

## Another Useful Representation of Stirling Number for Biology

Adrian Klllogjeri<sup>1</sup> & Pellumb Klllogjeri<sup>2\*</sup>

<sup>1</sup>*Department of Applied Econometrics, Risk Modeller at ZPG, London, United Kingdom*

<sup>2</sup>*Doctor of Philosophy in Mathematics and Computer Sciences, Statistics and Graph Theory, University Aleksander Xhuvani, Department of Mathematics, Elbasan, Albania*

**\*Correspondence to:** Dr. Pellumb Klllogjeri, Doctor of Philosophy in Mathematics and Computer Sciences, Statistics and Graph Theory, University Aleksander Xhuvani, Department of Mathematics, Elbasan, Albania.

### Copyright

© 2019 Dr. Pellumb Klllogjeri, *et al.* This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: 26 December 2018

Published: 13 March 2019

**Keywords:** *Partition of a Set; Partition of Number  $n$  in Accordance with Its Composition with Summands; Classes; Mathematical Biology; 4-Letter Alphabet; Protein Sequence*

### Abstract

In the paper “*Partition of a Set with  $N$  Elements into  $K$  Blocks with Number of Elements in Accordance with the Composition of Number  $N$  As a Sum of Any  $K$  Natural Summands (Another Representation of Stirling Number)*” - published in the International Journal of Advanced Computing [1], we have presented a new formula for computing the number of partitions of the set  $[n]$  into  $k$  blocks with respect to the representation of the number  $n$  as a sum of  $k$  natural summands (different or equal). The Stirling numbers of the second kind describe the number of ways a set with  $n$  elements can be partitioned into  $k$  disjoint, non-empty subsets.

The study is finalized with an extraordinary identity between the second type Stirling number and the number of partitions of number  $n$  in accordance with the composition of number  $n$  with summands. The new representation of Stirling numbers is very useful in biology for computing the number of partitions of number  $n$  which is the same result achieved by Stirling. Our paper is a

specific contribution for specific needs, particularly related to the computational problems in biology. Here will find answer all of those, researching in different fields of life and science, who are concerned about the number of partitions of number  $n$  into  $k$  parts with the same number of elements, with different number of elements or, several parts having the same number of elements and the others different numbers of elements.

## Introduction

Mathematical biology is a fast growing with many modern applications of mathematics. Since biology is becoming more quantitative, the use of mathematics in biology is very important. Recently developed experimental techniques in the biological sciences are generating a huge amount of quantitative data. The new quantitative technology requires methods of generating hypotheses and then testing them that rely heavily on sophisticated mathematical analyses. The application of mathematics in biology involves: making quantitative measurements, the collected biological data is used to develop mathematical models; hence approximate solutions of the models are obtained; the obtained solutions are used to make new predictions. Finally, the predictions can be further tested by new measurements - that is, the cycle is starting anew. Generally, the contribution of mathematics in biology is a mathematical modeling that address a biological process. The mathematical models encompass many biological processes such as neurological and cellular problems, ecology and evolutionary biology. Also, there are clustering applications from many different fields, such as, biological sciences, life sciences, medical sciences, behavioural and social sciences, earth sciences, engineering and information.

“To describe and understand biological properties, and seek precise biological structures, dynamics and interactions it is important to develop sensitive automatic computational methods for the identification of pseudo-periodic regions of sequences to find out typical patterns in biological sequences of multiple pseudo-repeats” [2].

To have a clear idea about this concept we introduce the following considerations.

Let  $A = \{a_1, a_2, \dots, a_m; a_i \in \mathbb{N}\}$  be a finite set of  $m$  letters.  $a_i$  is called a string or a letter of type  $i$  ( $1 \leq i \leq m$ ). Symbolic sequences, characterized by  $A$ , usually have a finite length:  $n$ . The strings play an important role in various fields, such as informatics, dynamical systems, biology, communication theory etc. For instance, the digital information underlying genetics, genomics, biochemistry, cell biology can be represented by a simple string of a 4-letter alphabet (four nucleotides: A, C, G, T) or a 20-letter alphabet (20 amino acids: A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y), which can be regarded as the ‘data structure’ of the life. [2,3].

Consider a protein sequence  $s = \text{AACDDEFGGHHHHIKL}$ . Such sequence has many partitions. Three of them are  $\{\text{AAC, DDEF, GGHH, HHIKL}\}$ ,  $\{\text{AACDDEF, GHHHHIKL}\}$  and  $\{\text{A, A, C, D, D, E, F, G, G, H, H, H, H, I, K, L}\}$ . The concept of pseudo-periodic partition is a generalization of that of atomic periodic partition. Based in this generalized concept, a biologist can investigate the pseudo-periodicity of a sequence that has no fixed weak-periodic pattern and no fixed period length. The algorithm for generating pseudo-periodic partitions of sequences is very useful to detect functional or structural domains and regions

of proteins. The number of data points in any data set is always finite, thereby the number of distinct partitions is finite. The number of different partitions for  $n$  observations into  $K$  groups is a Stirling number of the second kind [4-6], which is given by

$$\left\{ \begin{matrix} n \\ k \end{matrix} \right\} = \sum_{r=1}^k (-1)^{k-r} \frac{n^r}{r!(k-r)!}$$

This is a result from the theory of generating functions that defines the number of partitions of the set  $[n]$  into any  $k$  classes. This shows that enumeration of all possible partitions is impossible for even relatively small problems. In the case when the number of clusters is unknown, the number of different combinations is the sum of the Stirling numbers of the second kind [7,8].

DNA plays a major role in the process of inheritable alterations of a cell. As mentioned above, the digital information in cell biology can be represented by a simple string of a 4- letter alphabet. Erwin Chargaff (1905-2002) was deeply impressed by this discovery, and expressed his feeling saying "for I saw before me in dark contours the beginning of a grammar of Biology" (Chargaff, 1971) [9].

After the development of a method for the precise chemical characterization of nucleic acids, Chargaff, in 1950, observed, using current language, that in any double-stranded DNA segment, the Adenine (A) and Thymine (T) frequencies are equal, and so are the frequencies of Cytosine (C) and Guanine (G). (Chargaff, 1950). the k-word dictionary. The 4-letter alphabet consists of four letters  $\mathcal{A} = \{A, C, G, T\}$  [1,10].

"The properties of partitions of numbers extensively investigated by Hardy and Ramanujan have proved to be of outstanding mathematical interest. The first physical application known to us of the Hardy-Ramanujan asymptotic expression for the number of possible ways any integer can be written as the sum of smaller positive integers is due to Bohr and Kalckar\* for estimating the density of energy levels for a heavy nucleus" [11-13].

## Theoretical Issues of Partitioning

Let us consider the compositions of the natural number  $n$  as a sum of  $k$  natural numbers and, let assume that all possible and different compositions of the natural number  $n$  as a sum of  $k$  natural summands (in the sum can be found different and equal numbers) are those presented in the system of the identities ( $s$  different ways):

$$\begin{cases} n = a_{11} + a_{12} + \dots + a_{1k} \\ n = a_{21} + a_{22} + \dots + a_{2k} \\ \dots \\ n = a_{s1} + a_{s2} + \dots + a_{sk} \end{cases}, \forall a_{ij} \in N \tag{1}$$

All the compositions are different and each one of them represents a way of how  $n$  elements are distributed in  $k$  classes or blocks. In other words, in the system (1) we have all the ways that the set of  $n$  elements can be partitioned into  $k$  classes. All the ways are different from one another. So, in the system (1) are shown all the ways ( $s$  ways in all) that number  $n$  is composed as sum of  $k$  natural numbers. The respective partitions of number  $n$  related to the compositions of number  $n$  we call: *partitions in accordance with the composition of number  $n$  as a sum of  $k$  natural summands* [1].

**Theorem 1:** If all the possible and different compositions of the natural number  $n$  as a sum of  $k$  natural summands are those shown in the system of the equalities (1) then the number of all partitions of number  $n$  into  $k$  classes, where partitions are in accordance with the composition of number  $n$  as a sum of any  $k$  natural summands, is equal to the Stirling number of the second kind [4-6,14,15]. Holds true the identity:

$$\sum_{i=1}^s \left[ \prod_{j=1}^k \frac{1}{n_j - j + 1} \binom{n - \sum_{m=1}^{j-1} a_{im}}{a_{ij}} \right] = \sum_{r=1}^k (-1)^{k-r} \frac{r^n}{r!(k-r)!} \tag{2}$$

where,  $a_{i1} + a_{i2} + \dots + a_{ik} = n, i = 1, 2, \dots, s$  and

$$\frac{1}{n_j - j + 1} = \begin{cases} \frac{1}{n_j - j + 1} & \text{for every group of } a_{i,j} : a_{i,1} = a_{i,2} = \dots = a_{i,n_1}, a_{i,n_1+1} = \dots = a_{i,n_2}, \dots, a_{i,n_{l-1}+1} = \dots = a_{i,n_l} \\ 1 & \text{Otherwise} \end{cases}$$

We have used comma between indices in order to make the distinction between the natural numbers  $i$  and  $n_1, i$  and  $n_1+1$  and so on.

For simplicity, the natural numbers  $n_j$  represent the indices of the summands  $a_{ij}$  in the composition of the number  $n$ . They are such that:

$$n_j = \begin{cases} n_1 & \text{for } j = 1, 2, 3, \dots, n_1 \\ n_2 & \text{for } j = n_1 + 1, \dots, n_2 \\ \dots & \dots \\ n_l & \text{for } j = n_{l-1} + 1, \dots, n_l \end{cases} \quad (3)$$

Based on the order of the equalities in the expression for  $\frac{1}{n_j - j + 1}$ , it is clear that:

$$1 \leq n_1 \leq n_2 \leq \dots \leq n_l \leq k.$$

The above theorem is proved by comparing the sets that are represented by the quantities present on the two sides of the identity. The right side of the identity (2) represents the number of all different partitions of the set  $[n]$  into  $k$  classes including all the types of  $k$  classes. The left side of (2) represents the general number of the partitions of the same set  $[n]$ , but based on the classification of the classes in accordance with the different ways of the composition of the natural number  $n$  as sum of  $k$  summands.

Having into consideration the system of the identities (1) where we have all possible compositions of number  $n$  which are all different from one another, then the left side of (2) is the sum of the numbers of partitions into  $k$  classes and grouped in accordance with  $s$  different ways of the compositions of the natural number  $n$ . We reinforce that the right side of (2) is the general number of the partitions of  $n$  into  $k$  classes, but not grouped in accordance with the above compositions of number  $n$ . The identity (2) reflects two ways of computing the number of partitions of the set with  $n$  objects into  $k$  classes: one way, by counting the partitions in accordance with the different ways of the composition of the natural number  $n$  as sum of  $k$  summands and, the other way, by the method of generating functions [4,6,14,16]. Definitely, the above identity holds true.

**Remark:** The method that is applied to prove this theorem is unusual, it is not similar to the methods used in proving a mathematical statement. For this reason, this theorem is a special one and of special interest.

**Theorem 2:** The number of partitions of the set  $[n]$  into  $k$  classes is independent of the ways the classes are formed.

**Proof:** Consider one of the compositions of the natural number  $n$  shown in the system of equalities (1), which, for simplicity, we denote  $n = a_1 + a_2 + \dots + a_s, a_i \in N$ . Suppose that,  $a_i \neq a_j; \forall i, j \in \{1, 2, \dots, s\}$ . Now, we build

the partitions of number  $n$  into  $k$  classes in accordance with this composition and in accordance with the given order of the summands, i.e., firstly, the class with  $a_1$  elements, secondly, the class with  $a_2$  elements, ....., and lastly, the class with  $a_k$  elements. Then, the number of the partitions of number  $n$  into  $k$  classes in accordance with this composition and this order of its summands is:

$$\binom{n}{a_1} \cdot \binom{n-a_1}{a_2} \cdot \binom{n-a_1-a_2}{a_3} \cdots \binom{a_{k-1}+a_k}{a_{k-1}} \cdot \binom{a_k}{a_k} \tag{4}$$

Let us consider another order of its summands in this composition of  $n$ , for instance,  
 $n = a_i + a_3 + \dots + a_k + a_{k-1} + a_5 + a_2, \quad a_i \in N$

We build the classes of this partition in accordance with this order of the summands. Then, the number of partitions in accordance with the given composition of  $n$  and with the second order of its summands is:

$$\binom{n}{a_i} \cdot \binom{n-a_i}{a_3} \cdots \binom{n-a_i-a_3-\dots-a_k}{a_{k-1}} \cdots \binom{a_5+a_2}{a_5} \cdot \binom{a_2}{a_2} \tag{5}$$

Having into consideration that in the two cases we have the same summands of  $n$ , in the product of the fractions corresponding to the numbers of combinations that are factors in the two expressions (4) and (5), the denominators of the above fractions are of the form  $a_i!$ . The denominators of the two expressions differ only by the order of the factors in the products. So, the results of the productions in the denominators in the two expressions are equal.

The nominators of the fractions corresponding to the numbers of combinations are different, but it is obvious this phenomenon:

At expression (4), the nominator of the fraction corresponding to the number of the first combination has  $a_1$  successive factors starting with the greatest factor  $n$ ; the nominator of the fraction corresponding to the number of the second combination has  $a_2$  successive factors starting with the greatest among them  $n-a_1$ ; the nominator of the third fraction has  $a_3$  successive factors starting with the greatest among them,  $n-a_1-a_2$ ; .....,; the nominator of the last fraction has  $a_k$  successive factors starting with the greatest among them,  $n-a_1-a_2-\dots-a_{k-1} = a_k$  and ending with number 1. Grouping all the factors of the nominators of all the fractions corresponding to the numbers of combinations present in the expression (4) is got the product of the successive natural numbers starting with  $n$  and ending with the number 1. That is, the numerator of (4) is  $n!$

In the expression (5) it is obvious that, the nominator of the first fraction corresponding to the number of the first combination has  $a_i$  successive factors starting with the greatest among them,  $n$ ; the nominator of the

second fraction has  $a_3$  successive factors starting with the greatest among them,  $n-a_i$ ; and so on with the other fractions (as in the case of (4)).

Finally, grouping together all the factors of the nominators of the given fractions, is got the product of the successive natural numbers starting with n and ending with the number 1. The result is the same:  $n!$ . Consequently, the expressions (4) and (5) represent the same number. Making another grouping of the factors of the fractions present in (5), by applying the associative property for the multiplication, and in such a way that the first fraction has as nominator the product of  $a_1$  successive factors starting with number n and as its denominator: the factorial  $a_1!$ ; the second fraction of (5) have as nominator the product of  $a_2$  successive factors starting with the greatest  $n-a_1$  and as denominator the factorial:  $a_2!$ ; ...; continuing the same way with the other fractions, is got the same result: that of equality between (4) and (5). This conclusion is achieved for every other way of the composition of number n shown in the system of equalities (1). This way, it is proved the independence of the number of partitions of the set with n elements into k classes from the way the classes are formed and holds true the equality:

$$\binom{n}{a_1} \cdot \binom{n-a_1}{a_2} \cdot \binom{n-a_1-a_2}{a_3} \dots \binom{a_{k-1}+a_k}{a_{k-1}} \cdot \binom{a_k}{a_k} = \binom{n}{a_i} \cdot \binom{n-a_i}{a_3} \dots \binom{n-a_i-a_3-\dots-a_k}{a_{k-1}} \dots \binom{a_2}{a_2} \tag{6}$$

### Conclusion and the Advantages

The case when the cardinal of a set S (number n) is expanded as a sum of k natural summands is of specific importance. We have partitioned the set of n elements into k classes in accordance with the particular composition of number n. It is considered the combined case: partition of of the set [n] into k classes when number n is sum of k natural summands some of which are equal to one another. We proved the special theorem about an unusual identity, the proof of which is based on comparing the sets that are represented by the quantities present in the identity. In one side of the identity is Stirling number and on the other side is the new representation. The advantage of the new representation of the Stirling number is the following:

By expressing Stirling number in accordance with each composition of number n as a sum of k natural numbers, represented by the left side of the identity (2), we have at hand all the numbers of s different k-partitions [1].

If  $n = a_{11} + a_{12} + \dots + a_{1k}$  then the number of this first type k-partition is:

$$\prod_{j=1}^k \frac{1}{n_j - j + 1} \binom{n - \sum_{m=1}^{j-1} a_{1m}}{a_{1j}}$$

If  $n = a_{21} + a_{22} + \dots + a_{2k}$  then the number of this second type  $k$ -partition is:

$$\prod_{j=1}^k \frac{1}{n_j - j + 1} \binom{n - \sum_{m=1}^{j-1} a_{2m}}{a_{2j}}$$

.....

If  $n = a_{s1} + a_{s2} + \dots + a_{sk}$  then the number of this s type  $k$ -partition is:

$$\prod_{j=1}^k \frac{1}{n_j - j + 1} \binom{n - \sum_{m=1}^{j-1} a_{sm}}{a_{sj}}$$

The paper is an answer to all of those who are concerned about the number of partitions of number  $n$  into  $k$  parts with the same number of elements, with different number of elements or several parts having the same number of elements and the others different numbers of elements. They have to look at different parts of the new formula (LHS of (2)) and use the one they need.

As mentioned in introduction, the strings play an important role specially in biology. The digital information underlying genetics, genomics, biochemistry, cell biology can be represented by strings. Using the concept of partition, a biologist can investigate the pseudo-periodicity of a sequence that has no fixed weak-periodic pattern and no fixed period length. The algorithm for generating pseudo-periodic partitions of sequences is very useful to detect functional or structural domains and regions of proteins. The number of data points in any data set relates to the number of different partitions for  $n$  observations into  $K$  groups which is a Stirling number of the second kind.

### Bibliography

1. Pellumb Kllogjeri & Adrian Kllogjeri (2013). Partition of a Set with  $N$  Elements into  $K$  Blocks with Number of Elements in Accordance with the Composition of Number  $N$  as a Sum of Any  $K$  Natural Summands (Another Representation of Stirling Number, Recent science Publications archives 27702691, *International Journal of Advanced Computing*, 46(3) 1278-1284.

2. Lulang Li, Renchao Jin, Poh-Lin Kok & Honghui (2004). Pseudo-periodic partitions of biological sequences. *Bioinformatics*, 20(3), 295-306.
3. Kong, S. G., Fan, W. L., Chen, H. D., Hsu, Z. T., Zhou, N., *et al.* (2009). Inverse Symmetry in Complete Genomes and Whole-Genome Inverse Duplication. *PLoS one*, 4(11), e7553.
4. Rosen Kenneth, H. (2003). *Discrete Mathematics and its Applications*, Fifth Edition. Published by McGraw-Hill, New York, (pp. 301-349).
5. Slomson Alan (1991). *Introduction to Combinatorics*. 1<sup>st</sup> edition (Chapters on Partitions, Stirling's Approximation Partitions and Generating Functions).
6. Wilf Herbert, S. (1994). *Generating functionology*, Second Edition, 1994 by Academic Press, Inc (pp. 2-23).
7. Anderberg, M. R. (1973). *Cluster analysis for applications*. Academic Press, Inc., London, (p. 376).
8. Hardy, A. (1996). On the number of clusters. *Computational Statistics and Data Analysis*, 23(1), 83-96.
9. Chargaff, E. (1971). Preface to a Grammar of Biology: A hundred years of nucleic acid research. *Science*, 172(3984), 637-642.
10. Chargaff, E. (1950). Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia*, 6(6), 201-209.
11. Auluck, F. C., Chowla, S. & Gupta, G. (1942). On the maximum value of the number of partitions of  $n$  into  $k$  parts. *Journal of the Indian Mathematical Society*, 6, 105.
12. Auluck, F. C. & Kothari, D. S. (1946). Statistical mechanics and the partitions of numbers. *Proceedings of the Cambridge Philosophical Society*, 42(3), 272-277.
13. Anders Bjorner & Richard Stanley, P. (2010). *A Combinatorial Miscellany*. Electronic version, 6-23.
14. Kenneth, P. & Bogart (2004). *Combinatorics Through Guided Discovery*. Preface, 57-80.
15. Regnaud, B. J. T., Allenby, A. B. & Slomson (2011). *How to Count—An Introduction to Combinatorics*. Second Edition, by Taylor and Francis Group LLC, Printed in USA, International Standard Book Number 13, (pp. 51-68).
16. Leo Moser (2004). *An Introduction to the Theory of Numbers*. The Trillia Group, (pp. 1-6).