Section 6.3 Counting

Counting means assigning consecutive natural numbers to the elements of a set in a one-to-one fashion. Let us formulate counting in a mathematical style.

With n a natural number, let $[n)$ denote the set

$$[n) \underset{\text{def}}{=} \{k \varepsilon \mathbb{N} \mid k < n\} = \{0,\ 1,\ \ldots,\ n\text{-}2,\ n\text{-}1\};$$

$[n)$ is the primordial set with exactly n elements. If $n = 0$, $[n) = \emptyset$, the empty set; $[1) = \{0\}$, $[2) = \{0, 1\}$, etc.

We say that *the cardinality of* the set X *is* n, or that *the number of elements of* X *is* n, if there is a bijection $f:[n) \xrightarrow{\cong} X$; the function f provides the counting of X.

The question arises if one could have a bijection $f:[n) \xrightarrow{\cong} X$ and another $g:[m) \xrightarrow{\cong} X$ with the same set X, but with different n and m; if so, the notion of cardinality would not be well-defined. The answer to the question is "no"; the described situation is impossible. Namely, if we had that situation, $h \underset{\text{def}}{=} g^{-1} \circ f:[n) \longrightarrow [m)$ would be a bijection (see Chapter 1, p.25, where it is stated that the composite of two bijections is a bijection), and we would have a bijection between two different sets of the form $[n)$, contrary to the third of the following propositions:

If $h:[n) \longrightarrow [m)$ *is an injection,* $n \leq m$;
if $h:[n) \longrightarrow [m)$ *is a surjection,* $n \geq m$;
if $h:[n) \longrightarrow [m)$ *is a bijection,* $n = m$.

The proof of these assertions use induction; of course, the last part is a consequence of the two previous parts.

We call a set A *finite* if there exists $n \varepsilon \mathbb{N}$ such that the cardinality of A is n. That is, A

is finite if there exists a bijection of the form $[n) \xrightarrow{\cong} A$, with $n \in \mathbb{N}$. The cardinality of the set A is denoted by $|A|$; $|A|$ is defined just in case A is finite (in set theory, one talks about the cardinality of infinite sets too; we will not do so here). Thus, e.g.,

$$|[n)| = n \qquad (n \in \mathbb{N})$$

(exemplified by the identity function $1_{[n)} : [n) \longrightarrow [n)$).

A frequently applied method of finding out the cardinality of a set X is to find another set Y the cardinality of which is known, and to establish a bijection of X and Y; in this case we know that the cardinality of X is the same as that of Y. The principle is

If $|Y| = n$ *and* $f : Y \xrightarrow{\cong} X$, *then* $|X| = n$.

This is obvious, since, under the assumptions here, we have some $g : [n) \xrightarrow{\cong} Y$, and then $f \circ g : [n) \xrightarrow{\cong} X$.

One simple law concerning finiteness is that

any subset of a finite set is finite; *moreover, a proper subset of a finite set has a strictly smaller cardinality.*

The rigorous proof is by an induction: one proves by induction on the natural number n that any subset of a set of cardinality n is finite, in fact, of cardinality $\leq n$.

The three propositions stated above immediately generalize in the following forms:

If $h : A \longrightarrow B$ *is an injection,* $|A| \leq |B|$;
if $h : A \longrightarrow B$ *is a surjection,* $|A| \geq |B|$;
if $h : A \longrightarrow B$ *is a bijection,* $|A| = |B|$.

The third proposition expresses the fundamental fact of life according to which if we count a pile of pebbles on two occasions, and in the meantime, no pebble was added or taken away, then the numbers arrived at must be the same.

The first proposition is equivalent to the so-called **pigeon–hole principle,** according to which

if we have put n things into less than n holes, then in at least one hole, we have put at least two things.

Namely, let the set of the things be A and the set of holes B; $|A| = n$ and $|B| < n$; let the function $f:A \longrightarrow B$ map every thing in A to the hole it is put into; since $|A| > |B|$, f cannot be an injection, that is, there are $a_1 \neq a_2$ in A for which $f(a_1) = f(a_2)$, i.e., a_1 and a_2 are put into the same hole.

E.g., among thirteen people, there must always be at least two who were born in the same month. Among thirteen integers, there always are two distinct ones whose difference is divisible by 12: there are two that give the same remainder when divided by 12, and their difference is divisible by 12.

If $f:A \longrightarrow A$ is an injective function of a finite set A into itself, then f is a bijection; if $f:A \longrightarrow A$ is a surjective function of a finite set A into itself, then f is a bijection.

To see the first assertion, assume that $f:A \longrightarrow A$ is injective. Suppose f is not surjective, to derive a contradiction. There is some $a \in A$ such that $a \notin \text{range}(f)$. Then the same function f can be considered a function from A to $A-\{a\}$; that is, $f:A \longrightarrow A-\{a\}$; f so construed is still injective. But then we would have $|A| \leq |A-\{a\}|$, contradicting the fact that $A-\{a\}$ is a proper subset of A. This contradiction proves that f must be surjective.

The other half of the proposition is proved similarly; now, we take away an element from the domain, rather than from the codomain.

An application of the last-stated principle is the important proposition that

the congruence $ax \equiv b \pmod{n}$ *is solvable for the unknown* x *provided* $\gcd(a, n)=1$.

(For congruences, see section 2.2 in Chapter 2.)

To see this, first of all, recall that we denoted the set of all equivalence classes of the congruence $\bmod n$ by $\mathbb{Z}/n$, and that we proved that $\mathbb{Z}/n$ has exactly n elements; in particular, it is a finite set. Now, consider the mapping $f:\mathbb{N}/n \longrightarrow \mathbb{N}/n$ that takes $[x]$ to $[ax]$. Is f well-defined? For this, we need that if $[x]=[y]$, then $[ax]=[ay]$. But this is true: see Exercise 2 on page 45 of Chapter 2.

Under the assumption that $\gcd(a, n)=1$, f is an injective map: if $[ax]=[ay]$, then $ax \equiv ay \pmod{n}$, that is, $n \mid ax-ay=a(x-y)$; and since $\gcd(a, n)=1$, that is, a and n have no common prime factor, we must have that $n \mid x-y$, which means $x \equiv y \pmod{n}$, and so $[x]=[y]$.

By our last stated principle, f is surjective. This is what we want: the surjectivity of f means that for any $[b] \in \mathbb{Z}/n$ there exists $[x] \in \mathbb{Z}/n$ such that $f([x])=[b]$, that is, $[ax]=[b]$, that is, there is $x \in \mathbb{Z}$ such that $ax \equiv b \pmod{n}$.

The basic laws of counting connect operations on sets with operations on numbers. Here are the most important ones; the sets A, B, etc. are assumed to be finite.

$$|A \cup B| = |A| + |B| \qquad (1)$$

provided $A \cap B = 0$ (A *and* B *are disjoint*),

$$|A \times B| = |A| \cdot |B|, \qquad (2)$$

$$|B^A| = |B|^{|A|}. \qquad (3)$$

These laws contain the information that the operations of union, Cartesian product, and exponentiation, when applied to finite sets, result in finite sets again. The laws can be proved by appropriate inductions; e.g., the last one by induction on $|A|$. Let us see how this proof goes.

Basis Step. $|A| = 0$. In this case, A is empty, and there is exactly one function from $A = 0$ to (any set) B. Therefore, $|B^A| = 1$. Also, by the definition of exponentiation of numbers, $|B|^{|A|} = |B|^0 = 1$. This shows that the desired equality holds in this case.

Induction Step. $|A| = n + 1$, that is, there is a bijection $f:[n+1) \longrightarrow A$. Let $a = f(n)$, and let $A' = A - \{a\}$. The function f restricted to the subset $[n)$ of its domain, $g = f\restriction[n)$, is now a bijection from $[n)$ to A'; in particular, $|A'| = n$. Now, we set up a bijection

$$g : B^{A'} \times B \xrightarrow{\cong} B^A$$

as follows: to any pair $(s, b) \in B^{A'} \times B$ where s is a function $s:A' \longrightarrow B$ and $b \in B$, g assigns the function $t:A \longrightarrow B$ for which

$$t(x) = \begin{cases} s(x) & \text{if } x \neq a \text{ (hence, } x \,\varepsilon\, A') \\ b & \text{if } x = a \end{cases}$$

It is easy to check that g is indeed a bijection. It follows that

$$\begin{aligned} |B^A| = |B^{A'} \times B| &= |B^{A'}| \cdot |B| && \text{(by (2))} \\ &= |B|^n \cdot |B| && \text{(by the induction hypothesis,} \\ |B^{A'}| = |B|^n &\text{, since } |A'| = n \text{)} \\ &= |B|^{n+1} && \text{(by the laws } m^1 = m, \\ m^{n+\ell} = m^n \cdot m^\ell &\text{ for exponentiation of numbers)} \\ &= |B|^{|A|} \end{aligned}$$

as desired.

An application of the last fact is the proof of the relation

$$|\mathcal{P}(A)| = 2^{|A|},$$

or in words, the number of all subsets of an n-element set is 2^n. The reason for this is that the subsets of A are in a one-to-one correspondence with the functions $A \longrightarrow \{0, 1\}$: if $X \subseteq A$, we consider $\chi_X : A \longrightarrow \{0, 1\}$, the *characteristic function* of X, defined by

$$\chi_X(a) = \begin{array}{l} 1 \text{ if } a \in X \\ 0 \text{ if } a \notin X \end{array}.$$

Any function $\chi : A \longrightarrow \{0, 1\}$ is the characteristic function of a unique subset X of A, namely of $X = \{a \in A \mid \chi(a) = 1\}$. Thus, we have the bijection

$$\begin{array}{ccc} \mathcal{P}(A) & \xrightarrow{\cong} & \{0, 1\}^A \\ X & \longmapsto & \chi_X \end{array}$$

and therefore, $|\mathcal{P}(A)| = |\{0, 1\}^A| = 2^{|A|}$ as claimed.

The laws (1), (2), (3) are generalized to many-termed unions/sums and Cartesian products/products as follows. In what follows, I and each A_i are assumed to be finite sets.

Sum rule:

$$\left|\bigcup_{i \in I} A_i\right| = \sum_{i \in I} |A_i| \qquad \textit{provided the } A_i \textit{ are pairwise disjoint}:$$

$A_i \cap A_j = \emptyset$ *whenever* $i, j \in I$ *and* $i \neq j$.

Product rule:

$$\left|\prod_{i \in I} A_i\right| = \prod_{i \in I} |A_i|$$

(thus, the use of the same symbol $\prod$ for the product of numbers and the Cartesian product of sets is justified). The proofs of these identities are by induction on $|I|$. (1) is the special case of the sum rule when $|I| = 2$; (2) is the special case of the sum rule when $B = I$ and $A_b = \{b\} \times A$ (essentially, the case of equal-cardinality terms); (3) is the special case of the product rule when $I = A$, and $A_i = B$ for all $i \in I$.

The sum rule can be expressed in the following informal way. We have a set A which is *partitioned into* certain subsets A_i, for $i \in I$, or A is the *disjoint union* of the A_i's, meaning that $A = \bigcup_{i \in I} A_i$ and the A_is are pairwise disjoint. Note that this is the same as to say that every $a \in A$ belongs to A_i for exactly one index $i \in I$. To count the elements of A it suffices to count the elements of each A_i, and to add up the numbers obtained. We write $A = \dot{\bigcup}_{i \in I} A_i$ to indicate that A is the disjoint union of the A_i's.

To consider a kind of situation when the sum rule is useful, let $f: A \longrightarrow B$ an arbitrary function. Then the sets $f^{-1}(\{b\})$ when b runs over B form a partition of A: every $a \in A$ is in exactly one of the sets $f^{-1}(\{b\})$, namely the one for which $b = f(a)$. The sum rule says that $|A| = \sum_{b \in B} |f^{-1}(\{b\})|$. If we also assume that the sets $f^{-1}(\{b\})$ are all of equal cardinality, say m, then this says that $|A| = m \cdot |B|$.

The product rule is paraphrased as follows. An element of $\prod_{i \in I} A_i$ is the result of m independent choices ($m = |I|$), the i^{th} choice constrained to lie in the set A_i. The number of such compound selections consisting of m independent choices is the product of the numbers of the possibilities of the m individual choices.

The product rule has a generalized form which is the really useful version in practice. In this, we have selections in which the individual choices are not independent of each other, but the numbers of them are. We consider a subset A of a Cartesian product $\prod_{i<n} B_i$ determined as follows. The sequence $\langle a_i \rangle_{i<n}$ from $\prod_{i<n} B_i$ belongs to A iff each a_i belongs to a certain *constraint-set* $A(\langle a_j \rangle_{j<i})$, a subset of B_i depending on the segment $\langle a_j \rangle_{j<i}$ of the a_j preceding a_i. The essential assumption is that the cardinality of the

constraint-set $A(\langle a_j\rangle_{j<i})$ does not depend on $\langle a_j\rangle_{j<i}$, just on i; let us say, this cardinality is n_i:

$$|A(\langle a_j\rangle_{j<i})| = n_i,$$

at least when $\langle a_j\rangle_{j<i}$ is *properly constrained*: $a_j \in A(\langle a_k\rangle_{k<j})$ for all $j<i$.

In this case, $|A| = \prod_{i<k} n_i$. Let us call this rule the ***product rule for dependent selections.***

The product rule for dependent selections can be proved by induction on k, the length of the selections made.

Let us take a simple case illustrating the last mentioned rule. Let C be an alphabet of size n, and let us compute the number of strings in C^* in which there are no identical letters next to each other. The set of such strings being called A, A is a subset of $\prod_{i<k} C$ (we identify strings with sequences), and $\langle a_i\rangle_{i<k}$ from $\prod_{i<k} C$ belongs to A just in case for each i in the range $1\le i\le k$, we have $a_i \neq a_{i-1}$. In other words, in this case

$$A(\langle a_j\rangle_{j<i}) = \{a\in C \mid a \neq a_{i-1}\}$$

if $1\le i\le k$, and

$$A(\langle a_j\rangle_{j<0}) = C.$$

Thus, the numbers n_i are: $n_0 = n$, $n_i = n-1$ when $1\le i\le k$, and the desired number is $n\cdot(n-1)^{k-1}$.

It is customary to express the above argument in the following informal way. To have a string in which there are no two identical letters next to each other, we may take n different letters as the first letter of the string. But for the second letter, we can take only $n-1$, since the first one is now excluded. This says that the number of compound choices for the first two positions is $n(n-1)$. For the third letter we can again choose from $n-1$ letters, the ones that are different from the second letter, whatever that was; thus, there are $n(n-1)(n-1)$ possibilities for the segment in the first three positions. Etc.; the number of such strings of

length k is $n(n\text{-}1)^{k-1}$.

We can see that the informal argument actually reproves the product rule by induction on k.

Let us determine the cardinality of some important finite sets.

Suppose that $|A| = m$, $|B| = n$ *and* $m \leq n$. *Then the number of injections* $A \xrightarrow{\cong} B$ *between two given sets* A *and* B is

$$\prod_{i<m} (n\text{-}i) = n \cdot (n\text{-}1) \cdot \ldots \cdot (n\text{-}m\text{+}2)(n\text{-}m\text{+}1)$$

This can be easily shown by the product rule for dependent selections. First of all, we may assume without loss of generality that $A = [m)$. A function $A \longrightarrow B$ is a sequence $\langle b_i \rangle_{i<m}$ with each $b_i \in B$. The sequence $\langle b_i \rangle_{i<m}$ is an injection iff for all $i < m$, b_i differs from b_j for each $j < i$. This means that b_i in $\langle b_i \rangle_{i<n}$ is constrained to lie in the set

$$B(\langle b_j \rangle_{j<i}) = \{b \in B \mid b \neq b_j \text{ for all } j < i\}.$$

The latter set has cardinality $n - i$, since the b_j's are all distinct (the selection $\langle b_j \rangle_{j<i}$ being "properly constrained"), and hence, there are exactly i of them. We see that the cardinality of the constraint-set $B(\langle b_j \rangle_{j<i})$ is independent of the segment $\langle b_j \rangle_{j<i}$, it depends on i only. The product rule, for the variant for dependent selections, gives that the desired number is $\prod_{i<m} (n\text{-}i)$ as promised.

A special case of the last proposition, for the case $m = n$, is the following.

The number of bijections between two fixed sets of the same cardinality n *is* $n!$*; in particular, the number of permutations of a set of cardinality* n *is* $n!$.

Indeed, this follows from the previous proposition, since any injection from a set to another of the same cardinality is a bijection, as we stated above.

Note that the injections from $[m)$ into an alphabet A are the same as the strings in A^* of length m in which no letter is repeated; the proposition above gives a formula for the number of such strings.

Let $\binom{n}{k}$ (read: " *n-choose-k* ") denote the number of k-element subsets (more briefly: k-subsets) of an n-element set. Clearly, if $n < k$, then $\binom{n}{k} = 0$. Also, $\binom{n}{0} = \binom{n}{n} = 1$. We claim that

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad \textit{whenever} \quad k \leq n .$$

To show this, let us fix k and n, $k \leq n$. Let the set of all permutations of $[n)$ be called P, and let the set of all k-subsets of $[n)$ be S. We partition the permutations of $[n)$ into as many disjoint sets as there are k-subsets of $[n)$. Let $\sigma : [n) \xrightarrow{\cong} [n)$ be any permutation; consider the set of values of σ at the first k arguments $0, 1, \ldots, k-1$, that is, the set

$$X_\sigma \stackrel{\text{def}}{=} \{\sigma(0), \sigma(1), \ldots, \sigma(k-1)\} .$$

Since σ is one-to-one, X_σ is a k-subset of $[n)$. Consider the function

$$\begin{array}{rcl} f : P & \longrightarrow & S \\ \sigma & \longmapsto & X_\sigma \end{array} .$$

For any k-subset $X \in S$ *of* $[n)$, $f^{-1}(\{X\})$ consists of those permutations σ for which X_σ is the given set X;

$$f^{-1}(\{X\}) = \{\sigma \mid X_\sigma = X\} .$$

We **claim** that

$$|f^{-1}(\{X\})| = k!(n-k)! \qquad (4)$$

independently of X. This equality is based on the following general fact.

Suppose $A=A_1\dot\cup A_2$ and $B=B_1\dot\cup B_2$, assume that $|A_1| = |B_1|$, $|A_2| = |B_2|$ (and, as a consequence, $|A| = |B|$), and let us consider the set T of those bijections $\sigma:A\xrightarrow{\cong}B$ for which $\sigma[A_1]=B_1$, that is, σ maps A_1 onto B_1. Then, writing (temporarily) $\mathtt{Bij}(U, V)$ for the set of all bijections $U\xrightarrow{\cong}V$, we have a *bijective* mapping

$$\begin{array}{ccc} T & \xrightarrow{\quad\cong\quad} & \mathtt{Bij}(A_1, B_1)\times\mathtt{Bij}(A_2, B_2) \\ \sigma & \longmapsto & (\ \sigma{\restriction}A_1\ ,\ \sigma{\restriction}A_2\ )\ . \end{array} \qquad (4')$$

The point is that if the bijection $\sigma:A\xrightarrow{\cong}B$ maps (bijectively) A_1 onto B_1, then it necessarily maps the rest of A, A_2, bijectively onto B_2, the rest of B. In other words, if $\sigma\in T$, then $\theta_1=\sigma{\restriction}A_1\in\mathtt{Bij}(A_1, B_1)$ and $\theta_2=\sigma{\restriction}A_2\in\mathtt{Bij}(A_2, B_2)$. Conversely, if $\theta_1\in\mathtt{Bij}(A_1, B_1)$, $\theta_2\in\mathtt{Bij}(A_2, B_2)$, then σ defined by

$$\sigma(a) = \begin{cases} \theta_1(a) & \text{if } a\in A_1 \\ \theta_2(a) & \text{if } a\in A_2 \end{cases}$$

is a bijection $\sigma:A\xrightarrow{\cong}B$ for which $\sigma{\restriction}A_1=\theta_1$ and $\sigma{\restriction}A_2=\theta_2$.

In our application, $A=B=[n)$, $A_1=[k)$, $A_2=[n)-[k)$, $B_1=X$, $B_2=[n)-X$. Then the set T is what we called $f^{-1}(\{X\})$. Since $|A_1| =B_1| =k$, $|A_2| =B_2| =n-k$, we have $|\mathtt{Bij}(A_1, B_1)| =k!$, $|\mathtt{Bij}(A_2, B_2)| =(n-k)!$. The relation (4') therefore tells us that $|T| =k!\cdot(n-k)!$, as desired. This shows (4).

Since for each k-subset X of $[n)$, $f^{-1}(\{X\})$ is of the same cardinality, namely

$k!(n-k)!$, and P is partitioned into $\binom{n}{k}$ sets $f^{-1}(\{X\})$, we have

$$|P| = \binom{n}{k} \cdot k!(n-k)! \ .$$

But we know that $|P| = n!$. The desired expression for $\binom{n}{k}$ follows by dividing by $k!(n-k)!$.

The numbers $\binom{n}{k}$ are called the *binomial coefficients*, because their appearance in the

Binomial theorem:

$$(x+y)^n = \sum_{k \le n} \binom{n}{k} x^k y^{n-k} \qquad\qquad (n \in \mathbb{N},\ n \geq 1) \ .$$

This equality is immediate when one considers that in the product $(x+y)...(x+y)$ (n factors), when written out via the distributive law as a sum of monomials $x^k y^{n-k}$, the number of terms with exactly k x-factors (and hence exactly $n-k$ y-factors) is the same as the number of ways we can select k factors $(x+y)$ out of the n such; the latter number is, by definition, $\binom{n}{k}$.

The binomial coefficients satisfy many identities. One such is

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1} \ .$$

The reason for this is the fact that the set S of $k+1$-subsets of $[n+1)$ is partitioned into two disjoint subsets, S_1 and S_2, according to whether $X \varepsilon S$ does or does not contain the element n. The elements of S_1 are in one-to-one correspondence with the k-subsets of $[n)$: with $X \varepsilon S_1$, take away from X the fixed element n, and get a k-element subset of $[n)$. S_2 is nothing but the set of all $k+1$-subsets of $[n)$. Thus

$|S| = \binom{n+1}{k+1}$, $|S_1| = \binom{n}{k}$, and $|S_2| = \binom{n}{k+1}$,

and since $S = S_1 \dot{\cup} S_2$, the assertion follows.

The last-proved identity gives a recursive definition of the binomial coefficients. The successive calculation of the binomial coefficients is suggested by the *Pascal triangle*:

$$\begin{array}{c}
\binom{0}{0} \\
\binom{1}{0}\ \binom{1}{1} \\
\binom{2}{0}\ \binom{2}{1}\ \binom{2}{2} \\
\binom{3}{0}\ \binom{3}{1}\ \binom{3}{2}\ \binom{3}{3} \\
\binom{4}{0}\ \binom{4}{1}\ \binom{4}{2}\ \binom{4}{3}\ \binom{4}{4} \\
\binom{5}{0}\ \binom{5}{1}\ \binom{5}{2}\ \binom{5}{3}\ \binom{5}{4}\ \binom{5}{5} \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots
\end{array}$$

in which every coefficient is the sum of the two immediately above it, and in which all the values on the two sloping sides are equal to 1.

Substituting particular values for x and y in the binomial theorem, we get various identities involving the binomial coefficients. E.g., if we put $x = -1$, $y = 1$, we obtain

$$(-1 + 1)^n = 0 = \sum_{k \le n} (-1)^k \binom{n}{k} \qquad (n \ge 1),$$

that is,

$$\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \ldots + (-1)^{n-1}\binom{n}{n-1} + (-1)^n \binom{n}{n} = 0 .$$

Since $\binom{n}{0} = 1$, we may rewrite this as

$$\binom{n}{1} - \binom{n}{2} + \ldots + (-1)^{k+1}\binom{n}{k} + \ldots + (-1)^{n+1}\binom{n}{n} = 1\,,$$

that is,

$$\sum_{k=1}^{n} (-1)^{k+1}\binom{n}{k} = 1 \qquad (\, n \geq 1\,). \tag{5}$$

We will make use of the last identity in establishing the so-called *sieve principle*, or *inclusion/exclusion principle*.

The principle mentioned concerns the way one can compute the cardinality of a union of sets. The sum rule gives the answer when the sets involved are pairwise disjoint. In the general case, the answer involves the cardinalities of the various intersections of the given sets (which are all equal to 0 in the disjoint case).

Consider the special case of the union of two sets. We have

$$|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2|\,;$$

the reason is that "when we add up the cardinalities of A_1 and A_2, we count the elements in the intersection $A_1 \cap A_2$ twice; subtracting the cardinality of the intersection corrects this".

The case of three sets is like this:

$$|A_1 \cup A_2 \cup A_3| = |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| + |A_1 \cap A_2 \cap A_3|\,.$$

An argument justifying this would say that the corrections afforded by the three subtractions over-correct precisely for the elements that are in at least two of the double intersections; but these are exactly the elements which are in the triple intersection; hence, we have to compensate by adding the cardinality of that triple intersection.

We have to admit that these arguments, although intuitive, fall somewhat short of the ideal of a clear mathematical proof. Considering that the general case of an arbitrary number of sets is likely to be more involved, we are drawn to a more serious mathematical approach.

First of all, let us state the general result:

$$\left| \bigcup_{i=1}^{n} A_i \right| = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < i_2 < \ldots < i_k \le n} \left| A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k} \right|$$

or in a more detailed form

$$|A_1 \cup A_2 \cup \ldots \cup A_n| = \sum_{i=1}^{n} |A_i| - \sum_{1 \le i_1 < i_2 \le n} |A_{i_1} \cap A_{i_2}| + \sum_{1 \le i_1 < i_2 < i_3 \le n} |A_{i_1} \cap A_{i_2} \cap A_{i_3}| - \ldots + (-1)^{n+1} |A_1 \cap A_2 \cap \ldots \cap A_n| .$$

Here, e.g. the sum $\sum_{1 \le i_1 < i_2 \le n} |A_{i_1} \cap A_{i_2}|$ is taken over all pairs (i_1, i_2) of integers between 1 and n inclusive such that $i_1 < i_2$.

To prove this, we introduce the concept of *multiset*. Let X be a large set so that every set we may want to consider is a subset of X. A *multiset* is a function assigning a positive, negative or zero integer to every element of X; briefly, a function M from X to $\mathbb{Z}$, $M: X \longrightarrow \mathbb{Z}$. Intuitively, M is a "set" for which the things in X may be in M with various "multiplicities"; $M(x)$ is the *multiplicity of* x *in* M. E.g., with $X = \mathbb{N}$, we may consider the multiset M for which $M(n) = 0$ for all $n \ge 5$, and $M(0)=1$, $M(1)=-4$, $M(2)=0$, $M(3)=1$, $M(4)=2$. We consider only *finite* multisets, that is, ones in which only finitely many elements have a multiplicity different from 0.

A simple notation for concrete multisets follows the notation for functions; the multiset in the example may be denoted by

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & -4 & 0 & 1 & 2 \end{pmatrix} . \qquad (6)$$

It is understood that for any $x \varepsilon X$ not in the upper row of the notation, the multiplicity is 0.

Any ordinary set A (a subset of X) is considered as a multiset $\dot{A}$ for which

$$\dot{A}(x) = \begin{cases} 1 & \text{if } x \varepsilon A \\ 0 & \text{if } x \notin A \end{cases}.$$

In other words, $\dot{A}$ is the characteristic function of A as a subset of X. E.g., if $A = \{0, 2, 5\}$, then $\dot{A}$ is

$$\dot{A} = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

The *cardinality of* a multiset M, $|M|$, is, by definition, the sum of the multiplicities of the elements: $|M| = \sum_{x \varepsilon X} M(x)$; since we assume that only finitely many $M(x)$ are different from 0, the sum is a well-defined integer. E.g., in the example, $|M| = 0$, although M is far from being the same as the empty set.

Note that for a finite set A, the usual cardinality of A and the cardinality of it as a multiset are the same: $|A| = |\dot{A}|$.

We define *addition* of multisets by simply adding multiplicities: the multiset $M + N$ is defined by the equality

$$(M + N)(x) \underset{\text{def}}{=} M(x) + N(x)$$

In other words, the multiplicity of an element x in the sum-multiset $M+N$ is, by definition, the sum of the multiplicities of x in M and N.

E.g., for M as above, and for $A = \{0, 2, 5\}$, $M + \dot{A}$ is the multiset

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 2 & -4 & 1 & 1 & 2 & 1 \end{pmatrix}.$$

If a is an integer, $a \cdot M$ or more simply aM, (*scalar multiplication*) is the multiset for which

$$(aM)(x) \underset{\text{def}}{=} aM(x) .$$

E.g., the multiset $(-1)\dot{A}$ for A as above is $\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ -1 & 0 & -1 & 0 & 0 & -1 \end{pmatrix}$.

$-M$ means $(-1)M$, and $M - N$ means $M + (-1)N$. E.g., with M and A as before, $M - \dot{A}$ is $\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 0 & -4 & -1 & 1 & 2 & -1 \end{pmatrix}$.

The usual rules concerning addition and scalar multiplication (commutativity, associativity, etc) familiar from linear algebra are valid for addition and scalar multiplication of multisets, since they are inherited from those operations on numbers.

We have the following rules connecting cardinality and the operations just introduced:

$$|M + N| = |M| + |N| ,$$

$$|aM| = a|M| .$$

These are immediate from the definitions. As a consequence, the cardinality of a linear combination of multisets is the corresponding linear combination of the cardinalities of the terms.

The main point is the following equality of multisets: for any (ordinary) sets $A_1, A_2, \ldots, A_n$, we have

$$\left(\bigcup_{i=1}^{n} A_i\right)^{\cdot} = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < i_2 < \ldots < i_k \le n} (A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k})^{\cdot} . \quad (7)$$

To prove this, we take an arbitrary $x \varepsilon X$, and show that the multiplicity of x in the left-hand side equals the multiplicity of x in the right-hand side. If x does not belong to any of the A_i, that is, the multiplicity of x in the left side is 0, then it does not belong to any of the sets involved in the right side either, and thus its multiplicity on the right, being a sum of 0's,

is also 0. Let us then assume that x does belong to at least one A_i; thus, the multiplicity of x on the left is 1. Let those indices $\ell = 1, \ldots, n$ for which $x \varepsilon A_\ell$ be $\ell_1 < \ell_2 < \ldots < \ell_m$; in particular the number of these ℓs is m; $m \geq 1$. Let us also write

$$L \underset{\text{def}}{=} \{\ell_1, \ell_2, \ldots \ell_m\}.$$

Take an arbitrary selection $i_1 < i_2 < \ldots < i_k$ of indices between 1 and n (inclusive), and ask what the multiplicity

$$(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k})^{\cdot}(x) \tag{8}$$

is. Clearly, this is 1 or 0 depending on whether x does or does not belong to the set $A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}$. On the other hand, x belongs to the latter set if and only if x belongs to each one of the sets A_{i_1}, A_{i_2}, $\ldots A_{i_k}$, that is, if each of i_1, i_2, $\ldots$, i_k is the same as one of ℓ_1, ℓ_2, $\ldots \ell_m$, that is, if

$$\{i_1, i_2, \ldots i_k\} \subseteq L. \tag{9}$$

We have shown that (8) is equal to 1 if (9) holds; otherwise (8) is 0. Therefore, with a fixed k between 1 and n, the sum

$$\sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} (A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k})^{\cdot}(x)$$

equals the number of selections $i_1 < i_2 < \ldots < i_k$ for which (9) holds. But this number is nothing but the number of k-subsets of L, and this is $\binom{m}{k}$. It follows that the right-hand side of (7), when evaluated at x, equals $\sum_{k=1}^{n} (-1)^{k+1} \binom{m}{k}$, which is the same as $\sum_{k=1}^{m} (-1)^{k+1} \binom{m}{k}$, since $\binom{m}{k} = 0$ for $k > m$. By (5) and $m \geq 1$, the last sum is equal to 1. We have shown that the multiplicity of x on the right in (7) is 1; since the multiplicity on the left is also 1, we have proved (7).

Having proved (7), we may take the cardinality of the two multisets in (7). The cardinality of the left side is the same as the cardinality of the ordinary union-set. The cardinality of the right side may be taken term by term, as we pointed out above. The cardinalities of the intersection-multisets are just the cardinalities of the intersections as sets. We get the right-hand side expression in the framed equality; that equality is thus proved.

Note that the cases of two and of three sets stated earlier are the special cases of the general formula for $n = 2$ and $n = 3$.