Distributions and Stirling Numbers

Suppose there are n balls and k boxes. We determine the number of ways that the balls can be distributed among the boxes under a variety of conditions. Our main focus is on the case where the balls are distinguishable and no box can be left empty (the boxes may or may not be distinguishable).

Suppose that the balls are identical and the boxes are labelled $1, 2, \ldots, k$. Then, the number of ways that the balls can be distributed among the boxes equals the number of sequences of n zeros (corresponding to the balls) and $k-1$ ones (corresponding to dividers between "boxes" – the number of balls in box 1 equals the number of zeros before the first one, the number of balls in box 2 equals the number of zeros between the first and second ones, and so on). In turn, this equals $\binom{n+k-1}{n}$. If each box can hold at most one ball, then there are zero ways if $n > k$, and otherwise there are $\binom{k}{n}$ distributions (choose which boxes hold a ball). If no box can be left empty, then there are zero distributions when $n < k$ and $1 \times \binom{n-k+k-1}{n-k}$ (one way to take k of the identical balls and put one into each box).

Suppose the balls are labelled $1, 2, \ldots, n$ and the boxes are labelled $1, 2, \ldots, k$. Then, the number of ways that the balls can be distributed among the boxes equals the number of functions from the set of balls to the set of boxes ($f(i) =$ the label of the box containing ball i), which is k^n. If each box can hold at most one ball, then the number of ways equals the number of 1-1 functions from the balls to the boxes, which equals zero if $n > k$ and $n!/(n-k)!$ otherwise. If no box can be left empty, then the number of ways equals the number of onto functions from the balls to the boxes.

Suppose that the balls are labelled $1, 2, \ldots, n$ and the boxes are identical. We consider only the case where no box is left empty.

Definition (Stirling number of the second kind). For non-negative integers k and n, the Stirling number of the second kind, $S(n, k)$, equals the number of ways to distribute n labelled balls into k identical boxes so that no box is left empty.

The typical definition of the Stirling numbers of the second kind is that $S(n, k)$ is the number of ways to separate (or *partition*) a set of n objects into k disjoint non-empty subsets. What we have written above is equivalent, with the balls playing the role of the objects and the boxes playing the role of the subsets (the contents of the each box are the elements in that subset).

By the definition, for all $n \geq 1$ we have $S(n, 1) = 1$ and $S(n, n) = 1$. It also follows that $S(n, k) = 0$ unless $1 \leq k \leq n$.

We show that $S(n, 2) = (2^n - 2)/2 = 2^{n-1} - 1$. Each distribution of n labelled balls into two identical boxes determines $2! = 2$ distributions of the balls into two labelled boxes because there are 2! ways to designate the boxes as box 1 and box 2. Thus, $S(n, 2)$ equals one half of the number of ways to distribute n labelled balls into 2 labelled boxes. This equals the number of onto functions from an n-set to $\{1, 2\}$. We could appeal to (the unproved) Proposition F5, but in this case it is possible to count the distributions directly. The collection of balls in box 1 can be any non-empty proper subset $X \subseteq \{1, 2, \ldots, n\}$

($2^n - 2$ choices) and then the collection of balls in box 2 is $\{1, 2, \ldots, n\} - X$ (1 choice). Hence, the number of distributions of balls into labelled boxes is $2^n - 2$. Therefore, $S(n, 2) = (2^n - 2)/2 = 2^{n-1} - 1$.

Stirling numbers and onto functions. Each distribution of n labelled balls into k identical boxes determines $k!$ distributions of n labelled balls into k labelled boxes: one for each way of labelling the boxes with $1, 2, \ldots, k$. We know from above that the number of distributions of n labelled balls into k labelled boxes equals the number of onto functions from the set of balls to the set of boxes. Let's call this quantity $O(n, k)$. Thus, $k!S(n, k) = O(n, k)$ or, equivalently, $S(n, k) = \frac{1}{k!}O(n, k)$. Using Proposition F5 (which is not yet proved) this leads to the formula $S(n, k) = \frac{1}{k!}\sum_{i=0}^{k}(-1)^i\binom{k}{i}(k - i)^n$.

Example (counting using Stirling numbers). The number of ways in which a professor can distribute 6 different tasks among his 4 research assistants such that each assistant is assigned at least 1 task equals the number of onto functions from the tasks to the research assistants, which is $4!\,S(6, 4)$.

Example (counting using Stirling numbers). Let $A = \{a_1, a_2, \ldots, a_6\}$ and $B = \{b_1, b_2, b_3, b_4\}$. We use Stirling numbers of the second kind to count the number of onto functions $f : A \to B$ such that $f(a_6) = b_4$. There are two possibilities, either $f^{-1}(\{b_4\}) = \{a_6\}$ or $f^{-1}(\{b_4\}) \supset \{a_6\}$. In the first case, $f|_{A-\{a_6\}}$ is an onto function from $A - \{a_6\}$ to $B - \{b_4\}$, and there are $3!S(5, 3)$ possibilities for f. In the second case, $f|_{A-\{a_6\}}$ is an onto function from $A - \{a_6\}$ to B, and there are $4!S(6, 4)$ possibilities for f. Thus, by the rule of sum, the number of possibilities for f is $3!S(5, 3) + 4!S(6, 4)$.

Theorem SN1. Let n and k be integers with $n \geq k \geq 1$. Then,

$$S(n, k) = S(n - 1, k - 1) + kS(n - 1, k).$$

Proof.

The LHS counts the number of ways to distribute n labelled balls into k identical boxes so that no box is left empty.

The RHS also counts these distributions. For any distribution of n labelled balls into k identical boxes with no box left empty, either ball n is the only ball in its box, or it isn't. In the first case, the remaining $n-1$ balls are distributed among the remaining $k-1$ boxes, with no box left empty. There are $S(n - 1, k - 1)$ ways that this can happen. In the second case, the remaining $n - 1$ balls are distributed into k boxes with no box left empty ($S(n - 1, k)$ possibilities), and then ball n can be in any one of these k boxes (k possibilities). Thus, by the rule of product there are $kS(n, k)$ distributions in which ball n is not the only ball in its box. Finally, by the rule of sum, the number of distributions is $S(n - 1, k - 1) + kS(n - 1, k)$.

Since the LHS and RHS count the same thing (in different ways), they are equal. This completes the proof. ∎

Proposition SN1, together with the facts that $S(n, 1) = 1$ and $S(n, n) = 1$ allows us to compute "Stirling's Triangle" below. The number in row n and column k is $S(n, k)$. The interior entries of each subsequent row are derived from the row above: for $1 < k < n$ the entry in row n and column k equals k times the number immediately above it, plus the number one row up and one column over. For example, $S(5, 3) = 25 = 3 \cdot 6 + 7$.

"Stirling's Triangle:"

$$\begin{array}{cccccccc}
1 \\
1 & 1 \\
1 & 3 & 1 \\
1 & 7 & 6 & 1 \\
1 & 15 & 25 & 10 & 1 \\
1 & 31 & 80 & 65 & 15 & 1 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}$$

Exercise. Recall that we have used $O(n, k)$ to denote the number of onto functions from $\{a_1, a_2, \ldots, a_n\}$ to $\{b_1, b_2, \ldots, b_k\}$. Prove the following:
(a) $O(n, 1) = 1$ and $O(n, n) = n!$.
(b) $O(n, k) = 0$ unless $1 \leq k \leq n$.
(c) Use a combinatorial argument (i.e. a counting argument) similar to the proof of Theorem SN1 to show that $O(n, k) = kO(n-1, k-1) + kO(n-1, k)$.
(d) Use Theorem SN1 to prove the correctness of the recurrence in (c).
(e) Use the recurrence in (c) to write out the first four rows of the "onto triangle" for the numbers $O(n, k)$.

Exercise. Show that the number of ways to distribute n labelled balls among k identical boxes (some of which might be empty) is $\sum_{i=1}^{k} S(n, i)$. Also, argue that the number of ways to distribute n labelled balls among k identical boxes with at most one ball per box is 0 if $k < n$ and 1 otherwise.

As a matter of interest only, we mention that the remaining situation for distibutions of balls into boxes is that the n balls are identical and the k boxes are identical. In this case, the number of distributions of the balls into the boxes corresponds to the number of ways of writing n as a sum of k terms, some of which may be zero if empty boxes are allowed. These are called *partitions* of the integer n into at most k parts, and are well-studied in combinatorics and number theory. For example, if there are 10 identical balls and 4 identical boxes then $5 + 3 + 2 + 0$ represents the situation where one box has five balls, one has 3, one has 2, and one is empty.