

# Symmetries of the Multinomial and Hypergeometric Functions

Aresh Pourkavoos

June 6, 2022

You've probably come across the binomial formula:

$$\binom{n}{k} = \frac{n!}{(n-k)!k!},$$

where  $n \in \mathbb{N}$  and  $0 \leq k \leq n$ , representing the number of different  $k$ -element subsets of an  $n$ -element set. The formula generates entries of Pascal's Triangle:

				1				
			1		1			
		1		2		1		
	1		3		3		1	
	1	4		6		4		1
1		5	10		10	5		1
1	6		15	20		15	6	1

where  $n$  is the row number and  $k$  is the position within the row. One of the most striking features of Pascal's Triangle is its mirror symmetry: two entries on opposite sides of the same row are the same. This makes sense in terms of the subset definition: for every  $k$ -element subset, there is exactly one  $n - k$ -element subset, namely its complement. Expressed in terms of the binomial formula, though, this symmetry isn't immediately obvious.

$$\binom{n}{n-k} = \frac{n!}{(n-(n-k))!(n-k)!} = \frac{n!}{k!(n-k)!} = \frac{n!}{(n-k)!k!} = \binom{n}{k}$$

However, with a slight change of variables, we can express this symmetry much more clearly. Let

$$j = n - k, n = j + k,$$

so that

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} = \frac{(j+k)!}{j!k!} = f(j, k).$$

This way, finding the entry on the opposite side of the row is as easy as switching the order of the arguments, which does not affect the output thanks to commutativity:

$$f(k, j) = \frac{(k+j)!}{k!j!} = \frac{(j+k)!}{j!k!} = f(j, k).$$

Note also that the restriction on the value of  $k$  is gone: the function returns a result for all  $j, k \in \mathbb{N}$ . But, how do we interpret  $j$  and this new formula in general? In terms of the traditional formula,  $j$  is the number of elements *not* chosen. Instead, we'll place the complementary subsets with sizes  $j$  and  $k$  on equal footing: the formula returns the number of ways to split an appropriately-sized (namely,  $(j+k)$ -element) set into  $j$ - and  $k$ -element subsets. This interpretation generalizes well to more subsets. For example, the number of ways to split a  $(j+k+l)$ -element set into  $j$ -,  $k$ -, and  $l$ -element subsets is

$$f(j, k, l) = \frac{(j+k+l)!}{j!k!l!}$$

In the general case, with  $m$  subsets which have  $k_1$  through  $k_m$  elements, the formula becomes

$$f(k_1, \dots, k_m) = \frac{\left(\sum_{i=1}^m k_i\right)!}{\prod_{i=1}^m (k_i!)},$$

known as the multinomial coefficient. It represents the number of ways to split an appropriately-sized set into (labeled)  $k_1$ -, ..., and  $k_m$ -element subsets. To me, both its notation and its pronunciation, “the factorial of the sum over the product of the factorials,” rhyme in a satisfying way. The symmetry group of this function is the group of all possible permutations of its arguments, known as the symmetric group of order  $n$ , or  $S_n$ .

Another function which appears often in combinatorics is the hypergeometric distribution, which answers the following question:

Let there be  $N$  objects in a population,  $n$  of which have some property, e.g. they are green while the other  $N - n$  objects are red. A sample of  $K$  objects is drawn from the population without replacement. What is the probability that exactly  $k$  of them are green?

There are  $\binom{N}{K}$  samples that could be drawn, so this is the denominator of the probability. The number of samples where exactly  $k$  are green may be counted by looking at the red and green drawn objects separately. There are  $\binom{N-K}{n-k}$  ways to draw the correct number of red balls, and there are  $\binom{K}{k}$  ways to draw the correct number of green balls, so the numerator is the product of these. Thus the hypergeometric distribution, usually with  $k$  as the variable since the probabilities for all  $k$  sum to 1, is

$$\frac{\binom{N-K}{n-k} \binom{K}{k}}{\binom{N}{n}}.$$

To express this function in a more symmetric way, we can introduce some new variables as shown in the following table:

	Green	Red	Total
Not chosen	$j_1$	$j_2$	$J$
Chosen	$k_1$	$k_2$	$K$
Total	$n_1$	$n_2$	$N$

Here,  $n_1 = n$  and  $k_1 = k$ . Also, the sum of the first two entries in each row is the last entry, and likewise for the columns, so

$$J = N - K, \quad j_1 = n_1 - k_1, \quad j_2 = J - j_1, \quad k_2 = K - k_1, \quad n_2 = N - n_1.$$

Then the binomial coefficients may be written

$$\begin{aligned} \binom{N-K}{n-k} &= \binom{J}{j_1} = \frac{J!}{j_1!j_2!}, \\ \binom{K}{k} &= \binom{K}{k_1} = \frac{K!}{k_1!k_2!}, \\ \binom{N}{n} &= \binom{N}{n_1} = \frac{N!}{n_1!n_2!}. \end{aligned}$$

Thus the total expression for the hypergeometric probability is

$$\frac{J!K!n_1!n_2!}{j_1!j_2!k_1!k_2!N!} = \frac{(j_1 + j_2)!(k_1 + k_2)!(j_1 + k_1)!(j_2 + k_2)!}{j_1!j_2!k_1!k_2!(j_1 + j_2 + k_1 + k_2)!}.$$

Args of  $F$  have symmetry group  $D_4$ : symmetries of a square

$$\begin{aligned}
F \begin{pmatrix} j_1, & j_2, \\ k_1, & k_2 \end{pmatrix} &= \frac{f(j_1, j_2)f(k_1, k_2)}{f(j_1 + k_1, j_2 + k_2)} = \frac{\binom{j_1+j_2}{j_1} \binom{k_1+k_2}{k_1}}{\binom{j_1+j_2+k_1+k_2}{j_1+k_1}} = \frac{\binom{J}{j_1} \binom{K}{k_1}}{\binom{N}{n_1}} \\
F \begin{pmatrix} j_2, & j_1, \\ k_2, & k_1 \end{pmatrix} &= \frac{f(j_2, j_1)f(k_2, k_1)}{f(j_2 + k_2, j_1 + k_1)} = \frac{\binom{j_2+j_1}{j_2} \binom{k_2+k_1}{k_2}}{\binom{j_2+j_1+k_2+k_1}{j_2+k_2}} = \frac{\binom{J}{j_2} \binom{K}{k_2}}{\binom{N}{n_2}} \\
F \begin{pmatrix} k_1, & k_2, \\ j_1, & j_2 \end{pmatrix} &= \frac{f(k_1, k_2)f(j_1, j_2)}{f(k_1 + j_1, k_2 + j_2)} = \frac{\binom{k_1+k_2}{k_1} \binom{j_1+j_2}{j_1}}{\binom{k_1+k_2+j_1+j_2}{k_1+j_1}} = \frac{\binom{K}{k_1} \binom{J}{j_1}}{\binom{N}{n_1}} \\
F \begin{pmatrix} k_2, & k_1, \\ j_2, & j_1 \end{pmatrix} &= \frac{f(k_2, k_1)f(j_2, j_1)}{f(k_2 + j_2, k_1 + j_1)} = \frac{\binom{k_2+k_1}{k_2} \binom{j_2+j_1}{j_2}}{\binom{k_2+k_1+j_2+j_1}{k_2+j_2}} = \frac{\binom{K}{k_2} \binom{J}{j_2}}{\binom{N}{n_2}} \\
F \begin{pmatrix} j_1, & k_1, \\ j_2, & k_2 \end{pmatrix} &= \frac{f(j_1, k_1)f(j_2, k_2)}{f(j_1 + j_2, k_1 + k_2)} = \frac{\binom{j_1+k_1}{j_1} \binom{j_2+k_2}{j_2}}{\binom{j_1+k_1+j_2+k_2}{j_1+j_2}} = \frac{\binom{n_1}{j_1} \binom{n_2}{j_2}}{\binom{N}{J}} \\
F \begin{pmatrix} k_1, & j_1, \\ k_2, & j_2 \end{pmatrix} &= \frac{f(k_1, j_1)f(k_2, j_2)}{f(k_1 + k_2, j_1 + j_2)} = \frac{\binom{k_1+j_1}{k_1} \binom{k_2+j_2}{k_2}}{\binom{k_1+j_1+k_2+j_2}{k_1+k_2}} = \frac{\binom{n_1}{k_1} \binom{n_2}{k_2}}{\binom{N}{K}} \\
F \begin{pmatrix} j_2, & k_2, \\ j_1, & k_1 \end{pmatrix} &= \frac{f(j_2, k_2)f(j_1, k_1)}{f(j_2 + j_1, k_2 + k_1)} = \frac{\binom{j_2+k_2}{j_2} \binom{j_1+k_1}{j_1}}{\binom{j_2+k_2+j_1+k_1}{j_2+j_1}} = \frac{\binom{n_2}{j_2} \binom{n_1}{j_1}}{\binom{N}{J}} \\
F \begin{pmatrix} k_2, & j_2, \\ k_1, & j_1 \end{pmatrix} &= \frac{f(k_2, j_2)f(k_1, j_1)}{f(k_2 + k_1, j_2 + j_1)} = \frac{\binom{k_2+j_2}{k_2} \binom{k_1+j_1}{k_1}}{\binom{k_2+j_2+k_1+j_1}{k_2+k_1}} = \frac{\binom{n_2}{k_2} \binom{n_1}{k_1}}{\binom{N}{K}}
\end{aligned}$$