Lecture 8

Matrix-Tree Ingredients

This lecture introduces some ideas that we will need for the proof of the Matrix-Tree Theorem. Many of them should be familiar from Foundations of Pure Mathematics or Algebraic Structures.

Reading:
The material about permutations and the determinant of a matrix presented here is pretty standard and can be found in any number of places: the *Wikipedia* articles on *Determinant* (especially the section on $n \times n$ matrices) and *Permutation* are not bad places to start.

The remaining ingredient for the proof of the Matrix-Tree theorems is the *Principle of Inclusion/Exclusion*. It is covered in the first year module Foundations of Pure Mathematics, but it is also a standard technique in Combinatorics and so is discussed in many introductory books[1]. The *Wikipedia* article is, again, a good place to start. Finally, as a convenience for students from outside the School of Mathematics, I have included an example in Section 8.4.4 and an Appendix, Section 8.5, that provides full details of all the proofs.

8.1 Lightning review of permutations

First we need a few facts about permutations that you should have learned earlier in your degree.

Definition 8.1. *A* ***permutation*** *on n-objects is a bijection σ from the set $\{1, \ldots, n\}$ to itself.*

We'll express permutations as follows

$$\sigma = \begin{pmatrix} 1 & 2 & \ldots & n \\ \sigma(1) & \sigma(2) & \ldots & \sigma(n) \end{pmatrix}$$

where $\sigma(j)$ is the image of j under the bijection σ.

Definition 8.2. *The set* $\mathrm{fix}(\sigma)$ *is defined as*

$$\mathrm{fix}(\sigma) = \{j \mid \sigma(j) = j\}.$$

[1] See, for example, Dossey, Otto, Spence, and Vanden Eynden (2006), *Discrete Mathematics* or, for a short, clear account, Anderson (1974), *A First Course in Combinatorial Mathematics*.

8.1.1 The Symmetric Group S_n

One can turn the set of all permutations on n objects into a group by using *composition* (applying one function to the output of another) of permutations as the group multiplication. The resulting group is called the **symmetric group on n objects** or S_n and it has the following properties.

- The identity element is the permutation in which $\sigma(j) = j$ for all $1 \leq j \leq n$.
- S_n has $n!$ elements.

8.1.2 Cycles and sign

Definition 8.3 (Cycle permutations)**.** *A **cycle** is a permutation σ specified by a sequence of distinct integers $i_1, i_2, \ldots, i_\ell \in \{1, \ldots, n\}$ with the properties that*

- $\sigma(j) = j$ *if* $j \notin \{i_1, i_2, \ldots, i_\ell\}$
- $\sigma(i_j) = i_{j+1}$ *for* $1 \leq j < \ell$
- $\sigma(i_\ell) = i_1$.

*Here ℓ is the **length** of the cycle and a cycle with $\ell = 2$ is called a **transposition**.*

We'll express the cycle σ specified by the sequence $i_1, i_2, \ldots, i_\ell$ with the notation

$$\sigma = (i_1, i_2, \ldots, i_\ell).$$

and we'll say that two cycles $\sigma_1 = (i_1, i_2, \ldots, i_{\ell_1})$ and $\sigma_2 = (j_1, j_2, \ldots, j_{\ell_2})$ are **disjoint** if

$$\{i_1, \ldots i_{\ell_1}\} \cap \{j_1, \ldots j_{\ell_2}\} \;=\; \emptyset.$$

The main point about cycles is that they're like the "prime factors" of permutations in the following sense:

Proposition 8.4. *A permutation has a unique (up to reordering of the cycles) representation as a product of disjoint cycles.*

This representation is often referred to as the *cycle decomposition* of the permutation.

Finally, given the cycle decomposition of a permutation one can define a function that we will need in the next section.

Definition 8.5 (Sign of a permutation)**.** *The function* $\operatorname{sgn} : S_n \to \{\pm 1\}$ *can be computed as follows:*

- *If σ is the identity permutation, then* $\operatorname{sgn}(\sigma) = 1$.
- *If σ is a cycle of length ℓ then* $\operatorname{sgn}(\sigma) = (-1)^{\ell-1}$.
- *If σ has a decomposition into $k \geq 2$ disjoint cycles whose lengths are $\ell_1, \ldots, \ell_k$ then*

$$\operatorname{sgn}(\sigma) = (-1)^{L-k} \qquad \textit{where} \qquad L = \sum_{j=1}^{k} \ell_j.$$

This definition of $\operatorname{sgn}(\sigma)$ is equivalent to one that you may know from other courses:

$$\operatorname{sgn}(\sigma) = \begin{cases} 1 & \text{If } \sigma \text{ is the product of an even number of transpositions} \\ -1 & \text{otherwise} \end{cases}$$

8.2 Using graphs to find the cycle decomposition

Lest we forget graph theory completely, I'd like to conclude our review of permutations by constructing a certain graph that makes it easy to read-off the cycle decomposition of a permutation. The same construction establishes a bijection that maps permutations $\sigma \in S_n$ to subgraphs of K_n whose strongly-connected components are either isolated vertices or disjoint, directed cycles. This bijection is another key ingredient in the proof of Tutte's Matrix Tree Theorem.

Definition 8.6. *Given a permutation $\sigma \in S_n$, define the directed graph G_σ to have vertex set $V = \{1, \ldots, n\}$ and edge set*

$$E = \{(j, \sigma(j)) \mid j \in V \text{ and } \sigma(j) \neq j\}.$$

The following proposition then makes it easy to find the cycle decomposition of a permutation σ:

Proposition 8.7. *The cycle $(i_1, i_2, \ldots, i_l)$ appears in the cycle decomposition of σ if and only if the directed cycle defined by the vertex sequence*

$$(i_1, i_2, \ldots, i_l, i_1)$$

is a subgraph of G_σ.

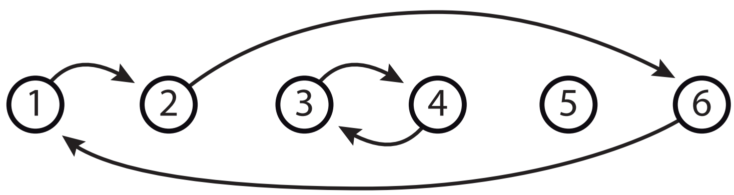
Example 8.8 (Graphical approach to cycle decomposition)**.** *Consider a permutation from S_6 given by*

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 6 & 4 & 3 & 5 & 1 \end{pmatrix}.$$

Then the digraph G_σ has vertex set $V = \{1, \ldots, 6\}$ and edge set

$$E = \{(1,2),\ (2,6),\ (3,4),\ (4,3),\ (6,1)\}.$$

A diagram for this graph appears below and clearly includes two disjoint directed cycles



Thus our permutation has $\mathrm{fix}(\sigma) = \{5\}$ *and its cycle decomposition is*

$$\sigma = (1, 2, 6)(3, 4) = (3, 4)(1, 2, 6) = (4, 3)(2, 6, 1),$$

where I have included some versions where the order of the two disjoint cycles is switched and one where the terms within the cycle are written in a different (but equivalent) order.

With a little more work (that's left to the reader) one can prove that this graphical approach establishes a bijection between S_n and that family of subgraphs of K_n which consists of unions of disjoint cycles. The bijection sends the identity permutation to the subgraph consisting of n isolated vertices and sends a permutation σ that is the product of $k \geq 1$ disjoint cycles

$$\sigma = (i_{1,1}, \ldots, i_{1,\ell_1}) \cdots (i_{k,1}, \ldots, i_{k,\ell_k}) \tag{8.1}$$

to the subgraph G_σ that has vertex set $V = \{v_1, \ldots, v_n\}$ and whose edges are those that appear in the k disjoint, directed cycles $C_1, \ldots, C_k$, where C_j is the cycle specified by the vertex sequence

$$\left(v_{i_{j,1}}, \ldots, v_{i_{j,\ell_j}}, v_{i_{j,1}} \right). \tag{8.2}$$

In Eqns. (8.1) and (8.2) the notation $i_{j,r}$ is the vertex number of the r-th vertex in the j-th cycle, while ℓ_j is the length of the j-th cycle.

8.3 The determinant is a sum over permutations

The Matrix-Tree theorems relate the number of spanning trees or arborescences to the determinant of a matrix and so it should not be surprising that another of our key ingredients is a fact about determinants. The standard recursive approach to computing determinants—in which one computes the determinant of an $n \times n$ matrix as a sum over the determinants of $(n-1) \times (n-1)$ submatrices—is equivalent to a sum over permutations:

Proposition 8.9. *If A is an $n \times n$ matrix then*

$$\det(A) = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) \prod_{j=1}^{n} A_{j\sigma(j)}. \tag{8.3}$$

Some of you may have encountered this elsewhere, though most will be meeting it for the first time. I won't prove it, as that would be too much of a diversion from graphs, but Eqn. (8.3) has the blessing of *Wikipedia*, where it is attributed to Leibniz, and proofs appear in many undergraduate algebra texts[2]. The very keen reader could also construct an inductive proof herself, starting from the familiar recursive formula.

Finally, I'll demonstrate that it works for the two smallest nontrivial examples.

[2] I found one in I.N. Herstein (1975), *Topics in Algebra*, 2nd ed., Wiley.

Example 8.10 (2×2 matrices)**.** *First we need a list of the elements of S_2 and their signs:*

Name	σ	sgn(σ)
σ_1	$\begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$	1
σ_2	$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$	-1

Then we can compute the determinant of a 2×2 matrix in the usual way

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

and then again using Eqn (8.3)

$$\begin{aligned}
\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} &= \sum_{k=1}^{2} \operatorname{sgn}(\sigma_k) \prod_{j=1}^{n} a_{j\sigma_k(j)} \\
&= \operatorname{sgn}(\sigma_1) \times a_{1\sigma_1(1)}a_{2\sigma_1(2)} + \operatorname{sgn}(\sigma_2) \times a_{1\sigma_2(1)}a_{2\sigma_2(2)} \\
&= (1) \times a_{11}a_{22} + (-1) \times a_{12}a_{21} \\
&= a_{11}a_{22} - a_{12}a_{21}
\end{aligned}$$

Example 8.11 (3×3 matrices)**.** *Table 8.1 lists the elements of S_3 in the form we'll need. First, we compute the determinant in the usual way, a tedious but straightforward business.*

$$\begin{aligned}
&\det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \\
&a_{11} \times \det \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} - a_{12} \times \det \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix} + a_{13} \times \det \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \\
&= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}
\end{aligned}$$

Then we calculate again using Eqn (8.3) and the numbering scheme for the elements of S_3 that's shown in Table 8.1.

$$\begin{aligned}
&\det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \sum_{k=1}^{6} \operatorname{sgn}(\sigma_k) \prod_{j=1}^{n} a_{j\sigma_k(j)} \\
&= \operatorname{sgn}(\sigma_1) \times a_{1\sigma_1(1)}a_{2\sigma_1(2)}a_{3\sigma_1(3)} + \cdots + \operatorname{sgn}(\sigma_6) \times a_{1\sigma_6(1)}a_{2\sigma_6(2)}a_{3\sigma_6(3)} \\
&= (1) \times a_{11}a_{22}a_{33} + (-1) \times a_{11}a_{23}a_{32} + (-1) \times a_{12}a_{21}a_{33} \\
&+ (1) \times a_{12}a_{23}a_{31} + (1) \times a_{13}a_{21}a_{32} + (-1) \times a_{13}a_{22}a_{31} \\
&= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}
\end{aligned}$$

σ_k	$\text{fix}(\sigma_k)$	Cycle Decomposition	$\text{sgn}(\sigma_k)$
$\sigma_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$	$\{1,2,3\}$	–	1
$\sigma_2 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$	$\{1\}$	$(2,3)$	-1
$\sigma_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$	$\{3\}$	$(1,2)$	-1
$\sigma_4 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$	$\emptyset$	$(1,2,3)$	1
$\sigma_5 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$	$\emptyset$	$(3,2,1)$	1
$\sigma_6 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$	$\{2\}$	$(1,3)$	-1

Table 8.1: *The cycle decompositions of all the elements in S_3, along with the associated functions* $\text{sgn}(\sigma)$ *and* $\text{fix}(\sigma)$.

8.4 The Principle of Inclusion/Exclusion

The remaining ingredient for the proof of the Matrix-Tree theorems is the *Principle of Inclusion/Exclusion*. As it is covered in a core first year module, none of the proofs in the rest of this lecture are examinable.

8.4.1 A familiar example

Suppose we have some finite "universal" set U and two subsets, $X_1 \subseteq U$ and $X_2 \subseteq U$. If the subsets are disjoint then it's easy to work out the number of elements in their union:

$$X_1 \cap X_2 = \emptyset \;\Rightarrow\; |X_1 \cup X_2| = |X_1| + |X_2|.$$

The case where the subsets have a non-empty intersection provides the simplest instance of the Principle of Inclusion/Exclusion. You may already know a similar result from Probability.

Lemma 8.12 (Inclusion/Exclusion for two sets)**.** *If X_1 and X_2 are finite sets then*

$$|X_1 \cup X_2| = |X_1| + |X_2| - |X_1 \cap X_2|. \tag{8.4}$$

Note that this formula, which is illustrated in Figure 8.1, works even $X_1 \cap X_2 = \emptyset$, as then $|X_1 \cap X_2| = 0$. The proof of this lemma appears in the Appendix, in Section 8.5.1.

8.4.2 Three subsets

Before moving to the general case, let's consider one more small example, this time with three subsets X_1, X_2 and X_3: we can handle this case by clever use

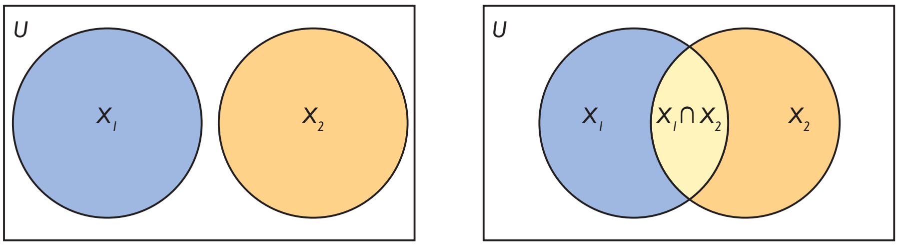


Figure 8.1: *In the example at left* $X_1 \cap X_2 = \emptyset$*, so* $|X_1 \cup X_2| = |X_1| + |X_2|$*, but in the example at right* $X_1 \cap X_2 \neq \emptyset$ *and so* $|X_1 \cup X_2| = |X_1| + |X_2| - |X_1 \cap X_2| < |X_1| + |X_2|$.

of Lemma 8.12 from the previous section. If we regard $(X_1 \cup X_2)$ as a single set and X_3 as a second set, then Eqn. (8.4) says

$$\begin{aligned}|(X_1 \cup X_2) \cup X_3| &= |(X_1 \cup X_2)| + |X_3| - |(X_1 \cup X_2) \cap X_3| \\ &= (|X_1| + |X_2| - |X_1 \cap X_2|) + |X_3| - |(X_1 \cup X_2) \cap X_3| \\ &= |X_1| + |X_2| + |X_3| - |X_1 \cap X_2| - |(X_1 \cup X_2) \cap X_3|\end{aligned}$$

Focusing on the final term, we can use standard relations about unions and intersections to say

$$(X_1 \cup X_2) \cap X_3 = (X_1 \cap X_3) \cup (X_2 \cap X_3).$$

Then, applying Eqn. (8.4) to the pair of sets $(X_1 \cap X_3)$ and $(X_2 \cap X_3)$, we obtain

$$\begin{aligned}|(X_1 \cup X_2) \cap X_3| &= |(X_1 \cap X_3) \cup (X_2 \cap X_3)| \\ &= |X_1 \cap X_3| + |X_2 \cap X_3| - |(X_1 \cap X_3) \cap (X_2 \cap X_3)| \\ &= |X_1 \cap X_3| + |X_2 \cap X_3| - |X_1 \cap X_2 \cap X_3|\end{aligned}$$

where, in going from the second line to the third, we have used

$$(X_1 \cap X_3) \cap (X_2 \cap X_3) = X_1 \cap X_2 \cap X_3.$$

Finally, putting all these results together, we obtain the analogue of Eqn. (8.4) for three subsets:

$$\begin{aligned}|(X_1 \cup X_2) \cup X_3| &= |X_1| + |X_2| + |X_3| - |X_1 \cap X_2| - |(X_1 \cup X_2) \cap X_3| \\ &= |X_1| + |X_2| + |X_3| - |X_2 \cap X_3| \\ &\quad - (|X_1 \cap X_3| + |X_2 \cap X_3| - |(X_1 \cap X_2 \cap X_3|) \\ &= (|X_1| + |X_2| + |X_3|) \\ &\quad - (|X_1 \cap X_2| + |X_1 \cap X_3| + |X_2 \cap X_3|) \\ &\quad + |X_1 \cap X_2 \cap X_3|. \qquad (8.5)\end{aligned}$$

Figure 8.2 helps make sense of this formula and prompts the following observations:

- Elements of $X_1 \cup X_2 \cup X_3$ that belong to exactly one of the X_j are counted exactly once by the sum $(|X_1| + |X_2| + |X_3|)$ and do not contribute to any of the terms involving intersections.

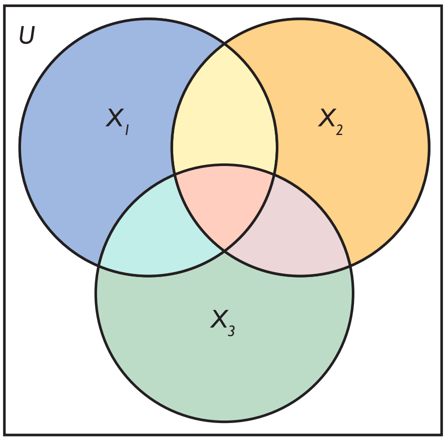


Figure 8.2: *In the diagram above all of the intersections appearing in Eqn. (8.5) are nonempty.*

- Elements of $X_1 \cup X_2 \cup X_3$ that belong to exactly two of the X_j are double-counted by the sum, $(|X_1| + |X_2| + |X_3|)$, but this double-counting is corrected by the term involving two-fold intersections.
- Finally, elements of $X_1 \cup X_2 \cup X_3$ that belong to all three of the sets are triple-counted by the initial sum $(|X_1| + |X_2| + |X_3|)$. This triple-counting is then completely cancelled by the term involving two-fold intersections. Then, finally, this cancellation is repaired by the final term, which counts each such element once.

8.4.3 The general case

The *Principle of Inclusion/Exclusion* generalises the results in Eqns. (8.1) and (8.5) to unions of arbitrarily many subsets.

Theorem 8.13 (The Principle of Inclusion/Exclusion)**.** *If U is a finite set and $\{X_j\}_{j=1}^n$ is a collection of n subsets, then*

$$\begin{aligned}
\left|\bigcup_{j=1}^n X_j\right| &= |X_1 \cup \cdots \cup X_n| \\
&= |X_1| + \cdots + |X_n| \\
&\quad - |X_1 \cap X_2| - \cdots - |X_{n-1} \cap X_n| \\
&\quad + |X_1 \cap X_2 \cap X_3| + \cdots + |X_{n-2} \cap X_{n-1} \cap X_n| \\
&\quad \vdots \\
&\quad + (-1)^{m-1} \sum_{1 \le i_1 \le \cdots \le i_m \le n} |X_{i_1} \cap \cdots \cap X_{i_m}| \\
&\quad \vdots \\
&\quad + (-1)^{n-1} |X_1 \cap \cdots \cap X_n| \qquad (8.6)
\end{aligned}$$

or, more concisely,

$$|X_1 \cup \cdots \cup X_n| \;=\; \sum_{I \subseteq \{1,\ldots,n\},\, I \neq \emptyset} (-1)^{|I|-1} \left| \bigcap_{i \in I} X_i \right| \tag{8.7}$$

The proof of this result appears in the Appendix, in Section 8.5.2 below.

8.4.4 An example

How many of the integers n with $1 \leq n \leq 150$ are coprime to 70? This is a job for the Principle of Inclusion/Exclusion. First note that the prime factorization of 70 is $70 = 2 \times 5 \times 7$. Now consider a universal set $U = \{1, \ldots, 150\}$ and the three subsets X_1, X_2 and X_3 consisting of multiples of 2, 5 and 7, respectively. A member of U that shares a prime factor with 70 belongs to at least one of the X_j and so the number we're after is

$$\begin{aligned}
|U| - |X_1 \cup X_2 \cup X_3| &= |U| - (|X_1| + |X_2| + |X_3|) \\
&\quad + (X_1 \cap X_2| + |X_1 \cap X_3| + |X_2 \cap X_3|) \\
&\quad - \; |X_1 \cap X_2 \cap X_3|. \\
&= 150 - (75 + 30 + 21) + (15 + 10 + 4) - 2 \\
&= 150 - 126 + 29 - 2 \\
&= 51
\end{aligned} \tag{8.8}$$

where I have used the numbers in Table 8.2 which lists the various cardinalities that we need.

Set	Description	Cardinality
X_1	multiples of 2	75
X_2	multiples of 5	30
X_3	multiples of 7	21
$X_1 \cap X_2$	multiples of 10	15
$X_1 \cap X_3$	multiples of 14	10
$X_2 \cap X_3$	multiples of 35	4
$X_1 \cap X_2 \cap X_3$	multiples of 70	2

Table 8.2: *The sizes of the various intersections needed for the calculation in Eqn. (8.8).*

8.5 Appendix: Proofs for Inclusion/Exclusion

The proofs in this section will not appear on the exam, but are provided for those who are interested or for whom the subject is new.

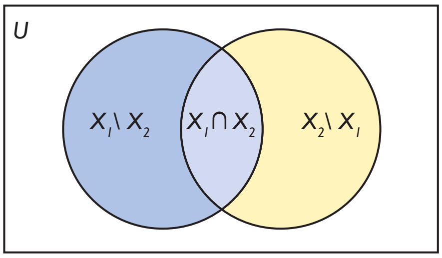


Figure 8.3: *Here $X_1 \backslash X_2$ and $X_1 \cap X_2$ are shown in shades of blue, while $X_2 \backslash X_1$ is in yellow.*

8.5.1 Proof of Lemma 8.12, the case of two sets

Recall that X_1 and X_2 are subsets of some universal set U and that we seek to prove that $|X_1 \cup X_2| = |X_1| + |X_2| - |X_1 \cap X_2|$.

Proof. Note that the sum $|X_1|+|X_2|$ counts each member of the intersection $X_1 \cap X_2$ twice, once as a member of X_1 and then again as a member of X_2. Subtracting $|X_1 \cap X_2|$ corrects for this double-counting. Alternatively, for those who prefer proofs that look more like calculations, begin by defining

$$X_1 \backslash X_2 = \{x \in U \mid x \in X_1, \text{ but } x \notin X_2\}.$$

Then, as is illustrated in Figure 8.3, $X_1 = (X_1 \backslash X_2) \cup (X_1 \cap X_2)$. Further, the sets $X_1 \backslash X_2$ and $X_1 \cap X_2$ are disjoint by construction, so

$$|X_1| = |X_1 \backslash X_2| + |X_1 \cap X_2| \qquad \text{or} \qquad |X_1 \backslash X_2| = |X_1| - |X_1 \cap X_2|. \tag{8.9}$$

Similarly, $X_1 \backslash X_2$ and X_2 are disjoint and $X_1 \cup X_2 = (X_1 \backslash X_2) \cup X_2$ so

$$\begin{aligned} |X_1 \cup X_2| &= |X_1 \backslash X_2| + |X_2| \\ &= |X_1| - |X_1 \cap X_2| + |X_2| \\ &= |X_1| + |X_2| - |X_1 \cap X_2| \end{aligned}$$

where, in passing from the first line to the second, we have used (8.9). The last line is the result we were trying to prove, so we are finished. □

8.5.2 Proof of Theorem 8.13

One can prove this result in at least two ways:

- by induction, with a calculation that is essentially the same as the one used to obtain the $n = 3$ case—Eqn. (8.5)—from the $n = 2$ one—Eqn. (8.4);
- by showing that each $x \in X_1 \cup \cdots \cup X_n$ contributes exactly one to the sum on the right hand side of Eqn. (8.7).

The first approach is straightforward, if a bit tedious, but the second is more interesting and is the one discussed here.

The key idea is to think of the the elements of $X_1 \cup \cdots \cup X_n$ individually and ask what each one contributes to the sum in Eqn. (8.7). Suppose that an element $x \in X_1 \cup \cdots \cup X_n$ belongs to exactly ℓ of the subsets, with $1 \leq \ell \leq n$: we will prove that x makes a net contribution of 1. For the sake of concreteness, we'll say $x \in X_{i_1}, \ldots, X_{i_\ell}$ where $i_1, \ldots, i_\ell$ are distinct elements of $\{1, \ldots, n\}$.

- As we've assumed that x belongs to exactly ℓ of the subsets X_j, it contributes a total of ℓ to the first row, $|X_1| + \cdots + |X_n|$, of the long sum in Eqn. (8.6).
- Further, x contributes a total of $-\binom{\ell}{2}$ to the sum in the row involving two-way intersections
$$-|X_1 \cap X_2| - \cdots - |X_{n-1} \cap X_n|.$$
To see this, note that if $x \in X_j \cap X_k$ then both j and k must be members of the set $\{i_1, \ldots, i_\ell\}$.
- Similar arguments show that if $k \leq \ell$, then x contributes a total of
$$(-1)^{k-1}\binom{\ell}{k} \;=\; (-1)^{k-1}\left(\frac{\ell!}{k!\,(\ell-k)!}\right)$$
to the sum in the row of Eqn. (8.6) that involves k-fold intersections.
- Finally, for $k > \ell$ there are no k-fold intersections that contain x and so x makes a contribution of zero to the corresponding rows in Eqn. (8.6).

Putting these observations together we see that x make a net contribution of
$$\ell - \binom{\ell}{2} + \binom{\ell}{3} - \ldots + (-1)^{\ell-1}\binom{\ell}{\ell} \tag{8.10}$$

This sum can be made to look more familiar by considering the following application of the Binomial Theorem:
$$\begin{aligned} 0 &= (1-1)^\ell \\ &= \sum_{j=0}^{\ell} (-1)^j (1)^{\ell-j}\binom{\ell}{j} \\ &= 1 - \ell + \binom{\ell}{2} - \binom{\ell}{3} + \ldots + (-1)^\ell\binom{\ell}{\ell}. \end{aligned}$$

Thus

$$0 \ = \ 1 - \left[\ell - \binom{\ell}{2} + \binom{\ell}{3} - \ \ldots \ + (-1)^{\ell-1}\binom{\ell}{\ell} \right]$$

or

$$\ell - \binom{\ell}{2} + \binom{\ell}{3} - \ \ldots \ + (-1)^{\ell-1}\binom{\ell}{\ell} \ = \ 1.$$

The left hand side here is the same as the sum in Eqn. (8.10) and so we've established that any x which belongs to exactly ℓ of the subsets X_j makes a net contribution of 1 to the sum on the right hand side of Eqn. (8.7). And as every $x \in X_1 \cup \cdots \cup X_n$ must belong to at least one of the X_j, this establishes the Principle of Inclusion/Exclusion.

8.5.3 Alternative proof

Students who like proofs that look more like calculations may prefer to reformulate the arguments from the previous section in terms of *characteristic functions* (sometimes also called *indicator functions*) of sets. If we define $\mathbb{1}_X : U \to \{0,1\}$ by

$$\mathbb{1}_X(s) = \begin{cases} 1 & \text{if } s \in X \\ 0 & \text{otherwise} \end{cases}$$

then we can calculate $|X|$ for a subset $X \subseteq U$ as follows:

$$\begin{aligned} |X| &= \sum_{x \in U} \mathbb{1}_X(x) \\ &= \left(\sum_{x \in X} \mathbb{1}_X(x) \right) + \left(\sum_{x \notin X} \mathbb{1}_X(x) \right) \\ &= \sum_{x \in X} \mathbb{1}_X(x) \end{aligned} \tag{8.11}$$

where, in passing from the second to third lines, I have dropped the second sum because all its terms are zero.

Then the Principle of Inclusion/Exclusion is equivalent to

$$\begin{aligned} \sum_{x \in X_1 \cup \cdots \cup X_n} \mathbb{1}_{X_1 \cup \cdots \cup X_n}(x) &= \sum_{I \subseteq \{1,\ldots,n\},\, I \neq \emptyset} (-1)^{|I|-1} \left| \bigcap_{i \in I} X_i \right| \\ &= \sum_{I \subseteq \{1,\ldots,n\},\, I \neq \emptyset} (-1)^{|I|-1} \sum_{x \in \bigcap_{i \in I} X_i} \mathbb{1}_{\bigcap_{i \in I} X_i}(x) \\ &= \sum_{k=1}^{n} (-1)^{k-1} \sum_{I \subseteq \{1,\ldots,n\},\, |I| = k} \left(\sum_{x \in \bigcap_{i \in I} X_i} \mathbb{1}_{\bigcap_{i \in I} X_i}(x) \right) \end{aligned}$$

which I have obtained by using of Eqn. (8.11) to replace terms in Eqn. (8.7) with the corresponding sums of values of characteristic functions.

We can then rearrange the expression on the right, first expanding the ranges of the sums over elements of k-fold intersections (this doesn't change the result since $\mathbb{1}_X(x) = 0$ for $x \notin X$) and then interchanging the order of summation so that the sum over elements comes first. This calculation proves that the Principle of Inclusion/Exclusion is equivalent to the following:

$$\begin{aligned}
\sum_{x \in X_1 \cup \cdots \cup X_n} & \mathbb{1}_{X_1 \cup \cdots \cup X_n}(x) \\
&= \sum_{k=1}^{n} (-1)^{k-1} \sum_{I \subseteq \{1,\ldots,n\},\, |I|=k} \left(\sum_{x \in X_1 \cup \cdots \cup X_n} \mathbb{1}_{\bigcap_{i \in I} X_i}(x) \right) \\
&= \sum_{x \in X_1 \cup \cdots \cup X_n} \sum_{k=1}^{n} (-1)^{k-1} \sum_{I \subseteq \{1,\ldots,n\},\, |I|=k} \mathbb{1}_{\bigcap_{i \in I} X_i}(x) \qquad (8.12)
\end{aligned}$$

Arguments similar to those in Section 8.5.2 then establish the following results, the last of which, along with Eqn. (8.12), proves Theorem 8.13.

Proposition 8.14. *If an element $x \in X_1 \cup \cdots \cup X_n$ belongs to exactly ℓ of the sets $\{X_j\}_{j=1}^n$ then for $k \leq \ell$ we have*

$$\sum_{I \subseteq \{1,\ldots,n\},\, |I|=k} \mathbb{1}_{\bigcap_{i \in I} X_i}(x) = \binom{\ell}{k} = \frac{\ell!}{k!\,(\ell-k)!}$$

while if $k > \ell$

$$\sum_{I \subseteq \{1,\ldots,n\},\, |I|=k} \mathbb{1}_{\bigcap_{i \in I} X_i}(x) = 0$$

Proposition 8.15. *For an element $x \in X_1 \cup \cdots \cup X_n$ we have*

$$\sum_{k=1}^{n} (-1)^{k-1} \sum_{I \subseteq \{1,\ldots,n\},\, |I|=k} \mathbb{1}_{\bigcap_{i \in I} X_i}(x) \;=\; \sum_{k=1}^{n} (-1)^{k-1} \binom{\ell}{k} \;=\; 1.$$

Lemma 8.16. *The characteristic function $\mathbb{1}_{X_1 \cup \cdots \cup X_n}$ of the set $X_1 \cup \cdots \cup X_n$ satisfies*

$$\mathbb{1}_{X_1 \cup \cdots \cup X_n}(x) \;=\; \sum_{k=1}^{n} (-1)^{k-1} \sum_{I \subseteq \{1,\ldots,n\},\, |I|=k} \mathbb{1}_{\bigcap_{i \in I} X_i}(x).$$