Second Lecture: 23/3

Theorem 2.1. (Binomial Theorem) *Let n be a non-negative integer. Then*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}. \tag{2.1}$$

Proof. When we fully expand $(x+y)^n$ there are a total of 2^n terms of the form $x^k y^{n-k}$, for some $0 \leq k \leq n$. This is because we can choose either x or y from each factor (2 choices) and there are n factors - so apply the Multiplication Principle.

For a fixed k, let us consider the number of times the term $x^k y^{n-k}$ occurs in the expansion. To get this term we must choose an x from k factors, and then a y from each of the remaining $n-k$ factors. There are $\binom{n}{k}$ choices for the k factors from which to choose x, hence this will be the number of times the term $x^k y^{n-k}$ occurs in the expansion. □

Terminology. Because of their appearence as coefficients in the Binomial Theorem, the numbers $\binom{n}{k}$ are usually referred to as *binomial coefficients*.

Remark 2.2. In the above proof we used the fact that ordinary multiplication of numbers is commutative - it allowed us to say that one got the same term $x^k y^{n-k}$ irrespective of which k factors one chose x from. Hence, there is no "binomial theorem" in a non-commutative ring, for example if x and y were matrices.

When computing binomial coefficients, the following recursive formula for them is often useful:

Proposition 2.3. (Pascal's identity)

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}. \tag{2.2}$$

There are many ways to prove this, but one way which I think gives "insight" (i.e.: explains how on earth one might *discover* such a formula rather than just *verify* it). This is the proof given below. But let me mention a couple of alternative proofs, which I will leave as exercises to the reader to work out in detail:

ALTERNATIVE 1: Use induction on a suitable quantity.
ALTERNATIVE 2: Use formula (1.8) and some algebraic manipulation.

Proof. The incisive proof involves *combinatorial* reasoning. Firstly, the LHS of (2.2) is, by definition, the number of ways to choose k distinct elements from an $(n+1)$-element set. Isolate one of the $n+1$ elements and consider two cases:

CASE 1: This element is among the k chosen. Then it remains to choose $k-1$ distinct elements from n. By definition, there are $\binom{n}{k-1}$ ways to do this.

CASE 2: This element is not among the k chosen. Then it remains to choose k distinct elements from n. By definition, there are $\binom{n}{k}$ ways to do this.

Cases 1 and 2 are obviously mutually exclusive (i.e.: disjoint) and it is an either/or situation so, by the addition principle, the total number of possibilities for the full choice of k elements is $\binom{n}{k-1} + \binom{n}{k}$. $\square$

Pascal's identity gives rise to the famous *Pascal's triangle*. Rather than me wasting time drawing a nice picture in `Latex`, just look at any of the zillions of Googleable pictures.

Example 2.4. Let us expand $(2x-y)^5$. Note that, when applying the Binomial Theorem as stated above, the role of "x" is now played by $2x$ and the role of "y" is played by $-y$. Thus,

$$(2x-y)^5 = \binom{5}{0}x^0y^5 + \binom{5}{1}x^1y^4 + \binom{5}{2}x^2y^3 + \binom{5}{3}x^3y^2 + \binom{5}{4}x^4y^1 + \binom{5}{5}x^5y^0. \tag{2.3}$$

The binomial coefficients are given by the corresponding row of Pascal's triangle, which reads as $1,\ 5,\ 10,\ 10,\ 5,\ 1$. Thus the expansion becomes

$$\begin{aligned}&(2x-y)^5 = \\ &= -1\cdot 1\cdot y^5 + 5\cdot(2x)\cdot y^4 - 10\cdot(4x^2)\cdot y^3 + 10\cdot(8x^3)\cdot y^2 - 5\cdot(16x^4)\cdot y + 1\cdot(32x^5)\cdot 1 = \\ &= -y^5 + 10xy^4 - 40x^2y^3 + 80x^3y^2 - 80x^4y + 32x^5.\end{aligned}$$

Example 2.5. *Compute the coefficient of $x^6y^7z^4$ in the expansion of $(x+y+z)^{17}$.*

SOLUTION: We now have three variables instead of two, but we can perform the same combinatorial reaoning as in our proof above of the Binomial Theorem. To get a term of the form $x^6y^7z^4$ in the expansion, one must choose an x from 6 of the 17 factors, then choose a y from 7 of the remaining 11 factors. There are $\binom{17}{6}$ possibilities for the factors from which to choose the x:s, then $\binom{11}{7}$ possibilities for the factors from which to choose the y:s. By the multiplication principle, the total number of possibilities is $\binom{17}{6} \times \binom{11}{7}$.

Remark 2.6. The same kind of reasoning can be extended to an arbitrary number of variables and one can formulate a general *Multinomial Theorem*. The only thing that is really more complicated in the general case is the notation, so I will leave it to yourselves to think how to write out the correct formulation, or just Google it.

Remark 2.7. If one has nothing better to do going home on the tram, one can use (1.8) and work out that $\binom{17}{6} \times \binom{11}{7} = 12376 \times 330 = 4084080$, so about 4 million. Binomial coefficients rapidly become large and often what one needs is just good estimates rather than exact values. So, suppose you had 2 minutes to save your life and had to estimate, up to a factor of 100, the value of, say, $\binom{63}{19}$. How would you do it ? I will leave this (hopefully intruiging !) question hanging and maybe come back to the topic of efficiently estimating binomial coefficients later.

Balls and Bins. Computer scientists *love* to talk in terms of placing balls in bins. There are essentially 4 different problems here, depending on whether or not the balls are *distinguishable*, and similarly for the bins. The problems are simplest when the bins are distinguishable, and we will provide answers below. The cases when the bins are indistinguishable will be returned to later - see Remark 2.13 below.

Proposition 2.8. *The number of ways to distribute n distinguishable balls among k distinguishable bins is k^n.*

Proof. There are k choices for the bin into which to place each ball, and there are n balls. Now apply the Multiplication Principle. □

Proposition 2.9. *The number of ways to distribute n indistinguishable (i.e.: identical) balls among k distinguishable bins is $\binom{n+k-1}{n} = \binom{n+k-1}{k-1}$.*

Proof. The difference here from the previous proposition is that, since the balls are now identical, all that matters is *how many* balls are placed in each bin. To see where the formula comes from, we can observe that there is a 1-1 correspondence between the possible distributions of the balls and sequences of $n + k - 1$ symbols of which n are identical "dots" and $k - 1$ are identical "dashes". For we can interpret the dashes as marking out where one "jumps" from one bin to the next. It's probably clearest with an example. Consider the following sequence of dots and dashes:

$$\bullet\,\bullet \mid \bullet\,\bullet\,\bullet \mid\mid \bullet \mid \bullet\,\bullet\,\bullet\,\bullet \mid \bullet\,.$$

This corresponds to a distribution of 11 balls into 6 bins, where the number of bins placed in bins $1 - 5$ is, respectively, $2,\ 3,\ 0,\ 1,\ 4,\ 1$.

To complete the proof, note that the number of sequences of n dots and $k - 1$ dashes is obviously $\binom{n+k-1}{n}$ since one just needs to choose in which n positions to place the dots. □

Example 2.10. Proposition 2.9 is used in statistical physics, where the balls are (elementary) particles and the bins are quantum energy levels. If you're interested, see `https://en.wikipedia.org/wiki/Bose-Einstein_statistics`

Example 2.11. Another interpretation of Proposition 2.9 is that $\binom{n+k-1}{n}$ is the number of solutions in non-negative integers to the equation

$$x_1 + x_2 + \cdots + x_k = n, \quad x_i \in \mathbb{N}_0. \tag{2.4}$$

For we can interpret x_i as the number of balls placed in bin number i. Solutions to (2.4) are usually referred to as *compositions of n into at most k parts*. The "at most" comes from the fact that the x_i are allowed to equal zero.

Example 2.12. Let $\mathbb{Z}^k$ denote the k-dimensional integer lattice, that is, the lattice of points in k-dimensional Euclidean space $\mathbb{R}^k$ all of whose coordinates are integers. We can interpret $\binom{n+k-1}{n}$ as the number of possible destinations of an n-step path in this lattice, starting from the origin and such that every step is in the positive direction along a coordinate axis. For we can interpret x_i in (2.4) as the number of steps taken along the i:th coordinate direction and note that, what determines where one ends up is simply the number of steps taken in each direction, not the order.

The technical terminology here would be to speak of the number of possible destinations for an n-step *simple, positively oriented, k-dimensional random walk*.

Remark 2.13. When counting the number of ways to place distinguishable balls into identical bins one encounters so-called *Stirling numbers*. When placing identical balls into identical bins one encounters so-called *partitions of n (into at most k parts)*. These problems are more difficult, no exact formulas are known but one can write down some recursive formulas. We will return to these topics in Lecture xx. See Chapter 12 of Biggs.

Inclusion-Exclusion Principle (also called Sieve Principle). This is a very elegant and useful generalisation of the addition principle to the case of sets that are not pairwise disjoint.

The case of two sets.

$$|A \cup B| = |A| + |B| - |A \cap B|. \tag{2.5}$$

The case of three sets.

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|. \tag{2.6}$$

There is a general pattern, given by the following result which we will prove next time:

Theorem 2.14. *(***Inclusion-Exclusion Principle***) Let $A_1, A_2, \ldots, A_n$ be finite sets. Then*

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{i=1}^{k} |A_i| - \sum_{i \neq j} |A_i \cap A_j| + \sum_{i \neq j \neq k} |A_i \cap A_j \cap A_k| - \cdots + (-1)^{n-1} |A_1 \cap \cdots \cap A_n|. \tag{2.7}$$