Topic 5: Basics of Probability*

September 13-15, 2011

1 Introduction

Mathematical structures like Euclidean geometry or algebraic fields are defined by a set of axioms. "Mathematical reality" is then developed through the introduction of concepts and the proofs of theorems. These axioms are inspired, in the instances introduced above, by our intuitive understanding, for example, of the nature of parallel lines or the real numbers. Probability is a branch of mathematics based on three axioms inspired originally by calculating chances from card and dice games.

Statistics, in its role as a facilitator of science, begins with the collection of data. From this collection, we are asked to make inference on the **state of nature**, that is to determine the conditions that are likely to produce these data. Probability, in undertaking the task of investigating differing states of nature, takes the complementary perspective. It begins by examining **random** phenomena, i.e., those whose exact outcomes are uncertain. Consequently, in order to determine the "scientific reality" behind the data, we must spend some time working with the concepts of the theory of probability to investigate properties of the data arising from the possible states of nature to assess which are most useful in making inference.

We will motivate the axioms of probability through the case of equally likely outcomes for some simple games of chance and look at some of the direct consequences of the axioms. In order to extend our ability to use the axioms, we will learn counting techniques, e.g, permutations and combinations, based on the multiplication principle.

A **probability model** has two essential pieces of its description.

- Ω, the **sample space**, the set of possible outcomes.
  - An **event** is a collection of **outcomes**. We can give explicitly define and event via its outcomes,

    $$A = \{\omega_1, \omega_2, \cdots, \omega_n\}$$

    or with a description

    $$A = \{\omega : \omega \text{ has property } \mathcal{P}\}.$$

    In either case, A is subset of the sample space, $A \subset \Omega$.
- P, the **probability** assigns a number to each event.

Thus, a probability is a function. We are familiar with functions in which both the domain and range are subsets of the real numbers. The domain of a probability function is the collection of all possible outcomes. The range is still a number. We will see soon which numbers we will accept as probabilities of events.

You may recognize these concepts from a basic introduction to sets. In talking about sets, we use the term **universal set** instead of sample space, **element** instead of outcome, and **subset** instead of event. At first, having two words for the same concept seems unnecessarily redundant. However, we will later consider circumstances in which we will want to consider a more complex situation which will combine ideas from sets and from probability. In these cases, having two expression for a concept will facilitate our understanding. A *Set Theory - Probability Theory Dictionary* is included at the end of this section to relate to the new probability terms with the more familiar set theory terms.

2 Equally Likely Outcomes

The essential relationship between events and the probability are described through the three **axioms of probability**. These axioms can be motivated through the first uses of probability, namely the case of equal likely outcomes.

If Ω is a finite sample space, then if each outcome is equally likely, we define the probability of A as the fraction of outcomes that are in A.

$$P(A) = \frac{\#(A)}{\#(\Omega)}.$$

Thus, computing $P(A)$ means counting the number of outcomes in the event A and the number of outcomes in the sample space Ω and dividing.

Exercise 1. *(a) Toss a coin.*

$$P\{\text{heads}\} = \frac{\#(A)}{\#(\Omega)} = \underline{\qquad}.$$

(b) Toss a coin three times.

$$P\{\text{toss at least two heads in a row}\} = \frac{\#(A)}{\#(\Omega)} = \underline{\qquad}$$

(c) Roll two dice.

$$P\{\text{sum is 7}\} = \frac{\#(A)}{\#(\Omega)} = \underline{\qquad}$$

Because we always have $0 \le \#(A) \le \#(\Omega)$, we always have

$$P(A) \ge 0 \qquad (1)$$

and

$$P(\Omega) = 1 \qquad (2)$$

This gives us 2 of the three axioms. The third will require more development.

Toss a coin 4 times. $\#(\Omega) = 16$
A = {exactly 3 heads} = {HHHT, HHTH, HTHH, THHH} $\#(A) = 4$

$$P(A) = \frac{4}{16} = \frac{1}{4}$$

B = {exactly 4 heads} = {HHHH} $\#(B) = 1$

$$P(B) = \frac{1}{16}$$

Now let's define the set C = {at least three heads}. If you are asked the supply the probability of C, your intuition is likely to give you an immediate answer.

$$P(C) = \underline{\qquad}.$$

Let's have a look at this intuition. The events A and B have no outcomes in common,. We say that the two events are **disjoint** or **mutually exclusive** and write $A \cap B = \emptyset$. In this situation,

$$\#(A \cup B) = \#(A) + \#(B).$$

If we take this **addition principle** and divide by $\#(\Omega)$, then we obtain the following identity:

If $A \cap B = \emptyset$, then

$$\frac{\#(A \cup B)}{\#(\Omega)} = \frac{\#(A)}{\#(\Omega)} + \frac{\#(B)}{\#(\Omega)}.$$

or

$$P(A \cup B) = P(A) + P(B). \tag{3}$$

Using this property, we see that

$$P\{\text{at least 3 heads}\} = P\{\text{exactly 3 heads}\} + P\{\text{exactly 4 heads}\} = \frac{4}{16} + \frac{1}{16} = \frac{5}{16}.$$

We are saying that any function P that accepts events as its domain and returns numbers as its range and satisfies (1), (2), and (3) can be called a **probability**.

If we iterate the procedure in Axiom 3, we can also state that if the events, $A_1, A_2, \cdots, A_n$, are mutually exclusive, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n). \tag{3'}$$

This is a sufficient definition for a probability if the sample space Ω is finite. However, we will want to examine infinite sample spaces and to use the idea of limits. This introduction of limits is the pathway that allows to bring in calculus with all of its powerful theory and techniques.

Example 2. *For the random experiment, flip a coin repeatedly until heads appears, we can write*

$$A_j = \{\text{the first head appears on the } j\text{-th toss}\}.$$

Then, each of the A_j are mutually exclusive and

$$\{\text{heads appears eventually}\} = A_1 \cup A_2 \cup \cdots \cup A_n \cup \cdots = \bigcup_{j=1}^{\infty} A_j = \{\omega; \omega \in A_j \text{ for some } j\}.$$

We would like to say that

$$P\{\text{heads appears eventually}\} = P(A_1) + P(A_2) + \cdots + P(A_n) + \cdots = \sum_{j=1}^{\infty} P(A_j) = \lim_{n\to\infty} \sum_{j=1}^{n} P(A_j).$$

This would call for an extension of Axiom 3 to an infinite number of mutually exclusive events. This is the general version of Axiom 3 we use when we want to use calculus in the theory of probability:

For mutually exclusive events, $\{A_j; j \geq 1\}$, then

$$P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j) \tag{3''}$$

Thus, statements (1), (2), and (3”) give us the complete axioms of probability.

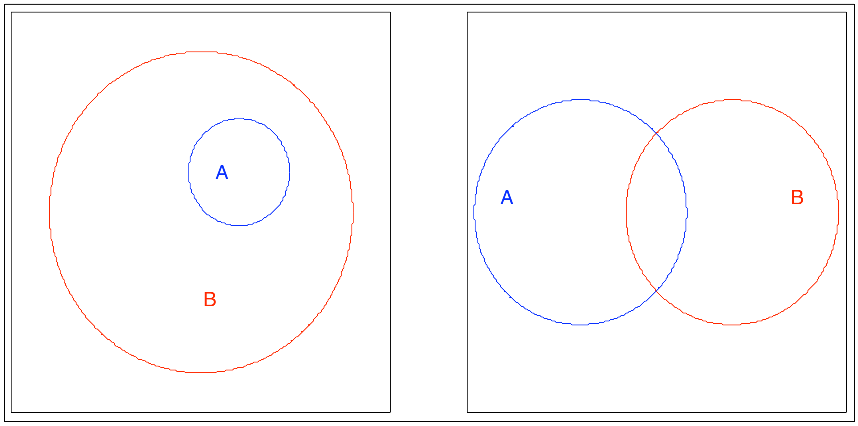

Figure 1: (left) **Difference** and **Monotonicity Rule**. If $A \subset B$, then $P(B \setminus A) = P(B) - P(A)$). (right) **The Inclusion-Exclusion Rule**. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Using area as an analogy for probability, $P(B \setminus A)$ is the area between the circles and the area $P(A) + P(B)$ double counts the lens area $P(A \cap B)$.

3 Consequences of the Axioms

Other properties that we associate with a probability can be derived from the axioms.

1. **The Complement Rule**. Because A and its **complement** $A^c = \{\omega; \omega \notin A\}$ are mutually exclusive

$$P(A) + P(A^c) = P(A \cup A^c) = P(\Omega) = 1$$

or

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

For example, if we toss a *biased* coin. We may want to say that $P\{\text{heads}\} = p$ where p is not necessarily equal to 1/2. By necessity,

$$P\{\text{tails}\} = 1 - p.$$

Toss a coin 4 times.

$$P\{\text{fewer than 3 heads}\} = 1 - P\{\text{at least 3 heads}\} = 1 - \frac{5}{16} = \frac{11}{16}.$$

2. **The Difference Rule** Write $B \setminus A$ to denote the outcomes that are in B but **not** in A. If $A \subset B$, then

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

(A and $B \setminus A$ are mutually exclusive and their union is B. Thus $P(B) = P(A) + P(B \setminus A)$.) See Figure 2 (left).

Exercise 3. *Give an example for which* $P(B \setminus A) \neq P(B) - P(A)$

Because $P(B \setminus A) \geq 0$, we have the following:

3. **Monotonicity Rule** If $A \subset B$, then $P(A) \leq P(B)$

We already know that for any event A, $P(A) \geq 0$. The monotonicity rule adds to this the fact that

$$P(A) \leq P(\Omega) = 1.$$

This, the range of a probability is a subset of the interval $[0, 1]$.

4. **The Inclusion-Exclusion Rule**. For any two events A and B,
$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad (4).$$
($P(A) + P(B)$ accounts for the outcomes in $A \cap B$ twice, so remove $P(A \cap B)$.) See Figure 2 (right).

   Exercise 4. *Show that the inclusion-exclusion rule follows from the axioms. Hint:* $A \cup B = (A \cap B^c) \cup B$ *and* $A = (A \cap B^c) \cup (A \cap B)$.

   Exercise 5. *Give a generalization of the inclusion-exclusion rule for three events.*

   Deal two cards.
   A = {ace on the second card}, B = {ace on the first card}
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$P\{\text{at least one ace}\} = \frac{1}{13} + \frac{1}{13} - \quad ?$$
   To complete this computation, we will need to compute $P(A \cap B) = P\{\text{both cards are aces}\}$.

5. **The Bonferroni Inequality.** For any two events A and B,
$$P(A \cup B) \le P(A) + P(B).$$

6. **Continuity Property.** Use the symbol $\subset$ to denote "contains in". If events satisfy
$$B_1 \subset B_2 \subset \cdots \text{ and } B = \bigcup_{i=1}^{\infty} B_i$$
   Then, by the monotonicity rule, $P(B_i)$ is an increasing sequence satisfying
$$P(B) = \lim_{i \to \infty} P(B_i). \qquad (5)$$
   Similarly, use the symbol $\supset$ to denote "contains". If events satisfy
$$C_1 \supset C_2 \supset \cdots \text{ and } C = \bigcap_{i=1}^{\infty} C_i$$
   Again, by the monotonicity rule, $P(C_i)$ is an decreasing sequence satisfying
$$P(C) = \lim_{i \to \infty} P(C_i). \qquad (6)$$

Exercise 6. *Establish the continuity property.* Hint: *For the first, let* $A_1 = B_1$ *and* $A_i = B_i \setminus B_{i-1},\ i > 1$ *in axiom (3"). For the second, use the complement rule and* **de Morgan's law**
$$C^c = \bigcup_{i=1}^{\infty} C_i^c$$

Exercise 7 (odds)**.** *The statement of* $a : b$ *odds for an event* A *indicates that*
$$\frac{P(A)}{P(A^c)} = \frac{a}{b}$$
Show that
$$P(A) = \frac{a}{a+b}.$$
So, for example, $1 : 2$ *odds means* $P(A) = 1/3$ *and* $5 : 3$ *odds means* $P(A) = 5/8$.

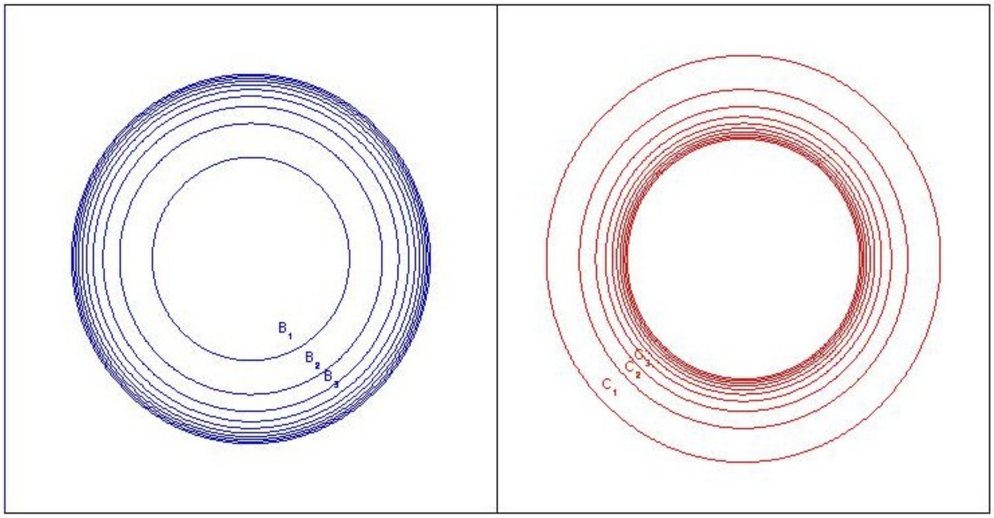

Figure 2: Continuity Property. (left) B_i increasing to an event B. Here, equation (5) is satisfied. (right) C_i decreasing to an event C. Here, equation (6) is satisfied.

4 Counting

In the case of equally likely outcomes, finding the probability of an event A is the result of two counting problems - namely finding $\#(A)$, the number of outcomes in A and finding $\#(\Omega)$, the number of outcomes in the sample space. These counting problems can become quite challenging and advanced mathematical techniques have been developed to address these issues. However, having some facility in counting is necessary to have a sufficiently rich number of examples to give meaning to the axioms of probability. Consequently, we shall develop a few counting techniques leading to the concepts of **permutations** and **combinations**.

We start with the **multiplication principle**.

Suppose that two experiments are to be performed.

- Experiment 1 can have n_1 possible outcomes and
- for each outcome of experiment 1, experiment 2 has n_2 possible outcomes.

Then together there are $n_1 \times n_2$ possible outcomes.

Example 8. *For a group of n individuals, one is chosen to become the president and a second is chosen to become the treasurer. By the multiplication principle, if these position are held by different individuals, then this task can be accomplished in*

$$n \times (n-1)$$

possible ways

Exercise 9. *Generalize the multiplication principle of counting to k experiments.*

Assume that we have a collection of n objects and we wish to make an **ordered arrangement** of k of these objects. Using the generalized principle of counting, the number of possible outcomes is

$$n \times (n-1) \times \cdots \times (n-k+1).$$

We will write this as $(n)_k$ and say n **falling** k.

4.1 Permutations

Example 10 (birthday problem). *In a list the birthday of k people, there are 365^k possible lists (ignoring leap year births) and*

$$(365)_k$$

possible lists with no date written twice. Thus, the probability, under equally likely outcomes, that no two people on the list have the same birthday is

$$\frac{(365)_k}{365^k} = \frac{365 \cdot 364 \cdots (365-k+1)}{365^k}$$

and

$$P\{\text{at least one pair of individuals share a birthday}\} = 1 - \frac{(365)_k}{365^k}$$

Here is a short table of these probabilities. A graph is given in Figure 3.

k	5	10	15	18	20	22	23	25	30	40	50	100
probability	0.027	0.117	0.253	0.347	0.411	0.476	0.507	0.569	0.706	0.891	0.970	0.994

The R *code and output follows. We can create an iterative process by noting that*

$$\frac{(365)_k}{365^k} = \frac{(365)_{k-1}}{365^{k-1}} \frac{(365-k+1)}{365}$$

Thus, we can find the probability that no pair in a group of k individuals has the same birthday by taking the probability that no pair in a group of $k-1$ individuals has the same birthday and multiplying by $(365-k+1)/365$. Here is the output for $k = 1$ to 45.

```
> prob=rep(1,45)
> for (k in 2:45){prob[k]=prob[k-1]*(365-k+1)/365}
> data.frame(c(1:15),1-prob[1:15],c(16:30),1-prob[16:30],c(31:45),1-prob[31:45])
   c.1.15. X1...prob.1.15. c.16.30. X1...prob.16.30. c.31.45. X1...prob.31.45.
1        1     0.000000000       16        0.2836040       31        0.7304546
2        2     0.002739726       17        0.3150077       32        0.7533475
3        3     0.008204166       18        0.3469114       33        0.7749719
4        4     0.016355912       19        0.3791185       34        0.7953169
5        5     0.027135574       20        0.4114384       35        0.8143832
6        6     0.040462484       21        0.4436883       36        0.8321821
7        7     0.056235703       22        0.4756953       37        0.8487340
8        8     0.074335292       23        0.5072972       38        0.8640678
9        9     0.094623834       24        0.5383443       39        0.8782197
10      10     0.116948178       25        0.5686997       40        0.8912318
11      11     0.141141378       26        0.5982408       41        0.9031516
12      12     0.167024789       27        0.6268593       42        0.9140305
13      13     0.194410275       28        0.6544615       43        0.9239229
14      14     0.223102512       29        0.6809685       44        0.9328854
15      15     0.252901320       30        0.7063162       45        0.9409759
```

Definition 11. *The ordered arrangement of all n objects is*

$$(n)_n = n \times (n-1) \times \cdots \times 1 = n!,$$

n **factorial**. *We take* $0! = 1$.

Exercise 12.

$$(n)_k = \frac{n!}{(n-k)!}.$$

4.2 Combinations

Write

$$\binom{n}{k}$$

for the number of number of different groups of k objects that can be chosen from a collection of n.

We will next find a formula for this number by counting the number of possible outcomes in two different ways. To introduce this with a concrete example, suppose 3 cities will be chosen out of 8 under consideration for a vacation. If we think of the vacation as visiting three cities in a particular **order**, for example,

New York then Boston then Montreal.

Then we are looking at permutations. This results in

$$(8)_3 = 8 \cdot 7 \cdot 6$$

choices.

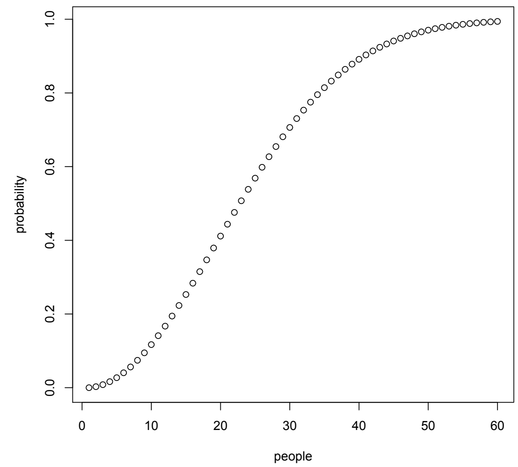

Figure 3: The Birthday Problem. For a room of containing n individuals. Plot of $P_n\{$at least one pair of individuals share a birthday$\}$.

If we are just considering the 3 cities we visit, irrespective of order, then these **unordered** choices are combinations. The number of ways of doing this is written

$$\binom{8}{3},$$

a number that we do not yet know how to determine. After we have chosen the three cities, we will also have to also pick an order to see the cities and so using the multiplication principle, we have

$$\binom{8}{3} \times 3 \cdot 2 \cdot 1 = \binom{8}{3} 3!$$

possible vacations if the order of the cities is included in the choice.

These two strategies are counting the same possible outcomes and so must be equal.

$$(8)_3 = 8 \cdot 7 \cdot 6 = \binom{8}{3} \times 3 \cdot 2 \cdot 1 = \binom{8}{3} 3! \quad \text{or} \quad \binom{8}{3} = \frac{8 \cdot 7 \cdot 6}{3 \cdot 2 \cdot 1} = \frac{(8)_3}{3!}.$$

Thus, we have a formula for $\binom{8}{3}$. Let's do this more generally.

Theorem 13.

$$\binom{n}{k} = \frac{(n)_k}{k!} = \frac{n!}{k!(n-k)!}.$$

The second equality follows from the previous exercise.

The number of ordered arrangements of k objects out of n is

$$(n)_k = n \times (n-2) \times \cdots \times (n-k+1).$$

Alternatively, we can form an ordered arrangement of k objects from a collection of n by:

1. First choosing a group of k objects.
   The number of possible outcomes for this experiment is $\binom{n}{k}$.

2. Then, arranging this k objects in order.
   The number of possible outcomes for this experiment is $k!$.

So, by the multiplication principle,

$$(n)_k = \binom{n}{k} \times k!.$$

Now complete the argument by dividing both sides by $k!$.

Exercise 14 (binomial theorem)**.**

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

Exercise 15. $\binom{n}{1} = \binom{n}{n-1} = n$. $\binom{n}{k} = \binom{n}{n-k}$. *Thus, we set* $\binom{n}{n} = \binom{n}{0} = 1$

The number of combinations is computed in R using `choose`. In the vacation example above, $\binom{8}{3}$ is determined by entering

```
> choose(8,3)
[1] 56
```

Theorem 16 (Pascal's triangle)**.**

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

To establish this identity, distinguish one of the n objects in the collection. Say that we are looking at a collection of n marbles, $n-1$ are blue and 1 is red.

1. For outcomes in which the red marble is chosen, we must choose $k-1$ marbles from the $n-1$ blue marbles. Thus, $\binom{n-1}{k-1}$ different outcomes have the red marble.

2. If the red marble is not chosen, then we must choose k blue marbles. Thus, $\binom{n-1}{k}$ outcomes do not have the red marbles.

3. These choices of groups of k marbles have no overlap. And so $\binom{n}{k}$ is the sum of the values in 1 and 2.

To see this using the example above,

$$\binom{8}{3} = \binom{7}{2} + \binom{7}{3}.$$

Assume that one of the 8 cities on the list includes New York. Then of the $\binom{8}{3}$ vacations, $\binom{7}{2}$ include New York and $\binom{7}{3}$ do not.

This gives us an iterative way to compute the values of $\binom{n}{k}$. Build a table of values for n (vertically) and $k \leq n$ (horizontally). Then, by the Pascal's triangle formula, a given table entry is the sum of the number directly above it and the number above and one column to the left. We can get started by noting that $\binom{n}{0} = \binom{n}{n} = 1$.

Pascal's triangle

	$k-1$	k	
$n-1$	$\binom{n-1}{k-1}$	$\binom{n-1}{k}$	← the sum of these two numbers
n		$\binom{n}{k}$	← equals this number

		k								
		0	1	2	3	4	5	6	7	8
	0	1								
	1	1	1							
	2	1	2	1						
	3	1	3	3	1					
n	4	1	4	6	4	1				
	5	1	5	10	10	5	1			
	6	1	6	15	20	15	6	1		
	7	1	7	21	35	35	21	7	1	
	8	1	8	28	56	70	56	28	8	1

Example 17. *For the experiment on honey bee queen - if we rear 60 of the 90 queen eggs, the we have*

```
> choose(90,60)
[1] 6.73133e+23
```

more than 10^{23} *different possible simple random samples.*

Example 18. *Deal out three cards. There are*

$$\binom{52}{3}$$

possible outcomes. Let x be the number of hearts. Then we have chosen x hearts out of 13 *and* $3-x$ *cards that are not hearts out of the remaining 39. Thus, by the multiplication principle there are*

$$\binom{13}{x}\cdot\binom{39}{3-x}$$

possible outcomes.

The probability of x hearts is the ratio of these two numbers. To compute these numbers in R *for* $x = 0, 1, 2, 3$*, the possible values for x, we enter*

```
> x<-c(0:3)
> prob<-choose(13,x)*choose(39,3-x)/choose(52,3)
> data.frame(x,prob)
  x       prob
1 0 0.41352941
2 1 0.43588235
3 2 0.13764706
4 3 0.01294118
```

Notice that

```
> sum(prob)
[1] 1
```

Exercise 19. *Deal out 5 cards. Let x be the number of fours. What values can x take? Find the probability of x fours for each possible value.*

5 Answers to Selected Exercises

1. (a) 1/2, (b) 3/8, (c) 6/36 = 1/6

3. Toss a coin 6 times. Let $A = \{\text{at least 3 heads}\}$ and Let $B = \{\text{at least 3 tails}\}$. Then

$$P(A) = P(B) = \frac{42}{64} = \frac{21}{32}.$$

Thus, $P(B) - P(A) = 0$. However, the event

$$B \setminus A = \{\text{exactly 3 tails}\} = \{\text{exactly 3 heads}\}$$

and $P(B \setminus A) = 20/64 = 5/16 \neq 0$.

4. Using the hint, we have that

$$P(A \cup B) = P(A \cap B^c) + P(B)$$
$$P(A) = P(A \cap B^c) + P(A \cup B)$$

Subtract these two equations

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

Now add $P(A)$ to both sides of the equation to obtain (4).

5. Use the associativity property of unions to write $A \cup B \cup C = (A \cup B) \cup C$ and use (4), the inclusion-exclusion property for the 2 events $A \cup B$ and C and then to the 2 events A and B,

$$\begin{aligned} P((A \cup B) \cup C) &= P(A \cup B) + P(C) - P((A \cup B) \cap C) \\ &= (P(A) + P(B) - P(A \cap B)) + P(C) - P((A \cap C) \cup (B \cap C)) \end{aligned}$$

For the final expression, we use one of De Morgan's Laws. Now rearrange the other terms and apply inclusion-exlcusion to the final expression.

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) - P(A \cap B) + P(C) - P((A \cap C) \cup (B \cap C)) \\ &= P(A) + P(B) + P(C) - P(A \cap B) - (P(A \cap C) + P(B \cap C) + P((A \cap C) \cap (B \cap C))) \\ &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \end{aligned}$$

The last expression uses the identity $(A \cap C) \cap (B \cap C)) = A \cap B \cap C$.

6. Using the hint and writing $B_0 = \emptyset$, we have that $P(A_i) = P(B_i) - P(B_{i-1})$ and that

$$\bigcup_{i=1}^n B_i = \bigcup_{i=1}^n A_i$$

Because the A_i are disjoint, we have by (3')

$$\begin{aligned} P\left(\bigcup_{i=1}^n B_i\right) &= P\left(\bigcup_{i=1}^n A_i\right) \\ &= P(A_n) + P(A_{n-1}) + \cdots + P(A_2) + P(A_1) \\ &= (P(B_n) - P(B_{n-1})) + (P(B_{n-1}) - P(B_{n-2})) + \cdots + (P(B_2) - P(B_1)) + (P(B_1) - P(B_0)) \\ &= P(B_n) - (P(B_{n-1}) - (P(B_{n-1})) - P(B_{n-2})) + \cdots + P(B_2) - (P(B_1) - (P(B_1)) - P(\emptyset) \\ &= P(B_n) \end{aligned}$$

because all of the other terms cancel. This is an example of a **telescoping sum**. Now use (3") to obtain

$$P\left(\bigcup_{i=1}^\infty B_i\right) = \lim_{n\to\infty} P(B_n).$$

For the second part. Write $B_i = C_i^c$. Then, the B_i satisfy the required conditions and that $B = C^c$. Thus,

$$1 - P(C) = P(C^c) = \lim_{i\to\infty} P(C_i^c) = \lim_{i\to\infty} (1 - P(C_i)) = 1 - \lim_{i\to\infty} P(C_i)$$

and

$$P(C) = \lim_{i\to\infty} P(C_i)$$

7. If

$$\frac{a}{b} = \frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}.$$

Then,

$$a - aP(A) = bP(A), \quad a = (a+b)P(A), \quad P(A) = \frac{a}{a+b}.$$

9. Suppose that k experiments are to be performed and experiment i can have n_i possible outcomes irrespective of the outcomes on the other $k-1$ experiments. Then together there are $n_1 \times n_2 \times \cdots \times n_k$ possible outcomes.

12.

$$(n)_k = n \times (n-1) \times \cdots \times (n-k+1) \times \frac{(n-k)!}{(n-k)!} = \frac{n \times (n-1) \times \cdots \times (n-k+1)(n-k)!}{(n-k)!} = \frac{n!}{(n-k)!}.$$

14. Expansion of $(x+y)^n = (x+y) \times (x+y) \times \cdots \times (x+y)$ will result in 2^n terms. Each of the terms is achieved by one choice of x or y from each of the factors in the product $(x+y)^n$. Each one of these terms will thus be a result in n factors - some of them x and the rest of them y. For a given k from $0, 1, \ldots, n$, we will see choices that will result in k factors of x and $n-k$ factors of y, i. e., $x^k y^{n-k}$. The number of such choices is the combination

$$\binom{n}{k}$$

Add these terms together to obtain

$$\binom{n}{k} x^k y^{n-k}.$$

Next adding these values over the possible choices for k results in

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

15. The formulas are easy to work out. One way to consider $\binom{n}{1} = \binom{n}{n-1}$ is to note that $\binom{n}{1}$ is the number of ways to choose 1 out of a possible n. This is the same as $\binom{n}{n-1}$, the number of ways to exclude 1 out of a possible n. A similar reasoning gives $\binom{n}{k} = \binom{n}{n-k}$.

19. The possible values for x are 0, 1, 2, 3, and 4. When we have chosen x fours out of 4, we also have $5-x$ cards that are not fours out of the remaining 48. Thus, by the multiplication principle, the probability of x fours is

$$\frac{\binom{4}{x} \cdot \binom{52}{5-x}}{\binom{52}{5}}.$$

To compute the numerical values for the probability of x fours:

```
> x<-c(0:4)
> prob<-choose(4,x)*choose(48,5-x)/choose(52,5)
> sum(prob)
[1] 1
> data.frame(x,prob)
  x         prob
1 0 6.588420e-01
2 1 2.994736e-01
3 2 3.992982e-02
4 3 1.736079e-03
5 4 1.846893e-05
```

6 Set Theory - Probability Theory Dictionary

Event Language	Set Language	Set Notation
sample space	universal set	Ω
event	subset	$A, B, C, \cdots$
outcome	element	ω
impossible event	empty set	$\emptyset$
not A	A complement	A^c
A or B	A union B	$A \cup B$
A and B	A intersect B	$A \cap B$
A and B are mutually exclusive	A and B are disjoint	$A \cap B = \emptyset$
if A then B	A is a subset of B	$A \subset B$