{3}$$

$$\sum_{j\le u} j\lambda_j \ge \theta u, \quad n_0 \le u \le n; \tag{4}$$

and

$$nP\big(\ell(\bar\xi)=n\big) \ge \theta_1, \quad n \ge n_0. \tag{5}$$

The Definition extends the notion of logarithmic assemblies which are reckoned by the relation $j\lambda_j \sim \theta'$ for some constant $\theta' > 0$ as $j \to \infty$. In this case, we would have asymptotic formulas instead of the bounds (3), (4), and (5). The logarithmic assemblies have been explored in [1], [6], [13] and [14]. The Definition also covers the cases discussed by the author in [10] and [11]. Note that the use of condition (4) for the interval $n_0 \le u \le n$ is motivated by the desire to include the case when λ_j, $1 \le j \le n$, also depend on n. As we shall see later, condition (5) is weaker than that used earlier (see e.g. assumption (6) in [10]), it is a Tauber type requirement of the applied analytic approach.

The main result of this paper is a total variation approximation. Let $\mathcal{L}(X)$ be the distribution of a r.v. X with respect to the appropriate probability measure. Afterwards the index r, $1 \le r \le n$, added to the vectors $\bar k(\sigma)$ and $\bar\xi$ will denote that only the first r coordinates are taken. Let $x_+ = \max\{x, 0\}$ for $x \in \mathbb{R}$ and $\ll$ be an analog of the symbol $\mathrm{O}(\cdot)$. In the sequel, the constants involved in these symbols and the newly introduced ones will depend only on n_0, Θ, θ, and θ_1.

Theorem 1 *Let $n \ge 1$ and $(\mathcal{A}_n, \mu_n)$ be weakly logarithmic. There exists a positive constant α such that*

$$\rho_{TV}\Big(\mathcal{L}\big(\bar k_r(\sigma)\big), \mathcal{L}(\bar\xi_r)\Big) := \sum_{\bar s \in \mathbb{Z}_+^r} \Big(\mu_n\big(\bar k_r(\sigma) = \bar s\big) - P(\bar\xi_r = \bar s)\Big)_+ \ll \Big(\frac{r}{n}\Big)^{\alpha} \tag{6}$$

for $1 \le r \le n$.

Theorem 1 has a wide spectrum of applications. To list but a few, we mention the value distribution problems for additive functions defined on the weakly logarithmic assemblies $\mathcal{A}_n$ which we were exploring in the recent time. Given a double index real sequence $\{h_j(k)\}, j \ge 1, k \ge 0$, such that $h_j(0) \equiv 0$, let us define an additive function $h : \mathcal{A}_n \to \mathbb{R}$ by setting

$$h(\sigma) = \sum_{j \le n} h_j\big(k_j(\sigma)\big).$$

Set $\alpha(n) \in \mathbb{R}$ and $\beta(n) > 0$ for some normalizing sequences.

The first problem concerns one-dimensional asymptotic distributions. One can ask under what conditions posed on $\{h_j(k)\}$ there exist $\alpha(n)$ and $\beta(n) > 0$ such that the distributions

$$\mu_n\big(h(\sigma) - \alpha(n) < x\beta(n)\big)$$

weakly converge as $n \to \infty$ to some limit distribution function. So far, the problem was examined and general sufficient conditions were obtained in the case of logarithmic classes of assemblies (see, for instance, [1] and more recent papers). It is worth to note here that the recent ideas demonstrated in [12] for permutations allow to prove also the necessity of conditions, provided that $\beta(n)$ is slowly oscillating at infinity. Combined with Theorem 1, these ideas yield extensions of the results for broader classes of structures.

The second problem is the invariance principle. Let $n \geq 1$ and $y_n : [0,1] \to \{1, \ldots, n\}$ be an increasing function, $y_n(0) = 0$ and $y_n(1) = n$. By taking the partial sums

$$h(\sigma, u) := \sum_{j \leq u} h_j\big(k_j(\sigma)\big)$$

and

$$H_n := H_n(\sigma, t) := \beta(n)^{-1}\Big(h\big(\sigma, y_n(t)\big) - \alpha\big(y_n(t)\big)\Big), \quad \sigma \in \mathcal{A}_n,$$

we define the paths in the Skorokhod $\mathbb{D}[0,1]$ space. By definition the latter is comprised from real-valued functions on $[0,1]$ which are right-continuous and have left-hand limits at each point; moreover, an appropriate topology is endowed (see [3]). We now ask when the distributions $\mu_n \circ H_n^{-1}$ weakly converge as $n \to \infty$. Despite to the fact that the dependent r.vs $k_j(\sigma), j \leq n$, are involved, if the limit process has independent increments (say, Brownian motion, stable or other Lévy processes), necessary and sufficient convergence conditions were obtained. For this, together with Theorem 1, we applied the technique developed in the cases of permutations (see [2] and the references therein).

Finally, generalizing the results of the paper [11], we have obtained general analogs of the law of iterated logarithm which have the form of the relations:

$$\lim_{m \to \infty} \limsup_{n \to \infty} \mu_n\Big(\max_{m \leq k \leq n} \beta(k)^{-1}\big|h(\sigma, k) - \alpha(k)\big| \geq 1 + \varepsilon\Big) = 0$$

and

$$\lim_{m \to \infty} \liminf_{n \to \infty} \mu_n\Big(\min_{m \leq k \leq n} \big|\beta(k)^{-1}\big(h(\sigma, k) - \alpha(k)\big) - a\big| \leq \varepsilon\Big) = 1$$

holding for every $\varepsilon > 0$ and $a \in [-1, 1]$. The conditions possed on $\{h_j(k)\}$ are close to that used in [11].

The proofs in full detail of the just mentioned probabilistic results will be published elsewhere. In the present note, we concentrate only on the proof of Theorem 1 and complete it in the last Section. Section 2, where the necessary analytic technique is exposed, is the main part of the paper.

2 The Analytic Part

To motivate the below presented asymptotic analysis and to make some hint on the proof of Theorem 1, we recall the following lemma reducing the problem to a one-dimensional case.

For $\bar{s} = (s_1, \ldots, s_n)$, we set $\ell_{ij}(\bar{s}) = (i+1)s_{i+1} + \cdots + js_j$ if $0 \leq i < j \leq n$. Moreover, let $\ell_r(\bar{s}) := \ell_{0r}(\bar{s})$, where $1 \leq r \leq n$. Then $\ell_n(\bar{s}) = \ell(\bar{s})$.

Lemma 1 *We have*

$$\rho_{TV}\Big(\mathcal{L}\big(\bar{k}_r(\sigma)\big), \mathcal{L}\big(\bar{\xi}_r\big)\Big) = \rho_{TV}\Big(\mathcal{L}\big(\bar{\xi}_r \big| \ell(\bar{\xi}) = n\big), \mathcal{L}\big(\bar{\xi}_r\big)\Big)$$
$$= \sum_{m \in \mathbb{Z}_+} P\big(\ell_r(\bar{\xi}) = m\big)\bigg(1 - \frac{P\big(\ell_{rn}(\bar{\xi}) = n - m\big)}{P\big(\ell(\bar{\xi}) = n\big)}\bigg)_+ . \tag{7}$$

Proof. See [1], p. 69.

One may expect that asymptotic formulas for either of the two local probabilities in the ratio on the right-hand side could be found by analyzing the Cauchy integral in the expression

$$P\big(\ell_{rn}(\bar{\xi}) = m\big) = \frac{1}{2\pi i}\int_{|z|=1} \frac{1}{z^{m+1}} \exp\Big\{\sum_{r<j\le n} \lambda_j (z^j - 1)\Big\} \mathrm{d}z. \tag{8}$$

Indeed, this has been the case dealing with the logarithmic assemblies (see [1], [13], and [14]). It appears that, under the weaker conditions listed in the Definition, the probabilities do not show asymptotic regularity; nevertheless, their ratio converges to unit uniformly in a large region for r and m. This is the idea of the proof of Theorem 1.

Now let us explore the sequences of functions

$$D(z) := \exp\Big\{\sum_{j\le n} \frac{d_j}{j} z^j\Big\} =: \sum_{k=0}^{\infty} D_k z^k \tag{9}$$

and

$$F(z) := \exp\Big\{\sum_{r<j\le n} \frac{d_j}{j} z^j\Big\} =: \sum_{k=0}^{\infty} F_k z^k,$$

where $d_j \ge 0$, $1 \le j \le n$, and $1 \le r \le n$ are arbitrary numbers, maybe, depending on n. Evidently, $F_0 = 1$ and $F_k = 0$ if $1 \le k \le r$. We seek an asymptotic formula for the ratio F_m/D_n as $n \to \infty$ with remainder term estimates holding uniformly for

$$0 \le r/n \le \delta, \qquad (1-\eta)n \le m \le n, \tag{10}$$

where δ and η are nonnegative parameters such that $\delta + \eta \le 1$. We now assume the following conditions:

$$d_j \le \Theta, \quad 1 \le j \le n; \tag{11}$$

$$\sum_{j\le u} d_j \ge \theta u, \quad n_0 \le u \le n; \tag{12}$$

and

$$nD_n \ge \theta_1 D(1), \quad n \ge n_0. \tag{13}$$

Having in mind the forthcoming application with $d_j = j\lambda_j$, we intentionally leave the same positive constants n_0, Θ, θ, and θ_1 appearing in (3), (4), and (5).

In the more traditional way, (see, for instance, [1], [6], [13] and [14] and the references therein), such a problem would have involved infinite series in the exponents. Avoiding conditions on their convergence in some disk and on the type of singularities on the boundary, stressing that we are allowing dependence on n for $D(z)$ and its coefficients, we have intentionally cut the terms over $j > n$.

In the just mentioned works, to obtain an asymptotic formula for F_m, some analytic properties of the generating functions or an asymptotic regularity of d_j as $j \to \infty$ were needed. They also implied an appropriate asymptotic regularity of D_n as $n \to \infty$. We now seek an asymptotic formula for the ratio when, instead of regularity conditions, lower and upper bounds are assumed. The idea to examine such a problem goes back to author's paper [8] where the condition that $d_j \asymp 1$ for $1 \le j \le n$ was used. Here and in the sequel, $a \asymp b$ means that $a \ll b$ and $b \ll a$. However, dealing with subsets of assemblies, the lower bound of d_j, $1 \le j \le n$, has appeared to be stringent. The present conditions (11), (12), and (13) open a broader perspective for applications. Finally, note that condition (11) implies $D_n \ll n^{\Theta-1}$ (see, for instance, [7]) and that the estimate $D_n \gg n^{\theta-1}$ follows from (12) and (13).

For brevity, set

$$e_r := \exp\bigg\{ -\sum_{j\le r} \frac{d_j}{j} \bigg\}.$$

Afterwards, all newly involved constants, if not stated otherwise, depend only on n_0, Θ, θ, and θ_1. Moreover, we assume that $n \to \infty$. We now formulate our analytic result.

Proposition 1 *If conditions* (11), (12), *and* (13) *are satisfied, then there exist constants* $0 < \delta_0 \le 1/2$, $C \ge 2$, *and* $0 < c < 1$ *such that*

$$F_m = e_r D_n \Big(1 + \mathrm{O}(\delta^c + \eta\delta^{-1} + r(\delta n)^{-1} \Big) \tag{14}$$

uniformly in region (10) *for all* δ *and* η *such that* $C/n \le \delta \le \delta_0$ *and* $0 \le \eta \le \delta$.

The result is nontrivial if $\delta^c + \eta\delta^{-1} + r(\delta n)^{-1}$ is less than some constant depending on that hidden in $\mathrm{O}(\cdot)$. For instance, if $\eta + (r/n) \le \delta^2$ and δ_0 is small enough, we have the lower and upper bounds, i.e. $F_m/e_r D_n \asymp 1$. Taking δ and η small and appropriately depending on $r/n \to 0$, from (14), we obtain the needed asymptotic formula of the type $F_m/e_r D_n = 1+$ Remainder. This is done in the proof of Theorem 1 given at the end of the note.

The proof of Proposition 1 is based on the ideas exposed in [9] and [16]. It begins with a rather typical splitting of the Cauchy integral

$$F_m = \frac{1}{2\pi i} \bigg(\int_{\Delta_0} + \int_{\Delta} \bigg) \frac{F(z)}{z^{m+1}} \,\mathrm{d}z =: J_0 + J,$$

where $\Delta_0 = \{z = \mathrm{e}^{i\tau} : |\tau| \le K/n\}$, $\Delta = \{z = \mathrm{e}^{i\tau} : K/n < |\tau| \le \pi\}$, and $1 \le K \le n$ is a parameter to be chosen later. As usual, the integral J along the large arc Δ is estimated. For this, we combine estimates of trigonometric polynomials together with Parseval's equality. In general, the assumed conditions are insufficient to integrate J_0, therefore we compare it with an analogous integral where the integrand has $D(z)$ in the place of $F(z)$. This allows us to compare F_m and D_n. Some useful lemmata have been proved in [10] under conditions holding now and listed in the proposition.

Lemma 2 *Let* $b_k \ge 0$ *for* $k \ge 0$ *and*

$$B(x) := \sum_{k=0}^{\infty} b_k x^k$$

converge in $x \in [0, 1)$. *Then*

$$\sum_{k\le s} b_k \le \mathrm{e}B(\mathrm{e}^{-1/s}), \quad s \ge 0.$$

Moreover, if

$$\frac{B'(x)}{B(x)} \le \frac{c_1}{1-x}$$

for all $0 \le x < 1$ *with a constant* $c_1 > 0$*, then*

$$\sum_{k \le s} b_k \gg B(\mathrm{e}^{-1/s})$$

for $s \ge 1$ *with a constant in* $\gg$ *depending on* c_1 *only.*

Proof. See Section 2.1 of [16].

This lemma gives another interpretation of condition (13).

Corollary 1 *If condition* (11) *is satisfied, then condition* (13) *is equivalent to*

$$D_n \gg \frac{1}{n} \sum_{k=0}^{n-1} D_k.$$

Proof. It suffices to combine the estimates in Lemma 2 with the recurrence relation

$$D_n = \frac{1}{n} \sum_{k=0}^{n-1} d_{n-k} D_k,$$

which can be obtained from (9) differentiating $D(z)$.

Lemma 3 *If conditions* (11) *and* (12) *are satisfied, then there exist positive constants* $\delta_0 < 1$ *and* $C \ge 2$ *such that*

$$\max_{T \le t \le \pi} \left| F(\mathrm{e}^{it}) \right| \ll e_r D(1) \delta^{c_2}$$

for $T := C(\delta n)^{-1}$ *and* $C/n \le \delta \le \delta_0$ *with some* $c_2 = c_2(C, \delta_0) > 0$*.*

Proof. This is Lemma 3 in [10].

Now we can complete the first task.

Lemma 4 *Let* $K := C/\delta$ *and* $C/n \le \delta \le \delta_0$*, where* C *and* δ_0 *are defined in Lemma* 3*. If conditions* (11), (12)*, and* (13) *are satisfied, then there exists a positive constant* c_3 *such that*

$$J \ll e_r D_n \delta^{c_3}$$

for $n/2 \le m \le n$*.*

Proof. Integrating by parts and using Lemma 3, we obtain

$$\begin{aligned} J &= \frac{1}{2\pi i m} \int_{\Delta} \frac{F(z)}{z^m} \Bigg(\sum_{r < j \le n} d_j z^{j-1} \Bigg) \, \mathrm{d}z + O\big(e_r D_n \delta^{c_2}\big) \\ &= \frac{1}{2\pi m} \Bigg(\sum_{\substack{r < j \le n \\ |m-j| \le L}} + \sum_{\substack{r < j \le n \\ |m-j| > L}} \Bigg) d_j \int\limits_{K/n \le |\tau| \le \pi} F(\mathrm{e}^{i\tau}) \mathrm{e}^{i\tau(j-m)} \, \mathrm{d}\tau \\ &\quad + O\big(e_r D_n \delta^{c_2}\big) =: \frac{1}{2\pi m} \sum_{\substack{r < j \le n \\ |m-j| \le L}} d_j I(m-j) + I + O\big(e_r D_n \delta^{c_2}\big), \end{aligned} \tag{15}$$

where $r \le L := \varepsilon n > n/K$ and $\delta/C < \varepsilon < 1$ to be specified later.

Firstly, we observe that, for $t \in \mathbb{N}$ and $r < t \le n$, by Lemma 2,

$$\begin{aligned} tF_t &= \sum_{r<j\le t} d_j F_{t-j} \le \Theta \sum_{k\le t} F_k \ll F(\mathrm{e}^{-1/t}) \asymp e_r D(\mathrm{e}^{-1/t}) \\ &\asymp e_r \exp\Big\{ \sum_{j\le t} \frac{d_j}{j} \Big\} = e_r D(1) \exp\Big\{ -\sum_{t<j\le n} \frac{d_j}{j} \Big\} \\ &\ll e_r n D_n (t/n)^\theta. \end{aligned} \tag{16}$$

Here, in the last step, we have applied the estimate

$$\sum_{t\le j\le n} \frac{d_j}{j} \ge \theta \log \frac{n}{t} - \Theta.$$

following from (11) and (12) via summation by parts. Secondly,

$$\begin{aligned} I(l) &= \sum_{k=0}^{\infty} F_k \int\limits_{K/n<|\tau|\le\pi} \mathrm{e}^{i(k-l)\tau}\,\mathrm{d}\tau \\ &= (2\pi - 2K/n)\mathbf{1}\{l \ge 0\} F_l - 2 \sum_{k\ge 0, k\ne l} F_k \frac{\sin \frac{K}{n}(k-l)}{k-l}. \end{aligned} \tag{17}$$

Hence

$$\begin{aligned} \frac{1}{m} \sum_{\substack{r<j\le n \\ |m-j|\le L}} d_j I(m-j) &\ll \frac{1}{n} \sum_{0\le l\le L} F_l + \frac{1}{n} \sum_{|l|\le L} \sum_{|k-l|>|l|/2} \frac{F_k}{|k-l|} \\ &+ \frac{1}{n} \sum_{1\le l\le L} \sum_{1\le|k-l|\le l/2} F_k \frac{|\sin(\frac{K}{n}(k-l))|}{|k-l|} =: S_1 + S_2 + S_3. \end{aligned} \tag{18}$$

Now, applying (16), we obtain

$$S_1 \ll (1/n) F(\mathrm{e}^{-1/L}) \ll e_r D_n \varepsilon^\theta$$

and

$$\begin{aligned} S_2 &\ll \frac{1}{n} \sum_{1\le|l|\le L} \Big(\frac{1}{|l|} \sum_{k<|l|/2} F_k + \sum_{(3/2)|l|<k\le n} \frac{F_k}{k} + \frac{1}{n} \sum_{k>n} F_k \Big) \\ &\ll e_r D_n \sum_{1\le l\le L} \frac{1}{l} \Big(\frac{l}{n}\Big)^\theta + \frac{1}{n} \sum_{k\le(3/2)L} F_k + \varepsilon \sum_{(3/2)L<k\le n} \frac{F_k}{k} + \varepsilon e_r D_n \\ &\ll \varepsilon^{\theta\wedge 1} e_r D_n + \varepsilon e_r n D_n \sum_{(3/2)L<k\le n} \frac{1}{k^2} \Big(\frac{k}{n}\Big)^\theta \\ &\ll \varepsilon^{\theta\wedge 1} \log(1/\varepsilon) e_r D_n. \end{aligned}$$

Here and afterwards, $a \wedge b := \min\{a, b\}$ for $a, b \in \mathbb{R}$.

Estimating the trigonometric sum S_3, together with (16), we also exploit the property $D(\mathrm{e}^{-1/k}) \asymp D(\mathrm{e}^{-1/l})$ if $|k-l| \le l$. This leads to

$$
\begin{aligned}
S_3 &\ll \frac{K}{n^2} \sum_{1 \le l \le L} \sum_{\substack{|k-l| \le l/2 \\ |k-l| \le n/K}} F_k + \frac{1}{n} \sum_{1 \le l \le L} \sum_{\substack{|k-l| \le l/2 \\ |k-l| > n/K}} \frac{F_k}{|k-l|} \\
&\ll \frac{Ke_r}{n^2} \sum_{1 \le l \le L} \sum_{\substack{|k-l| \le l/2 \\ |k-l| \le n/K}} \frac{D(\mathrm{e}^{-1/k})}{k} + \frac{e_r}{n} \sum_{1 \le l \le L} \sum_{\substack{|k-l| \le l/2 \\ |k-l| > n/K}} \frac{D(\mathrm{e}^{-1/k})}{k|k-l|} \\
&\ll \frac{Ke_r}{n^2} \sum_{1 \le l \le L} \frac{D(\mathrm{e}^{-1/l})}{l} \cdot \frac{n}{K} + \frac{e_r}{n} \sum_{1 \le l \le L} \frac{D(\mathrm{e}^{-1/l})}{l} \sum_{\substack{|k-l| \le l/2 \\ |k-l| > n/K}} \frac{1}{|k-l|} \\
&\ll \frac{e_r}{n} \sum_{1 \le l \le L} \frac{D(\mathrm{e}^{-1/l})}{l} \Big(1 + \log\Big(2 + \frac{lK}{n} \Big) \Big) \\
&\ll \Big(\log \frac{1}{\delta} \Big) e_r D_n \sum_{1 \le l \le L} \frac{1}{l} \Big(\frac{l}{n} \Big)^{\theta} \ll \Big(\log \frac{1}{\delta} \Big) e_r D_n \varepsilon^{\theta}.
\end{aligned}
$$

Now, inserting the estimates of S_i, $i = 1, 2, 3$, into (18), we obtain

$$
\frac{1}{m} \sum_{\substack{r < j \le n \\ |m-j| \le L}} d_j I(m-j) \ll \varepsilon^{\theta \wedge 1} (\log(1/\delta)) e_r D_n. \tag{19}
$$

To estimate the term I in (15), we again use integration by parts and Lemma 3. So applying Cauchy's inequality, we obtain

$$
\begin{aligned}
I &\ll \frac{1}{n} \sum_{L < |l| \le n} \frac{1}{|l|} \Big| \int\limits_{K/n \le |\tau| \le \pi} F'(\mathrm{e}^{i\tau}) \mathrm{e}^{-i(l-1)\tau} \, \mathrm{d}\tau \Big| + \frac{1}{n} \sum_{L < |l| \le n} \frac{1}{|l|} \max_{z \in \Delta} |F(z)| \\
&\ll \frac{1}{n\sqrt{L}} \Big(\sum_{1 \le |l| \le n} \Big| \int\limits_{K/n \le |\tau| \le \pi} F'(\mathrm{e}^{i\tau}) \mathrm{e}^{-i(l-1)\tau} \, \mathrm{d}\tau \Big|^2 \Big)^{1/2} \\
&\qquad + e_r D_n (\log(1/\varepsilon)) \delta^{c_2}.
\end{aligned}
$$

The integrals under the last sum are just the Fourier coefficients of an appropriate function; therefore, via Parseval's identity, we further have

$$
\begin{aligned}
I &\ll \frac{1}{n\sqrt{L}} \left(\int\limits_{K/n \le |\tau| \le \pi} |F'(\mathrm{e}^{i\tau})|^2 \, \mathrm{d}\tau \right)^{1/2} + e_r D_n (\log(1/\varepsilon)) \delta^{c_2} \\
&\ll \frac{1}{n\sqrt{L}} \left(\int\limits_{K/n \le |\tau| \le \pi} |F(\mathrm{e}^{i\tau})|^2 \Big| \sum_{r < j \le n} d_j \mathrm{e}^{i\tau j} \Big|^2 \mathrm{d}\tau \right)^{1/2}
\end{aligned}
$$

$$+ \quad e_r D_n (\log(1/\varepsilon))\delta^{c_2}.$$

By Lemma 3 and Parseval's identity again, we obtain

$$\begin{aligned} I &\ll e_r \delta^{c_2} \frac{D(1)}{n\sqrt{L}} \left(\int\limits_{|\tau|\le\pi} \Big| \sum_{j\le n} d_j \mathrm{e}^{i\tau j} \Big|^2 \mathrm{d}\tau \right)^{1/2} + e_r D_n (\log(1/\varepsilon))\delta^{c_2} \\ &\ll e_r D_n \varepsilon^{-1/2} \delta^{c_2}. \end{aligned}$$

Inserting this and (19) into (15), by taking $\varepsilon = \delta^{c_2/2}$, we easily complete the proof of the lemma:

$$J \ll \big(\varepsilon^{1\wedge\theta} \log(1/\delta) + \varepsilon^{-1/2}\delta^{c_2}\big) e_r D_n \ll \delta^{c_3} e_r D_n,$$

where $c_3 = (1 \wedge 2\theta)c_2/4$.

Corollary 1 *If conditions* (11), (12) *and* (13) *are satisfied,* $K = C/\delta$ *and* $C/n \le \delta \le \delta_0$*, then*

$$D_n = \big(1 + \mathrm{O}(\delta^{c_3})\big) \frac{1}{2\pi i} \int_{\Delta_0} \frac{D(z)}{z^{n+1}} \mathrm{d}z.$$

Proof. Apply Lemma 4 with $r = 0$ and $m = n$.

Proof of Proposition 1. Let $n \ge n_0$. It suffices to compare J_0 with the integral in Corollary 1. Recall the choice $K = C/\delta$. If $(1-\eta)n \le m \le n$, $0 \le \eta \le \delta$, and $z \in \Delta_0$, then

$$z^{-m} = z^{-n}\big(1 + \mathrm{O}(\eta\delta^{-1})\big).$$

Further,

$$\begin{aligned} F(z) &= e_r D(z) \exp\Big\{ \sum_{j\le r} \frac{d_j}{j}(1 - z^j) \Big\} \\ &= e_r D(z)\big(1 + \mathrm{O}(r/(\delta n)\big) \end{aligned}$$

uniformly in $z \in \Delta_0$. Hence by Lemma 4 and Corollary 1,

$$F_m = J_0 + \mathrm{O}(\delta^{c_3} e_r D_n) = e_r D_n\big(1 + \mathrm{O}(\delta^{c_3} + \eta\delta^{-1} + r(\delta n)^{-1}\big)$$

provided that $C/n \le \delta \le \delta_0$, $0 \le \eta \le \delta$, $0 \le r \le \delta n$, and $n \ge n_0$.

Since for $n \le n_0$ the Proposition trivially holds, we have finished its proof.

3 Proof of Theorem 1

We apply Proposition 1 with $d_j = j\lambda_j$. Keeping the previous notation, we see that

$$P\big(\ell_{rn}(\bar{\xi}) = n - m\big) = \frac{F_{n-m}}{e_r D(1)}, \qquad P\big(\ell(\bar{\xi}) = n\big) = \frac{D_n}{D(1)}.$$

Let n_0, C, δ_0, and c be as in Proposition 1. We may assume that $c \leq 1/3$. Take $\eta = (r/n)^{1/2}$, $\delta = (r/n)^{1/2(1+c)}$, $r \leq \delta_0^{2(1+c)} n =: c_4 n$, and $n \geq n_1 \geq n_0$ sufficiently large to assure $C/n \leq \delta \leq \delta_0$. For this choice, we obtain

$$\frac{P\big(\ell_{rn}(\bar{\xi}) = n - m\big)}{P\big(\ell(\bar{\xi}) = n\big)} - 1 \ll \Big(\frac{r}{n}\Big)^{c/2(1+c)}$$

uniformly in $0 \leq m \leq \eta n = \sqrt{rn}$.

The summands over $m > \sqrt{rn}$ in (7) contribute not more than

$$(rn)^{-1/2}\mathbf{E}\ell_r(\bar{\xi}) = (rn)^{-1/2} \sum_{j \leq r} j\lambda_j \leq \Theta(r/n)^{1/2}.$$

Hence, by (7), we obtain

$$\rho_{TV}\Big(\mathcal{L}\big(\bar{k}_r(\sigma)\big), \mathcal{L}(\bar{\xi}_r)\Big) \ll (r/n)^{c/2(1+c)},$$

provided that $1 \leq r \leq c_4 n$. Since the claim of Theorem 1 is trivial for $c_4 n < r \leq n$ or $n \leq n_1$, we complete the proof.

Acknowledgements

The author thanks the referees for their benevolent comments which have helped to improve the exposition of the paper.

References

[1] R. Arratia, A. D. Barbour and S. Tavaré, *Logarithmic Combinatorial Structures: a Probabilistic Approach*, EMS Monographs in Mathematics, EMS Publishing House, Zürich, 2003.

[2] G. J. Babu and E. Manstavičius, Limit processes with independent increments for the Ewens sampling formula, *Ann. Inst. Statist. Math.*, **54**, 607–620 (2002).

[3] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, New York, 1968.

[4] P. Flajolet and A. M. Odlyzko, Singularity analysis of generating functions, *SIAM J. Discrete Math.*, **3**(2), 216–240 (1990).

[5] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, 2009.

[6] J. Hansen, Order statistics for decomposable combinatorial structures, *Random Structures and Algorithms*, **5**, 517–533 (1994).

[7] A. Hildebrand and G. Tenenbaum, On some Tauberian theorems related to the prime number theorem, *Compositia Math.*, **90** 315–349 (1994).

[8] E. Manstavičius, Mappings on decomposable combinatorial structures: analytic approach, *Combinatorics, Probab. Computing*, **11**, 61–78 (2002).

[9] E. Manstavičius, An analytic method in probabilistic combinatorics, *Osaka J. Math.*, **46**, 273–290 (2009).

[10] E. Manstavičius, Total variation approximation and a functional limit theorem, *Monatshefte Math.*, **161**, 313–334 (2010).

[11] E. Manstavičius, Strong convergence on weakly logarithmic combinatorial assemblies, *Discrete Math.*, **311**, 463–477 (2011).

[12] E. Manstavičius, A limit theorem for additive functions defined on the symmetric group, *Lithuanian Math. J.*, **51**, 220–232 (2011).

[13] D. Stark, Explicit limits of total variation distance in approximations of random logarithmic assemblies by related Poisson processes, *Combinatorics, Probab. Computing*, **6**, 87–105 (1997).

[14] D. Stark, Total variation asymptotics for related Poisson process approximations of random logarithmic assemblies, *Combinatorics, Probab. Computing*, **8**, 567–598 (1999).

[15] A. M. Vershik, Statistical mechanics of combinatorial partitions, and their limit configurations. *Funct. Anal. Appl.* 30, 90-105 (1996).

[16] V. Zacharovas, Voronoi summation formulae and multiplicative functions, *Ramanujan J.*, **24**, 289–329 (2011).

[17] A. L. Yakymiv, On permutations with the cycle lengths in a given set, *Discrete Math.*, **1**, 125–134 (1989) (Russian).