Partitions and counting

Partitions

The problems to be studied in this chapter can be most conveniently described in terms of partitions of a set. A *partition* of a set $\mathcal{U}$ is a subdivision of the set into subsets that are disjoint and exhaustive, i.e., every element of $\mathcal{U}$ must belong to one and only one of the subsets. The subsets A_i in the partition are called *cells*. Thus $[A_1, A_2, \ldots, A_r]$ is a partition of $\mathcal{U}$ if two conditions are satisfied: (1) $A_i \cap A_j = \emptyset$ if $i \neq j$ (the cells are disjoint) and (2) $A_1 \cup A_2 \cup \ldots \cup A_r = \mathcal{U}$ (the cells are exhaustive).

Example 3.1 If $\mathcal{U} = \{a, b, c, d, e\}$, then $[\{a, b\}, \{c, d, e\}]$ and $[\{b, c, e\}, \{a\}, \{d\}]$ and $[\{a\}, \{b\}, \{c\}, \{d\}, \{e\}]$ are three different partitions of $\mathcal{U}$. The last is a partition into unit sets. ◇

The process of going from a fine to a less fine analysis of a set of logical possibilities is actually carried out by means of a partition. For example, let us consider the logical possibilities for the first three games of the World Series if the Yankees play the Dodgers. We can list the possibilities in terms cf the winner of each game as

$$\{\mathrm{YYY}, \mathrm{YYD}, \mathrm{YDY}, \mathrm{DYY}, \mathrm{DDY}, \mathrm{DYD}, \mathrm{YDD}, \mathrm{DDD}\}.$$

We form a partition by putting all the possibilities with the same number of wins for the Yankees in a single cell,

$$[\{\mathrm{YYY}\}, \{\mathrm{YYD}, \mathrm{YDY}, \mathrm{DYY}\}, \{\mathrm{DDY}, \mathrm{DYD}, \mathrm{YDD}\}, \{\mathrm{DDD}\}].$$

Thus, if we wish the possibilities to be Yankees win three games, win two, win one, win zero, then we are considering a less detailed analysis obtained from the former analysis by identifying the possibilities in each cell of the partition.

If $[A_1, A_2, \ldots, A_r]$ and $[B_1, B_2, \ldots, B_s]$ are two partitions of the same set $\mathcal{U}$, we can obtain a new partition by considering the collection of all subsets of $\mathcal{U}$ of the form $A_i \cap B_j$ (see Exercise 7). This new partition is called the *cross-partition* of the original two partitions.

Example 3.2 A common use of cross-partitions is in the problem of classification. For example, from the set $\mathcal{U}$ of all life forms we can form the partition $[P, A]$ where P is the set of all plants and A is the set of all animals. We may also form the partition $[E, F]$ where E is the set of extinct life forms and F is the set of all existing life forms. The cross-partition

$$[P \cap E, P \cap F, A \cap E, A \cap F]$$

gives a complete classification according to the two separate classifications. $\diamond$

Many of the examples with which we shall deal in the future will relate to processes which take place in stages. It will be convenient to use partitions and cross-partitions to represent the stages of the process. The graphical representation of such a process is, of course, a tree. For example, suppose that the process is such that we learn in succession the truth values of a series of statements relative to a given situation. If $\mathcal{U}$ is the set of logical possibilities for the situation, and p is a statement relative to $\mathcal{U}$, then the knowledge of the truth value of p amounts to knowing which cell of the partition $[P, \tilde{P}]$ contains the actual possibility. Recall that P is the truth set of p, and $\tilde{P}$ is the truth set of $\neg p$. Suppose now we discover the truth value of a second statement q. This information can again be described by a partition, namely, $[Q, \tilde{Q}]$. The two statements together give us information which can be represented by the cross-partition of these two partitions,

$$[P \cap Q, P \cap \tilde{Q}, \tilde{P} \cap Q, \tilde{P} \cap \tilde{Q}].$$

That is, if we know the truth values of p and q, we also know which of the cells of this cross-partition contains the particular logical possibility describing the given situation. Conversely, if we knew which cell contained the possibility, we would know the truth values for the statements p and q.

The information obtained by the additional knowledge of the truth value of a third statement r, having a truth set R, can be represented by the cross-partition of the three partitions $[P, \tilde{P}]$, $[Q, \tilde{Q}]$, $[R, \tilde{R}]$ This cross-partition is

$$[P\cap Q\cap R, P\cap Q\cap\tilde{R}, P\cap\tilde{Q}\cap R, P\cap\tilde{Q}\cap\tilde{R}, \tilde{P}\cap Q\cap R, \tilde{P}\cap Q\cap\tilde{R}, \tilde{P}\cap\tilde{Q}\cap R, \tilde{P}\cap\tilde{Q}\cap\tilde{R}]$$

Notice that now we have the possibility narrowed down to being in one of $8 = 2^3$ possible cells. Similarly, if we knew the truth values of n statements, our partition would have 2^n cells.

If the set $\mathcal{U}$ were to contain 2^{20} (approximately one million) logical possibilities, and if we were able to ask yes-no questions in such a way that the knowledge of the truth value of each question would cut the number of possibilities in half each time, then we could determine in 20 questions any given possibility in the set $\mathcal{U}$. We could accomplish this kind of questioning, for example, if we had a list of all the possibilities and were allowed to ask "Is it in the first half?" and, if the answer is yes, then "Is it in the first one-fourth?", etc. In practice we ordinarily do not have such a list, and we can only approximate this procedure.

Example 3.3 In the familiar radio game of twenty questions it is not unusual for a contestant to try to carry out a partitioning of the above kind. For example, he or she may know that he or she is trying to guess a city. He or she might ask, "Is the city in North America?" and if the answer is yes, "Is it in the United States?" and if yes, "Is it west of the Mississippi?" and if no, "Is it in the New England states?", etc. Of course, the above procedure does not actually divide the possibilities exactly in half each time. The more nearly the answer to each question comes to dividing the possibilities in half, the more certain one can be of getting the answer in twenty questions, if there are at most a million possibilities. ◇

Exercises

1. If $\mathcal{U}$ is the set of integers from 1 to 6, find the cross-partitions of the following pairs of partitions
   (a) $[\{1, 2, 3\}, \{4, 5, 6\}]$ and $[\{1, 4\}, \{2, 3, 5, 6\}]$.
   [Ans. $[\{1\}, \{2, 3\}, \{4\}, \{5, 6\}]$.]
   (b) $[\{1, 2, 3, 4, 5\}, \{6\}]$ and $[\{1, 3, 5\}, \{2, 6\}, \{4\}]$.
2. A coin is thrown three times. List the possibilities according to which side turns up each time. Give the partition formed by

putting in the same cell all those possibilities for which the same number of heads occur.

3. Let p and q be two statements with truth set P and Q. What can be said about the cross-partition of $[P, \tilde{P}]$ and $[Q, \tilde{Q}]$ in the case that

   (a) p implies q.

   [Ans. $P \cap \tilde{Q} = \emptyset$.]

   (b) p is equivalent to q.

   (c) p and q are inconsistent.

4. Consider the set of eight states consisting of Illinois, Colorado, Michigan, New York, Vermont, Texas, Alabama, and California.

   (a) Show that in three "yes" or "no" questions one can identify any one of the eight states.

   (b) Design a set of three "yes" or "no" questions which can be answered independently of each other and which will serve to identify any one of the states.

5. An unabridged dictionary contains about 600,000 words and 3000 pages. If a person chooses a word from such a dictionary, is it possible to identify this word by twenty "yes" or "no" questions? If so, describe the procedure that you would use and discuss the feasibility of the procedure. (One approach is the following. Use 12 questions to locate the page, but then you may need 9 questions to locate the word.)

6. Jones has two parents, each of his or her parents had two parents, each of these had two parents, etc. Tracing a person's family tree back 40 generations (about 1000 years) gives Jones 2^{40} ancestors, which is more people than have been on the earth in the last 1000 years. What is wrong with this argument?

The number of elements in a set

The remainder of this chapter will be devoted to certain counting problems. For any set X we shall denote by $n(X)$ the number of elements in the set.

Suppose we know the number of elements in certain given sets and wish to know the number in other sets related to these by the operations of unions, intersections, and complementations. As an example, consider the following problem.

Suppose that we are told that 100 students take mathematics, and 150 students take economics. Can we then tell how many take either mathematics or economics? The answer is no, since clearly we would also need to know how many students take both courses. If we know that no student takes both courses, i.e., if we know that the two sets of students are disjoint, then the answer would be the sum of the two numbers or 250 students.

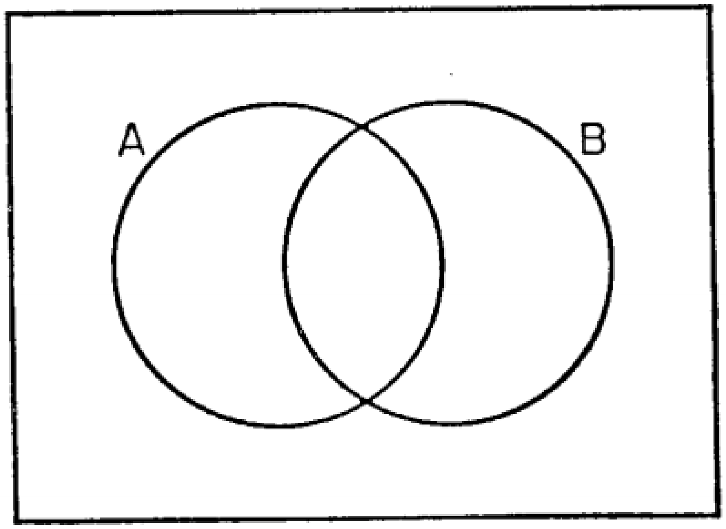

Figure 3.1: ◊

In general, if we are given disjoint sets A and B, then it is true that $n(A \cup B) = n(A) + n(B)$. Suppose now that A and B are not disjoint as shown in Figure 3.1. We can divide the set A into disjoint sets $A \cap \tilde{B}$ and $A \cap B$. Similarly, we can divide B into the disjoint sets $\tilde{A} \cap B$ and $A \cap B$. Thus,

$$n(A) = n(A \cap \tilde{B}) + n(A \cap B),$$

$$n(B) = n(\tilde{A} \cap B) + n(A \cap B).$$

Adding these two equations, we obtain

$$n(A) + (B) = n(A \cap B) + n(A \cap \tilde{B}) + n(\tilde{A} \cap B) + 2n(A \cap B).$$

Since the sets $A \cap B$, $A \cap \tilde{B}$, and $\tilde{A} \cap B$ are disjoint sets whose union is $A \cup B$, we obtain the formula

$$n(A \cup B) = n(A) + n(B) - n(A \cap B),$$

which is valid for any two sets A and B.

Example 3.4 Let p and q be statements relative to a set $\mathcal{U}$ of logical possibilities. Denote by P and Q the truth sets of these statements. The truth set of $p \vee q$ is $P \cup Q$ and the truth set of $p \wedge q$ is $P \cap Q$. Thus the above formula enables us to find the number of cases where $p \vee q$ is true if we know the number of cases for which p, q, and $p \wedge q$ are true. ◇

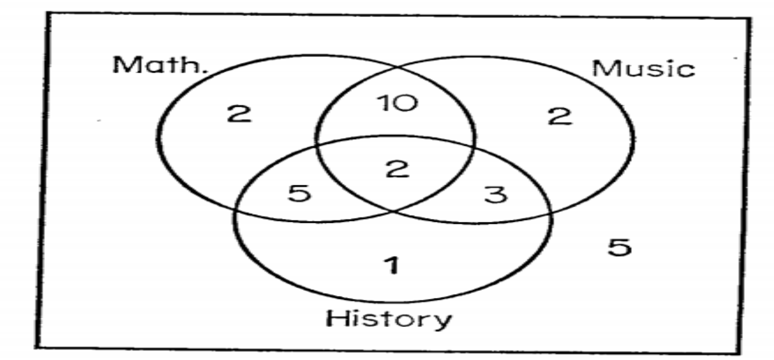

Figure 3.2: ◇

Example 3.5 More than two sets. It is possible to derive formulas for the number of elements in a set which is the union of more than two sets (see Exercise 6), but usually it is easier to work with Venn diagrams. For example, suppose that the registrar of a school reports the following statistics about a group of 30 students: 19 take mathematics. 17 take music. 11 take history. 12 take mathematics and music. 7 take history and mathematics. 5 take music and history. 2 take mathematics, history, and music. We draw the Venn diagram in Figure 3.2 and fill in the numbers for the number of elements in each subset working from the bottom of our list to the top. That is, since 2 students take all three courses, and 5 take music and history, then 3 take history and music but not mathematics, etc. Once the diagram is completed we can read off the number which take any combination of the courses. For example, the number which take history but not mathematics is $3 + 1 = 4$. ◇

Example 3.6 Cancer studies. The following reasoning is often found in statistical studies on the effect of smoking on the incidence of lung cancer. Suppose a study has shown that the fraction of smokers among those who have lung cancer is greater than the fraction of smokers among those who do not have lung cancer. It is then asserted that the fraction of smokers who have lung cancer is greater than the fraction of nonsmokers who have lung cancer. Let us examine this argument.

Let S be the set of all smokers in the population, and C be the

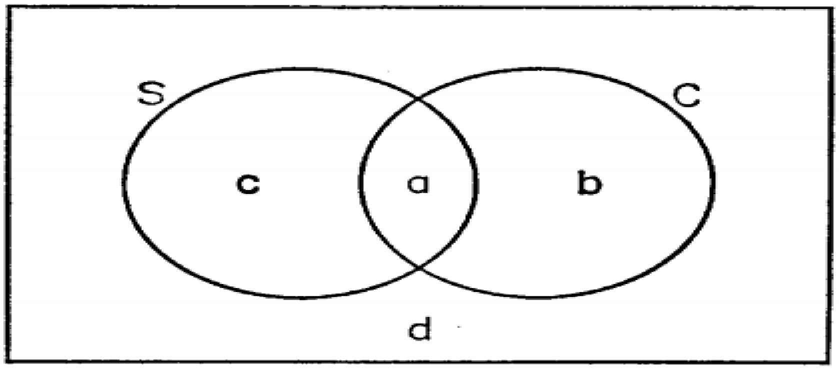

Figure 3.3: ◇

set of all people with lung cancer. Let $a = n(S \cap C)$, $b = n(\tilde{S} \cap C)$, $c = n(S \cap \tilde{C})$ and $d = n(\tilde{S} \cap \tilde{C})$, as indicated in Figure 3.3. The fractions in which we are interested are

$$p_1 = \frac{a}{a+b}, p_2 = \frac{c}{c+d}, p_3 = \frac{a}{a+c}, p_4 = \frac{b}{b+d},$$

where p_1 is the fraction of those with lung cancer that smoke, p_2 the fraction of those without lung cancer that smoke, p_3 the fraction of smokers who have lung cancer, and p_4 the fraction of nonsmokers who have cancer.

The argument above states that if $p_1 > p_2$, then $p_3 > p_4$. The hypothesis,

$$\frac{a}{a+b} > \frac{c}{c+d}$$

is true if and only if $ac + ad > ac + bc$, that is, if and only if $ad > bc$. The conclusion

$$\frac{a}{a+c} > \frac{b}{b+d}$$

is true if and only if $ab + ad > ab + bc$, that is, if and only if $ad > bc$. Thus the two statements $p_1 > p_2$ and $p_3 > p_4$ are in fact equivalent statements, so that the argument is valid. $\diamondsuit$

Permutations

We wish to consider here the number of ways in which a group of n different objects can be arranged. A listing of n different objects in a certain order is called a *permutation* of the n objects. We consider first the case of three objects, a, b, and c. We can exhibit all possible permutations of these three objects as paths of a tree, as shown in Figure 3.4. Each path exhibits a possible permutation, and there are six such paths. We could also list these permutations as follows: abc, bca, acb, cab, bac, cba. If we were to construct a similar tree for n objects, we would find that the number of paths could be found by multiplying together the numbers n, $n-1$, $n-2$, continuing down to the number 1. The number obtained in this way occurs so often that we give it a symbol, namely $n!$, which is read "n factorial". Thus, for example, $3! = 3 \cdot 2 \cdot 1 = 6$, $4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$, etc. For reasons which will be clear later, we define $0! = 1$. Thus we can say *there are* $n!$ *different permutations of* n *distinct objects.*

Example 3.7 In the game of Scrabble, suppose there are seven lettered blocks from which we try to form a seven-letter word. If the seven letters are all different, we must consider $7! = 5040$ different orders. $\diamondsuit$

Example 3.8 A quarterback has a sequence of ten plays. Suppose his or her coach instructs him or her to run through the ten-play sequence without repetition. How much freedom is left to the quarterback? He or she may choose any one of $10! = 3,628,800$ orders in which to call the plays. $\diamondsuit$

Example 3.9 How many ways can n people be seated around a circular table? When this question is asked, it is usually understood that two arrangements are different only if at least one person has a different person on the right in the two arrangements. Consider then one person in a fixed position. There are $(n-1)!$ ways in which the other people may be seated. We have now counted all the arrangements we wish to consider different. $\diamondsuit$

A general principle. There are many counting problems for which it is not possible to give a simple formula for the number of possible cases. In many of these the only way to find the number of cases is to draw a tree and count them (see Exercise 4). In some problems, the following general principle is useful.

If one thing can be done in exactly r different ways, for each of these a second thing can be done in exactly s different ways, for each of the first two, a third can be done in exactly t ways, etc., then the sequence of things can be done in the product of the numbers of ways in which the individual things can be done, i.e., $r \cdot s \cdot t$ ways.

The validity of the above general principle can be established by thinking of a tree representing all the ways in which the sequence of things can be done. There would be r branches from the starting position. From the ends of each of these r branches there would be s new branches, and from each of these t new branches, etc. The number of paths through the tree would be given by the product $r \cdot s \cdot t$.

Example 3.10 The number of permutations of n distinct objects is a special case of this principle. If we were to list all the possible permutations, there would be n possibilities for the first, for each of these $n - 1$ for the second, etc., until we came to the last object, and for which there is only one possibility. Thus there are $n(n-1)\dots 1 = n!$ possibilities in all. ◇

Example 3.11 If there are three roads from city x to city y and two roads from city y to city z, then there are $3 \cdot 2 = 6$ ways that a person can drive from city x to city z passing through city y. ◇

Example 3.12 Suppose there are n applicants for a certain job. Three interviewers are asked independently to rank the applicants according to their suitability for the job. It is decided that an applicant will be hired if he or she is ranked first by at least two of the three interviewers.

What fraction of the possible reports would lead to the acceptance of some candidate? We shall solve this problem by finding the fraction of the reports which do not lead to an acceptance and subtract this answer from 1. Frequently, an indirect attack of this kind on a problem is easier than the direct approach. The total number of reports possible is $(n!)^3$ since each interviewer can rank the men in $n!$ different ways. If a particular report does not lead to the acceptance of a candidate, it must be true that each interviewer has put a different applicant in first place. This can be done in $n(n-1)(n-2)$ different ways by our general principle. For each possible first choices, there are $[(n-1)!]^3$ ways in which the remaining men can be ranked by the interviewers. Thus the number of reports which do not lead to acceptance is $n(n-1)(n-2)[(n-1)!]^3$. Dividing this number by $(n!)^3$ we obtain

$$\frac{(n-1)(n-2)}{n^2}$$

as the fraction of reports which fail to accept a candidate. The fraction which leads to acceptance is found by subtracting this fraction from 1 which gives

$$\frac{3n-2}{n^2}$$

For the case of three applicants, we see that $\frac{7}{9}$ of the possibilities lead to acceptance. Here the procedure might be criticized on the grounds that even if the interviewers are completely ineffective and are essentially guessing there is a good chance that a candidate will be accepted on the basis of the reports. For n equal to ten, the fraction of acceptances is only .28, so that it is possible to attach more significance to the interviewers ratings, if they reach a decision. $\diamondsuit$

Exercises

1. In how many ways can five people be lined up in a row for a group picture? In how many ways if it is desired to have three in the front row and two in the back row?

   [Ans. 120;120.]

2. Assuming that a baseball team is determined by the players and the position each is playing, how many teams can be made from 13 players if

   (a) Each player can play any position?

   (b) Two of the players can be used only as pitchers?

3. Grades of A, B, C, D, or F are assigned to a class of five students.

   (a) How many ways may this be done, if no two students receive the same grade?

   [Ans. 120.]

   (b) Two of the students are named Smith and Jones. How many ways can grades be assigned if no two students receive the same grade and Smith must receive a higher grade than Jones?

   [Ans. 60.]

   (c) How many ways may grades be assigned if only grades of A and F are assigned?

   [Ans. 32.]

4. A certain club wishes to admit seven new members, four of whom are Republicans and three of whom are Democrats. Suppose the club wishes to admit them one at a time and in such a way that there are always more Republicans among the new members than there are Democrats. Draw a tree to represent all possible ways in which new members can be admitted, distinguishing members by their party only.

5. There are three different routes connecting city A to city B. How many ways can a round trip be made from A to B and back? How many ways if it is desired to take a different route on the way back?

[Ans. 9;6.]

6. How many different ways can a ten-question multiple-choice exam be answered if each question has three possibilities, a, b, and c? How many if no two consecutive answers are the same?

7. Modify Example 3.12 so that, to be accepted, an applicant must be first in two of the interviewers' ratings and must be either first or second in the third interviewers' rating. What fraction of the possible reports lead to acceptance in the case of three applicants? In the case of n?

Supplementary exercises.

. (a) How many four digit numbers can be formed from the digits 1, 2, 3, 4, using each digit only once?

(b) How many of these numbers are less than 3000?

[Ans. 12.]