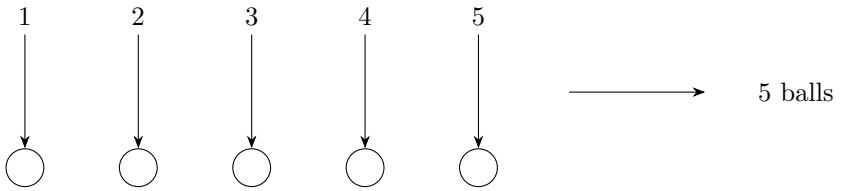
Discrete Math Notes

Professor Abramenko

Fall 2025

1 Counting and Maps Between Finite Sets

How do we count? We see a cluster of objects, e.g. balls. We then associate 1 with the first ball we see, 2 with the second, and so on, until no more balls are left. Then the last number we associated to a ball is the number of balls in the cluster – we have counted the balls.



Mathematically speaking, we set up a bijection between a (finite) set of natural numbers $\{1, 2, ..., n\}$ ($n = 5$ in the example above) and the elements in the set of objects we wanted to count. We are going to formalize this now. But first some notations.

1.0.1 Notations

- $\mathbb{N}$ = set of all natural numbers
$$= \{1, 2, 3, ...\}$$
- $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$
$$= \{0, 1, 2, ...\}$$
- For $n \in \mathbb{N} : \mathbb{N}_n = \{m \in \mathbb{N} | m \leq n\}$
$$\{1, 2, ..., n-1, n\}$$

Definition 1.1. Let S be any set and $n \in \mathbb{N}$.

(a) We say that S has size or cardinality n if there exists a bijective map (function) $f : \mathbb{N}_n \to S$. In that case, we write $|S| = n$. We also say that the empty set $\emptyset$ has cardinality 0 and we write $|\emptyset| = 0$.

(b) The set S is called finite if $|S| \in \mathbb{N}_0$ and infinite otherwise. If S is infinite, we also write $|S| = \infty$.

Question 1.1. Suppose S is a finite non-empty set. Is $|S|$ a well-defined natural number? Or is it possible that $|S| = n$ and $|S| = m$ with $n, m \in \mathbb{N}$ and $n \neq m$?

Theorem 1.2. Assume that $n, m \in \mathbb{N}$ and $f : \mathbb{N}_n \to \mathbb{N}_m$ is a function which is injective. Then $n \leq m$.

Proof. Induction bases: If $n = 1$ then the claim is instantly true since $1 \leq m \quad \forall m \in \mathbb{N}$.
Induction step: Assuming the claim is true for some $n = k \in \mathbb{N}$ (induction hypothesis), we have to show that it is also true for $k+1$. So let an injective map $f : \mathbb{N}_{k+1} \to N_m$ be given. We want to prove $k+1 \leq m$ and distinguish two cases. Note first that $k+1 \geq 2$ implies $m \geq 2$ since f is injective $\to m-1 \in \mathbb{N}$.

Case 1: $f(x) \neq m \forall x \in \mathbb{N}_k$
Then $f(x) \in \mathbb{N}_m \setminus \{m\} = \mathbb{N}_{m-1} \forall x \in \mathbb{N}_k$, and we can define a new function $f_1 : \mathbb{N}_k \to N_{m-1}$ by $f_1(x) = f(x) \forall x \in \mathbb{N}_k$. Since f is injective, also f_1 is injective. Hence by the induction hypothesis $k \leq m-1 \to k+1 \leq m$.

Case 2: $\exists x_0 \in \mathbb{N}_k$ with $f(x_0) = m$
Then, since f is injective and $x_0 \neq k+1$, we must have

$$f(k+1) \neq f(x_0) = m \to f(k+1) \in \mathbb{N}_m \setminus \{m\} = \mathbb{N}_{m-1} \quad (*)$$

Set $y_0 = f(k+1)$ and define a new function

$$f_2 : \mathbb{N}_k \to \mathbb{N}_{m-1} \text{ by } f_2(x) := \begin{cases} f(x) & \forall x \in \mathbb{N}_k \setminus \{x_0\} \\ y_0 & \text{if } x = x_0 \end{cases}$$

We verify that f_2 is injective:

If $x_1, x_2 \in \mathbb{N}_k \setminus \{x_0\}$ with $x_1 \neq x_2$, then

$$f_2(x_1) = f(x_1) \neq f(x_2) = f_2(x_2) \to f_2(x_1) \neq f_2(x_2)$$

The first inequality above follows from f being injective.
And if $x_1 \in \mathbb{N}_k \setminus \{x_0\}$, then $x_1 \neq k+1$ and hence (since f is injective)

$$f(x_1) \neq f(k+1) = y_0 \to f_2(x_1) = f(x_1) \neq y_0 = f_2(x_0), \text{ i.e. } f_2(x_1) \neq f_2(x_0)$$

Since f_2 is injective, we can apply the induction hypothesis and get $k \leq m-1 \to k+1 \leq m$ also in this case.

□

$(*)$ And also $f(x) \in \mathbb{N}_m \setminus \{m\} = \mathbb{N}_{m-1} \quad \forall x \in \mathbb{N}_k \setminus \{x_0\}$.

Corollary 1.3. If S is a finite set and $|S| = m, |S| = n$ with $m, n \in \mathbb{N}_0$, then $m = n$.

Proof. If $S = \emptyset$, we must have $|S| = 0$ and so $m = n = 0$. (There is no map $f : \mathbb{N}_k \to \emptyset$ for any $k \in \mathbb{N}$.) So let's assume $S \neq \emptyset$. Since $|S| = m$, there exists a bijective map $f : \mathbb{N}_m \to S$. Since also $|S| = n$, there exists another bijective map $g : \mathbb{N}_n \to S$. Because f is bijective, there exists the inverse function $f^{-1} : S \to \mathbb{N}_m$, which is also bijective. So we can compute the map $k := f^{-1} \circ g : \mathbb{N}_n \to \mathbb{N}_m$. As a composition of two bijective maps, h is also bijective, so in particular injective. Hence *Theorem 1.2* implies $n \leq m$. But the inverse $h^{-1} : \mathbb{N}_m \to \mathbb{N}_n$ is also bijective. Another application of *Theorem 1.2* yields $m \leq n$. Hence $n \leq m \leq n \to m = n$. □

Corollary 1.4. If T and S are two finite sets and $f : T \to S$ is an injective function, then $|T| \leq |S|$.

Proof. Set $n = |T|$ and $m = |S|$. If $n = 0$, there is nothing to show (since $0 \leq |S|$ always). If $T \neq \emptyset$, then also $S \neq \emptyset$ since there is no map $f : T \to \emptyset$. So we may assume $T, S \neq \emptyset$ and $n, m \in \mathbb{N}$. By assumption, there exist bijective maps $g : \mathbb{N}_n \to T$ and $h : \mathbb{N}_m \to S$. Then also $h^{-1} : S \to \mathbb{N}_m$ is bijective, and $h^{-1} \circ f \circ g : \mathbb{N}_n \to \mathbb{N}_m$ is injective, since it is a composition of injective maps. therefore *Theorem 1.2* implies $n \leq m$ □

Proposition 1.5. If A and B are finite sets and $A \bigcap B = \emptyset$, then $|A \bigcup B| = |A| + |B|$.

Proof. We may assume $A \neq \emptyset$ and $B \neq \emptyset$ (For $A = \emptyset$ or $B = \emptyset$ the claim is trivial). Set $k = |A| \in \mathbb{N}$ and $l = |B| \in \mathbb{N}$. Let $f : \mathbb{N}\neg \to A$ and $g \in \mathbb{N}_l \to B$ be bijective functions.

$$\text{Define } h : N_{k+l} \to A \bigcup B \text{ by } h(i) := \begin{cases} f(i) & \text{if } i \leq k \\ g(i-k) & \text{if } k+1 \leq i \leq k+l \end{cases}$$

One immediately checks that $h(\mathbb{N}_{k+1}) = A \bigcup B$, which means that h is surjective. But h is also injective: For $1 \leq i < j \leq k + l$ we have to show that $h(i) \neq h(j)$. In order to do this, we distinguish three cases.
Case 1:

$$j \leq k \to h(i) = f(i) \neq f(j) = h(j) \to h(i) \neq h(j)$$

The first inequality above comes from the fact that f is injective.
Case 2:

$$k + 1 \leq i \to h(i) = g(i-k) \neq g(j-k) = h(j) \to h(i) \neq h(j)$$

The first inequality above comes from the fact that g is injective.
Case 3:

$$1 \leq i \leq k \text{ and } k+1 \leq j \leq k+l \to$$
$$h(i) = f(i) \in A \text{ and } h(j) = g(j-k) \in B \to$$
$$h(i) \neq h(j) \text{ since } A \bigcap B = \emptyset$$

Therefore, h is bijective, by definition of cardinality, it follows that $|A \bigcup B| = k + l = |A| + |B|$ □

Corollary 1.6. If S is a finite set, and T is a proper subset of S (notation: $T \subset S$), then $|T| < |S|$.

Proof. Since the embedding map $j : T \to S, \quad j(x) = x \quad \forall x \in T$ is injective, we already know from 1.4 that $|T| \leq |S|$. Since $T \subset S$, it follows that $\exists s_0 \in S$ with $s_0 \notin T$. Set $T' := T \bigcup \{s_0\}$. By 1.5, $|T'| = |T| + |\{s_0\}| = |T| + 1$. But the same argument as above yields (since $T' \subseteq S$), $|T'| \leq S$. Hence $|T| + 1 \leq |S| \to |S| - 1 < |S|$. □

Proposition 1.7. Let T and S be two finite non-empty sets and $f : T \to S$ a function. Assume that $|T| = |S|$.

(a) If f is injective, then it is also surjective.

(b) If f is surjective, then it is also injective.

Proof. (a) Set $n = |T| = |S| \in \mathbb{N}$. We prove the claim by way of contradiction. Assume that f is not surjective. then the image $S' := f(T) = \{f(x)|x \in T\}$ is different from S. Hence $S' \subseteq S$, and so by 1.6 $|S'| < |S| = n$. Now consider the function $f' : T \to S'$, $f'(x) = f(x) \quad \forall x \in T$. Then f' is injective since f is injective. So by 1.4, $n = |T| \leq |S^{-1}| < n$, which is clearly a contradiction. So we must have $f(T) = S$, and f is surjective.

□

Proof of (b): *Exercise!*

Lemma 1.8. If $f : T \to S$ is a surjective map between two finite non-empty sets, then $|T| \geq |S|$.

Proof. Let $n = |S|$ and $g : \mathbb{N}_n \to S$ be a bijection. Set $s_i = g(i) \quad \forall 1 \leq i \leq n$. Then $S = \{s_1, ..., s_n\}$ and $s_i \neq s_j \quad \forall i \neq j$. Since f is surjective, there exists for each $1 \leq i \leq n$, an element $t_i \in T$ with $f(t_i) = s_i$. Note that $t_i \neq t_j \quad \forall i \neq j$ since $s_i \neq s_j \quad \forall i \neq j$. Now set $T' := \{t_i|1 \leq i \leq n\} \subseteq T \to |T'| \leq |T|$. But $h : \mathbb{N}_n \to T'$ with $h(i) = t_i \quad \forall 1 \leq i \leq n$ is bijective, and so $|T'| = n$. It follows that $|T| \geq |T'| = n = |S|$. □

We will now reformulate Corollary 1.4 and get one of the basic principles of combinatorics.

Corollary 1.9. "Pigeonhole Principle"
If T and S are two finite non-empty sets with $|T| > |S|$ and $f : T \to S$ is any map, then there exist $x_1, x_2 \in T$ with $x_1 \neq x_2$ and $f(x_1) = f(x_2)$.

Proof. By 1.4 f cannot be injective since $|T| > |S|$. But this precisely means that there exist two distinct elements x_1, x_2 of T which satisfy $f(x_1) = f(x_2)$. □

To formulate this principle in a less abstract way: if we have n letters which are to be put into m pigeonholes and $n > m$, then there is (at least) one pigeonhole which will receive ≥ 2 letters (T = set of letters, S = set of pigeonholes, f = map which assigns each letter to a specific pigeonhole, $x_1 \neq x_2 \hat{=}$ two distinct letters which end up in the same pigeonhole ($f(x_1) = f(x_2)$)).

Example 1.10. In a group of 13 (or more) people, there will always be 2 whose birthday is in the same month.

Here we apply 1.9 with T = set of people in the given group $\to |T| = 13$(or ≥ 13).
S = set of months of the year $\to |S| = 12$

$f : T \to S$ is defined by:
$f(x)$ = month in which person x has his/her birthday. So by 1.9, there are two different persons $x_1, x_2 \in T$ with $f(x_1) = f(x_2)$, i.e. x_1, x_2 have their birthday in the same month.

Example 1.11. In a set X of people with $|X| \geq 2$ there are two who have the same number of friends. Two clarifications: 1) A person is not considered a friend of him/herself. 2) Friendship is mutual: IF A is a friend of B, then also B is a friend of A.

We want to apply the pigeonhole principle.
Set $n = |X| \geq 2$ and $S = \mathbb{N}_{n-1} \bigcup \{0\} = \{0, 1, ..., n-1\}$. We define $f : X \to S$ by $f(x) := |\{y \in X | y \text{ is a friend of } x\}|$ since $|X| = |S| = n$, we cannot apply the pigeonhole principle immediately. However, we can make the following Observation:

- 0 and $n-1$ cannot both be elements of the image $f(x) = \{f(x) | x \in X\}$. This is because if there exists $x_0 \in X$ with $f(x_0) = n-1$, then x_0 is a friend of everybody in $X \setminus \{x_0\}$, and hence $f(x) \geq 1 \quad \forall x \in X$. Using this observation, there are two ways to finish the proof of our claim.

Method 1: By way of contradiction.
Assume that $f(x_1) \neq f(x_2)$ for all $x_1, x_2 \in X$ with $x_1 \neq x_2$. Then f is injective. Since $|X| = |S|$, 1.7 (a) now implies that f is also surjective. Hence $0 \in f(X)$ and $n-1 \in f(X)$ which contradicts our observation.
Method 2:
If $n-1 \notin f(X)$, we set $S_1 := \mathbb{N}_{n-1} \bigcup \{0\} = \{0, 1, 2, ..., n-2\}$ and define $f_1 : X \to S_1$ by $f_1(x) = f(x) \quad \forall x \in X$. Now we have $n = |X| > |S_1| = n-1$, and so by 1.9 there exist $x_1, x_2 \in X$ with $f_1(x) = f_1(x_2) \to f(x_1) = f(x_2)$. If $0 \notin f(X)$, we similarly set $S_2 = \mathbb{N}_{n-1} = \{1, ..., n-1\}$ and define $f_2 : X \to S_2$ by $f_2(x) = f(x) \quad \forall x \in X$. Then the pigeonhole principle yields the existence of $x_1, x_2 \in X$ with $x_1 \neq x_2$ and $f_2(x_1) = f_2(x_2) \to f(x_1) = f(x_2)$.

Remark 1.12. A similar observation as in 1.11 is also crucial in the example discussed in Section 6.3 of [**B**]. Namely, April cannot be the person who shook hands with 8 other people (why?) However, as opposed to 1.11, this observation does not immediately solve the problem.

Challenge: Read the text concerning this example on page 49 but do not turn the page in order to continue reading on page 50. Then try to figure out the solution for this problem. After coming up with an answer, you can then compare it with the one given on page 50 in [**B**]. Do you find it convincing?

Example 1.13. Given is a set $T \subseteq \mathbb{N}$ with $|T| = 12$. Show that there exist $x, y \in T$ with $x \neq y$ and $11 | x - y$ (i.e. 11 divides $x - y \leftrightarrow \exists q \in \mathbb{Z} : x - y = 11 \cdot q$).

The essential tool for solving this problem are remainders, see section 8.2 of [**B**]. More precisely, we will use the following Fact:
For any $a \in \mathbb{N}$, there exist uniquely determined $q, r \in \mathbb{N}_0$ with $a = q \cdot 11 + r$ and

$0 \leq r \leq 10$. (so r is the remainder after dividing a by 11).
Now set $S = \{r \in \mathbb{N}_0 | r \leq 10\}$ and define $f : T \to S$ by $f(a) = r$ if $\exists q \in \mathbb{Z} : a = 11q + r$ with $r \in S$, for any $a \in T$. Since $12 = |T| > 11 = |S|$, there exist $x, y \in T$ with $x \neq y$ and $f(x) = f(y) \to \exists q_1, q_2 \in \mathbb{Z}$ with $x = 11q_1 + r, y = 11q_2 + r$, where $r = f(x) = f(y) \in S \to$

$$x - y = 11(q_1 - q_2) \text{ with } q_1 - q_2 \in \mathbb{Z} \to 11|x - y$$

Example 1.14. We want to prove the following: Claim: For any $X \subseteq \mathbb{N}_{2n} (n \in \mathbb{N})$ with $|X| > n$, there exist $x_1, x_2 \in X$ with $x_1 \neq x_2$ and $x_1 | x_2$.

Proof. Set $Y = \{2m - 1 | m \in \mathbb{N}, 1 \leq m \leq n\} = \{1, 3, 5, ..., 2n - 1\}$ each $x \in X$ can be written in the form $x = 2^\alpha y$ with $a \in \mathbb{N}_0$ and $y \in Y$. (Use the prime factorization of x and pull out the highest power of 2 which divides x) Since y is odd, there is precisely one way to write x in the form $x = 2^\alpha y$ with $y \in Y$. Hence we can define a function $f : X \to Y$ by $f(x) = y$ if $x = 2^\alpha y$ and $y \in Y$. By the pigeonhole principle, there exist $x_1, x_2 \in X$ with $x_1 \neq x_2$ and $f(x_1) = f(x_2)$ because $|X| > n = |Y|$. If we set $y = f(x_1) = f(x_2) \in Y$, then $x_i = 2^{\alpha_i} y$ with $a_i \in \mathbb{N}_0$ for $i = 1, 2$. Since $x_1 \neq x_2$, we must have $a_1 \neq a_2$. Without loss of generality, we can assume $a_1 < a_2$ (otherwise we must renumber x_1 and x_2). Then $x_2 = 2^{\alpha_2 - \alpha_1} x_1$ with $2^{\alpha_2 - \alpha_1} \in \mathbb{N} \to x_1 | x_2$. □

Remark: There exists a subset $X \subseteq \mathbb{N}_{2n}$ with $|X| = n$ and $x_1 \nmid x_2 \quad \forall x_1, x_2 \in X$ with $x_1 \neq x_2$, e.g.

$$X = \{n, n + 1, ..., 2n - 1\} \quad (\text{or } X = \{n + 1, n + 2, ..., 2n\}$$

I will finish this chapter with some remarks on infinite sets. First an obvious one.

Proposition 1.15. The set $\mathbb{N}$ is infinite.

Proof. Obviously $\mathbb{N} \neq \emptyset$ and so $|\mathbb{N}| \neq 0$. Now assume by way of contradiction, that $|\mathbb{N}| = \mathbb{N}_n$ for some $n \in \mathbb{N}$. Then there exists a bijective map $f : \mathbb{N}_n \to N$, and hence also a bijective map $f^{-1} : \mathbb{N} \to N_n$. Now $j : \mathbb{N}_{n+1} \to N$ (the initial embedding of $\mathbb{N}_{n+1}$ onto $\mathbb{N}$) is injective, and hence also $f^{-1} \circ j : \mathbb{N}_{n+1} \to \mathbb{N}_n$ is injective. But this contradicts Theorem 1.2. Hence $|\mathbb{N}| \notin \mathbb{N}_0$, and so $\mathbb{N}$ is infinite. □

Corollary 1.16. Every set which contains $\mathbb{N}$ (e.g. $\mathbb{N}_\nvdash, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}, ...$) is infinite

Proof. This follows immediately from 1.15 since every subset of a finite set is finite. □

Infinite sets also have cardinalities but they are not numbers in the usual sense.

Definition 1.17. (a) We say that two sets A and B have the same cardinality, and we then write $|A| = |B|$, if there exists a bijective map $f : A \to B$.

(b) We write $|A| \leq |B|$ if there exists an injective map $f : A \to B$ and $|A| < |B|$ if $|A| \leq B$ and $|A| \neq |B|$.

(c) A set X is called countable if X is finite or $|X| = |\mathbb{N}|$ and uncountable if $|X| > \mathbb{N}$.

I will mention without proof a few interesting facts about infinite sets.

Facts 1.18. (a) If A is countably infinite, and $B \subseteq A$ is any infinite subset, then $|B| = |A|(= \mathbb{N})$.

(b) If A is uncountable and $B \subseteq A$ is countable, then $|A \setminus B| = |A|$.

(c) $\mathbb{Q}$ is countable and $\mathbb{R}$ is uncountable.

(d) For any set $X, |P(X)| > |X|$, where $P(X)$ is the power set $P(X) := \{Y | Y \subseteq X\}$.

(e) The Schröder–Bernstein theorem If two sets A and B satisfy $|A| \leq |B|$ and $|B| \leq |A|$, then $|A| = |B|$.

(f) A countable union of countable sets is countable.

2 Principles of Counting

The addition principle refers to the following:

Proposition 2.1. Assume that $A_1, ..., A_n$ are n finite sets ($n \in \mathbb{N}$) which are pairwise disjoint, i.e. $A_i \bigcap A_j = \emptyset \quad \forall i \neq j$. Then

$$|\bigcup_{i=1}^{n} A_i| = \sum_{i=1}^{n} |A_i|$$

Proof. The claim is proved most conveniently by induction on n.
$n = 1$ (induction basis): $|A_1| = \sum_{i=1}^{n} |A_i|$ is trivially true.
This proof is an example of an induction proof where the case $n = 2$ needs to be proved separately since it is used in the induction step.
$n = 2$: If A_1, A_2 are finite sets with $A_1 \bigcap A_2 = \emptyset$, then $|A_1 \bigcup A_2| = |A_1| + |A_2|$. This was proved in Proposition 1.5
Induction Step: Assume the claim is true for $n = k$ with $k \geq 2$. We want to show that is then also true for $n = k + 1$. So let $k + 1$ pairwise disjoint finite sets $A_1, ..., A_k, A_{k+1}$ be given. We have to show $|\bigcup_{i=1}^{k+1} A_i| = \sum_{i=1}^{k+1} |A_i|$. We want to apply the case $n = 2$, or rather Proposition 1.5, with $A = \bigcup_{i=1}^{k} A_i$ and $B = A_{k+1}$. But in order to do this, we need to verify the assumption $A \bigcap B = \emptyset$. Here we are using Exercise (1)(a) of Homework # 1 (and $A_i \bigcap A_{k+1} = \emptyset \quad \forall 1 \leq i \leq k$):

$$A \bigcap B = (\bigcup_{i=1}^{k} A_i) \bigcap A_{k+1} = \bigcup_{i=1}^{k} (A_i \bigcap A_{k+1}) = \bigcup_{i=1}^{k} \emptyset = \emptyset.$$

Hence Proposition 1.5 implies $|A \bigcup B| = |A| + |B|$. Now we are also using the Inductive Hypothesis which yields

$$|A| = |\bigcup_{i=1}^{k} A_i| = \sum_{i=1}^{k} |A_i|.$$

Putting things together, we get

$$|\bigcup_{i=1}^{k+1} A_i| = |\bigcup_{i=1}^{k} A_i \bigcup A_{k+1}| = |A \bigcup B| = |A| + |B| = \sum_{i=1}^{k} |A_i| + |A_{k+1}| = \sum_{i=1}^{k+1} |A_i|.$$

□

Corollary 2.2. If A and B are two finite sets, then $|A \bigcup B| = |A| + |B| - |A \bigcap B|$.

Proof. It is easy to verify that $A \bigcup B$ is the disjoint union of $A \setminus (A \bigcap B), B \setminus (A \bigcap B)$ and $A \bigcap B$ (left as exercise). Applying 2.1 with $n = 3$, we get

$$(*)|A \bigcup B| = |A \setminus (A \bigcap B) + B \setminus (A \bigcap B)| + |A \bigcap B|$$

Now A is obviously the disjoint union of $A \setminus (A \bigcap B)$ and $A \bigcap B$. Hence by 1.5, $|A| = |A \setminus (A \bigcap B)| + |A \bigcap B| \to |A \setminus (A \bigcap B)| = |A| - |A \bigcap B|$. Similarly, we get $|B \setminus (A \bigcap B)| = |B| - |A \bigcap B|$. Plugging this into $(*)$ above, we obtain

$$|A \bigcup B| = (|A| - |A \bigcap B|) + (|B| - |A \bigcap B|) + |A \bigcap B|$$

$$= |A| + |B| - |A \bigcap B|.$$

□

Corollary 2.3. ("Generalized Pigeonhole Principle")
If $f : X \to Y$ is a map between two finite non-empty sets and $|X| > r|Y|$ for some $r \in \mathbb{N}$, then there exists $y \in Y$ with $|f^{-1}(y)| \geq r + 1$.

Proof. Set $|X| = m, |Y| = n \in \mathbb{N}$. Let $y_1, ..., y_n$ be the n distinct elements of Y, i.e., $Y = \{y_1, ..., y_n\}$ and set $X_i := f^{-1}(y_i)$ for $1 \leq i \leq n$. Then $X = \bigcup_{i=1}^{n} X_i \quad (x \in X \to f(x) = y_i$ for some $1 \leq i \leq n \to x \in X_i)$ and $x_i \bigcap x_j = \emptyset \quad \forall i \neq j$ (it's impossible for any $x \in X$ to satisfy $f(x) = y_i$ and $f(x) = y_j$ since $y_i \neq y_j \quad \forall i \neq j$).
So we can apply the Proposition 2.1 and get:

$$m = |X| = |\bigcup_{i=1}^{n} X_i| = \sum_{i=1}^{n} |X_i|$$

Now assume, by way of contradiction, that $|x_i| \leq r \quad \forall 1 \leq i \leq n$. Then it followed that $m \leq \sum_{i=1}^{n} r = n \cdot r$, i.e. $|X| \leq r|Y|$. However, we are given that $|X| > r|Y|$, yielding a contradiction. Hence there exists (at least one) $y_i \in Y \quad (1 \leq i \leq n)$ with $|x_i| \geq r + 1$, i.e. $|f^{-1}(y_i)| \geq r + 1$. □

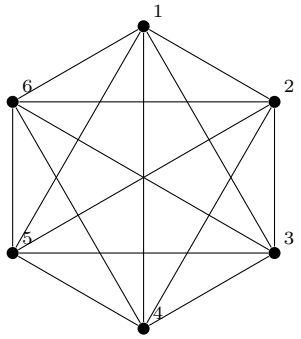
The pigeonhole formulation of 2.3 would go as follows: If m letters have to be distributed into n pigeonholes, and $m > r \cdot n$ for some $r \in \mathbb{N}$, then there exists a pigeonhole which receives at least $r + 1$ letters.

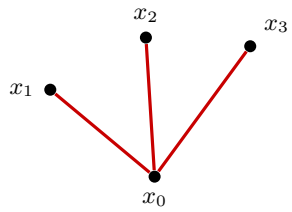
Example 2.4. In a group X of 100 people, there will always be at least 9 whose birthday is in the same month.
Set $Y = \{$ months of the year $\}$, so $|Y| = 12$, and consider $f : X \to Y$ with $f(x) =$ month in which person x has his/her birthday. Since $|X| = 100 > 96 = 8 \cdot 12 = 8 \cdot |Y|$, there exists a month y with $|f^{-1}(y)| \geq 8 + 1 = 9$, i.e. there are ≥ 9 people who have their birthday in month y.
A more sophisticated example is the following: (Here the application of 2.3 itself is trivial but it is not easy to see how the generalized pigeonhole principle can be applied at all.)

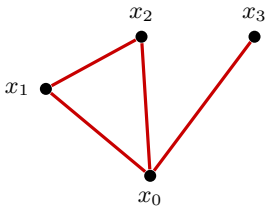
Example 2.5. Let X be any group of 6 people in which any two of them are either mutual friends or mutual strangers. Claim: there exists a subgroup of 3 people who are all mutual friends or are all mutual strangers.

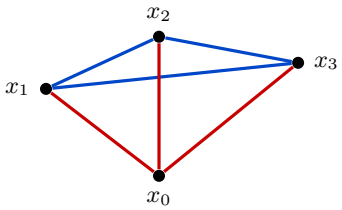
Graph Theoretic Representation: Assume that in the complete graph K_6, with 6 vertices and 15 edges all edges are colored either red or blue. Then there will be a red or blue triangle.

K_6 :









Poss. I

Poss. II

Proof. For $x_0 \in X$, set $X^1 := X \setminus \{x_0\}$ and define $f : X^1 \to Y := \{F, S\}$ by $f(x) = \begin{cases} F & \text{if is a friend of } x_0 \\ S & \text{if} x \text{ and } x_0 \text{ are strangers} \end{cases}$ Since $5 = |X'| > 4 = 2|Y|$, 2.3 tells us that there exists $y \in Y$ with $|f^{-1}(y)| \geq 2 + 1 = 3$. This means that there exist 3 distinct $x_1, x_2, x_3 \in X$ such that either x_0 and x_i are friends $\forall 1 \leq i \leq 3$ or x_0 and x_i are strangers $\forall 1 \leq j \leq 3$. This leads to two possibilities.
Case 1: $\exists 1 \leq i < j \leq 3$ such that x_i and x_j are friends. Then $\{x_0, x_i, x_j\}$ is a set of three mutual friends (red triangle).
Case 2: x_i and x_j are strangers $\forall 1 \leq i < j \leq 3$. Then $\{x_1, x_2, x_3\}$ is a set of 3 mutual strangers (blue triangle). □

We are now turning to the Multiplication Principle Let's first recall that the definition of the Cartesian product of two sets A and B :

$$A \times B := \{(a, b) | a \in A, b \in B\}$$

Here the (a, b) are ordered pairs, i.e, by definition $(a_1, b_1) = (a_2, b_2) \leftrightarrow a_1 = a_2$, and $b_1 = b_2 \quad (a_1, a_2 \in A, b_1, b_2 \in B)$. So if $a \neq b$, then $(a, b) \neq (b, a)$, whereas $\{a, b\} = \{b, a\}$. Note also that $\emptyset \times B = \emptyset$ (there are no pairs (a, b) with $a \in \emptyset$ and $A \times \emptyset = \emptyset$.

Proposition 2.6. If A and B are two finite sets, then $|A \times B| = |A| \times |B|$

Proof. This is clear if $A = \emptyset$ or $B = \emptyset$. So assume $m = |A| \in \mathbb{N}$ and $n = |B| \in \mathbb{N}$. Let $b_1, ..., b_n$ be the n distinct elements of B so that $B = \{b_1, ..., b_n\}$ the n distinct elements of B so that $B = \{b_1, ..., b_n\}$. Now set

$$A_i := A \times \{b_i\} = \{(a, b_i) | a \in A\} \subseteq A \times B \text{ for } 1 \leq i \leq n$$

We observe

1. $|A_i| = |A| = m \quad \forall 1 \leq i \leq n$ since $f_i : A \to A_i$ is bijective $(a \to (a, b_i))$.
2. $A_i \bigcap A_j = \emptyset \quad \forall i \neq j$ since $(a_1, b_i) \neq (a_2, b_j) \quad \forall a_1, a_2 \in A$ by definition of ordered pairs.
3. $A \times B = \bigcup_{i=1}^{n} A_i :\supseteq$ Follows from $A_i \subseteq A \times B \quad \forall 1 \leq i \leq n$ and $\subseteq$ is obtained as follows: Given any $x = (a, b) \in A \times B$, $b \in B$ implies $b = b_i$ for some $1 \leq i \leq n \to x = (a, b_{i_0}) \in A_{i_0} \subseteq \bigcup_{i=1}^{n} A_i$.

Now we can apply Proposition 2.1 and get

$$|A \times B| = |\bigcup_{i=1}^{n} A_i| = \sum_{i=1}^{n} |A_i| = \sum_{i=1}^{n} m = m \cdot n = |A| \cdot |B|$$

□

As in the step from Proposition 1.5 to Proposition 2.1, we can now also generalize 2.6 as follows:

Proposition 2.7. If $A_1, ..., A_n$ are n finite sets, then

$$|A_1 \times \cdots \times A_n| = \Pi_{i=1}^{n} |A_i|$$

Where $A_1 \times \cdots \times A_n$ is the set of ordered $n-$tuples $A_1 \times \cdots \times A_n = \{(a_1, ..., a_n) | a_i \in A_i \forall 1 \leq i \leq n\}$.

Proof. Exercise! □

Example 2.8. A typical license plate in Virginia contains a license "number" (code) which consists of a sequence of 3 letters followed by a sequence of 4 digits, each between 0 and 9. How many different license numbers of this type are there?
Set A = set of letters in our alphabet $= \{A, B, ..., Z\} \to |A| = 26$ $B = \{n \in \mathbb{N}, n \leq 9\} = \{0, 1, ..., 9\} \to |B| = 10$.

$$A_i = A \text{ for} 1 \leq i \leq 3, A_i = B \text{ for} 4 \leq i \leq 7 \to$$

possible license numbers $= |A_1 \times A_2 \times \cdots \times A_7| = \Pi_{i=1}^{7} |A_i|$ by proposition 2.7.

$$26^3 \cdot 10^4 = 175,760,000.$$

We are now going to discuss another important counting technique in connection with Cartesian products $A \times B$. Namely there are two ways to determine $|S|$ for any $S \subseteq A \times B$. Let's assume $|A| = m, A = \{a_1, ..., a_m\}, |B| = n, B = \{b_1, ..., b_n\} \quad (m, n \in \mathbb{N})$ We can associate an $m \times n$ matrix m with S, $M = (m_{i,j}) \quad 1 \leq i \leq m, 1 \leq j \leq n$. Where

$$m_{ij} = \begin{cases} 1 & \text{if } (a_i, b_j) \in S \\ 0 & \text{if } (a_i, b_j) \notin S \end{cases}$$

It is clear that $|S| = \sum_{i=1}^{m} \sum_{j=1}^{n} m_{ij}$. Now the double sum can be computed in two ways. For any $1 \leq i \leq m$, we can define the row sum:

$$r_i(S) = r_{a_i}(S) = \sum_{j=1}^{n} m_{ij} = |\{b \in B | (a_i, b) \in S\}|$$

and for any $1 \leq j \leq n$, we can define the column sum

$$c_j(S) = c_{b_j}(S) = \sum_{i=1}^{m} m_{ij} = |\{a \in A | (a, b)j) \in S\}|$$

. Then we get:

Proposition 2.9. $|S| = \sum_{i=1}^{m} r_i(S) = \sum_{j=1}^{n} c_j(S)$

Proof. $\sum_{i=1}^{m} r_i(S) = \sum_{i=1}^{m} \sum_{j=1}^{n} m_{ij} = |S|$ and also

$$\sum_{j=1}^{n} c_j(S) = \sum_{j=1}^{n} \sum_{i=1}^{m} m_{ij} = \sum_{i=1}^{m} \sum_{j=1}^{n} m_{ij} = |S|$$

□

Remark: With the notation of the proof of 2.6, we have $c_j(S) = |S \bigcap A_j|$. Since S is the disjoint union $S = \bigcup_{j=1}^{n} (S \bigcap A_j)$, we also obtain:

$$|S| = \sum_{j=1}^{n} |S \bigcap A_j| = \sum_{j=1}^{n} c_j(S)$$

from 2.1.

Example 2.10. In a class there are $m = 32$ boys and an unknown number n of girls. If each boy knows precisely 5 of the girls in class and each girl knows precisely 8 boys in class, what is n?
Set $A = \{\text{boys in the class}\}$, $B = \{\text{girls in the class}\}$ and $S = \{(a, b) \in A \times B | a \text{ knows } b\}$. We are given $|A| = 32$, $r_i(S) = |\{b \in B | \text{boy } a_i \text{ knows } b\}| = 5 \quad \forall 1 \leq i \leq 32$

$$c_j(S) = |\{a \in A | \text{girl } b_j \text{ knows } a\}| = 8 \quad \forall 1 \leq j \leq n$$

. Hence we can compute:

$$|S| = \sum_{i=1}^{m} r_i(S) = \sum_{i=1}^{32} 5 = 32 \cdot 5 = 160$$

and

$$|S| = \sum_{j=1}^{n} c_j(S) = \sum_{j=1}^{n} 8 = 8n \to 8n = 160 \to n = 20.$$

We are now turning to a statement which could be called the power principle. Notation: if $n \in \mathbb{N}$ and A is any set, then

$$A^n := \underbrace{A \times \cdots \times A}_{n \text{ times}} = \{(a_1, ..., a_n) | \forall a_i \in A \quad \forall 1 \leq i \leq n\}$$

So the set of all possible license plates in 2.8 can also be written in the form of $A^3 \times B^4$.

Corollary 2.11. $|A^n| = |A|^n \quad \forall n \in \mathbb{N}$, all finite sets$A$

Proof. This follows immediately from 2.7 by setting $A_1 = A_2 = \cdots = A_n = A$. □

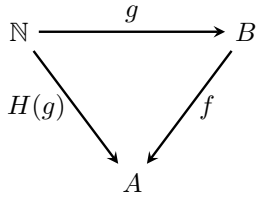
Example: $|\{0,1\}^n| = 2^n$
Another interpretation of A^n each ordered $n-$tuple $(a_1, ..., a_n) \in A^n$ can be identified with the function $f : \mathbb{N}_n \to A$ satisfying $f(i) = a_i \quad \forall 1 \leq i \leq n$. This idea will now be generalized.

Definition 2.12. For any two non-empty sets A and B, we define $A^B := \{f | f : B \to A \text{ is a map from } B \text{ to } A\}$.

Example: $A^{\mathbb{N}_n = A^n}$

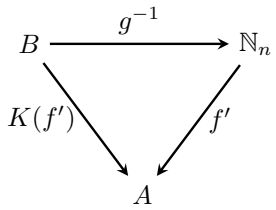
Proposition 2.13. For any two finite non-empty sets A and B, we get $|A^B| = |A|^{|B|}$.

Proof. Set $m = |A|$ and $|B| = n$ with $m, n \in \mathbb{N}$. We have to show $|A^B| = m^n$. Since $|B| = n$, there exists a bijective map $g : \mathbb{N}_n \to B$. We will use g in-order to define a bijective map $H : A^B \to A^{\mathbb{N}_n}$. Namely, for any $f \in A^B$, which is a map $f : B \to A$, we define $H(f) := f \circ g : \mathbb{N}_n \to A$, which means that $H(f) \in A^{\mathbb{N}_n}$



In order to show that H is bijective, we define a map

$$K : A^{\mathbb{N}_n} \to A^B, \qquad K(f') := f' \circ g^{-1} \quad \forall f' \in A^{\mathbb{N}_n}.$$



Now we verify

$$(K \circ H)(f) = K(H(f)) = K(f \circ g) = (f \circ g) \circ g^{-1} = f \circ (g \circ g^{-1}) = f \circ id_B = f$$

for all $f \in A^B$, i.e. $K \circ H = id_A B$, and

$$(H \circ K)(f') = H(K(f')) = H(f' \circ g^{-1}) = (f' \circ g^{-1}) \circ g = f' \circ id_{\mathbb{N}_n} = f'$$

For all $f' \in A^{\mathbb{N}_n}$, i.e., $H \circ K = id_A \mathbb{N}_n$. This shows that H is invertible, with $H^{-1} = K$, implying that H is bijective. It follows that $|A^B| = |A^{\mathbb{N}_n}|$. Now 2.11 implies

$$|A^B| = |A^{\mathbb{N}_n}| = |A^n| = |A|^n = m^n = |A|^{|B|}$$

□

Corollary 2.14. For any finite non-empty set X we get

$$|\{0,1\}^X| = 2^{|X|}$$

Next we want to set up a bijection between $\{0,1\}^X$ and the powerset $P(X) := \{A | A \subseteq X\}$. For this we need the following.

Definition 2.15. If X is any set (possibly infinite) and $A \subseteq A$, we associate the characteristic function $\chi_A : X \to \{0,1\}$ to A, which is defined by:

$$\chi_A = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \in X \setminus A \end{cases}$$

Special Cases: $\chi_A \equiv 1$ on X and $\chi_\emptyset \equiv 0$ on X.

Proposition 2.16. For any (possibly infinite) set X, the map

$$\chi : P(X) \to \{0,1\}^X \text{is bijective, and hence} |P(X)| = |\{0,1\}^X|$$

Proof. (a) χ is injective. Assume we have subsets $A, B \subseteq X$ which satisfy $\chi_A = \chi_B$. Then it follows for all $x \in X$:

$$x \in A \leftrightarrow \chi_A(x) = 1 \leftrightarrow \chi_B(x) = 1 \leftrightarrow x \in B$$

So A and B have the same elements and $A = B$ follows.
(b) χ is surjective.
Let $f \in \{0,1\}^X$ be given, i.e. $f : X \to \{0,1\}$ is a function. We want to show that f is in the image of χ. Set

$$A = \{x \in X | f(x) = 1\} \subseteq X$$

Then

$$f(x) = 1 = \chi_A(x) \quad \forall x \in A (\text{since } f(x) = 1 \text{ implies } x \in A)$$

and

$$f(x) = 0 = \chi_A(x) \quad \forall x \in X \setminus A (\text{since } f(x) = 0 \text{ implies } x \notin A)$$

So f and χ_A are both functions from X to $\{0,1\}$, and they satisfy $f(x) = \chi_A(x) \quad \forall x \in X$. Hence $f = \chi_A = \chi(A)$. □

Corollary 2.17. For any finite set X, $|P(X)| = 2^{|X|}$.

Proof. For $X = \emptyset$, we get

$$P(\emptyset)| = |\{\emptyset\} = 1 = 2^0 = 2^{|\emptyset|}$$

For finite $X \neq \emptyset$, 2.16 and 2.14 imply,

$$|P(X)| = |\{0,1\}^X| = 2^{|X}$$

□

Side-Remark Corollary 2.17 (*) is the reason why in some mathematical papers and books the power-set of X is denoted by 2^X (*) respectively, Proposition 2.16

Theorem 2.18. (G. Cantor) If X is any set, then there exists no surjective function $f : X \to P(X)$

Proof. If a surjective $f : X \to P(X)$ existed, then also $x \circ f : X \to \{0,1\}^x$ were surjective, since X is bijective by 2.16. So it suffices to show that there is no surjective map $G :\to \{0,1\}^X$. Given such a map G we construct $f : X \to \{0,1\}$ as follows

$$\text{For any } x \in X \text{ set:} f(x) = \begin{cases} 1 & \text{if } G(x)(x) = 0 \\ 0 & \text{if } G(x)(x) = 1 \end{cases}$$

Keep in mind that $G(x) \in \{0,1\}^X$ is a function from X to $\{0,1\}$. So $G(x)$ can be evaluated at x, and $G(x)(x) = 0$ or 1. But now $G(x) \neq f \quad \forall x \in X$ since (by definition of f) $G(x)(x) \neq f(x)$. Note that the functions $G(x)$ and f are distinct if there exists one element of the domain X (here the element is x) where the two functions differ.

$$f \neq G(x) \quad \forall x \in X \leftrightarrow f \notin G(X) \to G(X) \neq \{0,1\}^X$$

Hence $G : X \to \{0,1\}^X$ is not surjective, and no surjective function $F : X \to P(X)$ can exist. □

Corollary 2.19. For any (finite or infinite) set X, $|X| < |P(X)|$. In particular $P(\mathbb{N})$ is uncountable.

Proof. Since $f : X \to P(X), f(x) = \{x\} \quad \forall x \in X$ is obviously an injective map, we have $|X| \leq |P(X)|$. However, by Theorem 2.18, there cannot exist a surjective and in particular not a bijective map from X to $P(X)$. Hence $|X| \neq |P(X)|$, and $|X| < |P(X)|$ follows (see Def. 1.17). For $X = \mathbb{N}$, we obtain $|\mathbb{N}| < |P(\mathbb{N})|$, and so $P(\mathbb{N})$ is an uncountable set. □

Remark: It can be shown that $|\mathbb{R}| = |P(\mathbb{N})|$. Whether or not there exists a set S with $|\mathbb{N}| < |S| < |P(\mathbb{N})|$ is a statement which is independent of the standard axioms of set theory (see "continuum hypothesis"). Let's now return to A^n, where A is a finite non-empty set with $m = |A| \in \mathbb{N}$, namely as the set of

all functions from $\mathbb{N}_n$ to A, $A^{\mathbb{N}_n}$. Yet another interpretation is the following. A^n is the set of all ordered selections with no repetition of n elements from A. So $(a_1, ..., a_n) \in A^n$ is interpreted as the ordered selection of the elements $a_1, ..., a_n$ from A (it matters in which order these elements are chosen), where repetitions are allowed (i.e. we may have $a_i = a_j$ for some $i \neq j$). With this interpretation 2.11 becomes:

Corollary 2.20. If $|A| = m \in \mathbb{N}$, then there are precisely m^n ordered selections with repetition of n elements from A.

Question: What about ordered selections without repetitions?
An ordered selection without repetitions of n elements from A is an element $(a_1, ..., a_n) \in A^n$ with $a_i \neq a_j \forall i \neq j$. Equivalently, we have:

$$\{\text{ordered selections without repetitions of } n \text{ elements from } A\} = \{f \in A^{\mathbb{N}_n} | f \text{ is injective}\}$$

How many elements are there in this set?

Proposition 2.21. If $|A| = m \in \mathbb{N}$, then:

$$|\{f \in A^{\mathbb{N}_n} | f \text{ is injective}\}| = \begin{cases} \Pi_{l=m-n+1}^{m} l & \text{if } n \leq m \\ 0 \quad \text{if } n > m \end{cases}$$

Proof. First note that there are no injective maps $f : \mathbb{N}_n \to A$ if $n > m = |A|$ (pigeonhole principle; see 1.2 and 1.4 or 1.9).
For $1 \leq n \leq m$, the standard (slightly imprecise) argument goes as follows: for any injective $f \in A^{\mathbb{N}_n}$,

$$f(1) \in A \to |A| = m \text{ possibilities for } f(1)$$

$$f(2) \in A \setminus \{f(1)\} (\text{since } f(2) \neq f(1) \to m - 1 \text{poss. for } f(2)$$

$$\vdots$$

$$f(n) \in A \setminus \{f(1), ..., f(n-1)\} \to m - (n+1) \text{ poss. for} f(n)$$

Hence there are $m(m-1)\cdots(m-n+1)$ possible choices for an injective function $f : \mathbb{N}_n \to A$. A rigorous proof for this claim would use induction on n (or m).

For $n = 1$, any function from $N_1 = \{1\}$ to A is injective so that $\{f \in A^{\mathbb{N}|f \text{ is injective}} = |A^1| = |A| = m\}$.
For $n \geq 2$, we first list the m elements of A in some form and write $A = \{a_1, ..., a_m\}$. We then define

$$S := \{(x_1, ..., x_n) \in A^n | x_i \neq x_j \forall 1 \leq i \neq j \leq n\}$$

and

$$S_k := \{(x_1, ..., x_n) \in S | x_n = a_k\} \text{ for } 1 \leq k \leq m$$

Note that S is the disjoint union $S = \bigcup_{k=1}^{m} S_k$ and hence $|S| = \sum_{k=1}^{m} |S_k|$ by the addition principle 2.1.
Now for each k, the map $g_k : S_k \to S'_k$ with

$$S'_k = \{(y_1, ..., y_{n-1}) | y_i \in A \setminus \{a_k\} \forall i, y_i \neq y_j \forall 1 \leq i \neq j \leq n-1\}$$

and

$$g_k((x_1, ..., x_n)) = (x_1, ..., x_{n-1}) \text{ is bijective (verify this!)}$$

So that $|S_k| = |S'_k| \quad \forall 1 \leq k \leq m$. But

$$|S'_k| = (m-1)(m-2)\cdots((m-1)-(n-1)+1)$$

$$= (m-1)(m-2)\cdots(m-n+1)$$

By the induction hypothesis since $|A \setminus \{a_k\}| = m-1 \quad \forall 1 \leq k \leq m$. It follows that

$$|\{f \in A^n | f \text{ is injective}\}| = |S| = |\bigcup_{k=1}^{m} S_k| = \sum_{k=1}^{m} |S_k|$$

$$= \sum_{k=1}^{m} (m-1)\cdots(m-n+1) = m(m-1)\cdots(m-n+1)$$

as claimed. □

Definition 2.22. For $m, n \in \mathbb{N}$ with $n \leq m$, we set

$$(m)_n = m(m-1)\cdots(m-n+1) = \Pi_{l=m-n+1}^{m} l$$

and

$$m! = (m)_m = 1 \cdot 2 \cdot 3 \cdots m$$

$m!$ is called "m factorial". It is also common to define $0! := 1$.

Example 2.23. A coach has to put together a relay team consisting of 4 runners. He can choose among a group of 10 men. How many possible relay teams are there assuming that the order in which the men are running matters?
Answer: The task for the coach is to make an ordered selection without repetitions of 4 runners from a group of 10 men. According to Proposition 2.21, there are a total of $(10)_4 = 10 \cdot 9 \cdot 8 \cdot 7 = 5040$ of such selections possible.
An import special case occurs if $m = n$ in 2.21. Recall that a function $f : \mathbb{N}_n \to |A|$, where $|A| = n$, is injective if and only if f is bijective, see Proposition 1.7. Hence Proposition 2.21 immediately implies

Corollary 2.24. If $|A| = n \in \mathbb{N}$, then

$$|\{f \in A^{\mathbb{N}_n} | f \text{ is bijective}\}| = n!$$

Example 2.25. If a group of 7 people is asked to line up in a row, how many different rows can be formed with these 7 people?
Answer: Considering each row as an ordered selection without repetitions of the 7 elements of A = group of 7 people, 2.24 yields that there are $7! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 = 5040$ possibilities.

Definition 2.26. (a) For any non-empty set X, a permutation of X is a bijective map $f : X \to X$. The symmetric group of X is defined as

$$S(X) := \{f | f \text{ is a permutation of } X\}$$

We shall see later that $S(X)$ with composition as a group operation is indeed a group in the sense of (abstract) algebra.

(b) The symmetric group of degree n is defined as

$$S_n := S(\mathbb{N}_n) = \{f : \mathbb{N}_n \to \mathbb{N}_n | f \text{is bijective } \}$$

Plugging in $\mathbb{N}_n$ for A, 2.24 immediately implies

Corollary 2.27. $|S_n| = n! \forall n \in \mathbb{N}$ An immediate generalization of 2.27 is the following

Proposition 2.28. For any finite non-empty set X with $|X| = n$, we get $|S(X)| = n!$

Proof. Exercise! □

Hint: Set $A = X$ in 2.24. Then replace $\mathbb{N}_n$ by X and show, using a similar strategy as in the proof of 2.13, that $|\{f \in X^{\mathbb{N}_n} | f \text{ is bijective}\}| = |\{f' \in X^X | f' \text{ is bijective}\}|$.

3 Subsets and Binomial Coefficients

In chapter 2, we discussed ordered selections with and without repetitions, of elements from a finite set. In this chapter we will consider unordered selections of elements from a finite set. It is easier to start with unordered selections without repetitions. We will restrict our attention to finite sets.

Definition 3.1. For $r, n \in \mathbb{N}_0$ an $n-$set X with $|X| = n$. (So the only 0-set is $X = \emptyset$). An $r-$subset of an $n-$set X is a subset $Y \subseteq X$ with $|Y| = r$. Note that in this case 1.4 implies $r \leq n$.

Remark: An unordered selection without repetitions of r elements from an $n-$set X is the same as an $r-$subset of X. Selecting r different elements $x_1, ..., x_r$ from X without paying attention to the order in which these elements are chosen is the same as selecting the $r-$subset $\{x_1, ..., x_r\}$ of X.

Proposition 3.2. For $r, n \in N_0$ with $r \leq n$ and any $n-$set X, we obtain $|\{Y | Y \text{ is an } r - \text{subset of } X\}| = \frac{(n))r}{r!} = \frac{n!}{r!(n-r)!}$

Proof. Set $T = \{(x_1, ..., x_r) \in X^r | x_i \neq x_j \forall i \neq j\}$,

$$S = \{Y | Y \text{is an } r\text{-subset of } X\}$$

and define, $f : T \to S$ by $f((x_1, ..., x_r)) = \{x_1, ..., x_r\} \forall (x_1, ..., x_r) \in T$. We know from Proposition 2.21 that $|T| = (n)_r$, and we want to compute $|S|$. Since T is the disjoint union $T = \bigcup_{Y \in S} f^{-1}(Y)$ (Compare with the proof of 2.3), the addition principle 2.1 implies $|T| = \sum_{y \in S} |f^{-1}(Y)|$. But it is easy to see that for $Y = \{x_1, ..., x_r\} \in S$, i.e. $x_1, ..., x_r$ are r distinct elements of X, we get

$$f^{-1}(Y) = \{x_{\pi(1)}, ..., x_{\pi(r)} | \pi \in S_r\}$$

It now follows from Corollary 2.27 that $|f^{-1}(Y)| = |S_r| = r!$ for any $Y \in S$. Putting things together, we get:

$$(n)_r = |T| = \sum_{Y \in S} |f^{-1}(Y)| = \sum_{Y \in S} r! = |S| \cdot r! \to$$

$$|S| = \frac{(n)_r}{r!} = \frac{(n)(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{r!(n-r)!}$$

□

Example 3.3. (Illustrating the proof of 3.3)
$X = \mathbb{N}_5, n = 5, r = 3 \to$

$$|T| = (5)_3 = 5 \cdot 4 \cdot 3 = 60, |S| = \frac{(5)_3}{3!} = \frac{60}{6} = 10$$

$$Y = \{1, 3, 5\} \in S \to |f^{-1}(Y)| = 3! = 6$$

and

$$f^{-1}(Y) = \{(1,3,5), (3,1,5), (5,3,1), (5,1,3), (3,5,1), (1,5,3)\}$$

Definition 3.4. For $r, n \in \mathbb{N}_0$ with $r \leq n$ the binomial coefficient (or binomial number) $\binom{n}{r}$ ("n choose r") is defined as

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

One also sets $\binom{n}{r} = 0$ if $r > n$ or if $r \in \mathbb{Z}$ and $r < 0$.

Special Cases 3.5. (of binomial coefficients)
Assume $n \in \mathbb{N}_0$ and x is an $n-$set.

(a) $\binom{n}{0} = \frac{n!}{0!n!} = 1 = |\{0 - \text{ subsets of } X\}| = |\{\emptyset\}|$

(b) If $n \geq 1$, then $\binom{n}{1} = \frac{n!}{1!(n-1)!} = n = |\{1 - \text{ subsets of } X\}|$

(c) If $n \geq 2$, then $\binom{n}{2} = \frac{n!}{2!(n-2)!} = \frac{n(n-1)}{2} = |\{2-\text{subsets of} X\}| = |\{ \text{ edges of } K_n\}|$

(d) $\binom{n}{n} = \frac{n!}{n!0!} = 1 = |\{n - \text{subsets of} X\}| = |\{X\}|$

Corollary 3.6. For $r, n \in \mathbb{N}_0$ with $r \leq n$, $\binom{n}{r} = \binom{n}{n-r}$

Proof. This follows immediately from Definition 3.4 since

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

But it also follows from 3.3 since (where $|X| = n$)

$$|\{r - \text{subsets of } X\}| = |\{(n-r) - \text{subsets of } X\}|$$

Since $f : \{r-\text{subsets of } X\} \to \{(n-r)-\text{subsets of } X\}$ defined by $f(Y) = X \setminus Y$ for any r subset Y of X is bijective because

$$g : \{(n-r) - \text{subset of } X\} \to \{r - \text{subsets of } X\}$$

with $g(Z) = X \setminus Z \quad \forall (n-r)-$subsets of X is inverse to f (check this!). Note also that the addition principle 2.1 (or just 1.5) implies

$$|X| = |X \setminus Y| + |Y| \to |X \setminus Y| = |X| - |Y| \to X \setminus Y$$

is an $(n-r)-$subset of X if Y is an $r-$subset of X.

□

Corollary 3.7. $\sum_{r=0}^{n} \binom{n}{r} = 2^n$

Proof. If X is any $n-$set $e.g. X = \mathbb{N}_n$, then $|P(X)| = 2^n$ by 2.17. If we set $S_r = \{r - \text{subsets of } X\}$m then it is clear $P(X) = \sum_{r=0}^{n} s_r$ is a disjoint union. Hence $|P(X)| = \sum_{r=0}^{n} |s_r|$ by the addition principle. But $|S_r| - \binom{n}{r}$ by 3.3. It follows that $2^n = |P(X)| = \sum_{r=0}^{n} |s_r| = \sum_{r=0}^{n} \binom{n}{r}$. □

One of the most important properties of the binomial coefficient is the following which is also know as Pascal's Formula:

Lemma 3.8. For $n \in \mathbb{N}$ and $r \in \mathbb{N}_0$,

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$$

Proof. For $r > n$, we get $0 = 0 + 0$. According to definition 3.4. For $r = n$,

$$1 = \binom{n}{n} = \binom{n-1}{n-1} + \binom{n-1}{n} = 1 + 0$$

For $r = 0$

$$1 = \binom{n}{0} = 0 + 1 = \binom{n-1}{-1} + \binom{n-1}{0}$$

which is also true. Let's finally consider the main case $1 \leq r < n$:

$$\binom{n-1}{r-1}+\binom{n-1}{r} = \frac{(n-1)!}{(r-1)!(n-r)!}+\frac{(n-1)!}{r!(n-r-1)!} = \frac{r(n-1)+(n-r)(n-1)!}{r!(n-r)!} =$$

$$\frac{n(n-1)!}{r!(n-r)!} = \binom{n}{r}$$

This formula leads to a way to organize the binomial coefficients $\binom{n}{r}$ for $n, r \in \mathbb{N}_0$ with $r \leq n$ in the form of a triangle. □

Pascal's triangle 3.9.

level							
$n=0$			$\binom{0}{0}$				
$n=1$		$\binom{1}{0}$		$\binom{1}{1}$			
$n=2$	$\binom{2}{0}$		$\binom{2}{1}$		$\binom{2}{2}$		
$n=3$	$\binom{3}{0}$		$\binom{3}{1}$		$\binom{3}{2}$		$\binom{3}{3}$
$n-1$	$\binom{n-1}{0}$	$\binom{n-1}{1}$	$\cdots$	$\binom{n-1}{r-1}$	$\binom{n-1}{r}$	$\cdots$	$\binom{n-1}{n-1}$
n	$\binom{n}{0}$	$\binom{n}{1}$	$\cdots$		$\binom{n}{r}$	$\cdots$	$\binom{n}{n}$

Pascal's formula has many important consequences. Here is one which we apply in the following:

Lemma 3.10. For any $r \in \mathbb{N}_0$ and any $n \in \mathbb{N}_0$ we obtain

$$\sum_{k=0}^{r} \binom{k+n}{n} = \binom{n+r+1}{n+1}$$

Proof. Exercise! □

The next application of Pascal's formula is the

Theorem 3.11. Binomial Theorem for $n \in \mathbb{N}$ and any $x, y \in \mathbb{R}(\mathbb{C})$, we obtain

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

$$= x^n + nx^{n-1}y + \cdots + \binom{n}{r}x^{n-r}y^r + \cdots nxy^{n-1} + y^n$$

Proof. By induction on n.
$n = 1$

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

$n \geq 2$, by the induction hypothesis $(x+y)^{n-1} = \sum_{r=0}^{n-1}\binom{n-1}{r}x^{n-1-r}y^r$ This implies

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

$$= \sum_{r=0}^{n-1}\binom{n-1}{r}x^{n-r}y^r + \sum_{r=0}^{n-1}\binom{n-1}{r}x^{n-1-r}y^r =$$

$$(*) \quad x^n + \sum_{r=1}^{n-1}\binom{n-1}{r}x^{n-r}y^r + \underbrace{\sum_{r=0}^{n-2}\binom{n-1}{r}x^{n-1-r}y^{r+1}}_{=:S} + y^n$$

Using the substitution $k = r+1$, we can write the sum S as follows: $S = \sum_{k=1}^{n-1}\binom{n-1}{k-1}n^{-k}y^k$. But now we can again rename the summation index and get,

$$S = \sum_{r=1}^{n-1}\binom{n-1}{r-1}x^{n-r}y^r$$

using this in $(*)$, we obtain:

$$(x+y)^n = x^n\sum_{r=1}^{n-1}\binom{n-1}{r}x^{n-r}y^r + \sum_{r=1}^{n-1}\binom{n-1}{r-1}x^{n-r}y^r + y_n$$

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

By 3.8

$$= x^n + \sum_{r=1}^{n-1}\binom{n}{r}x^{n-r}y^r + y^n = \sum_{r=0}^{n}\binom{n}{r}x^{n-r}y^r$$

□

Corollary 3.12. For any $n \in \mathbb{N}$ and $x \in \mathbb{R}$ or $(\mathbb{C})$, we obtain

$$(x+y)^n = \sum_{r=0}^{n}\binom{n}{r}x^{n-r}\underbrace{=}_{3.6} = \sum_{r=0}^{n}\binom{n}{n-r}x^{n-r} = \sum_{r=0}^{n}\binom{n}{r}x^r$$

Remark 3.13. 3.12 immediately implies, with $x \geq 0$

$$(x+1)^n \geq 1 + nx \quad \forall x \in \mathbb{R}$$

It is less obvious, but it can also be proved that

$$(1+x)^n \geq 1 + nx \quad \forall n \in \mathbb{N} \forall x \in \mathbb{R}$$

With $x \geq -1$. ("Bernoulli's Inequality")

Corollary 3.14. (a) $\sum_{r=0}^{n} \binom{n}{r} = 2^n$

(b) $\sum_{r=0}^{n} \binom{n}{r} = \sum_{r=1}^{n} \binom{n}{r} = 2^{n-1}$

Proof. Both equalities follow from 3.12.

(a)

$$2^n = (1+1)^n = \sum_{r=0}^{n} \binom{n}{r} 1^r = \sum_{r=0}^{n} \binom{n}{r}$$

(b)

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

$$\sum_{r=0 \text{ (even)}} \binom{n}{r} - \sum_{r=1 \text{ (odd)}} \binom{n}{r} \to \sum_{r=0 \text{ even}} \binom{n}{r} = \sum_{r=1 \text{ (odd)}} \binom{n}{r} =: a$$

$$\to 2a = \sum_{r=0}^{n} \binom{n}{r} = 2^n \text{by (a)} \ \to a = 2^{n-1}$$

□

Remark 3.15. Using Proposition 3.3, 3.13 (b), can be reinterpreted as follows: For any $n-$set X, we get

$$|\{Y \in P(X) || Y| \text{ is even}\}| = |\{Y \in P(X) || Y| \text{ is odd}\}|$$

Lemma 3.16. $\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n} \quad \forall n \in \mathbb{N}$

Proof. Set $X = \mathbb{N}_{2n}, Y = \mathbb{N}_n$ and

$$Z = X \setminus Y = \{i \in \mathbb{N} | n+1 \leq i \leq 2n\}$$

Then obviously, X is the disjoint union of Y and Z. And for any subset $A \subseteq X$, $A$is the disjoint union of $A \bigcap Y$ and $A \bigcap Z$. We can introduce the following sets:

$$R := \{n - \text{subsets of} X\} = |R| = \binom{2n}{n} \text{ by3.3}$$

For any $k, l \in \mathbb{N}_0$ with $0 \leq k, l \leq n$ we define

$$S_k := \{k - \text{subsets of } Y\} \to |S_k| = \binom{n}{k} \text{ by 3.3}$$

$$T_l := \{l - \text{subsets of} Z\} \to |T_l| = \binom{n}{l} \text{ by 3.3}$$

$$U := \bigcup_{k=0}^{n} S_k \times T_{n-k}, \text{ which is obviously a disjoint union} \to$$

$$|U| \underbrace{=}_{2.1} \sum_{k=0}^{n} |S_k \times T_{n-k}| \underbrace{=}_{2.6} \sum_{k=0}^{n} |S_k| \times |T_{n-k}| = \sum_{k=0}^{n} \binom{n}{k} \binom{n}{n-k}$$

This implies, using 3.6, $|U| = \sum_{k=0}^{n} \binom{n}{k}^2$. To see that $|R| = |U|$ we define two maps:

$$f : R \to U \qquad g : U \to R$$

$$A \to (A \bigcap Y, A \bigcap Z) \qquad (B, C) \to B \bigcup C$$

. One easily verifies (do this!) that f and g are inverse to each other. Hence $|R| = |U| \to \binom{2n}{n} = \sum_{k=0}^{n} \binom{n}{k}^2$ □

Recap 3.17. What we have so far regarding selections in the following, $r, n \in \mathbb{N}$ and X is an $n-$set.

(a) $|\{\text{ordered selections with repetitions of } r\text{elements from} X\}|$

$$= |X^r| = n^r$$

(see 2.11 and 2.20)

(b) $|\{\text{ordered selections without repetition of } r \text{ elements from } X\}|$

$$= |\{(x_1, ..., x_r) \in X^r | x_i \neq x_j \quad \forall i \neq j\} = \begin{cases} (n)_r = \frac{n!}{(n-r)!} & \text{if} r \leq n \\ 0 \quad \text{if } r > n \end{cases}$$

(see 2.21)

(c) $|\{\text{unordered selections without repetition of } r \text{ elements from } X\}|$

$$= |\{r - \text{subsets of} X\}| = \begin{cases} \binom{n}{r} & \text{If} r \leq n \\ 0 & \text{If} r > n \end{cases}$$

(see 3.3)

(d) $|\{\text{unordered selections with repetition of } r \text{ elements from } X\}| =?$

Example In a bag, there are a red, a blue, and a green ball. If one retrieves one ball from the bag, puts it back in the bag, and then repeats this procedure 7 times, how many unordered selections with repetition of 8 balls do you get? For instance, $RBRBRBRB$ would be such a selection, which is considered to be equivalent to $RRRRBBBB$ but not to $RRRBBBBB$. Let's try to make the concept of an unordered selection with repetition more precise.

Definition 3.18. Let $X = \{x_1, ..., x_n\}$ be an $n-$set. An unordered selection with repetition of r elements from X is a sequence $[x_{i_1}, \ldots, x_{i_r}]$ with $i_j \in \{1, \ldots, n\}$ for $1 \leq j \leq r$, with the convention that

$$[x_{i_1}, \ldots, x_{i_r}] = \left[x_{i'_1}, \ldots, x_{i'_r}\right] \quad \text{if there exists } \pi \in S_r \text{ such that } i'_j = i_{\pi(j)} \; \forall\, 1 \leq j \leq r.$$

Example

$$n = 3, r = 4 \quad [x_2, x_1, x_2, x_1] = [x_1, x_2, x_1, x_2]$$

$$i_1 = 2, i_2 = 1, i_3 = 2, i_4 = 1; i'_1 = 1, i'_2 = 2, i'_3 = 1, i'_4 = 2$$

$$\pi \in S_4 \text{ with } \pi(1) = 2, \pi(2) = 1, \pi(3) = 4, \pi(4) = 3$$

Fortunately, there is an easy and unique way to write down these selections with repetitions, namely in the form

$$\left[\underbrace{x_1, ..., x_1}_{m_1 \text{ times}}, \underbrace{x_2, ..., x_2}_{m_2 \text{ times}}, ..., \underbrace{x_n, ..., x_n}_{m_n \text{ times}}\right]$$

Where $m_1, ..., m_n \in \mathbb{N}_0$ and $\sum_{j=1}^{n} m_j = r$. Here $m_j = 0$ simply means that x_j is not an element of the selection. In other words, there is a one-to-one correspondence between the set of all unordered selections with repetition of r elements from X and

$$\mathcal{N}(r, n) := \{(m_1, ..., m_n) \in \mathbb{N}_0^n | \sum_{j=1}^{n} m_j = r\}$$

we set

$$\mathbf{N}(r, n) := |\mathcal{N}(r, n)|, \text{ and the above discussion shows}$$

$$\mathbf{N}(r, n) = \{\text{unordered selections with repetition of } r \text{ elements from } X\}$$

Proposition 3.19. $N(r, n) = \binom{n+r-1}{n-1} \underbrace{=}_{3.6} \binom{n+r-1}{r} \quad \forall r \in \mathbb{N}_0, n \in \mathbb{N}.$

Proof. By induction on n.
$n = 1$:

$$\mathcal{N}(r, 1) = |\{m_1 \in \mathbb{N}_0 | m_1 = r\}| = |\{r\}| = 1 = \binom{r}{0}$$

To see how the induction works, let's do the cases
$n = 2$:

$$\mathcal{N}(r, 2) = |\{(m_1, m_2) \in \mathbb{N}_0^2 | m_1 + m_2 = r\}| \leftrightarrow m_1 = r - m_2$$

$$= \sum_{m_2=0}^{r} |\{(r-m_2, m_2)\}| = \sum_{m_2=0}^{r} 1 = r+1 = \binom{r+1}{1}$$

$n = 3$:

$$\mathcal{N}(r,3) = |\{(m_1,m_2,m_3) \in \mathbb{N}_0^3 | m_1+m_2+m_3 = r\}|$$

$$\sum_{m=3}^{r} |\{(m_1,m_2) \in \mathbb{N}_0^3 | m_1+m_2 = r-m_3\}|$$

Set $k = r - m_3$

$$= \sum_{k=0}^{r} |\{(m_1,m_2) \in \mathbb{N}_0^2 | m_1+m_2 = k\}|$$

$$= \sum_{k=0}^{r} \mathcal{N}(k,2) = \underbrace{=}_{n=2} \sum_{k=0}^{r} (k+1)\frac{(r+1)(r+2)}{2} = \binom{r+2}{2}$$

It's now clear how to generalize the argument used for the case $n = 3$ in order to carry out the general induction step. $n-1 \to n$:

$$\mathcal{N}(r,n) = |\{(m_1,...,m_n) \in \mathbb{N}_0^n | \sum_{j=1}^{n} m_j = r\}$$

$$= \sum_{m_n=0}^{r} |\{(m_1,...,m_{n-1}) \in \mathbb{N}_0^{n-1} | \sum_{j=1}^{n-1} m_j = r-m_n\}|$$

$$= \sum_{m_n=0}^{r} \mathcal{N}(r-m_1, n-1) = \sum_{k=0}^{r} \mathcal{N}(k, n-1)$$

By the I.H,

$$= \sum_{k=0}^{r} \binom{n+k-2}{n-2} \underbrace{=}_{3.10} \binom{n+r-1}{n-1}$$

□

Example 3.20. Picking 1 ball out of a bag with a red, a blue, and a green ball, putting it back in the bag, and doing this 8 times corresponds to an unordered selection with repetition of 8 elements from the set $X = \{\text{red ball, blue ball, green ball}\}$ with $|X| = 3$. Hence by 3.19m there are $\binom{3+8-1}{3-1} = \binom{10}{2} = \frac{10\cdot 9}{2} = 45$ possible outcomes.

Example 3.21. How many possible outcomes are there if one throws 3 dice simultaneously?

Clarification: We are only interested in the outcome of the throw not which die shows how many dots. For instance we do not distinguish between.

$$\begin{pmatrix} D_1 & D_2 & D_3 \\ 2 & 3 & 5 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} D_1 & D_2 & D_3 \\ 5 & 2 & 3 \end{pmatrix}$$

So the question is equivalent to determining the number of unordered selections with repetition of 3 elements from the set $X = \mathbb{N}_6$. Hence there are

$$\binom{6+3-1}{6-1} = \binom{8}{5} = \frac{8 \cdot 7 \cdot 6}{3!} = 56 \text{ such throws}$$

(See also Exercise 2 of section 11.2 on page 110 in **[B]**. The result 56 is given without explanation on that page, and the general amount for throws with n dice on page 413.)
We are now going to give yet another interpretation of binomial coefficients which naturally leads to a generalization namely multinomial coefficients.

Question 3.22. If one has n objects of two types, where the objects of the same type are indistinguishable, with r objects of type 1 (and hence $n - r$ objects of type 2), how many different ways are there to put these n objects into n different boxes?

Reformulation: How many elements of $\{1, 2\}^n$ have precisely r coordinates =1? (Objects: numbers 1 and 2; Boxes: n coordinates). Answer:Since these r out of the n coordinates to be filled by 1's this means that we have to choose an $r-$ subset of the $n-$set $\mathbb{N}_n$. Hence, by 3.3, there are $\binom{n}{r}$ possible choices.

Generalization 3.23. Suppose we are given $k, n \in \mathbb{N}$ and k numbers $r_1, ..., r_k \in \mathbb{N}_0$ satisfying $\sum_{j=1}^{k} r_j = n$.
How many elements of $\mathbb{N}_k^n$ have $r_1 1's, r_2 2's, ..., r_k k's$?

Definition 3.24. For $k, n \in \mathbb{N}$ and $r_j \in \mathbb{N}_0, \quad 1 \leq j \leq k$ with $\sum_{j=1}^{k} r_j = n$, we define the multinomial coefficient.

$$\binom{n}{r_1, ..., r_k} = \frac{n!}{r_1! \cdots r_k!}$$

Note that for $k = 2$, we get the special case:

$$\binom{n}{r_1, r_2} = \binom{n}{r_1} = \binom{n}{r_2} \quad \text{(recall that } r_1 + r_2 = n\text{)}$$

Proposition 3.25. For $k, n \in \mathbb{N}$ and $r_1, ..., r_k \in \mathbb{N}_0$ with $\sum_{j=1}^{k} r_j = n$, we obtain $|\{(a_1, ..., a_n) \in \mathbb{N}_k^n || \{1 \leq i \leq n | a_i = j\}| = r_j \quad \forall 1 \leq j \leq k\}| =$

$$\binom{n}{r_1, r_2, ..., r_k}$$

Proof. There are $\binom{n}{r_1}$ possible places for the coordinates to be = 1
Then there are $\binom{n-r_1}{r_2}$ remaining possible places for the coordinates to be = 2,
$\vdots$
Finally there is $\binom{n-r_1-r_2-...-r_{k-1}}{r_k} = 1$ options remaining for the coordinates

$= k$.
It follows that $|\{(a_1, ..., a_n) \in \mathbb{N}_k^n || \{1 \leq i \leq n | a_i = j\}| = r_j \quad \forall 1 \leq j \leq k\}| =$

$$\binom{n}{r_1}\binom{n-r_1}{r_2}\cdots\binom{n-r_1-...-r_{k-1}}{r_k} =$$

$$\frac{n!}{r_1!(n-r_1)!}\cdot\frac{(n-r_1)!}{r_2!(n-r_1-r_2)!}\cdots\frac{(n-r_1-...-r_{k-2})!}{r_{k-1}!(n-r_1-...-r_{k-1)!}}\cdot\frac{(n-r_1-...-r_{k-1})!}{r_k!0!}$$

$$= \frac{n!}{r_1!r_2!\cdots r_k!} = \binom{n}{r_1, r_2, ..., r_k}$$

□

Example 3.26. How many 11-letter "words" can be made from the letters of the word $ABRACADABRA$?
Here we have $n = 11$ and $k = 5$ (number of letters in the set $\{A, B, C, D, R\}$) with $r_1 = 5$ (= occurrence of A), $r_2 = 2, r_3 = 1, r_4 = 1, r_5 = 2 \rightarrow \binom{11}{5,2,1,1,2} = \frac{11!}{5!2!2!} = 83,160$ words can be formed from these letters.

We are now getting to the reason why these numbers are called " multinomial coefficients".

Theorem 3.27. Multinomial Theorem. For $n, k \in \mathbb{N}$ and any $x_1, ..., x_r \in \mathbb{R}(\mathbb{C})$ we obtain,

$$(x_1 + \cdots + x_k)^n = \sum_{\substack{r_1,...,r_k \in \mathbb{N}_0 \\ r_1+\cdots+r_k=n}} \binom{n}{r_1, \ldots, r_k} x_1^{r_1} \cdots x_k^{r_k}.$$

Proof. By induction on k.
$k = 1$:

$$\sum_{r_1=n} \binom{n}{r_1} x^{r_1} = \binom{n}{n} x_1^n = x_1^n$$

$k = 2$: This is the binomial theorem:

$$\sum_{\substack{r_1,r_2 \in \mathbb{N}_0 \\ r_1+r_2=n}} \binom{n}{r_1, r_2} x_1^{r_1} x_2^{r_2} = \sum_{r_1=0}^{n} \binom{n}{r_1, n-r_1} x_1^{r_1} x_2^{n-r_1}.$$

$$= \sum_{r_1=0}^{n} \binom{n}{r_1} x_1^{r_1} x_2^{n-r_1} \underbrace{=}_{3.11} (x_1 + x_2)^n$$

$k \geq 3$:

$$(x_1 + \cdots + x_k)^n = (x_1 + \cdots + x_{k-2} + (x_{k-1} + x_k))^n$$

By the Induction Hypothesis:

$$\sum_{\substack{r_1,...,r_{k-2},m \in \mathbb{N}_0 \\ r_1+...+r_{k-2}+m=n}} \binom{n}{r_1, ..., r_{k-2}, m} x_1^{r_1} ... x_{k-2}^{r_{k-2}} (x_{k-1} + x_k)^m$$

$$\overset{k=2}{=} \sum_{\substack{r_1,\dots,r_{k-2},\, m\in\mathbb{N}_0 \\ r_1+\cdots+r_{k-2}+m=n}} \binom{n}{r_1,\dots,r_{k-2},m} x_1^{r_1}\cdots x_{k-2}^{r_{k-2}} \sum_{\substack{r_{k-1}\in\mathbb{N}_0 \\ 0\le r_{k-1}\le m}} \binom{m}{r_{k-1}} x_{k-1}^{r_{k-1}} x_k^{m-r_{k-1}}.$$

$$\underbrace{=}_{\text{by the distributive law}} \sum_{r_1+\dots+r_{k-2}+m=n} \sum_{r_{k-1}+r_k=m} \binom{n}{r_1,\dots,r_{k-2},m}\binom{m}{r_{k-1}} x_1^{r_1}\cdots x_{k-2}^{r_{k-2}} x_{k-1}^{r_{k-1}} x_k^{r_k}$$

$$\underbrace{=}_{\text{by rewriting the previous sum:}} \sum_{r_1+\cdots+r_k=n} \binom{n}{r_1,\dots,r_{k-2},r_{k-1},r_k}\binom{r_{k-1}+r_{k-2}}{r_{k-1}} x_1^{r_1}\cdots x_k^{r_k}$$

$$= \sum_{r_1+\dots+r_k=n} \binom{n}{r_1,\dots,r_{k-2},r_{k-1},r_k} x_1^{r_1}\cdots x_k^{r} r_k$$

Since $\binom{n}{r_1,\dots,r_{k-2},r_{k-1}+r_k}\binom{r_{k-1}+r_k}{r_{k-1}}$

$$= \frac{n!}{r_1!\cdots r_{k-2}!(r_{k-1}+r_k)!}\frac{(r_{k-1}+r_k)!}{r_{k-1}!r_k!}$$

$$= \frac{n!}{r_1!\cdots r_{k-2}!r_{k-1}!r_k!} \underbrace{=}_{def.} \binom{n}{r_1,\dots,r_k}$$

□

Example 3.28. Compute $(x+y+z)^4$(for $x,y,z\in\mathbb{R}$). If we want to apply the multinomial theorem, we need to compute the multinomial coefficients $\binom{4}{a,b,c}$ for $a,b,c\in\mathbb{N}_0$ with $a+b+c=4$. Note that any permutation of a,b,c yields the same multinomial coefficient. So we may assume $a\ge b\ge c$.

$$4=4+0+0 \text{ gives us } \binom{4}{4,0,0} = \frac{4!}{4!0!0!} = 1$$

$$4=3+1+0 \text{ gives us } \binom{4}{3,1,0} = \frac{4!}{3!1!0!} = 4$$

$$4=2+2+0 \text{ gives us } \binom{4}{2,2,0} = \frac{4!}{2!2!0!} = 6$$

$$4=2+1+1 \text{ gives us } \binom{4}{2,1,1} = \frac{4!}{2!1!1!} = 12$$

If we write $(x+y+z)^4$ as a sum, we get

$$|\{(a,b,c)\in\mathbb{N}_0^3 | a+b+c=4\}| \underbrace{=}_{3.19} \binom{3+4-1}{3-1} = \binom{6}{2} = 15$$

terms of the form $x^a y^b z^c$. But recall that $\binom{4}{a,b,c} = \binom{4}{a',b',c'}$ if (a',b',c') is obtained from (a,b,c) by a permutation. That's why we only needed to compute 4 of the multinomial coefficients here. So the final result is the following:

$$(x+y+z)^4 = x^4+y^4+z^4$$
$$+4(x^3y+xy^3+x^3z+xz^3+y^3z+z^3y)$$
$$+6(x^2y^2+x^2z^2+y^2z^2)$$
$$+12(x^2yz+xy^2z+xyz^2)$$

There are, like for binomial coefficients, many formulas for multinomial coefficients. Let's just mention the following generalization of 3.14 (a).

Corollary 3.29. For any $k,n \in \mathbb{N}$, we obtain:

$$k^n = (1+\cdots+1)^n = \sum_{\substack{r_1,\ldots,r_k\in\mathbb{N}_0 \\ r_1+\cdots+r_k=n}} \binom{n}{r_1,\ldots,r_k}.$$

4 The Inclusion-Exclusion Principle, Euler's and Möbius' functions

In the following, $A_1,...,A_n$ are n finite sets. Under the assumption that $A_i \bigcap A_j = \emptyset \forall i \neq j$, we derived in 2.1 that $|\bigcup_{i=1}^n A_i| = \sum_{i=1}^n |A_i|$ (addition principle). Without assuming $A_1 \bigcap A_2 = \emptyset$, we observed in 2.2 that

$$A_1 \bigcup A_2 = |A_1| + |A_2| - |A_1 \bigcap A_2|$$

We are now going to deduce a generalized formula for $|\bigcup_{i=1}^n A_i|$, which is the most basic form of the inclusion-exclusion principle.

Definition and Remark 4.1. Given the finite sets $A_1,...,A_n$ and $k \in \mathbb{N}$ with $k \leq n$, we set

$$\alpha_k := \sum_{1\leq i_1<\cdots<i_k\leq n} |\bigcap_{j=1}^n A_{i_j}|$$

In particular

$$\alpha_1 = \sum_{i_1=1} |A_{i_1}| = \sum_{i=1}^n |A_i|$$

$$\alpha_2 = \sum_{1\leq i<l\leq n} |A_i \bigcap A_l| \text{ and } |\bigcap_{j=1}^n A_j| \quad (i_j = l \forall 1 \leq j \leq n \text{ if } k=n)$$

In general, the number of summands in α_k is

$$|\{(i_1,...,i_k) \in \mathbb{N}_n^k | 1 \leq i_1 < i_2 < \cdots < i_k \leq n\}|$$

$$= |\{k - \text{subsets of} \mathbb{N}_n\}| = \binom{n}{k} \text{ by } \underline{3.3}$$

(IF we require $1 \leq i_1 < \cdots < i_k \leq n$ for $k-$tuples $(i_1, ..., i_k) \in \mathbb{N}_n^k$, then they correspond bijectively to subsets $\{i_1, ..., i_k\} \subseteq \mathbb{N}_n$.)

Theorem 4.2. $|\bigcup_{i=1}^n A_i| = \sum_{k=1}^n (-1)^{k-1} \alpha_k$

Proof. By induction on n.
$n = 1$:

$$|A_1| = (-1)^0 \alpha_1 = |A_1|$$

$n = 2$:

$$|A_1 \bigcup A_2| = \sum_{k=1}^{2} (-1)^{k-1} \alpha_k = \alpha_1 - \alpha_2 = |A_1| + |A_2| - |A_1 - A_2| \text{ by 2.2}$$

We now carry out the induction step $n \to n+1$ for $n \geq 2$: Given $n+1$ sets $A_1, ..., A_{n+1}$, we have to show

$$|\bigcup_{i=1}^{n+1} A_i| = \sum_{k=1}^{n+1} (-1)^{k-1} \alpha_k$$

, where

$$\tilde{\alpha_k} := \alpha_k(A_1, ..., A_{n+1)=\sum_{1 \leq i_1 < \cdots < i_k \leq n+1} |\bigcap_{j=1}^{k} A_{i_j}| \text{ for } 1 \leq k \leq n+1$$

Writing

$$\bigcup_{i=1}^{n+1} A_i = \left(\bigcup_{i=1}^{n} A_i\right) \bigcup A_{n+1} \text{ and using the case } n = 2, \text{we first get}$$

$$|\bigcup_{i=1}^{n+1} A_i| = |\bigcup_{i=1}^{n} A_i| + |A_{n+1}| - |\left(\bigcup_{i=1}^{n} A_i\right) \bigcap A_{n+1} (*)$$

Note that by Example 1 of Homework # 1 $(\bigcup_{i=1}^n A_i) \cap A_{n+1} = \bigcup_{i=1}^n (A_i \bigcap A_{n+1})$

$$= \bigcup_{i=1}^{n} A_i', \text{ where } A_i' := A_i \bigcap A_{n+1} \quad \forall 1 \leq i \leq n$$

Using the inclusion hypothesis to $\bigcup_{i=1}^n A_i$ as well as $\bigcup_{i=1}^n A_i$ we get

$$|\bigcup_{i=1}^{n} A_i| = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} |\bigcap_{j=1}^{k} A_{i_j}|$$

and

$$|\bigcup_{i=1}^{n} A_i'| = \sum_{k=1}^{n}(-1)^{k-1} \sum_{1\le i_1<\cdots<i_k\le n} |\bigcap_{j=1}^{k} A_{i_j}'|$$

Now observe that $\bigcap_{j=1}^{k} A_{i_j}' = \bigcap_{j=1}^{k}(A_{i_j} \bigcap A_{n+1}) = \bigcap_{j=1}^{k} A_{i_j} \bigcap A_{n+1}$
Hence we can also write

$$|\bigcup_{i=1}^{n} A_i'| = \sum_{k=1}^{n}(-1)^{k-1} \sum_{1\le i_1<\cdots<i_k\le i_{k+1}=n+1} |\bigcap_{j=1}^{k+1} A_{i_j}|$$

$$(l = k+1) \quad = \sum_{l=2}^{n+1}(-1)^{l-2} \sum_{1\le i_1<\cdots<i_{l-1}<i_l=n+1} ||\bigcap_{j=1}^{l} A_{i_j}$$

whereas,

$$|\bigcup_{i=1}^{n} A_i| = \sum_{i=1}^{n}|A_i| + \sum_{k=2}^{n}(-1)^{k-1} \sum_{1\le i_1<\cdots<i_k\le n} |\bigcap_{j=1}^{k} A)i_j|$$

Plugging this all into $(*)$, we get

$$|\bigcup_{i=1}^{n+1} A_i| = \sum_{i=1}^{n+1}|A_i| + \sum_{k=2}^{n}(-1)^{k-1} \sum_{1\le i_1<\cdots<i_k=n} |\bigcap_{j=1}^{k} A_{i_j}| + \sum_{l=2}^{n+1}(-1)^{l-1} \sum_{1\le i_1<\cdots<i_l=n+1} |\bigcap_{j=1}^{l} A_{i_j}|$$

$$= \underbrace{\sum_{i=1^{n+1}}}_{=\tilde{\alpha_1}} |A_i| + \sum_{k=2}^{n+1}(-1)^{k-1} \sum_{1\le i_1<\cdots<i_k\le n+1} |\bigcap_{j=1}^{k}|A_{i_j}|$$

$$\sum_{k=1}^{n+1}(-1)^{k-1}\tilde{\alpha_k}$$

□

Corollary 4.3. $|A_1 + A_2 + A_3|$
$= |A_1| + |A_2| + |A_3| - |A_1 \bigcap A_2| - |A_1 \bigcap A_3| - |A_2 \bigcap A_3| + |A_1 \bigcap A_2 \bigcap A_3|$.

Remark 4.4. Note that in the sum in 4.2 each α_k is itself a sum of $\binom{n}{k}$ in terms by 4.1 hence the total number of summands in this expression for $|\bigcup_{i=1}^{n} A_i|$ is

$$\sum_{k=1}^{n}\binom{n}{k} = 2^n - 1$$

by 3.14(a)

Example 4.5. How many natural numbers ≤ 500 are divisible by 3 or 5 or by 7? Set

$$A = \{n \in \mathbb{N} | n \leq 500 \text{ and } 3|n\}$$

$$B = \{n \in \mathbb{N} | n \leq 500 \text{ and } 5|n\}$$

$$C = \{n \in \mathbb{N} | n \leq 500 \text{ and } 7|n\}$$

We have to compute $|A \bigcup B \bigcup C|$ First observe that

$$|A| = \lfloor \frac{500}{3} \rfloor = 166 \quad (A = \{3k | k \in \mathbb{N} \text{ and } k \leq 166\}$$

$$|B| = \lfloor \frac{500}{5} \rfloor = 100, \quad |C| = \lfloor \frac{500}{7} \rfloor = 71$$

$$|A \bigcap B| = |\{n \in \mathbb{N} | n \leq 500 \text{ and } 15|n\}| = \lfloor \frac{500}{15} \rfloor = 33$$

$$|A \bigcap C| = |\{n \in \mathbb{N} | n \leq 500 \text{ and } 21|n\}| = \lfloor \frac{500}{21} \rfloor = 23$$

$$|A \bigcap B| = |\{n \in \mathbb{N} | n \leq 500 \text{ and } 35|n\}| = \lfloor \frac{500}{35} \rfloor = 14$$

$$|A \bigcap B| = |\{n \in \mathbb{N} | n \leq 500 \text{ and } 105|n\}| = \lfloor \frac{500}{105} \rfloor = 4$$

Applying 4.3 we get

$$|A \bigcup B \bigcup C| = 166 + 100 + 71 - 33 - 23 - 14 + 4 = 271$$

Theorem 4.2 is often applied in order to determine how many elements of a given set X do not have any of a certain set of properties $P_1, ..., P_n$. In other words, if we set $A_i := \{x \in X | x \text{ has property } P_i\}$ for$1 \leq i \leq n$, then we need to determine $|X \setminus \bigcup_{i=1}^{n} A_i|$. 4.2 immediately implies the following

Corollary 4.6. If $A_i \subseteq X \quad \forall 1 \leq i \leq n,$ then

$$|X \setminus \bigcup_{i=1}^{n} A_i| = |X| - |\bigcup_{i=1}^{n} A_i| = |X| + \sum_{k=1}^{n} (-1)^k \alpha_k$$

with $\alpha_k = \sum_{1 \leq i_1 < \cdots < i_k \leq n} |\bigcap_{j=1}^{k} A_{i_j}| \quad \forall i \leq k \leq n$ as in 4.1.

Example 4.7. In a group of 73 students

52 can play the piano
25 can play the violin
20 can play the flute
17 can play piano and violin
12 can play piano and flute
7 can play violin and flute
1 can play all three instruments

How many of these 73 students can play neither piano, nor flute, nor violin?
Answer 73-52-25-20+17+12+7-1 = 11 students can neither play piano nor violin nor flute.

We are going to discuss another application of the inclusion-exclusion principle which refers to obtain elements of S_n (see 2.26).

Definition 4.8. A permutation $\pi \in S_n$ is said to be a derangement of $\mathbb{N}_n$ if $\pi(i) \neq i \forall i \in \mathbb{N}_n$. We set

$$D_n := \{\pi \in S_n | \pi \text{ is a derangement of } \mathbb{N}_n\} \subseteq S_n$$

and $d_n := |D_n|$

Examples $d_1 = 0, d_2 = 1, d_3 = 2$

Proposition 4.9. $d_n = n! \sum_{k=0}^{n} (-1)^k \frac{1}{k!} = n! \sum_{k=2}^{n} (-1)^k \frac{1}{k!}$

Proof. Set $A_i := \{\pi \in S_n | \pi(i) = i\}$ for $1 \leq i \leq n$. Then $D_n = S_n \setminus \bigcup_{i=1}^{n} A_i$ in order to apply 4.6, we need to compute $\alpha_k = \sum_{1 \leq i_1 < \cdots < i_k \leq n} |\bigcap_{j=1}^{k} A_{i_j}|$ for all $1 \leq k \leq n$. Now

$$\begin{aligned} |\bigcap_{j=1}^{k} A_{i_j}| &= |\{\pi \in S_n | \pi(i_j) = i_j \forall 1 \leq j \leq k\}| \\ &= |S(\mathbb{N}_n \setminus \{i_j | 1 \leq j \leq k\}| \\ &= (n-k)! \text{ by Proposition 2.8} \end{aligned}$$

Hence

$$\alpha_k = \sum_{1 \leq i_1 < \cdots < i_k \leq n} (n-k)! = \underbrace{=}_{4.1} \binom{n}{k} (n-k)! = \frac{n!}{k!}$$

it follows that $d_n = |S_n| + \sum_{k=1}^{n} (-1)^k \alpha_k =$

$$n! + \sum_{k=1}^{n} (-1)^k \frac{n!}{k!} = n! \sum_{k=0}^{n} (-1)^k \frac{1}{k!} = n! \sum_{k=2}^{n} (-1)^k \frac{1}{k!}$$

□

Example 4.10. (a) $d_4 = 4!(\frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!}) = 12 - 4 + 1 = 9$
Elements of D_4 : 4-cycles $(\cdot\cdot\cdot\cdot)$: 6
products of 2 disjoint transpositions $(\cdot\cdot)(\cdot\cdot)$: 3

(b) $d_5 = 5!(\frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!}) = 60 - 20 + 5 - 1 = 44$
Elements of D_5 : 5 cycles $(\cdot\cdot\cdot\cdot\cdot)$: 24
products of a 3-cycle and a transposition $(\cdot\cdot\cdot)(\cdot\cdot)$: 20

Remark 4.11. Since $\frac{1}{e} = e^{-1} = \sum_{k=0}^{\infty} = \frac{(-1)^k}{k!}, \sum_{k=0}^{n} \frac{(-1)^k}{k!}$ is already for small n a good approximation of $\frac{1}{e}$. It is in fact not difficult to show that d_n is, for all $n \in \mathbb{N}$, the integer closest to $\frac{n!}{e}$. $(e = 2.71828\cdots), e^{-1} = 0.367787\cdots)$.

Example 4.12. At a party, 10 gentlemen check their hats. In how many ways can their hats be returned so that

(a) No gentleman

(b) Precisely one gentleman gets the hat which he arrived in?

Answers

(a) Asks about the number of derangements of $\mathbb{N}_{10}$. So the aswer is $d_{10} = \sum_{k=2}^{10} \frac{10!}{k!} = 1,334,961 (\frac{n!}{e} = 1,334,960.916)$

(b) Set $A_i = \{\pi \in S_n | \pi(i) = i \text{ and } \pi(j) \neq j \forall j \neq i\}$. It is easy to verify that $|A_i| = |A_n| = d_{n-1} \forall 1 \leq i \leq n$. Since the union $\bigcup_{i=1}^{n} A_i$ is disjoint, we obtain

$$|\bigcup_{i=1}^{n} A_i| = \sum_{i=1}^{n} |A_i| = \sum_{i=1}^{n} d_{n-1} = nd_{n-1}$$

However, $\bigcup_{i=1}^{n} A_i = \{\pi \in S_n | \pi \text{ gives precisely one element of } \mathbb{N}_n\}$ So, in our example, the number of ways such that precisely one gentleman gets back his own has is

$$10 \cdot d_9 = 10! \sum_{k=0}^{9} (-1)^k \frac{1}{k!} = d_{10} - 1 \underbrace{=}_{(a)} 1,334,960$$

We will now introduce Euler's $\varphi-$function, whose computation is another application of the inclusion-exclusion principle.

Definition 4.13. Euler's function $\varphi : \mathbb{N} \to \mathbb{N}$ is defines as follows

$$\varphi(n) = |\{m \in \mathbb{N} | m \leq n \text{ and } \gcd(m, n) = 1\}|$$

Examples:

$$\varphi(1) = 1, \varphi(2) = 1, \varphi(3) = 2, \varphi(6) = 2$$

Lemma 4.14. For any prime number p and any $e \in \mathbb{N}$,

$$\varphi(p^e) = p^e - p^{e-1} = p^{e-1}(p-1) = p^e \left(1 - \frac{1}{p}\right).$$

Proof. If $m \in \mathbb{N}$, then $\gcd(m, p^e) = 1 \leftrightarrow \gcd(m, p) = 1 \leftrightarrow p \nmid m$. It follows that:

$$\varphi(p^e) = |\{m \in \mathbb{N}_{p^e} | \gcd(m, p^e) = 1\}|$$

$$= |\{m \in \mathbb{N}_{p^e} | p \nmid m\}|$$

$$|\mathbb{N}_{p^e} \setminus \{k \cdot p | k \in \mathbb{N} \text{ and } 1 \leq k \leq p^{e-1}\}|$$

$$= p^e - p^{e-1}$$

□

Lemma 4.14 can now be generalized as follows:

Theorem 4.15. If $n \in \mathbb{N}, n \geq 2$, has the prime factorization

$$n = p_1^{e_1} \cdots p_r^{e_r} = \Pi_{i=1}^r p_i^{e_i}, i.e. p_1, ..., p_r$$

are r pairwise distinct primes and $e_i \in \mathbb{N} \forall 1 \leq i \leq r$, then

$$\varphi(n) = n\Pi_{i=1}^r (1 - \frac{1}{p_i}) = \big(= \Pi_{i=1}^r (P_i^{e_i} - P_i^{e_i - 1}\big)$$

Proof. For $1 \leq i \leq r$, set $A_i := \{m \in \mathbb{N}_n | P_i | m\}$. Since, for any $m \in \mathbb{N}$, $\gcd(m, n) = 1 \leftrightarrow p_i \nmid m \forall 1 \leq i \leq r \leftrightarrow m \notin A_i \forall 1 \leq i \leq r$, we get

$$\varphi(n) = |\mathbb{N}_n \setminus \bigcup_{i=1}^r A_i|.$$

In order to apply 4.6, we need to determine the cardinalities $|\bigcap_{j=1}^k A_{i_j}$ for $1 \leq k \leq r$.

$$\begin{aligned} |\bigcap_{j=1}^k A_{i_j}| &= |\{m \in \mathbb{N}_n | P_{ij} | m \forall 1 \leq j \leq k\}| \\ &\{|m \in \mathbb{N}_n p_{i_1} \cdots P_{i_k} | m\}| \\ &|\{l p_{i_1} \cdots p_{i_k} | l \in \mathbb{N}, 1 \leq l \leq \frac{n}{p_{i_1} \cdots p_{i_k}}\}| \\ &= \frac{n}{p_{i_1} \cdots p_{i_k}} \end{aligned}$$

Using 4.6 it now follows that

$$\begin{aligned} \varphi(n) &= n + \sum_{k=1}^r (-1)^k \sum_{1 \leq i_1 < \cdots < i_k \leq r} \frac{r}{p_{i_1} \cdots p_{i_k}} \\ &= n(1 + \sum_{k=1}^r (-1)^k \sum_{1 \leq i_1 < \cdots < i_k \leq r} \frac{1}{p_{i_1}} \cdots \frac{1}{p_{i_k}} \\ &= n\Pi_{i=1}^r (1 - \frac{1}{p_i}) \end{aligned}$$

Where the last equation follows from the lemma below applies with $x_i = \frac{1}{p_i} \forall i$. □

Lemma 4.16. For all $r \in \mathbb{N}$, and all $x_1, ..., x_r \in \mathbb{R}$, we obtain

$$\Pi_{i=1}^r (1 - x_i) = 1 + \sum_{k=1}^r (-1)^k \sum_{1 \leq i_1 < \cdots < i_k \leq r} x_{i_1} \cdots x_{i_k}$$

Proof. Exercise! □

Example 4.17. (a) $n = 60 \to n = 2^2 \cdot 3 \cdot 5 \to \varphi(n) = (2^2-2)(3-1)(5-1) = 16$

(b) $n = 1000 \to 2^3 \cdot 5^3 \to \varphi(n) = (2^3 - 2^2)(5^3 - 5^2) = 4 \cdot 100 = 400.$

Euler's φ function is related to Möbius' μ-function, which we will introduce now.

Definition 4.18. The Möbius' function μ: $\mathbb{N} :\to \mathbb{N}_0$ is defined as follows $\mu(1) = 1$ and if $n \geq 2$, with prime factorization $n = p_1^{e_1} \cdots p_r^{e_r}$ (here $p_1, ..., p_r$ are again r distinct primes and $e_i \in \mathbb{N} \forall i$)

$$\mu(n) = \begin{cases} (-1)^r & \text{if } e_i = 1 \forall 1 \leq i \leq r \\ 0 & \text{if } e_i \geq 2 \text{ for at least one } i \end{cases}$$

Lemma 4.19. $\varphi(n) = \sum_{d|n} \mu(d) \frac{n}{d} \quad \forall n \in \mathbb{N}$

Proof. We first remark the notation $\sum_{d|n}$ means that the summation is taken over all $d \in \mathbb{N}_n$ which satisfy $d|n$. We also verify that $\varphi(1) = 1 = \mu(1)\frac{1}{1}$. Now let $n \geq 2$ and $n = p_1^{e_1} \cdots p_r^{e_r}$ be its prime factorization. In the proof of Theorem 4.15 it was shown that:

$$\varphi(n) = n + \sum_{k=1}^{r} (-1)^k \sum_{1 \leq i_1 < \cdots < i_k \leq r} \frac{n}{p_{i_1} \cdots p_{i_k}}$$

This can be rewritten as follows:

$$\varphi(n) = \mu(1)\frac{n}{1} + \sum_{k=1}^{r} \sum_{1 \leq i_1 < \cdots < i_k \leq r} \mu(p_{i_1} \cdots p_{i_r}) \frac{n}{p_{i_1} \cdots p_{i_r}}$$

$$= \sum_{f_1, ..., f_r \in \{0,1\}^r} \mu(p_1^{f_1} \cdots p_r^{f_r}) \frac{n}{p_1^{f_1} \cdots p_1^{f_r}}$$

$$\sum_{(f_1, ..., f_r) \in \mathbb{N}_0^r} \mu(p_1^{f_1} \cdots p_r^{f_r}) \frac{n}{p_1^{f_1} \cdots p_r^{f_r}}$$

$$\sum_{d|n} \mu(d) \frac{n}{d}$$

□

Lemma 4.20 is an important ingredient in the following Möbius' inversion formula

Theorem 4.20. If $f, g : \mathbb{N} \to \mathbb{R}(\mathbb{C})$ are functions which satisfy $f(n) = sum_{d|n} g(d)$, then $g(n) = \sum_{d|n} \mu(d) f(\frac{n}{d}) \quad \forall n \in \mathbb{N}$.

Theorem 4.21. *Proof.* Using the given formulas, we can compute

$$\sum_{d|n} \mu(d) f(\frac{n}{d}) = \sum_{d|n} \mu(d) \sum_{c|\frac{n}{d}} g(c) = \sum_{(c,d) \in S_1} \mu(d) g(c)$$

with

$$S_1 := \{(c,d) \in \mathbb{N}^2 | d|n \text{ and } c|\frac{n}{d}\}$$

Claim:

$$S_1 = S_2 := \{(c,d) \in \mathbb{N}^2 | c|n \text{ and } d|\frac{n}{c}\}$$

Proof of the claim:

(a) $S_1 \subseteq S_2$, $(c,d) \in S_1 \to d|n$ and hence $\frac{n}{d} \in \mathbb{N}$; also $c|\frac{n}{d} \to \exists k \in \mathbb{N} : kc = \frac{n}{d} \to kdc = n, c|n$ and $kd = \frac{n}{c} \to d|\frac{n}{c} \to (c,d) \in S_2$

(b) $S_2 \subseteq S_1$

$$(c,d) \in S_2 \to c|n\frac{n}{c} \in \mathbb{N}; d|\frac{n}{c} \to \exists l \in \mathbb{N} \text{ with}$$

$ld = \frac{n}{c} \to cld = n \to d|n$ and $cl = \frac{n}{d} \to c|\frac{n}{d} \to (c,d) \in S_1$. Since $S_1 = S_2$, we can now write:

$$\sum_{d|n} \mu(d) f(\frac{n}{d}) = \sum_{(c,d) \in S_2} \mu(d) g(c) = \sum_{c|n} \sum_{d|\frac{n}{c}} \mu(d) g(c) =$$

$$\sum_{c|n} g(c) \underbrace{\sum_{d|\frac{n}{c}} \mu(d)}_{=0 \text{ if } \frac{n}{c} \geq 2 \text{ by } \underline{4.20}} = g(n) \sum_{d|\frac{n}{n}} \mu(d) = g(n)\mu(1) = g(n)$$

□

With similar arguments as in the proof of Theorem 4.21 one can also prove the following converse of the Möbius' inversion formula:

Proposition 4.22. If $f, g : \mathbb{N} \to \mathbb{R}(\mathbb{C}, \mathbb{R})$ are functions which satisfy

$$f(n) = \sum_{d|n} \mu(d) g(\frac{n}{d}, \text{ then } g(n) = \sum_{d|n} f(d) \forall n \in \mathbb{N}$$

Proof. Exercise! □

Corollary 4.23. $\sum_{d|n} \varphi(d) = n \forall n \in \mathbb{N}$.

Proof. By 4.19 $\varphi(n) = \sum_{d|n} \mu(d) \frac{n}{d} \forall n \in \mathbb{N}$.
Applying 4.22 with $f = \varphi$ and $g = id_{\mathbb{N}}$ now yields $n = g(n) = \sum_{d|n} \varphi(d)$. □

5 Congruences

We begin this chapter with the repetition of some basic elementary number theory.

Theorem 5.1. For any $a, b \in \mathbb{Z}$ with $b \neq 0$, there exist uniquely determined $a \in \mathbb{Z}$ and $r \in \mathbb{N}_0$ with $a = bq + r$ and $0 \leq r < |b|$.

Proof. (a) Uniqueness If $a = bq_1 + r_1 = bq_2 + r_2$ with $q_1, q_2 \in \mathbb{Z}; r_1, r_2 \in \mathbb{N}_0$ and $0 \leq r_1, r_2 < |b|$, then $b(q_1 - q_2) = r_2 - r_1$ and $r_2 - r_1 \in \mathbb{Z}$ with $-|b| < r_2 - r_1 < |b|$. Since $b|r_2 - r_1$, the inequalities imply $r_2 - r_1 = 0 \to r_1 = r_2$;

$$\begin{cases} b(q_1 - q_2) = 0 \\ b \neq 0 \end{cases} \quad \to q_1 = q_2$$

(b) Existence If $a, b > 0$, this is proved in Theorem 8.2 on p. 66 in **[B]**. If $a = 0$ and $b \in \mathbb{Z} \setminus \{0\}$, we can write $0 = b \cdot 0 + 0q = r = 0$. If $a < 0$ and $b < 0$, then there exist $q', r' \in \mathbb{N}_0$ with

$$-a = bq' + r' \text{ and } 0 \leq r' < b$$

If $r' = 0$, we can write $a = bq + r$ with $r = r' = 0$ and $q = -q' \in \mathbb{Z}$. If $r' \neq 0$, then $0 < r' < b$ and hence $0 < r := b - r' < b$ os that $a = b(-q') - r' = b(-q' - 1) + (b - r') = bq + r$ with $q = -q' - 1 \in \mathbb{Z}$ and $r = b - r' \in \mathbb{N}$ with $0 \leq r < b$. Finally, if $b < 0$, we already know that there exist $q' \in \mathbb{Z}$ and $r \in \mathbb{N}_0$ with $a = |b|q + r$ and $0 \leq r < |b|$. Hence also, since $|b| = -b$, $a = bq + r$ with $q := -q', q \in \mathbb{Z}$.

□

Definition 5.2. We say that two integers $x, y \in \mathbb{Z}$ are congruent modulo m and write $x \equiv y \pmod{m}$ if $m|x - y$.

Examples:

$$2 \equiv 6 \pmod 4 \quad 4 \equiv 7 \pmod 3 \quad \cdots$$

Remark 5.3. Being congruent modulo m defines an equivalence relation on $\mathbb{Z}$. More precisely: If we define, for $x, y \in \mathbb{Z}$, xRy by $x \equiv y \pmod{m}$, then R is an equivalence relation on $\mathbb{Z}$.

Exercise Prove this! (Recapitulate section 7.2 in **B** if necessary).

Every equivalence relation R on a set S partitions S into equivalence classes, see section 7.3 in **[B]**.

Definition 5.4. The equivalence classes with respect to the equivalence relation " being congruent modulo m" are congruence classes. For $a \in \mathbb{Z}$, we denote its congruence class by $[a]_m$ or simply by $[a]$ if m is understood. Hence

$$[a]_m = \{b \in \mathbb{Z} | a \equiv b \pmod m\}$$

$$\{b \in \mathbb{Z} | \exists k \in Z \text{ s.t. } b = a + k \cdot m\}$$

$$\{a + k \cdot m | k \in \mathbb{Z}\}$$

Examples:

$$[0]_2 = \{\text{even integers}\} = [4]_2 = [6]_2 = \cdots$$

$$[1]_2 = \{\text{odd integers}\} = [3]_2 = [-1]_2 = \cdots$$

$$[2]_3] = \{..., -4, -1, 2, 5, ...\}$$

$$= \{2 + 3k | k \in \mathbb{Z}\}$$

Corollary 5.5. (of 5.1) There are precisely m different congruence classes modulo m, namely $[0]_m, ..., [m-1]_m$.

Proof. Given $a \in \mathbb{Z}$, there exist $q \in \mathbb{Z}$ and $r \in \mathbb{N}_0$ with $0 \leq r <$ such that $a = mq + r \to m|a - r \to a \equiv r \pmod{m} \to [a]_m = [r]_m$, with $r \in \{0, 1, ..., m-1\}$. Hence

$$\mathbb{Z} = \bigcup_{r=0}^{m-1} [r]_m$$

. And this union is disjoint (hence a partition of $\mathbb{Z}$). Since for $r_1, r_2 \in \mathbb{N}_{m-1} \cup \{0\} [r_1]_m = [r_2]_m$ implies $m|r_1 - r_2$ and $-m < r_1 - r_2 < m \to r_1 - r_2 = 0, r_1 = r_2$. Recall that two different equivalence classes are always disjoint. □

Definition 5.6. The set of congruence classes modulo m is denoted by $\mathbb{Z}_m$, i.e.

$$Z_m := \{[a]_m | a \in \mathbb{Z}\}$$

$$= \{[a]_m | a \in \mathbb{N}_0\}$$

$$\{[a]_m | a \in \mathbb{N}_0, 0 \leq a \leq m-1\}$$

In view of 5.5 $|\mathbb{Z}_m| = m$.

Using congruence classes, one can more nicely apply the pigeonhole principle to subsets of $\mathbb{Z}$ (not just $\mathbb{N}$), the generalizing Example 1.13 Exercise 5(a) of Homework # 1

Corollary 5.7. For any subset $s \subseteq \mathbb{Z}$ with $|S| > m$, there exist $x, y \in S$ with $x \neq y$ and $m|x - y$

Proof. Consider the function $f : S \to \mathbb{Z}_m$ defined by $f(x) = [x]_m$ for $x \in S$. By the pigeonhole principle, 1.9, there exist $x, y \in S$ with $x \neq y$ and $[x]_m = [y]_m$, since $|S| > m = |\mathbb{Z}_m|$. But $[x]_m = [y]_m$ means that

$$x \equiv y \pmod{m} \leftrightarrow m|x - y$$

□

We will now introduce some computation rules for congruences.

Lemma 5.8. Assume that $x_1, x_2, y_1, y_2 \in \mathbb{Z}$ satisfy $x_1 \equiv x_2 \pmod{m}$ and $y_1 \equiv y_2 \pmod{m}$. Then

(a) $x_1 + y_1 \equiv x_2 + y_2 \pmod{m}$

(b) $x_1 y_1 \equiv \pmod{m}$

(c) $x_1^n \equiv \pmod{x)_2^n} \pmod{m} \quad \forall n \in \mathbb{N}$

Proof. (a) We are given that $m|x_1 - x_2$ and $m|y_1 - y_2$. But then also $m|(x_1 - x_2) + (y_1 - y_2) = (x_1 + y_1) - (x_2 + y_2)$, which means that $x_1 + y_1 \equiv x_2 + y_2 \pmod{m}$.

(b) Since $m|x_1 - x_2$ and $m|y_1 - y_2$, we also have

$$m|(x_1 - x_2)y_1 + (y_1 - y_2)x_2 = x_1y_1 - x_2y_2 \to x_1y_1 \equiv x_2y_2 \pmod{m}$$

(c) This is most conveniently proved by induction on n

Proof. $n = 1$

$$x_1^1 = x_1 \equiv x_2 = x_2^1 \pmod{m} \text{ by assumption}$$

$n \geq 2$

$$x_1^n = x_1 \cdot x_1^{n-1} \text{ and } x_2^n = x_2 \cdot x_2^{n-1}, \text{ where again}$$

$$x_1 \equiv x_2 \pmod{m} \text{ by assumption and} x_1^{n-1} \equiv x_2^{n-1} \pmod{m}$$

By I.H. Applying (b) with $y_1 = x_1^{n-1}$ and $y_2 = x_2^{n-1}$, we get:

$$x_1y_1 = x_1^n \equiv x_2y_2 = x_2^n \pmod{m}$$

. □

□

As a first application of these computation rules for congruences, we derive some well know $(m = 3, 9)$ or not so well known $(m = 11)$, divsibility criteria.

Definition 5.9. If $x \in \mathbb{N}$ is given in the usual decimal notation

$$x = a_n a_{n-1} \cdots a_1 a_0, \text{ i.e. } x = \sum_{i=0}^{n} a_i 10^i$$

where, $a_i \in \mathbb{N}_9 \bigcup \{0\}$ and $a_n \neq 0$, we get:

$$\theta(x) := \sum_{i=0}^{n} a_i \text{ sum of digits}$$

$$\delta(x) := \sum_{i=0}^{n} (-1)^i a_i$$

Lemma 5.10. For $x = a_n a_{n-1} \cdots a_0 \in \mathbb{N}$ as in 5.9, we obtain

(a) $x \equiv \theta(X) \pmod 9$ and $x \equiv \theta(x) \pmod 3$. In particular $9|x \leftrightarrow 9|\theta(x)$ and $3|x \leftrightarrow 3|\theta(x)$.

(b) $x \equiv \delta(x) \pmod 1 1$. In particular, $11|x \leftrightarrow 11|\delta(x)$

Proof. (a) Observe that $10 \equiv 1 \pmod 9$ and $10 \equiv 1 \pmod 3$. Hence 5.8 implies $x = \sum_{i=0}^{n} a_i 10^i \equiv \sum_{i=0}^{n} a_i 1^i = \theta(x) \pmod 9$ and $\pmod 3$.
In particular, $9|x \leftrightarrow x \equiv 0 \pmod 9 \leftrightarrow \theta(x) \equiv 0 \pmod 9 \leftrightarrow 9|\theta(x)$, and similarly for 3 instead of 9.

(b) $10 \equiv -1 \pmod 1 1 \underbrace{\rightarrow}_{5.8} x = \sum_{i=0}^{n} a_i 10^i \equiv \sum_{i=0}^{n} a_i (-1)^i = \delta(x) \pmod 1 1$.

Hence, $11|x \leftrightarrow x \equiv 0 \pmod 1 1 \leftrightarrow \delta(x) \equiv 0 \pmod 1 1 \leftrightarrow 11|\delta(x)$.

□

Example 5.11. (a) Is 12345678 divisible by 9 (or 3)?

$$\text{Yes since, } \sum_{i=1}^{8} = \frac{8 \cdot 9}{2} = 4 \cdot 9 \text{ which is a multiple of 9}$$

Remark: For any permutation $\pi \in S_8$ and $x = \pi(1) \cdots \pi(8)$ (8-digit number), we also have $\theta(x) = \sum_{i=1}^{8} \pi(i) = \sum_{i=1}^{8} i = 4 \cdot 9$, so that x is divisible by 9 (and 3).

(b) $\delta(12345678) = 8-7+6-5+4-3+2-1 = 4 \neq 0 \pmod 1 1 \rightarrow 11 \nmid 12345678$. By contrast to (a), permuting the digits matters here: For instance

$$\delta(71643528) = 8 - 2 + 5 - 3 + 4 - 6 + 1 - 7 = 0 \rightarrow 11|71643528$$

Definition and Remark 5.12. One important consequence of 5.8 is that one can well-define $+$ and $\cdot$ on $\mathbb{Z}_m$ by

$$[x]_m + [y]_m := [x + y]_m$$

and

$$[x]_m \cdot [y]_m := [xy]_m \forall x, y \in \mathbb{Z}$$

This is well defined since it doesn't depend on the choice of representation of the congruence classes, if:

$$[x]_m = [x']_m \text{ and } [y]_m = [y']_m, \text{ then by} 5.8$$

$$[x' + y']_m = [x + y]_m \text{ and } [x'y']_m = [x + y]_m$$

With this addition and multiplication, $(\mathbb{Z}_m, +, \cdot)$ becomes a commutative ring with the multiplicative identity $[1]_m$.

We are now turning to the question which elements $[x]_m \in \mathbb{Z}_m$ have a multiplicative inverse, i.e., an element $[y]_m \in \mathbb{Z}_m$ with $[x]_m[y]_m = [1]_m$, in which case we will write $[x]_m^{-1} = [y]_m$. Examples:

$$[2]_3^{-1} = [2]_3, [2]_5^{-1} = [3]_5$$

$[2]_6$ does not exist since $[2]_6[y]_6 = [2y]_6 \neq [1]_6$ for all $y \in \mathbb{Z}$ because $2 \nmid 2y - 1 \rightarrow y \nmid 2y - 1 \rightarrow 2y \not\equiv 1 \pmod 6$

In this context, we should quickly repeat the definition and basic properties of the greatest common divisor $\gcd(a, b)$ of two integers $a, b \in \mathbb{Z}$.

Repetition 5.13. For $a, b \in \mathbb{Z}$ with $(a, b) \neq (0, 0)$, there exists a uniquely determined $d \in \mathbb{N}$ which satisfies:

(a) $d|a$ and $d|b$

(b) If $c \in \mathbb{Z}$ is such that $c|a$ and $c|b$ then $c|d$. d is called the greatest common divisor of a and b, and we write $d = \gcd(a, b)$. One also defines $\gcd(0, 0) := 0$.

(c) $\gcd(a, b) = \gcd(b, a) \quad \forall a, b \in \mathbb{Z}$

(d) $\gcd(a, 0) = |a| \quad \forall a \in \mathbb{Z}$

(e) $\gcd(a, b) = \gcd(|a|, |b|) \quad \forall a, b \in \mathbb{Z}$

(f) $\gcd(a, 1) = 1 \quad \forall a \in \mathbb{Z}$

(g) If $a, b \in \mathbb{N}$ and $a = \Pi_{i=1}^{r} p_i^{e_i}$, $b = \Pi_{i=1}^{r} p_i^{f_i}$ with r distinct primes $p_1, ..., p_r$ and $e_i, f_i \in \mathbb{N}_0 \forall 1 \leq i \leq r$, then $\gcd(a, b) = \Pi_{i=1}^{r} p_i^{\min(e_i, f_i)}$

The next property is an important consequence of the Euclidean Algorithm, which is a systematic method to compute $\gcd(a, b)$.

Theorem 5.14. For any $a, b \in \mathbb{Z}$, there exist $m, n \in \mathbb{Z}$ with $\gcd(a, b) = ma + nb$.

Proof. For $a = 0$ or $b = 0$, the claim is trivial. For instance $\gcd(a, 0) = |a| = m \cdot a + n \cdot 0$ for any $n \in \mathbb{Z}$ and $m = \begin{cases} 1 & \text{if } a \geq 0 \\ -1 & \text{if } a < 0 \end{cases}$ for $a, b > 0$, then the claim is proved in Theorem 8.4 on page 69 in [**B**]. And for arbitrary $a, b \in \mathbb{Z}$ with $a \neq 0 \neq b$ we obtain that there exist $m', n' \in \mathbb{Z}$ with

$$\gcd(a, b) = \gcd(|a|, |b|) = m'|a| + n'|b| = ma + nb$$

where we get $m = \begin{cases} m' & \text{if } a > 0 \\ -m' & \text{if } a < 0 \end{cases}$ and $n = \begin{cases} n' & \text{if } b > 0 \\ -n' & \text{if } b < 0 \end{cases}$ □

this allows us to prove the following

Proposition 5.15. For $x \in \mathbb{Z}, [x]_m \in \mathbb{Z}_m$ has an inverse in $\mathbb{Z}_m$ if and only if $\gcd(x, m) = 1$.

Proof. "Only if": Assume that there exist $y \in \mathbb{Z}$ with $[x]_m[y]_m = [1]_m \leftrightarrow xy \equiv 1 \pmod{m} \leftrightarrow m|xy - 1$ if there existed any prime number p with $p|x$ and $p|m$, then also $p|xy - 1$ and $p|x \rightarrow p| - 1$, a contradiction. Hence the only natural number which divides x and m is 1, implying $\gcd(x, m) = 1$.
"If": Now assume that $\gcd(x, m) = 1$, then by Theorem 5.14, there exist $k, l \in \mathbb{Z}$ with $1 = kx + lm \rightarrow m|1 - kx \rightarrow 1 \equiv kx \pmod{m} \leftrightarrow [1]_m = [kx]_m = [k]_m \cdot [x]_m$. Hence $[k]_m = [x]_m^{-1}$. □

Remark 5.16. There is a slight ambiguity in the formulation of Proposition 5.15. Namely, there are uniquely many elements $x' \in \mathbb{Z}$ which represent the congruence class $[x]_m$. Namely, $[x']_m = [x]_m$ if and only if $x' = x + km$ for some $k \in \mathbb{Z}$. But it is easily verified that

$$\gcd(x, m) = \gcd(x + km, m) \forall k \in \mathbb{Z} \quad \text{Exercise!}$$

Whether this gcd is 1 or not.

Definition and Corollary 5.17. We set $U_m := \{[x]_m \in \mathbb{Z} | [x]_m \text{ has an inverse in } \mathbb{Z}_m\}$ This set (which is in fact the group of units of the ring $\mathbb{Z}_m$) is often also denoted by $\mathbb{Z}_m^\times$. By 5.15

$$U_m = \{[x]_m | x \in \mathbb{N}_m \text{ and } \gcd(x, m) = 1\}$$

Here we use5.6: $[1]_m, ..., [m-1]_m, [m]_m = [0]_m$ are precisely the m distinct elements of $\mathbb{Z}_m$. It follows that $|U_m| = |\{x \in \mathbb{N}_m | \gcd(x, m) = 1\}| = \varphi(n)$. With Euler's φ function as introduced in 4.13. Before we prove Euler's theorem about the elements of U_m, we need a few more preparations.

Lemma 5.18. (a) If $x, y \in \mathbb{Z}$ satisfy $\gcd(x, m) = 1 = \gcd(y, m) = 1$, then $\gcd(xy, m) = 1$. It follows that U_m is closed under multiplication.

(b) If $a, b, c \in \mathbb{Z}$ satisfy $ab \equiv ac \pmod m$ and $\gcd(a, m) = 1$ then also $b \equiv c \pmod m$.

Proof. (a Recall that two integers are relatively prime (i.e. their gcd is 1) if they do not have a common prime factor. But the prime factors of xy are the prime factors of x together with the prime factors of y. So if x and y both have no common prime factors with m neither has y. For U_m it follows:

$$[x], [y] \in U_m \to \gcd(x, m) = \gcd(y, m) = 1 \to \gcd(xy, m) = 1 \to$$

by 5.15, 5.17

$$[x]_m \cdot [y]_m = [xy]_m \in U_m$$

(b It is convenient to use the invertibility of $[a]_m$ here. There exist $x \in \mathbb{Z}$ with $[x]_m[a]_m = [1]_m$ i.e. $xa \equiv 1 \pmod m$. Multiplying the given congruence

$$ab \equiv ac \pmod m$$

with x, we get

$$(xa)b \equiv (xa)c \pmod m$$

But since $xa \equiv 1 \pmod m$, applying 5.8, we get

$$1 \cdot b \equiv 1 \cdot c \pmod m, \text{ i.e. } b \equiv c \pmod m$$

as claimed.

□

Example 5.19. Without the condition $\gcd(a, m) = 1$, the "cancellation" $ab \equiv ac \pmod m \to b \equiv c \pmod m$ is usually incorrect. See for instance:

$$4 \cdot 4 = 16 \equiv 4 \cdot 1 \pmod 1 2, \text{ clearly} 4 \neq 1 \pmod 1 2$$

The next observation is related to Lemma 5.8

Lemma 5.20. For any $[a]_m \in U_m$, the multiplicative map

$$f = f_{[a]_m} : U_m \to U_m$$

Is bijective.

Proof. Note first that by 5.15 and 5.18 (a) f is really a map from U_m to U_m: If $[a]_m, [x]_m \in U_m$ for $a, x \in \mathbb{Z}$, then $\gcd(a, m) = \gcd(x, m) = 1$, hence also $\gcd(ax, m) = 1$, and hence $[a]_m \cdot [x]_m = [ax]_m \in U_m$.
There are two ways to show that f is bijective.
1^{st} Proof: Since U_m is finite, it suffices by Proposition 1.7 that f is injective.But this follows immediately from 5.18 (b)

$$\text{if, for } [x_1], [x]_2 \in U_m, f([x_1]_m) = f([x_2]_m), \text{then}$$

$$[a]_m[x_1]m = [a]_m[x_2]_m \leftrightarrow [ax_1]_m = [ax_2]_m \leftrightarrow$$

$$ax_1 \equiv ax_2 \pmod{m} \to x_1 \equiv x_2 \pmod{()} \text{ since } \gcd(a, m) = 1$$

$$\to [x_1]_m = [x_2]_m$$

2^{nd} Proof: Since $[a]_m \in U_m$, $[a]_m$ is invertible, and so there exists $b \in \mathbb{Z}$ with $[a]_m \cdot [b]_m = [1]_m$. It is now easy to verify (do this!) that the maps $f_{[a]_m}$ and $f_{[b]_m}$ are inverse to each other. Hence they are both bijective. □

Example 5.21. Consider U_9. Since $9 = 3^2$, $|U_9| = \varphi(9) = 3^2 - 3 = 6$. Explicitly $U_9 = \{[1], [2], [4], [5], [7], [8]\}$. (Since we use $m = 9$ in this example, I will omit the index 9 in the congruence classes.) Multiplication with [5] defines the following permutation of U_9: $f = f_{[5]} : U_9 \to U_9$

$$[1] \to [5]$$

$$[2] \to [1]$$

$$[4] \to [2]$$

$$[5] \to [7]$$

$$[7] \to [8]$$

$$[8] \to [4]$$

We can check that $f_{[2]}$ is the inverse of $f_{[5]}$. We are now in a position to state and prove Euler's Theorem.

Theorem 5.22. (Euler) For any $a \in \mathbb{Z}$, with $\gcd(a, m) = 1$, we obtain

$$a^{\varphi(n) \equiv 1 \pmod{m}}$$

Proof. Set $k = \varphi(m)$, and let $[x_1]_m, \cdots [x_k]_m$ be the different elements of U_m. One of them is $[a]_m$. For $f = f_{[a]_m}$ as in 5.20, $f(U_m) = U_m$, so

$$U_m = \{[x_i]_m | 1 \leq i \leq k\} = \{[ax_i]_m | 1 \leq i \leq k\}$$

Taking the product of all elements of U_m, we get

$$\Pi_{i=1}^k [x_i]_m = \Pi_{i=1}^k [ax_i]_m \leftrightarrow \left[\Pi_{i=1}^k x_i\right]_m = \left[\Pi_{i=1}^k (ax_i)\right]_m$$

$$\leftrightarrow \Pi_{i=1}^k x_i \equiv \Pi_{i=1}^k (ax_i) = a^k \Pi_{i=1}^k x_i \pmod{m}$$

Now all the factors on both sides of this congruence are relatively prime, and so using 5.18, we deduce

$$1 \equiv a^k = a^{\varphi(m)} \pmod{m}$$

□

Example 5.23. (5.21 continued)

$$a^6 \equiv 1 \pmod{9} \quad \forall a \in \{1, 2, 4, 5, 6, 7, 8\}, \text{e.g,}$$

$$64 = 2^6 \equiv 1 \pmod{9}, 5^6 = 15,625 \equiv 1 \pmod{9}$$

(That $9|15,625$ can also be checked using 5.10 (a)).

It is clear that Euler's theorem is very useful in order to determine congruences since one does not have to compute the powers $a^{\varphi(n)}$ explicitly. A special case of Euler's Theorem is the following which is also known as "Fermat's Little Theorem"

Corollary 5.24. For $a \in \mathbb{Z}$ and any prime p, we obtain

(a) $a^{p-1} \equiv 1 \pmod{p}$ if$p \nmid a$

(b) $a^p \equiv a \pmod{p}$ always.

Proof. If p is a prime, then $\varphi(p) = p - 1$, and $\gcd(a, p_=1$ if and only if $p \nmid a$. Hence (a) follows immediately from 5.22. If $p \nmid a$, then we can multiply both sides of the congruence $a^{p-1} \equiv 1 \pmod{p}$ with a and get $a^p \equiv a \pmod{p}$. But if $p|a$,then $a \equiv 0 \pmod{p}$ and also $a^p \equiv 0 \pmod{p}$, and so (b) holds for all $a \in \mathbb{Z}$. □

Example 5.25. For any $a \in \mathbb{N}$, the last digit of a^5 in its decimal notation is the same as the last digit of a. Argument: If $a = a_k, \cdots, a_1, a_0$ then $a \equiv a_0 \pmod{5}$ since $10 \equiv 0 \pmod{5}$. Similarly, if $a^5 = b_l, \cdots, b_1, b_0$, then $a^5 \equiv b_0 \pmod{5}$. Now $a \equiv a^5 \pmod{5}$ by Fermat's Little Theorem $\rightarrow a_0 \equiv b_0 \pmod{5}$. But also $a \equiv a^5 \pmod{2}$ (both, a and a^5 are either even or odd), and $a \equiv a_0 \pmod{2}, a^5 \equiv b_0 \pmod{2}$, imply $a_0 \equiv b_0 \pmod{2}$ as well. So $5|a_0 - b_0$ and $2|a_0 - b_0 \rightarrow$(since 2 and 5 are two distinct primes) $10|a_0 - b_0$. But the digits a_0 and b_0 are elements of $\mathbb{N}_9 \cup \{0\} \rightarrow a_0 = b_0$

Corollary 5.26. (of 5.24) For any prime number p, and any $a, b \in \mathbb{Z}$, we obtain

$$(a+b)^p \equiv a^p + b^p \pmod{p}$$

Proof. By 5.24(b), $(a+b)^p \equiv a+b \pmod{p}$, $a^p \equiv a \pmod{p}$ and $b^p \equiv b \pmod{p}$. It follows from 5.8(a) and the transitivity of congruences that $a^p + b^p \equiv (a+b)^p \pmod{p}$. □

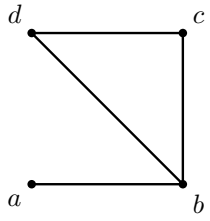
6 Graphs

Definition 6.1. A graph $G = (V, E)$ is a pair consisting of a non-empty set V and a (possibly empty) subset $E \subseteq P(V)$ with $|e| = 2 \forall e \in E$.

V is called the vertex set and E the edge set of G. If $e = \{x, y\} \in E$, we say that the vertices $x, y \in V$ are adjacent or that they are joined by the edge e. We also say that e contains the vertices x, y and that they lie on e. G is called finite if V (and hence also E) is finite. In this chapter, we will mainly consider finite graphs. Remark What is simply called a graph here (and in the textbook) is often called a "simple graph" or an undirected graph without loops and multiple edges in the mathematical literature. Directed graphs or "digraphs" as well as graphs which have loops (joining a vertex with itself or multiple edges (admitting more than one edge connecting two vertices) also have important applications but will currently not be considered here.

Visualization 6.2. Graphs are combinatorial objects but they can also be considered as geometric objects. Each vertex $x \in V$ is represented by a dot or node and each edge $e = \{x, y\}$ by a line connecting the two nodes x and y.

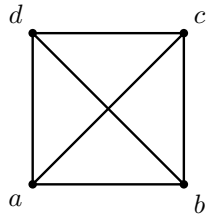
Example The graph $G = (V, E)$ with $V = \{a, b, c, d\}$ and $E = \{\{a, b\, \{b, c\}, \{c, d\}, \{b, d\}\}$ is identified with the geometric object (1-dimensional simlicial complex)



We already mentioned one type of graphs in connection with Ramsay numbers

Definition 6.3. The complete graph K_n of order n is the graph $K_n = (V_n, E_n)$ with $|V_n| = n$ and $E_n = \{2 - \text{subsets of } V_n\}$ by 3.3, $|E_n| = \binom{n}{2} = \frac{n(n-1)}{2}$. Conversely, it is easy to check (do this!) that if a graph $G = (V, E)$ satisfies $|V| = n$ and $|E| = \binom{n}{2}$, then G is the complete graph of order n.

Example $k_4 = \{\{a, b, c, d\}, E\}$ with $E = \{\{a, b\}, \{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}, \{c, d\}\}$ i.e.



K_n is an example of a regular graph, which we will introduce soon but first another.

Definition 6.4. If $G = (V, E)$ is a graph, the degree of a vertex $x \in V$ is defined as $\delta(x) = |D_x|$ where $D_x := \{e \in E | x \in e\}$

Example 6.5. (a) If $G = (V, E)$ is as in the example in 6.2, then $\delta(a) = 1, \delta(b) = 3, \delta(c) = 2, \delta(d) = 2$.

(b) If $G = K_n$, then $\delta(x) = n - 1$ for all vertices of K_n. Again, this is also a characterization of complete graphs: If $G = (V, E)$ with $|V| = n$ and $\delta(x) = n - 1 \forall x \in V$, then G is a complete graph of order n.

Definition 6.6. A Graph $G = (V, E)$ is called regular if there exists $r \in \mathbb{N}_0$ with $\delta(x) = r \forall x \in V$. In this case, G is called regular of degree r.

Examples:

(a) $G = (V, E)$ is regular of degree 0 if and only if $E = \emptyset$.

(b) If $G = (V, E)$ with $|V| = n$, then G is regular of degree $n - 1$ if and only if $G = K_n$.

An interesting connection between the degrees of the vertices of a finite graph is the following "handshaking lemma".

Proposition 6.7. If $G = (V, E)$ is a finite graph, then

$$\sum_{x \in V} \delta(x) = 2|E|$$

Proof. We set $S := \{(x, e) \in V \times E | x \in e\} \subseteq V \times E$ and use the counting strategy of Proposition 2.9, For all $x \in V$, we get

$$r_x(S) = |\{e \in E | x \in e\}| = \delta(x)$$

and for all $e \in E$, we get

$$c_e(S) = |\{x \in V | x \in e\}| = 2$$

Hence

$$|S| = \sum_{x \in V} \delta(x) = \sum_{e \in E} 2 = 2|E|$$

□

An immediate consequence of 6.7 is the following.

Corollary 6.8. If a finite graph $G = (V, E)$ is regular of degree r, then $r|V| = 2|E|$.

Example For $K_n = (V_n, E_n)$, we get $|V_n| = n$ and $|E_n| = \binom{n}{2} = \frac{n(n-1)}{2}$. K_n is $(n-1)$ regular, and so $(n-1)|V_n| = (n-1)n = 2|E_n|$ as stated in 6.8. Another consequence of 6.7 is the following

Corollary 6.9. If $G = (V, E)$ is finite and $V_{\text{odd}} = \{x \in V | \delta(x) \text{ is odd}\}$, then $|V_{\text{odd}}|$ is even.

Proof. Set $V_{\text{even}} := \{x \in V | \delta(x) \text{ is even}\}$. Then $V = V_{\text{even}} \cup V_{\text{odd}}$ is a disjoint union. 6.7 implies

$$\sum_{x \in V_{\text{odd}}} \delta(x) + \sum_{x \in V_{\text{even}}} \delta(x) = 2|E| \rightarrow \text{(In view of the definition of even)}$$

$$\sum_{x \in V_{\text{odd}}} \delta(x) \text{ is even} \rightarrow \text{(in view of the definition of } V_{\text{odd})} |V_{\text{odd}}| \text{ is even}$$

□

Special Cases 6.10. If $G = (V, E)$ is a finite regular graph of odd degree, then $|V|$ is even. This follows already from 6.8. So, for instance, there is no regular graph of degree 3 with precisely 5 vertices.

Remark: 6.9 can be formulated as follows: In a group V of people, the number of people who shake hands with an odd number of people in V is even. This gave 6.7 the name "Handshaking Lemma". Yet another consequence of 6.7, which will be useful is the following, is

Corollary 6.11. If $G = (V, E)$ is a finite graph and

$$m := \min\{\delta(x) | x \in V\}, \text{ then } m \leq 2\frac{2|E|}{|V|}$$

In particular, if $|E| < |V|$, then there exists a vertex $x \in V$ with $\delta(x) \leq 1$.

Proof. By definition of m, $\sum_{x \in V} \delta(x) \geq m|V|$ Hence 6.7 implies $m|V| \leq 2|E| \rightarrow m \leq 2\frac{|E|}{|V|}$. Therefore, if $|E| < |V|$, there exists $x \in V$ with $\delta(x) = m < 2 \rightarrow \delta(x) \leq 1$ (since $\delta(x) \in \mathbb{N}_0$. □

An important feature of graphs is that one can connect vertices by walking along edges.

Definition 6.12. A walk in a graph $G = (V, E)$ is a finite sequence $x_1, ..., x_k$ of vertices such that x_i and x_{i+1} are adjacent (so in particular distinct) for all $1 \leq i \leq k-1$. The walk $x_1, ..., x_k$ is called a path if $|\{x_1, ..., x_k\}| = l$. A cycle of length $k-1$ is called a walk $x_1, ..., x_k$ satisfying $k \geq 4$ and $x_1 = x_k$ and $|\{x_1, ..., x_k\}| = k-1$. (So that $x_1, ..., x_{k-1}$ is a path).

Definition 6.13. If $G = (V, E)$ is a graph and $x, y \in V$, we write $x \sim y$ if x and y can be joined by a walk in G, i.e., if there exists a walk $x_1, \ldots, x_k$ in G with $x_1 = x$ and $x_k = y$. It is easy to check (do this!) that $\sim$ defines an equivalence relation on V. The graph G is called connected if $x \sim y$ for all $x, y \in V$.

Remark 6.14. If two vertices x, y in a graph $G = (V, E)$ can be joined by a walk, they can also be joined by a path.

Proof. Let $x_1, ..., x_k = y$ be a walk of minimal length joining x and y in G. We claim that this walk is a path, i.e. $|\{x_1, ..., x_k\}| = k$. Suppose not. Then there exist $1 \leq i < j \leq k$ with $x_i \neq x_j$. But then $x_1, ..., x_i, x_{j+1}, ..., x_k$ is a walk joining x and y of shorter length $i + (k - i) - 1 = (k - 1) - (j - i) < k - 1$, contradicting the minimality of the original walk $x_1, ..., x_k$. Note that x_l and x_{l+1} are adjacent for $1 \leq l \leq i - 1$ and for $j + 1 \leq l \leq k - 1$ since $x_1, ..., x_k$ is a walk. Note also that x_i and x_{j+1} are adjacent since $x_i = x_j$ and x_j, x_{j+1} are adjacent vertices of the walk $x_1, ..., x_k$. The fact that $\sim$ is an equivalence relation leads to the following definition $\square$

Definition 6.15. If $G = (V, E)$ is a graph and $x \in V$, the connectedness component of x in G is the subgraph

$$G'(V', E') \text{ of} G \text{ defined by}$$

$$V' := [x] = \{y \in V | x \sim y\} \text{ and}$$

$$E' := \{e = \{y, z\} \in E | y, z \in V'\}$$

Remark 6.16. If $x, y \in V$ with $[x] \neq [y]$ (and hence $[x] \cap [y] = \emptyset$, see exercise (2) of homework # 8) and if $x' \in [x]$, $y' \in [y]$, then x' and y' are not adjacent.

Reason: If $\{x', y'\} \in E$, were an edge, then we had

$x \sim x' \sim y' \sim y \to x \sim y \to x \in [x] \cap [y] \to [x] = [y]$, contradiction our assumption.

Consequence If $\{[x_i]_{i \in I}\}$ are the distinct equivalence classes in V with respect to $\sim$ (for some suitable index set I), and

$$G_i = (\underbrace{[x_i]}_{=V_i}, E) \text{ with } E_i = \{e = \{y, z\} \in E | y, z \in [x_i]\}$$

and the corresponding connectedness components, then $E = \bigcup_{i \in I} E_i$ disjoint union, since no edge can connect two vertices which are elements of distinct equivalence classes. So

$$G = \bigcup_{i \in I} G_i$$

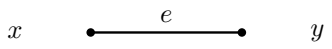
is also a disjoint union, in the sense that

$$V = \bigcup_{i \in I} V_i \quad \text{and } E = \bigcup_{i \in I} E_i \text{ are}$$

In view of]6.16, considerations about graphs can often be reduced to connected graphs, i.e. those which have just one connectedness component. We will now discuss a necessary and sufficient criterion for finite graphs to be connected. First the necessary one:

Proposition 6.17. If $G = (V, E)$ is a finite connected graph, then $|E| \geq |V| - 1$

Proof. By induction on $n = |V|$ $n = 1$: The graph $G = (\{x\}, \emptyset)$ with 1 vertex is connected by definition and here $|E = 0 \geq 1 - 1 = |V| - 1$.
$n = 2$ Superfluous for the proof but instructive:



If a graph G with two vertices x and y is connected, then $e = \{x, y\}$ must be an edge in fact the only possibly one. Then $G = (\{x, y\}, \{e\})$ and $|E| = 1 \geq 2 - 1 = |V| - 1$.
The induction step $n - 1 \to n$ for $n \geq 2$: If $G = (V, E)$ with $|E| \geq n = |V|$, we are done. So let's assume $|E| < n$. Then we have to show that in fact $|E| = n - 1$. By 6.11, there exists $x \in V$ with $\delta(x) \leq 1$. If we had $\delta(x) = 0$, then x were an isolated vertex and hence (there exists $y \in V$ with $y \neq x$ since $|V| \geq 2$) G were disconnected, which contradicts our assumption. Therefore, $\delta(x) = 1$. Hence there exists only one vertex $y \in V$ which is adjacent to x, and $e = \{x, y\}$ is the only edge which contains x. We now set $G' = (V', E')$ with $V' = V \setminus \{x\}$ and $E' = E \setminus \{e\}$. So that $|V'| = |V| - 1 = n - 1$ and $|E'| = |E - 1| < n - 1 = |V'|$: We observe that the graph G' is also connected: If $u, v \in V'$ are given, there exists a path (see (6.14)) $u = u_1, ..., u_k = v$ in G joining u and v since G is connected. We claim that this path is completely contained in G'. Suppose not. Then there exists $1 < i < k$ with $u_i = x$. But then $\{u_{i-1}, u_i\} = \{u_i, u_{i+1}\} = e$ since e is the only edge containing x, and $u_i = x$ is adjacent to u_{i-1} as well as u_{i+1} Now $e = \{u_{i-1}, u_i\} = \{u_{i-1}, x\}$ implies $u_{i-1} = y$ and $e = \{u_i, u_{i+1}\} = \{x, u_{i+1}\}$ implies $u_{i+1} = y$. Hence the walk $u_1, ..., u_{i-1} = y, u_i = x, u_{i+1} = y, ..., u_k$ contains a repetition, and is, therefore, not a path contradicting our assumption. It follows that $u_i \neq x \forall i$, and hence $u_i \in V' \forall i, \{u_i, u_{i+1}\} \in E' \forall 1 \leq i \leq k - 1$. Since $|V'| = n - 1$ and $|E'| < |V'|$, we can apply the Induction Hypothesis to $G' = (V' E')$ and obtain $|E'| = |V'| - 1 \leftrightarrow |E - 1| = (n - 1) - 1 \leftrightarrow |E| = n - 1 = |V| - 1$. □

Proposition 6.18. If $G = (V, E)$ is a finite graph with $|V| = n$, and $|E| > \frac{(n-1)(n-2)}{2}$, then G is connected.

Proof. If $n = 1$, then G is just a single vertex and hence completed by definition. For $n \geq 2$, the claim is again proved by induction on n using 6.11. Details are left as an Exercise. □

Remark 6.19. For every $n \in \mathbb{N}$ with $n \geq 2$, there exists a graph $G = (V, E)$ with $|V| = n$ and $|E| = \frac{(n-1)(n-2)}{2}$ which is disconnected (:= not connected),

namely the disjoint union of $K_{n-1} = (V_{n-1}, E_{n-1})$, $\Big($ recall that$|E_{n-1} = \binom{n-1}{2} = \frac{(n-1)(n-2)}{2}|\Big)$, and one isolated vertex.

We are now going to make it precise what it means that two graphs are "equivalently the same".

Definition 6.20. Two graphs $G = (V, E)$ and $G' = (V', E')$ are called isomorphic, and we write $G \widetilde{=} G'$, if there exists a bijective map $f : V \to V'$ which satisfies

$$(*)\{x, y\} \in E \to \{f(x), f(y)\} \in E' \quad \forall x, y \in V$$

If that is the case we call f an isomorphism between G and G' by slight abuse of notation, we will also say that $f : G \to G'$ is an isomorphism.

Remark 6.21. (a) If $f : G = (V, E) \to G' = (V', E')$ is an isomorphism, then so is f^{-1}. First note that $f^{-1} : V' \to V$ exists and is bijective since $f : V \to V'$ is bijective. Secondly, for any $x', y' \in V'$, we set $x = f^{-1}(x'), y = f^{-1}(y) \in V$ so that $f(x) = x'$ and $f(y) = y'$. Using $(*)$ for f, we then get the equivalences

$$\{x', y'\} \in E' \leftrightarrow \{f(x), f(y)\} \in E' \underbrace{\leftrightarrow}_{(*)} \{x, y\} \in E \leftrightarrow \{f^{-1}(x), f^{-1}(y')\} \in E$$

and this is precisely the condition $(*)$ for f^{-1}.

(b) Any isomorphism $f : G = (V, E) \to G' = (V', E')$ also induces a bijection from E onto E'. This is easy to verify ((Ex. 16) of HW # 10). Two isomorphic graphs share all essential properties. Below we are listing a couple of them.

Lemma 6.22. If $f : G = (V, E) \to G' = (V', E')$ is an isomorphism of graphs, then

(a) $|V| = |V'|$ (By definition)

(b) $|E| = |E'|$ (By 6.21(b))

(c) If $x_1, ..., x_k$ is a walk/path/cycle in G, then $f(x_1), ..., f(x_k)$ is a walk/path/cycle in G' (again by definition)

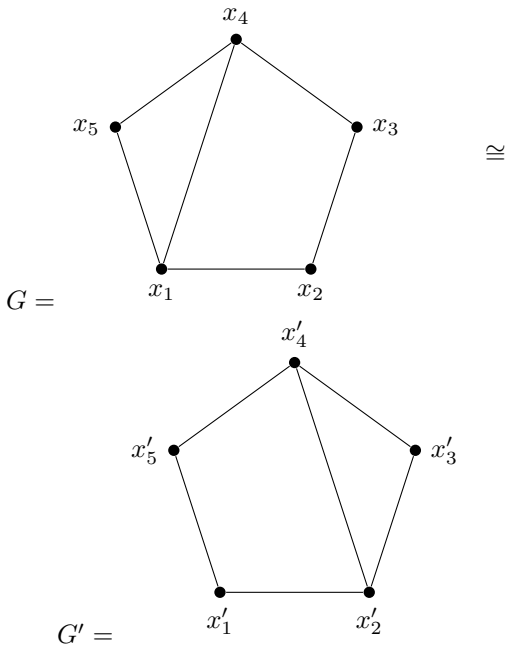
(d) G is connected if and only if G' is connected. More generally, every connectedness component of G is mapped by f bijectively onto a connectedness component of G'. (This follows from (c) $(*)$ in 6.20 and 6.21)

(e) $\delta(f(x)) = \delta(x) \forall x \in V$

(f) If $x, y \in V$ satisfy $x \sim y$, then $f(x), f(y) \in V'$ satisfy $f(x) \sim f(y)$ (by (c) above), and $d(x, y) = d(f(x), f(y))$, where $d(x, y) := \min\{l \in \mathbb{N}_0 |$ a walk of length l joining x and y in $G\}$ (this follows from (c), which applied to f yields $d(f(x), f(y)) \leq d(x, y)$, and applied to f^{-1} yields $d(x, y) \leq d(f(x), f(y))$.

Example 6.23. (a) Two complete graphs $G = (V, E)$ and $G' = (V', E')$ with $|V| = |V'|$ are isomorphic. In fact every bijection $f : V \to V'$ defines an isomorphism.

(b) If $G = (V, E)$ and $G' = (V', E')$ are regular of degree either 0 or 1, then they are isomorphic if and only if $|V| = |V'|$. If G and G' are both connected and regular of degree 2, then they are also isomorphic if and only if $|V| = |V'|$ (Exercise!).
Remark: For $r \geq 3$, two finite connected graphs which are regular of degree r need not be isomorphic (see also Ex. (3) of HW# 10).

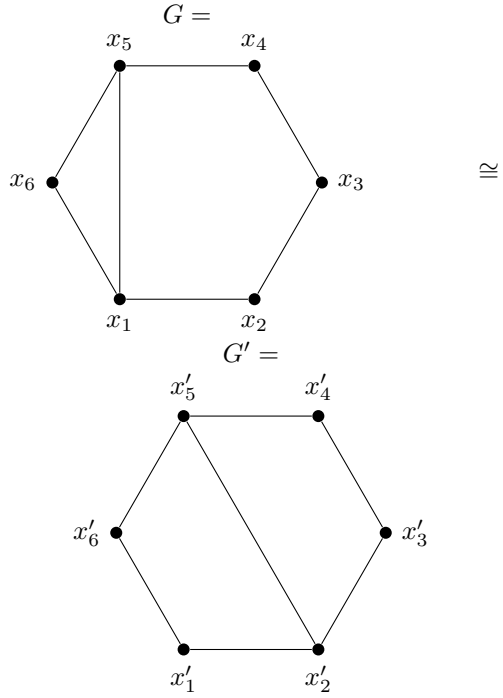


possible isomorphism f:

$$x_1 \mapsto x'_2$$
$$x_2 \mapsto x'_1$$
$$x_3 \mapsto x'_5$$
$$x_4 \mapsto x'_4$$
$$x_5 \mapsto x'_3$$

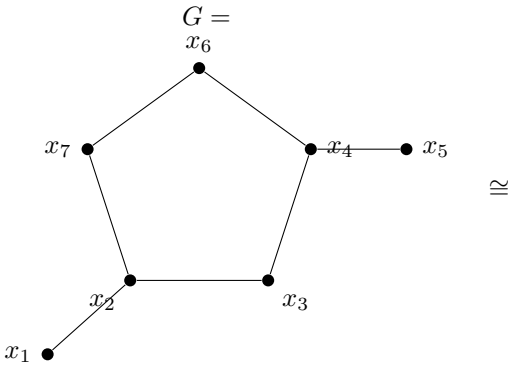
(c)

(d)

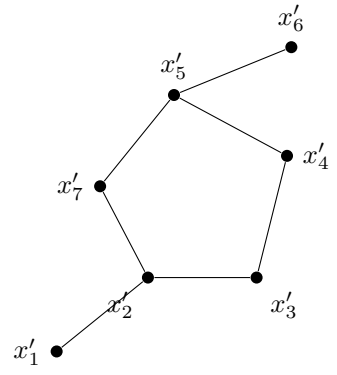


G and G' are not isomorphic since any isomorphism had to map the 3-cycle x_1, x_5, x_6, x_1 of G to the same 3-cycle of G'. However, G' doesn't have any 3 cycles.

(e)



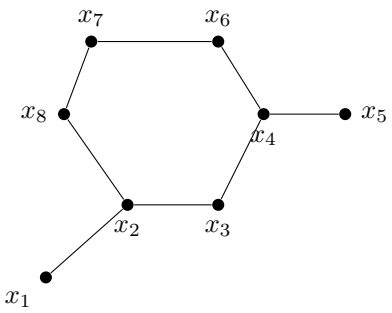
$G' =$



possible isomorphism f :

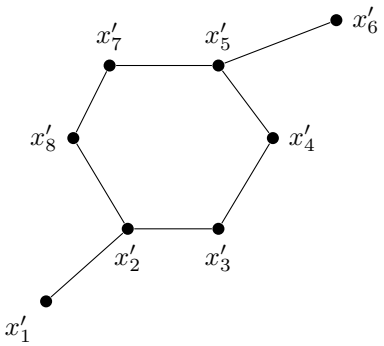
$$\begin{aligned} x_1 &\mapsto x'_1 \\ x_2 &\mapsto x'_2 \\ x_3 &\mapsto x'_7 \\ x_4 &\mapsto x'_5 \\ x_5 &\mapsto x'_6 \\ x_6 &\mapsto x'_4 \\ x_7 &\mapsto x'_3 \end{aligned}$$

(f) $G =$



$\not\cong$

$G' =$



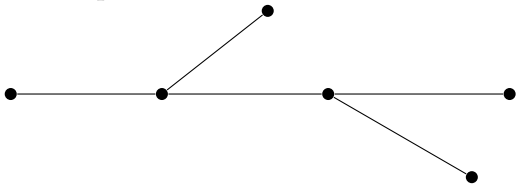
G and G' are not isomorphic: If there existed an isomorphism $f : G \to G'$, it had to map $\{x_1, x_5\}$ (as a subset of V, not an edge) to $\{x'_1, x'_6\}$, since in G the only vertices of degree 1 are x_1, x_5 and in G' the only vertices of degree 1 are x_1, x'_6. But $f(\{x_1, x_5\} = \{x'_1, x'_6\}$ implied by (6.22(f)). $d(x_1, x_5) = d(x'_1, x'_6)$, whereas $d(x_1, x_5) = 4$ and $d(x'_1, x'_6) = 5$.

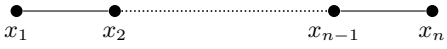
We are now introducing an important class of graphs.

Definition 6.24. A tree is a connected graph which does not have any cycles (of length ≥ 3).

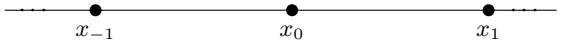
Examples:

Examples

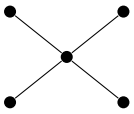




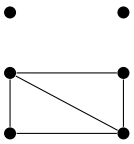
vertices: $x_1, \ldots, x_n$
edges: $\{x_1, x_2\}, \{x_2, x_3\}, \ldots$
$\ldots, \{x_{n-1}, x_n\}$



vertices: $x_i,\ i \in \mathbb{Z}$
edges: $\{x_i, x_{i+1}\},\ i \in \mathbb{Z}$



Non-examples



(2 isolated vertices)

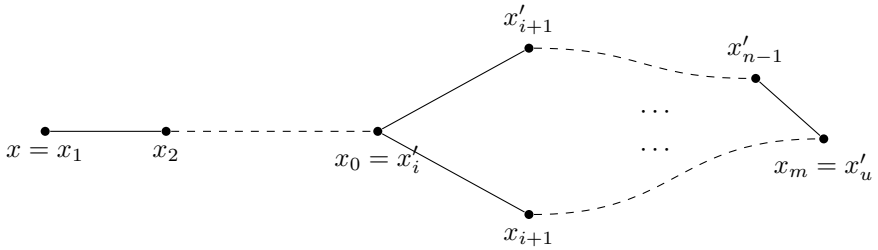
(contains a 3-cycle)

Proposition 6.25. For a (possibly infinite) graph $G = (V, E)$, the following two properties are equivalent:

(a) G is a tree.

(b) For any two vertices $x, y \in V$, there exists a unique path $x = x_1, ..., x_k = y$ which starts in x and ends in y (see 6.14).

Now assume there exists a second path. $x = x'_1, ..., x'_l = y$ and starts in x and ends in y. If it is not identical with the first path, there exists an $i \in \mathbb{N}$ with $1 \leq i < l$ such that $x_j = x'_j \forall 1 \leq j \leq i$ and $x_{i+1} \neq x'_{i+1}$.



(Note that we cannot have $i = k$, meaning $x_i = y$, because then $x'_1 = x_1, ..., x'_i = x_i$ would already be a path from x to y, contradicting $i < l$. Hence $i < k$, and x_{i+1} exists.) Since $x_k = y = x'_l$, there exists $i < m \leq k$ with $m = \min\{j \in \mathbb{N} | j > i \text{ and } x_j \in \{x'_1, ..., x'_l\}\}$ so $x_m = x'_n$ with $1 \leq n \leq l$. But here we must have $n > i$ as well since $m > i \to x_m \neq x_1 = x'_1, ..., x_m \neq x_j = x'_j$. Now consider the closed walk $x_j, x_{j+1}, ..., x_m = x'_n, x'_{n-1}, ..., x'_{i+1}, x'_i = x_i$. We claim that this walk is a cycle. It's clear that $x_j, x_{j+1}, ..., x_m$ are distinct as are $x'_i, x'_{i+1}, ..., x'_n$ since the walks are both parts of paths. But we also have $x_r \neq x'_s \forall i < r < m \forall i < s < n$ in view of the definition of m. Note also that this cycle is of length ≥ 3 since x_i, x_{i+1}, x'_{i+1} are 3 distinct vertices. (We do not, however, include the possibility that $m = i + 1$ or $n = i + 1$.) But the existence of a cycle of length ≥ 3 in a tree is impossible. Hence there cannot exist two different paths in G starting in x and ending in y. $(ii) \to (i)$: Since, for any $x, y \in V$, there exists a path joining them, G is connected. Now assume, by way of contradiction, that there exists a cycle $x_1, ..., x_k = x_1$ of length $k - 1 \geq 3$ in G. Then x_1, x_{k-1} and $x_1, ..., x_{k-1}$ are two different paths in G joining x_1 and x_{k-1}, contradicting (ii). Hence G is a connected graph without cycles of length ≥ 3, i.e. a tree.

Proposition 6.26. Assume that $G = (V, E)$ is a connected graph and $|V| \geq 2 (\leftrightarrow E \neq \emptyset)$

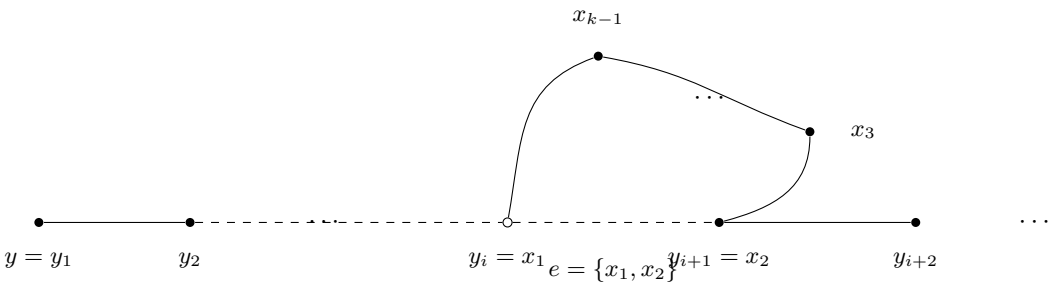
(a) G is a tree if and only if $(V, E \setminus \{e\})$ is disconnected for every edge $e \in E$.

(b) If G is a tree, then $(V, E \setminus \{e\})$ has precisely 2 connectedness components for any edge $e \in E$, which are both trees.

Proof. (a) First assume that G is a tree, and $e = \{x_1, x_2\} \in E$ with vertices $x_1, x_2 \in V$. If $(V, E \setminus \{e\})$ were connected, then there existed a path $p : x_1 = y_1, y_2, ..., y_k = x_2$ in $(V, E \setminus \{e\})$ joining x_1 and x_2 with $\{y_i, y_{i+1}\} \neq e \forall 1 \leq i < k$. But in G, there also exists the short path x_1, x_2 which joins x_1, x_2 and this path is definitely different (of shorter length than) P. However this contradicts 6.25. Hence $(V, E \setminus e)$ must be disconnected.

If on the other hand, $(V, E \setminus e)$ is disconnected for every $e \in E$, then G cannot contain any cycles: Assume, by way of contradiction, that $x_1, x_2, ..., x_k = x_1$ with $k \geq 4$ is a cycle in G. Now consider the edge $e = \{x_1, x_2\}$ and the graph $G' = (V, E \setminus e)$. Then G' would still be connected: Any two vertices $y, z \in V$ can be connected by a path $P := y =$

$y_1, ..., y_l = z \in G$. If this path does not contain e, then it is also a path in G'. But if e occurs in $p, ...$ let's say $e = \{y_i, y_{i+1}\} = \{x_1, x_2\}$, and without loss of generality $y_i = x_1, y_{i+1} = x_2$, then the path $y_1, ..., y_i, y_{i+1}, ..., y_k$ could be replaced with $y_1, ..., y_i = x_1, x_{k-1}, ..., x_2 = y_{i+1}, ..., y_k = z$ in G'. So $(V, E \setminus \{e\})$ would be connected, which contradicts our assumption on G. Hence G is connected without cycles, i.e. a tree.



Now assume that $G = (V, E)$ is a tree and $e = \{x_1, x_2\} \in E$. We claim that $V_1 = [x_1]$ and $V_2 = [x_2]$ are the two (sets of vertices of the two) connectedness components of the graph $G' = (V, E \setminus \{e\})$. It then follows immediately that they are trees (possibly with just one vertex) since they are connected without edges (a cycle in a connectedness component would also be a cycle in G). We have seen in (a) that x_1 and x_2 cannot be connected by a path in G'. Hence $x_1 \not\sim x_2$, and so $[x_1] \neq [x_2]$, which implies $[x_1] \bigcap [x_2] = \emptyset$. On the other hand, $[x_1] \bigcup [x_2] = V$: Let $y \in V$ be any vertex. Since G is connected, there exists a path $p : y = y_1, ..., y_k = x_1$ in G connecting y and x_1. If the edge e is not part of this path p, then it is a path in G', and hence $y \sim x_1$ in $G' \to y \in [x_1] = V_1$. Now suppose that e occurs in p. Since there are no repetitions in p, this is only possible if e is the last edge in this path, i.e. $y_{k-1} = x_2, y_k = x_1, e = \{y_{k-1}, y_k\} = \{x_2, x_1\}$. But then, again since there are no repetitions in P, e does not occur in the path $P' : y = y_1, ..., y_{k-1} = x_2$. This means that $y \sim x_2$ in G', and so $y \in [x_2] = V_2$. This shows that $V = V_1 \bigcup V_2$ (disjoint union), and V_1, V_2 are precisely two connectedness components of G'. □

Corollary 6.27. If $T = (V, E)$ is a tree, and $e_1, ..., e_n$ are n different elements of E, then $T' = (V, E')$ with $E' := E \setminus \{e_1, ..., e_n\}$ has precisely $n + 1$ different connectedness components.

Proof. Exercise! □

Theorem 6.28. A finite connected graph $G = (V, E)$ is a tree if and only if $|E| = |V| - 1$.

Proof. Let's first assume that the finite connected graph $G = (V, E)$ satisfies $|V| - 1 = |E|$. Then for any $e \in E$, $|E \setminus \{e\}| = |E| - 1 = |V| - 2 < |V| - 1$. Hence by Proposition 6.17 the graph $(V, E \setminus \{e\})$ is not connected. Therefore, Proposition 6.26(a) implies that G is a tree. This is of course also true if $|V| = 1$ and $|E| = 0$. Now assume that G is a tree with $n = |V| \geq 2$. (Again, the case

$|V| = 1$ and $|E| = 0$ is trivial.) Then $|E| \neq \emptyset$, there exists $e = \{x_1, x_2\} \in E$, and we can prove the claim by induction on n as follows: By Proposition 6.26(b) $G' = (V, E \setminus \{e\})$ has two connectedness components $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, where $V_i = [x_i]$ for $i = 1, 2$. Set $n_i = |V_i|$ for $i = 1, 2$. Since G_i is non-empty, $n_i \geq 1$ for $i = 1, 2$. And since V is the disjoint union of V_1 and V_2, $n = n_1 + n_2$. Hence $n_i < n$ for $i = 1, 2$. And since G_1 and G_2 are finite trees, our induction hypothesis implies $|E_j| = |V_j| - 1$ for $i = 1, 2$. But we also have that $E \setminus \{e\}$ is the disjoint union of E_1 and E_2 see (6.16). Therefore:

$$|E \setminus \{e\}| = |E_1| + |E_2| = (|V_1| - 1) + (|V_2| - 1) = n_1 + n_2 - 2 = n - 2$$

This implies

$$|E| = |E \setminus \{e\}| + 1 = (n - 2) + 1 = n - 1$$

□

Proposition and Definition 6.29. If $G = (V, E)$ is a finite connected graph, then there exists a subset $E' \subseteq E$ such that $T = (V, E')$ is a tree. T is called a spanning tree for G.

Proof. Set $n = |V|$ and $m = |E|$ by 6.17, $m \geq n - 1$. The claim will be proved by induction on m.

$m = n - 1 \to G$ is itself a tree by 6.28, and we can set $E' = E$

$m \geq n$: Again by 6.28, G is not a tree. Hence by 6.26(a)

There exists an edge $e \in E$ such that $G' = (V, E \setminus \{e\})$ is still connected. Since $|E \setminus \{e\}| = |E| - 1$, we can again apply the induction hypothesis to G'. It follows that there exists $E' \subseteq E \setminus \{e\} \subseteq E$ such that $T = (V, E')$ is a tree. □

Remark 6.30. (a) Interesting applications of spanning trees arise if we are given a weight function on the edges of the finite connected graph $G = (V, E)$. This leads to the M.S.T. (Minimal spanning tree) problem, see section 16.3 in [B].

(b) Every infinite connected graph also admits a spanning tree. However, a different proof is needed here. Instead of induction one uses "Zorn's Lemma".

7 Linear Recursions and Generating Functions

In mathematics, sequences $(a_n)_{n \in \mathbb{N}}$, are often defined by a recursion, i.e. a rule in which determines a_n if $a_{n-1}, ..., a_1$ are known.

Example 7.1. The sequence of derangement numbers $(d_n)_{n \in \mathbb{N}}$ satisfies the following recursions:

(a) $d_n = nd_{n-1}+(-1)^n$ for $n \geq 2$. (This follows immediately from Proposition 4.9.) Also $d_1 = 0$. Hence

$$d_2 = 2{\cdot}0{+}1 = 1, \quad d_3 = 3{\cdot}1{-}1 = 2, \quad d_4 = 4{\cdot}2{+}1 = 9, \quad d_5 = 5{\cdot}9{-}1 = 44, \quad d_6 = 6{\cdot}44{+}1 = 265, \ \ldots$$

(b) $d_n = (n-1)\big(d_{n-1} + d_{n-2}\big)$ for all $n \geq 3$, and $d_1 = 0$, $d_2 = 1$. This follows immediately from applying (a) for n and for $n-1$. Hence

$$d_3 = 2(1+0) = 2, \quad d_4 = 3(2+1) = 9, \quad d_5 = 4(9+2) = 44, \ \ldots$$

Note the following: If the recursion for a_n only involves a_{n-1}, then we need a_1 as a start recover the whole sequence $(a_n)_{n\in\mathbb{N}}$. If the recursion for a_n involves a_{n-1} and a_{n-2}, we need the values of a_1 and a_2 as a start. And so on.

Example 7.2. The Fibonacci numbers $(f_n)_{n\in\mathbb{N}}$ are defined by the recursion $(*) f_n = f_{n-1} + f_{n-2} \forall n \geq 3$. Together with $f_1 = 1, f_2 = 1 \to f_3 = 2, f_4 = 3, f_5 = 5, f_6 = 8, f_7 = 13, ...$

This is an example of a linear recursion, since the coefficients of f_{n-1} and f_{n-2} in $(*)$ above are constant (=1). By contrast, the recursions in 7.1(a) and 7.1(b) are not linear since the coefficients are non-constant factors of n; n in 7.1(a) and $n-1$ twice in 7.1(b). The recursions in 7.1 and 7.2 are homogeneous since they do not involve any constant summands. For the derangement numbers, we also have a "closed formula" for d_n which does not depend on the values d_i for $i < n$. Is there also a closed formula for the fibonacci numbers (and more generally for sequences $(a_n)_{n\in\mathbb{N}}$ defined by homogeneous linear recursions?)

Proposition 7.3. If $(f_n)_{n\in\mathbb{N}}$ is the Fibonacci sequence then

$$f_n = \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^n - \left(\frac{1-\sqrt{5}}{2} \right)^n \right) \forall n \in \mathbb{N}$$

Proof. By induction on n. Obviously, we want to use the recursion $f_n = f_{n-1} + f_{n-2} \forall n \geq 3$. But before this, we need to check the claim for $n = 1$ and $n = 2$.

$$n = 1: \quad \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right) - \left(\frac{1-\sqrt{5}}{2} \right) \right) = \frac{1}{\sqrt{5}} \cdot \sqrt{5} = 1$$

$$n = 2: \quad \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^2 - \left(\frac{1-\sqrt{5}}{2} \right)^2 \right)$$

$$= \frac{1}{4 \cdot \sqrt{5}} (1 + 2\sqrt{5} + 5 - (1 - 2\sqrt{5} + 5))$$

$$= \frac{1}{4 \cdot \sqrt{5}} 4 \cdot \sqrt{5} = 1$$

$n \geq 3$ We first observe that the roots of the quadratic equation:

$$x^2 = x + 1 \leftrightarrow x^2 - x + 1 = 0 \text{ are}$$

$$\alpha_1 = \frac{1}{2} + \sqrt{\frac{1}{4} + 1} = \frac{1}{2} + \sqrt{\frac{5}{4}} = \frac{1 + \sqrt{5}}{2}$$

$$\alpha_2 = \frac{1}{2} - \sqrt{\frac{1}{4} + 1} = \frac{1}{2} - \sqrt{\frac{5}{4}} = \frac{1 - \sqrt{5}}{2}$$

We have to prove $f_n = \frac{1}{\sqrt{5}}(\alpha_1^n - \alpha_2^n)$. Using the Induction Hypothesis, we get

$$f_{n-1} = \frac{1}{\sqrt{5}} \left(\alpha_1^{n-1} - \alpha_2^{n-2}\right)$$

and,

$$f_{n-2} = \frac{1}{\sqrt{5}} \left(\alpha_1^{n-2} - \alpha_2^{n-2}\right)$$

together these imply

$$f_n = f_{n-1} + f_{n-2} = \frac{1}{\sqrt{5}} \left(\alpha_1^{n-1} - \alpha_2^{n-2} + \alpha_1^{n-2} - \alpha_2^{n-2}\right)$$

$$= \frac{1}{\sqrt{5}} \left(\alpha_1^{n-2} (\alpha_1 + 1) - \alpha_2^{n-2} (\alpha_2 + 1)\right)$$

Since $\alpha_i^2 = \alpha_i + 1$ for $i = 1, 2$

$$\frac{1}{\sqrt{5}} (\alpha_1^n - \alpha_2^n)$$

as desired. □

Definition 7.4. We say that the sequence $(u_n)_{n \in \mathbb{N}}$ satisfies a homogeneous linear recursion of degree $k \geq 1$ if there exist $a_k \neq 0, a_1, ..., a_k \in \mathbb{R}$ such that $u_n = a_1 u_{n-1} + \cdots + a_k u_{n-k} \forall n > k (*)$. The sequence $(u_n)_{n \in \mathbb{N}}$ is then determined by the recursion $(*)$ and k given initial values $u_1 = c_1, ..., u_k = c_k$. The auxiliary or (characteristic) equation associated to $(*)$ is $x^k = a_1 x^{k-1} + \cdots + a_{k-1} + a_k$

$$\leftrightarrow x^k - a_1 x^k - \cdots - a_k x - a_k = 0$$

$$\leftrightarrow p(x) = 0, \text{ where } p(x) = x^k - \sum_{i=1}^{k} a_i x^{k-i}$$

is the characteristic or auxiliary polynomial to the recursion $(*)$ If $(f_n)_{n \in \mathbb{N}}$ is the fibonacci sequence, then the recursion $f_n = f_{n-1} + f_{n-2}$ for $n \geq 3$ is linear of degree 2 with $a_1 = a_2 = 1$ and $c_1 = c_2 = 1$. The auxiliary equation $x^2 = x + 1$ was used in proposition 7.3, and the strategy used there will be generalized soon. But let's first discuss the initial case $k = 1$.

Example 7.5. If $(u_n)_{n\in\mathbb{N}}$ satisfies $u_n = a_1 u_{n-1} \forall n \geq 2$ and $u_1 = c_1$, then $u_n = c_1 a^{n-1} \forall n \in \mathbb{N}$, as is easily shown by induction on n. Note that in this case a_1 is the unique solution of $x - a_1 = 0$. We are now turning to the case $k = 2$, an example of which is the fibonacci sequence.

Proposition 7.6. If $(u_n)_{n\in\mathbb{N}}$ satisfies the homogeneous linear recursion $u_n = a_1 u_{n-1} + a_2 u_{n-2}$ of degree $k = 2$ with $a_2 \neq 0$ and initial value values $u_1 = c_1$ and $u_2 = c_2$, and if the polynomial $p(x) = x^2 - a_1 x - a_2$ has two distinct roots α, β, then: $u_n = A\alpha^n + B\beta^n \forall n \in \mathbb{N}$, where A and B satisfy the equations $A\alpha + B\beta = c_1$ and $A\alpha^2 + B\beta^2 = c_2$. $(*)$

Proof. We first note that the system of two linear equations $(*)$ really has two uniquely determined solutions A and B: To begin with, 0 is not a root of $p(x)$ since $a_2 \neq 0$. Hence $\alpha \neq 0, \beta \neq 0$. And since $\alpha \neq \beta$, also $\alpha^2 - \alpha\beta = \alpha(\alpha - \beta) \neq 0$. Now $A\alpha + B\beta = c_1$ is equivalent to $B = \frac{c_1 - A\alpha}{\beta}$ and then the second equation in $(*)$ yields

$$A\alpha^2 + \beta^2 \frac{c_1 - A\alpha}{\beta} = c_2 \leftrightarrow A\alpha^2 + (c_1 - A\alpha)\beta = c_2 \leftrightarrow$$

$$A(\alpha^2 - \alpha\beta) + c_1\beta = c_2 \leftrightarrow A = \frac{c_2 - c_1\beta}{\alpha^2 - \alpha\beta} \to B = \frac{c_2 - c_1\alpha}{\beta^2 - \alpha\beta}$$

[intermediate example](fibonacci sequence): $a_1 = a_2 = 1, c_1 = c_2 = 1$,

$$\alpha = \frac{1+\sqrt{5}}{2}, \beta = \frac{1-\sqrt{5}}{2}, \alpha - \beta = \sqrt{5}, \alpha + \beta = 1$$

$$\frac{c_2 - c_1\beta}{\alpha^2 - \alpha\beta} = \frac{1-\beta}{\alpha(\alpha-\beta)} = \frac{1}{\alpha - \beta} = \frac{1}{\sqrt{5}}$$

and

$$\frac{c_2 - c_1\alpha}{\beta^2 - \alpha\beta} = \frac{1-\alpha}{\beta(\beta-\alpha)} = \frac{1}{\beta - \alpha} = -\frac{1}{\sqrt{5}}$$

Proof. It is now easy to prove the claim $u_n = A\alpha^n + B\beta^n$ by induction on n. For $n = 1, n = 2$, this equation is satisfied by the the choice of A and B. And for $n \geq 3$, we get, using the I.H. for $n-1$ and $n-2$

$$u_n = a_1 u_{n-1} + a_2 u_{n-2} = \alpha_1(A\alpha^{n-1} + B\beta^{n-1}) + \alpha_2(A\alpha^{n-2} + B\beta^{n-2})$$
$$= A(a_1\alpha^{n-1} + a_2\alpha^{n-2}) + B(a_1\beta^{n-1} a_2\beta^{n-2})$$
$$= A\alpha^{n-2}(a_1\alpha + a_2) + B\beta^{n-2}(a_1\beta + a_2)$$
$$= A\alpha^{n-2}\alpha^2 + B\beta^{n-2}\beta^2 (\text{ since } \alpha, \beta \text{ are roots of } x^2 - a_1 x + a_2)$$
$$= A\alpha^n + B\beta^n$$

□

Remark 7.7. (a) Even if $a_1, a_2, c_1, c_2 \in \mathbb{R}$ it's possible that $\alpha, \beta \in \mathbb{C} \setminus \{\mathbb{R}\}$.

(b) The formula $u_n = A\alpha^n + B\beta^n \forall n \in \mathbb{N}$ can of course also be written in the form

$$(*) u_n = A'\alpha^{n-1} + B'\beta^{n-1}$$

With $A' = A\alpha, B' = B\beta$. The advantage of $(*)$ is that A' and B' are slightly easier to compute than A, B since they satisfy the equations $A' + B' = c_1 \leftrightarrow B' = c_1 - A'$ and $A'\alpha + B'\beta = c_2 \leftrightarrow A'\alpha + (c_1 - A')\beta = c_2 \leftrightarrow A'(\alpha - \beta) + c_1\beta = c_2 \leftrightarrow A' = \frac{c_2 - c_1\beta}{\alpha - \beta}$ ($\alpha \neq \beta$ is of course again needed) and hence $B' = \frac{c_1\alpha - c_2}{\alpha - \beta} = \frac{c_2 - c_1\alpha}{\beta - \alpha}$

Example 7.8. Consider the sequence $(u_n)_{n\in\mathbb{N}}$ with the linear recursion $(*)$ $u_n = u_{n-1} - u_{n-2} \forall n \geq 3$ and initial values $u_1 = u_2 = 1$. The auxiliary equation is $x^2 = x - 1 \leftrightarrow x^2 - x + 1 = 0$, and its roots are $\alpha = \frac{1}{2} + \sqrt{\frac{1}{4} - 1} = \frac{1+\sqrt{-3}}{2}$ and $\beta = \frac{1}{2} - \sqrt{\frac{1}{4} - 1} = \frac{1-\sqrt{3}}{2}$. Note that $\alpha + \beta = 1$ and $\alpha - \beta = \sqrt{-3}$. The equations $A' + B' = 1$ and $A'\alpha + B'\beta = 1$ have the unique solution $A' = \frac{1-\beta}{\alpha-\beta} = \frac{\alpha}{\alpha-\beta} = \frac{\alpha}{\sqrt{3}}$

$$= \frac{1 + \sqrt{-3}}{2\sqrt{-3}} = \frac{\sqrt{-3} - 3}{-6} = \frac{3 - \sqrt{-3}}{6}$$

and

$$B' = \frac{1 - \alpha}{\beta - \alpha} = \frac{\beta}{-\sqrt{-3}} = \frac{1 - \sqrt{-3}}{-2\sqrt{-3}} = \frac{3 + \sqrt{-3}}{6}$$

It follows from 7.6 and 7.7 that

$$u_n = \frac{3 - \sqrt{3}}{6}\left(\frac{1 + \sqrt{-3}}{2}\right)^{n-1} + \frac{3 + \sqrt{-3}}{6}\left(\frac{1 - \sqrt{3}}{2}\right)^{n-1} \forall n \in \mathbb{N}$$

Which can also be written in the form

$$(**) u_n = \frac{1}{\sqrt{-3}}\left(\left(\frac{1 + \sqrt{-3}}{2}\right)^n - \left(\frac{1 - \sqrt{-3}}{2}\right)^n\right) \forall n \in \mathbb{N}$$

Since $\frac{1+\sqrt{-3}}{2\cdot\sqrt{-3}} = \frac{3-\sqrt{-3}}{6}$ and $\frac{1-\sqrt{-3}}{-2\sqrt{-3}} = \frac{3+\sqrt{-3}}{6}$, i.e., $A = \frac{1}{\sqrt{-3}}$ and $B = -\frac{1}{\sqrt{3}}$ in the notation of 7.6. A remarkable feature of the sequence (u_n) is that it is periodic, which (independently of the initial values) follows from $(*)$:

$$u_n = u_{n-1} - u_{n-2} = (u_{n-2} - u_{n-3}) - u_{n-2} = -u_{n-3} \forall n \geq 4$$

$$\rightarrow u_n = -(-u_{n-6)} = u_{n-6} \forall n \geq 7$$

so that

$$(u_n) = (u_1, u_2, u_3, u_4, u_5, u_6, u_1, u_2, u_3, u_4, u_5, u_6, ...)$$

The periodicity of (u_n) can also be deduced from $(**)$ as follows: Note that $\alpha = \frac{1}{2} + \frac{\sqrt{3}}{2}i = \cos\frac{\pi}{3} + i\sin\frac{\pi}{3} = e^{\frac{2\pi}{6}}$ is a (primitive) 6th root of unity, i.e., $\alpha^6 = 1$ (and $\alpha^m \neq 1$ for$1 \leq m \leq 5$), and $\beta = \frac{1}{2} - \frac{\sqrt{3}}{2}i = \bar{\alpha} = \alpha^{-1}$ since $|\alpha| = 1$.

Hence $(**)$ can also be written as $u_n = \frac{1}{\sqrt{-3}}(\alpha^n - \alpha^{-n})$ which in view of $\alpha^6 = 1$ implies for all $n \geq 7$

$$u_n = \frac{1}{\sqrt{-3}}(\alpha^{n-6}\alpha^6 - \alpha^{-(n-6)} \cdot \alpha^{-6}) = \frac{1}{\sqrt{-3}}(\alpha^{6n} - \alpha^{-(n-6)}) = u_{n-6}$$

Remark 7.9. (a) Example 7.8 shows that the same sequence $(u_n)_{n\in\mathbb{N}}$ can satisfy homogeneous linear recursions of different degrees, e.g., 2,3,6. But this shouldn't be a surprise since the argument on which 7.6 is based is precisely that every root α of the auxiliary equation $x^2 - a_1x - a_2 = 0$ yields a sequence $(v_n = \alpha^n)_{n\in\mathbb{N}}$ which satisfies the recursion $v_n = a_1v_{n-1} + a_2v_{n-2} \forall n \geq 3$ of degree 2 but also the recursion $v_n = \alpha v_{n-1}$ of degree 1.

(b) If $(u_n)_{n\in\mathbb{N}}$ is a sequence that satisfies a homogeneous linear recursion of degree 2, but no linear recursion of degree 1, then one can show that (u_n) is periodic if and only if the corresponding auxiliary equation has two distinct roots in $\mathbb{C}$, both of which are roots of unity. The proof of the "if" direction is easy (similar to the argument given at the end of 7.8), but the "only if" direction is not. In Exercise (4) of Homework #12, you should establish the weaker necessary condition for periodicity, namely that $|\alpha| = |\beta|$, where α and β are the two roots of the auxiliary equation.

Questions

1. How can we find a closed formula for (u_n) in the situation described in 7.6 if $p(x) = x^2 - a_1x - a_2$ has only one (double) root?

2. What can we say about homogeneous linear recursions of degree ≥ 3? The first question can be answered relatively quickly.

Proposition 7.10. If $(u_n)_{n\in\mathbb{N}}$ is as in 7.6 but $p(x)$ has only one (double) root α, i.e. $p(x) = x^2 - a_1x - a_2 = (x - \alpha)^2$, then

$$u_n = A\alpha^{n-1} + Bn\alpha^{n-1} \forall n \in \mathbb{N}$$

, where A and B are determined by equations $A + B = c_1$ and $(A + 2B)\alpha = c_2$.

Proof. Let's start by verifying that the last two equations have a unique solution for A and B if c_1, c_2 and α are given. $A + B = c_1, B = c_1 - A$.

$$(A + 2B)\alpha = c_2 \leftrightarrow A + 2(c_1 - A) = \frac{c_2}{\alpha} \leftrightarrow -A = \frac{c_2}{\alpha} - 2c_1$$

$$B = \frac{c_2}{\alpha} - c_1 \leftrightarrow A = 2c_1 - \frac{c_2}{\alpha}$$

Note that $\alpha \neq 0$ since $\alpha_2 \neq 0$ by assumption. We want to prove the formula for u_n by induction on n.

$$n = 1, 2 :$$

The equations $A + B = c_1 = u_1$ and $A\alpha + 2B\alpha = c_2 = u_2$ are satisfied by choice of A and B.

$$n \geq 3 :$$

In order to prepare for the induction step, we first observe that $x^2 - a_1x - a_2 = (x-\alpha)^2 = x^2 - 2\alpha x + \alpha^2$ which implies $a_1 = 2\alpha$ and $a_2 = -\alpha^2$. From this, we deduce $\forall n \in \mathbb{N}$, $(n-1)\alpha a_1 + (n-2)\alpha_2 = (n-1)\alpha(2\alpha) - (n-2)\alpha^2 = 2n\alpha^2 - 2\alpha^2 - n\alpha^2 + 2\alpha^2 = n\alpha^2$, which implies

$$(*) n\alpha^{n-1} = a_1(n-1)\alpha^{n-2} + a_2(n-2)\alpha^{n-3} \forall n \geq 3$$

Besides, we also have (as in 7.6)

$$\alpha^2 = a_1\alpha + a_2 \rightarrow (**)\alpha^{n-1} = a_1\alpha^{n-2} + a_2\alpha^{n-3} \forall n \geq 3$$

Using $(*)$ and $(**)$, and the I.H., we now obtain $\forall n \geq 3$:

$$u_n = a_1u_{n-1} + a_2u_{n-2} \underbrace{=}_{\text{I.H.}} a_1(A\alpha^{n-2} + B(n-1)\alpha^{n-2}) + a_2(A\alpha^{n-3} + B(n-2)\alpha^{n-3})$$

$$= A(a_1\alpha^{n-2} + a_2\alpha^{n-3}) + B(a_1(n-1)\alpha^{n-2} + a_2(n-2)\alpha^{n-3})$$

$$= A\alpha^{n-1} + Bn\alpha^{n-1}$$

By $(**)$ and $(*)$ $\square$

Example 7.11. Let the sequence $(u_n)_{n\in\mathbb{N}}$ be recursively defined by $u_n = 4(u_{n-1} - u_{n-2})$ with initial values $u_1 = u_2 = 1$. Then the auxiliary polynomial $p(x) = x^2 - 4x + 4 = (x-2)^2$ has the double root $\alpha = 2$. The equations $A + B = 1$ and $(A + 2B)\alpha = 1$ lead to $B = 1 - A$, $2(A + 2(1-A)) = 1 \leftrightarrow B = 1 - A, -A + 2 = \frac{1}{2} \leftrightarrow A = \frac{3}{2}, B = \frac{-1}{2}$. Applying 7.10, we get the following formula for u_n:

$$u_n = \frac{3}{2}2^{n-1} - \frac{1}{2}n2^{n-1} = 2^{n-3}(3-n) \forall n \in \mathbb{N}$$

We may check (for $n \geq 3$)

$$4(u_{n-1} - u_{n-2}) = 4(2^{n-3}(3-(n+1)) - 2^{n-4}(3-(n-2))$$

$$= 2^{n-2}(2(3-n+1) - 3 + n - 2) = 2^{n-2}(3-n) = u_n$$

as desired

Question 2 above is more difficult. Let's first discuss the case where the auxiliary polynomial has k distinct roots, if (u_n) satisfies a linear recursion of degree k.

Theorem 7.12. If $(u_n)_{n\in\mathbb{N}}$ satisfies the homogeneous linear recursion $u_n = a_1u_{n-1} + \cdots + a_ku_{n-k}$ of degree k with $a_k \neq 0$ and initial values $u_1 = c_1, ..., u_k = c_k$ and if the polynomial $p(x) = x^k - a_1x^{k-1} - \cdots - a_k$ has k distinct roots $\alpha_1, ..., \alpha_k$, then $u_n = \sum_{i=1}^{k} A_i\alpha_i^{n-1} \forall n \in \mathbb{N}$ where A_i are uniquely determined by the system of linear equations

$$(*) \sum_{i=1}^{k} \alpha_i^{i-1}A_i = c_i \forall 1 \leq i \leq k$$

sketch of proof: $(*)$ is an inhomogeneous system of linear equations for the k indeterminates $A_1, ..., A_k$ in matrix notation:

$$(*) \to \begin{pmatrix} 1 & \cdots & 1 \\ \alpha_1 & \cdots & \alpha_k \\ \vdots & \cdots & \vdots \\ \alpha_1^{k-1} & \cdots & \alpha_k^{k-1} \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_k \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}$$

And this solution has a unique solution since the Vandermonde Determinant $\det(m) = \Pi_{1 \leq i < j \leq k}(\alpha_j - \alpha_i)$ is $\neq 0$. (Here we are using that $\alpha_1, ..., \alpha_k$ are k distinct numbers.) Now since $\alpha_1, ..., \alpha_k$ are roots of $p(x)$, we obtain $\forall 1 \leq i \leq k$:

$$(**)\alpha_i^k = \sum_{j=1}^{k} a_j \alpha_j^{k-j} \to \alpha_i^{n-1} = \sum_{j=1}^{k} a_j \alpha_j^{n-j-1} \forall n \geq k+1$$

The formula $(***)$ $u_n = \sum_{i=1}^{k} A_i \alpha_i^{n-1}$ can now be proved by induction on n:

$$n \geq k : u_n = c_n = \sum_{i=1}^{k} A_i \alpha_i^{n-1}$$

follows from (*)

$$n \geq k+1 : \text{ We use the I.H applied to } u_{n-1}, ..., u_{n-k}, \text{yielding}$$

$$u_{n-j} = \sum_{i=1}^{k} A_i \alpha_i^{n-j-1} \forall 1 \leq j \leq k \to$$

$$u_n = \sum_{j=1}^{k} a_j u_{n-j} = \sum_{j=1}^{k} a_j \sum_{i=1}^{k} A_i \alpha_i^{n-j} = \sum_{i=1}^{k} A_i \alpha_i^n \left(\sum_{j=1}^{k} a_j \alpha_i^{-j} \right).$$

$$\sum_{i=1}^{k} A_i \sum_{j=1}^{k} a_j \alpha_j^{n-j-1} \underbrace{=}_{(**)} \sum_{i=1}^{k} A_i \alpha_i^{n-1}$$

Example 7.13. (continuation of 7.8)
Let $(u_n)_{n \in \mathbb{N}}$ be the recursion $u_n = u_{n-6} \forall n \geq 7$ of degree 6. With initial values $c_1 = 1, c_2 = 1, c_3 = 0, c_4 = -1, c_5 = -1, c_6 = 0$. This is the same sequence as in 7.8 but determined differently. Here the auxilary polynomial is

$$p(x) = x^6 - 1 = (x^3 - 1)(x^3 + 1) = (x-1)(x^2 + x + 1)(x+1)(x^2 - x + 1)$$

The roots of $p(x)$ are $\alpha_i = \alpha^i$ for $1 \leq i \leq 6$, where $\alpha = e^{\frac{2\pi i}{6}} = \frac{1}{2} + \frac{\sqrt{3}}{2} i$. As we noted in 7.8, $\alpha^6 = 1$ and is easy to check that with α also α^i is a root of $p(x)$ since $x^6 - 1 = 0 \leftrightarrow \alpha^6 = 1$ and $(\alpha^i)^6 = (\alpha^6)^i = 1^i = 1$. It is also easy to check that $\alpha_1, ..., \alpha_6$ are 6 different (composite) numbers, so that these must be

precisely the roots of $p(x)$ (a polynomial of degree 6 cannot have more than 6 roots). In other words,

$$p(x) = x^6 - 1 = \Pi_{i=1}^{6}(x - \alpha^i)$$

So 7.12 tells us that the closed formula for u_n looks like:

$$u_n = \sum_{i=1}^{6} A_i \alpha_i^{n-1}$$

where the A_i are determined by

$$(*) \sum_{i=1}^{6} A_i \alpha_i^{j-1} = c_j \quad \text{for } 1 \leq j \leq 6$$

Now it would be tedious to solve this system of equations from scratch. But as remarked in the proof of 7.12 (*) has a unique solution. And we already know this solution from 7.8. Note that

$$a_5 = \alpha^5 = \alpha^{-1} = \bar{\alpha} = \frac{1}{2} - \frac{\sqrt{3}}{2} i = \beta$$

with the notation from 7.8. And hence we get from 7.8 that:

$$A_1 + A_5 = 1$$

and $A_1\alpha_1 + A_5\alpha_5 = 1$ for $A_1 = \frac{3-\sqrt{3}}{6}$ and $A_5 = \frac{3+\sqrt{-3}}{6}$. We also saw in 7.8 that the sequence $(v_n)_{n\in\mathbb{N}})$ defined by $v_n = A_1\alpha_1^{n-1} + A_5\alpha_5^{n-1}$ satisfies $v_n = v_{n-6} \forall n \geq 7$ and $v_i = c_i \forall 1 \leq j \leq 6$. But these are precisely the conditions which determine our current sequence $(u_n)_{n\in\mathbb{N}}$. Hence we get for all $n \in \mathbb{N}$,

$$u_n = v_n = \sum_{i=1}^{6} A_i \alpha_i^{n-1}$$

with $A_2 = A_3 = A_4 = A_6 = 0$.

Example 7.14. Find a closed formula for the sequence $(u_n)_{n\in\mathbb{N}}$ which is a recursively defined by $u_n = 2u_{n-1} + u_{n-2} - 2u_{n-3}$ with initial values $c_1 = 1, c_2 = 2c_3 = 0$. We first solve the auxilary equation:

$$x^3 - 2x^2 - x + 2 = 0$$

which obviously has $\alpha = 1$ as one of its roots. After dividing the auxilary polynomial by $(x - 1)$, we get:

$$x^3 - 2x^2 - x + 2/(x - 1) = x^2 - x - 2$$

Now $x^2 - x - 2 = (x+1)(x-2)$, so that $\alpha_2 = -1$ and $\alpha_3 = 2$, are the remaining two roots of the auxiliary polynomial. The system of linear equations determining A_1, A_2, A_3 is:

$$\begin{cases} A_1 + A_2 + A_3 = 1, \\ A_1 - A_2 + 2A_3 = 2, \\ A_1 + A_2 + 4A_3 = 0 \end{cases} \iff \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 2 \\ 1 & 1 & 4 \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}.$$

The unique solution of this system of linear equations is

$$A_1 = 2, A_2 = -\frac{2}{3}, A_3 = -\frac{1}{3}$$

(This is standard linear algebra "reduced row form") It now follows from 7.12 that

$$u_n = 2 - \frac{2}{3}(-1)^{n-1} - \frac{1}{3}2^{n-1} \forall n \in \mathbb{N}$$

Things get again more complicated if the auxiliary polynomial has multiple roots. We will just state here how the closed formula can be obtained in this case.

Theorem 7.15. If a sequence $(u_n)_{n\in\mathbb{N}}$ satisfies the homogeneous linear equation $u_n = a_1 u_{n-1} + \cdots + a_k u_{n-k} (a_k \neq 0)$ of degree k, and if the auxiliary polynomial $p(x) = x^k - \sum_{i=1}^{k} a_i x^{k-i}$ has r distinct roots, $\alpha_1, ..., \alpha_r$ of respective multiplicities $m_1, ..., m_r$, i.e. $p(x) = \Pi_{i=1}^{r}(x - \alpha_i)_i^m$, then (u_n) is a linear combination of the sequences

$$(n^{j_i}\alpha_j^{n-1} \text{ for } 1 \leq i \leq r, 0 \leq j_i \leq m_j - 1 \forall i$$

(so these $m_1 + m_2 + \cdots + m_r = n$ sequences of this type), with coefficients which are uniquely determined by the initial values $c_i = u_i$ for $1 \leq i \leq k$.

We will now study an alternative way to derive formulas for u_n if the sequence $(u_n)_{n\in\mathbb{N}}$ satisfies a homogeneous linear recursion. This approach uses generating functions, which are an important tool in combinatorics (not limited to recursions). In this context it is common to start the indexing with 0 instead of 1, hence using $\mathbb{N}_0$ and not $\mathbb{N}$ as the index set.

Definition 7.16. The generating function associated to a sequence $(u_n)_{n\in\mathbb{N}_0}$ is the power series $U(x) = \sum_{n=0}^{\infty} u_n x^n$.

Remark 7.17. Addition and multiplication of power series. If $a = \sum_{n=0}^{\infty} a_n x^n$ and $b = \sum_{n=0}^{\infty} b_n x^n$, then $a + b := \sum_{n=0}^{\infty}(a_n + b_n)x_n$. And we define

$$a \cdot b = \sum_{n=0}^{\infty} c_n x^n$$

with

$$c_n = \sum_{m=0}^{n} a_m b_{n-m} \forall n \in \mathbb{N}_0$$

So in particular $c_0 = a_0 b_0 \to a \cdot b \neq 1$ if $a_0 = 0$ or $b_0 = 0$. Note: if $a_n = 0 \forall n > d$ and $a_d \neq 0$, for some $d \in \mathbb{N}_0$, then a is a polynomial of degree d.

$\{\sum_{n=0}^{\infty} a_n x^n | a_n \in \mathbb{R} \forall n \in \mathbb{N}_0\}$ together with addition and multiplication as defined above is a commutative ring with 1 (the latter since $a \cdot 1 = a$ for every power series a).

Fact / Exercise 7.18. $a = \sum_{n=0}^{\infty} a_n x^n$ is invertible, i.e., there exists $b = \sum_{n=0}^{\infty} b_n x^n$ with $ab = 1$, if and only if $a_0 \neq 0$

Example 7.19. (a) $\frac{1}{1-x} = (1-x)^{-1} = \sum_{n=0}^{\infty} x^n$ we verify $(1-x)\sum_{n=0}^{\infty} x^n = \sum_{n=0}^{\infty} c_n x^n$ with $c_0 = 1 \cdot 1 = 1$ and $c_n = 1 \cdot b_n - 1 \cdot b_{n-1} = 0 \forall n \geq 1$, where $b_n = 1 \forall n \in \mathbb{N}_0$. Hence $(1-x)\sum_{n=0}^{\infty} x^n = 1$.

(b) $\frac{1}{x-\alpha} = (x-\alpha)^{-1} = \sum_{n=0}^{\infty} -\frac{1}{\alpha^{n+1}} x^n = -\frac{1}{\alpha}\sum_{n=0}^{\infty} \frac{x^n}{\alpha^n} \forall \alpha \neq 0$ one can either verify, similarly as in (a), that $(-\alpha + x)\sum_{n=0}^{\infty} -\frac{1}{\alpha^{n+1}} x^n = 1$. Or one can observe that $\frac{1}{x-\alpha} = -\frac{1}{\alpha}\frac{1}{1-\frac{x}{\alpha}}$, and then plug in $\frac{x}{\alpha}$ for x in (a) and get $\frac{1}{1-\frac{x}{\alpha}} = \sum_{n=0}^{\infty} (\frac{x}{\alpha})^n$

Example 7.20. (a) For fixed $k \in \mathbb{N}$, set

$$u_n = |\{(e_1, ..., e_k) \in \mathbb{N}_0^k | \sum_{i=1}^{k} e_i = n\}| \forall n \in \mathbb{N}_0$$

We observed in 3.19 that $u_n = \binom{k+n-1}{k-1} \forall n \in \mathbb{N}_0$. Hence the associated generating function for $(u_n)_{n \in \mathbb{N}_0}$ is $U(x) = \sum_{n=0}^{\infty} \binom{n+k-1}{k-1} x^n$. We claim that $U(x) = \frac{1}{(1-x)^k}$.

Proof.

$$\frac{1}{(1-x)^k} = \frac{1}{1-x} \underbrace{\cdots}_{k\text{times}} \frac{1}{1-x}$$

$$(7.19(a)) \to = \left(\sum_{e_1=0}^{\infty} x^{e_1}\right) \cdots \left(\sum_{e_k=0}^{\infty} x^{e_k}\right) = \sum_{n=0}^{\infty} \left(\sum_{e_1 + \cdots + e_k = n} 1\right) x^n$$

$$= \sum_{n=0}^{\infty} |\{(e_1, ..., e_k) \in \mathbb{N}_0^k | \sum_{i=1}^{k} e_i = n\}| x^n = \sum_{n=0}^{\infty} u_n x^n = U(x)$$

□

(b) For $k \in \mathbb{N}$ and $\alpha \neq 0$, we get:

$$\frac{1}{(x-\alpha)^k} = \frac{(-1)^k}{\alpha^k} \frac{1}{(1-\frac{x}{\alpha})^k} \underbrace{=}_{(a)} \frac{(-1)^k}{\alpha^k} \sum_{n=0}^{\infty} \binom{k+n-1}{k-1} \left(\frac{x}{\alpha}\right)^n$$

In the following, we first show how we can compute the generating function $U(x) = \sum_{n=0}^{\infty} u_n x^n$ if the sequence $(u_n)_{n \in \mathbb{N}_0}$ satisfies a homogeneous linear recursion of degree k.

Proposition 7.21. Assume that the sequence $(u_n)_{n\in\mathbb{N}_0}$ satisfies

$$(*)u_n = a_1u_{n-1} + \cdots + a_ku_{n-k}\forall n \geq k$$

$a_k \neq 0$. If we set $D(x) = 1 - a_1x - \cdots - a_kx^k$, then there exists a polynomial $R(x)$ of degree $\geq k-1$ such that the generating function $U(x) = \sum_{n=0}^{\infty} u_nx^n$ satisfies the equation

$$U(x) = \frac{R(x)}{D(x)}$$

Proof. We verify that $D(x)U(x) := R(x)$ is a polynomial of degree $\leq k-1$. If we set $(1 - a_1 - \cdots - a_kx^k)\sum_{n=0}^{\infty} u_nx^n = \sum_{n=0}^{\infty} c_nx^n$, then we get for all $n \geq k$

$$c_n = u_n - a_1u_{n-1} - a_2u_{n-2} - \cdots - a_{k-1}u_{n-k+1} - a_ku_{n-k}$$

And hence, $c_n = 0\forall n \geq k$ by $(*)$. It follows that $R(x) = D(x)U(x) = \sum_{n=0}^{k-1} c_nx^n$, which is a polynomial of degree $\leq k-1$, $(= k-1$ if $c_{k-1} \neq 0)$. $\square$

Remark 7.22. (a) The coefficients $c_0, \cdots, c_{k-1}$ of $R(x)$ can easily be computed using $a_1, \cdots a_k$ and the initial values $u_0, \cdots, u_{k-1}$:

$$c_0 = u_0, c_1 = u_1 - u_1u_0, ..., c_k = u_{k-1} - a_1u_{k-2} - \cdots - a_{k-1}u_0$$

(b) Since the constant term of $D(x)$ is $1 \neq 0$, $D(x)$ is invertible by 7.18, i.e. there exist (uniquely determined) coefficients d_n such that $\frac{1}{D(x)} = \sum_{n=0}^{\infty} b_nx^n$ if one can compute these coefficients b_n, then the equation $\left(\sum_{n=0}^{k-1} c_nx^n\right)(\sum_{n=0}^{\infty} b_nx^n) = \sum_{n=0}^{\infty} u_nx^n$ can be used to derive a formula for u_n. Alternatively, instead of first computing $\frac{1}{D(x)}$ and then the product $R(x) \cdot \frac{1}{D(x)}$, one can also determine the power series $\frac{R(x)}{D(x)}$ directly if the roots of the polynomial $D(x)$ are known (see below).

(c) Recall that the auxiliary (or characteristic) polynomial associated with $(*)$ is $p(x) = x^k - a_1x^{k-1} - \cdots - a_k$. Determining the roots of $p(x)$ is equivalent to determining the roots of $D(x)$: Note first that both polynomials have nonzero constant terms so that 0 is not a root of either of them. We can write $p(x) = x^k(1 - a_1\frac{1}{x} - \cdots - a_k\frac{1}{x^k}) = x^kD(\frac{1}{x}) \rightarrow$

$$p(\alpha) = \alpha^kD(\alpha^{-1})\forall \alpha \neq 0 \rightarrow$$

$$\alpha \text{ is a root of } p \leftrightarrow \alpha^{-1} \text{ is a root of } D$$

Theorem 7.23. Assume that $\delta(x), g(x)$ are nonzero polynomials, and that the degree of f is strictly smaller than the degree of g. Assume further that $\alpha_1, ..., \alpha_r$ are the r distinct roots of g with respective multiplicities $m_1, ..., m_r$, i.e. $g(x) = c\Pi_{i=1}^{r}(x - \alpha_i)^{m_i}$ for some nonzero constant c. Then these polynomials $h_1, ..., h_r$ with $\deg h_i < m_i\forall 1 \leq i \leq r$ such that

1. $\frac{\delta(x)}{g(x)} = \frac{h_1(x)}{(x-\alpha_1)^{m_1}} + \cdots + \frac{h_r(x)}{(x-\alpha_r)^{m_r}}$ In particular, if k =deg g and g has k distinct roots $\alpha_1, ..., \alpha_k$ then there exist constants $c_1, ..., c_k$ such that

2. $\frac{f(x)}{g(x)} = \frac{c_1}{x-\alpha_1} + \cdots + \frac{c_k}{x-\alpha_k}$ Proof is available in Section 25.2 in **[B]**

How do we compute the polynomials $h_i(x)$ in 7.23(1), respectively, the constants c_i in 7.23(2). Let's start with the second question.

Lemma 7.24. If $g(x) = c\Pi_{j=1}^{r}(x - \alpha_j)$ with $c \neq 0$ and r distinct numbers $\alpha_1, \cdots, \alpha_r$, and if deg$f(x)$ <deg$g(x)$, then

$$(*)\frac{f(x)}{g(x)} = \sum_{j=1}^{r} \frac{c_j}{x - \alpha_j},$$

where $c_i = \frac{f(\alpha_j)}{c\Pi_{j\neq i}(\alpha_i - \alpha_j)} \forall 1 \leq i \leq r$.

Proof. Multiplying both sides of $(*)$ with $(x - \alpha_i)$, we get

$$\frac{f(x)}{c\Pi_{j\neq i}(\alpha_i - \alpha_j)} = c_i$$

Note that since $\alpha_j \neq \alpha_i \forall j \neq i$, $\alpha_i - \alpha_j \neq 0 \forall j \neq i$, and so we are allowed to divide by $\alpha_i - \alpha_j \forall j \neq i$. □

Example 7.25. Write $\frac{x^2+x+1}{x^3+x^2-4x-4}$ as a sum of partial fractions. So here $f(x) = x^2 + x + 1$ and $g(x) = x^3 + x^2 - 4x - 4$. First of all, we need to find the roots of $g(x)$. One immediately checks that $g(-1) = 0$, so that $g(x) = (x + 1)(h(x))$ for some polynomial $h(x)$ of degree 2. We compute $h(x)$ by long division:

$$x^3 + x^2 - 4x - 4/x + 1 = x^2 - 4$$

And obviously $x^2 - 4 = (x + 2)(x - 2)$ Thus, $g(x) = (x + 1)(x - 2)(x + 2)$ and its roots are $\alpha_1 = -1, \alpha_2 = -2, \alpha_3 = 2$. We know from 7.23 and 7.24 that $\frac{f(x)}{g(x)} = \frac{c_1}{x+1} + \frac{c_2}{x+2} + \frac{c_3}{x-2}$ with,

$$c_1 = \frac{f(\alpha_1)}{(\alpha_1 - \alpha_2)(\alpha_1 - \alpha_3)} = \frac{1}{1 \cdot (-3)} = -\frac{1}{3}$$

$$c_2 = \frac{f(\alpha_2)}{(\alpha_2 - \alpha_1)(\alpha_2 - \alpha_3)} = \frac{3}{(-1)(-4)} = \frac{3}{4}$$

$$c_3 = \frac{f(\alpha_3)}{(\alpha_3 - \alpha_1)(\alpha_3 - \alpha_2)} = \frac{7}{3 \cdot 4} = \frac{7}{12}$$

Hence,

$$\frac{x^2 + x + 1}{x^3 + x^2 - 4x - 4} = -\frac{1}{3(x + 1)} + \frac{3}{4(x + 2)} + \frac{7}{12(x - 2)}$$

(check this!)

The general case derived in 7.23(1) is more informed. If we write $h_i(x) = c_{i0} + c_{i1}x + \cdots + c_{im_j-1}x^{m_i-1}$ for $1 \leq i \leq r$ (note that $\deg h_i(x) < m_i \forall i$ and multiply (1) with $\Pi_{i=1}^r (x-\alpha_i)^{m_i}$ on both sides, then we get (since $\deg f(x) < \deg g(x) = \sum_{i=1}^r m_i$ and $\deg h_i(x)\Pi_{j\neq 1}(x-\alpha_j)^{m_j} < m_i + \sum_{j\neq i} m_j = \deg g(x) \forall 1 \leq i \leq r$) a system of $\sum_{i=1}^r m_i$ linear equations for the $\sum_{i=1}^r m_i$ unknown coefficients c_{ik_i}, $1 \leq i \leq r, 0 \leq k_i \leq m_i - 1$, which has a unique solution. Instead of describing this in full generality, I will only discuss one example where deg. $g(x) = 3$.

Example 7.26. Write $\frac{-6x^2-3x+1}{18x^3-3x^2-4x+1}$ as a sum of partial fractions. So here $f(x) = -6x^2 - 3x + 1$ and $g(x) = 18x^3 - 3x^2 - 4x + 1 = (1+2x)(1-3x)^2$(check this!). Hence $g(x)$ has the roots $\alpha_1 = -\frac{1}{2}$ with multiplicity $m_1 = 1$ and $\alpha_2 = \frac{1}{3}$ with multiplicity $m_2 = 2$. We want to write $\frac{f(x)}{g(x)}$ in the form

$$\frac{f(x)}{g(x)} = \frac{c_1}{x+\frac{1}{2}} + \frac{c_2}{x-\frac{1}{3}} + \frac{c_3}{(x-\frac{1}{3})^2} (*)$$

(Note that every polynomial of degree ≤ 1 can be written in the form $c_2(x - \frac{1}{3}) + c_3$. Multiplying $(*)$ with $(x+\frac{1}{2})(x-\frac{1}{3})^2$ on both sides we get

$$\frac{f(x)}{18} = c_1(x-\frac{1}{3})^2 + c_2(x+\frac{1}{2})(x-\frac{1}{3}) + c_3(x+\frac{1}{2})$$

$$\leftrightarrow (**) -\frac{1}{3}x^2 - \frac{1}{6}x + \frac{1}{8} = (c_1+c_2)x^2 + (-\frac{2}{3}c_1 + \frac{1}{6}c_2 + c_3)x + \frac{1}{9}c_1 - \frac{1}{6}c_2 + \frac{1}{2}c_3$$

So we have to solve the following system of linear equations

$$c_1 + c_2 = -\frac{1}{3} \quad 3c_1 + 3c_1 = -1$$

$$-\frac{2}{3}c_1 + \frac{1}{6}c_2 + c_3 = -\frac{1}{6} \leftrightarrow -4c_1 + c_2 + 6c_3 = -1$$

$$\frac{1}{9}c_1 - \frac{1}{6}c_2 + \frac{1}{2}c_3 = \frac{1}{18} \quad 2c_1 - 3c_2 + 9c_3 = 1$$

$$\iff \begin{pmatrix} 3 & 3 & 0 \\ -4 & 1 & 6 \\ 2 & -3 & 9 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} \iff \begin{array}{l} c_1 = \frac{2}{25}, \\ c_2 = -\frac{31}{75}, \\ c_3 = -\frac{2}{45}. \end{array}$$

An alternative computation of c_1, c_2, c_3 consists in a modification of the approach described in 7.2.4:

$$\frac{f(x)}{18(x+\frac{1}{2})(x-\frac{1}{3})^2 = \frac{c_1}{x+\frac{1}{2}}} + \frac{c_2}{x-\frac{1}{3}} + \frac{c_3}{(x-\frac{1}{3})^2} \rightarrow$$

$$\frac{f(-\frac{1}{2})}{18(-\frac{1}{2}-\frac{1}{3})^2} = c_1$$

and

$$\frac{f(\frac{1}{3})}{18(\frac{1}{3}+\frac{1}{2}) = c_3}$$

c_2 is determined by the equation $c_1 + c_2 = -\frac{1}{3}$. So we obtain:

$$\frac{f(x)}{g(x)} = \frac{-6x^2-3x+1}{18x^3-3x^2-4x-1} = \frac{\frac{2}{25}}{x+\frac{1}{2}} - \frac{\frac{31}{75}}{x-\frac{1}{3}} - \frac{\frac{2}{45}}{(x-\frac{1}{3})^2}$$

Applying 7.19(b) and 7.20(b), this can now be used to write $\frac{f(x)}{g(x)}$ as a power series.

$$\frac{f(x)}{g(x)} = \frac{2}{25}\frac{1}{x+\frac{1}{2}} - \frac{31}{75}\frac{1}{x-\frac{1}{3}} - \frac{2}{45}\frac{1}{(x-\frac{1}{3})^2}$$

$$= \frac{2}{25}\left(-\frac{1}{-\frac{1}{2}}\right)\sum_{n=0}^{\infty}\frac{1}{(-\frac{1}{2})^n}x^n - \frac{31}{75}(\frac{-1}{\frac{1}{3}})\sum_{n=0}^{\infty}\frac{1}{(\frac{1}{3})^n}x^n - \frac{2}{45}\frac{1}{(\frac{1}{3})^2}\sum_{n=0}^{\infty}(n+1)\frac{1}{(\frac{1}{3})^n}x^n$$

$$= \frac{4}{25}\sum_{n=0}^{\infty}(-2)^n x^n + \frac{31}{25}\sum_{n=0}^{\infty}3^n x^- \frac{2}{5}\sum_{n=0}^{\infty}(n+1)3^n x^n$$

$$= \sum_{n=0}^{\infty}(\frac{4}{25}(-2)^n + \frac{21}{25}3^n - \frac{2}{5}\frac{3^n}{n})x^n$$

Discussion 7.27. How to find formulas for u_n if $(u_n)_{n\in\mathbb{N}_0}$ satisfies a homogeneous linear recursion.

$$u_n = a_1u_{n-1} + \cdots + a_ku_{n-k}(a_k \neq 0)$$

of degree k using the generating function $U(x) = \sum_{n=0}^{\infty} u_n x^n$

Steps:

1. Find the roots of $D(x) = 1 - a_1x - \cdots - a_kx^k$ recall that these roots are the inverses of the roots of the auxiliary polynomial $x^k - a_1x^{k-1} - \cdots - a_k$; (see remark 7.22(c))
2. Compute $R(x) = r_0 + r_1x + \cdots + r_{k-1}x^{k-1}$, where $r_0 = u_0, r_1 = u_1 - a_1u_0, ..., r_{k-1} = u_{k-1} - a_1u_{k-2} - \cdots - a_{k-1}u_0$, where $u_0, u_1, \cdots, u_{k-1}$ are the k initial values of $(u_n)_{n\in\mathbb{N}_0}$
3. Develop, using step 1, $\frac{R(x)}{D(x)}$ into a sum of partial fractions
4. Use the previous step as well as the formulas in 7.19 and 7.20 to write down $\frac{R(x)}{D(x)}$ as a power series.

5. Use the previous step and the equation $U(x) = \frac{R(x)}{D(x)}$ (see 7.21) to find a formula for u_n for all $n \in \mathbb{N}_0$ by comparing the respective coefficients of x^n.

Example 7.28. Let $(u_n)_{n\in\mathbb{N}_0}$ be defined by the recursion, $u_n = 4u_{n-1}+3u_{n-2}-18u_{n-3}\forall n \geq 3$ with initial values $u_0 = u_1 = u_2$. So we can have $a_1 = 4, a_2 = 3, a_3 = -18$. Steps:

1. Here $p(x) = x^3 - 4x^2 - 3x + 18 = (x + 2)(x - 3)^2$ and hence $D(x) = 1 - 4x - 3x^2 + 18x^3$ has roots $\alpha_1 = -\frac{1}{2}$ of multiplicity $m_1 = 1$ and $\alpha_2 = \frac{1}{3}$ of multiplicity $m_2 = 2$.

2. $r_0 = u_0 = 1, r_1 = u_1 - \alpha_1 u_0 = -3$, $r_2 = u_2 - a_1u_1 - a_2u_0 = 1-4\cdot 1-3\cdot 1 = -6 \to R(x) = -6x^2 - 3x + 1$

3. (Steps 3 and 4)

$$\frac{R(x)}{D(x)} = \frac{-6x^2 - 3x + 1}{18x^3 - 3x^2 - 4x + 1} = \frac{2}{25}\frac{1}{x + \frac{1}{2}} - \frac{31}{75}\frac{1}{x - \frac{1}{3}} - \frac{2}{45}\frac{1}{(x - \frac{1}{3})^2} =$$

$$\sum_{n=0}^{\infty}\left(\frac{4}{25}(-2)^n + \frac{21}{25}3^n - \frac{2}{5}n3^n\right)x^n$$

By example 7.26

4. $u_n = \frac{4}{25}(-2)^n + \frac{21}{25}3^n - \frac{2}{5}n3^n \forall n \in \mathbb{N}_0$

Remark 7.29. (a) According to theorem 7.15, which we quoted without proof, the sequence (u_n) in 7.28 has to be a linear combination of the sequence $((-2)^n), (3^n), (n3^n)$. The fact that the sequences $((-2)^n)$ and (3^n) satisfy the recursion in 7.28 follows from the observation that -2 and 3 are roots of the polynomial $p(x) = x^3 - 4x^2 - 3x + 18$. The argument given for this conclusion in 7.12 still applies since for this it doesn't matter whether the roots are multiple or not. It is less clear that the sequence $n3^n$ satisfies the recursion of 7.28. But it can easily be verified as well: If $v_n = n3^n$ for all $n \in \mathbb{N}_0$, then

$$4v_{n-1} + 3v_{n-2} - 18v_{n-3} = 4(n-1)3^{n-1} + 3(n-2)3^{n-2} - 18(n-3)3^{n-3}$$

$$= (4n-4)3^{n-1} + (n-2)3^{n-1} - 2(n-3)3^{n-1}$$

$$= \big(4n - 4 + n - 2 - 2n + 6\big)3^{n-1} = (3n)3^{n-1} = n3^n = v_n.$$

(b) If $(u_n)_{n\in\mathbb{N}_0}$ satisfies the recursion

$$u_n = a_1u_{n-1} + \cdots + a_ku_{n-k}$$

with $a_k \neq 0$, the associated auxilary polynomial $p(x)$ is

$$p(x) = x^k - a_1x^{k-1} - \cdots - a_k \text{(see 7.14)}$$

$$\to p(x) = x^k(1 - a_1\frac{1}{x} - \cdots - a_k(\frac{1}{x})^k) = x^k D(\frac{1}{x})$$

With the polynomial $D(x) = 1 - a_1 x - \cdots - a_k x^k$ as in 7.21

$$\to D(\frac{1}{x}) = x^{-k} p(x)$$

and hence

$$D(x) = x^k p(\frac{1}{x})$$

So if $p(x) = \Pi_{i=1}^r (x-\alpha_i)^{m_i}$ is the factorization of $p(x)$ with r distinct roots $\alpha_1, ..., \alpha_r$ of respective multiplicities $m_1, ..., m_r$,n then $p(\frac{1}{x}) = \Pi_{i=1}^r (\frac{1}{x} - \alpha_i)^{m_i} \to$

$$(*) D(x) = x^k p(\frac{1}{x}) = x^k \Pi_{i=1}^r (\frac{1}{x} - \alpha_i)^{m_i}$$

Now note that all α_i are different from 0 since the constant term of $p(x)$ is $a_k \neq 0$. Also note that $k = \deg p(x) = \sum_{i=1}^r m_i$. Hence we can rewrite (*) as follows:

$$D(x) = \Pi_{i=1}^r (1 - \alpha_i x)^{m_i} = \Pi_{i=1}^r \alpha_i^{m_i} (\alpha_j^{-1} - x)^{m_i} =$$

$$c\Pi_{i=1}^r (\alpha_i^{-1} - x)^{m_i} \text{ with } c = \Pi_{i=1}^r \alpha_i^{m_i} \neq 0$$

This shows that the polynomial $D(x)$ has precisely r distinct roots $\alpha_1^{-1}, ..., \alpha_r^{-1}$, each with respective multiplicity m_i.

(c) The assertion in 7.15 that we have to consider (linear combinations of) the sequences $(n^i \alpha^n)_{n \in \mathbb{N}_0}$ for $i = 0, ..., m-1$ ifα is a root of multiplicity m of the auxiliary polynomial $p(x)$ come out of the blue. It can now be better motivated by Discussion 7.27. First of all, by the above Remark (b), α^{-1} is a root of multiplicity m of the polynomial $D(x)$. Then the part of the partial fraction decomposition $\frac{R(x)}{D(x)}$ involving α^{-1} is of the form

$$\sum_{k=1}^m \frac{c_k}{(x - \alpha^{-1})^k}$$

with constants $c_1, ..., c_m$. Writing this sum as a power series and using 7.20(b), we get (replacing α with α^{-1})

$$\sum_{k=1}^m c_k (-1)^k \alpha^k \sum_{n=0}^\infty \binom{k+n-1}{k-1} \alpha^n x^n =$$

$$\sum_{n=0}^\infty \underbrace{\sum_{k=1}^m c_k (-1)^k \alpha^k \binom{n+k-1}{k-1}}_{=:h_k(n)} \alpha^k x^n$$

Since for each $k \leq m$, $\binom{k+n-1}{k-1} = \frac{1}{(k-1)!}(n+k-1)(n+k-2)\cdots(n+1)$ is a polynomial of degree $k-1$ in n, $h_k(n)$ is a polynomial of degree $m-1$ in n. Hence there exist numbers $b_0, \cdots, b_{m-1}$ such that $h_k(n) = \sum_{j=0}^{m-1} b_j n^j$, which implies

$$\sum_{k=1}^{m} \frac{c_k}{(x-\alpha^{-1})^k} = \sum_{n=0}^{\infty} \left(\sum_{j=0}^{m-1} b_j n^j \right) \alpha^n x^n =$$

$$\sum_{n=0}^{\infty} \left(\sum_{j=0}^{m-1} b_j n^j \alpha^n \right) x^n$$

in view of step 5 of Discussion 7.27 this explains why a linear combination of the sequences $(n^j \alpha^n)_{n \in \mathbb{N}_0}$ for $j = 0, 1, \cdots, m-1$ occurs in the expression for (u_n).

8 Symmetric Group S_n

Recall definition 2.26(b)

$$S_n = \{f : \mathbb{N}_n \to \mathbb{N}_n | f \text{ is bijective}\}$$

So far we treated S_n as a set and, for instance, established that $|S_n| = n!$ in 2.27. However, S_n has also the structure of a group w.r.t composition as a group operation: If $f, g \in S_{n-1}$, then also $f \circ g \in S_n$ and $f^{n-1}, g^{-1} \in S_n$ id is the identity element of the group S_n, we have $f \circ id = id \circ f$ and $f \circ f^{-1} = f^{-1} \circ f = id \forall f \in S_n$. And, obviously, composition satisfies the Associative Law

$$(f \circ g) \circ h = f \circ (g \circ h) \forall f, g, h \in S_n$$

There are different ways to denote the elements of S_n. $f \in S_n$ is given the images $f(j)$ for $1 \leq j \leq n$. So we can write

$$f = \begin{pmatrix} 1 & 2 & \cdots & n \\ i_1 & i_2 & \cdots & i_n \end{pmatrix}$$

with $i_j = f(i_j)$ for $1 \leq j \leq n$. Note that $\{i_1, \cdots, i_n\} = \mathbb{N}_n$ since f is a permutation $\to f(\mathbb{N}_n) = \mathbb{N}_n$. An important type of permutations are cycles which are written in the form $(j_1, \cdots, j_k)$, where $\{j_1, \cdots, j_k\}$ is a $k-$subset of $\mathbb{N}_n$ and if $f = (j_1, \cdots, j_k)$, then

$$f(j_i) = j_{i+1} \text{ for } 1 \leq i \leq k-1, f(i_k) = i_1, \text{ and}$$

$$f(i) = i \forall i \in \mathbb{N}_n \setminus \{j, \cdots, j_k\}$$

Example

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 2 & 4 & 3 \end{pmatrix}$$

For each 1-cycle (j) is simply a different notation for id. 2-cycles are also called transpositions; they are of the form (ij) with $i \neq j \in \mathbb{N}_n$ so that $(ij)(k) = \begin{cases} j & \text{if } k = i \\ i & \text{if } k = j \\ k & \text{if } k \neq i, j \end{cases}$ Fact: Each permutation $f \in S_n$ can be written as a product of transpositions, i.e., $\exists$ permutations $\pi_1, \cdots, \pi_r \in S_n$ such that $f = \pi_1 \cdots \pi_r$. Fact 2: If $f = (j_1 \cdots j_k)$ is a $k\cdot$ cycle, then the order of $|f|$ of f is $|f| = k$. We also have $f = (j_k j_i \cdots j_{k-1}) = \cdots = (j_2 \cdots j_k j_1)$. Hence there are k different notations for the same $k-$cycle. Corollary. The symmetric group S_n has

$$\frac{(n)_k}{k} = \frac{n(n-1)\cdots(n-k+1)}{k}$$

k-cycles. Here $(n)_k$ is the number of injective maps from $\mathbb{N}_k$ to $\mathbb{N}_n$ (see 2.21), i.e. the cardinality of

$$\big\{(j_1, \ldots, j_k) \in \mathbb{N}_n^k \;\big|\; j_a \neq j_b \text{ for all } a \neq b\big\},$$

and each k-cycle in S_n is determined by exactly k such k-tuples.

Fact 3: Every $f \in S_n$ is a product of disjoint cycles, i.e. there exist cycles $\sigma_1, \cdots, \sigma_r \in S_n$ which are pairwise disjoint (i.e. for each $l \neq m$, if $\sigma_l = (j \cdots j_k)$ and $\sigma_m = (a_1 \cdots a_s)$ with $j_1, \cdots, j_k, a_1, \cdots, a_s \in \mathbb{N}_n$, then $\{j_1, \cdots, j_k\} \bigcap \{a_1, \cdots, a_s\} = \emptyset$ such that $f = \sigma_1 \circ \cdots \circ \sigma_r$. Additionally, this factorization is unique up to the permutation of the factors (one immediately checks that $\sigma_l \circ \sigma_m = \sigma_m \circ \sigma_l$ if σ_l and σ_m are disjoint cycles). This means that if $f = \sigma_1 \circ \cdots \circ \sigma_r$ and $f = \sigma_1' \circ \cdots \circ \sigma_r'$ with pairwise disjoint cycles $\sigma_1, \cdots, \sigma_r$, respectively, $\sigma_1', \cdots, \sigma_{r'}'$ in S_n, then $r = r'$ and there exists a permutation $\pi \in S_n$ such that $\sigma_i' = \sigma_{\pi(i)} \forall 1 \leq i \leq r$.

Example

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 4 & 1 & 2 & 5 & 3 \end{pmatrix} = (163)(24)(5) = (5)(24)(163) etc.$$

The fact that 5 occurs in a 1-cycle means that 5 is an element of $\mathbb{N}_6$ and (the only one) which is fixed by $f \in S_6$.

Definition If the decomposition of $f \in S_n$ into the product of disjoint cycles, there are m_i cycles of length i for $1 \leq i \leq n (m_i \in \mathbb{N}_0)$, than this is called the cycle type or simply the type of f. In **[B]**, the type of f is then denoted by $[1^m 2^{m_2} \cdots n^{m_n}]$.

Since every $j \in \mathbb{N}_n$ occurs in precisely one of the cycles of the factorization of f into disjoint cycles, we have $\sum_{i=1}^n m_i \cdot i = n$ ($m_i \cdot i$ is the number of elements in $\mathbb{N}_n$ which occur in one of the m_i cycles of length i). Of course m_i is often 0, for instance, if f is an $n-$cycle, then $m_n = 1$ and $m_i = 0$ for all $1 \leq i \leq n-1$. In the notation of the cycle type one often omits the i for which $m_i = 0$. For instance, in the example of $f \in S_6$ above, its cycle type would be $[1'2'3']$ Another more

descriptive notation for the cycle type which is often used is

$$(\cdot)(\cdot\cdot)(\cdots) \text{ for this } f \in S_6$$

Fact 4 $f \in S_n$ is a derangement of $\mathbb{N}_n$ (see 4.8) if and only if the number m_1 in its cycle type is 0, which precisely means that no element $j \in \mathbb{N}_n$ is fixed by f. (Obvious) We are now turning to the question of how many permutations of a given cycle type there are in S_n. Fact 5 If the given type is $[1^m, 2^m \cdots, n^{m_n}]$, then the number of elements of S_n having this type is

$$\frac{n!}{1^m 2^{m_2} \cdots n^{m_n} m_1! m_2! \cdots m_n!}$$

This follows from counting the number of possible cycles of length i (see this Corollary below fact 2) and the fact that cycles of the same length can be permuted arbitrarily in a product in which they occur; see section 12.5 in $[B]$ for a more detailed discussion. In practice, for small n, one isn't using this formula, but quickly doing case by case computations. Example (This is exercise (6) of Homework # 13) What are the possible cycle types for derangements of $\mathbb{N}_6$, and how many derangements are there of each of these types?
Answer: (also using Fact 4)

$$(\cdot\cdot\cdot\cdot\cdot\cdot) = [6'] \quad \frac{6!}{6} = 5! = 120$$

$$(\cdot\cdot)(\cdot\cdot\cdot\cdot) = [2'4'] \quad \frac{6\cdot 5}{2} \cdot \frac{4\cdot 3\cdot 2\cdot 1}{4} = 15 \cdot 6 = 90$$

$$(\cdots)(\cdots) = [3^2] \quad \frac{6\cdot 5\cdot 4}{3} \cdot (\frac{3\cdot 2\cdot 1}{3}) \cdot \frac{1}{2!} = 40$$

$$(\cdot\cdot)(\cdot\cdot)(\cdot\cdot) = [2^3] \quad \frac{6\cdot 5}{2} \cdot \frac{4\cdot 3}{2} \frac{2\cdot 1}{2} \frac{1}{5} = \frac{15}{265}$$

Note that, for instance, with type $(\cdots)(\cdots)$, if you already choose numbers for the first 3-cycle ($\frac{6\cdot 5\cdot 4}{3} = 40$ possibilities), then for the second 3−cycle you can only choose from the remaining 3 numbers from $\mathbb{N}_6$ (hence $\frac{3\cdot 2\cdot 1}{3} = 2$ remaining possibilities for the second 3 cycle). But since, for instance,

$$(123)(456) = (456)(123)$$

You have to divide by $2 = 2!$ in order to get the correct number of permutations of type $(\cdots)(\cdots)$. And in the last case, there are even $3! = 6$ possibilities of ordering the factors in the product and still getting to the same permutation. So by combining all permutations with cycle types where no 1-cycles are occuring, one obtains that there are 90+120+40+15 = 265 derangements of $\mathbb{N}_6$. This agrees with our computations of d_n in Chapter 4. The quickest way to compute d_n is by using the recursion $d_n = (n-1)(d_{n-1} + d_{n-2})$ (See exercise (5) of Homework # 7) Where we computed $d_4 = 9$ and $d_5 = 44$ in 4.10. Hence $d_6 = 5(44+9) = 5 \cdot 53 = 265$. As noted in 4.11, $\frac{n!}{e}$ is a good approximation for d_n (or equivalently, $\frac{1}{e}$ is a good approximation for $\frac{d_n}{n!} = \frac{d_n}{|S_n|}$) Hence we have $\frac{6!}{e} = \frac{720}{e} \approx 264.873$, $\frac{1}{e} = 0.36788$ and $\frac{d_6}{6!} = 0.368$.