1 Fun with the combination statistic

Of the 12 enumerative functions we've seen, the one with the most obvious utility is the *combination statistic*, a.k.a. the *binomial coefficient*:

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} = \frac{n(n-1)(n-2)\dots(n-k+2)(n-k+1)}{k!}$$

That this is even guaranteed to be an integer is slightly surprising, but how useful it is is even more surprising. Today we're going to look at a few of those uses and segue into a powerful combinatorial tool.

1.1 The multinomial/multicombinations

The binomial was useful for enumerating the ways that two different types of item could be arranged: recall our items-and-dividers arguments, for instance. By selecting k elements "of type 1" from a set of size n, we also select $n-k$ elements of "type 2". So even though a binomial appears to be a selection of a single class of item, it is implicitly a division of an ordered list between two classes. But what do we do when we're dividing a list into more than 2 classes?

Question 1: *How many anagrams are there of the word "MISSISSIPPI"?*

Answer 1: *Here we want to divide the positions of an 11-letter word into 4 classes of specified sizes: 4 positions for "I", 4 positions for "S", 2 for "P", and 1 for "M". We may start by choosing 4 of our ten positions as "I" locations: there are* $\binom{11}{4}$ *ways to do that. Now we have 7 remaining locations, so there are* $\binom{7}{4}$ *ways to choose positions for "S", now of the remaining 3 positions, 2 must be "P" and there are* $\binom{3}{2}$ *ways to do that; finally, the position of the "M" is forced, so in total there are* $\binom{11}{4}\binom{7}{4}\binom{3}{2} = 34650$ *ways to do this.*

Note that if we had assigned items to classes in a different order, we'd have different enumeration statistics, e.g., if we chose a position for "M", then the two "P"s, and then the four "I"s, we'd have $\binom{11}{1}\binom{10}{2}\binom{8}{4}$, which would have the same value.

In fact, this would give us a free combinatorial identity:

Proposition 1. $\binom{n}{k}\binom{n-k}{r} = \binom{n}{r}\binom{n-r}{k}$

This identity could also be proven algebraically with ease, and combinatorially based on the above logic. Because multiple-class selections are useful enough, we actually have a generalization of the combination statistic:

Definition 1. The *multinomial coefficient* is defined as follows, for $n = k_1 + k_2 + \cdots + k_r$:

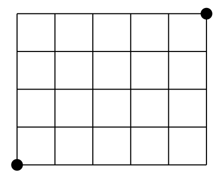
$$\binom{n}{k_1, k_2, \dots, k_r} = \binom{n}{k_1}\binom{n-k_1}{k_2}\binom{n-k_1-k_2}{k_3}\cdots\binom{k_{n-1}+k_n}{k_{n-1}}\binom{k_n}{k_n} = \frac{n!}{k_1!k_2!k_3!\cdots k_n!}$$

Note that the combination statistic is, as mentioned above, a division into clases of size k and $n-k$, so $\binom{n}{k} = \binom{n}{k, n-k}$. The permutation statistic ${}_nP_k$ can be considered as choosing $n-k$ non-selected elements, and then selecting each of the remaining elements for a distinguished class, so ${}_nP_k = \binom{n}{n-k,1,1,1,\dots,1}$.

We'll look at the multinomial in many of the same contexts we consider the binomial in.

1.2 Lattice paths

Paths through a lattice are a not-very-disguised application of the binomial coefficient. We might have a lattice as such:

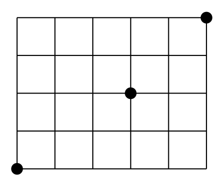


Question 2: *We know it would take 9 steps to get from the lower left corner to the upper right. How many ways are there of doing so?*

Answer 2: *A path can be uniquely matched with a sequence of instructions: each path consists of 4 upwards steps and 5 rightwards staps in some order. Thus, a path is uniquely determined by choosing 4 of the 9 steps to be upwards and filling the rest with rightwards steps. Thus there are* $\binom{9}{4} = 126$ *paths.*

Of course, we can find paths through a particular point, too, and use these to answer more sophisticated questions:

Question 3:



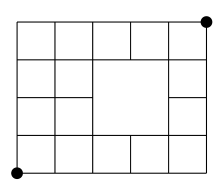
How many of the aforementioned 126 paths pass through the marked point?

Answer 3: *A path through the marked point can be thought of as a concatenation of two paths: one from the lower left to the waypoint, which consists of 2 upwards and 3 right steps, and a path from the waypoint to the upper right corner, which consists of 2 up and 2 right steps. There are* $\binom{5}{2}$ *ways to build the first path and* $\binom{4}{2}$ *ways to build the second, for a total of* $\binom{5}{2}\binom{4}{2} = 60$ *ways to construct a path assembled from the two of them.*

One might use this result to answer ancillary questions, for instance: what is the probability that a randomly selected path through this lattice will visit the waypoint? Clearly it is $\frac{60}{126} = \frac{10}{21}$.

A slightly subtler way to use the same result is as such:

Question 4: *How many paths are there through this lattice?*



Answer 4: *These are simply paths through the original lattice which* don't *pass through the point* $(3, 2)$. *We know how many paths there are total (126) and how many pass through that point (60), so the number that don't pass through that point is simply* $126 - 60 = 66$.

Note that the multinomial can be used here for more than 2 dimensions.

1.3 The Binomial (and Multinomial) Theorem

You are all probably familiar with the Binomial Theorem, which takes on one of the two following forms:

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}$$

$$(1+x)^n = \sum_{i=0}^{n} \binom{n}{i} x^i$$

It's from this statement that the name "binomial coefficient" comes, since the combination statistic is in fact a coefficient in a binomial expansion. Why should it be so? Perhaps the proof of the binomial theorem will illuminate matters.

Proposition 2. $(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}$.

Proof. We can prove this by induction. The case $(x+y)^0 = 1 = \sum_{i=0}^{0} \binom{0}{i} x^i y^{n-i}$ can be trivially verified.

Now we shall assume that $(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}$ and try to prove that $(x+y)^{n+1} = \sum_{i=0}^{n+1} \binom{n+1}{i} x^i y^{n+1-i}$:

$$\begin{aligned}
(x+y)^n &= \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i} \\
(x+y)(x+y)^n &= (x+y) \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i} \\
(x+y)^{n+1} &= \sum_{i=0}^{n} \binom{n}{i} x^{i+1} y^{n-i} + \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i+1} \\
(x+y)^{n+1} &= \sum_{i=0}^{n} \binom{n}{i} x^{i+1} y^{(n+1)-(i+1)} + \sum_{i=0}^{n} \binom{n}{i} x^i y^{n+1-i} \\
(x+y)^{n+1} &= \sum_{i=1}^{n+1} \binom{n}{i-1} x^i y^{n+1-i} + \sum_{i=0}^{n} \binom{n}{i} x^i y^{n+1-i} \\
(x+y)^{n+1} &= x^{n+1} y^0 + \sum_{i=1}^{n} \binom{n}{i-1} x^i y^{n+1-i} + x^0 y^{n+1} + \sum_{i=1}^{n} \binom{n}{i} x^i y^{n+1-i} \\
(x+y)^{n+1} &= x^{n+1} y^0 + \sum_{i=1}^{n} \left[\binom{n}{i-1} + \binom{n}{i} \right] x^i y^{n+1-i} + x^0 y^{n+1} \\
(x+y)^{n+1} &= x^{n+1} y^0 + \sum_{i=1}^{n} \binom{n+1}{i} x^i y^{n+1-i} + x^0 y^{n+1} \\
(x+y)^{n+1} &= \sum_{i=0}^{n+1} \binom{n+1}{i} x^i y^{n+1-i}
\end{aligned}$$

□

However, while that makes use of combinatorial identities, it doesn't give us a real combinatorial intuition of what the combination statistic is doing there in the first place. For better understanding, we might go to a combinatorial proof:

Proof. Let us expand $(x+y)(x+y)\cdots(x+y)$ making use only of the distributive law, but emphatically *not* commuting multiplication or collecting like terms. Then, since the ith binomial in the product contributes either an x or a y to each of the addends, the addends consist of all words of length n made of x and y. There will be 2^n such words (which can be confirmed by conisdering the result when $x = y = 1$), and there will be $\binom{n}{k}$ words with k xs and $n-k$ ys. Thus, on commuting, there will be $\binom{n}{k}$ terms of the form $x^k y^{n-k}$. □

This simple result is actually a key to a concept we will explore later! Here we have offered a free choice while simultaneously keeping a running total of how different choices are made; that is, multiplication by $(1+x)$ indicates a free choice betwen two alternatives, but stratifies based on how many times the second alternative was chosen.

Of course, a variant on the binomial coeffcent is the multinomial coefficient, which is effective in multinomials. One might, for instance, create a *trinomial theorem*:

$$(x+y+z)^n = \sum_{i+j+k=n} \binom{n}{i,j,k} x^i y^j z^k$$

But how about we get on the ball with that idea of "selection with running total"?

2 Generating functions

So we could read $\sum_{n=0}^{r} \binom{r}{n} x^n$ as encapsulating the number of ways to put n unlabeled balls in r labeled boxes, with a maximum of one per box, for a *range of values of* n. If we were to generalize this concept, we would have what is known as a *generating function*.

Definition 2. If a_n enumerates a number of ways to place, select, or otherwise arrange n identical objects, then $f(x) = \sum_{n=0}^{\infty} a_n x^n$ is called the *generating function* for a_n.

Some points of note: a generating function can frequently be a polynomial, as we saw above: the number of ways to put n balls in r boxes with no more than one per box is 0, so the power series described above would only have a finite number of terms (note: as a matter of convention, $\binom{n}{k} = 0$ for $k > n$; this is purely conventional rather than computational, since negative factorials are unevaluatable).

Also, given that this is an infinite series, note a question we do *not* ask: we don't actually care where the series converges! This is a purely algebraic construct, a *formal power series*; x represents only a variable. With rare exceptions we do not actually evaluate power series much.

This seems silly, though. Why is putting things in a power series more useful than putting them in an infinite sequence, or an infinite-dimensional vector, or any other information-organization structure?

The benefit of a power series shows up in a couple of moderately familiar principles:

Proposition 3. *If a_n enumerates a number of ways to place, select, or otherwise arrange n identical objects according to some set of rules A, and b_n enumerating likewise placement according to a completely disjoint ruleset B, and if c_n enumerates the number of ways to place n objects according to either A or B, then, clearly $c_n = a_n + b_n$, and in terms of generating functions:*

$$\sum_{n=0}^{\infty} c_n x^n = \sum_{n=0}^{\infty} a_n x^n + \sum_{n=0}^{\infty} b_n x^n$$

so we could represent the number of ways to perform one of two placement procedures using a version of the additive principle simultaneously adding over all values of n. But this could be done equally well with a vector. The big payoff of generating functions comes with the multiplicative principle:

Proposition 4. *If a_n enumerates a number of ways to place, select, or otherwise arrange n identical objects according to some procedure A, and b_n enumerating likewise placement according to a procedure B, and if c_n represents the number of ways to place n objects via* sequential performance of procedures A and B, *then $c_n = a_n b_0 + a_{n-1} b_1 + a_{n-2} b_2 + \cdots + a_0 b_n$. This is surprisingly easy to put in the language of generating functions:*

$$\sum_{n=0}^{\infty} c_n x^n = \sum_{n=0}^{\infty} (a_n b_0 + a_{n-1} b_1 + a_{n-2} b_2 + \cdots + a_0 b_n) x^n = \left(\sum_{n=0}^{\infty} a_n x^n\right)\left(\sum_{n=0}^{\infty} b_n x^n\right)$$

This is the big payoff, and it motivates our interpretation of the binomial theorem as a generating function: if we want to know how many ways there are to put n balls in r distinct boxes, with no more than 1 ball per box, we can think of this as a performance of r distinct placements of balls, each in a single box. For each box, we can place zero balls in one way, or one ball in one way, so the generating function associated with the number of ways to place balls in a single box is $1 + x$. Perform this procedure r times in sequence, and you have the generating function for the number of ways to place n balls in r boxes: $(1 + x)^r$.

2.1 Generating functions which are polynomial

This allows us to extend our understanding to more complicated variations on item-selection problems:

Question 5: *Of all 4-letter words one can make, how many have 1 vowel? How many have 2? How many have 3? How many have 4?*

Answer 5: *Let us consider the procedure of selecting a single letter, where the "objects" being tracked in this case are vowels. There are 26 ways to select the letter; 21 of them are not vowels, so they don't increase our count of tracked objects. 5 of them are vowels, associated with an object count of 1, so our generating function is $21x^0 + 5x^1$. We repeat this procedure 4 times, so the generating function representing our process as a whole is $(21 + 5x)^4 = 21^4 + 4 \cdot 21^3 \cdot 5x + 6 \cdot 21^2 \cdot 5^2 x^2 + 4 \cdot 21 \cdot 5^3 x^3 + 5^4 x^4$; the coefficients are our answers.*

We can also ask more advanced object-placement problems:

Question 6: *How many solutions are there to $x_1 + x_2 + x_3 = n$, for $0 \le x_i \le 2$? Alternatively, how many ways are there to distribute n balls among 3 distinct boxes, with no more than 2 balls per box?*

Answer 6: *Each box may have 0 balls placed in it, 1 ball placed in it, or 2 balls placed in it, so our generating function for an individual box-filling is* $1x^0 + 1x + 1x^2$. *Filling three boxes is thus represented by the g.f.* $(1 + x + x^2)^3 = 1 + 3x + 6x^2 + 7x^3 + 6x^4 + 3x^5 + x^6$.

Question 7: *A variation on the above: suppose we could place red, white, or blue balls, but still no more than 2 per box. How many such placements are there?*

Answer 7: *Here, each box may have 0 balls placed in it, 1 ball placed in it, or 2 balls placed in it, as above, but the number of ways to do so are not equal! There is only one way to buiild an empty box, but 3 ways to build a single-ball box, and 6 ways to build a 2-ball box, so our generating function for an individual box-filling is* $1x^0 + 3x + 6x^2$. *Filling three boxes is thus represented by the g.f.* $(1 + 3x + 6x^2)^3 = 1 + 9x + 45x^2 + 135x^3 + 270x^4 + 324x^5 + 216x^6$.

These are messy expansions above, but they're algebraic, which is a lot easier to brute-force than combinatorics is.

2.2 Infinite-series Generating Functions

All the examples we saw above were finite series, because the situations in question became impossible if n were large enough. In a lot of interesting generating function problems, however, we aren't so lucky, and as a result calculations require comfort with infinite series:

Question 8: *How many ways are there to distribute n identical balls among r distinguishable boxes with no restrictions on placement?*

Answer 8: *We did this with the twelvefold way, and we know up-front that the answer is* $\binom{n+r-1}{r-1}$. *But let's see what happens if we tackle it with generating functions!*

Into a single box, we can put 0 balls, or 1 ball, or 2 balls, or 3 balls, etc. The generating function associated with a single-box filling is thus $1 + x + x^2 + x^3 + x^4 + x^5 + \cdots$ *— already we get an infinite series! Fortunately, this is a familiar series, and one we can express as the nonpolynomial expression* $\frac{1}{1-x}$. *Thus, the generating function corresponding to an r-box placement is* $\frac{1}{(1-x)^r}$.

At this stage of answering the question, we seem to be stuck. But in fact, since we already knew the answer, we get a surprising series-expansion identity:

$$\frac{1}{(1-x)^r} = \sum_{n=0}^{\infty} \binom{n+r-1}{r-1} x^n$$

We will mostly find ourselves seeing some predictable forms for our distributions, which have convenient fcompact forms:

$$1 + x + x^2 + x^3 + \cdots = \frac{1}{1-x}$$

$$x^a + x^{a+1} + x^{a+2} + \cdots = \frac{x^a}{1-x}$$

$$1 + x + x^2 + x^3 + \cdots + x^b = \frac{1-x^{b+1}}{1-x}$$

$$x^a + x^{a+1} + x^{a+2} + \cdots + x^b = \frac{x^a - x^{b+1}}{1-x}$$

and most identical-ball distribution problems conform to these forms and can be answered using nothing but algebraic manipulation:

Question 9: *Box A must contain at least 3 balls. Box B must contain between 2 and 6 balls. Box C can contain any number of balls. How many ways are there to distribute n identical balls among the three boxes?*

Answer 9: *Box A corresponds to g.f.* $x^3 + x^4 + x^5 + \cdots = \frac{x^3}{1-x}$*. Box B corresponds to g.f.* $x^2 + x^3 + x^4 + x^5 + x^6 = \frac{x^2-x^7}{1-x}$*. Box C corresponds to g.f.* $1 + x + x^2 + x^3 + \cdots = \frac{1}{1-x}$*. Thus the generating function corresponding to a distriubtion into all 3 boxes is given by:*

$$
\begin{aligned}
\frac{x^3}{1-x}\frac{x^2-x^7}{1-x}\frac{1}{1-x} &= \frac{x^5-x^{10}}{(1-x)^3} \\
&= (x^5 - x^{10})\left(\sum_{n=0}^{\infty}\binom{n+3-1}{3-1}x^n\right) \\
&= \sum_{n=0}^{\infty}\binom{n+2}{2}x^{n+5} - \sum_{n=0}^{\infty}\binom{n+2}{2}x^{n+10} \\
&= \sum_{n=5}^{\infty}\binom{n-3}{2}x^n - \sum_{n=10}^{\infty}\binom{n-8}{2}x^n \\
&= \sum_{n=0}^{\infty}\binom{n-3}{2}x^n - \sum_{n=0}^{\infty}\binom{n-8}{2}x^n \\
&= \sum_{n=0}^{\infty}\left(\binom{n-3}{2} - \binom{n-8}{2}\right)x^n
\end{aligned}
$$