Section 6.3: Further Rules for Counting Sets

Often when we are considering the probability of an event, that event is itself a union of other events. For example, suppose there is a horse race with three different horses, x, y, and z, and we want to know the probability that either x or y will win. This can also be interpreted as the probability that either x or y will win, so is in fact a combination of the two different events "x wins" and "y wins". In this section, we consider how to determine such probabilities. As we have already noted, calculation of probabilities comes down to calculation of the size of subsets of a sample space, so our main task is to develop a method to count such sets.

1. The Addition Rule

We start with a rule which will allow us to count the size of a union of disjoint sets. Following this, we shall consider the more complicated case where we may have a union of non-disjoint sets. Recall that if A is a set, then $N(A)$ denotes its size.

Theorem 1.1. *Suppose a finite set A is equal to the union of k distinct mutually disjoint subsets $A_1, \ldots, A_k$. Then*

$$N(A) = N(A_1) + \cdots + N(A_k).$$

Corollary 1.2. *If S is a sample space of equally likely outcomes and E is an event which is the union of mutually disjoint events $E_1, \ldots, E_k$, then*

$$P(E) = P(E_1) + \cdots + P(E_k).$$

We illustrate with an example.

Example 1.3. How many digits from 1 through 999 have no repeated digits? What is the probability that a random chosen number from 1 to 999 has a repeated digit?

We can solve this problem by breaking up the set of integers A from 1 to 999 with non-repeated entries into three mutually distinct subset - A_1 which is the set of integers with no repeated entries with 1 digit, A_2 which is the set of integers with no repeated entries with 2 digits and A_3 which is the set of integers with no repeated entries with 3 digits. We now calculate the size of each of these sets individually.

- $N(A_1) = 9$ since no integer with one digit can possible have a repeated digit.
- $N(A_2) = 9 * 9 = 81$, since there are 9 choices for the first digit (any of the integers 1 through 9), and 9 choices for the second digit (the remaining 8 numbers which were not chosen for the first digit and the number 0).

- $N(A_3) = 9 * 9 * 8 = 648$ since there are 9 choices for the first digit (any of the integers 1 through 9), 9 choices for the second digit (the remaining 8 numbers which were not chosen for the first digit and the number 0) and 8 choices for the last digit (any of the numbers 0 through 9 which did not appear in the first two digits).

Thus $N(A) = N(A_1)+N(A_2)+N(A_3) = 9+81+648 = 738$. It follows that there are 261 digits with at least one repeated entry. Thus the probability that a random number from 1 to 999 has a repeated digit is

$$P = \frac{261}{999} = 0.26.$$

2. The Difference Rule

Next we consider a rule which will allow us to calculate the size of a difference of sets assuming that one set is contained in the other. Though the rule is fairly obvious, we shall prove it using our previous observations.

Theorem 2.1. *Suppose a finite set A and B are sets and $B \subseteq A$. Then*

$$N(A - B) = N(A) - N(B)$$

Consequently, if S is a sample space of equally likely outcomes and A and B are events with $B \subseteq A$, then the probability of A occurring but not B is

$$P(A - B) = P(A) - P(B).$$

Proof. Note that the sets $A - B$ and B will be mutually disjoint and $(A - B) \cup B = A$ since $B \subset A$. Therefore, using the addition rule,

$$N(A - B) + N(B) = N(A) \text{ or } N(A - B) = N(A) - N(B)$$

□

The following result which specifies how to calculate the probability of a given event not occurring is a corollary of this theorem.

Corollary 2.2. *If S is a sample space of equally likely outcomes and E is an event then*

$$P(E^c) = 1 - P(E).$$

Proof. Using the difference formula since E^c and E are disjoint, we have

$$N(U - E) = N(E^c) = N(U) - N(E).$$

Then we have

$$P(E^c) = \frac{N(U) - N(E)}{N(U)} = 1 - P(E).$$

□

We illustrate with some examples, the first of which is similar to the previous example.

Example 2.3. How many numbers between 1 and 999 have repeated digits?

Let A be the set of numbers without repeated digits. Then the set of numbers with repeated digits is A^c. We know from our previous example that the number of numbers from 1 to 999 without repeated digits is equal to 738. Therefore, if B is the set of numbers from 1 to 999 with repeated digits, we have

$$N(B) = N(A^c) = N(U) - N(A) = 999 - 738 = 261.$$

Example 2.4. How many seven digit telephone numbers are there without any repeated digits? How many with at least one repeated digit? What is the probability that a randomly chosen seven digit number will have no repeated digits?

Let U be the set of all seven digit numbers, let A be the subset of numbers with no repeated digits and let $B = A^c$ be the subset of numbers with repeated digits. We can calculate $N(A)$ using the multiplication rule. Specifically, we have 10 choices for the first digit, 9 for the second and so on. Thus $N(A) = 10 * 9 * 8 * 7 * 6 * 5 * 4 = 604,800$. Since $N(U) = 10^7$, we have

$$N(B) = N(A^c) = 10^7 - 604,800 = 9,395,200.$$

Finally, the probability that no digits will be repeated in a seven digit number is

$$P(A) = \frac{604,800}{10^7} = 0.061.$$

3. The Inclusion/Exclusion Rule

Finally we consider a rule which will allow us to calculate the size of a union of sets in general without any assumptions of intersections. In order to prove the formula we get, we shall use the results we have already considered.

Theorem 3.1. *(The Inclusion/Exclusion Rule) Suppose a finite set A and B are sets. Then*

$$N(A \cup B) = N(A) + N(B) - N(A \cap B)$$

Consequently, if S is a sample space of equally likely outcomes and A and B are events then the probability of either A or B occurring is

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Proof. Note that the set $A \cup B$ is the union of disjoint sets $A - B$ and B. Therefore, using the addition rule, we have

$$N(A \cup B) = N(A - B) + N(B).$$

It is easy to prove that $A-B = A-(A\cap B)$ and we also have $A \supseteq A\cap B$. Therefore, using the difference rule, we have

$$N(A \cup B) = N(A - B) + N(B) = N(A - A \cap B) + N(B)$$

$$= N(A) - N(A \cap B) + N(B) = N(A) + N(B) - N(A \cap B).$$

□

In fact, the inclusion/exclusion rule holds more generally for the union of any number of sets. We state the rule for three sets below, and it should be clear how this generalizes for more sets (which can be proved using induction).

Corollary 3.2. *(The Inclusion/Exclusion Rule) Suppose a finite set A, B, and C are sets. Then*

$$N(A \cup B \cup C) =$$

$$N(A)+N(B)+N(C)-N(A\cap B)-N(A\cap C)-N(B\cap C)+N(A\cap B\cap C)$$

We illustrate with some examples.

Example 3.3. How many integers from 1 through 1000 are multiples of 4 or 7? Determine the probability that a randomly chosen integer between 1 and 1000 is a multiple of 4 or 7.

Let A denote the set of integers between 1 and 1000 which are multiples of 4 and B the set of integers between 1 and 1000 which are multiples of 7. We shall calculate $N(A)$ and $N(B)$ using the counting methods we previoulsy considered. First, lining up the integers from 1 to 1000, and the corresponding multiples of 4, we have the following:

$$\begin{array}{cccccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & \cdots 997 & 998 & 999 & 1000 \\ & & 4\cdot 1 & & & & & 4\cdot 2 & & \cdots & & & 4\cdot 250 \end{array}$$

Therefore, there are $250 - 1 + 1 = 250$ integers which are multiples of 4. Likewise, for 7, we have

$$\begin{array}{ccccccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & \cdots & 994 & 995 & 996 & 997 & 998 & 999 & 1000 \\ & & & & & & 7\cdot 1 & \cdots & 7\cdot 142 & & & & & & \end{array}$$

Therefore, there are $142 - 1 + 1 = 142$ integers which are multiples of 7. The set of integers which are either multiples of 4 or 7 is the set $A\cup B$. Since $A\cap B$ is not empty, we cannot apply the addition rule and must use the inclusion/exclusion principle. In particular, to calculate $N(A\cup B)$, we need to determine $N(A\cap B)$. Note that $A\cap B$ is the set of all integers divisible by both 4 and 7, so will be the set of integers divisible by $28 = 4 \cdot 7$. We calculate the size of this set in the usual manner:

$$\begin{array}{cccccccccccc} 1 & \cdots & 27 & 28 & 29 & \cdots & 979 & 980 & 981 & \cdots & 999 & 1000 \\ & \cdot & & 1\cdot 28 & & \cdots & & 35\cdot 28 & \cdots & & & \end{array}$$

Therefore, there are $35 - 1 + 1 = 35$ integers which are multiples of 28. Therefore, using inclusion/exclusion, we have

$$N(A \cup B) = N(A) + N(B) - N(A \cap B) = 250 + 142 - 28 = 364$$

Thus there are 364 integers between 1 and 1000 which are divisible by either 4 or 7. This means that the probability of a randomly chosen integer between 1 and 1000 being divisible by 4 or 7 is

$$P(A \cup B) = \frac{364}{1000} = 0.364.$$

Example 3.4. (The Birthday Problem) Assuming all years have 365 days, and every day is an equally likely birthday, how large must n be so that in any randomly chosen group of n people, the probability that two or more have the same birthday is at least 1/2.

If there are a total of n people in a room, we can form a row of length n which contains each persons birthday. With this in mind, we let the sample space S be the set of all rows of length n of all possible birthdays. Since there are 365 days in the year, the size of this sample space will be 365^n. Next we shall calculate the total number of rows which do not have any repeated entries. Since the row has length n, there will be 365 choices for the first, 364 choices for the second (since it cannot be the same as the first). In particular, of the 365^n total rows,

$$365 \cdot 364 \cdots (365 - n + 1) = \frac{365!}{(365 - n)!}$$

will have no repeated entries. This means the number of rows with at least one repeated entry will be

$$365^n - \frac{365!}{(365 - n)!}.$$

Therefore, the probability of at least two people having the same birthday will be

$$P = 1 - \frac{365!}{(365 - n)! \cdot 365^n}$$

Setting $P > 1/2$, we get

$$1 - \frac{365!}{(365 - n)! \cdot 365^n} > \frac{1}{2} \text{ or } \frac{1}{2} > \frac{365!}{(365 - n)! \cdot 365^n}$$

These numbers are too big to solve by hand, but using a calculator, we get $n > 22$. This means that if a room contains at least 23 people, there is over a fifty percent chance that two of them have the same birthday!

Homework

(i) From the book, pages 330-333: Questions: 1, 5, 6, 10, 11, 17, 18, 22, 24, 27, 30