Reprinted from JOURNAL OF COMBINATORIAL THEORY, Series A
All Rights Reserved by Academic Press, New York and London

Vol. 35, No. 2, September 1983
Printed in Belgium

Three Problems in Combinatorial Asymptotics

HERBERT S. WILF*

Department of Mathematics, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Communicated by the Managing Editors

Received April 7, 1983

1. INTRODUCTION

We consider the following questions:

(A) How many different *sizes* of cycles does a permutation of n letters have (on the average, when n is large)?

(B) How many *different sizes* of parts does a partition of the integer n have (on the average, when n is large)?

(C) How many subspaces does an n-dimensional vector space over a finite field have, when n is large?

The answers are given, respectively, by

THEOREM A. *The average number of cycles exceeds the average number of distinct cycle lengths of an n-permutation by an amount that approaches the constant*

$$A = \sum_{m=2}^{\infty} \frac{(-1)^{m+1}}{m!} \zeta(m) = .6598155... \tag{1}$$

as n approaches infinity (ζ *is the Riemann zeta function*).

It is well known that the average number of cycles is $\log n + \gamma + o(1)$.

THEOREM B. *The average number of different sizes of parts that a partition of the integer n has is*

$$\bar{d}_n \sim \frac{\sqrt{6}}{\pi} n^{1/2} \qquad (n \to \infty). \tag{2}$$

* Research supported by the National Science Foundation.

0097-3165/83 $3.00

In fact, almost all partitions have

$$\frac{\sqrt{6}}{\pi} n^{1/2} + \omega(n)\, n^{1/4}$$

different sizes of parts, where $\omega(n)$ is any function that tends to infinity with n.

THEOREM C. *Let q be a fixed prime power, and let G_n be the number of vector subspaces of n-space over $GF(q)$. Then*

$$\begin{aligned} G_n &\sim C_0(q)\, q^{n^2/4} \qquad (n \to \infty, n \text{ even}), \\ &\sim C_1(q)\, q^{n^2/4} \qquad (n \to \infty, n \text{ odd}), \end{aligned} \tag{3}$$

in which the constants are given by

$$\begin{aligned} C_0(q) &= \sum_{-\infty}^{\infty} q^{-r^2} \Big/ \prod_{j \geq 1} (1 - q^{-j}), \\ C_1(q) &= \sum_{-\infty}^{\infty} q^{-(r-1/2)^2} \Big/ \prod_{j \geq 1} (1 - q^{-j}). \end{aligned} \tag{4}$$

2. PERMUTATIONS

We first consider the permutations of n letters. It is well known, from Riddell's formula or otherwise, that if $\varphi_n(S)$ is the number of these whose cycle lengths all lie in the set S of positive integers, then

$$\sum_{n \geq 0} \frac{\varphi_n(S)}{n!} x^n = \prod_{s \in S} e^{x^s/s}. \tag{5}$$

Let T be a fixed set of positive integers. If we multiply both sides of (5) by $(-1)^{|T|-|S|}$ and sum over all subsets $S \subseteq T$, we obtain

$$\sum_{n \geq 0} \frac{\psi_n(T)}{n!} x^n = \prod_{t \in T} (e^{x^t/t} - 1), \tag{6}$$

where now $\psi_n(T)$ is the number of n-permutations whose set of cycle lengths is *exactly* T.

Next sum (6) over all sets T such that $|T| = k$, to get

$$\sum_{n \geq 0} \frac{\rho_1(n, k)}{n!} x^n = \text{coeff}_{y^k} \left\{ \prod_{t \geq 1} (1 + y(e^{x^t/t} - 1)) \right\}.$$

Finally,

$$\sum_{n,k\geq 0} \frac{\rho_1(n,k)}{n!} x^n y^k = \prod_{t\geq 1} \{1 + y(e^{x^t/t} - 1)\} \tag{7}$$

is the generating function for $\rho_1(n, k)$, the number of n-permutations that have exactly k different cycle lengths.

Next, let $\bar{c}_n$ denote the average number of different cycle lengths of an n-permutation. If we apply $\partial/\partial y \log$ to (7) and set $y = 1$ we get

$$\sum_{n\geq 0} \bar{c}_n x^n = \frac{1}{(1-x)} \sum_{t\geq 1} (1 - e^{-x^t/t}) \tag{8}$$

and after matching coefficients we find

$$\bar{c}_n = \sum_{m=1}^{n} \frac{(-1)^{m+1}}{m!} \sum_{1\leq t\leq n/m} t^{-m}. \tag{9}$$

The term $m = 1$ contributes $\log n + \gamma + o(1)$ (in fact, that one term is exactly the average number of cycles), while the other terms give

$$\sum_{m=2}^{n} \frac{(-1)^{m+1}}{m!} \sum_{1\leq t\leq n/m} t^{-m} = \sum_{m=2}^{\infty} \frac{(-1)^{m+1}}{m!} \zeta(m) + o(1).$$

The above interchange of limiting processes is justified by the absolute convergence of the double series. Hence the average number of distinct cycle lengths of an n-permutation is

$$\bar{c}_n = \log n + \gamma - Q + o(1),$$

where Q is given by (1), completing the proof of Theorem A. Table I is a table of the $\rho_1(n, k)$.

TABLE I

The Number of n-Permutations with k Cycle-Lengths

k \ n	1	2	3	4	5	6	7	8	9	10
1	1	2	3	10	25	176	721	6406	42,561	436,402
2			3	14	95	424	3269	21,202	178,443	1,622,798
3						120	1050	12,712	141,876	1,418,400
4										151,200

3. Generalizations

The success of the method is traceable to the multiplicativity of the generating function on the right side of (5). Such g.f.'s, however, are very common. Suppose, in general, that $\varphi_n(S)$ is the number of combinatorial objects of "size" n whose "parts" all have sizes that lie in a certain set S. Suppose further that

$$\sum_{n\geq 0} \alpha_n \varphi_n(S)\, x^n = \prod_{s\in S} f_s(x^{-s}), \tag{10}$$

where $\{\alpha_n\}$ is a sequence of universal (=independent of S) constants. Then, as in (6),

$$\sum_{n\geq 0} \alpha_n \psi_n(T)\, x^n = \prod_{t\in T} (f_t(x^t) - 1) \tag{11}$$

and (7) becomes

$$\sum_{n,k\geq 0} \alpha_n \rho(n, k)\, x^n y^k = \prod_{t\geq 1} \{1 + y(f_t(x^t) - 1)\}. \tag{12}$$

This is the generating function for the number $\rho(n, k)$ of objects of size n whose parts have exactly k different sizes.

For another example, let $\rho_2(n, k)$ be the number of partitions of the integer n into not necessarily distinct parts of exactly k different sizes. Then (10) holds with $\alpha_n \equiv 1$ and $f_s(u) = (1 - u)^{-s}$, and so from (12) we get

$$\sum_{n,k\geq 0} \rho_2(n, k)\, x^n y^k = \prod_{t\geq 1} \left\{1 + \frac{yx^t}{1 - x^t}\right\}. \tag{13}$$

For a final example, let $\rho_3(n, k)$ be the number of partitions of the set $[n]$ that have exactly k different sizes of blocks. Then again (10) holds, this time with $f_s(u) = e^{u/s!}$ and $\alpha_n = 1/n!$, and we find

$$\sum_{n,k\geq 0} \frac{\rho_3(n, k)}{n!} x^n y^k = \prod_{t\geq 1} \{1 + y(e^{x^t/t!} - 1)\}. \tag{14}$$

Other examples might include labeled graphs and the distinct sizes of their connected components, rooted forests and the distinct sizes of their trees, etc.

4. The Average Number of Distinct Parts of a Partition of n

If we apply $\partial/\partial y$ log to (13), let $y = 1$, and match coefficients of powers of x, we find easily that the average number of different sizes of parts that occur in a partition of n is exactly

$$\bar{d}_n = (p(0) + p(1) + \cdots + p(n-1))/p(n), \tag{15}$$

where p is the partition function.

Now such a simple formula deserves a simple combinatorial proof, and here is one. For a partition π and an integer r, let $\delta(\pi)$ be the number of distinct parts of π and let $\chi(r, \pi)$ be 1 if r is a part of π, and 0 otherwise. Then

$$\begin{aligned} p(n)\,\bar{d}_n &= \sum_{\pi} \delta(\pi) = \sum_{\pi} \sum_{l \geqslant 1} \chi(l, \pi) \\ &= \sum_{l \geqslant 1} \left\{ \sum_{\pi} \chi(l, \pi) \right\} \\ &= \sum_{l \geqslant 1} p(n-l) \end{aligned}$$

as required.

From the Hardy–Ramanujan estimate

$$p(n) \sim \frac{1}{4\sqrt{3}n} \exp\left\{\pi \sqrt{\frac{2n}{3}}\right\}$$

Theorem B follows easily from (15) by summation (see [3], p. 341). The remark about "almost all" partitions in Theorem B follows from similar estimations applied to the generating function for the second moment.

As a subject for future research we originally suggested the study of the asymptotics of the excess of the average number of blocks over the average number of different block sizes of a partition of $[n]$. Andrew Odlyzko and Bruce Richmond solved that problem, and showed that the average number of different block sizes in a partition of $[n]$ is $\sim e \log n$. Their results will appear in [2].

5. The Number of Vector Subspaces

Here we investigate the sizes of the Gaussian coefficients

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \frac{(q^n - 1)(q^{n-1} - 1) \cdots (q^{n-k+1} - 1)}{(q^k - 1)(q^{k-1} - 1) \cdots (q - 1)} \tag{16}$$

and ıumbers $G_n = \sum_k \left[{n \atop k}\right]_q$. If q is a prime power, then (16) counensional subspaces of n-space over $GF(q)$, and G_n is the total subspaces, regardless of dimension. If q is not a prime powe still counts $k \times n$ matrices in reduced row echelon form over tters [1].

Nrom (16) we have

$$\left[{n \atop k}\right) \Big/ \prod_{1}^{\infty} (1-q^{-j}) \Big) \left[\prod_{n-k+1}^{n} (1-q^{-j}) \right] \left[\prod_{k+1}^{\infty} (1-q^{-j}) \right] \tag{17}$$

and ι once the upper estimate

$$\left[{n \atop k}\right]_q \leqslant q^{k(n-k)} \Big/ \prod_{j \geqslant 1} (1-q^{-j}). \tag{18}$$

Fstimate, we use the fact that

$$\prod (1-x_i) \geqslant 1 - \sum x_i \qquad (0 < x_i < 1)$$

and

$$\prod_{+1} (1-q^{-j}) \geqslant 1 - \sum_{n-k+1}^{n} q^{-j} \geqslant 1 - \sum_{n-k+1}^{\infty} q^{-j}$$

$$= 1 - \frac{q^{-(n-k)}}{(q-1)}$$

and

$$\prod_{k+1}^{\infty} (1-q^{-j}) \geqslant 1 - \frac{q^{-k}}{q-1}.$$

Fin the lower bound

$$\left[{n \atop k}\right]_q \geqslant \left(q^{k(n-k)} \Big/ \prod_{j \geqslant 1} (1-q^{-j})\right) \times \left\{1 - \frac{q^{-k}}{q-1} - \frac{q^{-(n-k)}}{q-1} + \frac{q^{-n}}{(q-1)^2}\right\} \tag{19}$$

forn coefficient.

ly the size of G_n, define the sum

$$S(n) = \sum_{k=0}^{n} q^{k(n-k)}. \tag{20}$$

To estimate $S(n)$, we have

$$S(n) = q^{n^2/4} \sum_{k=0}^{n} q^{-(k-n/2)^2} \tag{21}$$

and if n is even,

$$\sum_{k=0}^{2m} q^{-(k-m)^2} = 1 + 2 \sum_{r=1}^{m} q^{-r^2} \to \sum_{-\infty}^{\infty} q^{-r^2} \qquad (n \to \infty)$$

whence

$$S(n) \sim q^{n^2/4} \left(\sum_{-\infty}^{\infty} q^{-r^2} \right) \qquad (n \to \infty, n \text{ even}). \tag{22}$$

Similarly we find that

$$S(n) \sim q^{n^2/4} \left(\sum_{-\infty}^{\infty} q^{-(r-(1/2))^2} \right) \qquad (n \to \infty, n \text{ odd}). \tag{23}$$

Now we want to sum (19) to estimate G_n from below. We will encounter three sums that need estimating. First we will meet

$$\begin{aligned} S_2(n) &= \sum_{k=0}^{n} q^{k(n-k)-k} \\ &= q^{((n-1)/2)^2} \sum_{k=0}^{n} q^{-(k-(n-1)/2)^2} \\ &= O(q^{-n/2} S(n)) \end{aligned}$$

and similarly we will find

$$S_3(n) = \sum_{k=0}^{n} q^{(k(n-k)-(n-k)} = O(q^{-n/2} S(n)),$$

$$S_4(n) = \sum_{k=0}^{n} q^{k(n-k)-n} = O(q^{-n} S(n)).$$

Hence it follows that

$$G_n = \sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix}_q \geqslant \left(S(n) \Big/ \prod_{1}^{\infty} (1-q^{-j}) \right) \{1 + O(q^{-n/2})\}$$

and

$$G_n \leqslant S(n) \Big/ \prod_{1}^{\infty} (1-q^{-j}),$$

TABLE II

q	$C(q)$	$C(q)$
2	7.36477	7.36475
3	3.01976	3.01824
4	2.18989	2.18281
5	1.84551	1.82955
7	1.53747	1.49939
8	1.45506	1.40538

and therefore

$$\lim_{n\to\infty} \frac{G_n}{S(n)} = \prod_{j=1}^{\infty} (1-q^{-j})^{-1}. \tag{24}$$

If we combine this fact with estimates (22), (23) we find that Theorem C is proved.

The values of the constants $C_0(q)$, $C_1(q)$ in Eq. (4) appear in Table II.

An interesting feature of the asymptotics is the size of the central Gaussian coefficient. Recall that the middle binomial coefficient alone contributes $K/\sqrt{n}$ of the sum of all binomial coefficients of the same order. In the case of the Gaussian coefficients, the distribution is even more peaked. In fact, the central coefficient contributes a fixed positive fraction of the total. Precisely, our estimates above show that

$$\lim_{\substack{n\to\infty\\ n\text{ even}}} \begin{bmatrix} n \\ n/2 \end{bmatrix}_q \Big/ G_n = \left\{ \sum_{-\infty}^{\infty} q^{-r^2} \right\}^{-1},$$

$$\lim_{\substack{n\to\infty\\ n\text{ odd}}} \begin{bmatrix} n \\ (n-1)/2 \end{bmatrix}_q \Big/ G_n = q^{-1/4} \left\{ \sum_{-\infty}^{\infty} q^{-r^2} \right\}^{-1}.$$

For example, it is true that "about 47% of all vector subspaces of n-space over $GF(2)$ have the middle dimension $n/2$, if n is even."

Finally, estimates (18), (19) show that if $k = n/2 + x$, then for the Gaussian coefficients we have the estimate

$$\begin{bmatrix} n \\ (n/2)+x \end{bmatrix}_q \sim q^{n^2/4} q^{-x^2} \Big/ \prod_{j\geq 1} (1-q^{-j}) \qquad (n\to\infty).$$

In other words, the distribution of the Gaussian coefficients is asymptotically normal with bounded variance.

References

1. A. Nijenhuis, A. E. Solow, and H. S. Wilf, Bijective methods in the theory of finite vector spaces, *J. Combin. Theory Ser. A*, to appear.
2. A. M. Odlyzko and L. B. Richmond, On the number of distinct block sizes in partitions of a set, to appear.
3. P. Erdös and J. Lehner, The distribution of the number of summands in the partitions of a positive integer, *Duke Math. J.* **8** (1941), 335–345.

Printed by the St. Catherine Press Ltd., Tempelhof 41, Bruges, Belgium