Set Partitions

October 9, 2001

1 Introduction

A *set partition* is just that, a partition of a set into subsets. The subsets in a set partition are called *blocks*. Here we will study algorithms for enumerating, ranking, unranking, and generating random instances of set partitions. Set partitions and the combinatorial numbers that count them show up in a variety of unexpected places and have been the focus of significant amount of research. *Combinatorica* contains a function called `SetPartitions` that lists *all* set partitions of a given set.

```
In[1]:= SetPartitions[{a, b, c, d}]
Out[1]= {{{a, b, c, d}}, {{a}, {b, c, d}}, {{a, b}, {c, d}},
         {{a, c, d}, {b}}, {{a, b, c}, {d}}, {{a, d}, {b, c}}, {{a, b, d}, {c}},
         {{a, c}, {b, d}}, {{a}, {b}, {c, d}}, {{a}, {b, c}, {d}},
         {{a}, {b, d}, {c}}, {{a, b}, {c}, {d}}, {{a, c}, {b}, {d}},
         {{a, d}, {b}, {c}}, {{a}, {b}, {c}, {d}}}
```

Note that the order in which elements are written in blocks and the order of the blocks themselves does not matter, however it will be useful to define a canonical way of writing a set partition. Assuming a total order on a set X, a canonical way of writing a set partition of X is: write each subset in increasing order and write the subsets themselves in increasing order of their minimum elements. *Combinatorica* contains a function called `ToCanonicalSetPartition` that converts an arbitrary set partition into its canonical form.

```
In[2]:= ToCanonicalSetPartition[{{b}, {c, e}, {d, a}}]
Out[2]= {{a, d}, {b}, {c, e}}
```

Set partitions of $[n]$ with k blocks can be recursively generated by noting that a set partition can either contain the element 1 by itself, as a singleton block or 1 as part of a block with some other elements. Set partitions of the former type can be generated by generating all set partitions with $(k-1)$ blocks of $\{2, 3, \ldots, n\}$ and then inserting the block $\{1\}$ into each set partition. Set partitions of the latter type can be generated by first generating all set partitions with k blocks of $\{2, 3, \ldots, n\}$ and then inserting the element 1 into each subset in each set partition. *Combinatorica* contains a function `KSetPartitions` that implements this idea.

```
KSetPartitions[{}, 0] := {{}}
KSetPartitions[s_List, 0] := {}
KSetPartitions[s_List, k_Integer] := {} /; (k > Length[s])
KSetPartitions[s_List, k_Integer] := {Map[{#} &, s]} /; (k === Length[s])
KSetPartitions[s_List, k_Integer] :=
       Join[Map[Prepend[#, {First[s]}] &, KSetPartitions[Rest[s], k - 1]],
            Flatten[Map[Table[Prepend[Delete[#, j], Prepend[#[[j]], s[[1]]]],
                               {j, Length[#]}
```

```
                        ]&,
                        KSetPartitions[Rest[s], k]
                    ], 1
            ]

        ] /; (k > 0) && (k < Length[s])
```

```
In[3]:= KSetPartitions[{a, b, c, d, e}, 2]
Out[3]= {{{a}, {b, c, d, e}}, {{a, b}, {c, d, e}}, {{a, c, d, e}, {b}},
         {{a, b, c}, {d, e}}, {{a, d, e}, {b, c}}, {{a, b, d, e}, {c}},
         {{a, c}, {b, d, e}}, {{a, b, c, d}, {e}}, {{a, e}, {b, c, d}},
         {{a, b, e}, {c, d}}, {{a, c, d}, {b, e}}, {{a, b, c, e}, {d}},
         {{a, d}, {b, c, e}}, {{a, b, d}, {c, e}}, {{a, c, e}, {b, d}}}
```

2 Stirling Numbers of the Second Kind

All of this leads to the questions: (a) how many set partitions of $[n]$ are there? and (b) how many set partitions of $[n]$ are there with k blocks? We'll tackle the second question first.

The number of set partitions of $[n]$ having k blocks is a fundamental combinatorial number called the *Stirling number of the second kind*. We have already encountered combinatorial numbers of the first kind that James Stirling lent his name to. We use $\left\{ n \atop k \right\}$ to denote the number of set partitions of $[n]$ having k blocks. Knuth's advice is that we read $\left\{ n \atop k \right\}$ as "n subset k."

Let n be a positive integer. How many set partitions of $[n]$ are there with 0 blocks? Clearly, none and therefore $\left\{ n \atop 0 \right\} = 0$ for all positive integers n. It is also obvious that $\left\{ n \atop 1 \right\} = 1$. How many set partitions of $[n]$ are there with 2 blocks? Each element in $[n]$ can be thrown into one or the other block in a total of 2^n ways. From these, we remove the two cases in which all elements are assigned to one set to get $2^n - 2$ possibilities. We then divide by 2 to get the final answer $2^n - 1$ since block 1 and block 2 are indistinguishable.

The recursive algorithm outlined in the previous section for generating set partitions of $[n]$ with k blocks is a proof of the following "Pascal triangle" like identity:

$$\left\{ n \atop k \right\} = \left\{ n-1 \atop k-1 \right\} + k \left\{ n-1 \atop k \right\},$$

for all integers $k, n > 0$. Also $\left\{ n \atop 0 \right\} = \left\{ 0 \atop k \right\} = 0$ for integers $n > 0$ and $k > 0$ and we decree that $\left\{ 0 \atop 0 \right\} = 1$.

```
In[4]:= Table[StirlingSecond[n, k], {n, 10}, {k, 1, n}] // ColumnForm
Out[4]= {1}
        {1, 1}
        {1, 3, 1}
        {1, 7, 6, 1}
        {1, 15, 25, 10, 1}
        {1, 31, 90, 65, 15, 1}
        {1, 63, 301, 350, 140, 21, 1}
        {1, 127, 966, 1701, 1050, 266, 28, 1}
        {1, 255, 3025, 7770, 6951, 2646, 462, 36, 1}
        {1, 511, 9330, 34105, 42525, 22827, 5880, 750, 45, 1}
```

The recursive identity for Stirling numbers of the second kind can be verified in the above table. Each entry depends on two entries: one to its north-west and the other to its north. For example, the table tells us that $\left\{ {6 \atop 4} \right\} = 65$ and this is equal to $25 + 4 \cdot 10$. The second column in the table tells us that $\left\{ {n \atop 2} \right\} = 2^{n-1} - 1$, confirming what we showed earlier.

The above recursive formula can be used to compute Stirling numbers of the second kind. However, there is a beautiful identity involving Stirling numbers and binomial numbers, that leads to a faster algorithm:

$$\left\{ {n \atop k} \right\} = \frac{1}{k!} \sum_{i=0}^{k} (-1)^i \binom{k}{i} (k-i)^n. \tag{1}$$

In order to prove this identity we make a connection between Stirling numbers of the second kind and the number of *surjective functions* (onto functions) from one set to another. Let $S_{n,k}$ denote the set of surjections from the set $[n]$ to the set $[k]$. Note that every surjection $f : [n] \to [k]$ induces a set partition of $[n]$ into blocks $B_1, B_2, \ldots, B_k$, where $B_i = \{j \in [n] \mid f(j) = i\}$. However, each set partition of $[n]$ with k blocks corresponds to $k!$ surjections because every permutation of the blocks is a different surjection. This leads to the identity

$$\left\{ {n \atop k} \right\} = \frac{|S_{n,k}|}{k!}.$$

Now we focus on obtaining a formula for $|S_{n,k}|$ the number of surjections from $[n]$ to $[k]$.

To do this we resort to the inclusion-exclusion principle. Recall that we last used the inclusion-exclusion principle when we derived the formula $D_n = \sum_{i=0}^{n} (-1)^i n!/i!$ for the number of n-derangements. At that time we stated the inclusion-exclusion principle as follows. Let X be a finite set and let $X_1, X_2, \ldots, X_m$ be subsets of X. For any set $I \subseteq [m]$ of indices, let $X_I = \cap_{i \in I} X_i$. The inclusion-exclusion principle states that

$$\left| X - \bigcup_{i \in [m]} X_i \right| = \sum_{I \subseteq [m]} (-1)^{|I|} |X_I|.$$

To apply this principle to the current problem, we let X denote the set of all functions from $[n]$ to $[k]$ and let X_i, $1 \leq i \leq k$, denote the set of all functions from $[n]$ to $[k] - \{i\}$. So X_i is the set of functions that "miss" i. So $X - \cup_{i \in [k]} X_i$ is the set of functions that "miss" nothing. In other words, $X - \cup_{i \in [k]} X_i$ is $S_{n,k}$ the set of all surjections from $[n]$ to $[k]$. Now note that X_I is the set of functions from $[n]$ to $[k]$ that "miss" everything in I. In other words, X_I are the set of functions from $[n]$ to $[k] - I$. So $|X_I| = (k - |I|)^n$. Plugging this into the right hand side of the inclusion exclusion principle we get

$$|S_{n,k}| = \sum_{I \subseteq [k]} (-1)^{|I|} (k - |I|)^n.$$

Now notice that terms in the summation on the right hand side depend only on the size of $|I|$. So all terms corresponding to subsets I of the same size can be brought together and we get

$$|S_{n,k}| = \sum_{i=0}^{k} (-1)^i \binom{k}{i} (k-i)^n.$$

Substituting into Equation (1) we get identity we set out to prove. *Combinatorica* contains a function `StirlingSecond` that implements the above identity.

```
StirlingSecond[n_Integer,0] := If [n == 0, 1, 0]
StirlingSecond[0,k_Integer] := If [k == 0, 1, 0]

StirlingSecond[n_Integer?Positive, k_Integer?Positive] :=
        Sum [ (-1)^(k-i)*Binomial [k, i]*(i^n), {i, 1, k}]/k!
```

There are many ways of partitioning a 100-element set into blocks of 50, more than 10^{100} in fact!

```
In[5]:= StirlingSecond[100, 50]
Out[5]= 43098323700936634042151430154725869594352028961434061391244174113128O\
        3190588537831455982616599920139OO
```

3 A Generating Function for Stirling Numbers

Like in the case of integer partitions, we would like to derive generating functions for Stirling numbers of the second kind. In particular, for each positive integer k, we would like to derive a nice closed form expression for the generating function

$$\sum_{n=0}^{\infty} \left\{ {n \atop k} \right\} z^n.$$

This is possible, but a little hard to do. It is somewhat simpler to consider an *exponential generating function* for the sequence $(\left\{ {0 \atop k} \right\}, \left\{ {1 \atop k} \right\}, \left\{ {2 \atop k} \right\}, \ldots)$. This is defined as

$$S_k(z) = \sum_{n=0}^{\infty} \frac{\left\{ {n \atop k} \right\} z^n}{n!}.$$

The generating functions we have seen in the past are sometimes called *ordinary generating functions*, to distinguish them from exponential generating functions.

To derive a closed form expression for $S_k(z)$ we start by considering the product

$$(z + z^2 + z^3 + \cdots)^k.$$

What is the coefficient of z^n in the expansion of this product? To answer this we observe that the coefficient of z^n gets a contribution of 1 for every product $z^{i_1} \cdot z^{i_2} \cdots z^{i_k}$ where $i_1 + i_2 + \cdots + i_k = n$ and the term z^{i_1} is gotten from the first bracket, the term z^{i_2} is gotten from the second bracket, and so on. Thus the coefficient of z^n is the number of k-tuples $(i_1, i_2, \ldots, i_k)$ of nonzero integers such that $i_1 + i_2 + \cdots + i_k = n$.

Focus on a particular choice of $(i_1, i_2, \ldots, i_k)$. The numbers in this tuple can be treated as the sizes of k blocks in a set partition. For this particular choice of $(i_1, i_2, \ldots, i_k)$, how many set partitions can we get? To see this note that the first block the set partition can be “filled” in $\binom{n}{i_1}$ ways, the second in $\binom{n - i_1}{i_2}$ ways, the third in $\binom{n - i_1 - i_2}{i_3}$ ways and so on to yield

$$\binom{n}{i_1} \cdot \binom{n - i_1}{i_2} \cdot \binom{n - i_1 - i_2}{i_3} \cdots \binom{n - i_1 - i_2 - \cdots - i_{k-1}}{i_k} = \frac{n!}{i_1! i_2! \ldots i_k!}.$$

Now note that in doing this calculation we treated the blocks as distinguishable and therefore we have to divide by $k!$ to get the final answer. So the total number of set partitions corresponding to a particular choice of $(i_1, i_2, \ldots, i_k)$ is

$$\frac{1}{k!} \cdot \frac{n!}{i_1! i_2! \ldots i_k!}.$$

Furthermore, the total number of set partitions of $[n]$ with k blocks is

$$\sum_{(i_1,i_2,\ldots,i_k)} \frac{1}{k!} \cdot \frac{n!}{i_1!i_2!\ldots i_k!} \tag{2}$$

where the sum is over all choices of k-tuples $(i_1, i_2, \ldots, i_k)$ of nonzero integers that add up to n.

Now note that had we started with the product

$$\left(\frac{z}{1!} + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots\right)^k,$$

the coefficient of z^n in its expansion would get a contribution of $1/(i_1!i_2!\ldots i_k!)$ for every product $z^{i_1} \cdot z^{i_2} \cdots z^{i_k}$ where $i_1 + i_2 + \cdots + i_k = n$ and the term z^{i_1} is gotten from the first bracket, the term z^{i_2} is gotten from the second bracket, and so on. Thus the coefficient of z^n in this case would be

$$\sum_{(i_1,i_2,\ldots,i_k)} \frac{1}{(i_1!i_2!\ldots i_k!)}$$

where the sum is over all choices of k-tuples $(i_1, i_2, \ldots, i_k)$ of nonzero integers that add up to n. Denote this quantity by $C_{n,k}$. From (2) and the above observation we get

$$\frac{\left\{ n \atop k \right\}}{n!} = \frac{1}{k!} C_{n,k}.$$

From this we obtain that

$$S_k(z) = \frac{1}{k!}\left(\frac{z}{1!} + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots\right)^k = \frac{1}{k!}(e^z - 1)^k.$$

As an immediate application of this generating function result we provide an alternative proof of the identity

$$\left\{ n \atop k \right\} = \frac{1}{k!}\sum_{i=0}^{k}(-1)^i \binom{k}{i}(k-i)^n$$

that we proved earlier using the inclusion-exclusion principle. The proof is simply a matter of going through some algebra:

$$\begin{aligned}
S_k(z) &= \frac{1}{k!}(e^z - 1)^k \\
&= \frac{1}{k!}\sum_{i=0}^{k}\binom{k}{i} e^{zi}(-1)^{(k-i)} \text{ By the Binomial Theorem} \\
&= \frac{1}{k!}\sum_{i=0}^{k}\binom{k}{i}(-1)^{(k-i)}\left(\sum_{n=0}^{\infty}\frac{(zi)^n}{n!}\right) \\
&= \frac{1}{k!}\sum_{n=0}^{\infty}\sum_{i=0}^{k}\frac{z^n \cdot i^n}{n!}\binom{k}{i}(-1)^{(k-i)} \\
&= \frac{1}{k!}\sum_{n=0}^{\infty}\frac{z^n}{n!}\left(\sum_{i=0}^{k}\binom{k}{i}(-1)^{(k-i)} i^n\right).
\end{aligned}$$

Equating coefficients of like terms on the two sides we get

$$\left\{ n \atop k \right\} = \sum_{i=0}^{k}\binom{k}{i}(-1)^{(k-i)} i^n = \sum_{i=0}^{k}\binom{k}{i}(-1)^i (k-i)^n.$$

4 Bell Numbers

The total number of set partitions of $[n]$ is the nth *Bell number*, another fundamental combinatorial number, denoted B_n. Bell numbers are named after E. T. Bell, who is better known for his riveting accounts of the lives of famous mathematicians in *Men of Mathematics*, than for his contributions to combinatorics. Clearly, $B_n = \sum_{k=1}^{n} \left\{ {n \atop k} \right\}$ and this can be used to compute Bell numbers. However, the identity

$$B_n = \sum_{k=0}^{n-1} \binom{n-1}{k} B_k$$

provides an alternate way of computing Bell numbers, that is slightly faster. A simple bijection establishes the identity. Set partition of $[n]$ can be generated by choosing the subset containing 1 first and then prepending the chosen subset to set partitions of the rest of the elements. $\binom{n-1}{k}$ is the number of ways of choosing a $(k+1)$-subset containing 1 and B_{n-k-1} is the number of set partitions of the rest of the elements. Summed over all possible k and simplified using the symmetry of binomial numbers we get:

$$B_n = \sum_{k=0}^{n-1} \binom{n-1}{k} B_{n-(k+1)} = \sum_{k=0}^{n-1} \binom{n-1}{k} B_k.$$

In the following we show that B_{300} can be computed about twice as fast by using the recurrence relation instead of the sum of Stirling numbers.

```
In[6]:= {Timing[BellB[300];], Timing[Sum[StirlingSecond[300, k], {k, 300}];]}
Out[6]= {{4.39 Second, Null}, {7.37 Second, Null}}
```