Methods and Applications of Analysis
1 (2) 1994, pp. 119–142

ISSN 1073–2772

On the effective block size in Harper's theorem

William M. Y. Goh and Jet Wimp

ABSTRACT. Let s be a random set partition of $[n]$ and $X_n(\sigma)$ be the random variable marking the total number of blocks in σ. By employing a uniform probability on the sample space of random set partitions of $[n]$, L. Harper proved a central limit theorem for $X_n(\sigma)$.

We determine the *effective size* of the block size in Harper's theorem, that is, the minimal block size for which the conclusion of Harper's theorem is still maintained. This size is expressed as a quotient of the roots of two transcendental equations.

1. Introduction

The notion of effective size appears naturally in many problems in combinatorial enumeration. To illustrate the point we take the following example. Let σ be a random set partition of $[n]$, where $[n] = \{1, 2, \ldots, n\}$, the set of first n natural numbers. Let $X_n(\sigma)$ be the total number of blocks in σ. Also let u_n be the unique positive root of $ze^z = n$. Define the random variable

$$Y_n(\sigma) := \frac{X_n(\sigma) - M_n}{D_n}, \text{ where } M_n = \frac{n}{u_n} \text{ and } D_n = \sqrt{n/u_n^2}. \tag{1.1}$$

We assume a uniform probability on the sample space of random set partitions of $[n]$. The well-known theorem of Harper [4] states that

$$\operatorname{Prob}(Y_n \le x) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt \quad \text{as } n \to \infty. \tag{1.2}$$

In other words, we have a version of a central limit theorem. The "effective size" of the blocks is the optimal size $k = k(n)$, so that if we count only those blocks in a random set partition σ of sizes up to $k + L_n$ where L_n is an arbitrary function that goes to infinity as $n \to \infty$, call this count $X_n^{(k)}(\sigma)$, then we still obtain the conclusion of Harper's theorem,

$$\operatorname{Prob}\Big(\frac{X_n^{(k)} - M_n}{D_n} \le x\Big) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt. \tag{1.3}$$

To maintain asymptotic normality, it is not necessary to count the total number of blocks in a random set partition. We just have to count the number of blocks in σ of sizes up to $k + L_n$. We emphasize that in (1.2) and (1.3) the mean M_n and deviation

Received March 29, 1993, revised June 15, 1993.
1991 *Mathematics Subject Classification*: 41A60.
Key words and phrases: set partitions, asymptotics, central limit theorem, block size, Harper's Theorem.
The first author was supported by the National Science Foundation under Grant # DMS 9101753.

D_n are the same. Obviously k is not uniquely determined. Different k's may differ by a constant.

One can define the concept of effective size for other combinatorial configurations, for instance, the effective size of blocks in random ordered set partitions [2] or the effective size of primes in the Kac-Erdös theorem [5] on the distribution of $\omega(n)$. Both are interesting problems to study. This paper deals with the effective block size in a random set partition of $[n]$. Our major theorem is the following:

Theorem 1. *The effective size k of blocks in a random set partition of $[n]$ is $k = k(n) = u_n/r$ where r is the unique positive root of $z/2 - \ln z - 1 = 0$ in the unit interval* $(0, 1)$.

A rough estimate for r is $2.4 > 1/r > 2$. This means $1/r < e$. It is well-known that the mean of maximum block size of a random partition is $\sim e \ln n$, see [7]. Hence the effective size is smaller than the average maximum block size but the two are of the same asymptotic order. It is also well-known [6] that $u_n = \ln n - \ln\ln n + o(1)$. Therefore, the second-order term $(-1/r)\ln\ln n$ of the effective size cannot be dropped.

2. An analytical formulation of the problem

Our strategy for tackling the problem is straightforward. From the block index of a random set partition we have the following multivariate generating function (see Chapter 3 of [11]):

$$\sum_{n\geq 0} \frac{B(n)}{n!} E(t_1^{X_{n1}} t_2^{X_{n2}} \cdots t_n^{X_{nn}}) x^n = \exp\Big(\sum_{m\geq 1} \frac{t_m x^m}{m!}\Big), \tag{2.1}$$

where $X_{ni}(\sigma)$ is the number of blocks of size i in σ, $1 \leq i \leq n$. Let $E(X)$ denote the usual expectation of the random variable X, and $B(n)$ the n-th Bell number, i.e., the total number of set partitions of $[n]$. In (2.1), set $t_i = t$ for $1 \leq i \leq k$ and $t_i = 1$ for $i \geq k+1$. Thus

$$\sum_{n\geq 0} \frac{B(n)}{n!} E(t^{\overline{X}_{n,k}}) x^n = \exp\Big(t \sum_{1\leq m\leq k} \frac{x^m}{m!}\Big) \exp\Big(\sum_{m\geq k+1} \frac{x^m}{m!}\Big), \tag{2.2}$$

where the random variable $\overline{X}_{n,k} := \sum_{j=1}^{k} X_{n,j}$ marks the total number of blocks of sizes up to k.

Introducing $S_n(x) := \sum_{j=0}^{n} (x^j/j!)$, the n-th partial sum of e^x, and $R_{n+1}(x) := e^x - S_n(x)$, the remainder of e^x starting with the term x^{n+1}, we have

$$\sum_{n\geq 0} \frac{B(n)}{n!} E(t^{\overline{X}_{n,k}}) x^n = e^{t(S_k(x)-1)} e^{R_{k+1}(x)}.$$

By Cauchy's integral formula

$$E(t^{\overline{X}_{n,k}}) = \frac{n!}{B(n)} \frac{1}{2\pi i} \oint_C \frac{\exp\Big(t\big(S_k(x) - 1\big) + R_{k+1}(x)\Big)}{x^{n+1}}\, dx, \tag{2.3}$$

where C is any simple closed contour encircling the origin.

Now set $t = e^{\xi}$ in (2.3), ξ real,

$$E(e^{\xi \overline{X}_{n,k}}) = \frac{n!}{B(n)} \frac{1}{2\pi i} \oint_C \frac{\exp\left(e^{\xi}\left(S_k(x) - 1\right) + R_{k+1}(x)\right)}{x^{n+1}} \, dx.$$

We normalize $\overline{X}_{n,k}$ the same way as in Harper's theorem. Thus

$$E\left(\exp\left(\xi \frac{\overline{X}_{n,k} - M_n}{D_n}\right)\right) = e^{-\xi M_n / D_n} \frac{n!}{B(n)} \frac{1}{2\pi i} \oint_C \frac{\exp\left(e^{\xi/D_n}\left(S_k(x) - 1\right) + R_{k+1}(x)\right)}{x^{n+1}} \, dx. \quad (2.4)$$

Our goal is clear. We must find an optimal $k = k(n)$ so that the right-hand side of (2.4) approaches $e^{\xi^2/2}$ as $n \to \infty$. By the continuity theorem in probability [3] we then will have proved the theorem. The technique that is required to finish the proof is very involved. Basically, we apply the saddle point method to approximate the integral in (2.4). Unfortunately, the integrand is a complicated function of n and x. In such problems, uniformity of the approximation is always a major issue that must be confronted in the analysis.

3. Lemmas of approximation

This section collects some useful approximations that are relevant to the problem. Throughout this paper ξ is always held fixed. To simplify the matter, we need to define some notation. Let

$$I_{n,k} = \frac{1}{2\pi i} \oint_C \exp\left(g_{n,k}(x)\right) \frac{dx}{x}, \quad (3.1)$$

where

$$\begin{aligned} g_{n,k}(x) &= e^{\xi/D_n}\left(S_k(x) - 1\right) + R_{k+1}(x) - n \ln x \\ &= e^x - n \ln x + (e^{\xi/D_n} - 1) S_k(x) - e^{\xi/D_n}. \end{aligned}$$

A formal application of the saddle point method yields

$$g'_{n,k}(x) = e^x - \frac{n}{x} + (e^{\xi/D_n} - 1) S_{k-1}(x).$$

Proposition 1. *$g'_{n,k}(x)$ as a function of x defined on $(0, \infty)$ has a unique positive root $\rho_{n,k}$.*

Proof. Define

$$h_{n,k}(x) = x\left(e^x + (e^{\xi/D_n} - 1) S_{k-1}(x)\right) \quad (3.2)$$

on $(0, \infty)$. Differentiating with respect to x yields

$$h'_{n,k}(x) = \left(R_k(x) + e^{\xi/D_n} S_{k-1}(x)\right) + x\left(R_{k-1}(x) + e^{\xi/D_n} S_{k-2}(x)\right). \quad (3.3)$$

Since each term in the above is positive, $h'_{n,k}(x) > 0$ on $(0, \infty)$. Notice that $h_{n,k}(0) = 0$ and $h_{n,k}(\infty) = \infty$. This $h_{n,k}(x)$ is a continuous, strictly increasing function that maps $[0, \infty)$ to $[0, \infty)$. Hence, $h_{n,k}(x) = n$ has a unique positive root. □

Recall that u_n is the unique positive root of $ze^z = n$. The following proposition describes the asymptotic behavior of $\rho_{n,k}$.

Proposition 2.

$$\rho_{n,k} - u_n = -\frac{\xi}{D_n}\frac{S_{k-1}(\rho_{n,k})}{e^{\rho_{n,k}}} + \frac{\xi}{D_n u_n}\frac{S_{k-1}(\rho_{n,k})}{e^{\rho_{n,k}}} + O\Big(\frac{1}{D_n \ln^2 n}\Big), \tag{3.4}$$

uniformly for all positive integers k.

Proof. We first show that $\rho_{n,k} - u_n = o(1)$ as $n \to \infty$, uniformly for all $k \geq 1$.

(1) First we assume $\xi \geq 0$. Returning to (3.2), one can easily show that

$$xe^x \leq h_{n,k}(x) \leq h_{n,\infty}(x) = e^{\xi/D_n} xe^x.$$

Hence $u_n \geq \rho_{n,k} \geq \nu_n$ where ν_n is the unique positive root of

$$e^{\xi/D_n} xe^x = n. \tag{3.5}$$

Thus

$$|\rho_{n,k} - u_n| \leq |\nu_n - u_n|. \tag{3.6}$$

Observe that ν_n satisfies $\nu_n e^{\nu_n} = ne^{-\xi/D_n}$. Let u_t be the unique positive root of $ze^z = t$, $t > 0$. Then in terms of u_t we have

$$\nu_n = u_{ne^{-\xi/D_n}}. \tag{3.7}$$

Using the formula $u_t = \ln t - \ln\ln t + o(1)$, we get

$$\begin{aligned}\nu_n &= \ln(ne^{-\xi/D_n}) - \ln\ln(ne^{-\xi/D_n}) + o(1)\\ &= \ln n - \ln\ln n + o(1).\end{aligned} \tag{3.8}$$

This implies that

$$\nu_n - u_n = o(1). \tag{3.9}$$

Combining (3.9) with (3.6), we have

$$\rho_{n,k} - u_n = o(1), \quad \text{uniformly in } k. \tag{3.10}$$

(2) The case when $\xi < 0$ can be dealt with similarly.

Now we perform a "bootstrap" computation. Since $\rho_{n,k}$ is a root of $h_{n,k}(x) = n$, we have

$$n = \rho_{n,k}\big(e^{\rho_{n,k}} + (e^{\xi/D_n} - 1)S_{k-1}(\rho_{n,k})\big). \tag{3.11}$$

Also

$$n = u_n e^{u_n}. \tag{3.12}$$

Dividing (3.11) by (3.12) gives

$$1 = \frac{\rho_{n,k}}{u_n} e^{-u_n}\big(e^{\rho_{n,k}} + (e^{\xi/D_n} - 1)S_{k-1}(\rho_{n,k})\big).$$

Taking logarithms, we arrive at

$$0 = \ln\frac{\rho_{n,k}}{u_n} - u_n + \rho_{n,k} + \ln\big(1 + (e^{\xi/D_n} - 1)S_{k-1}(\rho_{n,k})e^{-\rho_{n,k}}\big). \tag{3.13}$$

Note that

$$\ln \frac{\rho_{n,k}}{u_n} = \ln\left(1 + \frac{\rho_{n,k} - u_n}{u_n}\right) = \frac{\rho_{n,k} - u_n}{u_n} + O\left(\left|\frac{\rho_{n,k} - u_n}{u_n}\right|^2\right),$$
$$\text{and} \quad 0 < \frac{S_{k-1}(\rho_{n,k})}{e^{\rho_{n,k}}} < 1.$$

Thus (3.13) simplifies to

$$0 = \frac{\rho_{n,k} - u_n}{u_n} + O\left(\left|\frac{\rho_{n,k} - u_n}{u_n}\right|^2\right) + (\rho_{n,k} - u_n)$$
$$+ (e^{\xi/D_n} - 1)\frac{S_{k-1}(\rho_{n,k})}{e^{\rho_{n,k}}} + O(|e^{\xi/D_n} - 1|^2)$$

or

$$0 = (\rho_{n,k} - u_n)\left(1 + \frac{1}{u_n} + O\left(\frac{1}{u_n^2}\right)\right) + \left(\frac{\xi}{D_n}\frac{S_{k-1}(\rho_{n,k})}{e^{\rho_{n,k}}} + O\left(\frac{1}{D_n^2}\right)\right), \quad (3.14)$$

uniformly for $k \geq 1$.

Solving (3.14) for $\rho_{n,k} - u_n$ gives

$$\rho_{n,k} - u_n = \left(\frac{-\xi}{D_n}\frac{S_{k-1}(\rho_{n,k})}{e^{\rho_{n,k}}} + O\left(\frac{1}{D_n^2}\right)\right)\left(1 + O\left(\frac{1}{u_n}\right)\right)$$
$$= \frac{-\xi}{D_n}\frac{S_{k-1}(\rho_{n,k})}{e^{\rho_{n,k}}} + O\left(\frac{1}{D_n \ln n}\right).$$

We can now write $\rho_{n,k} - u_n = (-\xi/D_n)(S_{k-1}(\rho_{n,k})/e^{\rho_{n,k}}) + \varepsilon_n$ with $\varepsilon_n = O(1/D_n \ln n)$. We plug this into (3.13) and bootstrap again. Finally, we get the desired result,

$$\varepsilon_n = \frac{\xi}{D_n u_n}\frac{S_{k-1}(\rho_{n,k})}{e^{\rho_{n,k}}} + O\left(\frac{1}{D_n \ln^2 n}\right). \quad \square$$

Proposition 3. (A)

$$\frac{S_k(kx)}{e^{kx}} = \delta + \sqrt{\frac{2}{\pi}}\frac{\zeta x}{x-1}\operatorname{Erfc}(\sqrt{k}\zeta)\left(1 + O\left(\frac{1}{\sqrt{k}}\right)\right), \quad \textit{where}$$
$$\delta := \begin{cases} 1, & 0 \leq x < 1 \\ 0, & x \geq 1 \end{cases}, \quad \textit{and } \zeta = |x - 1 - \ln x|^{1/2}, \quad (3.15)$$

uniformly for $x \geq 0$ *and*

$$\operatorname{Erfc}(\sigma) := \int_\sigma^\infty e^{-t^2}\,dt, \quad \sigma > 0. \quad (3.16)$$

(B) *Let* z *be in a compact set* K *of the open unit disk* $\{z \in \mathbb{C} : |z| < 1\}$. *Then we have*

$$\frac{S_k(kz)}{e^{kz}} = 1 - \frac{1}{\sqrt{2k\pi}}\frac{z}{1-z}(ze^{1-z})^k\left(1 + O\left(\frac{1}{k}\right)\right), \quad (3.17)$$

uniformly for all $z \in K$.

Proof. First we give the proof of (A). Recall that

$$S_n(x) := \sum_{k=0}^{n} \frac{x^k}{k!}.$$

We can write

$$\begin{aligned} S_n(x) &= \frac{x^{n+1}}{n!} \int_0^\infty e^{-xt}(1+t)^n \, dt \\ &= \frac{e^x}{n!} \int_x^\infty e^{-t} t^n \, dt. \end{aligned}$$

An integration by parts gives

$$S_n(x) = \frac{x^n}{n!} + e^x Q(n, x),$$

and

$$S_n(nx) = \frac{(nx)^n}{n!} + e^{nx} Q(n, nx),$$

in the notation of Temme [9]. The results in that reference give the asymptotic expansion,

$$S_n(nx) = \frac{(nx)^n}{n!} + e^{nx}\left\{ \frac{1}{\sqrt{\pi}} \operatorname{Erfc}[\varepsilon\sqrt{n}\zeta] + \frac{(e^{-x}xe)^n}{\sqrt{2\pi n}} \left[c_0 + \frac{c_1}{n} + \frac{c_2}{n^2} + \cdots \right] \right\},$$

where

$$\zeta := |x - 1 - \ln x|^{1/2}, \quad \varepsilon := \operatorname{sgn}(x-1), \quad c_0 = \frac{1}{x-1} - \frac{\varepsilon}{\sqrt{2}\zeta}.$$

This asymptotic expansion is *uniform* for $x \geq 0$. Although there may appear to be a problem near $x = 1$, there isn't, since,

$$c_0 = \frac{1}{3} + O[(x-1)].$$

Using Stirling's formula on the first term gives

$$\frac{S_n(nx)}{e^{nx}} = \frac{1}{\sqrt{\pi}} \operatorname{Erfc}[\varepsilon\sqrt{n}\zeta] + \frac{e^{-n\zeta^2}}{\sqrt{2\pi n}} \left(\frac{x}{x-1} - \frac{\varepsilon}{\sqrt{2}\zeta} \right) \left(1 + O\left(\frac{1}{n}\right) \right), \tag{3.18}$$

uniformly for $x \geq 0$. Now consider Gautschi's inequality [1, see 7.1.13],

$$\frac{1}{y + \sqrt{y^2 + 2}} < e^{y^2} \operatorname{Erfc} y \leq \frac{1}{y + \sqrt{y^2 + 4/\pi}}, \quad y \geq 0.$$

The usefulness of this inequality lies in its *uniformity.* We may write

$$e^{y^2} \operatorname{Erfc} y = \frac{1}{y + \sqrt{y^2 + a(y)}}, \quad \frac{4}{\pi} \leq a(y) < 2, \ y \geq 0.$$

For our purposes we rewrite the above result as

$$e^{-y^2} = \big(2y + b(y)\big) \operatorname{Erfc} y, \quad b(y) = \frac{a(y)}{y + \sqrt{y^2 + a(y)}}, \quad y \geq 0.$$

We see that

$$0 < b(y) \leq \sqrt{2}.$$

The second term may be written

$$\begin{aligned}
&\frac{e^{-n\zeta^2}}{\sqrt{2\pi n}}\Big(\frac{x}{x-1}-\frac{\varepsilon}{\sqrt{2}\zeta}\Big)\Big(1+O\Big(\frac{1}{n}\Big)\Big)\\
&\quad=\frac{1}{\sqrt{2\pi n}}\operatorname{Erfc}[\sqrt{n}\zeta]\Big(\frac{x}{x-1}-\frac{\varepsilon}{\sqrt{2}\zeta}\Big)(2\sqrt{n}\zeta+b(\sqrt{n}\zeta))\Big(1+O\Big(\frac{1}{n}\Big)\Big)\\
&\quad=\frac{1}{\sqrt{2\pi n}}\operatorname{Erfc}[\sqrt{n}\zeta]\Big(\frac{x}{x-1}-\frac{\varepsilon}{\sqrt{2}\zeta}\Big)\Big(\sqrt{\frac{2}{\pi}}\zeta+\frac{b(\sqrt{n}\zeta)}{\sqrt{2\pi n}}+O\Big(\frac{1}{n}\Big)\Big)\\
&\quad=\operatorname{Erfc}[\sqrt{n}\zeta]\Big\{\sqrt{\frac{2}{\pi}}\frac{\zeta x}{x-1}-\frac{\varepsilon}{\sqrt{\pi}}+O\Big(\frac{1}{\sqrt{n}}\Big)\Big\}. \qquad (3.19)
\end{aligned}$$

There are two cases to consider, $\varepsilon = 1$ and $\varepsilon = -1$. For the latter, we use the fact that

$$\operatorname{Erfc} y + \operatorname{Erfc}(-y) = \sqrt{\pi}.$$

Accounting for each case and putting (3.19) in (3.18) gives the final result,

$$\frac{S_n(nx)}{e^{nx}} = \delta + \sqrt{\frac{2}{\pi}}\frac{\zeta x}{x-1}\operatorname{Erfc}(\sqrt{n}\zeta)\Big(1+O\Big(\frac{1}{\sqrt{n}}\Big)\Big),$$

uniformly for $x \geq 0$, where

$$\delta := \begin{cases} 1, & 0 \leq x < 1;\\ 0, & x \geq 1.\end{cases}$$

Note that since

$$\zeta \doteq \frac{|x-1|}{\sqrt{2}}(1+O(x-1)),$$

the lead term above is continuous at $x = 1$, and yields, in fact, 1/2.

(B) follows from (A) when $z \geq 0$. The proof for a complex z can be found in [8], [10]. □

The following proposition is needed in Section 4.

Proposition 4.

$$\frac{S_k(kx)}{e^{kx}} \leq \frac{1}{2} + O\Big(\frac{1}{\sqrt{k}}\Big),$$

uniformly for $x \geq 1$.

Proof. Let $\zeta = |x - 1 - \ln x|^{1/2}$, $F_k(x) := x\operatorname{Erfc}(\sqrt{k}\zeta)$, and $H(x) := \sqrt{2/\pi}\,\zeta/(x-1)$. By Proposition 3, we have

$$\frac{S_k(kx)}{e^{kx}} = H(x)F_k(x)\Big(1+O\Big(\frac{1}{\sqrt{k}}\Big)\Big). \qquad (3.20)$$

It is easy to prove that $H(x)$ is a decreasing function on $[1, \infty)$. Our immediate goal is to prove that there exists a k_0 such that for all $k \geq k_0$, $F_k(x)$ is a decreasing function of x on $[1, \infty)$.

Now

$$F_k'(x) = \operatorname{Erfc}(\sqrt{k}\zeta) - \frac{\sqrt{k}}{2H(x)}e^{-k\zeta^2}. \qquad (3.21)$$

Note that

$$\operatorname{Erfc}(\sigma) := \int_\sigma^\infty e^{-t^2}\,dt \le e^{-\sigma^2} + e^{-(\sigma+1)^2} + e^{-(\sigma+2)^2} + \cdots$$
$$= e^{-\sigma^2}\sum_{j=0}^{\infty} e^{-2\sigma j - j^2} \le e^{-\sigma^2}\sum_{j=0}^{\infty} e^{-j^2}. \tag{3.22}$$

Since $H(x)$ is decreasing on $[1,\infty)$, using (3.21) and (3.22), we have

$$F_k'(x) \le e^{-k\zeta^2}\sum_{j=0}^{\infty} e^{-j^2} - \frac{\sqrt{k}}{2H(1)}e^{-k\zeta^2} < 0.$$

Thus there exists a k_0 such that for all $k \ge k_0$ $F_k(x)$ is strictly decreasing on $[1,\infty)$. This implies $H(x)F_k(x)$ is decreasing on $[1,\infty)$. Returning to (3.20), we see

$$\frac{S_k(kx)}{e^{kx}} \le H(1)F_k(1)\Big(1 + O\Big(\frac{1}{\sqrt{k}}\Big)\Big). \tag{3.23}$$

The proposition follows from observing that $H(1)F_k(1) = 1/2$. □

Recall the random variable $\overline{X}_{n,k} := \sum_{j=1}^{k} X_{n,j}$ which marks the total number of blocks of sizes up to k.

Proposition 5. *For the value of the expectation, we have*

$$E(\overline{X}_{n,k}) = S_k(u_n) + O\Big(\frac{S_k(u_n)}{u_n}\Big), \tag{3.24}$$

uniformly for $k \le \sqrt{2}u_n$.

Proof. Differentiating (2.3) with respect to t followed by setting $t = 1$, we have

$$E(\overline{X}_{n,k}) = \frac{n!}{eB(n)}\frac{1}{2\pi i}\oint_C \exp(e^x - n\ln x)\big(S_k(x) - 1\big)\frac{dx}{x}. \tag{3.25}$$

From this point on we follow deBruijn closely (see [6]). We shall use Szegö's approximation to tame the behavior of $S_k(x)$. Let $h(x) = e^x - n\ln x$. The saddle points are roots of $h'(x) = 0$, i.e., roots of $xe^x = n$. Thus u_n is a saddle point. The difficulty in proving the statement arises from non-uniformity and the unboundedness of $S_k(u_n)$. In the sequel, we shall show the contribution from all other saddle points is negligible. We may replace the integration contour C by a segment of the vertical line through u_n, and complete it to a closed contour by adding a large semi-circle. And if we make the radius R of the semi-circle tend to infinity, its contribution to the integral (3.25) tends to zero, the factor $S_k(x)/x^{n+1}$ being $O(R^{k-n-1})$ whereas $\exp e^x$ is bounded in the half-plane $\operatorname{Re} x \le u_n$. Therefore, the integral in (3.25) may be replaced by $\int_{u_n - i\infty}^{u_n + i\infty}$. Writing $x = u_n + iy$, we obtain

$$E(\overline{X}_{n,k}) = \frac{n!}{2\pi e B(n)}\exp(e^{u_n - n\ln u_n})\int_{-\infty}^{\infty}\exp(\psi(y))\Big(\frac{S_k(u_n+iy) - 1}{u_n + iy}\Big)\,dy, \tag{3.26}$$

where

$$\psi(y) = e^{u_n}\big((e^{iy} - 1) - u_n\ln(1 + iyu_n^{-1})\big).$$

Notice that $|\exp\psi(y)| = \exp(\operatorname{Re}\psi(y))$. We have to study

$$\operatorname{Re}\psi(y) = e^{u_n}\left(-1+\cos y - u_n \ln(1+y^2u_n^{-2})^{1/2}\right). \tag{3.27}$$

We now show that in (3.26) we can restrict ourselves essentially to the interval $[-\pi,\pi]$. First, we estimate the contribution for $y \geq \pi$. The situation when $y \leq -\pi$ can be dealt with similarly.

If $\pi < y < u_n$, then we have $\ln(1+y^2/u_n^2) > \frac{1}{2}y^2/u_n^2$ and $|S_k(u_n+iy)-1| < 2S_k(\sqrt{2}u_n)$ and therefore

$$\left|\int_\pi^{u_n} \exp\psi(y)\frac{S_k(u_n+iy)-1}{u_n+iy}\,dy\right| < \frac{2u_n}{|u_n+i\pi|}\exp\left(-\pi^2e^{u_n}/(4u_n)\right)S_k(\sqrt{2}u_n).$$

Note that $S_k(x) < e^x$ for all $x>0$ and all k. Thus

$$\begin{aligned}\left|\int_\pi^{u_n} \exp\psi(y)\frac{S_k(u_n+iy)-1}{u_n+iy}\,dy\right| &= O\Big(\exp\left(-\pi^2e^{u_n}/(4u_n)+\sqrt{2}u_n\right)\Big)\\ &= O\left(\exp(-e^{u_n}/u_n)\right).\end{aligned} \tag{3.28}$$

If $y > u_n$, we use $1+y^2/u_n^2 > 2y/u_n$. Putting $y = u_n x$, we get

$$\left|\int_{u_n}^\infty \exp\psi(y)\frac{S_k(u_n+iy)-1}{u_n+iy}\,dy\right| \leq 2\int_1^\infty \exp\left(\frac{-e^{u_n}u_n\ln(2x)}{2}\right)\frac{|S_k(u_n(1+ix))|}{(1+x^2)^{1/2}}\,dx. \tag{3.29}$$

We now use (A) in Proposition 3

$$\begin{aligned}|S_k(u_n(1+ix))| &\leq S_k(u_n\sqrt{1+x^2}) =\\ &\sqrt{\frac{2}{\pi}}e^{u_n\sqrt{1+x^2}}\frac{\zeta u_n\sqrt{1+x^2}}{u_n\sqrt{1+x^2}-k}\operatorname{Erfc}(\sqrt{k}\zeta)\left(1+O\left(\frac{1}{\sqrt{k}}\right)\right),\end{aligned} \tag{3.30}$$

where

$$\zeta = \left|\frac{u_n\sqrt{1+x^2}}{k}-1-\ln\frac{u_n\sqrt{1+x^2}}{k}\right|^{1/2}.$$

It is well-known that

$$\operatorname{Erfc}(\sigma) = \frac{1}{2}e^{-\sigma^2}\frac{1}{\sigma}\left(1+O\left(\frac{1}{\sigma^2}\right)\right) \quad \text{as } \sigma\to\infty.$$

Hence there exists an absolute constant A so that

$$\operatorname{Erfc}(\sigma) \leq \frac{A}{\sigma+1}e^{-\sigma^2} \quad \text{for all } \sigma > 0. \tag{3.31}$$

Putting $\sigma = \sqrt{k}\zeta$ in the above, we have

$$\operatorname{Erfc}(\sqrt{k}\zeta) \leq A\exp\left(-u_n\sqrt{1+x^2}+k+k\ln\frac{u_n\sqrt{1+x^2}}{k}\right)\frac{1}{1+\sqrt{k}\zeta}, \tag{3.32}$$

uniformly for $k \leq \sqrt{2}u_n$ and and $x \geq 1$.

Combining (3.29), (3.30), and (3.32) and using the estimate $\sqrt{1+x^2} \le 2x$ for all $x \ge 1$ we have

$$\left|\int_{u_n}^{\infty} \exp\psi(y)\frac{S_k(u_n+iy)-1}{u_n+iy}\,dy\right| = O\left(e^{k+k\ln(u_n/k)}\int_1^{\infty}\exp\left(\left(-\frac{1}{2}e^{u_n}u_n+k\right)\ln(2x)\right)dx\right). \quad (3.33)$$

It is easily seen that $\int_1^\infty (2x)^{-p}\,dx = O(e^{-p/2})$ $(p > 2)$ and $k\ln(u_n/k) \le u_n/e$, for all positive k. Therefore

$$\left|\int_{u_n}^{\infty} \exp\psi(y)\frac{S_k(u_n+iy)-1}{u_n+iy}\,dy\right| = O(e^{-n/5}). \quad (3.34)$$

Returning to (3.26) and making use of the fact [6] that

$$\frac{n!}{B(n)} = \frac{u_n^{n+1}\sqrt{2\pi(e^{u_n}+n/u_n^2)}}{\exp(e^{u_n}-1)}\left(1+O\left(\frac{1}{u_n}\right)\right), \quad (3.35)$$

we have

$$\begin{aligned} E(\overline{X}_{n,k}) &= \frac{u_n}{\sqrt{2\pi}}(e^{u_n}+n/u_n^2)^{1/2}\left(\int_{-\infty}^{\infty}\exp(\psi(y))\frac{S_k(u_n+iy)-1}{u_n+iy}\,dy\right)\left(1+O\left(\frac{1}{u_n}\right)\right) \\ &= \frac{u_n}{\sqrt{2\pi}}(e^{u_n}+n/u_n^2)^{1/2}\left(\int_{-\pi}^{\pi}\exp(\psi(y))\frac{S_k(u_n+iy)-1}{u_n+iy}\,dy\right)\left(1+O\left(\frac{1}{u_n}\right)\right), \end{aligned} \quad (3.36)$$

uniformly for $k \le \sqrt{2}u_n$.

Now consider the difference

$$\begin{aligned} D :&= \int_{-\pi}^{\pi}\exp(\psi(y))\frac{S_k(u_n+iy)-1}{u_n+iy}\,dy - \frac{S_k(u_n)}{u_n}\int_{-\pi}^{\pi}\exp(\psi(y))\,dy \\ &= \int_{-\pi}^{\pi}\exp(\psi(y))\frac{u_n(S_k(u_n+iy)-S_k(u_n))-iyS_k(u_n)}{u_n(u_n+iy)}\,dy. \end{aligned} \quad (3.37)$$

Take C_1 to be the contour $\{\, z : |z-u_n| = 2\pi \,\}$ and consider

$$\begin{aligned} S_k(u_n+iy)-S_k(u_n) &= \frac{1}{2\pi i}\oint_{C_1} S_k(z)\left(\frac{1}{z-(u_n+iy)}-\frac{1}{z-u_n}\right)dz \\ &= \frac{y}{2\pi}\oint_{C_1} S_k(z)\frac{dz}{(z-u_n)(z-u_n-iy)}. \end{aligned} \quad (3.38)$$

We have

$$\begin{aligned} |S_k(u_n+iy)-S_k(u_n)| &\le \frac{|y|}{\pi}\operatorname*{Max}_{z\in C_1}|S_k(z)| \\ &= \frac{|y|}{\pi}S_k(u_n+2\pi). \end{aligned} \quad (3.39)$$

Using (3.39) to estimate the difference D in (3.37), we have

$$\begin{aligned} D &\le \int_{-\pi}^{\pi} \exp\Big(\mathrm{Re}(\psi(y))\Big) \frac{u_n(|y|S_k(u_n+2\pi)/\pi) + |y|S_k(u_n)}{u_n(u_n-\pi)}\, dy \\ &\le \frac{2S_k(u_n+2\pi)}{\pi(u_n-\pi)} \int_{-\pi}^{\pi} \exp\Big(\mathrm{Re}(\psi(y))\Big)|y|\, dy \\ &= \frac{4S_k(u_n+2\pi)}{\pi(u_n-\pi)} \int_{0}^{\pi} \exp\Big(\mathrm{Re}(\psi(y))\Big) y\, dy. \end{aligned} \tag{3.40}$$

Recall from (3.27) that $\mathrm{Re}\,\psi(y) = e^{u_n}\big(-1+\cos y - u_n \ln(1+y^2u_n^{-2})^{1/2}\big)$. Hence

$$\begin{aligned} \int_0^{\pi} \exp\Big(\mathrm{Re}(\psi(y))\Big) y\, dy &\le \int_0^{\pi} \exp\big(e^{u_n}(-1+\cos y)\big) y\, dy \\ &= \int_0^{\pi/2} \exp\big(e^{u_n}(-1+\cos y)\big) y\, dy + \int_{\pi/2}^{\pi} (*)\, dy \\ &= \int_0^{\pi/2} \exp\big(e^{u_n}(-1+\cos y)\big) y\, dy + O\big(\exp(-e^{u_n})\big). \end{aligned} \tag{3.41}$$

We write

$$\begin{aligned} &\int_0^{\pi/2} \exp\big(e^{u_n}(-1+\cos y)\big) y\, dy \\ &\qquad = \int_0^{\pi/2} e^{u_n}(-\sin y) \exp\big(e^{u_n}(-1+\cos y)\big) \frac{y}{e^{u_n}(-\sin y)}\, dy. \end{aligned}$$

Integration by parts provides the estimate

$$\int_0^{\pi/2} \exp\big((e^{u_n})(-1+\cos y)\big) y\, dy = O(e^{-u_n}). \tag{3.42}$$

Combining (3.41) and (3.42) gives

$$\int_0^{\pi} \exp\big(\mathrm{Re}\,\psi(y)\big) y\, dy = O(e^{-u_n}). \tag{3.43}$$

By using (3.43), the integral in (3.40) may be estimated

$$D = O\bigg(\frac{S_k(u_n+2\pi)}{u_n} e^{-u_n}\bigg), \tag{3.44}$$

uniformly for $k \ge 1$.

We will require the following three equations:

$$\int_{-\pi}^{\pi} \exp\big(\psi(y)\big)\, dy = (2\pi e^{-u_n})^{1/2}\big(1+O(u_n^{-1})\big), \tag{3.45}$$

$$\int_{-\pi}^{\pi} \exp\big(\psi(y)\big) \frac{dy}{u_n+iy} = (2\pi e^{-u_n})^{1/2} u_n^{-1}\big(1+O(u_n^{-1})\big), \tag{3.46}$$

$$S_k(u_n+2\pi) \le S_k(u_n)\, e^{2\pi}, \tag{3.47}$$

where (3.45) can be found in [6] (see (6.2.4) in [6]), (3.46) can be established the same way as the previous result, and (3.47) can be proved by observing that

$$\ln S_k(u_n + 2\pi) - \ln S_k(u_n) = \int_{u_n}^{u_n+2\pi} \frac{S_{k-1}(x)}{S_k(x)}\,dx \le \int_{u_n}^{u_n+2\pi} 1\,dx = 2\pi. \tag{3.48}$$

Using this gives

$$D = O\left(\frac{S_k(u_n)}{u_n} e^{-u_n}\right), \tag{3.49}$$

uniformly for $k \ge 1$.

Putting (3.45), (3.46), and (3.49) into (3.36), we have

$$E(\overline{X}_{n,k}) = S_k(u_n) + O\left(\frac{S_k(u_n)}{u_n}\right),$$

uniformly for $k \le \sqrt{2}u_n$. □

The following proposition deals with the saddle point approximation to the integral $I_{n,k}$ in (3.1). The uniformity in k is the major concern. The usual saddle point method ignores the question of uniformity. Recall equation (3.1) and the definition of the saddle point $\rho_{n,k}$ (see Proposition 1).

Proposition 6.

$$I_{n,k} = \frac{\exp\left(e^{\xi/D_n}(S_k(\rho_{n,k}) - 1) + R_{k+1}(\rho_{n,k})\right)}{\rho_{n,k}^{n+1}\sqrt{2\pi g''_{n,k}(\rho_{n,k})}}\big(1 + o(1)\big), \tag{3.50}$$

uniformly for $\frac{1}{2}\ln n \le k \le \pi n^{5/12}$.

Proof. Decompose $I_{n,k}$ as follows:

$$I_{n,k} = \frac{1}{2\pi i}\oint_{|x|=1} \exp\big(g_{n,k}(x\rho_{n,k})\big)\frac{dx}{x} = I_1 + I_2, \tag{3.51}$$

where

$$I_1 = \frac{1}{2\pi i}\oint_{|\theta|\le\eta} \exp\big(g_{n,k}(x\rho_{n,k})\big)\frac{dx}{x},$$

and

$$I_2 = \frac{1}{2\pi i}\oint_{\eta<|\theta|\le\pi} \exp\big(g_{n,k}(x\rho_{n,k})\big)\frac{dx}{x}.$$

Here we choose $\eta = n^{-5/12}$. We shall show that I_1 gives the major contribution. First of all

$$I_1 = \frac{1}{2\pi}\int_{-\eta}^{\eta} \exp\big(g_{n,k}(e^{i\theta}\rho_{n,k})\big)\,d\theta. \tag{3.52}$$

The Taylor expansion of $g_{n,k}(z\rho_{n,k})$ at $z=1$ is

$$g_{n,k}(z\rho_{n,k}) = g_{n,k}(\rho_{n,k}) + g''_{n,k}(\rho_{n,k})\frac{\rho_{n,k}^2(z-1)^2}{2} + \frac{(z-1)^3}{2\pi i}\oint_{C_2}\frac{g_{n,k}(\zeta\rho_{n,k})}{(\zeta-z)(\zeta-1)^3}\,d\zeta,$$

where C_2 is an appropriate contour encircling 1, and z is in the interior of C_2. Letting $z=e^{i\theta}$ in the above gives

$$g_{n,k}(e^{i\theta}\rho_{n,k}) = g_{n,k}(\rho_{n,k}) - g''_{n,k}(\rho_{n,k})\frac{\rho_{n,k}^2\theta^2}{2} + R(\theta), \tag{3.53}$$

where

$$R(\theta) = \frac{g''_{n,k}(\rho_{n,k})\rho_{n,k}^2}{2}\left((e^{i\theta}-1)^2+\theta^2\right) + \frac{(e^{i\theta}-1)^3}{2\pi i}\oint_{C_2}\frac{g_{n,k}(\zeta\rho_{n,k})}{(\zeta-e^{i\theta})(\zeta-1)^3}\,d\zeta.$$

Substituting (3.53) into (3.52), we have

$$I_1 = \exp\bigl(g_{n,k}(\rho_{n,k})\bigr)\frac{1}{2\pi}\int_{-\eta}^{\eta}\exp\bigl(-g''_{n,k}(\rho_{n,k})\rho_{n,k}^2\theta^2/2\bigr)\exp\bigl(R(\theta)\bigr)\,d\theta. \tag{3.54}$$

To estimate $R(\theta)$, we use Proposition 2. Note that $g''_{n,k}(\rho_{n,k}) = e^{\rho_{n,k}} + n/\rho_{n,k}^2 + (e^{\xi/D_n}-1)S_{k-2}(\rho_{n,k})$. Thus

$$\left|\frac{g''_{n,k}(\rho_{n,k})\rho_{n,k}^2}{2}\left((e^{i\theta}-1)^2+\theta^2\right)\right| = O(\theta^3 n u_n) = O(n^{-1/4}u_n), \tag{3.55}$$

uniformly for $k\ge 1$. We now choose the contour $C_2=\{\zeta : |\zeta-1|=1/12\}$,

$$\left|\frac{(e^{i\theta}-1)^3}{2\pi i}\oint_{C_2}\frac{g_{n,k}(\zeta\rho_{n,k})}{(\zeta-e^{i\theta})(\zeta-1)^3}\,d\zeta\right| = O\left(\theta^3\oint_{|\zeta-1|=1/12}\frac{|g_{n,k}(\zeta\rho_{n,k})|}{|\zeta-e^{i\theta}|\,|\zeta-1|^3}\,|d\zeta|\right). \tag{3.56}$$

Since $g_{n,k}(\zeta\rho_{n,k}) = e^{\zeta\rho_{n,k}} - n\ln(\zeta\rho_{n,k}) + (e^{\xi/D_n}-1)S_k(\zeta\rho_{n,k}) - e^{\xi/D_n}$, on $|\zeta-1|=1/12$ we have

$$\begin{aligned}|g_{n,k}(\zeta\rho_{n,k})| &\le e^{13\rho_{n,k}/12} + O(n\ln\ln n) + O(e^{13\rho_{n,k}/12}/D_n)\\ &= O(e^{13u_n/12}) = O(n^{13/12}).\end{aligned} \tag{3.57}$$

Using (3.57), we have from (3.56) that

$$\left|\frac{(e^{i\theta}-1)^3}{2\pi i}\oint_{C_2}\frac{g_{n,k}(\zeta\rho_{n,k})}{(\zeta-e^{i\theta})(\zeta-1)^3}\,d\zeta\right| = O(n^{-5/4}n^{13/12}) = O(n^{-1/6}), \tag{3.58}$$

uniformly for $k\ge 1$.

Combining (3.55) and (3.58) gives

$$|R(\theta)| = O(n^{1/6}), \tag{3.59}$$

uniformly in k.

Returning to (3.54), we get

$$I_1 = \exp\bigl(g_{n,k}(\rho_{n,k})\bigr)\frac{1}{2\pi}\int_{-\eta}^{\eta}\exp\bigl(-g''_{n,k}(\rho_{n,k})\rho_{n,k}^2\theta^2/2\bigr)\,d\theta\,\bigl(1+O(n^{-1/6})\bigr). \tag{3.60}$$

Observe that $g''_{n,k}(\rho_{n,k})\rho_{n,k}^2 \sim e^{u_n}u_n^2$ uniformly in k and $\eta = n^{-5/12} > n^{-1/2}$. By the classical Laplace method, we obtain

$$\frac{1}{2\pi}\int_{-\eta}^{\eta} \exp\bigl(-g''_{n,k}(\rho_{n,k})\rho_{n,k}^2\theta^2/2\bigr)\,d\theta = \bigl(2\pi g''_{n,k}(\rho_{n,k})\rho_{n,k}^2\bigr)^{-1/2}\bigl(1+O(e^{-n^{1/6}/2})\bigr). \tag{3.61}$$

Substituting (3.61) into (3.60) gives

$$I_1 = \frac{\exp\Bigl(e^{\xi/D_n}\bigl(S_k(\rho_{n,k})-1\bigr)+R_{k+1}(\rho_{n,k})\Bigr)}{\rho_{n,k}^{n+1}\sqrt{2\pi g''_{n,k}(\rho_{n,k})}}\bigl(1+O(n^{-1/6})\bigr), \tag{3.62}$$

uniformly for $k \geq 1$.

Next, we shall show that I_2 is negligible compared with I_1. We have

$$I_2 = \frac{1}{2\pi}\int_{\eta<|\theta|\leq\pi} \exp\bigl(g_{n,k}(e^{i\theta}\rho_{n,k})\bigr)\,d\theta.$$

It follows that

$$\begin{aligned}|I_2| &\leq \frac{1}{2}\exp\bigl(g_{n,k}(\rho_{n,k})\bigr)\ \operatorname*{Sup}_{\eta<|\theta|\leq\pi}\left(\frac{\bigl|\exp\bigl(g_{n,k}(e^{i\theta}\rho_{n,k})\bigr)\bigr|}{\exp\bigl(g_{n,k}(\rho_{n,k})\bigr)}\right)\\ &= \frac{1}{2}\exp\bigl(g_{n,k}(\rho_{n,k})\bigr)\ \operatorname*{Sup}_{\eta<|\theta|\leq\pi}\ \frac{\exp\Bigl(\operatorname{Re}\bigl(g_{n,k}(e^{i\theta}\rho_{n,k})\bigr)\Bigr)}{\exp\bigl(g_{n,k}(\rho_{n,k})\bigr)}.\end{aligned} \tag{3.63}$$

Recall equation (3.1). We have

$$\begin{aligned}&\operatorname*{Sup}_{\eta<|\theta|\leq\pi}\ \frac{\exp\Bigl(\operatorname{Re}\bigl(g_{n,k}(e^{i\theta}\rho_{n,k})\bigr)\Bigr)}{\exp\bigl(g_{n,k}(\rho_{n,k})\bigr)}\\ &\qquad= \operatorname*{Sup}_{\eta<|\theta|\leq\pi}\ \exp\bigl\{e^{\xi/D_n}\bigl(\operatorname{Re} S_k(e^{i\theta}\rho_{n,k}) - S_k(\rho_{n,k})\bigr)\\ &\qquad\qquad + \bigl(\operatorname{Re} R_{k+1}(e^{i\theta}\rho_{n,k}) - R_{k+1}(\rho_{n,k})\bigr)\bigr\}\\ &\qquad\leq \operatorname*{Sup}_{\eta<|\theta|\leq\pi}\ \exp\bigl\{-e^{\xi/D_n}\bigl(S_k(\rho_{n,k}) - \operatorname{Re} S_k(e^{i\theta}\rho_{n,k})\bigr)\bigr\}\\ &\qquad\leq \exp\bigl\{-e^{\xi/D_n}\bigl(S_k(\rho_{n,k}) - \operatorname{Re} S_k(e^{i\eta}\rho_{n,k})\bigr)\bigr\}.\end{aligned} \tag{3.64}$$

To proceed further, we need some lower estimates:

$$S_k(\rho_{n,k}) - \operatorname{Re} S_k(e^{i\eta}\rho_{n,k}) = \sum_{j=0}^{k}\frac{1-\cos(jn^{-5/12})}{j!}\rho_{n,k}^j = 2\sum_{j=0}^{k}\frac{\sin^2\bigl(j/(2n^{5/12})\bigr)}{j!}\rho_{n,k}^j.$$

Since $j/(2n^{5/12}) \leq \pi/2$, by Jordan's inequality ($\sin x \geq 2x/\pi$ for $0 \leq x \leq \pi/2$), we have

$$\begin{aligned} S_k(\rho_{n,k}) - \operatorname{Re} S_k(e^{i\eta}\rho_{n,k}) &\ge 2\Big(\frac{2}{\pi}\Big)^2\Big(\frac{1}{2n^{5/12}}\Big)^2 \sum_{j=0}^{k} \frac{j^2\rho_{n,k}^j}{j!} \\ &\ge \frac{2}{\pi^2} n^{-5/6} S_k(\rho_{n,k}) \\ &\ge \frac{2}{\pi^2} n^{-5/6} S_k(u_n-1) \qquad \text{(use Proposition 2)} \\ &\ge \frac{2}{\pi^2} n^{-5/6} S_k(u_n) e^{-1} \\ &\ge \frac{2}{e\pi^2} n^{-5/6} S_{\lfloor(\ln n)/2\rfloor}(u_n), \end{aligned} \tag{3.65}$$

where $\lfloor(\ln n)/2\rfloor$ denotes the integer part of $(\ln n)/2$.

By (A) of Proposition 3, we find

$$S_{\lfloor(\ln n)/2\rfloor}(u_n) \ge n^{0.8465}. \tag{3.66}$$

Combining (3.65) and (3.66), we get

$$S_k(\rho_{n,k}) - \operatorname{Re} S_k(e^{i\eta}\rho_{n,k}) \ge \frac{2}{e\pi^2} n^{0.8465-5/6} \ge \frac{2}{e\pi^2} n^{0.0131}. \tag{3.67}$$

Putting (3.67) into (3.64) gives

$$\operatorname*{Sup}_{\eta<|\theta|\le\pi} \frac{\exp\Big(\operatorname{Re}\big(g_{n,k}(e^{i\theta}\rho_{n,k})\big)\Big)}{\exp\big(g_{n,k}(\rho_{n,k})\big)} \le \exp\Big(-e^{\xi/D_n}\frac{2}{e\pi^2}n^{0.0131}\Big), \tag{3.68}$$

which shows that I_2 is negligible compared to I_1. □

4. Determination of the effective block size

Note that the effective size maintains the expectation of the total number of blocks, i.e., the M_n in (1.1) must be at least asymptotic to the $E(\overline{X}_{n,k})$ in Proposition 5. According to Proposition 5, the effective size k must be such that $S_k(u_n) \sim M_n$. By Proposition 4, the effective size k cannot be less than or equal to u_n, hence $k > u_n$. In summary, we can say that the effective size k must satisfy

$$k > u_n, \tag{4.1}$$

$$S_k(u_n) = e^{u_n}\big(1+o(1)\big). \tag{4.2}$$

We shall find that the constraint $k \le \pi n^{5/12}$ is sufficient to allow us to analyze the problem. With this constraint on k we are in a position to use Proposition 6. From (2.4) and (3.1), we have

$$E\bigg(\exp\Big(\xi\frac{\overline{X}_{n,k}-M_n}{D_n}\Big)\bigg) = e^{-\xi M_n/D_n}\frac{n!}{B(n)}I_{n,k}. \tag{4.3}$$

For clarity, we introduce the notation

$$\frac{S_k(\rho_{n,k})}{e^{\rho_{n,k}}} = 1+\varepsilon_{n,k}, \quad \frac{S_{k-1}(\rho_{n,k})}{e^{\rho_{n,k}}} = 1+\tilde{\varepsilon}_{n,k}. \tag{4.4}$$

Because of Proposition 2 and (4.2), both $\varepsilon_{n,k}$ and $\tilde{\varepsilon}_{n,k}$ are $o(1)$ uniformly for $k > u_n$.

To simplify (4.3), we use (3.35) and Proposition 6. Rearranging factors in (4.3), we have

$$E\Big(\exp\Big(\xi\frac{\overline{X}_{n,k}-M_n}{D_n}\Big)\Big) = e^{-\xi M_n/D_n}\frac{u_n}{\rho_{n,k}}\frac{\sqrt{2\pi(e^{u_n}+nu_n^{-2})}}{\sqrt{2\pi g''_{n,k}(\rho_{n,k})}}\exp(e^{\rho_{n,k}}-e^{u_n})\Big(\frac{u_n}{\rho_{n,k}}\Big)^n$$
$$\times\exp\big(S_k(\rho_{n,k})(e^{\xi/D_n}-1)\big)\exp(1-e^{\xi/D_n})\big(1+o(1)\big). \tag{4.5}$$

Also we have

$$\frac{u_n}{\rho_{n,k}} = 1+o(1), \tag{4.6}$$

$$\frac{\sqrt{2\pi(e^{u_n}+nu_n^{-2})}}{\sqrt{2\pi g''_{n,k}(\rho_{n,k})}} = 1+o(1), \tag{4.7}$$

$$\exp(1-e^{\xi/D_n}) = 1+o(1). \tag{4.8}$$

Putting (4.6), (4.7), and (4.8) into (4.5) gives

$$E\Big(\exp\Big(\xi\frac{\overline{X}_{n,k}-M_n}{D_n}\Big)\Big)$$
$$= e^{-\xi M_n/D_n}\exp(e^{\rho_{n,k}}-e^{u_n})\Big(\frac{u_n}{\rho_{n,k}}\Big)^n\exp\big(S_k(\rho_{n,k})(e^{\xi/D_n}-1)\big)\big(1+o(1)\big), \tag{4.9}$$

uniformly for $\pi n^{5/12}\geq k>u_n$.

Observe the following:

$$\begin{aligned} e^{\rho_{n,k}}-e^{u_n} &= -e^{u_n}(1-e^{\rho_{n,k}-u_n}) \\ &= e^{u_n}(\rho_{n,k}-u_n)+\frac{1}{2}e^{u_n}(\rho_{n,k}-u_n)^2+O(e^{u_n}|\rho_{n,k}-u_n|^3). \end{aligned} \tag{4.10}$$

$$\begin{aligned}\Big(\frac{u_n}{\rho_{n,k}}\Big)^n &= \exp\Big(-n\ln\frac{\rho_{n,k}}{u_n}\Big) = \exp\Big(-n\ln\Big(1+\frac{\rho_{n,k}-u_n}{u_n}\Big)\Big) \\ &= \exp\Big(-\frac{n}{u_n}(\rho_{n,k}-u_n)+\frac{n}{2u_n^2}(\rho_{n,k}-u_n)^2+O\Big(\frac{n}{u_n^3}|\rho_{n,k}-u_n|^3\Big)\Big).\end{aligned} \tag{4.11}$$

Substituting (4.10) and (4.11) in (4.9), we have

$$E\Big(\exp\Big(\xi\frac{\overline{X}_{n,k}-M_n}{D_n}\Big)\Big)$$
$$= e^{-\xi M_n/D_n}\exp\Big(\frac{e^{u_n}}{2}(\rho_{n,k}-u_n)^2+\frac{n}{2u_n^2}(\rho_{n,k}-u_n)^2+o(1)\Big)$$
$$\times\exp\big(S_k(\rho_{n,k})(e^{\xi/D_n}-1)\big)\big(1+o(1)\big). \tag{4.12}$$

Now using Proposition 2 and (4.4), we have

$$\begin{aligned}\frac{e^{u_n}}{2}(\rho_{n,k}-u_n)^2+\frac{n}{2u_n^2}(\rho_{n,k}-u_n)^2+o(1)\\ &=\frac{u_n}{2}\xi^2(1+\tilde{\varepsilon}_{n,k})-\xi^2\big(1+o(1)\big)+\frac{1}{2}\xi^2\big(1+o(1)\big)+o(1)\\ &=\frac{u_n}{2}\xi^2+\frac{u_n}{2}\xi^2\tilde{\varepsilon}_{n,k}-\frac{1}{2}\xi^2+o(1). \qquad (4.13)\end{aligned}$$

Similarly, we have

$$\begin{aligned}S_k(\rho_{n,k})(e^{\xi/D_n}-1)\\ &=(e^{\rho_{n,k}}+e^{\rho_{n,k}}\varepsilon_{n,k})\left(\frac{\xi}{D_n}+\frac{1}{2}\frac{\xi^2}{D_n^2}+O\Big(\frac{1}{D_n^3}\Big)\right)\\ &=e^{\rho_{n,k}}\left(\frac{\xi}{D_n}+\frac{1}{2}\frac{\xi^2}{D_n^2}+O\Big(\frac{1}{D_n^3}\Big)\right)+e^{\rho_{n,k}}\varepsilon_{n,k}\left(\frac{\xi}{D_n}+O\Big(\frac{1}{D_n^2}\Big)\right)\\ &=\frac{e^{u_n}}{D_n}\xi-u_n\xi^2(1+\tilde{\varepsilon}_{n,k})+\frac{1}{2}u_n\xi^2+\xi^2+o(1)+e^{\rho_{n,k}}\varepsilon_{n,k}\left(\frac{\xi}{D_n}+O\Big(\frac{1}{D_n^2}\Big)\right). \qquad (4.14)\end{aligned}$$

These expressions hold uniformly for $\pi n^{5/12}\ge k>u_n$.

Putting (4.13) and (4.14) in (4.12) gives

$$\begin{aligned}E\left(\exp\Big(\xi\frac{\overline{X}_{n,k}-M_n}{D_n}\Big)\right)\\ &=\exp\left(-\frac{1}{2}u_n\xi^2\tilde{\varepsilon}_{n,k}+\frac{1}{2}\xi^2+e^{\rho_{n,k}}\varepsilon_{n,k}\left(\frac{\xi}{D_n}+O\Big(\frac{1}{D_n^2}\Big)\right)\right)(1+o(1)).\end{aligned}$$

We summarize our findings in the following proposition.

Proposition 7. *The Laplace transform of the random variable* $(\overline{X}_{n,k}-M_n)/D_n$*, i.e.,* $E\big(\exp(\xi(\overline{X}_{n,k}-M_n)/D_n)\big)$ *is equal to*

$$\exp\left(-\frac{1}{2}u_n\xi^2\tilde{\varepsilon}_{n,k}+\frac{1}{2}\xi^2+e^{\rho_{n,k}}\varepsilon_{n,k}\left(\frac{\xi}{D_n}+O\Big(\frac{1}{D_n^2}\Big)\right)\right)(1+o(1)), \qquad (4.15)$$

uniformly for $\pi n^{5/12}\ge k>u_n$*, where* $\tilde{\varepsilon}_{n,k}$ *and* $\varepsilon_{n,k}$ *are defined in* (4.4).

The presence of the term $\frac{1}{2}\xi^2$ is an indication of the asymptotic normality of the distribution. It is clear that the effective size k must make

$$-\frac{1}{2}u_n\xi^2\tilde{\varepsilon}_{n,k}+e^{\rho_{n,k}}\varepsilon_{n,k}\left(\frac{\xi}{D_n}+O\Big(\frac{1}{D_n^2}\Big)\right)=o(1).$$

Proposition 8. *If* $\pi n^{5/12}\ge k>u_n$ *and* $\liminf(u_n/k)\ge 3/4$*, then* $E\big(\exp(\xi(\overline{X}_{n,k}-M_n)/D_n)\big)$ *does not converge to* $e^{\xi^2/2}$.

Proof. We know already that $S_k(\rho_{n,k})/e^{\rho_{n,k}} = 1+\varepsilon_{n,k}$, where $\varepsilon_{n,k} = o(1)$. But this is not strong enough to force the conclusion. We need more detailed information about $\varepsilon_{n,k}$. Using Proposition 2 gives

$$\frac{S_k(\rho_{n,k})}{e^{\rho_{n,k}}} = \frac{S_k\big(u_n + O(1/D_n)\big)}{e^{u_n+O(1/D_n)}}.$$

By an argument similar to (3.48), one can show that

$$S_k\left(u_n + O\Big(\frac{1}{D_n}\Big)\right) = S_k(u_n)e^{O(\frac{1}{D_n})}.$$

Hence

$$\frac{S_k(\rho_{n,k})}{e^{\rho_{n,k}}} = \frac{S_k(u_n)}{e^{u_n}}\big(1+O(1/D_n)\big) = \frac{S_k(u_n)}{e^{u_n}} + O(1/D_n). \tag{4.16}$$

By (A) of Proposition 3 we find that

$$\frac{S_k(u_n)}{e^{u_n}} = 1 + \sqrt{\frac{2}{\pi}}\frac{\zeta x}{x-1}\operatorname{Erfc}(\sqrt{k}\zeta)\left(1+O\Big(\frac{1}{\sqrt{k}}\Big)\right), \tag{4.17}$$

where $x = u_n/k < 1$ and $\zeta = |x-1-\ln x|^{1/2}$.

Plugging (4.17) into (4.16) gives

$$\varepsilon_{n,k} = \sqrt{\frac{2}{\pi}}\frac{\zeta x}{x-1}\operatorname{Erfc}(\sqrt{k}\zeta)\left(1+O\Big(\frac{1}{\sqrt{k}}\Big)\right) + O(1/D_n). \tag{4.18}$$

Observe that $\varepsilon_{n,k} = o(1)$, $O(1/D_n) = o(1)$ and that $\zeta/(x-1)$ is bounded. Hence

$$\operatorname{Erfc}(\sqrt{k}\zeta) = o(1). \tag{4.19}$$

This implies

$$\sqrt{k}\zeta \to \infty \quad \text{as } k \to \infty. \tag{4.20}$$

Once we are guaranteed that the argument σ of $\operatorname{Erfc}(\sigma)$ tends to ∞, we can use the traditional asymptotics,

$$\operatorname{Erfc}(\sigma) = \frac{1}{2\sigma}e^{-\sigma^2}\big(1+O(1/\sigma^2)\big) \quad \text{as } \sigma \to \infty. \tag{4.21}$$

Thus

$$\begin{aligned}\operatorname{Erfc}(\sqrt{k}\zeta) &= \frac{1}{2\sqrt{k}\zeta}e^{-k\zeta^2}\left(1+O\Big(\frac{1}{k\zeta^2}\Big)\right)\\ &\geq \frac{1}{3}\exp\left(-k\Big(\frac{u_n}{k}-1-\ln\frac{u_n}{k}\Big)\right)\end{aligned} \tag{4.22}$$

for n large.

Substituting (4.18) and (4.22) into the term $e^{\rho_{n,k}}\varepsilon_{n,k}/D_n$ gives

$$\left|\frac{e^{\rho_{n,k}}\varepsilon_{n,k}}{D_n}\right| \geq \frac{1}{3}\sqrt{\frac{2}{\pi}}\left|\frac{\zeta x}{x-1}\right|\frac{e^{\rho_{n,k}}}{D_n}\exp\left(-k\Big(\frac{u_n}{k}-1-\ln\frac{u_n}{k}\Big)\right) - l_n, \tag{4.23}$$

where $|l_n| = O(e^{\rho_{n,k}}/D_n^2) = O(\ln n)$.

Note that $|\zeta x/(x-1)|$ is non-zero and bounded. Using Proposition 2, we have

$$\begin{aligned}\frac{e^{\rho_{n,k}}}{D_n}\exp\Big(-k\Big(\frac{u_n}{k}-1-\ln\frac{u_n}{k}\Big)\Big)&\\ &=\exp\Big(k\Big(1+\ln\frac{u_n}{k}\Big)-\frac{\ln n}{2}+\ln u_n+O(1/D_n)\Big)\\ &\geq \exp\Big(u_n(1+\ln 5/8)-\frac{\ln n}{2}+\ln u_n\Big)\quad (\text{use }\liminf u_n/k\geq 3/4)\\ &\geq \exp\big(u_n(0.5299-0.5)+\ln u_n\big)\\ &=\exp\big(u_n(0.0299)+\ln u_n\big). \end{aligned}\tag{4.24}$$

From (4.23) and (4.24), it is clear that $|e^{\rho_{n,k}}\varepsilon_{n,k}/D_n|\geq e^{0.0298u_n}$; the asymptotic magnitude is much larger than that of $-u_n/2\xi^2\tilde{\varepsilon}_{n,k}$ which is $O(\ln n)$. Hence

$$-\frac{1}{2}u_n\xi^2\tilde{\varepsilon}_{n,k}+e^{\rho_{n,k}}\varepsilon_{n,k}\Big(\frac{\xi}{D_n}+O\Big(\frac{1}{D_n^2}\Big)\Big)$$

is not $o(1)$. By Proposition 7, Proposition 8 is proved. □

According to Proposition 8, the possible effective size k must be such that $u_n/k\leq 3/4$. We shall use Proposition 3 to find it. We have

$$\frac{S_k(kx)}{e^{kx}}=1-\frac{1}{\sqrt{2k\pi}}\frac{x}{1-x}(xe^{1-x})^k\Big(1+O\Big(\frac{1}{k}\Big)\Big),\tag{4.25}$$

uniformly for $0<x\leq 3/4$.

Let $x=\rho_{n,k}/k$ in (4.25). We have

$$S_k(\rho_{n,k})=e^{\rho_{n,k}}-\frac{e^{\rho_{n,k}}}{\sqrt{2k\pi}}\frac{x}{1-x}(xe^{1-x})^k\Big(1+O\Big(\frac{1}{k}\Big)\Big).\tag{4.26}$$

Comparing (4.26) with (4.4) shows

$$\begin{aligned}e^{\rho_{n,k}}\varepsilon_{n,k}&=-\frac{e^{\rho_{n,k}}}{\sqrt{2k\pi}}\frac{x}{1-x}(xe^{1-x})^k\Big(1+O\Big(\frac{1}{k}\Big)\Big)\\ &=-\frac{1}{\sqrt{2\pi}}\frac{x}{1-x}\exp\Big(\rho_{n,k}+k\ln(xe^{1-x})-\frac{\ln k}{2}\Big)\big(1+O(1/k)\big).\end{aligned}\tag{4.27}$$

Thus

$$\begin{aligned}\frac{e^{\rho_{n,k}}\varepsilon_{n,k}}{D_n}&=-\frac{1}{\sqrt{2\pi}}\frac{x}{1-x}\exp\Big(\rho_{n,k}+k\ln(xe^{1-x})-\frac{\ln k}{2}-\frac{\ln n}{2}+\ln u_n\Big)\\ &\quad\times\big(1+O(1/k)\big).\end{aligned}\tag{4.28}$$

We must force $e^{\rho_{n,k}}\varepsilon_{n,k}/D_n$ to be $o(1)$. Because, in general, $e^{\rho_{n,k}}\varepsilon_{n,k}/D_n\geq u_n\tilde{\varepsilon}_{n,k}$, when this is done the term $-\frac{1}{2}u_n\xi^2\tilde{\varepsilon}_{n,k}$ also simultaneously becomes $o(1)$. Then, by Proposition 7, the asymptotic normality of the random variable $(\overline{X}_{n,k}-M_n)/D_n$ will follow. We now use Proposition 2 to simplify (4.28),

$$\begin{aligned}\frac{e^{\rho_{n,k}}\varepsilon_{n,k}}{D_n}&=-\frac{1}{\sqrt{2\pi}}\frac{x}{1-x}\exp\big(-k\ln k+k\ln u_n+k-\frac{\ln k}{2}-\frac{\ln n}{2}+\ln u_n+o(1)\big)\\ &\quad\times\big(1+O(1/k)\big).\end{aligned}\tag{4.29}$$

The optimal choice for k must satisfy

$$-k\ln k + k\ln u_n + k - \frac{\ln k}{2} - \frac{\ln n}{2} + \ln u_n = 0. \tag{4.30}$$

We emphasize that (4.30) is the equation for the effective size k. It is obvious from the expression (4.30) that if we replace k by $k + L_n$ where $L_n \to \infty$ as $n \to \infty$, then $e^{\rho_{n,k}}\varepsilon_{n,k}/D_n$ is $o(1)$. Consequently the random variable $(\overline{X}_{n,k} - M_n)/D_n$ is asymptotically normal.

5. Asymptotics for the effective size and the transitional distributions

To study equation (4.30), we first must investigate the positive roots of the equation

$$f_n(z) = \frac{1}{2}\ln n - \ln u_n, \tag{5.1}$$

where $f_n(z) := -z\ln z + z\ln u_n + z - \frac{1}{2}\ln z$.

Proposition 9. *For all sufficiently large n, $f_n(z) = \frac{1}{2}\ln n - \ln u_n$ has three positive roots. The relative maximum of $f_n(z)$ occurs approximately at u_n.*

The proof uses only elementary calculus, and we omit it.

Denote by μ_n the largest positive root of (5.1). Hence $\mu_n > u_n$ so μ_n is the effective block size of the problem. In order to study the asymptotics of μ_n, we make a change of variable, $y_n = u_n/\mu_n$. Plugging this into (5.1) we have

$$u_n \ln y_n + u_n - \frac{1}{2}y_n \ln\frac{u_n}{y_n} = y_n\Big(\frac{1}{2}\ln n - \ln u_n\Big). \tag{5.2}$$

Let $h_n(y) = u_n\ln y + u_n - \frac{1}{2}y\ln(u_n/y) - y(\frac{1}{2}\ln n - \ln u_n)$. By Proposition 9 the function $h_n(y)$ has a unique positive root y_n in the interval $(0,1)$ provided that n is sufficiently large.

Proposition 10. *y_n admits the asymptotic approximation*

$$y_n = r + \frac{r^2\ln(1/r)}{2-r}\frac{1}{u_n} + o\Big(\frac{1}{u_n}\Big), \tag{5.3}$$

where r is the unique positive root of $\frac{1}{2}z - \ln z - 1 = 0$ in $(0,1)$.

Proof. The main issue here is to show y_n has a limit. Its asymptotics are obtained by the same bootstrap method as used previously,

$$h_n(1/2) = u_n(1-\ln 2) - \frac{1}{4}\ln(2u_n) - \frac{1}{4}\ln n + \frac{1}{2}\ln u_n.$$

Since $1-\ln 2 > 1/4$, $h_n(1/2) > 0$ if n is large. Similarly, $h_n(1/3) < 0$ if n is large. Hence

$$\frac{1}{3} < y_n < \frac{1}{2} \quad \text{for all large } n. \tag{5.4}$$

Let $\overline{r} := \limsup y_n$. We have

$$\frac{1}{3} \le \overline{r} \le \frac{1}{2}. \tag{5.5}$$

Because the logarithm is a continuous function on $(0,1)$, we have

$$\limsup \ln y_n = \ln \overline{r}. \tag{5.6}$$

Now consider the equation satisfied by y_n:

$$\ln y_n + 1 = \frac{1}{2}\frac{y_n}{u_n}\ln\frac{u_n}{y_n} + y_n\Big(\frac{1}{2}\frac{\ln n}{u_n} - \frac{\ln u_n}{u_n}\Big). \tag{5.7}$$

Taking lim sup in (5.7), we get

$$\limsup(\ln y_n + 1) = \limsup\bigg(\frac{1}{2}\frac{y_n}{u_n}\ln\frac{u_n}{y_n} + y_n\Big(\frac{1}{2}\frac{\ln n}{u_n} - \frac{\ln u_n}{u_n}\Big)\bigg). \tag{5.8}$$

To simplify (5.8), we observe that $\limsup(a_n + b_n) = \limsup a_n + \limsup b_n$ if b_n is a null sequence or a constant sequence. Notice that

$$\frac{1}{2}\frac{y_n}{u_n}\ln\frac{u_n}{y_n} = o(1) \quad \text{and} \quad \lim\Big(\frac{1}{2}\frac{\ln n}{u_n} - \frac{\ln u_n}{u_n}\Big) = 1/2.$$

We have

$$\ln \overline{r} + 1 = \frac{1}{2}\overline{r}. \tag{5.9}$$

That is, $\overline{r}$ satisfies $z/2 - \ln z - 1 = 0$. Similarly, one shows that $\liminf y_n$ also satisfies $z/2 - \ln z - 1 = 0$. Since $z/2 - \ln z - 1 = 0$ has a unique root r in $(0,1)$, we find that $\lim y_n = r$.

Now we bootstrap. Let $y_n = r + \varepsilon_n$ with $\varepsilon_n = o(1)$ and plug this into (5.7),

$$\ln(r+\varepsilon_n) + 1 - \frac{1}{2}\frac{r+\varepsilon_n}{u_n}\ln\frac{u_n}{r+\varepsilon_n} - (r+\varepsilon_n)\Big(\frac{1}{2}\frac{\ln n}{u_n} - \frac{\ln u_n}{u_n}\Big) = 0. \tag{5.10}$$

To simplify (5.10), we use the following:

$$\frac{\ln n}{2u_n} - \frac{\ln u_n}{u_n} = \frac{1}{2} - \frac{\ln u_n}{2u_n}, \tag{5.11}$$

$$\begin{aligned} -\frac{r+\varepsilon_n}{2u_n}\ln\frac{u_n}{r+\varepsilon_n} &= -\frac{r}{2u_n}\ln\frac{u_n}{r} + \frac{r}{2u_n}\ln\Big(1+\frac{\varepsilon_n}{r}\Big) \\ &\quad - \frac{\varepsilon_n}{2u_n}\ln\frac{u_n}{r} + \frac{\varepsilon_n}{2u_n}\ln\Big(1+\frac{\varepsilon_n}{r}\Big). \end{aligned} \tag{5.12}$$

We plug (5.11) and (5.12) in (5.10) to get

$$\varepsilon_n\Big(\frac{1}{r} - \frac{1}{2}\Big) - \frac{r}{2u_n}\ln\frac{1}{r} + O(\varepsilon_n^2) + O\Big(\varepsilon_n\frac{\ln u_n}{u_n}\Big) = 0$$

or

$$\varepsilon_n \sim \Big(\frac{1}{r} - \frac{1}{2}\Big)^{-1}\frac{r\ln(1/r)}{2u_n} \quad \text{as } n \to \infty.$$

That is

$$\varepsilon_n = \frac{r^2\ln(1/r)}{2-r}\frac{1}{u_n} + o\Big(\frac{1}{u_n}\Big). \quad \square \tag{5.13}$$

Theorem 2. *The effective size*

$$\mu_n = \frac{u_n}{r} - \frac{\ln(1/r)}{2-r} + o(1). \tag{5.14}$$

Proof. Use Proposition 10 and the fact that $\mu_n = u_n/y_n$. □

The effective size μ_n has the property that it is optimal and guarantees the random variable $(\overline{X}_{n,\mu_n+L_n} - M_n)/D_n$ is asymptotically normal for all L_n satisfying $L_n \to \infty$ as $n \to \infty$. If L_n is a bounded sequence, then the corresponding behavior of $(\overline{X}_{n,\mu_n+L_n} - M_n)/D_n$ is called the *transitional behavior.* Let

$$k = \lfloor \mu_n \rfloor + L_n, \tag{5.15}$$

where $|L_n| \le M$ is a bounded integer sequence.

To study the transitional behavior, we return to (4.29)

$$\begin{aligned}\frac{e^{\rho_{n,k}\varepsilon_{n,k}\xi}}{D_n} &= -\frac{\xi}{\sqrt{2\pi}}\frac{x}{1-x}\exp\Big(-k\ln k + k\ln u_n + k - \frac{\ln k}{2} - \frac{\ln n}{2} + \ln u_n + o(1)\Big)\\ &\quad\times\big(1+O(1/k)\big)\end{aligned} \tag{5.16}$$

where $x = \rho_{n,k}/k$. Using (5.15) to simplify (5.16) gives

$$-k\ln k + k\ln u_n + k - \frac{\ln k}{2} - \frac{\ln n}{2} + \ln u_n = (L_n - \{\mu_n\})\ln r + o(1), \tag{5.17}$$

where $\{\mu_n\}$ denotes the fractional part of μ_n. Furthermore

$$\frac{x}{1-x} = \frac{r}{1-r} + o(1). \tag{5.18}$$

Substituting (5.17) and (5.18) into (5.16) gives

$$\frac{e^{\rho_{n,k}\varepsilon_{n,k}\xi}}{D_n} = -\frac{\xi}{\sqrt{2\pi}}\frac{r}{1-r}r^{L_n-\{\mu_n\}}\big(1+o(1)\big). \tag{5.19}$$

Put (5.19) in (4.15) of Proposition 7 and observe that $-\frac{1}{2}u_n\xi^2\tilde{\varepsilon}_{n,k} = o(1)$. Thus

$$E\Big(\exp\Big(\xi\frac{\overline{X}_{n,k}-M_n}{D_n}\Big)\Big) = \exp\Big(-\frac{\xi}{\sqrt{2\pi}}\frac{r}{1-r}r^{L_n-\{\mu_n\}}\Big)\exp\Big(\frac{1}{2}\xi^2+o(1)\Big)\big(1+o(1)\big).$$

This implies that

$$E\Big(\exp\Big(\xi\frac{\overline{X}_{n,k}-\overline{M}_n}{D_n}\Big)\Big) = \exp\Big(\frac{1}{2}\xi^2\Big) + o(1),$$

where

$$\overline{M}_n = M_n - \frac{r^{L_n+1-\{\mu_n\}}}{\sqrt{2\pi}(1-r)}D_n. \tag{5.20}$$

Thus we have proved the following:

Theorem 3 (The transitional distribution). *If $k = \lfloor \mu_n \rfloor + L_n$ where L_n is a bounded integer sequence, then the random variable $(\overline{X}_{n,k} - \overline{M}_n)/D_n$ is still asymptotically normal, where $\overline{M}_n$ is defined in* (5.20).

Remark. The transitional distribution is still normal, but the mean $\overline{M}_n$ is different from M_n. More interestingly, the mean $\overline{M}_n$ is an oscillatory function of n due to the following.

Proposition 11. *The sequence* $\{\mu_n\}$ *is dense in* $[0,1]$.

Proof. Let t be a large positive number, and let u_t be the unique positive root of $ze^z = t$. Thus when t is an integer n, u_t is reduced to u_n. Consider the equation in z

$$-z\ln z + z\ln u_t + z - \frac{1}{2}\ln z = \frac{1}{2}\ln t - \ln u_t. \tag{5.21}$$

Let $\mu(t)$ be the largest positive root of (5.21). Thus when t is an integer n, $\mu(t)$ is reduced to μ_n. To proceed we mention the following simple facts:

(a) $\mu(t)$ is well-defined for all sufficiently large t.
(b) $\mu(t)$ is a differentiable function of t.
(c) Since u_t satisfies $ze^z = t$, we have

$$u_t' = (e^{u_t} + t)^{-1}. \tag{5.22}$$

To show $\mu(t)$ is a strictly increasing function, we differentiate the equation below with respect to t:

$$-\mu(t)\ln\mu(t) + \mu(t)\ln u_t + \mu(t) - \frac{1}{2}\ln\mu(t) = \frac{1}{2}\ln t - \ln u_t.$$

Thus

$$\mu'(t)\left(-\ln\mu(t) + \ln u_t - \frac{1}{2\mu(t)}\right) = \frac{1}{2t} - \frac{1+\mu(t)}{t+tu_t}. \tag{5.23}$$

Since $\mu(t) > u_t$, we have

$$-\ln\mu(t) + \ln u_t - \frac{1}{2\mu(t)} < 0. \tag{5.24}$$

By Theorem 1 (when n is replaced by t the same conclusion still holds) we see

$$\frac{1+\mu(t)}{t+tu_t} \sim \frac{1}{tr} > \frac{1}{t}.$$

This implies that

$$\frac{1}{2t} - \frac{1+\mu(t)}{t+tu_t} < 0. \tag{5.25}$$

Hence $\mu'(t) > 0$ for all sufficiently large t so $\mu(t)$ is a continuous, strictly increasing function of t that tends to infinity as $t \to \infty$. Let $K(t)$ be its inverse so that

$$\mu\big(K(t)\big) = K\big(\mu(t)\big) = t. \tag{5.26}$$

Applying Theorem 1 to (5.23), we get

$$\mu'(t) \to 0^+ \quad \text{as} \quad t \to \infty. \tag{5.27}$$

Hence

$$K'(t) = \frac{1}{\mu'\big(K(t)\big)} \to \infty \quad \text{as} \quad t \to \infty. \tag{5.28}$$

Now given any a and b such that $0 < a < b < 1$, by the mean value theorem we have

$$K(b+l) - K(a+l) = K'(\xi)(b-a) \quad \text{for some } \xi \text{ between } a+l \text{ and } b+l.$$

By (5.28) $K'(\xi) \to \infty$ as $l \to \infty$. Hence there exist integers n_0 and l_0 such that

$$K(b+l_0) > n_0 > K(a+l_0).$$

This implies $\mu(K(b+l_0)) > \mu(n_0) > \mu(K(a+l_0))$, and $b + l_0 > \mu_{n_0} > a + l_0$. Consequently $\{\mu_{n_0}\} \in (a,b)$. □

References

1. M. Abramowitz and I. A. Stegun, eds., *Handbook of Mathematical Functions*, New York, 1965.
2. E. A. Bender, *Central and local limit theorems applied to asymptotic enumeration*, J. Combin. Theory Ser. A **15** (1973), 91–111.
3. J. Curtiss, *A note on the theory of moment generating functions*, Ann. Math. Statist. **13** (1942), 430–433.
4. L. Harper, *Stirling behavior is asymptotically normal*, Ann. Math. Statist. **38** (1967), 410–414.
5. M. Kac, *Probability methods in some problems of analysis and number theory*, Bull. Amer. Math. Soc. **55** (1949), 641–665.
6. N. G. de Bruijn, *Asymptotic Methods in Analysis*, Dover, New York, 1981.
7. V. N. Sachkov, *Random partitions of sets*, Theory Probab. Appl. **190** (1974), 184–190.
8. G. Szegö, *Uber eine Eigenschaft der Exponentialreihe*, Sitzungsber. Berl. Math. Ges. **23** (1924), 50–64.
9. N. M. Temme, *The asymptotic expansion of the incomplete gamma functions*, SIAM J. Math. Anal. **10** (1979), 757–766.
10. R. S. Varga, *Topics in Rational Interpolation and Approximation*, Les Presses d l'Université de Montreal, Montreal, Canada, 1982.
11. H. Wilf, *generatingfunctionology*, Academic Press, 1990.

Department of Mathematics and Computer Science, Drexel University, Philadelphia, Pennsylvania 19104, U. S. A.