# THE GENERATING FUNCTIONS OF STIRLING NUMBERS OF THE SECOND KIND DERIVED PROBABILISTICALLY

GEORGE KESIDIS,* *Pennsylvania State University*

TAKIS KONSTANTOPOULOS,** *University of Liverpool*

MICHAEL A. ZAZANIS,*** *Athens University of Economics and Business*

## Abstract

Stirling numbers of the second kind, $S(n, r)$, denote the number of partitions of a finite set of size $n$ into $r$ disjoint nonempty subsets. The aim of this short article is to shed some light on the generating functions of these numbers by deriving them probabilistically. We do this by linking them to Markov chains related to the classical coupon collector problem; coupons are collected in discrete time (ordinary generating function) or in continuous time (exponential generating function). We also review the shortest possible combinatorial derivations of these generating functions.

*Keywords:* Stirling number of the second kind; generating function; Markov chain; coupon collector problem

2010 Mathematics Subject Classification: Primary 05A18; 05-01
Secondary 60J10; 60J27

## 1. Introduction

The Stirling number of the second kind, $S(n, r)$, is defined as the number of ways to place $n$ labeled balls into $r$ unlabeled boxes so that no box is empty. Clearly, $S(0, r) = 0$ for all $r > 0$. Note that $S(n, 0) = 0$ if $n > 0$ because if you have one or more balls, you need at least one box to put them in. But $S(0, 0) = 1$ because it is certainly true that you need no boxes if you have nothing to place inside. Hence, $S(n, r)$ is defined for all integers $n, r \geq 0$ and, certainly, $S(n, r) = 0$ if $r > n$. The exponential and ordinary generating functions of the sequence $S(n, r)$, $n = 0, 1, \ldots$, are respectively given by

$$\sum_{n=0}^{\infty} S(n, r) \frac{t^n}{n!} = \frac{(e^t - 1)^r}{r!}, \tag{1}$$

$$\sum_{n=0}^{\infty} S(n, r) t^n = \frac{1}{(1/t - 1)(1/t - 2) \cdots (1/t - r)}. \tag{2}$$

See [3, Equations (24b) and (24c), Chapter 1]. For more information on Stirling numbers in relation to probability, see [2]. Mechanical demonstrations of these formulae can be based on

the recursion

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

which is easily derived by a standard argument: for any placement of $n$ labeled balls into $r$ unlabeled boxes, there is either a box containing ball $n$ only or not. But mechanical proofs may not be too satisfactory. In this article we provide new probabilistic proofs of the two identities. For the sake of completeness, we also review combinatorial proofs. A proof is called combinatorial if it can be derived as a result of equality of two sets. A proof is called probabilistic if it is a consequence of equality between two random variables. In real life, there is not always a clear distinction between mechanical, combinatorial, and probabilistic, but when we have a proof, we typically have a hunch of what type it is.

A combinatorial derivation of (1) is not very difficult; see Section 2. However, a combinatorial derivation of (2), given in Section 6, is more difficult; but the probabilistic proof given in Section 5 is simpler and more intuitive.

A few words about notation. If $A$ is a finite set then $|A|$ denotes its cardinality; $\mathcal{P}(A)$ denotes the set of all subsets of $A$; $\mathcal{P}_r(A)$ is the set of all $B \in \mathcal{P}(A)$ such that $|B| = r$; for a set (or logical clause) $\pi$, the indicator symbol $\mathbf{1}_\pi$ equals 1 on the set $\pi$ (when the clause $\pi$ is true) and 0 otherwise; if $f : A \to B$ is a function from the set $A$ into the set $B$ then $f^{-1}(b)$ is the set of all $a \in A$ such that $f(a) = b$; for a positive integer $n$, we let $[n] := \{1, 2, \ldots, n\}$; we also let $(n)_r := n(n-1)(n-2) \cdots (n-r+1)$ and $\binom{n}{r} = (n)_r / r!$. If $A_1$, $A_2$ are disjoint sets and $f_1$, $f_2$ functions with domains $A_1$, $A_2$, respectively, we let $f_1 \times f_2$ be the function with domain $A_1 \cup A_2$ which is equal to $f_1$ on $A_1$ and $f_2$ on $A_2$. The set of surjective (onto) functions $f : A \to B$ is denoted by $\mathrm{SUR}(A, B)$.

## 2. Combinatorial derivation of the exponential generating function

Recall the definition of $S(n, r)$ as the number of ways to place $n$ labeled balls into $r$ unlabeled boxes so that no box is empty. Therefore, the number of ways to place $n$ labeled balls into $r$ *labeled* boxes so that no box is empty is

$$W_{n,r} = r! \, S(n, r) \tag{3}$$

because, for each placement in unlabeled boxes, there are $r!$ corresponding placements in labeled boxes. If we think of boxes as colors then placing balls into boxes is the same as assigning colors to balls. With $r$ different colors available, suppose that $s_j$ balls have color $j$, $j = 1, \ldots, r$. Then the number of configurations is the multinomial coefficient

$$\binom{n}{s_1, \ldots, s_r} = \frac{n!}{s_1! \cdots s_r!}.$$

Hence,

$$r! \, S(n, r) = \sum_{s_1, \ldots, s_r \geq 1, \, s_1 + \cdots + s_r = n} \binom{n}{s_1, \ldots, s_r}. \tag{4}$$

The sum is over all integers $s_1, \ldots, s_r$ that sum to $n$ and are all positive because each color must be used at least once (no box can be left empty). This last identity is what is behind (1). The rest is mechanical. Multiply both sides by $t^n / n!$ and sum over all $n$ to obtain

$$\sum_{n \geq 0} r! \, S(n, r) \frac{t^n}{n!} = \sum_{n \geq 0} \sum_{s_1, \ldots, s_r \geq 1, \, s_1 + \cdots + s_r = n} \frac{t^n}{s_1! \cdots s_r!} = \sum_{s_1, \ldots, s_r \geq 1} \frac{t^{s_1 + \cdots + s_r}}{s_1! \cdots s_r!} \sum_{n \geq 0} \mathbf{1}_{s_1 + \cdots + s_r = n} \, .$$

The last sum is clearly equal to 1 because $s_1 + \cdots + s_r$ is a nonnegative integer. We are therefore left with a sum over $s_1, \ldots, s_r$ that, since 'variables separate', splits into a product:

$$\sum_{s_1, \ldots, s_r \geq 1} \frac{t^{s_1 + \cdots + s_r}}{s_1! \cdots s_r!} = \sum_{s_1 \geq 1} \frac{t^{s_1}}{s_1!} \cdots \sum_{s_r \geq 1} \frac{t^{s_r}}{s_r!} = (e^t - 1)^r.$$

The last equality follows from $e^t = 1 + \sum_{s \geq 1} t^s / s!$.

## 3. Probabilistic derivation of the exponential generating function

Consider the classical coupon collector problem. There are $M$ distinct coupons. Select $n$ coupons at random with replacement, and let $Y_n$ be the number of distinct coupons selected. Then the distribution of $Y_n$ is

$$\mathbb{P}(Y_n = r) = \frac{1}{M^n} \binom{M}{r} r! \, S(n, r). \tag{5}$$

To see why (5) holds, first note that our sample space is the set of all ordered $n$-tuples of coupons, a set of cardinality $M^n$. On the other hand, the subset of the sample space defined by the event $\{Y_n = r\}$ is the set of all ordered $n$-tuples that use exactly $r$ coupons. Given a subset of the set of coupons of cardinality $r$, there are $r! \, S(n, r)$ ordered $n$-tuples that use these $r$ coupons. Since there are $\binom{M}{r}$ ways to select the subset of coupons of size $r$, the formula follows.

Next observe that the sequence $Y_0 := 0, Y_1, Y_2, \ldots$ is a Markov chain. For background on Markov chains, see, e.g. [1]. We can think of this as the coupon selection process in discrete time. Let $N(t)$, $t \geq 0$, be an independent Poisson process with rate $M$ and consider the stochastic process $Y_{N(t)}$, $t \geq 0$, to be a continuous-time coupon selection process that is also a Markov chain. We can think of it in a different manner. Split the Poisson process $N$ into $M$ independent copies. Thus, let $N_1, \ldots, N_M$ be independent rate-1 Poisson processes. Then $N(t) = \sum_{j=1}^M N_j(t)$, $t \geq 0$, is a rate-$M$ Poisson process. Instead of having a single person collecting coupons at the ticks of the Poisson process $N$, we have $M$ people collecting coupons: person $j$ collects coupons of type $j$ at the points of $N_j$. Let $\tau_j$ be the first point of the Poisson process $N_j$. Then

$$\mathbb{P}(Y_{N(t)} = r) = \mathbb{P}(r \text{ of the } \tau_j \text{s are less than or equal to } t \text{ and the rest are greater than } t)$$

$$= \binom{M}{r} \mathbb{P}(\tau_1, \ldots, \tau_r \leq t) \mathbb{P}(\tau_{r+1}, \ldots, \tau_M > t)$$

$$= \binom{M}{r} (1 - e^{-t})^r (e^{-t})^{M-r}. \tag{6}$$

On the other hand, using (5) and the fact that $N$ is independent of $(Y_n)$,

$$\mathbb{P}(Y_{N(t)} = r) = \sum_{n=0}^{\infty} \frac{(Mt)^n}{n!} e^{-Mt} \mathbb{P}(Y_n = r) = \sum_{n=0}^{\infty} \frac{(Mt)^n}{n!} e^{-Mt} \frac{1}{M^n} \binom{M}{r} r! \, S(n, r). \tag{7}$$

Equating (6) and (7) yields (1).

## 4. Coda

Let $A$ and $B$ be two nonempty finite sets. The set $B^A$ of *all functions* from $A$ to $B$ has cardinality $|B|^{|A|}$.

The set of all *injective* (one-to-one) functions from $A$ to $B$ has cardinality

$$|B|(|B| - 1) \cdots (|B| - |A| + 1),$$

a number that is obviously 0 if $|A| > |B|$.

The set of all *surjective* (onto) functions from $A$ to $B$ has cardinality

$$|B|! \, S(|A|, |B|)$$

(a number that is obviously 0 if $|B| > |A|$), because we can think of a surjective function $f : A \to B$ as a placement of labeled balls (the elements of $A$) into labeled boxes (the elements of $B$) so that no box is left empty. See (3).

Each function $f : A \to B$ is obviously a surjective function onto some nonempty set, namely the set $f(A)$. Hence,

$$B^A = \bigcup_{C \subset B, \, C \neq \varnothing} \{f \in B^A : f(A) = C\} = \bigcup_{r=1}^{|B|} \bigcup_{C \subset B, \, |C| = r} \{f \in B^A : f(A) = C\},$$

a splitting of $B^A$ according to the image of each of its elements. Both unions are disjoint in the last display. Hence,

$$|B|^{|A|} = \sum_{r=1}^{|B|} \sum_{C \subset B, \, |C| = r} |C|! \, S(|A|, |C|)$$

$$= \sum_{r=1}^{|B|} r! \, S(|A|, r) \sum_{C \subset B, \, |C| = r} 1$$

$$= \sum_{r=1}^{|B|} r! \, S(|A|, r) \binom{|B|}{r},$$

and this verifies that (5) is a probability function.

## 5. Probabilistic derivation of the ordinary generating function

It suffices to prove (2) for $t = 1/M$, $M = 1, 2, \ldots$. Consider the probability distribution defined by (5) and sum over $n$:

$$\sum_{n=0}^{\infty} \mathbb{P}(Y_n = r) = \binom{M}{r} r! \sum_{n=0}^{\infty} S(n, r) M^{-n} = (M)_r \sum_{n=0}^{\infty} S(n, r) M^{-n}.$$

Here $(M)_r := M(M - 1) \cdots (M - r + 1)$. As mentioned earlier, the random sequence $Y_n$, $n = 0, 1, 2, \ldots$, is a Markov chain. We refer to the index $n$ as 'time'. We can easily compute the transition probabilities

$$\mathbb{P}(Y_{n+1} = r + 1 \mid Y_n = r) = \frac{M - r}{M}, \qquad r = 0, 1, \ldots.$$

Let $\sigma(r)$ be the total time that the chain remains in state $r$. Then $\sigma(r)$ is a geometric random variable with parameter $(M - r)/M$:

$$\mathbb{P}(\sigma(r) = k) = \left(\frac{r}{M}\right)^{k-1} \frac{M - r}{M}, \qquad k = 1, 2, \ldots.$$

Hence,
$$\mathbb{E}\sigma(r) = \frac{M}{M-r}.$$

But $\sigma(r) = \sum_{n=0}^{\infty} \mathbf{1}_{Y_n=r}$, and so

$$\mathbb{E}\sigma(r) = \sum_{n=0}^{\infty} \mathbb{P}(Y_n = r).$$

Hence,

$$\sum_{n=0}^{\infty} S(n,r)M^{-n} = \frac{1}{(M)_r}\frac{M}{M-r} = \frac{1}{(M-1)(M-2)\cdots(M-r)},$$

and this yields (2) for $t = 1/M$.

## 6. Combinatorial derivation of the ordinary generating function

The combinatorial derivation of the ordinary generating function (2) is more difficult. It is suggested as an exercise in Stanley's book; see [3, Exercise 16, Chapter 1, p. 46]. By multiplying both sides of (2) by $r!$ and expanding the right-hand side, we see that proving correctness of (2) is equivalent to showing that

$$r!\,S(n,r) = \sum_{k_1,\ldots,k_r \geq 1,\, k_1+\cdots+k_r=n} 1^{k_1}2^{k_2}\cdots r^{k_r}. \tag{8}$$

As noted in (4), the number $r!\,S(n,r)$ is the cardinality of the set $\mathrm{SUR}([n],[r])$ of surjections from a set of size $n$ onto a set of size $r$. We will show that the right-hand side of (8) counts the same thing by exhibiting a bijection

$$\Phi\colon \mathrm{SUR}([n],[r]) \to \bigcup_{k_1,\ldots,k_r \geq 1,\, k_1+\cdots+k_r=n} ([1]^{k_1} \times [2]^{k_2} \times \cdots \times [r]^{k_r}). \tag{9}$$

Let $f\colon [n] \to [r]$ be a surjective function. Consider the set $[n] = \{1,\ldots,n\}$ in its natural order, and let $i_1 < i_2 < \cdots < i_r$ be the points at which the function takes a new value for the first time. That is, let

$$i_1 := 1 \qquad i_{p+1} := \min\{i > i_m\colon f(i) \neq f(i_1),\ldots,f(i_p)\}, \quad p = 1,\ldots,r-1.$$

Then the numbers $f(i_1),\ldots,f(i_r)$ are just the numbers $1,\ldots,r$ in a different order. Let $L\colon [r] \to [r]$ then be defined by

$$L(f(i_p)) := p, \qquad p = 1,\ldots,r,$$

a relabeling of the values of $f$. Let

$$K_1 := [i_1,i_2), \qquad K_2 := [i_2,i_3), \quad \ldots, \quad K_r := [i_r,n],$$

and define $g_p \in [p]^{K_p}$ by

$$g_p(i) := L(f(i)), \qquad i \in K_p.$$

Finally, let

$$\Phi f := g_1 \times \cdots \times g_r.$$

(See the notational note at the end of Section 1.) It is easy to see that $\Phi$ is a bijection from $\mathrm{SUR}([n],[r])$ onto $\bigcup_{K_1,\ldots,K_r}([1]^{K_1} \times [2]^{K_2} \times \cdots \times [r]^{K_r})$, where the union is taken over disjoint intervals $K_1,\ldots,K_r$ such that $K_1 = [1,i_1)$, $K_2 = [i_2,i_3),\ldots,K_r = [i_r,n]$ for

some $1 = i_1 < i_2 < \cdots < i_r \leq n$, and given by the right-hand side of (9). Since all sets in the union are pairwise disjoint, its cardinality is given by the right-hand side of (8).

**Remark 1.** Note that, for the combinatorial derivation of the ordinary generating function, we used the ordering of $[n]$, something that was not done for the exponential generating function. Combining (4) and (8) we have

$$\sum_{s_1,\ldots,s_r \geq 1,\, s_1+\cdots+s_r=n} \binom{n}{s_1,\,\ldots,\,s_r} = \sum_{s_1,\ldots,s_r \geq 1,\, s_1+\cdots+s_r=n} 1^{s_1} 2^{s_2} \cdots r^{s_r},$$

a rather curious identity whose direct proof (that is, showing that the two sides are equal without showing that they are both equal to a known quantity) is unknown (to the authors).

## Dedication

We are pleased to have been given the opportunity to contribute to the final issue of *The Mathematical Scientist* and very happy that an anonymous referee is of the opinion that Professor Joe Gani (founding editor of *Journal of Applied Probability* and *Advances in Applied Probability*) 'would most certainly have enjoyed this paper'. We dedicate this paper to Joe Gani's memory.

## References

[1] BRÉMAUD, P. (2008). *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues.* Springer, New York.
[2] PITMAN, J. (2002). *Combinatorial Stochastic Processes* (Lecture Notes Math. **1875**). Springer, Berlin.
[3] STANLEY, R. P. (2011). *Enumerative Combinatorics*, Vol. 1. Cambridge University Press.