The Annals of Applied Probability
2000, Vol. 10, No. 2, 331–361

THE NUMBER OF COMPONENTS IN A LOGARITHMIC COMBINATORIAL STRUCTURE

BY RICHARD ARRATIA,[1] A. D. BARBOUR[2] AND SIMON TAVARÉ[1]

University of Southern California, Universität Zürich and University of Southern California

Under very mild conditions, we prove that the number of components in a decomposable logarithmic combinatorial structure has a distribution which is close to Poisson in total variation. The conditions are satisfied for all assemblies, multisets and selections in the logarithmic class. The error in the Poisson approximation is shown under marginally more restrictive conditions to be of exact order $O(1/\log n)$, by exhibiting the penultimate asymptotic approximation; similar results have previously been obtained by Hwang [20], under stronger assumptions. Our method is entirely probabilistic, and the conditions can readily be verified in practice.

1. Introduction. The joint distribution of the numbers $(C_1^{(n)}, \dots, C_n^{(n)})$ of cycles of sizes $1 \le i \le n$ in a uniformly chosen random permutation of n objects is also known as the Ewens Sampling Formula $\mathrm{ESF}_n(1)$; the more general $\mathrm{ESF}_n(\theta)$, $\theta > 0$, is obtained by weighting the distribution over the set of permutations by $\theta^{K_{0n}}$, where K_{0n} denotes the total number of cycles in the permutation. K_{0n} can be written as the sum of n independent Bernoulli random variables ξ_i with parameter $\theta/(\theta + i - 1)$:

$$K_{0n} = \sum_{i=1}^{n} \xi_i. \tag{1.1}$$

See Feller [13] and Rényi [28] for the case $\theta = 1$ and Watterson [34] for general θ. Hence, in particular, K_{0n} approximately has a Poisson distribution [8]:

$$d_{TV}(\mathscr{L}(K_{0n}), \mathrm{Po}\,(\kappa_{0n})) = O(1/\log n), \tag{1.2}$$

where $\kappa_{0n} = \sum_{i=1}^{n} \theta/(\theta + i - 1)$ and, for probability distributions P and Q on $\mathbb{Z}_+$,

$$d_{TV}(P, Q) := \sup_{A \subset \mathbb{Z}_+} |P(A) - Q(A)|.$$

Distributional approximations for K_{0n} have a long history. Goncharov [16, 17] showed that the number of cycles in a random permutation is asymptoti-

Received January 1999; revised August 1999.
[1]Supported in part by NSF grant DMS-96-26412.
[2]Supported in part by Schweizerischer NF Projekt Nr 20-50686.97.
AMS 1991 *subject classifications*. 60C05, 60F05, 05A16.
Key words and phrases. Logarithmic combinatorial structures, component counts, total variation approximation, Poisson approximation.

cally normally distributed as $n \to \infty$:

$$\frac{K_{0n} - \log n}{\sqrt{\log n}} \Rightarrow N(0,1). \tag{1.3}$$

Kolchin [22] derived approximations of the form

$$\mathbb{P}[K_{0n} = k] = \mathrm{Po}(\log n)\{k\}\ \{1 + o(1)\}, \tag{1.4}$$

uniformly with respect to $(\log n)^{-7/12}(k - \log n)$ in any finite interval, while Pavlov [27] established that

$$\mathbb{P}[K_{0n} = k] = \mathrm{Po}(\log n)\{k\}\left\{1 + O\left(\frac{|x|}{\sqrt{\log n}} + (\log n)^{-\delta}\right)\right\}, \tag{1.5}$$

for any fixed $\delta \in (0, 1/2)$, $k = \log n + x\sqrt{\log n}$, with $x = o(\sqrt{\log n})$. It follows from (1.5) that

$$d_{TV}(\mathscr{L}(K_{0n}), \mathrm{Po}\,(\log n)) = O\left((\log n)^{-1/2+\varepsilon}\right), \tag{1.6}$$

for any $\varepsilon > 0$, (1.3) then being an easy consequence. Analogous results for random mappings were proved by Stepanov [31], Kolchin [23] and Pavlov [27], for random mapping patterns by Mutafciev [25], for random polynomials over a finite field by Car [11], Hansen [18] and Arratia et al. [3], and for the irreducible factors of the characteristic polynomial of a matrix $T \in GL_n(F_q)$ by Goh and Schmutz [15]. Note that Brenti's ([10], Theorem 6.4.2) remarkable representation of the law of (K_{0n}) for random mappings as the law of $\sum_{i=1}^{n} \xi_i$ for independent Bernoulli random variables ξ_i implies a result analogous to (1.2), again from [8].

All these settings are particular examples of the larger logarithmic class of random decomposable combinatorial structures. By this, we mean random vectors $(C_1^{(n)}, \ldots, C_n^{(n)})$ satisfying the Conditioning Relation

$$\mathscr{L}(C_1^{(n)}, \ldots, C_n^{(n)}) = \mathscr{L}\left((Z_1, \ldots, Z_n) \,\Big|\, \sum_{i=1}^{n} iZ_i = n\right), \tag{1.7}$$

where $(Z_i,\ i \geq 1)$ are independent random variables on $\mathbb{Z}_+$, together with the Logarithmic Condition

$$\lim_{i\to\infty} i\mathbb{P}[Z_i = 1] = \lim_{i\to\infty} i\mathbb{E}Z_i = \theta, \tag{1.8}$$

for some $\theta > 0$. The best studied sub-classes, to which all of the examples above belong except for the random characteristic polynomials, are assemblies, where the Z_i have Poisson distributions; multisets, where the Z_i have negative binomial distributions; and selections, where the Z_i are binomially distributed. Random permutations are a logarithmic assembly, random polynomials a logarithmic multiset and random square free polynomials a logarithmic selection.

Flajolet and Soria [14] proved a central limit theorem for the number of components in considerable generality, as an application of their singularity analysis. Building on this, Hwang [20] has developed a powerful generating function approach to show that a Poisson approximation of the accuracy given in (1.2) can be established for a wide class of logarithmic assemblies, multisets and selections; furthermore, he identifies the leading asymptotic term in the total variation discrepancy. However, his method requires the distributions of the Z_i to come from families which are amenable to generating function arguments, as is indeed the case for those appearing in assemblies, multisets and selections. In this paper, we replace his approach by a probabilistic argument, which enables us to work with quite arbitrary distributions for the Z_i. We prove marginally more accurate approximations than Hwang's — we identify the penultimate approximating distribution — under conditions which are weaker than his, even for assemblies, multisets and selections, and which only involve simply calculated parameters of the distributions of the Z_i. We still use Fourier methods at one point, but only to make an accurate calculation in the context of the Ewens Sampling Formula, and not for the general distributions being approximated; our method for proving sharper approximations involves the refinement of Stein's method for perturbations of the Poisson distribution in Barbour and Xia [9].

Logarithmic combinatorial structures share a number of useful properties. First, it has been shown [29, 6, 30] under fairly general circumstances that, if $b = b(n) = o(n)$ as $n \to \infty$, then

$$d_{TV}(\mathscr{L}(C_1^{(n)}, \dots, C_b^{(n)}), \mathscr{L}(Z_1, \dots, Z_b)) \to 0. \tag{1.9}$$

In this case, the numbers of small components behave jointly like the independent random variables Z_i. In contrast, the sizes of the largest components $L_1^{(n)} \geq L_2^{(n)} \geq \cdots$ are essentially dependent; for any fixed r, the distribution of $n^{-1}(L_1^{(n)}, \dots, L_r^{(n)})$ is well approximated, locally as well as globally, by that of the r largest components of the Poisson–Dirichlet law PD(θ). See [21, 32, 2, 26] for particular structures and [19, 5] for general settings. For the distribution of $K_{0n} = \sum_{i=1}^{n} C_i^{(n)}$, the total number of components, both small and large components have to be considered simultaneously, and this makes the argument more complicated for our probabilistic approach; in contrast, K_{0n} is one of the easier quantities to treat by generating function methods.

A shortcut approach would be to take $b = n/\log n$ in (1.9), leading to a normal approximation for $\sum_{i=1}^{b} C_i^{(n)}$, and then to treat the remaining $\sum_{i=b+1}^{n} C_i^{(n)}$ needed for K_{0n} as a perturbation This indeed yields a central limit theorem, but since $\sum_{i=1}^{b} \operatorname{Var} C_i^{(n)} = O(\log n)$ and $\sum_{i=b+1}^{n} \mathbb{E} C_i^{(n)} = O(\log\log n)$, this approach would lead at best to an error estimate of order $O(\log\log n (\log n)^{-1/2})$ with respect to the usual metrics for weak convergence, an order much inferior to that in (1.2), and for the total variation metric it would yield nothing at all. Approximations such as (1.2) are more subtle, and are correspondingly more difficult to establish; in particular, even the difference between κ_{0n} and $\theta \log n$ is important for $\theta \neq 1$.

We establish two analogues of (1.2) in great generality. The first, Theorem 5.2, approximates the distribution of $K_{0n} = K_{0n}(C^{(n)})$ by the distribution of $K_{0,[\alpha n]}(Z)$, for a suitably chosen $\alpha = \alpha_\theta > 0$, where we use the notation

$$K_{vm}(X) = \sum_{i=v+1}^{m} X_i; \qquad T_{vm}(X) = \sum_{i=v+1}^{m} iX_i; \tag{1.10}$$

the error can be shown to be of order $O(1/\log n)$ under very mild conditions. The constant α_θ is given by $\exp\{\theta^{-1} - h_\theta\}$, where

$$h_t = \gamma + \Gamma'(t+1)/\Gamma(t+1) \tag{1.11}$$

is the t'th harmonic number $\sum_{j=1}^{t} 1/j$ if t is a positive integer; in the important special case where $\theta = 1$, it follows that $\alpha_\theta = 1$, and hence that $\mathscr{L}(K_{0n}) = \mathscr{L}(K_{0n}(Z) \mid T_{0n}(Z) = n)$ is then asymptotically close to the unconditional distribution of $K_{0n}(Z)$.

The second, Theorem 5.4, goes further than (1.2), showing that

$$\|\mathscr{L}(K_{0n}) - \nu_n\| = O(\log^{-3/2} n), \tag{1.12}$$

where

$$\nu_n\{s+1\} = \mathrm{Po}\,(\tau_n)\{s\}\left(1 + \tfrac{1}{2}a_n\tau_n^{-2}\{(s-\tau_n)^2 - \tau_n\}\right) \tag{1.13}$$

and

$$\tau_n = \sum_{i=1}^{n} \mathbb{E}Z_i - \theta h_\theta; \quad a_n = \sum_{i=1}^{n} (\mathrm{Var}\, Z_i - \mathbb{E}Z_i) - \theta^2 h'_\theta. \tag{1.14}$$

Here, $\|\cdot\|$ denotes the total variation norm, which it is more natural to use when signed measures, as ν_n may be, are involved: for probability measures Q_1 and Q_2,

$$\|Q_1 - Q_2\| = 2d_{TV}(Q_1, Q_2).$$

Under the conditions of Theorem 5.4, $\tau_n \sim \theta \log n$ and the limit $a_\infty = \lim_{n\to\infty} a_n$ exists and is finite. It is then easy to see that the measure ν_n is $O(1/\log n)$ away from the Poisson $\mathrm{Po}\,(\tau_n)$ distribution in total variation, whenever $a_\infty \neq 0$, so that no better rate than that given in (1.2) can be hoped for in general. For assemblies, multisets and selections, this second approximation holds provided that

$$d_{TV}(\mathscr{L}(Z_i), \mathrm{Po}\,(\theta/i)) = O((\log i)^{-2}). \tag{1.15}$$

The proof of these results still separates the treatment of the small and the large components. The small components are handled by proving in Section 4 that the convergence (1.9) takes place under extremely general conditions; as a consequence, the distributions of $K_{0b}(C^{(n)})$ and $K_{0b}(Z)$ can be matched, for suitable choice of b. In Section 3, we match the *conditional* distribution of $K_{bn}(C^{(n)})$ given $(C_1^{(n)}, \ldots, C_b^{(n)}) = (c_1, \ldots, c_b)$ to that of $K_{bn}(C^{*(n)})$ given $(C_1^{*(n)}, \ldots, C_b^{*(n)}) = (c_1, \ldots, c_b)$, where $C^{*(n)} = (C_1^{*(n)}, \ldots, C_n^{*(n)})$ has distribution $\mathrm{ESF}_n(\theta)$. It then remains to show that, for all $(c_1, \ldots, c_b)$ excluding a set of small probability, the conditional distribution of $K_{bn}(C^{*(n)})$ is close to the same fixed distribution, which is itself asymptotically Poisson. This final step is discussed in Section 2, and the complete approximation theorems in Section 5; the bulk of the detailed proof is deferred to Section 6.

We assume throughout that the Conditioning Relation (1.7) and the Logarithmic Condition (1.8) both hold, together with some supplementary condition. The weakest that we use, which strengthens the Logarithmic Condition slightly by implying some uniformity of decay in the tails of the distributions of the Z_i, but which imposes no extra requirement as to the rates of convergence in (1.8), is the Uniform Logarithmic Condition.

ULC(0): There exist constants $(e(i),\ i \geq 1)$ and $(c_s,\ s \geq 1)$ such that :

(i) $e(i) \downarrow 0$ as $i \to \infty$, and $D_1 = \sum_{s\geq 1} sc_s < \infty$;

(ii) $|i\mathbb{P}[Z_i = s] - \delta_{s1}\theta| \leq e(i)c_s$ for all $i, s \geq 1$.

For the sharper results, in which the orders of the error bounds are to be of a specified accuracy, some extra control is needed over the way in which the distributions of the Z_i, as is implicit in the Logarithmic Condition, come close to Be(θ/i). We choose to strengthen ULC(0) by requiring in addition either

$$\text{(1.16)} \quad \text{ULC(1):} \quad e(i) = O((\log i)^{-2}) \quad \text{and} \quad F(x) := \sum_{s>x} sc_s = O((\log x)^{-1})$$

or

$$\text{(1.17)} \quad \text{ULC}(r)\text{:} \quad e(i) = O((\log i)^{-2}) \ \text{ and } \ \sum_{s\geq 1} s^r c_s < \infty,$$

for some $r > 1$. These conditions are agreeably simple, but, as is often the case with moment conditions, could probably be slightly weakened. In the case of assemblies, multisets and selections, they are easy to verify, as is shown in the following proposition.

PROPOSITION 1.1. *For assemblies, multisets and selections, the Logarithmic Condition already implies that* ULC(0) *is satisfied, and* ULC(r) *is also satisfied for each* $r > 1$ *if, in addition,*

$$\text{(1.18)} \quad |i\mathbb{P}[Z_i = 1] - \theta| = O((\log i)^{-2}).$$

PROOF. For assemblies, where $Z_i \sim \text{Po}(\theta_i)$, and for selections, where $Z_i \sim \text{Bi}(m_i, p_i)$, observe that the Logarithmic Condition implies that, for $s \geq 2$,

$$\text{(1.19)} \quad i\mathbb{P}[Z_i = s] \leq \frac{i}{s!}(\mathbb{E}Z_i)^s \leq \frac{\theta_*^s}{s!i^{s-1}} \leq i^{-1}c_s,$$

where $c_s = \theta_*^s/s!$ and $\theta_* = \sup_{i\geq 1} i\mathbb{E}Z_i < \infty$; Condition ULC(0) follows automatically, with $e(i) = \max(i^{-1}, \sup_{j\geq i}|j\mathbb{P}[Z_j = 1] - \theta|)$, and ULC(r) also if $|i\mathbb{P}[Z_i = 1] - \theta| = O((\log i)^{-2})$.

For multisets, where $Z_i \sim \mathrm{NB}(m_i, p_i)$, the argument is somewhat more complicated. If $m_i \geq 1$ and $s \geq 2$, we have

$$(1.20)\qquad i\mathbb{P}[Z_i = s] \leq \begin{cases} \dfrac{i(m_i p_i)^s}{s!}\Big(1 + \dfrac{s}{m_i}\Big)^s \leq \dfrac{i}{s!}(2m_i p_i)^s, & \text{if } s \leq m_i, \\ ip_i^s 2^{m_i+s-1} \leq i(4p_i)^s \leq i(4m_i p_i)^s, & \text{if } s > m_i, \end{cases}$$

and $i\mathbb{P}[Z_i = s] \leq im_i p_i^s$ if $m_i < 1$; thus, whatever the value of m_i, we have

$$(1.21)\qquad i\mathbb{P}[Z_i = s] \leq i(4\theta_*/i)^s + \theta_* p_i^{s-1},$$

with θ_* as before. Now the Logarithmic Condition implies that

$$(1.22)\qquad \mathbb{P}[Z_i = 2]/\mathbb{P}[Z_i = 1] = \tfrac{1}{2}(m_i + 1)p_i \to 0$$

as $i \to \infty$, so that $p_i \to 0$ also. Hence, for $s \geq 2$ and $i > i_1 = \max(8\theta_*, i_0)$, where $i_0 = \max\{i : \ p_i > 1/2\}$, we have $i\mathbb{P}[Z_i = s] \leq e_0(i)c_s$ with $e_0(i) = \max(i^{-1}, p_i)$ and $c_s = 5\theta_* 2^{-(s-2)}$; hence $|i\mathbb{P}[Z_i = s] - \delta_{s1}\theta| \leq e(i)c_s$ for all $i > i_1$ and $s \geq 1$, with $e(i) = \sup_{j\geq i} \max(|j\mathbb{P}[Z_j = 1] - \theta|, e_0(j))$. The extension to all $i \geq 1$ is immediate, because $\mathbb{E}Z_i^r < \infty$ for all i, and ULC(r) is satisfied if $|i\mathbb{P}[Z_i = 1] - \theta| = O((\log i)^{-2})$. $\square$

2. Conditioning the Ewens Sampling Formula. Let $C^{*(n)}$ be distributed according to the Ewens Sampling Formula, the joint distribution resulting from the Conditioning Relation (1.7) when $Z_i = Z_i^* \sim \mathrm{Po}(\theta/i)$ for each i. In this section, we show that the conditional distribution of $K_{bn}(C^{*(n)})$ given $T_{bn}(C^{*(n)}) = l$ is order $O(\lambda_{bn}^{-1})$ close to the Poisson distribution $\mathrm{Po}(\theta\lambda_{bn} - \theta h_\theta + 1)$, uniformly in $n/2 \leq l \leq n$ and $0 \leq b \leq n/4$, where

$$\lambda_{bn} = \theta^{-1}\mathbb{E}(K_{bn}(Z^*)) = h_n - h_b;$$

the notation K_{bn} and T_{bn} is defined in (1.10) and h_t in (1.11). Note that the Conditioning Relation and the independence of the Z_i imply that we can equally consider the conditional distribution of $K_{bn}^* = K_{bn}(Z^*)$ given $T_{bn}^* = T_{bn}(Z^*) = l$; using the notation $X[r, s] = (X_r, \ldots, X_s)$ and suppressing the superscript $^{(n)}$, we have, for any $y \in \mathbb{Z}_+^n$,

$$\begin{aligned} &\mathbb{P}[C^*[b+1, n] = y[b+1, n] \mid T_{bn}(C^*) = l] \\ &\quad = \frac{\mathbb{P}[C^*[b+1, n] = y[b+1, n], T_{bn}(C^*) = l]}{\mathbb{P}[T_{bn}(C^*) = l]} \\ (2.1)\qquad &\quad = \frac{\mathbb{P}[Z^*[b+1, n] = y[b+1, n], T_{bn}(Z^*) = l, T_{0n}(Z^*) = n]}{\mathbb{P}[T_{bn}(Z^*) = l, T_{0n}(Z^*) = n]} \\ &\quad = \frac{\mathbb{P}[Z^*[b+1, n] = y[b+1, n], T_{bn}(Z^*) = l]\mathbb{P}[T_{0b}(Z^*) = n - l]}{\mathbb{P}[T_{bn}(Z^*) = l]\mathbb{P}[T_{0b}(Z^*) = n - l]} \\ &\quad = \mathbb{P}[Z^*[b+1, n] = y[b+1, n] \mid T_{bn}(Z^*) = l]. \end{aligned}$$

To derive the conditional distribution of K^*_{bn} given $T^*_{bn} = l$, note first that the *unconditional* distribution of K^*_{bn} is $\mathrm{Po}\,(\theta\lambda_{bn})$, and that, conditional on $K^*_{bn} = s$, the distribution of $Z^*[b+1, n]$ is multinomial:

$$\mathscr{L}(Z^*[b+1,n] \mid K^*_{bn} = s) = \mathrm{MN}\,(s;\, \mathscr{L}(U)), \tag{2.2}$$

where

$$\mathbb{P}[U = r] = 1/(r\lambda_{bn}), \quad b+1 \le r \le n. \tag{2.3}$$

Thus, conditional on $K^*_{bn} = s$, T^*_{bn} has the distribution of $W_s = \sum_{j=1}^{s} U_j$, where the $(U_j,\ j \ge 1)$ are independent and identically distributed with the distribution of U. Hence

$$\mathbb{P}[K^*_{bn} = s \mid T^*_{bn} = l] = \frac{\mathbb{P}[K^*_{bn} = s,\, T^*_{bn} = l]}{\mathbb{P}[T^*_{bn} = l]} = \mathrm{Po}\,(\theta\lambda_{bn})\{s\}\frac{\mathbb{P}[W_s = l]}{\mathbb{P}[T^*_{bn} = l]}, \tag{2.4}$$

and further progress depends on understanding the distribution of W_s. For the following approximation to the conditional probability in (2.4), whose proof is deferred to Section 6, we need to define the measure $\nu(\rho, c_1, c_2)$ with density

$$\nu(\rho, c_1, c_2)\{s\} = \mathrm{Po}\,(\rho)\{s\}\left(1 + c_1\rho^{-1}(s-\rho) + \tfrac{1}{2}c_2\rho^{-2}\{(s-\rho)^2 - \rho\}\right), \tag{2.5}$$

possibly signed, satisfying $\nu(\rho, c_1, c_2)\{\mathbb{Z}_+\} = 1$.

LEMMA 2.1. *Fix any* $0 < \gamma < 1$ *and* $1/2 < \alpha_1 < 1 < \alpha_2 < 3/2$, *and set* $b = [n^\gamma]$. *Then, uniformly in* $\alpha_1\theta\lambda_{bn} \le s \le \alpha_2\theta\lambda_{bn}$ *and in* $n/2 \le l \le n$, *we have*

$$\left|\frac{\mathrm{Po}\,(\theta\lambda_{bn})\{s+1\}\mathbb{P}[W_{s+1} = l]}{\mathbb{P}[T^*_{bn} = l]} - \nu(\theta\lambda_{bn}, -\theta h_\theta, \theta^2(h_\theta^2 - h'_\theta))\{s\}\right|$$
$$\le k\mathrm{Po}\,(\theta\lambda_{bn})\{s\}\left\{\lambda_{bn}^{-2}(1 + |s - \lambda_{bn}\theta|) + \lambda_{bn}^{-3}|s - \lambda_{bn}\theta|^3 + (1 - l/n)^{\bar{\theta}/2}\right\},$$

where $k = k(\theta, \gamma, \alpha_1, \alpha_2)$ *and* $\bar{\theta} = \min(\theta, 1)$.

As a result of this lemma, it is an easy matter to deduce that the conditional distribution of K^*_{bn}, given $T^*_{bn} = l$, is asymptotically well approximated by the same distribution for all l such that $n - l$ is much smaller than n, since λ_{bn} grows like $\log n$ as $n \to \infty$. An obvious candidate is $\nu(\theta\lambda_{bn}, -\theta h_\theta, \theta^2(h_\theta^2 - h'_\theta))$. For this choice, where both $c_1 = -\theta h_\theta$ and $c_2 = \theta^2(h_\theta^2 - h'_\theta)$ are fixed and $\rho = \theta\lambda_{bn} \asymp \log n \to \infty$, the measure $\nu(\rho, c_1, c_2)$ is a relatively small perturbation of the Poisson distribution $\mathrm{Po}\,(\rho)$. Note that $\nu(\rho, c_1, c_2)\{s\} > 0$ for all s in an interval $\alpha_1\rho \le s \le \alpha_2\rho$, where $0 < \alpha_1 < 1 < \alpha_2$ and $\alpha_l = \alpha_l(c_1, c_2)$, so that, by the properties of the Poisson distribution, there are *probability* measures ν' such that $\|\nu' - \nu(\rho, c_1, c_2)\| = O(e^{-\beta\rho})$, for some $\beta = \beta(c_1, c_2)$. It is also the case that, if σ, b_1 and b_2 are such that

$$\sigma + b_1 = \rho + c_1; \qquad b_2 - b_1^2 = c_2 - c_1^2, \tag{2.6}$$

then $\|\nu(\sigma, b_1, b_2) - \nu(\rho, c_1, c_2)\| \le k\rho^{-3/2}$ for some $k = k(c_1, c_2, b_1, b_2)$: see Barbour and Xia [9], Theorem 3.2.

THEOREM 2.2. *With* $b = [n^\gamma]$ *for* $0 < \gamma < 1$ *fixed, we have*

$$\|\mathscr{L}(K^*_{bn} - 1 \mid T^*_{bn} = l) - \nu(\theta\lambda_{bn}, -\theta h_\theta, \theta^2(h_\theta^2 - h'_\theta))\| = O(\log^{-3/2} n + (1 - l/n)^{\bar\theta/2})$$

for all $0 \le l \le n$.

PROOF. For $0 \le l < n/2$, it is clear that $\|\cdot\| \le 2$ is a good enough estimate. Otherwise, let ν denote $\mathscr{L}(K^*_{bn} - 1 \mid T^*_{bn} = l)$, so that, from Lemma 2.1,

$$|\nu\{s\} - \nu(\theta\lambda_{bn}, -\theta h_\theta, \theta^2(h_\theta^2 - h'_\theta))\{s\}| \le k_1 \zeta(s)$$

where

$$\zeta(s) = \mathrm{Po}\,(\theta\lambda_{bn})\{s\}\left\{\lambda_{bn}^{-2}(1 + |s - \lambda_{bn}\theta|) + \lambda_{bn}^{-3}|s - \lambda_{bn}\theta|^3 + (1 - l/n)^{\bar\theta/2}\right\},$$

for all $\alpha_1\theta\lambda_{bn} \le s \le \alpha_2\theta\lambda_{bn}$ and $n/2 \le l \le n$. However,

$$\sum_{s \ge 0} \zeta(s) \le k_2 \log^{-3/2} n + (1 - l/n)^{\bar\theta/2}, \tag{2.7}$$

and then, for any fixed $\alpha_1 < 1 < \alpha_2$,

$$\begin{aligned} &|\nu(\lambda, c_1, c_2)|\{[\alpha_1\lambda, \alpha_2\lambda]^c\} \\ &\quad \le \sum_{s \in [\alpha_1\lambda, \alpha_2\lambda]^c} \mathrm{Po}\,(\lambda)\{s\}\left(1 + \lambda^{-1}|c_1||s - \lambda| + \tfrac{1}{2}\lambda^{-2}|c_2|(s - \lambda)^2\right) \\ &\quad \le (1 + |c_1| + \tfrac{1}{2}|c_2|)\left\{\mathrm{Po}\,(\lambda)\{[0, \alpha_1\lambda]\} + \mathrm{Po}\,(\lambda)\{[\alpha_2\lambda - 2, \infty)\}\right\}, \end{aligned} \tag{2.8}$$

exponentially small as $\lambda \to \infty$. Then (2.7) and (2.8) with $\lambda = \theta\lambda_{bn}$ also imply that $\nu\{[\alpha_1\theta\lambda_{bn}, \alpha_2\theta\lambda_{bn}]^c\} = O(\log^{-3/2} n)$, and the theorem follows. □

An alternative version, which exploits the freedom of choice of ρ, c_1 and c_2 in (2.6), uses an expression which is obviously close to a Poisson distribution, in that c_1 is chosen to be zero. Let R_{bn} denote the distribution of $1 + W$, where W is a random variable with distribution differing as little as possible from $\nu(\rho_{bn}, 0, -\theta^2 h'_\theta)$, where

$$\rho_{bn} = \theta(h_n - h_b - h_\theta); \tag{2.9}$$

take

$$R_{bn}\{s+1\} = \mathbb{P}[W = s] = \mathrm{Po}\,(\rho_{bn})\{s\}\left(1 - \tfrac{1}{2}\rho_{bn}^{-2}\theta^2 h'_\theta\{(s - \rho_{bn})^2 - \rho_{bn}\}\right) \tag{2.10}$$

for $|s - \rho_{bn}| \le \gamma_\theta\rho_{bn}$, with $\gamma_\theta = \{2\theta^2 h'_\theta\}^{-1/2}$, and set

$$R_{bn}\{[\rho_{bn}(1 + \gamma_\theta)] + 1\} = 1 - R_{bn}\left\{[0, [\rho_{bn}(1 + \gamma_\theta)]]\right\},$$

noting that both

$$R_{bn}\left\{([\rho_{bn}(1 + \gamma_\theta)], \infty)\right\} \quad \text{and} \quad \sum_{j > [\rho_{bn}(1+\gamma_\theta)]} |\nu(\rho_{bn}, 0, -\theta^2 h'_\theta)|\{j\}(j + \rho_{bn})$$

are of smaller order than $O(\rho_{bn}^{-3/2})$.

COROLLARY 2.3. *If* $b = [n^\gamma]$ *for fixed* $0 < \gamma < 1$, *then*

$$d_{TV}(\mathscr{L}(K^*_{bn} \mid T^*_{bn} = l), R_{bn}) \le k(\log^{-3/2} n + (1 - l/n)^{\bar{\theta}/2})$$

for all $0 \le l \le n$, *for some* $k = k(\gamma, \theta)$.

3. The large components. In this section, we show that the conditional distribution of the sizes of the large components, given the sizes of the small components, is almost the same for all the logarithmic combinatorial structures under consideration. Using the notation $X[r, s] = (X_r, \dots, X_s)$, we prove that the conditional distribution of $C[b+1, n]$, given that $T_{bn}(C) = l$, is close to that of $C^*[b+1, n]$ given $T_{bn}(C^*) = l$, uniformly in $n/2 \le l \le n$ and $0 \le b \le n/4$, where C^* is distributed according to the Ewens Sampling Formula and all superscripts $^{(n)}$ are suppressed. As a result, the conditional distribution of $K_{bn}(C)$ is close to that of $K_{bn}(C^*)$, so that the approximations in Theorem 2.2 can be carried over to quite general logarithmic combinatorial structures.

As a first step, we show that the unconditional distributions of $Z[b+1, n]$ and $Z^*[b+1, n]$ are close. We assume throughout that Condition ULC(0) holds, and note that then

$$|\mathbb{P}[Z_i = 0] - (1 - \theta/i)| \le D_1 i^{-1} e(i); \quad |\mathbb{P}[Z_i = 1] - \theta/i| \le D_1 i^{-1} e(i) \tag{3.1}$$

and

$$\mathbb{P}[Z_i \ge 2] \le D_1 i^{-1} e(i) : \tag{3.2}$$

we define

$$E_L(b, n) = b^{-1} + e(b) + \sum_{i=b+1}^{n} i^{-1} e(i) + n^{-1} \log n. \tag{3.3}$$

LEMMA 3.1. *If* ULC(0) *holds, then*

$$d_{TV}(\mathscr{L}(Z[b+1, n]), \mathscr{L}(Z^*[b+1, n])) = O(E_L(b, n)), \tag{3.4}$$

uniformly in b *and* n.

PROOF. Take any $y \in \{0, 1\}^n$. Then

$$\begin{aligned}
\frac{\mathbb{P}[Z[b+1, n] = y[b+1, n]]}{\mathbb{P}[Z^*[b+1, n] = y[b+1, n]]} &= \prod_{i=b+1}^{n} \frac{\mathbb{P}[Z_i = 0]^{1-y_i} \mathbb{P}[Z_i = 1]^{y_i}}{e^{-\theta/i} (\theta/i)^{y_i}} \\
&\ge \prod_{i=b+1}^{n} (1 - D_1 i^{-1} e(i) - O(i^{-2}))(1 - D_1 \theta^{-1} e(i))^{y_i} \\
&\ge 1 - D_1 \sum_{i=b+1}^{n} i^{-1} e(i) - D_1 \theta^{-1} \sum_{i=b+1}^{n} e(i) y_i - O(b^{-1}).
\end{aligned}$$

Hence, for any $A \subset \mathbb{Z}_+^{n-b}$, it follows that

$$\begin{aligned}\mathbb{P}[Z[b+1,n] \in A] &\geq \mathbb{P}[Z[b+1,n] \in A \cap \{0,1\}^{n-b}]\\ &\geq \mathbb{P}[Z^*[b+1,n] \in A] - \mathbb{P}[Z^*[b+1,n] \in A \setminus \{0,1\}^{n-b}]\\ &\quad - D_1\theta^{-1} \sum_{i=b+1}^{n} e(i)\mathbb{P}[Z_i^* = 1] - O(E_L(b,n))\\ &= \mathbb{P}[Z^*[b+1,n] \in A] - O(E_L(b,n)),\end{aligned}$$

since $\mathbb{P}[\bigcup_{i=b+1}^{n}\{Z_i^* \geq 2\}] = O(b^{-1})$ and $\mathbb{P}[Z_i^* = 1] \leq i^{-1}\theta$. The lemma is now immediate. □

The next step is to compare the densities of $T_{bn}(Z)$ and $T_{bn}(Z^*)$. The random variables $T_{bn}(Z^*)$ have been well studied. As observed in [7],

$$l\mathbb{P}[T_{bn}(Z^*) = l] = \theta\mathbb{P}[T_{bn}(Z^*) < l - b], \tag{3.5}$$

and, as proved by Vervaat [33], $n^{-1}T_{0n}(Z^*) \xrightarrow{\mathscr{D}} X_\theta$, where X_θ has an everywhere positive density f_θ on $(0,\infty)$ which satisfies the equation

$$xf_\theta(x) = \theta\mathbb{P}[x-1 < X_\theta \leq x], \quad x > 0; \tag{3.6}$$

in particular,

$$\liminf_{n\to\infty} \mathbb{P}[T_{bn}(Z^*) < n/4] \geq \lim_{n\to\infty} \mathbb{P}[T_{0n}(Z^*) < n/4] = \mathbb{P}[X_\theta < 1/4] > 0. \tag{3.7}$$

In the next lemma, we show that the density of $T_{bn}(Z)$ has much the same properties.

LEMMA 3.2. *If* ULC(0) *holds, then*

$$(1)\colon l\mathbb{P}[T_{bn}(Z) = l] \leq \theta + O(E_L(1,n))$$

for all l *and* b. *Furthermore,*

$$(2)\colon \frac{\mathbb{P}[T_{bn}(Z) = l]}{\mathbb{P}[T_{bn}(Z^*) = l]} = 1 + O(E_L(b,n)),$$

uniformly in $n/2 \leq l \leq n$ *and* $0 \leq b \leq n/4$.

PROOF. Start with any $l, b \geq 0$. Writing T_{bn} for $T_{bn}(Z)$ and setting $T_{bn}^{(i)} = T_{bn} - iZ_i$, we have

$$\begin{aligned}l\mathbb{P}[T_{bn} = l] &= \mathbb{E}(T_{bn}1_{\{l\}}(T_{bn}))\\ &= \sum_{i=b+1}^{n}\sum_{s\geq 1} is\mathbb{P}[Z_i = s]\mathbb{E}\{1_{\{l\}}(T_{bn}^{(i)} + is)\}.\end{aligned} \tag{3.8}$$

Now, since conditioning gives

$$\mathbb{P}[T_{bn} = l - m] = \sum_{s\geq 0} \mathbb{P}[Z_i = s]\mathbb{P}[T_{bn}^{(i)} = l - m - is], \tag{3.9}$$

it follows that

$$|\mathbb{P}[T_{bn} = l - m] - (1 - \theta/i)\mathbb{P}[T^{(i)}_{bn} = l - m] - (\theta/i)\mathbb{P}[T^{(i)}_{bn} = l - m - i]|$$
$$\leq \frac{\theta}{i}\,\frac{D_1 e(i)}{\theta}\left\{\mathbb{P}[T^{(i)}_{bn} = l - m] + \mathbb{P}[T^{(i)}_{bn} = l - m - i]\right\} + O(i^{-1}e(i))$$
$$= O(i^{-1}e(i)),$$

and hence that

$$|\mathbb{P}[T^{(i)}_{bn} = l - m] - \mathbb{P}[T_{bn} = l - m]| = O(i^{-1}), \quad b + 1 \leq i \leq n$$

and

$$\mathbb{P}[T^{(i)}_{bn} = l - i] = \mathbb{P}[T_{bn} = l - i] - i^{-1}\theta\{\mathbb{P}[T_{bn} = l - 2i] - \mathbb{P}[T_{bn} = l - i]\} + O(i^{-2} + i^{-1}e(i)).$$

Using these last two estimates to simplify the right hand side of (3.8), we obtain

$$\begin{aligned} l\mathbb{P}[T_{bn} = l] &= \sum_{i=b+1}^{n} i\mathbb{P}[Z_i = 1] \\ &\quad\times\Big\{\mathbb{P}[T_{bn} = l - i] - i^{-1}\theta\{\mathbb{P}[T_{bn} = l - 2i] - \mathbb{P}[T_{bn} = l - i]\} \\ &\qquad + O(i^{-2} + i^{-1}e(i))\Big\} \\ &\quad + \sum_{i=b+1}^{n}\sum_{s\geq 2} is\mathbb{P}[Z_i = s]\{\mathbb{P}[T_{bn} = l - is] + O(i^{-1})\} \\ &= \theta\mathbb{P}[T_{bn} < l - b] + O(E_L(b,n)) + O\left\{\sum_{s\geq 2} s \max_{i\geq b+1} i\mathbb{P}[Z_i = s]\right\} \\ &= \theta\mathbb{P}[T_{bn} < l - b] + O(E_L(b,n)), \end{aligned}$$

in view of ULC(0), where we have used the fact that $\sum_{r\geq 0}\mathbb{P}[T_{bn} = r] = 1$. Part (1) now follows, since $E_L(b,n) \leq E_L(1,n)$ for all b.

Now take $n/2 \leq l \leq n$ and $0 \leq b \leq n/4$. Then it follows that

$$\frac{\mathbb{P}[T_{bn}(Z) = l]}{\mathbb{P}[T_{bn}(Z^*) = l]} = \frac{\mathbb{P}[T_{bn}(Z) < l - b] + O(E_L(b,n))}{\mathbb{P}[T_{bn}(Z^*) < l - b]} = 1 + O(E_L(b,n)),$$

from Lemma 3.1, (3.5) and (3.7). This proves Part (2). □

As a consequence of Lemma 3.2, we can also derive the following estimate, which is used in Section 5.

COROLLARY 3.3. *Uniformly in* $0 \leq b \leq n/4$, *for any* $0 < \beta \leq 1$,

$$\mathbb{E}(T^{\beta}_{0b}(C)\, I[T_{0b}(C) \leq n/2]) = O(b^{\beta}[1 + E_L(b,n)]).$$

PROOF. Direct calculation gives

$$\begin{aligned}
&\mathbb{E}(T_{0b}^{\beta}(C)\, I[T_{0b}(C) \le n/2]) \\
&\quad = \sum_{t=0}^{[n/2]} \frac{t^{\beta}\mathbb{P}[T_{0b}(Z) = t]\mathbb{P}[T_{bn}(Z) = n - t]}{\mathbb{P}[T_{0n}(Z) = n]} \\
&\quad = \sum_{t=0}^{[n/2]} \frac{t^{\beta}\mathbb{P}[T_{0b}(Z) = t]\mathbb{P}[T_{bn}(Z^*) = n - t]}{\mathbb{P}[T_{0n}(Z^*) = n]}\{1 + O(E_L(b, n))\} \\
&\quad \le 2 \sum_{t=0}^{[n/2]} \frac{t^{\beta}\mathbb{P}[T_{0b}(Z) = t]}{\mathbb{P}[T_{0n}^{*} \le n - 1]}\{1 + O(E_L(b, n))\},
\end{aligned}$$

where Lemma 3.2(2) and (3.5) give the last two lines. But

$$\begin{aligned}
\sum_{t=0}^{[n/2]} t^{\beta}\mathbb{P}[T_{0b}(Z) = t] &\le \mathbb{E}\left\{T_{0b}^{\beta}(Z)\right\} \le \{\mathbb{E}T_{0b}(Z)\}^{\beta} \\
&= \left(\sum_{i=1}^{b} i \sum_{s\ge 1} s\mathbb{P}[Z_i = s]\right)^{\beta} \le b^{\beta}(\theta + e(1)(1 + D_1))^{\beta},
\end{aligned}$$

by ULC(0), which, with (3.7), proves the corollary. □

Combining these two preliminary results, we can now prove the closeness of the conditional distributions of the sizes of the large components in an arbitrary combinatorial structure satisfying ULC(0) and in the Ewens Sampling Formula.

THEOREM 3.4. *For any combinatorial structure satisfying* ULC(0),

$$d_{TV}(\mathscr{L}(C[b+1, n] \mid T_{bn}(C) = l), \mathscr{L}(C^*[b+1, n] \mid T_{bn}(C^*) = l)) = O(E_L(b, n)),$$

uniformly in $n/2 \le l \le n$ *and* $0 \le b \le n/4$.

PROOF. We start much as in the proof of Lemma 3.1. For any $y \in \{0, 1\}^n$ such that $\sum_{i=b+1}^{n} iy_i = l$, we have

$$\frac{\mathbb{P}[C[b+1, n] = y[b+1, n] \mid T_{bn}(C) = l]}{\mathbb{P}[C^*[b+1, n] = y[b+1, n] \mid T_{bn}(C^*) = l]}$$

$$= \frac{\mathbb{P}[Z[b+1, n] = y[b+1, n]]}{\mathbb{P}[Z^*[b+1, n] = y[b+1, n]]}\, \frac{\mathbb{P}[T_{bn}(Z^*) = l]}{\mathbb{P}[T_{bn}(Z) = l]} \tag{3.10}$$

$$\ge 1 - (D_1/\theta) \sum_{i=b+1}^{l} e(i)y_i - O(E_L(b, n))$$

$$\ge 1 - (D_1/\theta) \left\{ \sum_{i=b+1}^{[l/2]} e(i)y_i + e([l/2]) \sum_{i=[l/2]+1}^{l} y_i \right\} - O(E_L(b, n)). \tag{3.11}$$

We thus find that, for any $A \subset \mathbb{Z}_+^{n-b}$,

$$
\begin{aligned}
&\mathbb{P}[C[b+1,n] \in A \mid T_{bn}(C) = l] \\
&\quad \geq \mathbb{P}[C^*[b+1,n] \in A \mid T_{bn}(C^*) = l] \\
&\quad\quad -\mathbb{P}\left[\bigcup_{i=b+1}^{n} \{C_i^* \geq 2\} \mid T_{bn}(C^*) = l\right] \\
&\quad\quad -(D_1/\theta) \sum_{i=b+1}^{[l/2]} e(i)\mathbb{P}[C_i^* = 1 \mid T_{bn}(C^*) = l] - O(E_L(b,n)),
\end{aligned} \tag{3.12}
$$

since, if $T_{bn}(C^*) = l$, then $\sum_{i=[l/2]+1}^{l} C_i^* \leq 1$. Hence there are two remaining elements to be bounded in (3.12).

First, using the Conditioning Relation, (3.5) and (3.9), we observe that, for $i \leq l/2$,

$$
\begin{aligned}
\mathbb{P}[C_i^* = 1 \mid T_{bn}(C^*) = l] &= \frac{\mathbb{P}[Z_i^* = 1]\mathbb{P}[T_{bn}^{(i)}(Z^*) = l - i]}{\mathbb{P}[T_{bn}^* = l]} \\
&\leq e^{-\theta/i}\frac{\theta}{i}\,\frac{e^{\theta/i}\mathbb{P}[T_{bn}^* < l-i-b]}{\mathbb{P}[T_{bn}^* < l-b]}\left(\frac{l}{l-i}\right) = O(i^{-1}),
\end{aligned}
$$

where T_{bn}^* denotes $T_{bn}(Z^*)$, so that

$$
(D_1/\theta) \sum_{i=b+1}^{[l/2]} e(i)\mathbb{P}[C_i^* = 1 \mid T_{bn}(C^*) = l] = O(E_L(b,n)). \tag{3.13}
$$

Then, by similar estimates,

$$
\begin{aligned}
\mathbb{P}[C_i^* = r \mid T_{bn}(C^*) = l] &= \frac{e^{-\theta/i}}{r!}\left(\frac{\theta}{i}\right)^r \frac{\mathbb{P}[T_{bn}^{(i)}(Z^*) = l - ir]}{\mathbb{P}[T_{bn}^* = l]} \\
&\leq \frac{1}{r!}\left(\frac{\theta}{i}\right)^r \frac{\mathbb{P}[T_{bn}^* < l-ir-b]}{\mathbb{P}[T_{bn}^* < l-b]}\left(\frac{(l/r)}{(l/r)-i}\right).
\end{aligned} \tag{3.14}
$$

If $i \leq l/2r$, we bound (3.14) by $\frac{2\theta^r}{r!i^r}$; if $l/2r < i < l/r$, we bound by $\frac{(2r\theta)^r}{r!l^r}\left(\frac{(l/r)}{(l/r)-i}\right)$, and if $i = l/r$, we bound by $\frac{(r\theta)^r}{r!l^r}\,\frac{l}{\mathbb{P}[T_{bn}^* < l-b]}$. Adding over the range $i \geq b+1$, this gives

$$
\begin{aligned}
&\mathbb{P}\left[\bigcup_{i=b+1}^{n} \{C_i^* = r\} \mid T_{bn}(C^*) = l\right] \\
&\quad \leq \frac{2\theta^r}{r!b^{r-1}} + \frac{(2e\theta)^r}{r^{3/2}l^{r-1}}\log(l/r) + \frac{(e\theta)^r}{r^{1/2}l^{r-1}\mathbb{P}[T_{0n}(Z^*) < n/4]},
\end{aligned}
$$

and hence, adding over $r \geq 2$, it follows that

$$
\mathbb{P}\left[\bigcup_{i=b+1}^{n} \{C_i^* \geq 2\} \mid T_{bn}(C^*) = l\right] = O(b^{-1} + n^{-1}\log n) = O(E_L(b,n)), \tag{3.15}
$$

uniformly in $n/2 \le l \le n$ and $0 \le b \le n/4$. Putting (3.13) and (3.15) into (3.12) gives the theorem. □

4. The small components. The purpose of this section is to prove the approximation (1.9) of the joint distribution of the sizes of the small components $(C_1^{(n)}, \ldots, C_b^{(n)})$ by that of the independent random variables $(Z_1, \ldots, Z_b)$, in the generality required in this paper. In order to achieve this, we first need two technical estimates. Define

$$b_2 = b_2(Z) = \max\{i : \mathbb{P}[Z_i = 0] < 1/2\}, \tag{4.1}$$

finite because of the Logarithmic Condition, and recall that

$$F(x) = \sum_{s>x} sc_s, \tag{4.2}$$

where the c_s are as for ULC(0).

LEMMA 4.1. *If* ULC(0) *holds, then*

$$m\mathbb{P}[T_{0b}(Z) = m] = O(n^{-1}b + F(n/4b)),$$

uniformly in $b_2 \le b \le m/2$ *and* $m \ge n/2$. *In particular*, $m\mathbb{P}[T_{0b}(Z) = m] = o(1)$ *if* $b = o(m)$, *and is of order* $O(b/m)$ *under* ULC(2).

PROOF. As for equation (3.8),

$$\begin{aligned} m\mathbb{P}[T_{0b} = m] &= \sum_{i=1}^{b} \sum_{s\ge 1} is\mathbb{P}[Z_i = s]\mathbb{P}[T_{0b}^{(i)} = m - is] \\ &\le \sum_{i=1}^{b} \sum_{s\ge 1} e(i)sc_s\mathbb{P}[T_{0b}^{(i)} = m - is] + \theta \sum_{i=1}^{b} \mathbb{P}[T_{0b}^{(i)} = m - i], \end{aligned} \tag{4.3}$$

where $T_{0b} = T_{0b}(Z)$ and $T_{0b}^{(i)} = T_{0b} - iZ_i$. Now, from Markov's inequality, for any $t > 0$,

$$\mathbb{P}[T_{0b}^{(i)} \ge t] \le \mathbb{P}[T_{0b} \ge t] \le t^{-1}\mathbb{E}T_{0b} = O(b/t), \tag{4.4}$$

uniformly in b and t, by the Logarithmic Condition and from the definition (1.10) of T_{0b}. Hence, for $1 \le i \le b_2$ in (4.3), we have

$$\begin{aligned} &\sum_{i=1}^{b_2} \left(\sum_{s\ge 1} e(i)sc_s\mathbb{P}[T_{0b}^{(i)} = m - is] + \theta\mathbb{P}[T_{0b}^{(i)} = m - i] \right) \\ &\quad \le \sum_{i=1}^{b_2} \left\{ e(1) \sum_{s=1}^{[m/2b_2]} sc_s + \theta \right\} \mathbb{P}[T_{0b}^{(i)} \ge m/2] + e(1)b_2 \sum_{s>m/2b_2} sc_s \\ &\quad \le b_2 \left\{ (\theta + e(1)D_1)4n^{-1}\mathbb{E}T_{0b} + e(1)F(n/4b_2) \right\} \\ &\quad = O(n^{-1}b + F(n/4b_2)), \end{aligned} \tag{4.5}$$

uniformly in $m \geq n/2$. For $i > b_2$ in (4.3), we use the elementary inequality $\mathbb{P}[T_{0b}^{(i)} = m - is] \leq 2\mathbb{P}[T_{0b} = m - is]$ to give

$$\theta \sum_{i=b_2+1}^{b} \mathbb{P}[T_{0b}^{(i)} = m - i] \leq 2\theta\mathbb{P}[T_{0b} \geq m/2] = O(b/n), \tag{4.6}$$

again by Markov's inequality, uniformly in $b \leq m/2$ and $m \geq n/2$. Finally, since $\sum_{i\geq 0} \mathbb{P}[T_{0b} = m - is] \leq 1$ for any s, it follows that

$$\begin{aligned}
&\sum_{i=b_2+1}^{b} \sum_{s\geq 1} e(i)sc_s\mathbb{P}[T_{0b}^{(i)} = m - is] \\
&\quad \leq 2e(1)\left\{\sum_{i=b_2+1}^{b} \sum_{s=1}^{[m/2b]} sc_s\mathbb{P}[T_{0b} = m - is] + \sum_{s>m/2b} sc_s\right\} \\
&\quad \leq 2e(1)\left\{D_1\mathbb{P}[T_{0b} \geq m/2] + \sum_{s>m/2b} sc_s\right\} \\
&\quad = O(n^{-1}b + F(n/4b)),
\end{aligned} \tag{4.7}$$

uniformly in $b_2 \leq b \leq m/2$ and $m \geq n/2$. Putting (4.5) – (4.7) into (4.3), the lemma follows. □

LEMMA 4.2. *If* ULC(0) *holds, then* $\liminf_{n\to\infty} n\mathbb{P}[T_{0n}(Z) = n] > 0$.

PROOF. Pick $b = b(n) \leq n/4$ in such a way that $b = o(n)$ and that b still grows fast enough to ensure that $E_L(b, n) \to 0$ as $n \to \infty$. This can be achieved, for instance, by taking

$$b(n) = \begin{cases} [n^{1/2}], & \text{if } e([\sqrt{n}]) \leq 4/\log^2 n, \\ \min\left([n\exp\{-\{e([\sqrt{n}])\}^{-1/2}\}], [n/4]\right), & \text{otherwise,} \end{cases}$$

and any *larger* sequences $b(n)$ still satisfying $b(n) = o(n)$ are also suitable. Then we have

$$\begin{aligned}
n\mathbb{P}[T_{0n}(Z) = n] &= \sum_{s\geq 0} n\mathbb{P}[T_{0b}(Z) = s]\mathbb{P}[T_{bn}(Z) = n - s] \\
&\geq n \sum_{s=0}^{[n/2]} \mathbb{P}[T_{0b}(Z) = s]\mathbb{P}[T_{bn}(Z^*) = n - s](1 - O(E_L(b, n))) \\
&\geq \tfrac{1}{2}n \sum_{s=0}^{[n/2]} \mathbb{P}[T_{0b}(Z) = s]\mathbb{P}[T_{bn}(Z^*) = n - s]
\end{aligned}$$

for all n large enough, by Lemma 3.2 (2). But now, from (3.5), for $s \leq n/2$,

$$n\mathbb{P}[T_{bn}(Z^*) = n - s] \geq \theta\mathbb{P}[T_{bn}(Z^*) < n - s - b] \geq \theta\mathbb{P}[T_{0n}(Z^*) < n/4]$$

is bounded away from 0, and $\mathbb{P}[T_{0b}(Z) < n/2] = 1 - O(n^{-1}b)$ from (4.4), completing the proof of the lemma. □

THEOREM 4.3. *If* ULC(0) *holds, then*

$$\Delta_1 = d_{TV}(\mathscr{L}(C[1,b]), \mathscr{L}(Z[1,b])) = O(E_S(b,n)),$$

uniformly in $0 \le b \le n/4$, *where*

$$E_S(b,n) = E_L(b,n) + n^{-1}b + F(n/4b); \tag{4.8}$$

$E_L(b,n)$ *is as defined in* (3.3) *and* $F(x)$ *as in* (4.2). *If* $b \asymp n^\gamma$ *for some* $0 < \gamma < 1$ *and* ULC(2) *holds, we have*

$$d_{TV}(\mathscr{L}(C[1,b]), \mathscr{L}(Z[1,b])) = O(1/\log n).$$

REMARK. Under ULC(0), if $b = b(n) = o(n)$ is such that $E_L(b,n) \to 0$, then $E_S(b,n) \to 0$ also. Hence $d_{TV}(\mathscr{L}(C[1,b]), \mathscr{L}(Z[1,b])) \to 0$ as $n \to \infty$ whenever $b = o(n)$ if ULC(0) holds, since $d_{TV}(\mathscr{L}(C[1,b]), \mathscr{L}(Z[1,b]))$ increases with b. This is a very weak condition for such convergence. For smaller values of b, the element $E_L(b,n)$ in the estimate of Δ is not accurate, but sharper rates under such circumstances are not needed in this paper; we are free to choose any (large) b that suits us, and we only require an approximation of order $O(1/\log n)$. With rather more work, sharper approximations of order $O(n^{-1}b + F(n/4b))$ can in fact be established in considerable generality [4].

PROOF. As in [4], from the conditioning relation, we have

$$\begin{aligned} 2\Delta_1 &= \sum_{r\ge 0} |\mathbb{P}[T_{0b}(C^{(n)}) = r] - \mathbb{P}[T_{0b}(Z) = r]| \\ &= \mathbb{P}[T_{0b} > n] + \sum_{r=0}^{n} p_r \left| 1 - \frac{\mathbb{P}[T_{bn} = n-r]}{\mathbb{P}[T_{0n} = n]} \right| \\ &\le \mathbb{P}[T_{0b} > n/2] + r_n^{-1} \sum_{r=[n/2]+1}^{n} p_r q_{n-r} \qquad (4.9) \\ &\quad + r_n^{-1} \sum_{r=0}^{[n/2]} p_r \left| \sum_{s\ge 0} p_s (q_{n-s} - q_{n-r}) \right|, \qquad (4.10) \end{aligned}$$

where we use the shorthand $p_t = \mathbb{P}[T_{0b} = t]$, $q_t = \mathbb{P}[T_{bn} = t]$ and $r_n = \mathbb{P}[T_{0n} = n]$, the argument Z being suppressed where possible. Hence it follows that

$$\begin{aligned} 2\Delta_1 \le{}& \mathbb{P}[T_{0b} > n/2] + r_n^{-1} \sum_{r=[n/2]+1}^{n} p_r q_{n-r} \\ &+ r_n^{-1} \sum_{r=0}^{[n/2]} p_r \left\{ \sum_{s=0}^{[n/2]} p_s |q_{n-s} - q_{n-r}| \right. \qquad (4.11) \\ &\left. + \sum_{s=[n/2]+1}^{n} p_s q_{n-s} + q_{n-r}\mathbb{P}[T_{0b} > n/2] \right\}. \end{aligned}$$

Most of the elements in (4.11) can be easily bounded. First, from Lemma 4.2, we have $r_n \geq cn^{-1}$ for some $c > 0$, and then

$$r_n^{-1} \sum_{r=0}^{[n/2]} p_r q_{n-r} \mathbb{P}[T_{0b} > n/2] \leq \mathbb{P}[T_{0b} > n/2] = O(b/n), \tag{4.12}$$

from (4.4). Furthermore, from Lemmas 4.1 and 4.2,

$$r_n^{-1} \sum_{r=[n/2]+1}^{n} p_r q_{n-r} \leq r_n^{-1} \max_{r>n/2} p_r = O(n^{-1}b + F(n/4b)). \tag{4.13}$$

Combining these estimates, we thus find that

$$2\Delta_1 = r_n^{-1} \sum_{r=0}^{[n/2]} \sum_{s=0}^{[n/2]} p_r p_s |q_{n-s} - q_{n-r}| + O(E_S(b, n)). \tag{4.14}$$

However, from Lemma 3.2,

$$\begin{aligned} &|(q_{n-s} - q_{n-r}) - (\mathbb{P}[T^*_{bn} = n - s] - \mathbb{P}[T^*_{bn} = n - r])| \\ &\quad = O\left(\max_{t \geq n/2} \mathbb{P}[T^*_{bn} = t]\, E_L(b, n)\right) = O(n^{-1} E_L(b, n)), \end{aligned} \tag{4.15}$$

since

$$\mathbb{P}[T^*_{bn} = t] = \theta t^{-1} \mathbb{P}[T^*_{bn} < t - b] \leq t^{-1}\theta$$

for all $t \geq 1$, by (3.5); and, again by (3.5), for $0 \leq s, r \leq n/2$,

$$\begin{aligned} &|\mathbb{P}[T^*_{bn} = n - s] - \mathbb{P}[T^*_{bn} = n - r]| \\ &\quad = \theta \left| \frac{\mathbb{P}[T^*_{bn} < n - s - b]}{n - s} - \frac{\mathbb{P}[T^*_{bn} < n - r - b]}{n - r} \right| \\ &\quad \leq \frac{2\theta}{n} |s - r| \frac{4\theta}{n} + \frac{4\theta |r - s|}{n^2} \\ &\quad \leq 4\theta n^{-2}(r + s)(1 + 2\theta), \end{aligned} \tag{4.16}$$

where, for the first inequality, we use (3.5) to give the bound $\mathbb{P}[T^*_{bn} = l] \leq 4\theta n^{-1}$ for $l \geq n/4$. Substituting (4.15) and (4.16) into (4.14) yields

$$\Delta_1 = O(E_L(b, n) + n^{-1}b + E_S(b, n)) = O(E_S(b, n)),$$

as required. Under ULC(2), with $b \asymp n^\gamma$, all that is needed is to check that $E_S(b, n) = O(1/\log n)$; note here that, under ULC(2), $F(x) = O(1/x)$. □

5. Theorems. We now prove the main results of the paper. The first approximation to the distribution of $K_{0n} = K_{0n}(C^{(n)})$ is derived by combining Corollary 2.3 and Theorems 3.4 and 4.3. The result is somewhat unwieldy, but has neater consequences. Let Q_{bn} denote the convolution of $\mathscr{L}(K_{0b}(Z))$ and R_{bn}, the latter defined as for Corollary 2.3.

LEMMA 5.1. *For any combinatorial structure satisfying* ULC(0), *if* $b = [n^\gamma]$ *for some fixed* $0 < \gamma < 1$, *then*

$$d_{TV}(\mathscr{L}(K_{0n}(C^{(n)})), Q_{bn}) = O(\lambda_{bn}^{-3/2} + E_S([n^\gamma], n)),$$

where $\lambda_{bn} = h_n - h_b$ *as before.*

PROOF. Writing $p_{kt}(X) = \mathbb{P}[K_{0b}(X) = k, T_{0b}(X) = t]$, and suppressing the superscript $^{(n)}$, direct calculation shows that

$$\begin{aligned}\Delta_2 &= 2d_{TV}(\mathscr{L}(K_{0n}(C)), Q_{bn})\\ &\le \sum_{k\ge 0}\sum_{t\ge 0}\sum_{s\ge 1}\Big|\mathbb{P}[K_{0b}(C) = k, T_{0b}(C) = t, K_{bn}(C) = s]\\ &\qquad\qquad -\mathbb{P}[K_{0b}(Z) = k, T_{0b}(Z) = t]R_{bn}\{s\}\Big|\\ &\le \sum_{k\ge 0}\sum_{t\ge 0} p_{kt}(C)\sum_{s\ge 1}|\mathbb{P}[K_{bn}(C) = s \mid K_{0b}(C) = k, T_{0b}(C) = t] - R_{bn}\{s\}|\\ &\quad + \sum_{k\ge 0}\sum_{t\ge 0}|p_{kt}(C) - p_{kt}(Z)|.\end{aligned}$$

Now the latter sum is just $2d_{TV}\big(\mathscr{L}(K_{0b}(C), T_{0b}(C)), \mathscr{L}(K_{0b}(Z), T_{0b}(Z))\big)$, which is bounded by $O(E_S(b, n))$ from Theorem 4.3. Furthermore, by the Conditioning Relation, as for (2.1),

$$\mathbb{P}[K_{bn}(C) = s \mid K_{0b}(C) = k, T_{0b}(C) = t] = \mathbb{P}[K_{bn}(Z) = s \mid T_{bn}(Z) = n - t].$$

Hence we have reached the estimate

$$\begin{aligned}\Delta_2 \le& \sum_{k\ge 0}\sum_{t\ge 0} p_{kt}(C)\\ &\quad\times\big\{2d_{TV}\big(\mathscr{L}(K_{bn}(Z) \mid T_{bn}(Z) = n - t), \mathscr{L}(K_{bn}(Z^*) \mid T_{bn}(Z^*) = n - t)\big)\\ &\qquad\qquad + 2d_{TV}\big(\mathscr{L}(K_{bn}(Z^*) \mid T_{bn}(Z^*) = n - t), R_{bn}\big)\big\}\\ &\quad + O(E_S(b, n)).\end{aligned}$$

For $0 \le t \le n/2$, the second of these differences is uniformly of order $O(\lambda_{bn}^{-3/2} + (t/n)^{\bar\theta/2})$, by Corollary 2.3; hence, from Corollary 3.3, its contribution is of order $O(\lambda_{bn}^{-3/2} + (b/n)^{\bar\theta}) = O(\lambda_{bn}^{-3/2})$. In the same range of t, the first of the distances is bounded by $O(E_L(b, n))$, from Theorem 3.4. Then, finally,

$$\begin{aligned}\sum_{k\ge 0}\sum_{t> n/2} p_{kt}(C) &= \mathbb{P}[T_{0b}(C) > n/2]\\ &\le n^{-1}\mathbb{E}T_{0b}(Z) + O(E_S(b, n)) = O(n^{-1}b + E_S(b, n)),\end{aligned}$$

by Theorem 4.3 and (4.4). This completes the proof of the lemma. □

THEOREM 5.2. *For any combinatorial structure satisfying* ULC(0), *we have*

$$d_{TV}\big(\mathscr{L}(K_{0n}(C^{(n)})), \mathscr{L}(K_{0,[\alpha n]}(Z))\big) = O(\log^{-1} n + E_S([n^\gamma], n)), \tag{5.1}$$

for any fixed $0 < \gamma < 1$, *where* $\alpha = \alpha_\theta = \exp\{\theta^{-1} - h_\theta\}$ *and* $E_S(b, n)$ *is as defined in* (4.8).

PROOF. By Lemma 3.1, and because

$$\left|\sum_{i=n}^{[\alpha n]} i^{-1} e(i)\right| = O(e(n \min\{\alpha, 1\})) = O(E_L(b, n))$$

for $b \le n \min\{\alpha, 1/4\}$, it follows that

$$d_{TV}\big(\mathscr{L}(K_{b,[\alpha n]}(Z)), \mathrm{Po}\,(\theta(h_{[\alpha n]} - h_b))\big) = O(E_L(b, n)). \tag{5.2}$$

But, from the definition of α, we have

$$\begin{aligned}\theta(h_{[\alpha n]} - h_b) &= \theta(h_{[\alpha n]} - h_n) + \theta(h_n - h_b)\\ &= \theta(\log \alpha + O(n^{-1})) + \theta\lambda_{bn} = \rho_{bn} + 1 + O(n^{-1}),\end{aligned}$$

uniformly in b, where ρ_{bn} is as in (2.9), and thus

$$d_{TV}\big(\mathrm{Po}\,(\theta(h_{[\alpha n]} - h_b)), \mathrm{Po}\,(\rho_{bn} + 1)\big) = O(n^{-1}) = O(E_L(b, n)). \tag{5.3}$$

Finally, an application of Stein's method for the Poisson distribution shows that $d_{TV}(\mathrm{Po}\,(\rho_{bn}+1), 1+\mathrm{Po}\,(\rho_{bn})) \le \rho_{bn}^{-1}$, and $d_{TV}(1+\mathrm{Po}\,(\rho_{bn}), R_{bn}) = O(\rho_{bn}^{-1})$ follows directly from the definition of R_{bn} preceding Corollary 2.3.

Combining (5.2) and (5.3) with Lemma 5.1, the estimate (5.1) is established. □

COROLLARY 5.3. *If condition* ULC(1) *holds, then*

$$d_{TV}(\mathscr{L}(K_{0n}(C^{(n)})), \mathscr{L}(K_{0,[\alpha n]}(Z))) = O(1/\log n).$$

PROOF. Simply note that $E_S([n^\gamma], n) = O(1/\log n)$ under ULC(1), because $e(i) = O((\log i)^{-2})$. □

We finish by showing when a Poisson perturbation approximation is appropriate. Here we make use of a version of Stein's method given in Barbour and Xia [9] for some simple signed compound Poisson measures and other related distributions, including the measures $\nu(\rho, c_1, c_2)$ introduced in Section 2. Approximation by signed Poisson and compound Poisson measures has been extensively studied in the classical settings, for instance in Kruopis [24] and Čekanavičius [12], who have shown that very accurate approximations can be obtained. Define $\rho_{bn} = \theta(h_n - h_b - h_\theta) = \theta\lambda_{bn} - \theta h_\theta$ as before in (2.9), and set

$$\mu_1 = \sum_{i=1}^{b} \mathbb{E} Z_i; \quad \sigma_1^2 = \sum_{i=1}^{b} \mathrm{Var}\, Z_i;$$

$$\mu_2 = \sum_{i=b+1}^{n} \mathbb{E} Z_i - \theta h_\theta; \quad \sigma_2^2 = \sum_{i=b+1}^{n} \mathrm{Var}\, Z_i - \theta h_\theta - \theta^2 h'_\theta.$$

Note that, under ULC(2),

$$|\rho_{bn} - \mu_2| = O(\log^{-1} b) \quad \text{and} \quad |\rho_{bn} - \theta^2 h'_\theta - \sigma_2^2| = O(\log^{-1} b). \tag{5.4}$$

Let

$$\tau_n = \mu_1 + \mu_2 = \sum_{i=1}^{n} \mathbb{E} Z_i - \theta h_\theta;$$

$$a_n = \sigma_1^2 + \sigma_2^2 - (\mu_1 + \mu_2) = \sum_{i=1}^{n} (\mathrm{Var}\, Z_i - \mathbb{E} Z_i) - \theta^2 h'_\theta. \tag{5.5}$$

Let ν_n be any probability distribution which satisfies

$$\nu_n\{s+1\} = \mathrm{Po}\,(\tau_n)\{s\}\left(1 + \tfrac{1}{2} a_n \tau_n^{-2}\{(s - \tau_n)^2 - \tau_n\}\right) \tag{5.6}$$

for $\tau_n \max\{1/2, (1 - |a_n|^{-1/2})\} \le s \le \tau_n \min\{3/2, (1 + |a_n|^{-1/2})\}$.

THEOREM 5.4. *For any combinatorial structure satisfying* ULC(3),

$$d_{TV}(\mathscr{L}(K_{0n}(C^{(n)})), \nu_n) = O(\log^{-3/2} n),$$

for ν_n defined through (5.5) *and* (5.6).

PROOF. Take $b = n^{1/2}$, and use Lemma 5.1. Then we just need to show that $\|\mathscr{L}(S) - \nu_n\| = O(\log^{-3/2} n)$, where $S = S_1 + S_2$ and S_1 and S_2 are independent, with $\mathscr{L}(S_1) = \mathscr{L}(K_{0b}(Z))$ and $S_2 \sim R_{bn}$.

Now, from the definition (2.10) of R_{bn},

$$\|\mathscr{L}(S_2 - 1) - \nu(\rho_{bn}, 0, -\theta^2 h'_\theta)\| = O(\log^{-3/2} n),$$

since $\rho_{bn} \asymp \log n$. Then, from Barbour and Xia [9], Theorem 3.2,

$$\|\nu(\rho_{bn}, 0, -\theta^2 h'_\theta) - \pi_{u_2, v_2}\| = O(\log^{-3/2} n),$$

where $u_2 = \rho_{bn} + \theta^2 h'_\theta$, $v_2 = -\theta^2 h'_\theta$ and $\pi_{u,v}$ is the (possibly signed) compound Poisson measure on $\mathbb{Z}_+$ with generating function

$$\hat{\pi}_{u,v}(z) := \sum_{r \ge 0} z^r \pi_{u,v}\{r\} = \exp\{u(z-1) + \tfrac{1}{2} v(z^2 - 1)\}. \tag{5.7}$$

On the other hand, from Barbour and Xia ([9], Corollary 4.4 and Proposition 4.6),

$$\|\mathscr{L}(S_1) - \pi_{u_1, v_1}\| = O(\log^{-3/2} n),$$

where $u_1 = 2\mu_1 - \sigma_1^2 \asymp \log n$ and $v_1 = \sigma_1^2 - \mu_1 = O(1)$; the condition ULC(3) is needed here as a prerequisite for applying these results. Since $\pi_{u_1, v_1} * \pi_{u_2, v_2} = \pi_{u_1 + u_2, v_1 + v_2}$, it thus follows that

$$\|\mathscr{L}(S_1 + S_2 - 1) - \pi_{u_1 + u_2, v_1 + v_2}\| = O(\log^{-3/2} n). \tag{5.8}$$

But now

$$\begin{aligned} u_1 + u_2 &= 2\mu_1 - \sigma_1^2 + \rho_{bn} + \theta^2 h'_\theta \\ &= 2(\mu_1 + \mu_2) - (\sigma_1^2 + \sigma_2^2) + O(\log^{-1} n) = \tau_n - a_n + O(\log^{-1} n) \end{aligned}$$

and

$$\begin{aligned} v_1 + v_2 &= \sigma_1^2 - \mu_1 - \theta^2 h'_\theta \\ &= (\sigma_1^2 + \sigma_2^2) - (\mu_1 + \mu_2) + O(\log^{-1} n) = a_n + O(\log^{-1} n), \end{aligned}$$

because of (5.4). Applying Barbour and Xia [9], Corollary 2.4, we thus have

$$\|\pi_{u_1+u_2, v_1+v_2} - \pi_{\tau_n - a_n, a_n}\| = O(\log^{-3/2} n),$$

and, again from Barbour and Xia ([9], Theorem 3.2),

$$\|\pi_{\tau_n - a_n, a_n} - \nu(\tau_n, 0, a_n)\| = O(\log^{-3/2} n).$$

Hence, from (5.8),

$$\|\mathscr{L}(S_1 + S_2 - 1) - \nu(\tau_n, 0, a_n)\| = O(\log^{-3/2} n),$$

and the theorem follows. □

COROLLARY 5.5. *For any combinatorial structure satisfying* ULC(3),

$$d_{TV}(\mathscr{L}(K_{0n}(C^{(n)})), 1 + \mathrm{Po}\,(\tau_n)) = \frac{|a_n|}{\sqrt{2\pi e}} + O(\log^{-3/2} n),$$

for τ_n and a_n defined in (5.5).

PROOF. The calculation based on the explicit expression for the density of ν_n is standard. □

REMARK. Theorem 3.4 of Barbour and Xia [9] can be used to prove more accurate approximations by shifted Poisson distributions. In particular, if a_n is an integer, then

$$d_{TV}\left(\mathscr{L}(K_{0n}(C^{(n)})), 1 - a_n + \mathrm{Po}\left(\sum_{i=1}^n \mathrm{Var}\, Z_i - \theta h_\theta - \theta^2 h'_\theta\right)\right) = O(\log^{-3/2} n).$$

Many other variations of the results above can be achieved, under slightly different conditions, by using similar methods. To take one example, if ULC(2) is satisfied, then

$$d_{TV}\left(\mathscr{L}(K_{0n}(C^{(n)})), \mathrm{Po}\left(\sum_{i=1}^n \mathbb{E} Z_i + 1 - \theta h_\theta\right)\right) = O(\log^{-1} n).$$

As observed in the introduction, ULC(r) holds for all assemblies, multisets and selections for which $|i\mathbb{P}[Z_i = 1] - \theta| = O((\log i)^{-2})$, for any value of $r > 1$. For these combinatorial classes, Hwang [20] needs the assumption that

$|i\mathbb{P}[Z_i = 1] - \theta| = O(i^{-\beta})$ for some $\beta > 0$, which, in the context of approximations with errors of logarithmic magnitude, is substantially stronger. It is not clear what assumptions about the distributions of the Z_i are required in general for his methods to be applicable, in the sense that the appropriate bivariate generating function belongs to his class $\mathscr{CM}$.

6. Proof of Lemma 2.1. In order to prove Lemma 2.1, we begin with Fourier inversion and integration by parts, to obtain

$$\begin{aligned} & n\mathrm{Po}\,(\theta\lambda_{bn})\{s+1\}\mathbb{P}[W_{s+1} = l] \\ & \quad = \mathrm{Po}\,(\theta\lambda_{bn})\{s+1\}\frac{1}{2\pi}\int_{-n\pi}^{n\pi} e^{-itl/n}\left\{\lambda_{bn}^{-1}\sum_{r=b+1}^{n} r^{-1}e^{itr/n}\right\}^{s+1} dt \qquad (6.1) \\ & \quad = \mathrm{Po}\,(\theta\lambda_{bn})\{s\}\frac{n\theta}{2\pi l}\int_{-n\pi}^{n\pi} e^{-itl/n}\left\{\lambda_{bn}^{-1}\sum_{r=b+1}^{n} r^{-1}e^{itr/n}\right\}^{s} V_{bn}(t)\,dt, \end{aligned}$$

where

$$V_{bn}(t) = \frac{e^{it/n}(e^{itb/n} - e^{it})}{n(1 - e^{it/n})};$$

note that

$$|V_{bn}(t)| \le \frac{\pi}{2}\min(1, 2|t|^{-1}) \le \frac{3\pi}{2(|t|+1)} \qquad (6.2)$$

in $|t| \le n\pi$, since

$$(2/\pi^2)x^2 \le 1 - \cos x = \tfrac{1}{2}|1 - e^{ix}|^2 \le \tfrac{1}{2}x^2 \qquad (6.3)$$

in $0 \le x \le \pi$. Similarly,

$$\begin{aligned} & n\mathbb{P}[T_{bn}(Z^*) = l] \\ & \quad = \frac{n\theta}{2\pi l}\int_{-n\pi}^{n\pi} e^{-itl/n}\exp\left\{-\theta\sum_{r=b+1}^{n} r^{-1}(1 - e^{itr/n})\right\} V_{bn}(t)\,dt. \qquad (6.4) \end{aligned}$$

We now approximate the integrals in (6.2) and (6.4) in a succession of lemmas. The argument is essentially straightforward, and rather tedious. We take $b = [n^\gamma]$ for some fixed γ, $0 < \gamma < 1$, and use the notation $O(E_M(n))$ to denote any quantity of order $n^{-\beta}$ for some fixed $\beta > 0$; this is very much smaller than the order $\lambda_{bn}^{-3/2} \sim \log^{-3/2} n$ to which we are working. Constants denoted by k with suffices depend only on the quantities θ, γ, α_1 and α_2; α_1 and α_2 are fixed, and satisfy $1/2 < \alpha_1 < 1 < \alpha_2 < 3/2$.

LEMMA 6.1. *For* $|t| \le n\pi$, *we have*

$$\left|\sum_{r=b+1}^{n} r^{-1}e^{itr/n}\right| \le (3\pi/2) + e^{-1} + \log^+(n/b|t|).$$

PROOF. Without loss of generality, take $t > 0$. Representing the sum as a complex integral, and choosing a ray as contour, we have

$$\left|\sum_{r=b+1}^{n} r^{-1}e^{itr/n}\right| = \left|\int_0^{e^{it/n}} \frac{u^b(u^{n-b}-1)}{u-1}\,du\right|$$
$$\leq \int_0^1 \frac{y^b|1-y^{n-b}e^{i(n-b)t/n}|}{|1-ye^{it/n}|}\,dy$$
$$= \int_0^1 y^b \left\{\frac{(1-y^{n-b})^2 + 2y^{n-b}(1-\cos[(n-b)t/n])}{(1-y)^2 + 2y(1-\cos[t/n])}\right\}^{1/2} dy.$$

For $n \leq |t| \leq n\pi$,

$$\{(1-y)^2 + 2y(1-\cos[t/n])\}^{1/2} \geq \{(1-y)^2 + 2y(1-\cos 1)\}^{1/2} \geq \sin 1,$$

and thus the integral is at most $2/\sin 1 \leq 3\pi/2$. For $|t| \leq n$, the integral is bounded piecemeal, since

$$\int_0^1 y^{b+(n-b-1)/2}\left\{\frac{1-\cos[(n-b)t/n]}{1-\cos[t/n]}\right\}^{1/2} dy \leq \int_0^1 y^{(n+b-1)/2}\pi(n-b)/2\,dy \leq \pi;$$

and then, from (6.3), since $1 \leq b \leq n$, we have

$$\int_{1-t/n}^1 \frac{y^b - y^n}{\sqrt{2y(1-\cos[t/n])}}\,dy \leq \frac{t}{n}\frac{1}{\sqrt{2(1-\cos[t/n])}} \leq \pi/2;$$

and finally

$$\int_0^{1-t/n}\left(\frac{y^b-y^n}{1-y}\right) dy = \sum_{r=b+1}^{n} r^{-1}(1-t/n)^r$$
$$\leq \int_b^n x^{-1}e^{-tx/n}\,dx$$
$$= \int_{tb/n}^t z^{-1}e^{-z}\,dz \leq e^{-1} + \log^+(n/bt). \qquad \square$$

COROLLARY 6.2. *If* $\alpha_1\theta\lambda_{bn} \leq s \leq \alpha_2\theta\lambda_{bn}$ *and* $\lambda_{bn} \geq 6((3\pi+1)/2 + e^{-1})$, *then*

$$I_{6.2} := \left|\int_{(n/b)^{1/2}\leq|t|\leq n\pi} e^{-itl/n}\left\{\lambda_{bn}^{-1}\sum_{r=b+1}^{n} r^{-1}e^{itr/n}\right\}^s V_{bn}(t)\,dt\right| = O(E_M(n)).$$

PROOF. From Lemma 6.1, if $\lambda_{bn} \geq 6((3\pi+1)/2 + e^{-1})$, then it follows that

$$\left|\lambda_{bn}^{-1}\sum_{r=b+1}^{n} r^{-1}e^{itr/n}\right|^s \leq (2/3)^{\alpha_1\theta\lambda_{bn}}$$

in $(n/b)^{1/2} \leq |t| \leq n\pi$ and in the given range of s, because then

$$\log^+(n/b|t|) \leq \tfrac{1}{2}\log(n/b) \leq \tfrac{1}{2}(\lambda_{bn}+1),$$

and hence the quantity within the moduli is at most 2/3. Using the bound on $V_{bn}(t)$ given in (6.2), the corollary follows. □

Now, for any $0 \le \varepsilon < 1$, define the complex valued function

$$G_\varepsilon(t) = \int_{\varepsilon t}^{t} \left(\frac{1-e^{iy}}{y}\right) dy = \int_{\varepsilon|t|}^{|t|} \left(\frac{1-\cos y}{y}\right) dy - i \,\mathrm{sign}(t) \int_{\varepsilon|t|}^{|t|} \frac{\sin y}{y}\, dy,$$

for $t \in \mathbb{R}$, noting also that $G_\varepsilon(-t) = \overline{G}_\varepsilon(t)$ and that $\mathscr{R}e\, G_\varepsilon(t) \ge 0$ for all t. The following property of G_ε that we use in what follows can be derived from known properties of the sine and cosine integrals ([1], pp 231–33):

$$G_* = \sup_{0\le\varepsilon<1} \sup_t |G_\varepsilon(t) - \min\{\log(|t|+1), \log(\varepsilon^{-1})\}| < \infty. \tag{6.5}$$

LEMMA 6.3. *With G_ε defined as above,*

$$\left|\lambda_{bn}^{-1} \sum_{r=b+1}^{n} r^{-1} e^{itr/n} - \left(1 - \lambda_{bn}^{-1} G_{b/n}(t)\right)\right| \le n^{-1}|t|(1 + 2\lambda_{bn}^{-1}).$$

PROOF. It is enough to take $t \ge 0$, since the result for $t < 0$ then follows by conjugation. First, it is immediate that

$$\sum_{r=b+1}^{n} r^{-1}(1 - e^{itr/n})$$
$$= \int_{t(b+1)/n}^{t(n+1)/n} \left(\frac{1-e^{iy}}{y}\right) dy - \sum_{r=b+1}^{n} \int_{(rt/n)}^{(r+1)t/n} \left\{\left(\frac{1-e^{iy}}{y}\right) - \left(\frac{1-e^{itr/n}}{(rt/n)}\right)\right\} dy.$$

Hence, by routine calculation, it follows that

$$\begin{aligned}
&\left|\lambda_{bn}^{-1} \sum_{r=b+1}^{n} r^{-1} e^{itr/n} - \left(1 - \lambda_{bn}^{-1} G_{b/n}(t)\right)\right| \\
&\quad = \lambda_{bn}^{-1} \left|\sum_{r=b+1}^{n} r^{-1}(1 - e^{itr/n}) - G_{b/n}(t)\right| \\
&\quad \le \lambda_{bn}^{-1} \left\{\int_{b|t|/n}^{(b+1)|t|/n} |y^{-1}(1-e^{iy})|\, dy + \int_{|t|}^{(n+1)|t|/n} |y^{-1}(1-e^{iy})|\, dy\right\} \\
&\qquad + \lambda_{bn}^{-1} \sum_{r=b+1}^{n} \int_{(rt/n)}^{(r+1)t/n} \left|\{(rt/n) - y\} + y e^{itr/n} - (rt/n)e^{iy}\right| (rt/n)^{-2}\, dy \\
&\quad \le \frac{2t}{n\lambda_{bn}} \\
&\qquad + \lambda_{bn}^{-1} \sum_{r=b+1}^{n} \int_{(rt/n)}^{(r+1)t/n} \left\{|y - (rt/n)|\,|1 - e^{itr/n}| + (rt/n)|e^{itr/n} - e^{iy}|\right\} \\
&\qquad\qquad \times (rt/n)^{-2}\, dy \\
&\quad \le \frac{2t}{n\lambda_{bn}} + \lambda_{bn}^{-1} \sum_{r=b+1}^{n} \frac{t}{2nr},
\end{aligned}$$

where we have twice used the inequality $|e^{iu} - e^{iv}| \le |u - v|$ for real u and v, proving the lemma. □

COROLLARY 6.4. *For $s \in \mathbb{Z}_+$ such that $\alpha_1 \theta \lambda_{bn} \le s \le \alpha_2 \theta \lambda_{bn}$,*

$$I_{6.4} := \left| \int_{-\sqrt{n/b}}^{\sqrt{n/b}} e^{-itl/n} \left\{ \left(\lambda_{bn}^{-1} \sum_{r=b+1}^{n} r^{-1} e^{itr/n} \right)^s - \left(1 - \lambda_{bn}^{-1} G_{b/n}(t)\right)^s \right\} V_{bn}(t)\, dt \right|$$
$$= O(E_M(n)).$$

PROOF. Take

$$a = \left(1 - \lambda_{bn}^{-1} G_{b/n}(t)\right), \qquad a + \varepsilon = \lambda_{bn}^{-1} \sum_{r=b+1}^{n} r^{-1} e^{itr/n},$$

and observe that $|a + \varepsilon| \le \lambda_{bn}^{-1} \sum_{r=b+1}^{n} r^{-1} = 1$; using the inequality

$$|(a+\varepsilon)^s - a^s| \le s|\varepsilon| \max\{|a|^s, |a+\varepsilon|^s\} \le s|\varepsilon|(1+|\varepsilon|)^s \le s|\varepsilon| e^{s|\varepsilon|}$$

for any $s \in \mathbb{Z}_+$, it follows from Lemma 6.3, in view of the range of t, and from (6.2) that

$$I_{6.4} \le k_{6.4}\, s \int_{-\sqrt{n/b}}^{\sqrt{n/b}} n^{-1}|t|\, e^{s/\sqrt{nb}} \min(1, |t|^{-1})\, dt = O((nb)^{-1/2} \log(n/b)),$$

which is enough. □

LEMMA 6.5. *If $\lambda \ge 1$, $G \in \mathbb{C}$ and $s \in \mathbb{Z}_+$ satisfy $\lambda^{-1}|G| \le 9/16$ and $(1/2)\theta\lambda \le s \le (3/2)\theta\lambda$, then*

$$D_G(s) := |e^{\theta G}(1 - \lambda^{-1}G)^s - \{1 - \lambda^{-1}G(s - \lambda\theta) + \tfrac{1}{2}\lambda^{-2}G^2((s-\lambda\theta)^2 - \lambda\theta)\}|$$
$$\le k_{6.5}(|G|+1)^6 \left\{\lambda^{-2}(1 + |s - \lambda\theta|) + \lambda^{-3}|s - \lambda\theta|^3\right\} e^{59\theta|G|/64}.$$

PROOF. By Taylor's expansion, writing $w = s - \lambda\theta$, we have

$$\begin{aligned} D_G(s) &= \big|\exp\{\theta G + (\lambda\theta + w)\log(1 - \lambda^{-1}G)\} \\ &\qquad -1 + \lambda^{-1}Gw - \tfrac{1}{2}\lambda^{-2}G^2(w^2 - \lambda\theta)\big| \\ &= \big|\exp\{-\lambda^{-1}Gw - \tfrac{1}{2}\lambda^{-2}G^2(\lambda\theta + w) + \eta_1\} \\ &\qquad -1 + \lambda^{-1}Gw - \tfrac{1}{2}\lambda^{-2}G^2(w^2 - \lambda\theta)\big|, \end{aligned}$$

where $|\eta_1| \le k_1 \lambda^{-2}|G|^3$, because of the restrictions on w and $|G|$. Then we have

$$\begin{aligned} &\big|\exp\{-\lambda^{-1}Gw\} - \{1 - \lambda^{-1}Gw + \tfrac{1}{2}\lambda^{-2}G^2w^2\}\big| \\ &\quad \le \frac{|w|^3|G|^3}{6\lambda^3} \exp\{\lambda^{-1}|G|\,|w|\} \le \frac{|w|^3|G|^3}{6\lambda^3} \exp\{\theta|G|/2\}, \end{aligned}$$

and

$$\left|\exp\{-\tfrac{1}{2}\lambda^{-2}G^2(\lambda\theta+w)\} - \{1-\tfrac{1}{2}\lambda^{-2}G^2(\lambda\theta+w)\}\right|$$
$$\le k_2\lambda^{-2}|G|^4\exp\{27\theta|G|/64\},$$

again by Taylor's expansion and because of the restrictions on w and $|G|$. The lemma follows upon multiplying and collecting terms. □

COROLLARY 6.6. *For* $(1/2)\theta\lambda_{bn} \le s \le (3/2)\theta\lambda_{bn}$ *and* $\lambda_{bn} \ge 16(G_* + 1)$, *we have*

$$I_{6.6} := \left|\int_{-\sqrt{n/b}}^{\sqrt{n/b}} e^{-itl/n}\{(1-\lambda_{bn}^{-1}G_{b/n}(t))^s - W_{b/n}(s,t)\}V_{bn}(t)\,dt\right|$$
$$\le\ k_{6.6}\{\lambda_{bn}^{-2}(1+|s-\lambda_{bn}\theta|) + \lambda_{bn}^{-3}|s-\lambda_{bn}\theta|^3\},$$

where

$$W_\varepsilon(s,t) := 1 - \lambda_{bn}^{-1}G_\varepsilon(t)(s-\lambda_{bn}\theta) + \tfrac{1}{2}\lambda_{bn}^{-2}G_\varepsilon(t)^2\{(s-\lambda_{bn}\theta)^2 - \lambda_{bn}\theta\}e^{-\theta G_\varepsilon(t)}. \tag{6.6}$$

PROOF. We use Lemma 6.5 to bound the difference within the braces. In the given range of t, (6.5) shows that

$$\left|e^{-\theta G_{b/n}(t)}\right| \le e^{\theta G_*}(1+|t|)^{-\theta}; \qquad |G_{b/n}(t)| \le G_* + \log(|t|+1), \tag{6.7}$$

and (6.2) is used to bound $|V_{bn}(t)|$. Combining these observations, we find a bound for $I_{6.6}$ of

$$k_1\int_0^{\sqrt{n/b}} (t+1)^{-\theta}(1+\log(t+1))^6$$
$$\times\left\{\lambda_{bn}^{-2}(1+|s-\lambda_{bn}\theta|) + \lambda_{bn}^{-3}|s-\lambda_{bn}\theta|^3\right\}(t+1)^{59\theta/64}(t+1)^{-1}\,dt,$$

which is of the required order; note that $\lambda_{bn} \ge 16(G_*+1)$ is enough to ensure that $\lambda_{bn}^{-1}|G_{b/n}(t)| \le 9/16\ |t| \le (n/b)^{1/2}$. □

LEMMA 6.7. *For* $\alpha_1\theta\lambda_{bn} \le s \le \alpha_2\theta\lambda_{bn}$, *we have*

$$I_{6.7} := \left|\int_{-\sqrt{n/b}}^{\sqrt{n/b}} e^{-itl/n}W_{b/n}(s,t)\{V_{bn}(t) - iG'_{b/n}(t)\}\,dt\right| = O(E_M(n)).$$

PROOF. Elementary Taylor estimates using (6.2) show that, for $|t| \le n\pi$,

$$|V_{bn}(t) - iG'_{b/n}(t)| = \frac{|e^{itb/n} - e^{it}|}{n|t|\,|1-e^{it/n}|}\,|in(1-e^{it/n}) - te^{it/n}|$$
$$\le \frac{2}{n|t|\,|1-e^{it/n}|}\,|e^{-it/n} - (1-it/n)| \le \frac{\pi n}{t^2}\,\frac{t^2}{2n^2} \le \pi/(2n).$$

Hence, recalling (6.6) and (6.7), it follows that

$$I_{6.7} \le k_1 \int_0^{\sqrt{n/b}} (t+1)^{-\theta}\{1+\log^2(t+1)\}n^{-1}\,dt$$
$$= O(n^{-1}(n/b)^{(1-\theta)/2}\log^2(n/b)),$$

which is enough. Note also that, t and $b \le n$,

$$|G'_{b/n}(t)| = |t^{-1}(e^{itb/n} - e^{it})| \le \min\{1, 2|t|^{-1}\} \le 3/(|t|+1). \tag{6.8}$$ □

LEMMA 6.8. *For $\alpha_1\theta\lambda_{bn} \le s \le \alpha_2\theta\lambda_{bn}$ and $n/2 \le l \le n$, we have*

$$I_{6.8} := \left|\int_{-\sqrt{n/b}}^{\sqrt{n/b}} (e^{-it} - e^{-itl/n})W_{b/n}(s,t)G'_{b/n}(t)\,dt\right| \le k_{6.8}(1-l/n)^{\bar\theta/2},$$

where $\bar\theta = \min(1, \theta)$ as before.

PROOF. In the given range of s and for $|t| \le \sqrt{n/b}$, from (6.6), (6.7) and (6.8), we have

$$|W_{b/n}(s,t)G'_{b/n}(t)| \le k_1(1+|t|)^{-\theta-1}\{1+\log^2(|t|+1)\};$$

furthermore, for $n/2 \le l \le n$, we have

$$|e^{-it} - e^{-itl/n}| \le \min(2, t(1-l/n)).$$

Now integrate $|e^{-it} - e^{-itl/n}|\,|W_{b/n}(s,t)G'_{b/n}(t)|$ with respect to t. If $\theta > 1$, this gives an upper bound for $I_{6.8}$ of

$$k_2 \int_0^\infty (1+|t|)^{-\theta-1}|t|(1-l/n)\{1+\log^2(|t|+1)\}\,dt = O(1-l/n);$$

if $\theta \le 1$ and $(1-l/n)^{-1} \le (n/b)^{1/2}$, split the integral into two parts to give

$$I_{6.8} \le k_3\left\{\int_0^{n/(n-l)} |t|(1-l/n)(1+|t|)^{-\theta-1}\{1+\log^2(|t|+1)\}\,dt\right.$$
$$\left. + \int_{n/(n-l)}^{\sqrt{n/b}} (1+|t|)^{-\theta-1}\{1+\log^2(|t|+1)\}\,dt\right\}$$
$$\le k_4(1-l/n)^{\theta}\{1+\log^2(n/(n-l))\} \le k_5(1-l/n)^{\theta/2};$$

the remaining case is simpler. □

LEMMA 6.9. *For $\alpha_1\theta\lambda_{bn} \le s \le \alpha_2\theta\lambda_{bn}$, we have*

$$I_{6.9} := \left|\int_{-\sqrt{n/b}}^{\sqrt{n/b}} e^{-it}\{W_{b/n}(s,t)G'_{b/n}(t) - W_0(s,t)G'_0(t)\}\,dt\right| = O(E_M(n)).$$

PROOF. Since

$$G_0(t) - G_{b/n}(t) = \int_0^{tb/n} \left(\frac{1-e^{iy}}{y}\right) dy,$$

we find that $|G_0(t) - G_{b/n}(t)| \le |t|b/n$ and that $|G_0'(t) - G_{b/n}'(t)| \le b/n$, and the remaining argument is straightforward, in view of (6.6), (6.7) and (6.8). □

LEMMA 6.10. *For $\alpha_1\theta\lambda_{bn} \le s \le \alpha_2\theta\lambda_{bn}$, we have*

$$\int_{\sqrt{n/b}}^{\infty} |W_0(s,t)G_0'(t)|\, dt = O(E_M(n)).$$

PROOF. Uniformly for all t and in the given range of s, we have

$$|W_0(s,t)G_0'(t)| \le k_1(1+|t|)^{-1-\theta}(1+\log^2(|t|+1)),$$

and the conclusion is immediate. □

Returning to the proof of Lemma 2.1, observe that Corollaries 6.2, 6.4 and 6.6, together with Lemmas 6.7–6.10, can be combined with (6.2) to give

$$\begin{aligned}
&\frac{n\mathrm{Po}\,(\theta\lambda_{bn})\{s+1\}\mathbb{P}[W_{s+1}=l]}{\mathrm{Po}\,(\theta\lambda_{bn})\{s\}}\\
&\quad= \biggl(\frac{n\theta}{2\pi l}\int_{-\infty}^{\infty} e^{-it-\theta G_0(t)}\Bigl\{1-\lambda_{bn}^{-1}G_0(t)(s-\lambda_{bn}\theta)\\
&\qquad\qquad + \frac{1}{2}\lambda_{bn}^{-2}G_0^2(t)[(s-\lambda_{bn}\theta)^2-\lambda_{bn}\theta]\Bigr\}G_0'(t)\,dt\\
&\qquad + O\Bigl\{\lambda_{bn}^{-2}(1+|s-\lambda_{bn}\theta|)+\lambda_{bn}^{-3}|s-\lambda_{bn}\theta|^3+(1-l/n)^{\bar\theta/2}\Bigr\}\biggr),
\end{aligned} \tag{6.9}$$

uniformly in $\alpha_1\theta\lambda_{bn} \le s \le \alpha_2\theta\lambda_{bn}$ and $n/2 \le l \le n$. An entirely similar chain of reasoning, but without needing Corollary 6.6, yields

$$n\mathbb{P}[T_{bn}(Z^*)=l] = \frac{n\theta}{2\pi l}\int_{-\infty}^{\infty} e^{-it-\theta G_0(t)}G_0'(t)\,dt + O(E_M(n)+(1-l/n)^{\bar\theta/2}). \tag{6.10}$$

These integrals can now be explicitly evaluated.

Integrating by parts, we first obtain that

$$\frac{\theta}{2\pi}\int_{-\infty}^{\infty} ie^{-it-\theta G_0(t)}G_0'(t)\,dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-it-\theta G_0(t)}\,dt = f_\theta(1),$$

where f_θ is the probability density of the random variable X_θ given in (3.6), and satisfies $f_\theta(1) = \theta\mathbb{P}[X_\theta \le 1]$ and

$$\frac{d}{d\theta}\mathbb{P}[X_\theta \le 1] = -h_\theta\mathbb{P}[X_\theta \le 1], \tag{6.11}$$

where h_θ is as in (1.11). Similarly, we obtain

$$\begin{aligned}
\frac{\theta}{2\pi}\int_{-\infty}^{\infty} ie^{-it-\theta G_0(t)}G_0(t)G_0'(t)\,dt &= \frac{1}{2\pi\theta}\int_{-\infty}^{\infty} e^{-it-\theta G_0(t)}(1+\theta G_0(t))\,dt\\
&= \{\theta^{-1}f_\theta(1) - f_\theta'(1)\},
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\theta}{2\pi}\int_{-\infty}^{\infty} ie^{-it-\theta G_0(t)}G_0^2(t)G_0'(t)\,dt \\
&= \frac{1}{\pi\theta^2}\int_{-\infty}^{\infty} e^{-it-\theta G_0(t)}(1+\theta G_0(t)+\tfrac{1}{2}\theta^2 G_0^2(t))\,dt \\
&= 2\{\theta^{-2}f_\theta(1)-\theta^{-1}f_\theta'(1)+\tfrac{1}{2}f_\theta''(1)\}.
\end{aligned}$$

Combining these results with (6.9) and (6.10) gives

$$\begin{aligned}
&\frac{\mathrm{Po}\,(\theta\lambda_{bn})\{s+1\}\mathbb{P}[W_{s+1}=l]}{\mathrm{Po}\,(\theta\lambda_{bn})\{s\}\mathbb{P}[T_{bn}(Z^*)=l]} \\
&\quad = 1+\lambda_{bn}^{-1}c_1(s-\lambda_{bn}\theta)+\frac{1}{2}\lambda_{bn}^{-2}c_2\{(s-\lambda_{bn}\theta)^2-\lambda_{bn}\theta\} \\
&\quad\quad +O\left\{\lambda_{bn}^{-2}(1+|s-\lambda_{bn}\theta|)+\lambda_{bn}^{-3}|s-\lambda_{bn}\theta|^3+(1-l/n)^{\bar{\theta}/2}\right\},
\end{aligned}$$

uniformly in $\alpha_1\theta\lambda_{bn}\le s\le\alpha_2\theta\lambda_{bn}$ and $n/2\le l\le n$, with

$$c_1=-\{\theta^{-1}f_\theta(1)-f_\theta'(1)\}/f_\theta(1)=-h_\theta$$

and

$$c_2=\left(2\{\theta^{-2}f_\theta(1)-\theta^{-1}f_\theta'(1)\}+f_\theta''(1)\right)/f_\theta(1)=-h_\theta'+h_\theta^2,$$

these last calculations following from (6.11). This completes the proof of Lemma 2.1. □

Acknowledgment. We thank Jim Pitman for directing us to the work of Brenti.

REFERENCES

[1] ABRAMOWITZ, M. and STEGUN, I. A. (1965). *Handbook of Mathematical Functions*. Dover Press, New York.
[2] ALDOUS, D. J. (1985). *Exchangeability and Related Topics. Lecture Notes in Math.* **1117** 1–198. Springer, New York.
[3] ARRATIA, R., BARBOUR, A. D. and TAVARÉ, S. (1993). On random polynomials over a finite field. *Math. Proc. Camb. Phil. Soc.* **114** 347–368.
[4] ARRATIA, R., BARBOUR, A. D. and TAVARÉ, S. (2000). *Logarithmic Combinatorial Structures*. To appear.
[5] ARRATIA, R. BARBOUR, A. D. and TAVARÉ, S. (1999). On Poisson-Dirichlet limits for random decomposable combinatorial structures. *Combin. Probab. Comput.* **8** 193–208.
[6] ARRATIA, R., STARK, D. and TAVARÉ. S. (1995). Total variation asymptotics for Poisson process approximations of logarithmic combinatorial assemblies. *Ann. Probab.* **23** 1347–1388.
[7] ARRATIA, R. and TAVARÉ, S. (1994). Independent process approximations for random combinatorial structures. *Adv. Math.* **104** 90–154.

[8] BARBOUR, A. D. and HALL, P. G. (1984). On the rate of Poisson convergence. *Math. Proc. Camb. Phil. Soc.* **95** 473–480.
[9] BARBOUR, A. D. and XIA, A. (1999). Poisson perturbations. ESAIM: P&S **3** 131–150.
[10] BRENTI, F. (1989). *Unimodal, Log-concave, and Pólya Frequency Sequences in Combinatorics.* Amer. Math. Soc., Providence, RI.
[11] CAR, M. (1982). Factorization dans $\mathbf{F}_q(X)$. *C.R. Acad. Sci. Paris Ser. I* **294** 147–150.
[12] ČEKANAVIČIUS, V. (1997). Asymptotic expansions in the exponent: a compound Poisson approach. *Adv. Appl. Probab.* **29** 374–387.
[13] FELLER, W. (1945). The fundamental limit theorems in probability. *Bull. Amer. Math. Soc.* **51** 800–832.
[14] FLAJOLET, P. and SORIA, M. (1990). Gaussian limiting distributions for the number of components in combinatorial structures. *J. Combin. Theory Ser. A* **53** 165–182.
[15] GOH, W. M. Y. and SCHMUTZ, E. (1991). A central limit theorem on $GL_n(F_q)$. *Random Structures Algorithms* **2** 47–53.
[16] GONCHAROV, V. L. (1942). On the distribution of cycles in permutations. *Dokl. Akad. Nauk SSSR* **35** 299–301.
[17] GONCHAROV, V. L. (1944). Some facts from combinatorics. *Izvestia Akad. Nauk. SSSR, Ser. Mat.* **8** 3–48. [See also: On the field of combinatory analysis. *Trans. Amer. Math. Soc.* **19** 1–46.]
[18] HANSEN, J. C. (1993). Factorization in $GF(q^m)[x]$ and Brownian motion. *Combin. Probab. Comput.* **2** 285–299.
[19] HANSEN, J. C. (1994). Order statistics for decomposable combinatorial structures. *Random Structures Algorithms* **5** 517–533.
[20] HWANG, H.-K. (1999). Asymptotics of Poisson approximation to random discrete distributions: an analytic approach. *Adv. Appl. Probab.* **31** 448–491.
[21] KINGMAN, J. F. C. (1977). The population structure associated with the Ewens sampling formula. *Theoret. Population Biology* **11** 274–283.
[22] KOLCHIN, V. F. (1971). A problem of the allocation of particles in cells and cycles of random permutations. *Theory Probab. Appl.* **16** 74–90.
[23] KOLCHIN, V. F. (1976). A problem of the allocation of particles in cells and random mappings. *Theory Probab. Appl.* **21** 48–63.
[24] KRUOPIS, J. (1986). Precision of approximations of the generalized Binomial distribution by convolutions of Poisson measures. *Lithuanian Math. J.* **36** 37–49.
[25] MUTAFCIEV, L. R. (1988). Limit theorem concerning random mapping patterns. *Combinatorica* **8** 345–356.
[26] MUTAFCIEV, L. R. (1990). Large components and cycles in a random mapping pattern. In *Random Graphs '87* (M. Karoński, J. Jaworski and A. Ruciński, eds.) 189–202. Wiley, New York.
[27] PAVLOV, A. I. (1988). Local limit theorems for the number of components of random permutations and mappings. *Theory Probab. Appl.* **33** 183–187.
[28] RÉNYI, A. (1962). On the extreme elements of observations. *MTA III. Oszt. Közl.* **12** 105–121. [Reprinted (1976) *Selected papers of Alfred Rényi* **3** 50-65. Akadémiai Kiadó, Budapest.]
[29] STARK, D. (1994). Total variation distance for independent process approximations of random combinatorial objects. PhD dissertation, Univ. Southern California.
[30] STARK, D. (1997). Total variation asymptotics for independent process approximations of logarithmic multisets and selections. *Random Structures Algorithms* **11** 51–80.
[31] STEPANOV, V. E. (1969). Limit distributions for certain characteristics of random mappings. *Theory Probab. Appl.* **14** 612–626.
[32] VERSHIK, A. M. and SHMIDT, A. A. (1977). Limit measures arising in the theory of groups I. *Theory Probab. Appl.* **22** 79–85.
[33] VERVAAT, W. (1972). *Success Epochs in Bernoulli Trials with Applications in Number Theory*. *Math. Centre Tracts* **42**. Math. Centrum, Amsterdam.
[34] WATTERSON, G. A. (1974). The sampling theory of selectively neutral alleles. *Adv. Appl. Probab.* **6** 463–488.

R. Arratia
S. Tavaré
Department of Mathematics
University of Southern California
Los Angeles, California 90089-1113
E-mail: rarratia@hto.usc.edu
stavare@hto.usc.edu

A.D. Barbour
Abteilung für Angewandte Mathematik
Universität Zürich
Winterthurerstrasse 190
CH-8057, Zürich
Switzerland
E-mail: adb@amath.unizh.ch