Lecture 20 - Permutations and Combinations II

6.042 - May 1, 2003

The last two lectures considered the problem of counting the number of elements in a finite set. We saw several basic techniques:

- The sum, product, and division rules.
- The use of bijections.
- Permutations and combinations.
- Inclusion-exclusion.

This lecture covers some common counting problems that can be solved by combining these basic tools in various ways.

Nevertheless, many of the counting problems you subsequently encounter will *not* fall into one of the categories presented here. So your goal should *not* be to memorize an arcane formula for each type of problem we cover. Instead, make sure that you understand how the formulas given here follow from the basic counting techniques that you're already seen. An ability to apply these fundamental tools effectively and in combination will carry you much further than memorizing a long list of bizarre formulas.

There is another theme in this lecture. A wide range of counting problems can be recast as questions about *sequences of symbols*. For example, earlier we constructed a bijection between subsets of an n-element set and n-bit strings. Then we counted the number of n-bit strings using the product rule. In this way, we found that an n-element set has 2^n subsets. This suggests a general approach to counting problems:

1. Use a bijection to recast the counting problem at hand into a problem about sequences of symbols.
2. Solve the sequence-of-symbols problem.

This method works well for problems of the permutation/combination variety, but it does not work well for problems with an inclusion-exclusion flavor. We'll apply this recipe many times over the course of the lecture.

1 Balls into Bins

Suppose that there are r balls, numbered $1, \ldots, r$, and there are n bins, labeled $B_1, \ldots, B_n$. Each ball must be placed in a bin. How many arrangements are possible? For example, if $r = 2$ and $n = 2$, then four arrangements are possible:

In each diagram above, bin B_1 is on the left and B_2 is on the right.

We can construct a bijection from balls-and-bins arrangements to sequences with r symbols drawn from $B_1, \ldots, B_n$. In particular, if ball i is placed in bin B_j, then B_j is the i-th symbol in the corresponding sequence. Thus, the four arrangements shown above correspond to the following four sequences:

$$(B_1, B_1) \qquad (B_1, B_2) \qquad (B_2, B_1) \qquad (B_2, B_2)$$

All that remains is to count the number of different sequences containing r symbols drawn from $B_1, \ldots, B_n$. By the product rule, we have:

$$\begin{aligned} |\{B_1, \ldots, B_n\}^r| &= |\{B_1, \ldots, B_n\}|^r \\ &= n^r \end{aligned}$$

This is the number of ways to arrange r distinguishable balls in n distinguishable bins.

2 r-Permutations with Repetition

Recall that an *r-permutation of a set T* is a sequence of r distinct elements of T. If the set T has n elements, then the number of r-permutations of T is:

$$P(n, r) = \frac{n!}{(n-r)!}$$

Now we consider a variant of r-permutations where the elements in the sequence are no longer required to be distinct.

Definition 1 *An* r-permutation with repetition of a set T *is a sequence of* r *elements of* T*.*

For example, if T is the set $\{A, B, C\}$, then the 2-permutations of T are:

$$\begin{array}{ccc} (A, B) & (A, C) & (B, A) \\ (B, C) & (C, A) & (C, B) \end{array}$$

In contrast, the 2-permutations of T with repetition are:

$$\begin{array}{ccc} (A, B) & (A, C) & (B, A) \\ (B, C) & (C, A) & (C, B) \\ (A, A) & (B, B) & (C, C) \end{array}$$

The last three sequences are new arrivals.

The set of all r-permutations with repetition of the set T is simply:

$$\underbrace{T \times T \times \ldots \times T}_{r \text{ terms}} \quad = \quad T^r$$

If $|T| = n$, then this set has size n^r by the product rule. Thus, the number of r-permutations with repetition of an n-element set is n^r.

Looking back, what we did in the preceding section was to construct a bijection from balls-and-bins arrangments to permutations with repetition of the set of bins. Not surprisingly, we found that there were n^r arrangements, as well.

3 Permutations of a Multiset

How many different ways are there to arrange the letters in the word $TABLE$? This is simply the number of permutations of the set $\{T, A, B, L, E\}$, which is $5! = 120$. Some of the exciting possibilities include:

$$(B, L, E, A, T) \qquad\qquad (A, B, E, L, T) \qquad\qquad (T, B, L, E, A)$$

How many different ways are there to arrange the letters in the word BOO? We might like to say that this is equal to the number of permutations of the "set" $\{B, O, O\}$. But this makes no sense, because the elements of a set are required to be distinct. We need some new terminology. A *multiset* is a set where the elements need not be distinct. A *permutation* of a multiset is a sequence that contains each element as many times as it appears in the multiset. In these terms, what we're asking for is the number of permutations of the multiset $\{B, O, O\}$. Clearly, there are three:

$$(B, O, O) \qquad\qquad (O, B, O) \qquad\qquad (O, O, B)$$

How many ways are there to arrange the letters in the word $BEEP$? That is, how many permutations are there of the multiset $\{B, E, E, P\}$? We could work through all the cases explicitly, but let's aim for a general formula. Suppose we make all the letters distinct by adding subscripts:

$$\{B, E_1, E_2, P\}$$

This is an ordinary 4-element set, so we know that the number of permutations is 4!. Now erasing the subscripts on the E's defines a 2-to-1 mapping from permutations of this set to permutations of the multiset $\{B, E, E, P\}$. For example:

$$\left.\begin{array}{c}(E_1, P, E_2, B)\\(E_2, P, E_1, B)\end{array}\right\} \text{ map to } \quad (E, P, E, B)$$
$$\left.\begin{array}{c}(E_1, E_2, B, P)\\(E_2, E_1, B, P)\end{array}\right\} \text{ map to } \quad (E, E, B, P)$$
$$\text{etc.}$$

By the division rule, the number of permutations of the multiset $\{B, E, E, P\}$ is $4!/2 = 12$.

How many different ways are there to arrange the letters in the word $MISSISSIPPI$? That is, how many permutations are there of the multiset $\{M, I, S, S, I, S, S, I, P, P, I\}$? Once again, let's initially make all the letters unique by adding subscripts:

$$\{M, I_1, S_1, S_2, I_2, S_3, S_4, I_3, P_1, P_2, I_4\}$$

This is an ordinary 11-element set, so there are 11! permutations. Now, erasing the subscripts on the P's defines a 2-to-1 mapping from permutations of this set to permutations of the multiset:

$$\{M, I_1, S_1, S_2, I_2, S_3, S_4, I_3, P, P, I_4\}$$

Next, suppose that we erase the subscripts on the S's. Prior to the erasure, the subscripts on the four S's could appear in 4! different orders. Sequences that differ only with respect to the order of those subscripts are identical after the erasure. Thus, erasing the subscripts on the S's defines a 4!-to-1 mapping from permutations of the multiset above to permutations of the multiset:

$$\{M, I_1, S, S, I_2, S, S, I_3, P, P, I_4\}$$

Similarly, erasing the subscripts on the I's defines another 4!-to-1 mapping to permutations of the multiset:

$$\{M, I, S, S, I, S, S, I, P, P, I\}$$

At last, this is what we wanted to count in the first place! By applying the division rule at each of the steps above, we find that the number of different ways to arrange the letters in the word $MISSISSIPPI$ is:

$$\frac{11!}{2 \cdot 4! \cdot 4!}$$

The reasoning we used suggests a general formula for the number of permutations of a multiset.

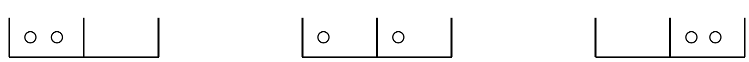
Fact 1 (Permutations of a Multiset) *Let M be a multiset with k different elements with multiplicities $n_1, n_2, \ldots, n_k$. Then the number of permutations of the multiset M is:*

$$\frac{(n_1 + n_2 + \ldots + n_k)!}{n_1!\ n_2!\ \ldots\ n_k!}$$

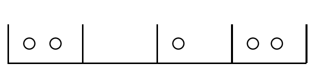
This formula is worth remembering. More importantly, make sure you understand the derivation. If you're going to follow the plan of recasting everything in terms of sequences of symbols, then you'd better know how to solve sequence-of-symbols problems!

4 Indistinguishable Balls into Bins

Suppose that there are r indistinguishable balls, and there are n bins, labeled $B_1, \ldots, B_n$. Each ball must be placed in a bin. How many arrangements are possible? The fact that balls are indistinguishable makes this problem quite different from the one we considered before, where the balls were numbered. For example, if $r = 2$ and $n = 2$, then only three different arrangements are possible, instead of four:



The solution to this problem uses a sneaky bijection, which is sometimes called *stars-and-bars*. We map each balls-and-bins configuration to a sequence of stars and bars by erasing the bottoms of the bins, erasing the leftmost wall and the rightmost wall, and replacing each ball with a star. For example:

 maps to $\star\star \mid \quad \mid \star \quad \mid \star\star$

This mapping allows us to recast the balls-and-bins problem as a sequence-of-symbols problem. More precisely, we have established a bijection between arrangements of r identical balls in n distinguishable bins and sequences containing r stars and $n-1$ bars. Note that the bars correspond to boundaries *between* bins; thus, an arrangement with n bins corresponds to a sequence with $n-1$ bars.

All that remains is to solve the sequence problem. That's easy: the sequences with r stars and $n-1$ bars are simply the permutations of a multiset with r stars and $n-1$ bars. According to our rule, the number of such permutations is:

$$\frac{(n+r-1)!}{r!\,(n-1)!} = \binom{n+r-1}{r}$$

There is another way to count the stars-and-bars strings: we must select r of the $n+r-1$ positions in the sequence to be stars, and the rest are bars. This selection can be done in $\binom{n+r-1}{r}$ ways, by our basic formula for r-combinations.

Either way, the number of ways to arrange r indistinguishable balls in n distinguishable bins turns out to be $\binom{n+r-1}{r}$.

5 r-Combinations with Repetition

Recall that an *r-combination of a set T* is a subset of T with exactly r-elements. Now we introduce a variation:

Definition 2 *An* r-combination with repetition of a set T *is a multiset of size r with elements drawn from T.*

For example, the 2-combinations with repetition of the set $T = \{A, B, C\}$ are the multisets:

$$\begin{array}{ccc} \{A, B\} & \{A, C\} & \{B, C\} \\ \{A, A\} & \{B, B\} & \{C, C\} \end{array}$$

How many r-combinations with repetition of an n-element set T are there?

This is another problem that can be solved by establishing a bijection to sequences of stars and bars. Order the n elements of T in some way, so that we can talk about the first element, second element, etc. Now map an r-combination with repetition to a sequence of stars and bars as follows. Draw $n-1$ bars. These bars define n segments. Each time the i-th element of T appears in the combination, put a star in the i-th segment.

For example, suppose $T = \{A, B, C, D, E\}$. The mappings of some 7-combinations with repetition of T to stars-and-bars sequences are shown below:

$$\begin{array}{lll} \{A, B, B, B, C, E, E\} & \text{maps to} & \star \mid \star\star\star \mid \star \mid\mid \star\star \\ \{B, B, C, C, C, D, D\} & \text{maps to} & \mid \star\star \mid \star\star\star \mid \star\star \mid \end{array}$$

In the first example, A appears once in the combination, and so there is one star in the first segment of the corresponding string. The element B appears three times, so there are three stars in the second segment, and so forth.

Thus, the number of r-combinations with repetition of an n-element set is equal to the number of sequences with r stars and $n - 1$ bars. As we saw before, this is equal to:

$$\binom{n + r - 1}{r}$$

Of course, there is nothing magical about stars and bars; we could equally well use sequences of x's and y's, but stars and bars are traditional.

Here is another way of looking at the problem. There is a bijection from r-combinations with repetition of an n-element set to arrangements of r-indistinguishable balls in n distinguishable bins: each time the i-th element appears in the combination, put a ball in the i-th bin. This bijection explains why we got the answer $\binom{n+r-1}{r}$ for both problems!

6 Distributions of Suits

Suppose that we distinguish the cards in a deck *only by suit.* Then the deck contains only four different types of card and contains 13 cards of each type. How many different 13-card hands are possible?

Here is one approach. The possible hands naturally correspond to the 13-combinations with repetition of the 4-element set $\{\spadesuit, \heartsuit, \diamondsuit, \clubsuit\}$. In the preceding section, we showed that the number of such combinations is:

$$\binom{4 + 13 - 1}{13} \quad = \quad 560$$

However, this approach sucks. Life is too short to spend time memorizing the definition of things like "r-combinations of an n-element set with repetition and extra pickles", let alone the associated formulas.

A better plan is to work from first principles. There is a bijection from hands to sequences of 3 bars and 13 stars: make one star for each spade, draw a bar, make a star for each heart, draw a bar, make a star for each diamond, draw a bar, make a star for each club. For example:

$$\spadesuit\spadesuit\spadesuit\spadesuit\heartsuit\heartsuit\heartsuit\diamondsuit\clubsuit\clubsuit\clubsuit\clubsuit\clubsuit \quad \text{maps to} \quad \star\star\star\star \mid \star\star\star \mid \star \mid \star\star\star\star\star$$

Now we've recast the card problem as a sequence-of-symbols problem, and we know how to solve sequence-of-symbols problems! The number of such stars-and-bars sequences is:

$$\frac{(13+3)!}{13!\ 3!} = 560$$

Because of the bijection, this is the number of different 13-card hands, if we distinguish cards only by suit.

6.1 Probability of a Distribution

We now know that there are 560 different hands. But these may not be equally likely. For example, surely a mix of suits is more likely than all spades. Let's compute the probability of being dealt a hand with exactly s spades, h hearts, d diamonds, and c clubs.

This problem is somewhat easier if we go back to regarding all 52 cards as distinct. Then the number of possible hands is $\binom{52}{13}$, and these are equally probable. All that remains is to count the number of hands with the right distribution of suits. We can choose the s spades in $\binom{13}{s}$ ways, the h hearts in $\binom{13}{h}$ ways, the d diamonds in $\binom{13}{d}$ ways, and the c clubs in $\binom{13}{c}$ ways. By the product rule, the number of hands with the proper suit distribution is the product of these four binomial coefficients. Therefore, the probability is:

$$\Pr\{(s,h,d,c)\} = \frac{\binom{13}{s}\binom{13}{h}\binom{13}{d}\binom{13}{c}}{\binom{52}{13}}$$

6.2 Indistinguishable Suits

We now know the probability of being dealt a specified number of spades, hearts, diamonds and clubs. Let's consider a slightly different question. Suppose that we're dealt a hand of 13 cards from a well-shuffled deck, and we look at the *multiset*:

$$\{\#\text{ spades}, \#\text{ hearts}, \#\text{ diamonds}, \#\text{ clubs}\}$$

Now, for example, a hand of 13 diamonds is equivalent to a hand of 13 spades; both map to the multiset $\{13,0,0,0\}$.

Each multiset corresponding to a 13-card hand has four elements that are in the range 0 to 13 that sum to 13. How many such multisets are there? Unfortunately, there is no tidy formula. But a case analysis shows that there are 39 possibilities.

Which multiset is the most likely? A natural guess is that the number of cards from each suit should be about the same, so the multiset $\{4, 3, 3, 3\}$ is most probable. There are four different suit distributions that yield this multiset, corresponding to the four ways to select the suit shared by 4 cards:

$$\begin{array}{cccc}
s = 4 & h = 3 & d = 3 & c = 3 \\
s = 3 & h = 4 & d = 3 & c = 3 \\
s = 3 & h = 3 & d = 4 & c = 3 \\
s = 3 & h = 3 & d = 3 & c = 4
\end{array}$$

Using this fact together with the formula for the probability of a suit distribution derived the preceding section, the probability of the multiset $\{4, 3, 3, 3\}$ is:

$$\begin{aligned}
\Pr\{\{4, 3, 3, 3\}\} &= \frac{\binom{13}{4}\binom{13}{3}\binom{13}{3}\binom{13}{3}}{\binom{52}{13}} \cdot 4 \\
&\approx \frac{1}{10}
\end{aligned}$$

Surprisingly, the most probable multiset is actually $\{4, 4, 3, 2\}$. The reason is that there are *twelve* different suit distributions that yield this multiset; there are four ways to choose the suit with 2 cards, and then three ways to choose the suit with 3 cards. Each of these twelve distributions is slightly less likely than each of the $\{4, 3, 3, 3\}$ distributions, but there are three times as many! The probability of this multiset is:

$$\begin{aligned}
\Pr\{\{4, 4, 3, 2\}\} &= \frac{\binom{13}{4}\binom{13}{4}\binom{13}{3}\binom{13}{2}}{\binom{52}{13}} \cdot 12 \\
&\approx \frac{1}{5}
\end{aligned}$$