Alabama Journal of Mathematics

Alabama Journal of Mathematics
Volume 47 (1) (2024), 1 − 22

Original Article

The Stirling Numbers of the Second Kind and Their Applications

S. Bagui*[1], K.L. Mehra[2]

[1] *Department of Mathematics and Statistics, University of West Florida, Pensacola, FL 32514, USA*
[2] *Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB T6G 2G1, Canada*

Abstract

The Stirling numbers of the second kind have numerous applications in mathematics and statistics. But these applications are not generally known to college mathematics and statistics students. This paper aims to propagate the use of the Stirling Number of the second kind to college mathematics and statistics students. We approach the Stirling numbers of the second kind very lucidly so that college students can grasp them easily. We derive a formula for the Stirling numbers of the second kind using the concept of one-to-one and onto functions. Using this formula, we derive a recurrence relation. By the multinomial theorem, we express a monomial r^n in terms of the Stirling numbers of the second kind and falling factorial of order r . Using this generating function, we gave a closed form of the sums of integral powers of integers. Finally, as applications in statistics, we offer closed form for nth (raw) moments of a few discrete distributions such as Binomial, Poisson, Geometric, and Negative Binomial.

Mathematics Subject Classification (2020). 05A15, 05A30, 11B73, 62E15

Keywords. Monomial, multinomial, integer powers, moments, discrete r.v.

1. Introduction

The Stirling numbers $S(n, r)$ of the second kind count the total number of distinct ways to partition a set of n different objects into r unordered - within and between - indistinguishable nonempty cells. In other words, it is the number of different ways to partition a set of n distinct elements into r unlabelled, unordered - within, nonempty subsets. These numbers are named after the Scottish mathematician James Stirling (1692-1770).

*Corresponding Author.
Email addresses: sbagui@uwf.edu
Received: November 15, 2023; Accepted: December 20, 2023

Now we will illustrate $S\left(n,r\right)$ by listing all possible partitions of a small initial set. Suppose we have a set S containing four (4) distinct elements, say, $S=\{A,B,C,D\}$. Consider finding the values of $S\left(4,2\right)$ and $S\left(4,3\right)$ by listing all their partitions. First, consider finding the value of $S\left(4,2\right)$ by manually listing all partitions. This has the following possibilities:

- $\{A\}\{B,C,D\}$, • $\{B\}\{A,C,D\}$, • $\{C\}\{A,B,D\}$, • $\{D\}\{A,B,C\}$,
- $\{A,B\}\{C,D\}$, • $\{A,C\}\{B,D\}$, • $\{A,D\}\{B,C\}$

The above list is partitioning four distinct objects into two *unlabeled* bags. So, by the above listing $S\left(4,2\right)=7$. But, if we consider partitioning four *distinct* objects into two *labelled* nonempty bags, this is done in

$$\begin{aligned}\sum_{\substack{i_1,i_2\geq 1\\ i_1+i_2=4}}\binom{4}{i_1,i_2}&=\binom{4}{1,3}+\binom{4}{2,2}+\binom{4}{3,1}\\&=\frac{4!}{1!3!}+\frac{4!}{2!2!}+\frac{4!}{3!1!}=4+6+4=14.\end{aligned}\tag{1.1}$$

We notice the following pattern here:

$$\sum_{\substack{i_1,i_2\geq 1\\ i_1+i_2=4}}\binom{4}{i_1,i_2}=14=(2)\,(7)=(2!)\,(7)=(2!)\,S\left(4,2\right)$$

Similarly, while considering $S\left(4,3\right)$ all possible partitions are

- $\{A\}\{B\}\{C,D\}$, • $\{A\}\{C\}\{B,D\}$, • $\{A\}\{D\}\{B,C\}$
- $\{B\}\{C\}\{A,D\}$, • $\{B\}\{D\}\{A,C\}$, • $\{C\}\{D\}\{A,B\}$

Thus, $S\left(4,3\right)=6$. Now partitioning four *distinct* abjects into three *labelled* nonempty bags gives the number

$$\sum_{\substack{i_1,i_2,i_3\geq 1\\ i_1+i_2+i_3=4}}\binom{4}{i_1,i_2,i_3}=\binom{4}{1,1,2}+\binom{4}{1,2,1}+\binom{4}{2,1,1}=12+12+12=36.$$

In this case, also, we observe a similar pattern

$$\sum_{\substack{i_1,i_2,i_3\geq 1\\ i_1+i_2+i_3=4}}\binom{4}{i_1,i_2,i_3}=36=(6)\,(6)=(3!)\,(6)=(3!)\,S\left(4,3\right)\tag{1.2}$$

The above relation holds because of changing unlabelled to labelled boxes, i.e., changing unordered arrangements to the ordered sequence.

Harris (1966) was the first to use Stirling numbers to calculate moments in statistics. Joarder and Mahmood (1997) gave an inductive derivation of Stirling numbers of the second kind and employed them to calculate the raw moments of discrete distributions. Kochler (1979) used the Stirling numbers of the second kind to obtain formulas for the moments of the Pearson goodness-of-fit statistic. Benyi and Manago (2005) considered a recursive formla for the moments of the Binomial with success probability equal to 1/2. Knoblauch (2008) utilized Stirling numbers of the second kind when considering raw moments of binomials using moment-generating functions. Very recently, Nguyen (2021) used Stirling numbers for raw momends of Binomial distibution and derived closed forms of sums of powers of integers. Trevino (2018) published a short note on the sums of powers of integers. Griffiths (2013) considered calculating raw moments of binomial recursively by utilizing Stirling numbers of the first kind.

The content of this paper is aimed primarily at senior undergraduate students and their instructors. The article is organized as follows: Section 2 reviews some preliminary results - some important combinatorial identities and inclusion & exclusion theorem - that will be used in the subsequent sections. Section 3 provides a derivation of Stirling numbers of the second kind $S(n, r)$ using the properties of one-to-one and onto functions, a new approach to establishing the well-known recursion relation among them (Theorem 3.1). In Section 4, we prove that the (so-called) 'generating function' of the sequence $\{S(n, r) : 1 \leq r \leq n\}$ equals the monomial r^n (Theorem 4.1), which result is then used to derive a closed (well-known) formula for the sum of the n-th powers of the first k positive integers (Theorem 4.2). Theorem 4.2 offers two proofs - combinatorial and probabilistic - to derive the closed form of the sum of powers of the first k positive integers. Section 5 provides the n-th order raw moments of several integer-valued discrete random variables, and finally, the closing Section 6 provides some concluding remarks.

2. Some Preliminaries

In this section, we state some preliminaries, including three combinatorial Identities 2.1 to 2.3, Multinomial formula, and an Inclusion-Exclusion Theorem 2.2, that are needed in subsequent sections to prove the intended results. For Identity 2.3, we provide a direct combinatorial proof as well as an alternative 'generating function' based supportive proof as verification.

Identity 2.1

Let n and k, $0 \leq k \leq n$, be any non-negative integers. Then the Pascal Triangle formula is the following well-known combinatorial identity

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}, \tag{2.1}$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ ($= 0$ if $n < k$) and $n! = n(n-1)\ldots(3)(2)(1)$, with $0! = 1$.

Identity 2.2

$$k\binom{n}{k} = n\binom{n-1}{k-1}. \tag{2.2}$$

Identity 2.3

An iterated form of Pascal's formula gives the following combinatorial identity

$$\sum_{m=0}^{k} \binom{m}{i} = \binom{k+1}{i+1}. \tag{2.3}$$

(The Identity 2.3 is known as *hockey stick identity* (Blitstein and Hwang (2014)).

We shall prove and add explanations below for only the Identity 2.3. The other two are straightforward and well-known. In the (partial) Pascal's triangle below, each element is the sum of the two numbers diagonally above it. The number $\binom{m}{i}$ appears in the mth row for $i = 0, 1, \ldots, m$. The mth row consists of the following numbers $\binom{m}{0}, \binom{m}{1}, \binom{m}{2}, \ldots, \binom{m}{m}$.

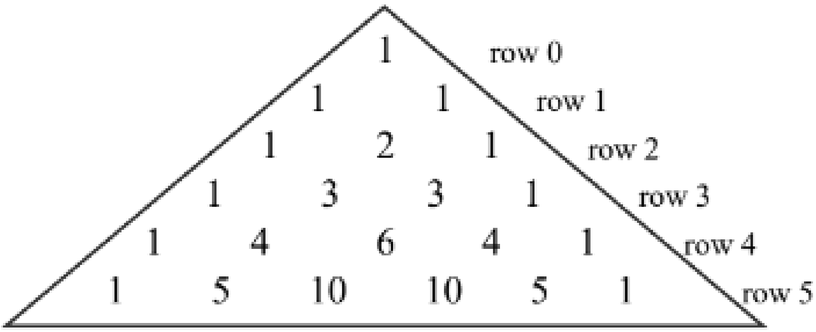


Figure 1. Pascal's Triangle

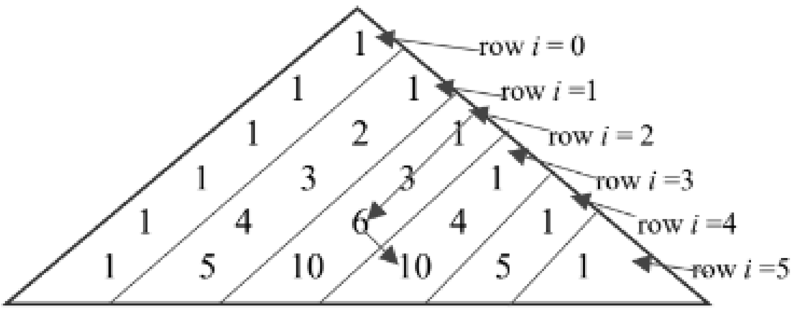


Figure 2. Diagonal split

The *hockey stick identity* can be visualized in the diagonal split of Pascal's triangle (Fig.2). In the i-th diagonal of Fig.2, the elements are $\binom{m}{i}$ for $m = i, i+1, \ldots, k$. In fact the sum of the first r elements of the i-th diagonal is equal to the r-th element of the $(i+1)$th diagonal, which equals to $\binom{k+1}{i+1}$, $k = r + i - 1$. An an illustration, consider in Fig.2 the sum of the first three ($r = 3$) elements of the 3^{rd} ($i = 2$) diagonal, which (in view of the equation (2.3)) equals

$$\binom{2}{2} + \binom{3}{2} + \binom{4}{2} = 1 + 3 + 6 = 10 = \binom{5}{3}.$$

the rth $\left(= 3^{\text{rd}}\right)$ element of the $(i+1)$th (4^{th}) diagonal.

Proof. Using the fact that $\binom{0}{i} = 0$, and a repeated application of Pascal's identity (2.1), we have obtained

$$\begin{aligned}
\sum_{m=0}^{k}\binom{m}{i} &= \binom{0}{i+1}+\binom{0}{i}+\binom{1}{i}+\binom{2}{i}+\binom{3}{i}+\ldots+\binom{k-1}{i}+\binom{k}{i}\\
&= \binom{1}{i+1}+\binom{1}{i}+\binom{2}{i}+\binom{3}{i}+\ldots+\binom{k-1}{i}+\binom{k}{i}\\
&= \binom{2}{i+1}+\binom{2}{i}+\binom{3}{i}+\ldots+\binom{k-1}{i}+\binom{k}{i}\\
&= \binom{3}{i+1}+\binom{3}{i}+\ldots+\binom{k-1}{i}+\binom{k}{i}\\
&= \binom{4}{i+1}+\binom{4}{i}+\ldots+\binom{k-1}{i}+\binom{k}{i}=\ldots\\
&= \binom{k-1}{i+1}+\binom{k-1}{i}+\binom{k}{i}=\binom{k}{i+1}+\binom{k}{i}\\
&= \binom{k+1}{i+1}, \text{ establishing the identity}
\end{aligned}$$

An interesting alternative proof using Generating Function (G.F.) is as follows: The G.F. of the LHS of (2.3) can be written for, $0 < x < 1$, as

$$\begin{aligned}
\sum_{i=0}^{\infty}\left\{\sum_{m=0}^{k}\binom{m}{i}\right\}x^i &= \sum_{m=0}^{k}\left\{\sum_{i=0}^{\infty}\binom{m}{i}x^i\right\} = \sum_{m=0}^{k}(1+x)^m = \frac{(1+x)^{k+1}-1}{1+x-1}\\
&= \frac{1}{x}\left[(1+x)^{k+1}-1\right] = \sum_{j=1}^{k+1}\binom{k+1}{j}x^{j-1}\\
&\text{, (setting } i=j-1)\\
&= \sum_{i=0}^{k}\binom{k+1}{i+1}x^i = \sum_{i=1}^{\infty}\binom{k+1}{i+1}x^i\\
&\left(\text{since } \binom{k+1}{i+1}=0 \text{ for } i>k\right)
\end{aligned}$$

Now equating the coefficients of x^i from the two sides yields at once

$$\sum_{m=0}^{k}\binom{m}{i} = \binom{k+1}{i+1}$$

The proof is complete. □

Theorem 2.1 (The Multinomial Theorem).

$$(x_1 + x_2 + \ldots + x_r)^n = \sum_{\substack{n_1,\ldots,n_r \geq 0 \\ n_1+\ldots+n_r=n}} \binom{n}{n_1, n_2, \ldots, n_r} x_1^{n_1} x_2^{n_2} \ldots x_r^{n_r} \tag{2.4}$$

the sum is over all non-negative integer-valued vectors $(n_1, n_2, \ldots, n_r)$ such that $n_1 + n_2 + \ldots + n_r = n$, and the numbers $\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1! n_2! \ldots n_r!}$ are known as multinomial coefficients.

Theorem 2.2 (Inclusion-Exclusion Principle).

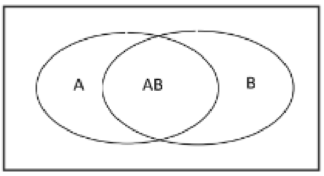
The number of objects of the set S which have at least one of the properties $P_1, P_2, \ldots, P_r$ *is given by*

$$\begin{aligned} |A_1 \cup A_2 \cup \ldots \cup A_r| = \sum_i |A_i| - \sum_{i<j} |A_i \cap A_j| + \sum_{i<j<k} |A_i \cap A_j \cap A_k| - \ldots \\ + (-1)^{r+1} |A_1 \cap A_2 \cap \ldots \cap A_r| \end{aligned} \tag{2.5}$$

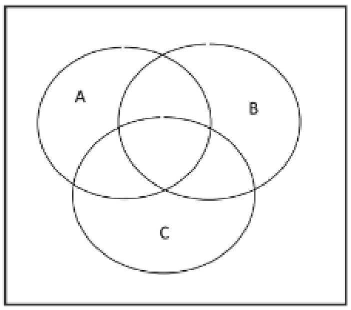
where $A_i = \{x : x \text{ in } S \text{ and } x \text{ has property } P_i\}$, $|A|$ *denotes the number of elements in the set* A, *and* $\sum_i$ *is the sum over* $\binom{n}{1}$ *terms,* $\sum_{i<j}$ *is the sum over* $\binom{n}{2}$ *terms,* $\sum_{i<j<k}$ *is the sum over* $\binom{n}{3}$ *terms, and so forth.*

The idea of the proof can be illustrated via the following two Venn Diagrams.

(i)



(ii)



Denote $A \equiv A_1, B \equiv A_2$, and $C \equiv A_3$.

From the Venn diagram (i), the number of elements in a union of two sets A_1 and A_2 can be written as

$$|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 A_2| = \sum_{i=1}^{2} |A_i| + (-1)^{2+1} P(A_1 A_2).$$

Similarly, from Venn diagram (ii), the number of elements in a union of three sets A_1, A_2, and A_3 can be stated as

$$\begin{aligned} |A_1 \cup A_2 \cup A_3| &= |A_1| + |A_2| + |A_3| - (|A_1 A_2| + |A_1 A_3| + |A_2 A_3|) + |A_1 A_2 A_3| \\ &= \sum_{i=1}^{3} |A_i| - \sum_{i<j}^{\binom{3}{2}} |A_i A_j| + (-1)^{3+1} |A_1 A_2 A_3| \end{aligned}$$

The above results can generalize the total number of elements in the union of r events as a sum of the number of elements in each set individually, minus the sum of the number of elements in each combination of two sets, plus the sum of the number of elements in each combination of three sets, and so on. A detailed proof of this theorem is given below:

Proof. A proof using Mathematical Induction can be completed with the help of identity (2.1). To see this, first note that

$$\begin{aligned} |A_1 \cup A_2| &= |\{x : x \in S \text{ and } x \text{ has at least one of properties } P_1 \text{ and } P_2\}| \\ &= |\{x : x \in S \text{ and has property } P_1\}| + |\{x : x \in S \text{ and has property } P_2\}| \\ &- |\{x : x \in S \text{ and } x \text{ has both properties } P_1 \text{ and } P_2\}| \\ &= |A_1| + |A_2| - |A_1 \cap A_2| \end{aligned} \tag{2.6}$$

so that, in view of (2.6), the assertion (2.5) does hold for $r = 2$. Assume now that his assertion (2.5) holds for the value $r = k$. We shall demonstrate below that then it holds for $r = k + 1$ also: Setting $A_j^* = A_j \cap A_{k+1},\ j = 1, 2, \ldots, k$, and by utilizing the proved and assumed facts, respectively, that equation (2.4) holds for $r = 2$ as well as k, we obtain

$$\begin{aligned} &|A_1 \cup A_2 \cup \ldots \cup A_{k+1}| = |A_1 \cup A_2 \cup \ldots \cup A_k| + |A_{k+1}| \\ &- |\,(A_1 \cup A_2 \cup \ldots \cup A_k) \cap A_{k+1}| \\ &= |A_1 \cup A_2 \cup \ldots \cup A_k| + |A_{k+1}| - |\,(A_1^* \cup A_2^* \cup \ldots \cup A_k^*)\,| \\ &= \left(\sum_{i=1}^{k} |A_i| + |A_{k+1}|\right) - \left(\sum_{1 \le i < j \le k} |A_i \cap A_j| + \sum_{i=1}^{k} |A_i^*|\right) \\ &+ \left(\sum_{1 \le i < j < l \le k} |A_i \cap A_j \cap A_l| + \sum_{1 \le i < j \le k}^{k} |A_i^* \cap A_j^*|\right) - \ldots \\ &+ (-1)^{k+1} \left(|A_1 \cap A_2 \cap \ldots \cap A_k| + \sum_{1 \le i_1 < \ldots < i_{k-1} \le k} |A_{i_1}^* \cap \ldots \cap A_{i_{k-1}}^*|\right) \\ &+ (-1)^{k+2} \left(|A_1^* \cap A_2^* \cap \ldots \cap A_k^*|\right) \\ &= \sum_{i=1}^{k+1} |A_i| - \sum_{1 \le i < j \le k+1} |A_i \cap A_j| + \sum_{1 \le i < j < l \le k+1} |A_i \cap A_j \cap A_l| - \ldots \\ &+ (-1)^{k+1} \sum_{1 \le i_1 < \ldots < i_k \le k+1} |A_{i_1} \cap \ldots \cap A_{i_k}| + (-1)^{k+2} |A_1 \cap A_2 \cap \ldots \cap A_{k+1}| \end{aligned} \tag{2.7}$$

the last equality followed by the utilization for each term in the precceding expression Identity 2.1. The equation (2.7) implies that (2.4) also holds for $r = k + 1$. By induction argument, therefore, (2.4) holds for all positive integers $r \ge 2$. The proof is complete. $\square$

3. Derivation of the Stirling Numbers of the Second Kind, *S(n,r)*

Let A and B be two finite sets, respectively, of n and r distinct elements with $n > r$; and consider all mappings from the set A to the set B. We shall demonstrate below that the 'Stirling Number of the Second Kind' has linkage to and determines the numbers of all 'onto' type and all 'into' type functions from set A to set B. We will derive the expression for $S\,(n, r)$ using the properties of these functions.

First, note that the number of all possible distinct functions from A to B is clearly r^n. Suppose now that we partition set A into r nonempty disjoint blocks (or unordered subsets) and then connect these blocks (or their individual elements) one-to-one to each of the r elements of set B. One such partition, upon permuting the r distinct elements of set B, leads to $r!$ 'onto' mappings. Since $S(n,r)$ is the number of such partitions of A, the total number of all distinct 'onto' mappings from A to B turns out to be $r!S(n,r)$. Accoringly,

$$\left.\begin{array}{l}\text{while the total number of all possible distinct functions}\\ \text{from A to B is r}^n\text{, the number of all 'onto' type function}\\ \text{from A to B is only r! S}(n,r)\end{array}\right\} \tag{3.1}$$

Further, note that the total number of 'into' type function from A to B, viz., those whose range misses at least one element of B, can be obtained from the inclusion-Exclusion Theorem 2.2 (see equation (2.5)) as

$$\binom{r}{1}(r-1)^n-\binom{r}{2}(r-2)^n+\binom{r}{3}(r-3)^n-\ldots+(-1)^r\binom{r}{r-1}1^n \tag{3.1a}$$

The above expression (3.1a) can be derived from (2.5) by defining
$A_1 = \{$ the set of those function from A to B whose range misses
exactly i elements of set $B\}$, and noting that

$$\left.\begin{array}{l}\text{the total number of all distinct functions from A to B}\\ \text{is \textit{equal to} the number of all 'onto' functions from A to B}\\ \text{\textit{plus} the number of all 'into' functions (not onto) from A to B}\\ \text{namely, those whose range misses at least one element of B.}\end{array}\right\} \tag{3.2}$$

Hence, by considerations (3.1) and (3.1a) and equation (3.2), we arrive at

$$\begin{aligned} r^n = &\binom{r}{1}(r-1)^n-\binom{r}{2}(r-2)^n+\binom{r}{3}(r-3)^n-\ldots\\ &+(-1)^r\binom{r}{r-1}1^n+r!S(n,r)\end{aligned} \tag{3.3}$$

which equation, namely, (3.3) at once yields

$$\begin{aligned} S(n,r)=\frac{1}{r!}&\left[r^n-\binom{r}{1}(r-1)^n+\binom{r}{2}(r-2)^n-\binom{r}{3}(r-3)^n+\ldots\right.\\ &\left.+(-1)^{r-1}\binom{r}{r-1}1^n\right]=\frac{1}{r!}\sum_{i=1}^{r}(-1)^i\binom{r}{i}(r-i)^n.\end{aligned} \tag{3.3a}$$

so that by setting $j = r - i$ on the RHS of (3.3a), we get a simpler expression (3.4) for $S(n,r)$, namely,

$$S(n,r)=\frac{1}{r!}\sum_{j=0}^{r}(-1)^{r-j}\binom{r}{j}j^n \tag{3.4}$$

Remark 3.1. The Stirling numbers $S(n,r)$ for all values of n and r always comply with the equalities: $S(n,0)=0,\ S(n,1)=1,\ S(n,n)=1$, and $S(n,r)=0$ for $n<r$. □

3.1. A Recursion Relation

The Stirling numbers $S(n,r),\ 1\le r\le n$, obey the following recursion relation

$$S(n,r)=S(n-1,r-1)+rS(n-1,r),\ r=1,2,\ldots,n \tag{3.5}$$

for all values of n and r. Various proofs are available in the literature (see, e.g. Berman and Fryer (1972), Joarder and Mahmood (1997), Brualdi (2010)). We will use (3.4) directly to prove the recurrence relation (3.5). To our knowledge, no one previously used the formula (3.4) to establish (3.5).

Theorem 3.1. $S(n,r)=S(n-1,r-1)+rS(n-1,r)$.

Proof By equation (3.4), we have

$$\begin{aligned}
S(n,r)&=\frac{1}{r!}\sum_{i=0}^{r}(-1)^{r-j}\binom{r}{j}j^n=\frac{1}{r!}\sum_{j=0}^{r-1}(-1)^{r-j}\binom{r}{j}j^n+\frac{r^n}{r!}\\
&=\frac{1}{r!}\sum_{j=0}^{r-1}(-1)^{r-j}\,r\binom{r-1}{j-1}j^{n-1}+\frac{r^n}{r!},\\
&\text{by Identity 2.2 in equation (2.2)}\\
&=\frac{1}{(r-1)!}\sum_{j=0}^{r-1}(-1)^{r-j}\left[\binom{r}{j}-\binom{r-1}{j}\right]j^{n-1}+\frac{r^n}{r!},\\
&\text{by Identity 2.1 in equation (2.1)}\\
&=\frac{1}{(r-1)!}\sum_{j=0}^{r-1}(-1)^{r-1-j}\binom{r-1}{j}j^{n-1}\\
&+\frac{r}{r!}\sum_{j=1}^{r-1}(-1)^{r-j}\binom{r}{j}j^{n-1}+\frac{r\cdot r^{n-1}}{r!}\\
&=\frac{1}{(r-1)!}\sum_{j=1}^{r-1}(-1)^{r-1-j}\binom{r-1}{j}j^{n-1}+\frac{r}{r!}\sum_{j=0}^{r}(-1)^{r-j}\binom{r}{j}j^{n-1}\\
&=S(n-1,r-1)+rS(n-1,r),\ \text{in view of equation (3.4).}
\end{aligned}$$

The proof is complete. □

The above proof has notable pedagogical value, especially for students (or instructors) who wish to avoid elaborate combinatorial arguments.

Preparing a table showing Stirling numbers of the second kind is straightforward. This can be done by using the recursion formula given in (3.5) for various values of n and r. Below, Table 1 shows $S(n,r)$ for $n,r=1,2,\ldots,7$.

$S(n,r)$'s occur as coefficients in all examples presented in Section 5, where n denotes the row and r the column of Table 1. For Example 1, we read the values of $S(4,1)$, $S(4,2)$, $S(4,3)$, and $S(4,4)$ from Table 1, row 4, as $S(4,1)=1,\ S(4,2)=7,\ S(4,3)=6,\ S(4,4)=1$.

Table 1. Stirling Numbers of the Second Kind, $S(n,r)$

n								
	0	1	2	3	4	5	6	7
0	1	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0
2	0	1	1	0	0	0	0	0
3	0	1	3	1	0	0	0	0
4	0	1	7	6	1	0	0	0
5	0	1	15	25	10	1	0	0
6	0	1	31	90	65	15	1	0
7	0	1	63	301	350	140	21	1

Alternatively Table 1 values can also be derived via equation (3.4) which yields $S(n,r)$ expressions for values of $r = 2,3,4,5,6,7$ as functions of n, given by

$$\begin{aligned}
S(n,2) &= 2^{n-1} - 1 \\
S(n,3) &= \frac{1}{6}\left[3^n - 3(2^n) + 3\right] \\
S(n,4) &= \frac{1}{24}\left[4^n - 4(3^n) + 6(2^n) - 4\right] \\
S(n,5) &= \frac{1}{120}\left[5^n - 5(4^n) + 10(3^n) - 10(2^n) + 5\right] \\
S(n,6) &= \frac{1}{720}\left[6^n - 6(5^n) + 15(4^n) - 20(3^n) + 15(2^n) - 6\right] \\
s(n,7) &= \frac{1}{5040}\left[7^n - 7(6^n) + 21(5^n) + 35(4^n) - 35(3^n) + 21(2^n) - 7\right].
\end{aligned} \tag{3.6}$$

and similarly for higher values of r leading to extended tables of type 1

4. Generating function of the Stirling number *S(n,r)*

Let n and r be any non-negative integers. The Stirling numbers of the second kind can be used to express the monomial r^n of r as a linear function of its so-called 'falling factorial' with Stirling numbers as coefficients: That is, for every non-negative integer n, the monomial r^n can be expressed as

$$r^n = \sum_{i=1}^{n} S(n,i)(r)_i = \sum_{i=1}^{n} S(n,i)\binom{r}{i}(i!), \tag{4.1}$$

where $(r)_i = r(r-1)(r-2)\dots(r-i+1)$ is known (defined) as the 'falling factorial of r up to i-th term'. The equation (4.1) is known as the generating function of the numbers $S(n,1)$, $S(n,2)\dots$, and $S(n,n)$. Multiple proofs are available in the literature (see, e.g., Brualdi (2010), Lint and Wilson(2001)). We will use the Multinomial theorem (Theorem 2.1) to prove this result. To the best of our knowledge, no one in the literature has used the multinomial formula to directly prove (4.1).

Theorem 4.1. Let n and r be any non-negative integers. Then

$$r^n = \sum_{i=1}^{n} S(n,i)(r)_i = \sum_{i=1}^{r} S(n,i)(r)_i, \text{ since } (r_i) = 0 \text{ for } i > r.$$

Proof. Write r^n as $r^n = (\overbrace{1+1+\ldots+1}^{r})^n$. Now, by Multinomial theorem (Theorem 2.1), we have

$$r^n = \sum_{\substack{i_j \geq 0 \\ i_1+\ldots+i_r=n}} \frac{n!}{i_1!i_2!\ldots i_r!} = \sum_{\substack{\text{at least one } i_j=0 \\ i_1+\ldots+i_r=n}} \frac{n!}{i_1!i_2!\ldots i_r!} + \sum_{\substack{\text{all } i_j \geq 1 \\ i_1+\ldots+i_r=n}} \frac{n!}{i_1!i_2!\ldots i_r!} \tag{4.2}$$

(divided into two parts, the first has at least one empty cell, the second has no empty cells).

Let us denote the first term by $U(n,r)$ and the second term by $T(n,r)$, i.e.,

$$U(n,r) = \sum_{\substack{\text{at least one } i_j=0 \\ i_1+\ldots+i_r=n}} \frac{n!}{i_1!i_2!\ldots i_r!} \text{ and } T(n,r) = \sum_{\substack{\text{all } i_j \geq 1 \\ i_1+\ldots+i_r=n}} \frac{n!}{i_1!i_2!\ldots i_r!} \tag{4.3}$$

Thus we rewrite (4.2) as

$$r^n = U(n,r) + T(n,r) \tag{4.4}$$

The number $T(n,r)$ can be viewed as the number of partitions of n *distinct* elements into r nonempty *ordered* blocks (or cells), wheras the Stirling number of the second kind $S(n,r)$ partitions n *distinct* elements into r *unordered* blocks (or cells). By ordering the r blocks in $S(n,r)$, we get $T(n,r)$ so that

$$T(n,r) = r!S(n,r) = S(n,r)(r)_r. \tag{4.5}$$

The number $U(n,r)$, on the other hand, stands for the sum of total of all partitions of n distinct elements into exactly j distinct cells corresponding to each $j = 1, 2, \ldots, (r-1)$ (i.e., with at least one empty block). We let j, $1 \leq j \leq r-1$, denote the number of non-zero among the non-negative integers $\{i_1, i_2, \ldots, i_r\}$ adding up to n, i.e., $i_1 + i_2 + \ldots + i_r = n$. This is possible in $\binom{r}{j}$ ways and, paying regard to their position, permuting these non-zero integers gives $\binom{r}{j}j!$ as the total number of such ways for each $j = 1, 2, \ldots, (r-1)$. Corresponding to each permutation of these possible set of j positive integral slots $(1 \leq j \leq r-1)$ among $\{i_1, i_2, \ldots, i_r\}$, there are $S(n,j)$ (known as the Stirling numbers of the second kind) ways n distinct elements can be partitioned into these j unordered nonempty cells with counts adding up to n. In view of the above explanation, we have

$$U(n,r) = \sum_{j=1}^{r-1} S(n,j) \binom{r}{j} j! = \sum_{i=1}^{r-1} S(n,j)(r)_j. \tag{4.6}$$

Thus, by (4.4), (4.5) , and (4.6), we have

$$\begin{aligned} r^n &= \sum_{i=1}^{r-1} S(n,i)(r)_i + S(n,r)(r)_r = \sum_{i=1}^{r} S(n,i)(r)_i \\ &= \sum_{i=1}^{n} S(n,i)(r)_i, \text{ since } (r)_i = 0 \text{ for } i > r. \end{aligned} \tag{4.6a}$$

The proof is complete. □

4.1. Sums of Integral Powers of Integers

The sum of integral powers of the first k natural numbers, that is, $1^n + 2^n + 3^n + \ldots + k^n$, $(n \geq 1)$, has been a topic of (mathematical) discussion for centuries. Many authors have worked on this topic. For related literature, the readers are encouraged to consult Beardon (1996), Mackiw (2000), Nguyen (2021), Spivey (2021), Schultz (1980), and Trevino (2018).

In this section, we will derive a closed form for the sums of the n-th power of first k natural numbers. We employ two methods. The first method we consider will be a combinatorial approach using the generating function (4.1). The second method will be a probabilistic method using a discrete probability distribution. Because of the simplicity of these methods, they can be easily adopted in classroom teaching.

Theorem 4.2.

$$1^n + 2^n + 3^n + \ldots + k^n = \sum_{i=1}^{n} i! S(n,i) \binom{k+1}{i+1} \tag{4.7}$$

Proof. Below, we offer two proofs: (I). A combinatorial proof, and (II). A Probabilistic proof.

(I). *(Direct combinatorial Approach)*. So far, no one has used the generating function (4.1) directly to prove the identity (4.7). By equation (4.1) (Theorem 4.1), we have

$$\begin{aligned}1^n + 2^n + 3^n + \ldots + k^n &= \sum_{r=1}^{k} r^n = \sum_{r=1}^{k}\left\{\sum_{i=1}^{n} S(n,i)(i!)\binom{r}{i}\right\}\\ &= \sum_{i=1}^{n} S(n,i)(i!)\sum_{r=1}^{k}\binom{r}{i}\\ &= \sum_{i=1}^{n} i! S(n,i)\binom{k+1}{i+1},\\ &\text{by the Identity 2.3 in equation (2.3).}\end{aligned}$$

This completes the first proof. □

(II). *(Probabilistic Approach)*. Let X be a discrete random variable (r.v.) with a probability function $P(X = x)$ with support $x = 0, 1, 2, \ldots$. Thus, by equation (4.1), we write

$$X^n = \sum_{i=1}^{n} S(n,i)(X)_i\,. \tag{4.8}$$

By taking expectations on both sides of (4.8), we establish a general relationship between n-th (raw) moments and i-th factorial moments of a discrete r.v. X, which is given below.

$$E(X^n) = \sum_{i=1}^{n} S(n,i)\, E\left[(X)_i\right]. \tag{4.9}$$

The expected value $E\left[(X)_i\right]$ in (4.9) denotes the i-th factorial moment of the r.v. X , and it is defined by

$$E\left[(X)_i\right] = E\left[X(X-1)\ldots(X-i+1)\right] = \sum_{x=0}^{\infty} x(x-1)\ldots(x-i+1)\, P(X=x) \tag{4.9a}$$

It is always straightforward to calculate the factorial moments of a discrete r.v. Let's consider the following probability mass function (pmf) of a r.v. X,

$$P(X = x) = \frac{1}{(k+1)}, \text{ for } x = 0, 1, \ldots, k. \tag{4.10}$$

The i-th factorial moment of the above pmf can be calculated as

$$\begin{aligned} E\left[(X)_i\right] &= \sum_{x=0}^{k} x(x-1)\ldots(x-i+1)\left(\frac{1}{(k+1)}\right) \\ &= \frac{1}{k+1}\sum_{x=i}^{k} x(x-1)\ldots(x-i+1) = \frac{1}{k+1}\sum_{x=i}^{k} i!\binom{x}{i} \\ &= \frac{i!}{k+1}\sum_{x=i}^{k}\binom{x}{i} = \frac{1}{k+1}i!\binom{k+1}{i+1}, \text{ by Identity 2.3 in equation (2.3).} \end{aligned} \tag{4.11}$$

Thus, by (4.9) and (4.11), we have

$$E(X^n) = \frac{1}{k+1}\sum_{i=1}^{n} i!S(n,i)\binom{k+1}{i+1}. \tag{4.12}$$

On the other hand, the n-th (raw) moment of X can be directly calculated as

$$\begin{aligned} E(X^n) &= \sum_{x=0}^{k} x^n P(X = x) = \sum_{x=0}^{k} x^n \left(\frac{1}{k+1}\right) \\ &= \frac{1}{k+1}\left(1^n + 2^n + \ldots + k^n\right). \end{aligned} \tag{4.13}$$

By equating (4.12) and (4.13), we have the desired result that

$$1^n + 2^n + 3^n + \ldots + k^n = \sum_{i=1}^{n} i!S(n,i)\binom{k+1}{i+1}.$$

The probabilistic proof is complete. □

Remarks 4.1

(i) When $n = 1$, the simplified version of (4.7) is given by

$$1 + 2 + 3 + \ldots + k = \sum_{i=1}^{1} S(1,i)(i!)\binom{k+1}{i+1} = \binom{k+1}{2} = \frac{k(k+1)}{2} \tag{4.14}$$

(ii) Simplified versions of (4.7) for $n = 2, 3, 4$ are given in the Appendix.

(iii) The pmf (4.10) can be viewed as the mixture distribution of Binomial and Uniform (Beta), i.e.,

$$P(X = x) = \int_0^1 \binom{k}{x} p^x (1-p)^{k-x}\, dp = \frac{1}{(k+1)}, \; x = 0, 1, \ldots, k.$$

Another way of deriving it would be: Throw randomly $k+1$ white balls onto the unit interval $(0,1)$, then choose one ball at random and paint it blue. Let X denote the number of white balls to the left of the blue ball. Since any one of the $k+1$ balls is equally likely to be selected for painting, thus

$$P(X=x)=\frac{1}{(k+1)},\ x=0,1,\ldots,k.$$

(see Blitzstein and Hwang (2014, p.353)).

5. Applications of *S(n,r)* in Calculating *n*-th Raw Moment of Discrete Random Variables

Moments provide information about the distribution's location, spead, shape, and peakedness. The location and spread depend on the first two moments, whereas shape and peakedness depend on the third and fourth moments. Generally, higher-order moments are not emphasized very much while reaching this topic. Standard textbooks such as Ghahramani (2019), Ross (2002), and Larson and Marx (2017) generally stop at the second moment. Calculating higher-order raw momenrs of discrete r.v.s could be very difficult or may not be possible. They are typically derived using recursion relation on raw moments or by taking derivatives of the moment-generating functions (mgf), Ross (2002). But computing higher order derivatives of mgfs can be very tedious with increasing n. However, deriving the factorial moments of integer-valued r.v.'s is very straightforward. The equation (4.9) shows n-th raw moments can be directly computed by combining Stirling numbers and the factorial moments. The advantage of this method is that it avoids calculating higher-order derivatives or cumbersom computations with recursion relations. The raw moments given by (4.9) are non-recursive closed-form for n-th raw moments of integer-valued r.v.'s. Also, raw moments can be used to get central moments. The following relationship holds between central moments and raw moments. The n-th central moments of an r.v. X is given by

$$E(X-\mu)^n=\sum_{i=0}^{n}(-1)^i\binom{n}{i}\left[E\left(X^i\right)\right]\mu^{n-i}, \tag{5.1}$$

where $\mu=E(X)$ represents the location, $\sigma^2=E(X-\mu)^2$ represents spread, Skew$(X)=E\left(\frac{X-\mu}{\sigma}\right)^3$ represents skewness, and Kurt$(X)=E\left(\frac{X-\mu}{\sigma}\right)^4-3$ represents the peakedness of distribution (see Blitzstein and Hwang (2014)).

(5.2)

Below, we illustrate n-th raw moments of Binomial, Poisson, Geometric, and Negative-Binomial r.v.'s.

Example 1. *(The Binomial).* Let X be a Binomial r.v. with parameters inetgers $k>0$ and p, $0<p<1$. The pmf of X is given by

$$P(X=x)=\binom{k}{x}p^x q^{k-x},\ \text{where } q=1-p,\ x=0,1,2,\ldots,k. \tag{5.3}$$

The i-th factorial moments of the r.v. X can be derived as

$$\begin{aligned} E\left[(X)_i\right] &= \sum_{x=0}^{k} \left[x(x-1)\dots(x-i+1)\right] \frac{k!}{x!\,(k-x)!} p^x (1-p)^{k-x} \\ &= \sum_{x=i}^{k} \frac{k!}{(x-i)!\,(k-x)!} p^x (1-p)^{k-x} \\ &= (k)_i\, p^i \sum_{y=0}^{k-i} \frac{(k-i)!}{y!\,(k-i-y)!} p^y (1-p)^{k-i-y},\ (y = x-i) \\ &= (k)_i\, p^i. \end{aligned} \tag{5.4}$$

Thus, by combining (4.9) and (5.4), the n-th Binomial moment is given by

$$E\left(X^n\right) = \sum_{i=1}^{n} S\left(n,i\right)(k)_i\, p^i. \tag{5.5}$$

The first four raw moments can be used for calculating a distribution's location, spread, skewness, and kurtosis via central moments. Below, we list the first four raw moments of the Binomial.

By equation (5.5) and Table 1, we have

$$\begin{aligned} &1.\ E(X) = S\left(1,1\right)(k)p = kp \\ &2.\ E\left(X^2\right) = S\left(2,1\right)(k)_1\, p + S\left(2,2\right)(k)_2\, p^2 = kp + (k)_2\, p^2 \\ &3.\ E\left(X^3\right) = S\left(3,1\right)(k)_1\, p^1 + S\left(3,2\right)(k)_2\, p^2 + S\left(3,3\right) p^3 \\ &\qquad\qquad = (k)_1\, p^1 + 3\,(k)_2\, p^2 + (k)_3\, p^3 \\ &4.\ E\left(X^4\right) = S\left(4,1\right)(k)_1\, p^1 + S\left(4,2\right)(k)_2\, p^2 + S\left(4,3\right) p^3 + S\left(4,4\right)(k)_4\, p^4 \\ &\qquad\qquad = kp + 7\,(k)_2\, p^2 + 6\,(k)_3\, p^3 + (k)_4\, p^4. \end{aligned} \tag{5.6}$$

Example 2. *(The Poisson).* Let X be a Poisson r.v. with parameter λ. The pmf of X is given by

$$P\left(X = x\right) = \frac{e^{\lambda}\lambda^x}{x!},\ x = 0, 1, 2, \dots \tag{5.7}$$

The i-th factorial moments can be computed as

$$\begin{aligned} E\left[(X)_i\right] &= \sum_{x=0}^{\infty} \left[x\left(x-1\right)\dots\left(x-i+1\right)\right] \frac{e^{-\lambda}\lambda^x}{x!} \\ &= \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^x}{(x-1)!} = \lambda^i \sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda^y}{y!} = \lambda^i. \end{aligned} \tag{5.8}$$

Thus, by the equations (4.9) and (5.8), we have

$$E\left(X^n\right) = \sum_{i=1}^{n} S\left(n,i\right)\lambda^i. \tag{5.9}$$

When $\lambda = 1$ it gives the *Bell number* $B_n = \sum_{i=1}^{n} S(n,i)$, which is the number of partitions of n distinct elements into nonempty indistinguishable (unordered) boxes, the number of boxes is not specified, but it can't exceed n. It is easy to see from Table 1 that $B_1 = 1, B_2 = 2, B_3 = 5, B_4 = 15, \ldots$. This number has applications in graph theory. It also has many real-life applications. For example, it can be used in counting the number of rhyme schemes for a stanza in poetry. A two-line stanza has $B_2 = 2$ possible rhyme schemes. There are $B_2 = 2$ ways two persons can sleep in unlabeled beds. In the *Tale of Genji* by Lady Shikibo Murasaki, the Japanese used diagrams depicting possible rhyme schemes of a five-line stanze $B_5 = 52$ as early as 1000 AD (see Gardner (1978, p.28)).

The first four raw moments of this distribution are listed below. For Stirling numbers, we used Table 1.

$$\begin{aligned}
1.\ E(X) &= S(1,1)\,\lambda = \lambda \\
2.\ E(X^2) &= S(2,1)\,\lambda + S(2,2)\,\lambda^2 = \lambda + \lambda^2 \\
3.\ E\left(X^3\right) &= S(3,1)\,\lambda + S(3,2)\,\lambda^2 + S(3,3)\,\lambda^3 = \lambda + 3\lambda^2 + \lambda^3 \\
4.\ E\left(X^4\right) &= S(4,1)\,\lambda + S(4,2)\,\lambda^2 + S(4,3)\,\lambda^3 + S(4,4)\,\lambda^4 \\
&= \lambda + 7\lambda^2 + 6\lambda^3 + \lambda^4.
\end{aligned} \tag{5.10}$$

Example 3. *(The Geometric)* Let X be a Geometrix r.v. with parameter p. Let X denote the number of failures before the first success. Thus, the pmf of X is given by

$$P(X = x) = pq^x,\ x = 0, 1, 2, \ldots,\ q = 1 - p. \tag{5.11}$$

The i-th factorial moments of the r.v. X can be derived as

$$\begin{aligned}
E\left[(X)_i\right] &= p \sum_{x=0}^{\infty} \left[x(x-1)\ldots(x-i+1)\right] q^x = p \sum_{x=i}^{\infty} \frac{x!}{(x-i)!} q^x \\
&= i!pq^i \sum_{y=0}^{\infty} \frac{(y+i)!}{y!i!} q^y = i!pq^i (1-q)^{-(i+1)} = i! \left(\frac{q}{p}\right)^i.
\end{aligned} \tag{5.12}$$

Thus, by (4.9) and (5.12), we have

$$E(X^n) = \sum_{i=1}^{n} S(n,i) \left(\frac{i!q^i}{p^i}\right). \tag{5.13}$$

Using (5.13) and Stirling numbers from Table 1, we list this distribution's first four raw moments below:

$$
\begin{aligned}
&1.\ E(X) = S(1,1)\left(\frac{q}{p}\right) = \frac{q}{p} \\
&2.\ E\left(X^2\right) = S(2,1)\left(\frac{q}{p}\right) + S(2,2)(2!)\left(\frac{q^2}{p^2}\right) = \left(\frac{q}{p}\right) + 2\left(\frac{q^2}{p^2}\right) \\
&3.\ E\left(X^3\right) = S(3,1)\left(\frac{q}{p}\right) + S(3,2)(2!)\left(\frac{q^2}{p^2}\right) + S(3,3)(3!)\left(\frac{q^3}{p^3}\right) \\
&\qquad = \left(\frac{q}{p}\right) + 6\left(\frac{q^2}{p^2}\right) + 6\left(\frac{q^3}{p^3}\right) \\
&4.\ E\left(X^4\right) = S(4,1)\left(\frac{q}{p}\right) + S(4,2)(2!)\left(\frac{q^2}{p^2}\right) \\
&\qquad + S(4,3)(3!)\left(\frac{q^3}{p^3}\right) + S(4,4)(4!)\left(\frac{q^3}{p^3}\right) \\
&\qquad = S(4,1)\left(\frac{q}{p}\right) + 14\left(\frac{q^2}{p^2}\right) + 36\left(\frac{q^3}{p^3}\right) + 24\left(\frac{q^3}{p^3}\right).
\end{aligned}
\tag{5.14}
$$

Example 4. *(The Negative Binomial).* Let X be a Negative Binomial r.v. with parameters p and k. Let X denote the number of failures before the k-th success. Thus, the pmf of X is given by

$$P(X = x) = p^k \binom{x+k-1}{x} q^x,\ x = 0, 1, 2, \ldots \tag{5.15}$$

The i-th factorial moment of this distribution is given by

$$
\begin{aligned}
E\left[(X)_i\right] &= p^k \sum_{x=0}^{\infty} \left[x(x-1)\ldots(x-i+1)\right] \frac{(x+k-1)!}{y!\,(k-1)!} q^x \\
&= p^k \sum_{x=i}^{\infty} \frac{(x+k-1)!}{(x-i)!\,(k-1)!} q^x = p^k \sum_{y=0}^{\infty} \frac{(y+i+k-1)!}{y!\,(k-1)!} q^{y+i},\ y = x - i \\
&= (i+k-1)_i\, p^k q^i \sum_{y=0}^{\infty} \frac{(y+i+k-1)!}{y!\,(i+k-1)!} q^y = (i+k-1)_i\, p^k q^i (1-q)^{-(i+k)} \\
&= (i+k-1)_i \left(\frac{p^i}{q^i}\right).
\end{aligned}
\tag{5.16}
$$

Thus, by (4.9) and (5.16), we have

$$E\left(X^n\right) = \sum_{i=1}^{n} S(n,i)\,(i+k-1)_i \left(\frac{q^i}{p^i}\right). \tag{5.17}$$

Using (5.13) and Stirling numbers from Table 1, we list this distribution's first four raw moments below.

1. $E(X) = S(1,1)(k)_1 \left(\frac{q}{p}\right) = \left(\frac{kq}{p}\right)$

2. $$E\left(X^2\right) = S(2,1)(k)_1 \left(\frac{q}{p}\right) + S(2,2)(k+1)_2 \left(\frac{q^2}{p^2}\right)$$
$$= (k)_1 \left(\frac{q}{p}\right) + (k+1)_2 \left(\frac{q^2}{p^2}\right)$$

3. $$E\left(X^3\right) = S(3,1)(k)_1 \left(\frac{q}{p}\right) + S(3,2)(k+1)_2 \left(\frac{q^2}{p^2}\right) + S(3,3)(k+2)_3 \left(\frac{q^3}{p^3}\right)$$
$$= (k)_1 \left(\frac{q}{p}\right) + 3(k+1)_2 \left(\frac{q^2}{p^2}\right) + (k+2)_3 \left(\frac{q^3}{p^3}\right)$$

4. $$E\left(X^4\right) = S(4,1)(k)_1 \left(\frac{q}{p}\right) + S(4,2)(k+1)_2 \left(\frac{q^2}{p^2}\right)$$
$$+ S(4,3)(k+2)_3 \left(\frac{q^3}{p^3}\right) + S(4,4)(k+3)_4 \left(\frac{q^4}{p^4}\right)$$
$$= (k)_1 \left(\frac{q}{p}\right) + 7(k+1)_2 \left(\frac{q^2}{p^2}\right) + 6(k+2)_3 \left(\frac{q^3}{p^3}\right) + (k+3)_4 \left(\frac{q^4}{p^4}\right).$$

6. Concluding Remarks

The Stirling numbers of the second kind are a useful and important set of mathematical numbers. They appear naturally in relating positive integral powers of positive integers to their falling factorials. This relation is given by the equation (4.1), with the RHS expression known as the generating function of the Stirling numbers $S(n,i)$'s. We proved this relation using the multinomial theorem. To the best of our knowledge, no one had earlier used this theorem directly for the proof. Additionally, using this generating function (4.1), we demonstrated a natural connection between $S(n,i)$'s and the sum of n-th powers of the first k positive integers and, thereby, how to compute the n-th (raw) moments of discrete random variables. Many researchers in the past worked on the sum of powers of positive integers (see Section 4.1). But, their proofs were beyond the level of undergraduate students. However, we have presented a proof that is easily understandable at the undergraduate level. Our proof is based on only the Generating Function (4.1) and the Hockey Stick Identity (2.3). We also have given an alternative probabilistic proof using a discrete uniform probability distribution. Because of our proof's simplicity, they can easily be employed in classroom teaching.

Deriving higher-order (raw) moments of discrete random variables is not easy. It is, however, straightforward to calculate the factorial moments of integer-valued random variables. we established a general relationship between the n-th (raw) moments and the i-th factorial moments using Stirling numbers, as given by (4.9), that yields higher-order (raw) moments readily. Using the n-th raw moments, one can derive the n-th central moments using equation (5.1). Higher-order moments, such as 3^{rd} and 4^{th}-order central moments, are needed to calculate the skewness and kurtosis of a population, and the corresponding sample statistic can be used to test the normality of data. Also, we utilized the n-th moment of a discrete uniform distribution to derive in closed form the sum of powers of the first k natural numbers. It is conventional to use raw moments to calculate the estimators based on the method of moments.

Finally, in Section 5, we derived in general closed-form, as examples, the expressions for the n-th raw moments for each of the Binomial, Poisson, geometric, and Negative Binomial distributions and also their explicit simplified expressions for $n = 1, 2, 3, 4$. Using equation (5.1) and these raw moments, we obtained central moments for $n = 2, 3, 4$ for each of these distributions and listed them in the Appendix. This article is aimed at senior undergraduate mathematics or statistics students and their instructors, who will find it informative and valuable.

Acknowledgements

The authors thank the editor and the referee for their constructive feedback on the earlier version of the paper, which led to the present improved version. We would also like to thank Daniel Bozeman for typing this paper in LaTeX.

References

[1] Beardon, A.E. (1996). Sums of powers of integers, *The American Mathematical Monthly*, 103, 201-213.

[2] Brualdi, R.A. (2010). *Introductory combinatorics*, 5^{th} edition, Pearson Prentice Hall, New Jersey.

[3] Berman, G. and Fryer, K.D. (1972). *Introduction to Combinatorics*, Academic Press, New York.

[4] Benyi, A. and Manago, S.M. (2005). A recursive formula for moments of a binomial distribution, *The College Mathematics Journal*, 36(1), 68-72.

[5] Gardner, M. (1978). Mathematical games, *Scientific American*, **238**(5), 24-32.

[6] Blitzstein, J.K. and Hwanng, J. (2014). *Introduction to Probability*, CRC Press, Boca Raton, FL.

[7] Ghahramani, S. (2019). *Fundamentals of probability with stochastic processes*, 4^{th} edition, CRC Press, Boca Raton, FL.

[8] Griffiths, M. (2013). Raw and Central moments of binomial random variables via Stirling numbers, *International Journal of Mathematical Education in Science and Technology*, **44**(2), 264-272.

[9] Harris, B. (1966). *Theory of Probability*, Addison-Wesley, Reading, Massachusetts.

[10] Joarder, A.H. and Mahmood, M. (1997). An inductive derivation of Stirling numbers of the second kind and their applications, *Journal of Applied Mathematics and Decision Sciences*, **1**(2), 151-157.

[11] Knoblauch, A. (2008). Closed-form expressions for the moments of the binomial probability distribution, *SIAM Journal on Applied Mathematics*, **69**(1), 197-204.

[12] Larsen, R.J. and Marx, M.L. (2017). *Introduction to mathematical statistics and its applications*, 6^{th} edition, Pearson Higher Education.

[13] Lint and Wilson (2001). *A course in combinatorics*, 2^{nd} edition, Cambridge University Press, Cambridge.

[14] Mackiw, G. (2000). A combinatorial approach to sums of integer powers, *Mathematics Magazine*, **73**(1), 44-46.

[15] Nguyen, D. (2021). A probabilistic approach to the moments of binomial random variables and application, *The American Statistician*, **75**(1), 101-103.

[16] Ross, S. (2002) *A first course in probability*, Pearson Education, New Delhi, India.

[17] Schultz, H.J. (1980). The sum of the kth powers of the first n integers, *The American Mathematical Monthly*, 87, 478-481.

[18] Spivey, M.Z. (2021). A combinatorial view of sums of powers, *Mathematics Magazine*, **94**(2), 125-131.
[19] Trevino, E. (2018). A short proof of powers formula, *The American Mathematical Monthly*, 125(7), p.659.

APPENDIX

Sums of Integer Powers of Integers

Given in (4.7) for $n = 2, 3, 4$. In all three cases, the simplifications are done using Table 1 and repeated use of Identity 2.1 and Identity 2.2.

(1) For $n = 2$,

$$\begin{aligned}
1^2 + 2^2 + 3^2 + \ldots + k^2 &= \sum_{i=1}^{2} S(2,i)(i!)\binom{k}{i+1} \\
&= S(2,1)(1!)\binom{k+1}{2} + S(2,2)(2!)\binom{k+1}{3} \\
&= \binom{k+1}{2} + 2\binom{k+1}{3} = \binom{k+2}{3} + \binom{k+1}{3} \\
&= \left(\frac{k+2}{k-1}\right)\binom{k+2}{3} + \binom{k+1}{3} \\
&= \frac{(2k+1)}{(k-1)}\frac{(k+1)k(k-1)}{3!} = \frac{k(k+1)(2k+1)}{6}.
\end{aligned}$$

(2) For $n = 3$,

$$\begin{aligned}
1^3 + 2^3 + 3^3 + \ldots + k^3 &= \sum_{i=1}^{3} S(3,i)(i!)\binom{k}{i+1} \\
&= S(3,1)(1!)\binom{k+1}{2} + S(3,2)(2!)\binom{k+1}{3} \\
&+ S(3,3)(3!)\binom{k+1}{4} = \binom{k+1}{2} + (3)(2)\binom{k+1}{3} \\
&+ 6\binom{k+1}{4} = \binom{k+1}{2} + 6\binom{k+2}{4} \\
&= \binom{k+1}{2} + \frac{(k+1)(k-1)}{2}\binom{k+1}{2} \\
&= \binom{k+1}{2}\left[1 + \frac{(k+2)(k-1)}{2}\right] \\
&= \binom{k+1}{2}\left[\frac{(k)(k+1)}{2}\right] = \left[\frac{k(k+1)}{2}\right]^2.
\end{aligned}$$

(3) For $n = 4$,

$$
\begin{aligned}
1^4 + 2^4 + 3^4 + \ldots + k^4 &= \sum_{i=1}^{4} S(4,i)(i!)\binom{k}{i+1} \\
&= S(4,1)(1!)\binom{k+1}{2} + S(4,2)(2!)\binom{k+1}{3} \\
&+ S(4,3)(3!)\binom{k+1}{4} + S(4,4)(4!)\binom{k+1}{5} \\
&= \binom{k+1}{2} + (7)(2!)\binom{k+1}{3} \\
&+ (6)(3!)\binom{k+1}{4} + (4!)\binom{k+1}{5}
\end{aligned}
$$

$$
\begin{aligned}
&= \binom{k+1}{2} + 14\binom{k+1}{3} + 36\binom{k+1}{4} + 24\binom{k+1}{5} \\
&= \binom{k+1}{2} + 2\binom{k+1}{3} + 12\left[\binom{k+1}{3}\right. \\
&\left. + 3\binom{k+1}{4} + 2\binom{k+1}{5}\right] \\
&= \frac{k(k+1)(2k+1)}{6} + 12\left[\binom{k+3}{5} + \binom{k+2}{5}\right] \\
&= \frac{k(k+1)(2k+1)}{6} + 12\left[\frac{k+3}{k-2}\binom{k+2}{5} + \binom{k+2}{5}\right] \\
&= \frac{k(k+1)(2k+1)}{6} + 12\left(\frac{2k+1}{k-2}\right)\binom{k+2}{5} \\
&= \frac{k(k+1)(2k+1)}{6} + \frac{k(k+1)(2k+1)(k+2)(k-1)}{10} \\
&= \frac{k(k+1)(2k+1)\left(3k^2+3k-1\right)}{30}.
\end{aligned}
$$

Central Moments and Their Usage

(1) The Binomial

$\mu = E(X) = kp$

$\sigma^2 = E(X-\mu)^2 = kpq$

$E(X-\mu)^3 = kpq(q-p)$

$E(X-\mu)^4 = 3k^2p^2q^2 + kpq(1-6pq)$

Skewness$(X) = E\left(\frac{X-\mu}{\sigma}\right)^3 = \frac{kpq(q-p)}{(kpq)^{\frac{3}{2}}} = \frac{q-p}{\sqrt{kpq}}$

Kurtosis$(X) = E\left(\frac{X-\mu}{\sigma}\right)^4 - 3 = \frac{3k^2p^2q^2+kpq(1-6pq)}{(k^2p^2q^2)} - 3 = \frac{1-6pq}{kpq}$.

(2) The Poisson

$\mu = E(X) = \lambda$

$\sigma^2 = E(X-\mu)^2 = \lambda$

$E(X-\mu)^3 = \lambda$

$E(X-\mu)^4 = 3\lambda^2 + \lambda$

$\text{Skewness}(X) = E\left(\frac{X-\mu}{\sigma}\right)^3 = \frac{\lambda}{(\lambda)^{\frac{3}{2}}} = \frac{1}{\sqrt{\lambda}}$

$\text{Kurtosis}(X) = E\left(\frac{X-\mu}{\sigma}\right)^4 - 3 = \frac{3\lambda^2+\lambda}{(\lambda^2)} - 3 = \frac{1}{\lambda}.$

(3) The Geometric

$\mu = E(X) = \frac{q}{p}$

$\sigma^2 = E(X-\mu)^2 = \frac{q}{p^2}$

$E(X-\mu)^3 = \frac{q(q+1)}{p^3}$

$E(X-\mu)^4 = \frac{\left[q\left(1+4q+q^2\right)+3q^2\right]}{p^4}$

$\text{Skewness}(X) = E\left(\frac{X-\mu}{\sigma}\right)^3 = \frac{\frac{q(1+q)}{p^3}}{\frac{(q)^{\frac{3}{2}}}{p^3}} = \frac{1+q}{\sqrt{q}}$

$\text{Kurtosis}(X) = E\left(\frac{x-\mu}{\sigma}\right)^4 - 3 = \frac{\frac{\left[3q^2+q\left(1+4q+q^2\right)\right]}{p^4}}{\frac{q^2}{p^4}} - 3 = \frac{1+4q+q^2}{q}.$

(4) The Negative Binomial

$\mu = E(X) = \frac{kq}{p}$

$\sigma^2 = E(X-\mu)^2 = \frac{kq}{p^2}$

$E(X-\mu)^3 = \frac{kq(q+1)}{p^3}$

$E(X-\mu)^4 = \frac{\left[kq\left(1+4q+q^2\right)+3k^2q^2\right]}{p^4}$

$\text{Skewness}(X) = E\left(\frac{X-\mu}{\sigma}\right)^3 = \frac{\frac{kq(1+q)}{p^3}}{\frac{(kq)^{\frac{3}{2}}}{p^3}} = \frac{1+q}{\sqrt{kq}}$

$\text{Kurtosis}(X) = E\left(\frac{X-\mu}{\sigma}\right)^4 - 3 = \frac{\frac{\left[3k^2q^2+kq\left(1+4q+q^2\right)\right]}{p^4}}{\frac{(k^2q^2)}{p^4}} - 3 = \frac{1+4q+q^2}{kq}.$