Random combinatorial structures: the convergent case

A. D. Barbour* and B. Granovsky†
Universität Zürich and Technion, Haifa

April 22, 2003

Abstract

This paper studies the distribution of the component spectrum of combinatorial structures such as uniform random forests, in which the classical generating function for the numbers of (irreducible) elements of the different sizes converges at the radius of convergence; here, this property is expressed in terms of the expectations of *independent* random variables Z_j, $j \geq 1$, whose joint distribution, conditional on the event that $\sum_{j=1}^{n} jZ_j = n$, gives the distribution of the component spectrum for a random structure of size n. For a large class of such structures, we show that the component spectrum is asymptotically composed of Z_j components of size j, $j \geq 1$, with the remaining part, of size $n - \sum_{j\geq 1} Z_j$, being made up of a single, giant component.

1 Introduction

In this paper, we consider the distribution of the asymptotic component spectrum of certain decomposable random combinatorial structures. A structure of size n is composed of parts whose (integer) sizes sum to n; we let

*Angewandte Mathematik, Winterthurerstrasse 190, 8057 Zürich, Switzerland; e-mail:adb@amath.unizh.ch.

†Department of Mathematics, Technion, Haifa, Israel; e-mail: mar18aa@techunix.technion.ac.il.

$C^{(n)} := (C_1^{(n)}, C_2^{(n)}, \ldots, C_n^{(n)})$ denote its component spectrum, the numbers of components of sizes $1, 2, \ldots, n$, noting that we always have $\sum_{j=1}^n jC_j^{(n)} = n$. For each given n, we assume that the probability distribution on the space of all such component spectra satisfies the Conditioning Relation:

$$\mathcal{L}(C^{(n)}) = \mathcal{L}\left((Z_1, Z_2, \ldots, Z_n) \,\middle|\, \sum_{j=1}^n jZ_j = n \right), \tag{1.1}$$

where $Z := (Z_j,\, j \geq 1)$ is a sequence of independent random variables, the same for all n; that is, for $y_1, y_2, \ldots, y_n \in \mathbf{Z}_+$,

$$\begin{aligned} &\mathbb{P}[(C_1^{(n)}, C_2^{(n)}, \ldots, C_n^{(n)}) = (y_1, y_2, \ldots, y_n)] \\ &\quad = \left\{ \mathbb{P}\left[\sum_{j=1}^n jZ_j = n \right] \right\}^{-1} \prod_{j=1}^n \mathbb{P}[Z_j = y_j]\, \mathbf{1}_{\{n\}}\left(\sum_{j=1}^n jy_j \right). \end{aligned} \tag{1.2}$$

This apparently curious assumption is satisfied by an enormous number of classical combinatorial objects, such as, for instance, permutations of n objects under the uniform distribution, decomposed into cycles as components, when the Z_j are Poisson distributed, with $Z_j \sim \mathrm{Po}\,(1/j)$; or forests of unlabelled unrooted trees under the uniform distribution, decomposed into tree components, when the Z_j are negative binomially distributed: see Arratia, Barbour & Tavaré (2002, Chapter 2)[ABT] for many more examples. However, such structures also arise in other contexts. For instance, the state of a coagulation–fragmentation process evolving in a collection of n particles can be described by the numbers $C_j^{(n)}$ of clusters of size j, $1 \leq j \leq n$, and if such a process is reversible and Markov, then its equilibrium distribution satisfies the conditioning relation for some sequence Z of random variables. In particular, under mass action kinetics, it follows that $Z_j \sim \mathrm{Po}\,(a_j)$, where $(a_j,\, j \geq 1)$ are positive reals, determined by the coagulation and fragmentation rates; see Whittle (1965), Kelly (1979, Chapter 8), Durrett, Granovsky & Gueron (1999) and Freiman & Granovsky (2002a).

In order to simplify the initial discussion of the asymptotics, it is helpful to impose some uniformity on the distributions of the Z_j. We do this by supposing that the sequence $(\mathbb{E}Z_j,\, j \geq 1)$ is regularly varying with exponent $\alpha \in \mathbb{R}$. Note that the distribution given in (1.2) remains the same if the random variables Z_j are replaced by 'tilted' random variables $Z_j^{(x)}$, where

$$\mathbb{P}[Z_j^{(x)} = i] = \mathbb{P}[Z_j = i]x^{ji}/k_j(x), \tag{1.3}$$

for *any* $x > 0$ such that

$$k_j(x) := \mathbb{E}\left\{x^{jZ_j}\right\} < \infty,$$

so that the right choice of x is essential, if $\mathbb{E}Z_j$ is to be regularly varying. Three ranges of α can then broadly be distinguished. The most intensively studied is that when $\alpha = 1$, and within this the logarithmic class, in which $\mathbb{E}Z_j \sim \mathbb{P}[Z_j = 1] \sim \theta j^{-1}$, for some $\theta > 0$: see the book [ABT] for a detailed discussion. For $\alpha > -1$, the expansive case, the asymptotics were explored for Poisson distributed Z_j in Freiman & Granovsky (2002a,b), with the help of Khinchine's probabilistic method, and particular models have been studied by many authors. Here, we treat the convergent case, in which $\alpha < -1$, in considerable generality. Our approach is quite different from the classical approach by way of generating functions, thereby allowing distributions other than the standard Poisson and negative binomial to be easily discussed. Note also that not all classical combinatorial structures fall into one of these three categories: random set partitions, studied using the Conditioning Relation by Pittel (1997), have Poisson distributed Z_j with means $x^j/j!$, which are never regularly varying, whatever the choice of $x > 0$.

As will be seen in what follows, a key element in the arguments is establishing the asymptotics of the probabilities $\mathbb{P}[T_{bn}(Z) = l]$ for l near n, where, for $y := (y_1, y_2, \ldots) \in \mathbf{Z}_+^\infty$,

$$T_{bn}(y) := \sum_{j=b+1}^{n} jy_j, \qquad 0 \leq b < n. \tag{1.4}$$

That this should be so is clear from (1.2), in which the normalizing constant is just $\{\mathbb{P}[T_{0n}(Z) = n]\}$, and is the only element which cannot immediately be written down. In the context of reversible coagulation–fragmentation processes with mass–action kinetics, the partition function c_n investigated by Freiman & Granovsky (2002a) is given by

$$c_n := \exp\left\{\sum_{j=1}^{n} a_j\right\} \mathbb{P}[T_{0n}(Z) = n], \tag{1.5}$$

explaining its relation to many of their quantities of interest. Now, in the expansive case, taking Poisson distributed Z_j with means $a_j \sim Aj^\alpha$, $\alpha > -1$, one has

$$\mathbb{E}T_{0n}(Z) \asymp n^{2+\alpha} \gg n \quad \text{and} \quad \mathrm{SD}\left(T_{0n}(Z)\right) \asymp n^{(3+\alpha)/2} \ll \mathbb{E}T_{0n}(Z).$$

The Bernstein inequality then implies that the probability $\mathbb{P}[T_{0n}(Z) = n]$ is extremely small, making a direct asymptotic argument very delicate. However, recall from (1.3) that the conditioning relation (1.1) delivers the same distribution for the combinatorial structure if the Poisson distributed random variables Z_j with means a_j are replaced by Poisson distributed random variables $Z_j^{(x)}$ with means $a_j x^j$, for *any* $x > 0$. Choosing $x = x_n$ in such a way that $\mathbb{E}T_{0n}(Z^{(x)}) = n$ makes the probability $\mathbb{P}[T_{0n}(Z^{(x)}) = n]$ much larger, and a local limit theorem based on the normal approximation can then be used to determine its asymptotics. The resulting component spectra typically have almost all their weight in components of size about $n^{1/(\alpha+2)}$, a few smaller components making up the rest.

For the logarithmic case, taking Poisson distributed Z_j with means $a_j \sim \theta/j$, $\theta > 0$, one has

$$\mathbb{E}T_{0n}(Z) \sim n\theta \quad \text{and} \quad \mathrm{SD}\,(T_{0n}(Z)) \asymp n,$$

so that no tilting is required. However, since $T_{0n}(Z) \geq 0$, these asymptotics also imply that $\mathcal{L}(n^{-1}T_{0n}(Z))$ is not close to a normal distribution — there is a different limiting distribution that has a density related to the Dickman function from number theory — and special techniques have to be developed in order to complete the analysis. Here, the component spectra typically have components of sizes around n^β for *all* $0 \leq \beta \leq 1$.

In the convergent case, taking Poisson distributed Z_j with means $a_j \sim Aj^\alpha$, $\alpha < -1$, the sequence of random variables $T_{0n}(Z)$ converges without normalization, and both the methods of proof and the typical spectra as $n \to \infty$ are again qualitatively different. We demonstrate that, for large n, the typical picture is that of small components whose numbers have the independent joint distribution of the Z_j, the remaining weight being made up by a *single* component of size close to n. This remains true without the Poisson assumption, under fairly weak conditions; for instance, our theory applies to the example of uniform random forests, where the asymptotic distribution of the size of the large component was derived using generating function methods by Mutafchiev (1998). Bell, Bender, Cameron and Richmond (2000, Theorem 2) have also used generating function methods to examine the convergent case for labelled and unlabelled structures, which, in our setting, correspond to Poisson and negative binomially distributed Z_j's, respectively; we allow an even wider choice of distributions for the Z_j. They use somewhat different conditions, and are primarily interested in whether or not the probability that the largest component is of size n has a limit as $n \to \infty$, though they also consider the limiting distribution of the number of components. Under our conditions, these limits always exist.

2 Results

We work in a context in which the random variables Z_j may be quite general, provided that, for large j, their distributions are sufficiently close to Poisson, the detailed requirements being given below: from now on, we use the notation $a_j := \mathbb{E}Z_j$. Since, in the convergent case, $a_j \to 0$, this mainly involves assuming that $\mathbb{P}[Z_j \geq 2] \ll \mathbb{P}[Z_j = 1]$ as $j \to \infty$, so that the Z_j can be thought of as independent random variables which usually take the value 0, and occasionally (but only a.s. finitely often) the value 1. This setting is broad enough to include a number of well known examples, including uniform random forests consisting of (un)labelled (un)rooted trees. In such circumstances, we are able to use a technique based on recurrence relations which are exactly true for Poisson distributed Z_j, and which can be simply derived using Stein's method for the compound Poisson distribution (Barbour, Chen and Loh, 1992). A corresponding approach is used in [ABT], though the detail of the argument here is very different.

Stein's method for the weighted sum $T^*_{bn} := T_{bn}(Z)$, when $Z_j \sim \mathrm{Po}\,(a_j)$ and $0 \leq b < n$, is based on the identity

$$\mathbb{E}\{T^*_{bn} g(T^*_{bn})\} = \sum_{j=b+1}^{n} j a_j \mathbb{E} g(T^*_{bn} + j), \tag{2.1}$$

which is true for all bounded functions $g : \mathbf{Z}_+ \to \mathbb{R}$. Here, in the convergent case, we write $a_j = j^{-1-q}\lambda(j)$, where $q = -1 - \alpha > 0$ and λ is a positive function obeying certain conditions to be specified later; we do not now necessarily assume that λ is slowly varying at infinity, though this assumption implies that conditions (2.8)–(2.10) below are automatically satisfied, and that L_s defined in (2.7) is finite. Taking $g = \mathbf{1}_{\{l\}}$, for any $l \geq b + 1$, it thus follows that

$$l\mathbb{P}[T^*_{bn} = l] \;=\; \sum_{j=b+1}^{n} j^{-q}\lambda(j)\mathbb{P}[T^*_{bn} = l - j]$$

$$= \sum_{j=b+1}^{l\wedge n} j^{-q}\lambda(j)\mathbb{P}[T^*_{bn} = l-j], \quad l \geq b+1; \qquad (2.2)$$

note that this recursion can also be deduced directly by differentiating the compound Poisson generating function, and equating coefficients. Recursion (2.2), coupled with the fact that $\mathbb{P}[T^*_{bn} = l] = 0$ for $1 \leq l \leq b$, successively expresses the probabilities $\mathbb{P}[T^*_{bn} = l]$ in terms of the probability $\mathbb{P}[T^*_{bn} = 0]$. In particular, if $l \leq n$ is large and if $\{j^{-q}\lambda(j)\}/\{l^{-q}\lambda(l)\}$ is close to 1 when j is close to l, it suggests that

$$l\mathbb{P}[T^*_{bn} = l] \approx l^{-q}\lambda(l)\mathbb{P}[T^*_{bn} < l-b-1] \approx l^{-q}\lambda(l),$$

giving the large l asymptotics for $\mathbb{P}[T^*_{bn} = l]$. Our approach consists of turning this heuristic into a precise argument, which can be applied also when the Z_j do not have Poisson distributions.

The Stein identity (2.1) is deduced from the Poisson Stein–Chen identity, which yields

$$\mathbb{E}\{Z_j g(T^*_{bn})\} = j^{-1-q}\lambda(j)\mathbb{E}\{g(T^*_{bn}+j)\}, \quad b+1 \leq j \leq n, \qquad (2.3)$$

when $Z_j \sim \mathrm{Po}\,(j^{-1-q}\lambda(j))$. Our first requirement is therefore to establish an analogue of (2.3) for more general random variables Z_j. In doing so, we shall suppose that each Z_j can be written in the form $Z_j = \sum_{k=1}^{r_j} Z_{jk}$ for some $r_j \geq 1$, where, for each j, the non-negative integer valued random variables $(Z_{jk},\, 1 \leq k \leq r_j)$ are independent and identically distributed. Clearly, this is always possible if we take $r_j = 1$. The reason for using such a representation is that the errors in our approximation to (2.2) can be shown to be smaller, if we take advantage of any divisibility in the distributions of the Z_j; in the Poisson and negative binomial cases, the Z_j are actually infinitely divisible, so that the r_j can be taken to be arbitrarily large. We now define $(\varepsilon_{js},\, s, j \geq 1)$ by setting

$$\begin{aligned} r_j\mathbb{P}[Z_{j1} = 1] &=: j^{-q-1}\lambda(j)(1-\varepsilon_{j1}); \\ r_j\mathbb{P}[Z_{j1} = s] &=: j^{-q-1}\lambda(j)\varepsilon_{js}, \quad s \geq 2, \end{aligned} \qquad (2.4)$$

so that then

$$0 \leq \varepsilon_{j1} = \sum_{s\geq 2} s\varepsilon_{js} \leq 1,$$

because $j^{-q-1}\lambda(j) = a_j = \mathbb{E}Z_j = r_j\mathbb{E}Z_{j1}$. We then assume that

$$0 \le \varepsilon_{js} \le \varepsilon(j)\gamma_s, \quad s \ge 2, \tag{2.5}$$

where

$$G := \sum_{s\ge2} s\gamma_s < \infty \quad \text{and} \quad \lim_{j\to\infty} \varepsilon(j) = 0; \tag{2.6}$$

we write $\varepsilon^*(j) := \max_{l\ge j+1} \varepsilon(l)$ and $r^*(j) := \min_{l>j} r_l$. For the subsequent argument, we need to strengthen (2.6) by assuming in addition that

$$G_q := \sum_{s\ge2} L_s s^{1+q}\gamma_s < \infty, \quad \text{where} \quad L_s := \sup_{l\ge s}\{\lambda(\lfloor l/s\rfloor)/\lambda(l)\}. \tag{2.7}$$

We also assume that

$$\lambda^+(l) := \max_{1\le s\le l} \lambda(s) = o(l^\beta) \quad \text{for any } \beta > 0; \tag{2.8}$$

$$L := \sup_{l\ge2} \max_{l/2<t\le l} \{\lambda(l-t)/\lambda(l)\} < \infty, \tag{2.9}$$

and that

$$\lim_{l\to\infty}\{\lambda(l-s)/\lambda(l)\} = 1 \quad \text{for all } s \ge 1; \tag{2.10}$$

we write $\Lambda_\beta := \max_{l\ge1} l^{-\beta}\lambda(l)$ for $\beta > 0$, and observe that

$$\mathbb{P}[Z_{jk} \ge 1 \text{ for } \infty \text{ many } j,k] = 0 \quad \text{and hence} \quad T_{0\infty}(Z) < \infty \text{ a.s.}, \tag{2.11}$$

from (2.4), (2.8) and the Borel–Cantelli lemma.

Now, writing $T_{bn} := T_{bn}(Z)$, and recalling that $Z_j = \sum_{k=1}^{r_j} Z_{jk}$, it is immediate that

$$\mathbb{E}\{Z_{j1}g(T_{bn})\} = \sum_{s\ge1} s\mathbb{P}[Z_{j1} = s]\mathbb{E}g(T_{bn}^{(j)} + js),$$

where $T_{bn}^{(j)} := T_{bn} - jZ_{j1}$, so that, with the above definitions,

$$\begin{aligned} \mathbb{E}\{T_{bn}g(T_{bn})\} &= \sum_{j=b+1}^{n} \mathbb{E}\{jZ_jg(T_{bn})\} \\ &= \sum_{j=b+1}^{n} jr_j \sum_{s\ge1} s\mathbb{P}[Z_{j1} = s]\mathbb{E}g(T_{bn}^{(j)} + js) \end{aligned} \tag{2.12}$$

$$
\begin{aligned}
= & \sum_{j=b+1}^{n} j^{-q}\lambda(j)\mathbb{E}g(T_{bn}+j) \\
& + \sum_{j=b+1}^{n} j^{-q}\lambda(j)\{(1-\varepsilon_{j1})\mathbb{E}g(T_{bn}^{(j)}+j) - \mathbb{E}g(T_{bn}+j)\} \\
& + \sum_{j=b+1}^{n}\sum_{s\geq 2} j^{-q}\lambda(j)s\varepsilon_{js}\mathbb{E}g(T_{bn}^{(j)}+js). \qquad (2.13)
\end{aligned}
$$

Taking $g = \mathbf{1}_{\{l\}}$ as before then gives the recursion

$$
\begin{aligned}
l\mathbb{P}[T_{bn}=l] = & \sum_{j=b+1}^{l\wedge n} j^{-q}\lambda(j)\mathbb{P}[T_{bn}=l-j] \\
& + \sum_{j=b+1}^{l\wedge n} j^{-q}\lambda(j)\{(1-\varepsilon_{j1})\mathbb{P}[T_{bn}^{(j)}=l-j] - \mathbb{P}[T_{bn}=l-j]\} \\
& + \sum_{j=b+1}^{\lfloor (l/2)\wedge n\rfloor}\sum_{s\geq 2} j^{-q}\lambda(j)s\varepsilon_{js}\mathbb{P}[T_{bn}^{(j)}=l-js], \qquad (2.14)
\end{aligned}
$$

which can be understood as a perturbed form of the recursion (2.2).

In order to show that the perturbation is indeed small, it is first necessary to derive bounds for the probabilities $\mathbb{P}[T_{bn}=s]$ and $\mathbb{P}[T_{bn}^{(j)}=s]$. However, since $\mathbb{P}[T_{bn}=s] \geq \mathbb{P}[Z_{j1}=0]\,\mathbb{P}[T_{bn}^{(j)}=s]$, we have the immediate bound

$$
\mathbb{P}[T_{bn}^{(j)}=s] \leq p_0^{-1}\mathbb{P}[T_{bn}=s], \quad s=0,1,\ldots, \qquad (2.15)
$$

where we assume that the distributions of the random variables Z_{j1} of (2.4) are such that

$$
p_0 := \min_{j\geq 1}\mathbb{P}[Z_{j1}=0] > 0. \qquad (2.16)
$$

Hence the following lemma is all that is required.

Lemma 2.1 *Suppose that conditions (2.5) – (2.10) and (2.16) are satisfied for some $q>0$. Then there exists a constant $K>0$, depending only on the distributions of the Z_j, such that*

$$
\mathbb{P}[T_{bn}=l] \leq K\lambda(l)l^{-1-q}, \qquad l\geq 1.
$$

Proof. For $1 \le l \le b$, the statement is trivial. For larger l, we proceed by induction, using the recursion (2.14), in which, on the right hand side, probabilities of the form $\mathbb{P}[T_{bn} = s]$ appear only for $s < l$, so that we may suppose that then $\mathbb{P}[T_{bn} = s] \le K\lambda(s)s^{-1-q}$ for all $1 \le s < l$. Under this hypothesis, we split the right hand side of (2.14) into three terms, which we bound separately; we take the first two lines together, and then split the third according to the value taken by js.

For the first term, we use (2.15), the induction hypothesis and conditions (2.8) and (2.9) to give

$$\begin{aligned}
&\sum_{j=b+1}^{l\wedge n} j^{-q}\lambda(j)(1-\varepsilon_{j1})\mathbb{P}[T_{bn}^{(j)} = l-j] \\
&\le \sum_{j=1}^{\lfloor l/2\rfloor} j^{-q}\lambda(j)p_0^{-1}\mathbb{P}[T_{bn} = l-j] + \sum_{j=\lfloor l/2\rfloor+1}^{l} j^{-q}\lambda(j)p_0^{-1}\mathbb{P}[T_{bn} = l-j] \\
&\le p_0^{-1}\lambda^+(\lfloor l/2\rfloor)KL\lambda(l)(2/l)^{1+q}\sum_{j=1}^{\lfloor l/2\rfloor} j^{-q} + p_0^{-1}L\lambda(l)(2/l)^q \\
&= K\lambda(l)l^{-q}\eta_0(l) + p_0^{-1}L\lambda(l)(2/l)^q, \qquad (2.17)
\end{aligned}$$

where

$$\eta_0(l) := p_0^{-1}2^{1+q}\lambda^+(\lfloor l/2\rfloor)Ll^{-1}\sum_{j=1}^{\lfloor l/2\rfloor} j^{-q} = o(1) \quad \text{as } l \to \infty.$$

For the second term, arguing much as before, we have

$$\begin{aligned}
&\sum_{j=b+1}^{\lfloor (l/2)\wedge n\rfloor}\sum_{s\ge 2}\mathbf{1}_{\{js\le\lfloor l/2\rfloor\}}j^{-q}\lambda(j)s\varepsilon_{js}\mathbb{P}[T_{bn}^{(j)} = l-js] \\
&\le \sum_{j=1}^{\lfloor l/2\rfloor}\sum_{s\ge 2}\mathbf{1}_{\{js\le\lfloor l/2\rfloor\}}j^{-q}\lambda(j)s\varepsilon_{js}p_0^{-1}KL\lambda(l)(2/l)^{1+q} \\
&\le \lambda(l)l^{-q}p_0^{-1}2^{1+q}\lambda^+(\lfloor l/2\rfloor)KLl^{-1}\sum_{j=1}^{\lfloor l/2\rfloor} j^{-q}\varepsilon(j)G \\
&\le \varepsilon^*(0)GK\lambda(l)l^{-q}\eta_0(l). \qquad (2.18)
\end{aligned}$$

For the third and final term, we have

$$\begin{aligned}
&\sum_{j=b+1}^{\lfloor (l/2)\wedge n\rfloor} \sum_{s\geq 2} \mathbf{1}_{\{\lfloor l/2\rfloor < js \leq l\}} j^{-q}\lambda(j)s\varepsilon_{js}\mathbb{P}[T_{bn}^{(j)} = l - js] \\
&\leq \sum_{s=2}^{l} \sum_{j=\lfloor l/2s\rfloor+1}^{\lfloor l/s\rfloor -1} j^{-q}\lambda(j)s\varepsilon_{js}\mathbb{P}[T_{bn}^{(j)} = l-js] + \sum_{s=2}^{l} \lfloor l/s\rfloor^{-q}\lambda(\lfloor l/s\rfloor)s\varepsilon_{\lfloor l/s\rfloor,s} \\
&= S_1 + S_2, \qquad (2.19)
\end{aligned}$$

say. Now

$$\begin{aligned}
S_1 &\leq \sum_{s=2}^{l} \sum_{j=\lfloor l/2s\rfloor+1}^{\lfloor l/s\rfloor-1} j^{-q}\lambda(j)s\varepsilon(j)\gamma_s p_0^{-1}K\lambda(l-js)(l-js)^{-1-q} \\
&\leq p_0^{-1}K\sum_{s=2}^{l}(l/2s)^{-q}L_sL\lambda(l)s\gamma_s\varepsilon^*(\lfloor l/2s\rfloor)R_q s^{-1-q/2}, \qquad (2.20)
\end{aligned}$$

where $R_q := \Lambda_{q/2}\sum_{t\geq 1} t^{-1-q/2}$, and this implies that

$$S_1 \leq K\lambda(l)l^{-q}\eta_1(l), \qquad (2.21)$$

where

$$\begin{aligned}
\eta_1(l) &:= p_0^{-1}LR_q2^q \min_{2\leq t\leq l}\left\{\varepsilon^*(\lfloor l/2t\rfloor)\sum_{s=2}^{t} s^{q/2}L_s\gamma_s + \varepsilon^*(0)\sum_{s\geq t+1} s^{q/2}L_s\gamma_s\right\} \\
&= o(1) \quad \text{as } l\to\infty,
\end{aligned}$$

in view of (2.6) and (2.7). For S_2, we have

$$\begin{aligned}
S_2 &\leq \sum_{s=2}^{l}\lfloor l/s\rfloor^{-q}\lambda(\lfloor l/s\rfloor)\varepsilon(\lfloor l/s\rfloor)s\gamma_s \\
&\leq \lambda(l)l^{-q}\sum_{s\geq 2} s^{1+q}L_s\gamma_s\varepsilon(\lfloor l/s\rfloor) \\
&:= \lambda(l)l^{-q}\eta_2(l), \qquad (2.22)
\end{aligned}$$

where $\eta_2(l) = o(1)$ as $l\to\infty$, again in view of (2.6) and (2.7).

Collecting these bounds, we can apply (2.14) to show that

$$l\mathbb{P}[T_{bn} = l] \leq \lambda(l)l^{-q}\{2^qLp_0^{-1} + \eta_2(l) + K[\eta_0(l)(1+\varepsilon^*(0)G) + \eta_1(l)]\}; \quad (2.23)$$

and this in turn is less than $K\lambda(l)l^{-q}$ provided that

$$K\{1 - [\eta_0(l)(1 + \varepsilon^*(0)G) + \eta_1(l)]\} > 2^q L p_0^{-1} + \eta_2(l),$$

which can be achieved uniformly for all $l \geq l_0$, for some large l_0, by choosing $K \geq 2^{q+1}Lp_0^{-1}$. As observed before, $\mathbb{P}[T_{bn} = l] = 0$ for $1 \leq l \leq b$. For $b + 1 \leq l \leq l_0$, we can suppose that $\mathbb{P}[T_{bn} = t] \leq K_{l-1}\lambda(t)t^{-1-q}$ for all $t \leq l-1$, and deduce from (2.23) that $\mathbb{P}[T_{bn} = t] \leq K_l\lambda(t)t^{-1-q}$ for all $t \leq l$, if we take

$$K_l = \max\{K_{l-1}, 2^q L p_0^{-1} + \eta_2(l) + K_{l-1}[\eta_0(l)(1 + \varepsilon^*(0)G) + \eta_1(l)]\};$$

this then completes the proof. $\diamond$

Equipped with the above lemma, we can now examine the recursion (2.14) in more detail, using the bounds derived in the course of proving Lemma 2.1, determining the asymptotics as $l, n \to \infty$ of the probabilities $\mathbb{P}[T_{bn} = l]$.

Theorem 2.2 *Suppose that conditions (2.5) – (2.10) and (2.16) are satisfied for some $q > 0$. Then, if $1 \leq l \leq n$, we have*

$$|\lambda^{-1}(l)l^{1+q}\mathbb{P}[T_{bn} = l] - 1| \leq H(l),$$

uniformly in $0 \leq b \leq l-1$, where $\lim_{l\to\infty} H(l) = 0$.

Proof. We use the recursion (2.14), observing that the contribution from the last line was bounded in the proof of Lemma 2.1 by

$$\lambda(l)l^{-q}\{\varepsilon^*(0)GK\eta_0(l) + K\eta_1(l) + \eta_2(l)\}, \tag{2.24}$$

uniformly in $0 \leq b \leq l-1$. We now need to examine the second line in more detail. First, note that, by Lemma 2.1, for $l \leq n$,

$$\begin{aligned}
&\sum_{j=b+1}^{l\wedge n} j^{-q}\lambda(j)\varepsilon_{j1}\mathbb{P}[T_{bn}^{(j)} = l - j] \\
&\leq \sum_{j=1}^{\lfloor l/2 \rfloor} j^{-q}\lambda(j)G\varepsilon(j)p_0^{-1}(2/l)^{1+q}KL\lambda(l) + p_0^{-1}G\varepsilon^*(\lfloor l/2 \rfloor)L\lambda(l)(2/l)^q \\
&\leq \lambda(l)l^{-q}(KG\varepsilon^*(0)\eta_0(l) + \eta_0'(l)),
\end{aligned} \tag{2.25}$$

where

$$\eta_0'(l) := 2^q p_0^{-1} G L \varepsilon^*(\lfloor l/2 \rfloor) = o(1) \quad \text{as } l \to \infty.$$

The remaining part of the second line of (2.14) is then bounded by

$$\begin{aligned}
&\left| \sum_{j=b+1}^{l \wedge n} j^{-q} \lambda(j) \{ \mathbb{P}[T_{bn}^{(j)} = l-j] - \mathbb{P}[T_{bn} = l-j] \} \right| \\
&\quad = \left| \sum_{j=b+1}^{l} j^{-q} \lambda(j) \{ \mathbb{P}[T_{bn}^{(j)} = l-j] - \sum_{s \geq 0} \mathbb{P}[Z_{j1} = s] \mathbb{P}[T_{bn}^{(j)} = l - j(s+1)] \} \right| \\
&\quad \leq \sum_{j=1}^{l} j^{-q} \lambda(j) \{ \mathbb{P}[Z_{j1} \geq 1] \mathbb{P}[T_{bn}^{(j)} = l-j] \\
&\qquad\qquad + \sum_{s \geq 1} \mathbb{P}[Z_{j1} = s] \mathbb{P}[T_{bn}^{(j)} = l - j(s+1)] \}. \qquad (2.26)
\end{aligned}$$

We now observe, using Lemma 2.1, (2.15), (2.4) and (2.9), that

$$\begin{aligned}
&\sum_{j=1}^{l} j^{-q} \lambda(j) \mathbb{P}[Z_{j1} \geq 1] \mathbb{P}[T_{bn}^{(j)} = l-j] \\
&\quad \leq \sum_{j=1}^{\lfloor l/2 \rfloor} r_j^{-1} j^{-1-2q} \lambda^2(j) p_0^{-1} K L \lambda(l) (2/l)^{1+q} \\
&\qquad\qquad + \{ r^*(l/2) \}^{-1} p_0^{-1} \{ L \lambda(l) \}^2 (2/l)^{1+2q} \\
&\quad := \lambda(l) l^{-q} \eta_3(l), \qquad (2.27)
\end{aligned}$$

where clearly $\eta_3(l) = o(1)$ as $l \to \infty$. Then we also have

$$\begin{aligned}
&\sum_{j=1}^{l} j^{-q} \lambda(j) \mathbb{P}[Z_{j1} = 1] \mathbb{P}[T_{bn}^{(j)} = l-2j] \\
&\quad \leq \sum_{j=1}^{\lfloor l/4 \rfloor} r_j^{-1} j^{-1-2q} \lambda^2(j) p_0^{-1} K L \lambda(l) (2/l)^{1+q} \\
&\qquad\qquad + \{ r^*(l/4) \}^{-1} p_0^{-1} \{ L^2 \lambda(l) \}^2 (4/l)^{1+2q} \\
&\quad := \lambda(l) l^{-q} \eta_4(l), \qquad (2.28)
\end{aligned}$$

again by Lemma 2.1, where also $\eta_4(l) = o(1)$ as $l \to \infty$. The remaining piece of the last term in (2.26) is split into two, as in the proof of the previous lemma, though the argument is a little simpler. The bound

$$\begin{aligned}
&\sum_{j=1}^{l} j^{-q}\lambda(j)\sum_{s\geq 2}\mathbf{1}_{\{j(s+1)\leq\lfloor l/2\rfloor\}}\mathbb{P}[Z_{j1}=s]\mathbb{P}[T_{bn}^{(j)}=l-j(s+1)]\\
&\quad\leq\ p_0^{-1}KL\lambda(l)(2/l)^{1+q}\sum_{j=1}^{l} r_j^{-1}j^{-1-2q}\lambda^2(j)G\varepsilon(j)\\
&\quad:=\ \lambda(l)l^{-q}\eta_5(l), \qquad (2.29)
\end{aligned}$$

with $\eta_5(l)=o(1)$ as $l\to\infty$, follows immediately. For the second part, we have

$$\begin{aligned}
&\sum_{j=1}^{l} j^{-q}\lambda(j)\sum_{s\geq 2}\mathbf{1}_{\{\lfloor l/2\rfloor<j(s+1)\leq l\}}\mathbb{P}[Z_{j1}=s]\mathbb{P}[T_{bn}^{(j)}=l-j(s+1)]\\
&\leq\ p_0^{-1}\sum_{s=2}^{l-1}\sum_{j=\lfloor l/2(s+1)\rfloor+1}^{\lfloor l/(s+1)\rfloor} r_j^{-1}\lambda^2(j)j^{-1-2q}\varepsilon(j)\gamma_s\mathbb{P}[T_{bn}=l-j(s+1)]\\
&\leq\ p_0^{-1}\sum_{s=2}^{l-1}\{r^*(l/2(s+1))\}^{-1}\varepsilon^*(\lfloor l/2(s+1)\rfloor)L^2L_s\lambda(l)\Lambda_{q/2}\{2(s+1)/l\}^{1+3q/2}\gamma_s\\
&\leq\ \{r^*(0)\}^{-1}\varepsilon^*(0)p_0^{-1}3^{1+3q/2}\sum_{s=2}^{l-1}L^2L_s\lambda(l)\Lambda_{q/2}(s/l)^{1+q}\gamma_s\\
&\leq\ \lambda(l)l^{-q}\eta_6(l), \qquad (2.30)
\end{aligned}$$

with

$$\eta_6(l):=\{r^*(0)\}^{-1}\varepsilon^*(0)p_0^{-1}3^{1+3q/2}L^2\Lambda_{q/2}G_ql^{-1}=o(1)\quad\text{as } l\to\infty.$$

Combining the results from (2.24) – (2.30), it follows from (2.14) that, for $l\leq n$,

$$l\mathbb{P}[T_{bn}=l]=\sum_{j=b+1}^{l} j^{-q}\lambda(j)\mathbb{P}[T_{bn}=l-j]+\lambda(l)l^{-q}\eta_7(l),$$

where $\eta_7(l)=o(1)$ as $l\to\infty$. Hence we deduce that

$$\begin{aligned}
&\lambda^{-1}(l)l^{1+q}\mathbb{P}[T_{bn}=l] \qquad (2.31)\\
&\quad=\ \mathbb{P}[T_{bn}\leq l-b-1]+\sum_{s=0}^{l-b-1}\left\{\frac{l^q\lambda(l-s)}{(l-s)^q\lambda(l)}-1\right\}\mathbb{P}[T_{bn}=s]+\eta_7(l).
\end{aligned}$$

In view of (2.10), we can find a sequence $s_l \to \infty$ such that $s_l = o(l)$ and

$$\max_{1\le s\le s_l} \left| \frac{\lambda(l-s)}{\lambda(l)} - 1 \right| = o(1) \quad \text{as } l \to \infty :$$

hence also

$$\sum_{s=0}^{s_l} \left| \frac{l^q \lambda(l-s)}{(l-s)^q \lambda(l)} - 1 \right| \mathbb{P}[T_{bn} = s] = \eta_8(l) = o(1) \quad \text{as } l \to \infty.$$

It then follows from (2.9) and (2.11) that

$$\begin{aligned} \sum_{s=s_l+1}^{\lfloor l/2 \rfloor} \left| \frac{l^q \lambda(l-s)}{(l-s)^q \lambda(l)} - 1 \right| \mathbb{P}[T_{bn} = s] &\le (2^q L + 1)\mathbb{P}[T_{bn} > s_l] \\ &\le (2^q L + 1)\mathbb{P}[T_{0\infty} > s_l] = \eta_9(l) = o(1) \quad \text{as } l \to \infty. \end{aligned}$$

For the remaining sum, we use Lemma 2.1 to give

$$\begin{aligned} &\sum_{s=\lfloor l/2 \rfloor+1}^{l-b-1} \left| \frac{l^q \lambda(l-s)}{(l-s)^q \lambda(l)} - 1 \right| \mathbb{P}[T_{bn} = s] \\ &\le KL\lambda(l)(2/l)^{1+q} \left\{ \frac{l}{2} + \sum_{s=1}^{\lfloor l/2 \rfloor} \frac{l^q \lambda(s)}{\lambda(l) s^q} \right\} \\ &\le KL2^q \Lambda_{q/2} \left\{ l^{-q/2} + (2/l) \sum_{s=1}^{\lfloor l/2 \rfloor} s^{-q/2} \right\} \\ &= \eta_{10}(l) = o(1) \quad \text{as } l \to \infty. \end{aligned} \tag{2.32}$$

Putting these estimates into (2.31), it follows that, for $1 \le l \le n$,

$$\lambda^{-1}(l) l^{1+q} \mathbb{P}[T_{bn} = l] = 1 - \mathbb{P}[T_{bn} > l - b - 1] + \eta_{11}(l), \tag{2.33}$$

where $\eta_{11}(l) = o(1)$ as $l \to \infty$. Finally, since also, for $b \le \lfloor l/2 \rfloor$,

$$\mathbb{P}[T_{bn} > l - b - 1] \le \mathbb{P}[T_{0\infty} > l/2] \to 0 \quad \text{as } l \to \infty,$$

whereas, for $\lfloor l/2 \rfloor < b < l$,

$$\begin{aligned} \mathbb{P}[T_{bn} > l - b - 1] &\le \mathbb{P}[T_{b\infty} > 0] \le \mathbb{P}[T_{\lfloor l/2 \rfloor,\infty} > 0] \\ &\le \sum_{j=\lfloor l/2 \rfloor}^{\infty} \lambda(j) j^{-1-q} \to 0 \quad \text{as } l \to \infty, \end{aligned} \tag{2.34}$$

it follows from (2.33) that

$$|\lambda^{-1}(l)l^{1+q}\mathbb{P}[T_{bn}=l]-1| \le H(l), \qquad 0 \le b \le l-1,$$

where $\lim_{l\to\infty} H(l)=0$, as required. $\diamond$

Remark. The assumption (2.16), that $p_0>0$, can be dispensed with, whatever the distributions of the Z_j, provided that (2.6) holds. Clearly, for some $m \ge 1$ and $t_1,\ldots,t_m$, we have

$$p_0' := \min\left\{\min_{j\ge m+1}\mathbb{P}[Z_{j1}=0],\ \min_{1\le j\le m}\mathbb{P}[Z_{j1}=t_j]\right\} > 0,$$

since $\lim_{j\to\infty}\mathbb{E}Z_j = 0$. Then, for $j \le m$ and $s > t_j$, we have the simple bound

$$\mathbb{P}[T_{bn}^{(j)}=l-js] \le \mathbb{P}[T_{bn}=l-j(s-t_j)]/\mathbb{P}[Z_{j1}=t_j],$$

which can be used as before, together with the induction hypothesis, to bound the right hand side of (2.14) in the proof of Lemma 2.1, provided that $s>t_j$. So, recalling (2.12) with $g=\mathbf{1}_{\{l\}}$, we write

$$\begin{aligned}
&\sum_{j=1}^m \mathbb{E}\{jZ_j\mathbf{1}_{\{l\}}(T_{bn})\}\\
&\quad= \sum_{j=1}^m jr_j\mathbb{E}\{Z_{j1}I[Z_j\le t_j]\mathbf{1}_{\{l\}}(T_{bn})\}\\
&\qquad\qquad + \sum_{j=1}^m jr_j\mathbb{E}\{Z_{j1}I[Z_j> t_j]\mathbf{1}_{\{l\}}(T_{bn})\}.
\end{aligned}$$

The second term is estimated exactly as before. The first is no larger than $\kappa\mathbb{P}[T_{bn}=l]$, where

$$\kappa := \sum_{j=1}^m jr_jt_j,$$

and hence can be taken onto the left hand side of (2.23) whenever $l \ge 2\kappa$; with these modifications, the proof of Lemma 2.1 can be carried through as before. The proof of Theorem 2.2 requires almost no modification, if p_0 is replaced by p_0'.

Remark. If the $(Z_j,\, j \geq 1)$ are infinitely divisible, then we can choose the r_j to be arbitrarily large for each fixed j, in the limit making $\eta_k(l) = 0$, $3 \leq k \leq 6$, and $p_0 = 1$. The limiting values as $r_j \to \infty$ of ε_{js}, for fixed j and $s \geq 1$, are *not* however in general zero.

Remark. The assumption (2.7) that G_q be finite is not just an artefact of the proofs. It appears in particular when bounding the quantity S_2 in (2.22) in the proof of Lemma 2.1, and is an element in the quantity $\eta_2(l)$, which contributes to $H(l)$ in Theorem 2.2. However, $l^{-1}S_2$ is of the same order as the probability that T_{0n} is composed of s components of equal sizes $\lfloor l/s \rfloor$, plus a small remainder, for some $s \geq 2$, and $G_q < \infty$ is the condition which ensures that this probability is of smaller order than $\lambda(l)l^{-1-q}$.

As is strongly suggested by the formula (1.2), Theorem 2.2, in giving the asymptotics of $\mathbb{P}[T_{0n}(Z) = n]$, is directly useful for establishing the asymptotic joint distribution of the entire component spectrum. This is given in the following theorem.

Theorem 2.3 *Suppose that conditions (2.5) – (2.10) and (2.16) are satisfied for some $q > 0$. Then*

$$\lim_{n\to\infty} d_{TV}(\mathcal{L}(C^{(n)}), Q_n) \to 0,$$

where Q_n is the distribution of $(Z_1, Z_2, \ldots, Z_n) + e(n - T_{0n}(Z))$, and $e(j)$ denotes the j'th unit n-vector if $j \geq 1$, and the zero n-vector otherwise.

Proof. As in [ABT, Lemma 3.1], it follows from the Conditioning Relation that, for any $b \leq n$,

$$\begin{aligned} d_{TV}(\mathcal{L}(C_1^{(n)}, \ldots, C_b^{(n)}), \mathcal{L}(Z_1, \ldots, Z_b)) \\ = \sum_{j\geq 0} \mathbb{P}[T_{0b} = j] \left\{ 1 - \frac{\mathbb{P}[T_{bn} = n-j]}{\mathbb{P}[T_{0n} = n]} \right\}_+ . \end{aligned} \tag{2.35}$$

Pick $b = b(n)$ with $n - b(n) \to \infty$, and observe that the right hand side of (2.35) is at most

$$\mathbb{P}[T_{0b} > j_n] + \mathbb{E}g_n(T_{0b}),$$

where $g_n(j) = 0$ for $j > j_n$ and where, for all n such that $H(n) < 1/2$,

$$0 \leq g_n(j) \leq \left| \frac{n^{1+q}\lambda(n-j)}{(n-j)^{1+q}\lambda(n)} - 1 \right| + 2^{1+q} L\, 2(H(n) + H(n-j)), \quad 0 \leq j \leq j_n,$$

from Theorem 2.2, provided that $0 \leq j_n \leq \lfloor n/2 \rfloor$ and that $j_n \leq n - b(n) - 1$. This implies in particular that $g_n(j)$ is uniformly bounded for sequences j_n satisfying these conditions. Now, from (2.10) and Theorem 2.2, it follows that $\lim_{n\to\infty} g_n(j) = 0$ for each fixed j. Since also $T_{0b} \leq T_{0\infty}$ a.s. and $T_{0\infty}$ is a.s. finite, it follows by dominated convergence that $\lim_{n\to\infty} \mathbb{E} g_n(T_{0b(n)}) = 0$, provided that $j_n \leq \min\{n - b(n) - 1, \lfloor n/2 \rfloor\}$. On the other hand,

$$\mathbb{P}[T_{0b(n)} > j_n] \leq \mathbb{P}[T_{0\infty} > j_n] \to 0,$$

so long as $j_n \to \infty$. Thus, taking for example $b(n) = \lfloor 3n/4 \rfloor$ and $j_n = \lfloor n/4 \rfloor - 1$, it follows that

$$d_{TV}(\mathcal{L}(C_1^{(n)}, \ldots, C_{b(n)}^{(n)}), \mathcal{L}(Z_1, \ldots, Z_{b(n)})) \to 0$$

as $n \to \infty$. On the other hand, we have $\sum_{j=\lfloor 3n/4 \rfloor+1}^{n} C_j^{(n)} \leq 1$ a.s., because $T_{0n}(C^{(n)}) = n$ a.s., by the definition of $C^{(n)}$. Hence, with $b(n)$ as above, we have $C_j^{(n)} = 0$ a.s. for all $j > b(n)$ if $T_{0b(n)}(C^{(n)}) = n$, while if $T_{0b(n)}(C^{(n)}) = t$ for some $t < n - b(n)$, then $C_{n-t}^{(n)} = 1$ and $C_j^{(n)} = 0$ for all other $j > b(n)$. This proves the theorem. $\diamond$

Theorem 2.3 has a number of immediate consequences.

Corollary 2.4
(a) For any fixed $k \geq 1$,

$$\mathcal{L}(C_1^{(n)}, \ldots, C_k^{(n)}) \to \mathcal{L}(Z_1, \ldots, Z_k) \quad \text{as} \quad n \to \infty.$$

(b) If $Y_n := \max\{j : C_j^{(n)} > 0\}$ and $K_n := \min\{j : C_j^{(n)} > 0\}$ are the sizes of the maximal and minimal components of the spectrum, then, as $n \to \infty$,

$$\mathcal{L}(n - Y_n) \to \mathcal{L}(T_{0\infty}(Z))$$

and

$$\mathbb{P}[K_n > b] \to \prod_{j=1}^{b} \mathbb{P}[Z_j = 0]$$

for any $b > 1$. In particular, it follows that

$$\lim_{n\to\infty} \mathbb{P}[Y_n = K_n = n] = \prod_{j\geq 1} \mathbb{P}[Z_j = 0]. \tag{2.36}$$

(c) The asymptotic distribution of the number of components X_n is given by

$$\mathcal{L}(X_n) \to \mathcal{L}\left(1 + \sum_{j\geq 1} Z_j\right).$$

Proof. All the assertions follow from Theorem 2.3, because $T_{0\infty}(Z) < \infty$ a.s. $\diamond$

Remark. The assertion (a) of the above corollary states the asymptotic independence of the numbers of components of small sizes, a fact that has also been established in [ABT] in the logarithmic case, and also in the Poisson setting for $q > 0$ in Freiman and Granovsky (2002b). This fact can be viewed as a particular manifestation of the heuristic general principle of asymptotic independence of particles in models of statistical physics.

Assertion (b) says that, as $n \to \infty$, the structures considered exhibit the gelation phenomenon; the formation, with positive probability, of a component with size comparable to n (see, for example, Whittle (1986), Ch. 13). Gelation also occurs in the logarithmic case [ABT], while it is not seen for $q > 0$ in the setting of Freiman and Granovsky (2002b). In this sense, $q = 0$ represents a critical value.

Now $\mathbb{P}[Y_n = n]$ is the probability that a structure is 'connected', as, for instance, in Bell, Bender, Cameron and Richmond (2000), who give a very general discussion of circumstances in which $\rho := \lim_{n\to\infty} \mathbb{P}[Y_n = n]$ exists, as well as giving a formula for the asymptotic distribution of X_n. They work in the settings of either labelled or unlabelled structures; in our terms, they assume that the Z_j have either Poisson or negative binomial distributions, respectively. Theorem 2.3 implies that ρ always exists under our conditions, and gives its value.

Example. Consider the uniform distribution over all forests of unlabelled, unrooted trees. The number m_j of such trees of size j was studied by Otter (1948), who showed that $m_j \sim c\rho^{-j} j^{-5/2}$, where $\rho < 1$, and gave values for both ρ and c. This combinatorial structure satisfies the conditioning relation with negative binomial random variables $Z_j \sim \mathrm{NB}\,(m_j, \rho^j)$, so that

$$\mathbb{P}[Z_j = s] = (1-\rho^j)^{m_j} \binom{m_j + s - 1}{s} \rho^{js}, \quad s \geq 0.$$

It thus follows that $\mathbb{E}Z_j = m_j\rho^j/(1-\rho^j) \sim cj^{-5/2}$, implying that our results can be applied with $\lambda(j) \to c$ and $q = 3/2$. Note that, if we take $r_j = 1$ for all j, we have

$$\mathbb{P}[Z_j = 2] = (1-\rho^j)^{m_j} \binom{m_j + 1}{2} \rho^{2j} \asymp (m_j\rho^j)^2,$$

so that $\varepsilon_{j2} \asymp j^{-5/2}$ as $j \to \infty$. On the other hand, negative binomial distributions are infinitely divisible, and other choices of r_j in (2.4) are possible: for each j, we can take $Z_{jk} \sim \mathrm{NB}\,(m_j/r_j, \rho^j)$, $1 \leq k \leq r_j$, for any choice of r_j. The corresponding values of ε_{js}, $s \geq 2$, are then given, using (2.4), by

$$\begin{aligned}
r_j\mathbb{P}[Z_{j1} = s] &= r_j(1-\rho^j)^{\frac{m_j}{r_j}} \binom{\frac{m_j}{r_j} + s - 1}{s} \rho^{sj} \\
&= r_j(1-\rho^j)^{\frac{m_j}{r_j}} \rho^{sj} \frac{(\frac{m_j}{r_j} + s - 1)\cdots(\frac{m_j}{r_j} + 1)\frac{m_j}{r_j}}{s!} \\
&= \{m_j\rho^j/(1-\rho^j)\}\epsilon_{js},
\end{aligned}$$

from which, for fixed j and $s \geq 2$, we deduce the limiting value

$$\epsilon^*_{js} = s^{-1}(1-\rho^j)\rho^{(s-1)j}$$

of ϵ_{js} as $r_j \to \infty$. Note that, as $j \to \infty$, $\varepsilon^*_{j2} \sim 2^{-1}\rho^j$ is of very much smaller order than the order $j^{-5/2}$ obtained for ε_{j2} when taking $r_j = 1$. As a result, the contributions to $H(l)$ arising from η'_0, η_1 and η_2, which enter in (2.24) and (2.25), are correspondingly reduced. As remarked earlier, letting $r_j \to \infty$ also allows us to take $p_0 = 1$ and $\eta_k(l) = 0$, $3 \leq k \leq 6$.

Similar arguments can be used for forests of unlabelled, rooted trees, now with $m_j \sim c'\rho^{-j} j^{-3/2}$. For forests of labelled, (un)rooted trees, $\mathcal{L}(T_{0\infty})$ is the compound Poisson distribution of $\sum_{j\geq 1} jZ_j$, where

$$Z_j \sim \mathrm{Po}\left(\frac{j^{j-2}}{j!e^j}\right) \quad \text{(unrooted)}; \qquad Z_j \sim \mathrm{Po}\left(\frac{j^{j-1}}{j!e^j}\right) \quad \text{(rooted)}.$$

The asymptotics of $\mathcal{L}(n - Y_n)$ then implied by Corollary 2.4 do not appear to agree with those of Mutafchiev (1998).

Acknowledgements

ADB was supported in this research by Schweizer Nationalfondsprojekt Nr. 20-67909.02. BG's research was supported by the Fund for the Promotion of Research at Technion.

References

- [1] R. Arratia, A. D. Barbour & S. Tavaré (2003) [ABT] *Logarithmic combinatorial structures: a probabilistic approach.* (to appear)

- [2] A. D. Barbour, L. H. Y. Chen & W.-L. Loh (1992) Compound Poisson approximation for nonnegative random variables using Stein's method. *Ann. Probab.* **20**, 1843–1866.

- [3] J. P. Bell, E. A. Bender, P. J. Cameron & L. B. Richmond (2000) Asymptotics for the probability of connectedness and the distribution of number of components. *Electron. J. Combin.* **7**, R33.

- [4] R. Durrett, B. Granovsky & S. Gueron (1999) The equilibrium behaviour of reversible coagulation–fragmentation processes. *J. Theoretical Probability* **12**, 447–474.

- [5] G. Freiman & B. Granovsky (2002a) Asymptotic formula for a partition function of reversible coagulation–fragmentation processes. *Israel J. Math.*, **130**, 259-279.

- [6] G. Freiman & B. Granovsky (2002b) Clustering in coagulation–fragmentation processes, random combinatorial structures and additive number systems: Asymptotic formulae and Zero-One law. *Preprint*, math.Pr/0207265, math@arXiv.org.
- [7] F. P. Kelly (1979) *Reversibility and stochastic networks.* Wiley, New York.
- [8] L. Mutafchiev (1998) The largest tree in certain models of random forests. *Rand. Struct. Alg.* **13**, 211–228.
- [9] R. Otter (1948) The number of trees. *Ann. Math.* **49**, 583–599.
- [10] B. Pittel (1997) Random set partitions: asymptotics of subset counts. *J. Comb. Theory A* **79**, 326–359.
- [11] P. Whittle (1965) Statistical processes of aggregation and polymerisation. *Proc. Cam. Phil. Soc.* **61**, 475–495.
- [12] P. Whittle (1986) *Systems in stochastic equilibrium.* Wiley, Chichester.