Discrete Mathematics lecture notes 9-1

October 31, 2013

25. Counting with multisets

Some of you have asked the reasonable question: Why can't a set contain a given element many times? The answer is somewhat unsatisfying,[1] but this just means that there's an opening for a new definition that allows for repeated elements.

Definition 1. A *multiset* is an unordered list of elements, repeats allowed. A multiset will be denoted $M = \langle \dots \rangle$ to distinguish it from an ordinary set.

One can more formally define a multiset M to be a pair (X, m), where X is a set and $m : X \to \mathbb{N}_{\geq 1}$ is a function. m is the *multiplicity* function, recording how many times a given element of the set X is included in the multiset M. So for example, if $X = \{a, b, c\}$ and $m(a) = 2$, $m(b) = 1$, $m(c) = 3$, then $M = \langle a, a, b, c, c, c\rangle$.[2]

Just as looking at k-element subsets of a set of size n gives rise to the binomial coefficients, we can investigate interesting counting problems using the notion of multisets of a certain size.

Definition 2. Denote by $\left(\!\binom{n}{k}\!\right)$ the number of k-element multisubsets[3] of a set of size n.

Example 3. If $n = 2$, consider the set $X = \{a, b\}$. Then the set of multisubsets of X of size 2 is $\{\langle a, a\rangle, \langle a, b\rangle, \langle b, b\rangle\}$, and the set of multisubsets of size 3 is $\{\langle a, a, a\rangle, \langle a, a, b\rangle, \langle a, b, b\rangle, \langle b, b, b\rangle\}$. Thus $\left(\!\binom{2}{2}\!\right) = 3$ and $\left(\!\binom{2}{3}\!\right) = 4$.

This example show that, unlike with binomial coefficients, we can have $\left(\!\binom{n}{k}\!\right) \neq 0$ even when $k > n$.[4] Let's begin to investigate these symbols.

Proposition 4. *The following are true:*

(a) $\left(\!\binom{n}{0}\!\right) = 1$ *for all* $n \in \mathbb{N}$.

(b) $\left(\!\binom{0}{k}\!\right) = 0$ *for all* $k \in \mathbb{N}_{\geq 1}$.

(c) $\left(\!\binom{n}{1}\!\right) = n$ *for all* $n \in \mathbb{N}$.

(d) $\left(\!\binom{1}{k}\!\right) = 1$ *for all* $k \in \mathbb{N}$.

(e) $\left(\!\binom{n}{2}\!\right) = \binom{n}{2} + n$ *for all* $n \in \mathbb{N}$.

Proof. I'll leave (a)-(c) as an exercise. For (d), consider the set $X = \{a\}$ with $|X| = 1$. Then a k-element multisubset of X must be $\langle \underbrace{a, a, \dots, a}_{k \text{ times}} \rangle$, so there is only one option.

For (e), fix a set X with $|X| = n$, and consider a multisubset $M = \langle a, b\rangle$ of X. There are two options: Either $a \neq b$, or $a = b$. If $a \neq b$, then M is in fact a *subset* of X; there are by definition $\binom{n}{2}$ subsets of X of size 2. On the other hand, if $a = b$, there are exactly $|X| = n$ such 2-element multisubsets of X, when the claim. $\square$

[1] "Because we said so."

[2] Which is the same as $\langle a, b, c, a, c, c, \rangle$ or $\langle c, c, b, a, a, c\rangle$ or ... because the list M is *unordered*.

[3] A *multisubset* of a set X is a multiset M, all of whose elements are also elements of X. One could define a *submultiset* of a multiset N, if one cared to. How might you make such definition?

[4] In fact, $\left(\!\binom{n}{k}\!\right) \neq 0$ for *any* $k \in \mathbb{N}$ so long as $n \neq 0$.

This Proposition allows us to begin to construct the giant, doubly infinite array of all symbols $\left(\!\binom{n}{k}\!\right)$: Indexing the columns by n and the rows by k, we have

$\left(\!\binom{n}{k}\!\right)$	0	1	2	3	4	$\ldots$	n
0	1	1	1	1	1	$\ldots$	
1	0	1	2	3	4	$\ldots$	
2	0	1	3	6	?		
3	0	1	?	?	?		
$\vdots$	$\vdots$	$\vdots$					
k							

In order to go any farther, we'll either have to find a closed form expression for $\left(\!\binom{n}{k}\!\right)$, or a recursive relationship that will allow us to compute the next entry of the table in terms of other entries we already know. Or both.

Lemma 5. *For $n, k \in \mathbb{N}_{\geq 1}$, we have* $\left(\!\binom{n}{k}\!\right) = \left(\!\binom{n-1}{k}\!\right) + \left(\!\binom{n}{k-1}\!\right)$.

Proof. This is the multichoose analogue of Pascal's Lemma, and in fact the proof of Pascal's lemma will (basically) work for us here as well. Consider: $\left(\!\binom{n}{k}\!\right)$ is the the number of k-element multisubsets of X, where $|X| = n$. Fix some favorite element $x \in X$. If M is a k-element multisubset of X, there are two options: x appears in M, or x does not appear in M.

If x is not an element of M, then in fact M is a k-element multisubset of $X - \{x\}$; as $|X - \{x\}| = n - 1$, there are exactly $\left(\!\binom{n-1}{k}\!\right)$ such multisubsets, giving us our first term.

On the other hand, if x is en element of M we can delete *one* copy of x from M to obtain a $(k-1)$-element multisubset of X. Conversely, given a $(k-1)$-element multisubset N of X, adding one copy of x will yield a multisubset of this second type. As there are by definition $\left(\!\binom{n}{k-1}\!\right)$ such N, the result follows. $\square$

This in fact allows us to completely fill in the array of multichoices, at least in the sense that we can go as far out as we feel like: To determine the entry in a certain spot, simply add the entries of the spot directly above and the spot directly to the left.[5] Filling out more:

$\left(\!\binom{n}{k}\!\right)$	0	1	2	3	4	5	$\ldots$	n
0	1	1	1	1	1	1	$\ldots$	
1	0	1	2	3	4	5	$\ldots$	
2	0	1	3	6	10	15	$\ldots$	
3	0	1	4	10	20	35	$\ldots$	
4	0	1	5	15	35	70	$\ldots$	
5	0	1	6	21	56	126	$\ldots$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	
k								

Now, a funny thing happens.[6] If you delete the first column and rotate the rest of the array $\pi/4$ clockwise, you'll end up with an infinite triangle with rows the diagonals of the above array. Moreover, you might even recognize this triangle as having exactly the same entries as Pascal's triangle, at least as far as we've computed.[7] We conclude that the multichoose coefficients $\left(\!\binom{n}{k}\!\right)$ must be closely related the standard binomial coefficients $\binom{m}{\ell}$, at least up to a relabeling.

[5] Note that this only makes sense if both the spots above and to the left actually exist, so that we're not on one of the edges of the array. But note that the edges are the ones we fully understand: The first row is entirely made up of 1s, and the first column has one 1 then nothing but 0s. Also note that these cases correspond to the situation where $n = 0$ or $k = 0$, which is exactly the case that the Lemma does not cover.

[6] Maybe not so funny if you looked a little bit more closely at what's going on in the proof of part (e) of Proposition 4. . . .

[7] Which, of course, means that they have the same entries *everywhere*. Can you prove this?

Proposition 6. *If $n, k \in \mathbb{N}$, $(n,k) \neq (0,0)$, then* $\left(\!\binom{n}{k}\!\right) = \binom{n+k-1}{k}$.

Proof. Fix a set X of order n, and let $\mathcal{M}$ denote the set of k-element multisubsets of X. We will find a bijection between $\mathcal{M}$ and the set of k-element sub*sets* (not submultisets) of a set of size $n+k-1$ (to be determined in the course of the proof); the result will follow.

Without loss of generality, we may assume that $X = [n] = \{1, 2, \ldots, n\}$. We think of the elements of X as the labels of n boxes. If $M \in \mathcal{M}$ is some k-element multisubset, for each time that a number i occurs in M, put one ball in the box with label i. I leave it to you to verify that we've actually constructed a bijection between $\mathcal{M}$ and the number of ways one can put k indistinguishable balls in n boxes.

Now, imagine that the boxes 1 through n are arranged linearly, which we might represent by simply recalling the vertical dividers between them:

$$|\underbrace{\quad\;}_{1}|\underbrace{\quad\;}_{2}|\cdots|\underbrace{\quad\;}_{n}|$$

Arrange the k balls that determine M linearly between these dividers; for example, if $n = 4$ and $M = \langle 1, 1, 3, 4, 4, 4\rangle$, the corresponding picture would be

$$|\underbrace{\bullet\;\bullet}_{1}|\underbrace{\quad\;}_{2}|\underbrace{\;\bullet\;}_{3}|\underbrace{\bullet\;\bullet\;\bullet}_{4}|$$

Of course, the linear arrangement means we don't have to explicitly label the ith box, so we might as well just represent the same by

$$|\bullet\bullet||\bullet|\bullet\bullet\bullet|$$

In other words, we can uniquely represent an element of M as a linear arrangement of k dots and $n+1$ dashes, so long as the first vertical dash appears to the left of all dots and the last vertical dash appears to the right of all dots. This means that the first and last dash aren't actually doing anything, so we might as well remove them, as no data is lost:

$$\langle 1, 1, 3, 4, 4, 4\rangle = \bullet\bullet\,||\,\bullet\,|\,\bullet\bullet\bullet \qquad \text{and} \qquad \langle 2, 2, 2, 3, 3, 4\rangle = |\,\bullet\bullet\bullet\,|\,\bullet\bullet|\bullet$$

So we actually have a bijection between $\mathcal{M}$ and the linear arrangements of k dots and $n-1$ dashes with no other conditions. Such an arrangement is in turn equivalent to choosing from the $n+k-1$ positions possible, the k that will be occupied with a dot.[8] But the number of such choices is just $\binom{n+k-1}{k}$, giving us the desired result. □

[8]Or the $n-1$ that will be occupied with a dash. Why does this give the same result?