Applied Mathematical Sciences, Vol. 5, 2011, no. 18, 875 - 881

A New Formula for the Number of Combinations and Permutations of Multisets

Željko Jurić

Department of Mathematics
Faculty of Mathematics and Natural Sciences
University of Sarajevo, Bosnia and Herzegovina
zjuric@utic.net.ba

Harun Šiljak

Department of Automatic Control and Electronics
Faculty of Electrical Engineering
University of Sarajevo, Bosnia and Herzegovina
hs14938@etf.unsa.ba

Abstract

A new formula for finding number of k-combinations of finite multisets is given. Its efficiency is compared to the formula already given by P.A. MacMahon. As a corollary, a formula for finding the number of k-variations of finite multisets is given. Advantages of such formula over the traditional generating functions method are also pointed out.

Mathematics Subject Classification: 05A19

Keywords: Enumerative combinatorics, combination, permutation, multiset

1 Introduction

Question of determining the number of combinations on multisets with finite number of elements is a classical problem in enumerative combinatorics (sometimes formulated as the number of k-combinations of an n-set with repetitions allowed, but with restrictions on the number of repetitions). In this paper we will introduce a new solution for it and compare it with the formula given by P.A. MacMahon in [2].

The first objective is to find the number of k-combinations (and later -permutations) in a multiset $A = \{m_1 \cdot a_1, m_2 \cdot a_2, \ldots, m_n \cdot a_n\}$ where m_i, $i = 1, \ldots, n$ are finite. We will denote this number as $C(k; m_1, m_2, \ldots, m_n)$.

Using this notation, MacMahon's formula from [2] can be written as

$$C(k; m_1, m_2, \ldots, m_n) = \\ = \sum_{p=0}^{n} (-1)^p \sum_{1 \le i_1 \le i_2 \le \cdots \le i_p \le n} \binom{n+k-m_{i_1}-m_{i_2}-\ldots-m_{i_p}-p-1}{n+k-m_{i_1}-m_{i_2}-\ldots-m_{i_p}-p}$$

which is equivallent to the following sum, where the summation is taken over all terms for which $n + k - m_{i_1} - m_{i_2} - \ldots - m_{i_p} - p > 0$

$$C(k; m_1, m_2, \ldots, m_n) = \\ = \sum_{p=0}^{n} (-1)^p \sum_{1 \le i_1 \le i_2 \le \cdots \le i_p \le n} \binom{n+k-m_{i_1}-m_{i_2}-\ldots-m_{i_p}-p-1}{n-1} \quad (1)$$

The total number of terms in (1) is easily obtained after noting that the inner sum has $\binom{n}{p}$ summands (not all are being summed, but all are evaluated), and since $p = 0, 1, \ldots, n$, the number of terms is equal to 2^n.

Since the number of terms in (1) is only dependent of n, this formula is convenient in cases when k is large, and n is small. In the following section we will show that our new formula is more efficient in cases when n is large.

2 The new formula

Theorem 1 *If $M = \max\{m_1, m_2, \ldots, m_n\}$ and $c(i)$ is the number of numbers m_p, $p = 1, \ldots, n$ which are not smaller than i then*

$$C(k; m_1, m_2, \ldots, m_n) = \\ = \sum \binom{c(i_1)}{\lambda_1} \binom{c(i_2)-\lambda_1}{\lambda_2} \cdots \binom{c(i_s)-\lambda_1-\lambda_2-\ldots-\lambda_{s-1}}{\lambda_s} \quad (2)$$

where summation is made over all representations $k = \lambda_1 i_1 + \lambda_2 i_2 + \cdots + \lambda_s i_s$, where $M \ge i_1 > i_2 > \cdots > i_s \ge 1$.

Proof. Observe a representation of integer k as $k = \lambda_1 i_1 + \lambda_2 i_2 + \cdots + \lambda_s i_s$, where $M \ge i_1 > i_2 > \cdots > i_s \ge 1$, λ_p, i_p are positive integers. Note that strict inequality implies that i_p are all different. Every such representation (note that due to the ordering we imposed, there are no representations which differ only in order of summands) encodes exactly one family k-combination in which some λ_1 elements are repeated i_1 times, etc. Now to determine the number of combinations in each family, observe that λ_1 elements which are i_1 times repeated can be chosen in $\binom{c(i_1)}{\lambda_1}$ ways. After that, the λ_2 elements can be chosen in $\binom{c(i_2)-\lambda_1}{\lambda_2}$ ways and so on - the λ_s elements can be chosen in $\binom{c(i_s)-\lambda_1-\lambda_2-\ldots-\lambda_{s-1}}{\lambda_s}$ ways, so by multiplicative principle we obtain (2). □

Number of terms in this formula is equal to the number of partitions of k in terms less or equal than M. Function giving that number is usually called partition function $Q(k, M)$, and its values are given by the sequence A026820 in [3]. Note that $Q(k, M) \leq Q(k, k)$ for all M, so the maximum number of terms in (2) is given by $Q(k, k)$ which is often simply written as $P(k)$ - number of partitions of k without restrictions on the size - values of $P(k)$ are given by the sequence A000041 in [3] (notation taken from [4]).

Using Hardy and Ramanujan's asymptotic relation

$$p(k) \sim \frac{\exp\left(\pi\sqrt{2k/3}\right)}{4k\sqrt{3}} \text{ as } k \to \infty$$

given in [1] we can estimate the number of terms in (2), if needed. Even without the estimate, it is clear that in case of large n and small k, formula (2) is much more efficient than (1).

n	k $(m = k)$	k $(m = \lfloor \frac{k}{2} \rfloor)$	k $(m = \lfloor \frac{k}{3} \rfloor)$	k Hardy approx. $(m = k)$
1	2	4	6	2
2	4	6	8	3
3	6	8	9	5
4	8	9	12	7
5	10	11	12	9
6	12	13	15	11
7	14	15	18	14
8	17	18	20	16
9	20	20	22	19
10	23	23	25	22
11	26	26	28	25
12	29	30	31	29
13	32	33	35	32
14	36	36	38	36
15	40	40	42	39
16	44	44	45	43
17	48	48	50	47
18	52	52	54	52
19	56	57	58	56
20	61	61	63	61
21	66	66	67	65
22	71	71	72	70

Table 1: Border values for k in efficient use of (2)

n	(1)	(2)
1	435	30
2	430	35
3	416	49
4	388	77
5	368	97
6	341	124
7	306	159
8	262	203
9	215	250
10	158	307
11	98	367
12	34	431
≥ 13	0	465
$\sum$	3451 (24.74%)	10499 (75.26%)

Table 2: Efficiency of (1) and (2)

For instance, Table 1 shows border values for k - if k is less than the value in the table for some n, (2) is more efficient than (1). m is taken to be in function of k, with values of k, $\left\lfloor \frac{k}{2} \right\rfloor$, $\left\lfloor \frac{k}{3} \right\rfloor$. Note that the results for various m are close, and that the last column - Hardy-Ramanujan relation approximates the first one for $m = k$ very well for larger k.

Furthermore, Table 2 shows results of efficiency analysis of both formulas. For all n in the interval $1 \leq n \leq 30$, we have checked which formula has less terms for all $1 \leq k \leq 30$, $1 \leq m \leq k$, so for $n = 1$, formula (1) was 435 times the more efficient one, while (2) was 30 times - and so on.

The data in Tables 1 and 2 is presented in corresponding charts (Figures 1 and 2, respectively).

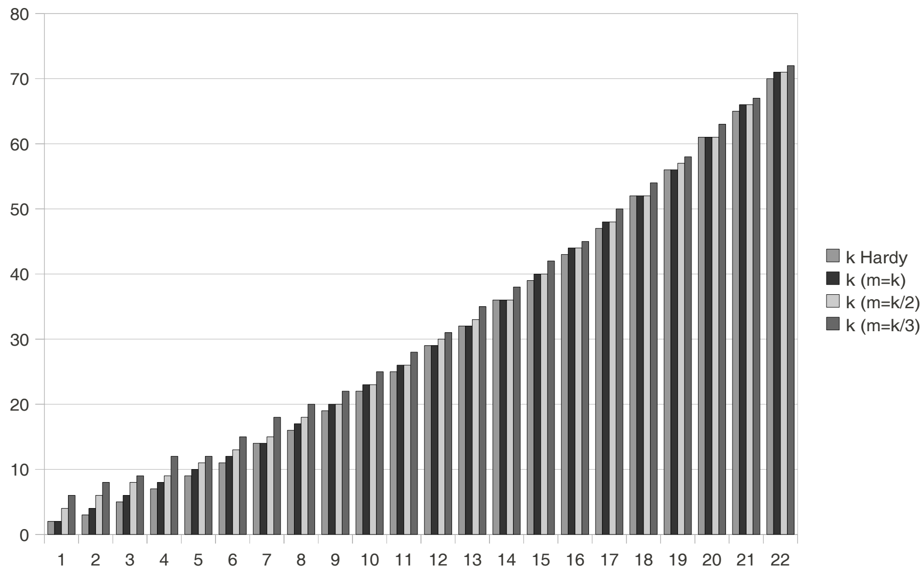

Figure 1: Chart representation of Table 1 data

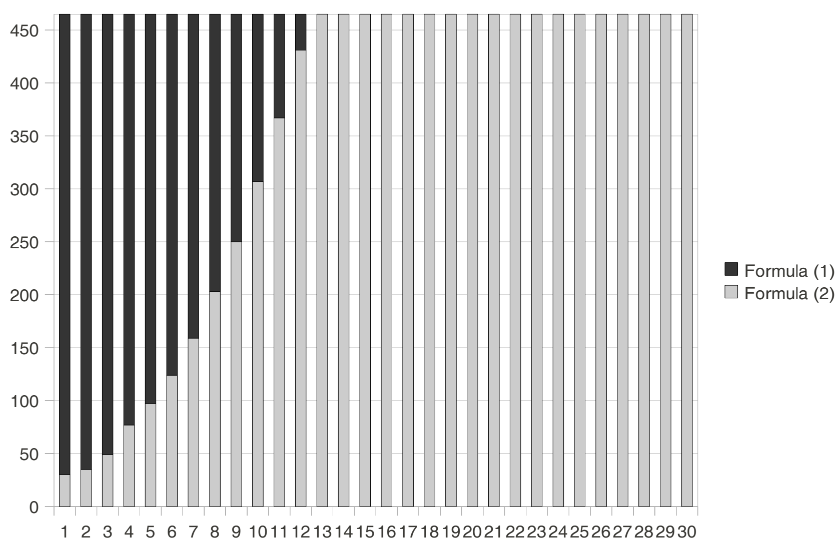

Figure 2: Chart representation of Table 2 data

As a corollary of Theorem 1, we obtain the following formula for k-permutations (note that $P(k; m_1, m_2, \ldots, m_n)$ denotes the number of k-permutations of a multiset with multiplicities m_i):

Theorem 2 *If* $M = \max\{m_1, m_2, \ldots, m_n\}$ *and* $c(i)$ *is the number of numbers* m_p, $p = 1, \ldots, n$ *which are not smaller than* i *then*

$$P(k;m_1,m_2,\dots,m_n)= \\ =\sum \binom{c(i_1)}{\lambda_1}\binom{c(i_2)-\lambda_1}{\lambda_2}\cdots\binom{c(i_s)-\lambda_1-\lambda_2-\dots-\lambda_{s-1}}{\lambda_s}\frac{k!}{i_1!^{\lambda_1}i_2!^{\lambda_2}\cdots i_s!^{\lambda_s}} \tag{3}$$

where summation is made over all representations $k=\lambda_1 i_1+\lambda_2 i_2+\dots+\lambda_s i_s$*, where* $M\ge i_1>i_2>\dots>i_s\ge 1$.

Proof. Starting with (2), we note that it is sufficient in each term of the sum to make all permutations with repetition where i_1 repeats λ_1 times, i_2 λ_2 times, etc. It is known that there are $\frac{k!}{i_1!^{\lambda_1}i_2!^{\lambda_2}\cdots i_s!^{\lambda_s}}$ such permutations, therefore, by multiplicative principle, (3) holds. □

Note that the formula (3) has the same number of terms as the formula (2), and hence appears to be efficient in finding the number of k-permutations.

When we say that, we bear in mind that there is no analogue formula to (1) which could be used for finding the number of k-permutations. Other known formulas used for this task have an enormous number of terms, such as

$$P(k;m_1,m_2,\dots,m_n)=\sum \binom{k}{i_1\quad i_2\quad \dots\quad i_n} \tag{4}$$

where summation is made over $i_1+i_2+\dots+i_n=k$, $0\le i_p\le m_p$, $p=1,\dots,n$. It is easily shown that (4) has $C(k;m_1,m_2,\dots,m_n)$ terms!

Finally, we must note that we are aware of the fact that problems of finding $C(k;m_1,m_2,\dots,m_n)$ and $P(k;m_1,m_2,\dots,m_n)$ are usually efficiently solved using generating functions. That way, $C(k;m_1,m_2,\dots,m_n)$ is found as the coefficient multiplying t^k in the expansion of generating function

$$\varphi(n;m_1,m_2,\dots,m_n;t)=\prod_{i=1}^{n}\sum_{j=1}^{m_i}t^j$$

while $P(k;m_1,m_2,\dots,m_n)$ equals the product of $k!$ and the coefficient multiplying t^k in the expansion of generating function

$$\psi(n;m_1,m_2,\dots,m_n;t)=\prod_{i=1}^{n}\sum_{j=1}^{m_i}\frac{t^j}{j!}$$

We will state four important advantages of formulas (2) and (3) compared to the generating functions procedure:

1. The first advantage is already stated in words 'formula' and 'procedure'. It is not possible to reduce generating functions method in a single formula which gives the result, while (2), on the contrary is exactly that - a single formula, not an algorithmic solution like generating functions.

2. Multiplying all terms in generating functions products is not needed - in each step we need to keep just the terms up to kth power - simple computation shows the number of multiplications is then at most $\frac{(n-1)(k+1)(k+2)}{2}$ multiplications. An analysis shows that our formula is more convenient in case of small k.

3. Following up on the previous argument: space complexity of generating functions procedure is $O(k)$, i.e. all coefficients up to kth have to be memorized. On the other hand, formulas (2) and (3) require no extra storage. That may be crucial when using it on a device with very limited storage capacity, such as programmable calculators - which can make the generating functions inadequate, while (2) and (3) are easily applied.

4. Last but not least: in pre-university combinatorics courses, generating functions are very seldom used, while the emphasis is put on formulas. Therefore, this formula is convenient for presenting in such non-sophisticated courses.

3 Conclusions

The formulas given by (2) and (3) represent a novel method in counting k-permutations and k-combinations of multisets. As we have shown, it is efficient in the applications.

References

[1] G. H. Hardy, S. Ramanujan, Asymptotic Formulae in Combinatory Analysis. Proc. London Math. Soc. 17(1918), 75-115

[2] P. A. MacMahon, Collected Papers - Vol. 1: Combinatorics, MIT Press, 1978.

[3] N. J. A. Sloane, The On-Line Encyclopedia of Integer Sequences, published electronically at http://oeis.org, 2010

[4] E.W. Weisstein, CRC Concise Encyclopedia of Mathematics, second ed., CRC Press, 2003.

Received: September, 2010