ADVANCES IN APPLIED MATHEMATICS **18,** 432–488 (1997)
ARTICLE NO. AM960523

On a Likely Shape of the Random Ferrers Diagram*

Boris Pittel

Department of Mathematics, Ohio State University, Columbus, Ohio 43210-1174

We study the random partitions of a large integer n, under the assumption that all such partitions are equally likely. We use Fristedt's conditioning device which connects the parts (summands) distribution to the one of a g-sequence, that is, a sequence of independent random variables, each distributed geometrically with a size-dependent parameter. Confirming a conjecture made by Arratia and Tavaré, we prove that the joint distribution of counts of parts with size at most $s_n \ll n^{1/2}$ (at least $s_n \gg n^{1/2}$, resp.) is close—in terms of the total variation distance—to the distribution of the first s_n components of the g-sequence (of the g-sequence minus the first $s_n - 1$ components, resp.). We supplement these results with the estimates for the middle-sized parts distribution, using the analytical tools revolving around the Hardy–Ramanujan formula for the partition function. Taken together, the estimates lead to an asymptotic description of the random Ferrers diagram, close to the one obtained earlier by Szalay and Turán. As an application, we simplify considerably and strengthen the Szalay–Turán formula for the likely degree of an irreducible representation of the symmetric group S_n. We show further that both the size of a random conjugacy class and the size of the centraliser for every element from the class are doubly exponentially distributed in the limit. We prove that a continuous time process that describes the random fluctuations of the diagram boundary from the deterministic approximation converges to a Gaussian (non-Markov) process with continuous sample path. Convergence is such that it implies weak convergence of every integral functional from a broad class. To demonstrate applicability of this general result, we prove that the eigenvalue distribution for the Diaconis–Shahshahani card-shuffling Markov chain is asymptotically Gaussian with zero mean, and variance of order $n^{-3/2}$. © 1997 Academic Press

* This research was supported through a partial reduction of teaching duties in Spring of 1995 and 1996, and by a grant from the NSA in Summer of 1996.

0196-8858/97 $25.00

INTRODUCTION

The problems concerning enumeration of various integer partitions of a given integer n have long occupied a central place in analytic number theory and enumerative combinatorics. The many reasons for such a preeminent role are made abundantly clear, for instance, in Andrews [1] and Stanley [26]. So it is all the more surprising that a pioneering work on probabilistic aspects of the partitions done by Erdös and Lehner in 1941 [7] appeared not so long ago. (During a recent conference in honor of Herb Wilf, the paper was mentioned by several speakers, and the participants were happy to see both the authors present.) Erdös and Lehner introduced a notion of random partition, postulating that every sample partition is assigned the same probability, $1/p(n)$, that is, where $p(n)$ is the total number of such partitions. Using the Hardy–Ramanujan asymptotic formula for $p(n)$, the authors found the limiting distribution of the number of summands (parts) and thus—by a classic duality—the limiting distribution of the largest summand. Afterward there appeared to be a lull until Erdös and Turán [9], Szalay and Turán [27–29], and Erdös and Szalay [8]. To a large extent, the three authors were motivated by the very natural connections that exist between the partitions and the symmetric group S_n of permutations on n letters. One such connection is a bijection between the integer partitions of n on one side and the set of the conjugacy classes of S_n on another side. A related, but considerably deeper, bijection exists between the partitions and the irreducible representations of S_n; see Ledermann [21], Diaconis [5], for instance. Among the host of results in the Hungarian series, we would specifically mention two asymptotic formulas, one for the centralizer size of an element from a random conjugacy class, in Erdös and Turán [9], another for the likely degree of randomly chosen irreducible representation, in Szalay and Turán [29]. The latter is based on the original formula of Frobenius [12], and its linear term factor depends on the value of a very complicated double integral. More recent advances were made by Wilf [34], who found a surprisingly simple derivation for the expected number of distinct sizes, and by Goh and Schmutz [14], who were able to show that this number is asymptotically Gaussian. Fristedt [11] undertook a systematic study of the random partitions that very fruitfully combined analytic and probabilistic tools. He introduced a conditioning device, conceptually analogous to the one by Shepp and Lloyd [25] for the cycles of a random permutation. (An alternative conditioning scheme for permutations, mappings, forests, and allocations has been championed by Kolchin [20] since the late sixties.) Here is Fristedt's approach. Let X_j denote the number of parts in the random partition of n $(j \geq 1)$. Let $\{Z_j\}_{j\geq 1}$ be a sequence of independent random variables such that, for some $q \in (0,1)$,

$$\mathrm{P}\{Z_j = i\} = (1 - q^j)(q^j)^i, \qquad i \geq 0;$$

so Z_j is geometric, with parameter q^j. It turns out that, conditioned on the event $\{\sum_j jZ_j = n\}$, the sequences $\{X_j\}$ and $\{Z_j\}$ are equidistributed, regardless of an actual value of q. Intuitively, one should choose q which maximizes $\mathrm{P}\{\sum_j jZ_j = n\}$, and Fristedt shows that $q = e^{-cn^{-1/2}}$ ($c := \pi/\sqrt{6}$), is almost optimal. (It is certainly optimal in terms of simplicity!) On the basis of this connection with $\{Z_j\}$, Fristedt was able to show that the Prohorov distance between $\{X_j\}_{j \leq t_n}$ and $\{Z_j\}_{j \leq t_n}$ approaches 0 if $t_n = o(n^{1/4})$. Besides having a significant heuristic value, this result represents a considerable strengthening of the Erdös–Lehner theorem. Also, the joint distribution of the $o(n^{1/4})$ largest parts turned out to be asymptotically close to the one of a certain time-homogeneous Markov chain. The fact that the first $o(n^{1/4})$ parts counts are asymptotic to the first $o(n^{1/4})$ Z_j's is close in spirit to the "independent process approximation"-type results obtained for the Ewens distributed permutations. (The uniformly random permutation is in this class.) The reader would do well to consult an expository paper by Arratia and Tavaré [2], who have initiated much of the research in that area. For a general class of combinatorial structures, which includes the random partitions, Arratia and Tavaré managed to obtain a surprisingly simple formula for the total variation distance (d_{TV}) between (the distributions of) the parts counts for the random structure and the independent variables that produce those counts, distribution-wise, upon being conditioned on their total weight value. Using an ingenious heuristic reasoning, Arratia and Tavaré formulated a general conjecture as to when one should expect that distance to be asymptotically small. (For the Ewens distributed permutations, Arratia, Stark, and Tavare [3] obtained the asymptotics that come tantalizingly close to confirming the conjecture of the two senior coauthors.) In a particular case of the random partition the conjecture states: given integer functions $j_1 = j_1(n)$, $j_2 = j_2(n)$, define

$$\mathbf{X}_1 = \{X_j\}_{j \leq j_1}, \qquad \mathbf{Z}_1 = \{Z_j\}_{j \leq j_1},$$
$$\mathbf{X}_2 = \{X_j\}_{j \geq j_2}, \qquad \mathbf{Z}_2 = \{Z_j\}_{j \geq j_2};$$

then

$$d_{\mathrm{TV}}(\mathbf{Z}_1, \mathbf{X}_1) \to 0, \qquad \text{iff } j_1/n^{1/2} \to 0;$$
$$d_{\mathrm{TV}}(\mathbf{Z}_2, \mathbf{X}_2) \to 0, \qquad \text{iff } j_2/n^{1/2} \to \infty.$$

Such a result—if true—would certainly allow one, for instance, to get an asymptotic distributional result for the first $o(n^{1/2})$ parts counts, as compared to $o(n^{1/4})$ in Fristedt [11]. Similarly, the counts of large parts, with size $\gg n^{1/2}$, would be nicely approximated by the independent variables Z_j, $j \gg n^{1/2}$. Of course, one would still have to fill the gap, that is, to study the parts of size $O(n^{1/2})$, to be able to describe an overall behavior of $\{X_j\}_{j\geq 1}$. In a nutshell, that is what the goal of the present paper is.

Leaving the complete formulations for the main body of the paper, here is an essence of our results. In Section 1 (Theorem 1) we use the Arratia–Tavaré formula to show that if there exist finite $a_i = \lim j_i/n^{1/2}$ then

$$d_{\mathrm{TV}}(\mathbf{Z}_i, \mathbf{X}_i) \to d_{\mathrm{TV}}(\mathscr{N}_i, \mathscr{N}), \qquad i = 1, 2.$$

Here $\mathscr{N}, \mathscr{N}_1, \mathscr{N}_2$ are three normals, with zero means and variances 1, $\sigma_1^2(a_1)$, $\sigma_2^2(a_2)$, respectively. We also find the approximations for the cases $a_1 = 0$, $a_2 = \infty$. The result fully confirms the Arratia–Tavare conjecture. The most technical step is the proof of a local limit theorem, with a remainder term, for $\sum_{j>j_1} jZ_j$, $\sum_{j<j_2} jZ_j$. This theorem, coupled with the Chernoff-type estimates and the optional stopping time theorem for an exponential martingale, allows us to obtain sharp probabilistic estimates for the total count $\tilde{X}(k)$ of the small parts of size k at least, small meaning $o(n^{1/2})$ $(k \ll n^{1/2})$, and the total count $X(k)$ of the large parts of size k at least, $k \gg n^{1/2}$, Lemmas 2 and 3. In Section 2, using the techniques from analytic number theory, we prove a proposition which basically asserts that, for $n^{\delta} \leq k_1 \leq k_2 \leq \infty$ $(\delta \in (\frac{3}{8}, \frac{1}{2}])$, the number of summands with sizes ranging from k_1 to k_2 is "sub-Gaussian." Actually a precursor of the limit theorem in the last section, the proposition is used in Section 2 to get the estimates for the parts counts $X(k)$ for k filling the midrange, i.e., at least n^{δ} and at most $n^{1/2}\log n$, Lemma 3. Combination of Lemmas 1–3 yields Theorem 2, which describes the probabilistic bounds of the counts $X(k)$, whence (by duality) those of the size ordered partition parts λ_k, for all essential values of k. (The bounds are analogous to those obtained by Szalay and Turán [27–29], but our methods appear to be much less technical.) As was observed by Vershik [33], the Szalay–Turán bounds show that, loosely speaking, the partition parts λ_k with high probability (whp) closely satisfy a deterministic equation

$$\exp\left(-\frac{c\lambda_k}{\sqrt{n}}\right) + \exp\left(-\frac{ck}{\sqrt{n}}\right) = 1, \qquad (*)$$

at least for "moderate" k's. (In 1952, Temperley [30] used the random partitions to model a crystal growing process. A heuristic argument based on statistical physics concepts led him to the equation $(*)$!) We use the bounds from Theorem 2 and the hook formula for the degree $d(\lambda)$ of the irreducible representation (Frame, Robinson, Thrall [10]), related to an integer partition λ of n, and show (Theorem 3) that

$$\log d(\lambda) = \log\sqrt{n!} - An + O_p\big(n^{3/4}\log^{3/2} n\big),$$

where

$$A = \frac{3}{2} - \gamma + \log c^{-1} - c^{-2}\sum_{j\ge 1}\frac{\log j}{j^2}, \qquad c := \frac{\pi}{\sqrt{6}},$$

and γ is the Euler constant. (The symbol $O_p(f(n))$ stands for $Y_n f(n)$, where Y_n is a random variable bounded in probability.) This significantly simplifies the linear term of the Szalay–Turán formula, and improves their remainder estimate, which was $O_p(n^{7/8}\log^4 n)$. We conjecture that the remainder is asymptotically Gaussian, with zero mean, and standard deviation of order $n^{3/4}$. We demonstrate the power of Theorems 1 and 2, proving (Theorem 4) that the centraliser size ξ for an element of a random conjugacy class satisfies

$$\log \xi = \frac{n^{1/2}\log^2 n}{4c} + \frac{n^{1/2}\log n}{2c}(X_n - \gamma - 2\log c),$$

where

$$\mathrm{P}(X_n \le x) \to e^{-e^{-x}}, \qquad \forall x \in \mathbb{R}.$$

This sharpens a single-term formula

$$\xi = \big(1 + o_p(1)\big)\frac{n^{1/2}\log^2 n}{4c},$$

due to Erdös and Turán [9]. Finally (Section 3), we parametrize (continuously) the ordered parts process, by introducing

$$V_n(t) = n^{-1/4}t\left(\lambda_{k(t)} - \frac{\sqrt{n}}{c}\log\frac{1}{t}\right),$$

for $t \in [0,1]$ not too close to either 0 or 1, and setting $V_n(t) \equiv 0$ for the extreme t's. The choice of the parametrization and the centering function is prompted by the Temperley–Vershik equation $(*)$, while the reason for scaling by $n^{1/4}$ is implicit in the proposition. In Lemma 4 we prove, among other properties, that the processes $V_n(t)$ are stochastically equicontinuous, uniformly on $[0,1]$, and then show (Theorem 5) that the finite-dimensional distributions of $V_n(t)$ converge weakly to those of a Gaussian process $V(t)$. Lemma 4 allows us to prove the second part of Theorem 5, namely that for a class of integral functionals F of a form

$$F(x) = \int_0^1 \phi(t, x(t))\, dt,$$

with ϕ behaving not too badly, we have $F(V_n) \to^{\mathscr{D}} F(V)$. Since $V(t)$ is Gaussian, this result is particularly useful for the linear integral functionals. As an illustration, we prove that the eigenvalue distribution for the Diaconis–Shahshahani card-shuffling Markov chain (Diaconis and Shahshahani [4], Diaconis [5]) is asymptotically Gaussian, with zero mean and standard deviation of order $n^{-3/4}$.

1. SMALL PARTS AND LARGE PARTS ...

Every partition of the integer n is uniquely characterized by the sequence of counts $\{X_j\}_{j\ge 1}$, where X_j is the total number of summands (parts) equal to j. So $X_j \ge 0$, $\sum_{j\ge 1} jX_j = n$; in particular, $X_j \equiv 0$ for $j > n$.

Let $Z = \{Z_j\}_{j\ge 1}$ be a sequence of independent random variables such that

$$\mathrm{P}(Z_j = i) = (1 - q^j)(q^j)^i, \qquad i \ge 0.$$

Fristedt [11] proved that, for an arbitrary q, the distribution of $\{X_j\}_{1\le j\le n}$ is the one of $\{Z_j\}_{1\le j\le n}$ conditioned on the event $\{R = n\}$, $R := \sum_{j=1}^{\infty} jZ_j = n$. He further proposed to select q such that $\mathrm{E}R \approx n$, as for this choice $\mathrm{P}(R = n)$ would be close to its maximum value. An almost optimal choice is $q = e^{-c/\sqrt{n}}$, $c = \pi/\sqrt{6}$, in which case

$$\mathrm{E}\, R = n + O(n^{1/2}), \qquad \mathrm{P}(R = n) = (1 + o(1)) \frac{1}{\sqrt[4]{96n^3}}; \qquad (1.1)$$

see Fristedt [11].

Given $j_1, j_2 \in \mathbb{N}$, introduce

$$\mathbf{Z}_1 = \{Z_j\}_{j\le j_1}, \qquad \mathbf{X}_1 = \{X_j\}_{j\le j_1}, \qquad R_1 = \sum_{j\le j_1} jZ_j,$$

$$\mathbf{Z}_2 = \{Z_j\}_{j_2\le j\le n}, \qquad \mathbf{X}_2 = \{X_j\}_{j_2\le j\le n}, \qquad R_2 = \sum_{j\ge j_2} jZ_j.$$

Introduce $d_{\mathrm{TV}}(\mathbf{Z}_i, \mathbf{X}_i)$, the total variation distance between $\mathbf{Z}_i$ and $\mathbf{X}_i$ $(i = 1, 2)$.

THEOREM 1. *Suppose* $\lim j_1 n^{-1/2} = a_1$, $\lim j_2 n^{-1/2} = a_2$ *exist.*

(1) *There exists* $a > 0$ *such that if* $a_1 \in (0, \infty)$, $a_2 \in [a, \infty)$ *then*

$$\lim d_{\mathrm{TV}}(\mathbf{Z}_i, \mathbf{X}_i) = d_{\mathrm{TV}}(\mathcal{N}_i, \mathcal{N}) \qquad (i = 1, 2),$$

where $\mathcal{N}, \mathcal{N}_1, \mathcal{N}_2$ *are three normals, with zero means, and variances equal to* 1, *and* σ_1^2, σ_2^2 *given by*

$$\sigma_i^2 = \lim \frac{\operatorname{Var} R - \operatorname{Var} R_i}{\operatorname{Var} R},$$

or explicitly

$$\sigma_1^2 = \frac{\pi}{\sqrt{24}} \int_{a_1}^{\infty} \frac{y^2 e^{-cy}}{(1 - e^{-cy})^2}\, dy, \qquad \sigma_2^2 = \frac{\pi}{\sqrt{24}} \int_0^{a_2} \frac{y^2 e^{-cy}}{(1 - e^{-cy})^2}\, dy.$$

(2) *If* $j_1 \to \infty$, $a_1 = 0$ *and* $a_2 = \infty$, $j_2 \le \chi n^{1/2} \log n$ $(\chi < (2c)^{-1})$, *then*

$$d_{\mathrm{TV}}(\mathbf{Z}_i, \mathbf{X}_i) \approx \frac{1}{4} \frac{\operatorname{Var} R_i}{\operatorname{Var} R} \cdot \mathrm{E}|1 - \mathcal{N}^2| = o(1) \qquad (i = 1, 2),$$

$$\frac{\operatorname{Var} R_1}{\operatorname{Var} R} \approx \frac{\pi}{8\sqrt{6}} \cdot \frac{j_1}{n^{1/2}},$$

$$\frac{\operatorname{Var} R_2}{\operatorname{Var} R} \approx \frac{1}{2} \left(\frac{j_2}{n^{1/2}} \right)^2 \exp\left(-c \frac{j_2}{n^{1/2}} \right).$$

Notes. (1) Fristedt [11] proved that, for $j_1 n^{-1/4} \to 0$, the *ratio* of the discrete densities of $\mathbf{Z}_1$ and $\mathbf{X}_1$ approaches 1, which certainly implies that $d_{\mathrm{TV}}(\mathbf{Z}_1, \mathbf{X}_1)$ approaches 0. He also obtained an analogous result for the $o(n^{1/4})$ largest parts, thus extending considerably a classical result of Erdös and Lehner [7]; it states that both the number of parts and the size of the largest part are asymptotic in probability to $(2c)^{-1}\sqrt{n} \log n$. (The latter makes the bound for j_2 in Theorem 1 rather natural.)

(2) The theorem confirms a conjecture by Arratia and Tavaré [2] that $d_{\mathrm{TV}}(\mathbf{Z}_1, \mathbf{X}_1) \to 0$ iff $j_1 n^{-1/2} \to 0$, and $d_{\mathrm{TV}}(\mathbf{Z}_2, \mathbf{X}_2) \to 0$ iff $j_2 n^{-1/2} \to \infty$.

Proof. A key element is the Arratia–Tavaré [2] formula

$$\begin{aligned} d_{\mathrm{TV}}(\mathbf{Z}_i, \mathbf{X}_i) &= \frac{1}{2} \mathrm{P}(R_i > n) \\ &\quad + \frac{1}{2} \sum_{r=0}^{n} \mathrm{P}(R_i = r) \left| \frac{\mathrm{P}(R_i^c = n - r)}{\mathrm{P}(R = n)} - 1 \right|, \end{aligned} \tag{1.2}$$

$$R^c := R - R_i.$$

Consider $i = 1$ first.

Step 1. By the definition of Z's and R_1,

$$\mathrm{E}\, R_1 = \sum_{j \le j_1} \frac{jq^j}{1-q^j} = n\left(\int_0^{j_1 n^{-1/2}} \frac{ye^{-cy}}{1-e^{-cy}}\,dy + O\big(j_1 n^{-1}\big)\right),$$

$$\sigma^2(R_1) = \sum_{j \le j_1} \frac{j^2 q^j}{(1-q^j)^2} \tag{1.3}$$

$$= n^{3/2}\left(\int_0^{j_1 n^{-1/2}} \frac{y^2 e^{-cy}}{(1-e^{-cy})^2}\,dy + O\big(j_1 n^{-1}\big)\right).$$

Let us prove that for every *fixed* $x \in \mathbb{C}$

$$\mathrm{E}\exp(xR_1^*) = \exp\left(\frac{1}{2}x^2 + O\big(|x|^3 j_1^{-1/2}\big)\right), \qquad R_1^* := \frac{R_1 - \mathrm{E}\, R_1}{\sigma(R_1)}. \tag{1.4}$$

(This means, in particular, that $R_1^* \Rightarrow \mathscr{N}$, and $\mathrm{E}(R_1^*)^k \to \mathrm{E}\mathscr{N}^k$, $\forall k \ge 1$.)

Denoting $\sigma = \sigma(R_1)$, we begin with

$$\mathrm{E}\exp\big(xR_1\sigma^{-1}\big) = \prod_{j \le j_1} \frac{1-q^j}{1-q^j e^{xj\sigma^{-1}}}.$$

Taking the (main branch) logarithm of the generic factor, and using $\log(1+z) = z - z^2/2 + O(|z|^3)$, we compute

$$-\log\left(1 + \frac{q^j(1-e^{xj\sigma^{-1}})}{1-q^j}\right) = \frac{q^j(e^{xj\sigma^{-1}}-1)}{1-q^j} + \frac{1}{2}\left(\frac{q^j(e^{xj\sigma^{-1}}-1)}{1-q^j}\right)^2$$
$$+ O\left(\frac{q^j|e^{xj\sigma^{-1}}-1|}{1-q^j}\right)^3$$
$$= \frac{x}{\sigma}\frac{jq^j}{1-q^j} + \frac{x^2}{2\sigma^2}\frac{j^2q^j}{(1-q^j)^2}$$
$$+ O\left(\frac{|x|^3}{\sigma^3}\frac{j^3q^j}{(1-q^j)^3}\right).$$

(That the z's we encounter all are $O(1)$ will be seen momentarily.) Adding the logarithms, and using (1.3), we get

$$\mathrm{E}\exp\big(xR_1\sigma^{-1}\big) = \exp\left[\frac{x}{\sigma}\mathrm{E}\, R_1 + \frac{x^2}{2} + O\left(\frac{|x|^3}{\sigma^3}\sum_{j \le j_1}\frac{j^3q^{2j}}{(1-q^j)^3}\right)\right]. \tag{1.5}$$

It remains to notice that σ^3 is of order $n^{3/2}j_1^{3/2}$, while the sum is close to

$$n^2 \int_0^{j_1 n^{-1/2}} \frac{y^3 e^{-cy}}{(1-e^{-cy})^3}\, dy = O\big(n^{3/2} j_1\big).$$

Thus the error term is $O(|x|^3 j_1^{-1/2})$. (This of course means that the error terms for the individual logarithms were small too, so that the usage of the expansion formula for $\log(1+z)$ was indeed legitimate.) Equation (1.4) is proved.

Step 2. Next, let us show that

$$\mathrm{P}(R_1^c = \mathrm{E}\, R_1^c + \Delta) = \frac{1}{\sqrt{2\pi\sigma_c^2}} \exp\left(-\frac{\Delta^2}{2\sigma_c^2}\right) + O(|\Delta| n^{-7/4} + n^{-5/4}) \tag{1.6}$$

($\sigma_c^2 := \mathrm{Var}\, R_1^c$), uniformly over Δ such that $\mathrm{E}\, R_1^c + \Delta \in \mathbb{N}$. (This is a variant of a local limit theorem with a remainder term.)

We begin with

$$f(u) := \mathrm{E}\exp(iuR_1^c) = \prod_{j>j_1} \frac{1-q^j}{1-q^j e^{iuj}}, \qquad u \in \mathbb{R} \tag{1.7}$$

(cf. Step 1). First, let $u = O(n^{-\delta})$, $\delta \in (\frac{2}{3}, \frac{3}{4})$. Analogously to (1.5), via $\log(1+z) = z - z^2/2 + z^3/3 + O(|z|^4)$, we obtain

$$f(u) = \exp\left[iu\,\mathrm{E}\, R_1^c - \frac{1}{2}u^2\sigma_c^2 - iu^3 S_3 + O\big(u^4 S_4\big)\right],$$

$$S_3 = \sum_{j>j_1} j^3 \left[\frac{q^j}{3!(1-q^j)} + \frac{q^{2j}}{2(1-q^j)^2} + \frac{q^{3j}}{3(1-q^j)^3}\right],$$

$$S_4 = \sum_{j>j_1} \frac{q^j j^4}{(1-q^j)^4}.$$

In particular,

$$S_3 = O\left[n^2 \int_0^\infty \frac{y^3 e^{-cy}}{(1-e^{-cy})}\, dy\right] = O(n^2),$$

$$S_4 = O\left[n^{5/2} \int_0^\infty \frac{y^4 e^{-cy}}{(1-e^{-cy})^4}\, dy\right] = O(n^{5/2}),$$

so that

$$|u|^3 S_3 + u^4 S_4 = O(n^{-(3\delta-2)}),$$

for u in question. Expanding $\exp[-iu^3 S_3 + O(u^4 S_4)]$, we have

$$f(u) = \exp\left(iu \,\mathrm{E}\, R_1^c - \frac{1}{2}u^2\sigma_c^{\,2}\right)\left[1 - iu^3 S_3 + O(u^4 n^2 + u^6 n^4)\right]. \quad (1.8)$$

We also need to bound $|f(u)|$ for the larger values of $|u|$. Let us prove that there exist $A, \alpha, \beta > 0$ such that

$$|f(u)| \le e^{-\alpha u^2 n^{3/2}}, \qquad \text{if } |u| \le \frac{1}{A\sqrt{n}}, \quad (1.9)$$

$$|f(u)| \le e^{-\beta\sqrt{n}}, \qquad \text{if } |u| \ge \frac{1}{A\sqrt{n}}. \quad (1.10)$$

By an elementary inequality

$$\left|\frac{1}{1-z}\right| \le \frac{1}{1-|z|}\exp[\mathrm{Re}\, z - |z|] \qquad (z \in \mathbb{C}, |z| < 1), \quad (1.11)$$

and (1.7), we write

$$|f(u)| \le \exp\left[-\sum_{j>j_1} q^j(1 - \cos uj)\right].$$

Now

$$\cos x \le 1 - \frac{x^2}{2!} + \frac{x^4}{4!}, \qquad x \in \mathbb{R},$$

so

$$\sum_{j>j_1} q^j(1 - \cos uj) \ge \frac{u^2}{2!}\sum_{j>j_1} q^j j^2 - \frac{u^4}{4!}\sum_{j>j_1} q^j j^4.$$

The sums on the right are asymptotic to

$$n^{3/2}\int_{j_1 n^{-1/2}}^{\infty} e^{-cy} y^2\,dy, \quad n^{5/2}\int_{j_1 n^{-1/2}}^{\infty} e^{-cy} y^4\,dy,$$

respectively. So, for A large enough there exists $\alpha > 0$ such that, if $|u| \le (A\sqrt{n})^{-1}$ and $n \ge n(A)$, then the sum on the left is at least $\alpha u^2 n^{3/2}$. This proves (1.9).

For $|u| \in [(A\sqrt{n})^{-1}, \pi]$, we have

$$(1-q)^2 = O(n^{-1}) = O(u^2) = O(1-\cos u).$$

Therefore, denoting $D = 1 - 2q\cos u + q^2$ and using $D = O(1-\cos u)$, we obtain

$$\begin{aligned}\sum_{j>j_1} q^j(1-\cos uj) &= \frac{q^{j_1+1}}{1-q} - \mathrm{Re}\,\frac{q^{j_1+1}e^{iuj_1}}{1-qe^{iu}}\\ &\geq q^{j_1+1}\left[\frac{1}{1-q} - \frac{1}{\sqrt{D}}\right]\\ &= \frac{2q^{j_1+2}}{1-q}\cdot\frac{1-\cos u}{\sqrt{D}\cdot\left(\sqrt{D}+1-q\right)}\\ &\geq \beta\sqrt{n},\end{aligned} \tag{1.12}$$

for sufficiently small $\beta > 0$. This proves (1.10).

With the relations (1.8)–(1.10) at hand, we turn directly to proving (1.6). As usual, we begin with the inversion formula

$$\mathrm{P}(R_1^c = m) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-imu} f(u)\,du, \qquad m \in \mathbb{N}.$$

Introduce $u_n = n^{-\delta}$. By (1.9), (1.10),

$$\left|\frac{1}{2\pi}\int_{|u|\geq u_n} e^{-imu} f(u)\,du\right| = O(e^{-\alpha n^{3/2-2\delta}}). \tag{1.13}$$

Consider $|u| \leq u_n$. Using (1.8), we write

$$\frac{1}{2\pi}\int_{|u|\leq u_n} e^{-imu} f(u)\,du = \int_1 + \int_2 + O\left(\int_3 + \int_4\right).$$

Here, setting $m = \mathrm{E}\,R_1^c + \Delta$, we have

$$\begin{aligned}\int_1 &:= \frac{1}{2\pi}\int_{|u|\leq u_n} e^{i\Delta u}\exp\left(-\frac{u^2}{2}\sigma_c^2\right)du\\ &= \frac{1}{\sqrt{2\pi\sigma_c^2}}\left[\exp\left(-\frac{\Delta^2}{2\sigma_c^2}\right) + O(e^{-u_n^2\sigma_c^2/2})\right],\end{aligned} \tag{1.14}$$

so that the remainder term is $O(e^{-\kappa n^{3/2-2\delta}})$, $\kappa > 0$. Further, since u^3 is odd and $|\sin y| \le |y|$,

$$\begin{aligned}\int_2 &:= -iS_3 \int_{|u| \le u_n} e^{i\Delta u} u^3 \exp\left(-\frac{u^2}{2}\sigma_c^2\right) du \\ &= O\left[S_3 |\Delta| \int_{-\infty}^{\infty} u^4 \exp\left(-\frac{\Delta^2}{2\sigma_c^2}\right) du\right] \qquad (1.15) \\ &= O(|\Delta| n^{-7/4}).\end{aligned}$$

Finally,

$$\begin{aligned}\int_3 &:= n^{5/2} \int_{-\infty}^{\infty} u^4 \exp\left(-\frac{\Delta^2}{2\sigma_c^2}\right) du \qquad (1.16) \\ &= O(n^{-5/4}),\end{aligned}$$

and

$$\begin{aligned}\int_4 &:= n^4 \int_{-\infty}^{\infty} u^6 \exp\left(-\frac{\Delta^2}{2\sigma_c^2}\right) du \qquad (1.17) \\ &= O(n^{-5/4}).\end{aligned}$$

Combining (1.14)–(1.17), we arrive at (1.6).

Note. Throughout Step 2, we have never used the condition $j_1 \to \infty$. This means that we can apply (1.6) for $j_1 = 0$, that is, for $R = \sum_{j \ge 1} jZ_j$. Recall that E $R = n + O(n^{1/2})$; also (analogously to (1.3)),

$$\begin{aligned}\sigma^2(R) &= n^{3/2}\left[\int_0^{\infty} \frac{y^2 e^{-cy}}{(1-e^{-cy})^2} dy + O(n^{-1/2})\right] \qquad (1.18) \\ &= \frac{\sqrt{24}}{\pi} n^{3/2} + O(n)\end{aligned}$$

(Fristedt [11]). Therefore, from (1.6) we obtain

$$\mathrm{P}(R = n) = \frac{1}{\sqrt{2\pi\sigma^2(R)}}\left(1 + O(n^{-1/2})\right) = \frac{1}{\sqrt[4]{96n^3}}\left(1 + O(n^{-1/2})\right) \qquad (1.19)$$

(cf. (1.1)).

Step 3. The rest is easy. We evaluate asymptotically the expression for $d_{\mathrm{TV}}(\mathbf{Z}_1, \mathbf{X}_1)$ in (1.1), using the estimates of Steps 1, 2. Begin with the first term, 0.5P $(R_1 > n)$. Since E $R \approx n$ and $j_1 = O(n^{1/2})$, it follows then from (1.2) that $n - \mathrm{E}\, R_1$ is of exact order n. Since $\sigma(R_1) = O(n^{3/4})$, applying the Markov inequality to R_1^* and using (1.4) with $x = \pm 1$, we obtain

$$\mathrm{P}(R_1 > n) = O(e^{-\kappa n^{1/4}}), \qquad \kappa > 0.$$

Turn to the sum in (1.2). For $r \le n$, using (1.6), (1.19),

$$\mathrm{E}\, R_1 + \mathrm{E}\, R_1^c = n + O(n^{1/2}),$$

we write

$$\left| \frac{\mathrm{P}(R_1^c = n - r)}{\mathrm{P}(R = n)} - 1 \right| = \left| \frac{\sigma(R)}{\sigma(R_1^c)} \exp\left(- \frac{(r - \mathrm{E}\, R_1)^2}{2\sigma^2(R_1^c)} \right) - 1 \right| + O\big(n^{-1}|r - \mathrm{E}\, R_1| + n^{-1/2}\big).$$

Therefore, using the notation $R_1^* = (R_1 - \mathrm{E}\, R_1)\sigma^{-1}(R_1)$,

$$d_{\mathrm{TV}}(\mathbf{Z}_1, \mathbf{X}_1) = \frac{1}{2}\mathrm{E}\left| \frac{\sigma(R)}{\sigma(R_1^c)} \exp\left[- \frac{(R_1^*)^2}{2} \big(\sigma^2(R)/\sigma^2(R_1^c) - 1 \big) \right] - 1 \right| + O\big(n^{-1}\sigma(R_1) + n^{-1/2}\big). \tag{1.20}$$

To evaluate the last expectation asymptotically, we use $R_1^* \Rightarrow \mathcal{N}$. Suppose first $\lim j_1 n^{-1/2} = a_1 \in (0, \infty)$. Then

$$\frac{\sigma^2(R_1^c)}{\sigma^2(R)} \to \sigma_1^2 := \frac{\pi}{\sqrt{24}} \int_{a_1}^{\infty} \frac{y^2 e^{-cy}}{(1 - e^{-cy})^2}\, dy.$$

Therefore

$$\begin{aligned} d_{\mathrm{TV}}(\mathbf{Z}_1, \mathbf{X}_1) &\to \frac{1}{2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left| \frac{1}{\sigma_1} \exp\left[- \frac{x^2}{2} (\sigma_1^{-2} - 1) \right] - 1 \right| \exp\left(- \frac{x^2}{2} \right) dx \\ &= \frac{1}{2} \int_{-\infty}^{\infty} \left| \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(- \frac{x^2}{2\sigma_1^2} \right) - \frac{1}{\sqrt{2\pi}} \exp\left(- \frac{x^2}{2} \right) \right| dx \\ &= d_{\mathrm{TV}}(\mathcal{N}_1, \mathcal{N}). \end{aligned} \tag{1.21}$$

Suppose $j_1 n^{-1/2} \to 0$. Then, by (1.3),

$$\sigma^2(R_1) = nj_1\big(1 + O\big(j_1 n^{-1/2}\big)\big) = o\big(\sigma^2(R)\big).$$

Therefore, using

$$\left|e^{-y} - (1-y)\right| \le y^2/2, \qquad \forall y \ge 0,$$

after simple manipulations we rewrite the expected value in (1.20) as

$$\begin{aligned}
&\frac{\sigma^2(R_1)}{4\sigma^2(R)} \cdot \mathrm{E}\left|1 - (R_1^*)^2\right| \\
&\quad + O\left[\frac{\sigma^4(R_1)}{\sigma^4(R)}\big(1 + E(R_1^*)^4\big) + (j_1/n)^{1/2}\right] \\
&\approx \frac{\pi}{8\sqrt{6}} \cdot \frac{j_1}{n^{1/2}} \cdot \mathrm{E}|1 - \mathscr{N}^2|. \qquad (1.22)
\end{aligned}$$

(We have used $\mathrm{E}(R_1^*)^k \to \mathrm{E}(\mathscr{N}^k)$, $\forall k \ge 1$.)

The case $i = 2$ is basically similar. So we will just briefly indicate why we need the conditions $a_2 \ge a$ and

$$j_2 \le \chi\sqrt{n}\log n, \qquad \chi < (2c)^{-1}. \qquad (1.23)$$

First of all, it can be shown that

$$\mathrm{E}\, R_2 = \big(1 + O(n^{-1/2})\big)n\int_{j_2 n^{-1/2}}^{\infty} \frac{ye^{-cy}}{1 - e^{-cy}}\, dy,$$

$$\sigma^2(R_2) = \big(1 + O(n^{-1/2})\big)n^{3/2}\int_{j_2 n^{-1/2}}^{\infty} \frac{y^2 e^{-cy}}{(1 - e^{-cy})^2}\, dy$$

(cf. (1.3)). Thus, for $j_2 n^{-1/2}$ bounded away from 0, $\sigma^2(R_2)$ is of an exact order

$$n^{3/2}\left(\frac{j_2}{\sqrt{n}}\right)^2 \exp\big(-cj_2/\sqrt{n}\big).$$

The analogue of (1.4) is then

$$\mathrm{E}\exp(xR_2^*) = \exp\left[\frac{1}{2}x^2 + O\left(\frac{e^{cj_2\sqrt{n}/2}}{n^{1/4}}|x|^3\right)\right],$$

$$R_2^* := \frac{R_2 - \mathrm{E}\, R_2}{\sigma(R_2)}.$$

So the remainder term is $o(1)$ precisely because the condition (1.23) holds.

Let $f(u)$ stand now for the characteristic function of R_2^c. Just as for R_1^c, we can prove (1.9), assuming that $j_2 \geq n^{1/2}$, say. To get (1.10), we use the bound (1.12) for $j_1 = 0$ to estimate

$$\begin{aligned}\sum_{j<j_2} q^j(1-\cos uj) &\geq \sum_{j=1}^{\infty} q^j(1-\cos uj) - \frac{q^{j_2}}{1-q}\\ &\geq \beta_0\sqrt{n} - (1+o(1))c^{-1}\sqrt{n}\,e^{-cj_2n^{-1/2}}\\ &\geq \frac{1}{2}\beta_0\sqrt{n},\end{aligned}$$

provided that $j_2 \geq a\sqrt{n}$, where $a \geq 1$ is sufficiently large. (Is it possible that (1.10) still holds whenever $j_2 n^{-1/2}$ is simply bounded away from 0?)

Once (1.9), (1.10) are established, the rest of Step 2 goes as before and we get (1.6) with R_2^c in place of R_1^c.

The analogue of (1.20) follows then directly, and so does (1.21) when $a_2 \in [a, \infty)$. Let $j_2 n^{-1/2} \to \infty$, assuming that (1.23) holds. In this case

$$\sigma^2(R_2) \approx c^{-1}n^{3/2}\left(\frac{j_2}{\sqrt{n}}\right)^2 \exp\left(-cj_2n^{-1/2}\right).$$

So, because of (1.23), the remainder term in the R_2-version of (1.20) is of order $n^{-1}\sigma(R_2)$, and we put it instead of the fraction $(j_1/n)^{1/2}$ into the R_2-version of (1.22), in which

$$\frac{\sigma^2(R_2)}{\sigma^2(R)} \gg n^{-1}\sigma(R_2),$$

again according to (1.23). This proves Theorem 1 for $i = 2$ also. ∎

Even though the increasing order of parts is quite natural, our primary goal is to study the likely shape of the Ferrers diagram, for which it is far more convenient to list the parts in decreasing order. So let λ_k denote the kth largest part in the random partition of n. Then we obtain the diagram as a plane array of n dots arranged into bottom aligned columns, so that the height of kth column is λ_k. Reading consecutively the numbers of dots in the rows of the array, beginning from the bottom row, we get the conjugate diagram (partition) $\{\lambda_k^*\}$. Since $\{\lambda_k\}$ is uniformly distributed on the set of all diagrams of size n, then so is $\{\lambda_k^*\}$. Notice that there is a simple connection between λ^* and the sums of X_j, namely

$$\lambda_k^* = X(k) := \sum_{j\geq k} X_j, \qquad k \geq 1. \tag{1.24}$$

Thus

$$\{\lambda_k\} \overset{\mathscr{D}}{\equiv} \{X(k)\}. \tag{1.25}$$

In particular,

$$\lambda_1 \overset{\mathscr{D}}{\equiv} X(1), \qquad \lambda_k - \lambda_{k+1} \overset{\mathscr{D}}{\equiv} X_k$$

(Erdös and Lehner [7], Fristedt [11]). Theorem 1 and (1.25) could be used to get easily a sharp probabilistic estimate of $X(k)$, whence of λ_k, for $k \gg \sqrt{n}$. However, we are interested primarily in the case $k = O(\sqrt{n})$. Fristedt [11] obtained a sharp distributional result for the first $o(n^{1/4})$ largest parts. As a first step toward achieving our goal, we can estimate now only

$$\tilde{X}(k) := \sum_{j=k}^{\underline{k}_n} X_j \qquad \left(\underline{k}_n := \left[n^{1/2}\log^{-1} n\right]\right),$$

for $k \le n^{\delta}$, and any $\delta < \frac{1}{2}$. The tail $X(\underline{k}_n + 1)$ and $X(k)$ $(k \ge n^{\delta})$ will be dealt with in the next section.

In the formulations below and elsewhere, we adopt the following notation. Let a family of random variables V_{nk} $(k \in K_n,\ n \ge 1)$, and a positive function $f(n,k)$ be such that $V_{nk}/f(n,k)$ is bounded in probability as n approaches infinity, *uniformly* for $k \in K_n$. We express this by writing

$$V_{nk} = O_p(f(n,k)), \qquad k \in K_n.$$

LEMMA 1. *If* $\log n \le k \le n^{\delta}$ *then*

$$\tilde{X}(k) = \tilde{E}(k) + O_p\left(\frac{n\log k}{k}\right)^{1/2}; \tag{1.26}$$

if $k \le \log n$ *then*

$$\tilde{X}(k) = \left[1 + O_p(\log^{-1} n)\right]\tilde{E}(k); \tag{1.27}$$

here

$$\tilde{E}(k) = \frac{\sqrt{n}}{c}\log\frac{1 - e^{-c\underline{k}_n n^{-1/2}}}{1 - e^{-ckn^{-1/2}}}.$$

(*So, for all* $k \le \underline{k}_n$, *the distribution of* $\tilde{X}(k)$ *is asymptotically concentrated around the deterministic number* $\tilde{E}(k)$, *but the degree of concentration appears to be higher for the larger values of* k.)

Consider now the large k's. Define

$$k_n = \left[\frac{\sqrt{n}}{2c} (\log n - 2 \log\log n - a_n) \right],$$

where $a_n \to \infty$ however slowly.

LEMMA 2. *If $a_n n^{1/2} \le k \le k_n$, then*

$$X(k) = E(k) + O_p\left(e^{-ckn^{-1/2}} n^{1/2} \log n\right)^{1/2}. \tag{1.28}$$

Proof of Lemma 1. Consider $\log n \le k \le n^{\delta}$. Introduce

$$\tilde{Z}(k) = \sum_{j=k}^{\underline{k}_n} Z_j. \tag{1.29}$$

Then, analogously to (1.3), using $k \ll \underline{k}_n$ we obtain

$$\begin{aligned} \mathrm{E}\,\tilde{Z}(k) &= \sum_{j=k}^{\underline{k}_n} \frac{q^j}{1-q^j} = \sqrt{n} \int_{kn^{-1/2}}^{\underline{k}_n n^{-1/2}} \frac{e^{-cy}}{1-e^{-cy}}\, dy + O(k^{-1} n^{1/2}) \\ &= \tilde{E}(k) + O(k^{-1} n^{1/2}), \end{aligned} \tag{1.30}$$

$$\begin{aligned} \sigma^2\left(\tilde{Z}(k)\right) &= \sum_{j=k}^{\underline{k}_n} \frac{q^j}{\left(1-q^j\right)^2} = \sqrt{n} \int_{kn^{-1/2}}^{\underline{k}_n n^{-1/2}} \frac{e^{-cy}}{\left(1-e^{-cy}\right)^2}\, dy \\ &= (1+o(1)) \frac{n}{ck}. \end{aligned} \tag{1.31}$$

Denoting $\sigma = \sigma(\tilde{Z}(k))$, we obtain (cf. (1.5))

$$\mathrm{E}\exp\left(x\tilde{Z}(k)\sigma^{-1}\right) = \exp\left[\frac{x}{\sigma}\mathrm{E}\,\tilde{Z}(k) + \frac{x^2}{2} + O\left(\sum_{j=k}^{\underline{k}_n} \frac{|x|^3}{\sigma^3} \frac{q^j}{\left(1-q^j\right)^3}\right)\right], \tag{1.32}$$

if x is such that each summand in the remainder is $O(1)$, uniformly for the given range of j. Since σ is of order $(n/k)^{1/2}$ (see (1.31)), this is indeed so if $x = O(k^{1/2})$. Assuming the last condition, and bounding the sum by the corresponding integral, we transform (1.32) for $x = O(k^{1/2})$ into

$$\mathrm{E}\left[\exp\left(x\sigma^{-1}\left(\tilde{Z}(k) - \mathrm{E}\,\tilde{Z}(k)\right)\right)\right] = \exp\left[\frac{x^2}{2} + O\left(|x|^3 k^{-1/2}\right)\right]. \tag{1.33}$$

Setting $x = \sqrt{6 \log k}$, by Markov's inequality we obtain then

$$P\left(\sigma^{-1}\left|\tilde{Z}(k) - \mathrm{E}\,\tilde{Z}(k)\right| \geq \sqrt{6 \log k}\right) \leq \exp\left[-3 \log k + O\left(k^{-1/2} \log k\right)\right]$$
$$\leq \frac{1}{k^2}.$$

Since the remainder in (1.30) is $o(\sigma)$, and $\sigma \approx (n/k)^{1/2}$, the last estimate implies that

$$P\left(\max_{\log n \leq k \leq n^{\delta}} \left|\tilde{Z}(k) - \tilde{E}(k)\right| \geq \left(\frac{7n \log k}{k}\right)^{1/2}\right) \leq \sum_{k \geq \log n} \frac{1}{k^2}$$
$$\leq \log^{-1} n. \quad (1.34)$$

Suppose $k \leq l_n := [\log n + 1]$. Introduce the reverse sequence $\tilde{Z}(l_n), \ldots, \tilde{Z}(1)$. It is easy to see that, given $u \in \mathbb{R}$, the sequence

$$\tilde{Y}(k) := \frac{e^{u\tilde{Z}(k)}}{\mathrm{E}\, e^{u\tilde{Z}(k)}}, \qquad 1 \leq k \leq l_n,$$

is a martingale,

$$\mathrm{E}\left(\tilde{Y}(k)\middle|\tilde{Y}(k+1), \ldots, \tilde{Y}(l_n)\right) = \tilde{Y}(k+1), \qquad 1 \leq k \leq l_n,$$

that is. Then, for any stopping time T (adapted to the reverse sequence), we have

$$\mathrm{E}\,\tilde{Y}(T) = \mathrm{E}\,\tilde{Y}(l_n) = 1 \quad (1.35)$$

(the optional stopping theorem, Durrett [6, Chap. 4]). Now, analogously to (1.32),

$$\mathrm{E}\, e^{u\tilde{Z}(k)} = \exp\left[u\,\mathrm{E}\,\tilde{Z}(k) + O\left(\sum_{j=k}^{\underline{k}_n} u^2 \frac{q^j}{\left(1-q^j\right)^2}\right)\right], \quad (1.36)$$

provided that the jth term in the sum is $O(1)$, uniformly for $k \leq j \leq \underline{k}_n$. Since $\underline{k} = o(n^{1/2})$, the condition is met if $u = O(n^{-1/2})$. For such u, we have

$$\sum_{j=k}^{\underline{k}_n} u^2 \frac{q^j}{\left(1-q^j\right)^2} = O\left(u^2 n \sum_{j=i}^{\infty} j^{-2}\right) = O(1),$$
$$u\left[E\tilde{Z}(k) - \tilde{E}(k)\right] = O(1),$$

whence, if we set $u = \pm n^{-1/2}$ in (1.36),

$$\mathrm{E}\exp\left[n^{-1/2}\left|\tilde{Z}(T) - \tilde{E}(T)\right|\right] = O(1).$$

Let T be the first (counting backward from l_n) moment $k \ge 1$ such that

$$\left|\tilde{Z}(k) - \tilde{E}(k)\right| \ge a_n\sqrt{n};$$

set $T = 1$, say, if no such k exists. Then

$$\begin{aligned}\mathrm{P}\left(\max_{1\le k\le l_n}\left|\tilde{Z}(k) - \tilde{E}(k)\right| \ge a_n\sqrt{n}\right) &\le e^{-a_n}\mathrm{E}\exp\left[n^{-1/2}\left|\tilde{Z}(T) - \tilde{E}(T)\right|\right]\\ &= O(e^{-a_n}).\end{aligned}$$

Since $a_n \to \infty$ however slowly, we see that

$$\begin{aligned}\tilde{Z}(k) - \tilde{E}(k) &= O_p(n^{1/2})\\ &= O_p\left(\frac{\tilde{E}(k)}{\log n}\right), \qquad 1 \le k \le l_n.\end{aligned} \tag{1.37}$$

(We have used

$$\tilde{E}(k) \approx \frac{\sqrt{n}}{c}\log\frac{k_n}{k} \approx \frac{\sqrt{n}}{2c}\log n, \qquad 1 \le k \le l_n.$$

The relations (1.34), (1.37) and Theorem 1(2) imply (1.26), (1.27). ∎

We omit the proof of Lemma 2 since it is very close to the proof above. In the next section we will prove Lemma 3, which—in addition to the range $k = O(\sqrt{n})$—provides the estimates for $X(k)$ with k only twice smaller than k_n.

2. ... AND INTERMEDIATE SIZE PARTS

We begin with an estimate which means that the random variables X_j with moderate j's still possess certain degrees of mutual independence.

Given $1 \le k_1 < k_2 \le n$, let $C = \sum_{j=k_1}^{k_2} X_j$ denote the total number of parts of sizes from the interval $[k_1, k_2]$. (By (1.25), $C \overset{\mathcal{D}}{\equiv} \lambda_{k_1} - \lambda_{k_2+1}$.) We consider the case

$$k_1 \ge n^{\delta}, \qquad \delta \in \left(\tfrac{3}{8}, \tfrac{1}{2}\right]. \tag{2.1}$$

(We use the symbol δ as in Lemma 1, since later Lemma 1 will be dovetailed with the next statement.) Select

$$\delta_1 \in \left(\tfrac{1}{2} - \tfrac{2}{3}\delta, \tfrac{1}{4}\right). \tag{2.2}$$

PROPOSITION. *For $\eta = O(n^{-\delta_1})$, uniformly over $n^{\delta} \le k_1 < k_2 \le \infty$,*

$$\mathrm{E}\left[\exp(\eta(C-m))\right] = O\left[\exp\left(\tfrac{1}{2}\eta^2\sigma_*^2\right)\right],$$

$$m := \sum_{j=k_1}^{k_2} \frac{q^j}{1-q^j}, \qquad \sigma_*^2 := \sum_{j=k_1}^{k_2} \frac{q^j}{(1-q^j)^2}. \tag{2.3}$$

Note. Thus, for $\eta = O(n^{-\delta_1})$, $k_1 \ge n^{\delta}$, the moment generating function (m.g.f.) E $e^{\eta C}$ is essentially bounded by the m.g.f. of the Gaussian variable with mean m and variance σ_*^2 of $\sum_{j=k_1}^{k_2} Z_j$.

Proof of Proposition. Let $\lambda = \{\lambda_1 \ge \lambda_2 \ge \cdots \lambda_m > 0\}$ be a generic finite sequence of positive integer. Introduce

$$C(\lambda) = \left|\{s : \lambda_s \in [k_1, k_2]\}\right|,$$

the number of the parts of λ with sizes from $[k_1, k_2]$. Then, mimicking the usual derivation of Euler's formula, we have:

For $|t| < 1$, $|x| \cdot |t|^j < 1$ ($k \in [k_1, k_2]$),

$$\begin{aligned}\sum_{n=0}^{\infty} t^n \sum_{\lambda : |\lambda| = n} x^{C(\lambda)} &= \prod_{j \notin [k_1,k_2]} (1 + t^j + t^{2j} + \cdots) \\ &\quad \cdot \prod_{j \in [k_1,k_2]} \left(1 + xt^j + (xt^j)^2 + \cdots\right) \\ &= \prod_{j \notin [k_1,k_2]} \frac{1}{1-t^j} \cdot \prod_{j=k_1}^{k_2} \frac{1}{1-xt^j}.\end{aligned}$$

Therefore, denoting the total number of partitions of n by $p(n)$, we write

$$\sum_{n=0}^{\infty} t^n p(n) \mathrm{E}(x^C) = \prod_{j \notin [k_1,k_2]} \frac{1}{1-t^j} \cdot \prod_{j=k_1}^{k_2} \frac{1}{1-xt^j}, \tag{2.4}$$

so

$$p(n)\mathrm{E}(x^C) = [t^n] \prod_{j \notin [k_1,k_2]} \frac{1}{1-t^j} \cdot \prod_{j=k_1}^{k_2} \frac{1}{1-xt^j}. \tag{2.5}$$

Set $x = e^{\eta}$, $\eta = O(n^{-\delta_1})$ (see (2.2) for the range of δ_1). In this case, the function on the right hand side of (2.4), (2.5) is analytic in a disc $\{t\colon |t| \le r = \exp(-\nu n^{-1/2})\}$, $\nu > 0$, since for $j \ge k_1$

$$xr^j \le \exp\left[O(n^{-\delta_1}) - \nu n^{\delta - 1/2}\right] < 1,$$

as $\delta_1 > \frac{1}{2} - \delta$. Using Cauchy's integral formula for the contour $\{t = re^{i\theta}\colon \theta \in (-\pi, \pi]\}$, we write

$$\begin{aligned} p(n)\mathrm{E}(x^C) &= \frac{1}{2\pi r^n}\int_{-\pi}^{\pi} e^{-in\theta}F_n(x, re^{i\theta})\,d\theta, \\ F_n(x,z) &:= \prod_{j=1}^{\infty}\frac{1}{1-z^j}\cdot\prod_{j=k_1}^{k_2}\frac{1-z^j}{1-xz^j}. \end{aligned} \tag{2.6}$$

In particular, the case $x = 1$ corresponds to the integral formula that led Hardy and Ramanujan [17] to their celebrated asymptotic formula for $p(n)$. Its simple corollary is

$$p(n) = \frac{e^{\pi\sqrt{2n/3}}}{4\sqrt{3}\,n}\left(1 + O(n^{-1/2})\right). \tag{2.7}$$

This result can be obtained in a short way via a remarkably simple formula due to Freiman (see Postnikov [24]):

$$\prod_{k=1}^{\infty}\frac{1}{1-e^{-kz}} = \exp\left[\frac{\pi^2}{6z} + \frac{1}{2}\mathrm{Log}\frac{z}{2\pi} + O(|z|)\right], \tag{2.8}$$

uniformly for $z \to 0$ within a corner $\{z\colon \mathrm{Im}\,z \le \epsilon\,\mathrm{Re}\,z,\ \mathrm{Re}\,z > 0\}, \epsilon > 0$ being fixed; Log denotes the main branch of the logarithmic function. (Freiman used (2.8) to obtain a weaker estimate

$$p(n) = \frac{e^{\pi\sqrt{2n/3}}}{4\sqrt{3}\,n}\left(1 + O(n^{-1/4+\epsilon})\right), \qquad \forall \epsilon > 0;$$

see Postnikov [24].) We will utilize the full power of (2.8) later. In the current proof we need only a special case $\mathrm{Im}\,z = 0$, which can be obtained in a direct fashion, via exponentiation of the product in (2.8) and application of the summation formulas. Using the inequality (1.11) and $|xr^j| \le 1$

$(j \geq k_1)$, we obtain

$$\begin{aligned}
\left|F_n(x, re^{i\theta})\right| &\leq F_n(x,r)\exp\left[\sum_{j\geq 1} r^j(\cos j\theta - 1)\right] \\
&\leq F_n(x,r)\exp\left[\operatorname{Re}\frac{1}{1-re^{i\theta}} - \frac{1}{1-r}\right] \\
&= F_n(x,r)\exp\left[\frac{2(r+r^2)(\cos\theta - 1)}{(1-r)(1-2r\cos\theta + r^2)}\right] \\
&\leq F_n(x,r)\exp\left(-\frac{a\theta^2}{n^{-3/2} + n^{-1/2}\theta^2}\right),
\end{aligned} \tag{2.9}$$

for a constant $a > 0$. Now

$$\begin{aligned}
\int_{-\pi}^{\pi} \exp\left(-\frac{a\theta^2}{n^{-3/2}+n^{-1/2}\theta^2}\right) d\theta &\leq \int_{|\theta|\leq n^{-1/2}} e^{-an^{3/2}\theta^2/2}\, d\theta \\
&\quad + \int_{n^{-1/2}\leq|\theta|\leq\pi} e^{-an^{1/2}/2}\, d\theta \\
&= O(n^{-3/4}).
\end{aligned} \tag{2.10}$$

Choosing $r = q (= e^{-cn^{-1/2}})$, with the help of (2.8) we easily obtain

$$q^{-n} \prod_{j=1}^{\infty} \frac{1}{1-q^j} = O\left(\frac{e^{\pi\sqrt{2n/3}}}{n^{1/4}}\right). \tag{2.11}$$

Combining (2.7), (2.9)–(2.11) we get from (2.6):

$$\mathrm{E}(x^C) = O\left(\prod_{j=k_1}^{k_2} \frac{1-q^j}{1-xq^j}\right).$$

Here, taking logarithms and using $x = e^\eta$, $\eta = O(n^{-\delta_1})$, we obtain

$$\begin{aligned}
\sum_{j=k_1}^{k_2} \log\frac{1-q^j}{1-xq^j} &= (x-1)\sum_{j=k_1}^{k_2}\frac{q^j}{1-q^j} + \frac{(x-1)^2}{2}\sum_{j=k-1}^{k_2}\frac{q^{2j}}{(1-q^j)^2} \\
&\quad + O\left(|x-1|^3 \sum_{j=k_1}^{k_2}\frac{q^{3j}}{(1-q^j)^3}\right) \\
&= \eta\sum_{j=k_1}^{k_2}\frac{q^j}{1-q^j} + \frac{1}{2}\eta^2\sum_{j=k_1}^{k_2}\frac{q^j}{(1-q^j)^2} \\
&\quad + O(n^{-(2\delta - 3/4 - 3\delta_1)}).
\end{aligned} \tag{2.12}$$

Indeed

$$|x-1|^3 \sum_{j=k_1}^{k_2} \frac{q^{3j}}{(1-q^j)^3} = O\left(n^{-3\delta_1+1/2} \int_{k_1 n^{-1/2}}^{\infty} \frac{e^{-3cy}}{(1-e^{-cy})^3}\, dy\right)$$
$$= O(n^{-3[\delta_1-(1/2-2\delta/3)]}),$$

and the contributions proportional to η^3 that come from $x-1$ and $(x-1)^3$ are of a lesser order of magnitude. ▮

As the first application, we can now fill the gap left open in Section 1.

LEMMA 3. *If*

$$n^\delta \le k \le \kappa_n, \qquad \kappa_n := \left[\frac{\sqrt{n}}{4c}\log n\right],$$

then

$$X(k) = E(k) + O_p\left(n^{1/2}(e^{ckn^{-1/2}}-1)^{-1}\log n\right)^{1/2}. \tag{2.13}$$

Note. So, as we had promised, the relation (1.28) is indeed a special case of (2.13) for $k \in [a_n n^{1/2}, k_n/2]$.

Proof of Lemma 3. By the proposition, with $k_1 = k$, $k_2 = n$,

$$\mathrm{E}\left[\exp(x(X(k)-m))\sigma_*^{-1}\right] = O\left(\exp(x^2/2)\right)$$
$$m = \sum_{j=k}^{n} \frac{q^j}{1-q^j}, \qquad \sigma_*^2 = \sum_{j=k}^{n} \frac{q^j}{(1-q^j)^2}, \tag{2.14}$$

provided that

$$|x| = O(\sigma_* n^{-\delta_1}). \tag{2.15}$$

The relation (2.14) is analogous to but simpler than (1.32). Further, it can be easily shown that, uniformly for $k \ge n^\delta$,

$$m = E(k) + O\left(\frac{q^k}{1-q^k}\right),$$

$$\sigma_*^2 = \frac{\sqrt{n}}{c}\frac{q^k}{1-q^k} + O\left(\frac{q^k}{(1-q^k)^2}\right)$$
$$= \frac{\sqrt{n}}{c}\frac{q^k}{1-q^k}\left[1+O(n^{-\delta})\right].$$

So, for $k \le \kappa_n$,

$$\sigma_* \ge \gamma_1 n^{1/4}, \qquad \gamma_1 > 0,$$

and (2.15) definitely holds if $|x| = O(n^{1/4-\delta_1})$. We also see that $m - E(k) = o(\sigma_*)$. So we set $x = \sqrt{4 \log n}$ and (just as in the proof of Lemma 1) arrive at (2.13). ∎

As a direct consequence of Lemmas 1–3 and (1.25), we get

THEOREM 2.

$$X(k), \lambda_k = \begin{cases} \left(1 + O_p\left(\log^{-1} n\right)\right) E(k) & (k \le \log n), \\ E(k) + O_p\left(nk^{-1} \log n\right)^{1/2} & (\log n \le k \le n^{1/2}), \\ E(k) + O_p\left(e^{-ckn^{-1/2}} n^{1/2} \log n\right)^{1/2} & (n^{1/2} \le k \le \kappa_n), \\ \left(1 + O_p\left(a_n^{-1}\right)\right) E(k) & (\kappa_n \le k \le k_n), \end{cases} \tag{2.16}$$

where

$$E(k) = \frac{\sqrt{n}}{c} \log \frac{1}{1 - e^{-ckn^{-1/2}}},$$
$$\kappa_n = \left[\frac{\sqrt{n}}{4c} \log n \right], \qquad k_n = \left[\frac{\sqrt{n}}{2c} (\log n - 2 \log \log n - a_n) \right], \tag{2.17}$$

with $a_n \to \infty$ *however slowly.*

The relations (2.16) are essentially analogous to the estimates obtained (among other results) by Szalay and Turán in the remarkable series of three papers [27–29]. The techniques differ substantially, though, ours being more probabilistic and noticeably less analytical, largely due to Fristedt's conditioning device and the Arratia–Tavaré conjecture justified in Theorem 1. Characteristically, Szalay and Turán did not use the connection stated in (1.25). It should be noted also that, neglecting the O_p-terms and extreme values of k, we can rewrite (2.16) loosely as

$$\exp\left(-\frac{c\lambda(k)}{\sqrt{n}}\right) + \exp\left(-\frac{ck}{\sqrt{n}}\right) \approx 1. \tag{2.18}$$

This important observation, based on the Szalay–Turán estimates, was made earlier by Vershik [33]. He also indicated striking parallels and dissimilarities with the asymptotic shape problem for the Plancherel distributed Ferrers diagrams (Logan and Shepp [22], Kerov and Vershik [31, 32]). The Plancherel distribution assigns to a partition λ, ($|\lambda| = n$), the probability measure proportional to the squared degree $d(\lambda)$ of a corresponding irreducible representation of the symmetric group S_n.

Szalay and Turán [29] used their estimates to study the likely magnitude of $d(\lambda)$, under the condition (assumed in the present paper) that λ be uniformly distributed. Using the classic formula of Frobenius [12],

$$d(\lambda) = n! \frac{\prod_{1 \le i \le j \le m}(\lambda_i - \lambda_j + j - i)}{\prod_{i=1}^{m}(\lambda_i + m - i)!},$$
$$\left(\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_m \ge 1, \sum_{i=1}^{m} \lambda_i = n\right), \tag{2.19}$$

they showed that

$$\mathrm{P}\left(\left|\log d(\lambda) - \log \sqrt{n!} + An\right| \le n^{7/8}\log^4 n\right) \ge 1 - n^{-1}, \tag{2.20}$$

with

$$A = -\frac{1}{2} - \log c + c^{-2}\int_0^\infty \frac{y \log y}{e^y - 1}\,dy$$
$$+ c^{-2}\int_0^\infty\int_0^\infty \log \frac{1}{1 - \dfrac{\log(1/(1-e^{-x-y}))}{y + \log(1/(1-e^{-x}))}}\,dx\,dy \tag{2.21}$$
$$\ge 0.02c^{-2} = 0.012\ldots .$$

The authors also made a "very risky" conjecture that there exist constants A_1, A_2 such that

$$\log d(\lambda) = \log\sqrt{n!} - An + A_1\sqrt{n}\log^2 n + A_2\sqrt{n}\log n$$
$$+ O_p\left(\sqrt{n}\log\log n\right), \tag{2.22}$$

stressing that—according to an argument due to Erdös—the error term cannot be improved to $O_p(\sqrt{n}/\log n)$.

Using $\lambda \overset{\mathscr{D}}{\equiv} \lambda^*$ and Theorem 2, we prove

THEOREM 3.

$$\log d(\lambda) = \log\sqrt{n!} - An + O_p\left(n^{3/4}\log^{3/2} n\right), \tag{2.23}$$

where

$$A = \frac{1}{2} + \int_0^\infty \frac{t \log t}{e^{ct} - 1}\,dt = \frac{3}{2} - \gamma + \log\frac{\sqrt{6}}{\pi} - \frac{6}{\pi^2}\sum_{j=1}^\infty \frac{\log j}{j^2} \tag{2.24}$$

and γ *is the Euler constant*; *numerically* $A = 0.1040493\ldots$. *Consequently, the double-integral term in* (2.29) *equals* 1.

Note. Contrary to the conjectured (2.22), we are inclined to believe that $n^{3/4}$ in the remainder term estimate in (2.23) is optimal. We would risk a conjecture that—just like R from Section 1—the random variable $\log[d(\lambda)/\sqrt{n!}\,]$ is asymptotically Gaussian, with mean $-An$, and standard deviation of order $n^{3/4}$.

Proof of Theorem 3. Our argument is based on an alternative expression for $d(\lambda)$, a so-called hook formula discovered by Frame, Robinson, and Thrall [10]:

$$d(\lambda) = \frac{n!}{\prod_{\square \in \lambda} h(\square)}. \tag{2.25}$$

Here the product is over all n unit cells (squares $\square$) in the diagram λ; for an (i, j)-cell (the intersection of ith row and jth column),

$$h(\square) = \lambda_i - j + \lambda_j^* - i + 1,$$

that is, $h(\square)$ is the number of cells in the hook comprised by $\square$ itself, and by the cells in ith row right of $\square$, and in jth column up from $\square$. (The reader is referred to Knuth [19, Chap. 6] for a detailed discussion of the enumerational−algorithmic aspects of the hook formula, and to Greene, Nijenhuis, and Wilf [15, 16] and Pittel [23] for its probabilistic interpretation. We note that the hook formula also was the starting point of analysis in Logan and Shepp [22] and Kerov and Vershik [31, 32].

In view of Stirling's formula for $n!$, it suffices to show that

$$\log \prod_{\square \in \lambda} h(\square) = \frac{1}{2} n \log n + nJ + O_p\left(n^{3/4}\log^{3/2} n\right), \qquad J := \int_0^\infty \frac{t \log t}{e^{ct} - 1}\,dt. \tag{2.26}$$

For the more explicit formula, we observe that

$$J = \frac{d}{dz}\left(\int_0^\infty \frac{t^{z-1}}{e^{ct} - 1}\,dt\right)\bigg|_{z=2},$$

and use the well-known formulas

$$\int_0^\infty \frac{\tau^{z-1}}{e^\tau - 1}\, dt = \Gamma(z)\zeta(z) \qquad (\mathrm{Re} > 1), \tag{2.27}$$

$$\Gamma(z) = \int_0^\infty \tau^{z-1} e^{-\tau}\, d\tau, \qquad \zeta(z) = \sum_{j\ge 1} \frac{1}{j^z},$$

$$\Gamma'(2) = 1 - \gamma, \qquad \zeta(2) = \frac{\pi^2}{6} (= c^2),$$

to obtain

$$J = 1 - \gamma - \log c - c^{-2} \sum_{j\ge 1} \frac{\log j}{j^2}.$$

So, turning to (2.26), we prove first that

$$\sum_{i=1}^{\kappa_n} |\lambda_i - E(i)| = O_p(n^{3/4}\log^{1/2} n),$$

$$\sum_{j=1}^{\kappa_n} |\lambda_j^* - E(j)| = O_p(n^{3/4}\log^{1/2} n). \tag{2.28}$$

Since $\lambda \overset{\mathscr{D}}{\equiv} \lambda^*$, we need to consider only the first sum. Write

$$\sum_{i=1}^{\kappa_n} |\lambda_i - E(i)| = \sum_1 + \sum_2 + \sum_3;$$

here (see Theorem 2)

$$\sum_1 := \sum_{i=1}^{[\log n]} |\lambda_i - E(i)| = O_p\left(\log^{-1} n \sum_{i=1}^{[\log n]} E(i)\right)$$
$$= O_p(n^{1/2}\log n),$$

$$\sum_2 := \sum_{[\log n]+1}^{[n^{1/2}]} |\lambda_i - E(i)| = O_p\left(n^{1/2}\log^{1/2} n \sum_{i=1}^{[n^{1/2}]} i^{-1/2}\right)$$
$$= O_p(n^{3/4}\log^{1/2} n),$$

$$\sum_3 := \sum_{[n^{1/2}]+1}^{\kappa_n} = O_p\left(n^{3/4}\log^{1/2} n \int_1^\infty e^{-cx}\, dx\right)$$
$$= O_p(n^{3/4}\log^{1/2} n).$$

So (2.28) follows. Also, again by Theorem 2,

$$\begin{aligned}\sum_{i\geq\kappa_n}\lambda_i &\leq \sum_{i=\kappa_n}^{k_n}\lambda_i + (\lambda_1^* - k_n)\lambda_{k_n}\\ &= O_p\left(\sum_{i\geq\kappa_n} n^{1/2}e^{-cin^{-1/2}} + n^{3/4}\log\log n\right)\\ &= O_p(n^{3/4}\log\log n),\\ \sum_{j\geq\kappa_n}\lambda_j^* &= O_p(n^{3/4}\log\log n).\end{aligned} \tag{2.29}$$

In the coming estimates we adopt the following notation: given $g\colon \mathbb{N}\to[0,\infty)$, we denote by g the set of cells whose upper right corners have coordinates satisfying $1\leq i\leq g(j)$, $1\leq j\leq g(i)$. Introduce

$$f(k)=\begin{cases}E(k) & (1\leq k\leq\kappa_n),\\ 0 & (k>\kappa_n).\end{cases}$$

By (2.28) and (2.29),

$$\begin{aligned}&\left|\sum_{\square\in\lambda}\log h(\square) - \sum_{\square\in f}\log h(\square)\right|\\ &\quad\leq \log n\left(\sum_{i=1}^{\kappa_n}|\lambda_i - E(i)| + \sum_{i>\kappa_n}\lambda_i\right.\\ &\qquad\qquad\left.+\sum_{j=1}^{\kappa_n}|\lambda_j^* - E(j)| + \sum_{j>\kappa_n}\lambda_j^*\right)\\ &\quad= O_p(n^{3/4}\log^{3/2}n).\end{aligned} \tag{2.30}$$

Further, introducing $h_0(\square)=E(i)-j+E(j)-i+1$ for an (i,j)-cell $\square\in f$, we have

$$\log h(\square) - \log h_0(\square) = O\left(\frac{|\lambda_i - E(i)|}{E(i)-j+\frac{1}{2}} + \frac{|\lambda_j^* - E(j)|}{E(j)-i+\frac{1}{2}}\right),$$

and, for instance, summing over j and then over i, we obtain

$$\sum_{\square\in f}\frac{|\lambda_i - E(i)|}{E(i)-j+\frac12} = O\left(\log E(1)\sum_{i=1}^{\kappa_n}|\lambda_i - E(i)|\right) = O_p(n^{3/4}\log^{3/2} n).$$

Therefore

$$\left|\sum_{\square\in f}\log h(\square) - \sum_{\square\in f}\log h_0(\square)\right| = O_p(n^{3/4}\log^{3/2} n),$$

and this estimate along with (2.30) delivers

$$\left|\sum_{\square\in \lambda}\log h(\square) - \sum_{\square\in f}\log h_0(\square)\right| = O_p(n^{3/4}\log^{3/2} n).$$

So it remains to estimate

$$T_n = \sum_{\substack{1\le i,j\le\kappa_n\\ i\le E(j),\, j\le E(i)}} \log(E(i)-j+E(j)-i+1).$$

We will refer to the summation region as $\mathscr{R}_n$. Notice that, for all (x, y) in the (i, j)-cell,

$$E(x)-y\ge E(i)-j\ge 0,\qquad E(y)-x\ge E(j)-i\ge 0,$$

since $E(\cdot)$ is decreasing. Therefore

$$\log(E(x)-y+E(y)-x+1) = \log h_0(\square) + O\left(\frac{1+E(x)-E(i)}{E(i)-j+\frac12} + \frac{1+E(y)-E(j)}{E(j)-i+\frac12}\right).$$

Using an inequality

$$\frac{1-e^{-u}}{1-e^{-v}} \le \frac{u}{v}\qquad (0<v\le u),$$

we obtain then

$$\iint_{(x,y)\in\square}\log(E(x)-y+E(y)-x+1)\,dx\,dy = \log h_0(\square) + O\left(\frac{1+n^{1/2}i^{-1}}{E(i)-j+\frac12} + \frac{1+n^{1/2}j^{-1}}{E(j)-i+\frac12}\right), \tag{2.31}$$

since, say,

$$\begin{aligned}\int_{i-1}^{i} (E(x) - E(i))\,dx &= \frac{\sqrt{n}}{c}\int_{i-1}^{i} \log\frac{1-e^{-icn^{-1/2}}}{1-e^{-xcn^{-1/2}}}\,dx\\ &\le \frac{\sqrt{n}}{c}\int_{i-1}^{i}\log\frac{i}{x}\,dx\\ &= \frac{\sqrt{n}}{c}\left[1+(i-1)\log(1-i^{-1})\right]\\ &\le \frac{\sqrt{n}}{ci}.\end{aligned}$$

Summing the bounds (2.31) for $(i, j) \in \mathscr{R}_n$, we easily get

$$T_n = \iint_{(x,y)\in R_n} \log(E(x) - y + E(y) - x + 1)\,dx\,dy + O\big(n^{1/2}\log^2 n\big); \tag{2.32}$$

here $(x, y) \in R_n$ iff $(x, y) \in \square$ such that $\square$'s upper right corner (i, j) is in $\mathscr{R}_n$. Introduce

$$H_n = \{0 < x, y \le \kappa_n : x \le E(y),\, y \le E(x)\};$$

the set $H_n \setminus R_n$ is covered by $O(n^{1/2}\log n)$ boundary cells. For every such (i, j)-cell, by (2.31),

$$\begin{aligned}0 &\le \iint_{(x,y)\in \square\cap R_n} \log(E(x) - y + E(y) - x + 1)\,dx\,dy\\ &= O\left(\log n + \frac{n^{1/2}}{i} + \frac{n^{1/2}}{j}\right),\end{aligned}$$

and—since each row and each column contains at most one boundary cell—the integral over $H_n \setminus R_n$ is $O(n^{1/2}\log^2 n)$. So (2.32) becomes

$$T_n = \iint_{(x,y)\in H_n} \log(E(x) - y + E(y) - x + 1)\,dx\,dy + O\big(n^{1/2}\log^2 n\big). \tag{2.33}$$

Next we want to extend integration to the whole

$$H := \{x, y > 0 : x \le E(y),\, y \le E(x)\}.$$

To this end, let us show that

$$\iint_{H\setminus H_n} \log(E(x) - y + E(y) - x + 1)\,dx\,dy = O(n^{3/4}\log n).$$

Using convexity of logarithm, the inequality

$$E(y) \le \frac{\sqrt{n}}{c}\log\frac{\sqrt{n}}{cy},$$

and denoting

$$p(x) = E(x) - x - \frac{\sqrt{n}}{c}\log\frac{cE(x)}{e\sqrt{n}} + 1,$$

we estimate half of the above integral:

$$\begin{aligned}
&\int_{\kappa_n}^{\infty}\left[\int_0^{E(x)} \log(E(x) - y + E(y) - x + 1)\,dy\right]dx\\
&\quad\le \int_{\kappa_n}^{\infty} E(x)\log\left[\frac{1}{E(x)}\int_0^{E(x)} E((x) - x + 1 + E(y))\,dy\right]dx\\
&\quad\le \int_{\kappa_n}^{\infty} E(x)\log p(x)\,dx\\
&\quad= O(n^{3/4}\log n).
\end{aligned}$$

The last equality holds since

$$p'(x) < 0, \qquad p(\kappa_n) \approx \frac{\sqrt{n}}{c},$$

$$E(x) \approx \frac{\sqrt{n}}{c}e^{-xcn^{-1/2}}, \qquad \text{uniformly for } x \ge \kappa_n.$$

(By working out a lower bound, we can show that $n^{3/4}\log n$ is the exact order of the integral.) Therefore (see (2.33)),

$$T_n = \iint_{(x,y)\in H} \log(E(x) - y + E(y) - x + 1)\,dx\,dy + O(n^{3/4}\log n). \tag{2.34}$$

Now for the fun! Introduce new variables

$$u = \frac{E(x) - y}{\sqrt{n}} = \frac{1}{c}\log\frac{1}{1 - e^{-xcn^{-1/2}}} - \frac{y}{\sqrt{n}} = \frac{1}{c}\log\frac{e^{-ycn^{-1/2}}}{1 - e^{-xcn^{-1/2}}},$$

$$v = \frac{E(y) - x}{\sqrt{n}} = \frac{1}{c}\log\frac{1}{1 - e^{-ycn^{-1/2}}} - \frac{x}{\sqrt{n}} = \frac{1}{c}\log\frac{e^{-xcn^{-1/2}}}{1 - e^{-ycn^{-1/2}}}.$$

Then the bounds for H become $u > 0$, $v > 0$. The Jacobian is given by

$$\frac{\partial(u,v)}{\partial(x,y)} = \frac{1}{n}\left(\frac{e^{-ycn^{-1/2}}}{1 - e^{-xcn^{-1/2}}} \cdot \frac{e^{-xcn^{-1/2}}}{1 - e^{-ycn^{-1/2}}} - 1\right) = \frac{1}{n}\left(e^{c(u+v)} - 1\right).$$

So the integral in (2.34) reduces to

$$\frac{1}{2}n\log n \iint_{u>0,\, v>0} \frac{du\,dv}{e^{c(u+v)} - 1} + n\iint_{u>0,\, v>0} \frac{\log(u + v + n^{-1/2})}{e^{c(u+v)} - 1}\,du\,dv$$
$$= \frac{I}{2}n\log n + nI_n.$$

Here, by (2.27),

$$I = \int_0^\infty \frac{t}{e^{ct} - 1}\,dt = \frac{1}{c^2}\Gamma(2)\zeta(2) = 1,$$

$$I_n = \int_0^\infty \frac{t\log(t + n^{-1/2})}{e^{ct} - 1}\,dt.$$

To estimate I_n, we approximate $\log(t + n^{-1/2})$ by $\log t$ and, introducing

$$G(t) = \int_0^t \frac{\tau}{e^{c\tau} - 1}\,d\tau,$$

integrate the remainder by parts:

$$I_n = \int_0^\infty \frac{t\log t}{e^{ct} - 1}dt + n^{-1/2}\int_0^\infty \frac{G(t)}{t(t + n^{-1/2})}dt.$$

The second integral equals

$$O\left(\int_0^1 \frac{dt}{t + n^{-1/2}} + \int_1^\infty t^{-2}\,dt\right).$$
$$= O(\log n).$$

Thus

$$I_n = J + O\big(n^{-1/2}\log n\big).$$

and (2.34) simplifies to

$$T_n = \frac{1}{2}n\log n + nJ + O\big(n^{3/4}\log n\big). \quad \blacksquare$$

Note. Clearly, from (2.23) and the proof we get

$$\frac{1}{n}\log\frac{d(\lambda)}{\sqrt{n!}} \xrightarrow{\text{P}} -\frac{1}{2} - J, \tag{2.35}$$

where

$$J = \iint_{\substack{0<x<f(y),\\ 0<y<f(x)}} \log(f(x) - y + f(y) - x)\,dxdy,$$

where

$$f(x) = c^{-1}\log\frac{1}{1-e^{-cx}}.$$

In essence, the relation (2.35) was claimed, without a proof, by Vershik [33].

A much simpler connection between partitions of n and the symmetric group S_n is based on the observation that with every partition $\lambda = \{X_j\}_{j\geq 1}$ one can associate a conjugacy class $C(\lambda)$ of S_n: it consists of all permutations with X_j cycles of length j ($j \geq 1$). This association establishes bijection between the partitions and the conjugacy classes. By Cauchy's formula,

$$\begin{aligned} c(\lambda) &:= |C(\lambda)| = \frac{n!}{\xi(\lambda)}, \\ \xi(\lambda) &:= \prod_{j\geq 1} j^{X_j} X_j!. \end{aligned} \tag{2.36}$$

The class function $\xi(\lambda)$ is of importance in its own right; it is the number of cosets of $C(\lambda)$ in S_n. It also equals the number of permutations that commute with a fixed permutation from $C(\lambda)$.

Under the uniform distribution on the set of partitions λ, $c(\lambda)$, $\xi(\lambda)$ are random variables. Erdös and Turán [9] proved that

$$\log \xi(\lambda) = \left(1 + o_p(1)\right)\frac{n^{1/2}}{4c}\log^2 n. \tag{2.37}$$

(By $o_p(1)$ we denote a random variable which converges to 0 in probability.) We strengthen (2.37), proving

THEOREM 4.

$$\log \xi(\lambda) = \frac{n^{1/2}\log^2 n}{4c} + \frac{n^{1/2}\log n}{2c}(X_n - \gamma - 2\log c); \tag{2.38}$$

here, for every $x \in \mathbb{R}$,

$$\mathrm{P}(X_n \le x) \to e^{-e^{-x}}.$$

Note. The appearance of the doubly exponential distribution came as a surprise, since we had somehow expected $\log \xi(\lambda)$ to be asymptotically Gaussian. (Notice that the limiting distribution is infinitely divisible nonetheless.) A partial reason for this "abnormal" phenomenon is that the smaller the parts are, the larger the role they play in shaping a likely behavior of $\xi(\lambda)$. In contrast, the dominant contribution to a typical value of $d(\lambda)$ appears to come from the moderate-sized parts.

Proof of Theorem 4. Here is our plan. First, we use Theorem 1(1) to show that the contribution to $\log \xi(\lambda)$ made by the variables X_j with $j = o(n^{1/2})$ is doubly exponential, with mean and variance respectively of orders $n^{1/2}\log^2 n$ and $n\log^2 n$. Second, we use Theorem 2 to show that the overall contribution by the remaining X_j is likely to be within the $o(n^{1/2}\log n)$-neighborhood of its mean. Combining the two results, we get the statement.

Let $a_n \to \infty$ however slowly, and $j_n = [n^{1/2}/a_n]$. Write

$$\begin{aligned}\log \xi(\lambda) &= \sum_{j\le j_n}\log\left(j^{X_j}X_j!\right) + \sum_{j>j_n}\log\left(j^{X_j}X_j!\right)\\ &= L_1(\lambda) + L_2(\lambda).\end{aligned}$$

1. By Stirling's formula,

$$\begin{aligned}L_1(\lambda) &= \sum_{j\le j_n} X_j\log\left(\frac{jX_j}{e}\right) + O\left(\sum_{j\le j_n}\log(X_j+1)\right)\\ &= \sum_{j\le j_n} X_j\log\left(\frac{jX_j}{e}\right) + O\left(a_n^{-1}n^{1/2}\log n\right),\end{aligned} \tag{2.39}$$

simply because $X_j \le n$. To get the limiting distribution of the sum in (2.39), we replace X's by Z's from Section 1 and consider

$$\mathscr{L}_1 = \sum_{j \le j_n} Z_j \log\left(\frac{jZ_j}{e}\right).$$

(Since $j_n n^{-1/2} \to 0$, by Theorem 1 $d_{\mathrm{TV}}(L_1, \mathscr{L}_1) \to 0$.) Let us break up $\mathscr{L}_1$ as follows:

$$\mathscr{L}_1 = \mathscr{L}_1^{(1)} + \mathscr{L}_1^{(2)},$$

$$\mathscr{L}_1^{(1)} = \frac{1}{2}\log n \sum_{j \le j_n} Z_j, \qquad \mathscr{L}_1^{(2)} = \sum_{j \le j_n} Z_j \log \frac{jZ_j}{en^{1/2}}.$$

Let us quickly dispense with $\mathscr{L}_1^{(2)}$. Given $j \le j_n$, denote $x_r = rj/\sqrt{n}$, $\Delta x_r = r/\sqrt{n}$, and—with the help of (2.27)—compute

$$\begin{aligned} \mathrm{E}\left(Z_j \log \frac{jZ_j}{en^{1/2}}\right) &= (1 - q^j) \sum_{r \ge 0} q^{jr} r \log \frac{rj}{en^{1/2}} \\ &= \frac{n}{j^2}(1 - q^j) \sum_{r \ge 0} e^{-cx_r} x_r \log \frac{x_r}{e}\, \Delta x_r \\ &= \frac{n}{j^2}(1 - q^j)\left(\int_0^\infty e^{-cx} x \log \frac{x}{e}\, dx + O(j/\sqrt{n})\right) \\ &= -\frac{\sqrt{n}}{cj}(\gamma + \log c) + O(1). \end{aligned} \tag{2.40}$$

Likewise, but more crudely,

$$\begin{aligned} \mathrm{E}\left(Z_j^2 \log^2 \frac{jZ_j}{en^{1/2}}\right) &= (1 - q^j) \sum_{r \ge 0} q^{rj} r^2 \log^2 \frac{rj}{en^{1/2}} \\ &= O\left((1 - q^j)\frac{n^{3/2}}{j^3}\right) = O\left(\frac{n}{j^2}\right). \end{aligned} \tag{2.41}$$

Using (2.40), (2.41), and recalling that $j_n = \sqrt{n}/a_n$, we see that

$$\mathrm{E}\,\mathscr{L}_1^{(2)} = -\frac{\gamma + \log c}{2c}\sqrt{n}\log n + O(\sqrt{n}\log a_n),$$

$$\operatorname{Var} \mathscr{L}_1^{(2)} \le \sum_{j \le j_n} \mathrm{E}\left(Z_j^2 \log^2 \frac{jZ_j}{en^{1/2}}\right) = O(n).$$

So the variance is negligible compared to the squared mean. Applying Chebyshev's inequality we obtain

$$\mathscr{L}_1^{(2)} = -\frac{\gamma + \log c}{2c}\sqrt{n}\log n + O_p\left(\sqrt{n}\log^{1/2} n\right), \tag{2.42}$$

say.

To get the limiting distribution of $\mathscr{L}_1^{(1)}$, we need to study the characteristic function

$$g_n(u) = \mathrm{E}\left(e^{it\mathscr{L}_1^{(1)}}\right), \qquad t = \frac{2c}{\sqrt{n}\log n}u.$$

By definition of $\mathscr{L}_1^{(1)}$ and independence of Z's,

$$g_n(u) = \prod_{j \le i_n} g_{nj}(u), \qquad g_{nj}(u) = \mathrm{E}\exp\left(\frac{icu}{\sqrt{n}}Z_j\right).$$

We compute

$$\begin{aligned} g_{nj}(u) &= \left(1 - q^j\right)\sum_{r \ge 0} q^{jr}\exp\left(\frac{icu}{\sqrt{n}}r\right) \\ &= \frac{1 - e^{crn^{-1/2}}}{1 - e^{c(iu-j)n^{-1/2}}} = \left(1 - \frac{e^{iucn^{-1/2}} - 1}{e^{cjn^{-1/2}} - 1}\right)^{-1} \\ &= \left(1 - \frac{iu}{j}\left(1 + O\left(n^{-1/2}\right)\right)\right)^{-1} \\ &= \left(1 - \frac{iu}{j}\right)^{-1}\left(1 + O\left(n^{-1/2}\right)\right). \end{aligned}$$

Then we can write

$$\begin{aligned} g_n(u) &= \left(1 + O\left(j_n n^{-1/2}\right)\right)\frac{\prod_{j=1}^{j_n} e^{iu/j}}{\prod_{j=1}^{j_n} e^{iu/j}(1 - iu/j)} \\ &= \left(1 + O\left(a_n^{-1}\right)\right)\exp\left[iu(\log j_n + \gamma)\right] \cdot e^{-\gamma iu}\Gamma(1 - iu) \\ &= \left(1 + O\left(a_n^{-1}\right)\right) \cdot e^{iu\log j_n}\Gamma(1 - iu). \end{aligned} \tag{2.43}$$

We have used here the relation

$$\prod_{j=1}^{\infty} e^{-z/j}(1 + z/j) = \left(e^{\gamma z}\Gamma(1+z)\right)^{-1}, \tag{2.44}$$

true for all $z \neq -1, -2, \dots,$ and an obvious estimate

$$\prod_{j \geq k} e^{-z/j}(1 + z/j) = 1 + O\left(z^2/k\right), \qquad i \to \infty.$$

Now it is also known, and easy to check, that

$$\Gamma(1 - iu) = \mathrm{E}\left(e^{iuX}\right), \qquad \mathrm{P}(X \leq x) = e^{-e^{-x}}.$$

(Probabilistically, the identity (2.44) means that

$$X \stackrel{\mathscr{D}}{\equiv} \gamma + \sum_{j \geq 1} \frac{Y_j - 1}{j},$$

where Y_j are i.i.d. exponentially distributed with parameter 1, hence X is infinitely divisible.) It follows then from (2.43) that

$$X_n := \frac{2c\mathscr{L}_1^{(1)}}{n^{1/2}\log n} - \log j_n \Rightarrow X,$$

and, in combination with (2.42) we have

$$\begin{aligned} \mathscr{L}_1 = \frac{\sqrt{n}}{2c}\log n \log j_n + \frac{\sqrt{n}}{2c}\log n\left(X_n - \gamma - \log c\right) \\ + O_p\left(n^{1/2}\log^{1/2} n\right). \end{aligned} \tag{2.45}$$

The same formula holds then for $L_1(\lambda)$.

2. Now consider $L_2(\lambda)$. We know that, with high probability (whp) as $n \to \infty$, $X_j = 0$ for $j \geq i_n = [(\sqrt{n}/2c)(\log n + a_n)]$. So whp

$$L_2(\lambda) = L_2^{(1)} + L_2^{(2)},$$

$$L_2^{(1)} := \sum_{j=j_n+1}^{i_n} X_j \log j, \qquad L_2^{(2)} := \sum_{j=j_n+1}^{i_n} \log\left(X_j!\right).$$

First, we have

$$L_2^{(1)} = \sum_{j=j_n+1}^{i_n} (X(j) - X(j+1))\log j$$
$$= X(j_n+1)\log(j_n+1) + \sum_{j=j_n+2}^{i_n} X(j)\log\left(\frac{j}{j-1}\right),$$

as $X(i_n+1) = 0$ whp. By Theorem 2,

$$X(j_n+1)\log(j_n+1) = c^{-1}\sqrt{n}\log\left(1 - e^{-c(j_n+1)n^{-1/2}}\right)^{-1}\log(j_n+1)$$
$$+ O_p\left(nj_n^{-1}\log n\right)^{1/2} \tag{2.46}$$
$$= \sqrt{n}\,c^{-1}\log\left(n^{1/2}c^{-1}j_n^{-1}\right)\log j_n + O_p\left(a_n^{-1}\sqrt{n}\log n\right).$$

The sum is bounded by $\Sigma_1 + \Sigma_2$; here, by Theorem 2,

$$\sum_1 := \sum_{j=j_n+2}^{k_n} X(j)\log\frac{j}{j-1} = O_p\left(\sum_{j\ge j_n+1} j^{-1}E(j)\right)$$
$$= O_p\left(n^{1/2}\int_{a_n^{-1}}^{\infty} x^{-1}\log(1-e^{-cx})^{-1}\,dx\right) = O_p\left(\sqrt{n}\int_{a_n^{-1}}^{1} x^{-1}\log(1/x)\,dx\right)$$
$$= O_p\left(\sqrt{n}\log^2 a_n\right);$$
$$\sum_2 := \sum_{j=k_n+1}^{i_n} X(j)\log\frac{j}{j-1} = O\left(k_n^{-1}X(k_n)(i_n-k_n)\right)$$
$$= O_p\left(k_n^{-1}E(k_n)(i_n-k_n)\right) = O_p\left(\sqrt{n}\log\log n\right)a_n\log a_n).$$

Therefore

$$L_2^{(1)} = \sqrt{n}\,c^{-1}\log\left(n^{1/2}c^{-1}j_n^{-1}\right)\log j_n + O_p\left(a_n^{-1}\sqrt{n}\log n\right). \tag{2.47}$$

Finally, consider $L_2^{(2)}$. We begin with

$$L_2^{(2)} = \sum_{\{j_n+1\le j\le i_n:\,X_j\ge 1\}}\ \sum_{k=1}^{X_j}\log k$$
$$= \sum_{k\ge 1}\log k\sum_{j=j_n+1}^{i_n} 1_{\{X_j\ge k\}} \tag{2.48}$$
$$= \sum_{k\ge 1}\log k\sum_{j=j_n+1}^{i_n} 1_{\{X_j\ge k,\ jk\le\sqrt{n}\,a_n\}}.$$

The last equality holds whp since

$$X_j \geq k \Rightarrow jk \leq X(j) \leq X(j_n),$$

and

$$X(j_n) = O_p(E(j_n)) = O_p\left(\sqrt{n}\log a_n\right).$$

Furthermore, notice that, for $jk < n$,

$$\mathrm{P}\left(X_j \geq k\right) = \frac{p(n-jk)}{p(n)}, \tag{2.49}$$

since there is an obvious bijection between partitions λ of n with $X_j(\lambda) \geq k$, and partitions λ of $n - jk$. For $jk \leq \sqrt{n}\,a_n$, the Hardy–Ramanujan formula (2.7) for $p(\cdot)$ transforms (2.49) into

$$\mathrm{P}\left(X_j \geq k\right) = O\left(e^{-bjkn^{-1/2}}\right), \qquad b = c/2.$$

Therefore the expected value of the sum in (2.48) is

$$\begin{aligned} O\left(\sum_{k\geq 1}\log k \sum_{j=j_n+1}^{i_n} e^{-bjkn^{-1/2}}\right) &= O\left(\sqrt{n}\sum_{k\geq 1} e^{-bj_nkn^{-1/2}}\log k\right) \\ &= O\left(\sqrt{n}\int_1^\infty e^{-bxa_n^{-1}}\log x\,dx\right) \\ &= O\left(\sqrt{n}\,a_n\log a_n\right). \end{aligned}$$

Hence

$$L_2^{(1)} = O_p\left(\sqrt{n}\,a_n\log a_n\right),$$

and adding the estimate (2.47) we get

$$L_2(\lambda) = \sqrt{n}\,c^{-1}\log\left(\sqrt{n}\,c^{-1}j_n^{-1}\right)\log j_n + O_p\left(a_n^{-1}\sqrt{n}\log n\right). \tag{2.50}$$

(Thus the leading term and the error estimate (2.46) determine their counterparts in (2.50).) Now the sum of the leading terms in (2.45) and (2.50) is

$$\begin{aligned} \sqrt{n}\,c^{-1}\log\left(nc^{-1}j_n^{-1}\right)\log j_n &= \sqrt{n}\,c^{-1}\left(\tfrac{1}{4}\log^2 n - \tfrac{1}{2}\log c\log n\right) \\ &\quad + O\left(n^{1/2}\log^2 a_n\right). \end{aligned}$$

So

$$\begin{aligned} L_1(\lambda) + L_2(\lambda) &= \frac{n^{1/2}\log^2 n}{4c} + \frac{\sqrt{n}\log n}{2c}\left(X_n - \gamma - 2\log c\right) \\ &\quad + O_p\left(a_n^{-1}\sqrt{n}\log n\right). \quad \blacksquare \end{aligned}$$

3. A FUNCTIONAL LIMIT THEOREM

Here we prove a theorem that may allow us to get the distributional results in the case when the functional $F(\lambda)$ depends primarily on the moderate-sized parts.

Introduce an integer-valued function

$$k(t) = \left[\frac{\sqrt{n}}{c}\log\frac{1}{1-t}\right] + 1, \qquad t \in (0,1);$$

so $k(t_1(n)) = 1$, where $t_1(n) := (c/2)n^{-1/2}$. For $t \in [t_1(n), 1)$, define

$$V_n(t) = n^{-1/4}t\left(\lambda_{k(t)} - \frac{\sqrt{n}}{c}\log\frac{1}{t}\right).$$

Also set $V_n(T) \equiv 0$, if $t < t_1(n)$, and $V_n(1) = 0$. (The choice of $k(t)$ and the centering function is motivated by the Temperley–Vershik's equation (2.18).) Clearly, since $\lambda_1 = O_p(n^{1/2}\log n)$,

$$V_n(t) = O_p\big(tn^{1/4}\log n\big). \tag{3.1}$$

Further, since $\lambda_k \equiv 0$ $(k > n)$, for $t \geq t_2(n) := 1 - e^{-cn^{1/2}}$ we have

$$V_n(t) = O(n^{1/4}e^{-c\sqrt{n}}). \tag{3.2}$$

For every n, the random function $V_n(t)$ is right-continuous on $[0,1)$ and the left-side limits $V_n(t-)$ exist for all $t \in (0,1]$.

LEMMA 4. *Introduce* $t^*(n) = n^{-\delta_0}$, *where* $\delta_0 \in (0, \frac{1}{8})$. *Then, given* $\varepsilon > 0$, $\mu > 0$, *and* $\nu \in (\frac{1}{4}, \frac{1}{2})$, $\rho \in (0, \frac{1}{12} - 2\delta_0/3)$,

$$V_n(t) = \begin{cases} O_p\big(tn^{1/4}\log n\big), & \text{if } t \leq n^{-\nu}, \\ O_p\big(t^{1/2}\log n\big), & \text{if } t \in \big[n^{-\nu}, t^*(n)\big], \end{cases}$$

$$\lim_{h\to 0}\limsup_{n\to\infty}\sup_{\{t,s:|t-s|\leq h\}} \mathrm{P}\big\{\big|V_n(t) - V_n(s)\big| \geq \varepsilon\big\} = 0, \tag{3.3}$$

$$\mathrm{E}\big|V_n(t)\big|^{\mu} = O\big(t^{\mu/2}(1-t)^{\mu/2} + n^{-\mu\rho}\big), \qquad t \geq t^*(n).$$

The second equation means that the random functions $V_n(\cdot)$ are stochastically equicontinuous, uniformly for $t \in [0,1]$.

Note. It is probably true that an even stronger property holds, namely

$$\lim_{h\to 0}\limsup_{n\to\infty}\mathrm{P}\left\{\sup_{|t-s|\leq h}\big|V_n(t) - V_n(s)\big| \geq \varepsilon\right\} = 0. \tag{3.4}$$

Proof of Lemma 4. (a) By (3.1),

$$V_n(t) = O_p\big(tn^{-1/4}\log n\big) = O_p\big(n^{-(\nu-1/4)}\log n\big), \qquad t \le n^{-\nu}. \quad (3.5)$$

Also, by Theorem 2 (the second equation in (2.16)), for $t \in [n^{-\nu}, t^*(n)]$,

$$\begin{aligned}\lambda_{k(t)} - \frac{\sqrt{n}}{c}\log\frac{1}{t} &= O_p\big(t^{-1} + n^{1/2}k(t)^{-1/2}\log^{1/2} n\big)\\ &= O_p\big(t^{-1} + t^{-1/2}n^{1/4}\log^{1/2} n\big).\end{aligned}$$

So

$$\begin{aligned}|V_n(t)| &= O_p\big(n^{-1/4} + t^{1/2}\log n\big)\\ &= O_p\big(t^{1/2}\log n\big) = O_p\big(n^{-\delta_0/2}\log n\big).\end{aligned}$$

So the first line in (3.3) is proven. Note that, for $t \le t^*(n)$,

$$V_n(t) = O_p(n^{-s}\log n) \qquad \big(s = \min\{\nu - \tfrac{1}{4}, \delta_0/2,\}\big).$$

(b) By (3.2) and the last comment, it suffices to prove a weaker version of the second part in (3.3), with $t, s \in [t^*(n), t_2(n)]$. Let us show that, uniformly for $t^*(n) \le t_1 \le t_2 \le t_2(n)$,

$$\mathrm{P}\{|V_n(t_2) - V_n(t_1)| \ge x\} \le \begin{cases} ae^{-cx^2/(8(t_2-t_1))}, & \text{if } x \le n^{\rho}(t_2 - t_1),\\ ae^{-cn^{\rho}x/8}, & \text{if } x \ge n^{\rho}(t_2 - t_1),\end{cases} \quad (3.6)$$

for some absolute constant $a > 0$. For $t \ge t^*(n)$, $k(t)$ is at least of order n^{δ}, where $\delta := \frac{1}{2} - \delta_0 > \frac{3}{8}$. Setting $k_i = k(t_i)$ $(i = 1, 2)$ and using the definition of $V_n(t)$, we write

$$\begin{aligned}V_n(t_2) - V_n(t_1) &= n^{-1/4}t_1\big((\lambda_{k_1} - \lambda_{k_2}) - (m_1 - m_2)\big)\\ &\quad - n^{-1/4}(t_2 - t_1)(\lambda_{k_2} - m_2) + O(n^{-1/4});\end{aligned} \quad (3.7)$$

here

$$\begin{aligned}m_i &:= \sum_{j=k_i}^{\infty}\frac{q^j}{1-q^j}\\ &= \frac{\sqrt{n}}{c}\log\frac{1}{t_i} + O\big(t_i^{-1}\big).\end{aligned} \quad (3.8)$$

By Cauchy's inequality, we obtain from (3.7): for $0 < v = o(n^{1/4})$,

$$\begin{aligned} &\mathrm{E}\exp\big(v\big|V_n(t_2) - V_n(t_1)\big|\big) \\ &\quad \le 2\,\mathrm{E}^{1/2}\Big(\exp\Big(2vt_1 n^{-1/4}\big|(\lambda_{k_1} - \lambda_{k_2}) - (m_1 - m_2)\big|\Big)\Big) \\ &\qquad \times \mathrm{E}^{1/2}\Big(\exp\Big(2v(t_2 - t_1)n^{-1/4}|\lambda_{k_2} - m_2|\Big)\Big). \end{aligned}$$

Notice that

$$\tfrac{1}{4} - \rho \in \big(\tfrac{1}{2} - \tfrac{2}{3}\delta, \tfrac{1}{4}\big).$$

So, for $v = O(n^{\rho})$, by Proposition (Section 2) (with $C = \sum_{j=k_1}^{k_2-1} X_j$ replaced by the equidistributed $\lambda_{k_1} - \lambda_{k_2}$), the first expectation is bounded, within a factor $O(1)$, by

$$\exp\left(\frac{4v^2t_1^2}{2n^{1/2}}\sigma_*^2\right) \le 2\exp\left(\frac{2}{c}v^2(t_2 - t_1)\right).$$

(Here we have used an easy estimate

$$\sigma_*^2 = \sum_{j=k_i}^{\infty} \frac{q^j}{(1 - q^j)^2} = \frac{\sqrt{n}}{c}\,\frac{1 - t_i}{t_i} + O\big(t_i^{-2}\big).) \tag{3.9}$$

Likewise, the same bound obtains for the second expectation. Therefore, there exists a constant $a > 0$ such that

$$\mathrm{E}\exp\big(v\big|V_n(t_2) - V_n(t_1)\big|\big) \le a\exp\left(\frac{2}{c}v^2(t_2 - t_1)\right). \tag{3.10}$$

Consequently, given $x > 0$,

$$\mathrm{P}\big\{\big|V_n(t_2) - V_n(t_1)\big| \ge x\big\} \le a\exp\left(\frac{2}{c}v^2(t_2 - t_1) - vx\right), \tag{3.11}$$

for every $v = O(n^{\rho})$. We see that

$$\bar{v} = v(x) := \frac{cx}{4(t_2 - t_1)},$$

which minimizes the exponent on the right, is at most n^{ρ} provided that $x \le x_0 := n^{\rho}(t_2 - t_1)$. Plugging $\bar{v}$ into (3.11) we get

$$\mathrm{P}\big\{\big|V_n(t_2) - V_n(t_1)\big| \ge x\big\} \le a\exp\left(-\frac{cx^2}{8(t_2 - t_1)}\right), \qquad x \le x_0. \tag{3.12}$$

For $x > x_0$, we choose $v = v(x_0)$ and easily obtain

$$\mathrm{P}\{|V_n(t_2) - V_n(t_1)| \ge x\} \le a \exp\left(-\frac{c}{8} n^\rho x\right). \tag{3.13}$$

So (3.6) is proved. Thus

$$\sup_{\{t,s\in[t^*(n),t_2(n)]:|t-s|\le h\}} \mathrm{P}\{|V_n(t) - V_n(s)| \ge \varepsilon\} \le ae^{-c\varepsilon^2/(8h)} + ae^{-cn^\rho\varepsilon/8},$$

so letting $h \downarrow 0$ and $n \to \infty$, in that order, we prove the second line in (3.3).

(c) Let $t^*(n) \le t \le t_2(n)$. As in (b), we write

$$V_n(t) = n^{-1/4} t\left(\lambda_{k(t)} - m\right) + O(n^{-1/4}),$$

$$m := \sum_{j=k(t)}^{n} \frac{q^j}{1-q^j},$$

and use Proposition to obtain

$$\mathrm{E}\exp\left(v|V_n(t)|\right) \le a \exp\left(\frac{v^2}{2c} t(1-t)\right), \qquad v = O(n^\rho)$$

(cf. (3.10)). Then, analogously to (3.12), (3.13),

$$P\{|V_n(t)| \ge x\} \le \begin{cases} ae^{-cx^2/(2t(1-t))}, & \text{if } x \le x_0, \\ ae^{-n^\rho x/2}, & \text{if } x \ge x_0, \end{cases}$$

where $x_0 = n^\rho t(1-t)/c$. Therefore

$$\begin{aligned} \mathrm{E}|V_n(t)|^\mu &\le a\mu \int_0^{x_0} x^{\mu-1} \exp\left(-\frac{cx^2}{2t(1-t)}\right) dx \\ &\quad + a\mu \int_{x_0}^{\infty} x^{\mu-1} \exp\left(-\frac{n^\rho x}{2}\right) dx \\ &= O\left(t^{\mu/2}(1-t)^{\mu/2} + n^{-\mu\rho}\right), \end{aligned}$$

which proves the third line in (3.3). ∎

Once we establish convergence of finite-dimensional distributions of $V_n(\cdot)$, Lemma 4 will guarantee convergence in distribution of $F(V_n)$ for a class $\mathscr{F}$ of the integral functionals F of a form

$$F(x) = \int_0^1 \phi(t, x(t))\, dt.$$

Here $\phi(t, x)$ is continuous for $(t, x) \in D := (0,1) \times \mathbb{R}$, and such that, for some $\mu > 0$, $\alpha < \mu/2 + 1$, $\beta < \mu/6 + 1$,

$$|\phi(t,x)| = O\left(\frac{|x|^{\mu}}{t^{\alpha}(1-t)^{\beta}}\right), \tag{3.14}$$

uniformly over D.

THEOREM 5. *Given $k \geq 1$ and $0 \leq t_1 < t_2 < \cdots < t_k \leq 1$,*

$$(V_n(t_1), \dots, V_n(t_k)) \xrightarrow{\mathscr{D}} (V(t_1), \dots, V(t_k)). \tag{3.15}$$

Here $V(\cdot)$ is a Gaussian process, with $\mathrm{E}\, V(t) \equiv 0$ and the covariance function given by

$$\begin{aligned} K(t_1, t_2) &= c^{-1}\left[t_1(1-t_2) - \tfrac{1}{2} l(t_1) l(t_2)\right], \qquad 0 \leq t_1 \leq t_2 \leq 1, \\ l(t) &= \frac{1}{c}\log\frac{t^t}{(1-t)^{1-t}}; \end{aligned} \tag{3.16}$$

each separable version of $V(\cdot)$ is continuous on $[0,1]$. Furthermore for every $F \in \mathscr{F}$,

$$\int_0^1 \phi(t, V_n(t))\, dt \xrightarrow{\mathscr{D}} \int_0^1 \phi(t, V(t))\, dt. \tag{3.17}$$

Notes. 1. We had not expected the bounds for α and β in (3.14) to be different. With extra work, we could get a better bound for β, namely $\beta < 3\mu/14 + 1$, but it would still be inferior to $\alpha < \mu/2 + 1$.

2. Had we proved (3.4), we would have been able to ascertain convergence in distribution of all the functionals F continuous in uniform metric, not just those from $\mathscr{F}$. One such functional is $\sup_t |x(t)|$, and so we would have concluded, at least, that $\sup_t |V_n(t)| = O_p(1)$. (Theorem 2 implies a weaker result, namely $\sup_t |V_n(t)| = O_p(\log n)^{1/2}$.)

COROLLARY. *Let*

$$F(x) = \int_0^1 \phi(t) x(t)\, dt,$$

where $\phi(t) = O(t^{-\alpha}(1-t)^{-\beta})$, $\alpha < \frac{3}{2}$, $\beta < \frac{7}{6}$. Then

$$F(V_n) \xrightarrow{\mathscr{D}} \mathscr{N}(0, \sigma^2), \tag{3.18}$$

where

$$\sigma^2 = \int_0^1 \int_0^1 \phi(t_1)\phi(t_2) K(t_1, t_2)\, dt_1\, dt_2. \tag{3.19}$$

Proof of Theorem 5 and *Corollary*. (a) It is enough to consider $t_1 > 0$, $t_k < 1$. For simplicity of notations, let $k = 2$. We will prove that for all *reals* v_1, v_2,

$$\mathrm{E}\, e^{v_1 V_n(t_1) + v_2 V_n(t_2)} \to \exp\left(\tfrac{1}{2}v_1^2 K(t_1, t_1) + v_1 v_2 K(t_1, t_2) + \tfrac{1}{2}v_2^2 K(t_2, t_2)\right). \tag{3.20}$$

(This certainly implies that

$$(V_n(t_1), V_n(t_2)) \xrightarrow{\mathscr{D}} (V(t_1), V(t_2)).)$$

Denote $k_b = k(t_b)$, $b = 1, 2$, and, given two reals u_1, u_2, set $x_b = \exp(u_b n^{-1/4})$, $b = 1, 2$. Then, just as for (2.6), we write

$$p(n)\mathrm{E}\left(x_1^{X(k_1)-X(k_2)} x_2^{X(k_2)}\right) = \frac{1}{2\pi r^N}\int_{-\pi}^{\pi} e^{-in\theta} F_n(\mathbf{x}, re^{i\theta})\, d\theta,$$

$$F_n(\mathbf{x}, z) := \prod_{j=1}^{\infty} \frac{1}{1 - z^j} \cdot \prod_{j=k_1}^{k_2 - 1} \frac{1 - z^j}{1 - x_1 z^j} \cdot \prod_{j=k_2}^{\infty} \frac{1 - z^j}{1 - x_2 z^j}; \tag{3.21}$$

here $r = \exp(-\nu n^{-1/2})$, $\nu > 0$. With this formula in place of (2.6), the argument begins similarly to the proof of Proposition, except that this time $\delta = \frac{1}{2}$, $\delta_1 = 0$. However, the radius r (ν, equivalently) will have to be chosen more carefully.

(b) Let us first dispense with the overall contribution to the value of the integral in (3.21) made by "large" θ's, with $|\theta| \geq \theta_n := n^{-3/4}\log n$. Just as for (2.9), we write

$$F_n(\mathbf{x}, re^{i\theta}) \leq F_n(\mathbf{x}, r)\exp\left(-\frac{a\theta^2}{n^{-3/2} + \theta^2 n^{-1/2}}\right),$$

for some positive $a > 0$. Considering separately $|\theta| \in [\theta_n, n^{-1/2}]$, and $|\theta| \geq n^{-1/2}$, we conclude that

$$\begin{aligned} r^{-n} &\int_{|\theta| \geq \theta_n} \left| F_n(\mathbf{x}, re^{i\theta}) \right| d\theta \\ &= O\left(r^{-n} F_n(\mathbf{x}, r) \int_{|\theta| \geq \theta_n} \exp(-an^{3/2}\theta^2/2)\, d\theta \right) \\ &\leq r^{-n} F_n(\mathbf{x}, r) \exp(-a \log^2 n/3). \end{aligned} \tag{3.22}$$

This estimate holds for every $r = e^{-\nu/n^{1/2}}$. As usual with a saddle point-type method, we should choose an r which is, asymptotically at least, a stationary point of $\rho^{-n}F_n(\mathbf{x}, \rho)$. Such an r will depend on $\mathbf{x}$! In view of (3.21) *and* (2.8), we are content to set $r = e^{-\tau}$, choosing τ close enough to a stationary point of

$$H(t, \mathbf{x}) := nt + \frac{\pi^2}{6t} + \sum_{j=k_1}^{k_2-1} \log \frac{1 - e^{-j}}{1 - x_1 e^{-tj}} + \sum_{j=k_2}^{\infty} \log \frac{1 - e^{-tj}}{1 - x_2 e^{-tj}},$$

that is, to a root of

$$H_t(t, \mathbf{x}) = n - \frac{\pi^2}{6t^2} - \sum_{b=1}^{2} (x_b - 1) S^{(b)}(t, x_b) = 0,$$

$$S^{(1)}(t, x) := \sum_{j=k_1}^{k_2-1} \frac{je^{-tj}}{(1 - x_1 e^{-tj})(1 - e^{-tj})},$$

$$S^{(2)}(t, x) := \sum_{j=k_2}^{\infty} \frac{je^{-tj}}{(1 - x_2 e^{-tj})(1 - e^{-tj})}.$$

The expression for H_t immediately suggests that we select

$$\tau = \tau^* \left(1 + \sum_{b=1}^{2} \frac{x_b - 1}{2n} S^{(b)}(\tau^*, x_b) \right), \qquad \tau^* := \frac{c}{\sqrt{n}}; \tag{3.23}$$

here $c = \pi/\sqrt{6}$. (So $e^{-\tau^*}$ equals q, the parameter from the previous sections!) A direct computation, based on

$$1 - xq^j = (1 + o(1))(1 - q^j), \qquad j \ge k_1,$$

shows then that

$$\begin{aligned} H_t(\tau, \mathbf{x}) &= n - n\left(1 + \sum_{b=1}^{2} \frac{x_b - 1}{2n} S^{(b)}(\tau^*, x_b) \right)^{-2} \\ &\quad - \sum_{b=1}^{2} (x_b - 1) S^{(b)}(\tau, x_b) \\ &= O\left[n^{-3/2} \sum_{b=1}^{2} S^{(b)}(\tau^*, x_b) \left(S^{(b)}(\tau^*, x_b) + \tau^* S_t^{(b)}(\tau^*, x_b) \right) \right] \\ &= O\left(n^{-3/2} \sum_{b=1}^{2} S^{(b)} \left(S^{(b)} + \tau^* T^{(b)} \right) \right). \end{aligned}$$

Here $S^{(b)}$ and $T^{(b)}$ denote $S^{(b)}(\tau^*, 1)$ and $S_t^{(b)}(\tau^*, 1)$, respectively, and it is straightforward that

$$S^{(b)} = O(n), \qquad T^{(b)} = O(n^{3/2}) \qquad (b = 1, 2). \tag{3.24}$$

Therefore

$$H_t(\tau, \mathbf{x}) = O(n^{1/2}). \tag{3.25}$$

(Notice that the derivative $H_t(t, \mathbf{x})$ at $t = \tau^*$ is considerably larger, of order $n^{3/4}$. As we will see shortly, the bound (3.25) is crucial for the argument.) Analogously, for t between τ^* and τ,

$$H_{tt}(t, \mathbf{x}) = \frac{\pi^2}{3t^2} - \sum_{b=1}^{2} (x_b - 1) S_t^{(b)}(t, \mathbf{x}) = \frac{2\sqrt{6}}{\pi} n^{3/2} + O(n^{5/4}). \tag{3.26}$$

Using (3.24) and (3.25), we obtain

$$\begin{aligned} H(\tau, \mathbf{x}) &= H(\tau^*, \mathbf{x}) - H_t(\tau, \mathbf{x})(\tau^* - \tau) - \frac{1}{2} H_{tt}(\tilde{t}, \mathbf{x})(\tau^* - \tau)^2 \\ &= \pi\sqrt{\frac{2n}{3}} + \sum_{j=k_1}^{k_2 - 1} \log \frac{1 - q^j}{1 - x_1 q^j} + \sum_{j=k_2}^{\infty} \log \frac{1 - q^j}{1 - x_2 q^j} \\ &\quad - \frac{c}{4n^2} \left(\sum_{b=1}^{2} S^{(b)} \right)^2 + O(n^{-1/4}), \end{aligned} \tag{3.27}$$

with the error term coming from (3.25) and $\tau^* - \tau = O(n^{-3/4})$. Here, expanding $\log(1 - x_1 q^j)$ in powers of $x_1 - 1$, and $x_1 - 1 = \exp(u_1 n^{-1/4}) - 1$ in powers of u_1, we obtain

$$\begin{aligned} \sum_{j=k_1}^{k_2 - 1} \log \frac{1 - q^j}{1 - x_1 q^j} &= (x_1 - 1) \sum_{j=k_1}^{k_2 - 1} \frac{q^j}{1 - q^j} + \frac{(x_1 - 1)^2}{2} \sum_{j=k_1}^{k_2 - 1} \frac{q^{2j}}{(1 - q^j)^2} \\ &\quad + O\left(|x_1 - 1|^3 \sum_{j=k_1}^{k_2 - 1} \frac{q^{3j}}{(1 - q^j)^3} \right) \\ &= u_1 n^{-1/4} \sum_{j=k_1}^{k_2 - 1} \frac{q^j}{1 - q^j} + \frac{1}{2} u_1^2 n^{-1/2} \sum_{j=k_1}^{k_2 - 1} \frac{q^j}{(1 - q^j)^2} \\ &\quad + O(n^{-1/4}). \end{aligned} \tag{3.28}$$

An analogous expansion holds for the second sum in (3.27). Also, with the help of $\tau - \tau^* = O(n^{-3/4})$, for the logarithmic term in the expansion (2.8) we have

$$\frac{1}{2}\operatorname{Log}\frac{\tau}{2\pi} = \log(24n)^{-1/4} + O(n^{-1/4}). \tag{3.29}$$

Combining (2.8) and (3.27)–(3.29), we obtain a key formula,

$$r^{-n}F_n(\mathbf{x}, r) = \frac{e^{\pi\sqrt{2n/3}}}{(24n)^{1/4}}\exp\left(\sum_{b=1}^{2}\left(u_b m_b n^{-1/4} + \frac{1}{2}u_b^2\sigma_b^2 n^{-1/2}\right) - \frac{c}{4n^2}\left(\sum_{b=1}^{2} u_b S^{(b)}\right)^2 + O(n^{-1/4})\right), \tag{3.30}$$

where

$$\begin{aligned} m_1 &:= \sum_{j=k_1}^{k_2-1}\frac{q^j}{1-q^j}, \qquad \sigma_1^2 := \sum_{j=k_1}^{k_2-1}\frac{q^j}{(1-q^j)^2},\\ m_2 &:= \sum_{j=k_2}^{\infty}\frac{q^j}{1-q^j}, \qquad \sigma_2^2 := \sum_{j=k_2}^{\infty}\frac{q^j}{(1-q^j)^2}. \end{aligned} \tag{3.31}$$

Thus, the bound (3.22) becomes explicit.

(c) Turn now to the small θ's, with $|\theta| \le \theta_n$. First of all, for those θ's,

$$H(\tau - i\theta, \mathbf{x}) = H(\tau, \mathbf{x}) + H_t(\tau, \mathbf{x})(-i\theta) + \tfrac{1}{2}H_{tt}(\tau, \mathbf{x})(-i\theta)^2 + O\left(n^2\theta_n^3\right),$$

since it can be demonstrated, analogously to (3.26), that

$$\sup\left\{\left|H_{ttt}(t, \mathbf{x})\right| : |t - \tau| \le \theta_n\right\} = O(n^2).$$

(For complex t in question, $H_{ttt}(t, \mathbf{x}) \sim -\pi^2/t^4$.) Besides,

$$\frac{1}{2}\operatorname{Log}\frac{\tau - i\theta}{2\pi} = \frac{1}{2}\log\frac{\tau}{2\pi} + O\left(\theta_n n^{1/2}\right).$$

Therefore

$$r^{-n}e^{-in\theta}F_n(\mathbf{x}, re^{i\theta}) = r^{-n}F_n(\mathbf{x}, r)\exp\left(-i\theta H_t(\tau, \mathbf{x}) - \frac{\theta^2}{2}H_{tt}(\tau, \mathbf{x}) + O(n^{-1/4})\right).$$

The rest is short. We integrate

$$\int_{|\theta|\le \theta_n} \exp\left(-i\theta H_t(\tau,\mathbf{x}) - \frac{\theta^2}{2} H_{tt}(\tau,\mathbf{x})\right) d\theta$$
$$= \sqrt{\frac{2\pi}{H_{tt}(\tau,\mathbf{x})}} \exp\left(-\frac{H_t^2(\tau,\mathbf{x})}{2H_{tt}(\tau,\mathbf{x})}\right) + O\left(H_{tt}^{-1/2} e^{-\theta_n^2 H_{tt}(\tau,\mathbf{x})/2}\right)$$
$$= \frac{\pi}{6^{1/4} n^{3/4}} \left(1 + O(n^{-1/4})\right).$$

(Indeed, by (3.25) and (3.26),

$$\frac{H_t^2(\tau,\mathbf{x})}{H_{tt}(\tau,\mathbf{x})} = O(n^{-1/2}), \qquad \theta_n^2 H_{tt}(\tau,\mathbf{x}) \ge a \log^2 n,$$

for some constant $a > 0$.) Since also

$$\int_{|\theta|\le \theta_n} \left| \exp\left(-i\theta H_t(\tau,\mathbf{x}) - \frac{\theta^2}{2} H_{tt}(\tau,\mathbf{x})\right) \right| d\theta \sim \frac{\pi}{6^{1/4} n^{3/4}},$$

we obtain

$$\int_{|\theta|\le \theta_n} r^{-n} e^{-in\theta} F_n(\mathbf{x}, re^{i\theta})\, d\theta = r^{-n} F_n(\mathbf{x}, r) \frac{\pi}{6^{1/4} n^{3/4}} \left(1 + O(n^{-1/4})\right). \tag{3.32}$$

Combining (3.21), (3.22), (3.30)–(3.32), and the Hardy–Ramanujan formula (2.7), we arrive at

$$\mathrm{E}\exp\left(u_1(X(k_1) - X(k_2))n^{-1/4} + u_2 X(k_2) X(k_2) n^{-1/4}\right)$$
$$= \exp\left(\sum_{b=1}^{2} \left(u_b m_b n^{-1/4} + \frac{1}{2} u_b^2 \sigma_b^2 n^{-1/2}\right) - \frac{c}{4n^2}\left(\sum_{b=1}^{2} u_b S^{(b)}\right)^2 + O(n^{-1/4})\right),$$

or, substituting $u_1 = \eta_1$, $u_2 = \eta_1 + \eta_2$,

$$\begin{aligned} &\mathrm{E}\exp\left(\sum_{b=1}^{2} u_b (X(k_r) - \mu(k_r)) n^{-1/4}\right) \\ &\quad = \exp\left(\frac{1}{2} \sum_{1\le b,\ \beta\le 2} f(k_b, k_\beta)\eta_b \eta_\beta + O(n^{-1/4})\right). \end{aligned} \tag{3.33}$$

Here

$$\mu(k) := \sum_{j=k}^{\infty} \frac{q^j}{1-q^j},$$

$$f(k_b, k_\beta) := n^{-1/2} \sum_{j=k_b \vee k_\beta} \frac{q^j}{(1-q^j)^2} - \frac{c}{2n^2} \sum_{j=k_b}^{\infty} \frac{jq^j}{(1-q^j)^2} \cdot \sum_{j=k_\beta}^{\infty} \frac{jq^j}{(1-q^j)^2}.$$

For $k = k(t)$, we have

$$\mu(k) = n^{1/2} \int_{kn^{-1/2}}^{\infty} \frac{e^{-cx}}{1-e^{-cx}}\,dx + O\left(\frac{n^{1/2}}{k}\right) = \frac{n^{1/2}}{c}\int_t^1 \frac{d\tau}{\tau} + O(t^{-1}) = \frac{n^{1/2}}{c}\log\frac{1}{t} + O(t^{-1}),$$

$$n^{-1/2} \sum_{j=k}^{\infty} \frac{q^j}{(1-q^j)^2} = \int_{kn^{-1/2}}^{\infty} \frac{e^{-cx}}{(1-e^{-cx})^2}\,dx + O\left(\frac{n^{1/2}}{k^2}\right) = \frac{1}{c}\int_t^1 \frac{d\tau}{\tau^2} + O(n^{-1/2}t^{-2}) = \frac{1}{c}\frac{1-t}{t} + O(n^{-1/2}t^{-2}), \tag{3.34}$$

$$n^{-1} \sum_{j=k}^{\infty} \frac{jq^j}{(1-q^j)^2} \approx \int_{kn^{-1/2}}^{\infty} \frac{xe^{-cx}}{(1-e^{-cx})^2}\,dx = \frac{1}{c^2 t}\log\frac{(1-t)^{1-t}}{t^t}.$$

Thus, setting $\eta_b = v_b t_b$ ($b = 1, 2$), from (3.33), (3.34) we obtain: by the definition of the process $V(\cdot)$,

$$\left(t_1 \frac{X(k(t_1)) - c^{-1}n^{1/2}\log(1-t_1)^{-1}}{n^{1/4}},\; t_2 \frac{X(k(t_2)) - c^{-1}n^{1/2}\log(1-t_2)^{-1}}{n^{1/4}}\right) \xrightarrow{\mathscr{D}} (V(t_1), V(t_2)),$$

and we recall that $\{X(k)\} \overset{\mathcal{D}}{\equiv} \{\lambda_k\}$. Furthermore, for $0 \le t_1 \le t_2 \le 1$,

$$\begin{aligned} \mathrm{E}\big(V(t_2) - V(t_1)\big)^2 &= K(t_1, t_1) - 2K(t_1, t_2) + K(t_2, t_2) \\ &= c^{-1}\Big[(t_2 - t_1) - (t_2 - t_1)^2 - \tfrac{1}{2}\big(l(t_1) - l(t_2)\big)^2\Big] \\ &\le c^{-1}(t_2 - t_1), \end{aligned}$$

so, since $V(t_2) - V(t_1)$ is Gaussian, with zero mean, we have

$$\mathrm{E}\big(V(t_2) - V(t_1)\big)^4 \le \mathrm{const}(t_2 - t_1)^2.$$

This inequality implies continuity of a separable version of $V(\cdot)$ on $[0,1]$ (Durrett [6]).

(d) Now let us prove the weak convergence of $F(V_n)$ for $F \in \mathcal{F}$. Without loss of generality, we assume that $\alpha, \beta > 1$.

Introduce $\varepsilon_m = 1/m$, $m \ge 1$, and write

$$\begin{aligned} F(V_n) &= \int_0^{\varepsilon_m} \phi(t, V_n(t))\, dt + \int_{\varepsilon_m}^1 \phi(t, V_n(t))\, dt \\ &\quad + \int_{1-\varepsilon_m}^1 \phi(t, V_n(t))\, dt \\ &= I_{nm} + K_{nm} + L_{nm}. \end{aligned} \tag{3.35}$$

We write

$$I_{nm} = \sum_{b=1}^{3} I_{nm}^{(b)},$$

where, using the bound for ϕ and Lemma 4,

$$\begin{aligned} I_{nm}^{(1)} &:= \int_{t_1(n)}^{n^{-\nu}} \phi(t, V_n(t))\, dt = O_p(n^{-s_1} \log^{\mu} n), \\ s_1 &= \nu(\mu - \alpha + 1) - \mu/4, \\ I_{nm}^{(2)} &:= \int_{n^{-\nu}}^{t^*(n)} \phi(t, V_n(t))\, dt = O_p(n^{-s_2} \log^{\mu} n), \\ s_2 &= \delta_0(\mu/2 + 1 - \alpha), \\ \mathrm{E}|I_{nm}^{(3)}| &= \int_{t^*(n)}^{\varepsilon_m} \mathrm{E}\big|\phi(t, V_n(t))\big|\, dt \\ &= O\left(\int_{t^*(n)}^{\varepsilon_m} (t^{\mu/2-\alpha} + n^{-\mu\rho} t^{-\alpha})\, dt\right) \\ &= O\big(\varepsilon_m^{\mu/2+1-\alpha} + n^{-\mu\rho}\varepsilon_m^{1-\alpha} + n^{-s_3}\big), \\ s_3 &= \min\big\{\delta_0(\mu/2 + 1 - \alpha), \mu\rho - \delta_0(\alpha - 1)\big\}. \end{aligned} \tag{3.36}$$

Since $\nu/2 > \alpha - 1$, s_2 is clearly positive, and so is s_1 if we select ν sufficiently close (from below) to $\frac{1}{2}$. Likewise, s_3 is made positive as well, by picking ρ arbitrarily close (from below) to $\frac{1}{12}$ and δ_0 sufficiently small to satisfy

$$\rho > \frac{\delta_0}{2}, \qquad \rho < \frac{1}{12} - \frac{2}{3}\delta_0.$$

For such a choice of ν, ρ, δ_0, we have

$$I_{nm} = O_p\big(\varepsilon_m^{\mu/2+1-\alpha} + n^{-\mu\rho}\varepsilon_m^{1-\alpha} + n^{-s}\big), \qquad s > 0. \tag{3.37}$$

Let us bound L_{nm}. Pick a $\chi \in (\frac{1}{2}, 1)$ and write

$$\begin{aligned} L_{nm} &= \int_{1-\varepsilon_m}^{1-n^{-\chi}} \phi(t, V_n(t))\, dt + \int_{1-n^{-\chi}}^{1} \phi(t, V_n(t))\, dt \\ &= L_{nm}^{(1)} + L_{nm}^{(2)}. \end{aligned}$$

First, analogously to (3.36),

$$\begin{aligned} \mathrm{E}\big|L_{nm}^{(1)}\big| &= O\big(\varepsilon_m^{\mu/2+1-\beta} + n^{-\mu\rho}\varepsilon_m^{1-\beta} + n^{-s_4}\big), \\ s_4 &= \min\{\chi(\mu/2 + 1 - \beta), \mu\rho + \chi(1-\beta)\}. \end{aligned} \tag{3.38}$$

Since $\mu/6 + 1 > \beta$, we can select ρ and χ so close (from below) to $\frac{1}{12}$ and (from above) to $\frac{1}{2}$, respectively, that $\mu\rho + \chi(1-\beta) > 0$. In that case s_4 is positive. Second, by the definition of $k(t)$,

$$k(t) \geq (1 + o(1))\,\chi\frac{\sqrt{n}}{c}\log n, \qquad \text{for } t \geq 1 - n^{-\chi},$$

whence, with high probability, $\lambda_{k(t)} \equiv 0$ for such t's, as

$$\lambda_1^* = \big(1 + o_p(1)\big)\frac{\sqrt{n}}{2c}\log n.$$

Therefore, whp,

$$V_n(t) = -n^{-1/4}t \cdot \frac{\sqrt{n}}{c}\log\frac{1}{t} = O\big(n^{1/4}(1-t)\big),$$

and consequently

$$\begin{aligned} \int_{1-n^{-\chi}}^{1} \big|\phi(t, V_n(t))\big|\, dt &= O_p\left(n^{\mu/4}\int_{1-n^{-\chi}}^{1}(1-t)^{\mu-\beta}\, dt\right) \\ &= O\big(n^{-(\mu/2+1-\beta)/2}\big). \end{aligned}$$

From this bound and (3.38) we infer

$$L_{nm} = O_p\big(\varepsilon_m^{\mu/2+1-\beta} + n^{-\mu\rho}\varepsilon_m^{1-\beta} + n^{-s'}\big), \qquad s' > 0,$$

and, combining it with (3.37), we conclude that, for some $s^* > 0$,

$$I_{nm} + L_{nm} = O_p\big(\varepsilon_m^{\mu/2+1-\alpha} + \varepsilon_m^{\mu/2+1-\beta} + n^{-\mu\rho}\big(\varepsilon_m^{1-\beta} + \varepsilon_m^{1-\alpha}\big) + n^{-s^*}\big). \tag{3.39}$$

As for K_{nm}, the process $\{V_n(t)\}_{t\in[\varepsilon_m,1-\varepsilon_m]}$ satisfies all the conditions of a theorem due to Gihman and Skorohod [13, Chap. 9, Sect. 7]. Specifically, the random functions $V_n(\cdot)$ are stochastically equicontinuous, uniformly on $[\varepsilon_m, 1-\varepsilon_m]$, and there exists a function $\psi(x)$, namely $\psi(x) = |x|^{\mu+1}$, such that $\psi(x) \to \infty$ as $|x| \to \infty$ and

$$\sup_n \sup_{t\in[\varepsilon_m,1-\varepsilon_m]} \mathrm{E}\,\psi\big(V_n(t)\big) < \infty, \qquad \lim_{a\to\infty} \sup_{t\in[\varepsilon_m,1-\varepsilon_m]} \sup_{|x|>a} \frac{|\phi(t,x)|}{\psi(x)} = 0.$$

By that theorem then

$$K_{nm} \xrightarrow{\mathscr{D}} K_m := \int_{\varepsilon_m}^{1-\varepsilon_m} \phi(t, V(t))\, dt, \qquad \text{as } n \to \infty. \tag{3.40}$$

Furthermore

$$\begin{aligned}\int_{[0,\,\varepsilon_m]\cup[1-\varepsilon_m]} \mathrm{E}\big|\phi(t,V(t))\big|\, dt &= O\left(\int_{[0,\,\varepsilon_m]\cup[1-\varepsilon_m]} \frac{E|V(t)|^{\mu}}{t^{\alpha}(1-t)^{\beta}}\, dt\right)\\ &= O\left(\int_0^{\varepsilon_m} t^{\mu/2-\alpha}\, dt + \int_{1-\varepsilon_m}^{1} t^{\mu/2-\beta}\, dt\right)\\ &= O\big(\varepsilon_m^{\mu/2+1-\alpha} + \varepsilon_m^{\mu/2+1-\beta}\big),\end{aligned}$$

since $\mathrm{Var}\, V(t) \le c^{-1} t(1-t)$, and $V(t)$ is Gaussian, with zero mean. So (Fubini theorem) $\phi(t, V(t))$ is almost surely integrable on $[0, \varepsilon_m] \cup [1-\varepsilon_m, 1]$ and

$$\mathrm{E}\left|\int_{[0,\,\varepsilon_m]\cup[1-\varepsilon_m,1\}} \phi(t,V(t))\, dt\right| = O\big(\varepsilon_m^{\mu/2+1-\alpha} + \varepsilon_m^{\mu/2+1-\beta}\big). \tag{3.41}$$

First letting $n \to \infty$ and then letting $m \to \infty$, by (3.39)–(3.41) we obtain that $F(V_n) \xrightarrow{\mathscr{D}} F(V)$. ∎

As an illustration, consider a function $r(\lambda) = \chi_\lambda(\tau)/d(\lambda)$, where $d(\lambda)$ is the dimension of the irreducible representation of S_n that corresponds to the partition λ, and $\chi_\lambda(\tau)$ is the value of the character of the representation at a transposition τ. According to Frobenius [12] (see Ingram [18]),

$$r(\lambda) = \frac{1}{n(n-1)} \sum_j \left[\lambda_j^2 - (2j-1)\lambda_j \right] = \frac{1}{\binom{n}{2}} \sum_j \left(\binom{\lambda_j}{2} - \binom{\lambda_j^*}{2} \right). \tag{3.42}$$

This function played an important role in the well-known analysis of the card-shuffling problem performed by Diaconis and Shahshahani [4]; see also Diaconis [5]. The prominence of this function in their proof is due to the remarkable fact that it determines the set of the eigenvalues of the transition probability matrix that describes the shuffling process. The authors were able to show that, except for the extreme partitions λ, the values of $r(\lambda)$ are quite small, which was a key element element of the argument. (To be sure, those rare partitions turned out to be influential enough to determine a concise threshold number of shuffles.)

Under the assumption that λ is distributed uniformly, $r(\lambda)$ is a random variable. Since $d(\lambda) = d(\lambda^*)$, it is clear from (3.42) that E $r(\lambda) = 0$. Let us study the asymptotic behavior of $r(\lambda)$.

Guided by Theorem 2, we set $\lambda_j = E(j) + R_j$ and obtain

$$\sum_j \left[\lambda_j^2 - (2j-1)\lambda_j \right] = \sum_{b=1}^{3} U_b + n,$$

$$U_1 := \sum_j \left[E^2(j) - 2jE(j) \right],$$

$$U_2 := \sum_j R_j^2,$$

$$U_3 := \sum_j \left[2(E(j) - j) R_j \right],$$

where

$$j \le \lambda_1^* = \left(1 + o_p(1)\right) \frac{\sqrt{n}}{2c} \log n.$$

It is easy to check that

$$U_1 = O_p\left(n \log^2 n\right) + \frac{n^{3/2}}{c^3} \int_0^\infty \left(\log^2 \frac{1}{1-e^{-y}} - 2y \log \frac{1}{1-e^{-y}} \right) dy \tag{3.43}$$
$$= O_p\left(n \log^2 n\right),$$

as the integral equals zero. Indeed, substituting $y = \log(1-t)^{-1}$,

$$\begin{aligned}\int_0^\infty \log^2 \frac{1}{1-e^{-y}}\,dy &= \int_0^1 \left(\log^2\frac{1}{t}\right)\frac{1}{1-t}\,dt\\ &= 2\int_0^1 \left(\log\frac{1}{1-t}\log\frac{1}{t}\right)\frac{1}{t}\,dt\\ &= 2\int_0^1 \left(\log\frac{1}{t}\log\frac{1}{1-t}\right)\frac{1}{1-t}\,dt,\\ 2\int_0^\infty y\log\frac{1}{1-e^{-y}}\,dy &= 2\int_0^1\left(\log\frac{1}{1-t}\log\frac{1}{t}\right)\frac{1}{1-t}\,dt.\end{aligned}$$

Using Theorem 2, we can show that

$$U_2 = O_p(n\log^2 n) \tag{3.44}$$

also. Turn finally to U_3. Setting $R(x) = R_{[x]+1}$, switching to integration, and substituting $x = (\sqrt{n}/c)\log(1-t)^{-1}$, we transform the formula for U_3 into

$$\begin{aligned}U_3 &= 2\int_0^\infty (E(x)-x)R(x)\,dx + O_p(n\log^2 n)\\ &= 2\int_0^\infty (E(x)-x)\big(\lambda_{[x]+1} - E(x)\big)\,dx + O_p(n\log^2 n)\\ &= 2\frac{n^{5/4}}{c^2}\int_0^1 V_n(t)\frac{1}{t(1-t)}\log\frac{1-t}{t}\,dt + O_p(n\log^2 n).\end{aligned}$$

Now the corresponding function

$$\phi(t,x) = \frac{2x}{c^2 t(1-t)}\log\frac{1-t}{t}$$

obviously satisfies the condition of Theorem 5, with parameters $\mu = 1$ and $\alpha = \beta = 1.01$, say. So, invoking (3.43), (3.44), we conclude that

$$n^{3/4} r(\lambda) \stackrel{\mathscr{D}}{\to} \mathscr{N}(0,\sigma^2),$$

with σ^2 given by (3.19) and (3.16). It follows, in particular, that typically $r(\lambda)$ is of order $n^{-3/4}$.

REFERENCES

1. G. E. Andrews, "The Theory of Partitions," Encyclopedia of Mathematics and Its Applications, Vol. 2, Addison–Wesley, Reading, MA, 1976.
2. R. Arratia and S. Tavaré, Independent process approximations for random combinatorial structures, *Adv. in Math.* **104** (1994), 90–154.
3. R. Arratia, D. Stark, and S. Tavaré, Total variation asymptotics for Poisson process approximations of logarithmic combinatorial assemblies, *Ann. Probab.* **23** (1995), 1347–1388.
4. P. Diaconis and M. Shahshahani, Generating a random permutation with random transpositions, *Z. Wahrsch. Verw. Gebiete* **57** (1981), 159–179.
5. P. Diaconis, "Group Representations in Probability and Statistics," Institute of Mathematical Statistics, Lecture Notes–Monograph Series, Vol. 11, Hayward, CA, 1988.
6. R. Durrett, "Probability: Theory and Examples," Wadsworth & Brooks/Cole, Belmont, CA, 1991.
7. P. Erdös and J. Lehner, The distribution of the number of summands in the partition of a positive integer, *Duke Math. J.* **8** (1941), 335–345.
8. P. Erdös and M. Szalay, On the statistical theory of partitions, *in* "Topics in Classical Number Theory" (G. Halász, Eds.), Vol. I, pp. 397–450, North-Holland, Amsterdam, 1984.
9. P. Erdös and P. Turán, On some problems of a statistical group theory, III, *Acta Math. Sci. Hungar.* **18** (1967), 309–320.
10. J. S. Frame, G. de B. Robinson, and R. M. Thrall, The hook graphs of the symmetric groups, *Canad. J. Math.* **6** (1954), 316–324.
11. B. Fristedt, The structure of random partitions of large integers, *Trans. Amer. Math. Soc.* **337** (1993), 703–735.
12. G. Frobenius, Über die charactere der symmetrischen gruppe, *Sitzungsber. Preuss. Akad. Berlin* (1903), 328–358.
13. I. I. Gihman and A. V. Skorohod, "The Theory of Stochastic Processes I," Springer-Verlag, New York, 1974.
14. W. M. Y. Goh and E. Schmutz, The number of distinct part sizes in a random integer partition, *J. Combin. Theory Ser. A* **69** (1995), 149–158.
15. C. Greene, A. Nijenhuis, and H. Wilf, A probabilistic proof of a formula for the number of Young tableaux of a given shape, *Adv. in Math.* **31** (1979), 104–109.
16. C. Greene, A. Nijenhuis, and H. Wilf, Another probabilistic formula in the theory of Young tableaux, *J. Combin. Theory Ser. A* **37** (1984), 127–135.
17. G. H. Hardy and S. Ramanujan, Asymptotic formulae in combinatory analysis, *Proc. London Math. Soc. (2)* **17** (1918), 75–115.
18. R. E. Ingram, Some characters of the symmetric group, *Proc. Amer. Math. Soc.* **1** (1950), 358–369.
19. D. E. Knuth, "The Art of Computer Programming," Vol. 3, "Sorting and Searching," Addison–Wesley, Reading, MA, 1973.
20. V. F. Kolchin, "Random Mappings," Optimization Software, Inc., New York, 1986.
21. W. Ledermann, "Introduction to Group Characters," 2nd ed., Cambridge Univ. Press, New York, 1987.
22. B. F. Logan and L. A. Shepp, A variational problem for random Young tableaux, *Adv. in Math.* **26** (1977), 206–222.
23. B. Pittel, On growing a random Young tableau, *J. Combin. Theory Ser. A* **14** (1986), 278–285.
24. A. G. Postnikov, "Introduction to Analytic Number Theory," Translation of Mathematical Monographs, Vol. 68, Amer. Math. Soc., Providence, RI, 1988.

25. L. A. Shepp and S. P. Lloyd, Ordered cycle lengths in a random permutation, *Trans. Amer. Math. Soc.* **12** (1966), 340–357.
26. R. P. Stanley, Ordered structures and partitions, *Mem. Amer. Math. Soc.* **119** (1972).
27. M. Szalay and P. Turán, On some problems of the statistical theory of partitions with application to characters of the symmetric group, I, *Acta Math. Acad. Sci. Hungar.* **29** (1977), 361–379.
28. M. Szalay and P. Turán, On some problems of the statistical theory of partitions with application to characters of the symmetric group, II, *Acta Math. Acad. Sci. Hungar.* **29** (1977), 381–392.
29. M. Szalay and P. Turán, On some problems of the statistical theory of partitions with application to characters of the symmeric group, III, *Acta Math. Acad. Sci. Hungar.* **32** (1978), 129–155.
30. H. N. Y. Temperley, Statistical mechanics and the partition of numbers, II. The form of crystal surfaces, *Proc. Roy. Soc. A* **48** (1952), 683–697.
31. A. M. Vershik and S. V. Kerov, The asymptotic of the Plancherel measure of the symmetric group and the limit form of Young tableaux, *Dokl. Akad. Nauk SSSR* **233** (1977), 1024–1027.
32. A. M. Vershik and S. V. Kerov, Asymptotic of the largest and the typical dimensions of irreducible representations of a symmetric group, *Functional Anal. Appl.* **19** (1985), 21–31.
33. A. M. Vershik, On a statistical sum associated with the Young diagram,'' *in* ''Differential Geometry, Lie groups and Mechanics, Vol. IX,'' pp. 20–29 (*Zap. Nauchn. Sem. LOMI* **164**), Nauka, Leningrad, 1987.
34. H. S. Wilf, Three problems in combinatorial asymptotics, *J. Combin. Theory Ser. A* **35** (1983), 199–207.