2nd Lecture: 18/1

Multinomial Theorem. The case of two variables you have probably all seen before.

Proposition 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.

Remark 2.2. Their appearence as coefficients in the Binomial Theorem explains why the numbers $\binom{n}{k}$ are referred to as *binomial coefficients*.

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 two results are useful:

Proposition 2.3. *For each $0 \leq k \leq n$ one has*

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

PROOF: The number of ways to choose k elements from n is the same as the number of ways to reject $n-k$ elements. Alternatively, use Proposition 1.5.

Proposition 2.4. (Pascal's identity) *For each $1 \leq k \leq n+1$ one has*

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}. \tag{2.3}$$

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 Proposition 1.5 and some algebraic manipulation.

NICEST PROOF: This proof involves *combinatorial* reasoning. Firstly, the LHS of (2.3) 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 $\binom{n}{k-1} + \binom{n}{k}$.

Along with the *initial conditions* $\binom{n}{0} = \binom{n}{n} = 1 \ \forall \, n$, eq. (2.3) gives a *recursive formula* for computing binomial coefficients. The numbers are usually represented in *Pascal's triangle*, which you can Google a picture of. The initial conditions give the 1s on the sides of the triangle, while (2.3) means that every other number is the sum of the two diagonally above it. The symmetry of the triangle about the perpendicular bisector corresponds to (2.2).

Example 2.5. 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,

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

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

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

We now generalize the Binomial Theorem to an arbitrary number of variables:

Theorem 2.6. (Multinomial Theorem) *Let $n,\ n_1,\ n_2,\ \ldots,\ n_k$ be non-negative integers such that $\sum_{i=1}^{k} n_i = n$. Then the coefficient of $x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k}$ in the expansion of $(x_1 + x_2 + \cdots + x_k)^n$ is*

$$\frac{n!}{\prod_{i=1}^{k} n_i!} \tag{2.4}$$

Remark 2.7. Note that this does indeed reduce to the Binomial Theorem in the case $k=2$. For then it says that the coefficient of $x_1^{n_1} x_2^{n_2}$ in the expansion of $(x_1+x_2)^n$ is $\frac{n!}{n_1!\,n_2!}$. Changing notation according to $x_1 \to x$, $x_2 \to y$, $n_1 \to k$, $n_2 \to n-k$ yields (2.1).

PROOF OF THEOREM 2.6.: As in the proof of Proposition 2.1, we begin by observing that when we fully expand $(x_1 + \cdots + x_k)^n$, there are a total of k^n terms. To get a term $\prod_{i=1}^{k} x_i^{n_i}$ one must choose x_i from n_i factors, for each $i = 1, \ldots, k$. Hence, the coefficient of this term is the number of n-letter "words" (a.k.a. strings) in which n_i of the "letters" are x_i, for each $i = 1, \ldots, k$. There are $n!$ permutations of the n letters but, for each i, the $n_i!$ internal permutations of the copies of x_i do not change the word. By MP, there are thus $P = \prod_{i=1}^{k} n_i!$ permutations which do not change the word, and hence $\frac{n!}{P}$ different words, q.e.d.

Exercise 2.8. (a) How many different words can one make using all the letters in

DISKRETMATEMATIK

(b) How many 5-letter words can one make which include all three Ts and otherwise only vowels ?
(c) If one chooses 3 of the 16 letters at random, what is the probability that one chooses 3 consonants ?
(d) Suppose you had available infinitely many copies of each of the 9 different letters D, I, S, K, R, E, T, M, A. In how many ways could you choose 13 letters, if order doesn't matter ?

SOLUTION: (a) Direct application of Theorem 2.6: $\frac{16!}{(2!)^5\, 3!}$.

(b) There are $\binom{5}{3} = 10$ ways to place the Ts. In the remaining two positions we can either (i) place two copies of the same vowel A, E or I or (ii) place two different vowels, in which case the order matters. By AP, there are $3 + 3 \cdot 2 = 9$ ways to place the vowels, hence by MP a total of $10 \cdot 9 = 90$ possible words.

(c) *1st solution*: Whenever the letters are chosen (uniformly) at random, the probability of some event E is given by

$$\mathbb{P}(E) = \frac{\text{numbers of choices in which } E \text{ occurs}}{\text{total number of possible choices}}.$$

The total number of possible choices of 3 letters is $\binom{16}{3}$. Since 10 of the 16 letters are consonants, the numerator is $\binom{10}{3}$. Hence, the desired probability is $\binom{10}{3}/\binom{16}{3}$.

2nd solution: We think in terms of conditional probabilities. Let E be the event that all 3 chosen letters are consonants. For each $i = 1, 2, 3$, let E_i be the event that the i:th letter chosen is a consonant. Thus $E = E_1 \wedge E_2 \wedge E_3$ and so we have

$$\mathbb{P}(E) = \mathbb{P}(E_1) \cdot \mathbb{P}(E_2 \,|\, E_1) \cdot \mathbb{P}(E_3 \,| E_1 \wedge E_2). \tag{2.5}$$

Since 10 of the 16 letters are consonants we have $\mathbb{P}(E_1) = \frac{10}{16} = \frac{5}{8}$. Given E_1, 9 of the remaining 15 letters are consonants, so $\mathbb{P}(E_2 \,|\, E_1) = \frac{9}{15} = \frac{3}{5}$. Given $E_1 \wedge E_2$, 8 of the remaining 14 letters are consonants, so $\mathbb{P}(E_3 \,|\, E_1 \wedge E_2) = \frac{8}{14} = \frac{4}{7}$.

Substituting into (2.5) we have $\mathbb{P}(E) = \frac{5}{8} \cdot \frac{3}{5} \cdot \frac{4}{7} = \frac{3}{14}$.

(d) Since order doesn't count, all that matters is how many times we choose each of the 9 different letters. Thus we can apply Proposition 1.9 with $k = 13$ and $n = 9$. Answer: $\binom{13+9-1}{13} = \binom{21}{13} = \binom{21}{8}$.

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.

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

The case of three sets.

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|. \tag{2.7}$$

There is a general pattern, given by the following result:

Theorem 2.9. **(Inclusion-Exclusion Principle)** *Let $A_1, A_2, \ldots, A_n$ be finite sets. Then*

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{i=1}^{k} |A_i| - \sum_{i \neq j} |A_i \cap A_j| + \sum_{i \neq j \neq k} |A_i \cap A_j \cap A_k| - \cdots + (-1)^{n-1} |A_1 \cap \cdots \cap A_n|. \tag{2.8}$$

In order to prove the Inclusion-Exclusion principle, we need a lemma:

Lemma 2.10. *Let $n \geq 1$. Then*

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0. \tag{2.9}$$

PROOF: Simply apply the Binomial Theorem (2.1) with $x = -1$, $y = 1$.

Remark 2.11. Alternatively, (2.9) says that if $n \geq 1$ then the sum of $\binom{n}{k}$ over all odd k equals the sum over all even k. In other words, it says that exactly half the subsets of an n-element set have an odd, resp. even number of elements. I leave it as an exercise for you to prove this “combinatorially”, i.e.: by describing an explicit bijection between the odd-size and even-size subsets of $\{1, \ldots, n\}$ for any $n \geq 1$.

PROOF OF THEOREM 2.9. Let x be any element of the union. It suffices to show that x is counted exactly once on the RHS of (2.8). Suppose x belongs to m of the n sets A_i, for some $1 \leq m \leq n$. The symmetry of the formula (2.8) means that we can assume, without loss of generality (WLOG), that x belongs to $A_1, A_2, \ldots, A_m$. Then, on the RHS of (2.8), x is counted $(-1)^{k-1}$ times for each k-element subset of $\{1, 2, \ldots, m\}$, where $1 \leq k \leq m$ (note that the empty subset does not appear). Hence, the total number of times x is counted is

$$\sum_{k=1}^{m}(-1)^{k-1}\binom{m}{k} = -\left(\sum_{k=0}^{m}(-1)^{k}\binom{m}{k} - (-1)^{0}\binom{m}{0}\right) \overset{\text{Lemma 2.10}}{=} -(0-1) = 1, \quad \text{v.s.v.}$$

Example 2.12. The I-E principle can be used to count the number of *surjective* functions $f : A \to B$, where A and B are two finite sets with $|A| \geq |B|$. See Homework 1.

Example 2.13. (Derangements) The set of all $n!$ permutations of $\{1, \ldots, n\}$ is called the *symmetric group of order* n and is denoted S_n.

Let $\pi \in S_n$, i.e.: it is a bijective function from $\{1, 2, \ldots, n\}$ to itself. We say that π is a *derangement* if $\pi(i) \neq i$ for all $i = 1, 2, \ldots, n$. In words, a derangement is a permutation with no fixed points. We denote by d_n the number of derangements in S_n. Then there is the following beautiful result:

$$\lim_{n\to\infty} \frac{d_n}{n!} = \frac{1}{e}. \tag{2.10}$$

One may think of (2.10) as saying that the probability of a random permutation of n objects being a derangement goes to $1/e$, as $n \to \infty$. This is usually popularised as the *hat problem* or the *cloakroom problem* - see Section 2 of the wiki article:
`https://en.wikipedia.org/wiki/Derangement`

To prove (2.10), let us being by recalling the Taylor series of the exponential function:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

Taking $x = -1$ we obtain

$$e^{-1} = \frac{1}{e} = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \tag{2.11}$$

Now back to derangements. For each $i = 1, 2, \ldots, n$, let $A_i := \{\pi \in S_n : \pi(i) = i\}$. Then, by definition, d_n is the cardinality of the set difference $S_n \setminus \cup_{i=1}^{n} A_i$. By I-E,

$$d_n = |S_n| - \sum_{i=1}^{n} |A_i| + \sum_{i\neq j} |A_i \cap A_j| - \cdots + (-1)^n |A_1 \cap \cdots \cap A_n|.$$

The important point is that each of the terms on the RHS is easy to compute. We have $|S_n| = n!$. Next take some A_i. Since i is fixed, a permutation belonging to A_i can act freely as a permutation of the remaining $n-1$ elements of $\{1, 2, \ldots, n\}$. Hence $|A_i| = (n-1)!$. Then take some $A_i \cap A_j$. A permutation in this set fixes both i and j, hence can freely permute the remaining $n-2$ elements of $\{1, 2, \ldots, n\}$. Thus $|A_i \cap A_j| = (n-2)!$. And so on ... We conclude that

$$d_n = n! - n \times (n-1)! + \binom{n}{2} \times (n-2)! - \binom{n}{3} \times (n-3)! + \cdots + (-1)^n \times \binom{n}{n} \times 0!$$

Using Proposition 1.5, this can be rewritten as

$$d_n = \frac{n!}{0!} - \frac{n!}{1!} + \frac{n!}{2!} - \frac{n!}{3!} + \cdots + (-1)^n \times \frac{n!}{n!} \Rightarrow$$

$$\Rightarrow \frac{d_n}{n!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!}$$

Taking limits as $n \to \infty$ and using (2.11), we obtain (2.10), v.s.v.

Remark 2.14. The factor $1/e$ appears as a limiting probability in a similar, equally famous, but considerably more subtle problem known colloquially as the *Secretary Problem*. See the wiki article if you're interested:
`https://en.wikipedia.org/wiki/Secretary_problem`

Pigeonhole Principle. This is usually presented somwhat informally as follows:

Pigeonhole Principle: If $n+1$ pigeons are to be distributed among n pigeonholes, then at least one pigeonhole must receive at least two pigeons.

A more precise formulation would be to say that if A, B are finite sets with $|A| > |B|$, then no function $f : A \to B$ can be injective.

Extended Pigeonhole Principle: If $k \cdot n + 1$ pigeons are to be distributed among n pigeonholes, then at least one pigeonhole must receive at least $k+1$ pigeons.

These very, very simple principles are used all over the place, often perhaps without realising it. We will see examples throughout the course. Here I give one rather simple and one not so simple application. The latter is amongst the most famous ones and also very clever, illustrating how sometimes in mathematics (as in life !), the key to solving a problem is to get the right idea, which may be based on something very simple but where it is not at all obvious how.

Example 2.15. I claim that amongst any 8 integers there must be a pair whose difference is a multiple of 7. Let A be the set of these 8 integers and let $B = \{0, 1, \dots, 6\}$. Define a map $f : A \to B$ by

$$f(x) = x \,(\text{mod } 7).$$

Now $|A| = 8$ and $|B| = 7$, so $|A| > |B|$ and hence, by PHP, the function f cannot be injective. Hence there must be a pair of numbers $a_1 \neq a_2$ such that $f(a_1) = f(a_2)$. But this means that $a_1 \equiv a_2 \pmod 7$ and hence that $a_1 - a_2$ is a multiple of 7, q.e.d.

Remark 2.16. In this example, we applied PHP to prove that there must *exist* a pair of numbers having the desired property. However, the argument gives us no insight as to how to *find* these two numbers, other than by an exhaustive search of all $\binom{8}{2} = 28$ pairs. This is typical of applications of PHP: they provide *pure existence arguments* which are *algorithmically* worthless.

Definition 2.17. A sequence $a_1, a_2, \ldots$ of distinct real numbers is said to be *monotone* if it is either strictly increasing or strictly decreasing.

Example 2.18. (Erdős-Szekeres theorem (1935)) *In any sequence of* $n^2 + 1$ *distinct real numbers, there must exist a montone subsequence of* $n + 1$ *numbers.*

Before beginning the proof, let us show that the result is "optimal" in the sense that it is possible to write down, for every $n \in \mathbb{Z}_+$, a sequence of n^2 distinct reals without any monotone subsequence of length $n + 1$. The idea is as follows:

1. Take any collection of n^2 distinct reals. Divide the n^2 numbers into n equally sized groups $G_1, G_2, \ldots, G_n$, each consisting of n numbers, such that G_1 consists of the smallest n numbers in the collection, G_2 contains the next n smallest and so on.
2. Write the entire collection from left to right as $G_1 \, G_2 \, \ldots \, G_n$.
3. Now rearrange the numbers inside each group G_i so that, from left to right, they form a decreasing sequence.

It is easy to see that the sequence thus obtained contains no monotone subsequence of length $n + 1$. Since the numbers are monotone decreasing within each group G_i, an increasing subsequence can contain at most one number from each group, hence at most n numbers in all. Similarly, since whenever $i < j$, every number in the group G_i is less than every number in the group G_j, a decreasing subsequence can only contain numbers from within a single group, hence at most n numbers.

PROOF OF ER-SZ THEOREM: Denote the sequence, written from left to right, as $x_1, x_2, \ldots, x_{n^2+1}$. For each $i = 1, 2, \ldots, n^2+1$, let $\mathcal{L}_i$ denote the length of the longest increasing subsequence whose first term is x_i. If there were no increasing subsequence at all of length $n + 1$ then, in particular, each $\mathcal{L}_i$ must be an integer among $1, 2, \ldots, n$. But there are $n^2 + 1$ of them so, by the Extended Pigeonhole Principle, some $n + 1$ of them must all be equal, i.e.: there must exist $1 \leq i_1 < i_2 < \cdots < i_{n+1} \leq n^2 + 1$ such that $\mathcal{L}_{i_1} = \mathcal{L}_{i_2} = \cdots = \mathcal{L}_{i_{n+1}}$. But now you just have to stop and think and unravel what this means: it means that the numbers $x_{i_1}, x_{i_2}, \ldots, x_{i_{n+1}}$ must form a decreasing subsequence of length $n + 1$. We're done !