A combinatorial approach to a probability problem

R. Fallah-Moghaddam
Department of Computer Science, Garmsar University
P. O. BOX 3588115589, Garmsar, Iran
r.fallahmoghaddam@fmgarmsar.ac.ir

ABSTRACT

Assume that A is a set with M elements. Also, consider that we intend to divide this set into p partitions. Some subsets may even be empty. A probabilistic question arises for us here. The question is:

How many members does the largest subset have on average?

In fact, we want to calculate the value of mathematical expectation of members of the largest subset. By using combinatorial method will prove that the expected value is equal to:

$$(\sum_{i=1}^{M}\sum_{j=1}^{p}(-1)^{j+1}\binom{p}{j}\frac{\binom{M+p-1-ij}{p-1}}{\binom{M+p-1}{p-1}})-1.$$

.

KEYWORDS: Combinatorial method, Probability, Partition, Mathematical expectation.

1 INTRODUCTION

In number theory and combinatorics, a partition of a positive integer n, also called an integer partition, is a way of writing n as a sum of positive integers. Two sums that differ only in the order of their summands are considered the same partition. If order matters, the sum becomes a composition.

A summand in a partition is also called a part. The number of partitions of n is given by the partition function $p(n)$. So $p(4) = 5$.

Partitions can be graphically visualized with Young diagrams or Ferrers diagrams. They occur in a number of branches of mathematics and physics, including the study of symmetric polynomials and of the symmetric group and in group representation theory in general. There are two common diagrammatic methods to represent partitions: as Ferrers diagrams, named after Norman Macleod Ferrers, and as Young diagrams, named after the British mathematician Alfred Young. Both have several possible conventions; here, we use English notation, with diagrams aligned in the upper-left corner.

The asymptotic growth rate for $p(n)$ is given by:

$$\log\big(p(n)\big) \sim C\sqrt{n} \;\; as\; n \to \infty$$

Where,

$$C = \pi\sqrt{\frac{2}{3}}.$$

The more precise asymptotic formula (see [1]):

$$p(n) \sim \frac{1}{4n\sqrt{3}} \exp\left(\pi\sqrt{\frac{2n}{3}}\right) as\; n \to \infty,$$

was first obtained by G. H. Hardy and Ramanujan in 1918 and independently by J. V. Uspensky in 1920. A complete asymptotic expansion was given in 1937 by Hans Rademacher. If A is a set of natural numbers, we let $p_A(n)$ denote the number of partitions of n into elements of A. If A possesses positive natural density α then

$$\log p_A(n) \sim C\sqrt{\alpha n} \quad \text{as} \;\; n \to \infty.$$

and conversely if this asymptotic property holds for $p_A(n)$ then A has natural density α. This result was stated, with a sketch of proof, by Erdős in 1942. (See [2] and [3])

One of the problems related to the partition theory is solving the equation

$$x_1 + x_2 + \cdots + x_r = n, where\; x_i \geq 0.$$

There is a famous formula that shows that the number of answers to this equation is equal to

$$\binom{n + r - 1}{r - 1}.$$

The problem that we are going to solve in this article is, in fact, a probabilistic discussion about the partition of a set.

Assume that A is a set with M elements. Also, consider that we intend to divide this set into p partitions. Some subsets may even be empty. A probabilistic question arises for us here. The question is:

How many members does the largest subset have on average?

In fact, we want to calculate the value of mathematical expectation of members of the largest subset. In the next section, by using combinatorial method will prove that the expected value is equal to:

$$(\sum_{i=1}^{M}\sum_{j=1}^{p}(-1)^{j+1}\binom{p}{j}\frac{\binom{M+p-1-ij}{p-1}}{\binom{M+p-1}{p-1}})-1.$$

2 MAIN RESULTS

As we said in the introduction, there is a close relationship between boundary issues and compositional issues, especially the issue of partition of a set. These boundary topics can easily be generalized to probabilistic topics. Before entering into the main discussion, it is better to pay attention to the following example.

The most beautiful example, is the calculating of the number e using combination tools. Jacob Bernoulli discovered this constant in 1683, while studying a question about compound interest. (See [4]). An account starts with $1.00 and pays 100 percent interest per year. If the interest is credited once, at the end of the year, the value of the account at year-end will be $2.00. What happens if the interest is computed and credited more frequently during the year? Bernoulli noticed that this sequence approaches a limit (the force of interest) with larger n and, thus, smaller compounding intervals. Compounding weekly ($n = 52$) yields $2.692597..., while compounding daily ($n = 365$) yields $2.714567... (approximately two cents more). The limit as n grows large is the number that came to be known as e. That is, with continuous compounding, the account value will reach $2.718281828...

After mentioning this preliminary example, we prepare to prove the main theorem.

Main Theorem. Assume that A is a set with M elements. Also, consider that we intend to divide this set into p partitions. Some subsets may even be empty. Therefore, the value of mathematical expectation of members of the largest subset is equal to:

$$(\sum_{i=1}^{M}\sum_{j=1}^{p}(-1)^{j+1}\binom{p}{j}\frac{\binom{M+p-1-ij}{p-1}}{\binom{M+p-1}{p-1}})-1.$$

Proof. Before the main proof, we need to prove some lemma. First, we claim that the number of integer answers of the equation

$$x_1+x_2+\cdots+x_p=M+p$$

Such that

$$\exists j: i \le x_j, for\ 1\le j\le p \text{ and } i\le M.$$

is equal to

$$\binom{M+p-1}{p-1}-\left(\sum_{j=0}^{p}(-1)^{j}\binom{p}{j}\binom{M+p-1-ij}{p-1}\right).$$

To complete the proof process, we use the principle of inclusion and exclusion. Set $X_t^{(i)}$ equals to solutions of the equation

$$x_1+x_2+\cdots+x_p=M+p$$

Such that

$$i<x_t, \quad \text{and} \quad i\leq M.$$

Using the principle of inclusion and exclusion, we have

$$|X_1^{(i)}\cup X_2^{(i)}\cup\ldots\cup X_p^{(i)}|=\binom{M+p-1}{p-1}-\left(\sum_{j=0}^{p}(-1)^{j}\binom{p}{j}\binom{M+p-1-ij}{p-1}\right),\text{ as we claim.}$$

Thus,

$$P(X_1^{(i)}\cup X_2^{(i)}\cup\ldots\cup X_p^{(i)})=1-\left(\frac{1}{\binom{M+p-1}{p-1}}\sum_{j=0}^{p}(-1)^{j}\binom{p}{j}\binom{M+p-1-ij}{p-1}\right).$$

We define Y_i the set of integer answers of the equation

$$x_1+x_2+\cdots+x_p=M+p$$

Such that

$$\exists j: i\leq x_j, for\ 1\leq j\leq p \quad \text{and} \quad i\leq M.$$

Therefore,

$$P(Y_i)=1-\left(\frac{1}{\binom{M+p-1}{p-1}}\sum_{j=0}^{p}(-1)^{j}\binom{p}{j}\binom{M+p-1-ij}{p-1}\right).$$

Hence,

$$P(Y_1) + P(Y_2) + \cdots + P(Y_M) = \sum_{i=1}^{M} \left(1 - \left(\frac{1}{\binom{M+p-1}{p-1}} \sum_{j=0}^{p} (-1)^j \binom{p}{j} \binom{M+p-1-ij}{p-1} \right) \right).$$

By a simple calculation, we obtain that

$$P(Y_1) + P(Y_2) + \cdots + P(Y_M) = \left(\sum_{i=1}^{M} \sum_{j=1}^{p} (-1)^{j+1} \binom{p}{j} \frac{\binom{M+p-1-ij}{p-1}}{\binom{M+p-1}{p-1}} \right).$$

Using the definition of mathematical expectation, we conclude that $P(Y_1) + P(Y_2) + \cdots + P(Y_M)$ is equal to the value of mathematical expectation of members of the largest subset of solutions of equation

$$x_1 + x_2 + \cdots + x_p = M + p$$

Such that

$$0 < x_i.$$

Therefore, if A is a set with M elements and we divide this set into p partitions, some subsets may even be empty, the value of mathematical expectation of members of the largest subset is equal to:

$$\left(\sum_{i=1}^{M} \sum_{j=1}^{p} (-1)^{j+1} \binom{p}{j} \frac{\binom{M+p-1-ij}{p-1}}{\binom{M+p-1}{p-1}} \right) - 1.$$

Notice that the number of integer answers of the equation

$$x_1 + x_2 + \cdots + x_p = M + p$$

Such that

$$0 < x_i,$$

is equal to the number of integer answers of the equation

$$x_1 + x_2 + \cdots + x_p = M$$

Such that

$$0 \le x_i.$$

□

3 ACKNOWLEDGEMENTS

The author thanks the Research Council of the University of Garmsar for support.

REFERENCES

[1] *Hardy, G.H. (1920).* Some Famous Problems of the Theory of Numbers. *Clarendon Press.*

[2] Nathanson, M.B. (2000). Elementary Methods in Number Theory. Graduate Texts in Mathematics. 195. Springer-Verlag.

[3] Erdős, Pál (1942). "On an elementary proof of some asymptotic formulas in the theory of partitions". Ann. Math. (2). 43 (3): 437–450.

[4] O'Connor, J J; Robertson, E F. "The number e". MacTutor History of Mathematics.