On the number of distinct block sizes in partitions of a set

A. M. Odlyzko
Bell Laboratories
Murray Hill, New Jersey 07974
USA

and

L. B. Richmond
Department of Combinatorics and Optimization
University of Waterloo
Waterloo, Ontario N2L 3G1
Canada

Abstract

The average number of distinct block sizes in a partition of a set of n elements is asymptotic to $e \log n$ as $n \to \infty$. In addition, almost all partitions have approximately $e \log n$ distinct block sizes. This is in striking contrast to the fact that the average total number of blocks in a partition is $\sim n(\log n)^{-1}$ as $n \to \infty$.

On the number of distinct block sizes in partitions of a set

A. M. Odlyzko
Bell Laboratories
Murray Hill, New Jersey 07974
USA

and

L. B. Richmond
Department of Combinatorics and Optimization
University of Waterloo
Waterloo, Ontario N2L 3G1
Canada

1. Introduction

A recent paper of H. Wilf [6] compares the number of distinct part sizes to the total number of parts in various combinatorial partition problems. It is well known and easy to prove that the average number of cycles of a permutation on n symbols is

$$\log n + \gamma + o(1) \quad \text{as} \quad n \to \infty ,$$

when $\gamma = 0.577...$ denotes Euler's constant. Wilf showed that the average number of distinct cycle sizes in a permutation on n letters is

$$\log n + \gamma - Q + o(1) \quad \text{as} \quad n \to \infty ,$$

where

$$Q = \sum_{n=2}^{\infty} \frac{(-1)^n}{n!} \zeta(n) = 0.65981... .$$

Thus in this case the numbers of parts and part sizes are almost the same.

The average number of parts in a partition of an integer n is known to be [3]

$$\sim \pi^{-1} (3/2)^{-1} n^{1/2} \log n \quad \text{as} \quad n \to \infty .$$

Wilf showed that the average number of distinct part sizes in a partition of n is

$$\sim \pi^{-1} 6^{1/2} n^{1/2} \quad \text{as} \quad n \to \infty .$$

Thus in this case the numbers of parts and of part sizes grow at slightly different rates.

The number of partitions of a set of n elements into k subsets is given by $S(n,k)$, the Stirling numbers of the second kind. The asymptotics of the $S(n,k)$ were known already to Laplace (see [1,5] for extensive bibliographies), and it follows from these asymptotic estimates that the average number of blocks in a partition of an n-element set is

$$\sim \frac{n}{\log n} \quad \text{as} \quad n \to \infty .$$

(Wilf has pointed out that this result can also be derived from the asymptotics of the Bell numbers and the recurrence for the Stirling numbers.) Wilf [6] derived a generating function for $B(n,k)$, the number of partitions of an n-element set with exactly k distinct block sizes, but he left open the problem of estimating $b(n)$, the average number of distinct block sizes. In this note we present two proofs that

$$b(n) \sim e \log n \quad \text{as} \quad n \to \infty .$$

Thus in this case there is a great difference between number of parts and part sizes. We also indicate how both our proofs can be easily adapted to show that most of the time the number of distinct part sizes is very close to $e \log n$ (i.e., the normal order is $e \log n$). The first proof is entirely self-contained apart from using the well-known formula for the asymptotics of the Bell numbers. The second proof relies on the general result of Hayman [4] about Taylor series coefficients of analytic functions.

In Section 2 we rederive Wilf's formula for the generating function of the $B(n,k)$. Our proofs are then presented in sections 3 and 4. With additional work it might be possible to obtain the complete distribution function of the $B(n,k)$.

2. Generating functions and preliminaries

We let $B(n,k)$ denote the number of partitions of an n-element set that have exactly k distinct sizes of blocks, and we let

$$B(n) = \sum_{k=1}^{n} B(n,k)$$

be the *n*-th Bell number, the total number of partitions. (We remark in passing that $B(n,k) = 0$ for k larger than approximately $(2n)^{1/2}$, which immediately indicates that the average numbers of blocks and block sizes have to be very different.)

Wilf's generating function for the $B(n,k)$, which he derives from his more general results [6], is

$$F(x,y) = \sum_{n,k \geq 0} \frac{B(n,k)}{n!} x^n y^k = \prod_{m=1}^{\infty} \{1 + y(\exp(\frac{x^m}{m!}) - 1)\} . \tag{2.1}$$

To prove it, we expand each of the exponentials on the right side of (2.1), and expand the product. We find that the coefficient of $n! x^n y^k$ in the resulting expansion is

$$\sum \frac{n!}{\prod_{i=1}^{k} l_i ! (m_i !)^{l_i}} , \tag{2.2}$$

where the sum is over choices of $l_1, \ldots, l_k > 0$, $m_1, \ldots, m_k > 0$, $\sum l_i m_i = n$. But each of the summands in (2.2) is the number of ways of choosing l_i blocks of size m_i from a set of n elements when the order of the blocks is irrelevant, which proves (2.1).

Setting $y = 1$ in (2.1) gives

$$F(x,1) = \sum_{n \geq 0} \frac{B(n)}{n!} x^n = \exp(e^x - 1) , \tag{2.3}$$

the well-known generating function for the Bell numbers.

Define

$$B_1(n) = \sum_k kB(n,k) ,$$

$$B_2(n) = \sum_k k^2 B(n,k) .$$

Then

$$\sum_n \frac{B_1(n)}{n!} x^n = \frac{\partial}{\partial y} F(x,y) \Big|_{y=1} = F(x,y) \cdot \sum_{m=1}^{\infty} \frac{\exp(\frac{x^m}{m!})-1}{1+y(\exp(\frac{x^m}{m!})-1)} \Big|_{y=1}$$
$$= F(x,1) \sum_{m=1}^{\infty} (1-\exp(-\frac{x^m}{m!})) , \tag{2.4}$$

and similarly

$$\sum_n \frac{B_2(n)}{n!} x^n = \frac{\partial}{\partial y} y \frac{\partial}{\partial y} F(x,y) \Big|_{y=1}$$
$$= F(x,1) \left[\{ \sum_{m=1}^{\infty} (1-\exp(-\frac{x^m}{m!})) \}^2 + \sum_{m=1}^{\infty} (1-\exp(-\frac{x^m}{m!})) \right. \tag{2.5}$$
$$\left. - \sum_{m=1}^{\infty} (1-\exp(-\frac{x^m}{m!}))^2 \right] .$$

To prove our result about the average number of block sizes, we will show that

$$b(n) = \frac{B_1(n)}{B(n)} \sim e \log n \quad \text{as} \quad n \to \infty . \tag{2.6}$$

To prove the result about normal order, it is sufficient to show that

$$\frac{B_2(n)}{B(n)} \sim \left[\frac{B_1(n)}{B(n)} \right]^2 , \tag{2.7}$$

since then the claimed result follows from Chebyshev's inequality. We do not present the details of the proof of (2.7), since they are analogous to the proofs of (2.6), although more involved.

Before proceeding to the proofs, we recall the asymptotic expansion of the Bell numbers (more precise results are known, see [2]):

$$\frac{B(n)}{n!} = \frac{1}{e\sqrt{2\pi}} \exp(\frac{n+1}{u_n} - (n+1) \log u_n - \frac{1}{2} u_n + o(1)) \quad \text{as} \quad n \to \infty , \tag{2.8}$$

where u_n is the unique positive root of

$$u_n e^{u_n} = n+1 , \tag{2.9}$$

so that

$$u_n = \log n - \log \log n + O\left(\frac{\log \log n}{\log n}\right) . \qquad (2.10)$$

This result is obtained by using Cauchy's formula

$$\frac{B(n)}{n!} = \frac{1}{2\pi i n!} \int_{|z|=u} F(z,1) z^{-n-1} dz$$

with $u = u_n$.

3. First proof

This proof shows, in essence, that the coefficient of z^n in the Taylor series of

$$f_k(z) = e^{e^z - 1}(1 - e^{-z^k/k!}) \qquad (3.1)$$

is approximately $B(n)/n!$ for $k \le e \log n$ and is negligible for $k > e \log n$.

First we make some preliminary observations. Since for $k \ge 1$

$$e^z - 1 = \sum_{n=1}^{\infty} z^n/n! \ , \qquad (3.2)$$

$$e^z - 1 - z^k/k! = \sum_{\substack{n=1 \\ n \ne k}} z^n/n! \ , \qquad (3.3)$$

both have Taylor series coefficients that are ≥ 0, we have for any $z \in \mathbf{C}$,

$$|e^z - 1 - z^k/k!| \le e^{|z|} - 1 - |z|^k/k! \ . \qquad (3.4)$$

Similarly, since the Taylor coefficients of (3.3) are ≥ 0 and less than or equal to those of (3.2), if

$$\exp(e^z - 1 - z^k/k!) = \sum_{m=1}^{\infty} b(k,m) z^m \ , \qquad (3.5)$$

then

$$0 \le b(k,m) \le B(m)/m! \ , \qquad (3.6)$$

and

$$|f_k(z)| \le f_k(|z|) \ .$$

We now proceed to the main part of the proof. Fix any $\varepsilon \in (0, 10^{-3})$. Consider $k \le (e-\varepsilon)\log n$, where n is taken sufficiently large (depending only on ε). The coefficient of z^n in $f_k(z)$ is $B(n) - b(k,n)$. By Cauchy's theorem,

$$b(k,n) = \frac{1}{2\pi i} \int_{|z|=u_n} \exp(e^z - 1 - z^k/k!)\, z^{-n-1}\, dz ,$$

and so by (3.4)

$$\begin{aligned} b(k,n) &\le u_n^{-n} \max_{|z|=u_n} \left|\exp(e^z - 1 - z^k/k!)\right| \\ &= \exp(e^{u_n} - 1 - u_n^k/k! - n \log u_n) \\ &= \sqrt{2\pi}\,(n!)^{-1} B(n) \exp(\log u_n + u_n/2 - u_n^k/k! + o(1)) \\ &\le (n!)^{-1} B(n) \exp(-u_n/4 + o(1)) \quad \text{as} \quad n \to \infty , \end{aligned}$$

since for $1 \le k \le (e-\varepsilon) \log n$, $u_n^k/k! \ge u_n$ (for n large enough). Therefore the coefficient of z^n in the expansion of

$$\sum_{k \le (e-\varepsilon)\log n} f_k(z) = e^{e^z - 1} \sum_{k \le (e-\varepsilon)\log n} (1 - e^{-z^k/k!})$$

is

$$\sim (e-\varepsilon)(n!)^{-1} B(n) \log n \quad \text{as} \quad n \to \infty . \tag{3.7}$$

Also, the corresponding coefficient for the range $(e-\varepsilon) \log n \le k \le (e+\varepsilon) \log n$ is in the range $[0, 2\varepsilon (n!)^{-1} B(n) \log n]$ by (3.6).

It remains to deal with $k \ge (e+\varepsilon) \log n$. If $k \ge 100\, \varepsilon^{-1} \log n$, then by Stirling's formula, on $|z| = u_n$ we have

$$|z^k/k!| = \frac{u_n^k}{k!} \le \left(\frac{e \log n}{k}\right)^k \le e^{-3k} ,$$

and therefore

$$|f_k(z)| \le \exp(e^{u_n} - 1 - 2k) .$$

If $(e+\varepsilon) \log n \le k \le 100\varepsilon^{-1} \log n$, then on $|z| = u_n$,

$$1-e^{-z^k/k!} = \sum_{1 \le m \le 100\varepsilon^{-1} k^{-1} \log n} (-1)^{m-1} \frac{z^{km}}{(k!)^m} + O(e^{-50 \log n}) .$$

Hence the coefficient of z^n in the Taylor expansion of

$$\sum_{k \ge (e+\varepsilon) \log n} f_k(z)$$

is

$$\frac{1}{2\pi i} \int_{|z|=u_n} \Big\{ \sum_{k \ge (e+\varepsilon)\log n} f_k(z) \Big\} z^{-n-1} dz = \sum_{\substack{k \ge (e+\varepsilon)\log n \\ k \le 100\varepsilon^{-1} \log n}} \sum_{m=1}^{100\varepsilon^{-1} k^{-1} \log n} \frac{(-1)^{m-1}}{(k!)^m} \frac{1}{2\pi i} \int_{|z|=u_n} e^{e^z-1} z^{km-n-1} dz$$

$$+ O(e^{\exp(u_n) - 5 \log n - n \log u_n}) , \qquad (3.8)$$

and the last term above is

$$O\left[\frac{B(n)}{n!} n^{-4} \right] . \qquad (3.9)$$

Again by Cauchy's theorem

$$\frac{1}{2\pi i} \int_{|z|=u_n} e^{e^z-1} z^{km-n-1} dz = \frac{B(n-km)}{(n-km)!} . \qquad (3.10)$$

We now conclude the proof by showing that

$$\frac{B(n-km)}{(k!)^m (n-km)!}$$

is small when compared to $B(n)/n!$ (and $(e+\varepsilon) \log n \le k \le 100\varepsilon^{-1} \log n$, $1 \le m \le 100\varepsilon^{-1}$, say).

Suppose that $(e+\varepsilon) \log n \le v \le 10^4 \varepsilon^{-2} \log n$. Then by (2.8),

$$\log \frac{B(n-v) n!}{B(n)(n-v)!} = -\frac{n+1}{u_n} + \frac{n-v+1}{u_{n-v}} + (n+1) \log u_n - (n-v+1) \log u_{n-v}$$
$$+ u_n/2 - u_{n-v}/2 + o(1) .$$

Now by (2.10),

$$u_{n-v} = u_n - v/n + O(\frac{v}{n \log n}) ,$$

so

$$\log \frac{B(n-v)n!}{B(n)(n-v)!} = (n+1)(\frac{1}{u_{n-v}} - \frac{1}{u_n}) - \frac{v}{u_{n-v}} + (n+1) \log \frac{u_n}{u_{n-v}}$$

$$+ v \log u_{n-v} + o(1)$$

$$= v \log u_n + O(1) .$$

Since for n sufficiently large, and $k \geq (e+\varepsilon) \log n$,

$$\log (k!)^m \geq m \,[k \log k - (1+\varepsilon/100)k] ,$$

we finally obtain

$$\log \left[\frac{B(n-km)n!}{k!^m B(n)(n-km)!} \right] \leq km \log u_n - mk \log k + (1+\varepsilon/100)km + O(1)$$

$$\leq km[\log \log n + o(1) - \log(e+\varepsilon) - \log \log n$$

$$+ (1+\varepsilon/100)] + O(1)$$

$$\leq -\varepsilon km/1000 \leq -\varepsilon 10^{-3} \log n .$$

Therefore

$$\sum_{\substack{k > (e+\varepsilon)\log n \\ k \leq 100\varepsilon^{-1} \log n}} \sum_{m=1}^{(100\varepsilon^{-1} \log n)/k} \frac{(-1)^{m-1}}{(k!)^m} \frac{1}{2\pi i} \int_{|z|=u_n} e^{e^z - 1} z^{km-n-1} dz$$

$$= O(\frac{B(n)}{n!} n^{-\varepsilon/2000}) . \qquad (3.11)$$

It follows from (3.7)-(3.11) that

$$\frac{B_1(n)}{B(n)} \sim e \log n \quad \text{as} \quad n \to \infty ,$$

which completes our first proof.

4. Second proof

We now prove our estimate using Hayman's results [4] concerning admissible functions. A function $f(z)$ is said to be admissible if (we need only consider the case f entire) with

$$a(v) = v \frac{f'(v)}{f(v)} = \frac{d \log f(v)}{d \log v} ,$$

$$b(v) = va'(v) = \frac{d^2 \log f(v)}{d^2 \log v} = v \frac{f'(v)}{f(v)} + v^2 \frac{f''(v)}{f(v)} - v^2 \left(\frac{f'(v)}{f(v)}\right)^2 ,$$

the following three conditions hold;

I) for some function $\delta(v)$ with $0 < \delta(v) < \pi$,

$$f(ve^{i\theta}) \sim f(v) e^{i\theta a(v) - \theta^2 b(v)/2}, \quad \text{as} \quad v \to \infty ,$$

uniformly for $|\theta| \le \delta(v)$, while

II) uniformly for $\delta(v) \le |\theta| \le \pi$

$$f(ve^{i\theta}) = o(f(v)/\sqrt{b(v)}), \quad \text{as} \quad v \to \infty$$

and finally

III) $b(v) \to \infty \quad \text{as} \quad v \to \infty$.

Lemma 1: The function

$$f(z) = e^{e^z - 1} \sum_{m=1}^{\infty} (1 - \varepsilon^{-z^m/m!})$$

is admissible.

Proof. The proof that III holds is immediate, since it is easily seen that $b(v) \sim v^2 e^v$ as $v \to \infty$ for this function.

We now establish that II holds for $\delta(v) = 2 \exp(-2v/5)$. Let $m_0 = m_0(v)$ be the largest integer such that

$$\frac{v^m}{m!} \geq v \quad \text{for} \quad m \leq m_0 \ .$$

We note that $m_0 \sim ev$ as $v \to \infty$. The argument of Section 3 shows that for $m \leq m_0$, $v = |z|$,

$$|\exp(e^z - 1 - z^m/m!| \leq \exp(e^v - 1 - v^m/m!) \leq \exp(e^v - 1 - v) \ .$$

Furthermore, if $z = ve^{i\theta}$, $\delta(v) = |\theta| \leq \pi$, then for large enough v,

$$\begin{aligned} \text{Re } e^z &= e^{v \cos \theta} \cos(v \sin \theta) \leq \exp(v \cos \delta(v)) \\ &\leq \exp(v(1 - \frac{1}{3}\delta(v)^2)) \\ &\leq \exp(v(1 - e^{-4v/5})) \leq e^v - ve^{v/5} \ , \end{aligned}$$

so

$$|\exp(e^z - 1)| \leq \exp(e^v - 1 - ve^{v/5}) \ .$$

Therefore for $m \leq m_0$, $z = ve^{i\theta}$, $\delta(v) \leq |\theta| \leq \pi$,

$$\begin{aligned} |\exp(e^z - 1) - \exp(e^z - 1 - z^m/m!)| &\leq \exp(e^v - 1)(\exp(-ve^{v/5}) + \exp(-v)) \\ &\leq 2 \exp(e^v - 1 - v) \ . \end{aligned} \tag{4.1}$$

Moreover for $m \geq m_0$, $|z|^m/m! < v$, so

$$|\exp(e^z - 1 - z^m/m!)| \leq |\exp(e^z - 1)| \exp(v) \ ,$$

and thus

$$|\exp(e^z - 1) (1 - \exp(-z^m/m!)| \leq 2 \exp(e^v - 1 + v - ve^{v/5}) \ . \tag{4.2}$$

Finally, if $m \geq v^2$, then

$$\left|\frac{z^m}{m!}\right| = O(e^{-m}) \ , \tag{4.3}$$

and so

$$|\exp(e^z - 1) (1 - \exp(-z^m/m!))| = O(\exp(e^v - m)) \ . \tag{4.4}$$

Applying (4.1) for $m \leq m_0$, (4.2) for $m_0 < m < v^2$, and (4.3) for $m \geq v^2$, we obtain

$$|\exp(e^z - 1) \sum_m (1 - \exp(-z^m / m!))| = O(\exp(e^v - v)) ,$$

which establishes II.

To complete the proof of the lemma, we need to show that I holds. Since

$$\exp(e^{ve^{i\theta}} - 1) \sim \exp(e^v - 1 + i\theta a(v) - \theta^2 b(v)/2)$$

as $v \to \infty$, uniformly for $|\theta| \le \delta(v)$, it will suffice to show that if

$$g(z) = \sum_m (1 - \exp(-z^m / m!)) ,$$

then in that same range for θ,

$$g(z) \sim g(v) \quad \text{as} \quad v \to \infty . \tag{4.5}$$

This part of the proof uses arguments similar to those of Section 3. Fix $\varepsilon > 0$. For $m \le (e - \varepsilon) v$, $|\theta| \le \delta(v)$, $z = ve^{i\theta}$,

$$\mathrm{Re}(z^m / m!) = (m!)^{-1} v^m \cos m\theta \ge \frac{1}{2} (m!)^{-1} v^m \ge e^{\varepsilon v / 10}$$

for large v, so

$$1 - \exp(-z^m / m!) = 1 + O(e^{-\varepsilon v}) , \tag{4.6}$$

where the constant implied by the O-notation depends only on ε. Also, for $m \ge (e + \varepsilon) v$,

$$\left| \frac{z^m}{m!} \right| = \frac{v^m}{m!} \le e^{-\varepsilon m / 10}$$

for large v, so

$$1 - \exp(-z^m / m!) = O(e^{-\varepsilon m / 10}) . \tag{4.7}$$

Finally, for $(e - \varepsilon) v \le m \le (e + \varepsilon) v$,

$$\mathrm{Re}\ z^m = v^m \cos m\theta > 0 ,$$

so

$$|\exp(-z^m / m!)| \le 1$$

and

$$|1 - \exp(-z^m/m!)| \le 2 \ . \tag{4.8}$$

Combining (4.5)-(4.7), we find that for $|\theta| \le \delta(v)$,

$$g(ve^{i\theta}) = ev + O(\log\ v) \ .$$

Since this holds for all $\varepsilon > 0$, we obtain (4.5), and in fact even the more precise statement that

$$g(ve^{i\theta}) \sim ev \tag{4.9}$$

as $v \to \infty$, uniformly for $|\theta| \le \delta(v)$.

□

We can now prove our result by applying Theorem I of [4], which gives (using (4.9))

$$\frac{B_1(n)}{n!} \sim \frac{ev_n}{v_n^n \sqrt{2\pi b(v_n)}} ,$$

where v_n is defined by $a(v_n) = n$. Theorem I of [4] also implies that

$$\frac{B(n)}{n!} \sim \frac{\exp(e^{u_n} - 1)}{u_n^n \sqrt{2\pi(u_n + u_n^2)\exp(u_n)}} ,$$

where u_n is defined by (2.9). Stirling's formula implies that

$$\begin{aligned} a(v_n) &= v_n e^{v_n} + \sum_{m=1}^{\infty} ((m-1)!)^{-1} v_n^m \exp(-v_n^m/m!) g(v_n)^{-1} \\ &= v_n e^{v_n} + O(v_n^{-1/2+\varepsilon}) \ . \end{aligned} \tag{4.10}$$

Next, (2.9) and (4.10) show that

$$(v_n - u_n) e^{v_n} + u_n (e^{v_n} - e^{u_n}) = O(v_n^{-1/2+\varepsilon}) \ ,$$

hence

$$v_n - u_n = O(e^{-v_n} v_n^{-1/2+\varepsilon}) \ .$$

This implies that

$$\exp(e^{v_n}) \;=\; \exp(e^{u_n} + O(v_n^{-1/2+\varepsilon})) \;=\; \exp(e^{u_n})(1 + O(v_n^{-1/2+\varepsilon})) \;,$$

$$v_n^n \;=\; u_n^n (1 + O(e^{-v_n})) \;,$$

and that

$$b(v_n) \;=\; (u_n + u_n^2) e^{u_n} (1 + O(e^{-u_n})) \;.$$

Finally, we find that

$$\frac{B_1(n)}{n!} \sim e v_n \, \frac{B(n)}{n!} \sim e u_n \, \frac{B(n)}{n!} \sim (e \, \log \, n) \, \frac{B(n)}{n!} \;,$$

which is our result.

REFERENCES

[1] E. A. Bender, Central and local limit theorems applied to asymptotic enumeration, J. Comb. Theory (A), *15* (1973), 91-111.

[2] N. G. de Bruijn, *Asymptotic Methods in Analysis*, North-Holland, 1958.

[3] P. Erdös and J. Lehner, The distribution of the number of summands in the partition of a positive integer, Duke Math. J. *8* (1941), 335-345.

[4] W. K. Hayman, A generalization of Stirling's formula, J. reine angew. Math. *196* (1956), 67-95.

[5] V. V. Menon, On the maximum of Stirling numbers of the second kind, J. Comb. Theory (A) *15* (1973), 11-24.

[6] H. S. Wilf, Three problems in combinatorial asymptotics, J. Comb. Theory (A), to appear.